International Journal of Interactive Multimedia and Artificial Intelligence

September 2014, Vol II, Number 7, ISSN: 1989-1660

You can never be overdressed or overeducated

Oscar Wilde

Special Issue on Multisensor user tracking and analytics to improve education and other application fields

http://www.ijimai.org

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE AND INTERACTIVE MULTIMEDIA
ISSN: 1989-1660 - VOL. II, NUMBER 7

Editor's Note

The International Journal of Interactive Multimedia and Artificial Intelligence provides an interdisciplinary forum in which scientists and professionals can share their research results and report new advances on Artificial Intelligence and Interactive Multimedia techniques.

This special issue, *Special Issue on Multisensor user tracking and analytics to improve education and other application fields*, concentrates on the practical and experimental use of data mining and analytics techniques, specially focusing on the educational area. The selected papers deal with the most relevant issues in the field, such as the integration of data from different sources, the identification of data suitable for the problem analysis, and the validation of the analytics techniques as support in the decision making process. The application fields of the analytics techniques presented in this paper have a clear focus on the educational area (where Learning Analytics has emerged as a buzzword in the recent years) but not restricted to it. The result is a collection of use cases, experimental validations and analytics systems with a clear contribution to the state of the art.

Gupta et al. [1] explores the use of analytics techniques towards the recognition of human activities (i.e. what a person is doing) using recorded videos as data source. The presented approach automatically detects human gait and uses clustering algorithms and Hu-Moments to construct activity templates that serves for the base of classification. According to the authors, the experimental results show a pretty high accuracy on activity recognition both at indoor and outdoor videos. This is a step forward in automatic activity identification, and provides a quite valuable help on determining if a user should be considered as data source for a certain activity (e.g. track routes for joggers).

Asawa et al. [2] utilizes a Support Vector Machine classifier for the detection of human emotions from different sources. In their study, the authors consider 3 discrete emotions (happy, anger and fear) and use audio/video recordings to determine the current most likely emotion of the recorded person, and the experimental results show a 93% of accuracy on emotion detection. This is a quite interesting result that leverages the potential of raw data for the automatic detection of high-level actions. That is, saying that a person is happy while, for example, doing a learning activity is a great indicator to predict his success in the task, much better than other measures such as the number of times they opened a digital resource.

Yuan et al. [3] analyses how dissemination activities influence the users' attitude in online communities. According to this research, communities have different user profiles where, typically, most of the users are lurkers (just consume information, do not contribute) while a low percentage are contributors. The role adopted by a user is not static, and may change over time if the users are encouraged to participate. This research uses automatic pattern recognition to classify users as lurkers or contributors and presents an experimental setting aimed to analyze how dissemination activities play a role in encouraging users' active participation.

Choudhary et al. [4] applies analytics techniques to provide recommendations to job seekers. Their analysis is feed by curriculum vitae information and by personality surveys. In their approach, they first use a questionnaire to determine the job seeker's MBTI score, which is later translated to the OCEAN model. A combination of the OCEAN classification and the CV information is analyzed in order to calculate the most suitable job for the seeker. The approach, tested with technical-skilled job seekers, showed the potential of the system to guide fresh graduates in the important decision of shaping their professional profile.

Picciano [5] contextualizes analytics methods in the learning area and provides a critical overview of the Learning Analytics field, with a special focus on blended learning situations. The author explains how learning analytics can benefit from the already existing work (e.g. with Herbert Simon's "bounded rationality" theory) with data-driven decision making models and therefore can help in the decision making process for teachers and learners. According to this critical review, analytics is not a panacea but can provide quite valuable tools to improve educational systems. Also, Picciano expresses his concerns related to the data gap that may take place in blended learning situations that may hinder the effective application of Learning Analytics tools.

Corbi et al. [6] carry out a thorough review of recommender systems and techniques, with a special emphasis in the recommendation model LIME. LIME combines categories (Learning, Interaction, Mentoring, Evaluation) and settings (Formal, Informal) to provide a balanced recommendation to the end-user, meaning the learner. In addition, they describe the implementation of LIME, called iLIME, a tutor-lecturer-crafted, rule-based recommender system; after stating the need for quality data capturing methods, the authors analyse how the online learning process can be monitored by using the Experience API. The iLIME recommender system allows teachers and tutors to design and personalize the student recommendations with a simple interface to express the rules that should trigger recommendations. The rule processor is fed by simple student action logs (i.e. user interaction, user profile, and user performance), classified by their role in the learning process. According to Corbi et al., the Experience API provides an elegant vehicle to retrieve, categorise and use the required learner actions from various sources.

Tobarra et al. [7] presents a system that integrates data from different sources in the educational scenario, in order to

DOI: 10.9781/ijimai.2014.270

analyse students' progress. This is a quite important topic, since one of the current problems with Learning Analytics is the difficulty to identify and capture the activity that takes place in all the diverse situations in which learning takes place. More specifically, the authors present a case of study where data from the LMS was combined with data from an automatic assessment system for virtual/remote laboratories, and they analyze two main questions: firstly, if the students are quitting or if they are active, secondly, if the activities were well designed. This contribution is a step forward in the application of analytics algorithms in the educational field.

Cortés et al. [8] foresee the future of higher education as a social learning environment, open and collaborative, where people construct knowledge in interaction with others. Thus, they present a project for the construction of a learning environment based on social networks, ubiquity and mobility, at the time that it suits the needs of a particular social, cultural and ethnographic context in Colombia. The inclusion of digital learning tools as the basis of the university learning environment enables data capturing, thus increasing the potential of analytics techniques.

Dr. Luis de-la-Fuente-Valentín
Prof. Dr. Daniel Burgos
Dr. Riccardo Mazza

REFERENCES

[1] J. P. Gupta, P. Dixit, N. Singh, V. B. Aemwal. "*Analysis of Gait Pattern to Recognize the Human Activities*", International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 7, pp. 7-16, 2014. DOI: 10.9781/ijimai.2014.271

[2] K. Asawa, P. Manchanda. "*Recognition of Emotions using Energy Based Bimodal Information Fusion and Correlation*", International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 7, pp. 17-21, 2014. DOI: 10.9781/ijimai.2014.272

[3] M. Yuan, M. Recker. "*Dissemination Matters: Influences of Dissemination Activities on User Types in an Online Educational Community*", International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 7, pp. 22-29, 2014. DOI: 10.9781/ijimai.2014.273

[4] R.S. Choudhary, R. Kukreja, N. Jain, S. Jain. "Personality and Education Mining based Job Advisory System", International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 7, pp.30-34, 2014. DOI: 10.9781/ijimai.2014.274

[5] A.G. Picciano. "*Big Data and Learning Analytics in Blended Learning Environments: Benefits and Concerns*", International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 7, pp. 35-43, 2014. DOI: 10.9781/ijimai.2014.275

[6] A. Corbi, D. Burgos, "*Review of current student-monitoring techniques used in elearning-focused recommender systems and learning analytics. The Experience API & LIME model case study*", International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 7, pp. 44-52, 2014. DOI: 10.9781/ijimai.2014.276

[7] L. Tobarra, S. Ros, R. Hernández, A. Robles-Gómez, A. C. Caminero, R. Pastor. "*Integration of multiple data sources for predicting the engagement of students in practical activities*" ", International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 7, pp. 53-62, 2014. DOI: 10.9781/ijimai.2014.277

[8] J.A.Cortés, J.O.Lozano. "*Social Networks as a Learning Environment for Higher Education*", International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 7, pp. 63-69, 2014. DOI: 10.9781/ijimai.2014.278

ACKNOWLEDGMENTS

This Special Issue has been supported by **UNESCO Chair on eLearning – UNIR**.

Contact details:

Headquarters / Rectorado
Gran Vía Rey Juan Carlos I, 41
26002 Logroño (La Rioja), España-Spain
Tel: +34 941 210 211

Madrid Premises / Delegación Madrid
Paseo de la Castellana 163, 8ª planta
28046 Madrid, España-Spain
Tel: +34 915 674 391

Web: **http://unesco-elearning.unir.net/**
Email:**unesco.elearning@unir.net**

Twitter**:@UNIRunescochair**

United Nations Educational, Scientific and Cultural Organization

UNESCO Chair on e-Learning Spain

UNIVERSIDAD INTERNACIONAL DE LA RIOJA

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

Analysis of Gait Pattern to Recognize the Human Activities

Jay Prakash Gupta[1], Pushkar Dixit[2], Nishant Singh[3], Vijay Bhaskar Aemwal[4]

[1]*Infosys Limited, Pune, India*
[2]*Faculty of Engineering and Technology Agra College, Agra, India*
[3]*Poornima Institute of Engineering and Technology, Jaipur, India*
[4]*SiemensInformation System, India*

***Abstract* – Human activity recognition based on the computer vision is the process of labelling image sequences with action labels. Accurate systems for this problem are applied in areas such as visual surveillance, human computer interaction and video retrieval. The challenges are due to variations in motion, recording settings and gait differences. Here we propose an approach to recognize the human activities through gait. Activity recognition through Gait is the process of identifying an activity by the manner in which they walk. The identification of human activities in a video, such as a person is walking, running, jumping, jogging etc are important activities in video surveillance. We contribute the use of Model based approach for activity recognition with the help of movement of legs only. Experimental results suggest that our method are able to recognize the human activities with a good accuracy rate and robust to shadows present in the videos.**

***Keywords* – Feature Extraction, Gait Pattern, Human Computer Interaction, Activity Recognition, Video Surveillance**

I. Introduction

THE goal of automatic video analysis is to use computer algorithms to automatically extract information from unstructured data such as video frames and generate structured description of objects and events that are present in the scene. Among many objects under consideration, humans are of special significance because they play a major role in most activities of interest in daily life. Therefore, being able to recognize basic human actions in an indispensable component towards this goal and has many important applications. For example, detection of unusual actions such as jumping, running can provide timely alarm for enhanced security (e.g. in a video surveillance environment) and safety (e.g. in a life-critical environment such as a patient monitoring system). In this paper, we use the concept of Gait for human activity recognition. The definition of Gait is defined as: “A particular way or manner of moving on foot”. Using gait as a biometric is a relatively new area of study, within the realms of computer vision. It has been receiving growing interest within the computer vision community and a number of gait metrics have been developed. We use the term Gait recognition to signify the identification of an individual from a video sequence of the subject walking. This does not mean that Gait is limited to walking, it can also be applied to running or any means of movement on foot. Gait as a biometric can be seen as advantageous over other forms of biometric identification techniques for the following reasons: unobtrusive, distance recognition, reduced detail, and difficult to conceal. This paper focuses on the design, implementation, and evaluation of activity recognition system through gait in video sequences. It introduces a novel method of identifying activities only on the basis of leg components and waist component. The use of waist below components for recognizing the activities makes it to achieve fast activity recognition over the large databases of videos and hence improves the efficiency and decreases the complexity of the system. To recognize the actions, we establish the features of each action from the parameters of human model. Our aim is to develop a human activity recognition system that must work automatically without human intervention. We recognized four actions in this paper namely walking, jumping, jogging and running. The walking activity is identified by the velocities of all components superior to zero but lesser than a predefined threshold. In case of jumping activity, every part of human moves only vertically and in the same direction either up or down. Therefore, jumping action can be identified by the velocities of all the three components to be near or equal to zero in horizontal direction but greater than zero in vertical direction. The only differences between jogging and running activities are that travelling speed of running is greater than jogging and other difference is of distance ratio between the leg components to the axis of ground. In case of running activity, speed of travelling is greater than jogging and the other difference is of distance ratio between leg components to the axis of ground.

The rest of the paper is structured as follows: Section 2 discusses the trend of activity recognition research area in the past decade which introduces the fundamentals of gait recognition systems and human activity recognition models; Section 3 presents the proposed work of human activity recognition using Gait; Section 4 analyzes and evaluates the empirical results of experiments to validate the proposed framework. Before evaluating the proposed system, some hypotheses are established and the evaluations are conducted

DOI: 10.9781/ijimai.2014.271

against these hypotheses; finally section 5 summarizes the novelties, achievements, and limitations of the framework, and proposes some future directions of this research.

II. LITERATURE REVIEW

In recent years, various approaches have been proposed for human motion understanding. These approaches generally fall under two major categories: model-based approaches and model-free approaches. Poppe has made a survey on vision based human action recognition [1]. When people observe human walking patterns, they not only observe the global motion properties, but also interpret the structure of the human body and detect the motion patterns of local body parts. The structure of the human body is generally interpreted based on their prior knowledge. Model-based gait recognition approaches focus on recovering a structural model of human motion, and the gait patterns are then generated from the model parameters for recognition. Model-free approaches make no attempt to recover a structural model of human motion. The features used for gait representation includes: moments of shape, height and stride/width, and other image/shape templates.

Leung & Yang reported progress on the general problem of segmenting, tracking, and labeling of body parts from a silhouette of the human [2]. Their basic body model consists of five U-shaped ribbons and a body trunk, various joint and mid points, plus a number of structural constraints, such as support. In addition to the basic 2-D model, view-based knowledge is defined for a number of generic human postures (e.g., “side view kneeling model,” “side horse motion”), to aid the interpretation process. The segmentation of the human silhouette is done by detecting moving edges. Yoo et al. estimate hip and knee angles from the body contour by linear regression analysis [3]. Then trigonometric-polynomial interpolant functions are fitted to the angle sequences and the parameters so-obtained are used for recognition.

In [4], human silhouette is divided into local regions corresponding to different human body parts, and ellipses are fitted to each region to represent the human structure. Spatial and spectral features are extracted from these local regions for recognition and classification. In model-based approaches, the accuracy of human model reconstruction strongly depends on the quality of the extracted human silhouette. In the presence of noise, the estimated parameters may not be reliable.

To obtain more reliable estimates, Tanawongsuwan and Bobick reconstruct the human structure by tracking 3D sensors attached on fixed joint positions [5]. However, their approach needs lots of human interaction because they have considered and identified only walking type of activity whereas our method has considered four type of activities and the performance is reasonable for each type of activity. Wang et al. build a 2D human cone model, track the walker under the Condensation framework, and extract static and dynamic features from different body part for gait recognition [6]. Their approach has fused static and dynamic features to improve the gait recognition accuracy but extraction of both static and dynamic features required more computation which lacks its applicability in real time scenario.

Zhang et al. used a simplified five-link biped locomotion human model for gait recognition [7]. Gait features are first extracted from image sequences, and are then used to train hidden Markov models for recognition. In [8], an approach for automatic human action recognition is introduced by using the parametric model of human from image sequences using motion/texture based human detection and tracking. They used the motion/texture of full body part whereas proposed approach used only the gait pattern of the lower body part which is more time efficient. Bobick & Davis interpret human motion in an image sequence by using *motion-energy* images (MEI) and *motion-history* images (MHI) [9]. The motion images in a sequence are calculated via differencing between successive frames and then thresholded into binary values. These motion images are accumulated in time and form MEI, which are binary images containing motion blobs. The MEI is later enhanced into MHI, where each pixel value is proportional to the duration of motion at that position. Moment-based features are extracted from MEIs and MHIs and employed for recognition using template matching. Because this method is based on the whole template matching instead of the only gait pattern of the legs, it does not take the advantage of recent development whereas we incorporated the matching only based on the gait analysis. Recent Gait studies for activity recognition suggest that gait is a unique personal characteristic, with cadence and cyclic in nature [10]. Rajagopalan & Chellappa [11] described a higher-order spectral analysis-based approach for detecting people by recognizing human motion such as walking or running. In their proposed method, the stride length was determined in every frame as the image sequence evolves.

TABLE 1. COMPARISON OF EXISTING APPROACHES

Ref.	Method	Advantage	Disadvantage	Uses
[8]	Gait recognition	Locomotion human model	Insensitive to noise	Indoor scenario
[9]	Model-based Action Recognition	Inclusion of motion texture	Poor performance in walking case	Indoor environment
[12]	Spectral analysis of human motion	Higher-order Spectral	Periodic detection	Differentiate between people and vehicular objects
[13]	View based motion analysis	Object models are not required	Need to reduce the distribution combinatory	Outdoor scenario
[27]	Activity recognition using smartphones	Real time application	More than one classifier reduces the accuracy	Indoor/Outdoor both

Vega and Sarkar [12] offered a novel representation scheme for view-based motion analysis using just the change in the relational statistics among the detected image features, without the need for object models, perfect segmentation, or part-level tracking. They modeled the relational statistics using the probability that a random group of features in an image would exhibit a particular relation. To reduce the representational combinatorics of these relational distributions, they represented them in a Space of Probability Functions (SoPF). Different motion types sweep out different traces in this space. They also demonstrated and evaluated the effectiveness of that representation in the context of recognizing persons from gait. But, there method requires multiple cameras from different viewpoints to model multi-view recognition system which requires extra setup and also computation, whereas the proposed approach is able to achieve high recognition performance from only a single viewpoint. Several other approaches and features used in [13-25] may be tied with gait analysis to predict the human actions. Human activity recognition using smartphones is also studied [26] but its recognition rate can be improved using gait analysis with more time efficiently. Table 1 compares the existing approaches.

III. PROPOSED METHODOLOGY

The proposed technique of human activity recognition is based on the foreground extraction, human tracking, feature extraction and recognition. Figure 1 shows the framework of the introduced human activity recognition system using Gait to identify four basic human activities (i.e. walking, running, jogging and jumping). The proposed method has following main steps: Foreground Extraction, Human Tracking, Feature Extraction and Activity Recognition. In this framework, the video is given as an input to the system from the activity database and frames are extracted from that video. The parametric model of human is extracted from image sequences using motion/texture based human detection and tracking. After that the results are displayed as the recognized activities like walking, running, jogging and jumping; and finally the performance of the method is tested experimentally using the datasets under indoor and outdoor environments.

A. Foreground Extraction

The first step is to provide a video sequence of an activity as an input in the proposed system from the dataset. That video contains a number of continuous frames. After that background subtraction technique is used to separate moving object present inside those frames. But these frames contain some noises which may lead to incurrent foreground subtraction. So first of all, we remove these noises. Some of the small noises are removed by using morphological image processing tools such as Erosion, Dilation, or Gaussian Filters. Generally, an object might be detected in several fragmented image regions. In that case, a region-fusion operation is needed. Two regions are considered to be the same object if they are overlapped or their distance less than a specific threshold value. With these constraints, the method is again very sensible to light condition, such as shadow, contrast changing and sudden changes of brightness.

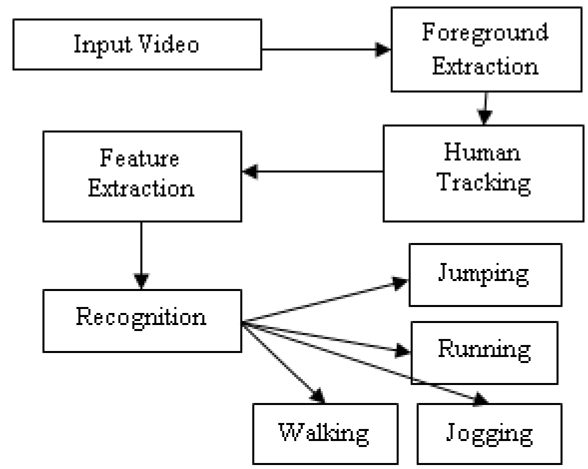


Fig. 1. Framework of Proposed System of Human Activity recognition

Intuitively, introducing some special characteristics of object, for instance texture properties, will probably improve the better results. Therefore, in the fusion process the color probability density of object's texture is additionally applied for computing the similarity between regions using Mean-shift algorithm [27]. This mixture of motion and texture of object for detection and tracking can reduce significantly noises and increases consequently the effectiveness of our tracking algorithm. However, there are always additive noises superposed with detected objects that will be eliminated later by human model constraints. The mean shift algorithm is a nonparametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. Hence, mean shift represents a general non-parametric mode finding/clustering procedure.

B. Human Tracking and Activity Recognition

In this phase, we apply Hu-moments [28] for shape analysis in which Zero- to third-order moments are used for shape recognition and orientation as well as for the location tracking of the shape. Hu-moments are invariant to translation, rotation and scaling. Hu derived expressions from algebraic invariants applied to the moment generating function under a rotation transformation. They consist of groups of nonlinear centralized moment expressions. The result is a set of absolute orthogonal (i.e. rotation) moment invariants, which can be used for scale, position, and rotation invariant pattern identification. The advantage of using Hu invariant moment is that it can be used for disjoint shapes. In particular, Hu invariant moment set consists of seven values computed by normalizing central moments through order three. In terms of central moment the seven moments are given as below:

$$M_1 = \eta_{20} + \eta_{02}$$
$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
$$M_4 = (h_{30} + h_{12})^2 + (h_{21} + h_{03})^2$$

$$M_5 = (h_{30} - 3h_{12})(h_{30} + h_{12})[(h_{30} + h_{12})^2 - 3(h_{21} + h_{03})^2] + (3h_{21} - h_{03})(h_{21} + h_{03})[3(h_{30} + h_{12})^2 - (h_{21} + h_{03})^2]$$

$$M_6 = (h_{20} - h_{02})(h_{30} + h_{12})^2 - (h_{21} + h_{03})^2 + [4h_{11}(h_{30} + h_{12})(h_{21} + h_{03})]$$

$$M_7 = (3h_{21} - h_{03})(h_{30} + h_{12})[(h_{30} + h_{12})^2 - 3(h_{21} + h_{03})^2] + (3h_{21} - h_{30})^2 (h_{21} + h_{03})[3(h_{30} + h_{12})^2 - (h_{21} + h_{03})^2]$$

These seven values given by Hu are used as a feature vector for centroid in the human model.

C. Feature Extraction

We employed a model based approach to extract the features. The extracted foreground that supposed to be a human is segmented into centroid and two leg components. We use Mean-shift algorithm again for computing the similar regions below the centroid of the human body for each leg components that will serve for tracking legs. We assume that with only these three components of human model the four basic actions could be identified correctly. The human model constraints are used for noise suppression. The three components namely centroid, left leg and right leg (i.e. vm1, vm2, vm3 respectively), are used in order to model parametric approach. The threshold concept is also used along with the defined method. Threshold calculation is applied as follows: Video sequences from the KTH and Weizmann datasets are normalized on the basis of number of frames and the time of a particular sequence for an activity. The threshold is calculated on the basis of a case study given in [29]. To recognize the actions, we establish the features of each action from the parameters of human model as follows: ***Walking feature*:** In case of walking action, every part of human move generally and approximately in the same direction and speed. Therefore, the walking activity can then be identified by the velocities of all components superior to zero but lesser than a predefined threshold for walking. Note that the significant difference between running and walking strides is that at least one of the feet will be in contact with the principal axis (ground) at any given time as shown in Figure 2 (a). ***Jumping feature*:** In case of jumping activity, every part of human moves only vertically and in the same direction either up or down [30-39]. Therefore, jumping action can be identified by the velocities of all the three components to be near or equal to zero in horizontal direction but greater than zero in vertical direction as shown in Figure 2(b). ***Jogging feature*:** The only differences between jogging and running activities were that travelling speed of running is greater than jogging and other difference is of distance ratio between the leg components to the axis of ground as shown in Figure 2(c). ***Running feature*:** Similarly in case of running activity, speed of travelling is greater than jogging and the other difference is of distance ratio between leg components to the axis of ground as shown in Figure 2 (d).

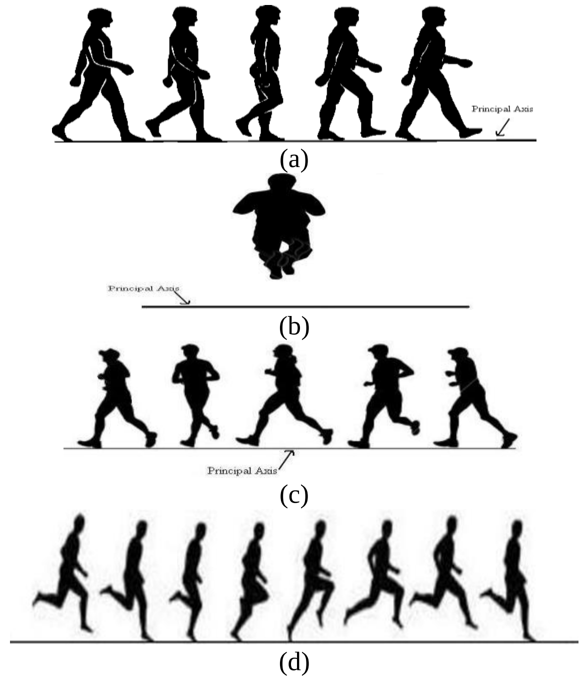
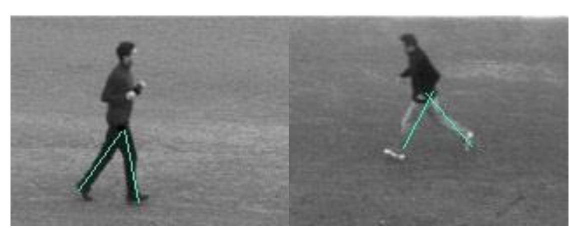


Fig. 2. Silhouette pattern for (a) Walking, (b) Jumping, (c) Jogging and (d) Running

Algorithm for Human Activity Recognition

1) Input is fed to the system as a single video sequence.
2) Frames are extracted from the input video, which are used for further processing.
3) Background subtraction technique is implemented to subtract background from the frames in order to obtain the foreground moving object.
4) Morphological operators are used to remove additional noises in the frames.
5) Mean-shift algorithm is used to track the human; based on the texture similarities in the frames.
6) Hu-moments are calculated to recognize the centroid of the tracked human. Again the Mean-shift algorithm is used to recognize each leg components of the model.
7) For feature extraction, model based approach is employed. The extracted foreground that supposed to be human is then segmented into centroid and the two leg components i.e., total three components.
8) The features of each action from the parameters of human model acts as the features for classifying all four activities (walking, jumping, jogging and running).
9) The features depend on the following criteria: Walking, Jumping, Jogging and Running.



(a) (b)


(c) (d)

Fig. 3. Templates of (a) jogging, (b) running, (c) walking, and (d) jumping for human activities.

IV. RESULTS AND DISCUSSIONS

This section analyses the various aspects of the proposed method. In activity recognition through gait, feature requirement is the main issue to model the human according to the parameters to fulfill the criteria.

A. Data Set Used

In order to evaluate our proposed approach of human activity recognition, we have used two datasets: (1) KTH Human Actions dataset (http://www.nada.kth.se/cvap/actions) and (2) Weizmann Actions dataset (http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html).

***KTH Human Actions dataset*:** KTH video dataset uses six types of human actions such as “walking”, “jogging”, “running”, “boxing”, “hand waving” and “hand clapping”, which were performed by 25 subjects in different scenarios with different clothing conditions as well.

The video sequences are down sampled to 160*120 pixels and an average length varying from 4 to 41 seconds. This dataset contains 2391 activity sequences. All videos are having static background with 25 fps. We use walking, jogging and running sequences of KTH actions data set for evaluation.

***Weizmann Actions dataset*:** Weizmann Actions dataset uses ten types of natural human actions such as “run,” “walk,” “skip,” “jumping-jack”, “jump-forward-on-two-legs”, “jump-in-place-on-two-legs”, “gallop sideways”, “wave-two-hands”, “wave-one-hand”, or “bend” which are performed by 9 different people in different scenarios with different clothing conditions as well. The video sequences are down sampled to 184*144 pixels and an average length varying from 2 to 4 seconds. This dataset contains 90 low resolution activity sequences. All the videos are having static background and running with 50 fps. We use walking, jogging and jumping sequences of Weizmann Actions dataset in this paper.

We have used templates of Mean Shift Clustering and Hu-Moments for jogging, running, walking and jumping activities as shaown in Figure 3. It is assumed that using centroid and two legs only these four activities can be identified.

B. Experimental Results

We have performed the human activity recognition experiments, with the proposed technique, on several videos, captured in outdoor and indoor environment. We have used two standard dataset namely KTH action dataset and Weizmann action dataset. In this paper, we have performed the experiments considering both indoor and outdoor scenario using KTH action dataset. But we have performed on only outdoor images of Weizmann action dataset.

1) Results on KTH dataset

Figure 4, 5 and 6 show the different frames of experimental results at different time instances on a standard KTH actions dataset. In Figure 4, first image of frame 5 shows that a human is walking. Second image of frame 5 shows the corresponding recognition result as walking with good accuracy. In Figure 5, first image of frame 10 shows that a human is jogging. Second image of frame 10 shows the corresponding recognition result as jogging. In Figure 6, first image of frame 3 shows that a human is running. Second image of frame 3 shows the corresponding recognition result as running with good accuracy.

2) Results on Weizmann dataset

To validate the robustness of our proposed method, we experimented on a standard Weizmann dataset. Figure 7, 8 and 9 shows the frame by frame result analysis of different human activity on this dataset at different time instances. In Figure 7, first image of frame 5 shows that a human is walking in outdoor environment. Second image of frame 5 shows the corresponding recognition result as walking with good accuracy.

In Figure 8, first image of frame 10 shows that a human is running in outdoor environment. Second image of frame 1 shows the corresponding recognition result as running with good accuracy. In Figure 9, first image of frame 1 shows that a human is jumping in outdoor environment. Second image of frame 1 shows the corresponding recognition result as jumping with good accuracy.

C. Result Analysis

Accuracy of proposed method is measured based on the number of frames recognized and number of frames not recognized by the following formulae:

$$Accuracy\,(\%) = \frac{No.of\;\; frames\, currectly\, recognized}{Total\; no.of\;\; video\;\; frames\, in\, a\, sequence} \times 100$$

Table 2 shows the accuracy of introduced approach over two large datasets with encouraging results; up to 95.01% of activities are recognized correctly in KTH dataset and 91.36% of activities are recognized correctly in Weizmann dataset. We have calculated the accuracy in both indoor and outdoor scenarios in the case of KTH dataset. Table 3 shows that the proposed method outperforms other existing methods.

Zhang et al. achieved 61% gait recognition accuracy over USF dataset of 4-7 activities using a simplified five-link biped locomotion human model [8]. Over indoor dataset of 5 activities, 93% accuracy is gained using the parametric model of human from image sequences using motion/texture based human detection and tracking [9]. Vega and Sarkar reported 90% accuracy using 3 actions over 71 subjects using the change in the relational statistics among the detected image features, without the need for object models, perfect segmentation, or part-level tracking [13]. Whereas, we are able to gain upto 95% and 91% accuracy using just gait analysis over KTH and Wiezmann datasets respectively. From the experimental results it is deduced that the introduced approach is more robust and able to achieve high accuracy over large datasets by considering more activities.

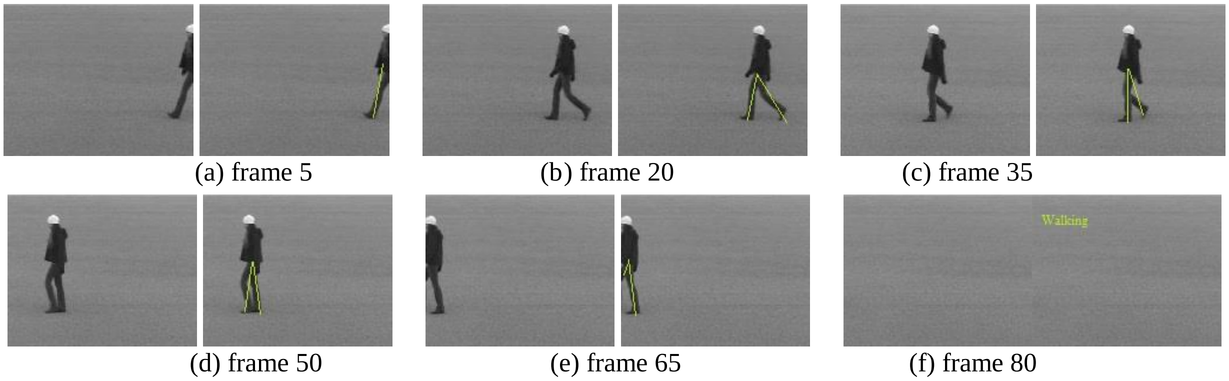


Fig. 4. Result on standard KTH dataset from of walking; first image shows input frame, second image shows corresponding output image; at the end, it recognize human activity as “Walking”.

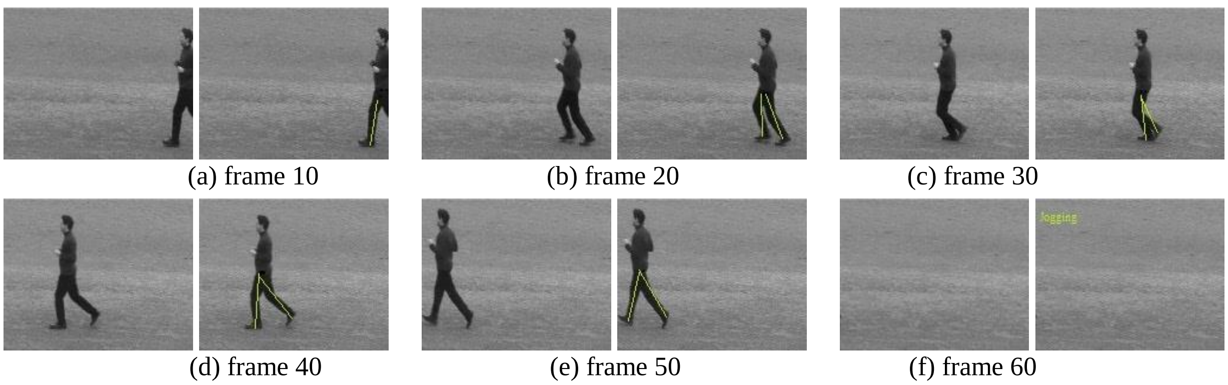
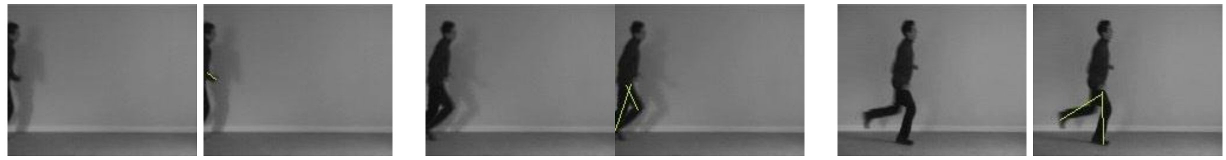


Fig. 5. Experimental result on standard KTH dataset of jogging; first image shows input frame, second image shows corresponding output image; at the end, it recognize human activity as “Jogging”.

TABLE 2. TABLE SHOWS THE RESULT ANALYSIS OF PROPOSED METHOD ON KTH HUMAN ACTIONS DATASET AND WEIZMANN ACTIONS DATASET ON THE BASIS OF FRAMES

Name of Dataset	Environment condition	Human Activities	Number of Frames	Number of Frames recognized	Recognition rate
KTH Dataset	Outdoor	Walking	1443	1434	99.3%
	Indoor	Walking	1415	1383	97.7%
	Outdoor	Jogging	1525	1425	93.4%
	Indoor	Jogging	1218	1157	94.9%
	Outdoor	Running	1089	980	89.9%
	Indoor	Running	1137	1080	94.9%
				Avg. %	**95.01%**
Weizmann Dataset	Outdoor	Walking	678	650	95.8%
	Outdoor	Running	588	552	93.8%
	Outdoor	Jumping	756	642	84.5%
				Avg.%	**91.36%**



(a) frame 3 (b) frame 6 (c) frame 12

(d) frame 18 (e) frame 24 (f) frame 30

Fig. 6. Result on standard KTH dataset of running; first image shows input frame, second image shows corresponding output image; at the end, it recognize human activity as “Running”.

(a) frame 5 (b) frame 20 (c) frame 35

(d) frame 45 (e) frame 60 (f) frame 75

Fig. 7. Experimental result on standard Weizmann dataset of walking; first image shows input frame, second image shows corresponding output image; at the end, it recognize human activity as “Walking”.

(a) frame 10 (b) frame 16 (c) frame 22

(d) frame 28 (e) frame 34 (f) frame 40

Fig. 8. Experimental result on standard Weizmann dataset of running; first image shows input frame, second image shows corresponding output image; at the end, it recognize human activity as “Running”.

(a) frame 1 (b) frame 5 (c) frame 9

(d) frame 13 (e) frame 17 (f) frame 21

Fig. 9. Experimental result on standard Weizmann dataset of jumping; first image shows input frame, second image shows corresponding output image; at *the* end of each sub-sequence it recognize human activity as “Jumping”.

TABLE 3. COMPARISON OF RESULTS WITH EXISTING METHODS

Method	Dataset	No. of Subjects	Number of Frames	Human Activities	Recognition rate
Proposed	KTH Dataset	25	7827	Walking Jogging Running	**95.01%**
	Weizmann Dataset	9	2022	Walking Jumping Running	**91.36%**
[8]	USF Dataset	75	2045	4-7 activities	**61%**
[9]	Indoor Dataset	-	9933	Standing Sitting Bending Walking Laying	**93%**
[13]	-	71	-	Walking Jogging Running	**90%**

In Figure 8, first image of frame 10 shows that a human is running in outdoor environment. Second image of frame 1 shows the corresponding recognition result as running with good accuracy. In Figure 9, first image of frame 1 shows that a human is jumping in outdoor environment. Second image of frame 1 shows the corresponding recognition result as jumping with good accuracy.

D. Result Analysis

Accuracy of proposed method is measured based on the number of frames recognized and number of frames not recognized by the following formulae:

$$Accuracy\,(\%) = \frac{No.of\ \ frames\ currectly\ recognized}{Total\ no.of\ \ video\ \ frames\ in\ a\ sequence} \times 100$$

Table 2 shows the accuracy of introduced approach over two large datasets with encouraging results; up to 95.01% of activities are recognized correctly in KTH dataset and 91.36% of activities are recognized correctly in Weizmann dataset. We have calculated the accuracy in both indoor and outdoor scenarios in the case of KTH dataset. Table 3 shows that the proposed method outperforms other existing methods.

Zhang et al. achieved 61% gait recognition accuracy over USF dataset of 4-7 activities using a simplified five-link biped locomotion human model [8]. Over indoor dataset of 5 activities, 93% accuracy is gained using the parametric model of human from image sequences using motion/texture based human detection and tracking [9]. Vega and Sarkar reported 90% accuracy using 3 actions over 71 subjects using the change in the relational statistics among the detected image features, without the need for object models, perfect segmentation, or part-level tracking [13]. Whereas, we are able to gain upto 95% and 91% accuracy using just gait analysis over KTH and Wiezmann datasets respectively. From the experimental results it is deduced that the introduced approach is more robust and able to achieve high accuracy over large datasets by considering more activities.

V. CONCLUSIONS

An efficient human activity recognition using gait technique based on model based approach is introduced in this paper which uses Mean shift clustering algorithm and Hu-Moments to construct the activity templates. This method has a promising execution speed of 25 frames per second and good activity recognition accuracy. The experimental results demonstrate that the proposed method accurately recognizes different activities in various video frames considering both indoor and outdoor scenarios while maintaining a high recognition accuracy rate. Currently our method determines key poses of each activity independently using parametric model only. Different activity classes may give similar key poses which may cause confusion and redundancy in recognition. More discriminative key poses can be applied jointly using some more refined and sophisticated algorithms such as Support Vector Machine (SVM). We found promising recognition performance more than 95% over 3-4 activities. Experimental results suggest that the proposed method outperforms other existing methods.

REFERENCES

[1] R. Poppe, "A survey on vision-based human action recognition," Image and Vision Computing, Vol. 28, No. 6, pp. 976-990, 2010.
[2] M.K. Leung and Y.H. Yang, "First Sight: A human body outline labeling system," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 17, No. 4, pp. 359–377, 1995.
[3] J.H. Yoo, M.S. Nixon and C.J. Harris, "Model-driven statistical analysis of human gait motion," In the Proceedings of the IEEE International Conference on Image Processing, pp. 285–288, 2002.
[4] L. Lee and W.E.L. Grimson, "Gait analysis for recognition and classification," In the Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 148–155, 2002.
[5] R. Tanawongsuwan and A. Bobick, "Gait recognition from time-normalized joint-angle trajectories in the walking plane," In the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 726–731, 2001.
[6] L. Wang, H. Ning, T. Tan and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 2, pp. 149–158, 2004.
[7] R. Zhang, C. Vogler and D. Metaxas, "Human gait recognition at sagittal plane," Image and Vision Computing, Vol. 25, No. 3, pp. 321–330, 2007.
[8] N. Nattapon, S. Nikom and K. Montri, "Model-based Human Action Recognition," In the Proceedings of SPIE, the International Society for Optical Engineering, pp. 111–118, 2008.
[9] A.F. Bobick and J.W. Davis, "The recognition of human movement using temporal templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 3, pp. 257–267, 2001.
[10] J. P. Gupta, N. Singh, P. Dixit, V. B. Semwal and S. R. Dubey, "Human Activity Recognition using Gait Pattern," International Journal of Computer Vision and Image Processing, Vol. 3, No. 3, pp. 31 – 53, 2013.
[11] A.N. Rajagopalan and R. Chellappa, "Higher-order spectral analysis of human motion," In the Proceedings of the International Conference on Image Processing, pp. 230–233, 2000.
[12] I.R. Vega and S. Sarkar, "Statistical motion model based on the change of feature relationships: human gait-based recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 10, pp. 1323–1328, 2003.
[13] S. R. Dubey and A. S. Jalal, "Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns," In proceeding of the Third International Conference on Computer and Communication Technology, pp. 346-351, 2012.
[14] S. R. Dubey and A. S. Jalal, "Species and variety detection of fruits and vegetables from images," International Journal of Applied Pattern Recognition. Vol. 1, No. 1, pp. 108-126, 2013.
[15] W.L. Romero, R.G. Crespo and A.C. Sanz, "A prototype for linear features generalization," International Journal of Interactive Multimedia & Artificial Intelligence, Vol. 1, No. 3 2010.
[16] K.S. Kumar, V.B. Semwal and R.C. Tripathi, "Real time face recognition using adaboost improved fast PCA algorithm," arXiv preprint arXiv:1108.1353, 2011.
[17] S. R. Dubey and A. S. Jalal, "Adapted Approach for Fruit Disease Identification using Images," International Journal of Computer Vision and Image Processing, Vol. 2, No. 3, pp. 44-58, 2012.
[18] H. Bolivar, A. Pacheco and R.G. Crespo, "Semantics of immersive web through its architectural structure and graphic primitives," International Journal of Interactive Multimedia & Artificial Intelligence, Vol. 1, No. 3, 2010.
[19] S. R. Dubey, P. Dixit, N. Singh, and J.P. Gupta, "Infected fruit part detection using K-means clustering segmentation technique," International Journal of Artificial Intelligence and Interactive Multimedia. Vol. 2, No. 2, pp. 65-72, 2013.
[20] N. Singh, S. R. Dubey, P. Dixit and J.P. Gupta, "Semantic Image Retrieval by Combining Color, Texture and Shape Features," In the Proceedings of the International Conference on Computing Sciences, pp. 116-120, 2012.
[21] S.J.B. Castro, R.G. Crespo and V.H.M. García, "Patterns of Software Development Process," International Journal of Interactive Multimedia & Artificial Intelligence, Vol. 1, No. 4, 2011.
[22] S. R. Dubey and A. S. Jalal, "Robust Approach for Fruit and Vegetable Classification," Procedia Engineering, Vol. 38, pp. 3449-3453, 2012.
[23] R.G. Crespo, S.R. Aguilar, R.F. Escobar and N. Torres, "Dynamic, ecological, accessible and 3D Virtual Worlds-based Libraries using OpenSim and Sloodle along with mobile location and NFC for checking in," International Journal of Interactive Multimedia & Artificial Intelligence, Vol. 1, No. 7, 2012.
[24] V.B. Semwal, V.B. Semwal, M. Sati and S. Verma, "Accurate location estimation of moving object in Wireless Sensor network," International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 1, No. 4 , pp. 71-75, 2011.
[25] C.J. Broncano, C. Pinilla, R.G. Crespo and A. Castillo, "Relative Radiometric Normalization of Multitemporal images," International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 1, No. 3, 2010.
[26] P. Siirtola and J. Röning, "Recognizing Human Activities User-independently on Smartphones Based on Accelerometer Data," International Journal of Interactive Multimedia & Artificial Intelligence, Vol. 1, No. 5, pp. 38-45, 2012.
[27] Y. Cheng, "Mean shift, mode seeking, and clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 8, pp. 790-799, 1995.
[28] M.K. Hu, "Visual pattern recognition by moment invariants," IRE Transactions on Information Theory, Vol. 8, No. 2, pp. 179-187, 1962.
[29] M.G. Gazendam and A.L. Hof, "Averaged EMG profiles in jogging and running at different speeds," Gait Postures, Vol. 25, No. 4, pp. 604–614, 2007.
[30] S. R. Dubey, N. Singh, J.P. Gupta and P. Dixit, "Defect Segmentation of Fruits using K-means Clustering Technique," In the Proceedings of the Third International Conference on Technical and Managerial Innovation in Computing and Communications in Industry and Academia. 2012.
[31] N. Singh, S. R. Dubey, P. Dixit and J.P. Gupta, "Semantic Image Retrieval Using Multiple," In proceeding of the 3rd International Conference on Technical and Managerial Innovation in Computing and Communications in Industry and Academia. 2012.
[32] K.S. Kumar, V.B. Semwal, S. Prasad and R.C. Tripathi, "Generating 3D Model Using 2D Images of an Object," International Journal of Engineering Science, 2011.
[33] V.B. Semwal, V.B. Semwal, M. Sati and S. Verma, "Accurate location estimation of moving object in Wireless Sensor network," International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 1, No. 4 , pp. 71-75, 2011.
[34] K.S. Kumar, S. Prasad, S. Banwral and V.B. Semwal, "Sports Video Summarization using Priority Curve Algorithm," International Journal on Computer Science & Engineering, 2010.
[35] S.R. Dubey, "Automatic Recognition of Fruits and Vegetables and Detection of Fruit Diseases", M.Tech Thesis, GLAU Mathura, India, 2012.
[36] M. Sati, V. Vikash, V. Bijalwan, P. Kumari, M. Raj, M. Balodhi, P. Gairola and V.B. Semwal, "A fault-tolerant mobile computing model based on scalable replica", International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 2, 2014.
[37] S.R. Dubey and A.S. Jalal, "Automatic Fruit Disease Classification using Images", Computer Vision and Image Processing in Intelligent Systems and Multimedia Technologies, 2014.
[38] D.D. Agrawal, S.R. Dubey and A.S. Jalal, "Emotion Recognition from Facial Expressions based on Multi-level Classification", International Journal of Computational Vision and Robotics, 2014.
[39] S.R. Dubey and A.S. Jalal, "Fruit Disease Recognition using Improved Sum and Difference Histogram from Images", International Journal of Applied Pattern Recognition, 2014.

Jay Prakash Gupta received his MTech in CSE in 2012 from GLA University Mathura, India and BTech in CSE in 2010 from B.S.A. College of Engineering and Technology Mathura, India. He is working as a System Engineer in Infosys Limited, Pune, India. His research interests include image processing, computer vision and pattern recognition.

Pushkar Dixit is currently working as an Assistant Professor in Computer Science and Engineering, FET Agra College, Agra, India. He received his MTech in Computer Science from GLA University Mathura, India and BTech in Computer Science from FET Agra College Agra, India. He has 3 year of teaching and research. His research interests include image processing, network security and computer networks.

Nishant Singh received his MTech in CSE in 2012 from GLA University Mathura, India and BTech in CSE in 2010 from B.S.A. College of Engineering and Technology Mathura, India. He has 1 year of teaching and research experience and currently, he is working as an Assistant Professor in Department of Computer Engineering and Applications, Poornima Institute of Engineering and Technology, Jaipur, India. His research interests include image processing, computer vision and pattern recognition.

Vijay Bhaskar Semwal has been severed as a PCC member in several International conferences and Journals. He was also head for 5th IEEE conference on big data at COER Roorkee. He is carrying more than 5 year of academic and industry experience. He served for various top organizations like Siemens, Newgen etc. Earlier he received his M.tech in wireless sensor domain and was awarded by gold medal. His major research work interest in Wireless Sensor Network, Artificial Intelligence, Image Processing, Computer Network &Security, and Design & Analysis of Algorithm, Machine Learning, and Information Retrieval Soft Computing

Recognition of Emotions using Energy Based Bimodal Information Fusion and Correlation

Krishna Asawa, Priyanka Manchanda[1]

[1]Department of Computer Science Engineering and Information Technology
Jaypee Institute of Information Technology, Noida, India

***Abstract* — Multi-sensor information fusion is a rapidly developing research area which forms the backbone of numerous essential technologies such as intelligent robotic control, sensor networks, video and image processing and many more. In this paper, we have developed a novel technique to analyze and correlate human emotions expressed in voice tone & facial expression. Audio and video streams captured to populate audio and video bimodal data sets to sense the expressed emotions in voice tone and facial expression respectively. An energy based mapping is being done to overcome the inherent heterogeneity of the recorded bi-modal signal. The fusion process uses sampled and mapped energy signal of both modalities's data stream and further recognize the overall emotional component using Support Vector Machine (SVM) classifier with the accuracy 93.06%.**

***Keywords* — Bimodal Fusion, Emotion Recognition, Intelligent Systems, Machine Learning, Energy Mapping**

I. Introduction

MULTI-SENSOR information fusion is a rapidly developing area of research and development which forms the foundation of intelligent robotic control. It comprises of methods and techniques which collect input from multiple similar or dissimilar sources and sensors, extract the required information and fuse them together to achieve improved accuracy in inference than that could be achieved by the use of a single data source alone. In this contribution, we discuss a novel approach to fuse heterogeneous datasets obtained from multiple sensors with the aim of analyzing the human's emotional behavior.

Emotions play an important role in human-to-human communication and interaction, allowing people to express themselves beyond the verbal domain. The ability to understand human emotions is desirable for the computer in some applications such as computer-aided learning or user-friendly on-line help. During an interaction, an individual uses multiple modalities such as eye gaze, hand gestures, facial expressions, body posture, and tone of voice. Human behavior is thus, inherently multimodal. In addition to its multimodal nature, the emotional state of an individual is also an integral component of human experience and plays a significant role in developing intelligent systems for human computer communication. It influences numerous phenomenons such as cognition, perception, learning, creativity and decision-making. Besides the problem solving, reasoning, perception and cognitive tasks, emotion recognition also plays a pivot role in functions which are essential for artificial intelligence.

Considering these two aspects of human behavior, we have designed and developed a technique to analyze and correlate bimodal data sets and further recognize the emotional component from these fused data sets. This new technology ensures a proper balance between emotion recognition and cognition tasks.

The existing fusion methods as listed in the section- related work, do not address how to bridge the heterogeneity present in the captured data, which corresponds to the individual modality. The energy based mapping method inspired from how the different sensed stimuli signals by humans, mapped to the corresponding energy onto designated areas of the brain. This method brings homogeneity among heterogeneous emotional cues by transforming them onto their corresponding energy levels. The achieved fusion accuracy of 93.06% can ensures a proper balance between emotion recognition and cognition tasks.

The rest of the paper has been organized in the following manner: in Section II, along with existing fusion approaches, we discuss an energy based method for fusion of multimodal data sets. In Section III, we explain the architectural framework of our model. The implementation of the solution is delineated in Section IV. Section V outlines the applications of this model. Lastly, we conclude the research study in Section VI.

II. Related Work

The wide use of multimodal data fusion technologies in versatile areas of application has invoked an ever increasing interest of researchers all over the globe. Multimodal data fusion techniques are used in numerous areas such as intelligent systems, robotics, sensor networks, video and image processing and many more. Multimodal data fusion can be performed at three levels: feature, decision and hybrid level fusion.

Feature level fusion has been used in [1] for fusing range of spatial cues with the relative assignment of linear weight to them. But they have unable to resolve the issue of how weights

DOI: 10.9781/ijimai.2014.272

should be assigned to justify relevance and importance of different cues.

Neti [2] have been performed decision level fusion for speaker recognition and speech event detection. They have analyzed audio features (e.g. phonemes) and visual features (e.g. visemes) independently to arrive at recognized decision according to single modality. Thereafter they have employed a linear weighted sum strategy to fuse these individual decisions. The authors have used the training data to determine the relative reliability of the different modalities and accordingly adjusted their weights. Where as in [3], for speaker identification, they have considered the results of different classifier at decision level fusion. From the speech corpora, a set of patterns are identified for each speaker on the basis of predefined features by two different classifiers. The majority decision regarding the identity of the unknown speaker is obtained by fusing the output scores of all the classifiers using a late integration approach.

A multimodal integration approach using custom defined rules has been suggested by, Holzapfel et al. [4]. They have shown smooth human - robot interaction in the kitchen setting by fusing results of speech and 3D pointing gestures. This multimodal fusion which is performed at the decision level based on the n-best lists generated by each of the event parsers. A close correlation in time of speech and gesture has been proved by this approach, but this is leading to the process time overhead to determine the best action based on n-best fused input.

In [5] two techniques viz (1) Gradient-descent-optimization linear fusion (GLF) and (2) the super-kernel nonlinear fusion (NLF) are suggested. Each of which does the optimal combination of multimodal information for video concept detection. In GLF, an individual kernel matrix is first constructed and then fused together based on a weighted linear combination scheme. Unlike GLF, the NLF method does nonlinear combination of multimodal information.

In [6], the authors have used NLF method and first construct an SVM for the individual modality as a classifier. Thereafter, for optimal combination of the individual classifier models a super kernel non-linear fusion is applied. Experiments conducted on TREC-2003 Video Track benchmark shows NLF has on average 3.0% better performance than GLF.

To classify image, Zhu et al. [7] have given a hybrid level multimodal fusion framework. They have used SVM to classify the images with embedded text within their spatial coordinates. The fusion process is done in two steps. Firstly, on the basis of low-level visual features, a bag-of-words model [8] is used to classify the given image. At the same time, the text detector records the existence of text in the image using text color, size, location, edge density, brightness, contrast, etc. In the second step, for fusing the visual and textual features together a pair-wise SVM classifier is used.

A time-delayed neural network employed by Cutler and Davis [9] for feature level multimodal data fusion in for locating the speaking person in the scene. This is being done by identifying the correlation between audio and visual streams.

In another work, related to detecting human activities Gandetto et al. [10] have used the Neural Network decision level fusion method to combine sensory data. An environment equipped with a heterogeneous network of state sensors for sensing CPU load, login process, and network load and cameras for sensing observation along with computational units working together in a LAN is considered for the experiment. Human activity is monitored by fusing the data from these two types of sensors at the decision level.

A framework is given in [16] which fuse textual and visual information. Author has proposed additional preprocessing before combining these modalities in a linear weighted fashion at the feature and scoring levels. The pre-processing called as latent semantic mixing, takes care about overlapping information among both modalities by mapping the bimodal feature space onto low dimensional semantic space.

In this paper, we propose a feature level linear weighted fusion model based on a human-inspired concept of brain energy mapping model. Humans collect sensory data via human biological senses (sight, hearing, touch, smell and taste) and map this data as energy stimuli onto designated regions of the brain. The brain then fuses them together to obtain an inference. This analogy is employed in designing the architectural framework of our work. This phenomenon is depicted in Fig 1.

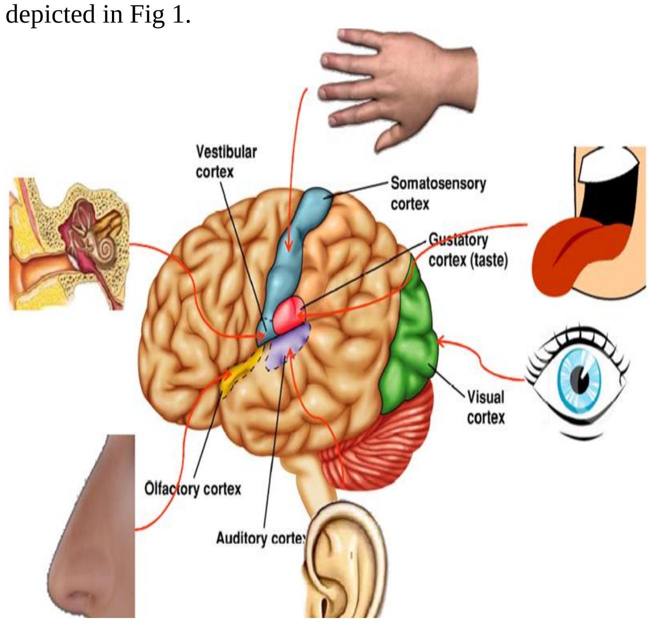


Fig. 1. The Brain Energy Mapping Model

III. ARCHITECTURAL FRAMEWORK

Figure 2 shows the overall architectural framework and computation stages as listed below.

- Step1: Obtain Bi-Modal Input stream
- Step 2: Split Bi-Modal Input into Audio and Video Components
- Step 3: Synchronized Sampling and processing of Audio and Video Components

- Step 4: Run two parallel process, each for audio and video.
- Audio thread performs segmentation and feature extraction for audio sample using praat tool.
- Video thread performs segmentation and facial feature extraction for video sample DAFL library.
- Step 5: Estimation of Audio and Video features energy.
- Step 6: Perform one of the following depending on user's input:
- Train SVM
- Test an unknown sample with trained SVM to predict emotion
- Step 7: Display emotion to the user

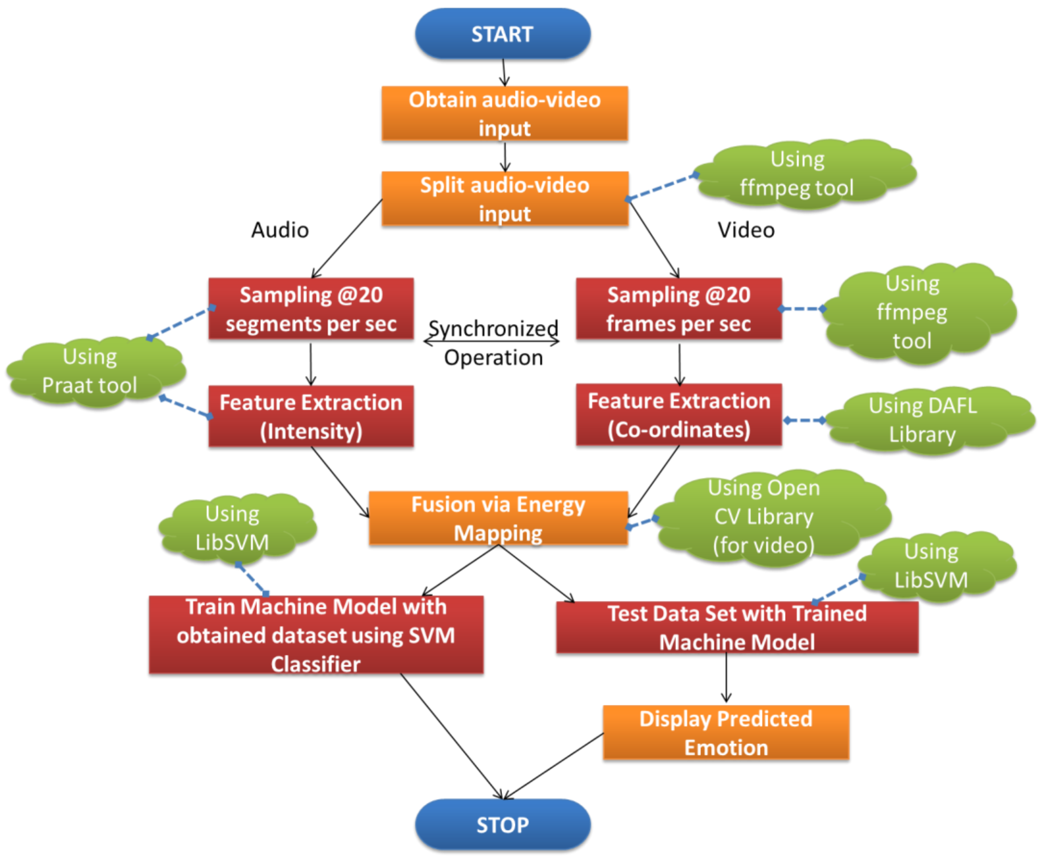


Fig 2. Architectural Framework of Bimodal Energy Based Fusion Model

Stage I – Data Pre-Processing

The bimodal inputs obtained and then split into two components – audio and video. Thereafter, audio processing and video processing is performed simultaneously and in synchronization. The synchronization is necessary to ensure that no data is lost and the audio and video samples at any particular instance are processed simultaneously.

The audio component is segmented at the rate of 20 samples (utterances) per second. Video Sampling is done at the rate of 20 frames per second.

Stage II – Feature Extraction

The prosodic feature mean intensity of the audio component between time 't1' and 't2' is computed as:

$$\frac{1}{(t2-t1)}\int_{t1}^{t2} x(t)dt \qquad (1)$$

where x(t) is intensity as function of time (in dB).

To compute the intensity, the values in the sound are first squared, then convolved with a Gaussian analysis window (Kaiser-20; sidelobes below -190 dB). The effective duration of this analysis window is 3.2 / (minimum_pitch), which guarantee that a periodic signal is analysed as having a pitch-synchronous intensity ripple not greater than 0.00001 dB.

The processing of video frames is done in two steps:

- Facial Feature extraction using the Discrete Area Filters (DAF) Library used to extract coordinates of 15 facial feature points. [13]
- Energy (gradient) computation (Fig 3) of extracted facial features co-ordinates using OpenCV Library.

A	B	C
D	E	F
G	H	I

$$energy(E) = \sqrt{xenergy^2 + yenergy^2}$$
$$xenergy = a + 2d + g - c - 2f - i$$
$$yenergy = a + 2b + c - g - 2h - i$$

Fig 3. Energy (Gradient) Computation

In Fig. 3, each lowercase letter represents the brightness (sum of the red, blue, and green values) of the corresponding pixel. To compute the energy of edge pixels, we consider that the image is surrounded by a 1 pixel wide border of black pixels (with 0 brightness).

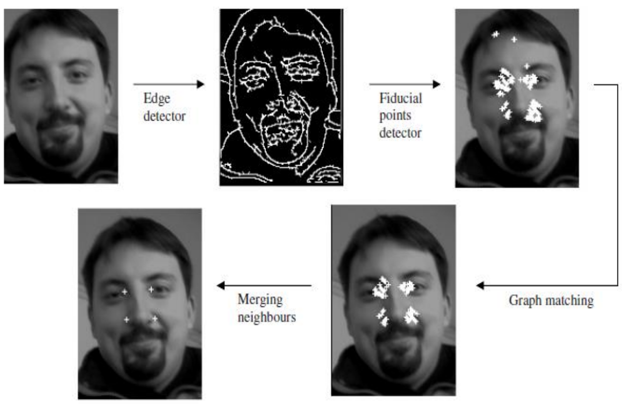


Fig 4. Facial Feature Detection using Discrete Area Filters

Fig 5. Bi-modal Input Processing

Stage III – Fusion via Energy Mapping

Our framework uses the technique of feature level linear weighted fusion. Consider a feature set <Ev, Ea>, where Ev is the total energy of video features and Ea is the total energy of audio features. The feature set is computed at intervals of 1 second for the bi-modal input.

$$E_a = wE_{a1} + wE_{a2} + \dots\dots + wE_{an} \qquad (2)$$

where,
$E_a$is the audio energy of the audio sample of 1sec duration.
E_{ai} is the audio energy of i^{th} sub-sample
w is the weight assigned to each sub-sample.
n is the number of sub samples = 20

(3)

$$E\_vi = w\_1 \ (E\_vio + E\_vi1 + E\_vi2 + E\_vi3) + w\_2 \ (E\_vi4 + E\_vi5 + E\_vi6 + E\_vi7) + w\_3 \ (E\_vi8 + E\_vi9 + E\_vi10) + w\_4 \ (E\_vi11 + E\_vi12 + E\_vi13 + E\_vi14)$$

where,
E_{vi} is the energy of i^{th} frame of the video sub-sample.
w_1 is weight assigned to left eye fiducial points.
w_2 is weight assigned to right eye fiducial points.
w_3 is weight assigned to nose fiducial points.
w_4 is weight assigned to mouth fiducial points.

$$w_4 > w_1, w_2 > w_3$$

$$E_v = wE_{v1} + wE_{v2} + \dots\dots\dots + wE_{vn} \qquad (4)$$

where,
E_v is the combined energy of the video sub-samples.
E_{vi} is the energy of i^{th} video sub-sample.
w is the weight assigned to each sub-sample.
n is the number of sub samples = 20

We further label the feature set with appropriate class (1 – Happy, 2 – Anger, 3 – Fear) depending on the emotional state of the user. This feature set is then used to train the machine model designed for predicting the mood of the user.

Stage IV – Emotion Prediction

The Support Vector Machine (SVM) classifier is used to predict the emotional state of the bimodal input. We use LibSVM[14] to develop the C-SVC (C - Support Vector Classification) SVM having RBF (Radial Basis Function - exp(-gamma*|u-v|^2)) kernel.

We further develop a machine model using C-SVC SVM and train it using the feature sets obtained in Stage III. The feature sets of the input to be tested are then labeled with an arbitrary label. These are then tested using the trained machine model. Finally, the predicted emotional state of the bi-modal input is displayed to the user.

IV. RESULTS AND ACCURACY CALCULATION

The energy based bimodal data fusion model was tested for the eNTERFACE[15] database with 3 discrete emotions that are happy, anger and fear. The specifications of the database are as follows:

- 648 samples
  - 43 subjects enacting 5 sentences of each of 3 emotions (happy, anger and fear)
- Samples having both male and female subject
- Frontal views with moderate lighting conditions
- Single person input

The 80: 20 ratios of the training and testing samples are considered for cross validation. The total samples for each emotion are 215. Three times process has been repeated with different sets of training and testing in the ratio of 80:20. On average in each round 485 samples are classifies correctly on the basis of 518 samples. The average percentage of classification for the three emotions is shown in the table 1.

TABLE 1. CONFUSION MATRIX (IN %)

	Predicted Emotion			
		Happy	Anger	Fear
Actual Emotion	Happy	92.59%	4.17%	3.24%
	Anger	1.85%	94.44%	3.70%
	Fear	1.85%	6.02%	92.13%

The model shows 93.06% accuracy for emotion recognition of Happy, Anger and Fear Emotions using energy mapping model.

V.CONCLUSION

In this research study, we have developed a tool to analyze and correlate bimodal data sets of emotional cues using energy based fusion model and further recognized the emotional component from these bimodal data sets using Support Vector Machine classifier. We have mapped the audio and video features of bimodal input to their corresponding energy levels. The model is tested for eNTERFACE 2005 database and an accuracy of 93.06% is obtained recognition of happy, anger and fear emotions. The tool developed for bimodal energy based fusion model can further be used as a wrapper tool to develop intelligent applications which require multimodal data fusion and emotion recognition, such as Real Time Emotion Recognition, Expressive Embodied Conversational Agent, Virtual Tutor, Questionnaire which analyse verbal and non-verbal behavior.

ACKNOWLEDGMENTS

The work reported in this paper is supported by the grant received from All India Council for Technical Education; A Statutory body of the Govt. of India. vide f. no. 8023/BOR/RID/RPS-129/2008-09.

REFERENCES

[1] Wang, J., Kankanhalli, M. S., Yan, W., & Jain, R. (2003). Experiential sampling for video surveillance. In First ACM SIGMM international workshop on Video surveillance (pp. 77-86).

[2] Neti, C., Maison, B., Senior, A. W., Iyengar, G., Decuetos, P., Basu, S., &Verma, A. (2000). Joint processing of audio and visual information for multimedia indexing and human-computer interaction(pp. 294-301).

[3] Radová, V., & Psutka, J. (1997). An approach to speaker identification using multiple classifiers. In Acoustics, Speech, and Signal Processing, ICASSP-97 (Vol. 2, pp. 1135-1138).

[4] Holzapfel, H., Nickel, K., & Stiefelhagen, R. (2004). Implementation and evaluation of a constraint-based multimodal fusion system for speech and 3D pointing gestures. In Proceedings of the 6th international conference on Multimodal interfaces (pp. 175-182).

[5] Wu, K., Lin, C.K., Chang, E., Smith, J.R. (2004) Multimodal information fusion for video concept detection. In: IEEE International Conference on Image Processing, Singapore (pp. 2391–2394).

[6] Adams, W. H., Iyengar, G., Lin, C. Y., Naphade, M. R., Neti, C., Nock, H. J., & Smith, J. R. (2003). Semantic indexing of multimedia content using visual, audio, and text cues. EURASIP Journal on Advances in Signal Processing (pp. 170-185).

[7] Zhu, Q., Yeh, M. C., & Cheng, K. T. (2006). Multimodal fusion using learned text concepts for image categorization. In Proceedings of the 14th annual ACM international conference on Multimedia (pp. 211-220).

[8] Li, F.F., Perona, P. (2005). A bayesian hierarchical model for learning natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington (vol. 2, pp. 524–531)

[9] Cutler, R., & Davis, L. (2000). Look who's talking: Speaker detection using video and audio correlation. In Multimedia and Expo, 2000. ICME 2000 (Vol. 3, pp. 1589-1592).

[10] Gandetto, M., Marchesotti, L., Sciutto, S., Negroni, D., Regazzoni, C.S. (2003). From multi-sensor surveillance towards smart interactive spaces. In: IEEE International Conference on Multimedia and Expo, Baltimore (pp. I:641–644).

[11] Bellard, F., & Niedermayer, M. (2012). FFmpeg. http://ffmpeg.org

[12] Boersma, Paul &Weenink, David (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.77, retrieved 18 May 2014 from http://www.praat.org/.

[13] Naruniec, J., &Skarbek, W. (2007). Face detection by discrete gabor jets and reference graph of fiducial points. In Rough Sets and Knowledge Technology Springer Berlin Heidelberg (pp. 187-194).

[14] Martin, Olivier, et al. 2006. The eNTERFACE' 05 Audio-Visual Emotion Database. Data Engineering Workshops, Proceedings.

[15] Chih-Chung Chang and Chih-Jen Lin (2006). LIBSVM : a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[16] [16] Nam Khanh Tran (2012) Multimodal Fusion for Combining

[17] Textual and Visual Information in a Semantic mode, Thesis submitted to Universitat Des Saarlandes.

Dr. Krishna Asawa presently working with Jaypee Institute of Information Technology (JIIT), Deemed to be University, NOIDA, INDIA in the capacity of Associate Professor. Dr. Krishna awarded Doctor of Philosophy (CSE) in 2002 from Banasthali Vidyapith, Deemed to be University, Banasthali, INDIA. Her area of interest and expertise includes Soft Computing and its applications, Information Security, Knowledge and Data Engineering. Before joining to the JIIT she worked with National Institute of Technology, Jaipur, INDIA and with Banasthali Vidyapith.

Ms. Priyanka Manchanda has completed her graduation in Computer Science and Engineering from Jaypee Institute of Information Technology (JIIT), Deemed to be University, NOIDA, INDIA in 2014. She is currently pursuing MS at Columbia University, New York.

Dissemination Matters: Influences of Dissemination Activities on User Types in an Online Educational Community

Min Yuan[1], Mimi Recker[1]

[1]*Utah State University*

***Abstract* — Emerging online educational communities provide spaces for teachers to find resources, create instructional activities, and share these activities with others. Within these online communities, individual users' activities may vary widely, and thus different user types can be identified. In addition, users' patterns of activities in online communities are dynamic, and further can be affected by dissemination activities. Through analyzing usage analytics in an online teacher community called the *Instructional Architect*, this study explores the influences of dissemination activities on the usage patterns of different user types. Results show that dissemination activities can play an important role in encouraging users' active participation, while the absence of dissemination activities can further increase participation inequality.**

***Keywords* — Educational technology, Learning systems, Online Communities, Pattern analysis**

I. Introduction

TEACHERS increasingly rely on the Internet to find online learning resources, create instructional activities using these resources, and then share these with others [1]-[4]. To help teachers in these tasks, several web-based tools, such as the *Instructional Architect*, the *Curriculum Customization Service*, and *Tapped In*, have been developed [5]-[7]. These tools are designed to help teachers' knowledge building processes, as well as to help support the development of online educational communities [8].

In an online educational community, a virtual space is provided for teachers and learners to seek information, ask questions, and interact with one and another [9]. In general, an online educational community contains the following four elements: *people* who create content and connect with each other, *computer systems* that mediate people's activities, *policies* that guide people's activities, and *purposes* that provide reasons and motivations for people to participate [10]. People participating in an online educational community typically have shared purposes, but their actual activities in the community can vary widely. For example, some teachers may actively collect resources and design instructional activities using these resources, some may willingly share their resources and teaching activities with other users, while others may simply engage in viewing other users' activities [11], [12].

As people engage in these different activities, they can be categorized into different user types. At a high level, two main categories of users have been identified in online communities: *lurkers,* who take on more non-participatory roles and principally view other members' activities and products; and *contributors,* who take on more active roles, create new content, and share with the community [13], [14]. Prior research has also detected that the lurker-contributor ratio in communities is often skewed, with substantially more lurkers than contributors [1], [14], [15], [16].

Further, patterns of activity over time in online educational communities are dynamic, resulting in different developmental paths [17]. For example, over time, one online community may thrive and grow with more user activity, while another may shrink (or even die) with fewer users and less participation [18]. Additionally, as time passes, some lurkers may follow a trajectory toward becoming contributors in a community [19].

However, despite prior research on characterizing user typologies in online educational communities, less work has focused on understanding user activity patterns, the evolution of patterns over time, and the resulting dynamics of online educational communities. As such, in this article, we report results from applying techniques from the emerging field of learning analytics to analyze usage patterns in an online educational community for teachers, called the *Instructional Architect* (IA.usu.edu). Understanding the evolution of user activities and the dynamics of a community is complex, as the analysis revolves around mining the massive amounts of data automatically generated by the community [7], [22]. Techniques from learning analytics offer approaches for analyzing these kinds of data, such as comparing users' number of logins and visit duration, analyzing user-generated content, and examining the relationships between users [23], [24]. Outcomes of such research can provide suggestions for dissemination activities that can promote particular users' activities in order to enhance the development and sustainability of online communities [20], [21].

In particular, this article reports results of longitudinal analyses of usage data automatically collected by the IA over two full school years. During the first school-year period

DOI: 10.9781/ijimai.2014.273

(2009-10), the development team conducted extensive dissemination activities; however, these had ceased by the second school-year period (2012-13). In this way, we explored the influences of dissemination activities on different IA user types, and whether these typologies changed after dissemination activities ended. By comparing the analytics of different user types during and after dissemination, this study identified which user types and what kind of activities were most affected when dissemination activities ended, thus providing insights on the sustainability of online communities.

II. Theoretical context

A. Online Educational Communities

Online educational communities have become an important part of teachers' lives, in that they can help teachers' seek instructional resources and interact with other teachers [13], [25]. Like any community, these online educational communities have different life cycles.

Researchers have provided a variety of definitions and descriptions of these life cycles. For example, [20] defined three stages in the life cycle: starting the online community, encouraging early online interaction, and moving to a self-sustaining community. [18] divided the life cycle into five stages: inception, creation, growth, maturity, and death.

Although the definitions are different, researchers have identified similar development trajectories in the evolution of online communities. For example, in the early stage of an online community, the technological components are developed and groups of users with similar purposes and needs begin to create content and/or interact with each other. At maturity or the self-sustaining stage, the community may have a large number of members and a large repository of content [18].

In addition, in analyzing the life cycles of communities, researchers have also focused on the role that dissemination plays in the development of communities. They noted that dissemination activities are important in encouraging users' early participation and interaction, in maintaining their interests over time, and in supporting the sustainability of communities [20], [18], [26].

In this vein, researchers have identified factors that can influence the development and sustainability of online community. For example, [27] listed two factors: the creation of content and the interaction between users. [9] analyzed two factors that appeared to determine the success of online communities: usability (how people can access, create, and use content), and sociability (how users can interact).

B. Users in Online Educational Communities

Research has also focused on identifying different user typologies based on users' participation practices in an online community. For example, [15] categorized users of the *Instructional Architect* based on their activity patterns using a probabilistic clustering algorithm. Results revealed three types of user groups, where each group had characteristic patterns in terms of its frequency in creating, viewing, and sharing content. Similarly, [16] investigated usage patterns in the *Curriculum Customization Service* by analyzing users' clickstream data. They found that some user types were characterized by viewing many interactive resources and shared resources, while other types were spending more time on viewing instructional materials and assessments.

By reviewing user patterns across several different online communities, [14] proposed the "90-9-1" rule. This rule divides users into three groups: 1) approximately 90% of the users are lurkers, who view other users' resources and products but do not contribute; 2) 9% of the users are intermittent contributors; and 3) 1% of the users are heavy contributors, who participate heavily and create most of the content in the community.

This "participation inequality" rate has been observed in several online communities. For example, [28] investigated the distribution of contributions made by authors in Wikipedia, and found that less than 10% of the total number of authors created more than 90% of the content. In analyzing user-generated content in nine popular websites (e.g., Amazon book review, Merlot.org, Slideshare.net), [29] found that the distribution of user-generated content similarly followed a "long-tail" distribution, thus providing further evidence of participation inequality.

This "participation inequality" phenomenon results in a skewed lurker–contributor ratio, as well as the "free riding" problem. In this phenomenon, users benefit from other users' activities without contributing anything in return [30]. If many users become "free riders" (or lurkers) in a community, participation and the number of resources created in the community will grow slowly, which may in turn negatively affect users' interest as well as the overall sustainability of the community [31].

To further examine participation inequality and the potential free riding problem, researchers have studied why lurkers may behave this way. Reasons for lurking include a desire for users to get to know the norms of a community before becoming contributors, a lack of familiarity with the community, a lack of reasons for contributing content, and technology barriers [11], [13]. In addition, by comparing lurkers and non-lurkers' activities, researchers have found that non-lurkers tend to have a desire for a greater variety of activities, such as getting answers to questions, participating in conversations, or offering expertise [11].

It is also important to note that users' activities in the communities often change over time. For example, lurkers can begin to create and share their products once they become more familiar with the functions of the online community and build trust with other users. Non-lurkers can become lurkers as they gradually lose interest or their needs are satisfied [18]. Thus, understanding how the activities of different types of users evolve over time, especially in response to changes in support or dissemination activities within the community, are needed.

III. Technological Context

The technological context for this study is the *Instructional Architect* (IA). The IA is a free, web-based tool that was first launched in 2001. Using iterative design approaches, the tool was improved several times and development stabilized in 2005, before the data were collected for the present article.

The IA enables teachers to use online educational resources to create, publish, and share instructional activities (called IA projects) within the IA online community [5], [32]. Figure 1 shows an IA project created by a teacher. This IA project, on the topic of the "Underground Railroad", provides text, maps, and links to supporting resources.

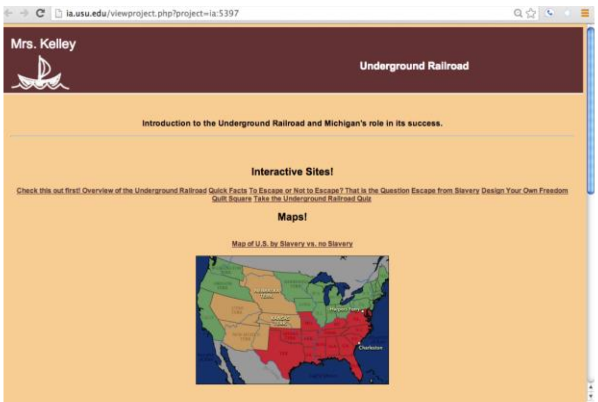


Fig. 1. A screenshot of a teacher-created IA project

Within the IA online educational community, users can engage in many different activities. Without logging in, any user can *browse* IA projects created and shared by other IA users. After logging in, a user can also *collect* online resources in his/hew own personal repository of online resources by using the 'My Resources' area of the IA to search for and save online resources from existing content repositories (e.g., the NSDL.org), or online content including web pages, pdf documents, or other public IA projects.

In the 'My Projects' area, teachers can *create* IA projects using online resources they have collected and annotate them with text. An IA project (a webpage) is then generated, which can then be used in a classroom activity. Finally, teachers can *share* IA projects by making them *public*, so that other users can easily view and copy them.

Since 2005, the IA has approximately 7,900 registered users, who have gathered over 75,600 online resources and created over 17,300 IA projects. Since August 2006, public IA projects have been viewed over 2.5 million times.

IV. Research Design

A. Research Design

This study analyzed the usage log files automatically collected by the IA in order to examine the evolution of the activity patterns of different user types. Two different time periods were examined: one in which dissemination activities were ongoing, and the other in which they had ended.

Since the launch of the IA, developers and researchers have taken many approaches for disseminating the tool to teachers. These included advertising online, offering teacher professional development workshops, and presenting at conferences. [32]. For example, between 2007 and 2011, a series of teacher workshops were conducted in several U.S. states, including South Dakota, Illinois, New York, and Utah. The workshops familiarized teachers with the IA, showed them how to design IA projects, and encouraged them to integrate these IA projects in their teaching.

In addition, members of the development team presented about the IA at several conferences, including the *International Conference on Educational Data Mining*, *Joint Conference on Digital Library*, the *Annual Meeting of the Association for Educational Communications Technology*, and the *Annual Meeting of the American Education Research Association* [33]-[36], as well as local, teacher-oriented conferences.

To examine the influence of dissemination activities on IA users types, this study compared the activities of different IA user types during two time periods: 1) the "*active dissemination*" period (9 months between 09/01/2009 - 05/31/2010), in which developers engaged in active dissemination activities, and 2) the "*no dissemination*" period (9 months between 09/01/2012 - 05/31/2013) in which dissemination activities had ended. Note that the nine-month period corresponds to the school year of U.S. teachers, our target users. It is also noteworthy that the activities we analyzed are IA users' naturally occurring behaviors, and *not* those of users specifically recruited to participate in a research study.

Specifically, this study had two research purposes: examining 1) how the IA community evolved and changed during and after dissemination activities, and 2) more specifically, how the activities of particular subsets of IA users also changed after dissemination activities ended. To align with these purposes, different user groups and data sources were used to address two research questions (see Table 1):

1. How did the activities of **IA visitors** change between the "active dissemination" period and the "no dissemination" period?
2. How did the activities of **lurkers** and **active contributors** change between the "active dissemination" period and the "no dissemination" period?

Usage activity in the IA is automatically collected by two complementary data sources: Google Analytics (GA) and the relational database powering the IA site (IADB). As a Google service, GA records the activities of all users in the IA website (which we call *IA visitors*). In particular, GA tracks visitors to the IA website, regardless of whether they have an account. In this analysis, we used seven metrics collected by GA (see Table 2) to analyze the activities of IA visitors.

TABLE 1
RESEARCH QUESTIONS AND DATA SOURCES

Research Questions	User Group Analyzed	Data Sources
RQ1	All visitors to the IA site	Google Analytics
RQ2	Users who created an IA account, in two groups: *lurkers* (did not create IA project) and *active contributors* (created IA projects)	IA database

B. Data Sources

TABLE 2
METRICS DESCRIBING ACTIVITIES OF IA VISITORS USING GOOGLE ANALYTICS

Metric	Description
# of visits	Number of visits to the website within a date range. A visit encompasses a set of interactions within the website (e.g. multiple page views).
# of new visits	Estimated number of the first-time visits.
# of unique visitors	Number of unduplicated (counted only once) visitors to the website within a date range.
# of page views	Total number of pages viewed, including repeated views of a single page.
Pageviews per visit	The ratio of total number of page viewed to number of visits.
Average visit duration	Average duration of a visit measured in seconds.
Bounce rate	Percentage of single-page visits (users who visit only one page of the website and then leave)

Note. Descriptions provided by Google Analytics.

TABLE 3
METRICS DESCRIBING ACTIVITIES OF USERS USING THE IADB

Metric	Description
# of logins	Number of times users log into the IA website within a date range
# of IA projects created	Number of IA projects created by users within a date range.
# of IA public projects created	Number of IA projects published within a date range.
# of IA projects copied	Number of IA projects copied from others within a date range.
# of online resources used	Number of online resources added to the IA projects within a date range

In contrast, the IADB records the activities of individual users who have registered for an account in the IA website. Using this data, we defined three categories of IA users for a particular time period: *lurkers*, who did not create IA projects; *contributors*, who created but did not share IA projects; and *active contributors*, who created and shared IA projects. In this analysis, we focused on two user types – *lurkers* and *active contributors*, and analyzed five metrics collected by the IADB capturing the activities of these users (see Table 3).

Also note that based on users' activities collected by IADB and GA, we assume that *lurkers* and *active contributors* are primarily teachers, while *IA visitors* come for the general Internet user base.

V. RESULTS

A. RQ1: Influence of Dissemination Activities on IA Visitors

Using the analytics from GA, Figures 2-8 compare the activities of IA visitors during the "active dissemination" and "no dissemination" periods (averaged monthly over the time period). Table 4 compares the activities of IA visitors averaged daily over these two time periods. The comparisons are made in terms of key GA analytics: the number of visits, new visits, unique visitors, *pageviews*, and *pageviews* per visit, as well as visit duration and bounce rate.

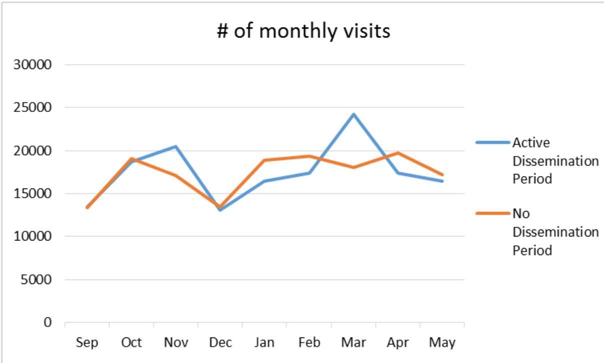


Fig. 2. Number of visits

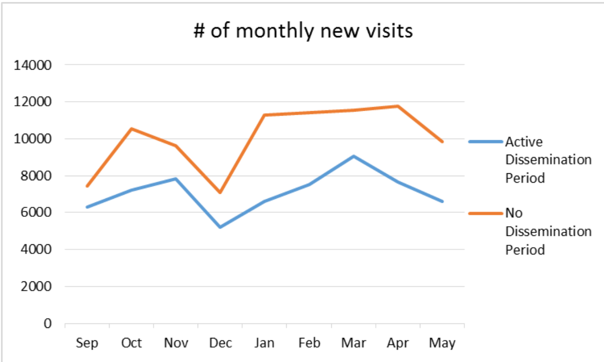


Fig. 3. Number of new visits

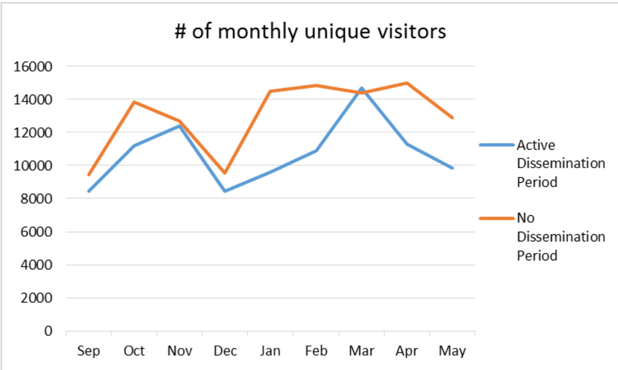


Fig. 4. Number of unique visitors

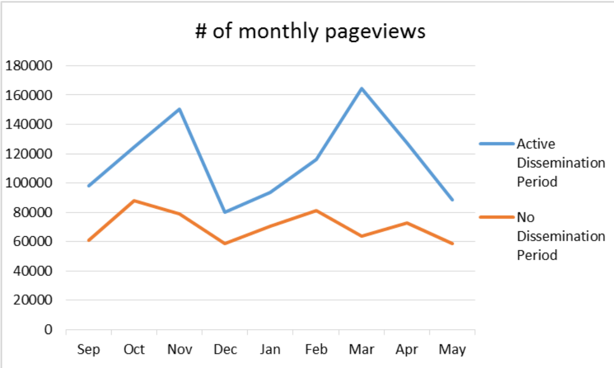


Fig. 5. Number of *pageviews*

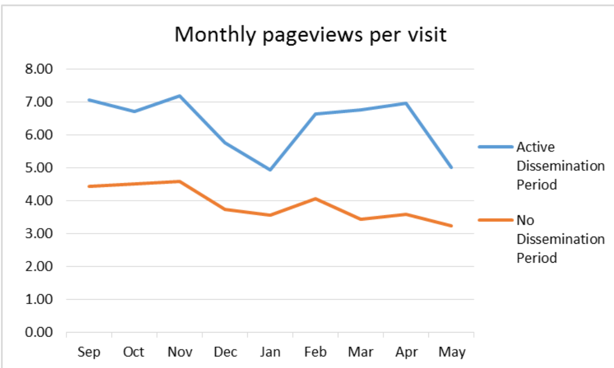


Fig. 6. Number of *pageviews* per visit

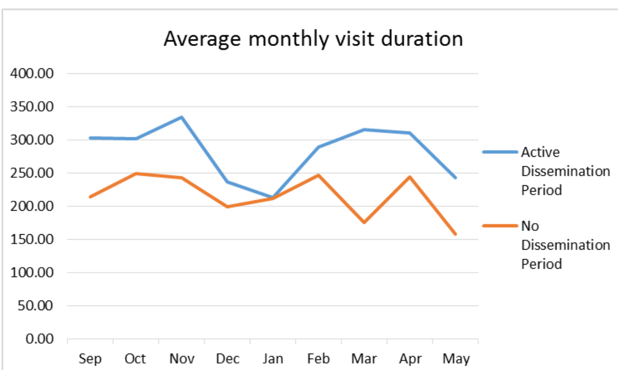


Fig. 7. Visit duration (measured in seconds)

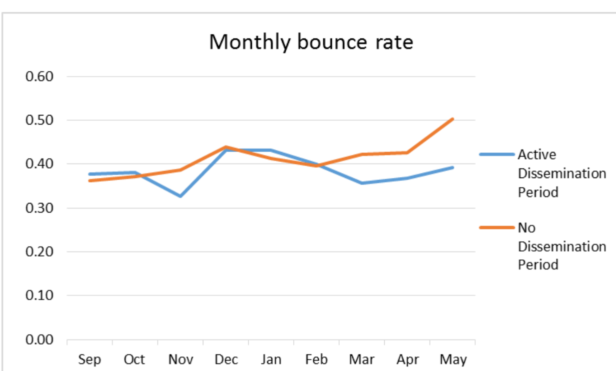


Fig. 8. Bounce rate

Due to non-normal distributions of the data, the Mann-Whitney test was used to compare whether visitors' activities between these two time periods were significantly different. As suggested by Figure 2, the overall number of visits did not differ significantly between these two time periods (U = 37129.50, p = .94). This suggests that dissemination activities had little effect on the overall number of visitors.

However, the number of new visits, the number of unique visitors, and the bounce rate all increased significantly during the subsequent "no dissemination" period (U = 24773.50, p < .001; U = 29879.00, p < .001; U = 29916.00, p < .001). In contrast, the number of *pageviews* and *pageviews* per visit, as well as average visit duration decreased significantly (U = 26917.50, p < .001; U = 12897.50, p < .001; U = 25048.00, p < .001). Taken together, these results suggest that while the overall number of visits stayed even between periods, the "active dissemination" period was characterized by more engaged visitors.

Note that one task of the IA is to help users to find useful online resources, and thus many IA projects contain links that lead users to resources outside the IA website (therefore inflating the bounce rate). It is plausible that the subsequent, "no dissemination" period was populated by more savvy users, who were quickly able to find desired resources. This would help explain the overall similar number of visits, coupled with decreased number of *pageviews*, visit duration, and higher bounce rate during this period.

TABLE 4
COMPARISON OF IA VISITORS' ACTIVITIES BETWEEN TWO TIME PERIODS

	Active dissemination			No dissemination		
	Mean	Median	SD	Mean	Median	SD
# of visits	577.70	641.00	364.73	572.48	607.00	338.52
# of new visits *	234.38	233.00	116.34	331.55	331.00	177.89
# of unique visitors *	355.03	371.00	193.73	429.33	442.00	236.35
# of *pageviews**	3821.04	3351.00	2937.54	2321.77	2345.00	1569.38
Pageviews per visit *	6.32	5.88	2.60	3.90	3.72	1.19
Average visit duration (seconds) *	282.39	265.51	125.30	215.04	206.74	98.20
Bounce rate *	.39	.38	.09	.41	.42	.08

* Difference between the two time periods is significant (Mann-Whitney test; p < .05)

B. RQ2: Influence of Dissemination Activities on Lurkers and Active Contributors

Table 5 compares the number of lurkers and active contributors between the two time periods, using analytics from the IADB. Recall that *lurkers* are defined as users who created an IA account but did not create any IA projects during the given time period. *Active contributors* are defined as users who created and shared IA projects during the given time period.

As shown in Table 5, after the dissemination activities ended, the number of lurkers increased while the number of active contributors decreased. Note that the number of active contributors in the "active dissemination" period was about six times greater than during the "no dissemination" period, suggesting that dissemination activities may have helped encourage users' active participation. The large drop of active contributors during the "no dissemination" period may exacerbate the free riding problems, as only a very small portion of users contributed IA projects during this period. In addition, a large increase can be seen in the lurker-active contributor ratio. This suggests that ceasing dissemination activities can lead to a more skewed lurker-active contributor ratio and thus aggravate participation inequality.

TABLE 5
THE NUMBER OF USERS IN EACH CATEGORY

	Active dissemination	No dissemination
# of lurkers	3908	6201
# of active contributors	547	92
Lurker-active contributor ratio	7 : 1	66 : 1

Evolution of Lurkers

Table 6 compares lurkers' mean number of logins between the two time periods, which significantly decreased after dissemination activities ceased (U = 1.20, $p < .001$). This suggests that during the subsequent "no dissemination" period, lurkers were less likely to log in, and thus less likely to make use of features in the IA community.

As can also be seen, the number logins for lurkers was very low. Note that one function of the IA community is to facilitate teachers' browsing existing IA project, and login is not required to view IA projects. As such, the low number of logins does not necessarily mean that lurkers viewed fewer projects or became inactive – they may simply have chosen to view IA projects without logging in. Unfortunately, our analytics do not enable us to track visitors who do not log in at the individual user level.

Evolution of Active Contributors

Compared to the "active dissemination" period, all five metrics for active contributors at the aggregated level declined during the "no dissemination" period. As can be seen from Table 7, they had fewer logins, created fewer IA projects, shared fewer IA projects, copied fewer IA projects from other users, and used fewer online resources in their IA projects.

However, a closer examination of active contributors' individual activities revealed a different picture. As shown in Table 8, during the "no dissemination" period, active contributors on average had significantly fewer logins (U = 21531.00, $p < .05$), and used significantly fewer online resources (U = 21269.00, $p < .05$). However, each active contributor on average created significantly more IA projects, shared significantly more of these, but copied significantly less (U = 18826.00, $p < .001$; U = 16673.50, $p < .001$; U = 20974.50, $p < .001$).

TABLE 6
COMPARISON OF LURKERS' MEAN # OF LOGINS BETWEEN TWO TIME PERIODS

	Active dissemination			No dissemination		
	Mean	Median	SD	Mean	Median	SD
# of logins	.09	0	1.01	.04	0	.96

TABLE 7
COMPARISON OF ACTIVE CONTRIBUTORS' ACTIVITIES AT AGGREGATE LEVEL BETWEEN TWO TIME PERIODS

	Active dissemination	No dissemination
# of logins	3440	474
# of IA projects created	1890	399
# (%) of IA public projects created	1194 (63%)	310 (77%)
# (%) of IA projects copied	422 (22%)	18 (4%)
# of online resources used	6509	1017

TABLE 8
COMPARISON OF ACTIVE CONTRIBUTORS' ACTIVITIES **AT INDIVIDUAL LEVEL** BETWEEN TWO TIME PERIODS

	Active dissemination			No dissemination		
	Mean	Median	SD	Mean	Median	SD
# of logins *	6.29	4.00	7.60	5.15	3.00	6.59
# of IA projects created *	3.46	2.00	3.82	4.34	5.00	2.91
# of IA public projects created *	2.18	1.00	2.92	3.37	2.50	2.60
# of IA projects copied *	.77	0	1.79	.20	0	.47
# of online resources used *	11.90	8.00	16.59	11.05	12.00	8.63

* Difference between the two time periods is significant (Mann-Whitney test; $p < .05$)

In sum, after the dissemination activities ended, the number of active contributors significantly declined, with a corresponding decline in the number of IA projects created, shared, and copied, and resources used. However, the remaining active contributors on average increased their levels of engagement in the community by creating and sharing significantly more IA projects.

VI. DISCUSSION AND CONCLUSION

This article described a study that examined the analytics automatically collected by usage logs in order to compare user activity patterns in an online educational community during and after dissemination activities. This study first provided an overall view of the community by exploring changes in *IA visitors*' activities during the two time periods. Second, this study focused on two types of IA users – *lurkers* and *active contributors* – and compared the dynamics of their activities in the community during the two time periods.

In comparing activities of IA visitors between the "active dissemination" and "no dissemination" period, we noted that the number of new visits and the number of unique visitors increased. This suggests that even though dissemination activities ended, the IA website attracted a growing number of new visitors and thus increased its audience size. This also suggests that users continue to find the IA online community useful for their tasks.

However, during the subsequent "no dissemination" period, the number of *pageviews*, *pageviews* per visit, and average visit duration decreased -- IA visitors viewed fewer IA projects and spent less time per visit. This could suggest that IA users are becoming more efficient in discovering information they desire. Alternatively, it could indicate that many IA projects were not visited, which makes content discovery a problem. Thus, the IA developers may consider user interface enhancements to recommend IA projects to users, so as to increase the number and variety of IA projects viewed by users [24].

We then compared users who have created an account in the IA in terms of two types of users: lurkers and active contributors. During the subsequent "no dissemination" period, the lurkers' number of logins decreased significantly, suggesting that they were less likely to consider themselves as members of IA community [11].

In comparing active contributors during the two time periods, we found that the number of active contributors dropped considerably during the "no dissemination" period. This resulted in an overall decrease in the amount of new content created in the community. However, on average, the remaining active contributors were much more engaged: they created more IA projects, and shared a higher percent of their IA projects. Thus while participation inequality increased after dissemination, the remaining active contributors were, plainly stated, more engaged contributors.

In sum, dissemination activities appear to play an important role in encouraging users' active participation in the IA community. With the absence of dissemination, while the overall number of visitors did not decrease, the lurker-active contributor ratio increased in the IA community. That is, participation inequality increased. Yet, those that remained active were more engaged contributors. Thus, at least for the IA community, it appears that dissemination is important in decreasing participation inequality and in increasing lurkers' sense of community, thereby contributing to the sustainability of the online community.

In conclusion, this study contributes to our understanding of how dissemination activities can influence the evolution of different user types in an online community. In addition, it shows how different kinds of analytics data can be used to help understand the dynamics of different user types. This, in turn, can help inform strategies for attracting new users, increasing the loyalty of existing users, and improving existing communities [12]. However, as this study only focused on one online educational community and contrasted user analytics during two relatively short time periods (9 months each), future research is needed.

ACKNOWLEDGMENT

The authors would like to thank the many users of the Instructional Architect. This material is based upon work supported by the National Science Foundation under Grant No. 0937630, and Utah State University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Portions of this research were previously presented at the 11th International Conference of the Learning Sciences (ICLS) in Boulder, Colorado, USA.

REFERENCES

[1] S. Abramovich, C.D. Schunn, and R. J. Correnti, "The role of evaluative metadata in an online teacher resource exchange," *Educational Technology Research and Development*, vol. 61, no. 6, pp. 863-883, 2013.

[2] D. E. Atkins, J. S. Brown, and A. L. Hammond (2007). A review of the open educational resources (OER) movement: Achievements, challenges, and new opportunities. Available: http://www.hewlett.org/Programs/Education/OER/OpenContent/Hewlett+OER+Report.htm

[3] C. L. Borgman, H. Abelson, L. Dirks, R. Johnson, K. R. Koedinger, M.C. Linn, C. A. Lynch, D. G. Oblinger, R. D. Pea, K. Salen, M. S. Smith, A. Szalay, "Fostering learning in the networked world: The cyberlearning opportunity and challenge, a 21st century agenda for the National Science Foundation". Report of the NSF Task Force on Cyberlearning. Virginia, US: NSF, 2008.

[4] M. Recker, A. Walker, S. Giersch, X. Mao, S. Halioris, B. Palmer, ... and M. B. Robertshaw, "A study of teachers' use of online learning resources to design classroom activities," *New Review of Hypermedia and Multimedia*, vol. 13, no. 2, pp. 117-134, 2007.

[5] M. Recker, "Perspectives on teachers as digital library users: Consumers, contributors, and designers," *D-Lib Magazine*, vol. 12, no. 9, 2006. Available: http://www.dlib.org/dlib/september06/recker/09recker.html

[6] T. Sumner and CCS Team, "Customizing science instruction with educational digital libraries," In *2010 Proceedings of the 10th annual joint conference on Digital libraries*, pp. 353-356.

[7] M. Schlager, U. Farooq, J. Fusco, P. Schank, P., and N. Dwyer, "Analyzing online social networking in professional learning communities: Cyber networks require cyber-research tools". *Journal of Teacher Education*, vol. 60, no. 1, pp. 86-100, 2009.

[8] R. J. Windle, H. Wharrad, D. McCormick, H. Laverty, and M. Taylor, (2010). Sharing and reuse in OER: experiences gained from open reusable learning objects in health. *Journal of Interactive Media in Education* [online]. 2010 (01). Available: http://jime.open.ac.uk/jime/article/viewArticle/2010-4/html

[9] J. Preece, "Sociability and usability in online communities: Determining and measuring success," *Behaviour & Information Technology*, vol. 20, no. 5, pp. 347-356, 2001.

[10] C. L. Hsu, and H. P. Lu, "Consumer behavior in online game communities: A motivational factor perspective," *Computers in Human Behavior*, vol. 23, no. 3, pp. 1642-1659, 2007.

[11] B. Nonnecke, D. Andrews, and J. Preece, "Non-public and public online community participation: Needs, attitudes and behavior," *Electronic Commerce Research*, vol. 6, no. 1, pp. 7-20, 2006.

[12] K. Panciera, R. Priedhorsky, T. Erickson, and L. Terveen, "Lurking? cyclopaths?: a quantitative lifecycle analysis of user behavior in a geowiki," In *2010 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1917-1926.

[13] J. Bishop, "Increasing participation in online communities: A framework for human–computer interaction", *Computers in human behavior*, vol. 23, no. 4, pp.1881-1893, 2007.

[14] J. Nielsen, (2006). Participation Inequality: Lurkers vs. Contributors in Internet Communities [online]. Available: http://www.nngroup.com/articles/participation-inequality/.
[15] B. Xu and M. Recker, "Teaching Analytics: A Clustering and Triangulation Study of Digital Library User Data", *Educational Technology & Society Journal*, vol. 15, no. 3, pp. 103-115, 2012.
[16] K. E. Maull, M. G. Saldivar, and T. Sumner, "Online Curriculum Planning Behavior of Teachers". In *2010 proceedings of EDM*, pp. 121-130.
[17] D. Maloney-Krichmar and J. Preece, "A multilevel analysis of sociability, usability, and community dynamics in an online health community," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 12, no. 2, pp. 201-232, 2005.
[18] A. Iriberri and G. Leroy, "A life-cycle perspective on online community success," *ACM Computing Surveys (CSUR)*, vol. 41, no. 2, pp. 11-40, 2009.
[19] B. Nonnecke and J. Preece, *Shedding light on lurkers in online communities*, Ethnographic Studies in Real and Virtual Environments: Inhabited Information Spaces and Connected Communities, Edinburgh, 1999, pp. 123-128.
[20] D. Andrews, J. Preece, and M. Turoff, M, "A conceptual framework for demographic groups resistant to online community interaction," in *2001 proceedings of the 34th Annual Hawaii International Conference on system sciences*, pp. 10-20.
[21] T. Elias, "Learning analytics: Definitions, processes and potential. Learning," *Learning*, vol. 23, pp. 134-148, 2011.
[22] G. Siemens and R. S. Baker, "Learning analytics and educational data mining: towards communication and collaboration," In *2012 proceedings of the 2nd international conference on learning analytics and knowledge*, pp. 252-254.
[23] R. Ferguson and S. B. Shum, "Social learning analytics: five approaches," In *2012 proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pp. 23-33.
[24] A. Bakharia, E. Heathcote, E., and S. Dawson, "Social networks adapting pedagogical practice: SNAPP," In: *Same Places, Different Spaces. Ascilite 2009*.
[25] J. Preece, and D. Maloney-Krichmar, "Online Communities". In J. Jacko and A. Sears, A. (Eds.) Handbook of Human-Computer Interaction, Lawrence Erlbaum Associates Inc. Publishers. Mahwah: NJ. pp. 596-620, 2003.
[26] M. Schlager, J. Fusco, and P. Schank, "Evolution of an online education community of practice". In K. A. Renninger and W. Shumar (Eds.), Building virtual communities: Learning and change in cyberspace, New York: Cambridge University Press, pp. 129-158, 2002.
[27] R. Farzan, J. M. DiMicco, and B. Brownholtz, "Spreading the honey: a system for maintaining an online community," In *2009 proceedings of the ACM 2009 international conference on Supporting group work*, pp. 31-40.
[28] F. Ortega, J. M. Gonzalez-Barahona, and G. Robles, "On the inequality of contributions to Wikipedia". In 2008 proceedings of the 41st Hawaii International Conference on System Sciences, pp. 304-304.
[29] X. Ochoa and E. Duval, "Quantitative analysis of user-generated content on the Web," In *2008 Proceedings of webevolve2008: web science workshop at WWW2008*, pp. 1-8.
[30] M. Feldman and J. Chuang, "Overcoming free-riding behavior in peer-to-peer systems," *ACM SIGecom Exchanges*, vol. 5, no. 4, pp. 41-50, 2005.
[31] L. Ramaswamy and L. Liu, "Free riding: A new challenge to peer-to-peer file sharing systems". In *2003 proceedings of the 36th Annual Hawaii International Conference on system sciences*, pp. 10-20.
[32] M. Recker, J. Dorward, D. Dawson, S. Halioris, Y. Liu, X. Mao, ... and J. Park, " You can lead a horse to water: teacher development and use of digital library resources". In *2005 Proceedings of the Joint Conference on Digital Libraries*, pp. 1–7.
[33] B. Xu and M. Recker, "Peer Production of Online Learning Resources: A Social Network Analysis," In *2010 Baker, R., Merceron, A., Pavlik, P.I. Jr. (Eds.). Proceedings of the 3rd International Conference on Educational Data Mining, pp. 315-316*.
[34] H. Leary, S. Giersch, A. Walker, and M. Recker, "Developing a Review Rubric for Learning Resources in Digital Libraries," In *2009 Proceedings of the Joint Conference on Digital Libraries, pp. 421-422*.
[35] L. Sellers, L. Ye, B. Robertshaw, M. Recker, and A. Walker, "Technology Integrated Professional Development: A Case Study of Junior High Science and Mathematics Teachers," presented at the Annual Meeting of the Associationfor Educational Communications Technology, Jacksonville, FL, Nov, 2011.
[36] A. Walker, M. Recker, B. Robertshaw, J.Olsen, L. Sellers, H. Leary, Y. Kuo, and L. Ye, "Designing For Problem Based Learning: A Comparative Study Of Technology Professional Development," presented at the Annual Meeting of the American Education Research Association, New Orleans, LA, Apr, 2011.

Min Yuan is currently pursuing her Ph.D from Utah State University, Logan, UT. Min got her bachelor's and master's degree from Nanjing Normal University, China. After that, she worked for four years at Yingtian College in China, and then started her Ph.D study at Utah State in 2011. Her research interests include teachers' evaluation of online resources, and online communities. She has several research papers in reputed international journal and conference.

Mimi Recker is Professor and Head of the Department of Instructional Technology & Learning Sciences, Utah State University, Logan, Utah. Mimi has a bachelor's degree in mathematics from the University of Pennsylvania. After a few years as a software engineer in Silicon Valley, she earned her PhD from the University of California, Berkeley. Mimi worked for two years at the Georgia Institute of Technology, and four years at Victoria University in New Zealand, then came to Utah State in 1998. Her research focuses on helping the education sector reap the benefits of cyber-learning. Her goals are to help provide teachers and learners with access to a network of high-quality, interactive and free online learning resources. Her research, largely funded by the National Science Foundation, has involved collaborations with a dynamic mix of faculty, post-doctoral students, and graduate students from Utah State University, as well as colleagues from around the world.

Personality and Education Mining based Job Advisory System

Rajendra S. Choudhary[1], Rajul Kukreja[1], Nitika Jain[1], Shikha Jain[1]

[1]*Jaypee Institute of Information Technology, Noida, India.*

***Abstract* — Every job demands an employee with some specific qualities in addition to the basic educational qualification. For example, an introvert person cannot be a good leader despite of a very good academic qualification. Thinking and logical ability is required for a person to be a successful software engineer. So, the aim of this paper is to present a novel approach for advising an ideal job to the job seeker while considering his personality trait and educational qualification both. Very well-known theories of personality like MBTI indicator and OCEAN theory, are used for personality mining. For education mining, score based system is used. The score based system captures the information from attributes like most scoring subject, dream job etc. After personality mining, the resultant values are coalesced with the information extracted from education mining. And finally, the most suited jobs, in terms of personality and educational qualification are recommended to the job seekers. The experiment is conducted on the students who have earned an engineering degree in the field of computer science, information technology and electronics. Nevertheless, the same architecture can easily be extended to other educational degrees also. To the best of the author's knowledge, this is a first e-job advisory system that recommends the job best suited as per one's personality using MBTI and OCEAN theory both.**

***Keywords* — E-job advisory system, Education mining, job recommendation system, Personality mining.**

I. Introduction

In today's world of internet so-called e-world, searching a dream job is very tedious. Lots of information, job proposals and job search engines are available. If we try to search one job, we will get a huge list of recommendation, but still finding out the most ideal and suitable one, that we will truly enjoy in the future, is very difficult.

The online job portals like naukri.com, indeed.com etc. provides two-way search options: one from the recruiter side where recruiter can filter the candidates who are matched with their expectations in terms of educational qualification. And the second is for the job seekers who upload their CV and look for the jobs best fit to their qualification. However, a recent research conducted by Cave [7] has shown that although, there is a range of skill sets involved in any tech career, the personality also puts a significant impact on the output generated at the workplace and the society, in general. For any job, the employer looks for candidate with some specific abilities. The discipline you studied is only one of the factors in the consideration. Every job demands an employee with some special qualities. For example, numbers of options are available for an (IT) engineering students like in our field of software development, testing, marketing, academics, research etc. The minimum educational qualification for all these jobs is an engineering degree in the field of computer science or information technology. As per survey conducted to see the recent trend, more than 42% of the students want to go for MBA after completing engineering degree, 40% want to become software developer and remaining wants to opt for testing, academics or others. This is what we call as dream career. But how much successful they become on getting their dream career? This is a big question. Research [2], [3], [4], [6], [7], [8], [10] conducted in this field has shown that the performance at the workplace highly depends upon your personality along with your knowledge. According to John Perry, a Senior Server Engineer and IT Architect at City of Mesa, Arizona [7], "Being an INTJ in an IT leadership position has been beneficial; this is due to the fact that we are pre-wired to create efficient organizations and systems just by our very nature."

Unfortunately, the personality traits which is an important attribute to decide job success, is missing in all widely used job recommending tools. To address this missing attribute, we present a novel approach for finding a relationship between the personality traits, education qualification and job for a candidate. The objective is to recommend the ideal jobs to a job seeker while conducting personality mining and education mining both. This will help in increasing the level of job satisfaction among employees and a better performance at the workplace. For personality mining, the job seeker has to fill a questionnaire. The design of this questionnaire is based upon Myers-Briggs Type Indicator (MBTI) [1]. On the basis of the response of the candidate, MBTI score is obtained. Then, these results are mapped on to the five personality dimensions (Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism) as suggested by OCEAN theory [9]. Why not OCEAN theory directly? In the model, MBTI is used to evaluate the personality of job seeker because the MBTI based questionnaire is available in renowned literatures. MBTI to

DOI: 10.9781/ijimai.2014.274

OCEAN mapping is done because as compared to MBTI, the dimensions proposed by OCEAN theory are closer to the personality traits required for suggesting jobs and are used by human resource team of any organization to describe the personality of an employee. Finally, the extracted personality is combined with the score of educational qualification of the candidate to suggest him best suited job rank-wise. In this paper, we have conducted the experimentation on students who have an earned engineering degree in the field of computer science, information technology and electronics. However, the same architecture can easily be extended to other educational degrees also.

To the best of the author's knowledge, this is a first e-job advisory system that recommends the job best suited as per one's personality and knowledge both.

The rest of this work is organized as follows. Section II discusses the personality theories used. Literature work is given in section III. Next our proposed model is described in section IV, followed by the implementations and results in section V and finally concluding the paper with future scope in section VI.

II. BACKGROUND

Personality mining refers to predicting the personality of a person based upon some feedback, questionnaire or using some social networking sites like Facebook, LinkedIn etc. In our proposed system, two personality theories are used: MBTI and OCEAN theory.

A. The Myers-Briggs Type Indicator (MBTI)

MBTI [1] assessment is a set of psychometric questionnaire designed to measure psychological preferences that is how people perceive the world and make decisions. In MBTI, personality traits are calculated in 8 different factors.

Extraversion (E)-(I) Introversion: Extrovert and Introvert refer to where people focus their attention to get their energy – either the OUTER or INNER world. Extroverts act first and think/reflect later whereas introverts think/reflect first and act later.

Sensing (S)-(N) Intuition: Sensing and intuition are the information-gathering (perceiving) functions. They describe how new information is understood and interpreted. Sensing people mentally live in the Present, attending to present opportunities. Intuitive people mentally live in the Future, attending to future possibilities.

Thinking (T)-(F) Feeling: Thinking and feeling are the decision-making (judging) functions. Thinking people instinctively search for facts and logic in a decision situation. Feeling people instinctively employ personal feelings and impact on people in decision situations.

Judgment (J)-(P) Perception: People also have a preference for using either the judging function (thinking or feeling) or their perceiving function (sensing or intuition) when relating to the outside world (extraversion). Judging people plan many of the details in advance before moving into action. Perceiving people are comfortable moving into action without a plan; plan on-the-go.

B. OCEAN Theory

OCEAN [9] is also known as Big Five personality theory. It suggests five broad domains or dimensions of personality that are used to describe human personality. *Openness:* It refers to the number of interests to which on is attracted and the depth to which those interests are pursued. High openness refers to a person with relatively more interests and, consequently, relatively less depth within each interest, while low openness refers to a person with relatively few interests and relatively more depth in each of those interests.

Conscientiousness: It refers to goal-directed behavior. High conscientiousness refers to a person who focuses intensely on his/her goals and exhibits the self-discipline associated with such focus. Low conscientiousness refers to one who is disorganized and distracted.

Extraversion: It refers to the number of relationships with which one is comfortable. High extraversion is characterized by a larger number of relationships and a larger proportion of one's time spent in enjoying them. Low extraversion is characterized by a smaller number of relationships and a smaller proportion of one's time spent in pursuing those relationships.

Agreeableness: It refers to one's general interpersonal orientation. High agreeableness describes a person who reacts to others with warmth and will bend to avoid conflict. Low agreeableness describes one who, in the extreme, only follows one's inner voice regardless of hurting others.

Neuroticism: Contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad, and tense. Emotional Stability refers to one's proneness to negative emotions and anxiety.

MBTI is a standard scale and a fixed MBTI questionnaire to measure the personality trait. So, in this paper we have first extract the personality from the feedback form of the job seeker on MBTI scale and then it is mapped over the OCEAN theory to interpret MBTI findings within a broader, more commonly shared conceptual framework.

III. RELATED WORK

For the last few decades, many researchers are working in the field of job recommendation system for job seekers and e-recruitment system for employers to make the system more intelligent and smart. In 2007, Meo [8] has proposed XML based multi-agent recommender system for online recruitment services. They have used XML as standard mechanism for user information representation. They also evaluated their work performance by doing comparisons from previous known results. But their system does not include job seekers personality rather they are recommending on the basis of seekers educational qualification only. Lounsbury [6] related personality traits and career satisfaction of human resource professionals, defining that optimism, emotional resilience, assertiveness and extraversion are powerful traits of managers or human resource professionals. Those with low levels of these traits were recommended to be coached to develop optimism-enhancing, to learn defensive pessimism strategies or engage in counseling or stress management programs. Nagarjuna [12] conducted a study for comparing engineers with commerce students, shows that engineering students are more self-reliant, realistic, responsible and emotionally tough. Engineering students are more socially aware, controlled, self-disciplined and perfectionists as compared to the commerce background students. Faliagka [3] proposed an online recruitment system using personality factors and ranked job seekers according to Analytic Hierarchy Process (AHP). Faliagka [2] proposed another recruitment system where they evaluated the performance of job seekers by extracting data from LinkedIn and blogs. But the system was not efficient as a very few people write blogs and this should not be used as personality mining criteria. Secondly, the proposed model is for recruiter only not for the job seekers. Moreover, in 2014, Cave [7] wrote an article on how he used MBTI theory in his organization. With the help of various examples, he explained the role of MBTI personality types in IT careers.

The conclusion is that there is a relationship between personality trait and ideal job for a job seeker to have better output at workplace. Some of the researchers have recognized the importance of personality and proposed the personality based e-recruitment system. However, no one has worked for job advisory system for job seekers.

IV. Our Model: E-PE Job

E-PE Job is a Job Advisory System based upon personality and education mining. The system suggests automated ranked jobs based on a set of criteria which can help the job seeker to search for a most appropriate job. In the current system, our main focus is on two criteria for job selection: personality mining and educational qualification. The system architecture is shown in figure 1. The whole system is arranged in three phases as explained below:

1) In the first phase, a set of questions are given from MBTI questionnaire to the job seeker to get information for personality analysis. The educational skills are extracted through online form or CV filled by the candidate. Applicant fills an online form with his most scoring subject and the subject of his interest. Candidate's desired job profiles are also asked.
2) In the second phase, data analysis is done from data gathered in phase 1. Firstly, the response to these questions is analyzed to extract the personality traits of the candidate. These questions extract different factors of personality traits of MBTI theory. After that MBTI score is mapped over five personality dimensions (Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism) as suggested by OCEAN theory. This is done because OCEAN five factors are the collection of characteristics found in nearly all personality and psychological tests and it gives more specific and more detailed personality traits of any candidate. The coefficient values for mapping MBTI to OCEAN (Big Five Personality traits) have been taken from [11]. The conversion process is as follows:

   Let

   MB_i : MBTI score vector for 8 personality traits
   M_i : vector that stores difference in values of (E-I), (S-N), (T-F), and (J-P)
   MO_{ij} : 2-D matrix store the MBTI to OCEAN mapping coefficients(as shown in table 1).
   C_{ij} : intermediate result
   O_j : OCEAN score vector

   Then

   $$M_i = MB_{2*i-1} - MB_{2*i} \quad \text{--- (1)}$$
   $$C_{ij} = M_i * MO_{ij} \quad \text{--- (2)}$$
   $$O_j = \Sigma_{i\text{-}> 1 \text{ to } 4} C_{ij} \quad \text{--- (3)}$$

Finally the value of different factors of OCEAN is obtained in the vector O_j.

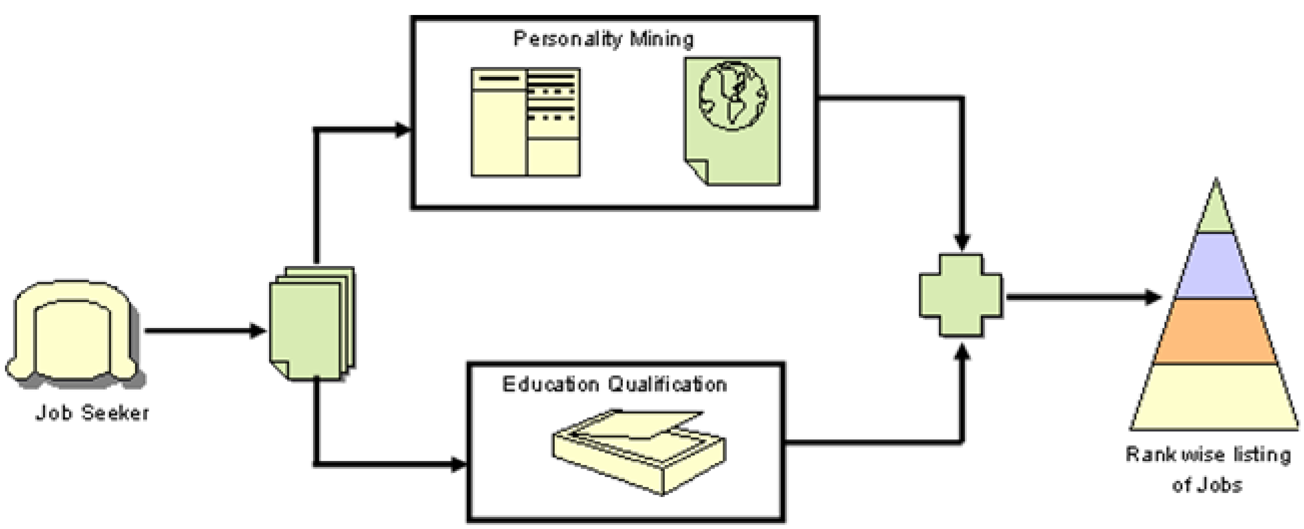


Fig 1 : E-PE Job : Proposed system architecture

TABLE I
CORRELATION FACTORS FOR MAPPING MBTI OVER OCEAN

	Extra-version	Openn-ess	Agree-ableness	Conscien tiousness	Neuro-ticism
E-I	-0.74	0.03	-0.03	0.08	0.16
S-N	0.10	0.72	0.04	-0.15	-0.06
T-F	0.19	0.02	0.44	-0.15	0.06
J-P	0.15	0.30	-0.06	-0.49	0.11

Moreover, based upon the education choices filled as shown in figure 2, different weight factors are assigned. For example, a weight factor of 40, 45 or 50 is assigned based on educational analysis. All desired jobs are given some priority and hence are given a weight of 40. Moreover, if most scored subject is same as subject of his interest, weight 50 is given otherwise they are given weight of 45 each. Finally, all the scores are summed up to get the aggregated score of educational qualification (E).

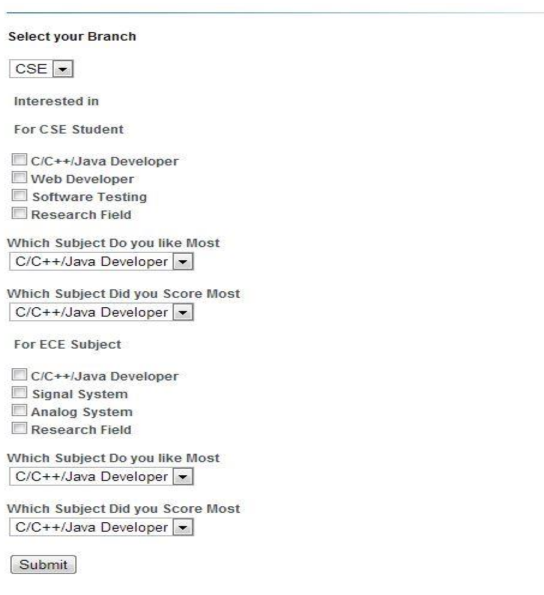


Fig 2: Snap shot of Questionnaire

3) In the third phase, coalescence is to be done by combining both educational analysis and personality traits analysis. Reference [5] has given the relationship between various parameters of the OCEAN theory with engineering degree based job profiles. In the proposed model, we have used this relationship as Wij.

To find the weight factor corresponding to each job profile (JO_i),

$JO_i = \Sigma_{i\text{-}>1 \text{ to } 5}((O_j + E)/ W_{ij})$ --- (4)

Finally, two weight factors, one calculated by the educational qualification and second by the personality mining are combined together (using equation 4) to get the compatibility index of the applicant with the corresponding job.

TABLE II
OCEAN SCORE MAPPING WITH ENGINEERING DEGREE BASED JOB PROFILES

	O	C	E	A	N
Web Developer	67.65	54.07	47.87	54.76	62.34
Software Developer	54.32	53.67	56.45	46.78	52.75
R & D	63.59	58.65	67.84	61.67	58.63
Data Analyst	45.56	47.66	51.53	56.45	46.85
Content Writer	64.67	56.43	61.29	52.43	57.93
Hardware Engineer	61.56	55.67	59.76	57.36	45.73

V. IMPLEMENTATION

The proposed system is implemented in real world to investigate its validation. The purpose of the investigation is to check the precision of personality mining and recommendation of job. The experimentation is conducted on the students who have just completed their engineering degree in computer science or information technology and are looking for the job. Here we are showing how the suitable job can be recommended to any job seeker.

Step wise implementation for one candidate is as follows:

1. *Data Collection*: Firstly, the candidate is asked to fill the MBTI based questionnaire in the specified time. Time limit is enforced so that the candidate cannot bluff. Moreover, each question is compulsory. Questions are picked randomly from the database from each category. After that he is asked to fill the education qualification details along with most scoring subject and the subject of his interest. The candidate's desired job profiles is also asked.
2. *Experimental Result:* Depending upon the response of the candidate, score is obtained for each of the eight attributes of MBTI. The value of the score varies between 4 and 20. The table 3 shows the MBTI score obtained by the candidate.

TABLE III
MBTI SCORE OF A CANDIDATE

Extrovert	13	Introvert	11
Sensing	14	Intuitive	13
Thinking	13	Feeling	11
Judging	15	Perceiving	12

3. Then, the obtained MBTI score is mapped over OCEAN theory using the table 1. This is done because dimensions proposed by OCEAN theory are closer to the personality traits required suggesting jobs than MBTI. The result after mapping is shown in figure 3 as a screen shot.
4. Moreover, score for educational qualification is obtained while considering education qualification, most scoring subject, subject of his interest and his desired job. All desired jobs are weighted 40. If most scored subject is same as subject of his interest, weight 50 is given otherwise they are given weight of 45 each.

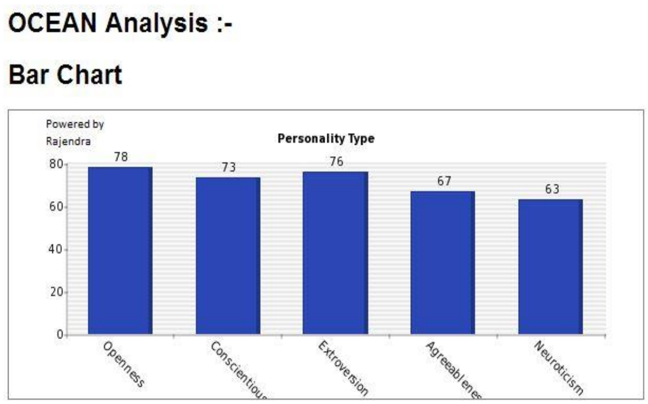


Fig 3. Result of MBTI to OCEAN Mapping for a candidate.

5. Finally, scores from personality mining and educational mining are coalesced together (as explained earlier) to find compatibility index of different jobs as per the recommendation of our model. Final output is shown in figure 4.

JOB CATEGORY	COMPATIBILITY INDEX
WEB DEVELOPER	78%
SOFTWARE DEVELOPER	75%
R & D	67%
DATA ANALYST	64%
CONTENT WRITER	62%
HARDWARE ENGINEER	66%

Fig 4. Screen shot of priority wise job recommendation for a candidate having engineering degree.

To validate the model, we collected a corpus data of 100 job seekers and estimated their personality and educational qualification in the same manner as explained above. Then, a list of six ideal jobs prioritize on the basis of compatibility index was suggested to all 100 applicants. Thereafter, they were asked for the feedback that how they feel about the suggested job. Is it actually the most suited one? In the feedback collection it was found that 51% applicants agreed with first job in the priority list, 22% applicant agreed with second job, 18% agreed with one of job between third to sixth option and remaining 9% disagreed with the recommendation.

VI. Conclusion and Future Scope

In the paper, E-PE Job (a Job Advisory System) based upon personality and education mining is proposed. The system implements automated ranking of jobs based on a set of criteria which can help the job seeker to search for a most appropriate job. The objective is to recommend the ideal jobs to a job seeker while conducting personality mining and education mining both. The model is successfully validated in the real world scenario. We have conducted the experimentation on 100 students who have earned an engineering degree in the field of computer science, information technology. However, the same architecture can easily be extended to other educational degrees also.

Next, we are working to build the same on a larger platform with much larger database. Moreover, we are working for extracting the personality factors more accurately by getting the data from some social networking sites.

References

[1] B. I. Myers, M. H. McCaulley, "A Guide to the Development and Use of the Myers-Briggs Type Indicator" (2nd ed.), Consulting Psychologists Press, 1995.

[2] E. Faliagka and A. Tsakalidis, "An integrated e-recruitment system for automated personality mining and applicant ranking", Internet Research, vol. 12(5), 2012, pp 551- 568.

[3] E. Faliagka, L. Kozanidis, S. Stamou, A. K. Tsakalidis, G. Tzimas, "A Personality Mining System for Automated Applicant Ranking in Online Recruitment Systems", ICWE , 2011, pp 379-382.

[4] F. Tian, S. Wang, C. Zheng, Q. Z. Abbrev, "Research on E-learner Personality Grouping Based on Fuzzy Clustering Analysis", CSCWD, 2008, pp 1035-1040.

[5] G. M. Luis, J. R. Castro, G. Licea, A. R. Diaz, and R. Salas, "Implementing Fuzzy Subtractive Clustering to Build a Personality Fuzzy Model Based on Big Five Patterns for Engineers", MICAI 2013, Part II, LNAI 8266, 2013, pp. 497–508.

[6] J. W. Lounsbury, R. P. Steel, L. W. Gibson, A. W. Drost, "Personality traits and career satisfaction of human resource professionals", Human Resource Development International, vol. 11(4), 2008, pp 351–366.

[7] K. Cave, "Leadership challenges for MBTI confirmed introverts", Available http://www.idgconnect.com/blog-abstract/7917/leadership-challenges-mbti-confirmed-introverts, 2014.

[8] P. Meo, g. Quattrone, G. Terracina and D. Ursino, "An XML-Based Multi-agent System for Supporting Online Recruitment Services"., IEEE Transactions on Systems, Man, and Cybernetics, vol. 37(4), 2007, pp.464-480.

[9] P. T. Costa Jr. and R. R. McCrae, "Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual", Odessa, FL: Psychological Assessment Resources, 1992.

[10] Q. Yang, J. Wang, "Research on Learners' Personality Mining Based on Improved Decision Tree Algorithm", In Proc of Second International Conference on Genetic and Evolutionary Computing, 2008.

[11] R. R. McCrae, P. T. Costa Jr, "Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality", Journal of Personality, vol. 57, 2008, pp 17-40.

[12] V. L . Nagarjuna, S. Mamidenna, "Personality Characteristics of Commerce and Engineering Graduates – A Comparative Study", Journal of the Indian Academy of Applied Psychology, vol. 34(2), 2008, pp 303–308.

Rajendra S. Choudhary was a graduate student at Jaypee Institute of Information Technology, Noida, India. His area of interest is e- learning and recommendation systems

Rajul Kukreja is a graduate student at Jaypee Institute of Information Technology, Noida, India. His area of interest is artificial intelligence and data mining.

Nitika Jain is a graduate student at Jaypee Institute of Information Technology, Noida, India. Her area of interest coginitive science, personality mining, e- learning.

Shikha Jain is assistant professor in Jaypee Institute of Information Technology, Noida, India. Her research area are affective computing, intelligent agents, information retrieval.

Big Data and Learning Analytics in Blended Learning Environments: Benefits and Concerns

Anthony G. Picciano[1]

[1] Graduate Center and Hunter College, City University of New York (CUNY)

***Abstract* — The purpose of this article is to examine big data and learning analytics in blended learning environments. It will examine the nature of these concepts, provide basic definitions, and identify the benefits and concerns that apply to their development and implementation. This article draws on concepts associated with data-driven decision making, which evolved in the 1980s and 1990s, and takes a sober look at big data and analytics. It does not present them as panaceas for all of the issues and decisions faced by higher education administrators, but sees them as part of solutions, although not without significant investments of time and money to achieve worthwhile benefits.**

***Keyworkds* — Blended learning, data-driven decision making, big data, learning analytics, higher education, rational decision making, planning.**

I. Introduction

IN May 2014, I was at North-West University in South Africa to lecture and conduct workshops on blended learning in higher education. The topics of my workshops related to conducting research in instructional technology, design of blended learning environments, MOOCs, and technology planning. For the technology planning session, administrators at North-West University shared with me a document that outlined its plan for integrating more technology, and specifically blended learning, into its academic programs. Among the strategies to be considered was the effective use of learning analytics to profile students and track their learning achievements in order to:

- identify at-risk students in a timely manner;
- monitor student persistence on a regular basis; and
- develop an evidence base for program planning and learner support strategies.

During the session, I was specifically asked to give my opinion about whether North-West University should invest in learning analytics technology at this time.

On May 28, 2014, one week after I returned to my home institution at the Graduate Center of the City University of New York (CUNY), I received an email from the University Director of Academic Technology, asking if I would comment on a white paper entitled, Blackboard Analytics and CUNY. This white paper outlined the potential for implementation of learning analytics software into the University's course/learning management system. It was sent to members of a committee examining the feasibility of this software for the university. The email specifically asked for one of the following responses:

- I think this is worth pursuing.
- I don't think this is worth pursuing.
- I am not sure.

and to provide comments in support of the choice.

North-West University and CUNY are 8,000 miles apart, on different continents, with very different missions, organizational structures, and academic programs. Yet, with regard to the acquisition and development of learning analytics software they were pretty much in the exact same situation.

Instructional technology is at the center of many discussions on college and university campuses across the globe. The Internet has permeated every aspect of our societies by its ubiquity, and has changed higher education as well. Online and blended learning, specifically, are being utilized with increasing regularity and are changing the way instruction is provided. In the United States, more than seven million students, approximately one-third of the higher education population, were enrolled in fully online college courses in 2013. [1] Millions more are enrolled in blended courses, although precise data on the extent of blended learning in American higher education is not to be found because of problems with definition and accurate data reporting at the individual college level. Precision aside, the changes brought on by online access to instruction is affecting the way our colleges and universities are being administered. Infusions of technology infrastructure, large-scale databases, and demands for timely data to support decision making have seeped into all levels of college leadership and operations. Data-driven decision making is evolving into a vastly more sophisticated concept known as big data which relies on software approaches generally referred to as learning analytics. Big data and learning analytics for instructional applications are still evolving and will take a few years to mature, although

DOI: 10.9781/ijimai.2014.275

their presence is already being felt and cannot be ignored. While big data and learning analytics are not panaceas for all of the issues and decisions being faced by higher education administrators, the hope is that they can become part of solutions and gracefully integrated into administrative and instructional functions. The purpose of this article is to examine the evolving world of big data and learning analytics in blended learning environments. Specifically, it will look at the nature of these concepts, provide basic definitions, and identify the benefits and concerns related to their development, implementation, and growth in higher education environments.

Administrative decision making processes have been evolving for decades and as more data were made available from integrated information systems, decisions became more rational, using data to support alternative courses of action. A new phenomenon, generally termed online learning, emerged in the 1990s and the early 2000s, that changed the way many faculty teach and students learn. As mentioned earlier, millions of students are learning online and entire colleges have been "built" to offer the entirety of their academic programs online. In addition, for most institutions, online technology is being integrated with face-to-face instruction in what is commonly being referred to as blended learning. The utilization of data-driven decision making in online learning environments has opened up new approaches and avenues for collecting and processing data on students and course activities whereby instructional transactions can be immediately recorded and added to an institutional database. Academic administration and evaluation, which in the past occurred away from the classroom, can now be integrated more closely into instructional activities.

II. Blended Learning

Blended learning environments present unique challenges to implementing learning analytics mainly because they have so many different facets and are difficult to define. They combine face-to-face instruction and online technology in myriad ways.

Blended learning is not one thing but comes in many different flavors, styles, and applications. It means different things to different people. The word "blended" implies a mixture rather than simply an attaching of components. When a picture is pasted above a paragraph of text, a presentation is created that may be more informative to the viewer or reader, but the picture and text remain intact and can be individually discerned. On the other hand, when two cans of different colored paints are mixed, the new paint will look different from either of the original colors. In fact, if the new paint is mixed well, neither of the original colors will continue to exist. Similar situations exist in blended learning. The mix can be a simple separation of part of a course into an online component. For instance, in a course that meets for three weekly contact hours, two hours might take place in a traditional classroom while the equivalent of one weekly hour is conducted online. The two modalities for this course are carefully separated, and although they may overlap, they can still be differentiated. In other forms of blended courses and programs, the modalities are not so easily distinguishable. Consider an online program that offers three online courses in a semester that all students are required to take. The courses meet for three consecutive five week sessions. However, students do a collaborative fifteen-week project that overlaps the courses. The students are expected to maintain regular communication with one another through email and group discussion boards. They are also required to meet face-to-face once a month on Saturdays where course materials from the online courses are further presented and discussed and some sessions are devoted to group project work. These activities begin to blur the modalities in a new mixture or blend where the individual parts are not as discernable as they once were. Add to this the increasing popularity of integrating videoconferencing, podcasting, YouTube videos, wikis, blogs, and other media into class work and the definition of blended learning becomes very fluid.

In the broadest sense, blended learning (see Figure 1) can be conceptualized as a wide variety of technology/media integrated with conventional, face-to-face classroom activities. This conceptualization serves as a guideline and should not be viewed as an absolute, limiting declaration. Also, it can apply to entire academic programs as well as individual courses.

III. Data-Driven Decision Making, Big Data, and Learning Analytics

The focus of this article is technology-based approaches that support decision making in blended learning environments. The simplest definition of the popular term "data-driven decision making" is the use of data analysis to inform courses of action involving policy and procedures. Inherent in this definition is the development of reliable and timely information resources to collect, sort, and analyze the data used in the decision making process. It is important to note that data analysis is used to inform and does not mean to replace entirely the experience, expertise, intuition, judgment, and acumen of competent educators. While decision making may be singly defined as choosing between or among two or more alternatives, in a modern educational organization, decision making is an integral component of complex management processes such as academic planning, policy making, and budgeting. These processes evolve over time, require participation by stakeholders, and most importantly, seek to include information which will help all those involved in the decision process.

Fundamental to data-driven decision making is a rational model directed by values and based on data. It is well-recognized, however, that a strictly rational model has limitations. An individual commonly associated with this concept and whose work is highly recommended for further reference, is Herbert Simon [2,3,4,5,6]. Simon was awarded the Nobel Prize in economics in 1978 for his research on decision making in organizations. His theory on the limits of

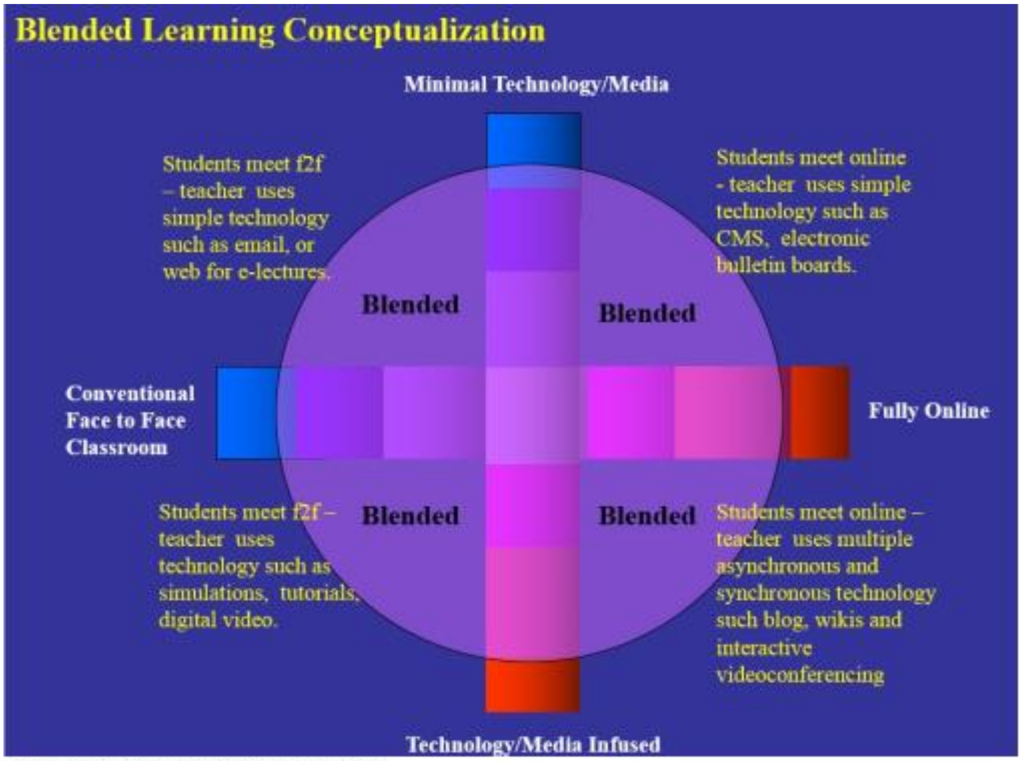


Fig. 1. Blended Learning Conceptualization

rationality, later renamed "bounded rationality," has as its main principle that organizations operate along a continuum of rational and social behaviors mainly because the knowledge necessary to function strictly according to a rational model is beyond what is available. Although first developed in the 1940s, this theory has withstood the test of time and is widely recognized as a fundamental assumption in understanding organizational processes such as decision making and planning [7,8,9]. More recently, modern computerized information systems are facilitating and instilling a greater degree of rationality in decision making in all organizations including colleges and universities. They support organizations and help them to adjust, adapt, and learn in order to perform their administrative functions. [10] While these systems are not replacing the decision maker, they surely are helping to refine the decision-making process.

Figure 2 illustrates the basic data-driven decision-making process. It assumes that decision making in education environments is fundamentally part of a social process. It also assumes that an information system is available to support the decision process, that internal and external factors not available through the information system are considered, and that a course or courses of action are determined. The information system in Figure 2 is a computerized database system capable of storing, manipulating, and providing reports from a wide variety of data. The decision process concludes with decision makers reflecting on and evaluating their decisions.

Terms related to data-driven decision making include data warehousing, data mining, and data disaggregation. Data warehousing essentially refers to a database information system that is capable of storing, integrating and maintaining large amounts of data over time. It might also involve multiple database systems. Data mining is a frequently used term in research and statistics which refers to searching or "digging into" a data file for information to understand better a particular phenomenon. Data disaggregation refers to the use of software tools to break data files down into various characteristics. An example might be using a software program to select student performance data by gender, by major, by ethnicity, or by other definable characteristics.

In recent years, two other terms, big data and analytics, have become important. Big data is a generic term that assumes that the information or database system(s) used as the main storage facility is capable of storing large quantities of data longitudinally and down to very specific transactions. For example, college student record keeping systems have maintained outcomes information on students such as grades in each course. This information could be used by institutional researchers to study patterns of student performance over time, usually from one semester to another or one year to another. In a big data scenario, data would be collected for each student transaction in a course, especially if the course was delivered electronically online. Every student entry on a course assessment, discussion board entry, blog entry, or wiki activity could be recorded, generating thousands of transactions per student per course. Furthermore, this data would be collected in real or near real time as it is transacted and then analyzed to suggest courses of action. Analytics software is evolving to assist in this analysis.

The generic definition of analytics is similar to data-driven decision making. Essentially it is the science of examining data to draw conclusions and, when used in decision making, to present paths or courses of action. In recent years, the definition of analytics has gone further, however, to incorporate elements of operations research such as decision trees and strategy maps to establish predictive models and to determine probabilities for certain courses of action. It uses data mining software to establish decision processes that convert data into actionable insight, uncover patterns, alert and respond to issues and concerns, and plan for the future. This might seem to be an overly complicated definition but the term "analytics" has been used in many different ways in recent years and has become part of the buzzword jargon that sometimes seeps into new technology applications and products. Goldstein and Katz (2005) in a study of academic analytics admitted that they struggled with coming up with a name and definition that was appropriate for their work. They stated that they adopted the term "academic analytics" for their study but that it was an "imperfect label."[11] Alias (2011) defined four different types of analytics that could apply to instruction including web analytics, learning analytics, academic analytics, and action analytics. [12] The trade journal, Infoworld, referred to analytics as:

> "One of the buzzwords around business intelligence software…[that]…has been through the linguistic grinder, with vendors and customers using it to describe very different functions.
>
> The term can cause confusion for enterprises, especially as they consider products from vendors who use analytics to mean different things…" [13]

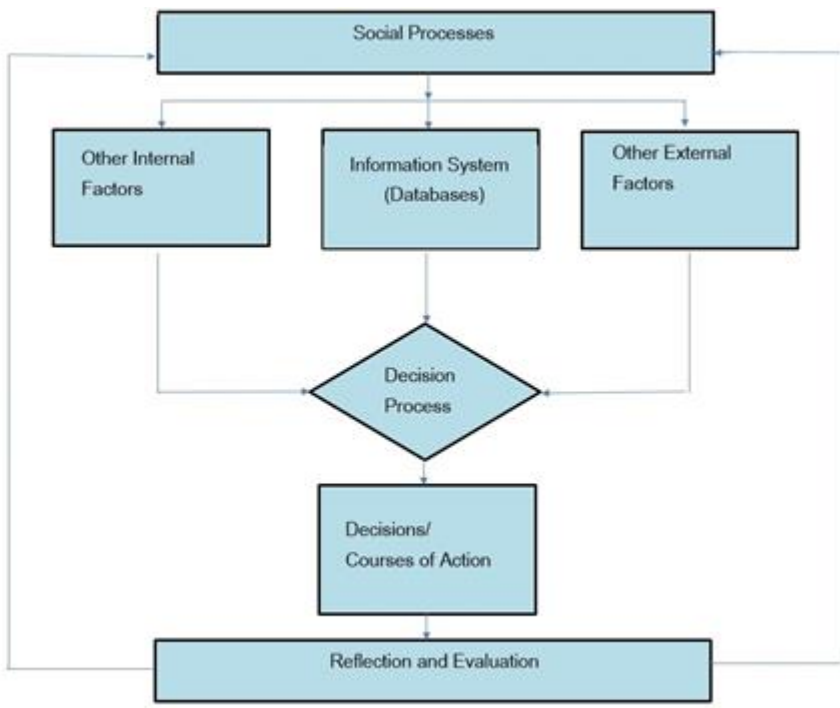


Fig. 2. The Data-Driven Decision-Making Process

Critical to the definition of analytics is the use of data to determine courses of action especially where there is a high volume of transactions. Common examples of analytics applications are examinations of Website traffic, purchases, or navigation patterns to determine which customers are more or less likely to buy particular products (i.e., books, movies) by ecommerce companies such as amazon.com or Netflix. Using these patterns, companies send personalized notifications to customers as new products become available. In higher education, analytics are beginning to be used for a number of applications that address student performance, outcomes, and persistence.

Big data concepts and analytics can be applied to a variety of higher education administrative and instructional applications including recruitment and admissions processing, financial planning, donor tracking and student performance monitoring. This article will focus on teaching and learning, and hence will specifically examine learning analytics.

To take advantage of big data and learning analytics, it is almost a requirement that transaction processing be electronic rather than manual. Traditional face-to-face instruction can support traditional data-driven decision-making processes, however, to move into the more extensive and time-sensitive learning analytics applications, it is important that instructional transactions are collected as they occur. This would be possible within a course management/learning management system (CMS/LMS). Most CMS/LMSs provide constant monitoring of student activity whether they are responses, postings on a discussion board, accesses of reading material, completions of quizzes, or some other assessment. Using the full capabilities of a basic CMS/LMS, a robust fifteen week online course would generate thousands of transactions per student. Real-time recording and analysis of these transactions could then be used to feed a learning analytics application. Not waiting for the end of a marking period or semester to record performance measures is critical to this type of application. Monitoring student transactions on a real-time basis allows for real-time decisions. Instructors may take actions or intervene in time to alert or assist students. A CMS/LMS or something similar therefore becomes critical for collecting and feeding this data into a "big" database for processing by a learning analytics software application. These instructional transactions should also be integrated with other resources such as student, course, and faculty data from the college information systems. Analytics software can then be used to analyze these transactions to establish patterns that are used to develop guidelines and rules for subsequent courses of action (see Figure 3). An important caveat is that the data accuracy should never be compromised in favor of timeliness of the data. Both accuracy and timeliness are required and need to be present in the learning analytics application.

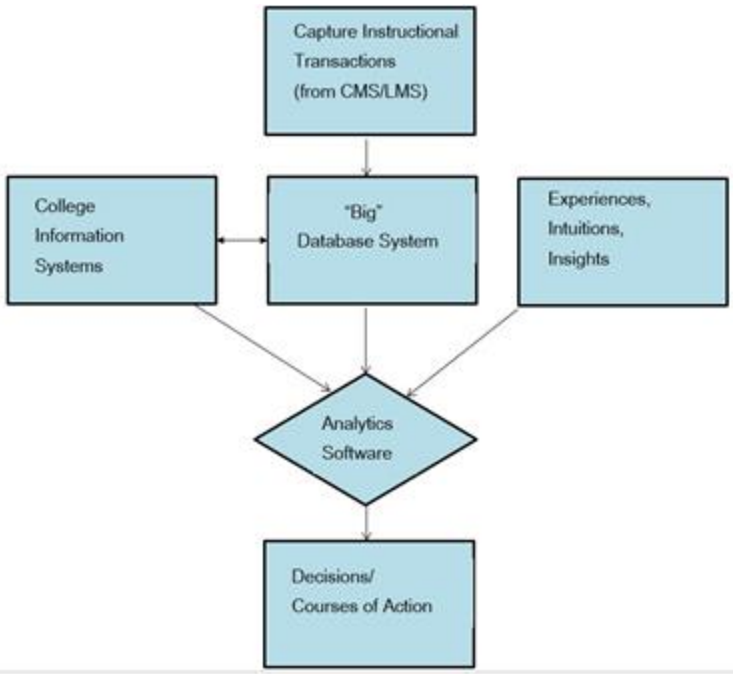


Fig. 3. Learning Analytics Flow Model

In a white paper published by IBM entitled Analytics for Achievement, eight categories of possible instructional applications utilizing analytics were described. The eight categories are as follows:

1. Monitoring individual student performance
2. Disaggregating student performance by selected characteristics such as major, year of study, ethnicity, etc.
3. Identifying outliers for early intervention
4. Predicting potential so that all students achieve optimally
5. Preventing attrition from a course or program
6. Identifying and developing effective instructional techniques
7. Analyzing standard assessment techniques and instruments (i.e. departmental and licensing exams)
8. Testing and evaluation of curricula. [14]

Of the above, monitoring individual student performance and course participation in a course is among the most popular type of learning analytics application. Anyone who has ever taught (face-to-face or online) will monitor student participation to determine engagement with the course material. Taking attendance is a time-honored classroom activity and most instructors will become concerned about students who have too many absences. Grades on quizzes and papers are also frequently monitored. A conscientious instructor will review his/her records and meet with those students who are not meeting the standards for the course. Many colleges have instituted mid-term reviews that provide students with indicators of their progress in a course. In online courses, CMS/LMSs routinely provide course monitoring statistics and rudimentary early warning systems that allow instructors to follow up with students who are not responding on blogs or discussion boards, not accessing reading materials, or not promptly taking quizzes. These course statistics are maintained in real-time and instructors can review them as often as they wish. Students who are not as engaged as they should be can be sent emails expressing concerns about their performance. None of these interventions requires learning analytics, however, these interactions can be enhanced significantly by expanding the amount and nature of the data collected. For example, a single student response on a discussion board can be analyzed through pattern recognition to determine the depth and quality of student engagement with the course material. The pattern used in this type of analysis are uncovered by examining thousands and tens of thousands of other student responses and evaluating sentences and phrases.

Examples of well-designed learning analytics-based student monitoring systems are Rio Salado Community College's Progress and Course Engagement (PACE) system, Northern Arizona University's Grade Performance System (GPS), and Purdue University's Course Signals System.

For purposes of this article, the Course Signals System, in particular, is a good example of learning analytics software because it is one of the first to be used in blended learning environments. It combines demographic information with online student interactions and produces a red, yellow or green light to show students how well they are doing in their courses -- and also provides that information to their professors who can intervene if necessary . Developed originally at Purdue University, Course Signals was licensed to SunGard Higher Education (now Ellucian) in 2010 to make it available to other colleges and universities. Large CMS/LMS providers such as Desire2Learn and Blackboard have modeled their own retention early warning systems after Purdue's work. [15] It has won a number of awards including the Campus Technology Innovators Award, Digital Education Achievement Award, and the Lee Noel and Randi Levitz Retention Excellence Awards. Course Signals has been used in online, face-to-face, and blended learning environments. It has been particularly popular in large-section size, blended, and flipped classroom courses. [16] While there have been several studies supporting the use of learning analytics software such as Course Signals for improving student retention, more research needs to be done. [17] Michael Caulfield, director of blended and networked learning at Washington State University at Vancouver, cautioned that the early research on the effectiveness of learning analytics on retention needs further verification and review. [18] The fact is that learning analytics as a tool for retention is still in its nascent stage. The Society for Learning Analytics Research (SoLAR) is an inter-disciplinary network of leading international researchers who are exploring the role and impact of analytics on teaching, learning, training and development. This society was established in 2011 and has held four conferences to discuss issues related to learning analytics research. To provide a vehicle for documenting the research, SoLAR established The Journal of Learning Analytics, a peer-reviewed, open-access journal, for disseminating research in this field. It provides a research forum within what George Siemens, the president of SoLAR calls "the messiness of science". [19] The first edition was published in June, 2014. The articles in this first edition address issues such as scaling- up learning analytics initiatives, the relationship between LMS/VLE usage and learning performance, the role of psychometric data to predict academic achievement, and the capacity to detect boredom through user log- data. All of these, while important, are just beginning to scratch the surface of effectiveness of learning analytics with respect to student performance and retention. Furthermore, there is practically no research that does cost-benefit comparisons of the large-scale implementation of learning analytics in blended learning or face-to-face environments. In sum, there is a long road ahead for researchers in this field and much study to be done.

IV. Benefits and Concerns

The New Horizon Report is published each year by The New Media Consortium and EDUCAUSE. It predicts six emerging technologies that are likely "to enter mainstream use" over the next five years. In the 2014 Report, the six technologies in rank order were identified as follows:

1. Growth of Social Media
2. Integration of Online, Blended and Collaborative Learning
3. Rise of Data-Driven Learning and Assessment
4. Shift from Students as Consumers to Students as Creators
5. Agile Approaches to Change
6. Evolution of Online Learning [20]

The ranking of these six technologies indicates that the first two will likely enter the mainstream in one to two years; the second two within three years; and the last two within five years or more. The Rise of Data-Driven Learning and Assessment (referring to learning analytics) was ranked third and indicates that this technology has potential and that widespread adoption is projected to be about three years away. This ranking also indicates that learning analytics need more exploration at this time and refinements before their adoption.

A. Benefits

Learning analytics can have significant benefits in monitoring student performance and progress. First, and at its most basic level, learning analytics software can mine down to the frequency with which individual students access a CMS/LMS, how much time they are spending in a course, and the number and nature of instructional interactions. These interactions can be categorized into assessments (tests, assignments, or exercises), content (articles, videos, or simulations viewed) and collaborative activities (blogs, discussion groups, or wikis).

Second, by providing detailed data on instructional interactions, learning analytics can significantly improve academic advisement related directly to teaching and learning. Learning analytics can improve the ability to identify at-risk students and intervene at the first indication of trouble. Furthermore by linking instructional activities with other student information system data (college readiness, gender, age, major), learning analytics software is able to review performance across the organizational hierarchy: from the student, to courses, to department, to the entire college. It can provide insights into individual students as well as the learning patterns of various cohorts of students.

Third, learning analytics software is able to provide longitudinal analysis that can lead to predictive behavior studies and patterns. By linking CMS/LMS databases with an institution's information system, data can be collected over time. Student and course data can be aggregated and disaggregated to analyze patterns at multiple levels of the institution. This would allow for predictive modeling that in turn, can create and establish student outcomes alert systems and intervention strategies.

In sum, learning analytics can become an important element in identifying students who are at risk and alerting advisors and faculty to take appropriate actions. Furthermore, it can do so longitudinally across the institution and can undercover patterns to improve student retention that in turn, can assist in academic planning.

B. Concerns

First, in order for big data and learning analytics applications to function well, data need to be accurate and timely. Learning analytics software works best for courses that are delivered completely electronically such as online courses. Traditional face-to-face courses that require significant data conversion time are problematic. Blended learning courses (part face-to-face and part online) likewise present data collection problems. Because blended learning courses vary so much in the nature of their delivery, learning analytics software can have significant data gaps. Instructional transactions that take place in the face-to-face environment will be lost unless the faculty member or teaching assistant is willing to manually enter them into the student information system.

The second, and perhaps the most serious concern, is that since learning analytics require massive amounts of data collected on students and integrated with other databases, colleges need to be mindful of privacy, data profiling, and the rights of students in terms of recording their individual behaviors. While college classes have always involved evaluating student performance and academic behavior, learning analytics take the recording of behavior to a whole new level and scope. As well-intentioned as learning analytics might be in terms of helping students succeed, this "big data" approach may also be seen as "big brother is watching" and, as such ,an invasion of privacy that some students would find objectionable. Precautions need be taken to ensure that the extensive data collection of student instructional transactions is not abused in ways that potentially hurt individuals. Vicky Gunn, Director of the Learning and Teaching Centre at the University of Glasgow, advises:

> "… it is clear that the growth of learning analytics needs a few up-front protocols of protection as soon as possible.. We should especially be considering…Ethical consent structures to enable students to know what is being gathered, when and how it will be used as well as opportunities for students to opt in/out." [21]

Third, there are not yet enough individuals trained to use big data and analytics appropriately. Experienced database administrators and designers capable of warehousing and integrating data across multiple files and formats are a necessity. In addition to the expertise needed to develop databases, instructional designers working with faculty will need to understand and derive insights into the student behaviors that are pertinent to the application at hand. There is also a need for institutional researchers, or others knowledgeable about statistics, decision trees, and strategy mapping, to develop algorithms that construct predictive models. College administrators may have to invest in consultants or undertake extensive professional development of their own staffs in order to develop appropriate applications. This will take time and additional resources and may or may not be worth the return on investment. Furthermore, because of the dearth of expertise, there may be a tendency to use instructional templates that are integrated into CMS/LMSs. These, although convenient, may be overly simplistic and should be considered with caution.

Fourth, a good deal of college and university student data may end up in larger governmental databases either at the state or national level. Bennett (2011) cautions that the United States is heading to an all-inclusive national K-20 database. [21] Federal education policies as promulgated by No Child Left Behind and Race to the Top funding have pushed many states to adopt comprehensive statewide student databases that could easily be the basis for establishing a national system. Furthermore, there is a certain amount of influence being exerted on the part of the U.S. Department of Education in favor of development of common database structures. Such a system might be beneficial but may also leave individuals vulnerable to privacy, data security and theft issues. In 2013, the people of the United States were awakened to the spying activities of the National Security Agency (N.S.A.) and the intelligence arms of other governments around the world. The problem became so bad that large Internet service companies such as Google, Facebook, and Yahoo invested hundreds of millions of dollars to seal up security systems that Edward J. Snowden revealed the N.S.A. had been exploiting. After years of cooperating with the U.S federal government, the goal of many of these companies is to thwart Washington as well as Beijing and Moscow. The users of "big data" and analytics need to be careful that these mega-database systems do not become the playground of exploitative individuals and organizations. [23]

Lastly, it might be beneficial to revisit the work of Herbert Simon and his theory on the limits of rational decision making that was mentioned earlier in this article. Herbert Simon was a life-long supporter of the use of computer technology to support decision making, including the application of artificial intelligence. At Carnegie-Mellon University where he taught for decades he was active in integrating artificial intelligence software in the learning sciences to improve instruction. Instructional data-driven decision making and learning analytics parallel Simon’s work in this area. In his honor, Carnegie-Mellon University established the Simon Initiative in 2013 to accelerate the use of learning science and technology to improve student learning. This initiative harnesses CMU's decades of learning data and research to improve educational outcomes for students. However, as database systems become bigger and as software such as learning analytics becomes more complex, a case can be made that the limits of rational decision making are being exceeded because of the plethora of information and data available. Simon was highly focused on the efficient use of data and is famously quoted as saying that too much information can consume its recipients and that “… a wealth of information creates a poverty of attention…” [24] Simon’s quote may be a most appropriate concern in the era of big data and learning analytics. Nathan Silver, an American statistician, echoed Simon in his 2012 bestseller, The Signal and the Noise…., and cautioned that in predictive models, there is a tendency to collect a lot of meaningless data (i.e., noise) creating the danger of poor predictions. [25]

V. Conclusion

This article started with a reference to two scenarios, one at North-West University in South Africa and one at the City University of New York. These institutions are very different, yet are facing similar decisions with regard to investing and acquiring learning analytic software. They are also similar in that, while they have some fully online academic programs, both are presently and for the foreseeable future investing heavily in blended learning. My recommendation to both institutions was that before committing to learning analytics they do a careful analysis of the costs related to acquiring this software. These would include not only direct costs such as software licenses and maintenance contracts, but also indirect costs to hire personnel and/or consultants to design and implement learning analytics applications. It would also include the feasibility and cost of data collection in blended learning environments where faculty or other personnel would be needed to provide accurate and timely data.

Colleges and universities around the world need to meet a number of challenges related to providing greater access to higher education. However, expanding access does not necessarily lead to expanding resources. To the contrary, higher education policy makers, while calling for more access, are limiting resources and instructional technology such as online and blended learning is being seen as an important vehicle for expanding access while containing costs. In its truest sense, expanded access does not just mean getting acceptance into college programs; it also means successful completion of degrees. Student attrition in many colleges and universities is at unacceptable levels and needs to be addressed as well. Data-driven decision making and learning analytics software have the potential to assist colleges in identifying and evaluating strategies that can improve retention. At the present time, however, these software are best suited for fully online environments, not face-to-face or blended learning environments. Nevertheless, as data-driven decision making enters the big data and learning analytics era, these new approaches, while not silver bullets, may be part of the solution. Higher education administrators would do well to consider the benefits, concerns, and costs iterated above when evaluating whether big data and learning analytics can be used in their institutions and determining the exact role they can play.

References

[1] Allen, I. E. & Seaman, J. (2014). “Grade Change: Tracking Online Learning Education in the United States.” Needham, MA: Babson Survey Research Group. http://www.onlinelearningsurvey.com/reports/gradechange.pdf
[2] Accessed June 6, 2014.
[3] Simon, H. A. (1945). Administrative Behavior. New York: Macmillan.
[4] Simon, H. A. (1957). Administrative Behavior (2nd ed.). New York: Macmillan.
[5] Simon, H. A. (1960). The New Science of Management Decision. New York: Harper & Row.
[6] Simon, H. A. (1979). Rational Decision Making in Business Organizations. American Economic Review, 69, 493–513.
[7] Simon, H. A. (1982). Models of Bounded Rationality. Cambridge, MA: MIT Press.
[8] Tyson, C. (2002). The Foundations of Imperfect Decision Making. Stanford, CA: Stanford Graduate School of Business Research Paper Series.
[9] Carlson, R. V., & Awkerman, G. (Eds.). (1991). Educational Planning: Concepts, Strategies and Practices. New York: Longman.
[10] Senge, P. M. (1990). The Fifth Discipline: The Art & Practice of the Learning Organization. New York, NY: Doubleday Currency.
[11] Dibello, A.J. & Nevis, E.C. (1998). How Organizations Learn. San Francisco: Jossey-Bass.
[12] Goldstein, P.J. & Katz, R. N. (2005). Academic Analytics: The Uses of Management Information and Technology in Higher Education. Boulder, CO: EDUCAUSE Center for Applied Research.
[13] Alias, T. (2011). Learning Analytics: Definitions, Processes, and Potential. Unpublished paper. http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf
[14] Kirk, J. (February 7, 2006). 'Analytics' Buzzword Needs Careful Definition. Infoworld. http://www.infoworld.com/t/data-management/analytics-buzzword-needs-careful-definition- 567 Accessed: June 3, 2014.
[15] IBM Software Group (2001). Analytics for Achievement. Ottawa, Ontario. http://public.dhe.ibm.com/common/ssi/ecm/en/ytw03149caen/YTW03149CAEN.PDF. Accessed: June 10, 2014.
[16] Feldstein, M. (November 2013). Course Signals Effectiveness Data Appears to be Meaningless (and Why You Should Care). Blog posting on e-Literate Website. http://mfeldstein.com/courssignals-effectiveness-data-appears-meaningless-care/ Accessed: August 9, 2014
[17] Iten, L., Arnold, K., Pistilli, M. (March 2008). Mining Real-time Data to Improve Student Success in a Gateway Course. Paper presented at the Eleventh Annual TLT Conference. West Lafayette, IN: Purdue University. http://www.itap.purdue.edu/learning/docs/research/TLT%2008%20presentation%20summary.pdf Accessed: August 9, 2014
[18] Pistilli, M.D., Arnold, K. & Bethune, M. (November 18, 2012). Signals: Using Academic Analytics to Promote Student Success. EDUCAUSE Review http://www.educause.edu/ero/article/signals-using-academic-analytics-promote-student-success. Accessed August 8, 2014.
[19] Straumsheim, C. (November 6, 2013). Mixed Signals. Inside Higher Education.

https://www.insidehighered.com/news/2013/11/06/researchers-cast-doubt-about-early-warning-systems-effect-retention. Accessed: August 9, 2014.
[20] Siemens, G. (June 2014). The Journal of Learning Analytics: Supporting and Promoting Learning Analytics Research. Journal of Learning Analytics, 1(1). http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/3908
[21] Accessed: August 11, 2014.
[22] Johnson, L., Adams-Becker, S., and Estrada, V. & Freeman, A. (2014). The NMC Horizon Report: 2014 Higher Education Edition. Austin, Texas: The New Media Consortium.
[23] Gunn, V. (May 7, 2014). Learning Analytics, Surveillance, and the Future of Understanding our Students. News Blog posting of the Society for Research in the Higher Education. http://srheblog.com/2014/05/07/learning-analytics-surveillance-and-the-future-of-understanding-our-students/ Accessed: August 11, 2014.
[24] Bennett, E. (2011). Moving Toward a National Education Database. Unpublished paper.
[25] Sanger, D.E. & Perlroth, N. (June 6, 2014). Internet Giants Erect Barriers to Spy Agencies. New York Times. http://www.nytimes.com/2014/06/07/technology/internet-giants-erect-barriers-to-spy-agencies.html?_r=0.Accessed: June 15, 2014.
[26] Simon, H.A. (1971). Designing Organizations for an Information-rich World in Martin Greenberger, Computers, Communication, and the Public Interest, Baltimore, MD: The Johns Hopkins Press.
[27] silver, N. (2012). The Signal and the Noise: Why So Many Predictions Fail — but Some Don't. New York: Penguin Press, Inc.

Anthony G. Picciano is a professor and executive officer in the Ph.D. Program in Urban Education at the Graduate Center of the City University of New York. He is also a member of the faculty in the graduate program in Education Leadership at Hunter College, and the doctoral certificate program in Interactive Pedagogy and Technology at the City University of New York Graduate Center. He has extensive experience in education administration and teaching, and has been involved in a number of major grants from the U.S. Department of Education, the National Science Foundation, IBM, and the Alfred P. Sloan Foundation.

Review of Current Student-Monitoring Techniques used in eLearning-Focused recommender Systems and Learning analytics. The Experience API & LIME model Case Study

Alberto Corbi, Daniel Burgos

UNIR Research,Universidad Internacional de La Rioja - UNIR

***Abstract* — Recommender systems require input information in order to properly operate and deliver content or behaviour suggestions to end users. eLearning scenarios are no exception. Users are current students and recommendations can be built upon paths (both formal and informal), relationships, behaviours, friends, followers, actions, grades, tutor interaction, etc. A recommender system must somehow retrieve, categorize and work with all these details. There are several ways to do so: from raw and inelegant database access to more curated web APIs or even via HTML scrapping. New server-centric user-action logging and monitoring standard technologies have been presented in past years by several groups, organizations and standard bodies. The Experience API (xAPI), detailed in this article, is one of these. In the first part of this paper we analyse current learner-monitoring techniques as an initialization phase for eLearning recommender systems. We next review standardization efforts in this area; finally, we focus on xAPI and the potential interaction with the LIME model, which will be also summarized below.**

***Keywords* — LIME model, eLearning, Conceptual Educational Model, Rule-based recommender system, Informal learning, Social interaction, Learning Tool Interoperability, User monitoring**

I. Introduction: Review of Recommender Engines, eLearning and Need for User Input Data

RECOMMENDER engines deliver suggestions based on collected information on preferences, general user behaviour and even items bought or content searched. Trendy online stores and services massively apply this approach ([HYPERLINK \l "1167344" 1]). The information can be obtained explicitly (by processing users' manual tiering) or implicitly, typically by monitoring users' behaviour, such as songs downloaded, applications launched, chat transcriptions, web sites visited, PDFs read, or ebooks transmitted to ePub readers (2]).

Recommenders can also make use of demographic info and social information (e.g., followers, e-friends, posts, replies, chat rooms, and others), as well as geographical location data or even health signals (e.g., pedometers, blood pressure).

Collaborative filters ([HYPERLINK \l "marlinmodeling" 3]) are very often used by recommender systems along with content-, knowledge- and social-based filters. Implementation of these filters has grown as access to the Internet has become more widespread in recent years. They can be used for any type of reachable media (e.g., movies, music, television, and books) and in many different scenarios, such as eLearning, e-commerce, mobile applications, search, dating, etc. These filters need to access as much of a user's navigation and behaviour history as possible in order to offer fine-tuned purchase options or action tips.

Memory-based methods use similarities and ratings from all users who have manually expressed their preferences/level of satisfaction on a given object/issue. These similarities represent the distance between two users and their tiered records. Model-based methods establish first the sets of similar users by using Bayesian classifiers, neural networks and fuzzy systems. Generally, commercial recommender engines use memory-based methods. On the other hand, model-based methods are usually associated with research environments, including eLearning. Hybrid techniques can also be applied and have been demonstrated to be of much importance to assist and guide users through systems. Hybrid recommenders merge different types of techniques in order to get the most out of each of them. Finally, we have rule-based recommenders, like LIME, which will be analysed below. In rule-based systems, a set of conditional filters are manually defined and triggered when necessary in order to deliver the appropriate recommendation to the user/learner.

The increase in the attention paid by the research community to recommender systems is striking, as has already been pointed out in 4]. Figure 1 shows, on the Y-axis, the number of cited papers from each year as of 2013. The size of each bubble corresponds to the number of proceeding articles for that given year. It can be noted that there is a peak of interest around 2009.

DOI: 10.9781/ijimai.2014.276

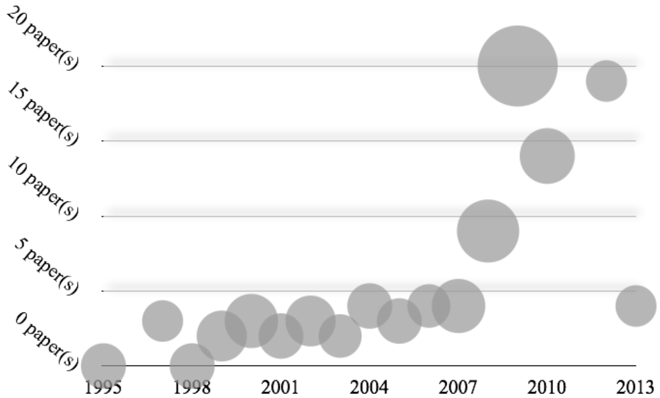


Fig. 1. # of papers and workshops related to subject *recommendation*

With the development of sophisticated eLearning environments and Learning Management Systems (LMS) ([HYPERLINK \l "abedour" 5]), *personalization* is also becoming an important feature. Personalized learning occurs when eLearning platforms are designed according to educational experiences that fit the needs, goals, and interests of each individual learner. Personalization can be achieved using different recommendation techniques, very similar to those just summarized. Ideally, recommender systems in eLearning environments should assist students in finding relevant learning actions and materials that perfectly match their profile and the best way towards self-education. The right time, the right context, and the right way are also critical. Recommenders should also keep learners motivated and enable them to complete their academic activities in an effective and efficient way. Personalization should take place, not only on enrolment-limited online campuses or *Small Private Online Courses* (site courses, college classes, student groups, etc.), but also on the now trendy MOOCs: *Massive Open Online Courses* environments (6], [HYPERLINK \l "mooeurope" 7]), where enrolment rate can be up to a few thousand students. In other words, a recommender system should have the ability to efficiently scale up or down independently of the number of students and without losing sight of the goal of improving individualized education.

Recommender systems (especially in eLearning) can also suffer from the *cold-start* problem. Cold start occurs when there is an initial lack of input data (ratings, logged actions from users, etc.) to trigger or initialize the appropriate algorithm. We can distinguish two main kinds of cold-start variants: *new item* and *new user* ([4]). The new-item problem arises because new items entered do not have initial ratings/inputs from users. Also, a priori, *new users* in a system might not yet have provided any input info, and therefore cannot receive any personalized recommendations.

Independently of the algorithm used, the identifiable potential issues (like cold start) and the scenario of application, recommender systems require input data in order to behave properly (8]). This data can be manually entered ([HYPERLINK \l "Bobadilla20111310" 9]) by the user (ratings, explicit opinions, etc.) or implicitly obtained by monitoring software. In an eLearning environment, the latter approach is more likely to be the chosen one.

We now list the most common techniques used for monitoring learners' actions in an LMS. The next sections will present the Experience API and other standardization efforts as new and modern ways of logging learner actions, chosen materials, student paths, etc., and serving them to recommender systems. Finally, we introduce the rule-based LIME model and discuss how can it be fed from an Experience API Learning Record Store repository (which we will also discuss) in order to properly operate and deliver rule-based recommendations to students.

II. Basic System-Dependent monitoring techniques

There exist three main different non-standard ways of interacting with Learning Management Systems (and electronic systems in general) and extracting user/learner data (also summarized in Figure 2):

A. Web Services

The first and most immediate way to obtain learner input data is through LMS-dependent web services and API calls. Modern LMS (10]) do usually offer simple, elegant, industry-standard and compelling ways (WSDL, SOAP, RPC and REST) of accessing their internal information and retrieving needed data. This approach has one main drawback: not every service needed is implemented and/or enabled by default. This could be easily tackled if we are granted access to the LMS infrastructure in order to add these missing *sockets* or activate existing *disabled-by-default* ones. However, this is not always possible in many scenarios (e.g., proprietary cloud-based campus environments). Another clear disadvantage is that developed web services are very unlikely to be compatible between two distinct LMS, making it necessary to re-code each of them for every platform and software version.

B. Scrapping

Web scrapping consists of, on the one hand, running automated HTTP(S) requests that retrieve the same pages and HTML documents as a user would fetch by operating a web browser manually ([HYPERLINK \l "6112910" 11]). On the other hand, after such requests have succeeded, data can be distilled, examined and applied to some sort of scripting/analytics. Most HTTP command line (CLI) client programs/libraries allow authentication and form submission, which is usually enough for most purposes. Although web scrapping seems the most compatible form of mechanized data-mining, we still face a minor problem: some LMS make huge use of Javascript for accessing resources and building routes to them. In this scenario, CLI web clients are not enough and should be superseded by what are known as headless web browsers, explained in previous studies (12], [HYPERLINK \l "Grigalis:jucs_20_2:unsupervised_structu" 13]). Such browsers are scriptable, run without any user interface, and best of all understand and can execute Javascript code without user intervention.

The result of a scrapping operation is usually an HTML file or a set of files of this kind, which should be processed afterwards (14]) in order to extract the desired monitoring information. As HTML is a descendant of XML, any XML parsing technique (XPath, XQuery, XSLT, etc.) and technology applies here, e.g., Nokogiri ([HYPERLINK \l "Hun13" 15]).

C. Raw database access

This is by far the most-often-seen method in the literature, which implies direct access to the system database. This approach has several advantages and disadvantages. The main advantage is speed, since no intermediaries, software layers or no other different APIs play a role in data retrieval (apart from the SQL engine and the APIs themselves). The most significant downside is possible database scheme migrations and incompatibilities as new versions of the server software are deployed.

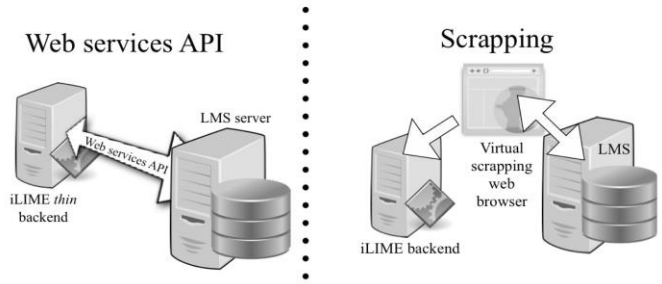


Fig. 2. Basic web monitoring techniques

These three techniques spring into action at some point or another of the student data-mining process, in different ways and with different goals. Some monitoring models, as seen in 16] and [HYPERLINK \l "Mazza04gismo:a" 17], make use of reports and logs derived from data contained in server temporary files. Other research efforts, such as the one presented in 18], have used closed-systems and setups, with their own specific monitoring methods and engines. The study presented in [HYPERLINK \l "5561329" 19] makes use of quiz results as input for a research recommender model. The authors of 20] and [HYPERLINK \l "6033004" 21] feed their recommenders with web-browsing behaviour. In22], the authors gain direct access to a Moodle instance database in order to boot their *Predictive a priori algorithm*. The model in [HYPERLINK \l "ElB10" 23] initially presents students with a test to identify his/her personality in Myers-Briggs dimensions. The authors in 24] suggest obtaining input data not only from the server and client sides, but also from proxy servers. In [HYPERLINK \l "Kardan:2012aa" 25], content recommendation needs each student to self-monitor him/herself: learners estimate different indexes themselves and compare them with actual values, which are retrieved by the system. The model presented in 26] uses the *AprioriAll* algorithm to immediately build sequences from server logs, which are used in conjunction with tags in order to deliver recommendations. The model in [HYPERLINK \l "conf/wec/WangH05" 27] also makes use of the AprioriAll algorithm using only web logs. In 28], again, only web-browsing activities of learners are monitored, but these are then subdivided into *web content mining*, *web structure mining* and *web usage mining* realms.

We also find learning research software prototypes, like the PSLC Datashop initiative from the Pittsburgh Science of Learning Center [HYPERLINK \l "Stamper:2011:MED:2026506.2026609" 29], which has defined its own XML DTD schema as a logging scaffold for their *Tutor* learning research platform. Some approaches rather build a dedicated tool or patch applied to a LMS, as in 30] with the MOCLog project for Moodle.

The Experience API and other standardization proposals for the monitoring phase, presented below, advocate a completely new and cohesive approach to this critical phase in the recommendation/learning analytics workflow.

II. Standard Specifications for Monitoring

The aforementioned non-standardized approaches to user/learner monitoring can be applied on fully controlled scenarios and research projects. However, they turn out to be unsatisfactory in real academic environments managed by third-party institutions.

There exist a few proposals that aim at standardizing the monitoring and logging of user actions. Almost all are based on the *Resource Description Framework*, or RDF [HYPERLINK \l "Pan09" 31]. The idea behind RDF is something called the *triple*. A triple can really be condensed to a plain sentence structure:

- subject
- phrase that characterizes a relationship
- object.

Example: Daniel – *is the author of* – this paper.

Triples are extremely useful and simple, and provide a grammar for the so-called *semantic web*.

Also, some of these specifications include some sort of software and database back-end service, linked APIs and query language that allow learning platforms to send and store monitoring data and third-party learning analytics software to query and retrieve analysable data. We summarize here the most important and paradigmatic monitoring specs:

The **Caliper** framework/**Sensor** API was proposed by the IMS Global Consortium and follows the triple metaphor. It is built around the following concepts (32]): *Learning Metric Profiles* that provide an activity-centric focus to standardize actions and related context; *Learning Sensor API* and *Learning Events*, which drive tools and an associated analytics service solution; and finally, Learning Tool Interoperability (LTI), which enhances and integrates standardized learning measurements with tool interoperability.

IEEE 1484.11.1/IEEE 1484.11.2 ([HYPERLINK \l "IEE05" 33]) provides a complex data model structure for tracking information on student interactions with learning content. Additionally, an API allows digital educational content coming from the LMS and third-party services to query and share collected information.

JSON Activity Streams (34]) is the name of the specification published by IBM, Google, MySpace, Facebook, VMware and Microsoft. Its goal is to provide sufficient metadata about an activity such that a consumer of the data can present them to a user in a rich human-friendly format. It does not provide a logging service, just the specification of the message format.

Finally, we also have the **Experience API**, which will be addressed in the next section.

Security and privacy models can also be applied in all specs cited above. Network communications can be encrypted and the subject can be anything but the learner's real name. Learning analytics researchers and logging storage implementers are responsible for the ethical usage of the compiled info coming from student monitoring. As with any other area related to digital mining, trust, accountability and transparency must always prevail ([HYPERLINK \l "Par14" 35]).

III. The Experience API Specification

The Experience API (or *xAPI* for short) is an eLearning monitoring specification developed by Rustici Software and the Advanced Distributed Learning Initiative (ADL), and is aimed at defining a data model for logging data about students' learning paths (36]). It also furnishes an API for sharing these data between remote systems, as we will see later. The Experience API allows, among other things, the tracking of games and simulations, real-world behaviour, learning paths and academic achievements. xAPI defines independent mechanisms, protocols, specifications, agreements and software tools for monitoring any imaginable scenario (Figure 3): from online campuses and student behaviour to workforce control ([HYPERLINK \l "6530268" 37]).

Fig. 3. Examples of usage of the Experience API

xAPI also uses JSON to transfer states/sentences to a central web service. This web service allows clients to read and write data in the form of *sentence objects* that share the foundations of the aforementioned triple scheme. In their simplest conception, sentences are in the form of *actor*, *verb* and *object/activity*, like the examples in Figure 4. A JSON xAPI message could resemble the following:

```
{"id": "3f2ef28f-ef1a-4a1f-9f5e",
 "actor": {
  "name": "Peter",
  "mbox": "mailto:some@new.user",
  "objectType": "Agent"
 },
 "verb": {
  "id": "http://.../verbs/solved",
  "display": {
   "und": "solved"
  }
 },
 "context": {
  "contextActivities": {
   "parent": [
    {
     "id": "http://../objects/problems",
     "objectType": "Activity"
    }
   ]
  }
 }}
```

More complex statement forms can be used and we will elaborate more on them in the next section. The set of verbs and objects an institution can work with is called *vocabulary*. Each institution can define its own vocabulary with no restriction as long as an URL links back each verb and object to a JSON stream describing it.

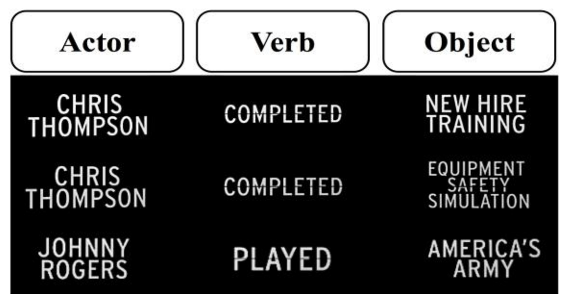


Fig. 4. Some examples of xAPI sentences

The Experience API was released, as version 1.0, in April 2013, and there are, as of today, over 100 adopters, projects and companies involved, such as those in Figure 5.

Fig. 5. Some adopters of the Experience API specification

The specification also contemplates a query API to help find logged statements, and performs some analytics (averages, aggregation, etc.) on the data. Finally, the Experience API is an open-source and free initiative, whose source code and specifications are open to anyone.

IV. Experience API LRS as an eLearning Monitoring Engine

The core of the Experience API is the Learning Record Store (LRS). The LRS is a specific module for data storage that allows an LMS (or any other social platform) to report tracking information on the learning experience. At any time, an LMS can send collected data over the network to an Experience API web service. An LRS is nothing more and nothing less than a wrapper or API software layer to a SQL database (initially, a PostgresSQL instance in the original Rustici implementation), as can be appreciated from Figure 6. This free LRS implementation was open-sourced by ADL (available at its Github repository) and is based on the Python computer language and on the publicly acclaimed Django web framework.

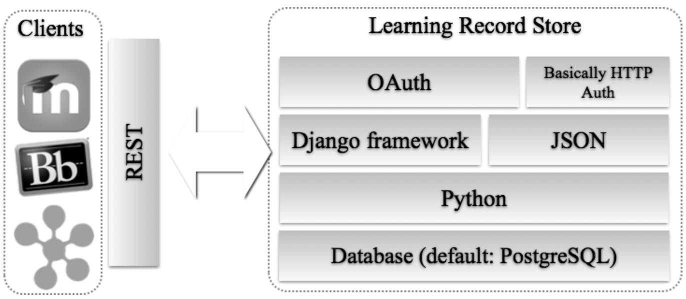


Fig. 6. Usual LRS software stack and interaction

The *learner (actor)*, *verb* and *object/activity* elements explained above are mandatory when talking to the LRS. However, they can be complemented with *result* and a *context* extra fields with additional information.

Students who interact with educational content via different systems or tools will leave traces in the LRS; each of these tools, if appropriately designed, will provide a totally different actor/user ID to preserve anonymity.

The *verb* element is a key part of an LRS communication, because it describes the action performed by the student. A URL must also be attached to the verb JSON property, pointing to its definition. This definition is composed of a name, a description, and a brief text suggesting plausible uses. In an eLearning environment, a verb is usually employed in its past tense form and could be something like: “read”, “tried”, “failed”, “passed”, “experienced”, etc.

The *object/activity* part of the statement refers to “what” was experienced in the action defined in the verb, and usually corresponds to the learning activity (webinar, wiki, chat room, forum, mail message, etc.). Objects/activities must also embody a URL pointing to their rationale, which can include other information such as a description of the learning activity, verbs that can apply, possible results and usage suggestions.

The *result* component provides the denouement to the statement. It includes score, level of success and completion fields.

The context part adds more details to the overall statement, like the relationship of the activity with other activities, its order in the learning stream, or the teacher’s name.

To every element in a sentence (actor, verb, context, etc.) sent to the LRS can be added, if needed, any type of pair key/value with extra information. It is even possible to add localization information so that an element can be perfectly identified in all possible languages.

As introduced in Figure 6, an LRS must also implement REST calls for data transfer (PUT, POST, GET and DELETE). The Experience API can make use of either OAuth or HTTP Basic Authentication when communicating with the outside world, ensuring a certified and secured dialogue between clients (usually an LMS) and the LRS service.

One of the key aspects of the LRS architecture is that it can be implemented in shared cloud ecosystems, allowing communications from very different eLearning platforms and academic institutions. In other words, monitoring data can be uniformly stored, allowing rapid, vast and democratic access to learning analytics information. Also, as LRS servers can integrate data from many different sources and from the same user/learner in a harmonized way, recommender systems can reduce the effects of possible cold-start scenarios.

Some companies are beginning to offer corporate cloud LRS services at different price tiers: Rustici Software, Saltbox, Learning Locker, Biscue, Clear, Grassblade, among others. Some also include compelling online analytics tools.

There exist some free LRS *hosting* services but mainly for testing and technology promotion purposes, and not applicable for research or production environments. It is worth mentioning the service run by ADL (lrs.adlnet.gov/xAPI) and the one deployed by Rustici Software (demo.tincanapi.com).

V. The LIME Model and the LRS

Now that we have reviewed the most prominent monitoring techniques and introduced a few recent efforts towards regulation, we should ask how a real recommender engine could work with and benefit from a specific RDF-based source. The Experience API and the LIME model, explained below, are chosen.

The LIME model, presented in 38], is a tutor-lecturer-crafted rule-based recommender grounded on four separate pedagogical components strongly evident in all stages of education (Figure 7):

- Learning, or what every learner needs to do in order to assimilate and build knowledge on his or her own.
- Interaction, or relationships established, activities and academic interaction between students, leading to the acquisition of knowledge and competencies.
- Mentoring, or what teachers/tutors give relevance to.
- Evaluation, or officially graded activities, in every single category above listed.

Lecturers-tutors must design a strategy for each of his/her courses. The model codifies this strategy for a course or class group by using settings and categories.

A *course setting* is the balance between *formal* and *informal* scenarios. In this context, *formal* means a regular academic programme with regular evaluation means (e.g. graded exams); *informal* means continuous evaluation and user activity inside the Learning Management System and every tool linked to it (e.g. Social Networks or repository). The system collects specific inputs from both settings, keeping an overall balance of 100%. For instance, if the designer requires just a *formal* setting, the balance should be *Informal*: 100% - *Formal*: 0%.

Furthermore, a learning scenario must be defined as the balance between the Learning, Interaction, Mentoring, and Evaluation, in combination with the Formal and Informal settings categories. In the LIME model, every category and setting are assigned with a specific weight (w_i), keeping an overall balance of 100%.

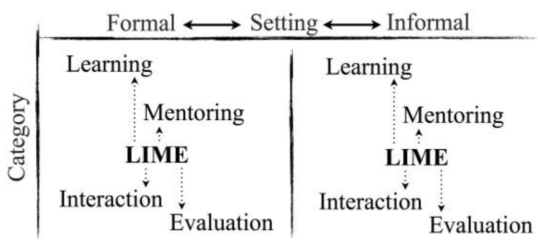
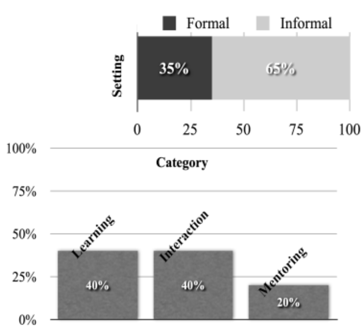


Fig. 7. Categories and settings in the LIME model

In the LIME model each input (action performed by a student in the eLearning platform or Social Network) is attributed a category and a weight, assigned by the teacher/tutor.

An example of model configuration for a specific site can be found in Figure 8. Based on these components, tutors can manually define and parameterize recommendation rules, which will only trigger a message to the student if conditions regarding categories, inputs and settings are met.



Inputs	Category (L-I-M)	Setting (In)Formal	EVAL
New content	L	I	
Forum post	I	F	
Problem set	L	F	✓
Direct message	I	I	
Progress test	M	F	
Chat message	I	I	
Profile update	I	F	
Public answer in forum	I	F	
Grade in test	M	F	✓

Fig. 8. Sample configuration of the LIME model for a specific course site

LIME is therefore a tutor-lecturer-crafted, rule-based recommender system for cloud-institutional learning environments (SPOCs or MOOCs), which contrasts with other recommendation paradigms reviewed in previous sections. LIME's goal is simply to improve learning efficiency, and to facilitate the learning itinerary of every student by a personalised recommendation set.

LIME can be fed from learner inputs in a variety of ways. However, our model can also be initialized with tracked data stored in a xAPI LRS instance/server if we make some assumptions.

How can LIME inputs be built out of information stored in the LRS? A LIME model input has to define an action and a context in which a learner performs this action:

- participation in chat
- answer in main forum thread
- message to tutor
- resolution of problem set
- formal broadcast mail to mates
- ratio of emoticons used in communications
- ...

xAPI verbs and objects, taken in an isolated way, are not sufficient. However, a joint entity composed of a verb plus an xAPI object makes more sense in our model, as shown in Figure 9:

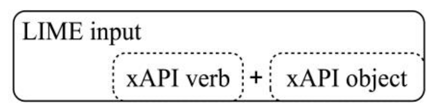


Fig. 9. LIME Inputs from xAPI sentences

As stated above, verbs and objects in the xAPI specification must be backed by JSON composites with information about meaning and usage tips. It is up to the implementer to define which verbs and objects best represent the scenario to be tracked and monitored. Let us take a look at the sample verbs and activities available on the official Experience API site (adlnet.gov/expapi). In Figure 10 are listed all the verbs and activities the LRS can store and their possible combinations to build a meaningful and compatible LIME input.

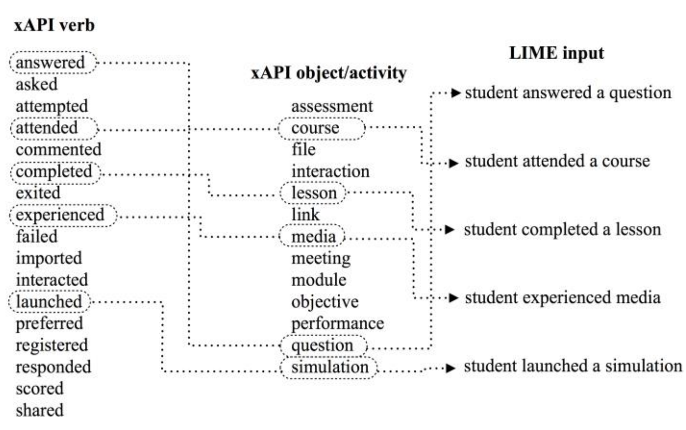


Fig. 10 From xAPI verbs and objects to LIME inputs

As explained in previous paragraphs and as part of the model configuration, each input should be assigned a weight (w_i), a category and a setting. These parameters should not reside on the LRS but on the LIME system's own configuration repository. In other words, LIME administrators should maintain an updated *equivalency list* between LRS vocabulary and LIME inputs. These inputs will then interplay with rules (Figure 11), which are, in turn, based on *predicate* *filtering*. Predicates are applied over collections of inputs and highly resemble W3C XQuery or ECMA LINQ, detailed in [HYPERLINK \l "Saigaonkar:2010:XFS:1858378.1858429" 39],40] and [HYPERLINK \l "Pardede:jucs_15_10:sqlxml_hierarchical_" 41].

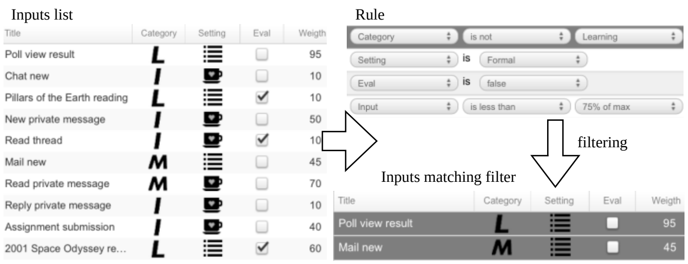


Fig. 11: Predicate filtering in LIME

As LIME was developed as a Basic Learning Tool Interoperability (Basic LTI) application, this equivalency list can even be stored in the LMS database through the LTI Settings API specification, part of LTI 1.0 and above. The model thus remains free from external configuration files or own database management. In order to save this list, it is only necessary to send a POST HTTP request like the one in the following example:

```
  POST http://server/imsblis/service/
  id=832823923899238
  lti_message_type=basic-lti-savesetting
  lti_version=LTI-1p0
  setting="participated+chat=message  in  chat
room; experienced+lesson=read text"
  oauth_callback=about:blank
  oauth_consumer_key=1213415
  oauth_nonce=14c6211cc66d87644f0855511
  oauth_signature=IkllkkZ1qfShYBYE+BhC
  oauth_signature_method=HMAC-SHA1
  oauth_timestamp=1338872426
  oauth_version=1.0
```

It is important to notice that LMS must be LTI compatible and support the Settings API protocol.

VI. LRS DATA AGGREGATION AND LIME RULES

Once LRS sentences are stored and an agreement between LIME inputs and these has been established, we have all the necessary ingredients to trigger recommender rules and deliver recommendations to students, if applicable. However, rules in LIME cannot operate upon atomic and individual LRS records, but only upon averages and aggregated substantial data, which offer a more equalized view of the learner situation. An example of this aggregation procedure is presented in Figure 12:

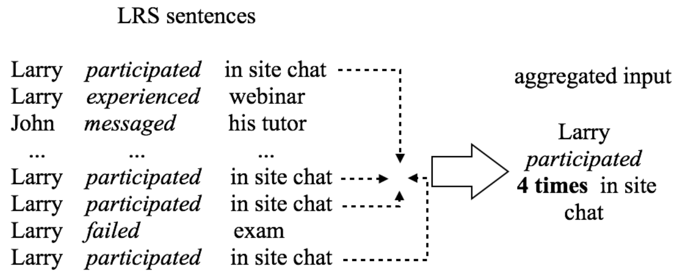


Figure 12: Aggregation of LRS sentences

Mathematically:

$$(\text{LIME input})_i = \frac{\sum_{j=0}^{n} (\text{LRS statement})_j}{w_i}$$

These aggregation operations are covered by the xAPI standard as well. The Experience API provides a query language to easily data-mine an LRS. For instance, the following code collects all the times the user "John" has tried an exam, and returns an aggregated result:

```
stmts.where(
    'actor.name = "John" and ('+
        'verb.id =
"http://adlnet.gov/expapi/verbs/passed"'+
        ' or '+
        'verb.id =
"http://adlnet.gov/expapi/verbs/failed"'+
    ')')
```

The default (and so far only) implementation of this query language is the ADL.Collection API, written in Javascript and ready to be used in browsers or on the server-side with NodeJS. There are two versions of this API: CollectionSync and CollectionAsync. They are almost the same, but the Async version runs the queries in a separate worker thread. The downside of this is that the statements must be serialized and passed into the worker, which can be slow. On the other hand, the user interface is more responsive.

VII. CONCLUSION

This paper describes incipient technologies and steps taken towards the dissemination of standardized monitoring engines. The engine mainly underlined in this paper is the Experience API, or xAPI for short. xAPI has been designed to store user data in a simple, centric, standard, client agnostic and powerful way. We also discuss the suitability of recommender systems in general and of the LIME recommender model in particular. LIME is a rule-based recommendation model. Rules in LIME require inputs (e.g. learner data and actions taken) that can be obtained in a variety of ways, like user tracking and interaction, user performance, or user profile.

We also perform a survey of the most common monitoring techniques and how they have been implemented in previous research projects related to recommender systems and learning analytics in general. With this review we illustrate there is no agreed way on how to register learner events. All mentioned techniques incorporate a certain percentage of dependency on the system software being monitored.

Finally, we present the required adaptations and modifications that xAPI sentences need in order to build LIME-compatible inputs and how those can be aggregated and mined in order to feed system rules. On rule execution, our model delivers suggestions to students and learners. The xAPI spec atomizes learner actions in verbs and objects, which must be syntactically combined in order to obtain the aforementioned inputs. These combinations must be designed and listed by the tutor/teacher and handed over to our model. We suggest this equivalency list resides in the LMS’s own database space, thanks to the LTI Settings API. The Experience API also offers native aggregation-statistical tools, which turn out to be of great help in this process.

Acknowledgment

This research is partially funded by UNIR Research (http://research.unir.net), Universidad Internacional de La Rioja (UNIR, http://www.unir.net), under the Research Support Strategy (2013-2015), Research Group TELSOCK.

VIII. References

[1] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," Internet Computing, IEEE, vol. 7, no. 1, pp. 76-80, Jan 2003.

[2] F. Ricci, L. Rokach, B. Shapira, and P.B. Kantor, Recommender Systems Handbook.: Springer, 2010.

[3] Benjamin Marlin, "Modeling User Rating Profiles For Collaborative Filtering," in NIPS'03, 2003.

[4] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender Systems Survey," Know.-Based Syst., vol. 46, pp. 109-132, Journal of Universal Computer Science 2013.

[5] M. Aberdour, "Open Source Learning Management Systems: Emerging open source LMS markets," 2007.

[6] Fran, Martin Ebner, Alexander Pohl, and Behnam Taraghi, "Interaction in Massive Courses," Journal of Universal Computer Science, vol. 20, no. 1, pp. 1-5, jan 2014.

[7] Y. Epelboin, "MOOC in Europe," UPMC-Sorbonne Université, 2013.

[8] Daniel Burgos, Colin Tattersall, and Rob Koper, "Representing Adaptive and Adaptable Units of Learning," in Computers and Education.: Springer Netherlands, 2007, pp. 41-56

[9] Jesus Bobadilla, Fernando Ortega, Antonio Hernando, and Javier Alcal, "Improving collaborative filtering recommender system results and performance using genetic algorithms," Knowledge-Based Systems, vol. 24, no. 8, pp. 1310-1316, 2011.

[10] M.AC. González, F.J.G. Penalvo, M.J.C. Guerrero, and M.A Forment, "Adapting LMS Architecture to the SOA: An Architectural Approach," in Internet and Web Applications and Services, 2009. ICIW '09. Fourth International Conference on, May 2009, pp. 322-327.

[11] S.K. Malik and S. A M Rizvi, "Information Extraction Using Web Usage Mining, Web Scrapping and Semantic Annotation," in Computational Intelligence and Communication Networks (CICN), 2011 International Conference on, Oct 2011, pp. 465-469.

[12] A Holmes and M. Kellogg, "Automating functional tests using Selenium," in Agile Conference, 2006, July 2006, pp. 6 pp.-275.

[13] Tomas Grigalis and Antanas , "Unsupervised Structured Data Extraction from Template-generated Web Pages," Journal of Universal Computer Science, vol. 20, no. 2, pp. 169-192, feb 2014

[14] H. Bosch et al., "Innovative filtering techniques and customized analytics tools," in Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on, 2009

[15] P. Hunter, Instant Nokogiri.: Packt Publishing Ltd., 2013.

[16] Angel A. Juan, Thanasis Daradoumis, Javier Faulin, and Fatos Xhafa, "SAMOS a Model for Monitoring Studentsand Groups; Activities in Collaborative eLearning," Int. J. Learn. Technol., vol. 4, no. 1/2, pp. 53-72, 2009.

[17] Riccardo Mazza and Christian Milani, "GISMO: a Graphical Interactive Student Monitoring Tool for Course Management Systems," in T.E.L.'04 Technology Enhanced Learning '04 International Conference. Milan, 2004, pp. 18-19.

[18] Jungsoon Yoo, Sung Yoo, Chris Lance, and Judy Hankins, "Student Progress Monitoring Tool Using Treeview," SIGCSE Bull., vol. 38, no. 1, pp. 373-377, 2006.

[19] S. Shishehchi, S.Y. Banihashem, and N.AM. Zin, "A proposed semantic recommendation system for e-learning: A rule and ontology based e-learning recommendation system," in Information Technology (ITSim), 2010 International Symposium in, vol. 1, June 2010, pp. 1-5.

[20] K. Takano and Kin Fun Li, "An Adaptive e-Learning Recommender Based on User's Web-Browsing Behavior," in P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2010 International Conference on, Nov 2010, pp. 123-131.

[21] K. Takano and Kin Fun Li, "An adaptive learning book system based on user's study interest," in Communications, Computers and Signal Processing (PacRim), 2011 IEEE Pacific Rim Conference on, Aug 2011, pp. 842-847.

[22] Enrique García, Crist, Sebasti, and Carlosde Castro, "An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering," User Modeling and User-Adapted Interaction, vol. 19, no. 1-2, pp. 99-132, 2009.

[23] El Hassan, A. and El Adani, M. El Bachari E., "Design of an Adaptive E- Learning Model Based on Learner’s Personality," Ubiquitous Computing and Communication Journal, vol. 5, 2010.

[24] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert Systems with Applications, vol. 33, no. 1, pp. 135-146, 2007.

[25] Ahmad A. Kardan, Nahid Ghassabzadeh Saryazdi, and Hamed Mirashk, "Learner Clustering and Association Rule Mining for Content Recommendation in Self-Regulated Learning," International Journal of Computer Science Research and Application, 2012.

[26] Boban Vesin, Mirjana Ivanovi, Aleksandra Kla, and Zoran Budimac, "Protus 2.0: Ontology-based semantic recommendation in programming tutoring system," Expert Systems with Applications, vol. 39, no. 15, pp. 12229-12246, 2012.

[27] Tong Wang and Pi lian He, "Web Log Mining by an Improved AprioriAll Algorithm.," in WEC (2), 2005, pp. 97-100.

[28] M. K. Khribi M. Jemni, "Toward a Hybrid Recommender System for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval," in World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 2017.

[29] John C. Stamper et al., "Managing the Educational Dataset Lifecycle with DataShop," in Proceedings of the 15th International Conference on Artificial Intelligence in Education, Berlin, Heidelberg, 2011, pp. 557-559.

[30] Riccardo Mazza, Marco Bettoni, Marco Far, and Luca Mazzola, "MOCLog--Monitoring Online Courses with log data," Proceedings of the 1st Moodle Research Conference, pp. 14-15, 2012.

[31] JeffZ. Pan, "Resource Description Framework," International Handbooks on Information Systems, 2009.

[32] IMS Global Learning Consortium Inc., "Learning Measurement for Analytics Whitepaper," 2013.

[33] IEEE, "Data Model for Content to Learning Management System Communication," IEEE Std 1484.11.1-2004, 2005.

[34] J and Atkins, M and Norris, W and Messina, C and Wilkinson, M and Dolin, R Snell, "JSON Activity Streams 1.0," 2011.

[35] Abelardo Pardo and George Siemens, "Ethical and privacy principles for learning analytics," British Journal of Educational Technology, vol. 45, no. 3, 2014.

[36] David Kelly and Kevin Thorn, "Should Instructional Designers Care About the Tin Can API?," eLearn, vol. 2013, no. 3, 2013.

[37] A del Blanco, A Serrano, M. Freire, I Martinez-Ortiz, and B. Fernandez-Manjon, "E-Learning standards and learning analytics. Can data collection be improved by using standard data models?," in Global Engineering Education Conference (EDUCON), 2013 IEEE, March 2013, pp. 1255-1261.
[38] Daniel Burgos, "L.I.M.E. A recommendation model for informal and formal learning, engaged," IJIMAI, pp. 79-86, 2013.
[39] Swati Saigaonkar and Madhuri Rao, "XML Filtering System Based on Ontology," in Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, New York, NY, USA, 2010, pp. 51:1--51:6.
[40] James Cheney, Sam Lindley, and Philip Wadler, "A Practical Theory of Language-integrated Query," SIGPLAN Not., vol. 48, no. 9, pp. 403-416, 2013.
[41] Eric Pardede, J. Wenny Rahayu, Ramanpreet Kaur Aujla, and David Taniar, "SQL/XML Hierarchical Query Performance Analysis in an XML-Enabled Database System," Journal of Universal Computer Science, vol. 15, no. 10, pp. 2058-2077, may 2009.

M.Sc. Alberto Corbi works as a senior researcher at the Technology-enhanced Learning & Social Networks (TELSOCK) research group and as a teaching assistant at the School of Engineering, both part of the International University of La Rioja. With a background in physics (M.Sc. in ocean-atmosphere interaction), computing and education, he is currently involved in research fields around recommender systems, eLearning standards and systems interoperability. Simultaneously, he is working on his PhD thesis around medical imaging at the Spanish Council for Scientific Research (CSIC).

Prof. Dr. Daniel Burgos works as Vice-Chancellor for Research & Technology and UNESCO Chair on eLearning at the International University of La Rioja (www.unir.net, http://research.unir.net). Previously he worked as Director of Education Sector and Head of User Experience Lab in the Research & Innovation Department of Atos, Spain, and as an Assistant Professor at Open Universiteit Nederland before that.
His research interests are mainly focused on Adaptive and Informal eLearning, Learning & Social Networks, eGames, and eLearning Specifications. He is or has been involved in a number of European-funded R&D projects, such as Intuitel, Hotel, Edumotion, Inspiring Science Education Stellar, Gala, IntelLEO, Go-MyLife, Grapple, Unfold, ProLearn, TenCompetence, EU4ALL, NiHao, Kaleidoscope, Sister, and ComeIn. Prof. Dr. Burgos holds degrees in Communication (PhD), Computer Science (Dr. Ing), Education (PhD), and Business Administration. Furthermore, he is a member of a number of Executive Boards of associations and professional clusters focused on educational technology and eLearning innovation, for example Telearc, Telspain, Menon Network, Adie, and others.

Integration of Multiple Data Sources for predicting the Engagement of Students in Practical Activities

Llanos Tobarra[1], Salvador Ros[1], Roberto Hernández[1], Antonio Robles-Gómez[1], Agustín C. Caminero[1], and Rafael Pastor[1]

[1]*Communication and Control System Department, from Spanish University for Distance Education (Universidad Nacional de Educación a Distancia, UNED)*

***Abstract* — This work presents the integration of an automatic assessment system for virtual/remote laboratories and the institutional Learning Management System (LMS), in order to analyze the students' progress and their collaborative learning in virtual/remote laboratories. As a result of this integration, it is feasible to extract useful information for the characterization of the students' learning process and detecting the students' engagement with the practical activities of our subjects. From this integration, a dashboard has been created to graphically present to lecturers the analyzed results. Thanks to this, faculty can use the analyzed information in order to guide the learning/teaching process of each student. As an example, a subject focused on the configuration of network services has been chosen to implement our proposal.**

***Keywords* — Learning Analytics (LA), Assessment and Evaluation Strategies, Virtual/Remote Laboratories, Collaborative Tools, Distance Education.**

I. Introduction

THE evaluation procedure is a key element within the process of learning. Basically, it allows faculty to check whether educative objectives are accomplished, not only by students, but also by all the participants involved in an educative program [25], such as pedagogical resources. As a consequence, lecturers are required to adapt the learning process to students' needs or preferences, reinforcing or extending it if necessary, according to the European Higher Education Area (EHEA) [27]. The importance of evaluation procedures is even greater at distance Universities since their students' learning process is different from that of face-to-face universities. In distant Universities, students must be more independent and self-demanding since there are no tight schedules, and this heavily affects the evaluation process. By means of evaluation, faculty can select the suitable learning results and adapt dynamically the subject contents to students [22].

On the other hand, adaptive hypermedia has been widely used for the development of customized Web-based courses in the field of Education [3]. Therefore, the students' learning process was guided, adapting both pedagogical resources and learning ways to specific user's features. Since lecturers adapt course materials to students' skills and usage data dynamically [15], they were able to acquire more knowledge in less time. ELM-ART [31] and TANGOW [4] are some examples of traditional educational adaptive systems. The students' interaction in these types of architectures is different from face-to-face students, as stated in [28]. In particular, students have to be able to adapt their communication way to the user interfaces of systems adapted to the students' needs [14].

It is also important to include collaborative issues taking into account the students' behavior. The most relevant research works related to adaptation in Computer Support for Collaborative Learning (CSCL) systems are COALE [9], WebDL [12], and COL-TANGOW [4]. COALE is a collaborative environment where different exercises are recommended to students. The main goal in WebDL is to facilitate user access to services. It focuses on adaptive support for navigation. COL-TANGOW is also a system that supports the dynamic generation of adaptive Web-based courses by selecting, at every step and for each student, the most suitable activities to be proposed.

Nowadays, the evolution of the Web 2.0 allows us to develop more sophisticated techniques to analyze more efficiently the students' learning process, in order to improve the learning contents and structure of a course. One of the most recent research areas is Learning Analytics (LA) [5], [7], [19] in order to discover and organize the information contained in the educational platform. Its main goal is to discover and organize the existing information in order to extract useful knowledge during the teaching/learning process.

Thus, this work is focused on a case of study in which two sources of information, AutoES (our automatic assessment system for virtual/remote laboratories) and the institutional Learning Management System (LMS), are aggregated to analyze the students' progress and their collaborative learning in virtual/remote laboratories. Guiding this process the following research questions arise:

1. Are the students engaged with the proposed practical activities or are they at risk of quitting the activities?

DOI: 10.9781/ijimai.2014.277

2. Can we create a system that helps to evaluate if the proposed activities are well-designed?

Within the context of the evaluation activities, two different learning processes have been detected. First, the practical experimentation in the virtual/remote laboratory with the virtual machines picked up by AutoES and, second, the students' knowledge creation through the discussion threats contained in the evaluation forums. Both of these learning processes are highlighted by lecturers when they are asked about how they perform the evaluation of students. So, there is a strong need to aggregate both data sources in order to answer to the aforementioned questions. In order to present these aggregated data to lecturers, a dashboard has been developed. This dashboard contains quantitative and qualitative information for lecturers about the students' experimental and collaborative progress during the evaluation procedure. These data will be validated by means of a set of learning indicators and their graphical visualization. In particular, the dashboard shows a set of evaluation events for each activity, the students' social network, the students' timeline for their activities, and some relevant metrics associated to them.

The remainder of this paper is organized as follows. Section II presents the different data sources that are aggregate in order to fulfill our research questions. After that, our proposal of the aggregated Learning Analytics dashboard, as a result of the integration of AutoES and the institutional LMS, is detailed in Section III. Section IV describes the visualization of the selected learning parameters, and Section V discusses the implications of this research work and some recommendations are given. Finally, Section V highlights our final remarks and suggests guidelines for future work.

II. Data Sources

The sources used in this work come from a self-evaluation system, named AutoES (AutoEvaluation System) [21], and the institutional Learning Management System (LMS). From this information, a data aggregation process will be done.

A. AutoES

The main objective of AutoES [21] is the management of the self-evaluation of practical activities with virtual/remote laboratories and the continuous assessment of the students' progress. It is a service-oriented application, which is considered as the latest generation of Internet-based platforms [20]. Using it, students will be able to perform a self-evaluation of their activities, which they performed with remote laboratories. Additionally, AutoES can solve all the errors made in the activity or configure it completely, with a penalty in the mark for the activity.

AutoES has several main benefits for the members of the learning community, especially within the field of distance higher education. First, it minimizes the response time in correcting students' practical activities, allowing the continuous evaluation process to be performed smoothly. Furthermore, it provides a more detailed monitoring of the students' progress, thereby reducing the time spent on the assessments themselves. The importance of these benefits is really significant, since the number of students enrolled in a course with a distance methodology can become very high. Thus, lecturers can focus on other tasks, such as dynamic adaptation of new activities to students' necessities or expanding the existing ones, which in turn improves the learning process more than devoting their time to correcting the students' activities.

AutoES is made of two different parts: the lecturers 'view and the students' view. From the lecturers' view, lecturers can perform subject management tasks such as selecting the activities for the subject, creating different groups with activities adapted to the students' level, checking students' progress by means of reports, etc. This view is presented by a Web application, named *LabManager*, which is accessible by lecturers through any Internet browser. For each particular student, Lab Manager provides last, maximum, and mean qualifications for each activity. Lecturers will be able to assign a student's final qualification according to these previous ones. It also includes the groups to which he/she belongs and the corresponding activities assigned. Note that the system allows lecturers to split up the subject's students by levels or types of activities. In addition, the system provides statistics about the student's run status and run time. From the learning process point of view, lecturers have several indicators of the students' performance, among others, number of tries for each activity, number of successful evaluations per activity and student, number of failed evaluations per activity and student, and a summary of the evaluation logs. Finally, a list of recent reports is stored for each of them, which can be checked by lecturers at will.

From the students' view, AutoES can automatically configure and/or evaluate a particular activity. Every time they check an activity, a report is created that summarizes the results of this checking. This report is presented each time the student checks an activity as a console message. So, students find out which parts of a particular activity are wrong and, additionally, AutoES can help them when they are not able to do a part of the activity. All this information will be automatically updated on the server side so that it can be used by lecturers to improve the learning process and to decide on the students' marks.

The architecture of AutoES is shown in Figure 1, including the interaction between its main elements, which are Web Client, Lab Manager, and Web Server.

Apart from the learning indicators that will be detailed in the next section, AutoES offers a set of parameters for each proposed activity and student:

1) Start date and time.
2) Finish date and time.
3) Number of successfully evaluations for that activity.
4) Number of failed evaluations for that activity.
5) Logs of errors.

These parameters are included in the aggregation data and in the dashboard.

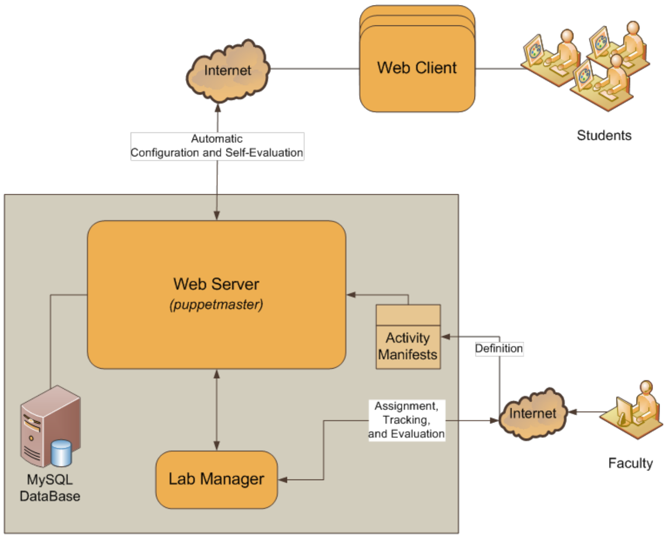


Fig. 1. Architecture of AutoES.

B. The Institutional LMS

The LA process is focused on all the information gathered from all the activities that are crucial for the lecturer's daily-work, especially when applying to a distance methodology. For this reason, a clear necessity of processing all this information appears in order to allow lecturers to extract interesting conclusions for the dynamic adaptation of the learning process to students. It is clear that the information provided by AutoES allows lecturers to have a partial view of the learning process, and it must be combined with the data contained within the LMS.

Therefore, there is a need to enrich the lecturer knowledge of the learning process through the information gathered by both AutoES and the LMS. After the revision of all the relevant educational tools inside the LMS, the forums have been pointed out as the most relevant information source for collaborative evaluation. The use of asynchronous on-line discussion forums is thought to be essential for the negotiation and exchange of ideas, as well as the development of critical thinking skills, all of which are important components of the collaborative learning process [10], [11], [16]. Furthermore, several studies have demonstrated a high correlation of students' participation levels in discussion forums with positive learning outcomes and knowledge constructions [23], [24].

In this sense, as a result of the integration of both learning environments, the aggregation dashboard can graphically show the students' progress both in an experimental and collaborative way at the same time. Therefore, lecturers can guide each student through the learning process based on his/her particular level of proficiency and grade her/him at the end of the term. In particular, the data aggregation, the computation of learning parameters, and their visualization are detailed below.

C. Data Aggregation

As explained before, within the context of the evaluation activities, we have found two learning processes. First, the practical experimentation with the virtual machines picked up by AutoES, and second, the students' knowledge creation through the discussion in the evaluation forums. These learning processes are highlighted by lecturers when they are questioned about how they perform the evaluation of students. Therefore, if we want to represent the learning process into our analysis, at least these two sources of information should be merged: AutoES' events and forums' evaluation messages.

There are other data sources that provide relevant data for the learning process. On one hand, we cannot extract further data from AutoES without changing the basis of the system. Nevertheless, the LMS can offer additional information, such as quizzes scores, activities' deadlines, time spent in the platform, and so on. As this work is a starting point of our research, we only consider the most relevant data sources, but in the future additional data should be aggregated in order to capture all the possible factors. According to this, there are several factors that cannot be obtained neither AutoES nor the LMS, such as personal conditions of students (social environment, health status…), and they may affect the learning process.

Because of the fact that both systems, AutoES and the LMS, have their own data representation, a database merging process is defined. So, in order to have the same representation for both databases, a generic register is created. Afterwards, the data from AutoES and forums are stored in the same database thus further computations are easier.

Each student's interaction is represented by a register within this "merging" process. A register is a structured data generated every time that a student performs an activity which happens at a particular time, and it could produce an output result. Each register contains the following data:

1) A register identifier.
2) The identifier of the user that generates the event.
3) The course to which the students belongs.
4) The type of activity that is represented by the register.
5) An associated report about the activity.
6) The practice associated with the activity.
7) The date and time when the activity takes place.

Thus, a student can produce a set of types of activities inside our learning context, namely:

1) Creation of a user at AutoES (called *created event*). Students enroll themselves dynamically, thus a register is created in this case. In this case, the report field is empty.
2) When a user starts AutoES tool (called *unchanged event*). The report field is empty.
3) A successful evaluation that produces a report as an output result (called *success event*). In this case, the report field contains a brief text that reports about the evaluation.
4) A failed evaluation that produces a report as an output

result (called *fail event*). In this case, the report field contains a brief summary about the errors that are found.

5) Publication of a new thread message inside the evaluation forum, where the message is the output result (called *init message event*). In this case, the report contains the posted message.
6) Response to a previous message inside the evaluation forum, where the message is the result (called *response to event*). In this case, the report contains the posted message.
7) Initiates a new activity, and the previous activity is finished (called *added to event*). The report field contains a reference to the finished activity and the activity field contains the identifier of the just started activity.
8) When a user gives up the AutoES tool (called *removed event*). The report field is empty in this type of register.

For this analysis, not all messages located at the forums are interesting. In this sense, previously to the merging process, the messages have been classified in several topic categories by using the cluster k- means algorithm and a bag-of-words approach. So, messages can be correlated with the evaluation activities due to their content and, additionally, filter which messages are relevant for the learning/evaluation process. Messages not related to the evaluation activities or whose contents are not relevant, are dropped from our study.

III. Description of Learning Indicators

Lecturers should evaluate not only the results, but also the experimental process conducted in the student's virtual/remote laboratory. During the learning process, it is very relevant to detect students at risk of quitting the activities and help them, because if nothing is done, they will not acquire the required learning skills. So, the main objective of this work is answering the following questions using the aggregation of the information of the previously described data sources.

1. Are the students engaged with the proposed practical activities or are they at risk of quitting the activities?
2. Can we create a system that helps to evaluate if the proposed activities are well-designed?

For that purpose, we have computed three indicators that represent their learning outcomes from the activities and four indicators correlated with the behavior of the student in the forums. All the indicators are graphically represented in the aggregation dashboard so lecturers can easily get an overview of the learning progress of each student.

The "*On time*" indicator is focused on the time spent on the realization of the evaluation activity. For the whole population, the average time to solve each activity is computed. Each student's time is compared to this average result by computing the student's corresponding z-score. A higher z-score means that the student is delayed with regard to his/her group and he/she is at risk of quitting. As oppose to this, a lower value means that he/she is solving the activities quickly. This indicator is usually higher at the beginning of the course, and it should decrease as the course goes by and the student is achieving the subject's objectives.

In a similar way, the second parameter called "*Failure rate*", is devoted to analyzing the number of failed evaluations for each student. The number of failed evaluations per activity, the time between failed evaluations, and the student's z-score, calculated by comparing each students' statistics with the average of all the students', are also calculated. A high z-score value means that the student has problems to solve the activity, so the lecturer should offer some additional help. It is also a source of frustration for the student and he/she may decide to quit. A lower value of this indicator means that the student has solved the activity with fewer problems than his/her classmates.

Finally, the third parameter called "*Success rate*" is correlated to successful evaluations. It is computed similarly to the failed evaluation parameter. The number of success evaluations per activity, the time between success evaluations, and the student's z-score, calculated by comparing each student's statistics with the average of all the students', are also calculated. A high z-score value means that the student has not problems to solve the activity. On the other hand, a low value for "*Success rate*" indicator in combination with a high value of "*Failure rate*" indicator could mean that the student is having trouble to solve the activities.

On the other hand, Social Network Analysis (SNA) provides a powerful mechanism for understanding how human relationships are created and developed, as well as detecting communication patterns and structures that should appear from these interactions [13], [26]. In [18], it is proved that there is a correlation between forum interactions and the students' performance. This correlation is also explored by the tool SNAPP [6] and Vercellone-Smith et al. [29] by means of social networks analysis. According to this, a Social Network (SN) can be represented as a directed graph in which nodes are individual or grouped users and links are the relationships among people. Nodes are also used to represent concepts, events, ideas, and other learning elements. These networks are usually built upon gathering and processing the information obtained from the LMS, where interactions among nodes are established in order to acquire new knowledge within a social community.

In our particular case, the creation of a SN graph for the analysis of educational communities is based on the messages published in discussion forums. More in detail, links between two nodes, where each node represents a particular student, are weighted with the amount of messages exchanged [6]. Thus, the analysis performed of the resulting social network allows lecturers to analyze the interest propagation of their group of students, as observed in Figure 3.

We have computed four basic indicators inside the social network that help lecturers analyze the student's progress and their level of proficiency.

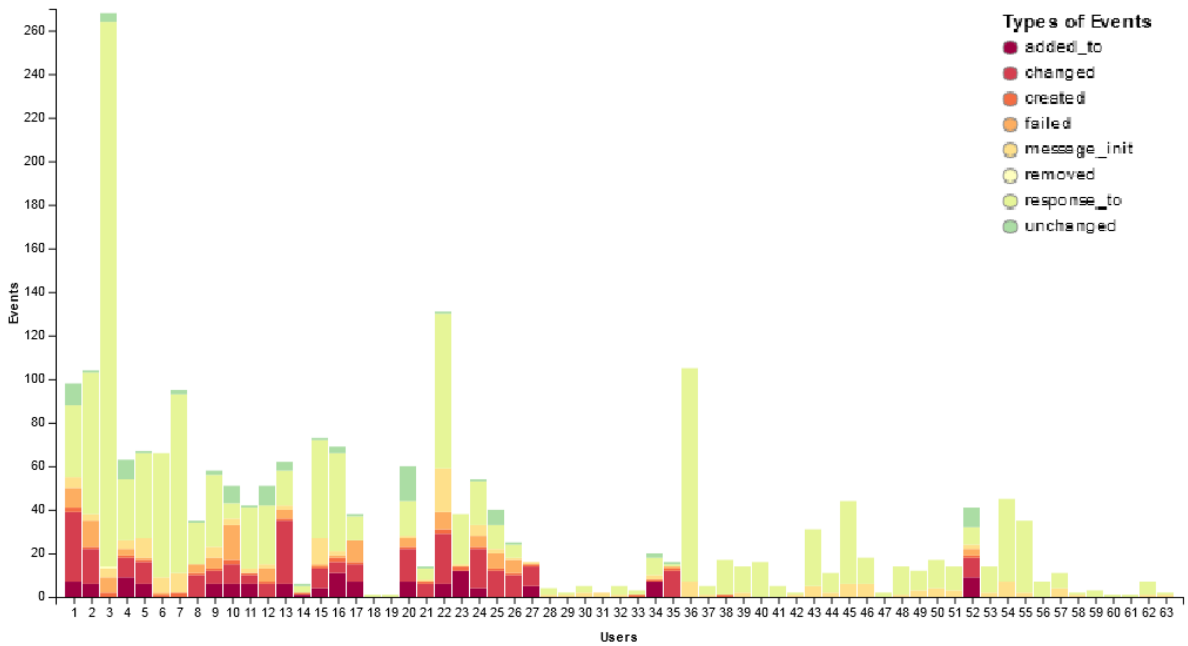


Fig. 2. Events per User by Including Students' Forums Interactions.

There are a variety of different measures to evaluate the importance, popularity, or social capital of a node within a social network:

1) *Degree centrality (interactivity)* focuses on individual nodes, it counts the number of edges that a node has. This value represents the interactivity level of the student; that is, how often the student posts in forums. This indicator could have several meanings that are qualified with the rest of learning indicators.
2) *Betweenness centrality (broker/hub)* of a node is the sum of the fraction of all-pairs shortest paths that pass through that node. Nodes that occur on many shortest paths between other nodes in the graph have a high betweenness centrality score and are more likely to behave as a hub or broker in the network. In this context, students tend to group into communities at forums. The students that behave as hubs or brokers in the social networks allow the exchange of ideas among communities due to the fact that they take part in several of them.
3) *Eigenvector centrality (neighborhood)* of a node, which is proportional to the sum of the centrality scores of its neighbors. A node is important if it is connected to other important nodes. A node with a small number of influential contacts may outrank one with a larger number of non-popular contacts. Thus, this parameter measures the relevance of the neighbors of a student. A better group of neighbors will help a student to create a better collaborative knowledge and it will encourage him to do the activities.

In addition to popularity measures, we pay attention to the *clustering coefficient (integration)* for each student. The bachelorhood of a node that represents a student is a set of nodes connected to it by an edge, not including itself inside the social network. The clustering coefficient of a node is the fraction of pairs of its neighbors that have edges between one another. Locally, this indicates how concentrated the neighborhood of a node is. A higher clustering coefficient means that the student has been exchanging messages with a high portion of the classroom.

The combination of these learning indicators allows lecturers to answer our research questions. According to the first question related to the *student's engagement with the activities*, an interactive student that presents a high degree centrality value with also a high value for "*Success rate*" indicator and a low value for "*Failure rate*" indicator means that the student is helping other students with the activities sharing his/her knowledge. It is common that each community is created with at least one student with these parameters. A particular case is if this student could have a high value as *betweennes centralities* because his/her answers are popular and he/she becomes member of more than one community. And it is also very frequent that the *eigenvector centrality value* of the student is as high as the clustering coefficient. So, this student is going to successfully pass the activities with high scores.

On the other hand, a very interactive student with a high value for *"On time"* indicator and a high value for indicator "*Failure rate*" could have serious problems in order to solve the activity and he/she is searching for help among his/her peers. Lecturers in this case must pay attention to the student and they should offer additional learning resources because he/she is at risk of quitting the activities. Students at risk could also have a low value of *betweenness centrality* and a low value of *eigenvector centrality* because he/she is not posting solutions of problems, which are popular messages.

Our second research question is easily answered starting from the previous results. If during the period of an activity, the number of students at risk of quitting computed with learning indicators has increased substantially, while during other activities the same students have a more successful performance, maybe lecturers should consider redesign the activity. That activity could be too complex or the resources for the development of the activity are not clear. This type of situation is often accompanied by an increase of the number of exchanged messages at the forums under the label of that activity. So, lecturers must be aware of the evolution of the performance of the students during the course. Further analysis over the interrelation of these parameters could help to detect automatically an activity, whose design is not correct by describing some threshold values. But, lecturers should supervise this classification. In addition, these value thresholds could be correlated to the learning context (subject, course…), thus it is a difficult task for automation.

IV. Visualization of Learning Parameters

In order to allow faculty to easily see these indicators and use them to guide the learning/teaching process of students, a dashboard with a Graphical User Interface (GUI) has been developed. Figures 3 and 4 show the most relevant lecturers' interfaces of our proposed LA dashboard.

When a faculty starts browsing in the home view of our LA dashboard, he/she can visualize several graphs, such as the one depicted at Figure 2, which summarize all the events. In addition to offering a global view of what happens in AutoES and evaluation forums at the same time, the lecturer can also observe (as a colored calendar) the set of events generated by each student, as presented in Figure 4(b). This way, a lecturer can easily verify the generated events, and why they are produced. An improvement will be that students and lecturers can compare this activity with the average activity of the course. So, students can be aware of their performance and adapt it in order to improve their learning outcomes.

As we mentioned above, we offer lecturers the possibility of examining the social network generated in the course. The size of students' node is directly proportional to his/her network degree. Additionally, the virtual students' communities represented by the social network are computed by following the Louvain method [2]. Students in the same community are colored with the same color. This visualization is represented at Figure 3.

Finally, we include a graphical visualization for the previously explained indicators; see Figure 4(a). There is a matrix with a cell for each pair <student, indicator>. If a student has a poor performance in an indicator, the cell representing it is colored in a darker red. On the other hand, if the performance of the student for an indicator is good, the indicator cell is colored with a dark blue color. So, lecturers can easily interpret the combined indicators through this graphical representation.

V. Discussion

This section discusses the implications of this research work by taking into account our current learning context.

A. Learning Context

In order to focus this work, AutoES will only use activities related to the configuration of network services. Its scope is much broader, since this system has been designed and implemented as a modular system, which is independent of the design and implementation of specific activities with remote laboratories. In this regard, we focus on the "Network Services Management in Operating Systems" (NetServicesOS) course belonging to the "Communications, Networks, and Content Management" post-graduate program at Spanish University for Distance Education (in Spanish, Universidad Nacional de Educación a Distancia – UNED). The duration of the subject is 15 weeks in the first semester of the academic year. The main goals of the NetServicesOS are the deployment and configuration of several network services for Windows and Linux operating systems, such as DNS, DHCP, FTP, Web, etc., using virtual machines (VMs).

Thanks to the use of AutoES, lecturers can track the progress of a large number of students and adapt dynamically the learning/teaching process. Students can also receive timely feedback on their activities – which was totally impossible with our traditional evaluation system based on explanation reports for each activity.

Since the UNED University follows a distance methodology, the main element of interaction among participants in the learning/teaching process (students and lecturers) are forums, which motivate the learning/teaching process of the subject and allow the formation of virtual social communities. Lecturers play a vital role in promoting a suitable learning space that motivates the interaction among students. In our particular case, lecturers provide students with a set of practical activities which require a great interaction among students to solve them. Lecturers have created a dedicated forum related to the activities for these purposes. The interactions in forums are also taken into account by lecturers when calculating students' final grades.

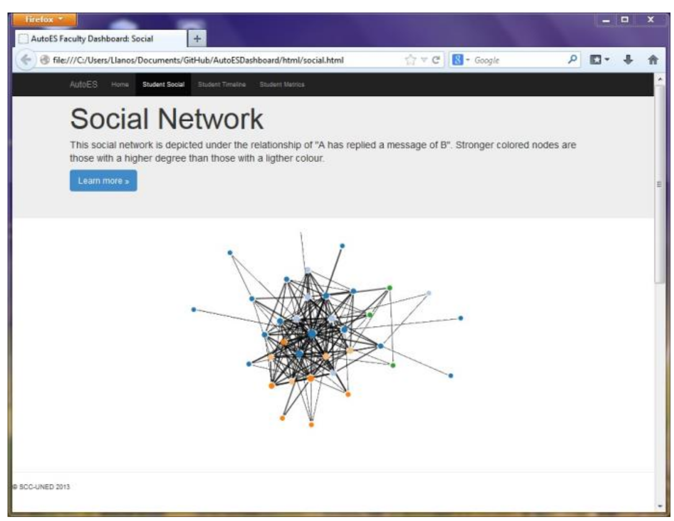
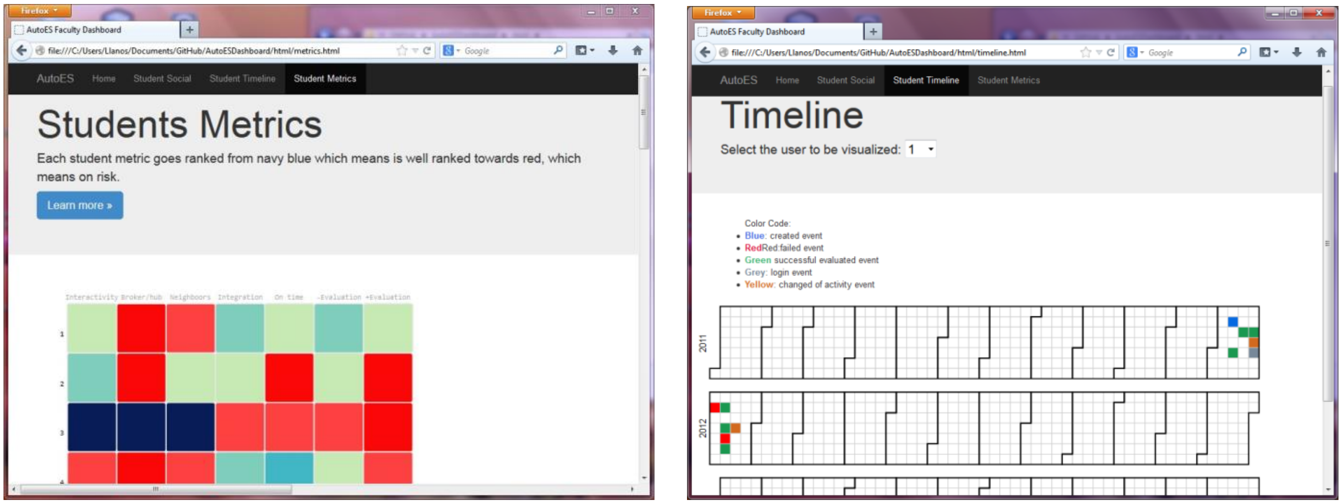


Fig. 3. Representation of the Social Network into the Dashboard.



(a) Table Summary of the Student's Metrics in the Dashboard.

(b) Calendar Interface by Representing Each Student's Event.

Fig. 4. Lecturers' Visualizations in our Proposed LA dashboard (Metrics and Calendar).

B. Results

After the cleaning phase previous to the data aggregation phase, an 83% percentage of messages are relevant to the analysis. The discarded messages are not related to the subject development. Instead, they are Christmas greetings, the place where students can buy/find the bibliography, or students introductions. These topics have a very low correlation with any subject topic, but they are very correlated to an external event. They are initially inactive, although they become very active within a particular time sub-window. After that, they become again inactive. Topic characterization and its impact in the learning outcomes have been widely studied at several works, such as [7], [29], and [30].

As a result of this merging process, there are 2179 events located in the final database, where 1583 are forum's events; this result is depicted in Figure 2. As stated above, there is a high percentage of information related to the student's learning process within the LMS. Therefore, it can be seen that the aggregation of the LMS data provides a large amount of information of great interest in order to guide our learning/teaching process. An initial approach of the statistical analysis of these events shows some relevant results, which are reflected in Figure 2. Each practice takes eleven days to be completed by a student in average.

If we pay attention to the correlation of messages and evaluation activities, we have found that most of the failed evaluations are followed by a message event; almost the 73% (see Figure 5). The visualization of Figure 5 shows the number of events per day. Each circle represents the amount of events. Thus, as the number of events is bigger, the circle is redder and its size is bigger. As we can see at day 4, as example, after the first occurrence of Failed events (which means that students fails the evaluation test), the number of events of "New threads" (which means a student has initiated a new thread in the forum) is increased. The following days the number of events of type "Responses", which mean that students are replying messages at the forums, is higher. This frequency analysis has been completed with a topic detection analysis in order to correlate forum messages and failure events.

Also, at least the 80% of students have posted a correlated message when they are moving from one activity to other. And 68% students who did not use AutoES have replied to the doubts of the AutoES students. This means that the doubts are more related to the development of the experiment itself than to the use of AutoES.

From the Social Network graph, depicted at Figure 3, three communities are detected: blue nodes, green nodes and orange nodes. Most of the students belong to the blue cluster, while the other two are smaller. Each node represents a student, and an arc between two students represents a message exchange. The size of the representation of the node indicates the popularity of the student. There is a clear big blue node in the center of the social network that is the lecturer, who plays a relevant role in the bigger community. In the same figure, we can clearly see that there are several students which have exchanged a very few messages. These students could be at risk and the lecturer should pay attention to the other learning indicators.

There are 36 students in the classroom who decided to work with AutoES. It is also relevant that students execute more than once the evaluation of each practice if he/she has obtained a successful evaluation. This fact can be easily detected with the timeline representation (see Figure 4(b)). As an example, the first activity, when a user is successfully evaluated, he/she is evaluated nine times in average. This situation occurs more often at the beginning of the use of AutoES rather than in the last part. So, students need a period of time for learn how to use the AutoES tool. Also, AutoES output must be improved in order to help students with this

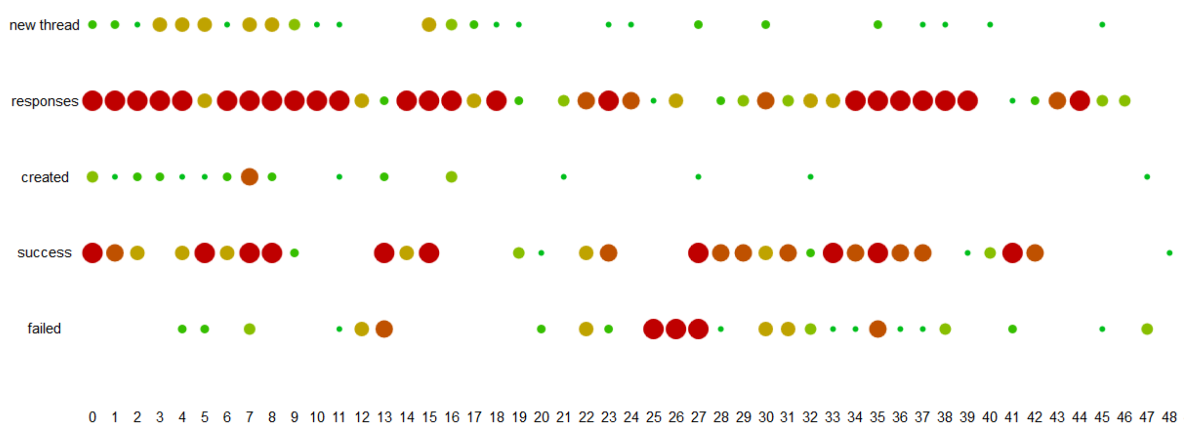


Fig. 5. Graph related to the Evolution of the Activity during the Course.

regard. Another relevant result is that failed evaluations are more common at the first activities rather than the last ones. While the average of failed evaluations of first practice is three by student; the average of failed evaluations of the last activity is 0.47.

There is high percentage of quitting the platform. At least 12 students stopped using the platform at the end of the course. Half of them have quitted the platform during the first activity. This fact is reflected at the dashboard Student Metrics (see Figure 4(a)), such as the learning indicators of the student 3 at Figure 4(a). This student has red values in the "*On time*", "*Successful rate*", "*Failure rate*" and *clustering coefficient* indicators. On the other hand, he has blue values *degree centrality*, *betweenness centrality* and *eigenvector centrality*. This student belongs to the orange community detected in the social network. It seems that he/she has sought for advice at the forum but he/she could not solve the activity. Finally, this student has stopped using AutoES.

On the other hand, at the same Figure 4(a), although student 1 has a red colored value for *betweenness centrality* and a pale value for *eigenvector centrality*, he/she has light green values for the rest of the learning indicators, which means he/she is successfully completing the activities with the help of forums. In fact, this student achieved a high ranking at the course

The activity of students at risk is in some cases average, with pale red values for "*On time*", "*Successful rate*" and "*Failure rate*" indicators, although most of them have strong red values. Out of the 12 students who quitted using AutoES, three of them, as the student 3 of our example, have blue values for *degree centrality*, *betweenness centrality* and *eigenvector centrality*. This means that students were searching for advice at the forums before quitting the activity. The rest of them also have pale red values for the seven learning indicators. Lecturers must detect these students and should offer them additional help to prevent their desertion from the platform.

The number of students that have stopped using the tool is around 30% of the total. On this fact, students were requested to fill in a survey about the tool. According to this, most of the students found AutoES useful and easy to use, as detailed in [21]. From the obtained feedback, the main drawback of AutoES was that students were not confident with the automated evaluation of the tool. This topic has also arisen in the forum messages. Thus, the activities design seems correct. But, it looks like the supporting documentation must be increased.

The obtained results from the research questions presented in this section have been validated with the real lecturers during the courses. There is a correlation among the score of the activities and the information obtained from the proposed system.

The proposed dashboard is useful with this regard. Moreover, students should have a reporting tool, such as a dashboard, that allows them to keep track of their learning process. Lecturers should periodically supervise the results of this dashboard in terms of the design of the activities.

VI. Conclusions and Future Work

This work integrates the information gathered from AutoES, a Learning Analytics (LA) system, and the most relevant tools of our institutional LMS. Therefore, lecturers are able to acquire new useful knowledge in order to improve the learning/teaching process of our subjects. As an example, a subject focused on the configuration of network services has been chosen to implement our approach. In particular, a graphical dashboard has been built from this integration and a set of learning parameters has been analyzed, so that lecturers can guide each student through the learning process based on his/her particular knowledge-level and grade her/him at the end of the term.

As a future work, we plan to improve the functionality of the system by developing alternative indicators for the analysis of the aggregated data from AutoES and forums' messages, this way improving the adaptation of the evaluation resources to achieve more intelligent curricula [17]. Additionally, we will also aggregate other information sources that can improve the vision of the learning process. Finally, different frameworks or contexts from EHEA, as the ones proposed by the ASEE Educational Research Methods (ERM) Division [1], could be explored in order to analyze if the results obtained are similar and/or there is a need of making some changes.

ACKNOWLEDGEMENTS

Authors would like to acknowledge the support of the following European Union projects: RIPLECS (517836- LLP-1-2011-1-ES-ERASMUS-ESMO), PAC (517742-LLP-1-2011-1-BG-ERASMUS-ECUE), EMTM (2011-1-PL1-LEO05-19883), and MUREE (530332-TEMPUS-1-2012-1-JO-TEMPUS-JPCR). Furthermore, we also thank the Community of Madrid for the support of E-Madrid Network of Excellence (S2009/TIC-1650).

REFERENCES

[1] American Society for Engineering Education Educational Research and Methods Division. Web page at http://erm.asee.org/. Date of last access: June 25, 2014.

[2] V. Blondel, J. Guillaume, R. Lambiotte, and E. Mech. Fast unfolding of communities in large networks. J. Stat. Mech, 2008.

[3] P. Brusilovsky and E. Millán. User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science, pages 3–53. Springer Berlin Heidelberg, 2007.

[4] R. M. Carro, A. Ortigosa, E. Martín, and J. H. Schlichter. Dynamic generation of adaptive web-based collaborative courses. In CRIWG, pages 191–198, 2003.

[5] T. N. M. Consortium. The horizon report (2011 edition). On-Line, 2011. Date of last access: June 25, 2014.

[6] S. Dawson, L. Lockyer, A. Bakharia, E. Heathcote, L. Macfadyen, P. Long, R. Phillips, and P. Poronnik. SNAPP. On-Line, 2011. Date of last access: June 25, 2014.

[7] L. P. Dringus and T. Ellis. Temporal transitions in participation flow in an asynchronous discussion forum. Computers and Education, 54(2):340–349, 2010.

[8] T. Elias. Learning analytics: Definitions, processes and potential. On-Line, 2011. Date of last access: June 25, 2014.

[9] N. Furugori, H. Sato, H. Ogata, Y. Ochi, and Y. Yano. COALE: Collaborative and adaptive learning environment. In Proceedings of CSCL 2002, pages 493–494, 2002.

[10] D. R. Garrison, T. Anderson, and W. Archer. Critical thinking, cognitive presence, and computer conferencing in distance education. American Journal of Distance Education, 15(1):7–23, 2001.

[11] D. R. Garrison and M. Cleveland-Innes. Facilitating cognitive presence in online learning: Interaction is not enough. American Journal of Distance Education, 19(3):133–148, 2005.

[12] E. Gaudioso and J. Boticario. Supporting personalization in virtual communities in distance education. World Scientific Publishing Company, 2002.

[13] R. Hanneman and M. Riddle. Introduction to social network methods (online textbook). On-Line, 2005. Date of last access: June 25, 2014.

[14] I.-H. Hsiao and P. Brusilovsky. The role of community feedback in the student example authoring process: An evaluation of annotex. British Journal of Educational Technology, 42(3):482–499, 2011.

[15] A. Kobsa, J. Koenemann, and W. Pohl. Personalized hypermedia presentation techniques for improving online customer relationships. The Knowledge Engineering Review, 16:111–155, 2001.

[16] J. B. Pena-Shaff and C. Nicholls. Analyzing student interactions and meaning construction in computer bulletin board discussions. Computers and Education, 42(3):243–265, 2004.

[17] A. Robles-Gómez, S. Ros, R. Hernández, L. Tobarra, A. C. Caminero, R. Pastor, M. Rodríguez-Artacho, M. Castro, E. SanCristóbal, and M. Tawfik. Towards an adaptive system for the evaluation of network services. In 2013 Frontiers in Education Conference - Energizing the Future, 2013. FIE'13. 43rd Annual, pages 1–7. IEEE, 2013.

[18] C. Romero, M. I. López, J. M. Luna, and S. Ventura. Predicting students' final performance from participation in on-line discussion forums. Computers and Education, 68:458–472, 2013.

[19] Pardo, D. Burgos, and C. D. Kloos. Monitoring student progress using virtual appliances: A case study. Computers and Education, 58(4):1058–1067, 2012.

[20] S. Ros, R. Hernández, A. Robles-Gomez, A. C. Caminero, Ll. Tobarra, and E. SanCristobal. Open service-oriented platforms for personal learning environments. IEEE Internet Computing, 17(4):26–31, 2013.

[21] S. Ros, A. Robles-Gómez, R. Hernández, A. Caminero, and R. Pastor. Using virtualization and automatic evaluation: Adapting network services management courses to the EHEA. IEEE Transactions on Education, 55(2):196–202, 2012.

[22] C. Saul and H.-D. Wuttke. Towards a high-level integration of interactive tools with e-assessments. In Proceeding of IEEE International Conference on Advanced Learning Technologies (ICALT), 2012.

[23] T. Schellens and M. Valcke. Fostering knowledge construction in university students through asynchronous discussion groups. Computer and Education, 46(4):349–370, 2006.

[24] T. Schellens, H. van Keer, M. Valcke, and B. de Wever. Learning in asynchronous discussion groups: a multilevel approach to study the influence of student, group and task characteristics. Behavior and Information Technology, 26(1):55–71, 2007.

[25] L. Schrum, M. D. Burbank, and R. Capps. Preparing future teachers for diverse schools in an online learning community: Perceptions and practice. The Internet and Higher Education, 10(3):204–211, 2007.

[26] J. P. Scott. Social Network Analysis: A Handbook. SAGE Publications, 2000.

[27] Spanish Government. Ministry of Education, Culture and Sports. What is bologna? (In Spanish). On-Line, 2009. Date of last access: June 25, 2014.

[28] Ll. Tobarra, A. Robles-Gómez, S. Ros, R. Hernández, and A. C. Caminero. Analyzing the students' behavior and relevant topics in virtual learning communities. Computers in Human Behavior, 31(2):659–669, 2014.

[29] P. Vercellone-Smith, K. Jablokow, and C. Friede. Characterizing communication networks in a web-based classroom: Cognitive styles and linguistic behavior of self-organizing groups in online discussions. Computers and Education, 59(2):222–235, 2012.

[30] E. Webb, A. Jones, P. Barker, and P. van Schaik. Using e-learning dialogues in higher education. Innovations in Education and Teaching International, 41(1): 93–103, 2004.

[31] G. Weber and P. Brusilovsky. ELM-ART: An adaptive versatile system for web-based instruction. International Journal of Artificial Intelligence in Education, 12:351–384, 2001.

Llanos Tobarra received her M.Sc. degree in Computer Science in 2004 and his Ph.D. in Computer Science in 2009, both from the University of Castilla-La Mancha, Albacete, Spain. She is currently a Lecturer at the Control and Communication Systems Department at the Spanish University for Distance Education, UNED. Her interests include security support in distributed systems and the analysis of social networks and remote labs for e-learning. She has co-authored more than 30 publications in international journals and conferences on these topics.

Salvador Ros (SM'07) received his M.Sc. degree in Physics, specialized in Control and Automatic Systems, in 1991 at Complutense University, Madrid, Spain, and his Ph.D. in Computer Science in 2012 at the Spanish University for Distance Education, UNED. He is Associate Professor at the Control and Communication Systems Department at UNED. For five years (May 2004-June 2009), he has been an Educational Technologies Manager at UNED. He managed the Virtual Campus and Multimedia Production at UNED. He has also been a Technological Projects Evaluator for the Scientific Investigation, Development and Technological Innovation, Spanish National Program. He is currently a Vice Dean of Technologies at School of Computer Science at UNED. He is a senior member of IEEE.

Roberto Hernández (SM'07) received his M.Sc. degree in Physics, specialized in Electronics, in 1989 at Complutense University, Madrid, Spain. He also received his Ph.D. in Sciences in 1994 at Spanish University for Distance Education, UNED, and Madrid, Spain. He is Associate Professor at

the Control and Communication Systems Department at UNED. He was the Dean of Technologies at School of Computer Science at UNED from 2005 to 2013. His research interests include quality of service support in distributed systems and development of infrastructures for e-learning. He has co-authored more than 60 publications in international journals and conferences on these topics. He is a senior member of IEEE.

Antonio Robles-Gómez (M'10) received his M.Sc. degree in Computer Science in 2004 and his Ph.D. in Computer Science in 2008, both from the University of Castilla-La Mancha, Albacete, Spain. He is an Assistant Professor at the Control and Communication Systems Department at the Spanish University for Distance Education, UNED. He teaches graduate and postgraduate courses related to the network interconnection and security domains. His research interests include quality of service support in distributed systems and development of infrastructures for e-learning. He has co-authored more than 35 publications in international journals and conferences on these topics. He is a member of IEEE.

Agustín C. Caminero (M'10) received his M.Sc. and Ph.D. degrees in Computer Science in 2004 and 2009 respectively, both from the University of Castilla-La Mancha, Albacete, Spain. After that, he was awarded with a Post-doctoral grant at Complutense University, Madrid, Spain. He is an Assistant Professor at the Control and Communication Systems Department at the Spanish University for Distance Education, UNED. His interests include quality of service support in parallel distributed computing systems, and development of infrastructures for e-learning. He has co-authored more than 35 publications in international journals and conferences on these topics. He is a member of IEEE.

Rafael Pastor (M'06) received his M.Sc. degree in Physics in 1994 at Complutense University, Madrid, Spain. He also received his Ph.D in 2006 at Spanish University for Distance Education, UNED, Madrid, Spain. He is an Associate Professor at the Control and Communication Systems Department at UNED. From 1994 to 2009, he worked at the UNED Computer Sciences Faculty, and as Innovation Manager of the Innovation and Development Centre of UNED. Since then he has been a General Manager, adding innovative services in the learning model of UNED. He is a member of the IEEE, Spanish Education Society, and .LRN Board Consortium.

Social Networks as Learning Environments for Higher Education

J.A.Cortés[1], J.O.Lozano[1]

[1]Systems Engineering Program, Cooperative University of Colombia

***Abstract* — Learning is considered as a social activity, a student does not learn only of the teacher and the textbook or only in the classroom, learn also from many other agents related to the media, peers and society in general. And since the explosion of the Internet, the information is within the reach of everyone, is there where the main area of opportunity in new technologies applied to education, as well as taking advantage of recent socialization trends that can be leveraged to improve not only informing of their daily practices, but rather as a tool that explore different branches of education research. One can foresee the future of higher education as a social learning environment, open and collaborative, where people construct knowledge in interaction with others, in a comprehensive manner. The mobility and ubiquity that provide mobile devices enable the connection from anywhere and at any time. In modern educational environments can be expected to facilitate mobile devices in the classroom expansion in digital environments, so that students and teachers can build the teaching-learning process collectively, this partial derivative results in the development of draft research approved by the CONADI in "Universidad Cooperativa de Colombia", "Social Networks: A teaching strategy in learning environments in higher education."**

***Keywords* — Collaborative learning; Digital Environments; mobile devices; Social Networks.**

I. Introduction

EDUCATIONAL environments, are immersed in the processes of innovation, which are framed in a set of social and technological transformations. These are given by the changes in information and communication, this is why the social relations and a new conception of relations technology-society identified trends in society. Communication networks introduced a technological configuration that enhances learning more flexible and, at the same time, the existence of new learning scenarios in particular as regards the use of social media in education.

Castells [3] defines Technology as "the use of scientific knowledge to specify ways of doing things in a reproducible manner." This technology has caused a profound revolution in all fields including Education, especially characterized by the appearance of multimedia devices and a dramatic expansion of such as Telecommunications networks. The speed of information processing grows constantly and almost unlimited storage capacity.

Currently, there are terms such as e- Learning, e-Commerce, e -Business, etc., extending the terms m -Learning, m - Commerce related to mobile environments and finally comes to personal atmosphere defined as PLE (Personal Learning Enviroment); these are the different ways to characterize the population living with technology and are the further evolution of structures and components thereof whenever required. These generations and environments that have been incorporated into their lifestyles will be located in the third wave posed Toffler [19] as the " Information Society " supported by advances in information technology and telematics. With the advancement of telecommunications is expected to be greater participation of individuals in the production of information; production with concern Cartier [2] investigated and called the term media, whose object of study is the content traveling the net and how they can be interpreted in a more meaningful way to integrate various means of expression such as text, sound, images defined as static and dynamic Multimedia. These changes and concepts are reflected in the concept of service integration and favorable to smart growth "smart" communications devices that are being used throughout the Knowledge Society and Information Technologies.

A. Research Problem

Social networks are structures composed of groups of people, which are connected by one or more types of relationships, such as friendship, kinship, common interest or shared knowledge.

Today, virtual learning environments (VLE), provides a space for academic interaction mediated by information technology and telecommunications (ICT's), which offer many features, resources and tools for collaborative work, making it a good tool for development of formative research, as frequent interaction among members generates diversity of ideas, approaches and insights that lead to the achievement of a joint and meaningful learning.

Within virtual learning environments, are several services that enable you to perform the educational process by encouraging the learning of students or users. These development platforms have allowed adapt educational environments such as LMS (Learning Management System) or learning system manager. Currently exist various digital platforms which are used massively in educational environments due to its low cost ; the use of software platforms such as Blackboard or Moodle, to allow virtual support the academic, automating these processes together have enabled the emergence of new models of teaching and learning; these models have allowed each student to have

DOI: 10.9781/ijimai.2014.278

personal computers available for exclusive use in any environment, in addition to the resources of the institution Hunt [9], as their own . These new scenarios in which students interact with information networks creating interactive generation where the use of " Netpods " and social networks , however not taken into account the institutions for their educational processes, is common that the students and executives continue to consume Blackberries, SmartPhones and Tablets in their daily work , while schools and other educational institutions remain in a primitive "pre- digital " state, due to disuse of distributed equipment Gagné [5].

Understanding technology as a support to improve educational processes, means that institutions regularly do a review of their learning environments (data centers, licenses, software, broadband, electronic library, laboratories, etc.). What it is to take stock: what is, what is obsolete, what needs to be renewed or updated? This knowledge, ultimately, will allow institutions to have a true picture of their technological capacity and act promptly without incurring higher costs.

This project aims to define models that have access to the virtual learning environments (VLE) through social networks, in order to extend the benefits of the students in the Cooperative University of Colombia, these settings allow you to manage learning through online courses generated by teachers, where students have greater access to courses and information, these computing resources are expected to generate new knowledge

The information society produces spaces of flows such as technology, places and people called by Castells real virtuality: time without time and without space [3]. These concepts described in "The global village of McLuhan," McLuhan [11] where the presence and incorporation of these technologies into educational models allow us to reduce the time and distance in communication processes. Valuable contributions on these concepts arise and emerge new features such as the cyber society, Joyanes [10] (Formation of social networks) cyberculture (Knowledge of the culture of the society in Red) and cyberspace (feeling in the same spaces in different places).

The development of virtual courses is known in various ways and with varying purposes such as Virtual Education, e - learning, e- trainning, among others, but still varied views about the issue persists; Institutions prefer to acquire technological platforms such as WebCT, Learning Space, Blackboard, Moodle among others, and the institutions are counted starting from scratch development to support the development of courses on NET, Driscoll [4]. It seems that the constant was no longer invest in developing a technology platform but rather lead them to those efforts that teachers begin their process of building materials and can locate trouble on the platform , initiating a "virtual dialogue" with Barabasi [1] , Driscoll [4] students .

In this context the question arises on which this study will answer:

What would be the Mediations supported in Connectivism and social networks, which could be appropriated as differentiators in today's learning environments in higher education?

II. METHODOLOGY(MATERIALS AND METHODS)

A. Hypothesis

The hypothesis proposed research is: Current developments and likewise investigations have relied on traditional pedagogical schools that have been oriented logography , however latest studies and research in virtual learning environments have concluded the iconography and connectivist environments are aspects that due to technological development , are impacting today's learning environments. We believe in this concept, it is necessary to investigate how new developments on the Internet, in particular how social networks are impacting the people and in particular in education, in order to define new educational models that reinforce learning in Higher Education.

Methodology

This research is descriptive qualitative ethnographic court, as it seeks to establish as new generations of students entering higher education using new technologies; where the proposed development is constructed with the use of tools and technologies used by students and described in the background such as forms of connectivity to new Internet services , web development service through LMS and application semantic Web and Social Networks as primary implementation tools for the implementation of effective and efficient services in academic consultations.

For Valles [20], qualitative research is one where the quality`s study of the activities, relationships, business, media, materials or instruments in a given situation or problem is studied. The emphasis is to document all information that is given daily in a given situation or scenario, observe and carry out full and continuous interviews, trying to get the minimum of detail being investigated.

The analyzed population were different semesters of the Faculty of engineering students, and area teachers, different instruments were used in gathering information, in the same way as it is a research of qualitative cutting, were used statistical evaluation tools such as Atlas TI (evaluation in qualitative analysis Software).

The sample was taken on about 35 students and 20 teachers from different semesters of the Faculty of Engineering was applied to a survey; 30 teachers from different educational institutions working in different educational levels underwent a group interview.

Phase 1: Information Gathering and dissemination activities

After selecting the appropriate research design and adequate sample, data on the variables involved in the research were collected, the data is classified and the variables involved in the process were determined, observations were recorded and the data were coded with order to have grounds to establish the instruments. It was very important to find support teachers who formed for this purpose a team of teachers, and helped both of these tasks in preparation , as well as propaganda of the activities in the student community which had incidence direct or indirect.

Techniques and instruments for data collection.

In this phase was designed each one of the instruments used for gathering information such as: the development of surveys, design group interviews and activities that had to do with the proper format of the behavior of individuals who participated in the process, just as processes were determined application of tools and means of collecting training. To carry out the study and information gathering used instruments listed below:

- Surveys
- Semi-structured individual and group interview

Methods and Procedures

In order to discover the concepts and relationships of data (interviews, observations) found, then organize and carry out the analysis to the following:

1. To identify and characterize the categories from sociology and psychology allow linking learning, learning styles and technologies used by students in the new learning environments, we use a theoretical review, supported on different texts authors consulted experts on the subject.
2. Selection and constitution of the group from a survey conducted in the Cooperative University , as a result it was decided to work with students from different semesters and teachers from both the University and faculty who work in Educational Institutions Secondary level facilities the study, which students completed their informed consent.
3. With the purpose of identifying emerging categories in particular learning styles and technologies of communication and used social networks, as well as the environments of learning in different educational institutions, applied surveys and group interviews.
4. The identification of categories of styles of learning and different learning environments, as well as technologies, was based on data from surveys and interviews, which were supported by arrays of relationship. The interpretation of the data was made at a later stage with Atlas TI.

Phase 2: Adaptation of Tools and Implementation

Theoretical review

The study of learning environments supported by technologies, requires that the methodology should be consistent with the theoretical framework, thus establishing the foundations of learning and learning styles and the use of technologies supported by different authors as Papert [14] and Siemens [16] selected by the researchers, we find that the patterns we detect and how we do it in the research on learning with the support of the mass media and connectivism.

Ethnographic method that sets a reflective process and allows approach and establish a trust relationship with the student group allowing inquiry through communication was considered. Authors such as Taylor & Bogdan[18], reflexive ethnography Hammersley & Atkinson[8], is very important for the group of researchers, understand and assess this methodology should be dynamic, and each instrument is considered as a prototype arrangement changes were reviewed that establish the objectives of the research.

The Survey

We consider the use of the survey to do a scan on the stakeholders of the University Cooperative Researchers case Bogotá. This investigation aimed to consider the profile of the groups, the relationship could be established with them to do the work and an exploration of the means used to access knowledge through learning. Was performed in 2 groups of 35 students each and two groups of 30 teachers. Knowledge and use of social networks, virtual learning environments and networks: The question and explores three areas of information questionnaire.

This instrument is applied in physical and the student fills out the form printed in the same format giving the option to fill in the digital format also provided and these were sent to a digital mailbox researcher. The information obtained was tabulated in Excel and recurring topics were entered in the answers.

Group interview

Valles [20] said that the group interview is to expose a group of people to a semi -structured interview guideline, not addressed to an individual but to a group where some free stimuli and sometimes structured to allow established as a structured and free response questions through this structure.

It is in the group interview the possibility of approaching the group of teachers, this strategy allows to inform aspects of research actors, their interests, as well as allowing an everyday space group is allowed to explore the opinions, beliefs, representations on the topics of study.

Two groups of 30 teachers of Informatics Master of Education, for a conversation which was set for an hour and a half where they felt about the concepts of teaching and learning, Knowledge, Intelligence and Virtual Learning Environments were selected. The dialogue was developing with driving interviewer, allowing diverse opinions that were recorded and then applied a survey where personal concepts of each of the participants were embodied.

Phase 3: Data analysis and presentation of results

At this stage there was an emphasis on understanding and interpretation of quantitative and qualitative methods of analysis, the proper interpretation of the data, coding the data, defining research categories and their relationships, generation diaries field, the use of statistical tools and the definition of an appropriate methodology for this research. This step was essential to use a number of techniques and statistical procedures for the collection of information by researchers, was the basis for the evaluation and validation of the effectiveness of strategies designed and necessary process feedback.

Phase 4: Defining Models in social learning environments

As a result of analysis, the results were validated in order to determine standard and suitable for use in environments learning strategies, which required the collaboration of teachers and policy programs, in order to validate the proposed adequate to different teaching practices and the use of technology in educational process

III. RESULTS

The results that are part of this investigation and to allow a methodological proposal to consolidate ICT mediated, social networking and multimedia tools in structuring a virtual learning environment, in accordance with the present trends. In this approach we analyze how the structure should be in virtual courses, detailing each of its parts, especially the theories of David Merrill[12] and Robert Gagné[5] precursors of instructional design; how should be learning in environments connectivist, what should be the new styles of teaching - learning and how educational models should be used in these new environments.

A. Connectivist Learning Environments.

The Horizon Report [6] has raised the technology adoption trends in learning environments for the next years, which are summarized as follows:

1. Knowledge is decentralized, the amount of resources that are available online that allow the production and distribution of content in multiple ways to facilitate the acquisition of knowledge both teachers and students.
2. Technology continues to affect our way of life in all environments, the digital divide is diversified with more products related to access and digital literacy skills, informational, media literacy, creating new environments inequality gaps
3. Technology is not only a medium that has been used to train students, but has become a means of communication and relationships, and a ubiquitous, transparent part of their lives. Social relations is one that has been felt more impact, especially in education. Communication between all stakeholders in education has become more open, multidisciplinary, and multisensory and becomes integrated gradually into all our activities.
4. Teachers and educational institutions are gradually making inroads into digital technologies. So, are increasingly beginning teachers to use in their educational practices different technological resources, from email to those provided in the web such as social networks.
5. How we think about learning environments is changing. In traditional education, learning environments are associated with physical spaces and presentiality. Today, however, the "spaces" where students learn are becoming more community and interdisciplinary time and are supported by technologies associated with communication and virtual collaboration. The time and space are transformed to combine the classroom with the virtual, blurring the boundaries between the two worlds, which are experienced by students as one.
6. Current technologies rely increasingly on cloud computing structures , supporting the information technology tends to be decentralized , deployment of cloud applications and services are changing not only the way you configure and use the software and data storage , but also how to conceptualize these functions. No matter where we store our work; what matters is that our information is accessible no matter where we are or what device they have chosen.

B. Teaching Styles.

The proposal of a new distance education model emphasizes mostly on employment of connectivism and ITC in this type of education. Today is orienting the educational process to the use of technology and new learning models. That according to Siemens [16] in his lecture “Connectivism: Creativity and innovation in a complex world”, emphasized that education should aim to promote the development of creativity and innovation in students. This suggested that students should be involved in creating learning content constantly looking to learn creating something new. The current education system reduces the listed capacity, according to the expert, who noted that technology in the classroom allows students to be co-creator and an active participant in their learning.

Teaching styles are linked to the peculiar way that each educator to implement and lead teaching their students. The concept of teaching style or style of education focuses not only on learning, but also in the way of how the individual undertakes, aims or combine various educational experiences. Therefore, the teaching style must be a social environment.

C. Connectivism as way Learning Network

If knowledge is changing so quickly and if such important challenges such as climate change or global warming are tackled , if the technology is changing every day , ¿how can we prepare our students today to take on the challenges of tomorrow? . Today we are moving from a model where education systems and the types of courses we teach are created in advance in the institutions and the student comes to the classroom after we have already created textbooks, materials and resources.

One of the things we have to do, is stop treating intelligence as something that exists inside a person 's head, but rather realize that intelligence exists as a result of contextual knowledge of our environment and relations , social media and technology for our students and ourselves are participants and which are collaborative members. The very structure of learning creates connections in neural networks can be found in the form of linking ideas and ways in which we connect with people and information sources. Our expertise lies in the connections we form, either with others or with information sources such as databases and information systems available on the Web. This appears as a connectivism learning theory for the digital age

In his article “Connectivism: A learning theory for digital age”, Siemens [17] summarizes his theory on the following principles: among which can be highlighted which have to do with learning and its relationship with the networks, the taking of decisions and connections among ideas and concepts.

D. Designing Courses

Structure of virtual courses

To design the Courseware or supported in Virtual environments, now called VLO (Virtual Learning Objects) courses we rely on theories of Instructional Design. The teaching and learning processes are possible because, through an appropriate instructional design, media are used to facilitate students in rich learning situations. Technology is the means, not the end. Unable to assess the usefulness of a specific technology without verifying instructional design.

For proper operation of a virtual course should include the following parameters:

- Building a theoretical model of instructional design that supports the development of the course
- Establish activities that enable online interaction
- Designing the navigation tools
- Structure course information such that the user quickly place
- Design a suitable user interface

The development of the course was considered the following: design course, design of instruction, characterization of the instructor, characterization of the student, the platform components.

E. Design Platform

To build the platform the XP methodology was used, this is an agile methodology focused on enhancing interpersonal relationships as a key to success in software development, promoting teamwork, worrying about learning developers, and fostering a good working environment. XP is based on continuous feedback between the customer and the development team, communication between all participants in the solutions implemented simplicity and courage to face the changes. XP is defined as particularly suitable for projects with very vague and changing requirements, and where there is a high technical risk.

Step 1. Research or planning

Gather all the information regarding the project requirements in order to get to know the specifications of the problem. For this stage theoretical information of each of the key elements of the research, which correspond to Student Virtual Learning Environment and Social Media, was collected. This information was used to design the best way to know that teachers present concepts of the Cooperative University of Colombia on the subject, so propose a solution to the problem recorded in chapter 1 and allows interaction thereof.

Step 2. Analysis

Validate the information gathered above to specify the problem to be solved. Once the survey asks the following analysis and conclusion in general that allowed validating the trend in social networks and knowledge of virtual learning environments were generated.

Step 3. Design and Build

The conceptualization of the information gathered in the investigation phase was conducted.

- Design of the images used in the software.
- GUI design.
- Development of prototype software programming on Moodle.
- List of use cases.

Step 4. Simulation

At this stage all the necessary components were integrated to operate the project. Were designed and set up the courses, the interface was done with social networks and the tests were performed in order to test its operation. Simulation software in the emulator which is located in www.jairolozano.com/uccvirtual.

Fig1. Configuration and integration of Moodle with social networks. Source: Authors

IV. Conclusion

This paper presents the results of research by determining an appropriate structure to develop and integrate virtual learning environments in network. With advances in technology and its foray into the information society and knowledge, it is undeniable that online education is expanding its worldwide coverage , so you cannot be oblivious to its structure and consolidation that allow not only ownership its architecture, but of teaching and learning strategies that flow from them . Productivity Internet, where you see a preview is from 2010, from the semantic web concepts to the detriment of conceptualizations of web 2.0 for this purpose the following activities were performed:

- A proposal for a course mediated by the use of multimedia technologies, Software and Connectivity Social Networking was designed to improve learning environments in higher education
- Some appropriate strategies for using social media in learning environments were defined
- A learning environment where multimedia technologies and social networks such as mediation in the teaching-learning process used was structured.
- Measuring instruments defined methodological spaces in social networks supported in connectivism, which will enable teachers and students to improve the quality of teaching and learning processes.

- Of the above, and as a result of the interviews, are evident new roles, methods, trends and architectures in different Virtual Learning Environments (VLE), these are:

Trends in digital learning environments.

- Teacher training for using digital media in teaching and learning remains a challenge. Know and understand the educational potential of these technologies promote their use in the classroom. The training of teachers from a holistic perspective that incorporates the use of technology resources as an inseparable part of the practice of teaching and learning is the first condition for significant incorporation of digital media in all educational levels.
- Comprehensive change management in higher education must be understood from a systemic and transformative approach that contributes to economic growth, human development and social cohesion. While educational policies cannot be imposed, it is the responsibility of those who have been chosen for this consider, reflect and make decisions to promote the necessary changes; otherwise, we risk that they never occur. This includes a change of role in forcing educational institutions to avoid reflections that everything remains the same, allowing shoot tangible and sustained changes. A redefinition of the educational model that includes new ways to generate, manage and transmit knowledge is required.
- Digital literacy must become an essential skill of the teaching profession. Although there is general agreement on its importance, training in techniques and skills related to the digital realm remains an exception in teacher education programs. The based tools and platforms skills and standards have proven to be something ephemeral, given that digital literacy has less to do with tools like the thought: digital skills have multiple faces (technology, information, multimedia content, and digital identity) and require a comprehensive way to be faced.
- The training of students in the use of new media and audiovisual communication languages is critical. Students need new knowledge and skills in the field of writing and communication, other than those that were needed a few years ago only. Increasingly it is necessary to possess technological expertise to collaborate globally and be able to understand the content and design of new media. For this reason, must be integrated into the curriculum new literacies, and their evaluation, which requires understanding, in its entirety, the meaning and scope of these new skills and competencies.
- Using technology to appropriate treatment of information and knowledge building is still too rare. A key challenge is to not only reflect on the use of emerging technologies by themselves, but put them in the dialectic of information processing to solve complex problems of society, being one of the challenges of higher education. It is not only incorporate technologies or not, but to put forward the needs of student understanding and thinking new ways of working with complex reality we face , to be able to build knowledge about the same .
- Adapting teaching practices to the requirements of the digital society and knowledge is required. Technologies place the student as the protagonist and author in different spaces, but their role is still predominantly receptor in the contexts of formal education. The underlying this phenomenon is that it cannot reduce the proliferation of the use of technology, since many other sociocultural aspects are driving change in existing education and labor practices. The low velocity in the appropriation of technology by the education sector may be due, among other causes, to which teachers are trained as users and not as leaders in the design and implementation of the use of technology for educational purposes. These trends and challenges have a profound effect on the way we experience with emerging technologies and how they implement and use in the educational world.

Acknowledgements

The research is part of the final report of the research project supported by the Cooperative University of Colombia. JA Cortés, Cooperative University of Colombia, Bogotá, Colombia

References

[1] A. L. Barabási. Linked: The New Science of Networks, Cambridge, MA, Perseus Publishing. 2002.
[2] M. Cartier. La médiatique. Editions du Laboratoire de Télématique. Université du Québec à Montréal. Montréal, Canada. 1980
[3] M. Castells (pp 31-419).La era de la información. La Sociedad Red. Tercera Edición. Vol. 1 España: Alianza Editorial. 2005.
[4] M. Driscoll. Psychology of Learning for Instruction. Needham Heights, MA, Allyn & Bacon. 2000.
[5] R. M. Gagné. Las Condiciones del Aprendizaje. Madrid: Ed, Interamericana. 1979.
[6] I. García, I. Peña-López,; L.Johnson, , R.Smith, , A. Levine, , & K. Haywood. Informe Horizon: Edición Iberoamericana 2010. Austin, Texas: The New Media Consortium. 2010
[7] J.C. Gleick. The Making of a New Science. New York, NY, Penguin Books, 1987.
[8] M. Hammersley, & P. Atkinson. Etnografía. Métodos de investigación.2001
[9] T. Hunt. Desarrollar la capacidad de aprender. La respuesta a los desafíos de la era de la información. Editorial Urano. Barcelona. 1997
[10] L. Joyannes. Cibersociedad. México: Ed. Mc Graw Hill. 1997
[11] M. McLuhan. La Aldea Global. Madrid: Ediciones GEDISA. 1995
[12] D. Merril. Educational Technology, New York: Li & Jones, 1991.
[13] N. Negroponte. Ser Digital. Ed. Atlántida. 1995.
[14] S. Papert. La informática en el aula: Agentes de Cambio.By Seymour Papert This article appeared in The Washington Post Education Review Sunday, October 27, 1996 The Washington Post Revisión de la Educación Domingo, 27 de octubre 1996
[15] H. Rheingold. La Comunidad Virtual. Ed. Addison Wesley. 1993.
[16] G. Siemenes. Conectivismo: Creatividad e innovación en un mundo complejo Universitat de València (España), coordinado por Beatríz Gallardo, George Siemens, Dolors Capdet y Paz Villar. 2011.

[17] G. Siemens. Conectivismo: Una teoría de aprendizaje para la era digital. 2004.
[18] S.J.Taylor & R. Bogdan. Introducción a los métodos cualitativos de investigación social. 1996.
[19] A. Toffler. La Tercer Ola. España: Plaza & Janes.1980.
[20] M. Valles. Técnicas Cualitativas de investigación de Social. 1996.

Jairo Augusto Cortes Mendez, the Systems Engineer Incca of Colombia University, Specialist Educational Multimedia Antonio Nariño University, Master in Education Management from the University of the Andes – Colombia - Bogotá, Master in Information Society and Knowledge of the Universitat Oberta de Catalunya – UOC, España, Diploma of Advanced Studies - DEA in Systems Engineering and Automation-UOC. Currently teaching full-time Cooperative University of Colombia - Bogotá.

Oswaldo Jairo Lozano. Systems Engineer District University Specialist Educational Multimedia University Antonio Nariño, Specialist Networks and Telecommunications Cooperative University of Colombia, Master in Telematics from the Central University of Las Villas - Cuba, Master of Teaching at the University of the Salle. He is currently a full-time Teacher of the Cooperative University of Colombia - Bogotá.

www.elasticbox.com
info@elasticbox.com
blog.elasticbox.com
Twitter: elasticbox

elasticbox

Deploy applications, not servers.

Challenges In Cloud Computing

Automation

- How long to deploy an application?
- What version do I use?
- How do I upgrade applications?

Portability

- How do I change providers?
- What is being used?
- How much does it cost?

Auto-Scaling

- Can my application auto-scale?
- How do I configure auto-scaling?

Disaster Recovery Planning

- Can my application tolerare faults?
- How do I recover my system?

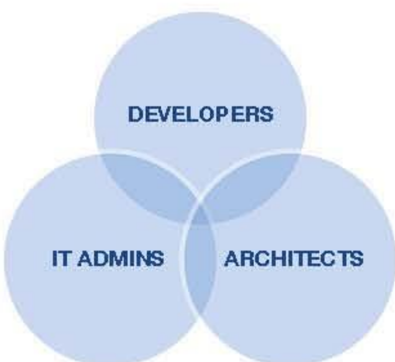
ElasticBox Solution

Automation

- ✔ Automatic Deployments
- ✔ Automatic Configuration
- ✔ Disaster Recovery

Runtime Environment

- ✔ Application Scaling
- ✔ Fault Tolerance
- ✔ Resource Clean-up
- ✔ Replication



Framework Design

- ✔ Architecture Policies
- ✔ Versioning
- ✔ Platform Management

Infrastructure Control

- ✔ Cost Analysis
- ✔ Policy Management
- ✔ Traceability

amazon webservices™ Windows Azure hp rackspace openstack EUCALYPTUS