Applying Bayesian Regularization for Acceleration of Levenberg-Marquardt based Neural Network Training

Azizah Suliman[1], Batyrkhan S. Omarov[1,2] *

[1] College of Computer Science & Information Technology, Universiti Tenaga Nasional, Kuala Lumpur (Malaysia)
[2] International Information Technologies University, Almaty (Kazakhstan)

unir

LA UNIVERSIDAD
EN INTERNET

Abstract

Neural network is widely used for image classification problems, and is proven to be effective with high successful rate. However one of its main challenges is the significant amount of time it takes to train the network. The goal of this research is to improve the neural network training algorithms and apply and test them in classification and recognition problems. In this paper, we describe a method of applying Bayesian regularization to improve Levenberg-Marquardt (LM) algorithm and make it better usable in training neural networks. In the experimental part, we qualify the modified LM algorithm using Bayesian regularization and use it to determine an appropriate number of hidden layers in the network to avoid overtraining. The result of the experiment was very encouraging with a 98.8% correct classification when run on test samples.

Keywords

I. Introduction

CURRENTLY the theories of artificial neural networks (ANN) are interdisciplinary in nature and it is one of the fastest growing disciplines which are used in various scientific and applied fields. ANN has been successfully applied in a wide range of applications - from household appliances to complex computer systems. Artificial neural network methods are also widely used in classification problems. A classification problem is a task to include the sample to one of several disjoint sets. When solving classification problems, ANN should include the existing object characteristics (observable data) to one or more specific classes.

However, there are a lot of problems that remain open in classification. For example, network topology selection problem, determining the number of hidden layers and neurons interpretation of weighting coefficients and bias, their evaluation of optimality, etc. The main idea for a neural network research was to develop mathematical and software tools for modeling processes of human thinking in solving various applied problems. Solving the recognition problems, using conventional methods with strict algorithms and limitations, is not possible. An automated system which solves such problems should not be programmed, it should be learnt. Thus, in our research, in order to recognize and classify we chose Levenberg-Marquardt algorithm.

The classical Levenberg-Marquardt algorithm copes poorly with the situation where the training set contains elements that stand out from the

* Corresponding author.

E-mail address:batyahan@gmail.com

general population. However such types of problems do occur in the practice tasks. In this study, we apply a modification in Levenberg-Marquardt algorithm on the basis of Bayesian regularization to make the algorithm more acceptable for practical tasks in recognition and classification, develop direct distributed neural network that trained modified LM algorithm and test it for recognition and classification problems.

II. Related Works

There are a lot of research to improve Levenberg-Marquardt algorithm to train neural networks, in last years, for example, (An Ru et. al., 2016) in [1] suggested improving the LM algorithm by direct calculation of quasi-Hessian and adopted gradient vector, that does not require storage of the Jacobian matrix. In own case, it causes to reduce the computation time of a neural network training. However this kind of improvement cannot decrease the testing error. Henri P. Gavin, uses LM method for nonlinear least squares curve-fitting problems [2]. In his research, he considers the Levenberg-Marquardt curve-fitting method as a combination of methods for minimization such as the Gauss-Newton method and the gradient descent method.

Thus, Levenberg-Marquardt method improvement comes from improvements of the named two methods. Gauss-Newton method assumes that the least square function is locally quadratic, and it causes to reduce the squared error. The LM method is more similar to the method of gradient descent, when its parameters are far from their optimal value. In contrast, the LM method is more similar to Gauss-Newton method, in case of its parameters are close to their optimum. So, in [2], Henri P. Gavin explains these methods and represents the applying of software to solve problems of approximating of nonlinear least squares. Also, Liyan

Qi proposed LM-method with self adjusting parameters for solving a nonlinear system of equations [3]. By their approach, at each step, the LM parameter μk is automatically adjusted on the basis of the correlation between actual reduction and predicted reduction. Thus, Under the BD-regular condition, they prove that PSA-LMM is locally superlinearly convergent, for semi-smooth equations and locally quadratically convergent for strongly semi-smooth equations. In [4], Haddout and Rhazi try solving the problem of non-linear least squares, based upon LM and Gauss–Newton methods by minimizing the sum of squares of errors between the data and model prediction.

Some researchers use Levenberg-Marquardt algorithm for problems related to prediction. In [5], Murat Kayri did a comparative analysis to Levenberg-Marquardt and Bayesian regularization algorithms from the point of view of predictive abilities.

Despite its effectiveness, the LM algorithm requires large computational cost. To provide the practical suitability, the method demands to diminish the working time for neural networks of large size.

In our research we propose our approaches to improve the method and illustrate experimental results to prove the improvement of the algorithm.

III. Architecture

To provide image recognition and classification using neural network training we create a program complex. Flowchart of the program complex is presented in Fig. 1. Firstly, we create a mathematical model of the Levenberg-Marquardt algorithm and its improvement. After, using the developed mathematical model, we construct a neural network that, following, is trained and its performance is evaluated. After getting the result we compare the improved LM algorithm with classical LM method and the improvements of the other researches.
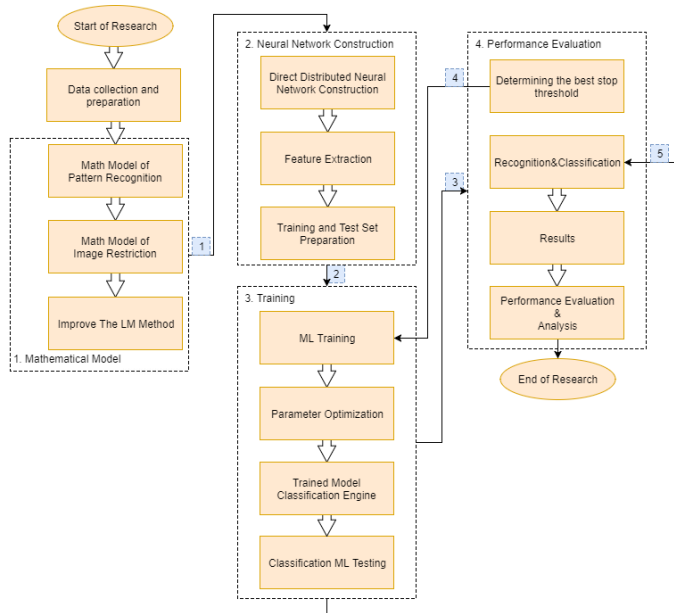

Fig. 1. Architecture of the Image Recognition and Classification Framework.

IV. Bayesian Regularization Based Improvement of LM Method

The standard LM algorithm copes poorly with the situation where the training set contains elements that stand out from the general population, usually such a situation arises when using data obtained by experiment. It means that, the algorithm is inconvenient for practical training. In order to provide the practical suitability, we propose Applying Bayesian Regularization for Acceleration of Levenberg-Marquardt based Neural Network Training. Moreover, applying Bayesian Regularization allows accelerating the LM based Neural Network Training. The essence of the approach is transition from searching the minimum point of mean square error to searching the minimum point of the function expressed by equation (1).

$$F(Y) = \alpha E_\theta + \beta E_D \qquad (1)$$

Here E_D – network error, E_θ –the sum of the squares of the network weights, α and β –hyperparameters.

As a result, the algorithm seeks to minimize the network error and to prevent unlimited growth of its weights. The weight of the neural network can be considered as random variables and their distribution density can be expressed by the formula:

$$P(\theta \,|\, D, \alpha, \beta, M) = \frac{P(D \,|\, \theta, \beta, M) P(\theta \,|\, \alpha, M)}{P(D \,|\, \alpha, \beta, M)} \qquad (2)$$

here D – training set, M – neural network model (in our case, it is feed forward neural network that learnt on the basis of Levenberg-Marquadradt algorithm), θ - weight vector of the neural network. $P(\theta \,|\, \alpha, M)$ - priori probability, reflecting our knowledge of the initial weights of the network. $P(D \,|\, \theta, \beta, M)$ - likelihood function, which is the probability that the neural network with weights θ correctly responds to a set D. $P(D \,|\, \alpha, \beta, M)$ - normalization factor, that provides the equality of the total probability 1. If we assume the training set is noisy with Gaussian noise and the network weights distribution is Gaussian distribution, then formula (2) will be transformed as:

$$P(\theta \,|\, D, \alpha, \beta, M)$$
$$= \frac{\dfrac{1}{Z_\theta(\alpha)} \dfrac{1}{Z_D(\beta)} e^{-(\alpha E_\theta + \beta E_D)}}{P(D \,|\, \alpha, \beta, M)} = \frac{e^{-F(\theta)}}{Z_F(\alpha, \beta)} \qquad (3)$$

Here,

$$Z_D(\beta) = \left(\frac{\pi}{\beta}\right)^{\frac{p}{2}}, \quad Z_\theta(\alpha) = \left(\frac{\pi}{\alpha}\right)^{\frac{S}{2}} \qquad (4)$$

Normalization coefficient can be expressed from the formula (5):

$$P(D \,|\, \alpha, \beta, M) = \frac{Z_F(\alpha, \beta)}{Z_\theta(\alpha) Z_D(\beta)} \qquad (5)$$

Coefficient Z_F remains unknown. However it can be approximated by using the same assumptions as in the Levenberg-Marquardt method as equation (6):

$$Z_F \approx (2\pi)^{\frac{S}{2}} \sqrt{\left\|H^{-1}\right\|} e^{-F(\theta)} \qquad (6)$$

here H^{-1} - inverse matrix to approximate Hessian matrix. Then

$$\alpha = \frac{\gamma}{2E_\theta}, \quad \beta = \frac{p - \gamma}{2E_D} \qquad (7)$$

Here $\gamma = S - 2\alpha\, trace([H]^{-1})$, $\gamma \in [0, S]$ - parameter reflecting the number of weights of a neural network taking part in decreasing a function $trace(H^{-1})$ that is the sum of the diagonal elements of the inverse matrix to approximate Hessian matrix.

In this work to calculate hyperparameters α and β we use the formula proposed by Jan Poland in the work [6, 7]:

$$\gamma = S - \alpha\, trace\left([H]^{-1}\right)$$

$$\alpha = \frac{\gamma}{2E_W + trace([H]^{-1})}$$

$$\beta = \frac{p\,N}{2E_d} \tag{8}$$

As a result, neural network is less susceptible to fluctuation in the training set and accurately approximates the function that is given by the training set.

V. Experiment Results

To test the proposed approach, three well known problems such as face recognition, lung cancer classification and object classification problem, have been applied. To test the problem the data set was divided into two parts, a training set and a test set. The training set data were 70% of instances of each species, and in the test were about 30% of instances data. Fig. 2 demonstrates samples of the each training dataset.

Face recognition experiments were carried out upon the basis of a facial images database by Dr. Libor Spacek [16]. The database contains face images of 395 different people, with 20 images of each person. The test sample consists of 20 images of each person (only 20 × 395 = 7900 images).

For image classification, in the presented research we use JSRT database [17] that includes 147 digital images of lung cancer image samples of two types (small-cell and non-small-cell types), with the 512x512 size.

For objective recognition as human, car, and other objects, we use Omnidirectional and panoramic digital images dataset. The dataset includes 30 omnidirectional images for human detection and 50 omnidirectional images for car detection [18].
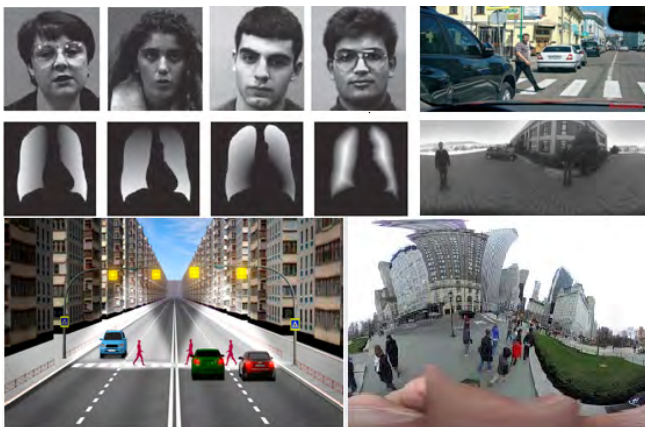

Fig. 2. Samples of the training dataset.

In particular, the problem of choosing the number of hidden layer neurons was considered in [7-13, 19-21]. In [11] authors state that the optimal number of neurons in the hidden layer (N^h) should be calculated from the formula (9).

$$N^h = \sqrt{N^{(i)} N^{(o)}} \tag{9}$$

Here $N^{(i)}$ is number of input layer neurons, $N^{(o)}$ is the number of output layer neurons.

So, we should check this in practice. In this work, we carried out a practical study of the influence of the number of neurons in the hidden layer of the neural network in the learning rate and recognition quality. As a selection criteria of number of neurons, number of training epochs of neural network and recognition quality were chosen.
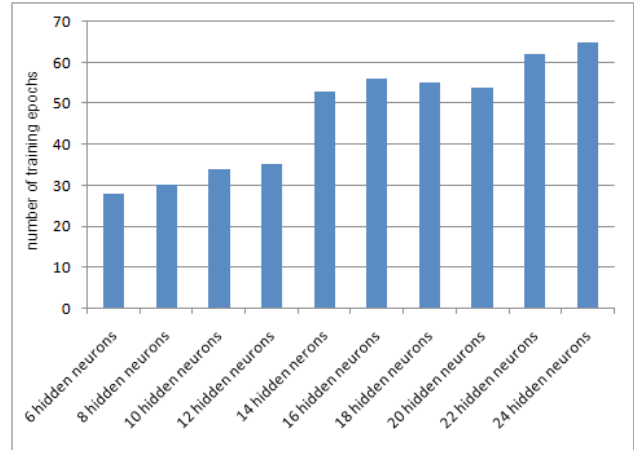

Fig. 3. Variable number of hidden neurons and learning rate.

Fig. 3 shows the ratio of neural network training epochs, when the number of neurons in the hidden layer varied from 6 to 24, in increments of 2 neurons. As it can be seen from the ratio, number of teaching epochs for the neural networks with different number of neurons in the hidden layer, the increase of this number reduces the speed of learning. As a result, not only the number of training epochs increase due to the growth of the Jacobian matrix of the neural network, also the total training time increases. However, this does not mean that the neural network with 6 neurons in the hidden layer will give the best results.

The graphs in Fig. 4 confirm that the minimization of the number of neurons in the hidden layer of the neural network does not improve the recognition quality. 9 and 15 neurons in the hidden layer give the best result. However, the number of training epochs in neural network with 18 hidden layers is substantially greater than others. Therefore, we assume that the best results in the "learning rate recognition quality" ratio are given by the neural network with 9 neurons in the hidden layer.
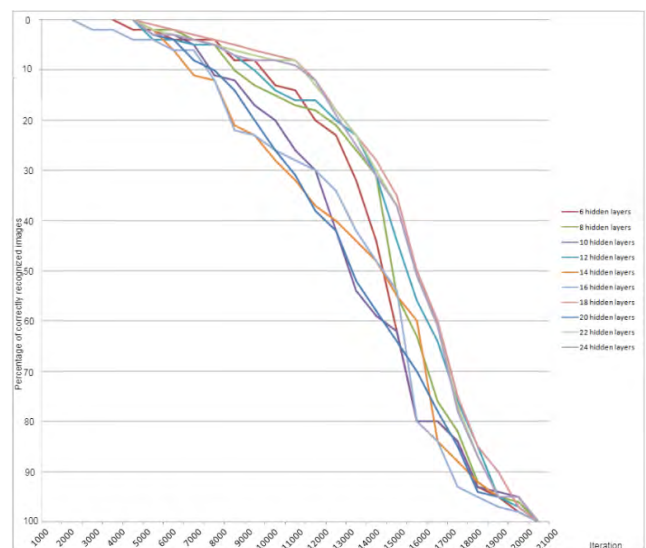

Fig. 4. Testing of neural networks with different number of neurons in the hidden layer.

For objective recognition as human, car, and other objects, we use Omnidirectional and panoramic digital images dataset. The dataset includes 30 omnidirectional images for human detection and 50 omnidirectional images for car detection [18].
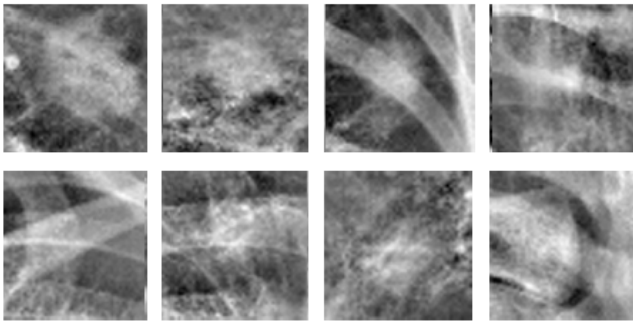

Fig. 5. Classification example of lung cancer image samples.

The results of the classification experiments consider correctly recognized training and test sets that were classified by using modified LM algorithm and standard LM algorithms, which are proposed by other researchers. As the experiment results show, modified Levenberg-Marquardt algorithm gives the highest training and test accuracy with the minimum sensitivity. Fig. 5 illustrates the classification of small-cell and non-small-cell lung cancer images of JSRT database. Table I illustrates the results of comparison with the other researchers.

TABLE I.
PERFORMANCE EVALUATION OF THE IMPROVEMENT IN THE LM ALGORITHM

Classifier	Percentage of Correctly Classified Training Samples%	Percentage of Correctly Classified Test Samples%	Train Validate to Test Partition %
The Proposed LM	99.3	98.8	60-40
Satish and Ritu, 2014 [11]	92.8	87.5	-
Aimi et. al., 2013 [12]	94.75	92.53	70-30
Devesh, 2014 [13]	84.36	-	-
Muhammad et. al, 2013 [14]	90.53	88.4	-
Abdel-Zaher et.al, 2016 [15]	99.03		54.76-45.24

VI. Conclusion

A mathematical model of pattern recognition using neural networks was proposed. Modification of Levenberg-Marquardt algorithm using Bayesian regularization was described and tested. Using the modified method feed forward neural network was constructed and tested for classification problems. The proposed method fulfills better in digital image classification, and maintains a good quality trade-off between the classification rate and sensitivity. The proposed method is effective from the point of view of computationally cost. Accordingly, the proposed method can be a valuable tool for medical data analysis and classification.

The predicted method better suits the classical problem, and also supports a trade-off between sensitivity and specificity. The proposed method is also effective in cooperation. Therefore, the proposed method can be a useful tool for classification in the field of data mining.

References

[1] AN Ru, LI Wen Jing, Han Hong Gui. QIAO Jun Fei. An Improved Levenberg-Marquardt Algorithm with Adaptive Learning Rate for RBF Neural Network. Proceedings of the 35th Chinese Control Conference July 27-29, 2016.

[2] Henri P. Gavin. The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems. March 22, 2017.

[3] Liyan Qi, Xiantao Xiao and Liwei Zhang. A Parameter-Self-Adjusting Levenberg-Marquardt Method For Solving Nonsmooth Equations. Journal of Computational Mathematics Vol.34, No.3, 2016, 317–338.

[4] Soufiane Haddout and Mbarek Rhazi. Levenberg-Marquardt's and Gauss-Newton algorithms for parameter optimisation of the motion of a point mass rolling on a turntable. European Journal Of Computational Mechanics Vol. 24 , Iss. 6, 2015.

[5] Murat Kayri. Predictive Abilities of Bayesian Regularization and Levenberg–Marquardt Algorithms in Artificial Neural Networks: A Comparative Empirical Study on Social Data. Mathematical and Computational Applications, May 2016.

[6] Altayeva, A.B., Omarov, B.S., Aitmagambetov, A.Z., Kendzhaeva, B.B., Burkitbayeva, M.A.. Modeling and exploring base station characteristics of LTE mobile networks. Life Science Journal 11(6),31, pp. 227-233. 2014.

[7] Poland J. On the Robustness of Update Strategies for the Bayesian Hyperparameter alpha. Jan Poland, November, 2001.

[8] Omarov, B., Suliman, A., Kushibar, K. Face recognition using artificial neural networks in parallel architecture. Journal of Theoretical and Applied Information Technology 91 (2), pp. 238-248. (2016). Islamabad.

[9] Omarov, B., Suliman, A., Tsoy, A. Parallel backpropagation neural network training for face recognition. Far East Journal of Electronics and Communications. Volume 16, Issue 4, December 2016, Pages 801-808. (2016).

[10] A. Altayeva, B. Omarov, H.C. Jeong, Y.I. Cho. Multi-step face recognition for improving face detection and recognition rate. Far East Journal of Electronics and Communications 16(3), pp. 471-491.

[11] Satish Saini, Ritu Vijay. 2014. Optimization of Artificial Neural Network Breast Cancer Detection System based on Image Registration Techniques. International Journal of Computer Applications (0975 – 8887) Volume 105 – No. 14, November 2014.

[12] Aimi Abdul Nasir, Mohd Yusoff Mashor, and Rosline Hassan. 2013. Classification of Acute Leukaemia Cells using Multilayer Perceptron and Simplified Fuzzy ARTMAP Neural Networks. The International Arab Journal of Information Technology, Vol. 10, No. 4, July 2013.

[13] Devesh Batra, 2014. Comparison Between Levenberg-Marquardt And Scaled Conjugate Gradient Training Algorithms For Image Compression Using MLP . International Journal of Image Processing (IJIP), Volume (8) : Issue (6) : 2014.

[14] Muhammad Ibn Ibrahimy, Md. Rezwanul Ahsan, Othman Omran Khalifa. 2013. Design and Optimization of Levenberg-Marquardt based Neural Network Classifier for EMG Signals to Identify Hand Motions. MEASUREMENT SCIENCE REVIEW, Volume 13, No. 3, 2013.

[15] Abdel-Zaher Ahmed, Eldeib Ayman. 2016. Breast cancer classification using deep belief networks. Expert Systems With Applications, 46 (2016) 139–144.

[16] Dr Libor Spacek. 2016. Collection of Facial Images. Designed and maintained by Dr Libor Spacek. Electronic resource. http://cswww.essex.ac.uk/mv/allfaces/index.html, June 2016.

[17] JSRT Database. 2016. Public Interest Incorporated Association Japanese Society of Radiological Technology. Electronic resource. http://www.jsrt.or.jp/jsrt-db/eng.php, 2016.

[18] Pinz A. 2016. Human, car, other object database, Electronic resource, http://cvrg.iyte.edu.tr/datasets.htm, 2016.

[19] Boulid, Y., A. Souhar, and M. E. Elkettani, "Handwritten Character Recognition Based on the Specificity and the Singularity of the Arabic Language", International Journal of Interactive Multimedia and Artificial

Intelligence, vol. 4, issue Regular Issue, no. 4, pp. 45-53, 2017.

[20] Harish, B. S., and S. V. A. Kumar, "Anomaly based Intrusion Detection using Modified Fuzzy Clustering", International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, issue Regular Issue, no. 6, pp. 54-59, 2017.

[21] Reguieg, S., and N. Taghezout, "Supporting Multi-agent Coordination and Computational Collective Intelligence in Enterprise 2.0 Platform", International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, issue Regular Issue, no. 6, pp. 70-80, 2017.

Azizah Suliman

Azizah Suliman is an Associate Professor at College of Information Technology, Universiti Tenaga Nasional, Malaysia. She received her BCS from Southern Illinois University, USA in 1985, MCS from Universiti Malaya in 1990 and PhD in Computer Science (AI) from Universiti Putra Malaysia in 2011. Her interests includes Soft Computing, Parallel and GPU Programming, Image Processing and Embedded Systems.

Batyrkhan S. Omarov

Batyrkhan S. Omarov is an Assistant Professor at International Information Technologies University, Almaty, Kazakhstan. He received his BCS and MSC from Al-Farabi Kazakh National University, Almaty, Kazakhstan in 2008 and 2010 respectively. From 2014 till 2017 did PhD research at Unversiti Tenaga Nasional, Kuala Lumpur, Malaysia. Research Interests include Computer Vision, Machine Learning and Data Science.