다층신경망을 위한 온라인방식 학습의 개별학습단계 최적화 방법

이태승[0] 최호진
한국항공대학교 한국정보통신대학교
thestaff@hitel.net hjchoi@icu.ac.kr

Local-step Optimization in Online Update Learning of Multilayer Perceptrons

Tae-Seung Lee[0] Ho-Jin Choi
Hankuk Aviation University Information and Communication University

Abstract

A local-step optimization method is proposed to supplement the global-step optimization methods which adopt online update mode of internal weights and error energy as stop criterion in learning of multilayer perceptrons (MLPs). This optimization method is applied to the standard online error backpropagation (EBP) and the performance is evaluated for a speaker verification system.

I. Introduction

It is well known that the standard error backpropagation (EBP) algorithm for training multilayer perceptrons (MLPs) converges very slowly to a good local optimum. Many researches have been conducted to resolve the difficult problem and achieved substantial fruits [1]. Most of the learning-improving methods have their original ideas on manipulating properties of cost function in the space of network weights and may be referred to as the global-step optimization of learning. Comparably, attempts to reduce the amount of learning data and shorten processing time of epochs have been also sought especially in pattern recognition tasks which have abundant redundancy within learning data [2-3]. The kinds of methods may be quoted as the local-step optimization of learning.

Among the two update modes of weights, online and offline, the online mode has shown very rapid convergence in pattern recognition tasks due to its characteristic to utilize the rich redundancy [4]. However, in spite of the characteristic the awful amount of learning data for high recognition rate is an obstacle to fast learning. In this paper, we attempt to adopt the principles of the local-step optimization and the online update, making MLP learning short. Since the method to be proposed belongs to the local-step optimization, it can be combined with any global-step optimization methods adopting online update mode of weights and error energy as stop criterion. However, we will apply the local-step optimization method to the standard online EBP in this paper.

II. Local-step Optimization Method

MLPs learn models of learning by establishing decision boundaries that discriminate the model areas. If patterns of models are fully presented in an iterative manner and the internal weights of an MLP are adjusted so that all patterns of each model are classified into their corresponding model, and the decision boundaries will finally be settled within the optimal positions.

The basic idea of the proposed local-step optimization method comes from the fact that the patterns far behind decision boundaries in each model cannot contribute to the establishment of decision boundaries any more. The useless patterns resulting from the proceeding of learning generate trivial local gradients and the exclusion of the patterns does not affect the effective progress of learning. The usefulness of given patterns during an epoch may be determined from their approaching to the objective error energy to be set before learning. An epoch is the duration that all the learning patterns are given to be processed and the determination to stop learning is raised on the end of each epoch. The learning accomplishment of an epoch in the EBP of online update mode is measured by average error energy e_{avg} for the entire learning patterns P as follows:

$$e_{avg}(t) = \frac{1}{P}\sum_{n=1}^{P} e(n) \quad (1)$$

here, t stands for epoch count and $e(n)$ for a half summation of error energies from all output nodes for given pattern n. Learning continues until the average error energy is less than learning objective E_{obj}:

$$\begin{cases} w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n), & if\ e_{avg}(t) \geq E_{obj} \\ Stop, & otherwise \end{cases} \quad (2)$$

here, w_{ij} stands for weighted link from computational node j to node i. Including the usefulness concept to learning patterns, eqn. 2 can be expanded into the following formula:

$$\begin{cases} w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n), & if\ e_{avg}(t) \geq E_{obj},\ e(n) >= \psi \\ Skip\ weights\ update, & if\ e_{avg}(t) \geq E_{obj},\ e(n) < \psi \\ Stop, & if\ e_{avg}(t) < E_{obj} \end{cases} \quad (3)$$

here, ψ stands for threshold to skip weights update and designates the distance between pattern n and decision boundary. High ψ increases the number of the patterns whose weights update is skipped so that the computation for the skipped patterns would be saved from the processing during each epoch. But the excess will exclude the effective patterns to learning and the average error energy hardly meets learning objective so that learning would be entirely long on the contrary. Therefore, to validate eqn. 3 the following condition should be contented with:

$$\sum_{n=1}^{O} e(n) \approxeq \sum_{n=1}^{O} e'(n) \quad (4)$$

here, $e'(n)$ stands for error energy that the skipped pattern meeting the second condition in eqn. 3 generates and O for the number of the patterns skipped. The meaning of eqn. 4 is that the error energy which the skipped patterns should generate if eqn. 3 were not applied to them must be equal to the error energy generated when eqn. 3 is applied. When the value of the right term in eqn. 4 is greater than that of the left term, the patterns that cannot be learned as much as the difference between the left and the right will lose the opportunity to be learned and an ineffective learning will come out as a result.

Bearing the condition of eqn. 4 in mind, a rational ψ can be defined as below:

$$\psi = \lambda \cdot E_{obj}, \qquad 0 < \lambda \leq 1 \quad (5)$$

here, λ stands for threshold to compensate for the errors from error-born patterns, making e_{avg} finally decreased to E_{obj}. High λ leads the weights update of many patterns, whose errors are below ψ, to be skipped but more epochs will be necessary to decrease some errors from error-born patterns. Therefore, it is important to select the highest λ guaranteeing the same or near number of learning epochs to that by the original learning algorithm. Typically, an appropriate λ is some value below 0.5. Skipping weights update by this optimization criterion induces the effective reduction of unnecessary epoch-relevant computation and results in the local-step optimization in online EBP.

III. Performance Evaluation and Discussion

To evaluate the performance of the proposed optimization method, we use the speaker verification system, speech database, and evaluation conditions developed by Lee et al. [5]. In this evaluation, an MLP contains input vector of fifty elements, one hidden layer consisting of two computational nodes, and one output node. During learning, around 580 patterns are processed per epoch for complete background speakers set of 29 speakers. Evaluation results are those averaged for 1,400 times of learning, each corresponding to the individual enrollment of 40 speakers using different 35 words. Measurements are obtained in terms of the numbers of epochs and learning patterns and equal error rate (EER).

The comparisons in performance for the online EBP (OnEBP) and the proposed optimization method are presented in Table 1 and Table 2. In Table 1, the performances according to two different learning rates 0.5 and 1.0 are described with the same objective error energy 0.005 and $\lambda = 0.1$ for the proposed optimization method. The full learned patterns (Full Lrn. Pat.) means those that pass through the thorough learning procedure (viz. forward and backward propagations) and the update skipped patterns (Up. Skip. Pat.) those whose updates of weights are skipped to make learning short. The converted number (Conv. No.) designates the number of the full learned patterns as converted from update skipped patterns and the converting is made by multiplying 0.4 to the number of update skipped patterns. The figures in the table show that the optimal learning rate may be different in each method (viz. 0.5 for online EBP and 1.0 for the proposed optimization method) but with either optimal learning rate the effect of reducing the number of full learned patterns can be taken by the proposed optimization method. From the results, we know that the proposed optimization method uses only 73 % of the patterns to complete the learning over the online EBP when applying their optimal learning rates.

Table 1: Comparison of the best performances of online EBP and the proposed optimization method

Learning Method	OnEBP		Proposed	
Learning Rate	0.5	1.0	0.5	1.0
EER	1.59 %	1.64 %	1.59%	1.61 %
No. of Epochs	301	313	304	315
No. of Full Lrn. Pat.	174,185	180,969	123,713	90,033
No. of Up. Skip. Pat.	NA	NA	52,086	92,665
Conv. No. of Full Lrn. Pat.	174,185	180,969	144,547	127,099

The performances with 1.0 of learning rate are given as to the number of background speakers in Table 2. The number of background speakers determines the amount of learning patterns during an epoch. This experiment was designed to examine the relationship of the proposed optimization method to redundancy within learning patterns. From the results, it is shown that the larger the amount of learning patterns, the more effective the proposed optimization method. It is also meaningful that all the error rates achieved by the proposed optimization method do not be deteriorated over those by online EBP. A good error rate or recognition rate is the most important performance in pattern recognition tasks.

Table 2: Comparison of the performances of online EBP and the proposed optimization method according to the numbers of background speakers

No. of BG Speakers	10	20	29
EER for OnEBP	2.83 %	2.24 %	1.64 %
EER for Proposed	2.83 %	2.20 %	1.61 %
No. of Full Lrn. Pat. for OnEBP	32,951	85,969	180,969
Conv. No. of Full Lrn. Pat. for Proposed	27,908	66,585	127,099
Ratio of Proposed over OnEBP	84.7 %	77.5 %	70.2 %

The operational mechanism of the proposed optimization method can be inferred from the aspects to error rate and the number of full learned patterns during the learning of a typical sample as viewed in Fig. 1. In this figure, it is found that high λ makes more updates skipped but increases epoch count as well to lower the errors from error-born patterns. It is also noted that the error energy curve of the proposed optimization method with a proper λ goes down similar to that of online EBP, so both local-step optimization and learning reliability can be obtained at a time.

IV. Conclusion

The proposed local-step optimization method was able to achieve the original goal to optimize the processing time for each step and shorten entirely learning duration only if its combining global-step optimization method adopts online update mode of weights and uses error energy as stop criterion of learning. The method proved advantageous in that the means is very simple and the optimization of local-step is effective without any drops in recognition rate. If a great number of learning patterns are required to get an exact classifying performance, the proposed local-step optimization method can show a definite effect.

References

1 Demuth, H., and Beale, M. : 'Neural Network Toolbox User's Guide Version 4' (The MathWorks, Inc., 2001)

2 Hara, K., Nakayama, K., and Karaf, A. A. M. : 'A Training Data Selection in On-line Training for Multilayer Neural Networks', Proc. IEEE Int. J. Conf. on Neural Networks, May 1998, Anchorage, Vol. 3, pp. 2247-2252

3 Lee, T., Choi, S., Choi, W., Park, H., Lim, S., and Hwang, B. : 'Qualitative Discriminative Cohort Speakers Method to Reduce Learning Data for MLP-Based Speaker Verification Systems', Lecture Notes in Computer Science, 2003, **2690**, pp. 1082-1086

4 LeCun, Y. : 'Generalization and Network Design Strategies', Department of Computer Science, University of Toronto, 1989.

5 Lee, T., and Hwang, B. : 'Continuants Based Neural Speaker Verification System', Lecture Notes in Artificial Intelligence, 2004, **2972**, pp. 89-98

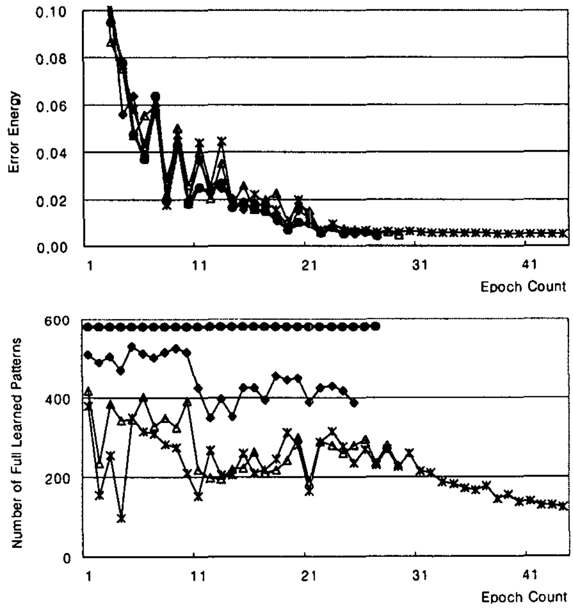

Fig. 1 Operational behaviors of the proposed optimization method in terms of the curves of error energy and the number of full learned patterns

- For the proposed with $\lambda = 0.1$
- For the proposed with $\lambda = 0.4$
- For the proposed with $\lambda = 0.6$
- For online EBP