Expert Systems With Applications 63 (2016) 249–266

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A fast learning method for streaming and randomly ordered multi-class data chunks by using one-pass-throw-away class-wise learning concept

Prem Junsawang*, Suphakant Phimoltares, Chidchanok Lursinsap

Advanced Virtual and Intelligent Computing (AVIC) Research Center, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Patumwan, Bangkok, Thailand

ARTICLE INFO

Article history:
Received 5 June 2015
Revised 30 June 2016
Accepted 1 July 2016
Available online 5 July 2016

Keywords:
Classification
Versatile elliptic basis function
Incremental learning algorithm

ABSTRACT

Presently, the amount of data occurring in several business and academic areas such as ATM transactions, web searches, and sensor data are tremendously and continuously increased. Classifying as well as recognizing patterns among these data in a limited memory space complexity are very challenging. Various incremental learning methods have proposed to achieve highly accurate results but both already learned data and new incoming data must be retained throughout the learning process, causing high space and time complexities. In this paper, a new neural learning method based on radial-shaped function and discard-after-learn concept in the data streaming environment was proposed to reduce the space and time complexities. The experimental results showed that the proposed method used 1 to 95 times fewer neurons and 1.2 to 2,700 times faster than the results produced by MLP, RBF, SVM, VEBF, ILVQ, ASC, and other incremental learning methods. It is also robust to the incoming order of data chunks.

1. Introduction

With the advancement of current technology, tremendous amount of data in various fields such as business, science, medicine, and others have been generated. These data may overwhelm the storage as well as memory space. Obviously, they make the analysis and classification of these tremendous data complex in terms of time and space complexities. The present classical algorithms for analyzing the data are not efficient enough to cope with this new emerging and challenging problem. Classifying patterns existing among data is one of the processes in analysis. This process concerns the problems of developing an efficient learning method to achieve high accurate results within an acceptable time complexity. Classification of patterns can be applied to many machine intelligence studies such as face recognition (Chen, Han, Wang, & Fan, 2011; Wang, Li, & Zhang, 2008), object recognition (Li, Bebis, & Bourbakis, 2008; Serratosa, Alquézar, & Amézquita, 2012), pattern recognition (Khunarsal, Lursinsap, & Raicharoen, 2013; Maglogiannis et al., 2008; Melin & Castillo, 2013; Xinjun, 2010) and pattern classification (Abrahams, Coupey, Zhong, Barkhi, & Manasantivongs, 2013; Khanmohammadi & Chou, 2016; Nguyen, Khosravi, Creighton, & Nahavandi, 2015). Among the proposed learning methods, it was found that neural learning method is rather more efficient and also practical than other methods. It can handle different patterns of data distribution in a high dimensional space and also behaves as either a linear function or a nonlinear function. This ability makes neural learning method possibly achieve a rather low space complexity in terms of number of neurons as well as very high accuracy.

Typically, there are two types of neural learning methods having been proposed. The first type is called *batch* learning (Khanmohammadi & Chou, 2016; Khunarsal et al., 2013; Nguyen et al., 2015). It assumed that a sufficient amount of training data set is presented and a fixed network structure is set up. Furthermore, the testing data are also assumed to have the same statistical distribution as that of the training data (Giraud-Carrier, 2000). No new incoming data are allowed to involve or occur after the training process. Although the batch learning usually provides satisfactory results in classification, it is not suitable to handle the tremendous amount of data available because of time consuming and limited storage capacity. Wilson and Martinez (2003) studied the general inefficiency of batch training for gradient descent learning and concluded that the batch training was not a practical approach for a large training data set. Moreover, the batch learning is not suitable to handle the situation when tremendous amount of new data are generated in every second such as banking

* Corresponding author.
E-mail addresses: jpremlc@gmail.com (P. Junsawang), suphakant.p@chula.ac.th (S. Phimoltares), lchidcha@gmail.com (C. Lursinsap).

http://dx.doi.org/10.1016/j.eswa.2016.07.002
0957-4174/

transactions, biological data, etc. This may be due to the assumption that no new data is allowed to be added during the training process. To cope with the learning time of the batch type, many techniques were proposed to speed up the learning process. The first approach is to reduce the number of dimensions of data (Patra, Widjaja, Das, & Ang, 2005; Yan, Ma, & Zhu, 2006) and also the size of training set (López, Gagné, Castellanos-Dominguez, & Orozco-Alzate, 2015; Verbiest, Derrac, Cornelis, Garca, & Herrera, 2016). Another approach is to modify the learning process in terms of error function (Hua, Chungb, Wanga, & Yinga, 2012; Jiang, Deng, Wang, & Zhang, 2003). These approaches are rather helpful to increase the accuracy but the training time is uncontrollable to reach the desired accuracy.

The concept of *incremental learning* or *continuous learning* was introduced to tackle the aforementioned issues by gradually adding neurons and adjusting the weights according to the training data (Langley, 1995). The goal is to achieve the linear time and space complexities of the learning process. However, the concept did not include the aspect of allowing new incoming data to the training process. The set of training data is fixed during the process. A few pre-processing steps were proposed to improve the performance of incremental learning. These steps concerned the feature extraction. Hall, Marshall, and Martin (1998) introduced a constructive method based on adding incrementally observations to an eigenspace model called *incremental principal component analysis* (IPCA). A modified version of IPCA was proposed by simultaneously performing feature extraction and classification (Ozawa, Toh, Abe, Pang, & Kasabov, 2005). Pang, Ozawa, and Kasabov (2005) proposed an incremental linear discriminant analysis called ILDA in which both between-class and within-class scatter matrices are computed incrementally. They compared the proposed ILDA to original LDA and showed that their method was better and effectively capable of evolving a discriminant eigenspace over a fast and large data stream. Note that every neuron uses the same activation function and only a single network is obtained after the training process.

To further improve the performance, the combination of different types of classifiers was suggested and known as *ensemble learning*. Polikar, Upda, Upda, and Honavar (2001) proposed an ensemble classifier called *Learn++* but the obtained classifier was sensitive to parameter set-up. Wilson and Martinez (2003) proposed the gradient descent incremental learning which is significantly faster than original batch learning but there was no apparent difference in accuracy. Probabilistic RBF or PRBF network was proposed to handle the classification problems (Constantinopoulos & Likas, 2006) by sequentially adding a new component for stationary environment until no component containing data points belonging to more than one class. Duan, Shaob, Houa, Hea, and Zenga (2009) proposed an incremental learning algorithms for Lagrangian Support Vector Machines (LSVM) in both sequential and chunk-incremental learning in which sequential-incremental learning refers to only one sample at a time in each epoch of learning and chunk-incremental learning refers to more than one sample in each epoch. Their results showed that LSVM was faster and more efficient than other sequential and chunk-incremental learning methods based on LSVM. Yi and Wu (2011) presented an incremental SVM based on reserved set for network intrusion detection to reduce the training time. Additionally, memory-base learning methods were presented as incremental learning by which some training data are accumulated incrementally such as an evolving clustering method called ECM (Kasabov, 2002) and a fast prototype-based nearest neighbor classifier called ASC (Shen & Hasegawa, 2008).

Although, these incremental learning techniques claimed a faster speed than that of non-incremental learning, each datum must be repeatedly used to adjust the weights of the network. This leads to the problem of uncontrollable number of repetitions. These methods either required access the learned data many times, forgot the prior knowledge, or could not handle a new incoming class. Therefore, they are not suitable to apply in many practical applications such as data mining, robotics, intrusion detection, business transactions, analyzing real-time satellite images. The several sets of data are presented during the learning process in forms of one-by-one or chunk-by-chunk with various sizes. To solve this new scenario, the learning should be conducted incrementally in one pass called *one-pass incremental learning*. The term *one pass* means that the training data are used or accessed only once for a learning process (Kasabov, 2002).

In the past decade, several incremental learning methods have been proposed under one-pass learning concept to reduce the learning time on large scale data. Sequential and chunk-incremental principal component analysis were developed under one-pass environments to handle large-scale classification problem (Ozawa, Pang, & Kasabov, 2008). The performance of the proposed method was evaluated in terms of classification accuracy and learning time. The results showed that chunk-incremental learning could reduce the learning time effectively as compared with sequential-incremental learning and the chunk-incremental learning could obtain major eigenvectors with fairly good approximation. Jaiyen, Lursinsap, and Phimoltares (2010) proposed a new learning method to speed up the convergence rate in almost linear time based on the structure of *versatile elliptical basis function* (VEBF) neural network. The VEBF is one type of radial-shape function. The proposed learning was conducted under the discard-after-learn concept and one-pass-throw-away learning concept. Their experimental results showed that the classification accuracies were comparable to multilayer perceptron (MLP) and radial basis function (RBF) trained by traditional batch learning. Although, the time and space complexities of VEBF was the lowest among all compared methods but the performance was very susceptible to the orderings effect of an incoming datum. Another disadvantage is that the training samples must be learned one by one even if a chunk of training sample is available at a time. This caused inefficiency in computations because the eigenvector and eigenvalue computation in PCA must be performed to each training sample in the chunk. The method was not suitable to handle incoming data chunk such as banking transaction, intrusion detection, and emerging data on the internet. Xu, Shen, and Zhao (2012) proposed an incremental learning vector quantization (ILVQ) algorithm for pattern classification. The ILVQ outperformed other incremental learning in terms of accuracy and compression ratio. Liu and Ban (2015) applied incremental self-organizing neural network under one pass learning for clustering problem. The compared results showed the superior performance of the proposed algorithm in learning robustness, efficiency, working with outliers without requiring the predefined number of clusters. Recently, Ciarelli, Oliveira, and Salles (2012) and Ciarelli and Oliveira (2015) proposed the incremental learning method called the evolving Probabilistic Neural Network (ePNN) which is an on-line incremental learning method. The method is based on Gaussian Mixture Model and Expectation Maximization (EM) algorithm. Zhou, Zheng, Hu, Xu, and You (2016) proposed a local on-line learning method. In their work, a multiple hyperplane passive aggressive algorithm was integrated with on-line clustering technique. The experimental results achieved notably better performance without using kernel approximation and second order modeling. Fan, Song, and Shrestha (2016) proposed a kernel on-line learning method with adaptive kernel width. This kernel width could be adapted automatically. The simulation results showed that the proposed algorithm could adapt the training data with different initial kernel width. Its performance was better in both accuracy and learning time compared with the kernel algorithms with a fixed ker-