Neurocomputing 133 (2014) 258–270

Contents lists available at ScienceDirect

Neurocomputing

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

Imbalanced evolving self-organizing learning

Qiao Cai [a], Haibo He [b,*], Hong Man [a]

[a] Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA
[b] Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881, USA

ARTICLE INFO

Article history:
Received 17 September 2012
Received in revised form
1 November 2013
Accepted 4 November 2013
Communicated by B. Hammer
Available online 10 January 2014

Keywords:
Self-organizing map
Genetic Algorithm
Imbalanced learning

ABSTRACT

In this paper, a hybrid learning model of imbalanced evolving self-organizing maps (IESOMs) is proposed to address the imbalanced learning problems. In our approach, we propose to modify the classic SOM learning rule to search the winner neuron based on energy function by minimally reducing local error in the competitive learning phase. The advantage of IESOM is that it can improve the classification performance through obtaining useful knowledge from the limited and underrepresented minority class data. The positive and negative SOMs are employed to train the minority and majority class, respectively. Based on the original minority class, the positive SOM evolves into a new stage that might discover novel knowledge. The purpose of convergent evolution is to recurrently search the fitness value via minimal mean quantization error in the feature space, which can motivate the offspring individuals to move toward the center of positive SOM so as to form more explicit boundary. The iterative learning procedure is used to adaptively update the incremental feature maps and create more minority instances to facilitate learning from imbalanced data. The effectiveness of the proposed algorithm is compared with several existing methods under various assessment metrics.

1. Introduction

Imbalanced learning [1] is a popular research topic for a wide range of applications in data mining domain. For the real-world datasets, most of the examples perform the "normal" characteristic while the minority part of data fall in the "abnormal" group. The previous work [3] showed that the detection of abnormal data can reduce the computational cost for a training classifier. For imbalanced data learning, the minority class labels are supposed to equally assigned to these abnormal data. The representative imbalanced data are generally involved in a variety of aspects. For instance, for financial analysts, a huge amount of data are processed and analyzed by various statistical tools, but the existence of the abnormal data brings about more difficulties to precisely learn the useful knowledge from a given dataset. Other typical examples of imbalanced learning include biomedical data analysis (e.g., cancer detection), spam detection, among others. The number of data might be too limited to learn the type of rare cases that are obtained under different conditions. Typically, the ratio of the minority class to majority class is remarkable from 1:100 to 1:100,000 [6]. In this paper, computation intelligence approaches are investigated to tackle these challenging issues.

Most previous work focused on the binary classification problems [2]. The others [33,34] also tried to employ the multi-class data and define the class with a small number of data as the minority class while the other data are merged into the majority class. Although the minority class can be recognized by classifiers, the artificial majority class might be more likely to be misclassified. The knowledge of imbalanced data is complex especially when we solve the multi-class problems, since the amounts of some data classes are the same or similar to each other, which increases the difficulty to artificially select the minority class. The imbalanced learning problems can be summarized as two categories: absolute imbalance and relative imbalance [4]. The absolute imbalance occurs in the situation when the minority instances are significantly scarce and implicit, whereas the dataset with relative imbalance can show explicit data distribution but still rare quantity for minority examples. The characteristic of rare instances exists in the typical imbalance where the limited representative data lead to difficult learning regardless of between-class imbalance. The other form of imbalance is within-class imbalance. It concentrates on the representative data distribution for the sub-concepts within a class. The within-class imbalance problem seems to be more difficult than the datasets with the concepts in a similar characteristic [5].

A variety of solutions have been proposed to address the imbalanced learning. To understand this issue comprehensively, most of the state-of-the-art methods are generalized as the following categories. A critical and comprehensive survey on imbalanced learning can be found in [1].

(a) Sampling methods: Random oversampling for minority instances and undersampling for majority instances can facilitate change of the distribution for original dataset [7]. The informed undersampling using KNN [8] is also presented to achieve

* Corresponding author.
E-mail address: he@ele.uri.edu (H. He).

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.11.010

undersampling. To overcome the disadvantages of the basic sampling methods, the synthetic minority oversampling technique (SMOTE) [9] is to select one from the nearest neighbors for each original minority example, and generate synthetic minority data, based on the linear interpolations between the original examples and randomly selected nearest neighbors. Borderline-SMOTE [10] only generates synthetic data for the minority instances near the border rather than every original minority instance. ADASYN [11] is proposed to adaptively create the different quantities of synthetic data corresponding to the density distribution. Other sampling strategies are integrated with ensemble learning techniques [26,49] to address the imbalanced learning issue. The SMOTEBoost [12] algorithm is achieved via combining SMOTE with Adaboost.M2. RAMOBoost [24] adjusts the sampling weights of minority class examples based on the data distributions.

(b) Cost-sensitive learning methods: Unlike various sampling strategies to change the data distribution, the goal of cost-sensitive learning [35,37] is to calculate the costs for misclassification through different cost matrices. AdaCost [36] adopts the cost-sensitive learning with boosting. Cost-sensitive decision tree [13] can prune the scheme for imbalanced data with misclassification costs through specifying decision threshold. Cost-sensitive neural network models [29,38] can be widely applied for imbalanced learning.

(c) Kernel-based learning methods: The kernel-based learning approaches include many state-of-the-art techniques for the application of data mining domain [25,30]. A Granular Support Vector Machines-Repetitive Undersampling (GSVM-RU) algorithm [14] carries out the iterative learning procedure based on GSVM. Kernel-boundary alignment (KBA) [23] is proposed to modify the kernel matrix via a kernel function based on the distribution of imbalanced data. There is another typical kernel-based learning algorithm for maximizing area under curve (AUC) of the receiver operating characteristic (ROC) graph [42].

(d) Active learning methods: These active learning methods [39–41] are traditionally adopted to handle the special issues relevant to training data without class labels (unlabeled data). As mentioned in [16], the criteria of termination for active learning methods are investigated to apply for the class imbalance issues on word sense disambiguation (WSD) through maximal confidence and minimal error.

In this paper, the hybrid learning model of imbalanced evolving Self-Organizing Maps (IESOM) with Genetic Algorithms is proposed to solve the imbalanced learning issue. Unlike Kohonen's original learning rule, the proposed method improves the way to search the winner neuron by stochastic gradient descent on energy function, which can minimally reduce local error in the competitive learning of SOM. Additionally, the advantage of IESOM can improve the classification performance by learning the implicit knowledge from the subset of the limited and underrepresented minority data.

The remainder of this paper is organized as follows. The fundamental mechanisms are introduced in Section 2. Section 3 presents the IESOM algorithm and framework and discusses its advantages on the imbalanced learning problem. Based on this, the performance of the proposed approach is compared with other common methods for imbalanced learning problems in Section 4. We also present more detailed analysis and discussion on the effectiveness of the proposed method. Finally, a conclusion and a brief discussion of future work are provided in Section 5.

2. The fundamental mechanisms

In this section, the fundamental knowledge is introduced to support the proposed method. The detailed information can also facilitate us to comprehensively understand the crucial steps for the framework design.

2.1. Self-organizing map

Traditional SOM [17,32] is a powerful unsupervised learning approach that tries to stimulate synaptic neurons to search the most similar reference vector for the training instances, which can be considered to simply represent the input layer. Generally speaking, the procedure of unsupervised SOM [20,50] is viewed as nonlinear transformation from higher dimensional input space to lower dimensional (one- or two-dimensional) map lattice, involving the local search based on interactively lateral influence of competitive neurons. The vector quantization technique is mentioned in Eq. (1) to search the best matching unit (BMU) by minimizing quantization error. Here we use X to represent dataset, and assume that the vector $x_i \in X \subset R^n$ is the input example and m_j is a reference vector. These two vectors have the identical dimensions in the feature space

$$i(x) = \arg\min_j \| x - m_j \| \tag{1}$$

The topology-preserving map arranges specific geometric structure for the neurons on the feature grid to search the data clusters via competitive learning. The winner neurons are on the centroids of the topological neighborhood [27] of cooperative neurons. The similarities between winner neuron and synaptic neuron are configured by the lateral interactions. A common kernel function for topological neighborhood is Gaussian function $h_{j,i(x)}$, which denotes the degree of interactive connections between two adjacent neurons. In Eq. (2), σ and $d_{j,i(x)}$ represent the radius of the topological neighborhood and the lateral distance between the winner neuron $i(x)$ and synaptic neuron j based on the position information of the feature map, respectively

$$h_{j,i(x)}(t) = \exp\left(-\frac{d_{j,i(x)}^2}{2\sigma^2(t)} \right) \tag{2}$$

$$d_{j,i(x)} = \| r_j - r_{i(x)} \| \tag{3}$$

$$\sigma(t) = \sigma_0 \left(\frac{\sigma_1}{\sigma_0} \right)^{t/N_c} \tag{4}$$

where t represents the training time; r_j is a 2-D position vector of neuron j; N_c denotes the convergence iterations; σ_0 and σ_1 are the initial and terminal neighborhood radii, respectively.

Based on the previous learning stages, the feature lattice can be generated in terms of adjusting each reference neuron weight $m_j(t)$ as follows:

$$m_j(t+1) = m_j(t) + \alpha(t) h_{j,i(x)}(t)(x(t) - m_j(t)) \tag{5}$$

where α is the learning rate and $\alpha(t) = \alpha_0(\alpha_1/\alpha_0)^{t/N_c}$.

There are some key features contributing to its learning capability: the competitive computation between input data and winner neuron, the selected neighborhood function, the number of neurons as well as the learning rate.

2.2. Energy function based self-organizing map

As the most novel form of unsupervised learning in the neural networks, SOMs implement topological preservation and dimensionality reduction by smoothing the clusters regarding to the feature lattice [28]. However, the original SOM algorithm does not optimize the energy function [18,31]. Although SOM involves a trade-off between the quantization accuracy and topological smoothness, no explicit learning rule merges these two properties into an energy function. To modify the original learning rule, the