Optik 125 (2014) 1178–1184

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Human cognitive paradigm and its application in semi-supervised learning

Chun Zhang*, Dongsheng Li, Junan Yang, Aixia Yong

Electronic Engineering Institute, Key Laboratory of Electronic Restriction, 230037 Hefei, China

ARTICLE INFO

Article history:
Received 29 March 2013
Accepted 30 July 2013

Keywords:
Human cognitive paradigm
Semi-supervised learning
Behavioral learning
Tri-training
Co-training
Data editing

ABSTRACT

In many practical data mining applications such as web page classification, unlabeled training examples are readily available but labeled ones are fairly expensive to obtain. Therefore, semi-supervised learning algorithms such as Tri-training have attracted much attention. However, mislabeling the unlabeled data during the learning process is an inevitable problem and harms the performance improvement of the hypothesis. To solve this problem, a novel human cognitive paradigm is constructed for semi-supervised learning in this paper. In detail, based on local distribution of feature space, the majority voting scheme is substituted by an estimation of the probability of sample to belong to a certain class as an efficient strategy for data editing. It considers the form of the underlying probability distribution in the neighborhood of a point to identify and remove the mislabeled data. Validation of the proposed method is performed with extensive experiments. Results demonstrate that compared with Tri-training method, our method can more effectively and stably exploit unlabeled data to enhance the learning performance.

1. Introduction

Semi-supervised learning has become an attractive topic in machine learning and data mining, since the labeled data for supervised learning are expensive to obtain and the large amount of unlabeled data are readily to collect [1]. It exploits unlabeled data in addition to the limited labeled ones to improve the performance [2]. Many semi-supervised classification approaches have been proposed, such as the EM based mixture generative model [3,4], transductive support vector machine (TSVM) [5], and graph based regularization methods [6].

A prominent achievement in this area is the co-training paradigm proposed by Blum and Mitchell [7], which trains two classifiers separately on two different views, i.e. two independent sets of attributes, and uses the predictions of each classifier on unlabeled examples to augment the training set of the other. Then Zhou and Li [8] proposed the Tri-training algorithm which uses three classifiers to perform the co-training. It does not require the constraints on attributes, nor does it require the constraints on special classifiers or the cross-validation. Therefore, Tri-training has more applications [9,10].

However, many researchers have noted a common problem in co-training style algorithms, that is, the performance of semi-supervised learning are usually not stable because the unlabeled examples may often be wrongly labeled during the learning process [3,8]. Moreover, Tri-training may suffer more from this problem. How to effectively identify and filter the mislabeled noise data during the co-training iteration has great significance to enhance the classification performance of co-training style approaches.

Currently, Defense Advanced Research Projects Agency (DARPA) is soliciting innovative research proposals in the area of machine learning for electronic warfare applications and sets up the behavioral learning for adaptive electronic warfare (BLADE) program [11] in 2010. At the same time, more and more research fruits hold the viewpoint that human behavioral learning can effectively improve the performance of machine learning [12–14]. Inspired by these booming trends, this paper constructed human cognition paradigm as a tool to resolve the inevitable mislabeling problem in Tri-training. In detail, we provide a "translation" of relevant terms from machine learning to human cognition paradigm. In human cognitive paradigm, we consider the form of the underlying probability distribution in the neighborhood of a point to identify and remove the mislabeled data. Thus the revised semi-supervised learning algorithm is called HCP-Tri-training (human cognitive paradigm based Tri-training).

The rest of this paper is organized as follows. Section 2 briefly describes the co-training process of Tri-training. Section 3 presents the human cognitive paradigm and its application in data editing. Then in Section 4, the HCP-Tri-training algorithm is presented. Section 5 performs extensive experiments on the proposed HCP-Tri-training method. Finally, we provide some concluding remarks and suggestions for future work in Section 6.

* Corresponding author. Tel.: +86 13965122924.
E-mail address: zhangchuncwins@163.com (C. Zhang).

0030-4026/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.07.136

2. Co-training process of Tri-training

The pseudo code of Tri-training is presented in [8]. During the co-training process of Tri-training, some unlabeled data are firstly labeled by co-labeling, and then these newly labeled data are used to update training set and perform re-training when some sufficient condition is satisfied. The co-labeling and re-training process for each individual classifier is repeated until none of three individual classifiers changes. The co-labeling and re-training in each co-training round are showed as follows.

2.1. Co-labeling

Let **L** and **U** denote the original labeled and unlabeled set respectively, which are drawn independently from the identical underlying data distribution. In Tri-training, three different classifiers, i.e. H_i, H_j, H_k, are initially trained from three bootstraps of **L** respectively. Then the co-labeling iteration is performed as follows: For every unlabeled data x in U, if H_i and H_j agree on labeling it as $H_i(x)$, then x becomes newly labeled one for the third classifier H_k. Thus all newly labeled data from **U** like x are copied into **L** with new labels and forms new candidate training set of H_k. Here, H_i and H_j act as a joint classifier denoted by H_i & H_j. Similarly, the new candidate training sets of H_i and H_j are formed by their corresponding joint classifier.

Because the new candidate training set might be used to refine the individual classifier in the followed re-training step, if the newly labeled data is wrongly labeled by the joint classifier, the third classifier will obtain a new training data with noise label, which is harmful to its refinement. Therefore, Zhou and Li [8] derived a sufficient condition to decide whether the new candidate training set should be used for re-training.

2.2. Sufficient condition for re-training

In Tri-training, the sufficient condition for re-training aims to ensure that the classification accuracy of individual classifier could be improved after it is re-trained by the new training set.

The sufficient condition is derived from the finding of Angluin and Laird [15] on the PAC property of hypothesis learned from noisy training examples. That is, the hypothesis minimizing the disagreement with the sequence of training examples will close to the ground-truth hypothesis with the PAC property, if the size m of noisy training set satisfies:

$$m = \frac{c}{\varepsilon^2(1-2\eta)^2} \tag{1}$$

where c is a constant, ε is the hypothesis worst-case error rate and $\eta(<0.5)$ is the noise rate on training set. This equation is reformed as the following utility function:

$$u = \frac{c}{\varepsilon^2} = m(1-2\eta)^2 \tag{2}$$

Obviously, this utility function indicates $u \propto 1/\varepsilon^2$.

According to Eq. (2), the object of re-training in Tri-training is to ensure the classification error rate ε of hypothesis can be reduced iteratively; meanwhile the size m of new training set for each individual classifier can be iteratively increased.

Let $\mathbf{L}_{i,t}$ and $\mathbf{L}_{i,t-1}$ denote the newly labeled training subset for H_i from **U** by the joint classifier H_j & H_k in the tth and $(t-1)$-th co-training round respectively, where all members of $\mathbf{L}_{i,t}$ will be put back in **U** as unlabeled ones in the tth round. Thus the training set for H_i in the tth and $(t-1)$-th round are $\mathbf{L}\cup\mathbf{L}_{i,t}$ and $\mathbf{L}\cup\mathbf{L}_{i,t-1}$, whose sizes are $\left|\mathbf{L}\cup\mathbf{L}_{i,t}\right| = \left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t}\right|$ and $\left|\mathbf{L}\cup\mathbf{L}_{i,t-1}\right| = \left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t-1}\right|$. Further, let η_L denote the noise rate on the original labeled set **L** and let $\widehat{e}_{i,t}(<0.5)$ denote the error rate upper bound of H_j & H_k on $\mathbf{L}_{i,t}$, then the noise rate on $\mathbf{L}\cup\mathbf{L}_{i,t}$ denoted by:

$$\eta_{i,t} = \frac{\eta_L\left|\mathbf{L}\right| + \widehat{e}_{i,t}\left|\mathbf{L}_{i,t}\right|}{\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t}\right|} \tag{3}$$

And with Eq. (2), the utility of H_1 in tth round denoted by $u_{i,t}$ could be reformed as:

$$u_{i,t} = (\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t}\right|)(1-2\eta_{i,t})^2 = (\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t}\right|) \times \left(1 - 2\frac{\eta_L\left|\mathbf{L}\right| + \widehat{e}_{i,t}\left|\mathbf{L}_{i,t}\right|}{\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t}\right|}\right)^2 \tag{4}$$

Similarly, $u_{i,t-1}$ can be computed as:

$$u_{i,t-1} = (\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t-1}\right|)(1-2\eta_{i,t-1})^2 = (\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t-1}\right|) \times \left(1 - 2\frac{\eta_L\left|\mathbf{L}\right| + \widehat{e}_{i,t-1}\left|\mathbf{L}_{i,t-1}\right|}{\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t-1}\right|}\right)^2 \tag{5}$$

As shown in Eq. (2), since u is in proportion to $1/\varepsilon^2$, it can be derived that if $u_{i,t} > u_{i,t-1}$ then $\varepsilon_{i,t} > \varepsilon_{i,t-1}$, which implies that H_i can be improved through utilizing $\mathbf{L}_{i,t}$ in its training. This condition can be expressed as Eq. (6) by comparing Eqs. (4) and (5):

$$(\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t}\right|)\left(1 - 2\frac{\eta_L\left|\mathbf{L}\right| + \widehat{e}_{i,t}\left|\mathbf{L}_{i,t}\right|}{\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t}\right|}\right)^2 > (\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t-1}\right|) \times \left(1 - 2\frac{\eta_L\left|\mathbf{L}\right| + \widehat{e}_{i,t-1}\left|\mathbf{L}_{i,t-1}\right|}{\left|\mathbf{L}\right| + \left|\mathbf{L}_{i,t-1}\right|}\right)^2 \tag{6}$$

Considering that η_L can be very small and assuming $0 \le \widehat{e}_{i,t}, \widehat{e}_{i,t-1} < 0.5$, then the first term on the left hand of Eq. (6) is bigger than its correspondence on the right hand if $\left|\mathbf{L}_{i,t-1}\right| < \left|\mathbf{L}_{i,t}\right|$, while the second term on the left hand is bigger than its correspondence on the right hand if $\widehat{e}_{i,t}\left|\mathbf{L}_{i,t}\right| < \widehat{e}_{i,t-1}\left|\mathbf{L}_{i,t-1}\right|$. These restrictions can be summarized into the condition shown in Eq. (7), which is used in tri-training to determine when an unlabeled example could be labeled for a classifier.

$$0 < \frac{\widehat{e}_{i,t}}{\widehat{e}_{i,t-1}} < \frac{\left|\mathbf{L}_{i,t-1}\right|}{\left|\mathbf{L}_{i,t}\right|} < 1 \tag{7}$$

Although sufficient condition (7) can partially help compensate the mislabeling problem in Tri-training, this problem is still serious especially when the initial labeled set is very limited. In order to more effectively resolve this problem and improve the stability of classification performance, the human cognition paradigm would be instantiated and equipped into individual classifier.

3. Human semi-supervised learning

Do passive experiences help students learn things, in addition to the explicit instructions received from teacher? Intuitively, the answer appears to be “yes.” Perhaps surprisingly, there is little study on this question. Clearly, passive experiences are nothing more than unlabeled data, and it seems likely that humans exploit such information in ways similar to how semi-supervised learning algorithms in machines do. In this Section, we demonstrate human cognition paradigm and its application in semi-supervised learning.

3.1. Human learning style

In many real world situations, humans are exposed to a combination of labeled data and far more unlabeled data when they