

Online Cost Sensitive Domain Knowledge Adaption

Dhanashree M. Patil

Department of Computer Engineering
JSPM's Imperial College of Engineering and
Research, Wagholi, Pune

R.N. Phursule

Department of Computer Engineering
JSPM's Imperial College of Engineering and
Research, Wagholi, Pune

ABSTRACT

In online learning, classifier is updated as records get available sequentially with the help of time. The misclassifications depends on the class which has been wrongly classified. Above both concepts are largely studied in data mining. In Cost sensitive online learning classification, cost of the misclassification is considered during online learning of the classifier. In domain knowledge adaption, Knowledge (Classifier) from source domain is adapted to target domain considering cost of misclassification. So accuracy of the system is increased. In proposed work cost of the misclassification is considered to adapt the knowledge from source domain to target domain. The proposed system uses the concept of cost sensitive online learning based knowledge transfer. In existing knowledge transfer system i.e. in OTL cost of misclassification is not considered while transferring knowledge from source domain to target domain. But in proposed system cost of misclassification is considered while updating classifier on sequentially coming data. To evaluate the approach spam base dataset is used and the implemented system gives better results.

Keywords

Cost sensitive classification, online learning, Domain Adaption, Naive Bayes classifier.

1. INTRODUCTION

Classification is the vital task of the data mining [8]. Machine learning, neural network, pattern recognition also uses classification approach as subpart of their theory. In classification main aim is to label a test instance (Vector of attributes) to one of the finite class or category. While classification task most of the classifier techniques focuses to reduce the number of misclassification. Specifically, traditional understanding is that classifier is better in performance if its error rate is less. This understanding assumes that misclassification cost of the all class is same but in much application cost of misclassification varies according to the class. Significant work has done on how to consider the cost of the particular class while learning and classifying i.e. how to make classification learning algorithms cost sensitive. To solve above problems in cost sensitive classification two more meaningful parameters are proposed in the literature survey [1]. Weighted sum of the sensitivity and specificity [1] and weighted misclassification cost [1]. In recent years, many efforts are devoted to make batch classifiers cost sensitive. However these batch classification algorithms are poor in scalability and efficiency. In the area of the data mining and machine learning there is need to develop efficient and scalable algorithms for mining of massive rapidly growing data. Online learning and cost sensitive classifications are broadly studied separately in the literature and very less research in Cost sensitive online learning. In [1] new framework Cost sensitive online learning is proposed for

online learning classification task which considers cost of the misclassification by optimizing a predefined cost sensitive measures and gives theoretical description for the same. Framework consist two Cost sensitive online classification using gradient descent (CSOGD) algorithms which aims to minimize weighted misclassification cost and maximize the weighted sum. Experiments in [1] shows that CSOGD is best for cost sensitive online learning. Online learning is one of the directions to develop such algorithms. Main Aim of the online learning algorithm is incrementally learn some prediction model to make correct prediction on the sequence of the examples. Many online learning algorithms are available in literature such as Passive Aggressive Online learning [5], popular Perceptron algorithm [6] etc. Most of the current online learning algorithms are not suitable for the cost sensitive classification. In current online learning algorithms suffer loss function does not considered at the time of updating classifier i.e. current online learning algorithms are cost insensitive. In data mining, there is need to develop robust classifiers whose performance does not affect by on the mismatch distributions. To tackle this challenge online transfer learning or domain adaption was introduced [9]. Goal of the online transfer learning is to transfer knowledge from some source site to some target site with online learning. Key motivation of the online transfer learning utilizes knowledge of the source domain at target domain while online learning. Work in [9] both cases 1) Class distribution of the source class and target is different. In this paper, we propose an algorithm for online knowledge transfer which considers the cost of the misclassification at the time knowledge transfer at target site. The proposed system applies OCSKT algorithms for online knowledge transfer.

2. PROPOSED SYSTEM

2.1. Motivation

Despite being studied extensively, most existing online learning techniques are unsuitable for cost sensitive classification tasks, an important problem for data mining which has to address varied misclassification costs. In current system there are very few algorithms which consider cost of the misclassification in knowledge transfer so there is open scope to work on this concept.

2.2. Proposed System

The proposed system have best algorithm OCSKT for cost sensitive online learning classification and there is need of knowledge transfer algorithm which will consider cost of the misclassification. And the approach used in OCSKT is "how to consider the cost of the misclassification" for online knowledge transfer. It optimizes the cost sensitive measures i.e. weighted sum of the sensitivity and specificity, and minimizes the weighted misclassification cost for cost sensitive knowledge transfer. In each learning round, learner

receives the instance for prediction and learner guesses its class label (PL). Actual label (AL) of that instance is fetched and compared with predicted label. If predicted class and actual class are not same learner suffers loss. Loss is calculated with particular suffer loss function, if suffered loss is greater than zero then prediction model is updated. In the case of Binary classifier, prediction results in following four types

- 1) Prediction is True positive (TP) if AL=PL =+1
- 2) False positive (FP) if PL=+1 and AL= -1
- 3) True negative (TN) if PL=AL= -1
- 4) False negative if PL=-1 and AL = +1

Mp is number of false negatives and Mn is number of false positives. Tp denotes number of positive examples and Tn denotes the number of the negative examples. Sensitivity is defined as the ratio between the number of true positives Tp - Mp and the number of positive examples.

$$\text{Sensitivity} = \frac{T_p - M_p}{T_p}$$

Specificity is defined as the ratio between the number of true negatives Tn-Mn and the number of negative examples.

$$\text{Specificity} = \frac{T_n - M_n}{T_n}$$

Measures to calculate the performance of the cost sensitive classification are sum and cost.

$$\text{Sum} = N_p * \text{Sensitivity} + N_n * \text{Specificity}$$

Where $N_p + N_n = 1$ and both must be greater than or less than zero and one respectively are trade off parameter between specificity and sensitivity. Generally higher is the sum higher is the accuracy of the classifier. There are two different loss functions in the proposed system.

$$L1(w; (x, y)) = \max(0, (\rho * I(y = 1) + I(y = -1)))y(w: x) \dots \text{Equation(1)}$$

$$L2(w; (x, y)) = (\rho * I(y = 1) + I(y = -1)) * \max(0, 1 - y(w: x)) \dots \text{Equation(2)}$$

Both loss functions considers cost of the misclassification using is where $\rho = N_p.T_n / N_n.T_p$ for the maximization of the weighted sum and $\rho = C_p/C_n$ for minimization of misclassification cost. We will use one of these loss functions to calculate suffer loss in online cost sensitive knowledge transfer and same loss function in update function to update the classifier.

2.3. System Architecture

Figure 1 shows the system architecture of the cost sensitive online knowledge transfer.

Training dataset 1 : This is the labeled training dataset at site1 with set of features $F = (f_1, \dots, f_n)$ with class distribution D. Classifier learning algorithm takes batch of Training dataset D to build classifier. Classifier 1 is build by training dataset 1 TD1 and CLA 1 at site 1

Training dataset 2 : This is the labeled training dataset at site 2 with set of features $F = (f_1, \dots, f_n)$ with class distribution D.

Domain Adaption: In Domain Adaption step, classifier 1 from side 1 is utilized. To predict the class of the instance X_t

at t round the combination of classifier c1 and c2 is used. To combine classifier at round t it will use w_{1t} and w_{2t} weight parameters for respective classifiers used. Following ensemble function is used to predict label (PL) of the instance.

$$PL = \text{sign}(w_{1t} * \sigma(c_1(X_t)) + w_{2t} * \sigma(c_2(X_t)) / 0.5)$$

.....Equation(3)

At the beginning $w_{1,1} = w_{2,1} = 1/2$ $\sigma(x)$ is normalization functions. Then suffer loss is calculated using equation 1. If suffer loss is greater than one then classifier 2 is updated.

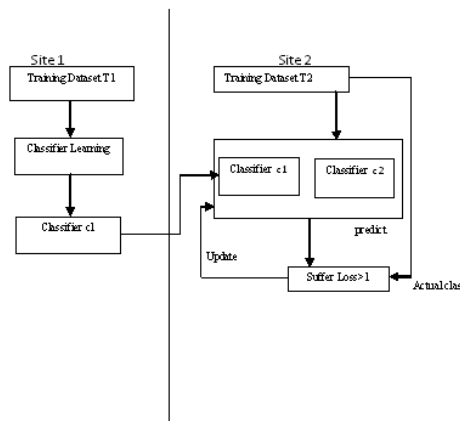


Figure 1: System Architecture

3. RESULTS AND DISCUSSION

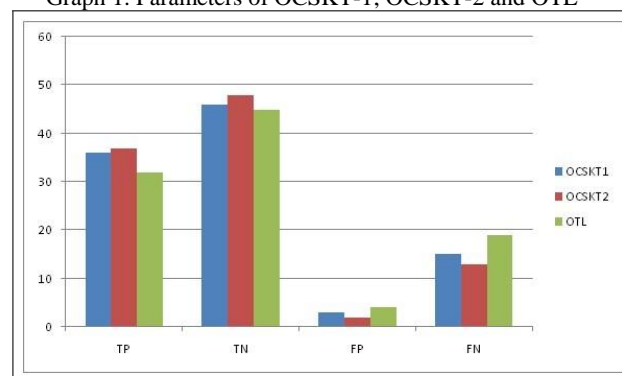
In this proposed system initially for classifier 1 at side 1, 1000 records are stored .And for Classifier 2 at another side for initializing that classifier 10 records have used. After combining both classifiers total 2000 records have used .And for testing 100 records have used.

Table 1. Comparison Parameters of OCSKT-1, OCSKT-2 and OTL

Attributes	OCSKT-1	OCSKT-2	OTL
TP	35	37	32
TN	45	47	42
FP	4	3	7
FN	15	12	19

Above table 1 shows comparison parameter values for OCSKT-1,OCSKT-2 and OTL algorithm.

Graph 1. Parameters of OCSKT-1, OCSKT-2 and OTL



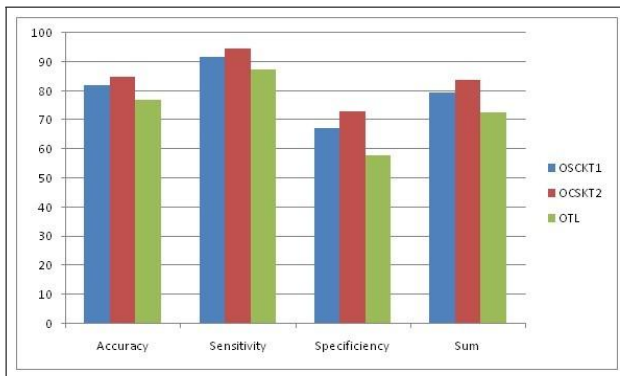
Above graph 1 Shows the True Positive , True negative, False positive, False negative values for OCSKT-1,OCSKT-2 and OTL algorithm.

Table 2. Comparison Of Accuracy, Sensitivity, Specificity and Sum of OCSKT-1,OCSKT-2 and OTL.

Parameters	OCSKT-1	OCSKT-2	OTL
Accuracy	82	84	76
Sensitivity	92	92	86
Specificity	68	72	58
Sum	78	82	72

Above table 2 shows performance comparison for OCSKT-1, OCSKT-2 and OTL algorithm.

Graph 2. Performance comparison of OCSKT-1, OCSKT-2 and OTL



Above graph 2 shows the performance comparison of OCSKT-1, OCSKT-2, OTL algorithm.

4. CONCLUSION

In this paper, the proposed system having two variations of the online cost sensitive knowledge transfer algorithm. This algorithm transfers the knowledge from source domain to target domain by considering the cost of misclassification. And these two variations of algorithms are compared with existing OTL algorithm. This proposed algorithm gives better performance than existing OTL algorithm. This system is used where dataset at target domain is very less or it is very difficult or expensive to get the data. Proposed system considers the case where the target site attributes list must be subset of the source site attribute list. In future work this

system will work on the case where target site attribute list is not subset of source site attribute list.

5. ACKNOWLEDGMENT

We would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are also thankful to reviewer for their valuable suggestions. We also thank the college authorities for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

6. REFERENCES

- [1] Cost Sensitive Online Classification, Jialei Wang, Peilin Zhao, and Steven C.H. Hoi, Member, IEEE, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 10, OCTOBER 2014
- [2] Meta cost: A general method for making classifiers cost sensitive.
- [3] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, fscore and ROC: A family of discriminate measures for performance evaluation," in Proc. AAI , Hobart, TAS, Australia, 2006, pp. 1015-1021
- [4] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in Proc. 15th ECML, Pisa, Italy, 2004, pp. 3950.
- [5] K. Crammer, O. Dekel, J. Keshet, S. Shalev Shwartz, and Y. Singer, "Online passive-aggressive algorithms," J. Mach. Learn. Res. , vol. 7, pp. 551-585, Mar. 2006.
- [6] Y. Y. Freund and R. E. Schapire, "Large margin classification using the perceptron algorithm," Mach. Learn., vol. 37, no. 3, pp. 277-296, 1999.
- [7] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," Psych. Rev. , vol. 65 no. 6, pp. 386-408, 1958.
- [8] From data mining to knowledge discovery: an overview, Usma Fayyad, 1996
- [9] OTL: A Framework of Online Transfer Learning, Peilin Zhao zha0106ntu.edu.sg, Steven C.H. Hoi 2010
- [10] Y. Li and P. M. Long, "The relaxed online maximum margin algorithm," in Proc. NIPS , 1999, pp. 498504
- [11] Y. Li, H. Zaragoza, R. Herbrich, J. Shawe Taylor, and J. S. Kandola, "The perceptron algorithm with uneven margins," in Proc. 19th ICML , San Francisco, CA, USA, 2002, pp. 379-386.
- [12] "An algorithm for transfer learning in heterogeneous environment" Argyriou, Andreas 2008