ISSN: 2710-0588

Transfer Learning Algorithm Combined with Hierarchy Correlation of Data

Hua Pang[1], Long Wang[2, *]

[1] College of Mathematics and Systems Science, Shenyang Normal University, Shenyang 110034, China

[2] College of Information, Liaoning University, Shenyang 110036, China

Abstract

An improved transfer learning algorithm is proposed for data with hierarchical correlation which is introduced in the classical transfer learning algorithm TrAdaBoost. In the initial weight assignment phase, the hierarchical relationship between data is considered, and weight restriction are added. The experimental results show that the correct, precision and recall rate of the recommendation method based on the improved algorithm reach more than 85% on the learning resource sample data with hierarchical correlation.

Keywords

Transfer learning; Hierarchical Correlation; Weight Distribution; Weight Adjustment.

1. Introduction

In recent years, with the promotion of global informatization, the amount of data is increasing day and night. In modern society, data mining and application have penetrated into every industry field. The arrival of large data has brought convenience to people and brought new challenge simultaneously, More and more social circles have shifted their focus to data mining and machine learning. with people's increasing demands for artificial intelligence. Machine learning algorithms have been used in various fields of social development and playing a more important role. With the continuous expansion of the application field, traditional machine learning methods have been unable to meet a variety of new learning tasks, resulting in a variety of new machine learning methods. transfer learning is one of them. At present, the study of transfer learning has received the attention of more and more scholars, it become an important topic and research hotspot of machine learning.

Traditional machine learning methods require that training data and test data are in the same distribution, and that there are enough training samples available, that is, a large amount of mark data is needed to train high-precision learners. For the new test data, it is not necessarily the same as the target data. The model trained by the original training data will lose its function, and the model must be relabeled to retrain a new model, which requires an expensive price. Therefore, it becomes very meaningful to study how to use those original data. transfer learning is a new machine learning method to solve this problem.

Transfer learning refers to applying the knowledge learned in the original task to new tasks to help the learning of new tasks. For example, learning how to learn the accordion on the keyboard would be easier. Learning the Chinese language course will also help in learning the history course. The basic idea of transfer learning is [1]: "Transfer" useful knowledge gained from related fields into new areas of learning tasks. Among them, the related field is called the source field, and the new field is called the target field.

ISSN: 2710-0588

2. Related Works

Currently, researchers have proposed many transfer learning algorithms [2-5]. From the perspective of data processing, transfer learning mainly includes case-based transfer learning and feature representation-based transfer learning. In the case where the source domain and target domain are distributed differently but the content is similar, the transfer effect of the instance-based transfer learning algorithm is more obvious. Dai et al. [6] extended the classical Boosting ensemble learning algorithm AdaBoost and proposed the TrAdaBoost algorithm which has transfer learning ability. The basic idea of TrAdaBoost is to mix the source domain with the target domain for training, and adjust the weights of the samples that were misclassified after each iteration of training. If the misclassified sample belongs to the target domain, the sample is considered to be more important, so the weight of the sample is increased, and the influence of the sample in the next iterative training is improved; if the sample that is incorrectly classified belongs to the source domain, the sample is considered to be not important, so reduce the weight of the sample and reduce the impact of the sample on the next iterative training.

TrAdaBoost is a classic case-based transfer learning algorithm. Boosting method is used in the algorithm to continuously adjust the weights of training samples through iterative methods to find the most useful sample data for the classification model, thereby improving the accuracy of the classification model. The specific flow of TrAdaBoost algorithm is as follows:

Step1: set D = Ds ∪ Dt

One source training sample set:

Ds = {(d1, C(d1)), (d2, C(d2)), ……(dm,C(dm))},

Target domain training sample set :

Dt= {(dm+1, C(dm+1)), (dm+2,C(dm+2)), ……(dn,C(dn))},

C(di) ϵ{0,1}, Indicates the category to which it belongs.

Step2:Set the algorithm iteration number to R(R<m), Initialize iteration count r = 0.

Step3:The sample weight is initialized and the sample initial weight is:

$$w_i^0 = \begin{cases} 1/m, & 1 \le i \le m \\ 1/n-m, & m+1 \le i \le n \end{cases} \tag{1}$$

Step4:r =r+1,

Step5:The sample weights are normalized and the sample weights are adjusted to:

$$w_i^r = \frac{w_i^{r-1}}{\sum_{i=1}^{n} w_i^{r-1}} \tag{2}$$

Step6:Call Weak Learning algorithm and train to get the classifier:

Step7:Calculate the error rate, the error rate of the classifier:

$$\varepsilon^r = \sum_{i=m+1}^{n} \frac{w_i^r \times \left| h^r(d_i) - c(d_i) \right|}{\sum_{i=m+1}^{n} w_i^r} \tag{3}$$

ISSN: 2710-0588

Set:

$$\varepsilon^r = \begin{cases} \varepsilon^r & \varepsilon^r < \frac{1}{2} \\ \frac{1}{2} & \varepsilon^r \geq \frac{1}{2} \end{cases} \tag{4}$$

Step8:Adjust sample weights based on error rate:

$$w_i^r = w_i^r \times \begin{cases} 1, & h^r(d_i) = c(d_i) \\ \beta_s, & h^r(d_i) \neq c(d_i) \text{ and } 1 \leq i \leq m \\ (\beta_t^r)^{-1}, & h^r(d_i) \neq c(d_i) \text{ and } m+1 \leq i \leq n \end{cases} \tag{5}$$

Among then:

$$\beta_s = \frac{1}{1 + \sqrt{2 \times \ln \frac{m}{R}}} \tag{6}$$

$$\beta_t^r = \frac{\varepsilon^r}{1 - \varepsilon^r} \tag{7}$$

Step9:If iteration count , go to step4,Otherwise get the final classifier:

$$h^R(d) = \begin{cases} 1, & \sum_{r=\frac{R}{2}}^{R} \ln(\frac{1}{\beta_t^r}) \times h^r(d) \geq \frac{1}{2} \times \sum_{r=\frac{R}{2}}^{R} \ln(\frac{1}{\beta_t^r}) \\ 0, & \text{other} \end{cases} \tag{8}$$

TrAdaBoost is a simple and effective instance-based transfer learning algorithm. However, in practical applications, this method also has some areas for improvement.

3. Correlation-based Transfer Learning

3.1. Data Hierarchy

The hierarchically related data processed by this paper can be represented by the tree structure shown in Figure 1 The root node A represents the domain of data, the internal node BCDEFG represents the hierarchical structure of data, and the leaf node represents sample data.

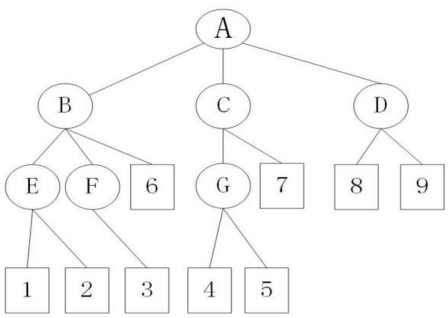

Figure 1. Data hierarchy

ISSN: 2710-0588

3.2. Initial Weight Distribution

The TrAdaboost algorithm distributes the initial weights equally to all sample data. If there is a hierarchical correlation between sample data, the average initial weights are obviously not reasonable enough. If there is a hierarchical correlation between the sample data, then the sample represented by the leaf node with smaller distance from the root node is more important, and the sample represented by the leaf node with larger distance from the root node is less important. Therefore, the definition of the correlation between the sample and the domain ρ_i is as follows:

$$\rho_i = \frac{\sum_{j=1}^{n} G_j}{n * G_i + \sum_{j=1}^{n} G_j} \tag{9}$$

The correlation measure is added to the sample weight initialization to reflect the sample level correlation.

$$w_i^0 = \begin{cases} (1/m) \times \rho_i, & 1 \le i \le m \\ (1/n-m) \times \rho_i, & m+1 \le i \le n \end{cases} \tag{10}$$

3.3. Weight Adjustment Strategy

In the weight adjustment phase, the TrAdaboost algorithm only adjusts the misclassified samples and does not consider the correlation between the samples. For the sample data with hierarchical correlation, the correlation function f(L) is introduced in the sample weight adjustment stage, and the weights of multiple data samples related to the misclassified samples are adjusted within a certain range, which will effectively improve the efficiency of the algorithm. The smaller the distance Lki between the current sample node i and the wrong-divided sample node k, the greater the correlation between the samples; the larger the distance Lki between the current sample node i and the wrong-divided sample node k, the correlation F between samples. The smaller. Function f(L) has the following properties:

f(L) ϵ{0,1}.

f(L)monotonically decreasing, where xi ϵ [0,+∞].

f (0)=1.

The progressive deceleration of f(L) gradually becomes slower and the curve becomes smoother and smoother.

According to this definition function:

$$f(L) = \begin{cases} 1, L = 0 \\ 1 - \ln(1 + \frac{1}{L})^L, L \neq 0 \end{cases} \tag{11}$$

At the same time, in the sample weight adjustment process, if the sample weight is too large or too small, the classification ability of the classifier will be reduced. Therefore, the weight restriction is added to the weight adjustment. If the weight is greater than the upper threshold u, the weight is adjusted to u if the weight is less than If the lower threshold l is set, the weight is adjusted to l. In summary, the eighth step of the TrAdBoost algorithm is modified to the

ISSN: 2710-0588

weights of all samples whose error rate is adjusted and the distance of the wrong sample is smaller than the threshold λ :

$$w_k^{r+1} = w_k^r \times \begin{cases} 1, & h^r(d_i) = c(d_i) \\ 1-(1-\beta_s) \times f(L_{ki}), & h^r(d_i) \neq c(d_i) \text{ and } 1 \leq i \leq m \\ 1+\left((\beta_t^r)^{-1}-1\right) \times f(L_{ki}), & h^r(d_i) \neq c(d_i) \text{ and } m+1 \leq i \leq n \end{cases} \tag{12}$$

$$w_k^{r+1} = \begin{cases} u, & w_k^{r+1} > u \\ l, & w_k^{r+1} < l \\ w_k^{r+1}, & \text{other} \end{cases} \tag{13}$$

4. Experimental Analysis

This article applies the transfer learning algorithm to the teaching resource recommendation service in the network teaching platform.

On the above Data set, training tests were performed using the TrAdaBoost algorithm, respectively, and compared with the transfer learning algorithm proposed in this paper, the results shown in Table1 were obtained.

Table 1. ALGORITHM Comparison

Data set	TrAdaBoost			This Algorithm		
	CorrectRate	Precision	Recall	CorrectRate	Precision	Recall
0	81.7%	81.3%	81.1%	85.20%	85.70%	86.10%
1	81.4%	81.0%	81.2%	85.30%	86.10%	85.40%
2	82.0%	80.7%	81.5%	86.10%	86.00%	86.20%
3	80.9%	82.2%	81.9%	86.20%	86.30%	85.20%
4	81.3%	81.9%	82.3%	85.50%	85.90%	85.70%
5	81.1%	81.5%	81.1%	85.60%	86.10%	85.10%
6	82.0%	81.4%	80.5%	86.20%	86.20%	85.10%
7	80.7%	81.2%	81.9%	85.10%	85.40%	88.20%
8	81.2%	81.3%	82.2%	86.40%	85.30%	85.20%
9	81.6%	82.0%	80.8%	86.10%	85.90%	85.10%
Mean	81.4%	81.5%	81.5%	85.8%	85.9%	85.7%

5. Conclusion

For the sample data with hierarchical correlation, this paper proposes a transfer learning algorithm that combines the hierarchical correlation of data. This algorithm improves the weight initialization and weight adjustment methods of the classic transfer learning algorithm TrAdaBoost. The experimental results show that the algorithm can effectively achieve the purpose of transfer learning. For the data with hierarchical correlation, the algorithm has more advantages than the TrAdaBoost algorithm. High accuracy and efficiency.

References

[1] PAN S J, YANG QIANG. A Survey on Transfer Learning[J]. IEEE Trans on Knowledge and Data Engineering, 2010, 22(10): 1345 - 1359.

[2] PARDOE D, STONE P. Boosting for regression transfer[A]. Piscataway:IEEEInc Press,2010:863-870.

ISSN: 2710-0588

[3] S.J.Pan, V.W.Zheng, Q.Yang, and D.H.Hu. Transfer learning for wifi-based indoor localization[C], the Proceedings of the Workshop on Transfer Learning for Complex Task of the 23rd AAAI Conference on Artificial Intelligence, Chicago, Illinois, USA, July 2008.

[4] J.Ramon, K.Driessens, and T.Croonenborghs. Transfer learning in reinforcement learning problems through partial policy recycling[C], the Proceedings of the 18th European conference on Machine Learning, Berlin, Heidelberg: Springer-Verlag, 2007, pp. 699–707.

[5] M.E.Taylor and P.Stone. Cross-domain transfer for reinforcement learning[C], the Proceedings of the 24th international conference on Machine learning. New York, USA: ACM, 2007, pp. 879–886.

[6] DAI WENYUAN, YANG Q, XUE G. Boosting for transfer learning[A]. New York:Academic Press,Inc, 2007:193-200.