

Improving SCL Model for Sentiment-Transfer Learning

Songbo Tan

Institute of Computing Technology
Beijing, China

tansongbo@software.ict.ac.cn

Xueqi Cheng

Institute of Computing Technology
Beijing, China

cxq@ict.ac.cn

ABSTRACT

In recent years, Structural Correspondence Learning (SCL) is becoming one of the most promising techniques for sentiment-transfer learning. However, SCL model treats each feature as well as each instance by an equivalent-weight strategy. To address the two issues effectively, we proposed a weighted SCL model (W-SCL), which weights the features as well as the instances. More specifically, W-SCL assigns a smaller weight to high-frequency domain-specific (HFDS) features and assigns a larger weight to instances with the same label as the involved pivot feature. The experimental results indicate that proposed W-SCL model could overcome the adverse influence of HFDS features, and leverage knowledge from labels of instances and pivot features.

1 Introduction

In the community of sentiment analysis (Turney 2002; Pang et al., 2002; Tang et al., 2009), transferring a sentiment classifier from one source domain to another target domain is still far from a trivial work, because sentiment expression often behaves with strong domain-specific nature.

Up to this time, many researchers have proposed techniques to address this problem, such as classifiers adaptation, generalizable features detection and so on (DaumeIII et al., 2006; Jiang et al., 2007; Tan et al., 2007; Tan et al., 2008; Tan et al., 2009). Among these techniques, SCL (Structural Correspondence Learning) (Blitzer et al., 2006) is regarded as a promising method to tackle transfer-learning problem. The main idea behind SCL model is to identify correspondences among features from different domains by modeling their correlations with pivot features (or generalizable features). Pivot features behave similarly in both domains. If non-pivot features from different domains are correlated with many of the same pivot features, then we assume them

to be corresponded with each other, and treat them similarly when training a sentiment classifier.

However, SCL model treats each feature as well as each instance by an equivalent-weight strategy. From the perspective of feature, this strategy fails to overcome the adverse influence of high-frequency domain-specific (HFDS) features. For example, the words “stock” or “market” occurs frequently in most of stock reviews, so these non-sentiment features tend to have a strong correspondence with pivot features. As a result, the representative ability of the other sentiment features will inevitably be weakened to some degree.

To address this issue, we proposed Frequently Exclusively-occurring Entropy (FEE) to pick out HFDS features, and proposed a feature-weighted SCL model (FW-SCL) to adjust the influence of HFDS features in building correspondence. The main idea of FW-SCL is to assign a smaller weight to HFDS features so that the adverse influence of HFDS features can be decreased.

From the other perspective, the equivalent-weight strategy of SCL model ignores the labels (“positive” or “negative”) of labeled instances. Obviously, this is not a good idea. In fact, positive pivot features tend to occur in positive instances, so the correlations built on positive instances are more reliable than that built on negative instances; and vice versa. Consequently, utilization of labels of instances and pivot features can decrease the adverse influence of some co-occurrences, such as co-occurrences involved with positive pivot features and negative instances, or involved with negative pivot features and positive instances.

In order to take into account the labels of labeled instances, we proposed an instance-weighted SCL model (IW-SCL), which assigns a larger weight to instances with the same label as the involved pivot feature. In this time, we obtain a combined model: feature-weighted and instance-weighted SCL model (FWIW-SCL). For the sake

of convenience, we simplify “FWIW-SCL” as “W-SCL” in the rest of this paper.

2 Structural Correspondence Learning

In the section, we provide the detailed procedures for SCL model.

First we need to pick out pivot features. Pivot features occur frequently in both the source and the target domain. In the community of sentiment analysis, generalizable sentiment words are good candidates for pivot features, such as “good” and “excellent”. In the rest of this paper, we use K to stand for the number of pivot features.

Second, we need to compute the pivot predictors (or mapping vectors) using selected pivot features. The pivot predictors are the key job, because they directly decide the performance of SCL. For each pivot feature k , we use a loss function L_k ,

$$L_k = \sum_i (p_k(x_i)w^T x_i - 1) + \lambda \|w\|^2 \quad (1)$$

where the function $p_k(x_i)$ indicates whether the pivot feature k occurs in the instance x_i ,

$$p_k(x_i) = \begin{cases} 1 & \text{if } x_{ik} > 0 \\ -1 & \text{otherwise} \end{cases},$$

where the weight vector w encodes the correspondence of the non-pivot features with the pivot feature k (Blitzer et al., 2006).

Finally we use the augmented space $[x^T, x^T W]^T$ to train the classifier on the source labeled data and predict the examples on the target domain, where $W=[w_1, w_2, \dots, w_k]$.

3 Feature-Weighted SCL Model

3.1 Measure to pick out HFDS features

In order to pick out HFDS features, we proposed Frequently Exclusively-occurring Entropy (FEE). Our measure includes two criteria: occur in one domain as frequently as possible, while occur on another domain as rarely as possible. To satisfy this requirement, we proposed the following formula:

$$f_w = \log(\max(P_o(w), P_n(w))) + \log\left(\frac{\max(P_o(w), P_n(w))}{\min(P_o(w), P_n(w))}\right) \quad (2)$$

where $P_o(w)$ and $P_n(w)$ indicate the probability of word w in the source domain and the target domain respectively:

$$P_o(w) = \frac{N_o(w) + \alpha}{N_o + 2 \cdot \alpha} \quad (3)$$

$$P_n(w) = \frac{N_n(w) + \alpha}{N_n + 2 \cdot \alpha} \quad (4)$$

where $N_o(w)$ and $N_n(w)$ is the number of examples with word w in the source domain and the target domain respectively; N_o and N_n is the number of examples in the source domain and the target domain respectively. In order to overcome overflow, we set $\alpha=0.0001$ in our experiment reported in section 5.

To better understand this measure, let’s take a simple example (see Table 1). Given a source dataset with 1000 documents and a target dataset with 1000 documents, 12 candidate features, and a task to pick out 2 HFDS features. According to our understanding, the best choice is to pick out w_4 and w_8 . According to formula (2), fortunately, we successfully pick out w_4 , and w_8 . This simple example validates the effectiveness of proposed FEE formula.

Table 1: A simple example for FEE

Words	$N_o(w)$	$N_n(w)$	FEE	
			Score	Rank
w_1	100	100	-2.3025	6
w_2	100	90	-2.1971	4
w_3	100	45	-1.5040	3
w_4	100	4	0.9163	1
w_5	50	50	-2.9956	8
w_6	50	45	-2.8903	7
w_7	50	23	-2.2192	5
w_8	50	2	0.2231	2
w_9	4	4	-5.5214	11
w_{10}	4	3	-5.2337	10
w_{11}	4	2	-4.8283	9
w_{12}	1	1	-6.9077	12

3.2 Feature-Weighted SCL model

To adjust the influence of HFDS features in building correspondence, we proposed feature-weighted SCL model (FW-SCL),

$$L_k = \sum_i (p_k(x_i) \sum_l \delta_l w_l x_{il} - 1) + \lambda \|w\|^2 \quad (5)$$

where the function $p_k(x_i)$ indicates whether the pivot feature k occurs in the instance x_i ;

$$p_k(x_i) = \begin{cases} 1 & \text{if } x_{ik} > 0 \\ -1 & \text{otherwise} \end{cases},$$

and δ_l is the parameter to control the weight of the HFDS feature l ,

$$\delta_l = \begin{cases} \eta & \text{if } l \in Z_{HFDS} \\ 1 & \text{otherwise} \end{cases}$$

where Z_{HFDS} indicates the HFDS feature set and η is located in the range $[0,1]$. When “ $\eta=0$ ”, it indicates that no HFDS features are used to build the correspondence vectors; while “ $\eta=1$ ” indicates that all features are equally used to build the correspondence vectors, that is to say, proposed FW-SCL algorithm is simplified as traditional SCL algorithm. Consequently, proposed FW-SCL algorithm could be regarded as a generalized version of traditional SCL algorithm.

4 Instance-Weighted SCL Model

The traditional SCL model does not take into account the labels (“positive” or “negative”) of instances on the source domain and pivot features. Although the labels of pivot features are not given at first, it is very easy to obtain these labels because the number of pivot features is typically very small.

Obviously, positive pivot features tend to occur in positive instances, so the correlations built on positive instances are more reliable than the correlations built on negative instances; and vice versa. As a result, the ideal choice is to assign a larger weight to the instances with the same label as the involved pivot feature, while assign a smaller weight to the instances with the different label as the involved pivot feature. This strategy can make correlations more reliable. This is the key idea of instance-weighted SCL model (IW-SCL). Combining the idea of feature-weighted SCL model (FW-SCL), we obtain the feature-weighted and instance-weighted SCL model (FWIW-SCL),

$$L_k = \gamma \cdot \sum \rho(\psi(k), \psi(x_i)) (p_k(x_i) \sum_l \delta_l w_l x_{il} - 1) + \lambda \|w\|^2 + (1-\gamma) \cdot \sum (1 - \rho(\psi(k), \psi(x_j))) (p_k(x_j) \sum_l \delta_l w_l x_{jl} - 1) \quad (6)$$

where γ is the instance weight and the function $p_k(x_i)$ indicates whether the pivot feature k occurs in the instance x_i ;

$$p_k(x_i) = \begin{cases} 1 & \text{if } x_{ik} > 0 \\ -1 & \text{otherwise} \end{cases}$$

and δ_l is the parameter to control the weight of the HFDS feature l ,

$$\delta_l = \begin{cases} \eta & \text{if } l \in Z_{HFDS} \\ 1 & \text{otherwise} \end{cases},$$

where Z_{HFDS} indicates the HFDS feature set and η is located in the range $[0,1]$.

In equation (6), the function $\rho(z,y)$ indicates whether the two variables z and y have the same non-zero value,

$$\rho(z,y) = \begin{cases} 1 & \text{if } z = y \text{ and } z \neq 0; \\ 0 & \text{otherwise} \end{cases};$$

and the function $\psi(z)$ is a hinge function, whose variables are either pivot features or instances,

$$\psi(z) = \begin{cases} 1 & \text{if } z \text{ has a positive label} \\ 0 & \text{unknown} \\ -1 & \text{if } z \text{ has a negative label} \end{cases}.$$

For the sake of convenience, we simplify “FWIW-SCL” as “W-SCL”.

5 Experimental Results

5.1 Datasets

We collected three Chinese domain-specific datasets: Education Reviews (Edu, from <http://blog.sohu.com/learning/>), Stock Reviews (Sto, from <http://blog.sohu.com/stock/>) and Computer Reviews (Comp, from <http://detail.zol.com.cn/>). All of these datasets are annotated by three linguists. We use ICTCLAS (a Chinese text POS tool, <http://ictclas.org/>) to parse Chinese words.

The dataset Edu includes 1,012 negative reviews and 254 positive reviews. The average size of reviews is about 600 words. **The dataset Sto** consists of 683 negative reviews and 364 positive reviews. The average length of reviews is about 460 terms. **The dataset Comp** contains 390 negative reviews and 544 positive reviews. The average length of reviews is about 120 words.

5.2 Comparison Methods

In our experiments, we run one supervised baseline, i.e., Naïve Bayes (NB), which only uses one source-domain labeled data as training data.

For transfer-learning baseline, we implement traditional SCL model (T-SCL) (Blitzer et al., 2006). Like TSVM, it makes use of the source-domain labeled data as well as the target-domain unlabeled data.

5.3 Does proposed method work?

To conduct our experiments, we use source-domain data as unlabeled set or labeled training set, and use target-domain data as unlabeled set or testing set. Note that we use 100 manual-annotated pivot features for T-SCL, FW-SCL and W-SCL in the following experiments. We select

pivot features use three criteria: a) is a sentiment word; b) occurs frequently in both domains; c) has similar occurring probability. For T-SCL, FW-SCL and W-SCL, we use prototype classifier (Sebastiani, 2002) to train the final model.

Table 2 shows the results of experiments comparing proposed method with supervised learning, transductive learning and T-SCL. For FW-SCL, the Z_{HFDS} is set to 200 and η is set to 0.1; For W-SCL, the Z_{HFDS} is set to 200, η is set to 0.1, and γ is set to 0.9.

As expected, proposed method FW-SCL does indeed provide much better performance than supervised baselines, TSVM and T-SCL model. For example, the average accuracy of FW-SCL beats supervised baselines by about 12 percents, beats TSVM by about 11 percents and beats T-SCL by about 10 percents. This result indicates that proposed FW-SCL model could overcome the shortcomings of HFDS features in building correspondence vectors.

More surprisingly, instance-weighting strategy can further boost the performance of FW-SCL by about 4 percents. This result indicates that the labels of instances and pivot features are very useful in building the correlation vectors. This result also verifies our analysis in section 4: positive pivot features tend to occur in positive instances, so the correlations built on positive instances are more reliable than the correlations built on negative instances, and vice versa.

Table 2: Accuracy of different methods

	NB	T-SCL	FW-SCL	W-SCL
Edu->Sto	0.6704	0.7965	0.7917	0.8108
Edu->Comp	0.5085	0.8019	0.8993	0.9025
Sto->Edu	0.6824	0.7712	0.9072	0.9368
Sto->Comp	0.5053	0.8126	0.8126	0.8693
Comp->Sto	0.6580	0.6523	0.7010	0.7717
Comp->Edu	0.6114	0.5976	0.9112	0.9408
Average	0.6060	0.7387	0.8372	0.8720

Although SCL is a method designed for transfer learning, but it cannot provide better performance than TSVM. This result verifies the analysis in section 3: a small amount of HFDS features occupy a large amount of *weight* in classification model, but hardly carry corresponding sentiment. In another word, very few top-frequency words degrade the representative ability of SCL model for sentiment classification.

6 Conclusion Remarks

In this paper, we proposed a weighted SCL model (W-SCL) for domain adaptation in the context of sentiment analysis. On six domain-transfer tasks, W-SCL consistently produces much better performance than the supervised, semi-supervised and transfer-learning baselines. As a result, we can say that proposed W-SCL model offers a better choice for sentiment-analysis applications that require high-precision classification but hardly have any labeled training data.

7 Acknowledgments

This work was mainly supported by two funds, i.e., 0704021000 and 60803085, and one another project, i.e., 2004CB318109.

References

- Blitzer, J. and McDonald, R. and Fernando Pereira. Domain adaptation with structural correspondence learning. EMNLP 2006.
- DaumeIII, H. and Marcu, D. Domain adaptation for statistical classifiers. Journal of Artificial Intelligence Research, 2006, 26: 101-126.
- Jiang, J., Zhai, C. A Two-Stage Approach to Domain Adaptation for Statistical Classifiers. CIKM 2007.
- Pang, B., Lee, L. and Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. EMNLP 2002.
- Sebastiani, F. Machine learning in automated text categorization. ACM Computing Surveys. 2002, 34(1): 1-47.
- S. Tan, G. Wu, H. Tang and X. Cheng. A novel scheme for domain-transfer problem in the context of sentiment analysis. CIKM 2007.
- S. Tan, Y. Wang, G. Wu and X. Cheng. Using unlabeled data to handle domain-transfer problem of semantic detection. SAC 2008.
- S. Tan, X. Cheng, Y. Wang and H. Xu. Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis. ECIR 2009.
- H. Tang, S. Tan, and X. Cheng. A Survey on Sentiment Detection of Reviews. Expert Systems with Applications. Elsevier. 2009, doi:10.1016/j.eswa.2009.02.063.
- Turney, P. D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. ACL 2002.