Sentiment Classification via Integrating Multiple Feature Presentations

Yuming Lin, Jingwei Zhang, Xiaoling Wang, Aoying Zhou
Institute of Massive Computing
East China Normal University
200062 Shanghai, China
{ymlinbh, gtzhjw}@gmail.com, {xlwang, ayzhou}@sei.ecnu.edu.cn

ABSTRACT

In the bag of words framework, documents are often converted into vectors according to predefined features together with weighting mechanisms. Since each feature presentation has its character, it is difficult to determine which one should be chosen for a specific domain, especially for the users who are not familiar with the domain. This paper explores the integration of various feature presentations to improve the classification accuracy. A general two phases framework is proposed. In the first phase, we train multiple base classifiers with various vector spaces and use these classifiers to predict the class of testing samples respectively. In the second phase, the previous predicted results are integrated into the ultimate class via *stacking* with SVM. The experimental results demonstrate the effectiveness of our method.

Categories and Subject Descriptors

I.2.7 [**Natural Language Processing**]: Text Analysis - *Sentiment Analysis*

General Terms

Algorithms, Experimentation

Keywords

sentiment classification, multiple feature presentations, base classifiers, integrating

1. INTRODUCTION

More and more people prefer to write reviews and share their opinions on e-commerce sites and microblogging. Determining the overall sentiment of such user-generated data is valuable to many applications like recommendation systems, business and government intelligence, etc. In the bag of words framework, the documents are often converted into vectors based on predefined feature presentation including feature type and features weighting mechanism, which is critical to classification accuracy. The major feature types contain unigrams, bigrams and the mixtures of them, etc. The features weighting mechanism mainly includes presence, frequency, tf*idf and its variants. Pang et al. [4] have compared the performances of different learning methods, and reported that the unigram presence with SVM is superior to the others. Our preliminary experiments (See Table 1 for more details) show that none of the feature presentations can always outperform the others for all domains. Thus, it is a challenge to determine which feature presentation should be used, especially for the users who aren't familiar with the analyzed domains.

In this paper we put forward a general framework to integrate multiple base classifiers for a binary task of classifying sentiment into positive and negative classes. The base classifiers are trained by samples expressed using different feature presentations. Our solution consists of two phases: the training phase and the integrating phase. In the first phase, multiple feature presentations are applied on modeling the samples, and the base classifiers are trained by these transformed samples. In the second phase, we integrate the predictions generated by base classifiers with assembling technology such as *voting* and *stacking*[5]. Compared with the traditional classification with single feature presentation, our framework offers two major advantages: first, we don't need to determine which feature presentation should be chosen for classification tasks; second, the accuracy of our framework is always higher than that of single feature presentation integrated. In our framework, the task of training base classifiers can be processed in parallel easily.

2. THE PROPOSED METHOD

The n training samples are converted into n vectors according to m feature presentations. Let $X_i = (x_1^i, \ldots, x_n^i)^T$ denote such a vectors space using the i^{th} feature presentation, where x_j^i is the row vector of the j^{th} training sample. l_{s_k} is the k^{th} sample's class. $Y_i = (y_1^i, \ldots, y_r^i)^T$ is similar to X_i, excepting y denotes the row vector of testing sample. Our framework includes the following steps:

1. Converting training samples into m vector spaces with m feature presentations, and using m them to train m base classifiers respectively;(Line 2–4 in Algorithm 1)

2. training a integrating classifier based on the predicting results of m base classifiers; (Line 5–9)

3. Using m base classifiers to predict the classes of r testing samples independently, generating m predicting results for each testing sample;(Line 12–13)

4. For each test sample, the predicting results of m base classifiers are integrated into one ultimate result by *stacking* with SVM. (Line 14)

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

In Algorithm 1, the parameter k refers to the k-fold cross validation procedure is used to generate meta-level training samples for the integrating classifier $\tilde{c}$.

```
Algorithm 1. Sentiment Classification by Stacking
Input:the training vector set $X=\{\langle X_1,l_{s_1}\rangle,\dots,\langle X_n,l_{s_n}\rangle\}$,
      the testing vector set $Y = \{Y_1,\dots,Y_r\}$,
      $k$
Output:the ultimate predicted result set $L = \{l_1,\dots,l_r\}$
for $i$=1 to $k$ do
    $X^i_{meta} = X^i_{meta} \cup \{n/k$ samples never chosen in $X\}$;
    generate the base classifier set $C_i$ on $X - X^i_{meta}$;
    $C = C \cup \{C_i\}$;
    for $j$=1 to $n/k$ do
        for $t$=1 to $m$ do      //$m$ feature presentations
            $\widehat{l}^t_{s_j} = c_t(x^t_j)$;      //$c_t \in C_i$
        $T_{meta} = T_{meta} \cup \{\langle\langle\widehat{l}^1_{s_j},\dots,\widehat{l}^m_{s_j}\rangle, l_{s_j}\rangle\}$;
train the integrating classifier $\tilde{c}$ on $T_{meta}$;
select the $C_h$ with the highest accuracy in $C$;
for $i = 1$ to $r$ do
    for $j$=1 to $m$ do
        $\widehat{l}^j_i = c_j(y^j_i)$;              //$c_j \in C_h$
    $L = L \cup \{\tilde{c}(\langle\widehat{l}^1_i,\dots,\widehat{l}^m_i\rangle)\}$;
return $L$;
```

3. EXPERIMENT

To validate the effectiveness and robustness of proposed method, a real-world dataset reorganized by Blitzer et al.[1] is prepared for experiments, which consists of reviews of books (**B**), DVDs (**D**), electronics (**E**) and kitchen appliances (**K**) from Amazon. Each domain contains 1000 positive and 1000 negative reviews. Five-fold cross validation is applied. The LIBSVM [2] with a linear kernel is used to implement SVM algorithm. We focus on three main feature types (unigrams, bigrams and the mixture of them) and feature weighting mechanisms (frequency, presence and Δtf*idf [3]). Multiple feature presentations can be integrated in our framework, but in our experiments we only consider seven combinations described in Table 1, where "uni_f" refers to the unigrams with frequency, "bi_p" to bigrams with presence and "Δtf*idf" to unigrams with Δtf*idf, and so on.

3.1 Preprocess

All punctuation marks and the terms occurring less than 5 times are eliminated but the stop words retained. Similar to [4], we remove the negatory words and append the tag "NOT_" to the words following the negatory word in the sentence. For instance, the sentence "It doesn't work smoothly." would be altered to become "It NOT_work NOT_smoothly.".

3.2 Results and Discussion

Table 1 reports the classification accuracies using various feature presentations separately in different domains. The accuracy shown in bold font corresponds to the optimal feature presentation in the corresponding domain. For example, the bold 87.00 means the unigrams with presence achieves the highest accuracy in kitchen appliances domain. These results verify our motivation, that is no one feature presentation can outperform the others in all domains. Thus we propose a framework to integrate multiple feature presentations for improving the classification accuracy.

Table 1: The classification accuracy (%) using different feature presentations

	uni_f	bi_f	uni+bi_f	uni_p	bi_p	uni+bi_p	Δtf*idf
B	79.90	75.75	**80.45**	78.25	75.55	79.50	79.85
D	79.15	74.00	79.80	78.60	74.45	80.25	**83.00**
E	83.80	79.90	83.85	83.10	80.45	**85.15**	84.40
K	84.95	80.40	86.65	**87.00**	80.55	86.50	86.20

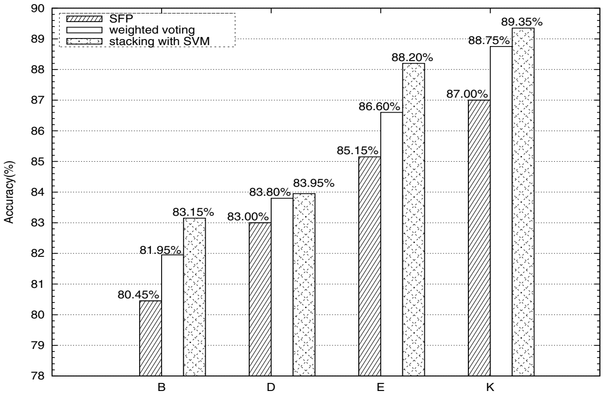

Figure 1: Comparison on accuracies

The single feature presentations (**SFP**) with highest accuracy in each domain are treated as baselines. The weighted voting scheme is a simple way to assembling multiple classifiers, and we treat it as a competitor of the stacking with SVM method. Figure 1 illustrates the comparison of above three methods. For sentiment classification in various domains, on the one hand, the proposed method outperforms the single feature presentation used, on the other hand the stacking with SVM is superior to the weighted voting.

4. ACKNOWLEDGMENTS

This work was supported by the 973 project (No. 2010CB328106), NSFC grant (No. 61170085 and 60925008), Program for New Century Excellent Talents in China (No. NCET-10-0388) and Innovation Program of Shanghai Municipal Education Commission.

5. REFERENCES

[1] J. Blitzer, M. Dredze, and F. Pereira. Biographies, bollywood, boo-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, pages 440–447, June 2007.

[2] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[3] J. Matineau and T. Finin. Delta tfidf: An improved feature space for sentiment analysis. In *ICWSM*, pages 258–261, May 2009.

[4] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using maching learning technique. In *EMNLP*, pages 79–86, July 2002.

[5] B. Ženko, L. Todorovski, and S. Džeroski. A comparison of stacking with mdts to bagging, boosting, and other stacking methods. In *ICDM*, pages 669–670, December 2001.