Imbalanced Sentiment Classification with Multi-Strategy Ensemble Learning

Zhongqing Wang, Shoushan Li*, Guodong Zhou, Peifeng Li and Qiaoming Zhu
Natural Language Processing Lab
Soochow University
Suzhou, China
{wangzq870305, shoushan.li}@gmail.com,{gdzhou, pfli, qmzhu}@suda.edu.cn

Abstract—**Recently, sentiment classification has become a hot research topic in natural language processing. But most existing studies assume that the samples in the negative and positive categories are balanced, which might not be true in real applications. In this paper, we investigate sentiment classification tasks where the class distribution of the samples is imbalanced. To handle the imbalanced problem, we propose a multi-strategy ensemble learning approach to this problem. Our ensemble approach integrates sample-ensemble, feature-ensemble, and classifier-ensemble by exploiting multiple classification algorithms. Evaluation across four domains shows that our ensemble approach outperforms many other popular approaches that handling imbalanced classification problems, such as re-sampling and cost-sensitive approaches, and is proven effective for imbalanced sentiment classification.**

Keywords: **sentiment classification; imbalanced classification; ensemble learning**

I. INTRODUCTION

The purpose of sentiment classification is to predict the positive and negative polarities of a given text [15]. The computational treatment of sentiment classification has recently attracted a great deal of attention, such as online product review classification [4], opinion summarization [13], and opinion retrieval [20].

Supervised learning approaches have been widely employed and proven effective in sentiment classification in the literature [15], but they normally assume the balance between negative and positive samples, which may not keep in practice. Actually, many sentiment classification applications involve imbalanced class distributions in that the sample number of one class in the training data is much larger than the other class. For clarity, the class with more samples is referred to as *majority class* (MA) and the other class with fewer samples is referred to as *minority class* (MI).

Many standard methods, such as re-sampling [2], one-class classification [7], cost-sensitive learning [21], and ensemble learning [5], have been proposed to solve imbalanced class distribution problems. Among them, ensemble learning approach has been one of the most promising approaches for handling the imbalanced classification problems given its excellent performance improvement. Previous studies mainly focus on sample-ensemble learning: first generating either multiple over-sampling MI data sets [3] or under-sampling MA data sets [11], and then training multiple classifiers with the obtained data sets (These classifier are referred to as member classifiers). The reported performances of such approaches outperform many other standard approaches.

However, it is well known that the diversity of the member classifiers is a key issue for the success of ensemble learning [8]. Note that the training data for each classifier contains almost the same MI samples in sample-ensemble learning. Although the class imbalance problem has been solved, the diversities of the member classifiers are considered inadequate due to the same training samples in the MI.

In this paper, we propose a novel multi-strategy ensemble learning approach to maximize the diversities among the member classifiers. Our approach integrates the feature-ensemble strategy and classifier-ensemble strategy into the existing sample-ensemble strategy to guarantee the high diversity of the involved member classifiers. Specifically, the feature-ensemble strategy exploits random subspace generation method to generate two feature subspace classifiers, whereas classifier-ensemble strategy trains the member classifiers by employing different classification algorithms, such as Naïve Bayes and Maximum Entropy. Experimental results show that our approach significantly outperforms various existing methods as well as the method of using sample-ensemble strategy alone for imbalanced classification.

The remainder of this paper is organized as follows. Section II overviews the related work in sentiment classification. Section III presents our multi-strategy ensemble learning framework in handling imbalanced class distribution. Section IV gives the experimental results. Finally, Section V draws the conclusion and outlines the future work.

II. RELATED WORK

Sentiment classification can be performed on words, sentences or documents [18]. This paper focuses on document-level sentiment classification.

In the literature, there are mainly two kinds of approaches on document-level sentiment classification: term-counting approaches (lexicon-based) and machine learning approaches (corpus-based).

For term-counting approaches, a sentiment classifier is derived from a set of labeled sentiment words called seed words, instead of labeled documents. Specifically, the relationship between each two words are extracted based on certain knowledgebase resources such as WordNet and unlabeled corpora. The sentiment polarity of a text is then predicted by calculating the semantic orientation of the words in the text using such relationship [17].

Compared to unsupervised leaning approaches, supervised learning methods have become more popular since the pioneer work on sentiment classification by Pang et al.

* Corresponding Author

978-0-7695-4554-7/11 $26.00 © 2011 IEEE
DOI 10.1109/IALP.2011.28

[15]. They often yield better performances due to the availability of labeled data. In particular, various kinds of information have been explored to improve the bag-of-words model, such as contextual features [16], document subcomponent information [14], and polarity shifting [10].

The imbalanced class distribution problem is a new research area in sentiment classification. Despite the fact that Liu et al. [12] propose a simple probability-based term weighting scheme to better distinguish documents in the MI samples on imbalanced text categorization. However, few work has focus on the imbalanced problem in sentiment classification.

III. Multi-Strategy Ensemble for Imbalanced Sentiment Classification

Given N samples in the training data including N_+ positive samples and N_- negative samples, most of the existing studies assume the balance between the number of positive samples and the number of negative samples in sentiment classification, i.e., $N_+ = N_-$, which may not hold in real applications. Normally, there are more samples in one class than the other in the training data, i.e., $N_+ << N_-$ or $N_+ >> N_-$.

A. Sample-Ensemble Strategy

Under-sampling is an effective strategy to deal with class imbalanced problem [1,19]. However, the major problem of under-sampling is that it discards many potentially useful MA samples. Therefore, the sample-ensemble learning strategy is proposed to fully exploit all MA samples [12].

Given the MI training set S_{MI} and the MA training set S_{MA}, the under-sampling approach randomly selects a sample subset S'_{MA}, where $|S'_{MA}|<|S_{MA}|$ and $|S'_{MA}|=|S_{MI}|$.

In sample-ensemble strategy, several subsets $S'_{MA1}, S'_{MA2} ..., S'_{MAn}$ are independently selected from S_{MA}. For each S'_{MAi} $(1 \le i \le n)$, the condition of $|S'_{MAi}|=|S_{MI}|$ is satisfied just as under-sampling does. Given the multiple data sets, multiple classifiers C_i (i=1, 2, ..., n) are trained with each S'_{MAi}. The whole dataset of S_{MI} and all generated classifiers are then combined for the final decision [12]. Generally, the generated classifiers for combination are called member classifiers.

B. Feature-Ensemble Strategy

An important issue in ensemble learning is the diversity of classifiers, i.e., the member classifiers should be different from each other as much as possible. However, in sample-ensemble learning, each classifier contains almost the same MI samples. Therefore, we first propose feature-ensemble strategy to enlarge the diversity of the member classifiers.

Specifically, we use Random Subspace Generation (RSG) approach to generate feature subspace classifiers in feature-ensemble learning. RSG is an ensemble technique proposed by Ho [6]. Normally, multiple classifiers (also called subspace classifiers) can be first constructed in random subspaces using modified training sets and are then combined using a simple majority voting strategy for supervised learning [6].

In sentiment classification, a document is usually modeled as one bag-of-words and represented as a vector of features. For each subset of training data generated by sample-ensemble learning, we first randomly select half of the features to train one subspace classifier and leave the remaining half to train another subspace classifier. This subspace generation approach would make the member subspace classifiers quite different from each other although the MI samples in each subspace classifier remain the same.

C. Classifier-Ensemble Strategy

Besides the feature-ensemble strategy, we also propose a classifier-ensemble strategy to guarantee the diversity of the involved classifiers. The classifier-ensemble strategy involves more than one classification algorithm. That is to say, multiple classification algorithms are prepared and the classifier of each subset as well as each data set is trained with a random selected classification algorithm. In our experiments, we employ three algorithms: Naïve Bayes, SVM, and Maximum Entropy.

D. Multi-Strategy Ensemble learning

The objective of all of the above strategies is to fully utilize the training data and to guarantee the diversity of the multiple classifiers. To take full advantage of these strategies, we propose a novel approach which integrates all of the above strategies. The whole algorithm of our multi-strategy ensemble learning approach for imbalanced sentiment classification is illustrated in Figure 1.

Input: The MI training set S_{MI} and the MA training set S_{MA}
Output: A combination classifier
Algorithm:
1) $i \leftarrow 0$
repeat T times:
2-1) $i \leftarrow i+1$
2-2) randomly select a subset S'_{MAi} from S_{MA} and keep $|S'_{MAi}|=|S_{MI}|$
2-3) generate two feature subspaces from $S'_{MAi} \cup S_{MI}$
2-4) randomly choose a classification algorithm to train two subspace classifiers C_{i1} and C_{i2}
3) combine all generated classifiers C_{ik} (i=1,2,...,T, k=1,2)

Figure 1. Multi-Strategy Ensemble Learning Framework

In detail, we first adopt under-sampling approach to generate multiple sets of imbalanced initial training data (step 2-2); second, we generate two feature subspaces for each subset of the training data (step 2-3); third, a classification algorithm is randomly chosen to train each subspace classifier (step 2-4); finally, all generated classifiers are combined for the final decision (step 3).

Table II: PERFORMANCES OF THE IMBALANCED CLASSIFICATION ALGORITHMS

Domain	Full Training	Cost Sensitive	One Class	Term Weight	Re-Sampling		Ensemble Learning			
					Under-	*Over-*	*Sample-*	*Feature-*	*Classifier-*	*Multi-Strategy*
Book	0.624	0.765	0.466	0.381	0.768	0.671	0.797	0.799	0.797	**0.814**
DVD	0.630	0.770	0.512	0.403	0.742	0.698	0.765	0.784	0.787	**0.797**
Electronic	0.688	0.771	0.497	0.566	0.753	0.707	0.757	0.786	**0.797**	0.795
Kitchen	0.740	0.763	0.580	0.654	0.778	0.752	0.785	0.818	0.830	**0.834**
Aver.	0.670	0.767	0.514	0.501	0.760	0.707	0.776	0.797	0.803	**0.810**

IV. EXPERIMENTAL STUDIES

The data collection consists of four domains: Book, DVD, Electronic, and Kitchen[1]. For each domain, we randomly select a set of training data with 1000 negative samples, keeping the same imbalanced ratio as the whole data. That is to say, the numbers of positive samples in the four domains are different. For example, as the imbalanced ratio of the Books domain is 7.29, the number of the positive samples is 7290 (1000×7.29). For the test data in each domain, we randomly select 200 negative and 200 positive samples.

TABLE I: CONFUSION MATRIX FOR A TWO-CLASS PROBLEM

	Positive Prediction	Negative Prediction
Positive class	True Positive (*TP*)	False Negative (*FN*)
Negative class	False Positive (*FP*)	True Negative (*TN*)

Table I illustrates the basic performance measures built over a 2×2 confusion matrix, where *TP* and *TN* denote the number of correctly classified positive and negative samples respectively, while *FP* and *FN* denote the number of misclassified positive and negative samples, respectively. We adopt the popular geometric mean (*G-mean*) to measure the results [9]:

$$G-Mean = \sqrt{TP_{rate} \times TN_{rate}}$$

$$\text{Where } TP_{rate} = \frac{TP}{TP+FN}, TN_{rate} = \frac{TN}{TN+FP}$$

Where TP_{rate} is the true positive rate (also called positive recall or sensitivity) and TN_{rate} is the true negative rate (also called negative recall or specificity)

The Maximum Entropy classifier and Naïve Bayes classifier are implemented using the *Mallet*[2] *tool*, while SVM is implemented using the SVM-*light*[3] tool, and the features are unigram words with Boolean weights. For thorough comparison, various settings are implemented:

1) Full-training: directly throwing all the training data for training.

2) Cost-sensitive classification: performing cost-sensitive classification [21] with the *lib*-SVM[4] *tool*. The cost weight for a MA sample is set to the imbalanced ratio between the MI and MA samples in each domain while the cost weight for a MI sample is 1.

3) One-class classification: performing one-class classification [7] with the *lib*-SVM *tool*.

4) Term weighting: adopting the approach proposed by Liu et al. [12].

5) Re-sampling approach: implementing two most popular re-sampling approaches:

a) Over-sampling: performing over-sampling by randomly selecting the MI samples.

b) Under-sampling: performing under-sampling by randomly selecting the MA samples.

6) Ensemble-learning approach: implementing three kinds of ensemble learning approach and the multi-strategy ensemble learning:

a) Sample-ensemble learning: independently sampling several subsets from the MA samples and developing multiple classifiers based on the combination of each subset with the MI samples.

b) Feature-ensemble learning: exploits random subspace generation method to generate two feature subspaces in training two individual classifiers for each subset generated by sample-ensemble learning.

c) Classifier-ensemble learning: random choose different classification method for each subset generated by sample-ensemble learning.

d) Multi-strategy ensemble learning: combining feature-ensemble learning and classifier-ensemble learning with sample-ensemble learning as discussed in Section III.

Since most settings involve random selection of samples, we run 10 times for each setting and report the average performance of the 10 runs.

Table II overviews the performance of the four domains of imbalanced sentiment classification, for clarity, only *G-mean* is reported. It shows that almost all the specifically designed methods outperform full-training except term weighting. This is mainly due to the fact that term weighting fails to take the sample imbalance into account. Both under-sampling and cost-sensitive classification always perform better than full training, one-class classification, and over-sampling. These badly performed approaches are more likely to classify test samples as the MA class and thus perform badly in terms of *G-mean.*

Consistently, ensemble-learning always performs much better than the other approaches. Among the ensemble approaches, feature-ensemble learning and classifier-ensemble learning significantly outperforms sample-ensemble learning alone. Our approach (multi-strategy) achieves the best performance on average, which confirms its effectiveness to enlarge the diversities of the member classifiers. Significance test shows that our approach is

[1] The data are from multi-domain sentiment dataset v2.0. http://www.seas.upenn.edu/~mdredze/datasets/sentiment/.
[2] http://mallet.cs.umass.edu/
[3] http://svmlight.joachims.org/
[4] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

significantly better than sample-ensemble learning with *t*-test (at 0.01 level).

V. Conclusion and Future Work

In this paper, we investigate sentiment classification tasks suffering from imbalanced class distribution. In particular, a multi-strategy ensemble learning approach is proposed to solve this problem. In our approach, we integrate sample-ensemble with under-sampling, feature-ensemble with random subspace generation, classifier-ensemble with random selection of classification algorithms. Evaluation shows that our approach performs remarkably better than some other standard approaches, such as re-sampling approach and cost-sensitive approach and also performs significantly better than using sample ensemble alone.

In our future work, we will consider many other issues such as feature selection and active learning in imbalanced sentiment classification.

VI. Acknowledgments

The research work described in this paper has been partially supported by NSFC grants, No. 90920004, No. 61070123, No. 61003153, No. 60970056 and No. 61003155 and Open Projects Program of National Laboratory of Pattern Recognition.

References

[1] Barandela R., J.S. Sánchez, V. García, and E. Rangel. 2003. Strategies for Learning in Class Imbalance Problems. Pattern Recognition, vol.36, pp.849–851, 2003.

[2] Chawla N., K. Bowyer, L. Hall, and W. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research, vol.16, pp.321–357, 2002.

[3] Chawla N., A. Lazarevic, L. Hall, and K.W. Bowyer. 2003. SMOTEBoost: Improving Prediction of the Minority Class in Boosting. In Proceedings of Seventh European Conf. Principles and Practice of Knowledge Discovery in Databases, pp. 107-119.

[4] Cui H., V. Mittal, and M. Datar. 2006. Comparative Experiments on Sentiment Classification for Online Product Reviews. In Proceedings of AAAI-06, pp.1265-1270.

[5] Guo H. and H. Viktor. 2004. Learning from Imbalanced Data Sets with Boosting and Data Generation: the Databoost-im Approach. ACM SIGKDD Explor. Newsl., vol.6(1), pp.30-39, 2004.

[6] Ho T. 1998. The random subspace method for constructing decision forests, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.20(8), pp.832-844, 1998

[7] Juszczak P. and R. Duin. 2003. Uncertainty Sampling Methods for One-Class Classifiers. In Proceedings of ICML-03, Workshop on Learning with Imbalanced Data Sets II, pp.81–88.

[8] Kittler J., M. Hatef, R. P.W. Duin, and J. Matas.On Combining Classifiers . IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.20,226-239,1998.

[9] Kubat M. and S. Matwin. 1997. Addressing the Curse of Imbalanced Training Sets: One-Sided Selection. In Proceedings of ICML-97, pp.179–186.

[10] Li S., S. Lee, Y. Chen, C. Huang, and G. Zhou. 2010. Sentiment Classification and Polarity Shifting. In Proceedings of COLING-10, pp.635-643.

[11] Liu X., J. Wu, and Z. Zhou. 2006. Exploratory Under Sampling for Class Imbalance Learning. In Proceeding of ICDM-06, pp. 965-969.

[12] Liu Y., H. Loh, and A. Sun. 2009. Imbalanced Text Classification: a Term Weighting Approach. Expert System with Applications, vol.36(1), pp.690–701, 2009.

[13] Lloret E., A. Balahur, M. Palomar, and A. Montoyo. Towards Building a Competitive Opinion Summarization System. In Proceedings of NAACL-09 Student Research Workshop and Doctoral Consortium, pp.72-77.

[14] McDonald R., K. Hannan, T. Neylon, M. Wells, and J. Reynar. 2007. Structured Models for Fine-to-coarse Sentiment Analysis. In Proceedings of ACL-07, pp.432-439.

[15] Pang B., L. Lee, and S. Vaithyanathan. 2002.Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of EMNLP-02, pp.79-86.

[16] Riloff E., S. Patwardhan, and J. Wiebe. 2006. Feature Subsumption for Opinion Analysis. In Proceedings of EMNLP-06, pp.440-448.

[17] Turney P. 2002. Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of reviews. In Proceedings of ACL-02, pp.417-424.

[18] Wan X. 2009. Co-Training for Cross-Lingual Sentiment Classification. In Proceedings of ACL-IJCNLP-09, pp.235–243.

[19] Yen S. and Y. Lee. 2009. Cluster-Based Under-Sampling Approaches for Imbalanced Data Distributions. Expert Systems with Applications. vol.36, pp.5718-5727, 2009

[20] Zhang M. and X. Ye. 2008. A Generation Model to Unify Topic Relevance and Lexicon-based Sentiment for Opinion Retrieval. In Proceedings of SIGIR-08, pp.411-418.

[21] Zhou Z. and X. Liu. 2006. Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem. IEEE Transaction on Knowledge and Data Engineering, vol.18, pp.63–77, 2006.