Semi-Supervised Learning for Imbalanced Sentiment Classification

Shoushan Li[†], Zhongqing Wang[†], Guodong Zhou[†*] and Sophia Yat Mei Lee[‡]

[†] Natural Language Processing Lab
Soochow University, China
{shoushan.li, wangzq870305}
@gmail.com, gdzhou@suda.edu.cn

[‡] Department of CBS
The Hong Kong Polytechnic University
sophiaym@gmail.com

Abstract

Various semi-supervised learning methods have been proposed recently to solve the long-standing shortage problem of manually labeled data in sentiment classification. However, most existing studies assume the balance between negative and positive samples in both the labeled and unlabeled data, which may not be true in reality. In this paper, we investigate a more common case of semi-supervised learning for imbalanced sentiment classification. In particular, various random subspaces are dynamically generated to deal with the imbalanced class distribution problem. Evaluation across four domains shows the effectiveness of our approach.

1 Introduction

Sentiment classification aims to predict the sentiment polarity of a text [Pang et al., 2002] and plays a critical role in many Natural Language Processing (NLP) applications [Liu et al., 2005; Wiebe et al., 2005; Cui et al., 2006; Lloret et al., 2009]. Although supervised learning methods have been widely employed and proven effective in sentiment classification in the literature [Pang et al., 2002], they normally depend on a large amount of labeled data, which usually involves high cost in labor and time. To overcome this problem, various semi-supervised learning methods are proposed to effectively utilize a small scale of labeled data along with a larger amount of unlabeled data [Dasgupta and Ng, 2009; Wan, 2009; Li et al., 2010].

However, all the existing semi-supervised learning methods assume the balance between negative and positive samples in both the labeled and unlabeled data, and none of them consider a more common case where the class distribution is imbalanced, i.e., the number of positive samples is quite different from that of negative samples in both the labeled and unlabeled data. For clarity, the class with more samples is referred as the *majority class* (*MA*) and the other class with fewer samples is referred as the *minority class* (*MI*). In fact, semi-supervised learning on imbalanced classification is rather challenging: at least, there exist two basic issues to be solved. On the one hand, imbalanced classification requires a specifically-designed classification algorithm. Trained on the imbalanced labeled data, most classification algorithms tend to predict test samples as the *majority class* and may ignore the *minority class*. Although many methods, such as re-sampling [Chawla et al., 2002], one-class classification [Juszczak and Duin, 2003], and cost-sensitive learning [Zhou and Liu, 2006], have been proposed to solve this issue, it is still unclear as to which method is more suitable to handle the imbalanced problem in sentiment classification and whether the method is extendable to semi-supervised learning. On the other hand, given the classification algorithm and the unlabeled data, which method is effective for capturing the inherent information in the unlabeled samples to improve the performances? Unfortunately, the issue of semi-supervised learning on imbalanced data sets has not been carefully studied in the literature.

In this study, we adopt under-sampling (a typical re-sampling approach) to deal with the imbalanced problem in sentiment classification due to its fine performance in our empirical study (A detailed comparative study can be found in Section 5.2). The under-sampling approach randomly selects a subset of the *MA* samples from the initial training data (the given labeled data) and then combines them with all the *MI* samples to form a new initial training set. Accordingly, given the new balanced initial training data, any existing semi-supervised learning method, such as co-training with personal/impersonal views [Li et al., 2010], can be used to make use of the unlabeled samples. However, one obvious limitation of such method is that under-sampling throws out many *MA* samples which might be useful for further semi-supervised learning. Therefore, to make better use of the given labeled data, we first perform under-sampling several times on the *MA* samples to obtain multiple subsets. Then the *MA* samples in each subset, together with all the *MI* samples, form a set of initial training data. Finally, multiple sets of initial training data are used to train multiple classifiers, which work together via an ensemble to select confident samples from the unlabeled data, in the same way as co-training.

One key problem with co-training is that it makes a strong assumption that the two feature sets involved should be

[*] Corresponding author

conditionally independent given a class [Blum and Mitchell, 1998]. Although some studies have been done to relax this assumption with weaker ones [Balcan et al., 2005], a basic intuitive requirement, that the involved classifiers should be different enough from each other, should be met so that they can complement each other. Note that the initial training data for each classifier contains the same *MI* samples. Another problem is that, as more and more samples from the labeled data are iteratively picked to train the involved classifiers, the difference between the involved classifiers may become smaller and smaller during the co-training process. Therefore, to guarantee the diversity of the involved classifiers, we propose a novel random subspace generation method to dynamically generate two feature subspaces in training two individual classifiers given one set of the training data.

The remainder of this paper is organized as follows. Section 2 overviews the related work in semi-supervised sentiment classification as well as imbalanced classification. Section 3 discusses the task of imbalanced sentiment classification. Section 4 presents our approach of semi-supervised learning for imbalanced sentiment classification. Section 5 evaluates the experimental results. Finally, Section 6 draws the conclusion and outlines the future work.

2 Related Work

This section gives a brief overview of the related work in both semi-supervised sentiment classification and imbalanced classification.

2.1 Semi-supervised Sentiment Classification

Generally, sentiment classification methods can be categorized into three types: unsupervised [Turney, 2002], supervised [Pang et al., 2002], and semi-supervised [Sindhwani and Melville, 2008]. Compared to supervised and unsupervised methods, semi-supervised methods for sentiment classification become more and more popular due to their making use of both the labeled and unlabeled data. This paper mainly focuses on semi-supervised methods for sentiment classification.

One kind of semi-supervised methods for sentiment classification is to utilize prior lexical knowledge in conjunction with the labeled and unlabeled data. For example, Sindhwani and Melville [2008] jointly analyzed the sentiment of documents and words based on a bipartite graph representation of the labeled and unlabeled data while Li et al. [2009] employed some simple update rules to make use of tri-factorization of the term-document matrix. It is rather common that such methods require a high-quality lexicon with the polarity of words properly defined.

Another kind of semi-supervised methods for sentiment classification is to employ some bootstrap techniques, such as self-training [Yarowsky, 1995] and co-training [Blum and Mitchell, 1998]. Among them, co-training has been proven more effective than self-training [Wan, 2009; Li et al., 2010]. The key issue of applying co-training is to find a suitable set of different views. For instance, Wan [2009] regarded two different languages (i.e., English and Chinese) as two views while Li et al. [2010] considered personal and impersonal texts as two views. This paper employs the co-training technique and generates different views from random feature subspaces.

Among others, Dasgupta and Ng [2009] integrated various methods, such as spectral clustering, active learning, transductive learning, and ensemble learning, in semi-supervised sentiment classification.

To our best knowledge, no existing semi-supervised methods consider the class imbalance problem in sentiment classification.

2.2 Imbalanced Classification

Imbalanced classification, as a challenging learning problem, has been widely studied in several research areas, such as machine learning [Kubat and Matwin, 1997], pattern recognition [Barandela et al., 2003], and data mining [Chawla et al., 2004], at either data level or algorithmic level [Chawla et al., 2004].

At the data level, different forms of re-sampling, such as over-sampling and under-sampling, are proposed. Specifically, over-sampling aims to balance the class populations through replicating the *MI* samples [Chawla et al., 2002] while under-sampling aims to balance the class populations through eliminating the *MA* samples [Barandela et al., 2003; Yen and Lee, 2009].

At the algorithmic level, specific learning algorithms, such as cost-sensitive learning, one-class learning, and ensemble learning [Juszczak and Duin, 2003; Guo and Viktor, 2004; Zhou and Liu, 2006], are proposed. For more details, please refer to the comprehensive survey by He and Garcia [2009].

However, all the existing studies on imbalanced classification only focus on supervised imbalanced classification. Until now, there are no semi-supervised methods reported on imbalanced classification.

3 Imbalanced Sentiment Classification

Given 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, i.e., $n_+ = n_-$, which may not hold in real applications. Normally, there are more samples in one class than the other class in the training data. i.e., $n_+ << n_-$ or $n_+ >> n_-$.

To better understand the imbalanced class distribution problem in sentiment classification, let us examine a popular data set collected by Blitzer et al. [2007]. This data set contains four domains, each of which consists of 1,000 positive and 1,000 negative documents. In fact, this balanced data is not the real collection from websites but is extracted from another data set[1] with more documents. We re-collect all documents of these four domains from the original data set and show their real class distributions in Table 1. As we can see in Table 1, all the class distributions are imbalanced, with

[1] The data is from multi-domain sentiment dataset v2.0. http://www.seas.upenn.edu/~mdredze/datasets/sentiment/ .

the imbalanced ratios (n_+/n_-) ranging from 3 to 8. Interestingly, the number of negative samples is normally much smaller than that of positive samples, largely due to two reasons: (1) people tend to publish their opinions on popular products, which are more likely positive; (2) there may exist many flaunt positive reviews from the product companies and dealers. Although there may exist some flaunt negative reviews from opponents, the number is much less. In this sense, negative reviews may be more valuable than positive reviews for potential buyers.

Domain	n_+	n_-	n_+/n_-
Book	425,159	58,315	7.29
DVD	69,175	11,383	6.08
Electronic	15,397	4,316	3.57
Kitchen	14,290	3,784	3.78

Table 1: Class distributions of positive and negative samples across four typical domains

One straightforward way to handle imbalanced classification is applying a re-sampling method, which balances the positive and negative classes either by over-sampling the *MI* samples or by under-sampling the *MA* samples. In particular, random under-sampling has empirically been shown to be the most effective re-sampling method for dealing with imbalanced classification [Japkowicz and Stephen, 2001] and thus adopted in this paper.

4 Semi-supervised Learning for Imbalanced Sentiment Classification

This paper employs the co-training technique in semi-supervised learning for imbalanced sentiment classification. In particular, different views are generated from random (feature) subspaces.

4.1 Random Subspace Generation

Random Subspace Generation (RSG) is an ensemble technique proposed by Ho [1998]. Assume $X=(X_1,X_2,...,X_n)$ the training data and X_i a *m*-dimensional vector $X_i=(x_{i1},x_{i2},...,x_{im})$, described by *m* features. RSG randomly selects *r* ($r<m$) features and thus obtains an *r*-dimensional random subspace of the original *m*-dimensional feature space. In this way, a modified training set $X^S=(X_1^S,X_2^S,...,X_n^S)$ consisting of *r*-dimensional samples $X_i^s=(x_{i1}^S,x_{i2}^S,...,x_{ir}^S)$ $(i=1,...,n)$ is generated. Normally, multiple classifiers (called subspace classifiers) can be first constructed in random subspaces X^S using modified training sets and then combined using a simple majority voting strategy for supervised learning [Ho, 1998] or employed to select confidently unlabeled samples using a co-training technique in semi-supervised learning [Wang et al., 2008].

In sentiment classification, a document is usually modeled as one bag-of-words and represented as a vector of features, each of which is measured using the weight of the corresponding word (also called term) $(t_1,...,t_m)$ with *m* the number of terms. For the *r*-dimensional random subspace, only *r* terms are utilized to generate the feature vector. In our implementation, half of the features are randomly selected to generate a subspace (i.e., $r=m/2$) to keep a decent performance of the subspace classifiers. In particular, we randomly select half of the features to train one subspace classifier and leave the remaining half to train another subspace classifier.

4.2 Semi-supervised Learning with Dynamic Subspace Generation

As mentioned in Section 3, under-sampling is an effective way to handle imbalanced classification. Given the balanced initial training data after under-sampling, random subspace generation can naturally be applied to generate two subspace classifiers in co-training.

The major problem of the above technique is that it discards many potentially helpful *MA* samples. To fully utilize the training data, we iteratively perform under-sampling without duplication until there are not enough samples to form a set of balanced *MA* and *MI* samples [Liu et al., 2009]. As a result, multiple sets of initial training data are available and each data set is employed to generate two subspace classifiers.

Input:
Labeled data *L* containing n_+ positive and n_- negative samples;
Unlabeled data U
Output:
Automatically labeled sample set *A*
Procedure:
(a) Initialize $A=\varnothing$
(b) Perform under-sampling *K* times to get *K* sets of balanced initial training data with $K=(\text{int})\ (n_+/n_-)$
(c) Loop for *N* iterations:
  c1) Initialize the sample set $B=\varnothing$ which contains the most confidently labeled samples in each iteration
  c2) For *i* =1 to *K*:
    c2-1) Generate random two subspaces of the whole feature space
    c2-2) Train two subspace classifiers C_{i1} and C_{i2} with the *i*-th set of the initial training data
    c2-3) Use C_{i1} and C_{i2} to select one positive and one negative sample with most confidence, and put them into *B*
  c3) Put all samples in *B* into each set of the initial training data
  c4) $A=A\cup B$

Figure 1: Co-training with Multiple Under-sampling and Dynamic Subspace Generation

However, since all these subspaces have the same *MI* training data at first, the *MA* training data may become more similar when more unlabeled data are added. Therefore, instead of using only one set of subspaces, we dynamically generate the two subspaces during the iteration process in the co-training algorithm. Figure 1 illustrates the whole algorithm. For clarity, we refer to random subspace generation with only a fixed subspace splitting as static subspace generation while the one with dynamically-changing subspaces as dynamic subspace generation. The difference between these two methods lies in the position of Step (c2-1). If Step (c2-1) is moved before c), the dynamic subspace generation algorithm becomes a static one.

Basically, there are two main advantages of our dynamic strategy over the static strategy in generating random subspaces. First, the dynamic strategy makes the involved subspace classifiers quite different from each other even when the training data becomes similar after some iterations. Second, considering that the most helpful features (e.g., sentimental words) for sentiment classification usually account for a small portion, it is possible that one random subspace might contain few useful features. When this happens in the static strategy, the corresponding subspace classifier will perform badly in selecting correct samples from the unlabeled data. This makes semi-supervised learning fail. In comparison, the dynamic strategy can avoid this phenomenon.

5 Experiments

In this section, we will systematically evaluate our semi-supervised learning proposal for imbalanced sentiment classification on the data collection as described in Section 3.

5.1 Experimental Setting

The data collection consists of four domains: Book, DVD, Electronic, and Kitchen[2]. For each domain, we randomly sample an initial training data with 100 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 Book domain is 7.29, the number of the positive samples is around 729 (100×7.29). For the test data in each domain, we randomly extract 400 negative samples and 400 positive samples and leave all the remaining samples as the unlabeled data in semi-supervised learning.

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

Table 2: Confusion matrix for a two-class problem

Table 2 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, while *FP* and *FN* denote the number of misclassified positive and negative samples, respectively. We adopt the popular geometric mean $G\text{-}mean = \sqrt{TP_{rate} \times TN_{rate}}$, where $TP_{rate} = TP/(TP + FN)$ is the true positive rate (also called positive recall or sensitivity) and $TN_{rate} = TN/(TN + FP)$ is the true negative rate (also called negative recall or specificity). One advantage of *G-mean* is that it maximizes the accuracy of each of the two classes in order to balance both classes at the same time [Kubat and Matwin, 1997].

Besides, the Maximum Entropy classifier implemented within the Mallet[3] tool is adopted (except the one-class and cost-sensitive classifier as described in Section 5.2), and the features are unigram words with Boolean weights.

5.2 Experimental Results

Supervised Learning for Imbalanced Sentiment Classification

In this subsection, we report the performances of supervised methods for imbalanced sentiment classification (The unlabeled data is not involved). For thorough comparison, various kinds of supervised methods are implemented including:

1) **Full-training**: using all the training data for training.
2) **Over-sampling**: performing over-sampling by randomly selecting the *MI* samples.
3) **Under-sampling**: performing under-sampling by randomly selecting the *MA* samples.
4) **One-class classification**: performing one-class classification [Juszczak and Duin, 2003] with lib-SVM tool[4].
5) **Cost-sensitive classification**: performing cost-sensitive classification [Zhou and Liu, 2006] with lib-SVM 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.

Figure 2 shows the performance of these supervised methods on the four domains. It shows that almost all the specifically-designed methods outperform **full-training.** Among them, under-sampling always performs best while the other methods are more likely to classify test samples as the *MA* class and thus perform badly in terms of *G-mean*.

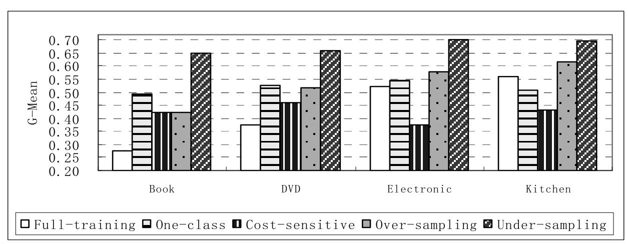

Figure 2: Performances of different supervised methods for imbalanced sentiment classification

Semi-supervised Learning for Imbalanced Sentiment Classification

In this subsection, we report the performance of our semi-supervised method for imbalanced sentiment classification. Since the unlabeled data are imbalanced while the selected samples in each iteration are balanced, we can hardly perform co-training until all unlabeled data are automatically labeled (there are not enough balanced data for

[2] This dataset is freely available for research purpose. For details, please contact the corresponding author.

[3] http://mallet.cs.umass.edu/

[4] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

selection). Instead, we only run 50 iterations and record the corresponding results. For comparison, we also give the results of two baselines. One is to use only one set of balanced initial training data and then perform co-training with two random subspaces. This method is denoted as coTraining-underSampling in Figure 3. The other is to use all sets but only use fixed subspaces (static subspace generation method). This method is denoted as coTraining-Static in Figure 3.

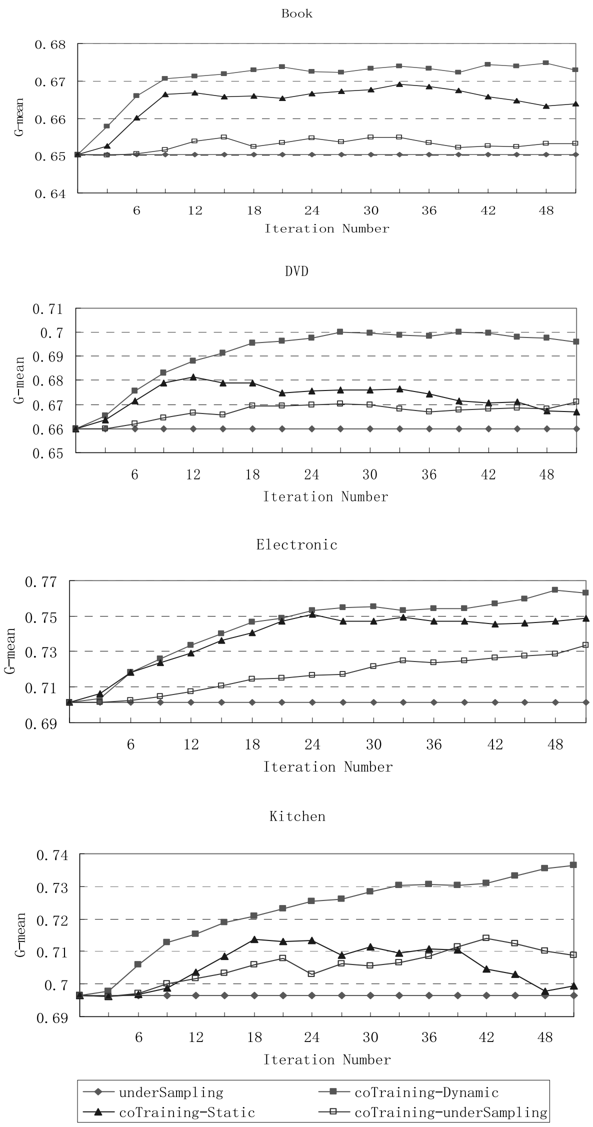

Figure 3: Performances of different semi-supervised methods on imbalanced sentiment classification

In Figure 3, we can see that our approach (coTraining-Dynamic) robustly outperforms under-sampling and the two baselines in all the four domains. Since these methods involve random selection of samples, we run 10 times for them and find that the improvements over under-sampling and the two baselines are significant with paired t-test (at the 0.05 level).

Comparison with the State-of-the-Art

For comparison, we implement a state-of-the-art semi-supervised method for sentiment classification, as proposed by Li et al. [2010], which employs personal and impersonal views in co-training. In particular, we employ it in static subspace generation. Figure 4 illustrates the superiority of our semi-supervised method due to its effectively adapting to semi-supervised learning on imbalanced data.

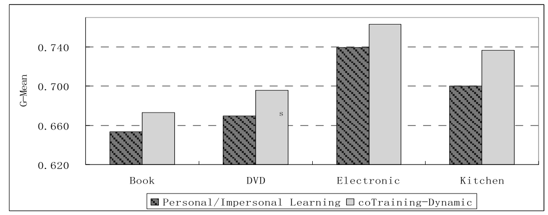

Figure 4: Comparison with the State-of-the-Art [Li et al. 2010]

6 Conclusion

In this paper, we address semi-supervised learning for imbalanced sentiment classification. We first adopt under-sampling to generate multiple sets of balanced initial training data and then propose a novel semi-supervised learning method based on random subspace generation which dynamically generates various subspaces in the iteration process to guarantee enough variation among the involved classifiers. Evaluation shows that our semi-supervised method can successfully make use of the unlabeled data and that dynamic subspace generation significantly outperforms traditional static subspace generation.

To the best of our knowledge, this is the first work that systematically addresses the imbalanced class distribution problem in sentiment classification, especially under semi-supervised learning.

In this study, we only focus on the imbalanced problem in the initial labeled data. Questions as to how to build a stop criterion in controlling the iteration number in co-training and how to make full use of the unlabeled data in the *majority class* still remain great challenges in our future study.

Acknowledgments

The research work described in this paper has been partially supported by three NSFC grants, No. 61003155, No. 60873150 and No. 90920004 and Open Projects Program of National Laboratory of Pattern Recognition.

References

[Balcan et al., 2005] M. Balcan, A. Blum, and K. Yang. Co-training and Expansion: Towards Bridging Theory and practice. *Advances in Neural Information Processing Systems*, vol.17, pp.89-96, 2005.

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

[Blitzer et al., 2007] J. Blitzer, M. Dredze, and F. Pereira. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. In *Proceedings of ACL-07*, pp.440-447, 2007.

[Blum and Mitchell, 1998] A. Blum and T. Mitchell. Combining Labeled and Unlabeled Data with Co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, pp. 92-100, 1998.

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

[Chawla et al., 2004] N. Chawla, N. Japkowicz, and A. Kotcz. Editorial: Special Issue on Learning from Imbalanced Data Sets. *SIGKDD Exploration Newsletter*, vol.6(1):pp.1-6, 2004.

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

[Dasgupta and Ng, 2009] S. Dasgupta and V. Ng. Mine the Easy and Classify the Hard: Experiments with Automatic Sentiment Classification. In *Proceedings of ACL-IJCNLP-09*, pp.701-709, 2009.

[Guo and Viktor, 2004] H. Guo and H. Viktor. 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.

[He and Garcia, 2009] H. He and E. Garcia. Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering In Knowledge and Data Engineering*, vol.21(9), pp.1263-1284, 2009.

[Ho, 1998] T. Ho. The Random Subspace Method for Constructing Decision Forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20(8), pp.832-844, 1998.

[Japkowicz and Stephen, 2001] N. Japkowicz and S. Stephen. The Class Imbalance Problem: A Systematic Study. *Intelligent Data Analysis*, vol.6, pp.429–450, 2001.

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

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

[Li et al., 2009] T. Li, Y. Zhang, and V. Sindhwani. A Non-negative Matrix Tri-factorization Approach to Sentiment Classification with Lexical Prior Knowledge. In *Proceeding of ACL-IJCNLP-09,* pp.244-252, 2009.

[Li et al., 2010] S. Li, C. Huang, G. Zhou, and S. Lee. Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification. In *Proceedings of ACL-10*, pp.414-423, 2010.

[Liu et al., 2005] B. Liu, M. Hu, and J. Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proceedings of WWW-05*, pp.342-351, 2005.

[Liu et al., 2009] X. Liu, J. Wu, and Z. Zhou. Exploratory under-sampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39(2), pp.539-550, 2009.

[Lloret et al., 2009] E. Lloret, 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, 2009.

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

[Sindhwani and Melville, 2008] V. Sindhwani and P. Melville. Document-Word Co-regularization for Semi-supervised Sentiment Analysis. In *Proceedings of ICDM-08*, pp.1025-1030, 2008.

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

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

[Wang et al., 2008] J. Wang, S. Luo, and X. Zeng. A Random Subspace Method for Co-training. In *Proceedings of IJCNN-08*, pp.195–200, 2008.

[Wiebe et al., 2005] J. Wiebe, T. Wilson, and C. Cardie. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, vol.39(2/3), 2005.

[Yarowsky, 1995] D. Yarowsky. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In *Proceedings of ACL-05*, pp.189-196, 1995.

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

[Zhou and Liu, 2006] Z. Zhou and X. Liu. 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.