Improving Sentiment Analysis Through Ensemble Learning of Meta-level Features

Rana Alnashwan, Adrian O'Riordan, Humphrey Sorensen, and Cathal Hoare

Computer Science, Western Gateway Building,
University College Cork, Cork, Ireland

{r.alnashwan, a.oriordan, sorensen, hoare}@cs.ucc.ie

Abstract. In this research, the well-known microblogging site, Twitter, was used for a sentiment analysis investigation. We propose an ensemble learning approach based on the meta-level features of seven existing lexicon resources for automated polarity sentiment classification. The ensemble employs four base learners (a Two-Class Support Vector Machine, a Two-Class Bayes Point Machine, a Two-Class Logistic Regression and a Two-Class Decision Forest) for the classification task. Three different labelled Twitter datasets were used to evaluate the effectiveness of this approach to sentiment analysis. Our experiment shows that, based on a combination of existing lexicon resources, the ensemble learners minimize the error rate by avoiding poor selection from stand-alone classifiers.

Keywords: Opinion Mining, Sentiment Analysis, Lexicon, Machine Learning, Twitter.

1 Introduction

Today, the vast amount of data available online can have considerable value for society when they are assessed as part of opinion mining analyses. Therefore, finding the right techniques and models for the sentiment analysis of big data has become a crucial activity in order to obtain greater value from the data available. The objective of the study is to maximize the potential of these kinds of data on the Internet, as sentiment can be analyzed in order to ascertain trends and inform decisions on various subjects.

Some researchers use meta-level features while others use ensemble learning, but not in combination. The main contribution of this paper is in investigating the effectiveness of using a combination of existing lexicon resources as meta-level features in ensemble learning for sentiment classification. This offers advantages over using either a single lexicon resource or a single classifier.

The remainder of this paper is structured as follows: section 2 surveys approaches to sentiment classification that relate to our work; section 3 describes our classification approach, for which the experimental test and results are provided in section 4; section 5 addresses the conclusion and potential extension of the work.

2 Related Work

2.1 Twitter Sentiment Analysis

Microblogging web services have now become an important source for gathering a variety of information for sentiment analysis [21]. This is due to the nature of these services, whereby people can communicate with others by sharing their opinions, publicizing their status, joining with other people who have similar interests, making online friends, expressing political or religious views, and providing positive and negative reactions to a variety of topics [1], [12]. Twitter is the most popular microblogging service and has shown significant growth since it was launched in October 2006 [12]. Microblogging, and, more particularly, Twitter, is a valuable source for sentiment analysis, as a large number of tweets contain sentiment information [21], [27]. Twitter is a challenging platform for analysts because it contains informal text and it is hard to trace specific events, as people can post about anything and everything [15]. Another reason is that what people discuss online is very different from what is found in, for example, newspapers [17]. An increasing number of opinion-mining researchers are focusing their attention on tweet sentiment analysis, a sample of which is shown in Table 1.

2.2 Sentiment Analysis Methods

Recently, a number of approaches, techniques and methods have been applied across different tasks to address the sentiment analysis classification problem. According to Wang et al. [27], sentiment analysis relies on two main methods: natural language processing techniques and machine learning approaches.

There has been much work on natural language processing techniques to identify sentiment analysis for texts. For example, Deng and Liu [8] find opinions in product reviews using linguistic rules, whereas Nasukawa and Yi [19] focus their research on syntactic parsing and sentiment lexicons. Although rule-based methods for identifying sentiment polarity and targets are effective, the major drawbacks are that they cannot be extended without expert knowledge and the coverage of the rules is not satisfactory [27]. Wang et al. [27] compare machine learning and rule-based methods and assert that machine learning approaches usually score higher for recall due to the strong generalization ability of classifiers. Moreover, Pang et al. [22] show that machine learning approaches have a good level of accuracy, about 83% having greater accuracy than the human-generated baseline in their results.

Researchers have applied stand-alone supervised machine learning and/or hybrid classification approaches for tweet sentiment analysis, as presented in the summary in Table 1. For instance, Khan et al. [13] apply the use of a hybrid scheme using first an Enhanced Emoticon Classifier (EEC), second an Improved Polarity Classifier (IPC), and third SentiWordNet Classifier (SWNC) methods. In a similar context, Balage Filho et al. [3] apply a hybrid classification approach that has two emoticon lexicons as their rule-based classifier and SentiStrength as a lexicon-based classifier. The third classifier is a Support Vector Machine (SVM), a machine learning classifier. The study assigns

a confidence threshold in each of the classifiers to achieve the overall confidence level required.

Approaches that integrate sentiment lexicon resources as features in supervised classification schemes have also been studied. For example, in Kouloumpis et al. [15], the authors propose a supervised method using different feature sets: word n-gram, part of speech (POS) and a lexicon. Another study [4] combines 13 existing sentiment analysis methods and resources as a feature set in a supervised classifier focused on different aspects, such as polarity, strength and emotion. Sentiment analysis was performed using three different machine learning algorithms. The result shows that a lexicon-based approach is best for polarity classification, while part-of-speech approaches are more suitable for subjectivity classification.

The idea of combining multiple supervised learners to obtain predictive classification has also been explored by the research community. Ensemble learning is a relatively new approach that is gaining the attention of research in sentiment classification tasks. Thelwall et al. [24] apply an ensemble learning algorithm approach, stacked generalization, to sentiment classification. They achieved good results by employing five different supervised learning techniques to three different domains. Clark and Wicentwoski [5] apply another ensemble learning approach – a combination of multiple Naïve Bayes classifiers, whereby each has a single feature, e.g. n-gram, sentiment lexicon, part of speech, emoticons and assigning weight to words that have repeated letters. Wang et al. [26] apply three different ensemble approaches, namely, bagging, boosting and random subspace; five supervised learning algorithms were used as base classifiers with a bag-of-words feature.

3 Classification Approach

In this section, we describe the sentiment classification approach that has been used in this research. We focus on polarity: a binary classification of positive or negative.

Our proposed approach relies on two models for sentiment classification, as shown in Figure 1. First, a set of combinations of sentiment analysis methods and lexicons forms a feature vector for each tweet. Second, an ensemble method uses a supervised approach.

3.1 Tweet Sentiment Representation

Feature Hashing. The model used in feature representation was feature hashing with n-grams. The feature-hashing model converts streams of words into a set of integer features and vectors thereof, by creating a hashing dictionary that consists of n-gram features calculated using the terms repeated in the text. One advantage of using feature hashing is that it reduces the dimensional space for the supervised learning machine by representing text documents as numeric feature vectors. The feature hashing is set to a bitsize of 10 in hashing each n-gram.

Table 1. Selected previous studies in sentiment analysis

Classification with lexicon and/or learning algorithms						
Study	**Year**	**Classification type**	**Feature set**	**Lexicon**	**Classifier**	**Dataset**
[4]	2014	Subjectivity and polarity	Meta-level features and part-of- speech features	-	Naïve Bayes, Logistic, Perceptron and Support Vector Machine	Stanford Twitter Sentiment (STS), Sanders and SemEval
[11]	2016	Positive, negative and neutral	Combining popular "off-the-shelf" sentiment analysis methods	-	Support Vector Machine and Random Forests	Tweets, movie and product reviews, opinions and comments in news articles
[13]	2014	Positive, negative and neutral	POS	Emoticons, Bing Liu, Bill McDonald, Senti-WordNet	-	Twitter streaming API
[3]	2013	Positive, negative and neutral	POS	Emoticons, SentiS-trength lexicon	Support Vector Machine	SemEval Twitter dataset
[15]	2011	3-way classification	n-gram, MPQA subjectivity lexicon, part-of-speech features and emoticons	-	AdaBoost.MH Algorithim	Hashtagged Twitter dataset (HASH), the emoticon Twitter dataset (EMOT) and iSieve Corporation (ISIEVE)

Ensemble learning				
Study	Year	Base learner	Ensemble method	Dataset
[24]	2013	Naïve Bayes, centroid-based classification, K Nearest Neighbor, Maximum Entropy model and Support Vector Machines	Stacked generalization	Book reviews, hotel reviews and notebook reviews
[5]	2013	Naïve Bayes	Confidence-weighted voting scheme	SemEval Twitter dataset
[26]	2014	Naïve Bayes, Maximum Entropy, Decision Tree, K Nearest Neighbor and Support Vector Machine	Bagging, boosting, and random subspace	10 different reviews
[6]	2014	Logistic Regression, Random Forest, Support Vector Machine and Multinomial Naïve Bayes	Majority voting	Sanders, Stanford Twitter Sentiment Corpus, Obama-McCain Debate (OMD), Health care reform (HCR)

Meta-level Features. Meta-level features are output for each method and lexicon resource for sentiment analysis. These resources and methods ascertain the polarity of each tweet. The number of features in each lexicon resource can be calculated by finding the matching words in the text and the lexicon resources. The results of adding these values are then represented as a feature vector. These features are summarized in Table 2.

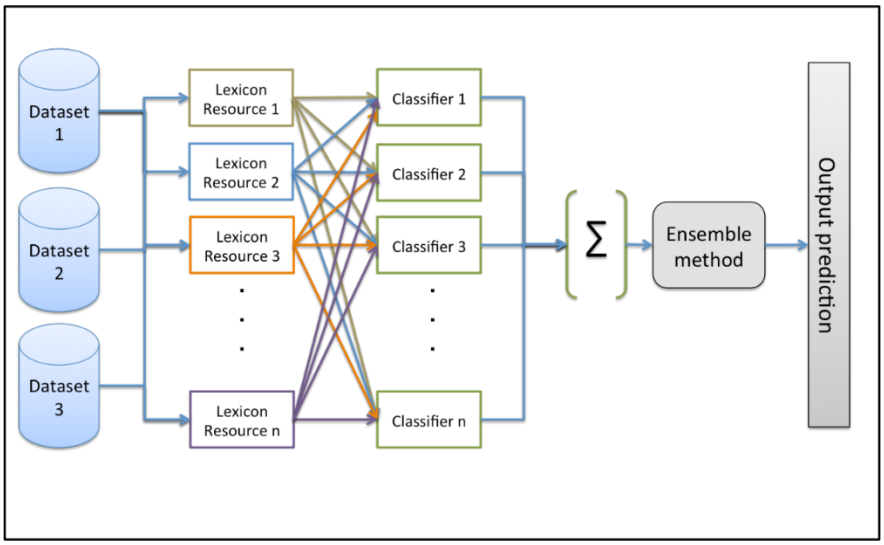

Fig. 1. Schematic of the approach used in this research

SentiWordNet. SentiWordNet 3.0 is a lexicon source for sentiment classification and opinion mining developed by Baccianella et al. [2] and is an improved version of SentiWordNet 1.0, proposed by Esuli and Sebastiani [9]. SentiWordNet 3.0 is based on WordNet 3.0, which classifies all part-of-speech into groups of synonyms, which are named synsets. SentiWordNet annotates all synsets with a value between 1 and 0 to indicate the positivity, negativity or neutrality of each synset. This lexicon was developed using semi-supervised classification and a random walk process [2]. The lexicon is freely available to researchers.

We extracted two features from the SentiWordNet lexicon: the positive value and negative value.

Bing Liu Lexicon. We employed Bing Liu's lexicon resource [16], which includes misspelled words, slang and some morphological variants. The lexicon has 2,006 positive and 4,783 negative words.

The positive and negative features were extracted from each tweet that matched Bing Liu's lexicon.

AFINN. AFINN-111 is an improved version of AFINN-96. The original version was called ANEW (Affective Norms for English Words) and was developed before the

widespread use of microblogging platforms [20]. It was generated using people's psychological reactions.

We extracted two features, positivity and negativity, corresponding to the rating values of all words in tweets that matched the AFINN-111 lexicon.

NRC-Hashtag. The NRC-hashtag sentiment lexicon resource was proposed by Muhammad et al. [18]. The lexicon was created by adopting the use of hashtags of emoticon words, such as #angry, #joy and #sadness in tweets [18].

Using this lexicon, we extracted positive and negative features by matching words in the NRC-hashtag lexicon with tweets and then adding those values.

Sentiment140 Lexicon. Sentiment140 and NRC-hashtag were created by the same group [18] and have the same format. Sentiment140 focuses on emoticon labels instead of hashtags to indicate positive or negative. The researchers used 1.6 million tweets to develop this lexicon.

We extracted the feature, as we did with NRC-hashtags.

Sentiment140 Method.[1] The Sentiment140 method is an Application Program Interface (API) for assigning tweets their polarity. It was generated by Nielsen [20], who used a supervised learning technique on 1.6 million tweets, the same corpus as the Sentiment140 lexicon. Emoticons in tweets and noisy data were considered for sentiment analysis classification.

One feature was extracted from the Sentiment140 method: one output value for each tweet, in contrast with the Sentiment140 lexicon.

SentiStrength. SentiStrength is a web application for automatic sentiment analysis that evaluates the strength of sentiment in short texts [25]. It uses supervised and unsupervised learning methods.

We extracted three features from the SentiStrength resource: positive, negative and polarity features.

Table 2. Features of the lexicon resource

Lexicon source	Number of features extracted	Range of values
SentiWordNet	2 (positive and negative)	{0, …., 1}
Bing Liu	2 (positive and negative)	{0,1}
AFINN	2 (positive and negative)	{-5, …… ,5}
NRC-hashtag	2 (positive and negative)	{-∞, ….. , ∞}
Sentiment140 lexicon	2 (positive and negative)	{-∞, ….. , ∞}
Sentiment140 method	1 (method output)	{0,2,4}
SentiStrength	3 (positive, negative and polarity)	{-1,1}

[1] http://www.sentiment140.com/

Pre-processing. As a result of the characteristics of the language used on Twitter, some pre-processing steps were required in order to reduce the dimensionality of the feature space. The first step involves removing links, punctuation, special characters and digits and replacing them with white space. Then, all capital letters are converted to lower case to unify the data format. Finally, letters that are repeated more than twice in sequence are reduced to a sequence of two, as reducing them to one would lead to error. For example, "greeeeeeat" or "greeeat" is converted to "greeat".

3.2 Classifier Ensemble for Tweet Sentiment Analysis

Ensemble learning is a technique in machine learning that trains multiple learners to solve the same problem [28]. The multiple learners that are employed in an ensemble are called base learners [28]. According to Dietterich [7], there are three significant reasons for using an ensemble base: 1) statistical: when the result relies on a combination of classifiers, this can reduce the chance of selecting the wrong classification; 2) computational: some of the learning algorithms are based on a local search, where it is possible to become stuck in local optima — by applying an ensemble, the running of the local search can start with a number of different classifiers that can achieve better approximations than any single classifier; and 3) representational: when the hypothesis space does not present an appropriate target function, whereas an ensemble can expand that space to give a better approximation.

Employing an ensemble approach will not always guarantee a better result than the best base learner [6]. However, using a combination of classifiers will decrease the error rate from selecting a poor classifier by outperforming random selection [6].

In our experiments, we focused on supervised learning approaches in which tweets and all the extracted features described previously were fed in as vectors of sentiment features. In order to evaluate the proposed approach, labelled data are needed to train the model and evaluate its performance. We concentrated on a polarity prediction task, task, identifying whether the features were positive or negative.

Four classifiers were used as our base learners to fulfill the sentiment classification task. The classifiers were: a Two-Class SVM, a Two-Class Bayes Point Machine, Two-Class Logistic Regression, and a Two-Class Decision Forest. These classifiers were selected because they have been widely used in previous sentiment analysis research, as well as having high performance and diversity compared with other classifiers. In our experiment, classifiers were built and trained to predict unseen data (the test data). To obtain an effective ensemble, two elements should be considered: the diversity and accuracy of each classifier [14], [28]. The different decision boundaries of base learners lead to uncorrelated errors. The ensemble approach should outperform a random selection of base learners.

After the base learners were trained, our ensembles were developed by a majority voting method, which is one of the most common ensemble methods in classification tasks [28].

4 Experimental Evaluation

4.1 Datasets

To evaluate the effectiveness of our approach, we considered three labelled datasets.

Stanford Twitter Sentiment (STS). Stanford Twitter Sentiment was proposed by Go et al. [10]. The dataset contains 1.6 million tweets that are automatically labelled to positive and negative according to emoticon. We randomly selected 12,000 tweets.

SemEval-2016. This dataset was provided by the Semantic Evaluation of Systems (SemEval-2016) challenge. This involved the undertaking of challenging tasks by researchers who are interested in semantic analysis problems. Each tweet was annotated manually to positive, negative or neutral.

Health Care Reform (HCR). A health care reform (HCR) labelled dataset was created by Speriosu et al. [23]. It was collected from extracted tweets that had the hashtag "#hcr". The authors then annotated a subset of collected data for the polarity classes.

Table 3. Dataset statistics

	Positive	Negative	Total
STS	5,999	6,001	12,000
SemEval	4,385	1,415	5,800
HCR	542	1,380	1,922

4.2 Experimental Setup

We conducted our experiment using Microsoft Azure, an integrated cloud service. In practice, we used the Azure machine learning cloud computing platform to run the Two-Class SVM, the Two-Class Bayes Point Machine, Two-Class Logistic Regression and the Two-Class Decision Forest.

In some of our data, as shown in Table 3, the number of positive and negative tweets was unbalanced, so we performed resampling with replacements to avoid biasing the classifiers towards one specific class.

4.3 Results

The analysis was carried out in three phases: 1) constructing the ensemble, 2) applying meta-level features to each classifier, and 3) combining them using meta-level features on each base learner for the ensemble. We compared the results from the three approaches with stand-alone classifiers with feature hashing, which was set as our baseline. We evaluated the model that combined the meta-level approach with the ensemble

approach to address the potential for improving performance. We evaluated the approaches using STS, SemEval-2016 and HCR for the polarity classification task.

The results of the polarity classification tasks are shown in Table 4. The best baseline classifier for the SemEval and HCR datasets is the Two-Class Decision Forest, whereas, for the Stanford dataset, it is Two-Class Logistic Regression. The meta-level approach outperformed the baseline by just over 5% in accuracy and by F measurement and average in both the Stanford and SemEval datasets. The meta-level improvements are fewer in the HCR dataset, which indicates that the HCR dataset is easier to classify than Stanford or SemEval. We also observed in the meta-level approach that the Two-Class Decision Forest scored best in both the SemEval and HCR datasets. However, Two-Class Logistic Regression was best with the Stanford dataset.

According to the outcomes, the ensemble with meta-level features shows better results compared with the original ensemble. The ensembles in both cases scored better for accuracy than the average of the base learners. Thus, applying our approach could avoid the common classification task problem of the poor selection of classifier. Furthermore, by considering the average of the polarity tasks, we observe that there is no significant difference between the best classifier, the Two-Class Decision Forest, and the proposed ensemble approach, which scored 82.4% and 81.0%, respectively.

5 Conclusions and Future Work

We conducted a series of experiments on sentiment classification in social media text using ensemble learning methods. Each base learner in the ensemble used meta-level feature extraction. The features covered a combination of several existing lexicon and method resources for sentiment analysis. Moreover, feature hashing was used in the representation of tweets. The experiments investigated three datasets to verify the effectiveness of the present approach across different data. Our experiment results show that such ensemble classifiers can minimize the error rate by avoiding poor selection from the stand-alone classifiers, which is an effective way of ensuring stability. In addition, using the meta-level feature mitigated problems associated with the sparsity of the data. In that context, the meta-level ensemble approach can achieve promising results.

We believe that our approach can be relevant to other social media analysis and any other classifier could easily be integrated into the proposed framework. As for future work, the classification task could be expanded by considering neutral text. In addition, the proposed approach can also be expanded by evaluating different ensemble methods (voting schemes) or by considering other lexicon resources and methods in sentiment analysis to boost classifier performance.

Table 4 Polarity classification performance

		Stanford dataset					SemEval dataset					HCR				
		Accuracy	Precision	Recall	F	Average	Accuracy	Precision	Recall	F	Average	Accuracy	Precision	Recall	F	Average
Baseline	Two-Class Support Vector Machine	0.643	0.629	0.668	0.648	0.647	0.729	0.737	0.730	0.733	0.732	0.721	0.654	0.705	0.678	0.690
	Two-Class Bayes Point Machine	0.659	0.653	0.656	0.654	0.656	0.759	0.767	0.757	0.762	0.761	0.788	0.724	0.792	0.757	0.765
	Two-Class Logistic Regression	0.667	0.654	0.682	0.668	0.668	0.758	0.762	0.765	0.763	0.762	0.759	0.695	0.753	0.723	0.733
	Two-Class Decision forest	0.630	0.617	0.652	0.634	0.633	0.852	0.891	0.808	0.847	0.850	0.820	0.748	0.857	0.799	0.806
	Average	0.650	0.638	0.665	0.651	0.651	0.775	0.789	0.765	0.776	0.776	0.772	0.705	0.777	0.739	0.749
Ensemble	Ensemble classifiers (majority voting)	0.660	0.648	0.676	0.662	0.662	0.802	0.818	0.786	0.802	0.802	0.812	0.754	0.815	0.783	0.791
Meta-level	Two-Class Support Vector Machine	0.766	0.758	0.769	0.763	0.764	0.816	0.821	0.817	0.819	0.818	0.758	0.697	0.740	0.718	0.728
	Two-Class Bayes Point Machine	0.762	0.750	0.775	0.762	0.762	0.828	0.837	0.824	0.831	0.830	0.811	0.750	0.818	0.783	0.791
	Two-Class Logistic Regression	**0.791**	**0.784**	**0.794**	**0.789**	**0.790**	0.843	0.848	0.845	0.846	0.846	0.790	0.733	0.782	0.757	0.766
	Two-Class Decision Forest	0.771	0.766	0.771	0.768	0.769	**0.913**	**0.951**	**0.873**	**0.911**	**0.912**	**0.834**	0.755	**0.890**	**0.817**	**0.824**
	Average	0.773	0.765	0.777	0.771	0.771	0.850	0.864	0.840	0.852	0.852	0.798	0.734	0.808	0.769	0.777
Combining	Ensemble classifiers (majority voting)	0.783	0.777	0.784	0.780	0.781	0.873	0.889	0.858	0.873	0.873	0.832	**0.788**	0.818	0.803	0.810

References

1. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment Analysis of Twitter Data. In: Proceedings of the Workshop on Languages in Social Media, 30–38. ACL (2011).
2. Baccianella, S.,Esuli, A., Sebastiani, F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. LREC, 10, 2200–2204 (2010).
3. Balage Filho, P.P., Pardo, T.A.S.: NILC_USP: A Hybrid System for Sentiment Analysis in Twitter Messages. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), 2, 568–572 (2013).
4. Bravo-Marquez, F., Mendoza, M., Poblete, B.: Meta-level Sentiment Models for Big Social Data Analysis. Knowledge-Based Systems, 69, 86–99 (2014).
5. Clark, S., Wicentwoski, R.: SwatCS: Combining Simple Classifiers with Estimated Accuracy. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), 2, 425–429 (2013).
6. da Silva, N.F.F., Hruschka, E.R., Hruschka, E.R.: Tweet Sentiment Analysis with Classifier Ensembles. Decision Support Systems, 66, 170–179 (2014).
7. Dietterich, T.G.: Ensemble Methods in Machine Learning. In: Multiple Classifier Systems, pp. 1-15. Springer, Berlin Heidelberg (2000).
8. Ding, X., Liu, B.: The Utility of Linguistic Rules in Opinion Mining. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 811–812, ACM (2007).
9. Esuli, A., Sebastiani, F.: SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In: Proceedings of LREC, 6, 417–422 (2006).
10. Go, A., Bhayani, R., Huang, L.: Twitter Sentiment Classification Using Distant Supervision. CS224N Project Report, Stanford, 1, 12 (2009).
11. Gonçalves, P., Dalip, D.H., Costa, H., Gonçalves, M.A., Benevenuto, F.: On the Combination of "Off-the-Shelf" Sentiment Analysis Methods. SAC, 1158–1165 (2016).
12. Java, A., Song, X., Finin, T., Tseng, B.: Why We Twitter: Understanding Microblogging Usage and Communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, 56–65, ACM (2007).
13. Khan, F.H., Bashir, S., Qamar, U.: TOM: Twitter Opinion Mining Framework Using Hybrid Classification Scheme. Decision Support Systems, 57, 245–257 (2014).
14. Ko, A.H.-R., Sabourin, R., de Souza Britt, A.: Combining Diversity and Classification Accuracy for Ensemble Selection in Random Subspaces. In: Neural Networks, 2006. IJCNN'06. International Joint Conference on, 2144–2151, IEEE (2006).
15. Kouloumpis, E., Wilson, T., Moore, J.D.: Twitter Sentiment Analysis: The Good the Bad and the OMG! ICWSM, 11, 538–541 (2011).
16. Liu, B. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5, 1–167 (2012).
17. Lloyd, L., Kaulgud, P., Skiena, S.: Newspapers vs. Blogs: Who Gets the Scoop? In: AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, 117–124 (2006).
18. Mohammad, S.M., Kiritchenko, S., Zhu, X.: NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. arXiv preprint arXiv:1308.6242 (2013).
19. Nasukawa, T., Yi, J.: Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In: Proceedings of the 2nd International Conference on Knowledge Capture, 70–77, ACM (2003).
20. Nielsen, F.Å.: A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. arXiv preprint arXiv:1103.2903 (2011).

21. Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. LREC, 10, 1320–1326 (2010).
22. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, 79–86, ACL (2002).
23. Speriosu, M., Sudan, N., Upadhyay, S., Baldridge, J.: Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph. In: Proceedings of the First workshop on Unsupervised Learning in NLP, 53–63, ACL (2011).
24. Su, Y., Zhang, Y., Ji, D., Wang, Y., Wu., H.: Ensemble Learning for Sentiment Classification. In: Chinese Lexical Semantics, pp. 84–93. Springer, Berlin Heidelberg (2012).
25. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment Strength Detection for the Social Web. Assoc. Inf. Sci. Technol. 63, 163–173 (2012).
26. Wang, G., Sun, J., Ma, J., Xu, K., Gu, J.: Sentiment Classification: The Contribution of Ensemble Learning. Decis. Support Syst. 57, 77–93 (2014).
27. Wang, X., Wei, F., Liu, X., Zhou, M., Zhang, M.: Topic Sentiment Analysis in Twitter: A Graph-Based Hashtag Sentiment Classification Approach. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, 1031–1040, ACM (2011).
28. Zhou, Z.H.: Ensemble Methods: Foundations and Algorithms. CRC Press, Boca Raton (2012).