An Ensemble Model using a BabelNet Enriched Document Space for Twitter Sentiment Classification

Semih Sevim
Kocaeli University Computer Engineering Departmen, Kocaeli, TURKEY
E-mail: semih.sevimm@gmail.com

Sevinç İlhan Omurca and Ekin Ekinci
Kocaeli University Computer Engineering Departmen, Kocaeli, TURKEY
E-mail: silhan@kocaeli.edu.tr, ekin.ekinci@kocaeli.edu.tr

Abstract—With the widespread usage of social media in our daily lives, user reviews emerged as an impactful factor for numerous fields including understanding consumer attitudes, determining political tendency, revealing strengths or weaknesses of many different organizations. Today, people are chatting with their friends, carrying out social relations, shopping and following many current events through the social media. However social media limits the size of user messages. The users generally express their opinions by using emoticons, abbreviations, slangs, and symbols instead of words. This situation makes the sentiment classification of social media texts more complex. In this paper a sentiment classification model for Twitter messages is proposed to overcome this difficulty. In the proposed model first the short messages are expanded with BabelNet which is a concept network. Then the expanded and the original form of the messages are included in an ensemble learning model. Consequently we compared our ensemble model with traditional classification algorithms and observed that the F-measure value is increased.

Index Terms—Twitter sentiment classification, ensemble learning, Semantic enrichment, BabelNet.

I. INTRODUCTION

Today, people mostly prefer to communicate and socialize on the Web. As a result of this, social media websites such as Facebook, Twitter, LinkedIn, Youtube, Instagram and so on, have exploded in popularity with tremendous impact on society in several ways. For example, in many elections around the world social media has played an important role. Many of the companies measure their customer satisfaction via social media postings; therefore, adjust the advertisement of their products to develop a better sales strategy. Apart from these, social media websites can be used to improve the people's career and business prospects. In summary, social media is an important and major source of information to enhance the strategies range from the business and health to politics and the education system in the competitive world. Thus, to obtain the view of public from these mediums there is a need for an automated analysis system.

The sentiment classification task is extremely useful in social media analysis as it allows us to gain the common opinion behind social media messages and is broadly studied in the literature [1]. Indeed, sentiment classification is the task of learning positive or negative opinion have been expressed on certain product, organization, event or individual from opinionated documents and required in-depth analysis.

In recent years, sentiment classification from social networks has become necessary due to containing huge amount of opinionated documents. In social networks, users usually write short-texts which are also rife with emoticons, abbreviations, slangs and symbols to express their opinions. The main characteristic of a short text is the text length which is no longer than 200 characters. The sparsity and un-standardability are the two main drawbacks of short text messages. If a text sample is short, it means that the features of it are very sparse and also these limited features frequently contain non-standard terms, noise terms and abbreviations [2]. With such a text compilation, sentiment classification is becoming a more complex problem because short text examples naturally do not provide sufficient statistics or associations for classifiers. In the other hand, text classification methods generally use word statistics and associations to provide a judgment about which terms are similar or fall into which classes. In this case, short-text classification problem becomes a crucial research area in social media analysis [3].

To address the above mentioned issues, several approaches for short text classification are presented in the literature. Faguo et al. [2] proposed short text classification based on statistics and rules. Zhou et al. [4] devised a hybrid model of character-level and word-level

features based on recurrent neural network with long short-term memory in order to reduce the impact of word segmentation and improve the overall performance of Chinese short text classification. Wang et al. [5] used better feature space selection to overcome sparseness problem of the short text classification. Sriram et al. [6] expanded the features of twitter texts by extracting additional author information to classify tweets with a better accuracy. For short texts in Chinese, Wensen et al. [7] used Wikipedia concepts and Word2vec to expand the features. Chen et al. [8] used LDA and K-Nearest Neighbor algorithm to improve short text classification. Sang et al. [9] expanded short texts using word embedding for classification.

Original social media messages can be preprocessed and expanded by several ways which are presented and applied as above mentioned. After these initial processes the social media messages become inputs to machine learning algorithms. Among the machine learning methods, ensemble learning allow us not only to combine multiple methods both for the preprocessing and the classification stages but also to achieve a better hypothesis. The underlying intuition of ensemble learning is to build a strong learner from a bunch of weak learner. An ensemble learning system is proposed to perform sentiment classification for twitter messages which are limited by 140 characters in length. With these kinds of short messages, language is evolving faster than ever before. In text mining systems, extracting features from text, figuring out which features are relevant, and selecting an accurate classification algorithm are fundamental questions [10].So the accurate representation of short text has become an important step of any text mining task for social media.

In this paper, the first key point of the proposed ensemble learning model is the representation of short text messages and the second one is to decide a suitable classifier. For the first key point, three subsets of the original datasets are established to construct the extended feature space. The first one is constituted only by the raw data, the second one is constituted by adding the concepts obtained from Babelnet to the raw data. The third one is constituted by the concepts obtained from Babelnet and only the extended terms by Babelnet. For the second key point, we focus on the same base learners to provide diversity which is called homogeneous classifier ensembles. The well established classification algorithms such as Naïve Bayes (NB), Support vector machines (SVM) and Decision tree (DT) are selected respectively for the weak learners of our three different ensemble learning models. The experimental results which are obtained by three public twitter datasets demonstrate that our proposed method outperforms existing classification methods in literature.

The remainder of the paper is organized as follows. In section 2, related work is summarized. In section 3, applied text preprocessing and expansion methods are defined. In section 4, methodological background of classification algorithms; in section 5, proposed ensemble model are given. In section 6 experimental results are reported and finally in section 7 we summarize the main points of our work and discuss the contributions.

II. Related Work

In recent years, analyzing opinion oriented human behavior has attracted the attention of researchers in artificial intelligence and natural language processing. In the literature, there are many studies on the design of methods for sentiment classification of text; however, the numbers of studies which use ensemble of classifiers are limited. Akhilesh at. al. [11] used Naive Bayes method to analyze Assembly Election on twitter users data. Indrajit at. al. [12] classified tweets into three categories, news, sport and movies by using topic modelling which is an unsupervised machine learning method. Deebha at. al. [13] formed opinion mining model according to polarity of words and examined effect of negation words. Xia et al. [14] proposed an ensemble technique for sentiment classification by integrating different feature sets and classification algorithms to boost the overall performance. Li et al. [15] used a heterogeneous ensemble learning method for Chinese sentiment classification. Su et al. [16] presented an ensemble model for sentiment classification of reviews. And their experiments showed that stacking is consistently effective compared with majority voting. Filho and Pardo [17] combined rule-based, lexicon-based and machine learning approaches for twitter sentiment classification. Hassan et al. [18] proposed the bootstrapping ensemble framework (BPEF) which provided better performance for classifying Twitter sentiments. Wang et al. [19] compared the performance of bagging, boosting, and random subspace methods based on Naive Bayes, Maximum Entropy, Decision Tree, K Nearest Neighbor, and Support Vector Machine learners for sentiment classification. They showed that random subspace has the better comperative results. Fersini et al. [20] proposed a novel ensemble method based on Bayesian Model Averaging for sentiment analysis of user reviews. They showed their model outperformed the other approaches such as bagging. Chalothorn and Ellman [21] introduced an ensemble model for twitter sentiment analysis and they formed their model by the majority voting of four base learners such as Support Vector Machine, Naive Bayes, SentiStrength and Stacking. Wang et al. [19] and Xia et al. [14] showed that the ensemble methods provide an effective way to represent short twitter texts in different ways and also improve the performance of individual base learners for sentiment classification. Cotelo et al. [22] combined the structural information and textual concepts of tweets with direct combination, stacked generalization and Multiple Pipeline Stacked Generalization methods. Isidoros Perikos and Ioannis Hatzilygeroudis [23] presented a sentiment analysis system for automatic recognition of emotions in text, using an ensemble model based on a Naïve Bayes, a Maximum Entropy learner and a knowledge-based tool. Corrêa Jr. et al. [24] proposed a tweet classification system based on ensemble of different text representations to obtain better

classification performance and used Support Vector Machines and Logistic regression as base learners.

III. Feature Engineering

A. Text Preprocessing

To perform short text classification, firstly some preprocessing steps should be carried out. The most prominent characteristic of short text classification systems, such as sentiment classification in Twitter, is the shortness of the sample texts. Besides, symbols, emoticons, character repetitions, hashtags, letters in capital and abbreviations have very common usage in these kinds of texts. In this case, a lexical normalization step is needed as one of the preprocessing steps. The lexical normalization process translates nonstandard messages to Standard English and for this aim we have used a text normalization dataset generated by the Computer Science Department, The University of Texas at Dallas in this study[25]. While an original tweet in Hobbit dataset is like "my feet r killin n im dead but the hobbit was soooo good i love !!"; after the lexical normalization this tweet is become like "my foot be kill and in dead but the hobbit be so good love". In another normalization example a sample tweet in StsGold which is like "@wale I'm gonna havta temp stop fllwing u while ur talkin abt kobe bc I loveeeeeeee him & I'm taking it personal and I like lebron 2." is converted to "i going to have to temporary stop flow you while you talk about kobe because i love he camp am take it personal and like lebron to".

In our study, we have removed emoticons from the tweet messages which are regarded as noisy labels as another preprocessing step. We have also removed urls, user mentions, retweet markers, as well as punctuations and digits which are not meaningful for our classification task. Then, all letters in capital were converted to lowercase letters. Consequently an input text for our classification system rifes with hashtags and preprocessed body text only. After these steps, the stemming is applied to body text by using Stanford NLP tools [26].

B. Concepts Generation with BabelNet

For a more accurate classification system the restricted feature space of the sample input messages are extended by using BabelNet repository which is the largest multilingual semantic network and encyclopedic dictionary composed by concepts of WordNet and Wikipedia [27]. Concepts are defined as units of knowledge, each with a unique meaning. In our proposed system, each sentence in a sample message is processed by using BabelFy API [28] and concepts from BabelNet repository are added to the obtained the feature space as new features of the related sentence. However, the stop words are kept in their original form and they are not extended by the related concepts.

For a natural language processing (NLP) system Word Sense Disambiguation (WSD) is a very challenging task in order to develop an accurate system. Expanding the

term space with the term related concepts can be a way for handling ambiguity. For example, "I am taking aspirin for my cold" and "weather is cold today" are two sentences which contain "cold" term. In the first sentence the term "cold" refers to a disease, however, in the second it refers to an environmental condition. When we try to get concepts of these two sentences from BabelFy; it gives a description like "A mild viral infection involving the nose and respiratory passages" for the cold term in the first sentence and a description like "Having a low or inadequate temperature or feeling a sensation of coldness or having been made cold" for the cold term in the second sentence. Besides, term related concepts induce the increase in classifier accuracy due to generating joint feature space for terms which have common senses.

C. Feature Weighting

Term frequency-inverse document frequency (tf-idf) weighting scheme, which is very commonly used in text mining applications, is used to exploit the distinctiveness or importance value of a term in a document collection [29]. While tf computes the frequency of each term in a specific document, idf computes the the importance of a term for document collection. Here, a document can be viewed as a matrix, $D_{ij} = |d_{ij}|_{mxn}$, n denotes the number documents and where m denotes the number of different of terms in the whole documents. d_{ij} represents the weight of term t_i in document d_j. The formula of tf-idf is shown in the equation (1):

$$d_{ij} = tf_{ij} \times idf_{ij} = tf_{ij} \times \log \frac{N}{n_i} \qquad (1)$$

Where tf_{ij} equals the number of time t_i appears in document d_j. N is the number of documents and t_i occurs in n_i documents. We use standard tf-idf to form input for machine learning algorithms, so each tweet collection is converted to a feature matrix and then classification is performed.

IV. Classification Algorithms

Classification is a supervised learning task, which is one of the major issues in text mining applications. The task of the classification is to assign new unobserved data to one of predefined classes by learning relationship between inputs and outputs (class labels) through available data [30]. By using available data, which is called as training data, a classification model is comprised and new unobserved data, which is called as test data, is used to validate this model. In literature there are many algorithms to perform classification in text mining area. However, to yield a better classification performance, it's important to use a proper algorithm. In this paper NB, SVM, and DT classifiers are used due to their ability of classifying text data.

A. Naïve Bayes

With its simplicity and high classification accuracy NB has gained great importance and is applied to many domains including sentiment classification [31]. The assumption behind NB, which is called class conditional independence, is that all words are independent from each other within each class. The assumption of class conditional independence provides high accuracy and speed when the number of features is large. Naïve Bayes uses Bayes Theorem to compute the probability of every instance for each predefined class to assign these instances to one out of a set of the predefined classes. Based on class conditional independence, the computation of probability can be presented as the following:

$$P(X \mid C_i) = \prod_{k=1}^{n} P(X_k \mid C_i) \qquad (2)$$

where $X=\{X_1, X_2, ..., X_n\}$ is feature vector composed of words and $C_1, C_2, ..., C_m$ are class labels.

B. Sequential Minimal Optimization

Sequential Minimal Optimization (SMO) is one WEKA version of Support Vector Machine (SVM), breaks the training problem into quadratic programming (QP) sub-problems by using heuristics, was suggested by Platt [32]. SMO is an iterative algorithm so chooses the smallest possible sub-problem to solve at every step. The QP could be given as follows [33]:

$$\max \sum_{i=1}^{l} \alpha_i - \sum_{i=1}^{l}\sum_{j=}^{l} \alpha_i \alpha_j y_i y_j k(X_i X_j) \qquad (3)$$

$$subject\ to\ 0 \le \alpha_i \le c,\ i = 1, ..., l \qquad (4)$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \qquad (5)$$

In the Equation (3), $k(X_i X_j)$ is the kernel function and α_i is the Lagrange multiplier; in the Equation (4), c is the regularization parameter. Equations above should satisfy Karush-Kuhn-Tucker (KKT) conditions in the Equation (6) and $Q_{ij}=y_i y_j k(X_i X_j)$ is positive semi-definite.

$$\begin{aligned} \alpha_i = 0 &\Rightarrow y_i f(X_i) \ge 1 \\ 0 < \alpha_i < c &\Rightarrow y_i f(X_i) \ge 1 \\ \alpha_i = c &\Rightarrow y_i f(X_i) \le 1 \end{aligned} \qquad (6)$$

Due to easy implementation, be faster and better scaling properties SMO has gained great attention for sentiment classification problems [34].

C. Decision Tree

Based on divide and conquer approach decision tree is defined as tree structured data structure [35]. There are different algorithms for generating decision tree such as C4.5, ID3, J48, AD Tree, BF Tree, CART and many more. In this study, Quinlan's C4.5 implementation J48 is preferred which is provided by the open-source software WEKA. Information gain calculated to select the best splitting attribute. J48 decision tree supports a variety of input data types for example nominal, textual and numeric and yields higher performance for sentiment classification tasks [36].

V. ENSEMBLE LEARNING SYSTEM

Machine learning methods have been extensively used in sentiment analysis due to their predictive performance. While the ordinary machine learning approaches that try to learn one hypothesis from the training data, ensemble methods try to construct a set of hypotheses and combine them for prediction. Ensemble learning uses a set of models which are called base learners in order to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone [19].

A labeled dataset for training, the base inducer which forms a classifier, the diversity generator which is responsible for generating diverse classifiers, and the combiner which is responsible for combining the results of the base classifiers are four building blocks that constitute a typical ensemble model [37]. Further to that, to achieve an optimal ensemble model, accuracy and diversity are two vital requirements [38]. The more diverse the base learners are, more accurate ensemble model is obtained due to diversified classifiers lead to uncorrelated classifications and improve the classification accuracy [39,40].

Methods for building ensembles can be divided into two main dimensions: how classifiers are combined and how the learning process is done [39] As for the first dimension, when combining classifiers with weights, a fixed or dynamically determined weight is assigned a learner. The majority voting is the most popular weighting combining method for base learners. In this method, a new instance is labeled due to the most frequent vote obtained from base learners [37]. Further, in weighted combination, the predictions of base classifiers are linearly aggregated. When combining classifiers with meta learning techniques, the predictions of classifiers are used as features of a meta learning model. According to [14] in trained methods, the weighted combination shows a slight superiority when compared to meta-classifiers and also simpler for sentiment classification task. Therefore we used majority voting model instead of a meta learning model our experiments.

The majority voting model combines the predictions from different classifiers using a simple fixed rule which counts the predictions of base classifiers and assigns the input to the most predicted class by classifiers. By this rule, with an ensemble model which has k=1,…,K base classifiers, the class label of the jth data point is

determined as in the equation (7) [43].

$$c_j = \sum_{k=1}^{K} I(\arg\max(c_{kj}) = j) \qquad (7)$$

In the literature ensemble learning methods can be divided into two categories: instance partitioning methods such as Bagging and Boosting and feature partitioning methods such as Random Subspace [42]. There is not a global criteria to select a certain ensemble technique for improving the performance of a sentiment analysis task [43]. Thus, as for the second dimension, the learning process is realized with Bagging method and the well known classification algorithms are selected as weak learners of the ensemble model in this study.

In the sentiment classification of short texts task, the main issue for improvement of an ensemble model accuracy is how diversified the input text are represented rather than which classifiers are chosen as base learners. Therefore in this work, three different homogeneous classifier ensembles are formed and the well known classifiers such as NB, SVM and DT are applied as homogeneous classifiers respectively in these models. The proposed ensemble model is shown in figure 1.
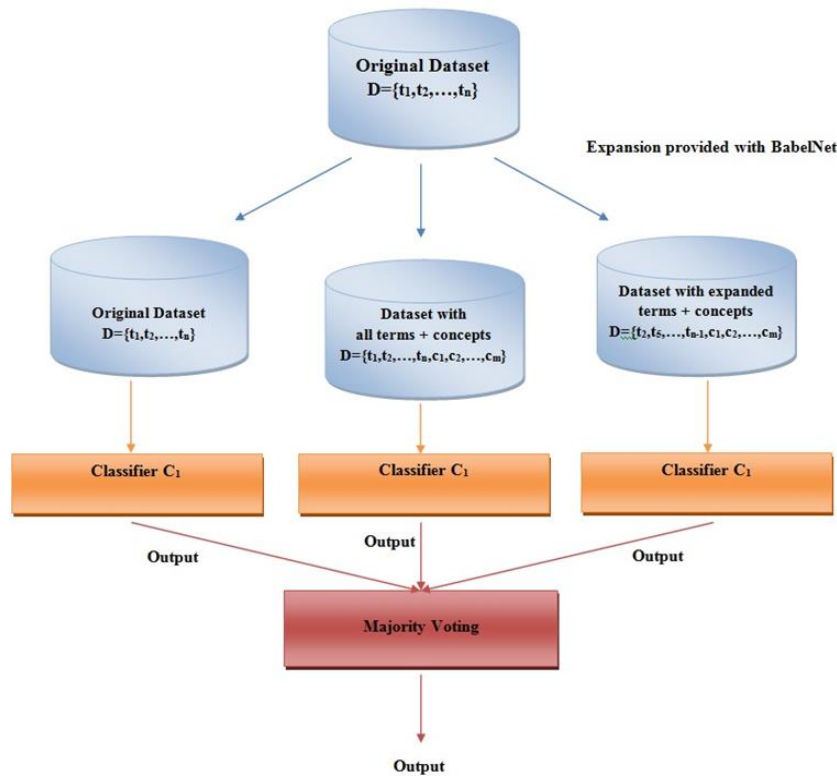

Fig.1. Proposed Ensemble Model

As it is seen from figure 1, three new training datasets for three base learners are generated from original input data. The original data form (raw data) is kept as the training data of first base learner. As for the second one, the raw data and the concept words obtained from Babelnet are jointly used. As for the third one, the concepts words obtained from babelnet and only the terms which are expanded by BabelNet are jointly used. In the data of third base learner, the terms which are not considered by Babelnet are ignored. Thus only the most important terms and their concept extensions are considered as a feature space. Eventually by these base learners, the diversity of an ensemble model has been provided. As the next phase, the decided classification algorithms are used as base learners. And finally, to obtain the class label for an input message, classification outputs of three homogeneous classifiers are combined by majority voting method.

VI. Experimental Results

A. Datasets

Table 1. Description of Datasets

Dataset	#Positive	#Negative	Theme
Hobbit	354	168	Movie
IPhone6	371	161	Smartphone
UMICH	3995	3091	Movie
Archeage	724	994	Game
StsGold	596	739	General
Ststest	183	177	General

We have employed six datasets which are widely evaluated in similar studies to assess the performance of

the proposed ensemble model. The whole datasets are in English, publicly available, non-encoded and involve tweets which are collected through Twitter API and labeled as negative or positive by using a Labeling tool. Table 1 summarizes the positive and negative tweet counts and theme of each datasets.

B. The Results Evaluation

As the first stage of the proposed ensemble model, we have generated three training sets to provide the diversity. The three columns of table2 explain the feature spaces of these training sets respectively. The hobbit dataset originally constituted by 954 terms and this original form is used as the first training dataset of proposed ensemble model. 2182 terms which are obtained by 954 original terms and the related concepts are used as the features of second training dataset of model. By eliminating non extended terms of the original dataset 1992 features are used as the features of third dataset of model. With these three feature sets, diversity for the proposed ensemble model is provided.

Table 2. Feature Extensions

	Feature Space		
	Original	All terms+Concepts	Expanded terms+ Concepts
Hobbit	954	2182	1992
Iphone6	1085	2616	2414
Archeage	1963	4120	3848
UMICH	1650	3615	3387
StsGold	2522	5346	4975
StsTest	1027	2516	2372

After the comprising feature sets by extracting the semantically meaningful terms from twitter messages the model selection for training process must be realized. In this step, the well known classification algorithms such as NB, SVM and C4.5 are selected as weak learners of the ensemble model. In NB, SVM, C4.5 and ensemble experiments 80% of input data is used as training set and the remaining 20% is used as test set. The f-measure results using 10-fold cross validation are reported as in table 3.

Table 3. Classification results

	Classification Methods			
	NB	SVM	C4.5	ENSEMBLE
HOBBIT	0,824	0,834	0,83	**0,916**
IPHONE6	0,772	0,774	0,696	**0,839**
UMICH	0,960	**0,991**	**0,991**	**0.990**
ARCHEAGE	0,751	0,825	0,775	**0,837**
STSGOLD	0,710	0,785	0,714	**0,791**
STSTEST	0,690	0,617	0,709	**0,717**

Bold values in table 3 indicate the best classification performances. When we examine the results, it is clearly realized that the proposed ensemble model presents an overall outstanding performance to any of the other used individual classifiers. Only with UMICH dataset, the ensemble cannot increase the performances of the base

classifiers. The accuracy of base classifiers for this dataset is already high such as 0.991 or 0.96. Therefore the classifiers actually cannot be considered as weak classifiers for this dataset.

When the datasets are examined in terms of theme, it can be clearly seen Sts-Gold and Sts-Test represent several domains. This causes a decline in classification performance because of domain-dependent sentiment words. Polarity of these words such as "short" change from domain to domain. When we mention about "battery life" of a smartphone for the review sentence "The battery life is very short", sentiment word "short" has negative polarity. "The response time of the game is very short." is about "response time" of a game and "short" has positive polarity for this review sentence. The effect of domain dependent sentiment words comes out in the classification results in Table 3. For Sts-Gold and Sts-Test the proposed ensemble model cannot increase the classification results as much as it does in the remaining datasets.

While an ensemble model compared to any other classification models, ensemble quite likely provides a superior performance due to combination of weak classifiers. The main drawback of an ensemble model is being time consuming. However if the time consuming tasks such as model selection and training can be preformed offline then the ensemble model is highly recommended due to its superior accuracy results.

VII. Conclusion

In this paper, we have proposed an ensemble model to classify English Twitter messages according to sentiments by using word embedding representation. We indicated the need for word embedding like methods for short text classification problems; since the obtained results showed that providing word embedding by babelnet is an adequate way to represent main concepts related with the terms. The embedded knowledge extracted from Babelnet contributed the overall performance a short text classification system. Another main contribution of this study is that we propose a more accurate classifier by applying an ensemble model which includes the sub datasets which are constituted with not only the terms in original twits but also the concepts of these terms. We compared our ensemble model with widely used classification algorithms such as NB, SVM and C4.5. The comperative results which are obtained from six public datasets confirm that our ensemble model provides a more accurate classifier for short text messages.

References

[1] S. Sun, C. Luo and Junyu Chen, "A Review of Natural Language Processing Techniques for Opinion Mining Systems", In Information Fusion, vol. 36, pp. 10-25, 2017. doi:10.1016/j.inffus.2016.10.004

[2] Z. Faguo, Z. Fan, Y. Bingru and Y. Xingang, "Research on Short Text Classification Algorithm Based on Statistics and Rules," 2010 Third International Symposium on

Electronic Commerce and Security, Guangzhou, pp. 3-7, 2010. doi:10.1109/ISECS.2010.9

[3] I. Taksa, S. Zelikovitz and A. Spink, "Using Web Search Logs to Identify Query Classification Terms," Information Technology, 2007. ITNG '07. Fourth International Conference on, Las Vegas, NV, pp. 469-474, 2007. doi: 10.1109/ITNG.2007.202

[4] Y. Zhou, B. Xu, J. Xu, L. Yang, C. Li and B. Xu, "Compositional Recurrent Neural Networks for Chinese Short Text Classification," 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Omaha, NE,pp.137-144, 2016. doi: 10.1109/WI.2016.0029

[5] M. Wang, L. Lin and F. Wang, "Improving Short Text Classification through Better Feature Space Selection", 2013 Ninth International Conference on Computational Intelligence and Security, Leshan, pp. 120-124, 2013. doi: 10.1109/CIS.2013.32

[6] B. Sriram, D. Fuhry, E. Demir and M. Demirbas, "Short Text Classification in Twitter to Improve Information Filtering", In Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval, pp.841-842, 2010. doi:10.1145/1835449.1835643

[7] L. Wensen, C. Zewen, W. Jun and W. Xiaoyi, "Short Text Classification Based on Wikipedia and Word2vec", 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, pp. 1195-1200, 2016. doi: 10.1109/CompComm.2016.792489

[8] Q. Chen, L. Yao and J. Yang, "Short Text Classification Based on LDA Topic Model", 2016 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, pp.749-753, 2016. doi: 10.1109/ICALIP.2016.7846525

[9] L. Sang, F. Xie, X. Liu and X. Wu, "WEFEST: Word Embedding Feature Extension for Short Text Classification", 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, pp. 677-683, 2016. doi: 10.1109/ICDMW.2016.0101

[10] A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. Passonneau, "Sentiment Analysis of Twitter Data Proceedings of the Workshop on Languages in Social Media", Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 30-38, 2011.

[11] A. K. Singh, D. K. Gupta and R. M. Singh, "Sentiment Analysis of Twitter User Data on Punjab Legislative Assembly Election", I.J. Modern Education and Computer Science, vol. 9, pp. 60-68, 2017. doi: 10.5815/ijmecs.2017.09.07

[12] I. Mukherjee, S. Sahana and P.K. Mahanti, "An Improved Information Retrieval Approach to Short Text Classification", I.J. Information Engineering and Electronic Business, vol.4, pp. 31-37, 2017. doi: 10.5815/ijieeb.2017.04.05

[13] D. Mumtaz and B. Ahuja, "A Lexical Approach for Opinion Mining in Twitter", I.J. Education and Management Engineering, vol. 4, pp. 20-29, 2016. doi: 10.5815/ijeme.2016.04.03

[14] R. Xia, Chengqing Zong and Shoushan Li, "Ensemble of Feature Sets and Classification Algorithms for Sentiment Classification", In Information Sciences, vol. 181(6), pp. 1138-1152, 2011. doi:10.1016/j.ins.2010.11.023

[15] W. Li, W. Wang and Y. Chen, "Heterogeneous Ensemble Learning for Chinese Sentiment Classification", Journal of Information and Computational Science, 9(15), pp. 4551-4558, 2012.

[16] Y. Su, Y. Zhang, D. Ji, Y. Wang and H. Wu, "Ensemble Learning for Sentiment Classification", 13th Chinese Conf.

on Chinese Lexical Semantics (CLSW'12), Berlin, pp.84-93, 2012. doi:10.1007/978-3-642-36337-5_10Ava

[17] P.P.B. Filho and T.A.S. Pardo, "NILC USP: A Hybrid System for Sentiment Analysis in Twitter Messages", Semeval 2013, Atlanta, Georgia, pp. 568–572, 2013.

[18] A. Hassan, A. Abbasi and D. Zeng, "Twitter Sentiment Analysis: a Bootstrap Ensemble Framework", SocialCom, Alexandria, VA, pp. 357-364, 2013. doi: 10.1109/SocialCom.2013.56

[19] G. Wang, J. Sun, J. Ma, K. Xu and J. Gu, "Sentiment Classification: The Contribution of Ensemble Learning", Decision Support Systems, vol. 57, pp. 77-93, 2014. doi: 10.1016/j.dss.2013.08.002

[20] E. Fersini, E. Messina and F.A. Pozzi, "Sentiment Analysis: Bayesian Ensemble Learning", Decision Support Systems, vol. 68, pp. 26-38, 2014. doi:10.1016/j.dss.2014.10.004

[21] T. Chalothorn and J. Ellman, "Simple Approaches of Sentiment Analysis via Ensemble Learning", Information Science and Applications Lecture Notes in Electrical Engineering, Information Science and Applications, pp. 631-639, 2015. doi: 10.1007/978-3-662-46578-3_74

[22] J.M. Cotelo, F.L. Cruz, F. Enríquez and J.A. Troyano, "Tweet Categorization by Combining Content and Structural Knowledge", Information Fusion, vol. 31, pp.54-64, 2016. doi:10.1016/j.inffus.2016.01.002

[23] I. Perikos and I. Hatzilygeroudis, "Recognizing Emotions in Text Using Ensemble of Classifiers", Engineering Applications of Artificial Intelligence, vol. 51, pp. 191-201, 2016. doi: 10.1016/j.engappai.2016.01.012

[24] E. A. Corrêa Jr., V. Q. Marinho and L. B. dos Santos, "NILC-USP at SemEval-2017 Task 4: A Multi-view Ensemble for Twitter Sentiment Analysis", 2017. doi:abs/1704.02263,2017.

[25] http://www.hlt.utdallas.edu/~yangl/data/Text_Norm_Data _Release_Fei_Liu/, Accessed July 2017

[26] C. D Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J Bethard, and D. McClosky, "The Stanford Corenlp Natural Language Processing Toolkit", Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60, 2014. doi: 10.3115/v1/P14-5010

[27] R. Navigli and S. P. Ponzetto, "BabelNet: The Automatic Construction, Evaluation and Application of a Wide-coverage Multilingual Semantic Network", Artificial Intelligence, vol. 193, pp. 217-250, 2012. doi: 10.1016/j.artint.2012.07.001

[28] A. Moro, F. Cecconi and R. Navigli. "Multilingual Word Sense Disambiguation and Entity Linking for Everybody.", Proc. of the 13th International Semantic Web Conference, Posters and Demonstrations (ISWC 2014), Riva del Garda, Italy, pp. 25-28, 2014.

[29] S. İ. Omurca, S. Baş and E. Ekinci, "An Efficient Document Categorization Approach for Turkish Based Texts", International Journal Of Intelligent Systems And Applications In Engineering, 3(1), pp. 7-13, 2015. doi: 10.18201/ijisae.94177

[30] S. İ. Omurca and E. Ekinci, "An Alternative Evaluation of post Traumatic Stress Disorder with Machine Learning Methods," 2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA), Madrid, pp. 1-7, 2015. doi: 10.1109/INISTA.2015.7276754

[31] H. Parveen and S. Pandey, "Sentiment Analysis on Twitter Data-set using Naive Bayes Algorithm," 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT),

Bangalore, pp. 416-419, 2016. doi: 10.1109/ICATCCT.2016.7912034

[32] J. Jenkins, W. Nick, K. Roy, A. Esterline and J. Bloch, "Author identification using Sequential Minimal Optimization", SoutheastCon 2016, Norfolk, VA, pp. 1-2, 2016. doi: 10.1109/SECON.2016.7506654

[33] L. Yan, Y. Zhang, Y. He, S. Gao et al. "Hazardous Traffic Event Detection Using Markov Blanket and Sequential Minimal Optimization (MB-SMO)", Sensors, Basel, Switzerland, 16(7) , 2016. doi: 10.3390/s16071084

[34] Michael Gamon, "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis", Proceedings of the 20th International Conference on Computational Linguistics, 2004. doi: 10.3115/1220355.1220476

[35] E. Ekinci and H. Takçı, "Elektronik Postaların Adli Analizinde Yazar Analizi Tekniklerinin Kullanılması", 2012.

[36] P. C. S. Njølstad, L. S. Høysæter, W. Wei and J. A. Gulla, "Evaluating Feature Sets and Classifiers for Sentiment Analysis of Financial News", 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, pp. 71-78, 2014. doi: 10.1109/WI-IAT.2014.82

[37] L. Rokach, "Ensemble-based classifiers", Artificial Intelligence Review, vol. 33, pp. 1-39, 2010. doi:10.1007/s10462-009-9124-7

[38] T. Windeatt and G. Ardeshir, "Decision Tree Simplification For Classifier Ensembles". International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 18(5), pp.749-776, 2004. doi:10.1142/S021800140400340X

[39] T. K. Ghosh J, "Error Correlation and Error Reduction in Ensemble Classifiers", Connection science, special issue on combining artificial neural networks: ensemble approaches, 8(3-4), pp. 385-404, 1996. doi:10.1080/095400996116839

[40] X. Hu, "Using Rough Sets Theory and Database Operations to Construct a Good Ensemble of Classifiers for Data Mining Applications", Proceedings 2001 IEEE International Conference on Data Mining, San Jose, CA, pp. 233-240, 2001. doi: 10.1109/ICDM.2001.989524

[41] L. Rokach, "Ensemble Methods for Classifiers", The Data Mining and Knowledge Discovery Handbook, pp. 957-980, 2005. doi: 10.1007/0-387-25465-X_45

[42] ZH. Zhou, "Ensemble Methods: Foundations and Algorithms", Chapman & Hall, 2012.

[43] O. Araque, I. Corcuera-Platas, J. Fernando Sánchez-Rada and Carlos A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications", In Expert Systems with Applications, vol. 77, pp. 236-246, 2017, doi: 10.1016/j.eswa.2017.02.00

Authors' Profiles

Semih Sevim is postgraduate at the Kocaeli University Computer Engineering Department in Turkey, interested in data mining, natural language processing and sentiment analysis.

Sevinç İlhan Omurca is an associate professor at the Kocaeli University Computer Engineering Department in Turkey. She has Ph.D. at the Kocaeli University Electronics and Communication Engineering. Her main research interest includes text mining, sentiment analysis, natural language processing, machine learning and data mining earned.

Ekin Ekinci is Research Assistant of Computer Engineering Departmant at Kocaeli University in Turkey. She has received her BEng. in Computer Engineering from Çanakkale Onsekiz Mart University in 2009 and ME. in Computer Engineering from Gebze Technical University in 2012. She is currently working towards Ph.D. degree in Computer Engineering from Kocaeli University. Her main research interest includes text mining, sentiment analysis, natural language processing and machine learning.