

Generating Word and Document Embeddings for Sentiment Analysis

Cem Rifkı Aydın, Tunga Güngör, and Ali Erkan

Boğaziçi University
Computer Engineering Department
Bebek, Istanbul, Turkey, 34342
cemrifkiaydin@gmail.com, gungort@boun.edu.tr, alierkan@gmail.com

Abstract. Sentiments of words differ from one corpus to another. Inducing general sentiment lexicons for languages and using them cannot, in general, produce meaningful results for different domains. In this paper, we combine contextual and supervised information with the general semantic representations of words occurring in the dictionary. Contexts of words help us capture the domain-specific information and supervised scores of words are indicative of the polarities of those words. When we combine supervised features of words with the features extracted from their dictionary definitions, we observe an increase in the success rates. We try out the combinations of contextual, supervised, and dictionary-based approaches, and generate original vectors. We also combine the word2vec approach with hand-crafted features. We induce domain-specific sentimental vectors for two corpora, which are the movie domain and the Twitter datasets in Turkish. When we thereafter generate document vectors and employ the support vector machines method utilising those vectors, our approaches perform better than the baseline studies for Turkish with a significant margin. We evaluated our models on two English corpora as well and these also outperformed the word2vec approach. It shows that our approaches are cross-lingual and cross-domain.

Keywords: Sentiment Analysis, Opinion Mining, Word Embeddings, Machine Learning

1 Introduction

Sentiment analysis has recently been one of the hottest topics in natural language processing (NLP). It is used to identify and categorise opinions expressed by reviewers on a topic or an entity. Sentiment analysis can be leveraged in marketing, social media analysis, and customer service. Although many studies have been conducted for sentiment analysis in widely spoken languages, this topic is still immature for Turkish and many other languages.

Neural networks outperform the conventional machine learning algorithms in most classification tasks, including sentiment analysis [1]. In these networks, word embedding vectors are fed as input to overcome the data sparsity problem and make the representations of words more “meaningful” and robust. Those embeddings indicate how close the words are to each other in the vector space model (VSM).

Most of the studies utilise embeddings, such as word2vec [2], which take into account the syntactic and semantic representations of the words only. Discarding the sentimental aspects of words may lead to words of different polarities being close to each other in the VSM, if they share similar semantic and syntactic features.

For Turkish, there are only a few studies which leverage sentimental information in generating the word and document embeddings. Unlike the studies conducted for English and other widely-spoken languages, in this paper, we use the official dictionaries for this language and combine the unsupervised and supervised scores to generate a unified score for each dimension of the word embeddings in this task.

Our main contribution is to create original and effective word vectors that capture syntactic, semantic, and sentimental characteristics of words, and use all of this knowledge in generating embeddings. We also utilise the word2vec embeddings trained on a large corpus. Besides using these word embeddings, we also generate hand-crafted features on a review basis and create document vectors. We evaluate those embeddings on two datasets. The results show that we outperform the approaches which do not take into account the sentimental information. We also had better performances than other studies carried out on sentiment analysis in Turkish media. We also evaluated our novel embedding approaches on two English corpora of different genres. We outperformed the baseline approaches for this language as well. The source code and datasets are publicly available¹.

The paper is organised as follows. In Section 2, we present the existing works on sentiment classification. In Section 3, we describe the methods proposed in this work. The experimental results are shown and the main contributions of our proposed approach are discussed in Section 4. In Section 5, we conclude the paper.

2 Related Work

In the literature, the main consensus is that the use of dense word embeddings outperform the sparse embeddings in many tasks. Latent semantic analysis (LSA) used to be the most popular method in generating word embeddings before the invention of the word2vec and other word vector algorithms which are mostly created by shallow neural network models. Although many studies have been employed on generating word vectors including both semantic and sentimental components, generating and analysing the effects of different types of embeddings on different tasks is an emerging field for Turkish.

Latent Dirichlet allocation (LDA) is used in [3] to extract mixture of latent topics. However, it focusses on finding the latent topics of a document, not the word meanings themselves. In [4], LSA is utilised to generate word vectors, leveraging indirect co-occurrence statistics. These outperform the use of sparse vectors [5].

Some of the prior studies have also taken into account the sentimental characteristics of a word when creating word vectors [6–8]. A model with semantic and sentiment components is built in [9], making use of star-ratings of reviews. In [10], a sentiment lexicon is induced preferring the use of domain-specific co-occurrence statistics over the word2vec method and they outperform the latter.

¹ <https://github.com/cemrifki/sentiment-embeddings>

In a recent work on sentiment analysis in Turkish [11], they learn embeddings using Turkish social media. They use the word2vec algorithm, create several unsupervised hand-crafted features, generate document vectors and feed them as input into the support vector machines (SVM) approach. We outperform this baseline approach using more effective word embeddings and supervised hand-crafted features.

In English, much of the recent work on learning sentiment-specific embeddings relies only on distant supervision. In [12], emojis are used as features and a bi-LSTM (bidirectional long short-term memory) neural network model is built to learn sentiment-aware word embeddings. In [13], a neural network that learns word embeddings is built by using contextual information about the data and supervised scores of the words. This work captures the supervised information by utilising emoticons as features. Most of our approaches do not rely on a neural network model in learning embeddings. However, they produce state-of-the-art results.

3 Methodology

We generate several word vectors, which capture the sentimental, lexical, and contextual characteristics of words. In addition to these mostly original vectors, we also create word2vec embeddings to represent the corpus words by training the embedding model on these datasets. After generating these, we combine them with hand-crafted features to create document vectors and perform classification, as will be explained in Section 3.5.

3.1 Corpus-based Approach

Contextual information is informative in the sense that, in general, similar words tend to appear in the same contexts. For example, the word *smart* is more likely to cooccur with the word *hardworking* than with the word *lazy*. This similarity can be defined semantically and sentimentally. In the corpus-based approach, we capture both of these characteristics and generate word embeddings specific to a domain.

Firstly, we construct a matrix whose entries correspond to the number of cooccurrences of the row and column words in sliding windows. Diagonal entries are assigned the number of sliding windows that the corresponding row word appears in the whole corpus. We then normalise each row by dividing entries in the row by the maximum score in it.

Secondly, we perform the principal component analysis (PCA) method to reduce the dimensionality. It captures latent meanings and takes into account high-order cooccurrence removing noise. The attribute (column) number of the matrix is reduced to 200. We then compute cosine similarity between each row pair w_i and w_j as in (1) to find out how similar two word vectors (rows) are.

$$\cos(w_i, w_j) = \frac{w_i^\top w_j}{\|w_i\| \|w_j\|} \quad (1)$$

Thirdly, all the values in the matrix are subtracted from 1 to create a dissimilarity matrix. Then, we feed the matrix as input into the fuzzy c-means clustering algorithm. We chose the number of clusters as 200, as it is considered a standard for word embeddings in the literature. After clustering, the dimension i for a corresponding word indicates the degree to which this word belongs to cluster i . The intuition behind this idea is that if two words are similar in the VSM, they are more likely to belong to the same clusters with akin probabilities. In the end, each word in the corpus is represented by a 200-dimensional vector.

In addition to this method, we also perform singular value decomposition (SVD) on the cooccurrence matrices, where we compute the matrix $M^{PPMI} = U\Sigma V^T$. Positive pointwise mutual information (PPMI) scores between words are calculated and the truncated singular value decomposition is computed. We take into account the U matrix only for each word. We have chosen the singular value number as 200. That is, each word in the corpus is represented by a 200-dimensional vector as follows.

$$w_i = (U)_i \quad (2)$$

3.2 Dictionary-based Approach

In Turkish, there do not exist well-established sentiment lexicons as in English. In this approach, we made use of the TDK (Türk Dil Kurumu - “Turkish Language Institution”) dictionary to obtain word polarities. Although it is not a sentiment lexicon, combining it with domain-specific polarity scores obtained from the corpus led us to have state-of-the-art results.

We first construct a matrix whose row entries are corpus words and column entries are the words in their dictionary definitions. We followed the boolean approach. For instance, for the word *cat*, the column words occurring in its dictionary definition are given a score of 1. Those column words not appearing in the definition of *cat* are assigned a score of 0 for that corresponding row entry.

When we performed clustering on this matrix, we observed that those words having similar meanings are, in general, assigned to the same clusters. However, this similarity fails in capturing the sentimental characteristics. For instance, the words *happy* and *unhappy* are assigned to the same cluster, since they have the same words, such as *feeling*, in their dictionary definitions. However, they are of opposite polarities and should be discerned from each other.

Therefore, we utilise a metric to move such words away from each other in the VSM, even though they have common words in their dictionary definitions. We multiply each value in a row with the corresponding row word’s raw supervised score, thereby having more meaningful clusters. Using the training data only, the supervised polarity score per word is calculated as in (3).

$$w_t = \log \frac{\frac{N_t}{N} + 0.01}{\frac{N'_t}{N'} + 0.01} \quad (3)$$

Here, w_t denotes the sentiment score of word t , N_t is the number of documents (reviews or tweets) in which t occurs in the dataset of positive polarity, N is the number of all the words in the corpus of positive polarity. N' denotes the corpus of negative polarity. N'_t and N' denote similar values for the negative polarity corpus. We perform normalisation to prevent the imbalance problem and add a small number to both numerator and denominator for smoothing.

As an alternative to multiplying with the supervised polarity scores, we also separately multiplied all the row scores with only +1 if the row word is a positive word, and with -1 if it is a negative word. We have observed it boosts the performance more compared to using raw scores.

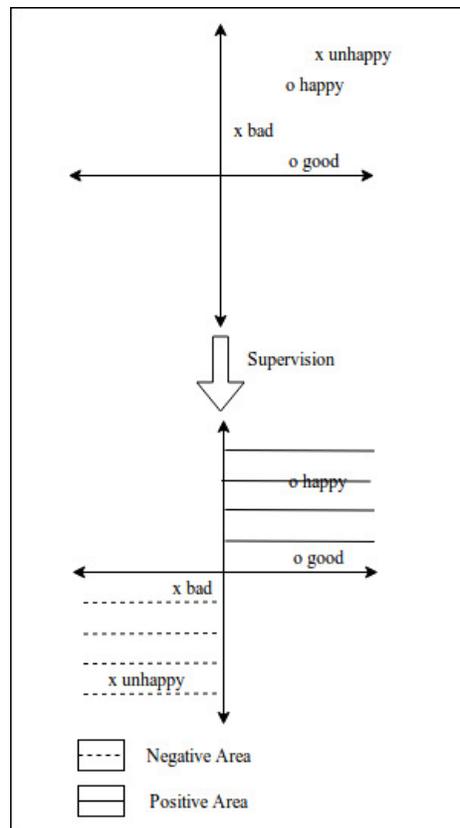


Fig. 1. The effect of using the supervised scores of words in the dictionary algorithm. It shows how sentimentally similar word vectors get closer to each other in the VSM.

The effect of this multiplication is exemplified in Figure 1, showing the positions of word vectors in the VSM. Those “x” words are sentimentally negative words, those

“o” words are sentimentally positive ones. On the top coordinate plane, the words of opposite polarities are found to be close to each other, since they have common words in their dictionary definitions. Only the information concerned with the dictionary definitions are used there, discarding the polarity scores. However, when we utilise the supervised score (+1 or -1), words of opposite polarities (e.g. “happy” and “unhappy”) get far away from each other as they are translated across coordinate regions. Positive words now appear in quadrant 1, whereas negative words appear in quadrant 3. Thus, in the VSM, words that are sentimentally similar to each other could be clustered more accurately. Besides clustering, we also employed the SVD method to perform dimensionality reduction on the unsupervised dictionary algorithm and used the newly generated matrix by combining it with other subapproaches. The number of dimensions is chosen as 200 again according to the U matrix. The details are given in Section 3.4. When using and evaluating this subapproach on the English corpora, we used the SentiWordNet lexicon [14]. We have achieved better results for the dictionary-based algorithm when we employed the SVD reduction method compared to the use of clustering.

3.3 Supervised Contextual 4-scores

Our last component is a simple metric that uses four supervised scores for each word in the corpus. We extract these scores as follows. For a target word in the corpus, we scan through all of its contexts. In addition to the target word’s polarity score (the self score), out of all the polarity scores of words occurring in the same contexts as the target word, minimum, maximum, and average scores are taken into consideration. The word polarity scores are computed using (3). Here, we obtain those scores from the training data.

The intuition behind this method is that those four scores are more indicative of a word’s polarity rather than only one (the self score). This approach is fully supervised unlike the previous two approaches.

3.4 Combination of the Word Embeddings

In addition to using the three approaches independently, we also combined all the matrices generated in the previous approaches. That is, we concatenate the reduced forms (SVD - U) of corpus-based, dictionary-based, and the whole of 4-score vectors of each word, horizontally. Accordingly, each corpus word is represented by a 404-dimensional vector, since corpus-based and dictionary-based vector components are each composed of 200 dimensions, whereas the 4-score vector component is formed by four values.

The main intuition behind the ensemble method is that some approaches compensate for what the others may lack. For example, the corpus-based approach captures the domain-specific, semantic, and syntactic characteristics. On the other hand, the 4-scores method captures supervised features, and the dictionary-based approach is helpful in capturing the general semantic characteristics. That is, combining those three approaches makes word vectors more representative.

3.5 Generating Document Vectors

After creating several embeddings as mentioned above, we create document (review or tweet) vectors. For each document, we sum all the vectors of words occurring in that document and take their average. In addition to it, we extract three hand-crafted polarity scores, which are minimum, mean, and maximum polarity scores, from each review. These polarity scores of words are computed as in (3). For example, if a review consists of five words, it would have five polarity scores and we utilise only three of these sentiment scores as mentioned. Lastly, we concatenate these three scores to the averaged word vector per review.

That is, each review is represented by the average word vector of its constituent word embeddings and three supervised scores. We then feed these inputs into the SVM approach. The flowchart of our framework is given in Figure 2. When combining the unsupervised features, which are word vectors created on a word basis, with supervised three scores extracted on a review basis, we have better state-of-the-art results.

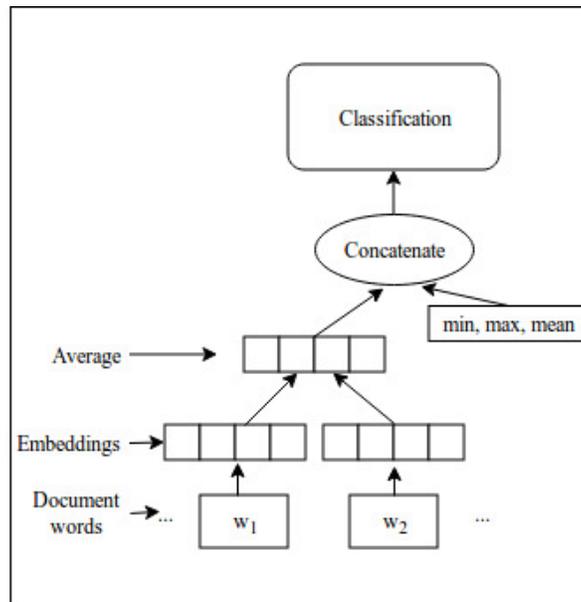


Fig. 2. The flowchart of our system.

4 Datasets

We utilised two datasets for both Turkish and English to evaluate our methods.

For Turkish, as the first dataset, we utilised the movie reviews which are collected from a popular website². The number of reviews in this movie corpus is 20,244 and the average number of words in reviews is 39. Each of these reviews has a star-rating score which is indicative of sentiment. These polarity scores are between the values 0.5 and 5, at intervals of 0.5. We consider a review to be negative if the score is equal to or lower than 2.5. On the other hand, if it is equal to or higher than 4, it is assumed to be positive. We have randomly selected 7,020 negative and 7,020 positive reviews and processed only them.

The second Turkish dataset is the Twitter corpus which is formed of tweets about Turkish mobile network operators. Those tweets are mostly much noisier and shorter compared to the reviews in the movie corpus. In total, there are 1,716 tweets. 973 of them are negative and 743 of them are positive. These tweets are manually annotated by two humans, where the labels are either positive or negative. We measured the Cohen’s Kappa inter-annotator agreement score to be 0.82. If there was a disagreement on the polarity of a tweet, we removed it.

We also utilised two other datasets in English to test the cross-linguality of our approaches. One of them is a movie corpus collected from the web³. There are 5,331 positive reviews and 5,331 negative reviews in this corpus. The other is a Twitter dataset, which has nearly 1.6 million tweets annotated through a distant supervised method [15]. These tweets have positive, neutral, and negative labels. We have selected 7,020 positive tweets and 7,020 negative tweets randomly to generate a balanced dataset.

5 Experiments

5.1 Preprocessing

In Turkish, people sometimes prefer to spell English characters for the corresponding Turkish characters (e.g. *i* for *ı*, *c* for *ç*) when writing in electronic format. To normalise such words, we used the Zemberek tool [16]. All punctuation marks except “!” and “?” are removed, since they do not contribute much to the polarity of a document. We took into account emoticons, such as “:)”, and idioms, such as “kafayı yemek” (lose one’s mind), since two or more words can express a sentiment together, irrespective of the individual words thereof. Since Turkish is an agglutinative language, we used the morphological parser and disambiguator tools [17, 18]. We also performed negation handling and stop-word elimination. In negation handling, we append an underscore to the end of a word if it is negated. For example, “güzel değil” (not beautiful) is redefined as “güzel_” (beautiful_) in the feature selection stage when supervised scores are being computed.

5.2 Hyperparameters

We used the LibSVM utility of the WEKA tool. We chose the linear kernel option to classify the reviews. We trained word2vec embeddings on all the four corpora using the

² <https://www.beyazperde.com>

³ <https://github.com/dennybritz/cnn-text-classification-tf>

Table 1. Accuracies for different feature sets fed as input into the SVM classifier in predicting the labels of reviews. The word2vec algorithm is the baseline method.

Word embedding type	Turkish (%)		English (%)	
	Movie	Twitter	Movie	Twitter
Corpus-based + SVD (U)	76.19	64.38	66.54	87.17
Dictionary-based + SVD (U)	60.64	51.36	55.29	60.00
Supervised 4-scores	89.38	76.00	75.65	72.62
Concatenation of the above three	88.12	73.23	73.40	73.12
Corpus-based + Clustering	52.27	52.73	51.02	54.40
word2vec	76.47	46.57	57.73	62.60
Corpus-based + SVD (U) + 3-feats	88.45	72.60	76.85	85.88
Dictionary-based + SVD (U) + 3-feats	88.64	71.91	76.66	80.40
Supervised 4-scores + 3-feats	90.38	78.00	77.05	72.83
Concatenation of the above three + 3 feats	89.77	72.60	77.03	80.20
Corpus-based + Clustering + 3-feats	87.89	71.91	75.02	74.40
word2vec+ 3-feats	88.88	71.23	77.03	75.64

Gensim library [19] with the skip-gram method. The dimension size of these embeddings are set at 200. As mentioned, other embeddings, which are generated utilising the clustering and the SVD approach, are also of size 200. For c-mean clustering, we set the maximum number of iterations at 25, unless it converges.

5.3 Results

We evaluated our models on four corpora, which are the movie and the Twitter datasets in Turkish and English. All of the embeddings are learnt on four corpora separately. We have used the accuracy metric since all the datasets are completely or nearly completely balanced. We performed 10-fold cross-validation for both of the datasets. We used the approximate randomisation technique to test whether our results are statistically significant. Here, we tried to predict the labels of reviews and assess the performance.

We obtained varying accuracies as shown in Table 1. “3 feats” features are those hand-crafted features we extracted, which are the minimum, mean, and maximum polarity scores of the reviews as explained in Section 3.5. As can be seen, at least one of our methods outperforms the baseline word2vec approach for all the Turkish and English corpora, and all categories. All of our approaches performed better when we used the supervised scores, which are extracted on a review basis, and concatenated them to word vectors. Mostly, the supervised 4-scores feature leads to the highest accuracies, since it employs the annotational information concerned with polarities on a word basis.

As can be seen in Table 1, the clustering method, in general, yields the lowest scores. We found out that the corpus - SVD metric does always perform better than the clustering method. We attribute it to that in SVD the most important singular values are taken into account. The corpus - SVD technique outperforms the word2vec algorithm for some corpora. When we do not take into account the 3-feats technique, the corpus-based SVD method yields the highest accuracies for the English Twitter dataset. We

show that simple models can outperform more complex models, such as the concatenation of the three subapproaches or the word2vec algorithm. Another interesting finding is that for some cases the accuracy decreases when we utilise the polarity labels, as in the case for the English Twitter dataset.

Since the TDK dictionary covers most of the domain-specific vocabulary used in the movie reviews, the dictionary method performs well. However, the dictionary lacks many of the words, occurring in the tweets; therefore, its performance is not the best of all. When the TDK method is combined with the 3-feats technique, we observed a great improvement, as can be expected.

Success rates obtained for the movie corpus are much better than those for the Twitter dataset for most of our approaches, since tweets are, in general, much shorter and noisier. We also found out that, when choosing the p value as 0.05, our results are statistically significant compared to the baseline approach in Turkish [11]. Some of our subapproaches also produce better success rates than those sentiment analysis models employed in English [12, 13]. We have achieved state-of-the-art results for the sentiment classification task for both Turkish and English. As mentioned, our approaches, in general, perform best in predicting the labels of reviews when three supervised scores are additionality utilised.

We also employed the convolutional neural network model (CNN). However, the SVM classifier, which is a conventional machine learning algorithm, performed better. We did not include the performances of CNN for embedding types here due to the page limit of the paper.

As a qualitative assessment of the word representations, given some query words we visualised the most similar words to those words using the cosine similarity metric. By assessing the similarities between a word and all the other corpus words, we can find the most akin words according to different approaches. Table 2 shows the most similar words to given query words. Those words which are indicative of sentiment are, in general, found to be most similar to those words of the same polarity. For example, the most akin word to *muhteşem* (gorgeous) is *10/10*, both of which have positive polarity. As can be seen in Table 2, our corpus-based approach is more adept at capturing domain-specific features as compared to word2vec, which generally captures general semantic and syntactic characteristics, but not the sentimental ones.

6 Conclusion

We have demonstrated that using word vectors that capture only semantic and syntactic characteristics may be improved by taking into account their sentimental aspects as well. Our approaches are cross-lingual and cross-domain. They can be applied to other domains and other languages than Turkish and English with minor changes.

Our study is one of the few ones that perform sentiment analysis in Turkish and leverages sentimental characteristics of words in generating word vectors and outperforms all the others. Any of the approaches we propose can be used independently of the others. Our approaches without using sentiment labels can be applied to other classification tasks, such as topic classification and concept mining.

Table 2. Most similar words to given queries according to our corpus-based approach and the baseline word2vec algorithm.

Query Word	Corpus-based	word2vec
Muhteşem (<i>Gorgeous</i>)	10/10	Harika (<i>Wonderful</i>)
Berbat (<i>Terrible</i>)	Vasat (<i>Mediocre</i>)	Kötü (<i>Bad</i>)
Fark (<i>Difference</i>)	İlginç (<i>Interesting</i>)	Tespit (<i>Finding</i>)
Kötü (<i>Bad</i>)	Sıkıcı (<i>Boring</i>)	İyi (<i>Good</i>)
İyi (<i>Good</i>)	Güzel (<i>Beautiful</i>)	Kötü (<i>Bad</i>)
Senaryo (<i>Script</i>)	Kurgu (<i>Plot</i>)	Kurgu (<i>Plot</i>)

The experiments show that even unsupervised approaches, as in the corpus-based approach, can outperform supervised approaches in classification tasks. Combining some approaches, which can compensate for what others lack, can help us build better vectors. Our word vectors are created by conventional machine learning algorithms; however, they, as in the corpus-based model, produce state-of-the-art results. Although we preferred to use a classical machine learning algorithm, which is SVM, over a neural network classifier to predict the labels of reviews, we achieved accuracies of over 90 per cent for the Turkish movie corpus and about 88 per cent for the English Twitter dataset.

We performed only binary sentiment classification in this study as most of the studies in the literature do. We will extend our system in future by using neutral reviews as well. We also plan to employ Turkish WordNet to enhance the generalisability of our embeddings as another future work.

Acknowledgments

This work was supported by Boğaziçi University Research Fund Grant Number 6980D, and by Turkish Ministry of Development under the TAM Project number DPT2007K12-0610. Cem Rifkî Aydın is supported by TÜBİTAK BİDEB 2211E.

References

1. Goldberg, Y.: A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research* **1510** (2016) 345–420
2. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. *CoRR* **1301** (2013) 1–12

3. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *Journal of Machine Learning Research* **3** (2003) 993–1022
4. Turney, P.D., Pantel, P.: From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* **37** (2010) 141–188
5. Cortes, C., Vapnik, V.: Support-vector networks. *Machine Learning* **20** (1995) 273–297
6. Li, F., Huang, M., Zhu, X.: Sentiment analysis with global topics and local dependency. In: *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, Association for Computational Linguistics (2010) 1371–1376
7. Lin, C., He, Y.: Joint sentiment/topic model for sentiment analysis. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management, ACM* (2009) 375–384
8. Boyd-Graber, J., Resnik, P.: Holistic sentiment analysis across languages: Multilingual supervised latent dirichlet allocation. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics (2010) 45–55
9. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, Association for Computational Linguistics (2011) 142–150
10. Hamilton, W.L., Clark, K., Leskovec, J., Jurafsky, D.: Inducing domain-specific sentiment lexicons from unlabeled corpora. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics (2016) 595–605
11. Ertugrul, A.M., Önal, I., Acartürk, C.: Does the strength of sentiment matter? A regression based approach on turkish social media. In: *Natural Language Processing and Information Systems - 22nd International Conference on Applications of Natural Language to Information Systems, NLDB 2017, Liège, Belgium, June 21-23, 2017, Proceedings.* (2017) 149–155
12. Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., Lehmann, S.: Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In: *EMNLP, Association for Computational Linguistics* (2017) 1615–1625
13. Tang, D., Wei, F., Qin, B., Liu, T., Zhou, M.: Coooolll: A deep learning system for twitter sentiment classification. In: *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014.* (2014) 208–212
14. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D., eds.: *LREC, European Language Resources Association* (2010)
15. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. *Processing* (2009) 1–6
16. Akin, A.A., Akin, M.D.: Zemberek, an open source nlp framework for turkic languages. *Structure* **10** (2007) 1–5
17. Sak, H., Güngör, T., Saraçlar, M.: Morphological disambiguation of turkish text with perceptron algorithm. In: *Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2017)*, CICLing Press (2007) 107–118
18. Sak, H., Güngör, T., Saraçlar, M.: Turkish language resources: Morphological parser, morphological disambiguator and web corpus. In: *Proceedings of the 6th International Conference on Advances in Natural Language Processing (GoTAL '08)*, Springer-Verlag (2008) 417–427
19. Řehůřek, R., Sojka, P.: Software Framework for Topic Modelling with Large Corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, Valletta, Malta, ELRA (2010) 45–50 <http://is.muni.cz/publication/884893/en>.