

Comparing Sentiment Analysis Models to Classify Attitudes of Political Comments on Facebook (November 2016)

Chaya Liebeskind, Karine Nahon, Yaakov HaCohen-Kerner, Yotam Manor

Abstract—This paper is a preliminary study which compares nine ML methods of sentiment analysis aimed towards classifying a corpus of 5.3 million messages of the public on Facebook pages of incumbent politicians. Two sentiments were examined: the general attitude of a comment and the attitude of the comment towards the content of a political post. Our results show that Logistic Regression outperformed the other eight ML models in terms of accuracy and F-measures. Also, we found that n-gram representation performed best. An interesting finding is a difference in success rate when classifying attitude in general vs. attitude towards the content in the political context.

Index Terms— Machine Learning, Political discourse, Sentiment analysis, Social media

I. INTRODUCTION

RESEARCH about the use of social media platforms, such as Facebook and Twitter, by politicians has increased in recent years. These studies examined patterns of behavior of politicians, characteristics of the relationships between politicians and other groups like journalists, celebs and influencers, success and failure factors of political use, propagation of political information in social media and more. This paper adds to the rich literature of politicians and social media by comparing nine Machine Learning (ML) methods of sentiment analysis in an attempt to classify a large corpus of 5.3 Million posts of users replying to politicians (Israeli Member of Knesset, hereafter MKs), posted on Facebook during 2014-2015. This is the first phase of a larger project aimed towards establishing an explanatory model for commenting positively on politicians posts on facebook. The goal of this first phase is to choose the best method for classifying automatically such a big corpus of comments on political posts, in order to be able later run statistical tests to develop an explanatory model of such comments.

¹ Manuscript received August 12, 2016.

Chaya Liebeskind is with the Jerusalem College of Technology, Lev Academic Center, Israel (e-mail: liebchaya@gmail.com).

Karine Nahon is with the Interdisciplinary Center Herzliya, Israel and University of Washington (e-mail: karineb@uw.edu)

Yaakov HaCohen-Kerner is with the Jerusalem College of Technology, Lev Academic Center (e-mail: kerner@jct.ac.il).

Yotam Manor is with the Hebrew University (e-mail: yotammanor@gmail.com).

In this research, we adopt a supervised ML approach. First, we obtained a user comments dataset annotated with sentiment. We distinguish between two sentiment classification tasks: General attitude and Attitude towards the content of the post. Second, we represent each comment as a vector of features. Our feature set include both Facebook depended features, such as "like" and emojis counts, and text-based features. We compare five different text representation approaches, i.e., word, lemma, character n-grams, dictionary-based and extended dictionary-based, by training a classifier to distinguish among sentiment labels, analyzing the relevant features and predicting sentiments for new comments.

The contribution of this study is derived by several factors: the dataset is derived from a large corpus (~5.3 Million messages posted over 2 years on Facebook), the comparison of two different sentiment classification tasks, and it is the first work in NLP on Hebrew Facebook for classification purposes.

II. THEORETICAL BACKGROUND AND LITERATURE

A. Politicians on Social Media

Social media has an important impact on public discourse, and is a major player in political context by users and politicians. Comparative literature survey shows that the use of social media among politicians is constantly increasing in democracies, such as Britain [1], New Zealand [2], Australia [3], the US [4] and Israel [5], while also political participation on social media has increased. In the context of our study two main streams of research which examine political discourse on social media are relevant. One, research that focuses on information flows around political content, and on analysis of relationships among users. For example, Kushin and Kitchener focused on political groups on Facebook and found that the representation of viewpoints was highly skewed in favor of discussion among likeminded participants (homophily) [6]. This homophilous tendency has been reported in other studies which examined other platforms such as Twitter and blogs [7, 8]. Second, research that focuses on sentiment in context of political discourse. For example, Robertson et al studied political discourse on Facebook while focusing on two politicians for 22 months and found that positive comments decreased over time, while negative comments increased [9]. This is similar to the findings of other researchers who showed that the political discourse is dominated by a small portion of users and has a large negative rhetoric laced with sarcasm and humor [10], and that online political discussion tends to

contain a significant level of uncivil discussion [6]. Stieglitz and Dang-Xuan have shown that emotionally charged Twitter messages tend to be shared more often and more quickly compared to neutral ones [11]. Our project enters at this domain. It contributes to the literature by examining the comments of the public on a large corpus of data (5.3 Million messages) collected for two years on posts of political incumbent in Israel (MKs).

B. Sentiment Analysis

When Automatic sentiment analysis addresses the tasks of automatically identifying, extracting, and analyzing subjective information in natural language texts. The general aim is to determine the author's opinion about a specific topic. Most sentiment analysis studies address marketing and commercial tasks, such as extracting opinions from customer reviews [12–14], movie reviews [15, 16], and product reviews [17, 18].

Simultaneously, there is increasing interest in the sentiment analysis of the social web. Sentiment analysis enables to know what people think about specific topic and to perform analysis in order to plan future actions. There is a widespread variety of studies concerning sentiment analysis of posts in various social forums such as: blogs, Facebook, and Twitter.

Tsytarau and Palpanas [19] reviewed the development of sentiment analysis and opinion mining during the last years, and also discussed the evolution of a relatively new research direction, namely, contradiction analysis. The authors supplied an overview of the most popular sentiment extraction algorithms, used in subjectivity analysis and to compare between them. They also introduced an overview of the most popular opinion mining datasets and data sources. According to their analysis, the trends of the past years show an increasing involvement of the research community, along with a drive towards more sophisticated and powerful algorithms. They tried to identify several interesting open problems, and to indicate several promising directions for future research.

Various general approaches have been proposed for the sentiment classification task. Two of the main approaches are the ML and the Dictionary approaches. In our study, we used both the ML and the Dictionary approaches.

The ML approach is composed of two general steps: (1) learn the model from a training corpus, and (2) classify a test corpus based on the trained model [17, 20, 21]. Various ML methods have been applied for sentiment classification. For instance, Pang and Lee applied three ML methods: Naive Bayes (NB), Maximum Entropy (ME) and Support Vector Machines (SVM) [22]. Pang and Lee [22] combined SVM and regression (SVR) modes, with metric labelling. Glorot et al. [23] applied a deep learning method for [large-scale](#) sentiment classification. Moraes et al. [24] empirically compared between SVM and ANN for document-level sentiment classification.

The Dictionary approach is based on a pre-generated dictionary that contains sentiment polarities of single words, such as the Dictionary of Affect of Language², the General

Inquirer³, the WordNet-Affect⁴, or the SentiWordNet [25]. Polarity of a sentence or document is usually computed by averaging the polarities of individual words. Most of the dictionary methods aggregate the polarity values for a sentence or document, and compute the resulting sentiment using simple rule-based algorithms [26]. More advanced systems, such as the Sentiment Analyzer introduced by Yi et al. [21], and the Linguistic Approach by Thet et al [16], extract sentiments precisely for some target topics using advanced methods that exploit domain-specific features, as well as opinion sentence patterns and Part-Of-Speech tags.

Some studies applied both the ML and the Dictionary approaches. For example, Ortigosa et al. [27] introduced their system, called SentBuk, which is able to extract information about the student's sentiments from the messages they write in Facebook with high accuracy. SentBuk retrieves messages written by users in Facebook and classifies them according to their polarity (positive, neutral or negative), extracts information about the users' sentiment polarity according to the sent messages, models the users' regular sentiment polarity, and detects significant emotional changes. The classification method implemented in SentBuk combines lexical-based and ML methods. SentBuk obtained an accuracy result of 83.27% using this classification method. Thelwall, et al. [28] described and assessed the SentiStrength 2 as a general sentiment strength detection algorithm for the social web. Their software primarily uses direct indications of sentiment. The results from six diverse social web data sets (MySpace, Twitter, YouTube, Digg, Runners World, BBC Forums) indicate that their software is better than a baseline approach for all data sets in both supervised and unsupervised cases. SentiStrength 2 is not always better than ML approaches that exploit indirect indicators of sentiment, and is particularly weaker for positive sentiment in news-related discussions. In general, SentiStrength 2 is robust enough to be applied to a wide variety of different social web contexts.

III. METHODS

We compare nine ML methods on a manually coded dataset (N=577) in order to find the best suitable algorithm for classifying comments in political pages of incumbent politicians on Facebook. Once we find the best method we can then classify automatically the entire corpus. The corpus is comprised of posts of 84 (out of 120) MK members (n posts = 33,537), and the comments of ~2.9M users (n of comments = ~5.3M).

A. Preparing the dataset for Analysis

We study two main variables: ATTITUDE and ATTITUDE_TOWARDS_CONTENT_OF_THE_POST.

ATTITUDE: The general attitude conveyed in a comment to a political message. The general attitude focuses on the vibe of the comment. For example – if a comment strengthens the

² <http://www.hdcus.com/>

³ <http://www.wjh.harvard.edu/~inquirer/>

⁴ <http://wndomains.fbkc.eu/wnaffect.html>

opposite view of an MK post, this will still be considered as a general attitude that is positive.

ATTITUDE_TOWARDS_CONTENT_OF_THE_POST: The Attitude of the comment towards the political post denotes whether the commenter support or oppose the political content of the post (1=Positive, 2=Negative, 3=Neutral, 4= Not Applicable, that is the comment does not relate to the post of the MK, 99=Unclear/Undefined)

Initially, we manually coded 100 comments by 3 coders. Coding manually political messages is complicated as the same text may reflect multiple attitudes towards multiple stakeholders. Therefore, we needed 3 rounds of manual coding

in order to reach a satisfactory reliability level. In each one of the rounds the coders discussed the disagreements and refined the coding scheme to reach a better agreement. In the 3rd round we calculated Fleiss' Kappa to measure reliability of agreement for two variables: attitude (0.78) and attitude towards content of the post (0.82). A Fleiss Kappa between 0.6-0.8 is considered a 'substantial agreement', and >0.8 'almost perfect agreement'[30]. Once we reached a high level of agreement, one coder continued and manually coded 612 comments. The comments were chosen respective to their distribution in the main corpus (see table 1).

TABLE I
THE DISTRIBUTION OF THE ATTITUDES IN THE SAMPLED DATASET

Variables	Positive	Negative	Neutral	Not Applicable	Unclear
ATTITUDE	233	327	47	-	5
ATTITUDE_TOWARDS_CONTENT_OF_THE_POST	221	122	16	243	10

For the preparation of our dataset we omitted the unclear category and comments which were not written in Hebrew or in English. Finally, the dataset that we ran was N=577.

B. Supervised attitude classification

In this research, we adopt a supervised Machine Learning (ML) approach for classifying Facebook comments. We next describe the collected information from the text and Facebook properties and how we incorporate it as features within the ML framework.

Feature Sets. We next detail how the special characters of Facebook, e.g. emojis, found useful in prior work, are encoded as features and describe different text representations, which we have explored, for feature extraction.

Facebook-based Features. In the last decade, the necessity of incorporating Emojis' information in automated sentiment classification of informal texts was proven [14, 31–34]. Therefore, we encoded each Emoji as separate feature and counted the number of its occurrences in the comment. Next, using Facebook API, we extracted additional three Facebook depended features: the number of "likes" that the comment got, the number of comments on the comment, and a Boolean feature, which indicates whether the commentator also "liked" the status. Another two features that we defined are the number of occurrences of the MK writer of the post and the number of occurrences of other MKs, either aliens or rivals of the post writer.

Text-based Features. First, we define two general text-based features: the number of words in the comment and the number of characters in the comment. Then, following the rationale of Aisopos et al. [35] that the higher the number of punctuations is, the more likely is the corresponding comment to be subjective, we encoded common punctuations (with frequency > 10) as features by counting their normalized

number of occurrences. In Twitter, Aisopos et al. found that while exclamation marks constitute a typical annotation for positive sentiments, question marks usually express a negative feeling. The defined punctuation features allow us to explore whether these findings are also valid in our setting.

Next, we investigate five types of text representations:

1. Unigram/Word representation - Each of the words in the comment is considered as a feature. The score of the feature is the word number of occurrences in the comment divided by the comment length (termed normalized word count).
2. Lemma representation- We lemmatized all the comments using a Part-of-Speech (PoS) tagger [36]. Then, each of the comments' lemmas is a feature scored by the normalized lemma count.
3. Character n-grams representation - Each comment is considered as a *character n-grams*, i.e., strings of length n. For example, the character 3-grams of the string "character" would be: "cha", "har", "ara", "rac", "act", "cte", and "ter". Since there is much less character combinations than word combinations, this representation overcomes the problem of sparse data that arises when using word representation. On the other hand, this representation still produces a considerably larger feature set. Previous work on short informal data showed that character n-gram features can be quite effective for sentiment analysis [35, 37]. This is due to the tendency of noise and misspellings to have smaller impact on substring patterns than on word patterns. Therefore, in this representation, we considered each of the character n-grams of the comment as a feature and scored it by its normalized count in the comment.
4. Dictionary-based representation - We combine the dictionary approach, which relies on a pre-built dictionary

that contains opinion polarities of words, with our ML approach. Our features are the dictionary words scored by their normalized count. We used the intersection of the seed sentiment list with the manually extended list of 85 positive words and 83 negative words generated by HaCohen-Kerner and Badash [38].

5. Extended dictionary-based representation- We extended our dictionary with Facebook sentiment words by applying a statistical measure of word co-occurrence. Assuming that words that occur frequently together are topically related [39], for each sentiment word in the original dictionary (described in the previous dictionary-based representation), we extracted the 20 most similar word using *Dice coefficient* [40] and an unannotated corpus of over than 4 million comments. Then, an annotator selected the sentiment words from these candidate lists. We increased the size of our Hebrew dictionary (the extended sentiment list) from 177 words to 830 words (327 positive words and 503 negative words). Our features are the dictionary words scored by their normalized count. Since the dictionary was generated from the Facebook corpus, the extracted sentiment words are typical to Facebook. We recognized two interesting type of words: slang sentiment words such

as "king" and "stupid", and sentiment words from events that affect political discourse such as "terrorist attack" and "unemployed".

IV. RESULTS

We used nine ML methods to combine the features in a supervised classification framework: Random Forest, Decision Tree, Bagging, Adaboost, Bayes Network, Supported Vector Machine (SVM), Logistic Regression and Multilayered Perceptron. We estimated the accuracy rate of each ML method by a 10-fold cross-validation test. We ran these ML methods by the WEKA platform [41, 42] using the default parameters. To reduce the number of features in the feature sets, we tried to filter out non-relevant features using two well-known feature selection methods: Information gain (InfoGain, IG) [43] and Correlation-based Feature Subset (CFS) [44]. The second method had better performance. Therefore, the results presented in this section include CSF feature selection which significantly improved the accuracy of all the configurations. (We detail the important features, which were selected by the CSF feature selection for the best configurations in Table 5 of the analysis Section).

TABLE II
COMPARISON OF RESULTS OBTAINED BY NINE ML METHODS

#	ML Method	ATTITUDE		ATTITUDE_TOWARDS_CONTENT	
		Accuracy (%)	F-Measure	Accuracy (%)	F-Measure
1	Random Forest	74	0.713	60	0.589
2	Decision Tree (J48)	71	0.699	61	0.565
3	Bagging	73	0.712	63	0.598
4	AdaBoost (M1)	69	0.67	61	0.54
5	Bayes Network	71	0.693	60	0.55
6	Logistic Regression	78	0.771	66	0.64
7	Multilayered Perceptron	75	0.744	62	0.595
8	SVM (SMO)	72	0.709	62	0.579
9	SVM (LibSVM)	69	0.673	61	0.54

TABLE III
COMPARISON OF RESULTS OBTAINED BY FIVE TEXT-BASED REPRESENTATIONS

#	Representation	ATTITUDE		ATTITUDE_TOWARDS_CONTENT	
		Accuracy (%)	F-Measure	Accuracy (%)	F-Measure
1	Word\Unigram	78	0.771	66	0.64
2	Lemma	77	0.761	64	0.619
3	Character n-grams	80	0.801	67	0.665
4	Dictionary-based	74	0.733	61	0.554
5	Extended dictionary-based	75	0.742	62	0.562

Table 2 shows the performances of the different ML methods on the feature set of Facebook and the state-of-the-art word representation. The best ML method was Logistic Regression. Therefore, we have performed further experiments using only this method.

In this research, we investigated five types of text representations (Section 3): unigram\word representation,

lemma representation, character n-grams representation, dictionary-based representation and extended dictionary-based representation. The attitude classification results of the Logistic Regression algorithm using each of these representations are presented in the left side of Table 3. The character n-grams representation (n=3) yielded the best accuracy result (80%). The advantage of the representation over the extended dictionary-based representation is notable (5%) and is statistically significant according to the McNamar

test [45] at level 0.05. Even though, we extended our dictionary using statistical co-occurrence measure, the dictionary coverage is still limited. We consider utilizing a semi-automatic iterative scheme to increase the recall of the dictionary [46].

The results of the attitude towards content classification results of the Logistic Regression algorithm are presented in the right side of Table 3. The best results (67%) were obtained using the character n-grams representation (n=2 and n=3). However, these results are significantly lower than the results

of the attitude classification. The task of attitude towards content classification is difficult and more sophisticated text understanding approaches, e.g. semantic similarity between the post and the comment, should be applied.

We experiment three configurations of the character n-grams representations: n=2, n=3 and a combination of n=2 and n=3. Table 4 shows a comparison of the character n-grams configurations for the two classification tasks. The optimal configurations of the tasks were different.

TABLE IV
A COMPARISON OF THE CHARACTER N-GRAMS CONFIGURATIONS

#	Character n-grams	ATTITUDE		ATTITUDE_TOWARDS_CONTENT	
		Accuracy (%)	F-Measure	Accuracy (%)	F-Measure
1	n=2	74	0.737	64	0.625
2	n=3	80	0.801	62	0.577
3	n=2 and n=3	74	0.75	67	0.665

V. ANALYSIS

We used the information obtained by the CFS selection method to better understand which features have more influence on the classification accuracy. Table 5 presents information on the features, which were selected by the CFS method for the best configuration for each of the classification tasks. The Boolean feature, which indicates whether the commentator also "liked" the status was informative for both of the tasks. Although the name of the writer of the post was not selected by the attribute towards content classifier, the character 2-grams "MK", which

indicates a mention of a politician, was selected. No emoji feature was selected for any of the tasks. Only some of the selected features were informative, namely formed a word of two or three letters in English or Hebrew. For example, in the top-20 selected features, both classification tasks selected the English word "age" along with the Hebrew words "already" and "next". Additional selected features for the attitude classification task were the Hebrew words "law", "father", "her", "for them", "past" and "white". Additional selected features for the attitude towards content classification task were the Hebrew words "sex", "it", "no" and two plural suffixes of two letters.

TABLE V
SELECTED FEATURES FOR THE BEST CONFIGURATIONS

Task	# feat.	Facebook-based	Text-based	General
Attitude classification	62	COMMENTOR_LIKED, MK_WRITER_OF_POST	HE: 51 EN: 6	Special chars: "and " Comment length (number of words)
Attitude towards content	43	COMMENTOR_LIKED	HE: 20 (n=2), 14 EN: 3 (n=2), 4 (n=3)	Special chars: /

TABLE VI
ATTITUDE CLASSIFICATION: CONFUSION MATRIX

Positive	Negative	Neutral	
177	48	0	Positive
34	270	14	Negative
4	14	16	Neutral

TABLE VII
ATTITUDE TOWARDS CONTENT CLASSIFICATION: CONFUSION MATRIX

Positive	Negative	Not applicable	
129	69	19	Positive
26	212	13	Negative
30	31	48	Not applicable

Table 5 we complete our analysis by presenting the confusion matrixes of the best classification results. Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class.

Table 6 shows that most of the classification errors were due to incorrect classification of positive comments as negative (48) and vice versa (34). Most of the incorrectly classified neutral comments were classified as negative. No positive comment was classified as neutral.

Table 7 shows that most of the classification errors were due to incorrect classification of positive comments as negative (69) and vice versa (26). However, there was no difference between the number of comments that are not forwarded to the content which were classified as positive (30) and the number of these comments which were classified as negative (31).

VI. CONCLUSIONS

In this paper, we presented two sentiment classification tasks: General attitude and Attitude towards the content of the post. We combined Facebook-based and text-based features in supervised ML algorithms. We obtained that classifying the attitude towards the content is significantly more difficult. For both of the tasks, we found that the character n-grams model text representation outperformed other four representations. This is the first work in NLP on Hebrew Facebook for classification purposes.

We further plan to explore word embedding for text representations, where words are mapped to vectors of real numbers. Methods of word embedding mathematically reduce the dimension of the words' vector to a continuous vector with a lower dimension. The dimension reduction is often implemented by one of the following methods: neural networks, dimensionality reduction on the word co-occurrence matrix and probabilistic models.

In addition, to increase the performance of the attitude towards context classification, we plan to add features which calculate the textual and semantic similarities between the post and comment text.

ACKNOWLEDGMENT

We would like to express our deep gratitude to Avital Day, our research assistant, for her help in programming and carrying out the research experiments. This work was partially funded by an internal research grant from Jerusalem College of Technology, Lev Academic Center.

REFERENCES

1. Williamson, A.: MPs on Facebook. Hansard Society, London (2009).
2. Busby, C., Bellamy, P.: New Zealand Parliamentarians and online social media. New Zealand Parliament's Parliamentary Library (2011).

3. Bruns, A., Burgess, J.: #Ausvotes: How twitter covered the 2010 Australian federal election, <http://search.informit.com.au/documentSummary;dn=627330171744964;res=IELHSS>.
4. Smith, A.: Cell Phones, Social Media and Campaign 2014, <http://www.pewinternet.org/2014/11/03/cell-phones-social-media-and-campaign-2014/>, (2014).
5. Haleva-Amir, S.: Personal Web Applications in the Service of Knesset Members: Personal Israeli Politics in the Digital Era, (2014).
6. Kushin, M.J., Kitchener, K.: Getting political on social network sites: Exploring online political discourse on Facebook. *First Monday*. 14, (2009).
7. Nahon, K., Hemsley, J.: Homophily in the Guise of Cross-Linking Political Blogs and Content. *Am. Behav. Sci.* 0002764214527090 (2014).
8. Colleoni, E., Rozza, A., Arvidsson, A.: Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *J. Commun.* 64, 317–332 (2014).
9. Robertson, S.P., Douglas, S., Maruyama, M., Semaan, B.: Political Discourse on Social Networking Sites: Sentiment, In-group/Out-group Orientation and Rationality. *Inf Polity*. 18, 107–126 (2013).
10. Mejova, Y., Srinivasan, P., Boynton, B.: GOP Primary Season on Twitter: "Popular" Political Sentiment in Social Media. In: *WSDMI3* (2013).
11. Stieglitz, S., Dang-Xuan, L.: Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *J. Manag. Inf. Syst.* 29, 217–248 (2013).
12. Hu, M., Liu, B.: Mining opinion features in customer reviews. In: *AAAI*. pp. 755–760 (2004).
13. Bross, J., Ehrig, H.: Automatic construction of domain and aspect specific sentiment lexicons for customer review mining. In: *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. pp. 1077–1086. ACM (2013).
14. Kiritchenko, S., Zhu, X., Mohammad, S.M.: Sentiment analysis of short informal texts. *J. Artif. Intell. Res.* 50, 723–762 (2014).
15. Kennedy, A., Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifters. *Comput. Intell.* 22, 110–125 (2006).
16. Thet, T.T., Na, J.-C., Khoo, C.S., Shakthikumar, S.: Sentiment analysis of movie reviews on discussion boards using a linguistic approach. In: *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. pp. 81–84. ACM (2009).
17. Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: *Proceedings of the 12th international conference on World Wide Web*. pp. 519–528. ACM (2003).
18. Cui, H., Mittal, V., Datar, M.: Comparative experiments on sentiment classification for online product reviews. In: *AAAI*. pp. 1265–1270 (2006).
19. Tsytsarau, M., Palpanas, T.: Survey on mining subjective data on the web. *Data Min. Knowl. Discov.* 24, 478–514 (2012).
20. 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-Volume 10*. pp. 79–86. Association for Computational Linguistics (2002).
21. Yi, J., Nasukawa, T., Bunesco, R., Niblack, W.: Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In: *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*. pp. 427–434. IEEE (2003).
22. Pang, B., Lee, L.: Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: *Proceedings of the 43rd annual meeting on association for computational linguistics*. pp. 115–124. Association for Computational Linguistics (2005).
23. Glorot, X., Bordes, A., Bengio, Y.: Domain adaptation for large-scale sentiment classification: A deep learning approach. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. pp. 513–520 (2011).
24. Moraes, R., Valiati, J.F., Neto, W.P.G.: Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Syst. Appl.* 40, 621–633 (2013).

25. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proceedings of LREC. pp. 417–422. Citeseer (2006).
26. Zhu, J., Zhu, M., Wang, H., Tsou, B.K.: Aspect-based sentence segmentation for sentiment summarization. In: Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion. pp. 65–72. ACM (2009).
27. Ortigosa, A., Martín, J.M., Carro, R.M.: Sentiment analysis in Facebook and its application to e-learning. *Comput. Hum. Behav.* 31, 527–541 (2014).
28. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment strength detection for the social web. *J. Am. Soc. Inf. Sci. Technol.* 63, 163–173 (2012).
29. Mohammad, S.M., Kiritchenko, S., Zhu, X.: NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. *ArXiv Prepr. ArXiv13086242*. (2013).
30. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. *Biometrics*. 33, 159–174 (1977).
31. Boia, M., Faltings, B., Musat, C.-C., Pu, P.: A:) is worth a thousand words: How people attach sentiment to emoticons and words in tweets. In: Social Computing (SocialCom), 2013 International Conference on. pp. 345–350. IEEE (2013).
32. Hogenboom, A., Bal, D., Frasinca, F., Bal, M., de Jong, F., Kaymak, U.: Exploiting emoticons in sentiment analysis. Presented at the (2013).
33. Zhao, J., Dong, L., Wu, J., Xu, K.: Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1528–1531. ACM (2012).
34. Liu, B.: Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* 5, 1–167 (2012).
35. Aisopos, F., Papadakis, G., Tserpes, K., Varvarigou, T.: Content vs. context for sentiment analysis: a comparative analysis over microblogs. In: Proceedings of the 23rd ACM conference on Hypertext and social media. pp. 187–196. ACM (2012).
36. Adler, M., Goldberg, Y., Gabay, D., Elhadad, M.: Unsupervised Lexicon-Based Resolution of Unknown Words for Full Morphological Analysis. In: ACL. pp. 728–736 (2008).
37. Raaijmakers, S., Kraaij, W.: A shallow approach to subjectivity classification. In: ICWSM (2008).
38. HaCohen-Kerner, Y., Badash, H.: Positive and Negative Sentiment Words in a Blog Corpus Written in Hebrew. In: the 20th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems. *Procedia Computer Science*, 733-743 (2016).
39. Schutze, H., Pedersen, J.O.: A Cooccurrence-Based Thesaurus and Two Applications to Information Retrieval. *Inf. Process. Manag.* 33, 307–18 (1997).
40. Smadja, F., McKeown, K.R., Hatzivassiloglou, V.: Translating collocations for bilingual lexicons: A statistical approach. *Comput. Linguist.* 22, 1–38 (1996).
41. Witten, I.H., Frank, E.: *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann (2005).