2012 IEEE 12th International Conference on Data Mining Workshops

Learning Domain-Specific Polarity Lexicons

Gulsen Demiroz*, Berrin Yanikoglu*, Dilek Tapucu*†, and Yucel Saygin*
* Faculty of Engineering and Natural Sciences, Sabancı University, Istanbul, Turkey
†Dept. of Computer Engineering, Izmir Institute of Technology, Izmir, Turkey
Email: {gulsend, berrin, dilektapucu, ysaygin}@sabanciuniv.edu

***Abstract*—Sentiment analysis aims to automatically estimate the sentiment in a given text as positive or negative. Polarity lexicons, often used in sentiment analysis, indicate how positive or negative each term in the lexicon is. However, since creating domain-specific polarity lexicons is expensive and time-consuming, researchers often use a general purpose or domain-independent lexicon. In this work, we address the problem of adapting a general purpose polarity lexicon to a specific domain and propose a simple yet effective adaptation algorithm. We experimented with two sets of reviews from the hotel and movie domains and observed that while our adaptation techniques changed the polarity values for only a small set of words, the overall test accuracy increased significantly: 77% to 83% in the hotel dataset and 61% to 66% in the movie dataset.**

***Keywords*—sentiment analysis; lexicon adaptation; polarity detection; machine learning; natural language processing**

I. Introduction

Sentiment analysis aims to estimate the sentiment in textual documents as positive, or negative. It has become a popular research area in recent years, due to wide range of applications and commercial interest to the problem. Automatically analyzing product reviews for instance is of great interest to companies to understand what their customers are thinking about a product or a specific aspect of a product. A common approach in sentiment analysis is to try to estimate the sentiment expressed in the review from the polarity (sentiment orientation) of the words within the text.

The problem of identifying the polarity of words have been addressed in [1] where authors apply a clustering method to determine the polarity of adjectives. Similarly authors in [2] use a set of seed words and clustering methods to find the polarity of adjectives in a corpus. More recently, polarity lexicons, such as the SentiWordnet[3], have been built for sentiment analysis. A polarity lexicon can be constructed manually [4], using heuristics [5], [6] or by machine learning techniques [3]. In [7], they discuss three main approaches for opinion lexicon building: manual approach, dictionary-based approach, and corpus-based approach. The major shortcoming of the manual approach is the cost (time and effort) to hand select words to build such a lexicon. There is also the possibility of missing important words that could be captured with automatic methods. Dictionary-based approaches work by expanding a small set of seed opinion words, with the use of a lexical resource such as the WordNet [8]. The main drawback of these approaches is that the resulting lexicon is not domain specific. Corpus-based approaches can overcome these problems by learning a domain-specific lexicon using a domain corpus of labeled reviews.

The most commonly used polarity lexicon is the SentiWordNet, where a word is associated with a negative polarity to indicate its negative sentiment orientation, a positive polarity to indicate its positive sentiment orientation, and an objective polarity to indicate its neutrality. The basic assumption with the polarity lexicons is that a word's polarity is the same across different domains, which may not be true for all the words. For example, the word "*small*" has a polarity of -0.25 in SentiWordNet which is appropriate in the hotel domain like in the review sentence "*The hotel had really small rooms*". However, when the review is about a digital camera, the word "*small*" should have a positive polarity as in the review sentence "This camera is great as it has a small size".

The idea of updating the polarities of words in a given lexicon has been investigated before. In [9] authors stress the importance of contextual polarity to differentiate from the prior polarity of a word. They extract contextual polarities by defining several contextual features. In [10], a double propagation method is used to extract both sentiment words and features, combined with a polarity assignment method starting with a seed set of words. In [11], which has been the main motivation for this work, authors use linear programming to update the polarity of words based on specified hard or soft constraints. For instance, if a word has negative polarity in the domain-independent lexicon but appears together with positive words in general, then its polarity is updated to positive, to minimize the costs imposed by the soft constraints. Another application of linear programming appears in [12] to learn a sentiment lexicon which is not only domain specific but also aspect-dependent. Lastly, a recent work expands a given dictionary of words with known polarities by first producing a new set of synonyms with polarities and using these to further deduce the polarities of other words [13].

In this work we propose several variations of a simple method which is based on the delta tf-idf concept [14], to adapt a domain-independent polarity lexicon to a specific domain. We use SentiWordNet, a domain independent lexicon, as a baseline for demonstrating the effectiveness of the proposed method. Specifically, we adapt SentiWordNet to two different domains in order to obtain two domain-specific lexicons. We

This work was developed in the context of UBIPOL (Ubiquitous Participation Platform for Policy Making) project funded by European Commission, FP7.

978-0-7695-4925-5/12 $26.00 © 2012 IEEE
DOI 10.1109/ICDMW.2012.120

then compare the sentiment classification accuracies obtained with SentiWordNet and the new domain specific lexicons.

The basic idea for domain adaptation is to learn the domain-specific polarities from labeled reviews in a given domain. In order to do that, we analyze the occurrence of the words in the lexicon in positive and negative reviews in a given domain. If a particular word occurs significantly more in positive reviews than in negative reviews, then we assume that this word should have positive polarity for this domain, and vice versa. We propose a couple of alternatives for the update mechanism of a word's polarity. The proposed approaches allow us to adapt a domain-independent lexicon such as SentiWordNet, for a specific domain by updating the polarities of only a small subset of the words. However, we also show that this small set of updated words has a significant contribution to sentiment analysis accuracy.

Our paper is organized as follows: Section II describes sentiment analysis with a domain-independent lexicon. Section III presents the quality of the adapted lexicon. Section IV describes experimental results and error analysis. Finally, in Section V we draw some conclusions and propose future extension of this work.

II. Sentiment Analysis with a Domain-Independent Lexicon

We show the advantages of adapting a domain-independent polarity lexicon by comparing two approaches: (1) sentiment analysis using a domain-independent lexicon as explained in this section, and (2) sentiment analysis using the adapted lexicon exactly like the first approach. The adaptation process is explained in Section III.

The polarity lexicon we use as the domain-independent lexicon is the SentiWordNet that consists of a list of words with their POS tags and three associated polarity scores $< pol^-, pol^=, pol^+ >$ for each word [3]. The polarity scores indicate the measure of negativity, objectivity and positivity, and they sum up to 1. Some sample scores are provided in Table I from SentiWordNet.

TABLE I
SAMPLE ENTRIES FROM SENTIWORDNET

Word	Type	Negative	Objective	Positive
sufficient	JJ	0.75	0.125	0.125
comfy	JJ	0.75	0.25	0.0
moldy	JJ	0.375	0.625	0.0
joke	NN	0.19	0.28	0.53
fireplace	NN	0.0	1.0	0.0
failed	VBD	0.28	0.72	0.0

A. Word polarity

As many other researchers have done, we simply select the dominant polarity of a word as its polarity and use the sign to indicate the polarity direction. The dominant polarity of a word w, denoted by $Pol(w)$, is calculated as:

$$Pol(w) = \begin{cases} 0 & \text{if} \quad max(pol^=, pol^+, pol^-) = pol^= \\ pol^+ & \text{else if} \quad pol^+ \geq pol^- \\ -pol^- & \text{otherwise} \end{cases} \tag{1}$$

In other words, given the polarity triplet $< pol^-, pol^=, pol^+ >$ for a word w, if the objective polarity is the maximum of the polarity scores, then the dominant polarity is 0. Otherwise, the dominant polarity is the maximum of the positive and negative polarity scores where pol^- becomes $-pol^-$ in the average polarity calculation. For example, the polarity triplet of the word "sufficient" is <0.75,0.125,0.125>; hence Pol("sufficient") $= -0.75$. Similarly, the polarity triplet of the word "moldy" is <0.375,0.625,0.0>; hence Pol("moldy") $= 0$.

An alternative way for calculating dominant polarity could be to completely ignore the objective polarity $pol^=$ and determine the $Pol(w_i)$ of the word to be the maximum of pol^- and pol^+. With this method, the dominant polarity of the word "moldy" would be -0.375 instead of 0. However, we preferred the first approach as more appropriate, since many words appear as objective or dominantly objective in SentiWordNet.

B. Review polarity

For estimating the sentiment in a review, we use a simple approach that computes the average review polarity and makes a decision based on this score. This is done by first applying the Stanford NLP tool [15] to all the reviews in order to extract the POS tags of each word. Then, we compute the average polarity of the review using the dominant polarity of each word in the review using Eq. 2, using only words with POS tags JJ*(Adjective), RB*(Adverb), NN*(Noun), and VB*(Verb) which have dominant polarity positive or negative (we do not count the objective polarity words as their dominant polarity is 0).

$$\text{Average review polarity}(R) \quad = \tfrac{1}{|R|} \textstyle\sum_{w_i \in R} Pol(w_i) \tag{2}$$

The reviews with average polarity score greater than a threshold of zero are classified as Positive, while others are classified as Negative (zero score is classified as Negative). This threshold was found experimentally to give a roughly equal number of mistakes in positive and negative review classification.

III. Adapting a Domain-Independent Lexicon

Our purpose is to a update the polarities of words in a given polarity lexicon to adapt them to specific domains. In this section we describe our approach for domain adaptation together with the methods we used for selecting the set of adapted words for sentiment classification.

A. Finding domain specific words

For adapting the general purpose lexicon, we update the polarity of a word, if it's occurrence in labeled reviews strongly indicate one class, while SentiWordNet would suggest the other class. For instance if a word's dominant polarity is negative, but it occurs very often in positive reviews and not very often in negative ones, we update its dominant polarity. In order to see which words in the domain appear more in a particular class of reviews, compared to the other class, we first compute the tf-idf (term frequency - inverse document frequency) scores of each word separately for positive and negative review classes. The $tf(w, c)$ counts the occurrence of word w in class c, while $idf(w)$ is the proportion of documents where the word w occurs, discounting very frequently occurring words in the whole database (e.g. 'not', 'be') [16]. There are quite a few variants of tf-idf computations [17], and the tf-idf variant we use is denoted as $tf.idf$ and computed as:

$$tf.idf(w_i, +) = log_e(tf(w_i, +) + 1) * log_e(N/df(w_i)) \quad (3)$$

$$tf.idf(w_i, -) = log_e(tf(w_i, -) + 1) * log_e(N/df(w_i))$$

where the first term is the scaled term frequency (tf) and the second term is the scaled inverse document frequency (idf). The term $df(w_i)$ indicates the document frequency which is the number of documents in which w_i occurs and N is the total number of documents (reviews in our case) in the database.

In Eq. 4, we define a new measure for polarity adaptation of words, called $(\Delta tf)idf$. It estimates whether the polarity of a word should be adjusted, considering its occurrence in positive and negative reviews separately.

$$(\Delta tf)idf(w_i) = tf.idf(w_i, +) - tf.idf(w_i, -) \quad (4)$$

$$= [tf(w_i, +) - tf.idf(w_i, -)] \times idf(w_i)$$

Our new measure is similar to the *Delta TFIDF* term defined in [14] for calculating the polarity scores of words. As shown in Eq. 5, *Delta TFIDF*(w_i, d) score of a word w_i in document d considers the difference in the document frequencies of that word in positive and negative corpora. Then, these scores are summed for each word in document d, to obtain a sentiment value for the document.

In contrast, $(\Delta tf)idf(w_i)$ of word w_i considers the difference between the *term* frequencies of the word w_i in positive and negative reviews.

$$\text{Delta TFIDF}(w_i, d) = tf(w_i, d) \times \quad (5)$$

$$[idf(w_i, +) - idf(w_i, -)]$$

In this process we excluded words with POS tags containing "PRP" or "DT" to exclude stop words such as "the","I","a", etc. A portion of the resulting features are shown in Table II.

B. Updating word polarities

When we observe a disagreement between the SentiWordNet polarity and the $(\Delta tf)idf$ score of a word, we consider changing its polarity. For instance in Table II, the word "joke" has a positive polarity, while its $(\Delta tf)idf$ score is negative, indicating that it occurs more in negative reviews. Similarly, the word "comfy" has a negative polarity, while its $(\Delta tf)idf$ score is positive, indicating that it occurs more in positive reviews.

For deciding on the new polarity of a word where a mismatch is observed, there are a couple of alternatives:

- *Flip*: Using the opposite polarity of the word (if the negative polarity of a word was dominant, we switch to its positive polarity and vice versa)
- *ObjectiveFlip*: Switching the objective polarity words to either negative or positive; similarly switching the negative or positive word to objective instead of its opposite polarity as done in *Flip*.
- *Shift*: Shifting the polarity of a word toward the other pole (as in adverbs, [18], [19].)
- *DeltaScore*: Computing the new polarity based on the $(\Delta tf)idf$ score of the word.

We tried two of these combinations (*Flip* and *DeltaScore*) and report results in large databases in two separate domains, as described in Section 4. As can be seen in Table VI, in our experiments we observed that Flip has updated the polarity of the word "joke" in the TripAdvisor dataset [20]. The SentiWordNet dominant polarity of the word "joke" was +0.53 and the updated polarity is -0.41. Indeed, the word "joke" appeared in a sentences like "Check in was a joke,...." where it contributes to negative sentiment.

C. Extent of the Updates

For choosing how many words to update, there can be a couple of different alternatives:

- *Top-k%*: changing the polarity of the top-k% of the words showing a mismatch. For this option, we ranked the words with respect to decreasing $|(\Delta(tf))idf|$ scores and examined the top-k% of the list for sign disagreements between the $(\Delta tf)idf$ scores and the SentiWordNet polarity.
- *Threshold*: changing the polarity of all the words below/above a fixed threshold where a disagreement occurs (e.g. $(\Delta tf)idf < -10$ OR $(\Delta tf)idf > 10$)
- *Iterative*: changing the polarity of a word one at a time using hill-climbing, where the change would be accepted if it was found to improve accuracy on the validation set.

We tried all three approaches, but report the first two as the third option is too slow and not better than the others. For the *Top-k%* selection, we tried top-5 and top-10%. For the *Threshold* selection, we tried two runs with different positive and negative threshold value ranges that will enable a good number of words to be picked.

Notice that all of these update methods can also include the *ObjectiveFlip* approach where *Top-k%* would be modified to have a *Middle-k%* and *Threshold* would have two threshold

TABLE II
$(\Delta tf)idf$ COMPUTATION

w_i	$tf(w_i,+)$	$tf(w_i,-)$	$idf(w_i)$	$\Delta(tf)(w_i)$	$Pol(w_i)$	Result
treat	21	0	4.96	15.34	0.77	
exceeded	15	0	5.23	14.51	0	
neat	19	2	5.01	9.51	0.48	Agreement
dirty	19	249	2.65	-6.7	-0.47	
smelly	0	24	4.79	-15.41	-0.75	
fireplace	37	1	4.57	13.46	0	
comfy	72	13	3.64	6.01	-0.75	
pleased	41	11	4.09	5.13	-0.5	Disagreement
joke	5	40	4.29	-8.25	0.53	

values to determine the middle range of words such that $2 < (\Delta tf)idf < -1$.

IV. Experimental Evaluation

We implemented and tested the proposed polarity adaptation techniques on real review data sets to assess their impact on the overall sentiment classification. Following sections detail the data sets used and the results obtained on these data sets.

A. Data Sets

We tested our system in two different domains: hotel and movie reviews. For the hotel domain, we extracted 6000 reviews from a snapshot of the TripAdvisor site which was prepared by [20]. In order to have an equal number of positive and negative reviews, we randomly collected samples from this larger set, resulting in 1500 positive and 1500 negative reviews in Train and Test datasets. In the hotel reviews, a review may have a rating of 1,2,3,4,5 where we assume that the reviews with rating 1, and 2 are negative and the reviews with rating 4, and 5 are positive. We did not consider the reviews with rating 3 for training since they do not have a strong polarity.

For the movie review domain, we use the movie review data provided by [21], consisting of 2000 reviews. In order to have an equal number of positive and negative reviews, we randomly split this database into two sets, Train and Test datasets, each containing 500 positive and 500 negative reviews. In the movie reviews, a review is marked with label ”-” to indicate that it is a negative review and marked with ”+” to indicate that it is a positive review.

B. Results

The results for two datasets are shown in Tables III and IV. We have tried Flip, and DeltaScore updating methods, with top-5% and top-10% of all the words. We also tried the Threshold update with different threshold values for picking the words to flip.

As can be seen in these tables, the updates all show improvement over the alternative of using SentiWordNet polarity values without doing any adaptation. These improvements are comparable to those in [11] where around 2% accuracy had been obtained using an adaptation done by linear programming.

TABLE VII
POLARITY SCORES: BEFORE AND AFTER UPDATE

Word	Type	SentiWordnet	Flip	DeltaScore
failed	VBD	0	-0.28	-0.73
garbage	NN	0	-0.125	-0.47
joke	NN	0.53	-0.19	-0.41
ludicrous	JJ	0.56	-0.125	-0.36
implausible	JJ	0.44	-0.25	-0.27
laughable	JJ	0.56	0	-0.21
courage	NN	-0.5	0.375	0.22
comfy	JJ	-0.75	0	0.30
complicated	JJ	-0.625	0.125	0.32
sufficient	JJ	-0.75	0.125	0.50
treat	NN	0	0.25	0.77
gem	NN	0	0.15	0.81

V. Conclusion and Future Work

This work aimed at finding out how we can adapt an existing polarity lexicon to a specific domain by learning new polarity orientations for the words by looking at how they are used in a particular domain. Although the proposed method is very simple and efficient, the use of the resulting adapted lexicons have increased review sentiment classification accuracy in both of the tested domains.

For future work, we are going to test the proposed methods on a larger dataset in different domains and with more lexicons. We also plan to incorporate this polarity adaptation approach to our open source sentiment analysis system SARE [22], which may be accessed through http://sentilab.sabanciuniv.edu/sare.

References

[1] V. Hatzivassiloglou and K. R. Mckeown, “Predicting the semantic orientation of adjectives,” in *Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 1997, pp. 174–181.

[2] J. M. Wiebe, “Learning subjective adjectives from corpora,” in *AAAI*, 2000, pp. 735–740.

[3] P. Turney, “Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews,” 2002, pp. 417–424.

[4] S. R. Das, M. Y. Chen, T. V. Agarwal, C. Brooks, Y. shee Chan, D. Gibson, D. Leinweber, A. Martinez-jerez, P. Raghubir, S. Rajagopalan, A. Ranade, M. Rubinstein, and P. Tufano, “Yahoo! for amazon: Sentiment extraction from small talk on the web,” in *8th Asia Pacific Finance Association Annual Conference*, 2001.

[5] S. min Kim, “Determining the sentiment of opinions,” in *Proceedings of COLING*, 2004, pp. 1367–1373.

[6] S. Baccianella, A. Esuli, and F. Sebastiani, “Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining,” in *Proc. of LREC*, 2010.

TABLE III
CLASSIFICATION RATES ON TRIPADVISOR DATASET

Update Method	Training	Test	Training (no 3-stars)	Test (no 3-stars)
None (using SentiWordnet)	76.03	75.13	78.10	77.25
After %5 *Flip*	77.33	75.87	79.15	77.76
After %10 *Flip*	78.23	76.53	80.94	79.32
After %5 *DeltaScore*	80.4	78.03	82.16	80.12
After %10 *DeltaScore*	82.37	**80.27**	84.85	**82.72**
After Threshold (≥ 5 or ≤ -10) *Flip*	77.80	76.33	79.93	78.30
After Threshold (≥ 5 or ≤ -5) *Flip*	78.27	76.53	80.94	79.32

TABLE IV
CLASSIFICATION RATES ON THE MOVIE DATASET

Update Method	Training	Test
None (using SentiWordnet)	60.0	61.3
After %5 *Flip*	60.8	62.6
After %10 *Flip*	62.7	63.9
After %5 *DeltaScore*	68.9	64.1
After %10 *DeltaScore*	73.0	**65.8**
After Threshold (≥ 10 or ≤ -5) *Flip*	60.5	62.0
After Threshold (≥ 5 or ≤ -5) *Flip*	61.6	63.1

TABLE V
EXAMPLE WORD FLIPS FROM TRIPADVISOR DATASET

Word	SentiWordNet	Updated	Review Context	Review Polarity
failed	0	-0.73	The speciality restaurants tried to be americanized but **failed** horribly.	2
joke	0.53	-0.41	Check in was a **joke**, our room ...	1
comfy	-0.75	0.30	The room was big enough with a large **comfy** bed.	5
sufficient	-0.75	0.49	Simple but **sufficient** complimentary breakfast (coffee, fruit, ...) left us satisfied.	4
treat	0	0.77	The lounge was a wonderful **treat** each morning and afternoon.	5
gem	0	0.81	What a **gem**!	5

TABLE VI
EXAMPLE WORD FLIPS FROM MOVIE DATASET

Word	SentiWordNet	DeltaScore	Review Context	Review Polarity
garbage	0	-0.47	... and i tend to shy away from watching such **garbage**...	-
ludicrous	0.56	-0.36	... are so over the top, nonstop, and too **ludicrous** for words...	-
implausible	0.44	-0.27	This movie was just completely **implausible**...	-
laughable	0.56	-0.21	This is a **laughable** 1977 rip off of king kong (1976), ...	-
courage	-0.5	0.22	Few filmmakers have the **courage** and sheer audacity...	+
complicated	-0.625	0.32	It is notable for introducing one of the first **complicated** gay characters...	+
jarring	-0.625	0.38	The tune is haunting, but it is also completely **jarring**.	+

[7] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis." in *Mining Text Data*, C. C. Aggarwal and C. Zhai, Eds. Springer, 2012, pp. 415–463.

[8] C. Fellbaum, Ed., *WordNet An Electronic Lexical Database*. Cambridge, MA ; London: The MIT Press, May 1998. [Online]. Available: http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=8106

[9] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," *Computational Linguistics*, pp. 399–433, 2009.

[10] G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding domain sentiment lexicon through double propagation," in *Proceedings of the 21st international jont conference on Artifical intelligence*, ser. IJCAI'09. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2009, pp. 1199–1204.

[11] Y. Choi and C. Cardie, "Adapting a polarity lexicon using integer linear programming for domain specific sentiment classification," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 2009, pp. 590–598.

[12] E. C. Dragut, C. Yu, P. Sistla, and W. Meng, "Construction of a sentimental word dictionary," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, ser. CIKM '10. New York, NY, USA: ACM, 2010, pp. 1761–1764.

[13] Y. Lu, M. Castellanos, U. Dayal, and C. Zhai, "Automatic construction of a context-aware sentiment lexicon: an optimization approach," in *Proceedings of the 20th international conference on World wide web*, ser. WWW '11. New York, NY, USA: ACM, 2011, pp. 347–356.

[14] J. Martineau and T. Finin, "Delta TFIDF: An improved feature space for sentiment analysis," in *ICWSM*, 2009.

[15] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, ser. HLT '05. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 339–346.

[16] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," in *Communications of the ACM*, 1975, pp. 613–620.

[17] G. Paltoglou, S. Gobron, M. Skowron, M. Thelwall, and D. Thalmann, "Sentiment analysis of informal textual communication in cyberspace," 2010.

[18] S. Li, S. Y. M. Lee, Y. Chen, C.-R. Huang, and G. Zhou, "Sentiment classification and polarity shifting," in *Proceedings of the 23rd International Conference on Computational Linguistics*, ser. COLING '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 635–643.

[19] D. Ikeda, H. Takamura, L. arie Ratinov, and M. Okumura, "Learning to shift the polarity of words for sentiment classification," in *Proceedings*

of the Third International Joint Conference on Natural Language Processing (IJCNLP, 2008.

[20] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data: A rating regression approach," *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 783–792, 2010.

[21] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the ACL*, 2004, pp. 271–278.

[22] M. Hussaini, A. Kocyigit, D. Tapucu, B. Yanikoglu, and Y. Saygin, "An aspect-lexicon creation and evaluation tool for sentiment analysis researchers," in *ECMLPKDD*, 2012.