Identifying Opinion Subgroups in Arabic Online Discussions

Amjad Abu-Jbara
Department of EECS
University of Michigan
Ann Arbor, MI, USA
amjbara@umich.edu

Ben King
Department of EECS
University of Michigan
Ann Arbor, MI, USA
benking@umich.edu

Mona Diab
Department of Computer Science
George Washington University
Washington DC, USA
mtdiab@gwu.edu

Dragomir Radev
Department of EECS
University of Michigan
Ann Arbor, MI, USA
radev@umich.edu

Abstract

In this paper, we use Arabic natural language processing techniques to analyze Arabic debates. The goal is to identify how the participants in a discussion split into subgroups with contrasting opinions. The members of each subgroup share the same opinion with respect to the discussion topic and an opposing opinion to the members of other subgroups. We use opinion mining techniques to identify opinion expressions and determine their polarities and their targets. We opinion predictions to represent the discussion in one of two formal representations: signed attitude network or a space of attitude vectors. We identify opinion subgroups by partitioning the signed network representation or by clustering the vector space representation. We evaluate the system using a data set of labeled discussions and show that it achieves good results.

1 Introduction

Arabic is one of the fastest growing languages on the internet. The number of internet users in the Arab region grew by 2500% over the past 10 years. As of January 2012, the number of Arabic-speaking internet users was 86 millions. The recent political and civic movements in the Arab World resulted in a revolutionary growth in the number of Arabic users on social networking sites. For example, Arabic is the fastest growing language in Twitter history [1].

This growth in the presence of Arab users on social networks and all the interactions and discussions that happen among them led to a huge amount of opinion-rich Arabic text being available. Analyzing this text could reveal the different viewpoints of Arab users with respect to the topics that they discuss online.

When a controversial topic is discussed, it is normal for the discussants to adopt different viewpoints towards it. This usually causes rifts in discussion groups and leads to the split of the discussants into subgroups with contrasting opinions. Our goal in this paper is to use natural language processing techniques to detect opinion subgroups in Arabic discussions. Our approach starts by identifying opinionated (subjective) text and determining its polarity (positive, negative, or neutral). Next, we determine the target of each opinion expression. The target of opinion can be a named entity mentioned in the discussion or an aspect of the discussed topic. We use the identified opinion-target relations to represent the discussion in one of two formal representations. In the first representation, each discussant is represented by a vector that encodes all his or her opinion information towards the discussion topic. In the second representation, each discussant is represented by a node in a signed graph. A positive edge connects two discussants if they have similar opinion towards the topic, otherwise the sign of the edge is nega-

[1] http://semiocast.com/publications/2011_11_24_Arabic_highest_growth_on_Twitter

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 829–835,
Sofia, Bulgaria, August 4-9 2013.

tive. To identify opinion subgroups, we cluster the vector space (the first representation) or partition the signed network (the second representation).

We evaluate this system using a data set of Arabic discussions collected from an Arabic debating site. We experiment with several variations of the system. The results show that the clustering the vector space representation achieves better results than partitioning the signed network representation.

2 Previous Work

Our work is related to a large body of research on opinion mining and sentiment analysis. Pang & Lee (2008) and Liu & Zhang (2012) wrote two recent comprehensive surveys about sentiment analysis and opinion mining techniques and applications.

Previous work has proposed methods for identifying subjective text that expresses opinion and distinguishing it from objective text that presents factual information (Wiebe, 2000; Hatzivassiloglou and Wiebe, 2000a; Banea et al., 2008; Riloff and Wiebe, 2003).

Subjective text may express positive, negative, or neutral opinion. Previous work addressed the problem of identifying the polarity of subjective text (Hatzivassiloglou and Wiebe, 2000b; Hassan et al., 2010; Riloff et al., 2006). Many of the proposed methods for text polarity identification depend on the availability of polarity lexicons (i.e. lists of positive and negative words). Several approaches have been devised for building such lexicons (Turney and Littman, 2003; Kanayama and Nasukawa, 2006; Takamura et al., 2005; Hassan and Radev, 2010). Other research efforts focused on identifying the holders and the targets of opinion (Zhai et al., 2010; Popescu and Etzioni, 2007; Bethard et al., 2004).

Opinion mining and sentiment analysis techniques have been used in various applications. One example of such applications is identifying perspectives (Grefenstette et al., 2004; Lin et al., 2006) which is most similar to the topic of this paper. For example, in (Lin et al., 2006), the authors experiment with several supervised and statistical models to capture how perspectives are expressed at the document and the sentence levels. Laver et al. (2003) proposed a method for extracting perspectives from political texts. They used their method to estimate the policy positions of political parties in Britain and Ireland, on both economic and social policy dimensions.

Somasundaran and Wiebe (2009) present an unsupervised opinion analysis method for debate-side classification. They mine the web to learn associations that are indicative of opinion stances in debates and combine this knowledge with discourse information. Anand et al. (2011) present a supervised method for stance classification. They use a number of linguistic and structural features such as unigrams, bigrams, cue words, repeated punctuation, and opinion dependencies to build a stance classification model. In previous work, we proposed a method that uses participant-to-participant and participant-to-topic attitudes to identify subgroups in ideological discussions using attitude vector space clustering (Abu-Jbara and Radev, 2012). In this paper, we extend this method by adding latent similarity features to the attitude vectors and applying it to Arabic discussions. In another previous work, our group proposed a supervised method for extracting signed social networks from text (Hassan et al., 2012a). The signed networks constructed using this method were based only on participant-to-participant attitudes that are expressed explicitly in discussions. We used this method to extract signed networks from discussions and used a partitioning algorithm to detect opinion subgroups (Hassan et al., 2012b). In this paper, we extend this method by using participant-to-topic attitudes to construct the signed network.

Unfortunately, not much work has been done on Arabic sentiment analysis and opinion mining. Abbasi et al. (2008) applies sentiment analysis techniques to identify and classify document-level opinions in text crawled from English and Arabic web forums. Hassan et al. (2011) proposed a method for identifying the polarity of non-English words using multilingual semantic graphs. They applied their method to Arabic and Hindi. Abdul-Mageed and Diab (2011) annotated a corpus of Modern Standard Arabic (MSA) news text for subjectivity at the sentence level. In a later work (2012a), they expanded their corpus by labeling data from more genres using Amazon Mechanical Turk. Abdul-Mageed et al. (2012a) developed SAMAR, a system for subjectivity and Sentiment Analysis for Arabic social media genres. We use this system as a component in our approach.

3 Approach

In this section, we present our approach to detecting opinion subgroups in Arabic discussions. We propose a pipeline that consists of five components. The input to the pipeline is a discussion thread in Arabic language crawled from a discussion forum. The output is the list of participants in the discussion and the subgroup membership of each discussant. We describe the components of the pipeline in the following subsections.

3.1 Preprocessing

The input to this component is a discussion thread in HTML format. We parse the HTML file to identify the posts, the discussants, and the thread structure. We transform the Arabic content of the posts and the discussant names that are written in Arabic to the Buckwalter encoding (Buckwalter, 2004). We use AMIRAN (Diab, 2009), a system for processing Arabic text, to tokenize the text and identify noun phrases.

3.2 Identifying Opinionated Text

To identify opinion-bearing text, we start from the word level. We identify the polarized words that appear in text by looking each word up in a lexicon of Arabic polarized words. In our experiments, we use Sifat (Abdul-Mageed and Diab, 2012b), a lexicon of 3982 Arabic adjectives labeled as positive, negative, or neutral.

The polarity of a word may be dependant on its context (Wilson et al., 2005). For example, a positive word that appears in a negated context should be treated as expressing negative opinion rather than positive. To identify the polarity of a word given the sentence it appears in, we use SAMAR (Abdul-Mageed et al., 2012b), a system for Subjectivity and Sentiment Analysis for Arabic social media genres. SAMAR labels a sentence that contains an opinion expression as positive, negative, or neutral taking into account the context of the opinion expression. The reported accuracy of SAMAR on different data sets ranges between 84% and 95% for subjectivity classification and 65% and 81% for polarity classification.

3.3 Identifying Opinion Targets

In this step, we determine the targets that the opinion is expressed towards. We treat as an opinion target any noun phrase (NP) that appears in a sentence that SAMAR labeled as polarized (positive or negative) in the previous step. To avoid the noise that may result from including all noun phrases, we limit what we consider as an opinion target, to the ones that appear in at least two posts written by two different participants. Since, the sentence may contain multiple possible targets for every opinion expression, we associate each opinion expression with the target that is closest to it in the sentence. For each discussant, we keep track of the targets mentioned in his/her posts and the number of times each target was mentioned in a positive/negative context.

3.4 Latent Textual Similarity

If two participants share the same opinion, they tend to focus on similar aspects of the discussion topic and emphasize similar points that support their opinion. To capture this, we follow previous work (Guo and Diab, 2012; Dasigi et al., 2012) and apply Latent Dirichelet Allocation (LDA) topic models to the text written by the different participants. We use an LDA model with 100 topics. So, we represent all the text written in the discussion by each participant as a vector of 100 dimensions. The vector of each participant contains the topic distribution of the participant, as produced by the LDA model.

3.5 Subgroup Detection

At this point, we have for every discussant the targets towards which he/she expressed explicit opinion and a 100-dimensions vector representing the LDA distribution of the text written by him/her. We use this information to represent the discussion in two representations. In the first representation, each discussant is represented by a vector. For every target identified in steps 3 of the pipeline, we add three entries in the vector. The first entry holds the total number of times the target was mentioned by the discussant. The second entry holds the

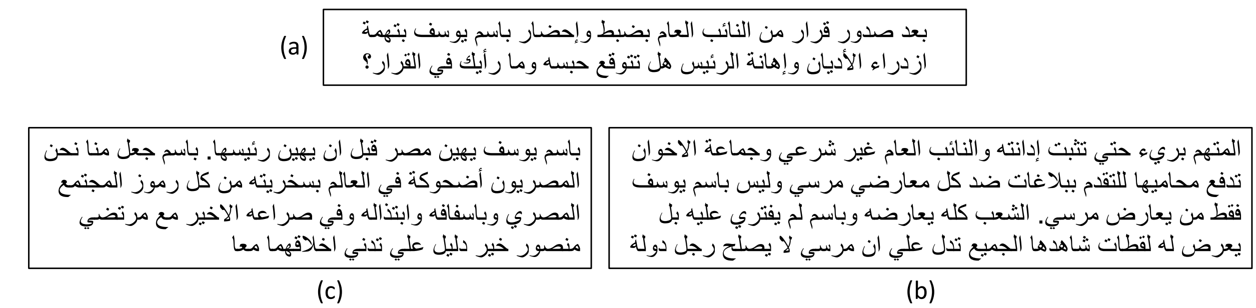

Figure 1: An example debate taken from our dataset. (a) is the discussion topic. (b) and (c) are two posts expressing contrasting viewpoints with respect to the topic.

number of times the target was mentioned in a positive context. The third entry holds the number of target mentions in a negative context. We also add to this vector the 100 topic entries from the LDA vector of that discussant. So, if the number of targets identified in step 3 of the pipeline is t then the number of entries in the discussant vector is $3 * t + 100$.

To identify opinion subgroups, we cluster the vector space. We experiment with several clustering algorithms including K-means (MacQueen, 1967), Farthest First (FF) (Hochbaum and Shmoys, 1985; Dasgupta, 2002), and Expectation Maximization (EM) (Dempster et al., 1977).

The second representation is a signed network representation. In this representation, each discussant is represented by a node in a graph. Two discussants are connected by an edge if they both mention at least one common target in their posts. If a discussant mentions a target multiple times in different contexts with different polarities, the majority polarity is assumed as the opinion of this discussant with respect to this target. A positive sign is assigned to the edge connecting two discussants if the number of targets that they have similar opinion towards is greater than the targets that they have opposing opinion towards, otherwise a negative sign is assigned to the edge.

To identify subgroups, we use a signed network partitioning algorithm to partition the network. Each resulting partition constitutes a subgroup. Following (Hassan et al., 2012b), we use the Dorian-Mrvar (1996) algorithm to partition the signed network. The optimization criterion aims to have dense positive links within groups and dense negative links between groups.

The optimization function is as follows:

$$F(C) = \alpha \times |NEG| + (1 - \alpha) \times |POS| \quad (1)$$

where C is the clustering under evaluation, $|NEG|$ is the number of negative links between nodes in the same subgroup, $|POS|$ is the number of positive links between nodes in different subgroups, and α is a parameter that specifies importance of the two terms. We set α to 0.5 in all our experiments.

Clusters are selected such that $P(C)$ is minimum. The best clustering that minimizes $P(C)$ is found by moving nodes around clusters in a greedy way starting with a random clustering. To handle the possibility of finding a local minima, the whole process is repeated k times with random restarts and the clustering with the minimum value of $P(C)$ is returned. We set k to 3 in all our experiments.

4 Data

We use data from an Arabic discussion forum called Naqeshny.[2]. Naqeshny is a platform for two-sided debates. The debate starts when a person asks a question (e.g. which political party do you support?) and gives two possible answers or positions. The registered members of the website who are interested in the topic participate in the debate by selecting a position and then posting text to support that position and dispute the

[2] www.Naqeshny.com

opposing position. This means that the data set is self-labeled for subgroup membership. Since the tools used in our system are trained on Modern Standard Arabic (MSA) text, we selected debates that are mostly MSA. The data set consists of 36 debates comprising a total of 711 posts written by 326 users. The average number of posts per discussion is 19.75 and the average number of participants per discussion is 13.08. Figure 1 shows an example from the data.

5 Experiments and Results

We use three metrics to evaluate the resulting subgroups: Purity (Manning et al., 2008), Entropy, and F-measure. We ran several variations of the system on the data set described in the previous section. In one variation, we use the signed network partitioning approach to detect subgroups. In the other variations, we use the vector space clustering approach. We experiment with different clustering algorithms. We also run two experiments to evaluate the contribution of both opinion-target counts and latent similarity features on the clustering accuracy. In one run, we use target-opinion counts only. In the other run, we use latent similarity features only. EM was used as the clustering algorithm in these two runs. Table 1 shows the results. All the results have been tested for statistical significance using a 2-tailed paired t-test. The differences between the results of the different methods shown in the table are statistically significant at the 0.05 level. The results show that the clustering approach achieves better results than the signed network partitioning approach. This can be explained by the fact that the vector representation is a richer representation and encodes all the discussants' opinion information explicitly. The results also show that Expectation Maximization achieves significantly better results than the other clustering algorithms that we experimented with. The results also show that both latent text similarity and opinion-target features are important and contribute to the performance.

System	**Purity**	**F-Measure**	**Entropy**
Signed Network	0.71	0.67	0.68
Clustering - K-means	0.72	0.70	0.67
Clustering - EM	**0.77**	**0.76**	**0.50**
Clustering - FF	0.72	0.69	0.70
Opinion-Target Only	0.67	0.65	0.72
Text Similarity Only	0.64	0.65	0.74

Table 1: Comparison of the different variations of the proposed approach

6 Conclusion

In this paper, we presented a system for identifying opinion subgroups in Arabic online discussions. The system uses opinion and text similarity features to encode discussants' opinions. Two approaches were explored for detecting subgroups. The first approach clusters a space of discussant opinion vectors. The second approach partitions a signed network representation of the discussion. Our experiments showed that the former approach achieves better results. Our experiments also showed that both opinion and similarity features are important.

Acknowledgements

This research was funded in part by the Office of the Director of National Intelligence, Intelligence Advanced Research Projects Activity. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the of?cial views or policies of IARPA, the ODNI or the U.S. Government.

The authors would like to thank Basma Siam for her help with collecting the data.

References

Ahmed Abbasi, Hsinchun Chen, and Arab Salem. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst.*, 26(3):12:1–12:34, June.

Muhammad Abdul-Mageed and Mona Diab. 2011. Subjectivity and sentiment annotation of modern standard arabic newswire. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 110–118, Portland, Oregon, USA, June. Association for Computational Linguistics.

Muhammad Abdul-Mageed and Mona Diab. 2012a. Awatif: A multi-genre corpus for modern standard arabic subjectivity and sentiment analysis. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios

Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).

Muhammad Abdul-Mageed and Mona Diab. 2012b. Toward building a large-scale arabic sentiment lexicon. In *Proceedings of the 6th International Global Word-Net Conference, Matsue, Japan.*

Muhammad Abdul-Mageed, Sandra Kübler, and Mona Diab. 2012a. Samar: a system for subjectivity and sentiment analysis of arabic social media. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '12, pages 19–28, Stroudsburg, PA, USA. Association for Computational Linguistics.

Muhammad Abdul-Mageed, Sandra Kuebler, and Mona Diab. 2012b. Samar: A system for subjectivity and sentiment analysis of arabic social media. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 19–28, Jeju, Korea, July. Association for Computational Linguistics.

Amjad Abu-Jbara and Dragomir Radev. 2012. Subgroup detection in ideological discussions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Jeju, Korea, July. The Association for Computational Linguistics.

Pranav Anand, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 1–9, Portland, Oregon, June. Association for Computational Linguistics.

Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2008. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In *LREC'08*.

Steven Bethard, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Dan Jurafsky. 2004. Automatic extraction of opinion propositions and their holders. In *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*, page 2224.

Tim Buckwalter. 2004. Issues in arabic orthography and morphology analysis. In *Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages*, Semitic '04, pages 31–34, Stroudsburg, PA, USA. Association for Computational Linguistics.

Sanjoy Dasgupta. 2002. Performance guarantees for hierarchical clustering. In *15th Annual Conference on Computational Learning Theory*, pages 351–363. Springer.

Pradeep Dasigi, Weiwei Guo, and Mona Diab. 2012. Genre independent subgroup detection in online discussion threads: A study of implicit attitude using textual latent semantics. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 65–69, Jeju Island, Korea, July. Association for Computational Linguistics.

A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38.

Mona Diab. 2009. Second generation tools (amira 2.0): Fast and robust tokenization, pos tagging, and base phrase chunking. In Khalid Choukri and Bente Maegaard, editors, *Proceedings of the Second International Conference on Arabic Language Resources and Tools*, Cairo, Egypt, April. The MEDAR Consortium.

Patrick Doreian and Andrej Mrvar. 1996. A partitioning approach to structural balance. *Social Networks*, 18(2):149–168.

Gregory Grefenstette, Yan Qu, James G Shanahan, and David A Evans. 2004. Coupling niche browsers and affect analysis for an opinion mining application. In *Proceedings of RIAO*, volume 4, pages 186–194. Citeseer.

Weiwei Guo and Mona Diab. 2012. Modeling sentences in the latent space. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 864–872, Jeju Island, Korea, July. Association for Computational Linguistics.

Ahmed Hassan and Dragomir Radev. 2010. Identifying text polarity using random walks. In *ACL'10*.

Ahmed Hassan, Vahed Qazvinian, and Dragomir Radev. 2010. What's with the attitude?: identifying sentences with attitude in online discussions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1245–1255.

Ahmed Hassan, Amjad Abu-Jbara, Rahul Jha, and Dragomir Radev. 2011. Identifying the semantic orientation of foreign words. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2*, HLT '11, pages 592–597, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. 2012a. Extracting signed social networks from text. In *Workshop Proceedings of TextGraphs-7: Graph-based Methods for Natural Language Processing*, pages 6–14, Jeju, Republic of Korea, July. Association for Computational Linguistics.

Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. 2012b. Signed attitude networks: Predicting positive and negative links using linguistic analysis. In *Submitted to the Conference on Emprical Methods in Natural Language Processing*, Jeju, Korea, July. The Association for Computational Linguistics.

Vasileios Hatzivassiloglou and Janyce Wiebe. 2000a. Effects of adjective orientation and gradability on sentence subjectivity. In *COLING*, pages 299–305.

Vasileios Hatzivassiloglou and Janyce M Wiebe. 2000b. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th conference on Computational linguistics-Volume 1*, pages 299–305. Association for Computational Linguistics.

Hochbaum and Shmoys. 1985. A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10(2):180–184.

Hiroshi Kanayama and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *EMNLP'06*, pages 355–363.

Michael Laver, Kenneth Benoit, and John Garry. 2003. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(02):311–331.

Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on?: identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, pages 109–116. Association for Computational Linguistics.

Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In Charu C. Aggarwal and ChengXiang Zhai, editors, *Mining Text Data*, pages 415–463. Springer US.

J. B. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.

Ana-Maria Popescu and Orena Etzioni. 2007. Extracting product features and opinions from reviews. In *Natural language processing and text mining*, pages 9–28. Springer.

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *EMNLP'03*, pages 105–112.

Ellen Riloff, Siddharth Patwardhan, and Janyce Wiebe. 2006. Feature subsumption for opinion analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 440–448. Association for Computational Linguistics.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 226–234, Suntec, Singapore, August. Association for Computational Linguistics.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *ACL'05*, pages 133–140.

Peter Turney and Michael Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21:315–346.

Janyce Wiebe. 2000. Learning subjective adjectives from corpora. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 735–740.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT/EMNLP'05*, Vancouver, Canada.

Zhongwu Zhai, Bing Liu, Hua Xu, and Peifa Jia. 2010. Grouping product features using semi-supervised learning with soft-constraints. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1272–1280. Association for Computational Linguistics.