Towards Cross-Language Sentiment Analysis through Universal Star Ratings

Alexander Hogenboom, Malissa Bal, Flavius Frasincar, and Daniella Bal

Abstract The abundance of sentiment-carrying user-generated content renders automated cross-language information monitoring tools crucial for today's businesses. In order to facilitate cross-language sentiment analysis, we propose to compare the sentiment conveyed by unstructured text across languages through universal star ratings for intended sentiment. We demonstrate that the way natural language reveals people's intended sentiment differs across languages. The results of our experiments with respect to modeling this relation for both Dutch and English by means of a monotone increasing step function mainly suggest that language-specific sentiment scores can separate universal classes of intended sentiment from one another to a limited extent.

Key words: Sentiment analysis, cross-language analysis, star ratings

1 Introduction

Today's Web enables people to produce an ever-growing amount of virtual utterances of opinions in any language. Anyone can write reviews and blogs, post messages on discussion forums, or publish whatever crosses one's mind on Twitter at any time. This yields a continuous flow of an overwhelming amount of multi-lingual data, containing traces of valuable information – people's sentiment with respect to products, brands, etcetera. As recent estimates indicate that one in three blog posts [15] and one in five tweets [12] discuss products or brands, the abundance of user-generated content published through such social media renders automated cross-language information monitoring tools crucial for today's businesses.

Alexander Hogenboom · Malissa Bal · Flavius Frasincar · Daniella Bal
Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR, Rotterdam, the Netherlands, e-mail: `hogenboom@ese.eur.nl`, `malissa.bal@xs4all.nl`, `frasincar@ese.eur.nl`, `daniella.bal@xs4all.nl`

Sentiment analysis comes to answer this need. Sentiment analysis refers to a broad area of natural language processing, computational linguistics, and text mining. Typically, the goal is to determine the polarity of natural language text. An intuitive approach involves scanning a text for cues signaling its polarity. Existing approaches typically consist of language-specific parts such as sentiment lexicons, i.e., lists of words and their associated sentiment, possibly differentiated by Part-of-Speech (POS) and/or meaning [2], or components for, e.g., identifying the POS or lemma of words. Yet, each language-specific sentiment analysis approach typically produces sentiment scores for texts in its reference language. Intuitively, these scores should be comparable across languages, irrespective of the techniques used to get these scores – provided that these techniques adhere to the same constraints in that they, e.g., produce a score on a continuous scale between -1 (negative) and 1 (positive). However, sentiment scores are not directly comparable across languages, as they tend to be affected by many different language-specific phenomena [3].

Therefore, we propose to perform cross-language sentiment analysis not by using the sentiment scores associated with natural language content per se, but by involving another way of measuring sentiment – sentiment classification by means of star ratings. In such ratings, which are commonly used in, e.g., reviews, a more positive sentiment towards the topic of a text is typically reflected by a higher number of stars associated with the text. Sentiment scores are affected by the way people express themselves in natural language, whereas star ratings are (universal) classifications of the sentiment that people actually intend to convey.

In this paper, we aim to gain insight in the relation between language-specific sentiment scores and universal star ratings in order to be able to compare sentiment scores across languages. As such, we benefit from the robust and fine-grained type of analysis that traditional, lexicon-based sentiment analysis techniques offer [24], while using universal star ratings in order to scale the obtained language-specific sentiment scores so that sentiment normalization across languages can be realized.

The remainder of this paper is structured as follows. First, we discuss related work on cross-language sentiment analysis in Sect. 2. We then propose a method for making language-specific sentiment scores comparable across languages through universal classifications of intended sentiment in Sect. 3. A discussion of insights following from an evaluation of our method is presented in Sect. 4. Last, we conclude and propose directions for future work in Sect. 5.

2 Cross-Language Sentiment Analysis

In a recent literature survey on sentiment analysis [21], the current surge of research interest in systems that deal with opinions and sentiment is attributed to the fact that, in spite of today's users' hunger for and reliance upon on-line advice and recommendations, explicit information on user opinions is often hard to find, confusing, or overwhelming. Many sentiment analysis approaches exist, yet the topic of cross-language sentiment analysis has been relatively unexplored.

Among popular bag-of-word approaches to sentiment analysis in an arbitrary language (typically English), a binary encoding of text, indicating the presence or absence of specific words, has initially proven to be an effective representation [20]. Later research has focused on different vector representations of text, including vector representations with additional features representing semantic distinctions between words [26] or vector representations with sophisticated weighting schemes for word features [19]. Such vector representations are typically used by machine learning algorithms in order to score a piece of natural language text for its associated sentiment or to classify it as either positive or negative.

The alternative lexicon-based approaches typically exhibit lower classification accuracy, but tend to be more robust across domains [24]. Also, lexicon-based approaches can be generalized relatively easily to other languages by using dictionaries [17]. Recently proposed lexicon-based sentiment analysis techniques range from rather simple [9, 10] to more sophisticated approaches that take into account structural or semantic aspects of content, for instance by means of a deeper linguistic analysis focusing on differentiating between rhetorical roles of text segments [8, 23, 24].

These existing approaches may work very well for the language they have been designed for, yet applying the state-of-the-art in sentiment analysis on entirely new languages has been shown to have its challenges, as each language may require another approach [18]. Existing research on sentiment analysis in different languages has been focused mainly on how to create new sentiment analysis methods with minimal effort, without losing too much accuracy with respect to classifying a text as either positive or negative. The focus of existing research varies from creating sentiment lexicons [11, 25] to constructing new sentiment analysis frameworks [1, 5, 6, 7, 18] for languages other than the reference language.

Moens and Boiy [18] have analyzed the creation of different sentiment analysis frameworks for different languages, while requiring minimal human effort for developing these frameworks. For any of their considered languages, Moens and Boiy [18] recommend a three-layer sentiment analysis framework in order to realize a fast way of computing sentiment scores. The first layer is very fast in its computations, yet does yield very accurate sentiment scores. When the result of a computation is not of a desired level of accuracy, the text is processed by the second, more precise, but also slower, computation layer. This process is repeated on the third layer. If still no accurate score is computed, the score of layer two is kept. The results of Moens and Boiy [18] indicate that the specifics of the configurations of such frameworks differ per language.

Rather than creating language-specific sentiment analysis frameworks, Bautin, Vijayarenu, and Skiena [4] have proposed to analyze cross-lingual sentiment by means of machine translation. They use machine translation in order to convert all considered texts into English and subsequently perform sentiment analysis on the translated results. By doing so, the authors assume that the results of the analysis on both the original text and the translated text are comparable and that the errors made by the machine translation do not significantly influence the results of the sentiment analysis.

However, the quality of machine translation may very well have an influence on the quality of the output of the sentiment analysis on the translated text, as low-quality translations do not typically form accurate representations of the original content and hence are not likely to convey the sentiment of the original text. Moreover, when focusing on developing sentiment lexicons for other languages by means of, for instance, machine translation, the need for distinct sentiment analysis approaches for different languages [18] is largely ignored.

Therefore, Bal et al. [3] have recently proposed a framework in which the sentiment in documents written in multiple languages can be assessed by means of language-specific sentiment analysis components. By means of this framework, the sentiment in documents has been compared with the sentiment conveyed by their translated counterparts. These experiments have shown that sentiment scores are not directly comparable across languages, as these scores tend to be affected by many different language-specific phenomena. In addition to this, Wierzbicka [27, 28] has argued that there is a cultural dimension to cross-language sentiment differences, as every language imposes its own classification upon human emotional experiences, thus rendering English sentiment-carrying words artifacts of the English language rather than culture-free analytical tools.

In this light, we are in need of a more universal way of capturing sentiment in order to be able to compare sentiment expressed in different languages. In this paper, we assume that star ratings can be used for this purpose. A higher number of stars associated with a piece of sentiment-carrying natural language text is typically associated with a more positive sentiment of the author towards the topic of this text. As such, star ratings are universal classifications of the sentiment that people actually intend to convey, whereas traditional sentiment scores tend to reflect the sentiment conveyed by the way people express themselves in natural language. Intuitively, both measures may be related to some extent, yet to the best of our knowledge, the relation between language-specific sentiment scores and universal sentiment classifications has not been previously investigated.

3 From Sentiment Scores to Star Ratings

As traditional sentiment analysis techniques are guided by the natural language used in texts, they allow for a fine-grained linguistic analysis of conveyed sentiment. In addition, they are rather robust as they take into account the actual content of a piece of natural language text, especially when involving structural and semantic aspects of content in the analysis [8, 23]. Yet, this language-dependency thwarts the cross-language comparability of sentiment thus identified.

Typically, traditional approaches are focused on assigning a sentiment score to a piece of natural language text, ranging from, e.g., -1 (negative) to 1 (positive). In order to support amplification of sentiment, such as "very good" rather than "good", sentiment scores may also range from, e.g., -1.5 (very negative) to 1.5 (very positive). Ideally, a sentiment score of for instance 0.7 would have the same meaning in both English and, e.g., Dutch, yet research has shown that this is not typically the case [3]. Therefore, in order to enable cross-language sentiment analysis while enjoying the benefits of traditional language-specific sentiment analysis approaches, a mapping from language-specific scores of conveyed sentiment to universal classifications of intended sentiment is of paramount importance.

In our current endeavors, we assume a five-star rating scale to be a universal classification method for an author's intended sentiment, i.e., consensus exists with respect to the meaning of each out of five classes. These classes are defined on an ordinal scale, i.e., a piece of text that is assigned five stars is considered to be more positive than a piece of text that belongs to the class of documents with four stars. Additionally, we assume that higher language-specific sentiment scores are associated with star ratings, which we model as a monotonically increasing step function. As such, we assume texts with, e.g., four stars to have higher sentiment scores than texts belonging to the three-star class.

Given these assumptions, we can construct language-specific sentiment maps for translating language-specific sentiment scores into universal star ratings. In each mapping, we consider five star segments, where we define a star segment as a set of sentiment-carrying natural language texts that have the same number of stars assigned to them. These five star segments are separated by a total of four boundaries, the position of which is based on the sentiment scores associated with the texts in each segment.

An intuitive sentiment map is depicted in Fig. 1. One could expect the one-star and five-star classes to be representing the extreme negative and positive cases, respectively, i.e., covering respective sentiment scores below -1 and above 1. The class of documents associated with three stars would intuitively be centered around a sentiment score of 0, indicating a more or less neutral sentiment. The classes of two-star and four-star texts would then cover the remaining ranges of negative and positive scores, respectively, in order to represent the rather negative and positive natural language texts, respectively. Many alternative mappings may exist for, e.g., different domains or languages. Mappings may for instance be skewed towards positive or negative sentiment scores or the boundaries may be unequally spread across the full range of sentiment scores.

The challenge is to find an optimal set of boundaries for each considered language in order to enable cross-language sentiment analysis by mapping language-specific sentiment scores, reflecting the sentiment conveyed by the way people express themselves in natural language, to universal star ratings, reflecting the intended sentiment. The goal of such an optimization process is to minimize the total costs c_b associated with a given set of boundaries b. We define these costs as the sum of the number of misclassifications $\varepsilon_i(b)$ in each individual sentiment class $i \in \{1, \ldots, 5\}$, given the set of boundaries b, i.e.,

$$c_b = \sum_{i=1}^{5} \varepsilon_i(b) . \tag{1}$$

Fig. 1 Intuitive mapping from sentiment conveyed by natural language to universal star ratings.

This optimization process, yielding a set of boundaries associated with the least possible number of misclassifications, is subject to the constraint that the boundaries must be non-overlapping and ordered, while being larger than the sentiment score lower bound s_l and smaller than the sentiment score upper bound s_u, i.e.,

$$s_l < b_1 < b_2 < b_3 < b_4 < s_u. \tag{2}$$

Finding an optimal set of boundaries is not a trivial task, as many combinations exist and the boundaries are moreover interdependent. Once an arbitrary boundary is set, it affects the possible locations of the other boundaries. Furthermore, classes may not be perfectly separable in the sole dimension of sentiment scores.

For example, let us consider the separation problem presented in Fig. 2, where documents in *Segment A* need to be separated from those in *Segment B* by means of boundary B. The two segments however exhibit some overlap, which prevents the segments from being perfectly separable. Yet, some solutions are better than others in this scenario. For instance, in the intersection of *Segment A* and *Segment B*, boundary B_1 would result in all five documents from *Segment A* being erroneously classified as *Segment B* documents, one *Segment B* document being classified as a *Segment A* document, and only three documents being classified correctly in *Segment B*. Boundary B_2 on the other hand would yield only three misclassifications in *Segment A*, one misclassification in *Segment B*, and two and three correct classifications in *Segment A* and *Segment B*, respectively.

Many algorithms can be used in order to cope with such issues. One may want to consider using a greedy algorithm in order to construct a set of boundaries. Alternatively, heuristic or randomized optimization techniques like genetic algorithms may be applied in order to explore the multitude of possible solutions. Finally, if the size of the data set allows, a brute force approach can be applied in order to assess all possible boundary sets at a certain level of granularity.

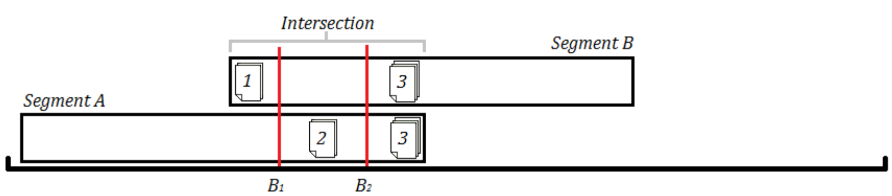

Fig. 2 Separating two segments of documents from one another by means of tentative boundaries B_1 and B_2. In the intersection of both segments, both *Segment A* and *Segment B* contain three documents with an equally high sentiment score. *Segment A* contains two additional documents with a lower score, whereas *Segment B* contains another document with an even lower score.

By using our proposed method, the sentiment conveyed by people's utterances of opinions in natural language can first be accurately analyzed by means of state-of-the-art tools tailored to the language of these texts. The sentiment scores thus obtained can subsequently be transformed into universal star ratings by means of language-specific sentiment maps. We can thus make language-specific sentiment scores comparable across languages by mapping these scores to universal classifications of intended sentiment.

4 Evaluation

Our proposed approach of making language-specific sentiment scores comparable across languages by means of mapping these scores to universal star ratings, can be used to perform several analyses, as depicted in Fig. 3. First, the sentiment score of documents can be compared across languages (1). This has already been done in previous research endeavors, which have revealed that this type of cross-language sentiment analysis is not the most promising one, as sentiment scores as such do not appear to be directly comparable across languages [3]. A more suitable analysis is an exploration of how language-specific sentiment scores can be converted into universal star ratings (2) and how such mappings differ across languages (3). Therefore, we focus on this type of analysis in this paper. The (interpretation of) star ratings could also be compared across languages (4), yet this falls outside of the scope of our current endeavors, as we assume star ratings to be universal and comparable across languages.

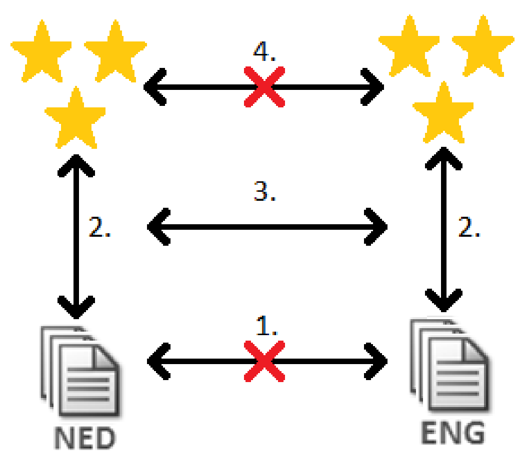

Fig. 3 Considered comparisons between Dutch (*NED*) and English (*ENG*) pieces of sentiment-carrying natural language text.

In our analysis, we consider two sets of similar documents. One set consists of 1,759 short movie reviews in Dutch, crawled from various web sites [13, 14]. The other collection consists of 46,315 short movie reviews in English [16, 22]. Each review in our data sets contains a maximum amount of 100 words. Each review has been rated by its respective writer on a scale of one to five or ten stars, depending on the web site, where more stars imply a more positive verdict. We have converted all document ratings to a five-star scale by dividing all scores on a ten-star scale by two and rounding the resulting scores to the nearest integer. This process results in a data set in which, for both considered languages, the documents are approximately normally distributed over five star classes, while being slightly skewed towards the higher classes.

The documents in our data set are first analyzed for the sentiment conveyed by their text by means of an existing framework for lexicon-based sentiment analysis in multiple languages, more specifically English and Dutch [3]. This framework is essentially a pipeline in which each component fulfills a specific task in analyzing the sentiment of an arbitrary document. For each supported language, this framework first prepares documents by cleaning the text – i.e., converting the text to lowercase, removing diacritics, etcetera – and performing initial linguistic analysis by identifying each word's POS as well as by distinguishing opinionated words and their modifiers from neutral words. After this preparation process, each document is scored by sum-aggregating the sentiment scores of the opinionated words, while taking into account their modifiers, if any.

Scoring each document in our data set for the sentiment conveyed by its text yields a set of 1,759 two-dimensional data points for Dutch and 46,315 similar two-dimensional data points for English, each of which represents a paired observation of a language-specific sentiment score and the associated universal star rating of intended sentiment. These data points can be used to construct a mapping between sentiment scores and star ratings for each considered language. As the size of our data set allows for it, we use a brute force approach in our current endeavors, where we assess the performance in terms of number of misclassifications for all possible combinations of boundaries, with a step size of 0.1.

The resulting sentiment score ranges per star class are reported in Table 1. These ranges are averages over all folds of our 10-fold cross-validation. The results in Table 1 indicate that sentiment maps may have different characteristics for different languages. For instance, the Dutch sentiment map appears to be more equally spread than the English sentiment map. Additionally, more than in Dutch documents, moderate sentiment scores in English documents are typically already associated with extreme star ratings. This effect hold is more apparent in positive ratings than in negative ratings.

When using the boundaries thus obtained for classifying pieces of opinionated natural language text into one out of five star categories solely based on the sentiment score conveyed by the text itself, the performance of the constructed sentiment maps turns out to differ per language as well. The 10-fold cross-validated overall classification accuracy on Dutch documents equals approximately 20%, whereas the overall classification accuracy on English documents equals about 40%. This

Table 1 Language-specific sentiment score intervals associated with each considered number of stars for both Dutch and English.

Stars	Dutch sentiment scores	English sentiment scores
1	$[-1.5, -0.5]$	$[-1.5, -0.4]$
2	$(-0.5, -0.1]$	$(-0.4, 0.0]$
3	$(-0.1, 0.4]$	$(0.0, 0.2]$
4	$(0.4, 0.9]$	$(0.2, 0.4]$
5	$(0.9, 1.5]$	$(0.4, 1.5]$

observation suggests that the sentiment conveyed by natural language text may in some languages be a better proxy for intended sentiment than in other languages. As such, more (latent) aspects of opinionated pieces of natural language text, such as structural aspects or emoticons, may need to be taken into account when converting language-specific sentiment scores into universal star ratings in order to better facilitate cross-language sentiment analysis.

5 Conclusions and Future Work

In this paper, we have proposed to facilitate cross-language sentiment analysis by comparing the sentiment conveyed by natural language text across languages by using these language-specific sentiment scores to classify pieces of sentiment-carrying natural language text into universal star ratings. We have shown that the way natural language reveals people's sentiment tends to differ across languages, as the relation between sentiment conveyed by natural language and intended sentiment is different for the two languages considered in our current work.

The results of our initial experiments with respect to modeling this relation for each language by means of a monotone increasing step function mainly suggest that the sole dimension of language-specific sentiment scores can separate universal classes of intended sentiment from one another to a limited extent. As such, we have made first steps towards cross-language sentiment analysis through universal star ratings, yet our results warrant future research.

In future research, we consider relaxing some of our assumptions in order for the mapping between language-specific sentiment scores and universal star ratings to be more accurate. For instance, we could consider dropping the monotonicity constraint and allow for a non-linear relation between sentiment scores and star ratings. Last, more aspects of content other than the associated sentiment score, e.g., emoticons, could be used as proxy for star ratings in order to facilitate sentiment analysis across languages.

Acknowledgements We would like to thank Teezir (`http://www.teezir.com`) for their technical support and fruitful discussions. The authors of this paper are partially supported by the Dutch national program COMMIT.

References

1. Abbasi, A., Chan, H., Salem, A.: Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. ACM Transactions on Information Systems **26**(3) (2008)
2. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: 7th Conference on International Language Resources and Evaluation (LREC 2010), pp. 2200–2204. European Language Resources Association (2010)
3. Bal, D., Bal, M., van Bunningen, A., Hogenboom, A., Hogenboom, F., Frasincar, F.: Sentiment Analysis with a Multilingual Pipeline. In: Web Information System Engineering, *Lecture Notes in Computer Science*, vol. 6997, pp. 129–142. Springer (2011)
4. Bautin, M., Vijayarenu, L., Skiena, S.: International Sentiment Analysis for News and Blogs. In: 2nd International Conference on Weblogs and Social Media (ICWSM 2008, pp. 19–26. AAAI Press (2008)
5. Dau, W., Xue, G., Yang, Q., Yu, Y.: Co-Clustering Based Classification. In: 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2007), pp. 210–219. Association for Computing Machinery (2007)
6. Dau, W., Xue, G., Yang, Q., Yu, Y.: Transferring Naive Bayes Classifiers for Text Classification. In: 22nd Association for the Advancement of Articifial Intelligence Conference on Artificial Intelligence (AAAI 2007), pp. 540–545. AAAI Press (2007)
7. Gliozzo, A., Strapparava, C.: Cross Language Text Categorization by Acquiring Multilingual Domain Models from Comparable Corpora. In: ACL Workshop on Building and Using Parallel Texts (ParaText 2005), pp. 9–16. Association for Computational Linguistics (2005)
8. Heerschop, B., Goossen, F., Hogenboom, A., Frasincar, F., Kaymak, U., de Jong, F.: Polarity Analysis of Texts using Discourse Structure. In: 20th ACM Conference on Information and Knowledge Management (CIKM 2011), pp. 1061–1070. Association for Computing Machinery (2011)
9. Heerschop, B., Hogenboom, A., Frasincar, F.: Sentiment Lexicon Creation from Lexical Resources. In: 14th International Conference on Business Information Systems (BIS 2011), *Lecture Notes in Business Information Processing*, vol. 87, pp. 185–196. Springer (2011)
10. Heerschop, B., van Iterson, P., Hogenboom, A., Frasincar, F., Kaymak, U.: Analyzing Sentiment in a Large Set of Web Data while Accounting for Negation. In: 7th Atlantic Web Intelligence Conference (AWIC 2011), pp. 195–205. Springer (2011)
11. Hofman, K., Jijkoun, V.: Generating a Non-English Subjectivity Lexicon: Relations that Matter. In: 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009), pp. 398–405. Association for Computing Machinery (2009)
12. Jansen, B., Zhang, M., Sobel, K., Chowdury, A.: Twitter Power: Tweets as Electronic Word of Mouth. Journal of the American Society for Information Science and Technology **60**(11), 2169–2188 (2009)
13. Korte Reviews: Korte Reviews (2011). Available online, `http://kortereviews.tumblr.com/`
14. Lemaire: Lemaire Film Reviews (2011). Available online, `http://www.lemairefilm.com/`
15. Melville, P., Sindhwani, V., Lawrence, R.: Social Media Analytics: Channeling the Power of the Blogosphere for Marketing Insight. In: 1st Workshop on Information in Networks (WIN 2009) (2009)
16. Metacritic: Metacritic Reviews (2011). Available online, `http://www.metacritic.com/browse/movies/title/dvd/`
17. Mihalcea, R., Banea, C., Wiebe, J.: Learning Multilingual Subjective Language via Cross-Lingual Projections. In: 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007), pp. 976–983. Association for Computational Linguistics (2007)
18. Moens, M., Boiy, E.: A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts. Information Retrieval **12**(5), 526–558 (2007)

19. Paltoglou, G., Thelwall, M.: A study of Information Retrieval weighting schemes for sentiment analysis. In: 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010), pp. 1386–1395. Association for Computational Linguistics (2010)
20. Pang, B., Lee, L.: A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on Minimum Cuts. In: 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004), pp. 271–280. Association for Computational Linguistics (2004)
21. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval **2**(1), 1–135 (2008)
22. Short Reviews: Short Reviews (2011). Available online, `http://shortreviews.net/browse/`
23. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics **37**(2), 267–307 (2011)
24. Taboada, M., Voll, K., Brooke, J.: Extracting Sentiment as a Function of Discourse Structure and Topicality. Tech. Rep. 20, Simon Fraser University (2008). Available online, `http://www.cs.sfu.ca/research/publications/techreports/#2008`
25. Wan, X.: Co-Training for Cross-Lingual Sentiment Classification. In: Joint Conference of the 47th Annual Meeting of ACL and the 4th International Join Conference on Natural Language Processing of the AFNLP (ACL 2009), pp. 235–243. Association for Computational Linguistics (2009)
26. Whitelaw, C., Garg, N., Argamon, S.: Using Appraisal Groups for Sentiment Analysis. In: 14th ACM International Conference on Information and Knowledge Management (CIKM 2005), pp. 625–631. Association for Computing Machinery (2005)
27. Wierzbicka, A.: Alternative Linguistics: Descriptive and Theoretical Modes, chap. Dictionaries vs. Encyclopedias: How to Draw the Line, pp. 289–316. John Benjamins Publishing Company (1995)
28. Wierzbicka, A.: Emotion and Facial Expression: A Semantic Perspective. Culture Psychology **1**(2), 227–258 (1995)