The many aspects of fine-grained sentiment analysis
An overview of the task and its main challenges

Orphée De Clercq

LT[3], Language and Translation Technology Team
Ghent University
Ghent, Belgium
Email: orphee.declercq@ugent.be

***Abstract*—In this survey paper, the task of aspect-based sentiment analysis is defined in close detail. We explain how this fine-grained task actually comprises several subtasks and focus on the domain of customer reviews. We reveal which datasets have been made publicly available and describe the state of the art on the subtasks of aspect term extraction, aspect term classification and aspect polarity classification. We conclude this survey by listing some of the main challenges the domain is still facing which illustrate that this task is far from solved.**

***Keywords*–sentiment analysis; user-generated content; natural language processing.**

I. Introduction

With the arrival of Web 2.0 technologies, online communication has become commonplace. These allow site visitors to add content, called *user-generated content* [1]. Examples include forums and message boards, blogs, review sites, e-commerce platforms, but also social networking sites such as Facebook or Twitter. Not only are these a new means of interpersonal and community-level communication, they have also become an important resource for gathering subjective information.

When we need to make a decision about the purchase of a car or cell phone, a travel destination to go to, or a good restaurant to visit, we are typically interested in what other people think. Before Web 2.0, we asked for opinions from friends and family. With the explosive growth of user-generated content on the Web in the past few years, however, it has become possible to go online and find recommendations or check the experience of other customers, e.g., for a particular restaurant to have lunch at. Instead of relying on anecdotal evidence from friends, we have access to a handy overview of the main aspects of that restaurant enabling us to answer that one crucial question: 'Will I like it?'

The same applies from the perspective of companies, governments and organizations. To know the sentiments of the general public towards its brand, products, policies, etc., an organization no longer needs to resort to opinion polls or surveys. Most of that information is already available online, in the form of user-generated content. In previous studies, user-generated content has been used by companies to track how their brand is perceived by consumers [2], for market prediction [3] or to determine the sentiment of financial bloggers towards companies and their stocks [4]; by individuals who need advice on purchasing the right product or service [5] and by nonprofit organizations, e.g., for the detection of suicidal messages [6].

As the amount of online information has grown exponentially, so has the interest in new text mining techniques to handle and analyze this growing amount of subjective text. One of the main research topics is sentiment analysis, also known as opinion mining. The objective of sentiment analysis is the extraction of subjective information from text, rather than factual information. Originally, it focused on the task of automatically classifying an entire document or sentence as positive, negative or neutral. This more coarse-grained level of analysis, however, does not allow to discover what people like and dislike exactly [7].

Often, users are not only interested in people's general sentiments about a certain product, but also in their opinions about specific features, i.e., parts or attributes of that product. One way to do this is by applying aspect-based sentiment analysis (ABSA). Aspect-based (or feature-based) sentiment analysis systems [8] focus on the detection of all sentiment expressions within a given document and the concepts and aspects (or features) to which they refer. Such systems do not only try to distinguish the positive from the negative utterances, but also strive to detect the target of the opinion, which comes down to a very fine-grained sentiment analysis task and "almost all real-life sentiment analysis systems in industry are based on this level of analysis" [7, p10].

In this paper, we first define the task of aspect-based sentiment analysis in detail, with a special focus on the analysis of customer reviews. In Section 2 we explain which datasets have been made available in the framework of SemEval, a well-known workshop in the Natural Language Processing (NLP) community. Next, we move on to discuss the state of the art when applying supervised machine learning techniques to the various subtasks. In Section 4 we explain which challenges still need to be tackled in the near future after which we conclude this survey (Section 5).

II. Definition

Several surveys of the field of sentiment analysis are available, such as [9] or [10]. However, the books by [7], [11] are more recent and extensive summaries of this rapidly evolving field. Liu offers a comprehensive definition of what an *opinion* is:

> "An opinion is a quintuple, $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, where e_i is the name of an entity, a_{ij} is an aspect

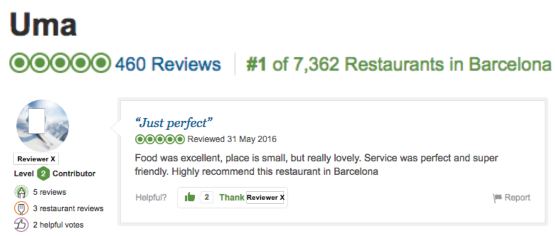

Figure 1. Review from a particular restaurant in Barcelona that was posted on TripAdvisor.

> of e_i, s_{ijkl} is the sentiment on aspect a_{ij} of entity e_i, h_k is the opinion holder, and t_l is the time when the opinion is expressed by h_k. The sentiment s_{ijkl} is positive, negative, or neutral, or expressed with different strength/intensity levels.” [11, pp19-20]

Following this definition, sentiment analysis thus consists of automatically deriving these opinion quintuples from texts and it comprises various subtasks. We will now explain each of these tasks based on an example review presented in Figure 1.

1) **Entity extraction and categorization**: Extract all entity expressions in a document collection, and categorize or group synonymous entity expressions into entity clusters (or categories). In our example, the collection consists of restaurant reviews and the entity presented here is ‘Uma’, belonging to the category *Restaurants*.

2) **Aspect extraction and categorization**: Extract all aspect expressions of the entities, and categorize these aspect expressions into clusters. These aspects can be both explicit and implicit. In our example, we can find out which aspects of this restaurant are mentioned while reading through the review. The explicit aspects are ‘food’, ‘place’ and ‘service’. Implicitly, the final sentence says something about the restaurant in general. If we classify all these aspect expressions into categories, these could be: *Food*, *Ambience*, *Service*, and *Restaurant* respectively.

3) **Opinion holder extraction and categorization**: Extract opinion holders –h_k– for opinions from text or structured data and categorize them. In our example this can easily be derived from the metadata accompanying the review, i.e., we know who wrote the review. Because of privacy concerns the username was anonymized to ‘Reviewer X’.

4) **Time extraction and standardization**: Extract the times when opinions are given and standardize different time formats, t_l. This information can also be easily derived from the time stamp attached to the review: the review was written on 31 May 2016.

5) **Aspect sentiment classification**: Determine whether an opinion on an aspect a_{ij} is positive, negative or neutral, or assign a numeric sentiment rating to the aspect, s_{ijkl}. We can read that the food and service were evaluated positive as well as the restaurant in general. Though the reviewer did note that the place is small -which might hint at a negative sentiment- this is countered in the next part, which clearly indicates that there is a positive ambience.

The quintuples that can be derived from our example are: `(Uma, `*`Food`*`, positive, Reviewer X, May-31-2016)`, `(Uma, `*`Ambience`*`, positive, Reviewer X, May-31-2016)`, `(Uma, `*`Service`*`, positive, Reviewer X, May-31-2016)` and `(Uma, `*`Restaurant`*`, positive, Reviewer X, May-31-2016)`.

This framework has been called many names such as feature-based, topic-based, entity-based or target-based sentiment analysis, but is currently most-known under the name of aspect-based sentiment analysis. It should be noted that any real-life application will have to be able to process many reviews at once and thus a very important final step is to aggregate all aspects and sentiments over an entire document collection.

The focus of this survey is on customer reviews, in this genre one can derive the entity, opinion holder and time as such from the metadata, which is why the main focus will be on the second and fifth subtask. Actually, this second task consists of two subsequent steps: aspect term extraction and aspect term categorization. In this respect, we follow the task decomposition as suggested by the organizers of three Semantic Evaluation tasks on aspect-based sentiment analysis [8], [12], [13].

III. DATASETS

When it comes to customer reviews and aspect-based sentiment analysis, systems have been developed for a variety of domains, such as movie reviews [14], reviews for electronic products, e.g., digital cameras [15] or netbook computers [16], and restaurant reviews [16], [17]. As always when research is performed on individual datasets, true advancements in the field cannot be properly evaluated.

Though several benchmark datasets had already been made publicly available, such as the product reviews dataset of Hu and Liu [15] or the restaurant reviews dataset of [17], it was not until the International Workshop on Semantic Evaluation devoted attention to the task that this problem was tackled. Parts of the previously-mentioned English datasets were extracted an re-annotated for SemEval2014 Task 4 [8] and SemEval 2015 Task 12 [12]. Last year, seven other languages were also included in a third run of the task, i.e., SemEval 2016 Task 5 [13]. Table 1 presents an overview of all the annotated data that is available in different languages and domains so far.

TABLE I. OVERVIEW OF THE BENCHMARK SEMEVAL DATASETS

Domain	*Subdomain*	*Language*	*#Sentences*
Electronics	Camera	Chinese	8040
	Laptops	English	3308
	Phones	Chinese	9521
	Phones	Dutch	1697
Hotels		Arabic	6029
Restaurants		Dutch	2286
		English	2676
		French	2429
		Russian	4699
		Spanish	2951
		Turkish	1248
Telecom		Turkish	3310

Noteworthy is that all this data has been annotated using the same annotation guidelines.[1] Basically the annotation process consists of three incremental steps. First, all explicit and implicit targets -the word or words referring to a specific entity or aspect- are annotated. Next, these targets are assigned to

[1]These guidelines are available at http://goo.gl/wOf1dX.

domain-specific clusters of aspect categories, and in the final step the sentiment expressed towards every aspect is indicated. Three main polarities are distinguished: positive, negative and neutral.

These shared tasks can be perceived as online data competitions: during a specific time frame training data is released allowing NLP teams from all over the world to work on the same problem. In a final stage unseen test data is released, usually for one to three days and each team can submit their system's output. This output is then evaluated for all teams in the same manner which facilitates meaningful comparisons of different techniques.

IV. STATE OF THE ART

In this section we discuss the state of the art, our main focus is on supervised machine learning techniques performed on English data. For more information on unsupervised and hybrid techniques we refer to the survey by [18] and for an overview of the current approaches to languages other than English, we refer to the workshop proceedings of SemEval 2016 Task 5 [19].

A. Aspect Term Extraction

For the task of aspect term extraction (ATE), the most popular and successful approaches are based on frequency and supervised learning [8], [11]. Hu and Liu [15] introduced the task of aspect-based sentiment analysis and constructed the first strong baseline for aspect term extraction by identifying all nouns and noun phrases based on part-of-speech tags and counting frequencies. They only kept the frequent nouns and noun phrases using a frequency threshold. In subsequent research, this method was improved by incorporating pruning mechanisms based on pointwise mutual information, meronymy discriminators (e.g., for the camera class these would be 'camera has', 'camera comes with',...) and exploiting the WordNet hierarchy [20]. Another improvement was to only include those noun phrases that occur in sentiment-bearing sentences or in certain syntactic patterns [21] or to use the C-value measure which allows to also extract multi-word aspects [22]. A combination of this frequency baseline with continuous vector space representations of words [23] has also proven effective in the work of Pavlopoulos and Androutsopoulos [24].

Using supervised learning, the most dominant method is to approach the ATE task as a sequential labeling task [11]. Following the IOB2 notation for Named Entity Recognition [25] the aspect term in the annotated training data is labeled with 'B' indicating the beginning of an aspect term, 'I' indicating the inside of an aspect term and 'O' indicating the outside of an aspect term. The two systems achieving the best performance for this subtask in SemEval 2015 Task 12 used this approach. In [26] (which was actually based on preliminary work [27]), a classifier was trained using Conditional Random Fields (CRF), and in [28] a designated Named Entity Recognition system was used. Both systems implemented typical named entity features such as word bigrams, trigrams, token shape, capitalization, name lists, etc. For SemEval 2016, subsequent work by Toh and Su [29] found that using the output of a Recurrent Neural Network (RNN) as additional features is beneficial for the labeling tasks. More specifically the Bidirectional Elman-type RNN model [30] captures long-range dependencies.

B. Aspect Term Categorization

The next task is to group aspect terms into categories, known as aspect term categorization. The majority of existing research combines similar aspect terms into aspect groups without starting out from a predefined set of aspect categories. The most common approaches are to aggregate synonyms or near-synonyms using WordNet [31], statistics from corpora [32], [33], or semi-supervised learning, or to cluster aspect terms using (latent) topic models [34], [16]. In other research domain-specific taxonomies have been used to aggregate related terms or hierarchical relations between aspect terms [35]. More recently, a multi-granular aspect aggregation method was introduced in the work of [36] by first calculating the semantic relatedness between two frequent aspect terms and then performing hierarchical agglomerative clustering to create an aspect term hierarchy.

All the above-mentioned approaches assume that the list of aspect categories is unknown and has to be aggregated from scratch. In this respect, the task definition as proposed in the aspect-based SemEval tasks differs in that several predefined and domain-specific categories have to be predicted, thus transforming the aggregation task into a multiclass classification task. The two systems achieving the best results on this individual subtask in SemEval 2015 Task 12 both used classification to this purpose, respectively individual binary classifiers trained on each possible category which are afterwards entered in a sigmoidal feedforward network [26] and a single Maximum Entropy classifier [37], respectively. When it comes to the features that were exploited by these systems especially lexical features in the form of bag-of-words (such as word unigrams and bigrams [26] or word and lemma unigrams [37]) have proven successful. The best system [26] also incorporated lexical-semantic features in the form of clusters learned from a large corpus of reference data, whereas the second-best [37] applied filtering heuristics on the classification output and thus solely relied on lexical information for the classification. As is the case for many NLP problems, the added value of deep learning is becoming more apparent for this task as well. For SemEval 2016 Toh and Su [29] found that when their sigmoidal feedforward network is enhanced with the probability output of a Deep Convolutional Neural Network (CNN) [38] as additional features, the performance increases. Moreover, ablation experiments revealed that these CNN features contribute the most to performance.

C. Aspect Term Polarity Classification

The final task is aspect term polarity classification. In the context of aspect-based sentiment analysis, the sentiment polarity has to be determined for each mentioned aspect term of a target entity. Existing sentiment analysis systems can be divided into lexicon-based and machine learning approaches. Lexicon-based methods (see [39] for an overview) determine the semantic orientation of a piece of text based on the words occurring in that text. Crucial in this respect, are sentiment or subjectivity lexicons allowing to define the semantic orientation of words. Lexicons comprise various sentiment or opinion words together with their strength and overall polarity. The word *wonderful*, for example, indicates a positive sentiment, whereas the word *terrible* has a negative connotation. Many subjectivity lexicons were constructed in the past, mainly for English, such as the well-known MPQA lexicon [40] or

SentiWordNet [41], but also for other languages, such as the Pattern [42] and Duoman [43] lexicons for Dutch.

Machine learning approaches to sentiment analysis make use of classification algorithms, such as Naïve Bayes or Support Vector Machines trained on a labeled dataset [10]. This dataset can be extracted from existing resources such as reviews labeled with star ratings [44] or manual annotations [45]. Crucial in this respect is the engineering of a set of effective features [11]. Current state-of-the-art approaches model a variety of contextual, lexical and syntactic features [46], allowing them to capture context and the relations between the individual words. Though deep learning techniques have also been applied to this subtask, mainly in the form of word embeddings [23], for SemEval 2016 the best performing system relied solely on (advanced) linguistic features [47].

According to Liu [11], the key issue is to determine the scope of each sentiment expression within aspect-based sentiment analysis. The main approach is to use parsing to determine the dependency relations and other relevant information, as done in [48] where a dependency parser was used to generate a set of aspect dependent features, or in [49] where each feature is weighted based on the position of the feature relative to the target aspect in the parse tree. With respect to the SemEval tasks it has been shown that general purpose systems used to classify at the sentence level are very effective, which even seems to hold when testing on out-of-domain data [12] or on other languages [13]. However, we do believe that this is inherent to the customer reviews used for the SemEval tasks, these reviews do not contain many conflicting sentiments within one sentence. This brings us to one of the challenges in the field, i.e., domain adaptation, on which we will elaborate in the next section.

V. CHALLENGES

Though research on sentiment analysis has flourished in the past decade, the problem is far from solved. Excellent books and surveys have been published which also devote much attention to the various challenges that lie ahead, see [7] and [50] for recent and extensive overviews. In this section we discuss some of the main challenges.

The focus on consumer reviews in this survey and in most of the research performed on aspect-based sentiment analysis already hints at one challenge, namely **domain adaptation**. Consumer reviews are very product-oriented and the aspect expressions that have to be extracted almost exclusively consist of nouns or noun phrases. Moreover, when someone writes a review the text will almost always include an opinion. In reality, however, large chunks of non-opinionated text co-occur with opinionated text and also verbal expressions or a variety of words can be used to refer to certain aspects. Think for example of political tweets or discussion forums. Nevertheless, it cannot be ignored that domain-knowledge is crucial for aspect-based sentiment analysis. The importance of lexical features in the classification tasks is obvious and if you have the time to compile a different lexicon for each domain you will be able to solve about 60% of the cases [7].

Even when not focussing on reviews, most of the text that is processed in the field of sentiment analysis is **user-generated content** (UGC) which is very different from standard text. Though this UGC is often highly expressive because many emoticons and techniques such as flooding (the repetition of various characters the place emphasis, *loooool*) can be used, it is also full of misspellings, grammatical errors, abbreviations, ... which hinder automatic text processing because the tools used for this are originally trained on standard text [51]. Especially if we consider the importance of lexical features, deviations from the standard can already have a large impact. In this respect promising research has been performed by Van Hee et. al. [52], they investigate to what extent the performance of a sentiment classifier can be further improved by applying a complex normalisation system as a preprocessing step. This normalisation system automatically translates noisy into standard text and the results reveal that this approach is beneficial, especially when testing on unseen data.

One can definitely say that UGC also allows for more **creative language use**, such as sarcasm, irony, humour and metaphor. These are all very difficult to interpret for natural language processing systems. In this respect, we see more research emerging. In 2015, for example, a SemEval shared task was organized on detecting sentiment in tweets rich in metaphor and irony [53]. The tweets provided for this task, however, were almost all ironic and negative and thus did not represent a realistic distribution of sarcastic messages in a random Twitter stream. It will be interesting to see how research in this direction is performed. Interesting in this respect is also the idea to construct a knowledge base including stereotypes and commonly used similes. According to Schouten and Frasincar [18] this evolution to more concept-centric approaches combined with machine learning will give rise to much better algorithms, not only for discovering irony but also for sentiment analysis in general.

As [54] phrases it: “sentiment analysis requires a **deep understanding** of the explicit and implicit, regular and irregular, and syntactic and semantic language rules.” Extracting and classifying explicit sentiment might seem straightforward, however, in reality words are hardly ever used in isolation and whenever sentence composition comes into play both form and context can alter the intended sentiment dramatically. in this respect research is emerging on the impact of those small negation and modification words, which reveals that these are crucial to include [55]. Implicit sentiment is even more complex, much can be read between the lines and even factual statements can evoke different opinions when used in different domains [56]. Moreover, in aspect-based sentiment analysis for example a certain aspect can be referred to with a pronoun or other synonymous phrases, which brings us to the task of coreference resolution. Though many survey studies have claimed that the recognition of coreference is crucial for successful aspect-based sentiment analysis [11], [57] not much research has been performed in this direction. When it comes to this deep understanding, the field is in high expectations of the surge of deep learning techniques. It will be interesting to see whether these new techniques are apt to the task.

VI. CONCLUSION

In this survey the focus has been on aspect-based sentiment analysis of consumer reviews. We have defined the task in close detail and have explained the state of the art for the subtasks of aspect term extraction, aspect term classification and aspect term polarity classification. We have discussed some of the main challenges the field still needs to overcome, such as domain adaptation, processing user-generated and creative language, solving some of the more NLP-hard problems. An interesting evolution to follow in this respect, will be the move towards deep learning in the field of Natural Language Processing.

References

[1] M.-F. Moens, J. Li, and T.-S. Chua, Eds., Mining user generated content. Chapman and Hall/CRC, 2014.

[2] J. Zabin and A. Jefferies, "Social media monitoring and analysis: Generating consumer insights from online conversation," Aberdeen Group Benchmark Report, Aberdeen Group, Tech. Rep., 2008.

[3] T. O. Sprenger, A. Tumasjan, P. G. Sandner, and I. M. Welpe, "Tweets and trades: The information content of stock microblogs," European Financial Management, vol. 20, no. 5, 2014, pp. 926–957.

[4] N. O'Hare, M. Davy, A. Bermingham, P. Ferguson, P. Sheridan, C. Gurrin, and A. F. Smeaton, "Topic-dependent sentiment analysis of financial blogs," in Proceedings of the 1st International Conference on Information and Knowledge Management Workshop on Topic-sentiment Analysis for Mass Opinion (TSA-2009), 2009, pp. 9–16.

[5] M. Dabrowski, T. Acton, P. Jarzebowski, and S. O'Riain, "Improving customer decisions using product reviews - CROM - Car Review Opinion Miner," in Proceedings of the 6th International Conference on Web Information Systems and Technologies (WEBIST-2010), 2010, pp. 354–357.

[6] B. Desmet, "Finding the online cry for help: automatic text classification for suicide prevention," PhD, Ghent University, 2014.

[7] B. Liu, Sentiment Analysis - Mining Opinions, Sentiments, and Emotions. Cambridge University Press, 2015.

[8] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "Semeval-2014 task 4: Aspect based sentiment analysis," in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval-2014), 2014, pp. 27–35.

[9] J. G. Shanahan, Y. Qu, and J. Wiebe, Eds., Computing Attitude and Affect in Text: Theory and Applications, ser. the Information Retrieval Series. Springer, 2006, no. 20.

[10] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval., vol. 2, no. 1-2, 2008, pp. 1–135.

[11] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, 2012, pp. 1–167.

[12] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis," in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 486–495.

[13] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryiğit, "Semeval-2016 task 5: Aspect based sentiment analysis," in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016, pp. 19–30.

[14] T. T. Thet, J.-C. Na, and C. S. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," Journal of Information Science, vol. 36, no. 6, 2010, pp. 823–848.

[15] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining (KDD-2004), 2004, pp. 168–177.

[16] S. Brody and N. Elhadad, "An unsupervised aspect-sentiment model for online reviews," in Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-2010), 2010, pp. 804–812.

[17] G. Ganu, N. Elhadad, and A. Marian, "Beyond the stars: improving rating predictions using review text content," in Proceedings of the 12th International Workshop on the Web and Databases (WebDB-2009), 2009, pp. 1–6.

[18] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, 2016, pp. 813–830.

[19] S. Bethard, M. Carpuat, D. Cer, D. Jurgens, P. Nakov, and T. Zesch, Eds., Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). Association for Computational Linguistics, 2016.

[20] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Proceedings of the 2005 Conference on Human Language Technology and Empirical Methods in Natural Language Processing (EMNLP-2005), 2005, pp. 339–346.

[21] S. Blair-Goldensohn, T. Neylon, K. Hannan, G. A. Reis, R. Mcdonald, and J. Reynar, "Building a sentiment summarizer for local service reviews," in Proceedings of the WWW-2008 workshop on NLP in the Information Explosion Era (NLPIX-2008), 2008.

[22] J. Zhu, H. Wang, B. K. Tsou, and M. Zhu, "Multi-aspect opinion polling from textual reviews," in Proceedings of the 18th Association for Computing Machinery Conference on Information and Knowledge Management (CIKM-2009), 2009, pp. 1799–1802.

[23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," CoRR, 2013.

[24] J. Pavlopoulos and I. Androutsopoulos, "Aspect term extraction for sentiment analysis: New datasets, new evaluation measures and an improved unsupervised method," in Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM-2014), 2014, pp. 44–52.

[25] E. Tjong Kim Sang, "Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition," in Proceedings of the 6th Conference on Natural Language Learning (COLING-2002), 2002, pp. 155–158.

[26] Z. Toh and J. Su, "NLANGP: Supervised machine learning system for aspect category classification and opinion target extraction," in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), June 2015, pp. 496–501.

[27] Z. Toh and W. Wang, "DLIREC: Aspect term extraction and term polarity classification system," in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval-2014), 2014, pp. 235–240.

[28] I. n. San Vicente, X. Saralegi, and R. Agerri, "EliXa: A Modular and Flexible ABSA Platform," in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 748–752.

[29] Z. Toh and J. Su, "NLANGP at SemEval-2016 Task 5: Improving Aspect Based Sentiment Analysis using Neural Network Features," in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016, pp. 282–288.

[30] P. Liu, S. Joty, and H. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings." in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP-2015), 2015, pp. 1433–1443.

[31] Y. Liu and S. Lin, "Log-linear models for word alignment," in Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL-2005), 2005, pp. 459–466.

[32] H.-H. Chen, M.-S. Lin, and Y.-C. Wei, "Novel association measures using web search with double checking," in Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics (COLING - ACL-2006), 2006, pp. 1009–1016.

[33] D. Lin and X. Wu, "Phrase clustering for discriminative learning," in Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-2009), 2009, pp. 1030–1038.

[34] I. Titov and R. McDonald, "A joint model of text and aspect ratings for sentiment summarization," in Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-2008), 2008, pp. 308–316.

[35] N. Kobayashi, K. Inui, and Y. Matsumoto, "Extracting aspect-evaluation and aspect-of relations in opinion mining," in Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP - CoNLL-2007, 2007, pp. 1065–1074.

[36] I. Pavlopoulos, "Aspect based sentiment analysis," PhD, Department of Informatics, Athens University of Economics and Business, 2014.

[37] J. Saias, "Sentiue: Target and aspect based sentiment analysis in

SemEval-2015 Task 12," in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), June 2015, pp. 767–771.

[38] A. Severyn and A. Moschitti, "UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification." in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015), 2015, pp. 464–469.

[39] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational Linguistics, vol. 37, no. 2, 2011, pp. 267–307.

[40] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in Proceedings of the 2005 Conference Empirical Methods in Natural Language Processing (EMNLP-2005), 2005, pp. 347–354.

[41] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC-2010), 2010, pp. 2200–2204.

[42] T. De Smedt and W. Daelemans, "Vreselijk mooi! Terribly beautiful: a subjectivity lexicon for Dutch adjectives," in Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC-2012), 2012, pp. 3568–3572.

[43] V. Jijkoun and K. Hofmann, "Generating a non-English subjectivity lexicon: Relations that matter," in Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2009), 2009, pp. 398–405.

[44] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP-2002), 2002, pp. 79–86.

[45] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," Computer Intelligence, vol. 39, no. 2, 2005, pp. 165–210.

[46] L. D. Caro and M. Grella, "Sentiment analysis via dependency parsing," Computer Standards & Interfaces, vol. 35, no. 5, 2013, pp. 442–453.

[47] C. Brun, J. Perez, and C. Roux, "XRCE at SemEval-2016 Task 5: Feedbacked Ensemble Modeling on Syntactico-Semantic Knowledge for Aspect Based Sentiment Analysis," in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016, pp. 277–281.

[48] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011), 2011, pp. 151–160.

[49] E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," Information Retrieval, vol. 12, no. 5, 2009, pp. 526–558.

[50] S. M. Mohammad, "Challenges in sentiment analysis," in A Practical Guide to Sentiment Analysis, D. Das, E. Cambria, and S. Bandyopadhyay, Eds. Springer, 2016.

[51] J. Eisenstein, "What to do about bad language on the internet," in Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2013, pp. 359–369.

[52] C. Van Hee, M. Van de Kauter, O. De Clercq, E. Lefever, B. Desmet, and V. Hoste, "Noise or Music? Investigating the Usefulness of Normalisation for Robust Sentiment Analysis on Social Media Data," Expert Systems with Applications, submitted.

[53] A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes, "Semeval-2015 task 11: Sentiment analysis of figurative language in twitter," in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 2015, pp. 470–478.

[54] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intelligent Systems, vol. 28, no. 2, 2013, pp. 15–21.

[55] S. Kiritchenko and S. Mohammad, "The effect of negators, modals, and degree adverbs on sentiment composition," in Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2016, pp. 43–52.

[56] M. Van de Kauter, D. Breesch, and V. Hoste, "Fine-grained analysis of explicit and implicit sentiment in financial news articles," Expert Systems with applications, vol. 42, no. 11, 2015, pp. 4999–5010.

[57] R. Feldman, "Techniques and applications for sentiment analysis," Communications of the ACM, vol. 56, no. 4, 2013, pp. 82–89.