Sentiment Analysis and Opinion Mining

Okoro Jennifer Chimaobiya
Mrs. Hari Priya
MsIT, Jain College, 9th Block Jayanagar.
Bangalor-560069, India.

Abstract- **Sentiment analysis and opinion mining is the field of study that analyses people's opinions, sentiments, evaluations, attitudes, and emotions from written language. It is one of the most active research areas in natural language processing and is also widely studied in data mining, Web mining, and text mining. In fact, this research has spread outside of computer science to the management sciences and social sciences due to its importance to business and society as a whole. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. For the first time in human history, we now have a huge volume of opinionated data recorded in digital form for analysis.**

1. INTRODUCTION

Opinion mining is a type of natural language processing for tracking the mood of the public about a particular product. Opinion mining, which is also called sentiment analysis, involves building a system to collect and categorize opinions about a product. Automated opinion mining often uses machine, a type of artificial intelligence, to text for sentiment. Opinion mining can be useful in several ways. It can help marketers evaluate the success of an ad campaign or new product launch, determine which versions of a product or service are popular and identify which demographics like or dislike particular product features. For example, a review on a website might be broadly positive about a digital camera, but be specifically negative about how heavy it is. Being able to identify this kind of information in a systematic way gives the vendor a much clearer picture of public opinion than surveys or focus groups do, because the data is created by the customer. There are several challenges in opinion mining. The first is that a word that is considered to be positive in one situation may be considered negative in another situation. Take the word "long" for instance. If a customer said a laptop's battery life was long, that would be a positive opinion. If the customer said that the laptop's start-up time was long, however, that would be is a negative opinion. These differences mean that an opinion system trained to gather opinions on one type of product or product feature may not perform very well on another. A second challenge is that people don't always express opinions the same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In opinion mining, however, "the movie was great" is very different from "the movie was not great". Finally, people can be contradictory in their statements. Most reviews will have both positive and negative comments, which is somewhat manageable by analysing sentences one at a time. However, the more informal the medium (twitter tweets or blog posts for example), the more likely people are to combine different opinions in the same sentence. For example: "the movie bombed even though the lead actor rocked it" is easy for a human to understand, but more difficult for a computer to parse. Sometimes even other people have difficulty understanding what someone thought based on a short piece of text because it lacks context. For example, "That movie was as good as his last one" is entirely dependent on what the person expressing the opinion thought of the previous film.

2. LITERATURE REVIEW.

As the Internet and Web technologies continue to grow and expand, the space and scope in the area of information retrieval is also expanding. Hence, researchers take a keen interest in solving the problems associated with OM, which is one of the subareas related to information retrieval and knowledge discovery from the Web. OM is considered an interesting area of research due to its many applications in society. Over the past few years, the ubiquitous dependency on e-marketing, e-business, e-banking, product recommendations, political reviews, and other social activities has attracted research communities worldwide. Special attention has been given to customer mining of reviews as they seek information from the Web about a product and/or the product's reputation. A number of sub-areas of this topic have been explored and extensive research has been reported on each of the sub-problems (Tsytsarau and Palpanas, 2011 and Zhai et al., 2011).

Despite numerous research efforts, the current OM studies and applications still have limitations and margins for improvement. Accordingly, OM suffers from a number of problems, such as accuracy, scalability, quality, standard of data, natural language understanding comprehension, among others.

Some of the major challenges related to natural language processing, such as context dependency, semantic relatedness and ambiguity, have made OM difficult. As practical applications require high accuracy, some of the work must still be performed manually because of the challenging problems with the NLP. For example, the problem of ambiguity, context dependency, and complex and vague sentences require further attention to improve the accuracy of the data analyses. While private blogs are an important source of data for OM, the blog posts are typically written informally and are highly diverse and thus subject to inaccuracies and misinterpretations in analysis.

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCRIT - 2016 Conference Proceedings

To execute the OM process, opinions are collected from the World Wide Web. The Web is a huge and diverse source of information that collects and summarizes opinions from a diverse, multi-dimensional, and redundant data source and, as such, it poses a tremendous challenge for a number of reasons. As a result of these issues, opinion collections are currently limited to specific websites, or opinions are collected on a large scale in an ad hoc fashion from different sites and then processed. On-line analytical processing systems are only possible if there is an efficient system to aggregate and summarize the large collection of text (Tsytsarau and Palpanas, 2011).

Most of the existing research regarding opinion mining is domain dependent, which limits the scope of the application as well as the generalization of the information. Machine-learning systems, which are domain dependent, require that data be manually labelled, a difficult task to manage. Hence, generalized domain independent algorithms are needed for the automatic identification and classification of opinion components.

Scalability of the data is another major challenge in the field of OM. The main goal of OM research is to provide a search engine on the Internet that provides fast, accurate, and well-summarized results of queries regarding opinions of people about anything and everything in the world. However, the limited speed, the huge volume of data and the high dimensionality of the data do not allow for a desirable solution. Thus, complex NLP and text processing algorithms as well as scalable solutions are needed to alleviate these overwhelming concerns and to improve efficiency.

Also presenting a huge challenge in the face of OM is the availability and accessibility of a standard dataset. Few data are currently available to facilitate the classification, bench marking and analysis of the derived text. The absence of a standard of measure that evaluates the results of the overall steps of the OM process remains a concern as well because the existing measurement techniques conduct only partial evaluations, such as simple aggregation of data. Performing such aggregations with respect to opinions is not sufficient for a qualitative analysis of opinions as it is also essential to conduct an analysis of conflicting opinions. Tsytsarau and Palpanas (2011) such an analysis of conflicting opinions is termed contradictory analysis (Tsytsarau and Palpanas, 2011), which is a new direction in the field of OM. Thus, to date, little research has been conducted in this area (Choudhury et al., 2008).

Another main challenge in this area is the quality of reviews. Because the Web is openly accessible to everyone, anyone can post a review, a situation that brings into question the quality of a review or opinion. When individuals are making decisions based on the reviews accessed from the Web, it is important that the reviews be credible and of high quality. However, only limited work has been conducted on opinion quality determination. For example, some researchers who have explored this issue have used the profiles of the reviewers as a means to verify the quality of a review (Lu et al., 2010). Because of an increasing trend to use online reviews when determining a product's reputation, stakeholders are including spam reviews on their sites to enhance their products' reputations. Therefore, as it is necessary to identify spam reviews, some studies have focused on spam detection (Chen et al., 2009, Jindal and Liu, 2007 and Lim et al., 2010). Even so, this task remains a challenging problem. "An accepted source for information or advice is either an expert on the subject, or a persuasive force to check the quality of opinions that they are believable and trustworthy" (Conrad et al., 2008). Open forums and blogs often suffer from a lack of expertise and the inability to present text in the appropriate way.

Opinions are collected in two formats, i.e., structured questions–answers (Kim and Hovy, 2005) and plain text (Somprasertsri, 2010). Mining opinion from structured data is not the main issue, however. Rather, opinion mining from unstructured text is the problem that invites numerous challenges (Liu, 2010a and Liu, 2010b). For example, the identification of the opinion components, context dependency, word sense ambiguity, multilingual effects, and noise in the text, etc. are concerns that are still challenging NLP and affecting opinion mining efficiency.

One of the important problems of OM is the identification of opinion targets from unstructured text. The opinion target is defined as the entity or features of an entity about which an opinion is expressed. The sub-tasks related to opinion target identification include opinion identification, the relevancy of features and features classification, which depends on natural language processing and computation techniques as described in the background study (Somprasertsri, 2010). Another problem is domain dependency, which can be a problem when the target features that are relevant to a specific domain take on different meanings or interpretations when in a different domain. Accordingly, creating a knowledge base for each domain with relevant features and attributes is a difficult but real concern. Hence, generalized procedures are used to identify and disregard the domain dependency of features (Balahur and Montoyo, 2008, Ben-David et al., 2007 and Qiu et al., 2009).

3. RESEARCH METHODOLOGY.

The analysis of existing opinion-related dimensions can be performed at various levels of granularity. Some applications consider the whole document as a single entity for opinion analysis, while other applications focus on sentence level and still other applications focus on the expression or phrase level and term level. The finest-grain level is the term level analysis.

Document level opinion summarization is a broad level of opinion mining, which is sometimes referred to as topic level opinion mining. This level summarizes the opinion about a given topic. Topic-based opinion summarization sums up the overall positive and negative opinions expressed in documents. Hence, the system of opinion mining visualizes the opinion scores according the positive and the negative scores. While various approaches have been employed for document level opinion mining (Dave et al., 2003, Kim and Zhai, 2009, Pang et al., 2002 and Turney, 2002), the following steps are normally

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCRIT - 2016 Conference Proceedings

followed (an overview of the overall process is presented in figure 1).

1. Extract all opinion terms after pre-processing the document.
2. Classify the opinion terms as positive/negative.
3. If the number of positive opinion terms exceeds the number of negative opinion terms, the document is considered to express a positive opinion; if the reverse holds, the document is considered to express a negative opinion.
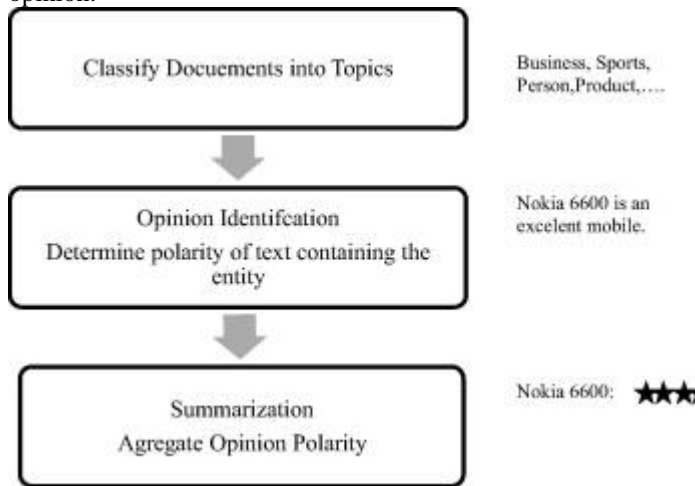

Figure 1 Overview of topic based opinion summarization (Kim and Zhai, 2009).

Turney (2002) discussed an interesting model for review ranking called "thumbs up or thumbs down" whereby an unsupervised model for document polarity identification based on lexical resources is presented. Turney (2002) posits that for any input document d having terms T where each term t belongs to T, if the polarity is (1, 0, −1) where 1 represents positive polarity, 0 represents neutral polarity and −1 represents negative polarity, then if the sum of the polarities of all terms is greater than 0, the document is considered positive; if the sum of the polarities of all terms is less than 0, then the document is negative; and if the sum is equal to 0, then the document is considered to be neutral. The model is defined as given below.
equation(2)

$$Polarity(d) = \begin{cases} Positive & \sum_{t \in T(d)} o(t) > 0 \\ Negative & \sum_{t \in T(d)} o(t) < 0 \\ Neutral & \sum_{t \in T(d)} o(t) = 0 \end{cases}$$

Concentrating on the corpus free approach for review classification, Pang et al. (2002)employed three machine-learning algorithms (naïve Bayesian, maximum entropy, and support vector machine) to rank the documents. However, as this method requires training regarding interpretation of data collected from rated reviews, the problem of domain dependency and a pre-knowledge base remains unsolved.

Dave et al. (2003) formulated a model for review classification based on features for machine learning and classification. Their approach depends on a manually annotated corpus whereby each of the annotated corpuses is described by features related to positivity and negativity. The test document is classified through an annotated corpus using similarity scores. The classifier depends on information retrieval techniques for feature extraction and scoring. As such, this paper proposed that a group of sentences or a full review can provide a more reliable analysis than an individual sentence as a sentence-based performance analysis is limited due to noise and ambiguity. Chen et al. (2006) described a model for review classification. Their work is based on a set of research questions regarding opinions or reviews.

1. What are the differences between positive and negative reviews?
2. What is the origin of a particular opinion?
3. How do these opinions change over time?
4. To what extent can differentiating features be identified from an unstructured text?
5. How accurately can these features predict the category of a review?

These study first analysed terminology variations in a huge number of reviews based on syntactic, semantic, and statistical associations and used term variation patterns to represent underlying topics. This method uses a log likelihood ratio test algorithm to select the most predictive terms, and thus, they are potentially exploited for classification of conflicting reviews. The proposed algorithm indicates approximately 70% accuracy in the conflicting review classification.

Finn and Kushmerick (2006) described an approach to classify documents as either subjective or objective. This paper proposed an automatic genre analysis, i.e., distinguishing documents according to style. This method investigates the use of machine learning for automatic genre classification. Furthermore, these authors introduced the concept of domain transfer through genre classifiers so the classifier could be used for multiple topics in a single document. This paper used different features when building genre classifiers for multiple-topic domain classification.
Kim and Zhai (2009) described a novel model for the summarization of contradictory opinions. This model requires that two sets of opinion-oriented sentences (positive and negative) be extracted from input documents and then, based on these sets of sentences, the algorithm generates a comparative summary of the opinion. This framework relies on measuring the content similarities and contrast similarities of the sentences.
"Recent solutions for sentiment analysis have relied on feature selection methods ranging from lexicon-based approaches where the set of features are generated by humans, to approaches that use general statistical measures

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCRIT - 2016 Conference Proceedings

where features are selected solely on empirical evidence" (Duric and Song (2011)).
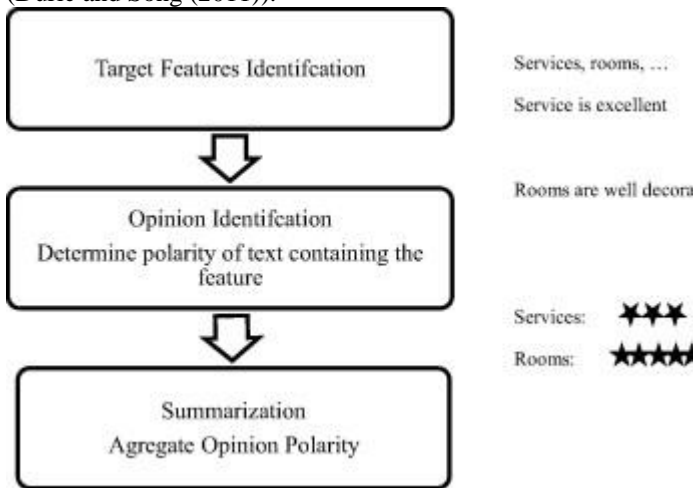

Figure 2.

4. IMPLEMENTATION.

There are some topics that work under the umbrella of SA and have attracted the researchers recently. In the next subsection, three of these topics are presented in some details with related articles.

4.1. Emotion detection

Sentiment analysis is sometimes considered as an NLP task for discovering opinions about an entity; and because there is some ambiguity about the difference between opinion, sentiment and emotion, they defined *opinion* as a transitional concept that reflects attitude towards an entity. The *sentiment* reflects feeling or emotion while emotion reflects attitude.

It was argued by Plutchik, that there are eight basic and prototypical emotions which are *joy*, *sadness*, *anger*, *fear*, *trust*, *disgust*, *surprise*, *and anticipation*. Emotions Detection (ED) can be considered a SA task. SA is concerned mainly in specifying positive or negative opinions, but ED is concerned with detecting various emotions from text. As a Sentiment Analysis task, ED can be implemented using ML approach or Lexicon-based approach, but Lexicon-based approach is more frequently used.

ED on a sentence level was proposed by Lu and Lin. They proposed a web-based text mining approach for detecting emotion of an individual event embedded in English sentences. Their approach was based on the probability distribution of common mutual actions between the subject and the object of an event. They integrated web-based text mining and semantic role labelling techniques, together with a number of reference entity pairs and hand-crafted emotion generation rules to recognize an event emotion detection system. They did not use any large-scale lexical sources or knowledge base. They showed that their approach revealed a satisfactory result for detecting the positive, negative and neutral emotions. They proved that the emotion sensing problem is context-sensitive.

Using both ML and Lexicon-based approach was presented by Balahur et al. They proposed a method based on common-sense knowledge stored in the emotion corpus (EmotiNet) knowledge base. They said that emotions are not always expressed by using words with an affective meaning i.e. *happy*, but by describing real-life situations, which readers detect as being related to a specific emotion. They used SVM and SVM-SO algorithms to achieve their goal. They showed that the approach based on EmotiNet is the most appropriate for the detection of emotions from contexts where no affect-related words were present. They proved that the task of emotion detection from texts such as the ones in the emotion corpus ISEAR (where little or no lexical clues of affect are present) can be best tackled using approaches based on common-sense knowledge. They showed that by using EmotiNet, they obtained better results compared to the methods that employ supervised learning on a much greater training set or lexical knowledge.

Affect Analysis (AA) is a task of recognizing emotions elicited by a certain semiotic modality. Neviarouskaya et al. have suggested an Affect Analysis Model (AAM). Their AAM consists of five stages: symbolic cue, syntactical structure, word-level, phrase-level and sentence-level analysis. This AAM was used in many applications presented in Neviarouskaya work, and.

Classifying sentences using fine-grained attitude types is another work presented by Neviarouskaya et al. They developed a system that relied on the compositionality principle and a novel approach dealing with the semantics of verbs in attitude analysis. They worked on 1000 sentences from http://www.experienceproject.com. This is a site where people share personal experiences, thoughts, opinions, feelings, passions, and confessions through the network of personal stories. Their evaluation showed that their system achieved reliable results in the task of textual attitude analysis.

Affect emotion words could be used as presented by Keshtkar and Inkpen using a corpus-based technique. In their work, they introduced a bootstrapping algorithm based on contextual and lexical features for identifying paraphrases and to extract them for emotion terms, from nonparallel corpora. They started with a small number of seeds (WordNet Affect emotion words). Their approach learned extraction patterns for six classes of emotions. They used annotated blogs and other data sets as texts to extract paraphrases from them. They worked on data from live journals blogs, text affect, fairy tales and annotated blogs. They showed that their algorithm achieved good performance results on their data set.

Ptaszynski et al. have worked on text-based affect analysis (AA) of Japanese narratives from Aozora Bunko. In their research, they addressed the problem of person/character related affect recognition in narratives. They extracted emotion subject from a sentence based on analysis of anaphoric expressions at first, then the affect analysis procedure estimated what kind of emotional state each character was in for each part of the narrative.

Studying AA in mails and books was introduced by Mohammad. He has analysed the Enron email corpus and proved that there were marked differences across genders

in how they use emotion words in work-place email. He created lexicon which has manual annotations of a word's associations with positive/negative polarity, and the eight basic emotions by crowd-sourcing. He used it to analyse and track the distribution of emotion words in books and mails. He introduced the concept of emotion word density by studying novels and fairy tales. He proved that fairy tales had a much wider distribution of emotional word densities than novels.

4.2. Building resources

Building Resources (BR) aims at creating lexica, dictionaries and corpora in which opinion expressions are annotated according to their polarity. Building resources is not a SA task, but it could help to improve SA and ED as well. The main challenges that confronted the work in this category are *ambiguity of words*, *multilingualism*, *granularity* and *the differences in opinion expression among textual genres*.

Building Lexicon was presented by Tan and Wu. In their work, they proposed a random walk algorithm to construct domain-oriented sentiment lexicon by simultaneously utilizing sentiment words and documents from both old domain and target domain. They conducted their experiments on three domain-specific sentiment data sets. Their experimental results indicated that their proposed algorithm improved the performance of automatic construction of domain-oriented sentiment lexicon.

Building corpus was introduced by Robaldo and Di Caro. They proposed Opinion Mining-ML, a new XML-based formalism for tagging textual expressions conveying opinions on objects that are considered relevant in the state of affairs. It is a new standard beside Emotion-ML and WordNet. Their work consisted of two parts. First, they presented a standard methodology for the annotation of affective statements in the text that was strictly independent from any application domain. Second, they considered the domain-specific adaptation that relied on the use of ontology of support which is domain-dependent. They started with data set of restaurant reviews applying query-oriented extraction process. They evaluated their proposal by means of fine-grained analysis of the disagreement between different annotators. Their results indicated that their proposal represented an effective annotation scheme that was able to cover high complexity while preserving good agreement among different people.

Boldrini et al. have focused on the creation of EmotiBlog, a fine-grained annotation scheme for labelling subjectivity in non-traditional textual genres. They focused on the annotation at different levels: document, sentence and element. They also presented the EmotiBlog corpus; a collection of blog posts composed by 270,000 token about three topics in three languages: Spanish, English and Italian. They checked the robustness of the model and its applicability to NLP tasks. They tested their model on many corpora i.e. ISEAR. Their experiments provided satisfactory results. They applied EmotiBlog to sentiment polarity classification and emotion detection. They proved that their resource improved the performance of systems built for this task.

Building Dictionary was presented by Steinberger et al. In their work they proposed a semi-automatic approach to creating sentiment dictionaries in many languages. They first produced high-level gold-standard sentiment dictionaries for two languages and then translated them automatically into a third language. Those words that can be found in both target language word lists are likely to be useful because their word senses are likely to be similar to that of the two source languages. They addressed two issues during their work; the morphological inflection and the subjectivity involved in the human annotation and evaluation effort. They worked on news data. They compared their triangulated lists with the non-triangulated machine-translated word lists and verified their approach.

4.3. Transfer learning

Transfer learning extracts knowledge from auxiliary domain to improve the learning process in a target domain. For example, it transfers knowledge from Wikipedia documents to tweets or a search in English to Arabic. Transfer learning is considered a new cross domain learning technique as it addresses the various aspects of domain differences. It is used to enhance many Text mining tasks like text classification, sentiment analysis, Named Entity recognition, part-of-speech tagging etc.

In Sentiment Analysis; transfer learning can be applied to transfer sentiment classification from one domain to another or building a bridge between two domains. Tan and Wang proposed an Entropy-based algorithm to pick out high-frequency domain-specific (HFDS) features as well as a weighting model which weighted the features as well as the instances. They assigned a smaller weight to HFDS features and a larger weight to instances with the same label as the involved pivot feature. They worked on education, stock and computer reviews that come from a domain-specific Chinese data set. They proved that their proposed model could overcome the adverse influence of HFDS features. They also showed that their model is a better choice for SA applications that require high-precision classification which have hardly any labelled training data.

Wu and Tan have proposed a two-stage framework for cross-domain sentiment classification. In the first stage they built a bridge between the source domain and the target domain to get some most confidently labelled documents in the target domain. In the second stage they exploited the intrinsic structure, revealed by these labelled documents to label the target-domain data. They worked on books, hotels, and notebook reviews that came from a domain-specific Chinese data set. They proved that their proposed approach could improve the performance of cross-domain sentiment classification.

The Stochastic Agreement Regularization algorithm deals with cross-domain polarity classification. It is a probabilistic agreement framework based on minimizing the Bhattacharyya distance between models trained using two different views. It regularizes the models from each view by constraining the amount by which it allows them to disagree on unlabelled instances from a theoretical model. The Stochastic Agreement Regularization algorithm

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCRIT - 2016 Conference Proceedings

was used as a base for the work presented by Lambova et al. Which discussed the problem of cross-domain text subjectivity classification. They proposed three new algorithms based on multi-view learning and the co-training algorithm strategy constrained by agreement. They worked on movie reviews and question answering data that came from three famous data sets. They showed that their proposed work give improved results compared to the Stochastic Agreement Regularization algorithm.

Diversity among various data sources is a problem for the joint modelling of multiple data sources. Joint modelling is important to transfer learning; that is why Gupta et al. have tried to solve this problem. In their work, they proposed a regularized shared subspace learning framework, which can exploit the mutual strengths of related data sources while being unaffected by the effects of the changeability of each source. They worked on social media news data that come from famous social media sites as BlogSpot, Flicker and YouTube and also from news sites as CNN, BBC. They proved that their approach achieved better performance compared to others.

6.   CONCLUSION

This work presents an in-depth background study about opinion mining. The subject has attracted considerable attention since the 1990s, specifically with respect to subjectivity analysis and lexical resource generation. Based on web content and the advancements of Web 2.0 technology, this study indicates that considerable attention has been given to opinion mining in the last few years. This study exploits social networks and web blogs, the most popularly employed sources for opinion retrieval, to examine opinion representation, opinion mining models, opinion components, and related problems. A number of computational models and linguistic features related to opinion mining, component analysis and opinion-target identification are thoroughly discussed

REFERENCES

[1]   Abbasi et al., 2008 Ahmed Abbasi, Hsinchun Chen, Arab Salem **Sentiment analysis in multiple languages: feature selection for opinion classification in web forums** ACM Trans. Inf. Syst., 26 (3) (2008), pp. 1–3

[2]   Abulaish et al., 2009 M. Abulaish, M.N. Doja, T. Ahmad **Feature and opinion mining for customer review summarization** Pattern Recognition and Machine Intelligence, Springer Berlin Heidelberg (2009), pp. 219–224

[3]   Balahur and Montoyo, 2008 Balahur, A., Montoyo, A., 2008. A feature dependent method for opinion mining and classification. Paper presented at the International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE 08.

[4]   Baroni and Vegnaduzzo, 2004 Baroni Marco, Vegnaduzzo Stefano, 2004. Identifying subjective adjectives through web-basedmutual information paper presented at the German conference on natural language processing KONVENS-04.