Sentiment Analysis at Document Level

Conference Paper · August 2016

DOI: 10.1007/978-981-10-3433-6_20

3 authors:

Salima Behdenna
University of Oran 1 Ahmed Ben Bella, Oran, Algeria

Fatiha Barigou
Université Oran 1, Ahmed Benbella, Oran, Algeria

Ghalem Belalem
University of Oran1, Ahmed Ben Bella

Sentiment Analysis at Document Level

Salima Behdenna, Fatiha Barigou and Ghalem Belalem

Department of Computer Science, Faculty of Sciences, University of Oran 1 Ahmed Ben Bella, PB 1524 El M'Naouer, Oran, Algeria (31000)
{behdennasalima, fatbarigou, ghalem1dz}@gmail.com

Abstract Sentiment analysis becomes a very active research area in the text mining field. It aims to extract people's opinions, sentiments, and subjectivity from the texts. Sentiment analysis can be investigated at three levels: document level, sentence level and aspect level. An important part of research effort focuses on the sentiment analysis at the document level, including works on opinion classification of reviews. This survey paper tackles a comprehensive overview of the last update of sentiment analysis at document level. The main target of this survey is to give nearly full image of sentiment analysis techniques at this level. In addition, some future research issues are also presented.

Keywords: opinion mining, sentiment analysis, opinion, review, document, machine learning.

1 Introduction

Now days many applications and platforms on the web allow us to deposit views, to share the sentiments and opinions on a variety of topics. Given the importance of this information in several areas (political, commercial or individual), it would be interesting to treat opinions automatically. The term "sentiment analysis" is used to refer to the automatic processing of opinions, sentiments and subjectivity in texts. This field is known as the opinion mining [16] or sentiment analysis [7].

Sentiment analysis or opinion miming is an extremely active field of research in natural language processing (NLP), which allows extracting the opinions from a set of documents relevant to a given topic. Sentiment analysis can be performed at three different levels [7]:

- Document level analysis: at this level, the task is to determine the general opinion of the document. It considers the whole document as a basic information unit (talking about one topic). This level of analysis assumes that each document expresses opinions on a single entity. Thus, it is not applicable to documents that evaluate or compare several entities.

- Sentence level analysis: the task at this level is to determine if each sentence has expressed a positive, negative or neutral opinion. This level of analysis is closely related to subjectivity classification that distinguishes the sentences (called objective sentences) expressing factual information and sentences (called subjective sentences) expressing subjective opinions. In this case, treatments are twofold; firstly identify if the sentence is carrying or not an opinion, then assess if the opinion is positive or negative. But the main difficulty comes from the fact that objective sentences can be carrying opinion.
- Aspect level analysis: this level performs a finer analysis and requires the use of natural language processing. It is based on the idea that the opinion is characterized by a polarity which could be positive or negative and a target (of opinion). In this case, treatments are twofold: first identify the entity and aspects of the entity in question, and then evaluate the opinion on each of the aspects.

Sentiment analysis has many applications ranging from identifying consumer sentiment towards products [7] (this information can give companies valuable information as to the satisfaction or dissatisfaction of their consumers and it is also immensely valuable for consumers in their decisions to purchase a particular product) to voters' reaction to political adverts. Other application areas in which sentiment analysis can be very useful are: Business Intelligence, Recommendation systems, etc.

The rest of this paper is arranged as follows: Section 2 provides background information and related work of sentiment analysis at the document level. Then a comparative study between different works is presented in section 3. Last section concludes our study and discusses some future directions for research.

2 Related Work

The field of sentiment analysis is vast and knows a spectacular growth because of the commercial challenges. A relatively exhaustive state of art was drawn up in 2008 by Pang and Lee [16] They focused on the applications and challenges in sentiment analysis. They mentioned the techniques used to solve each problem in sentiment analysis.

In his book, Bing Liu [7] presents a synthesis of works in the field of sentiment analysis. It updates the state of the art of Pang and Lee (2008) and distinguishes three levels of analysis: analysis at document level, at sentence level and at aspect level.

In [1], the authors provide a summary of the technical approaches based on machine learning such as Naïve Bayes, maximum entropy, SVM, neural networks and decision tree, and semantic orientation approaches used in that field.

Existing approaches for sentiment analysis are grouped into three main categories: Supervised approaches, unsupervised approaches and semi supervised approaches.

2.1 Supervised approaches

Those approaches require labeling a corpus in advance (positive, negative or neutral) [28]. The main features used are: words, bigram, tri-gram, part of speech and polarity. Several supervised-based techniques are used in the field of sentiment analysis, but two of them appear to provide the best results. These are the Support Vector Machines (SVM) and Naïve Bayes (NB) classifiers [9,21,25]

Pak and Paroubek [13] developed a new sub-graph-based representation extracted from syntactic dependency trees. They represent a text as a collection of sub-graphs, where the nodes are words (or word classes) and arcs the syntactic dependencies between them. Such representation avoids the loss of information associated with the use of "bag of words" models, the latter being based only on collections of n-grams of words. Thus, approaches based on n-grams cannot correctly identify the complex sentiment expressions. The authors use the Incremental Parser (XIP) to construct the dependency tree. They tested the model on a set of French reviews of video games, developed as part of the DOXA project [32] on opinion mining. Thus, they were able to show that SVM classifier using features built from sub graphs, extracted from dependency trees, gives better results than traditional systems based on unigram.

In [17], the authors propose the use of neural network to learn an effective model of sentiment classification. They compared their work with an SVM model using the multi-thematic Amazon corpus. The experiment results show identical performance.

In [30], the objective of the work is the classification of Chinese mobile reviews with SVM and NB classifiers; these reviews are much shorter on average than those of the PC. The scoring used in this work is iTunes' score. ITunes has a score of 1 to 5 points, the reviews with 1 or 2 points are marked as negative, reviews with 4 and 5 points are marked as positive and those with 3 points are marked as neutral. The results show that the NB classifier is better than the SVM classifier.

Vinodhini and Chandrasekaran [26] evaluated the PCA (Principle Component Analysis) application effect with the two methods of sentiment classification: SVM and NB. The experiments are performed on product reviews. The performances are improved using the PCA as a method of reduction of features.

Several works exist for sentiment analysis at document level treating movie reviews. Pang [14], is the first who experiments this approach with machine learning. The proposed method which proved to be good in text categorization, did not achieve good performance for the sentiment classification. It also demonstrated that the binary representation is more significant than the frequency representation.

Zhang et al. [28] used Classification by minimizing the error (CME) to assign a score of opinion to each sentence of the blog. They then defined some features based on subjectivity and the relevance in all the sentences on the blog. A classifier SVM is used to assign a score to each document based on the values of the features defined. So blogs are classified according to their final score based on the relevancy score multiplied by the score of opinion.

A freely available corpus Arab (OCA) for sentiment analysis is proposed by Rushdi - Saleh et al [19]. The OCA corpus consist of 500 Arabic reviews collected from specialized movie-related Arabic webpages, 250 of them are considered as positive and the rest as negative opinions. In addition, various experiments were conducted on this corpus using machine learning algorithms such as NB and SVM.

Govindarajan [6] proposed a hybrid method of classification based on the coupling of NB and genetic algorithm (GA). In this method, first the two basic classifiers NB and GA are built to assign a score of opinion, and then the classification of a new review is done by combining the predictions of two basic classifiers, by a majority vote. The author used a set of 2000 movie reviews, 1000 marked positive and 1000 marked negative (they have been extracted from the corpus of Bo Pang). The hybrid method is compared with the two base classifiers NB and GA and their performance is analyzed in terms of precision. The results showed that the hybrid method has improved the performance compared to the basic methods.

Nguyen et al. [10] proposed a new type of feature named "rating-based feature" and evaluate it on movie reviews. This feature is based on the fact that scores (in which users use to categorize entities in reviews) could provide useful information to improve the performance of classification of opinion. For a review with no associated score, the authors use a regression model to predict the score. They combine rating-based feature with unigram, bigram, and tri-gram.

In [23], the authors propose a model for sentiment analysis of movie reviews by a combination of natural language processing and machine learning approaches. First, various schemes of pre-processing are applied to the data set. Secondly, the behavior of the two classifiers, NB and SVM is studied in combination of different schemes of feature selection. The results of classification show clearly that linear SVM give more precision than NB classifier.

In [5], the objective of the author is analyzing the opinions in Arabic Tweets with the presence of dialectical words. Dialectical words were translated to their corresponding words of the Modern Standard Arabic (MSA) by the use of dialect lexicon. Both NB and SVM classifiers were used to determine the polarity of tweets. These classifiers build their classification models from the two versions of the same data set. The first version consists of tweets that contain dialectical words and the second version consists of tweets containing translated dialectical words. The results show that the replacement of dialectical words with their corresponding words MSA improves the accuracy of classifiers (3%).

2.2 Unsupervised Approaches

Unsupervised approaches exploit a sentiment lexicon which is either built independently of any corpus (built from existing dictionaries), or generated from the corpus (words containing opinion are extracted directly from the corpus). The objective of these lexicons is to index the most words carrying possible opinion. So we can further divide unsupervised methods into dictionary-based and corpus-based relative to how the lexicon is built. If a document contains many subjective words, then it is considered as a document containing opinions [8] [11].

Turney presents a simple algorithm for classifying reviews as recommended or not recommended [24]. The classification of a review is predicted by the average of the semantic orientation of the sentences in the review that contain adjectives or adverbs. A sentence has a positive semantic orientation when it has good associations and a negative semantic orientation when it has bad associations. The semantic orientation of a sentence is calculated as mutual information (PMI Point wise Mutual Information) between the given sentence and the word "excellent" and also between the given sentence and the word "poor". Finally, the review is classified according to the average orientation of the sentences it contains.

To overcome the problem of domain dependency in the sentiment analysis, Rothfels and Tibshirani [18] propose an approach for treating movie reviews using the automatic selection of items with positive/ negative opinions. They choose two seed reference sets, one positive and one negative to calculate the semantic orientation. Semantic orientation is calculated as the PMI mutual information between the given sentence and the seed reference set.

Baloglu and Aktas [2] proposed a lexicon-based approach for sentiment analysis. This approach is divided into three phases. The first phase is the crawling phase; the data are collected from blogs on the Web. The second phase is the analysis phase, in which the data are analyzed to extract useful information (predefined keywords) and uses SentiWordNet to determine the sentiment score of each keyword and finally the review is classified based on the average of these scores. The third phase is the visualization, in which information is displayed to better understand the results.

A sentiment analysis system named Document based Sentiment Orientation System is proposed in [20]. It uses an unsupervised approach which determines the sentiment orientation of the movie reviews and Word Net to identify synonyms and antonyms of opinion word list. Negation is also handled in the proposed system. The documents are classified as positive, negative or neutral. The approach provides a summary of the total number of positive and negative documents.

2.3 Semi-Supervised Approaches

Semi-supervised approaches combine the strengths of machine learning and lexicon-based approaches by taking into account the linguistic processing of lexicon-based approaches before starting the learning process as in machine learning approaches.

Ohana and Tierney [12] assesses the use of SentiWordNet to the task of document level sentiment classification using the Polarity data set of film reviews. Initially, the vocabulary has been employed to calculate the score of positive and negative terms found in a document and determine the sentiment direction. Then this method was improved by the construction of a relevant feature by using SentiWordNet as a source and applied to the SVM classifier. The results indicate SentiWordNet could be utilized as an important resource for sentiment classification tasks.

An ensemble learning method based behavior-knowledge space BKS is proposed in [29], in which four basic classifiers are used, tow unsupervised; SWS (single weighted sum of sentiment words) and WSC (weighted sum of sentiment words is based concepts), and two supervised: SVM and k-nearest neighbors (KNN). The experiment results not only explain the effectiveness of the proposed method, but also show that this method is much higher than the basic classifiers. In addition, this method is better than the voting method.

Two different ways to combine the analysis of discourse RST (Rhetorical Structure Theory) with the sentiment analysis are proposed in [3]: (i) a recurrent neural network on the structure of the RST and (ii) a reweighting discourse units. They show that the reweighting discourse units can lead to substantial improvements for the sentiment analysis based-lexicon, and show that the recurrent neural network using RST structure offers significant improvements over the basic classification methods.

3 Comparative study

We use the following components to make a comparative study;

- Approach used
- Technique used
- Data source used
- Feature construction used
- Quantitative Evaluation

As we have seen in the related work, the three main classes of approaches used for sentiment analysis are: supervised approach, unsupervised approach, and semi-supervised approach. Table 1 below categorizes the surveyed works into these classes

Table 1. Comparative study by approach

Approach	Ref
Supervised	[6], [10], [13], [14], [17], [19], [23], [26], [28], [30], [5]
Unsupervised	[2], [18], [20], [24]
Semi-supervised	[3], [12], [29]

From Table 1, the supervised approach is the dominant one compared to the other approaches (semi-supervised, unsupervised approaches).

Sentiment classification techniques are usually distinguished based on approach been used. Several machine learning algorithms are used as a technique for document classification. Prominent methods are: NB, Maximum Entropy, KNN, and SVM. The unsupervised learning methods are divided into dictionary-based and corpus-based. The surveyed works used different techniques. We have collected the various techniques used and presented the result in Table 2.

Table 1. Comparative study by technique

Technique used	Ref
SVM	[10], [13]
SVM & NB	[5], [19], [23], [26], [30]
SVM & NB & Maximum Entropy	[14]
SVM & CME	[28]
SVM & RNA	[17]
SVM & GA	[6]
Dictionary-based	[2], [20], [24]
Corpus-based	[18]
Dictionary –based & SVM	[12]
Dictionary –based & SVM & KNN	[29]
RNA	[3]

From Table 2, the most used methods seem to be based on SVM and NB, for the supervised approach. For unsupervised approach, dictionary-based approaches are the most frequent across the majority of works.

Several data sources used in the sentiment classification studies, but in the sentiment classification at document level the most used is the reviews. Table 3 below presents data sources used in the surveyed different works.

Table 2. Comparative study by data sources

Data Sources	Ref
Reviews of video games	[13]
Amazon reviews	[17]
Mobile reviews of We Chat	[30]

Product reviews	[26]
Movie reviews	[3], [6], [10], [12], [14], [18], [20], [23]
Blogs	[2], [28]
Corpus OCA	[19]
Epinions review	[24]
Hotel reviews	[29]
Tweets	[5]

From Table 3, The movie reviews are widely used in the field of sentiment classification at document level.

The goal of feature construction is to select good features for sentiment classification. Many features are considered for sentiment analysis: unigram, bigram, n-gram, POS and opinion words. Table 4 presents the feature construction used of the surveyed works.

Table 3. Comparative study by feature construction

Feature Construction	Ref
Sub graphs extracted from syntactic dependency trees	[13]
POS	[17]
n-gram	[30]
unigram	[26]
unigram, bigram, POS	[14]
unigram, bigram, trigram	[19]
bigram	[6]
rating-based feature, unigram, bigram, trigram	[10]
TF-IDF, unigram, bigram, trigram	[23]
POS, semantic orientation, PMI	[24]
4-gram, semantic orientation, PMI	[18]
POS, opinion words	[20]
POS	[12]
reweighting discourse units	[3]

From Table 4, the selected features most used are: Unigram, bigram, POS, semantic orientation and PMI.

The performance of different works used for opinion mining is evaluated by calculating various metrics like accuracy, precision, recall and F-measure. The Accuracy is calculated by using the following equation.

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FP+FN} \quad (1)$$

With

- TP: True Positive.
- TN: True Negative.
- FP: False Positive.
- FN: False Negative

The following figure (Fig. 1) Summarize the accuracy of works using movie review data set [31][1].

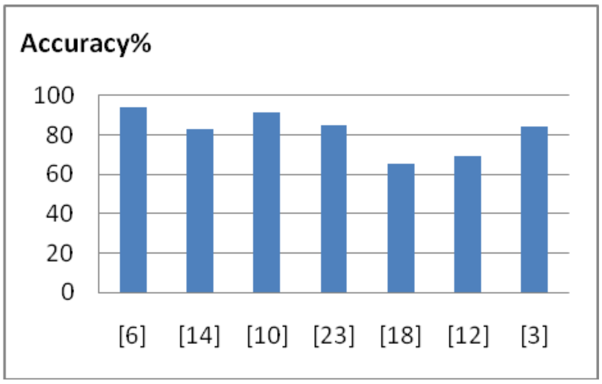

Fig. 1. Sentiment analysis for movie review dataset

From the figure (Fig. 1), we find that work [6], which used a hybrid method of classification based on the coupling of NB and GA, gave the best results in terms of accuracy. Even for work [10], which used a new type of feature named “rating-based feature”.

4 Conclusion and Future Work

The number of documents expressing opinions is constantly increasing on the World Wide Web. Sentiment analysis at document level provides an overall opinion of the document on a single entity. In this article, we have presented an overview of related work of sentiment analysis at document level, mainly the approach of machine learning is considered as dominance at this level. The main classifiers used are SVM and NB. The more text representation used is "bag of words" representation, but supervised approaches using n-grams features cannot properly modeled the negation, and cannot correctly identify the complex sentiment expressions due to the loss of information incurred when representing texts with bag of words models. Most of the work in the field uses movie review data for classification. The classification of the documents is not always relevant:

---

[1] Note : where multiple results were provided in these works, we selected only the best results

- In many applications, the user needs to know what aspects of entities are liked and disliked by consumers, but this level of classification can not extract them.
- It is not easily applicable to documents (such as forum discussions, blogs, and news articles) that evaluate or compare several entities.
- Different emotions on the different aspects of an entity cannot be extracted separately.

It is therefore necessary to move at sentence level, i.e., to classify sentiment expressed in each sentence. However, there is no fundamental difference between document and sentence level classifications because sentences are just short documents.

In the future work, more efforts would be done to improve the performance measures, for this purpose, there is a need for finer-grained granularity at the aspect level. Furthermore, the languages that have been studied mostly are English. Presently, there are very few researches conducted on sentiment classification of other languages like Arabic. We intend to propose a new method for aspect based sentiment analysis of Arabic texts.

References

1. Anitha, N., Anitha, B., Pradeepa, S.: Sentiment Classification Approaches – A Review. J. IJIET. Vol. 3 , 2319 – 1058 (2013)
2. Baloglu, A., Aktas, M.S.: An Automated Framework for Mining Reviews from Blogosphere. J. Advances in Internet Technology, vol. 3 234-244, (2010)
3. Bhatia, P., Ji, Y., Eisenstein, J.: Better Document-level Sentiment Analysis from RST Discourse Parsing. In: Empirical Methods in Natural Language Processing pp. 2212–2218. EMNLP, Lisbon (2015)
4. Chen, Y.F., Miao, D.Q., Li, W., Zhang, Z.F.: Semantic Orientation Computing Based on Concepts. J. CAAI Trans. vol. 6, 489–494 (2011)
5. Duwairi, R.M.: Sentiment Analysis for Dialectical Arabic. In: 6th ICICS International Conference on Information and Communication Systems, pp. 166 - 170 (2015)
6. Govindarajan, M.: Sentiment Analysis of Movie Reviews using Hybrid Method of Naive Bayes and Genetic Algorithm. J. Advanced Computer Research. 139, 2249-7277 (2013)
7. Liu, B.: Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, (2012)
8. Mishne, G., Multiple Ranking Strategies for Opinion Retrieval in Blogs. TREC Blog Track , TREC, (2006)
9. Nilesh, M.S., Shriniwas Deshpande, Vilas Thakre, "Survey of Techniques for Opinion Mining. J. IJCA. 13, 30-35 (2012)
10. Nguyen, D.Q., Nguyen, D.Q., Vu, T., Pham, S.B.: Sentiment Classification on Polarity Reviews: An Empirical Study Using Rating-based Features. In: 5th Workshop on Com-

pu-tational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp 128–135. Maryland (2014)

11. Oard, D. W., Elsayed, T., Wang, J., Wu, Y., Zhang P., Abels, E. G., Lin, J. J., Soergel, D.: TREC 2006 at Maryland: Blog, Enterprise, Legal and QA Tracks. TREC, (2006)
12. Ohana, B., Tierney, B.: Sentiment Classification of Reviews Using SentiWordNet. In: 9th. IT&T Conference, pp. 22-23 Dublin Institute of Technology, Dublin (2009)
13. Pak, A., Paroubek, P. : Classification en polarité de sentiments avec une représentation textuelle à base de sous-graphes d'arbres de dépendances. TALN (2011)
14. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Empirical Methods in Natural Language Processing, pages 79-86. EMNLP (2002)
15. Pang, B., Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In: 42th Annual Meeting of the Associatoin for Computational Linguistics, pp. 271-278. ACL, Barcelona (2004).
16. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and TrendsR in Information Retrieval Vol. 2,pp.1–135 (2008)
17. Rafrafi, A., Guigue,V., Gallinari, P.: Réseau de neurones profond et SVM pour la classi-fi-cation des sentiments. In: CORIA, pp. 121-133 (2011)
18. Rothfels, J., Tibshirani, J.: Unsupervised sentiment classification of English movie reviews using automatic selection of positive and negative sentiment items. CS224N-Final Project (2010)
19. Rushdi-Saleh, M., Martín-Valdivia, M.T, Ureña-López, L.A., Perea-Ortega, J.M: OCA: Opinion corpus for Arabic. J. ASIS&T. 62, 2045–2054 (2011)
20. Sharma, R., Nigam, S., Jain, R.: Opinion Mining Of Movie Reviews At Document Level. J. IJIT. 3, (2014)
21. Sindhu, C., ChandraKala, S.: A Survey On Opinion Mining And Sentiment Polarity Clas-sification. J. IJETAE. 3, (2013)
22. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.N., Potts, C.: Recursive deep models for semantic compositionality over a sentiment tree bank. In: Empir-ical Methods for Natural Language Processing. EMNLP, (2013)
23. Tripathi, G., Naganna, S.: Feature Selection And Classification Approach For Sentiment Analysis. J. MLAIJ. 2201, (2015)
24. Turney, P. D.: Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: 40th annual meeting of the Association for Computational Linguistics, pp. 417–424. ACL, Philadelphia(2002)
25. Vinodhini, G., Chandrasekaran, R. M.: Sentiment Analysis and Opinion Mining: A Survey. J. IJARCSSE, 2277, 282-292 (2012)
26. Vinodhini, G., Chandrasekaran, R.M.: Effect of Feature Reduction in Sentiment analysis of online reviews. J. IJARCET, 2278 – 1323 (2013)
27. Wang, S., Manning, C.D.: Baselines and bigrams: simple, good sentiment and topic clas-sification". In: 50th Annual Meeting of the Association for Computational Linguistics, pp. 90 – 94. ACL,(2012)
28. Zhang, Q., Wang, B.,Wu, L., Huang, X.: FDU at TREC 2007 : Opinion Retrieval of Blog Track. In : E. M. Voorhees, , L. P. Buckland (eds), TREC 2007, vol. Special Publication ,pp.500- 274, National Institute of Standards and Technology (NIST), Maryland(2007)

29. Zhang, Z., Miao, D., Wei, Z., Wang, L.: Document-Level Sentiment Classification Based on Behavior-Knowledge Space Method . In: ADMA, pp. 330–339, Heidelberg (2012)
30. Zhang, L., Hua, K., Wang, H., Qian, G., Zhang, L.: Sentiment Analysis on Reviews of Mobile Users. In: 11th International Conference on Mobile Systems and Pervasive Com-puting, pp. 458 – 465. Procedia Computer Science 34, (2014).
31. http://www.cs.cornell.edu/people/pabo/movie-review-data/.
32. https://www.projet-doxa.fr/index.php