

Sentiment Analysis using Aspect Level Classification

Priyanka Patil¹, Pratibha Yalagi²

¹Department of Computer Science and Engineering, Walchand Institute of Technology, Solapur, Maharashtra, India

²Department of Information Technology, Walchand Institute of Technology, Solapur, Maharashtra, India

ABSTRACT

The natural language text is analyzed by using sentiment analysis and classified into positive, negative or neutral based on the human emotions, sentiments, opinions expressed in the text. The user reviews and comments on movies on the web are increasing day by day. And to make a decision in movie planning, these reviews are useful for other users. To perform manual analysis of a huge number of reviews is practically not possible. Hence, to solve such problem we require an automated approach of a machine that mines the overall sentiment or opinion polarity from the reviews. This paper analyses the movie reviews using sentiment analysis and text mining techniques. In this technique, Sentiment scores are assigned to the aspects with respect to the words used. Sentiment analysis is performed at three different levels. These are documented level, sentence level, and aspect level. Most of the previous work is done in the document or sentence-level sentiment analysis. We focus on the aspect level opinion mining of movie reviews. From a given set of reviews of a movie, we get a sentiment profile. By using SentiWordNet different approaches are proposed for sentiment analysis. It contains two-word phrases and linguistic rules together for opinion mining. The Send score algorithm is devised to perform this function. For attempting to mine and understand the user's feedback data, we perform an aspect level sentiment extraction. To predict the polarity of the reviews, a priority-based algorithm forms the rule base for the classifier. First, we perform a cleanup on the review data, then Send score algorithm is utilized to generate the aspect document matrix.

Keywords: Aspect-Level Sentiment Analysis, Opinion Mining, Text Mining, Sentiment Profile, Send Score

I. INTRODUCTION

Nowadays, customer reviews have become an important source of information for every field where trying to understand customer provided feedback. Therefore user review is essential in almost all the fields. The common users are becoming more and more comfortable with the Web, an increasing number of people are writing reviews on it. Hence, the number of reviews a system receives from users grows rapidly and internet slang words make it difficult to comprehend. Today's review analysis systems mostly determine the polarity of a sentence as a whole. However, the reader might be interested to know the aspect-wise summary of opinions instead of the overall picture. Sentiment analysis is the process which analyzes text to identify positive and negative opinions. Sentiment analysis is the task of finding subjectivity in natural language. From the areas of natural language processing, machine learning, data

mining and sentiment analysis approaches are drawn. Sentiment Analysis is the computational study of people's emotions, opinions and attitudes toward an entity. The entity can be individuals, topics or events. And these topics are most likely to be covered by the reviews. The two terms sentiment analysis or opinion mining is interchangeable. It expresses a mutual meaning. But, according to some researchers, opinion mining and sentiment analysis have slightly different notions. Sentiment Analysis identifies the sentiment expressed in a text then analyzes people's opinion about an entity while Opinion Mining extracts and analyzes it. Therefore, the target of sentiment analysis is to find opinions, identify the sentiments they express, and then classify their polarity.

Initially, sentiment analysis was directed towards determining the general polarity of a certain sentence or document. But the interest has recently shifted towards a

more qualitative analysis, which aims to detect the different features or aspects of a topic towards which opinions are expressed. For example, if we are interested in the analysis of a movie review to capture the opinions of the reviewers towards aspects such as the plot, the cinematography, or the performance of the actor. In aspect-level sentiment analysis, most challenging part is that a system needs to detect the relevant aspects. And then these can be associated with a polarity. This paper presents the analysis of unstructured data using movie reviews as a case study. This is accomplished by performing sentiment analysis on the review data. Our proposed system for movie reviews gives an aspect wise summary rather than the overall sentiment of a review.

II. RELATED WORK

In the recent years, Sentiment analysis has been an area under study. Many researchers have contributed a lot to this study. One of the notable contributors is Bing Liu [1]. His book on Natural Language Processing focused on the mining of the huge volume of texts with opinions. Li Zhang, Weiran Xu, Si Li [2], proposed a method using a slack function, for object extraction and aspect identification. For that, he used dictionary based method to analyze the sentiment for the aspect. M. Ikonomaki, S. Kotsiantis, V. Tampakas [3], in their paper, proposed the various machine learning algorithms which are used to perform classification.

Su Su Htay and Khin Thidar Lynn [4], proposed a novel idea for using Pattern Knowledge to find the sentiment words. In this paper customer reviews are taken as input and opinion summary is produced as output. Theresa Wilson, Janyce Wiebe and Paul Hoffmann [5], together presented a system to automatically identify the contextual polarity for a large set of sentiment terms. It is performed by achieving phrase-level sentiment analysis. Xiaowen Ding, Philip S. Yu [6] presented an algorithm on opinion orientation. To solve the problem by utilizing external evidence and linguistic rules, the algorithm was used which provide a holistic lexicon-oriented methodology. Chee Kian Leong a, Yew Hair Lee b, Wai Keong Mak [7], together devised a system that is capable of mining SMS texts with emoticons. S.L. Ting, W.H. Ip, Albert H.C. Tsang [8] shown naïve Bayes classification is one on the simplest classifiers to learn and implement by the highlighting performance of naïve Bayes classification in their paper.

Hu and Liu [9] introduced the most challenging part of aspect-based sentiment analysis, where polarity is not assigned to documents or sentences, but to single aspects discussed in them. Qiu et al. [10] continued to achieve the idea in which opinion words can be used to detect aspects, by focusing on single reviews. Brody and Elhadad [11] have tried to extract topics as aspects. To determine the polarity of each topic/aspect, they start from a set of seed opinion words. And by using a label propagation algorithm, they propagate their polarities to other adjectives. For that, they used a label propagation algorithm. Zhu et al. [12] use product aspects and some aspect-related terms as input for their algorithm. And then new aspect-related terms have discovered by applying a bootstrapping algorithm based on co-occurrence between new candidate terms and seed terms. Abbasi et al. [13] proposed sentiment analysis techniques for classification of web forum postings in multiple languages by utilizing of stylistic and syntactic features. Also, they introduced entropy weighted genetic algorithm (EWGA). This is a hybrid genetic algorithm which uses the information gain heuristic to improve feature selection.

III. PROPOSED WORK

The sentiment classification has three levels, document-level, sentence-level, and aspect-level. In this paper, we have explored a new SentiWordNet based scheme for both aspect-level and document-level sentiment classification. The document-level classification use different linguistic features, those are ranging from Adverb+Adjective combination to Adverb+Adjective+Verb combination. In this paper, we have also devised a new domain specific heuristic for aspect-level sentiment classification of movie reviews. For this scheme, we locate the opinionated text around the desired aspect/feature in a review and compute its sentiment orientation. This is performed for all the movie reviews. And from all those reviews, the sentiment scores on a particular aspect are then aggregated. This process is carried out for all aspects under consideration and then a summarized sentiment profile of the movie on all aspects is presented in an easy to visualize and understandable pictorial form.

We have used the publicly available library SentiWordNet [15]. The SentiWordNet approach

contains sentiment score for each opinion containing the terms of the text by a lookup in its library. In this lexical resource, each term t is occurring in WordNet is associated with three numerical scores. These are $obj(t)$, $pos(t)$, and $neg(t)$, which describing the objective as positive and negative polarities of the term t , respectively. For computing these three scores, the results produced by eight ternary classifiers are combined. For use SentiWordNet, first, we need to extract relevant opinionated terms which then lookup for their scores in the SentiWordNet library. To make use of SentiWordNet we need to take a lot of decisions regarding the linguistic features which are used. They decide how much weight is to be given to each linguistic feature, and for taking sentiment scores the aggregation method is used. We have implemented the SentiWordNet algorithmic formulation for both aspect-level and document-level sentiment classification.

The sentiments are the words or sentences which represent the view, emotion or opinion that is expressed. It can be positive, negative or neutral. The proposed Sentiment Analysis on movie reviews is based on three important parts viz Data Extraction, feature extraction and classification.

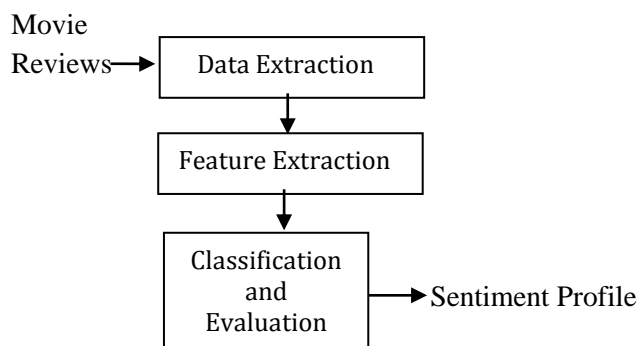


Figure 1: System Architecture of proposed system

The following steps will show the process of the proposed system which is discussed in paper shown in fig [1]:

1. Data Extraction: The data extraction contains three steps viz, data collection, parsing and reviews collection. The first step is data collection. The social networking site consists of various blogs and they relate to various topics worldwide. Instead of taking whole blogs with reviews, we will rather search for a particular topic as a

movie, and download all its web pages. Then extract them in the form of text files.

In this system, we have used dependency parsing to capture the relation between feature and their associated opinions. The idea is to capture the association between any specific feature and expressions of opinion that come together to describe that feature. Among the large set of data, reviews are collected by identifying sentiments.

2. Feature Extraction: In feature extraction, initially POS tagging is performed. The approaches based on SentiWordNet aim the term profile of the review document and extract terms having a desired POS label (such as adjectives, adverbs or verbs). This indicates that before applying the SentiWordNet based formulation; the POS tagger should be applied to review the text. It tags each term occurring in the review text. Then some selected terms with their desired POS tag are extracted and the sentiment score of each extracted term is obtained by using the SentiWordNet library. The scores for all terms in a review are then aggregated using weighting and aggregation scheme. Before sending all words to classifier minimize the whole bag of the word. The last step is feature selection in which the desired features are identified for topics.

3. Classification and Evaluation: The classification is done by the SentiWordNet based publicly available library that provides positive, negative and neutral scores for words. The output is presented in sentiment profile.

Aspect-level Sentiment Analysis

The reasonable measure of positivity or negativity expressed in a review is document-level sentiment classification. However, it may be a good idea to explore the sentiment of the reviewer about various aspects of the item in selected domains, expressed in that review. But, practically most of the reviews have a mixture of positive and negative sentiment about different aspects of the item. Hence, it may be difficult and inappropriate to implore on an overall document-level sentiment polarity that is expressed in a review for the item. Therefore, the document-level sentiment classification is not a suitable, complete and comprehensive measure for detailed analysis of aspects of the item under review.

The aspect-level sentiment analysis allows analyzing the positive and negative aspects of an item. Such analysis is often domain specific. The aspect-level sentiment analysis includes the following: (a) identification of aspects is to be analyzed, (b) locating the opinionated content about identified aspect in the review, and (c) determination the sentiment polarity of views expressed about an aspect.

In this paper movie review is a focused domain, so we explored the aspect-level sentiment analysis of the movie reviews. In that, the first step is to identify which aspects should be considered in movie domain. Sometimes a particular aspect is given by different words such as acting, screen presence, screenplay. Hence we created an aspect vector for all aspects under consideration. After creation of aspect vectors, a review is parsed sentence-by-sentence. For each sentence, we search for the presence of opinion about an aspect. If we found, then use the SentiWordNet based approach to determine its sentiment polarity. So this is performed for all the sentences in a review and subsequently for all reviews of a movie. To obtain an opinionated analysis of a particular aspect the derived scores for that aspect from all the reviews of a movie are aggregated. Thus first and opinionated content about an aspect in a review located and then the SentiWordNet based approach is used to compute its sentiment polarity. We used the SentiWordNet (AAC) scheme for this purpose. When an aspect indicating term is located, we first lookup up to 5-gram backward for the occurrence of adjectives or adverb+adjective combines. If we do not find the term, we search up to 5-gram forward for their occurrence. The lookup terminates at 5-gram or sentence boundary whichever is encountered first in both cases. Then for these terms, the sentiment polarity is computed using the SentiWordNet based formulation for AAC.

IV. CONCLUSION

In this paper, we have applied the sentiment analysis techniques on movie reviews domain. These techniques are used to analyze the feedback given by other users for movie reviews. Existing Systems are generally capable of obtaining the polarity for documents or sentences as a whole. The proposed method is a one that determines the polarity of every aspect in a multi-aspect sentence. In this paper, we proposed feature-based heuristic scheme

for aspect-level sentiment classification of movie reviews. The aspect level sentiment classification produces an easy and accurate sentiment profile of a movie based on various aspects. The aspect-level sentiment profile result is identical to the document-level sentiment classification of movie reviews. Though, the aspect-level sentiment profile produces a more accurate and focused sentiment summary of a particular movie and is more useful for the users. We designed aspect-level sentiment analysis, which is a novel and unique way for obtaining a complete sentiment profile of a movie based on multiple reviews on different aspects of evaluation. The derived sentiment profile is easy to understand, informative, and extremely useful for users. Also, the algorithm used for aspect-level sentiment profile is very simple, fast in producing results, quick to implement, and does not require any previous training. It produces very useful and detailed sentiment profile of a movie on different aspects. This part of the implementation can also be used as an add-on step in movie recommendation systems that use content-filtering, collaborative-filtering or hybrid approaches. The sentiment profile can be used as an additional filtering step in the design of movie recommender systems. This profile is a valuable form of sentiment analysis and subsequent form of information expressed by a large number of users about a movie.

V. REFERENCES

- [1] Liu. B, *Sentiment analysis and subjectivity*, In N. Indurkha & F. J. Damerau (Eds.), *Handbook of natural processing language* (pp. 627-666), Boca Raton: CRC Press, 2010.
- [2] Li Zhang, Weirman Xu, Si Li, *Aspect Identification and sentiment analysis based on NLP*, Proceedings of IC-NIDC 2012, 660-664, 2012.
- [3] M. Ikonomaki, S. Kotsiantis , V. Tampakas, *Text Classification Using Machine Learning Techniques*, WSEAS transactions on computers, Issue 8, Volume 4, , pp. 966-974, 2005.
- [4] Su Su Htay and Khin Thidar Lynn, *Extracting Product Features and Opinion Words using Pattern Knowledge in customer reviews*, Hindawi Publishing Corporation The scientific World Journal, Article ID 394758, (1-5), 2013.

- [5] Theresa Wilson, Janyce Wiebe and Paul Hoffman, *Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis*.
- [6] Xiaowen Ding, Philip S. Yu , *A Holistic Lexicon Based Approach to Opinion Mining*, 2007.
- [7] Chee Kian Leong, Yew Haur Lee b, Wai Keong Mak, Mining sentiments of SMS texts for teaching evaluation, *Expert Systems with applications* 39:2584-2589, 2012.
- [8] S.L. Ting, W.H. Ip, Albert H.C. Tsang, Is Naive Bayes a Good Classifier for Document Classification? *International Journal of Software Engineering and Its Applications* Vol. 5, No. 3, 2011.
- [9] Hu M., Liu B., Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168–177. ACM, 2004.
- [10] Qiu G., Liu B., Bu J., Chen C., Expanding domain sentiment lexicon through double propagation. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence*, volume 9, pages 1199–1204, 2009.
- [11] Brody S., Elhadad N., An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 804–812, Association for Computational Linguistics, 2010.
- [12] Zhu J., Wang H., Zhu M., Tsou B. K., Ma M., Aspect-based opinion polling from customer reviews. *IEEE Transactions on Adjective Computing*, 2(1):37–49, 2011.
- [13] A. Abbasi, H. Chen, and A. Salem, “Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums,” In *ACM Transactions on Information Systems*, vol. 26 Issue 3, pp. 1-34, 2008.
- [14] Tsung-Hsien Chiang, Hung-Yi Lo, Shou-De Lin, *A Rankingbased KNN Approach for Multi-Label Classification*, 2012.
- [15] Senti WordNet, available at <http://www.sentiwordnet.isti.cnr.it>