

User Modeled Aspect based Aspect Sentiment Analysis

Rupal Bhargava

Department of Computer Science and
Information System

Birla Institute of Information and Technology
and Science, Pilani

Pilani, India

Yashvardhan Sharma

Department of Computer Science and
Information System

Birla Institute of Information and Technology
and Science, Pilani

Pilani, India

Abstract— Rapid increase in blogs, forum and social networking sites has drastically changed the way people communicate and express their opinions. This trend had inspired many research works targeting at the automated analysis of opinions. In this paper, we present aspect based sentimental approach which will bring out the attitude of the document according to user priority. The objective is to classify review into polarity class considering preferences of the user. As we are dealing with aspect based sentimental analysis, we will classify the polarity of the sentiments for each aspect and then, on the basis of user priority we will decide the overall polarity. The objective or irrelevant text is filtered out with respect to the given query. The proposed system does not require any labeled data, as raw text is taken as input in the form of multiple documents. The only supervision taken under consideration is using WordNet, Part of Speech Tagger. Our System achieves considerable improvement over the baseline and has better accuracy compared to the existing system, on the same dataset.

Keywords — Sentiment analysis, opinion mining, feature extraction, sentiment classification

I. INTRODUCTION

Sentimental analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Sentiments and insinuations play a vital role in our lives. Before we take a decision, we appraise the opinions of others before finally reaching a cessation. With the advancement of forums, blogs and social networking sites, there had been a considerable increase in sharing of opinions and sentiments on the wires. Cybnauts exhibit really divergent views on a certain subject, and many a times end up agreeing to a disagreement. Task of Sentiment Analysis is to find opinions of people, identify the sentiment they express and then classify their polarity. There are mainly three classification levels in sentiment analysis i.e. document level, sentence level, and aspect level sentiment analysis.

Document level Sentiment Analysis considers the whole document as a basic information unit and then classifies the document as expressing a positive or negative sentiment. Sentence level Sentiment Analysis initially finds the subjective sentences in the document and then processes each sentence for the opinion. Classifying a document at either document or sentence level does not provides the necessary detailed opinions on all the aspects of the entity, so here the Aspect Level sentiment analysis comes into picture. Aspect level Sentiment analysis aims to classify the sentiment with respect to the specific aspect of the entities.

Substantial amount of research has been done in automated text analysis, viewpoint analysis and/or opinion extraction. These methods bring out the overall attitude of the document but it fails to detect the beliefs about individual aspects of the subtopics. Also they are unable to relate the extracted views to a specific issue properly, as it gives an overall sentimental analysis rather than providing a pertinent one. Immense increase in information overload of view-points on the Internet calls for other methods to extract the best useful information for the motive of decision making. Opinion mining or Viewpoint Excavation has been treated as a classification technique which stratifies documents or products as good/bad, positive/negative etc.

II. LITERATURE REVIEW

Sentiment Analysis is the computational treatment of sentiments and emotions expressed in a text [5]. There are four main tasks that need to performed for the same i.e. opinion identification, feature extraction, sentiment classification and visualization. Sentiment analysis at both the document level and sentence level has been too coarse to determine precisely what users like or dislike. In order to address this problem, sentiment analysis at the attribute level is aimed at extracting opinions on product's specific attributes from reviews.

Pang et al. [8] summarized the extracts by examining the alliance between subjectivity detection and polarity

classification which had the same amount of polarity information as that of the full review. Zhang et al [15] proposed a work which used a keyword matching strategy to identify and tag product features in sentences. Mukherjee et al. [7] have used the Wikipedia knowledge to filter the irrelevant objective text and proposed a weakly supervised approach for sentiment classification of movie reviews. Popescu et al [9] developed an unsupervised information extraction system called OPINE, which extracted product features and opinions from reviews. OPINE first extracts noun phrases from reviews and retains those with frequency greater than an experimentally set threshold and then assesses those by OPINE's feature assessor for extracting explicit features. The assessor evaluates a noun phrase by computing a Point-wise Mutual Information score between the phrase and meronymy discriminators associated with the product class.

Ranade et al. [11] summarized online debates by extracting highly topic relevant and sentiment rich sentences. This was done by extracting Topic relevant, document relevant and Sentiment Relevant features present in the debates. Zhai et al. [16] proposed an approach of Opinion Feature Extraction based on Sentiment Patterns, which takes into account the structure characteristics of reviews for higher values of precision and recall. With a self-constructed database of sentiment patterns, sentiment pattern matches each review sentence to obtain its features, and then filters redundant features regarding relevance of the domain, statistics and semantic similarity. Virmani et al. [13] proposed an algorithm which clubbed aspect level with the opinion value and sentiment value to help conclude summarized value of remark about a student. Aspect tree is constructed which has different level and weights assigned to each branch to identify level of aspect. Wang et al. [14] proposed a feature based vector model and a weighing algorithm for sentiment analysis in Chinese product reviews. Also a feature extraction method based on dependency parsing is presented to identify the corresponding aspects that opinion words modify. Somprasertsri et al. [12] dedicated their work to properly identify the semantic relationships between product features and opinions. His approach is to mine product feature and opinion based on the consideration of syntactic information and semantic information by applying dependency relations and ontological knowledge with probabilistic based model.

Feczko et al. [2] had used sentiment analysis to analyze user product reviews for multiple products, identify and parse the positive and negative viewpoints and display the aggregated information in a user friendly and practically useful manner but they have not worked on the feature level. Hu et al. [3] extracted the product features, identified the opinion sentences, and summarized the results. Their feature extraction algorithm is based on heuristics that depend on feature terms' respective occurrence counts. They use association rule mining based on the Apriori algorithm to extract frequent item sets as explicit product features. Their work is closely related to ours, but they have not considered

the preferences of the user. Also they have just summarized the extract to find the opinionated sentences. Mukherjee et al. [6] have focused mainly on the feature extraction process to identify the opinions for product review. This work is different from ours by the perspective of text summarization.

III. PROPOSED SYSTEM

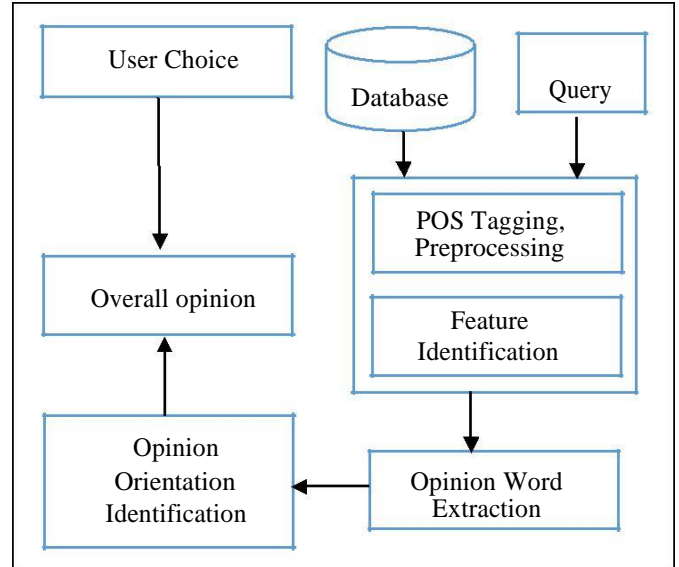


Figure 1 Proposed System

Figure 1. Presents the proposed algorithm for the sentiment classification system. Inputs to the system is the search query in form of name of that topic and user preferences (not compulsory) and output is the overall opinion about the topic considering the user's aspects. Database is searched for the documents related to the query. Documents then retrieved are used for the further process. Usually features are nouns or noun phrases. Hence POS tagging is crucial. POS tagging is done using Stanford POS tagger. Some preprocessing of words is also performed which includes removal of stop words and stemming.

Frequently occurring features are then extracted using knearest neighbor technique. This is the important phase, as in absence of any prior knowledge about the domain of the document we need to grab hold of all the features that can possibly be of that domain. Extraneous features can be pruned by the user if needed.

After obtaining the complete feature list we extract corresponding opinions about individual features. This is done by identifying the subjective sentences i.e. the sentences that contain some opinion. WordNet [10] is used to identify the opinionated words which are generally adjectives. An option is also provided to the user if he/she also wants to consider the objective sentences, because there are sentences that may present an implicit opinion about something. All the opinions are extracted for the target feature with the help of

dependency graph built for clustering, for the purpose of feature extraction.

Opinion orientation are then identified with the help of SentiWordNet [1]. Also sentiment score is calculated for each target feature using all the extracted opinions. Overall opinion about the topic or subject is decided via considering user preferences and the score of individual target feature.

IV. RESULTS

The proposed technique has been implemented in java. We conducted our experiment using dataset by Hu and Liu, the customer reviews of 5 electronic products: 2 digital camera (Nikon, Canon), 1 DVD Player (Apex), 1 Mp3 (Zen Xtra), and a cellular phone (Nokia).

Each sentence is tagged in data set with a feature and sentiment orientation of the sentence with respect to the feature. The review documents were cleaned to remove tags. After that NLP and preprocessing is done. Our system is then applied to perform sentiment analysis through summarization as explained in methodology. For each product features are extracted and system is evaluated for the opinions. Results are dependent on the features extracted.

Table 1 gives precision, recall, f-measure and accuracy results of the proposed system. These results of sentiment analysis are highly dependent on the feature extraction method used. The average precision, recall, f-measure and accuracy are also mentioned in the table.

In our proposed system we have given user an alternative to find opinions according to his own preferences as shown in fig.2 i.e. if user has any preferences overall sentiment will be provided to the user on the basis his priority. Our system has a good accuracy in predicting the sentiment analysis. Hence our method of finding feature based sentiment effectively.

Table 1 Results

Product	Precision	Recall	F-Measure	Accuracy
Apex	0.777778	0.875	0.823529	0.714286
Canon	0.777778	0.777778	0.777778	0.636364
Nikon	0.767442	0.868421	0.814815	0.711538
Nokia	0.78125	0.892857	0.833333	0.75
Zen Xtra	0.756757	0.965517	0.848485	0.756098
Average	0.772201	0.875915	0.819588	0.713657

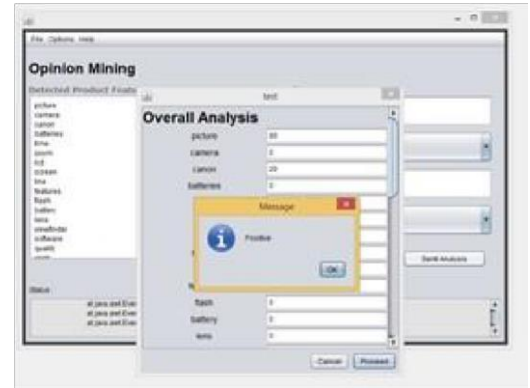


Figure 2 Result for overall opinion mining

V. CONCLUSION AND FUTURE WORK

With the development of internet, a new platform has been introduced for people to express their attitudes, opinions and feelings. Such an explosion of views on internet have demanded a way to extract and analyze this treasure of information; hence Sentiment analysis has become one important field of study. Our experimental results indicate that the proposed techniques are very promising in performing their jobs. In this paper we proposed a feature extraction based novel approach for sentiment analysis. The objective is to provide feature/subtopic based summary with its opinion for multiple documents. Our proposed model considers the user opinions to calculate the overall opinion about the subject. Our experimental results demonstrated that the proposed approach is effective and promising. In our future work, we plan to improve our technique by finding the objective sentences explicitly that indicate towards an implicit opinion. Beside this we will work on some machine learning techniques for opinion orientation identification.

REFERENCES

- [1] Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In LREC (Vol. 10, pp. 2200-2204).
- [2] Feczko, Matthew, Andrew Schaye, M. Marcus, and A. Nenkova. "SentiSummary: Sentiment Summarization for User Product Reviews." In proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 265-271.
- [3] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168-177. ACM, 2004.
- [4] Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL 2003, pp. 252-259.
- [5] Liu, Bing. "Handbook of natural language processing." Sentiment Analysis and Subjectivity, (2010): 627-667.
- [6] Mukherjee, Subhabrata, and Pushpak Bhattacharyya. "Feature specific sentiment analysis for product reviews." Computational Linguistics and Intelligent Text Processing. Springer Berlin Heidelberg, 2012. 475-487.
- [7] Mukherjee, Subhabrata, and Pushpak Bhattacharyya. "Wikisent: weakly supervised sentiment analysis through extractive summarization with wikipedia." Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2012. 774-793.
- [8] Pang, Bo, and Lillian Lee. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." In Proceedings of the 42nd annual meeting on Association for Computational Linguistics, p. 271. Association for Computational Linguistics, 2004.
- [9] Popescu, Ana-Maria, and Oren Etzioni. "Extracting product features and opinions from reviews." Natural language processing and text mining. Springer London, 2007. 9-28.
- [10] Princeton University "About WordNet." WordNet. Princeton University. 2010.
- [11] Ranade, Sarvesh, Jayant Gupta, Vasudeva Varma, and Radhika Mamidi. "Online debate summarization using topic directed sentiment analysis." In Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining, p. 7. ACM, 2013.
- [12] Somprasertsri, Gamgarn, and Pattarachai Lalitrojwong. "Mining FeatureOpinion in Online Customer Reviews for Opinion Summarization." J. UCS 16.6 (2010): 938-955.
- [13] Virmani, Deepali, Vikrant Malhotra, and Ridhi Tyagi. "Aspect Based Sentiment Analysis to Extract Meticulous Opinion Value." arXiv preprint arXiv: 1405.7519(2014).
- [14] Wang, Hanshi, et al. "Feature-based Sentiment Analysis Approach for Product Reviews." Journal of Software 9.2 (2014): 274-279.
- [15] Zhang, Kunpeng, Ramanathan Narayanan, and Alok Choudhary. "Voice of the customers: mining online customer reviews for product featurebased ranking." Proceedings of the 3rd conference on online social networks. USENIX Association, 2010.
- [16] Zhai, Yongyong, et al. "Extracting Opinion Features in Sentiment Patterns." Information Networking and Automation (ICINA), 2010 International Conference on. Vol. 1. IEEE, 2010.