

Customer Sentiment Classification and Rating System based on Products Reviews

Monika¹ Sandeep Garg²

¹Research Scholar ²Assistant Professor

^{1,2}Department of Computer Science & Engineering

^{1,2}RPIIT, Bastara, Karnal, India

Abstract— The advent of social media and ecommerce has brought the era of a new age business and its customer base is growing exponentially every year. Social media has revolutionized the new-age customer's decision making through the myriads of sources available to them like online feedback or reviews, forum discussions, blogs and Twitter on the web. Social media has revolutionized the new-age customer's decision making through the myriads of sources available to them like online feedback or reviews, forum discussions, blogs and Twitter on the web. This paper focuses on extracting the features from product reviews taken from amazon.com orebay.in sites given by reviewers to state their opinions. This is done at aspect level of analysis using ontology. Then it determines whether they are positive or negative thereby giving a scaling system to identify the effectiveness of a product. The scaling system can be in the form of a star rating system. Output of such analysis is then summarized. The results generated by proposed methodology will be compared to the base research by Arindam Chaudhury et.al in [7].

Key words: Natural Language Processing, Ontology, Text Mining, Sentiment Analysis, Classification Algorithm

I. INTRODUCTION

In this era of internet one is surrounded by digital world with lot of information on the web. Now getting the opinions or sentiments about a topic such as reviews on software, product, films, music, and books etc. are quite easy. There is a large volume of opinion texts available about these different entities on different related sites which is a clear evidence of audience' sentiments on it.

Every aware customer, now before deciding about a new product, movie or concept, has the liberty to review the feedbacks given by previous customers. This is certainly a big boost for the authentic products saleability in emarket. These feedbacks are also important from perspective of sellers to reconsider their product and strategies. Now, the problem of automating this process of sentiment gathering is that the review text length and the number of reviews is so large in some cases that the process of distillation of knowledge a very difficult task.

A. Sentiment Classification

There are three types of opinion mining approaches.

1) Feature level or Phrase level

This level addresses the features of products, classifies them and then the reviews for those features are taken separately.

2) Sentence level

This level focuses on the comments or reviews about the product. The customer gets to know about the different feedbacks from previous customers. Here, the emphasis is on the difference between the subjective and objective information. The subjective information represents the

opinion, may be negative or positive and the objective information is the fact.

3) Document Level

Here, the complete document under review is about a single product and by a single customer. Hence, the sentiment mining from these reviews is not as fruitful because the document represents views of a single person which can easily be a biased view.

B. Feature Extraction

While trying to extract the particular feature from a review or comment, the first thing is how to represent the review. [6] One way to do that is the conventional VSM (Vector space model). The VSM does the job in a quantitative way by representing the documents with vectors. The review text or comment is first tokenized and then added to a vector. The vector is filtered for the stop words. Then, each word in the vector is converted to its corresponding word stem. The effectiveness of these words is evaluated by using quantitative methods and their statistical information including the frequency of each word stem, or another similar measure, in the comment as the corresponding element in the vector used to represent that text. The following steps are performed for text classification.

1) Parsing the documents and case-folding

This step removes all abbreviations and non-alpha characters from the comment. Case-folding means converting all the characters in a text into the same case. In this research, we use Stanford parser to parse our sentences to disintegrate them into different parts of grammar. Now we can exclude the unwanted words

2) Removing Stopwords

There are words in English which are used to provide structure to the language like conjunctions, articles, pronouns and prepositions. Such words which occur very frequently and carry no useful information about the content are called stopwords. So, remove such words from the review comment.

3) Stemming

Stemming is the process for reducing derived words to their stem or root. Porter stemmer utilizes suffix stripping. Porter's stemming algorithm steps:

- Remove the plurals and suffixes like -ed or -ing from each token.
- Convert each existence of y to i when another vowel in stem.
- The double suffixes are mapped to single ones: -ization, -ational etc.
- The suffixes like -full, -ness etc. are either removed or properly dealt with.
- Remove the suffixes like -ant, -ence, etc.
- Gets rid of a final -e.

4) Term Weighting

In this step, weight is assigned to a word based on number of times it occurs in the comment. This method is called term frequency and inverse term frequency which is a traditional method to assign a weight to the words.

The explicit feature can be extracted as:

- Depending on frequency of nouns and noun phrases
- Based on the relations between Opinion and Target.

In [1], Nie and Liu have proposed anelite method that gives a new domain ontology called Fuzzy Domain Sentiment Ontology Tree (FDSOT), which integrates domain sentiment knowledge into the analysis approach to perform feature-level opinion mining. They achieve significant accuracy in sentiment classification by utilizing the prior sentiment knowledge of our ontology. Our approach here is driven by the same method.

II. RELATED WORK

Similar clustering job on web documents was done by Roul et.al, [2] where they propose a Tf-Idf based Apriori for clustering the web documents. They then rank the documents in each cluster using Tf-Idf and similarity factor of documents based on the user query much like the way we are going to rate the products on reviews gained. Maryam et.al[3] has illustrated two different approaches for feature extraction. Rule-based algorithm and HAC algorithm. Feature ranking will be done using MAX opinion score algorithm and opinion score obtained from SentiWordNet. Support Vector Machine (SVM) which is a supervised learning technique is used in feedback analysis system which accepts the responses given by students as input preprocess it and lastly applies term weighting algorithm by Sheetal Pereira et.al[4]. After applying term weighting algorithm it displays analysis to the particular faculty.

Tian Xia et. al [6] in 2011, proposed Vector Space Model. This model represents a document by a vector of terms extracted from each document along with their associated weights representing the relevance of each term in the document and in the entire collection of documents. Chaudhary et. al in [7] proposed and developed a hierarchical bidirectional recurrent neural network (HBRNN) in order to characterize sentiment specific aspects in review data available at DBS Text Mining Challenge. HBRNN is developed by extending RNN and BRNN so that accuracy and efficiency are improved. HBRNN predicted aspect sentiments vector at review level. HBRNN is optimized by fine tuning different network parameters and compared with methods like long short term memory (LSTM) and bidirectional LSTM (BLSTM). The methods are evaluated with highly skewed data. All models are evaluated using precision, recall and F1 scores. The results on experimental dataset indicate superiority of HBRNN over other techniques.

Ravi et. al in [8] presented a rigorous survey on sentiment analysis, which portrayed views presented by over one hundred articles published in the last decade regarding necessary tasks, approaches, and applications of sentiment analysis. Several sub-tasks needed to be performed for sentiment analysis which in turn could be accomplished using various approaches and techniques. This survey covering published literature during 2002–2015, was organized on the basis of sub-tasks to be performed, machine learning and natural language processing techniques used and applications

of sentiment analysis. The paper also presented open issues and along with a summary table of a hundred and sixty-one articles. Deepshikha et. al in [9], focussed on extracting the features from bank reviews taken from mouthshut.com and myBankTracker.com sites given by reviewers to state their opinions. This was done at aspect level of analysis using ontology. Then it determined whether they are positive or negative. Output of such analysis was then summarized. According to author, the average human reader had difficulty identifying relevant sites and extracting and summarizing the opinions in them. Automated sentiment analysis systems were thus needed. The research used a combination approach of domain ontology and Stanford dependency relation which intended to enhance the sentiment classification. By using this approach one could view the strength or the weakness of the features of a particular bank in more detail.

III. BASE RESEARCH

Arindam Chaudhuri et. al in [7] held a supervised approach that built on techniques from information retrieval. The algorithm populated an inverted index with pseudo-documents that encoded dependency parse relationships extracted from the sentences in the training set. Each record stored in the index was annotated with the polarity and domain of the sentence it represents. When the polarity or domain of a new sentence had to be computed, the new sentence was converted to a query that was used to retrieve the most similar sentences from the training set. The retrieved instances were scored for relevance to the query. The most relevant training instant was used to assign a polarity and domain label to the new sentence. While the results on well-formed sentences were encouraging, the performance obtained on short texts like tweets demonstrated that more work was needed in this area.

IV. PROPOSED WORK

A. Sentiment Analysis for Product rating through Ontology. (SAPRO)

SAPRO is aimed at classification from multiple reviews on various products from any online source (amazon.com from 28-Feb-2014 to 15-Jul-2014 in this case). The data received from Amazon is in JSON format and needs to be parsed by our java application for implementing the proposed methodology. The JSON review format is as below:

```
{"reviewerID": "A2IBPI20UZIR0U", "asin": "1384719342", "reviewerName": "cassandra tu \"Yeah, well, that's just like, u...\", \"helpful\": [0, 0], \"reviewText\": \"Not much to write about here, but it does exactly what it's supposed to. filters out the pop sounds. now my recordings are much more crisp. it is one of the lowest prices pop filters on amazon so might as well buy it, they honestly work the same despite their pricing.\", \"overall\": 5.0, \"summary\": \"good\", \"unixReviewTime\": 1393545600, \"reviewTime\": \"02 28, 2014\"}
```

The process flow diagram of our work is shown in the figure below.

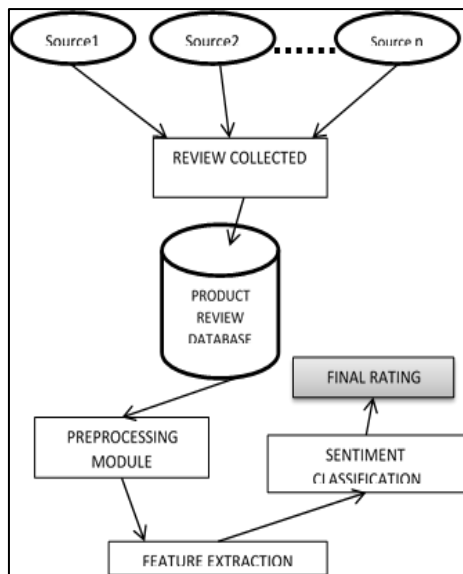


Fig. 1: Demonstration of SAPRO process

The objective of the work is to perform sentiment rating by extracting the features from product reviews given by reviewers to state their opinions. For each review:

- 1) Select a sentence from the review.
- 2) Parse the sentence using Stanford Parser.
- 3) Find features (target words) using ontology within that sentence and filter out the unwanted words.
- 4) If the feature is found then using type dependencies extract the sentiment carrying words.
- 5) For extracting the sentiment carrying word use sentiment extraction algorithm.
- 6) Check Sword List and domain sword list to find sentiment carrying word.
- 7) Extract polarity using SentiWordNet3.
- 8) Assign net polarity to target sentiment word.
- 9) Classify and aggregate result.

V. CONCLUSION, LIMITATIONS AND FUTURE SCOPE

SAPRO considers the end users' perspective while addressing the Sentiment Analysis problem. Since the volume of reviews or comments is normally so large that it is practically not feasible to go through all the reviews so as to get useful information. The research uses a combination approach of domain ontology and Stanford dependency relation which intends to enhance the sentiment classification. By using this approach one can view the strength or the weakness of the features of a particular product in more detail.

We see in this research several directions for future research, most of them based on our work's limitations. First, our work relies heavily on the amazon reviews sample collected for a specific small period of time. Second, the reviews were for a single category of products. So this makes, our work more suitable for this particular category of products. The future methodology should be such that it applies to all categories and it is tested across different entity types like banks, hotels, products, and mobile apps. So the dependency on a category ontology has to be reduced.

REFERENCES

- [1] Nie, Liu, Wang, Song in "The Opinion Mining Based on Fuzzy Domain Sentiment Ontology Tree for Product Reviews" Journal of SW Vol. 8 No.11, 2013.
- [2] Roul, Devanand, Sahay in "Web Document Clustering and Ranking using Tf-Idf based Apriori Approach", International Conference on Advances in Computer Engineering & Applications, 2014.
- [3] Maryam, SeemaKoulkur [3] in "Feature ranking in sentiment analysis", International Journal of Computer Applications, 2014.
- [4] Sheetal Pereira, Uday Joshi in "Implementation of SVM technique in feedback analysis system", International journal of computer applications, 2014.
- [5] Chandhana Surbhi in "Natural language processing future", International conference on optical imaging sensor, 2013.
- [6] Tian Xia in "Improved VSM text classification by title vector based document representation method" ICCSE Pg 210-213, 2011.
- [7] Arindam Chaudhuri and Soumya K. Ghosh, "Sentiment Analysis of Customer Reviews Using Robust Hierarchical Bidirectional Recurrent Neural Network", Advances in Intelligent Systems and Computing 464, Springer 2016.
- [8] Kumar Ravi and VadlaMani Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications", Knowledge-Based Systems 89-91, Elsevier 2015.
- [9] Deepshikha Chaturvedi and Shalu Chopra, "Customers Sentiment on Banks", International Journal of Computer Applications (0975 – 8887) Volume 98– No.13, July 2014.
- [10] Mauro Dragoni and Fondazione Bruno Kessler, "SHELLFBK: An Information Retrieval-based System For Multi-Domain Sentiment Analysis", Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 502–509, IEEE 2015.