IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 1, Ver. IV (Jan – Feb. 2015), PP 13-17
www.iosrjournals.org

Feature Based Semantic Polarity Analysis Through Ontology

Dr.M.S.Anbarasi[1], Sarathkumar.B[2], Bomanna jaswanth[2],
Shahul Nikamuthul asiqine.H[2]

1(Department of Information Technology , Pondicherry Engineering College, Pondicherry ,India.)
2(Department of Information Technology , Pondicherry Engineering College, Pondicherry ,India.)

Abstract: *Opinion mining, a trending research area where customers feels that opinions of others are always important for making decisions while purchasing the products. Here the problem is to collect those opinions and preprocess them. Thus, the customer reviews have been collected and after that Ontology is constructed to structure the information available in unstructured text reviews and to exploit semantic relation between them. Then we extract the features of the product and their opinions further polarities (positive, negative or neutral) of different features of certain digital products are identified. Opinions of different features of productivity customer are classified and summarized by an enhanced opinion mining technique. The performance of the system is evaluated by metrics such as precisions, recall and F-measure. This information provided to users will be more helpful to make decisions before buying a product.*
Keywords: *ontology,opinion mining,feature extraction,pos tagging,polarity identification*

I. Introduction

Opinion mining is a highly trending research field that comprises natural language processing, computational linguistics and text analysis techniques with the aim of extracting various kinds of added-value and informational elements from users' opinions. The task of organizing user feedback and retrieving information is known as opinion mining. Customers' satisfaction is the key to the manufactures for the success of their business, the opinions (or) feedbacks which are given by the customers are necessary to know about their products in the market.

Approaches of opinion mining where user reviews has been collected and classified into two approaches: text classification and information extraction approaches. In the former, researchers have been exploring techniques for polarity identification for entire document as either positive vs. negative [Bing Liu 2008; Pang and Lee, 2004; L.Zhuang 2006, etc.]. The latter, on the other hand, describes about the products or reviews at feature level, which helps both customer and developer to know in detail about their strength and weakness.

Time is running from pin to pin, people preferences to the online shopping is comparatively more than generic method of in store shopping. Supporting to this many review sites and online websites is helping to the customers with consisting of the feedbacks and reviews.

Many researchers are being focusing on polarity identification of user reviews from years even though the accuracy is not achieved. Every document consists of both positive and negative feedbacks ,according to many opinion mining approaches the document which consisting most positive preferable sentences is considered as positive feedback ,in the same way for negative feedback. The former approach (Zhou, L., & Chaovalit, 2008; Neviarouskaya, Prendinger, & Ishizuka, 2009; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011) is based on opinion words and extracting information about the positive ,negative and neutral documents.

In fact, opinions classified at the document or sentence level determines the rating of the product or whether the product is good, bad or neutral. A product which is identified by these mining techniques as good does not mean that the user has positive opinions on all aspects or features. Likewise, a bad product does not mean that the user dislikes everything about the object. In a document (e.g., a product review), the user typically writes about both the positive and negative aspects of the object, although the general sentiment toward that object may be positive or negative (Ahmad & Doja, 2012). To obtain such detailed aspects, it is necessary to perform feature-based opinion mining in an attempt to identify the features in the opinion and classify the sentiments of the opinion for each of these features (Feldman, 2013).

An ontology can be defined as an ''explicit, machine-readable specification of a shared conceptualization'' (Studer, Benjamins, & Fensel, 1998). Many authors utilizing ontologies for defining a domain so that it serves many purpose. As well as the relations among these terms and are now applied in various fields, like agent and knowledge management systems and e-commerce platforms (Gómez-Pérez & Corcho, 2002).

. In this paper we propose an innovative opinion mining methodology that takes advantage of new Semantic Web-guided solutions and ontology to enhance the results obtained with traditional natural language processing techniques and sentiment analysis processes. In this paper first we will collect the customer review data, those data will be processed. Further we will extract the features through pos tagging and frequency calculation. Finally the main goal of all these works is to taking into account about each aspect of the products, classifying each feature are important and giving polarities and semantic scores to it .

II. Headings

1. Data Collection

The basic step we should perform is collection of data.To determine the polarity of the sentences, based on aspects, large numbers of reviews are collected from the Web. There are lots of websites on the Internet where the large numbers of customer reviews are available. We used One of the popular website in social network is Amazon (www.amazon.com) to collect the reviews from web.

2. Data preprocessing

Data processing should be done to those reviews which are having been collected from the website. But in those reviews some unwanted information is available. We should remove those information like URL links (e.g. http://semantic.com), Twitter user names (e.g.@lifestyle – with symbol @ indicating a user name),Twitter special words (such as "RT"6) and emoticons. But those information with "RT" should be collected, RT means retweet which is an useful information, an aspect that more retweets means more people are discussing about that aspect. User reviews can be a mix of both subjective and objective sentence. Subjective sentences are mostly based on facts so it should not be considered for the further process, and only objective sentences are to taken for the remaining process. Then pos tagging module is done for the extraction of the features which is the next step of the process. A tool of the name Stanford POS tagger tool (http://nlp.stanford.edu/software/tagger.shtml) is utilized for tagging purpose. It is a piece of software which is used to read those human reviews and assigns parts of speech for each word of a sentence. So, that feature extraction is possible and then Term frequency, dispersion, deviation is to be calculated. These terms are useful in determining the polarities of the features.

III. Frequency calculation

The TF/IDF (TF-Term Frequency, IDF-Inverse Document Frequency) algorithm is the best algorithm which is well known for weighting scheme for terms in the information retrieval. The TF/IDF score (or, weight) which depends upon the occurrences and it is also considers the length of the document and the query terms, it is nothing but a term increases with the number of occurrences in a document (**TF** component). The score will also increase with the rarity of the term across the entire collection (**IDF** component) and

3.1Document Frequency (DF)

The ratio of the document that contains the term t to the total number of documents that contains term t

3.2Inverse Document Frequency (IDF)

The IDF component acts to discriminate between informative and non-informative query terms. Those terms that have a high IDF are considered more informative, because they rarely occur in the collection. On the other hand, terms that have a low IDF are considered uninformative, since they occur in many documents. As the number of documents in a collection increases, IDF becomes increasingly important in order to discriminate between those documents that contain non-informative query terms and those that contain high informative query terms.

This value is obtained by the taking the total number of documents in the collection and then differentiating then dividing it with the former gives the value of idf

$$\mathbf{idf = log_{10}(} \mid D \mid / df_t)$$

3.3 Term Frequency (TF)

The TF component, which is often normalized in some way with respect to the document length, is used to discriminate between documents that contain a query term several times and those that contain the term many times. This makes the assumption that documents that contain more mentions of a given query term and are more "about" the given term and therefore are more likely to be relevant to the query. Note that this can be a poor assumption particularly when collection size increases and it becomes nosier .Repeated words might be because of irritation towards the product The TF component becomes more important when documents grow longer that the length of the document extended.

This measurement is a simple one:

$$1 + \log_{10}(tf_{t,d})$$

IV. Ontology Creation

Creating Ontology is long and time consuming process which is also requires more human effort for this creation. To reduce the time and effort for this length process of ontology, the ontology Learning has been developed which is useful to know the concepts and relations that exists in particular domain. This will be achieved with the use of certain tools like natural language (NLP) process tools which includes a morphologic analyzer, a part-of-speech (POS) tagger, and a chunk parser, unstructured customer statements on a specific product can be translated, analyzed and modeled automatically.

Ontology applications include natural language generation, intelligent information integration, semantic-based access to the Internet and extracting information from texts. However, the most important contribution of ontologies is the key role they play in the development of the Semantic Web. The Semantic Web is an extension of the current Web, where information is given a well-defined meaning, encouraging cooperation among human users and computers (Berners-Lee, Hendler, & Lassila, 2001). Ontologies serve as the primary means of knowledge representation in the Semantic Web. Although various ontology languages have emerged, the currently dominant standards are RDF/S (Resource Description Framework Schema) and OWL (Web Ontology Language). Additional points of motivation for preferring the use of ontologies in an application include: (a) analyzing domain knowledge and separating the latter from operational knowledge, (b) Enabling the reuse of domain knowledge, (c) Making domain assumptions explicit, and (d) Sharing a common understanding of the information structure among people and/or software agents.

Ontology construction makes us to know the various levels of abstraction through the customer statements about the characteristics of the product. Construction of ontology can be done both on manually and on machine learning. It can be done with the help tools which include editors, checkers to extract relation among the entities of the product.

The first step in ontology creation is classifies the attributes it is nothing but it extracts the features of the product. Then it will proceeds with the sub groups (or) other attributes to existing attributes as features, it forms an ontology structure in the hierarchical form with the different aspects of the products. There is also chance of existing similar text repeating in a review those are not being considered in the classification. In the available text document it also makes text into three categories of structured text, unstructured text and semi structured text. Each word in the review is verified and classified. After ontology creation then it provides semantic score to each aspect with the various enhanced opinion mining techniques. It also makes the domain to reuse .

V. Opinion mining

After analyzing the corpus by means of NLP techniques, extracting the relevant features and identifying the features polarity, the framework proposed here provides an innovative opinion mining mechanism. The opinion mining module described in this section is based on vector analysis and enables an effective feature sentiment classification. In the opinion mining process it generates the opinions about the products based on the review analysis and polarities.

After extracting the features of the product through the ontology then we should semantic score and polarities i.e. either it can be positive (or) negative (or) neutral. Those semantic scores are assigned scores with the help of n-grams technique.

There are various tools available for providing polarities to the features of products, among them one of the best tools Sentiword.Net is used in this paper. The tool provides polarities to the words of the reviews. Positive, Negative and neutral score are being calculated with the average words that are present in the review, if more positive words it is considered as positive score, consists of negative words is considered as Negative score and neutral score is defined with their sum less than one.

N Gram's technique of opinion mining is utilized for to identify that how many words nearly and prefers to the features of the products. It consists of N_Gram After, N_Gram Before, N_Gram Around which gives the detail which specifies the features of the products. So, that accurate semantic score as being calculated semantic score and polarities. The words which indicates positive is considered as supporting to feature and negative words which are blaming considers the negative words.

We will provide the score on a scale -5 to 5. If the calculated score from range below zero it indicates Negative score and polarity will be negative. If the score is zero then it is neutral polarity, and the score is more than zero it is Positive score. The customers are being satisfied with the semantic score which is provided with this paper.

Since the polarity is calculated for each features separately the techniques of N grams is applied to the collected words of each features in the structure that is created by the ontology .To provide a sentiment score for each words ,we use sentiwordnet . The feature polarity is calculated using SentiWordNet 3.0 (Baccianella & Esuli, 2010) (SWN). SENTIWORDNET is the result of the automatic annotation of all the synsets of WORDNET according to the notions of "positivity", "negativity", and "neutrality". Each synset s is associated to three numerical scores Pos(s), Neg(s), and Obj(s) which indicate how positive, negative, and "objective" (i.e., neutral) the terms contained in the synset .Different senses of the same term may thus have different opinion-related properties.

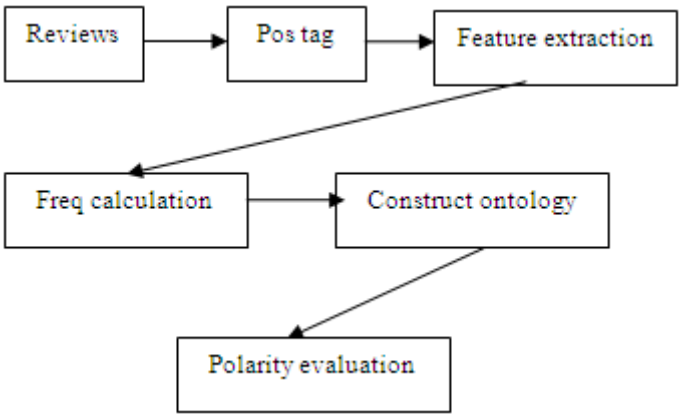
INTENDATIONS and EQUATIONS

Algorithm

Input: User reviews
Output : Opinion Results and ontology

Process :
Tagging process → POS TAGGING
Extraction of Features f1,f2,f3……fn;
Word count=Sum (Number of words in each document);
For each word w
Term Frequency=Number of occurrences of each word w/Total number of words in the document.
End loop
For each word w
Opinion Strength Calculation = Sign(SET(p))(CS(p)/Set(p))
Inverse Document Frequency=ln(R/RCA)*gamma
Gamma=1/ln(R)
Sentence=OpinionStrength*InverseDocumentFrequency*
Degree
End loop
For each word w
Opinion Score=Averagepolarity*populationweight*weight
End loop
Output ontology file.

VI. Figures and Tables



Reference

[1] EfstratiosKontopoulos,ChristosBerberidis,Theologos Dergiades, Nick Bassiliades, " Ontology-based sentiment analysis of twitter posts,"Expert Systems with Applications 40, Elsevier (2013) 4065–4074.
[2] Abd. Samad Hasan Basaria,Burairah Hussina,I. Gede Pramudya Anantaa, Junta Zeniarjab, " Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization", Elsevier, Procedia Engineering 53 (2013) 453 – 462.
[3] Li Chen, Luole Qi, Feng Wang, " Comparison of feature-level learning methods for mining online consumer reviews", Elsevier, Expert Systems with Applications 39 (2012) 9588–9601.

[4] Baldoni, M., Baroglio, C., Patti, V., & Rena, P. (2012). From tags to emotions Ontology-driven sentiment analysis in the social semantic web. Intelligenza Artificiale, 6(1), 41–54.
[5] De Kok, D., & Brouwer, H. (2012). Natural language processing for the working programmer. E-book available under the creative commons attribution 3.0 License (CC-BY). 2011,access: November 2012.
[6] Moreo, A., Romero, M., Castro, J. L., & Zurita, J. M. (2012). Lexicon-based commentsoriented news sentiment analyzer system. Expert Systems with Applications, 39(10), 9166–9180.
[7] Zhang, L., Liu, B., Lim, S., & O'Brien-Strain, E. (2010). Extracting and ranking product features in opinion documents. In Proceedings of the 23rd international conference on computational linguistics: Posters (pp. 1462–1470). Association for Computational Linguistics .
[8] Ning, L., Guanyu, L., & Li, S. (2010). Using formal concept analysis for maritime ontology building. In Proceedings of the 2010 international forum on information technology and applications (IFITA '10) (pp. 159–162). New York, USA: Springer.Vol.2