

Aspect based Opinion Mining from Restaurant Reviews

Chinsha T C
PG Scholar
Dept. of Computer Science
Govt. Engineering College
Thrissur, India

Shibily Joseph
Asst. Professor
Dept. of Computer Science
Govt. Engineering College
Thrissur, India

ABSTRACT

Opinion mining or sentiment analysis analyses the text written in a natural language about a topic and classify them as positive negative or neutral based on the human's sentiments, emotion, opinions expressed in it. Nowadays user reviews and comments on travels on the web are increasing day by day. These comments are useful for other users to make a decision in travel planning. The manual analysis of such huge number of reviews is practically impossible. To solve this problem an automated approach of a machine to mine the overall sentiment or opinion polarity from the reviews is needed. Opinion mining can be done at three different levels, which are document level, sentence level and aspect level. Most of the previous work is in the field of document or sentence level sentiment analysis. This paper focus on the aspect based opinion mining of restaurant reviews, i.e. given a set of reviews of a restaurant we get a sentiment profile of its important features automatically. A different approach proposed for opinion mining which uses SentiWordNet, two word phrases and linguistic rules together for opinion mining.

General Terms

Natural Language Processing, Sentiment Analysis.

Keywords

Opinion Mining, Sentiment Analysis, Aspect Based Opinion Mining, Tourism domain, Aspect Extraction, SentiWordNet.

1. INTRODUCTION

The recent proliferation of web2.0 applications, users now have many opportunities to express opinions and can share experiences on the internet [5]. Due to the explosive growth of social media (e.g., reviews, forum discussions, blogs, Twitter and comments) on the Web, individuals and organizations are increasingly using the content in these media for decision making. This trend makes attention of organizations and researches around the world towards opinion mining area. This provides a strong motivation for research and also offers many challenging research problems. This is a hard problem to be solved because natural language is highly unstructured in nature. The interpretation of the meaning of a particular sentence by a machine is very difficult. But the usefulness of the opinion is increasing day by day. Machines must be made to interpret and understand human emotions and feelings. Sentiment analysis and opinion mining are approaches to implement the same.

The problem of opinion mining can be solved to a satisfactory level by manual training. But an efficient fully automated system for opinion mining which needs no manual intervention has not been introduced yet. This is mainly because of the challenges in this field.

Nowadays user reviews and comments on travels on the web are an important information source in travel planning. Such types of reviews are restaurant reviews, which help people in

travel planning. Due to the lack of work in this tourism domain, restaurant reviews are selected as domain for this project. This paper aims to implement an aspect based opinion miner for tourism domain, which automatically finds important features or aspects (e.g., food, service of restaurants etc.) and its opinion (i.e. opinion on food, service like aspects are how much positive, negative and neutral). Such a way it will create a sentiment profile of each restaurant, which can be further used to compare and select restaurants at a particular location by a traveller. This paper propose a different approach for opinion mining which makes use of adjectives, adverbs, linguistic rules and Sentiwordnet [12]. The paper also presents the proposed system architecture and the description about the sequence of steps for implementation.

2. RELATED WORKS

Opinion mining comprises the area of natural language processing, computational linguistics and data mining. Opinion mining can be done at various levels, which is document level, sentence level and aspect level [1]. In document level, overall opinion polarity of the document is calculated and classified as positive or negative. In sentence level, each sentence in the document is analysed and determine the opinion expressed in a sentence as positive, negative, or neutral. In opinion mining, the term aspect means important features of products rated by customers (For example, in the case of restaurant food, service, cleanliness etc.). The product and restaurant reviews are a mixture of positive and negative opinion about different aspects. It needs more fine-grained analysis of reviews to mine these mixed opinions, aspect level perform this task. Hence aspect based opinion mining is preferred in this work.

The core tasks in aspect based opinion mining is aspect identification, aspect based opinion word identification and its orientation detection [1]. For example, consider a review of a restaurant, "The environment is nice but food is bad". First step is to identify the aspect, which is environment and food, then find aspect related opinion word, which are nice and bad. Then detect its orientation, i.e. whether that opinion word expresses positive or negative opinion. By analysing the above example we get that environment has positive opinion and the food has positive opinion. The proposed method finds positive and negative value of each opinion with the help of SentiWordNet [12].

There are many research works in aspect identification task [10] [4] and opinion mining. Some of them are given below. Different approaches used by them are supervised learning, unsupervised learning, lexicon based and natural language techniques.

In [2] a supervised machine learning approach is used for sentiment analysis. It is the first paper, used this approach. Unigram and bigram used as features for classification, which classify movie reviews as positive or negative. Learning methods used are Naive Bayes, maximum entropy and

support vector machines. Conclusion made was, machine learning techniques are quite good in comparison to the human baselines for opinion based text classification and relatively svm performs better. But supervised learning is sensitive to the quantity and quality of the training data and may fail when training data are insufficient. Hence unsupervised learning method proposed by researchers, which is preferred in this work.

In [3] a simple unsupervised learning algorithm is used for classifying a review as recommended (thumbs up) or not recommended (thumbs down). Classifications are done using some fixed syntactic patterns, which are composed on part-of-speech (POS) tags. Semantic orientation of the words was calculated by using the mutual information between the given phrase and the word excellent minus the mutual information between the given phrase and the word poor. If the average semantic orientation of its phrase is positive then review was classified as recommended otherwise not recommended. The algorithm was evaluated on 410 reviews from Epinions and achieved the average accuracy of 74%.

In [4] natural language processing and Data mining techniques used for mining and summarization process of customer reviews of a product. Aspects, aspect related opinions are identified and classified as positive or negative. Aspect based a text summary also produced. No visualization tools are used. Current work aims to use visualization tool for producing summary.

In [5] a linguistic approach is used for opinion mining. For each sentence in a review, it generates a dependency tree, and splits the sentence into clauses. Then it determines the contextual sentiment score for each clause utilizing grammatical dependencies of words and the prior sentiment scores of the words derived from SentiWordNet and domain specific lexicons. Negation is also delicately handled in this study. Automatic aspect extraction not included in this work.

In [6] aspect based sentiment analysis done using support vector machine classifier. This work combines the use of dependency parsing, co-reference resolution and SentiWordNet together for the sentiment analysis. Average accuracy of 77.98% is obtained.

In [7] a new feature based heuristic was used for aspect level sentiment classification. The authors proposed AAACV algorithm based on SentiWordNet to find the sentence level aspect score. This scheme calculates 'Adverb+Adjective' and 'Adverb+Verb' combined score using SentiWordNet. Accuracy of 78.7% obtained over movie review dataset. Current method is based on this algorithm.

In [8] authors extend the Bing Liu's aspect-based opinion mining technique to apply it to the tourism domain. The score of opinion on aspects is calculated based on some linguistic rules. They discover consumer preferences about tourism products using some statistics.

In [9] a modular software is proposed named OpinionZoom, that helps users to understand the vast amount of tourism opinions disposed all over the Web in an easy manner. Authors used the same method in [8] but it produces novel graphic summaries of opinions. They successfully implemented and tested OpinionZoom, The proposed method use SentiWordNet and linguistic rules to find the aspect score.

3. PROPOSED METHOD

The architecture of the proposed opinion mining model is given in the Figure 1. The following sub sections describe each sequent steps in detail.

3.1 Data Collection

The data collection module collects the restaurant reviews from the web. A review crawler needs to be developed for extracting the reviews from the web, which parses the HTML webpages containing the opinions and extracts the reviews of a particular restaurant. The extracted reviews should store in a database, which is used as the input to the aspect based opinion mining system. The restaurant reviews are available in the Trip Advisor [14] site.

3.2 Preprocessing

The proposed technique involves pre-processing of reviews to improve the accuracy of opinion mining process and to avoid the unnecessary processing overhead. The pre-processing steps include removal of unnecessary characters.

3.3 Aspect Extraction

Aspects are the important features rated by the reviewers. For the aspect extraction, the idea in [4] is used. An aspect may be a single word or a phrase. For example the service and quality of food is the different aspects of a restaurant. In [4] says, in most cases nouns and noun phrases are the aspects. Therefore, searching of noun and noun phrases are needed. To extract the important aspects, first find the Part-Of-Speech tag of words in each sentence of a review then identify the nouns and noun phrases. For example, tags like NN (noun), NNS (noun plural), NNP (proper noun, singular) etc. of word need to be identified. Then find its frequency, most frequent ones are chosen and infrequent ones are discarded. After group the synonym words and creates an aspect dictionary of important aspects. Sanford POS Tagger [13] can be used for tagging purpose.

3.4 Subjectivity/Objectivity Classification

All the sentences in reviews do not express an opinion. A sentence of the review is analysed only when it contains an opinion. Such subjective sentences should be identified and other (objective) sentences should be removed in order to avoid further processing overhead. Subjective sentences are identified by looking presence of aspects and opinion words. This can be done with the help of Sentiwordnet, which contains opinion words, and aspect dictionary, which contains important aspects (created earlier).

3.5 Aspect related Opinion Words Identification

This module identifies the aspect related opinion words. Reviews are taken from the database then analyses each review sentence and look up in the aspect dictionary. If any aspects are present in sentence its corresponding opinion words should be identified. The POS information of a word like adjective, adverb, noun and verb are used for extracting the opinion words in a sentence. If any aspects are present in a sentence, then extract adjective, adverb, noun and verb by searching 5-gram forwards and backwards from the aspect position [7].

3.6 Orientation Detection

Orientation of an opinion on each aspect is identified in this step, i.e., polarity scores of opinion on aspect. Proposed method uses SentiWordNet for assigning priority scores to opinion words, which is a dictionary of opinion words. In it

each synset of terms in Wordnet has given positive and negative scores. Word sense disambiguation is not considered here. In order to get contextual information of a sentence the patterns [3] is extracted such as Adjective-Noun, Adverb-Adjective, Noun-Noun, Noun-Adjective, Adverb-Verb combinations. Scores of this pattern and its combined score in a sentence is calculated based on algorithm in [7] and priority scores from SentiWordNet. The linguistic rules [1] [8] is also

used for finding contextual scores. If negation words like not, never, etc. are present the polarity should be reversed.

For example, “I am not satisfied with the service”, in the above sentence the opinion word satisfied has positive polarity, but due to the presence of negation word ‘not’ the polarity changed to negative. Such negation words should consider before calculating sentence score. But rule, Too rules in [8] also to be considered.

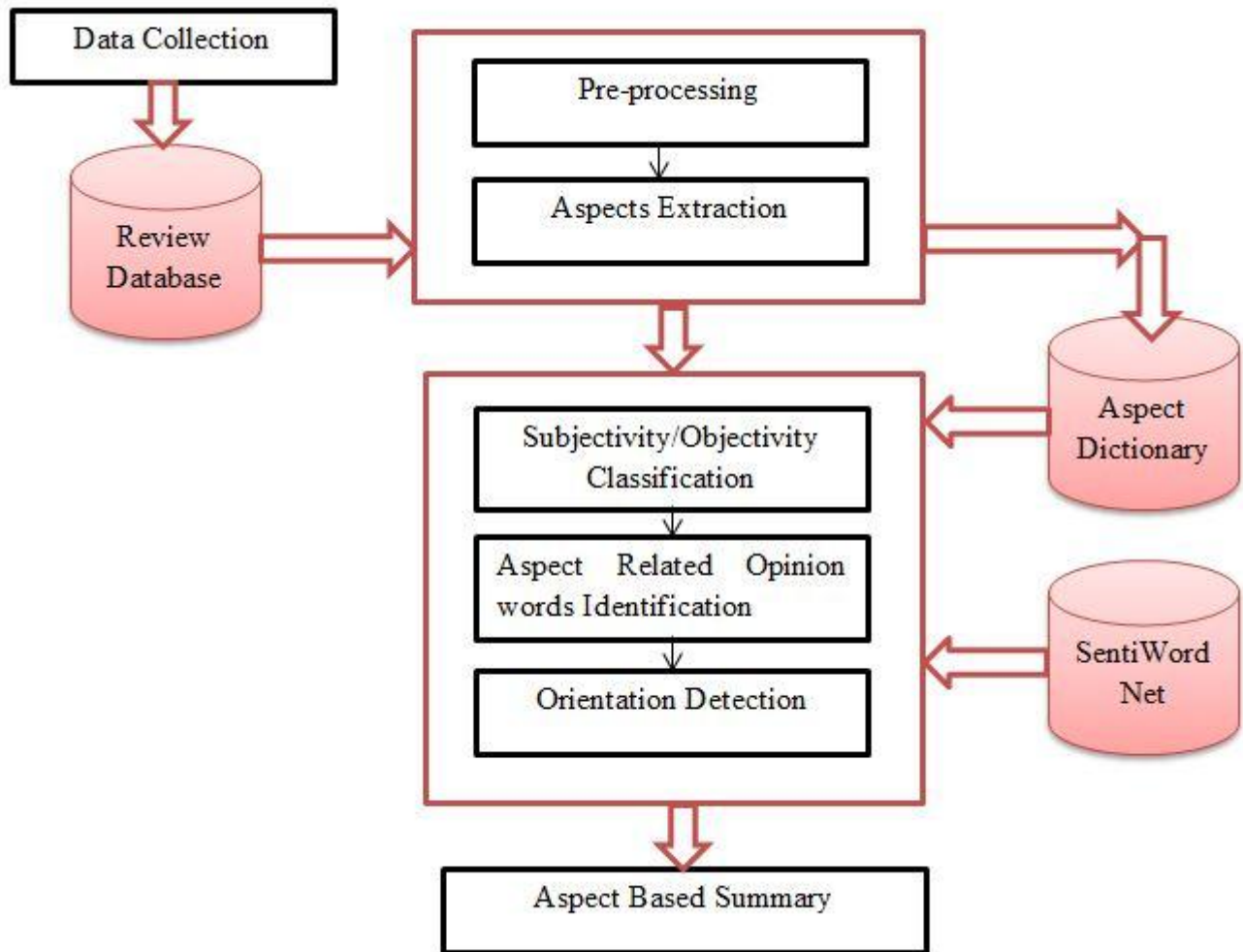


Fig 1: Proposed Opinion Mining Model

3.7 Aspect based Summary

This module aggregates the scores of each aspect in all reviews and produce an aspect based summary. Positive and negative scores of aspects are separately aggregated; hence we get an aggregate positive score and negative score of aspect. By using this result a sentiment profile of each restaurant can be created. Visualization tools can be used for this purpose, which will be very useful for customers for knowing the opinion on different features of a restaurant also get an idea that, each aspect is how much positive and how much negative. The scores of opinions on each aspect in all reviews can be aggregated using the formula given below [6].

For each aspect j of the restaurant,

$$\text{Aggregate_Positive_polarity}[j] = \sum_i \text{Positive_Pol}_{i,j}$$

$$\text{Aggregate_Negative_polarity}[j] = \sum_i \text{Negative_Pol}_{i,j}$$

In the above formula ‘ i ’ represent a review.

4. PERFORMANCE EVALUATION

The performance of the proposed system should be evaluated for understanding the effectiveness of our system. For that, firstly need a suitable test dataset. The performance of the proposed method can be evaluated by building a suitable annotated test set of restaurant reviews. Restaurant reviews are available in [14]. These reviews need to be collected by developing a crawler. Then the collected reviews annotated suitably with positive and negative polarities in sentence basis. This can be done manually. This annotated dataset can be further used for evaluation purpose. Evaluation can be done by measuring accuracy, precision and recall.

Accuracy is one of the measures, which is used for evaluating the performance of the system. Accuracy is the portion of all true predicted instances against all predicted instances, which

is the measure of how much is correct. The formula for accuracy is given below.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

In the above formula TP is the true positive instances. TN is the true negative instances. FP is the false positive instances and FN is the false negative instances.

Precision and recall are the other two measures, which are extended versions of accuracy. Precision measures exactness, i.e. it is the number of reviews in test set is correctly labelled as positive out of total number of reviews classified as positive. Recall measures completeness, i.e. it is the number of reviews in test set correctly labelled as positive out of total number of reviews that truly are positive. Corresponding formulas for precision and recall is given below.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

After calculating precision and recall, F-measure can be calculated using these values. F-measure is the harmonic mean of precision and recall, i.e. it is a combined form of precision and recall. This is a strict metric. If we get high value to the F-measure, it means the system architecture is reasonable and the proposed methods are effective. F-Measure is calculated using the formula given below.

$$\text{F-Measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. CONCLUSION

Opinion mining comprises the area of natural language processing, computational linguistics and data mining. The problem of opinion mining is a research area in which many hard problems are to be addressed. The restaurant reviews on travels on the web are important information source, which help people in travel planning. This paper aims to implement an aspect based opinion miner for tourism domain such as restaurant reviews, which automatically finds important aspects and opinions of a restaurant by analysing reviews, then create a sentiment profile of each restaurant, which can be further used to compare and select restaurants at a particular location by a traveller. This paper propose an approach, which use SentiWordNet, two word phrases and linguistic rules together for opinion orientation detection, with automatic acquisition of aspects.

In this work only explicit aspects are considered and word sense disambiguation is ignored. Next work aims to include the same and also the analysis of different type of sentences like conditional, comparative sentences in order to improve the accuracy of opinion mining.

6. REFERENCES

- [1] Bing Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies 5 (2012), no. 1, 1-167.
- [2] Lee L. Vaithyanathan Pang, B., Thumbs up? sentiment classification using machine learning techniques, In Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol. 10, Association for Computational Linguistics, 2002, pp. 78-86.
- [3] Peter D Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, In Proceedings of the 40th annual meeting on association for computational linguistics, Association for Computational Linguistics, 2002, pp. 417-424.
- [4] Minqing Hu and Bing Liu, Mining and summarizing customer reviews, Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2004, pp. 168-177.
- [5] Thet, Tun Thura, et al. Sentiment analysis of movie reviews on discussion boards using a linguistic approach. Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, ACM 2009.
- [6] Varghese R and Jayasree M, Aspect based sentiment analysis using support vector machine classifier, Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on, IEEE, 2013, pp. 1581 - 1586.
- [7] Singh, V. K., et al. Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification. Automation, Computing, Communication, Control and Compressed sensing (iMac4s) 2013 International Multi-Conference on. IEEE, 2013.
- [8] E. Marrese-Taylor, J. D. Velásquez, F. Bravo-Marquez, and Y. Matsuo, Identifying customer preferences about tourism products using an aspect-based opinion mining approach, in Knowledge-Based and Intelligent Information and Engineering Systems. Procedia Computer Science, 2013.
- [9] Marrese-Taylor, Edison, Juan D. Velásquez, and Felipe Bravo-Marquez. Opinion Zoom: A Modular Tool to Explore Tourism Opinions on the Web. Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on. Vol. 3. IEEE, 2013.
- [10] A.M. Popescu and O. Etzioni, Extracting product features and opinions from reviews, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005.
- [11] Ohana, Bruno, and Brendan Tierney. Sentiment classification of reviews using SentiWordNet. 9th. IT & T Conference. 2009.
- [12] <http://sentiwordnet.isti.cnr.it/>.
- [13] <http://nlp.stanford.edu/software/tagger.shtml/>.
- [14] www.tripadvisor.in/
- [15] Bakliwal, Akshat, et al. Towards Enhanced Opinion Classification using NLP Techniques. Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), IJCNLP. 2011.