ISSN:0975-9646

Tourview: Sentiment Based Analysis on Tourist Domain

Deepanshi Sharma, Achal Kulshreshtha, Priyanka Paygude
IT Department, Bharati Vidyapeeth University, India

***Abstract*— Due to rapid growth in tourism, tourists like to find information regarding different places. With a rapid growth in web, tourists generally share their reviews on social website. These websites have become major source of information for tourism but due to huge data and mixed emotions it becomes difficult for a tourist to make a decision regarding the travelling. For this problem we proposed the system "tourview" a tool that offers a set of summarization methods to help users digest the vast availability of opinions in an easy manner. The main objective behind this system is that tourist can easily extract subjective and useful information. This system also reduces the time required for searching and easily supports growth of tourism.**

***Keywords*— Opinion Mining; Natural Language Processing; Sentiment Analysis; Customer Review**

INTRODUCTION

Nowadays, tourists check opinions and experiences published by other travelers on different web platforms when planning their own vacations. Travelers from all around the world share their experiences on websites which have become major sources of information for tourists. When trying to analyze web opinions, tourists are often confused because of large amount of reviews and mixed sentiments. Considering this problem, we propose TourView, a tool that offers a set of summarization methods to help users digest the vast availability of opinions in an easy manner. The main objective behind this system is that tourist can easily extract subjective and useful information. This system also reduces the time required for searching and easily supports growth of tourism.

Objective of this project is to process the reviews, of tourist places, collected from a website based on sentiment analysis. And then segregate those reviews based on different categories like weather, peak season, expenses incurred, transport availability, nearby location etc. Also provide suggested places to the user based on their requirements and shows top most review of a particular place.

I. BACKGROUND

A. Opinion Mining

Opinion Mining (OM) is one of the challenging issues which have taken a lot of efforts by researcher to solve. There are a number of tasks in OM. Classification of tourist reviews into the positive, negative and neutral classes (also known as semantic orientation) is one of these tasks, that helps product businesses to easily identify semantic orientation of their products services. On the other hand, as a result of the growth of the web, people are able to express their views and opinions. In addition, the number of reviews has grown rapidly which makes it hard for user to find correct opinions of the particular place. Therefore, automated opinion discovery is needed.

One of the existing methods for finding positive and negative orientation of person reviews was called: holistic lexicon-based approach to opinion mining. This approach utilizes semantic orientation of customer reviews based on each product features. The holistic lexicon-based approach uses some linguistic rules to classify customer reviews semantically. This approach was good enough and gave reasonable results but it did not consider the strength levels of customer opinion. Moreover, classification of customer reviews into the positive, negative and neutral classes (semantic classification) helps manufacturers to understand reviews easily; however it does not help manufacturers to understand the priority of reviews in each class.

B. Sentiment Analysis

Sentiment analysis is the study of opinions, and emotions toward entities, events and their attributes. In the past few years, it has been successful in attracting a great deal of attentions from both academia and industry due to many challenging research problems and a wide range of applications. Opinions are important because whenever we need to make a decision we want to hear opinion of other people. This is true for a person as well as for organizations.

However, there was almost no computational study on opinions before the invention of web technologies because there was little opinionated text available. In the past, when a person had to make a decision, he/she was required to ask for opinions from friends and families. Whenever an organization wanted opinions of the public about its products and services, it used to conduct various surveys. However, with the explosive growth of the social media contents on the websites in the recent past, the world has been transformed. Discussion forums, blogs, and social network sites have now enabled people to post reviews of products and express their views. Now, if you want to buy a product, you are no longer limited to asking your friends and families because there are many user reviews on the websites. For a company, it may no longer need to conduct surveys or focus groups in order to gather consumer opinions about its products and those of its competitors because there is a plenty of such information publicly available. However, finding sites containing opinions and monitoring them can still be a tough task because there are large numbers of opinions. In a number of instances, these are hidden in blogs and also in long forum posts. It is difficult for a human reader to find relevant sites, extract related information, understand, summarize and organize them into usable forms. Thus, we need automated opinion discovery and summarization systems. Sentiment analysis is not a single task, but it is a problem containing many sub-problems viz-a-viz objects identification, opinion orientation classification and integration feature extraction and synonym grouping. Here, focus is towards the methods that find information of the new challenges raised by sentiment-aware applications, as compared to those that already exist in old fact-based analysis techniques.

C. Natural Language Processing

Natural language processing (NLP) is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (natural) languages.

Below are the most commonly researched tasks in NLP which are included in Stanford CoreNLP package.

- Part-of-speech tagging
- Automatic summarization
- Parsing
- Natural language understanding
- Machine translation
- Named entity recognition (NER)
- Optical character recognition (OCR)
- Natural language generation

D. Customer Review

A customer review is a review of product or service made by the customer who has purchased the product or service. Customer reviews is a customer feedback on various sites. There are also dedicated review sites, some of which use customer review as well as or instead of professional reviews. Other users can also grade reviews for their usefulness or accuracy.

II. RELATED WORK

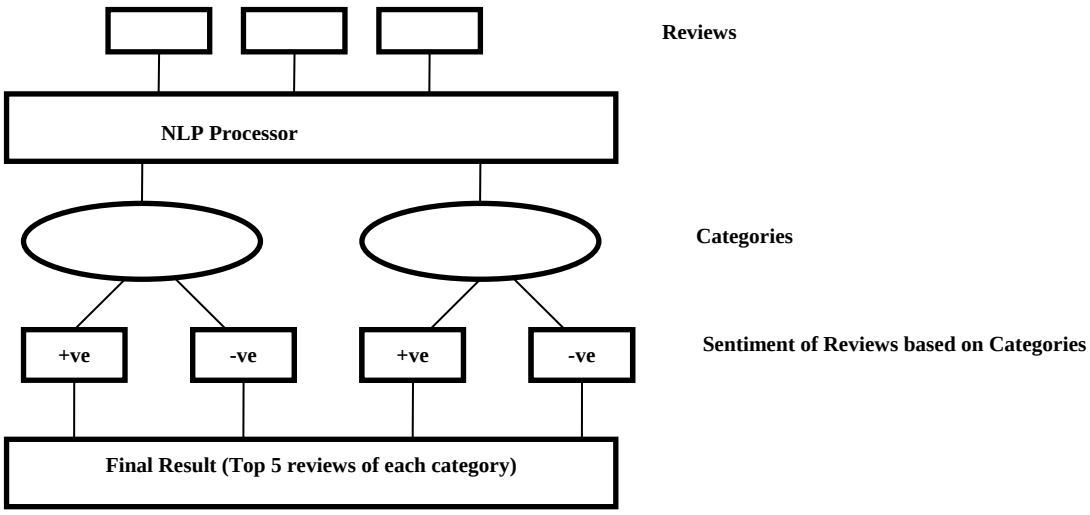
Sentiment analysis has been extensively investigated during the last years mainly for English language. Currently, existing approaches can be split into two main groups: methods based on the combination of lexical resources and Natural Language Processing (NLP) techniques; and machine learning approaches. In [1] paper, authors has also introduced the use of lexical databases, of user reviews in Spanish for the various sectors like accommodation, food and beverage, for Sentiment Analysis. A global sentiment score has been calculate based on the negative and positive words which appear in the review and using the mentioned lexicon database. They have tested algorithm with users online reviews that have been acquired from TripAdvisor.

In this paper [2], author proposes OpinionZoom, modular software that helps users in an easy manner to understand the vast amount of tourism opinions disposed all over the Web. Author was successful to implement and test OpinionZoom, which encompasses the situation of the tourism industry in Los Lagos, the Lake District, in Chile. Results showed the effectiveness of the designed proposal when applied to solving this specific industry's issues.

The paper [3] discusses Bing Liu's opinion mining approach based on aspects to apply it on the tourism domain. This approach is based on the fact that different users refer to different kinds of products very differently when writing reviews on the Web. As physical product reviews are focused in this paper, it was difficult to apply them on tourist domain. After a detailed research of product reviews on tourism available online, we found these features and then model them in this extension, hence highlighting the use NLP-based rules which were both new and more complex for the required tasks of sentiment classification at the aspect-level. The task of opinion visualization, summarization and proposing of new methods to show the users vast availability of opinions in an easy manner is also done.The paper also showed the development of an architecture for aspect-based opinion mining tool, which was then used analyze opinions from TripAdvisor for the tourism industry in Los Lagos. Results proved that this extension performed better than Liu's model for tourism domain by improving Accuracy and Recall for the tasks of subjective and sentiment classification. Also, the approach is largely effective in finding the sentiment orientation of opinions, which obtained an F-measure of 92% for this task. However, the algorithms on an average extracted about 35% of the explicit aspect expressions by using non-extended approach for this task. Finally, the results showed the effectiveness of this design when it was applied to solve the industry's specific issues in the Los Lagos, as almost 80% of the users using this tool considered that the tool adds valuable information to their business.

III. PROPOSED SYSTEM

A. Architectural Diagram



B. Implementation

•Module I- Review Collection

Input Reviews from the user.

•Module II- Review Processing

The system processes each review in the following manner:

Fragmenting each sentence

Tagging words with their respective POS

Annotating tagged words with associated sentiments

Analyzing each sentiment as positive or negative

•Module III- Categorization

The system then categorizes each review into different categories like weather, peak season, expected expenditure etc.

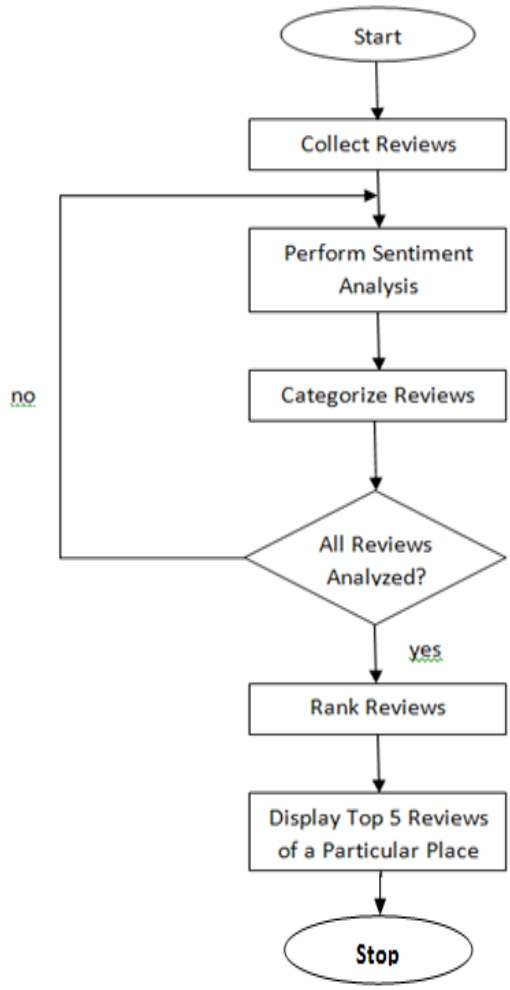
•Module IV- classification

Classify each review as positive or negative according to its category.

•Module V- Result Declaration

Display the final result (The Recommended Review) as per the criteria given by user.

C. Flow Diagram



IV. CONCLUSION

We will be using Stanford's Natural Language Processor to process the reviews of tourist places. And then will classify and segregate the reviews based on their categories like weather, peak season etc. so that user can easily extract useful information. We will also be providing top 5 reviews of all categories of a tourist place to the user. The main objective behind this system is that tourist can easily extract subjective and useful information. This system also reduces the time required for searching and easily supports growth of tourism.

REFERENCES

[1] I. F. Akyildiz, X. Wang, and W. Wang, "Wireless Mesh Networks: A Survey", Computer Networks and ISDN Systems, Vol.47, Issue-2, **2005**, pp.**445-487**.

[2] I. F. Akyildiz, and X. Wang, "A Survey on Wireless Mesh Networks", IEEE Radio Communications, Vol.43, Issue-3, **2005,** pp.**23-30.**

[3] M. Lee et al., "Emerging Standards for Wireless Mesh Technology", IEEE Wireless Communications, Vol.13, Issue-4, **2006**, pp**.56-63**.

[4] N.B. Salem, and J-P Hubaux, "Securing Wireless Mesh Networks", IEEE Wireless Communications, Vol.13, Issue-2, **2006**, pp.**50-55**.

[5] S. Han, E. Chang, L. Gao, T. Dillon, T., Taxonomy of Attacks on Wireless Sensor Networks, in the Proceedings of the 1st European Conference on Computer Network Defence (EC2ND), University of Glamorgan, UK, Springer Press, SpringerLink Date: December **2007.**

[6] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," Ad Hoc Networks 1, **2003**, pp. **293-315**.

[7] Y. Yang, Y. Gu, X. Tan and L. Ma, "A New Wireless Mesh Networks Authentication Scheme Based on Threshold Method," 9th International Conference for Young Computer Scientists (ICYCS-2008), **2008**, pp. **2260-2265**