IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 9, November 2014.
www.ijiset.com
ISSN 2348 – 7968

Sentiment Analysis in Tourism

[1]Aurchana.P, [2]Iyyappan.R , [3]Periyasamy.P ,

[1]Assistant Professor, Dept.of M.CA, Sri Manakula Vinayagar Engineering College, Puducherry

[2,3]PG Student , Dept. of M.CA, Sri Manakula Vinayagar Engineering College, Puducherry

ABSTRACT

The important part to gather the information is always seems as what the people think. Users express their views and opinions regarding products and services. These opinions are subjective information which represents user's sentiments, feelings or appraisal related to the same.Every day, millions of people travel around the globe for business, vacations, sightseeing, or other reasons. An astronomical amount of money is spent on tickets, accommodations, food, transportation, and entertainment. Tourism is an information-based business where there are two types of information flow. One flow of information is from the providers to the consumers or tourists.

This is information about goods that tourists consume such as tickets, hotel rooms, entertainments, and so forth. The other flow of information which follows a reverse direction consists of aggregate information about tourists to service providers. In this Chapter we will discuss the information flow about the behaviour of tourists. When the aggregated data about the tourists is presented in the right way, analysed by the correct algorithm, and put into the right hands, it could be translated into meaningful information for making vital decisions by tourism service providers. Data mining can be a very useful tool for analysing tourism-related data.

TOURISM DATA MINING

Usually two types of machine learning activities are common in tourism association learning and classification learning.

In association learning, the learning method searches for associations or relation- ships between features of tourist behaviour. For example, the algorithm may try to find out if tourists who are interested in shopping also prefer to stay near the centre of a city. That is, there is no specific target variable in this type of data mining, and so this is popularly known as unsupervised learning.

A second style of machine learning is classification learning. This learning scheme takes a set of classified examples from which it discovers a way of classifying unseen examples. This is a form of supervised learning, in which there is a specific target variable. For example, by using classification analysts may be interested to classify tourists into two groups' high spenders and low spenders for luxury items. In this case the target variable isexpenditure on luxury items. Based on a set of demographic and other variables the classification algorithm will establish the specific attributes of a tourist that qualify them as a high spender or a low spender. Next, we describe the various machine learning techniques used in tourism data mining.

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 9, November 2014.
www.ijiset.com
ISSN 2348 – 7968

USES OF DATA MINING IN TOURISM

The three main uses of data mining techniques in the tourism industry are:

(1) forecasting expenditures of tourists,

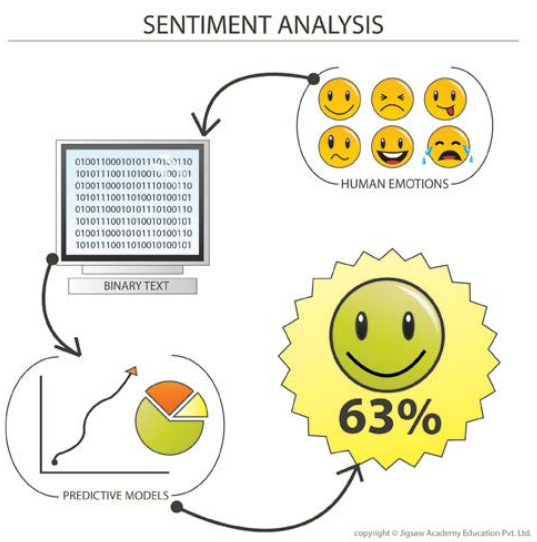
(2) Analysing profiles of tourists, and

(3) Forecasting number of tourist arrivals. In the following sections examples are presented to demonstrate how data

Mining techniques are used to support these activities.

SENTIMENT ANALYSIS

Sentiment analysis, also known as opinion mining, is the analysis of the *feelings* (i.e.



Attitudes, emotions and opinions) behind the words using natural language processing tools. It's looking beyond the number of Likes, Shares or Comments you get on an ad campaign, product release, blog post, and video to understand *how* people are responding to it. Was the review positive? Negative? Sarcastic? Ideologically biased?

Opinion target, opinion holder and opinion are the definitions used to extracting opinions from different online sources. An opinion can be expressed in two types.

1. Direct opinion,

2. Comparative opinion.

All the opinions are stored in a document. Following are the steps to extracting the opinions.

- Identify the objects.
- Feature extraction and synonym grouping Opinion orientation determination.
- Integration.

Existing System:

Existing approaches are based on different supervised and unsupervised methods using opinion words and phrases and the grammar information. One key issue is to identify opinion words and phrases (such as good, bad, poor, or great), which are instrumental to sentiment analysis. However, there are seemly an unlimited number of expressions

That people use to express opinions, and in different domains, they can be significantly different. Even in the same domain, the same word might indicate different opinions in different contexts.

Proposed System:

Our work builds on previous studies focusing on the relationship between the discussions held in firm-specific finance Web forums and public stock behaviour. However, instead of assuming a shareholder view of participants in a finance Web forum as in previous research, and considering them to be uniformly representative of investors, we adopted a stakeholder perspective. This perspective more accurately represents the diversity of the constituency groups participating in the Web forum and closely aligns the analysis with the corporation's stakeholder theory.

To address the broad questions posed in this research, and guided by the literature reviewed, we developed a framework for analysis with four major stages: stakeholder analysis, topical analysis, sentiment analysis, and stock modelling. During the stakeholder analysis stage, we identified the stakeholder groups participating in Web forum discussions. In the topical analysis stage, the major topics of discussion driving communication in the Web forum are determined.

The sentiment analysis stage consists of assessing the opinions expressed by the Web forum participants in their discussions. Finally, in the stock modelling stage, we examine the relationships between various attributes of Web forum discussions and the firm's stock behaviour.

Definition (Opinion):

An opinion is a quadruple, (g, s, h, t),

Where g is the opinion (or sentiment) target, s is the sentiment about the target, h is the opinion holder and t is the time when the opinion was expressed

Definition (entity):

An entity e is a product, service, topic, issue, person, organization, or event. It is described with a pair, e: (T, W), where T is a hierarchy of parts, sub-parts, and so on, and W is a set of attributes of e.

Definition (aspect category and aspect expression): An aspect category of an entity represents a unique aspect of the entity, while an aspectexpression is an actual word or phrase that appears in the text indicating an aspect category. Each aspect category (or simply aspect) should also have a unique name in a particular application.

The process of grouping aspect expressions into aspect categories (aspects) is called **Aspect categorization**: Aspect expressions are usually nouns and noun phrases but can also be verbs, verb phrases, adjectives, and adverbs. The following definitions are useful (Hu and Liu, 2004).

Definition (explicit aspect expression): Aspect expressions that are nouns and noun phrases are called explicit aspect expressions. For example, "picture quality" in "The picture quality of this camera is great" is an explicit aspect expression.

Definition (implicit aspect expression): Aspect expressions that are not nouns or noun phrases are called implicit aspect expressions.

For example: "expensive" is an implicit aspect expression in "This camera is expensive." It implies the aspect price. Many implicit aspect expressions are adjectives and adverbs that are used to describe or qualify some specific aspects, e.g., expensive (price), and reliably (reliability). They can also be verb and verb phrases, e.g., "I can install the software easily." "Install" indicates the aspect installation. Implicit aspect expressions are not just adjectives, adverbs, verbs and verb phrases; they can also be very complex, e.g., "This camera will not easily fit in a coat pocket." Here, "fit in a coat pocket" indicates the aspect size (and/or shape).

Comparative Opinions

Problem Definitions:

A comparative sentence expresses a relation based on similarities or differences of more than one entity. There are several types of comparisons. They can be grouped into two main categories: gradable comparison and non-gradable comparison

Gradable comparison: Such a comparison expresses an ordering relationship of entities being compared. It has three sub-types:

1. Non-equal gradable comparison: It expresses a relation of the type greater or less than that ranks a set of entities over another set of entities based on some of their shared aspects, e.g., "Coke tastes better than Pepsi." This type also includes preference, e.g., "I prefer Coke to Pepsi."

2.Equative comparison: It expresses a relation of the type equal to that states two or more entities are equal based on some of their shared aspects, e.g., "Coke and Pepsi taste the same."

3. Superlative comparison: It expresses a relation of the type greater or less than all others that ranks one entity over all others, e.g., "Coke tastes the best among all soft drinks." **Non-gradable comparison:** Such a comparison expresses a relation of two or more entities but does not grade them.

There are three main sub-types:

1. Entity A is similar to or different from entity B based on some of their shared aspects, e.g., "Coke tastes differently from Pepsi."

2. Entity A has aspect a1, and entity B has aspect a2 (a1 and a2 are usually substitutable), e.g., "Desktop PCs use external speakers but laptops use internal speakers."

3. Entity A has aspect a, but entity B does not have, e.g., "Nokia phones come with earphones, but iPhones do not."

In English, comparisons are usually expressed using comparative words (also called **comparatives**) and superlative words (also called **superlatives**).Comparatives are formed by adding the suffixes and superlatives are formed by adding the suffix -est to their base adjectives and adverbs. For example, in "The battery life of Nokia phones is longer than Motorola phones," "longer" is the comparative form of the adjective "long." "longer" (and "than") here also indicates that this is a comparative sentence. In "The battery life of Nokia phones is the longest," "longest" is the superlative form of the adjective "long", and it indicates that this is a superlative sentence.

English also has irregular comparatives and superlatives, i.e., more, most, less, least, better, best, worse, worst, further/farther and furthest/farthest.

Comparative keywords used in non-equal gradable comparisons can be further grouped into two categories according to whether they express increased or decreased quantities, which are useful in sentiment analysis. Increasing comparative: Such a comparative expresses an increased quantity, e.g., more and longer. Decreasing comparative: Such a comparative expresses a decreased quantity, e.g., less and fewer.

Types of Data, Features and Detection

Review content: The actual text content of each review. From the content, we can extract linguistic features such as word and POS n-grams and other syntactic and semantic clues for deceptions and lies. However, linguistic features may not be enough because one can fairly easily craft a fake review that is just like a genuine one. For example, one can write a fake positive review for a bad restaurant based on his true experience in a good restaurant.

Meta-data about the review: The data such as the star rating given to eachreview, user-id of the reviewer, the time when the review was posted, the time taken to write the review, the host IP address and MAC address of the reviewer's computer, the geo-location of theReviewer, and the sequence of clicks at the review site. From such data, we can mine many types of abnormal behavioural patterns of reviewers and their reviews.

For example, from review ratings, we may find that a reviewer wrote only positive reviews for a brand and only negative reviews for a competing brand. Along a similar line, if multiple user-ids from the same computer posted a number of positive reviews about a product, these reviews are suspicious. Also, if the positive reviews for a hotel are all from the nearby area of the hotel, they are clearly not trustworthy.

Product information: Information about the entity being reviewed, e.g., the product description and sales volume/rank. For example, a product is not selling well but has many positive reviews, which is hard to believe.

IMPLEMENTATION MODULES:

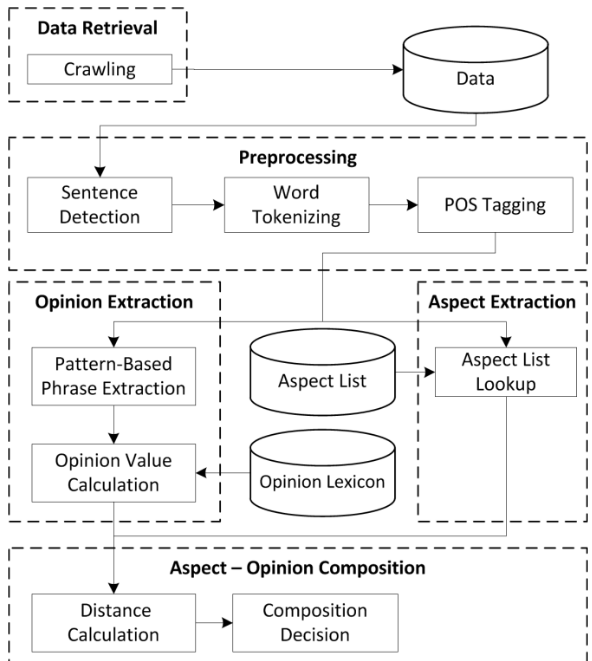
- ✓ Posting opinions
- ✓ Object identification
- ✓ Feature extraction
- ✓ Opinion-orientation determination
- ✓ Integration

MODULE DESCRIPTION:

Posting opinions: In this module, we get the opinions from various people about business, e-commerce and products through online. The opinions may be of two types. Direct opinion and comparative opinion. Direct opinion is to post a comment about the components and attributes of products directly. Comparative opinion is to post a comment based on comparison of two or more products. The comments may be positive or negative.

Object identification:

In general, people can express opinions on any target entity like products, services, individuals, organizations, or events. In this project, the term object is used to denote the target entity that has been commented on. For each comment, we have to identify an object. Based on objects, we have to integrate and generate ratings for opinions.



The object is represented as "O". An opinionated document contains opinion on set of objects as $\{o_1, o_2, o_3 \ldots o_r\}$.

Feature extraction:

An object can have a set of components (or parts) and a set of attributes (or properties) which we collectively call the features of the object. For example, a cellular phone is an object. It has a set of components (such as battery and screen) and a set of attributes

ISSN 2348 – 7968

(such as voice quality and size), which are all called *features* (or *aspects*). An opinion can be expressed on any feature of the object and also on the object itself.

With these concepts in mind, we can define an object model, a model of an opinionated text, and the mining objective, which are collectively called the *feature-based sentiment analysis model*. In the *object model*, an object "O" is represented with a finite set of features,

$$F = \{f_1, f_2,\ldots,f_n\}$$

Which includes the object itself as a special feature. Each feature f_i ?F can be expressed with any one of a finite set of words or phrases $W_i = \{w_{i1},w_{i2}, \ldots, w_{im}\}$

Which are the feature's *synonyms*.

Opinion-orientation determination:

Direct opinion:

A *direct opinion* is a quintuple $(o_j, f_{jk}, oo_{ijkl}, h_i, t_l)$,

where o_j is an object,

f_{jk} is a feature of the object o_j,

oo_{ijkl} is the orientation of the opinion on feature f_{jk} of object o_j,

h_i is the opinion holder, and

t_l is the time when the opinion is expressed by h_i.

The opinion orientation oo_{ijkl} can be positive, negative, or neutral.

Comparative opinion:

A *comparative opinion* expresses a preference relation of two or more objects based their shared features. A comparative opinion is usually conveyed using the comparative or superlative form of an adjective or adverb, such as "Coke tastes better than Pepsi."

Integration:

Integrating these tasks is also complicated because we need to match the five pieces of information in the quintuple. That is, the opinion oo_{ijkl} must be given by opinion holder h_i on feature f_{jk} of object o_j at time t_l .To make matters worse, a sentence might not explicitly mention some pieces of information, but they are implied using pronouns, language conventions, and context. Then generate ratings based on above tasks. Thus we can clearly see how holders view the different features of each product.

Algorithm

1.OpinionFeatureExtraction

Input: a collection of reviews, D={d1, d2, ..., d}, and a set of commonly used seed adverbs, adverbList, emotion verb

list V, minimum support threshold minsup, minimum

similarity threshold minsim, and length of n-gram n

Output: a set of opinions, P, and a set of features, F

1. P = ExtractOpinion(D, adverbList)

2. F=£, tempAdj = P;

3. Do

4. tempF=£;

5. For each sentence s of each review d in D

6. For each adj in tempAdj

7. nounPhrases=ExtractNoun(s,adj);

8. For each n-gram phrase nph in nounPhrases

9. insertnph into tempF and accumulate support if

it is repeated;

10. if f2FOP such that f.feature is null, f.s_id=s.s_id

andf.adj=adj

11. f.feature=nph

12.else insert(d.user_id,d.item_id,s.s_id,nph,adj) into

FOP;

13. If tempF is empty return;

14. tempAdj=£;

15. For each sentence s of review d inD

16. For each adj extracted from s based on tempF.

17. if(adj is not in P)
18. insertadj into tempAdj;
19. insert (d.user_id, d.item_id, s.s_id, £,adj) intoFOP;
20. P = P tempAdj, F = F tempF;
21. while (tempAdj – £)
22. for each sentence s of review d that contains no featurein F and no opinion in P
23. if s contains any emotion verb v2V
24. insert (d.user_id,d.item_id,s.s_id, £,v) into FOP
25. nounPhrase=ExtractNounE(s,v); //extract n-gramnoun phrases around verb
v
26. Insert nounPhrase into F and accumulate support for repeated ones;
27. For each n-gram nph in nounPhrase
28. insert (user_id,item_id,s_id,nph,v) into FOP;
29. MergeSynonymFeature(FOP,F, minsim); //merge the
synonym features in F
30. Remove from FOP and F features that have support lessthan threshold minsup;

CONCLUSION

Sentiment detection has a wide variety of applications in e- commerce. It helps in classifying, summarizing reviews and in other real time applications. This paper focuses on the survey of techniques and approaches that promise to enable mining of implicit aspects from travellers reviews. Challenges still exist in the area of implicit aspect identification. Finding accurate solutions is one of the main existing issues. Very few attempts have been made to extract implicit features. Implicit aspects play a significant role in determining sentiments from customer reviews. It has been observed that extracting implicit aspects in addition to explicit features has resulted in increase in sentiment analysis results.

REFERENCES:

1.International Journal of Computer Applications (0975 – 8887) Volume 56– No.13, October 2012 Classification of Opinion Mining Techniques.

2.International Journal of Computer Applications (0975 – 8887) Volume 98– No.4, July 2014

Implicit Aspect Identification Techniques for Mining Opinions: A Survey.

3.Electronic Commerce Research and Applications Combining user preferences and user opinions for accurate recommendation 1567-4223/$ - see front matter _ 2012 Elsevier B.V. All rights reserved.

4.Identifying Customer Preferences about Tourism Products using an Aspect-Based Opinion Mining Approach.

5.Implicit Aspect Identification Techniques for Mining Opinions: A Survey- International Journal of Computer Applications (0975 – 8887) Volume 98– No.4, July 2014

6.Sentiment Analysis and Opinion Mining - April 22, 2012 ,Bing Liu liub@cs.uic.edu.

7. Opinion Mining Bing Liu Department of Computer -Science University of Illinois at Chicago liub@cs.uic.edu.

8.Data mining techniques and applications – A decade review from 2000 to 2011Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Hsiao.

9. Opinion Mining Textual Datasets Patrick Moran Bethany Herwaldt, Jeffrey Salter Shaina Race, Ralph Abbey Dr. Carl Meyer.

10. Opinion Mining: Exploiting the Sentiment of the Crowd
Diana Maynard Adam Funk Kalina Bontcheva
University of Sheffield, UK.

Reference sites:

1. https://semantria.com/sentiment-analysis
2. http://www.alchemyapi.com/products/alchemylanguage/sentiment-analysis/
3. http://www.alchemyapi.com/api/sentiment-analysis/
4. http://text-processing.com/demo/sentiment/
5. http://sentistrength.wlv.ac.uk/
6. https://class.coursera.org/nlp/lecture/145
7. http://www.streamcrab.com/
8. www.tripadviosor.com
9. http://tourism.pondicherry.gov.in/