NavaJyoti, International Journal of Multi-Disciplinary Research
Volume 3, Issue 1, August 2018

Sentimental Analysis of a Product & its Reviews Using Machine Learning Technique

Monisha.B, ShreeLakshmi.P

[1]*Dept of BCA, The Bangalore Social and Educational Institute of Management Studies. Bangalore, India*
[2]*Dept of BCA, The Bangalore Social and Educational Institute of Management Studies. Bangalore, India*

ABSTRACT

Due to the more volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form, much of the current research is concentrating more on the area of sentiment analysis of the products Using Decision tree and Chi-square technique. Now a days people are forced to build a system and understand and represented in an electronic text in means of results. A perfect method for predicting sentiments could enable us, to extract opinions from the internet and predict online customer's preferences, which could prove valuable for economic or marketing research. This paper presents a survey covering on Sentimental analysis of the product and the reviews using machine learning technique.

Keywords: opinion mining, corpora, Decision tree, sentimental analysis, machine learning technique.

INTRODUCTION

Product mining (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Generally speaking, product analysis aims to determine the attitude of a speaker or a writer with respect to some product and the attitude may be his or her judgment or evaluation affective state or the intended emotional communication. The Web contains a wealth of opinions about products, which are expressed in newsgroup posts, review sites, and elsewhere. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. As a result, the problem of "opinion mining" has seen increasing attention over the last three years from (Turney, 2002; Hu and Liu, 2004) and many others. This survey covers techniques and approaches that promise to directly enable opinion.

Oriented information-seeking systems. Our focus is on methods that seek to address the new challenges raised by sentiment-aware applications, as compared to those that are already present in more traditional fact-based analysis. Our beliefs and perceptions of reality, and the choices we make, are largely conditioned on how others see and evaluate the world. For this reason, when we need to make a decision we often seek out the opinions of others. This is true not only for individuals but also for organizations

CHALLNEGE IN OPINION MINING

Opinion mining is a relatively recent discipline that studies the extraction of opinions using Artificial Intelligence and/or Natural Language Processing techniques. More informally, it's about extracting the opinions or sentiments when given a piece of text. This provides a great source of unstructured information especially opinions that may be useful to others, like companies and their competitors and other consumers. For example, someone who wants to buy a camera, can look for the comments and reviews from someone who just bought a camera and commented on it or written about their experience or about the camera manufacturer. He can get feedback from customer and can make the decision.

Also a manufacturing company can improve their products or adjust the marketing strategies.

Opinion Mining needs to take into account how much influence any single opinion is worth. This could depend on a variety of factors, such as how much trust we have in a person's opinion, and even what sort of person they are. It may differ from person to person like an expert person and any non-expert person. There may be spammers. Also we need to take into account frequent vs. infrequent posters, Consider a following segment of few sentences talked about iPhone.(1) My friend bought an iPhone a few days ago. (2) It was such a nice phone. (3) The touch screen was really cool. (4) The voice quality was clear too. (5) However, my mother was mad with me as I did not tell her before I bought it. (6) She also thought the phone was too expensive, and wanted me to return it to the shop"[28] the question is: what we want to mine or extract from this review? The first thing that we notice is that there are several opinions in this review. Sentences (2), (3) and (4) express some positive opinions, while sentences (5) and (6) express negative

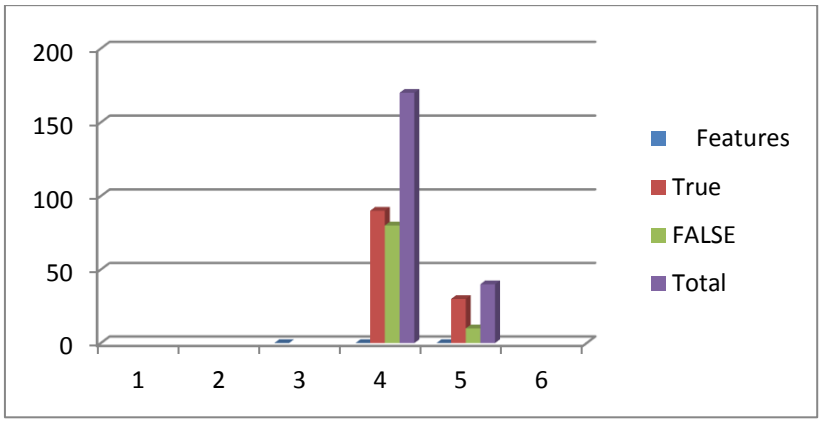
Opinions or emotions. Then we also notice that the opinions all have some targets. The target of the opinion in sentence (2) is the iPhone as a whole, and the targets of the opinions in sentences (3) and (4) are \touch screen" and \voice quality" of the iPhone respectively. The target of the opinion in sentence (6) is the price of the iPhone, but the target of the opinion/emotion in sentence (5) is \me", not iPhone. Finally, we may also notice the holders of opinions. The holder of the opinions in sentences (2), (3) and (4) is the author of the review (\I"), but in sentences (5) and (6) it is \my mother"[28]. From the above example we understand that in general, opinions can be expressed about anything, e.g., a product, a service, an individual, an organization, an event, or a topic, by any person or organization. We use the entity to denote the target object that has been evaluated. An entity is a product, service, person, event, organization, or topic. The entity consist of components or parts, sub-components, and so on, and there are set of attributes of entity. Each component or sub-component also has its own set of attributes [28].For example, a particular brand of cellular phone is an entity, e.g., iPhone. It has a set of components, e.g., battery and screen, and also a set of attributes, e.g., voice quality, size, and weight. The battery component also has its own set of attributes, e.g., battery life, and battery size. There are two main types of opinions: regular opinions and comparative opinions. Regular opinions are often referred to simply as opinions in the research literature. A comparative opinion expresses a relation of similarities or differences between two or more entities, and/or a preference of the opinion holder based on some of the shared aspects of the entities. A comparative opinion is usually expressed using the comparative or superlative form of an adjective or adverb, although not always. An opinion is simply a positive or negative sentiment, attitude, emotion or appraisal about or an aspect of the entity from an opinion holder. Positive, negative and neutral are called opinion orientations also called sentiment orientations, semantic orientations, or polarities [28].

EXPERIMENTAL RESULTS

In this paper we are Extracting the experimental results by taking two different method i,e Decision Tree Technique and Chi-Square technique. Table.1 shows some of the sentimental analysis of the Amazon product using different parameters.

Table 1 Sentimental analysis of product

Features Product 1	**True**	**False**	**Total**
Positive	90	80	170
Negative	**30**	10	40



Graph 1 Sentimental Analysis of a Product

a. Decision tree:

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning.

A decision tree consists of three types of nodes:[1]

Decision nodes – typically represented by squares

Chance nodes – typically represented by circles
End nodes – typically represented by triangles

The decision tree can be linearized into **decision rules**, where the outcome is the contents of the leaf node, and the conditions along the path form a conjunction in the if clause. In general, the rules have the form:
if (condition1)*and* (condition2) *and(* condition3) *then* outcome.

Decision rules can be generated by constructing association rules with the target variable on the right. They can also denote temporal or causal relations. An Association rule is a **rule**-based machine learning method for discovering interesting relations between variables in large databases.Figure1 shows the pictorial representation of the decisions for the sentimental analysis of a product

.

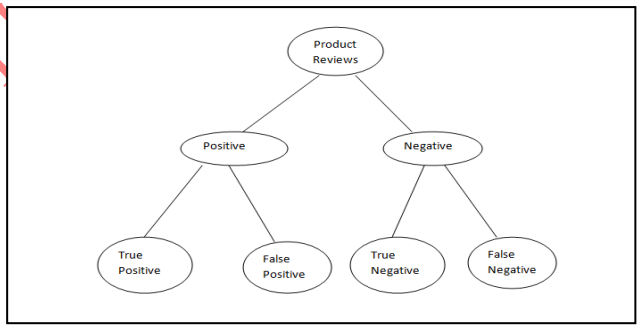


Figure.1.Sentimental Analysis of product using Decision Tree

The decision tree technique is the easiest way to extract the result from the set of reviews. We have applied the decision tree technique for our sentimental analysis of the product by this analysis we are able to take the proper decision that is either in a positive way or else in a negative way.

b. **Chi-squared test**

chi-squared test, also written as χ^2 test, is any statistical hypothesis test wherein the sampling distribution of the test statistic is a chi-squared distribution when the null hypothesis is true. Without other qualification, 'chi-squared test' often is used as short for *Pearson's* chi-squared test.

Chi-squared tests are often constructed from a sum of squared errors, or through the sample variance. Test statistics that follow a chi-squared distribution arise from an assumption of chi-squared test, also written as **χ^2 test**, is any statistical hypothesis test wherein the sampling distribution of the test statistic is a chi-squared distribution when the null hypothesis is true. Without other qualification, 'chi-squared test' often is used as short for *Pearson's* chi-squared test.

Chi-squared tests are often constructed from a sum of squared errors, or through the sample variance. Test statistics that follow a chi-squared distribution arise from an assumption of independent normally distributed data, which is valid in many cases due to the central limit theorem. A chi-squared test can be used to attempt rejection of the null hypothesis that the data are independent.

Formula

Expectation (AB) $=\frac{A\times B}{N}$

Where 'A' represents Positive

'B' represents Negative

Expectation (AB) =170*40/210

=1800/210 =32.38%

We have done the calculation by taking the reviews, so according to the expectation of a particular product, the expectation value should reach more than 70% (according to our experimental result), by the calculation we are getting 32.38% is the expectation level it has not reached the expectation level of human being so through this survey we can able to say the expectation value will not be suitable to take the proper decision about the product

APPLICATIONS

Opinions are so important that whenever one needs to make a decision, one wants to hear others' opinions. This is true for both individuals and organizations. The technology of opinion mining thus has a tremendous scope for practical applications. Opinion mining is useful for Individual consumers. If one wants to purchase a product, it is useful to see a summary of opinions of existing users so that he/she can make an informed decision. This is better than reading a large number of reviews to form a mental picture of the strengths or weaknesses of the product. He/she can also compare the summaries of opinions of competing products, which is even more useful. Study of reviews about the product is important for organizations and businesses. This is important for the organizations to make improvements in the products. For example, it is critical for a product manufacturer to know how consumers perceive its products and those of its competitors. This information is not only useful for marketing and product benchmarking but also useful for product design and product developments. Manufacturing companies can even think of increasing or decreasing the manufacturing of some product. Opinion mining is also useful for the advertisement companies. These companies can get an idea about the market. The type of products people liked most, what is the overall thinking of people about something; all such points can be extracted using opinion or review mining. This is important for creating and designing advertisement by the advertisement company.

CONCLUSION

Opinion mining is a new field of study. This is important because, in this competitive world, every customer try to compare multiple products before purchasing. Also the organization needs customer opinion about their products to being the competition and to put improvements in their products.

This is a recent trend in research also. While the development of the opinion mining tools described shows very much work in progress and initial results are promising, but still it requires a lot of refinements

Since this is a study of sentiments of a person, so it requires a lot of precision. Whenever any person talk about something, then the context in which he is talking and how the sentence is formed may change the phrasing method to catch the exact opinion said by that person.

By the analysis of two methods that is decision tree and chi square method, the result of decision tree technique is more effective and very easy to calculate when compared to chi square technique. Through this conclusion we would like to conclude that the decision tree technique will be more effective

REFERENCES

[1]Abraham Kandel George Meghabghab. Search Engines, Link Analysis and User's Web Behaviour. Springer, 2008.
[2] Arthur M. Langer. Analysis and Design of Information Systems. Springer; 3rd edition, 2007.
[3] Brad Ediger. Advanced Rails. O G Popa, 2008.
[4 Brian Clifton.Advanced Web Metrics with Google Ana- lytics.Sybex, 2 edition, 2010.
[5] Chunping Li Lili Zhao. Ontology based opinion mining for movie reviews. KSEM 2009, 2009.
[6] David A. Freedman. Statistical Models: Theory and Prac-tice. Cambridge University Press, 2 edition, 2009.
[7] David Heinemeier Hansson Leonard Richardson, Sam Ruby. Restful Web Services. O'Reilly Media; 1 edition, 2007
[8] David M. PennockKushal Dave, Steve Lawrence. Min- ing the peanut gallery: Opinion extraction and semanticclassi_cation of product reviews. WWW'03, 2003. 34, 36
[9] E cient and E ective Clustering Methods for Spatial Data Mining RT Ng, J Han - Proc. of, 1994
[10] Eibe Frank Ian H. Witten. Data Mining: Practical Ma- chine Learning Tools and Techniques. Morgan Kaufmann; 2 edition, 2005.
[11] HAN," Efficient and effective clustering methods for spatial data mining R Ng - Proceedings of the 20th Conference on VLDB Cited by 2 - Related articles
[12]HoracioSaggionKalinaBontcheva Christian LeiboldAdam Funk, Yaoyong Li. Opinion analysis for business intelligence applications. OBI'08, 2009.
[13] Liu Bing Hu Minqing. Mining and summarizing cus-tomer reviews. KDD'04, 2004.vii, 27, 30, 31, 32, 33, 34, 35, 38, 59
[14] Liu Bing. Web Data Mining. Springer, 2007
[15] Lynne Schrum Gwen Solomon. Web 2.0: New Tools, New Schools. International Society for Technology in Educa-tion; First Edition edition, 2007.
[16] Max Bramer. Principels of Data Mining.Springer, 2007. 21
[17] Michael Milton. Head First Data Analysis: A Learner's Guide to Big Numbers, Statistics, and Good Decisions. O'Reilly Media, 1 edition, 2009.
[18] Michael Papazoglou. Web Services: Principles and Tech- no logy. Prentice Hall; 1 edition, 2007
[19] Mining frequent patterns without candidate generation J Han, J Pei, Y Yin ACM Sigmod Record 29 (2)
[20] Mining frequent patterns without candidate generation: A frequent-pattern tree approachJ Han, J Pei, Y Yin, R MaoData mining and knowledge discovery
[21] Oren Etzioni Ana-Maria Popescu. Extracting product features and opinions from reviews.EMNLP-05, 2005. 34
[22] Paul Kimmel. UML Demysti_ed. McGraw-Hill Osborne Media; 1 edition, 2005.
[23] Philip S. Yu Ding Xiaowen, Liu Bing. A holistic lexicon- based approach to opinion mining. WSDM'08, 2008.vii, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 39, 58, 59, 87
[24] RakeshAgrawal and RamakrishnanSrikant.Fast algo- rithms for mining association rules. 1994.
[25] Russ Olsen. Design Patterns in Ruby. Addison-Wesley, 2008.
[26] Simon Corston-Oliver Eric Ringger Michael Gamon, An-thonyAue. Pulse: Mining customer opinions from free text. IDA, 2005. 34
Language.O'Reilly, 2008.