International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 07 | July-2016 www.irjet.net p-ISSN: 2395-0072

Enhance Digital Marketing Using Sentiment Analysis and End User Behavior

Sagar A. Kinholkar [1], Prof. K. C. Waghmare [2]

[1]*Student, Dept. of Computer Engineering, PICT, Pune, India*
[2]*Professor, Dept. of Computer Engineering, PICT, Pune, India*

Abstract - *In case of digital marketing, an end user behavior is completely unforeseeable. In order to earn maximal profit through digital marketing, it is necessary to correctly find out the behavior pattern of the prospect end user. Tracking the end user activities and analysis on their opinions helps to identify a probable interest of the end user. This can be achieved with various data mining methods. Sentiment analysis aims to find the attitude of a speaker or a writer with respect to some topic or the overall context of a document. To solve the given problem a framework can be created which track end user behavior and performed sentiment analysis on their opinion.*

Key Words: Data Mining, Business Intelligence, Opinion mining, Sentiment Analysis, Web Application.

1. INTRODUCTION

Digital Marketing play an important role in industry or any organization. Without marketing no one can sell their product. Digital Marketing is the process of building and maintaining customer Relationships through online activities to generate sales and/or capture Customers that are searching on the Internet for answers. Digital Marketing also helps to increase profit and business revenues. The main objectives of Digital Marketing are Maximize return on investment (ROI), Reach the right audience, Engage with your audience etc. It is Cost effective (only time and effort) [9].

Just putting advertisement or other marketing documents on Internet is not sufficient. It will not help to increase profit and business revenues. The proper use of digital marketing should be done. We require more techniques along with Digital Marketing to increase profit. We can track user activities on website. It will help us to find interest of end user at some extent. Also with this tracking activities technique's, Sentiment analysis can be used to find interest of a prospect end user. Sentiment Analysis aims to determine the attitude of the author of a specific piece of content with respect to the topic of interest. Comments and content can be referred to as Positive, Negative, Neutral or have no sentiment at all [11]. Sentiment analysis is also called as Opinion Mining.

Sentiment analysis (SA) is a computational study of opinions, sentiments, emotions, and attitude expressed in texts towards an entity [11]. Sentiment analysis (also called opinion

mining, review mining or appraisal extraction, attitude analysis) is the task of detecting, extracting and classifying opinions, sentiments and attitudes concerning different topics, as expressed in textual input [12,14]. SA helps in accomplishing different objectives like watching public mood in regards to political development, market knowledge [13], the estimation of consumer loyalty, movie sales prediction and many more.

Ever growing use of the Internet and online activities (like visiting, conferencing, reconnaissance, ticket booking, online exchanges, e-business, blogging, clicks streams, etc.) drives us to extract, change, load, and investigate a large quantity of structured and unstructured data, at quick paces, referred to as Big Data. Such information can be examined using a combination of Data Mining, Web Mining, and Text Mining methods in different real life applications.

Find behavior of the end user on the web is a critical task. A huge amount of working data is produced when end-user connect with web environment. This data is stored in various form and it may be useful in finding the interest of the end-user. Pointing out the prospect end-user in a web environment is a challenging task [10]. The main goal of the behavior analysis is to find out the interest of a prospect end-user towards particular things e.g. any product, movie, or organization. It is difficult to find individuals opinion about the organization. The organization, in this case, may be educational institutes or other.

Now a days, a large number of educational institute are Spending money on advertisement. The primary goal of these institutes are to increase the number of candidates in that institutes. Each and every educational institute can't meet their goal just by giving advertise on the internet.

2. MOTIVATION

In a case of business, it is necessary to understand the need and preferences of an end user. End user's activities and behavior helps to develop an effective marketing strategy. To build more advanced insights on market trends and customer behavior requires to understand of end user opinion.

It necessary to bring positive engagement to a customers. Grow business opportunities and revenues is the primary goal of any organization and to achieve this goal digital

marketing along with advanced methods are required. It increases customer responses, conversions and clicks, and to decrease churn. Each customer's score informs actions to be taken with that end user.

Following are some motivational factor that need to be consider in order to maintain positive engagement with user.

1. Necessary to understand the need and preferences of an end user.
2. Develop an effective marketing strategy.
3. To bring positive engagement to a customers.
4. Grow business opportunities and revenues.

3. METHODOLOGY

In the last one and half decades, research groups, the educated community, open and administration commercial enterprises are working thoroughly on Sentiment Analysis, also called Opinion Mining, to extract and analyze public mood and views. A bag-of-words strategy poses various problems in analyzing the sentiments of the opinions or reviews. Today, most of the research have been focused on developing highly developed approaches for Sentiment Classification.

Pang and Lee(2003)(2004) came up with a subjectivity classifier create a min-cut based classification with the help of Naive Bayesian classifier in order to divide subjective portions and objective portions of the text. A Large number of approaches that was based on the subjective portions of the text. Actually affect the polarity of the document. The subjectivity extracts all subjective portion from the text and then passed on to the classifier typically SVM to decide the overall polarity of the text. Using this approach, Pang and Lee showed that taking the subjectivity extracts out and used sentiment analysis on it gives better accuracies than applying to the complete text [1, 2].

SentiFul: A Lexicon for Sentiment Analysis (2011). In this approach, a WordNet dictionary is used to give a score to the particular term.[4] An aspect- based opinion polling algorithm (2011) is used to find out the different aspect of a customer reviews. One review may contain different aspects. Identification of these aspects is important. It uses a multi-aspect bootstrapping method and an aspect-based segmentation model [3].

Research works have been carried out for social network analysis and sentiment analysis for identifying mood of the people with respect to some product. Sentiment analysis study has been carried over two decades. Is text positive or negative, this classification is done using words (Tokenization) (2012), i.e. cool is positive and unpleasant is negative using polarity to create of positive and negative sentiment [5]. Sentiment tokenization has problems while handling HTML and XML markup, Twitter markup, Numbers are difficult to handle. Also other approaches like Part-of-speech tagging, n-grams (2010) effectively find sentiment of twitter comments using the machine learning techniques [6].

Different sentiment lexicon providing classes of positive, negative, strong, pronoun, quantifier and many more which are used to create WordNet of positive and negative sentiment lexicon. SENTIWORDNET (2011) is produced based on the automatic annotation of all the synsets of WORDNET. This is done according to the concept of positivity, negativity, and neutrality. Each synsets is corresponding to three numerical scores Pos(s), Neg(s), and Obj(s) which state that how positive, negative, and neutral the terms contained in the synsets [4].

4. FEATURE SELECTION IN SENTIMENT CLASSIFICATION

To do sentiment analysis, it is necessary to extract a feature from the text, reviews, or documents. Some of the feature are,

Terms presence and frequency: In this approach word and their frequency counts. It also specify whether the term is present or not. It uses term frequency weights to specify importance of the term.

Parts of speech (POS): In this approach, adjective from the text or reviews are find out.it is important to identify opinions.

Opinion words and phrases: These words are used in opinions to express someone's feelings.

Negations: The negative words are also used to express negative opinion about different aspects.

4.1 Feature selection methods

4.1.1 Pointwise Mutual Information (PMI)

Pointwise Mutual information (PMI) based measure [3]. PMI between two term term1 and term2 is calculated with the help of following formula:

$$\mathrm{PMI}(term_1 , term_2) = \log_2 \left(\frac{Pr(term_1 \wedge term_2)}{Pr(term_1) \cdot Pr(term_1)} \right)$$

Where,

P_r (term1 ^ term2) is a probability of how often term1 and term2 co-occurrence. P_r (term1) is probability of occurrence of term1.

P_r (term2) is probability of occurrence of term2.

For every term a sentiment association score is generated

Score (term) = PMI (term, Positive) - PMI (term, Negative)

If Score (term) > 0, then word is positive.

If Score (term) < 0, then term is negative.

4.2 Sentiment classification techniques

To categorized documents, reviews and opinion into different classes, some classifier are used in Sentiment Analysis.

4.2.1 Naive Bayesian

Naive Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computation involved and, in this sense, is considered "naive" [17].

4.2.2 Bayesian Network (BN)

Bayesian networks (BNs), also known as belief networks (or Bayes nets for short), belong to the family of probabilistic graphical models (GMs). These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics [16].

4.2.3 Maximum Entropy Classifier (ME)

The Max Entropy classifier is a probabilistic classifier which belongs to the class of exponential models. Unlike the Naive Bayes classifier that we discussed in the previous article, the Max Entropy does not assume that the features are conditionally independent of each other. The MaxEnt is based on the Principle of Maximum Entropy and from all the models that fit our training data, selects the one which has the largest entropy. The Max Entropy classifier can be used to solve a large variety of text classification problems such as language detection, topic classification, sentiment analysis and more. [15]

4.2.4 Support Vector Machines Classifiers (SVM)

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships.
Support Vector Machine (SVM) is primarily a classier method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables [18].

4.2.6 Decision tree classifiers

Decision Tree Classifier is a very popular classification technique. Decision Tree Classifier uses precisely created questions about the test record. Every time it get an answer until a proper class label of the record is found.
The decision tree classifiers composed of number of test questions and conditions in a tree format.

4.3 Lexicon based approach

In case of Lexicon based approach lexicons are used to determine the polarity of a text.
The lexicon-based approach involves calculating orientation for a document from the semantic orientation of words or phrases in the document [2].

5. CONCLUSIONS

The Feature selection methods and sentiment classification techniques help to find out user interest and opinion about the different aspects effectively. Sentiment analysis is an evolving field with a variety of use applications. Although sentiment analysis tasks are challenging due to their natural language processing origins, much progress has been made over the last few years due to the high demand for it.

ACKNOWLEDGEMENT

We express our profound gratitude to P. T. Kulkarni, Principal, Pune Institute of Computer Technology. We express our deep felt gratitude to Dr. R.B. Ingle, Head of the Department, computer engineering.

REFERENCES

[1] Pang Bo and Lillian Lee Sentiment Classification using Machine Learning Techniques", Elsevier Inc. 2003.

[2] Pang Bo and Lillian Lee Sentiment Analysis Using Subjectivity summarization based on Minimum Cuts", Elsevier Inc.2004.

[3] Jingbo Zhu, Member, IEEE, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou, Member, IEEE, and Matthew Ma, Senior Member, IEEE Aspect-Based Opinion Polling from Customer Reviews IEEE Transactions on ffective computing, vol. 2, no. 1, january-march 2011.

[4] Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka,SentiFul: A Lexicon for Sentiment Analysis IEEE transactions on affective computing, vol. 2, no. 1, january-march 2011.

[5] Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou Movie Rating and Review Summarization IEEE transactions on systems, man, and cyberneticspart c: applications and reviews, vol. 42, no. 3, may 2012.

[6] Brendan OConnor Ramnath Balasubramanyan "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series" Proceedings of the Fourth International AAAI Conference on Web logs and Social Media, Washington, DC, pp. 511-519 2010.

[7] Jaworska, J., Sydow, M.: behavioral Targeting in On-line Advertising: An Empirical Study, 2008, Accepted for the 9th International Conference on Web Information Systems Engineering (Wise 2008), Auckland, New Zealand, September 1-4, 2008 (to be printed in LNCS, Springer).

[8] Pablo Loyola n, PabloE.Roma n, JuanD.Velasquez, Predicting web user behavior using learning-based ant colony optimization , Department of Industrial Engineering,University Chile,Republication ,Santiago, Chile 2011.

[9] Newcomb, K.:Search Marketing Shows Strength in 2006, search engine watch.com, 2007.

[10] Jiang Long, Mo Yu, Ming Zhou, Xiaohua Liu and Tisjun Zhao, "Targetdependent twitter sentiment classification" In Preceeding of the 49 Annual Meeting of the Association for Computational Linguistics:Vol. 1, pp 151-160, 2011.

[11] W. Medhat, Sentiment analysis algorithms and applications: a survey, Ain Shams Eng. J. (2014), http://dx.doi.org/10.1016/j.asej.2014.04.011.

[12] A. Balahur, Methods and Resources for Sentiment Analysis in Multilingual Documents of Different Text Types, PhD Thesis, University of Alicante, Spain, 2011, 273p.

[13] Y.M. Li, T.-Y. Li, Deriving market intelligence from microblogs, Decis. Support Syst. 55 (2013) 206217.

[14] A. Montoyo, P. Martnez-Barco, A. Balahur, Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments, Decis. Support Syst. 53 (2012) 675679.

[15] http://blog.datumbox.com/machine-learning-tutorial-the-max-entropy-text-classifier

[16] http://www.rusnauka.com/3_ANR_2012/Economics/8_99606.doc.htm

[17] https://tom.host.cs.st-andrews.ac.uk/ID5059/L15-LeungSlides.pdf

[18] http://www.statsoft.com/Textbook/Support-Vector-Machines