ORIGINAL RESEARCH PAPER
Volume - 11 | Issue - 01 | January - 2022 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

A PERCEPTION CLASSIFICATION USING HYBRID FEATURE EXTRACTION

Engineering

Dr. C. Suganthi Associate Professor/ CSE, Muthayammal College of Engineering, Rasipuram – 637408.

ABSTRACT

The estimation of product quality is more significant during the purchasing of online products. Therefore, many opinion mining and sentiment classification methods were introduced to purchase the best products through online shopping. But, these classification methods haven't attained the effective product classification with best reviews and ratings. Several popular classification algorithms estimated along with three filtering Methods. Presented a hybrid technique to extract the product features. In this method, association rules and point-wise mutual information is combine. Large amount of features of products, not only to increase the time of computation but also to increase accuracy of classification. Optimize the amount of marketing funds spent on each customer, and just making a binary decision on whether to market to him. General Purpose Emotion Lexicons (GPELs) that associate with emotion categories remain a valuable resource for emotion analysis. Proposed a hybrid feature extraction method PCA (Principle Component Analysis) using lexicon-based method to classify and separate the products from the large set of different products depending on their features, best product ratings and positive reviews.

KEYWORDS

Dimensionality Reduction, Feature Extraction, Machine Learning, Principal Component Analysis (PCA).

INTRODUCTION

Machine learning is the subset of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning is the scientific study of algorithms and statistical models that computer systems use to perform specific task without using explicit instructions relying on patterns and inference instead. Data mining is the field study within machine learning and focus on data analysis through unsupervised learning. Sentiment analysis (SA) is the task of determining the emotional state of a writer based on their written texts by considering the polarity of keywords used in the writing. The meaning of keywords could be positive, negative, or neural. For example by using words like 'happy', 'angry', 'sad' or 'indifferent'. SA is done by detecting the contextual polarity of documents using semantic orientation technique and machine learning. This can limit the classifying of binary positive and negative sentiments, or rate emotional levels on a rank.

For instance, a case study of mobile review was used in which reviews were rated on appreciation of the mobile phone from 1 to 5. Various filtering techniques reduce the original training data to different degrees, leading to our name of Hierarchical Classification for our model. Additionally, this allows us to explore the relationship between document topic, polarity words, and high- frequency words. The dramatic development of Web 2.0 technology in the past few years has greatly changed people's life styles especially people's shopping patterns have experienced a significant change. People who want to purchase something new products mainly focus on viewing a large number of reviews about the relative products or services beforehand, so they can make reliable decisions about product. Mining valuable information from these product reviews not only provides some necessary purchase information for the potential consumers but also helps producers track the feedbacks of users on time.

The feedback and review contributes producers to maintain the good specification of products and improve the negative reviews about products timely, and finally make them gain competitiveness in the future. Viral marketing mainly focus on promoting a good product in society. This "word-of-mouth" advertising can be much more cost effective than traditional methods since it leverages the customers themselves to carry out most of the promotional effort.

Most of the branded products are making advertisement in online or on social media websites such as YouTube as an example. A small company develops a product and wants to market it through an online social network. It has a limited amount of money such that it can only select a small number of initial users in the network to use it. The company aims that the initial users to love the product and start influencing their friends on the social media to use it, and their friends would impudence their friends of friends and so on, and thus through the word-of-mouth effect a large population in the social network would adopt the application.

PROBLEM DEFINITION

In our daily life when we are going to buy a product, we take opinion of our friends and we are influenced by them in our decision making to buy the product. Opinion is the view or judgment about something. Most recent researches exploited that seed word that are unique, and lead to low robust. Online information is becoming progressively dynamic and the emergence of online social media and user generated content further results in experience. It is hard for a person or an organization to get the latest trends and outline the general opinions about products due to the huge diversity and size of social media that builds the need of automated checking and real time opinion extraction and opinion mining.

But due to fake reviews and easy editing of reviews given by a person, result in the drop out on sale of the product. So many best products are not yet identified in markets. This results in loss of money and standard of starting companies. The security of user reviews is not assured this leads to editing of review and some unethical work. By using fake review customers are easily attracted and results in fake product. Originality of a product is not established. Customer satisfaction and quality of product is the main reason for buying a good product. But in some scenarios the products failed to meet up the customer needs by analyzing the fake reviews.

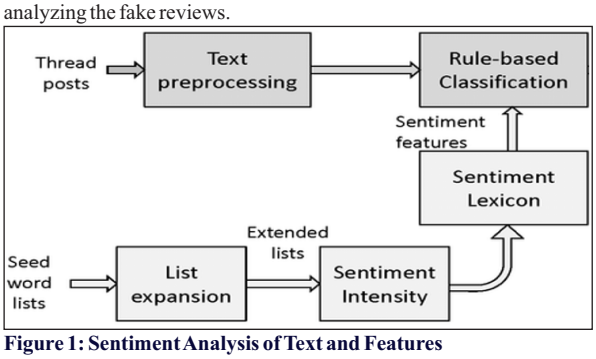


Figure 1: Sentiment Analysis of Text and Features

PROPOSED SYSTEMS

The explosion of user made viral advertisements in online websites and social media lead to many new opportunities in the marketing field with a huge explosion of good products. The PCA method will be exploited to extract the product features and can be used to extract the opinion-based features for obtaining of efficient neighbor's product ratings. In this process, the individual products have been categorized according to the best fit neighboring product ratings with unique index ID.

Sentiment Analysis is an example of machine learning. Using sentiment analysis the emotion of the review is obtained from the text. This method uses some special algorithms like T-SNE (T-distributed Stochastic Neighbor Embedding), PCA (Principal Component Analysis) and dimensionality reduction are used to recommend the product for users. Dimensionality Reduction is used to reduce the number of dimensions of the products which are already classified by

feature selection and feature extraction methods. Feature Extraction is a process of dimensionality reduction by which an initial data is reduced to manageable groups for processing.

Feature Selection process which automatically selects those features that contribute to our prediction or output which you are interested in. When using feature selection it is necessary to separate the relevant features in our data, having irrelevant features decrease the accuracy of models. The reviews and features of product are classified by feature selection and extraction process. Then dimensionality reduction is used to group the classified data set in a single dimension. Product recommendation generates the suggestion for items or content a specific user would like to purchase.

Then the rating for product is obtained from the customer and stored as a data set. Viral marketing is made followed by this process, which is based on rating and features. The products are influenced by the employees of the company in social media and friends of employees and friends of friends. This makes the product with good selling rate and more profit for company. Fake products and money wastage can be avoided by this project.

IMPLEMENTATION

SENTIMENT ANALYSIS

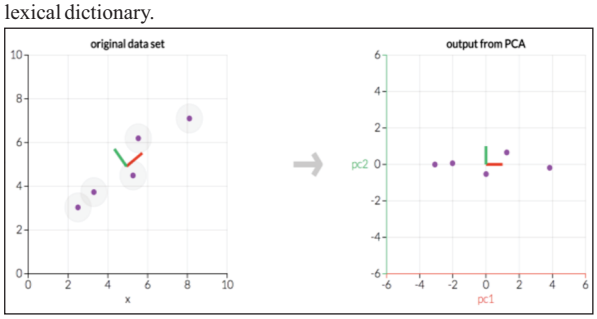
Sentiment analysis being a multifaceted problem has a lot of economic stakes if the results are better for the desired objectives. As we have studied the literature in this area, there is one thing that is clearly understood and expressed in that there are many open-ended problems at different strata. In our proposed work, we have used the dependency information, which is a relation between words in a dependency tree for a sentence parsed by a syntactic parser and established that features derived from the dependency relation produces better results. The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral. Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

PRODUCT RECOMMENDATION

Given a set of symbol sequences, a fixed-length vector representation for each symbol can be learned in a latent space by exploiting the context information among symbols, in which “similar” symbols will be mapped to nearby positions. If treat each product ID as a word token, and convert the historical purchase records of a user into a time stamped sequence, can then use the same methods to learn product embeddings. Unlike matrix factorization, the order of historical purchases from a user is naturally captured. User may be recommended with best product with best price to get good review.

PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. PCA is a best choice for dimensionality reduction and visualization for datasets with a large number of features. Using PCA algorithm has to classify individual products using consumer ratings. For example here going to classify any one product for mining the process through SVM based on lexical dictionary.



DIMENSIONALITY REDUCTION

Dimensionality reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables. Approaches can be divided into feature selection and feature extraction. Finally using both PCA and T-SNE (T-distributed Stochastic Neighbor Embedding) algorithm we can isolate individual products and randomly selecting any one of the product with unique index ID. Finding a best fit of that product neighboring using Cosine and Euclidean with accuracy through reducing high dimensional data. Similarly the process repeated for all the individual products for finding accuracy.

CONCLUSIONS

This project mainly focuses on making a viral marketing of the new products which is very good in quality and specification. It reduces the fake products selling in online and increase the selling rate of customers loved products. It maximizes the profit rate of a company by increasing its sales rate. It maintains a good relationship between marketing agents and customers and the product developer. Accurate segmentation is used so it is easy to make correct decision. Lexicon analysis makes the measurement of sentiment level makes the accurate classification of products. This project makes profit for the product developer, marketing agencies and the companies. In future the project can be further extended.

REFERENCES:

[1] W. Chen, C. Wang, and Y. Wang, “Scalable influence maximization for prevalent viral marketing in large-scale social networks,” in Proc. 16th ACM SIGKDD International Conf. on Knowledge Discovery Data Mining, Jul. 2010, pp. 1029–1038.
[2] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing,” in Proc. 8th ACM SIGKDD International Conf. on Knowledge Discovery Data Mining, Jul. 2002, pp. 61–70.
[3] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in Proc. 9th ACM SIGKDD International Conf. on Knowledge Discovery Data Mining, Aug. 2003, pp. 137–146.