Eleyon SOUTH ASIAN JOURNAL OF ENGINEERING AND TECHNOLOGY

Full Length Article

E-Commerce Recommendation over Big Data Based on early reviewers for effective product marke ting Prediction Rates

V.S.Suresh Kumar[a], A.Vijaya Rao[b], V.Vijay[b], D.Nagarjun[b], G.Thangavel[b]

a, Assistant professor,Department of computerscience, Nandha collage of technology, Erode-638052, Tamilnadu, India
b,Department of computerscience, Nandha collage of technology, Erode-638052, Tamilnadu, India
b,Department of computerscience, Nandha collage of technology, Erode-638052, Tamilnadu, India
b,Department of computerscience, Nandha collage of technology, Erode-638052, Tamilnadu, India
b,Department of computerscience, Nandha collage of technology, Erode-638052, Tamilnadu, India

***Corresponding Author**
sureshkmar@nandhatech.org
(V.S.Sureshkumar)
Tel.: +91 8056552243

DOI:

ABSTRACT: Online reviews is the important source of information for users before selecting a product or making a decision. Early reviews of a product tend to have a high impact on the subsequent product sales.In this paper we study the behavior characteristics early reviewers through their posted early reviews.At first we divided the product lifetime into three stages (Early, majority and laggards).A person who post a reviews in early stage is consider as early reviewers. The Early reviewers are the first one who respond to the product at the beginning stage. We quantitatively characterize early reviewers based on their rating behaviors. We use k-means with PageRank to predicting the early reviewers.

Keywords: E-commerce; reviews; feature identification; opinion mining.

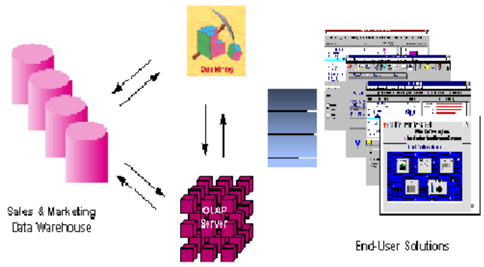
Introduction

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors,allowing business to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. The Foundations of Data Mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature: Massive data collection Powerful multiprocessor computers Data mining algorithms Commercial databases are growing at unprecedented rates.

1.How Data Mining Works

How exactly is data mining able to tell you important things that you didn't know or what is going to happen next? The technique that is used to perform these feats in data mining is called modeling. Modeling is simply the act of building a model in one situation where you know the answer and then applying it to another situation.That you don't. For instance, if you were looking for a sunken Spanish galleon on the high seas the first thing you might do is to research the times when Spanish treasure had been found by others in the past. You might note that these ships often tend to be found off the coast of Bermuda and that there are certain characteristics to the ocean currents, and certain routes that have likely been taken by the ship's captains in that era. You note these similarities and build a model that includes the characteristics that are common to the locations of these sunken treasures. With these models in hand you sail off looking for treasure where your model indicates it most likely might be given a similar situation in the past. Hopefully, if you've got a good model, you find your treasure.

1.1 An Architecture for Data Mining:



1.2 Related work

In this section, the details of the proposed system are going to be present. In fig.2. The flow chart is describing the overview of our proposed system. Firstly we are going to collected all the reviews of the consumer from those reviews the aspects are to be identified and opinions are collected and then data preprocessing is done to remove all the noisy words from the collected opinions. After data gets classified by using data classification, the most ranking products are to be collected according to term frequency and opinions collected. Simultaneously are going to get the best rated product.In this algorithm we combine the three techniques.

a. Frequency based method

Frequency-based method is the method which is used in our aggregate ranking algorithm, in which it gives the features according to term frequency of the product .This method takes only the frequency of the term and which will impact on the customer opinions on the particular product ,it helps in rating the product. There are some usual features of the product will appear frequently those are consider as the important features.

b. Corelation- based method

Correlation-based method, which measures the correlation between the reviews on particular products and the final rankings. It ranks the aspects based on the number of cases when such two kinds of opinions are consistentCorrelation-based method ranks the aspects by simply counting the consistent cases between reviews on particular products and the final rankings. It ignores to model the uncertainty in the generation of overall ratings, and thus cannot achieve satisfactory performance.

c. Hybrid method

Hybrid method , that captures both aspect frequency and the correlation The hybrid method simply aggregates the results from the frequency-based and correlation-based methods, and cannot boost the performance effectively.

1.3Advantages

By aggregating these things we can achieve the high accuracy and efficiency and we can classify the items in efficient manner. We are going to give the highest ranking product directly without reading all the reviews.

2.Existing system

The previous methods are using the fuzzy Kano model or traditional navel margin-based methods to get the rating from the users.It using the sentiment collaborative filtering algorithm in fuzzy Kano model. By using this model they display the different types of smiley photos to the users to express their reviews.Based on this method the product reviews are calculated.

3.Proposed system

To overcome the problems in the previous versions the text based reviews are introduced.In this paper the reviews are divided into three stages (Early, majority and laggards).The person who posting the review in the first stage is said to be Early reviews.This Early reviews are the important for the product success and failure.Byanalyzing the early reviewers by using their reviews we can predict the product future.K-means with PageRank is used to analyzing the early reviewers.

4. Conclusions

In this paper, we have studied the novel task of early reviewer characterization and prediction on two real-world online review datasets. Our empirical analysis strengthens a series of theoretical conclusions from sociology and economics. We found that (1) an early reviewer tends to assign a higher average rating score; and (2) an early reviewer tends to post more helpful reviews. Our experiments also indicate that early reviewers' ratings and their received helpfulness scores are likely to influence product popularity at a later stage. In our current work, the review content is not considered. In the future, we will explore effective ways in incorporating review content into our early reviewer prediction model. Currently, we focus on the analysis and prediction of early reviewers, while there remains an important issue to address, i.e., how to improve product marketing with the identified early reviewers. We will investigate this task with real e-commerce cases in collaboration with e-commerce companies in the future.

References

[1] J. McAuley and A. Yang, "Addressing complex and subjective product-related queries with customer reviews," in WWW, 2016, pp. 625–635.

[2] N. V. Nielsen, "E-commerce: Evolution or revolution in the fast- moving consumer goods world," nngroup. com, 2014.

[3] W. D. J. Salganik M J, Dodds P S, "Experimental study of in- equality and unpredictability in an artificial cultural market," in ASONAM, 2016, pp. 529–532.

[4] J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Image- based recommendations on styles and substitutes," in SIGIR, 2015, pp. 43–52.

[5] K. Sarkar and H. Sundaram, "How do we find early adopters who will guide a resource constrained network towards a desired distribution of behaviors?" in CoRR, 2013, p. 1303.

[6] D. Imamori and K. Tajima, "Predicting popularity of twitter ac- counts through the discovery of link-propagating early adopters," in CoRR, 2015, p. 1512.

[7] X. Rong and Q. Mei, "Diffusion of innovations revisited: from social network to innovation network," in CIKM, 2013, pp. 499–508.

[8] B. Liu, Sentiment Analysis - Mining Opinions, Sentiments, and Emo- tions. Cambridge University Press, 2015.

[9] A. A. Saurabh Kataria, "Distributed representations for content- based and personalized tag recommendation," in ICDM Workshop- s, 2015, pp. 1388–1395.

[10] S. L. G. X. Dongjing Wang, ShuiGuang Deng, "Improving mu- sic recommendation using distributed representation," in WWW, 2016, pp. 125–126.

[11] P. Wang, J. Guo, Y. Lan, J. Xu, S. Wan, and X. Cheng, "Learning hi- erarchical representation model for nextbasket recommendation," in SIGIR, 2015, pp. 403–412.

[12] S. Chen and T. Joachims, "Modeling intransitivity in matchup and comparison data," in WSDM, 2016, pp. 227–236.

[13] S. Usami, "Individual differences multidimensional bradey-terry model using reversible jump markov chain monte carloalgorith- m," Behaviormetrika, vol. 37(2), p. 135155, 2010.

[14] T. Liu, Learning to Rank for Information Retrieval. Springer, 2011.

[15] Z. Cao, T. Qin, T. Liu, M. Tsai, and H. Li, "Learning to rank: from