

A Novel Data-Mining Approach for Review Analysis of Products Using Web Application

Vijayaragavi.V¹, Vishnupriya.R², Suresh.M³

^{1, 2, 3} UG SCHOLAR, Department of Computer Science, Arjun college of Technology, Coimbatore

Abstract: A network based shopping is a trending business that has increased the rate of the frequenter's interest of shopping. This technology facilitates the user to buy their products over the internet. The main thing to be considered for a person who is willing to buy a product online is to analysis the review that is being given by other customers. The Data Mining technique is used for fetching and analysis of a product reviews from various online websites.

I. INTRODUCTION

The present system of review has many problems such as feedback calculation is not accurate, not convenient for visualization and limited opportunities for the user to discuss their experience. To overcome our approach relies on word frequency statistics and employs an exploratory analysis stage based on Self Organizing Maps. Subsequent modelling using network analysis methods was used to determine influential users among the forum members. These findings can open new avenues of research into a speedy data gathering, reviews, and analysis that provides more featured outputs and solutions for public health and important feedback for the manufacturer.

II. RELATED WORKS

Creating centralized web site where the care is taken to provide the registration page for user and admin. Admin will have to add product and also will be able to view all the necessary detail, will ensure that all reviews can be used in the accuracy for the particular product. User privilege is also controlled by the admin.

III. COMPETITIVE ADVANTAGE OF THE SYSTEM

- A. Social media, from private messaging to mass Medias and networks, is providing unbound chances for patients to discuss their experiences with drugs and devices.
- B. Our approach relies on word frequency statistics and employs an exploratory analysis stage based on Self Organizing Maps.
- C. Vector quantization algorithms reduce the feature space's size without losing information for identifying clusters in the classification space

IV. PROBLEMS WITH THE CURRENT SYSTEM

Past studies have used computational approaches to for sentiment extraction of user opinion. Further literature searches have revealed that companies have scoured social media platforms to assess user sentiment. While the data-mining literature is extensive, none have identified influential users, and how forum relationships affect the opinions and behaviors of other users.

- A. Feedback calculation is not accurate.
- B. Not convenient for user visualization.
- C. Limited opportunities for users discuss their experiences.

V. PROBLEMS WITH THE CURRENT SYSTEM

Past studies have used computational approaches to for sentiment extraction of user opinion. Further literature searches have revealed that companies have scoured social media platforms to assess user sentiment. While the data-mining literature is extensive, none have identified influential users, and how forum relationships affect the opinions and behaviors of other users.

- A. Feedback calculation is not accurate.
- B. Not convenient for user visualization
- C. Limited opportunities for users discuss their experiences.

VI. SALIENT FEATURES OF THE APPLICATION

A. *Data Mining and Text Mining*

Data mining aims to analyze a set of given data or information in order to identify novel and potentially useful patterns. Text mining aims to extract useful knowledge from textual data or Documents

B. *Machine Learning and Data Analysis Paradigms*

Machine Learning and Data Analysis Paradigms have been designed to acquire knowledge manually from human experts, which can be a very time-consuming and labour intensive process. To address this issue, machine learning algorithms have been refined to achieve knowledge automatically from examples or source data.

C. *Hybrid Approach*

The boundaries between the different paradigms are usually unclear and many systems have been built to combine different approaches. It is not surprising to find that many, data mining, and text mining systems adopt such a hybrid approach.

D. *Evaluation Methodologies*

The accuracy of a learning system has to be checked before it becomes useful. Bounded limitations of data often makes estimating accuracy a difficult task. Deciding a good evaluation methodology is very imperative for machine learning systems development

E. *Accuracy*

The level of precision in the proposed system will be elevated as there is a provision of the automated accounts for each user. All functions would be performed accurately and this ensures that the information coming from the server is accurate.

F. *Reliability*

The reliability of the proposed system is be high because the above mentioned reasons. The reason for the increased reliability of the system is that now there would be proper storage of information.

VI. SYSTEM ARCHITECTURE

A. *Objective*

Subsequent modeling using network analysis methods was used to determine influential users among the forum associates. These findings can open advanced avenues of scrutiny into rapid data collection, feedback, and analysis that can enable improved outcomes and solutions for public health and important feedback for the manufacturer.

B. *Methodology*

Creating centralized web site where the care is taken to analysis the reviews of the product from several online shopping websites. Each product will be rated according to the review. Creating the planned modules that are admin panel which includes user registration, review gathering, review analysis, etc

C. *Work Plan*

To check the possible reviews that has been got from the customers and also to do the transaction for the given concern. Where it mainly uses the web interface to get the output.

VII. MODULE DESCRIPTION

A. *Administrator*

The administrator plays a major role in customizing web services and thereby helping users to gather information or purchase products according to their own needs. The admin can login to the website using his username and password. Once the admin has logged in, he can add, delete or update several products and information needed by the registered visitors of the website .The web services include online book purchase, online shopping of mobile phones, online information, and education counseling information providers.

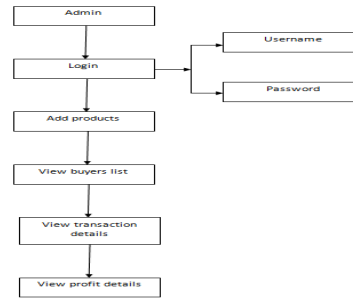


Fig 8.1

B. User Profile Creation

Users can generate an account in the website by providing various specifics such as username, password, name, age, gender, date of birth, income status, caste category etc... Once after the successful account creation he/she can login to the website using his username and password. The personal details provided by the user are crucial for online shopping. When the user searches a product such as a book or mobile phone, the search results will be displayed. The users can also edit their profile information and update the database.

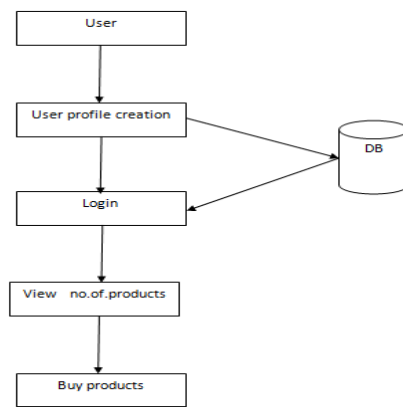


Fig 8.2

C. Purchase/ Feedback

The last module of this project is purchase/feedback. When the customers searching for a product are satisfied with the search result, they can make payment through online using credit card by clicking on the product. The click event will take them to the payment gateway. The customers can add description as feedback.

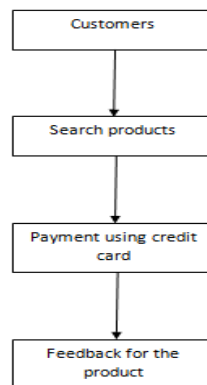


Fig8.3

D. Risk Management

Risk Management is another important and challenging module in this project. In this module based on the user feedback, we can extract the positive and negative feedback words for estimating particular feedback are either positive or negative or intermediate category. We can see the all user feedback for all products. The accuracy of the prediction depends on the information a user provides at the stage of feedback. Data mining technology can generate new business opportunities. Data mining automates the process of finding predictive information in large databases. The risk management is categories 3 ways in the project.

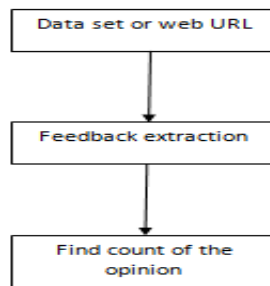


Fig 8.4

IX. SURVEY OF EXISTING SYSTEMS

Past studies have used computational approaches to for sentiment extraction of user opinion. Corley et al. analyzed how influenza posts by users in social media correlated to patient reporting data.

- A. Feedback calculation is not accurate.
- B. Not convenient for user visualization.
- C. Limited opportunities for users discuss their experiences.

X. TECHNOLOGIES USED

A. Java servlet

A Java servlet is a Java program that extends the capabilities of a server. Although servlets can react to any types of appeals, they most frequently implement applications presented on Web servers. *Java Servlets* are server-side *Java* program modules that process and answer client appeals and implement the *servlet* interface. It aids in enhancing Web server functionality with minimal overhead, maintenance and support. A *servlet* acts as an intermediary between the client and the server.

B. JavaScript Object Notation

JavaScript Object Notation is a text-based open Standard designed for human-readable data interchange which is derived from the JavaScript scripting language; it is a language for representing simple data structures and associative arrays, called objects.

C. Graph partitioning algorithms:

The current model of a social network to model future connections within the network. Viral marketing uses word-of-mouth effect by measuring the interactions among customers and carefully marketing individuals with the most social connections. Newsgroups discussions take advantage of 'response' relationships based on how often people respond to messages they agree (or disagree) with using graph partitioning algorithms.

- 1) *Vector quantization algorithms*: Vector quantization algorithms reduce the feature space's size without losing information for identifying clusters in the classification space.
- 2) *Stemming algorithm*: A stemming algorithm is a procedure of linguistic normalization, in which the variant forms of a word are shortened to a common form, for example,

connection
connections

connective ---> connect
connected
connecting

It is significant to acknowledge that we use stemming with the intention of improving the performance of IR systems. It is not an exercise in etymology or grammar. Literally from an etymological or grammatical point of view, a stemming algorithm is liable to make many mistakes. In addition, stemming algorithms - at least the ones presented here - are pertinent to the written, not the spoken, form of the accent or the language .

For some of the world's languages, Chinese for example, the concept of stemming is not applicable, but it is certainly significant for the diverse languages of the Indo-European group. In these languages words favor to be stable at the front, and to vary at the end:

-ion
-ions
connect-ive
-ed
-ing

The variable part is the 'ending', or 'suffix'. Removing these endings off is called 'suffix stripping' or 'stemming', and the residual part is called the stem.

XI. FUTURE SCOPE

The application discussed above is a humble effort to bring more effectiveness to the whole system of Online shopping and reviewing of a product. Yet there is a scope for modification and up gradation in the future.

XII. CONCLUSION

To conclude, this paper successfully created a new design for A Novel Data-Mining Approach for Review Analysis of a Product Using Web Application. In this web application user can perform gathering of reviews about the particular product and also analysis the reviews and give the appropriate review to the user. We hope that the web application will provide comfort for the user as to which all the details is being provided to the user.

REFERENCES

- [1] Ochoa, A. Hernandez, L. Cruz, J. Ponce, F. Montes, L. Li, and L. Janacek. "Artificial Societies and Social Simulation Using Ant Colony, Particle Swarm Optimization and Cultural Algorithms," New Achievements in Evolutionary Computation, Edition of book, Vol. , P. Korosec, , Ed : , p. 267-297, 2010.
- [2] W. Cornell and W. Cornell. (2013). How Data Mining Drives Pharma: Information as a Raw Material and Product [Webinar]. Available: <http://acswebinars.org/big-dat>
- [3] L. Toldo, "Text Mining Fundamentals for Business Analytics," presented at the 11th Annual Text and Social Analytics Summit. Boston, MA, 2013
- [4] L. Dunbrack. "Pharma 2.0 – Social Media and Pharmaceutical Sales and Marketing," in Health Industry Insights, 2010, p.
- [5] C. Corley, D. Cook, A. Mikler, and K. Singh. "Text and Structural Data Mining of Influenza Mentions in Web and Social Media," Int. J. Environ. Res. Public Health, Vol. 7, 596-615, Feb. 2010
- [6] L. Getoor and C. Diehl. "Link mining: a survey," SIGKDD Explor. Newsl., vol. 7, pp. 3—12, Dec. 2005
- [7] Q. Lu. And L. Getoor, "Link-based Classification." In Proc. Of the 20th Int. Conf. on Machine Learning (ICML). Washington, D.C., 2003, pp. 496-503
- [8] A. Ng, A. Zheng, and M. Jordan, "Stable algorithms for link analysis," in Proc. of the SIGIR Conf. on Information Retrieval. New Orleans, Louisiana, 2001, pp. 258-266
- [9] B. Taskar, M. Wong, P. Abbeel, and D. Koller, "Link Prediction in Relational Data," in Advances in Neural Information Processing Systems (NIPS), Vancouver, B.C., 2003
- [10] D. Liben-Nowell and J.M. Kleinberg, "The link prediction problem for social networks,"Journal of the American Society for Information Science and Technology, Vol. 57, pp. 556-559, May 2007
- [11] Z. Lacroix, H. Murthy, F. Naumann, and L. Raschid, "Links and Paths through Life Sciences data sources," in Proc. of the 1st Int. Workshop on Data Integration in the Life Sciences (DILS), Leipzig, Germany., 2004, pp. 203 211
- [12] J. Noessner, M. Niepert, C. Meilicke, and H. Stuckenschmidt, "Leveraging Terminological Structure for Object Reconciliation" in The Semantic Web: Research and Applications, Heidelberg, Berlin: Springer, 2010, pp.334-348
- [13] M.E.J. Newman, "Detecting community structure in networks," European Physical Journal, vol. 38, pp. 321-330, March 2004.
- [14] J. Huan and J. Prins, "Efficient Mining of Frequent Subgraphs in the Presence of Isomorphism," in Proc. Of the 3rd IEEE Int. Conf. on Data Mining (ICDM'03), Melbourne, Florida. 2003, pp. 549-552
- [15] D. Hand, "Principles of Data Mining," Drug Safety, vol. 30, pp. 621-622, July 2007