

Online Review Mining for Forecasting Sales

¹C.Pandi, ²K.Suresh Babu

¹ PG Scholar Department of Computer Science, St.Peter's University, Chennai, India, pandic24@gmail.com

² Assistant Professor, Department of Computer Science, St.Peter's University, Chennai, India,
sureshstudy2010@gmail.com

Abstract

The growing popularity of online product review forums invites people to express opinions and sentiments toward the products. It gives the knowledge about the product as well as sentiment of people towards the product. These online reviews are very important for forecasting the sales performance of product. In this paper, the online review mining techniques discussed commercial product. Sentiment PLSA which is responsible for finding hidden sentiment factors in the reviews and ARSA model used to predict sales, performance. An Autoregressive Sentiment and Quality Aware model (ARSQA) also in consideration for to build the quality for predicting sales performance. This paper proposes clustering and classification based algorithm for sentiment analysis.

Keywords: PLSA, ARSA, ARSQA

1. INTRODUCTION

In this paper, automatic data mining technique for predicting sales performance on a product quality domain. I tackle the problem of mining reviews for predicting product sales performance. The product feedback was got from the manufacturer, admin and the user who used the products and the admin automatically re-ranks the quality of the product using the feedbacks given. This makes the product user's to determine the real quality of the product and they will be satisfied.

The manufacturer initially gives their feedback to the main admin, and then the main admin will Re-rank according to the quality which displayed to the new user's. Then the new user's feedback is also obtained in the text format so that the admin will again Re-rank the product from the users point.

To upload the product in the Admin, the manufacturers have to create an account with the admin. To create an account the manufacturers have to provide their username, password and product id, rate of the product, expiration data etc... This information is stored in the database of the admin. Once the manufacturer creates an account they are able to upload their product in the admin.

Once the Manufacturer uploads the product in the admin, the product will be displayed in the admin. Also all the manufacturers and their products information will be stored in the admin. Also the admin will store all the records of the manufacturers they are accessing the admin.

Once the manufacturer uploads the products information in the admin, they will give the feedback about their product. Mostly these are not trustable. So the admin will check the quality of the product and leave the feedback regarding the product. This feedback will be trustable because the feedback is given by the admin and not by the manufacturers. So that the ranking is applied and the product will be listed as per the admin's feedback.

Once the user bought the product and leaves the feedback regarding the product. Once the server got the feedback from the user, the server will re-rank the product based on their feedback. After it was re-ranked, the server will update the product quality as per feedback.

If the user search of the product deals generally, the admin will display the deals of the day for the particular product from the different manufacturers. So that the user can see the deals and purchase the products by using the deals.

ARCHITECTURAL DIAGRAM



Fig.1 Architecture Diagram

2. RELATED WORK

2.1 Online Review Mining

Review Mining is one of the growing mining sectors. It is very predictive for analysis review. Many online blog and social networking sites are available, where many people are expressing their review with respect to product and movie. If we considering these reviews then it is very helpful to increase sales performance.

2.2 Domain Driven Task

Domain-driven data mining generally targets actionable knowledge discovery in complex domain problems. It aims first to utilize and mine many aspects of intelligence for example, in-depth data, domain expertise, and real-time human involvement as well as process, environment, and social intelligence. Domain-driven data mining works to expose next generation methodologies for actionable knowledge discovery, identifying how KDD can better contribute to critical domain problems in theory and practice. It uncovers domain-driven techniques to help KDD strengthen business intelligence in complex enterprise applications. Domain Driven task is categorized into three level. These three are Human Intelligence, Domain intelligence, Network intelligence.

2.3 Human Intelligence

In this domain driven task, we considered number of blog. These blogs are considered with opinions of use. Opinions of user or information of other thing is posted on blog. So blogs are the way to express opinion of people. So this is one way to get the people sentiment analysis.

2.4 Knowledge discovery from database

It is a part of KDD Knowledge discovery from database. We discover knowledgeable data by mining raw data. raw data is considered as input to the system. By using this data apply data mining algorithm so that knowledge is discovered from it. This knowledge is of the pattern or a collective analysis report. This is very important for business to increase profit of organization. Here actionable knowledge is considered to the knowledge discovery. AKD is an iterative optimization process toward the actionable pattern, considering surrounding business environment and Problem states.

2.5 Clustering and Classification

Clustering can be considered the most important unsupervised learning problem, so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be “the process of organizing objects into groups whose members are similar in some way”. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters. So our problem solution is not using kind of supervised learning. By using clustering, the records of the product data is collected simultaneously and made available for analysis.

3. S PLSA

Sentiment Probabilistic Latent Semantic Analysis in which a review can be considered as being generated under the influence of a number of hidden sentiment factors. For the said purpose numbers of blogs are available. Separating number of blogs from other blog is very tedious task. Categorize each one blog and consider it only those blog which is relating with product review. If we considering word review pair then each word expressing the positive and negative comments. Initial task is to separate the sentiment in different categories like positive, negative, average. Number of other word is present in to the sentiment .those unnecessary words are increasing calculation. Optimization is done to illuminate unnecessary calculation. The people can observe their opinions and increase their feel to use the product if the reviews are good. If the product is bad then the reviews are bad then it will become serious impact for the product. Reviews posted in online are so important so that it is directly affecting the product’s sales and the bad reviews about the movie failed to get the right place in the rank in the product database

Proposed Work

In present day’s online purchasing has become very popular. When the product is ordered by seeing its specification and the look the user may not be aware whether it is of good quality or not. The websites which exists also propagates based on the manufacturer views.

In this paper the website the aggregates and ranks the products based on the user reviews and feedback. The admin of this site receives both manufacturers and users feedback information based on the product rank, which is very useful for the future users to buy the products of their desire and quality.

Experimental Results

By doing this type of experiments the best quality products are detected in all category with proof i.e. after the usage of products only the customer will provide their feedbacks. So it will be easy to find out the good quality products without wasting our time in comparison.

CONCLUSION

The increasing use of online reviews as a way of conveying views and comments has proved a unique way to find sales performance and derive business intelligence. In this paper, the study shows the problem of predicting sales performance using sentiment information mined from reviews. The outcome of this generates knowledge from mined data that can be useful for forecasting sales. S-PLSA is useful for analysis of sentiment that helps us to classify different categorization of sentiments in blogs. By using ARSA, we can easily predict sales performance.

REFERENCES

- [1] Mukherjee, A., Liu, B., Wang, J., Glance, N., Jindal, N. Detecting Group Review Spam. Dept of CS. Technical Report, UIC, 2011
- [2] Bing Liu, Junhui Wang, Natalie Glance and Nitin Jindal. Detecting Group Review Spam . WWW 2011, March 28– April 1, 2011, Hyderabad, India.
- [3] Lim, E.P., Nguyen, V.A., Jindal, N., Liu, B., and Lauw, H. Detecting product review spammers using rating behavior. CIKM, 2010.
- [4] N. Jindal, B. Liu, and E.-P. Lim. Finding unusual review patterns using unexpected rules. CIKM, 2010. [11] Siddu P. Algur, Amit P. Patil, P.S Hiremath and S. Shivashankar. Conceptual level Similarity Measure based Review Spam Detection. IEEE 2010.
- [5] Wei Jin, Hung Hay Ho and Rohini K. Srihari. Opinion Miner: A Novel Machine Learning System for Web Opinion Mining and Extraction. KDD’09, June 28– July 1, 2009, Paris, France. ACM 2009

- [6]. D. Gruhl, R. Guha, R. Kumar, J. Novak, and A. Tomkins, "The Predictive Power of Online Chatter," Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery in Data Mining (KDD), pp. 78-87, 2005.
- [7]. A. Ghose and P.G. Ipeirotis, "Designing Novel Review Ranking Systems: Predicting the Usefulness and Impact of Reviews," Proc. Ninth Int'l Conf. Electronic Commerce (ICEC), pp. 303-310, 2007.
- [8]. Y. Liu, X. Huang, A. An, and X. Yu, "ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs," Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 607-614, 2007.
- [9]. Y. Liu, X. Yu, X. Huang, and A. An, "Blog Data Mining: The Predictive Power of Sentiments," Data Mining for Business Applications, pp. 183-195, Springer, 2009
- [10]. L. Cao, C. Zhang, Q. Yang, D. Bell, M. Vlachos, B. Taneri, E. Keogh, P.S. Yu, N. Zhong, M.Z. Ashrafi, D. Taniar, E. Dubossarsky, and W. Graco, "Domain-Driven, Actionable Knowledge Discovery," IEEE Intelligent Systems, vol. 22, no. 4, pp. 78-88, July/Aug. 2007.
- [11]. L. Cao, Y. Zhao, H. Zhang, D. Luo, C. Zhang, and E.K. Park, "Flexible Frameworks for Actionable Knowledge Discovery," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 9, pp. 1299-1312, Sept. 2009.
- [12]. Xiaohui Yu, Jimmy Xiangji Huang, "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain" IEEE Trans. Knowledge and Data Eng., Vol. 24, No. 4, April 2012.
- [13] R. Mikut, M. Reischl, "Data Mining Tools," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol.1, no. 5, pp. 431-443, 2011.
- [14] A. Faro, D. Giordano, F. Maiorana, "Mining Massive Datasets by an Unsupervised Parallel Clustering on a GRID: Novel Algorithms and Case Study," Future Generation Computer Systems, vol. 27, no. 6, pp. 711-724, 2011.
- [15] F. Maiorana, G. Fazio, "Knowledge Discovery from Text on a Cloud Architecture and its Application to Bioinformatics," in Proc. 9th International Conference on Biomedical Engineering, IASTED, 2012.