

Mining Online Reviews in Websites for Predicting Sales Performance

¹Divagar.S, ²Joseph Raymond V

¹M.Tech Scholar, ²Assistant Professor

Department of Information Technology, SRM UNIVERSITY, CHENNAI, India

¹divagarkrr@gmail.com, ²joesphraymondv@gmail.com

Abstract— Nowadays people posting reviews are increasing and it gives us the knowledge to buy the products. People can express their sentiments through reviews posted in websites. By these reviews we can predict the movie sales performance. The question is how the sentiment factor can be analyzed in the reviews? Reviews are analyzed by Sentiment PLSA which helps us to find the unknown factors in the reviews and helps us to classify the reviews. Then we have to predict the sales performance, for the prediction regression is suitable way to do and we suggest an autoregressive sentiment (ARSA) Aware Model. To improve the prediction accuracy to enhance the best quality factor we also consider an autoregressive sentiment and quality aware model (ASQA) to build the quality for mining reviews sales performance in movie domain.

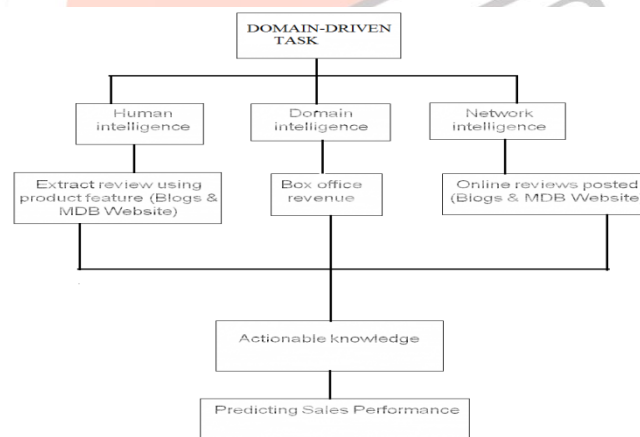
Index Terms— Keywords: OnlineReview mining, hidden sentiment analysis, prediction, S-PLSA, ARSA

I. INTRODUCTION

Nowadays people posting reviews are increasing and it gives us the knowledge to buy the products. People can express their sentiments through reviews posted in websites.

The actionable knowledge developed by using the average of number of quality the quality reviews presented and also the number of the people rated the reviews in the websites and mdb websites. The actionable knowledge is the last part and which can be developed by the base models and algorithms which is used to effectively predict the sales performance and which can be shared to all the peoples across the world. Predicting sales performance[1][2] is completely a domain driven task, it gives us to analyze the box office revenues. Such that the actionable knowledge can be developed by the sentiments and the quality reviews also plays as an important role here. Some wrong reviews also affect the prediction but only we are taking the whole reviews so it will not be impact the sales of the movie.

II. OVERALL PROCESS



III. RELATED WORK

A. Synthesize Human Intelligence

In this part of human intelligence we have to build blogs [4] and we have to post the reviews of the users dynamically. Any users can post the reviews in blogs and it will be an open source and the knowledge can be easily shared. Related work we need here is to collect all the data, relevant to the movie database and also it should be updated regularly so that all the people can post the reviews and it makes the sales performance to predict easily. Here we can set the reviews date and time and it becomes a factor for predicting past sales performance too. This human intelligence is also useful for actionable knowledge and which influences the sales of the movie. The reviews posted in the blogs will give us the intense details of people's opinion and it is further classified into positive and negative using SVM classifier. The opinions posted are collected as a bag of words which is useful for SVM classification.

B, Domain Intelligence

With the use of Domain Intelligence we develop the movie database or we can take the data from IMDB sites. In the MDB (movie database) sites we can update the box office movie data (revenue) and we can make use of the box office revenue for predicting sales performance of the movie by comparing with the previous past sales of the specific movie. In the MDB (movie database) site the admin will update the opening box office movies and also updates how much each movie getting revenue. The admin also updates the details of upcoming movies and people can also post their reviews about the movie. The Domain Intelligence plays an important role for predicting sales performance by giving valuable information like weekend revenue collection, gross revenue collection and also gives the information how many weeks the movie runs. Unlike blogs, in MDB sites only the registered users can post and rate the reviews there. The movie database site also helps us to observe opinions of the upcoming and ongoing movies which will act the reviews to know about the movies well.

C. S-PLSA

S-PLSA helps us to analyse the sentiments [7] presented in the reviews which helps in to understand the emotions of the reviewers. The people can observe their opinions and increase their feel to watch the movie if the reviews are good. If the movie is bad then the reviews are bad then it will become serious impact for the movie. Reviews posted in online are so important so that it is directly affecting the movie's sales and the bad reviews about the movie failed to get the right place in the box office in the movie database site. S-PLSA also finds the hidden sentiments in the reviews which is that some words does not give the meaning directly and it gives the polarity for the each and every review. In the existing system there was lack in classifying the review correctly because it failed to find the hidden sentiment factors present inside the review, LSA (Latent Semantic Analysis) [6] had been used in the existing system. The polarity of the review cannot be find in the existing system and so we approached a new system that we propose S-PLSA First we have to train the reviews in the backend of positive comments and negative.

For Positive product feature set calculation:

Positive product feature set = Number of positive words in the comments / Total number of words

For Negative product feature set calculation:

Negative product feature set = Number of Negative words in the comments / Total number of words

Then for the sentiment mining the comments or reviews have to be crawled from the blogs and MDB sites.

IV. ACTIONABLE KNOWLEDGE

In this work The actionable knowledge is the final part of predicting sales performance of the movie. After we finished training the reviews and classification of the reviews we further use ARSA [3] which is used to predict the movie sales by using the average of box office revenue, average of public sentiment, average of quality review. These three averages will give a percentage by which we can easily identify which movie is best. In ARSA we also can check the previous sales of the movie which really helps us to know more about the movie. The average of box office revenue is that we calculate how much tickets have been sold from the time of movie release and the average of public sentiment is calculated from the "x of y people have found this comment useful" and average of quality review is calculated by using the ranking of the reviews.

D. Results



Ongoing Movie : okok
Quality Review using S-PLSA
 Brief Summary
 Postive Comment
 No comments about Santhanam,
 he simply rocked in each and
 every scene
 Negative Comment
 udhayanidhi dancing was not
 good

Fig 1 Classification of Reviews using S-PLSA

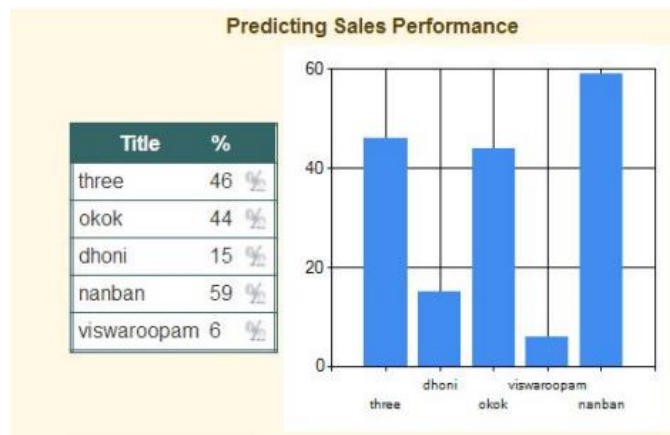


Fig 2 Predicting Sales Performance using ARSA.

V. CONCLUSION

In this project we have showed how the reviews are mined and how they are useful for predicting the sales performance in the market domain. With the use of S-PLSA we can mine the sentiments hidden inside the reviews and classify as positive and negative reviews using the polarity for the reviews. Thus S-PLSA is helpful for finding the quality of the review. After S-PLSA we use the crawling process to extract the reviews from the blogs and MDB sites and ARSA is used. With the use of ARSA we can easily predict the movie sales performance and we can produce the actionable knowledge from it.

VI. FUTURE EXTENSION

Predict sales performance in the other domains like consumer electronics, mobile phones and computers based on the user reviews posted in the websites like flipkart, snapdeal etc.

REFERENCES

- [1] D.Gruhl, R.Guha, R. Kumar, J. Novak, and A. Tomkins, "The Predictive Power of Online Chatter," Proc. 11th ACM SIGKDD Conf."Knowledge Discovery in Data Mining (KDD)", pp. 78-87, 2005.
- [2] 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
- [3] Y. Liu, X. Huang, A. An, and X. Yu, "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.
- [4] 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, pringer, 2009.
- [5] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information Diffusion through Blogspace," Proc. 13th Int'l Conf. World Wide Web (WWW), pp. 491-501, [6] 2004..
- [7] T. Hofmann, "Probabilistic Latent Semantic Analysis," Proc.Uncertainty in Artificial Intelligence (UAI), 1999.
- [8] C. Whitelaw, N. Garg, and S. Argamon, "Using Appraisal Groups for Sentiment Analysis," Proc.14th ACM Int'l Conf. Information and Knowledge management (CIKM), pp. 625-631, 2005.