INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

A SURVEY ON: PREDICTION ALGORITHMS FOR ONLINE PRODUCT RECOMMENDATION

[1]Archana Singh, [2]Avinash Sharma, [3]S. R. Yadav, [4]Pushpa Singh

Abstract: - As the number of internet users are increasing day by day. This tends for most of the market to move towards online shopping. People are investing much time on social networking sites, so that is the best place to analyze its behavior and show advertisement for the products of their interest. Some of website provides user rating for different product but they do not recommend any user to purchase. So bridging the gap between user social behavior and product requirement is done. In this paper a brief survey of product prediction algorithm is done. Paper has discussed different features for increase the efficiency of the work.

Keywords:-Data mining, Social network, Product rating, Product recommendation.

I. INTRODUCTION

As electronic media has included many things in daily life. One of its facilities is internet this reduced many work such as transfer of data in form of text, audio, video, etc. As numbers of users are increasing day by day, which attract different company to sale there products. This online market for advertisement and selling is playing very drastic role. There are different types of service provider in the market in order to attract users, so more the number of user more will be the brand value.

These regular visits of millions of customer at few points not only attract companies but many of researchers are also getting different field of research. One of most popular research is next node prediction, as most of social networking websites are doing this for making strong bonding between there users. Beside this some of sites are collecting reviews from different users from various social sites to come and share their experience for any product, movies, etc. Analyses of those comments or reviews are done by researcher for better understanding of the product market value is done by researcher. On the basis of this analysis if users product purchasing is predict by the system this is a new research area that is product prediction base on social networking sites. [2, 5, 9].

Some of websites like epinion, flixter, etc. are platform where user from different community come and share their view co-comment each other on different topics. By this website increase and decrease the rating of the product. So base on the same product a relation is develop between unknown users.

Many of Social networking sites like facebook, twitter, etc. are platform for users to maintain their social relation with each other by sharing message, voice, video, etc. with each other. A combination of both product review and social network sites is done in much work like [8, 10].

It is known that people buy those products that are suggested by some other person who has trust on it. As advertisement for particular community increase product selling if target community is correct. So product prediction work will identify those communities for advertising. As chance of purchasing that kind of product is high.

Objective

This work deal with the problem of predicting the purchase behaviors of social media users who have unknown history on an e-commerce website (cold start). More in detail, aim at predicting which product categories (e.g. electronics) the user will buy from by using solely information derived from the social network. Such a predictive system would help in several practical scenarios, including:

- Build a cold start recommender system, by providing high-level recommendations to social media users who connect for the first time to an e-commerce website.
- Improve existing product recommendation engines that can guide the recommender system to find domains of interest for the user.
- Provide e-commerce companies with tools for targeted social media campaigns.
- Correlate data from different domain websites like social network (Facebook) and reviewer (Epinion).
- Increase accuracy of product prediction system.

Proposed System use jaccard coefficients because it developed relation between neighborhoods. It commonly used similarity metric in information retrieval [24]|measures the probability that both x and y have a feature f, for a randomly selected feature f that either x or y has. While other prediction methods like regression, K-Nearest neighbor, develop relation among numeric data only. In Jaccard no discrete value is required as in Bayes classifier. So considering these facts jacccard coefficient values are used for summarization of trust. This trust is developing by social relation features, where bonding between users are identified by jaccard coefficient value.

II. RELATED WORK

Szabo et. al. [10] has proposed a model where six features are utilize for the ranking of the product. It was named as AFFrank which include feature like affinity rank history, product community size, evolution distance, member connectivity, social context and average rating.

Katz [2] in similar fashion, construct the probabilistic product prediction model. This model utilize social network as well as user recommendation, product acceptance as well as review of user for different similar products.

Freddy Chong et. al. [1] has proposed Social-Union, a method which combines similarity matrices derived from heterogeneous explicit or implicit SRNs. Moreover, an effective weighting strategy of SRNs influence based on their structured density. This work also generalizes our model for combining multiple social networks. This work perform an extensive experimental comparison of the proposed method against existing rating prediction and product recommendation algorithms, using synthetic and two real data sets (Epinions and Flixter).

B. Sarwar et. al. [6] has proposed, a CF Collaborative filtering model has been developed. Which utilize item-item similarity, instem of item user similarity It has been obtained that product recommendation base on item similarity is high.

Kleinberg et. al. [4] has evolved weight similarity matrix between the user and item for product recommendation. It is clear that once product gets high rating then chance of acceptance is also high. But this required manual work as recommendation rating need user choice.

J. Herlocker et. al. [3] has developed a graph which is of Unipart and bipart pattern. So model develops in this paper use product recommendation and social network dataset. This combination as highly increases the product recommendation accuracy.

J. Konstan et. al. [5] has review different researcher work on the various evaluation strategies, this paper present empirical results for the of accuracy metrics on one dataset. Metrics within each equivalency class were strongly correlated, while metrics from different equivalency classes were uncorrelated.

Huberman and others [2] studied the social interactions on Twitter to reveal that the driving process for usage is a sparse hidden network underlying the friends and followers, while most of the links represent meaningless interactions.
Java et al [7] investigated community structure and isolated different types of user intentions on Twitter.

Jansen and others [3] have examined Twitter as a mechanism for word-of-mouth advertising, and considered particular brands and products while examining the structure of the postings and the change in sentiments. However the authors do not perform any analysis on the predictive aspect of Twitter.

III. PREDICTION METHODS

In this section, discussion of some methods used in prediction:

Regression method: In this method combination of different inputs are done like from dependent and independent variable, then social network input, product rating, etc. As it is known that regression model is linear or non-linear. Here linear model play good role as compare to non-linear one. So prediction model need to be linear from non-linear model. Here in-order to analyze work like sentiment analysis prediction model be non-linear as in [9]. But in case of prediction base on numbers it will be linear. So analysis of early post and current post done by users are done at different sites. Those posts are collect from the sites and put in prediction model for analysis.

Bayes classifier: Bayes classifier is a probabilistic classifier using Bayes' theorem. Based upon the priori probability of the prediction event, Bayes classifier uses the Bayesian formula to calculate its posterior probability, that the object belongs to the result classes, and then select the class with the largest posterior probability, as the event is most likely to have that result. If the prediction result is discrete, the Bayes classifier can be applied directly. Otherwise, the prediction result must be discretized first [8]. This classifier has an assumption that the predictors

must be conditionally independent. There is no solid evidence always that the discussed metrics satisfy this assumption.

K-nearest neighbor classifier: K-nearest neighbor classifier, one of the simplest machine learning algorithms, tries to cluster the objects according to their distance to others. Commonly we use the Euclidean distance and Manhattan distance. For two entities p={$p1$,$p2$,…,pn} and q={$q1$,$q2$,…,qn} with n-dimensional feature vector. Then, the entity, which is being predicted, is assigned to the cluster including most of its k-nearest neighbors. Finally, the entity is predicted to have the same output as the entities in its cluster.

Artificial Neural network: Artificial neural network is a computational model [11] to simulate the human brain. An artificial neural network consists of lots of artificial neurons. And these neurons could belong to many interconnected group, including input layer, hidden layer and output layer. The input layer is responsible for receiving raw data and transmitting them to the next layer. The output layer will give us the final prediction result. Thus using artificial neural network to do prediction, our major task is choosing the network structure and designing the hidden layer. In addition to using them to predict directly, the Self Organizing Map (SOM), one kind of artificial neural network, could be used for features' dimensionality reduction for further analysis.

Decision tree: Decision tree is a visual technique in data mining and machine learning. Travelling from root node to leaf, one entity will get the prediction result. Classification tree and regression tree are two basic and major types of decision trees. Classification tree analysis is applied when the prediction output is discrete classes. And regression tree is used when predicted outcome is continuous value. Unlike the artificial neural network being a black box model, the decision tree is a white box model, which could be relatively easily explained. Besides, decision tree works well with dummy variables and empty variables.

Model based prediction: This possibly is the hardest way to do prediction. We have to build a mathematical model on the object before prediction, which requires deep insight into the object. At this point we do not know enough about social media to develop effective models for them. Even though there is some progress in modeling [7], model-based prediction remains an open and challenging topic.

IV. SOCIAL NETWORKS AND PURCHASE BEHAVIORS

Some research has investigated the broader topic of how social network influences users in their purchases. In [5] empirically demonstrate that a user's friends exercise \peer pressure": if friends widely adopt a product, the user is more likely to buy it. [11] Study the trading dynamics on the e-commerce social network Taobao. They show that buyers are more likely to purchase from sellers that friends in their network have already bought from (information passing). They prove that when a buyer has to decide from which seller to buy a product, the social network has a bigger influence on the decision than the seller's ratings and the price of the product.

As different online social network have different purpose so features of the website are also vary from network to network. Let us consider the facebook as the base of the online social networks. In this there are many different events like comment, like, tag, unlike, write on wall, etc. So event act as feature of the user. This can be understand as the when user like, comment, send message then link of the graph not generate, even when user send friend request also then also the new link is not generate between them. But once one user accept or conform the other user friend request then only new link of the graph of that online social network is develop between those user only.

Now the problem is what will be the feature for analyzing the behavior. For this one approach which is adopt in this paper is to make a dataset of the events generate by the user with the number of time it occur. Let us consider one example that Node = {U1, U2, U3….Un}, Link = {L1, L2, L3……..Ln}.

Recommendation on social media is a fairly new topic. Most work focuses on suggesting interesting content items (e.g. URLs, pictures, posts) or new friends. Social media recommender systems di_er from classical ones in that they often leverage existing social relations to boost the recommendations.

V. EVALUATION PARAMETER

In order to evaluate results there are many parameter such as accuracy, precision, recall, F-score, etc. Obtaining values can be put in the mention parameter formula to get better results.

$$\Pr ecision = \frac{True\_Positive}{True\_Positive + False\_Positive}$$

$$\mathrm{Re}call = \frac{True\_Positive}{True\_Positive + False\_Negative}$$

$$F\_Score = \frac{2 * \Pr ecision * \mathrm{Re}call}{\Pr ecision + \mathrm{Re}call}$$

In above true positive value is obtain by the system when the input article is in favour of B disputant and system also says that article is in favour of the B disputant. While in case of false positive value is obtain by the system when the input article is in favour of B disputant and system says that article is in favour of the C disputant.

VI. CONCLUSION

Mining World Wide Web has necessitated the users to make use of automated tools to locate desired information resources and to follow and assess their usage pattern. Web item pre- has been widely used to reduce the user confusion problem. Proposed model is mostly used for prediction in social network because of its high accuracy. It is a powerful method for arranging users' choices with there social relation into clusters according to their similarity. This survey helps to develops a techniques for overcoming the issues of web item prediction. However, research of the web item prediction is just at its beginning and much deeper understanding needs to be gained.

VII. REFERENCES

[1] Chong Tat Chua, Hady W. Lauw, and Ee-Peng Lim. "Generative Models for Item Adoptions Using Social Correlation". IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 9, SEPTEMBER 2013.

[2] Katz, Leo. (1953) A new status index derived from sociometric analysis. Psychometrika, 18(1):39-43.

[3] J. Herlocker, J. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proc. ACM SIGIR Conf.*, pages 230–237, 1999.

[4] Liben-Nowell, David, and Kleinberg, Jon. (2007). The Link Prediction Problem for Social Networks . Journal of the American Society for Information Science and Technology, 58(7):1019-1031.

[5] J. Herlocker, J. Konstan, L. Terveen, and J. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. on Information Systems*, 22(1):5–53, 2004.

[6] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proc. WWW Conf.*, pages 285–295, 2001.

[7] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and Passivity in Social Media," in Proceeding of the 22th international conference on World Wide Web - WWW '11, 2011, p. 113-114.

[8] R. Sharda and D. Delen, "Predicting box-office success of motion pictures with neural networks," Expert Systems with Applications, vol. 30, no. 2, pp. 243-254, Feb. 2006.

[9] W. Zhang and S. Skiena, "Improving Movie Gross Prediction through News Analysis," in 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, 2009, vol. 30, no. 2, pp. 301-304.

[10] G. Szabo and B. a. Huberman, "Predicting the popularity of online content," Communications of the ACM, vol. 53, no. 8, p. 80, Aug. 2010.

[11] S. Kak, Y. Chen, L. Wang, "Data Mining Using Surface and Deep Agents Based on Neural Networks," in Proceedings of the Sixteenth Annual Americas' Conference on Information Systems, 2010.