A Brief Literature Survey on: Online Product Purchasing on User Behavior

Monika Pal
Asst. Professor
St. Francis Institute Of Technology
Mumbai-400103

ABSTRACT

Online Social Rating Networks such as Epinions and Flixter, allow users to form handful constructive social networks, through their daily routine like recommending on the corresponding products, or similarly co-rating products. The preponderance of preceding work in Rating prognosis and Recommendation of products mainly takes into account ratings of users on products. However, in Social Rating Networks users can also construct their precise social network by reckoning each other as friends. In this paper, a perusal of different techniques for product prediction is generated.

General Terms

Collaborative Filtering, Jaccard-Coeffiecient, Latent Factors.

Keywords

Product predilection, cordial or social network, link prediction, Node Neighborhood, Item adoption..

1. INTRODUCTION

In our day to day life we come across choices and options. What to eat? Which movie to watch over the weekend? Which camera to buy? The decision making space is expanding as we progress. Without having previous knowledge of the domain space it is difficult to make a final decision. Hence people are dependent on advises or suggestions from their friends or the advice of experts.

Over the decade internet has added a new boon to everyone's life. There is large mass of data procurable on the web. The challenge of information overload is faced by both retailers and consumers. The retailers have started using algorithmic approach towards the information that is to be shown to the users. The retailers found this technique more practical. If a retailer displays more relevant content according to the consumer's perspective then a consumer is more likely to show interest in purchasing. In order to find solution to this problem the area of recommendation organization has emerged. Researchers have developed various algorithms and systems which are spreading in the world of E-Commerce by online vendors like Amazon.com, Ebay.com, and Netflix.com. For example, Netflix price always comes to mind in this context. Hence recommendation organizations have now become popular in the market and research community. In this work we propose a trust based recommendation system by incorporating the idea of social networking. Social impact plays an essential ingenue in human decision making process. The concern of trust in recommendation system is anatomized as the other users in our current user's network are his/her friends. Also, the current user can specify how much level of trust they grant on other users. This way the second concern of impudence level

on friend's rating as per your product predilection is also resolved.

2. RELATED WORK

Collaborative filtering occupying algorithms that use user-item rating undergo from sparse and asymmetry of rating data. Herlocker, et al. [5] put forward an exceptional overview of the goals, datasets, and breakthrough of collaborative filtering systems. This drove researchers to delve into other data sources that can be comprehended in such algorithms. Using eloquent data about an item was first explored by Balabanovic et al. [6]. Melville et al. [7] enhanced CF by using content of a movie, e.g., movie category. Breese et al. [15] used a Bayesian clustering model to group users placed on their ratings. Ungar and Foster [16] also used a Bayesian approach to group users based on their predilictions. Exactness in predilictions made recommendation system was augmented by Pazzani [8] by using hybrid methods both of user data (probability information) and product data (content). Hybrid systems which combine content and association have also been suggested in which various weights are determined on the addition of similarity [16]. Melville et al. [18] proposed a general framework for content-boosted Collaborative Filtering, where content-based predictions are applied to convert a sparse user ratings matrix into a full ratings matrix, and then a CF method is used to provide recommendations.

The applicability of trust to recommender systems has been established in several research studies. Ziegler and Lausen [12] showed that a correlation between trust and user similarity in an empirical study of a real online community. All Consuming one was an online community where users rate books. The authors showed that users were significantly more similar to their trusted peers than to the population as a whole. This work was extended in [22] which augmented the analysis of the All Consuming community and added an analysis. Within that community, results also showed a strong correlation between trust and similarity in movie ratings. The second result used the Film Trust system where users have stated how much they trust their friends in a social network and also rated movies. In this recommender system, trust is used in place of the Pearson correlation coefficient to generate predictive ratings. Results showed that when the users rating of a movie is different than the average rating, it is likely that the recommended rating will more closely reflect the user's tastes. The awareness towards building social relationships with the intention of making up social communities all the way through the Internet has been drastically increased in recent years. Social network based recommendation system are gaining popularity these days. The recommendations made by friends were showing better prediction results than the recommendations made by systems [10]. Several researchers have investigated exploring social networks and trust in

particular. This is due to the fact that human decision making process is primarily affected by peers having similar taste preferences[11].In an empirical study conducted by Ziegler and Lausen [12] showed a correlation between trust and user similarity. Abdul- Rahman and Hailes [13] showed that in a predefined context, such as movies, users develop social connections with people who have similar preferences. Walter et al. [4] propose the use of social network information in recommendation systems and analyze the impact of trust dynamics on the performance of such a system.

3. RESEARCH OBJECTIVE

In this work we propose a trust based recommendation system by incorporating the idea of social networking. Social impact plays an essential ingenue in human decision making process. The concern of trust in recommendation system is anatomized as the other users in our current user's network are his/her friends. Also, the current user can specify how much level of trust they grant on other users. This way the second concern of impudence level on friend's rating as per your predilection is also resolved. The recommendation system proffered in this research paper is Social Network based Recommendation System.

3.1 Social Networking Features

Choosing an appropriate feature set is the most critical part of any machine learning algorithm. For link prediction, each data point corresponds to a pair of vertices with the label denoting their link status, so the chosen features should represent some form of proximity between the pair of vertices. In existing research works on link prediction, majority of the features are extracted from the graph topology. Also, some works develop a feature set constructed from a graph evolution model. Besides these, the attributes of vertices and edges can also be very good features for many application domains.

The features that are based on graph topology are the most natural for link prediction. Here call them graph-topological feature. In fact, many works [9, 1] on product prediction concentrated only on the graph topological feature-set. Typically, they compute the similarity based on the node neighborhoods or based on the ensembles of paths between a pair of nodes. The advantage of these features are that they are generic and are applicable for graphs from any domain. Thus, no domain knowledge is necessary to compute the values of these features from the social network. However, for large social networks, some of these features may be computationally expensive. Below we explain some of the popular graph topological features under two categories: (1)

Node neighborhood based and (2) Path based.

3.2 Node Neighborhood Based Features

3.2.1 Common Neighbors

For two nodes, x and y , the size of their common neighbors is defined as $|X \cap Y|$. The idea of using the size of common neighbors is just an attestation to the network transitivity property. In simple words, it means that in social networks if vertex x is connected to vertex z and vertex y is connected to vertex z , then there is a heightened probability that vertex x will also be connected to vertex y.

3.3 Jaccard Co-efficient

The common neighbors metric is not normalized, so one can use the Jaccard Coefficient, which normalizes the size of common neighbors as below:

$$Jaccard\ Coefficient(x,y) = \frac{|X \cap Y|}{|X \cup Y|}$$

Conceptually, it defines the probability that a common neighbor of a pair of vertices x and y would be selected if the selection is made randomly from the union of the neighbor-sets of x and y . So, for high number of common neighbors, the score would be higher.

However, from the experimental results of four different collaboration network s, Liben-Nowell et. al. [6] showed that the performance of Jaccard coefficient is worse in comparison to the number of common neighbors.

3.4 Shortest Path Distance

The fact that the friends of a friend can become a friend suggests that the path distance between two nodes in a social network can influence the formation of a link between them. The shorter the distance, the higher the chance that it could happen. But, also note that, due to the small world phenomenon, mostly every pair of nodes is separated by a small number of vertices. So, this feature sometimes does not work that well.

Katz. Leo Katz proposed this metric in [3]. It is a variant of shortest path distance, but generally works better for link prediction. It directly sums over all the paths that exist between a pair of vertices x and y . But, to penalize the contribution of longer paths in the similarity computation it exponentially damps the contribution of a path by a factor of β l, where l is the path length. The exact equation to compute the Katz value is as below:

$$Katz(x,y) = \sum_{l=1}^{\infty} \beta^l \cdot |paths_{x,y}^l|$$

Where path(x,y) is the set of all paths of length l from X to y. Katz generally works much better than the shortest path since it I s based on the ensemble of all paths between the nodes X and y.

3.4.1 Hitting Time

The concept of hitting time comes from random walks on a graph. For two vertices, x and y in a graph, the hitting time, H(x,y) defines the expected number of steps required for a random walk starting at x to reach y . Shorter hitting time denotes that the nodes are similar to each other, so they have a higher chance of linking in the future. Since this metric is not symmetric, for undirected graphs the commute time,

Cx,y = Hx,y + Hy,x , can be used. The benefit of this metric is that it is easy to compute by performing some trial random walks. On the downside, its value can have high variance; hence, prediction by this feature can be poor.

3.5 Problem Formulation

Users face many choices on the web when it comes to choosing which product to buy, which video to watch, and so on. In making adoption decisions, users rely not only on their own preferences, but also on friends. Many of the researches in previous work for the different combination of features for prediction of item adoption.

But in [1] it has utilize a Social Correlation Framework that incorporates the social correlation matrix C in the generation of user-item adoption links. Within this framework, paper propose two generative models: Sequential Generative Model and Unified Generative Model. The Sequential Generative

Model learns C in two sequential steps, first employing LDA to learn the parameters of the user and item latent factors, followed by learning C based on those parameters.

The Unified Generative Model learns simultaneously with the user and item latent factors in a principled, and unified way. The framework and two generative models are novel contributions over the previous state of the art that relies only on user and item latent factors (e.g., LDA).

But the prediction accuracy can be increase by including some more features which are not taken in the work. As the social network contain large number of information which is utilize for the adoption of item.

Users face many choices on the web when it comes to choosing which product to buy, which video to watch, and so on. In making adoption decisions, users rely not only on their own preferences, but also on friends. Many of the researches in previous work for the different combination of features for prediction of item adoption.

But in [1] it has utilize a Social Correlation Framework that incorporates the social correlation matrix C in the generation of user-item adoption links. Within this framework, paper propose two generative models: Sequential Generative Model and Unified Generative Model. The Sequential Generative Model learns C in two sequential steps, first employing LDA to learn the parameters of the user and item latent factors, followed by learning C based on those parameters.

The Unified Generative Model learns simultaneously with the user and item latent factors in a principled, and unified way. The framework and two generative models are novel contributions over the previous state of the art that relies only on user and item latent factors (e.g., LDA).

But the prediction accuracy can be increase by including some more features which are not taken in the work. As the social network contain large number of information which is utilize for the adoption of item.

4. CONCLUSION

As the increase of social media, number of internet user has increase in large number. Most of the young people send lots of time on internet, so this attract most of the company for there product branding. As user spend time and watch different small text, image or video advertisement. This makes an image of that product in their mind and chance of that product selling increases. Now next step is positive publicity of the people on internet by mentioning product review is done different website. So analysis of product review and user social connect takes to research of product predict.

So researchers get new field for mining that is product prediction. Web item prediction has been widely used to reduce the user confusion problem. This work has focus on product prediction where new combination of Jaccard base social network utilization is done with probabilistic function LDA. Here by including the social media features efficiency of product prediction get highly increases. Experiment done on real dataset and compare with existing method. Results shows that with the increase in features for Jaccard coefficient prediction accuracy has increase. Although research in this field is just a start, it is required to develop an adaptive algorithm as per social network.

As research is never ending process, so lots of work are still present in this work as well. Such as rating of product is done

on no criteria, but it should some of technical, environmental, etc. kind of criteria. So as per user requirement product prediction get stronger. Combination of other feature such as company profile has also make a great impact on the user choice so inclusion of that feature will also increase the product prediction efficiency. As results are showing less than one percent of accuracy due to large number of options so by using some kind of filter, few of products can be remove and chance of accurate prediction can be improve.

5. REFERENCES

[1] Freddy Chong Tat Chua, Hady W. Lauw, and Ee-Peng Lim. ―Generative Models for Item Adoptions Using Social Correlation‖. IEEE transaction on knowledge and data engineering, vil. 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] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proc. 4th ACM RecSys Conf*, 2010.

[7] L. Kartuz. A new index derived from social analysis. *Psychometrika*, 18(1), 1953.

[8] H. Li, S. Bhowmick, and A. Sun. Affrank: Affinity-driven ranking of products in online social rating networks. Journal of the American Society for Information Science and Technology 2012.

[9] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In Proc. 12th CIKM Conf., 2003.

[10] P. Massa and P. Avesani. Trust-aware collaborative filtering for recommender systems. In Proc. Federated Int. Conf. on The Move to Meaningful Internet:CoopIS, DOA, ODBASE, pages 492–508, 2004.

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

[12] Rodriguez, S. (2009). Consumer Behavior Report. *Economic Climate Shifts Consumers Online,* 3.

[13] Nielson. (2010). *Global Trends in Online Shopping.*USA: Nielson.

[14] Massa, P. & Bhattacharjee, B. (2004). Using Trust in Recommender Systems: An Experimental Analysis, Trust Management: Second International Conference, 2004, Oxford, UK, March 29-April 1, 2004, Volume 2995, 2004.

[15] Abdul-Rahman, A. and Hailes, S., Using Recommendations for Managing Trust In Distributed

Systems. In Proceedings of IEEE Malaysia International Conference on Communication, 1997.

[16] John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering, Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1998.

[17] Melville P., Mooney R.L. , Nagarajan R., Content-Boosted Collaborative Filtering for Improved Recommendations, In proceedings of Eighteenth national conf. of Artificial Intelligence, 2002.

[18] Prem Melville, Raymond J. Mooney, and Ramadass Nagarajan. Contentboosted collaborative filtering for improved recommendations. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02), 2002.

[19] P. Resnick and H.R. Varian, Recommender systems, In proceedings of Communications of the ACM, 1997, 40(3).

[20] Mladenic, D.: Text-learning and Related Intelligent Agents: A Survey, IEEE Intelligent Systems and their applications, 1999, Vol. 14(4).

[21] Jeffrey D.Ullman, Jure Leskovec, Anand Rajaraman, Mining of Massive Datasets, Cambridge University Press.

[22] Byeong Man Kim & Qing Li & Chang Seok Park & Si Gwan Kim & Ju Yeon Kim, A new approach for combining content-based and collaborative filters, J Intell Inf Syst 2006, Vol 27.

[23] Nitin Agarwal, Ehtesham Haque, Huan Liu, and Lance Parsons, Research Paper Recommender Systems: A Subspace Clustering Approach, In International Conference on Web-Age Information Management.

[24] C. Basu, H. Hirsh, and W. Cohen. Recommendation as Classification: Using social and content-based

information in recommendation. In Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98).

[25] Mahmood, T., Ricci, F.: Improving recommender systems with adaptive conversational strategies. In: C. Cattuto, G. Ruffo, F. Menczer (eds.) Hypertext, pp. 73–82. ACM (2009).

[26] Burke, R.: Hybrid web recommender systems. In the Adaptive Web, 2007 Vol 4321.

[27] Zhou, Daniel and Resnick, Paul, Assessment of conversation co-mentions as a resource for software module recommendation, In Proceedings of RecSys '09. ACM, New York, NY.

[28] Luhn, H. P, A statistical approach to mechanized encoding and searching of literary information, In IBM Journal of Research and Development, 1957, Vol 1(4).

[29] Dennis, S. F, In M. E. Stevens, V. E. Giuliano, & L. B. Heilprin (Eds.), Statistical association methods for mechanized documentation, Symposium Proceedings (Miscellaneous publication 269). Washington, DC: National Bureau of Standards.

[30] Salton, G., & Buckley, C., Weighting approaches in automatic text retrieval, In Information Processing and Management, 1988, Vol. 24(5).

[31] Sparck-Jones, K., A statistical interpretation of term specificity and its application in retrieval, In Journal of Documentation, 1972, Vol. 28(1).

[32] Robertson, S. E., & Sparck-Jones, K, Relevance weighting of search terms, In Journal of the American Society of Information Science, 1976 Vol. 27.

[33] Slonim, N., & Tishby, N, Document clustering using word clusters via the information bottleneck method, In Proceedings of the 23rd international conference on research and development in information retrieval, 2000.