International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 4, Issue 12, December 2015 **ISSN 2319 - 4847**

Integration of Jaccard Coefficient With LDA for User Product Prediction by Social Network

Hritika Jain[1], Dr. Bhupesh Gaur[2] , Prof. R M Sharma[3]

[1]M. Tech. Scholar TIT Bhopal

[2]HOD & Professor, Department of CSE, TIT Bhopal

[3]Asst. Prof., Department of CSE, MLCRP Vishwavidyalaya

ABSTRACT

Online Social Rating Networks such as Epinions and Flixter, allow users to form several implicit social networks, through their daily interactions like co-commenting on the same products, or similar co-rating products. This paper work on the user product prediction base on the social network as well as product rating. Here new concept of jaccard coefficient is involved in the work, where a different feature of the user from social graph is use for prediction. Results shows that proposed work has high precision value of 0.25 is achieved while accuracy of 0.0714 as compared to previous approach in [1], where precision is 0.081 while accuracy is only 0.0202.

Keywords: Data-relation, Jaccard Coefficient, Product prediction, Social network, Social features

1.INTRODUCTION

Digtal world has reduce the cost of developing and maintaining the social relation between users. Due to this reason many social networking media are growing day by day like facebook, linkedin , etc. This digital gathering attracts most of the e-shops where different products are sale based on the user choice. Some of social sites like Epinions and Flixter have attracted huge attention for the product rating and recommendation. In such sites, users often belong to different social network implicitly or explicitly for their interpersonal relation. As in Epinions user add each other as friend users' permission for discussing about any topic or product. Therefore a large number of Uniparty friendship network take part in this. However, besides the explicit friendship relations between the users, there are also other implicit relations. For example, users can co-comment on products and they can co-rate products.

This attracts many researchers to work for developing product prediction model. This is like a system which analyzes the user behavior based on the user social network or past purchasing, or friend recommendations [1, 6, 8].

Many of the researchers are working in this field of product prediction that is a combination of two or more social network. So Recommendations is based on the friendship network among users, and the network groups. However, as they mentioned their method focuses only on path counts and does not exploit other features and network characteristics which can be informative for link formation.

In addition, even though a social network is informative, particular features may be irrelevant and noisy for a specific user. For these reasons, some an effective automatic weighting strategy of the social networks influence based on their structured density. So the local (i.e. user's profile density) and global (i.e. network's density) characteristics of multimodal social graphs is used. Based on these characteristics, for each target user analogously calibrate the influence of each social network. For example, a user could have very few friends in the friendship network, but many interactions in co-commenting or co-rating products (i.e. user-items rating network). In such a case, the weighting strategy of model promotes the information given by the user-item rating network. Finally, generalize model for combining multiple social networks. Social network are highly vulnerable which can be understand as:

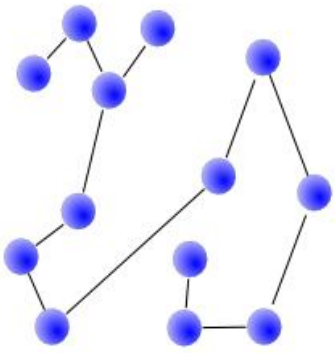

Figure. 1: Social network graph at time t.

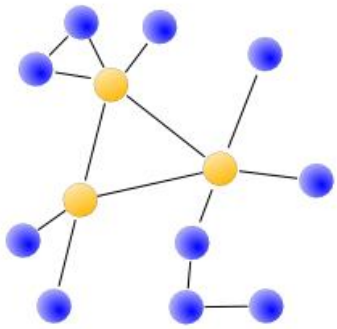

Figure. 2: Social network graph at time t'.

In order to understand it in a better way this can be understand by Figure.1 &Figure. 2 here yellow nodes of a Figure is going to produce the new link between them. This make change in the structure of the graph. Therefore new features get generate and accuracy of the system get decrease as this dynamic change reduce system efficiency [6, 10]. As in Figure. 1 there is no direct link between yellow nodes at time t, but at time t' there is a direct link between them so finding of the new edge before t' in the graph is next link prediction, in the social network

1.1 Problem Identification

There are different techniques developed for prediction which is based on the makov modal, soft computing techniques like ant bee colony, latent dirichlet function, etc. [1]. Out of these a perfect combination of function is required for increasing the prediction ratio. Combination of social features is also required for increasing the prediction accuracy. As per the latest product market new feature involvement is always done, so an adaptive algorithm is required.

2.RELATED WORK

H. Liet. Al.[8] has proposed a model where six features are utilize for the ranking of the product. It was named as AF rank which include feature like affinity rank history, product community size, evolution distance, member connectivity, social context and average rating.

Katz[7] in similar fashion, construct the probabilistic product prediction model. This model utilizes 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. Sarwaret. al. [11] has proposed,a CF Collaborative filtering model has been developed. Which utilize item-item similarity, instead of item user similarity. It has been obtained that product recommendation base on item similarity is high.

Kleinberget. 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. Herlockeret. 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. Konstanet. 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.

//NEED TO MAKE TABLE IN PROPER WAY

Table 1 Comparison table of different papers.

TECHNIQUES	PROBLEM	ADVANTAGES	DISADVANTAGES
IN [4] WEIGHT SIMILARITY MATRIX	WEIGHT NEED TO BE ADJUST AS PER DATASET	RECOMMENDATION OF HIGH RATE PRODUCTS IS ACCURATE	MANUAL WORK

IN [11] COLLABORATIVE FILTERING	UTILIZE ONLY ITEM-ITEM RELATION	LESS COMPLEX	LESS EFFICIENT
UNIFIED GENERATIVE MODEL [1]	USER SOCIAL RELATION IS NOT SUFFICIENTLY UTILIZE	UTILIZE ITEM-USER SOCIAL RELATION	PREDICTION ACCURACY IS LOW

2.1 Feature Selection

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 features 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}.

3.PROPOSED WORK

In this work product is predict, with the use of different relation such as user product relation user-user relation. Based on this relation a new combination of features is use for the prediction of product that will be purchase by user. So fig. 2 represents the steps of proposed work.

3.1 User-User Dataset

In this dataset user-user feature relation is present. This can be understand as user U1 has some relation with U2 in terms of {Like, comment, share image, share video, message, share comment, friend request, same group, common friends, video chat, text chat, etc.}, then number of time these activity done by the user is count in the dataset for U2 by U1 is store.

3.2 Pre-Processing

As dataset contain number of feature between user so conversion of dataset as per working environment is done in this step here dataset is arrange into matrix form where first two column represent user is while rest of column represent the feature count value present in dataset. If zero present in the column then it shows that that feature is not use by the specify user ids in first two column.

$$UUD \leftarrow \text{Pre\_processing}(UUD)$$

3.3 Jaccard Coefficient

The value of the features is is in integer form and differ person to person, so the Jaccard Coefficient are generate from the pre-processed dataset:

$$\text{Jaccard-coefficient}(x,y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}$$

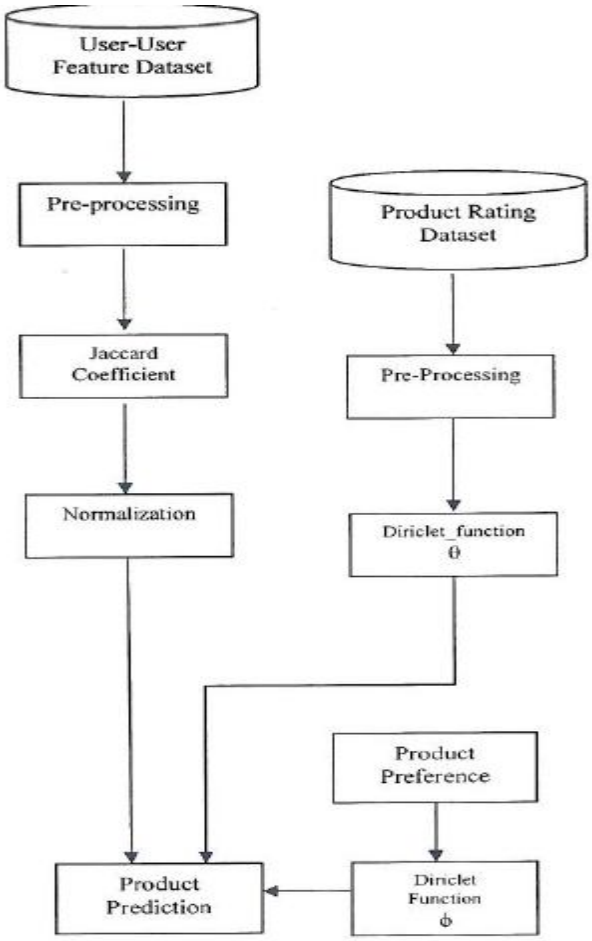

Figure. 3: Block diagram of Proposed Work

Conceptually, it defines the probability that a common neighbor of a pair of user 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. This can be understand as let X∩Y has like feature value 5, while XUY has like value 50 then Jaccard Coefficient value is 0.1.

JC[]←Jaccard_Coefficient(UUD)

3.4 Normalization

In this step matrix obtain after the jaccard coefficient need normalization. As different feature has different priority so to put all feature in same scale normalization is required.

So if W is the weight matrix, then following step will do normalization.

Loop 1:n

JC[n] ← JC[n]*W

EndLoop

3.5 Product Rating Dataset

In this dataset product rating feature is present. This can be understand as user U1 has either use or have knowledge or its review for any product id P1then rate it on the basis of his thought such as {best, very good, better, good, ok}.

3.6 Pre-Processing

As dataset contain number of rating between user and product so conversion of dataset as per working environment is done in this step here dataset is arrange into matrix form where first column represent user-id second represent product-id while third us for rate. For giving rate instead of giving any text rate values are provide for each class. If zero present in the column then it shows that that product is not use by the specify user ids.

UPD ←Pre_processing(UPD)

3.7 Latent Dirchlet Algorithm

Here with the help of this function dirrchlet will give a value as a relation between the user and user or item which is base on the UPD rating dataset.

θ← LDA(UPD)

In the similar fashion each product has its own product preference, so by the use of LDA one more relation is introduce.

Φ←LDA(Product_preference)

3.8 Product Prediction

This is the final step here user product prediction is done on the basis of social graph (user-user dataset), user-item relation, item preference.

In this step each user Xn who is friend of user Yj where j=1,2….t where t is number of X friends.

Loop 1: n
  Loop 1:j
    P[j] ←θj*□*JCn
  EndLoop
EndLoop

Now this P has j number of entries. So the maximum value index in P will be the final product id.

3.9 Proposed algorithm:

Input: UUD, UPD, Product_Preference
Output: Product_prediction

1. UUD ←Pre-Processing(UUD)
2. JC[]←Jaccard_Coefficient(UUD)
3. Loop 1:n
4. JC[n] ← JC[n]*W
5. EndLoop
6. UPD ←Pre_processing(UPD)
7. θ← LDA(UPD)
8. Φ← LDA(Product_preference)
9. Loop 1: n
10. Loop 1:j
11. P[j] ←θj*□*JCn
12. If P[j]>T
13. [x y]←Find_Friend(UPD[j], n) // Find friend of user n for product j
14. P[j] = P[j]*(x*C1 + y*C2) // x is Number of j product user who are n friend and y is other user of product j
15. EndIf
16. EndLoop
17. EndLoop

4.EXPERIMENT AND RESULT

4.1 Experimental Setup

This section presents the experimental evaluation of the proposed work. All algorithms and utility measures were implemented using the MATLAB tool. The tests were performed on an 2.27 GHz Intel Core i3 machine, equipped with 4 GB of RAM, and running under Windows 7 Professional.

4.2Dataset

The Epinions dataset contains

- 49,290 users who rated a total of
- 139,738 different items at least once, writing
- 664,824 reviews.
- 487,181 issued trust statements.

Users and Items are represented by anonimized numeric identifiers.

The dataset consists of 2 files: first file contains the ratings given by users to items, second file contains the trust statements issued by users.

4.3 Evaluation Parameter

To test outcomes of the work following are the evaluation parameter such as Precision, Recall and F-score.

Precision = TP / (TP+ FP)

Recall = TP / (TP + TN)

F-score = 2 * Precision * Recall / (Precision + Recall)

where TP : True Positive
TN: True Negative
FP : False Positive

Area Under the Curve (AUC):

With the help of precision and recall value AUC value is calculated that is term as the Area under the precision, recall curve.

4.4 Results

Results are compare with the previous work in [1] which is term as previous work in this paper.

Table 2 Comparison results of previous work with proposed work for 1000 user and 5 product

Values	Previous [1]	Proposed
Precision	0.0811	0.5625
Recall	0.0234	0.4091
F-Measure	0.0364	0.4737

Table 3 Comparison results of Previous work with proposed work for 1000 user and 5 product.

Values	Previous [1]	Proposed
Accuracy	0.0185	0.1151
Error Rate	0.9815	0.8949

Table 4 Comparison results of Previous work with proposed work for 500 user and 5 product.

Values	Previous [1]	Proposed
Precision	0.1852	0.6000
Recall	0.0221	0.4091
F-Measure	0.0395	0.4865

Table 5 Comparison results of Previous work with proposed work for 500 user and 5 product.

Values	Previous [1]	Proposed
Accuracy	0.0202	0.1227
Error Rate	0.9798	0.8773

.

It has been observed by table 2, 3, 4 & 5 that by product prediction of proposed work is better as compare to previous one, as precision value is higher. It is observed that as the size of the dataset increases then number of user and there chance of generating product prediction get less. This is due to the confusion or the randomness of user.

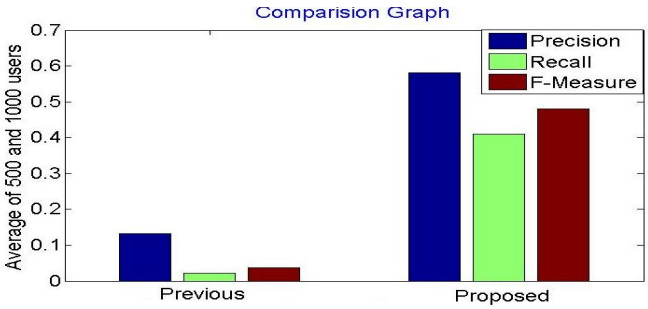

Figure 4 Average of precision, Recall, F-measure.

It has been observed by figure 4 that by product prediction of proposed work is better as compare to previous one, as average of different evaluation parameter from table 1 and table 2 is higher. It is observed that as the size of the datset increases or decrease values are always above from previous work

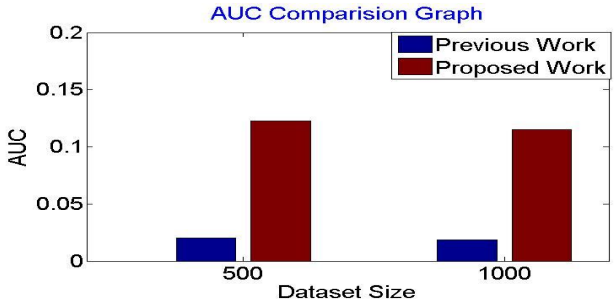

Figure 5 Comparison of AUC value at 1000 and 500 user.

Figure 5 represent the Area under the curve value where area is plot between precision, recall. These graphs also represent same things that with propose work is better as compare to previous one as area under the proposed work is more as compare to previous one.

5.CONCLUSION

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 comparision is done with existing method. Result shows that with the increase in features for Jaccard coefficient prediction accuracy has increase. It is obtained that proposed work has high precision value of 0.25 is achieved while accuracy of 0.0714 as compared to previous approach in [1], where precision is 0.081 while accuracy is only 0.0202.

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., pages 135–142, 2010.
[7] L. Kartuz. A new index derived from social analysis. Psychometrika, 18(1):39–43, 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.