IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Sorensen Similarity based User Recommendation by Social Network and Product Rating

Mohammed Affan

MBA, Marketing

School of Business and Management

CHRIST (Deemed to be University), Bengaluru, India

Abstract: The internet has initiated as a crucial part of everyone's daily life, as a vast amount of information is present on the web. Consumers, as well as retailers, tolerate the problems of the overload. Here the retailers keen to use the algorithm work process for deciding the material for the users. The proposed work is explained in detail with the help of a block diagram. This work has focused on product prediction where a new combination of Jaccard-based social network utilization is done with probabilistic function LDA. Hereby, including the social media features, the efficiency of product prediction gets highly increased.

I. INTRODUCTION

In today's scenario, people have become very possessive towards choices & options such as what should we intake as food what to buy on electronic devices. The decision-making space is getting increased as we are going ahead. In the absence of prior knowledge regarding the domain space, it's hard to get to the final decision [1, 2]. Therefore, mainly people depend on expert advice and their friends for making the decision.

At present, the internet has initiated as a crucial part of everyone's daily life, as a vast amount of information is present on the web. Consumers, as well as retailers, tolerate the problems of the overload. Here the retailers keen to use the algorithm work process for deciding the material for the users. When the retailer shows highly related content, this leads to increasing the customer interest in buying it. For getting the solution to this problem, the researchers have generated several algorithms and some systems, these becoming very famous such as Amazon.com, Ebay.com & Netflix.com. Therefore, the recommendation system is getting renowned in the research and commercial community [3].

In a consumer behavior report, "One in four consumers shows that they give more time online because they found it economy high percent of consumers said they spend the same amount [4]. As hundreds of e-commerce websites have no history of the users, understanding their purchasing behavior is a big task. So, the prediction of a product as per the user choice is made by learning its social behavior [5]. So, developing a method that gives product prediction is the main objective of this work. This predictive method would help in several functional areas, 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.
- Increase efficiency of the present product prediction systems or engines as they can guide the recommender system to find domains of interest for the user.
- Develop e-commerce companies with tools for targeted social media campaigns.
- Correlate data from different domain websites like social networks (Facebook) and reviewers (Epinion).
- Increase accuracy of the product prediction system.

II. Related Work

In [6] has proposed Social-Union, a method that 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 performs 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).

In [7], the proposed approach models the human sense of the relationships between objects based on their appearances. The approach is based on the human perception of visual connections between products. This human perception is modeled as a network inference graph problem. It is thus capable of recommending clothes and accessories together with excellent subjective performance.

In [8] paper proposed a knowledge service framework based on a case set. Three knowledge retrieval methods are designed based on parts keywords, customer orders, and manufacturing processes. Additionally, a VSMbased (vector space model) knowledge recommender method is proposed using a similarity matching algorithm to improve the efficiency of knowledge acquisition and transfer.

In [9], the paper has put forward a Structural Balance Theory-based Recommendation (i.e., SBT-Rec) approach. In the concrete, (I) user-based recommendation: we look for the target user's "enemy" (i.e., the users having opposite preference with target user); afterward, we determine the target user's "possible friends," according to "enemy's enemy is a friend" rule of Structural Balance Theory, and recommend the product items preferred by "possible friends" of target user to the target user. (II) likewise, for the product items purchased and selected by the target user, we determine their "possibly similar product items" based on Structural Balance Theory and recommend them to the target user.

In [10], a review-aware cross-domain recommendation algorithm, called RACRec, is proposed to address the fully cold-start problem in the field of product recommendation. Firstly, reviews are dynamically selected by using the adjacency matrix. Secondly, domain-specific preference vectors and domain-shared preference vectors of the cold start user are extracted by a migration model. On the other hand, the product feature vector in the target domain, which is generated from review texts by encoder and decoder, is combined with preference vectors of the cold-start user to make the rating prediction.

In [11], the author proposed a novel collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering. Firstly, the score matrix is preprocessed with normalization and dimension reduction to obtain denser score data. Based on these processed data, the clustering principle is generated, and dynamic evolutionary clustering is implemented. Secondly, the search for the nearest neighbors with the highest similar interest is considered. A measurement about the relationship between users is proposed, called user correlation, which applies the satisfaction of users and the potential information. In each user group, user correlation is applied to choose the nearest neighbors to predict ratings.

III. PROPOSED WORK

This proposed work is explained in detail with the help of a block diagram. Here, the block diagram shows the steps of the proposed work. The proposed work algorithm again defines the same thing. Work has included different formulas with the example of its input and output. Explanation of some dataset formats is also discussed as preprocessing of those is done in proposed work.

Feature Selection

As different online social networks have other purposes, the website features also vary from network to network. Let us consider Facebook as the basis of online social networks. There are many different events like comment, like, tag, unlike, write on the wall, etc. Hence, events act as a feature of the user. This can be understood as when a user likes a comment, send a message, then the link of the graph not generate, even when the user sends friend request also then also the new connection is not generated between them. But once one user accepts or conforms to the other user friend request, then only a new link of the graph of that online social network is developed between those users only.

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

User-User Dataset

The user dataset contains social relation data of the different users, including social features of that network. The social part includes different relations like { share image, Like, share video, comment, text chat, video chat, same group, common friends, friend request, message, etc.}. Here counting of the different features is present between two users.

Pre-Processing

Raw datasets contain various features and formats, so the conversion of data as per the proposed work environment is done in this step. Here data is organized into columns and rows where each column represents a different feature count between the users. Here first two columns are the user's identification number. If the feature value is zero, then that feature is not utilized by that pair of users.

UUD ← Pre_processing(UUD)

Sorensen Similarity

The value of the features is in integer form and differ from person to person, so the Sorensen Similarity [12]: generate from the preprocessed dataset:

$$SS = \frac{N(x) \cap N(y)}{d(x) + d(y)}$$

Where N(x) is a number of the neighbor of x, and d(x) is the degree of x. Sorensen Similarity is the ratio of common neighbor between x, y to the sum of nodes degree.

Product Rating Dataset

This dataset contains product market value estimated by the continuous analysis of the different products on the same evaluation parameters. Here five years of market analysis is done for each product. A rating of product on the scale of {poor, good, very good, etc.} is done.

Pre-Processing

Here dataset is preprocessed for conversion of scale text value into numeric values as numeric value is required in the proposed work. Here user rating is also identified by the dataset in front of the product. Conversion of data as per the proposed work environment is done in this step. Here, the dataset is arranged into a matrix form where the first column represents user-id, the second represents product-id, while the third is for rate. For giving rate, instead of giving any text rate, values are provided for each class. If zero is present in the column, it shows that the specified user ids do not use that product.

UPD ←Pre_processing(UPD)

Latent Dirichlet Algorithm

As per the values obtained from the user-user and user product data, one more relation is maintained between them by the Dirichlet function.

P← LDA(UPD)

The product has its own market value, which needs to be related by LDA again and represented by P.

P←LDA(Product_preference)

Product Rating Dataset

This dataset contains product market value estimated by the continuous analysis of the different products on the same evaluation parameters. Here five years of market analysis is done for each product. Rating R of product on the scale of {poor, good, very good, etc.} is done.

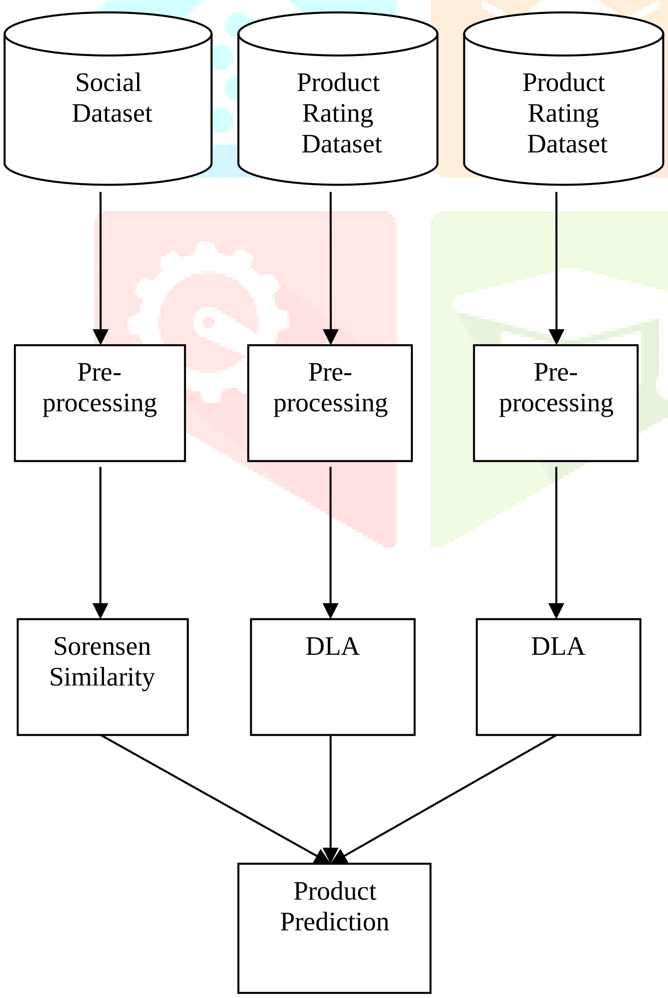

Fig. 1 Block diagram of Proposed Work

Product Prediction

In this step, utilization of both the network value features is done. This can be understood as let X is a friend of p number of users. So, if q friend of X suggestion is required, then it can be evaluated by:

```
Loop 1: p
        Loop 1: q
                P[q] ←SS*P*R
        EndLoop
EndLoop
```

IV. Experiment And Result

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 a 2.2 GHz Dual-Core Intel Core i7 notebook, equipped with 8 GB of RAM and running under Windows 10 Professional.

Dataset

The Epinions dataset contains [13].

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

Anonymized numeric identifiers represent users and Items.

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

Evaluation Parameter

The evaluation parameters such as Precision, Recall, and F-score to test the work following are the evaluation parameters [14].

Precision = TP / (TP+ FP)

Recall = TP / (TP + FN)

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

Accuracy = (TP + TN) / (TP +TN +FP +FN)

Error Rate = 1-Accuracy

Where TP : True Positive

TN : True Negative

FP: False Positive

FN: False Negative

Area Under the Curve (AUC):

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

Results

The proposed model SSPR (Sorensen Similarity Product Recommendation) is compared with the existing model proposed in [6]. The comparison is made on different dataset sizes.

Table 1 Precision value-based comparison of product recommendation.

Dataset	IASO [6]	SSPR
Set 1	0.2708	0.7143
Set 2	0.2184	0.5833
Set 3	0.0466	0.2813

Table 2 Recall value-based comparison of product recommendation.

Dataset	IASO [6]	SSPR
Set 1	0.1074	0.3636
Set 2	0.0745	0.3043
Set 3	0.0745	0.3913

Table 3 F-Measure value-based comparison of product recommendation.

Dataset	IASO [6]	SSPR
Set 1	0.1538	0.444
Set 2	0.111	0.4
Set 3	0.0573	0.3273

Table 1, 2 and 3 shows that the Sorensen similarity model in the SSPR model for product recommendation has increased the precision, recall and F-measure values. It is also found that with an increase in dataset size, the importance of parameter gets decreases. The use of social features in work for product recommendation has increased the prediction accuracy.

Table 4 Accuracy value-based comparison of product recommendation.

Dataset	IASO [6]	SSPR
Set 1	0.0833	0.2857
Set 2	0.0588	0.25
Set 3	0.0295	0.1957

Table 4 has a high accuracy value as compared to the previous model IASO [6]. This paper has utilized the Sorensen similarity model in the SSPR model for product recommendation. The use of social features in work for product recommendation has increased the prediction accuracy.

Table 5 Error rate value-based comparison of product recommendation.

Dataset	IASO [6]	SSPR
Set 1	0.9167	0.7143
Set 2	0.9412	0.75
Set 3	0.9705	0.8043

Table 5 has a lower error rate value as compared to the previous model IASO [6]. This paper has utilized the Sorensen similarity model in the SSPR model for product recommendation. The use of social features in work for product recommendation has increased the prediction accuracy. Product rating and preference value have reduced this parameter value.

V. Conclusion

Researchers get a new field for mining that is product prediction. Web item prediction has been widely used to reduce the user confusion problem. This work has focused on product prediction where a new combination of Jaccard-based social network utilization is done with probabilistic function LDA. Hereby including the social media features efficiency of product prediction get highly increases. The experiment is done on a real dataset and compared with the existing method. Results show that with the increase in features, for Jaccard coefficient prediction accuracy has increased. Although research in this field is just a start, it must develop an adaptive algorithm as per social network.

VI. References

1. R. D. Evans, J. X. Gao, N. Martin, and C. Simmonds, ‘‘A new paradigm for virtual knowledge sharing in product development based on emergent social software platforms,’’ Proc. Inst. Mech. Eng. B, J. Eng. Manuf., vol. 232, no. 13, pp. 2297–2308, Nov. 2018.

2. J. Zammit, J. Gao, R. Evans, and P. Maropoulos, ‘‘A knowledge capturing and sharing framework for improving the testing processes in global product development using storytelling and video sharing,’’ Proc. Inst. Mech. Eng. B, J. Eng. Manuf., vol. 232, no. 13, pp. 2286–2296, Nov. 2018.

3. R. Mishra, P. Kumar, and B. Bhasker, ‘‘A Web recommendation system considering sequential information,’’ Decis. Support Syst., vol. 75, pp. 1–10, Jul. 2015.

4. T. Liu, H. Wang, and Y. He, ‘‘Intelligent knowledge recommending approach for new product development based on workflow context matching,’’ Concurrent Eng., vol. 24, no. 4, pp. 318–329, Dec. 2016.

5. J. Song, H. Zhan, J. Yu, Q. Zhang, and Y. Wu, ‘‘Enterprise knowledge recommendation approach based on contextaware of time-sequence relationship,’’ Procedia Comput. Sci., vol. 107, no. 2017, pp. 285–290, 2017.

6. 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.

7. R. He and J. McAuley, ‘‘Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering,’’ in Proc. 25th Int. Conf. World Wide Web, 2016.

8. Z. Wu, L. Li and H. Liu, "Process Knowledge Recommendation System for Mechanical Product Design," in *IEEE Access*, vol. 8, pp. 112795-112804, 2020.

9. L. Qi *et al.*, "Structural Balance Theory-Based E-Commerce Recommendation over Big Rating Data," in *IEEE Transactions on Big Data*, vol. 4, no. 3, pp. 301-312, 1 Sept. 2018,

10. Y. Jin, S. Dong, Y. Cai and J. Hu, "RACRec: Review Aware Cross-Domain Recommendation for Fully-Cold-Start User," in *IEEE Access*, vol. 8, pp. 55032-55041, 2020,

11. Chen, J., Zhao, C., Uliji *et al.* Collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering. *Complex Intell. Syst.* **6,** 147–156 (2020).

12. T. Sorensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biol. Skr., 5, 1948.

13. http://www.trustlet.org/downloaded_epinions.html.

14. Asela Gunawardana, Guy Shani "A Survey of Accuracy Evaluation Metrics of Recommendation Tasks". Journal of Machine Learning Research 10 2009.