International Journal of Pure and Applied Mathematics
Volume 119 No. 12 2018, 13241-13250
ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

NEIGHBORHOOD-BASED APPROACH OF COLLABORATIVE FILTERING TECHNIQUES FOR BOOK RECOMMENDATION SYSTEM

Sivaramakrishnan N , Subramaniyaswamy V*, Arunkumar S, Renugadevi A, Ashikamai KK

School of Computing, SASTRA Deemed University, Thanjavur. India

*Corresponding Author

Abstract

Recommendation System or Recommender System help the user to predict the "rating" or "preference" a user would give to an item. Recommender systems in general helps the users to find content, products, or services (such as digital products, books, music, movie, TV programs, and web sites) by combining and analyzing suggestions from other users, which mean rating from various people, and users. These recommendation systems use analytic technology to calculate the results that a user is willing to purchase, and the users will receive recommendations to a product of their interest. The aim of the System is to provide a recommendation based on users likes or reviews or ratings. Recommendation system comprises of content based and collaborative based filtering techniques. In this paper, collaborative based filtering has been used to get the expected outcome. The expected outcome has been achieved through collaborative filtering with the help of correlation techniques which in turn comprises of Pearson correlation, cosine similarity, Kendall' s Tau correlation, Jaccard similarity, Spearman Rank Correlation, Mean-squared distance, etc. This paper tells about which similarity metrics such us Pearson correlation (PC), constrained Pearson correlation (CPC), spearman rank correlation (SRC) which is good in the context of book recommendation system and then applied with neighborhood algorithm.

Keywords: Pearson correlation, spearman correlation, constrained Pearson correlation, K nearest neighborhood algorithm, recommendation systems.

1. Introduction

In early 20's the recommendation system was a new technique, but nowadays almost each and every e-commerce website recommended items are based on recommendation system which may be under content- based or collaborative-based technique. Under the mentioned technique, the items are recommended for the users which increase their sales product and provide a satisfaction to them for the product which he/she buys.

By gathering information according to the interest of the users, automatic prediction has been done by the process of filtering and

such filtering is called as Collaborative filtering. Collaborative filtering recommends items based on computing similar items of the user which uses neighborhood algorithm (K-nearest neighbors). In Collaborative filtering technique, rating matrix is put up to group the users of similar interest [1]. The rating matrix comprises of items and users which is similar to general matrix which in turn comprises of columns and rows. The rating matrix is usually sparse since all the users may not have shown interest to all the items specified. In order to make predictions, the similar users and similar items could be determined with the help of neighborhood concept. The neighborhood technique comprises of item-based and user-based models. If the clustering of the items is based on similar users who have similar interest then it comes under user-based models. If the clustering of the items is based on similar rating which are given by similar user then it comes under item-based model [2].

The above-mentioned technique (K-nearest neighbor) is a machine learning algorithm which is used to find similar users among clusters of many users by their likes or ratings or reviews given by each specific user and make predictions according to the shortest distance among the users. K-nearest neighbor algorithm measures distance to determine the "closeness" of instances and gives output. Our proposal focuses on the development of methods that produce items' representations based on user's ratings for recommender systems. Our aim is to give best recommended books for the user; in spite we use whatever algorithm or metrics [2].

2. General Algorithm

Book recommendation System general approach is as follows:

1. To predict the rating of the user for an arbitrary book then active user has to be chosen.

2. To weigh all of the users in the data set weighting technique has to be used.

3. Based on the closest similarity to the active user, a set of users has to be picked up.

4. According to the subset selected in step 3, prediction has to be done by the user.

In this paper, In step1, rating predictions are done by the user in the order in which they appear.

In step2, the different similarity weighting schemes are 1) Pearson similarity, 2) spearman similarity, and 3) Constrained Pearson similarity. Here weighting schemes are used to observe which users are similar and which are not.

In step3, parameter 'k' is used which denotes the number of users chosen here. This is the 'k' users that correspond to the active users who are chosen in this step. We can make prediction to the books based on the weighted average only when we have k most similar users to the active users. It is defining as follows:

$$s(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - \bar{r}_v)^2}} \quad (1)$$

Where

u is the active user

v is any other user

N is the set of k most similar users to u

$r_{v,i}$ is the v's rating of book i

r is user's average book rating

s(u,v) is the similarity between active user u and user v. [3]

3. Related Works

One of the fast-growing businesses is e-commerce, in which recommendation plays an important role for the benefit of customer, consumer and who sells the items (vendor). Research articles published for recommendation systems are very useful, which help the researches to keep track in their research field. Their evaluation becomes better and helps to determine their individual strengths and weakness by proposing more recommendation approaches [12-22]. There are many techniques and metrics, but the question arise which evaluation constitutes a good recommendation system [23-36]. In research fields authors don't prefer evaluation standards often, three quarters of evaluation published in User modeling, but User-Adapted interaction had serious shortcomings in their evaluation, so the researchers not often see alone the evaluation but also prefer the out coming results. [4]

For Collaborative filtering the general formula for statistical was first published in the group lens project in which PC was defined for the basis of weights. (Resnick ET al.1994) The correlation between a and i

$$w(a,i) = \frac{\sum_j (v_{a,j} - \overline{v}_a)(v_{i,j} - \overline{v}_i)}{\sqrt{\sum_j (v_{a,j} - \overline{v}_a)2 \ \ \sum_j (v_{i,j} - \overline{v}_i)2}} \qquad (2)$$

Where,

j - Item both the user a and i rated. [5]

3.1 Table: rating matrix (from 0-5))

Users	I	II	III	IV
User A	2	4	3	2
User B	4	3	5	3
User C	3	4	?	?
User D	5	2	4	4
User E	2	1	3	5

In the above table representation values:

I - Wuthering heights

II - Nineteen eighty-Four

III - Nineteen eighty-Four

IV - The Lord Of Rings

Consider the ratings in the above rating matrix table. Suppose if we need to find User C's prediction for the Book the Lord of the rings, the steps as follows:

C's mean rating is 3. 5. There are four users who have rated the book The Lord of rings so the users with neighborhood are (A, B, D, and E),

The average ratings of user A = 2.75,

The average ratings of user B = 3.75,

The average ratings of user D = 3.75,

The average ratings of user E = 2.75.

s (C, A) = 1

s(C,B) =1

s (C, D) = -1

s(C,E) = 1

Bellcore video recommender (Hill et al.1995) and the Ringo music recommender (shardanand and maes 1995) they expanded the idea of group lens project, in which Ringo said CPC shows better results for computing similarity weights. (Breese ET al.1998) performed empirical analysis with many neighborhood algorithms in collaborative filtering, in which PC and cosine vector similarity are compared where correlation shows better performance. [7]

4 . Existing Work

4.1 Pearson Correlation Coefficient:

The intersection of the books is taken by the active user u and another user v with the help of PC coefficient. Then it computes the statistical correlation between the two active users who are mentioned above. By PC coefficient, the similarity between the two users are defined below as

$$s(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - \bar{r}_v)^2}} \quad (3)$$

Here, $r_{u,i}$ denotes movie *i*'s *ranking* amongst user *u*'s ratings.

4.2 Mean Square Distance

The sum of the squared difference between the user u and user v ratings of their common books which is divided by the number of common books. This metrics finds the similarity of tastes among the users. The mean squared distance scheme is defined as:

$$s(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - r_{v,i})2}{|I_u \cap I_v|} \quad [8]$$

4.3 Cosine Similarity

The cosine similarity approach is different from other metrics, to find the similarity among the users. Unlike other statistical approaches, here each and every user is considered as a vector of his/her ratings given to movies, Thus the distance between these two vectors are find by cosine similarity. The cosine similarity formula is defined as:

$$s(u,v) = \frac{r_u . r_v}{\|r_u\| \|r_v\|} \quad (6)$$

r_u Rating vectors for movies by user u

r_v Rating vectors for movies by user v [9]

4.4 Spearman Rank Correlation

In spearman correlation approach we find rank of the user rather than finding similarity among the user. Here the movies which are given higher ratings are given rank as 1 and the movies which are given lower ratings are then assigned to higher ranks. Books which have same ranks are given as average rank for their respective positions.

$$s(u,v) = \frac{\sum_{i \in I_u \cap I_v} (k_{u,i} - \bar{k}_u)(k_{v,i} - \bar{k}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (k_{u,i} - \bar{k}_u)2 \ \sum_{i \in I_u \cap I_v} (k_{v,i} - \bar{k}_v)2}} \quad (7)$$

$k_{u,i}$ is the i's ranking for the ratings given by user u. [10]

5. Proposed Work

In this paper we have done these similarity metrics using a book data set.

First, predictions have been made for all the users and the active users at the same time. We would calculate for every book the active user has given the ratings in order to find the similarity measure for active user and rest other users.

Second, we would find similarity of the user by using different approaches such as

1- PC coefficient
2- SRC coefficient
3- CPC coefficient

This step is done to find similar users similar and users whom are not similar

Then after computing k number of similar users for the active user then we find the values for every book using weighted average technique.

5.1 Constrained Pearson correlation

CPC is more similar to PC. In CPC instead of calculating mean we calculate median [10]. First it identifies the correlation coefficient between the active user and other user. Then the user whose value which is greater than certain threshold is found and then similar users are identified [12]

$$CPCC_{sim(u,v)} = \frac{\sum_{i \in I} (r_{u,i} - r_{med})(r_{v,i} - r_{med})}{\sqrt{\sum_{i \in I} (r_{u,i} - r_{med})2 + \sum_{i \in I} (r_{v,i} - r_{med})2}} \quad (8)$$

Where

$r_{u,i}$ - The average rating of user u for the given item i.

$r_{v,i}$ - The average rating of user v for the given item i.

r_{med} - Median value for the rating.

The other two approaches like PC and SRC coefficient are produced.

6. Results and conclusion

Most of the studies have been told that PC coefficient gives better results. In this paper we have compared three correlation coefficients (PC coefficient, CPC coefficient, SRC coefficient) for neighbor-based approach. The results have been evaluated for all the three correlations for neighbors ranging from 2 to 9 and then evaluated using mean absolute error.

The evaluations have been done for book recommender system based on ratings of the book given by the uses and the book they have rated. For this book recommender system, we have concluded that spearman correlation coefficient works best and having mean absolute error less than 1. The other two correlation coefficient works well which is having mean absolute error rate less than 1.2 which is acceptable (Recommendation whose mean absolute error greater than 2 is said to give false recommendations). In this system PC coefficient and CPC produces nearly same results with minimal difference in the mean

absolute error. Hence for book recommender system SRC coefficient would give better results than other two correlations as mentioned above. The results have been plotted as graph and shown below.

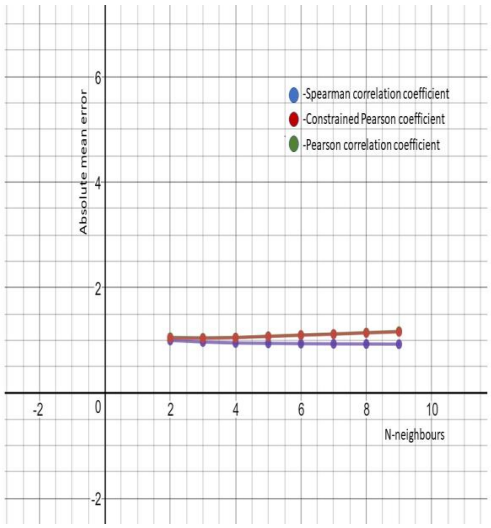

Fig 1: graph that consists of SRC, CPC and PC

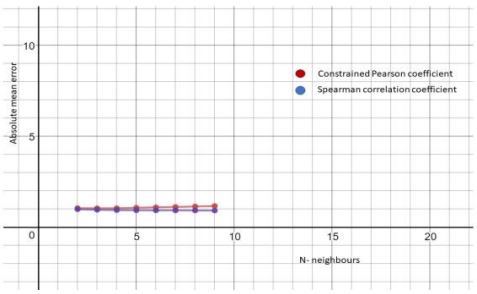

Fig 2: graph between SRC and CPC

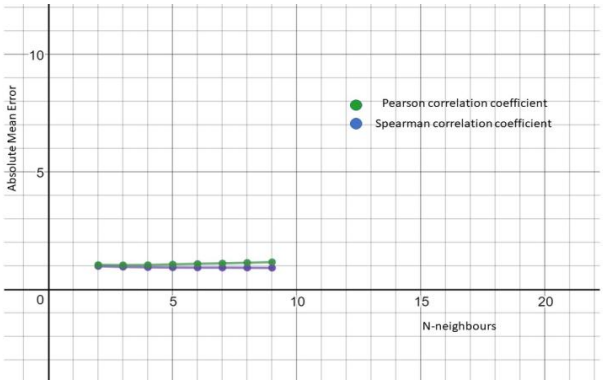

Fig 3: graph between SRC and PC

Our book dataset consists of (User_Id, Book_Id, Rating), From rating given by each user we calculate PC, SRC and CPC.

From those results we also calculate mean absolute error for each correlation and plotted as a graph which are shown above

From the above given figures:

1) Fig: 1 A graph is plotted against SRC, CPC and PC with N-neighbours at x-axis and absolute mean error at y-axis, (To differentiate PC and CPC the graph has been plotted separately as fig 2 and 3)

2) Fig: 2 A graphs is plotted against SRC and CPC with N-neighbours at x-axis and absolute mean error at y-axis

3) Fig: 3 A graphs is plotted against SRC and PC with N-neighbours at x-axis and absolute mean error at y-axis

7. References

[1] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, Item-based Collaborative Filtering Recommendation Algorithms, Appears in WWW10, May 1-5, 2001, Hong Kong.

[2] Hua-Ming Wang, Ge Yu, Personalized recommendation system K- neighbor algorithm optimization, International Conference on Information Technologies in Education and Learning (ICITEL 2015)

[3] Melville, Mooney, Nagarajan, Content Boosted Collaborative Filtering for Improved Recommendations, Proceedings of the Eighteenth National Conference on Artificial Intelligence, Edmonton, Canada, 2002.

[4] Joeran Beel, Stefan Langer, Marcel Genzmehr, Bela Gipp, Corinna Breitinger, Andreas Nürnberger, Research Paper Recommender System Evaluation: A Quantitative Literature Survey, 12 October 2013

[5] John S. Breeze, David Heckerman, Carl Kadie, Empirical Analysis of Predictive Algorithms for Collaborative Filtering, Microsoft Research Redmond, WA 98052-6399

[6] Michael D. Ekstrand, John T. Riedland, Joseph A. Konstan, Collaborative Filtering Recommender System, Foundation and Trends in Human Computer Interaction, Vol. 4, No. 2 (2010) 81–173.

[7] Herlocker, Konstan, Riedl, An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms, Informational retrieval, January 2002.

[8] Nirmal Krishnan, Emily Brahma, Survey of Neighborhood-based Collaborative Filtering Techniques for a Movie Recommendation Engine, 4 October 2016

[9] Ajay Agarwal, Minakshi Chauhan, Similarity Measures used in Recommender Systems: A Study, International Journal of Engineering Technology Science and Research IJETSR, ISSN 2394 – 3386 Volume 4, Issue 6 June 2017.

[10] Upendra Shardanand, Pattie Maes, Social Information Filtering: Algorithms for Automating "Word of Mouth", ACM.

[11] Taner Arsan, Efecan Köksal, Zeki Bozkuş, Comparison Of Collaborative Filtering Algorithms with Various Similarity Measures For Movie Recommendation, International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.6, No.3, June 2016

[12] Logesh, R., Subramaniyaswamy, V., Vijayakumar, V., Gao, X. Z., & Indragandhi, V. (2017). A hybrid quantum-induced swarm intelligence clustering for the urban trip recommendation in smart city. Future Generation Computer Systems, 83, 653-673.

[13] Subramaniyaswamy, V., & Logesh, R. (2017). Adaptive KNN based Recommender System through Mining of User Preferences. Wireless Personal Communications, 97(2), 2229-2247.

[14] Logesh, R., & Subramaniyaswamy, V. (2017). A Reliable Point of Interest Recommendation based on Trust Relevancy between Users. Wireless Personal Communications, 97(2), 2751-2780.

[15] Logesh, R., & Subramaniyaswamy, V. (2017). Learning Recency and Inferring Associations in Location Based Social Network for Emotion Induced Point-of-Interest Recommendation. Journal of Information Science & Engineering, 33(6), 1629–1647.

[16] Subramaniyaswamy, V., Logesh, R., Abejith, M., Umasankar, S., & Umamakeswari, A. (2017). Sentiment Analysis of Tweets for Estimating Criticality and Security of Events. Journal of Organizational and End User Computing (JOEUC), 29(4), 51-71.

[17] Indragandhi, V., Logesh, R., Subramaniyaswamy, V., Vijayakumar, V., Siarry, P., & Uden, L. (2018). Multi-objective optimization and energy management in renewable based AC/DC microgrid. Computers & Electrical Engineering.

[18] Subramaniyaswamy, V., Manogaran, G., Logesh, R., Vijayakumar, V., Chilamkurti, N., Malathi, D., & Senthilselvan, N. (2018). An ontology-driven personalized food recommendation in IoT-based healthcare system. The Journal of Supercomputing, 1-33.

[19] Arunkumar, S., Subramaniyaswamy, V., & Logesh, R. (2018). Hybrid Transform based Adaptive Steganography Scheme using Support Vector Machine for Cloud Storage. Cluster Computing.

[20] Indragandhi, V., Subramaniyaswamy, V., & Logesh, R. (2017). Resources, configurations, and soft computing techniques for power management and control of PV/wind hybrid system. Renewable and Sustainable Energy Reviews, 69, 129-143.

[21] Ravi, L., & Vairavasundaram, S. (2016). A collaborative location based travel recommendation system through enhanced rating prediction for the group of users. Computational intelligence and neuroscience, 2016, Article ID: 1291358.

[22] Logesh, R., Subramaniyaswamy, V., Malathi, D., Senthilselvan, N., Sasikumar, A., & Saravanan, P. (2017). Dynamic particle swarm optimization for personalized recommender system based on electroencephalography feedback. Biomedical Research, 28(13), 5646-5650.

[23] Arunkumar, S., Subramaniyaswamy, V., Karthikeyan, B., Saravanan, P., & Logesh, R. (2018). Meta-data based secret image sharing application for different sized biomedical images. Biomedical Research,29.

[24] Vairavasundaram, S., Varadharajan, V., Vairavasundaram, I., & Ravi, L. (2015). Data mining-based tag recommendation system: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5(3), 87-112.

[25] Logesh, R., Subramaniyaswamy, V., & Vijayakumar, V. (2018). A personalised travel recommender system utilising social network profile and accurate GPS data. Electronic Government, an International Journal, 14(1), 90-113.

[26] Vijayakumar, V., Subramaniyaswamy, V., Logesh, R., & Sivapathi, A. (2018). Effective Knowledge Based Recommeder System for Tailored Multiple Point of Interest Recommendation. International Journal of Web Portals.

[27] Subramaniyaswamy, V., Logesh, R., & Indragandhi, V. (2018). Intelligent sports commentary recommendation system for individual cricket players. International Journal of Advanced Intelligence Paradigms, 10(1-2), 103-117.

[28] Indragandhi, V., Subramaniyaswamy, V., & Logesh, R. (2017). Topological review and analysis of DC-DC boost converters. Journal of Engineering Science and Technology, 12 (6), 1541–1567.

[29] Saravanan, P., Arunkumar, S., Subramaniyaswamy, V., & Logesh, R. (2017). Enhanced web caching using bloom filter for local area networks. International Journal of Mechanical Engineering and Technology, 8(8), 211-217.

[30] Arunkumar, S., Subramaniyaswamy, V., Devika, R., & Logesh, R. (2017). Generating visually meaningful encrypted image using image splitting technique. International Journal of Mechanical Engineering and Technology, 8(8), 361–368.

[31] Subramaniyaswamy, V., Logesh, R., Chandrashekhar, M., Challa, A., & Vijayakumar, V. (2017). A personalised movie recommendation system based on collaborative filtering. International Journal of High Performance Computing and Networking, 10(1-2), 54-63.

[32] Senthilselvan, N., Udaya Sree, N., Medini, T., Subhakari Mounika, G., Subramaniyaswamy, V., Sivaramakrishnan, N., & Logesh, R. (2017). Keyword-aware recommender system based on user demographic attributes. International Journal of Mechanical Engineering and Technology, 8(8), 1466-1476.

[33] Subramaniyaswamy, V., Logesh, R., Vijayakumar, V., & Indragandhi, V. (2015). Automated Message Filtering System in Online Social Network. Procedia Computer Science, 50, 466-475.

[34] Subramaniyaswamy, V., Vijayakumar, V., Logesh, R., & Indragandhi, V. (2015). Unstructured data analysis on big data using map reduce. Procedia Computer Science, 50, 456-465.

[35] Subramaniyaswamy, V., Vijayakumar, V., Logesh, R., & Indragandhi, V. (2015). Intelligent travel recommendation system by mining attributes from community contributed photos. Procedia Computer Science, 50, 447-455.

[36] Vairavasundaram, S., & Logesh, R. (2017). Applying Semantic Relations for Automatic Topic Ontology Construction. Developments and Trends in Intelligent Technologies and Smart Systems, 48.