

Neural Network Classification for user Profile Learning over Digital Library Recommendation Engine

Disha Sharma^{1*}, Sumit Kaur² and Diksha³

¹CSE, Chandigarh University, Gharuan, Mohali - 140413, Punjab, India; er.dishasharma21@gmail.com

²Department of CSE, Chandigarh University, Gharuan, Mohali - 140413, Punjab, India; Sumit.bhangu87@gmail.com

³Department of Applied Sciences, Chandigarh University, Gharuan, Mohali - 140413, Punjab, India; dikshanagpal91@gmail.com

Abstract

Objectives: To propose a hybrid recommendation engine to make perfect order of recommendations for online digital library portals. **Methods/Statistical Analysis:** The proposed model combines the content-based learning based upon the neutralized factors along with the collaborative learning for producing the recommendations on the basis of the inter-user similarity, which is learned from the user's profile representation vector. The hybrid recommendation algorithm is intended to produce more relevant recommendations for the digital library users. **Findings:** The proposed has been tested over the different sizes of the data. The proposed model has been designed for the discovery of the similar entities using the neural network classification and the top five entities are utilized for the calculation of the missing profile values to compute the recommendations over the digital library systems. The proposed model's performance has been analyzed and studied using the statistical performance measuring parameters which elaborates the system performance from various perspectives. The experimental results have shows the robust performance of the proposed model in the terms of mentioned parameters. **Application/Improvements:** The proposed model has been improved the recommendations by the means of the calculating the inter user relevance to evaluate the recommendations for the user in the scope.

Keywords: Digital Library, Hybrid Recommendation Algorithm, Neural Network, Recommendation Engine, User Relevance

1. Introduction

For the comparison of position of one particular book with the other books, Recommendation system is used and it also helps to give higher and lower rank to the books. It helps user to choose best option from the list. As the internet usage become higher, online websites for buying and selling of the books, songs, furniture, videos apparels become higher. To find best site among a list of sites, recommendation system helps a lot. With the help of multiple factors, recommendation system is assigned. For assigning a rank, following are the multiple ways:

Standard competition Recommendation: In the procedure of ranking, this system compares books and gives same rank to similar books. Ranking is computed according to the higher value of product popularity factors.

Modified recommendations based upon competition: This model evaluates the semantic gaps in the differentiation based weights. On the basis of this model, the recommendations are considered in the weight based ranking computed in the inter-product competition.

Dense Recommendation Model: This model does not evaluate the semantic gap, but relies upon the formatted ranking planned on the basis of the selective measures.

*Author for correspondence

This system evaluates the rankings of similar products in the similar levels^{1,2}.

Ordinal Recommendation Model: In this model, the book data is inter-compared and the different ranks are assigned to each of the member enlisted within the product list. This model utilizes the rule-based evaluation for the preparation of the e-book rankings

Fractional Recommendation Model: In this system³, after comparing same book appoints same rank, but the rank is the average of the book. For example, “2 4.6 5.6 7 9”.

Without the pre-provided or pre-defined direction for the user evaluation, the semantic ranking module evaluates and reserves the application residing in the computing storage, which is made easy to understand and e-book listing information is attained in the form of the product page list. The e-book ranking algorithm is utilized for the preparation and sharing of the pre-developed semantic ranking lists in the semantic manner^{4,5}. It Dynamically merges and automatically updates the recommendation after performing comparison of new added book with the existing list of books⁶.

In⁷ different techniques in collection development in the field of digital environment are defined. It briefly describes the variation that came across in acquisition, retrieval and also the storage of information due to development in various techniques. It also portrayed the problems faced by the readers and librarians and how the development has overcome all those problems.

The proposed model defined in⁸ for the distribution of information in digital library university, he has proposed a quality based recommender system. To judge the relevance this system use quality of the product. In the recommendation process, system estimates the product quality and assumes this estimation as a new factor. To access the important data, this system helps a lot as it easily calculate product quality. The proposed system is tested in digital library university, and designed by using fuzzy technique.

The proposed model defined in⁹ is based upon the adaptive and semantic online digital e-book list over the online library database. To choose best e-book from many e-books this system helps the user by providing many different and specified features. With the help of these algorithms, communities come to know the weakness and hence improve their services. The major problem with this algorithm is that it include less number of features but give very accurate results. In this model, the multiple

query merger scheme has been proposed for the resource evaluation and combination formation. The division class based query evaluation model is known as the “skyline package query model”, which is utilized for the querying of the e-book packages to prepare the product listings.

The model proposed in¹⁰ they work on Online Libraries especially on schema.org. The schema.org based directional and conceptual approach for the calculation of the ranking direction and conceptual ranking structure has been studied in this model. For the various types of products, the demand and approach plays the important role in the calculation of the ranking lists.

In¹¹ the author has discussed an impactful way for a well-organized and exposure to an automated technique for the license plates perception.

In¹² the LBS systems are discussed and also what are the various requirements for the accuracy of the system. One of the most popular method i.e. fingerprint method is discussed and various smart-phone’s sensors are analyzed and based on various experiments it is tried to found out which phone’s sensor is best for indoor fingerprint positioning.

The model has been proposed in¹³ for the upgradation of web page rank, they introduce a back propagation neural network on Intelligent web mining. The web is enhancing at very high speed, because of enhancing information resources. For finding the necessary data from web, large size increase problems. With the help of personal web search, this problem can be defeat but the user has to give his private information to preserve privacy.

2. Experimental Design

The proposed algorithm has been designed for the adaptive ranking models for the preparation of the online digital library recommendation lists. The design of this model is utilized for the content-based model, which is performed in the multi-layered architecture to produce the more relevant and clear results from the library contents containing the information about the books. Various product features such as book category, author, publisher, popularity and other similar factors are utilized for the decision calculation over the content based recommendation engine. The digital library model is entitled to provide the book information, which is provided by the Application Programming Interface (API) from the Google books. The Google books are the large book data-

base, which contains the various e-books from the various categories. The content based recommendation is further amalgamated with the collaborative filtering, which analyzes the user similarity with the other user profiles to find the most appropriate matches. The most appropriate matches are aligned within in the collaborative filtering and the final search results are represented on the basis of the final listing obtained after the application of neural network over the input data.

The proposed model is based upon the probabilistic neural network, which is based upon the concept of the back propagation (bpb) for the elimination or selection of the entities by analyzing the training data in comparison with the input testing data. The entity weights are calculated automatically for the inclusion of the entity classification for the incorporation of the probability based evaluation of the entity matches. The neural network weights are computed with flexibility of the neuron relationships after evaluating the gradient descent based final decision logic evaluation. The probability method can be given by the following equation:

$$B_{ij}(k+1) = Q_{ij}(k) + m \frac{\partial B}{\partial b_{ij}}$$

Where the Q denotes the training coefficients, B gives the function for cost (or cost function), ∂ gives the initial value of probability, m denotes the total number of input arguments and k denotes the current value of the incremental counter or index counter. This model computes the flexi-weights for the ranking based equation modeling, the training samples have been analyzed for the calculation of product polarization and determined e-book ranking preparation in the semantic form. This model utilizes the supervised classification model based upon the neural computing for the preparation of collaboration user-profile based learning for the final decision computation from the input training matrix. The softmax function has been utilized as the activation function for the neural computing algorithm, which is utilized to determine the cost function for the realization of the probabilistic classification. The following equation satisfies the softmax function based fitness evaluation:

$$b_j = \frac{\exp(b)}{\sum_t \exp(b_t)}$$

Where b_j gives the overall class probability for the input classes of and b and b_k denotes the total number of unit argument inputs to the neural model, which is

controlled by the range variable of j and k, which are utilized to handle the indexing factor over the input data. The entropy function (also called cost entropy function) is utilized for the inclusion of the primary function for the successful running of the neural algorithm based probabilistic classification. The entropy function evaluates the pattern matching cost by analyzing the pattern strength and produces the strong features as the final results. The cost entropy function is defined with the following equation:

$$CE = -\sum d_j \log(e_j)$$

Where denotes the probability of the target unit defined by the indexer value j. The is the entity to define the output probability based upon the indexer value j and computed once the activation function is applied. Under the proposed model, the proposed knowledge-driven bpb-neural network with probabilistic approach in incorporated over the input book data for the calculation of the recommendation lists from the input data. The overall algorithms can be defined in the following set of the algorithms used in the hierarchical architecture:

Algorithm 1: BPP-Probabilistic Neural network for digital library recommendation engine

Input 1: Book Data

Input 2: Local parameters

Input 3: Global parameters

Input 4: User browsing history

Input 5: Other users profile recommendations

Output: Content Ranking Matrix

1. Acquisition of the library information matrix
2. Extract and align the book information data from the input library information matrix
3. Obtain the number of rows and columns from the input data matrix
4. Begin the sequential iteration with counter I and capitalizing value(C) equals number of rows
 - a. Extract the Ith rows from the input data information matrix
 - b. Acquire the local parameters for the given row
 - c. Acquire the global parameters for the given row
 - d. Align up the input local parameters and prepare the singular value representation (SVR)
 - e. Align up the input global parameters and prepare the singular value representation (SVR)
 - f. Obtain the singular value formation from step 4(e) and 4(f).
 - g. Update the content based indexing matrix (CIM) with local and global SVR factors

- h. Update the current row in CIM with book id
- i. If I equals C
 - i) Quit the iteration
 - ii) Return the CIM
- j. Otherwise
- k. Go to 4(a)
5. Verify the user's browsing history
- 6 If user history is empty
 - a. Store CIM information to hybrid indexing matrix (HIM)
 - b. Return HIM
7. Otherwise
 - a. Continue
 - b. Input the user browsing history \rightarrow UBH
 - c. Create the quadrupled factor matrix (QFM) containing author rating, book rating, publisher rating and online rating \rightarrow QFM
 - d. Initialize the neural network paradigm
 - e. Initialize the number of layers equals 10
 - f. Assign the multi-objective activation function
 - g. Acquire the Other users profile recommendations matrix (OUPR) as training matrix
 - h. Apply the probabilistic neural network classification.
 - i. Return the matching users list
 - j. Shortlist the five top listed users in the similarity matrix
 - k. Assign the top five rows to the mini-matrix
 - l. Perform the column-wise averaging factor over the mini-matrix
 - m. Assign in the new vector to the Collaborative Recommendation Factors (CRF) vector
 - n. Return the CRF
8. Evaluate the CIM against the CRF
9. Performing the re-sorting over the recommendation indexing in CIM and return HIM
10. Rearrange the HIM according to CRF
11. Return the HIM

Algorithm 2: Probabilistic Neural Network (PNN) method

Input 1: X (Training Matrix)

Input 2: T (Testing Vector)

Input 3: Activation Function

Input 4: Number of Layers

- Assign the activation function to the PNN and assign \rightarrow phi
- Initialize the offshoot value for the activation function over {phi} and return \rightarrow phiA

- Initiate the neural network with probabilistic back propagation behavior
- Acquire X and rearrange this to the matrix of input nodes (denoted with i)
 - Initiate the iterative function over every input object (i)
 - Prepare the output vector for each input object as derivative of X
 - Return the computational cost
- Initialize the hidden layers
- Perform the processing over the current input node matrix
 - Perform the computation over each object
 - Prepare the final results with following equation
$$Res_j = \hat{a}_i \phi(Q_{ji} \times res_i)$$
- Initialize the size of output matrix with equal number of neurons (k)
 - Perform the computation over each neuron denoted with k
 - $Res_j = \hat{a}_i \phi(W_{ji} \times res_i)$

Algorithm 3: Neural Network Activation Function

- Acquire the name of activation function
- Initialize the activation function information
- Initialize the activation of the input and output layers
- Perform the computation over the input object i based upon the current neuron k
 - Compute the result
- Perform the computation for every hidden neuron denoted with j
 - Compute the result
- Initialize the hidden layers
- Perform the processing over the current input node matrix
 - Compute the result
 - **Return the final results with Res= {res_k}**

Algorithm 4: Output Error Algorithm

- Initialize the error calculation module
- Perform the iterative operation over each input object
 - $CE_{in} = \hat{a}_{kfactor} (T_k - R_k)^2$
- Compute the averaging factor over CE_{in}
- Return the final error vector CE_{in}

3. Workflow Analysis of the Neural Network

The features are acquired and assigned to the training and testing matrix prior to the computational module of neural network. The training and testing matrices along with the number of input layers is passed to the neural network classifier as shown in Figure 1:

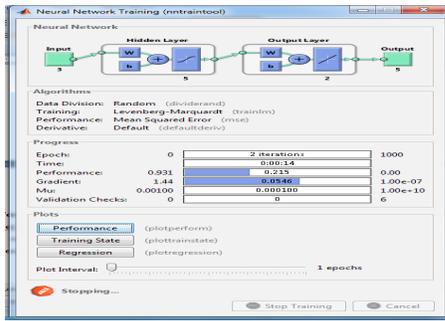


Figure 1. Neural network window.

3.1 Performance Plot

The following graph shows the production of user system and also displays the Mean Square Error (mse) of training, testing and validation. In the below graph blue color shows training, red color displays testing and green line display validation. In Figure 2 Best validation performance is shown by lowest means square and the best match in this graph is at 11th epoch.

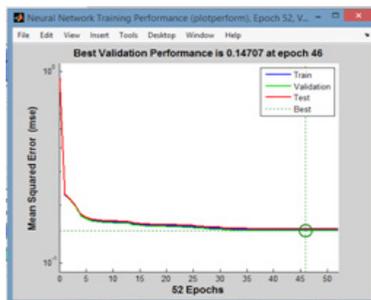


Figure 2. Performance Plot.

3.2 Confusion Matrix

For training, testing and validation performance of the full system, this matrix introduce the real and the forecast value of the network. In this Figure 3 matrix large number of right response is mentioned by green square and wrong is given by red square. Overall performance is displayed in blue square.

3.3 ROV Curve

This is Receiver Operating Characteristic (ROC) curve. In the case of this problem network work very well. In all axes, ROC curve is representing by colored lines. This is the graphical rendering of the real positive rate (sensitivity) versus the wrong positive rate (1 - specificity) as the threshold is changed. With 100% specificity and 100% sensitivity, ideal test display points in upper-left corner is shown in Figure 4.

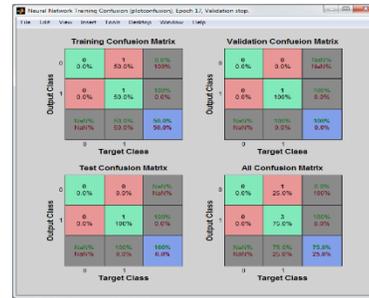


Figure 3. Confusion Matrix.

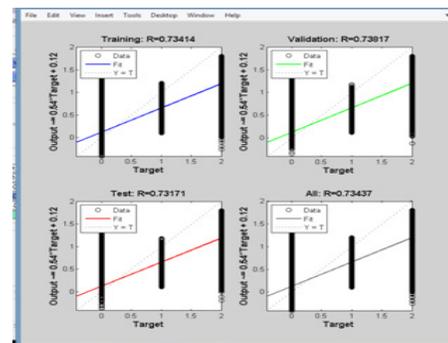


Figure 4. ROV Curve.

3.4 Training State

At each loop this displays the gradient value of back propagation. A very near point of goal is displayed by Gradient value. Loops whose Mean Square Error values are enhanced, they are known as validation fails loops. After 6 regular validation fails, MATLAB automatically stops working is shown in Figure 5.

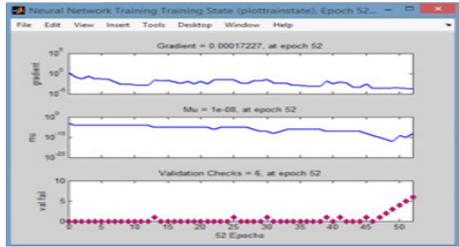


Figure 5. Training State.

3.5 Error Histogram

Graphs displays blue color for training, red color for testing and green color for validation. The testing and training errors are shown in first bin. In last bin validation error occur and on 7th bin one training error occur is shown in Figure 6.

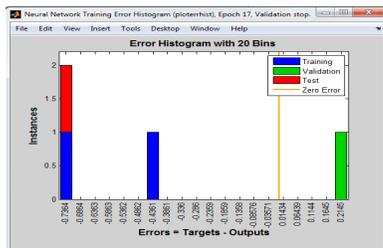


Figure 6. Error Histogram.

4. Conclusion

The proposed model is based upon the probabilistic user profile learning using the neural network for the collaborative learning. The user profile learning method is utilized for the purpose of determining the missing entities in the user's profile over the digital library portals. The proposed recommendation algorithm utilizes the user's browsing history to prepare the testing data for the neural network classifier. The digital library user base is utilized for the preparation of the training data for the proposed model. The other user's recommender engine profile parameters have been obtained from the recommendation engine's user profiles. The similar profiles are shortlisted according to the similarity rank produced after the neural network classification. Afterwards, the averaging factor is applied over the shortlisted rows from the training data and the missing profile entities of the current user's profile are fulfilled and the recommendation lists are prepared according to the similarity ranking produced between the user's profile entities and the e-book

factorization. A variety of the chain experiments has been conducted for the in-depth evaluation of the proposed model design. The experimental results have proved the efficiency and robustness of the proposed model in preparing the semantic lists for the product ranking over the digital library systems in the terms of gradient, time complexity and confusion matrix parameters obtained from the neural network results.

5. References

- Green DT, Pearson JM. The examination of two web site usability instruments for use in B2C Online Libraries organizations. *Journal of Computer Information Systems*. 2009; 49(4):19–32.
- Scioscia S, Floriano F, Ruta M, Loseto G, Gramegna F, SaverioIeva S, Pinto A, Sciascio ED. A Mobile Matchmaker for the Ubiquitous Semantic Web. *International Journal on Semantic Web and Information Systems (IJSWIS)*. 2014; 10(4):77–100.
- Wang T, Lin Y. Accurately predicting the success of B2B ecommerce in small and medium enterprises. *Expert Systems with Applications*, published by Elsevier. 2009; 36(2):2750–8.
- Verma V, Neha N, Malhotra D, Malhotra M, Singh J. Online Libraries Website Recommendation Using Semantic Web Mining and Neural Computing. *Procedia Computer Science ELSEVIER*. 2015; 45:42–51.
- Lazarica M, Lungu I. Aspect eprivind proiect are asistmelor de comert electronic, ASE Publishing House. 2007; 147 pp.
- Mital M, Monika M, AshisPani A, Ramesh R. Determinants of choice of semantic web based Software as a Service: An integrative framework in the context of e-procurement and ERP. *Computers in Industry*. 2014; 65(5):821–27.
- Kavitha R. Collection Development in Digital Libraries: Trends and Problems. *Indian Journal of Science and Technology*. 2009 Dec; 2(12):1–6.
- Tejeda-Lorente T, Alvaro A, Porcel C, Peis E, Sanz R, Herrera-Viedma H. A quality based recommender system to disseminate information in a university digital library. *Information Sciences*. 2014; 261:52–69.
- Sessoms S, Matthew M, Anyanwu K. Enabling a Package Query Paradigm on the Semantic Web: Model and Algorithms. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems*. Springer Berlin Heidelberg, 2014; 8420:1–32.
- Hepp H, Martin M. The Web of Data for Online Libraries: Schema. org and Good Relations for Researchers and

- Practitioners. In *Engineering the Web in the Big Data Era*, Springer International Publishing. 2015; 146(7):723–27.
11. Reji PI, Dharun VS. License plate detection and recognition using vertical based edge detection algorithm and radial basis function neural network. *Indian Journal of Science and Technology*. 2015 Oct; 8(26):1–5.
 12. Yim J, Joo J. Smartphone Sensor Value Pattern Analysis with Neural Network. *Indian Journal of Science and Technology*. 2015 Oct; 8(26):1–7.
 13. Malhotra M, Dhairya D. Intelligent web mining to ameliorate Web Page Rank using Back-Propagation neural network. In *Confluence the Next Generation Information Technology Summit (Confluence)*, 2014 5th International Conference IEEE, Noida. 2014. p. 77–81.