K-Nearest Neighbor (kNN) based Product Ranking Model with Minimum Cost Evaluation

Manpreet Kaur¹, Dr. Raman Chadha²
¹Research Scholar, ²Professor and Head,
CGCTC, Jhanjeri, Mohali
¹manu.ks90@gmail.com, Mohali, ²dr.ramanchadha@gmail.com

Abstract—The product ranking model is the model to create the ranking lists over the product listing data obtained from the e-commerce portals. The product ranking models are utilized to produce the variety of product ranking lists in order to fulfill the requirement to show the multiple lists over the e-commerce ranking lists. The e-commerce portals use the variety of lists to increase the usability and the accessibility by creating the maximum visibility of the output results, which are shown for the similar product ranking, products suggested by other users, most selling products, newly added products, etc. A perfect product ranking model must be capable of producing the content-based and collaborative indexing over the input data. The content-based product ranking models are utilized to produce the product ranking by evaluating the input query by the users. The collaborative filtering is the algorithm used to create the product ranking lists according to the user’s profile and other similar users, friends, etc. In this thesis, the proposed model has been designed by using the effective combination of the content-based and collaborative filtering methods for the realization of the robust product ranking model.

Indexed Terms—Product ranking, e-commerce ranking, k-nearest neighbor, non-probabilistic classification

I. INTRODUCTION

E-commerce stands for Electronic commerce. It is that the shopping for and commercialism of products and services employing a network resembling web. E-commerce is employed in on-line searching, bill payments, ebooks, e-mail, etc. The businesses offer these services square measure Amazon, flipkart, quikr, snapdeal, olacabs and paytm. E-commerce is principally used as a result of it reduces dealings value and improve level of client services. It conjointly has worldwide market accessibility. Numerous classes of e-commerce are: business-to-business, business-to-customer, customer-to-customer or customer-to-business. The ranking system is extremely helpful particularly in e-commerce. A number of the advantages of ranking are:

- It helps the shoppers to choose that choice to be chosen among the multiple choices and in less quantity of your time.
- The personalised e-commerce ranking system has redoubled the trust and loyalty relationship between customers and web site homeowners.
- Provides comparison between totally different corporations and websites that increase competition between the web site homeowners which ends up in higher and improved services.

Ranking product may be a method to rank numerous product. First off one product is graded then grouping of product are performed and also the teams of product is graded. Let’s say let’s take a product Mobile phones, they there are sorted consistent with their corporations like Samsung, Nokia, Apple, HTC etc. And that they are sub sorted consistent with their models and at every step ranking is performed.

Semantic internet is employed to store knowledge within the system while not the steering of human and might be simply clear to humans in type of websites. In linguistics internet knowledge stores are created on internet. It develops the common framework that allows to share the information among corporations, communities or applications. The linguistics internet is amalgamated with the ranking system and during this dynamic consolidation is performed. In dynamic consolidation, if a brand new product is entered in an exceedingly graded list then, it'll mechanically generate the rank of the merchandise by examination it with different product within the list and then it updates the ranking.

II. LITERATURE REVIEW

Hepp, Martin et. Al. Has worked for the e-commerce notably schema.org and good relations for researchers and practitioners on the online of the info. Within the paper, the author has given the introduction and first steerage on the abstract structure of schema.org. They need created the patterns for demand and possession that embody the range of things like piece of furniture, apparels, natural philosophy devices, cosmetics, books, etc. And have created a full tool chain for manufacturing and overwhelming the actual information. The author have conjointly mentioned the subject like authentication (e.g. With webid), identity, access control; information management problems from the publisher and client perspective and micropayment services. The drawback of this application is restricted to micro-data that isn't applicable for e-commerce product ranking system.

Sessoms, Matthew, and Kemafor Anyanwu has worked model and algorithms for sanctionative a Package question criterion on the linguistics net. The package question is that the combination of multiple queries that helps to induce resource combination on linguistics net. The taxonomic category of such queries is “skyline package queries”. In distinction to package queries on one relative models, the RDF model have injected the challenge of deciding the skyline package of ternary relations over multiple joins. The various combination of latest operators for skyline package queries relative question operators and RDF information storage models have developed the four ways for analysis. The author lacks within the use of further techniques for optimization love prefetching additionally because the integration of top-k techniques.
Mital, Monika et. Al. has used the back propagation neural network on Intelligent net mining to upgrade website rank. Due to the rise in data resources, the online is developing at quicker rate. However its large size will increase the difficulties throughout the analysis method in extracting the desired data from net. This drawback are often overcome by exploitation the personalised net search however the user should provide his personal data to keep up privacy. Within the paper, the author addresses all the on top of mentioned problems by exploitation the back propagation neural network for implementing page ranking method.

Mital, Monika et. Al. have planned The integrative framework within the context of e-commerce and ERP to spot determinant of selection for saas. During this paper, the author has tried to classify, determine and rank the size that are influencing saas sourcing call. Exploitation extended AHP (analytic hierarchy process) technique, the framework is analyzed that helped in distinguishing quality and prices so weights criteria are known exploitation the info that was collected by eight users and nine service suppliers of saas supported ERP and e-procurement.

III.EXPERIMENTAL DESIGN

The proposed model has been designed for the product ranking and listings over the product data obtained from the e-commerce portals. The multivariate ranking of the products plays the vital role in the listing of the products. The products are required to be listed in the multiple types according to the requirement. The proposed model has been primarily defined in the two major stages for the product ranking according to the content based ranking and user’s preference based ranking. The first phase of the ranking is computed on the basis of the product properties and popularity given in the form of product rating, manufacturer rating, popularity and other similar factorization. The proposed model evaluates the product entities on the basis of the individual product properties as per given in the database extracted from the online source. The proposed model evaluates the product properties in the factorization method called popularity, accessibility and trust (PAT) assessment or evaluation. The proposed model design for the content based ranking based upon the PAT features has been enlisted in the following algorithm design:

Algorithm 1: The content based product ranking

1. Acquire the input data
2. Extract the popularity, access and trust features for all products
3. Apply the averaging mean factor (AMF) over the popularity, access and trust features for each product entry
4. Add all of the AMF factors to create the cumulative mean
5. Update the product list with the cumulative means
6. Apply the content ranking based upon the cumulative mean value
7. Sort the product entries in the database according to the cumulative mean value
8. Return the ranking list

Algorithm 2: Collaborative user-similarity based ranking

1. Acquire the product data with PAT features
2. Acquire the current user’s browsing history
3. Compute the weighted mean based upon the browsing history
4. Construct the weighted mean vector from the product ranking data
5. Acquire the all user mean vector database (AUMVD)
6. Initiate the k-nearest neighbor model factors
7. Run the iteration for all entries in the AUMVD
   a. Obtain the current user’s mean vector
   b. Obtain the mean vector from AUMVD on current index
   c. Compute the squared distance between the two vectors in (a) and (b)
   d. Update the similarity data matrix (SDM)
   e. Update the kNN classification result data
   f. If the last entry on AUMVD
      i. Return the iteration
8. Sort SDM in the descending order
9. Extract the N user list from the top
10. Acquire the overall weighted mean vector (OWMV) from the top N users
11. Run the iteration for all entries in the product list
    a. Obtain the current product vector
    b. Obtain the OWMV
    c. Compute the squared distance between the two vectors in (a) and (b)
    d. Update the Product Similarity Matrix (ProSiM)
    e. Update the kNN classification result data
    f. If the last entry on product list
       i. Return the iteration
12. Sort SDM in the descending order
13. Return the ProSim Matrix as the result

IV.RESULT ANALYSIS

The proposed model named PAT model, which has been compared against the existing model of product ranking based upon web parameters over the given e-commerce product list in this simulation scenario with similar structure and environment. The detailed analysis of the simulation results has been conducted by analyzing the results obtained from the existing and proposed model simulations. The proposed model has been described efficient and effective while evaluated on the basis of the projected resources and entropy for the coalition of the network load and uniqueness respectively. Minimum of the 10 percent improvement has been observed in the favor of the proposed model when compared on the basis of various performance parameters in the variety of simulation scenarios.

The adverse situations have been analyzed theoretically and the effective in the results has been observed in the proposed model to mitigate the errors and repetitions with the new product ranking model than the existing model. The newly proposed scheme has been primarily evaluated for the utilization of the resources while using the ranking model the given product data in the simulation. The utilization of the resources, measured in the form of projected resources has been recorded significantly lower than existing model, which clearly indicates the robustness of the proposed model. The following figure 1 indicates the robust performance of the proposed model.
Additionally, the entropy parameter signifies the higher uniqueness level in the ranking data using the proposed model than the existing model. The proposed model evaluation over the entropy has been described in the following figure 2. The high uniqueness of the product table data increases the additional robustness to the ranking models associated with the arrangement of the entries in the perfect form over the given set of data. The result of the proposed model has been observed nearly double than the existing model, which can be clearly indicated by the following figure:

V. CONCLUSION

The individual entity relationship based similarity evaluation has been utilized for the content based filtering, which utilizes the popularity, accessibility and trust (PAT) factors based content filtering algorithm. The PAT features have added the flexible and robust content-filtering model.

The k-nearest neighbor classification model has been utilized for the collaborative filtering model for the evaluation of the similarity of the other users, when the new users are enlisted over the e-commerce engines. The historical data of the existing users is evaluated using the averaging factors to compute the first level test vector,
which is further classified with the k-nearest neighbor model for the production of the collaborative filtering. The proposed model results have been evaluated in the form of various performance parameters associated with the time complexity, duplicate entry evaluation and resource usage. The proposed model has outperformed the existing model on the basis of all of the above listed parameters during the performance evaluation phase.

REFERENCES


