



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume3, Issue3)

Available online at www.ijariit.com

Hybrid Recommendation Model With Nearest Neighbor Classification Based Collaborative Approach

Ramandeep Sharma

Swami Devi Dyal Institute of Engineering & Technology
Ramandeepsharma256@gmail.com

Samarth Kapoor

Swami Devi Dyal Institute of Engineering & Technology
kavi.samarth@gmail.com

Abstract: *The e-commerce product ranking models are designed to handle the very large volumes of the data involved in the database. Formally, the multifactor ranking models are incorporated over the online portals, which are capable of producing the multivariate lists. The product lists are prepared on the basis of various features, which include the popularity, accessibility, and trust based factors which are associated with the e-commerce products for the realization of the content-based filtering over the e-commerce portals. In this paper, the multivariate and multifactor ranking algorithm has been proposed in order to solve the problems related to the low entropy, duplication, and unbalanced feature analysis. The proposed model design is entirely based upon the popularity, visitor density, number of customers and security analysis based factors of the e-commerce pages containing the product lists. The multifactor values are organized in the different columns containing the different kinds of information, which are converted to the normalized and compatible values to create the data uniformity. The proposed model offers the collaborative index-based product ranking model over the dense e-commerce databases. The proposed model has been designed to use the collaborative filtering based upon the k-nearest neighbor algorithm. The proposed model has been undergone the various experiments for the performance evaluation based on the time complexity, resource utilization, and other similar factors.*

Keywords: *Product Ranking, Recommendation Model, Recommender System, Ranking Algorithm.*

INTRODUCTION

Along over two decades of research and commercial development, ranking systems have proved to be a successful technology to overcome the information overload that burdens users in modern online media. The inherent possibility of dealing with diverse sources of information, such as the content of the items and the collaborative and social interactions among users and between users and a system, has enabled the development of rich strategies based on each of these evidence, deriving content-based, collaborative, and social filtering ranking approaches. Furthermore, as each particular type of ranking technique has its own limitations and weaknesses, hybrid strategies have been proposed that combine the suggestions generated by different techniques in different ways. The success of ensemble approaches has been recently evidenced in the Netflix prize, where the top classified teams used different forms of ranking ensembles. There are, however, general limitations remain unsolved and are still considered as open research problems in the field. We have mentioned the sparsity of the information (either in the forms of content-based attributes, collaborative ratings, and social connections), and the new user problem, but other problems, not related to a specific ranking technique, have been identified in the literature, and deserve special attention by themselves, such as the need of contextualization, the explanation of the rankings, and the efficiency in computing rankings.

LITERATURE REVIEW

Neha Verma et. al. have worked on the E-commerce Website Ranking Using Semantic Web Mining and Neural Computing. The authors have discussed the design of the new semantic web-based E-commerce page ranking model based on neural network. Hepp, Martin et. al. has worked on the Web of Data for E-Commerce: Schema.org and GoodRelations for Researchers and Practitioners. In this tutorial, the authors have given a comprehensive overview and hands-on training on the advanced conceptual structures of schema.org for e-commerce, including patterns for ownership and demand, and have present the full toolchain for producing and consuming respective data. Sessoms, Matthew, and Kemafor Anyanwu have worked on enabling a Package Query Paradigm on the Semantic Web: Model and Algorithms. The authors have introduced the concept of a *Package Query* for querying for resource combinations on the Semantic Web. Malhotra, Dheeraj et. al have worked on Intelligent web mining to ameliorate Web Page Rank using Back-Propagation neural network. The speedy expansion of web is enjoyable because of the increase in information resources

but at the same time, its huge size and interference of SEOs in search process lead to increased difficulty in extracting relevant information from the web. Furukawa, Takao et. al. have worked on the identification of the evolutionary process of emerging technologies. The authors have developed a chronological network analysis of World Wide Web conference sessions. In the evolutionary process of emerging web-based technology, the convergent session nodes that recapitulate past research topics and the divergent session nodes in the networks play significant roles in promoting dynamic interactions among research topics. Scioscia, Floriano et. al. have worked on the Mobile Matchmaker for the Ubiquitous Semantic Web. This paper presents Mini-ME, a novel mobile inference engine designed from the ground up for the Swot.

EXPERIMENTAL DESIGN

In general, any product contains the useful, popularity & trust to include the robust ranking information. The system has to differentiate between both of the ranking and popularity systems. Consider the below product where the person reading a book is the useful information and the background, people and the market are the unwanted data. The system has to group together the repeated pattern to identify the objects from the product list.

4.1.1 Product rating model: This model has primary popularity factors like user rating, product rating and other similar factors. Most of the e-commerce portals make use of the accessibility and popularity factors to construct the semantic ranking building under the recommendation systems. This model uses the Cartesian coordinate system for the realization of the product rating model. The popularity factors in this model are called “Additive primaries” because desired popularity factors can be produced by adding them together.

4.2.1 PAT ranking model: PAT product ranking model stands for popularity, accessibility, and trust factor model. This model describes popularity factors in terms of their user rating, product rating, and trust evaluation. This model offers a more intuitive representation of the relationship between popularity factors. Basically, a popularity model is a specification of the coordinate system and a subspace within that, where each popularity factor is represented in single value factor.

4.2.1.1 Popularity: Popularity represents the dominant weight of the product by using the weight of the product on the e-commerce systems. It is the term for the term popularity factors and their evaluation. Popularity is expressed from 0 to 1 when neutralized using the statistically neutralized methodology. It represents popularity of product rating, user rating number of purchases. Eventually, all factors can be mixed from these basic parameters for the hybridization of the factors.

4.2.2.1 Accessibility: Accessibility represents the dominance of the total hits and links from where the product page is accessed or forwarded. Also, the Alexa popularity and Google page rank value adds the additional effort to the product ranking model. It can also be thought as the intensity of the product hits and it's floating around the web in terms of anchors hits. It is defined as the degree of accessibility and number of links. A highly popular object is a total hit, whereas a highly low accessibility makes the product a total flop one. When there is no popularity in the product, it is considered as the fresh listing or un-accessed listing.

4.2.3.1 Trust: The trust of the product describes the intensity of the security of the product listed on the e-commerce listing. In other words, the trust factor is defined as a relative security factor for the product page, where the hackers cannot add the fake statistics, overhauled popularity and the overloaded anchor links across the web. The trust of the product is given by the point of security between the product and the line connecting the product with its users and the security parameters associated with the product hosted in the database.

4.3.1 Popularity, Accessibility, and Trust based product ranking model: PAT factors are said to lie within a triangle of popularity, accessibility and trust of the listed products, whose vertices are defined by the three primary popularity factors in PAT model. The popularity is given by P, A by accessibility and T for trust factor for the product listing methods for the product rating model.

Algorithm 1: Product ranking model based on PAT

1. Acquisition of Query Product
 2. Obtain the PAT values
 3. Formulate the Co-occurrence Matrix (CCM) for containing the PAT factors
 4. For every product in the listing
 - a. Acquisition of the popularity factors
 - b. Acquisition of the accessibility factors
 - c. Acquisition of the trust factors
 - d. Normalize the individual PAT indices for averaging factor.
 - e. Formulate CCM
 - f. Compare and PAT CCM matrix with the query or the product in focus
 - g. Calculate the matching factors for the product listing using Product Matching Index (PMI).
 5. Reset PMI in the descending order for the most to the least popular product listing.
 6. Return the product ranking list (PRL) to the system.
-

The existing model has been improved by using the hybrid approach for the purpose of recreation of the robust ranking model. The local and global parameters have been utilized using the PAT (popularity, accessibility, and trust) based factors. The k-nearest

neighbor algorithm has been utilized for the classification of the users under the collaborative approach. The proposed collaborative approach has been utilized for the user similarity matrix for the purpose of finding the neighbor entity relationship. The following algorithmic steps have been followed to achieve the k-NN classification goal over the PAT features:

Algorithm 2: PAT based k-NN classification for the user similarity evaluation

1. Restructure the PAT based data ranking list
2. Pass the evaluation rate to the k-NN method
3. Get the size in number of rows of the ranking list
4. If matrix does not match the input sequence
 - a. Return the program with the collaborative ranking
5. Acquire the user history information from the web access log database
6. If the user history is empty
 - a. Return the program with the collaborative ranking
7. Acquire the user training data (UTD)
8. Load the historical data into the runtime memory
9. Perform the averaging factor over the historical data
 - a. Return the virtual user profile vector (VUPV)
10. Start the iteration for similar evaluation
11. Acquire the averaged PAT feature for the current row
12. Acquire the virtual user profile vector
13. Evaluate the individual distance between individual components
 - a. Return the similarity matrix
14. Compute the cumulative phase distance between the VUPV and UTD
15. Update the kNN similarity matrix
 - a. Return the similarity matrix
16. Obtain the most similar candidates from the similarity vector
17. Perform the averaging factor to prepare the representative matrix (RM)
18. Re-compute the content based ranking list against the RM
19. Restructure the content based ranking list
20. Return the final product ranking matrix

RESULT ANALYSIS

5.1 Results for scenario 1 with 100 test cases

Projected resources: The projected resource has been evaluated for the measurement of the utilization of the resources over the given ranking model for the e-commerce portals. The high performance is indicated by the lower value of the projected resources computed from the simulation environment and higher value indicates the lower performance. The proposed model has been considered better than the existing model as it has been measured with the lower value for projected resources over the given simulation scenario. The comparison of the evaluated results has been performed over the results obtained from the existing and proposed models. The performance evaluation has been performed on the basis of projected resources and entropy. Proposed model has been proved itself as the better model than existing ranking model. The proposed model has been proved to be efficient than the existing model on the basis of both the performance parameters in the first scenario with 100 product entries.

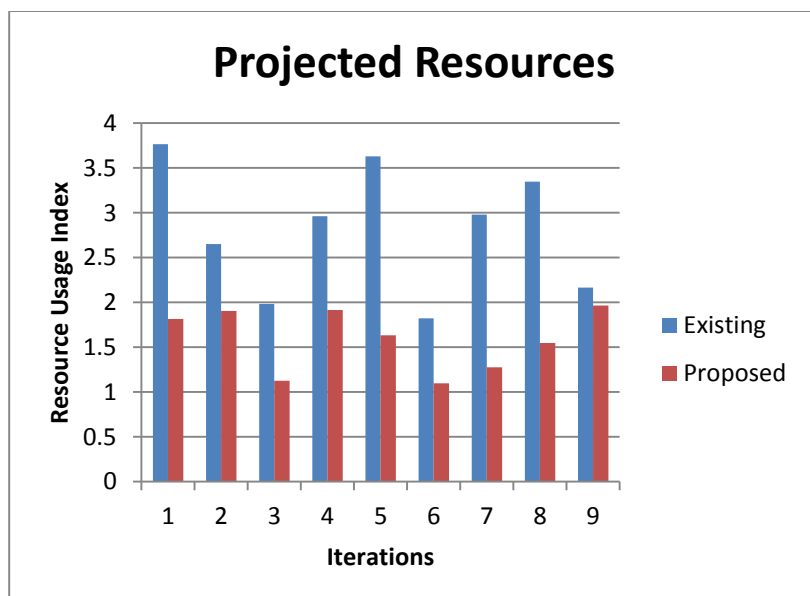


Figure 5.1: Projected Resources based comparison for scenario 1

Entropy: The ranking efficiency, size of the population and the uniqueness of the entities in the given product list is measured by using the entropy parameter. The unique data decreases the risk of repeated entries in the semantic ranking lists, which has been strongly observed from the proposed model simulation. The consistently high entropy justifies the strength of the proposed ranking model. The detailed results for entropy can be seen below:

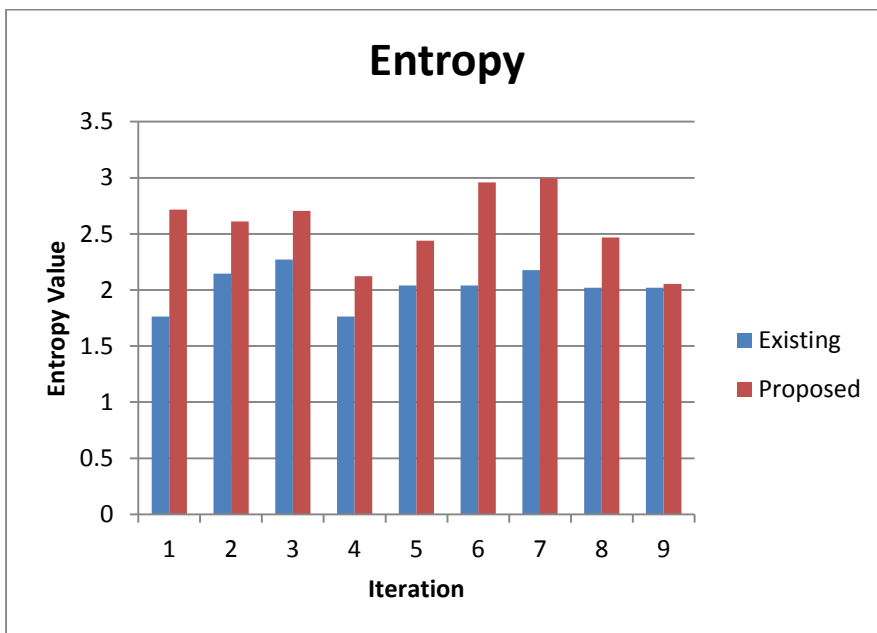


Figure 5.2: Entropy-based comparison for scenario 1

5.2 Results for scenario 2 with 500 test cases:

Projected resources: The high performance in this scenario is indicated by the lower value of the projected resources computed from the simulation environment and higher value indicates the lower performance. The proposed model has been considered better than the existing model as it has been measured with the lower value for projected resources over the given simulation scenario of semantic ranking over the e-commerce portals. The comparison of the evaluated results has been performed over the results obtained from the existing and proposed model with 500 product entries. The performance evaluation has been performed on the basis of projected resources and entropy. The proposed PAT model has been proved itself as the better model than existing models. The PAT model has been proved to be efficient than existing on the basis of both the performance parameters.

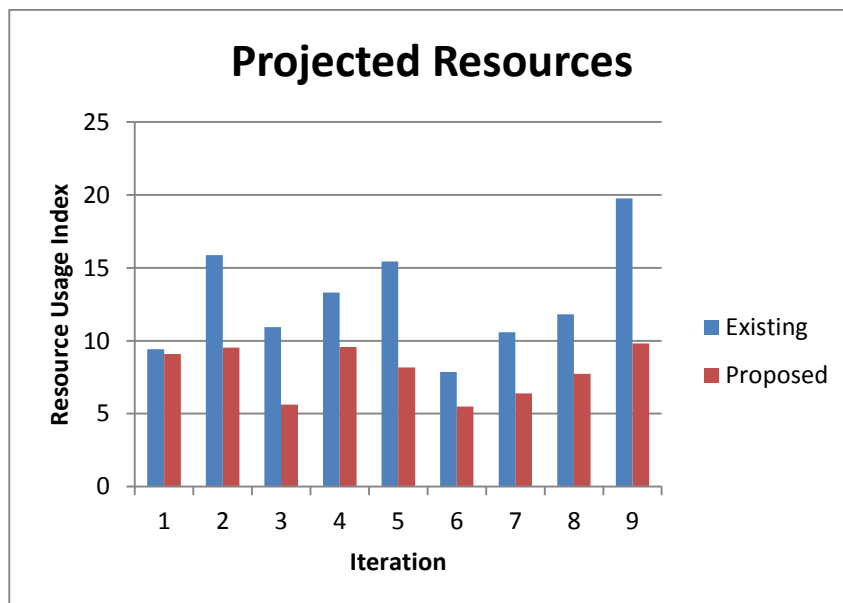


Figure 5.3: Projected Resources based comparison for scenario 2

Entropy: The uniqueness observed from the obtained results has signified the decrease in the probability of the repetitive entries in the given product list, which has been strongly observed from the proposed model simulation. The consistently high entropy in the proposed in comparison with existing model justifies the strength of the proposed semantic ranking model. The detailed results for entropy can be seen below:

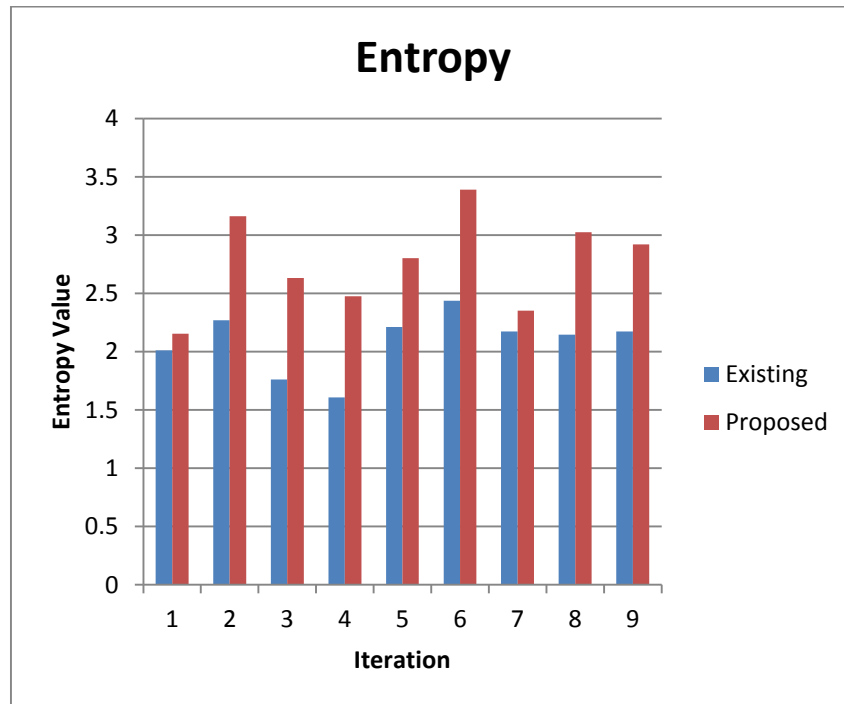


Figure 5.4: Entropy-based comparison for scenario 2

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

- [1] Verma, Neha, Dheeraj Malhotra, Monica Malhotra, and Jatinder Singh. "E-commerce Website Ranking Using Semantic Web Mining and Neural Computing." *Procedia Computer Science* 45 (2015): pp. 42-51, ELSEVIER.
- [2] Hepp, Martin. "The Web of Data for E-Commerce: Schema.org and GoodRelations for Researchers and Practitioners." In *Engineering the Web in the Big Data Era*, pp. 723-727. Springer International Publishing, 2015.
- [3] Sessoms, Matthew, and Kemafor Anyanwu. "Enabling a Package Query Paradigm on the Semantic Web: Model and Algorithms." In *Transactions on Large-Scale Data-and Knowledge-Centered Systems XIII*, pp. 1-32. Springer Berlin Heidelberg, 2014.
- [4] Malhotra, Dhairya. "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 Conferences*, pp. 77-81. IEEE, 2014.
- [5] Furukawa, Takao, Kaoru Mori, Kazuma Arino, Kazuhiro Hayashi, and Nobuyuki Shirakawa. "Identifying the evolutionary process of emerging technologies: A chronological network analysis of World Wide Web conference sessions." *Technological Forecasting and Social Change* 91 (2015): 280-294.
- [6] Scioscia, Floriano, Michele Ruta, Giuseppe Loseto, Filippo Gramegna, Saverio Ieva, Agnese Pinto, and Eugenio Di Sciascio. "A Mobile Matchmaker for the Ubiquitous Semantic Web." *International Journal on Semantic Web and Information Systems (IJSWIS)* 10, no. 4 (2014): 77-100.
- [7] Mital, Monika, Ashis Pani, and Ram Ramesh. "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* 65, no. 5 (2014): 821-827.
- [8] D.T. Green and J. M. Pearson, "The examination of two website usability instruments for use in B2C e-commerce organizations," *Journal of Computer Information Systems*, Vol. 49, No. 4, 2009, pp. 19-32

- [9] T. Wang and Y. Lin, "Accurately predicting the success of B2B e-commerce in small and medium enterprises," *Expert Systems with Applications*, Vol. 36, No. 2, published by Elsevier, 2009, pp. 2750–2758.
- [10] M. Lazarica and I. Lungu, *Aspecte privind proiect area sistemelor de concert electronic*, ASE Publishing House, 2007, pp. 147.
- [11] <http://www.networkcomputing.com/917/917f2.html>
- [12] I. Ivan, L. Săcuiu and D. Milodin, "Integrarea aplicațiilor de comerț electronic," *Revista Română de Informatică și Automatică*, Vol. 19, No. 2, 2009, pp. 23 – 42.
- [13] Y. Lee and K. Kozar, "Investigating the effect of website quality on e-business success: An analytic hierarchy process (AHP) approach," *Decision Support Systems*, Vol. 42, No. 3, published by Elsevier, 2006, pp. 1383–1401.
- [14] S. Värlander, "Online information quality in experiential consumption: An exploratory study," *Journal of Retailing and Consumer Services*, Vol. 14, No. 5, 2007, pp. 328-338.
- [15] E. K. Delice and Z. Gungor, "The usability analysis with heuristic evaluation and analytic hierarchy process," *International Journal of Industrial Ergonomics*, Vol. 39, No.6, 2009, pp. 934–939.