

Enhanced Swarm based Optimized Recommendation System

Shivani Sharma
M-Tech IT
CDAC, Mohali

ABSTRACT

Recommendation system is needed to provide personalized information to the user to improve the searching of the user based upon their searching method. In this work, an architecture that integrates product information with user's access log data and then generates a set of recommendations for that particular user. This architecture uses more than one data mining algorithms like clustering and pattern matching algorithms. In previous work, for Clustering they are using K-mean Clustering & other algorithm for pattern matching is Boyer Moore Pattern Matching Algorithm. To enhance this work, an optimization algorithm will be used. Here Particle Swarm optimization algorithm has been used which is a soft computing algorithm of Artificial Intelligence. It will help to improve the results on the basis of intelligence. In this work a database of mobiles is created manually and based on there technical specifications better mobiles are recommended to the user based on intelligence. It will firstly create clusters on the basis of some similarity and then based on random population and technical specifications probability will be calculated. The product with least probability will be kept on top and then sort in top to bottom order.

Keywords

Data Mining, Recommendation system, KNN, PSO, Precision, Recall

1. INTRODUCTION

The ascent in amount of information of learning over web in past couple of years has brought on the developing danger of information flooding that progressively has made the matter of getting to pertinent information to the clients. Conjointly with the climb in assortment of web locales and destinations, website admin notice it hard to plan the substance as per the client's need. The information interest of web clients are regularly worked out by assessing client's web route conduct. Web Usage Mining (WUM) is utilized to concentrate information from net client's entrance logs by utilizing information preparing procedures. One in every one of the uses of WUM is Recommendation framework that is customized data separating system acclimated either check regardless of whether an unmistakable client can support a given thing or to detect a posting of things which may be of awesome importance to the client.

Recommender structures help customers find and assess things of venture. Proposal Systems can use information digging methodologies for making recommendations using taking in picked up from the movement and characteristics of the customers. The design of an online web suggestion framework taking into account web use mining fundamentally comprises of three stages: Data Pre-handling, Pattern discovery and creating proposals. Information pre-preparing and Pattern recognition stages are performed logged off and the proposals are created on the web. Information pre-preparing includes changing the web access logs and client

profiles into arrangement fitting for the framework. Design identification includes utilizing information mining systems like bunching, successive example mining or affiliation principle mining. Current suggestion frameworks show certain impediments, for example, insight, versatility, adaptability, restricted exactness. The impediment of these frameworks is that the new client gets the suggestions just on the premise of his present route.

There are many definitions of recommender systems. One of the first was presented by Paul Resnick and Hal R. Varian in 1997. They claim that "in a typical recommender system, people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients" [1]. These systems are usually defined in terms of their functionality as the systems or agents that suggest the products to the users who purchase products on e-commerce sites [3]. The recommender systems help the consumer to make the decision what to buy. The recommender systems can be classified because of the level of personalization into non-personalized and personalized methods (Figure1). The former methods do not consider the characteristics and preferences of the customers, whereas the latter tightly depends on the user profile.

The main goals of the recommender systems are as follow:

- To cope with information overload
- To help all customers (new, frequent, and infrequent) to make decisions what products to buy, which news to read next, which movie is worth watching, etc.
- To convert observers to buyers
- To build credibility through community and maintain the loyalty of the customers
- To inviting customers to come back
- To enhance e-commerce sales and cross-sell

Recommender Systems collect and exploit various types of information about users, products, and interactions between users and products to generate a personalized list of items that fits user's current needs. This information may be gathered while the user is interacting with the system or can be acquired from other systems the user has interacted with. The most commonly used information about the interaction between users and products is the set of subjective ratings assigned by the users to previously experienced products. The system then uses these ratings for predicting the ratings for products not yet experienced. The products with the highest estimated ratings are then recommended to a user.

2. PROPOSED METHOD

The ascent in measure of data over web in most recent couple of years has brought on the developing danger of data flooding which thusly has made the issue of getting to important information to the clients. Additionally with the climb in number of sites and site pages, website admins think

that its testing to detail the substance as per the client's need. The data interest of the online clients can be made sense of by assessing client's web route conduct. Web Usage Mining (WUM) is utilized to concentrate information from Web clients access logs by utilizing Data Mining Techniques. One of the uses of WUM is Recommendation framework which is customized data separating system used to either figure out if a specific client will endorse a given thing or to distinguish a rundown of things which can be of noteworthy significance to the client.

Recommender systems help customers find and assess things of speculation. Proposal Systems can use information digging techniques for making recommendations using taking in picked up from the movement and characteristics of the customers. The design of an online web proposal framework in view of web use mining essentially comprises of three stages: Data Pre-handling, Pattern identification and producing suggestions. Information pre-preparing and Pattern discovery stages are performed logged off and the proposals are created on the web. Information pre-handling includes changing the web access logs and client profiles into configuration fitting for the framework. Design discovery includes utilizing information mining procedures like bunching, consecutive example mining or affiliation standard mining. Current proposal frameworks display certain constraints, for example, knowledge, versatility, adaptability, restricted precision. The drawback of these frameworks is that the new client acquires the suggestions just on the premise of his present route.

The past work which considered defeats these hindrances by executing a half and half design that incorporates item data with client's entrance log information and afterward produces an arrangement of suggestions for that specific client. In this work, an engineering that coordinates item data with client's entrance log information and after that produces an arrangement of proposals for that specific client. This design utilizes more than one information mining calculations like grouping and example coordinating calculations. For Clustering they use K-mean Clustering & other algorithm for pattern matching is Boyer Moore Pattern Matching Algorithm. To enhance this work, an optimization algorithm will be used. Here Particle Swarm optimization algorithm has been used which is a soft computing algorithm of Artificial Intelligence. It will help to improve the results on the basis of intelligence.

2.1 Proposed Model

The proposed model spotlights on the above four goals which are useful in enhancing the productivity of the framework and are for all intents and purposes actualized utilizing C#.Net environment.

In this recommendation system a dataset of 77 different phones has been utilized in total and user always get 5 recommendations when he/she will select any phone. Results of the proposed framework are calculated in terms of precision, recall and F-measure.

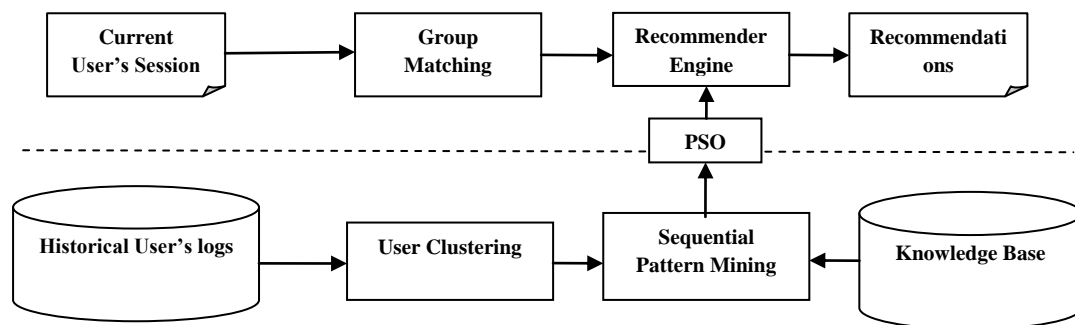


Fig. 1: Basic Design of the System

Fig. 1 shows the basic design of the system where User's historical information is divided into clusters using KNN clustering algorithm. These clusters are matched with knowledge base using boyer's moore pattern matching algorithm. These matched patterns are passed through optimized where Particle Swarm Optimization is used for pattern matching and passed these outputs to recommender engine so that users gets answer for his/her query.

2.2 Algorithm Level Design

The Algorithm design involves the following steps:

User logs → Clustering → K-Mean

1. Cluster the data into K groups
2. Select k point at random as cluster centres
3. Assign objects to their closest cluster centre according to Euclidean distance formula
4. Calculate the centroid or mean off all objects in each cluster
5. Repeat step 2,3,4 until the same points are assigned to each cluster in consecutive rounds

Knowledge Base →

Input text with n character and pattern with m character.

1. Compute function last
2. $I \leftarrow m-1$
3. $J \leftarrow m-1$
4. Repeat
 - If $P[j]=T[i]$ then
 - If $j=0$ then return i
 - Else
 - $I \leftarrow i-1$
 - $J \leftarrow j-1$
 - Else
 - $I \leftarrow i+m-\min(j,i+\text{last}[T[i]])$
 - $J \leftarrow m-1$
 - Until $i > n-1$
 - Return no match
5. If matched

PSO → Best Match

- I. For each particle(matches return by Boyee Moore initialize particle
- II. Do
 - a) For each particle
 - Calculate Fitness Values
 - If the fitness value is better than the best fitness value(pbest) in history.
 - Set current value as the new pbest.
 - end
 - b) .For Each Particle:
 - Find in the particle neighbourhood the particle with best fitness.
 - Calculate particle velocity
 - Apply the velocity constriction
 - Update particle position
 - Apply position constriction

While max iteration criteria is not attained.

3. PERFORMANCE AND RESULT ANALYSIS

This recommendation system containing 77 different phone samples in total and user always get 5 recommendations when he/she will select any phone. For theoretical analysis this case study has been analyzed. In this 10 users has been used both base type & enhanced recommendation system. In base type system, k-mean clustering is been used for recommendation but in enhanced this clustering selection has been filtered or optimized using particle swarm optimization which is based on artificial intelligence. Table 1 gives the analysis user wise where TP, FP, FN and TN rates has been analyzed. TP defines the true positive rate which means the number of phones which are rated good and also that are actually good whereas FP defines the false positive rate which means items rated good but actually that are bad and not to be recommended. The other are FN that is false negative which means items are actually good but treated as bad and TN means true negative which treated the things bad which are actually bad.

Table 1: User Wise TP, FP, TN and FN rates

Users	TP		FP		FN		TN	
	BASE	ENHANCED	BASE	ENHANCED	BASE	ENHANCED	BASE	ENHANCED
User 1	3	4	1	0	0	0	1	1
User 2	2	3	1	0	1	0	1	2
User 3	4	4	0	0	0	0	1	1
User 4	3	3	0	1	1	1	1	0
User 5	4	3	0	0	1	0	0	2
User 6	3	2	0	1	0	0	2	2
User 7	2	2	1	0	1	0	1	3
User 8	4	4	0	0	0	1	1	0
User 9	3	4	0	0	1	0	1	1
User 10	2	4	1	1	1	0	1	0

Table 2 gives the total count of TP, FP, FN and TN rate for all users on the basis of Rated Good and Actually Good and Rated Bad and Actually Bad.

Table 2: TP, FP, FN and TN rate

	Actually Good		Actually Bad	
	K-mean clustering Based	Enhanced PSO based	K-mean clustering Based	Enhanced PSO based
Rated Good	30	33	4	3
Rated Bad	6	2	10	12

Precision

In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the query:

$$Precision = \frac{| \{relevant\ documents\} \cap \{retrieved\ documents\} |}{| \{retrieved\ documents\} |}$$

Accuracy considers all recovered archives, yet it can likewise be assessed at a given cut-off rank, considering just the highest results returned by the framework.

$$Precision = \frac{tp}{tp + fp}$$

Recall

Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\text{Recall} = \frac{| \{ \text{relevant documents} \} \cap \{ \text{retrieved documents} \} |}{| \{ \text{retrieved documents} \} |}$$

For instance for content hunt on an arrangement of records review is the quantity of right results separated by the quantity of results that ought to have been returned

$$\text{Recall} = \frac{tp}{tp + fn}$$

F-Measure

In factual investigation of parallel arrangement, the F_1 score (likewise F-score or F-measure) is a measure of a test's exactness. It considers both the exactness p and the review r of the test to register the score: p is the quantity of right positive results partitioned by the quantity of all positive results, and r is the quantity of right positive results separated by the quantity of positive results that ought to have been returned. The F_1 score can be deciphered as a weighted normal of the accuracy and review, where a F_1 score achieves its best esteem even from a pessimistic standpoint at 0. The customary F-measure or adjusted F-score (F_1 score) is the symphonious mean of accuracy and review:

$$F - \text{Measure} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

Comparative Analysis

Table 3 shows the results value of the comparative analysis between K-mean clustering based and Enhanced PSO Based on the basis of **Precision ,recall, F-Measure**

Table 3: Comparative analysis

	K-mean clustering Based	Enhanced PSO based
Precision	0.88	0.92
Recall	0.83	0.94
F-Measure	0.854	0.924

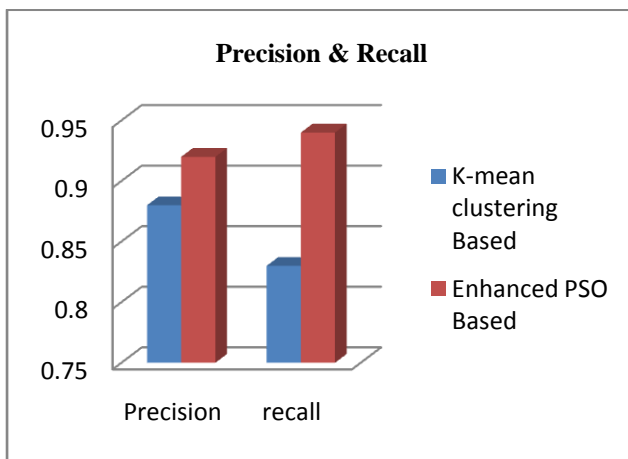


Fig. 2: comparison of Precision and recall

Fig 2 shows the comparative of precision and recall values between the two methods K-mean clustering based and Enhanced PSO Based and Fig. 3 shows the F-measure based comparative analysis of K-mean clustering based and Enhanced PSO Based method.

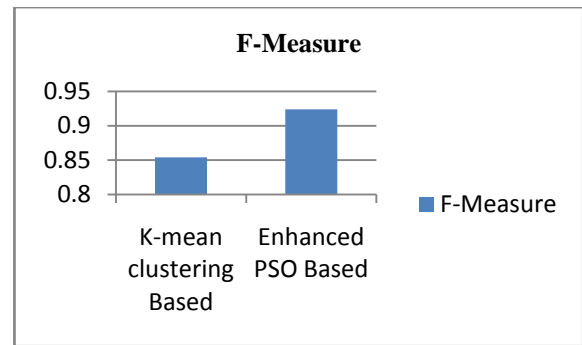


Fig. 3: Comparison of F-measure

4. CONCLUSION

Recommender frameworks help clients discover and evaluate things of investment. Recommendation Systems can utilize data mining strategies for making suggestions utilizing learning gained from the activity and qualities of the clients. In this work, an architecture that integrates product information with user's access log data and then generates a set of recommendations for that particular user. This architecture uses more than one data mining algorithms like clustering and pattern matching algorithms. For Clustering they use K-mean Clustering & other algorithm for pattern matching is Boyer Moore Pattern Matching Algorithm. To enhance this work, an optimization algorithm is used. Here Particle Swarm optimization algorithm has been used which is a soft computing algorithm of Artificial Intelligence. For performance analysis, a case study is taken under consideration and hence results are evaluated. Results concludes that enhanced recommendation system based on PSO improve the accuracy of the system.

Future work could go in the direction to test and analyze the performance of the proposed architecture with different database. In future, hybrid architecture might be added to enhance features of the recommendation system.

5. REFERENCES

- [1] Sneha Y.S, G. Mahadevan, Madhura, :An Online Recommendation System based on web usage mining and semantic web using LCS Algorithm", IEEE 2011
- [2] Sneha Y.S, G. Mahadevan, Madhura, : A Personalized Product Based Recommendation System Using Web Usage Mining and Semantic Web", International Journal of Computer Theory and Engineering (IJCTE) Vol. 4, No. 2, April 2012
- [3] Mehrdad Jalali1, Norwati Mustapha, Md. Nasir B Sulaiman, Ali Mamat, "A Web Usage Mining Approach Based on LCS Algorithm in Online Predicting Recommendation Systems" 12th International Conference Information Visualisation IEEE 2008.
- [4] SuleymanSalim et al "Using Semantic Information for web usage mining based recommendation" International Conference IEEE 2009.
- [5] Xin Sui, Suozhu Wang, Zhaowei Li "Research on the Model of Integration with Semantic Web and Agent Personalized Recommendation System "Proceedings of the 2009 13th International Conference on Computer Supported Cooperative Work in Design.
- [6] A. Apostolico, "String editing and longest common subsequences", Handbook of Formal Languages, Vol. 2.

- Linear Modeling: Background and Application, Springer Verlag, Berlin, chapter 8, 1997, pp. 361–398.
- [7] RVVSV Prasad, V Valli Kumari “A Categorical Review of Recommender Systems” , International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.5, September 2012.
- [8] Hadi Khosravi Farsani, and Mohammadali Nematbakhsh “A Semantic Recommendation Procedure for Electronic Product Catalog”, World Academy of Science, Engineering and Technology 22 2006.
- [9] R.Suguna, D, Sharmila, “Clustering Web log Files - A Review”, International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 4, April – 2013 ISSN: 2278- 0181
- [10] Himangni Rathore, Hemant Verma, “Analysis on Recommended System for Web Information Retrieval Using HMM”, International Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4, November 2014.
- [11] R.Baraglia, F.Silvestri,” An online recommender system for large Web sites”, Web Intelligence, IEEE/WIC/ACM, 2004, pp. 20–24.
- [12] Sneha Y S, G. Mahadevan, Madhura, “A Personalized Product Based Recommendation System Using Web Usage Mining and Semantic Web”, 2011 International Conference on Information and Computer Application (ICICA 2011).
- [13] A. Singh and S. Sharma, “Role of Page ranking algorithm in Searching the Web: A Survey”, International Journal of Engineering & Technology, Management and Applied Sciences, Vol. 1, Issue 1, June 2014.
- [14] P. Bari and P. M. Chawan. "Web Usage Mining", Journal of Engineering Computers & Applied Science, Vol. 2, No. 6, pp. 34-38, 2013
- [15] R. Sharma and K. Kaur, "Review of Web Structure Mining Techniques using Clustering and Ranking Algorithms", International Journal of Research in Computer and Communication, IJRCCCT, Vol. 3, No. 6, pp. 663-668, 2014.
- [16] C. Tsai, C. Lai and M. Chiang, "Data mining for internet of things: A survey," IEEE Communication Surveys & Tutorials, Vol. 16, No. 1, 2014.
- [17] F. S. Gharehchopogh and Z. A. Khalifelu, "Analysis and evaluation of unstructured data: text mining versus natural language processing" Application of Information and Communication Technologies (AICT), 2011 5th International Conference on. IEEE, pp. 1-4, 2011.
- [18] C. Kaur, and R. R. Aggarwal, "Reference Scan Algorithm for Path Traversal Patterns.", International Journal of Computer Applications, Vol. 48. Pp no. 20-25, 2012
- [19] V. Losarwar, M. Joshi, “Data Preprocessing in Web Usage Mining”, International Conference on Artificial Intelligence and Embedded Systems (ICAIES'2012), pp. 15-16, 2012.
- [20] Esra Saraç, Selma Ayşe Özel, “Web Page Classification Using Firefly Optimization” IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA), PP 1-5, 19-21 June 2013.
- [21] Abdelhakim Herrouz, Chabane Khentout, Mahieddine Djoudi, “Overview of Web Content Mining Tools” The International Journal of Engineering And Science (IJES), Volume 2, Issue 6, 2013.
- [22] Kira Radinsky, Eric Horvitz, “Mining the Web to Predict Future Events” http://research.microsoft.com/en-us/um/people/horvitz/future_news_wsdm.pdf.
- [23] Monika Yadav, Mr. Pradeep Mittal, “Web Mining: An Introduction” International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013 .
- [24] Xin-She Yang, Xingshi He, “Firefly algorithm: recent advances and applications” Int. J. Swarm Intelligence, Vol. 1, No. 1, 2013.
- [25] Pikakshi Manchanda, Sonali Gupta, Komal Kumar Bhatia, “IOSR Journal of Computer Engineering (IOSRJCE), Volume 4, Issue 1 (Sep-Oct. 2012), PP 20-25”.