

# Recommendation Framework using Pattern Searching Mechanism, HD Technique, Social Data

Mr. Saurabh R. Deshpande

Department of Information Technology  
Sinhgad Technical Education Society's, SKNCOE,  
Pune, India

Prof. Jyoti R. Yemul

Department of Information Technology  
Sinhgad Technical Education Society's, SKNCOE,  
Pune, India

**Abstract**—Data on the web is growing exponentially. Also, users of internet are dependent on it for their day-to-day activities like internet banking, shopping and many more. For E-Commerce businesses to grow accurate recommendation of product suites is necessary to attract new customers and retain existing ones. Typically, existing recommendation techniques are based on Collaborative Filtering which are dependent on rating data which may be unavailable in most of the cases. Due to this, recommendations generated are less accurate. In this paper, a new recommendation technique is used which will increase accuracy of the recommendations generated. Though accuracy and time taken to generate recommendations varies according to searched query, but, approximately accuracy is improved by 50-55% in proposed system.

**Keywords**—Collaborative Filtering, Data Mining, E-Commerce, Recommendations

## I. INTRODUCTION

“We are living in information age”; everyday data is getting generated exponentially and extracting knowledge out of that data is becoming a challenge. Web data mining is a technology that aims to provide interesting patterns from large amounts of data and knowledge discovery from data. Customers are using E-commerce websites for online shopping which contains insight information about customer behavior, likes, dislikes, preferences and priorities. Recommendations has got immense importance so as to improve customer experience and increase in business.

Currently available recommendation systems has reached limitations in terms of accuracy because of dependency on user-item rating data. This rating data is unavailable most of the time. In proposed methodology, a new recommendation technique is suggested which uses frequent pattern matching and Heat-Diffusion algorithm along with the inclusion of social media dataset so as to increase accuracy of the recommendations generated.

In this paper Section 1 contains basic information about domain and introduction of the problem and its solution. Section 2 contains contains an overview of existing system. Section 3 includes a description of system architecture. Section 4 contains an explanation about proposed system.

Section 5 describes implementation. Section 6 contains results. Section 7 conclusion of this system and future work.

## II. RELATED WORK

### A. Collaborative Filtering (CF)

Collaborative Filtering is a technique used by many recommendation engines. Neighborhood based and Model based are the two approaches of collaborative filtering. [2]

Neighborhood based approaches are applied for the predictions and are widely used in commercial CF systems. Example includes used-based approach and item-based approach. User-based approach analyzes rating data of similar users and generate recommendations for active user on that basis. While an item-based approach, items rated by active user are analyzed and recommendations are generated. [1]

In model based approach datasets are trained previously and based on already trained model predictions are made. Example of model based approach includes the clustering model – Google News – categories of news like “business”, “sports”, “technology” are predicted using the heading words of the news. Words in the heading are applied to already trained model and news is placed in the respective cluster. As more and more content are added to the cluster model becomes more clever and recommends more accurately. [3][5]

Recommendations based on CF techniques depends rating matrix containing user specific ratings. Though, most of the times rating data are unavailable as information on the Web is less structured and more diverse. Also for available rating data question of veracity cannot be addressed. [8][9]

### B. Query Suggestion

Query suggestion is a valuable technique so as to recommend relevant queries to user. User searches information by providing queries to search engines; if search engine would able to predict what user may want to search then it will save much time and predictions will be made accurately. [11]

Model enriches itself by the addition of more and more queries to it, and it learns on its own through the customer behavior. [4]

C. Image Recommendation

Image recommendation is another interesting and most widely used recommendation application on the web. Usually, such systems asks active user to rate some images from and within different categories to find users like and preferences and based on ratings data it then displays similar images which are more likely to the user. [9]

In general, no matter what type of source data is, the proposed recommendation framework can be applied to most of recommendation tasks on the web, which will give more relevant results.

III. SYSTEM ARCHITECTURE

In new system architecture, novel scheme to generate recommendation is proposed. For this, DFS algorithm is used to traverse web graph data, Apriori algorithm to decrease time complexity and social data is used to increase accuracy of the system. [1][12]

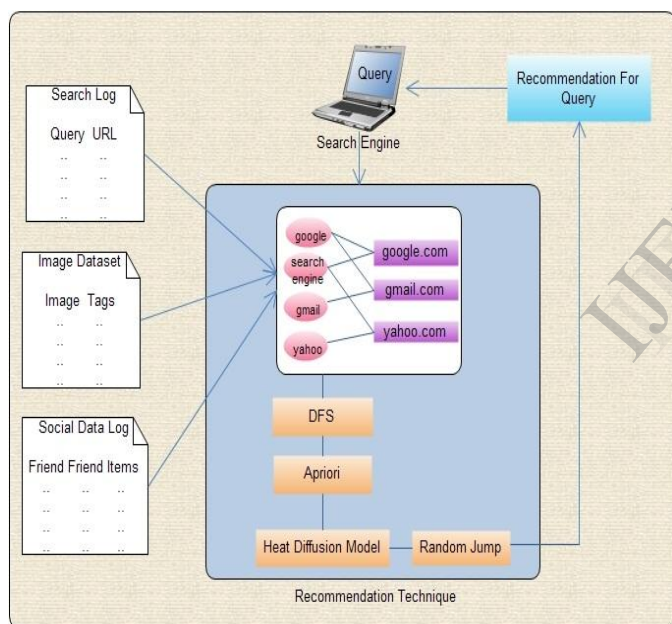


Fig. 1. Architecture of proposed system

- User can search for the query through search text box
- By referring to the query, DFS algorithm can extract subgraph from main Graph 'G'
- To retrieve frequently occurred patterns from the subgraph, Apriori algorithm is used
- By calculating HD values on extracted final subgraph, final recommendations are generated.

Here in figure 1, architecture flow is shown and is implemented accordingly.

IV. PROPOSED SYSTEM

Proposed system for generating recommendations allows user to use it as a general framework for recommendation. Here, in this paper it is demonstrated and results are calculated by considering a sample dataset.

New system uses DFS and apriori algorithm to extract subgraph related to searched query and heat values are calculated for each query-URL relationship.

A. Heat Diffusion

Heat Diffusion is a physical phenomenon; in physical medium, heat always flows from a position with higher temperature to lower temperature. In the same way, queries are also interrelated with each other in someor the other way. [1]

B. Graph Diffusion

Graph is a data structure that is most suited for web data as, relationship between the nodes is strongly establishes using graphs. Heat Diffusion technique can be applied to established graph nodes assuming information propagation on web graphs.[12]

C. Random Jump

According to heat diffusion technique, heat can only propogate through the links that are connect nodes in a graph. But in practical scenario, random relations do exist even though nodes are not directly connected. To capture such relations random jump technique is used. [1]

D. Building Graph for Recommendation

Here to build the generalised recommendation framework , sample query-URL dataset is used. Relationship is established between searched queries and clicked URLs. Sample dataset is as shown in table.

TABLE I. SAMPLE DATASET

ID	QUERY	URL	RANK
368	TWITTER	HTTP://WWW.FACEBOOK.COM	3
368	TWITTER	HTTP://EN.WIKIPEDIA.ORG/WIKI/TWITTER	1
1248	IPHONE	HTTP://WWW.APPLE.COM/IPHONE	4
1248	IPHONE	HTTP://WWW.YOUTUBE.COM/WATCH?P=OFXXG	3
2598	GOOGLE	HTTP://WWW.GOOGLE.COM	6
2598	GOOGLE	HTTP://WWW.GMAIL.COM	8
2598	GOOGLE	HTTP://WWW.YOUTUBE.COM	7

543	MOBILE	WWW.SNAPDEAL.COM/PRODUCTS/MOBILESPHONES	2
543	MOBILE	GADGETS.NDTV.COM/MOBILES/ALL-BRANDS	6

This sample dataset is now represented in terms of graph datastructure. Queries & URLs are considered as nodes of bipartite graph and edge from query to URL exists if user has clicked URL 'u' after issuing query 'q'.

Bipartite Graph,  $B_{q1} = (V_{q1}, E_{q1})$ , where,  $V_{q1} = Q \cup L$ ,  $Q = \{q1, q2, \dots, qn\}$ , and  $L = \{l1, l2, \dots, ln\}$ .  $E_{q1} = \{(qi, lj) \mid \text{there is an edge from } qi \text{ to } lj\}$ .

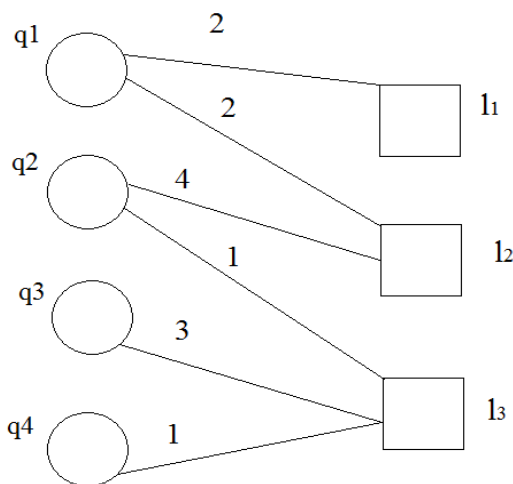


Fig. 2. Undirected query-URL bipartite graph

Undirected graphs cannot directly processed with heat diffusion as it cannot predict interpret results accurately. Hence the undirected bipartite is converted into directed bipartite graph.

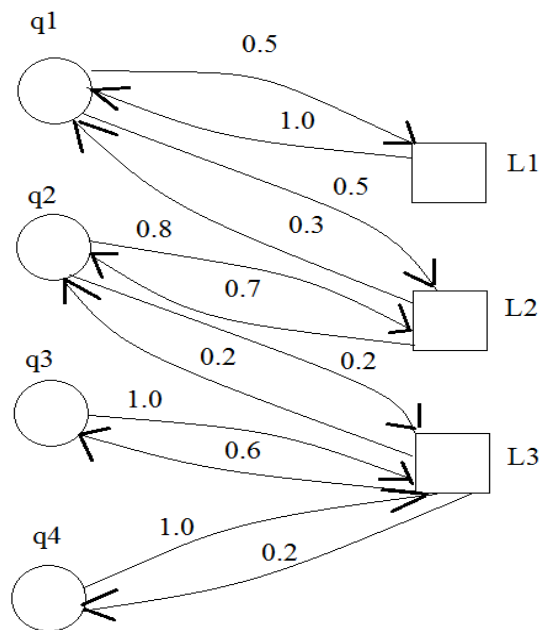


Fig. 3. Directed query-URL bipartite graph

The weight on query-URL edge is normalized by the number of times that query is issued, while the weight on a directed URL-query edge is normalized by the number of times URL is clicked. After the conversion of graph, query suggestion algorithm is designed. [1][12]

E. Query Suggestion Algorithm

- 1) A converted bipartite graph  $G = (V^+ \cup V^*, E)$  consists of query set  $V^+$  and URL set  $V^*$
- 2) Given a query  $q$  in  $V^+$ , a subgraph is constructed by using depth-first-search in  $G$
- 3) A subgraph is given to apriori algorithm to retrieve frequently appearing query and clicked URLs
- 4) Start the diffusion process using,  $f(1) = e^{ar}f(0)$
- 5) Top-K queries with the largest values in  $f(1)$  are displayed as suggestions. [12]

V. IMPLEMENTATION

Proposed system is implemented as a browser based application with username and password authentication mechanism.

A. Procedural Steps

- 1)  $U$  is the set of users,  $U = \{u1, u2, u3\}$
- 2)  $D$  is the set of data,  $D = \{d1, d2\}$   
 $d1 = \{q, u\}$ ,  $q = \text{query}$ ,  $u = \text{clicked URL for the query}$   
 $d2 = \{u1, u2, i\}$ ,  $u1 = \text{user}$ ,  $u2 = \text{user related to } u1$  &  $i = \text{related entity of } u1$
- 3)  $Q$  be the main set of entered query,  $Q = \{q1, q2, q3\}$

- 4)  $SYS = \{DX, DF, AP, BG, HD, HV\}$   
 DX = It Data Extractor which extract the data from the dataset  
 DF = Search query using DFS in database  
 AP = Filter the results of DFS using Apriori Algorithm  
 BG = It generate the Bipartite graph by considering query and url as a node  
 HD = Heat Diffusion find out H-D matrix for query and H-D with Random Jump matrix  
 HV = It is Heat Vector which suggest the final recommendation for the given query in particular order

- 5) P be the set of processes  
 $P = \{P1, P2, P3, P4\}$   
 $P1 = \{e1, e2\}$   
 where,  
 $e1 = i|j$  database designing from the dataset  
 $e2 = j|j$  show all clicks through data from the database  
 $P2 = \{e1, e2, e3, e4\}$   
 where,  
 $e1 = i|j$  Take the Query from the user  
 $e2 = i|j$  Search query using DFS  
 $e3 = j|j$  Filter the results of DFS using Apriori  
 $e4 = j|j$  generate the directed bipartite Graph

- 6) Graph  $G = \{E, V\}$  where,  $V = \{v1, v2, v3\}$  be the set of vertex and  $E = \{(v1, v2), (v2, v3)\}$  be the set of edges

- 7)  $P3 = \{e1, e2\}$   
 where,  
 $e1 = i|j$  find out the similarity information propagation on Web graphs  
 $e2 = j|j$  Find out H-D matrix for query and H-D with Random Jump matrix  
 $F_i(t) = \text{heat at node } V_i \text{ at time } t$

- 8)  $P4 = \{e1, e2\}$   
 where,  
 $e1 = i|j$  find out the Heat Vector  
 $e2 = j|j$  suggest final recommendation on the base of Heat Vector in given order  
 $F_i(t) = \text{heat at node } V_i \text{ at time } t$

## VI. RESULTS & DISCUSSIONS

Experimental evaluation of recommendation framework is given in this section. In proposed system, evaluation is done by taking sample dataset and results are shown by issuing a query – ‘aa’. Time and Accuracy are the parameters that are targeted for comparison.

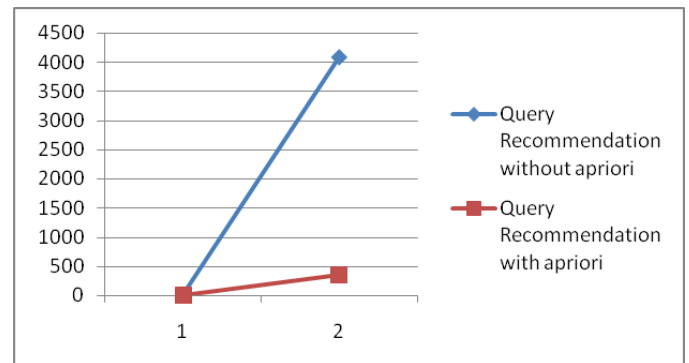


Fig. 4. Graph comparing results of query suggestions by considering Apriori algorithm and without Apriori algorithm

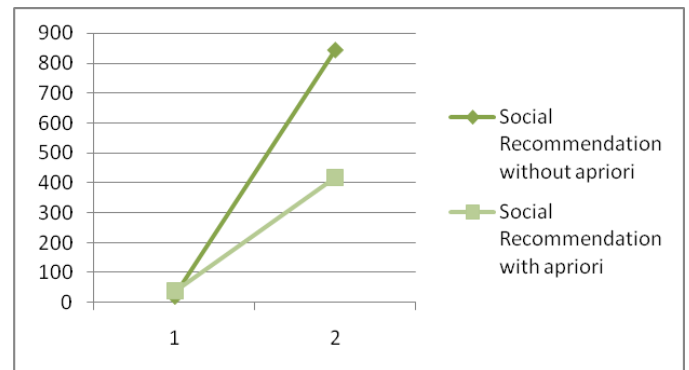


Fig. 5. Graph comparing results of social recommendation by considering Apriori algorithm and without Apriori algorithm

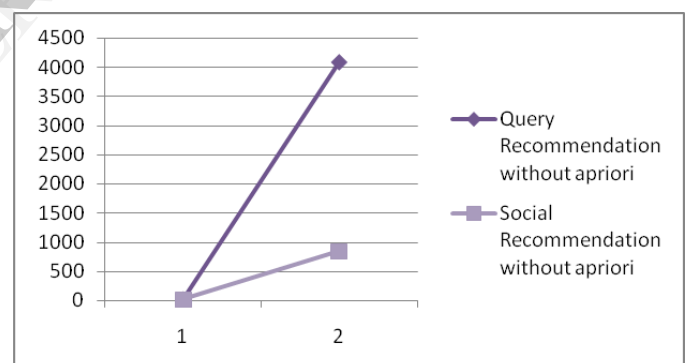


Fig. 6. Graph comparing results of query suggestions by considering sample social media dataset without apriori algorithm

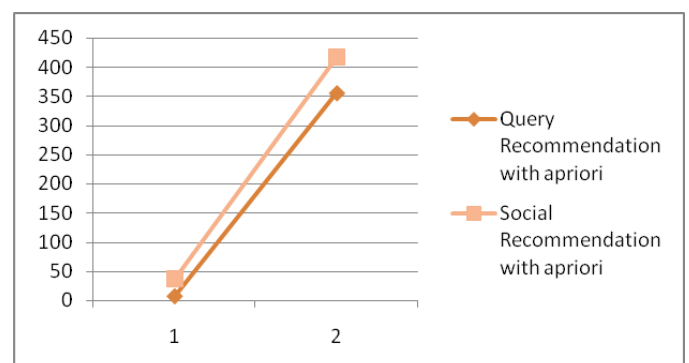


Fig. 7. Graph comparing results of query suggestions by considering sample social media dataset with apriori algorithm

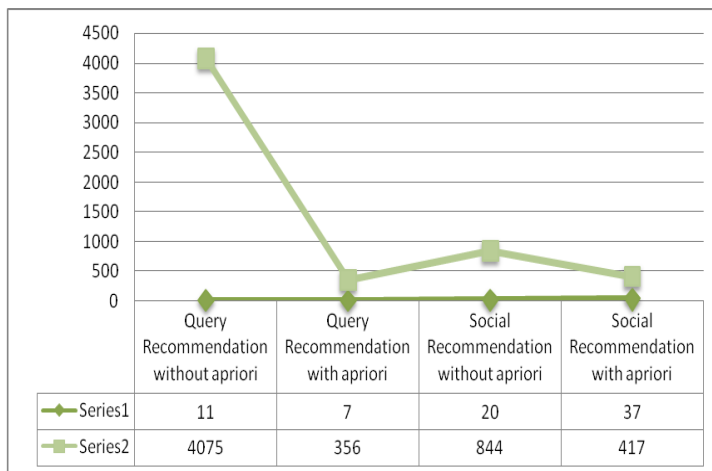


Fig. 8. Graph comparing results of query suggestions by considering query-URL dataset & sample social media dataset with & without using apriori algorithm

As results shown in above graphs, due to inclusion of apriori algorithm time taken to generate query recommendation is significantly reduced. Also by considering social media dataset recommendation includes various sources which are previously exists only in the form of text. Overall performance of the system is improved by 60-65% interms of time and accuracy.

## VII. CONCLUSION AND FUTURE SCOPE

Existing system uses collaborative filtering technique to generate recommendation. The respective technique uses user-item rating data which is unavailable most of the times which affects accuracy of the system. In this paper, a general framework for recommendation using DFS, Apriori algorithm, HD technique and social media data is proposed that reduces time taken to generate recommendation and increases accuracy of the recommendation. Due to this, the overall performance of the system is improved by 60-65% footnotes.

In future, the proposed system can be built using new generation big data technology by adding more feature to UI like query suggestion, event processing.

## ACKNOWLEDGEMENT

The authors would like to thank all the unknown reviewers for their valuable comments and suggestions.

## REFERENCES

- [1] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithm", ACM, MAY 2001
- [2] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, H. Li, "Context-Aware Query Suggestion by Mining Click-Through and Session Data", ACM ON KDD, 2008
- [3] H. Cui, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma, "Query Expansion by using user logs", IEEE, July/August 2003
- [4] Ma, I. King, and M. Rung-Tsong Lyu, "Mining Web Graphs for Recommendations", IEEE Transaction on Knowledge and Data Engineering, VOL. 24, No. 6, June 2012
- [5] H. Mase, and H. Ohwada, "A Collaborative Filtering Incorporating Hybrid-Clustering Technology", ICSAI 2012
- [6] H. Ma, H. Yang, M. R. Lyu and I. King, "Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates selection", ACM, October 2008
- [7] H. Ma, H. Yang, I. King, and M. Rung-Tsong Lyu, "SoRec: Social Recommendation Using Probabilistic Matrix Factorization
- [8] J. Yu, K. Xie, H. Zhao, F. Liu, "Prediction of User Interest Based on Collaborative Filtering for Personalized Academic Recommendation", IEEE, 2012
- [9] N. Abdullah, Y. Shlomo Geva, "Integrating Collaborative Filtering and Search-based Techniques for Personalized Online Product Recommendation", 11<sup>th</sup> IEEE International Conference on Data Mining, 2011
- [10] N. Craswell and M. Szummer, "Random Walks on the Click Graph", ACM, 2007
- [11] S. Hui, Lu. Pengyu, and Z. Kai, "Improving Item-Based collaborative Filtering Recommendation System with Tag", IEEE, 2012
- [12] Saurabh R. Deshpande, and Jyoti R. Yemul, "Web Graph Recommendation Technique using Heat Diffusion Method and Apriori Algorithm", Cyber Times International Journal of Technology & Management, vol. 7, Issue 1, October 2013-March 2014
- [13] W. Xia, L. He, J. Gu, K. He, and L. Ren, "Boosting Collaborative Filtering Based on Missing Data Imputation Using Item's Genre Information", IEEE, 2009