Personalized Recommendation System for Medical Assistance using Hybrid Filtering

Archana B. Salunke
M.E.computerscience
Plot 7,Harsidhi soci,Jalgaon Road,Aurangabad,Maharashtra

Smita L. Kasar
AssistantProfessor
MGM'sJawaharlalNehru Engg.College,Aurangabad

ABSTRACT
Recommender systems assist the consumers of service oriented environment to find out and select the most suitable services from a large number of available ones. Proposed paper is based on Personalized Recommendation System for medical assistance using keyword extraction. User can search doctor's profiles or hospital names according to doctor and hospital attributes. Natural Language Processing (NLP) is used to process user's ratings and reviews to compute system ratings. Depending on users rating and reviews, profiles are recommended. Medical-based Personalized Recommendation System computes similarity between given and collected attribute by using top-k query which is used to recommend each doctor profile and hospital name for each attribute in information retrieval. Personalized Recommendation system for medical assistance yields 0.06 satisfactions and 0.02 accuracy.

General Terms
Keyword Extraction ,Hybrid Filtering

Keywords
Recommendation System, Personalization, Profile, Natural Language Processing (NLP), XML, Top-k query.

1. INTRODUCTION
A Recommender System is a kind of data filtering system that tries to suggest a set of data items to the users that may be their preferred ones. Recommendation systems provide strong search engines to find out user favorite data from huge data available on Internet and World Wide Web. They have become fundamental applications in electronic commerce and information access, providing suggestions that effectively prune large information spaces so that users are directed toward those items that best meet their needs and preferences. Information retrieval [1] model of Web data which adds Metadata has been analyzed to identify a way to investigate the information origins effectively. Web recommendation system [2] is a specific type of information filtering system that aims to predict the user favorite data. User can approach the needed information from this model speedily and precisely. Peoples search for doctors and hospital, based on services provided by them. So modern interests focus on increasing accuracy of recommendation.

Now days, interest in web-based recommendation system is increasing because personalized services provided by these recommendation system are highly used in term internet shopping [5], [7]. They can be applied in variety of applications. Providing personalized of fitted information is required and essential for applying new marketing strategies as web personalization, customer relationship management (CRM),one to-one marketing [8]. That's the reason of Information Filtering had been studied. This proposed recommendation system for doctor profile and hospital name provides accurate and efficient results according to doctor specialty and hospital type. Proposed recommendation system can also be applied to some other development of the recommendation system of practical area. Proposed recommendation system provides recommendation for doctors and hospitals according to user reviews and ratings. Keywords present in user reviews are extracted and are analyzed for recommendation. Also proposed system yields high level of satisfaction and accuracy rate with reflecting doctor and hospital profile.

2. LITERATURE SURVEY
2.1 Recommendation System
Recommendation systems [3], [4] are a subclass of information filtering system that explore to anticipate the 'rating' or 'preference' that a user would give to an item. Recommender systems have changed the way people find products, information, and even other people [6]. The technology behind recommender systems has evolved over the past 20 years into a rich collection of tools that enable the practitioner or researcher to develop effective recommenders. Recommendation system can also be referred as 'Information Filtering Technology'. Recommendation systems used to find doctor profiles and hospital names according to user ratings rapidly and exactly. 'Extractor' abstracts the data frequently required by users from huge database by performing preprocessing of data [9]. Extracted data get stored in predefined XML templates for further processing. Data collected in information extraction process is emerges as the document. Document contains environment for learning the rules of information extraction and extracting required information. Such document can be categorized into three types as 'Structured documents', 'Unstructured documents', 'Semi-structured documents'. As type of document varies, method for learning the rules also varies [10].

Collaborative Filtering
Collaborative filtering approach is widely used to design recommendation systems. This approach depends on gathering and examining a huge amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users. A key benefit of the collaborative filtering technique is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items [11]. When building a model from a user's profile, a distinction is often made between explicit and implicit forms of data collection.

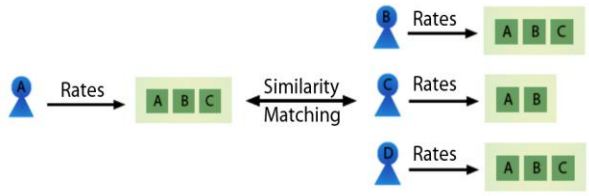

Figure 1: Collaborative filtering in Recommendation System

Content-based filtering
Content based filtering is based on description of the item and a profile of the user's preference. In a content-based recommender system, keywords are used to describe the items; beside, a user profile is built to indicate the type of item this user likes [13]. This technique has its origin in information retrieval and information filtering research.

Hybrid Filtering
Hybrid filtering tries to merge different techniques to mutually remove their weakness, so the proposed Medical based personalized recommendation systems have combined top-k query algorithm, keyword extraction and user-based personalization. Aim of top-k query is to deliver most accurate answers to the users query from large dataset. Top-k query algorithm is used to search out precise records from the given record set that matches the filtering keyword, and arrange them according to their scores. Top-K query algorithm uses scores documents against keywords. Here, top-k query is used to retrieve doctor and hospital name.

In keyword extraction process, Keywords are abstracted from hospital type and doctor specialty and are matched against each tuple stored in the database. The searched text entered for searching is marked as keyword and the low priority words are truncated and the process keyword extraction is done on the remaining words. Hybrid filtering also contains on the fly XML generation. The result generated from any search is structured and stored in an XML file. These files are created upon every single search done by user and admin. The structured formed within the XML file is called as Steiner tree

3. PROPOSED SYSTEM

Proposed method consists of Userlnterface, Extractor, Filter, KeyFinder, Profile Manager, and Database Manager. UserInterface phase contacts with users, takes input from users and generates appropriate outputs for user requests. Following figure shows system architecture:

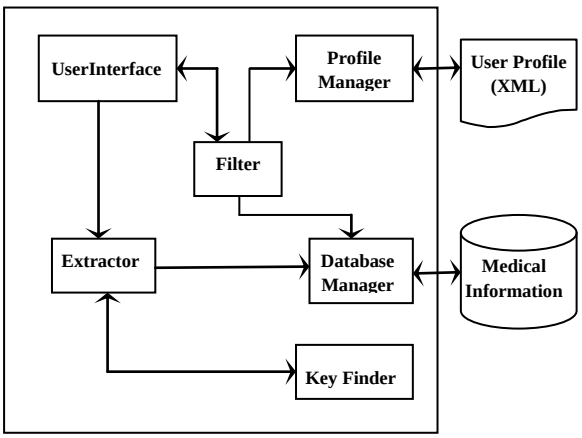

Figure 2: System Architecture of proposed recommendation system

Extractor extracts keywords, from doctor name, specialty, hospital name, address. If some record does not contain any keyword, it could be extracted proper keywords in names via KeyFinder. Filter carried out filtering process to refine the collected information which all gets stored in Database. UserProfile contains personal information of doctors and hospitals along with its address, which can be display using XML files. Profile manager contains all filtered data of doctors and hospitals form database.

Filter performs filtering process to provide proper records of doctor and hospitals to users among the stored records on user's chosen specialty and hospital type via UserProfile which contains personal preference information with XML forms, and provides list of refining doctor and hospital namesto users.UserProfile has been stored the user favorite information by topic. When users select some record with UserInterface, they reflect user's preference information by weight value of upgrade.

User Provide reviews and ratings based on Natural Language Processing (NLP).Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human–computer interaction. Many challenges in NLP involve natural language understanding, that is, enabling computers to derive meaning from human or natural language input, and others involve natural language generation.

Depending upon the reviews and ratings of users in combination with system generated ratings, records are ranked. Rank of records used to list out doctor profile or hospital name according to user interest. Registered User can search for doctor profile or hospital name in two different ways as other user's ratings and system generated ratings. Based on user ratings or system ratings, profiles are recommended to users depending on doctor specialty and hospital type. Unregistered users can view all details of doctor profiles and hospital names. They are also able to view average rating provided by user and system.

Keywords are extracted from collected information which contains doctor's profile, hospital name, services provided, specialties. So that extracted keywords are further used for recommendations to other users. Proposed recommendation system also provides recommendation for doctors and hospitals according to user reviews and ratings. Keywords present in user reviews are extracted and are analyzed for recommendation. User can also able to view system generated ratings.

3.1 Top-k query

Information retrieval systems use various approaches to rank query answers. Users are more concerned about the most important i.e., top-k query answers in the potentially huge answer space. Different emerging applications demands for competent support for top-k queries. For example, in the context of the Web, the effectiveness and efficiency of meta search engines, which combine rankings from different search engines, are highly related to efficient rank aggregation methods [21]. Similar applications are present in the perspective of information retrieval and data mining. Most of these applications compute queries that involve joining and aggregating multiple inputs to provide users with the top-k results.

Aim of top-k query is to retrieve best answers from a potentially very large result set. The definition of top-k queries requires a system able to "rank" objects. In proposed Medical-based Personalized Recommendation System, top-k query is used to rank the doctor and hospitals records. Records are ranked according to both user and system ratings so that user can get accurate recommendations.

3.2 Keyword extraction

In order to get keywords section from content, it assume as follows:

1) Keywords section can exists in doctor's profile.

2) Keywords section can exists in hospital profile.

3) Keywords section can exist in user ratings and reviews.

This can easily calculate degree of reflection for user posts by using following equation:

$$\text{Re } flex = \frac{\#KeywordTitle}{\#TitleTerm}$$

Where, title term includes all noun and compound noun.

3.3 User based personalization

UserProfile is composed hierarchical architecture as Figure 3.

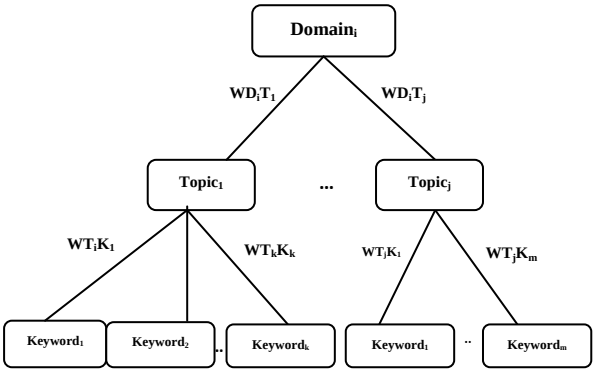

Figure 3: Architecture of UserProfile

Where, WD_iT_j = weight of domain i to Topic j.

WT_jK_k = weight of topic j to keyword k.

Generally, each user profile stored in XML format which further forms hierarchical architecture as shown in fig.3. Preferred information can be easily indicated by domain and topic of the users. It can be traced easily preference information by Domain and Topic of the users, adapt easily with a variation of signal information by forming hierarchical architecture. The update of Medical Profile is made for every click of the records collected by topic. Whenever a profile is selected, it makes increasing of a frequency of attributes and keyword then, recalculates a rate of each occurrence and reflects to Doctor or Hospital Profile.

```
−<Doctors>
 −<Record id="1">
    <Dr_Name>P.jagtap</Dr_Name>
    <Dr_Code>AUR4567</Dr_Code>
    <Specialist>orthopedic</Specialist>
    <Hospital>Rajeev Hospital</Hospital>
    <City>Aurangabad</City>
    <Address>Nutan Colovy, Himayat Nagar, Aurangabad</Address>
    <Contact>9687756453</Contact>
  </Record>
 −<Record id="2">
    <Dr_Name>A B Kadethankar</Dr_Name>
    <Dr_Code>AB12</Dr_Code>
    <Specialist>Orthodontist</Specialist>
    <Hospital>Eknath Hospital</Hospital>
    <City>Aurangabad</City>
    <Address>Plot no 11 Garkheda Parisar Aurangabad</Address>
    <Contact>998877667</Contact>
  </Record>
 −<Record id="3">
    <Dr_Name>A Bhattacharya</Dr_Name>
    <Dr_Code>AB13678</Dr_Code>
    <Specialist>Physician</Specialist>
    <Hospital>Sai Hospital</Hospital>
    <City>Aurangabad</City>
    <Address>Rajurkar Colony, Natsaang Road, Aurangabad</Address>
    <Contact>9856325698</Contact>
  </Record>
```

Figure 4: Example of Doctors Profile

4. RESULTS AND ANALYSIS

Satisfaction and accuracy for measuring system's performance defined as following:

$$SAT = \frac{\#SatisfiedDoctors / Hospital\ Profiles}{\#\text{Re}commendedDoctors / Hospital\ Profiles}$$

SAT represents ratio of the number of recommended doctor or hospital profiles for users to the number of satisfied doctor or hospital profiles for given doctor specialty or hospital type.

$$ACC = \frac{\#\text{Re}commendedSat + Not\,\text{Re}commendedUnsat}{TotalSaved\ Profiles}$$

Accuracy ACC represents ratio of total number of saved doctor or hospital profiles to recommended satisfied profiles plus not-recommended, unsatisfied profiles for a given doctor specialty or hospital type.

Table 1 and table 2 shown below describe the overall doctor and hospital rating. Average rating is computed by combining user ratings and system ratings.

Table 1: Doctor Ratings

Sr. No	Name of Doctor	No. Of positive review	No. Of Negative Review	Combination of Positive and Negative Reviews	User Rating	System Rating	Average Rating
1	A kasliwal	3	3	5,5,1,2,1,5	19	9	4.66
2	A Birjdar	3	3	3,1,5,5,4,1	19	9	4.46
3	A Sarda	2	3	1,2,4,5,1	13	10	4.4
4	A Jawlekar	2	3	4,1,2,1,5	13	7	4
5	A B Kadethankar	4	1	4,5,3,5,1	18	16	6.8

Table 2: Hospital Ratings

Sr. No	Name of Hospital	No. Of positive review	No. Of Negative Review	Combination of Positive and Negative Reviews	User Rating	System Rating	Average Rating
1	Abhasi Hospital	2	3	3,5,1,1,4	14	7	4.4
2	Arun Hospital	2	3	2,3,5,5,2	17	7	4.8
3	Bhaki pooja Hospital	2	4	3,1,1,5,5,1	15	10	4.16
4	HP Hospital	3	2	3,1,5,5,4	18	9	5.4
5	ITC Hospital	5	1	3,5,5,5,4,4	26	13	6.66

For measuring the satisfaction ratio, threshold value of recommended profiles is fifty. But for measuring the accuracy does not have any restrictions on the profiles recommended to users as every record saved on local can be treated as recommended. Following figure describes whole user's average of SAT and ACC. Value of SAT generated as 0.06 and value of ACC generated as 0.02.

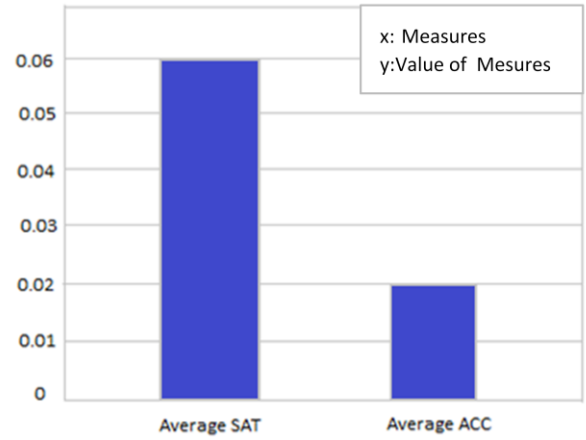

Figure 7: Average SAT and ACC

4. CONCLUSION

People spend more time and need various efforts to search doctor and hospitals according to the specialties provided. Proposed method provides accurate and efficient recommendations to both registered and unregistered users depending on doctor's specialty and hospital types. Through keyword extraction process, proposed system extracts keywords from user ratings and reviews and user this keywords for recommendation of doctor and hospital name to other users. Proposed recommendation system provides flexibility to users to search information of doctor and Hospital. Experiments carried out describe the accuracy and efficiency of recommendation system. Proposed system yields high level of satisfaction and accuracy rate with reflecting doctor and hospital profile. Future work can be implemented for labs and pathologies. Also in future, recommendation values can be based on geographical distance between hospitals and patients.

5. REFERENCES

[1] Bamshad Mobasher, Honghua Dai, Tao Luo, Yuqing Sun and Jiang Zhu, "Integrating Web Usage and Content Mining for More Effective Personalization, " Electronic Commerce and Web Technologies LCNS, vo.1875, pp.165-176, 2000.

[2] Ibrahim Cingil, Asuman Dogac, Ayca Azgin, "A broader approach to personalization," Communications of the ACM, vo.43, Issue 8, pp. 136-141, 2000.

[3] Nitin Agarwal, Ehtesham Haque, Huan Liu, and Lance Parsons, "Research Paper Recommender Systems: A Subspace Clustering Approach, " WAIM 2005, LNCS 3739, pp. 475-491, 2005.

[4] Gediminas Adomavicius, Alexander Tuzhilin, "User Profiling in Personalization Applications through Rule Discovery and Validation, " ACM, pp.377-381, 1999.

[5] Bamshad Mobasher, Robert Cooley, Jaideep Srivastava, "Automatic personalization based on Web usage mining", Communications of the ACM, vo.43, Issue 8, pp.142-151, 2000.

[6] Joseph Kramer, Sunil Noronha, John Vergo, "A User-centered design approach to personalization" Communications of the ACM, vo.43, Issue 8, pp.44-48, 2000.

[7] Marco Gori, Augusto Pucci, "Research Paper Recommender Systems: A Random-Walk Based Approach, " Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 2006.

[8] Masashi Shimbo, Takahiko Ito, Yuji Matsumoto "Evaluation of Kernel-based Link Analysis Measures on Research Paper Recommendation, " JCDL '07 Proceedings of the 7th ACM/IEEE-CSjoint conference, 2007.

[9] Bela Gipp, Joran Beel, and Christian Hentschel, "Scienstein: A Research Paper Recommender System, " In Proceedings of the International Conference on Emerging Trends in Computing (iCETiC'09), pp. 309-315,2009.

[10] Andre Vellino, "Recommending Journal Articles with PageRank Ratings, " Recommender Systems 2009.

[11] A. Naak, H. Hage, and E. A(meur, "A Multi-criteria Collaborative Filtering Approach for Research Paper Recommendation in Papyres, " MCETECH 2009, LNBIP 26, pp. 25-39, 2009.

[12] Zhiping Zhang, Linna Li, "A Research Paper Recommender System based on Spreading Activation Model, " IEEE, 2010 2nd International Conference on Information Science and Engineering (ICISE), pp.928-931, 2010.

[13] Worasit Choochaiwattana, "Usage of Tagging for Research Paper Recommendation" 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), vo.2, pp.439-442, 2010.

[14] Chenguang Pan, Wenxin Li, "Research Paper Recommendation with Topic Analysis" 2010 International Conference On Computer Design And Appliations (ICCDA 2010), vo.4, pp.264-268, 2010.

[15] Pijitra Jomsri, Siripun Sanguansintukul, Worasit Choochaiwattana, "A Framework for Tag-Based Research Paper Recommender System: An IR Approach, " 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, 2010.

[16] Kazunari Sugiyama, Min-Yen Kan, "Scholarly Paper Recommendation via User's Recent Research Interests, " JCDL'10 Proceedings of the 10th annual joint conference on Digital libraries, 2010.

[17] Cristiano Nascimento, Alberto H. F. Laender, Marcos Andre Gonc;:alves, Altigran S. da Silva, "A Source Independent Framework for Research Paper Recommendation, " JCDL '11, 2011.

[18] Kiyoko Uchiyama, Akiko Aizawa, Hidetsugu Nanba, Takeshi Sagara, "OSUSUME: Cross-lingual Recommender System for Research Papers, " CaRR 2011, Proceedings of the 2011 Workshop on Context awareness in Retrieval and Recommendation, 201l.

[19] Felice Ferrara, Nirmala Pudota, and Carlo Tasso, "A Keyphrase-Based Paper Recommender System" IRCDL 2011, CCIS 249, pp. 14-25, 2011.

[20] Manabu Ohta, Toshihiro Hachiki, Atsuhiro Takasu, "Related Paper Recommendation to Support Online Browsing of Research Papers, " IEEE , 2011 Fourth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT), pp.130-136 , 2011.

[21] Jianhua Feng, Guoliang Li, and Jianyong Wang, “Finding Top-k Answers in Keyword Search over Relational Databases Using Tuple Units”, Ieee Transactions On Knowledge And Data Engineering, Vol. 23, No. 12, December 2011.