Hybrid Recommender System for Therapy Recommendation

V. Vishwajith[1], S. Kaviraj[2], R. Vasanth[3]

Department of Computer Science and Engineering,

Dr. Mahalingam College of Engineering & Technology, Pollachi, Coimbatore, India[1,2,3]

Abstract: This system provides data-driven therapy recommendation for the patients. Therapy will be recommended to a patient by analysing the response from previous records which are similar to the given patient's record. Two methods for therapy recommendation, namely, Collaborative Recommender and Demographic-based Recommender, were proposed. Both algorithms aim to predict the individual response to different therapy options using diverse patient data and recommend the therapy which is assumed to provide the best outcome for a specific patient and time, that is, consultation. Both methods are evaluated using a clinical database incorporating patients suffering from the autoimmune skin disease psoriasis. The Collaborative Recommender proves to generate both better outcome predictions and recommendation quality. However, due to sparsity in the data, this approach cannot provide recommendations for the entire database. In contrast, the Demographic-based Recommender performs worse on average but covers more consultations. Consequently, both methods profit from a combination into an overall recommender system. In addition to the above proposed system, new Model-based Recommender is proposed and it is compared with the above system to check its efficiency. Model-based Recommender is also proposed to enhance the efficiency of recommendation. Data mining brings the concept of artificial intelligence, data structures, statistics, and database together. It is a high demand area because many organizations and businesses can benefit from it. The large volume of daily captured data in healthcare institutions and out-of-hospital settings opens up new perspectives for Data Mining in healthcare. Due to the amount of the data, its high dimensionality and complex interdependencies within the data, an efficient integration of the available information is only possible using technical aids. So, data-driven Clinical Decision Support Systems (CDSS) are designated to assist physicians or other health professionals during clinical decision-making.

Keywords: Therapy Recommendation, Hybrid Recommender System, Collaborative Filtering, Demography Based Filtering, Model Based Recommender, Clinical Decision Support System

I. INTRODUCTION

The large volume of daily captured data in healthcare institutions and out-of-hospital settings opens up new perspectives for healthcare. Due to the amount of that data, its high dimensionality and complex interdependencies within the data, an efficient integration of the available information is only possible using technical aids. In this regard, data-driven CDSS can be expected to take a major role in future healthcare. Generally, CDSS are designated to assist physicians or other health professionals during clinical decision-making. CDSS are demanded to be integrated into the clinical workflow and to provide decision support at time and location of care. Data-driven CDSS, in particular, make use of data-mining and machine-learning techniques to extract and combine relevant information from patient data, in order to provide assistance for diagnosis and treatment decisions or even to be used in clinical quality control based on large-scale data.

A. Objective

To recommend therapy for a patient by analysing the response from previous records which are similar to the given patient's record.

B. Overview

This system provides data-driven therapy recommendation for the patients. Therapy will be recommended to a patient by analysing the response from previous records which are similar to the given patient's record. Two methods for therapy recommendation, namely, Collaborative Recommender and Demographic-based Recommender, are proposed. Both algorithms aim to predict the individual response to different therapy options using diverse patient data and recommend the therapy which is assumed to provide the best outcome for a specific patient and time, that is, consultation. The proposed methods are evaluated using a clinical database incorporating patients suffering from the autoimmune skin disease psoriasis. The Collaborative Recommender proves to generate both better outcome predictions and recommendation quality.

II.   RELATED WORKS

A.   Recommendation Using Collaborative Filtering

Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). The underlying assumption of the collaborative filtering approach is that if a person A has the same opinion as a person B on an issue, A is more likely to have B's opinion on a different issue than that of a randomly chosen person. For example, a collaborative filtering recommendation system for television tastes could make predictions about which television show a user should like given a partial list of that user's tastes (likes or dislikes).Note that these predictions are specific to the user, but use information gleaned from many users.

In the more general sense, collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. Applications of collaborative filtering typically involve very large data sets. Collaborative filtering methods have been applied to many different kinds of data including: sensing and monitoring data, such as in mineral exploration, environmental sensing over large areas or multiple sensors; financial data, such as financial service institutions that integrate many financial sources; or in electronic commerce and web applications where the focus is on user data, etc. The remainder of this discussion focuses on collaborative filtering for user data, although some of the methods and approaches may apply to the other major applications as well.

The motivation for collaborative filtering comes from the idea that people often get the best recommendations from someone with tastes similar to themselves. Collaborative filtering encompasses techniques for matching people with similar interests and making recommendations on this basis.

B.   Demographic-Based Recommender

The demographic approach is based on the assumption that different demographic niches have different tastes in items. The system therefore recommends items to users based on their demographic profiles, such as age, gender, language and location. Many modern recommender systems use hybridization, which combines two or more recommender techniques to gain better performance than when the systems are implemented individually. Demographic filtering is mostly employed in hybrid systems together with other types of recommender techniques in order to enhance prediction accuracy. The strength of the demographic filtering technique is that the new user problem does not apply to this type of recommender system since they do not need a list of ratings from a new user to make recommendations. However, according to earlier research, the major problem with demographic systems is that demographic data in combination with item rating.

Demographic information can be used to identify the types of users that like a certain object. For example, shows information on the age, gender, education, etc. of people that rated a restaurant together with their rating of the restaurant. One might expect to learn the type of person that likes a certain restaurant. Similarly, Life Style Finder attempts to identify one of 62 pre-existing clusters to which a user belongs and to tailor recommendations to users based upon information about others in this cluster. Obtaining demographic information can be difficult. Life Style Finder enters into dialog with the user to help categorize the users are difficult to acquire. The demographic-based and collaborative filtering approaches hybridization had been introduced by researches for improving the recommendation quality rather than solving "cold-start problem". A group of researchers have applied a hybrid model-based approach on movie domain using user demographic data to enhance the recommendation suggestion process, it classified the genres of movies based on user demographic attributes, such as user age (kid, teenager or adult), student (yes or no), have children (yes or no) and gender (female or male).

C.   Content-Based Collaborative Filtering

The content-based approach is when the system learns a user's taste from items the user previously rated highly, and uses this to recommend new similar items. If a user for example has rated movies that belong to a specific genre, or in which a specific actor has acted in, then the recommender system learns to suggest movies for the user of that genre or actor. Advantages with content-based filtering include the independence of other users; recommendations depend solely on the user's own previous preferences. Furthermore, content-based recommenders are also capable of recommending new items that have not been rated by any users before since recommendations are made exclusively based on its features, and not on other users' ratings. Finally, content-based recommenders are only effective when enough ratings from an user have been collected. Thus, for a new user with no rating history, the system will not be able to provide recommendations.

Basically, these methods use an item profile (i.e., a set of discrete attributes and features) characterizing the item within the system. The system creates a content-based profile of users based on a weighted vector of item features. The weights denote the importance of each feature to the user and can be computed from individually rated content vectors using a variety of techniques. Simple approaches use the average values of the rated item vector while other

sophisticated methods use machine learning techniques such as Bayesian Classifiers, cluster analysis, decision trees, and artificial neural networks in order to estimate the probability that the user is going to like the item.

Content-based recommendation engine works with existing profiles of users. A profile has information about a user and their taste. Taste is based on user rating for different items. Generally, whenever a user creates his profile, Recommendation engine does a user survey to get initial information about the user in order to avoid new user problem. In the recommendation process, the engine compares the items that are already positively rated by the user with the items he didn't rate and looks for similarities. Items similar to the positively rated ones will be recommended to the user. Here, based on user's taste and behavior a content-based model can be built by recommending articles relevant to user's taste. Such a model is efficient and personalized yet it lacks something.

D. An Approach to Data Mining in Healthcare: Improved K-means Algorithm

Clustering is a division of data into groups of similar objects. Each group, called cluster, consists of data objects that are similar between themselves and dissimilar to objects of other groups. Dissimilarities are assessed based on the attribute values describing the objects. Often, distance measures are used. Representing data by fewer clusters necessarily lose certain fine details, but achieves simplification. K-means is a popular clustering method, but it also has disadvantages. One is the fixed number of clusters must be specified as an input to the algorithm. Moreover, the initial randomly choice data points as cluster means can result in different final clusters. That means each rerun will produce a different result. Determining the number of clusters is usually so hard to achieve a good clustering result. A number of researchers used method based on information obtained during the K-means clustering operation itself to select the number of clusters, K. It is a clustering algorithm that partitions a dataset of N items into K clusters iteratively. The algorithm operates through the following steps:
1. Initialize by partitioning the data into K clusters.
2. Re-assign every datapoint to the closest cluster center.
3. Update the clusters' centers.
4. Repeat step 2-3 until convergence where the cluster centers no longer change.

The output of the algorithm is a set of K centroids. For each centroid, the sum of the squared distance to every datapoint in that cluster is minimized. In the healthcare industry, determining the number of groups of patients is an important problem and it requires high precision. One way to find out the appropriate number of clusters k is running the algorithm with various k, then choose k that the result clusters are the best.

E. Hypothetical Recommendation

Explanation and dynamic feedback given to a user during the recommendation process can influence user experience. Despite this, many real-world recommender systems separate profile updates and feedback, obfuscating the relationship between them. This paper studies the effects of what is called hypothetical recommendations. These are recommendations generated by low cost, exploratory profile manipulations, or "what-if" scenarios. In particular, the effects of dynamic feedback from the recommender system on profile manipulations, the resulting recommendations and the user's overall experience. Results from a user experiment (N=129) suggest that dynamic feedback improves the effectiveness of profile updates, when dynamic feedback is present, users can identify and remove items that contribute to poor recommendations, profile update tasks improve perceived accuracy of recommendations and trust in the recommender, regardless of actual recommendation accuracy.

III. METHODOLOGY

The algorithms described in the following aim at recommending the potentially most effective systemic therapy for a given patient and consultation. The collaborative filtering idea is transferred to the therapy recommendation domain, considering therapies as items and therapy response as a user's preference. In a preceding prediction step, individual therapy outcome is estimated for all available therapies that have not yet been applied to the patient. The two recommender approaches proposed in this work differ in the information used to represent consultations. The applied Collaborative Recommender algorithm uses solely outcome from all preceding consultations to represent a consultation. The hybrid Demographic based Recommender approach on the other hand is taking additionally all available patient describing data into account. Both recommender engines suffer from drawbacks depending on the data properties which the other approach is capable of compensating for. Therefore, an ensemble of recommenders is introduced combining both recommender engines

A. Collaborative Recommender Algorithm

In this approach, the consultation under consideration is only represented by the affinity values related to therapies applied up to this consultation. The underlying assumption is that the therapy applied to a given patient within the therapy history and the associated outcome reincorporates information about that respective patient and consultation which can then be transferred to patients with similar therapy history.
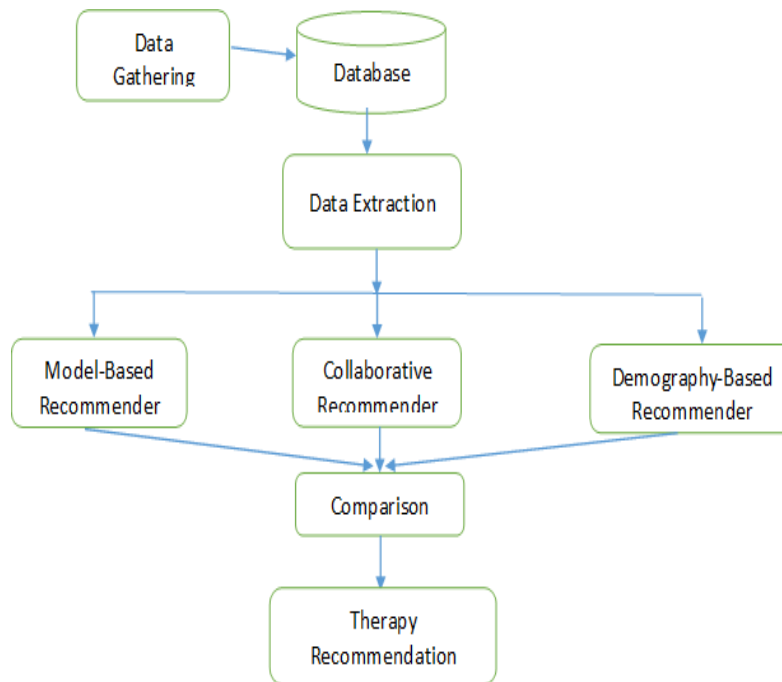
Figure 1 Block Diagram of Hybrid Recommender System

Here, all attributes respected for similarity computation, that is, affinity entries for previously applied therapies, are of ratio scale and within the same range of values. However, the affinity matrix is characterized by only intermittently available entries, that is, resulting in sparse data vectors to be compared. Three similarity metrics proposed in the recommender system literature are investigated and compared in this work. Vector Similarity, which originates from information filtering using vector space models, is widely used in collaborative filtering algorithms. Vector Similarity simply computes the cosine of the angle between two vectors $r_{v,m}$ and $r_{k,m}$ representing two consultations $v \in V$ and $k \in V$, respectively. Furthermore, the degree of linear relationship between two vectors can be quantified using the Pearson Correlation derived from a linear regression model. This metric relies on the assumption that a linear relationship must exist and the errors are independent and have a probability distribution with zero mean and constant variance. However, these assumptions are often violated in the context of collaborative filtering data which can deteriorate the outcome accuracy. To overcome these stated model assumptions, the Spearman Rank Correlation.

B. Demographic-based Recommender Algorithm

To overcome the limitations related to the above described collaborative filtering approach concerning lacking information and cold start, the Collaborative Recommender is extended to utilizing all patient describing information summarized in Table 1 to represent a consultation. The straightforward underlying assumption here is that the available patient describing data carries sufficient information for facilitating meaningful comparisons between consultations. However, as already stated , the patient describing data employed for consultation comparison is not only sparse but the attributes involved into the similarity calculation are characterized by inhomogeneity, that is, are of various level of measurement (dichotomous, nominal, ordinal, interval, and ratio scaled). The similarity measure utilized in this work facilitating both, handling missing values and varying levels of measurement, is the Gower Similarity Coefficient. Here, the level of measurement of the individual attributes is respected for each attribute comparison. Furthermore, the Gower coefficient offers the opportunity to control the individual attribute's impact on overall similarity by assigning specific weights to attributes.

IV. RESULTS

A. Dataset

Table 1 Meta Data – Treatment Table

Field	Type	NULL	KEY
Treatment_id	char(7)	NO	PRI
Treatment_name	varchar(50)	NO	
Methodology	varchar(100)	NO	

sTable 2 Meta Data – Patient Table

Field	Type	NULL	KEY
Patient_id	char(6)	NO	PRI
Name	varchar(25)	YES	
Age	int(11)	YES	
Gender	varchar(7)	YES	
Height	int(11)	YES	
Weight	int(11)	YES	
Blood_group	varchar(6)	YES	
Sugar_level	int(11)	YES	
Disease_id	char(5)	YES	MUL
Treatment_id	char(7)	YES	MUL

Table 3 Meta Data – Disease Table

Field	Type	NULL	KEY
disease_id	char(5)	NO	PRI
disease_name	varchar(20)	NO	
disease_type	varchar(20)	NO	

Table 4 Meta Data – Consultation Table

Field	Type	NULL	KEY
Id	int(11)	NO	PRI
patient_id	char(6)	YES	MUL
consultation_date	Date	YES	
development_in_face	int(11)	YES	
development_in_hands	int(11)	YES	
development_feet	int(11)	YES	
self_assessment	int(11)	NO	
Effectiveness	Float	NO	
days_count	int(11)	NO	

In table 1 shows 3 attributes—Treatment_id, Treatement_name and Methodology. Treatment_id is considered as a char type and also it is a primary key, Treatment_name and Methodology are varchar type.   Table 2 contains 10 attributes, there are Patient_id, Name, Age, Gender, Height, Weight, Blood_group, Sugar_level, Disease_id, Treatment_id. Age, Height, Weight and Sugar level are considered as int type. Other attributes are considered as character type. In table 3 describes about disease details and table 4 is consultation table that describes therapy details about the patient and patients with similar disease are grouped based on the algorithm used.

B.   Evaluation Metrics
In this work, two different evaluation metrics are considered. On the one hand, the individual recommender engine or system yields to predict the response to specific therapies. If the prediction meets the real therapy response, the system can provide the medical practitioner with a reliable support for his decision- making based on the estimation. To quantify the difference between estimated response and real response, the RMSE for a specific consultation is computed between provided affinity entries and predictions. RMSE reflects the rating error in the same value domain as the actual affinity measure with large errors having more impact.

$$\text{RMSE}_v = \frac{1}{Mv}\sqrt{\sum_{m=1}^{Mv}(p_{v,m} - r_{v,m})^2}$$

On the other hand, N top-ranked therapies are usually selected from the affinity predictions for a consultation under consideration and presented to the user. For evaluating recommendation quality, decision support accuracy metrics are commonly utilized. Initially, all consultations are divided into those cases where at least one of the top N therapies recommended were actually applied in the respective consultation and those where the therapies were not compliant. Furthermore, therapies were considered having good outcome if the affinity assigned to that therapy complies with $r_{v,m}$ ≥ 0 5 and $r_{v,m}$ < 0 5 otherwise which leads to the definitions summarized in Table 5.

The outcome-driven precision describes the ratio of all therapies recommended by the system for a consultation v ∈V, that is, top N therapies, which were applied and show good response, that is, are considered successful, and is defined as

$$\text{Precision @N}v = \frac{TP_v}{TP_v + FP_v}$$

Table 5 Outcome-driven evaluation definitions.

	Good Outcome	Bad Outcome
Recommendations compliant	TP	FP
Recommendations not compliant	FN	TN

C.   Summary of Results

The Collaborative Recommender Algorithm has shown good results but there are some drawbacks in the algorithm, to overcome these drawbacks Demographic-based Recommender is introduced. Together they have given better results. Regarding recommendation precision, the ground truth is obtained from all consultations having one or more therapies which showed good response. The Demographic-based Recommender affinity prediction error is significantly higher. All recommendations which could not be provided by the Collaborative Recommender due to missing information were imputed by a weighted averaging Demographic-based Recommender showing the best performance The Collaborative Recommender utilizing basic collaborative filtering algorithms, considering only therapy outcome from previously applied therapies for consultation representation, outperforms the Demographic-based Recommender approach. The weighted averaging Collaborative Recommender method taking the similarity weight into account demonstrates better performance than simple averaging over all neighborhood sizes studied in this work.

CONCLUSION

In this work, Therapy Recommendation is done by analyzing the response from previous records which are similar to the given patient's record. Even though there is an extensive impact of recommender systems in other domains, application Data Mining in healthcare are still rare to date. Dependent on the data employed for determining similarity between consultations and therapy outcome estimation, two approaches were compared. For both algorithms, a Collaborative Recommender approach and Demographic-based Recommender, various variations were studied concerning similarity metric, considering credibility of the computed similarity and aggregation of the respected information for estimating potential therapy response. All algorithms were evaluated using the accuracy of predicting the outcome and according to the precision with which the top 3 recommended therapies meet the ground truth. However, as ground truth only therapies were accepted which have shown good response, therapies with bad outcome were neglected. Therefore, evaluation could only be done on a fraction of the already rather limited database. However, it is assumed that the performance of the approaches studied in this work depend substantially on the amount of available data and will improve considerably if scaled to larger datasets.

REFERENCES

[1]. Sebastian Zaunseder, "Therapy Decision Support Based on Recommender System Methods", Hindawi Journal of Healthcare Engineering, Volume 2017, March 2017.
[2]. P.T.Kavitha, Dr.T.Sasipraba, "HealthCare Decision Support System using Distributed Data Mining based on Ranking", International Journal of Engineering and Innovative Technology (IJEIT), Volume 01, Issue 06, June 2012.
[3]. Dinh Thuan Nguyen, Gia Toan Nguyen, and Vu Tuan Nguyen Lam, " An Approach to Data Mining in Healthcare: Improved K-means Algorithm", Journal of Industrial and Intelligent Information, Volume 01, Issue 01, March 2013.
[4]. Liu H, Kong X, Bai X, Wang W, Bekele T. M, and Xia F, "Context-based collaborative filtering for citation recommendation", IEEE Access, Volume 03,September 2015.
[5]. James Schaffer, Tobias Hollerer, John O'Donovan, "Hypothetical Recommendation: A Study of Interactive Profile Manipulation Behavior for Recommender Systems", Dept. of Computer Science,University of California, Santa Barbara, 2015.
[6]. S.Saraswathi, Dr. Mary Immaculate Sheela," A Comparative Study of Various Clustering Algorithms in Data Mining ", International Journal of Computer Science and Mobile Computing, Volume 03, Issue 11, November 2014.
[7]. Sangeeta Oswal1, Gokul Shah, " A Study on Data Mining Techniques on Healthcare Issues and its uses and Application on Health Sector ",Department of MCA Vivekanand Education Society Institute of Technology, Chembur, Mumbai, India , Volume 07, Issue 06 ,June 2017.
[8]. Joonseok Lee, Mingxuan Sun, Guy Lebanon, "A Comparative Study of Collaborative Filtering Algorithms", IEEE Transactions on Knowledge and Data Engineering, May 2012.

[9]. P. IndiraPriya, Dr. D.K.Ghosh, "A Survey on Different Clustering Algorithms in Data Mining Technique ", International Journal of Modern Engineering Research (IJMER), Volume 03, Issue 01, January 2013.

[10]. Nikita Kamble, Manjiri Harmalkar, Manali Bhoir, Supriya Chaudhary, "Smart Health Prediction System Using Data Mining", International Journal of Scientific Research in Computer Science Engineering and Information Technology, Volume 02, Issue 02 ,April 2017.

[11]. Prof. Mamta Sharma, Farheen Khan, Vishnupriya Ravichandran," Comparing Data Mining Techniques Used For Heart Disease Prediction", International Research Journal of Engineering and Technology (IRJET), Volume 04, Issue 06, June 2017.

[12]. Sagar Bhise, Prof. Sweta Kale, "Effieient Algorithms to find Frequent Itemset Using Data Mining", International Research Journal of Engineering and Technology (IRJET), Volume 04, Issue 06 , June 2017.

[13]. S. Celik, O. Yilmaz, "Comparison of Different Data Mining Algorithms For Prediction Of Body Weight From Several Morphological Measurements In Dogs", The Journal of Animal & Plant Science, February 2017.

[14]. Sundararajan S, Dr.Karthikeyan S, "Study On Spatial Data Clustering Algorithms In Data Mining",International Journal Of Engineering And Computer Science Volume 01,Issue 01 ,October 2012.

[15]. Yang Zhang, Simon Fong,Jinan Fiaidhi, and SabahMohammed,"Real-Time Clinical Decision Support System with Data Stream Mining ",Hindawi Publishing Corporation Journal of Biomedicine and Biotechnology,Volume 2012, May 2012.