Application of recommender system methods for therapy decision support

Felix Graber, Hagen Malberg, Sebastian Zaunseder, Stefanie Beckert, Denise Kuster, Jochen Schmitt

Angaben zur Veröffentlichung / Publication details:

Graber, Felix, Hagen Malberg, Sebastian Zaunseder, Stefanie Beckert, Denise Kuster, and Jochen Schmitt. 2016. "Application of recommender system methods for therapy decision support." In *IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), 14-16 September 2016, Munich, Germany*, edited by Christoph Thuemmler, Chunxue Bai, and Alois Paulin, 1–6. Piscataway, NJ: IEEE. https://doi.org/10.1109/healthcom.2016.7749495.

Application of Recommender System Methods for Therapy Decision Support

Felix Gräßer, Hagen Malberg and Sebastian Zaunseder
Institut für Biomedizinische Technik
Technische Universität Dresden
Dresden, Germany
Felix.Graesser@tu-dresden.de

Stefanie Beckert, Denise Küster and Jochen Schmitt
Zentrum für Evidenzbasierte Gesundheitsversorgung
Universitätsklinikum Dresden
Dresden, Germany
Jochen.Schmitt@uniklinikum-dresden.de

Susanne Abraham
Klinik und Poliklinik für Dermatologie
Universitätsklinikum Dresden
Dresden, Germany
Susanne.Abraham@uniklinikum-dresden.de

***Abstract*—In this paper two approaches for facilitating therapy decision support are proposed and compared. Both approaches, the *Collaborative Recommender* and hybrid *Demographic-based Recommender*, are based on recommender system methods which origin from the field of product recommendation in e-commerce applications. An exemplary dataset comprising health record excerpts of patients suffering from the skin disease psoriasis is used for evaluating both approaches. The approaches estimate the outcome of a subset of systemic therapies to support the medical practitioner in making therapy decisions for a specific patient and time, i.e. consultation under consideration. Both systems proved to work and are capable of assisting medical practitioners prospectively with making appropriate therapy decisions.**

I. Introduction

Recommender systems are described as systems making personalized product suggestions or guiding users to individually useful objects within a large space of possible options by applying statistical and knowledge discovery techniques based on previously recorded data [1]. Today, recommender systems are an accepted and widespread technology used by many market leaders of various industries (e.g. Amazon[1], Netflix[2], Spotify[3]) and online merchandising cannot be imagined without it anymore. The field of recommender systems has evolved considerably over the years yielding extremely sophisticated and specialized methods depending on domain, purpose and personalization level [2].

In contrast to traditional classification algorithms in machine learning, recommender system techniques are robust on sparse datasets and no extensive data collections or domain knowledge are required. A taxonomy of recommendation algorithms can be made differing between content-based [3], collaborative filtering [1] and more sophisticated hybrid approaches [4]. All approaches have in common to convert estimations of a user's preference for items into recommendations using explicit or implicit previous ratings as expressions of preference. While the content-based approach links preference to item attributes the collaborative filtering method considers the rating of other users in the system to make personalized prediction on an active user's preference.

The large volume of daily captured data in healthcare institutions and even in out-of-hospital settings opens up new perspectives to improve clinical practice. However, due to the amount of data, its high dimensionality and complex interdependencies within the data, an efficient integration of the available information cannot be expected from physicians or other health professionals without aids. Therefore, intelligent clinical decision support systems (CDSS) can play a significant role providing such assistance with making diagnosis and treatment decisions or even in clinical quality control [5]. The interpretative analysis of large-scale clinical information e.g. from electronic health records (EHR) which is characterized by its volume, complexity and dynamics form a promising basis for such systems. CDSS usually rely on manually encoded rule based expert knowledge which creates a bottleneck during development and updating. Using automatically extracted associations based on data mining methods facilitates more individual and even unknown patterns to be revealed.

Employment of CDSS and data mining applications for therapy decision support is still very limited which we believe to be closely connected with limitations of typical machine learning methods as stated beforehand. In contrast, collaborative filtering algorithms are capable of handling sparse and heterogeneous data. Additionally, they are permitting insight into the decision making process making results interpretable. Nevertheless, to our knowledge there is only work loosely related to the idea of using collaborative filtering for CDSS including research on collaborative filtering and studies from interdisciplinary and medical communities on CDSS. There are few works applying collaborative filtering algorithms for disease risk or mortality prediction [6][7][8] and on a nursing care plan recommender system [9]. However, we are not aware of any work on therapy or treatment recommendation systems applying collaborative filtering techniques.

[1] www.amazon.com
[2] www.netflix.com
[3] www.spotify.com

In this work we study transferring the idea of collaborative filtering to the domain of CDSS by developing a recommender system aiming at predicting the adequacy of various therapies for a given patient at a given time. Therefore, two methodologies for therapy adequacy estimation, a *Collaborative Recommender* and a hybrid *Demographic-based Recommender*, are compared. A physician can incorporate that information into his decision on the therapy to be chosen. The exemplary recommender system is developed targeting therapy recommendations for patients suffering from the autoimmune skin disease psoriasis.

II. Data

The data for exemplary development and evaluation of the therapy recommender system are excerpts from health records provided by the Clinic and Polyclinic for Dermatology, University Hospital Dresden. The dataset comprises $V = 1072$ consultations from 192 patients suffering from various types of psoriasis. For each consultation patient and therapy describing attributes were provided containing demographic data, comorbidities, state of health as well as current and previous local and systemic treatment. The data was manually extracted from health records and stored in a mySQL database. Within a careful revision corrupted and invalid data was corrected or eliminated and padded where information was missing but could be assumed to be constant over consultations for a specific patient. Overall, the data contains $A = 15$ patient describing attributes for each consultation and $T = 3$ therapy attribute for each applied treatment. The available data is of various levels of measurement.

Despite of data padding, attributes are just intermittently available, leading to inhomogeneous and sparse matrices. Table I and Table II summarize patient attributes and therapy information, respectively. All attributes are supplied with scale of measurement, range of values and availability relative to all consultations. In case of comorbidities and therapies the availability is related to all applied comorbidities or therapies, respectively.

Previous treatment is the collection of all relevant therapies applied to a patient preceding his or her first consultation included in the database. Current treatment comprises all therapies applied within the last two weeks preceding the respective consultation. There are both local and systemic therapies contained in the available datasets. However, this study focuses on recommending the most effective systemic therapy out of $M = 15$ available therapies of this class for a given consultation.

Therapies are provided along with an therapy effectiveness indicator which represents the subjective assessment, an objective health state improvement indicator and adverse effects. The objective health state improvement indicator relates to the severity of psoriasis quantified by the Psoriasis Area and Severity Index (PASI). This score ranging from 0 (no disease) to 72 (maximal disease) combines both the skin area affected and the severity of lesions in a single score [10]. Change of PASI between two consecutive consultations (ΔPASI) is assigned as objective health state improvement indicator attribute to all therapies applied between those consultations.

TABLE I
Patient describing attributes

Attribute	Scale	Range	Availability %
Patient Data			
Gender	nominal	1,2	100
Weight	interval	50 - 165	50
Size	interval	99 - 204	36
Year of Birth	interval	1931 - 1998	100
Family Status	dichotomous	0,1	0
Education	ordinal	1,2,3,4,5	0.3
Profession	nominal	1,2,3,4	1.4
Planned Child	nominal	1,2,3	8
Year of First Diagnosis	interval	1950 - 2014	90
Family Anamnesis	ordinal	1,2,3	50
Type of Psoriasis	nominal	1,2,3,4,5,6	100
Comorbidities			
Comorbidity	nominal	1,2,3,...,34	-
Status	ordinal	1,2,3	100
Under Treatment	dichotomous	0,1	45
State of Health			
PASI Score	interval	0 - 43	69

TABLE II
Therapy describing attributes

Attribute	Scale	Range	Availability %
Systemic Therapy	nominal	1,2,3,...,15	-
Effectiveness	ordinal	1,2,3	98
ΔPASI	interval	-37 - 25	42
Adverse Effect	dichotomous	0,1	100
Local Therapy	nominal	1,2,3,...,7	-
Effectiveness	ordinal	1,2,3	90
ΔPASI	interval	-37 - 25	42
Adverse Effect	dichotomous	0,1	100

III. Methodology

A. System Overview

The subsequently described algorithm aims to recommend the 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. Effectiveness, ΔPASI as well as absence of adverse effects are employed for deducing an affinity measure representing the user ratings.

Therefore, in a first prediction step personalized therapy affinity for therapies that have not been applied to a patient are estimated. Available data is used to identify

the most similar consultations to the consultation under consideration to predict a patient's individual affinity to various therapies using therapy ratings within the aggregated nearest neighbours. In a subsequent recommendation step therapies can be ranked according to the affinity measures assigned to the single therapies and the top N ranked entry or entries can be recommended. Here, two approaches were compared, that differ in the information used and the representation of the consultation profile: On the one hand, we apply a *Collaborative Recommender* algorithm using solely the affinity measures from all preceding consultations as consultation representation. On the other hand, we apply a hybrid *Demographic-based Recommender* approach taking additionally all available patient describing data into account for consultation representation to compute the affinity estimation.

In the subsequent sections the affinity computation is detailed, both similarity metrics are described and the actual affinity estimation algorithm is presented.

B. Affinity Model

In conventional recommender systems' applications both explicit and implicit data is collected. Such data comprises explicit ratings or clicked items, items being part of the shopping basket or visited pages, respectively. All such information reflects the user's preference. Here, *affinity* is mathematically quantifying a rating to a therapy by representing effectiveness, ΔPASI and absence of adverse effects. Therefore, preference is modeled using a weighted sum of three parameters attained from effectiveness ($f_{1,v,m}$), change of the PASI score (ΔPASI), i.e. current PASI compared to the PASI of the previous consultation ($f_{2,v,m}$), and adverse effects ($f_{3,v,m}$). Weight w_t allows to vary the impact of the three described components but are all set to one here. As indicated in Table II, not all applied therapies in the dataset are provided with all of the three attributes. Suchlike missing data is encountered by normalization with the sum of weights w_t where $\delta_{t,v,m}$ is set to zero in case of missing values and set to one otherwise. Thus, the affinity of a given patient and consultation $v \in V$ to a therapy $m \in M$ is modeled as

$$r_{v,m} = \frac{\sum_{t=1}^{T} \delta_{t,v,m} \cdot w_t \cdot f_{t,v,m}}{\sum_{t=0}^{T} \delta_{t,v,m} \cdot w_t}$$

where all three affinity components (effectiveness, adverse effects and ΔPASI Score) are mapped to the domain $0 < f_{t,v,m} \leq 1$. Effectiveness is factorized with a constant value resulting in the nominal values 0.25 (poor), 0.5 (moderate) and 0.75 (good). The metric ΔPASI Score is mapped by a negative sigmoid function adjusted to the domain $0..1$ as shown in Figure 1 facilitating large impact on small PASI Score variations declining with increasing absolute value. Finally, the binary adverse effect indicator is mapped to -0.5 if any adverse effect is present and 0 otherwise to penalize the overall affinity measure if adverse effects have occurred. The mapping rules can be summarized as

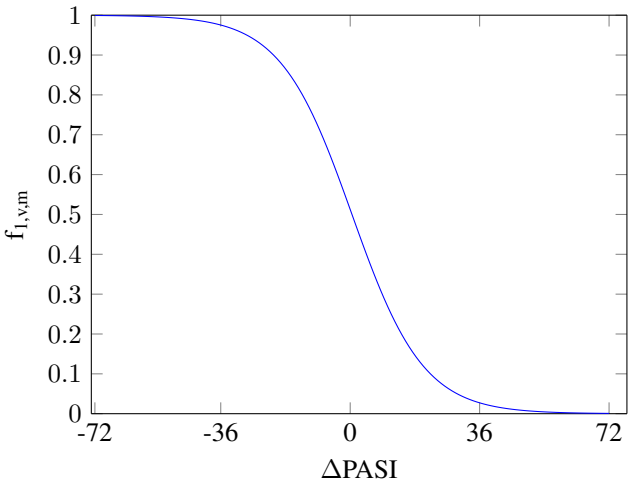

Fig. 1. Sigmoid function mapping the ΔPASI Score to the domain $0 < f_{t,v,m} \leq 1$.

$$f_{1,v,m} = 0.25 \cdot \text{effectiveness}$$

$$f_{2,v,m} = 1 - \frac{1}{1 + e^{-0.1 \cdot \Delta\text{PASI}}}$$

$$f_{3,v,m} = \begin{cases} -0.5 & \text{if adverse effect} \\ 0 & \text{otherwise} \end{cases}$$

C. Similarity Metrics

Based on similarity between consultations, both algorithms *Collaborative Recommender* and *Demographic-based Recommender* estimate affinity on therapies which were not yet applied to a patient and are thus not yet rated. Similarity computation has critical impact on the recommender system output and is highly dependent on the consultation representation. The basic idea of the similarity computation between two consultations $v \in V$ and $k \in V$ is to calculate similarity $w_{v,k}$ between all consultation pairs considering only therapies $m \in M$ which are co-rated in both consultations.

The distinction between the two aforementioned recommendation techniques results from the information which used for consultation representation and consequently the similarity metric applied. In this section both similarities and the data considered for similarity computation are detailed.

1) Collaborative Recommender: Here, the consultation profile is only represented by the therapies and related affinities applied up to the consultation under consideration. The assumption is made that the therapy applied to a given patient within the medical history and the associated outcome reincorporates information about that respective patient and consultation that can be transferred to patients with similar medical history.

The attributes respected for similarity computation are all

interval scaled and within the same range of values hence the *Vector Similarity*

$$w_{v,k} = \frac{\sum_{m=1}^{M} r_{v,m} \cdot r_{k,m}}{\sqrt{\sum_{m=1}^{M} r_{v,m}^2} \cdot \sqrt{\sum_{m=1}^{M} r_{k,m}^2}}$$

can be utilized which is widely used in collaborative filtering algorithms. Vector Similarity origins from information retrieval and 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. The resulting overall recommender algorithm is related to a user-user Collaborative-Filtering Recommender [2] correlating between consultations as users to recommend therapies as items.
As can be seen in Table II, only a small fraction of all available therapies being applied per patient and consultation and only a reduced number of therapy preference parameters are available for those applied therapies. This results in an extremely sparse affinity matrix which is relied upon when computing similarity. Another drawback of this approach is the *cold start* problem occurring when a new patient is included into the system. Lacking information makes it difficult or even impossible to find similar consultations [2].

2) Demographic-based Recommender: To overcome the limitations related to the above described collaborative filtering approach, the *Collaborative Recommender* is extended to utilizing all patient describing information summarized in Table I to form a consultation profile. The underlying assumption is that the available data is carrying sufficient information for facilitating meaningful comparisons between consultations.
The attributes involved into the similarity calculation are characterized by sparsity and inhomogeneity, i.e. are of various level of measurement (nominal, ordinal, interval scale). The similarity measure utilized in this work facilitating both handling missing values and varying levels of measurement is the *Gower Similarity Coefficient* [11]. Here, overall similarity $w_{v,k}$ is computed out of the individual attribute similarities $\rho_{v,k,m}$ depending on their presence $\delta_{v,k,m}$ and assigned weights w_m

$$w_{v,k} = \frac{\sum_{m=1}^{M} \delta_{v,k,m} \cdot w_m \cdot \rho_{v,k,m}}{\sum_{m=1}^{M} \delta_{v,k,m} \cdot w_m}$$

The computed overall similarity is normalized with the sum of weights w_m of present values. $\delta_{v,k,m}$ is set to zero in case of missing values and set to one otherwise. The definition of $\rho_{v,k,m}$ differs depending on the data type: For similarity computation between ordinal and interval scaled values the Manhattan distance normalized to the attribute range is utilized, whereas for nominal or dichotomous attributes Simple Matching (M-coefficient) or Jaccard similarity coefficient (S-coefficient) are applied, respectively.

D. Affinity Estimation

For calculating an affinity estimate $r_{v,m}$ for a consultation under consideration $v \in V$ and therapy $m \in M$ a simple weighted average

$$p_{v,m} = \frac{\sum_{k=1}^{K} r_{k,m} \cdot w_{v,k}}{\sum_{k=1}^{K} |w_{v,k}|}$$

is used. The summations are over all K most similar consultations $k \in K$ for therapy m with $w_{v,k}$ being the weight between consultation v and k representing similarity and $r_{k,m}$ being the affinity during consultation k on therapy m. The number of nearest neighbours K included in the computation is crucial and needs to be chosen cautiously. Furthermore, when computing the affinity prediction previous consultations of the same patient are discarded.

IV. Evaluation

Recommending therapies differ in two essential aspects from e-commerce applications being the traditional application of recommender systems. Firstly, the described affinity measure reflecting user ratings are not representing user taste but rather are related to a patient's response on a specific therapy. Secondly, it is not the patient making the choice for an item but the medical practitioner which of course incorporates much more information in his decision, like exclusion rules and medication cost. To model such determinants is possible, but out of scope of this work.
In this work, a recommender system yields to predict the outcome of different therapies. If the prediction meets the reality, the system can provide the physician a support for its decision based on this estimated outcome. Metrics to measure success of the recommender system thus quantify the difference between estimated outcome and real outcome. Here, the Root Mean Squared Error (RMSE) for a specific consultation is computed considering only the available 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 [12].

$$RMSE_v = \frac{1}{M_v} \sqrt{\sum_{m=1}^{M_v} (p_{v,m} - r_{v,m})^2}$$

A. Results

Initially, the affinity matrix comprising all $V = 1072$ available consultations can be computed in advance to the affinity prediction step and recommendation procedure. Figure 2 shows an excerpt of the affinity matrix demonstrating 25 randomly selected consultations. Its density, i.e. the occupied fields, amounts to 17.25%. Using the aforementioned affinity or the patient describing consultation data two similarity matrices are computed, respectively.

To study the performance of the proposed method a leave one out cross validation (LVOOCV) is run, meaning that

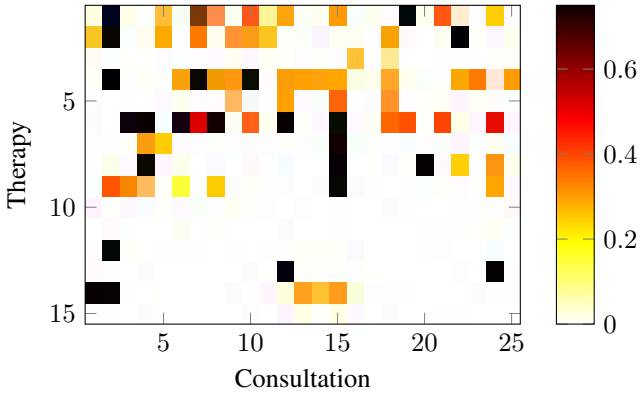

Fig. 2. Affinity Matrix for 25 randomly selected consultations.

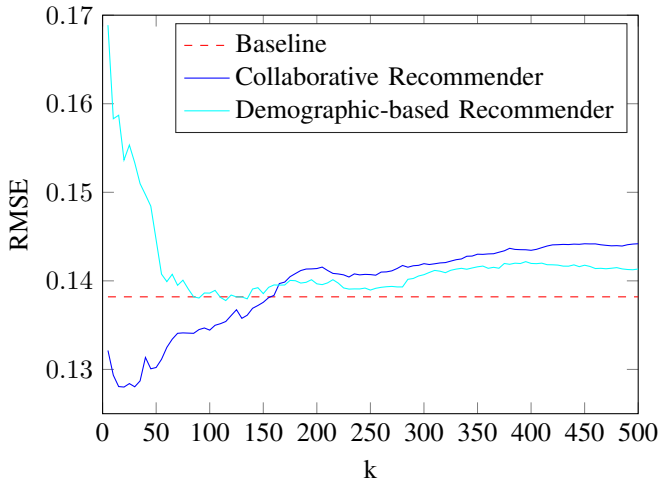

Fig. 3. RMSE of affinity estimates computed using Collaborative Recommender, Demographic-based Recommender and baseline for various k. The baseline is defined as mean affinity for each therapy.

each consultation is evaluated individually using all other patients' consultation information. The mean of the $V = 1072$ evaluation results make the overall RMSE.

The number of aggregated nearest neighbours k effects distinctly the affinity estimates as can be seen in Figure 3. The numerical results are summarized in Table III for a selection of included nearest neighbours $k = 5, 25, 50, 125, 250$ and 500. Both methods differ considerably in their behaviour. In case of the *Collaborative Recommender* approach the prediction error is small for just few nearest neighbours having a minimum at around $k = 20$, before the RMSE monotonically increases with increasing k. The performance of the hybrid *Demographic-based Recommender* affinity estimation is poor for just few neighbours but the prediction error declines quickly until reaching a minimum at around $k = 115$ nearest neighbours. However, this approach is only narrowly capable of outperforming the baseline error which is defined as the RMSE between mean affinity for each therapy computed from all available consultations and the applied therapies for a consultation under consideration. As detailed in Table III besides the aforementioned behaviour it can be shown that in both cases variance (standard deviation in brackets) of the estimated affinity over all evaluated consultations decreases with increasing number of neighbours included in the computation.

TABLE III
MEAN RMSE AND STANDARD DEVIATION FOR VARIOUS k NEAREST NEIGHBOURS

k	RMSE
Collaborative Recommender	
5	0.1321 (0.1235)
25	0.1284 (0.1031)
50	0.1302 (0.0999)
125	0.1367 (0.0930)
250	0.1407 (0.0891)
500	0.1442 (0.0860)
Demographic-based Recommender	
5	0.1689 (0.1473)
25	0.1553 (0.1206)
50	0.1446 (0.1064)
125	0.1382 (0.0922)
250	0.1390 (0.0867)
500	0.1413 (0.0887)

V. CONCLUSION

We have presented the application of recommender system algorithms in the context of therapy decision support. To the best of our knowledge - despite the immense impact of recommender systems in other domains - applications of recommender systems to the medical domain are rare to date. Two approaches were compared differing in the data used for determining similarity between consultations and computing affinity estimates. A *Collaborative Recommender* employing basic collaborative filtering algorithms outperforms the hybrid *Demographic-based Recommender* approach integrating patient describing data into the computation. The similarity computation underlying the *Demographic-based Recommender* is affected unfavourably by less relevant information included into the calculation whereas more important factors have too little effect. Nevertheless, including more information into the recommendation is essential to overcome the limitations in cases were no medical history or just little information is available for a specific patient under consideration. Furthermore, it is assumed that performance of both approaches will improve considerably if scaled to larger datasets.

Therefore, future work will concentrate on improving the *Demographic-based* system using attribute weighting methods and on more sophisticated hybrid approaches [4]. Also regarding the *Collaborative Recommender* extensions, i.e. rating normalization techniques, respecting inverse user frequency or time variance, significance weighting depending on data availability and extracting latent factors using singular

value decomposition [2], will be studied and applied to much more comprehensive datasets. Moreover, RMSE is solely respecting the affinity estimation error but neglects the fraction of available affinity estimates. Therefore, in future work additional evaluation metrics will be considered respecting actual recommendation quality additionally to affinity prediction error.

ACKNOWLEDGMENT

This work is part of the project *Therapieempfehlungssystem* which is funded by the *Roland Ernst Stiftung für Gesundheitswesen*.

REFERENCES

[1] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," pp. 285–295, 2001.

[2] X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Advances in Artificial Intelligence*, vol. 3, no. Section 3, pp. 1–19, 2009.

[3] M. J. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artificial Intelligence Review*, vol. 13, no. 5, pp. 393–408, 1999.

[4] R. Burke, "Hybrid Recommender Systems : Survey and," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[5] P. E. Beeler, D. W. Bates, and B. L. Hug, "Clinical decision support systems." *Swiss medical weekly*, vol. 144, no. 5, p. w14073, 2014.

[6] D. a. Davis, N. V. Chawla, N. a. Christakis, and A. L. Barabási, "Time to CARE: A collaborative engine for practical disease prediction," *Data Mining and Knowledge Discovery*, vol. 20, no. 3, pp. 388–415, 2010.

[7] S. Hassan and Z. Syed, "From netflix to heart attacks: Collaborative filtering in medical datasets," *IHI'10 - Proceedings of the 1st ACM International Health Informatics Symposium*, pp. 128–134, 2010.

[8] M. Komkhao, J. Lu, and L. Zhang, "Determining Pattern Similarity in a Medical Recommender System," *Data and Knowledge Engineering*, pp. 103–114, 2012.

[9] L. Duan, W. N. Street, and E. Xu, "Healthcare information systems: data mining methods in the creation of a clinical recommender system," *Enterprise Information Systems*, vol. 5, no. 2, pp. 169–181, 2011.

[10] T. Fredriksson and U. Pettersson, "Severe Psoriasis Oral Therapy with a New Retinoid," *Dermatologica*, vol. 157, no. 4, pp. 238–244, 1978.

[11] J. C. Gower, "A General Coefficient of Similarity and Some of Its Properties," *Biometrics*, vol. 27, no. 4, pp. 857–871, 1971.

[12] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5–53, 2004.