Neighborhood optimization for therapy decision support

Felix Gräßer, Hagen Malberg, Sebastian Zaunseder

Angaben zur Veröffentlichung / Publication details:

Gräßer, Felix, Hagen Malberg, and Sebastian Zaunseder. 2019. "Neighborhood optimization for therapy decision support." *Current Directions in Biomedical Engineering* 5 (1): 1–4. https://doi.org/10.1515/cdbme-2019-0001.

Felix Gräßer*, Hagen Malberg, and Sebastian Zaunseder

Neighborhood Optimization for Therapy Decision Support

https://doi.org/10.1515/cdbme-2019-0001

Abstract: This work targets the development of a neighborhood-based Collaborative Filtering therapy recommender system for clinical decision support. The proposed algorithm estimates outcome of pharmaceutical therapy options in order to derive recommendations. Two approaches, namely a Relief-based algorithm and a metric learning approach are investigated. Both adapt similarity functions to the underlying data in order to determine the neighborhood incorporated into the filtering process. The implemented approaches are evaluated regarding the accuracy of the outcome estimations. The metric learning approach can outperform the Relief-based algorithms. It is, however, inferior regarding explainability of the generated recommendations.

Keywords: Clinical Decision Support System, CDSS, Therapy Recommender System, Neighborhood Optimization.

1 Introduction

Clinical decision support systems (CDSS) are intended to provide assistance for personalized diagnosis and treatment decisions [1]. Suchlike systems are expected to play an increasingly important role in future healthcare. Especially data-driven approaches, employing data mining and machine learning techniques to exploit the large volume of daily captured and widely unused clinical data, promise to open up new perspectives. In contrast to expert systems, which derive recommendations or suggestions using knowledge stored in rule sets (if-then rules), data-driven approaches are supposed to be capable of extracting knowledge automatically from the available data [1]. However, in order to facilitate a high degree of acceptance among medical practitioners, such approaches are required to provide reliable and interpretable decision support. This work aims at developing a CDSS which supplies the attending physician with individualized and patient-specific treatment recommendations. To that end, this work transfers methodologies from the field of Recommender Systems (RS) research to the CDSS domain. RS are widely applied in other domains such as e-commerce or music and movie streaming services. In those applications, sophisticated and specialized approaches were developed over the recent years to provide a target user with personalized product recommendations [2]. Such methods can be capable of meeting both the stated reliability and interpretability requirements. Specifically, this work deals with finding a similarity function optimized for the data at hand, which is fundamental for the widely used class of neighborhood-based RS algorithms.

***Corresponding author: Felix Gräßer,** Institute of Biomedical Engineering, Technical University Dresden, Dresden, Germany, e-mail: felix.graesser@tu-dresden.de
Hagen Malberg, Institute of Biomedical Engineering, Technical University Dresden, Dresden, Germany
Sebastian Zaunseder, Department of Information Technology, University of Applied Sciences and Arts Dortmund, Dortmund, Germany

2 Materials and Methods

2.1 Psoriasis Data

The CDSS algorithms proposed in this work are evaluated on the basis of a clinical dataset consisting of 1242 consultation representations X from 239 patients suffering from various types of the skin disease Psoriasis [3]. Each consultation representation x incorporates the individual therapy history, as well as demographic and condition related data, adding up to a total of 125 attributes. The level of measurement of the present data ranges from binary and nominal qualitative attributes to ordinal and ratio scaled quantitative attributes. The overall objective of the therapy RS is to predict the numerically decoded outcomes y of 7 systemic pharmaceutical therapy options based on x in order to provide the treating physician with a ranked list of therapies.

2.2 Collaborative Filtering for Therapy Decision Support

Deriving recommendations based on the local neighborhood of a target user is a straightforward and efficient approach denoted as Collaborative Filtering (CF) [2]. CF identifies users with similar taste by comparing purchase histories or product ratings and derives potentially most preferred products. This approach was transferred to therapy recommendation in our previous work [3]. Here, consultations where compared using representations as introduced above in order to derive treat-

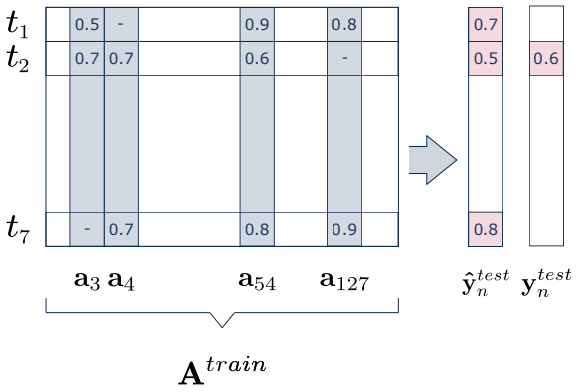

Fig. 1: Outcomes $\hat{y}_n^{test}$ of treatment options t are estimated for a test consultation n based on all outcomes observed in the treatment history of the K most similar training data consultations. A^{train} accumulates the numerically encoded outcomes of all previously applied treatment options, i.e. the treatment history of the training consultations. In A^{train} rows represent treatment options and columns training data consultations.

ment recommendations with potentially good outcome. Therefore, the numerically decoded outcomes $y_n^{test} \in [0, 1]$ of potential treatment options, ranging from bad to good response, are estimated for a test patient and consultation n. To do so, outcomes observed in the treatment history of the K most similar consultations to n are averaged for each therapy option as pictured in figure 1. The outcome-consultation matrix A^{train} accumulates outcomes of ever applied treatments, i.e. the treatment history for each training consultation. Finally, the Root Mean Squared Error (RMSE) can be computed between outcome estimate $\hat{y}_n^{test}$ and actually observed outcome y_n^{test} to evaluate the estimation accuracy.

Both, the local neighbourhood of n included into the estimation and the coefficients, for calculating the weighted average of the observed outcomes are defined by a similarity measure s_k^n for each training consultation k. Here, a similarity function $s(x_n^{test}, x_k^{train})$ defines s_k^n for test and training consultation representations x_n^{test} and x_k^{train}.

3 Neighborhood Optimization

3.1 Similarity Assumptions

The similarity function $s(x_n^{test}, x_k^{train})$ itself and the impact of attributes incorporated into the similarity computation determine the computed outcome estimate in the CF setting. In this work we compare two methods that both automatically adapt $s(x_n^{test}, x_k^{train})$ to the data at hand in order to find an appropriate neighborhood and averaging coefficients. Both investigated approaches assume a supervised classification problem where each instance is associated with a distinct class. In the present setting this corresponds to a priori assumptions regarding similarity or dissimilarity. Each consultation is characterized by a numeric outcome indicator associated with the applied treatment option and unknown outcome for all other options which have not been applied (unobserved ground truth). Consequently, assumptions regarding similarity or dissimilarity between a pair of consultations can only be derived from those consultations which applied therapies in common and for which in both cases outcome is known. Figure 2 assumes a training consultation representation x_n^{train} which is associated with a treatment which showed good response (> 0.5). Thus, neighboring consultations x_k^{train} are labeled as similar to x_n^{train} if the same treatment is present in the outcome-consultation vector a_k^{train} holding the treatment history of x_k^{train} and if this treatment also has shown good outcome. Conversely, neighboring consultations are labeled as dissimilar to consultation n if the same treatment is present in a_k^{train} but this treatment has shown bad response (≤ 0.5). Regarding neighboring consultation representations x_k^{train} for which is true that the in n applied therapy was never applied, no information regarding the similarity label is available.

3.2 Attribute Weighting

We assume that individual attributes are of varying importance concerning the similarity between consultations or even are entirely irrelevant. The baseline metric for computing similarity is the *Gower coefficient* as already successfully applied in [3]. The *Gower coefficient* differentiates between data types and facilitates attribute weighting when quantifying similarity. A widely and successfully used class of feature weighting and selection algorithms which exploit the concept of similarity are Relief-based algorithms (RBAs) [4]. In this work, the RBA approach is adapted to the aforementioned similarity assumptions. Within an iterative process, a random target consultation representation x_i^{train} is drawn from the training data X^{train} and, based on this sample, each dimension w^d of an attribute weight vector w is updated according to equation 1.

$$w^d = w^d + s^d_{Hits} - s^d_{Misses} \tag{1}$$

Here, in accordance with figure 2, s_{Hits} is the average similarity between target i and the K closest consultations which are also labeled as similar according to the definitions from above. s_{Misses} is the average similarity between target i and the K closest consultations which are labeled as dissimilar. As the attribute weight vector is initialized with 0, attributes whose weights become negative are assumed to be irrelevant or unfavourable and are neglected when computing $s(x_n^{test}, x_k^{train})$. The optimal free parameters, namely the

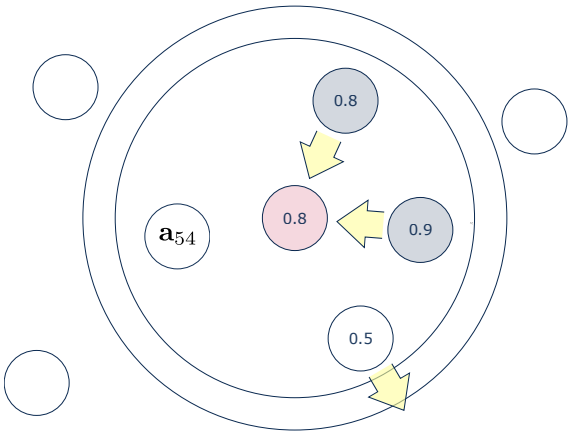

Fig. 2: Neighboring consultation representations x_k^{train} of x_n^{train} with same treatment applied and same outcome (> 0.5) are considered as similar (grey) and vice versa (white). For consultations associated with differing treatment no information about similarity is available (a_{54}). LMNN intents to cause the target sample x_i^{train} to be surrounded by samples of the same class while being separated from samples of different classes.

neighborhood size K and the number of iterations are determined with cross validation.

3.3 Metric Learning

Metric learning based algorithms assume that not only importance of the individual attributes but also the multivariate distribution of the data as well as correlations among attributes have crucial impact on the similarity computation and hence impact the outcome estimation. Here, as baseline metric the *Euclidean distance* is employed to derive similarity between the rescaled consultation representations. *Mahalanobis distance* additionally considers the distribution of the data by measuring distance in standard deviations along the principal components of the data when computing *Euclidean distance*.

$$d(x_n^{train}, x_k^{train}) = (x_n^{train} - x_k^{train})^\top \mathbf{M}(x_n^{train} - x_k^{train}) \tag{2}$$

However, instead of employing the inverse covariance matrix as global transformation $\mathbf{M}$, generalized *Mahalanobis metrics* can incorporate additional constraints. The *Large Margin Nearest Neighbor* (LMNN) algorithm proposed by [5] learns such a generalized *Mahalanobis metric* and is especially intended for neighborhood-based classification algorithms. The overall intention of the LMNN approach is to learn $\mathbf{M}$ such-like that it causes the target sample i to be surrounded by samples k of the same class while being separated from samples of different classes as pictured in figure 2. The loss function, which is optimized to learn $\mathbf{M}$, consists of two competing objectives ϵ_{pull} and ϵ_{push} whose relative impact is controlled using a meta parameter ν. Firstly, large average distances $d(x_i^{train}, x_k^{train})$ between x_i^{train} and the K closest consultation representations x_k^{train} labeled as similar, i.e. the *target neighbors*, are penalized according to equation 3.

$$\epsilon_{pull}(\mathbf{M}) = \sum_{i,k} d(x_i^{train}, x_k^{train}) \tag{3}$$

Secondly, small distances between i and consultations which are labeled as dissimilar and which invade the perimeter (plus a unit margin) established by the *target neighbors*, are penalized according to equation 4. The *hinge loss* $[z]_+ = max(z, 0)$ ensures that only invading consultations contribute to the loss function.

$$\epsilon_{push}(\mathbf{M}) = \sum_{i,k,l} [d(x_i^{train}, x_k^{train}) + 1 - d(x_i^{train}, x_l^{train})]_+ \tag{4}$$

Analogously to the RBA, free parameters such as the neighborhood size K, impact ratio of the two competing objectives ν, and the learning rate μ need to be determined.

4 Evaluation and Results

4.1 Nested Cross Validation

As the consultations of the individual patients cannot be regarded to be independent (i.i.d.), a patient-wise evaluation scheme is applied in this work. Hence, to make most of the available data and to ideally provide an unbiased estimate of the true generalization error a *nested cross-validation* approach is applied for model selection and evaluation. The outer loop (outer cv) implements a leave-one-patient-out cross validation which in each iteration holds out all consultation of one test patient p for evaluation. The inner loop implements a 5-fold cross-validation (inner cv) including the remaining patients' consultations for model selection. To avoid bias due to sample dependencies, also within the inner loop consultations from the same patient never enter different folds. Within this inner loop, the cv performance is calculated for all possible model variants (grid search) and the best performing model parameters are selected. Finally, the RMSE is computed between predicted and actually observed outcome for the hold out consultations of p using all the remaining patients' consultations to compute the outcome estimates.

4.2 Outcome Estimate Accuracy

Figure 3 shows the two baseline metrics *Gower coefficient* and *Euclidean distance* as well as the best performing vari-

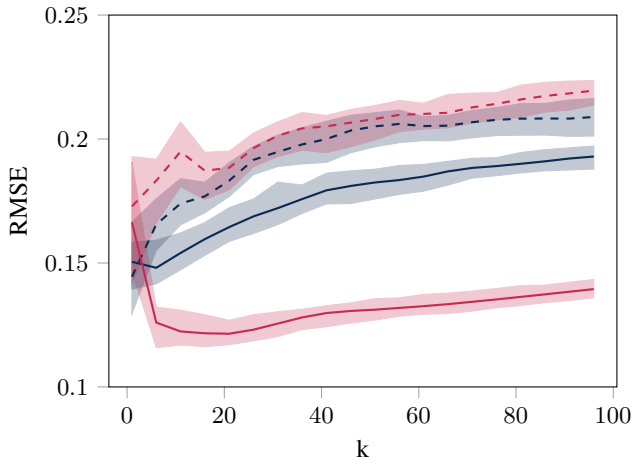

Fig. 3: Mean and value range of the cross-validation RMSE between estimated and observed outcome comparing the two baseline metrics *Gower coefficient* (- - -) and *Euclidean distance* (- - -) and the two optimization strategies RBA (——) and LMNN (——), respectively. RMSE is computed for a neighborhood size range $K \in [1, 100]$.

ant of each of the proposed neighborhood optimization approaches, respectively. Regarding attribute weighting (RBA), the best performance was obtained at a local neighborhood $K_{RBA} = 15$ and iterating once over each training consultation. Regarding the metric learning approach (LMNN), the best performance was obtained with a local neighborhood size $K_{LMNN} = 10$. Furthermore, the inner cv yields best results for setting the impact ratio of the two competing objectives as $\nu = 0.5$, and the learning rate $\mu = 0.001$. For each method, the average inner cv results for all test patients p and over a range of included nearest neighbors $K = [1, 100]$ is shown. As can be seen, the RMSEs vary among test partitions. When choosing the neighborhood size K it must be considered that RMSE can only be computed if there is any overlap of therapies applied in the test consultation and in the therapy history of incorporated neighbors which is not always given for small K. Therefore, a lower boundary for the selected neighborhood size $K \geq 10$ is defined which none of the found K fall below. The outer cv results (table 1), however, do not show the same performance. Two opposing phenomena can be observed. Firstly, in comparison with the inner cv, within the outer loop there is a larger training data set to select the most similar consultations from, which results in overall better performing baseline models. Secondly, as the optimized similarity functions are learned on the entire training folds, the functions are subject to overfitting and the inner cv results may be biased.

Tab. 1: Mean and standard deviation of the outer cv results evaluating each test patient p by applying the best performing model parameters determined in the inner cv.

Method	RMSE	Method	RMSE
Gower	0.1379 (0.1069)	Euclidean	0.1347 (0.1083)
RBA	0.1336 (0.1112)	LMNN	0.1410 (0.1209)

5 Conclusion

Generally, the neighborhood-based CF approach for therapy recommendation allows to estimate therapy outcome which can be utilized to provide decision support. Inspection of this neighbourhood can serve as a basis for explaining and interpreting recommendations. The inner cross validation results show that the estimation of the outcome score varies depending on the method employed. Utilizing data type dependent functions for computing distance or similarity as done by the *Gower coefficient* proves to be beneficial in comparison with *Euclidean distance*. This approach can be further improved by assigning appropriate weights to attributes. Learning a transformation matrix M which, besides only scaling individual attributes is also capable of rotating the basis of the consultation representations outperforms the attribute weighting approach. However, attribute weighting bears, in contrast to the LMNN algorithm, the additional potential to reveal insights into determining factors regarding outcome and, equally to the *Gower coefficient*, is applicable to representations with missing values which is a pervasive challenge in the medical domain. Nevertheless, the differences in the inner and outer cv results show that the limited data set in combination with the applied evaluation strategy causes the inner evaluation loop to be biased and doesn't provide a reliable indicator for model selection.

References

[1] Berner ES. Clinical Decision Support Systems. Springer International Publishing; 2016.
[2] Ricci F, Rokach L, Shapira B, Kantor P. Recommender Systems Handbook. Springer US; 2011.
[3] Gräßer F, Beckert S, Küster D, Schmitt J Malberg, H, Zaunseder S, et al. Therapy Decision Support Based on Recommender System Methods. Journal of Healthcare Engineering; 2017.
[4] Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore J. Relief-Based Feature Selection: Introduction and Review. CoRR; 2017.
[5] Weinberger KQ, Blitzer JC, Saul LK. Distance Metric Learning for Large Margin Classification. Journal of Machine Learning Research; 10: 207–244.