A Comparison of Collaborative Filtering Methods for Medication Reconciliation

Huanian Zheng, Rema Padman, Daniel B. Neill

The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

Abstract

Medication Reconciliation has emerged as a major patient safety goal in the management of medication errors and prevention of adverse drug events. The medication reconciliation process supports the task of detecting and correcting potential mistakes in a patient's medication list so that physicians can make correct, consistent, timely and safe prescribing decisions. Maintaining an accurate list of a patient's medications is a very challenging task for which the current solution is a process driven approach. In prior work, we proposed a promising data driven approach through the use of collaborative filtering algorithms to improve the accuracy of the medication list. This is analogous to the framework used by online retailers to recommend relevant products to customers. In this paper, we extend our original framework to include other types of patient information, develop some new collaborative filtering approaches and test them using medication data from a long-term care clinic. The results are encouraging and suggest several promising directions for the future, including embedding these methods in current medication reconciliation processes and evaluating them in actual clinical settings.

Keywords: Patient Safety, Medication Reconciliation, Collaborative Filtering, Machine Learning

Introduction

Medication reconciliation is the process of creating the most accurate list of all medications that a patient is taking [1, 2]. Given the high human and financial costs resulting from errors in prescribing and dispensing medications, patient safety efforts have focused on medication reconciliation as an important step in improving the current situation [2]. Adverse drug events are adverse outcomes associated with some medication prescribing decisions. A physician's prescribing decision depends on medical knowledge combined with data from many sources that include a patient's demographic information and clinical history such as diagnoses, allergies, and prescription and non-prescription medications. In practice, this knowledge is often incomplete and this lack of relevant information can negatively affect the physician's prescribing decision, resulting in adverse drug events.

The typical medication reconciliation process includes collecting a patient's complete and accurate medication history, clarifying its appropriateness, and documenting any changes in medication orders through reconciliation of differences. While many of these tasks were traditionally processed manually, introducing errors of their own, information technology is beginning to streamline this process.

Electronic health records and prescribing systems currently enable organizations to store and transmit patient's medication records more conveniently and accurately. However, recent studies indicate that there are still significant discrepancies between medication histories derived from any number of valid sources, all of which may be utilized during medication reconciliation [3, 4]. Given the availability of vast amounts of prescription data in these electronic repositories, in prior work, we proposed a data driven approach to address these discrepancies through the use of collaborative filtering (CF) algorithms to improve the accuracy of the medication list [1]. This is analogous to the framework used by online retailers to recommend relevant products to customers. In this paper, we extend our original framework to include other types of patient information, such as specific demographic features, to develop some new collaborative filtering approaches and test them using medication data from a long-term care clinic.

Model Formulation and Evaluation

Medication data can be represented as a compact model in a variety of ways. Each patient's medication information can be represented in the form of a drug list, which is the set of all drugs that has ever been recorded for this patient. The complete medication list of all patients is then a sparse binary matrix, M = $\{m_{ij}\}$, for patients i = 1…N, j=1…M, where

$$m_{ij} = \begin{cases} 1 & \text{if drug j occurs in patient i's list} \\ 0 & \text{other wise} \end{cases}$$

This adjacency matrix representation is analogous to the data representation in the domain of collaborative filtering and has been used in many machine learning papers [5, 6]. As described in [1], another valid definition is to consider a patient's medication information as a list l_i, where l_i constitutes the set of drugs e_j for a given patient i and $e_j \in l_i$ if and only if $m_{ij} = 1$. Besides the drug information, the dataset also provides us with the patients' demographic information, such as the gender and age for each patient. Furthermore, using the Center for Disease Control's Ambulatory Care Drug Database System (www2.cdc.gov/drugs), the drug can also be classified by its generic chemical name, or more generally, as a member of a therapeutic class.

Other relevant notation is as follows:

$D = \{d_1, d_2, \dots, d_m\}$ denotes the complete set of drugs

$L_{Train} = \{l_1, l_2, \dots, l_n\}$ denotes the set of training lists

$\hat{L} = \{\widehat{l_1}, \widehat{l_2}, \dots, \widehat{l_{\hat{n}}}\}$ denotes a set of partial lists

L_i denote the subset of lists that contains drug d_i

L_{ij} denote the subset of lists that contains both drug d_i and d_j

In this paper, we make prediction about a missing drug based on what other drugs have been prescribed for this patient as indicated by the medication list available for this patient. Demographic information is also used to refine the result in order to make more accurate predictions.

The evaluation of the methods uses the standard AllBut1 metric used in the collaborative filtering domain, where the algorithms are compared based on how well they predict the voting of a particular user on one of the items given the user's voting information on other items [7]. Analogously, we compare our algorithms on how well they predict a patient's missing medication (one that has been removed prior to testing) given the patient's remaining medication list. Given the similarity between the two prediction problems, we can adapt and apply many algorithms which have been widely used in CF domain to our medication reconciliation problem. In each case, the algorithm assigns a score for each drug not observed in the patient's list. These drugs are then sorted in decreasing order based on this score and the one with the highest score is assumed to be the most probable missing drug from the partial list.

Collaborative Filtering (CF) Methods

In this section, we introduce the algorithms to apply to our medication reconciliation problem and test their performance. We start with some methods that use only the medication information. Following this, we discuss methods which also use demographic and diagnosis information. Thus, the algorithms applied to the datasets with demographic and diagnosis information is extensions of the medication focused methods.

Naïve Bayes

Naïve Bayes learner is one of the most practical Bayesian learning methods in the literature [8]. It applies to learning tasks where each instance x is described by a conjunction of attribute values with the simplifying assumption that the attribute values are conditionally independent given the target value. In our case, instances are a patient's drug list in which each entity comes from the same drug set, D. We assume that the probability of occurrence of drugs on the list is conditionally independent given another observed drug.

cGraph

The cGraph algorithm [9] is a new graph-based method in link analysis research, which assumes that links are generated based on an unknown underlying graph structure that captures the pair wise relationships between entities. Thus, we assume that drugs appearing on the same list are based on the same, unknown underlying graph structure. The cGraph algorithm approximates the underlying graph using weighted counts of co-occurrences that are accumulated during a single scan of the dataset. In our experiment the edge from drug i to j is approximated as:

$$W_{ij} = P(d_j|d_i) = \frac{\sum_{l:\, l \in L_{ij}} \left(\frac{1}{|l|-1}\right)}{|L_i|} \qquad (1)$$

Here |l| is the size of the list.

At testing phase, we generate the scores for possible missing drugs using the random tree generation model presented in [8]. Thus the probability that d_i is the missing drug in a partial list

$\widehat{I_p}$ is:

$$P(d_i|\widehat{I_p}) = \begin{cases} \frac{1}{|\widehat{I_p}|}\sum_{d_j \in \widehat{I_p}} \frac{P(d_i|d_j)}{\sum_{d_k \notin \widehat{I_p}} P(d_k|d_j)}, & d_i \notin \widehat{I_p} \\ 0, & d_i \in \widehat{I_p} \end{cases} \quad (2)$$

This is the score we assign to each possible missing drug.

D-Sim

Item-Based recommendation [6] is another common technique to solve a collaborative filtering problem. Unlike k-nearest neighbors, item-based method first analyzes the user-item matrix to identify relationships between different items, and then uses these relationships to indirectly compute recommendations for users. Retrieved from the patient-drug matrix, each column represents the corresponding drug, which is thought of as a binary vector in the N dimensional patient-space. The similarity between two drugs is measured by computing the cosine of the angle between two vectors. Formally, similarity between drug i and j is given by

$$sim(i,j) = \cos(\vec{i},\vec{j}) = \frac{\vec{i}\cdot\vec{j}}{||\vec{i}||_2 * ||\vec{j}||_2} \quad (3)$$

Once we get the similarity between each pair of drugs, scores for each drug is computed by summing the similarity between the target drug and every active drug in the patient's medication list.

Demographic Method

An obvious method to use the gender and age information for patients is to group patients into similar categories. The premise is that patients in the same age group are more likely to take similar drugs. Most of the patients in the dataset are between 40 and 100 and there are very few patients under 40. Here we discuss two simple methods to group patients of similar age. First we partition the training lists into 7 groups, with the age range for each group being: 1-40, 41-50, 51-60, 61-70, 71-80, 81-90, and 91 and above. Then during both training and testing phases, we only compute on one of the subsets the testing patient $\hat{p}$ belongs to. An alternative way of using age information is to use a dynamic partitioning method without pre-computation on the training set. Instead of having fixed age range in each subset, we pick those records of patients who are in the interval of age $[age(\hat{p}) - 5, age(\hat{p}) + 5]$ from the training set and run the algorithm on this subset. When $age(\hat{p})$ is too small (less than 45) or too large (greater than 85), we have to retrieve all the patients from 1-40 or 91 and above because the number of patients in these age intervals is quite small.

We use gender information using the simple and straightforward approach of creating two separate subsets, one for only male and another for only female patients. We include these in our experiments to evaluate their impact on prediction accuracy.

Drug-Therapeutic Class Method

As mentioned earlier, each drug can be generalized to a therapeutic class. Within the same clinic, there are many fewer therapeutic classes than there are individual drugs. Therefore, predicting a therapeutic class is a relatively simpler problem than predicting the brand drug name. The hybrid method assumes that if the prediction accuracy for the missing therapeutic class is high, this information may be used to direct prediction of the missing drug. To be more specific, before predicting the exact drug, we use the collaborative filtering method to generate a list of therapeutic classes that are ordered according to the probability to be the missing one for current patient's medication list. After we get an ordered list of possible drugs, any drugs that do not belong to the top 5 possible therapeutic classes are removed from the recommendation list.

Results

In addition to the algorithms described in the previous section, we also implemented three methods described in [1]: Co-occurrence, K-Nearest neighbors, and Popular, to establish a baseline for comparison of our methods. To evaluate these methods on the prediction task, we used medication data from an online pharmacy that provides medications to long-term care clinics in the Eastern United States. The clinic we chose for our experiments is the largest individual clinic in our data set. It contains medication records for 701 patients and 318 different drugs occur in these records.

Table 1 summarizes the results of our methods applied to medication-only data. Each row gives the proportion of patients whose missing drug is ranked 1st / within top 5 in the ordered list of candidate drugs generated by the corresponding algorithm. The top result and any other result that is not significantly different from the top result at $\alpha = 0.05$ are highlighted in bold.

Table 1: Results for medication-only dataset

Methods	Top 1	Top 5
Co-occurrence	0.3283	**0.4216**

KNN	0.3244	0.3898
Popular	0.3226	**0.4181**
N-Bayes	0.3123	0.4011
cGraph	0.2389	0.3238
D-Similarity	**0.3421**	**0.4231**

Among all these methods, D-Similarity shows the highest performance and cGraph the lowest, compared to the baseline methods. cGraph algorithm has only about 24% chance to predict the correct drug at first try. Although technically cGraph is a derivative of co-occurrence with special weighting criterion, its performance is worse than the original co-occurrence method. All of the other methods give similar prediction accuracy. A detailed analysis of the results indicates that both the percentage accuracy and the specific recommended drugs for each case are similar for these methods. In other words, for the same patient, if one of above methods cannot guess the correct drug in a few tries, other methods generally cannot either. The reason for this may be that there are some overwhelmingly popular drugs in this clinic. In particular, the 5 most popular drugs appear in about 90% of the patients' medication lists, while most of the remaining drugs only appear in about 10 lists. Consequently, all the methods are more likely to recommend these frequent drugs.

The Top 1 prediction accuracy of methods using patients' demographic information is presented in Table 2. Columns A-D represents the results for the following settings:

A. Fixed age interval with gender information

B. Fixed age interval without gender information

C. Dynamic age grouping with gender information

D. Dynamic age grouping without gender information

The results show that grouping patients of the same gender does not help improve the performance. Since the proportion of men in the dataset is relatively small (27.5%), making predictions solely based on this subset of records is difficult because of lack of sufficient drug information. The two ways of using age information have similar impact on performance, and both have worse performance than the original methods.

Table 2: Results for demographic method

Methods	**A**	**B**	**C**	**D**
Co-occur	**0.2981**	**0.2871**	**0.2761**	0.2566
KNN	0.2822	0.2692	**0.2878**	**0.2867**
Popular	0.2719	**0.2812**	0.2598	0.2612
N-Bayes	0.2558	0.2498	0.2626	0.2345
cGraph	0.1928	0.2128	0.2082	0.1989
D-Sim	**0.3011**	**0.2803**	**0.2867**	**0.2910**

Table 3 presents the prediction accuracy for the missing therapeutic class. This clinic has only 83 different therapeutic classes compared to 318 distinct drugs, so we would expect better results for our methods. The gains are not significant with respect to top 1 accuracy. However, Top 5 prediction accuracy is substantially higher for predicting therapeutic classes as compared to predicting specific drugs.

Table 3: Results for therapeutic-class prediction

Methods	**Top 1**	**Top 5**
Co-occurrence	**0.3572**	**0.5823**
KNN	0.3312	0.5426
Popular	**0.3505**	0.5695
N-Bayes	**0.3505**	0.5731
cGraph	0.3189	0.5078
D-Similarity	**0.3605**	**0.5918**

We choose D-sim method to recommend therapeutic classes which are used as a filter prior to the actual missing drug prediction since it produces higher accuracy than all the other methods. Table 4 shows the accuracy of drug prediction after removing drugs which do not belong to any of the Top 5 therapeutic classes identified in the preprocessing step. Compared with the results in Table 1, we observe that this method increases the original prediction accuracy by as much as 5% in the best case. This method also has the advantage that it can recommend both a specific therapeutic class and a specific drug to a prescribing physician, not just one or the other.

Table 4: Results for Drug-Therapeutic Class Method

Methods	**Top 1**	**Top 5**

Co-occurrence	**0.3617**	**0.4720**
KNN	0.3417	0.4501
Popular	**0.3577**	**0.4693**
N-Bayes	0.3312	0.4561
cGraph	0.2712	0.4021
D-Similarity	**0.3632**	**0.4751**

Discussion and Conclusions

In general, the goal of using collaborative filtering algorithms is to infer the underlying structure of the data sample available in the given domain. The medication reconciliation domain is a good field for this assumption since drugs are prescribed for particular diseases and some pairs of drugs often appear together while some others never do. Our experimental results suggest that simple CF approaches, such as D-Similarity that use only drug information can do a relatively good job at guessing missing drugs. On the other hand, simply introducing additional information about patients, such as demographics does not guarantee improved predictions.

One easy and effective multi-stage solution is to predict the therapeutic class in the first stage and use this prior information to refine drug prediction results. This technique improves the performance, and produces additional useful information for end users, without introducing too much complexity. Combining these predictions with their potential consequences, adverse or otherwise, can produce more insightful recommendations for physicians at the point of prescribing.

One major limitation of our current research is the implicit assumption about the accuracy of the training data. Given the concerns about quality of healthcare data in actual practice, the uncertainty about training data needs to be incorporated into the prediction methods. Another limitation is the homogeneous nature of the data. The patients' records from a long-term care center are relatively homogenous in terms of their drug regimens and their diagnoses. In order to increase diversity of demographic and medication information, we plan to evaluate current and new algorithms using data from multiple, disparate clinics and test the methods in real decision making settings.

Acknowledgements

The authors would like to thank C. Miller and M. Schaefer from Millennium Pharmacy Systems for providing the data used for this study. We are also grateful to S. Hasan for sharing valuable knowledge and insights on this problem.

References

[1] Hasan S, Duncan GT, Neill DB and Padman R. Towards a Collaborative Filtering Approach to Medication Reconciliation. AMIA Symposium Proceedings, pp. 288-292, 2008.

[2] Aspden P, Wolcott J, Bootman L, and Cronenwett LR. Preventing Medication Errors. National Academies Press, Washington DC, 2006.

[3] Glintborg B, Andersen SE, and Dalhoff K. Insufficient communication about medication use at the interface between hospital and primary care. BMJ (16:1), pp. 34-39, 2007a.

[4] Yang JC, Tomlinson G, and Naglie G. Medication lists for elderly patients. J. of Gen. Int. Med. (16:2), pp. 112-115, 2001.

[5] Goldenberg A, Kubica J, and Komarek P. A comparison of statistical and machine learning algorithms on the task of link completion. Proc. of the ACM SIGKDD Workshop on Link Analysis for Detecting Complex Behavior, 2003.

[6] Sarwar B, Karypis G, Konstan J, and Reidl J. Item-based collaborative filtering recommendation algorithms. In The 10th International World Wide Web Conference, pp. 285-295, 2001.

[7] Breese J, Heckerman D, and Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In The Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, pp. 43-52, 1998.

[8] Hastie T, Tsibshirani R, and Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2001: Springer.

[9] Kubica J, Moore A, Cohn D, and Schneider J. Finding underlying connections: A fast graph-based method for link analysis and collaboration queries. Proc. of 20th International Conference on Machine Learning, 2003.

Address for Correspondence

Huanian Zheng, The Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

Email: huanianz@andrew.cmu.edu