Article from

Predictive Analytics and Futurism

December 2016
Issue 14

Collaborative Filtering for Medical Conditions

By Shea Parkes and Ben Copeland

Recent trends in health care legislation have led to a rise in risk-bearing health care provider organizations, such as accountable care organizations (ACOs). Entrusted with the care of thousands of patients, these organizations must leverage data-driven approaches to population health management in order to improve quality of care and reduce costs.

One area of concern for data-driven analysis involves the accuracy of a patient's clinical documentation. Efforts to improve accuracy in a population's clinical records are often referred to as clinical documentation improvement or coding improvement. From a clinical standpoint, the benefit from coding improvement is obvious. A patient record that contains the entirety of the patient's illnesses will result in a more appropriate treatment plan.

However, there can be financial incentives in coding improvement. Alternative payment models often account for the health status of a patient population, through the use of risk scores, when reimbursing a health care provider for services. A more accurate clinical record ensures that risk-bearing health care providers are appropriately compensated when they care for sicker or healthier populations.

Coding improvement initiatives often start by looking through a given patient's records for explicit evidence of conditions that did not make it into the official diagnosis information: conditions coded on claims in prior years, or mentioned in the unstructured text of an electronic medical record. After these explicit sources of coding improvement are exhausted, more analytical methods can evaluate a patient for comorbidities to consider adding (or removing). One approach is to find explicit evidence of missed codings in large reference data sets and train predictive models that can be then be applied to other, potentially slimmer sources. This can work well for predicting specific chronic conditions in a population, even when only a short claims history is available.

Collaborative filtering can provide a different approach to identifying uncoded conditions by identifying common clinical patterns among patients in a population. Analysts can then make patient-level lists of conditions to review based upon comorbidities experienced by similar patients. The collaborative filtering approach works well at giving personalized lists of potential comorbid conditions from the patient perspective.

WHAT IS A COLLABORATIVE FILTERING SYSTEM?

If you have ever viewed a product on Amazon or watched a show on Netflix, then you have been a part of a collaborative filtering system, also known as a recommender system. Collaborative filtering systems are commonly applied to help users identify potentially interesting products among a large list of options, through the use of historical viewing or rating information. For example, Netflix will recommend certain shows to you based on your previous viewings. These recommendations are built using viewing or rating data from other users who have viewed the same shows as you.

Collaborative filtering often takes three forms: user-based, item-based, or matrix factorization. User-based collaborative filtering seeks to find users that have rated items similarly, and predict preferences for other items that similar users liked. Item-based collaborative filtering seeks to find similarities among items themselves, and then suggest items that are similar to a user's highly rated items. Matrix factorization estimates latent factors for each user and item and then uses these latent factors to find items that hopefully align with a user's preferences.

For an illustration of collaborative filtering in a clinical setting, consider the hypothetical patient panel in Figure 1.

FIGURE 1: EXAMPLE PATIENT PANEL

Condition	Patient 1	Patient 2	Patient 3	Patient 4
Diabetes	X	X		
Hypertension	X	X		
Coronary Artery Disease			X	
Hyperlipidemia		X	X	X
COPD			X	X

Patient 1 appears to be most similar to Patient 2. Thus, for Patient 1, hyperlipidemia might be considered as a potential comorbidity. Likewise, Patient 4 is most similar to Patient 3, so coronary artery disease might be considered as a potential comorbidity.

FIGURE 2: EXAMPLE PATIENT PANEL, CONDITIONS TO CONSIDER

Condition	Patient 1	Patient 2	Patient 3	Patient 4
Diabetes	X	X		
Hypertension	X	X		
Coronary Artery Disease			X	O
Hyperlipidemia	O	X	X	X
COPD			X	X

The preference inputs in collaborative filtering may take two forms: explicit ratings or implicit ratings. Explicit ratings are generated when the users themselves identify their preferences, such as giving a rating to a movie or a product. While explicit ratings carry a higher level of confidence for a user's preference, they are often not available. More commonly, implicit ratings are inferred from a user's actions, such as viewing a movie or buying a product.

The implementation explored in this article utilizes an implicit rating, matrix-factorization model to identify relative likelihood ratings for uncoded conditions. Each patient is a "user," with potential comorbid conditions being suggested as the "items." Implicit condition confidence values, or ratings, are inferred from the medical history of each patient in a population. These patient, condition, and confidence inputs are processed to generate latent factors for each patient and condition. These latent factors, an abstract representation of similarities among patients and conditions, can be combined to generate a rating for each patient-condition pairing.

The hypothetical example in Figure 3 illustrates using the estimated latent factors to generate condition ratings for a single patient.

FIGURE 3: CONDITION RATINGS BASED ON ESTIMATED LATENT FACTORS

Latent Factor	Patient	Diabetes	Hypertension	COPD	Menopause
1	**0.8**	0.2	0.3	0.1	-1.0
2	**0.4**	0.6	0.8	0.1	0.1
3	**-0.5**	0.1	-0.1	-0.1	0.1
4	**0.6**	-0.2	0.2	0.5	-0.1
Patient Rating	---	**0.23**	**0.73**	**0.47**	**-0.87**

A condition's rating for a given patient is calculated as the dot product of the patient's latent factors and the respective condition's latent factors (e.g., Diabetes Rating = 0.8x0.2 + 0.4x0.6 + -0.5x0.1 + 0.6x-0.2). Here, hypertension would be identified as the most likely potential comorbidity to consider. While latent factors are not easily interpretable, one could roughly associate each latent factor with a patient characteristic. Latent factor 1 could be gender-related because it has a strong coefficient for menopause. Latent factor 2 may be related to blood pressure, considering the high coefficients of both diabetes and hypertension, while latent factor 4 may be related to lung issues. Most real matrix factorization models use so many latent factors it would not be reasonable to try to actually attach interpretations to them.

A matrix factorization approach provides some useful benefits. The model is fast and simple to train, and thus can realistically be tuned to find unique relationships for each patient population. There are implementations available in cluster computing frameworks that gain additional speed by distributing the calculations (e.g., Apache Spark). Matrix factorization works well with the sparse nature of patient condition information (e.g., most patients only have a handful of conditions). Finally, the comorbid nature of many conditions can be naturally expressed via latent factors (e.g., a latent factor related to cardiovascular disease can usefully explain many conditions).

FEATURE ENGINEERING

There are two important considerations for generating useful input data: which features will be used, and how will confidence values for these features be determined. The features chosen here are a combination of historical condition information and demographic information. These features and their confidence values are generated from a patient population's clinical history.

For condition features, diagnoses in a patient's clinical history are grouped into clinically meaningful categories, or conditions, using the Clinical Classifications Software (CCS) of the Agency for Healthcare Research and Quality (AHRQ). Pa-

tients who are seen for the same condition multiple times are given a higher confidence value. More confidence is given for conditions that have been coded more recently. Additionally, more confidence is given for conditions that were coded in an inpatient setting rather than an outpatient setting.

The two main demographic features are age and gender. Unlike condition features, demographic features are given the same confidence level across all patients. The confidence value is determined such that demographic importance does not overpower condition information. However, these confidence values must also be large enough that gender-specific and age-specific conditions are modeled appropriately.

FITTING THE MODEL

The two most important hyper-parameters are lambda, the regularization parameter, and rank, the number of latent factors. Lambda should be strong enough to avoid overfitting in the training data, while also still allowing for meaningful personalization in predictions. Rank must be high enough to allow for meaningful groupings in latent factors, while avoiding the computational burden of higher rank models.

The goal is to identify the hyper-parameter values that are most useful for identifying uncoded comorbidities. To accomplish this, a tuning data set that excludes the most recent months of data is created. The hold-out data is analyzed to find conditions coded for the first time in a patient's medical history. A variety of models are trained on the tuning dataset with different hyper-parameter values. For each model, the hold-out data is used to calculate the percentage of newly coded conditions appearing in each patient's 10 highest-rated uncoded conditions. Using the best performing hyper-parameter values, a final model is trained with all of the available data to make up-to-date patient-level lists of the highest-rated conditions.

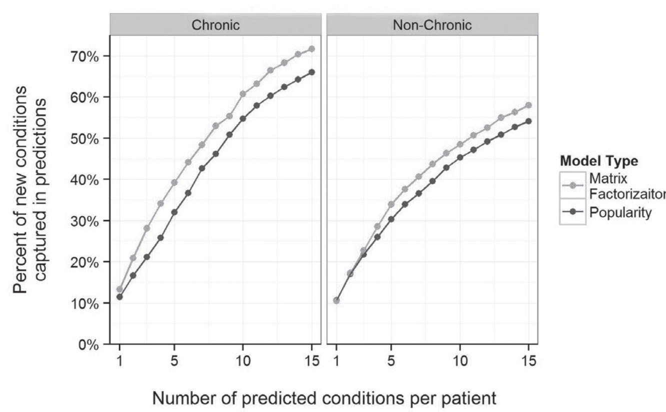
This whole tuning process is fast enough to calibrate hyper-parameters for each unique patient population.

MODEL PERFORMANCE

When using any advanced analytics, it is always important to have a useful baseline model to compare against. For a collaborative filtering model, the most basic reference model would be a simple **popularity** model that identifies the population's most common conditions, excluding conditions that have already been coded for a patient. For example, a popularity model would identify the most common condition, such as hypertension, as the highest-rated condition to consider for all patients that do not already have hypertension coded.

The illustration in Figure 4 compares model accuracy on a sample population for the collaborative filtering model (Matrix Factorization) versus the simpler Popularity model. The vertical axis shows the estimate of accuracy discussed above: the percentage of newly coded conditions from the hold-out set that were among the predicted conditions for each patient. The horizontal axis displays accuracy for different numbers of predicted condi

FIGURE 4: MODEL ACCURACY ON TWO-MONTH HOLD-OUT



The left side focuses on chronic conditions, which are more likely to go uncoded if they are not the primary reason that a patient seeks care. The right side focuses on non-chronic conditions. Because of the higher intensity level required in care, non-chronic conditions are more likely to be coded at the time the illnesses arise. For both the chronic and non-chronic conditions, the matrix factorization model consistently outperforms the popularity model.

CASE STUDY

This case study will examine model inputs and model results for a sample patient with diabetes. For this patient, the input features, the top 10 highest-rated conditions, and a breakdown of the contribution towards the highest-rated condition will be explored.

DIABETES PATIENT

The table in Figure 5 shows the input features and their respective confidence values. Demographic features are given a constant confidence value, whereas the confidence values for condition features are a factor of the patient medical history.

FIGURE 5: INPUT FEATURES AND CONFIDENCE VALUES

FEATURE	CONFIDENCE
Age- 45	
Gender- Male	
Subscriber Relationship- Policyholder	
Diabetes mellitus with complications- Chronic	
Essential hypertension	
Disorders of lipid metabolism	
Other nutritional; endocrine; and metabolic disorders- Chronic	
Diabetes mellitus without complication- Chronic	

The table in Figure 6 shows the top 10 highest-rated conditions and their relative ratings for this patient. The ratings are determined through a recombination of latent factors for the patient and the respective condition.

FIGURE 6: HIGHEST-RATED CONDITIONS

HIGHEST-RATED CONDITIONS	RATING
Thyroid disorders- Chronic	
Mood disorders- Chronic	
Anxiety disorders- Chronic	
Other upper respiratory disease- Chronic	
Esophageal disorders- Chronic	
Nutritional deficiencies- Chronic	
Other nervous system disorders- Chronic	
Osteoarthritis- Chronic	
Spondylosis; intervertebral disc disorders; other back problems- Chronic	
Asthma	

The table in Figure 7 breaks down the relative contribution for the highest-rated condition, thyroid disorders. A condition's rating can be decomposed into contributions from each of the input features, based on the feature's confidence value and latent factors.

FIGURE 7: CONTRIBUTING FEATURES, THYROID DISORDERS

THYROID DISORDERS - CONTRIBUTING FEATURES	CONTRIBUTION
Subscriber Relationship- Policyholder	
Essential hypertension	
Diabetes mellitus with complications- Chronic	
Age- 45	
Disorders of lipid metabolism	
Other nutritional; endocrine; and metabolic disorders- Chronic	
Diabetes mellitus without complication- Chronic	
Gender- Male	

The demographic features have a high contribution to the rating, which is partially due to the high confidence value associated with these features. Hypertension and diabetes are other strong contributing factors. Male gender appears to be slightly negatively associated with thyroid disorders.

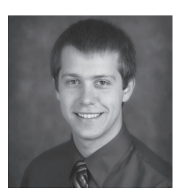
CONCLUSION

These lists of potential comorbid conditions can be used in a number of work-flows. Most importantly, these condition lists could be used to remind clinicians of common comorbidities to consider coding at the time of service.

In addition to identifying new conditions, the same model can be used to identify potential outliers in the conditions that have already been coded. Estimated ratings for existing conditions can be calculated, and those with extremely low values might represent codings that should be reconsidered to ensure there was not perhaps a mistake during data entry.

Accurately documenting a patient's clinical status will be increasingly important as more health care providers enter into alternative payment arrangements. Provider organizations face a growing scrutiny on the quality and cost of care. As a result, advanced analytics must find their way into daily workflows. Collaborative filtering systems provide a unique perspective toward coding improvements that produce useful suggestions of uncoded conditions. ■

Shea Parkes, FSA, MAAA, is an actuary at Milliman in Indianapolis. He can be reached at *shea.parkes@milliman.com.*



Ben Copeland is a data scientist and actuarial student at Milliman in Indianapolis. He can be reached at *ben.copeland@milliman.com*