Artificial Intelligence in Medicine 65 (2015) 89–96

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Predicting readmission risk with institution-specific prediction models

Shipeng Yu[a,*], Faisal Farooq[a], Alexander van Esbroeck[b], Glenn Fung[c], Vikram Anand[a], Balaji Krishnapuram[a]

[a] Siemens Healthcare, 51 Valley Stream Parkway (Mail Code E09), Malvern, PA 19355, USA
[b] Department of Computer Science and Engineering, University of Michigan, 500 S State Street, Ann Arbor, MI 48109, USA
[c] American Family Insurance, 302 N Walbridge Avenue, Madison, WI 53714, USA

ARTICLE INFO

Keywords:
Predictive modeling
Readmission risk prediction

ABSTRACT

Objective: The ability to predict patient readmission risk is extremely valuable for hospitals, especially under the Hospital Readmission Reduction Program of the Center for Medicare and Medicaid Services which went into effect starting October 1, 2012. There is a plethora of work in the literature that deals with developing readmission risk prediction models, but most of them do not have sufficient prediction accuracy to be deployed in a clinical setting, partly because different hospitals may have different characteristics in their patient populations.
Methods and materials: We propose a generic framework for institution-specific readmission risk prediction, which takes patient data from a single institution and produces a statistical risk prediction model optimized for that particular institution and, optionally, for a specific condition. This provides great flexibility in model building, and is also able to provide institution-specific insights in its readmitted patient population. We have experimented with classification methods such as support vector machines, and prognosis methods such as the Cox regression. We compared our methods with industry-standard methods such as the LACE model, and showed the proposed framework is not only more flexible but also more effective.
Results: We applied our framework to patient data from three hospitals, and obtained some initial results for heart failure (HF), acute myocardial infarction (AMI), pneumonia (PN) patients as well as patients with all conditions. On Hospital 2, the LACE model yielded AUC 0.57, 0.56, 0.53 and 0.55 for AMI, HF, PN and All Cause readmission prediction, respectively, while the proposed model yielded 0.66, 0.65, 0.63, 0.74 for the corresponding conditions, all significantly better than the LACE counterpart. The proposed models that leverage all features at discharge time is more accurate than the models that only leverage features at admission time (0.66 vs. 0.61 for AMI, 0.65 vs. 0.61 for HF, 0.63 vs. 0.56 for PN, 0.74 vs. 0.60 for All Cause). Furthermore, the proposed admission-time models already outperform the performance of LACE, which is a discharge-time model (0.61 vs. 0.57 for AMI, 0.61 vs. 0.56 for HF, 0.56 vs. 0.53 for PN, 0.60 vs. 0.55 for All Cause). Similar conclusions can be drawn from other hospitals as well. The same performance comparison also holds for precision and recall at top-decile predictions. Most of the performance improvements are statistically significant.
Conclusions: The institution-specific readmission risk prediction framework is more flexible and more effective than the one-size-fit-all models like the LACE, sometimes twice and three-time more effective. The admission-time models are able to give early warning signs compared to the discharge-time models, and may be able to help hospital staff intervene early while the patient is still in the hospital.

1. Introduction

It is well known that healthcare costs in the United States are imposing an increasing burden on the federal budget, yet the quality of care provided is arguably not adequate for many patient groups, such as individuals with multiple chronic conditions. Recently this has received serious attention and resulted in the healthcare reform legislation known as the Patient Protection and Affordable Care Act (PPACA), which was one of the most aggressive approaches to tackle this problem in the past several decades.

* Corresponding author. Tel.: +1 4843563424.

http://dx.doi.org/10.1016/j.artmed.2015.08.005
0933-3657/

Among the many cost drivers of healthcare in the U.S., *hospital readmissions* (or *rehospitalizations*) contribute to a significant proportion of total inpatient spending. Generally, a hospital readmission is defined as an admission to a hospital within a certain time frame (which can be 7, 15, 30, 60, 90 days or even as long as one year), following an original (*index*) admission and discharge. A readmission can occur at either the same hospital or a different hospital and can involve planned or unplanned surgical or medical treatments. In a popular study published in the *New England Journal of Medicine*, it was reported that 19.6% of the Medicare beneficiaries who had been discharged from a hospital were readmitted within 30 days, 34.0% within 90 days, and 56.1% within one year [1]. In another study performed by the Medicare Payment Advisory Commission (MedPAC), it was found that 17.6% of hospital admissions resulted in readmissions within 30 days of discharge, accounting for $15 billion in Medicare spending [2]. Also around 76% of readmissions were flagged as potentially avoidable.

As part of the PPACA legislation, the Center for Medicare and Medicaid Services (CMS) proposed a provision called the Hospital Readmission Reduction Program (HRRP), which is intended to reduce hospital readmissions. It is designed to penalize hospitals that have *excessive readmissions* at 30 days, i.e., higher-than-average 30-day readmission rate after risk adjustment. Initially CMS focused on three conditions, viz. heart failure (HF), acute myocardial infarction (AMI) and pneumonia (PN), and will add several other conditions starting 2015. Starting October 1, 2012, the penalty for each hospital was capped at 1% of the total Medicare reimbursement, and will gradually grow to 3% in 2015 and beyond. A large amount of money is at stake due to the HRRP program; more than 2000 hospitals were penalized in fiscal year 2013, with the total forfeited amount reaching $280 million [3].

There is a large body of research on reducing readmissions. Most of these efforts focus on improving discharge process and/or care transitions, e.g., making sure patients are well educated about their follow-up care and home medications, transitioning discharge information to the primary care doctors, carrying out home visits or follow-up phone calls to check patients' status. These interventions are resource-intensive and are not reimbursed the majority of the time. Thus, a critical step to make the intervention successful is to estimate the risk of patients being readmitted. This type of readmission risk assessment could be used to help target the delivery of the resource-intensive interventions to only the patients at greatest risk. Ideally, models designed for this purpose would provide clinically relevant stratification of readmission risk and yield information early enough during the hospitalization to trigger a transitional care intervention, many of which involve discharge planning and often commence well before hospital discharge. One popular approach is the LACE model [4], which is a simple yet effective readmission risk profiling tool at discharge time. A recent survey paper [5] performed a systematic review of risk prediction models for hospital readmissions. As noted, most approaches do not have sufficient prediction accuracy to be deployed in a clinical setting. Part of the reasons is that hospitals are known to have different characteristics in their patient populations, and the one-model-fits-all strategy may not work optimally. Note that not only does the disease (case) mix vary amongst hospitals, different hospitals also capture different patient characteristics. For example, ambulatory information may not be available in some hospitals in electronic format.

In this paper we discuss our recent work on a general framework for institution-specific readmission risk prediction. It extracts past patient data from the specific hospital or health system including, for instance, demographics, labs, medications, ICD and CPT codes, etc. It also identifies which patients were readmitted to the same hospital within a pre-defined number of days (typically 30 days) of discharge. Then it combines all the available information for each patient and builds a statistical model to predict readmission (to the same hospital). If a condition-specific risk prediction model is desired, the framework can adjust the model fitting only to the patients that have that condition. We showcase how a support vector machine (SVM) based classification approach and a Cox regression based prognosis approach can be applied in this context. For a new patient, the final model is able to predict a risk score indicating the likelihood of him/her being readmitted. We present the results of the proposed framework on three large hospitals in the U.S., and compare with LACE scores to show the effectiveness of the approach. We also develop models leveraging data at admission and discharge time, respectively, which allows us to see the additional benefit (if any) to predict the risk at discharge time.

The contribution of this paper is three-fold:

- To our knowledge this is the most comprehensive experimental study on institution-specific and condition-specific readmission risk predictions. We built different models at admission time and discharge time, leveraging different data that were available, and experimented on HF, AMI, PN as well as all-cause all-condition readmission risk prediction.
- We applied a classification method (SVM) and a prognosis analysis (Cox regression) to the data, and performed a systematic comparison with existing popular approach. While we outperform the competing methods, we also see the limitation of current readmission risk prediction approaches, which is due mostly to the unavailability of other relevant data (such as the socioeconomic variables).
- In addition to the classification performance such as the area under the ROC curve (AUC), we also measured ranking performance, using precision and recall at the top of the ranked patients based on the calculated readmission risk. The ranking related measures are arguably better at reflecting the real clinical value of the predictive model, since most institutions would only have resources to focus on the patients with the highest readmission risks.

We acknowledge that SVM and Cox regression are well-established methods, but we believe that using them and the methodical implementation to the particular readmission problem is new. In addition, the complete new holistic view of the problem and the need to move to institution-specific models is also major contribution of this paper.

The paper is organized as follows. Section 2 lists some related work on readmission risk prediction. Section 3 presents the modeling approach we used in this work. Section 4 shows detailed experimental setup and results, and Section 5 concludes the paper with discussions.

2. Related work

In the last ten years, there have been numerous studies that attempted to model the risk of readmissions, with accuracies (measured by AUC or c-statistic in validation sets) ranging approximately from 0.6 to 0.78. Most of the models focused in specific subpopulation. For instance, in [6] a model was created to identify heart failure patients at risk for 30-day readmission or death using an extensive dataset comprised of 12 different cohorts with 1100–14,500 patients each. For a comprehensive survey of models for prediction of readmissions for heart failure patients, please refer to [7]. Similarly in [8] a model was tailored to a specific subpopulation (patients older than 65 years). This model used Medicare inpatients 65 years or older from the general U.S. population, with a relatively small data set (1400 patients, 700 for training and 700 for validation), and the goal was to predict 30-day complicated