Expert Systems with Applications 42 (2015) 7110–7120

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Predictive modeling of hospital readmissions using metaheuristics and data mining

Bichen Zheng [a], Jinghe Zhang [a], Sang Won Yoon [a,*], Sarah S. Lam [a], Mohammad Khasawneh [a], Srikanth Poranki [b]

[a] Department of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY 13902, United States
[b] United Health Services Hospitals, Binghamton, NY 13903, United States

ARTICLE INFO

Article history:
Available online 8 May 2015

Keywords:
Neural networks
Support vector machine
Particle swarm optimization
Hospital readmission
Risk prediction

ABSTRACT

This research studies the risk prediction of hospital readmissions using metaheuristic and data mining approaches. This is a critical issue in the U.S. healthcare system because a large percentage of preventable hospital readmissions derive from a low quality of care during patients' stays in the hospital as well as poor arrangement of the discharge process. To reduce the number of hospital readmissions, the Centers for Medicare and Medicaid Services has launched a readmission penalty program in which hospitals receive reduced reimbursement for high readmission rates for Medicare beneficiaries. In the current practice, patient readmission risk is widely assessed by evaluating a LACE score including length of stay (L), acuity level of admission (A), comorbidity condition (C), and use of emergency rooms (E). However, the LACE threshold classifying high- and low-risk readmitted patients is set up by clinic practitioners based on specific circumstances and experiences. This research proposed various data mining approaches to identify the risk group of a particular patient, including neural network model, random forest (RF) algorithm, and the hybrid model of swarm intelligence heuristic and support vector machine (SVM). The proposed neural network algorithm, the RF and the SVM classifiers are used to model patients' characteristics, such as their ages, insurance payers, medication risks, etc. Experiments are conducted to compare the performance of the proposed models with previous research. Experimental results indicate that the proposed prediction SVM model with particle swarm parameter tuning outperforms other algorithms and achieves 78.4% on overall prediction accuracy, 97.3% on sensitivity. The high sensitivity shows its strength in correctly identifying readmitted patients. The outcome of this research will help reduce overall hospital readmission rates and allow hospitals to utilize their resources more efficiently to enhance interventions for high-risk patients.

1. Introduction

Healthcare has become one of the largest industries globally, and as such, it consumes a large amount of resources. In recent years hospital readmission has become a major topic of discussion in the U.S. healthcare system due to significant unnecessary costs associated with it. In 2004 about one-fifth of the Medicare beneficiaries were readmitted to hospitals within 30 days of discharge. It was estimated that the unplanned readmission of Medicare patients cost $17.4 billion (Jencks, Williams, & Coleman, 2009). Many of the preventable readmissions were related to low quality of care during patient stays in the hospital, as well as to poor arrangement of the discharge process (Malnick, Balla, & Schattner, 2008). Hospital readmission rate is thus recognized as a quality indicator of inpatient care for which effective, preventative interventions can be implemented (Hasan et al., 2010). The Centers for Medicare and Medicaid Services (CMS) has launched a readmission payment reduction program in which hospitals are financially penalized when Medicare patients are rehospitalized within 30 days of discharge (Centers for Medicare & Medicaid Services, 2012a). Thus, it is advantageous for hospitals to reduce their readmission rates by using effective and efficient interventions during patient stays and the discharge process. Currently, the finalized readmission penalty program focuses on acute myocardial infarction (AMI), heart failure (HF), and pneumonia (PN) since the readmissions from these diagnoses are more

* Corresponding author. Tel.: +1 607 777 5935; fax: +1 607 777 4094.
E-mail addresses: bzheng6@binghamton.edu (B. Zheng), jzhang51@binghamton.edu (J. Zhang), yoons@binghamton.edu (S.W. Yoon), sarahlam@binghamton.edu (S.S. Lam), mkhasawn@binghamton.edu (M. Khasawneh), Srikanth_Poranki@uhs.org (S. Poranki).

http://dx.doi.org/10.1016/j.eswa.2015.04.066
0957-4174/

common, expensive, and preventable (Centers for Medicare & Medicaid Services, 2012c; QualityNet, 2012). Various interventions are implemented to reduce readmission rates, including enhanced education for patients during the discharge process, medication reconciliation, follow-ups, etc. (Koehler et al., 2009).

Considering that healthcare resources (including physicians, nurses, and other medical resources) are very costly and limited, it is impractical and inappropriate for hospitals to provide equal efforts and interventions for all patients. Therefore, a prediction model that can be used to identify high-risk patients in advance could greatly benefit healthcare providers by enabling them to target resources on risky patients and, by extension, reduce the overall readmission rate (Centers for Medicare & Medicaid Services, 2012b). Once a particular patient is identified as high-risk, intensive interventions can be made to prevent a potential readmission. To corroborate this, one study found that tele-monitoring high-risk patients and corresponding private health plans enabled a 15% reduction in readmissions at a home healthcare facility (Minott, 2008).

However, the process of identifying patients who are very likely to be readmitted within 30 days of discharge is very difficult based on clinical expertise. This is due to the complex causes of readmission, such as a patient's health condition, quality of inpatient care and social determinants. Therefore, the objective of this research is to model the readmission patterns appropriately to predict the likelihood of readmission accurately. To describe the implicit patterns that lead to readmission and non-readmission, there are two clusters of approaches: analytical modeling and data mining. Since the readmission patterns, i.e. the relationship between predictors and dependent variables, are unknown, it is impractical to build an analytical model for accurate pattern description. However, historical data provides good evidence of those implicit patterns. Consequently, researchers proposed the concept and various algorithms of data mining and machine learning to capture hidden patterns from data.

Risk assessment models have been proposed to address the readmission problem for patients with various conditions such as general medicine patients and stroke and heart failure patients. In this research, the readmission rate of HF patients in a community hospital is studied. The majority of past research in hospital readmission used cohort study, logistic regression, and scoring systems to address the problem (Ross, Mulvey, & Stauffer, 2008). In general, existing risk-prediction models of hospital readmission perform poorly, according to the review research conducted by the Department of Veterans Affairs in 2011 (Kansagara, Englander, & Salanitro, 2011).

In this study, classification models that use neural networks, random forest (RF) and support vector machines (SVM) are proposed to predict the readmission risk of a particular HF patient. The remainder of this paper is structured as follows: Section 2 discusses the related literature in risk prediction modeling, especially those applied to assess patients' readmission risks. Proposed methodologies are described in Section 3; in Section 4, experiments are conducted to train and test those classification models, and the result analyses are discussed to compare the quality of those classifiers. Finally, the summary of this research and future work are addressed in Section 5.

2. Literature review

Risk-prediction models are broadly implemented in clinical and medical fields to support diagnostic decision-making. These include risk-prediction models for the risk assessment of breast cancer, type 2 diabetes, cardiovascular disease, and mortality for critically-ill hospitalized adults, as well as many others (Lindstrom & Tuomilehto, 2003; Siontis, Tzoulaki, Siontis, & Ioannidis, 2012). There are two types of risk-prediction models regarding breast cancer: identifying the risk that a patient will develop breast cancer over a certain time period and estimating the probability that a breast cancer-related gene mutation will occur in an individual (Claus, Risch, & Thompson, 1994; Parmigiani, Berry, & Aguilar, 1998). According to the risk-assessment results, a high-risk patient will be referred to intensive interventions and attention (e.g. screening and counseling) to prevent potential breast cancer. These models can help reduce the mortality rate among high-risk patients and can control the cost and complications for low-risk patients (Domchek et al., 2003). Noticeably, data-driven machine learning algorithms have been introduced into various medical decision-support domains, including cancer diagnosis (Mukti & Ahmed, 2013; Nahar, Imam, Tickle, Ali, & Chen, 2012; Zheng, Yoon, & Lam, 2014), cardiovascular abnormality detections (Sufi & Khalil, 2011), risk prediction (Siontis et al., 2012), etc. Data-oriented risk-prediction models have become effective tools that help medical decision-making and offer a number of benefits to both healthcare providers and patients.

As an important quality indicator of healthcare services, the high hospital readmission rate has attracted increasing attention and effort from the government, healthcare institutions, insurance payers and patients. Risk-prediction models can help prevent avoidable readmissions and eventually reduce the overall readmission rate. Various methodologies and techniques have been used to develop risk-prediction models for hospital readmission. A brief overview of the previous studies in readmission prediction is presented in Table 1.

Among the proposed methods, a cohort study and statistical models such as logistic regression and Cox proportional hazards regression, are the most common methods to identify risk factors. After those are used, weighted scoring systems are developed to measure the readmission risk of patients based on significant risk factors (Hasan et al., 2010; Whitlock et al., 2011). Cohort studies are commonly used in clinical areas in which groups are tracked from risk factor (exposure) to disease (outcome) in order to identify the correlation between them. As a longitudinal study, the exposure-disease association is determined with a higher quality and less bias. However, in a cohort study, it is expensive and difficult to achieve a high degree of similarity in the control group and to compensate for class imbalance in real cases (Grimes & Schulz, 2002; The Himmelfarb Health Sciences Library, 2011). In addition, logistic regression is a popular classification approach, especially when the outcome is binary. As one of the risk-prediction models, the LACE score has already been implemented in some hospitals. It is developed to predict unplanned readmission and mortality based on a prospective cohort study. This index considers four independent variables, including length of stay (L), acuity level of admission (A), comorbidity condition (C), and use of emergency rooms (E). A LACE score has been developed to evaluate and assess the patient readmission risk based on the LACE index assuming the linear relationship among the four variables. For instance, a LACE score can be obtained by summing up the values of those four variables (van Walraven et al., 2010). A threshold is set up to determine patient readmission risk based on clinics' specific circumstances to classify patients into different risk groups.

The risk-prediction models listed above employ different variables and target different diagnosis-related groups (DRGs). However, considering the ease of implementation and the difficulty of collecting medical and healthcare data, a model with fewer variables is more applicable. In general, those models are not capable of providing an accurate prediction of the readmission risk of a particular patient. Some perform poorly with an accuracy of less than 50%, and very few of them can predict correctly in over 70%