Predicting hospitalization likelihood using analytical modeling

Highlights

- Hospitalization results not only in additional costs but also increases the risk of hospital-acquired infections, cause disabilities to the patient and results in financial loss for the patient.
- A predictive analytics solution for predicting the likelihood of hospitalization can help reduce costs associated with hospitalization and also assist the hospitals to focus on the overall care management.

Preventing hospitalization is one of the key objectives of the healthcare systems. Hospitalization results not only in additional costs but also increases the risk of hospital-acquired infections, cause disabilities to the patient and results in financial loss for the patient. Safety studies show that additional hospitalization, litigation costs, infections acquired in hospitals, lost productivity and medical expenses cost few countries as much as US$19 billion annually.[1] Identifying the patterns in patient parameters leading to hospitalization can help in pre-empting hospitalization, devising patient care plans for handling patient inflow at hospitals and providing better patient care services.

The availability of electronic health records (EHR) facilitates the development of prediction models that identify patients at risk for hospitalization. This paper discusses the application of analytical modeling for predicting likelihood of hospitalization for a cohort of diabetes patients.

Background

Chronic diseases such as diabetes, congestive heart failure, hypertension and chronic obstructive pulmonary disease are among the most common, expensive and preventable of all health problems. According to Centers for Disease Control and Prevention (CDC), seven out of ten deaths among Americans each year are from chronic diseases.[2]

Advanced analytics modeling techniques can help in identifying patients at risk of hospitalization by analyzing the patient's clinical and laboratory test results, hospital facilities availed and prescription data. By identifying the patients at risk, hospitals can be better equipped to provide a better care for the patient. This may include availability of skilled physicians, design of a dedicated custom care plan to meet the patients and strategies (like altering medication regimen or conducting additional laboratory tests) to prevent or reduce the impact of the disease and improve the overall wellness.

Choice of diabetes for the study

Diabetes affects nearly 25 million Americans, and the number is expected to grow substantially every year. According to a report released by the Agency for Healthcare Research and Quality (AHRQ), diabetes is a disease that's costing Americans $83 billion a year in hospital fees — 23 percent of total hospital spending.[3] Diabetes is often present with other concomitant diseases such as high blood pressure and coronary heart diseases. Therefore, hospitalization likelihood of diabetes patients is an important case study as it directly impacts the overall wellness of the patient.

Predictive analytics in healthcare

Predictive analytics solutions are comprised of statistical techniques that are able to learn patterns present in historical data. The obtained knowledge can be subsequently applied to detect or predict trends in new data. The IBM® SPSS® Modeler platform is an extensive predictive analytics platform that is designed to bring predictive intelligence to decisions made by individuals, groups, systems and the enterprise.

The SPSS Modeler platform can be used to develop analytical models for healthcare providers that can be used in the diagnosis and treatment of various diseases. The data required for the analytical models can come from the EHR. For example, knowing in advance whether a group of patients are at a low or high risk for a disease or condition, healthcare providers can create targeted treatment measures for different cohorts. Healthcare providers can also devise different strategies to keep low risk patients at low risk, while mitigating the risk associated with high risk patients.

Hospitalization likelihood analytics— Modeling approach

Multiple approaches have been proposed to identify patients at risk for hospitalization. In one of the proposed solutions Outcomes Assessment Information Set (OASIS), plan of care, medications and medical record information from an urban home health agency were used to build and validate a predictive hospitalization model.[4]

This paper describes an implementation of risk of hospitalization for diabetes patients using the patient's clinical and laboratory test results, hospital facilities availed and prescription data. The model is designed to predict the likelihood of hospitalization for a patient in the near future (for example, the next 30 days). The 30-day period was chosen as an example to demonstrate the analytical capability of predicting the hospitalization likelihood of a patient.

The data preparation

We prepared a sample data set of approximately 2,000 patients consisting of claims information along the broad dimensions: prescription fills, current procedural terminology (CPT) codes, hospital facilities availed, international classification of diseases (ICD) codes, lab tests and demographics.

At the start of the evaluation period, when the patient visited the treatment facility and took the HbA1C test, the HbA1C test result value was termed as the "Baseline HbA1C Test result." To keep his/her HbA1C levels in control, the patient was prescribed medication (oral hypoglycemic) and HbA1C testing at periodic intervals. Data on HbA1C test results during a 6 to 12 month period was considered as the "Within Evaluation Period Levels." At the end of the evaluation period, the patient took another HbA1C test and the HbA1C test result was termed as "Evaluation HbA1C Test result." In other words, the patient's historical data was reconstructed such that his "Baseline HbA1C Test result" was termed as his starting state, from where he underwent medication regimen and laboratory testing procedures, leading him to the end state, the "Evaluation HbA1C Test result." This reconstructed data was used to build and train the predictive model.

The analysis was restricted to patients with at least 6 months of medication history. Having defined the evaluation period for each patient, we set the flag for "hospitalization" for a patient — that is, if a patient has availed "in-patient" hospital services such as ambulatory services or surgery, then we set the hospitalization flag to "TRUE" for the patient. After the flag was set for each patient, we divided the patient data into training and testing data sets.

Building the model

We built a classification model using the patient's clinical and laboratory test results, hospital facilities availed and prescription data. The model was built using the training data set and tested on the testing data set. For the test data set, we built the "misclassification matrix" to evaluate the accuracy of the model. The misclassification matrix we obtained was:

	Actual State		
	Percentage of cases	Not Hospitalized	Hospitalized
Predicted State	Not Hospitalized	85.9	14.1
	Hospitalized	19.3	80.69

In the model we built with our data set of close to 2,000 patients, we obtained an accuracy of 86 percent for non-hospitalization cases and an accuracy of close to 81 percent for hospitalization cases. In other words, 86 percent of the non-hospitalization cases and 81 percent of the hospitalization cases were classified accurately.

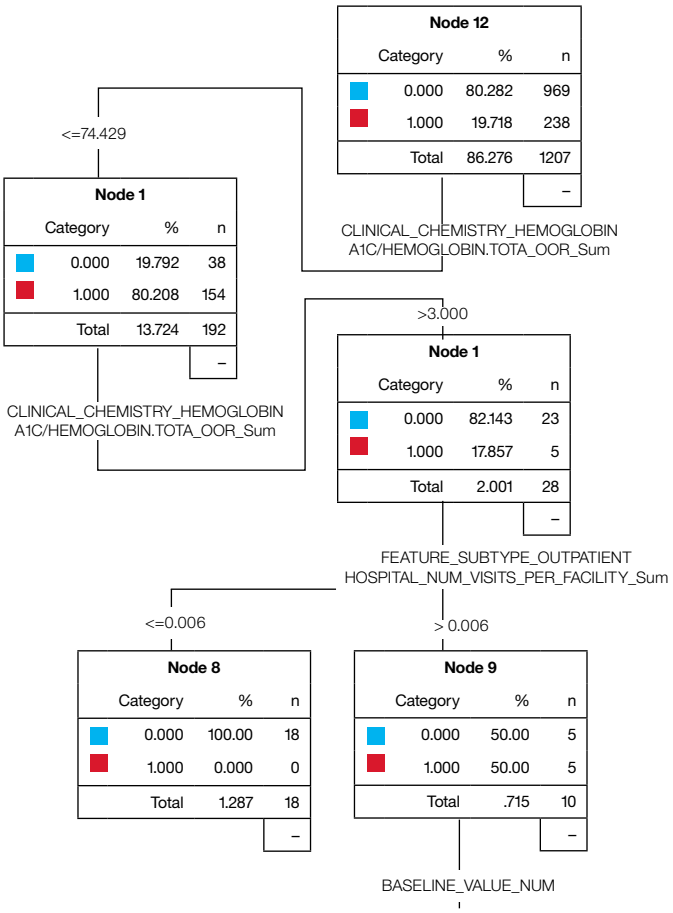
We also calculated a propensity score for patients who are predicted as "will be hospitalized." Propensity scores indicate the likelihood of hospitalization. Sample propensity scores are as shown:

	Hospitalization		
PID	Actual	Predicted	Propensity Score
4107137	1	1	0.418
6755819	1	1	0.89
8132228	1	1	1

For the patient with ID 4107137, we predicted him as "at risk of hospitalization." The probability of hospitalization is 41.8 percent. Similarly for patient with ID 6755819, the corresponding probability is 89 percent.

The classification model also gave the decision tree and the set of rules that can lead to the hospitalization of a particular patient.

A sample decision tree and rule set is as below:



IBM—Virtually unmatched commitment and proven solutions

IBM has deep expertise in managing and integrating complex systems. IBM specialists have subject matter experience in healthcare, life sciences, bioinformatics and social care. By bringing these together, IBM is enabling a foundation for transforming 21st century care systems.

Conclusion

A predictive analytics solution for predicting the likelihood of hospitalization can help reduce costs associated with hospitalization and also assist the hospitals to focus on the overall care management. The classification model we designed gave high accuracy rates and can be customized to patients suffering from ailments other than diabetes. The model also gives the "rules" (other ailments, out of range tests) that lead to hospitalization which gives a heads up for physicians to do a course correction in the treatment, prescribe additional tests etc to prevent a patient from getting hospitalized.

References

1. http://www.who.int/features/factfiles/patient_safety/patient_safety_facts/en/index6.html

2. "Maximize the Value of Population Management Programs With Advanced Analytics"— April 2014. http://www.beckershospitalreview.com/healthcare-information-technology/maximize-the-value-of-population-management-programs-with-advanced-analytics.html

3. "Diabetes: Hospital Bills Cost U.S. $83 Billion A Year, " August 2010. http://www.kaiserhealthnews.org/stories/2010/august/19/diabetes-cost-ft.aspx

4. Development and testing of an analytic model to identify home healthcare patients at risk for a hospitalization within the first 60 days of care. http://www.ncbi.nlm.nih.gov/pubmed/18032198

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
April 2015

GBW03324-USEN-00