A Predictive Model of Patient Readmission Using Combined ICD-9 Codes as Engineered Features

Robert P. Yerex, Ph.D.

UVA Medical Center Accountable Care Organization

To be presented at the 2015 Federal Committee on Statistical Methodology Research Conference (December 2015)

# Abstract

The timing of post-discharge care is a significant factor in reducing unplanned hospital readmission. Statistical learning techniques can be applied to the development of models that predict the likelihood of patient readmission during the critical 30 day post-discharge period. The accuracy of these models is dependent on the quantity and quality of data used for training and validation. For Medicare and Medicaid patients who are members of an Accountable Care Organization (ACO), the Centers for Medicare and Medicaid (CMS) provides detailed claims based data that can be used, if appropriately collated, and transformed. This involves the identification and creation of useful features, which when included in the model, increases its predictive strength. Creation of derived features (feature engineering) is a process in which a large number of base dimensions (n) are combined to create a smaller features set (n\* << n), reducing the complexity of the model, while retaining its inherent information value. The inpatient admission diagnoses, in the form of ICD9 codes, are an example of high dimensionality attributes found in the CMS claims data. Patterns inherent in the combinations of these codes can be used to create an engineered feature. In this study, a taxonomy of patterns of patient diagnoses was developed that was then used as a feature within a random survival forest model that predicts the hazard function (where the hazard event is unplanned readmission) of an individual patient for the first 30 days post discharge. Over the ensuing 30 days after release from hospital, a patient’s likelihood of readmission can be dynamically estimated based on the remaining portion of the hazard curve. Inclusion of the multiple diagnoses feature increased model accuracy to the point where it could be effectively used as a tool for targeting post-discharge patient care.