Semi-Supervised Functional Status Capture for At-Risk Readmit Patients

Andrew Placona, MS and Shantanu Phatakwala, MS
Evolent Health, INC.

1 Introduction

Thirty day all cause hospital re-admission rates continue to be a major public health problem in spite of recent improvements. According to a report by American Hospital Association, the all cause 30 day re-admission rate fell to 17.5% after remaining flat at 19% for years. Although care management programs have been proven effective in reducing readmission risk, there is still a question whether or not these programs identify the right patients. A meta-analysis discovered that out of 21 predictive models for hospital readmission risk only 6 had a c-statistic above 0.7 . Capturing functional data has often been a limitation. Functional data can be found in the Electronic Medical Record (EMR), but little research exists on its data quality. This paper examines using a semi-supervised approach to capture functional status based on the number of documents that describe negative functional status prior to a hospitalization.

2 Methods

The data for 1143 patients were obtained from a metropolitan health center. Negative functional status was defined as the sum of D documents within set S, which included any document that described negative functional status or the inability to perform Activities of Daily Living (ADL), a known readmission risk factor. A list of seed words were fed to Sentiword net to obtain synonyms to improve classification similar to sentiment classification when word does not appear in the corpus. Documents were tokenized via Python using NLTK, fed through a negation handler, and scanned for key words. A document was classified as a “negative” ADL document if it had a stemmed version of one of the seed words or its synonyms. The sum of total “negative” ADL documents that occurred before the index admission but within the one year time period was calculated for each patient.

Three evaluation techniques were use to evalute algorithm effectiveness: a confusion matrix to determine algorithm accuracy classifying “positive” or “negative” ADL phrases, bivariate analysis of the resulting variable to the outcome, and comparsion of two GDBT decision tree models to see the impact of the new variable. C-stats, plots of the predictions, and relative influence plots between the two models were compared to evaluate the impact.

3 Results

The algorithm had a ~1.5x lift compared to chance. In addition, the presence of negative ADL documents was associated with being 1.94x as likely to being readmited to this hospital. Lastly, the inclusion of the number of negative ADL documents improved the model c-stat by 0.03 on the validation sample.

4 Discussion

The immediate takeaway is that functional status can be approximated by counting the number of documents that negatively described functional status. the impact on predictive models is modest at best. The main finding may not be the incremental model performance, but rather add evidence that functional information is a strong indicator of readmission risk. Clearly the readmission problem is more complicated than adding one variable. Although functional status improves the model, it is part of a larger equation. What this analysis does show is that more granular and diverse data is needed to improve the existing readmission algorithms.

The main limiting factor was sample size. With a larger, more representative sample size, it could be possible to tell exactly what the modeling impact is in functional information. Another limiting factor was the intial seeding words. At this point, without a gold standard annotated sample, it is difficult to determine whether this algorithm captures all of the ADL statuses contained within the documents. However, this paper suggests that the previous limitation of getting patient level, real time pyscho-social and socio-economic information may be coming to end. With further research done to fully flush out these key aspects to readmission, it will be possible to determine whether or not readmission models can perform well or if readmissions can be predicted at all.