Dr. Sumanth Swaminathan
Revon Systems Inc
04/04/2016

Mathematical Prediction of Physician Triage of Asthma

Asthma is a lung condition that imposes a significant burden on patients' daily lives. Escalations of this condition (or exacerbations) are a frequent trigger of physician and hospital visits, which are both costly and distressing to patients. The need for novel solutions that limit the impact of exacerbations on global health is abundantly apparent.

One emerging approach to addressing asthma exacerbation is early detection by way of mobile app technology. Many of these apps, however, utilize rule based decision frameworks, which are constantly hampered by the size of the variable space involved in triage and diagnosis.

We are interested in developing a mathematical model that predicts the appropriate triage (urgency of illness) for an asthma patient based on patient health characteristics. In particular, we hope to train a machine learning type model on physician generated triage data and use that to make sample predictions. Some common supervised learning classification methods that have been employed on related problems include ensemble decision trees (Random Forests), Gaussian Naïve Bayes, and KNN. Some of our major goals and questions include:

1. What are the most important patient health features or combination of features for predicting an accurate patient triage?
2. Why do those particular features or combination of features matter the most?
    a. Can we understand the temporal component of these features (does the temporal change in these features matter or can we make predictions on features at a snapshot of time?)
3. Why does the particular machine learning method that is selected to solve our problem perform better than alternative possibilities (lower error, faster execution, ease of explanation, etc)?
4. What insights can be drawn from the physician triage data itself? Are there nontrivial trends in physician diagnosis that can be brought to light?

5. What data visualization techniques best represent the models, and how might you tune the visualization to convey different aspects of the features and functionality?
6. How do you represent the probability accuracy for the factors that affect the outcome and instill the appropriate level of confidence that the results are trustable and of high quality?
7. Can you suggest ways of using feedback from incorrect algorithm triage to feedback into the current predictor to improve future performance (retraining protocols, real time retraining, etc)?

NOTE: The preference is that all development work is done in Python 3. If these are not available, R and Matlab are acceptable as well (ordered by preference).