MPI 2017 Executive Summary for Revon Systems, Inc.
Predicting Exacerbation and Associated Triage in COPD Patients

This executive summary complements a presentation of results given at the 33rd Annual Mathematical Problems in Industry (MPI) Workshop in Newark, NJ on June 23, 2017. A full report will be delivered electronically to Revon Systems, in addition to being posted on the MPI 2017 website.

Group participants: Alireza Bagheri, Valeria Barra, Lance Bryant, Karolina Dubovskaya, Eli Goldwyn, Elvira Kadaub, Richard O. Moore, Daniel Pasut, Ryan Pellico, Lou Rossi, Vrushaly Shinglot, Axel Turnquist.

Presenters: Sumanth Swaminathan, Klajdi Qirko

Overview. Revon Systems has designed an expert system, delivered via a smart phone app, that provides recommendations to users on what action is appropriate given reported symptoms related to chronic obstructive pulmonary disease (COPD) . These symptoms are analyzed against a variety of background data including age, height, and weight, medical data including preexisting conditions such as asthma and diabetes, and vital sign data such as temperature, heart rate, and forced expiratory volume (FEV).

The Revon working group at MPI was tasked with a number of objectives, including improving the quality of triage recommendations from the expert algorithm, identifying trends and salient features in the doctor and patient data that might help with triage, and forming a predictive model for when an exacerbation of a patient's symptoms might occur.

Supervised Machine Learning. The Revon app currently uses a supervised machine learning approach where the algorithm is trained to reproduce the consensus triage opinion of several doctors on sets of data generated from a number of virtual patients. Some of the data are not currently used in the training phase, including a second triage opinion provided by each doctor along with confidences associated with each triage opinion. Also, the training data are incomplete in several fields, posing obvious difficulties for any machine learning method.

To incorporate the second triage opinions and associated confidences, the working group constructed proxy models using logistic regression and random forests of decision trees. These models were validated against the consensus first triage opinion on a separate validation data set after being trained on the primary data set using (i) the unweighted first triage opinions, (ii) the first triage opinions weighted by the doctors' confidence values, (iii) a weighted sum of the first and second triage opinions, and (iv) a bootstrapped set of resampled opinions with distribution determined by the doctors' confidences. These trials yielded the following observations:

- The logistic regression model performed slightly better in matching the validation set when trained using first triage opinions weighted by the confidence values compared with unweighted first triage opinions.
- Training the random forests model on a weighted sum of triage opinions produced substantially worse performance relative to training on the first triage opinions.

To assess the impact of missing data, the working group investigated (i) replacing the missing data with fixed values of 0 (corresponding to the introduction of a zero category in the nominal case), and (ii) replacing the missing data with the average value of the missing variable type. A neural network with three hidden layers of twenty neurons each was constructed using the augmented data sets, using either the first triage opinions or a confidence-weighted average of both triage opinions as the training targets. The following points are based on a numerical analysis of the neural networks trained on one set of data and validated against another:

- Insertion of the average value performed better than insertion of zeroes in all cases considered, in terms of both correlation coefficient and mean-square error.

- Use of just the first triage opinions provided better correlation and lower error in the training phase than use of the weighted average of both triage opinions.

- None of the neural networks provided acceptable skill in the validation phase despite showing promise during the training phase.

Unsupervised Machine Learning. Revon is interesting in complementing its doctor-informed supervised learning approach with unsupervised learning methods applied to the data in search of patterns that can be mined for diagnostic value.

The working group used the k-means clustering algorithm with $k = 1, \ldots, 4$ to generate clusters that they could then compare with the clusters labeled by the consensus triage opinion. This resulted in the following observations:

- All clusters produced an overlap with the triage consensus better than that provided by random chance.

- The most effective clustering was pairwise ($k = 2$), where repeated clusterings successfully segregated each of the four triage categories.

Predictive modeling. In addition to automating triage, Revon is interested in predictive models that anticipate likely exacerbations based on time series of the user's medical data. The working group formulated a model postulating vital sign data, specifically temperature, pulse rate, and oxygenation, as independent variables. The time evolution of these independent variables was modeled as a random walk and observations were incorporated into an estimate of their mean and covariance through the extended Kalman filter. Observations were assumed to be provided by direct noisy measurements of the variables and by registering symptoms such as severe cough, changes in sputum color and volume, and shortness of breath, whose dependence on the independent variables was modeled through logistic regression. The state estimate then provides a probability of future exacerbations conditioned on the time series of medical data and based on a particular definition of exacerbation, such as two consecutive days with at least two manifested symptoms. This work resulted in the following observations:

- A regression performed on severe cough was successfully predicted 12 of 18 incidents, but patient data for other symptoms was found to be insufficiently rich for a meaningful regression.

- Using only severe cough as the definition of an exacerbation, the extended Kalman filter trained on a portion of the time series was successful in predicting exacerbations in the remaining data with 37.5% accuracy.

Final Comments.

- The working group struggled throughout the week with lack of uniformity in the data (e.g., different formats for time stamps) and lack of clarity in label names. Adoption of a uniform standard would greatly assist analysis.

- The machine learning methods described above have different implementations for nominal and ordinal data. The Revon data set offers many opportunities to exploit these alternative treatments to improve the quality of its algorithm.