Review: An Integrated Data Mining Approach to Real-time Clinical Monitoring and Deterioration Warning

Layla Pournajaf
Department of Mathematics and Computer Science
Emory University
February 12, 2013

**Presentation Comment**  I can present this paper, but I am not sure if it is useful.

**Summary**  Yi Mao et. al. in the paper "An Integrated Data Mining Approach to Real-time Clinical Monitoring and Deterioration Warning" proposed a general health deterioration prediction system by integrating existing mining methods to study several features of two main clinical time-series heart-rate and oxygen saturation. The features included features of individual time series such as DFA, entropy, energy, and variance and combined features including correlation and cohesion. Then, they used a forward feature selection method to select the best set of features that contribute to the results. Finally, they applied several classification methods such as SVM and Logistic Regression to train data. Among these classifiers, their Logistic Regression model outperformed the others was approved for a clinical trial as the authors stated in the paper.

**Positive Points**
- The paper studies several features of time-series that reflects different characteristics of time-series. This study can be useful for other clinical and non-clinical mining processes dealing with time-series. Combined features of
time-series might be interesting for other studies such as event prediction in traffic studies.

- The paper takes advantage of several classification methods. It studies both non-linear classifiers and linear classifiers for the extracted features, showing non-linear classifiers outperform the others. This information along with how they dealt with unbalanced training set using under-sampling methods can be utilized again for other studies such as event mining in traffic studies.

**Negative Points**

- Few experiments are conducted which does not provide details on the effect of each feature, specificity the combined features. These details might be useful for more improvements.

- Domain knowledge can offer a high impact on choosing and extracting features for these types of study, but this was not used or mentioned in the paper.

**Questions**

- The paper does not mention time windows and how early they can predict deterioration. Answering to this question might be useful for further improvements of this approach.

- The paper suggests that their approach is different from previous works because it is not limited to study an specific disease. One can argue, training a system for an specific disease would benefit from a vast domain knowledge that might result in better accuracy. Therefore, Is it possible to feed this general system with such domain knowledge? For example, a set of possible diseases that might result in sudden death can be identified to extract more features from data.