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

Authors: Yi Mao, Wenlin Chen, Yixin Chen, Chenyang Lu, Marin Kollef, Thomas Bailey

Presentation: Layla Pournajaf
Department of Mathematics and Computer Science
Emory University
Feb 21, 2013
Motivation

- Every year, 4-17% of patients undergo cardiopulmonary or respiratory arrest while in hospitals.

- Lots of these patients could have been saved if warning of serious clinical events could be provided early, before its occurrence rather than when it is happening.

- Early prediction based on real-time electronic monitoring data has become an apparent need in many clinical areas.
Research Gap: General Prediction

- Most prior studies focus on some specific disease prediction.
  - McQuatt et al. introduced a decision tree method to analyze head injury.
  - Loforte et al. found statistical indexes for detecting sepsis by investigating the relationship between heart rate and respiration.
  - Khosla et al. applied multiple machine learning techniques for stroke prediction.
Purpose: General Mortality Prediction

- Build **classification methods** to monitor real-time signal of **heart rate** and **oxygen saturation rate** of patient.
- Issue **early warning** alerts before **clinical deterioration/death**.
- This system enables at-risk patients to be timely checked and treated by healthcare professionals in order to prevent potential deterioration and death.
Data

- In most hospitals, only intensive care units (ICUs) are equipped with real-time electronic medical sensors.

- The authors previously has proposed a real-time data sensing (RDS) system, which enables patients’ vital signs data in general hospital units to be collected via wireless sensors. (Still in small scale trial)
Overview of The Approach

- **Preprocess** the data set by removing abnormal values.
- **Extract features** from the patient’s collected real-time vital sign time series, including heart rate and oxygen saturation rate.
- **Apply feature selection techniques** to select the most relevant and discriminative features.
- **Apply classification algorithms** to train the model, and evaluate the prediction performance by cross validation.
Features

- Extracted from patients’ vital sign time series - heart rate and oxygen saturation rate.

- 34 features, including features within single time series and features linking the two time series.
Features: Detrended Fluctuation Analysis (DFA)

- (DFA) is a method for determining the statistical self-similarity of a signal.

- Mathematically, DFA is a scaling analysis method to reveal long-range power-law correlation exponents in noisy time series.

- It is most suitable for non-stationary time series with slowly varying trends, such as heart rate and oxygen saturation rate.
The DFA of a time series is calculated as the average fitting error over all segments in different scales.

Given a time series \( \{x(i)\}, 1 \leq i \leq N \), integration is performed to convert the original time series as follows:

\[
y(j) = \sum_{i=1}^{j} [x(i) - \langle x \rangle], 1 \leq j \leq N
\]

where

\[
\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x(i)
\]
Features : Detrended Fluctuation Analysis (DFA)

- Next, the integrated time series $y(i)$ ($i = 1, 2, ...N$) is divided into boxes of equal length $n$. A polynomial function $y(n)$ is fitted to each box of length $n$ by minimizing the least square error.

$$ F(n) = \sqrt{\frac{1}{N} \sum_{j=1}^{N} [y(j) - y_n(j)]^2} $$

- Typically, $F(n)$ increases with $n$ and follows the power law:
Features: Detrended Fluctuation Analysis (DFA)
Features : Approximate entropy (ApEn)

- It is a measurement designed to quantify the degree of regularity versus unpredictability.
- It quantifies the unpredictability of fluctuations in a time series.
Features: Spectral analysis

- Spectral analysis is another major method for analyzing clinical time-series data.
- For the calculation of the power spectra, the time series was resampled at 3.4 Hz using interpolation.
- The mean value, the standard deviation value was subtracted from the time series before applying the Fast Fourier Transformation (FFT).
Features: Spectral analysis
Features : First order features

Mean ($\mu$),
Standard deviation ($\sigma$),
Kkewness ($\gamma_1$),
kurtosis ($\gamma_2$)
Features: First order features

- Positive Skewness
Features:
First order features

- Negative Kurtosis
- Positive Kurtosis
Features : Second order features

- Co-occurrence features in one dimensional time series

- $Q = 10$ in this paper
Features: Second order features

Co-occurrence features: **energy** \((E)\), **entropy** \((S)\), **correlation** \((\text{COR})(\rho_{x, y})\), **inertia** \((F)\), and **local homogeneity** \((LH)\)

\[
E = \sum_{i=1}^{Q} \sum_{j=1}^{Q} c(i, j)^2
\]

\[
S = \sum_{i=1}^{Q} \sum_{j=1}^{Q} c(i, j) \cdot \log(c(i, j))
\]

\[
\rho_{x, y} = \frac{\sum_{i=1}^{Q} \sum_{j=1}^{Q} (i - \mu_x)(j - \mu_y)c(i, j)}{\sigma_x \cdot \sigma_y}
\]

\[
F = \sum_{i=1}^{Q} \sum_{j=1}^{Q} (i - j)^2 c(i, j)
\]

\[
LH = \sum_{i=1}^{Q} \sum_{j=1}^{Q} \frac{1}{1 + (i - j)^2} c(i, j)
\]
Features : Cross-sign features

- **Linear Correlation** : indicates the strength and direction of a linear relationship between two random variables.

\[
\gamma_{1,2} = \frac{E[(X_1(t) - E(X_1(t)))(X_2(t) - E(X_2(t)))]}{\text{Var}[X_1(t)] \cdot \text{Var}[X_2(t)]}
\]
Features : Cross-sign features

- **Coherence**: provides both amplitude and phase information about the frequencies held in common between the two time series.

\[
C_{1,2} = \frac{\phi_{x_1x_2}}{[\phi_{x_1x_1} \cdot \phi_{x_2x_2}]^{\frac{1}{2}}}
\]

where \(\phi_{x_1x_2}\) is the cross-spectral density, and \(\phi_{x_1x_1}\) and \(\phi_{x_2x_2}\) are autospectral densities.
Feature Selection

- Forward feature selection algorithm.
- Metrics: AUC and F-Score

\[ F(i) \equiv \frac{(\overline{x}_i^{(+)} - \overline{x}_i)^2 + (\overline{x}_i^{(-)} - \overline{x}_i)^2}{\frac{1}{n_{+1}} \sum_{k=1}^{n_+} (x_{k,i}^{(+)} - x_i^{(-)})^2 + \frac{1}{n_{-1}} \sum_{k=1}^{n_-} (x_{k,i}^{(-)} - x_i^{(-)})^2} \]
Classification Algorithms

- Support vector machine
  - The key idea is to learn an optimal hyperplane that can separate the training data set with maximum margin
  - Linear and Non-Linear

- Logistic regression
  - Logistic regression is a model for predicting the probability of an event, which can also be used for binary classification.
Data

- MIMIC II (Multiparameter Intelligent Monitoring in Intensive Care) database which contains comprehensive clinical data from tens of thousands of Intensive Care Unit (ICU) patients.

- Among 772 records, 175 are from visits belong the positive (death) set.
Skewed Data Problem

- **Undersampling**: The idea is to combine the minority class with only a subset of the majority class each time to generate a sampling set, and take the ensemble of multiple sampled models.

- **Exploratory undersampling**: The idea is to iteratively remove those samples that can be correctly classified by a large margin to the class boundary by the existing model.
Evaluation Criteria

- **AUC** (Area Under receive operating characteristic Curve),
- **PPV** (Positive Predictive Value) : the proportion of patients who actually suffer deterioration/death, among the candidate patients who are warned by the system
- **NPV** (Negative Predictive Value),
- **Sensitivity**, the proportion of patients who are warned correctly, among the patients who actually suffer deterioration/death
- **Specificity**
Results: The performance comparison of different features.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Features</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF Kernel SVM</td>
<td>DFA</td>
<td>0.6332</td>
</tr>
<tr>
<td></td>
<td>DFA+Cross Feature</td>
<td>0.6565</td>
</tr>
<tr>
<td></td>
<td>DFA+Cross Feature+ApEn</td>
<td>0.6753</td>
</tr>
<tr>
<td></td>
<td>All Features</td>
<td>0.7090</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>DFA</td>
<td>0.5370</td>
</tr>
<tr>
<td></td>
<td>DFA+Cross Feature</td>
<td>0.5731</td>
</tr>
<tr>
<td></td>
<td>DFA+Cross Feature+ApEn</td>
<td>0.5974</td>
</tr>
<tr>
<td></td>
<td>All Features</td>
<td>0.7402</td>
</tr>
</tbody>
</table>
Results: The performance comparison of different features.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Selected features</th>
<th>AUC</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression (AUC)</td>
<td>23</td>
<td>0.7844</td>
<td>0.9483</td>
<td>0.5208</td>
<td>0.7692</td>
<td>0.8567</td>
</tr>
<tr>
<td>Logistic Regression (F-score)</td>
<td>26</td>
<td>0.7592</td>
<td>0.9483</td>
<td>0.5104</td>
<td>0.7656</td>
<td>0.8540</td>
</tr>
<tr>
<td>SVM (AUC)</td>
<td>5</td>
<td>0.7752</td>
<td>0.9654</td>
<td>0.4852</td>
<td>0.8041</td>
<td>0.8651</td>
</tr>
<tr>
<td>SVM (F-score)</td>
<td>4</td>
<td>0.7736</td>
<td>0.9497</td>
<td>0.4833</td>
<td>0.7163</td>
<td>0.8652</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>AUC</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression + all features</td>
<td>0.7402</td>
<td>0.9483</td>
<td>0.3646</td>
<td>0.7000</td>
<td>0.8185</td>
</tr>
<tr>
<td>Logistic Regression + all features + exploratory undersampling</td>
<td>0.7767</td>
<td>0.9500</td>
<td>0.4615</td>
<td>0.9000</td>
<td>0.6440</td>
</tr>
<tr>
<td>Logistic Regression + exploratory undersampling + feature selection</td>
<td><strong>0.8082</strong></td>
<td>0.9473</td>
<td>0.4865</td>
<td>0.9000</td>
<td>0.6546</td>
</tr>
</tbody>
</table>
Results: The 10 highest-weighted variables of our final logistical regression model

<table>
<thead>
<tr>
<th>Feature</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>local homogeneity of heart rate</td>
<td>-14.50</td>
</tr>
<tr>
<td>standard deviation of oxygen saturation</td>
<td>10.20</td>
</tr>
<tr>
<td>entropy of oxygen saturation</td>
<td>10.17</td>
</tr>
<tr>
<td>low frequency of heart rate</td>
<td>8.62</td>
</tr>
<tr>
<td>local homogeneity of oxygen saturation</td>
<td>7.77</td>
</tr>
<tr>
<td>LF/HF of oxygen saturation</td>
<td>4.53</td>
</tr>
<tr>
<td>inertia of heart rate</td>
<td>3.86</td>
</tr>
<tr>
<td>entropy of heart rate</td>
<td>2.97</td>
</tr>
<tr>
<td>low frequency of oxygen saturation</td>
<td>-2.89</td>
</tr>
<tr>
<td>mean of oxygen saturation</td>
<td>-2.86</td>
</tr>
</tbody>
</table>
Discussion

- The paper does not mention time windows and how early they can predict deterioration!
- Why limited to Oxygen saturation and heart rate? What about other data such as blood pressure?
- 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.