Pulse classification

machine learning approach

Uwe Sterr
ST2C GmbH

2018-11-08
Outline

- Task description
- Introduction to machine learning (ML) process
  - Feature engineering
  - Algorithms
  - Performance measurement
- ML process applied to task
- ML results
Task description

- Detect pulses, avoid false detection

Puls data

Measured distortion data
Outline

- Task description
- Introduction to machine learning (ML) process
  - Feature engineering
  - Algorithms
  - Performance measurement
- ML process applied to task
- ML results
Machine learning process

EDA => exploratory data analysis
source http://www.feat.engineering/intro-intro.html#the-model-versus-the-modeling-process
Feature engineering

- Variables that go into the model are called:
  - Predictors
  - Features
  - Independent variables

- Quantity being model called:
  - Prediction
  - Outcome
  - Response
  - Dependent variable
Machine learning is function mapping

Build a model so it maps the unknown function $f(x)$

\[
\text{outcome} = f(\text{features}) = f(X_1, X_2, \ldots, X_p) = f(X)
\]

\[
\hat{Y} = \hat{f}(X)
\]

\[
\text{outcome} = f(\text{features}) = f(X_1, X_2, \ldots, X_p) = f(X)
\]

\[
\hat{Y} = \hat{f}(X)
\]
Outline

- Task description
- Introduction to machine learning (ML) process
  - Feature engineering
  - Algorithms
  - Performance measurement
- ML process applied to task
- ML results
Algorithms

- Logistic regression
- Tree based
  - Random forest
  - Gradient boosted trees
- Support Vector Machines (SVM)
- Neural Networks (NN)
  - Convolutional NN (CNN)
  - Recurrent NN (RNN)
Logistic regression

Regression limited to outcome between "0" and "1"

\[ \text{logistic}(\eta) = \frac{1}{1 + \exp^{-\eta}} \]

\[ P(Y = 1 \mid X_i = x_i) = \frac{1}{1 + \exp^{-(\beta_0 +}} \]

\[ \{1 + \exp^{-(\beta_0 +} \]
Tree based

- **Random forest**
  - Each tree is independent of previous trees
  - For each intersection only subset of features is used
    - Trees are uncorrelated

- **Gradient boosted trees**
  - Boosting ⇒ algorithm focuses on weak predictions
  - Each tree is based on all previous trees
Support Vector Machines (SVM)

- Maximal margin classifier
- Using separating hyper-plane
  - Hyper-plane is a $p-1$ dimensional subspace in $p$-dimensional space
  - In two dimensional space $\Rightarrow$ hyper-plane is line
  - Hyper-plane divides features space in two regions $\Rightarrow$ binary classification
SVM as optimization problem

Constrained optimization problem

\[
\begin{align*}
\text{maximize} & \quad M \\
\text{subject to} & \quad \sum_{j=1}^{p} \beta_j^2 = 1, \\
& \quad y_i (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}) \geq M \\
& \quad \text{for all } i = 1, \ldots, N.
\end{align*}
\]
Outline

- Task description
- Introduction to machine learning (ML) process
  - Feature engineering
  - Algorithms
  - Performance measurement
- ML process applied to task
- ML results
Performance measurement

Confusion matrix

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>Actual class</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cat</td>
<td>Cat</td>
<td>True Positives (TP)</td>
</tr>
<tr>
<td>Cat</td>
<td>Non-cat</td>
<td>False Positives (FP)</td>
</tr>
<tr>
<td>Non-cat</td>
<td>False Negatives (FN)</td>
<td>True Negatives (TN)</td>
</tr>
</tbody>
</table>

Based on the four elements of the confusion matrix various metrics are defined, for details check https://en.wikipedia.org/wiki/Confusion_matrix
Metrics based on confusion matrix elements

- Sensitivity \( \Rightarrow \) P(event predicted | event observed)

\[
\text{Sensitivity} = \frac{\text{Sample with the event and predicted to have the event}}{\text{Samples having the event}} = \frac{TP}{TP+FN}
\]

- Specificity \( \Rightarrow \) P(no-event predicted | no-event observed)

\[
\text{Specificity} = \frac{\text{Sample without the event and predicted as nonevents}}{\text{Samples without the event}} = \frac{TN}{TN+FP}
\]
Receiver operating characteristic (ROC)

- Visual representation of confusion matrix

- Includes for various thresholds
  - Sensitivity
  - Specificity

- AUC \(\Rightarrow\) area under curve
  - The higher the better
  - \(0 < \text{AUC} < 1\)
Confusion matrix and ROC

**Densitiy plot of predicted probability**

- **Ground truth = Negative**
- **Ground truth = Positive**

**Receiver Operating Characteristic (ROC)**

- **TN**
- **FP**
- **FN**
- **TP**

**Confusion Matrix**

- **TP Rate: 0.92**
- **FP Rate: 0.38**
- **FN Rate: 0.08**
- **TN Rate: 0.62**

Threshold: 0.43
Confusion matrix and ROC for hit

Ground truth = No Puls

Ground truth = Puls

Density plot of predicted probability

Receiver Operating Characteristic (ROC)

Confusion Matrix

Puls Rate: 0.92
False Puls Rate: 0.38
Missed Puls Rate: 0.08
No Puls Rate: 0.62

threshold: 0.43
R Plots

Density plot of predicted probability

Receiver Operating Characteristic (ROC)

Confusion Matrix

threshold: 0
Outline

- Task description
- Introduction to machine learning (ML) process
  - Feature engineering
  - Algorithms
  - Performance measurement
- ML process applied to task
- ML results
Create augmented labeled data

How to label data as being positive?
Create augmented labeled data

How to label data as being positive?

- Create augmented hits
- Vary parameters
Features of time signals
Features generated

- Sample values of window
- Dynamic time warp (window)
- Min(window)
- Max(window)
- Median(window)
- Variance(window)
Features generated

[Graph showing various features like CasMaxExp_1, CasMaxExp_2, etc.]
Dynamic time warp example

Shows which samples of query and template (red line) are compared.
Threeway plot

Shows in the center graph which indexes are connected
Dynamic time warp (DTW) for signal

Color and size correspond to DTW value, Time step: 1

Window
Algorithm

- Start with simplest algorithm
- Use simple algorithm for feature engineering
- Use more complex algorithm if result is unsatisfactory
Outline

- Task description
- Introduction to machine learning (ML) process
  - Feature engineering
  - Algorithms
  - Performance measurement
- ML process applied to task
- ML results
## ML results log reg for measured data

### Validation data

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>Actual class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Puls</td>
<td>33847</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>No-Puls</td>
<td>0</td>
<td>16924</td>
</tr>
</tbody>
</table>

### Test data

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>Actual class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Puls</td>
<td>8461</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>No-Puls</td>
<td>0</td>
<td>4230</td>
</tr>
</tbody>
</table>

### Predicted class

<table>
<thead>
<tr>
<th>Actual class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Puls</td>
<td>True Puls</td>
<td>False Puls</td>
</tr>
<tr>
<td>No-Puls</td>
<td>Missed Puls</td>
<td>No Puls</td>
</tr>
</tbody>
</table>
### Validation data

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>Actual class</th>
<th>Puls</th>
<th>No-Puls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puls</td>
<td>33847</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>No-Puls</td>
<td>0</td>
<td>16924</td>
<td></td>
</tr>
</tbody>
</table>

### Test data

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>Actual class</th>
<th>Puls</th>
<th>No-Puls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puls</td>
<td>8461</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>No-Puls</td>
<td>0</td>
<td>4230</td>
<td></td>
</tr>
</tbody>
</table>
ML results for measured data ROC

- ROC of logistic regression
  - Perfect separation of two classes
  - No need for more complex algorithm
# ML results for SNR = 18dB

## Logistic regression

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>Actual class</th>
<th>Pulse</th>
<th>No-Pulse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulse</td>
<td>4077</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>No-Pulse</td>
<td>190</td>
<td>185459</td>
<td></td>
</tr>
</tbody>
</table>

## Gradient boosted trees

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>Actual class</th>
<th>Hit</th>
<th>No-Hit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>4266</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>No-Hit</td>
<td>1</td>
<td>185634</td>
<td></td>
</tr>
</tbody>
</table>

## Support vector machine

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>Actual class</th>
<th>Hit</th>
<th>No-Hit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>4267</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>No-Hit</td>
<td>0</td>
<td>185634</td>
<td></td>
</tr>
</tbody>
</table>
ML results for SNR = 18dB ROC

- ROC of logistic regression
  - Not perfect separation of two classes
  - Need more complex algorithm => Gradient boosted trees
ML results for SNR = 18dB, compare models

- ROC, Sensitivity and Specificity for GBM and LogReg vs cross validation
## Optimise ML hyper parameter

<table>
<thead>
<tr>
<th>Simulated Annealing</th>
<th>• Optimize ML hyper parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML algorithm</td>
<td>• Run training</td>
</tr>
<tr>
<td>Evaluate metric</td>
<td>• Using test data</td>
</tr>
</tbody>
</table>
Want to meet ML people from academia and industrie

- Machine Learning User Group Stuttgart MLUGS
  - Technically oriented
  - 419 members

- Autonomen RoboCar bauen
  - hands on, build a real world system
  - 153 members
  - https://www.meetup.com/Esslingen-Makerspace/

- Stuttgart AI
  - More concepts and industrie presenting themself
  - 979 members
  - https://www.meetup.com/StuttgartAI/