Report of Coding Test

Xu Yumeng
Faculty of Science
National University of Singapore
16-Oct-2019

1.Workflow

Read and segment training data and testing data

↓

Shuffle and split training data: 70% training data, 30% validation data

↓

Extract Features: simple features and wavelets features

↓

Choose models using cross-validation (10-fold) : Logistic regression and SVM

↓

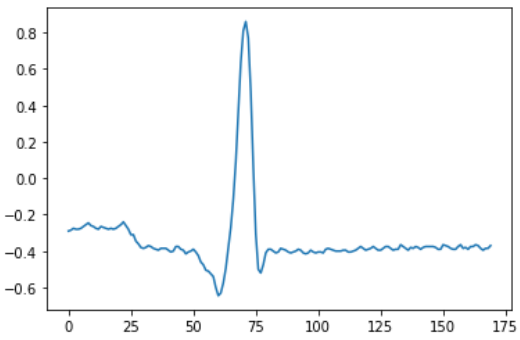
Build a machine learning model (SVM, Gaussian kernel)

↓

Export results to WFDB format (*.test)

2. Methods

2.1 Segmentation based on peaks



According to the heart rates of the training data and testing data, one beat was segmented from 70 data points in front of its peak to 100 data points after its peak.

2.2 Extraction features from each segment

a. Simple features: Each segment was split into 3 parts. Min, max, mean and median values were calculated for each part. In total, 12 simple features were generated for each segment.

b. Wavelets features: Wavelets transform is a common method for ECG signals filtering and feature extraction. ECG signals were processed using Haar wavelets transform and coefficients of level-1 Haar wavelets were used as wavelets features.

2.3 Choosing a machine learning algorithm

Logistic regression and SVM both are common and efficient algorithms for binary classification. 10-fold cross-validation was conducted for chosen a more appropriate algorithm from them.

2.4 Training, testing and evaluation of the model

Training dataset was split into training data and validation data in a ratio of 70%:30%. Since the number of “V” beats is much smaller than the number of other beats (around 1:10 in average), AUC value was adopted as a metric for evaluating the performance of model besides accuracy.

3.Results

3.1 10-Fold Cross-validation for Logistic Regression and SVM

Machine learning algorithms	Cross-validation score
Logistic Regression	0.9974077858193032
SVM	0.9993695025312157

3.2 Classification report for SVM on validation dataset

	Precision	Recall	F1 score	support
Others	1.00	1.00	1.00	3951
“V” beats	0.99	0.99	0.99	332
Accuracy			1.00	4283
Macro average	0.99	0.99	0.99	4283
Weighted average	1.00	1.00	1.00	4283

AUC value = 0.993469702652676