ECG Heartbeat Classification

Nik Dennler (nik.dennler@uzh.ch)

March 19, 2020

1 Introduction

This project aims to differentiate between different heartbeat patterns and to detect and classify anomalies. Two different ECG data sets are investigated. After exploring the data via scatter plots and clustering approaches, different recurrent neural network architectures are tested and compared with a 1D-CNN baseline[1]. Finally, we investigate a transfer learning approach, where we train a network on the larger data set, then fine-tune and test for the smaller data set.

2 Data Exploration

The first data set is the MIT-BIH Arrhythmia Database[2], consisting of 48 half-hour ambulatory ECG recordings of healthy patients and of patients with different arrhythmic disorders. The recordings were clipped to single heartbeats and those manually categorized into one of five classes, resulting in 109446 labeled samples. The second data set is a subset of the PTB Diagnostic ECG Database[3], which consists of ECG recordings from 200 patients, from which 148 are diagnosed as Myocardial Infarction (MI) and 52 healthy control. Also here, the recordings were clipped to single heartbeats and binary classified, resulting in 14552 labeled samples. In all experiments, the authors have used ECG lead II re-sampled to 125 Hz, where subsequently the samples where cropped, down-sampled, normalized and zero-padded to the fixed dimension of 188. For the training, evaluation and testing of the network, the data was split into train and validation sets. Figure 1 shows the heartbeat patterns of the two data sets for the different classes. Figure 2 shows the data after the dimensionality is reduced with Principal Component Analysis (PCA) projected to two dimensions. For clustering, k-Means and Gaussian Mixture Model approaches have been applied on the 99% most relevant principal components, where both were unsuccessful in separating the classes accurately.

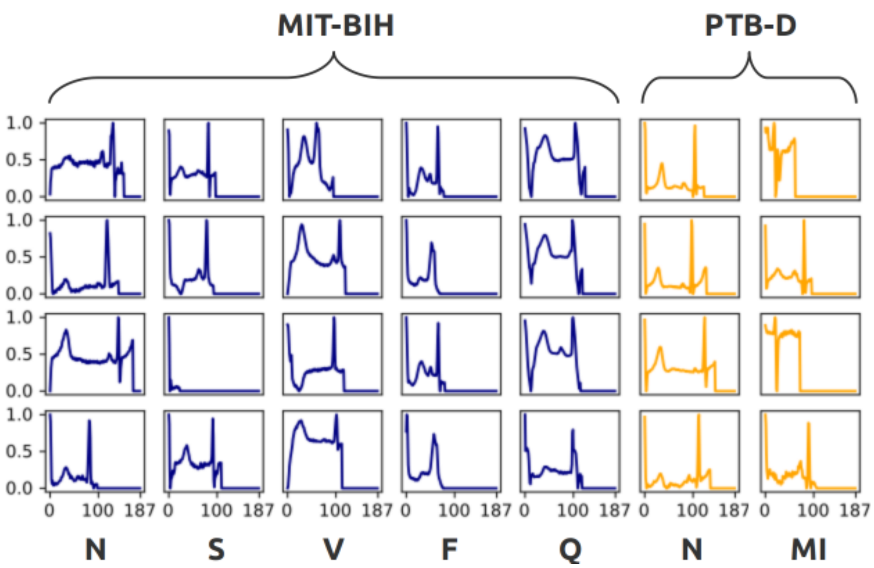

Figure 1: Single heartbeat ECG recording samples of different data set and different classes. Blue (left) are samples of the MIT-BIH data set, orange (right) are samples of the PTB-D data set. The different letters stand for different anomaly classes.

[1] Kachuee et al., "ECG Heartbeat Classification: A Deep Transferable Representation", 2018

[2] Moody et al., "The impact of the MIT-BIH Arrhythmia Database", 2001

[3] Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals", 2003

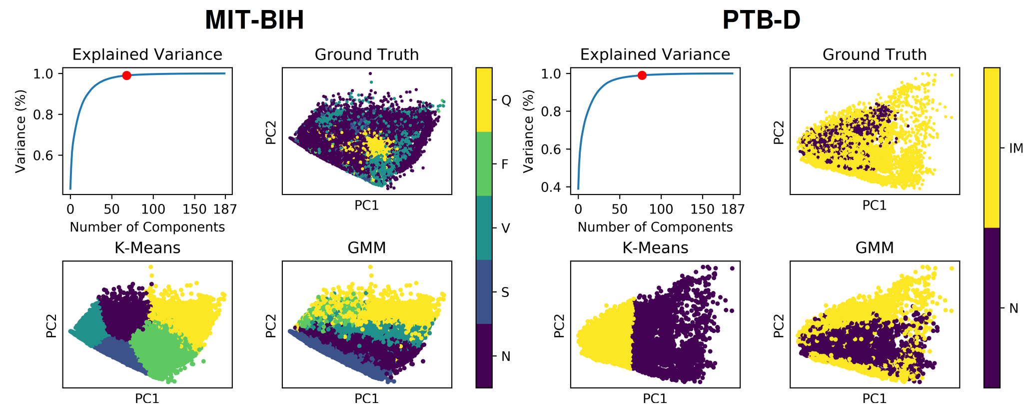

Figure 2: PCA dimensionality reduction plots. Left are samples of the MIT-BIH data set, right the ones of the PTB-D data set. Top left: Nr. of PCs vs explained variance. Red dot signalizes 99% explained variance, which determines the number of PCs used for the dimensionality reduction. Top right and bottom plots: PCA dimensionality reduction, projected onto two dimensions. Colors correspond to classes, determined by the ground-truth, k-means clustering and Gaussian-mixture model classification.

MIT-BIH		**PTB-D**			
Model	Accuracy	Model	Accuracy	AUROC	AUPCR
1D CNN (baseline)	0.9853	1D CNN (baseline)	0.9284	0.9057	0.9351
RNN (LSTM)	0.8384	RNN (LSTM)	0.7472	0.6386	0.7831
RNN (GRU)	0.9860	RNN (GRU)	0.7221	0.5	0.7221
Bidirectional RNN (LSTM)	0.9868	Bidirectional RNN (LSTM)	0.9605	0.9495	0.9647
Bidirectional RNN (GRU)	**0.9879**	Bidirectional RNN (GRU)	**0.9777**	**0.9728**	**0.9810**
Transfer Learning	–	Transfer Learning	0.9330	0.9232	0.9482

Table 1: Results for the different model. All shown metrics where evaluated on a test set that has not been seen during training or validation. Bold denotes the best results.

3 Experiments and Results

3.1 Individual Learning

Different recurrent neural network (RNN) architectures where investigated and tested against a 1D-CNN baseline network. After a quick manual hyperparameter search on the training set of the MIT-BIH dataset, we trained and tested two recurrent networks and two bidirectional recurrent networks, using LSTM and GRU gates. Each network consists of two recurrent layers with 64 and 128 neurons and dropout rate of $r = 0.2$ respectively, then three fully connected layers with 64, 16 and 8 ReLu neurons respectively and finally one fully connected layer with the neuron count equal to the number of outputs or equal to one in the case of binary classification. The Adam optimizer with an initial learning rate of 0.001 was chosen. For the MIT-BIH data set, the loss function is sparse categorical cross-entropy loss, while for the PTB-D dataset it is binary cross-entropy loss. The results of the metrics evaluated on the test set can be seen in Table 1.

3.2 Transfer Learning

We also investigated in transfer learning, i.e. in using a pre-trained model for a new task, aiming to perform well without the requirement of many labels or hours of training time. The best performing model of the MIT-BIH data set was imported (here the Bidirectional RNN with GRU gates), the last three layers clipped and four layers added: three fully connected layers with 64, 16 and 8 ReLu neurons respectively and the ultimate layer with one neuron and sigmoidal activation. All old layers where frozen and only the new ones trained. The results of the metrics evaluated on the test set can be seen in Table 1.

4 Conclusion

Our bidirectional RNN model with GRU gates performs best and significantly better than the baseline model for both data sets. This could be explained by the consideration of the network of past and future data events, but also by the resulting larger number of parameters. The transfer learning approach did beat the baseline but did not perform as well as the best tested model. Considering that the training time was much shorter, this is still a positive outcome.