Detection Of Myocardial Infarction Using ICA

Kužílek J[1], Spilka J[1], Lhotská L[1],Hanuliak M[2]
[1]Department of Cybernetics, FEE CVUT, [2]Medical Technologies CZ a. s.
kuzilj1@fel.cvut.cz

Abstract. Independent Component Analysis (ICA) is a powerful tool for processing of multi-dimensional signals and it was successfully deployed in many different applications. Mostly in EEG signal preprocessing - it was used for artefacts removal (mostly blink artefacts). Our work aimed at ECG infarction data. We tried to find differences between components of normal and infarction data, based on features calculated from these components. Feature set was then passed to k-means clustering algorithm and resulting clusters were observed and analysed.

1 Introduction

Myocardial infarction is one of the most prevalent heart diseases in the Czech Republic. It is mostly due to narrowed coronary arteries causing insufficient oxygen supply of myocardium. State-of-the-art methods developed to diagnose myocardial infarction achieve good results on experimental/testing databases but in the real-world applications they perform worse. Most of these methods are based on decision rule classifiers using features calculated on averaged beat of 10 s long electrocardiographic (ECG) record [5].
Our research aim is to use ICA to find differences between waveform of components in infarction and non-infarction data. ICA was successfully applied for EEG analysis as well as for ECG analysis, such as ischemia [8] or atrial fibrillation [6] detection.

2 Independent component analysis and usage on bio-signals

ICA is described in many publications [1, 7, 2, 4, 3]. It is a method, which searches source signals from knowledge of their mixture. In case of our problem we can assume that heart damaged by myocardial infarction makes different environment for signals propagation than healthy heart and waveforms in component are different too.
ICA represents one solution of Blind Source Separation(BSS) problem. There are several definitions of ICA, but all of them assume linear combination of source signals (sometimes called components):

$$\mathbf{X} = \mathbf{AS}, \tag{1}$$

where $\mathbf{X}$ is a matrix of mixture of source signals, $\mathbf{A}$ is mixing matrix, which characterizes environment through which source signals pass, and $\mathbf{S}$ is matrix of source signals. $\mathbf{X}$ and $\mathbf{S}$ got size n x m, where n is number of source signals and m is length of record in samples. Mixture matrix $\mathbf{A}$ is of size n x n, where n is number of source signals. Generally we assume, that number of components and measured signals does not need to be the same. Figure 1 show schematic representation of mixing process.
Components can be obtained using the following expression:

$$\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} = \mathbf{WX}, \tag{2}$$

where matrix $\mathbf{W}$ is inverse matrix to matrix $\mathbf{A}$. We can see, that component search is reduced to search of matrix $\mathbf{W}$.
The separation is trying to create components, that would be independent and their linear combination should be the original signal $\mathbf{X}$. This is done by an iterative algorithm, which maximizes function of independence F. Function F measures independence of each component on other components.

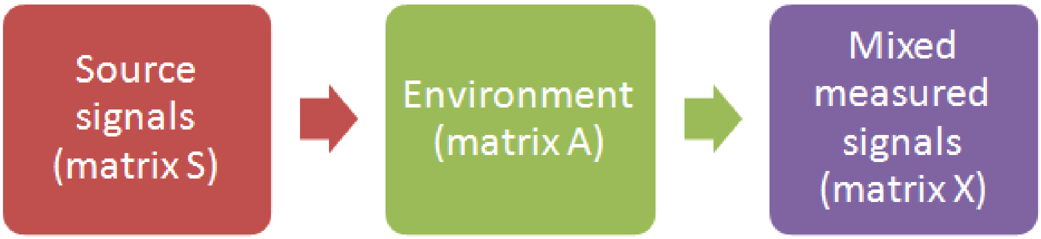

Fig. 1: Schematic representation of a signal mixing process.

ICA has some disadvantages - it can not guarantee order of components and their energy. Second disadvantage can be partially avoided by data preprocessing.

As we said before, ICA was successfully deployed on biomedical data, in most cases EEG signals for removal of blink artefacts, which were created by eye opening and closing during recording of signals. These signals are independent on normal brain activity and they are separated in one component. So removing them is easy, corresponding component is replaced by zeros.

Another application field of ICA are ECG signals. ICA was deployed for detection of atrial fibrillation detection and ischaemia detection. In the first case fibrillation is an independent activity and it is mostly separated to one component. In the second case it separated ST segment from data to one component and this component is then used for detection of ischaemia. Our research idea came from this second application - when ST segment was separated, then we could expect separation of T-wave and we could see differences between normal and infarction data.

3 Method

The first step of analysis was data pre-processing. We used band pass 2nd order Butterworth filter in order to remove noise. Without pre-processing ICA separates noise to one or more components and we loose information about basic components of ECG signal. They are mixed together in remaining components. So pre-processing made beat waveform more separable. Filtration removed izoline drift(low frequencies) and fast noise such as muscle or electric grid brum. After filtration ICA was applied to data and thus we acquired independent components. The next step of analysis was to compute the component parameters. The set of features we used was as follows:

- mean of the component
- zero crossings of the component
- energy in range 1-8 Hz of the component
- 1st and 2nd spectral momentum of the component
- correlation between component and triangle-shaped signal

These features were selected because of their good representation of QRS complex and T wave. And they made clustering possible.

These features were used to cluster components into groups with the same properties. We clustered infarction and non-infarction data separately. For each group (infarction and non-infarction) we applied k-means clustering algorithm and clustered data into 4 clusters. We tried different number of clusters but lower number of clusters did not provide any information and, on the other hand, more clusters only splitted data from one cluster to more – only creating sub-clusters of these four. The last step of analysis was comparison of infarction groups that were created by clustering. This was done by observation only and we searched for clusters with similar components. K-means clustering was chosen because we do not know any expert

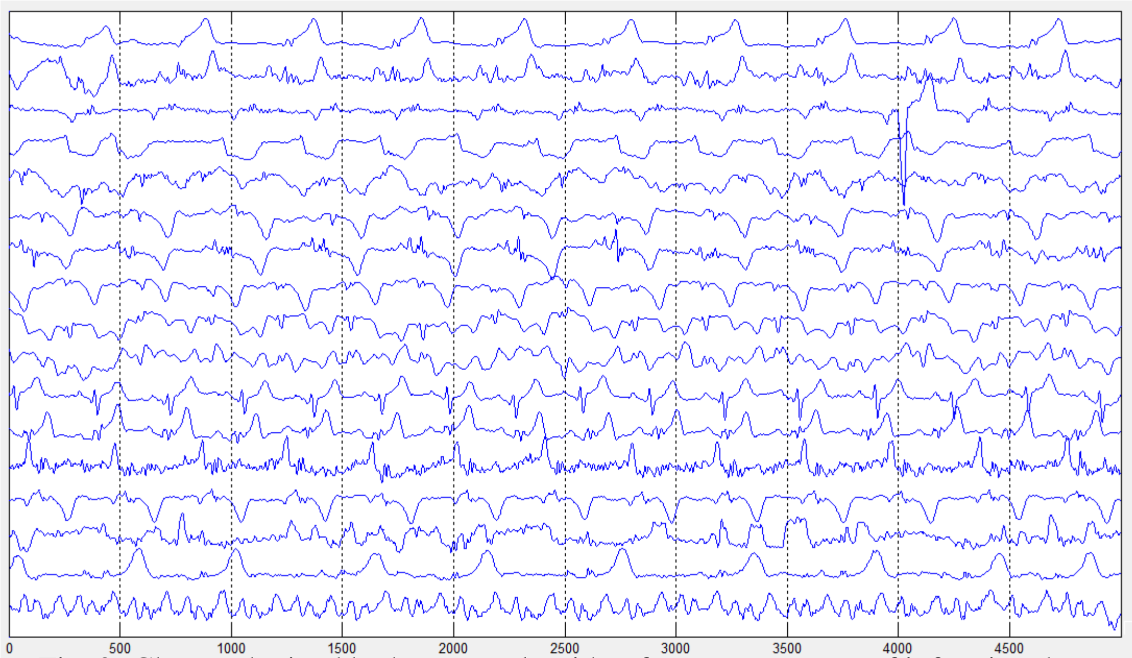

Fig. 2: Cluster obtained by k-means algorithm from components of infarction data.

classification and we need to test our feature set for classification of components obtained by ICA.

4 Results

We used small group of data that was selected from a large database provided by Medical Technologies CZ a. s. This database consists of more than 6000 rest ECG recordings. We selected 20 myocardial infarction records and 20 non-infarction records because our experiment was only preliminary analysis that should reveal possibilities of ICA in analysis of myocardial infarction. This means that we selected 320 components (8 from each record) for cluster-analysis. Selection was done randomly.
It was observed that QRS complex was divided into 3 different components each containing one wave from QRS complex. This was in accord with expectation and assumptions of other researchers. T-wave is separated in its own component, too. Components from group of infarction data were clustered more specifically than non-infarction data. The T-waves components were mostly in one cluster. The cluster contained 17 from 20 T-wave components. You can see it in Figure 2. Components containing QRS complex were in two different clusters. This is because of features selection. Different components (Q-, R-, S-, T-wave components) from group of non-infarction data were mixed into all 4 groups. There was not any significant cluster containing only one type of data.

5 Discussion

The performed observations showed that infarction data had more significant T-wave components than non-infarction data. These components were separated mostly in one cluster, so we can assume that they have more significant values of features than other components which were mixed together in different clusters. These observations were done only on small dataset and must be proved on larger dataset consisting of the whole database. But observation also proved that we can create detection algorithm based on our presented features and ICA algorithm which in combination with traditional rule-based detection algorithms make better solution of myocardial infarction detection. K-means algorithm gave us first basic insight in processed data and in our final solution it will be replaced by another classification algorithm, such as Support Vector Machines or perceptron based classification in combination with Adaboost algorithm.

6 Conclusions

We made first analysis of myocardial infarction database, which is not a standard one and covers more data than other usual experimental databases. Our goal was to compare components of infarction and non-infarction recordings. It was shown that infarction records had more significant T-wave component than non-infarction. And they were clustered by k-means algorithm in one cluster. Other components were scattered through all other clusters. Based on knowledge of basic myocardial infarction characteristics, we can assume, that T-wave components are characteristic for infarction data only. We will perform analysis on larger dataset and create ICA based detector of myocardial infarction in our future work.

7 Acknowledgement

Research described in the paper has been supported by the ČVUT Grant SGS10/279/OHK3/3T/13, research program No. MSM 6840770012 "Transdisciplinary Research in Biomedical Engineering II" of the CTU in Prague and CTU MediaLab Foundation.

References

[1] Pierre COMON, *Independent component analysis, a new concept?*, Signal Processing **36** (1994), 287 – 314.

[2] J. Michael HERRMANN and Fabian J. THEIS, *Statistical analysis of sample-size effects in ica*, (2007).

[3] Aapo HYVÄRINEN, Juha KARHUNEN, and Erkki OJA, *Independent component analysis*, Wiley, 2001.

[4] Aapo HYVÄRINEN, Jaakko SÄRELÄ, and Ricardo VIGÁRIO, *Spikes and bumps: Artefacts generated by independent component analysis with insufficient sample size*.

[5] P. M. RAUTAHARJU, L. P. PARK, B. R. CHAITMAN, F. RAUTAHRJU, and Z. M. ZHANG, *The novacode criteria for classification of ecg abnormalities and their clinically significant progression and regression.*, J Electrocardiol **31** (1998), no. 3, 157–187 (eng).

[6] José Joaquín RIETA, Francisco CASTELLS, César SÁNCHEZ, Vicente ZARZOSO, and José MILLET, *Atrial activity extraction for atrial fibrillation analysis using blind source separation*, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING **51** (2004), 1176 – 1186.

[7] Jaakko SÄRELÄ and Ricardo VIGÁRIO, *Overlearning in marginal distribution-based ica: Analysis and solutions*, Journal of Machine Learning Research **4** (2003), 1447–1469.

[8] Telemachos STAMKOPOULOS, Konstantinos DIAMANTARAS, Nicos MAGLAVERAS, and Michael STRINTZIS, *Ecg analysis using nonlinear pca neural networks for ischemia detection*, IEEE TRANSACTIONS ON SIGNAL PROCESSING **46** (1998), 3058 – 3067.