Proceedings of the 8th International Conference on Applied Informatics
Eger, Hungary, January 27–30, 2010. Vol. 2. pp. 129–137.

Clustering of ECG Signals

Anikó Vágner[a], István Juhász[b], Ádám Balázs[c], József Kuk[d]

[a]University of Debrecen, Faculty of Informatics
e-mail: vagner.aniko@inf.unideb.hu
[b]University of Debrecen, Faculty of Informatics
e-mail: juhasz.istvan@inf.unideb.hu
[c]University of Debrecen, Faculty of Informatics
e-mail: broothy@gmail.com
[d]University of Debrecen, Faculty of Informatics
e-mail: kukjozsef@gmail.com

Abstract

Holter electrocardiographic (ECG) signals are ambulatory long-term registers used to detect heart diseases. These signals normally include more, than one channels and its duration is up to 24 hours. The principal problem of the cardiologists is the manual inspection of the whole Holter ECG to find all those beats which morphology differs from the normal beats.

In this paper we present our method. We apply firstly a preprocessing to represent a signal with characteristic points. In this way we work with a set of points with a few dimensions instead of the signals. Secondly we apply a grid clustering technique to reduce the number of points. Finally we use a special density-based clustering algorithm, named Optics. We can also visually identify the beats or a partition of points. In this way we can perform manual clustering. With this method, the result is easy ECG analysis and optimization of time of processing.

We created a program used C# programming language in .NET environment. The main functionality of our program is: preprocessing and visualization, manual or graphical clustering and automatic clustering.

Keywords: clustering, signal processing, application development

1. Introduction

There are new developments and improvements in biosignal recording devices. These are containing acquired signals increased in quality. This is an important problem both on the processing and in the storage. In addition, these kinds of records provide a huge amount of information.

A cooperative research came into being between University of Debrecen Faculty of Informatics and Labtech Ltd. in March 2009. Labtech is a little company; it develops and manufactures ECG systems. The Labtech improves the software that belongs to the ECG systems. The software has many parts. Our task was to develop software in .NET environment. This is clustering heartbeats on the basis of files prepared by Labtech.

On Figure 1 is an original ECG record. The physicians work with this record. The record has two or three channels. It is 24 hours long that means approximately one hundred thousand heartbeats.

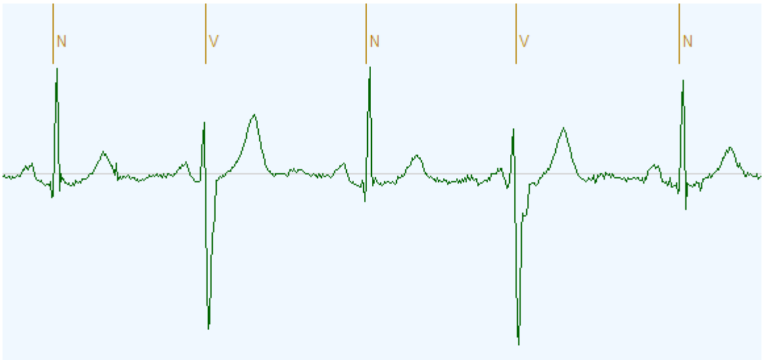

Figure 1: ECG signals

The annotations show us the place of heartbeat in the ECG record. On Figure 1 the red lines show us the annotations. The letter on the line marks the character of heartbeat. This character is not important in our project.

Our task was clustering the similar ECG signals. We want to make easier the job of the physician. The physician doesn't need to examine all signals. It is enough to examine each cluster of ECG signals. He or she can choose easier beats which have morphology differ from the normal beats. [3]

2. Main results

At first we worked on signaling. On Figure 2 you can see a visualization screen. We split the ECG record into signals using the annotation. We draw the signals on one another. And we get the left picture on the Figure 2.

On the left figure there are all signals of one channel of a record. The darker the part of the figure is, the more signals are on the given place. You can realize that there are various signal types on this figure.

Our aim to separate the various signal types into clusters. There are two clusters on the right side.

On the left figure you can notice a characteristic cluster. Many signals belong to this cluster, because it's dark. This cluster is in the right upper figure. There is another characteristic cluster on the left figure. This cluster has also many signals.

This cluster is on the right lower figure. You can notice that the two clusters are different, so we can't take them into one cluster.

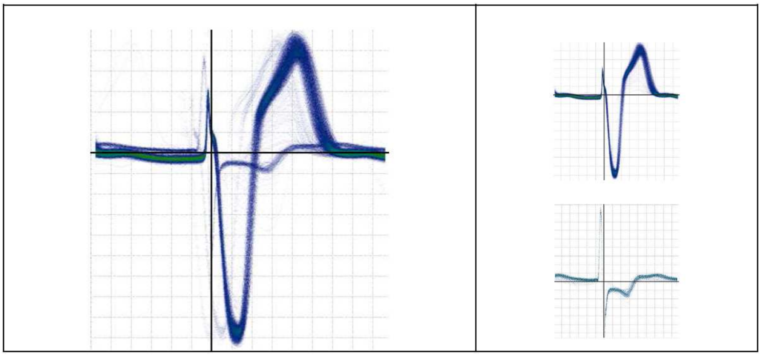

Figure 2: Visualization

And you can notice some other clusters on the left figure, too. These clusters have fewer signals. There aren't these clusters on the right.

Beside the original ECG's we got filtered signals from Labtech, which essentially are the derivative of the original signals. This derivative describes the shape of original record well. You can see filtered signals on Figure 3.

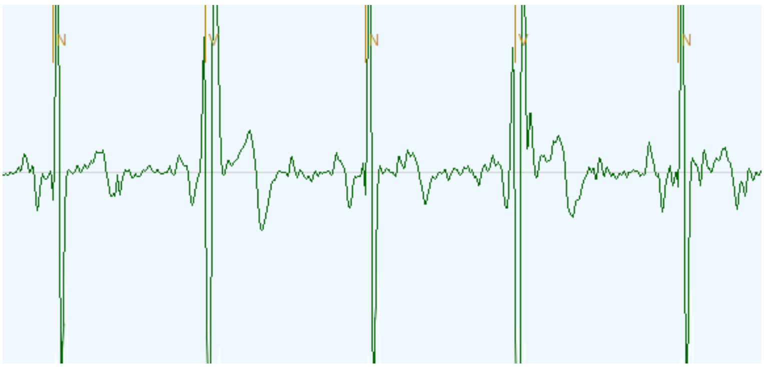

Figure 3: Filtered ECG signals

The second task was the preprocessing. Our first idea in clustering ECG signals is to not work on filtered signals, but to find characteristic points. In our algorithm we find the reference point first. The reference point is the first intersection of the x axis and the signal, after the annotation. It is the green point on the Figure 4. Next we find the extreme values after and before the reference point. The red points represent the extreme values and its places.

We transform the three points into a coordinate system in such a way, that the green point is at the origin. So the two red points represent the filtered signal. In this way we have information about the width and height of the filtered (and so of the original) signal. After this we work with a pair of two-dimensional points instead of the signal. If we plot all pair of points in a coordinate system, we get a set of points (Figure 5). It is easier to cluster the set of points, than to cluster the signals.

Figure 4: Representing a signal with characteristic points

We transform the three points into a coordinate system in such a way, that the green point is at the origin. So the two red points represent the filtered signal. In this way we have information about the width and height of the filtered (and so of the original) signal. After this we work with a pair of two-dimensional points instead of the signal. If we plot all pair of points in a coordinate system, we get a set of points (Figure 5). It is easier to cluster the set of points, than to cluster the signals.

Now, our task is to cluster the set of points.

To the clustering we have to keep in mind, that the set of points is not totally random. The pair of points is in the first and third quarter or in the second and fourth quarter.

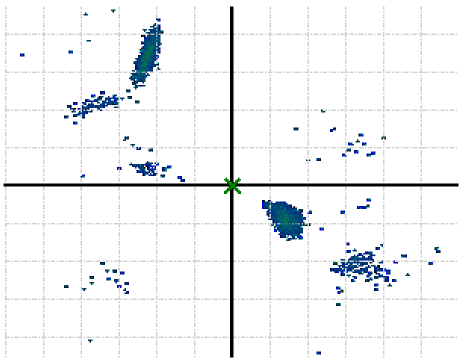

Figure 5: The set of characteristic points

From the well-known types of clustering algorithms ([4], [5], [6]) we apply the grid-based method. In this method we transform the points into grids, and later we work just with these grids. The main advantage of this method is the speed. In our case many points in the set of point have the same coordinate or they are close to each other. With the grid-based method we can radically reduce the number of points. The runtime of our algorithm was also radically reduced after applying of the grid-based method. In each grid we count how many points get there, and this number used as a similarity metrics.

The most clustering methods can build only cluster, with elliptic shape. The density-based methods can discover arbitrarily shaped cluster. The basic idea of density-based clustering is that for each point of a cluster the neighborhood of a given radius has to contain at least a minimum number of points. With density-based methods, density can be defined to be the number of values in a predefined unit area in the data space. The purpose of this kind of clustering is to group points from each high-density region into a cluster respectively and ignore the objects in low-density regions. [4]

We can identify arbitrarily shaped cluster, when we inspect the Figure 6. Point pairs circled with red are represent certain clusters. Circled ones with green are not represent clusters.

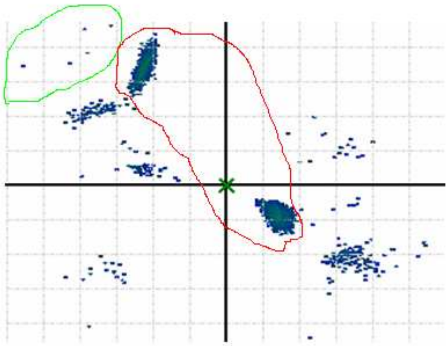

Figure 6: Clusters

These methods are dynamic methods; we don't need to give the number of clusters. The clustering changes based on parameters, like the radius and the threshold. One of the density-based methods is the Optics [1], [2].

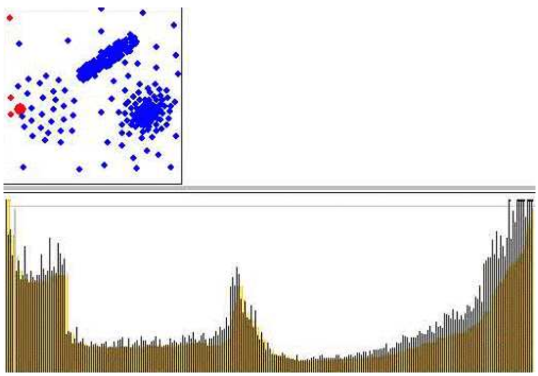

Figure 7: The Optics method

The Optics method orders the points and assigns reachability distance value to each point. On the upper of Figure 7 there are the points need to be clustered, and bottom is a belonging two dimensional plot, with the ordering of the points on the x-axis and the reachability distance on the y-axis. The points, outside of clusters are at the end of the order, and the given reachability distance are very big.

We identify the clusters with giving a threshold value for the reachability distance. On the Figure 7 the horizontal line represents the threshold. Each valley, that crosses the line, is made a cluster. Now is one cluster. If the line was moved down, more clusters would be identified. In our algorithm we use this method, but we customized it.

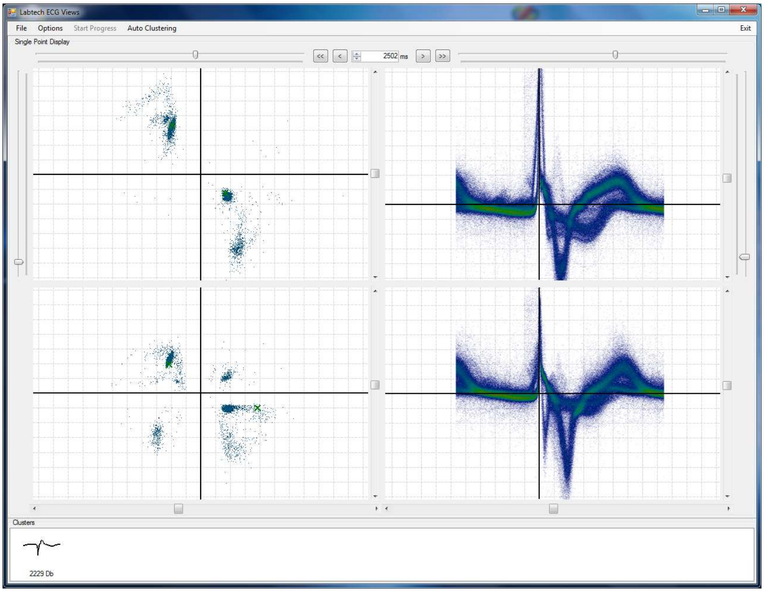

Figure 8: The visualization screen

3. Main functionality of C# program

3.1. Preprocessing and visualization

On the right side of the screen on Figure 8 there are the signals drawn on one another. On the left side there is the set of points. Now we work with two channels. At the bottom there are the templates. A template is the median of the signals, belong to the cluster. Now there is only one template, because this is an initial state.

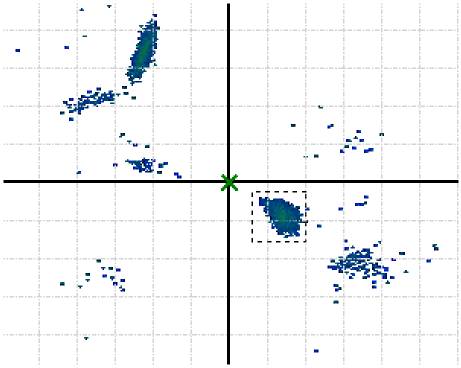

Figure 9: Manual clustering

3.2. Manual clustering

By manual clustering you can see the set of points belonging to one cluster. You cut with the mouse the points belonging to a cluster. See Figure 9.

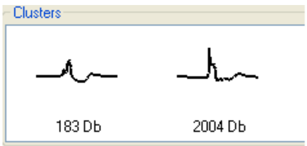

Figure 10: Templates after manual clustering

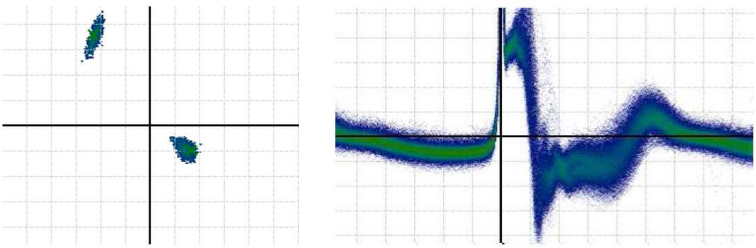

Figure 11: One cluster and the belonging signals

The original set of points is divided into two clusters (Figure 10). A template is the median of signals. You can study the set of points and the curves, if you click to the template. You see in the Figure 11, that the curves belonging to a template build really one cluster. You can continue to divide the clusters.

You can do the same thing with the curves, too. Namely you can cut the part of the curves too, and they are also divided into two clusters.

3.3. Automatic clustering

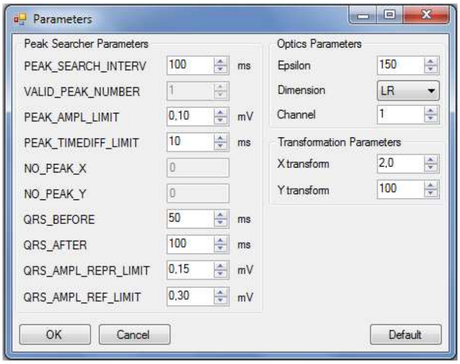

Figure 12: Parameters for the automatic clustering

There are parameters for the automatic clustering on the Figure 12. The parameters on the left side control finding of reference point and extreme values. The transformation parameters are the grid parameters. In the optics parameters: Epsilon is the radius in the optics method. (The dimension is not important any more). The channel is the channel of ECG record.

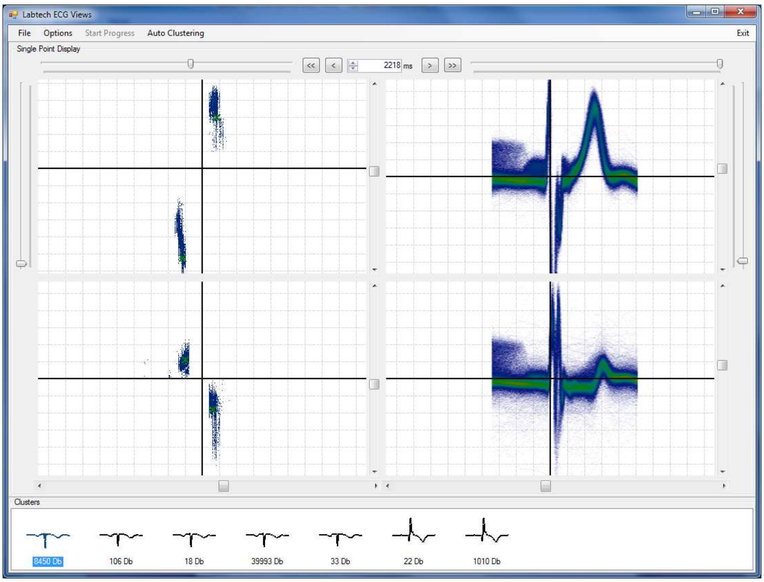

Figure 13: Automatic clustering

The program creates the templates represents the clusters after setting of parameters. First it uses the grid-based method, then the modified Optics method. The result can be seen on Figure 13. On the bottom side there are the templates. If you click to each template, the set of points and the curves belonging to the cluster appear.

The clusters can be refined.

4. Future work

Our experiences until now shows that it is enough to find one left and one right extreme values. Later, if it is necessary, we can work with more extreme values in more dimensions. The prototype is ready, now we develop the product.

References

[1] M. ANKERST, M. M. BREUNIG, H.-P. KRIEGEL, and J. SANDER, Optics: ordering points to identify the clustering structure, in SIGMOD '99: Proceedings of the 1999 ACM SIGMOD international conference on Management of data, (New York, NY, USA), pp. 49-60, ACM Press, 1999.

[2] `http://www.dbs.informatik.uni-muenchen.de/Forschung/KDD/Clustering/index.html`

[3] PAU MICÓ, D. CUESTA, and D. NOVÁK, Clustering Improvement for Electrocardiographic Signals, Image Analysis and Processing - ICIAP 2005 LNCS 3617/2005 pp. 892-899, Springer Verlag, 2005.

[4] S. B. KOTSIANTIS and P. E. PINTELAS, Recent Advances in Clustering: A Brief Survey, Transactions on Information Science and Applications, 2004,1(1) pp. 73–81.

[5] P. BUNGKOMKHUN and S. AUWATANAMONGKOL, Grid-based Supervised Clustering - GBSC, World Academy of Science, Engineering and Technology 60 pp. 536–543, 2009.

[6] R. XU and D. WUNSCH II, Survey of Clustering Algorithms, IEEE Transactions on Neural Networks, Vol. 16, No. 3, pp. 645–678, MAY 2005.

Anikó Vágner, István Juhász, Ádám Balázs, József Kuk
University of Debrecen, Faculty of Informatics
4032 Debrecen, Egyetem tér 1.