Nature Inspired Clustering Methods in The Electrocardiogram Interpretation Process in Cardiology

Bursa M[1], Lhotska L[1], Travnicek Z.[1]

[1]BioDat Research Group, Gerstner laboratory, CTU in Prague,

[2]Institute of Intermedia, CTU in Prague

bursam@fel.cvut.cz

Abstract. The paper focuses on the field of artificial intelligence techniques and their use in biomedical data processing. It concentrates on the clustering techniques inspired by various ant colonies and other nature concepts. The paper evaluates the use of the following nature inspired methods: Ant Colony inspired Clustering, Ant Colony inspired method for Decision Tree generation, Radial Basis Function Neural Networks with different learning algorithms and compare them to classical approaches, such as k-means and hierarchical clustering. The methods have been evaluated using the annotated MIT-BIH database. Use of the Dynamic Time Warping measure improved Sensitivity about 0.7 % and Specificity about 0.9 % when compared to classical feature extraction. The best-performing method is the agglomerative hierarchical clustering (Se=94.3, Sp=74.1), however it is practically unusable as it is memory and computational demanding. Acceptable results (complexity vs. error) have been obtained by the Ant-Colony inspired method for Decision tree generation (Se=93.1, Sp=72.8).

1 Introduction

Arrhythmia diagnosis and ECG interpretation is an important non-invasive step in clinical diagnosis process. The presence of significant heart arrhythmias is an important marker of cardiac death. In our work we study the use of nature inspired techniques in the process of ECG interpretation.

During the analysis of Holter recording a huge amount of data is presented to the cardiologist. In commercial systems, mainly the template-matching method is used. The speed of final record evaluation and automated analysis is crucial. In clinical practice, 5 minutes of automated off-line data processing for 24 hour recording is considered as acceptable [1].

In this paper, the ant-colony inspired methods are used for biomedical data clustering.

2 Clustering Methods

The clustering algorithms can be classified as *stochastic* or *deterministic*. The clustering optimization (minimization of an objective function) can be accomplished using traditional techniques or through a random search of the state space driven by metaheuristics. The state space consists of all possible labelings.

2.1 Radial basis function (RBF) artificial neural networks

Radial basis function (RBF) neural networks (RBF-NN) have been introduced by Broomhead and Lowe in 1998 [2]. The model is inspired by the locally tuned response observed in biologic neurons that can be found in several parts of the nervous system, for example cells in the auditory systems which are selective to small bands of frequencies, in visual cortex, etc. Such locally tuned neurons show response characteristics bounded to a small range of the input space.

Using a supervised learning method, the RBF neural networks can perform as a clustering method that divides the incoming data into homogeneous clusters. In this paper we have evaluated the following nature inspired learning methods: the ACO_R method [3] inspired by ant paradigm and the Particle Swarm Optimization (PSO) algorithm [4] inspired by the behavior of bird flocks or fish schools. The evaluation of the nature inspired learning methods has been published in [5].

2.2 Ant inspired clustering

The clustering of dead bodies by certain ant species inspired Deneubourg et al. [6]. In their article they published biologically inspired model for clustering by a group of homogeneous agents. The ants possess only local perceptual capabilities and they can also compare whether the objects are similar or not to the object they are carrying.

Ant Colony inspired clustering [7] method belongs to the group of nature inspired methods (together with neural networks, self-organizing maps, swarm intelligence, evolutionary computing, etc.). It is a stochastic metaheuristics, which uses a similarity measure and the paradigm of ant species, and which is able to collect similar items (broods, dead bodies) together.

The model can be described as follows: First the data vectors are randomly scattered onto a two-dimensional grid (usually a toroidal one). Ants (agents) are then randomly placed onto the grid. In each iteration step an ant searches its neighborhood and computes a probability of picking up a vector (if the ant is unloaded and steps onto a vector) or a probability of dropping down a vector (if the ant is loaded and steps onto a free grid element). The probability of picking up a vector is higher when an ant steps on a vector that introduces heterogeneity in its surrounding. The probability of dropping down the vector is higher when an ant carrying a vector steps into an area of similar vectors. These rules lead to formation of clusters in the 2D grid which can be easily obtained.

2.3 Ant colony optimization for decision tree generation

The ACO_DTree method [8, 9] implemented uses a hybridization of Ant Colony Optimization method [10] with evolutionary approach to generate and optimize the population of binary decision trees. The algorithm maintains a population of decision trees and the whole population is evolved through time. The best solutions contribute into a pheromone matrix that is used to create new solutions and worst solutions are removed. The optimization of decision values of the trees is performed using the PSO algorithm [4]. Using a supervised training algorithm it can be used to perform clustering in a similar way as the RBF-NN.

The pseudo-code of the ACO_DTree method (see below) is included for clarity. The important steps are described in the following text.

```
1    Generate initial population of solutions
2    Repeat until stopping criterion is reached
3      Create new solutions (pheromone driven)
4      Evolve population (mutation)
5      Evaluate population (determine fittness)
6      Select best individuals
7      Pheromone evaporation
8      Daemon actions (pheromone laying, adapt.)
9    End Repeat
```

2.3.1 Decision Tree Construction

By a classification tree we mean hereby a tree-like structure composed of similar nodes. Each node can have left and right sub node. Each node is represented by a decision rule with two parameters (feature index $feature_j$ and decision value $decVal$) which can be described in the following way for an item s_i:

```
1  if (s(i).getFeature(j) < decVal)
2    classifyToLeftBranch
3  else
4    classifyToRightBranch
```

The same applies to the root node. The tree is constructed as follows: first, a random feature index is selected (for the root node). The decision value is also chosen randomly. Then, for each subnode, the subnode is created with certain probability (decreasing with the increasing level in the tree). The feature index is selected using the pheromone matrix: the edges which have proved to be successful are preferred. The decision value is determined randomly. The random selection of decision value does not present a problem, because even the randomly selected values perform acceptable solution. However, the population is then trained – the decision values are optimized (see later).

3 Data

The real ECG signals come from widely used MIT-BIH database [11] which contains annotated records. Description of the data can be found in the work of Chudacek et al.[1]. For the sake of simplicity, only two major distinct classes are often used: normal cardiac action and abnormal cardiac action (mainly premature ventricular contraction beats or left or right bundle branch block beats). The classification into more classes is nearly impossible due to lack of data (abnormal heart action signal) in some signals. By omitting minor abnormal heart actions, more records from the MIT-BIH database can be processed.

From the ECG signal, the following eight features have been automatically extracted[1]: (i) amplitudes of Q, R, S, positive T and negative T wave and (ii) amplitude ratio of Q/R, R/S and R/T waves.

For data clustering, the features have been normalized to the interval of $\langle 0.0; 1.0 \rangle$. Along with the features a signal window of 401 samples has been stored (containing the beat which was used for feature extraction) thus making it available to perform a DDTW distance evaluation.

4 Results

Table 1 shows an average sensitivity and specificity for all methods evaluated. Sensitivity is computed as $Se = TP/(TP + FN)$, specificity as $Sp = TN/(FP + TN)$. These measures are commonly used in clinical statistics.

The method has been (after preliminary parameter-tuning tests on smaller datasets) applied to the MIT-BIH database with more than 80.000 records.

The best results have been achieved by the hierarchical agglomerative method, which is the only method not using centroid approach. It is however very demanding. The nature inspired methods, however, outperformed the basic k-means algorithm both in specificity and sensitivity and achieved more stable results (in term of standard deviation).

Using the DTW measure improved Se about 0.7 % and Sp about 0.9 % when compared to classical feature extraction for a #106 signal. This leads to the overall Se=92.8 and Sp=90.1 which is satisfactory as no other feature extraction has been used. The disadvantage is the computational complexity of the measure.

The best-performing has been the agglomerative hierarchical clustering (Se=94.3, Sp=74.1), however it is practically unusable as it is memory and computational demanding. Acceptable results (complexity vs. error) have been obtained by the Ant-Colony inspired method for Decision tree generation (Se=93.1, Sp=72.8).

5 Discussion and conclusions

The studied nature inspired methods provide an excellent heuristics when compared to the exact and exhausting methods which justifies their use in the process of ECG data processing.

The results have shown that the performance of the nature inspired method is comparable and in some cases outperforms the classical methods. Agglomerative clustering requires a lot of resources (memory and processor time). K-Means method provides faster results, but easily gets stuck in the local minima. The advantage of Ant Clustering and Kohonen network is that

Tab. 1: Method Comparison

Method/Result	Measure	SE	SP
K-Means	L2	77	65
Agglomerative	L2	94	74
Ant Colony Clustering	L2	79	68
Kohonen SOM	L2	79	66
Ant Colony Clustering	DTW	79	69
RBF (ACO_R)	L2	87	71
RBF (PSO)	L2	89	68
ACO_DTree	L2	93	72

it can determine natural clusters within data. The drawback (mainly for Kohonen network) is its high computing resources consumption.

The main advantage of the (ACO_DTree) is that it produces a white-box structure, which can be understood also by physicians. It can be also used in long-term processing, as the application is straigthforward and simple. It also reduces the memory needed (in the application phase), as it can process the data in a serial manner.

The DTW measure produced better results than the Euclidean measure (using automatically extracted features). This is a significant result, which (in combination of the measures) should lead to better overall results as the solutions obtained are not identical. The future work would concentrate on combination of these two approaches to obtain better results.

The ant colony inspired methods can be applied in many stages of the ECG processing process. However, only some process stages are really suitable for such methods. The ant colony methods can find only an approximation of the best solution. Thus some compromise has to be accepted, which can be a problem for non-informed physicians. The evaluation of the ECG clusters must be validated by cardiologists in the final stage.

Very important step in the method implementation and evaluation process is the visualization of data and relevant processes in the algorithm run. The resulting clusters can be visualized and further classified. Great advantage is a 3D visualization that allows natural insight into inner functionality within the method and provides additional information. In the experiment we have the oportunity to utilize a 3D stereo (full stereo) projection onto three planes (front, left and right). Such visualization is used nowadays in many operating micromanipulators (robots) and is very important for medical doctors.

Acknowledgements

References

[1] Chudacek, V., Lhotska, L.: Unsupervised creation of heart beats classes from long-term ecg monitoring. Conference: Analysis of Biomedical Signals and Images. 18th International EURASIP Conference Biosignals 2006. Proceedings. **18** (2006) 199–201

[2] Broomhead, D., Lowe, D.: Multivariable functional interpolation and adaptive networks. Complex Systems **2** (1988) 321–355

[3] Socha, K., Dorigo, M.: Ant colony optimization for continuous domains. European Journal of Operational Research **185(3)** (2008) 1155–1173

[4] Kennedy, J., Eberhart, R.C.: Particle swarm optimization. Proceedings IEEE International Conference on Neural Networks **IV** (1995) 1942–1948

[5] Bursa, M., Lhotska, L.: Nature inspired methods in the radial basis function network learning process. Springer Berlin/Heidelberg, LNCS **5163** (2008) 829–838

[6] Deneubourg, J.L., Goss, S., Franks, N., Sendova-Franks, A., Detrain, C., Chretien, L.: The dynamics of collective sorting robot-like ants and ant-like robots. In: Proceedings of the first international conference on simulation of adaptive behavior on From animals to animats, Cambridge, MA, USA, MIT Press (1990) 356–363

[7] Lumer, E.D., Faieta, B.: Diversity and adaptation in populations of clustering ants. From Animals to Animats: Proceedings of the 3th International Conference on the Simulation of Adaptive Behaviour **3** (1994) 501–508

[8] Bursa, M., Huptych, M., Lhotska, L.: Ant colony inspired metaheuristics in biological signal processing: Hybrid ant colony and evolutionary approach. Biosignals 2008-II. Setubal: INSTICC Press **2** (2008) 90–95

[9] Bursa, M., Lhotska, L., Macas, M.: Hybridized swarm metaheuristics for evolutionary random forest generation. Proceedings of the 7th International Conference on Hybrid Intelligent Systems 2007 (IEEE CSP) (2007) 150–155

[10] Dorigo, M., Blum, C.: Ant colony optimization theory: A survey. Theoretical Computer Science Issues 2–3 **344** (2005) 243–278

[11] Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation **101**(23) (2000) e215–e220