

# Nature Inspired Concepts in the Electrocardiogram Interpretation Process

M Bursa, L Lhotska

Czech Technical University in Prague, Prague, Czech Republic

## Abstract

In this paper we compare and evaluate the use of the following 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 hierarchical clustering and *k*-means. We have evaluated the methods on the annotated MIT-BIH database. In the case of Ant Colony inspired clustering we have also studied the Dynamic Time Warping (DTW) measure.

The DTW measure improved *Se* about 0.7 % and *Sp* about 0.9% when compared to classical feature extraction for a #106 signal. 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).

## 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.

Holter recording is a long-term ECG recording used as a non-invasive diagnostic aid in coronary disease diagnostics and also in heart disease prevention and treatment. It is usually recorded over a time period longer than 24 hrs using a reduced lead system. Nowadays the Holter device can store high resolution recordings (with higher sampling frequency and sample resolution) which allows the investigation and monitoring such markers as microvolt T-wave alternans [1], etc.

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 24hour recording is considered as acceptable [2].

Data clustering, in literature referred as *cluster analysis*, numerical taxonomy, typological analysis, etc., is a common unsupervised learning technique aimed at accumulation of similar patterns into groups (clusters): partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share certain common aspect which is defined by the defined distance measure. Review of main clustering classes (partitioning methods, hierarchical methods, density-based methods and grid-based methods) can be found in [3].

Nature inspired metaheuristics play an important role in the domain of artificial intelligence, offering fast and robust solutions to many problems, although they belong to the branch of approximative methods. Stochastic nature inspired metaheuristics have interesting properties which make them suitable to be used in data mining, data clustering and other application areas. We concentrate on the state-of-the-art nature methods inspired by the social behavior of insect communities, by the swarm intelligence, brain processes and other real nature processes (physiological, genetic, communication, information processing, and many other).

## 2. Metrics considered

A key part in the clustering process is to properly use relevant distance measure(s). A simple measure as Manhattan distance (L1 norm) can be used. A more common measure used is Euclidean distance (L2 norm). These two metrics (L1 and L2) are special cases of more general Minkowski distance where  $p = 1$  and  $p = 2$  respectively. The Minkowski distance is defined as follows:

$$L_p(\mathbf{x}, \mathbf{y})_p = \left( \sum_{i=1}^d |\mathbf{x}_i - \mathbf{y}_i|^p \right)^{1/p} = \|\mathbf{x} - \mathbf{y}\|_p \quad (1)$$

which is defined for any two vectors  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_d)$  and  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_d)$  of the same dimension  $d$  (where every

( $i$ -th) component  $x_i \in \mathcal{R}$  and  $y_i \in \mathcal{R}$ ). The drawback in the direct use of Minkowski metrics is the tendency of the largest-scaled feature to dominate the others. Solutions to this problem include normalization or weighting policies. In case of nominal or symbolic data, a mapping to discrete numeric domain is often used.

Many other distance metrics have been proposed in the literature, for example the Chi-square metric [4], the Mahalanobis metric [5], the Cosine Similarity metric [6], the Quadratic metric [7], the Modified Value Difference metric [8], Hausdorff metric [9, 10], etc.

In the clustering methods described, the Euclidean metric (also called  $L2$  norm) is the most often used. Note that for comparison purposes and in order to speed up the comparison process, the computation of square root is not necessary.

## 2.1. Dynamic time warping (DTW) metrics

The DTW algorithm [11] technique allows comparison of time series which are moderately shifted (or varied) in the time dimension (direction of time axis). The limit case is the L2 distance, where the warp path reduces into a straight line (supposed that both vectors have the same length). Figure 1 and Figure 2 shows an example of time warping measure applied to two ECG beats. Each figure shows two compared signals (one of them is rotated by  $90^\circ$  counter-clockwise for better visualization). The segmented line between signals represents the warp path and begins in the upper left corner and runs to the bottom right corner. The signals in the Fig. 1 belong to the same group (both are normal (non-pathological) heartbeats). Note, that the warp path is not very distant from the diagonal line in this case.

DTW metrics can be defined for any two vectors  $\mathbf{x} = (x_1, \dots, x_d)$  and  $\mathbf{y} = (y_1, \dots, y_d)$  of the same dimension  $d$  (for simplicity) as follows:

$$DTW(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^d |x_i - y_{\delta[i]}|^2} \quad (2)$$

if the L2 norm is used for the individual features (other measure can be also used). The  $\delta$  vector contains the warping path (assigns each index  $i$  an index value  $\delta[i] \in \{1, \dots, d\}$  for the vector  $\mathbf{y}$ . Note that if  $\delta = \{1, 2, \dots, d\}$ , we obtain the L2 norm.

The drawback of this method is that it can perform unintended warping by finding the minimal-cost warping path (correctly from the view of the method). See Fig. 2, where two non-similar signals are matched using the DTW metrics. Note, how the warp path approaches the extremities instead of sticking to the diagonal when compared to Fig. 1. The Sakoe-Chiba band [12] and Itakura band [13]

which constraint the warping path in a certain (parametrically defined) distance from the diagonal are commonly used to overcome this issue.

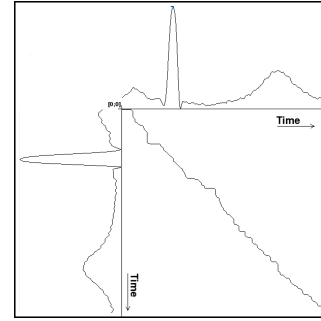


Figure 1. Dynamic time warping (DTW) producing correct (wanted) path (both signals belong to the same group).

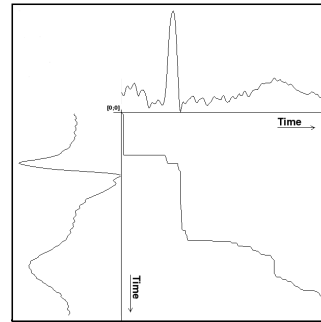


Figure 2. Dynamic time warping (DTW) algorithm producing incorrect warping (unwanted – the signals belong to different groups).

## 3. Nature inspired methods

### 3.1. Radial basis function (RBF) artificial neural networks

Radial basis function (RBF) neural networks (RBF-NN) have been introduced by Broomhead and Lowe in 1998 [14]. 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.

Radial basis function networks consist of a layer of units with radial basis activation function (RBF) and an output layer of linear summation unit(s). As the RBF, often Gaussian activation functions are used, therefore the corresponding units are called *Gaussian (kernel) units*. Each

Gaussian unit  $c$  has an associated vector  $\mathbf{m}_c \in \mathcal{R}^n$ , indicating the position of the Gaussian in the  $n$ -dimensional input vector space, and a standard deviation  $\sigma_c$ .

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 [15] inspired by ant paradigm and the Particle Swarm Optimization (PSO) algorithm [16] inspired by the behavior of bird flocks or fish schools. The evaluation of the nature inspired learning methods has been published in [17].

### 3.2. Ant inspired clustering

The clustering of dead bodies by certain ant species inspired Deneubourg et al. [18]. 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 [19] 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.

### 3.3. Ant colony optimization for decision tree generation

The  $ACO\_DTree$  method [20,21] uses a hybridization of Ant Colony Optimization method [22] 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 [16]. Using a supervised training algorithm it can be used to perform clustering in a similar way as the RBF-NN.

## 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. All results have been first clustered into four classes (the data contained four natural classes); final classification used the majority vote rule. Only the best results are presented in the table.

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).

Table 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

## 5. Discussion and conclusions

The studied nature inspired methods provide an excellent heuristics over the exact 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 out-

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

Ant Colony Clustering method can be efficiently used together with DTW metrics for time series clustering. However, some restriction on the warping path has to be defined. Wider restriction band does not necessarily mean better results, experimentally a band width of 10 samples was considered as appropriate to the signal length. Note that DTW measure increases comparison time (it is an  $O(n^2)$  algorithm in comparison to L2 metrics ( $O(n)$ )). Centroid methods are not suitable for the Derivative Dynamic Time Warping metrics used as they must define and find appropriate center (with 401 dimensions in our case).

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.

## Acknowledgements

This research project has been supported by the research program number MSM 6840770012 Transdisciplinary Research in the Area of Biomedical Engineering II of the CTU in Prague, sponsored by the Ministry of Education, Youth and Sports of the Czech Republic. This work has been developed in the BioDat research group <http://bio.felk.cvut.cz>.

## References

- [1] Bloomfield DM, Bigger JT, Steinman RC, Namerow PB, Parides MK, Curtis AB, Kaufman ES, Davidenko JM, Shinn TS, Fontaine JM. Microvolt t-wave alternans and the risk of death or sustained ventricular arrhythmias in patients with left ventricular dysfunction. *J Am Coll Cardiol* 2006;47(2):456–463.
- [2] 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* 2006;18:199–201.
- [3] Jain A, Murty MN, Flynn PJ. Data clustering: A review. *ACM Computing Surveys* 1999;31:264–323.
- [4] Wilson D, Martinez T. Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research* 1997;6:1–34. URL <http://www.jair.org>.
- [5] Mahalanobis P. On the generalised distance in statistics. *Proceedings of the National Institute of Science of India* 1936;12:49–55.
- [6] Salton G, McGill M. *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1983.
- [7] Fukunaga J, Beauregard R. An optimal global nearest neighbor metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984;6(3):314–318.
- [8] Cost S, Salzberg S. A weighted nearest neighbor algorithm for learning with symbolic features. *Mach Learn* 1993;10:57–78.
- [9] Huttenlocher D, Klanderman D, Rucklidge A. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* September 1993;15(9):850–863.
- [10] Alt H, Behrends B, Blömer J. Approximate matching of polygonal shapes (extended abstract). In *SCG '91: Proceedings of the seventh annual symposium on Computational geometry*. New York, NY, USA: ACM Press. ISBN 0-89791-426-0, 1991; 186–193.
- [11] Myers CS, Rabiner LR. A comparative study of several dynamic time-warping algorithms for connected word recognition. *The Bell System Technical Journal* 1981;607:1389–1409.
- [12] Sakoe, Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics Speech and Signal Processing* 1978;ASSP-26, no.1:43–9.
- [13] Itakura F. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics Speech and Signal Processing* 1975;ASSP-23, no.1:670–72.
- [14] Broomhead D, Lowe D. Multivariable functional interpolation and adaptive networks. *Complex Systems* 1988;2:321–355.
- [15] Socha K, Dorigo M. Ant colony optimization for continuous domains. *European Journal of Operational Research* 2008; 185(3):1155–1173.
- [16] Kennedy J, Eberhart RC. Particle swarm optimization. *Proceedings IEEE International Conference on Neural Networks* 1995; IV:1942–1948.
- [17] Bursa M, Lhotska L. Nature inspired methods in the radial basis function network learning process. *LNCS Springer in press* 2008;.
- [18] Deneubourg JL, 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. ISBN 0-262-63138-5, 1990; 356–363.
- [19] Lumer ED, 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* 1994;3:501–508.
- [20] Bursa M, Huptych M, Lhotska L. Ant colony inspired metaheuristics in biological signal processing: Hybrid ant colony and evolutionary approach. *Biosignals 2008 II Setbal INSTICC Press* 2008; 2:90–95.
- [21] 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.
- [22] Dorigo M, Blum C. Ant colony optimization theory: A survey. *Theoretical Computer Science Issues* 2 3 2005;344:243–278.

Address for correspondence:

Miroslav Bursa  
Dept. of Cybernetics, CTU in Prague, Technicka 2, 166 27  
Prague, Czech Republic  
[bursam@fel.cvut.cz](mailto:bursam@fel.cvut.cz)