

Artificial Intelligence in Biomedical Engineering

DRAGOS AROTARITEI
dragos.arotaritei@umfiasi.ro

University of Medicine and Pharmacy “Grigore T. Popa” Iasi
Department of Medical Bioengineering
Romania

May 2016



Introduction

- Artificial intelligence (AI) is an area of computer science, which has been developed since the 1950s, specialized in dealing with problems considered difficult by traditional computer scientists through the use of knowledge and of probabilities and other kinds of uncertainties.
- AI techniques can be applied to solve complex problems in biomedical engineering.
- AI subfields
 - Neural Networks
 - Evolutionary Computation,
 - Vision
 - Robotics
 - Expert Systems
 - Speech Processing
 - Planning,

-
- Machine Learning Algorithms can create new medical knowledge (supervised and unsupervised)
 - Natural Language
 - Fuzzy systems and approximate reasoning
 - Hybrid systems
- Machine Learning algorithms
 - UC Irvine Machine Learning Repository
 - <http://archive.ics.uci.edu/ml/>
 - Pattern recognition - a branch of machine learning that focuses on the recognition of patterns

Feature extraction

- ❑ Feature extraction is an intermediary step between preprocessing data and the process (classification, clusterization, rule generation, etc.)
- ❑ Two problems arise after data is collected: dimensionality of each item from data and the number of items. Sometimes both problems are possible to be solved in a single step.
- ❑ The digital information is usually structured in vectors and matrix having sometimes very large dimension. As consequence, manipulation of these types of data requires a dimensionality reduction: principal component analysis (PCA), linear discriminant analysis (LDA), canonical correlation analysis (CCA), singular value decomposition (SVD), dimensionality reduction via kernels.
- ❑ The second problems refers to selection of features from all the features that are relevant for our goal.

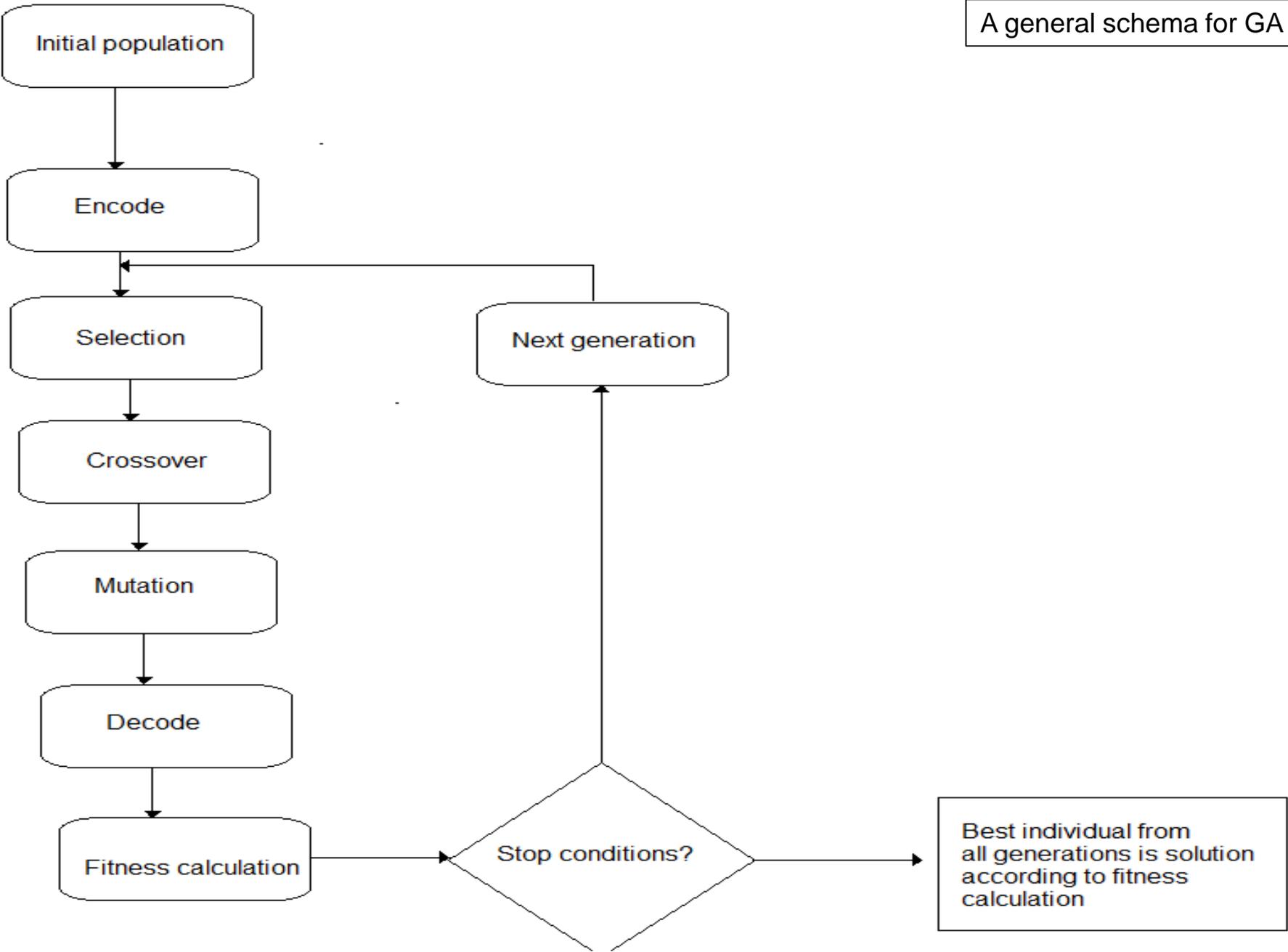
-
- independent component analysis (ICA), ISOMAP with extensions, kernel principal component analysis (kernel PCA), Latent semantic analysis (LSA), Partial least squares regression (PLS regression), orthogonal transformations
orthogonal transformation is a linear transformation $T : V \rightarrow V$ on a real inner product space V , that preserves the inner product), nonlinear dimensionality reduction (NLDR) - manifold learning algorithms
 - There are many overlapping algorithms that map the both objectives.
 - The feature extraction is can be completely automated very rarely, the ability of author plays an important role in identification and selection of possible relevant feature and the method (new or classic one) that leads to results.

Evolutionary Algorithms

- ❑ Subset of evolutionary computation
- ❑ The most frequently usage: optimization (single objective and multiobjective)
- ❑ Genetic algorithms (based on Darwinian evolution) – the most used in numerical optimizations
- ❑ Differential evolution – Based on vector differences
- ❑ Ant colony optimization – suitable for path optimization problems
- ❑ Particle swarm optimization
- ❑ Genetic programming – suitable for finding a mathematical formula for relationship among variables based on experimental data
- ❑ Memetic algorithms - most MAs can be interpreted as a search strategy in which a population of optimizing agents cooperate and compete using local strategies
- ❑ Other non-Darwinian algorithms: Baldwinian Evolutionary Algorithms, Lamarckian Evolutionary Algorithms, Genetic Local Search

GA (Genetic Algorithms)

- ❑ Used most frequently for numerical optimization problems
- ❑ GA belong to heuristic search that mimics the natural evolution (Darwinian evolution)
- ❑ A large area of applications including Biomedical Engineering, Clustering, Computational Biology, Image processing (Dense pixel matching).
- ❑ Explanations of the power of genetic algorithms is given by Holland's schema theorem (fundamental theorem of genetic algorithms). Low-order schemata with above-average fitness increase exponentially in successive generations. A schema is a template that identifies a subset of strings with similarities at certain string positions¹.
- ❑ The chromosome is used to code de information. The usually conversion is value to binary string of '0's and '1's. Each value '0'/'1' is named allele.



-
- ❑ *Encode/Decode* is shortcut for *convert values to chromosome/ convert chromosome to values*.
 - ❑ The algorithm start with an initial number of solutions (initial population) represented by chromosomes.
 - ❑ The key problem in GA is the definition of fitness, that usually means the function minimization (e.g. mean squared error), *fitness* = y . The algorithm remain valid for maximum of function if we replace *fitness* = $1-y$.
 - ❑ The GA operators are: Selection, Crossover and Mutation.
 - ❑ Selection
 - Roulette wheel selection - This fitness level is used to associate a probability of selection with each individual chromosome
$$p_i = \frac{f_i}{\sum_{i=1}^N f_i}$$
 - Tournament selection - tournament selection is often implemented in practice due to its lack of stochastic noise
 - Stochastic universal sampling
 - Reward-based selection

-
- Boltzmann selection
 - Truncation selection
 - Fair coin (a sequence of independent Bernoulli trials with probability 1/2 of success on each trial), $p = \text{rand}()$ in Matlab. If $p < 0.5$, $k = '1'$ else $k = '0'$. This is often mentioned as Coin flipping.
 - Crossover (mate).
 - One-point crossover
 - Two-points crossover
 - “Cut and splice“ – results in change of the length in children chromosomes
 - Uniform Crossover (UX) - uses a fixed mixing ratio between two parent
 - Half Uniform Crossover (HUX) - only half of the bits that are different will be exchanged using number of different bits (Hamming distance) between the parents. The half of this number is the number of bits is exchanged between parents to form the childs.
 - Other that not mimics the nature (Partially-mapped crossover (PMX), Order crossover (OX))



1 1 0 1 0 0 1 0 0 0

0 0 1 0 0 1 1 0 1 1

one point crossover

1 1 0 0 0 1 1 0 1 1

0 0 1 1 0 0 1 0 0 0

crossover point

1 1 0 1 0 0 1 0 0 0

0 0 1 0 0 1 1 0 1 1

two points crossover

1 1 0 0 0 1 1 0 1 1

0 0 1 1 0 0 1 0 0 0

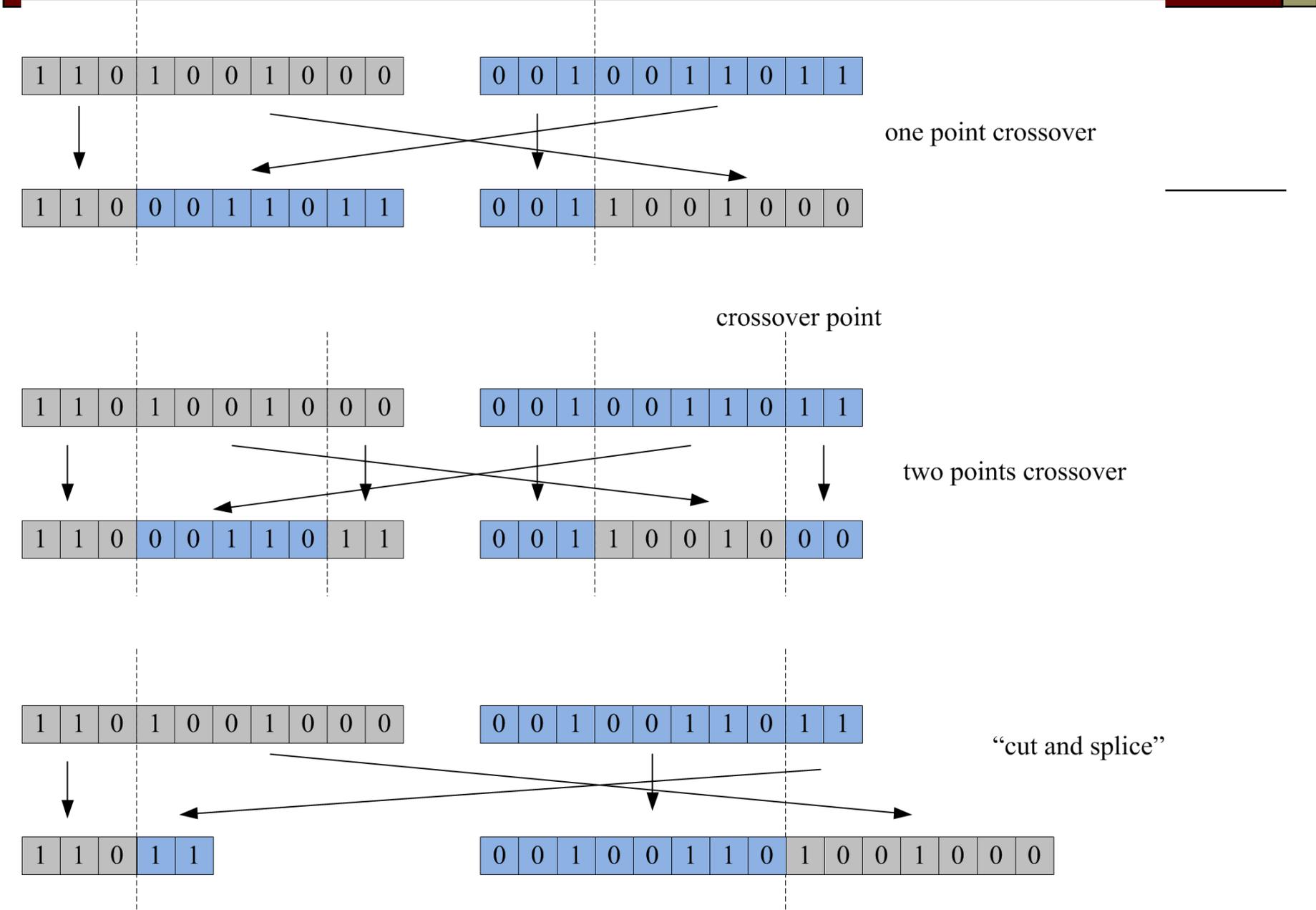
“cut and splice”

1 1 0 1 0 0 1 0 0 0

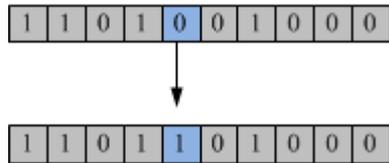
0 0 1 0 0 1 1 0 1 1

1 1 0 1 1

0 0 1 0 0 1 1 0 1 0 0 1 0 0 0



-
- *Mutation* is a genetic operator used to maintain genetic diversity of population

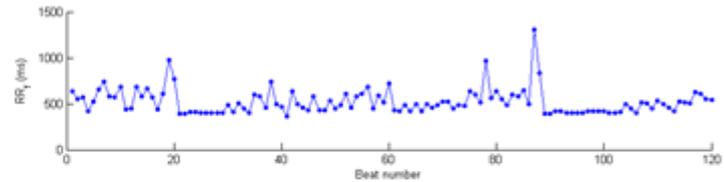
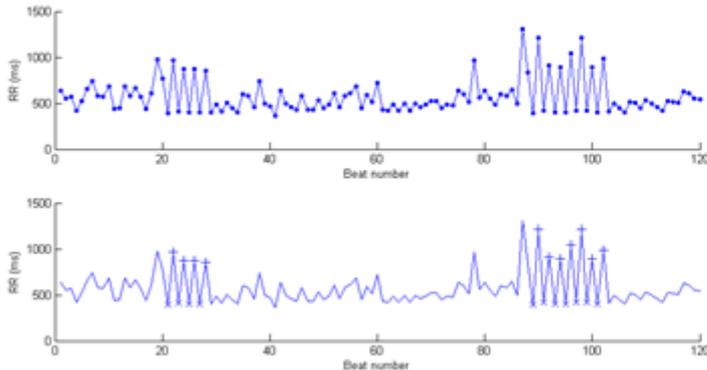


- Mutation occurs during evolution according to a probability defined by user. The usually mutation is made by '1' \leftrightarrow '0'
- Depending on genome (an organism that have a complete set of DNA, that is all the genes) type mutation can be:
 - Bit string mutation
 - Uniform
 - Non-uniform
 - Gaussian
 - Boundary (used for integer and float genes)

-
- Problems and limitations
 - Premature convergence
 - Elitism selection can produce a uniform population (diversity is at low level) and population cannot evolve toward a optimal solution
 - For high dimensional problems fitness can be expansive
 - GA algorithms do not scale well with complexity
 - Convergence to local optima instead of global optima (solution the penalty method)
 - For other specific problem some alternative methods can have a faster speed of convergence, e.g. integer linear programming.
 - *Multiobjective* problems can create difficulties in defining a fitness function that can have conflicting objectives – Pareto optimal
 - Constraints (inequalities, boundaries of topologic surfaces, of n-dimensional boundaries) can be included in GA
 - Other operators has been proposed to improve GA.

Parameter optimizations²

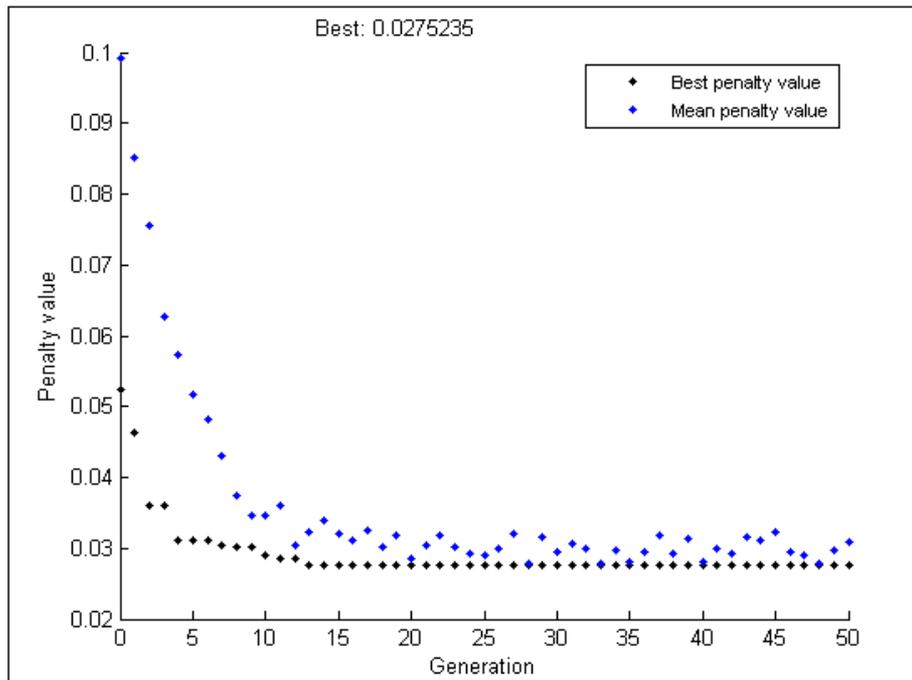
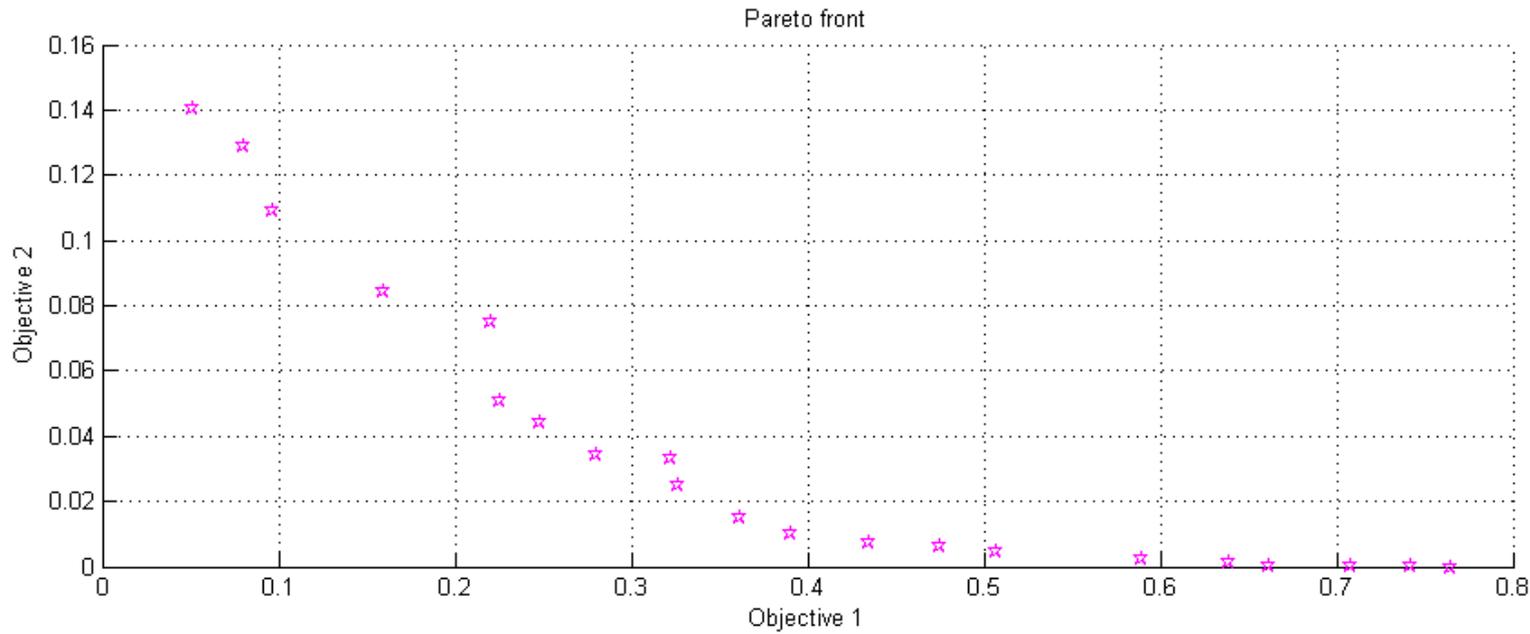
- One of the most important tasks for AF detection is the parameter set for classifier that are usually thresholds involved in rules or other parameters like window length.
- Preprocessing stage (Beat sequence from 8219 MIT AFIB database with ectopic beats and replacement of them)



²D. Arotaritei, C. Rotariu, Parameter Optimizations in Detection of Atrial Fibrillation Based on Heart Rate Interval, BIOMEPE 2015 - Conference on Bio-Medical Instrumentation and related Engineering and Physical Sciences, 2015, Athenes. Greece

-
- The set of parameters are focused on length of window L because an optimal window can produce a relevant value of randomness of time series that can be interpreted as AF
 - A set of parameters can be defined as $P = [L, p_1, p_2, \dots, p_h]$. Optimization of some parameters can be conflictual that is if optimization of objective using one parameter can produce a decrease of optimized value for other parameters. A tradeoff using optimized curve of values can be used to choose a correct tradeoff (e.g. Pareto front).
 - The selected parameters are³: L – length of window, Pth_1 – threshold from where the segment is considered to be AF (number of AF segments/number of total segments), $RMSt$ – threshold of Root Mean Square of Successive Differences, $TPRt$ threshold of Turning Point Ratio, SEt - threshold of Shannon Entropy and TKt - threshold of Teager-Kaiser energy (six parameters).

-
- ❑ An vector of HRV descriptors are good for detection of AF if sensitivity S_e and specificity S_p are both enough close to 1.0 for a correct classification.
 - ❑ Genetic Algorithms proved to be a good choice for optimization problem, single objective or multi-objective.
 - ❑ The parameters are set to be limited by lower values [32, 0.1%, 0.01, 0.01%, 0.01, 0.01] and upper values [480, 99.99%, 0.999, 99.99%, 0.999, 0.999].
 - ❑ The chromosome coded for first value is mapping linearly into integer values between 32 and 480 (logarithmic mapping is a solution for some types of problems).
 - ❑ We start with a population of 50 individuals and after 40 iterations the solution is given in next figure. The multi-objective function is minimization of bot functions $1 - S_e$ and $1 - S_p$. The results are $S_e = 90.17$ and $S_p = 90.12$.

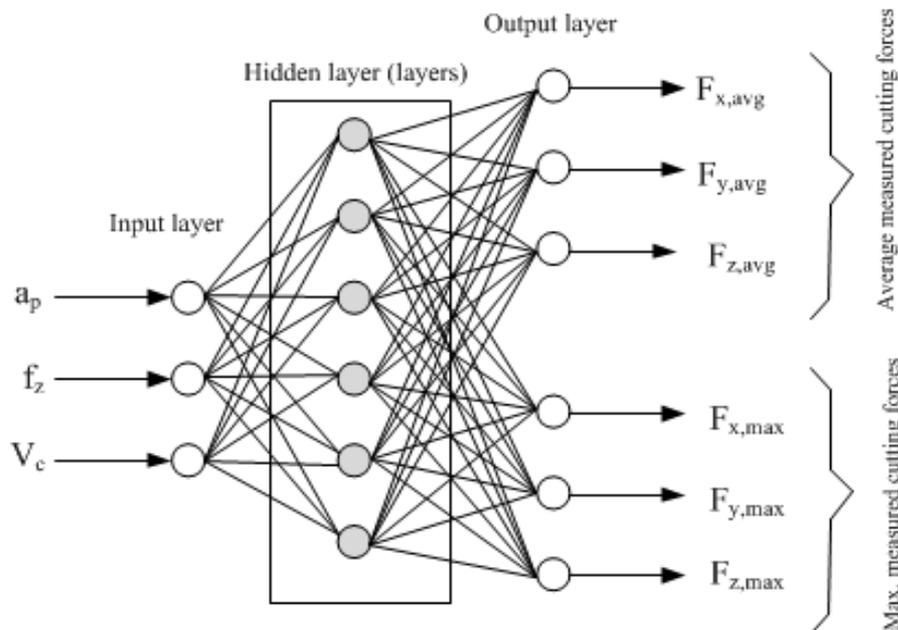


Single objective GA, $f_{fitness} = 1 - \min(S_e, S_p)$
 fusion of two objectives (if we have more than two objectives, the graphic become more difficult).

Artificial Neural Networks (ANN or shortly NN)

- ❑ Neural networks are models inspired by biological neural networks. Main features: neuron transfer function, architecture, neuron connections.
- ❑ Main usage: function approximation, classifier, pattern recognition, prediction.
- ❑ ANN proven to be universal approximators (Haykin S.).
- ❑ In 1989, Cybenko G. demonstrated, using a Kolmogorov's older result that multilayer feed-forward network with a single hidden layer, which contains finite number of hidden neurons, is a universal approximator. The demonstration has been made for sigmoid activation function.
- ❑ In the most common sense, the universal approximator can approximate any non-linear function with a desired precision if its architecture is large enough.
- ❑ Other NNs architectures have been proved to be also universal approximators: radial basis function (RBF) (Park J., and Sandberg I.W.), recurrent neural networks (RNN) (Schäfer A.M., and Zimmermann H.G.) and Kohonen maps.

- Recurrent Neural Networks (RNN) in various architectures and connection among neurons are the subject of dynamic nonlinear modelling and prediction with very good results.
- The most common architecture is feedforward Neural network (multilayer perceptron)



a_p (axial depth of cut [mm]), f_z (feed per tooth [mm/tooth]), and V_c (cutting speed [m/min]). The output layer has six neurons corresponding to average cutting forces and maximum measured forces

The output of neuron in the layer k is given by an activation function that takes into account all the output of neurons of layer $k - 1$, weighted and biased by neuron parameters.

$$x_i^k = \varphi \left(\sum_{j=1}^{n_j} w_{ij}^k x_j^{k-1} + b_i^k \right) = \varphi(w^T x + b)$$

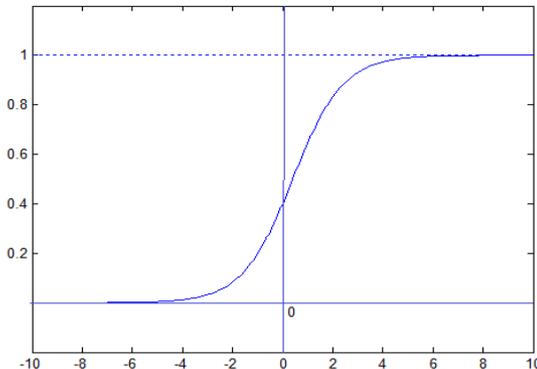
$$\varphi(x) = 1/(1 + e^{-x+b})$$

Model of nonlinear phenomenon using ANN

- Grinding is an essential procedure in orthopedic surgery.
- High-speed miniature grinding wheels are often used to remove material.
- E.g. Stryker CORE U Drill with wheel, Diamond Round Bur type (round shape)



- Tabular data are available from experiments
- The inputs are linearly scaled in range $[-1, +1]$ in order to prevent saturation of sigmoidal functions.



Learning algorithm: backpropagation with clustering data.

Backpropagation – gradient based algorithm

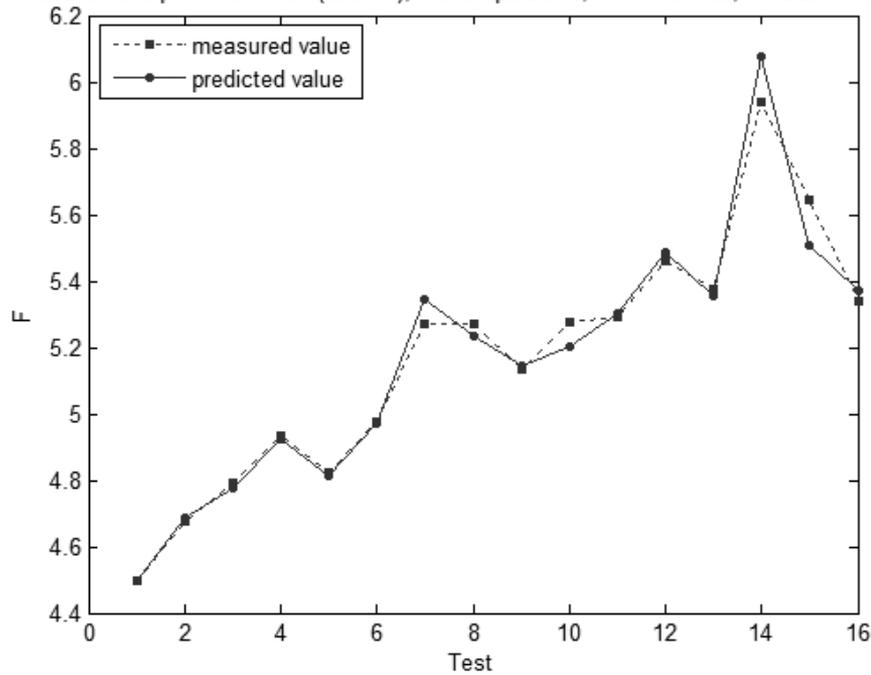
Two stages: feedforward data

backpropagation error

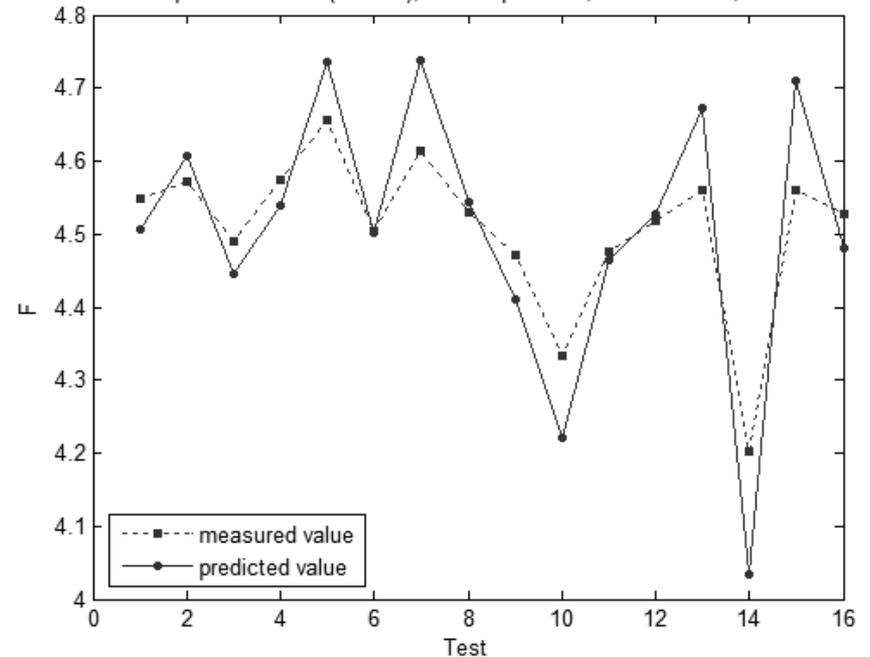
Weights are initially set to random values and

updated in these two stages by $w_{ij}^k(t+1) = w_{ij}^k(t) + \Delta w_{ij}^k(t)$

Measured and predicted force (case A), model quadratic, $R^2 = 0.97485$, RMSE = 0.081444



Measured and predicted force (case B), model quadratic, $R^2 = 0.36367$, RMSE = 0.1079



R^2 - coefficient of determination, proportion of the variance in the dependent variable that is predictable from the independent variable

Conclusions

- ❑ Heuristic algorithms are a very useful solution in many biomedical applications
- ❑ Preprocessing experimental data plays an important role in success of chosen algorithm
- ❑ Feature selection are most often depending on the author's choice
- ❑ If n -dimensional data are for two or more classes are spatially too close, some operation (e.g. orthogonalization) are necessary.
- ❑ Some GA not truly mimic the nature. Some operations and operators are introduced based statistical and mathematical consideration in order to find a acceptable solution for a given problem (minimization and maximization)
- ❑ Genetic algorithms start with a random "gene" sets (chromosomes). In the real world an organism with random genes would not live.
- ❑ The initial population is like a set amino acids form a kind of primordial soup

References

- Nikolaos G. Bourbakis, Artificial Intelligence Methods and Applications, World Scientific, 1992.
- Stuart J. Russell and Peter Norvig, Artificial Intelligence A Modern Approach, Pearson, 2009.
- Mariusz Flasiński, Introduction to Artificial Intelligence, Springer, 2016.
- <http://infolab.stanford.edu/~ullman/mmds/ch11.pdf>
- Zhisong PAN , Zhantao DENG, Yibing WANG, Yanyan ZHANG, Dimensionality reduction via kernel sparse representation, Front. Comput. Sci., 2014, 8(5): 807–815
- Zbigniew Michalewicz Genetic Algorithms + Data Structures = Evolution Programs, Springer, 1996
- David E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Professional, 1989
- Carlos Coello Coello, Gary B. Lamont, David A. van Veldhuizen, Evolutionary Algorithms for Solving Multi-Objective Problems, Springer, 2007.
- Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, Foundations of Machine Learning, The MIT Press, 2012.
- George J. Klir, Bo Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall, 1995
- Simon Haykin, Neural Networks: A Comprehensive Foundation (2nd Edition), Prentice Hall, 1998.
- <https://arxiv.org/ftp/arxiv/papers/1408/1408.0101.pdf>
- <http://www.lcc.uma.es/~ccottap/papers/handbook03memetic.pdf>
- ¹J. Holland, Adaptation in Natural and Artificial Systems, The MIT Press, 1992.
- S. Dash, K.H. Chon, S. Lu, E.A. Raeder, “Automatic Real Time Detection of Atrial Fibrillation”, Annals of Biomedical Engineering, vol. 37, issue 9, pp. 1701-1709, 2009.

Conclusions

- ❑ A gene is a segment of DNA containing the code used to synthesize a protein. A chromosome contains hundreds to thousands of genes.
- ❑ The number of individuals in population at each generation is fixed GA. Some researchers proposed GA schemes using variable size of population during evolution
- ❑ The length of chromosome are the same for all individuals. There are some proposals that use a variable length genotype (the organism itself), SAGA.
- ❑ Algorithms from Artificial Intelligence are not a universal solution for optimized problems - sometime other solution based on non-heuristic models can be more efficient.



Thank you for your attention!

Σας ευχαριστώ για την προσοχή σας