

Artificial Immune System for Medical Data Classification

Wiesław Wajs¹, Piotr Wais², Mariusz Świącicki³,
and Hubert Wojtowicz²

¹ Institute of Automatics,
University of Mining and Metallurgy, Kraków, Poland
wwa@ia.agh.edu.pl

² National Technical University in Krosno
waisp@poczta.onet.pl
hubwoj@pwsz.krosno.pl

³ Institute of Computer Modelling,
Cracow University of Technology, Kraków, Poland
mswiecic@uck.pk.edu.pl

Abstract. The article presents application of artificial immune algorithms in classification of vectorized medical data sets. Artificial immune network was created and trained for the purpose of arterial blood gasometry parameters (pH, pCO₂, pO₂, HCO₃) classification. Training data originates from the Infant Intensive Care Unit of the Polish – American Institute of Pediatrics, Collegium Medicum, Jagiellonian University in Cracow.

1 Introduction

Biological phenomena, and particularly organic processes are of a dynamic character. Diagnosis and therapy are of the same character. In approaching diagnosis and making therapeutic decision phenomena happening in time are investigated. Every doctor knows that it is rare to make a decision about diagnosis and treatment on the basis of only one clinical observation. Usually it is based on several patient's examinations, regular analysis of many biophysical and biochemical parameters, or imaging examination. On a particular level of these examinations preliminary diagnosis is determined, and later a final one. Basing on determined diagnosis and on the evaluation of the sickness process dynamics decision about the treatment is taken.

Initial stabilization of the infants state is a difficult task. To achieve it the doctor analyses repeatedly many parameters related to the patient's health condition. These parameters are: birth anamnesis, physiological examinations (body's weight, diuresis), results of additional examinations (biochemical, micro biological, imaging) and readings from the monitoring instruments (pulseximeter, cardio monitor, invasive measurement of the arterial pressure, respiratory mechanics monitor). It isn't rare when doctor appraises simultaneously over fifty variables. Analysis of this amount of data is hard and requires experience, all the more if the decision about the treatment should be made quickly. In result of carried out analysis doctor makes decision about the treatment expecting positive results, which are expressed by the desired changes

in the results of additional examinations, readings of the monitoring instruments and physical examination. The whole process can be verified by comparing it to the model of respiratory insufficiency progress carried out by the doctor. Creation of this model is based on theoretical and empirical knowledge found in scientific literature and also on the doctor's experience. Intensification of gas exchange disorders are best reflected by arterial blood gasometry parameters examined in the context of currently used methods of ventilation support. For that reason as an output values four directly measured parameters of arterial blood gasometry: pH, pCO₂, pO₂, HCO₃ and oxygen parameter (pO₂/FiO₂) were chosen. Alkalis deficiency and hemoglobin oxygen saturation were omitted as derivatives of the four parameters mentioned above.

2 An Artificial Immune Network for Signal Analysis

Main goal of the system is a classification of signals and to achieve it several problems had to be solved. First of them is connected with an algorithm of learning the immune network. The next problem is related to the structures of data, which are responsible for representation of signals.

The input signal for the system, is interpreted as an antibody (Ab) so the task of immune network is to find an antigen Ag that will be suitable for Ab. The Ag-Ab representation will partially determine which distance measure shall be used to calculate their degree of interaction. Mathematically, the generalized shape of a molecule (m), either an antibody (Ab) or an antigen (Ag), can be represented by a set of real-valued coordinates $m = \langle m_1, m_2, \dots, m_L \rangle$, which can be regarded as a point in an L -dimensional real-valued space.

$$D = \sqrt{\sum_{i=1}^L (ab_i - ag_i)^2} \quad (1)$$

The affinity between an antigen and an antibody is related to their distance that can be estimated via any distance measure between two vectors, for example the Euclidean or the Manhattan distance. If the coordinates of an antibody are given by $\langle ab_1, ab_2, \dots, ab_L \rangle$ and the coordinates of an antigen are given by $\langle ag_1, ag_2, \dots, ag_L \rangle$, then the distance (D) between them is presented in equation (1), which uses real-valued coordinates. The measure distances are called Euclidean shape-spaces.

Algorithm used in the research is described in papers [1] [2] by de Castro and Von Zuben. Modified version of immune net algorithm was adapted to signals classification. The learning algorithm lets building of a set that recognizes and represents the data structural organization. The more specific the antibodies, the less parsimonious the network (low compression rate), whilst the more generalist the antibodies, the more parsimonious the network with relation to the number of antibodies. The suppression threshold controls the specificity level of the antibodies, the clustering accuracy and network plasticity.

2.1 Overview and Analysis of the Research

Currently immune network performs unsupervised pattern recognition of the gathered medical data. In the process of further research description of the data classes will be

added, which is prepared by an expert medicine doctor. The idea is to categorize patients according to the hospitalization group they are initially assigned to by the doctor. Next goal after creating a reliable classification system is to develop a time series predicting system to help doctors in forecasting patient's condition ahead in time.

A major issue in the research is the scarcity of available data from hospitalization. Efficient system of new data acquisition is one of the main concerns and to achieve it in the future a creation of an agent system [3] is planned for gathering information through the Internet from several hospitals and adding it to the database.

3 Multidimensional Medical Data Classification

In the given example use of the AIS is presented in the unsupervised pattern recognition of the medical data. Using a database, functioning for the few years on the Infant Intensive Care Ward of the Polish – American Institute of Pediatrics Collegium Medicum Jagiellonian University of Cracow, a network was created, which final task is going to be the classification of arterial blood gasometry parameters. In the process of training previous values of gasometry, respirator settings and surfactant administration were used as an input data. Training of the network consists of two phases. First phase is a learning of the network. Training data set is comprised of blood parameters values starting from the time (t) – time series are of the same length. Second phase tests the network's generalization abilities by presenting it input vectors from the test dataset. Elements of the training dataset – time series are interpreted by the network as antigens, which are stimulating particular antibody creating an artificial immune network. Training dataset for the artificial immune network comprises of pattern series, which are represented by the vectors. As a result of this time series included in the dataset can be interpreted as a point coordinates in n-dimensional space, which n-value is dependant on the pattern series length.

4 Summary

In the paper functioning of the artificial immune system and example of its application in unsupervised pattern recognition of medical multidimensional data was presented. Perspectives for creating a reliable classification and time series modeling systems based on the artificial immune algorithm were discussed.

References

- [1] De Castro, L. N., Von Zuben, F. J. (2000a), An Evolutionary Immune Network for Data Clustering, Proc. of the IEEE SBRN, pp. 84-89.
- [2] De Castro, L. N., Von Zuben, F. J. (2000b), The Clonal Selection Algorithm with Engineering Applications, GECCO'00 – Workshop Proceedings, pp. 36-37.
- [3] Wajs Wiesław, Wais Piotr, Autonomous Agent for Computer System and Computerized System of Automatics, Some Analogy and Difference, International Workshop Control and Information Technology, IWCIT 2001.