E11-2003-189

I. Antoniou[1,2], V. V. Ivanov[1,3], R. V. Polozov[4],
E. Yarevsky[1,5], P. V. Zrelov[3,5]

METHODS AND ALGORITHMS FOR IDENTIFICATION OF RARE EVENTS

[1]International Solvay Institutes for Physics and Chemistry,
CP-231, ULB, Bd. du Triomphe, B-1050 Brussels, Belgium
[2]Department of Mathematics, Aristoteles University of Thessaloniki,
54006 Thessaloniki, Greece
[3]Laboratory of Information Technologies, Joint Institute
for Nuclear Research, 141980 Dubna, Russia
[4]Institute for Theoretical and Experimental Biophysics,
Moscow Region, 142290 Puschino, Russia
[5]Institute Superiéure de Technologie, 6 rue Coudenhove-Kalergi,
L-1359, Luxembourg

1 Introduction

The problem of time-series analysis aiming at the identification of changes in series dynamics and the prediction of its further development is of great interest in economics, meteorology and many other areas of science and society. Complexity of the problems related to this area gives rise to a variety of mathematical approaches directed to their analysis and solution. However, up to now there have been developed only a few theoretically completed results and effective practical tools for solution of such tasks. The development of realistic mathematical models, which capture the characteristic features of the analyzed complex dynamical systems and processes, plays a very important role in these studies. Such models may serve as a basis for elaboration of mathematical tools which provide effective analysis and control of the object under study.

This paper is devoted to the study of the current state of the time-series data analysis and prediction problem. We discuss a few approaches which can be used for the solution of a pattern recognition problem for events corresponding to various hypotheses.

The paper is organized as follows. In Section 2 we analyse selection algorithms based on statistical good-of-fit criteria. In Section 3 we present identification methods based on feed-forward artificial neural networks including the Chebyshev neural network which has been shown to be very efficient in the time series approximation and prediction. In Section 4 we discuss the basis of the artificial immune systems (AIS) and then briefly present discrimination schemes based on the AIS. As the methods based on the AIS are not very well-known, in Section 5 we consider main elements of the self-nonself selection scheme in some detail. Section 5 consists of concluding remarks and discussion.

2 Selection algorithms based on statistical good-of-fit criteria

The testing of the experimental data correspondence to some theoretical hypotheses is one of the most important part of the data analysis. In order to present the main concept of the hypothesis testing, let us recall some definitions. The hypothesis which can be formulated without any additional assumptions is called a simple hypothesis. The hypothesis which consists of a few simple hypotheses is called a complex hypothesis. In order to present the hypothesis testing, it is enough to consider only the simple hypothesis [1].

Suppose we want to test the hypothesis H_0 (called the null-hypothesis) against the alternative hypothesis H_1 using a set of experimental data. Let X be some function of observables, calling the *test statistics*, and W be the space of all possible values of X. We divide W into *critical* w and *admissible* $(W - w)$ regions so that if the values of function X hit in the region w, then the null-hypothesis is not correct. Thus, the choice of the testing criterion H_0 is reduced to the choice of the testing statistics X and the critical region w.

The size of the admissible region is usually chosen in order to get the prescribed *significance level* α, determined as probability of X to hit into w, when the hypothesis H_0 is valid:

$$P(X \in w|H_0) = \alpha. \tag{1}$$

Therefore, α is the probability that H_0 is rejected while it is correct.

The efficiency of a testing criterion depends on its ability to separate the given hypothesis H_0 from the alternative hypothesis H_1. The measure of usefulness of a criterion is given by a *criterion power*. The criterion power is determined as the probability $1 - \beta$ of X to hit into the critical region when H_1 is correct:

$$P(X \in w|H_1) = 1 - \beta. \tag{2}$$

In other words, β is the probability of X to hit into the admissible region if the alternative hypothesis is correct:

$$P(X \in W - w|H_1) = \beta. \tag{3}$$

There exist two different kinds of errors in the hypothesis testing:

a) of first order error (or *loss*): rejection of the null-hypothesis when it is correct. The probability of such error is equal to α;

b) of second order error (or *admixture*): acceptance of the null-hypothesis when it is not correct. The probability of such an error is equal to β.

The test criteria that check the correspondence of pre-assigned hypothesis (the null-hypothesis H_0) against all possible alternative hypotheses are called the *goodness-of-fit* criteria [1]. Such criteria test experimental data against the density function which corresponds to the hypothesis H_0, in accordance with which the testing data must be distributed.

Motivated by a practical point of view, here we will consider only the criteria which are independent of the form of the testing distribution. The most efficient criteria are based on comparison of the distribution function $F(x)$ corresponding to the null-hypothesis H_0 with the empirical distribution function $S_n(x)$ [1]:

$$S_n(x) = \begin{cases} 0, & \text{if } x < x_1; \\ i/n, & \text{if } x_i \le x \le x_{i+1}, \quad i = 1, \ldots, n-1. \\ 1, & \text{if } x_n \le x, \end{cases} \tag{4}$$

Here $x_1 \leq x_2 \leq \ldots \leq x_n$ is the ordered sample (*variational series*) of the size n constructed on the basis of observations of the variable x.

The testing statistics is a measure of "distance" between the theoretical $F(x)$ and empirical $S_n(x)$ distribution functions. The well-known goodness-of-fit test, the Smirnov-Cramer-Mises criterion (also known as Ω^2-criterion [2, 3]), is based on the statistics

$$\Omega_n^2 = \int_{-\infty}^{\infty} [S_n(x) - F(x)]^2 f(x) dx, \tag{5}$$

where $f(x)$ is the density function corresponding to the null-hypothesis H_0. Such sort of statistics are also known as *non-parametric* statistics.

In paper [4], there have been suggested and investigated a new class of non-parametric statistics

$$\Omega_n^k = n^{k/2} \int_{-\infty}^{\infty} [S_n(x) - F(x)]^k f(x) dx, \tag{6}$$

which generalize the statistics (5). The values of statistics (6) can be calculated with the simple algebraic formula

$$\Omega_n^k = -\frac{n^{\frac{k}{2}}}{k+1} \sum_{i=1}^{n} \left\{ \left[\frac{i-1}{n} - F(x_i) \right]^{k+1} - \left[\frac{i}{n} - F(x_i) \right]^{k+1} \right\}. \tag{7}$$

These statistics have a higher power for the bigger parameter k, and they are more convenient for analysis when the alternative hypothesis has a two-sided form.

As it has been mentioned above, the goodness-of-fit criteria constructed on the basis of statistics (7) are usually applied for the testing of the correspondence of each sample (event) to the distribution known *apriori*.

On the basis of the Ω_n^k criteria, a very efficient procedure has been developed and applied for selection of rare multidimensional events [5, 6, 7]. After minor modifications, this scheme can be used for time-series data processing in order to detect rare (abnormal) events. The modified algorithm has the following steps:

1. The time-series to be analysed is transformed ("normalized") so that the contribution of a dominant distribution (in most cases this distribution concerns with the background process) is described by the distribution function $F_b(x)$.

2. Each sample, composed of values pertaining to the transformed series, is tested with the Ω_n^k goodness-of-fit criterion for correspondence to the $F_b(x)$ hypothesis. In this process the abnormal events, which do not comply with the null-hypothesis, correspond to large absolute values of the Ω_n^k-statistic, resulting in their clustering in the critical region.

3. Events that happen to be in the critical region are further subjected to a second test in accordance with items a) and b). The only difference in the

second test is that now it is the abnormal (signal) events that are collected in the admissible region (using the corresponding distribution function $F_s(x)$). This results in the additional suppression of *background* events in the series under investigation.

The statistical goodness-of-fit criterion discussed above is very efficient in the identification of rare events, because it is powerful and statistically justified. However, to apply this criterion, one has to construct a distribution function corresponding to the analysed process and to determine the size and the preparation procedure for the analysed sample.

3 Identification methods based on feed-forward artificial neural networks

There exist two different approaches for identifying the abnormal events which can be realized on the artificial neural network (ANN) basis.

- The first approach is the classification of individual events represented by empirical samples of finite volumes pertaining to one of the different partial distributions composing the distribution to be analyzed.
- The second approach uses the ability of the ANN to approximate and then to predict the analysed time-series.

A layered *feed-forward* neural network is one of the most convenient tools for constructing the classifier of empirical samples of finite volumes [8, 9, 10] and for approximating the values of unknown real-valued functions [11, 12, 13].

The most-known feed-forward neural network for constructing the classifier of empirical samples of finite volumes is a multi-layer perceptron (MLP). The MLP consists of an input layer corresponding to the data to be processed, an output layer giving the results, and hidden layers. The network scheme is presented in Fig. 1.

In Fig. 1, x_k, h_j and y_i denote the input, hidden and output neurons, respectively. w_{jk} are the weights of the connections between the input neurons and the hidden neurons, and w_{ij} are the weights of the connections between the hidden and the output neurons. The signals $a_j = \sum_k w_{jk}x_k$ and $a_i = \sum_j w_{ij}h_j$ are fed to the inputs of the hidden and output neurons, respectively. The output signals of these neurons are determined as $h_j = g[(a_j+\theta_j)/T]$ and $y_i = g[(a_i+\theta_i)/T]$, where $g(a/T)$ is a transfer function, T is the *"temperature"* determining its slope, and θ is the threshold of the corresponding node. Typically, $g(a/T)$ is a sigmoid, for example, of the form $g(a/T) = \tanh(a/T)$.

The adaptation of the MLP to the problem to be solved is called the neural network *training* or *learning*. The learning in the case of the *back-propagation* algorithm [14] is performed by the minimization of the error functional E with respect

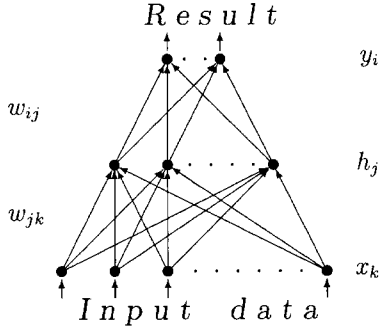

Figure 1: Scheme of the multi-layer perceptron with one hidden layer

to the weights w_{ij}, w_{jk}:

$$E = \frac{1}{2}\sum_p [\vec{y}(x_p) - \vec{f}(x_p)]^2. \tag{8}$$

Here $p = 1, 2, ..., N_{train}$ is the number of training patterns, $\vec{f}(x_p)$ is the target value of the output signal.

The values of the output signals from the hidden and output neurons are functions of the corresponding weights. In order to minimize the error functional, an iterative procedure is used. The changes of the weights w_{ij} and w_{jk} at each iteration are given by:

$$\Delta w_{ij} = -\eta \delta_i g'(a_i) h_j + \alpha \Delta w_{ij}^{old}. \tag{9}$$

and

$$\Delta w_{jk} = -\eta \sum_i w_{ij} \delta_i g'(a_i) g'(a_j) x_k + \alpha \Delta w_{jk}^{old}, \tag{10}$$

where the value δ_i is determined from the equation $\delta_i = y_i - f(x_i)$. In Eqs. (9) and (10), η is the parameter controlling the learning speed [14], $\alpha \Delta w_{ij}^{old}$ and $\alpha \Delta w_{jk}^{old}$ are the moments, which suppress oscillations at the network output. The procedure of the neural network training goes on until an acceptable correspondence between the output signals and the target values is reached. The MLP network is the very efficient tool for classifying events, although its learning speed and power of recognition critically depend on the choice of a method of input data encoding.

A comparative study of classifiers based on the goodness-of-fit criteria Ω_n^k and the MLP were carried out in [15, 16]. It has been shown when only the parameters of the dominant distribution are known, the goodness-of-fit criteria Ω_n^k serve as a suitable tool for the recognition of events corresponding to various distributions. The repeated application of the Ω_n^k-criteria permits extracting of the contributions for any number of partial distributions from the resultant spectrum observed in experiment. Then, if necessary, the neural network can be employed with the estimated

parameters of the constituent distributions. It is worth noting that the Ω_n^k-criterion usage is substantially quantitative, while the results yielded by the ANN are only qualitative.

In Ref. [11] Lapedes and Farber studied the ability of the MLP networks to reconstruct and predict time series, which is an important problem in economics, meteorology and many other areas. Using the time series produced by the logistic map (Fig. 2) as a model, they have demonstrated that the neural networks with sigmoidal neurons in the hidden layer can be used for prediction of highly chaotic time series with an accuracy which is of orders of magnitude better as compared to conventional methods. They have also shown that the neural network performs

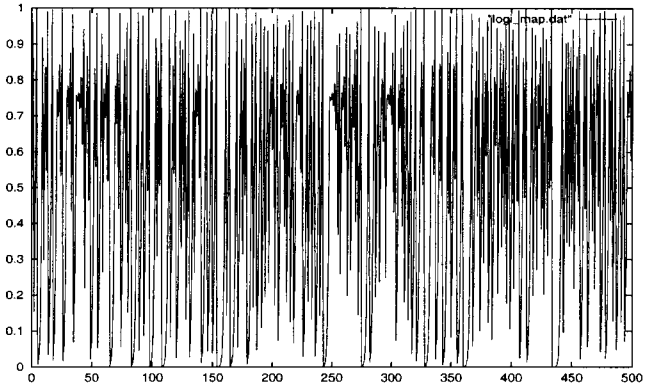

Figure 2: Time series corresponding to the logistic map equation

well because it globally approximates a learning map by performing a kind of a generalized mode decomposition of the map.

In Ref. [17] a new approach was proposed for the approximation of one-dimensional functions based on a Chebyshev Neural Network (CNN). The CNN realizes the expansion of a function in the orthogonal Chebyshev polynomials of the first kind. The expansion coefficients are computed during the network training where arbitrary points (for instance, measured in experiment) from the function domain are used.

The basic concept of the CNN network is presented below. Let us consider the function $f(x)$ defined on a finite set of x values: $f(x_0)$, $f(x_1)$, ..., $f(x_n)$. If the values of $f(x)$ at intermediate x values are required for the solution of a problem, it is convenient to construct an interpolating function $\varphi(x)$. This function should be easy to calculate, and it should approximate $f(x)$ with some degree of accuracy in its domain.

We will look for interpolating functions which can be expanded in terms of the orthogonal Chebyshev polynomials T_n of the first kind:

$$\varphi_n(x) = c_0\,T_0(x) + c_1\,T_1(x) + c_2\,T_2(x) + ... + c_n\,T_n(x). \tag{11}$$

where

$$T_n(x) = \cos(n \arccos x), \qquad |x| \leq 1.$$

For $n = 0$, $T_0(x) = 1$, and for $n = 1$, $T_1(x) = \cos(\arccos x) = x$. The Chebyshev polynomials T_n satisfy the recurrence relation:

$$T_{n+1}(x) = 2\,x\,T_n(x) - T_{n-1}(x).$$

From the orthogonality of the Chebyshev polynomials [20] one obtains

$$c_i = \sum_{k=0}^{n-1} f(x_k)\,T_i(x_k) / \sum_{k=0}^{n-1} T_i^2(x_k), \qquad 0 \leq i \leq n, \tag{12}$$

where

$$x_k = \cos\left\{\frac{(2k+1)\,\pi}{(2n)}\right\}, \qquad 0 \leq k \leq n-1$$

are the roots of $T_n(x)$.

Formula (12) allows one to calculate the expansion coefficients c_k of (11) if the values of the function $f(x)$ at the nodal points x_k are known. However, in practice this is hardly possible.

In order to avoid this difficulty and to calculate the coefficients c_i of expansion (11) with arbitrary set of experimental points, it is possible to implement the feed-forward Neural Network with the architecture represented in Fig. 3.

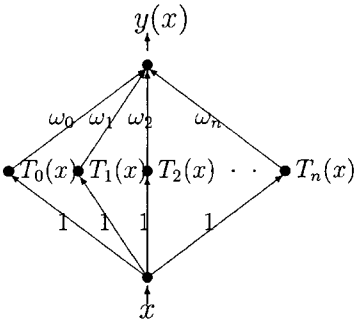

Figure 3: Architecture of the feed-forward Chebyshev neural network

The network has one input neuron, to which the argument x is applied, a single layer of hidden neurons and one output neuron, from which the computed value of function $y(x)$ is obtained.

The argument x is transferred from the input neuron to the neurons of the hidden layer. i-th neuron of the hidden layer transforms the received signal in accordance with the transfer function $g_i(x)$ which is a Chebyshev polynomial:

$$g_i(x) = T_i(x), \qquad \text{where} \quad i = 0, \ldots, n.$$

Then, the sum of weighted signals from the neurons of the hidden layer

$$a = \sum_{i=0}^{n} \omega_i \cdot T_i(x) \tag{13}$$

is transferred to the output neuron, which transforms it according to the function $g(a) = a$.

The CNN network permits realization of expansion (11). When the weights of the connections between the input neuron and the neurons in the hidden layer are all set to 1, the weights ω_i will play the role of the expansion coefficients c_i. The number of neurons in the hidden layer coincides with the number of terms in Eq. (11) and determines the accuracy of the function approximation.

The weights ω_i are calculated during the neural network *training* using the *back-propagation* algorithm. The correction to the weights ω_i at k-th step is given by the following expression

$$\Delta\omega_{i,k} = -\eta\Delta E_{i,k} + \alpha\Delta\omega_{i,k-1},$$

where

$$\Delta E_{i,k} = \sum_{p}[y(x_p) - f(x_p)]T_{i-1}(x_p).$$

This approach permits calculating of the expansion coefficients during the network training process, for which arbitrary points (for instance, measured in experiments) from the function domain are used. The neural network provides the accuracy of the function approximation practically coinciding with the accuracy that can be achieved within the traditional approach, when the values of the function at the nodal points are known.

The more detailed study (based on the logistic map analysis) has shown (see [18]) that, compared with the conventional MLP, the CNN network provides a 50-fold improvement in the approximation. As the new approach provides a better approximation of the time series, one has new possibilities for a long-term prediction.

Figure 4 shows the behaviour of the deviation of the predicted value from the actual value when the MLP network (a) and the CNN network (b) are used for the long-term prediction. One can clearly see that the MLP is reliable only for 3 iterations, while the CNN can go up to 9 iterations.

4 Discrimination schemes based on artificial immune systems

In this section we describe the main features of the natural immune system (NIS) which can be used for information processing. We briefly present the negative selection algorithm while its analysis is given in the next section.

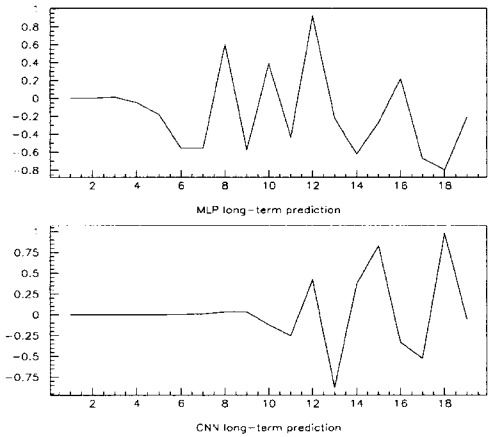

Figure 4: The behavior of the deviation of the predicted value from the actual value when the MLP network (a) and the CNN network (b) are used for the long-term prediction

The main mission of the NIS is to defend a body against pathogenic organisms, cells or molecules [21, 22]. In order to do this, the immune system must perform a pattern recognition in a highly effective way: to distinguish organisms of the body (*"self"*: $\approx 10^6$) from foreign ones (*"nonself"*: $> 10^{16}$) [23].

The NIS is realized as a multi-layered system. At the first level the *skin* protects a body against infection. The second level is physiological, where *physical-chemical conditions* (*pH* and *temperature*) provide inappropriate living conditions for pathogens. Once pathogens have entered the body, they are handled by *the innate immune system* and by the *adaptive immune response*.

The *innate immune system* consists primarily of circulating scavenger cells such as macrophages that ingest extracellular molecules and materials, clearing the system of both debris and pathogens. The *adaptive immune response* is responsible for immunity that is adaptively acquired during the lifetime of the body. It can be viewed as a distributed detection system which consists of white blood cells, called *lymphocytes*. Lymphocytes function as small independent detectors that circulate through the body in the blood and lymph systems. Lymphocytes can be viewed as *negative detectors*, because they detect only non-self patterns: molecular bonds formed between a pathogen and receptors that cover the surface of the lymphocyte.

The ability to detect most pathogens requires a huge diversity of lymphocyte receptors, which can be partly achieved by generating lymphocyte receptors through genetic process that provides a huge amount of randomness. However, even if receptors are randomly generated, there are not enough lymphocytes in the body to

provide a complete coverage of all pathogen patterns. It is estimated that there are 10^8 different lymphocyte receptors in the body at any given time, which must detect potentially 10^{16} different foreign patterns [24, 25].

Protection is made *dynamic* by *continual circulation* of lymphocytes through the body, and by a *continual turnover* of the lymphocyte population. Lymphocytes are typically short-lived (a few days) and are continually replaced by new lymphocytes with new randomly generated receptors.

Protection is made more specific by *learning* and *memory*. If the NIS detects a pathogen that has not been encountered before, it undergoes a primary response, during which it "learns" the structure of the specific pathogen, i.e. it evolves a set of lymphocytes with high affinity for that pathogen, through a process called *affinity maturation*. This produces a large number of lymphocytes that have high affinity for a particular pathogen, which accelerates its detection and elimination.

To summarize, the NIS has many features that are desirable from a computer science standpoint. The system is massively parallel and its functioning is truly distributed. Individual components are disposable and unreliable, yet the system as a whole is robust (reliable). Previously encountered infections are detected and eliminated quickly, while novel intrusions are detected on a slower time scale, using a variety of adaptive mechanisms. The system is autonomous, controlling its own behaviour both at the detector and effector levels. Each immune system detects infections in slightly different ways, so pathogens that are able to evade the defenses of one immune system cannot necessarily evade those of every other immune system.

The natural immune system is a subject of great research interest because of its powerful information processing capabilities [26]. The key features of the immune system which are important for the field of information processing may be summarized as follows: Recognition, Feature extraction, Diversity, Learning, Memory, Distributed detection, Self-regulation, Threshold mechanism, Co-stimulation, Dynamic protection, Probabilistic detection. Other related features like adaptability, specificity, self-tolerance, differentiation etc. also perform important functions in immune response. All these remarkable information-processing properties of the immune system reflect important aspects in the field of computation.

The rapidly emerging field called Artificial Immune Systems (AIS) (also called *Immunological Computation*) explores different immunological mechanisms and their relation to information processing and problem solving [26]. So far, Artificial Immune Systems have received very little attention as compared to other techniques based on biological metaphors, such as neural networks and evolutionary algorithms [27].

S.Forrest et. al. [28] developed a simple and very effective computer *negative-selection algorithm* for change detection based on the principles of self-non-self discrimination in the immune system.

This algorithm can be briefly described in following steps:

- Define *self* as a collection S of strings of the length l over a finite alphabet, a collection that needs to be protected or monitored. For example, S may

be normal pattern (program, data file) of activity, which is segmented into equal-sized substrings[1].

- Generate a set R of *detectors*, each of which fails to match any string in S. Instead of exact or perfect matching[2], the method uses a *partial matching rule*, in which two strings match if and only if they are identical for at least r contiguous positions, where r is a suitable chosen parameter (as described in [28]).

- Monitor S for changes by continually matching the detectors in R against S. If any detector ever matches, then a change is known to have occurred, because the detectors are designed to not match any of the original strings in S.

This algorithm relies on three important points: (1) each copy of the detection algorithm is unique, (2) detection is probabilistic, and (3) a robust system should detect (probabilistically) any foreign activity rather than looking for specific known patterns of changes. Further studies [29, 30] show many insights of the algorithm. The algorithm seems to have many potential applications in change-detection, some of them are discussed below.

Based on this algorithm, S.Forrest and her group at the University of New Mexico started to work on a research project with a long-term goal to build an artificial immune system for a computer [28, 31, 32]. Their computer immune system has to protect a computer against unauthorized use of computer facilities, maintain the integrity of data files, and prevent the spread of computer viruses. Their first results had shown feasibility and perspectives of this new immunological approach to anomaly detection for the networked and distributed computing environment.

Dasgupta and Forrest [33] experimented with several time series data sets (both real and simulated) to investigate the performance of the negative selection algorithm [28] for detecting anomaly in the data sets. The objective of this work is to develop an efficient detection algorithm that can be used for revealing any changes in steady-state characteristics of a system or a process. In these experiments, the notion of self is considered as the normal behaviour patterns of the monitored system[3]. So, any deviation that exceeds an allowable variation in the observed data, is considered as an anomaly in the behaviour pattern. This approach relies on sufficient enough sample of normal data (that can capture the semantics of the data

[1]This is analogous to the way, proteins are broken up by the immune system into smaller subunits, called peptides, to recognize by T-cell receptors [21].

[2]For strings of any significant length a perfect matching is highly improbable, so a partial matching rule is used which rewards more specific matches (i.e., matches on more bits) over less specific ones. This partial matching rule reflects the fact that the immune system's recognition capabilities need to be fairly specific in order to avoid confusing self molecules with foreign molecules.

[3]It is assumed that the normal behaviour of a system or a process can often be characterized by a series of observations over time. Also the normal system behaviour generally exhibit stable patterns when observed over a time period.

patterns) to generate a diverse set of detectors, that probabilistically detect changes without requiring prior knowledge of anomaly (or faulty) patterns.

They applied the algorithm for "The Tool Breaking Detection" in a milling operation [33]. The tool breakage detection problem is formulated as a problem of detecting temporal changes in the cutting force pattern that results from a broken cutter. That is, the new data patterns are monitored to check whether or not the current pattern is different from the established normal pattern, where a difference (i.e. a match in the complement space) implies a change in the cutting force dynamics.

This detection algorithm was successful in detecting the existence of broken teeth from simulated cutting force signals in a milling process. The results suggest that the approach can be used as a tool for the automated monitoring of safety-critical operations.

5 Analysis of main elements of the self-nonself selection scheme

In this section we consider the approach to the anomaly detection problem based on the negative-selection algorithm. It is reduced to the problem of detecting whether or not an analysed pattern, represented as a string, implies a change in the normal behaviour patterns. Hence, we will analyse the main elements of self-nonself selection scheme:

- encoding algorithm,
- detector set generation,
- matching rules and estimation schemes,

together with their limitations and connections with the approaches discussed in sections 2 and 3.

5.1 Encoding time series data

The pre-processing of raw time series data can be considered as constructing an alternative representation of the data while preserving the information content. Furthermore, any change that exceeds allowable variation in the data pattern should ideally be reflected in the representative space. This can be a problem when very small changes in real-valued data need to be monitored. To handle this, an encoding method is used that maps real-valued data into a discrete form. An analogue value is first normalized with respect to a defined fixed range to determine the interval in which it belongs, and then the interval is encoded into binary form. However, if the value falls outside range (MIN, MAX), it will encode to all 0's or all to 1's depending on which side of the range it crossed. Accordingly, if each data item is encoded by

m binary digits (which may be chosen according to the desired precision), then there would be $2^m - 2$ different intervals between the maximum (MAX) and minimum (MIN) ranges of data. Thus, an analogue value x, $MIN \leq x \leq MAX$, corresponds to the binary string representing n_a where $MIN + n_a \cdot d \leq x \leq MIN + (n_a + 1) \cdot d$. Here the interval size $d = (MAX - MIN)/(2^m - 2)$, and n_a can vary from 1 to $2^m - 2$.

Evidently, this procedure cannot be used as a universal algorithm for raw data pre-processing. It corresponds to one particular choice of the function S_n (4), namely $x_{i+1} - x_i = \text{const}$. At the same time, the way how the original data are transformed is crucial for the efficiency of any approach as we have already mentioned in section 3.

5.2 Detector set generation

In the general description of the algorithm [28], the candidate detectors are generated *randomly* and then tested to see if they match any self string. If a match is found, the candidate is rejected. This process is repeated until a desired number of detectors are generated. A probabilistic analysis is used to estimate the number of detectors that are required to provide a certain level of reliability. The major limitation of the random generation approach appears to be computational difficulty of generating valid detectors, which grows exponentially with the size of self. Also for many choices of the length l and the matching parameter r, and compositions of self, the random generation of strings for detectors may be prohibitive.

A more efficient detector generation algorithm has been proposed by Helman and Forrest [34]. This algorithm runs in linear time with the size of self. Other methods for generating nonself detectors, with varying degrees of computational complexity, have also been suggested [30].

These studies have demonstrated some principal problems related to the generation of the effective detectors set. These problems concern the size of a representative set of detectors and the so-called holes in the detector set [32, 35].

It must be mentioned that the procedure of the detector set generation is equivalent in some sense both to the construction of the distribution function $F_b(x)$ corresponding to the null-hypothesis testing (section 2) and to the ANN training on a historical data set (section 3). So, the data encoding method, which has been applied to raw data, plays a very important role for determining what approach is the most adequate one for the solution of a problem.

5.3 Matching rules and estimation schemes

In order to explain the main idea behind matching rules, we will discuss a partial matching rule based on a pre-specified degree of similarity [33]. In order to measure this similarity, we use an r contiguous matching rule between two strings of equal

length. Thus, for any two strings x and y, $match(x, y)$ is true if x and y agree (match) on at least r contiguous locations ($r \leq l$), as illustrated in Fig. 5.

X: bcabcbad
Y: dcabdcba

Figure 5: Illustration of the matching rule: x and y are two strings defined over the four-letter alphabet a, b, c, d. X and Y match at three contiguous locations (underlined). Thus, $match(x, y)$ is true for $r \leq 3$ and false for $r > 3$.

The partial matching rule provides a detector with a capability of detecting sample strings in its neighbourhood according to the threshold value r. This is demonstrated in Fig. 6 for the binary string. Fig. 6 shows that the coverage of

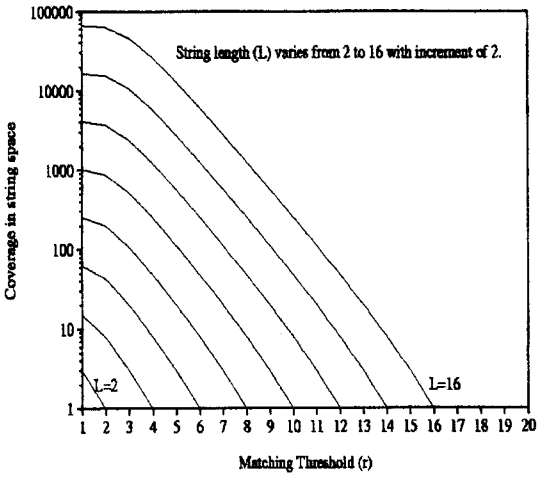

Figure 6: The dependence (in log scale) of the number of points that can be covered by each binary string (of the defined length in its string space) against the different matching threshold r

strings of a fixed length increases exponentially with the decrease of r. Though the maximum coverage can be achieved with $r = 1$, the generated detectors will probably be matched with many self strings resulting in false detection. On the other hand, a perfect matching (for $r = l$) implies that symbols are identical at each location in two strings. Thus, a very large number of detectors is needed to detect patterns in the non-self space. An optimal r value estimates a reasonable size detector set for the success of this method.

When a non-overlapping set of detectors is generated with a suitable matching threshold, each detector can serve as a distinct novelty pattern class in the non-self space. However, in case of overlapping detectors, multiple detectors may be

activated for a sample (abnormal) pattern, and one needs more detectors to provide sufficient coverage in the non-self space.

Thus, one can see that the procedure of non-self object recognition, as well as the decision making function are not yet well-elaborated. This procedure has rather qualitative than quantitative character and, in this sense, it is close to the pattern recognition procedure based on the ANN approach. Therefore, the most developed and statistically justified approach is the approach based on the goodness-of-fit criteria basis (section 2) where the matching is realized on the Ω_n^k-statistics and the threshold is defined by a critical limit corresponding to a chosen confidence level α.

6 Conclusions

In this section we shortly compare the methods discussed above and present some directions for future investigations.

The neural network approach (in particular, ART networks) have also been successfully applied to the problem of detecting a tool breakage in milling operations [36, 37, 38, 39]. Preliminary results of both ANN and AIS approaches qualitatively coincide, though there are some important similarities and differences between them. However, a careful quantitative comparison of these two approaches is an important area of further research.

As the CNN network gives an approximation which is linear with respect to the expansion coefficients, other methods like the least square approximation can be used in the same situation. The comparison of those methods with the CNN can give a deeper insight into the accuracy, stability and convergence of the approximation and reveal their strengths and weaknesses. It may lead to elaboration of more effective and accurate methods for the series prediction.

The comparative analysis of a powerful procedure developed in [4, 5, 6] on the Ω_n^k basis for rare events selection (section 2) and the negative-selection algorithm based on immunological principles (section 4) shows that both approaches are intended for the solution of the same problem, namely, the identification of rare abnormal patterns (events). Moreover, taking into account the similarities discussed above, one can find that main ideas of both schemes are very close. In addition, the Ω_n^k-criterion *collects* all patterns that do not confirm the null-hypothesis (these are so-called "non-self" patterns in the immunological sense) into the critical region corresponding to the chosen significance level α.

Among the presented approaches, the AIS-based methods are rather new and therefore not so well developed. Hence, there are big research perspectives in this field. For example, searching for efficient techniques for preparation of "self" sets, generation algorithms and parametric models is very important. However, the accurate comparison of these methods with other well-established methods is absolutely necessary to be able to demonstrate the advantages of the AIS and to find appropriate applications.

Acknowledgments

We are grateful to Prof. I. Prigogine and Prof. V. G. Kadyshevsky for encouragement and support.

This work has been supported in part by the European Commission in the frame of the Information Society Technologies (IST) program, the IMCOMP project (IST-2000-26016) and by the Luxembourg Ministry of Culture, High Education and Research under Grant BFR01/069.

References

[1] W.T. Eadie, D. Dryard, F.E. James, M. Roos and B. Sadoulet: *Statistical Methods in Experimental Physics*, North-Holland Pub.Comp., Amsterdam-London, 1971.

[2] H. Cramer: *Mathematical Methods of Statistics*, University of Stockholm, 1946.

[3] G.V. Martinov: *Omega-squared criteria*, Moscow, "Nauka", 1978 (in Russian).

[4] V.V. Ivanov and P.V. Zrelov, Int. J. Comput. & Math. with Appl., vol. **34**, No. 7/8, (1997)703-726; JINR Communication P10-92-461, 1992 (in Russian).

[5] P.V. Zrelov and V.V. Ivanov: *The Small Probability Events Separation Method Based on the Smirnov-Gramer-Mises Goodness-of-Fit Criterion.* In: "Algorithms and Programs for Solution of Some Problems in Physics". Sixth Volume. Preprint KFKI-1989-62/M, Budapest, Hungary. 1989, p.127-142.

[6] P.V. Zrelov and V.V. Ivanov: *The Relativistic Charged Particles Identification Method Based on the Goodness-of-Fit ω_n^3-Criterion.* Nucl. Instr. and Meth. in Phys. Res., **A310** (1991) 623-630.

[7] P.V. Zrelov, V.V. Ivanov, V.I. Komarov, A.I. Puzynin, A.S. Khrykin: *"Simulation of Experiment on the Investigation of the Processes of the Subtheshold K^+ Production"*. JINR Preprint, P10-92-369, Dubna, 1992; *"Mathematical Modeling"*, v.**4**, No.**11**, 1993, c.56-74 (in Russian).

[8] S. Haykin, "Neural Networks: A Comprehensive Foundation", Prentice-Hall, Inc., 1999.

[9] C. Peterson and Th. Rögnvaldsson, in Proc. II Int. Workshop on *"Software Engineering, Artificial Intelligence and Expert Systems in High Energy Physics"*. New Comp. Tech. in Phys. Res. II, edited by D. Perret-Gallix, World Scientific, (1992)113.

[10] B. Denby, in Proc. II Int. Workshop on *"Software Engineering, Artificial Intelligence and Expert Systems in High Energy Physics"*. New Comp. Tech. in Phys. Res. II, edited by D. Perret-Gallix, World Scientific, (1992)287.

[11] A. Lapedes and R. Farber: *"Nonlinear Signal Processing using Neural Networks: Prediction and System Modeling"*, Los Alamos Report LA-UR 87-2662 (1987).

[12] R.D. Jones, Y.C. Lee, C.W. Barnes, G.W. Flake, K. Lee, P.S. Levis and S. Quin: *"Function Approximation and Time Series Prediction with Neural Networks"*, Los Alamos Report LA-UR 90-21 (1990).

[13] D.T. Pham and L. Xing, "Neural Networks for Identification, Prediction and Control", Springer-Verlag Berlin (1995).

[14] D.E. Rumelhart, G.E. Hinton, R.J. Williams: *"Learning Internal Representations by Error Propagation"* in D.E. Rumelhart and J.L. McClelland (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations. MIT Press (1986).

[15] V.V. Ivanov, in Proc. IV Int. Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, April 3-8, 1995, Pisa, Italy; "New Computing Techniques in Physics Research IV", edited by B. Denby & D. Perret-Galix, "World Scientific", (1995)765.

[16] A.Yu. Bonushkina et al, Int.J.Comput. & Math. with Appl., vol. **34**, No. 7/8, (1997)677-685.

[17] V. Basios, A.Yu. Bonushkina, V.V. Ivanov: *"On a Method for Approximating One-Dimensional Functions"*, Int. J. Comput. & Math. with Appl., Vol. 34, No 7/8, pp. 687 - 693, 1997.

[18] I. Antoniou, P. Akritas and V.V. Ivanov: *Identification and Prediction of Discrete Chaotic Maps Applying a Chebyshev Neural Network*, "Chaos, Solitons and Fractals", 11 (2000) 337-344.

[19] E.A. Jackson, Perspectives of Nonlinear Dynamics, Cambridge University Press, 1989.

[20] I.S. Berezin and N.P. Zhidkov, *"Computing methods"*, Vol. I, (Translated by O.M. Blunn), Pergammon Press, 1965; this is a translation of the original Russian "*Metody Vychislenii"* published by Fizmatgiz, Moscow, 1959.

[21] N.K. Jerne: *"The Immune System"*, Scientific American, **229(1)**: 52-60, 1973.

[22] R.A. Goldsby, T.J. Kindt and B.A. Osborne: *Kuby Immunology*, 4th ed., W.H. Freeman and Company, (2000).

[23] J.K. Inman, *The antibody combining region: Speculations on the hypothesis of general multispecificity*, Theoretical Immunology, 1978.

[24] S. Tonegawa, *Somatic generation of antibody diversity*, Nature, **302**: 575–581, 1983.

[25] D.G. Osmond, *The turn-over of B-cell populations*, Immunology Today, **14(1)**: 34–37, 1993.

[26] D. Dasgupta: *"An Overview of Artificial Immune Systems and Their Applications*, In: Artificial Immune Systems and Their Applications, Springer-Verlag Berlin Heidelberg 1999, 3-21, 1999.

[27] D. Dasgupta and Nii Attoh-Okine: *Immunity-Based Systems: A Survey*, In: Proc. of the IEEE Int. Conf. on Systems, Man, and Cybernetics, Orlando, October 12-15, 1997.

[28] S. Forrest, A.S. Perelson, L. Allen, and R. Cherukuri: *Self-Nonself Discrimination in a Computer.* In: Proc. of IEEE Symposium on Research in Security and Pravicy, pp. 202-212, Oakland, CA, 16-18 May 1994.

[29] P. D'haeseleer: *"An immunological approach to change detection: theoretical results"*, In Proc. of IEEE Symposium on Research in Security and Privacy, Oakland, CA, May 1996.

[30] P. D'haeseleer, S. Forrest and P. Helman: *"An immunological approach to change detection: algorithms, analysis and implications"*, In Proc. of IEEE Symposium on Research in Security and Privacy, Oakland, CA, May 1996.

[31] S. Forrest, S. Hofmeyr, A. Somayaji, and T. Longstaff: *A sense of self for UNIX processes.* In: Proc. of the 1996 IEEE Symposium on Computer Security and Pravicy, IEEE Press, 1996.

[32] S. Hofmeyr, and S. Forrest: *Immunizing Computer Networks: Getting All the Machines in Your Network to Fight the Hacker Disease.* In: Proc. of the 1999 IEEE Symposium on Computer Security and Pravicy, IEEE Press, 1999.

[33] D. Dasgupta and S. Forrest: *"Tool Breakage Detection in Milling Operations using a Negative-Selection Algorithm"*, Technical Report CS95-5, Department of Computer Science, University of New Mexico, 1995.

[34] P. Helman and S. Forrest: *"An Efficient Algorithm for Generating Random Antibody Strings"*, Technical Report No. CS94-7, Department of Computer Science, University of New Mexico, 1994.

[35] S. Hofmeyr: *"An Immunological Model of Distributed Detection and Its Application to Computer Security"*, PhD Thesis, 1999.

[36] Cihan H. Dagli and P. Poshyanonda: *"Artificial Neural Networks for Intelligent Manufactoring"*, Chapter 3, p. 57, Chapman & Hall, 1994.

[37] S. Rangwala and D. Dornfeld: *"Sensor Integration Using Neural Networks for Intelligent Tool Condition Monitoring"*, Journal of Engineering for Industry, 112:219-228, August 1990.

[38] I.N. Tansel and C. McLaughlin: *"Detection of tool breakage in milling operations-I. The time series analysis approach"*, International Journal of Machine Tools & Manufacturing, 33(4):531-544, 1993.

[39] I.N. Tansel and C. McLaughlin: *"Detection of tool breakage in milling operations-I. The neural network approach"*, International Journal of Machine Tools & Manufacturing, 33(4):545-558, 1993.

Received on October 8, 2003.

Антониоу Я. и др.
E11-2003-189

Методы и алгоритмы для идентификации редких событий

Рассмотрено несколько различных подходов для идентификации редких событий, нестандартного поведения динамических систем и изменений в динамике временных рядов. Проанализированы алгоритмы отбора на основе статистических критериев согласия, методы распознавания на основе прямоточных искусственных нейронных сетей и схемы дискриминации на основе искусственных иммунных систем. Обсуждаются сильные и слабые стороны различных подходов и дается их сравнительный анализ.

Работа выполнена в Лаборатории информационных технологий ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2003

Antoniou I. et al.
E11-2003-189

Methods and Algorithms for Identification of Rare Events

Several different approaches to the identification of rare events, abnormal behaviour of dynamical systems and changes in series dynamics are considered. Selection algorithms based on the statistical good-of-fit criteria, identification methods based on the feed-forward artificial neural networks and discrimination schemes based on the artificial immune systems are analyzed. We discuss strengths and weaknesses of the methods and give their preliminary comparison.

The investigation has been performed at the Laboratory of Information Technologies, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2003

Макет *Т. Е. Попеко*

Подписано в печать 11.11.2003.
Формат 60 × 90/16. Бумага офсетная. Печать офсетная.
Усл. печ. л. 1,43. Уч.-изд. л. 2,22. Тираж 310 экз. Заказ № 54174.

Издательский отдел Объединенного института ядерных исследований
141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.
E-mail: publish@pds.jinr.ru
www.jinr.ru/publish/