IMMUNITY-BASED NEURAL NETWORKS TO MACHINE LEARNING FOR COMPLEX PREDICTION PROBLEMS

ANDRE A. KELLER [1]

[1] *Research Unit for Statistics, Analysis, and Multidisciplinary Modeling, University Paris 1 Panthéon-Sorbonne, Paris, France*

Abstract: The complexity of the phenomena observed and the massive quantity of recorded high-dimensional data (such as with climate patterns) justified the use of automatic data-driven exploratory techniques. Machine learning allows us to analyze and predict such volumes of complex data. Machine learning is widely used in astronomy, bioinformatics, medicine, meteorology. The search for ever greater efficiency of the necessary calculations combined with advances in computing led researchers to design hybrid systems using more efficient approaches in certain features and functions of machine learning. Zhang et al. [Int. J. Forecast. 14 (1998) 35-62] described artificial neural networks as 'data-driven self-adaptive methods' with few assumptions about the pattern. Machine learning proved to have faster performances with metaheuristic algorithms like artificial immune systems (AIS). An immune model has the characteristics of a complex network whose nodes interact dynamically. The immune principles consist of mechanisms for defense against external threats (foreign antigens). This paper introduces the main concepts of such hybrid systems. Two types of complex applications from the literature illustrate the scope of such an approach. The first application is to solve a multi-objective optimization problem. The second application relates to the prediction problems of landslide displacements.

Keywords: Machine Learning, Neural Network, Artificial Neural Network, Immune Principles, Artificial Immune System, Multi-objective Optimization Problem, Prediction Problem, Landslide Displacement Prediction

1 INTRODUCTION

The requirement for in-depth and automated observation of complex phenomena resulted in a massive and varied set of data to be exploited. To this set was added mechanically other generated data accompanying and resulting from a multitude of connected objects (e.g., smartphones, cameras, personal computers, wireless sensors networks). Various fields of activity such as biology, astrophysics, finance, and marketing gather such large and complex datasets. Gudivada [Gudivada et al., 2015] gave notably that example of NASA's solar dynamics observatory (SDO) which captured its 100 millionth image on January 2015. All in all, the result is an accelerated accumulation of all kinds of databases with high-dimensional, heterogeneous and complex information.

The attractiveness of this cheaper data compared to traditional human surveys inevitably questioned the specialists on the capacities and the performances of the existing means of calculation. The task of imagining an immense and powerful system was all the more difficult because it was necessary to consider a long chain of processing including generation of data, acquisition of these data, storage and statistical exploitation. This exploitation, using traditional and more advanced methods, should allow the understanding of a complex domain, and if possible a projection in evolution. Most traditional data mining algorithms use sequential and centralized calculations [Fayyad et al,1996]. However, some algorithms allow a parallel computation distributed on several machines that increases the speed of computations. These new processing methods, in addition to the advances in computer science, will not be enough to counter the new challenges of large databases. Renewal will, therefore, go through a paradigm shift. That of machine learning will prevail [Cantú-Paz, 1998] [Cantú-Paz and Goldberg, 2000] [Luo and El-Baz, 2018].

The recent definition of 'big data' comes in four words starting with the letter 'V' (The so-called 4 Vs). These words that characterize these big datasets are data 'Volume', 'Velocity,' 'Veracity,' and 'Value' [Gudivada et al., 2015]. The volume of the current data is measured in terabytes turns in petabytes according to the observers. The major discrepancies in data quality are attributable to their possible contradictions, incomplete character, imprecision, subjective nature, redundancy, bias and noise [Gudivada et al., 2015]. The big data operators are thus confronted with trade-offs among desired scalability, availability, performance and security.

Examples of big databases can be found in the field of security. Acharjya and Kause [Acharjya and Kause, 2016] reviewed big data analytics. In this field, the data to be used are monitored in real-time using video detection from surveillance cameras. Sunny [Sunny, 2017] reviewed different methods for fire detection [Foggia et al, 2015]. The methods focus on spatio-temporal features and specific properties of fire like color, shape variation, motion. They are part of a multi-expert system where this information is combined. The fire detection process consists of data generation, feature extraction, and classification according to a naïve Bayesian classifier (i.e., a probabilistic machine learning algorithm) [Borges and Izquierdo, 2010].

Data analytics methods have different objectives than

traditional analyzes. Trends in big data analytics were studied by Kambatla et al. [Kambatla et al, 2014] and Ma et al. [viewed]. The goals of big data analytics include mainly dimensionality reduction and discovering hidden features. Dimensionality reduction (DR) speeds up data processing. DR can be obtained by using linear methods like principal component analysis (PCA), linear discriminant analysis, or independent component analysis (ICA) [Hyvärinen, 2013]. The efficient methods of data mining are reported by Tsai et al. [Tsai et al, 2015]. The methods include sampling and data condensation, density-based and grid-based approaches, divide-and-conquer, incremental learning, etc. The process of knowledge discovery (KDD) in databases consists of the following operations: gathering, selection, pre-processing (e.g., detecting, cleaning, filtering, etc), transformation (reducing the data complexities, sampling, coding), data mining, and evaluation [Tsai et al, 2015]. Representative algorithms for data mining problems are clustering and classification [Shalev-Shwarz and Ben-David, 2012].

This paper introduces for didactic purposes the paradigms and algorithms of machine learning and related approaches. The paper attempts to show the ability of these methods to deal with complex problems. *Section 2* deals with the concepts and techniques of machine learning for feature selection and classification. *Section 3* addresses an algorithm inspired by the principles of the human immune system. This section is on immune system prin ciples, implementation of the algorithm, with application to classification problems. *Section 4* is devoted to the study and prediction of complex phenomena of real life. The application to illustrate this issue is the prediction of landslides with important material and human issues.

2 MACHINE LEARNING TECHNIQUES FOR FEATURE SELECTION AND DATA CLASSIFICATION

The adoption of innovative business practices has in theory been linked to an opportunity for increasing the total value in a given market [Bakos, 1998]. Scott-Morton [Scott-Morton, 1991] explored the influence of information technology in the 1990s. Business Opportunities and Technology Advances are two forces that could affect the structure of a market. The combined effect of these forces will drive value chain transformations, thus causing the emergence of new business models and market structures.

2.1 Statistical Machine Learning

The instruments of statistical learning to understand data and modeling are of two kinds of supervised and unsupervised learning [James et al, 2014].

James et al. described supervised learning as a process of constructing a statistical model involving estimating and predicting an output from one or more inputs. The main classical methods in this category are regression and classification. The distinction between regression and classification is due to the quantitative (numerical value) or qualitative (value in one category) nature of the variables used. Indeed, to predict a qualitative response for an observation is equivalent to classifying this observation by attributing it one category.Thus, linear regression of least squares is used for a quantitative response, while logistic regression can correspond to a binary qualitative response (in two categories) [James et al, 2014, pp.28-20]. Other classifiers are linear discriminant analysis and K-nearest neighbors.

The goal is to predict a response y (an output) measured on n observations by using a set of p associated features (inputs, attributes or predictors) $X=\left(x_1,x_2,\ldots,x_p\right)$ measured on the same number of observations. Formally, we may write the regression model as $\hat{y}=\hat{f}\left(x_1,\ldots,x_p\right)$ where $\hat{f}$ denotes an estimate for f and $\hat{y}$ the prediction for y. For a multiple linear regression, regressing y on X, we write the linear relationship as the approximate model $y\approx\beta_0+\beta_1x_1+\cdots+\beta_px_p$. The logistic model formalizes the relationship between the conditional probability $p(X)=\Pr\left(y=k\middle|X=x_0\right)$ and the p predictors $X=\left(x_1,\ldots,x_p\right)$ either by retaining the following logit or log-odds formulation

$$\log\left(\frac{p(X)}{1-p(X)}\right)=\beta_0+\beta_1x_1+\cdots+\beta_px_p$$

A naïve Bayes classifier is a probabilistic machine learning. It can be used in a variety of classification tasks supposing that the predictors (or features) are independent. Suppose m categories (or classes) $C_1,C_2,\ldots,C_m$. By Bayes' theorem, the posterior probability (i.e, the conditional distribution of y given X for category C_k) is

$$p\left(C_k\middle|X\right)=\frac{p\left(X\middle|C_k\right)p\left(C_k\right)}{p(X)}$$

where the probability of the predictors X is given by $p(X)=p\left(x_1\right)\times\cdots\times p\left(x_p\right)$. The posterior probability $p\left(X\middle|C_k\right)$ of X conditioned on C_k is based on more information than the prior probability $p\left(C_k\right)$. The class for which $p\left(C_k\middle|X\right)$ is maximized is the maximum posteriori hypothesis [Han et al, 2012, pp. 350-355].

The K-nearest neighbors (KNN) applies when the required conditional distribution required for calculation is not known and must be estimated. Other classifiers like SVMs (Support Vector Machine) and feedforward NNs (Neural Networks) are used to classify images [Lemley et al, 2016].

2.2 Feature Extraction Techniques

Lemley et al [Lemley et al, 2016] enumerated some feature extraction methods. These exploratory data techniques are notably Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Histogram of Oriented Gradient (HOG). This introductive presentation is restricted to both PCA and ICA methods.

a) PCA method consists of reducing a large set of characteristic features observed to a smaller set. PCA method is a statistical method that finds correlations between data features. This method involves unsupervised approaches since it only involves the features $X_1, X_2, \ldots, X_p$ without associating a response Y.

The 1st principal component of a set of features $X_1, X_2, \ldots, X_p$ is the normalized linear combination as follows $Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \cdots + \phi_{p1}X_p$ such that $\sum_{j=1}^{p} \phi_{j1}^2 = 1$. First principal component loading vector $(\phi_{11}, \phi_{21}, \ldots, \phi_{p1})^T$ solves a constrained optimization problem for which the sample variance is maximized subject to the normalization constraint [James et al, 2013, pp. 374-285].

PCA makes possible the visualization of the variables taken two by two. For 10 initial variables ($p = 10$), one should examine $\binom{10}{2} = 45$ scatter plots. Principal components $Z_1, Z_2, \ldots, Z_m$ can be aggregated [Keller and Samson, 1986] or used as predictors in a Principal Component Regression (PCR). The first components are the uncorrelated variables of linear regression using the least squares.

b) ICA method was attributed to Jutten and Hérault [Jutten and Hérault, 1991] [Hérault et al, 1985]. It must be pointed out here that the independence of the extraction signals in ICA (and not only their non-correlation) is related to the use of the higher statistics in the nonlinear functions of the learning rule (not only the classical statistics of order 2) [Duvaut, 1990]. Jutten and Hérault [Jutten and Hérault, 1991] compared experimentally the two methods PCA and ICA.

Thus, the signal received by an antenna (or another sensor) is the superimposition of elementary signals emitted by other sources located in its field of reception. The difficulty comes from extracting these different sources from the signals received. Jutten and Hérault proposed a blind identification procedure (without a priori information). They used a fully interconnected neural network with learning abilities.

Hyvärinen [Hyvärinen, 2013] surveyed this probabilistic method for finding maximally independent and non-Gaussian components. ICA approach has two stages: a preliminary whitening stage and an estimation stage. The model is able to recover the original signals constituting a message that had mixed them.

Let the observed variables and independent components be respectively $x_i(t), i = 1, \ldots, n$ and $s_j(t), j = 1, \ldots, m$. The model is $x_i(t) = \sum_{i=1}^{m} a_{ij} s_j(t)$ or $x = As$ in matrix form. The identifiability supposes that three conditions are met: statistically independence, non Gaussian distributions and invertibility of squared matrix A. In the preliminary whitening step the $n \times T$ matrix data X is linearly transformed by $Z = VX$ (V is found by PCA) so that $Z = VX$ is white (i.e. $\frac{1}{T} ZZ^T = I$). The model is $Z = VAS = \tilde{A}S$. Then, transformed mixing matrix $\tilde{A}$ is the estimated by maximizing an objective function that is related to a non-Gaussianity measure of the components [Hyvärinen, 2013]. Figure 1 pictures measured signals (a) and recovered sources by using ICA (b).

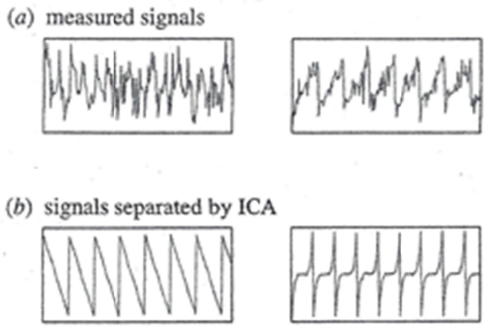

Figure 1 Measured signals and recovered sources by using ICA (extracted figures of Figure 1 from [Hyvärinen, 2013, p. 3])

2.3 Evolutionary Machine Learning

In practice, machine learning, metaheuristic algorithms, and distributed computing combine. Machine learning proved to have faster performances when combined with metaheuristic algorithms. Colanzi et al. [Colanzi et al, 2011] adapted the following three different bio-inspired meta-heuristics in the clustering problem: genetic algorithms (GAs), ant colony optimization (ACO), and artificial immune system (AIS). Carneiro et al. [Carneiro et al, 2019] an algorithm combining network-based data classification and PSO (Particle Swarm Optimization).

3 LEARNING-BASED ARTIFICIAL IMMUNE CLASSIFIER

Darwish [Darwish, 2018] surveyed bio-inspired algoritms for computing. Computational models inspired by the human

immune system correspond to systems that are particularly compliant with the complexity of the phenomena to be described, analyzed and modeled. These adaptive models have decentralized structures for parallel and distributed computing. Immune systems have the ability of learning to recognize patterns, to remember the patterns encountered, and to deduce an overall behavior of local situations encountered [Dasgupta, 1999] [Coello and Cortés, 2005] [Dasgupta and Gonsalez, 2003] [Dasgupta et al, 2011].

3.1 Immune System Principles

Nomenclature

AIC	Artificial immune classifier ;
AIS	Artificial immune system;
Antibody (Ab)	Soluble protein molecule secreted by B-cells in response to an antigen;
Antigen (Ag)	Foreign substance in the body recognized by the immune system;
APC	Antigen presenting cells;
B-cells	White blood cells derived from the bone marrow that developed into plasma;
IS	Immune system;
Lymphocytes	White blood cells produced in the lymphoide organs which are essential in the immune defense;
Lymphokines	Chemical substances secreted by lymphocytes to regulate the immune response;
MHC	Major histocompatibility complex;
Peptides	Short chains of amino acids linked by peptide (amide) bonds;
Plasma cells	Main source of antibodies, these cells are white blood cells participating in the immune defenses.
T-cells	

The immune system defends a vertebrate body from foreign pathogens. Aickelin and Dasgupta [Aickelin and Dasgupta, 2005] present the distinction between two types of immune systems (ISs). Innate (or acquired) IS refers to unchanged existing mechanisms for the detection and destruction of invasive organisms. Adaptive IS, on the other hand, is able to respond instantly to recognized invasive organisms (presented in the past and memorized).

Figure 2 pictures the process of biological immune mechanisms on which the AIS algorithm is based. Suppose that a foreign antigen is detected (top of the figure). Antigens are recognized by antibodies receptors in the immune system. When an antibody matches an antigen the corresponding B-cell is stimulated to proliferate by cloning . The immune system's ability to detect foreign (unknown) antigens is the negative selection mechanism. This mechanism provides a tolerance for self-cells (those of the body). The basic principle of the selection mechanism is that only those cells that recognize the antigen proliferate. In this response adapted to the antigen, T-cells play an essential regulatory role. This process is known as the clonal selection principle. New cloned cells mutate with high rates to increase their receptor population. This genetic mutations generates antibodies with higher affinity for antigens. This phase corresponds to a hypermutation process. Abs-cells become memory cells which are activated in later attacks by similar antigens. They are more efficient secondary response. We are in the presence of an elitist mechanism retaining the best solutions found throughout the process ([Frank, 1996], [Forrest and Hofmeyr, 2000]).

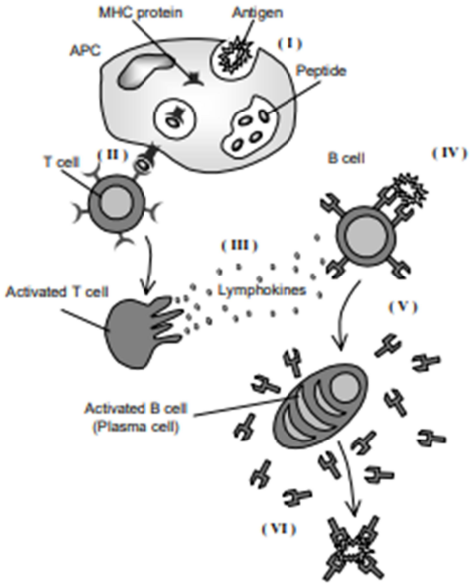

Figure 2. Acquired Human Immunological Mechanisms for the Defense Against Foreign Pathogens (Reprint of Figure 6 from [deCastro , and von Zuben, 1999, p. 12])

3.2 Implementation of Immunological Algorithms

The areas of implementation of AIS are extensive and diverse [Shen et al, 2008]. Cisar et al [Cisar et al, 2014] mention domains in their chronological order of application: *anomaly detection* (reagent recognition in chemical analysis, image analysis, prediction of infections, analysis of medical data, computer security) , *machine learning* (pattern recognition, conceptual learning, clusterin and data classification), *optimization* (optimization of multivariate numerical functions, multi-objective, combinatorial optimization).

Three types of decisions are needed to implement a basic AIS: initializing and encoding scheme of antibodies and antigens, defining an affinity measure between antibodies and antigens, configuring selection and mutation processes [Colanzi et al, 2011]. The CLONALG and aiNet algorithm were developed by de Castro and von Zuben [de Castro and von Zuben, 2000]. Both algorithms evolve a population towards a set of effective detectors. On the other hand, the items of the population interact in the immune network-based aiINet (Artificial immune network) algorithm. These interactions are due to the suppression mechanisms of items that are identical or similar to those of training data (within a threshold) [Timmes et al, 2008].

a) *Encoding Scheme of Cells and Affinity Measure*. For most problems, antibodies and antigens are strings of binary (or real) numbers for which the length is the number of variables and the position the variable identifier. The variables (or cells) can be compared by a simple Pearson correlation coefficient or by another distance indicator. Other distances can the standard Euclidean metric $D(x,y)=\sqrt{\sum_{j=1}^{n}\left(x_j-y_j\right)^2}$ or the recognition ball of

radius ε $B_\varepsilon(x)=\{y: D(x,y)<\varepsilon\}$. For the classical Euclidean distance, the smaller the distance, the stronger is the affinity connecting two items. For the recognition ball, two items inside each ball have a strong affinity [Timmis, 2008]. In other applications, we will use the negative selection according to which close or even identical items must be eliminated. Here, higher the affinity of antibody to the antigen, more clones will be generated.

b) *Configuring Selection and Mutation Processes*. The number of clones is proportional to the affinity to the antigen. Using a round operator, we my write [Colanzi et al, 2011] $\text{numClones}=\text{round}(\beta n/i)$ where β denotes a clonal multiplicator, n the amount of antibodies, and $i\in\{1,\ldots,n\}$ the antibody affinity-based current ranking. The amount of antibody characteristics that are affected is calculated by $p=\left(\frac{1}{\rho}\right)\exp(-f)$, where ρ is the mutation rate of antibody and f the normalized affine function of the antibody to antigen.

c) *CLONALG Pseudocode*. The aiNet pseudocode is shown by Timmis et al [Timmes et al, 2008]. Colanzi et al [Colanzi et al, 2011] described the following steps for CLONALG:

1. Random generation of an antibody population to counter a set of antigens;
2. Calculation of an affinity measure f for antibodies to antigens;
3. Selection of a subset of n antibodies with the strongest affinities for cloning;
4. Creation of a C clones population whose number is in increasing relation with their greater affinity for antigens;
5. Mutation of the C clones population according to a rate in inverse proportion to that of their affinity to the antigens, and new C^* matured clones population resulting therefrom;
6. Evaluation of the C^* matured clones population with new measure f^* of their affinity;
7. Selection of the n matured antibodies (with higher affinities) to compose the next population generation;
8. Elimination of the worst antibodies in the population and their replacement by random selection.
9. The process repeats until a stop condition is met.

The CLONALG algorithm initially intended to solving machine learning and pattern recognition problems was adapted later to solve optimization problems [de Castro and von Zuben, 2002] [Čisar et al, 2014]. De castro and von Zuben [de Castro and von Zuben, 2009] used a Bayesian AIS for solving multi-objective optimization problems. Chu et al. [Chu et al, 2008] described the flowchart of the immune algorithm by adapting the interpretation of successive operations to the problem of optimization. Figure 3 pictures the different steps of an immune algorithm.

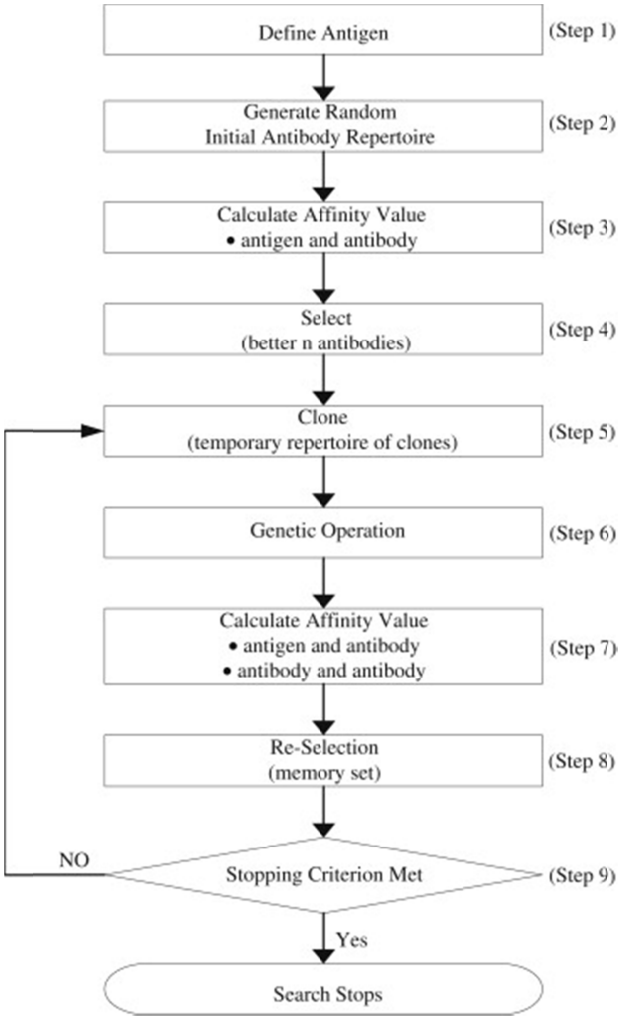

Figure 3. Flowchart of the immune algorithm [reprint of Figure 1 from [Chu et al, 2008] DOI: 10.1016/j.mcm.2008.02.008)

Step 1 defines the antigens (decision variables, objective function and constraints of an optimization problem). The antigens are recognized by antibody receptors (trial solutions of a variable, optimal solutions). *Step 2* generates an initial antibody repertoire by random coding. Thus the real-number coding of the decision variables $x_i, i=1,\ldots,n$ defined each on the intervals $[a_i,b_i]$ has for expression $x_i=a_i+\gamma(b_i-a_i)$ where γ is a random number uniformly distributed between 0-1. *Step 3* evaluates antibody-antigen affinity (based on objective function and possible constraint violations). The degree of violation penalizes the affinity of the antibodies. The affinity Ag_i is calculated according to the formula $Ag_i=(vio_i+f_i)^{-1}$. *Step 4* selects the n best antibodies according to their affinities. This step of the procedure is followed by a cloning operation under

either *Step 5* or *Step 9*, depending on the origin of the antibodies to be cloned. *Step 5* refers to cloning of the n antibodies selected in the previous step. The set C of the clones is placed in a temporary repertoire. *Step 6* generates new antibodies, which can come from complete genetic operations such as crossover and mutation. The crossover would get new antibodies by mixing their genetic material. The mutation that randomly changes the components of an antibody makes it possible to obtain a greater diversity. The antibody-antigen affinities are evaluated at this step. The purpose of *Step 7* is to maintain the diversity of the antibodies between them. This antibody-antibody affinity is obtained by comparing the antibodies of the current register with best antibodies of the memory M. The antibody-antibody affinity Ab_j can be expressed as $Ab_j = \left(1 + D_j\right)^{-1}$ where D_j is the Hamming distance. *Step 8* generates the memory set M. It collects antibodies (trial solutions) with high antibody-antigen affinity to recognize antigens (optimal solutions). In *Step 9* the calculation stops if the test relating to the stopping criterion is satisfied. Chu et al [Chu et a, 2008] also proposed A modified immune algorithm combining AIS with the genetic algorithm GA with higher search capabilities. GA is used as a pre-processor to screen the initial antibodies repertoires.

3.3 Artificial Immune Classifier

Classification methods consists of two main steps. A learning step and a classification step. A classifier is built in the learning step from a training set, and the model is used in the classification step to predict class labels for the data [Han et al, 2012, pp. 328-330]. The AIS algorithm are adaptive systems that found classification methods [Zhang and Yi, 2007]. It is used in practice for the detection of industrial anomalies or computers. AIS classifier with K-means belongs to the unsupervised classification methods for which training data are not needed . It is not indicated to which class each training tuple $\left(x_1, \ldots, x_n\right)$ belongs. Colanzi et al [Colanzi et al, 2011] showed that AIS is one of the bio-inspired metaheuristics for improving conventional clustering methods (such as the K-means algorithm). Clustering techniques explore similar patterns and collect them into categories or clusters. The solutions do not require prior knowledge about the data to be clustered. A clustering problem is reduced to the minimization of the sum of the Euclidean distances between each object and the center of the belonging cluster. The optimization problem is

$$\min \sum_{i=1}^{o} \sum_{j=1}^{c} w_{ij} \sqrt{\sum_{a=1}^{n} \left(x_{ia} - c_{ja}\right)^2}$$

where o and c denote the number of data objects and centers, respectively; x_{ia} is the value of a th attribute of i th data object; c_{ja} is the value of a th attribute of j th cluster center; $w_{ij} \in \{0,1\}$ is an associated value taking 1 if the object is grouped in cluster j, and 0 otherwise. Zhong et al and al [Zhong et al, 2007] proposed unsupervised AIC (artificial immune classification) [Zhong et al, 2006] and supervised AIC [Zhong et al, 2007] for remote-sensing imagery. Karakose [Karakose, 2013] compared the flowcharts of AICs with supervised-unsupervised learning models. Karakose proposed a reinforcement-based AIC. This approach is similar to a self-learning scheme using the mechanisms of clonal selection and memory cells. Unlike other algorithms, the reinforcement learning algorithm calculates the set of probabilities of the present state to determine the next state.

4 COMPLEX PREDICTION PROBLEMS

4.1 Self-Adaptive Data Driven Methods

a) *Conventional Trading Strategies*.

Abbreviations

AIC: Akaike information criterion*; AR(p):* *p*-order autoregressive model; *ARCH(q):* autoregressive conditional heteroskedasticity model; *ARIMA(p,d,q)*: *d*-degree differenced mixte autoregressive integrated moving average with orders *p,q* respectively; *ARMA(p,q): mixed p*-order autoregressive-*q*-order moving average model*; BIC: Bayesian information criterion; BJ: Box and Jenkins criterion; CAT:* criterion autoregressive transfer function*; EMA*: exponential moving average; *FPE*: final prediction error; *GARCH(p,q):* generalized autoregressive conditional heteroskedasticity model; *HQ:* Hannan-Quin criterion; *MA(q)*: *q*-order moving average; *L-MA*: long-term moving average; *S-MA:* short term moving average; *MACD*: moving average convergence/divergence; *SMA:* simple moving average; *WMA:* weighted moving average.

Machine learning can relate to the time series that are available in large scale. A temporal series refers to a sequence of ordered historical measures observed at regular intervals of time (yearly, monthly, daily, hourly, for example). Predicting future values from observed data from the past is a common concern of many scientific disciplines (meteorology, telecommunications, finance, etc.) [Bontempi et al, 2013].

The trading strategies are mainly based on a variety of weighted MA (WMA, EMA) or non-weighted (SMA) techniques, ARMA models, ARCH models, GARCH models, as well as logit regression [Dunis and Williams, pp. 10-20].

A naïve trading strategy should assume that the more recent rate of return y_t is the best predictor of the next future rate of return $\hat{y}_{t+1}$ such as $\hat{y}_{t+1} = y_t$. Techniques based on moving averages are more widely used. They smooth out short-term fluctuations. Let *n* daily closing prices be $y_1, \ldots, y_n$. A simple 5-day moving average is

$$MA = \frac{1}{5}\left(y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}\right)$$

for $t = 3, n-2$ with $n-4$ terms. Let two MAs be S-MA (for short term MA) and L-MA (for long term MA) with different lengts n, m respectively with $n \ll m$. MACD trading strategy refers to a decision rule for taking position in the market. Thus, if the S-MA curve intersects the L-MA curve from below, a short-term position is taken and conversely. A weighted average of last *n* prices WMA is such that the weighting decreases with each previous prices,

such as with $n=5$. WMA at t *is*

$$\frac{5}{N}y_{t-2}+\frac{4}{N}y_{t-1}+\frac{3}{N}y_t+\frac{2}{N}y_{t+1}+\frac{1}{N}y_{t+2}$$

where N denotes the triangular number $n(n+1)/2$. The set of these decreasing weights is $\left\{\frac{5}{15},\frac{4}{15},\frac{3}{15},\frac{2}{15},\frac{1}{15}\right\}$. The weighting of an exponential moving average (EMA) decreases exponentially, giving more weight to the recent prices. Moving averages SMA and EMA are illustrated in Figure 4.

Figure 4. Chart representing simple and exponential moving averages [drawn from StockCharts https://school.stockcharts.com/doku.php?id=technical_indicator:moving_avearges]

The ARMA(p,q) model with p autoregressive terms and q moving-average terms is

$$y_t=c+\varepsilon_t+\sum_{i=1}^{p}\phi_i y_{t-i}+\sum_{i=1}^{q}\theta_i\varepsilon_{t-i}$$, where c is a constant, ε_t is a white noise, $\phi_1,\ldots,\phi_p$ and $\theta_1,\ldots,\theta_q$ are parameters. In an ARCH(q) process, the errors term $\varepsilon_t=\sigma_t z_t$ is composed of two elements: a white noise z_t and a time-dependent standard deviation σ_t. The error variance series σ_t^2 is specified as $\sigma_t^2=\alpha_0+\sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2$ where $\alpha_0>0$ and $\alpha_i\geq 0, i>0$. A GARCH process assumes that the error variance is an ARMA model.

Logit estimation is a regression technique to determine the probability of an upward or downward movement. This model takes the form

$$y_t^*=\beta_0+\beta_1 x_{1t}+\cdots+\beta_p x_{pt}+\varepsilon_t$$

where the dependent variable y_t^* is observed trough the dummy variable

$$y_t=\begin{cases}1, \text{if } y_t^*>0\\ 0, \text{otherwise}\end{cases}.$$

The automatic selection of ARMA models [Keller and Mourad, 1987] and the update of the Kalman filter are essential complementary tools. The model selection is based on goodness of fit criteria, such as AIC, BIC, CAT, FPE and HQ. Keller and Mourad [Keller and Mourad, 1987] compared these criteria by using Monte Carle experiments. The preference order of these criteria found for an AR (1) was as follows

$$AIC \prec FPE \prec BIC \prec HQ \prec CAT \prec BJ$$

In the forecasting process, new available data requires a fast recomputation of the model. Kalman filter allows re-estimating the model with added sample data. The original model and the additional restrictions are described by the two vector equations $y=X\beta+a$ and $y_k=X_k\beta+a_k$ [Brockwell and Davis, 1987] [Granger and Newbold, 1986].

b) *Neural Network Regression Models*. The purpose of an NNR model is to determine the prediction of an output from a nonlinear transfer function of its weighted inputs.

Figure 5 shows a NNR network suitable for time series prediction. The nodes (or neurons) are distributed over three superimposed layers: an input layer, an output layer, and a hidden intermediate layer. The nodes of two successive layers are completely interconnected. The model of Figure 5 thus describes a three-layer network (an input layer, an intermediate hidden layer, and an output layer). The nodes on each are distributed as follows: 5 input nodes denoted by $X_t=\left(x_t^{(1)},x_t^{(2)},\ldots,x_t^{(n)}\right), n=5$ where $x_t^{(i)}$ denotes the level of activity of the *i*th node, 2 intermediate nodes denoted $b_j, j=1,2$ which are elementary processing units, and a single output node provided with an operator. The weighted entries related to the b_j s' entries. The set of weights is $W=\{W_1,W_2,\ldots,W_n\}$ where $W_j=\left(w_{1j},w_{2j},\ldots,w_{nj}\right)$, each elementary weight representing the strength of a connection. A saturation activation function (sigmoid, hyperbolic tangent or bilogistic sigmoid) depending on the inputs makes it possible to calculate the outputs of each unit b_j. We have

$$b_j=F\left(\sum_{i=1}^{n}x_t^{(i)}w_{ij}-w_0\theta_j\right)$$ where the introduced correction term $w_0\theta_j$ designates an input bias modulated by the associated weight w_{0j}.

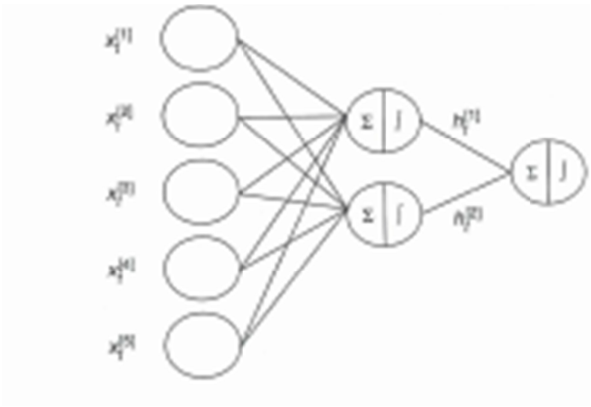

Figure 5. Fully Connected Single Output NNR (reprint of Figure 1.14 from [Dunis and Williams, 2003, p.24])

At the beginning of the process input weights are randomly drawn. The process results in a calculated value that is compared to the observed value. This gap can evaluate the capacity of a model of reproducing the data. Some statistical measures are expressed by root mean square error (RMSE) or by other indicators (average MAE error, MAPE, Theil-U coefficient or directional change CDC [Keller, 1989, p.259-260]). Subsequently, the RMSE is minimized by adjusting the input weights.

c) *Predicting Financial Time Series*. Keller [Keller, 1989 pp. 251-252] listed the applications for forecasting exchange rates in 1967-1984. The data are daily spot rates and forward rates. The instruments are Box-Jenkins models [Keller and Maréchal, 1986], ARIMA models, vector AR, and regression models. Dunis and Huang [Dunis and Huang, 2003, pp. 129-162] cited the following contributions. Baillie and Bollerslev (1990) analyzed the volatility of FX market hourly data from ARIMA and GARCH models [Keller, 1994]. West and Cho (1995) evaluated the predictive capabilities of the GARCH, AR models and non-parametric models on a weekly basis.

The use of neural models (NNRs and RNNs) was extensively used in the studies of the 1990s [Dunis and Huang, 2003]. Kuan and Liu (1995) proposed a two-step RNN for forecasting exchange rates. Tentil (1996 applied the RNNs to predict the USD / DEM, Franses and Van Homlen (1998) used NNRs to predict daily exchange rate data. Dunis and Huang [Dunis and Huang, 2003] used NNRs and RNNS for forecasting and trading currency volatility with application to daily GBP/USD and USD/JPY in the period 1993-1999. The results were benchmarked against GARCH(1,1), implied volatility, and forecast combinations

4.2 Grey Prediction Models

In systems theory, a system can be defined by a color that represents the information we hold about the system. So a system is a *black box* when we do not have any information on the characteristics and equations that govern the system. On the contrary, a white system is a system that can be fully known and described. A Grey system thus occupies an intermediate place with known and unknown information. The (stochastic or fuzzy) uncertainties that affect most real-life models make it possible to place them in this class of gray systems. Kayacan et al. [Kayacan et al, 2010] noted the growing interest of this approach in the early 1990s in far eastern countries (Australia, China, Japan) and the USA.

The theory of the grey system was introduced by Deng [Deng, 1982, 1989]. The model consists of predicting an uncertain system of imperfect information with little data. The method does not directly use raw data but generates data to build a model. The model uses a cumulative generator operator (AGO) to reduce the randomness of the data. This approach, whose parameters are unknown, proved to be superior to conventional statistical models.

Other intelligent approaches in time series analysis have been proposed such as the Kalman filter, fuzzy systems, hidden Markov models and vector support machines. Kayacan et al [Kayacan et al, 2010] give examples of hybridization between classical methods and gray system theory. Wang [Wang, 2002] combined fuzzification techniques with Grey theory to predict stock prices. Huang and Jane [Huang and Jane, 2009] combined a moving average autoregressive exogenous (ARX) with Grey system theory and rough set (RS) for stock market forecasting and portfolio selection. Hajnoori and al [Hadjnoori et al, 2013] also presented a stock price prediction using a grey fuzzy method.

a) *Grey Characteristics*. Gray models are time series prediction models. They predict future values in a time series using only the most recent data set. We make sure that the data are all positive (by applying a first-order accumulated generating operator 1-AGO to the original data). These models come down to differential equations whose coefficients are variables of time. The model is recalculated when new data is available. The general model is denoted $GM(n,m)$ where n denotes the order of the difference equation and m the number of variables. Grey model $GM(1,1)$ is the most used model.

b) *Grey Model First Order One Variable (GM(1,1))*. Let the original data be $x^{(0)} = \left(x^{(0)}(1),\ldots,x^{(0)}(n)\right), n \geq 4$. Using 1-AGO $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1,\ldots n$, we obtain the series $x^{(1)}$. The Grey generated model is the first-order Grey differential equation

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \tag{4.1}$$

where a denotes the Grey developmental parameter and b the Grey control parameter. These parameters are estimated by least-squares method and initial condition $x^{(1)}(1) = x^{(0)}(1)$. The solution to Eq.(4.1) is $\hat{x}^{(1)}(k) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-a(k-1)} + \frac{b}{a}, k = 2,3,\ldots$. It is the time response equation of GM(1,1). Using the inverse accumulated generating operator IAGO, we obtain the

predicted value of the primitive data at time $(k+1)$

$$\hat{x}^{(0)}(k+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-ak}\left(1-e^{a}\right) \quad .$$

The predicted value of the primitive data at time $(t+p)$ is

$$\hat{x}^{(0)}(k+p)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-a(k+p)}\left(1-e^{a}\right)$$

(see [Tien, 2009], [Wang, 2002]).

c) *Grey Verhulst Model [Wen and Huang, 2004].* The Verhulst model describes an increasing process with a saturation region. The grey Verhulst model can be defined as $\frac{dx^{(1)}(t)}{dt}+ax^{(1)}(t)=b\left(x^{(1)}(t)\right)^{2}$ for which E. Kayacan et al [Kayacan et al, 2010] gave the complete solution. The solution of $x^{(1)}$ at time k is [Kayacan et al, 2010]

$$\hat{x}^{(1)}(k)=\frac{ax^{(0)}(1)}{bx^{(0)}(1)+\left(a-bx^{(0)}(1)e^{a(k-1)}\right)} \quad .$$

Applying $x^{(1)}=IAGO.GM(1,1).AGO.x^{(0)}$ we obtain the solution of $x^{(0)}(t)$ at time k [Kayacan et al, 2010]

$$\hat{x}^{(0)}(k)=\frac{ax^{(0)}(1)\left(a-bx^{(0)}(1)\right)}{\left(bx^{(0)}(1)+\left(a-bx^{(0)}(1)\right)\right)e^{a(k-1)}} \times\frac{\left(1-e^{a}\right)e^{a(k-2)}}{\left(bx^{(0)}(1)+\left(a-bx^{(0)}(1)\right)\right)e^{a(k-2)}}.$$

In this case the saturation point at $\frac{a}{b}$ limits the prediction value.

4.3 *Monitoring and Predicting Natural Hazards*

a) *Monitoring and predicting Natural Hazards*. Bogue [Bogue, 2012] provided an overview of the systems installed to monitor and predict natural phenomena with often disastrous consequences. These damaging phenomena include earthquakes, landslides displacements, tsunamis, hurricanes, typhoons and tornadoes, volcanic eruptions and other dangerous phenomena. The probably augmentation of the frequency of these phenomena creates an urgency of observatory installations for real-time monitoring and forecasting such phenomena. Bogue [Bogue, 2012] described some systems in the United States and Japan, and mentions the more recent observation satellites.

Three applications can illustrate the type of models used in the prediction of natural elements such as earthquakes, water filling of a reservoir and the protection of water quality of a lake. The first illustrative application is the prediction of earthquakes by using a conventional back propagation (BP) neural network model [Hu et al, 2012]. The second illustrative application combines the techniques Grey GM(1,1), BP artificial neural networks, and Markov to predict the incoming flow to a reservoir [Hua, 2012]. The third illustrative application is on the protection of high quality fresh water (Vrana Lake in Croatia). The authors [Hrnjica and Bonacci, 2019] of this study have monthly measurements over 38 years and hold a forecast horizon of 6 months to a year. This study uses two types of Artificial Neural Network (ANN): the Recurrent Neural Networks (RNN) with Long-Short Term Memory (LSTM) and the Feed Forward Neural Network (FNN).

b) *Landslide Forecasting Data*. Landslides are movements of earth, rocks and debris along a (mountainous) slope. They are caused by intense rains, earthquakes, storms. Human and material damage can be considered and very expensive. These reasons justified the interest of monitoring and anticipating the occurrence of such hazardous latent phenomena. Figure 6 illustrates the Danba landslide in the eastern margin of the Tibetan Plateau on the right bank of the Dadu river (see Xie et al. [Xie et al, 2017] for a geological description of this area).

Figure 6 Danba landslide location in the eastern margin of the Tibetan Plateau (reprint of Figure 2 from S. Xie et al [Xie et al, 2017])

The available experimental data come from several monitoring points and cover 76 observation periods. The sequence of landslide deformations over these periods is shown in Figure 7. The landslide deformation is nonlinear, non-stationary and random with a sharp peak [Xie et al, 2017].

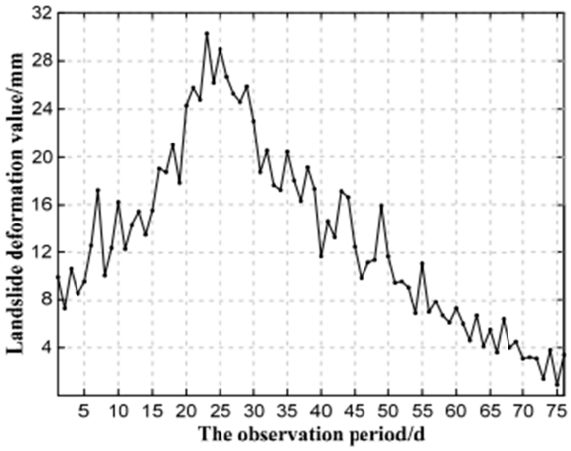

Figure 7 Deformation sequence of the Danba landslide displacement (reprint of Figure 3 from S. Xie et al [Xie et al, 2017])

c) *Landslide Forecasting Models*. To analyze and predict such chaotic time series, Xie et al. [Xie et al, 2017] proposed a combination of techniques including the modified ensemble empirical mode decomposition (MEEMD) technique, the approximate entropy method, and the WLS-SVM model (weighted least square support vector machine). Xie et al. [Xie et al, 2017] compared their model to other prediction models such as with variants of the wavelet neural network, SVM (support machine vector), variants of the LS-SVM approach, and WLS_SVM.

Gray model GM(1,1) is widely used landslide prediction problems. Li et al [Li et al, 2007] applied a center approach Grey GM(1,1) model to predict landslide deformations. Some authors also proposed improvements to the basic model or combined it with other forecasting approaches. Wang [Wang, 2017] renewed the conventional gray GM(1,1) model by observing that a greater accuracy will be obtained with the most recent data. The principle is that the most recent data introduced removes the oldest data. Ma and Li [Ma and Li, 2016] generalized the model GM(1,1) and extended it to discrete gray systems. The generalized grey model (GGM(1,1)) is characterized by the following original data $x^{*(0)} = \left(x^{*(0)}(1), x^{*(0)}(2), \ldots, x^{*(0)}(n)\right)$ where $x^{*(0)}(k) = x^{(0)}(k) + h, k = 1, 2, \ldots, n$ with $h \in \mathbb{R}$. The whitenization differential of GGM(1,1) is defined and we derive the time response equation. Second order grey model GM(2,1) is an important expansion of GM(1,1). Xu and Dang [Xu and Dang, 2015] presented the methodology of GM(2,1). Wang et al [Wang et al, 2013] applied a GM(2,1) model to predict landslide deformations.

5 CONCLUSION

This paper provides an overview of the modeling trends of complex phenomena whose chaotic observed data required appropriate approaches and models. The fields of finance and geology illustrate this study. Gray model GM(1,1) is widely used landslide prediction problems.

REFERENCES

Acharjya D.P. and Kauser A.P. 2016, "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools", *International Journal of Advanced Computer Science and Applications,* **7**, 2, pp. 511-518.

Aickelin U. and Dasgupta D. 2005, "Artificial Immune Systems". In Burke E.K. and Kendall G. (Eds.), *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, Springer Science+Business Media, New York, pp. 3-9.

Bakos Y. 1998, "The Emerging Role of Electronic Internet", *Communications of the ACM*, August. http://dx.doi.org/10.1.1.83.180&rep=rep1&type=pdf

Bogue R. 2012, "Monitoring and Predicting Natural Hazards in the Environment", *Sensor Review*, **32**, 1, pp. 4-11.

Bontempi G. Taieb S. and Le Borgne Y-A. 2012, "Machine Learning Strategies for Time Series Forecasting". In Anfaure MA and Zimány (Eds.), Business Intelligence, *Lecture Notes in Business Information Processing*, **138**, Springer, Berlin Heidelberg, pp. 62-77, http://dx.doi.org/10.1007/978-3-642-36318-4_3

Borges P.V.K. and Izquierdo E. 2010, "A Probabilistic Approach for Vision-Based Detection in Videos", *IEEE Transactions on Circuits and Systems for Video Technology*, **20**, 5, pp. 721-731, http://dx.doi.org/10.1109/TCSVT.2010.2045813

Brockwell P.J. and Davis R.A. 1991, *Time Series: Theory and Methods*, Second edition, Springer Verlag, New York, NY.

Cantú-Paz E. 1998, "A Survey of Parallel Genetic Algorithms", *Calculateurs Parallèles, Réseaux et Systèmes Répartis*, **10**, 2, pp. 141-171.

Cantú-Paz E. and Golberg D.E. 2000, "Efficient Parallel Genetic Algorithms: Theory and Practice", *Computer Methods in Applied Mechanics and Engineering*, **186**, pp. 221-238.

Carneiro M.G. Cheng R. Zhao L. and Jin Y. 2019, "Particle Swarm Optimization for Network-Based Data Classification", *Neural Networks*, **110**, pp. 243-255.

de Castro P.A.D. and von Zuben, F.J. 2009, "MOBAIS: A Bayesian Artificial Immune System for Multi-Objective Optimization", *Journal of Mathematical Modelling and Algorithms,* **8**, 2, pp.151-173.

de Castro L.N. and von Zuben F.J. 2002, "Learning and Optimization Using the Clonal Selection Principle", *IEEE Transactions on Evolutionary Computations*, Special Issue on Artificial Immune Systems **6**, 3, pp.239-251.

de Castro L.N. and von Zuben F.J. 2000, "An Evolutionary Immune Network for Data Clustering". In *Proceedings of the Sixth*

Brazilian Symposium on Neural Networks, Rio de Janeiro, Brazil, pp. 84-89.

de Castro L.N. and von Zuben F.J. 1999, "Artificial Immune Systems: Part I- Basic Theory and Applications", *Technical Report*, TR-DCA 01/99 (95 pages).

Chandra A. and Yao X. 2006,"Evolving Hybrid Ensembles of Learning Machines for Better Generalization", *Neurocomputing*, **69**, pp. 686-700.

Chu C-W. Lin M-D. Liu G-F. and Sung Y-H. 2008, "Application of Immune Algorithms on Solving Minimum-Cost Problem of Water Distribution Network", *Mathematical and Computer Modelling*, **48**, 11-12, pp. 1888-1900. http://dx.doi.org/10.1016/j.mcm.2008.02.0008

Čisar P. Čisar S.M. and Markoski B. 2014, "Implementation of Immunological Algorithms in Solving Optimization Problems", *Acta Polytechnica Hungarica*, **11**, 4, pp.225-239.

Coello C.A.C. and Cortés N.C. 2005, "Solvi Optng Multiobjective Optimization Problem Using an Artificial Immune System", *Genetic Programming and Evolvable Machines*, **6**, 2, pp. 163-190.

Colanzi T.E. Klewert W.K.G. Assunção W.K.G Pozo A.T.R Vendramin A.C.B Pereira D.A.B. Zorzo C.A. and de Paula Filho P.L. 2011,"Application of Bio-inspired Metaheuristics in the Data Clustering Problem", *Clei Electronic Journal*, **14**, 3, 18 pages.

Darwish A. 2018, "Bio-Inspired Computing: Algorithms Review, Deep Analysis, and the Scope of Applications", *Future Computing and Informatics Journal*, **3**, pp. 231-246.

Dasgupta D. Yu S. and Nino F. 2011, "Recent Advances in Artificial Immune Systems: Models and Applications", *Applied Soft Computing*, **11**, pp. 1574-1587.

Dasgupta D. (Ed.) 1999, *Artificial Immune Systems.* Springer-Verlag, Berlin, Heidelberg.

Dasgupta D. 1999, "An Overview of Artificial Immune Systems and Their Applications". In D. Dasgupta (Ed.), *Artificial Immune Systems*. Springer-Verlag, Berlin, Heidelberg.

Dasgupta D. and González 2003, "Artificial Immune Systems (AIS) Research in the Last Five Years". In the 2003 Congress on Evolutionary Computation, 8-12 December, , Camberra, Australia.

Deng J-L. 1989, "Introduction to Grey Systems Theory", *The Journal of Grey System*, **1**, pp.1-24.

Deng J-L. 1982, "Control Problems of Grey Systems", *Systems and Control Letters*, **5**, pp.288-294.

Dunis C.L. and Williams M. 2003, "Applications of Advanced Regression Analysis for Trading and Investment". In Dunis C.L., Laws J. and Naïm P. (Eds.), *Applied Quantitative Methods for Trading and Investment*, John Wiley & Sons, Chichester UK, pp. 1-40.

Dunis C.L. and Huang X. 2003, "Forecasting and Trading Currency Volatility: An Application of Recurrent Neural Regression and Model Combination". In Dunis C.L., Laws J. and Naïm P. (Eds.), *Applied Quantitative Methods for Trading and Investment*, John Wiley & Sons, Chichester UK, pp. 129-162.

Duvaut P. 1990, Principles of Source Separation Methods Based on Higher Order Statistics (in French: "Principes des méthods de séparation fondées sur les moments d'ordre supérieur"), *Traitement du Signal*, **7**,5, Special issue on Nonlinear Non-Gaussian, pp. 407-418.

Fayyad U. Piatetsky-Shapiro G. and Smyth P. 1996, "From Data Mining to Knowledge Discovery in Databases", *AI Magazine*, **17**, 3, pp. 37-54.

Foggia P. Saggese A. and Vento M. 2015, "Real-Time Fire Detection for Video-surveillance Applications Using a Combination of Experts Based on Color, Shape, and Motion", *IEEE Transactions on Circuits and Systems for Video Technology*, **25**, 9, pp. 1545-1556. http://dx.doi.org/10.1109/TCSVT.2015.2392531

Forrest S. and Hofmeyr S.A. 2000,"Immunology as Information Processing". In Segel L.A. and Cohen I. (Eds), *Design Principle of the Immune System for the Immune System and Other Distributed Autonomous Systems*, Oxford, University Press, pp.341-387.

Frank S.A. 1996, "The Design of Natural and Artificial Adaptive Systems". In Rose M.R. and Lauder G.V. (Eds.), *Adaptation*, Academic Press, San Diego, CA.

Granger C.W.J. and Newbold P. 1986, *Forecasting Economic Time Series*, Second edition, Academic Press, Orlando San Diego New York.

Gudivada V.N. Baeza-Yates R. and Raghavan V.V. 2015, "Big Data: Promises and Challenges", *Computer,* **48**, 3, pp. 20-23.

Hajnoori A. Amiri M. and Alimi A. 2013, "Forecasting Stock Price Using Grey-Fuzzy Technique and Portfolio Optimization by Invasive Weed Optimization Algorithm", *Decision Science Letters*, **7**, pp. 175-184.

Han J. Kamber M. and Pei J. 2012, *Data Mining: Concepts and Techniques*, Third Edition, Morgan Kaufmann Publishers, Waltham MA, Amsterdam, Boston.

Hérault J. Jutten C. and Ans B. 1985, Detection of Primitive Quantities in a Composite Message by a Neuro-Mimetic Computing Architecture in Unsupervised Learning (in French:"Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé"). In *Proceedings Xème Colloque GRETSI*, Nice, France, May 20-24, pp. 1017-1022.

Hu W-S. Wang H. and Nie H.L. 2012, "Regional Short-Term Earthquake Prediction Model Based on BP Neural Network", *Applied Mechanics and Materials*, **166-169**, pp. 2309-2314.

https://doi.org/10.4028/www.scientific.net/AMM.166-169.

Hua G. 2012, "Application of Combined Grey Forecasting Model in the Prediction of Incoming Flow", *Agriculture Research in the Arid Areas*, **1**.

Huang K.Y. and Jane C-J. 2009, "A Hybrid Model for Stock Market Forecasting and Portfolio Selection Based on ARX, Grey System and RS Theories", *Expert Systems with Applications*, **36**, pp. 5387-5392.

Hrnjica B. and Bonacci O. 2019, "Lake level Prediction using Feed Forward and Recurrent Neural Networks", *Water Resources management*, 33, 7, pp. 2471-2484. https://doi.org/10.1007/s11269-019-02255-2

Hyvärinen A. 2013,"Independent Component Analysis: Recent Advances", *Philosophical Transactions, A (Mathematical Physical and Engineering Sciences)*, **371**, pp. 1-19. http://dx.doi.org/10.1098/rsta.2011.0534

James G. Witten D. Hastie T. and Tibshirani R. 2014, *An Introduction to Statistical Learning with Applications in R*, Springer, New York, Heidelberg, Dordrecht, London. On

Jutten C. and Hérault J. 1991, "Blind Separation of Sources, Part I: An Adaptive Algorithm Based on Neuromimetic Archtecture, *Signal Processing*, **24**, 1, pp. 1-10. http://dx.doi.org/10.1016/0165-1684(91)90079-X

Kambatla K. Kollias G. Kumar V. and Grama A. 2014, "Trends in Big Data Analytics", *Journal of Parallel and Distributed Computing,* **74**, 7, pp. 2561-2573.

Karakose M. 2013, "Reinforcement Learning Based Artificial Immune Classifier", *The Scientific World Journal*, 2013, **1**, 7 pages. http://dx.doi.org/10.1155/2013/58184

Kayacan E. Ulutas B. and Kaynak O. 2010, "Grey System Theory-based Models in Time Series Prediction", *Expert Systems with Applications*, **37**, pp. 1784-1789.

Keller A. 1994, ARCH and GARCH Modeling of Conditional Heteroskedasticity for Financial Time Series (in French: "Modélisation ARCH et GARCH de l'hétéroscédasticité conditionnelle des séries chronologiques financières"), IAE Research Seminar, UVHC University, November 10, Valenciennes (France).

Keller A. 1989, "Advanced Time-Series Analysis". In Dunis C. and Feeny M. (Eds), *Exchange Rate Forecasting*, Woodhead-Faulkner, New York, London, Cambridge UK, pp. 207-299.

Keller A. and Mourad M. 1987, "Automatic Selection Procedures of ARIMA Models: A Comparison of Alternative Criteria", *First International Conference on Industrial and Applied Mathematics*, June 29-July 3, Paris.

Keller A. and Maréchal B. 1986, "Box-Jenkins Forecasting Techniques for the Daily Franc/Dollar Echange Rate: Foreign Exchange Market's Efficiency Re-Examined", *Fifth International Forecast Symposium: Expanding the Domain and Using the Forecast*, June 15-18, Paris.

Keller A. and Samson P. 1984, Constructing Aggregated Time Series by Factorial Analysis of Multiple Cyclical Indices (in French: "Construction d'indicateurs synthétiques par l'analyse factorielle d'indices conjonturels multiples"), *Prévision et Analyse Economique (Cahiers du GAMA)* **5**, 3, pp. 7-52.

Lemley J. Abdul-Wahid S. Banik D. and Andonie R. 2016, "Comparison of Recent Machine Learning Techniques for Gender Recognition from Facial Images". In *Proceedings of the 27th Modern Artificial Intelligence and Cognitive Science*, Dayton OH, USA, pp. 97-102.

Li X. Kong J.M. Wang C.H. and Wang C.H. 2007, "Application of Center Approach Grey GM(1,1) Model Prediction of Landslided Deformation with a Case Study", *Journal of Engineering Geology*, **15**, pp. 673-676.

Luo J. and El Baz D. 2018, "A Survey on Parallel Genetic Algorithms for Shop Scheduling Problems". In *Proceedings of IEEE International Parallel and Distributed Processing Symposium*, May 21-25, Vancouver BC, Canada, pp. 629-636.

Ma C. Zhang H.H. and Wang X. 2014, "Machine Learning for Big Data Analytics in Plants", *Trends in Plant Science,* **19**, 12, pp.798-808.

Ma D-Y. and Li J-L. 2016," Two Novel Grey System Models and Their Applications on Landslide Forecasting" *Journal of Control Science and Engineering,* 2016, article ID 5727160, 6 pages. http://dx.doi.org/10.1155/2016/5727160

Scott-Morton M.S. (Ed.) 1991, *The Corporation of the 1990s: Information Technology and Organizational Transformation*, Oxford University Press, New York.

Shalev-Shwartz S. and Ben-David S. 2014, *Understanding Machine Learning*. Cambridge University Press, Cambridge UK, New York, Melbourne, New Dehli.

Shen X. Gao X.Z. and Bie R. 2008, "Artificial Immune Networks: Models and Applications", *International Journal of Computational Intelligence Systems*, **1**, 2, pp. 168-176.

Sunny M. 2017, "Survey on Various Fire Detection Methods", *International Research Journal of Engineering and Technology*, **4**, 2, pp. 344-347. https://www.irjet.net/archives/V4/i2/IRJET-V41268.pdf

Tien T-L. 2009, "A New Grey Prediction Model FGM(1,1)", *Mathematical and Computer Modelling*, **49**, pp. 1416-1426.

Timmis J. Hone A. Stibor T. and Clark E. 2008, "Theoretical advances in artificial immune systems", *Theoretical Computer Science*, **403**, 1, pp. 11-32.

Tsai C.-W. Lai C.-F. Chao H.-C. and Vasilakos A.V.

2015, ”Big Data Analytics: a Survey”, *Journal of Big Data*, **2**, 2, pp. 1-32.

Wang Y-F. 2002, “Predicting Stock Price Using Fuzzy Grey Prediction System”, *Expert Systems with Applications*, **22**, 1, pp. 33-39.

Wang Z. 2017, “Application of Renewal Gray GM(1,1) Model to Prediction of Landslide Deformation”, *Journal of Geoscience and Environment Protection*, **5**, pp. 148-154.

Wang D. Huang X. Wang M.D. and Liu T.B. 2013, “Application of Gray GM(2,1] Model to Prediction of Landslide Deformation”, *Hydrogeology & Engineering Geology*, **40**, pp.52-54.

Wen K-L. and Huang Y-F. 2004, “The development of Grey Verhulst Toolbox and the Analysis of Population Saturation State in Taiwan-Fukien”. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, **6**, October 10-13, The Hague, Netherlands, pp. 5007-5012.

Xie S. Liang Y. Zheng Z. and Liu H. 2017, “Combined Forecasting Method of Landslide Deformation Based on MEEMD, Approximate Entropy, and WLS-SVM”, *International Journal of Geo-Information*, **6**, 5, pp. 1-16. http://dx.doi.org/10.3390/ijgi6010005

Xu N. and Dang Y-G. 2015, “An Optimized Grey GM(2,1) Model and Forecasting of Highway Subgrade Settlement”, *Mathematical Problems in Engineering*, **2015**, Article ID 606707, 6 pages. http://dx.doi.org/10.1155/2015/606707

Zhang G-Q. Patuwo B.E. and Hu M.Y. 1998, ”Forecasting with Artificial Neural Networks: the State of the Art”, *International Journal of Forecasting*, **14**, pp. 35-62.

Zhang C and Yi Z. 2007,”An Artificial Immune Network Model Applied to Data Clustering and Classification”. *Proceedings of 4th International Symposium on Neural Networks*, June 3-7, Nanjing, China.

Zhong Y. Zhang L. Huang B. and Li P. 2006, “An Unsupervised Artificial Immune Classifier for Multi/Hyperspectral Remote Sensing Imagery”, *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 2, pp. 420-431.

Zhong Y. Zhang L. Gong J. and Li P. 2007, “A Supervised Artificial Immune Classifier for Remote Sensing Imagery”, *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 12, pp. 3957-3966.

THE AUTHOR

André A. Keller (Prof. PhD) is associate member in the Research Unit for *Statistics, Analysis and Multidisciplinary Modeling (SAMM)* in the University Paris 1 Panthéon-Sorbonne. He is a Principal Editor of the *Journal of Computational and Applied Mathematics* (Elsevier). He is also Editor of the journal *WSEAS Transactions on Mathematics*. He has also held teaching positions in the University of Paris, and other universities in France and foreign countries. Prof. Keller taught notably optimization techniques, econometric modeling, applied mathematics and game theory. He presented numerous Plenary Lectures at International Conferences. His research interests include discrete mathematics, circuit analysis, control theory, time series analysis, fuzzy logic, game theory, time-delay systems, multi-objective optimization, metaheuristic algorithms. His publications include articles and books. A recently published books was entitled "*Time-Delay Systems with Economic Dynamics & Control"* (Lambert Academic Publishing, Saarbrücken, Germany, 2010). Another published book addressed the "*Mathematical Optimization Terminology: A Comprehensive Glossary of Terms*" (a publication of Elsevier/Academic Press). Two parts of the book "*Multi-Objective Optimization in Theory and Practice: I: Classical Methods*" and "*II: Metaheuristic Algorithms*" were published in 2017 and 2019, respectively (Bentham Science Publishers, Sharjah UAE). A book in preparation is on Bayesian optimization to machine learning by Springer. Prof. Keller is a member of the *American Association for Science and Technology* (MSCIT). He reviewed numerous scientific papers for international journals of major publisher, such as Elsevier, Hindawi, Springer, World Academic Publishing, World Scientific, WSEAS Press.