DE GRUYTER OPEN
1862
RIGA TECHNICAL UNIVERSITY
Information Technology and Management Science
ISSN 2255-9094 (online)
ISSN 2255-9086 (print)
December 2016, vol. 19, pp. 45–52
doi: 10.1515/itms-2016-0010
https://www.degruyter.com/view/j/itms

Neural Network Modelling for Sports Performance Classification as a Complex Socio-Technical System

Ivars Namatēvs[1], Ludmila Aleksejeva[2], Inese Poļaka[3]
[1–3] *Riga Technical University*

Abstract **– Extraction of meaningful information by using artificial neural networks, where the focus is upon developing new insights for sports performance and supporting decision making, is crucial to gain success. The aim of this article is to create a theoretical framework and structurally connect the sports and multi-layer artificial neural network domains through: (a) describing sports as a complex socio-technical system; (b) identification of pre-processing subsystem for classification; (c) feature selection by using data-driven valued tolerance ratio method; (d) design predictive system model of sports performance using a backpropagation neural network. This would allow identifying, classifying, and forecasting performance levels for an enlarged data set.**

Keywords **– Classification, data pre-processing, multi-layer neural networks, sports performance.**

I. Introduction

The prediction of sports performance is carried out using different methodological approaches. The first and most common approach found in the literature has to do with the use of traditional statistical methods, such as linear discriminant analysis, multivariate discriminant analysis [1], multiple linear regressions [2], and probit and logit model [3]. The next step in solving the classification problems starts with the establishment of induction methods. Some of the most popular of this kind are recursive, ID3, C4.5, CN2, C5.0, kNN (k-nearest neighbours algorithm), and structural equation modelling, which compares theoretical models to datasets [4].

The current trend in the intelligent nonlinear system modelling research is concerned with the integration of artificial intelligence (AI) tools: intelligent agents, neural networks, evolutionary algorithms, fuzzy technology, and support vector machines [5] in a framework for solving complex adaptive problems [6].

There are not a lot of studies referring to the evaluation of sports performance, especially, based on a socio-technical system by putting into practice artificial neural networks (ANNs). On the other hand, this method has been used and proven useful in other fields, such as medicine, engineering, and economics.

Although ANNs have been successfully applied to a wide range of supervised and unsupervised learning problems, they have not often been applied for data mining settings, in which the comprehensibility of learned models and the time required to induce models from large datasets are two fundamental considerations [7].

The rest of this article is organised as follows. In section II, the applicability of neural network methods to the task of data mining based on socio-technical systems theory is considered. Specifically, one may be willing to consider complexity and nonlinearity using neural networks. In section III, related studies using trained neural networks on sports performance are uncovered. Section IV covers a comprehensible system architecture consisting of the proposed pre-process and classification subsystems for sports performance classification. Conclusions and scope for future research are mentioned in Section V.

II. Theoretical Considerations

Conceptually, performance of the socio-technical system is assessed based on the use of the system, the relevant data mining concept and the artificial neural model.

A. Concept of Complex Socio-Technical System Modelling

The fundamental problems of social and technical sciences are the study of nonlinearity, complexity and randomness [8]. These constructs can be evaluated based on inherent performance, decision-making, and coordination structures and networks in different industries, i.e., size, aggregation, communication among human population, and evaluation of their efficiency and effectiveness towards different markets.

It is now recognised that many of our ecological, social, economic, and political problems are also of a global, complex, and nonlinear nature [9]. Today the fast-changing environment presents a challenge for complexity research practitioners, who rely on the systems theory [10]. Indeed, we are surrounded by and integrated into nonlinear, unsmooth and random systems or systems of systems, where rules and networks whether technical, environmental, immune or social could not be easily predicted and controlled [11].

System models, which involve people and technology in the workplace, are called socio-technical systems (STSs). These systems include a network of users, developers, information technology at hand, and the environments in which the system will be used and supported [12]. The systems include complexity as an important paradigm across disciplines in science, sports, and business, which should include system modelling. Complexity theory to the study of STS describes a complex system for which it is difficult, if not possible, to reduce the number of parameters or characterising variables without losing its essential functional properties [13].

It seems that the concept of complexity under consideration can improve the methods of modelling and the design of a complex STS. It is pointed out that while studying complexity one should take into consideration: (a) ontological levels of complexity; (b) the variety and quantity of constituent elements of a complex system along with their interrelation; (c) the increasing cognitive resources needed to make sense of more complex systems.

There are seven varieties of theories of complexity listed in [8]: computational complexity, Kolmogorov complexity and algorithmic information theory, stochastic complexity, descriptive complexity theory, information-based theory of complexity, descriptive complexity theory, Diophantine complexity, and nonlinear dynamic complexity.

A truly complex system would be completely irreducible. This means that it would be impossible to derive a simplified model from this system without losing relevant properties [10]. The process of modelling is quite challenging as it includes many entities that cooperate among themselves. According to a decision making problem, the architecture of system modelling can be constituted by one model or an ensemble of models. The unifying concept resulted in the study of STS as a non-linear system. Moreover, most soft and hard systems are nonlinear in nature and characterised as complex.

Considering sports as an STS will provide a basis for applying methods and techniques to develop a model describing performance as socio-technical events that occur at multiple levels [14]. Figure 1 summarises the definitions of each STS component and the nature of relationships among the components.

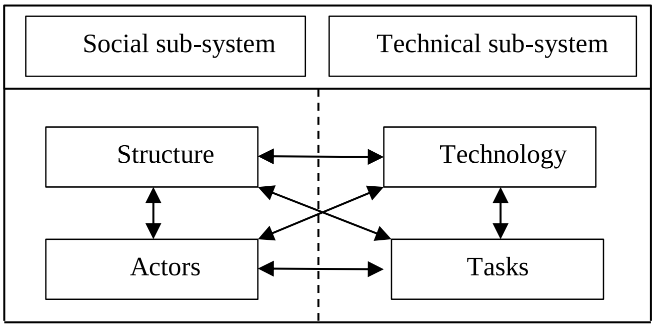


Fig. 1. Socio-technical system [15].

Hence, nonlinear system modelling has attracted a lot of attention in many areas such as classification and decision making, process monitoring, control, optimisation, and prediction [16]. System identification and performance, in connection with a decision-making process, are modelling procedures, which are based on mathematical representation of mapping of a large amount of input and output and of a large amount of data for complex dynamical systems [17].

B. Data Mining in Sports

Data mining presents building intelligent descriptive and predictive models incorporating the relationship between the description of a situation and a result related to the situation [4], [18]. Vercellis [19] has noted that data mining is based on inductive learning methods, whose main purpose is to derive general rules starting from a set of available examples. Data mining methodology, techniques and tools help develop a knowledge based system that can assist sport trainers in decision making [20]. The data mining system is classified on the basis of functionalities. The major data mining functionalities are classification, prediction, clustering, association rule mining, and characterisation [18]. Johnson and Wichern [21] emphasise five main categories of data mining, including classification, prediction, clustering, market basket analysis, and description.

The task of the classification is to discover the optimal class membership $f: D \rightarrow C$ from the given dataset $D = \{t_1, \dots, t_i, \dots, t_n\}$, where t_i is ith data record and class set $C = \{c_1, \dots, c_j, \dots, c_y\}$. The t_i are usually m-dimensional vectors, which are called input variables [22].

There are two different distinctions of data classification problem: (a) dichotomous where the class labels either 0 or 1 are assigned to an unknown data item; (b) model P(c|t), which outputs not only a class label for a data item, but also a probability of class membership [21].

One of the challenges in building classifiers is that it is typically high dimensional, large M with a relatively few samples, N. Supervised learning is used for classification, whereas unsupervised learning is used for clustering [23].

Ofoghi et al. [24] when analysing sports performance data indicate that there are three main attributes, which are of interest to sport scientists, namely, ranking, time, and scores. The sports performance analysis involves: (a) major data mining methods, i.e., classification, clustering, relationship modelling, rule mapping; (b) data mining techniques, i.e., k-means, regression analysis, Bayesian networks, neural networks; (c) sports performance analysis requirements, i.e., performance prediction, demand analysis, real-time decision making. A comprehensive sports performance analysis involves different data mining methods and techniques. Table I summarises [25] the major use of four data mining methods examined
in the paper.

TABLE I
DATA MINING METHODS AND TECHNIQUES

Method	Technique	Sport
Classification	Naïve Bayes	track cycling
	linear discriminant analysis	rugby
	linear discriminant analysis	rowing
	linear discriminant analysis	long jump
Clustering	k-means	track cycling
	k-means	golf
	self-organising maps	basketball
	mixed	decathlon
Relationship modelling	linear regression	swimming
	neural networks	swimming
	neural networks	aerobics
	neural networks	football
Rule mining	association rules	basketball
	association rules	table tennis

Consequently, the mapping, which starts from the sport domain to the data mining domain for performance analysis, serves a starting point to combine sport science as a STS and computer sciences. As the present paper focuses on the study of neural network modelling, the data mining method classification has been chosen.

To sum up, the objective of classification is to predict (the class) user-specified goal attribute based on the values of other attributes, called the predictive (feature) attributes. The goal attribute might be the prediction of whether or not a sportsman is ready for competition, while the predictive feature attributes might be the stress level or consumption of the supplement.

There are many different approaches to solve classification tasks – the most popular ones [25] are to apply artificial neural networks (ANNs) and linear genetic programming (LGP).

C. Artificial Neural Network for Sports Performance

Artificial neural networks (ANNs) can be described as an extremely simplified model of brain cells that cooperate with each other to perform the desired function [26].

The complexity of brain neurons is highly abstracted when modelling artificial neurons. The main advantages of ANN are: (a) the opportunity to retrieve hidden information that allows solving complex problems; and (b) the ability to generalise and produce both linear and non-linear outputs [27].

Neural network methods are not commonly used for data mining tasks because they often produce incomprehensible models and require long training times. However, using two approaches of neural network learning algorithms: (a) rule extraction and (b) easy-to-understand networks, one would be able to produce comprehensible models that do not require excessive training times [7].

For several decades researchers from many scientific disciplines areas have been putting their efforts for designing ANN predictive models considering the following problems: pattern classification, clustering and categorisation, function approximation, optimisation, content-addressable memory, and control [28]. ANNs represent one of the most successful identification techniques used to model nonlinear dynamics, complexity and randomness of systems. ANNs represent a modern branch of automatic control theory, which has existed for several years and suggests an alternative solution to this problem. ANNs have been effectively used for approximating complex nonlinear functions [29]. However, they cannot cope well with feature interaction. ANNs are treated as black box learning and it is difficult for humans to understand or interpret the classification explicitly.

In this paper, a classifier is developed that predicts readiness or unreadiness for performance using the rough set data-driven valued tolerance ratio method and backpropagation learning algorithm (RS-BPNN).

III. Related Studies

Pfeiffer and Hohmann [30] have shown that training science views itself as an integrated and applied science, developing practical measures founded on the scientific method. Here, the interrelations between different variables or variable sets are mostly of a nonlinear character. In these cases, methods like neural networks, e.g., the pattern recognising methods of self-organising Kohonen feature maps or similar instruments to identify interactions might be successfully applied to analyse data. Two examples are given, in which neural networks are employed for pattern recognition. While one investigation deals with the detection of sporting talent in swimming, the other is based on game sports research, identifying tactical patterns in team handball. The third and last example shows how an artificial neural network can be used to predict competitive performance in swimming.

Memmert and Perl [31] outline a framework for analysing types of individual development of creative performance based on neural networks. Consequently, two kinds of sport-specific training programs for the learning of game creativity in real field contexts have been investigated. By using neural networks, it is now possible to distinguish between five types of learning behaviours in the development of performance, the most striking ones being “up-down” and “down-up”.

By using a multi-layer-perceptron neural network, a model to predict the flight of javelins has been developed by Maier et al. [32]. The input parameters to the model are three release angles and the velocity at release, while the output is the distance reached. The neural network model was found to predict actual flights of javelins within 5 %, with a mean difference between the model and real throws of 2.5 %. The model was used to generate maps of distances reached for different combinations of release parameters.

Dut-Mazumder et al. [33] have explored the factors that best explain the performance of association football teams. According to the dynamical system analysis, movement patterns in team sports exhibit nonlinear self-organising features. Artificial neural networks as nonlinear processing tools are becoming increasingly popular to investigate the coordination of participants in sports competitions. Typical ANN learning algorithms can be adapted to perform pattern recognition and pattern classification. Particularly, dimensionality reduction by a Kohonen feature map (KFM) can compress chaotic high-dimensional datasets into low-dimensional relevant information. Such information would be useful for developing effective training drills that should enhance self-organising coordination among players.

Silva et al. [34] present a study on the use of neural network to create realistic models of swimming performance prediction. To support the creation of swimming performance prediction models, the authors have identified the factors explaining the performance in the 200 meters individual medley and 400 meters front crawl events in young swimmers. The authors suggest that the neural network tool can be a good technique in the resolution of performance modelling problems.

Iyer and Sharda [35] have investigated team selection for international sports competitions, which requires predicting performance of individual athletes. Researchers have described the use of neural networks to rate players and select specific players for a cricket competition. The neural networks are employed to predict each cricketer’s performance in the future based upon their past performance. The classifier of the cricketers into three categories – performer, moderate and failure – has been developed. The neural network models have been progressively trained and tested using four sets of data. The trained neural network models have then been applied to generate a forecast of the cricketer’s near term performance. The results have shown that the neural networks can indeed provide valuable decision support in a team selection process.

The comparison of the studies performed by different researchers clearly shows that a neural network can be used as a comprehensive tool for sports performance analysis. Choosing the most appropriate model means the creation of a comprehensive system architecture based on STS and ANN design.

IV. System Architecture

The proposed system architecture describes a model bringing together sports performance analysis requirements and data mining classification method through the scope of socio-technical components. The ANN is used to analyse sports performance for decision making.

A. System Architecture

The structure of the proposed system is organised to combine the data and mathematical model to help decision makers, i.e., trainers, sport agents, in their work. The process of decision making is divided into three main stages. The data management module that includes a database is designed to contain the data required to evaluate sports performance. The exclusion stage is the pre-process subsystem, where compatibility and restrictions are identified. Finally, the evaluation stage is the classification subsystem, where the ANN classifies feasible attributes on the basis of performance criteria in order to identify the preferred decision as the best opportunity. The architecture of the proposed system is shown in Fig. 2.

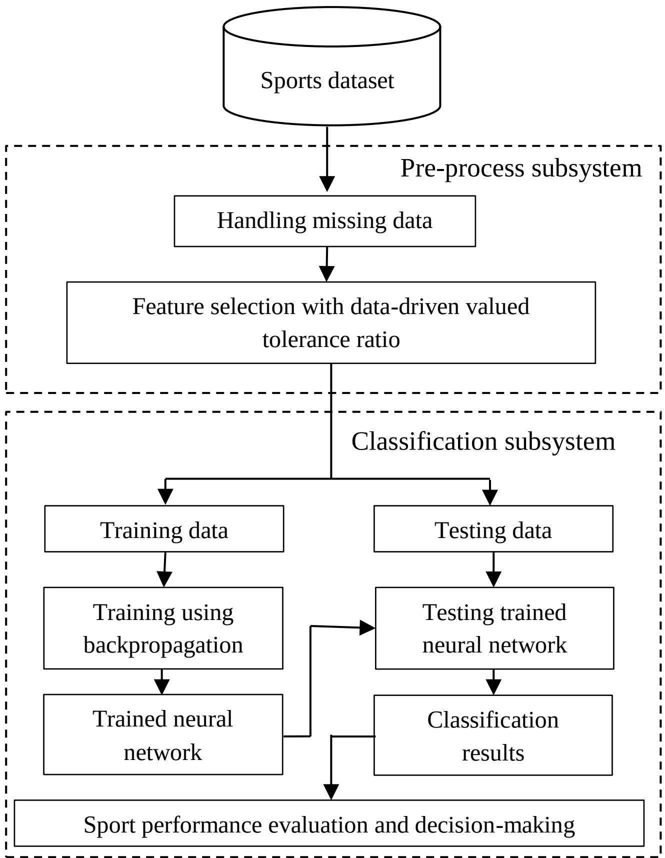


Fig. 2. System architecture.

The proposed system functions by using sports dataset, which consists of four sub-datasets revealing the training and competition objectives for high performance. Missing values of the sports dataset are handled by the mean imputation method. Special attention is given to data reduction based on the novel data-driven valued tolerance relation method for feature selection. The classifier of the sports performance evaluation and decision making will be developed by using a two-step backpropagation neural network. Finally, the classification accuracy and sensitivity will be evaluated.

B. Dataset Description

First of all, the attributes of sports are specified as classification variables. Data are recorded based on four main components of STS: *actors, structure, tasks, and technology constructs.* Using the STS system modelling approach, four datasets are proposed to be used. These datasets include: $\mathcal{D}_1$: sportsmen's expressions, i.e., stress level, medical descriptors; $\mathcal{D}_2$: structure expressions, i.e., field activity, stadium; $\mathcal{D}_3$: task expressions, i.e., athletic training, stamina training; and $\mathcal{D}_4$: technology expressions, i.e., sportswear, sport gadgets [36].

The selection of the most informative attributes for each dataset is carried out using correlation-based feature selection. This method is based on correlation searching for attributes, which have high correlation with class attributes and the smallest possible inter-attribute correlation.

C. Handling Missing Values and Normalisation

Data pre-processing techniques [37] include data cleaning (replacing missing values by manually entering values or using some constant values or using some constant values instead of the missing values), data reduction (discretisation of attribute values and determining attribute informativeness), and data transformation (normalisation, construction of new attributes and attribute aggregation).

Three different types of missing values can be distinguished [38]: a) missing completely at random (MCAR); b) missing random (MAR), and c) not missing at random. When opting for dealing with MCAR, the most commonly used methods are mode imputation, mean imputation, median imputation, most frequent value imputation, and bypassing.

In the proposed study, missing values will be handled as follows. According to the MCAR type, the *mean imputation* will be used. This method replaces the missing values x_i with the mean of the existing values for the corresponding attribute i if the attribute is ordinal, and with mode if the attribute is nominal [38]. Juhola and Lurikalla [39] found good results for classification in two-class datasets with up to 20–30 % missing values, but showed that it was not sensible to use a dataset with higher percentage of missing values.

The data transformation process consists of data normalisation. Data are normalised by using *z-score normalisation with standard deviation.* This method is widely used in data mining because it does not require defining the maximum and minimum limits to attributes [4]. The equation for normalised value of attribute $\grave{A_i}$:

$$A_i^{`} = \frac{A_i - \bar{A}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_I - \bar{A})^2}}, \quad (1)$$

where $\bar{A}$ – the mean value of attribute A;
$A_i^{`}$ – the i-th value of attribute A;
n – a number of records in the dataset.

D. Feature Selection

Attribute reduction is a key concept in the rough set theory. An attribute reduct can be interpreted as a minimal set of attributes that can satisfy some special criteria. In general, there are two strategies of reduct and feature selection: (a) algorithmic approach seeking quick algorithms to compute the reducts; (b) decision approach finding appropriate features for classification problems. The main purpose of feature selection is to reduce the number of features used in the classification while maintaining acceptable classification accuracy [40]. For instance, information gain criterion is one of the commonly used criteria for feature selection.

Feature selection, the process of selecting relevant features, has been widely used in model construction using the rough set theory [41]. In the sports performance approach, a novel *data-driven valued tolerance ratio* (DVT) method for feature selection is proposed based on the idea of data-driven data mining. No additional feature selection is performed and all embeddings are calculated directly from the provided data.

The corresponding thresholds need to be determined before defining an attribute reduct. There are two strategies applied to set threshold value: (a) subjective method where the values are provided by users or domain experts; (b) objective method where the values of the given measures are based on the entire set of attributes [42].

Any information system is $< U, AT >$, where U is a non-empty finite set of objects, called universe, AT is a set of attributes. The tolerance ratio T_B [43] is defined as follows:

$$T_B = \{< x, y > \in U^2 \,|\forall b \in B\} = \\ = *\vee\, b(y) = *\, b(x) = b(y))\}, \quad (2)$$

where B is a subset of the attribute set AT with $B \subseteq AT$. For any x in U, the tolerance class $T_B(x)$ can be calculated, for any y in $T_B(x)$, the tolerance degree $P_B(x, y)$ can be calculated, where $\lambda_{\max}^{x}$ is the maximum, and $\lambda_{\min}^{x}$ is the minimum threshold, i.e.,

$$\lambda_{\max}^{x} = \max\{P_B(x,y)|y \in T_B(x)\}, \quad (3)$$

and

$$\lambda_{\min}^{x} = \min\{P_B(x,y)|y \in T_B(x)\}. \quad (4)$$

For any x in U, there are $\lambda_{\max}^{x}$ and $\lambda_{\min}^{x}$. Accordingly, it can be assumed that

$$\lambda_{\max} = \min\{\lambda_{\max}^{x}|y \in U\}, \quad (5)$$

and

$$\lambda_{\min} = \max\{\lambda_{\min}^{x}|y \in U\}. \quad (6)$$

Therefore, the threshold in the value tolerance ratio can be calculated as follows:

$$\lambda = \min\{\min\{\max\{P_B(x,y)\}, \max\{\min\{P_B(x,y)\}\}\}. \quad (7)$$

Algorithm for threshold selection λ:

Input: the attribute subset $B \subseteq AT$
Step 1: For any x in U calculate its tolerance class $T_B(x)$.
Step 2: For any x in U calculate $\lambda_{\max}^{x}$ and $\lambda_{\min}^{x}$ using (3) and (4).
Step 3: Calculate $\lambda_{\max}$ and $\lambda_{\min}$ using (5) and (6).
Step 4: $\lambda = \min\{\lambda_{\max}, \lambda_{\min}\}$ using (7).

Data-driven valued tolerance (DVT) can be defined as follows:

$$DVT_B^{\lambda} = \{< (x,y) > \in U^2 | P_B(x,y) \geq \lambda\} \cup I_U, \quad (8)$$

where I_U is the identity relation on U.

$$I_U = \{< x, y > |x \in U\}. \quad (9)$$

According to (8) and (9), the lower approximation of an object set X can be calculated as follows:

$$\underline{DVT_B^{\lambda}}(X) = \bigcup_{x \in : DVT_B^{\lambda}(x) \subseteq X} DVT_B^{\lambda}(x), \quad (10)$$

and the upper approximation

$$\overline{DVT_B^{\lambda}}(X) = \bigcup_{x \in X} DVT_B^{\lambda}(x). \quad (11)$$

Considering an incomplete decision table $< U, AT = C \cup D >$, a condition attribute set $B \subseteq C$, then the positive region of the decision table with reference to B is defined, where U / D is a partition of U with reference to D

$$POS_{DVT_B^{\lambda}}(D) = \bigcup_{Y \in U/D} \underline{DVT_B^{\lambda}}(Y). \quad (12)$$

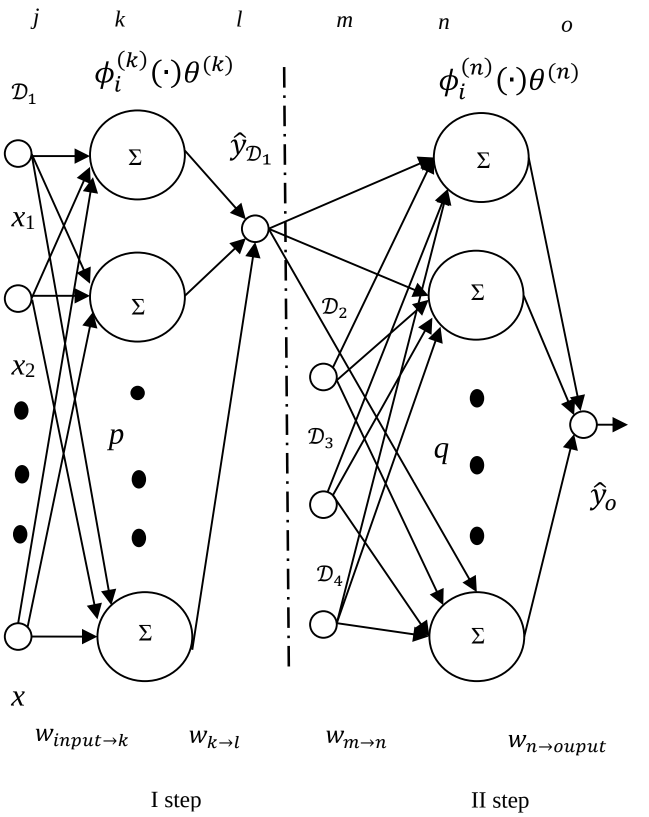
Algorithm for feature selection:

Step 1: Find the lower approximation of each class $\underline{DVT_B^{\lambda}}(X)$ using (10).
Step 2: Find the upper approximation of each class $\overline{DVT_B^{\lambda}}(X)$ using (11).
Step 3: Compute positive region of the set of attributes by means of $POS_{DVT_B^{\lambda}}(D)$ using (12).
Step 4: Find tolerance ratio T_B of a positive region of the set of attributes (2).
Step 5: Select attributes as reduct.

E. Data Classification

The classification process consists of two steps: (a) training dataset; and (b) testing dataset. Cross validation technique is used to randomly portion the pattern set into a training set and a test (validation) set. For the proposed experimental analysis, the pattern set is divided into a training set to train the network and a testing set in the ratio of 70 to 30 [44].

After executing the pre-processing stage that provides clean datasets for both training and testing processes, the data are forwarded to the third classification stage, which is modelled on a feedforward neural network with error backpropagation. In other words, the attributes after the feature selection are connected to the input of backpropagation neural (BPNN) network. The architecture of the proposed BPNN is shown in Fig. 3.



where $\mathcal{D}_1, \ldots, \mathcal{D}_4$ – datasets;
$x_1, \ldots, x_i$ – inputs;
j, *m* – input layers;
k, *n* – hidden layers;
l – intermediate layer;
p, *q* – hidden layers nodes;
$\phi_i^{(k)}, \phi_i^{(n)}$ – the sigmoidal function;
$\theta^{(k)}, \theta^{(n)}$– the bias function;
l, *o* – output layers;
$\hat{y}_{\mathcal{D}_1}$ – intermediate output;
$\hat{y}_o$ – output.

Fig. 3. Architecture of BPNN for sports performance.

The structure of BPNN consists of two steps, which include the following: firstly, the feature selection by calculating the threshold and tolerance ratio; secondly, each unit of k layer receives selected $\mathcal{D}_1$ dataset attributes from the input layer j. There is one hidden layer k with p nodes with sigmoid activation function $\phi_i^{(k)}(\cdot)\theta^{(k)}$. Thirdly, the intermediate layer l receives the results of the classification and forwards them further to Step II where the n layer also receives selected attributes from three other datasets $\mathcal{D}_2$., $\mathcal{D}_3$, and $\mathcal{D}_4$. There is one hidden layer n with q nodes with sigmoid activation function $\phi_i^{(n)}(\cdot)\theta^{(n)}$. Finally, there is layer o where the results of output $\hat{y}_o$ provide with a linear activation function.

F. Classification Accuracy

The classifier accuracy (13), classifier sensitivity (14), and classifier specificity (15) determine and evaluate the correctness of classes classified within the embedding and calculated (for two classes) as follows [45]:

$$\Phi^{Acc} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\ \%, \tag{13}$$

where ϕ^{Acc}– the classifier accuracy;
TP – the number of true positives;
TN – the number of true negatives;
FP – the numbers of false positives;
FN – the number of false negatives.

$$\Phi^{Sen} = \frac{TP}{TP+FN} \times 100\%, \tag{14}$$

where ϕ^{Acc}– the classifier sensitivity, and

$$\Phi^{Spec} = \frac{TN}{TN+FP} \times 100\%, \tag{15}$$

where ϕ^{Acc}– the classifier specificity.

Besides, the receiving operating performance metrics as a graphical measurement tool can be determined.

V. Conclusion

The aim of this study has been to show the applicability of socio-technical system theory and structurally connect the sports and multi-layer artificial neural network domains.

The study has shown that the use of the artificial neural network technology for solving the classification tasks can create realistic models of sports performance prediction based on the previously selected criteria in sports performance related issues.

A narrative review of data mining for sports performance is given. To cover all aspects of sports performance analysis, the task requires data pre-processing and classification.

Further research should empirically validate and:

- determine the most suitable classifier for sports performance evaluation;
- develop a comprehensive system for complex nonlinear sports performance;
- consider and choose the number of neurons in hidden layers;

- extend the proposed methodology to obtain more than two performance classes;
- calculate the accuracy of the neural network and compare it with previous data from the literature.

Finally, it can be considered that the neural network tool can be a good approach in the resolution of complex problems, such as performance modelling and talent identification in a wide variety of sports.

References

[1] A. Field, *Discovering Statistics using SPSS,* 3rd ed. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage Publications, 2009, pp. 585–626.

[2] N. L. Leech., K. C. Barett and G. A. Morgan, *IBM SPSS for Intermediate Statistics,* 5th ed. New York and London: Routledge, Taylor & Francis Group, 2015, pp. 109–143.

[3] J. S. Cramer, *Logit Models from Economics and other Fields,* Cambridge: Cambridge University Press, ch. 9. 2003. [E-book] Available: EBSCO e-book. https://doi.org/10.1017/CBO9780511615412

[4] S. Tuffery, *Data mining and Statistics for Decision Making*, University of Rennes, France: John Wiley & Sons, 2011. https://doi.org/10.1002/9780470979174

[5] S. Russel, and P. Norvig, *Artificial Intelligence,* 3rd ed. UK: Pearson Education, 2014.

[6] K. J. Dooley, "A Complex Adaptive Systems Model of Organizational Change," *Nonlinear Dynamics, Psychology, and Life Sciences,* vol. 1 issue 1, pp. 69–97, Jan. 1997. https://doi.org/10.1023/A:1022375910940

[7] M. W. Craven and J. W. Shavlik, "Using Neural Networks for Data Mining," *Future Generation Computer Systems* special issue on Data Mining. [Online]. Available: http://web.cs.iastate.edu/~honavar/nn7.pdf [Accessed: Sep. 2, 2016].

[8] S. Zambelli and D. A. R. George, Eds., *Nonlinearity, Complexity and Randomness in Economics. Towards Algorithmic Foundations for Economics.* UK: Wiley-Blackwell, 2012. https://doi.org/10.1002/9781118300442

[9] K. Mainzer, "Challenges of Complexity in the 21st Century. An Interdisciplinary Introduction," *European Review*, vol. 17, issue 2, pp. 219–236, 2009. https://doi.org/10.1017/S1062798709000714

[10] W. B. Arthur, *Complexity and the Economy,* Oxford: Oxfords University Press, 2015.

[11] C. Homes, *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems,* University of Amsterdam: Cambridge University Press, 2013. https://doi.org/10.1017/CBO9781139094276

[12] W. Scacchi, "Socio-Technical Design," *The Encyclopaedia of Human-Computer Interaction.* [Online]. Available: https://pdfs.semanticscholar.org/1869/e6b64a18936f5da83773b947986c0d5fb9de.pdf [Accessed: Aug. 12, 2016].

[13] B. Pavard and J. Dugdale, "The Contribution of Complexity theory to the study of socio-technical cooperative systems," Université Paul Sabatier, Toulouse, France. [Online]. Available: http://membres-lig.imag.fr/dugdale/papers/Contribution%20of%20complexity%20theory1.pdf [Accessed: Aug. 12, 2016].

[14] M-H. Chang and J. E. Harrington Jr., "Agent-based Models of Organization," *Handbook of Computational Economics. Agent-based Computational Economics,* Amsterdam, North-Holland: Elsevier, 2006, vol. 2., ch. 26, pp. 1273–1337. https://doi.org/10.1016/S1574-0021(05)02026-5

[15] A. J. Hester, "Measuring Alignment within Relationships among Socio-Technical System Components: A Study of Wiki Technology Use," in *Proceedings of the 50th annual conference on Computers and People Research, SIGMIS-CPR '12,* 2012, pp. 147–155. https://doi.org/10.1145/2214091.2214132

[16] F. Li, J. Qiao, H. Han and C. Yang, "A Self-organizing Cascade neural network with Random Weights for Nonlinear System Modeling," *Applied Soft Computing Journal,* vol. 42, pp. 184–193, May 2016. https://doi.org/10.1016/j.asoc.2016.01.028

[17] T. Mcdermott, W. Rouse, S. Goodman and M. Loper, "Multi-level Modelling of Complex Socio-Technical Systems," *Procedia Computer Science,* vol. 16, pp. 1132–1141, 2013. https://doi.org/10.1016/j.procs.2013.01.119

[18] J. Han and M. Kamber, *Data Mining Concepts and Techniques*, 2nd ed. San Francisco, USA: Morgan Kuafmann, 2011.

[19] C. Vercellis, *Data Mining and Optimization for Decision Making.* Politecnico di Milano, Italy: Wiley, 2009.

[20] R. P. Schumaker, O. K. Solienman and H. Chen, *Sports Data Mining* (Integrated Series in Information Systems). vol. 26, 2010. https://doi.org/10.1007/978-1-4419-6730-5

[21] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis,* 6th ed. New Jersey: Perason Prentice Hall, 2007.

[22] A. Sukovs, L. Aleksejeva, K. Makejeva un A. Borisovs, *Datu ieguve. Pamati.* Riga: Riga Technical University, 2006.

[23] A. J. C. Trappey, C. V. Trappey, A.-C. Chang, F.-L. Lee, H.-C. Hsieh and M.-H. Chao, "Promoting and Positioning Entertainment Artist Using Clustering and Classification Approaches," *International Journal of Electronic Business Management,* vol. 10, no. 4, 2012, pp. 274–285.

[24] B. Ofoghi, J. Zeleznikowa, C. MacMahon and M. Raab, "Data Mining in Elite Sports: A review and a Framework," *Measurement in Physical Education and Exercise Science,* vol. 17, 2013, pp. 171–186. Available: EBSCO. https://doi.org/10.1080/1091367X.2013.805137

[25] S. Provorovs and A. Borisov, "Use of Linear Genetic Programming and Artificial Neural Network Methods to Solve Classification Task," *Information Technology and Management Science,* vol. 49, pp. 133–138, Dec. 2011. https://doi.org/10.2478/v10143-011-0055-9

[26] A. Petrakova, M. Affenzeller and G. Merkurjeva, "Heterogenous versus homogeneous Machine Learning Ensembles," *Information Technology and Management Science,* vol. 18, pp. 135–140, Dec. 2015. https://doi.org/10.1515/itms-2015-0021

[27] K.-L. Du and M. N. S. Swamy, *Neural Networks and Statistical Learning,* London, UK: Springer, 2014. https://doi.org/10.1007/978-1-4471-5571-3

[28] M. F. Musso, E. Kyndt, E. C. Cascallar and F. Dochy, "Predicting General Academic Performance and Identifying the Differential Contribution of Participating Variables Using Artificial Neural Networks," *Frontline Learning Research,* vol. 1, issue 1, pp. 42–71, 2013. https://doi.org/10.14786/flr.v1i1.13

[29] L. L. S. Linhares, A. I. R. Fontes, A. M. Martins, F. M. U. Araújo, and L. F. Q. Silveira, "Fuzzy Wavelet Neural Networks Using a Correntropy Criterion for Nonlinear System Identification," *Mathematical Problems in Engineering,* 2014. https://doi.org/10.1155/2015/678965

[30] M. Pfeiffer and A. Hohman, "Applications of Neural Networks in Training Science," *Human Movement Science,* vol. 31, issue 2, pp. 344–359. https://doi.org/10.1016/j.humov.2010.11.004

[31] D. Memmert and J. Perl, "Game Creativity Analysis Using Neural Networks," *Journal of Sports Science,* vol. 27, issue 2, pp. 139–149, 2009. https://doi.org/10.1080/02640410802442007

[32] K. D. Maier, V. Wank, K. Bartonietz and R. Blickhan, "Neural network based models of javelin flight: prediction of flight distances and optimal release parameters," *Sports Engineering*, vol. 3, issue 1, pp. 57–63, Feb. 2000. https://doi.org/10.1046/j.1460-2687.2000.00034.x

[33] A. Dut-Mazumder, C. Button, A. Robins and R. Bartlett, "Neural Network Modelling and Dynamical System Theory," *Sports Medicine,* vol. 41, issue 12, pp. 1003–1017, Dec. 2011. https://doi.org/10.2165/11593950-000000000-00000

[34] A. J. Silva, A. M. Costa, P. M. Oliveira, V. M. Reis, J. Saavedra, J. Perl, A. Rouboa and D. A. Marinho, "The Use of Neural Network Technology to Model Swimming Performance," *Journal of Sports Science & Medicine*, vol. 6, issue 1, pp. 117–125, March 2007. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3778687/ [Accessed: September 10, 2016].

[35] S. R. Iyer and R. Sharda, "Prediction of athletes performance using neural networks: An application in cricket team selection" *Expert Systems with Applications,* part 1, vol. 36 issue 3, pp. 5510–5522. https://doi.org/10.1016/j.eswa.2008.06.088

[36] L. Grinčikaitė-Samuolė, "Attitudes of High-Level Female Sprinters Towards Factors Influencing the Training System," *Baltic Journal of Sports and Health Science,* vol. 100, issue 1, pp. 10–16, 2016.

[37] A. Kirshners, I. Polaka and L. Aleksejeva, "Gastric Cancer Risk Analysis in Unhealthy Habits Data with Classification Algorithms," *Information Technology and Management Science*, vol. 18, pp. 97–102, Dec. 2015. https://doi.org/10.1515/itms-2015-0015

[38] R. Janssen, P. Spronck, P. Dibbets and A. Arntz, "Frequency Ratio: A Method for Dealing with Missing Values within Nearest Neighbour Search," *Journal of System Integration*, vol. 6, no. 3, 2015. https://doi.org/10.20470/jsi.v6i3.233

[39] M. Juhola and J. Laurikkala, "Missing Values: How Many can They be to Preserve Classification Reliability?" *Artificial Intelligence Review*, vol. 40, issue 3, pp. 231–245, 2013. https://doi.org/10.1007/s10462-011-9282-2
[40] X. Jia, L. Shang, B. Zhou and Y. Yao, "Generalized Attribute Reduct in Rough Set Theory," *Knowledge-Based Systems*, vol. 91, pp. 204–218, Jan. 2016. https://doi.org/10.1016/j.knosys.2015.05.017
[41] J. H. Miller, "Active Nonlinear Tests (AnTs) of Complex Simulation Models," *Management Science,* vol. 44, no. 6, pp. 820–830, June 1998. https://doi.org/10.1287/mnsc.44.6.820
[42] G. Wang, W. Wu, L. Guan and F. Hu, "Data-driven Valued Tolerance Relation Based on the Extended Rough Set," *Fundamenta Informaticae*, vol. 132, no. 3, pp. 349–363, 2014. https://doi.org/10.3233/FI-2014-1048
[43] M. Kryszkiewicz, "Rough Set Approach to Incomplete Information Systems," *Information Science,* vol. 112, pp. 39–49, Dec. 1998. https://doi.org/10.1016/S0020-0255(98)10019-1
[44] K.-S. Leung, K. H. Lee, J.-F. Wang et al. "Data Mining on DNA Sequences of Hepatitis B Virus," *Transactions on Computing Biology and Bioinformatics,* vol. 8, no. 2, pp. 428–440, March/April 2011. https://doi.org/10.1109/TCBB.2009.6
[45] G. Lee, D. E. R. Bucheli and A. Madabhushi, "Adaptive Dimensionality Reduction with Semi-Supervision (AdReSS): Classifying Multi-Attribute Biomedical Data," PLoS ONE, vol. 11, no. 7, July 2016. https://doi.org/10.1371/journal.pone.0159088

Ivars Namatēvs holds Mg. sc. ing. degree from Riga Technical University and *MBA* degree from Riga Business School. His research interests include data mining methods, artificial neural networks and their application, as well as agent-based modelling, genetic algorithms and support vector machines.
The most important previous publications: I. Namatēvs. (2015). Concept analysis of complex adaptive systems. *International Scientific Forum: Proceedings of XVI International Scientific Conference: Towards Smart, Sustainable and Inclusive Europe: Challenges for Future Development.* Riga, Latvia, 28–30 May 2015.
E-mail: ivars@turiba.lv

Ludmila Aleksejeva received her Dr. sc. ing. degree from Riga Technical University in 1998. She is an Associate Professor at the Department of Modelling and Simulation, Riga Technical University. Her research interests include decision making techniques and decision support system design principles, as well as data mining methods and tasks, and especially collaboration and cooperation of the mentioned techniques.
The most important previous publications: Gasparovica-Asite, M., Polaka, I., Aleksejeva, L. (2015). The impact of feature selection on the information held in bioinformatics data. *Information Technology and Management Science*. vol. 18, pp. 115–121. Available from: doi:10.1515/itms-2015-0018. Plinere, D., Borisovs, A., Aleksejeva, L. (2015). Interaction of software agents in the problem of coordinating orders. *Automatic Control and Computer Sciences (AC&CS).* vol. 49, issue 5, pp. 268–276. Available from: doi: 10.3103/S0146411615050089. Aleksejeva, L., Užga-Rebrovs, O., Borisovs, A. (2012). *Fuzzy Classification and Clustering. Textbook*. Riga: RTU Press, 248 p.
E-mail: ludmila.aleksejeva_1@rtu.lv

Inese Poļaka received her Dr. sc. ing. degree from Riga Technical University in 2014. She works as a Lecturer at the Institute of Information Technology of Riga Technical University (Latvia) and Leading Researcher at the Faculty of Medicine, University of Latvia. Main research interests include data mining, machine learning, classifiers, evolutionary algorithms and their applications, as well as bioinformatics and biostatistics.
Some of her most important previous publications: Bērziša, S., Poļaka, I., Šūpulniece, I., Grabis, J., Ozoliņš, E., Meiers, E. (2016.) Method for decomposition of monolithic enterprise applications. *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, United States of America, New York, 4–8 April 2016. Pisa: ACM, pp. 1210–1213. ISBN 978-1-4503-3739. Poļaka, I., Gasparoviča-Asīte, M., Borisovs, A. (2014) Genetic algorithm and tree based classification in bioinformatics. *MENDEL 2014: 20th International Conference on Soft Computing: Proceedings*. Vol. 20, Czech Republic, Brno, 25–27 June 2014. Brno: Brno University of Technology, pp. 21–26. ISSN 1803-3814.
E-mail: inese.polaka@rtu.lv