PREDICTING COMPETITIVE SWIMMING PERFORMANCE

Robert Wilk,[1, A, B] Olga Fidos-Czuba,[2, B] Łukasz Rutkowski,[2, B, D] Krzysztof Kozłowski,[2, B] Piotr Wiśniewski,[2, B, D] Adam Maszczyk,[3, D] Arkadiusz Stanula,[3, D] Robert Roczniok[3, A, B, C, D]

[1] Department of Water Sports, The Jerzy Kukuczka Academy of Physical Education, Katowice, Poland
[2] Phd student. The Jerzy Kukuczka Academy of Physical Education, Katowice, Poland
[3] Department of Statistics and Methodology, The Jerzy Kukuczka Academy of Physical Education, Katowice, Poland
[A] Study Design; [B] Data Collection; [C] Statistical Analysis; [D] Manuscript Preparation

Address for correspondence:
Robert Roczniok
The Jerzy Kukuczka Academy of Physical Education, Department of Statistics and Methodology, Katowice, Poland,
Mikołowska 72a, 40-065 Katowice, Poland
E-mail: r.roczniok@awf.katowice.pl

Abstract. The aim of this study was to present the results of analyses conducted by means of complementary analytic tools in order to verify their efficacy and the hypothesis that Kohonen's neural models may be applied in the classification process of swimmers. A group of 40 swimmers, aged 23 ±5 years took part in this research. For the purpose of verification of usefulness of Kohonen's neural models, statistical analyses were carried out on the basis of results of the independent variables (physiological and physical profiles, specific tests in the water). In predicting the value of variables measured with the so called strong scale regression models, numerous variables were used. The construction of such models required strict determination of the endogenous variable (Y – results for swim distances of 200 m crawl), as well as the proper choice of variables in explaining the study's phenomenon. The optimum choice of explanatory variables for the Kohonen's networks was made on the grounds of regression analysis. During statistical analysis of the gathered material neural networks were used: Kohonen's feature maps (data mining analysis). The obtained model has the form of a topological map, where certain areas can be separated, and the map constructed in this way can be used in the assessment of candidates for sports training.

Key words: Kohonen feature map, swimming performance, sports selection, regression

Introduction

Dynamical systems theory (DST) has the power to potentially unify existing sub-disciplines such as sports biomechanics, notational analysis, motor control, physiology and psychology under one macroscopic platform (Glazier 2010). Stergiou et al. (2004) proposed several nonlinear tools to study specific features of complex human movements, e.g. Lyapunov exponent (LyE), correlation dimension (CoD), approximate entropy (ApEn) and ANNs. Such tools, which operate with few assumptions about the structure of the dataset, are undoubtedly useful and their application to complex datasets will become increasingly widespread in the future (Dutt-Muzumder et al. 2011).

Many researchers confirm the growing significance of the process of modeling with regard to the utilization of Artificial Neural Networks (ANNs) in the optimization of selection and training processes. The occurrence of linear and non-linear relationships between variables led scientists to develop Artificial Neural Networks for modeling and prediction (Haykin 1994; Zadeh 2002; Mester and Perl 1999; Perl et al. 2002).

The results of conducted research confirm the findings of Bartlett 2006, Maszczyk et al. (2011) concerning the usage of perceptron networks in classification of athletes. Maszczyk et al. (2012) showed the large usefulness of non-linear neural models in the process of prediction, which was also confirmed in this research. The use of clustering which uses Kohonen's network can be very useful from the practical point of view – in spite of the rather incomprehensive (as not directly defined) operating objectives of this network and despite the

lack of forced operating direction. Due to the protection made by this network we are able to better understand data, which in turn gives the possibility of improving the process of its further analysis (Roczniok et al. 2007).

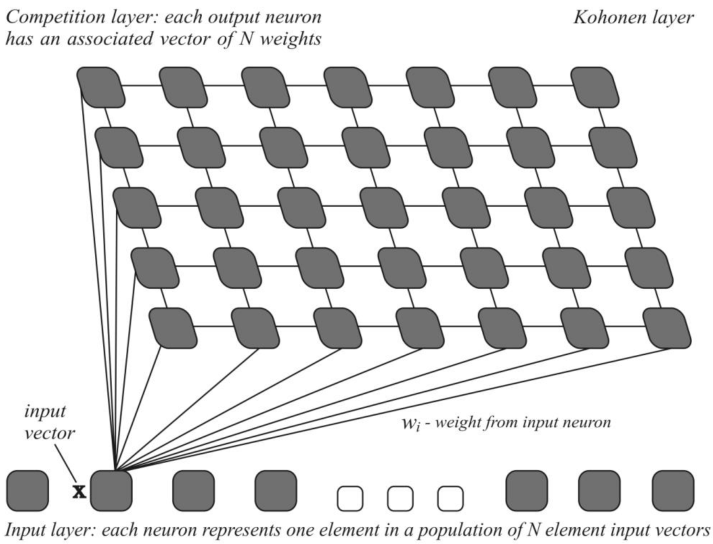

Figure 1. Kohonen feature map of input patterns

Self organizing maps are a class of the ANN model, based on a method called competitive learning, where the output nodes compete amongst themselves to be activated on a per-group basis. For presented inputs, the output node that wins the competition is called a 'winnertakes- all neuron' (Haykin 1999). Each node represents process types (section of a match), and each cluster represents a class of a similar process type. Based on competitive learning, the architecture of the network can be fabricated to develop its model such as that of a KFM and the Willshaw and von der Malsburg model (Konen et al. 994). The term KFM comes from the capability to recognize patterns or clusters in the data without supervision/target data (Kohonen 1997). A KFM is an essential tool for analysing dynamical movement patterns in sports. The KFM architecture compresses surplus high-dimensional inputs to a low-dimensional structure (e.g. one- or two-dimensional [2-D]). Dimensionality reduction is performed to recognize and validate structures visually, yet preserve nonlinear topological relationships in the data sets (Perl and Dauscher 2006). This helpful feature retains the relevant information and discards irrelevant information in high-dimensional datasets, which is typical of dynamical systems. They consist of an 'input layer' and a presumed 'competition layer' (Figure 1). The weights of the connections from the input nodes to a single node in the competition layer are interpreted as a reference vector in the input space, i.e. a self-organizing map represents a set of vectors in the input space and one vector for each node in the competition layer (Dutt-Mazumder et al. 2011).

The aim of this study was to present the results of analyses conducted by means of complementary analytic tools in order to verify their efficacy and the hypothesis that Kohonen's neural models may be applied in the classification process of swimmers.

Methods

Participants

A group of 40 swimmers, aged 23 ±5 years took part in this research. Measurements were performed in the Human Performance Laboratory of the Academy of Physical Education in Katowice and at the swimming pool. In predicting the value of variables measured with the so called strong scale regression models, numerous variables were used. The construction of such models required strict determination of the endogenous variable (Y – results of the 200 m crawl), as well as the proper choice of variables in explaining the study's phenomenon. The optimum choice of explanatory variables for the Kohonen's networks was made on the grounds of regression analysis. During statistical analysis of the gathered material neural networks were used: Kohonen's feature maps (data mining analysis). All study procedures were approved by the Bioethics Committee for Scientific Research at the Academy of Physical Education in Katowice. The research subjects were informed of the aim of the study and experimental risks. All statistical analyses were carried out on a PC using the statistical package STATISTICA 10.0.

Data collection and tools of statistical analyses

Methods of the model experiment and direct observation were used during the research. The structure of the following variables was used: RXn^nY^n, with one multi-valued dependent variable (Y^n), and n multivalent independent variables (Xn^n) taking into consideration the principle of randomization (R). The results of the 200 m crawl was the dependent variable (Y) in all tests. Independent variables for multivariate analysis were obtained by measuring the different characteristics of swimmers in the following groups: anthropometric measurements, body mass and body composition were then evaluated by electrical impedance (Inbody 720, Biospace Co., Japan). Two hours after a light breakfast, a ramp cycloergometer test (0,5W/s) was administered to determine aerobic capacity. During the test, heart rate, minute ventilation (VE), oxygen uptake (VO_2) and expired carbon dioxide (CO_2) were continuously measured using a MetaLyzer 3B-2R stationary spiroergometer (Cortex, Germany). Fingertip capillary blood samples for the assessment of lactate (LA) concentration (Biosen C-line Clinic, EKF-diagnostic GmbH, Germany) were drawn before the test, as well as during the 3rd, 6th, 9th, and 12th min of recovery. In the second day of evaluation the Wingate test (0.45 Nm/kg) was administered to determine anaerobic capacity. Fingertip capillary blood samples for the assessment of lactate (LA) concentration were drawn before the test, as well as in the 4rd, 8th, min of recovery. Resting blood samples were drawn from the antecubical vein to determine hematological variables (hemoglobin concentration (HGB), haematocrit value (HCT), number of erythrocytes (RBC) (Advida 2120, Siemens, Germany)). All together there were 30 independent variables and one dependent variable Y – result of the 200 m crawl swim.

Results

The analysis of the signs showed that a higher level (VO_2max, Realtive Peak Power, Relative Mean Power, haemoglobin, erythrocyte number) results were significantly related distances of the 200 m crawl swim. With the increase of BMI and Percent Body Fat values the swimmers obtain worse results at the distance of 200 m.

Table 1. Structural parameters of the regression equation for the dependent variable Y – result for the 200 m crawl swim

Variable	beta	St. error beta	B	St. error beta	t	p
Intercept			97.77	15.47	6.32	0.00003
Relative PeakPower (W/kg)	–2.17	0.49	–4.69	1.05	–4.46	0.00100
Erythrocyte (mln/ul)	–0.26	0.07	–4.88	1.23	–3.96	0.00200
Hemoglobin (g/dl)	–0.35	0.12	–1.76	0.57	–3.08	0.00900
PBF %	0.21	0.09	0.19	0.08	2.36	0.03500
Realtive VO_2max (ml/min/kg)	–0.75	0.22	–0.40	0.11	–3.47	0.00400
MeanPower (W/kg)	1.23	0.43	4.21	1.46	2.89	0.01300
BMI	0.30	0.13	0.55	0.24	2.27	0.04100

Kohonen Feature Map Learning Algorithm

The aim of a KFM is to activate different nodes of the network to respond similarly to inputs (Haykin 1999). Node weights are initialized to small random values or sampled evenly by the two largest principal component eigenvectors. An initial neighborhood radius is defined and the subsequent distance between each input and the output node is computed, according to the given equation (Perl and Dauscher 2006):

$$d_i = \sum_{i=1}^{N-1} \left\{ x_i(t) - w_{ij}(t) \right\}^2,$$

where $x_i(t)$ = input to node i at time t and $w_{ij}(t)$ = weight from input node i to output node j at time t; N = neighbouring nodes.

For each input vector the ‘winning node’ is determined. There are two methods; maximal dot product: $i^* = \text{argmax}\,(w_i, x)$; minimal Euclidean distance: $i^* = \text{argmin}\,(w_i, x)$. The KFM training phase consists of input vectors (e.g. coordinates of players) in a random order. The KFM training consists of weight updates of ‘winner node’ i* and its neighbouring nodes in N_{i^*i} (t) are updated according to the learning equation (Dutt-Muzumder et al. 2011).

$$w_{\downarrow}ij(t + 1) = w_{\downarrow}ij(t) + \eta(t) \times N_{\downarrow}(i^*i)\ (t) \times [x_{\downarrow}i(t) - w_{\downarrow}ij(t),$$

where $w_{ij}(t)$ = initial weight; $w_{ij}(t + 1)$ = updated weight; $\eta(t)$ = learning rate, which decreases with time and $N_{i^*i}(t)$ = neighbourhood function (Lippmann 1987).

The neighbourhood function can be symmetric (e.g. rectangular and Gaussian) or anti-symmetric:

$$ü_{ü}\ (\ ) = \begin{cases} 1, & \text{if} \quad d_M(i * i) \le \lambda(t) \\ 0, & \text{otherwisw} \end{cases}$$

where the rectangular neighbourhood function is based on the Manhattan distance between nodes d_M (i*i)where λ (t) is the neighbourhood parameter. The Gaussian neighbourhood function:

$$\frac{-d_E^2(i^*, i)}{\lambda^2(t)},$$

where the neighbourhood function is based on the Euclidean distance between nodes d_E (i*, i). The Gaussian function is preferred over the rectangular, since it gives smoother mapping from input points to weight coordinates (Leonedes 1998). Regardless of the neighbourhood functional form, both shrink with time. The training can be stopped after it undergoes a number of predefined iterations (t_{max}) defined by $t = t_{max}$.

Results

Basing on calculated regression equation coefficients (st. error B), it can be stated that such qualities as: Relative Peak Power, Relative Mean Power, Realtive VO_2max, Erythrocyte number, Hemoglobin, BMI, PBF % have considerable influence on predicting the dependent variable Y – result of the 200 m crawl swim for the studied subjects. If the value of Relative Peak Power variable decreases by one point (one unit of this quality), then the swim result at 200 m will improve by 4.69 s, assuming that the remaining variables remain unchanged. Basing on variables obtained by regression analysis, the Kohonen's model was constructed.

The colour saturation scale was used in addition to numbering the neurons representing particular cases (Figure 2). This scale was created upon identifying the map areas by means of familiar results of the dependent variable Y – result of the 200 m crawl (very good: below 1 : 53.50; average: 1 : 53.50 – 1 : 59.99 and poor: above 1 : 59.99). The classification table (Figure 2) created on the basis of learning sets for Y – result of the 200 m crawl swim, in the Kohonen's network can be useful in assessing new objects, not presented during learning. This results from evident, clear ordering of the explained variable value corresponding to particular neurons - one point to large areas of the map containing neighbouring neurons corresponding to approximate values of the explained variable. This indicates that there is a relation between the assumed set of explanatory variables and the explained variable. One can notice that if at the beginning of the study our candidate is classified to neurons with the biggest colour saturation, it may be expected that the candidate will be among the best swimmers in his group. If, on the other hand, our candidate is classified to neurons with the lowest colour saturation, then one can expect this subject to be among the athletes that reach the worst results in that age group. In particular neurons, the neuron point average was also given, which served as the basis for map area identification.

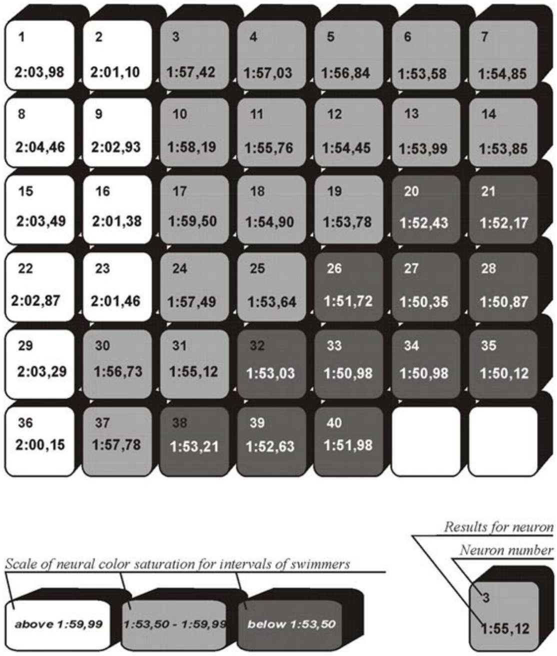

Figure 2. Structure of the studied set of swimmers obtained by means of the Kohonen's map – Y sports results

Discussion and conclusions

Recent theoretical contributions to the theory of talent in sport have clearly shown that a complex and longitudinal framework is necessary to successfully address talent identification and the talent promotion issue in most sports (Gagne 1985; Abot and Collins 2004). The nature of recruitment and selection of an athlete consists in finding the vector of the candidate's abilities with respect to each stage of sports training. Therefore, the selection process can be optimized by creating a large source of information on a candidate's sport abilities with as few examined features as possible, using a regression model and a neural model (Maszczyk et al. 2012). As talent development is a complex non-linear process, and the different components of early talent make-up not only change over time, but can also mutually suppress or enhance each other, linear models like discriminate analysis can only approximate the non-linear talent development within a very small range of the future performance output. Because of this, neural networks also seem to be appropriate tools for talent detection purposes (Philippaerts et al. 2008). Studying dynamical systems using equations pertaining to perturbations has some practical disadvantages, since these equations are practically limited to weak nonlinearities (Beek and Beek 1998). To overcome this limitation, graphical methods such as Kohonen feature maps (KFMs) to analyze nonlinear behavior have had an increasing impact (Dutt-Muzumder et al. 2011). Due to their pattern detection ability, such methods as e.g. the Self-organizing Kohonen Feature Map may allow to predict the future success of talents by revealing distinct patterns in the individual sets of sport specific dispositions.

The basic objective of this study was to establish Kohonen's neural models assisting in the classification process in swimming. The use of clustering where the Kohonen's network is applied can be very useful from the practical point of view – in spite of the rather incomprehensive (as not directly defined) operating objectives of this network and the lack of the forced operating direction. Due to the projection made by this network we are able to better understand the data, which in turn gives the possibility of improving the process of its further analysis. Based upon such great possibilities of data mining analysis, the Kohonen's network can be also used in the selection of candidates for competitive sports. The results of the conducted study confirmed Bartlett's (2006), Tidow's (2000), Lees (2002) and Murakami's et al. (2005) findings concerning the use of neural networks in classification of sports results.

By recognizing the clusters existing in the multidimensional data vector names can be attributed to them. Thus, the Kohonen's network acquires the possibility of their classification according to internal logic of the data itself, and not on the grounds of arbitrary criteria. Based upon such great possibilities of data mining analysis, we can use the Kohonen's network also in the selection of candidates for competitive sports training. The obtained model has the form of a topological map, where certain areas can be separated. Upon determining map areas corresponding to particular athletes or their groups, we can identify the determined map areas. Therefore, it is necessary to specify the average sports development level achieved by the athletes represented by each of the neurons. The map constructed in this way can be used in the assessment of candidates for sports training. On the grounds of a characteristic, a subject is assigned to one of the determined classes. One can expect that the candidate will achieve the level of sports development similar to the average development level of subjects qualified to this group during network learning. In conclusion, models based on Kohonen's networks showed that by use of independent variables, they could accurately group subjects into categories which after a year achieve very good, average or poor results. This implies that these models may be used for data mining analysis, which aims at assisting in the recruitment process of candidates for the javelin throw. Roczniok et al. (2007) conducted a study on a group 140 swimmers from the Silesia Macroregion in Poland found that models based on Kohonen's networks showed that by use of independent variables, they could accurately group subjects into categories which after a year, achieve very good, average and very weak performances. This implies that these models may be used for data mining analysis, which is aimed at assisting recruitment of candidates for sport swimming. Hohman and Seidel (2003) show that neural networks are able to recognize global patterns of different talent make-ups, they are a worthwhile tool in the detection of talents under the condition of non-linear talent development. Hence, from a dynamical systems point of view, a successful neural network modeling may be interpreted as a representation of deviations of the different states of the system from equi-probability, in our case the identification of swimmers performance. This is a very interesting aspect of the modeling of competitive performances, because the non-linear dynamic systems perspective is rapidly emerging as one of the dominant meta-theories in the natural sciences, and there is also reason to believe that in the future it will eventually provide a more general integrative understanding in training science, as well.

This objective of this paper was to determine whether the talent development outcome can be modeled by means of the nonlinear mathematical method of artificial neural networks.

We conclude that the results of this research confirm that Kohonen's Feature Map can be used for optimization of swimmers classification and prediction of sports results. On the grounds of analysis of standardized Beta values one can say that the biggest predictability for the distance of the javelin throw was shown by such qualities as: Relative Peak Power, Relative Mean Power and Relative VO_2max.

References

Abbot A., Collins D. A theoretical and empirical analysis of a 'State of the Art' talent identification model. High Ability Studies. 2012; 13 (2): 157–178.

Abbot A., Collins D. Eliminating the dichotomy between theory and practice in talent identification and development: considering the role of psychology. Journal of Sport Sciences. 2004; 22 (5): 395–408.

Bartlett R. Artificial intelligence in sports biomechanics. New dawn or false hope? Journal of Sports Science and Medicine. 2006; 5: 474–479.

Beek P.J., Beek W.J. Tools for constructing dynamical models of rhythmic movement. Hum Mov Sci. 1998; 7: 301–342.

Dutt-Mazumder A., Button C., Robins A., Bartlett R. Neural Network Modelling and Dynamical System Theory. Sports Med. 2011; 41 (12): 1003–1017.

Gagne F. Giftedness and talent: Reexamining a reexamination of the definitions. Gifted Child Quarterly. 1985; 19: 103–112.

Glazier P.S. Game, set and match? Substantive issues and future directions in performance analysis. Sports Med. 2010; 40 (8): 625–34.

Haykin S. Neuralnetworks, a comprehensive foundation. Macmillan College Publishing Company. New York 1994.

Haykin S. Neural networks: a comprehensive foundation. 2nd ed. Prentice-Hall Inc. Upper Saddle River (NJ) 1999.

Hohmann A., Seidel I. Scientific Aspects of Talent Development. International Journal of Physical Education. 2003; 40 (1): 9–20.

Kohonen T. Self-organizing maps. New York: Springer 1997.

Konen W., Maurer T., Von der Malsburg C. A fast dynamic link matching algorithm for invariant pattern recognition. Neural Networks. 1994; 7: 1019–1030.

Lees A. Technique analysis in sports: a critical review. Journal of Sports Sciences. 2002; 20: 813–828.

Leondes C.T. Algorithm and architectures. San Diego (CA); Academic Press 1998.

Lippmann RP. An introduction to computing with neural nets. IEEE ASSP Magazine 1987; 4 (3): 422, http://hawk.cs.csuci.edu/William.Wolfe/UCD/engineering/cse/Graduate/courses/CSC5542/ Lippmann.pdf (21.09.2011).

Maszczyk A., Zając A., Ryguła I. A neural network model approach to athlete selection. Sports Engineering. 2011; 13 (2): 83–93.

Maszczyk A., Roczniok R., Waśkiewicz Z., Czuba M., Mikołajec K., Zając A., Stanula A. Application of Regression and Neural Models to Predict competitive swimming performance. Perceptual and Motor Skills. 2012; 114 (2): 610–626.

Mester J., Perl J. Unconventional simulation and empirical evaluation of biological response to complex high training loads. Sports Science, Rome 1999.

Murakami M., Tanabe S., Ishikawa M., Isolehto J., Komi V., Ito A. Biomechanical analysis of the javelin throwing at 11th World Championships in Athletics in Helsinki. International Association of Athletics Federation. New Studies in Athletics. 4. 2005.

Perl J., Lames M., Glitsh U. Modellbildung in der sport-wissenschaft. Beitrage zr Lehr und Forchung im Sport, Bd 132, Schorndorf, Hofmann. 2002.

Perl J., Dauscher P. Dynamic pattern recognition in sport by means of artificial neural networks. In: Begg R. Palaniswami M. editors. Computational intelligence for movement science. Hershey (PA): Idea Group Publishing. 2006: 299–318.

Philippaerts R.M, Coutts A., Vaeyens R. Physiological Perspectives on the Identification and Development of Talented Performers in Sport. In: Talent Identification and Development. The Search for Sporting Excellence, eds. R. Fisher, R. Bailey. Berlin: ICSSPE, 2008: 49–67.

Roczniok R., Ryguła I., Kwaśniewska A. The use Kohonen's neural networks in the recruitment process for sport swimming. Journal of Human Kinetics. 2007; 17: 75–88.

Stergiou N., Buzzi L., Kurz MJ. Nonlinear tools in human movement. In: Innovative analysis of human movement, ed. N. Stergiou. Champaign (IL): Human Kinetics. 2004: 66–77.

Tidow G. Challenge decathlon – barriers on the way to becoming the "king of athletes". Part I. International Association of Athletics Federation. New Studies in Athletics. 2000; 15.

Zadeh L. From computing with numbers to computing with words. International Journal of Applied Math and Computer Science. 2002; 12 (3): 307–324.

Cite this article as: Wilk R., Fidos-Czuba O., Rutkowski Ł., Kozlowski K., Wiśniewski P., Maszczyk A., Stanula A., Roczniok R. Predicting Competitive Swimming Performance. Central European Journal of Sport Sciences and Medicine. 2015; 9 (1): 105–112.