Journal of Human Kinetics volume 17, 2007, 75-88

The Use of Kohonen's Neural Networks in the Recruitment Process for Sport Swimming

by

Robert Roczniok[1], Igor Ryguła[2], Anna Kwaśniewska[3]

Being a long-term and complex process, contemporary sport training requires continuous inflow of information conditioning its effectiveness. This information is used in training control, understood as the issue of decisions based on information concerning the condition of the control process and knowledge of control objective (Raczek 1989). There are many factors influencing the effectiveness of the training process, and the sport selection process is one of them. Its essence consists in determining the talent vector of the candidate for sport training, which allows for increasing the effectiveness of sport training (Ryguła 2002).

Key words: *neural networks, sport selection, Kohonen's maps*

1 - Department of Methodology and Statistics, Academy of Physical Education in Katowice
2 - Department of Theory of Sport, Academy of Physical Education in Katowice
3 - Student of doctoral studies at the Medical University of Silesia

Introduction

Instruments which would allow to define personal predispositions for achieving high sport performance have been sought now for many years (Kozioł 1984). Successive phases of training different selection criteria, which are to help in determining the contestant's chances to achieve high performance (Płatonow 1997, Raczek 1989) are being applied. Statistical and mathematical forecasting methods (Komor 1982, Mester, Perl, 1999, Perl and et al. 2002) are becoming more and more significant in this area. Multidimensional exploration techniques, which are only sporadically used in the area of sport science can be included in them. They are connected with recognition of objects and their grouping (Taxonomic analysis) Witt, Burton 1996), concentration analysis, factor analysis and discriminatory analysis (Ryguła 2003).

By using multidimensional exploration techniques for optimizing the recruitment process, we have the possibility of grouping objects or grouping qualities. In both cases we find subsets of more homogenous taxonomic units. In relation to qualities, the subsets carry similar information on contestants (Ryguła 2003). In consequence, we have the possibility of identifying certain types of contestants, which can lead to the optimization of recruitment and which will allow the use of different training loads.

The basic purpose of the study was to verify the usefulness of Kohonen's neural network in assisting the sport swimmer recruitment process for the 50m sprinter and 800 m endurance events.

Objective of the study

The basic objective of this study was to establish neural form models assisting the recruitment process in sport swimming. The detailed objective consisted of creating an application based on neural networks which would form the basis for selecting swimmers that would achieve categorical performances of very good, good and poor a taxonomic division within homogenous subgroups.

Study queries

Accomplishment of the foregoing objectives implicates the formulation of the following study queries:

1. Which of the analysed qualities have the biggest prognostic value?
2. What will the structures be like of constructed neural models effecting exploratory data analysis?

3. How to use the obtained neural models?

Study Hypothesis

The content of formulated study queries lead to assuming the following study hypothesis:

H: One assumes the hypothesis that on the grounds of the contents contained in Ryguła's scientific theories and connection systems, as well as gathered and accurately described experimental data, it is possible to create an application that will help in making the selection for the training of juniors in sport swimming.

Material and Methods

In order to obtain an answer to the study queries posed earlier and to verify the formulated hypothesis, versatile, statistical analysis of junior swimmers' measurement results had been carried out. Results obtained by a group of 140 swimming contestants from the Silesia Macroregion aged 10 ± 0.7 years formed the material for analysis. Study objects were collected by using intentional selection technique, assuming the criterion of learning three sport swimming styles correctly.

The main studies were preceded by a year's fitness preparation . The study group participated in training sessions three times a week - twice in the swimming pool (45 minutes) and once in the gym (90 minutes). The session were of general and complementary nature. The purpose of the occupations at the swimming pool was to learn swimming.

The conducted studies had the nature of empirical and exploratory analyses of experimental charakter. For this purpose, experiment methods and direct participatory observation were used (Ryguła 2003). Differences in the structure of training loads formed the experimental factor. Optimum control was effected in conformity with procedure described in the works by Cholewa (1998) and Ryguła (2000). The study problem was solved on the grounds of empirical and prognostic study types, by using the model of statistical tests, that is two multiple-valued dependent variables Y (50 m (scores freestyle), 800 m (scores freestyle), n independent multiple-valued variables –Xn" Y2".

The constructed model did not take into consideration the influence of the biological development level on sport training effects. In assessing the potential possibilities of candidates the assumption was accepted on the specific nature of sport training in swimmers carried out according to the methodology described in the literature (Colwin 1992; Costill et al. 1995; Płatonow 1997).

Techniques and Tools

Looking for the answer to formulated study queries, measurements were made of many qualities that will fulfil the role of independent variables. Such was the role of measurement results of properties referring to body build qualities, general and special physical fitness. Performance on 50 m and 800 m distances converted into scores according to triathlon charts (measurements made after one year's training) are the explained variables (quality indexes). The numerical values of all variables were determined during measurement of the following qualities:

Independent variables: all variables were made after one year's fitness preparation when recruiting for sport swimming.

1. 50 m (points) - performance on a distance of 50 m with the front crawl stroke converted into points according to triathlon charts
2. <u>800m (points)</u> - performance on a distance of 800 m with front crawl stroke converted into points according to triathlon charts
3. <u>25 m (s)</u> - performance on a distance of 25 m with front crawl stroke
4. <u>25m RR (s)</u> – performance on a distance of 25 m with front crawl stroke, work of the arms only with a board between the legs.
5. <u>25 NN (s)</u> – performance on a distance of 25 m with front crawl stroke with aboard with only legs working
6. <u>Turn (s)</u> – speed of reaction of the whole body and space orientation ability. Tumble turn to front crawl stroke. The study group participant stands at a distance of 7.5 m from the turn wall, on getting a signal he swims to the turn wall with maximum speed, performs a somersault turn – glides and returns with the front crawl stroke to to take-off point (i.e. 7.5 m from the wall). The test is carried out three times and timed with accuracy up to 0.01 second. The time of the best trial is recorded.
7. <u>Take-off jump (s)</u> – speed of reaction of the whole body. Take-off jump from the pillar for front crawl stroke. The study group participant stands in the starting position on the pillar, and upon getting the signal, performs a starting jump for the front crawl stroke and swims 7.5 m. The trial is carried out three times a timed with accuracy up to 0,01 second. The time of the best trial is recorded.
8. <u>Glide test (s)</u> – ability to keep a state of balance. The study group participant stands with his back to the turn wall and on getting the signal glides into the crawl stroke at a distance of 7.5 m from the wall. The test is carried out three times and timed with accuracy up to 0.01 second. The time of the best trial is recorded.
9. <u>Swimming coordination index</u>: calculated according to the equation

$$WSKP = \frac{T25RR + T25NN}{2} - T25kraul$$

10. Long jump from take-off point (cm) – jumping ability. The standing long jump trial with two-leg push was used for assessing jumping ability. The study group participant stands with his legs spread out as wide as the hips before the marked line and swings both arms and leaps forward as far as he can. The result is the distance from the starting line to the last mark where the feet contact the ground. The best result from 2 trials with reading accuracy up to 1 cm is recorded (Grabowski, Szopa 1991) The result of this trial was assumed as the dynamic force index. It is important because it decides about the distance of the take-off jump and has considerable influence on the turn quality; furthermore, many authors mention it as a good criterion for assessing speed and force abilities (Bartkowiak 1976; Bułgakowa 1986, Raczek 1989).
11. RR technique (n cycles) – swimming step with crawl stroke on the breasts with a board between the legs. It consists in assessing the distance swum during one single movement. The better the contestant's technique, the bigger the distance covered in one movement cycle.
12. Technique style (n cycles) - swimming step front crawl stroke. A specialist swimming test consists in assessing the distance swum during one single movement. The covered distance also indirectly shows the power of arm thrust and the so called "feeling of the water". Due to little experience and the young age of the contestants, the trial was slightly changed. We analysed the number of full cycles of RR movement on a distance of 25 m freestyle. Take-off was from the water as to exclude the effect of the take-off jump (Bartkowiak 1976; Bułgakowa 1986, Training material PZP 1992).
13. Body height (cm)
14. Body mass (kg)
15. Foot length (cm)- distance between antropometric points pternion and akropodion [Drozdowski 1998].
16. Length of hand (cm) - projection distance between antropometric points stylion and phalangion (Drozdowski 1998).
17. Arm span (cm) – measurement of arm span was assessed by the side reach technique- the largest spreading of the upper limbs is determined by the distance between antropometric points of daktylion of the right and left hand, with spread out upper limbs [Drozdowski 1998].
18. Vital lung capacity (ml)
19. Rohrer's coefficient (%)

    This coefficient was calculated according to the following formula:

$$WSKR = \frac{\text{body mass (gms)}}{height^3 (cm)^3} x100$$

20. BMI (%)

BMI - Body Mass Index: value of relationship between body mass expressed in kilograms and square of height expressed in metres, (British Journal of Nutrition 1991).

Dependent variables: The measurement of dependent variables was carried out every 6 weeks, however, taken for analysis was the result after one year's training from the time of recruitment, additionally preceded by a yearly cycle of fitness preparation. The choice of one or two variables called quality indexes - the value of which at the end of the training cycle should be minimal (time of swimming a given distance) was needed for constructing the model.

The role of quality indexes was played by:

50m (points) – Y_1 performance on a distance of 50m front crawl converted into points according to triathlon charts.

800m (points) - Y_2 performance on a distance of 50m front crawl converted into points according to triathlon charts.

Statistical Analysis

In predicting the value of variables measured with so called strong scale regression models, numerous variables were used. The construction of such models required strict determination of endogenous variables (50m, 800m) as well as the right choice of variables in explaining the study's phenomenon. In conformity with criteria of the optimum model, the choice of variables explaining the regression model consists in putting them together in such a way that they are correlated with the explained variable and not correlated with other explanatory variables.

The results of trials and tests obtained during measurement (20 variables) were assumed as explanatory variables. The optimum choice of explanatory variables for the regression model was made on the grounds of a stepwise progressive regression algorithm. For determining functional dependencies of variables computer graphics and mean square approximation methods were used (Buja, Tukey 1991).

During statistical analysis of gathered material neural networks were used: Kohonen's maps (data mining analysis).

Results

Table 1

Structural parameters of regression equation for dependent variable Y_1 –50m

Measured variables	*Beta*	*St. error. beta*	*B*	*St. error B*	*t(135)*	*p level*
Intercept			158.74	2.5171	63.062	0.0000
Glide	-0.3751	0.1077	-26.56	7.6252	-3.483	0.0007
Foot length	0.7980	0.1094	56.51	7.7479	7.294	0.0000
Body height	0.6113	0.1011	43.29	7.1619	6.044	0.0000
25m RR	-0.3182	0.1034	-22.53	7.3204	-3.078	0.0025

Summing up of dependent variable regression: Y_1-50m
$r=0.91$; $r^2= 0.828$; Corrected $r^2= 0.823$
$F(4.135)=162.72$; $p<.00000$; standard estimation error: 29.783

Table 2.

Structural parameters of regression equation for dependent variable Y_2 –800m

Studied qualities	*Beta*	*St. Error beta*	*B*	*St. error B*	*t(131)*	*p level*
Intercept			80.321	1.127	71.256	0.0000
50m	-0.0042	0.0814	-0.174	3.386	-0.051	0.9590
Vital lung capacity	0.2550	0.0405	10.605	1.684	6.298	0.0000
800m	-0.2455	0.0499	-10.210	2.076	-4.917	0.0000
Hand length	0.2339	0.0445	9.728	1.852	5.252	0.0000
25m RR	-0.4234	0.0712	-17.613	2.961	-5.949	0.0000
Rohrer's index	0.1004	0.0297	4.177	1.237	3.378	0.0010
Standing long jump	0.1500	0.0372	6.241	1.549	4.028	0.0001
RR technique	0.1310	0.0458	5.450	1.903	2.864	0.0049

Summing up of dependent variable of regression: Y_2-800m
$r= 0.95$; $r^2= 0.90$; Corrected $r^2= 0.89$
$F(8.131)=152,61$; $p<.00000$; Standard estimation error: 13.338

Based on calculated regression equation coefficients (st. error B) we can asses that such qualities as *Glide, Foot Length, Body height and RR 25m* have considerable influence on predicting the dependent variable Y_1 – 50 m for the studied contestants, whereas such features as: *50 mm, 800m, RR 25m, Long Jump, RR Technique, Hand length, Vital Lung Capacity, Rohrer's Index* have significant effect on predicting the dependent variable Y_2 - 800m for the tested subjects. If the value of a variable changes by one point (one unit of this quality) then the value of our explained variable Y_1 – 50m decreases by -26.56 point, assuming that the remaining variables remain unchanged.

Untill now we have been dealing with linear dependencies existing among analysed variables. Nevertheless they are not always of the same nature in sport, which is why further on clustering was applied with the use of Kohonen's network.

Upon reducing variables the clustering method was used in order to partition a given tested set into homogenous taxonomic subsets. The feature that diversifies clustering methods from standard classification methods was the fact that in the former, class standards (i.e., results of particular cases) are unknown before starting the studies. Analysis was carried out on a set of 140 learning cases, described by four and eight independent variables successively for dependent variables Y_1 – 50m and Y_2 – 800m.

Variables describing the studied cases for distance Y_1- 50m: (25m RR, Glide, Body Height, Foot Length); and Y_2 – 800m: (50m, 800m, 25m RR, Long Jump from marked-out spot, RR Technique, Hand Length, Vital Lung Capacity, Rohrer's Index). The data consisted, therefore, of 4 and 8 explanatory variables and one explained variable, which constituted the case number. This explained variable was not used during teaching but was only used for describing the formed groups.

The classification tables established on the basis of learning sets for Y_1 – 50 m and Y_2 – 800 m distances in Kohonen's network can be useful in assessing new objects, not presented during learning. This results from evident, clear ordering of 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. Tables 3 and 4 a colour saturation scale was used in addition to numbering the neurons representing particular cases. This scale was created upon identifying the map areas by means of familiar results of dependent variables Y_1 – 50m (Table 3) and Y_2 – 800 m (Table 4). One can notice that, if our candidate is classified on recruitment to neurons with the biggest colour saturation, one can expect that the candidate will be among the best swimmers in his group after a year's training. If, on the other hand our candidate is classified to neurons with the lowest colour saturation, then one can expect this subject to belong to the weakest contestants in this age group. In particular neurons, the neuron point average was also given, which served as the basis for map area identification.

Table 3

Structure of studied set of junior swimmers obtained by means of Kohonen's map – Y_1 – 50m distance

1 58,66	2 47,00	3	4	5 101,00	6	7	8 218,00	9 224,50	10 314,00	11 314,00	12 233,75
13	14	15	16	17 78,00	18	19	20	21 230,00	22 309,00	23 240,00	24 233,75
25	26 75,00	27 70,50	28 122,00	29 233,75	30 123,00	31 180,00	32 220,00	33	34	35 180,00	36 268,66
37 76,00	38	39 70,00	40	41 166,50	42 175,00	43 157,00	44	45 218,50	46 218,00	47 257,00	48 264,00
49 72,00	50 38,00	51 131,50		53	54 136,50	55 116,00	56 214,00	57 217,50	58	59	60 245,50
61 64,00	62 105,00	63	64 105,00	65 108,00	66 233,75	67 118,00	68 218,00	69	70 214,00	71	72 234,50
73 102,00	74	75 106,00	76	77 167,00	78	79	80 173,00	81 173,66	82 239,00	83	84 200,66
85 131,50	86	87	88 145,00	89 147,00	90 145,00	91	92	93	94 158,00	95	96 155,00
97	98 97,00	99 101,00	100 128,00	101	102 132,00	103 145,00	104	105 227,00	106 230,00	107 234,50	108
109	110 83,00	111	112	113 128,00	114 140,00	115 185,50	116 233,00	117 140,50	118	119	120 245,00
121 86,75	122	123 109,00	124 112,00	125	126 140,00	127 142,00	128	129 235,00	130 238,00	131 245,00	132 209,33
133	134	135	136	137	138	139	140				

Scale of neural colour saturation for score intervals

<0-50)	<50-100)	<100-150)	<150-200)	<200-250)	<250-300)	<300-350)

Neuron number – 12

Point average for neuron – 233,75

Table 4

Structure of studied set of junior swimmers obtained by means of Kohonen's map – Y_2 – 800m distance

1 85,50	2	3 55,00	4 54,50	5	6 31,00	7 27,00	8 30,66	9 35,00	10 51,66	11	12 26,66
13	14 99,00	15	16	17	18	19	20	21	22 25,00	23	24 45,33
25 106,00	26	27 85,00	28	29 80,00	30 35,00	31 30,00	32 65,00	33 75,50	34 51,50	35	36 26,00
37	38	39	40	41	42	43 68,00	44 64,00	45	46	47	48
49 119,75	50 152,00	51 120,00	52	53 94,66	54 93,00	55 76,66	56	57 74,00	58	59	60 31,33
61	62	63 115,00	64	65 96,00	66	67	68	69	70	71 33,50	72
73 102,75	74	75 137,33	76 138,00	77 130,66	78	79 43,50	80	81 57,00	82	83 32,00	84
85	86	87 161,50	88	89	90 76,00	91	92	93	94	95	96 34,66
97 134,50	98 143,00	99 167,50	100	101 145,00	102 130,00	103	104 66,50	105 62,00	106 57,33	107	108 41,66
109	110	111	112 142,00	113	114 136,50	115	116	117	118	119	120
121 137,00	122	123 143,00	124	125 109,00	126 101,00	127 110,00	128 107,00	129 95,00	130 90,00	131	132 33,66
133	134	135	136	137	138	139	140				

Scale of neural colour saturation for score intervals

<0--50)	<50-75)	<75-100)	<100-125)	<125-150)	<150-175)

Neuron number: 12 — *Point average for neuron*: 233,75

The calculation of average was necessary because sometimes several subjects with specified score performance were assigned to one neuron. For both distances, Kohonen's network singled out fairly homogenous subsets of the best, average and weakest subjects. Therefore, one can say that these networks can be a very useful application in the recruitment of sport swimmers.

By analysing the value of explanatory variables corresponding to "good" neurons one can determine the required qualities in subjects of this group of variables.

Discussion

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 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.

Upon recognizing the clusters existing in the multidimensional data vector names can be attributed to them. Due to this Kohonen's network acquires the possibility of their classification according to internal logic of the data itself, and not on the grounds of some fancy arbitrary criteria. Based upon such great possibilities of data mining analysis, we can use Kohonen's network also in the selection of candidates for sport training. The obtained model has the form of a topological map, where certain areas can be separated. Upon determining map areas corresponding to particular sportsmen or their groups, we can identify the determined map areas. To do this it is necessary to specify the average sport development level achieved by the sportsmen represented by each of the neurons. The map constructed in this way can be used in the assessment of candidates for sport training. On the grounds of a characteristic the subject is assigned to one of the determined classes. One can expect that the candidate will achieve the level of sport development similar to the average development level characteristic of persons qualified to this group during network learning. To sum up, 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.

Conclusions

1. On the grounds of analysis of standardized *Beta* values one can say that the biggest predictability for the distance of 50m was shown by such qualities as: foot length and body height. Whereas for the 800m distance 25mRR and vital lung capacity were best predictors.

2. If the candidate for swimming training is qualified for neurons with the biggest colour saturation then this subject will be among the best swimmers in the group after a one year training program.
3. Models based on Kohonen's networks can be used in data mining analysis, which is aimed at assisting recruitment for sport swimming.

References

Bartkowiak E. Swimming training. WSiT. Warszawa, 1976.

Bułhakowa N. Ż. Otbor i podgotowka junnych płowcow. Fizkultura i Sport, Moskwa, 1986.

Buja A., Tukej P. A. Computing and Graphics in Statistics. Springer – Verlag, New York, 1991.

Colwin C.M. Swimming into 21-st Century. Leisure Press, Champaign, Illinois, 1992.

Costil D.L., Maglischo E.W., Richardson A.B. 1995. Swimming. Blackwell Scientific Publication, Oxford.

Drozdowski Z. Antrophometry in physical education. Poznań, 1998.

Grabowski H., Szopa J. Physical proficiency European Test. „EUROFIT", AWF, Kraków, 1991.

Komor J.A. Application of medling methods in sports. Department of Sport, Warszawa, 1982.

Kozioł R. System –cybernetic approach in training and sports fight. Scientific books, AWF, Kraków, nr 49, 1984.

Mester J., Perl J. Unconventional simulation and empirical evaluation of biological response to complex high training loads. [In:] Parisi P., Pigozzi F., Prinzi G. (Eds.): Sport Science'99 in Europe. Rome, 1999

Perl J., Lames M., Glitsch U. Modellbildung in der Sport-wissenschaft. Beiträge zur Lehr und Forchung im Sport, Bd 132, Schorndorf, Hofmann, 2002.

Płatonow W.N. Professional training in swimming. Biblioteka Trenera. Warszawa, 1997.

Raczek J.: Youth training in professional sports system. AWF, Katowice, 1989. AWF Katowice, 1989.

Ryguła I.(red) Elements of theory, methodology diagnosis and optimalization of the sports training. AWF, Katowice 2002.

Ryguła I. Research process in sciences about sport. AWF, Katowice, 2003.

Wit A., Barton G. Artificial neural networks. Sport wyczynowy nr 9. 1996r.

Address for correspondence:

Dr inż. Robert Roczniok
Department of Methodology and Statistics,
Academy of Physical Education in Katowice
72a Mikołowska str., Katowice, Poland
e-mail: r.roczniok@awf.katowice.pl
phone: +48 32 207 51 39, Fax: +48 32 207 52 00

Authors submited their contribution of the article to the editorial board.
Accepted for pinting in Journal of Human Kinetics vol. 17/2007 on march 2007.