Available online at www.sciencedirect.com
ScienceDirect
Procedia - Social and Behavioral Sciences 117 (2014) 482 – 487
Procedia
Social and Behavioral Sciences

ICSPEK 2013

Application of Neural and Regression Models in Sports Results Prediction

Adam Maszczyk[a]*, Artur Gołaś[a], Przemysław Pietraszewski[a], Robert Roczniok[a], Adam Zając[a], Arkadiusz Stanula[a]

[a]*Department of Sports Training, Jerzy Kukuczka Academy of Physical Education in Katowice, ul. Mikołowska 72a 40-065, Poland*

Abstract

The investigation was aimed at comparing regression and neural models with respect to their accuracy of predicting sports results. The present study involved a group of 116 javelin throwers, aged 18±0.5 years. The statistical analysis was initially done by the Shapiro-Wilk normality test and by the homogeneity test. The correlation matrix and regression analysis revealed four predictors (cross step, specific power of the arms and the trunk, specific power of the abdominal muscles and the grip power).Subsequently, non-linear regression models as well as neural models were built. Thus, to verify our models, the sports results were predicted for the group of 20 javelin throwers from the Polish National Team and tested by comparing the model-generated predictions with their actual data. The non-linear regression models and perceptron networks structured as 4-3-1, demonstrated their capacity for making generalizations and predicting sports results. Moreover, the difference in the value of absolute errors was 12.68 m (between true and estimated performances in the group of 20 Polish javelin throwers), thus favouring the neural models. The analysis of the above data clearly shows that the neural model does better at predicting sports results than the regression model. Therefore, the investigation demonstrated a significantly greater accuracy of prediction for perceptron models.

Selection and peer-review under responsibility of ICPESK 2013.

Key words: sport selection, linear models, non-linear models, Artificial Neural Networks, optimization

1. Introduction

Neural networks can be employed wherever a relationship between explanatory variables (inputs) and explained variables (outputs) exists (Haykin, 1994; Maszczyk, 2011). However, they are especially useful for

* Corresponding author. Adam Maszczyk; Tel.: 48604641015.
E-mail address: a.maszczyk@awf.katowice.pl

1877-0428 © 2013 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of ICPESK 2013.
doi:10.1016/j.sbspro.2014.02.249

seeking very complex input-output relationships, which are difficult to capture using statistical methods that are usually employed in such cases (e.g., the analysis of relationships or the separation of taxonomically homogenous groups). Given that the relationships between variables may be either linear or non-linear, recently, Artificial Neural Networks (ANN) were used more frequently to identify their actual nature. Today, this tool is frequently used for solving the modelling and prediction problems (Maier et al., 2000; Lees, 2002; Maszczyk, 2012).

Most relationships in sports science are, unfortunately, not linear. Each unit change in the x variable will not always bring about the same change in the y variable (Barton & Lees, 1993; Zehr, 2005; Hamilton, 2009; Woodman et al., 2010; Jones et al., 2010). Thus, researchers must using nonlinearity tools for describe the problem (i.e., nonlinear regression or neural models). But which could be better in supporting and optimizing athlete recruitment?

The investigation was intended to determine which variables offer the most information and qualify for performing the role of explanatory variables in the regression and neural models. An attempt was made to resolve the question whether regression models or artificial neural networks (ANN) predict sports results more precisely and so better support and optimize the athlete recruitment and selection processes of the 18-year-old javelin throwers.

2. Material and Methods

2.1. Participants

The presented study involved a group of 116 javelin throwers, aged 18±0.5 years. The core investigation was preceded by 12 months of general physical fitness training. The subjects participated in training three times a week. Besides general conditioning the athletes were trained to throw the javelin from a full run-up. The experimental factor was represented by differently structured training workloads assigned to individual athletes.

2.2. Data collection and statistical analysis

The data on the first 70 javelin throwers that were built into the non-linear regression model and neural model were derived from the measurements of the athletes' features made in May 2009 and 2010, the data set was subdivided into three series: learning series (40 cases), validation series (15 cases) and test series (15 cases). Then, to enhance the model, 26 new training cases were added (athletes at the same age, in the process of training), whose independent variables were taken from the measurements made in April 2010, and estimated again (46 cases – learning series, 25 – validation series and 25 – test series).

The research problem was solved using empirical and predictive investigations, with

the following model of statistical research: dependent variable Y (the distance of the javelin throw from a full run-up), and polytomic independent variables Xnn Yn.

The results of the trials and tests were used as the 41 explanatory variables and one dependent variable Y - the distance of the javelin throw.

To find the relationships between all investigated features a correlation matrix was calculated, while the statistical significance of particular explanatory variables (X) with respect to the explained variable (Y) was found by determining the vector of correlation. To determine the optimal set of predictors, the vector R0 was determined for the explanatory variables and the vector R1 for the correlations generated by the vector R0 of variables showing significant correlation with the explained variable Y.

The functional relationships between the variables were found by means of computer graphic techniques and midpoint quadratic approximation. This approach allowed for determining eight predictors which significantly improved the model's explained variable Y- the distance of the javelin throw. However, four variables were removed from the model following statistical testing: Body Cell Mass Index, body mass, specific power of the shoulder girdle and specific power of the shoulders and the trunk (hypothesis testing -significance testing and statistical verification of structural parameters of regression equation for dependent variable Y- within the meaning of the equation: sign (r(xj,y))=sign (aj)).

Ultimately, the regression equation was re-estimated with the remaining four explanatory (statistically significant) variables:

- The cross step with assuming the throwing stance (calculated using unit t, motion correctness assessed on a scale from 1 to 5) – variables CSATS expressed in seconds.
- Specific power of the arms and the trunk: a 2-kg forward medicine ball throw from an upright sitting position (measurement with accuracy of 5 cm) – variable SPAT expressed in meters.
- Specific power of the abdominal muscles: the maximal number of forward bends performed by a subject lying on the back during 10 s (n full cycles) – variable SPAM.
- Grip power - was measured using a dynamometer (Smedley Hand Dynamometer, Stoelting Co, Wood Dale, Ill) at exams 1 and 2 with midlife strength determined as the average of the best results in these 2 exams – unit N/kg variable GP.

But, constructed graphs of variables showed the nonlinearity of this problem. Thus, with this optimal sets of explanatory variables assembled, the construction of the nonlinear regression model begin.

Variable Y (the distance of a javelin throw from a full run-up - averages of three throws performed after a 30-minute warm-up) was selected as the model's explained variable.

Mean and standard deviations (SD) were calculated for all variables. The Kolmogorov-Smirnov test of normality and Levine's test of homogeneity of variance were performed to verify the normality of the distribution.

For generalization and prediction of sport results Multilayer Perceptron neural models were used (MLP). The networks were trained using the Levenberg-Marquardt algorithm. The training process was iterative (in successive training epochs (iterations), weights and thresholds were appropriately modified to reduce the total network error). The level of significance for all analyses was set at $p \leq 0.05$.

All statistical analyses in both groups of athletes were carried out on a PC using the statistical packages STATISTICA 9.1, STATISTICA Neural Networks module (Release 4.0E) and Excel 2010 from Microsoft Office 2010.

3. Results

Table 1 presents the results of non-linear regression analysis using optimal explanatory variables. Using the same variables (independent variables that were significantly associated with performance), the nonlinear perceptron models (multilayer perceptron - MLP) were constructed with the following structures: 4-2-1, 4-3-2-1 (four input neurons [variables], one or two hidden layer [with two and three neurons, respectively] and one outcome) and 4-3-1 (four input neurons [variables], one hidden layer [with three neurons] and one outcome). For networks 4-2-1, and 4-3-2-1 values of S. D. ratio for validation series might not be satisfactory. Normalized Root Mean Squared Error (NRMSE), for learning series (0.468), and values similarly seen in the validation and test data (0.275 and 0.386), were too height and not satisfactory to claim that this model adjusted well to learning data. Network 4-3-2-1 reached better results than 4-2-1. The Standard Deviation Ratio for learning and validation data was: 0.276; 0.285, and 0.279 for test series. Results for networks 4-2-1 and 4-3-2-1, showed problems of decreased ability for generalization. However, the value in validation and test series, and the correlation coefficient in those groups (0.96), implicated a necessity of building more models with a larger number of neurons in a hidden layer, which could approximately fit better into the network and learning data in the first set (Kurtz and Stergiou, 2005).

Table 1. Regression statistics of non-linear regression model for Y- distance of javelin throw (four predictors)

Variables	BETA	Error st. Beta	B	Error st. B	t	p
Intercept			34.101	4.302	3.471	0.027
Cross step with assuming the throwing stance	-0.224	0.052	-6.242	2.258	-2.571	0.025
Specific power of the arms and the trunk	0.147	0.013	1.636	0.411	1.363	0.024
Specific power of the abdominal muscles	0.185	0.043	0.371	0.025	1.440	0.019
Grip power	0.521	0.021	0.412	0.037	3.413	0.018

Table 2. Regression statistics of assessment of non-linear neural models for dependent variable Y- distance of javelin throw

MLP 4-3-2-1			
Data Standard	Learning series	Validation series	Test series
Normalized Root	0.124	0.113	0.138
R	0.965	0.941	0.966

Finally, the use of architecture 4-3-1 brought a breakthrough. In the group of 18 year-olds javelin throwers, the quality measures for this network (built for the first 70 cases) were 0.298 for the training subset, 0.284 for the validation subset and 0.278 for the test subset. The results pointed to a good fit between the model and the training data. However, with 46 new training cases added to the model, and following model re-estimation, the results improved. The network quality measure for the training subset, demonstrated an even better fit between the network and the training data. Regarding for new 4-3-2-1 networks, the NRMSE for the learning series was 0.124, and for the validation and test series 0.113 and 0.138, respectively (see Table 2).

What's more the difference in the absolute error values was 12.68 m (between true and estimated performances vs. models predictions in verification group of 20 Polish javelin throwers), favouring the neural models (see Table 3).

Table 3. Prediction errors for Y- distance of javelin throw

N	Network Error	Absolute Network Error	Regression Error	Absolute Regression Error
20	-2.31	16.77	-2.13	29.45

4. Discussion

This investigation was primarily aimed to identify the efficiency and predictive usefulness of artificial neural networks treated as an athlete recruitment tool in contrast to the widely used regression models. Indirectly, to accomplish the intended goals, an attempt was made to resolve which variables were most informative and qualified for playing the role of the models' explanatory variables. These variables significantly influenced throwing distances in the analyzed group of the 18-year-old javelin throwers. When the model's parameters for the young javelin throwers were being interpreted, it was found that increasing grip power by a unit increased the throwing distance by nearly 41 cm, while diminishing the time needed to perform cross step with assuming the throwing stance improved the result by approximately 6.2 m, assuming the regression error to be 2.2 m. Regarding the other variables, increasing, specific power of the arms and the trunk improved the throwing distance by 1.6 m; whereas increased abdominal strength yielded approximately 37 additional cm. Of significance, was the finding that the variable cross step with assuming the throwing stance (determining athlete's specific fitness) had the greatest predictive value and, specific power of the arms and the trunk best determined the maximal strength performance of selected groups of muscles. The above results are consistent with the theory of sport, where the javelin throw is described as a speed and strength event, (Tidow, 2000; Bompa, 2000; Hatton, 2005; Murakami et al., 2005).The neural model had better goodness of fit for athletes achieving medium or weak results. The negative total error of the network indicates that the model makes larger errors in athletes who throw the javelin further. The analysis of the above data clearly shows that the neural model better predicts sports results than the regression model, confirming also the Bartlett et al. findings (1996), whose neural models provided predictions of better quality than the multiple regression models. Murakami et al. (2005) indirectly proved that neural models are capable of better predictions than nonlinear or linear regression models. Therefore, the investigation demonstrated a significantly greater accuracy of prediction for the perceptron models.

5. Conclusions

The results of the investigation into the group of 18 year olds javelin throwers show that the created neural models offer much higher quality of prediction than the nonlinear regression model (absolute network error 16.77 m versus absolute regression error 29.45 m). The optimal set of variables that are the most informative and so usable as the explanatory variables of the nonlinear regression models and neural models consists of: *cross step with assuming the throwing stance, specific power of the arms and the trunk, specific power of the abdominal muscles* and *grip power*. The investigation's results explicitly demonstrate that neural models are a tool which is far superior and offers better optimization possibilities in predicting sports results, athlete recruitment and selection processes, than the widely applied regression models.

References

Bartlett, R., Müller, E., Lindinger, S., Brunner, F., & Morriss, C. (1996) Three-dimensional evaluation of the kinematic release parameters for javelin throwers of different skill levels, Journal of Applied Biomechanics, 12: 7-14.
Bompa, T.O. (2000). Training guidelines for young athletes (2nd ed.). Champaign, Ill: Human Kinetics Publishers.
Hatton, L. (2005). Optimizing the javelin throw in the presence of prevailing winds., United Kingdom: University of Kingston. Retrieved Jan 28, 2007 from http://www.leshatton.org/Documents/jav2007a_paper.pdf.
Haykin, S. (1994). Neural networks, a comprehensive foundation. New York: Macmillan College Publishing Company.
Lees, A. (2002). Technique analysis in sports: a critical review. Journal of Sports Sciences, 20, 813- 828.
Maier, K., Wank, V., Bartonietz, K., & Blickham, R. (2000) Neural network based models of javelin flight. Prediction of flight distances and optimal release parameters, Sports Engineering, 3: 57-63.
Maszczyk A., Roczniok R., Waśkiewicz Z., Czuba M., & Mikołajec K. (2012). Application of regression and neural models to predict competitive swimming performance. Perceptual and Motor Skills, 114(2): 610-626.

Maszczyk A., Zając A. and Ryguła I. (2011). A neural network model approach to athlete selection. Sports Engineering, 13:83-93.
Murakami, M., Tanabe, S., Ishikawa, M., Isolehto, J., Komi, V., Ito, A. (2005) Biomechanical analysis of the javelin throwing at 11th World Championships in Athletics in Helsinki, New Studies in Athletics, 4:11-21.
Tidow, G. (2000). Challenge decathlon – barriers on the way to becoming the "King of Athletes". Part I. IAAF: New Studies in Athletics, 15, 243–256.