A Fuzzy-based Software Tool Used to Predict 110m Hurdles Results During the Annual Training Cycle

Krzysztof Przednowek[1], Krzysztof Wiktorowicz[2], Tomasz Krzeszowski[2] and Janusz Iskra[3]
[1]*Faculty of Physical Education, University of Rzeszow, Rzeszow, Poland*
[2]*Faculty of Electrical and Computer Engineering, Rzeszow University of Technology, Rzeszow, Poland*
[3]*Faculty of Physical Education and Physiotherapy, Opole University of Technology, Opole, Poland*

Keywords: 110m Hurdles, Predictive Models, Fuzzy Systems, R Programming Language.

Abstract: This paper describes a fuzzy-based software tool for predicting results in the 110m hurdles. The predictive models were built on using 40 annual training cycles completed by 18 athletes. These models include: ordinary least squares regression, ridge regression, LASSO regression, elastic net regression and nonlinear fuzzy correction of least squares regression. In order to compare them, and choose the best model, leave-one-out cross-validation was used. This showed that the fuzzy corrector proposed in this paper has the lowest prediction error. The developed software can support a coach in planning an athlete's annual training cycle. It allows the athlete's results to be predicted, and in this way, for the best training loads to be selected. The tool is a web-based interactive application that can be run from a computer or a mobile device. The whole system was implemented using the R programming language with additional packages.

1 INTRODUCTION

Nowadays a variety of computer tools and methods play an important role in sport training. Both competitors and coaches are looking for new solutions that can support the training process. One aspect of such support can be the use of regression models to predict results. Prediction can be used to calculate performance results (Edelmann-Nusser et al., 2002; Maszczyk et al., 2011; Przednowek et al., 2014) or to identify sporting talent (Papić et al., 2009; Roczniok et al., 2013). For example, in the paper (Edelmann-Nusser et al., 2002), the authors use artificial neural networks to predict swimmers' competitive performance. The neural models were cross-validated and the results show that the modeling was very precise. The paper (Przednowek et al., 2014) describes the use of linear and nonlinear multivariable models as tools to predict 400m hurdles results. Another paper (Haghighat et al., 2013) presents a review of data-mining techniques that are used for prediction in various sporting disciplines.

Despite the existence of methods for prediction in sport, there is lack of tools that could be used by a coach during the training process, particularly in the 110m hurdles. Available software such as Kinovea, Physics Toolkit and SkillSpector can be used for the biomechanical analysis of human motion based on video sequences (Omorczyk et al., 2014; Sañudo et al., 2014; Gavojdea, 2015). For example, Sañudo et al. (Sañudo et al., 2014) use Kinovea software to determine the mean propulsive velocity and the maximal velocity during a bench press. In the paper (Gavojdea, 2015), Kinovea and Physics Toolkit are used to analyze the double salto backward tucked. Another application, named Lince (Gabin et al., 2012) is used in the design of observational systems, video recording, the calculation of data quality and the presentation of results. In another paper (Randers et al., 2010), the authors compare different multi-camera systems used for football match analysis. Papic et al. (Papić et al., 2009) developed a fuzzy expert system for scouting and evaluating young sports talent. A similar system is presented in (Louzada et al., 2016), where the authors carried out talent identification in soccer using a web-oriented expert system. In the 110m hurdles, the coach can use the tool to estimate the parameters of hurdle clearance (Krzeszowski et al., 2016). These parameters are estimated using the particle swarm optimization algorithm and they are based on analysis of the images recorded with a 100 Hz camera.

From the literature review, it can be seen that there is a need to develop tools supporting sports training. The main contribution of this paper is, therefore to

Przednowek, K., Wiktorowicz, K., Krzeszowski, T. and Iskra, J.
A Fuzzy-based Software Tool Used to Predict 110m Hurdles Results During the Annual Training Cycle.
DOI: 10.5220/0006043701760181
In *Proceedings of the 4th International Congress on Sport Sciences Research and Technology Support (icSPORTS 2016)*, pages 176-181
ISBN: 978-989-758-205-9

Table 1: Description of variables used to construct the models.

Variable	Description	$\bar{x}$	x_{min}	x_{max}
y	Predicted 110m hurdles result [s]	14.02	13.26	15.13
x_1	Age [years]	21.9	18.0	28.0
x_2	Body height [cm]	187.3	181.0	195.0
x_3	Body mass [kg]	77.8	71.0	83.0
x_4	Body mass index	22.1	20.3	23.5
x_5	Current 110m hurdles result [s]	14.33	13.34	15.40
x_6	Maximal and technical speed [m]	12513	5800	17970
x_7	Technical and speed exercises [m]	5925	2470	10200
x_8	Speed and specific hurdle endurance [m]	11961	3150	20400
x_9	Pace runs [m]	64087	25780	100300
x_{10}	Aerobic endurance [m]	328631	80600	550000
x_{11}	Strength endurance [m]	20638	1850	46595
x_{12}	Strength of lower limbs [kg]	291119	96400	658600
x_{13}	Trunk strength [amount]	38442	5240	145000
x_{14}	Upper body strength [kg]	3352	1630	4850
x_{15}	Explosive strength of lower limbs [amount]	1244	0	2214
x_{16}	Explosive strength of upper limbs [amount]	656	213	1850
x_{17}	Technical exercises – walking pace [min]	456	130	1110
x_{18}	Technical exercises – running pace [min]	574	195	1450
x_{19}	Runs over hurdles [amount]	778	362	1317
x_{20}	Hurdle runs in varied rhythm [amount]	1077	320	1850

develop a fuzzy-based software tool for results prediction in the 110m hurdles. This tool is created as a web application that can be run from a computer or a mobile device. It allows training loads to be planned across the annual training cycle so the athlete achieves their expected results.

2 MATERIAL

The training data contain 40 records. These records were collected from 18 highly trained athletes (mean result in the 110m hurdles: 14.02 s) aged between 18 and 28. The athletes were members of the Polish National Team. Each record contains an athlete's parameters and that athlete's training program across the annual training cycle. The models for result prediction were build using 21 variables (Tab. 1). The input variables $x_1 - x_5$ represent the athlete's parameters, the input variables $x_6 - x_{20}$ represent the training loads and the output variable y represents the predicted 110m hurdles result. The training loads were classified on the basis of work (Iskra and Ryguła, 2001), but it should be noted that this classification can be formulated in different ways. In this paper, the values of these loads are the sum of all the loads of the same type realized during the annual training cycle. The 110m hurdles results were registered before and after the cycle.

3 PREDICTIVE MODELS

3.1 Problem Formulation

We considered the regression problem with p inputs (predictors) X_j and one output (response) $\hat{Y}$. The goal was to build the predictive model $\hat{Y} = f(X_1, \ldots, X_p)$ based on a data set containing n observations in the form of pairs $(\mathbf{x}_i, y_i)$, where $i = 1, \ldots, n$, $p = \dim(\mathbf{x})$. In this paper, we use:

- linear models in the form of *ordinary least squares* (OLS), ridge regression, *least absolute shrinkage and selection operator* (LASSO) and elastic net regression,
- nonlinear model in the form of *fuzzy rule base system* (FRBS).

The detailed description of the linear models can be found in (Wiktorowicz et al., 2015). The fuzzy model is described in the next section.

Due to the small amount of data ($n = 40$), the models are compared using the *leave-one-out cross-validation* method (Arlot and Celisse, 2010). The idea of this method is based on the separation of subsets of learning data from the data set. Each subset is formed by removing only one record from the data set, which becomes the testing pair. The predictive quality of a model is expressed by the *root of the mean square error of cross-validation* (RMSE_{CV}) calculated as

$$\mathrm{RMSE}_{\mathrm{CV}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{-i})^2} \quad (1)$$

where $\hat{y}_{-i}$ is the output of a model constructed on the data set after removing the pair $(\mathbf{x}_i, y_i)$. Moreover, we use the fitness measure expressed by the *root mean square error of training* ($\mathrm{RMSE}_{\mathrm{T}}$) defined as

$$\mathrm{RMSE}_{\mathrm{T}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (2)$$

where $\hat{y}_i$ is the output of a model constructed on the full data set.

3.2 Fuzzy Regression Model

The fuzzy model proposed in this paper is constructed in the following steps.

1. Cross-validation of the OLS model

$$\hat{Y} = f_{\mathrm{OLS}}(X_1, \ldots, X_p) \quad (3)$$

for the data $(\mathbf{x}_i, y_i)$. The error in the ith step of cross-validation has the form

$$d_i = y_i - \hat{y}_{-i} \quad (4)$$

where $\hat{y}_{-i} = f_{\mathrm{OLS}}(\mathbf{x}_{-i})$.

2. Constructing the fuzzy (nonlinear) model

$$\hat{D} = f_{\mathrm{FUZZY}}(X_1, \ldots, X_p) \quad (5)$$

for the data $(\mathbf{x}_i, d_i)$. This model predicts the errors obtained in Step 1, that is it determines $\hat{d}_i = f_{\mathrm{FUZZY}}(\mathbf{x}_i)$. The best fuzzy model can be chosen on the basis of cross-validation conducted, for example, with varying numbers of fuzzy sets.

3. Cross-validation of the OLS model with the corrected error in the form

$$d_i^{new} = y_i - (\hat{y}_{-i} + \hat{d}_i) \quad (6)$$

where $\hat{y}_{-i}$ and $\hat{d}_i$ are determined by (3) and (5), respectively.

3.3 Comparison of Models

The regression models were calculated in R (R Core Team, 2016). The `lm.ridge` function from the "MASS" package (Venables and Ripley, 2002) was used to calculate the OLS and the ridge regressions. The LASSO regression and the elastic net regression were obtained with the `enet` function included in the "elastic net" package (Zou and Hastie, 2016). The fuzzy regression model was calculated using `frbs.learn` from the "frbs" package (Riza et al., 2015). The learning method was the Wang-Mendel algorithm (Wang and Mendel, 1992).

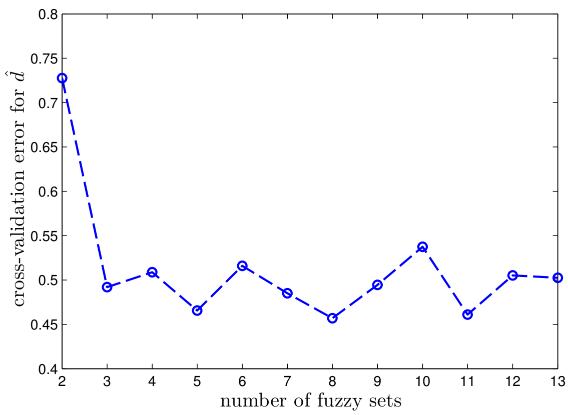

Figure 1: Cross-validation error for $\hat{d}$ as a function of the number of fuzzy sets. The smallest error is obtained for eight sets.

The parameters of the applied models are shown in Table 2. In the fuzzy model, five Gaussian membership functions are used, the t-norm is "minimum", the defuzzification is the "weighted average method", and the implication is "minimum". The number of fuzzy sets was determined by calculating the cross-validation errors. These errors are shown in Fig. 1 as a function of the number of fuzzy sets (changing from 2 to 13). From Fig. 1 it is seen that the best model is obtained for eight sets. The errors $\mathrm{RMSE}_{\mathrm{CV}}$ and $\mathrm{RMSE}_{\mathrm{T}}$ for the models under consideration are presented in Table 3. It shows that the proposed fuzzy regression model has the lowest $\mathrm{RMSE}_{\mathrm{CV}}$ and the high-

Table 2: Parameters of models.

Regression	Parameters
OLS	—
RIDGE	`lambda = 16.1`
LASSO	`lambda = 0, s = 0.04`
ENET	`lambda = 0.16, s = 0.56`

FUZZY	`method.type = "WM"`
	`num.labels = 8`
	`type.mf = "GAUSSIAN"`
	`type.tnorm = "MIN"`
	`type.defuz = "WAM"`
	`type.implication.func = "MIN"`

Table 3: Summary of errors.

Regression	$\mathrm{RMSE}_{\mathrm{CV}}$ [s]	$\mathrm{RMSE}_{\mathrm{T}}$ [s]
OLS	0.3807	0.1302
RIDGE	0.2276	0.1641
LASSO	0.2397	0.1495
ENET	0.1996	0.1562
FUZZY	0.0851	0.2852

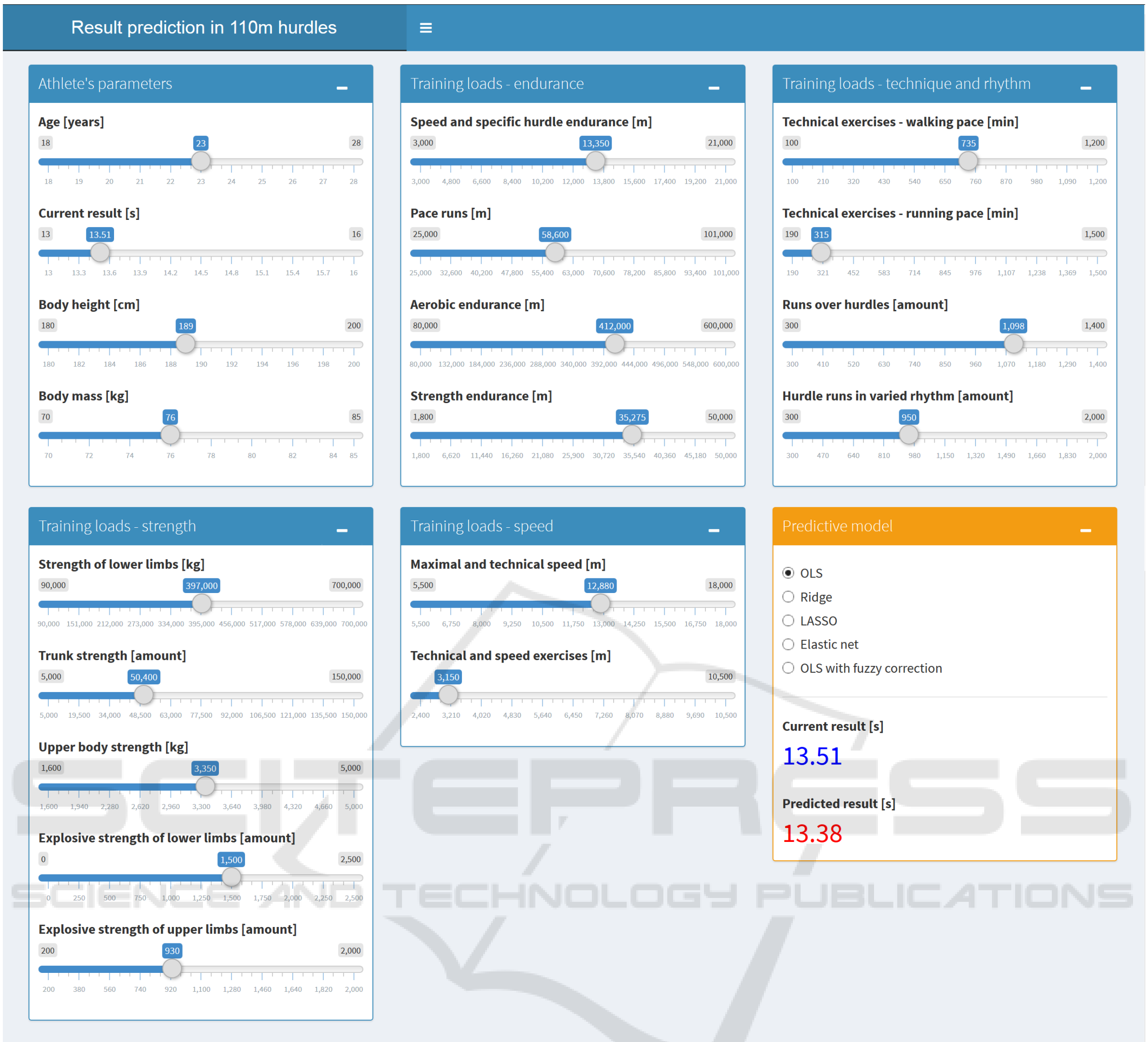

Figure 2: Screenshot of the application for result prediction in 110m hurdles.

est $RMSE_T$. It means that it can predict the result better than the linear models, but it has the worst fit to the data.

4 GRAPHICAL USER INTERFACE

The graphical user interface was implemented in R language using the libraries `shiny`, `shinythemes` and `shinydashboard`. This interface is a web-oriented application and therefore it requires only a web browser and an Internet connection to be used. The current version of the developed system is available on `https://hurdles.shinyapps.io/prediction`. The application consists of two tabs labeled "Result prediction" and "About".

The "Result prediction" tab is used for entering data and for result prediction (Fig. 2). The input variables are grouped into five boxes: "Athlete's parameters", "Training loads – endurance", "Training loads – technique and rhythm", Training loads – strength", and "Training loads – speed". The value of each input can be modified using appropriately scaled sliders. For example, the box "Training loads – endurance" presented in Fig. 3 has five sliders for changing the endurance training loads. Each slider has its range determined on the basis of the minimum and maximum values in the database (Tab. 1). For instance, the slider "Pace runs" ranges from 25000 m to 101000 m with each step equal to one meter.

In the last box, labeled "Predictive model" (Fig. 4), the user can choose one of the developed re-

Figure 3: Screenshot of the box for entering endurance training loads.

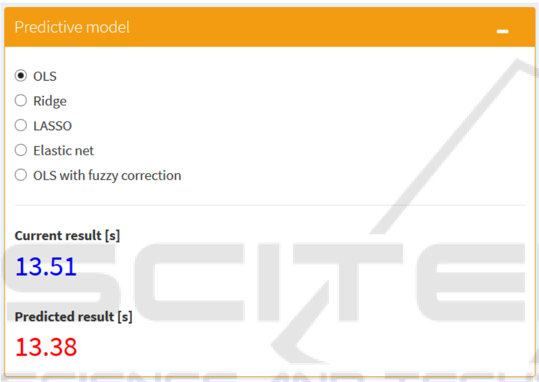

Figure 4: Screenshot of the box for result prediction.

gression models. Two `textOutput` fields display the current and predicted results. Prediction of the result is performed automatically after changing the value in any box in this tab. In this way, the user can modify training loads and observe the changes that occur in the expected result. The training program should correspond to the inputs specified in Tab. 1.

The "About" tab contains information about the application and the authors.

5 CONCLUSIONS

In this paper a fuzzy-based software tool for result prediction in the 110m hurdles was presented. The prediction is based on the following models: OLS regression, ridge regression, LASSO regression, elastic net regression and OLS regression with fuzzy correction. The best prediction was obtained by the proposed fuzzy model, but it has the lowest fitness to the data. The parameters of the models can be also validated in future using an independent group of athletes with different training conditions.

The whole application, composed of the predictive models and the graphical user interface, was created in R programming language. The simple interface allows an athlete's parameters and training loads to be changed. In this way, the coach can predict the expected result and select individual components of the training for a given athlete.

Further work will focus on the development of the proposed application, which involves implementing individual user accounts, the preparation of an athlete databases and creating reports. In addition, a new computational module will be developed for generating training loads.

ACKNOWLEDGEMENTS

This work has been supported by the Polish Ministry of Science and Higher Education within the research project "Development of Academic Sport" in the years 2016-2018, project no. N RSA4 00554.

REFERENCES

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79.

Edelmann-Nusser, J., Hohmann, A., and Henneberg, B. (2002). Modeling and prediction of competitive performance in swimming upon neural networks. *European Journal of Sport Science*, 2(2):1–10.

Gabin, B., Camerino, O., Anguera, M. T., and Castañer, M. (2012). Lince: Multiplatform sport analysis software. *Procedia - Social and Behavioral Sciences*, 46:4692 – 4694.

Gavojdea, A.-M. (2015). The impact of using the specialized software on studing the landings at uneven bars. In *Conference proceedings of eLearning and Software for Education (eLSE 2015)*, number 3, pages 346–352.

Haghighat, M., Rastegari, H., and Nourafza, N. (2013). A review of data mining techniques for result prediction in sports. *Advances in Computer Science: an International Journal*, 2(5):7–12.

Iskra, J. and Ryguła, I. (2001). The optimization of training loads in high class hurdlers. *Journal of Human Kinetics*, 6:59–72.

Krzeszowski, T., Przednowek, K., Wiktorowicz, K., and Iskra, J. (2016). Estimation of hurdle clearance parameters using a monocular human motion tracking method. *Computer Methods in Biomechanics and Biomedical Engineering*, 19(12):1319–1329. PMID: 26838547.

Louzada, F., Maiorano, A. C., and Ara, A. (2016). iSports: A web-oriented expert system for talent identification in soccer. *Expert Systems with Applications*, 44:400 – 412.

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

Omorczyk, J., Nosiadek, L., Nosiadek, A., and Chwała, W. (2014). Use of biomechanical analysis for technical training in artistic gymnastics using the example of a back handspring. In Urbanik, C., Mastalerz, A., and Iwańska, D., editors, *Selected problems of biomechanics of sport and rehabilitation*, volume II, pages 104–115. Józef Piłsudski University of Physical Education in Warsaw.

Papić, V., Rogulj, N., and Pleština, V. (2009). Identification of sport talents using a web-oriented expert system with a fuzzy module. *Expert Systems with Applications*, 36(5):8830 – 8838.

Przednowek, K., Iskra, J., and Przednowek, K. H. (2014). Predictive modeling in 400-metres hurdles races. In *2nd Int. Congress on Sport Sciences Research and Technology Support - icSPORTS 2014*, pages 137–144. SCITEPRESS, Rome, Italy.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Randers, M. B., Mujika, I., Hewitt, A., Santisteban, J., Bischoff, R., Solano, R., Zubillaga, A., Peltola, E., Krustrup, P., and Mohr, M. (2010). Application of four different football match analysis systems: A comparative study. *Journal of Sports Sciences*, 28(2):171–182.

Riza, L. S., Bergmeir, C., Herrera, F., and Benítez, J. M. (2015). frbs: Fuzzy rule-based systems for classification and regression in R. *Journal of Statistical Software*, 65(6):1–30.

Roczniok, R., Maszczyk, A., Stanula, A., Czuba, M., Pietraszewski, P., Kantyka, J., and Starzyński, M. (2013). Physiological and physical profiles and on-ice performance approach to predict talent in male youth ice hockey players during draft to hockey team. *Isokinetics and Exercise Science*, 21(2):121–127.

Sañudo, B., Rueda, D., Pozo-Cruz, B. D., de Hoyo, M., and Carrasco, L. (2014). Validation of a video analysis software package for quantifying movement velocity in resistance exercises. *Journal of Strength and Conditioning Research*.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York.

Wang, L. X. and Mendel, J. M. (1992). Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427.

Wiktorowicz, K., Przednowek, K., Lassota, L., and Krzeszowski, T. (2015). Predictive modeling in race walking. *Computational Intelligence and Neuroscience*, 2015:9. Article ID 735060.

Zou, H. and Hastie, T. (2016). *Package "elasticnet"*. CRAN.