Modeling children's mathematical gift by neural networks and logistic regression

Margita Pavlekovic a, Mirta Bensic b, Marijana Zekic-Susac c,*

a Faculty of Teacher Education, University of J.J. Strossmayer in Osijek, L. Jegera 12, 31000 Osijek, Croatia
b Department of Mathematics, University of J.J. Strossmayer in Osijek, Cajev trg 3, 31000 Osijek, Croatia
c Faculty of Economics, University of J.J. Strossmayer in Osijek, Gajev trg 7, 31000 Osijek, Croatia

A R T I C L E   I N F O

Keywords:
Mathematical gift
Logistic regression
Neural networks

A B S T R A C T

The purpose of the paper was to extract important features of children’s mathematical gift by using neural networks and logistic regression, in order to create a model that will assist teachers in elementary schools to recognize mathematically gifted children in an early stage, therefore enabling further development and realization of that gift. The initial model was created on the basis of a theoretical background and heuristic knowledge on giftedness in mathematics, including five components: (1) mathematical competencies, (2) cognitive components of gift, (3) personal components that contribute gift development, (4) environmental factors, and (5) efficiency of active learning and exercising methods, as well as grades and out-of-school activities of pupils in the fourth year of elementary school. The three neural network classification algorithms were tested in order to extract the important variables for detecting mathematically gifted children. The best neural network model was selected on the basis of a 10-fold cross-validation procedure. The model was also investigated by the logistic regression. Important predictors detected by two methods were compared and analyzed. The results show that both methods extract similar set of variables as the most important, including grades in mathematics, mathematical competencies of a child regarding numbers and calculating, but also grades in the literature, and environmental factors.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Previous research (Johnson, 2007) emphasized the need for accurate detection and further development of mathematical gift. Mathematical giftedness of children in elementary schools was detected by using mathematical tests such as SAT-Math, as well as scientifically approved standard Raven progressive matrices in the process of psychological evaluation of a child (O’Boyle et al., 2005; Pind, Eyriún, Gunnarsdóttir, & Johannesson, 2003). In schools where psychologists are not available, teachers usually use mathematical competencies as the only criterion for determining a child’s gift. Pavlekovic, Zekic-Susac, and Djurdjevic (2009) created an expert system that uses five components of mathematical gift, identified upon theoretical background (Sterberg, 2001; Tannenbaum, 1983; Terman & Oden, 1959; Vlahovic-Stetic, 2006) and heuristics, such as (1) mathematical competencies, (2) cognitive components of gift, (3) personal components that contribute gift development, (4) environmental factors, as well as (5) efficiency of active learning and exercising methods. Their research shows that MathGift ES detected more children as gifted than teachers did in their estimations, and that the expert system estimations are more similar to psychologists’ estimations. The model did not include grades of the pupils which are often used by teachers in gift detection.

This paper aims to extend the previous research (Pavlekovic et al., 2009) by including pupils’ grades into the model, as well as to test quantitative methods based on learning theory in order to avoid the dependence on heuristics and human expert. To identify important selectors that teachers usually use to identify gifted children, it was challenging to test an intelligent method that has abilities of dealing with nonlinear functions in the purpose of prediction, classification, and association. Therefore, three neural network (NN) algorithms were tested in order to classify children in one of the two gift categories. The model is aimed to learn psychological findings, and to be used as a part of a decision support system in schools where psychologist’s estimations are not available.

A survey was conducted at 10 Croatian elementary schools where the psychologists’ and teachers’ estimations were obtained for each child in the sample. The results of the neural network model and logistic regression model were compared in terms of important predictors. The advantages and limitations of both approaches are also discussed.

The output of the model was binary defined representing two categories: (1) mathematically gifted children and (2) mathematically
non-gifted children. An empirical research was conducted in 2006, including pupils of age 10 (fourth grade) in 10 elementary schools in Osijek. The estimations were compared using statistical tests.

The structure of the paper is the following: Section 2 contains a review of previous research in the area, followed by the description of neural network and logistic regression methodology used in the paper. The data about examiners are described in a separate section. After the results, the conclusion and guidelines for future research are given.

2. Review of previous research

A number of authors has tested various statistical methods and neural networks in education, but mainly in measuring the potential of students with regard to their study performance as an important step of the admission process in a school. Gorr, Nagin, and Szczypula (1994) predicted student grade point averages (GPAs) by using linear regression, stepwise polynomial regression, and NNs, and then compared the predictions with an index used by an admissions committee for predicting student GPAs in professional school. Their results show that none of the tested methods was significantly better than the practitioners’ index. Hardgrave, Wilson, and Walstrom (1994) investigated NNs in predicting students’ success in a graduate program. They showed that non-parametric procedures such as neural networks perform at least as well as traditional methods and are worthy of further investigation.

Wilson and Hardgrave (1995) tested different classification and regression methods such as discriminant analysis, logistic regression and neural networks, in predicting graduate student success in an MBA program, and shown that non-parametric procedures, such as neural networks, perform at least as well as traditional methods and are worthy of further investigation in that area. Zelensnikov and Nolan (2001) created a decision support system based on fuzzy logic and predicted rules to assist teachers in grading essays. Paliwal and Kumar (2009) used neural networks and traditional statistical techniques, such as regression analysis, discriminant analysis, factor analysis, and logistic regression to predict academic performance of business school graduate students. Their results show that similar set of predictors is extracted by all tested methods. Statthacopoulou, Magoulas, Grigoriadou, and Samarakou (2005) propose to use the methodology of NNs and fuzzy logic for an advanced student diagnosis process in an intelligent learning system. Their model enables a system to “imitate” teacher in diagnosing student characteristics, and in selecting the learning style that suits those characteristics. The system is tested in learning vector construction in physics and mathematics. Results obtained by the system are compared to the recommendations of a group of experienced teachers, showing that the system is able to manage the diagnostic process, especially for marginal cases, where it was difficult even for teacher to bring accurate evaluation of student. Canales, Pena, Peredo, Sossa, and Gutierrez (2007) developed an adaptive and intelligent web-based education system (WBES), which takes into account individual student learning requirements and enables the usage of different techniques, learning styles, learning strategies, and ways of interaction.

The above shows that neural networks were frequently used in addition to traditional statistics in previous research on student success. However, the area of detecting mathematical gift was not investigated enough. O’Boyle et al. (2005) investigated brain activation of mathematically gifted male adolescents, but there is a lack of research of mathematical giftedness in elementary schools. Teachers usually rely on mathematical tests and their subjective judgments in detecting mathematically gifted children. Pavlekovic et al. (2009) compared the efficiency of expert systems and neural networks in detecting a child’s mathematical gift. They showed that both methods can serve as an efficient tool in detecting mathematically gifted children, on the basis of five basic components of mathematical gift, which consisted of totally 60 input variables. As an extension to that previous paper, this research adds some new input variables to the model describing pupils’ grades and out-of-school activities, and compares the basic feature of mathematical gift extracted by neural networks and logistic regression.

3. Description of data and sampling procedure

The initial data sample for the research included 257 children of the fourth grade at 10 elementary schools in Osijek, Croatia at December 2006. Parents’ permissions to do psychological evaluations were obtained for 106 pupils. After excluding missing data, 105 pupils were used for modeling. The input space consisted of the five components of mathematical gift identified in previous research (Pavlekovic, Zekic-Susac, & Djurdjevic, 2007), as well as of the data describing pupil’s grades, and additional out-of-school activities in which a pupil was engaged. The five components of mathematical gift were (1) mathematical competencies, (2) cognitive components of gift, (3) personal components that contribute the gift development, (4) strategies of learning and exercising, and (4) environmental factors. The group of mathematical competencies was represented by the level at which a pupil deals with (a) numbers, (b) measurements, (c) shapes, and (d) solving mathematical problems. The estimation of a pupil’s level of dealing with those competencies was given by a pupil’s teacher. Teachers participating in the research were previously instructed on the estimation criteria. In addition to teacher estimations on the above competencies, a pupil’s grade point average, individual grades in the course of mathematics and the literature, as well their out-of-school activities were also used as input data. Therefore, a total number of 18 input variables were used, and their descriptive statistics is presented in Table 1.

The output variable represented a pupil’s gift in mathematics estimated by school psychologists who were specialized in education of gifted children. Although psychologists found four categories of mathematical gift, in order to pay special attention to gifted children, categories were further regrouped into two main groups: (1) group of pupils found to be mathematically gifted, consisted of pupils assigned to categories 1 and 2 by psychologists, further referred as “gifted” pupils, and (0) group of pupils that were not found to be mathematically gifted, consisted of pupils assigned to categories 3, and 4, further referred as “non-gifted” pupils. The structure of psychologists’ findings is presented in Table 2.

It can be seen from Table 2 that more than 50% of pupils in the sample were recognized as mathematically gifted by psychologists. The reason for such high proportion of gifted lies in the fact that parents’ permissions to do psychological evaluations were obtained mostly for children with a high grade point average (50% of pupils in the final sample have the rounded grade point average of 5). Therefore, the initial sample (n = 257) was representative, but the final sample (n = 105) consisted of mainly very good and excellent pupils (mean of the GPA = 4.7). In spite of such non-representative final sample, the research was continued since it was assumed that (1) the population of gifted pupils remained in the final sample, and (2) if the model is able to accurately recognize gifted among such above-average pupils, it should also be useful while dealing with the whole population.

In order to train and test neural networks, the total sample was randomly divided into three subsamples such that 80% of data was used for training the network, 10% of data was used to find the optimal learning time and network structure in a cross-validation
procedure, while the remaining of 10% of data was used to finally test the network. The distribution of gifted and non-gifted pupils in all three samples is presented in Table 3. The logistic modeling was conducted by using the whole sample.

Generalization ability of the neural network models was tested in a 10-fold cross-validation procedure where the initial sample was divided in 10 equally sized parts; one part was used for testing the network, and the rest of the sample for training. After producing the result, another part of the sample was used for testing, and remaining for training. The procedure was repeated 10 times, and the results were used to estimate a generalization error of a NN model (Masters, 1995).

4. Modeling methodology

4.1. Neural network methodology

Neural networks (NNs) as an artificial intelligence method have been successfully used for classification, prediction, and association in a number of business areas (Li, 1994), as well as in social sciences (Detienne, Detienne, & Joshi, 2003) and other domains. In terms of computational complexity, NNs can handle problems that require an iterative usage of data to detect patterns, where lots of example data are present, and when it is difficult to specify a parametric model (Detienne et al., 2003). NNs as a non-parametric method have the ability to overcome the proportionality and linearity constraints imposed by parametric methods (Hu, Zhang, Jiang, & Patuwo, 1999). In this paper, three NN algorithms were tested: the multilayer perceptron (MLP), the radial basis function network (RBFN), and the probabilistic neural network (PNN).

MLP is a general purpose feedforward network, and one of the most frequently used NN algorithms. In order to optimize the error function it uses the classical backpropagation algorithm based on deterministic gradient descent algorithm originally developed by Paul Werbos in 1974, extended by Rumelhart, Hinton, Williams (in Masters, 1995). Since the main disadvantage of the backpropagation algorithm is the danger of local minima, the conjugate gradient algorithm is also tested in order to overcome this limitation (Masters, 1995). Conjugate gradient is combined with the classical backpropagation such that backpropagation is used in first 100 epochs, while the conjugate gradient is used in the next 500
epochs. The standard delta rule was used for the weight adjustment, and the learning rate was dynamically optimized during the learning process (ranged from 0.09 to 0.01, while the momentum was set to 0.3).

RBFN is based on a clustering procedure for computing distances among each input vector and a center, represented by the radial unit. The ability of RBFN with one hidden layer to approximate any nonlinear function is proved by Park and Sandberg (in Karayiannis & Weigun, 1997). Michelli (in Karayiannis & Weigun, 1997) showed how this network can produce an interpolating surface which passes through all the pairs of the training set. RBFN algorithm uses Euclidean distance and Gaussian transfer function in the hidden layer which maps the output of the distance function according to:

$$f(x) = \varphi(||x - c||) = e^{-\frac{1}{2}}$$

(1)

where \( x \) is an input vector, \( c \) is the center determined by a clustering algorithm, and parameter \( \sigma \) is determined by the nearest neighbor technique. A pruning procedure for gradually decreasing the number of hidden units is used in our experiments. The initial number of hidden units is set to the size of the training sample. Since the MLP and RBFN are primarily designed for the prediction type of problems, a softmax activation function was added in the output layer of both networks in order to obtain probabilities in the output.

The PNN algorithm was tested due to its fast learning and efficiency in classification. It is a stochastic-based NN developed by Specht (in Karayiannis & Weigun, 1997). Masters (1995) suggests PNN as a good selection for classification problems when there are outliers in data (Masters, 1995), and when the learning speed is important, since this algorithm does not learn iteratively, it passes only once through the dataset. The architecture of the PNN is built upon Bayes' classifier using the Parzen window estimator to estimate the probability distributions of the class samples (Patterson, 1995). Parzen's technique estimates a "sphere-of-influence" function for separating an unknown point from the known training sample point. Such function has a higher value if the distance is close and converges to zero if the distance becomes large. Taking the sum of this function for all known training set members, and classifying the unknown point into the population with the largest sum is the main idea of the PNN algorithm. Parzen's estimated density function is (Masters, 1995):

$$g(x) = \frac{1}{n\sigma} \sum_{i=1}^{n} W\left(\frac{x - x_i}{\sigma}\right)$$

(2)

where \( n \) is the sample size, \( x \) is an input vector, \( \sigma \) is the scaling parameter that controls the width of the area of influence of the distance, and \( W \) is the weighting function, for which the Gaussian function is usually used.

The output layer of all three models consisted of one neuron (valued as 1 for the insolvent companies, and 0 for solvent companies). Sensitivity analysis was performed on the test sample in order to determine the significance of input variables to the model. One hidden layer is used in all NN models in our experiments. Regarding the number of hidden units, the method of pruning is used which eliminates the weights lower than a threshold (0.05 in our experiments) input and hidden units at the end of the training process in order to produce smaller and faster networks with equivalent performance. The initial number of hidden units was set to 37 in MLP networks, and to the size of the training sample in RBFN and PNN. Overtraining is avoided by a split-sample process which alternatively trains and tests the network (using a separate test sample) until the performance of the network on the test sample does not improve for \( n \) number of iterations. The maximum number of training epochs was set to 1000. The generalization ability of all three NN models is determined by a 10-fold cross-validation procedure.

### 4.2. Logistic regression methodology

Logistic regression (LR) modeling (see, e.g. Harrel, 2001) is widely used for the analysis of multivariate data involving binary responses. It provides a powerful technique analogous to multiple regression and ANOVA for continuous responses. Since the likelihood function of mutually independent variables \( Y_1, ..., Y_n \) with outcomes measured on a binary scale is a member of the exponential family with \( \log \left( \frac{e^\beta}{1 + e^\beta} \right) \) as a canonical parameter \( (\pi_1 \text{ is a probability that } Y_j \text{ becomes 1}) \), the assumption of the logistic regression model is a linear relationship between a canonical parameter and the vector of explanatory variables \( x_j \) (dummy variables for factor levels and measured values of covariates):

$$\log \left( \frac{\pi_j}{1 - \pi_j} \right) = x_j^T \beta$$

(3)

This linear relationship between the logarithm of odds and the vector of explanatory variables results in a nonlinear relationship between the probability of \( Y_j \) equals 1 and the vector of explanatory variables:

$$\pi_j = \exp \left( x_j^T \beta \right) \left(1 + \exp \left( x_j^T \beta \right) \right)^{-1}$$

(4)

A special characteristic of the dataset in this research was the small number of cases compared to the number of predictor variables used in the model. Also, the collinearity appeared among some of the crucial predictors. As the sample consisted of 57 gifted pupils and 48 non-gifted pupils, to assure the reliability of the model, the number of predictors (or candidate predictors) for giftedness in the logistic regression modeling should not have been greater than 5. For these reasons, four different data reduction methods were used in the modeling procedure, such as the principal component analysis, the cluster analysis, forward selection, and backward selection procedures. They were performed in order to extract important predictors and to find the best logistic regression model.

### 5. Model selection and results

#### 5.1. Neural network results

According to Liu (1995) and Anders and Korn (1999), model selection should be performed on the basis of a generalization error. Some of the well-known methods for testing the generalization ability of the models are n-fold cross-validation, jackknifing and bootstrapping (Hu et al., 1999; Masters, 1995). We use the cross-validation due to its simplicity, and because it produces no statistical bias of the result since each tested sample is not the member of the training set. We perform a 10-fold cross-validation procedure (or leave \( k \) cases out), and compute the average of the predictive errors, which is used to estimate the generalization error. This estimate is also used as the model selection criterion. According to Witten and Frank (2000), extensive tests on numerous datasets have shown that 10 is sufficient value for \( n \) in the \( n \)-fold cross-validation. The procedure of cross-validation is performed according to a slightly modified description of Masters (1995) including the following steps: (1) the in-sample data were divided into 10 equally sized independent subsamples, (2) each NN model is estimated 10 times, each time using the different set of nine subsamples for training, and tested on 1 sample that was left out of training, (3) 10 different results were obtained for each NN model, and (4) the average result, i.e. the average hit rate is computed for
all omitted training cases. The model with the lowest average error is selected as the best model.

The 10-fold CV procedure was performed on each of the three NN algorithms: MLP, RBFN, and Probabilistic. The results of the three NN algorithms are presented in Table 4. The average hit rate of all children in the test sample as well as separate hit rates of gifted and non-gifted children are computed on each of 10 test samples in the 10-fold CV procedure. The average of the average hit rate across all 10 test samples is used to select the best NN model.

When the results of the MLP networks are observed, Table 4 shows that the MLP average hit rates across 10 samples in the 10-fold CV procedure varied from 59.52% to 100%. The average of the average hit rate 71.98%. It can be noticed that the MLP network is more successful in recognizing gifted than non-gifted children. The average hit rate across 10 samples is higher for gifted children (78.28%) than for non-gifted children (65.69%). The results of the RBFN algorithm are less successful, due to the fact that its clusters are not well distributed among the children. The PNN algorithm produced the average hit rate of 68.06%, which is lower than correspondent hit rates of the MLP algorithm.

5.2. Logistic regression results

The statistical analyses of the data were performed by using SAS 9.1 procedures. The procedures VARCLUS, PRINCOMP and PRINQUAL were used for data reduction before model development, while the logistic regression modeling was performed by the procedure LOGISTIC. The best logistic regression model was constructed in that way was the best model we have reached on the observed dataset among all models that used pre-model development analysis. Cluster summary report for these four clusters is shown in Table 5.

Clusters and their standardized scoring coefficients are given in Table 6. Proportion of total variation explained by the first cluster was 0.3273, by the first two clusters 0.4576, by the first three clusters 0.5604, and the proportion of total variation explained by all clusters was 0.6242. It should also be mentioned that the first three clusters were positively inter-correlated while the last cluster was negatively correlated with cluster 1 and cluster 3.

<table>
<thead>
<tr>
<th>Test sample in CV procedure</th>
<th>Average hit rate (%)</th>
<th>Hit rate of gifted (%)</th>
<th>Hit rate of non-gifted (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60.00</td>
<td>100.00</td>
<td>20.00</td>
</tr>
<tr>
<td>2</td>
<td>61.91</td>
<td>57.14</td>
<td>66.67</td>
</tr>
<tr>
<td>3</td>
<td>59.52</td>
<td>85.71</td>
<td>33.33</td>
</tr>
<tr>
<td>4</td>
<td>75.00</td>
<td>50.00</td>
<td>100.00</td>
</tr>
<tr>
<td>5</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>6</td>
<td>70.00</td>
<td>40.00</td>
<td>100.00</td>
</tr>
<tr>
<td>7</td>
<td>64.29</td>
<td>100.00</td>
<td>28.57</td>
</tr>
<tr>
<td>8</td>
<td>87.50</td>
<td>100.00</td>
<td>75.00</td>
</tr>
<tr>
<td>9</td>
<td>75.00</td>
<td>50.00</td>
<td>100.00</td>
</tr>
<tr>
<td>10</td>
<td>66.67</td>
<td>100.00</td>
<td>33.33</td>
</tr>
<tr>
<td>Average hit rate across 10 samples</td>
<td>71.98</td>
<td>78.28</td>
<td>65.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>R-squared with own cluster</th>
<th>Standardized scoring coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1 NUMBERS</td>
<td>0.7858</td>
<td>0.1203</td>
</tr>
<tr>
<td>MEASURES</td>
<td>0.7183</td>
<td>0.1203</td>
</tr>
<tr>
<td>SHAPES</td>
<td>0.7147</td>
<td>0.1203</td>
</tr>
<tr>
<td>PROBLEMS</td>
<td>0.6199</td>
<td>0.1203</td>
</tr>
<tr>
<td>COGN</td>
<td>0.6563</td>
<td>0.1203</td>
</tr>
<tr>
<td>ENVIRON</td>
<td>0.5769</td>
<td>0.1203</td>
</tr>
<tr>
<td>ACTIV</td>
<td>0.7477</td>
<td>0.1203</td>
</tr>
<tr>
<td>GPA</td>
<td>0.6851</td>
<td>0.1203</td>
</tr>
<tr>
<td>MATH</td>
<td>0.7077</td>
<td>0.1203</td>
</tr>
<tr>
<td>LIT</td>
<td>0.7097</td>
<td>0.1203</td>
</tr>
</tbody>
</table>

| Cluster 2 ACTIV_MI | 0.5021 | 0.44073 |
| ACTIV_LANG | 0.6261 | 0.44073 |
| ACTIV_LIT | 0.5915 | 0.44073 |

| Cluster 3 PERSONAL | 0.3556 | 0.39400 |
| ACTIV_S | 0.3528 | 0.39400 |
| ACTIV_M | 0.4340 | 0.39400 |
| ACTIV_SCI | 0.4748 | 0.39400 |

| Cluster 4 ACTIV_A | 1.0000 | 1.00000 |

The logistic regression model for detecting mathematical gift which used cluster 1, cluster 2, cluster 3, and cluster 4 as predictor variables had the properties summarized in Tables 7–9.

It can be seen from Table 9 that the best LR model produced the average hit rate of 76.81%. The accuracy in recognizing mathematically gifted children was 80.7%, while the model was less accurate in recognizing non-gifted children (hit rate of non-gifted was 72.92%).

The above results show that the best LR model produced higher hit rates than the best NN model. The accuracy in classifying gifted children was 80.70% vs. 78.28%. Due to the fact that neural networks and logistic regression used different modeling procedures, their results are not directly comparable. The LR model was obtained by using the whole dataset. Since the main focus of the paper was to extract important features for recognizing mathematically gifted children, the most important predictors are compared and commented in the next section.

6. Feature extraction

In order to get better insight into the importance of input variables in modeling mathematical gift, the sensitivity analysis was performed on the best NN model by using the test data. The results are presented in Table 10, where variables that have sensitivity ratio 1 or higher, have a positive effect to the performance of the model, while the variables with the ratio 1 or lower have no effect or have a negative effect to the performance. It can be seen that the first five most important variable are: (1) MATH – pupil’s grades in mathematics, (2) NUMBERS – the level of pupil’s success with numbers, (3) LIT – pupil’s grades in the literature, (4) ENVIRON – the level of support obtained by environment, and (5) pupil’s activities in foreign languages. It can be noticed that the pupil’s GPA, as well as some mathematical competencies, such as SHAPES – the level of dealing with shapes, PROBLEMS – the level of solving mathematical problems, belong the group of feature that have less influence to the mathematical gift, but the sensitivity ratio of those variables is above 1, indicating its valuable contribution to the output variable. The variables that could be omitted from the model, according to sensitivity analysis, are: ACTIV_S – activities in sports, ACTIV_M – activities in music, ACTIV_S – activities in the literature and drama, ACTIV – active learning and exercising component, COGN – cognitive component, and ACTIV_A – activities in fine arts.

The best LR model extracted four main cluster components where the cluster 1 has the highest importance (see Table 7) and was the only one found statistically significant on the level of 0.05. This cluster consisted of 10 variables: NUMBERS, MEASURES, SHAPES, PROBLEMS, COGN, ENVIRON, ACTIV, GPA, MATH, and LIT. Among these 10 variables, four of them describe mathematical competencies, three of them describe a pupil’s grades (GPA, MATH, and LIT), while others belong to the group of cognitive factors, environmental factors, and out-of-school activities that influence mathematical gift. All the variables belonging to cluster 1 had R-squared value with own cluster higher than 0.5, and they together form a set of important features for recognizing mathematical gift. The model did not evident a statistically significant influence of other variables on the observed dataset.

If the extracted features of two different methodological models were compared, it could be seen that eight variables extracted by the NN model were also included in the cluster 1 components of the best LR model. The list of predictors extracted by both models is shown in Table 11.

Sets of variables describing mathematical competencies and environmental factors were extracted with a high influence to...
the output by both NN and the LR model. It is interesting that the grade of a pupil in the literature course is also among the important predictors, as well as the grade point average. Out-of-school activities showed no significant influence to the output by both NN and the LR model.

7. Discussion and conclusion

The paper deals with modeling children’s mathematical gift by using neural networks and logistic regression. The initial model included five components: (1) mathematical competencies, (2) cognitive components of gift, (3) personal components that contribute gift development, (4) environmental factors, and (5) efficiency of active learning and exercising methods, as well as grades and out-of-school activities of pupils in the fourth year of elementary school. The three neural network classification algorithms were selected on the basis of a 10-fold cross-validation procedure. The model was also tested by the logistic regression. The results of the best neural network model and the logistic regression model were compared on the basis of model accuracy and the selection of important variables. Although the logistic regression shows better accuracy in terms of the average hit rate of giftedness category, the neural network model is more successful in recognizing gifted pupils. Both models extract a similar set of input variables as relevant. Therefore, a set of features relevant for recognizing mathematical gift can be extracted and emphasized for taking into consideration while deciding on a child’s mathematical gift. Besides mathematical competencies and grades, important features include environmental factors showing that the environment (school, family, and community) with its measures can also affect realization of a child’s mathematical gift.

Future research could focus on improving the accuracy of the best model, as well on testing some other methods such as support vector machines, decision trees and other intelligent methods in addition to statistics. The suggested model could be used as a decision support tool for teachers in schools as well as a basis for further research in this area.

References


