Assessment of Features of Cognitive Functions and the Social Sphere of Children and Adolescents Using Data Analysis Methods

Alexey I. Molodchnekov[1,2], Leyla S. Namazova-Baranova[3] and Georgy A. Karkashadze[3]

[1] Federal research center "Computer science and Control" of RAS, Russia
[2] Peoples friendship University (RUDN), Russia
[3] Central Clinical Hospital of RAS, Russia
aim@tesyan.ru

Abstract. The paper describes the results of the analysis of population studies among schoolchildren to assess their cognitive functions. The object of research was the cognitive, social, and other characteristics of children and adolescents. Methods of machine learning and statistics were used for data analysis. As a result, dependencies between the social and cognitive parameters of children and school performance were identified.

Keywords: artificial intelligence, machine learning, cognitive.

CEUR-WS.org/Vol-2790/paper23.pdf

1 Introduction

Studies aimed to find relationships between the processes of various areas of mental activity (cognitive, emotional -personal and social) in children and identifying predictors of successful or not successful social functioning are an actual area of modern neuroscience. For example, a study of 366 children under 10 in Italy showed that nonverbal intelligence was statistically associated with the ability to recognize emotions in 3 phases of development. The use of this model showed a significant effect of the cognitive aspect on the reflexive phase [1]. A study by Chinese scientists showed that the emotional understanding of words completely in children completely mediated the effect of age on empathy [2]. American scientists have studied early personality and environmental predictors of the development of empathy in young children, as well as the relationship of empathy to prosocial behavior with peers at a later age. The results suggest that both parental and personality characteristics are related to the development of empathy in early childhood and may contribute to the later prosocial behavior of children with peers [3].

However, population studies on this issue are relatively small and usually focus on the connection of two or three indicators, for example: the relationship between social perception tasks and language tests [4], the relationship of emotional intelligence with traditional violence and cyber violence [5], the connection between childhood stresses and personality disorders [6] or the relationship of behavioral problems with motor

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

and intellectual disorders [7] - that is, they are narrowly targeted. Other population studies include a larger number of psyche analysis points, but solely for assessing other factors, and not for analyzing intrasystemic relationships and dependencies, for example: using indicators of cognitive abilities, perception of emotions and social behavior to confirm the constructiveness and validity of the emotional intelligence assessment test in Great Britain [8]; assessment of cognitive abilities and mental health of a population of children from closely related marriages in India [9]; genetic influence on the relationship between simultaneously measured intelligence and academic performance in childhood in representative cohorts from England and the Netherlands [10].

This article presents the results of an assessment of population-based studies of the cognitive functions of children 11 and 15 years old in several regions of Russia. To analyze the data obtained, methods of machine learning and statistical data analysis were used.

2 Assessment of Population-Based Studies

As part of the RFBR project, a population-based study of the cognitive functions of children and adolescents aged 11 and 15 was conducted. The following cognitive features of children were assessed using special methods: "The volume of short-term auditory-speech memory", "Visual-figurative thinking", "Verbal-logical thinking", "Constructive praxis", "Percentage of correct answers according to Mnemotest", "Average number errors by Mnemotest", "Attention to the arrangement of numbers", "The amount of delayed auditory-speech memory". In addition to these indicators, the following were assessed: duration of sleep, time of falling asleep, headaches, time spent at the computer, playing sports and others. All participants underwent: cognitive testing, psychological examination, examinations of a neurologist, pediatrician, allergist, ENT, ophthalmologist, orthopedist, ultrasound of the abdominal cavity, urinary system, thyroid gland, ECG, spirometry and laboratory examination. The study involved more than two thousand people from several regions of Russia. There were 1000 children are 11 years old and 1012 children are 15 years old. As a result, more than 20 different parameters were collected for each subject. The following tasks were set:

1. check how many groups students can be divided by their cognitive functions;
2. evaluate the degree of influence of different cognitive indicators on the division into groups;
3. check the relationships between different groups of indicators in each group.

To solve the tasks, the following methods were successively applied:

1. clustering of data obtained from the results of testing the cognitive functions of children;
2. the use of classification algorithms to identify cognitive traits that affect the breakdown into classes;
3. the use of classification algorithms to identify non-cognitive signs that affect the assignment of a child to a particular class;

The use of regression analysis to identify the relationships between different parameters of groups (cognitive activity, emotional and personal characteristics and indicators of socialization of schoolchildren) of children and their breakdown into groups.

2.1 Data Clustering

Clustering was carried out only on the base of the results of cognitive research. Clustering was evaluated in two ways.

The first was to review the results of clustering by experts. These results showed that in both clusters there are children who can be very well attributed to one of the classes, and there are children who are in the border zone. This is due to the fact that in some indicators children have good results, in others it is worse. You can see it on figure 9. Moreover, the indicators themselves are completely different.

To select the number of clusters and assess the quality of clustering, a silhouette coefficient was used [11]. To cluster children, the k-means algorithm was used. The figures 1-8 below show the values of this coefficient for a group of 15-year-old and 11-year-old children, for the number of clusters from 2 to 5.

It can be seen that it takes the highest value for two clusters, i.e. this group of children is best divided into two classes.

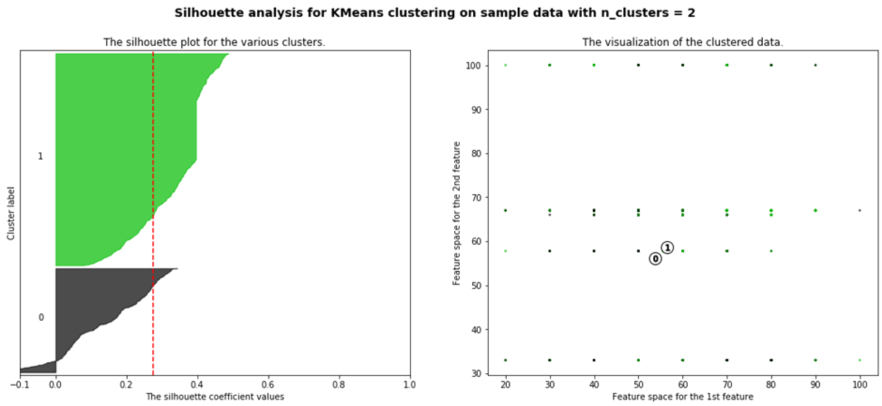

Fig. 1. Silhouette analysis for 2 clusters in 15-year-old children group.

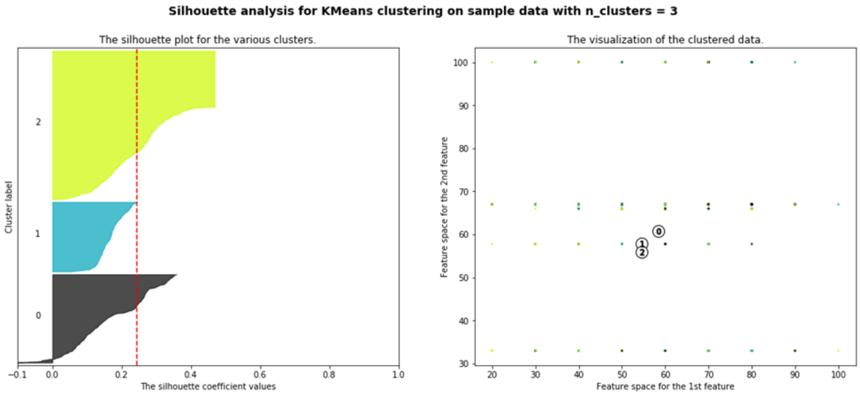

Fig. 2. Silhouette analysis for 3 clusters in 15-year-old children group.

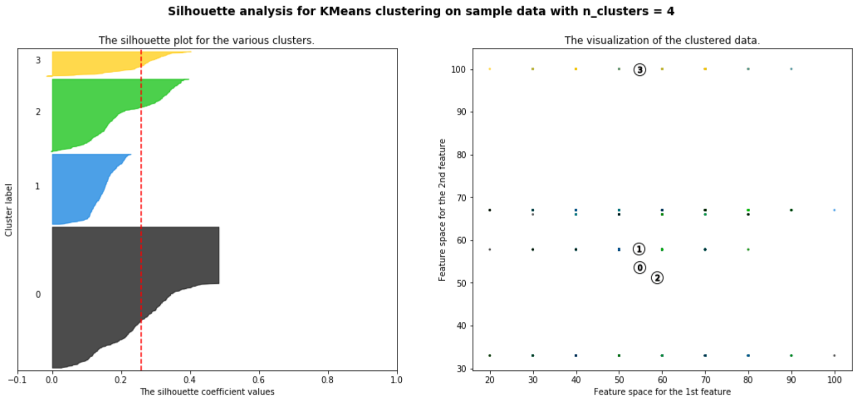

Fig. 3. Silhouette analysis for 4 clusters in 15-year-old children group.

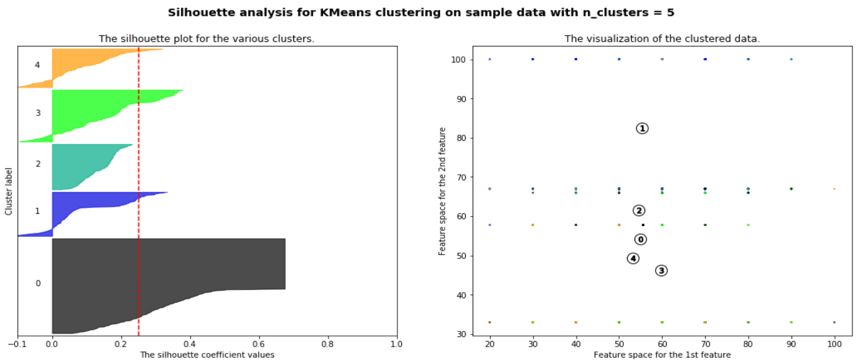

Fig. 4. Silhouette analysis for 5 clusters in 15-year-old children group.

Results for 11-year-old children group are shown bellow.

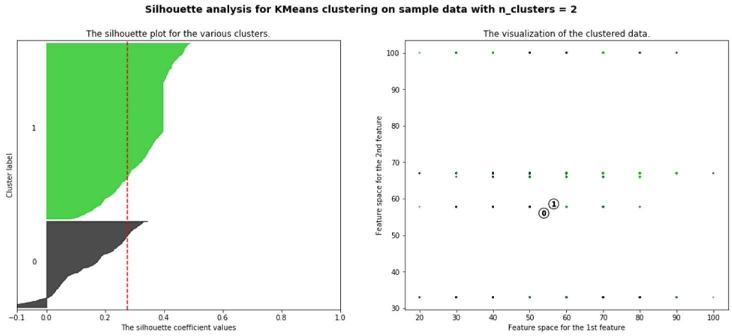

Fig. 5. Silhouette analysis for 2 clusters in 11-year-old children group.

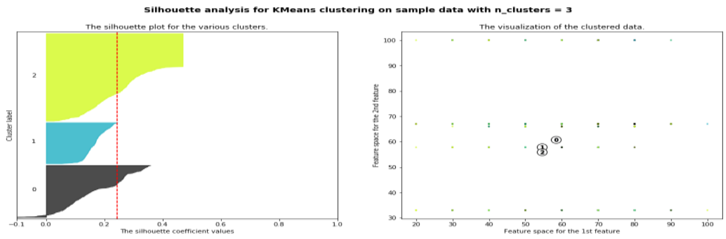

Fig. 6. Silhouette analysis for 3 clusters in 11-year-old children group.

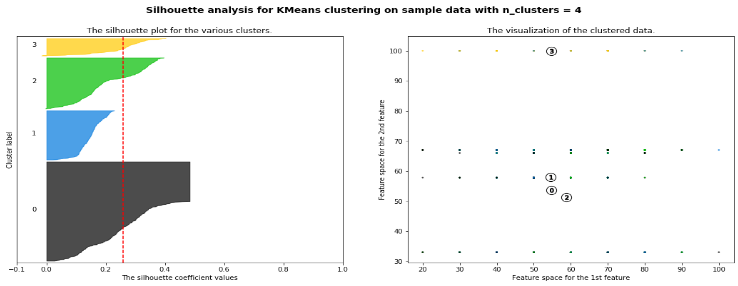

Fig. 7. Silhouette analysis for 4 clusters in 11-year-old children group.

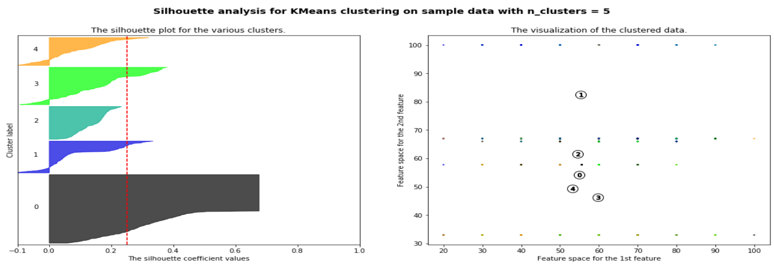

Fig. 8. Silhouette analysis for 5 clusters in 11-year-old children group.

Clustering based on the results of testing cognitive functions made it possible to distinguish two clusters in each age group. Figure 9 shows the location of objects of each class in two-dimensional space, 15-year-olds on the left, 11-year-olds on the right. The TSNE algorithm was used to reduce the size of the vectors.

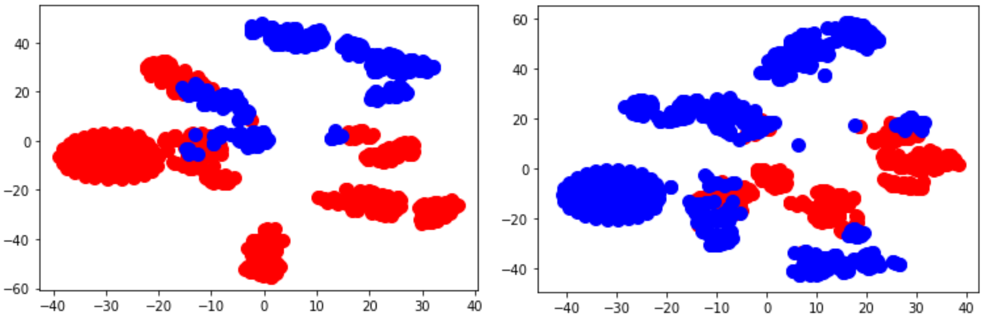

Fig. 9. Location of objects of each class in two-dimensional space.

The first cluster (shown in blue in the figure and indicated by the number 0) was characterized by higher cognitive indicators, primarily in terms of constructive praxis and verbal-logical thinking. The second cluster (shown in red in the figure and indicated by the number 1) was associated with lower cognitive parameters, also for the most part in constructive praxis and verbal-logical thinking.

2.2 Get the Most Important Features for Children Assigning to the Classes

To identify the features that make the greatest contribution to dividing children into classes, the following classification algorithms were used: random forest and the support vector method. Using these algorithms, models were constructed for dividing children into classes separately based on the results of cognitive testing and non-cognitive attributes. The results of assessing the quality of the machine learning models by the f1-measure are showed in the table 1.

Table 1. Results of ML algorithms tests.

	Random forest		SVM	
	Cognitive features	Non cognitive features	Cognitive features	Non cognitive features
11-year-old	0,96	0,7	0,99	0,76
15-year-old	0,94	0,65	0,99	0,79

The use of clustering and classification methods showed that children in both groups are well divided into two classes according to the results of cognitive studies. Moreover, for 15-year-olds, the main cognitive indicators for separation are: “Constructive praxis, accuracy of completing tasks%”, “Verbal-logical thinking, accuracy of completing tasks%”, “Mnemotest,% of correct answers”. For 11-year-olds: “Constructive praxis, accuracy of completing tasks%”, “Constructive praxis - average time for completing tasks, points" "Verbal-logical thinking, accuracy of completing tasks%". In addition to these signs, non-cognitive signs were identified that allow children to be divided into classes. Among these indicators, the most significant were: Sleep time, lessons with tutors, lessons, football, hockey, equestrian sports, time spent on computer games and some others.

2.3 Assessment of the Relationships Between Features of Different Groups

To identify the relationships between groups of features among themselves, methods of statistical analysis were used. Correlation matrices were built, on the basis of which the relationships and dependencies between the signs were identified. To construct the matrices, the Pearson correlation coefficient (r) was used.

First of all, an analysis of the compatibility of features among the participants was carried out. The goal was to determine how certain social and personal features are grouped in the examined population. Such a grouping can make it possible to divide the population into different socio-cultural, socio-personal typologies. As a result, also some identified combinations of features can be considered as certain socio-cultural markers. Especially this can concern those socio-cultural features that cannot be directly formulated in the questionnaire, for example, the material situation of the family, or the level of family attention to extracurricular needs of children. However, only the high compatibility of these signs could reliably confirm their suitability as markers of the material situation of the family.

At the first stage, we separately analyzed the compatibility of various sports and various hobbies. For both age groups, a very high correlation is established between tennis and hockey ($r = 0.810$ for 15-year olds). Moreover, it is maximum and even exceeds the compatibility of such similar species as roller-skating and ice-skating. Such a result requires a search for features that combine tennis and ice hockey. These sports vary in seasonality and nature (team and non-team playing sports).

The next step was to divide all the features into groups according to the proximity of the correlation value relative to each other and other parameters (clustering of features). The results demonstrate the grouping of sports, most pronounced in 15 year olds, according to which three clusters can be distinguished. The first cluster is roller-skating, ice-skating, and classic skiing, adjacent to them are skateboarding, downhill skiing and table tennis. The second cluster consists of: football, basketball and volleyball, which adjoins a bunch of tennis, hockey. It is easy to see that all these are playing sports, with the overwhelming majority being team sports. The third cluster is represented by martial arts, equestrian sports and swimming. These are a variety of individual sports. The identification of three clusters, which are so clearly distinguished by the nature of sports activity, indicates regular differences in the basic psycho-emotional characteristics of the participants, otherwise such a clustering would be impossible. This allows us to consider three clusters (conditionally "skating", "gamers" and "individuals"), as marking the different sports psychotype of the participants. Thus, behind each of the established sports psychotypes are certain emotional and characterological components of the emerging personality. In this regard, further analysis will be of particular interest, how these sports psychotypes will differ in cognitive, somatic and other social and personal parameters.

Interestingly, a similar, but more diffuse, clusterization was observed for 11 year olds. They distinguished not three, but two clusters: "game types" and "speed skating." Individual sports were evenly scattered around the edges of these two clusters. That is, we can say that at the age of 11, the division into sports psychotypes is already present, but more generalized, and in the future by the age of 15 it becomes more pronounced. This confirms that this clustering is based on the characteristics of the emerging personality, which, as is well known, as they grow older, they form from fragments into integral forms.

A similar algorithm was used to analyze the compatibility of various non-sports hobbies. Strong ties in the compatibility of various hobbies were not revealed in any of the age groups. Eleven-year-olds see a gradation in groups of combinations: 1) music, dance, an art school - conditionally a group of art; 2) robotics, modeling - conditionally a group of technicians; 3) computer programming, photo and video equipment - a conditionally digital group. At 15 year olds, the grouping is changing. Two clusters are clearly distinguished, and they are not identical in composition to 11-year-olds: now robotics is grouped with computer programming, and photo and video technology is grouped with modeling. The latter group is attracted by music and less obviously dancing.

Next, an analysis was made of the compatibility of the social signs, but not of the emotional-personal ones, since the relationship between social and emotional-personal manifestations is the subject of a separate analysis. For estimation, we used the calculation of the Pearson correlation coefficient (r). A value from 0.5 to 1 was estimated as a strong bond, from 0.2 to 0.5 - as a moderately strong bond, from 0.2 to -0.2 - as a lack of bond, from -0.2 to -1 - as a negative bond.

First of all, it should be noted that there is a strong connection between computer games playing on weekends and school days (r = from 0.703 to 0.280 for 4 different amounts of time): those who plays more time on weekends they also play more time on school days - and vice versa. This indicates that increased gaming interest is not subject to substantial regulation - there are very few situations where there are few games on school days and a lot on weekends.

A relationship was found between the amount of time devoted to computer games and the use of the Internet for non-gaming purposes (r = from 0.421 to 0.208 for 5 different volumes of time). It is maximally expressed for adolescents who devote up to 1 hour to games and the Internet on school days. This suggests that the enthusiasm for computer games and the use of the Internet by adolescents are interrelated phenomena.

Passion for computer games is associated with a lack of sports interests. Teenagers who spend 3 to 4 hours and more than 4 hours on computer games on weekends, as well as more than 3 hours on school days, do not engage in any sports regularly (r = 0.430; 0.325; 0.239). Interestingly, children who play sports regularly at the section level find a twofold connection with computer games on weekends: a moderately strong connection with a lack of enthusiasm for computer games (r = 0.346), as well as a moderately strong connection with computer games from 2 to 3 hours in days off (r = 0.349) - it is possible that these can be sports simulators. Teenagers who are fond of sports at the level of domestic / street competitions are often associated with computer games on weekends from 1 to 2 hours (r = 0.462). In general, these data indicate that sports enthusiasm does not exclude enthusiasm for computer games. But a strong passion for computer games is poorly combined with sports activities. Also partly a strong passion for sports is not compatible with computer games in general.

Adolescents who do not play sports regularly also do not get involved in summer tourism with a sports orientation (r = 0.255), which is logical. In turn, those who are not fond of sports tourism do not attend any summer camps (r = 0.335). Apparently, in this case we are talking about a certain typology of an individual character and family structure. A moderate positive relationship between the passion for music and skiing (r = 0.206) is shown, which is a very interesting fact for further interpretations.

It was found that the time to fall asleep, up to 10pm, is moderately strongly associated with the duration of sleep from 8 to 9 hours (r = 0.241), while the time later to fall asleep, after 11pm, was associated with a short sleep less than 8 hours (r = 0.397).

A direct relationship between the level of cognitive functions and success in school education in mathematics, literature and the Russian language is revealed. According to the results, it was found that higher cognitive parameters are more common in children who sleep more than 8 hours, have non-sporting passions, and ski. Lower cognitive parameters are more common in children who have smoking experience, play sports at the section level, engage in dancing, hockey, and martial arts. No connections were found between the level of cognitive functions and classes with tutors, most unsportsmanlike and sporting hobbies, visiting summer camps, the volume of computer games on weekends, and the time of falling asleep at night.

It was established that success in schooling positively correlates with a duration of sleep of more than 8 hours, unsportsmanlike hobbies, skiing. At the same time, success in schooling is negatively correlated with smoking experience, sports at the section level, dancing and hockey.

3 Conclusion

The article describes the results of population studies. Models are built for dividing children into classes according to their level of cognitive function. The cognitive characteristics that make the greatest contribution when dividing students into classes are determined. The relationships between different groups of characteristics of children in these classes are revealed. The research results show the contribution of somatic and social factors to the formation of children's mental health and the success of their education.

The reported study was funded by the Russian Foundation for Basic Research (project No. 17-29-02501 ofi_m).

References

1. Albanese O., De Stasio S., Di Chiacchio C., Fiorilli C., Pons F.: Emotion comprehension: the impact of nonverbal intelligence. J Genet Psychol. 171(2):101-115 (2010)
2. Li, Y., Yu, D.: Development of Emotion Word Comprehension in Chinese Children from 2 to 13 Years Old: Relationships with Valence and Empathy. PLoS One. 10(12):e0143712 (2015)
3. Taylor, Z. E., Eisenberg N., Spinrad, T.L., Eggum, N.D., Sulik, M.: The relations of ego-resiliency and emotion socialization to the development of empathy and prosocial be-havior across early childhood. Emotion. 13(5):822-831 (2013)
4. Loukusa, S., Makinen, L., Kuusikko-Gauffin, S., Ebeling, H., Moilanen, I.: Theory of mind and emotion recognition skills in children with specific language impairment, autism spectrum disorder and typical development: group differences and connection to knowledge of grammatical morphology, word-finding abilities and verbal working memory. Int J Lang Commun Disord. 49(4):498-507 (2014)
5. Fenning, R. M., Baker, J. K., Baker, B., L., Crnic, K. A.: Parenting children with borderline intellec-tual functioning: a unique risk population. Am J Ment Retard. 112(2): 107-121 (2007)
6. Afifi, T. O., Mather, A., Boman, J., Fleisher, W., Enns, M. W., Macmillan, H., Sareen, J.: Child-hood adversity and personality disorders: results from a nationally representative pop-ulation-based study. J Psychiatr Res. 45(6):814-822 (2011)
7. Vrijmoeth, C., Monbaliu, E., Lagast, E., Prinzie, P.: Behavioral problems in children with mo-tor and intellectual disabilities: prevalence and associations with maladaptive per-sonality and marital relationship. Res Dev Disabil. 33(4):1027-1038 (2012)
8. Mavroveli, .S, Petrides, K. V., Sangareau, Y., Furnham, A.: Exploring the relationships be-tween trait emotional intelligence and objective socio-emotional outcomes in child-hood. Br J Educ Psychol. 79(Pt 2):259-272 (2009)
9. Fareed, A.: Estimating the inbreeding depression on cognitive behavior: a popula-tion based study of child cohort. PLoS One. 9(12):e115393 (2014)
10. Calvin, C. M., Deary, I. J., Webbink, D., Smith, P., Fernandes, C., Lee, S. H,. Luciano, M., Visscher, P. M.: Multivariate genetic analyses of cognition and academic achievement from two popu-lation samples of 174,000 and 166,000 school children. Behav Genet. 42(5):699-710 (2012)
11. Aranganayagi, S., Thangavel, K.: Clustering categorical data using silhouette coefficient as a relocating measure. Proc. International Conference on Computational Intelligence and Multi-media Applications (ICCIMA 2007). V. 2., pp. 13-17. IEEE (2007)