Data Mining Analyses of Learning Preferences Among Gifted Students

Zdena Lustigova[1], Veronika Novotna[2]

Abstract

This paper presents selected results of data mining analysis of educational and psychological/psychometric data, collected within 5 years on the group of 91 children, who participated in a specific online educational program (math, programming and science) for gifted children, organized by Charles University in Prague. The information record for each student is represented by 151 variables, of both categorical and quantitative nature, describing 1) demographic characteristics, 2) personal characteristics, including motivation and intelligence 3) behavioral and action records and 4) particular educational results and learning paths. The study is based on comparison of students with very similar personal characteristics like motivation and intelligence and their study interests and results. Students, who chose the right courses, that matched their nature of talent and personal characteristics, succeed. Those, who for some reason choose the course that does not match his /her nature of talent, are more likely to fail. Both, course selection and connected success or failure within selected course, is affected mostly by all components of intelligence, except of crystalline, and also by selected components of motivation, especially Dominance, Independence and Persistence. The nature of talent seems to be determined besides factors mentioned above by selected components of creativity.

1. Introduction

The education domain offers a fertile ground for many interesting and challenging data mining applications. These applications can help both educators and students to improve the quality of education. In [2] Ma, Liu, Wong and others deal with Gifted Education Program (GEP) of the Ministry of Education (MOE) in Singapore. They focus mainly on better selection of students for remedial classes, since traditional methods choose too many participants, which increase the teaching load of the instructors and slow down the real GEP students.

Identification of gifted children using neural networks within the online environment is also the topic of the paper [3], where Bae, Ha and Park offer a special questionnaire and use it to measure the implicit capabilities of giftedness and to cluster the students with similar characteristics. The neural network and data mining techniques are applied to extract a type of giftedness, their characteristics, and their learning path. Kamath and Srimani in [4] deals with gifted children performance analysis using generic data mining model, that validates the accuracy and efficiency of the learning model and leads, according the authors, to more reliable and authentic predictions. Cluster analysis as a specific technique was used few times by Parker ([5] and [6]) to identify perfectionism. A nationally gathered sample of 820 academically talented sixth graders took the Multidimensional Perfectionism Scale, and scores were cluster analyzed using both hierarchical and non-hierarchical cluster analysis with cross-validation. A three-cluster solution was indicated. Parent perceptions of the children were consistent with the students' self-perceptions. The construct of perfectionism was primarily associated with conscientiousness and secondarily with agreeableness and neurosis. Both studies did not identify statistically significant differences between gifted students and the general cohort.

[1] Charles University in Prague Czech Republic

[2] Charles University in Prague Czech Republic

2. Description of a research problem

2.1 Research design

The presented research is based on descriptive analysis of data collected during 5 years (2006-2011) on the group of 91 children, participating in the specific online teaching and learning program for gifted children in Czech Republic.

The collected information record for each case (child) was represented by 151 variables, both categorical and metric nature, describing on one side 1) demographic characteristics (age, gender, school, town, nationality, etc.), 2) personal characteristics, including motivation and intelligence, 3) behavioral and action records, including individual decision making records (e.g. course set selection, preferences in teacher selection, etc.) and 4) their learning paths based mainly on particular educational results.

Personal characteristics, e.g. 17 components of motivation, were obtained on the base of Achievement Motivation Inventory (AMI) psychological test, developed originally by Schuler & Prochaska [10] under the German name Leistungsmotivationsinventar (LMI) in 2000, but largely used since 2002 under the abbreviation AMI [11]. The terminology used in the following reflects approaches and tests, mentioned above.

The values obtained by AMI and assessing 17 dimensions of work/study related achievement motivation were represented by variables Var37-53: Compensatory Effort, Competitiveness, Confidence in Success, Dominance, Eagerness to Learn, Engagement, Fearlessness, Flexibility, Flow, Goal Setting, Independence, Internality, Persistence, Preference for Difficult Tasks, Pride in Productivity, Self-Control, Status Orientation.

The intelligence and its components (represented by V22-28: IQ_crystalized, IQ_fluid, IQ_verbal, IQ_numeric, IQ_figural, IQ_memory, IQ_knowledges) were measured by professional psychologists on the base of WISC-IV tests.

The tests for measuring the level of critical thinking and its components (inference, recognition of assumptions, deduction, interpretation, evaluation of arguments) as well as the test for measuring creativity and its components (fluence, flexibility, originality and elaboration) and the test for emotional intelligence were involved. Cooperating psychologists decided to consider emotional intelligence as a prerequisite for successful learning (inspired by Goleman, e.g. [13]). Social component could not be quantified.

In all four cases of psychological/psychometric tests, the scores by components, raw scores, total scores and normalized scores were available and used according the research purposes and the way of data processing.

The whole data set included more than 12000 data cells, both measured and recorded. Approximately one eighth (12%) of expected full data set (13740 values) was unavailable (1740 missing data cells). The data unavailability reasoned mainly from external sources, e.g. parents' awareness of their child's IQ or personality tests or lack of teachers' online records, concerning particular student's educational results.

2.2 Data processing method

Besides descriptive statistics cluster analysis was used to identify homogeneous subgroups of variables or cases in a given population of all 151 variables and 91 gifted students.

Cluster analysis as a main task of exploratory data mining, was used for grouping sets of objects/variables in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups. Cluster analysis is used wherever the researchers do not know the number of groups in advance, but wants to identify and then to analyze group membership.

Within the study just two from existing three general approaches to cluster analysis were used, namely hierarchical clustering and two-stage cluster analysis. Specific approaches, integrating classification and association rule mining, etc., were adopted from [7, 8, 9].

Before interpretation, cluster validity was estimated on the base of three criteria:

1) Cluster size,
2) Meaningfulness,
3) Criterion validity.

3. Selected results

Approximately 12 research questions were set up, focused mostly on the existence of a meaningful clusters among students according different combination of variables, e.g.:

1) Motivational characteristics of students and their a) Course selection (areas of interest) during 5 years, b) Study results and specific learning paths, c) Gender.
2) Intellectual characteristics (as a whole or particular components of intelligence, including creativity and EMI) of students and their a) Course selection (areas of interest), b) Study results and specific learning paths, c) Gender.
3) Mixed influence of Motivation and its components, intelligence (as a whole) or particular components, creativity, emotional intelligence, gender, etc.…

Other questions were focused on difference in achieved results between boys and girls, different types of intelligence, etc. Some questions were hypothesized, some not. Many new appeared during the data processing itself.

Basically, there seems to be strong relationship between course topic selection and the factor of **Dominance,** which might be summarized as following:

- The interest in Biology (and biology connected subjects e.g. environmental studies) were connected with less dominant individuals, while
- Physics (and physics connected subjects) proved to be selected by individuals with much higher dominance.

In connection with the higher dominance also both programming and math appeared (in particular years), but the relationship was not as strong as in case of physics.

The **"Goal setting"** and "I**ndependence"** variables proved a relatively high predictive value for course selection, too. "**Independence"** has caused the formation of two clusters in every particular year and during the whole program as well as Dominance. Again, the first cluster was dominated by biology and biology connected topics, the second cluster was dominated by different items in different years, but mostly connected with physics, followed by programming, and also by math. The study results (success or failure) were greatly determined by "**Persistence"**.

The influence of gender was not proved, neither rejected. In case of course selection and gender variables two clusters appeared, but their quality was not the best, see Fig. 1). In other words, although among those interested in biology were no boys at all, among „physicist“ 20% of girls appeared.

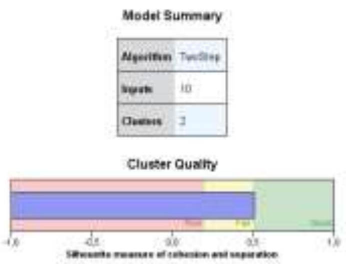

Fig.1 Example of results obtained. The quality of the two clusters formatted on the base of Course Selection variables V6-14 and variable Gender (V15).

The second group of questions revealed that almost all components of intelligence (V22-28, as well as the total IQ scores (V21), caused the formation of two clusters. The most significant effect both verbal intelligence and numerical intelligence proved. No, respectively low, effect was revealed for components IQ-crystalline, respectively IQ-knowledge. The success in particular courses was strongly influenced by verbal and numerical IQ components, while the others, including total IQ score, seem to play less important role in Talnet population. Quite strong influence was also proved for selected components of creativity, e.g. elaboration, fluency and flexibility.

4. Conclusions

The course selection prediction seems to be affected of all the studied variables mostly by:

- Intelligence, by all of its components except of crystalline. The most significant impact was observed in case of verbal and numerical component.
- Motivation, namely following components: Dominance, Independence, Persistence and Goal setting, where the first 3 affect the course selection, while the last one the number of courses taken and finished.

The course selection is not influenced by critical thinking total score, neither by any of its components.
The success in a selected course is related mostly to the nature of particular student's talents. If the students chose the appropriate courses (courses that matched their nature of talent), they succeed. An individual who, for some reason, chooses the course that does not match his/her nature of talent is more likely to fail.
The nature of talent (as mentioned before) seems to be determined by:

- All components of intelligence, except of crystalline;
- Selected components of motivation, namely Dominance, Independence and Persistence;
- Selected components of creativity (elaboration, fluency, flexibility).

References

[1] Mooi,E., Sarstedt,M. (2011) A concise guide to market research. The process, data and methods using IBM SPSS statistics.
[2] Ma,Y., Liu,B., Wang, Ch.(2000) Targeting the Right Students Using Data. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. Boston, Massachusetts, USA — August 20 - 23, 2000.Pp.457-464
Available online ftp://ftp.cse.buffalo.edu/users/azhang/disc/disc01/cd1/out/papers/kdd/p458-ma.pdf
[3] Bae, S., Ha, S.H., Park, S.C. (2006). Identifying gifted students and their learning paths using data mining techniques. *WIT transaction on State of the Art in Science and Engineering.* Vol 2. WIT press. Available online http://library.witpress.com/pages/PaperInfo.asp?PaperID=18312
[4] Srimani, P.K., Kamath, A. (2012) Data Mining Techniques for the Performance Analysis of a Learning Model – A Case Study. *International Journal of Computer Applications*. Vol. 53– No.5.
[5] Parker, W.D. (1997). An Empirical Typology of Perfectionism in Academically Talented Children. *Am Educ Res J* September 21, Vol. 34. No. 3.
[6] Parker, W.D.(1996). The Incidence of Perfectionism in Gifted Students. *Gifted Child Quarterly*, Vol. 40. No. 4.
[7] Ling, C. and Li C. (1998). Data mining for direct marketing: problems and solutions. In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining KDD-98.
[8] Liu, B., Hsu, W. and Ma, Y. (1998). Integrating classification and association rule mining. In. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. KDD-98.
[9] Liu, B., Hsu, W. and Ma, Y. (1999). Mining association rules with multiple minimum supports. In Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining. KDD-99.
[10] Schuler,H., Prochaska,M. (2000). Entwicklung und Konstruktvalidierung eines berufsbezogenen Leistungsmotivationstests. In: Diagnostica. Vol. 46. No.2.
[11] Schuler, H., Thornton, G.C., Frintrup, A., Mueller-Hanson, R. (2002). Achievement Motivation Inventory (AMI). Göttingen, Bern, New York: Hans Huber Publishers
[12] Goleman, D. (1998). Working With Emotional Intelligence. New York, NY. Bantum Books.
[13] Lustigova, Z., Dufresne, A., & Courtemanche, F. (2010). Acquiring Physiological Data for Automated Educational Feedback in Virtual Learning Environments. *New Educational Review*, Vol. 21. No2.