Ostravská univerzita v Ostravě

Pedagogická fakulta, Katedra informačních a komunikačních technologií

Estimating the optimal number of initial learning paths to GLM module

Jacek Grudzień, Michal Burda

Abstract

The article describes the preparations for establishing a new, adaptive lesson module. It focuses on the group of students, which allows to determine the number of initial learning paths. It includes a discussion on several methods of clustering and the optimal number of learning paths.

1. Introduction

During the work on a new lesson module (GLM) [1], giving the possibility of implementation into the LMS Moodle and LMS Claroline systems, the requirement of preparation learning paths tailored to the needs of students has emerged. GLM can operate according to two strategies discussed widely in work [2]. The first of these involves the preparation of teaching materials according to the linear scheme, the second – the composition the initial paths of learning.

Creation of the initial path of teaching requires a thorough knowledge of students [3], which in practice leads to find the answers to some questions. First of all, it is necessary to find out whether they are smart or having difficulties students. It is also obligatory to specify the characteristics of students' skills in searching for analogies, making generalizations, memorizing or orientation in mathematical and geometrical problems [4]. Research should bring the answer to how many learning paths must be established and what content they should carry.

2. Research methodology

The research sample was 167 students randomly selected from all students in the school. The structure of intelligence Amthauer's test was performed on a sample of students and the test results were treated using the statistical package R [5].

Histograms of the test components were made and then the hierarchical clustering methods using the Manhattan metric: Ward [6] and Complete-linkage [7, 8, 9] were conducted separately. Based on the results of hierarchical clustering the optimal number of groups was chosen. The degree of similarity of every element to the proposed cluster was verified. The results are shown on the figures.

3. Results and discussion

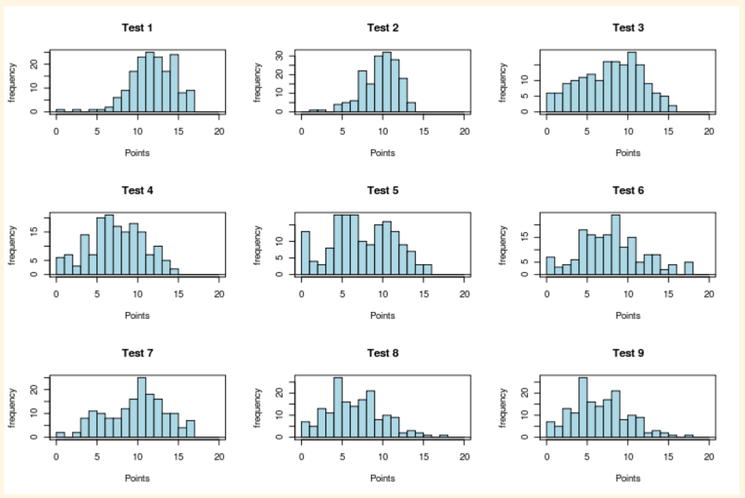

Figure 1 Component histograms of tests carried out on 167 students - raw data.

Figure 1 shows the raw data of the intelligence test structure. Based on the data presented in the histograms some hypotheses about the population of students can be drawn.

In the test 1 determining knowledge of a simple information from the surrounding world, the results are outstanding and demonstrates students' high ability to update their knowledge. Test 2 shows that students are able to find common characteristics and properties of objects or concepts and compare them. The average test results 7 and 9 of the whole population suggest an intermediate ability to think logically and remember information.

The results of the tests 3, 4, 5, 6 and 8 below average show a weak ability to inference and generalization (3, 4), poor skills of solving text exercises (5), lack the ability to make calculations (5, 6) or detecting dependence between the numbers (6). Very poor 8th test results indicate the absence of ability to work with three-dimensional objects among the entire population.

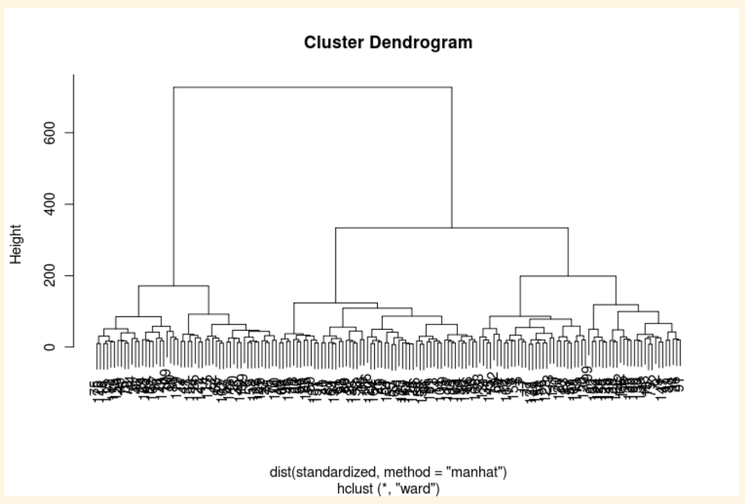

Figure 2 Ward's hierarchical clustering method with using the Manhattan metric.

After the hierarchical clustering using Ward's method and Manhattan metrics (Figure 2) population was divided into roughly equal groups of 57, 52 and 58 representatives.

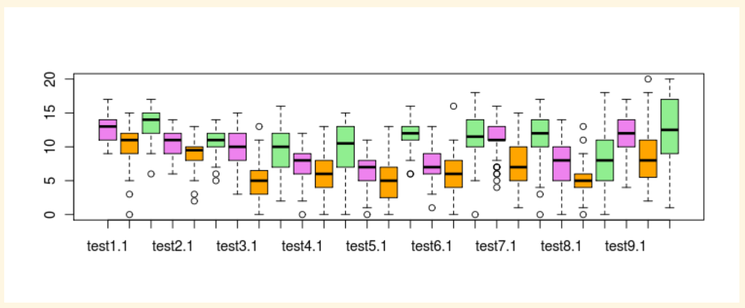

Figure 3 The split of the population into 3 clusters. Ward's hierarchical clustering method and the Manhattan metric. Count of clusters: 1 (violet) - 57, 2 (orange) - 52, 3 (green) – 58.

These groups are shown in the boxplot in Figure 3. As can be seen from the graph the population of students is not parted due to special skills (for example, due to mathematical ability, linguistic, or spatial) but because of the general level of intelligence. In other words, the population of students was divided into bright and average students and students with poorly developed intellectual components.

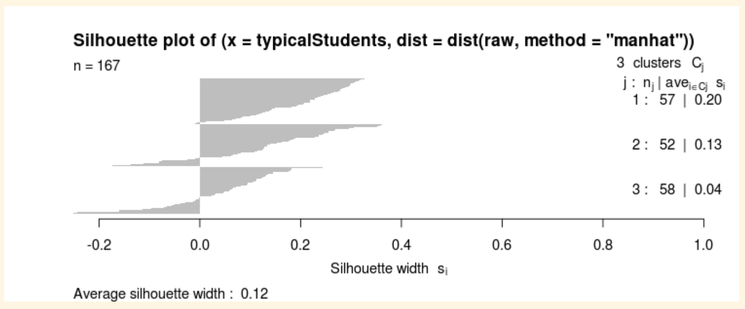

Figure 4. The confirmation the validity of a designated group.

Looking for matching objects to clusters shown in Figure 4, it is possible to confirm the validity of a designated group.

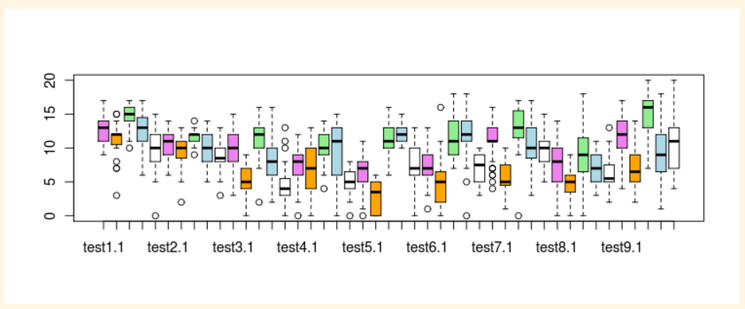

Figure 5 The division into 5 clusters. Ward's hierarchical clustering method and the Manhattan metric. Count of clusters: 1 (violet) - 57, 2 (orange) - 28, 3 (green) - 31 4 (blue) - 27, 5 (white) – 24.

The division of the students' population according to the directional abilities can be caused only by increasing the number of clusters. This situation is illustrated in Figure 5, where the cluster of intermediate students (1) remained unchanged, while the cluster of clever students is splitted into two parts: the cluster of thoroughly intelligent students (3) (i.e., having well-developed intellect components in each direction) and the cluster of math and computing-oriented ones (4). A cluster of students with poorly developed intellectual components is divided into the cluster of students poorly generalizing the concepts (5) and the cluster of students with reduced computing ability (2). It can be assumed that the cluster thoroughly intelligent (3) represents the students with broad interests, reading books and skillfully using the new media.

The cluster of mathematically and computationally oriented is represented by students prefering natural sciences and engineering subjects, which must gain more general knowledge.

Intermediate students appearing in the cluster (1) should be allowed to the harmonious development of all skills. For students with underdeveloped component of generalization (5) should be paid attention to incentive their work. The last group of students (2) requires the help of specialists.

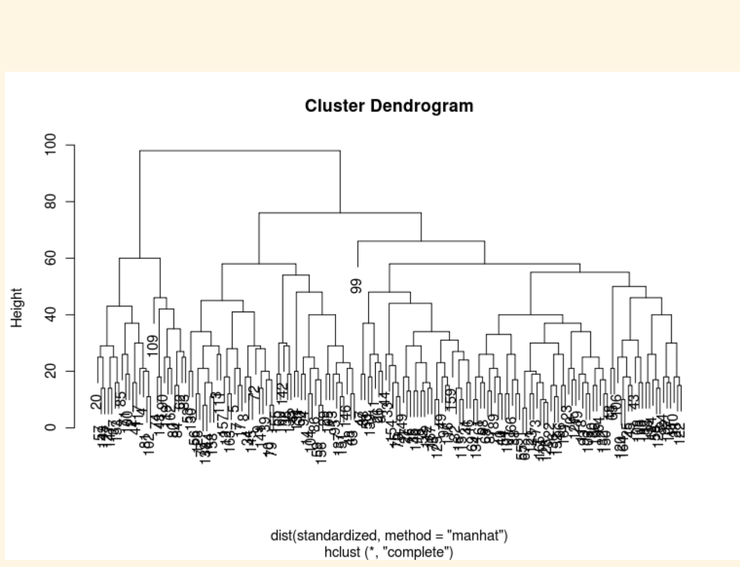

Figure 6 Complete linkage hierarchical clustering method with using the Manhattan metric.

A similar situation occurs after the clustering Complete-linkage method (Figure 6). Population estimated by this method can also be divided into three main clusters.

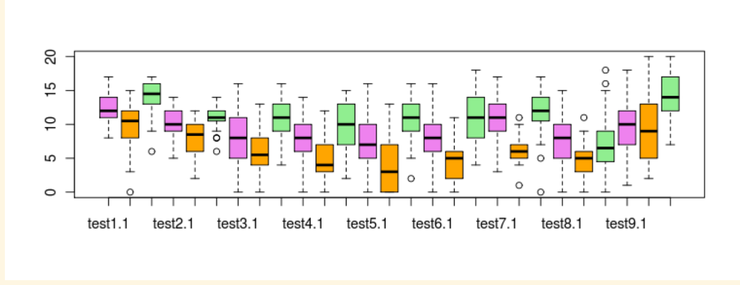

Figure 7 The split of the population into 3 clusters. Complete-linkage hierarchical clustering method and the Manhattan metric. Count of clusters: 1 (violet) – 93, 2 (orange) - 26, 3 (green) – 48.

Based on a boxplot (Figure 7) can be estimated that the initial division does not run due to the dominant type of intelligence, but because of the general level of intelligence.

In other words, clustering by Complete-linkage method leads to a distribution of almost identical to the Ward's method. The only change is the multiplicity of clusters - Complete-linkage method classified 93 students as an intermediate (cluster 1 – violet), 48 as a bright (cluster 3 – green) and 26 as weak (cluster 2 – orange).

Theoretically, the number of generated clusters determines the number of initial learning paths implemented to the GLM module. However, in practice, it should be considered the time-consuming preparation of each path. This means that the learning paths should not be multiplied. The optimal number oscillates around 2 – 4.

The division of the students' population has a decisive influence on the preparation of teaching materials. Initial prediction that the population is devided by the nature of intelligence tended toward the creation of student ability-oriented materials. It was assumed that depending on the type of intelligence learning units will be dominated by one of the media of communication. The results of the calculations, however, suggest just the opposite – the materials contained in the learning paths should include a similar number of media types, have analogous structure and differ only by the degree of difficulty.

Bibliography

[1] GRUDZIEŇ, Jacek Lessons with the use of artiffcial intelligence. In *Theoretical and Practical Aspects of Distance Learning : Use of E-learning in the Training of Professionals in the Knowledge Society*. Eugenia Smyrnova-Trybulska. I. Cieszyn - Katowice : Studio Noa Ireneusz Olsza, 2010. p. 5. ISBN 978-83-60071-30-4.

[2] GRUDZIEŇ, Jacek. Rozvoj informatických kompetencí pomocí umělé inteligence. In *Information and Communication Technology in Education: Proceedings*. Rožnov pod Radhoštěm : University of Ostrava, Pedagogical Faculty, 2011. [not publicated].

[3] ZOUBEK, Lukáš; BURDA Michal Computer Analysis of Learning Characteristics. In *Theoretical and Practical Aspects of Distance Learning : Use of E-learning in the Training of Professionals in the Knowledge Society*. Eugenia Smyrnova-Trybulska. I. Cieszyn - Katowice : Studio Noa Ireneusz Olsza, 2010. p. 5. ISBN 978-83-60071-30-4

[4] Смирнова, Евгения Николаевна. Развитие важнейших компонентов интеллекта на основе комплексного использования НИТ при обучении математике в старшей школе [Текст] : дис... канд. пед. наук: 13.00.02 / Смирнова Евгения Николаевна ; *Украинский гос. педагогический ун-т им. М.П.Драгоманова*. - К., 1996. - р 258 л.

[5] *R Project's website* [online]. 2011 [cit. 2011-08-14]. The R Project for Statistical Computing. Available from WWW: <http://www.r-project.org/index.html>.

[6] WARD, Joe H., Jr., Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*. March 1963, 58, p. 236-244. Available from: <http://iv.slis.indiana.edu/sw/data/ward.pdf>.

[7] SØRENSEN, Thorvald . A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons.. *Biologiske Skrifter / Kongelige Danske Videnskabernes Selskab.* 1948, 5, p. 1-34

[8] LUKASOVÁ, Alena ; ŠARMANOVÁ, Jana. *Metody shlukové analýzy*. Praha : Technické literatury, 1985. p. 67-72.

[9] KABACOFF, Robert I. . *Quick-R : acccessing the power of R* [online]. 2011 [cit. 2011-08-14]. Cluster Analysis . Available from WWW: <http://www.statmethods.net/advstats/cluster.html>

Conclusions

The presented GLM module can be improved on the basis of this research. The current research assumes the preparation of learning paths with contents relevant to the population of students. It was found that:

- change the hierarchical clustering method does not affect the basic division of the population of students,
- taking into account the matching objects to cluster the three initial learning paths will be prepared in the GLM module,
- didactic material on the paths should have a similar structure and different degrees of difficulty.

At the same time, the clustering attempts to verify the conclusions drawn on the basis of the results above are planned to be carried out. Further tests will include hierarchical clustering using Euclidean metric and non-hierarchical clustering.