Analysing Student Spatial Deployment in a Computer Laboratory

VLADIMIR IVANČEVIĆ,
University of Novi Sad, Serbia
MILAN ČELIKOVIĆ,
University of Novi Sad, Serbia
AND
IVAN LUKOVIĆ
University of Novi Sad, Serbia

Although data mining techniques are constantly improved and made more easy to use, there is still little application of related methods in the investigation of student seating choices and respective implications. Therefore, in this paper we analyse characteristics and effects of student spatial deployment in a computer laboratory by using data mostly gathered through electronic student testing system. For a first semester programming course, we examine relations between student assessment results and student choices of seating locations in a laboratory, as well as search for regularities in the spatial deployment, such as students' test location changes throughout the semester and the connection they share with the success on tests. Obtained results demonstrate the existence of patterns, whose meaning is also discussed, and emphasise the importance of considering student spatial choices in the study of understanding and improving student behaviour and learning.

Key Words and Phrases: student spatial deployment, student seating, mining assessment data

1. INTRODUCTION

Research on seating arrangements of students in classrooms has always been a familiar topic within the scientific community. Although the latest data analysis and investigation techniques have found their way into the field known as educational data mining (EDM), that has not led to substantial improvements in the study of student spatial deployment in a classroom, which may seem unusual since computers can contribute much to the automatisation of data gathering and shortening of the duration of the study.

In this paper, we propose an approach which focuses on the usage of data mining (DM) techniques in the examination of student spatial deployment in a classroom. Our main goal is to demonstrate how to prepare data and extract information necessary for the modern research on student seating locations. Analyses were conducted using data from a basic programming course organized at the Faculty of Technical Sciences (FTS) of University of Novi Sad and they include: investigation of student assessment results with respect to the student seating choices in the classroom; and investigation of migration and changes in student seating choices in the classroom with respect to assessment results.

2. RELATED WORK

Different studies concerning implications of choosing different seating locations in a classroom at the university level often have seemingly contradicting results. Some claim that seating location does not influence grades (Kalinowski et al, 2007) or that the impact

Authors' addresses: V. Ivančević, Department of Computing and Control, Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia. E-mail: dragoman@uns.ac.rs; M. Čeliković, Department of Computing and Control, Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia. E-mail: milancel@uns.ac.rs; I. Luković, Department of Computing and Control, Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia; E-mail : ivan@uns.ac.rs

is so weak that it can be considered unimportant (Montello, 1988). Others state that the seating location is linked to a better willingness to participate in classes (Çinar, 2010) or that the seating in the back leads to lower grades when compared to the front (Gossard et al, 2006). Moreover, even with the initially randomized seating arrangement, there was a positive impact of the seating in the front of the lecture hall (Perkins et al, 2005).

Unfortunately, the use of computers and data mining in those circumstances has been scarce, which could be attributed to the fact that an infrastructure for automatically tracking student seating locations is still not available in most classrooms, so the process of recording seating choices is impractical and labourious. In order to avoid the manual data collection, we relied on the automatic logging system available in the laboratory.

3. THE ENVIRONMENT

3.1 The Laboratory

The analysed course was held in a laboratory (see Fig. 1) specifically designed for computer science education at FTS (Rakić et al, 2007). It features 32 student work locations, each with a single personal computer uniquely identified by a number in the range of 200 to 231, and a separate computer for the instructor. In the same figure, the typical positions of teaching assistants are marked by stars: the leading teaching assistant (LTA) standing at the front of the classroom and presenting a topic; and the supporting teaching assistant (STA) moving around the room. Student evaluation in the course was conducted using a software for computer-based student testing developed at the same laboratory (Živanov et al, 2008), and monitored by teaching assistants, both directly and with the help of a surveillance system. In order to be analysed, the student work locations were divided according to their distances from the LTA in 3 location groups (see Fig. 1): *Front-Zone* (11 locations: 200-202, 210-212, 220-224); *Mid-Zone* (11 locations: 203-206, 213-216, 225-227); and *Back-Zone* (10 locations: 207-209, 217-219, 228-231).

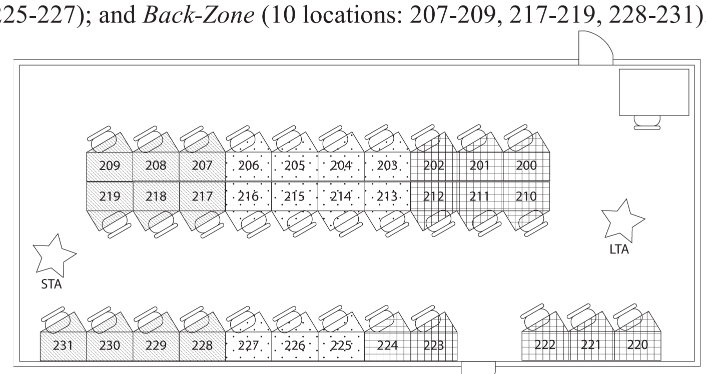

Fig. 1. An overview of the laboratory

3.2 Data Description

We examined students' results achieved on tests conducted as a part of the *Programming Languages and Data Structures* course, the first year mandatory course of *Computer and Control* curriculum carried out at FTS in the winter semester of academic year 2010/2011. The data encompassed 3 tests with 153 students involved and 430 individual assessments. These data were extracted from the testing system logs and include: course, test, and student group identification; basic student data (id, first name, and last name); and test data (achieved and maximum score, computer IP address, and completion time).

4. ANALYSING SPATIAL DEPLOYMENT

Further in this section we present our analyses conducted through *Oracle Data Miner* software tool. As we attempt to make general conclusions relying on 3 tests data, we need to make one assumption about the environment: *student seating choices at tests are representative of their seating choices during the non-test classes*. This claim is supported by the practical observations made during the course and by the findings presented later showing that many students chose the same location or the one nearby.

4.1 Analysing Student Scores with Respect to Location

For each location in the laboratory, we determined the number of students who sat at that location during the assessments (*SCN*), as well as the sum of all the test scores made by those students when they sat there (*SCSUM*). Next, for each location we calculated its corresponding average test score *SCAVG* as the quotient of *SCSUM* and *SCN*. Since FTS uses a six-point grade scale from 5 (unsatisfactory) to 10 (excellent), these single test scores (integers from the range of -5 to 10) can also be interpreted as grades. We clustered the locations by their *SCAVG* and *SCN* values using an enhanced version of the *k*-Means algorithm (Oracle, 2008) and obtained 4 groups: *s1* (characterised by wide score range and low *SCN*), *s2* (low *SCAVG* and high *SCN*), *s3* (high *SCAVG* and moderate *SCN*), and *s4* (moderate *SCAVG* and moderate *SCN*). Their properties are shown in Table 1, together with member distributions among 3 location zones (*Front*, *Mid*, and *Back*).

Table I. Locations clustered by average test score and individual test count

<table>
<tr><th>Clust.</th><th>Size</th><th>Confidence[%]</th><th>Support Count</th><th>Front[%]</th><th>Mid[%]</th><th>Back[%]</th></tr>
<tr><td rowspan="2">s1</td><td>6</td><td>100</td><td>6</td><td>17</td><td>33</td><td>50</td></tr>
<tr><td colspan="6">Rule: 4.295 <= SCAVG <= 8.21 and 10.0 <= SCN <= 12.5</td></tr>
<tr><td rowspan="2">s2</td><td>6</td><td>83.3</td><td>5</td><td>17</td><td>50</td><td>33</td></tr>
<tr><td colspan="6">Rule: 4.73 <= SCAVG <= 5.60 and 14.0 <= SCN <= 14.5</td></tr>
<tr><td rowspan="2">s3</td><td>5</td><td>80.00</td><td>4</td><td>60</td><td>20</td><td>20</td></tr>
<tr><td colspan="6">Rule: 6.91 <= SCAVG <= 7.34 and 12.5 <= SCN <= 15.0</td></tr>
<tr><td rowspan="2">s4</td><td>15</td><td>86.67</td><td>13</td><td>40</td><td>33</td><td>27</td></tr>
<tr><td colspan="6">Rule: 5.6 <= SCAVG <= 6.91 and 12.5 <= SCN <= 14.5</td></tr>
</table>

The highest scores (*s3*) are mostly present in the *Front*, while the lower scores (*s2*) are featured more in the *Mid* and less in the *Back*. It is interesting to note that the number of locations with the low *SCN* increases when going from the front to the back (*s1*). Cluster with the moderate *SCAVG* and moderate *SCN* (*s4*) is the largest, as well as the most balanced among the zones. This may suggest that the test scores are dependent on other factors, although the link to location LTA distance is observable in smaller clusters.

Additionally, we utilised the *Anomaly Detection* technique (Oracle, 2008) which employs the single-class Support Vector Machine (SVM) method (Boser, 1992) to detect atypical locations with respect to their *SCAVG* and *SCN* values. We found 4 anomalous locations: *216*, *219*, *223*, and *231*. Location *219* was anomalous since it has a high score with low test count, while the other seats got labelled because of the associated low scores. One common trait is that three of them (*219*, *223*, and *231*) are border locations (only one direct neighbouring location). This spatial information is not an inherent part of *SCAVG*, nor *SCN*, but it emerges from the findings. The farthest locations (*219* and *231*) also stand out because of their low *SCN*, possibly owing to lower visibility and audibility.

If we arrange the locations into 8 equal-size bins by their distance from LTA, we can plot the dependency between the average test scores expressed in points and the location LTA distance in centimetres (see Fig. 2) which shows that the highest test scores are linked to front locations, although the farthest seats also have above average values.

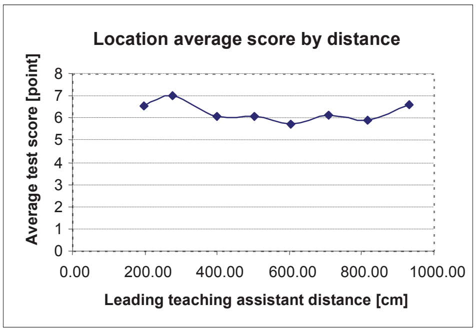

Fig. 2. Location average test score by leading teaching assistant distance for all three tests

4.2 Analysing Student Migration within the Laboratory

For each of the 131 students who had done all three tests in the course, we calculated the average test score and the number of different seats occupied during the tests. Next, we divided students into 3 groups according to the associated number of different seats. Group size, as well as average test score, is shown in Table 3. The findings demonstrate a one level higher grade of the group of students who never changed the seat when compared to those who changed the seat each test. Furthermore, as the number of associated seats increases, both the number of students and average score decrease. This shows that nearly half of the students (47%) stay at the same location, while the rise in the frequency of seat change corresponds to the drop in test scores.

Table III. Student average scores and count by the number of chosen locations

Number of chosen locations	Student count	Average test score
1	61	6.72
2	46	6.06
3	24	5.63

In addition to that, we calculated location positions in the laboratory, as well as the distance (d) that student covered during n analysed tests (n=3):

$$d = \sum_{i=1}^{n} \sqrt{\left(\frac{\sum_{j=1}^{n} x_j}{n} - x_i\right)^2 + \left(\frac{\sum_{j=1}^{n} y_j}{n} - y_i\right)^2}$$

where: x_i , y_i – x and y coordinates of the location where student sat during the test i. Next, using the enhanced k-Means algorithm, we clustered the students according to the respective distance covered and number of different locations associated with the student, thus obtaining *distance clusters*: *d1* (1 location, no distance covered), *d2* (3 locations, any distance), *d3* (2 locations, longer distance), and *d4* (2 locations, shorter distance). Moreover, we clustered the same students according to their summed test scores and formed 4 *student test score clusters* (given from lowest to highest scores): *gLow*, *gMid*, *gMid2*, and *gHigh*. Finally, for each distance cluster we determined how many of its members belong to each of the four test score clusters (see Table 4).

Obtained results support previous findings. Many students who did not change the location a single time (*d1*) are present in the two groups with highest test scores (*gMid2* and *gHigh*). Those who changed their location for each test (*d2*) are best represented by the groups with lower scores (*gLow* and *gMid1*). Students who changed the location only once have more pronounced medium scores (*d3*) or scores more balanced among the whole score range (*d4*). The last finding could represent all those students whose typical location was occupied by someone else, so they had to select another seat close to the original one. As this could happen to any student, the differences between the percentages in this cluster are the smallest - all score ranges are more or less equally represented. Moreover, clusters *d1* and *d4* represent students who always keep the same seat (*d1*) or do not move far from their preferred seat (*d4*). Since they make 71% of the population, the student location switch inertia is further confirmed.

Table IV. Covered distance clusters by student test score clusters

Distance cluster	**Size**	***gLow*** **[%]**	***gMid1*** **[%]**	***gMid2*** **[%]**	***gHigh*** **[%]**
d1	61	21.31	18.03	24.59	36.07
d2	24	33.33	33.33	20.83	12.5
d3	14	14.29	28.57	35.71	21.43
d4	32	28.12	25	18.75	28.12

5. CONCLUSION

In the paper, we discuss the application of DM techniques for the investigation of student seating arrangements in the classroom. The proposed approach uncovered practical results which show that with the increased location distance from the instructor, scores tend to drop to a stable level and that border locations are outliers in terms of associated test scores and occupancy. Furthermore, the analyses of classroom migration demonstrate that students who do not change the seating location have, on average, a one grade level higher score than the others. Also, many students gravitate towards such behaviour.

Information acquired through these means could be used for the development of an algorithm for the identification of optimal seating arrangements which would diminish negative effects of a bad seating choice on student learning. Further work could also include the examination of factors influencing student seating choices, as well as. the integration of the proposed analyses with the existing student testing system.

ACKNOWLEDGEMENTS

The research was supported by Ministry of Science and Technological Development of Republic of Serbia, Grant III-44010.

REFERENCES

BOSER, B. E., GUYON, I. M., AND VAPNIK, V.N. 1992. A training algorithm for optimal margin classifiers, In *Proceedings of the 5th Annual ACM Workshop on COLT*, 144-152

ÇINAR, İ. 2010. Classroom geography: Who sit where in the traditional classrooms?, *The Journal of International Social Research*, 3(10), 200-212

GOSSARD, M. H., JESSUP, E., AND CASAVANT, K. 2006. Anatomy of a classroom: An exploratory analysis of elements influencing academic performance, *NACTA Journal*, 6, 36-39

KALINOWSKI, S., AND TAPER, M. L. 2007. The effect of seat location on exam grades and student perceptions in an introductory biology class, *Journal of College Science Teaching*, 1, 54-57

MONTELLO, D. R. 1988. Classroom seating location and its effect on course achievement, participation, and attitudes, *Journal of Environmental Psychology*, 8, 149-157

ORACLE 2008. *Oracle Data Mining Concepts, 11g Release 1 (11.1)*, B28129-04, Oracle

PERKINS, K. K. AND WIEMAN C. E. 2005. The surprising impact of seat location on student performance, *The Physics Teacher*, 43, 30-33

RAKIĆ, P., STRIČEVIĆ, L., ŽIVANOV, Ž., SUVAJDŽIN, Z., AND HAJDUKOVIĆ, M. 2007. Computer classroom: Deployment and exploitation, *Info M*, 6(21), 9-13.

ŽIVANOV, Ž., RAKIĆ, P., STRIČEVIĆ, L., PUŠIĆ, B., SUVAJDŽIN, Z., AND HAJDUKOVIĆ, M. 2008. Computer aided student examination, *Info M*, 7(25), 45-53.