Grouping Students for Cooperative and Collaborative Learning: challenges and trends in virtual learning environments

Flavius L. Gorgônio
Federal University of Rio Grande do Norte (UFRN)
Caicó, Brazil
flgorgonio@gmail.com

Yuri K. N. Silva
Federal University of Rio Grande do Norte (UFRN)
Caicó, Brazil
yuriikeelvin@gmail.com

Karliane M. O. Vale
Federal University of Rio Grande do Norte (UFRN)
Natal,, Brazil
karlianev@gmail.com

Huliane M. Silva
Federal University of Rio Grande do Norte (UFRN)
Natal, Brazil
hulianeufrn@gmail.com

Abstract— **Educational Data Mining is concerned with developing new methods and techniques for discovering knowledge from educational databases in order to contribute to better understanding academic environment, and can be used to support decision making. Clustering methods has been used in educational data mining to segment classes into homogeneous and/or heterogeneous groups of students. This approach improves cooperative and collaborative interactions among students and it contributes significantly to the mutual learning process. This is particularly important in a virtual learning environment, where there is no physical interaction between students, and mutual learning is through virtual interactions, chats, cooperative and collaborative activities and exercises developed in groups. This paper reviews methods and techniques used to grouping students in productive and collaborative teams and investigates challenges and trends to use these approaches in virtual learning environments.**

Keywords— cooperative and collaborative learning; virtual learning environments; clustering methods.

I. Introduction

Data mining can be defined as the process of knowledge discovery – in the form of useful and interesting patterns – from large amounts of data, where data sources can include internal databases, data warehouses, data repository, social networks and internet [1]. Although data mining is only one step in the overall knowledge discovery process, some authors often use it as a synonym for the entire process. Other steps include data pre-processing (data selection, data cleaning, data integration and data transformation) and post-processing (interpretation and evaluation of the obtained results) tasks.

Over the last few years, data mining methods and techniques have been used in a wide variety of areas, including education. Educational data mining addresses the use of data mining algorithms in educational databases in order to discover useful knowledge that contributes to a better understanding of the academic environment and can be used to support decision making.

A Virtual Learning Environment (VLE) can be described as an integrated academic environment in which students can apply for admission, enroll in the classes offered by VLE, access a course, take tests, and interact with the teachers as well as classmates, in a learning environment mediated by computers and digital technologies over the internet [6][7]. In addition, VLEs include a set of tools for administrators, teachers and students performing their duties and implements specific modules for routine administrative functions of an academic environment.

VLE incorporates text, audio, video, graphics, and animation in a multimedia computer environment to offer a powerful education ambiance centered in interactivity. Because of all of these features offered, VLE typically accumulates a large amount of data, because they can store all the information about students and their interaction with the academic environment in log files and data sets [8].

Information stored in VLE databases can be used to modify the environment and the students' learning process. While linear (or traditional) VLEs use a simple and predefined sequence of learning materials that are equal to every single user, adaptable VLEs enable the user to actively and adaptively specify and modify the learning process. In order to do so, VLEs use automated models to identify users' context and to be adjusted to the individual requirements of each user [9].

Cooperative and collaborative learning plays an important role in the learning process, since students collectively construct their knowledge through a constant exchange of information, points of view, questioning, resolution of questions and mutual evaluation. Cooperative and collaborative learning goes beyond working together, it requires team work with well-defined roles to ensure the success of the group [10].

Despite the advantages of cooperative and collaborative learning, the strategy used to formation of groups is critical, since the heterogeneous groups are pointed out by some researchers as more effective in the learning process. From the pedagogical point of view, the distribution of students in heterogeneous groups increases the knowledge sharing between peers and, consequently, improves mutual learning, since individuals can share different types of knowledge [11].

This paper reviews methods and techniques used to grouping students in productive and collaborative teams and investigates challenges and trends to use such approaches in virtual learning environments. The remainder of this paper is organized as follows: Section II presents a brief review about data mining, cluster analysis, and cooperative and collaborative learning concepts. Section III describes the main differences between homogeneous and heterogeneous grouping strategies, which is followed, in Section IV, by a state-of-the-art review about how to proceed to grouping students into homogeneous or heterogeneous groups. Finally, Section V provides a brief discussion, some final conclusions and directions for future works in this area.

II. Theoretical Review

A. Data Mining Tasks

Data mining tasks are usually classified into two categories: descriptive and predictive mining tasks [2][3]. Descriptive mining tasks characterize properties of the data in a target data set, in order to find patterns that describe correlations, trends, clusters, trajectories and anomalies in the analyzed data. Predictive mining tasks perform induction on the current data in order to make predictions, such as to predict the value of a particular attribute based on the values of other attributes.

Predictive models include classification and regression analysis tasks. Classification is the process of finding a model or function that describes and distinguishes data classes. The model is used to predict the class label or object for which the class label is unknown [2]. Regression analysis differs from classification task because output information is a continuous numerical value or a vector of such values rather than a discrete class [4]. Regression analysis are usually taken using statistical and mathematical models.

Descriptive models include the mining of frequent patterns, associations, and correlations; cluster analysis; and outlier analysis. The task of mining frequent patterns is commonly used to identify the occurrence of patterns that are frequent in a database and that can describe the behavior of the objects involved by identifying associations or correlations between attributes. On the other hand, the task of cluster analysis is related to the identification of objects with similar characteristics, whereas the analysis of outliers (or anomalies) seeks to identify objects whose behavior deviates significantly from the others, to the point of suggesting special attention.

B. Cluster Analysis

Cluster analysis, or simply clustering, is the process of partitioning a set of objects into subsets, named clusters, such that objects are similar to one another in the same cluster, yet dissimilar to objects in other clusters [2]. To group objects into clusters, clustering algorithms compare their attributes using similarity or dissimilarity measures. Clustering can be regarded as a data mining task associated with data description activities, having a wide range of applications.

The task of clustering is particularly important because it is often the first step in data mining analysis. By using this task, for example, it is possible to identify groups of related records that can be used as a starting point for exploring future relationships [5]. Additional analysis using standard analytical and other data mining techniques can determine the characteristics of these groups with respect to some desired outcome. In the context of this work, cluster analysis is used as a strategy to identify students with similar characteristics, based on their attributes.

Formally, cluster analysis can be described as the follow: given X, a set of *n* input patterns:

$$\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n\}$$

where each x_i represents a *p*-dimensional vector, denoted by:

$$\mathrm{x}_i = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)^T \in \mathbb{R}^p$$

and each measure x_{ij} represents an attribute of the dataset, a clustering process attempts to seek a *k*-partition of X, denoted by:

$$\mathrm{C} = \{\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_k\}, \text{ where } k \leq n$$

such that the following priorities are met:

i. $\mathrm{C}_i \neq \emptyset;\ i = 1, \ldots, k$

ii. $\bigcup_{i=1}^{k} \mathrm{C}_i = \mathrm{X}$

iii. $\mathrm{C}_i \cap \mathrm{C}_j = \emptyset;\ i, j = 1, \ldots, k;\ i \neq j$

The task of identifying similar objects in a dataset requires the adoption of a distance metric between the objects that can determine the proximity between them. There are two types of distance metrics: similarity shows the similitude between items, i.e., the greater the similarity, the more alike (or near) the items are. Dissimilarity measures the difference between items, the greater the dissimilarity, the more different (or far) they are [11].

C. Cooperative and Collaborative Learning

The terms 'cooperative learning' and 'collaborative learning' are used interchangeably in some cases, but they are not the same [15]. In collaborative learning, the emphasis is on the process of working together, while the achievement of the product is central to cooperative learning [10]. However, in order to avoid misunderstandings, most authors consider that in most cases, the terms cooperative and collaborative learning are used interchangeably.

Cooperative and collaborative learning has been carried out since the middle of the sixteenth century, through homogeneous groups, with students of the same sex and age, grouped in the same classroom and conducted by one or more teachers. However, other forms of organization of students into groups were also accepted in cases of population with a reduced number of students, such as several smaller heterogeneous groups within the large homogeneous group, even though these practices were viewed negatively [14].

Recently, the growth and popularization of distance learning has made possible several innovative ways to encourage collaborative and cooperative learning in virtual learning environments. Workgroups that use computer-mediated cooperation and collaboration result, in most cases, in improvements in performance, in interaction among members, and in the development of critical thinking [16].

III. Homogeneous and Heterogeneous Grouping

The problem of optimizing the grouping of students in homogeneous and/or heterogeneous teams has been widely studied in literature. Both cooperative and collaborative learning can be influenced by the way groups are organized, whether in the form of homogeneous or heterogeneous groups. If, on the one hand, homogeneous groups tend to maintain concentration and focus, reducing external interference, on the other hand, heterogeneous groups promote the dissemination of new ideas and the sharing of new knowledge [17][18].

Heterogeneous groups obtain better results in terms of background knowledge, sharing of ideas, and in the exercise of social inclusion, ethnicity and gender policies. However, homogeneous groups are better at achieving specific goals, while heterogeneous groups tend to be more creative and more innovation oriented [24].

Despite the benefits related to the practice of workgroups, whether in homogeneous or heterogeneous groups, the difficulty in distributing the students of a class in groups in the most homogeneous/heterogeneous way possible is still a challenge for most educators, who have difficulties to make use of these strategies [11]. Therefore, it is necessary to evaluate techniques and methods that assist the educators in the distribution of students in groups.

IV. Some Strategies for grouping students

The formation of groups that are too homogeneous and non-standardized, due to an inadequate distribution of students in groups, as well as the complexity of distribution of students in ideal groups, make it difficult to share knowledge among students. This statement may explain the fact that there are many more articles addressing the use of heterogeneous than homogeneous groups.

Good and Marshall criticize the grouping of students by ability and state that teachers who group students by ability need to carefully evaluate their reasons for grouping [25]. They conclude by stating that higher quality and more thoughtful teaching of mixed groups of learners can lead to better outcomes than it can be done by fragmentary teaching of a given number of different groups.

Glass also advocates that adoption of mixed or heterogeneous ability or achievement groups offer several advantages, among which: i) less able students are reduced risk of being stigmatized; ii) expectations of the teachers for all students are maintained at higher levels; and iii) opportunities for more able students to assist less able peers in learning can be realized [26].

Silva, Silva and Gorgônio present a strategy capable of forming heterogeneous teams by using traditional clustering algorithms, such as k-means and self-organizing maps [11]. This strategy is divided in two stages: in stage 1, the clustering algorithm is applied on the selected attributes, organizing individuals in accordance with the similarity which they have to each other. In stage 2, a distribution algorithm is applied, which allocates similar individuals into distinct groups, favoring the formation of heterogeneous groups. Subsequently, this work is continued by the author in her Bachelors thesis [21].

Vellido, Castro and Nebot carry out an extensive review on the use of several cluster analysis algorithms (k-means, fuzzy c-means, self-organizing maps and generative topographic mapping) applied in the grouping of data in VLEs [12]. Authors divide dozens of articles reviewed into three major categories: i) works devoted to group e-learning material based on similarities; ii) works that attempt to group students according to their navigational and/or behavior in the VLE; and iii) works that proposes clustering analysis as part of an e-learning strategy but do not present any practical application results.

Kay, Koprinska and Yacef used three data mining approaches (mirroring visualization, sequential pattern mining, and clustering) to analyze student workgroup interaction data in the context of a software development project course [13]. In this work, authors analyzed the activity and interaction among the members, in addition to using clustering algorithms to identify members with similar characteristics, identifying the individuals who had leadership and initiative profiles. These results enable project managers to track the progress of the projects and to identify previously possible problems and difficulties in their execution.

Graf and Bekele explore the formation of heterogeneous groups for intelligent collaborative learning systems. In this paper, the authors present a mathematical approach for forming heterogeneous groups of students based on their personality traits and performance. In sequence, they implement the proposed mathematical approach by using the ant colony optimization algorithm [19]. Results shown that the algorithm finds stable solutions close to the optimum for different datasets.

Lin, Huang and Cheng propose a strategy, named automatic group composition system for composing collaborative learning groups, using a particle swarm optimization based algorithm [20]. In this paper, the authors seek to balance the level of individual knowledge of each student and their interests, grouping criteria that are often conflicting, but are commonly considered by the instructors when forming work groups. To address this problem, this study formulates a group composition problem to model the formation of collaborative learning groups that satisfy the two grouping criteria. A novel approach based on particle swarm optimization is proposed for composing well-structured collaborative learning groups with the objective of maximizing the learning performance of all students participating in the collaborative learning environment.

Moreno, Ovalle and Vicari proposed a different approach for group formation in collaborative learning through the use of genetic algorithms [22]. According to the authors, the main feature of such a method is that it allows for the consideration of as many student characteristics as may be desired, translating the grouping problem into a multiobjective optimization one. The main goal of this approach was to obtain inter homogeneous groups, which are as similar as possible to the general characteristics of the total sample of students, but also considering the heterogeneity inside each one. Results obtained were compared with three others methods: an exhaustive method that enumerates all possible groups guaranteeing that the optimal fitness value is found; a random method, where students were randomly attributes to groups; and a self-organizing proceeding, where students could choose their own groups.

Alhunitah and Menai also investigated the same theme, but using a different strategy. In their article named "Solving the Student Grouping Problem in e-Learning Systems Using Swarm Intelligence Metaheuristics" [23], the authors propose the use of a metaheuristic called particle swarm optimization (PSO) to solve the problem, comparing it with two other metaheuristics: ant colony system and artificial bee colony. The results presented show that the PSO algorithm obtained better groups with more heterogeneity.

However, although most educators defend the use of heterogeneous classes as a way to facilitate learning, there are still several contexts in which the use of homogeneous classes is advocated. Because there is no consensus among educators that heterogeneity is the only correct alternative, there are some studies that investigate the formation of homogeneous groups.

Krueger and Casey argue the use of homogeneous groups for applied research [27]. Hooper and Hannafin evaluated the effects of heterogeneous versus homogeneous grouping on the learning of progressively complex concepts and results indicated that no significant differences in achievement were found between the two grouping methods [28].

Shields conducted a comparative study of student attitudes and perceptions in homogeneous and heterogeneous classrooms, and concluded in her study that students in heterogeneous classes reported lower teacher expectations, less academic learning time, less homework, and less teacher feedback [29]. In addition, she found that in homogeneous groups, teachers can individualize the pace, process, and products required of students and she further reported that teachers of homogeneous classes for gifted students tended to require students to engage in longer term, research style assignments, rather than frequent, lower level cognitive assignments generally given to a regular class [30].

Bikarian conducted a large study about the effects of heterogeneous or homogeneous grouping on reading achievement. The author describes advantages and disadvantages of grouping students both homogeneously and heterogeneously, demonstrating that it is neither true than one is better than the other [30]. The paper also presents some factors to be considered regarding the grouping of students, including group size, gender of group, composition, the task or activity to be completed, interaction, and if adult presence will be needed for the given assignment.

Kuo, Chu and Huang conducted an experiment on a university English course to compare the learning performance and learning interest of the students assigned to the collaborative learning groups based on the learning style-based homogeneous and heterogeneous grouping strategies [31]. This study found that although the learning performance of the homogeneous groups was not significantly different from that of the heterogeneous groups, the mean of the post-test in the homogeneous groups outperformed that of the heterogeneous groups.

Finally, Ary, Jacobs, Sorensen and Walker, in their book "Introduction to Research in Education", describe some strategies for homogeneous selection of students in scientific experiments, comparing this approach with random assignment and random matching [18].

V. Discussion and Conclusions

Currently, cooperative and collaborative learning plays an important role in the learning process and teachers in classrooms are finding ways to engage students in these activities. When confronted with the need to group students to perform activities in a cooperative or collaborative way, teachers undoubtedly need to make choices about how the class should be divided and these choices can directly influence the learning process, making this activity easier or more difficult to perform.

Scientific research shows that both cooperative and collaborative learning can be influenced by the way groups are organized, whether in the form of homogeneous or heterogeneous groups. Methods traditionally adopted for group allocation, such as random assignment or self-organizing proceeding, where students can choose their group by affinity with other students, are inefficient since it produces groups with very different profiles.

Most of the time, it is responsibility of the teacher to assign students to their respective groups. Even if the teacher wants to make this choice manually, the complexity of finding the ideal combination that maximizes the criteria of homogeneity or heterogeneity of all groups makes this task impossible in an acceptable time.

This article presented several strategies to group students, either homogeneously or heterogeneously, discussing advantages, disadvantages or limitations of each of the approaches presented. Several of the reviewed articles present computational strategies to optimize the process of assigning students to their respective groups, seeking to facilitate the grouping task.

Considering that most of the available articles in literature give preference to grouping in a heterogeneous way, since this approach favors collaborative learning through the sharing of knowledge by group members, a greater number of articles with this approach was presented. However, in many cases, it is possible to choose the homogeneous approach, simply by using the proposed procedure in reverse.

As future work, it is suggested the development of computational tools that use strategies evaluated with two main objectives: i) to facilitate the work of teachers, helping them in the assignment task of students in homogeneous or heterogeneous groups; and ii) provide mechanisms to compare such strategies in order to validate their use in other contexts not yet explored.

REFERENCES

[1] D. Larose and C. Larose. Discovering Knowledge in Data: An introduction to data mining, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2014.

[2] J. Han, M. Kamber, and J. Pei. Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA: Morgan Kaufmann Publishers, 2011.

[3] P.-N. Tan, M. Steinbach, and V. Kumar. Introduction to Data Mining. Harlow, Essex: Pearson Education, 2014.

[4] I. Witten, E. Frank, M. Hall, and C. Pal. Data Mining: Practical machine learning tools and techniques. 4th ed. Cambridge, MA: Morgan Kaufmann Publishers, 2016.

[5] S. Sumathi and S. Sivanandam. Introduction to Data Mining and its Applications. Berlin: Springer, 2006.

[6] A. Kumar, R. Pakala, R. Ragade, and J. Wong, "The Virtual Learning Environment System," Frontiers in Education Conference, FIE '98. 28th Annual, Tempe, AZ, USA, vol 2, pp. 711-716, 1998.

[7] J. Weiss. "Introduction: Virtual Learning and Learning Virtually," In: The International Handbook of Virtual Learning Environments, vol I, J. Weiss, J. Nolan, J. Hunsinger, and P. Trifonas, Eds. Dordrecht, The Netherlands: Springer, 2006, pp. 1-33.

[8] A. Zafra, C. Romero, and S. Ventura, "Multi-Instance Learning versus Single-Instance Learning for Predicting the Student's Performance," In: Handbook of Educational Data Mining, C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker, Eds. Boca Raton, FL: CRC Press, 2011, pp. 187-200.

[9] D. Müller. Design Characteristics of Virtual Learning Environments: A theoretical integration and empirical test of technology acceptance and is success research. Saarbrücken, Germany: Springer, 2013.

[10] R. Hernández, "Collaborative Learning: Increasing Students' Engagement Outside the Classroom," US-China Education Review A, vol.2, n.9, 2012, pp. 804-812.

[11] H. Silva, C. Silva, and F. Gorgônio, "A Self-Organizing Map Based Strategy for Heterogeneous Teaming," In: Applications of Self-Organizing Maps. M. Johnsson, Org. Rijeka: InTech, 2012, pp. 89-117.

[12] A.Vellido, F. Castro, and A. Nebot, "Clustering Educational Data," In: Handbook of Educational Data Mining, C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker, Eds. Boca Raton, FL: CRC Press, 2011, pp. 75-92.

[13] J. Kay, I. Koprinska, and K. Yacef, "Educational Data Mining to Support Groups Work in Software Development Projects," In: Handbook of Educational Data Mining, C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker, Eds. Boca Raton, FL: CRC Press, 2011, pp. 173-185.

[14] A. Zabala. A Prática Educativa: Como ensinar. Porto Alegre, RS: Artmed, 1998.

[15] A. Alipour, "Cooperative versus Collaborative Learning," International Journal of Research in Linguistics, Language Teaching and Testing, vol. 1, Issue 4, 2016, pp. 153-158.

[16] L. Kupczynski, M. A. Mundy, J. Goswami, and V. Meling, "Cooperative Learning in Distance Learning: A mixed methods study," International Journal of Instruction, vol. 5, n. 2, 2012, pp. 81-90.

[17] F. Lu, and M. Anderson, "Peer Effects in Microenvironments: The benefits of homogeneous classroom groups," Journal of Labor Economics, vol. 33, n. 1, 2015, pp. 91–122.

[18] D. Ary, L. Jacobs, C. Sorensen, and D. Walker. Introduction to Research in Education. Belmont, CA: Cengage Learning, 2013.

[19] S. Graf, and R. Bekele, "Forming Heterogeneous Groups for Intelligent Collaborative Learning Systems with Ant Colony Optimization," In: Intelligent Tutoring Systems. Berlin, Heidelberg: Springer, 2006.

[20] Y.-T. Lin, Y.-M. Huang, and S.-C. Cheng, "An Automatic Group Composition System for Composing Collaborative Learning Groups Using Enhanced Particle Swarm Optimization," Computers & Education, vol. 55, n. 4, 2010, pp. 1483-1493.

[21] H. Silva, "Uma Estratégia Baseada em Análise de Agrupamento para Formação de Equipes Heterogêneas com Foco na Troca de Conhecimentos e no Aprendizado Mútuo," Bachelor Thesis, Federal University of Rio Grande do Norte, Brazil, 2012.

[22] J. Moreno, D. Ovalle, and R. Vicari, "A Genetic Algorithm Approach for Group Formation in Collaborative Learning Considering Multiple Student Characteristics," Computers & Education, vol. 58, n. 1, 2012, pp. 560-569.

[23] H. Alhunitah and M. Menai, "Solving the Student Grouping Problem in e-Learning Systems Using Swarm Intelligence Metaheuristics." Computer Applications in Engineering Education, vol. 24, n. 6, 2016, pp. 831-842.

[24] M.-I. Dascalu, C.-N. Bodea, M. Lytras, P. Pablos, and A. Burlacu, "Improving e-Learning Communities Through Optimal Composition of Multidisciplinary Learning Groups," Computers in Human Behavior, Vol. 30, Jan. 2014, pp 362-371.

[25] T. Good and S. Marshall, "Do students learn more in heterogeneous or homogeneous groups?," In: The Social Context of Instruction: Group Organization and Group Processes, P. Peterson, L. Wilkinson, and M. Halliman, Eds. New York: Academic Press, 1984, pp. 15-38.

[26] G. Glass, "Grouping students for instruction," In: School Reform Proposals: The Research Evidence. A. Molnar, Ed. Greenwich, CT: Information Age Publishing, 2002, pp. 95-112.

[27] R. Krueger and M. A. Casey. Focus Groups: A practical guide for applied research, 5th ed., Singapore: Sage publications, 2014.

[28] S. Hooper, M. Hannafin, "Cooperative CBI: The Effects of Heterogeneous versus Homogeneous Grouping on the Learning of Progressively Complex Concepts," Journal of Educational Computing Research, vol. 4, n. 4, 1988, pp. 413-424.

[29] C. Shields, "A comparison study of student attitudes and perceptions in homogeneous and heterogeneous classrooms," Roeper Review, vol. 24, n. 3, 2002, pp. 115-119.

[30] S. Bikarian, "The Effects of Heterogeneous or Homogeneous Grouping on Reading Achievement", Master Thesis, Sierra Nevada College, Nevada, USA, 2009.

[31] Y.-C. Kuo, H.-C. Chu, and C.-H. Huang, "A Learning Style-based Grouping Collaborative Learning Approach to Improve EFL Students' Performance in English Courses," Educational Technology & Society, vol. 18, n. 2, 2015, pp. 284-298.