Chapter 5

A Self – Organizing Map Based Strategy for Heterogeneous Teaming

Huliane M. Silva, Cícero A. Silva and
Flavius L. Gorgônio

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/52776

1. Introduction

Even though education and knowledge are processes inherent to human development and are present in all cultures since the earliest and most remote age, the educational model adopted today had its origins in Ancient Greece, where the first signs of appreciation of cultural knowledge took place [1]. At the time, the most traditional way to prepare young individuals of any social class to social integration was through individual processes of teaching, whether in daily life, with their parents, or in contact with masters and artisans. Although, the more privileged classes enjoyed other types of learning, such as access to reading, writing and other areas of knowledge, this process was always conducted on an individual basis.

The need to generalize the teaching of reading, writing and the so-called general culture among the less privileged social strata caused an increase in the number of students in relation to the number of teachers available. This fact prompted educators that time to seek a teaching model that could bring knowledge from the educators to a maximum number of individuals at the same time. Given this need, the Greeks developed the earliest forms of grouping students in order to maximize their teaching activities [2].

The school and the way students are organized in the classroom have also undergone various transformations throughout history [2]. Initially, they were organized in large groups in a single classroom, and guided by a teacher or tutor who had different concepts that he judged to be of common interest, combined with specific content, targeted to smaller groups or individual students. Later, new forms of organization of schools and classrooms emerged, such as the structuring of the content presented according to age, the division of

students into fixed and/or mobile groups along with learning from interaction with other individuals through the formation of study groups.

Throughout the evolution of the teaching-learning process, interaction between members of a group in order to encourage mutual learning, has become increasingly valued. Currently, learning from the development of team activities is very encouraging, since it facilitates the sharing of experiences and ideas among group members and allows the realization of some activities that are not possible to be carried out individually. From the socio-educational viewpoint, it is considered a means to promote socialization and cooperation among different levels, in order to solve problems of group dynamics and facilitate learning among peers.

From the pedagogical point of view, the distribution of students in heterogeneous teams allows the exchange of knowledge among peers and, consequently, enhances mutual learning, given that individuals can share different kinds of knowledge. However, the procedures commonly adopted by teachers and educators in the process of forming academic teams usually do not favor such heterogeneity, since in most cases, students choose their own teams considering their affinities and common interests. In other instances, it is the teacher who leads the process of teaming through some selection criterion, which can range from random choice (through a "draft") to appointing some students to be team leaders, trying to better distribute students within teams, and thus make them more heterogeneous.

In this context, a problem arises: how to partition a set of n students into k teams, maximizing the heterogeneity among the members of each team, to allow students share their individual knowledge with each other? This chapter presents a strategy to partition a class into several teams that enables the formation of heterogeneous teams, with the goal of enabling knowledge sharing and mutual learning on the part of the team members. The proposed strategy is based on using of well-known clustering algorithms, such as self-organizing maps and K-means algorithm, and using the Davies-Bouldin cluster validity index to measure the results.

The remainder of the chapter is presented in the following way. Section 2 presents a literature review on the process of forming heterogeneous teams and its difficulties in educational settings. This second section was divided into three parts: the first argues about the importance of team work, the second discusses about the complexity of teaming process, while the third presents several works related to the area of educational data mining. Section 3 introduces the clustering process and its stages, according to three different approaches and presents the clustering algorithms to be used in the proposed strategy: self-organizing maps and k-means, and a brief discussion of similarity measures and clustering validation indices, with the presentation of the Davies-Bouldin index. Section 4 presents and discusses the strategy proposed in this paper. Section 5 presents the methodology used in the experiments, a brief discussion of the databases used and discusses the results obtained. Finally, section 6 presents the conclusions and proposals for future works.

http://dx.doi.org/10.5772/52776

2. Theoretical basis

2.1. The importance of team work

Society is changing faster and faster; along with these changes new ways of social living, interacting with people, teaching and learning are emerging. The way the teaching-learning process is currently conducted in the classroom is not what it used to be some years ago, since teaching and learning have been undergoing several modifications. The use of new information and communication technologies have contributed significantly to these changes, offering new ways of learning and abandoning the old ways of studying, which is becoming, – if not less important – at least as one more among many other alternatives available.

Traditional pedagogy was based on transmission of general culture, i.e., the great discoveries of mankind, as well as aiming on the formation of reasoning and training the mind and will [3]. The methods and practices adopted in this approach overburden the student with merely memorized knowledge, without seeking to establish relations with the everyday life and without encouraging the formation of critical thinking and intellectual capacities.

According to traditional pedagogy, the teaching activity is focused on the teacher who explains and interprets the matter for the students to understand. Besides, this approach assumes that students, by listening and doing repetitive exercises, end up memorizing the subject in order to later reproduce it, whether it is through by questions from the teacher or via tests [3]. This old paradigm, currently seen as outdated, was based on the knowledge transmission performed by teachers, on memorization and competitive and individualistic learning by the students, more and more out of favor [4].

Nowadays, it is defended that the practice of teaching centered on the teacher is not the best approach; and that the methodologies and practices widely adopted by teachers – which are based on repetition and rote memorization – undermined the objectives of traditional pedagogy. That is why, over time, difficulties found in the teaching process became more evident and this methodology, gradually, was modified.

In opposition to this approach, this new pedagogy changes the focus towards students instead of teachers. Therefore, students now become the center of school activity and are placed under favorable conditions in order to learn by themselves from their own needs. In this conception what becomes crucial is the issue of learning. The challenge posed to teachers is changing teaching axis towards paths that lead to learning, and also make it essential for teachers and students to be in a constant process of continual learning [4].

Methodologies such as collaborative learning and cooperative learning have often been advocated in the academic field, since, it recognizes these methodologies to have potential to promote a more active learning, by encouraging critical thinking, development of capabilities in interaction, information exchange and problem solving, as well as development of the self-regulation of the teaching-learning process [5].

Using these methods of learning is not something new. For years, educators have been using these practices of collaborative and cooperative learning, along with group work, because

they believed in the potential that these methods have to prepare students to face work demands [5]. Despite being longstanding methodologies, the terms collaborative learning and cooperative learning are often confused in literature. Both have similar definitions but they are different in theoretical and practical perspectives.

The terms collaboration and cooperation can be differentiated as it follows: Collaboration as a certain philosophy of interaction and personal lifestyle, where individuals are responsible for their actions, including learning and respect regarding skills and contributions of each member of the group, while cooperation is a structure of interaction designed to facilitate the achievement of a specific product or goal by means of people working together in teams [6]. The same author also discusses the differences of both terms when used in the classroom. In the cooperative model, it is the teacher who retains full control of the class. Although students work in groups, the teacher is responsible for guiding the tasks performed in the room, i.e., it is a teacher focused process. However, the collaborative model is more open; the responsibility of guiding the tasks is on the group itself and not on the teacher – who can be consulted – but the group needs to interact in order to achieve the shared goal.

It is, then, possible to say that both the concepts of cooperation and collaboration are applied to group activities. Although they possess fundamental differences concerning the dynamics of working together, their goals are common and both practices are complementary, representing, therefore, opposition to the teacher-centered education system, on which Traditional Pedagogy is based [5]. The school has considerably developed lately and, therefore, the need for teamwork becomes more and more a matter of métier than a personal choice. This need occurs due to multiple reasons, such as the increasing intervention in school, by educators, psychologists and educational psychologists, division of pedagogic work in primary school, the evolution towards learning cycles among other reasons justifying the need for teamwork [7].

Through group or team work, therefore, it is possible for people to get into contact with different visions of worlds, learn to socialize knowledge, listen to and give opinions on a particular subject, accept suggestions, develop a group mentality and be proactive, among others. The same author also stresses that teamwork, in general, makes the activity more enjoyable and enables the realization of a common activity, with common goals, allowing the enrichment experiments and experiences [8].

It is common to find in literature different terms for the concepts of group and team. A team can be defined as a group gathered around a common project, whose implementation involves various ways of agreement and cooperation [7]. In [9], team is defined as a group of people, who besides working together, cooperate one with another, sharing common goals. Thus, we can say that team is a group of people working together towards a common goal. In this work, in order to keep terminology simple and following the standard adopted by [8], there will be no differentiation made about them, even though the term team is used more often than group.

http://dx.doi.org/10.5772/52776

2.2. The teaming process

Teamwork is a practice commonly adopted in carrying out various tasks, whether simple or complex. In [7], it is argued that teamwork is a matter of skill, and also requires conviction that cooperation is a professional value. Teamwork can be justified based on three main reasons: i) the collaborative work allows assigning tasks to be performed simultaneously by team members, enabling the completion of the activity in a shorter period of time; ii) the existence of activities that have no possibility of being individually performed, because they demand multiple skills; iii) teamwork allows the exchange of experiences among team members, encouraging mutual learning.

The first two reasons justify the use of teamwork in the commercial sector, while the latter justifies its use in academia. In fact, the main reason for the usage of teamwork in academia is to allow the exchange of knowledge among its members, enabling knowledge pre-existing or acquired during the learning process by each individual to be shared by others.

Although in the commercial sector teaming is driven by productivity, where the most efficient teams are those that produce faster results, in academic fields, the goal is mutual learning and knowledge sharing, even though the final result is achieved in a longer period of time. Thus, many researchers argue that the more heterogeneous teams in a learning process are, the more exchange of knowledge among its members is going to happen, fostering mutual learning [2,10,11].

In academia teachers may adopt different ways to teach a class, as well as they may use different techniques to evaluate students in the performance of activities. A common way of evaluation is promoting teamwork, which is used as a means for students to carry out academic activities. In this process the students are grouped into different teams and each team is responsible for executing the task assigned by the teacher.

However, an important question, to which not always due attention is paid by teachers is the method used to determine which students are going to be part of each team. In academia, such process can be influenced by several factors, especially by preference and personal motivation from the students themselves. The most commonly used methods are:

a. Mutual choice: each team is chosen by its own members, usually subject to a minimum and maximum amount of participants that is defined by the educator. The main advantage of this method is usually the affinity between team members. The main disadvantage is that the method tends to form too homogeneous teams, where members have a profile very similar to each other's;

b. Random choice: the teams are chosen at random, usually through some kind of random draw or lottery. Despite allowing a less homogenous distribution than in the previous method, we cannot guarantee that all teams remain equally heterogeneous, given the large amount of possibilities of dividing a class into teams. Another disadvantage of this method is that students often have little or no affinity with each other, which makes the connectedness of the group difficult;

c. Choice guided by the educator: in this case, it is the teacher who defines the teams, seeking to balance these teams and make the more heterogeneous possible, considering the prior knowledge he or she has on the students, while they can meet individual preferences of students to participate in either of the teams. This method has the same problem as the previous one, given an explosion of combinatorial possibilities to choose the teams. Therefore, not all the teams are equally balanced.

One of the problems associated with the development of activities in teams is little engagement and commitment of some members with the performance of activities. This may partly be caused by deficiencies in the teaming process. Teams whose individuals have very similar profiles, i.e. very homogeneous teams, tend to gather students with the same abilities and limitations, so that there will always be activities that none of the individuals in the team have the skills necessary to perform it. Another common problem in this process is that it tends to lead to the formation of a few teams composed of individuals who possess academic performance quite above average and other teams with individuals who have performance below the average, contributing to a certain segregation of students based on their academic performance.

Analyzing the three teaming methods presented above, it is possible to see that none of them directly addresses these problems, i.e. none of them guarantees the heterogeneity of the teams. Taking the method of random choice as an example, where teams are formed from a random selection, and considering a classroom composed of n students, the number of different possibilities of dividing the class into$k$ teams, each consisting of n/k members, is given by the equation(1),derived from the combinatorial analysis:

$$C_{n,k} = \frac{n!}{k!(n-k)!} \tag{1}$$

If it is noticed that teams can have any number of students, flexibility normally allowed by some teachers, the number of possibilities is greater, since it corresponds to the number of possible partitions of a set, being given by the Stirlingnumber of the second kind [12], given by the equation (2),

$$S_n^{(k)} = \sum_{k=1}^{n} \frac{1}{k!} \sum_{i=0}^{k} \left[(-1)^{k-i} C_{k,i} i^n \right] \tag{2}$$

where,n represents the number of students and k represents the number of teams.

For purposes of illustration of how these values can be extremely large, even considering relatively small-sized classes, Table 1 shows the number of different possible ways to divide a class with n students into k teams.

http://dx.doi.org/10.5772/52776

Number ofstudents	Numberof teams	Number of fix-sized teams	Number ofvariable-sized teams
10	2	45	511
12	3	220	86926
12	4	495	611501
25	5	53130	2.4 x 1015
50	5	2118760	7.4 x 1032

Table 1. Quantity of different configuration of students in teams

The values shown in Table 1 demonstrate that, even though teachers use mechanisms to measure the heterogeneity of each team formed, the complexity of finding the ideal combination of students and teams for maximizing the criteria of heterogeneity by performing an exhaustive search in the set of possible solutions makes this task impossible to be performed in a feasible time frame. Thus, a possible alternative to circumvent these difficulties is the use of computational methods in finding approximate (quasi-optimal) solutions which, although not the ways of doing it, represent a viable possibility for solving the problem.

2.3. Educational data mining

The term data mining may be defined as a set of automated techniques for exploration of large data sets in order to discover new patterns and relationships that, due to the volume of data, would not be easily discovered by human beings with bare eye, due to great amount of data. Data mining isdefinedas a process of automatic discovery of useful information in large data warehouses[13,14]. In [15],authors describe it as a process of extracting information that emerged from the intersection of three areas: classical statistics, artificial intelligence and machine learning, which can be used both to identify and describe past events and analyze and predict future trends.

The methods and data mining techniques have been applied to a wide variety of subject areas, such as commercial and industrial sectors, the analysis and understanding of data from research institutions, in medicine and bioinformatics, in text analysis as well as in identification of feelings and opinions on social networks, among others. More recently, researchers in the field of educational computing have been using these techniques in order to investigate problems in computer-mediated learning environments, including the identification of factors that affect learning and developing more effective educational systems. [16-18].

This new area of research, called educational data mining, is primarily focused on developing methods for exploring data sets collected in educational settings [19]. Thus, the area of educational data mining uses computational techniques derived from traditional data mining – classification, regression, density estimation and clustering being some of them – in order to provide mechanisms to optimize the learning process [20].

Literature review shows a growing number of recently published works on this subject, where researchers have sought, in computing, solutions to problems encountered in education. In this context, data mining has been widely used to solve problems with difficult resolution and great importance, not only related to teaming, rather including, also, several other areas of education [14]. One can cite, for example, the development of a methodology for student monitoring based on objective tests on the web [21] and the use of data mining techniques to find association rules and extract patterns about information of students [22], among other works.

In [23], it isshownan agent architecture, integrated to a distance education environment, as a way to solve the problem of formation of collaborative groups, allowing the establishment of the roles that individuals in a group will play in the development of a collaborative activity. To perform the work referred above, the author uses an agent modeled with genetic algorithms, which enables the formation of collaborative study groups in distance learning courses via the web. Finally, the author demonstrates, through the results, that the teams formed from the proposed approach in the work had a superior performance in their activities, compared to the ones that formed teams at random.

Another work in the context of distance education is presented in [24], in which data mining techniques are used in order to identify the profile of students at risk of dropout or failure, and then generate alerts that aware and assist teachers/tutors with monitoring and interacting with these students. Thus, the author proposed an architecture for virtual learning environments – based on information extracted through processes of data mining – in order to identify students with characteristics and behaviors that can be considered as belonging to risk group (dropout and/or failure). The results obtained from the use of the architecture described in the work proved satisfactory, since the warnings contributed positively in the communication and involvement of teachers with students, providing an educational action that improved quality of education in this scenario.

Two works stand out in the literature due to the use of clustering to identify individuals with similar profiles and seek the formation of homogeneous teams, contrary to the purpose described in this chapter. The first aims to identify groups of students with similar profiles in a classroom, in order that the teacher can make use of a differentiated pedagogy adequate to meet groups of students having the same learning difficulty [25]. This method was applied to students in regular classroom teaching and the data were collected from forms filled out by students, in which they identify their degree of certainty in the understanding of every topic addressed by the teacher. The authors cite the use of algorithms K-means and Self-Organizing Maps for these experiments, stating that such algorithms are very useful in the formation of homogeneous teams of students and the identification of groups of similar students in a particular class is an important tool when the teacher wants to apply a differentiated pedagogy on these groups.

Another study which uses educational data mining and also statistical techniques of clustering is presented in [26], which aims to identify and generate homogeneous groups to perform tasks in educational settings. The main objective of the study is to research and implement a clustering tool for distance education platforms, in order to allow the in-

http://dx.doi.org/10.5772/52776

crease of interactions among students with similar profiles in virtual learning environments, allowing better conditions for the learning desired in these environments. According to the author, the methodology adopted for undertaking the work has proved satisfactory, meeting the expected results, since interaction in distance education environments occurred more easily.

In [27], the authors conducted a study focused on improving education, trying to identify a new and smaller set of variables that may influence the quality of teaching and learning the discipline of mathematics, so that mathematics teachers improve activities undertaken in the classroom. In this context, the technique of clustering was useful because, according to the authors, a large amount of information was obtained through the data collected via questionnaires, and this information would be meaningless unless they were classified into groups which one can handle, therefore the advantages of applying a Ward clustering algorithm, in order to group the variables.

This brief literature review revealed some papers belonging to the growing and diverse field of research in educational data mining. The following section describes the task of clustering in the context of data mining, as well as two of the most widely used clustering algorithms that process.

3. Clustering

The task of analyzing and clustering similar objects in a given group, taking into consideration one or more common characteristic(s) existent among them, is an important activity inherent to human behavior, since it, in a general way, permits the organization of objects or everyday activities. People are, daily, faced with the need to group a set of data: either at a supermarket, organizing products complying with the criteria of category or brand; in organizing books in a bookcase, following an order according to subjects, or even the choice of friends in social network, taking into account, for example, the affinity between them – such as belonging to the same classroom at school or even musical taste. Thus, the clustering is often performed intuitively and ends up unnoticed by the user.

3.1. Definitions

Cluster may be defined as a set of cohesive entities, so that internal entities (belonging to the group) are more similar to each other, and more different from external entities (not belonging to the group) [28]. Thus, clustering may be understood as a technique able to divide a data set into one or more sub-sets, taking into account the similarity existing among its elements. However, far from a consensus, this is not the only definition adopted for the term, it is common to find in literature a variety of definitions for this technique, result of studies performed by different researchers in different areas where clustering can be applied [29,31].

Clustering is a statistical technique with general use, applied in different fields of knowledge and widely used in activities involving data analysis. Some of the numerous applica-

tions of clustering in different contexts include their use: in psychology, to identify different types of depression; in biology, to identify groups of genes with similar functions; in medicine, to detect patterns in spatial or temporal distribution of a particular disease; in sales, to identify customer profiles and determine sales strategies, among others [14,29].

Most of its applications is the analysis of large databases on which there is limitedor nonexistent information about its structure and the main goal of its use is precisely to allow in understanding and description of data unknown up to then [12,32]. Thus, clustering can be regarded as a data mining task associated with data description activities, having a wide range of applications. However, it is necessary to be careful in its use, for instance, in analyzing attributes that make up the database and determine in advance the goals desired with the application, to thereby obtain satisfactory results.

3.2. Stages in clustering

The clustering process is usually comprised of several steps, and some authors present these stages more succinctly [28,29], while others have to do it in a more detailed way, divided into more stages [30]. Figure 1 presents the five steps included in the clustering process, as described in [29], which includes the following stages: data preparation, proximity, clustering, validation and interpretation of results, described below:

i. First stage: data preparation involves aspects related to the pre-processing of data, as well as adequate representation for being used by a clustering algorithm;

ii. Second stage: called proximity, it is consisted of the proximity measures proper to the application, as well as the information you want to obtain from data extraction. These measures can be classified as a measure of similarity and dissimilarity;

iii. Third stage: formation of clusters is the central stage of the clustering process. It is at this stage that one or more clustering algorithms are applied on the data in order to identify structures existing in the same cluster;

iv. Fourth stage: the validation consists of assessing the results. In general, it determines if the clusters obtained are significant, i.e., if the solution obtained is representative to the set of analyzed data and the expected solution;

v. Fifth stage: the interpretation refers to the process of examining and labeling each cluster according to its goals, describing its nature. The interpretation goes beyond a simple description, since it still corresponds to a validation process of the clusters found based on the initial hypotheses, as well as other subjective assessments that are of interest to the specialist.

In [28], five steps to the clustering process are also presented, namely: development of the dataset, data preprocessing and standardization, cluster identification, cluster interpretation and, finally, conclusions. These steps, as described below and illustrated in Figure 2, have several similarities with the process described in [29], although some activities described in a particular stage of a process happen in a different stage in another process.

http://dx.doi.org/10.5772/52776

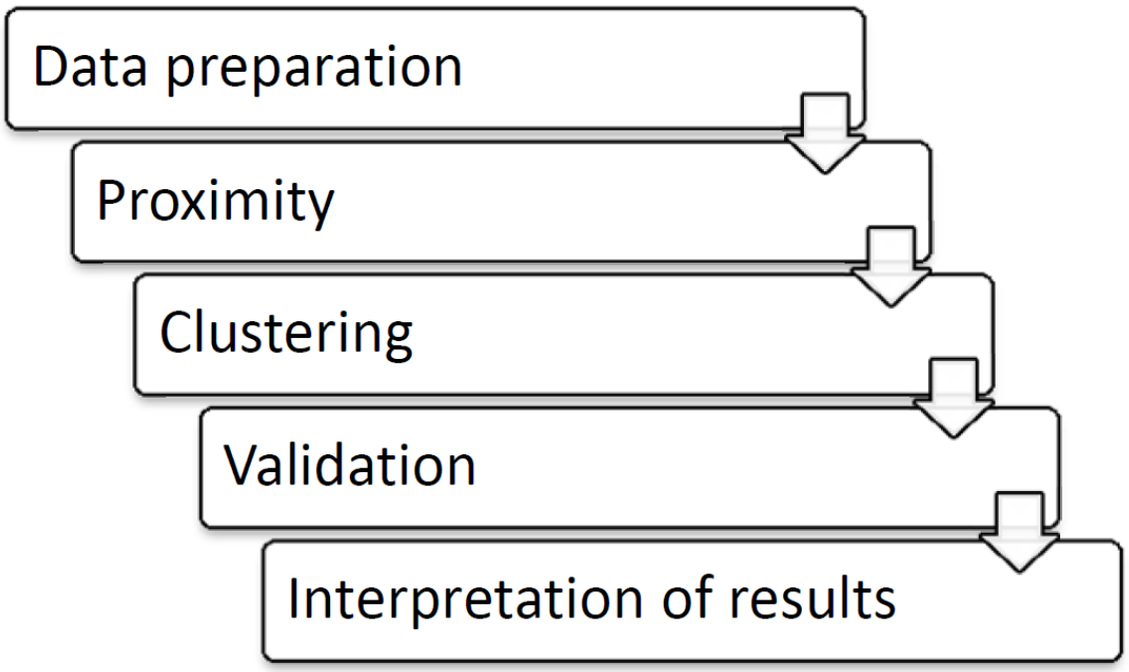

Figure 1. Stages in the clustering process, according to [29]

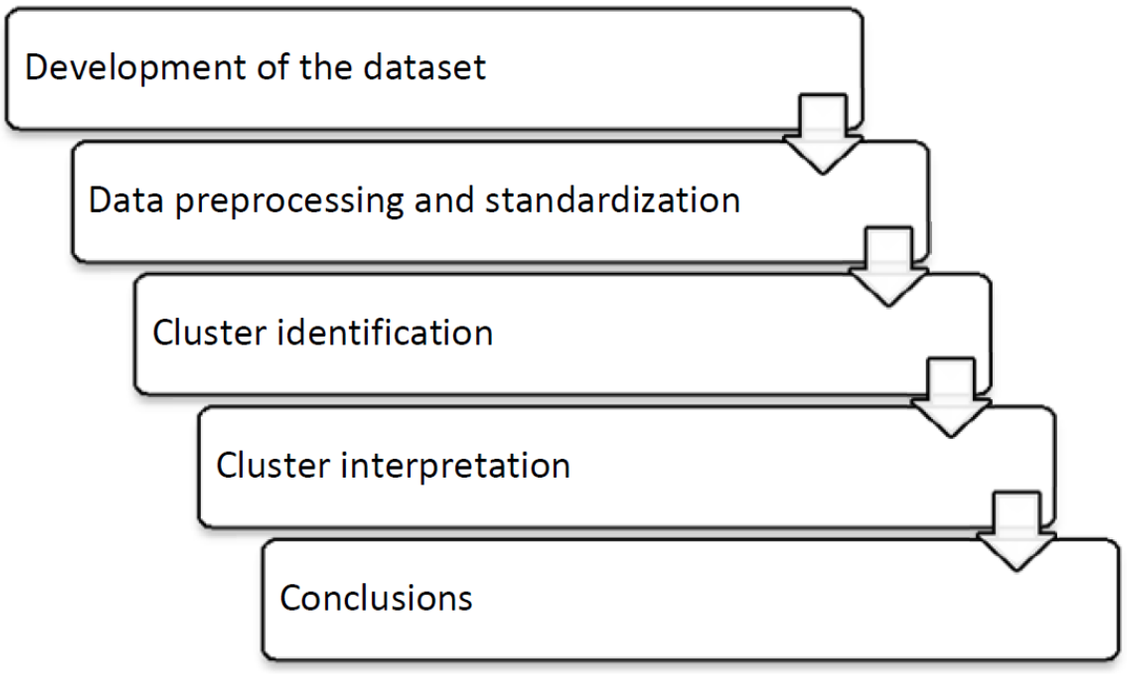

Figure 2. Stages in the clustering process, according to [28]

i. First stage: the development of the data set includes the problem definition, and the choice of the data to be analyzed;

ii. Second stage: the pre-processing is the stage of data preparation, which includes the standardization of variables used in the process;

iii. Third stage: the stage cluster identification consists of applying an algorithm to the data set, resulting in a cluster structure;

iv. Fourth stage: the stage of interpretation must be performed by specialists, who analyze the characteristics used in the cluster to verify the relevance of the obtained results and, if necessary, suggest modifications in the data, followed by reapplication of the previous stages;

v. Fifth stage: the final stage corresponds to the interpretation of results and formulation of conclusions, focusing on the regularities implicit in the results.

In [30], it is described a third clustering process, based on a model slightly different, with six stages,is shown in Figure 3 and described below:

i. First stage: in this stage the objectives to be achieved with the task of clustering and the selection of variables used to characterize the clusters are defined. Objectives cannot be separated from the variable selection, because the researcher restricts the possible results through selected variables;

ii. Second stage: in this stage some matters regarding the procedures to be adopted in case of outliers detection are evaluated, and decisions are taken about how to measure the similarity of objects and if there is any need for data standardization of;

iii. Third stage: in this stage, it is performed an evaluation of the assumptions that were made during the previous steps, which concerns the representativeness of the sample and the impact of variable multicollinearity in the clustering process;

iv. Fourth stage: in this stage cluster definition is performed, where it is necessary to determine which algorithm is used, the number of clusters to be formed, and identify, from the results obtained, if it will be necessary to set the clustering process again;

v. Fifth stage: This stage involves the interpretation of the obtained clusters, where the specialist will examine each cluster formed for the purpose of appointing or designating a label that accurately describes its fundamental characteristics;

vi. Sixth stage: This stage is responsible for validating the solution obtained and by clusters of clusters found. Validation aims to ensure that the solution of clusters is representative for the general population, and thus is generalizable to other objects and stable over time. The profile of clusters involves the description of the characteristics of each cluster to explain how they may differ in important dimensions.

http://dx.doi.org/10.5772/52776

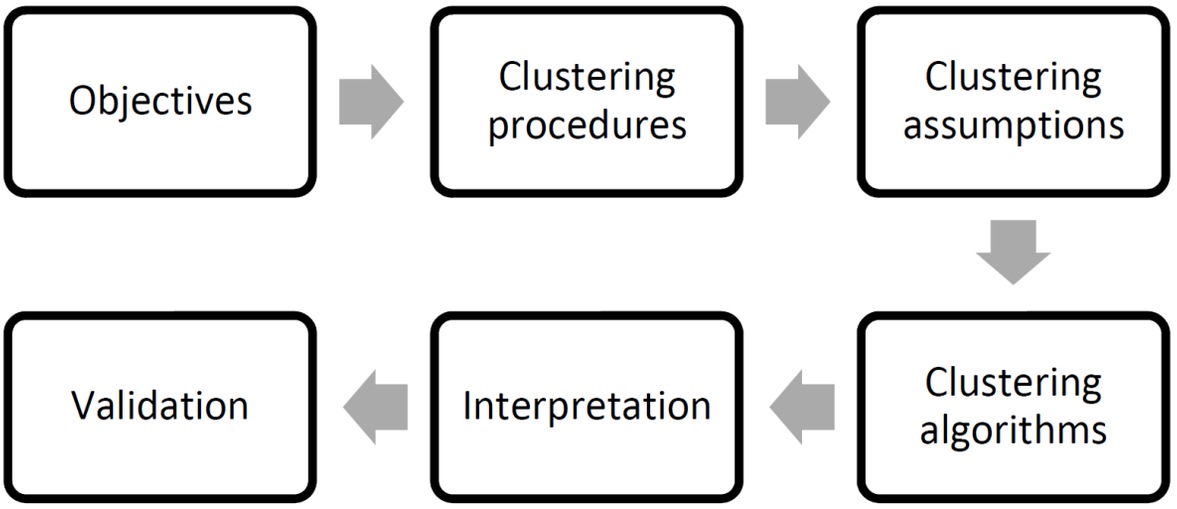

Figure 3. Stages in the clustering process, according to [30]

3.3. Similarity and dissimilarity measures

The task of identifying similar items from the existing ones in an input set requires the adoption of a metric distance between the items 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, more alike (or near) the items are. Dissimilarity measures the difference between items, the greater the dissimilarity, the more different (or far) they are [31].

Considering each item of the input set as a vector in the *p*-dimensional space, a distance function between two items x_i and x_j of the set X may be defined as in equation (3):

$$d : X \times X \to \mathbb{R} \tag{3}$$

$$d_{ij} = d(x_i, x_j)$$

where d_{ij} is a real value associated with each pair of items in the input set and is calculated from a measure of similarity (or dissimilarity) that meets the following assumptions:

i. $d(x_i, x_j) = d(x_j, x_i), \quad \forall x_i, x_j \in X$

ii. $d(x_i, x_j) \geq 0, \quad \forall x_i, x_j \in X$

iii. $d(x_i, x_j) = 0 \leftrightarrow x_i = x_j, \quad \forall x_i, x_j \in X$

iv. $d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j), \quad \forall x_i, x_j, x_k \in X$

In the literature, various metrics of similarity and dissimilarity are presented which meet these conditions. The choice of a metric is associated with characteristics of the input set, such as the nature of the variables (discrete, continuous, binary, etc.), the scale of measure-

ments (nominal, ordinal, intervallic, etc.)., The format of the clusters in *p*-dimensional space (spherical, square, elliptical, etc..) and even the preference of the researcher [33, 34].

In this work, the similarity metric used was Euclidean distance, because it is the most widely used in classification and clustering tasks, which is a generalization of the distance between two points on a Cartesian plane and is given by the square root of the sum of squares of differences of values of each attribute. Mathematically, it is defined by:

$$d_{ij} = \left(\sum_{f=1}^{p} \left| x_{if} - x_{jf} \right|^2 \right)^{1/2} \tag{4}$$

wherex_i and x_j are two input vectors in *p*-dimensional space and x_{if} corresponds to the f^{th} attribute of the vectorx_i.

3.4. Metrics for the evaluation of results

One of the difficulties in clustering tasks is to measure whether the results are satisfactory, since in most cases, not much is known about the data being analyzed. Several metrics have been proposed for evaluation of results in clustering tasks [31,35-42], most of them are based on the application of cluster validation indices, which measure the average intra-cluster distances (between objects belonging to the same cluster) and inter-cluster (between objects belonging to different clusters). According to [52], the index most used for this purpose are: Silhouette index, Dunn index and the Davies-Bouldin index, and among these, the Davies-Bouldin index is more robust for use in tasks whose data sets have hyperspherical clusters, with no outliers, features common in applications that use the K-means and SOM algorithms.

Being $C = \{C_1, C_2, \cdots, C_k\}$a partition of the input set X. The Davies-Bouldin index for the partition C_i is calculated as defined in equation (5):

$$db(i) = \frac{1}{K} \sum_{i=1}^{K} R_i \tag{5}$$

where$K$ is the number of existing partitions and R_i is the relative similarity between the cluster C_i and the other clusters. The similarity R_{ij} between clusters C_i and C_j is computed as described in equation (6):

http://dx.doi.org/10.5772/52776

$$R_{ij} = \max_{i \neq j} \frac{\left(\frac{e_i}{\sqrt{n_i}}\right) + \left(\frac{e_j}{\sqrt{n_j}}\right)}{d_{ij}} \tag{6}$$

where d_{ij} is the distance between the mean element (centroid) of the clusters*i* and *j*, $n_k$is the number of elements of the cluster *k*and $e_k$is the average square distance between elements in cluster *k*and its centroid, given byequation (7):

$$e_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \left(x_i - w_\xi\right)^2 \tag{7}$$

wheren_k is the number of elements in cluster*k*, $x_i$is an element in cluster *k*and w_ξ represents the centroid in cluster*k*.

3.5. Self-organizing mapsalgorithm

Self-organizing maps (SOM) are a class of neural networks for unsupervised, collaborative and competitive learning, which have been widely used in automatic data classification tasks, visualization of high dimension data and dimensionality reduction [43]. Self-organizing maps, like other clustering algorithms, are used to identify clusters of objects based on similarities found in their attributes, i.e., features. Thus, in the end of a clustering process, it is possible to identify which objects have greater similarity to each other and which are more different.

The architecture of a SOM neural network is extremely simple, consisting of only two layers of neurons (Figure 4). The first input layer, comprising a vector with p neurons, is the dimensionality of the input set (i.e., the number of features of the data table). Each input neuron is connected to all neurons of the next layer. The second layer, also known as the output layer, the map which represents the set of input will be projected, and comprises a set of neurons, usually arranged in the form of a vector (unidimensional) or a matrix (two-dimensional), where each neuron is connected only to its neighbors.

During the training phase of a SOM neural network, each representative of the input set is randomly selected and presented to the input layer of the network. An activation function computes the similarity between the input vector and all neurons of the map. The neuron of the output layer which is most similar to the input neuron is declared the winner and their synaptic weights, as well as the synaptic weightsof their neighbors, are updated. The process is repeated with the other vectors of the input set, several times, until the network is trained.

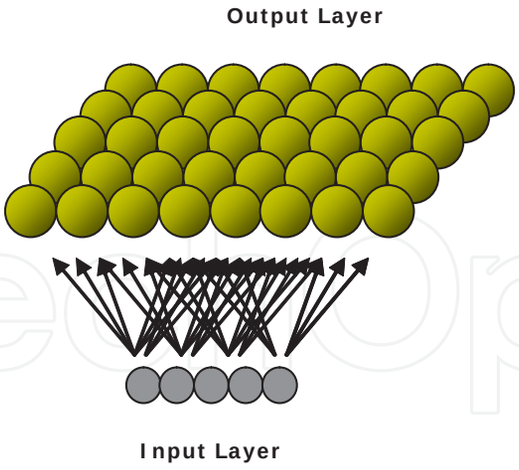

Figure 4. Architecture of a SOM neural network

The similarity function commonly used to calculate the distance between the input vector and the neuron network is the Euclidean distance as shown in equation (8) given by:

$$d_{ij} = \left(\sum_{f=1}^{p} \left| x_{if} - w_{jf} \right|^2 \right)^{1/2} \tag{8}$$

wherex_i is an input vector in the *p*-dimensional space,w_j is a neuron of the output layer and x_{if} represents the f^{th} attribute of the vectorx_i.

To identify the winning neuron (*bmu*, i.e., best match unit), it is necessary to check all the neurons of the output layer, in order to identify which of them has the shortest distance to the input vector, by using the equation (9):

$$w_{\xi} = \min\left(d_{ij} \right) \tag{9}$$

wherew_{ξ} represents the winner neuron and d_{ij} is the Euclidean distance between an element of the input set and an output layer neuron.The synaptic weights of the neuron and its neighborhood are updated using the equation (10):

$$m_i\left(t+1\right) = m_i\left(t\right) + h_{ci}\left(t\right) \cdot \left[x\left(t\right) - m_i\left(t\right) \right] \tag{10}$$

where*t* represents time, $x(t)$represents any element in the input set and h_{ci} determines the neighborhood radius to be modified, usually being reduced while the training algorithm progresses.

http://dx.doi.org/10.5772/52776

In pattern recognition tasks, which are a major application for this type of network, being the winner neuron means to be the most similar neuron, from the existing in the output map, to the value presented to the input of the network. The winner neuron has, along with its neighborhood, its values enhanced, so that if the same input is subsequently presented to the network, that region of the map will be further enhanced.

3.6. K-means algorithm

Initially proposed in [44], the K-means is a partition clustering algorithm, one of the most known and used in clustering tasks, especially due to its simplicity and easy implementation.

As with other clustering algorithms, the goal of K-means algorithm is to cluster a set of n items into k groups, based on a given similarity measure, which is usually the Euclidean distance. The basic idea of the K-means clustering is based on the centroids, which are the average of a group of points. Its training process takes place considering all the vectors in each iteration, and the process is repeated until convergence [45]. Convergence occurs when there is no change in value of the centroids or when the processing reaches the limit of iterations, normally very high. At the end of processing, each element is said to belong to the cluster represented by its centroid.

Then the K-means algorithm is described, presenting its stages, as follows:

1. Set the value of k, corresponding to the number of groups of the sample;
2. Randomly select a set of centroid to represent the k groups;
3. Calculate a matrix of distances between each set of data elements and each centroid;
4. Assign each element to its nearest centroid;
5. Recalculate the value of each centroid from the average values of the elements belonging to this centroid, generating a new matrix of distances;
6. Return to step 4 and repeat until convergence.

The K-means algorithm has linear complexity O(npk), where n and p are, respectively, the number of elements and the dimensionality of the data set, and k is the number of desired clusters. The K-means has good scalability, since the values of p and k are, in most cases much smaller than n [46]. In addition, being based on the principle of vector quantization, the algorithm works well on compact, hyperspherical and well defined clusters.

Among the disadvantages of K-means there is a need to provide a pre-set value to k, the number of clusters, which often is done randomly. The main strategy to overcome this difficulty is to run the algorithm several times, for different values of k and measure up the cohesion of clusters detected by cluster validation indices. In [47], several other techniques are presented to approach this problem.

In [48], it is indicated as the main disadvantage of the K-means the fact that it is a nondeterministic algorithm, strongly influenced by both the initialization values as well as small

changes in the training set, which can influence major alterations in solution which the algorithm converges, which makes this algorithm a rather unstable one. As the choice of initial values of the centroids is usually done at random or from elements that compose the set of input data, this strategy is widely criticized and some changes have been proposed to improve the performance of this algorithm [28].

In [49], it is emphasized that the K-means is not an appropriate method to deal with non-convex shaped clusters or of different sized clusters as well as being very sensitive to noise and distortion (outliers), so that a small number of data having such characteristics can significantly influence the values of the centroids.

Despite all the criticism, K-means is one of the most studied clustering algorithms, having a large number of variants that differ in small details, such as in the way of selecting the initial centroids, in calculating the similarity between the centroids and elements of the input set and the strategies used to compute the centroid of each cluster [49].

Examples of variations of the K-means are K-modes, which uses the concept of fashion, rather than average, to cluster categorical data; and K-medoids, which uses real components of the input set to represent the cluster centroids, reducing the influence of noise and distortion. In addition, other algorithms that were later developed, such as LBG, Expectation-Maximization and SOM, share common ideas with the K-means.

4. The proposed strategy

The problem addressed earlier in this chapter concerns the formation of heterogeneous teams, aiming to encourage integration of students with different profiles and thus promote knowledge sharing and mutual learning. However, clustering algorithms, as described in the previous section, act in a contrary way, identifying clusters of objects based on common features and similarities found in their attributes, i.e., these algorithms identify homogeneous groups. What at first glance may seem contradictory is resolved through the use of a strategy of teaming that promotes diversity in each team.

The strategy of this approach can be divided into two stages: in the first stage, clustering algorithms are used to identify individuals having a similar academic profile, according to a selection criterion, such as performance at school activities; in the second stage, an algorithm for the distribution of students into teams is applied, which allocates students with similar profile in different teams, favoring heterogeneity of teams.

Clustering tasks using K-means algorithm tend to establish a direct relationship between the number of centroids and the expected number of clusters, so that each centroid represents a group of individuals. Unlike this, self-organizing maps generally utilize a two-dimensional grid, with a much higher number of neurons than the expected number of groups, which allows obtaining more detailed results than those obtained with K-means centroids. Taking this point in consideration, self-organizing maps have a superior performance than K-means

in clustering tasks, since they provide information about the proximity between objects in the results presented.

However, while the K-means algorithm, at its output, provides labels corresponding to each object in the input set, allowing the direct relationship of each object to the group it belongs to, self-organizing maps provide more subjective information, suggesting that objects that are mapped to a single neuron or adjacent neurons in the output map, have a close relationship in the input set and belong to the same group. Thus, the association of objects from the input set to the clustering they belong to is not performed directly.

One of the approaches traditionally used to label the elements of the input set in clustering tasks which use the SOM algorithm is to perform a new clustering process on the neurons of the map in order to identify groups of neurons and assign similar elements that are associated with those neurons as belonging to a same cluster. This approach is presented in [54], using K-means algorithm to segment the output map of the SOM algorithm in distinct k regions, where k represents the number of desired groups.

A similar approach is proposed in this paper, which uses a combination of SOM and K-means to segment the input set, corresponding to the students in the class, in k groups, where k represents the desired number of students on each team. Then the strategy is applied to separate the teams, which selects one element from each group for the formation of a heterogeneous team. Figure 5 summarizes the process, which is detailed below:

1. Initially, the data of the students are gathered in a single set, from which a subset of attributes to represent each individual is selected;
2. In stage 1, the SOM algorithm is applied on the selected attributes, organizing individuals in accordance with the similarity which they have to each other. Also in this stage, K-means algorithm is applied on the SOM obtained results in order to segment the groups obtained;
3. In stage 2, a distribution algorithm is applied, which allocates similar individuals into distinct groups, favoring the formation of heterogeneous groups;
4. In step 3, final adjustments are made and each team is allocated.

5. Used methodology and obtained results

In order to validate the strategy proposed in this chapter, this section presents the results of using this approach on two databases selected for the experiments: the Iris database and a real database with academic performance of undergraduates from the course of Bachelorship in Information Systems at Federal University of Rio Grande do Norte, superior education institution located in the northeastern region of Brazil.

Iris is one of the most popular data sets publicly available and has been widely used in testing algorithms for pattern recognition, machine learning and data mining. Although this da-

tabase is not related to the context of applications proposed in this chapter, it waschosendue to its being a dataset widely known and used, whosereference values are known a priori and can be used for validity comparison of the proposed strategy.

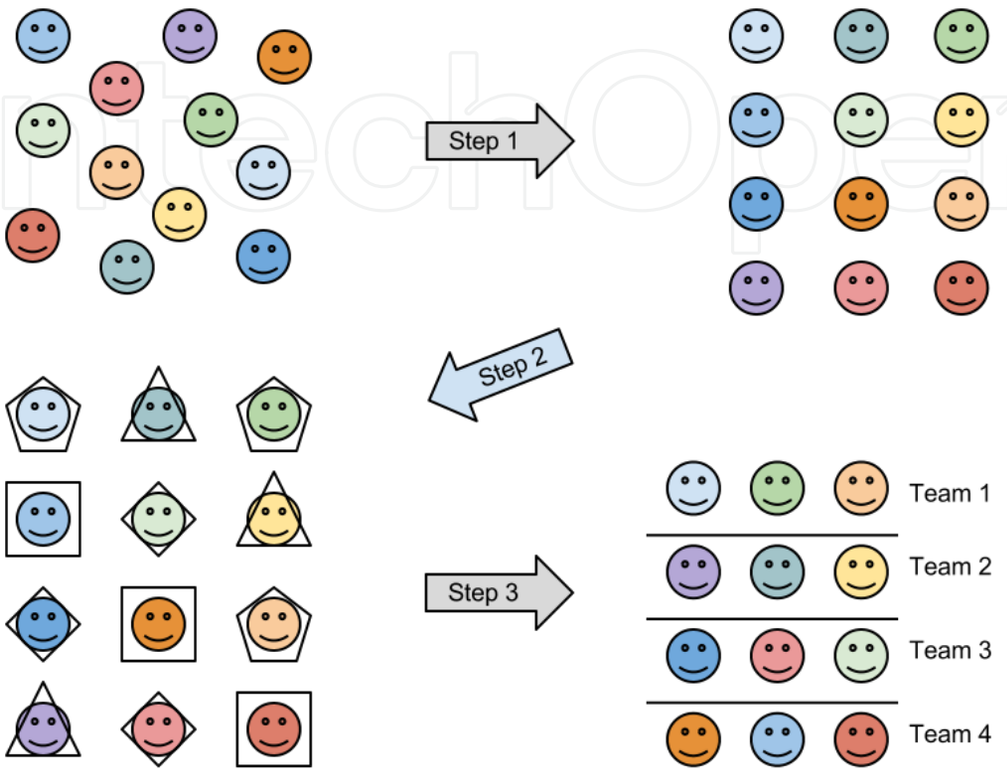

Figure 5. The proposed strategy

This database has 150 instances containing data from measurements of the width and length of three species of the flower Iris, namely, *Setosa*, *Versicolor* and *Virginica* [53]. Each instance of the base has four attributes, corresponding to length and width of sepal, length and width of petal, as well as additional information about class and order number that are not considered in the experiments. The 150 instances are equally divided, so that each species has 50 records. A sample of Iris database is shown in Table 2.

Instance	Sepal length	Sepal width	Petal length	Petal width	Class
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
3	4.7	3.2	1.3	0.2	Setosa
...	...	...	...	...	...
150	5.9	3.0	5.1	1.8	Virginica

Table 2. Sample of the Iris dataset structure

http://dx.doi.org/10.5772/52776

In the experiments with the Iris database, the main objective was to determine whether the strategy worked correctly, actually forming heterogeneous groups, consisting of instances belonging to different species. For the experiments described here, we considered only the four attributes related to the length and width of sepals and petals, and ignored the attributes related to the number and class to which the instance belongs.

Initially, the experiment simulated the process of teaming in the classroom, which is usually conducted by draw, with groups being formed randomly. For this, the Iris dataset was divided into 50 groups, each containing three instances of the database. Then, the process of teaming was repeated with the same numbers as the previous experiment, but applying the strategy proposed in this paper.

For the proposed approach, the data were originally submitted to the SOM algorithm, and then the map obtained at the output of SOM was segmented using the K-means algorithm. All experiments in this paper were implemented from the use of the package SOM Toolbox 2.0 [54]. In all cases, the size of the maps was established automatically from estimates made by the algorithm available on the implementation of the SOM Toolbox, which also used the method of linear initialization of maps [43] and batch training. For training the SOM, we used sheet shaped maps, with 11 x 6 neurons dispersed in hexagonal shape. Figure 6 shows the maps obtained during the experiment. The left map represents the U-matrix obtained directly from the SOM algorithm, while the map on the right shows the segmentation of neurons derived from the application of K-means algorithm.

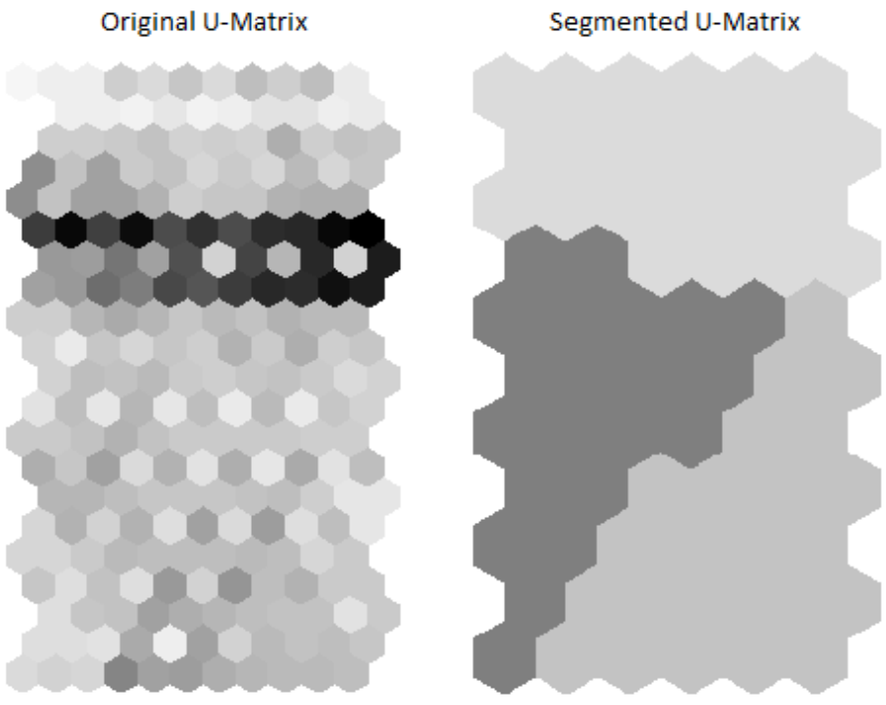

Figure 6. Original and segmented U-Matrix relating to Iris dataset

The comparison between the two approaches was performed both qualitatively and quantitatively. Through visual observation, we found that many of the groups formed by random strategy, had two or three elements belonging to the same class, suggesting the presence of homogeneous groups in the formation of teams. By repeating the experiment with the proposed approach, this condition of existence of more than one instance of a group belonging to the same class is minimized, being reduced to a few instances, from classification errors of the algorithm. The Iris dataset has an interesting characteristic, the class *Setosa* can be linearly separated from the others, but the classes *Versicolor* and *Virginica* are not linearly separable and, in general, clustering algorithms for classification and clustering make mistakes in erroneously assigning some instances belonging to these classes.

From the quantitative point of view, intra-cluster and inter-cluster dispersion measures were used to measure the heterogeneity of the groups formed using the Davies-Bouldin index (db). Table 3 presents the results of minimum, maximum and average db index, and the standard deviation of these measures, obtained in 20 executions of the algorithms, using both approaches.

Methodology	Minimum db index	Maximum db index	Average db index	Standard deviation
Homogeneous clustering	0.61	0.68	0.64	0.02
Random approach	10.06	14.16	12.12	1.13
Proposed strategy	18.63	21.28	19.82	1.35

Table 3. Heterogeneity of groups for the Iris dataset measured by the Davies-Bouldin index

Once demonstrated the applicability of the proposed strategy for the formation of heterogeneous groups, based on experiments performed with Iris dataset, the second set of experiments used a real dataset, named Students dataset, within the context of the problem discussed in the beginning of the chapter. The dataset used contains information on the academic performance of a group of students in a particular class, in various disciplines of the undergraduate program in Information Systems UFRN.

The Students dataset comprises 43 samples, corresponding to the students comprising the examined group and 39 attributes were considered, corresponding to the course subjects. The performance of each student is expressed as a score between 0.0 and 10.0. If the student has not attended a particular discipline, that discipline is scored as 0.0. Since there is no prior information about this dataset, we do not know the number of clusters available. A sample of the database students is shown in Table 4.

The experiments performed with Students dataset were conducted in analogous manner to that performed with the Iris dataset. In this case, the main objective was to determine whether the proposed strategy could form heterogeneous teams composed of students with different profiles and different academic performance. As in the previous experiment, two approaches were taken, the first using a random teaming process, and the second, applica-

tion of SOM and K-means clustering algorithms and then a strategy for distributing students with similar performances in different teams. The results were also compared through the same criteria used previously, qualitative analysis of the teams formed, based on comparison of profiles of the selected students on the same team, and quantitative assessment, measured through the use of the Davies-Bouldin index. Table 5 presents the results of minimum, maximum and average intra-cluster and inter-cluster dispersion measures,and standard deviation of these measures, obtained in 20 executions of the algorithms, using both approaches and the db index to measure the heterogeneity of the groups formed.

Instance	Discipline 1	Discipline 2	Discipline 3	...	Discipline 39
Student 1	8.4	9.8	9.5	…	0.0
Student 2	6.4	8.0	7.2	…	0.0
Student 3	9.8	9.2	9.3	…	10.0
...	…	…	…	…	…
Student 39	8.5	7.2	7.1	…	0.0

Table 4. Structure of Students dataset

For training the SOM, sheet shapedmaps, with 11 x 6 neurons dispersed in hexagonal shape were used. In all cases, the size of the maps was set automatically from estimates made by the algorithm, available on the implementation of the SOM Toolbox, which also used the method of linear maps startup and batch training.

Methodology	Minimum db index	Maximum db index	Average db index	Standard deviation
Random approach	2.87	4.09	3.49	0.32
Proposed strategy	4.01	5.04	4.52	0.45

Table 5. Heterogeneity of groups for the Students dataset measured by the Davies-Bouldin index

6. Conclusions and final thoughts

Throughout human history, there are several approaches that contributed to the improvement of teaching and learning. However, virtually all of these approaches have one thing in common: the ability of humans to learn from their peers. Within this context, the development of team activities is often a common practice in society, adopted in performing various daily tasks. In school, this practice has been widely used due to its fostering mutual learning. In fact, the formation of heterogeneous teams facilitates the sharing of ideas and experi-

ences among members of a team, allowing the exchange of knowledge between them and carrying out of activities that are not likely to be done individually.

However, the procedures commonly adopted by teachers in the classroom for teaming do not always contribute to knowledge exchange and mutual learning. Teams formed at random or from affinities between its members do not favor the heterogeneity. Furthermore, individuals with the same academic profile and who have knowledge in the same areas have less information and content to provide and share with each other. Thus the process of teaming must be guided so as to prioritize heterogeneity among members of the teams.

The use of computational tools to solve problems in the area of education has been an increasingly common practice. In this context, a research field that has received recent attention is the educational data mining, which seeks to use data mining techniques in order to investigate problems that affect learning, as well as the development of educational systems. Such surveys are presented as an alternative to solving these problems that are focused primarily on exploring the dataset collected in educational settings. However, analyzing the literature available in the area, one can identify a lack of algorithms and tools to improve the process of academic teaming, since most of the available algorithms search homogeneous groups.

Thus, this paper presents a strategy capable of forming heterogeneous teams by using traditional clustering algorithms, such as K-means and self-organizing maps, contributing to the process of forming study groups and conducting works in academia. By using cluster validation indices, such as Bouldin-Davies index, the results obtained from the experiments carried out show that the teams formed by the use of the proposed strategy are more heterogeneous than those obtained with the methods conventionally used in classroom, such as random or affinity-based approaches, demonstrating its efficiency in the formation of heterogeneous groups of objects, both in educational and other datasets.

Future work may include optimizations in the proposed strategy, in order to even more heterogeneous teaming to be achieved. Using genetic algorithms to organize teams during the second stage of the strategy appears to be a viable alternative to evaluate different possible combinations of individuals, thus promoting heterogeneity. On the other side, the use of other clustering algorithms, more stable and with improved performance, can also contribute to better results in the team allocation process. Finally, assessments in relation to learning and performance of students through the process of developing team activities can prove the greater efficiency of utilization of diverse teams, compared to homogeneous teams.

Author details

Huliane M. Silva, Cícero A. Silva and Flavius L. Gorgônio

*Address all correspondence to: flavius@ufrnet.br

Laboratory of Computational Intelligence Applied to Business Federal Universityof Rio Grande do Norte, Caicó, RN, Brazil

References

[1] Fonseca MJ. A Paideia Grega Revisitada. RevistaMillenium 1998;3(9) 56-72.

[2] Zabala A. A Prática Educativa: Como Ensinar. Porto Alegre: Artmed; 1998.

[3] Libâneo JC. Didática. São Paulo: São Paulo; 1994.

[4] Behrens MA. Projetos de Aprendizagem Colaborativa num Paradigma Emergente. In: Moran JM, Masetto MT, Beherens MA. (eds.) Novas Tecnologias e Mediação Pedagógica. Campinas: Papirus; 2000. p67-132.

[5] Torres PL, Irala EAF. Aprendizagem Colaborativa. In: Torres PL. (ed.) Algumas Vias para Entretecer o Pensar e o Agir. Curitiba: SENAR-PR; 2007. p65-97.

[6] Panitz T. Collaborative versus cooperative learning: A comparison of the two concepts which will help us understand the underlying nature of interactive learning, Cooperative Learning and College Teaching 1997;8(2) 1-13.

[7] Perrenoud P. Dix NouvellesCompétences pour Enseigner. Paris: ESF Éditeur; 1999.

[8] Colenci AT. O ensino de engenharia como uma atividade de serviços: a exigência de atuação em novos patamares de qualidade acadêmica. MSc thesis. Universidade de São Paulo; 2000.

[9] Maybi S. Team Building: comoconstruirequipeseficazes. Specialist thesis. Universidade de Passo Fundo; 2000.

[10] Gillies RM. Cooperative Learning: Integrating Theory and Practice. Thousand Oaks: Sage Publications; 2007.

[11] Millis B, Rhem J. Cooperative Learning in Higher Education: Across the Disciplines, Across the Academy. Sterling: Stylus Publishing; 2010.

[12] Costa JAF. Classificação automática e análise de dados por redes neurais auto-organizáveis. DSc thesis. UniversidadeEstadual de Campinas; 1999.

[13] Amorim T. Conceitos, técnicas, ferramentas e aplicações de Mineração de Dados para gerar conhecimento a partir de bases de dados. Undergraduatethesis. Universidade Federal de Pernambuco; 2006.

[14] Tan PN, Steinbach M, Kumar V. Introduction to Data Mining. Boston: Addison Wesley; 2005.

[15] Sferra HH, Corrêa AMCJ. Conceitos e Aplicações de Data Mining: Data Mining ConceptsandApplications. Revista de Ciência&Tecnologia 2003;11(22): 19-34.

[16] Ha SH, Bae SM, Park SC. Web mining for distance education. In: IEEE Engineering Management Society (eds.) ICMIT 2000: Management in the 21st Century: proceedings of the IEEE International Conference on Management of Innovation and Technology, v2, p715-719, ICMIT2000, 12-15 Nov 2000, Orchard Hotel, Singapore. IEEE Engineering Management Society; 2000.

[17] Machado AP, Ferreira R, Bittencourt II, Elias, E, Brito P, Costa E. Mineração de Texto em Redes Sociais Aplicada à Educação a Distância. Colabor@ - Revista Digital da CVA 2010; 6(23). http://pead.ucpel.tche.br/revistas/index.php/colabora/article/download/132/115 (accessed 20 May 2012).

[18] Paiva R, Bittencourt II, Pacheco H, Silav AP, Jaques P, Isotani S. Mineração de Dados e a Gestão Inteligente da Aprendizagem: Desafios e Direcionamento. In: SBC proceedingsofthe I Workshop de Desafios da Computação Aplicada à Educação, DesafIE'2012, 17-18 July 2012, Curitiba, Brazil. Curitiba: UFPR; 2012.

[19] Baker RSJ, Isotani S, Carvalho AMJB. Mineração de Dados Educacionais: Oportunidades para o Brasil. RevistaBrasileira de InformáticanaEducação2011;19(2) 3-13.

[20] Baker RSJ. Date Mining for Education. In: McGaw B, Peterson P, Baker E. (eds.) International Encyclopedia of Education. Oxford: Elsevier; 2010. p112-118.

[21] Zaina LAM, Ruggiero WV, Bressan, G. Metodologia para Acompanhamento da Aprendizagem através da Web, Revista Brasileira de Informática na Educação 2004;12(1): 20-28. http://www.lbd.dcc.ufmg.br/colecoes/rbie/12/1/002.pdf (accessed 22 May 2012).

[22] Milani F, Camargo SS. Aplicação de Técnicas de Mineração de Dados na Previsão de Propensão à Evasão Escolar. In: Congresso Sul Brasileiro de Computação: proceedingsofthe V Congresso Sul Brasileiro de Computação, V SULCOMP, 29 Sept - 1 Oct 2010, Criciúma, Brazil. Criciúma: Ed. UNESC; 2010.

[23] Silveira SR. Formação de grupos colaborativos em um ambiente multiagente interativo de apredizagem na internet: um estudo de caso utilizando sistemas multiagentes e algoritmos genéticos. DScthesis. Universidade Federal do Rio Grande do Sul; 2006.

[24] Kampff AJC. Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à prática docente. DSc thesis. Universidade Federal do Rio Grande do Sul; 2009.

[25] Pimentel EP, França V, Omar N. A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. In: Sampaio FF, Motta CLR, Santoro FM (eds.) Proceedingsofthe XIV Simpósio Brasileiro de Informática na Educação, SBIE'2003, 12-14 November 2003, Rio de Janeiro, Brazil. Rio de Janeiro: NCE/IM/UFRJ; 2003.

[26] Azambuja S. Estudo e implementação da análise de agrupamento em ambientes virtuais de aprendizagem. MScthesis. Universidade Federal do Rio de Janeiro; 2005.

[27] Zanella A, Lopes LFD. Melhoria da qualidade do ensino através da análise de agrupamento. In: ABEPRO (eds.) Proceedingsofthe XXVI Encontro Nacional de Engenharia de Produção, ENEGEP'2006, 9-11 October 2006, Fortaleza, Brazil. Fortaleza: ABEPRO; 2006.

[28] Mirkin B. Clustering for Data Mining: A Data Recovery Approach. Boca Raton: Chapman and Hall/CRC; 2005.

[29] Faceli K, Lorena AC, Gama J, Carvalho ACPLF. Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina. Rio de Janeiro: LTC; 2011.

[30] Hair Jr. JF, Anderson RE, Tatham RL, Black WC. Multivariate Data Analysis. Upper Saddle River: Prentice Hall; 2005.

[31] Frei F. Introdução à Análise de Agrupamento: Teoria e Prática. São Paulo: Unesp; 2006.

[32] Gorgônio FL. Uma arquitetura para análise de agrupamentos sobre bases de dados distribuídas aplicadas a segmentação de mercado. DScthesis. Universidade Federal do Rio Grande do Norte; 2009.

[33] Kasznar IK, Gonçalves BML. Técnicas de Agrupamento: Clustering. EletroRevista: Revista Científica e Tecnológica 2007;6(20): 1-5. http://www.ibci.com.br/20Clustering_Agrupamento.pdf (accessed 22 May 2012).

[34] Bussab WO, Miazaki ES, Andrade DF. Introdução à análise de agrupamento. São Paulo: ABE/IME/USP; 1990.

[35] Kuncheva LI. Combining Pattern Classifiers: Methods and Algorithms. New Jersey: John Wiley & Sons; 2004.

[36] Pölzlbauer G. Survey and comparison of quality measures for self-organizing maps. In: Paralič J, Pölzlbauer G, Rauber A. (eds.) Proceedings of the Fifth Workshop on Data Analysis, WDA'04, 24-27 June 2004, VysokéTatry. Slovakia: Elfa Academic Press; 2004.

[37] Salazar Giron EJ, Arroyave G, Ortega Lobo O. Evaluating several unsupervised class-selection methods. In: Perez Ortega G, BranchBedoya, JW (eds.) Memorias Encuentro de Investigación sobre Tecnologías de Información Aplicadas a laSolución de Problemas: EITI-2001, Medellín: Universidad Nacional de Colombia, 2001. p1-6.

[38] Salazar Giron EJ, Vélez AC, Mario Parra C, Ortega Lobo O. A cluster validity index for comparing non-hierarchical clustering methods. In: Ortega Lobo O, BranchBedoya JW. (eds.) Memorias Encuentro de Investigación sobre Tecnologías de Información Aplicadas a laSolución de Problemas: EITI-2002, Medellín: Universidad de Antioquia, 2002. p115-120.

[39] Shim Y, Chung J, Choi, I. A comparison study of cluster validity indices using a non-hierarchical clustering algorithm. In: IEEE Computer Society Press (eds.) Proceedings

of the International Conference on Computational Intelligence for Modeling, Control and Automation CIMCA2005, 28-30 November 2005, Vienna, Austria; 2005.

[40] Kim M, Ramakrishna RS. New Indices for Cluster Validity Assessment. Pattern Recognition Letters 2005;26(5) 2353-2363.

[41] Gonçalves ML, Netto MLA, Costa JAF, ZulloJr J. Data clustering using self-organizing maps segmented by mathematic morphology and simplified cluster validity indexes. In: International Neural Network Society (eds.) proceedings of IEEE International Joint Conference on Neural Networks, IJCNN'06, 16-21 July 2006, Vancouver, Canada. Piscataway: IEEE Xplore; 2006.

[42] Saitta S, Raphael B, Smith IF. A bounded index for cluster validity. In: Perner P (ed.) LNCS: Lecture Notes in Artificial Intelligence 4571: proceedings of the 5th International Conference on Machine Learning and Data Mining in Pattern Recognition, MLDM'2007, 18-20 July 2007, Leipzig, Germany. Berlin: Springer-Verlag; 2007

[43] Kohonen T. Self-Organizing Maps. Berlin: Springer; 2001.

[44] MacQueen JB. Some methods for classification and analysis of multivariate observations. In: Le Cam LM, Neyman J. (eds.) Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Jun 21-Jul 18 1965 and Dec 27 1965-Jan 7 1966, Berkeley, USA. Berkeley: University of California Press; 1967.

[45] Linde Y, Buzo A, Gray RM. An Algorithm for Vector Quantizer Design. IEEE Transactions on Communications 1980;28(1) 84-95.

[46] Xu R, Wunsch II D. Survey of Clustering Algorithms. IEEE Transaction on Neural Networks 2005;16(3) 645-678.

[47] Chiang MMT, Mirkin B. Experiments for the number of clusters in K-means. In: Neves J, Santos MF, Machado JM (eds.) LNCS: Progress in Artificial Intelligence 4874: proceedings of the 13th Portuguese Conference on Artificial Intelligence, EPIA'2007, 3-7 December 2007, Guimaraes, Portugal. Berlin: Springer-Verlag; 2007.

[48] Leisch F. Ensemble methods for neural clustering and classification. PhD Thesis. TeschnischeUniversität Wien; 1998.

[49] Han J, Kamber M. Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufmann; 2006.

[50] Ultsch A. Knowledge Extraction from Self-Organizing Neural Networks. In: Opitz O, Lausen B, Klar R. (ed.) Information and classification. Berlin: Springer-Verlag; 1993. p301-306.

[51] Davies DL, Bouldin DW. A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence 1979;1(2) 224-227.

[52] Villanueva WJP, Vonzuben FJ. Índices de validação de agrupamentos. In: Wu ST (ed.) Proceedingsofthe I Encontro dos Alunos e Docentes do Departamento de En-

genharia de Computação e Automação Industrial, EADCA'2008, 12-13 March 2008, Campinas, Brazil. Campinas: UNICAMP; 2008.

[53] UCI repository of machine learning databases. Department of Information and Computer Science, University of California, Irvine, CA, USA. http://www.ics.uci.edu/~mlearn/MLRepository.html (accessed 28 July 2012).

[54] Vesanto J, Alhoniemi E. Clustering of the self-organizing map. IEEE Transactions Neural Networks 2000;11(3) 586–600.

IntechOpen