Applied Learners Clustering Methodology in the Massive University

[1]Samir Anter and [2]Ahmed Zellou
[1]FST Mohammedia, Hassan II University, Casablanca, Morocco
[2]ENSIAS, Mohammed V University, Rabat, Morocco

Abstract: Working in a group, especially, in the academic massive context is a privileged moment of exchange and interaction between learners as well as between learners and the trainer. It is also important to customize the learning process according to the skills and requirements of the each group of learners. We present in this study an algorithm called LeCA (Learner's Clustering Algorithm) to grouping learners according to their profiles. For this, we defined a model of a user profile as well as metrics to calculate the distance between different profiles and also between a profile and a group.

Key words: Clustering, e-Learning, collaborative learning, per-sonalized learning, LeCA, clusters apprentices, massive context

INTRODUCTION

Group learning: In the academic environment, working in groups around a common activity is favoured in all teachings. The group is the basic structure of different teaching practices privileged today (cooperative learning or collaborative learning) whether in the standard educational, hybrid or fully online (e-Learning). Characterized by interactive dimensions, participatory and active these pedagogical practices fully identify with the social and constructivist approach (social constructivist) learning and acquisition of knowledge.

Researcher of this approach, Vygotsky highlighted the importance of social interaction in the development of knowledge, particularly through the concept of Proximal Development Zone (PDZ) (Chaiklin, 2003; Harland, 2003). It introduced the concept of "Potential development" of the individual, representing the difference between the current cognitive level of the latter, determined by its ability to solve a problem alone and the level of potential development as determined through the problem solving by this individual when helped by adults or collaborates with more advanced peers. This educational approach in which each student participates and uses its cognitive capital, advantageously and in a complementary manner, facilitates the acquisition and appropriation of knowledge by integrating a group activity and to give the learner a active role within the group in particular and in the learning process in general.

In addition to the cognitive and active dimensions induced by this approach it also has the benefit of placing the learner at the center of the learning process, increase motivation, cultivate their independent learning, finally to enable him to different internalize-learning nature of such behaviours and attitudes related to social relations. Over all the social constructivist approach would allow learning to learn which is metacognitive dimension untapped by the practices of conventional teaching.

Grouping strategy: Many studies on the formation of groups suggest two ways: spontaneous pairing or imposed pairing (Depover *et al.*, 2004). Spontaneous pairing means that the learner freely makes the choice of belonging to a group while the contrasted pairing means that the group constitution is imposed by a responsible actor (teacher, facilitator, mentor, coordinator) before starting training. Generally, the first combination mode (spontaneous) is a deliberate choice of learners affinity to their peers, often for their ability in the proposed activity their geographical proximity, seeking more similarities within the group, acceptance or mutual recognition.

The second grouping method, resulting from a selection made by a responsible actor, considers contrasting individuals on one or more predefined characteristics: level of expertise or performance, career goals, linguistic or cultural choice. The objective is to create heterogeneous groups while maintaining a balance between cognitive, social and emotional interdependence and positive interdependence (Depover *et al.*, 2004).

Forming groups according to such modalities can be relatively easy in the small effective not exceeding a

---

Corresponding Author: Samir Anter, FST Mohammedia, Hassan II University, Casablanca, Morocco

hundred students. Beyond this number, the task of constitution of groups of learners is often reduced to a cut in numbers alphabetically or by region/geographical area without or little consideration affinities of academic, professional or even personal data learners. The democratization of education has made teaching by groups of students a very responsive approach. Indeed, universities welcome students from more diverse in their social and cultural backgrounds as open access institutions are beginning to generate a disturbing failure rate (Romainville, 2000).

Thus, universities are forced to question not only their traditional pedagogical practices but also their methodology of research, especially, the organization of learners into groups (Romainville, 2000). In the present and future academic institutions, the organization of learners into working groups can be guided by the deliberate and concerted pedagogical strategy by training leaders, particularly in the choice of grouping criteria or profile learners.

In this context of large size, automating the creation of groups is inevitable. In this ultimate objective we propose in this study a new approach based on clustering algorithms. Is defined as a data organization operation (observations, attributes, vectors, etc.) in groups or clusters according to a similarity criterion (Jain and Dubes, 1988; Hartigan, 1975). In many cases, the similarity criteria depends on the data or context and there is no measure that is universally best for all kinds of clustering problems. The key is that it should be determined carefully. Indeed, the results are conditioned by this choice. However, before discussing the approaches to calculating similarity it will be necessary to define for each learner, the characteristics to be treated. In other words, it is essential to determine a profile for each learner.

Literature review: Clustering approaches consists of forming groups of objects that are characterized by a great similarity between them on one hand and by little similarity to the objects of other groups on the other hand (Hartigan, 1975; Steinbach *et al.*, 2000; Ng *et al.*, 2002). The latter is as mentioned above, based on the similarities between different profiles. Different existing clustering approaches using three components, the profile learners, the distance or similarity between objects and the clustering algorithms.

Learner profile: The learner profile was created to adapt the tutor-learner interaction according to each learner. It is defined by Py (1998) as a set of information describing

the learner and generally referring to his knowledge, skills and behaviors. It includes all the information collected about the learner before or during the learning process. This information generally has naked content a heterogeneous nature and structure which implies a difficulty of reuse in different contexts, hence, the standardization.

Some researhers suggest that the learner's profile has a dynamic appearance. In other words, it must take into account the improvement of the learner's knowledge at each stage of learning. Mendelsohn and Dillenbourg (1993) in this case, the profile must have two types of information those that are static (personal information for example) and those that are dynamic and that may change during the learning process. This second type of information is updated automatically depending on the interaction of the learner with the system or by a tutor. As for Self (1990) he provides a profile model as a vector of 4 dimensions, expertise, knowledge, particularities (like his character) and the history of learning.

Different profile standardization works has been proposed. These models include, PAPILearner (Public and Private Information Learner specification) (Dolog *et al.*, 2004) which is a standard proposed by the learner model working group IEEE, IMS-LIP (IMS Learner information package) (Dagger *et al.*, 2002) selected 2004 to make it the basis of a European standard of learner data and IMS-RDCEO (IMS Reusable Definition of Competency or Educational Objective). In terms of content each of these standards specifies a set of information describing a learner that should be included in the profile. These data are relatively common to the standard and can be divided into several components.

- Identification; Contact details of the learner, marital status, institutions, etc
- Security; Passwords, login, security keys. Degrees and obtained certificates
- Preferences; Language, etc
- Skills; Skills, knowledge, notes, etc
- Goals and interests; Objectives, recreation and interest
- Productions; Works and activities
- Others; Relationships with other learners, history, etc

Once, this information is specified it becomes necessary to find a way to organize them in order to exploit them especially in the phase calculating distances or similarities between learners. Different approaches was proposed as presented.

Similarity and distance measurement: A metric can be considered as a distance measure if it satisfies the four conditions (R1-R4):

$$R1\text{-}d(x, y) > 0$$
$$R2\text{-}d(y, y) = 0$$
$$R3\text{-}d(x, y) = d(y, x)$$
$$R4\text{-}d(x, y) \leq d(x, z) + d(z, y)$$

where, $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$ are two objects represented by two vectors and $d(x, y)$ the distance between them. Distances and approaches similarity measurement pro-posed in the literature are very diverse. Among them we can mention: the Euclidean distance, Manhattan distance proposed by Carmichael and Sneath the Minkowski distance (Zhou *et al.*, 2013) Cosinus, Dice, Jaccard, Gower coefficient (Gower, 1971). In the following study we will limit ourselves to the description of the ones we think are most appropriate to our context.

Euclidean distance: The Euclidean distance is originally designed for geometric problems and was then adapted to all object-related problems defined by numeric variables. It's widely used in clustering problems and it considered as the default measure used by the k-means algorithm. It is defined for two objects: $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$ by:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Jaccard: Jaccard is one of the metrics to measure the similarity between objects defined by quantitative variables. It is defined for the two objects: $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$ as follows:

$$Jaccard(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} x_i y_i}$$

Gower coefficient: The metrics defined above can only be used for objects defined by variables of the same type. However, in reality, several objects can be defined by different variables. For example, a learner can be defined by: his name his birthday his marks his age, etc. In this case, it is necessary to adopt a methods that can supports objects defined by variables of different types.

The Gower coeffient is proposed to measure the similarity between objects defined by various variables (Gower, 1971). Thus, for the 2 objects: $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$ as follows:

$$Gower(x, y) = \frac{\sum_{i=1}^{n} w(x_i - y_i) s(x_i - y_i)}{\sum_{i=1}^{n} w(x_i - y_i)}$$

where, $s(x_i, y_i)$ is the degree of similarity between x_i and y_i and $w(x_i, y_i)$ is equal to 1 when x_i and y_i are defined and 0 otherwise.

Clustering: Clustering is the organization of data (observations, attributes, vectors, etc.,) into groups or homogeneous clusters. This is usually done in order to find a simplified representation of the original data, so that, we can explore them easily. The problem of clustering has been addressed in many contexts and in many disciplines (statistics, image processing, artificial intelligence, pattern recognition, data compression, etc). For all these reasons this is an important step in the data analysis process. This reflects its wide audience and usefulness as one of the most fundamental steps for understanding, learning and analyzing data.

Clustering methods are generally supervised or unsupervised. A clustering method is called supervised if it has knowledge of the classifying data (the number of clusters to obtain, for example) in advance, unlike unsupervised method for which no information is given in advance. Moreover, these methods can also be divided into 2 major groups, classification by partitioning and hierarchical classification.

In addition to the two approaches described above a third approach may be considered. This is the hybrid approach using the first two ones. The latter consists first in partitioning the objects according to the hierarchical approach before using the first approach to improve the groups obtained.

Partitioning clustering approach: This method is generally based on single partitioning algorithms. The most famous is k-means which is one of the most popular clustering algorithms. It was introduced by Cox in 1957 and named like this by Forgey (1965), MacQueen (1967), Ball and Hall (1967).

It is a supervised and non-hierarchical hard clustering algorithm (not a fuzzy). Although, it was

proposed 50 years ago it still remains one of the most used algorithms for clustering. The ease of implementation, simplicity, efficiency and empirical success are the main reasons for its popularity. The principle of k-means is that it partitions a set of points into k clusters where each cluster is represented by its centroid. It tries to minimize the sum of the distances of points to different clusters. Let $x = \{x_1, x_2, ..., x_n\}$ be a set of n d-dimensional points divided into k clusters $C = \{c_1, c_2, ..., c_K\}$ where v_k is the centroid of the cluster c_k such as:

$$\bigcap_{i=1}^{k} c_i = \varnothing$$

And:

$$\bigcup_{i=1}^{k} c_i = x$$

The intra-cluster distance is defined by:

$$J(c_k) = \sum_{x_i \in c_k} |x_i - u_k|^2$$

Thus, the objective of k-means clustering is to minimize the sum of the intra-cluster distances given by:

$$J(c) = \sum_{j=1}^{k} \sum_{x_i \in c_k} |x_i - u_j|^2$$

The k-means algorithm starts with an initial partition of k clusters and then exchanges the points between clusters while minimizing the value of J(c). The main steps of k-means are the following:

Algorithm 1; k-mean algorithm:

Step 1: Select an initial partition of k clusters and repeating steps (2) and (3) until the value of J (c) stabilizes

Step 2: Calculate the distance from different objects to different centroids and assigning them to the nearest cluster

Step 3: Recalculate the centroids of different clusters

Figure 1 illustrates the application of the algorithm on an example of four objects to divide into two clusters. As we can see, the centroids of the cluster are badly chosen at the beginning but the convergence of the algorithm corrects this choice:

- Initial state
- Choice of centroids
- Assignment of points to the nearest clusters
- Recalculating the centroids
- Reallocation of points to different clusters
- Recalculating centroids

It to be noted that the k-means versions proposed by Forgey (1965) and MacQueen (1967) are different. Indeed, the MacQueen algorithm recalculates the new centroids each time an object is transferred from one cluster to another unlike the forgy algorithm that only recalculates the centroids once all objects have been affected.

Hierarchical clustering: A hierarchy can be seen as a set of nested partitions (Fig. 2). Graphically, a hierarchy is often represented by a tree structure also said
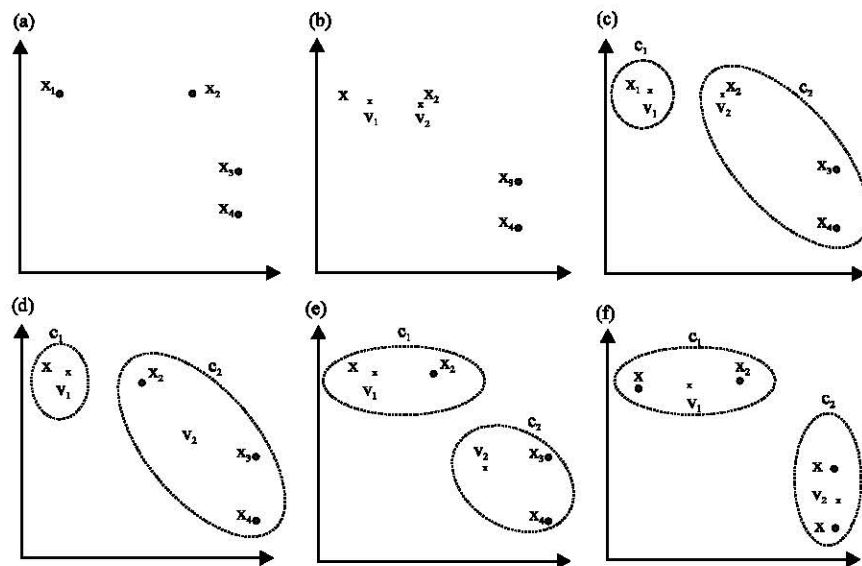

Fig. 1: a-f) Illustration of the k-means algorithm

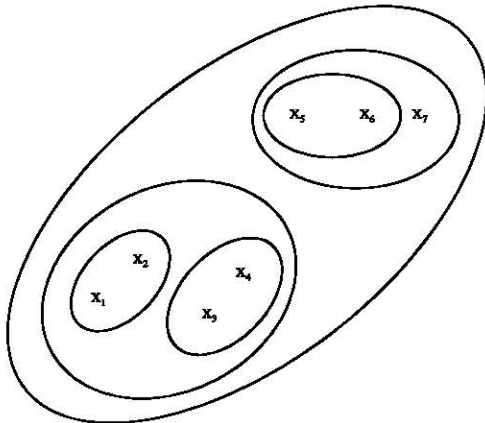## MATERIALS AND METHODS

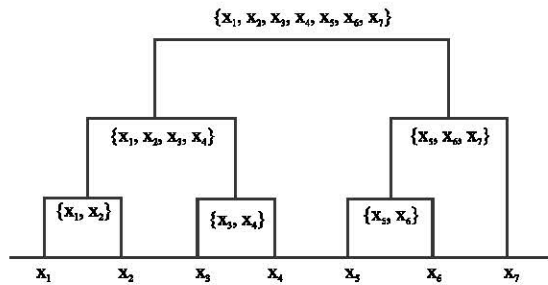Fig. 2: Nested representation

Fig. 3: Dendrogram representation

dendrogram Fig. 3. The principle of hierarchical algorithms is to build a cluster tree (or dendrogram) as presented in Fig. 3:

- The root of the tree is formed by the cluster $x = \{x_1, x_2, ..., x_n\}$ containing all of the objects to be classified
- Each node of the tree represents a cluster $ck \subset x$
- The leaves of the tree correspond to singletons $\{x_1\}$ $\{x_2\}, ..., \{x_n\}$

The method of hierarchical partitioning may appear in two different forms: a top-down approach called divisive and rarely used and a bottom-up approach known agglomerative. The upward decomposition starts with a set of classes each of which is formed by a single object. At each step, the two closest classes are merged until a termination condition is satisfied. The top-down approach against starts with all objects in a single class. At each iteration a class is split into two ones, until a stop condition is satisfied. The fusion or decomposition is based on a degree of similarity or dissimilarity.

Learner clustering: As, we explained upstream, group work and cooperation or collaboration among students has become a necessity in the last decade. This consolidation is also crucial for customization of the learning process based on skills and constraints of the group. For this reason, we propose in this study an algorithm for the formation of these groups based on the learner profile. Thus, our approach is divided into 3 stages. The first will be to define the model of a user profile and the second will define the metric to calculate the distance between learners (their profiles) and also the distance of a learner to a group before organizing them into final groups in a last step.

Learner's profile definition: As we mentioned at the beginning of this study, the different approaches for the standardization of learning profile are almost all agreed on a set of information that should contain. This information are those relating to the identification, security, certification and degrees obtained the preferences of the learner his skills his productions and other different information. Thus, we consider a learner profile consisting of:

- Identification; Name and surname, etc
- Security; Login, passwords or security keys
- Degrees; The title
- Qualifications; The field of education
- Qualities; Represents the level of mastery of content taught during the preparation of the degree
- Languages; The languages spoken
- Accessibility; The information on potential learner disabilities and their learning preferences
- Goals; The goals and objectives
- Interest; Interests, sports, hobbies, etc
- Skills; The skills and knowledge acquired
- Activities; Different activities of the learner
- Relations; Relationship that may exist with other learners
- Progress; The learner's progress during the learning process. Geographical information

In the implementation, the information contained in the learner's profile is represented by a vector $x = \{x_1, x_2, ..., x_n\}$ where, x_i is a variable representing a member of the profile.

In the context of grouping learners these variables should not be treated the same way. Indeed, two learners

who have no centroid of interest in common must not be treated in the same way as in the case of two learners who are far away from each other in terms of skills. To overcome this problem we propose to associate a weight p_i to each variable x_i according to its importance and the context.

Some variables are special they require special treatment. An example of these variables is the language. Indeed, two learners who do not speak any language in common can cause a blockage in the learning process if they are assigned to the same group. In the rest of this study this type of variable is called intolerable variables.

Distance computing: The assignment of a student to a group is based on its distance to the latter. Thus, it is necessary in the first step to calculate the distances between different profiles.

Distance between two profiles: As mentioned in the previous study, the learner profile comprises variables of different kinds. Some are quantitative, others qualitative. This implies that we can not be limited to a single distances from those proposed at the beginning of this study but rather to combine several.

Due to its simplicity, efficiency and popularity we pro-pose to use the Jaccard distance (1-Jaccard (x-y)) for qualitative variables and the euclidean distance for quantitative variables. Thus, the distance $d(x, y)$ between two profiles $x = \{x_1, x_2, ..., x_n\}$ and $y = \{y_1, y_2, ..., y_n\}$ is given by:

$$d(x,y) = \frac{\sum_{i=1}^{n} w_i d(x_i - y_i)}{\sum_{i=1}^{n} w_i}$$

Distance between a profile and a cluster: Once the distance between two profiles is defined it becomes necessary to define the distance between a profile and a cluster. Let $x = \{x_1, x_2, ..., x_n\}$ be a profile and $c = \{c_1, c_2, ..., c_m\}$ a cluster of profiles. The distance between x and c is given by:

$$d(x,c) = \frac{1}{m} \sum_{i=1}^{m} d(x,c_i)$$

Intracluster distance: The intracluster distance is the average distance between profiles in pairs. Thus, for the cluster $c = \{c_1, c_2, ..., c_n\}$ the intracluster distance is given by:

$$J(c) = \frac{2}{n(n-1)} \sum_{j=1}^{n} \sum_{i=j}^{n} d(c_i, c_j)$$

Degree of homogeneity: The problematic addressed in this study is indeed the construction of the groups with the aim of the collective construction of knowledge in collaborative learning. A question can be asked in this step: what types of groups we will form most homogeneous or most heterogeneous?.

The problem is not new, it still divides the teaching staff as well as the scientific community. Differences in a group not only increase the motivation of the learners but also allow them to interact with each other by creating an exchange atmosphere that allows them to benefit from the experiences and knowledge of their colleagues to build their own knowledge. However, a very great heterogeneity at the group level can cause a blocking because there would be no listening between the learners of the same group.

PTo take advantage of the rich heterogeneity of groups while avoiding blockage we must reach a compromise by forming sufficiently homogeneous groups but not enough. For this reason, we have defined a function that represents the degree of heterogeneity defined for each cluster by:

$$h : c \to [0,1]$$
$$c_i \to h(c_i)$$

Thus, $h(c_i) = 1$ means that the cluster c_i is very hetero-geneous and $h(c_i) = 0$ means that it is very homogeneous.

Learner's clustering: In the context of collaborative learning, groups consist of a limited number of learners (N-Max) where the herogenity of a group does not exceed a given (H-Threshold).

Thus, we propose the LeCA algorithm (Learner Clustering Algorithm). It is a hierarchical algorithm that proceeds by successive divisions as long as the number of learners per group (N-MAX) is not reached and/or the threshold of heterogeneity (H-Threshold) is not exceeded. To divide subgroups, LeCA uses the k-means algorithm.

The choice of a hierarchical algorithm using k-means is justified by their simplicity their efficiency and also by the fact that, we can not apply clustering by partitioning which requires prior knowledge of the number of classes.

In the following, we are going to present our algorithm. To do this, we have defined firstly, a division function that receives as input a set of profiles c which will be divided into two subsets c_1 and c_2 by adopting the principle of k-means that is in 3 stages.

Algorithm 2; Learner's clustering algorithm:
Step 1: Selecting an initial partition of two sets c_1 and c_2
Step 2: Calculating the distance of different profiles to the sets c_1 and c_2 and then assigning them to the nearest one
Step 3: Repeating step (2) until the value of $J(c_1)+J(c_2)$ becomes stable

When the division function is defined, it only remains to define the algorithm LeCA. The latter receives as input a set of c profiles and returns the subsets for which the degree of homogeneity is less than H-threshold and the number of profiles is less than N-AMax. This is a recursive algorithm and consists of 2 steps:

Algorithm 3; Recursive algorithm:
Step 1; Divide the set c into two subsets c_1 and c_2 if the division conditions are satisfied
Step 2; Apply the algorithm LeCA on each one of the subsets c_1 and c_2

RESULTS AND DISCUSSION

Simulation and evaluation: To test our approach, we have developed a tool that allows us to do it on two levels. First, we were able to present an example of use and the results obtained. Second, to compare the results provided by our approach with those provided by teachers to whom we have asked to organize certain randomly generated profiles into subgroups according to criteria of their choice.

In order to evaluate our approach, we generated different sets of user profiles then we asked ten experienced teachers in the field of teaching to propose configurations (coefficients, maximum number of attributes per group) that they find reasonable.

Based on this configuration, we used our algorithm on one hand and on the other hand we asked teachers to group these profiles according to the configurations they proposed at the beginning.

From the results obtained by the professors on one side and those returned by our tool on the other side we calculated the difference in terms of the number of groups obtained in each case. This is represented by a difference percentage calculated by the following equation:
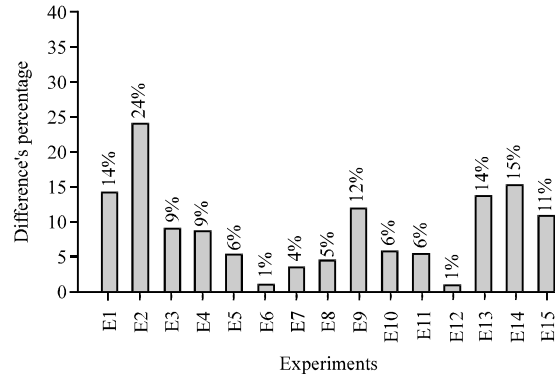
$$\frac{|N_1 - N_2|}{N_1 + N_2}$$

Fig. 4: The number of the obtained groups; Difference in the group number
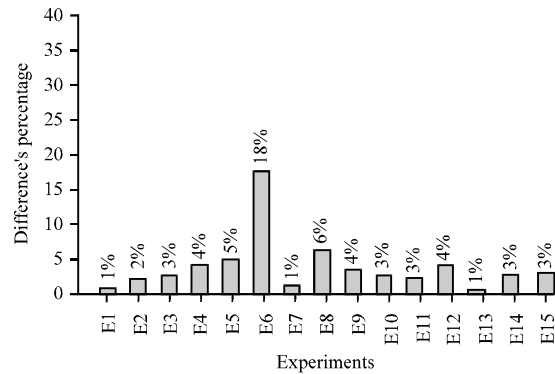

Fig. 5: LeCA vs. experimentation; Difference between group

Where:
N_1 = The number of groups given by teachers
N_2 = The number of groups given by our tool

Thus, we have plotted the graph shown in the Fig. 4. As we can see, the difference between the number of groups returned by the simulation tool of our approach and that of the groups created by the teachers is at worst 24%.

In a second step, we calculated the differences between the groups obtained by our approach. To do this we calcu-lated the percentage of displaced learners (i.e., the number of learners who are not assigned to the same group by our algorithm on the one hand and by teachers on the other hand). So, we have plotted the diagram shown in Fig. 5.

As can be seen, the percent difference between the groups formed by our algorithm and those formed by the teachers is still below 6% with the exception of the E6 experiment where it is 18%. This proves that we can adopt it as a solution for the formation of groups especially in a

context where the number of learners is very important that it becomes impossible for a teacher to do it manually.

CONCLUSION

Working in groups in an e-Learning context, provides an opportunity for monitoring a reduced number of learners. Also, it is a privileged moment of exchange and interaction with the trainer. It is especially an opportunity to adopt a social constructivist approach to teaching. To this end, we have proposed in this study a new approach to forming these groups. The proposed methodology is indeed divided into three stages. In the first, we describe the learner profile. Each profile is represented by attributes vector which are associated weight expressing degree of importance each coefficient.

In a second step, we define for each pair of profiles a distance in the form of degree of heterogeneity. These measurements were used to calculate the distance between a profile and a group. Thus, we begin to build groups. To do this, we have proposed an algorithm we called LeCA (Learner's Clustering Algorithm). This is a hierarchical clustering algorithm which adopts the divisive approach. In other-words, it starts with a set of profiles to partition. This set is divided into two groups based on the principle of k-means. We start this operation on each of the obtained groups until at achieving a number of attributes desired in each group or a tolerable heterogeneity is achieved.

We then proposed an algorithm called Learner's Clustering Algorithm (LeCA). It is a recursive hierarchical classification algorithm which adopts the divisive approach. In other words, it starts with a group of learners to partition. This group is divided into 2 subgroups based on the principle of k-means. We will repeat this operation on each of the groups obtained until we reach a desired number of learners or when a tolerable heterogeneity is reached.

REFERENCES

Ball, G.H. and D.J. Hall, 1967. A clustering technique for summarizing multivariate data. Syst. Res. Behav. Sci., 12: 153-155.

Chaiklin, S., 2003. The Zone of Proximal Development in Vygotsky's Analysis of Learning and Instruction. In: Vygotsky's Educational Theory in Cultural Context, Kozulin, A., B. Gindis, V.S. Ageyev and S.M. Miller (Eds.). Cambridge University Press, Cambridge, England, UK., pp: 39-64.

Dagger, D., V. Wade and O. Conlan, 2002. Towards a Standards-Based Approach to e-Learning Personalization using Reusable Learning Objects. In: E-Learn: World Conference on e-Learning in Corporate, Government, Healthcare and Higher Education, Driscoll, M. and T. Reeves (Eds.). Association for the Advancement of Computing in Education (AACE), Montreal, Canada, ISBN:978-1-880094-46-4, pp: 210-217.

Depover, C., J.J. Quintin and B. De Lievre, 2004. [Analysis of the effects of two types of group formation in a hybrid distance learning device (In French)]. Rev. Int. Technol. Pedagogie Universitaire, 1: 38-44.

Dolog, P., N. Henze, W. Nejdl and M. Sintek, 2004. Personalization in distributed E-learning environments. Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers and Posters, May 19-21, 2004, ACM, New York, USA., ISBN:1-58113-912-8, pp: 170-179.

Forgey, E., 1965. Cluster Analysis of multivariate data: Efficiency versus interpretability of classification. Biometrics, 21: 768-780.

Gower, J.C., 1971. A general coefficient of similarity and some of its properties. Biometrics, 27: 857-871.

Harland, T., 2003. Vygotsky's zone of proximal development and problem-based learning: Linking a theoretical concept with practice through action research. Teach. Higher Educ., 8: 263-272.

Hartigan, J.A., 1975. Algorithm CHAID: Clustering Algorithms. John Wiley, New York.

Jain, A.K. and R.C. Dubes, 1988. Algorithms for Clustering Data. Prentice-Hall Inc., Englewood Cliffs, NJ., USA.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Symp. Math. Statist. Prob., 1: 281-297.

Mendelsohn, P. and P. Dillenbourg, 1993. [The Development of Intelligent Computer-Assisted Education]. In: Intelligence Naturelle, Intelligence Artificielle: Symposium de l'Association de Psychologie Scientifique de Langue Française (Rome, 1991), Ny, J.F.L. (Ed.). Presses Universitaires de France, Paris, France, ISBN:9782130452713, pp: 231-256 (In French).

Ng, A.Y., M.I. Jordan and Y. Weiss, 2002. On spectral clustering: Analysis and an algorithm. Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, December 03-08, 2001, Vancouver, British Columbia, Canada, pp: 849-856.

Py, D., 1998. [Some artificial intelligence methods for student modeling (In French)]. Sci. Tech. Educatives, 5: 1-20.

Romainville, M., 2000. [Failure in the Mass University]. L'Harmattan, Paris, France, (In French).

Self, J.A., 1990. Bypassing the Intractable Problem of Student Modelling. In: Intelligent Tutoring Systems: At the Crossroads of Artificial Intelligence and Education, Frasson, C. and G. Gauthier (Eds.). Ablex Publishing, New York, USA., pp: 107-123.

Steinbach, M., G. Karypis and V. Kumar, 2000. A comparison of document clustering techniques. KDD. Workshop Text Min., 400: 525-526.

Zhou, L., H. Chen and J. Liu, 2013. Continuous ordered weighted distance measure and its application to multiple attribute group decision making. Group Decis. Negotiation, 22: 739-758.