An Unsupervised Method for Discovering How Does Learners' Progress toward Understanding in MOOCs

Ali El mezouary, Brahim Hmedna, Omar Baz

Abstract: Massive Open Online Course (MOOC) seems to expand access to education and it present too many advantages as: democratization of learning, openness to all and accessibility on a large scale, etc. However, this new phenomenon of open learning suffers from the lack of personalization; it is not easy to identify learners' characteristics because their heterogeneous masse. Following the increasing adoption of learning styles as personalization criteria, it is possible to make learning process easier for learners. In this paper, we extracted features from learners' traces when they interact with the MOOC platform in order to identify learning styles in an automatic way. For this purpose, we adopted the Felder-Silverman Learning Style Model (FSLSM) and used an unsupervised clustering method. Finally, this solution was implemented to clustered learners based on their level of preference for the sequential/global dimension of FSLSM. Results indicated that, first: k-means is the best performing algorithm when it comes to the identification of learning styles; second: the majority of learners show strong and moderate sequential learning style preferences.

Keywords: MOOC; learning styles; FSLSM; sequential/global learning styles automatic detection; clustering.

I. INTRODUCTION

Nowadays, MOOCs environments constitute one of the most used platforms in the field of e-learning. This new form of online teaching and learning has encouraged a large number of learners around the world to register for free in several courses with different themes [1]. MOOCs were categorized by Downes into two main types based on different learning theories: networks of distributed online resources (cMOOCs) such as found in the MOOC entitled "Connectivism and Connective Knowledge" (CCK08) and structured learning pathways centralized on digital platforms (xMOOCs) such as Coursera, edX and Udacity [1], [2]. One of the main challenges of MOOCs is the extremely low completion rate, which is mostly below 13 %, typically ranging from 2% to 10% for a course [3]–[5]. However, some MOOCs are nowadays well designed and effectively operated by outstanding teaching teams [6]. MOOC environments capture and store large data sets from learners' activities that can provide insight into the learning processes [7]. The MOOC learning activities include viewing a video, reading or posting on the forum and undertaking assignments. MOOC learners are heterogeneous in terms of their skills, background knowledge and preferences. Indeed, the data collected can help us understand how learners learn. Therefore, instead of presenting the same learning content to all learners, it is possible to design systems that are able to provide learners with learning content that is more adapted to their learning style [8]. For example, some learners prefer text to learn (verbal learners), while others prefer pictures or videos (visual learners) [9]. Multiple learning styles models have been proposed in the literature [10]. In particular, the Felder–Silverman Learning Style Model (FSLSM) is well recognized in engineering education [11]. The FSLSM performs a four-dimension classification: active/reflective, sensing/intuitive, verbal/visual and sequential/global. Several studies have emphasized the importance of adapting content to learners' learning styles in order to increase satisfaction, improve academic performance (effectiveness) and save time (efficiency) [12]. In addition, teachers can benefit from understanding learners' learning styles because they can provide more effective interventions [13]. In order to identify learners' learning styles, many systems require them to fill out a questionnaire (44 questions), which is inappropriate, because when there are too many questions or the learner does not know the importance or future use of questionnaire, the learner will choose at will answer. To overcome this problem, various automated methods that can identify learners' learning styles based on their behavior in the learning environments have been developed [14]. By using these automatic approaches, it is possible to track changes in learners' learning styles during the learning process. Moreover, teachers finally have a chance to personalize their instruction, and learners can learn in a way that matters to them [15], [16]. The study presented in this paper, consists in proposing an automatic approach to identify learners' learning styles for the sequential/global dimension of FSLSM [17]. To achieve this goal, some characteristics of MOOCs need to be taken into account:

- The heterogeneity of MOOC learners is high due to their large-scale population (massiveness) [18], [19]. This feature is a big challenge in identifying learning styles; therefore, it raises a serious question: how to cluster large heterogeneous learner profiles according to their learning styles?

Manuscript received on February 09, 2021.
Revised Manuscript received on March 03, 2021.
Manuscript published on March 30, 2021.
* Correspondence Author
Ali El mezouary*, IRF-SIC Laboratory, Faculty of Science, Ibn Zohr University, Agadir-Morocco. Email: a.elmezouary@uiz.ac.ma
Brahim Hmedna, IRF-SIC Laboratory, Faculty of Science, Ibn Zohr University, Agadir-Morocco. Email: brahim.hmedna@edu.uiz.ac.ma
Omar Baz, Department, IRF-SIC Laboratory, Faculty of Science his/her department, Ibn Zohr University, Agadir-Morocco. Email: o.baz@uiz.ac.ma

- The openness of MOOC is one of the major factors that impacts learning in that these environments give learners the freedom to choose their learning objectives and define which activities are most suitable for them [20].

This openness may contribute significantly to revealing learners' learning styles by allowing them to interact with the activities they prefer.

In our study, in order to justify the choice of the quality k-means algorithm, we compared the clustering results obtained by this algorithm with those of three other algorithms (K-means, MiniBatch, Birch and Agglomerative). The traces analyzed are from the edX course entitled "Statistical Learning (Stat, Winter 2015)", from a Stanford University MOOC.

This study tries to answer the following research questions: RQ1: how can learners' learning styles be automatically identified in the MOOC platform? RQ2: what features are most important for this identification? RQ3: how can we cluster learner profiles?

II. BACKGROUND

In this section, we first give an overview of learning styles and introduce the most commonly used model to identify them. Then, we describe different approaches that can be used to identify learners' learning styles.

A. Learning Styles Concepts and Models

Learning style reflects the aspect of difference between learners in terms of the type of instruction that is most appropriate for them [21]. For example, some learners with logical dominant intelligence learn better when learning materials are presented through numbers, reasoning of concepts [9]. Learning style is a holistic model that provides a wide range of learning directions and makes the same instructional method beloved by some learners and hated by others [22]. Learning style is a category under a broader umbrella of learner characteristics; it has been utilized in psychology and pedagogy since 1930. Therefore, it does not have a single definition. According to [23], learning style is the preferred way of using one's ability to learn. For Felder, learning style is a concept that reflects "the desirable way in which the learner receives and processes information".

Different models of learning styles have been proposed by several psychologists. In this sense, we can distinguish: Felder and Silverman model [11], Dunn and Dunn model [25], VAK/VARK model [26], Kolb's model [27] and Honey and Mumford's model [28]. According to the classification model proposed by Coffield et al. (2004) named families of learning styles, the authors identify 71 learning style models and categorize them into five families: (i) Constitutionally-based, including visual, auditory, kinesthetic, tactile. (ii) Cognitive structure, including patterns of ability. (iii) A relatively stable personality type. (iv) Flexibly stable learning preferences. (v) Learning approaches and strategies.

A learning style model is comprised of a number of dimensions (ordering, processing, transforming …). Each dimension is a composition of antagonist learning styles. For example, the processing dimension has two learning styles: active and reflective. The ordering dimension has two learning styles: sequential and random. Each learning style has its own unique characteristics, where learners prefer to interact with specific resources and favor performing certain specific activities. Learning style is a part of the learner model. In addition, it contains other characteristics or parameters such as prior knowledge, skills and cognitive style. Felder and Silverman's learning style model (FSLSM) was developed to help teachers in their teaching methodology to engineering learners [11]. According to this model, learners are classified into four dimensions: (i) *The perception dimension*: Sensing learners tend to be concrete, practical, methodical and oriented toward facts and hands-on procedures. Intuitive learners are more comfortable with abstractions and are oriented toward theories and underlying meaning. (ii) *The input dimension*: Visual learners prefer materials such as graphs, charts or videos, while verbal learners prefer words either written or spoken. (iii) *The processing dimension*: Active learners prefer to learn by doing, experimentation and collaboration, while reflective learners prefer to think and absorb the information alone or in small groups. (iv) *The understanding dimension*: Sequential learners prefer information to be provided in a linear (serial) fashion and tend to make small steps through learning material. Global learners tend to make larger leaps and tend to require seeing the "big picture" before understanding the topic. The Felder and Silverman Index is one of the most widely used methods of distinguishing learning styles of learners. This index consists of 44 questions (11 questions for each dimension). Each question contains two exclusive options (a & b), so as to distinguish learner preferences for each dimension in the form of values between +11 and -11. In order to distinguish at a more precise level the learners' preferences, three categories of preferences are distinguished for each dimension, as shown in "Fig 1" [29]. The category "Balanced preference" (score is between 3 and -3). The category "Moderate preference" (score is between -5 and -7, or between 5 and 7). The "Strong preference" category (score is between 9 and 11, or between -9 and -11).

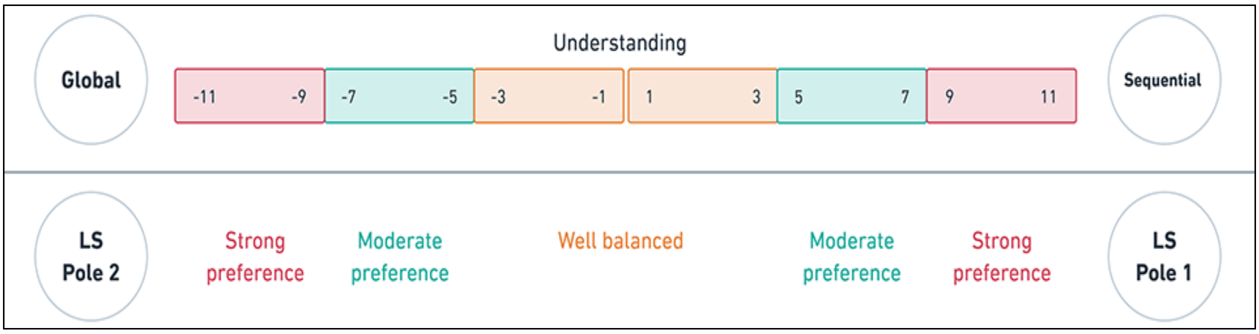

Fig. 1.Scales of learning style dimension (understanding dimension)

For our study, we chose to use the FSLSM because several works that we have studied in this sense, have proven the validity and reliability of this model in assessing learning styles [30]. Among other reasons that we identified and that allowed us to justify this choice we can cite the following: i) the four dimensions of this model are distinct and independent [31]. ii) The FSLSM allows to categorize in a granular way the learning styles, representing each dimension on a scale from -11 to +11 [32].

B. Approaches of Detecting Learning Styles

The process of identifying learning styles is illustrated in “Fig. 2”. There are two types of approaches for learning style detection, namely collaborative and automatic approaches [33]. The principle of the first approaches is to present learners with a questionnaire to complete. In principle, this questionnaire should reflect one of the recommended learning style models. In order to overcome the above-mentioned drawbacks faced by using such questionnaire, researchers have proposed several artificial intelligence techniques to detect automatically learning styles of learners. These approaches consist in collecting and analyzing the learners' traces during different interactions between them and the system.

For automatic approaches we distinguish between literature-based and data-based approaches.

- The literature-based approach uses the learners’ traces to get hints about their learning styles. By applying a set of predefined rules from the literature on learners' interactions with the system, a model that computes learning styles is built [34]. The main advantage is that, the constructed model can be exploited and applied on the data collected in any other course.
- The Data-driven approaches involve building a model that mimics the learning styles questionnaire. It constructs a model by using learners’ behaviours to feed the artificial intelligence classification algorithm such as: (i) Bayesian technique [35]–[37]. (ii) Neural network [38]–[41]. (iii) Decision tree method [42]–[44]. (iv) Naïve Bayes [40], [45]. (v) Reinforcement learning [46], [47]. (vi) Markov model [48], [49]. The main advantage of these methods is that they use live data to classify learners, which makes them very accurate and allows the system to track and update learners' learning styles.

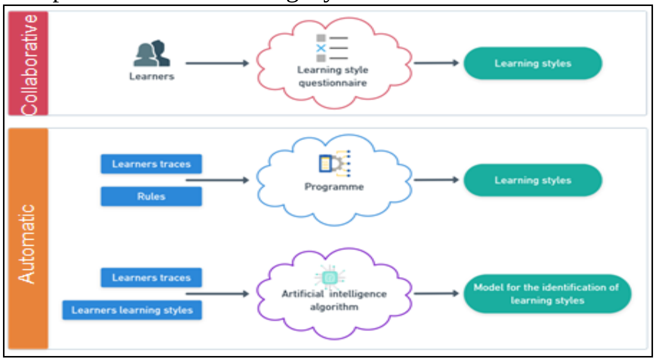

Fig. 2.Approaches of detecting Learning Styles

III. METHOD

In this section, we first present the MOOC dataset on which the approach has been established. Second, we describe the details of features extracted from the MOOC traces. Finally, the clustering implementation is provided.

The architecture shows in “**Error! Reference source not found.**”, represent the overall process of our approach for automatic identification of learners' learning styles. This architecture implements several phases.

- For each learning style of the understanding dimension of the FSLSM, after the extraction of learner traces from learner logs, we preprocessed it, and then we extracted features from learner's traces that will help us to predict their learning style. The feature vector of each learner was passed as input to the unsupervised clustering method. Finally, the learners are distributed according to their degree of preference on four clusters (Very Weak, Weak, Moderate and Hight).

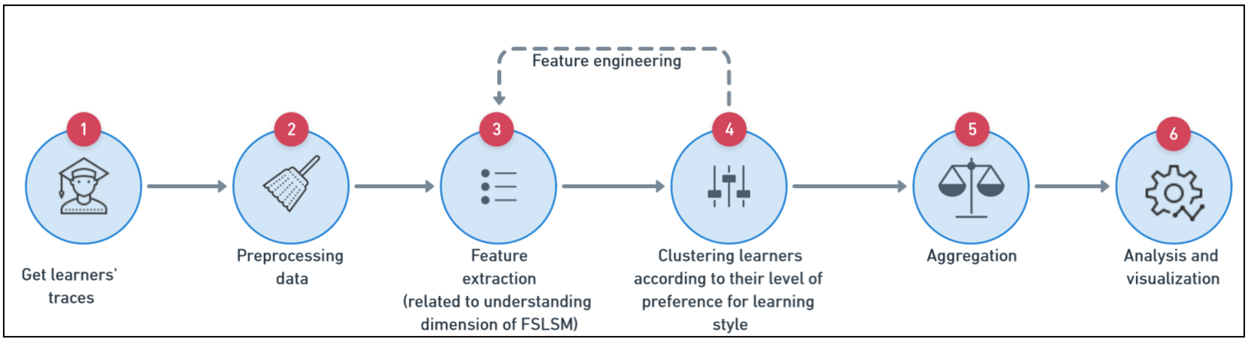

Fig. 3.Architecture of Proposed Methodology

A. Requirements for Implementation

K-means algorithms can be implemented using *Scikit-learn*[1], which is a Python module for machine learning built on top of *SciPy*. To install *Scikit-learn* we can use one command-line in Python: *'pip install scikit-learn'* or download and install *Anaconda*[2] Python distribution, which we use in this case. In addition to *Scikit-learn*, we also need the following libraries: (i) *NumPy*[3]: is the primary package for scientific computing through Python. (ii) *Matplotlib*[4]: is a very versatile library for producing plots and other two-dimensional data visualizations. (iv) *Pandas*[5]: provides high-level data structures and functions designed to make working with structured or tabular data fast, easy, and expressive.

In this study, we used Pandas instead of the *Hadoop*[6] ecosystem for two reasons: (1) the size of our dataset is 6 gigabytes; hence, our dataset can be stored easily in memory (2) there is no need for a cluster.

B. Dataset

In our study we analyzed a dataset that we retrieved from the Center for Advanced Research through Online Learning (CAROL[7]). Specifically, data collected from the course "Statistical Learning (Stat, Winter 2015)". We note that this course was distributed over 9 weeks. During this period, 32,209 learners registered and generated approximately 18,475,724 events. Examples of event include viewing video lectures, attempting graded quizzes and homework assignments, and participating in course forum discussions.

The clickstream events were generated whenever a learner interacted with the MOOC platform. Each event was described by a set of attributes such as interaction types (*event_type*), date and time of interaction (*timestamp_event*), and the learner identifier (*anon_screen_name*). The Dataset was anonymized for the purpose of privacy protection of learners.

C. Data Pre-Processing

Given the raw data offered by CAROL, we first extract traces generated between 19/01/2015 and 06/04/2015, and perform a data cleaning process to improve the quality of the data, viz removing some columns that contain redundant information. After that, we associated each event with the week in which it was executed.

Based on the "*Resource_display_name*" feature, which contains names of the video, assignment, or teaching module associated with the action, we generated a new feature "*Resource_format_name*" by mapping every resource with its format.

In the following subsection, we will briefly explain how learners' features related to each learning style are extracted, normalized and reduced.

1) Feature Selection

Our goal in this step is to be able to represent each learner by a vector of learning style characteristics. This process is a difficult task, requiring multidisciplinary knowledge. It involves preparing the features to feed into the learning algorithms. The recursive feature elimination method is one of the most used methods to select the most effective features. This method consists in progressively eliminating the less discriminating features [50].

For this purpose, our feature sectioning process is based: First, on a comparative study we conducted to help us identify what learners with different learning styles prefer when using an e-learning system. For this, we chose four relevant research works [29], [39], [51], [52], as mentioned in Table 1. Second, on the hypotheses we developed about learners' characteristics regarding the comprehension dimension of FSLSM.

This dimension is used to categorize learners according to their progress in the learning sequence (sequential learning vs. global learning). As shown in Table 2, the navigational mode of learners can reveal hints about their learning style. Sequential learners learn each concept in a linear (step-by-step) manner; they often use the "Next" and "Back" navigation buttons to move between units of the course. They also view the answers to their quizzes in more detail. Global learners, on the other hand, try to have a more global view of the concept by learning in a holistic way. They frequently access to the course presentation page (outlines), so they tend to go directly to a specific unit (seq_goto) or a specific part of the video (seek_video).

1 http://scikit-learn.org
2 https://www.anaconda.com
3 http://www.numpy.org
4 https://matplotlib.org
5 http://pandas.pydata.org
6 http://hadoop.apache.org
7 https://datastage.stanford.edu

Retrieval Number: 100.1/ijitee.E86730310521
DOI: 10.35940/ijitee.E8673.0310521

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

Table- I: Comparison of understanding features related to some data-driven study

		Features	Study			
			Villaverde	**Latham**	**Garcia**	**Graf**
Understanding dimension	**Sequential**	Ques_detail				*
		Choose to be guided through the steps of solving a problem		*		
		Navigation_step_by_step	*		*	
		Exam_res_high			*	
	Global	Outline_visit				*
		Outline_stay				*
		Ques_overview				*
		Ques_interpret				*
		Ques_develop				*
		Navigation_skip	*		*	*
		Navigation_overview_visit				*
		Navigation_overview_stay				*
		Choose to solve a problem straight away		*		

Table- II: Relevant features for understanding dimension

	LS	Features	Resources	Description
Understanding dimension	**Sequential**	# seq_next	Navigation	Number of times the learner navigated to the next unit in a sequence
		# seq_prev	Navigation	Number of times the learner navigated to the previous unit in a sequence
		Δ sequential_navigation	Navigation	= (# seq_next + # seq_prev) / (# seq_next + # seq_prev + # seq_goto)
		# dist_unit_visit	Navigation	Number of distinct units (pages) visited
		# show_answer	Quiz	Number of visits to the answers to problems
		# page_close	Navigation	Number of pages closed
	Global	# progress_show	Navigation	Number of visits to progress page
		# seek_video	Video	Number of seek in videos
		# seq_goto	Navigation	Number of times the learner skipped to specific units
		Δ global_navigation	Navigation	= # seq_goto / (# seq_next + # seq_prev + # seq_goto)
		# outline_visit	Outlines	Number of visits to outlines

2) Feature Normalization

The performance of many machine-learning algorithms is sensitive to the scales of features [53]. If we do not maintain a uniform distribution of the value of each feature, some of them will have a significant impact on the output model. Therefore, it is necessary to normalize the measurement units of features before starting the modeling process. To carry out this normalization, we used MinMax method, which consists in bringing all the values in the interval [0 1], by setting the minimum to 0 and the maximum to 1. Thus, the normalized value of x is formulated as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

D. Clustering Analysis

1) k-Means Clustering Algorithm

The clustering purpose is to identify groups or clusters of learners showing similar learning styles patterns. For it, we must respect the two following criteria: (1) minimize the distance between elements of the same class (intra-class distance, similarity) and (2) maximize the distance between clusters (inter-class distance, separability) [54]. Clustering analysis is useful in domains where the dataset is unlabeled because it allows finding the hidden structure in the data. In the literature, several works in the field of clustering have emerged, which have given rise to several clustering algorithms, namely: Density-based clustering, connectivity-based clustering, centroid-based clustering, etc.

In what follows, we will focus on centroid-based clustering: k-means, which uses euclidean distance as its similarity measure:

$$distance\,(x, y) = \left(\sqrt{\textstyle\sum_{i=1}^{d}(x_i - y_i)^2}\right)^2 \qquad (2)$$

Where x and y are d-dimensional feature vectors.

The purpose of this algorithm is to partition a set of n points $\{x_1, x_2, x_3, \ldots, x_n\}$ into k groups $\{c_1, c_2, c_3, \ldots c_k\}$ while maximizing the separation of the clusters and the compactness of the elements of the same cluster. K-means clustering algorithm is illustrated in "Fig.4":

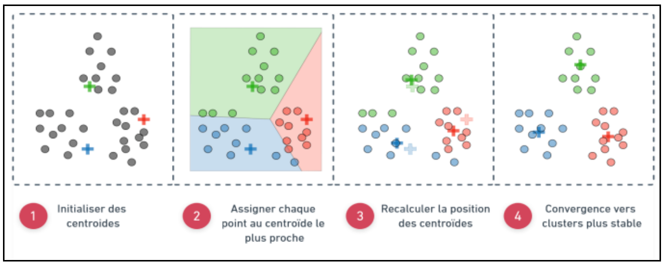

Fig. 4.Illustration of the K-Means Algorithm.

- (1) we choose K, the number of clusters, and randomly add K centroids to the feature space, because K-means is highly sensitive to the initial position of centroids [55], we applied the K-means++ algorithm.

- (2) We then compute the distance from each data point x to each centroid c using a given metric, such as the Euclidean distance in our case. Then, we assign the closest centroid to each data point.
- (3) For each centroid, we calculate the average feature vector of the data points labeled with it, and these average feature vectors become the new locations of the centroids.
- (4) We re-calculate the distance from each data point to each centroid, modify the assignment and repeat the procedure until the assignments are stabilized.

2) Estimation of the Optimal Number of Clusters

Many clustering algorithms, including k-means, require the specification of the optimal number of clusters (k) as an input parameter. Indeed, estimating this k-value represents a major challenge. Based on the solutions existing in the literature, to determine the number k, we have chosen to use the elbow method [56], which consists in tracing a graph with in x-axis the number of clusters k {1,2,3,4,5,...}, and in y-axis the sum of squared errors (SSE) [57]. Indeed, the best k estimated by this method is located at the point of the curve where an elbow is formed.

$$SSE = \frac{1}{n}\sum_{i=1}^{k}\sum_{x\in Cluster_i}\left|x - Centroid_i\right|^2 \qquad (3)$$

According to “Fig. 5 as an application of this method, on the characteristics of learners' preferences according to the comprehension dimension of the FSLSM, we found that the curve formed an elbow at the level of 4 clusters.

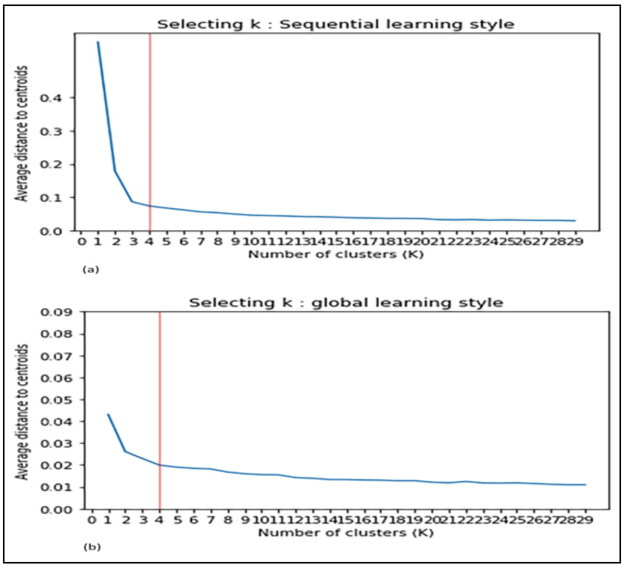

Fig. 5.Elbow method applied (a) on the features of sequential learning styles; and (b) on the features of global learning styles.

3) Clustering Results

Once the optimal value of k is estimated for both the sequential and global learning styles, the k-means algorithm was used to cluster the learners. Therefore, the learners are divided into 4 clusters according to their degree of preferences: Very Weak (C1); Weak (C2); Moderate (C3) and High (C4). Table 3 and Table 4 show the values of features according to sequential learning styles and global learning styles respectively

According to the results shown in the Table 3 and Table 4, whether for the sequential learning style or for the global learning style, learners with very low preferences constituted the largest group (54.4% and 85.4% learners). Indeed, the average values of all the very low preference cluster features for both styles are the lowest. This can be explained by the fact that a significant number of the learners were dropped out of this MOOC. While, the mean values of almost feature at the low preference cluster features are moderate for both learning styles. On the other hand, when comparing the Moderate and High preference clusters respectively for the two styles, we notice that, the sum of the percentages of learners from these two clusters in the sequential learning style (16.7%) is greater than that in the global learning style (7.7%).

4) Clustering Quality Evaluation

Once learners have been clustered according to their preferences for each learning style, the quality of this clustering should be evaluated, although in unsupervised learning it is difficult to evaluate the performance of a clustering model, especially when there are no reference labels.

There are two main evaluation strategies. Internal evaluation criteria can be used to evaluate clustering results in terms of the intended clustering properties (compactness and separability). External evaluation criteria, on the other hand, allow a cluster to be compared with a reference (expected result) - for example a class label.

As shows in Table 5, we compared the values of these indices against the clustering results of four different clustering algorithms, and we found that the values computed for these two indices that are related to the results returned by the k-means algorithm are better compared to the remaining values of the other three algorithms.

Table- III: Sequential learning style clusters

Clusters	C_1	C_2	C_3	C_4
Features (mean)				
# seq_next	0,00	6,75	28,74	84,01
# seq_prev	0,00	2,53	9,79	27,91
avg_sequential_navig	0,20	1,00	0,98	0,94
# dist_seq_visit	2,24	7,98	35,30	70,95
# showanswer	0,04	3,51	16,05	48,38
# page_close	4,62	24,99	96,74	242,83
# Learners	17508	9277	1974	3421
% Learners	54,4%	28,8%	6,1%	10,6%
Cluster labels	**Very Weak**	**Weak**	**Moderate**	**High**

Table- IV: Global learning style clusters

Clusters	C_1	C_2	C_3	C_4
Features (mean)				
# progress	1,24	14,56	25,85	56,02
# seek_video	7,38	88,69	108,55	235,93
# seq_goto	0,02	0,88	7,63	18,81
Avg_global_navig	0,00	0,01	0,08	0,09
# outline	4,53	30,23	25,60	58,89
# Learners	**27474**	**2223**	**1841**	**642**
% Learners	**85,4%**	**6,9%**	**5,7%**	**2,0%**
Cluster labels	**Very Weak**	**Weak**	**Moderate**	**High**

In our case, as there are no reference labels available, we simply use internal evaluation criteria such as the Silhouette index [58] and the Calinski-Harabasz index [59] to evaluate the quality of our clustering results and to justify the choice of the k-means algorithm. A brief definition of these two indices is as follows:

- **Silhouette index**: a composite index reflecting the compactness and separation of clusters; a larger average Silhouette index indicates a better overall quality of the clustering result.
- **Calinski-Harabasz index**: this index measures the degree of intra-cluster similarity and the degree of inter-cluster separation. Using this index, clustering results are of better quality when its value maximum.

E. Aggregation Process: Balance of Learning Styles

The aggregation phase is an essential step in determining each learner's degree of preference for understanding dimension of the FSLSM. The k-means clustering algorithm allowed us to categorize each pair (learner and learning style) by assigning it a label reflecting its level of preference for each learning style such as very weak, weak, moderate or high. The understanding dimension of FSLSM model is composed of two poles (sequential/global). These are bipolar and complementary. For example, a learner with a high preference for the sequential learning style simultaneously shows a weak preference for the global learning style. On the other hand, if the preferences are similar (high or weak), the learner has a balanced learning style ("Fig. 6"). Based on this balance of learning styles, we establish a grid of all combinations that quantifies the degree of dominance of each learning style, as shown in table 6. By merging the two vectors of learning styles features using the balance of learning style, we obtain a vector of features relative to the understanding dimension, as illustrated in "Fig. 7".

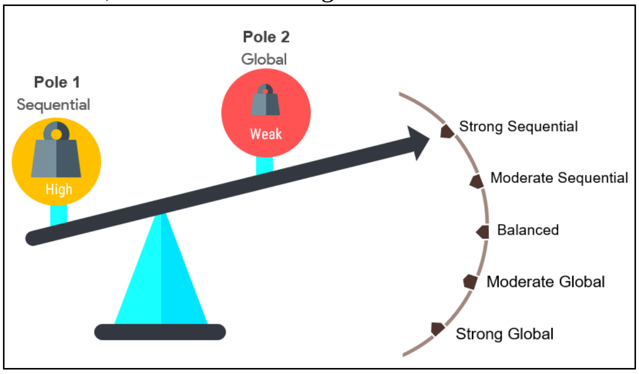

Fig. 6.Balance of learning styles

Table- V: Internal evaluation metrics for clustering

Validation index	Calinski-Harabasz (CH)				Silhouette (SI)			
Algorithms	K-means	MiniBatch	Birch	Agglomerative	K-means	MiniBatch	Birch	Agglomerative
Sequential	**239064**	239022	218663	232308	**0.81**	0.80	0.80	0.80
Global	**26101**	24182	18263	22722	**0.82**	0.74	0.80	0.80

Table- VI: Balance of learning styles

		Sequential LS			
		Very Weak (0)	**Weak (1)**	**Moderate (2)**	**High (3)**
Global LS	**Very Weak (0)**	Balanced	Moderate Sequential	Strong Sequential	Strong Sequential
	Weak (-1)	Moderate Global	Balanced	Moderate Sequential	Strong Sequential
	Moderate (-2)	Strong Global	Moderate Global	Balanced	Moderate Sequential
	High (-3)	Strong Global	Strong Global	Moderate Global	Balanced

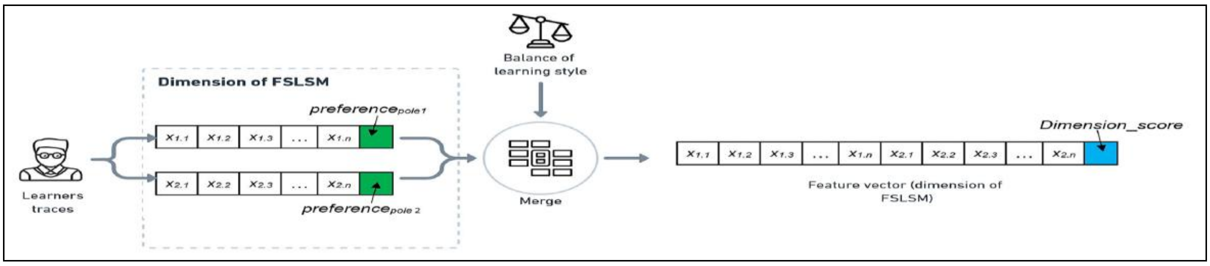

Fig. 7.Aggregation process

IV. RESULTS & DISCUSSION

For the features identified for the understanding dimension of FSLSM, we collected 18,475,724 traces from 32,209 learners taking "Statistical Learning" MOOC course. A series of data pre-processing were done to clean, wrangle and extract features from the learners' traces. The data was normalized into the interval [0 1].Then multiple clustering algorithms such as *K-means*, *MiniBatch*, *Birch* and *Agglomerative* were applied. We applied aggregation process in order to quantify the degree of dominance of each learning style. The distribution of learning styles in this dimension (understanding) is presented in figure 9.

From the "**Error! Reference source not found.**8", we can observe that the majority of learners in this study had a sequential learning style. These learners in general performed better than global learners [60]. This is consistent with the findings of similar studies [61] which showed that global learners are potentially at risk of dropping out. Therefore, more attention should be given to global learners in the learning process. [62] Believes that global learners can benefit from group discussions. In light of this finding, teachers need to create more pedagogical resources with a higher level of granularity.

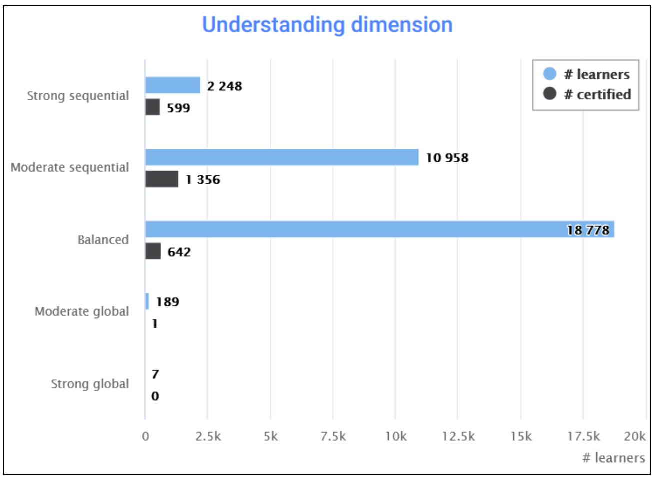

Fig. 8.Distribution of learning styles

V. CONCLUSION

In this paper, we have attempted to identify automatically the sequential/global learning styles of learners through their behaviors, while they are interacting with the MOOC platform. Recently, different machine learning and data mining techniques have been applied in this context. The study we conducted proposes an unsupervised clustering technique to discover how does learners' progress toward understanding in MOOCs. The traces of 32,209 learners enrolled in one of Stanford University's MOOCs in 2015 were analyzed and used to test and validate our approach. The results of this study show that the majority of learners who are enrolled in this MOOC possess sequential learning style preferences.

REFERENCES

1. S. Downes, « Places to go: Connectivism & connective knowledge », Innovate: Journal of Online Education, vol. 5, no 1, p. 6, 2008.
2. M. Cisel et É. Bruillard, « Chronique des MOOC. (French) », Sticef, vol. 19, p. 1-16, 2012.
3. W. Xing et D. Du, « Dropout prediction in MOOCs: Using deep learning for personalized intervention », Journal of Educational Computing Research, vol. 57, no 3, p. 547-570, 2019.
4. T. R. Liyanagunawardena, P. Parslow, et S. Williams, « Dropout: MOOC participants' perspective », 2014.
5. M. Kloft, F. Stiehler, Z. Zheng, et N. Pinkwart, « Predicting MOOC dropout over weeks using machine learning methods », in Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs, 2014, p. 60-65.
6. J. Reich, « MOOC completion and retention in the context of student intent », EDUCAUSE Review Online, 2014.
7. L. S. Settouti, Y. Prié, A. Mille, et J.-C. Marty, « Systèmes à base de traces pour l'apprentissage humain », Communication in the international TICE, Technologies de l'Information et de la Communication dans l'Enseignement Supérieur et l'Entreprise, 2006.
8. X. Tsortanidou, C. Karagiannidis, et A. Koumpis, « Role and Value of Learning Theories in the Area of Adaptive Educational Hypermedia Systems Incorporating Learning Styles », International Journal of Emerging Technologies in Learning (iJET), vol. 13, no 1, p. 93-108, 2018.
9. H. Pashler, M. McDaniel, D. Rohrer, et R. Bjork, « Learning Styles: Concepts and Evidence », Psychological Science in the Public Interest, vol. 9, no 3, p. 105-119, déc. 2008, doi: 10.1111/j.1539-6053.2009.01038.x.
10. F. Coffield, D. Moseley, E. Hall, K. Ecclestone, et others, Learning styles and pedagogy in post-16 learning: A systematic and critical review. Learning and Skills Research Centre London, 2004.
11. R. M. Felder et L. K. Silverman, « Learning and teaching styles in engineering education », Engineering education, vol. 78, no 7, p. 674-681, 1988
12. S. Graf et K. Kinshuk, « Providing adaptive courses in learning management systems with respect to learning styles », in E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 2007, p. 2576-2583.
13. W. L. H. Wong, « A Study of Language Learning Style and Teaching Style Preferences of Hong Kong Community College Students and Teachers in English for Academic Purposes (EAP) Contexts », 2015, Consulté le: mars 29, 2017. [En ligne]. Disponible sur: http://ir.canterbury.ac.nz/handle/10092/11661.
14. S. Graf, T.-C. Liu, et others, « Identifying learning styles in learning management systems by using indications from students' behaviour », in Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on, 2008, p. 482-486, Consulté le: avr. 01, 2017. [En ligne]. Disponible sur: http://ieeexplore.ieee.org/abstract/document/4561743/.
15. T. J. Foriska, « The Principal as Instructional Leader: Teaming with Teachers for Student Success. », Schools in the Middle, vol. 3, no 3, p. 31-34, 1994.
16. M. Zajac, « Using learning styles to personalize online learning », Campus-wide information systems, vol. 26, no 3, p. 256-265, 2009.
17. S. Graf et T.-C. Liu, « Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach », Journal of Educational Technology & Society, vol. 12, no 4, p. 3, 2009.
18. I. Chuang et A. D. Ho, « HarvardX and MITx: Four Years of Open Online Courses–Fall 2012-Summer 2016 », 2016.
19. D. Koller, A. Ng, C. Do, et Z. Chen, « Retention and intention in massive open online courses: In depth », Educause review, vol. 48, no 3, p. 62-63, 2013.
20. A. Bakki, L. Oubahssi, S. George, et C. Cherkaoui, « A Model to Assist Pedagogical Scenario Building Process in cMOOCs », in Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on, 2017, p. 5-7.
21. J. Li, S. Han, et S. Fu, « Exploring the relationship between students' learning styles and learning outcome in engineering laboratory education », Journal of Further and Higher Education, vol. 43, no 8, p. 1064-1078, sept. 2019, doi: 10.1080/0309877X.2018.1449818.
22. R. Oxford, « Language learning styles and strategies: Concepts and relationships », IRAL - International Review of Applied Linguistics in Language Teaching, vol. 41, no 4, p. 271-278, 2003, doi: 10.1515/iral.2003.012.

23. R. J. Sternberg, « Allowing for Thinking Styles. », Educational leadership, vol. 52, no 3, p. 36-40, 1994.
24. R. M. Felder, « Matters of style », ASEE prism, vol. 6, no 4, p. 18-23, 1996.
25. R. S. Dunn, K. J. Dunn, et G. E. Price, Learning style inventory. Price Systems Lawrence, KS, 1979.
26. M. T. C. HELENA, « Styles of Learning Based on the Research of Fernald, Keller, Orton, Gillingham, Stillman, Montessori and Neil D Fleming », 2017.
27. D. Kolb, Experiential Learning: turning experience into learning. New Jersey: Prentice Hall, 1984.
28. P. Honey et A. Mumford, Using your learning styles. Chartered Institute of Personnel and Development, 1986.
29. S. Graf, Kinshuk, Q. Zhang, P. Maguire, et V. Shtern, « Facilitating Learning Through Dynamic Student Modelling of Learning Styles », in Towards Learning and Instruction in Web 3.0, Springer, New York, NY, 2012, p. 3-16.
30. T. A. Litzinger, S. H. Lee, et J. C. Wise, « A study of the reliability and validity of the Felder-Soloman Index of Learning Styles », in Proceedings of the 2005 American Society for Education Annual Conference & Exposition, 2005, p. 1-16.
31. P. E. Ramírez-Correa, F. J. Rondan-Cataluña, J. Arenas-Gaitán, et J. L. Alfaro-Perez, « Moderating effect of learning styles on a learning management system's success », Telematics and Informatics, vol. 34, no 1, p. 272-286, 2017.
32. R. M. Felder et R. Brent, « Understanding student differences », Journal of engineering education, vol. 94, no 1, p. 57-72, 2005.
33. P. Brusilovsky, « Methods and techniques of adaptive hypermedia », User Model User-Adap Inter, vol. 6, no 2-3, p. 87-129, juill. 1996, doi: 10.1007/BF00143964.
34. J. Feldman, A. Monteserin, et A. Amandi, « Automatic detection of learning styles: state of the art », Artificial Intelligence Review, vol. 44, no 2, p. 157-186, 2015.
35. B. G. Aslan, Ö. Öztürk, et M. M. İnceoğlu, « Effect of bayesian student modeling on academic achievement in foreign language teaching (university level english preparatory school example) », 2014.
36. C. Carmona, G. Castillo, et E. Millán, « Designing a dynamic bayesian network for modeling students' learning styles », in Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on, 2008, p. 346-350.
37. P. García, A. Amandi, S. Schiaffino, et M. Campo, « Evaluating Bayesian networks' precision for detecting students' learning styles », Computers & Education, vol. 49, no 3, p. 794-808, 2007.
38. J. Bernard, T.-W. Chang, E. Popescu, et S. Graf, « Learning style Identifier: Improving the precision of learning style identification through computational intelligence algorithms », Expert Systems with Applications, vol. 75, p. 94-108, 2017.
39. J. E. Villaverde, D. Godoy, et A. Amandi, « Learning styles' recognition in e-learning environments with feed-forward neural networks », Journal of Computer Assisted Learning, vol. 22, no 3, p. 197-206, juin 2006, doi: 10.1111/j.1365-2729.2006.00169.x.
40. F. Rasheed et A. Wahid, « Learning Style Detection in E-learning Systems Using Machine Learning Techniques », Expert Systems with Applications, p. 114774, févr. 2021, doi: 10.1016/j.eswa.2021.114774.
41. R. Zatarain-Cabada, M. L. Barrón-Estrada, V. P. Angulo, A. J. García, et C. A. R. García, « Identification of felder-silverman learning styles with a supervised neural network », in Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence, Springer, 2010, p. 479-486.
42. « Learning style detection in E-learning systems using machine learning techniques », Expert Systems with Applications, vol. 174, p. 114774, juill. 2021, doi: 10.1016/j.eswa.2021.114774.
43. I. Karagiannis et M. Satratzemi, « An adaptive mechanism for Moodle based on automatic detection of learning styles », Education and Information Technologies, vol. 23, no 3, p. 1331-1357, 2018.
44. T. Sheeba et R. Krishnan, « Automatic Detection of Students Learning Style in Learning Management System », in Smart Technologies and Innovation for a Sustainable Future, Springer, 2019, p. 45-53.
45. J. Feldman, A. Monteserin, et A. Amandi, « Detecting students' perception style by using games », Computers & Education, vol. 71, p. 14-22, 2014.
46. M. T. C. HELENA, « Styles of learning based on the research of fernald, keller, orton, gillingham, stillman, montessori and neil d fleming », 2017.
47. F. A. Dorça, L. V. Lima, M. A. Fernandes, et C. R. Lopes, « Comparing strategies for modeling students learning styles through reinforcement learning in adaptive and intelligent educational systems: An experimental analysis », Expert Systems with Applications, vol. 40, no 6, p. 2092-2101, 2013.
48. H. Fasihuddin, G. Skinner, et R. Athauda, « Towards adaptive open learning environments: Evaluating the precision of identifying learning styles by tracking learners' behaviours », Education and Information Technologies, déc. 2015, doi: 10.1007/s10639-015-9458-5.
49. M. S. Hasibuan, L. E. Nugroho, P. I. Santosa, et S. S. Kusumawardani, « A Proposed Model for Detecting Learning Styles Based on Agent Learning », International Journal of Emerging Technologies in Learning (iJET), vol. 11, no 10, p. 65, oct. 2016, doi: 10.3991/ijet.v11i10.5781.
50. R. Kohavi et G. H. John, « Wrappers for feature subset selection », Artificial intelligence, vol. 97, no 1-2, p. 273-324, 1997.
51. A. Latham, K. Crockett, D. McLean, et B. Edmonds, « A conversational intelligent tutoring system to automatically predict learning styles », Computers & Education, vol. 59, no 1, p. 95-109, août 2012, doi: 10.1016/j.compedu.2011.11.001.
52. P. Garcia, A. Amandi, S. Schiaffino, et M. Campo, « Evaluating Bayesian networks' precision for detecting students' learning styles », Computers & Education, vol. 49, no 3, p. 794-808, 2007.
53. A. Coates et A. Y. Ng, « Learning Feature Representations with K-Means », in Neural Networks: Tricks of the Trade, Springer, Berlin, Heidelberg, 2012, p. 561-580.
54. R. D. Brown et Y. C. Martin, « Use of structure- activity data to compare structure-based clustering methods and descriptors for use in compound selection », Journal of Chemical Information and Computer Sciences, vol. 36, no 3, p. 572-584, 1996.
55. D. Arthur et S. Vassilvitskii, « k-means++: The advantages of careful seeding », in Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, 2007, p. 1027-1035.
56. D. J. Ketchen et C. L. Shook, « The application of cluster analysis in strategic management research: an analysis and critique », Strategic management journal, vol. 17, no 6, p. 441-458, 1996.
57. M. Anandarajan, C. Hill, et T. Nolan, « Cluster analysis: modeling groups in text », in Practical Text Analytics, Springer, 2019, p. 93-115.
58. P. J. Rousseeuw, « Silhouettes: a graphical aid to the interpretation and validation of cluster analysis », Journal of computational and applied mathematics, vol. 20, p. 53-65, 1987.
59. T. Caliński et J. Harabasz, « A dendrite method for cluster analysis », Communications in Statistics-theory and Methods, vol. 3, no 1, p. 1-27, 1974.
60. W. W. F. Lau et A. H. K. Yuen, « Exploring the effects of gender and learning styles on computer programming performance: implications for programming pedagogy », British Journal of Educational Technology, vol. 40, no 4, p. 696-712, 2009, doi: https://doi.org/10.1111/j.1467-8535.2008.00847.x.
61. J. L. Ross, M. T. Drysdale, et R. A. Schulz, « Cognitive learning styles and academic performance in two postsecondary computer application courses », Journal of Research on Computing in Education, vol. 33, no 4, p. 400-412, 2001.
62. M. Terry, « Translating learning style theory into developmental education practice: An article based on Gregorc's cognitive learning styles », Journal of College Reading and Learning, vol. 32, no 2, p. 154-176, 2002.

AUTHORS PROFILE

Ali El mezouary, Received his Ph.D and his Master's degree in Applied Mathematics and Computer science from Faculty of Science, University Ibn Zohr, Agadir-Morocco, in 2013, and 2008 respectively. Since July 2014, he is an assistant professor in the computer science department of the Higher School of Technology of the University Ibn Zohr in Agadir-Morocco. Since 2015, he is a member of the IRF-SIC Laboratory of the faculty of science of the same university. His thesis is part of a general problem of the educational semantic Web. His research area include: Learning system design; Adaptive Learning Systems; Semantic Web; Semantic Research; Educational Semantic Web; Ontology, Learning Analytics and Learning Styles.

Brahim Hmedna, Received his Ph.D and Master's degree in computer science from faculty of science, sniversity Ibn Zohr, Agadir-Morocco, in 2020, and 2012 respectively. Currently, He served as professor of computers science at the Heigh School. He is a member of IRF-SIC Laboratory. He supervises two doctoral theses in the same laboratory. His research interests include: MOOC; Data-science; Learning analytics, learning style; Frustration detection and satisfaction in MOOCS.

Omar Baz, He is a senior professor of mathematics at higher school of technology of Ibn Zohr University. Agadir-Morocco. He is a member of thee-Learning center of the same University. He is a member of IRF-SIC laboratory at the same University. He is a trainer at the AUF. His research interests include: E-learning; Learning styles; Adaptive Learning Systems and statistics. Currently, he directs 4 theses in the IRF-SIC Laboratory.