CrossMark

A fuzzy classification approach for learning style prediction based on web mining technique in e-learning environments

Ouafae El Aissaoui[1] · Yasser El Alami El Madani[2] · Lahcen Oughdir[1] · Youssouf El Allioui[3]

Abstract

Adaptive E-learning platforms provide personalized learning process relying mainly on learning styles. The traditional approach to find learning styles depends on asking learners to self-evaluate their own attitudes and behaviors through surveys and questionnaires. This approach presents several weaknesses including the lack of self-awareness of learners of their own preferences. Furthermore, the vast majority of learners experience boredom when they are asked to fill out the corresponding questionnaire. Besides that, traditional approach assumes that learning styles are fixed, and cannot change over time. In this paper, we propose a generic approach for detecting learning styles automatically according to a given learning styles model. In fact, our approach does not depend on a specific LSM. This work consists of two major steps. First, we extract learning sequences from learners log files using web usage mining techniques. Second, we classify the extracted learners' sequences according to a specific learning style model using clustering algorithms. To perform our approach we use Felder-Silverman Model as LSM and Fuzzy C-Means as a clustering algorithm. We have conducted an experimental study using a real-world dataset. The obtained results show that our approach outperforms traditional approach and provides promising results.

Keywords Web usage mining · Clustering algorithms · Fuzzy C-means clustering algorithm · Adaptive E-learning systems · Learning style · Learning style prediction · Learning object

✉ Youssouf El Allioui
   yelallioui@gmail.com

[1]   LSI, FPT, Sidi Mohamed Ben Abdellah Univesity, Fez, Morocco

[2]   ENSIAS, Mohammed V University, Madinat Al Irfane, B.P.: 713, Agdal, Rabat, Morocco

[3]   LS3M, FPK, Hassan First University, B.P. : 145, 25000 Khouribga, Morocco

1 Introduction

With the rapid growth of the Internet, E-learning platforms have become a promising solution to support learners to enhance their skills and capabilities to achieve their intended learning outcome. This fact encouraged educational institutions to develop e-learning platforms by creating new techniques and tools to provide their educational contents. However, early-developed platforms did not obtain good results since they considered learners as similar and consequently provide the same content in the same way. For this reason, many researchers have been encouraged to shift the focus from standardizing the design and the delivery of learning contents, to employing learner characteristics in learning process, creating a new current called adaptive-E-learning. Adaptive E-learning corresponds to techniques and tools developed to personalize learning process based on learner profile in order to improve his/her learning achievement. It is worth mentioning that both researchers and learners agree that taking into account learner profiles increase greatly the effectiveness of learning process (Abdullah et al. 2015). Generally, learner profile consists of learner characteristics such as his/her preferences, capacities, goals, learning styles, etc. (Brusilovsky and Millán 2007). Learning style remains one of the major characteristics that affect learning achievement of a given learner since it helps systems to personalize the learning process according to learner preferences. Learning styles can be defined as a description of the learner's attitudes and behaviors, which determine his/her preferred way of learning (Honey and Mumford 1992). They are represented as stable indicators, which describe how learners perceive, interact with, and responds to the learning environment (Keefe 1987). Learning styles have been characterized in different ways based on a variety of theoretical models. It is worth noting that several scales and classifications are in use such as Kolb's model (Kolb and Kolb 2005), Felder and Silverman Model (Felder and Silverman 1988), etc. Most of these models are similar and focus on various stables indicators related to environmental preferences, sensory modalities, personality types, and/or cognitive styles (Cook and Smith 2006). One of the main challenges to building reliable adaptive E-learning system is the identification of learners learning styles. The traditional detection way relies mainly on static method, which consists of asking learners to self-evaluate their own attitudes and behaviors through surveys and questionnaires. One of the well-known questionnaires is the index of Learning Styles[1] (ILS), which were developed by Felder and Silverman. However, this technique is not effective since it assumes that learning styles are fixed, and cannot change over time. Second, almost all learners are bored to fill out long questionnaires because it requires an additional effort and concentration. Moreover, static methods assume that learners are aware of their preferred way of learning, which is not always the case. In order to address these limitations, many works have been done in automatic learning style detection (El Aissaoui et al. 2018) (Feldman et al. 2015)(Latham et al. 2013) (El Allioui n.d.). Despite that, effective detection of learning styles still an open issue.

In this work, we propose a generic approach for detecting learning styles in an automatic manner through web mining techniques. This approach does not depend on a specific LSM thanks to its ability to be adapted to any learning style model. The

---

[1] https://www.webtools.ncsu.edu/learningstyles/

proposed learning style detection approach relies on three major steps: First, we extract and analyze learner sequences from learner log files stored on web servers using a web mining technique. In our case, learner sequences represent the learner activities in e-learning system. Second, we provide a mapping between learning objects and a given LSM in order to find the appropriate learning style of the concerned learner. Actually, the learner behavior and the learning style model are matched through learning objects. In fact, learning objects are the main characteristics, which describe the learning behavior while learning objects are used to categorize the learning styles. Finally, in order to detect, automatically, the learner learning style, we launch a clustering algorithm according to the classes of the used learning style Model. For our case study, we opt for Felder-Silverman learning styles Model (FSLSM) (Felder and Silverman 1988) since it is the most appropriate to implement adaptive e-learning portals (Graf and Kinshuk 2009)(Kuljis and Liu 2005). In addition, we use a fuzzy c-means clustering algorithm (FCM) to categorize learner sequences according to the 16 FSLS. FCM is able to manage effectively the uncertainty of data (Kannan et al. 2012). Indeed, clustering process may cause the overlapping membership of sequences, which will affect the goodness of clusters.

The remainder of this paper is organized as follows: Section 2 provides useful background. In section 3 we present the related work. Section 4 describes the proposed approach; Section 5 shows experiments and result analysis. Finally, we give some conclusions and discuss future directions.

2 Preliminaries

In this section, we will introduce some concepts that adaptive learning approaches are based on. These are the following concepts: 1) web usage mining (WUM), 2) learning style model (LSM), 3) Felder-Silverman learning style Model (FSLSM), 4) Learning Object (LO) and 5) the fuzzy c-means algorithm (FCM).

2.1 Web usage mining

Web usage mining (WUM) refers to the process of discovering and extracting automatically usage patterns from web data access sites (Cooley et al. 1997)(Mobasher 2006). It is based on identifying browsing patterns by analyzing the user's navigational behavior. The main data sources for WUM consists of log files, which contain user interaction information when users access resources stored in web servers. Web log files are plain text files which stores row data about user activities such as user IP Address, User Name, URL, etc.(Facca and Lanzi 2003).

According to (Cooley et al. 1999), the usage mining process consists of three steps: 1) data pre-processing, 2) pattern discovery and 3) pattern analysis. The first step consists of cleaning data and transforming it into a form that can be easily analyzed. The second step involves extracting hidden knowledge and patterns from the prepared data using machine learning and data mining techniques. Pattern Analysis is the final step, which focuses on extracting interesting patterns from the output of the pattern discovery process by keeping only the relevant ones.

2.2 Learning style model

The learner model (LM) is the most important component in an adaptive e-learning system (Riad et al. 2009), because of its ability to represent the characteristics of the learner according to which the educational system provides recommendations (Murray and Pérez 2015).

Among important characteristic that a LM can contain is that everyone has their own learning style (LS) (Murray and Pérez 2015). A LS refers to the preferred way in which the learner grasp and treat information, and it is considered as one of the main components of the LM since it can describe the learner's behavior when he interacts with the e-learning environment (Kolb and Kolb 2005) (Curry 1981).

According to the behavior of the learners during the learning process we can classify them into many learning style categories; these categories can be used to create a learning style model (LSM). However, there are many LSMs such as Kolb, 4MAT, VAK... and Felder-Silverman which is considered as the most popular one thanks to its ability to quantify students' learning style (Feldman et al. 2015).

2.3 Felder-Silverman model

According to the previous researches (Graf and Kinshuk 2009)(Kuljis and Liu 2005), the FSLSM is the most used in adaptive e-learning systems and the most appropriate to implement them. The FSLSM presents four dimensions with two categories for each one, where each learner has a dominant preference for one category in each dimension: processing (active/reflective), perception (sensing/intuitive), input (visual/verbal) and understanding (sequential/global).

Active (A) learners prefer to process information by interacting directly with the learning material, while reflective (R) learners prefer to think about the learning material. Active learners also tend to study in group, while the reflective learners prefer to work individually.

Sensing (Sen) learners tend to use materials that contains concrete facts and real-world applications, they are realistic and like to use demonstrated procedure and physical experiments. While the intuitive (I) learners prefer to use materials that contains abstract and theoretical information, they tend to understand the overall pattern from a global picture and then discovering possibilities.

The visual (Vi) learners prefer to see what they learn by using visual representations such as pictures, diagrams, and charts. While the verbal (Ve) learners like information that are explained with words; both written and spoken.

Sequential (Seq) learners prefer to focus on the details by going through the course step by step in a linear way.in the opposite, the global (G) learners prefer to understand the big picture by organizing information holistically.

2.4 Learning object

We can define a learning object (LO) as an elementary unit of learning which construct a learning content. The LO can take various digital resources forms as text, image, video, audio, etc. and it is characterized by its ability to be reusable, interoperable, durable, and accessible.

The reuse of LOs helps the educational institutes, on one side, to decrease the cost of creating them and on the other side to find useful contents by looking for them in dedicated repositories which contain collected LOs.

To be easily found, a LO has to be tagged and described with metadata .the most pertinent standards that are used to describe a LO are: MPEG-7, Dublin Core, and LOM (Keown 2007) (Mardis and Jo Ury 2008).

2.5 Fuzzy C-means clustering algorithm

FCM is an unsupervised learning algorithm, developed by Dunn in 1973 (Dunn 1973) and improved by Bezdek in 1983 (Peizhuang 1983). FCM is widely used for the reason that it is able to manage effectively the uncertainty of data (Kannan et al. 2012). Indeed, clustering process may cause the overlapping membership of data, which will affect the goodness of clusters.

FCM is a method of clustering which allows divide a set of data points into a predefined number of classes. Formally, this algorithm is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{n}\sum_{j=1}^{c} d_{ij}^{m} {F_i{-}C_j}^2, \quad 1 \leq m \leq \infty \tag{1}$$

Where $m \Re \ and > 1$, d_{ij} is the degree of membership of F_i in the cluster C_j, F_i is the ith of d-dimensional measured data and C_j is the d-dimension center of the cluster. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership d_{ij} using Eq. (2):

$$d_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\left\|F_i{-}C_j\right\|}{\left\|F_i{-}C_k\right\|}\right)^{\frac{2}{m-1}}} \tag{2}$$

and the cluster centers C_j by Eq. (3):

$$C_j = \frac{\sum_{i=1}^{n} d_{ij}^{m} * F_i}{\sum_{i=1}^{n} d_{ij}^{m}} \tag{3}$$

3 Related work

Many approaches have been proposed to automatically detect learners' learning styles based on machine learning techniques. First, In order to avoid intentional or unintentional wrong answers, and to save students' time on filling in a questionnaire.

Castro et al. (2007) provide a taxonomy of e-learning problems/ tasks to which Data Mining techniques have been applied, including, for instance: 1) learners' classification based on their learning performance and similar e-learning system usage, 2) course

adaptation and learning recommendations based on the learners' learning behaviors, 3) approaches that treat the evaluation of learning material and educational web-based course, 4) interaction optimization and e-learning platform navigation, and 5) feedback to both teachers and learners of e-learning content and detection of atypical students' learning behavior, etc.

Although authors of (Castro et al. 2007) aimed to make this general survey as complete as possible, they may have failed to find and identify some works, in the literature, proposing a learning style prediction approaches based on web mining techniques.

Abdullah et al. (2015) conducted a study to evaluate impact that learning style models can play on performance of learners in e-learning environment for providing recommendations for learners, instructors, and contents of online courses.

Abdullah et al. (2015) used Naïve-Bayes Tree as classification method to classify the learning style of learners as per FSLSM. The data collected through Blackboard LMS is analyzed using classification techniques in Weka tool. The results show that NB Tree classifier provided the highest correct value (69.697%) and it could be applied to develop FSLSM while taking into consideration learners' preference. Moreover, with the proposed approach learners' performance increased by more than 12%.

In the light of this fact, (Feldman et al. 2015) studied current directions made in the field of automatic learning styles detection. The authors reviewed, analyzed and summarized main findings from the studied works. They describe some variables that can be monitored in e-learning systems and which can be considered as useful insights to detect learning styles according to the categories defined in FSLSM.

In (Latham et al. 2013), authors have proposed a method which incorporates Artificial Neural Network technique based on a Multi-Layer Perceptron to automatically identify the learners' learning styles. The proposed method combines a set of behavior characteristics extracted from the tutoring conversation to increase the precision of the learning styles prediction. The classification step relies on classifying learners' data based on the four FSLSM dimensions: Processing (Active/Reflective) and Understanding (Sequential/Global).

Authors in (Chang et al. 2009) presented a learning style classification mechanism to classify and identify students' learning styles. 117 elementary school students are classified. The students were classified according to three learning styles: 1) dilatorily type, 2) transitory type and persistent type. This classification is based on students' behavior and the time spent on learning objects. Authors combined genetic algorithms (GA) with an improved k-nearest neighbor (K-NN) to classify students according to their learning styles.

Under the paradigm of fuzzy-logic, several approaches have been proposed such as (Jegatha Deborah et al. 2015), when Deborah et al. propose to use fuzzy rules to handle the uncertainty in the learning style predictions. They used the Gaussian membership function based fuzzy logic for learning of C programming language course contents. This approach is efficient in handling uncertainty in the learner's behavior through the information collected from their profile information and user activities in the e-learning system. It could be used in cases where there is incomplete user information.

In (Hogo 2010), Hogo has presented the use of two different fuzzy clustering techniques, FCM and kernelized FCM, to find the learner's categories and predict their

profiles. Fuzzy clustering reflects the learner's behavior more than crisp clus-tering. The author have mentioned five different types of learners such as 1) Regular students, 2) Bad students, 3) Worker students, 4) Casual students and 5) Absent students. Hogo has not considered any standard learning style model for learning styles of the learners. The usage data of students is captured based on questionnaire approach of FSLSM. Authors proved that the ability of KFCM was better than FCM in predicting the e-learners behaviour.

Hmedna, Mezoury, Baz, & Mammass (2016) introduced an automatic student modeling approach for identifying and tracking learners' learning styles based on their behavior and actions during a MOOC according to FSLSM. They proposed the exploitation of users' knowledge through neural network models to increase learners' engagement and satisfaction. This makes it possible to 1) identification of learners who have the same learning style, 2) form clusters based on learners' profiles and 3) do research recommendations of appropriate resources for each cluster though navigation-al support.

However, almost all works have taken into account the specific features of learning styles based on a particular learning style model. In fact, learning style models influence greatly the produced results. In addition, despite the positive impact that learning style models can play in recommendation phase to enhance the learning process, they are often ignored.

As can be noticed, all the previous works relied on a learning style model in their approaches. Most of the proposed approaches used the FSLSM's dimensions and considered that there are 8 learning styles where each one corresponds to a dimension's category. In reality, there are sixteen learning style combinations obtained by combing one category from each dimension. Similarly to the previous work; we have also relied on the FSLSM, but by considering sixteen LSC.

4 Our approach

After extraction of learners' behaviors using Web Log analysis approach, we will follow the following steps: 1) description of general model of our approach, 2) mapping the learning objects onto FSLSM, 3) finally we describe some evaluation metrics aiming at measure the goodness of our method.

4.1 Methodology

An adaptive E-learning system takes into account the learner's learning style and provides contents to the learners based on their preferred learning styles which are identified using the learners' sequences. To identify a learner's learning style we have to rely on a standard learning style model such as FSLSM, where the captured sequences can be labeled with a specific learning style combination using an unsupervised algorithm. In order to implement the unsupervised algorithm, the learner's sequences which have been extracted from the log file, must be transformed to the input of that algorithm.

The learning sequence of a learner is defined by the various learning objects accessed by that specific learner during a session. Each sequence contains the sequence

id, session id, learner id, and the set of learning objects accessed by the learner in a session.

After detecting the learners' sequences, we classify them according to the chosen LSM by assigning a specific learning style combination to each sequence.

Figure 1 resumes our proposed approach in its general form:

FSLSM's feature values and the sequences of learners are given as an input to the clustering algorithm. Based on the feature values, the sequences are grouped together.

In this work, we employ FCM algorithm. FCM is widely used thanks to its ability to manage effectively the uncertainty of data. Indeed, clustering process may cause the overlapping membership of sequences, which will affect the goodness of clusters.

Starting with random initialization of the membership values for each sequence from the manually selected clusters, the clusters are converged by recursively updating the cluster centers and membership function in Eq. (2) and Eq. (3). This is to minimize the objective function in Eq. (1). Convergence stops when the overall difference in the membership function between the current and previous iteration is smaller than a given epsilon value, ε. After the convergence, and defuzzification is applied. Each sequence is assigned to a specific cluster according to the maximal value of its membership function. The learner's learning style is grouped based on the center value selection. The center values are calculated based on the feature values and assigned to each cluster. Finally, all the sequences are grouped into sixteen clusters of FSLSM categories and each sequence is assigned the label as per FSLSM. The steps of the FCM algorithm are illustrated in Algorithm 1 as shown in Table 1.

4.2 LOs – FSLSM mapping

In order to match a LOs to LSM, we should have relied on a FSLSM. there are several reasons for choosing this model, one important reason, is that the FSLSM considers that the student's learning style can be changed unexpectedly and in a non-deterministic way (Graf and Kinshuk 2009), therefore, our approach aims to update the student's learning style dynamically after each interaction with the adaptive e-learning systems. According to the previous researches (Kuljis and Liu 2005), the FSLSM is the most
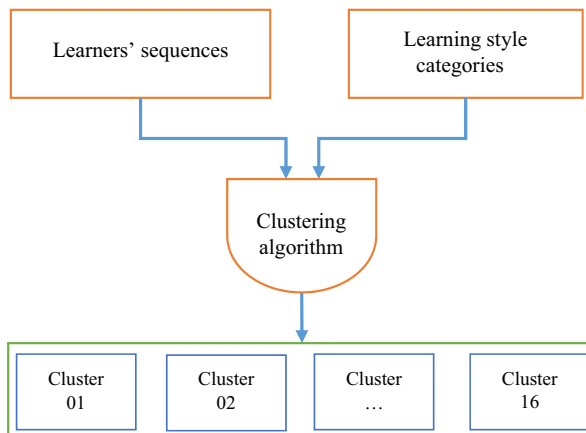

Fig. 1 General model

Table 1 FCM Algorithm

Step 1: Initialize the number of clusters:

m = 2

$\varepsilon = 0.005$

M_{ij} = degree of member ship function x_i in cluster j

F_i = input sequence

C_j = center of the cluster

Step 2: Initialize the membership function randomly

Step 3: repeat

 Update the cluster center, C_j using **Eq. (3)**

 Update the membership function M_{ij} using **Eq. (2)**

Until $\left\| M_{ij}^{k+1} - M_{ij}^{k} \right\| < \varepsilon$

used in adaptive e-learning systems and the most appropriate to implement them. According to FSLSM there are four dimensions, where each dimension contains two opposite categories, and each learner prefers a specific category in each dimension. Thus, to identify the learner' learning style; we have to determine a combination composed by one category from each dimension. As a result we will obtain sixteen combinations:

```
Learning Styles Combinations (LSCs) = { (R,I,Ve,G),
(A,I,Ve,G), (R,Sen,Ve,G), (A,Sen,Ve,G), (R,I,Vi,G),
(A,I,Vi,G),(R,Sen,Vi,G), (A,Sen,Vi,G), (R,I,Ve,Seq),
(A,I,Ve,Seq), (R,Sen,Ve,Seq), (A,Sen,Ve,Seq), (R,I,Vi,Seq),
(A,I,Vi,Seq), (R,Sen,Vi,Seq), (A,Sen,Vi,Seq)} .
```

Basically, we suppose that each LSC reflects the preferred LOs that are accessed by a learner during the learning process, so in order to identify the LSC for each learner, we first have to match the Los with its appropriate LSC. Based on the matching table presented in our previous work (El Aissaoui et al. 2018) (El Allioui n.d.); we have obtained the following Table 2, where the mapped learning objects are considered a feature values of the FCM algorithm.

4.3 Evaluation metrics

In order to measure the goodness of our method and evaluate the performance of the classifier used in our approach, we have used the confusion matrix technique. The confusion matrix technique is a specific table layout that summarizes the number of correct and incorrect predictions in each class, and it is used to compute several validation metrics.

We suppose that we have a confusion matrix with n classes; the following equations show how to compute the total number of false negative (TFN), false positive (TFP), true negative (TTN), and the true positive (TTP).

FN is the number of instances the classifier predicted as negative but they are positive. It is calculated as in Eq. (4):

Table 2 Learning objects as per FSLSM

Cluster ID	Combination	Videos	PPTs	Demo	Exercise	Assignments	PDFs	Announcements	References
C01	(R, I, Ve, G)	3	2	0	1	1	3	2	3
C02	(A, I, Ve, G)	3	2	1	2	2	2	1	2
C03	(R, Sen, Ve, G)	3	1	0	1	1	3	2	2
C04	(A, Sen, Ve, G)	3	1	1	2	2	2	1	1
C05	(R, I, Vis, G)	3	2	0	1	1	1	1	4
C06	(A, I, Vi, G)	3	2	1	2	2	2	0	3
C07	(R, S, Vi, G)	3	1	0	1	1	1	1	3
C08	(A, S, Vi, G)	3	1	1	2	2	1	0	2
C09	(R, I, Ve, Seq)	3	2	0	1	1	3	2	3
C10	(A, I, Ve, Seq)	3	2	1	2	2	2	1	2
C11	(R, Sen, Ve, Seq)	3	1	0	1	1	3	2	2
C12	(A, Sen, Ve, Seq)	3	1	1	2	2	2	1	1
C13	(R, I, Vi, Seq)	3	2	0	1	1	2	1	4
C14	(A, I, Vi, Seq)	3	2	1	2	2	2	0	3
C15	(R, Sen, Vi, Seq)	3	1	0	1	1	2	1	3
C16	(A, Sen, Vi, Seq)	3	1	1	2	2	1	0	2

Cluster ID	Examples	Practical Material	Forum	Topic list	Images	Charts	Email	Sequential
C01	0	0	1	2	0	0	1	0
C02	0	0	1	2	0	0	1	0
C03	1	1	0	1	0	0	1	0
C04	1	1	0	1	0	0	1	0
C05	0	0	1	2	1	1	0	0

Table 2 (continued)

Cluster ID	Examples	Practical Material	Forum	Topic list	Images	Charts	Email	Sequential
C06	0	0	1	2	1	1	0	0
C07	1	1	0	1	1	1	0	0
C08	1	1	0	1	1	1	0	0
C09	0	0	1	1	0	0	1	1
C10	0	0	1	1	0	0	1	1
C11	1	1	0	0	0	0	1	1
C12	1	1	0	0	0	0	1	1
C13	0	0	1	1	1	1	0	1
C14	0	0	1	1	1	1	0	1
C15	1	1	0	0	1	1	0	1
C16	1	1	0	0	1	1	0	1

$$FN_i = \sum\nolimits_{\substack{j=1 \\ j \neq i}}^{n} x_{ij} \tag{4}$$

FP is the number of instances the classifier predicted as positive but they are negative. It is calculated as in Eq. (5):

$$FP_i = \sum\nolimits_{\substack{j=1 \\ j \neq i}}^{n} x_{ji} \tag{5}$$

TN is the number of negative instances the classifier correctly identified as negative. It is calculated as in Eq. (6):

$$TN_i = \sum\nolimits_{\substack{j=1 \\ j \neq i}}^{n} \sum\nolimits_{\substack{k=1 \\ k \neq i}}^{n} x_{jk} \tag{6}$$

TP is the number of positive instances the classifier correctly identified as positive. It is calculated as in Eq. (7):

$$TTP_{all} = \sum\nolimits_{j=1}^{n} x_{jj} \tag{7}$$

In order to evaluate the performance of a classifier, the following measures can be computed for each class i based on the equations described above. In particular, this concerns four validation metrics: Accuracy (A), Precision (P), Recall (R) and F-Measure (F1).

Precision (P): it is also called "Positive Predictive value", it represents the fraction of true positive instances among the predicted positive instances. Precision is calculated as in Eq. (8).

$$P_i = \frac{TP_i}{TP_i + FP_i} \tag{8}$$

Recall (R): it is also called "Sensitivity", it is the proportion of positive instances that are correctly classified as positive. Recall is calculated as in Eq. (9).

$$R_i = \frac{TP_i}{TP_i + FN_i} \tag{9}$$

Accuracy (A): it defines the rate at which a model has classified the records correctly. Accuracy is calculated as in Eq. (10).

$$A = \frac{TTP_{all}}{total\ number\ of\ classifications} \tag{10}$$

F_1 − Measure: Since the class distribution is uneven, it is always better to calculate the F_1 score. The F_1 score provides the weighted average of precision and recall. Hence it considers both False Positive (FP) & False Negative (FN). F_1 − Measure is calculated

as in Eq. (11)

$$F_1-\text{Measure} = \frac{2(R \, X \, P)}{(R + P)} \qquad (11)$$

5 Experiment and results

To measure the effectiveness of our proposal, we conducted an experiment on students from computer science and Management school.

The learners' behavior data was collected from the E-learning platform of Sup'Management Group.[2] The online educational system was developed based on Moodle platform. Moodle[3] is a free Open Source software used to help educators and students to enhance the learning process.

A total of 126 students participated to this study. First, students filled in the questionnaire[4] based on the index of learning styles[5] (ILS) developed by Richard M. Felder and Barbara A. Soloman (1988). Second, we collected 1235 sequences from the E-learning platform, which represent the learners' behavior. The number of sequences varied from 1 to 35 sequences per students.

The sequences of learners constitute the input of the FCM clustering algorithm. The Center values for each cluster are computed based on the feature values based on mapping of learning objects to FSLSM categories. The result of the clustering is shown in Fig. 2.

1235 similar sequences are considered for clustering based on K-means and FCM algorithms. Center values for each cluster are computed based on the feature values i.e. mapping of learning objects to FSLSM categories. The result of clustering as shown in Fig. 2.

The sequences are labeled as per 16 categories of FSLSM. The total number of sequences clustered by FCM algorithm is 1298 sequences, there are more than the 1235 input sequences as some of the sequences belong to more than one cluster based on the feature values.

5.1 Goodness of clusters

In our work, we compare the goodness of our approach with the K-means algorithm presented in (El Aissaoui et al. 2018) using four factors: 1) Accuracy, 2) Precision, 3) Recall and 4) F1-Measure. All these evaluation metrics are presented in section 4.3.

The FCM algorithm is used to classify the sequences as per the 16 categories of FSLSM. The algorithm is launched for various numbers of Iterations (I) and the accuracy obtained and compared with K-means algorithm, as shown in Table 3.

---

[2] https://www.supmanagement.ma/fc/login/index.php
[3] https://moodle.org/
[4] https://goo.gl/forms/iVoTLzHa7Bp4n46x1
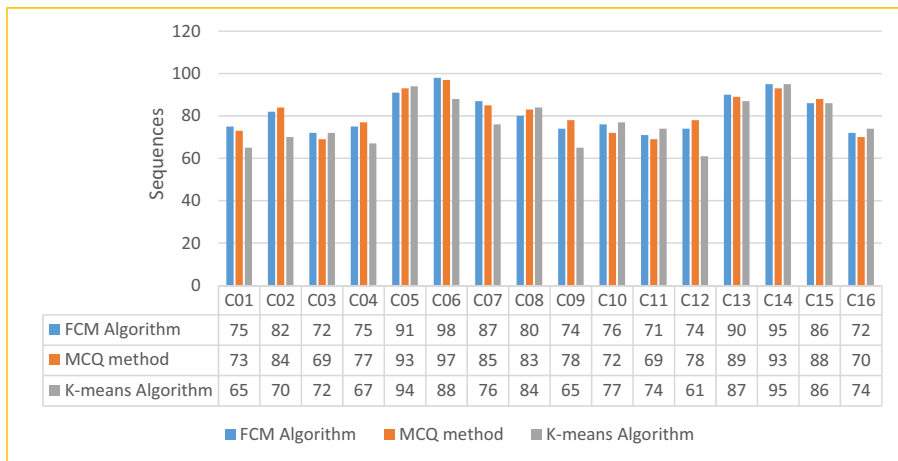[5] https://www.webtools.ncsu.edu/learningstyles/

Fig. 2 Result of clustering of learners' sequences using FCM, MCQ and K-means methods

According to values defined in the Table 3, the accuracy of FCM algorithm is 95.93% for 200 iterations. So, if the number of iteration increases, the accuracy may increase but algorithm will take more time for execution.

In order to support the correctness of our algorithm and to measure its scalability, we compare its time complexity with the time complexity of K-means clustering algorithm presented in (El Aissaoui et al. 2018).

The time complexity of K-means algorithm (Xu and Wunsch 2005) is $O(ndci)$ and the time complexity of FCM (Almeida and Sousa 2006) is $O(ndc^2i)$ where n represent the number of data points, d is the number of clusters, c is the number of dimensions and i is the number of iterations. In our case, $n = 1235$, $d = 16$, $c = 4$ and $i = \{60, 100, 200\}$. We have computed and compared the time complexity of both algorithms and we obtain the results displayed as Table 4.

Finally, from all obtained results we may conclude that FCM algorithm produces best results than K-Means algorithm but it still requires more computation time. This is due to the fuzzy measures calculations involvement in the FCM algorithm.

Table 3 Result of clustering algorithms

Iteration	FCM algorithm				K-means algorithm			
	Accuracy	Precision	Recall	F-Measure	Accuracy	Precision	Recall	F-Measure
≈60	88.61%	89.93%	92.35%	88.97%	74.31%	75.83%	78.34%	76.16%
≈100	89.54%	91.32%	93.81%	90.78%	75.90%	77.81%	80.23%	78.33%
≈200	96.89%	97.27%	99.44%	98.53%	78.83%	79.90%	83.10%	80.12%

Table 4 Time complexity of k-means and FCM according to number of iterations

I	Time complexity of K-means algorithm $O(ndci)$	Time complexity of FCM algorithm $O(ndc^2i)$
60	47,4 s	758,7 s
100	79 s	1264 s
200	158 s	2529 s

6 Conclusion

In this work, we have proposed a generic approach for detecting learning styles automatically. It relies on two major steps. The first step is based on extracting learning sequences from learners log files using web usage mining techniques while the second step attempt to classify the extracted learners sequences according to a specific learning style model using clustering algorithms. Our proposal is generic due to its ability to be adapted to a given LSM and to be implemented by a given clustering algorithm. In this paper, we focused on FSLSM, the most widely used LSM in E-learning environment. To perform our approach we use FCM as a clustering algorithm to manage effectively the uncertainty of data. Using a real-world dataset, we conducted an experimental study. Findings show that our approach outperforms traditional approach and provides good result results. As future work, we will investigate the use of other machine learning algorithms to improve the accuracy of the learning style detection.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Abdullah, M., Bashmail, R. M., Daffa, W. H., Alzahrani, M., & Sadik, M. (2015). The impact of learning styles on Learner's performance in E-learning environment. *International Journal of Advanced Computer Science and Applications, 6*(9), 24–31. https://doi.org/10.14569/IJACSA.2015.060903.

Almeida, R. J., & Sousa, J. M. C. (2006). Comparison of fuzzy clustering algorithms for classification. In *International Symposium on Evolving Fuzzy Systems* (Vol. 00, pp. 112–117). Ambleside: IEEE Computer Society. https://doi.org/10.1109/ISEFS.2006.251138.

Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web*. Heidelberg: Springer, Berlin. https://doi.org/10.1007/978-3-540-72079-9_1.

Castro, F., Vellido, A., Nebot, À., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. In L. C. Jain, R. A. Tedman, & D. K. Tedman (Eds.), *Evolution of Teaching and Learning Paradigms in Intelligent Environment (Vol. 62)* (pp. 183–221). Heidelberg: Springer, Berlin. https://doi.org/10.1007/978-3-540-71974-8_8.

Chang, Y.-C., Kao, W.-Y., Chu, C.-P., & Chiu, C.-H. (2009). A learning style classification mechanism for e-learning. *Computers & Education, 53*(2), 273–285. https://doi.org/10.1016/j.compedu.2009.02.008.

Cook, D. A., & Smith, A. J. (2006). Validity of index of learning styles scores: Multitrait-multimethod comparison with three cognitive/learning style instruments. *Medical Education, 40*(9), 900–907. https://doi.org/10.1111/j.1365-2929.2006.02542.x.

Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the World Wide Web. In *Proceedings Ninth IEEE International Conference on Tools with Artificial*

Intelligence (pp. 558–567). Newport Beach: IEEE Comput. Soc. https://doi.org/10.1109/TAI.1997.632303.

Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Knowledge and Information Systems, 1*(1), 5–32. https://doi.org/10.1007/BF03325089.

Curry, L. (1981). Learning preferences in continuing medical education. *Canadian Medical Association Journal, 124*(5), 535–536 Retrieved from http://www.cmaj.ca/content/124/5/535.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics, 3*(3), 32–57. https://doi.org/10.1080/01969727308546046.

El Aissaoui, O., El Madani El Alami, Y., Oughdir, L., & El Allioui, Y. (2018). Integrating web usage mining for an automatic learner profile detection: A learning styles-based approach. In *2018 international conference on intelligent systems and computer vision (ISCV)* (pp. 1–6). Fez: IEEE. https://doi.org/10.1109/ISACV.2018.8354021.

El Allioui, Y. (n.d.). Advanced Prediction of Learner's Profile based on Felder Silverman Learning Styles using Web Usage Mining approach and Fuzzy C-Means Algorithm. *International Journal of Computer Aided Engineering and Technology*, "*in press*"(forthcoming), 18. Retrieved from http://www.inderscience.com/info/ingeneral/forthcoming.php?jcode=ijcaet.

Facca, F. M., & Lanzi, P. L. (2003). Recent Developments in Web Usage Mining Research. In *International Conference on Data Warehousing and Knowledge Discovery* (pp. 140–150). Prague: Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-45228-7_15.

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering Education, 78*(7), 674–681 Retrieved from http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS-1988.pdf.

Feldman, J., Monteserin, A., & Amandi, A. (2015). Automatic detection of learning styles: State of the art. *Artificial Intelligence Review, 44*(2), 157–186. https://doi.org/10.1007/s10462-014-9422-6.

Graf, S., & Kinshuk. (2009). Advanced adaptivity in learning management systems by considering learning styles. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 3, pp. 235–238). Milan: IEEE Computer Society. https://doi.org/10.1109/WI-IAT.2009.271.

Hogo, M. (2010). Evaluation of e-learners behaviour using different fuzzy clustering models: A comparative study. *ArXiv Preprint ArXiv:1003.1499, 7*(2), 131–140. Retrieved from https://arxiv.org/abs/1003.1499.

Honey, P., & Mumford, A. (1992). Book reviews: The manual of learning styles. *Management Learning, 14*(2), 147–150. https://doi.org/10.1177/135050768301400209.

Jegatha Deborah, L., Sathiyaseelan, R., Audithan, S., & Vijayakumar, P. (2015). Fuzzy-logic based learning style prediction in e-learning using web interface information. *Sadhana - Academy Proceedings in Engineering Sciences, 40*(2), 379–394. https://doi.org/10.1007/s12046-015-0334-1.

Kannan, S. R., Ramathilagam, S., & Chung, P. C. (2012). Effective fuzzy c-means clustering algorithms for data clustering problems. *Expert Systems with Applications, 39*(7), 6292–6300. https://doi.org/10.1016/j.eswa.2011.11.063.

Keefe, J. W. (1987). *Learning style: Theory and practice*. National Association of secondary school principals. Retrieved from https://eric.ed.gov/?id=ED286873.

Keown, R. (2007). Learning objects: What are they, and why should we use them in distance education? *Distance Learning, 4*(4), 73–77 Retrieved from https://search.proquest.com/openview/8b9398df073c66bd0265444c0e5bd3f3/1?pq-origsite=gscholar&cbl=29704.

Kolb, A. Y., & Kolb, D. A. (2005). The Kolb Learning Style Inventory — Version 3 . 1 2005 Technical Specifications. *LSI Technical Manual*, 1–72. https://doi.org/10.1016/S0260-6917(95)80103-0.

Kuljis, J., & Liu, F. (2005). A comparison of learning style theories on the suitability for elearning. In IASTED international conference on web technologies, applications, and services (pp. 191–197). Calgary: IASTED/ACTA Press. Retrieved from http://dblp.uni-trier.de/db/conf/iastedWTAS/wtas2005.html#KuljisL05.

Latham, A., Crockett, K., & Mclean, D. (2013). Profiling Student Learning Styles with Multilayer Perceptron Neural Networks. In *2013 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 2510–2515). https://doi.org/10.1109/SMC.2013.428.

Mardis, L. A., & Jo Ury, C. (2008). Innovation – An LO library: Reuse of learning objects. *Reference Services Review, 36*(4), 389–413. https://doi.org/10.1108/00907320810920360.

Mobasher, B. (2006). Web Usage Mining. In *Web data mining: Exploring hyperlinks, contents, and usage data* (pp. 1216–1220). https://doi.org/10.1007/978-3-642-19460-3.

Murray, M. C., & Pérez, J. (2015). Informing and performing: A study comparing adaptive learning to traditional learning. *The International Journal of an Emerging Transdiscipline, 18*(1), 111–125 Retrieved from http://www.inform.nu/Articles/Vol18/ISJv18p111-125Murray1572.pdf.

Peizhuang, W. (1983). Pattern recognition with fuzzy objective function algorithms (James C. Bezdek). *SIAM Review, 25*(3), 442–442. https://doi.org/10.1137/1025116.

Riad, A. M., El-Minir, H. K., & El-Ghareeb, H. A. (2009). Review of e-learning systems convergence from traditional systems to services based adaptive and intelligent systems. *Journal of Convergence Information Technology, 4*(2), 108–131 Retrieved from https://pdfs.semanticscholar.org/20b0/c04db26298d3e4ab4eebbf76dc030b5f3753.pdf.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*. IEEE Computer Society. https://doi.org/10.1109/TNN.2005.845141.