DOI: 10.15514/ISPRAS-2024-36(1)-10

From Interaction Data to Personalized Learning: Mining User-Object Interactions in Intelligent Environments

J.G. Hernández-Calderón, ORCID: 0000-0002-8014-8819 <*guillermohernandez02@uv.mx*>
E.I. Benítez-Guerrero, ORCID: 0000-0001-5844-4198 <*edbenitez@uv.mx*>
J.R. Rojano-Cáceres, ORCID: 0000-0002-3878-4571 <*rrojano@uv.mx*>
C. Mezura-Godoy, ORCID: 0000-0002-5386-107X <*cmezura@uv.mx*>

Facultad de Estadística e Informática, Universidad Veracruzana
Av. Xalapa S/N, Col. Obrero Campesina, C.P. 91020,
Xalapa, Veracruz México.

Abstract. The aim of this work is to contribute to the personalization of intelligent learning environments by analyzing user-object interaction data to identify On-Task and Off-Task behaviors. This is accomplished by monitoring and analyzing users' interactions while performing academic activities with a tangible-intangible hybrid system in a university intelligent environment configuration. With the proposal of a framework and the Orange Data Mining tool and the Neural Network, Random Forest, Naive Bayes, and Tree classification models, training and testing was carried out with the user-object interaction records of the 13 students (11 for training and two for testing) to identify representative sequences of behavior from user-object interaction records. The two models that had the best results, despite the small number of data, were the Neural Network and Naive Bayes. Although a more significant amount of data is necessary to perform a classification adequately, the process allowed exemplifying this process so that it can later be fully incorporated into an intelligent educational system to contribute to build personalized environments.

Keywords: intelligent learning environments; user-behavior identification; user-object interaction data; data mining.

For citation: Hernández-Calderón J.G., Benítez-Guerrero E., Rojano-Cáceres J.R., Mezura-Godoy C. From Interaction Data to Personalized Learning: Mining User-Object Interactions in Intelligent Environments. *Trudy ISP RAN/Proc. ISP RAS*, vol. 36, issue 1, 2024. pp. 157-174. DOI: 10.15514/ISPRAS-2024-36(1)-10.

Full text: Hernández-Calderón J.G., Benítez-Guerrero E., Rojano R., Mezura-Godoy C. From Interaction Data to Personalized Learning: Mining User-Object Interaction Data for Student Modeling in Intelligent Learning Environments. *Programming and Computer Software*, 2023, Vol. 49, No. 8, pp. 657–670. DOI: 10.1134/S036176882308008X.

Acknowledgements. The first author gratefully acknowledges CONAHCYT for scholarship No. 421557 for graduate studies.

От простого взаимодействия к персонализированному обучению: исследование взаимодействий пользователя с объектом в интеллектуальном окружении

Х.Г. Эрнандес-Кальдерон, ORCID: 0000-0002-8014-8819 <*guillermohernandez02@uv.mx*>
Э.И. Бенитес-Герреро, ORCID: 0000-0001-5844-4198 <*edbenitez@uv.mx*>
Х.Р. Рохано-Касерес, ORCID: 0000-0002-3878-4571 <*rrojano@uv.mx*>
К. Месура-Годой, ORCID: 0000-0002-5386-107X <*cmezura@uv.mx*>

Факультет статистики и информатики Университета Веракруса,
Мексика, 91020, Веракрус, Халапа, проспект Халапа, квартал Обреро-Кампесина.

Abstract. Цель этой работы состоит в развитии степени персонализации интеллектуальных учебных сред путем анализа данных взаимодействия пользователя с объектом, чтобы определить его поведение при решении задачи и в перерывах в работе (On-Task/Off-Task). Это достигается путем мониторинга и анализа взаимодействия пользователей при выполнении учебных заданий с гибридной системой, адаптированной для интеллектуальной университетской среды. Предложенная конфигурация системы, дополненная инструментарием добычи данных Orange и классификаторами на основе моделей нейронной сети, случайного леса, наивного классификатора Байеса и дерева классификации было проведено обучение взаимодействиям с объектами и тестирование 13 студентов (11 для обучения и два для тестирования), что позволило выявить представительные последовательности действий. Несмотря на небольшое количество данных, удалось понять, что наилучшие результаты показали две модели – нейронная сеть и наивный классификатор Байеса. Хотя для адекватного выполнения классификации необходим более значительный объем данных, проведенный опыт позволил лучше понять процесс. Впоследствии его можно будет полностью включить в интеллектуальную образовательную систему, что позволит внести вклад в создание персонализированных сред.

Ключевые слова: интеллектуальное окружение обучения; идентификация поведения пользователя; данные по взаимодействию пользователя с объектом; извлечение данных.

Для цитирования: Эрнандес-Кальдерон Х.Г., Бенитес-Герреро Э.И., Рохано-Касерес Х.Р., Месура-Годой К. От простого взаимодействия к персонализированному обучению: исследование взаимодействий пользователя с объектом в интеллектуальном окружении. Труды ИСП РАН, том. 36, вып. 1, 2024. стр. 157-174 (на английском языке). DOI: 10.15514/ISPRAS-2024-6(1)-10.

Полный текст: Эрнандес-Кальдерон Х.Г., Бенитес-Герреро Э.И., Рохано Р., Месура-Годой К. От простого взаимодействия к персонализированному обучению: извлечение данных о взаимодействии пользователя с объектом для моделирования поведения студентов в интеллектуальном окружении. Programming and Computer Software, 2023, т. 49, № 8, стр. 657–670 (на английском языке). DOI: 10.1134/S036176882308008X.

Благоларности. Первый автор благодарит Национальный совет Мексики по науке и технике (CONAHCYT) за стипендию № 421557 для обучения в аспирантуре.

1. Introduction

Human behavior refers to how an individual responds to various stimuli in their environment, physically and mentally. Computationally, it is possible to determine user behavior from observation and comparative analysis, and it is conceivable to classify their behavior by monitoring the development of a person's activities. In an intelligent environment, behavior patterns can be obtained from analyzing user interactions data [1]. For a computer system, modeling the behavior of users and understanding signs of the patterns it identifies involves various tasks, the first of which is to monitor and collect data related to user interactions, and then analyze the detected behavior signals considering the context.

Generally, behaviors are detected initially by identifying movements or events that happen within the environment. These are low-level and do require temporal or contextual knowledge to be placed

[2]. In intelligent environments, this information is collected by sensors when any movement or interaction is detected; then, the set of registered events and their space-time relationships give rise to activities.

There exist different challenges to building behavior-aware intelligent environments. An example of these challenges is the development of robust approaches for recognizing activities [3]. Another situation to face is that when detecting human activities, they are complex to model and dynamic because they constantly change and evolve with the user [2].

One approach considers that a critical factor for interpreting human behavior relates to the moment the actions are performed and their duration [2]. Moreover, concurrent activities (executed at the same time) [4] and sequential activities can be recognized [5]. In ubiquitous computing applications, advances in activity recognition have allowed passing from low-level recognition to identifying daily life activities.

This paper presents an approach for identifying user behavior through the analysis of user-object interactions. This is addressed in the context of a case study related to students solving academic activities using a Tangible-Intangible hybrid system that allows detecting and collecting user-object interaction data. The main contribution of this paper is the analysis of user-object interaction data to find relevant information related to user behavior while performing an academic activity in a technologically enhanced environment.

The remainder of this paper is organized as follows. Section 2 presents background information related to behavior representation, behavior detection, and behavior analysis. Section 3 describes a framework for analyzing user-object interaction data, that encompasses 3 phases: 1) Extraction and processing, 2) Transformation, and 3) Pattern recognition and evaluation; further explained in Sections 4, 5, and 6 respectively. Section 7 presents the discussion and limitations. Finally, Section 8 concludes this paper.

2. Background

Human behavior refers to how an individual responds to various stimuli in their environment, physically and mentally. Computationally, it is possible to determine user behavior from observation and comparative analysis, and it is conceivable to classify their behavior by monitoring the development of a person's activities. In an intelligent environment, behavior patterns can be obtained from analyzing user interactions data [1]. For a computer system, modeling the behavior of users and understanding signs of the patterns it identifies involves various tasks, the first of which is to monitor and collect data related to user interactions, and then analyze the detected behavior signals considering the context.

This section describes the characteristics of the styles used in this document.

2.1 Behavior representation

In the educational field, different research projects have defined student behavior from different perspectives and classifications, the most general being On-Task and Off-Task behaviors [6-7].

On-Task behavior is defined as the activity performed by the student that complies with the instructions given by the teacher for the task or lesson; for example, listening to the teacher's instructions, writing, reading a text, talking to the teacher, and looking up words in the dictionary. In addition, it defines Off-Task behavior as any activity performed by the student that is not directly associated with the task in the way that it was instructed, for example, leaving the classroom, excessive erasing, drawing on the desk, or looking at a point in space [6]. Another definition of On-Task behavior refers to the student's attention during class time. While Off-Task behavior, as student inattention during class time as self-distractions, interactions with their peers, environmental distractions, or others that include things or actions that are not classifiable into previous categories [7].

In addition, it is possible to find that behaviors studied and classified in educational settings include affective states such as boredom, commitment, and concentration [8], mental states such as flow experience [9], and dedication or absorption during tasks [10-11].

2.2 Behavior detection

Different research projects have studied human-object interaction detection. One example is the improvement of the accuracy while detecting interactions between persons and objects using computer vision and a graph neural network [12], or a graph model-based algorithm [13]; an application for smart glasses that assists workers in an industrial site recognizing human-object interactions [14]; as a contribution in the visual understanding field [15-16], or to help to solve the problem of missing human behavior objects [17]. It can be observed that behavior detection is related to the technologies available in the intelligent environment configuration, the sensors installed in the background, or can be related to the data sources, such as the case of images, videos, interactions with software systems [18], cameras for collecting images, microphones to record sounds within the classroom [19], and the dialogue between student and teacher and wearable devices [20] through which it is possible to identify tasks and student behavior in a classroom or digital learning environments as intelligent tutors. These data provide information that identifies individual and group behaviors [21]. Although observation methods for identifying student behaviors predominate in the educational field, there are efforts to identify them from the interaction data that the student has with a computer program, as observed in the work of [6, 8, 22, 23].

However, there is no defined record structure to store user interaction data in intelligent educational environments. Even though works such as [13, 24-31] use sensor data to address the situations of interest of each job, no defined structure is observed that they share among themselves.

2.3 Behavior analysis

Several approaches are observed to analyze data generated by users in intelligent environment settings. These include visual representations and analysis routines that allow the teacher to track the performance of one or several students to track their progress [27, 32-33]; comparison, and classification. Analysis of the accelerometer data to identify activities of daily living using a Gaussian mixture model and Gaussian mixture regression [34] to classify the data at runtime. Also, the use of multi-agent systems to analyze user behavior [35] and logistic regression models to predict, at runtime, behaviors of lack of commitment or gaming the system behavior.

Different research projects have used various techniques for data analysis to identify user behavior. In intelligent environment settings, these techniques include but are not limited to the Hidden Markov Model (HMM), probabilistic hierarchical models of human behavior using HMM, conditional random fields, and dynamic Bayesian networks. Other works, such as those by [36] and [37], consider artificial neural networks for monitoring and predicting activities of daily living; On the other hand, [38] proposes a hybrid inference approach to detect abnormal user behavior, and [39] uses sequential pattern mining. Also, for the analysis of human behavior, [40] has proposed using a temporal structure or a set of actions over time with T-patterns, broadly used by [41] and [42] or sequential patterns through the GSP algorithm, proposed by Skirant [43]. This algorithm allows the discovery of sequential patterns with a minimum support specified by the user, where the support of a pattern is the number of data sequences containing that pattern. It makes multiple passes over the data [43].

3. Framework for analyzing user-object interaction data

This section describes a proposal, based on [44-45], for a framework to analyze user-object interaction data for student modeling in intelligent learning environments. It consists of three phases: data acquisition and storage, transformation, and analysis.

Extraction and processing. The main objective of this phase is to collect and store user-object interaction data in an intelligent learning environment setting from a Tangible User Interface, the TanQuery prototype [46]. In TanQuery, there are three different areas where student interactions take place and are identified: the assignment area, the work area, and the result area. In the assignment area, the user assigns a value to the tokens he will use, depending on the token type and the information available in the database. In the work area, the user places the tokens to build a query tree and the system can identify the objects that compose it (root, left child, right child, or attributes); this distribution of objects is evaluated, and relational algebra expressions are constructed. The relational algebra expression (and the equivalent SQL expression) as well as the results of its evaluation are displayed in the result area. Fig. 1 shows the configuration of the TanQuery prototype.

During the activity, the student interacts with different tokens (objects) and generates different records of user-object interactions. This user-object interaction data is stored in a local database with a structure based on the physical layer of the intelligent desktop conceptual model [44] and considers the identity of each object visible to the system, the type of object (relation, attribute, operator, among others), the position (x, y), and the degrees used to calculate the location of the object in the application domain. Also, this record considers the domain information in the root, right child, left child, parent, and attributes, each for each component of the query tree; it also considers the date and interaction time.

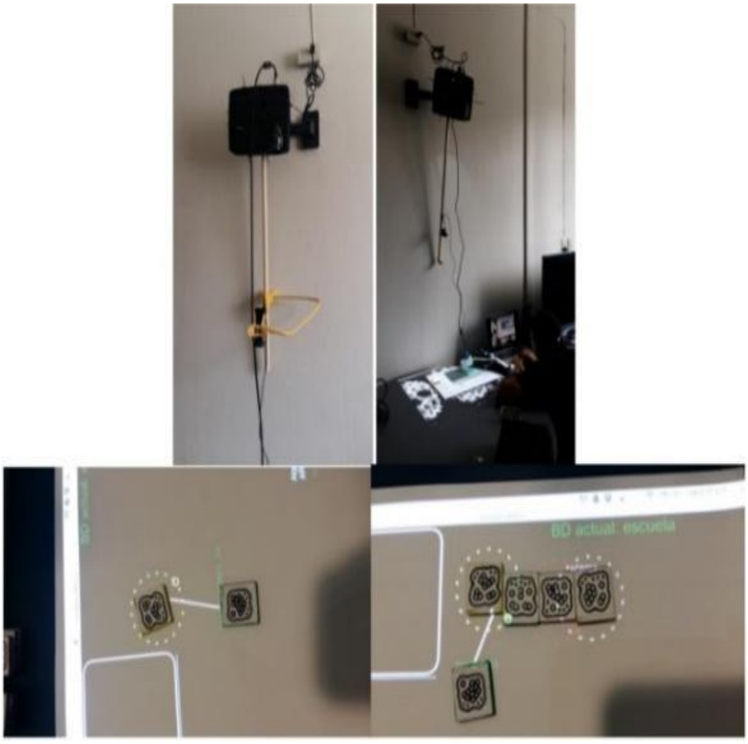

Fig. 1. Implementation of TanQuery prototype

Transformation. This phase transforms the interaction data into user behaviors considering an interval. This phase considers a model of the student's behavior during the activity allowing, based on the presence or absence of activities in an observation period, to identify whether the student has On-Task or Off-Task behavior. Once the behavior is defined, the same behaviors in contiguous intervals generate behavior sequences with different durations. Furthermore, during this phase, the data is preprocessed as input for the GSP algorithm.

Pattern recognition and evaluation. With a sequential algorithm, a process to identify the student behavioral patterns is carried out to identify representative behavioral sequences of each performance quadrant and train a classifier to condense the sequences of student behavior so that the intelligent environment can readily apprehend similarities and dissimilarities.

4. Data extraction and processing

For this study, a group of 16 students (12 male, 4 female) with similar ages (Max: 24 y/o, Min: 20 y/o, Average: 21 y/o), and basic relational database knowledge, all the participants were enrolled in the fifth semester of undergraduate studies. Also, and a teacher (Male with 14 years of experience) from the educational program related to information technology participated voluntarily (using a non-probabilistic sampling scheme) from a university in the southeast of Mexico, they were enrolled in the Databases subject. The participants were asked to participate in an open call, in which the evaluation process and the instruments to be used are explained. Based on the Belmont Report [47] on interaction with human beings, they were asked to request signing a letter as a requirement of consent. The study sample included participants whose reported age was 20 years or older, who had the same credit advancement in the Bachelor of Computational Technologies from the same class, and who were enrolled in the educational experience databases.

All participants interacted individually to solve relational algebra exercises using the TanQuery prototype, while a system stores the interaction data in a database table with the user-object interaction record structure. The connection to the database management system and the insertion in the online records are made each time the user interacts with the objects within the intelligent environment. It is sought that these are always available so that any component of the system and the environment can use them if required. The organization of these data is according to the user who generates them and the session in which they execute the activities. Table 1 illustrates an example of user-object interaction records.

Table 1. Example of User-Object interaction records

Name	Example	Example	Example	Example
ID	8	20	24	45
Type	Relation	Attribute	Attribute	Operator
X position	865	1080	1171	993
Y position	692	864	937	795
Value	STUDENT_INF	NOM_EST	EMAIL_EST	PROYECTION
Rotating degrees	0	358	355	86
Action	updateTUIObj	updateTUIObj	updateTUIObj	updateTUIObj
Root	PROYECTION	PROYECTION	PROYECTION	PROYECTION
Right child	-	-	-	-
Left child	-	-	-	8
Parent node	45	45	20	-
Attributes	-	24	-	20
Date	06/04/17	06/04/17	06/04/17	06/04/17
Time	12:52:57	12:52:57	12:52:57	12:52:57

5. Transformation

Once the dataset is created, the system finds the ObservedBehavior, Interval, StartTime, and EndTime columns. The possible intervals are obtained considering the analysis interval. For this work, an interval of 7 seconds is configured. Based on the interaction record and the number of possible intervals, an iterative process is carried out that obtains the records generated by the student to determine if he manifests an On-Task behavior or an Off-Task behavior.

After the analysis is finished, is possible to know the observed behavior, the analysis interval, and the moment of the student's interaction in which the interval begins and ends; an example is shown in Table 2.

Subsequently, the user-object interaction data for each student is recovered to identify the On-Task and Off-Task behaviors manifested during the activity to find sequences of student behavior. Also, the characteristics of the activity and the student's result are stored manually in the database to calculate the student's performance in the activity.

To execute the analysis, the behavior sequence structure IS modified by calculating the duration, in intervals, of the behavior and assigning the prefix ONT for On-Task behavior and OFFT for Off-Task behavior. For example, the behavior sequence of student 1, where he manifests the On-Task behavior from intervals 9 to 14, was represented as ONT6, and the Off-Task behavior during interval 15 was represented as OFFT1. To identify the student's performance, the number of hits obtained by the student is recovered. The total time in the activity is calculated by subtracting the time of the last record and the time of the first record. From the activity information, the Minimum Expected Grade (MEG), defined as the minimum score to pass the activity, and the maximum time to develop the activity are recovered (both defined by the professor); these values serve as parameters of the defineStudentTaskPerformance function [44].

Table 2. Identification of student behavior intervals

Observed behavior	Interval	Start time	endTime
On-Task	1	06/04/2017 12:48:07	06/04/2017 12:48:37
On-Task	2	06/04/2017 12:48:37	06/04/2017 12:49:07
On-Task	3	06/04/2017 12:49:07	06/04/2017 12:49:37
On-Task	4	06/04/2017 12:49:37	06/04/2017 12:50:07
On-Task	5	06/04/2017 12:50:07	06/04/2017 12:50:37
On-Task	6	06/04/2017 12:50:37	06/04/2017 12:51:07
Off-Task	7	06/04/2017 12:51:07	06/04/2017 12:51:37
On-Task	8	06/04/2017 12:51:37	06/04/2017 12:52:07
On-Task	9	06/04/2017 12:52:07	06/04/2017 12:52:37
On-Task	10	06/04/2017 12:52:37	06/04/2017 12:53:07
On-Task	11	06/04/2017 12:53:07	06/04/2017 12:53:37
On-Task	12	06/04/2017 12:53:37	06/04/2017 12:54:07
On-Task	13	06/04/2017 12:54:07	06/04/2017 12:54:37

The student performance quadrant scheme allows to classify the students considering the activity's results and time. Each quadrant represents the confluence between the time developed in the activity and the result obtained according to the following:

- Quadrant A: the grade obtained by the student is greater than the MEG, and the time it took to complete the task could be more optimal.
- Quadrant B: the grade obtained by the student is greater than the MEG, and the time it took her to complete the task is greater than optimal.
- Quadrant C: the grade obtained by the student is below the MEG, and the time it took her to complete the task could be more optimal.
- Quadrant D: the grade obtained by the student is below the MEG, and the time it took her to complete the task is greater than optimal.

6. Pattern recognition and evaluation

The objective of this stage is to calculate a classification model from the data related to the student's behaviors. Since the sequences have different lengths for each student (see Fig. 2), it was necessary to characterize all the sequences in the same way. One possible way was to characterize each sequence by the number of On-Task and Off-Task behaviors present in the sequence; however, with this way of doing it, the temporary nature of the sequences was lost. As an alternative, we decided to characterize a sequence by subsequences, which we call motifs. Thus, all the possible motifs were calculated from all the sequences of students' behaviors and, later, each sequence was characterized in terms of the presence or absence of a motif in it, thus representing all the sequences in a homogeneous way. In the end, each student is represented by characterizing their sequence in terms of motifs plus the performance quadrant to which they belong. The data represented in this way were analyzed using machine learning techniques to arrive at the classification model. Section 6.1 details the characterization of the sequences in terms of motifs and section 6.2 explains the analysis using machine learning techniques.

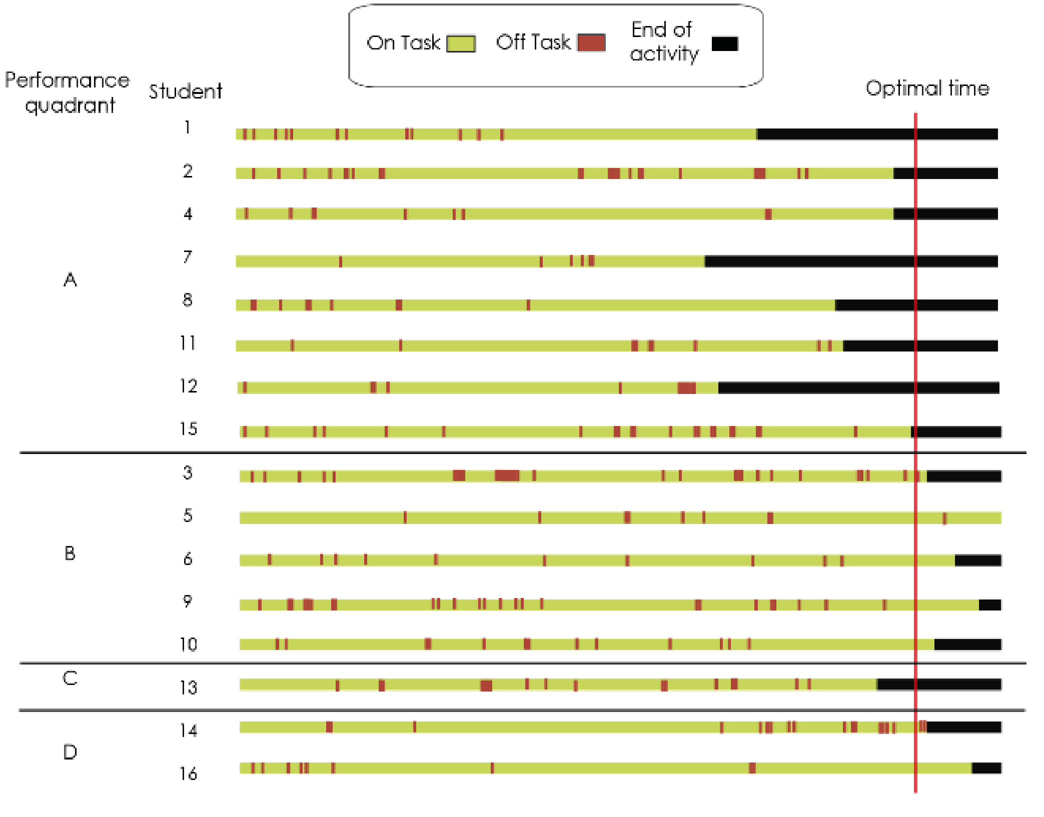

Fig.2. Participant's behavioral sequences

6.1 Characterization of sequences

Once having represented the behavioral sequences considering the performance quadrant and student behavior, the behavioral sequences were ordered alphabetically to assign them a numerical value, shown in Table 3, and they were substituted in the sequences of each participant. The set of behavior sequences for each quadrant was sent as a parameter to the GSP algorithm, and the sequential patterns and support for each one was obtained. The complete number sequences of quadrants A, B, C, and D students were 213.

Table 3. Example sequence of student 1's behavior representation (Rep)

Participant	Interval		Behavior	Rep..	idSec
	Begin	End			
Student1	1	2	On-Task	ONT2	22
Student1	3	4	Off-Task	OFFT2	3
Student1	5	6	On-Task	ONT2	22
Student1	7	8	Off-Task	OFFT2	5
Student1	9	14	On-Task	ONT6	51
Student1	15	15	Off-Task	OFFT1	1
Student1	16	18	On-Task	ONT3	30
Student1	19	20	Off-Task	OFFT2	3
Student1	21	37	On-Task	ONT17	19
Student1	38	38	Off-Task	OFFT1	1
Student1	39	41	On-Task	ONT3	30

The Minimum Support Value (MinSup) represented as a percentage of occurrence (support) of the transactions containing all the pattern items in the GSP algorithm was set to 50%. For quadrant A, 56 sequences were found with support equal to or greater than four and equal to or less than eight. In quadrant B, 74 sequences with support equal to or greater than three and equal to or less than five were found. In performance quadrants C and D, no sequences with support equal to 50% were found. However, 83 sequences with a support value equal to two were found. It is observed that the patterns with a higher support number contain fewer behavioral sequences, and those with a lower support number are made up of more complex sequences.

An example of the behavioral patterns of quadrant A is shown in Table 4. It can be noted that the shortest patterns contain only one behavior sequence, and the longest contains five. The highest support found in the patterns is eight. The patterns with this support manifest the OFFT1 sequence, which is the presence of the Off-Task behavior during a seven-second interval at different times of the activity, but no more than three times during the entire activity. Sequences with a support of 6 involve the behavior Off-Task with a longer duration (OFFT2) or the presence of OFFT1 four times during the entire activity. Behavior patterns with less support (minimum four) are also observed that involve the occurrence of sequences of behavior Off-Task for two or more periods of seven seconds (OFFT2, OFFT3) and the presence of behaviors On Task and Off-Task at time intervals greater than 21 seconds (ONT3, ONT4). Considering the scope of application, this is interpreted as students in this quadrant spending very little time outside of the activity and combining short periods of inactivity with periods of activity of intermediate duration. Most of the time, they interact and say they are doing the activity continuously. In other words, they spend the most time doing something.

Table 4. Example of behavioral patterns found in quad-A

Sequence	Support
OFFT1,	8.0
OFFT1, OFFT2 ,	6.0
OFFT1, OFFT1, ONT3,	5.0
ONT2, OFFT2, OFFT1, OFFT1,	4.0

Table 5 summarizes the results of measuring the student's performance and identifying the behaviors, showing the occurrence of behaviors (On-Task and Off-Task) manifested by each student during the interaction session grouped by performance quadrant.

In quadrant B, the shortest patterns contain one behavior sequence, while the longest contains ten. The highest support found is five for the pattern with a single sequence of behavior and three for the pattern with the most extended sequence. The patterns with longer sequences and support three are shown below. In them, it is observed that the OFFT1 behavior sequence predominates, and it is observed in the pattern with a length of eight that this Off Task behavior sequences occur after the student manifests an On-Task behavior for 28 seconds. (ONT4). It is interpreted that the students of quadrant B, during the activity, manifest a more significant amount of dead time. When they return to do it, they do so for more extended periods. Compared to the students in performance quadrant A, the students in performance quadrant B interrupt a more significant number of times, for short periods, performing the activity. After that, they begin to perform it for a longer time. In other words, they tend to manifest periods of inactivity more frequently, followed by periods of activity periodically, causing them to take longer to perform. However, they could have answered the activity exercises correctly.

Table 5. Student behaviors and performance quadrant

Participant	Performance Quadrant	On-Task	Off-Task
Student1	A	183	14
Student2	A	220	30
Student3	B	227	34
Student4	A	241	9
Student5	B	284	3
Student6	B	262	10
Student7	A	171	6
Student8	A	215	12
Student9	B	253	27
Student10	B	250	14
Student11	A	218	11
Student12	A	170	13
Student13	C	224	18
Student14	D	245	18
Student15	A	234	23
Student16	D	269	9

In quadrant C, no patterns with support more significant than three were found because only one student was assigned in this quadrant based on his performance in the activity. It is necessary to have more students be able to identify sequences of behavior in common.

In quadrant D, all the patterns found have support two and range in length from one to eight behavior sequences. The most representative sequence in most patterns is ONT9, which means that the student manifested On Task behavior for 63 consecutive seconds during the activity. It is essential to mention that students from quadrants A and B do not manifest this sequence. In other words, students in the F quadrant spend more time in short periods of inactivity than doing the task and occasionally tend to resume the activity for long periods.

An analysis of the sub-sequences found from quadrants A and B was performed to allow the system to identify the performance quadrant based on the sequence of behaviors manifested during the activity. For this sequential analysis, the authors considered the works of [48-49].

The procedure for the analysis of each student sequence is as follows:

- The sub-sequence identifier algorithm finds all quadrants' sub-sequences in the sequence.
- Each sub-sequence occurrence is stored, a matrix is generated, and this matrix is sent to the classifier.

The sequences of the students are those identified when the atomic behaviors manifested during the study session were coupled. The sub-sequences used for this exercise are those found by the GSP algorithm from performance quadrants A and B. The sub-sequences of quadrants C and D weren't considered due to the small number of students who, according to the calculation of their performance, have 1 and 2 students, respectively.

Table 6 presents an example of sub-sequence organization considering the motif approach. Subsequently, the sub-sequences found were counted for each student (13 students, eight from quadrant A and five from quadrant B). The students from C and D quadrants were not considered because their GSP patterns were not representative due to the small number of students belonging to those quadrants. Subsequently, these were organized, considering the structure of the motifs of the article [48] and the analysis process [48-49], resulting in the information being organized as presented in Table 7. A total of 128 sub-sequences correspond to the performance quadrants A and B after removing two repeated sequences in both performance quadrants.

Table 6. Example of sub-sequences and motif organization

Sub-sequence	Motif
OFFT1	motif1
OFFT2	motif2
OFFT3	motif3
ONT2	motif4
…	
OFFT1, OFFT1, OFFT1, OFFT1, OFFT2	motif53
OFFT1, OFFT1, OFFT1, ONT3, OFFT1	motif54
OFFT1	motif55
OFFT2	motif56

Table 7. Example of motif(m) occurrence in each student (St) sequence

St	m1	m2	m3	m4	m8	m10	m13
1	6	4	0	2	15	20	12
2	8	3	2	2	28	7	3
3	14	1	1	1	91	3	10
4	5	2	0	1	10	3	5
5	3	0	0	0	3	0	0
6	10	0	0	0	45	0	0
7	4	1	0	1	6	0	0
10	8	3	0	1	28	12	7
11	5	0	2	0	10	0	0
12	3	1	0	0	3	2	0
15	9	4	2	1	36	6	6

6.2 Analysis of sub-sequences of student behavior

Once the occurrence of each behavior sub-sequence in each sequence of the 13 students was recovered, two students were randomly drawn. The remaining 11 were used to train the classifier using the Orange Data Mining tool, version 3.30. 1. The data on sub-sequence occurrence from the two students separated from the original group were used to assess whether the classification was adequate. The results of the different classification models in the Orange Data Mining tool are shown in Table 8. There we can see that of the four selected models (Neural Network, Random Forest, Naive Bayes, Tree), the two models that had the best results, despite the small data, were the Neural Network and Naive Bayes.

Various feature elimination methods were used to know the most representative sub-sequences of quadrants A and B. The authors used ReliefF and Information Gain as scoring methods. The results are shown in Table 9.

One of the approaches to identify the performance quadrant while the student is performing the activity is through decision trees, defined by [44] as "a prediction model that, given a set of data, diagrams of logical constructions are made, which serve to represent and categorize a series of conditions that occur successively, for the resolution of a problem".

Table 8. Evaluation results

Model	AUC	CA	F1	Precision	Recall
Tree	0.214	0.455	0.398	0.354	0.455
Random Forest	0.143	0.273	0.285	0.315	0.273
Neural Network	0.500	0.636	0.636	0.636	0.636
Naive Bayes	0.500	0.636	0.636	0.636	0.636

Table 9. Best-scored sub-sequences using the Scoring Methods (SM): ReliefF (RF) and Information Gain (IG)

SM	Id	Sub-sequence	Value
RF	58	ONT20	0.192
RF	67	ONT10, ONT4	0.178
RF	80	ONT4, ONT10, OFFT1	0.163
IG	58	ONT20	0.468
IG	67	ONT10, ONT4	0.468
IG	80	ONT4, ONT10, OFFT1	0.370

Finally, to identify in an intelligent educational environment, at runtime, the student's performance quadrant based on the interaction with objects, it is necessary to find a set of rules that, incorporated into the ambient intelligence system, allow personalized support without waiting to finish the activity. A decision tree and the data set of the 11 previously selected students were used to find them.

Fig. 3 shows the resulting decision tree, from which the following rules are identified:

- If (motif1 > 9) = QuadrantedperformanceB.
- If (motif1 ≤ 9)&(motif58 ≤ 0) = QuadrantedperformanceA.
- If (motif1 ≤ 9)&(motif58 > 0) = QuadrantedperformanceB.

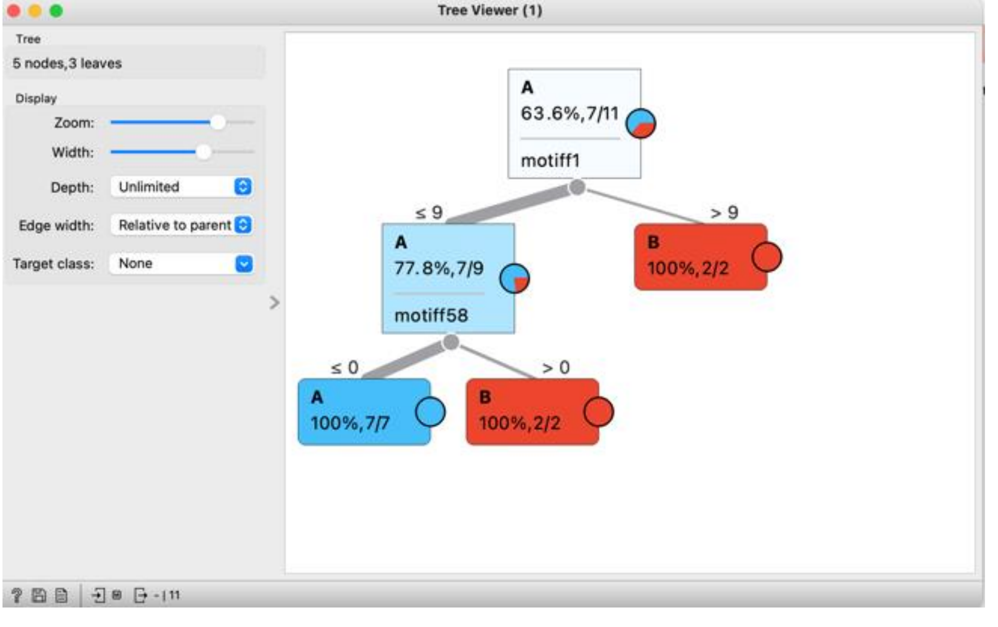

Fig. 3. Decision tree

The rules obtained from the decision tree are likely used so that, during the activity, the intelligent environment can generate personalized interventions that improve student performance. For example, if it is identified from the interaction records (that the student generates during the activity) that the student may result in low performance, the environment would intervene to help him improve his final performance in a personalized way.

7. Discussion

Including personalization mechanisms in an intelligent educational environment will allow the environment or educational systems to execute prompt interventions by identifying students in at-risk situations and aiding based on teaching-learning experiences [50].

To achieve this, it is necessary to consider contextual information about the student (academic and personal) to identify the student's learning needs, why they need to learn it, and the most appropriate way to learn it. With this, it is possible to design a personalized learning path for each student that, through a non-intrusive follow-up supported by technology, would allow the identification of deviations in those paths and, therefore, students in risk situations.

Knowing the student's behaviors while performing the activity would allow the intelligent educational environment to improve student performance by helping them move from a C or D quadrant, which performs below the desired level, to an A or B. An intelligent educational environment would provide support materials or interventions on the activity's topics.

The process described in this work can be helpful to incorporate personalization mechanisms into educational environments and to enhance customization services or applications like affective feedback systems [51], identifying the competency level of students [52], monitoring students while writing academic tasks [53], and learning monitoring and customization in mobile learning platforms [54]. To analyze data from user interaction with objects or with educational software and provide customized learning materials for each student considering their strengths, needs, skills, interests, and academic performance.

8. Conclusions and future work

The proposal addressed in this work contributes to identifying user behaviors through monitoring the user-object interactions manifested when performing academic activities and monitoring through a non-intrusive observation of the activities carried out in intelligent educational environments. Mainly, when addressing the case study, it is intended to contribute to the improvement of teaching-learning processes by supporting its users by identifying their behaviors and associating them with the performance of the academic task they are performing to identify needs presented by students or to identify ways to improve their performance in a university intelligent learning environment setting.

It was necessary to analyze recent works on intelligent environments, ubiquitous computing, and ambient intelligence that address identifying user behaviors and those applied in the educational field. This topic is partially covered in the educational field, so it was in our interest to work on it.

During the data-collection process, light conditions changed from natural to artificial to avoid problems related to the detection of markers. Also, a local database server was set up to avoid misconnection and data-storage problems.

The test results showed that it is possible to measure the student's performance quadrant and identify her performance by analyzing her interactions with objects. Also, a higher concentration of students is observed in the A and B performance quadrants, and fewer students in the C and D quadrants. However, this is an unexpected result due to the balance of the quadrants for the classification of students and subsequent discovery of patterns; it is a positive result that indicates that the students, for the most part, perform well in terms of having a higher-than-expected result in the activity. Most students exhibited On Task behavior for most of the activity; it is possible to interpret this as a greater involvement of the participants in the task. Nevertheless, it is possible to develop further studies in a controlled environment where the student is in a more familiar environment for him and with more freedom.

Although it is not possible to generalize the results from such a limited sample, observable indications and differences are obtained in the development of the activity of the students in each of the quadrants. Students in quadrants A and C show more behavior Off Task at the beginning and end of the activity, while students in quadrant B show this behavior throughout the development of the activity. In the two students in performance quadrant D, the development of the activity is different.

Concerning the use of observed behavior patterns, identifying the occurrence of sub-sequences in each behavior sequence of the students in quadrants A and B was carried out due to having a more significant number of students. Using the Orange Data Mining tool and the classification models Neural Network, Random Forest, Naive Bayes, and Tree, training, and testing were done with the data of the 13 students (11 for training and two for testing).

The number of participants for this work is small and it is not possible to generalize results. Due to this very limited number of data, the performance of the classifiers is low and can be improved by increasing the sample. Therefore, the results obtained are not conclusive, so more experiments with a larger number of participants will be needed to obtain more reliable results.

Future work considers the evaluation with a larger number of participants, also the incorporation of the analysis process into a system to automatically identify student behavior and performance during the activity to provide customized assistance during the learning process.

References

[1]. G. Virone et al., Behavioral Patterns of Older Adults in Assisted Living, in IEEE Transactions on Information Technology in Biomedicine, vol. 12, no. 3, pp. 387-398, May 2008, doi: 10.1109/TITB.2007.904157.

[2]. Pantic Maja, et al. Human computing and machine understanding of human behavior: A survey. Proceedings of the 8th international conference on Multimodal Interfaces. 2006.

[3]. Favela Jesus. Activity, Behavior, and Context: The ABC of Pervasive Healthcare Research. Intelligent Environments (Workshops). 2012.

[4]. Nef Tobias et al. Evaluation of three state-of-the-art classifiers for recognition of activities of daily living from smart home ambient data. Sensors 15.5 (2012): 11725-11740.

[5]. Vinh La The, et al. Semi-Markov conditional random fields for accelerometer-based activity recognition. Applied Intelligence 35 (2011): 226-241.

[6]. Moore Dennis W. et al. Increasing on-task behavior in students in a regular classroom: Effectiveness of a self-management procedure using a tactile prompt. Journal of Behavioral Education 22 (2013): 302-311.

[7]. Godwin Karrie E. et al. Off-task behavior in elementary school children. Learning and Instruction 44 (2016): 128-143.

[8]. Doddannara Lakshmi S. et al. Exploring the relationships between design, students' affective states, and disengaged behaviors within an ITS. Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings 16. Springer Berlin Heidelberg, 2013.

[9]. Wang Li-Chun, Ming-Puu Chen. The effects of game strategy and preference-matching on flow experience and programming performance in game-based learning. Innovations in Education and Teaching International 47.1 (2010): 39-52.

[10]. Schaufeli Wilmar B., Arnold Bakker. UWES Utrecht work engagement scale preliminary manual. Occupational Health Psychology Unit (2003).

[11]. Bumbacco C., Scharfe E. (2023). Why attachment matters: first-year post-secondary students' experience of burnout, dIsengagement, And drop-out. Journal of College Student Retention: Research, Theory & Practice, 24(4), 988-1001.

[12]. Zhang J., Mohd Yunos Z, Haron H. Interactivity Recognition Graph Neural Network (IR-GNN) Model for Improving Human–Object Interaction Detection. Electronics. 2023; 12(2):470. https://doi.org/10.3390/electronics12020470

[13]. Qing Ye, Xiuju Xu, Human-object interaction detection based on graph model, Proc. SPIE 12610, Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022), 126100A (28 April 2023); https://doi.org/10.1117/12.2671248
[14]. Mazzamuto Michele, et al. A wearable device application for human-object interactions detection. International Conference on Computer Vision Theory and Applications (VISAPP). 2023.
[15]. Luo Tianlun, et al. From detection to understanding: A survey on representation learning for human-object interaction. Neurocomputing (2023): 126243.
[16]. Kim, Sanghyun, Deunsol Jung, Minsu Cho. Relational Context Learning for Human-Object Interaction Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
[17]. Ye Qing, et al. Human object interaction detection based on feature optimization and key human-object enhancement. Journal of Visual Communication and Image Representation 93 (2023): 103824.
[18]. Zidianakis Emmanouil, et al. Building a sensory infrastructure to support interaction and monitoring in ambient intelligence environments. Distributed, Ambient, and Pervasive Interactions: Second International Conference, DAPI 2014, Held as Part of HCI Interational 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings 2. Springer International Publishing, 2014.
[19]. D'Mello Sidney K. et al. Multimodal capture of teacher-student interactions for automated dialogic analysis in live classrooms. Proceedings of the 2015 ACM on international conference on multimodal interaction. 2015.
[20]. Belapurkar Neha, Sagar Shelke, Baris Aksanli. The case for ambient sensing for human activity detection. Proceedings of the 8th International Conference on the Internet of Things. 2018.
[21]. Jayarajah Kasthuri et al. Need accurate user behaviour? pay attention to groups!. Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing. 2015.
[22]. Baker Ryan Shaun et al. Off-task behavior in the cognitive tutor classroom: When students "Game the system". Proceedings of the SIGCHI conference on Human factors in computing systems. 2004.
[23]. Baker Ryan. Modeling and understanding students' off-task behavior in intelligent tutoring systems. Proceedings of the SIGCHI conference on Human factors in computing systems. 2007.
[24]. Leonidis Asterios et al. An intelligent task assignment and personalization system for students' online collaboration. Universal Access in Human-Computer Interaction. Applications and Services: 6th International Conference, UAHCI 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part IV 6. Springer Berlin Heidelberg, 2011.
[25]. Sanish Rai, Xiaolin Hu. Behavior pattern detection for data assimilation in agent-based simulation of smart environments. 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). Vol. 2. IEEE, 2013.
[26]. Wang Qunbo, Wenjun Wu, Yuxing Qi. A Learning Analytic Model for Smart Classroom. Web and Big Data: APWeb-WAIM 2018 International Workshops: MWDA, BAH, KGMA, DMMOOC, DS, Macau, China, July 23–25, 2018, Revised Selected Papers 2. Springer International Publishing, 2018.
[27]. Korozi Maria et al. LECTOR: towards reengaging students in the educational process inside smart classrooms. Intelligent Human Computer Interaction: 9th International Conference, IHCI 2017, Evry, France, December 11-13, 2017, Proceedings 9. Springer International Publishing, 2017.
[28]. White Ryen W., Ahmed Hassan Awadallah, Robert Sim. Task completion detection: A study in the context of intelligent systems. Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2019.
[29]. Montebello Matthew. Assisting education through real-time learner analytics. 2018 IEEE Frontiers in Education Conference (FIE). IEEE, 2018.
[30]. Matsui Kanae, Tatsuhiko Kasai, Keiya Sakai. Challenges for data collecting of teacher and student'behavior in different types of class using video and wearable device. 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR). IEEE, 2019.
[31]. Prabono Aria Ghora, Seok-Lyong Lee, Bernardo Nugroho Yahya. Context-based similarity measure on human behavior pattern analysis. Soft Computing 23 (2019): 5455-5467.
[32]. Biswas Gautam, Brian Sulcer. Visual exploratory data analysis methods to characterize student progress in intelligent learning environments. 2010 International Conference on Technology for Education. IEEE, 2010.
[33]. G. Mathioudakis et al. Amiria: Real-time teacher assistance tool for an ambient intelligence classroom. Proceedings of the Fifth International Conference on Mobile, Hybrid, and Online Learning (eLmL 2013). 2013.

[34]. Barbara Bruno et al. Analysis of human behavior recognition algorithms based on acceleration data. 2013 IEEE International Conference on Robotics and Automation. IEEE, 2013.
[35]. Valérian Guivarch et al. Hybrid system to analyze user's behaviour. 2016 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2016.
[36]. Louis Atallah, Guang-Zhong Yang. The use of pervasive sensing for behaviour profiling – a survey. Pervasive and mobile computing 5.5 (2009): 447-464.
[37]. Zacharoula Papamitsiou, Eirini Karapistoli, Anastasios A. Economides. Applying classification techniques on temporal trace data for shaping student behavior models. Proceedings of the sixth international conference on learning analytics & knowledge. 2016.
[38]. Valeria Soto-Mendoza, et al. Detecting abnormal behaviours of institutionalized older adults through a hybrid-inference approach. Pervasive and Mobile Computing 40 (2017): 708-723.
[39]. Andreas D. Lattner, et al. Sequential pattern mining for situation and behavior prediction in simulated robotic soccer. RoboCup 2005: Robot Soccer World Cup IX 9. Springer Berlin Heidelberg, 2006.
[40]. Magnus S. Magnusson. Discovering hidden time patterns in behavior: T-patterns and their detection. Behavior research methods, instruments, & computers 32.1 (2000): 93-110.
[41]. Maurizio Casarrubea et al. T-pattern analysis for the study of temporal structure of animal and human behavior: a comprehensive review. Journal of neuroscience methods 239 (2015): 34-46.
[42]. Carlos Santoyo et al. Observational analysis of the organization of on-task behavior in the classroom using complementary data analyses. Anales de Psicología, 2017, vol. 33, num. 3, p. 497-514 (2017).
[43]. Ramakrishnan Srikant, Rakesh Agrawal. Mining sequential patterns: Generalizations and performance improvements. Advances in Database Technology—EDBT'96: 5th International Conference on Extending Database Technology Avignon, France, March 25–29, 1996 Proceedings 5. Springer Berlin Heidelberg, 1996.
[44]. Hernández-Calderón, José Guillermo et al. A System to Match Behaviors and Performance of Learners From User-Object Interactions: Model and Architecture. International Journal of Information Technologies and Systems Approach (IJITSA) 12.2 (2019): 82-103.
[45]. Magdalena Cantabella et al. Analysis of student behavior in learning management systems through a Big Data framework. Future Generation Computer Systems 90 (2019): 262-272.
[46]. José Antonio Xohua-Chacón, Edgard Iván Benítez-Guerrero, Carmen Mezura-Godoy. A tangible system for learning relational algebra. Revista Colombiana de Computación 19.1 (2018): 39-55.
[47]. United States. National Commission for the Protection of Human Subjects of Biomedical, and Behavioral Research. The Belmont report: ethical principles and guidelines for the protection of human subjects of research. Vol. 1. Department of Health, Education, and Welfare, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978.
[48]. Krisztian Buza, Lars Schmidt-Thieme. Motif-based classification of time series with bayesian networks and svms. Advances in Data Analysis, Data Handling and Business Intelligence: Proceedings of the 32nd Annual Conference of the Gesellschaft für Klassifikation eV, Joint Conference with the British Classification Society (BCS) and the Dutch/Flemish Classification Society (VOC), Helmut-Schmidt-University, Hamburg, July 16-18, 2008. Springer Berlin Heidelberg, 2010.
[49]. Jia-Min Ren, Jyh-Shing Roger Jang. Time-constrained sequential pattern discovery for music genre classification. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011.
[50]. Vanessa Alejandra Camacho Vázquez. Detección de emociones negativas en las redes sociales., Ph.D. Thesis, 2017.
[51]. Samantha Jiménez, R. Juárez-Ramírez, Víctor H Castillo, Alan Ramírez Noriega. The Role of Personality in Motivation to use an Affective Feedback System. Program Comput Soft 47, 793–802 (2021). https://doi.org/10.1134/S0361768821080156.
[52]. J. R. Aguilar-Cisneros, R. Valerdi, B. P. Sullivan. Students' Systems Thinking Competency Level Detection through Software Cost Estimation Concept Modeling. Program Comput Soft 48, 499–512 (2022). https://doi.org/10.1134/S0361768822080060.
[53]. Kochegurova E. A., Martynova Y. A. Aspects of Continuous User Identification Based on Free Texts and Hidden Monitoring. Program Comput Soft 46, 12–24 (2020). https://doi.org/10.1134/S036176882001003X.
[54]. H. Del Ángel-Flores, E. López-Domínguez, Y. Hernández-Velázquez et al. Usability Evaluation of a Mobile Learning Platform Focused on Learning Monitoring and Customization based on a Laboratory Study. Program Comput Soft 48, 583–597 (2022). https://doi.org/10.1134/S0361768822080102.

Информация об авторах / Information about authors

Хосе- Гильермо ЭРНАНДЕС-КАЛЬДЕРОН имеет степень PhD мексиканского Университета Веракруса в области компьютерных наук. Профессор факультета статистики и информатики этого университета. Сфера научных интересов: человеко-машинное взаимодействие, искусственный интеллект, совместные вычисления и визуализация.

José-Guillermo HERNÁNDEZ-CALDERÓN – Ph. D. in Computer Science from the University of Veracruz in México. Professor at the Faculty of Statistics and Informatics of the University of Veracruz in Mexico. Research interests: Human Computer Interaction, Artificial Intelligence, Collaborative Computing, and Visualization.

Эдгар Иван БЕНИТЕС-ГЕРРЕРО – доктор компьютерных наук Гренобльского университета во Франции. Профессор факультета статистики и информатики Университета Веракруса в Мексике. Научные интересы: взаимодействие человека с компьютером, искусственный интеллект, совместные вычисления, управление данными и визуализация.

Edgard Ivan BENÍTEZ-GUERRERO – Ph. D. in Computer Science from the University of Grenoble in France. Professor at the Faculty of Statistics and Informatics of the University of Veracruz in Mexico. Research interests: Human Computer Interaction, Artificial Intelligence, Collaborative Computing, Data Management and Visualization.

Хосе Рафаэль РОХАНО-КАСЕРЕС имеет степень PhD Технологического института фундаментальных исследований Доктор наук в области компьютерных наук в Институте технических наук, кампус Куэрнавака, Монтеррей. Также имеет степень магистра по искусственному интеллекту Университета Веракрузана. В настоящее время штатный профессор факультета статистики и информатики университета Веракрузана. Сфера научных интересов: семантическая глобальная информационная сеть, инструменты для совместной работы и изучение проблем нетрудоспособных людей.

José Rafael ROJANO-CÁCERES – PhD in computer science from the Instituto Tecnológico de Estudios Superiores Monterrey campus Cuernavaca. Master in Artificial Intelligence from the Universidad Veracruzana. Currently Full Time Professor at the Faculty of Statistics and Informatics of the Universidad Veracruzana. His research areas include semantic web, collaborative tools, and disability.

Кармен МЕСУРА-ГОДОЙ получила степень доктора компьютерных наук в Университете Савойи во Франции. Профессор факультета статистики и информатики Университета Веракруса в Мексике. Основные исследовательские интересы: человеко-машинное взаимодействие, совместные вычисления, визуализация и мультиагентные системы.

Carmen MEZURA-GODOY – PhD in Computer Science from the University of Savoie in France. Professor at the Faculty of Statistics and Informatics of the University of Veracruz in Mexico. Main research interests: Human-Computer Interaction, User eXperience-UX, Computer Support Collaborative Work, Visualization and Multiagent Systems.