IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 6, 2019, accepted July 2, 2019, date of publication July 10, 2019, date of current version August 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927860

User Experience Evaluation Using Mouse Tracking and Artificial Intelligence

KENNEDY E. S. SOUZA[1], MARCOS C. R. SERUFFO[2], HAROLD D. DE MELLO JR.[3], DANIEL DA S. SOUZA[4], AND MARLEY M. B. R. VELLASCO[5]

[1]Post-Graduate Program in Anthropic Studies in the Amazon, Federal University of Para, Castanhal 68740-222, Brazil
[2]Post-Graduate Program in Anthropic Studies in the Amazon, Federal University of Para, Belem 66075-110, Brazil
[3]Department of Electrical Engineering, Rio de Janeiro State University, Rio de Janeiro 20940-200, Brazil
[4]Post-Graduate Program in Electrical Engineering, Federal University of Para, Belem 66075-110, Brazil
[5]Post-Graduate Program in Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro 22451-900, Brazil

Corresponding author: Marcos C. R. Seruffo (seruffo@ufpa.br)

This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, in part by the University of Para (UFPA), in part by the National Council of Technological and Scientific Development (CNPq), and in part by the Carlos Chagas Filho Research Support Foundation (FAPERJ).

ABSTRACT Business platform models frequently require continuous adaptation and agility to allow new experiences to be created and delivered to customers. To understand user behavior in online systems, researchers have taken advantage of a combination of traditional and recently developed analysis techniques. Earlier studies have shown that user behavior monitoring data, as obtained by mouse tracking, can be utilized to improve user experience (UX). Many mouse-tracking solutions exist; however, the vast majority is proprietary, and open-source packages do not provide the resources and data needed to support UX research. Thus, this paper presents: 1) the development of an interaction monitoring application titled Artificial Intelligence and Mouse Tracking-based User eXperience Tool (AIMT-UXT); 2) the validation of the tool in a case study conducted on the Website of the Brazilian Federal Revenue Service (BFR); 3) the definition of a new relationship pattern of variables that determine user behavior; 4) the construction of a fuzzy inference system for measuring user performance using the defined variables and the data captured in the case study; and 5) the application of a clustering algorithm to complement the analysis. A comparison of the results of the applied quantitative methodologies indicates that the developed framework was able to infer UX scores similar to those reported by users in questionnaires.

INDEX TERMS User interfaces, computer science, ergonomics, artificial intelligence.

I. INTRODUCTION

According to ISO 9241-210 [1], user experience (UX) includes all the users' emotions, beliefs, preferences, perceptions, physical, and psychological responses, as well as behaviors and achievements, that occur before, during, and after use of a product or service. UX is, therefore, the direct representation of the human factor in the context of the development of products and services on digital platforms. UX has become an increasingly prominent aspect of systems development, following the evolution of business and process models.

Thus, data collection tools for user-application interactions (through techniques such as eye tracking and mouse tracking, for example) have become critical for the success of a Web service. Evaluation and monitoring systems are capable of providing statistical descriptions that allow usage patterns to be identified, so that they can be used as a reference for the development of systems, ranging from use recommendations, through product offer mechanisms, to data traffic predictions at the network management level. Allied with these techniques, computational intelligence can be used to correlate data and assist user behavior identification.

The analysis of user interaction in Web systems is an important premise for user satisfaction evaluation tools and may even support modifications to increase the UX level. In the present article, a systematic method of evaluating UX by using metrics obtained from mouse tracking in combination with computational intelligence techniques is proposed. Traditional methods for performing such assessments are

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Babu Thanikanti.

based on the analysis of the responses of a group of users to items on a satisfaction questionnaire. In the methodology proposed in the present article, an evaluation score consists of the performance parameters of users for task completion, which are directly correlated with the UX perception.

The results obtained in this research were compared with those of a classic UX evaluation method to verify the relationship between the UX that the user reports and the UX as measured using data collection and correlation tools. The main contribution is the development of a structure based on computational intelligence techniques that allows the UX to be inferred from user performance parameters. This structure is called the Artificial Intelligence and Mouse Tracking-based User eXperience Tool (AIMT-UXT). Secondary contributions include the development and application of a tool for collecting data from mouse-tracking data, a case study using the Website of the Federal Revenue Service of Brazil (FRSB), and a comparison of the scores of the methodologies used in the UX evaluation.

This article is organized as follows. Section II presents studies related to this research. Section III describes the developed tool, data collection, and analysis methods. In section IV, a case study and the analysis of the obtained results are presented. In section VI, the conclusion is presented.

II. RELATED WORK

This section presents the main research studies in the literature that support the development of the methods described in this article. The gaps that were found in UX evaluation methods in the literature review and that are considered in the proposed method are highlighted.

A. USER EXPERIENCE EVALUATION

Several means for evaluating UX for computer systems and Websites exist, and techniques involving cardiac monitoring, eye tracking, fixation of attention, etc. may be applied. The application of eye and mouse tracking has been investigated for almost 20 years.

In their study reported in [2], the authors identified a strong relationship between the position of the user's gaze and the position of the user's cursor on a computer screen during Web browsing. These results attest to the possibility of evaluating the UX exclusively from mouse tracking.

In [3], a Website evaluation tool called WebTracer was proposed, which can record eye movements, the operational data of a user, and the screen image of the pages visited through the use of eye and mouse tracking techniques, and can also reproduce navigation operations. In addition to confirming the findings reported in [2], the paper presents optimized techniques for capturing and transmitting data in terms of processing resources and network throughput.

In the study in [4], a tool was developed for recording all mouse movements on a Web page and used to analyze and investigate mouse usage trends and behaviors. The obtained results allowed content providers to increase the interface's design effectiveness.

The research reported in [5] showed the identification of user behaviors for UX prediction resulting from the analysis of mouse usage patterns. The presented results allow user frustration and attempt to read a text to be inferred with high precision.

In a similar cognitive approach for improving UX, in the study presented in [6] the authors recorded mouse patterns to understand the manner in which the user interacts with the design of sites. They concluded that differences in the content that users search on a site can result in large differences in the number of times users move the mouse.

As an additional approach for UX evaluation using the mouse-tracking technique, in [7] a solution for capturing the mouse movements of users on Web pages to identify areas of interest was proposed. The application was developed to process client server requests quickly and thus optimizes server resources.

In the literature review, it was found that the most recent UX evaluation studies used recurring commercial tools based on mouse tracking. These include MouseFlow,[1] HotJar,[2] and CrazyEgg.[3] However, in addition to service fees, such tools include an internal implementation method, that is, they require laborious adaptations for integration with the systems being studied. Furthermore, they do not allow access to the source code, a feature that restricts the depth of investigations. Conversely, non-proprietary tools, mostly *freeware*, have disadvantages that include limitations on the types of navigation data that can be tracked, the recording and playback of test sessions per user, and the number of sites and pages that can be tested. Some, such as MetricBuzz,[4] still host *scripts* on third-party sites, which can cause security issues and conflicts with secure sockets layer (SSL) certificates. Thus, the characteristic of these non-proprietary systems that definitively made it impossible to use them in our study is the fact that they execute simple analyses, usually involving only statistical descriptions of the data.

Although dozens of visual analysis tools exist, such as those mentioned above, many monitor only user session data. Our study included the development of our own solution with an effective focus on UX and user behavior on a site. AIMT-UXT includes heat map features and the recording of activities based on mouse tracking, which can be easily exported and subsequently served as a basis for the application of fuzzy logic complemented by a clustering algorithm in the UX evaluation. Although it is possible to identify in the literature a few studies in which fuzzy systems were used to measure UX [8]–[13], in none of these was a framework such as the present one developed and in only a few were the results compared with those of other methodologies.

[1] https://mouseflow.com/
[2] https://www.hotjar.com/
[3] https://www.crazyegg.com/
[4] https://www.metricbuzz.com/

B. EVALUATION MODELS USING QUESTIONNAIRES

In the literature review, references to UX evaluation methods based on questionnaires and applied in several areas were identified, according to [14].

In [15], the authors presented a comparison of the following five questionnaire methods, which they considered to be adaptable to evaluations for Websites: the System Usability Scale (SUS) [16], Questionnaire for User Interface Satisfaction (QUIS) [17], Computer System Usability Questionnaire (CSUQ) [18], Words [19], and a model described in [15], named Our Questionnaire. The comparison showed that, among the models tested, the SUS method achieved the highest accuracy level with the smallest number of samples.

In addition, the authors of [20] concluded that the SUS method is reliable and capable of jointly measuring learning and usability, which are directly correlated with the user's performance. In the study presented in [21], based on the analysis of approximately 1,000 results obtained with the SUS method, the authors determined the reliability and effectiveness of this method in terms of measuring the usability of a wide variety of products and services.

In the study reported in [22], data were collected from 262 users of applications for comparison with evaluations previously registered using the SUS method. The results showed that the previous application UX is reflected positively in the evaluation ascribed by the data collection tool. From the conclusions presented in the mentioned papers, it can be understood that the SUS method is a classical reference metric for UX evaluation. This motivated us to compare the results obtained by our UX framework with those obtained by the SUS method.

C. EVALUATION FACTORS AND ARTIFICIAL INTELLIGENCE

Data on users' interaction with computational systems contain relationships and implicit characteristics that can be discovered by applying artificial intelligence algorithms. In the last six years, such techniques, most frequently fuzzy logic models, have been applied in UX problems.

In a previous paper [23], a fuzzy logic model based on graphical interfaces was proposed to predict five levels of usability in applications. This model used three input variables, but the authors did not explain their measurement method and the results were not compared with those of other UX evaluation techniques. The authors of [24] adopted the same five levels to evaluate Websites and proposed a fuzzy model with five inputs, namely, Navigation, Page Composition, Effectiveness, Efficiency, and Satisfaction, obtained from three different sources, including questionnaires. The final results were compared with Webby Award data.

Addressing usability optimization as a user interface development process, the authors of [13], [25], [26] also utilized fuzzy logic. In [13], a framework that quantifies user interface usability by means of a Mamdani fuzzy system for selecting a transformation process that can generate semi-automatically a user interface having optimal usability was presented. Also using the Mamdani method, the authors of [26] quantified the conflicts among usability attributes, because the results of manual assessment of required usability factors can lead to critical ambiguities for the development of more appropriate and usable software systems. However, in the literature survey presented in [25] the five key factors that affect the development of e-commerce Websites were identified. The Websites' usability was measured by using the neuro-fuzzy inference system (ANFIS) together with questionnaires.

Specifically for enhancing the quality of mobile applications, in [27] 12 usability factors that were extracted from 10 usability evaluation models based on questionnaires, including the SUS, were examined. In this study, fuzzy association rules were generated from the results of the usability survey questionnaires and the patterns were used to obtain the knowledge from the users' experience to improve usability.

Adopting other datamining techniques, the authors of [28] proposed the use of a biclustering algorithm to extract information from the daily online activities of virtual campus users. The results showed that the knowledge extracted from log files with statistical measures helped to provide better usability and adaption to user preferences. For improving UX, in the study in [29] user context information was also extracted, but for mobile applications. The data extracted by Google Services API were applied to a non-informed classification technique for identifying navigation patterns and allowing the application to be adapted to the context of use and user characteristics.

More recently, artificial neural networks (ANNs) were applied in UX classification tasks. In particular, in the study in [30] the concept of Website similarity as perceived by users was explored with the goal of facilitating the reuse of good Web designs. The authors concluded that ANN models with user impression-related inputs had a greater effect on user-assessed similarity of Websites than those with user interface intrinsic inputs. In [31], software that was developed to collect selected metrics related to the visual complexity of Web pages was described. These metrics were used in training an ANN user behavior model to predict the users' subjective perception of the Webpages orderliness and complexity. In the authors' opinion, this approach can aid Web designers to produce Websites that attain higher levels of the users' subjective appreciation.

Although, as mentioned above, many different UX evaluation methodologies have been developed, the application of artificial intelligence techniques is still incipient. Gaps exist that we sought to resolve in the present study. Thus, the goal of this study was to provide a comprehensive UX assessment and analysis solution. We developed a tool based on mouse tracking, AIMT-UXT, that captures the user performance parameters and uses them in a fuzzy model to infer a score for each user. Then, an additional intelligent technique, a clustering algorithm, is applied to identify users with similar characteristics. This methodology allows the inference of the fuzzy system to be confirmed. Finally, the results were compared to those of a classic user evaluation method, the SUS questionnaire. An innovative framework was used to facilitate the comparison of results.

III. ARTIFICIAL INTELLIGENCE AND MOUSE TRACKING-BASED USER EXPERIENCE TOOL

The AIMT-UXT tool was developed to obtain the user's interactions with the mouse and subsequently to analyze them using computational intelligence techniques. The tool is composed of three modules: single-view, heatmap, and datafuzzy. The software allows the collection, organization, and processing of data through the architecture presented in Figure 1.

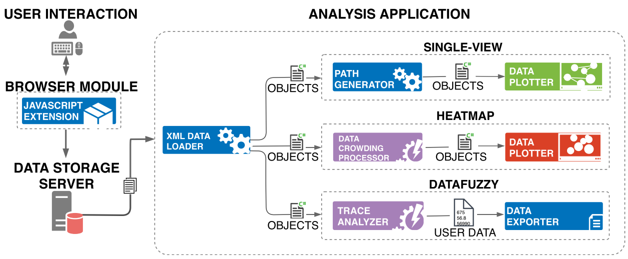

FIGURE 1. Artificial intelligence and mouse tracking-based user experience tool architecture.

This technology arrangement allows the flexibility obtained through the use of PHP and JavaScript, which allows data capture and storage modules to operate on a multi-platform basis, to be balanced with the capability of the data analysis modules to access proprietary native libraries and resources with high performance through C#.

As shown in Figure 1, the architecture is divided into three parts: the browser module, data storage server, and analysis application. In general, the interaction data are captured by the browser module, which groups the data and sends them to the storage server, responsible for transcribing, organizing, and storing the collected data. This data is used by the analysis application modules (single-view, heatmap, and datafuzzy) to generate information about system usage and user behavior. The modules are described in detail in the following sections.

A. BROWSER MODULE

The browser module is an extension of the Web browser that allows interaction data to be captured while the user performs a certain task. Then, user interactions with the mouse act as a trigger for recording the data of the accessed page, the movement of the mouse itself, and the page object with which the user is interacting. The recordings are performed at a minimum interval of 500 ms, defined empirically to allow bandwidth savings for data transmission and still preserve the accuracy of data capture.

Figure 2 illustrates the browser module's architecture, in which the data collection procedure performed by AIMT-UXT consists of three steps. The browser engine i) receives the data of the Web page to be accessed, ii) renders the interface, and iii) shows to the user the front end. After the completion of this process, the browser engine injects into the JavaScript code from the front end the functions necessary to capture user interaction with the interface. The data collected in the front end are sent to the data grouping system, where the data grouping and user unique identification and encoding in JSON are performed, for subsequent transmission through HTTP requests to the storage server.

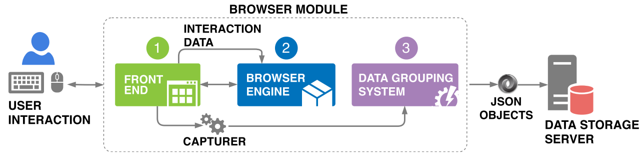

FIGURE 2. Detailed view of browser module.

For intercommunication between the stages and modules, JSON is used, because it is a standard that provides high-level data structuring, which allows at the same time interoperability between different languages and easy coding and decoding, even when performed manually. The HTTP protocol is used because of its easy implementation and use.

B. STORAGE SERVER

The data sent by the browser module are received by the storage server. Through an application written in PHP, the data are decoded, transcribed to PHP objects, separated by single user IDs, encoded in XML, and stored in separate directories by the calibrated Web application domain for subsequent use in the analysis application.

C. ANALYSIS APPLICATION

The analysis application, as indicated in Figure 1, includes three modules that, using the DirectX graphical API, can generate from the data decoded by the XML loader component different representations of data received from the storage server. These are the single-view module, which builds individual views of interactions, the heatmap module, which is responsible for grouping data and generating compiled views of multiple samples, and the datafuzzy module, which articulates an action strategy based on a set of linguistic rules.

The operation of the single-view module is divided into two stages, starting with the reception of the data of each sample, encoded in XML. Stage one is responsible for sorting the data into Scroll, Click, and Wait, preparing for the next component view. Then, these are represented graphically in stage two, by means of coordinate points on the screens captured during the analysis.

The heatmap module is also divided in two stages. After receipt of the XML data of multiple samples for decoding and transcription for C# objects by the XML loader, the data stored in the memory pass to the first stage, responsible for identifying agglomerations of coordinated points and assigning them scores accordingly. After they have been appropriately processed, the objects pass to the second stage, in which, through the coordinates, they are positioned on the captured screens of the calibrated system and defined with colors according to the score assigned in the previous stage. The result of this processing is a heatmap cluster.

The datafuzzy module, as well as the previous modules, depends on the receipt of the data of multiple samples that are decoded by the XML loader. In the first stage, the obtained data are submitted to a process of identification and quantification of user behaviors based on their chronological order. The processed information is sent to the second stage, which organizes it in preparation for export and submission to the subsequent processing performed through a fuzzy computer intelligence system external to AIMT-UXT, which is discussed in the next section.

IV. CASE STUDY

The Website of the Federal Revenue Service of Brazil was used in our case study.[5] Dozens of tax services are provided by the Website, including the income declarations of individuals and legal entities, the main sources of the tax collected by the Brazilian government.

Users were selected based on the random sampling method [32]. This method considers a subgroup of individuals (a sample) chosen from a larger group (a population). Each user was chosen entirely randomly, ensuring that a user had the same probability of being selected at any time during the sampling period. A total of 21 users participated in the experiment.

All the users in the experiments were university exact sciences and humanities students. The age range of the students was 20 to 25 years-old. A brief interview (pre-test) was conducted to identify their previous knowledge about the subject of the test. Therefore, it was considered that the knowledge base of the sample space was not discrepant.

The tests were conducted on three computers with Ubuntu 16 and the Google Chrome browser. The AIMT-UXT browser module, which was responsible for capturing interaction data, was installed. Each execution of the test occurred without any interference from other users or researchers involved in the study, aiming to leave the interface as the only entity to guide the user to the completion of the set tasks. Each task execution was considered finished only at the moment of its completion or when the user declared he/she was withdrawing.

A. VARIABLES

To infer the UX, it was necessary to define variables that can evaluate the performance of the user in executing the tasks. Table 1 shows the variables considered relevant for this case study.

The interaction record can be analyzed using the trace analyzer component of the datafuzzy module. The values for the input variables that allow the user's performance in a task to be measured and the corresponding UX to be evaluated are stored in this component.

[5] http://idg.receita.fazenda.gov.br

TABLE 1. Description of the variables.

Variable	Description
Completion time	Time elapsed from complete rendering of the page to completion of the task by the user.
Actual search distance	Distance in pixels traveled by the mouse cursor from fully rendering the page until the user clicked on an object.
Ideal search distance	Distance in pixels in a straight line from the initial mouse position to the position of the object clicked by the user.
Search ratio	Percentage of actual search distance of the ideal search distance. It is particularly useful for identifying deviations of attention.
Total number of clicks	Number of times the user clicked to complete the task.
Click decision delay	Time between the positioning of the mouse cursor over an item until the click.
Page replays	Number of times the user entered a page already viewed.
Page returns	Number of times the user entered a page, ran a maximum of two more pages ahead, and then returned to the initial page.

TABLE 2. Description of tasks performed by users.

Task	Description
1	Find the Declaration of Income Tax of Individuals on the FRSB Website.
2	Find access to the SIMPLES NACIONAL portal through the FRSB website.
3	Initiate the procedure of Tax Audit - Service.
4	Start the registration process in the e-CAC Portal.

B. TASKS

User tasks were defined that allowed the UX to be evaluated in this case study. The tasks were selected using the criterion of relevance to the population. Therefore, considering the growth trend in performing income tax returns through mobile devices [33], the notable expansion of the microenterprise segment [34], the large volume of requests for anticipation of statement analysis for taxpayers who are retained in the audit, but still weren't summoned [35], and the requirement of registration for use of most FRSB services, we selected four activities, described in Table 2.

Each task was initially performed by one of the authors of this study to obtain reference values for the variables used in the study, presented in Table 3 (s=seconds; px=pixels):

Table 3 presents the values of variables considered ideal for performing the defined tasks, which served as a basis for comparison with the users' registered values for the same tasks.

C. QUESTIONNAIRE

After the tests, a self-assessment questionnaire was administered to understand the user's perception when performing the tasks. As a result of the literature review in section II, the SUS questionnaire was selected as the comparative evaluation method.

TABLE 3. Reference values for the four tasks.

Variable	Task 1	Task 2	Task 3	Task 4
Completion time	5 s	4.5 s	4 s	4.5 s
Actual search distance	1947 px	2205 px	1290 px	1522 px
Ideal search distance	45 px	65 px	753 px	45 px
Search ratio	4326%	3392%	171%	3382%
Total number of clicks	2	2	3	2
Click decision delay	1 s	1 s	1 s	1 s
Page replays	0	0	0	0
Page returns	0	0	0	0

The respondents of the SUS questionnaire indicate their answers on a Likert scale [36] ranging from "Strongly Disagree" to "Strongly Agree." The SUS questionnaire contains 10 statements related to usability aspects, alternating between positive and negative affirmations [37].

The SUS questionnaire was administered after the completion of each of the four tasks. The results of each user were calculated using the method defined in [16], resulting in a score between 0 and 100, where 0 corresponds to a poor usage experience and 100 to a good user experience. The grades resulting from the application of the SUS method were compared with the evaluations obtained from the fuzzy inference of the AIMT-UXT, described in the following subsection.

D. FUZZY LOGIC

Fuzzy sets theory, introduced by Zadeh [38] to handle vague, imprecise, and uncertain problems, has been used as a modeling tool for complex systems that can be controlled by humans but are difficult to define precisely.

Because of these characteristics, fuzzy logic can be conveniently applied for UX evaluation. It initially involves the construction of the fuzzification interface in which the inputs are mapped to the fuzzy sets, represented by the membership functions. In this case study, we used trapezoidal and triangular functions, where the minimum and maximum intervals of the records observed for each variable were previously defined. Figure 3 presents the linguistic variables with the corresponding degrees of input membership functions.

After the fuzzification of each input, the process of inference of the UX evaluation was initiated. To achieve this, a hierarchical model was constructed, taking into account the relationship between the input variables presented in Figure 4.

The hierarchy of variables presented in Figure 4 served to facilitate the formulation of the rules that describe the UX evaluation process. The Mamdani inference method [39] was used for fuzzy reasoning and the center of area defuzzification method was applied.

From a set of 90 rules that were formulated, a fuzzy system was obtained (a fuzzy score), considering the measured values of the eight input variables and one output variable.

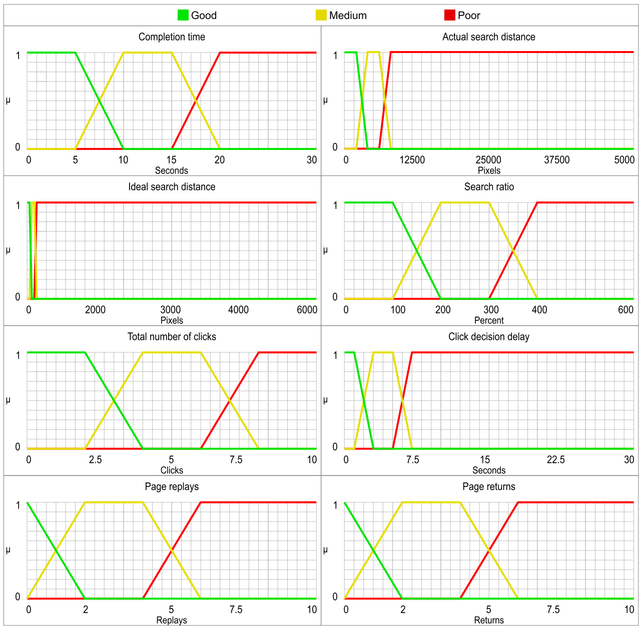

FIGURE 3. Input membership functions.

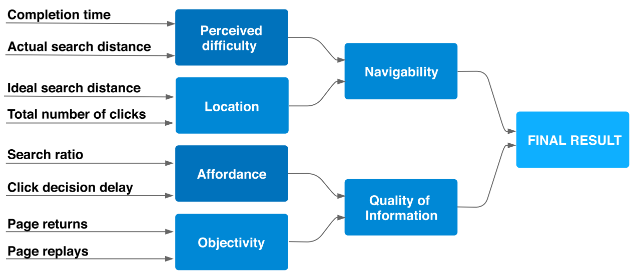

FIGURE 4. Hierarchy of variables.

E. CLUSTERING

Clustering methods can be used to separate records in a dataset into subsets or clusters so that elements of a cluster share common properties that distinguish them from elements in other clusters. Clustering algorithms can help identify natural groups in a dataset, using a certain similarity measure.

In this case study, the values of the fuzzy score were input to a grouping method. Thus, using the TensorFlow tool [40] and a machine learning algorithm for visualization of the clusterization called t-distributed stochastic neighbor embedding (t-SNE) [41], the obtained scores were grouped by similarity.

The t-SNE algorithm is a dimensionality reduction method well-suited for embedding high-dimensional data for visualization in a low dimensional space. It models each high-dimensional object by two or three-dimensional points such that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability [41].

Thus, the interaction logs of all users in the four tasks captured by the AIMT-UXT were loaded into TensorFlow. Then, the t-SNE algorithm performed the reduction of the

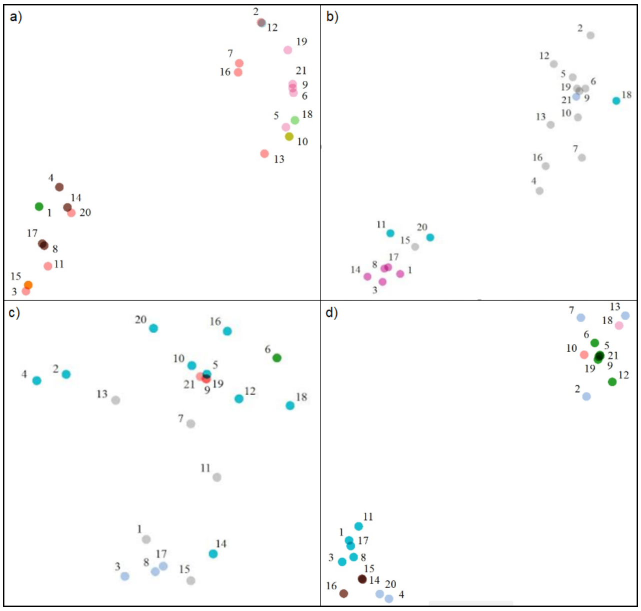

FIGURE 5. Clustering obtained using the t-distributed stochastic neighbor embedding algorithm.

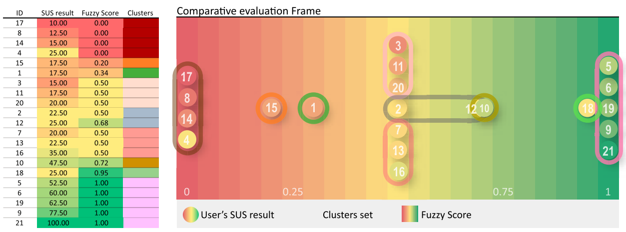

FIGURE 6. Results set for Task 1.

eight input variables to two, distributing the users according to the planes shown in Figure 5. Each quadrant represents the result of the algorithm for each of the four tasks, respectively, in Figures 5a), 5b), 5c), and 5d).

In Figures 5a), 5b), 5c), and 5d), it can be observed that the users, represented by indexes 1 to 21, are associated with a colored circle that indicates the resulting fuzzy score value. It can be observed that, although users present different distributions in each task, it is possible to identify, in general, the formation of two large groups, defined as the users with a poor UX and a satisfactory UX.

V. RESULTS AND DISCUSSION

The interaction data captured by the AIMT-UXT and processed to produce evaluation scores in the fuzzy system, together with the clusters visualized by applying the t-SNE algorithm, were compared to the results of the SUS method, the classic UX technique. Figures 6, 7, 8, and 9, corresponding to Tasks 1, 2, 3, and 4, respectively, allow a comparison of the results of the different methodologies.

Next to each table, the columns of which identify the grades assigned by the SUS and the fuzzy score and the colors of the clusters that belong to each user, we present the graphical representation of these results. These representations

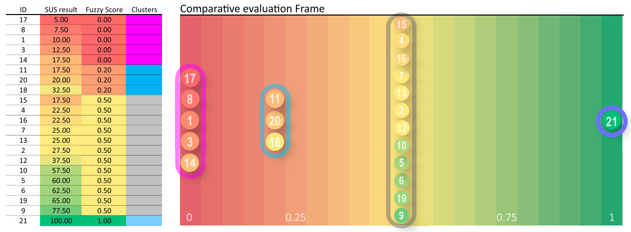

FIGURE 7. Results set for Task 2.

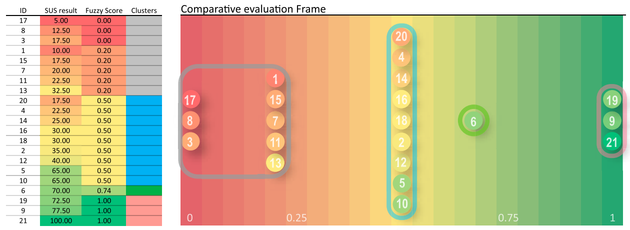

FIGURE 8. Results set for Task 3.

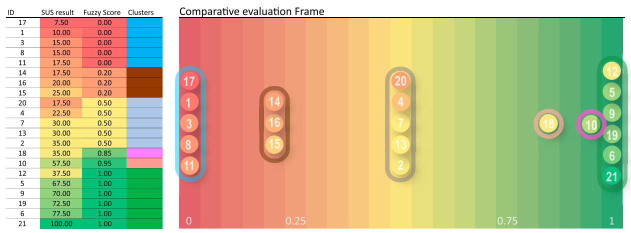

FIGURE 9. Results set for Task 4.

consist of frames with a color scale, composed of a gradient from red to green, representing the values of the fuzzy score, from 0 to 1 in ascending order. Additional important elements of this representation are as follows.

- The numbered circles arranged in each frame identify the users;
- The positions of the circles on the horizontal axis indicate the obtained fuzzy score;
- The color of each circle is defined by the result obtained using the SUS method, by means of a chromatic representation similar to that used for the fuzzy score; however, it varies between 0 and 100;
- The delimitations around the user representations indicate the existence of a cluster; i.e., the users contained in these clusters presented similar characteristics.

Figure 6 presents, at the bottom, the legend of this representation, with the indications of the SUS, fuzzy score, and cluster results.

An analysis of all the figures verified that certain users presented more consistent evaluations; for example, Users 8, 17, and 21 belonged to the same groups in all tasks. In addition, the results of the corresponding fuzzy score and SUS are consistent with the groups to which the users were allocated; i.e., the value assigned by the user in the SUS method is in tune with the value measured by the proposed fuzzy model. It is also noted that, despite the "horizontal dispersion" in the evaluation of some users, such as Users 3, 7, and 10, these users presented equivalent evaluations with SUS and fuzzy methodologies. This indicates that, in fact, the performances, and consequently the UX, were close in different tasks.

It is also noted that, in the case of a minority of users, such as Users 4 and 18, who presented close evaluations in terms of the SUS result and fuzzy score values (except in Task 1), the clustering algorithm failed to identify the correct level of similarity, placing them in different groups.

In general, it was established that the fuzzy-based AIMT-UXT system constitutes a methodology that can produce a subjective assessment of users' UX from their performance records on tasks. The identification of groups of users with similar performances was, for the most part, successfully performed using an additional computational intelligence technique, a clustering algorithm.

In addition, the reliability of AIMT-UXT is evidenced by a comparison of its results with those of the classical method of UX evaluation: in most cases, the scores ascribed by the SUS methodology are consistent with the fuzzy score. It was also noted that users were gathered in well-defined groups, which indicates that those who reported good UX via SUS also had positive fuzzy-based AIMT-UXT results, being grouped in clusters that were in general well defined. This indicates that users who reported a good UX, also had positive evaluation results through the fuzzy system. Similarly, users with a poor UX, for the most part, had a poor evaluation via the fuzzy system. Thus, the proposed tool provides a graphic resource for visualizing results in different methodologies, making it possible to identify the behavior of UX easily and intuitively.

It should be noted that in the reported case study the sample was small-scale, consisting of only 21 users. However, AIMT-UXT is generic and flexible and can be extended to any number of users, providing UX by means of its computational intelligence capabilities, unlike other existing solutions. In addition, it should be noted that the AIMT-UXT tool is free and open source, and its integration is simple, which allows the evaluation of diverse online systems without the restrictions present in similar solutions in the market.

VI. CONCLUSION

In this paper, a quantitative UX evaluation methodology was proposed and a case study of its application in the Brazilian government tax services Website was presented. The results were obtained by monitoring the interactions of users on the Web interface, including mouse movements and navigation parameters, using a tool developed for this purpose, AIMT-UXT. The evaluation was obtained by applying artificial intelligence methods (fuzzy logic and clustering), which are integrated in the tool.

To validate the data capture and analysis capabilities of the AIMT-UXT tool, the evaluations assessed by this tool were compared with those of a traditional and subjective UX method.

The values returned by the fuzzy inference system on the interaction records of a set of users were congruent with the values obtained with the SUS method. In addition, in most cases, the evaluations were consistent with the clusters identified by the t-SNE algorithm. This result shows that the AIMT-UXT is a promising tool for UX measurement from user performance records for tasks.

As contributions of this study, we highlight the following. The development and application of a free, open-source tool for monitoring user interactions in Web systems without the restrictions pertaining to similar market solutions, the use of artificial intelligence techniques integrated in the tool to measure the UX from user performance parameters, the validation of the methodology with tasks performed on the Website of the Federal Revenue Service of Brazil, and an easy and intuitive comparison with a classic UX-based questionnaire method.

In future work, we intend to investigate additional factors that may influence UX and implement additional computational intelligence algorithms, including self-organizing maps, to improve clustering.

Additionally, we intend to develop a new version of the tool, with features that allow the data of users belonging to the poor UX group to be processed to adapt the system interface automatically, to realize an improvement in the UX perceived by this group of users.

REFERENCES

[1] (Mar. 2018). *ISO 9241-11:2018—Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts*. [Online]. Available: https://www.iso.org/standard/63500.html

[2] M. C. Chen, J. R. Anderson, and M. H. Sohn, "What can a mouse cursor tell us more?: Correlation of eye/mouse movements on Web browsing," in *Proc. Extended Abstr. Hum. Factors Comput. Syst. (CHI EA)*, New York, NY, USA, 2001, pp. 281–282. doi: 10.1145/634067.634234.

[3] N. Nakamichi, M. Sakai, J. Hu, K. Shima, M. Nakamura, and K. Matsumoto, "Web-tracer: Evaluating Web usability with browsing history and eye movement," in *Proc. 10th Int. Conf. Hum.-Comput. Interact. (HCI Int.)*, 2003, pp. 813–817.

[4] F. Mueller and A. Lockerd, "Cheese: Tracking mouse movement activity on Websites, a tool for user modeling," in *Proc. Extended Abstr. Hum. Factors Comput. (CHI EA)*, New York, NY, USA, 2001, pp. 279–280. doi: 10.1145/634067.634233.

[5] V. Navalpakkam and E. Churchill, "Mouse tracking: Measuring and predicting users' experience of Web-based content," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI)*, New York, NY, USA, 2012, pp. 2963–2972. doi: 10.1145/2207676.2208705.

[6] M. D. Smucker, X. S. Guo, and A. Toulis, "Mouse movement during relevance judging: Implications for determining user attention," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, New York, NY, USA, 2014, pp. 979–982. doi: 10.1145/2600428.2609489.

[7] L. Čegan and P. Filip, "Advanced Web analytics tool for mouse tracking and real-time data processing," in *Proc. IEEE 14th Int. Sci. Conf. Inform.*, Nov. 2017, pp. 431–435.

[8] H. O. Nyongesa, T. Shicheng, S. Maleki-Dizaji, S. T. Huang, and J. Siddiqi, "Adaptive Web interface design using fuzzy logic," in *Proc. IEEE/WIC Int. Conf. Web Intell. (WI)*, Oct. 2003, pp. 671–674.

[9] C. B. Cakir and B. Oztaysi, "A model proposal for usability scoring of websites," in *Proc. Int. Conf. Comput. Ind. Eng.*, Jul. 2009, pp. 1418–1422.

[10] L. Liang, X. Deng, and Y. Wang, "Usability measurement using a fuzzy simulation approach," in *Proc. Int. Conf. Comput. Modeling Simulation*, Feb. 2009, pp. 88–92.

[11] W. Liu, D. Huang, and Y. Zhang, "Research on fuzzy comprehensive evaluation of user experience," in *Proc. IEEE Youth Conf. Inf. Comput. Telecommun.*, Nov. 2010, pp. 122–125.
[12] Y. Zhou, S. Niu, and S. Wang, "Research on user experience evaluation methods of smartphone based on fuzzy theory," in *Proc. 7th Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, vol. 1, Aug. 2015, pp. 122–125.
[13] M. Hentati, L. B. Ammar, A. Trabelsi, and A. Mahfoudhi, "A fuzzy-logic system for the user interface usability measurement," in *Proc. 17th IEEE/ACIS Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distrib. Comput. (SNPD)*, May/Jun. 2016, pp. 133–138.
[14] E. Folmer and J. Bosch, "Architecting for usability: A survey," *J. Syst. Softw.*, vol. 70, nos. 1–2, pp. 61–78, Feb. 2004. doi: 10.1016/S0164-1212(02)00159-0.
[15] T. S. Tullis and J. N. Stetson, "A comparison of questionnaires for assessing website usability," in *Proc. Usability Prof. Assoc. Conf.*, vol. 1, 2004, pp. 1–12.
[16] J. Brooke, "SUS—A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.
[17] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI)*, New York, NY, USA, 1988, pp. 213–218. doi: 10.1145/57167.57203.
[18] J. R. Lewis, "IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use," *Int. J. Hum.-Comput. Interact.*, vol. 7, no. 1, pp. 57–78, Jan. 1995. doi: 10.1080/10447319509526110.
[19] J. Benedek and T. Miner, "Measuring desirability: New methods for evaluating desirability in a usability lab setting," *Proc. Usability Professionals Assoc.*, vol. 2003, nos. 8–12, p. 57, 2002.
[20] J. Sauro, *A Practical Guide to System Usability Scale: Background, Benchmarks Best Practices*. Scotts Valley, CA, USA: CreateSpace, 2011.
[21] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *J. Usability Stud.*, vol. 4, no. 3, pp. 114–123, 2009.
[22] S. McLellan, A. Muddimer, and S. C. Peres, "The effect of experience on system usability scale ratings," *J. Usability Stud.*, vol. 7, no. 2, pp. 56–67, Feb. 2012. [Online]. Available: http://dl.acm.org/citation.cfm?id=2835476.2835478
[23] T. Kushwaha and O. P. Sangwan, "Prediction of usability level of test cases for GUI based application using fuzzy logic," in *Proc. Next Gener. Inf. Technol. Summit (4th Int. Conf.)*, Sep. 2013, pp. 83–86.
[24] N. Chaudhary and O. P. Sangwan, "Multi criteria based fuzzy model for website evaluation," in *Proc. 2nd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2015, pp. 1798–1802.
[25] T. Singh, S. Malik, and D. Sarkar, "E-commerce website quality assessment based on usability," in *Proc. Int. Conf. Comput. Commun. Automat. (ICCCA)*, Apr. 2016, pp. 101–105.
[26] K. Gulzar, J. Sang, M. Ramzan, and M. Kashif, "Fuzzy approach to prioritize usability requirements conflicts: An experimental evaluation," *IEEE Access*, vol. 5, pp. 13570–13577, 2017.
[27] M. A. Kabir, O. A. M. Salem, and M. U. Rehman, "Discovering knowledge from mobile application users for usability improvement: A fuzzy association rule mining approach," in *Proc. 8th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Nov. 2017, pp. 126–129.
[28] F. Xhafa, S. Caballe, L. Barolli, A. Molina, and R. Miho, "Using bi-clustering algorithm for analyzing online users activity in a virtual campus," in *Proc. Int. Conf. Intell. Netw. Collaborative Syst.*, Nov. 2010, pp. 214–221.
[29] T. D. O. S. Auad, L. F. C. Mendes, V. Ströele, and J. M. N. David, "Improving the user experience on mobile apps through data mining," in *Proc. IEEE 20th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2016, pp. 158–163.
[30] M. Bakaev, V. Khvorostov, S. Heil, and M. Gaedke, "Evaluation of user-subjective Web interface similarity with Kansei engineering-based ANN," in *Proc. IEEE 25th Int. Requirements Eng. Conf. Workshops (REW)*, Sep. 2017, pp. 125–131.
[31] M. A. Bakaev, T. A. Laricheva, S. Heil, and M. Gaedke, "Analysis and prediction of university websites perceptions by different user groups," in *Proc. 14th Int. Sci.-Tech. Conf. Actual Problems Electron. Instrum. Eng. (APEIE)*, Oct. 2018, pp. 381–385.
[32] W. Banuenumah, F. Sekyere, and K. A. Dotche, "Field survey of smart metering implementation using a simple random method: A case study of new Juaben municipality in Ghana," in *Proc. IEEE PES PowerAfrica*, Jun. 2017, pp. 352–357.
[33] BRASIL. (2017). *IRPF 2017 Receita Recebeu 28.524.560 Declarções No Prazo*. Accessed: Jun. 12, 2018. [Online]. Available: https://goo.gl/CUXBNm
[34] BRASIL. (2018). *Quantidade de Optantes Simples Nacional (Incluindo Simei)*. Accessed: Jun. 12, 2018. [Online]. Available: https://goo.gl/6CyGzj
[35] BRASIL. (2017). *Caiu Na Malha Fina? Conheça o E-Defesa*. Accessed: Jun. 12, 2018. [Online]. Available: https://goo.gl/kEu44f
[36] R. Likert, "A technique for the measurement of attitudes," *Arch. Psichol.*, vol. 22, no. 140, pp. 5–55, 1932.
[37] J. Brooke, "SUS: A retrospective," *J. Usability Stud.*, vol. 8, no. 2, pp. 29–40, 2013.
[38] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.
[39] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," *Proc. Inst. Elect. Eng.*, vol. 121, no. 12, pp. 1585–1588, Dec. 1974.
[40] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Design Implement. (OSDI)*, Berkeley, CA, USA, 2016, pp. 265–283. [Online]. Available: http://dl.acm.org/citation.cfm?id=3026877.3026899
[41] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

KENNEDY E. S. SOUZA received the bachelor's degree in information systems from Federal University of Para (UFPA), Castanhal, Para, Brazil, in 2019, where he is currently pursuing the master's degree in anthropic studies in the Amazon.

From 2016 to 2018, he, while an undergraduate, was a scientific initiation fellow, being a member of the Systems Development Laboratory (LADES). He is currently a master's degree Fellow. His research interests include human-computer interaction, usability, user experience, and social technologies.

Mr. Souza is an Associate Member of the Brazilian Computer Society (SBC).

MARCOS C. R. SERUFFO received the Technology's degree in data processing from the University Center of the State of Para (CESUPA), in 2004, and the master's degree in computer science from the Federal University of Para, in 2008, where he also received the Ph.D. degree in electrical engineering with an emphasis in applied computing.

He is currently an Associate Professor (level IV) of the Federal University of Para (UFPA). He is also permanent Professor of Postgraduate Program in Anthropic Studies in the Amazon (PPGEAA), a Collaborator Professor of the Postgraduate Program in Electrical Engineering (PPGEE), and a Collaborator Professor of the Postgraduate Program in Computer Science (PPGCC). He is a Researcher with the Operational Research Laboratory (LPO).

He is the author of more than 80 scientific papers. His research interests include computing science, social technologies, digital TV, computer networks, access technologies, and human-computer interaction.

Dr. Seruffo is an Associate Member of the Brazilian Computer Society (SBC).

HAROLD D. DE MELLO JR. received the bachelor's and master's degrees in electrical engineering from the Federal University of Para (UFPA), in 1999 and 2006, respectively, and the Ph.D. degree in electrical engineering from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), in 2014.

He is currently a Professor with the Department of Electrical Engineering, Faculty of Engineering, Rio de Janeiro State University (UERJ). He is the author of 18 scientific papers. His research interests include artificial intelligence, applied computing, control and automation, telecommunications and information systems, and signal processing.

DANIEL DA S. SOUZA received the bachelor's degree in information systems and the master's degree in electrical engineering with emphasis in applied computing from the Federal University of Pará (UFPA), in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree.

He is currently a member of the Operational Research Laboratory (LPO). His research topics are focused on human-computer interaction, user experience, software engineering, and computer networks.

Mr. Souza is an Associate Member of the Brazilian Computer Society (SBC).

MARLEY M. B. R. VELLASCO received the bachelor's and master's degrees in electrical engineering from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil, and the Ph.D. degree in computer science from the University College London (UCL).

She is currently the Head of the Computational Intelligence and Robotics Laboratory (LIRA), PUC-Rio. She is the author of four books and more than 400 scientific papers in the area of soft computing and machine learning. She has supervised more than 35 Ph.D. Thesis and 85 M.Sc. Dissertations. Her research interests include computational intelligence methods and applications, including neural networks, fuzzy logic, hybrid intelligent systems (neuro-fuzzy, neuro-evolutionary and fuzzy-evolutionary models), robotics and intelligent agents, applied to decision support systems, pattern classification, time-series forecasting, and control, optimization, and data mining.