UXmood - A Tool to Investigate the User Experience (UX) Based on Multimodal Sentiment Analysis and Information Visualization (InfoVis)

Abstract:

Evaluating User Experience (UX) is not a trivial task, and UX specialists have used a variety of tools to ana- lyze data collected from user tests, which causes difficulty in synchronizing the data. This paper presents UXmood, a tool that condenses multiple distinct data types (audio, video, text, and eye-tracking) in a dashboard of coordinated visualizations to ease the analysis process and allow to manage several projects where each project has several logs of user interaction. The tool replays sessions of tests and uses a combination of different sentiment analysis techniques to present a suggestion of user sentiment at any given time during the tasks. The visualizations support brushing and details-on-demand interactions and are synchronized with a temporal slider, allowing analysts to see specific moments of the tests freely. Also, the uses of the sentiment analysis in the collected data may improved the qualitative analysis of UX.

Published in: 2019 23rd International Conference Information Visualisation (IV)

Date of Conference: 02-05 July 2019

Date Added to IEEE *Xplore*: 26 August 2019

ISBN Information:

ISSN Information:

DOI: 10.1109/IV.2019.00038

Publisher: IEEE

Conference Location: Paris, France

SECTION I.

Introduction

According to ISO 9241-110:2010, user experience (UX) is defined as a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service [1]. However, many researchers and practitioners affirm that the UX definition goes beyond that, including a mix of social, physiological, and psychological concepts [2]. For some researchers, the key concept is emotion (or feeling [3]): how a person feels about using a product. The UX definition of this line of thought relates human perception, action, and cognition [4].

The UX evaluation methods consider both quantitative and qualitative metrics, with a higher focus on the latter. According to Law et al. [4], methods to collect qualitative data include questionnaires, interviews, user observations, video-recordings, collages, photographs, body movements (eye-tracking), psycho-physiological clues, protocols (such as think-aloud), and others.

The analysis of subject aspects of UX is not an easy task. Traditional methods such as questionnaires, retrospective self-reports, and observations may collect incomplete information or conceal the real emotional experience of the user during the test of products, systems or services. This is a current challenge in the area of UX evaluation, and one method that has stood out to solve this issue is sentiment analysis.

The process of collecting subjective sentiment data happens through automated techniques that can extract these data from many sources, such as video, audio, and texts [5]. According to Hussain [6] these extraction process can be classified into: Physiological-based Emotion Recognition for understanding the emotional engagement of user while the user interacts with the system (eye tracking); Video-based Emotion Recognition through facial expression and body language analysis for producing insights in human emotional reactions, such as fear, happiness, sadness, surprise, anger, disgust, or neutrality; and Audio-based Emotion Recognition for measuring human emotions by analyzing the human voice collected through a microphone while using the system, such as anger, sadness, and happiness.

In this context, we present UXmood (Fig. 1): a tool to help in the qualitative analysis of UX through a dashboard of coordinated visualizations on sentiment data, automatically generated through multimodal sentiment analysis of user-interaction data collected during evaluation tests.

The tool can be used in real time, while the user interaction occurs, or offline, as a replay of the test section. Currently, the tool collects and synchronizes data from video, audio, interactions, and eye-tracking.

The dashboard visualizations are coordinated through a temporal slider so that the analyst can see volunteer's reactions in specific moments. In any given moment or time interval the analyst can see the user's sentiment, gaze coordinates, duration of eye fixations, accuracy in performing the task, and aloud comments. Besides, the sentiments can be displayed in polarized form (positive, neutral, or negative) or emotions (angry, fear, happy, sad, surprise, disgust, contempt, and neutral), depending on the analysts' goal.

Fig. 1.

Overview of uxmood: Region (a) presents the menu to insert tests and projects; region (b) displays the controller of videos and animation; region (c) introduce emotions and polarity legends; region (d) present a sentiment analysis data in a gantt chart; region (e) displays an interaction video overlaid a scatter plot of eye tracker log and sentiment analysis; region (f) introduces the scan path of the same data; region (g) presents a facial expression video overlaid a sentiment icon; region (h) displays only interaction video; region (i) introduce a word cloud for transcription data; region (j) presents a stacked bar visualization for a duration of sentiment analysis.

Show All

The tool can also be used in the Web, allowing more specialists in user interaction, usability, and experience to contribute to the analysis. It also allows managing files into projects, where each one may contain interaction files from several users so that the analyst can reason about them simultaneously.

This paper is organized as follows: Section II shows studies related to UX in the three main areas of this work: sentiment analysis, InfoVis, and Eye Tracking; Section III describes the tool prototype architecture and explains how the data is input, processed, and visualized through a number of InfoVis techniques and interactions (such as filters and details-on-demand). Section IV concludes the work by presenting a discussion about developed functions and future work directions.

SECTION II.

Related Works

This section presents studies about the use of sentiment analysis, information visualization, and eye-tracking, in the context of UX evaluation.

A. UX Based on Multimodal Sentiment Analysis

Sentiment analysis uses data from several sources (such as facial video, audio, and text) to infer a person's sentiment, and can be used to identify opinions and emotional states of volunteers during user tests [7].

Hussain et al. [6] use multimodal sentiment analysis to identify the emotional state of volunteers during UX evaluation tasks. They conclude that sentiment analysis is a non-invasive method to detect emotional state, and thus avoids interfering on both the volunteer's cognitive state and their performance on tasks.

Setchi and Asikhia [7] elaborated a method to infer emotion from audio and images of UX tests in the context of shopping. They use a camera to take photos from the volunteer's field of view and a microphone to capture audio obtained from applying the Think-Aloud protocol. Their method uses a lexical dictionary (i.e., a “sentiment dictionary”) to process text content, and uses the camera images to identify the referred products.

Voigt et al. [8] applied sentiment analysis to evaluate UX in the context of distance learning. Their method processed text content from the student's social media posts.

B. Infovis to Present UX

Hussain et al. [6], produced a service to generate visualizations for data in the UX context. Their system offers visualization of user interactions in real time or replay mode.

Figueiras [9] presented qualitative metrics to the evaluation of UX in the context of information visualization. Information visualization evaluations focus majorly on quantitative measures, such as task duration and accuracy; but qualitative aspects, such as learning curve and commitment to the task, are equally important [9].

C. UX Based on Eye Tracker

Róbert et al. [10] uses eye-tracker data to evaluate UX within web environments, using metrics such as number of fixations and dwell time. They divided the screen into areas of interest, and measured how the user visually interact with them.

Fu et al. [11] use fixations and saccades to identify user visual behavior in different InfoVis interfaces. They analyzed the collected data with a scanpath visualizations to reason about UX during the interaction.

Fig. 2.

Uxmood application structure and modules.

Show All

D. UXmood Contributions

The UXmood uses multimodal sentiment analysis (audio, video, and text) to infer emotions and polarizations of volunteers during user tests, helping analysts to reason about data collected from UX evaluations.

UXmood supports eye-tracker data. Each gaze measure in the log contains a timestamp so the tool can associate each of them with the sentiment inferred at that time by the multimodal classifier, adding a new layer of information to the scanpath and scatterplot displayed in the dashboard.

The tool works by replaying the user tests while presenting visualizations on both the collected data (audio, video, text, and eye-tracking) and the generated sentiment data (chronological emotion and polarization of the user sentiment), enabling online sharing and collaboration of an enhanced test data in a single tool.

SECTION III.

UXmood Tool Prototype

The tool aims to help UX analysts in testing software by replaying user interactions while showing visualizations on the sentiment analysis of collected data. Sentiment can be defined as polarization (positive, negative, neutral) and emotion (angry, fear, happy, sad, surprise, disgust, contempt and neutral). The tool is implemented in Python on its server-side and runs on browsers, using the javascript library D3 to generate visualizations in HTML.

he tool has three main modules: (1) the input module, that uploads, stores, and synchronizes data, (2) the processing module, that manipulates data, transcripts audio, and generates sentiment data, and (3) the visualization module, that plots charts to display the collected and generated data and enables analysts to interact with UXmood. Fig. 2 shows this structure.

A. Input Module

The input module manages the data collected during the software test. The analyst needs to input at least one of the following data: user's face video, user's audio, video capture of the test screen, a tasks log file, and eye tracker log. The tasks log must have the start and end timestamps for each task in the test, and the eye tracker log must contain the gaze coordinates, a boolean indicating fixations, and the timestamp of each measurement. The tool accepts the most used formats, such as MP4, MOV, and AVI (for videos), WAV and MP3 (for audios). The log files must be in CSV format.

The analyst uploads files through a formulary, as shown in Fig. 3. If the data is not already synchronized (i.e., if the capture did not initiate at the same time), the start timestamp of each capture must be specified to correlate files chronologically. He must specify the display's resolution for the eye tracker data.

Fig. 3.

Input upload formulary.

Show All

After uploading the data, the UXmood synchronizes the video data with the eye tracker log. First, it calculates the test elapsed time using the given timestamps, then it verifies if the duration of the input videos are coherent: If the videos are longer, they are cut to the duration of the test; if they are shorter, an error message is reported, as the data was probably either corrupted or wrongly uploaded. The tool then filters the eye tracking data to include only measures in the video time window and to remove entries that fall outside the display resolution, as they are both irrelevant to the analysis.

If the log file with task start-end timestamps was not input with the rest of the data, it can be inserted later in a formulary. With this information, UXmood can generate visualizations and sentiment data for each task separately.

After the data input and synchronization, the tool passes the data to the processing module to generate sentiment analysis.

B. Processing Module

The processing module uses multimodal sentiment analysis classifiers, which are algorithms to infer the user sentiment during the tests from recorded audio and video. The module consists of four classifiers: video, audio, text, and multimodal. Following is a brief explanation of how such classifiers work in UXmood.

Fig. 4.

Gantt chart - timeline of sentiments. (a) Shows emotions in the default view, and in (b) the polarized sentiment is shown but without the data outside task intervals.

Show All

The video classifier processes a user-face video during the test. It analyses a frame every 250 ms since Yan et al. [12] defines that micro facial-expressions takes a quartersecond to be formed [12], and so a higher frame-rate would be redundant. Each frame is classified in a convolutional neural network, as studied by Arriaga et al. [13], and outputs the sentiment that the user is expressing during that frame. The model was trained using the FER+ face-image dataset [14].

The audio classifier process an audio record of the user during the test. The system slice the audio file into five-second sections, then an algorithm developed by [15] uses signal processing techniques to extract features (e.g., pitch, log energy, and Mel-frequency cepstral coefficients) and uses SVM (Support Vector Machine) to classify the user-sentiment at each moment.

The text classifier transcripts each audio slice into text using google's speech-recognition algorithm. Then, the Vinay et al. [16] algorithm classifies each slice using an SVM to infer user emotion, but not polarization. To complement this data, we also use the SentiwordNet [17] lexical dictionary to classify sentiment polarization.

The multimodal classifier unifies the output of previous classifiers into one final sentiment. Mehrabian [18] proposed the 7-38-55 rule that suggests that only 7% of communication is on spoken words (text), 38% is on voice timbre, tone, and volume (audio), and 55% is on corporal expressions (video). The multimodal classifier internally uses the 7-38-55 rule to unify the three classified sentiments into a single sentiment that better describe the user in that frame.

C. Visualization Module

This module is responsible for generating visualizations for both the input data and the generated sentiment data. The next section explains the chosen InfoVis techniques and the data that they represent. Then we show the implemented user-interactions for the tool.

1) Infovis Techniques

UXmood presents the collected and generated data in a visualization dashboard with the following techniques. All visualizations use color to encode the inferred sentiment.

The Gantt chart works as a sentiment timeline, chronologically displaying the output of classifiers. Fig. 4(a) (also in Fig.1-D]) shows the Gantt chart configured as an emotion timeline. Each horizontal section encodes color to the sentiment inferred by the video, audio, text, and multimodal classifiers. The annotation above the columns indicates the start and end timestamps for each task.

A stacked bar chart shows the total proportion of inferred sentiment on the collected data. Fig. 5 (also in Fig.1-J]) shows this visualization displaying emotion proportions.

This chart aims to give an overview of the general sentiment of the user during the whole test.

Fig. 5.

The stacked bar chart of classified emotions.

Show All

A wordcloud shows the main words said by the participant during the test, with higher size words being the most spoken. As words can be said in different contexts, each can have multiple sentiments associated. The wordcloud uses the multimodal classifier's most inferred sentiment, of all occurrences of a given word to encode its color. Fig. 6 (also in Fig.1-I]) shows an example of this visualization (in the user-spoken language). The user said the word “Região” in eight sentences, where the majority of them occurred during a “disgust” classification.

A scanpath shows eye-tracked user fixations on the display. Fig. 7 (also in Fig.1-F]) shows this visualization. The size of each circle represents the duration of fixations, and the color indicates the multimodal classifier's most inferred sentiment during fixation.

Fig. 6.

Wordcloud of user-spoken words.

Show All

Fig. 7.

Scanpath showing user's fixation with the associated inferred sentiment.

Show All

A scatterplot shows eye-tracked points as a chronological animation. Fig. 8 (also in Fig.1-E and H]) shows this visualization. Each dot represents an entry in the eye tracker log, and the color indicates the multimodal classifier's inferred sentiment. While the test is replaying, only the last two seconds of eye-tracked data are plotted on the scatterplot, to trace recent movements. However, after the replay is done, the full data is drawn to give the analyst an overview of the user's sentiment in each area of the display. The scatterplot is drawn upon the current frames of the screen video, which are updated as the replay continues.

Fig. 8.

All of the points the volunteer looks at with their emotions (left) and the same information for a particular moment in the video.

Show All

A sentiment icon displayed at the top left corner of the user-face video indicates the video classifier's inferred emotion at each frame of the video, updating as the replay continues. Fig. 9 (also in Fig.1-G]) illustrates this visualization where the inferred user emotion is happy.

2) Interactions

Fig. 1, at the start of the paper, presents an overview of the UXmood dashboard with all its visualizations. Fig. 10 (also in Fig.1-B]) shows one of the tool components: a replay controller for the analyst to play, pause, fast forward and rewind the replay of the user test.

Fig. 9.

User-face video with sentiment icon.

Show All

Fig. 10.

The controller element to operate the replay of the user test.

Show All

The tool also supports a set of interactions to filter or transform the displayed data. Thus, the analyst can switch between sentiment polarity or emotion, filter out data that lies outside tasks' time window, define the classifier that is used to infer sentiment (video, audio, text, or multimodal), filter out fixations whose duration is below a given threshold (in ms), and to filter out a set polarities or emotions.

The analyst can set these configurations at the window shown in Fig. 11, clicking on the settings tab. Switching between polarity and emotion can help the analyst to evaluate the visualizations with the desired level of detail about the inferred sentiments.

Fig. 11.

The visualization settings window.

Show All

Filtering out data that lies outside tasks' time range help eliminating unwanted data, such as the preliminary steps of the tasks (Fig. 4(b) shows the effect of filtering unwanted data). Defining the classifier that infers emotions helps the analyst to evaluate different collected sources separately. Filtering out fixations below a given threshold may generate insights about visual elements that most caught the user's attention.

Filtering out a set of polarities or emotions helps the analyst to understand the elements of the tested software that caused more specific reactions. The analyst can apply those filters through the checkbox displayed on the sentiment subtitles, as shown in Fig. 12 (also in Fig.1-C]).

Besides these configurations, it is also possible to apply the brushing and get details on demand by hovering the mouse over visualization elements. Fig. 13 illustrates this interaction in Gantt chart: hovering a bar in the video chart shows a frame of the user at that moment, and hovering a bar on the text Gantt chart shows the transcription of the audio at that moment. This hovering also works as a brushing interaction, highlighting all elements that occurred in that time frame.

Fig. 12.

Legend and filters, being filter of emotions (left) and polarities (right).

Show All

Fig. 13.

Detais on-demand for the video and text gantt chart

Show All

SECTION IV.

Conclusions and Future Work

This work featured UXmood, a tool that uses visualization panel and sentiment analysis algorithms, to better understand the user's feelings and actions during the tests. The tool automatically generates insights about the user emotional state from videos, audios, and texts, through multimodal classifiers, analyzing UX more precise compared to traditional methods that are usually invasive and more error-prone.

The tool also supports Eye-Tracking data, making possible to know the area the user was looking when he felt in a certain way and how much time he looked at it, providing more information about the tested software elements.

UXmood has the possibility of synchronizing data from various sources, visualizing a diversity of data and at different times, highlighting the eye gaze data with the video and emotions

Many future works can be explored, such as the development of new visualization techniques for sentiment data, new UX evaluation methodologies, and the inclusion of new sensors to enhance multimodal sentiment analysis precision such as electroencephalography (EEG), electrocardiogram (ECG), and electromyography (EMG). Automatic identification of phrases in speech intervals will be the next step in continuing this work.

Also as future work, an evaluation with specialists is necessary, having a defined scenario, aiming to identify problems in the interface, interaction, and ambiguity of users' feelings. In the same way, the possibility of visualizing and analyzing UX makes it possible to identify metrics to evaluate UX in a different context of the user.