Predicting User Preference in Pairwise Comparisons Based on Emotions and Gaze

S. Angelastro(✉), B. Nadja De Carolis(✉), and S. Ferilli(✉)

Dipartimento di Informatica, Università di Bari, Bari, Italy
{sergio.angelastro,berardina.decarolis,stefano.ferilli}@uniba.it

Abstract. Emotions have an impact to almost all decisions. They affect our choices and are activated as feedback during the decision process. This work aims at investigating whether behavior patterns can be learned and used to predict the user's choice. Specifically, we focused on pairwise image comparisons in a preference elicitation experiment, and exploited a Process Mining approach to learn preferences. We proposed and evaluated a strategy based on experienced emotions and gaze behaviour, whose results show promising prediction performance.

Keywords: User modeling · Emotion analysis · Gaze behavior · Process mining

1 Introduction

In human decision-making both rational thinking and emotions have an important role. Although decisions are supposed to be a consequence of an intentional and rational behavior, emotions strongly impact choices as they may affect the logical reasoning. Loewenstein and Lerner [11] proposed a decision-making model in which emotions are divided into two types: those anticipating future emotions and those immediately experienced while deliberating and deciding. Therefore, when making a choice, emotions experienced during the decision process can be used as feedback about one's preferences. In the field of Recommender Systems (RSs), preference learning is a necessary step to produce good recommendations. Usually this process is based on explicit feedback, however, recent trends use approaches based on implicit feedback inferred by analyzing user's behavior during the interaction [18]. Nowadays, it has been recognized that successful recommendations need to take into account user perceptions of what is being recommended that in some way elicits interest, perplexity, curiosity, or, more generally, an emotional response in the user. Early studies have shown the potential for improving RSs by analyzing clicking behavior [1] or incorporating eye tracking data [19]. Moreover, behavioral data enables the analysis on user high-level decision-making processes [8]. Also emotions experienced during the interaction with a RSs reflect liking/disliking of content and, therefore, can be used as an

F. Wotawa et al. (Eds.): IEA/AIE 2019, LNAI 11606, pp. 253–261, 2019.
https://doi.org/10.1007/978-3-030-22999-3_23

implicit relevance feedback [15]. Following this approach, our hypothesis is that by detecting behavior patterns typical of the decision-making process it is possible to learn a model to predict future choices.

To address this issue, we investigated how both gaze behavior and emotions, detected from facial expressions, could be used to this aim. Therefore, we performed experiments in which pairwise items' image comparisons was used as a preference elicitation method. Then, to learn the model, a process mining approach has been used and, specifically, the WoMan framework. It includes prediction features that have proven to be effective in several domains, and that may support the aims of this paper. Experiments show that the prediction task, based on both emotions and gaze, yields good results. Specifically, in the 81% of the cases the system correctly classifies the final user's choice before half of the process span, and in just 15% it is unable to provide a suggestion since, in the current phase of the research, the user indecision modeling is missing.

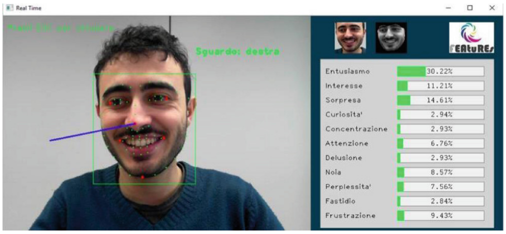
This paper is structured as follows. The next two sections present two frameworks for cognitive emotions recognition and gazes detection, and one for learning and predict preferences. Then, Sect. 4 reports the details of the proposed approach and comments on the experimental outcomes. Finally, in the last section, some conclusions and future work issues are discussed.

2 Emotions Recognition and Gaze Detection

Emotions are part of our everyday living and influence decision-making. The work described in [14] emphasizes that one of the four roles played by emotions in decision-making is to provide information through the display of both positive and negative emotions that arise directly from the options being considered by the decision maker [16]. In particular, according to [3], we are interested in secondary emotions, since decision-making is a cognitive process.

In the context of RSs, the interaction is mainly performed through a display in front of which the user looks at items to make a decision about what to select. Then it is feasible to recognize emotions from facial expressions. To this aim, we developed a Facial Expressions Recognition (FER) system specifically trained on a suitable dataset. The system is called **FEAtuREs** (Facial Expression Analysis for Recognition of Emotions), able to analyze facial expressions both from recorded video and in real time [5]. It follows a commonly used pipeline in FER systems [4]. It performs facial expression recognition on a single image considering as a region of interest the whole face. The set of descriptors used in FEAtuREs is based on the Histogram of Oriented Gradients (HOG) [2]. The classification of the facial expression is done by a multi-class Support Vector Machine (SVM) adopting the "one-against-one" strategy. The final prediction is returned by a voting system among all the classifiers.

The system has been trained and validated on a dataset that integrates three different ones: (i) "EU-Emotion Stimulus Set (EESS)" [12]; (ii) "The Cambridge Mindreading (CAM) Face-Voice Battery" [9]; (iii) "The Cambridge Mindreading Face-Voice Battery for Children (CAM-C)" [10]. In particular, we selected eleven



Preference: a		
timestamp	**gaze**	**emotion**
201811062104048	left	boredom
201811062104054	left	interest
201811062104059	left	interest
201811062104065	left	interest
20181106210407	right	boredom
201811062104076	right	boredom
201811062104082	right	boredom

Fig. 1. (a) FEAtuREs interface (left) - (b) User's behavior example: a is chosen (right)

cognitive emotions that were mentioned in literature as relevant to the decision-making process. The output of this selection is a set of 4184 images whose distribution is the following: enthusiasm (498), interest (340), surprise (295), curiosity (453), concentration (495), attention (374), disappointment (370), boredom (270), perplexity (369), discomfort (461), frustration (259). The average accuracy is of 92% calculated using k-fold evaluation with k = 10. The interface for cognitive emotions recognition is illustrated in Fig. 1(a). Besides emotion recognition, FEAtuREs performs also gaze detection and tracking to detect, in the pairwise comparison, the side of the screen (left or right) the user is looking at. It has been implemented using Dlib and OpenCV functions[1]. Then, a specific instance of a decision-making process is bounded between a starting point, when the system shows for the first time the item's pairs, and an ending point, when a choice is made. While the user is engaged in this process, the flow of emotions and gazes is gathered and collected by FEAtuREs. The user preference choice process is labeled as a or b depending on the selected item. In Fig. 1(b) is shown an example of the user behavior, gazes and emotions with the timestamp in which they occur, collected before the item a is chosen.

3 Process Mining: The WoMan Framework

While Process Mining and Management (PMM) techniques [17] have been typically motivated by, and exploited in, business and industrial domains, they have been recently used for other application fields, as well. In this paper we aim at checking whether they can be effective for user preference prediction, where the user behavior is seen as a process of emotions and gazes. So, a quick recall of PMM basics may be helpful here. A *process* consists of actions performed by agents, formally specified by a *workflow*, which defines their allowed compositions[2]. A process execution is described in terms of *events* associated to the performed activities. A *case* is a particular execution of activities compliant to a given workflow. Case *traces* consist of lists of events associated to time points. A *task* is a generic piece of work, and an *activity* is its actual execution.

[1] Head pose estimation using OpenCV and Dlib: https://www.learnopencv.com/.
[2] Sequential, parallel, conditional, or iterative composition.

Specifically, we adopted WoMan (Workflow Management), an incremental, declarative [13], and logic-based PMM framework [6]. In addition to more typical PMM tasks, such as process mining and supervision, it proved able to support the prediction task[3] in many application domains (see [7] for more technical details). WoMan takes as input trace elements consisting of 6-tuples $\langle T, E, W, P, A, O\rangle$, where T is the event timestamp, E is the type of the event (one of 'begin' or 'end' of 'activity' or 'process'), W is the name of the reference workflow, P is the case identifier, A is the activity name, and O is the progressive occurrence number of A. WoMan models describe the structure of workflows using two elements: **task** (a kind of activities that is allowed in the process) and **transition** (the allowed connections between set of tasks). WoMan consists of several modules. The learning module, **WIND** (Workflow INDucer), learns or refines a process model according to a case. The supervision module, **WEST** (Workflow Enactment Supervisor and Trainer), takes the case events as long as they are available, and returns information about their compliance with the currently available model for the process they refer to. While in supervision mode, WoMan can make several kinds of predictions. Specifically, when the enacted process is unknown, **WOGUE** (Workflow Guesser) returns a ranking (by confidence) of a set of candidate process models.

4 Process Mining for Preference Learning and Prediction

Approach. We used WoMan to face two problems: *preferences learning*, by modeling user's behavioral aspects on a pairwise comparison, and *preference prediction*, guessing future user's choices during the decision-making process.

User's behavior can be defined as a flow of emotions and gazes, framed in the context of a decision-making process. This flow can be seen as a process consisting of activities, i.e., emotions and/or gazes, performed (implicitly) by agents, i.e., users[4], and a process concerning the preference's choice.

We devised and tested several strategies to deal with these kinds of information. Due to space constraints, we report and evaluate only one, namely a strategy based on a combination of gazes, proved to be a powerful source of information in RSs area [19], and emotions. For convenience, we will refer with $D_g = \{left, right\}$ to the domain of gazes, and with $D_e = \{annoyance, attention, boredom, concentration, curiosity, disappointment, enthusiasm, frustration, interest, perplexity, surprise\}$ to the domain of emotions. The idea is to consider gazes and emotions together, in order to give them equal importance, by transforming D_a and D_g into a unique domain $D_{a,g}$ defined as their Cartesian product, and containing all the possible syntactic concatenations. The number of possible tasks will be higher than the ones which consider domains individually, carrying more informative power at the expense of the process complexity.

[3] Given a workflow and an intermediate status of a process execution, the goal is predicting how the execution might proceed, or what kind of process is being enacted, among a set of candidates.

[4] Users will be ignored since the interest is not on user's profiling.

Example 1. Consider the user's behavior in Fig. 1(b). The event trace of the case, named c_1, of the process a (user preference) obtained by applying our strategy is:

$$\begin{array}{l}
\langle 201811062104048, begin_of_process, a, c_1, start, 1\rangle \\
\langle 201811062104048, begin_of_activity, a, c_1, boredom_left, 1\rangle \\
\langle 201811062104054, end_of_activity, a, c_1, boredom_left, 1\rangle \\
\langle 201811062104054, begin_of_activity, a, c_1, interest_left, 1\rangle \\
\langle 201811062104059, end_of_activity, a, c_1, interest_left, 1\rangle \\
\qquad\qquad\qquad\qquad \ldots \\
\langle 201811062104082, begin_of_activity, a, c_1, boredom_right, 3\rangle \\
\langle 201811062104082, end_of_activity, a, c_1, boredom_right, 3\rangle \\
\langle 201811062104082, end_of_process, a, c_1, stop, 1\rangle
\end{array}$$

Note that the name of the activity carries both gaze and emotion information. Whenever a new activity X is detected a *begin_of_activity* event is stored, just after being terminated the previous activity, if any, by recording an *end_of_activity* event. Several occurrences of the same activity are progressive numbered.

In order to learn preferences from user's behavior, the module **WIND** is applied, discovering one model for each preference, containing tasks and transitions. Since there is no concurrency relation between activities, a transition happen whenever a new activity in the event trace is detected, and consists in stepping from the current to the next one.

The experiments concern the evaluation of the proposed strategy by assessing whether performance on the task of preference prediction in the decision-making process at least overcomes a total random classifier.

Datasets and Models Description. Datasets have been created by collecting the user's behaviors, while they made a choice on each image pairs sourced from:

IAPS[5] (International Affective Picture System) domain, an images database designed to provide a standard for studies in emotions domain; 34 users (15 females and 19 males) for 18 pairs of images, paired by opposite valence (positive or negative), and randomly placed to the left or to the right. They produced 280 cases of choice, 145 for a and 135 for item b;

ComPro (COMmercial PROducts) domain, images selected in the domain of commercial products taken from several areas, e.g., clothing, food and art; 42 users (18 females and 24 males) for 28 pairs of images, paired by category and not fully conflicting. They produced 578 cases, where 275 concern the item a and 303 the item b.

We applied the proposed strategy to transform user's cases into event logs, and discover the process models they refer to. Table 1 reports some statistics on each experimental event log and model: number of cases (#cases) and events (#events) (also on average per case), tasks (#task) and transitions (#trans). ComPro involves more cases than IAPS, allowing to discover many more different

[5] http://csea.phhp.ufl.edu/Media.html.

transitions, while #task is almost the same since few cases are sufficient to mine the most of them. The overall #events in ComPro is higher than the one in IAPS, as it involves more cases, while the average length is almost the same. The number of discovered tasks and transitions between the 2 processes, for both datasets, is quite comparable.

Table 1. Datasets and models statistics

Domain	Process *a*					Process *b*				
	#cases	#events		#task	#trans	#cases	#events		#task	#trans
		Overall	Avg	Overall	Overall		Overall	Avg	Overall	Overall
IAPS	145	2842	19.6	15	109	135	2872	21.27	17	141
ComPro	275	5920	21.53	22	184	303	6434	21.23	22	185

Table 2. Preference prediction statistics

Dataset	Performance						
	Acc		P	R	F-score	Indecision	Avg sequence
	I	D					
IAPS	**81%**	**95%**	**95%**	**96%**	**95%**	**15%**	19,56
ComPro	62%	80%	79%	81%	80%	22%	21,38

Performance Evaluation. The experimental procedure was as follows. A 10-fold cross validation procedure was run on each event log (translated dataset). Process models was learned from each training set by applying the WoMan's module WIND [6]. Finally they was used as a reference to call WOGUE on each event in the test sets to predict which kind of process was in execution. As long as case events were tested, WOGUE returned a ranking of the set of candidate processes, using the one with highest Mean Reciprocal Rank as case label.

Table 2 reports datasets on the row headings and corresponding average performance on the columns. Column *Acc* (for *Accuracy*) reports the ratio of cases that WOGUE has correctly classified, distinguishing between the one under indecision, *Acc-I*, i.e., even when it is unable to assign a label, and the one when a decision was made, *Acc-D*. The column *Indecision* reports the ratio of cases in which WOGUE didn't assigned a label. Columns *P* (for *Precision*), *R* (for *Recall*), and *F-score* report classical predictive measures (only on labeled cases), and finally, column *Avg Sequence* reports the average length, in terms of number of events, of the tested cases. The winning strategies are reported in bold.

Our strategy, given its powerful representation formalism, outperforms, in all datasets, a total random classifier, being *Acc* greater than 50%. Moreover, predictive measures, summarized by the *F-score*, yielded a 95% for IAPS, and a 80% for ComPro, which means that it can fairly predict preferences.

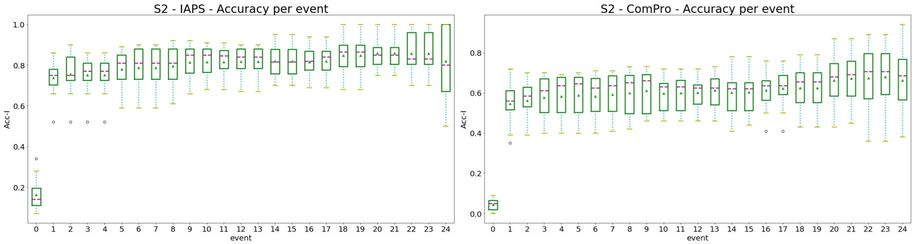


Fig. 2. **IAPS** - (left) and **ComPro** - (right) for predictions.

Specifically, the prediction task in IAPS is 95% accurate in assigning labels, and it is still high (81%) even under indecision (15% of *Indecision*). It is 95% precise (P) in making predictions, correctly covering the 96% (R) of the cases. As regards the ComPro dataset, performance are not as good as on IAPS, nevertheless it yielded about 80% of *Acc-D*, P, and R, when it decided a label, while *Acc-I* is 62% (being *Indecision* value equal to 22%). This may be due to both challenging elicitation skill of ComPro and lack of a user's indecision model.

Graphs in Fig. 2 show the 10-folds *Acc-I* trend, in form of box plots' sequence, per *event*. Since the *Avg Sequence* in Table 2 is between 19–21, and since the interest is on guessing as early as possible the user's choice, only the first 25 events are shown. The *Acc-I* value converges rapidly after few events to the one in Table 2, for both datasets, therefore it is quite reliable at the beginning. Specifically, the accuracy on IAPS is about 80% after 6 events (a third of *Avg Sequence*), and the accuracy on ComPro is about 60%–70% within 10 events (a half of *Avg Sequence*), which are noteworthy results. Low accuracy values are outliers and occur especially at the beginning, while good ones are in boxes[6], in general distributed around the mean. Since boxes in graphs are skewed to the maximum up to a half of the sequence, it is very likely to have a correct prediction in this time span. These results confirmed that both emotions and gazes are important in modeling the user's behavior for preference prediction. Considering trends in Fig. 2 and the average sequence length in Table 2, the system can provide reliable predictions within a half of the sequence. However a user's indecision behavior model would help to increase the overall accuracy.

5 Conclusions and Future Directions

Emotions are fundamental to almost all decisions. This work investigated the use of the WoMan Process Mining framework for preference learning and prediction. We focused on pairwise image comparisons in a preference elicitation experiment, assessing the effectiveness of a strategy which takes into account gaze behaviour and emotions. Results confirm that it significantly outperforms prediction of a total random classifier. In future work, we plan to exploit other

[6] The range between the first and third quartile.

strategies and to perform new experiments with more users. We also plan to ask a marketing expert to suggest how to select suitable items to be shown in the pairwise comparison. Then we will integrate our approach in a game in which the system will guess as early as possible the user's choice.

References

1. Agichtein, E., Brill, E., Dumais, S.: Improving web search ranking by incorporating user behavior information. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 19–26. ACM (2006)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
3. Damasio, A., et al.: Descartes' Error: Emotion, Reason, and the Human Brain. GP Putnam's Sons, New York (1994)
4. Del Coco, M., Carcagnì, P., Palestra, G., Leo, M., Distante, C.: Analysis of HOG suitability for facial traits description in FER problems. In: Murino, V., Puppo, E. (eds.) ICIAP 2015. LNCS, vol. 9280, pp. 460–471. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23234-8_43
5. D'Errico, F., Paciello, M., De Carolis, B., Vattanid, A., Palestra, G., Anzivino, G.: Cognitive emotions in e-learning processes and their potential relationship with students' academic adjustment. Int. J. Emot. Educ. **10**(1), 89–111 (2018)
6. Ferilli, S.: WoMan: logic-based workflow learning and management. IEEE Trans. Syst. Man Cybern.: Syst. **44**, 744–756 (2014)
7. Ferilli, S., Esposito, F., Redavid, D., Angelastro, S.: Predicting process behavior in WoMan. In: Adorni, G., Cagnoni, S., Gori, M., Maratea, M. (eds.) AI*IA 2016. LNCS (LNAI), vol. 10037, pp. 308–320. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49130-1_23
8. Glaholt, M.G., Reingold, E.M.: Eye movement monitoring as a process tracing methodology in decision making research. J. Neurosci. Psychol. Econ. **4**(2), 125 (2011)
9. Golan, O., Baron-Cohen, S., Hill, J.: The Cambridge Mindreading (CAM) face-voice battery: testing complex emotion recognition in adults with and without asperger syndrome. J. Autism Dev. Disord. **36**(2), 169–183 (2006)
10. Golan, O., Sinai-Gavrilov, Y., Baron-Cohen, S.: The Cambridge Mindreading Face-voice Battery for Children (CAM-C): complex emotion recognition in children with and without autism spectrum conditions. Mol. Autism **6**(1), 22 (2015)
11. Loewenstein, G., Lerner, J.S.: The role of affect in decision making. Handb. Affect. Sci. **619**(642), 3 (2003)
12. O'Reilly, H., et al.: The EU-emotion stimulus set: a validation study. Behav. Res. Methods **48**(2), 567–576 (2016)
13. Pesic, M., van der Aalst, W.M.P.: A declarative approach for flexible business processes management. In: Eder, J., Dustdar, S. (eds.) BPM 2006. LNCS, vol. 4103, pp. 169–180. Springer, Heidelberg (2006). https://doi.org/10.1007/11837862_18
14. Pfister, H., Böhm, G.: The multiplicity of emotions: a framework of emotional functions in decision making. Judgm. Decis. Mak. **3**(1), 5 (2008)
15. Tkalčič, M., De Carolis, B., de Gemmis, M., Odić, A., Košir, A. (eds.): Emotions and Personality in Personalized Services: Models, Evaluation and Applications. HIS. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-31413-6

16. Tkalčič, M., Maleki, N., Pesek, M., Elahi, M., Ricci, F., Marolt, M.: A research tool for user preferences elicitation with facial expressions. In: Proceedings of the Eleventh ACM Conference on Recommender Systems, pp. 353–354. ACM (2017)
17. van der Aalst, W., et al.: Process mining manifesto. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM 2011. LNBIP, vol. 99, pp. 169–194. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-28108-2_19
18. White, R.W., Jose, J.M., Ruthven, I.: Comparing explicit and implicit feedback techniques for web retrieval: TREC-10 interactive track report. In: Proceedings of the Tenth Text Retrieval Conference (TREC-10), pp. 534–538 (2002)
19. Xu, S., Jiang, H., Lau, F.: Personalized online document, image and video recommendation via commodity eye-tracking. In: Proceedings of the 2008 ACM Conference on Recommender Systems, pp. 83–90. ACM (2008)