Leveraging implicit gaze-based user feedback for interactive machine learning

Omair Bhatti[1 [0000−0001−7983−2384]], Michael Barz[1,2 [0000−0001−6730−2466]], and Daniel Sonntag[1,2 [0000−0002−8857−8709]]

[1] German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany
[2] Oldenburg University, Oldenburg Germany

Abstract. Interactive Machine Learning (IML) systems incorporate humans into the learning process to enable iterative and continuous model improvements. The interactive process can be designed to leverage the expertise of domain experts with no background in machine learning, for instance, through repeated user feedback requests. However, excessive requests can be perceived as annoying and cumbersome and could reduce user trust. Hence, it is mandatory to establish an efficient dialog between a user and a machine learning system. We aim to detect when a domain expert disagrees with the output of a machine learning system by observing its eye movements and facial expressions. In this paper, we describe our approach for modelling user disagreement and discuss how such a model could be used for triggering user feedback requests in the context of interactive machine learning.

Keywords: interactive machine learning· eye tracking· gaze· confusion detection· emotion detection· user disagreement

1 Introduction

Applying machine learning to a new problem or a new domain usually requires a machine learning practitioner to collect a large amount of labelled samples, select representative / discriminating features, and choose an appropriate learning algorithm to model the concepts at hand. In contrast, interactive machine learning enables users, also without a background in machine learning to train a model in a fast-paced, incremental manner [1]. A user can steer the behaviour of the machine learning model by continuously providing feedback, e.g., upon requests from the system. However, repeated feedback queries, such as trivial yes/no questions, can be perceived as frustrating and annoying [6].

This may lead to reduced user trust in model outputs and deteriorate a user's impression of a model's accuracy [12]. Previous research discussed guidelines and rules for developing IML systems and their interfaces to avoid such problems [8, 11, 24]. Dudley and Kristensson [8] propose to reduce the number of interactions by triggering feedback requests for questions of high relevance to the system.

We propose to limit feedback requests to situations in which a user disagrees with the output of a machine learning models by observing the eye movements and facial

expression of that user. In this work, we specify what user disagreement with a machine learning model means, we describe our planned user study and approach for modelling user disagreement based on gaze and facial expressions, and discuss how interactive machine learning systems may benefit from such a model.

2 Background

We hypothesise that user disagreement stems from negative affective states such as frustration, confusion or disappointment. Therefore, we examine the previous literature on how to detect these affective states using implicit user feedback and how they relate to user disagreement. Previous research has shown that human gaze and facial expressions can be used for affect recognition [16, 25] and generally are sources for implicit user feedback [3, 4]. Lallé et al. [15] introduce predictors for the state of *confusion* leveraging gaze from a user. According to D'Mello and Graesser [9], confusion "is hypothesized to occur when there is a mismatch between incoming information and prior knowledge [...], thereby initiating cognitive disequilibrium" (p. 292). Therefore we hypothesize that user confusion can be an indicator for a user's disagreement with the output of a model. Pollak et al. [20] use facial emotion recognition to detect user satisfaction and dissatisfaction where positive emotional feedback corresponds to satisfaction and negative to dissatisfaction. To detect user disagreement, we look for situations in which the user is *confused* or *dissatisfied* by the model output. We plan an experiment to where we push the user to disagree with the model's output while his gaze and facial expression are recorded.

2.1 Confusion Detection

User confusion occurs when a mismatch exists between prior user knowledge and incoming information [9]. Early research on confusion detection origins from the field of educational computing [5, 7], where predictors for confusion leverage facial expression of students, the posture of students or students interface actions and their studying behaviour. Pachman et al. [18] propose the usage of gaze data for confusion prediction in digital learning. In their study, the participants are presented with a puzzle, and while solving it, their gaze is recorded. The authors aim to detect the buildup of confusion during the problem-solving process. On the other hand, we focus on the immediate affective state of confusion resulting from the user processing the information of the model's output. Detecting this type of *immediate* confusion is especially relevant in the field of Human-Computer-Interaction (HCI) since user experience and user satisfaction decreases when the user is in such a confused state [17]. Pentel [19] introduces a predictor for confusion, using mouse movement recorded when playing a simple game. The author show that an SVM model trained on mouse movements can successfully predict user confusion, but it is restricted to the generated game and thus difficult to generalise. [21] create a predictor for confusion based on gaze data on their persona information visualisation tool. The visualisation contains multiple areas of interest (AOIs) with different types of information about a persona. Their model achieves an accuracy of 80%

of confusion predictions using the number of fixation, the length of transition paths between AOIs, and the user's demographic data as features. In a follow-up work, Salminen et al. [22] train a model using gaze-based data only, achieving an accuracy of 70%. The accuracy increases to 99% when the model includes demographic data as features. This indicates that the demographic data correlates with confusion in their recorded dataset. Including demographic data as features leads to a significant improvement of the model to 99% in accuracy. The authors state that most instances of confusion occur for non-experienced, old males which indicates that trust correlates with age and gender (demographic features). However, this suggests that demographic features can be used to model how often confusion appears in different user groups but not for real-time monitoring of confusion.

Lallé et al. [15] created a predictor for confusion during interaction with their interactive data visualisation tool ValueChart. The tool's goal is to assist users to make the best suitable decision (finding rental property) based on their preference. In a study with 136 participants, the authors collect gaze and mouse movement data while a user performs tasks on ValueChart. The user can report confusion by clicking a button on the top right corner of the visualisation tool. Their confusion prediction model achievs a accuracy of 61% using a Random Forest Classifier. A more recent contribution from the same group [23] uses deep learning based on eye movements to predict confusion on the same dataset as [15]. Instead of using features calculated from the eye-tracking data, the authors suggest using the raw sequential gaze data and feeding it to a Recurrent Neural Network (RNN), allowing the RNN to pick up discriminators for classifying confusion that would otherwise be lost when using calculated features. Using deep learning, their model outperforms their previous work (61% vs 82%), and their results suggest that deep learning in combination with raw sequential gaze data is a feasible option for affect recognition. A possible limitation is their self-report button for reporting instances of confusion. It can influence the user's gaze because of its placement in the interface. Therefore it is important to provide a non-distracting way to self-report confusion. A trigger placed in the hand of the participants could be a solution.

2.2 Leveraging Emotion Detection for ML

Using implicit emotional feedback for artificial agents is a recent idea, and only a few publications have explored it. Pollak et al. [20] investigated whether emotional feedback from a user can serve as the reward function for a reinforcement learning agent. The reward function corresponds to the user's level of satisfaction inferred from facial emotion recognition. The emotions are classified as negative ('angry', 'disgust', 'fear', 'sad'), positive ('happy') or neutral ('neutral', 'surprise') [10]. In their experiment, the user controls a drone's movement, which then, based on the emotional feedback, learns whether it took the correct corresponding action. Their initial finding suggests that incorporating emotional feedback into the reward function of a reinforcement learning agent can be used to teach an agent. The author indicate that there is a considerable individual difference between participants' strength of emotional feedback, which makes it harder to differentiate between positive and negative feedback. Therefore, we plan to gather facial expressions and gaze data in our study to have a multimodal solution for user disagreement detection.

Krause and Vossen [14] suggest the use of implicit triggers based on user confusion or uncertainty for explanations in human-agent interaction. They argue that explanations should not only be provided when the user explicitly asks for it but also when the agent detects that the user is uncertain or confused. Their work also lists other possible implicit triggers such as conflicts between a user's beliefs and the agent or misunderstanding its output. These triggers are similar to those we propose to detect since they also describe user disagreement with the model's output, but instead of triggering an explanation, we query the user for feedback

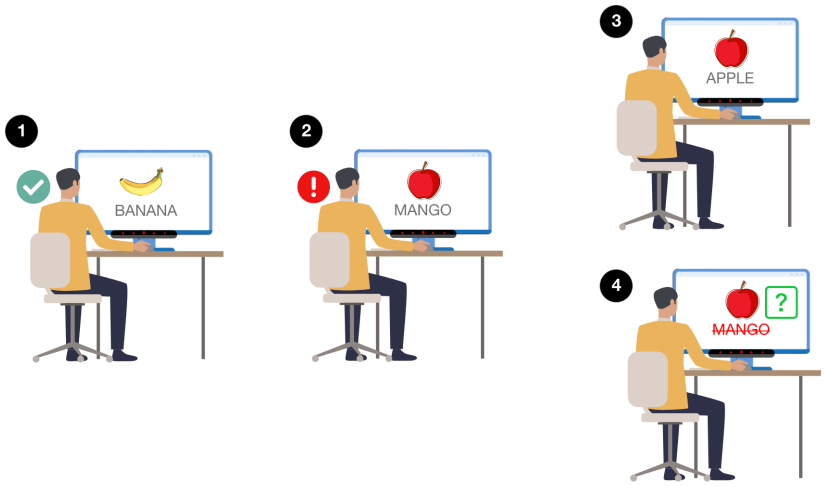

Fig. 1. (1) User interacts with an IML system; (2) our predictor picks up that the user disagrees with the output; (3) the IML system reacts with returning alternative solution or (4) triggers a feedback request

3 Method

In this section we describe our approach, how we plan to collect data for user disagreement, how we plan to create a predictor for user disagreement and how the predictor can be leveraged in IML.

3.1 Data Collection

We plan to conduct an experiment to collect the data necessary to create an effective user disagreement predictor. The experiment will collect the users' gaze data using an eye-tracker and simultaneously record their facial expressions using a video camera. We plan to show a series of images containing an object and its corresponding label,

simulating an output of an object detection model. We will randomly include images containing objects with wrong labels. These instances lead to user disagreement. We want to minimise influences on gaze and facial expression; therefore, the participant uses a trigger placed in his hand to confirm user disagreement. The participant will see one object-label pair at a time for a certain amount of time. If the user presses the trigger in his hand, we stop the image sequence and confirm that he disagrees with the output. A possible extension of the study is to show an image depicting a scene and a caption describing it. To robustly record and synchronize the data we intend to use our multisensor-pipeline (MSP) framework for prototyping multimodal-multisensor interfaces based on real-time sensor input [2].

3.2 Disagreement Detection Model

The features for our planned detection model will be sourced from the eye-tracker and the video camera recording the user. Based on previous research, we list the features we hypothesise to be relevant for user disagreement detection (see Table 1).

Table 1. List of features collected from previous research relevant for user disagreement detection

	Feature	**Source**
Eye-tracker	Number of fixations	[15, 21]
	Fixation durations	[15, 21]
	Length of the transition paths between AOIs (image and label)	[18, 21, 22]
	Image of the scan path visualising the last seconds of eye movement before the user self-reports disagreement	[23]
	Raw sequential gaze data as time-series	[23]
Video camera	Emotion detection based on (FER2013)	[7, 9, 13]
	Body posture/ movement	[5, 7]

3.3 Application in IML

User feedback is essential for interactive machine learning. It helps IML systems to become 'lifelong' learners. Hence the importance to enable users to provide feedback by creating effective interfaces and human-agent interactions. A crucial aspect is **when** to trigger feedback requests since repeatedly asking for feedback can be perceived as frustrating [6] and also reduces trust and impression of model accuracy [12]. Therefore, we try to provide implicit user feedback to the IML system with our proposed user disagreement predictor. The feedback from our predictor consists of a confidence value of detecting user disagreement and the gaze scan path leading to his affective state. The IML system then can react either by showing an alternative solution or triggering a request asking the user for explicit feedback. Figure 1 depicts an example of such a pipeline with a IML system for image captioning. When our predictor detects confusion, the IML system gets notified that the user disagrees with the captioning provided

for the image. Further, it also receives the previous scan path leading to disagreement. The IML system can return an alternative captioning or explicitly ask the user for correction.

3.4 Limitations

We intend to use a remote eye tracking system for gaze estimation in our disagreement detection system. For this, the interaction screen must be instrumented with an additional piece of hardware that requires a user-specific calibration. Also, individual differences of users' eye movements when expressing disagreement need to be considered. They could have a negative impact on the generalizability of our approach.

4 Conclusion

We have shown the motivation and need for detecting when to ask a user for feedback. The following steps will be to conduct the planned study, collect the dataset, and create a user disagreement detection model using the features we collected from previous works. Further, we will use the detection model as a trigger for querying feedback by integrating it into an IML system.

Acknowledgements This work was funded by the *German Federal Ministry of Education and Research* (BMBF) under grant number *01JD1811C* (GeAR).

References

[1] Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T.: Power to the people: The role of humans in interactive machine learning. AI Magazine **35**(4), 105–120 (Dec 2014), https://doi.org/10.1609/aimag.v35i4.2513

[2] Barz, M., Bhatti, O.S., Lüers, B., Prange, A., Sonntag, D.: Multisensor-pipeline: A lightweight, flexible, and extensible framework for building multimodal-multisensor interfaces. In: Companion Publication of the 2021 International Conference on Multimodal Interaction, p. 13–18, ICMI '21 Companion, Association for Computing Machinery, New York, NY, USA (2021), ISBN 9781450384711, https://doi.org/10.1145/3461615.3485432, URL https://doi.org/10.1145/3461615.3485432

[3] Barz, M., Bhatti, O.S., Sonntag, D.: Implicit estimation of paragraph relevance from eye movements. Frontiers Comput. Sci. **3**, 808507 (2021), https://doi.org/10.3389/fcomp.2021.808507, URL https://doi.org/10.3389/fcomp.2021.808507

[4] Barz, M., Stauden, S., Sonntag, D.: Visual search target inference in natural interaction settings with machine learning. In: Bulling, A., Huckauf, A., Jain, E., Radach, R., Weiskopf, D. (eds.) ETRA '20: 2020 Symposium on Eye Tracking Research and Applications, Stuttgart, Germany, June 2-5, 2020, pp. 1:1–1:8, ACM (2020), https://doi.org/10.1145/3379155.3391314, URL https://doi.org/10.1145/3379155.3391314

[5] Bosch, N., Chen, Y., D'Mello, S.: It's written on your face: Detecting affective states from facial expressions while learning computer programming. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) Intelligent Tutoring Systems, pp. 39–44, Springer International Publishing, Cham (2014), ISBN 978-3-319-07221-0

[6] Cakmak, M., Chao, C., Thomaz, A.L.: Designing interactions for robot active learners. IEEE Transactions on Autonomous Mental Development **2**(2), 108–118 (2010), https://doi.org/10.1109/TAMD.2010.2051030

[7] D'Mello, S.K., Craig, S.D., Graesser, A.C.: Multimethod assessment of affective experience and expression during deep learning. Int. J. Learn. Technol. **4**(3/4), 165–187 (oct 2009), ISSN 1477-8386, https://doi.org/10.1504/IJLT.2009.028805, URL https://doi.org/10.1504/IJLT.2009.028805

[8] Dudley, J.J., Kristensson, P.O.: A review of user interface design for interactive machine learning. ACM Trans. Interact. Intell. Syst. **8**(2) (jun 2018), ISSN 2160-6455, https://doi.org/10.1145/3185517, URL https://doi.org/10.1145/3185517

[9] D'Mello, S.K., Graesser, A.C.: Confusion. In: International handbook of emotions in education, pp. 299–320, Routledge (2014)

[10] Ekman, P., Friesen, W.V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., et al.: Universals and cultural differences in the judgments of facial expressions of emotion. Journal of personality and social psychology **53**(4), 712 (1987)

[11] Ghajargar, M., Persson, J., Bardzell, J., Holmberg, L., Tegen, A.: The UX of Interactive Machine Learning. Association for Computing Machinery, New York, NY, USA (2020), ISBN 9781450375795, URL https://doi.org/10.1145/3419249.3421236

[12] Honeycutt, D., Nourani, M., Ragan, E.: Soliciting human-in-the-loop user feedback for interactive machine learning reduces user trust and impressions of model accuracy. Proceedings of the AAAI Conference on Human Computation and Crowdsourcing **8**(1), 63–72 (Oct 2020), URL https://ojs.aaai.org/index.php/HCOMP/article/view/7464

[13] Khaireddin, Y., Chen, Z.: Facial emotion recognition: State of the art performance on fer2013. arXiv preprint arXiv:2105.03588 (2021)

[14] Krause, L., Vossen, P.: When to explain: Identifying explanation triggers in human-agent interaction. In: 2nd Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence, pp. 55–60 (2020)

[15] Lallé, S., Conati, C., Carenini, G.: Predicting confusion in information visualization from eye tracking and interaction data. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, p. 2529–2535, IJCAI'16, AAAI Press (2016), ISBN 9781577357704

[16] Lim, J.Z., Mountstephens, J., Teo, J.: Emotion recognition using eye-tracking: Taxonomy, review and current challenges. Sensors **20**(8) (2020), ISSN 1424-8220, https://doi.org/10.3390/s20082384, URL https://www.mdpi.com/1424-8220/20/8/2384

[17] Nadkarni, S., Gupta, R.: A task-based model of perceived website complexity. MIS Quarterly **31**(3), 501–524 (2007), ISSN 02767783, URL http://www.jstor.org/stable/25148805

[18] Pachman, M., Arguel, A., Lockyer, L., Kennedy, G., Lodge, J.: Eye tracking and early detection of confusion in digital learning environments: Proof of concept. Australasian Journal of Educational Technology **32**(6) (Dec 2016), https://doi.org/10.14742/ajet.3060, URL https://ajet.org.au/index.php/AJET/article/view/3060

[19] Pentel, A.: Patterns of confusion: Using mouse logs to predict user's emotional state. In: Cristea, A.I., Masthoff, J., Said, A., Tintarev, N. (eds.) Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd Conference on User Modeling, Adaptation, and Personalization (UMAP 2015), Dublin, Ireland, June 29 - July 3, 2015, CEUR Workshop Proceedings, vol. 1388, CEUR-WS.org (2015), URL http://ceur-ws.org/Vol-1388/PALE2015-paper5.pdf

[20] Pollak, M., Salfinger, A., Hummel, K.A.: Teaching drones on the fly: Can emotional feedback serve as learning signal for training artificial agents? arXiv preprint arXiv:2202.09634 (2022)

[21] Salminen, J., Jansen, B.J., An, J., Jung, S.G., Nielsen, L., Kwak, H.: Fixation and confusion: Investigating eye-tracking participants' exposure to information in personas. In: Proceedings of the 2018 Conference on Human Information Interaction I&' Retrieval, p. 110–119, CHIIR '18, Association for Computing Machinery, New York, NY, USA (2018), ISBN 9781450349253, https://doi.org/10.1145/3176349.3176391, URL https://doi.org/10.1145/3176349.3176391

[22] Salminen, J., Nagpal, M., Kwak, H., An, J., Jung, S.g., Jansen, B.J.: Confusion prediction from eye-tracking data: Experiments with machine learning. In: Proceedings of the 9th International Conference on Information Systems and Technologies, icist 2019, Association for Computing Machinery, New York, NY, USA (2019), ISBN 9781450362924, https://doi.org/10.1145/3361570.3361577, URL https://doi.org/10.1145/3361570.3361577

[23] Sims, S.D., Conati, C.: A neural architecture for detecting user confusion in eye-tracking data. In: Proceedings of the 2020 International Conference on Multimodal Interaction, p. 15–23, ICMI '20, Association for Computing Machinery, New York, NY, USA (2020), ISBN 9781450375818, https://doi.org/10.1145/3382507.3418828, URL https://doi.org/10.1145/3382507.3418828

[24] Zacharias, J., Barz, M., Sonntag, D.: A survey on deep learning toolkits and libraries for intelligent user interfaces (2018)

[25] Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(1), 39–58 (2009), https://doi.org/10.1109/TPAMI.2008.52