2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW) | 978-1-6654-5490-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ACIIW57231.2022.10085996

Interactive Machine Learning for Multimodal Affective Computing

Rajesh Titung
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, USA
rt7331@rit.edu

***Index Terms*—affective computing, interactive machine learning, machine teaching, active learning**

I. Introduction

Affective computing [1] is an expanding field that is taking on new forms with the development of more powerful computing devices and better modality fusion techniques. However, most datasets – and their resulting models – are limited to basic modalities (speech, text, and video) [2]. Additionally, the lack of labeled emotion data inclusive of comprehensive modalities has been a hindrance to further development in affective computing [3]. Furthermore, emotional expressions can be inherently ambiguous, thus resulting in multiple equally valid representations in expression and perceptions [4]–[6]. Such ambiguity creates challenges for the modalities required to capture the affective expression and the perceived affective states (annotated labels).

My doctoral work aims to explore interactive Machine Learning (iML), which allows iterative interaction between the learner/model and one or more human users, to address issues such as data scarcity and emotional expression-perception ambiguity that are relevant for affective computing. I focus on affective computing to predict and assess human emotions using natural language [7], and natural language $+X$ [8], [9] which includes additional modalities such as eye gaze, speech prosody, facial expressions, and biophysical modalities that can be captured unobtrusively. Because language is a central means of expressing emotions, I consider language as the fundamental modality. And I chose it as the base modality for my research. My current research plan is to explore how to apply iML for affective computing that integrates NLP/speech with other modalities to address resource and subjectivity issues and challenges such as model bias.

So far, I have explored the effectiveness of interactive Machine Learning for text modality, for the affective computing task of text-based emotion prediction [10], using simulation-based meta-learning as experimental methodology [11], building on prior work [12]. The next steps will involve iML with human subject experiments, with the goal of developing a framework that elicits and labels human emotions instantly and uses human input to build more robust affect computing models interactively. I also focus on challenging and complex affect states, which are currently understudied.

The planned research questions for this research are:

1) How can interactive machine learning be effectively used for affective computing that integrates NLP/speech with other modalities?
2) What is the most efficient way of incorporating human users in the loop to enable customization to individual users and a generalization to multiple users and many emotions?
3) How can an iML system incorporate Machine Teaching strategies (an iML approach) that approximate actual teaching-learning experiences?

I will narrow down some of the research questions as I progress with my research.

II. Background and Prior Works

The iML approach can be studied under Active Learning (AL) and interactive Machine Teaching (MT) approaches [13] considering who is responsible for selecting the instances used for training the learner. In Active Learning, the learner iteratively selects highly informative instances to be labeled from a pool or stream of unlabeled instances using various query strategies [14] which are passed to the human user for annotation and later used for training. Furthermore, in Machine Teaching, the human user is in the central role of selecting highly informative instances to be used for training and then uses a subset of the instances selected to train the learner based on various conditions known as teaching criteria [15]–[17]. With the selection of highly informative instances, these iML methods allow for achieving good performance with limited data and many other benefits. Machine Teaching has not been studied for affective computing yet.

Interactive Machine Learning allows iterative interaction between the users and the learning system (learner), thus accelerating the learning process. Iterative user involvement in generating new data instances and selecting highly informative instances using human knowledge helps to enhance modeling resource efficiency. So, interactive Machine Learning can help to build highly efficient affective models in resource-scarce

Acknowledgment: This material is based in part upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

affective computing tasks [18]. In addition, user involvement can help build models robust to ambiguous emotional cues (consider joy and frustration: people tend to smile both when they are happy or frustrated [19]). Continual human involvement-based learning also allows incorporating more emotions and features over time [12].

Poria et al. [2] reported that 90% of studies in affective computing focus on the textual, audio, and visual modalities. Yet, a single or few modalities may be insufficient, or in certain circumstances, even misleading for predicting emotions (especially in case of near emotions like boredom, sleepiness, and sadness) [20], [21].

Using iML-based Active Learning has been explored for unimodal settings to some extent. Zhang et al. [22], Wu and Huang [23], and Abdelwahab and Busso [24] used Active Learning in various settings for speech-based emotion recognition. Similarly, Alm [7] has explored AL with evolutionary computation of emotional voices, while Washington et al. [25] and others have used facial expressions and AL for emotion recognition. Hazer-Rau et al. [26], Kansizoglou et al. [27], and others have reported on the use of AL for affective computing using more than one modality.

Thus prior work shows promise, but there is still much space for exploration in terms of more comprehensive modalities, emotional ambiguity, model biases, customizability, and so on. I plan to work on more modalities, explore various AL query strategies, and build strategies specific to modalities and fusion techniques. There is a lack of work on Machine Teaching for affective computing, with previous work being mostly concentrated on cybersecurity [28]. I especially plan to focus on interactive Machine Teaching [17] for the affective computing paradigm because of its theoretical advantages over Active Learning [15].

III. Scope of the Work

Rouast et al. [3] reported that the lack of labeled datasets is impeding the adoption of deep learning for affective computing. The ambiguity of features representing human emotions from conflicting cues (in near emotions, groups of associated emotions [29], or quite distinct emotions [30]) poses an issue for building a robust affective computing system. Some affective states are over-studied, while some are under-studied in affective computing. Moreover, emotional expressions are connected to cultural contexts and customization can assist with capturing this complexity.

My work aims to develop a framework that allows teacher-learner interaction for a resource-efficient robust affective computing system. I aim to study and introduce more teaching strategies and teachers that can be generalized or customizable for multiple emotional states. I am starting with verified emotion elicitation stimuli for two affective states, frustration and surprise, and plan to use stimuli to elicit targeted emotions and label the collected data on the fly from the users and integrate them into an iML framework in an online learning scenario. I aim to build a flexible framework to include more affective states and more emotion elicitation stimuli.

IV. Methodology and Proposed Solutions

The first task in building the framework is to compile verified emotion elicitation stimuli. For gathering verified stimuli, an ongoing experiment is collecting multimodal data from human subjects during emotion elicitation using multiple stimuli meant for eliciting the targeted emotions. Each task is followed by self-rating (and rating for partners in dyadic tasks) as the label of the data. The study will verify the effectiveness of the elicitation tasks in eliciting the target emotions based on statistical comparisons of the subject's ratings across tasks and emotions. I will also use the collected data to test the effectiveness of interactive Machine Learning strategies like Active Learning and Machine Teaching. Finally, I will also use the gathered data as a seed dataset to initially train the learner in the planned framework.

The verified elicitation tasks will be used later in the planned iML framework with an interactive mobile interface that can be applied flexibly in various contexts. Subjects can interact with the stimuli via the interface, enabling emotion elicitation and data generation, which is concurrently labeled by the subjects. The labeled data will be provided to the learner when teaching criteria are met. The user also gets feedback from the learner, which ensures iterative emotion modeling. Building on the work of Liu et al. [17], I plan to explore the possibility of various types of teachers with various knowledge and scope, mimicking actual human teachers in real or simulated settings. I also plan to explore more teaching criteria customizable to affective computing settings.

V. Preliminary Work and Tentative Plan

A CHI 2022 workshop (*The Future of Emotion in Human-Computer Interaction*) and a NAACL 2022 workshop [11] (*Bridging Human-Computer Interaction and Natural Language Processing*) accepted my early work with my advisor. The CHI workshop contribution (non-archivable) investigated the roles of emotion in human-AI collaboration and ways to involve human interaction to enhance emotion learning. The NAACL workshop paper explored the use cases of Active Learning and interactive Machine Teaching for text-based affective computing tasks. I have also studied various modes and types of emotion elicitation tasks and filtered effective tasks for the data collection experiment described above.

I plan to complete the data collection from subjects and analyze all the collected data and use them for preliminary testing and emotion modeling. After that, I plan to work on designing and creating an interface for the framework, potentially leveraging an interface previously built in my lab for the text modality (expanded to multi-modality). Finally, I plan to work on simulating several types of teachers and having actual human users along with various teaching criteria and how to include all of them in the planned framework effectively.

VI. Contributions to Date and Expected

I have analyzed the applicability and benefits of iML for affective computing in preliminary research. I have demonstrated that Active Learning works efficiently for text modality-based affective computing tasks. Moreover, using transformer-based models and iML together appears to allow for achieving higher performance resource-efficiently. I simulated teachers in interactive Machine Teaching using an Active Learning strategy. I explored different machine teaching criteria for the first time (from the study of prior works) in affective computing. I am working on data collection using stimuli for frustration and surprise generation in a lab setting, and it is ongoing.

One of my expected contributions is to develop a technical framework that is inclusive of comprehensive multimodal data and generalizable to multiple affective states. In addition, my work will also result in new data resources. I also plan to design better teachers and teachers' integration into the framework and design teaching criteria specific to an affective computing problem. Finally, achieving a robust and flexible framework can allow for high scalability. Another by-product will be the supporting software built for the framework that will be able to be used for similar tasks.

Ethical Impact Statement

This work would seek to create a framework that achieves learning in a resource-efficient way. Interactive Machine Learning can reduce the annotator's labeling workload and enable energy-efficient models, resulting in models with a low carbon footprint. The data is being collected in an IRB-approved study with informed consent. However, there are a few possible limitations of the current study. The research is being conducted at a university, which may limit the age and variety of the participants. To ensure gender diversity, we plan to have at least 50% female subjects and aim to include subjects of diverse backgrounds.

Acknowledgment

I would like to thank my advisor Prof. Cecilia O. Alm for her guidance and contribution to all the work described above.

References

[1] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.

[2] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98–125, 2017.

[3] P. V. Rouast, M. T. Adam, and R. Chiong, "Deep learning for human affect recognition: Insights and new developments," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 524–543, 2019.

[4] V. Sethu, E. M. Provost, J. Epps, C. Busso, N. Cummins, and S. Narayanan, "The ambiguous world of emotion representation," *arXiv preprint arXiv:1909.00360*, 2019.

[5] E. Ferracane, G. Durrett, J. J. Li, and K. Erk, "Did they answer? Subjective acts and intents in conversational discourse," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, Jun. 2021, pp. 1626–1644. [Online]. Available: https://aclanthology.org/2021.naacl-main.129

[6] C. O. Alm, "Subjective natural language problems: Motivations, applications, characterizations, and implications," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, 2011, pp. 107–112.

[7] C. Alm, *Affect in Text and Speech*. VDM Verlag, 2009.

[8] N. Kaushik, R. Bailey, A. Ororbia, and C. O. Alm, "Eliciting confusion in online conversational tasks," in *9th Int'l Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2021, pp. 1–5.

[9] J. E. Shea, C. O. Alm, and R. Bailey, "Contemporary multimodal data collection methodology for reliable inference of authentic surprise," in *IEEE Western New York Image and Signal Processing WS (WNYISPW)*, 2018, pp. 1–5.

[10] C. O. Alm and R. Sproat, "Emotional sequencing and development in fairy tales," in *Affective Computing and Intelligent Interaction*, J. Tao, T. Tan, and R. W. Picard, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 668–674.

[11] R. Titung and C. Alm, "Teaching interactively to learn emotions in natural language," in *Proceedings of the Second Workshop on Bridging Human–Computer Interaction and Natural Language Processing*. Seattle, Washington: Association for Computational Linguistics, Jul. 2022, pp. 40–46. [Online]. Available: https://aclanthology.org/2022.hcinlp-1.6

[12] A. Tegen, P. Davidsson, and J. A. Persson, "Towards a taxonomy of interactive continual and multimodal learning for the internet of things," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC '19 Adjunct. New York, NY, USA: Association for Computing Machinery, 2019, p. 524–528. [Online]. Available: https://doi-org.ezproxy.rit.edu/10.1145/3341162.3345603

[13] A. Tegen, P. Davidsson, and J. Persson, "A taxonomy of interactive online machine learning strategies," in *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2020, Ghent, Belgium, September 14-18, 2020, Proceedings, Part II*, ser. Lecture Notes in Computer Science, F. Hutter, K. Kersting, J. Lijffijt, and I. Valera, Eds., vol. 12458. Springer, 2020, pp. 137–153. [Online]. Available: https://doi.org/10.1007/978-3-030-67661-2_9

[14] B. Settles, "Active learning literature survey," 2009.

[15] X. Zhu, "Machine teaching: An inverse problem to machine learning and an approach toward optimal education," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, Mar. 2015. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/9761

[16] X. Zhu, A. K. Singla, S. Zilles, and A. N. Rafferty, "An overview of machine teaching," *arXiv*, vol. abs/1801.05927, 2018.

[17] W. Liu, B. Dai, A. Humayun, C. Tay, C. Yu, L. B. Smith, J. M. Rehg, and L. Song, "Iterative machine teaching," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ser. ICML'17. JMLR.org, 2017, p. 2149–2158.

[18] A. Holzinger, "Interactive machine learning for health informatics: When do we need the human-in-the-loop?" *Brain Informatics*, vol. 3, no. 2, pp. 119–131, 2016.

[19] M. Hoque and R. W. Picard, "Acted vs. natural frustration and delight: Many people smile in natural frustration," in *2011 IEEE International Conference on Automatic Face Gesture Recognition (FG)*, 2011, pp. 354–359.

[20] A. Chin, A. Markey, S. Bhargava, K. S. Kassam, and G. Loewenstein, "Bored in the USA: Experience sampling and boredom in everyday life." *Emotion*, vol. 17, no. 2, p. 359, 2017.

[21] J. Danckert, T. Hammerschmidt, J. Marty-Dugas, and D. Smilek, "Boredom: Under-aroused and restless," *Consciousness and Cognition*, vol. 61, pp. 24–37, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105381001730569X

[22] Z. Zhang, J. Deng, E. Marchi, and B. Schuller, "Active learning by label uncertainty for acoustic emotion recognition," in *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France*, 2013.

[23] D. Wu and J. Huang, "Affect estimation in 3D space using multi-task active learning for regression," *IEEE Transactions on Affective Computing*, 2019.

[24] M. Abdelwahab and C. Busso, "Active learning for speech emotion recognition using deep neural network," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2019, pp. 1–7.

[25] P. Washington, C. Mutlu, A. Kline, C. Hou, K. Dunlap, J. Kent, A. Husic, N. Stockham, B. Chrisman, K. Paskov *et al.*, "An exploration of active learning for affective digital phenotyping," *arXiv preprint arXiv:2204.01915*, 2022.

[26] D. Hazer-Rau, S. Meudt, A. Daucher, J. Spohrs, H. Hoffmann, F. Schwenker, and H. C. Traue, "The uulmMAC database—a multimodal affective corpus for affective computing in human-computer interaction," *Sensors*, vol. 20, no. 8, p. 2308, 2020.
[27] I. Kansizoglou, L. Bampis, and A. Gasteratos, "An active learning paradigm for online audio-visual emotion recognition," *IEEE Transactions on Affective Computing*, 2019.
[28] S. Alfeld, X. Zhu, and P. Barford, "Explicit defense actions against test-set attacks," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, ser. AAAI'17. AAAI Press, 2017, p. 1274–1280.
[29] P. Ekman, "Basic emotions," *Handbook of Cognition and Emotion*, vol. 98, no. 45-60, p. 16, 1999.
[30] M. E. Hoque, D. J. McDuff, and R. W. Picard, "Exploring temporal patterns in classifying frustrated and delighted smiles," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 323–334, 2012.