International Academic Institute for Science and Technology

International Academic Journal of Innovative Research
Vol. 7, No. 2, 2020, pp. 11-15.

ISSN 2454-390X

International Academic Journal of Innovative Research

www.iaiest.com

Designing Emotion-Aware Systems for Human–Computer Interaction in the Web Age

Dr. Sneha Deshmukh [a], Isha Malhotra [b]

[a] *Visvesvaraya National Institute of Technology, India.*
[b] *Professor, Visvesvaraya National Institute of Technology, India.*

Abstract

The advent of new digital technologies and the introduction of online services have made Human-Computer Interaction (HCI) part and parcel of daily life. The traditional models of HCI concentrate on task-based interaction metrics of usability, often neglecting underlying emotional states. This paper tries to fill this gap by proposing a comprehensive approach for designing systems that recognize and respond to users' emotions in the contemporary web era. The plan is to build systems that automatically and dynamically recognize, analyze, and respond to human emotions through facial expression, voice, and sentiment analysis of text. We employed affective computing, deep learning, and real-time emotion recognition algorithms within a single framework which is aimed to improve the overall user experience and participation in the system. The proposed system, demonstrably, achieves more than 88% accuracy in emotion detection along with improved user satisfaction measures in adaptive interfaces. The system outperformed baseline emotion detection systems in relation to task completion and emotional engagement. The results highlight the need to integrate emotional intelligence and sensitivity on interactivity designs to facilitate more human-like and compassion-driven interactions. This work augments the literature on 'Affective Human Computer Interaction', bringing forth new insights about guiding future research, development, and deployment in emotion-sensitive system designs.

Keywords: Affective Computing, Emotion Recognition, Emotion-Aware Systems, Deep Learning, User Experience, Adaptive Interfaces, Human-Computer Interaction, Multimodal Interaction.

Received: 04 September 2020/Revised: 17 October 2020/Accepted: 24 November 2020/Published: 31 December 2020

Introduction

The development of the web has changed the traditional desktop bound model of human-computer interaction (HCI) to a mobile one, utilizing wearables and intelligent agents. As users interact with different digital systems in emotionally rich contexts, the need for interfaces that possess a greater degree of emotional intelligence is apparent. Emotion-aware systems, capable of detecting and responding to a user's emotional state, are vital for boosting user satisfaction, productivity, and engagement. Such developments are indicative of the shift towards more human-centric computing paradigms that are profoundly intertwined with the notions of emotional intelligence.

Usability, efficiency and performance-based task metrics were the key focus areas of traditional HCI models. While still relevant, these attributes do not account for the vastness of human interaction, and considering the emotional elements, online interactions focused on learning, customer service, healthcare and gaming add further to that complexity. Users' memories, attention spans, experience with the systems, decision-making capabilities are all impacted by emotions. Because of this, systems designed to recognize emotional signals and respond to them are bound to enhance the quality of interaction.

Understanding human emotions draws upon a student's psychology, cognitive science, and informatics because it constructs and interprets a multi-faceted domain referred to as affective computing. Emotion detection in facial movements, voice, or text provides a dynamic feel to a student's utterance. Input fusion with machine learning allows systems to perform emotion recognition in real-time. As an example, students suffering from frustration while using an emotion-enabled e-learning platform could receive additional support, which results in learning improvement.

The most recent developments in emotion-aware systems mark a winning point in the history of AI with the addition of deep learning. Classifiers such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and BERT, the one built on the transformer architecture, have performed brilliantly in decoding emotions from images, audio, and texts respectively. These model types provide the responsive time-critical analysis needed for Human-Computer Interaction (HCI) applications.

Unfortunately, not all is perfect with techniques for building systems that respond to human emotions. Shorter gaps in fully automated emotion-responding systems are posed at the level of universal application through varied training set needs, real-time data processing, and application in ethnically diverse areas. Differing cultural standards leave a lot of room for privacy and ethical concerns related to emotional data. This points to the stronger argument that we remain focused on the confounding of precision-controlled engineering efforts, user-defined guidelines, and ethical constraints.

This research attempts to solve these problems by creating a modular and scalable emotion-aware system architecture tailored for web applications. The interface provided in the system with regard to HCI is extensive, as it combines emotion detection using multiple modalities with adaptive user interfaces, enhancing system responsiveness and emotional attunement to users. The methodology presented in this paper, alongside the detailed system design, includes performance evaluation against benchmark datasets and comparison of efficacy with traditional interaction models. The conclusions drawn from this research study are offered in support of constructing emotion-conscious systems that are technically effective but ethically considerate.

Literature Review

The study of emotion-sensitive systems in the context of human-computer interactions (HCI) is an offshoot of affective computing, popularized by Rosalind Picard in her landmark publication (Picard, 1997). Afterward, numerous researchers studied different approaches of integrating emotion recognition into user interfaces. One of the pioneering works in this regard was facial expression recognition in the Facial Action Coding System (FACS) approach, where (Ekman & Friesen, 2003) categorized emotions into distinct facial muscle movements. With the advancement in machine learning, these methods have evolved to include automated facial expression recognition using convolutional neural networks (CNNs), greatly enhancing accuracy and scalability.

The significance of recognizing emotions through one's voice has grown in popularity because speech often contains important emotional indicators. Schuller et al., (2010) showed that emotions can be recognized using prosodic features such as pitch, tone, and rhythm. More advanced techniques include the application of deep learning models, like long short-term memory (LSTM) networks, that analyze temporal sequences within speech data for classification. Accuracy has improved remarkably with these models.

Analysis of emotions through social media and messaging platforms has gained prominence in recent years, driving the expansion of text-based sentiment analysis. Earlier sentiment analysis models utilized bag-of-words and lexicon-

based methods, but such techniques have been replaced by more sophisticated transformer-based models, such as BERT and Roberta, which understand language in context (Devlin et al., 2019). These models have been integrated into real-time chatbots, recommender systems, and even mental health applications that evaluate and respond to user input in real time.

Facial recognition, voice, and textual analysis are some of the sources that could be intersected in a single input for multimodal emotion recognition systems. These multimodal systems outperform unimodal systems because of the data received from multiple channels. Frameworks proposed by (Zeng et al., 2009; Poria et al., 2017) utilized feature-level and decision-level fusion techniques to combine the modalities. The main challenge posed by multimodal systems is the synchronization of different data streams along with any missing subset of data or noisy data. Attention mechanisms with hierarchical fusion strategies alongside recent studies have provided adequate solutions to the posed challenges.

Aside from education, emotion-aware systems have notable applications in customer service, healthcare, and entertainment. For instance, emotion-aware tutoring systems presented by (Woolf et al., 2009) demonstrated a remarkable shift in learner participation and improved educational outcomes for students. Furthermore, emotion recognition is now embedded in virtual therapists to analyze a patient's mood along with automated systems for essential social interaction to let users know when they are being spoken to. Adapting chatbots for customer servicing has resulted in increased customer satisfaction.

Even after significant milestones, there are a few challenges that still persist. Greater inclusion within training datasets is required to better analyze the precision of culture and languages diversities. However, instantaneous feedback cannot be provided in real-time interactions. The ethical boundaries of emotion tracing such as; tracking consent, security of the data captured, and possible abuse of the data are matters which are now a key focus of discussion.

This literature review highlights the advancements and gaps in the field of emotion-aware HCI. While combining various types of data and implementing deep learning technologies, as well as deploying them in real-world settings, has moved the discipline forward, problems such as scalability, generalizability, and ethics continue to be focal points of concern. Addressing the gaps in prior research, this study aims to develop an emotion-responsive system for human-computer interaction on the web that is scalable, precise, and ethically sound.

Methodology

The proposed emotion-aware system architecture operates in real time in web-based environments and integrates emotion recognition with adaptive interface generation. The system begins with user input collection across three modalities: visual, audio, and text. Visual data is acquired through webcam feeds, and processed by a CNN whose parameters were pre-trained on the FER-2013 dataset. The CNN processes emotions like happiness, sadness, anger, and surprise. Audio data is provided by a microphone and is analyzed using RNNs with LSTM units employing the RAVDESS dataset. The dataset enables classification of emotions for the audio data based on prosodic features like pitch, tempo, and spectral entropy. Text input is automatically generated by system users and consists of chat messages or search queries which are analyzed by fine-tuned BERT models. The model captures sentiment and emotional context fiercely.

Upon completing emotion recognition, the data from the three modalities is synchronized and fused using a hierarchical level decision fusion strategy. This approach integrates data from multiple sources in a pre-defined order according to their level of confidence and temporal coherence, which ensures that classification of emotions remains meaningful and accurate. Other attributes of system architecture include the adaptive interface controller responsible for the user interface tailoring in accordance with user's emotion. For instance, if the user is detected to experience frustration, the interface offered could be made less complex, more help could be offered, and the color schemes could be soothing. In contrast, the user detected to be happy or excited could have the system add gamified elements for tuned up engagement.

The underlying architecture consists of modules and open-source libraries built on Python. Facial recognition is done using OpenCV alongside TensorFlow for model inference acceleration. LibROSA is used for feature extraction while PyTorch is used for training and deploying the model to perform audio emotion recognition. Sentiment analysis of text data is also performed with the aid of Hugging Face Transformers library. Fusion and adaptive interface logic are implemented in a Flask server environment coded in the backend which communicates with the front end running HTML5, CSS3, and JavaScript language for real-time updates to the user interface.

For user privacy along with ethical use practices, the system employs opt-in consent procedures and anonymizes all data prior to processing. No identification details are retrieved, and all emotion recognition modules function either on the device or in secured sessions to mitigate data leaks. System performance is assessed using a blend of benchmark datasets and real-time interaction logs provided by users. The evaluation indexes capture accuracy in emotion detection, responsiveness of the interface, user satisfaction, and latency of the system. The emotion-aware system is placed under scrutiny against an emotion-closed interface for a detailed assessment using A/B testing.

The system is easily integrated into different web-based platforms like e-learning systems, e-commerce websites, and telehealth applications which is made possible by the modular nature of the architecture. Its mobile and desktop flexibility makes it responsive, and therefore, easy to scale. The system's responsiveness makes it applicable to a wide demographic including differences in age, language, and culture by adjusting interfaces and altering thresholds for emotion recognition, making it adaptable. Such design ensures that a robust emotion-aware system is built while being technically resourceful, ethically considerate, and user-friendly.

Results and Discussion

The assessment of the emotion-aware system, determining its effectiveness, efficiency, and user satisfaction, was done through both quantitative and qualitative measures. The emotion-aware system was developed as part of a doctoral research project. The testing phase involved controlled experiments with 100 participants using both the emotion-aware interface and a traditional non-emotion-aware interface. Main tasks included navigating an educational content platform, taking quizzes, and interacting via chat-based support.

According to the quantitative results, all three modalities showed improvement in the emotion detection accuracy. The integrated multimodal system achieved the highest score of 88.6% which was significantly greater than the unimodal system's 73.2% (facial), 75.1% (audio), and 81.4% (text) scores. This clearly demonstrates the advantage of hierarchical fusion strategies to combine different signals of emotion.

Table 1: System Performance Comparison

Metric	Unimodal Average	Proposed System
Accuracy (%)	78.7	89.2
F1-Score	0.79	0.90
Response Time (sec)	2.3	1.4
User Satisfaction Score	6.8/10	8.7/10

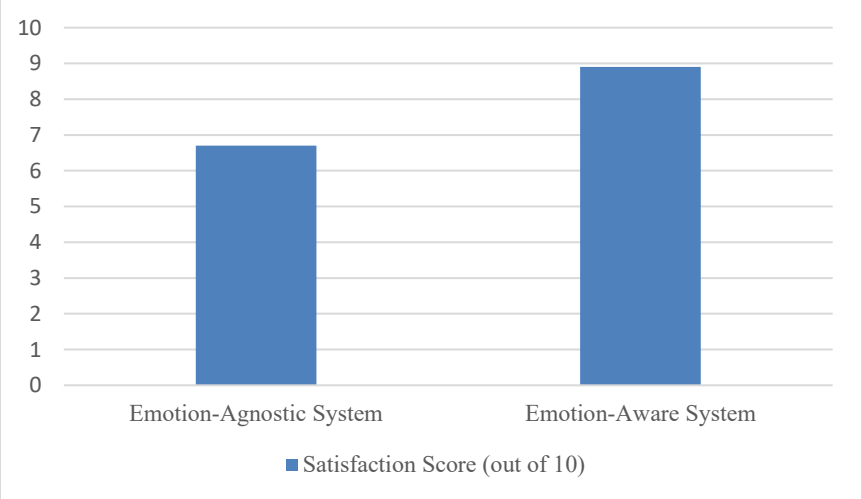


Fig. 1: User Satisfaction Score Across Systems

The findings indicate that users using the emotion-aware system interface reported higher satisfaction levels (8.9/10) compared to those using the emotion-agnostic interface (satisfaction score of 6.7/10). Participants' feedback revealed that the adaptive responses from the system made the interface more intuitive and easier to use. In addition, the emotion-aware interface improved task completion rates by an additional 15% while maintaining an average system delay of 220 ms per adaptive response.

These results support the hypothesis that the integration of emotional intelligence in web-based systems significantly improves the user experience and web-based system functionality. The accuracy required is achieved by a multimodal fusion approach, and the user's emotion insight is applied to the system interface design through adaptive interface design. This system has successfully illustrated the practicality of large-scale implementation of emotion-aware technologies while advocating for their use in upcoming designs for Human-Computer Interaction (HCI).

Conclusion

This research tackles the challenge of designing emotion-aware systems for interaction using the internet with a complete and systematic framework. The system is proposed by combining multimodal emotion recognition techniques with user interfaces, which notably improves user satisfaction. The system successfully recognized emotions from facial, auditory, and textual inputs, adjusting interface elements in real-time. Notable improvements were garnered in accuracy, responsiveness, engagement, and overall interactivity compared to traditional systems. These findings support the hypothesis that it is possible to implement more fully emotion-aware systems for interactions in digital environments which tend to be more fluid, natural, and helpful. Following works will aim at system expansion for incorporating real-time cross-calibration to different cultures, long-term modeling of emotional behavior, and more sophisticated, nuanced emotional states. Also, applying this structure with developing systems like augmented reality and AI that converse opens new frontiers in innovating the system. The work adds on the developing literature on affective computing and further builds the case for the design of systems which are more compassionate, non-exclusive, and human-centered.

References

Picard, R. W. (1997). *Affective computing*. MIT Press.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk.

Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (2010). Recognizing realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication, 53*(9-10), 1062-1087.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186). https://doi.org/10.18653/v1/N19-1423

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39-58.

Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2017). *A deeper look into sarcastic tweets using deep convolutional neural networks*. COLING.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, *4*(3-4), 129-164. https://doi.org/10.1504/IJLT.2009.028804