Multimodal Affective Computing in the Large-scale Pre-trained Model Era

A Special Issue of Pattern Recognition

1 Aims and Scopes

With the urgent demand for generalized deep models, many pre-trained large models are proposed, such as bidirectional encoder representations (BERT), vision transformer (ViT), generative pre-trained transformers (GPT), and further more. These models are being constantly adapted to many tasks to support our daily work and lives. For example, GPT has been widely used in many areas, such as machine translation, text summarization, and question answering. Inspired by the success of these models in single domains, the development of Vision-Language Models (VLMs) enables training unified models with dual-modal interactions. The CLIP (Contrastive Language-Image Pre-Training) model is trained based on large-scale image and language data from the internet, and it has been proven capable of extending to many downstream tasks without finetuning. Inspired by this, various multimodal pre-trained large models are being developed to facilitate artificial general intelligence (AGI).

Recently, a significant demand for emotion representation and understanding in artificial intelligence and multimedia affective analysis is attracting increasing research efforts from both academia and industries. With the rapid development of multimedia and social network technology, people get used to sharing their daily lives and opinions online, which results in a rapid growth in user-generated content including text, images, audio, and videos. Such data usually contain personal emotions and sentiments from which meaningful information can be discovered, e.g., depression and suicide tendencies which can be used for public safety. Under this circum-

stance, large-scale pre-trained models have been adapted to affective computing tasks gradually. Large Language Models (LLMs), such as ChatGPT, have been used for various affective computing problems, such as aspect extraction, aspect polarity classification, and opinion extraction via prompting in a zero-shot manner, however, their performances are still not comparable with existing specialized or finetune models, and biases are discovered from applying pre-trained LLMs on prompt-based sentiment and emotion analysis. Additionally, compared with LLMs, the application of large audio and video models (LAMs and LVMs) in affective computing is still at an early stage, rare studies of adapting models pre-trained on large-scale general-purpose data for vision-based affective computer problems are conducted. Existing methods are mostly for pre-training models on emotion and sentiment-related datasets using the image and language modalities with the contrast learning framework, but their capabilities of adapting to different affective computing tasks still require further exploration. Besides, there is still room for investigating the usage of other modalities and relevant combinations for affective analysis, such as vision-speech and vision-language-speech.

Multimodal affective analysis using large-scale pre-trained models is rather challenging due to the following reasons. Domain knowledge learned from general data is different from affective computing, so adequate approaches are needed for extracting desired domain knowledge, such as prompt engineering, efficient finetuning, and zero-shot learning for downstream task adaptation. Moreover, emotions are often jointly expressed and perceived through multiple modalities, effective pre-training frameworks can collaboratively extract information from multimodality such as vision, speech, and language need to be explored. Furthermore, as emotion is a subjective concept, affective analysis involves a multi-disciplinary understanding of human perceptions and behaviors. Consequently, extensive efforts are still needed to facilitate the application of large-scale pre-trained multimodal models for affective computing tasks.

This special issue seeks original contributions reporting the most recent progress on affective computing with foundation models pre-trained on large-scale multi-modal data. It targets a mixed audience of researchers and product developers from several communities, i.e., multimedia, machine learning, psychology, artificial intelligence, and further more.

The topics of interest include, but are not limited to:

- Affective content understanding with models pre-trained using large-scale uni-modal text, images, or speech data
- Affective content understanding with models pre-trained using large-scale multi-modal data
- Foundation models trained on large-scale multimodal affective data
- Frameworks for model training with large-scale data for affective computing problems
- Fusioin strategies for pre-training and fine-tuning affective foundation models

The topics of interest include, but are not limited to:

- Prompt engineering for using large-scale pre-trained models for affective computing problems
- Zero-shot and few-shot learning with large-scale pre-trained models for affective computing problems
- Emotion-aware artificial intelligence generated content
- Large-scale affective data collection and annotation
- Evaluation metric design for affective computing
- Affective computing-based applications in entertainment, robotics, education, health care, and biometrics, *etc.*

2 Important Dates

- Submission deadline: December 31, 2025
- First notification: March 31, 2026
- Revision submission: May 31, 2026
- Notification of acceptance: August 31, 2026
- Anticipated publication: October 2026

3 Paper Submission

Authors should prepare their manuscripts according to the Guide for Authors of Pattern Recognition available at https://www.sciencedirect.com/journal/pattern-recognition/publish/guide-for-authors and submit their papers at the submission page https://www.editorialmanager.com/PR/default.aspx. All the papers will be peer-reviewed following the rigorous Pattern Recognition reviewing procedures. Manuscripts should not be published or currently submitted for publication elsewhere. Submitted manuscripts should not have been published previously, nor be under consideration for publication elsewhere. If the submission is an extended work of a previously published conference paper, please include the original work and a cover letter describing the changes that have been made.

4 Guest Editorial Board