Title:	Final Project Report
Course Name:	Artificial Intelligence Applications
Course Number:	**(60613MEEOS-CME0327)**
Supervised by:	Dr. ABDULKADİR KAYIKLI
Department:	Computer Engineering
Students Name:	Yousef Rayyan
Student Id:	1904010031
Date:	Semester 2, 2024

Literature Review on Face and Voice Emotion Recognition Using Python

Introduction

Emotion recognition from facial expressions and voice is rapidly gaining significance with its increasing application in human-computer interaction, security, and entertainment. Several techniques have been applied to improve the absolute accuracy and efficiency of emotion recognition systems using Python libraries and other related deep learning frameworks.

Emotion Recognition from Speech

Koolagudi and Rao (2012) reviewed emotion recognition from speech, emphasizing prosodic and spectral features. Solutions of different machine learning techniques used in speech emotion recognition were discussed, with accompanying thought points and challenges faced by the authors, such as speaker dependency and cultural differences. This review of the works emphasized the necessity for robust methods of feature extraction and classification in the improvement of the performance of the SER systems.

Another job is by Rastogi, Anand, and Sharma (2023), which delved into using the Librosa library within Python for feature extraction in speech emotion recognition. They presented an overview of the effectiveness of deep learning models in emotion classification from speech signals, particularly in detecting frustration and annoyance. Their findings demonstrated evidence for the utility of Python-based libraries in supporting instance-efficient SER systems.

Issa, Demirci, and Yazici (2020) developed a framework for speech emotion recognition using deep convolutional neural networks. Their study indicated the advantages of employing CNNs as a method of feature extraction and classification of SER, which provided a significantly more accurate result compared to other conventional methods. They also discussed the fusion of speech and image data for multimodal emotion recognition.

Facial Emotion Recognition

Minaee et al. (2021) devised an attentional-based convolutional network-equipped facial expression recognition system named "Deep-Emotion." Such a current approach will give prominence to critical regions of the face to improve the accuracy of emotion detection using an attention mechanism. The system implementation was in Python and had high performance on the benchmark datasets.

Ko (2018) presented a review based on visual information for facial emotion recognition. The discussions presented in this paper are related to different techniques for feature extraction and classification, both traditional and deep learning models used in the same, based on which Ko emphasized the importance of Python libraries like OpenCV and TensorFlow in system development for FER.

The research conducted by Ali, Khatun, and Turzo (2020) used neural networks for detecting emotion from faces. They developed a convolutional neural network using the Keras library in Python for recognizing emotions from facial expressions. Their study demonstrated how widely used and effective Python frameworks were in developing FER systems.

Multimodal Emotion Recognition

Another research study by Ley, Egger, and Hanke (2019) focused on emotion detection systems and the use of facial and vocal features. They applied their Python scripts to analyze sound files. They later combined the output with video data to work toward increasing the accuracy of recognition of emotions.

The research by Li et al. (2021) on multimodal emotion recognition was part of a project aimed at developing a new open-source Python toolbox, MindLink-lumpy. This toolbox aims to reduce inaccuracies by integrating various data sources like facial expressions and speech. Their study highlighted multimodal datasets and the role of Python in coming up with such versatile emotion recognition systems.

Maningo et al. (2020) developed an intelligent space system with face and speech emotion recognition to determine the selection of music that befits the present emotional state of the user. The implemented system had practical applications through Python functions in the enhancement of user experience with the use of multimodal emotion recognition.

Conclusion

To sum up, the literature reviewed discusses advancements in the recognition of emotion through speech and facial expression based on Python. Much research has shown how efficient deep learning models and Python libraries are in raising the accuracy and time complexity of the emotion recognition system. Multimodal data integration for enhanced robustness of such systems opens up a way to more sophisticated and, hence practical applications in human-computer interactions.

References

- Koolagudi, S. G., & Rao, K. S. (2012). Emotion recognition from speech: a review. International Journal of Speech Technology, 15(2), 99–117. doi:10.1007/s10772-011-9125-1

- Minaee, S., Minaei, M., & Abdolrashidi, A. (2021). Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. Sensors, 21, 3046. https://doi.org/10.3390/s21093046

- Ko, B.C. (2018). A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors, 18, 401. https://doi.org/10.3390/s18020401

- Ali, M. F., Khatun, M., & Turzo, N. A. (2020). Facial emotion detection using neural network. Retrieved from ResearchGate

- Rastogi, R., Anand, T., & Sharma, S. K. (2023). Emotion Detection via Voice and Speech Recognition. International Journal of Research in Engineering, Science and Management. Retrieved from IGI Global

- Issa, D., Demirci, M. F., & Yazici, A. (2020). Speech emotion recognition with deep convolutional neural networks. Biomedical Signal Processing and Control. Elsevier. Retrieved from ScienceDirect

- Ley, M., Egger, M., & Hanke, S. (2019). Evaluating Methods for Emotion Recognition based on Facial and Vocal Features. Retrieved from ResearchGate

- Li, R., Liang, Y., Liu, X., Wang, B., & Huang, W. (2021). MindLink-eumpy: an open-source python toolbox for multimodal emotion recognition. Frontiers in Human Neuroscience. Retrieved from Frontiers

- Maningo, J. M. Z., Bandala, A. A., & Vicerra, R. R. P. (2020). A Smart Space with Music Selection Feature Based on Face and Speech Emotion and Expression Recognition. Retrieved from IEEE Xplore