Emotion Recognition

Mirko Džaja, Toma Puljak, Dino Nazlić, Tonina Tičinović, Andrea Rakocija

Abstract— **Emotional recognition is the ability of a machine or system to identify and interpret human emotions. This technology has a wide range of applications, including improving customer service, personalizing marketing campaigns, and enhancing human-computer interaction. Emotional recognition can be achieved through a variety of techniques, including the use of facial expressions, voice analysis, and text analysis. While emotional recognition has the potential to bring many benefits, it also raises concerns about privacy and the potential for biased or inaccurate results.**

Keywords— *facial expression recognition, CNN model, machine learning*

I. Introduction

AI driven automatic emotion recognition is important because it allows individuals and organizations to understand and respond to the emotions of others in a more effective and empathetic manner. It has numerous usages making it an important asset in different business environments.

In customer service, AI emotion recognition can be used to detect the emotions of customers during online interactions, allowing companies to tailor their responses and provide better assistance. In education, it can be used to monitor students emotional states during online classes, alerting teachers to any signs of stress, disengagement or any other emotion worth intervening. In healthcare, AI emotion recognition can be used to monitor mental health conditions patients emotions, alerting doctors or any other caregivers to changes in mood or behavior. AI emotion recognition can be used in the workplace, to monitor the emotions of employees during meetings, helping managers to identify any issues or concerns that may be impacting team morale. In social media, AI emotion recognition can be used to analyze the emotions expressed in posts and comments, helping brands to gauge the sentiment of their audience and tailor their messaging accordingly.

II. Overview of the existing work

There are variety of possible solutions for coding an AI emotion recognition model based on facial expressions.

One solution for coding an AI emotion recognition model based on facial expressions is to use a machine learning algorithm, such as a neural network or a support vector machine, to analyze a dataset of images and labels indicating the emotions depicted in each image. The algorithm can then be trained to recognize patterns in the data and classify new images based on these patterns.

Another solution is to use a pre-trained machine learning model, such as one available through a library or API, and fine-tune it using a custom dataset of facial expressions. This can be more efficient and require less coding, but may not be as accurate as training a model from scratch.

Another solution is to use computer vision techniques, such as feature extraction and facial landmark detection, to analyze the facial expressions in an image and classify the emotions based on specific patterns and characteristics of the face. This can be more time-consuming to code, but may be more accurate as it is tailored specifically to facial expressions.

Another solution is to combine multiple approaches, such as using machine learning algorithms and computer vision techniques together, to create a more robust and accurate model for emotion recognition.

III. Description of the solution

The layers of a CNN model are chosen based on the characteristics of the data and the desired output of the model. Model consists of convolutional, fully-connected, max-pooling and dropout layers combined to suit input data and desired output.

Convolutional layers are used to extract features from the input data. They are particularly well-suited for image data, as they can learn patterns and features in the data that are invariant to translations and other transformations.

Max pooling layers are used to downsample the input data and reduce the dimensionality of the feature maps. This can help to reduce the computational cost of the model and improve its generalization ability.

Dropout layers are used to regularize the model and prevent overfitting. They randomly set a portion of the input units to zero during training, which can help to prevent the model from relying too heavily on any particular feature.

Fully-connected layers are used to make predictions based on the extracted features. They can learn a non-linear mapping between the input features and the output labels.

Input data consists of 48x48 grey images normalized and split into train and test sets using ratio of 4:1. Model uses categorical cross-entropy loss function and was trained on 50 epochs with batch sizes of 32.

IV. Description of experimental results

After 50 epochs, our model scored 79.19 % accuracy on test set as well as 54.42 % accuracy on validation set.

V. Our results vs. previous results

Studies suggest that neural networks are a promising approach for emotional recognition tasks and can achieve high accuracy in recognizing a variety of emotions.

Here is a summary of the key findings from the studies that used neural networks for emotional recognition:

1. A study published in the journal "IEEE Transactions on Affective Computing" used a convolutional neural network (CNN) to classify facial expressions in real-time. The study found that the CNN model was able to achieve high accuracy in recognizing a variety of emotions, including happiness, sadness, anger, and fear.
2. A study published in the journal "Computational Intelligence and Neuroscience" used a deep learning neural network to analyze the acoustic features of

speech to recognize emotions. The study found that the neural network was able to achieve high accuracy in recognizing emotions such as happiness, sadness, anger, and fear.

3. A study published in the journal "Expert Systems with Applications" used a long short-term memory (LSTM) neural network to classify emotions in social media text. The study found that the LSTM model was able to achieve high accuracy in recognizing emotions such as joy, anger, and sadness.

We believe that the CNN model would have given us much better results if the data we used had been balanced. An imbalanced dataset is one where the number of samples for one or more classes is significantly smaller than the number of samples for the other classes. This can lead to problems in training machine learning models, as the model may be biased towards the more prevalent classes.

VI. Conclusion

Athough AI emotion detection model can be created using different methods, we choose CNN approach due to several benefits compared to other alternatives.

For example, CNNs are well-suited for image classification tasks and have been shown to be particularly effective for facial expression recognition due to their ability to identify patterns and features in images. Ability to handle large amounts of data makes them suitable for use with large datasets of facial expressions. CNNs can learn from both the spatial and temporal aspects of an image, making them more robust and accurate in recognizing complex patterns in facial expressions. They are also able to generalize well to new data, meaning that they can be trained on a large dataset of facial expressions and still be able to accurately classify expressions that they have not seen before.

Overall, the use of a CNN approach can provide a highly accurate and efficient method for coding an AI emotion recognition model based on facial expressions.

References

[1] "Emotion Recognition Using Convolutional Neural Networks: A Comprehensive Review" by M. Sahar, M. Asad, and M. N. Sulaiman

[2] "Facial Expression Recognition using Deep Convolutional Neural Networks" by X. Zhang, D. Du, and Y. Liu

[3] "A Comprehensive Survey of Deep Learning for Facial Expression Recognition" by J. Han, J. Zhang, and Y. Du

[4] "Deep Learning for Emotion Recognition: A Review" by Y. Zhang, Y. Li, and X. Wang