JOURNAL OF
EMERGING INVESTIGATORS

Article

Training neural networks on text data to model human emotional understanding

Preetha Sathish[1], Ajay Sathish Preetha[1]
[1] University High School, Irvine, California

SUMMARY

Recent advancements in artificial intelligence (AI) have revolutionized the field of computer science. Different subsectors of AI, like natural language processing (NLP) models, generative AI, computer vision, autonomous and recommendation systems, cybersecurity, quantum computing, etc., have helped automate human tasks, resulting in a tremendous amount of time and energy being saved. Despite the massive development of AI, all AI models lack one major factor, which is emotion. How can emotion be built into AI in order for it to develop the emotional intelligence of the human brain to interpret and understand emotions so that it could create more human-friendly interactions? In this work, we hypothesized that training neural networks to predict emotions using text-based sentiment analysis will lead to significant improvements in AI's abilities to classify emotional states. By using NVIDIA CUDA Toolkit and TensorFlow, we were able to create a sentiment prediction model that achieved an accuracy of 94% and predicted the six basic emotions of joy, sadness, anger, fear, love, and surprise. Concluding this research, we observed that neural networks can develop the habit of recognizing emotions. This can be further fed into complex AI algorithms and systems to fine-tune emotional intelligence resulting in more natural interactions, benefiting humans in the future.

INTRODUCTION

Evidence of the development of human emotions has been found in our ancestors six million years ago (1). Despite the tremendous amount of change that has occurred in the human world, emotions are one key factor that remains the same in humans even today. The six basic emotions of human life — joy, sadness, anger, fear, love, and surprise — have been permeating humans forever. With the recent emergence of artificial intelligence (AI), there have been a variety of AI models trained to accompany humans with multitude of tasks like learning, art and video generation, task automation, organization and presentation automation (2). Despite AI's almost flawless sense of intelligence, one of the biggest drawbacks that it has is comprehending emotions (3, 4). This creates challenges for humans to interact with AI, as humans have a difficult time communicating their needs and emotions to these systems in their everyday lives. For example, when hospitals use AI with a lack of emotional intelligence, it struggles to comprehend whether a person is happy or angry and to understand their thoughts (4). This can result in improper care of a patient which can be hurtful. For example, a mischaracterization of a patient's emotions, like happiness and pain, can lead to the AI administrating medication like painkillers to the patient even though the patient is not in pain but rather happy. This could lead to the increase of the patient's stress levels, prolong suffering, and escalate their symptoms ultimately harming the patient's life.

Recently, machine learning models have been developed by scientists that could recognize human emotions (5). This process starts with selecting an emotion model (EM) or creating custom neural networks. The data is given by using a standard sentiment dictionary where there are various forms of input like text, speech, and images. This data is further processed by vectorization, where the EM associates different emotions with a range of numerical values. This is further fed into a machine learning algorithm where the model can be trained to produce a formula to predict emotions.

However, there is limited research on the effectiveness of AI comprehending emotions in high-stress environments like hospitals, as mentioned in the example above (4, 6). It is not known how well AI can interpret emotions in order to properly interact with humans in aiding them to complete a task. The idea behind the creation of emotional machine learning models is to produce machines that feel more human-friendly for humans to work with (4, 6). Thus, we hypothesized that training neural networks to predict emotions using text-based sentiment analysis will lead to significant improvements in AI's abilities to classify emotional states. Following the completion of our research, our emotion model achieved an accuracy of 94% in predicting emotions. With this information, future AI models could be further refined to improve emotional intelligence to accurately predict human emotions in high-stress environments, like hospitals for example, enhancing patient care and establishing greater emotional Human-AI relationships.

RESULTS

We first searched the internet for emotional datasets that could be used to train machine learning models. We came across a GitHub repository, DAIR.AI Emotion Dataset which consisted of a file with a piece of text (e.g., I feel so enraged but helpless at the same time) and the classified emotion (e.g., anger) labeled next to it for example. This dataset represented the six basic types of emotions: joy, sadness, anger, fear, love, and surprise, and each of the respective numbers of samples for each type of emotion (**Figure 1**) (4).

Next, we vectorized the emotions category of the dataset so that it could be fed into the custom machine learning neural network that we were going to build. We developed a custom

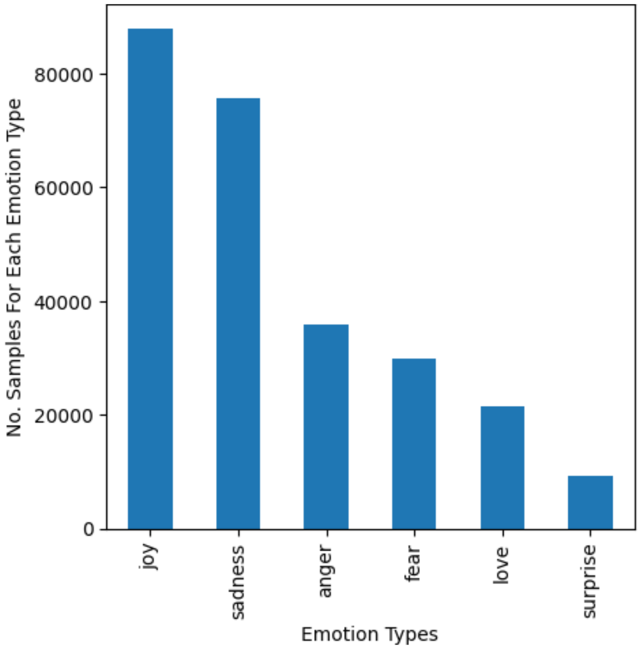

Figure 1: Emotion dataset graph. The six basic types of emotions: joy, sadness, anger, fear, love, and surprise are shown with the respective number of samples for each type of emotions. The dataset used to train this model was fetched from the DAIR.AI Emotion Dataset repository (12). This dataset was originally crafted for Natural Language Processing (NLP) Tasks for the use of emotion classification.

neural network based on TensorFlow Keras library (7). Keras is a high-level application programming interface (API) used to build deep learning neural networks. Keras consists of two main components called layers and models. Layers are a major component of network topology in AI that perform a simple input/output transformation. Models are simply a directed acyclic graph (DAG) consisting of layers (7). A DAG is a one-way flowchart where there are components called layers, which are the building blocks of AI which process data in a step-by-step manner to help the model recognize and make sense of the information being provided.

By using a system of models and layers, we built a neural network using the TensorFlow Keras Library that trained on the vectorized dataset. This model is a sequential model which is typically used in text classification tasks. This is especially useful as data is allowed to be processed in a linear manner, where the output of one layer is the input of another layer which is well suited for text classification tasks, where the data needs to be trained through a series of transformations like embedding, feature extraction, and classification etc. This model combines an embedding layer, convolutional layers, Long Short-Term Memory (LSTM) layers, and fully connected (dense) layers which all work together in order to understand meanings, extract the useful meanings and features from a piece of text, capture relationships between the words and their context, and classify text based into specific categories based on their learned features, in this case emotions. This model has a total of 4,332,230 parameters which can be trained. The training process of this model involved feeding the dataset through the emotion model in multiple epochs, complete rounds of training the dataset through an algorithm. As we continued training and evaluating the emotion model, we observed that there were steady improvements in the accuracy of the emotion model which capped out at about 94% overall (**Figure 2**). However, as we continued to further

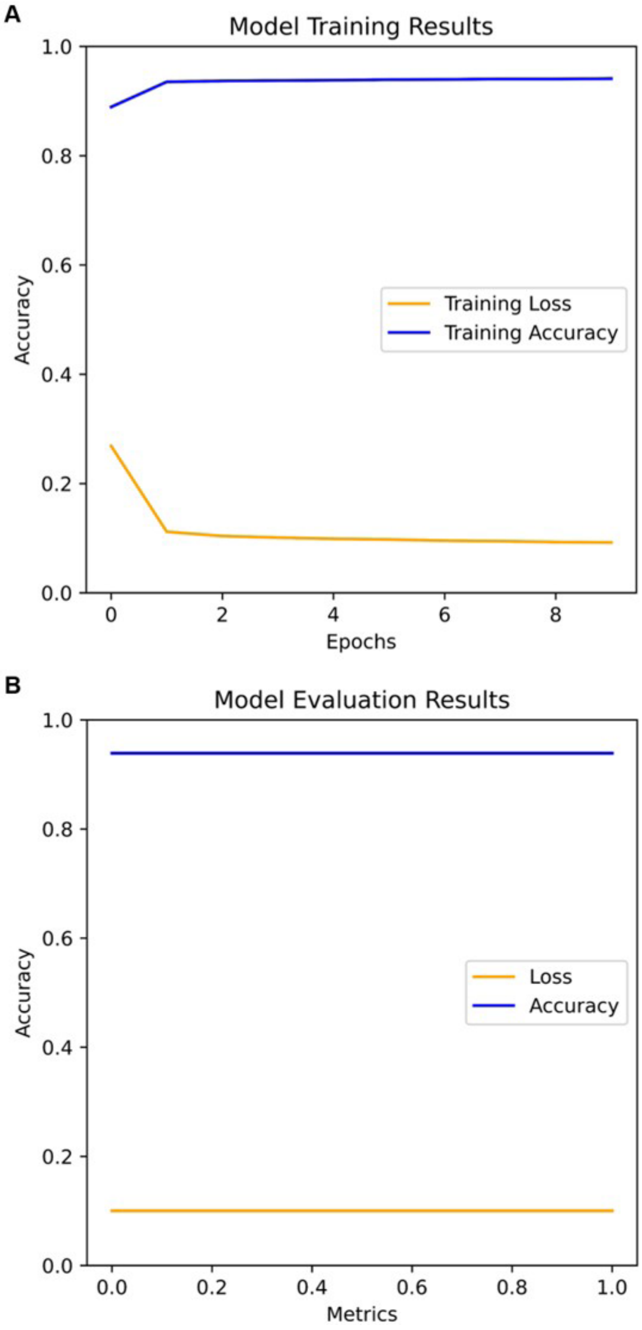

Figure 2: Training and evaluation results. A) The findings from the training process, along with the accuracy and loss metrics after three, five, and ten epochs of model training are presented. The trend of increasing accuracy and decreasing loss with more epochs proves the model's ability to learn information from the provided dataset. **B)** The evaluation results, including the accuracy and loss metrics after three, five, and ten epochs are presented based on the model's performance in predicting emotions from newly encountered data. This was taken from the 25% test split of the dataset. We can see that the model has performed very well on all 3 trials, achieving more than 90% with losses only about 1% each time, proving that the model is able to comprehend emotions very well.

https://doi.org/10.59720/24-014

increase our model's performance, we faced a multitude of hardware issues. Further attempts and the increased number of epochs resulted in overheating issues in our hardware as well as frequently crashing our system which we tried our best to resolve through reducing batch sizes, optimizing model parameters, and implementing early stopping to avoid overtraining.

DISCUSSION

This experiment showed us that neural networks could be created and trained in order to develop emotional interpretations of given prompts, in this demonstration, text prompts specifically. Contrary to other sentiment analysis methods which use a rule-based approach we have used lexicon-based methods to evaluate emotions. A rule-based method only uses three values to evaluate emotions: "-1" for negative emotions, "0" for neutral emotions, and "+1" for positive emotions. On the other hand, lexicon-based methods for emotion detection involve assigning specific emotions (such as happiness, sadness, and joy, etc.) to certain words or phrases. A value or score is then assigned to these words or phrases to indicate how closely they are associated with a specific emotion. The model then uses these parameters to predict a text's overall emotion while analyzing a given text prompt. Thus, the lexicon-based method provides a better understanding of the fact that artificial intelligence is able to predict emotions, providing actual labels, on a deeper level compared to the traditional rule-based method (8).

Despite the success of the emotion model in predicting emotions, we did encounter a few issues in our research. One issue that we found out while testing out this model is that it had a tendency in choosing the emotions joy and sadness more compared to the rest. It had less accuracy in predicting the rest of the emotions. We believe this is due to the increased number of samples of joy and sadness compared to the rest of the emotions in the dataset that we used (**Figure 1**). We also believe that a larger dataset could have been used to train the model to achieve more accurate results, something that we lacked. A larger dataset would further enable us to train the model with a much more balanced and extensive set of information preventing the emotion bias while also teaching the model to predict emotions more accurately when presented with a variety of scenarios and text prompts. The lack of good hardware was the second biggest issue that we ran upon while testing the model. Better hardware would have prevented the computer from crashing while training the model, ensure smoother workflow and runtimes, and would enable us to train bigger datasets ensuring an even better accuracy and precision of the model.

An example use case would be using emotion AI models in healthcare. One potential application is using emotional text analysis to analyze patient descriptions of their illnesses to tailor prescriptions and administer care based on their emotional states (9). For instance, this would be used with people who have dementia (5). People with dementia have a hard time understanding their feelings and communicating them to their caregivers. This puts a tremendous amount of pressure on caregivers on deciding what they need and how they are feeling. Such emotion AI models are able to analyze biometrics or psychometrics like facial expression, speech, or behavior to predict the state of being, further aiding the caregiver in administering the patient more easily. They can also increase compassion toward caregivers as well as increasing their stamina (5). In essence, this scenario highlights just one of the vast potentials of emotion AI models.

By conducting this experiment, we determined that neural networks do have the ability to predict emotions. Our research aims to provide a starting base to the discoveries of artificial intelligence and ability to understand emotions with its impressive 94% accuracy in predicting emotions using text data using the improved lexicon-based method. Emotion AI models can also be paired with other plugins provided by providers, like OpenAI for example, and their services like their photo and video generation, and text-to-speech (TTS) etc., and other multipurpose artificial intelligence systems and services to provide more natural and seamless interactions (10, 11).

MATERIALS AND METHODS

The dataset used in this experiment was collected from the GitHub repository DAIR.AI Emotion Dataset (https://github.com/dair-ai/emotion_dataset) (12). The data was cleaned and preprocessed and the emotions were replaced with numerical values: "anger": 0, "fear": 1, "joy": 2, "love": 3, "sadness": 4, "surprise": 5, according to the lexicon-based method. We used Anaconda to create custom environments that could be populated with specific packages associated with our task and be able to maintain and control them (2). Packages that we used to conduct our research were NumPy, pandas, tensorflow, and matplotlib. We also used NVIDIAs CUDA toolkit, which allowed us to use our GeForce GTX 1650 GPU on our HP - Pavilion Gaming Laptop and TensorFlow together to train the model. The integrated development environment (IDE) used to write code was Jupyter Notebook.

We started by taking the cleaned and preprocessed dataset and split it into 75:25 train to split ratio named, "Training Sentences & Training Labels" and "Testing Sentences & Testing Labels". These were then tokenized and padded to create uniform-length arrays. Tokenization was performed with a vocabulary size of 40,000. An embedding dimension of 100 was attached and an out-of-vocabulary (OOV) tokenizer was used to handle unseen words. The tokenized sentences and labels were then fed into the model created using TensorFlow Keras API. Then the model was created using TensorFlow's Keras API where we created a sequential model where individual layers were created and stacked together using embedding, convolution, Long Short-Term Memory (LTSM), and dense layers to create the neural network. The model was thus trained, evaluated, and finally tested. Weights, layers, and hyper-parameters were adjusted to achieve the best performance and intended output of the model during this experiment.

For the accuracy of 94%, we trained the model for 10 epochs, implemented an early stopping mechanism with a patience of 3 epochs, and an automatic restoration of the best weights that were observed during the training process in order to further improve the accuracy of training in the future. Evaluation was done through TensorFlow's prebuilt evaluate function which evaluated the 25% of the dataset which we separated for testing in order to provide insights to how well the model is able to predict emotions through unseen data. Here, a sentence could be provided to test and the model would predict a numerical value according to the respective emotional category. Further, the numerical values

https://doi.org/10.59720/24-014

will be remapped to the original emotional labels: “anger”, “fear”, “joy”, “love”, “sadness”, and “surprise”, allowing for a more exact and interactive validation of the model’s emotional classification based on user-input.

Code for this project can be found at github.com/Aj-Cdr/Development-of-Neural-Networks-as-Predictive-Sentiment-Analysis-Models.

Received: January 9, 2024
Accepted: August 5, 2024
Published: March 14, 2025

REFERENCES

1. Spikins, P.A., et al. “From Homininity to Humanity: Compassion from the Earliest Archaics to Modern Humans.” *Time and Mind*, vol. 3, no. 3, Jan. 2010, pp. 303–325, https://doi.org/10.2752/175169610x12754030955977.
2. “Getting Started with Anaconda Distribution.” *Anaconda.com*, 2024, docs.anaconda.com/free/anaconda/getting-started/what-is-distro/. Accessed 20 Oct. 2024.
3. Liu, Ruibo, et al. “Modulating Language Models with Emotions.” *ArXiv (Cornell University)*, 1 Jan. 2021, https://doi.org/10.18653/v1/2021.findings-acl.379. Accessed 26 Aug. 2023.
4. García, Rosa A. “A Systematic Literature Review of Modalities, Trends, and Limitations in Emotion Recognition, Affective Computing, and Sentiment Analysis.” *Applied Sciences*, https://doi.org/10.3390/app14167165.
5. Hockley, Janine, et al. “To AI or Not to AI: That Is the Question in Mental Health Nurse Recruitment.” *Issues in Mental Health Nursing*, vol. 45, no. 11, 29 Apr. 2024, pp. 1–4, https://doi.org/10.1080/01612840.2024.2341043 . Accessed 18 June 2024.
6. Kapoor, Amit, and Vishal Verma. “EMOTION AI: UNDERSTANDING EMOTIONS through ARTIFICIAL INTELLIGENCE.” *International Journal of Engineering Science and Humanities*, vol. 14, no. Special Issue 1, 17 May 2024, pp. 223–232, https://doi.org/10.62904/0vcbvb24.
7. TensorFlow. “Keras | TensorFlow Core | TensorFlow.” *TensorFlow*, 2019, www.tensorflow.org/guide/keras.
8. Devika, M.D., et al. “Sentiment Analysis: A Comparative Study on Different Approaches.” *Procedia Computer Science*, vol. 87, 2016, pp. 44–49, https://doi.org/10.1016/j.procs.2016.05.124.
9. Hindelang, Michael, et al. “Transforming Health Care through Chatbots for Medical History-Taking and Future Directions: Comprehensive Systematic Review.” *JMIR Medical Informatics*, vol. 12, 29 Aug. 2024, p. e56628, https://doi.org/10.2196/56628. Accessed 21 Sept. 2024.
10. “ChatGPT Enterprise.” *Openai.com*, 2023, openai.com/chatgpt/enterprise.
11. “Research.” *Openai.com*, 2024, openai.com/news/research. Accessed 20 Oct. 2024.
12. Saravia, Elvis, et al. “CARER: Contextualized Affect Representations for Emotion Recognition.” *ACLWeb*, Association for Computational Linguistics, 1 Oct. 2018, aclanthology.org/D18-1404/.