IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁙ Impact Factor 8.102 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125238

Emotion Recognition from Formal Text (Poetry)

Prof. Vishal Walunj[1], Ankit Choudhary[2], Ameya Kale[3], Mangesh Gade[4]

Guide, Department of Computer Engineering, D. Y. Patil Institute of Engineering & TechnologyAmbi, Pune[1]

Department of Computer Engineering, D. Y. Patil Institute of Engineering & TechnologyAmbi, Pune[2-4]

Abstract: The classification of emotional states in poetry or formal texts has received less attention from experts in computational intelligence than informal textual content, such as SMS, email, chat, and online user reviews.

This work introduces a technique for classifying emotional states in poetry using cutting-edge Artificial Intelligence technology known as Deep Learning in order to fill this knowledge gap. To analyse the poetry corpus and categorise the text into different emotional states, such as love, joy, hope, grief, anger, and others, the system uses an attention-based C-LSTM model.

I. INTRODUCTION

The classification of ideas, sentiments, and emotional states has drawn interest from specialists in a number of disciplines, including natural language processing, computational linguistics, and artificial intelligence. Machines are capable of analysing both formal and informal textual content. Poems, novels, essays, plays, and official or legal documents are examples of formal text, whereas SMS, conversation, and social media posts are examples of informal text.

Due to its complexity, identifying and categorising emotional experiences in formal language, especially poetry, can be difficult. However, emotional states and themes from both formal and casual text have been successfully extracted and analyzed using machine learning techniques, such as multi-label emotion categorization. This is particularly helpfulfor categorising mixed data, which combines two or more languages.

Poetry can be divided into two emotional categories using Support Vector Machines (SVM) and a BiLSTM classifier. However, by utilising the Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), and the Attention mechanism of deep learning, an Attention-based C-LSTM model can further enhance this categorization. This methodology enables a more precise classification of emotional states in poetry by allowing poetry to bedivided into up to five emotional classes. This is a substantial advance in the classification ofemotional states in formal language and is an extension of the baseline study.

II. RELATED WORK

A brief summary of pertinent studies on the classification of emotions in poetry text is given in this section. Using machine learning approaches, various academics have recently done studies in the area of emotion identification. For instance, Sreeja and Mahalakshmi [1] created an emotional state detection system from poetry text by dividing it into several emotion categories and using the Naive Bayes machine learning classifier to categorise the results.

The classification of emotions from poems has also been attempted using other methods, such as contrasting SVM and Naive Bayes algorithms. However, the small dataset size is a problem that all models and methods struggle with.

It could be beneficial to incorporate many contexts or domains to increase the model's adaptability. For instance, adding phonemic features might allow the model to categorise emotions from an audio clip, like music. Additionally, the model might be used as a translator between two people if a speech-to-text converter is added.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁙ Impact Factor 8.102 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125238

III. LITERATURE SURVEY

SNo.	Journal, Year	Author	Title	Abstract
1.	IEEE, 2021	Iqra Ameer, Geigori Sidorov	Multi-Label Emotion Classification on Code-Mixed Text: Data and Methods	Using ML, DL, Transfer based learning to classify emotion inmixed data
2.	EEE, 2021	Chanf Liu, Taiao Liu, Shuojue Yang,and Yajun Du	Individual Emotion Recognition Approach Combine Gated Recurrent Unit with Eotion	Using BI-GRU to Classify emotion.
3.	IEEE, 2020	Hassan Alhuzali andSophia	Improving Textual Emotion Recognition Based on Intra- and Inter-	Using Variant Triplet center loss (VTCL) for emotion
4.	IEEE, 2019	Erdenebileg Batbaatar, Meijing Li, and Keun Ho Ryu	Semantic Emotion Neural Network for Emotion Recognition from Text.	Using SENN model Which can utilize both sematic/syntactic and emotional relationship information by

Table 1: Literature Survey

IV. METHODOLOGY AND MATERIALS

1. Depending on the size of the dataset and the complexity of the ML models, the hardware requirements for this project employing ML models can change, however some typical requirements include:

- **CPU:** To accommodate the computational demands of running ML algorithms, a multi-core CPU with a high clock speed is advised.
- **Memory:** It is advised to have at least 8 GB of RAM, while additional memory may beneeded depending on the size of the dataset and the complexity of the models.
- **Storage:** To store the dataset and the software needed for the project, a solid-state drive (SSD) with enough storage space is advised.
- A graphics processing unit (GPU) might be needed for the computationally expensive deep learning models.
- **Operating System:** A 64-bit operating system, such as Windows 10 or Ubuntu, isrecommended to support the software and libraries required for the project.

2. Depending on the size of the dataset and the complexity of the models used, the software needs for this project utilising ML models can vary, however some typical requirements include:

- **Programming language:** To implement the ML models, a programming language is required, such as Python or R.
- **Machine Learning Libraries:** A selection of ML libraries and software, including Pandas, NumPy, NLTK, Matplotlib, scikit-learn, etc. These libraries offer pre-made training, validation, and prediction algorithms and routines.
- **Software for data analysis and visualisation:** Tools for data analysis and visualisationinclude R Studio

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁙ Impact Factor 8.102 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125238

and Jupyter Notebook. This program can be used to examine the dataset, get it ready for modelling, and display the outcomes.

- **Database management software:** The dataset may need to be stored and managed using a database management system, such as SQLite or MySQL.

3. System Overview

This project is carried out in the same manner as any machine learning project, which consists ofthe steps listed below:

- **Data collection** is one of the most crucial processes in the training of a deep learning model. To train a model without introducing undertraining, the dataset should be sizable enough.

- **Data Pre-processing:** It's critical to ensure that the data is fit for use before implementing the deep learning model. As a result, in order to ensure that the output of our model is accurate, we must remove any noise, outliers, incorrect, or incomplete data.

- **The words are then converted into numbers in the following stage.** So, fundamental pre-processing operations including stop-word removal, lowercase conversion, and tokenization are carried out. Following tokenization, a vocabulary is created that converts word sequencesinto integer sequences, where each number stands for a different word in the vocabulary.

- **Feature Representation and Extraction:** Each word is converted into an embedding vector to help the model learn, and features are then retrieved from these embedded vectors, whichare then used as input.
- The input is then placed through a multi-label classification process (more than twoclassifications are considered multi-label).

4. System Architecture

This section describes the model's suggested architecture for categorizing poetry into different emotional states, such as happiness, rage, terror, melancholy, etc. The model presents the outcome after receiving text input via a Python GUI application. For this project, an alternative would be to create a web-based frontend that would communicatewith the backend using HTTP request/response.

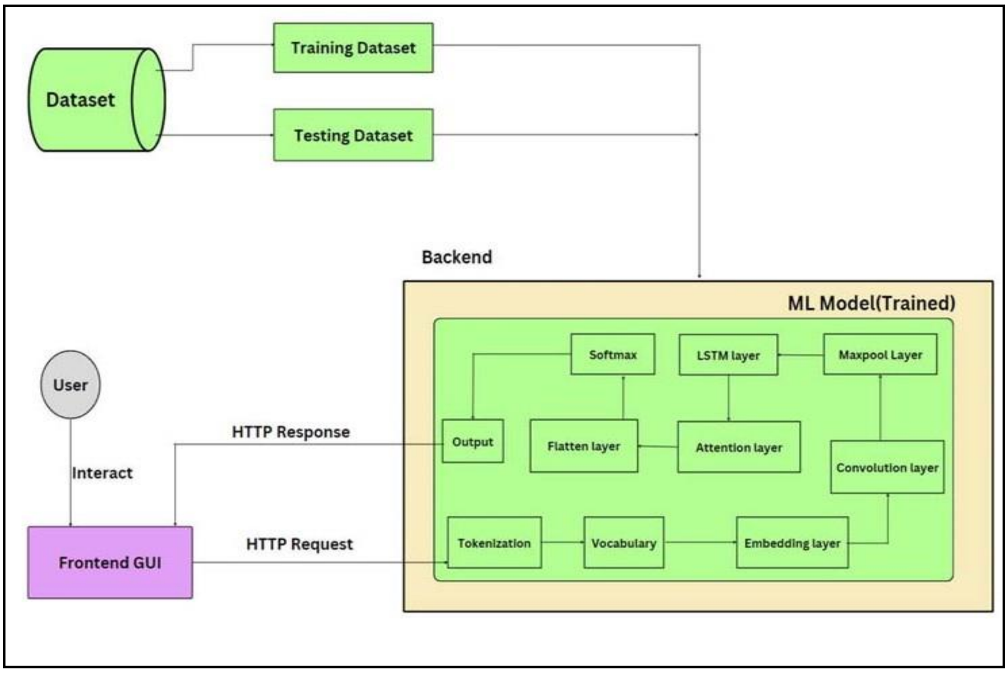

Figure 1: System Architecture

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125238

The suggested architecture has numerous layers, including:

- **Embedding Layer:** This layer represents text using pre-trained 300-dimensional word vectors taken from Wikipedia articles. The pre-trained vectors already capture the similarity of words with comparable meanings, making this technique superior to training word vectorswith our tiny dataset.

- **Deep Network:** This layer compresses the series of embedding vectors into a representation that fully encapsulates the text's information. For this portion of the network, an RNN (or a version like LSTM/GRU) is typically utilised. Overfitting can be avoided by including dropout.

- **Fully Connected Layer:**
   - This layer translates the RNN, LSTM, or GRU's deep representation into the finaloutputi classes or class scores.
   - It consists of batch normalization, fully connected layers, and optional dropout layersfor regularization.

- **Output Layer:** Depending on the issue at hand, this layer employs either Sigmoid (for binary classification) or SoftMax (for both binary and multi-class classification).

V. MATHEMATICAL MODEL

A convolutional layer is used in the C-LSTM (Convolutional LSTM) algorithm, a variation of the LSTM (Long Short- Term Memory) technique, to extract spatial information from input data. The following are some mathematical expressions for the C-LSTM model:

Let h_t be the hidden state vector, c_t be the cell state vector, and x_t be the input vector at time step t.
The convolutional layer creates a set of feature maps (f_k) from the input vector (x_t) after applyinga series of filters (w_k). The formula for calculating each feature map is: f_k = w_k * x_t, where * stands for the convolution operation.

The hidden state and cell state vectors are then updated by the LSTM layer in the manner describedbelow:

(W_i * f_t + U_i * h_t-1 + b_i) = i_t (W_f * f_t + U_f * h_t-1 + b_f) = f_t (W_o * f_t + U_o * h_t-1 + b_o) = o_t
W_c * f_t + U_c * h_t-1 + b_c = g_t i_t * g_t + f_t * c_t * c_t-1 = c_t
W_i, W_f, W_o, W_c, U_i, U_f, W_o, and W_c are weight matrices, b_i, b_f, W_o, and W_c are bias vectors, is the sigmoid function, and tanh is the hyperbolic tangent function. Backpropagation through time and gradient descent can be used to train the model to minimize a given loss function.

VI. ADVANTAGES

The following are some of the most frequently cited advantages of emotional classification of formal language (poems) using ML:
1) Scalability: Manual formal data analysis and organisation are slow and significantlyless effective.
2) In a handful of seconds, ML can accurately categorise a sizable quantity of texts.
3) Real-time analysis: This methodology can be used to quickly extract observations from formal material, such as court papers, in addition to poems.
4) This methodology can also assist those without a background in literature inunderstanding poetry and gaining interest in the material they are seeking to read.
5) This methodology can aid readers in appropriately classifying the material byminimizing whatever bias they may have.

VII. LIMITATIONS

ML approach for classifying emotions in formal language (poetry) has numerous benefits, but italso has certain drawbacks. Below is a list of a few of such difficulties:

- **Scheduling takes time:** Since training and testing the model require time, it takes some time until it is completely operational.
- **Computationally demanding:** Running the model can be computationally demanding because it uses

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125238

a lot of RAM, computing power, and other resources.

• **Skilled interpretation of outcomes:** Even though the model can provide results, a skilled person is still required to interpret the results correctly.

• **Language barrier:** If two poems are written by two different people, they will differ from one another. Therefore, it might be problematic when the same words appear in multiple poems because they can have diverse meanings and be categorized under various emotions.

VIII. CONCLUSION AND FUTURE WORK

This study's goal was to investigate several strategies for categorizing emotional states in formal language, particularly poetry, using machine learning algorithms. The study improved its grasp of the issue and determined which algorithms would be useful by examining earlier research. The model will undergo additional analyses following its deployment to determine its correctness and spot areas for improvement in order to guarantee that it produces accurate andworthwhile findings.

IX. REFERENCES

[1] P. S. Sreeja and G. S. Mahalakshmi, "Emotion recognition from poems by maximumposterior probability," International Journal of Computer Science and Information Security, Volume 14, pages 36–43.
[2] "Punjabi poetry classification: The test of 10 machine learning algorithms," J. Kaur and J. R. Saini methods," in Proc. 9th International Conference on Machine Learning and Computing (ICMLC), 2017, pp. 1–5.
[3] "Sad or glad?," by G. Mohanty and P. Mishra. "Corpus creation for Odia Poetry with Sentiment Polarity Information," in Proceedings of the 19th International Conference on Computer-Assisted Language Learning (CICLing), Hanoi, Vietnam, 2018.
[4] "Analysing Sentiment in Classical Chinese Poetry," Y. Hou and A. Frank, 9th SIGHUM Workshop on Language, Technol. Cultural Heritage, Social Sciences, and Humanities (LaTeCH), 2015, pp. 15–24.
[5] "SemEval- 2015 task 11: Sentiment analysis of figurative language in Twitter," in Proc. 9th Int. Workshop Semantic Eval. (SemEval), 2015, pp. 470-478. A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes.
[6] "Automated analysis of Bangla poetry for classification and poet identification," in Proc. 12th Int. Conf. Natural Lang. Process., Dec. 2015, pp. 247-253. G. Rakshit, A. Ghosh, P.Bhattacharyya, and G. Haffari.
[7] "Emotion classification in Arabic poetry using machine learning," International Journal of Computer Applications, vol. 65, no. 16, May 2013. O. Alsharif, D. Alshamaa, and N. Gheim.
[8] "Semantic-Emotion Neural Network for Emotion Recognition from Text," IEEE August 2019, by E. Batbaatar, M. Li, and K. H. Ryu.
[9] "Individual Emotion Recognition Approach Combined Gated Recurrent Unit with Emotion Distribution Model," IEEE December 2021; C. Liu, T. Liu, S. Yang, and Y. Du.

