Deployment of the deep learning fusion method to emotional semantic evaluation of natural language

Gulista Khan* | Sanjeev Kumar Mandal* | Sunil Sharma*

*Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India, Associate Professor, College of Computing Science and Information Technology.

Abstract The emotional semantic evaluation of natural language plays a crucial role in sentiment analysis. Deep learning methods have shown great potential in capturing the complex relationships between words and emotions. This paper proposes a deep learning fusion method for deploying emotional semantic evaluation. The technique combines multiple deep learning architectures to capture local and global contextual information, including Bidirectional Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM) networks, and Self-Attention mechanisms. Pretrained GloVe word embedding's utilized to enhance word representation. A novel fusion layer combines the outputs of individual models; employing self-attention means to assign weights dynamically. This allows the model to weigh the importance of different representations in the final prediction. Benchmark movie review (MR) for sentiment analysis and emotion classification tasks are used to evaluate the proposed method. Experimental results demonstrate superior performance compared to individual deep learning models and traditional feature-based approaches. The proposed fusion method effectively captures the nuances of emotional semantics in natural language, leading to more accurate and nuanced evaluations.

Keywords: deep learning, GRU, LSTMs, self-attention mechanisms

1. Introduction

Emotions fundamentally impact human communication and comprehension. They affect how we see the world, how we act, and how we want to perceive the world (Patil et al 2019). There has been a rise in interest in comprehending and assessing the emotional content encoded in natural language as computational linguistics (ClinLing) and natural language processing (NLP) have developed. The dynamic semantic evaluation of natural language is the task of evaluating the emotional valence, intensity, and attitude provided by text or audio data. It seeks to identify and extract emotional cues from verbal phrases, from straightforward assertions to intricate narratives.

The investigation of emotional semantics in natural language has wide-ranging effects (Shaheed et al 2021). Businesses can determine client happiness and sentiment toward their products or services by using sentiment analysis techniques, such as recognizing the emotional undertone of customer reviews or social media remarks. Dynamic semantic analysis can be used in psychotherapy and mental health research to analyse and track emotional states and identify potential mental health problems.

Manual annotation or language conventions are frequently used in traditional techniques for emotional semantic evaluation (Ayzeren et al 2019). These techniques, however, have a limited ability to scale and require considerable time and labor. Researchers have been able to create automated systems that can learn from and infer emotional content from vast volumes of text data thanks to the development of machine learning and deep learning techniques.

Traditional methods for evaluating dynamic semantics frequently rely on manual annotation or language conventions (Sommerfeldt et al 2019). These techniques, however, have a limited potential to scale and are time- and labor-intensive. Researchers have created automated systems to learn from vast amounts of text data and infer emotional content. This has been made possible by developing machine learning and deep learning techniques.

Recurrent neural networks (RNNs) and attention mechanisms, in particular, have shown promise in recent developments in deep learning for capturing the nuanced emotional semantics found in natural language (Shopon et al 2021). These models can focus on the temporal dynamics and environmental factors influencing emotional expression.

Our research aims to create a technique that accurately captures the subtleties of emotional semantics in natural language. By utilizing various architectures' advantages and including local and global contextual data, we seek to outperform our deep learning models and conventional feature-based techniques.
The additional divisions of this article are as follows: Part 2 introduces related works, Part 3 discusses the methodology, Part 4 assesses the efficiency of the proposed method, and Part 5 concludes the paper.

2. Related works

Guarino et al. (2022) created a method employing open-ended semantic questions to evaluate psychological variables, including emotions, thoughts, and attitudes. The findings demonstrated competitive or greater reliability and validity compared to rating scales, indicating that the semantics questioning method had good statistical features. According to these results, natural language-based semantic measurements may enhance and extend the use of conventional rating scales in the size and description of psychological variables. A study (Svetlakov et al 2021) proposed the Semantic-Emotion Neural Network (SENN), a unique neural network design that addresses the shortcomings of current emotion recognition models in NLP. It comprises two subnetworks: a CNN for word interactions and emotional aspects and a single-directional LSTM for background information and semantic linkages. The SENN model has shown much better emotion identification than existing techniques.

The study (Dargan and Kumar 2020) used bigrams, a semantic distance, to uncover new information on topic flow and conceptual cohesiveness in samples of continuous English. The distance measurements were tested against data from simulated verbal fluency, and bigram length norms were created using a substantial corpus of text. The method shows potential in elucidating how semantics are processed in real-world tales and bridges the gap between small-scale single-word research and more thorough discourse analysis. Jain and Kanhangad (2019) proposed a graph-based, semisupervised method for capturing and modelling the various language subtleties and events involved in textual emotion expression. Our approach provides the building blocks for developing contextualized affect representations by generating rich structural descriptors. They use word embedding to improve these representations and test their effectiveness on different emotion recognition tasks. The experimental results show that our approach is superior to state-of-the-art methodologies. Sharma et al. (2021) compared several methods for determining sentiment or mood in interactions between therapists and clients during psychotherapy sessions. The researchers used a database of 97,497 statements made during psychotherapy sessions to carry out their investigation. The BERT model was trained using sentiment assessments made by people. By comparing the individual kappa values of each model, the researchers assessed the performance of each. These technological developments offer a viable way for researchers to conduct extensive research on sentiment analysis in psychotherapy, previously constrained by time-consuming and manual procedures.

Guarino et al. (2023) suggested a big data sentiment analysis technique based on sensitive information subjects. The aim is to use natural language processing technology to extract sentiment tendencies from the massive volume of text data produced on the Internet. The process described in the study uses a neural network model to combine semantic subject data into a text representation. The attention mechanism in the model aids in determining the significance of every sentence in the text. In the age of big data, the model's adaptability is further improved by incorporating sentiment dictionary labelling, making it a useful tool for monitoring public opinion and sentiment analysis. Pereira et al. (2023) investigated the application of NLP (natural language processing) methods to collect and analyse emotions expressed on social networking platforms. With the ability to gather data continuously across time and reduce the input of data and other stressor-related errors, the method has advantages over more conventional approaches. By analysing social media data, NLP can be a potent tool for clinical analytics and healthcare informatics, giving significant insight into patients' sentiments and emotions.

The objective of Casanova et al. (2021) was to create a fresh deep neural network sentiment analysis model. The model optimizes grid search-based hyperparameters and integrates long short-term memory (LSTM) with convolutional neural networks (CNNs). The study's findings show that in terms of accuracy for sentiment analysis, the suggested LSTM-CNN-grid search-based neural network model performs better than the baseline models. A study (Cascone et al 2020) employs a neural network model to integrate topic semantics into a representation of text. Adding a system for attention, which uses a context-aware vector to determine each word's weight, allows for integration. The experimental findings in the research show how well the suggested model works to increase the validity of sentiment assessment results.

Terhörst et al. (2019) suggested a theoretical framework for mining messages to extract fine-grained information on disasters, such as affected parties, damaged infrastructure, and disrupted services. First, they employ LSTM networks to categorize tweets about disasters into two categories to obtain higher accuracy by preserving long-term semantic dependencies.

3. Proposed Method

3.1. Dataset- MV (movie review)

A total of 10662 favorable and negative movie reviews are included in this dataset, with each review consisting of one sentence. Predicting sentences' positive and negative feelings is difficult.

3.2. Text Embedding
We are aware that neural networks cannot be fed text directly. First, we must translate the text’s words into numbers. Therefore, using the acquired representation of the expression vectors as input, we pretrain an unsupervised embedding vector from GloVe2 to produce the word embedding.

If the vocabulary count of the dataset is \( k \) and the input sentence \( Y \) has \( o \) words with an average word embedding dimension \( ofn \), the matrix of embedding \( B \) will have a dimension space of \( J^*k \). As a result, here is how the sentence is represented in the input:

\[
Y(s_1, s_2, s_3, \ldots, s_n), Y \in J^*o
\]  

\( J^* \) is the length of the space of every word in the lexicon.

3.3. Bi-GRU

The recurrent gated unit (GRU) was first proposed as a standard recurrent neural network (RNN) variation. With inputs \( Y_s \) and \( L_s \), GRU computes \( G_s \) for each point \( t \) as follows:

\[
j_s = \sigma(G_f y_s + R_f l_s - 1) \hspace{1cm} (2)
\]

\[
R_s = \sigma(G_r y_s + R_r l_s - 1) \hspace{1cm} (3)
\]

where \( l_s, j_s \) and \( ut \) stand for the reset gate, update gate, and d-dimensional hidden state, respectively. The GRU has three parameters: \( G_s, G_r, G_f, b_r \) and \( R_s, R_r, R \). The sigmoid function, or elementwise production, is represented by the symbol. We represent the preceding context for a word at \( s \) using the hidden state from the forward GRU \( h_s \) and the following context for that word at \( s \) using the hidden state from the backwards GRU, \( l_s \), which reverses the text encoding. The bidirectional contextual encoding of \( y_s \) is concatenated into the string \( l_s = [l_s; l_s] \) and used as the Bi-GRU layer’s output at time \( s \).

3.4. LSTM

Long short-term memory (LSTM) is an architectural type of recurrent neural network (RNN). It was created to solve the issue of vanishing gradients, which prevents typical RNNs from capturing long-term dependencies in sequential data. In this scenario, the angles gradually disappear or increase exponentially over time. The input gate decides the amount of received data that should be stored in each memory cell. Each memory cell component obtains a value between 0 and 1 after being fed the current input and the prior concealed state through a sigmoid activation function. The forget gate determines information removed from the memory cell. It uses the prior hidden state and the current input as inputs, applies a sigmoid activation function, and outputs a value between 0 and 1 for each memory cell component using the current input and the prior hidden state as inputs and a sigmoid activation function. This gate aids in the selective output from the memory cell of pertinent information.

\[
l_s = \sigma(W_i x_s + U_i h_s - b_i) \hspace{1cm} (4)
\]

\[
f_s = \sigma(W_f x_s + U_f h_s - b_f) \hspace{1cm} (5)
\]

\[
o_s = \sigma(W_o x_s + U_o h_s - b_o) \hspace{1cm} (6)
\]

3.5. Self-attention

In neural networks, particularly in models such as transformers, a process known as self-attention, also known as scaled dot-product attention, is employed to record significant links between various parts in a sequence. Self-attention enables the concurrent processing of all sequence components, in contrast to recurrent neural networks (RNNs), which only process rows sequentially. Each sequence element is split into query, key, and value vectors during self-attention. These vectors are constructed from the input sequence to calculate attention scores between various items. The attention scores determine the relevance or importance of each component with others. The dot-product operation, scaling, and softmax normalization are used to calculate attention scores.

After being received, the attention scores are utilized to weigh the appropriate value vectors. The contextual information for each element in the sequence is then captured by adding the weighted values to create a weighted representation. This enables the model to suppress useless or noisy input and prioritize relevant elements. Compared to conventional recurrent architectures such as GRU (gated recurrent unit) and LSTM (long short-term memory), self-attention has some advantages. Because attention ratings can give larger weights to significant parts regardless of where they are in
the sequence, it makes it possible to capture long-range relationships more effectively. Self-attention also enables parallel computation, making it quite effective for handling lengthy arrangements.

4. Result

The effectiveness of the suggested and current methods is assessed in this section. The parameters are recall, accuracy, precision, and MAE. FSS-GCN (Rodgers et al 2021) and GCNN-LSTM (Fenu et al 2021) are the current processes.

Accuracy is a metric for how well a model forecasts the results or labels of a specific dataset. It is computed by dividing the total number of forecasts made by the number of accurate predictions, commonly expressed as a percentage. The accuracy value is greater than that of our proposed method. Figure 1 depicts the accuracy outcome. By comparison, it shows that the value of our proposed method (proposed method 96%) is superior to existing methods (FSS-GCN 83%, GCNN-LSTM 90%).

![Figure 1 Outcome of accuracy.](image1)

The accuracy of categorization or prediction models is evaluated using a statistical metric called precision. Out of all expected positive events (true positives plus false positives), it determines the proportion of events that were accurately anticipated to be positive (true positives). Figure 2 depicts the precision outcome. By comparison, it shows that our proposed method (proposed method 95%) is better than existing methods (FSS-GCN 81%, GCNN-LSTM 89%).

![Figure 2 Outcome of precision.](image2)

Recall, also known as sensitivity or true positive rate, expresses how many positive occurrences the model correctly accepted. The proportion of real positives to all real and erroneous negatives is calculated. The recall value is larger than that of our proposed method. Figure 3 depicts the recall outcome. By comparison, it shows that our proposed method (proposed method 93%) is greater than existing methods (FSS-GCN 85%, GCNN-LSTM 88%).

![Figure 3 Outcome of recall.](image3)
The acronym MAE stands for mean absolute error. It is a typical metric to assess a collection’s average error size or deviations between expected and observed values. MAE measures how closely the predictions match the actual data without considering the error’s direction. MAE is less than other values. Figure 4 depicts the MAE outcome. By comparison, it shows that our proposed method (proposed method 0.6) is improved over existing methods (FSS-GCN 1.8, GCNN-LSTM 1.3).

5. Conclusion

This paper proposes a deep learning fusion method approach for evaluating emotional semantics in natural language. The technique gathers local and global contextual data using long short-term memory (LSTM) networks, self-attention mechanisms, and bidirectional gated recurrent units (GRUs), allowing for a more thorough comprehension of the connections between words and emotions. To improve word representation, trained GloVe embeddings of words are also used. Benchmark movie review (MR) for sentiment evaluation and emotional task classification is used to assess the proposed technique. According to the experimental results, it performs better than single machine learning models and conventional feature-based methods. The fusion method’s skillful capture of the intricacies connected with emotional meaning in natural language allows for a better assessment of sentiment and attitude. This study demonstrates the possible use of deep learning fusion techniques for the textual representation of intricate emotional interactions. It improves sentiment evaluation and emotion classification tasks by proposing a more sophisticated and precise method for interpreting and deciphering emotion semantics in natural language. Large-scale, high-quality datasets are essential for training deep learning models. Emotional semantic assessment necessitates labelled data with precise emotional annotations, but these may be few or challenging to acquire. Emotional semantic evaluation can be improved by incorporating other modalities, including text, audio, and visual data. The development of deep learning fusion techniques that successfully integrate several modalities to increase the precision and robustness of emotional analysis can be the subject of future research.
Ethical considerations

Not applicable.

Declaration of interest

The authors declare no conflicts of interest.

Funding

This research did not receive any financial support.

References


Svetakov M, Hodashinsky I, Slezkin A (2021) Gender, age and number of participants effects on identification ability of EEG-based shallow classifiers. In 2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), pp. 0350-0353. IEEE.