International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Segmentation of Neural Text and Its Application to Sentiment Analysis

R. Hendra Kumar[*1], C. Kusuma Latha[*2]
[*1] Associate Professor, [*2] M.tech Student
[1,2] KMM Institute of Technology and Sciences, Andhra Pradesh, India

Article Info
Publication Issue :
Volume 8, Issue 6
November-December-2022
Page Number : 125-130
Article History
Accepted: 05 Nov 2022
Published: 15 Nov 2022

ABSTRACT

Poetry and formal texts have gotten less attention in recent years from experts in artificial intelligence than informal textual content such as SMS, email, chat, and online user reviews. Using Deep Learning, this study proposes a text-based emotional state categorization system. The text corpus is used to construct an attention-based Bi-LSTM model and CNN, further which are compared for their accuracy. There are a number of distinct emotional states that which can be classified from the text using the suggested method. These states include neutral, joy, fear, sadness.

Keywords: Deep learning, Emotion recognition, Bi-LSTM, CNN, Formal text, Emotional states.

I. INTRODUCTION

There are a huge number of comment texts on the Web due to the fast growth of Internet technology and social networks in recent years. When it comes to analysing the emotional tendencies of remarks, artificial intelligence technologies [1] can aid. As a component of artificial intelligence, sentiment analysis is a highly useful tool for determining the sentiment trend of comments. Words have varied contributions to categorization in sentiment analysis. Experts in disciplines such as natural language processing, computational linguistics, and artificial intelligence have been interested in the categorization of views, feelings, and emotional states. It is possible for a machine to evaluate formal and casual writing [2]. While the formal textual content consists of poetry, novels, essays, novels, plays, and official/legal documentation, the informal textual content consists of SMS messages, chat messages, and social media posts [1,3], among other things.

In text categorization (also known as text classification), free-text texts are assigned predetermined categories. He or she can give conceptual perspectives of document collections, and he or she may be used in the actual world. Reports in health care organisations, for example, are commonly categorised using taxonomies of illness categories, surgical procedures, insurance reimbursement codes, etc.

The use of machine learning algorithms to extract and analyse emotional states and themes from poetry writing has become increasingly popular [4] in recent years. But such research are restricted by tiny datasets and a limited range of emotional states. Traditional machine learning approaches have been employed to

detect emotional states from poetry text in previous research, which used restricted datasets that had been tagged with a small number of emotion classes. To categorise poetry material into two emotional categories, one research [3] employed a Support Vector Machine (SVM) and a BiLSTM classifier.

Hence BiLSTM model is used which can help close this gap. We also utilised deep learning's Attention mechanism. To improve the accuracy of the categorization of emotional states from text, we also utilise five emotion classes, which is an expansion of baseline work.

In this framework, section 1 discuss about Introduction. Related work will be discussed in section 2, section 3 describes about our proposed, where section 4 provides results of our work and we will conclude the paper in section 5.

II. RELATED WORKS

A lot of academics have been working on "Emotion recognition" using machine learning approaches in recent years. Using multiple emotion categories to categorise poetry material, [1] created a technique for recognising emotional states from poetry text. The Nave Bayes machine learning classifier is used for this purpose. Despite the fact that some poetry are misclassified, the results are promising.

To analyse tweets with respect to different linguistic dimensions like sarcasm, metaphor, and irony, [5] applied multiple the ML classifiers like Decision Tree and Naïve Bayes. The 11-point scale was used to perform the comparison and up to the mark, results are achieved. Similar to work [1], in which they face an issue of limited dataset size, where [5] also used a limited dataset. Both studies aimed to handle this limitation in the future by the extension of the dataset. SVM and Nave Bayes algorithms were used [6] to categorise poetry according to their emotion category. The results of a tiny dataset were encouraging. Incorporating phonemic elements might enhance the system's versatility. It is possible to determine a person's sentiments about music based on their moods and topics.

An audio input system was proposed [7]. A support vector machine classifier is used to achieve the best results in this regard. In spite of this, mood classification can be refined. The automated analysis of poetry in Arabic, Chinese, German, Malay, Persian and other languages has been done in the past. According to [8] Arabic poetry is divided into several kinds such as Fakhr, Heja, Retha, and Ghazal. Machine learning techniques are used to classify the Arabic poetry according to the categories listed above. SVM and Ss-CNN algorithms are used [9] to do sentiment analysis at the sentence-level based on German novels, and the results are impressive. The SVM classifier outperformed the S-CNN classifier in terms of accuracy. A supervised ML method was suggested in [10]. In experiments, a variety of classifiers were employed. A collection of emotions, on the other hand, can be improved in the future in order to achieve more hopeful outcomes.

Author & Year	Proposed	Finding/Outcomes
P. S. Sreeja and G. S. Mahalakshmi	Emotion recognition from poems by maximum posterior probability	Recognition of emotions from poem texts using the probability
A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes, 2015	Sentiment analysis of figurative language in Twitter	Analysis of sentiment from the twitter data
G. Rakshit, A. Ghosh, P. Bhattacharyya, and G. Haffari, 2015	Automated analysis of Bangla poetry for classification and poet	Classification and analysis of Bangla poetry

	identification	
K. Bischoff, C. S. Firan, R. Paiu, W. Nejdl, C. Laurier, and M. Sordo, 2009	Music mood and theme classification-a hybrid approach	Classification of theme and mood of music
O. Alsharif, D. Alshamaa, and N. Ghneim, 2013	Emotion classification in Arabic poetry using machine learning	Using machine learning for the classification of emotions from Arabic poetry

III. Methodology

The procedure to develop our system is clearly described in this section.

- For our suggested system, we'll use text data to classify emotions.
- Vector space dimensions and class numbers are determined by analysing the data.
- Before pre-processing, the data is divided into training and testing.
- A word tokenizer called NLTK is used for the tokenization of data when the pre-processing is complete.
- As a result, the unique tokens are tallied, padded, and the label length is transformed to integers and classified.
- Pretrained word2vec from file are imported and embedding matrix is created, which will be mapped later in our corpus to existing word vector.
- Word vectors are used to create the embedding layer, and the model is trained using Bi-LSTM and CNN.
- After training, the model is utilised for evaluation, and a confusion matrix of anticipated emotions is shown on the model's plotter.

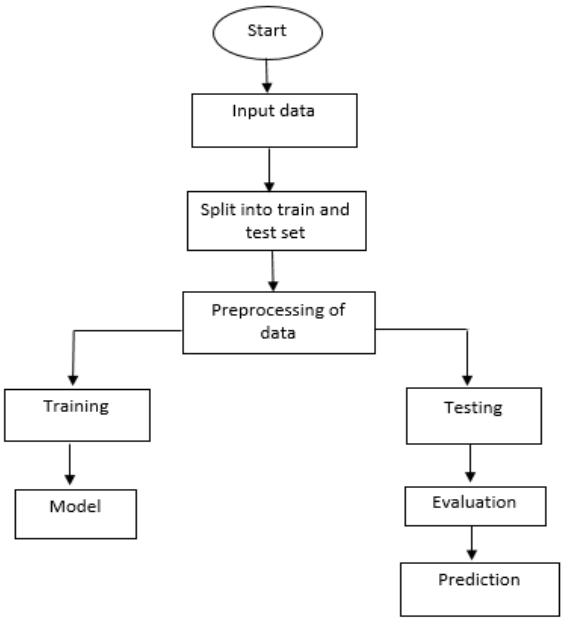

Fig 1. Proposed method block diagram

Long Short-Term Memory (LSTM)

LSTM is an artificial recurrent neural network (RNN) architecture in deep learning. With the LSTM, there are feedback connections, as opposed to the typical feedforward neural network. In addition to processing single data points (such as pictures), it is able to process large datasets (such as speech or video). When it comes to tasks like handwriting identification, speech recognition, and anomaly detection in network traffic (or IDSs), LSTM can be used (intrusion detection systems).

Cells, input gates, output gates, and forget gates make up a typical LSTM unit. There are three gates in the cell that govern information flow in-and-out of it, and the cell may store data for any period of time.

These networks are ideal for categorising, processing, and generating predictions based on time series data, because there might be unpredictable delays between key occurrences in the time series. When training typical RNNs, the vanishing gradient problem might

occur. LSTMs were created to solve this problem. There are a number of situations where the relative insensitivity of the LSTM to gap length is an advantage over RNNs, hidden Markov models, and other sequence learning techniques

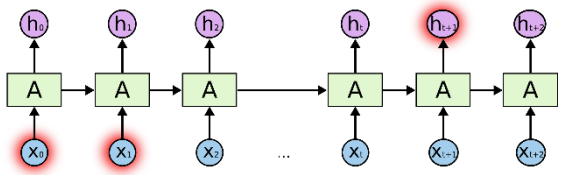

Fig 2. LSTM network diagram

Recurrent Neural Network (RNN)

An RNN is a type of artificial neural network in which the connections between nodes form a directed graph along a time sequence. This gives it the ability to display temporal dynamic behaviour, which is a great benefit. With the help of their internal state (memory), RNN's can handle varying length sequences of data. Because of this, they may be used in tasks such as handwriting recognition or speech recognition.

For all intents and purposes, the phrase "recurrent neural network" is used interchangeably with the terms "finite impulse" and "infinite impulse." Temporal dynamic behaviour may be seen in both kinds of networks. Unlike an infinite impulse recurrent network, which cannot be unrolled, a finite impulse recurrent network is a directed cyclic graph that can be unrolled and substituted with a strictly feedforward neural network.

Infinite impulse recurrent networks and finite impulse recurrent networks may both store extra states, and the storage can be under direct control by the neural network. In some cases, the storage can be replaced by a different network or graph, if it includes time delays or feedback loops. This type of regulated state is referred to as gated state or gated memory, and is seen in LSTMs and gated recurrent units. Feedback Neural Network (FNN).

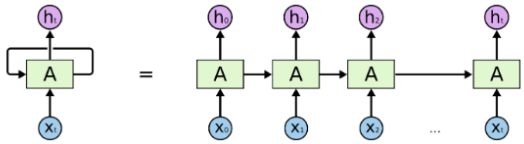

Fig3. Unrolled RNN diagram

Convolutional neural network

As a deep learning algorithm, the convolutional network is mostly utilised for viewing and analysing pictures. Basically, it's a neural network-based system. According to the weights shared by the kernels of convolution that are used to scan buried layers, they are also called differently. This network is widely utilised in various applications based on image and video recognition. These are used for segmentation, classification, and image analysis in a variety of different disciplines, including as medicine and engineering. Mathematics is mostly responsible for the naming of this convolution. Where this operation is used for the matrix multiplications in the layers of the network.

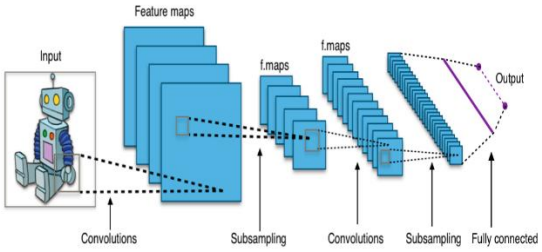

Fig 4. CNN Architecture

In the image above, you can see CNN's organisational structure. This neural network has input, hidden, and output layers. Activation is employed to hide input and output in neural networks, which is why the intermediate layers of feed forward are called hidden layers. Certain layers are also included in this hidden layer for convolution operations. In addition, these layers are used for operations such as multiplication and product operations. Commonly, ReLU is used as a function for the activation of an object. This is in addition to pooling, completely connected and normalising.

4. Results and Discussions

Here we'll discuss about the evaluations or predictions that were achieved by using the approach described above and how they were assessed.

After completing the training part, we plot accuracy and confusion matrix plots that which will be used to evaluate our models. The confusion matrix shows the classified emotional states.

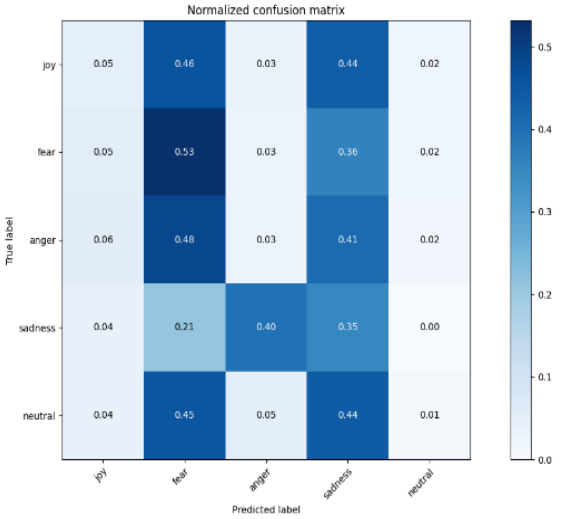

Fig 5. Confusion matrix plot using Bi-LSTM

The above figure shows the classified emotional states by using our proposed method.

IV. CONCLUSION

This study provides the classification of emotions from the text using the deep learning technique Bi-LSTM. That which classifies the 5 types of emotions from the text. Here Bi-LSTM and CNN are used for training the data. From the comparison of both the models, Bi-LSTM gives better accuracy.

V. REFERENCES

[1]. P. S. Sreeja and G. S. Mahalakshmi, "Emotion recognition from poems by maximum posterior probability," Int. J. Comput. Sci. Inf. Secur., vol. 14, pp. 36–43, 2016.

[2]. J. Kaur and J. R. Saini, "Punjabi poetry classification: The test of 10 machine learning algorithms," in Proc. 9th Int. Conf. Mach. Learn. Comput. (ICMLC), 2017, pp. 1–5.

[3]. G. Mohanty and P. Mishra, "Sad or glad? Corpus creation for Odia poetry with sentiment polarity information," in Proc. 19th Int. Conf. Comput. Linguistics Intell. Text Process. (CICLing), Hanoi, Vietnam, 2018.

[4]. Y. Hou and A. Frank, "Analysing sentiment in classical Chinese poetry," in Proc. 9th SIGHUM Workshop Lang. Technol. Cultural Heritage, Social Sci., Hum. (LaTeCH), 2015, pp. 15–24.

[5]. A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes, "SemEval-2015 task 11: Sentiment analysis of figurative language in Twitter," in Proc. 9th Int. Workshop Semantic Eval. (SemEval), 2015, pp. 470–478.

[6]. G. Rakshit, A. Ghosh, P. Bhattacharyya, and G. Haffari, "Automated analysis of Bangla poetry for classification and poet identification," in Proc. 12th Int. Conf. Natural Lang. Process., Dec. 2015, pp. 247–253.

[7]. K. Bischoff, C. S. Firan, R. Paiu, W. Nejdl, C. Laurier, and M. Sordo, "Music mood and theme classification-a hybrid approach," in Proc. ISMIR, Oct. 2009, pp. 657–662.

[8]. O. Alsharif, D. Alshamaa, and N. Ghneim, "Emotion classification in Arabic poetry using machine learning," Int. J. Comput. Appl., vol. 65, p. 16, May 2013.

[9]. A. Zehe, M. Becker, F. Jannidis, and A. Hotho, "towards sentiment analysis on German literature," in Proc. Joint German/Austrian Conf. Artif. Intell. Cham, Switzerland: Springer, 2017, pp. 387–394.

[10]. L. Barros, P. Rodriguez, and A. Ortigosa, "Automatic classification of literature pieces by emotion detection: A study on Quevedo's poetry," in Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact, Sep. 2013, pp. 141–146.

[11].S. Soumya, S. Saju, R. Rajan, and N. Sebastian, "Poetic meter classification using TMS320C6713 DSK," in Proc. Int. Conf. Signal Process. Commun. (ICSPC), Jul. 2017, pp. 23–27.

[12].A. Almuhareb, I. Alkharashi, L. A. Saud, and H. Altuwaijri, "Recognition of classical Arabic poems," in Proc. Workshop Comput. Linguistics Literature, 2013, pp. 9–16.

[13].A. Rahgozar and D. Inkpen, "Bilingual chronological classification of Hafez's poems," in Proc. 5th Workshop Comput. Linguistics for Literature, 2016, pp. 1–21.

[14].F. Can, E. Can, P. D. Sahin, and M. Kalpakli, "Automatic categorization of ottoman poems," Glottotheory, vol. 4, no. 2, pp. 40–57, Jan. 2013.

[15].C. Jareanpon, W. Kiatjindarat, T. Polhome, and K. Khongkraphan, "Automatic lyrics classification system using text mining technique," in Proc. Int. Workshop Adv. Image Technol. (IWAIT), Jan. 2018, pp. 1–4.

[16].Rang, "Poetry classification using support vector machines," J. Comput. Sci., vol. 8, no. 9, pp. 1441–1446, Sep. 2012.

[17].W. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," Expert Syst. Appl., vol. 41, no. 4, pp. 1742–1749, Mar. 2014.

[18].TensorFlow Text Classification Using Attention Mechanism. Accessed: Jan. 25, 2020. [Online]. Available: https://androidkt.com/tensorflow-text classification-attention-mechanism/

[19].Keras Documentation: Embedding. Accessed: Jan. 2, 2020. [Online]. Available: https://keras.io/layers/embeddings/

[20].J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional CNN-LSTM model," in Proc. 54th Annu. Meeting Assoc. Comput. Linguistics, 2016, pp. 225–230.

Cite this article as :

R. Hendra Kumar, C. Kusuma Latha, "Segmentation of Neural Text and Its Application to Sentiment Analysis", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 125-130, November-December 2022.
Journal URL : https://ijsrcseit.com/CSEIT228612