

CLASSIFICATION OF POETRY TEXT INTO EMOTIONAL STATE USING DEEP LEARNING TECHNIQUE

S.Anand, M. Greeshma, M. Yedu kondalu, G. Vinod Babu , M. Ratnakar

Associate Professor, Priyadarshini Institute of Technology & Science, AP, India.

Under Graduate, Priyadarshini Institute of Technology & Science, AP, India.

ABSTRACT

The classification of emotional states from poetry or formal text has received less attention by the experts of computational intelligence in recent times as compared to informal textual content like SMS, email, chat, and online user reviews. In this study, an emotional state classification system for poetry text is proposed using the latest and cutting edge technology of Artificial Intelligence, called Deep Learning. For this purpose, an attention-based C-Random forest model is implemented on the poetry corpus. The proposed approach classifies the text of poetry into different emotional states, like love, joy, hope, sadness, anger, etc. Different experiments are conducted to evaluate the efficiency of the proposed system as compared to other state-of-art methods as well as machine learning and deep learning methods. Experimental results depict that

the proposed model out performed the baselines studies with 88% accuracy. Furthermore, the analysis of the statistical experiment also validates the performance of the proposed approach.

1 INTRODUCTION

The classification of opinions, sentiments and emotional states has gained the attention of experts from different fields like natural language processing, computational linguistics and computational intelligence [1]. There are two types of writings that can be analyzed by machine: formal and informal [2]. The formal textual content pertains to poetry, novels, essays, novel, plays, and official/legal documentation, whereas the informal textual content is about SMS, chat, and social media posts. Due

to complex nature of the formal text (poetry), detection and classification of emotional states is a challenging task. For instance, the verse “And the sunlight clasps the earth, And the moonbeams kiss the sea:”, taken from the poem “Love Philosophy” (Shelley) conveys a love emotion. The manual strategy for detecting emotional states expressed by the poet in the poetry text is difficult and timeconsuming.

In recent times, machine learning techniques have been applied successfully for extracting and analyzing emotional states and themes from poetic text [1]–[3]. However, small datasets labeled with a limited number of emotional states are the major limitations of such studies [1], [3]. The existing studies on the detection of emotional states from poetry text have used traditional machine learning techniques [1], [4] with limited datasets tagged with a small number of emotion classes. One of the study [3] conducted on emotion classification from poetry text has used one machine learning classifier, namely Support Vector Machine (SVM) and a Random forest classifier, for classifying poetry text into two emotion classes. This gap can be bridged by investigating Attention-based C-Random forest model, which can take advantage of both the Convolutional Neural Network (CNN), Bidirectional Long Short Term Memory (Random forest), as well as, we also exploited the Attention mechanism of deep learning.

Therefore, produced and recorded in the type of text data have to be operated by computers. For that reason obtaining and using meaningful information from these data will facilitate the work of people. Text mining is multi discipline; it makes use of such as data mining, artificial intelligent, natural language processing and information retrieval. Text classification is one of the most widely studied areas in text mining. It is a supervised learning problem. To solve this problem, firstly, data separated into two parts called train and test data. In the next step, classification algorithms applied training data. In the last step classifier models predict label of text by means of learning model. There are many text-mining studies in the literature. Spam mail detection: Email is the simplest and cheapest communication tools in daily life. Most of the received emails are the text. These mails are two parts, named spam and nonspam. Some advertisement companies or bad people prefer to send spam mails to users. Spam mails give rise to waste of time and slow down daily work. Therefore,

spam mails have to be found and not delivered to the users. Spam mail detection is binary text classification application [1, 2].

Web page classification: Web pages occur in HyperText Markup Language (HTML) files. There are some information based on text in these files. If text-mining techniques apply to these files, web pages may classify by the computers [3].

Author detection: Text-mining methods apply to the books or column so that author of texts can be recognized [4, 5]. There are very different studies apart from the above popular topics. [6], automatic text summarization [7], automatic question answering systems [8], sentiment analysis on text [9, 10], music genre classification on lyrics [11, 12], television-rating prediction with social media [13] and automatic news article classification

Poetry is a form of literary art in which language is used for its aesthetic and evocative qualities in addition to, or in lieu of, its' apparent meaning. Classical poetry showed works of art with diverse styles from the history development of past literary, tradition and the nature of the data is unique and creative. Poetries are usually mean to deliver expression such as love, kindness and dignity. Thus, there are various categories of poetries. However, efforts in performing automatic classification of poetries are very rare. This is because the forms and features of poetry text are different from normal text as bound by a factor of lines, stanzas, rhyme, elements of style and beauty of sound and rhythm. Furthermore, poetries are usually in the form of short textual paragraphs with little discriminative value word features for automatic classification purposes. Therefore, the classification of poetries proved to be a challenging task. Classification of poetry is important particularly in information retrieval (or poetry retrieval) as its retrieval is not according to a simple keyword matching but involves the context, classes and themes of the poetries. Apart from that, the ability to recognize poetry from prose is potentially important, particularly for search machines with particular emphasis for poetry mining.

2 RELATED WORK

Pantun or the Malay poetries are categorized by audience, shape and theme (Bakar, 1983). Audience means the division of the poetry from the perspective of both parties,

who recited and heard the poetry according to their age. Therefore, poetries are appropriate to be ‘heard by’ or

‘recited for’ children, adults or the elderly. The shape refers to the division of poetry from the point of length lines structure either consisting of two, four, six and up to sixteen lines. While the theme refers to the categorization of poetry based on philosophical concept, experience, emotion, interpretation and human understanding. For example, poetry that expresses love to God is classified under the ‘religion’ theme. Malay poetry consists of short sentences in which each line of poetry has only four to seven words. The lines of Malay poetry are formed in pairs by alternating between each line of poetry. Typically, the numbers of lines in each verse of poetry are either two, four, six, eight or twelve lines.

- Pantun or the Malay poetries are categorized by audience, shape and theme (Bakar, 1983). Audience means the division of the poetry from the perspective of both parties, who recited and heard the poetry according to their age.

Therefore, poetries are appropriate to be ‘heard by’ or

- ‘recited for’ children, adults or the elderly. The shape refers to the division of poetry from the point of length lines structure either consisting of two, four, six and up to sixteen lines. While the theme refers to the categorization of poetry based on philosophical concept, experience, emotion, interpretation and human understanding. For example, poetry that expresses love to God is classified under the ‘religion’ theme. Malay poetry consists of short sentences in which each line of poetry has only four to seven words. The lines of Malay poetry are formed in pairs by alternating between each line of poetry. Typically, the numbers of lines in each verse of poetry are either two, four, six, eight or twelve lines.
- Computer graphics
- Basic game development
- Mapping and geography (GIS software)

3 METHODOLOGY

1) DATA ACQUISITION

To train a deep learning model one of the most important steps is to collect the data.

For this purpose, we have used the dataset acquired from which consists of 9142 posts.

2) PREPROCESSING

To implement the deep learning model, the next step is to transform the words into numbers. So some of the basic preprocessing steps such as stop-words removal, conversion to lowercase, and tokenization, are performed. After tokenization, a vocabulary is built which transforms the sequences of words into the sequences of integers, where each integer represents a specific word in a vocabulary.

3) FEATURE REPRESENTATION

To enable the model to learn, each word is further transformed into an embedding vector by using the Keras embedding layer.

4) FEATURE EXTRACTION

In this module, the proposed model extracts the n-gram features from the input received from the previous embedding layer.

4. RESULT

LOADING THE LIBRARIES

```

C:\Windows\System32\cmd.exe - python code.py
Microsoft Windows [Version 10.0.22631.3447]
(c) Microsoft Corporation. All rights reserved.

C:\Users\neham\OneDrive\Desktop\New folder>python code.py
Number of instances: 7480
| Classifier | Training Accuracy | Test Accuracy |
|-----|-----|-----|

```

ALGORITHM EVALUTION

```

C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.22631.3447]
(c) Microsoft Corporation. All rights reserved.

C:\Users\neham\OneDrive\Desktop\New folder>python code.py
Number of instances: 7480
| Classifier | Training Accuracy | Test Accuracy |
|-----|-----|-----|
| SVC | 0.9867513 | 0.4512832 |
C:\Users\neham\AppData\Local\Programs\Python\Python312\Lib\site-packages\sklearn\svm\_classes.py:31: FutureWarning: The default value of 'dual' will change from 'True' to '
'auto' in 1.5. Set the value of 'dual' explicitly to suppress the warning.
  warnings.warn(
C:\Users\neham\AppData\Local\Programs\Python\Python312\Lib\site-packages\sklearn\svm\_base.py:1237: ConvergenceWarning: Liblinear failed to converge, increase the number of
iterations.
  warnings.warn(
| LinearSVC | 0.9988392 | 0.5768717 |
| RandomForestClassifier | 0.9988392 | 0.5541444 |
| DecisionTreeClassifier | 0.9988392 | 0.4578877 |
Joy (1. 0. 0. 0. 0. 0.) 1084
anger (0. 0. 1. 0. 0. 0.) 1080
sadness (0. 0. 0. 1. 0. 0.) 1079
fear (0. 1. 0. 0. 0. 0.) 1078
disgust (0. 0. 0. 0. 1. 0.) 1057
guilt (0. 0. 0. 0. 0. 1.) 1057
shame (0. 0. 0. 0. 0. 1.) 1045
This looks so impressive GUILT
I have a fear of dogs FEAR
My dog died yesterday SADNESS
I love you anymore..! JOY

C:\Users\neham\OneDrive\Desktop\New folder>
C:\Users\neham\OneDrive\Desktop\New folder>

```

5.CONCLUSION

To categorize English poetry text within multiple emotion classes, we have exploited a deep learning technique namely, Attention-based C-Random forest model. For

experimentation, a benchmark dataset is used with an extension into the emotion classes: Alone, Hope, Nature, and Surprise along with their respective poems.

6.REFERENCE

- [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, “Analyzing 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," *Glottology*, 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/tensorflowtextclassificationattentionmechanism/>
- [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.
- [21] Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.