



GUJARATI POETRY CLASSIFICATION BASED ON EMOTIONS USING DEEP LEARNING

Bhavin Mehta
Assistant Professor,
Gujarat Technological University,
Ahmadabad, Gujarat, India

Bhargav Rajyagor
Assistant Professor,
Gujarat Technological University,
Ahmadabad, Gujarat, India

Abstract - Poetries are the way to express the word that represents emotions and thoughts with the use of any language in the world and the computational linguistic study to the emotion recognition from poetry is an overwhelming and complex task too. Ultimate goal of this study is to disclose emotions through Gujarati poetries with the use of variety of characteristic there within Gujarati poems. Study presents a novel perspective in sentiment capture as of Gujarati Poems. 'Kavan' Gujarati poems collection was by hand interpreted upon the Indian idea mentioned in 'Navarasa'. In the collection, 300+ poems were categorized into nine feeling mentioned in 'Navarasa', respectively 'વીરસ', 'હાસ્યસ', 'ક્રુણસ', 'શૃંગારસ', 'રોદ્રસ', 'ભયાનાક્રસ', 'બીભસ્તસ', 'અદ્ભુતસ', and 'શાંતસ'. The Zipfian allotment is part a family of related discrete power law probability distributions, which is used to identify the probability of one 'rasa' from the given poem.. Deep Learning has colossal set of pattern recognition tools that can be used for natural language processing and have incredible potential to locate a solution for challenging machine learning problem. Result accuracy was found up to ~87.62% of emotion classification task from Gujarati poetry corpus.

Keywords: Gujarati Poetry; Navarasa; Emotion recognition; Deep Learning; NLP

I. INTRODUCTION

Gujarati speech is not just admired in the globe however as well is employed as formal speech in Gujarat, in Gujarati feelings and Literature is greatly connected with each other. Feelings found in Gujarati literature are greatly associated with poetry. Poetries are expressed with full of emotions as main basis. Poetry awakens abundant emotional states in reader. In poetry, poet can communicate through their emotions like, happiness, sadness, frustration, pain, love, and with faith. Poet places their emotions in poetry through imagery power, suitable selection words, rhyme, and rhythm. Emotions must be mined and poetry must be classified, from the invoked emotional states, for ease of recovery to the reader.

Our proposed system accepts poetry as input and returns emotion invoked from that poetry as the result with the help of deep learning concepts, which is purely based on the Indian idea of 'Navarasa' mentioned in the 'Natyashastra'[1]. Our Concept of 'Navarasa' is similar to

that mentioned in 'Rasa theory' by sage Bharat[2]. 'Rasa' i.e. emotions, that are generated by the mental state 'bhava'. Nine 'Rasas' are the main essence of any art and literature.

In this study author's main contributions is to effort towards collecting more than 300 different Gujarati poems and identify the human emotions using automated system through the concept of deep learning.

Gujarati is an Indian speech resident to the Gujarat, a state of India. Gujarati is broadly oral language in India, with sixth position and most broadly oral speech in the world, with twenty sixth position[3].

From Zaverchand Meghani to Sairam Dave, there have been many poets who have contributed to Gujarati language literature. To form an emotion detection system, Gujarati poetry corpus is required.

'Kavan' is the collection of 300+ Gujarati poetry and probably the first for Gujarati language. Poetry represents one of the basic nine emotions dictated in 'Navarasa' theory. Name of emotions can be used for training the system and locate the emotion class of test poetry. Proposed emotion detection system uses Zipf's law and presents number of occurrences of word statistically in poetry that matches exactly with one 'rasa'.

II. LITERATURE REVIEW

An extraordinary job for feeling detection from poetry in foreign speech and Indian speech noticed, but no such task seen for specifically Gujarati speech. From literature review found out that: most feeling search is basis primarily on non poetic literature or verse form. Jasleen and Saini [1] have used SVM and NB for emotion detection from Punjabi poetry with 72.02% accuracy by SVM. They have mentioned 'Kavi' Punjabi poetry collection, manually annotated based on Indian concept of 'Navarasa'. They also suggested working on the improvement of the miscategorized poetry.

Bafna and Saini[4] worked for token extraction and comparison by two methods and on two corpora. They have used 'BaSa' and Zipf's law as method and achieved token comparison between both the methods.

Tiple and Thomas[5] carried out an mood and emotion detection in North Indian Classical Music. Extracting emotions by the feature like Raga, Rasa, Taal, and can be mapped to mood. They have suggested method for mapping musical feature to emotions using classifier like Support Vector Regression, Support Vector Machine and



Gaussian Mixture Model with accuracy above 90%; can be applied to Indian Classical Music and Navarasa Model.

Bawa and Josan[6] have introduced a technique for automatically classifying varied types of poem according to mood. They come with three types of mood namely happy/joy, sad, and anticipation. Mood of poem from identical dataset is classified using Maximum Entropy and Naive Bayes Classifier with result 67.85% and 73.3% respectively. Ahmad, Asghar, Alotaibi and Khan[7] developed an emotional state classification system for poetry text and contain use an interest basis model is implemented on their own poems collection, resulted with 88% accuracy. Bafna and Saini [8] presents poem categorization using suitable algorithm, on their poem collection and tried to identify category like “Bal Geet”, “Updesh Geet” or “Bhajan”, resulted with 83.33% accuracy. Audichya and Saini[9] with the usual speech procedure with “Alankaar” in Hindi poems, carried out work on “Alankaar”, and have build a organized hierarchical arrangement of Hindi “Alankaar” forms & sub forms, then work to recognize a small number of forms. They attained 97% effectiveness in coverage the initial outcome with an typical performance time of 0.002S.

Haider, Eger, Kim, Klinger and Menninghaus[10] presented a system to identify feelings from their poem collection, have used pre-trained BERT model for processing. Jasleen and Saini[11] presented Punjabi Poems Classification using ten best Machine Learning algorithms, have found outcome for Punjabi Poems categorization that 4 algorithms like, HP, NB, KNN and SVM, by correctness of different percentages. Audichya and Saini[12] developed a system, verse form recognition by systematic classification method of Hindi poems, to identify “Chhand” which covers 53 “Chhands” and have achieved 95.02% overall accuracy, and the overall failure rate was 4.98%.

Ahmed, Ali, Hasan, and Mohammed[13] have proposed: categorization of the current Arabic poems by ML for four forms: lover poetry, Islamic poetry, common poetry and biased poetry, study uses Naive Bayes, Support Vector Machine, and Linear Support Vector method for classification. They have achieved 64%, 17.75%, 72% precision result for NB, SNM, and LSV respectively.

Rajyagor and Rakholia[14] derived Handwritten Character Recognition Using Deep Learning, which uses Deep Learning (DP) as one of the method for recognition of handwritten Gujarati text from image. Rajyagor and Rakholia[15] implemented a system for Tri-level handwritten text segmentation techniques for Gujarati language, used 500+ handwritten sentences written by different people, have obtained the accuracy for character-level segmentation is 82%, word-level is 90% and for the line-level segmentation is 87%.

III. DATASET

Authors have created their own dataset named ‘Kavan’. This dataset has 300+ poems written by well known Gujarati literature poets. In this ‘Kavan’ Gujarati

poetry corpus authors have collect poets who wrote poetry using different ‘Rasa’s’.

‘Kavan’ name is used for Gujarati poetry corpus as it is another name of ‘Kavita’ in Gujarati literature. ‘Kavan’ is to express emotions in poetries in Gujarati literature, and most suitable name for the Gujarati poetry corpus.

કવન

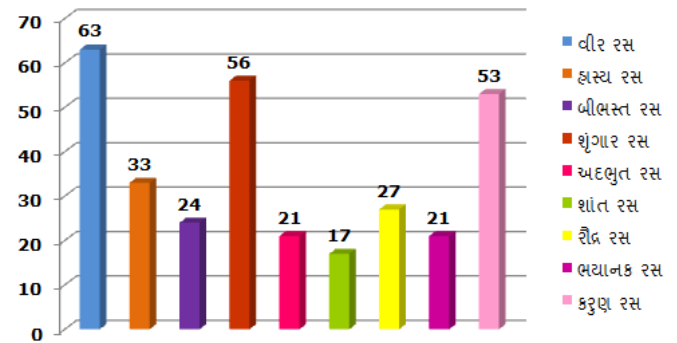


Figure 1 'Kavan' Dataset

IV. METHODOLOGY

In our proposed methodology, our first task is data collection and creation of Gujarati poetry corpus, from the poetry of different poets of Gujarati literature. We have completed this as mentioned in dataset. Zipf's law is employed to identify some identical tokens located in Gujarati poetry corpus.

NLTK in “Python” programming language is used for text processing applied to achieve NLP operations like

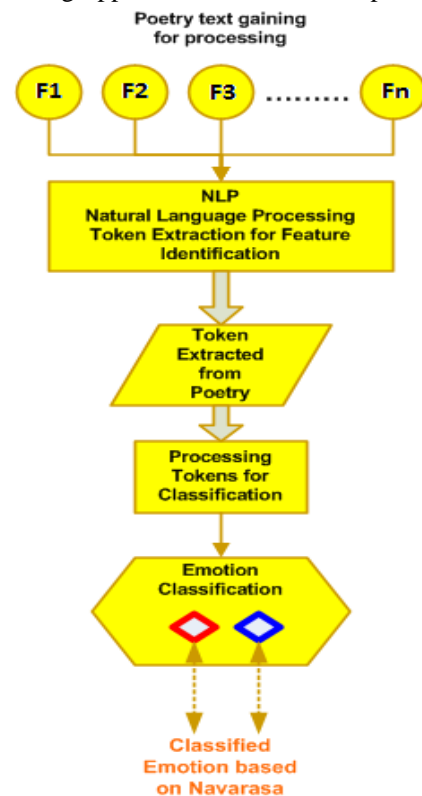


Figure 2 Emotion Classification Workflow

tokenization, labeling and so on.

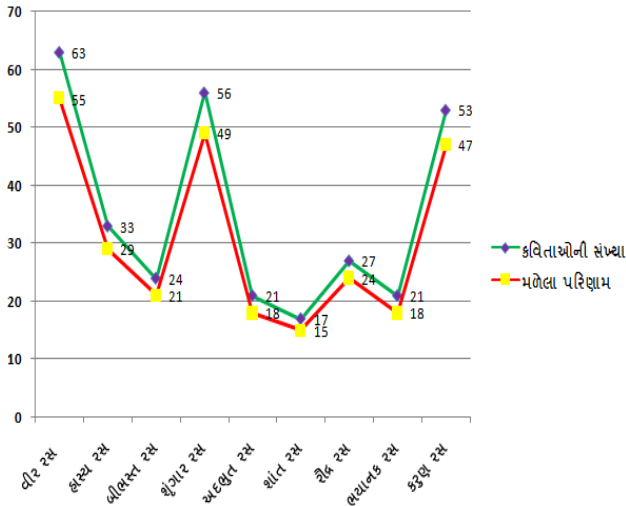


Figure 5 Derived Results

The method automatically takes one poem after another as input. It will then be processed to compare words in the poem with the words in the metadata to determine the ‘Rasa’ of the poem.

After processing and determining the ‘Rasa’ of the poetry on the basis of the Indian concept of ‘Navarasa’, output will be displayed as follows:

```

4:Kavan/Bhanushankar_Vyas/bhanushankar_1.txt
મળેલ કાવ્યનો રસ: હાસ્ય રસ
5:Kavan/Zaverchand_Meghani/Meghani_1.txt
મળેલ કાવ્યનો રસ: વીર રસ
6:Kavan/Zaverchand_Meghani/Meghani_2.txt
મળેલ કાવ્યનો રસ: વીર રસ
7:Kavan/Zaverchand_Meghani/Meghani_3.txt
મળેલ કાવ્યનો રસ: વીર રસ
8:Kavan/Zaverchand_Meghani/Meghani_4.txt
મળેલ કાવ્યનો રસ: વીર રસ
9:Kavan/Zaverchand_Meghani/Meghani_5.txt
મળેલ કાવ્યનો રસ: વીર રસ
10:Kavan/Zaverchand_Meghani/Meghani_6.txt
મળેલ કાવ્યનો રસ: બીભસ્ત રસ
11:Kavan/Zaverchand_Meghani/Meghani_7.txt
મળેલ કાવ્યનો રસ: વીર રસ
12:Kavan/Zaverchand_Meghani/Meghani_8.txt
મળેલ કાવ્યનો રસ: કરુણ રસ
13:Kavan/Befam/Befam_1.txt
મળેલ કાવ્યનો રસ: શૃંગાર રસ
14:Kavan/Befam/Befam_2.txt
મળેલ કાવ્યનો રસ: કરુણ રસ
15:Kavan/Befam/Befam_3.txt
મળેલ કાવ્યનો રસ: અદ્ભુત રસ
16:Kavan/Befam/Befam_4.txt
મળેલ કાવ્યનો રસ: શૃંગાર રસ
    
```

Figure 6 Identified Rasa from Poetry

VI. CONCLUSION

By the review of literature, the Authors have found that there was limited research study for classification of Gujarati poetry on the basis of the concept of ‘Navarasa’.

The emotions contained in the poem have strong connection with the poem; poets have penned the poetries with lots of emotions. In this study authors have experimented to achieve an automatic system for classification of poetries on the basis of emotions present in poem with the Indian concept of ‘Navarasa’. A collection of 300+ poems called ‘Kavan’ Gujarati poetry corpus has been created by the authors themselves.

On the Basis of result achieved, it is concluded that our system performed remarkably for the emotion detection and classification from Gujarati poetry.

VII. REFERENCES

- [1] J. R. Saini and J. Kaur, “Kāvi: An Annotated Corpus of Punjabi Poetry with Emotion Detection Based on ‘Navrasa,’” *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 1220–1229, 2020, doi: 10.1016/j.procs.2020.03.436.
- [2] Wikipedia, “[https://en.wikipedia.org/wiki/Rasa_\(aesthetics\)](https://en.wikipedia.org/wiki/Rasa_(aesthetics)),” 2021. .
- [3] W. Contributors, “Gujarati language. In Wikipedia.” https://en.wikipedia.org/w/index.php?title=Gujarati_language&oldid=1027979135.
- [4] P. B. Bafna and J. R. Saini, “An application of Zipf’s law for prose and verse corpora neutrality for Hindi and Marathi languages,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 3, pp. 261–265, 2020, doi: 10.14569/ijacsa.2020.0110331.
- [5] B. Tiple and P. Thomas, “Analysis of Features for Mood Detection in North Indian Classical Music - A Literature Review,” *Ijrect*, vol. 6, no. 6, pp. 181–185, 2017.
- [6] G. Josan and S. Bawa, “Automatic Mood Classification of Punjabi Poems Using Supervised Approach,” no. August, 2018, [Online]. Available: <https://www.researchgate.net/publication/326995938%0AAutomatic>.
- [7] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and S. Khan, “Classification of Poetry Text into the Emotional States Using Deep Learning Technique,” *IEEE Access*, vol. 8, pp. 73865–73878, 2020, doi: 10.1109/ACCESS.2020.2987842.
- [8] P. Bafna and J. R. Saini, “Hindi Poetry Classification using Eager Supervised Machine Learning Algorithms,” *2020 Int. Conf. Emerg. Smart Comput. Informatics, ESCI 2020*, pp. 175–178, 2020, doi: 10.1109/ESCI48226.2020.9167632.



- [9] M. K. Audichya and J. R. Saini, "Towards Natural Language Processing with Figures of Speech in Hindi Poetry," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 3, pp. 128–133, 2021, doi: 10.14569/IJACSA.2021.0120316.
- [10] T. Haider, S. Eger, E. Kim, R. Klinger, and W. Menninghaus, "PO-EMO: Conceptualization, annotation, and modeling of aesthetic emotions in German and english poetry," *Lr. 2020 - 12th Int. Conf. Lang. Resour. Eval. Conf. Proc.*, no. 2001, pp. 1652–1663, 2020.
- [11] J. Kaur and J. R. Saini, "Punjabi poetry classification: The test of 10 machine learning algorithms," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1283, pp. 1–5, 2017, doi: 10.1145/3055635.3056589.
- [12] M. K. Audichya and J. R. Saini, "Stanza Type Identification using Systematization of Versification System of Hindi Poetry," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 1, pp. 142–153, 2021, doi: 10.14569/IJACSA.2021.0120117.
- [13] M. A. Ahmed, R. A. Hasan, A. H. Ali, and M. A. Mohammed, "The classification of the modern Arabic poetry using machine learning," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 17, no. 5, pp. 2667–2674, 2019, doi: 10.12928/TELKOMNIKA.v17i5.12646.
- [14] B. Rajyagor and R. Rakhliya, "Handwritten Character Recognition using Deep Learning," *Int. J. Recent Technol. Eng.*, vol. 8, no. 6, pp. 5815–5819, Mar. 2020, doi: 10.35940/ijrte.F8608.038620.
- [15] B. Rajyagor and R. Rakholia, "Tri-level handwritten text segmentation techniques for Gujarati language," *Indian J. Sci. Technol.*, vol. 14, no. 7, pp. 618–627, 2021, doi: 10.17485/ijst/v14i7.2146.