

SURVEY OF MARATHI POEM CLASSIFICATION USING MACHINE LEARNING

Prof. R. A. Deshmukh
Professor
Rajarshi Shahu College of
Engineering
Maharashtra, INDIA
Pune - 411033
radesh19@gmail.com

Namrata Chavan
Undergraduate Student
Rajarshi Shahu College of
Engineering
Maharashtra, INDIA
Pune - 411033
chavan.namrata97@gmail.com

Sayali Gole
Undergraduate Student
Rajarshi Shahu College of
Engineering
Maharashtra, INDIA
Pune - 411033
sayaligole3@gmail.com

Kumar Adarsh
Undergraduate Student
Rajarshi Shahu College of
Engineering
Maharashtra, INDIA
Pune - 411033
adarshaman1472@gmail.com

Suraj Kore
Undergraduate Student
Rajarshi Shahu College of
Engineering
Maharashtra, INDIA
Pune - 411033
surajkore47@gmail.com

Abstract

Now days there are various poems with different meaning, so we have generalize our idea for poet identification and poem classification. In this paper, we provide a survey and comparative analysis of existing techniques for poem classification like machine learning and lexicon-based approaches, together with evaluation metrics. Using various machine learning algorithms and NLP techniques, we will do research on Poem data set and classify according to it. We have also discussed general challenges in finding ambiguous word. The analysis of poem and its computation has various application and help user for understanding in this work, we have first identified the poem details from dataset then we have classified the poem according to its sentiments and then find the meaning of ambiguous word.

Keywords: *NLP, Machine Learning, sentiments, lexicon-based approaches.*

I. INTRODUCTION

Poetry is kind of creative expression that makes uses of aesthetics and rhythmic quality of language. The poems have many meaning to derive from its line and are often very specific towards something. Poems are having various kinds of sentiments to show like devotional, happy, etc. It is seventh most spoken language and has rich tradition. Our work is to identify the poet of poem and its sentiments along with words meaning which is ambiguous. As we are working with Marathi language we will see the work done by poets like kusumagraj, Guru Thakur, etc. in poetic rich tradition. In the poetry analysis basic thing is required for the analyzing the poetry by the words and lines. And classify that poem according to its sentiments based on words and lines. That contain the all the mood related terminologies in the poems. That we are required to identify based on the classification of the poem. By using the identification of the poet and their work we can perform the poem classification. It is done by using the various types of algorithms and the terminology. From the past the poems

and the songs are the very popular in the society. The every poem and the song have the deep meaning in it. We are developing the project for the identification of poem and the classify the type of the poem, we are motivated to this project on this basis. By classifying the poem and are able to understand the type of poem and poet. In this approach we classify the poem in the form of mood and the type.

II. RELATED WORK

Classification and identification of poems and their poets has been previously have been done in language like English, Bangla, Punjabi. In Bangla poetry svm classifier is used to find poem and its features like syntactic, orthographic features, etc [1].

The present model provides a way to classify the poem. The model proposed by Anupam ghosh , Geetanjali Rakshit have used subject based classification of poem in bangla language using its syntactic and semantic features and use of stylometric features for classification in its different attributes[1].

The model proposed by Navinder Kaur, Amandeep Verma [2] focusses on a Authorship Attribution in Punjabi language which is used to find author of collected data and find performance metrics based on data present in dataset.

The model proposed by Xia Li, Bin Wu [3] focusses on a finding unknown word using active distance, transform probability on song poetry and used Chi-square test.

III. LITERATURE REVIEW

The “Geetanjali Rakshit, Anupam Ghosh, Pushpak Bhattacharyya, GholamrezaHaffari” used Subject based classification using its syntactic and semantic features of Bangla poem and used stylometric features to find poems orthographic features, syntactic features, etc. [1]. We are going to refer which factors they have used for classification the poem and finding its orthographic, syntactic and lexical feature.

“Navinder Kaur, Amandeep Verma” performed authorship attribution to fin author from unseen text and used classification algorithm to find the author of testing data and analyzed accuracy, precision ,recall factors[2]. From this paper we have learned how to train the classifier model and how to evaluate its performance.

“Pandian*, V. V. Ramalingam and R. P. Vishnu Preet” used author work in past for classification of unknown Tamil poem and used Decision tree algorithm to find classifier accuracy [4], and we are going to implement it in better way for our system.

“Zhou GuoDong” used MIIM model and maximum entropy and chunking strategy to find unknown words [7], we are going to prefer the strategy used in it for checking its performance.

For unknown word detection “Xia Li, Bin Wu, Bailing Zhang” have used word embedding technique for classification and use of active distance [3]. Then results are stored in dictionary to analyze its meaning.

Jasleen Kaur, Jatinder Kumar R. Saini” uses concept of NLP, Machine learning, KNN for classification of Punjabi poem [5]. Results show that each classification algorithm has different accuracy; we will analyze same for our system and use best algorithms. Referred papers have been mentioned in Table 1.

SNo.	Title	Dataset	Algorithm	Performance	Remarks	Limitation
1.	Automated Analysis of Bangla Poetry for Classification and Poet Identification [1]	Bangla 1341 poem, 4 classes.	SVM classifier, Naïve Bayes.	Accuracy: 56.8%	With the use of stylistic feature the accuracy increase to 92.3% and lexical features is best for poet identification.	Result show Swadesh was often confused with Pooja word i.e. word disambiguation.
2.	Authorship Attribution of	Punjabi 856 poem, Poet	SVM Classifier,	Accuracy:79% For character	Poet wise training and	Failure of the classifier using

	Punjabi Poetry using SVM Classifier [2]	wise poem is distributed.	Tf-idf, weighting.	trigram features.	testing of poem and word ngram and char ngram features is used.	word based features.
3.	Unknown Word Detection in Song Poetry [3]	20000 Chinese song poetry.	SVM classifier, Chi square Test, Segmentation algorithm.	Accuracy of original tokenizer is 79.1% and while adding unknown word accuracy raise to 86.44%.	The unknown word is searched and added to dictionary and its meaning is found.	This work doesn't work for Ancient Chinese.
4.	Authorship Identification for Tamil Classical Poem (MukkoodarPallu) using Bayes Net Algorithm [4]	800 Tamil Poem with 25 features in 4 set.	Bayes Net algorithm, C4.5 decision tree algorithm, Weka tool.	Accuracy 94.1% using Bayes net Algorithm.	Author detection is done using known work of authors.	Tamil language are still anonymous for classification.
5.	Punjabi Poetry Classification: The Test of 10 Machine Learning Algorithms [5]	240 Punjabi Poem in 4 category i.e. NLP, LIPA, RORE, PHSP.	Hyper pipe, KNN, SVM Naïve Bayes.	Highest accuracy is achieved by using SVM i.e. 58.79%.	Linguistic Parameter is used to enhance the performance of poetry classification task.	Accuracy is low if decision tree and bagging is used.
6.	SVM based classification metrics for poetry style [6]	413 poems in 2 classes i.e. Bold and constrained positive, Graceful and restrained negative.	SVM, Naïve Bayes, Vector Space Model.	Accuracy: 88.6% using SVM.	Feature based accuracy using Feature number and Feature items for training and testing the poem.	4 pieces of song-ci not used in training corpus directly used in classifier.
7.	A Chunking Strategy towards Unknown Word Detection in Chinese Word Segmentation [7]	Two attributes PK and CTB used having training and testing data specified in PK for training 1100k words and 17k words for testing.	Maximum Matching algorithm.	Accuracy: 80% For unknown word.	Chunking strategy is used along with error driven approach.	More improvement needed in ambiguity resolution.
8.	A contrastive analysis of English and Bangla phonemics [8]	Vowels, Consonant phonemes of two languages are used.	Similarity and dissimilarity of two languages.	Spectral similarity as well as difference in phonemic system.	Clear picture of sound. Its differ feature lateral palatals and lateral approximation.	Actual difference can be better treated.
9.	Emotion Classification in	1231 Arabic Poem with 4	SVM, C5.0 Naïve	Accuracy: 49.15 using C5.0	It tested susceptibility	The choice of Arabic corpus

	Arabic Poetry using Machine Learning [9]	features Fakhi,Heza, Ghazal, Retra.	Bayes Hyper pipe.		of Arabic poetry to emotion classification.	is limited as it is written in standard Arabic not in numeric diabetes.
10.	A Meter Classification System for Spoken Persian Poetries[10]	136 poetries utterances with 12 Persian meters.	SVM classification algorithm.	Accuracy :91%	Automatic meter detection algorithm was evaluated and implemented.	Incorrect detection of meter type because of speaker wrong utteration.

Table 1 Literature Review

The work in Indian regional languages like Tamil, Punjabi, and Bangla provided us brief idea about poet identification, poem summarization and unknown word detection through various techniques and classification algorithm. We have referred various methods for author and poem detection like we can detect author using known work of author [4].

The chunking strategy and feature based extraction used in gave some briefing about unknown word detection and its analyzation according to poet point of view [6][7].

The emotion classification methods used in can be used in sentiment analysis according poet nature and their view as it is not done in many of regional language of India [9].

Further Automatic metric detection can be done using SVM classifier [10], and we will try to remove the incorrect detection in our future implementation of our idea.

IV. PROPOSED SYSTEM

In this proposed system we have first added the poem in database in the form of key value pair and then we have added search methodology in it to find poem details like poet, number of line in poem, etc., then we have classified the poem to find its sentiments using classification algorithm like Naïve Bayes or SVM classifier then we are going to find the meaning of ambiguous word.

Fig.1 below is the block diagram of the proposed system which shows us how actually the events take place for the poem in dataset.



Fig.1: Block Diagram of Proposed System

Objectives of the proposed system are as follows:

- This idea depicts poem identification, poem classification (love, nature, etc.).
- Finding the ambiguous word and implementing the recommendation system.
- This will help user to understand actual meaning of poem with its meaning.

The Fig.2 below displays how the system will work.

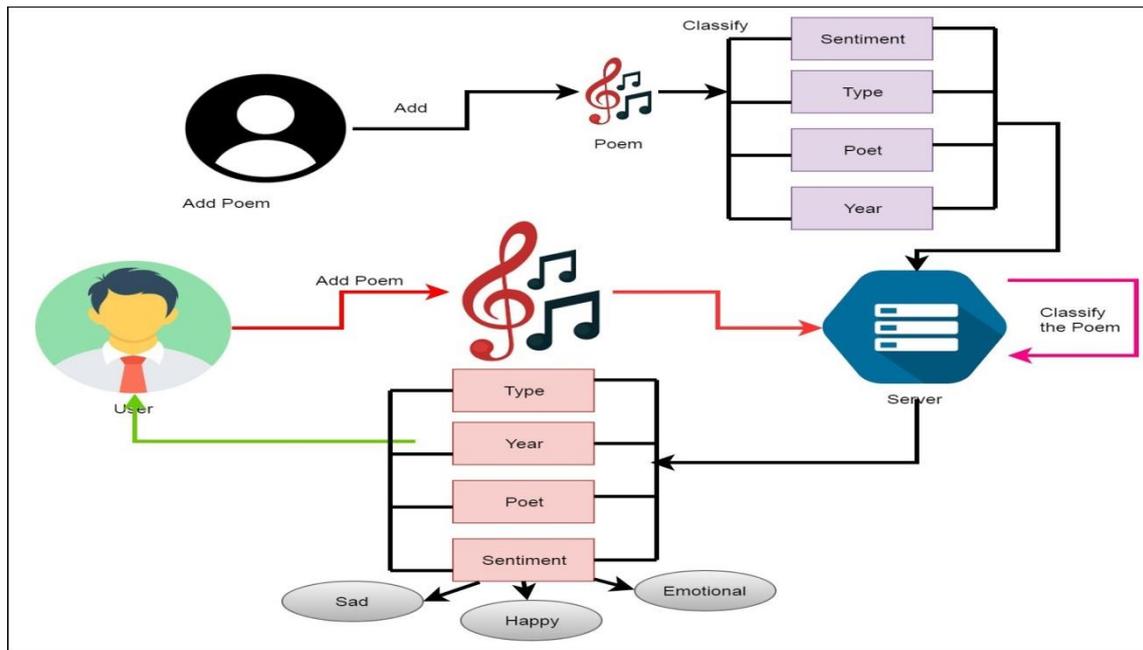


Fig.2: Architecture of the proposed system

Working of the proposed system:

- The poems will be added to the dataset.
- Then through the dataset the poet of the particular poem will be searched along with poet name, the result will also consist number of lines of poem and the year in which it was written.
- Then by using classification algorithm poem will be classified according to its sentiments or types.
- The result will be shown to user.

The proposed system will enable user to find poem based on category wise classification and poet based classification. The user will understand different nature of poem and unique feature, with the help of word which categorize poem into different category like nationalism, love, etc. User will understand the peculiar feature of every category. The user will also understand actual meaning of words in poem that user may find difficult for understanding according to poet point of view. The proposed system will be beneficial to user who has little understanding of Marathi language. It will also enhance poet reputation by analyzing their work in Marathi language and their style of writing the poem and enhancing their style of poetry.

V. CONCLUSION AND FUTURE SCOPE

In this model we are developing system that will classify the poem based on various features like love, devotional, etc. We will conduct poet identification as well as we will find ambiguous word meaning. With some preliminary investigation, we observed that words alone are not perpetually enough for classifying poems into classes, because of poets typically resorting to symbolism. We have a tendency to be in a position to determine the writer properly, many of the times using various classification algorithm. This model can be explored for the dataset containing N number of poems and we can use this analogy and methodology for essay, story classification as well. We can use Deep learning in our idea for future implementations.

REFERENCES

- [1] Geetanjali Rakshit, Anupam Ghosh, Pushpak Bhattacharyya, Gholamreza Haffari "Automated Analysis of Bangla Poetry for Classification and Poet Identification" .
- [2] Navinder Kaur, Amandeep Verma "Authorship Attribution of Punjabi Poetry using SVM Classifier" .
- [3] Xia Li, Bin Wu, Bailing Zhang "Unknown Word Detection in Song Poetry".
- [4] Pandian*, V. V. Ramalingam and R. P. Vishnu Preet "Authorship Identification for Tamil Classical Poem (Mukkoodar Pallu) using Bayes Net Algorithm".
- [5] Jasleen Kaur, Jatinder Kumar R. Saini "Punjabi Poetry Classification: The Test of 10 Machine Learning Algorithms".
- [6] Zhong shilte, Wen-Ting Liang "SVM based classification metrics for poetry style".
- [7] Zhou GuoDong "A Chunking Strategy towards Unknown Word Detection in Chinese Word Segmentation".
- [8] Dr. Binoy Barman "A contrastive analysis of English and Bangla phonemics".
- [9] Ouais Aishray, deema Alshamaa, Nada Ghrem "Emotion Classification in Arabic Poetry using Machine Learning".
- [10] Saeid Hamidi, Farbod Razzazi "A Meter Classification System for Spoken Persian Poetries".