ISSN 1870-4069

A Study on Significance on Features in Emotion Recognition System for Poems

Sreeja Ponnarassery Sreenivasan, G S Mahalakshmi

Anna University, Department of Computer Science and Engineering,
Chennai, India

srj_ps@yahoo.com, gsmaha@annauni.edu

Abstract. Poem is a type of literature designed to express, concepts, emotions, and experiences in an excellent way. The main aim of this work is to recognize emotion automatically, with an emphasis on exploring features of poems composed in English. This is an innovative approach to emotion recognition from poems. A Poem Emotion Recognition System (PERS) is developed to identify emotions from the poems, classified into nine emotions, based on *Navarasa* under *Rasa Theory* which is described in *Natyashastra* written by *Bharatha Muni*. The nine basic emotions such as *Love, Sad, Anger, Hate, Fear, Surprise, Courage, Joy, and Peace classified as Navarasa.* The poems are mined from the web and extracted ten features which will help to identify the emotion depicted by poems. The main contribution of this paper is the feature engineering. The evaluation contains measuring the performance of different feature sets across Naive Bayes classification. This experiment explains grouping of similar features gives different results, and it shows the combination of all feature gives a better result. Similarly, logistic regression identified the significant features in each emotion category.

Keywords. Poem emotion recognition corpus, emotion analysis, Naive Bayes classifier, logistic regression, maximum likelihood probability.

1 Introduction

A surge in rapidly increasing subjective expressions in text media have triggered keen interest in methods that automatically identify opinions, emotions, and sentiments in text. This proposed work explores methods for automatic emotion recognition from poetry written in English. In this research work, a Poem Emotion Recognition System (PERS) has been developed to explore the possibilities and limitations of automatic emotion recognition. Emotions are classified, based on the 'Navarasa' described in the 'Natyashastra' [17]. Navarasa comprises nine basic emotions namely love, sadness, anger, hate, fear, surprise, courage, joy, and peace.

Navarasa is based on Rasa Theory given by Bharatha Muni in Natyashastra. NatyaShastra is an Indian text dated between 2nd century BC and 2nd century AD that

analyzes all features of performing art [8]. The main contribution of the research work is as follows:

1. Extraction of Linguistic, Orthographic, Statistical, Semantic and Poetic Features.
2. Development of a poem emotion recognition model using Maximum Posterior Probability and identification of significant features.
3. Identification of most contributing feature in emotion recognition by Logistic Regression.

The proposed PERS approached the problem in two ways, utilizing machine learning methods: developing an emotion model based on the Maximum Posterior Probability and Logistic Regression Method. The rest of this work is as follows, Section 2 provides an overview of related work, Section 3 presents the proposed methodology, and Section 4 gives the details of evaluation results. Finally, conclusions are presented in Section 5.

2 Literature Survey

Classification is a data analysis task in which the model or classifier predicts the category label. A text underscores the emotional state of the writer and evokes emotions in the reader. The emotion of the text can be interpreted in different ways using assorted computational models. As Literary Arts comprises many emotions, these literature pieces especially poems can be used for the task of Emotion Recognition, which is very challenging in computational point of view. Over the past semi-century, there have been multiple approaches to emotion recognition from text, and many emotion recognition types of research based on probabilistic approaches.

Alm et al. [1] used a simple Natural Language Parser for keyword spotting, phrase length measurement, and emotion identification. Wu et al. [26] used semantic labels and a Separable Mixture Model to identify emotions. They manually generated rules for emotion, semantic labels, and attitudes with the help of emotion-generation rules, semantic labels and attitudes. Emotion association rules are automatically derived using the Apriori algorithm.

Strapparava and Mihalcea [23] constructed a large dataset of news headlines annotated for six basic emotions [6] such as anger, disgust, fear, joy, sadness, and surprise. They proposed the Latent Semantic Analysis (LSA) and the Naïve Bayes classifier and evaluated several knowledge-based methods for the automatic identification of these emotions in text. Minato et al. [15] constructed a Japanese Emotion Corpus and identified emotions automatically through an analysis of the corpus. The advantage is that it can yield high precision. However, its disadvantage is that it is impossible to determine the emotion of words that are not in the corpus. Das and Bandyopadhyay [3] used the conditional random field classifier to recognize emotion in sentences in Bengali blogs.

Perfors et al. [16] discussed, in detail, issues and applications of the Bayesian approach in cognitive science. Bielza et al. [2] proposed multi-dimensional Bayesian network classifiers which are probabilistic graphical models that systematize class and feature variables as a class subgraph, a feature subgraph and a bridge (from class

to feature) subgraph. Yoon and Chung [27] proposed a classifier based on the Nave Bayes theorem and defined emotions in in a two-level and three-level class. They represented emotions on the arousal and valence dimensions. Lei et al. [11] described how a context-aware system could be easily constructed in different domains by mashing up Bayesian network fractions independently designed or learned. Lee et al. [10] extracted context information using the Bayesian network.

Steyerberg et al. [22] used logistic regression for clinical decision making that requires estimates of the likelihood of a dichotomous outcome in individual patients. In this study, they compared alternative strategies in 23 small subsamples from a large dataset of patients with an acute myocardial infarction and developed predictive models for 30-day mortality. D'Mello et al. [4] explored the reliability of detecting a learner's affect from conversational features extracted from interactions with the Auto Tutor, an Intelligent Tutoring System that helps students learn by holding a conversation in natural language. They used multiple regression analysis and confirmed the hypothesis that dialogue features could expressively predict the affective states of confusion, boredom, frustration, and flow.

3 Poem Emotion Recognition System

This section discusses the proposed novel Poem Emotion Recognition System (PERS). In this paper PERS consists of two experiments. One is the machine learning classification model, Naive Bayes Classifier and a prediction model, Logistic Regression are detailed. Identifying emotions from poems is a classification issue, from the viewpoint of text mining. An interesting task that offers diverse challenges for classification such as the following:

3.1 Poem Emotion Recognition Corpus

Despite the availability of several lexicons in emotion analysis, those that help to identify emotions from poetry are few and far between. The lexicons that are available are not specifically poetry-centric and consequently fail to focus on poetic features. Emotion classification is based on the Navarasa described in the Natyasastra [17]. Navarasa, to recapitulate, consists of nine primary emotions: love, sadness, anger, hate, fear, surprise, courage, joy, and peace. Although there are many text corpora for emotion recognition, we are unaware of the existence of a text corpus for poetry, based on the said nine emotions.

The corpus PERC[1] created is from an exhaustive collection of poetry especially that of Indian poets during the period 1850-2016 and this corpus is publically available now. The novelty of this research is the creation of a corpus using poems mined from the web[2] and evaluated by experts in the field.The corpus has a data size

[1] http://dx.doi.org/10.17632/n9vbc8g9cx.1

[2] http://100-poems.com, http://www.poetry-chaikhana.com, http://www.poetry-chaikhana.com, http://allpoetry.com

of 736 poems by ten leading poets from the period 1850 to the present day. The average number of words per poem, across the eight poets, ranges from 74-284. Table 1 shows the details of the corpus, such as the poet's name and the number of poems collected for each of the poets. Table 2 depicts the number of poems available in each emotion category.

Definition 1:

We define the PERC as a set of P poems of pairs (p, E) where p is a poem and E is the emotion from a set of features. Our collection is drawn from the work of ten leading poets of the period, including Rabindranath Tagore, Sarojini Naidu, Aurobindo, Ananda Murthy, Darshan Singh, Jiddu Krishnamurthi, Lalan, Nazrul Islam, Kamala Das, and Meena Kandasamy. The poems have been amassed from several well-known sites as the first step towards their collection and selection.

Inter-rater reliability [7] is one of the most efficient methods for the evaluation of the corpus. Since the corpus has been assembled by human experts rather than a machine learning technique, we calculate inter-rater reliability using the Fleiss kappa measurement primarily to study the closeness between the annotations. Since our corpus has acquired an inter-rater agreement value of 0.4816, we have proved that the corpus created shows moderate agreement [14], which is reasonably good.

The corpus has a data size of 736 poems by ten leading poets from the period 1850 to the present day. Table 1 shows the details of the corpus, such as the poet's name and the number of poems collected for each of the poets. Table 2 details the number of poems available in PERC for each emotion category.

Table 1. Corpus Details.

Poet	No. of Poems	Poet	No. of Poems
Rabindranath Tagore	284	Darshan Singh	20
Aurobindo	110	Anandamurthy	16
Sarojini Naidu	90	Lalan	20
Jiddu Krishnamurthy	20	Kamala Das	69
Nazrul Islam	32	Meena Kandasamy	75
Total 736			

Table 2. Poem Details.

Emotion	No. of Poems	Emotion	No. of Poems
Anger	52	Love	154
Courage	64	Peace	82
Fear	54	Sadness	126
Hate	54	Surprise	50
Joy	100	Total	736

3.2 Feature Extraction

Each instance that provides inputs to PERS is characterized by its values on a fixed, predefined set of features or attributes. To aid proficiency, Features clustered into five groups: linguistic, statistical, semantic, orthographic and poetic.

Linguistic Features: Linguistic features include nouns, verbs, adjectives, and adverbs. Words in grammatical classes show significant results in emotion recognition. Stanford Tagger [25] is used for POS tagging, and extracted noun, verb, adjective and adverb classes. These features comprise a dichotomy that is Boolean.

Statistical Features: This feature set includes term frequency and inverse document frequency, which are the most common features used in classification problems. Term Frequency (TF) and Inverse Document Frequency (IDF) are the two statistical features used in this Emotion Recognition model [19]. Generally, TF of a term can be defined as number of times a term occurs in a document and Document Frequency (DF), is defined as no. of documents having that term. Inverse Document Frequency (IDF) is calculated [19] by Equation (1):

$$IDF = log \frac{Total\ No\ of\ Documents}{DF}. \quad (1)$$

Orthographic Features. Orthographic features include negation words, the title of the poem, the last line of the poem, and the repeated lines/refrain. The negation words saved in a list are extracted through mapping. These negation words act as emotion modifiers that play an important role in emotion identification. The title of the poem is an essential feature in Emotion Recognition (ER) as it offers clues about the content of the poem. The last line irrefutably concludes the poem. This line is, most of the time, the one that carries the most emotion in the poem. The repeated lines that constitute the refrain are another strategic feature, with the repetition underscoring their significance in the poem.

Poetic Features. In this set, only two features similes and metaphors are targeted. The rules that extract these two features are explained below.

Rule 1: If the $word_i$ is like, and the POS tag of like is not equal to the verb, the $word_{i-1}$ and the $word_{i+1}$ are compared, and a simile is found to exist.

Rule 2: If the $word_i$ is as the $word_{i-1}$ and the $word_{i+1}$ two are compared, a simile is found to exist.

Rule 3: If the $word_i$ is are, is, was and were, and the $word_{i+1}$ is not in the present continuous tense form of the verb, a metaphor is found to exist.

Example 1:

"Let thy hue-winged lyrics hover like birds
Over the swirl of the heart's sea. "
- Musa Spiritus, Aurobindo.[3]

In this example, a simile is used when lyrics are compared to birds.

Example 2:

I had gone a-begging from the door in the village path, When thy golden
Chariot appeared in the distance like a gorgeous dream, And I wondered who

[3] https://allpoetry.com/Musa-Spiritus

Was this king-of-all kings!"

- I Had Gone A-begging, Rabindranath Tagore[4]

In this example, the golden chariot is a metaphor for riches and wealth. The chariot, Gilded in gold, indicates that the man to whom it belonged was immensely wealthy.

Semantic Features. Approaches to identifying emotions have evolved from keyword-based methods to semantic-based methods. Semantic-based approaches use semantic-based dictionaries or a semantic knowledge base. They help to identify the conceptual information associated with emotions. A conceptual model is illustrated for concept identification from poems that principally relies on the conceptual information extracted from ConceptNet [13].

ConceptNet [13] is a semantic network of commonsense knowledge made up of over a million simple sentences contributed by volunteers on the Open Mind Common Sense website [12]. This problem is approached by adding a weight that defines a function *C(S, wt)* to identify the concept of a given poem [21]. *S* is a set of relations *S= a, r, c* where *a* is a word in the poem, *r* the relation, and *c* the concept. *wt* is the weight assigned to each relation in ConceptNet. The concept can be defined as follows.

Definition 2:

Concept is defined as a key value pair, *c = (cid; wt)*, where *cid* refers to the unique id of the concept and *wt* is the weight assigned to the respective concept [21].

3.3 Poem Emotion Recognition Using Naive Bayes Classifier

The classification technique in Naive Bayes Classifier is based on the maximum posterior probability value. Bayesian classification is based on Baye's theorem. From Section 2, it is evident that Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases. Naive Bayesian classifiers assume that the effect of a feature value on a given class is independent of the other feature values. This assumtion is called class conditional independence [23]. This work is the extension of Sreeja and Mahalakshmi [24] where emotion is recognized by maximum posterior probability using bag of word feature. In this paper, different set of features and its combinations are used to analyze the emotion recognition process. The Naive Bayes classifier is used to identify emotion from text and conducted this research with a different set of features. The process of identifying emotion from a poem can be defined as follows:

Definition 3:

Let *TP* be a set of training poems contains set of words and their associated emotion labels. Each word is represented by a feature vector, $X_i = (x_1, x_2, x_3 \ldots x_t)$, depicting *t* values made on the token from *t* features, respectively, $x_1, x_2, \ldots x_t$. Let *p* be a test Poem have X_m words. The emotion classes $E_1, E_2 \ldots E_9$ are love, sadness, courage, anger, hate, joy, fear, peace, surprise. Given testing Poem, *p*, and the classifier will predict that *p* belongs to emotion having highest posterior probability, conditioned on

[4] https://allpoetry.com/poem/11404303-I-Have-Gone-a-Begging-From-Door-to-Door

p. That is, the Naive Bayesian classifier predicts that poem p belongs to the class E_i, if and only if by Equation (2):

$$PR(E_i|p) > PR(E_j|p) \quad for\ 1 \le j \le 9, j \ne i\ , \qquad (2)$$

where $p = (X_1, X_{2,\ldots,} X_m)$. Thus maximizes $(E_i|p)$. The class E_i for which $PR(E_i|p)$ is maximized and called the maximum posterior hypothesis. By Bayes' theorem, by Equation (3):

$$PR(E_i \mid p) = \frac{PR(p \mid E_i) \times PR(E_i)}{PR(p)}. \qquad (3)$$

As *PR(p)* is constant for all classes, it can be maximized using Equation (4):

$$PR(E_i \mid p) = \arg\max_{i \in (1..9)} PR(p \mid E_i) \times PR(E_i)\ , \qquad (4)$$

$PR(E_i)$ is calculated by Equation (5):

$$PR(E_i) = N(E_i) \Big/ \sum_{i=1}^{9} N(E_i)\ , \qquad (5)$$

with many features, it would be computationally expensive to compute $PR(p|E_i)$. To reduce computation, naive assumption of class conditional independence is made by Equation (6). It means that there are no dependency relationships among the features:

$$PR(p \mid E_i) = \prod_{K=1}^{m} (X_K \mid E_i) \qquad (6)$$

$$= PR(X_1 \mid E_i) \times PR(X_2 \mid E_i) \times \ldots\ldots \times PR(X_m \mid E_i)\ .$$

$$PR(X_K \mid E_i) = \prod_{j=1}^{n} (x_{Kj} \mid E_i) \qquad (7)$$

$$= PR(x_{K1} \mid E_i) \times PR(x_{K2} \mid E_i) \times \ldots\ldots \times PR(x_{Kj} \mid E_i)\ .$$

The probabilities $PR(x_{K1}| E_i)$, $PR(x_{K2}| E_i)$, …. $PR(x_{Kj}| E_i)$ are found from the training set, where x_{Kj} refers the K^{th} word's j^{th} feature value.

The main disadvantage of the Naive Bayes classifier is its conditional independence Domingos and Pazzani [25], the assumption that the effect of an attribute value on a given class is independent of the values of other attributes. The other issues are data scarcity [26] and imbalanced classes. To identify the most contributing feature in poem emotion recognition Logistic Regression model is applied.

3.4 Poem Emotion Recognition Using Logistic Regression

In general, regression is a method to predict the value of a dependent variable from one or more independent variables. There are different forms of regression. In this section, the research carried out in logistic regression is discussed. Excel-Solver[5], is

[5] https://www.solver.com/

used to calculate the Logistic Regression Function for each class. The steps involved in calculating Logistic regression function are as follows.

1. Creation of an excel sheet containing numeric feature values.

In Figure[6] 1, column B gives the title of the poem. Columns C to L show the number of nouns, verbs, adjectives, and adverbs; the number of words in the title; the number of repeated words; the number of negation words; and the number of metaphors and similes in the poem. Since the logistic regression is carried out separately for each class, a trick is used, setting the output equal to 1 for the training instances that belong to the class and 0 for those that do not.

2. Sort the dependent variable to make the data evident.

3. From the given input, x_1, x_2, x_3, ... x_n the 'Logit' equal to the expression shown on the right-hand side of Equation (8):

$$Logit\ L = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n. \quad (8)$$

The explanatory variables are certain nouns, verbs, adjectives, adverbs, title words, repeating words, last-line words, negation words, similes, and metaphors. The *Logit L* can be written as:

$$Logit\ L = b_0 + b_1noun + b_2verb + b_3adjective + b_4adverb + b_5title - word + b_6repeating - word + b_7last - lineword + b_8negation + b_9simile + b_{10}metaphor,$$

where *Logit L* is a link function states the relation between predictor and the mean of the distribution function, b_0 is the intercept from the regression Equation, b_1 *to* b_{10} are regression coefficients or decision variables. The decision variables $b_0 ... b_{10}$ are set to 0.1. The solver will adjust the decision variables during the optimization process.

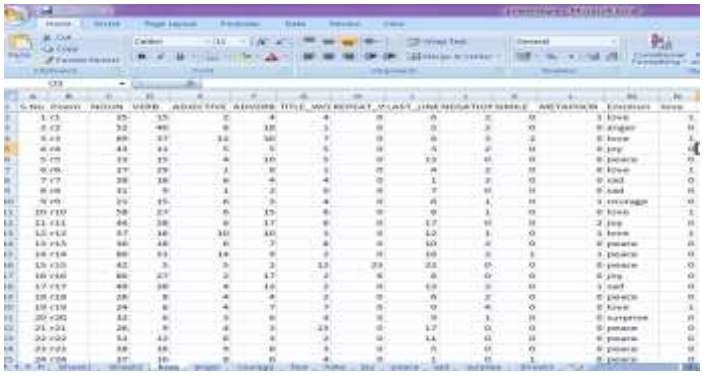

Fig. 1. Data for Logistic Regression Model.

[6] http://dx.doi.org/10.17632/n9vbc8g9cx.1

4. Calculate the $\boldsymbol{e^L}$ for each record and *PR(X)* by Equation (9):

where *'e'*=2.718281.

$$PR(X) = e^L/(1 + e^L), \quad (9)$$

where $Logit\ L = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n$.

5. Calculate the Log-likelihood Function (LL) using Equation (10):

$$PR(Y_i = y_i|x_{1i}, x_{2i}, \ldots . x_{ni}). \quad (10)$$

Equation (10) is the conditional probability that the predicted dependent variable y_i equals the observed value Y_i for the given independent variables $x_{1i}, x_{2i}, \ldots . x_{ni}$. It can be calculated by the following Equation (11):

$$PR(Y = y|X) = PR(X)^y[1 - PR(X)]^{(1-y)}. \quad (11)$$

Taking natural logarithm on both sides of Equation (11) leads to Equation (12):

$$\ln[PR(Y = y|X)] = y.\ln[PR(X)]\,.\,(1 - y)\ln[1 - PR(X)]. \quad (12)$$

Log-likelihood function LL is the sum of terms for all data records by the following Equation (13):

$$LL = \sum Y_i PR(X_i) + (1 - Y_i)(1 - PR(X_i)). \quad (13)$$

6. Calculate the *Maximum Log-Likelihood (MLL)* function using Solver.

The objective of the Logistic regression is to find the decision variables that maximize the *LL* function to produce the *MLL*. The Solver adjusts the decision variables GRG Nonlinear solving method. Solver was run several trials to obtain the optimum solution. The optimal solution is calculated by equation (14) to (22) for each emotion category.

$$Love = 0.3736 + 0.0042x_1 - 0.0056x_2 - 0.026x_3 + 0.0063x_4 + 0.0023x_5 + 0.0036x_6 - 0.0091x_7 - 0.0121x_8 + 0.0157x_9 + 0.0871x_{10} \quad (14)$$

$$Anger = 0.0137 + 0.0003x_1 + 0.0008x_2 + 0.0005x_3 + 0.0004x_4 - 0.0011x_5 - 0.0004x_6 - 0.0013x_7 - 0.0001x_8 + 0.0055x_9 + 0.0022x_{10} \quad (15)$$

$$Courage = 0.0402 + 0.0007x_1 - 0.0014x_2 - 0.0008x_3 + 0x_4 + 0.0026x_5 - 0.002x_6 + 0.0011x_7 + 0.0034x_8 - 0.005x_9 - 0.0069x_{10} \quad (16)$$

$$Fear = 0.0114 + 0.0003x_1 + 0.0012x_2 - 0.0006x_3 - 0.0015x_4 - 0.0013x_5 + 0.0005x_6 - 0.0003x_7 - 0.0006x_8 + 0.0045x_9 + 0.0049x_{10} \quad (17)$$

$$Hate = 0.0062 + 0.0002x_1 - 0.0001x_2 - 0.001x_3 + 0.0001x_4 + 0.0001x_5 - 0.0001x_6 - 0.0003x_7 - 0.0008x_8 - 0.0019x_9 - 0.001x_{10} \quad (18)$$

$$Joy = 0.1358 + 0.0001x_1 + 0.0012x_2 + 0.0026x_3 - 0.0084x_4 + 0.0026x_5 + 0.0013x_6 + 0.0088x_7 - 0.0045x_8 - 0.0003x_9 - 0.0047x_{10} \quad (19)$$

$$Peace = 0.1969 - 0.0024x_1 + 0.0044x_2 + 0.0149x_3 - 0.0059x_4 - 0.0008x_5 - 0.0016x_6 - 0.0009x_7 + 0.0044x_8 - 0.0374x_9 - 0.0555x_{10} \quad (20)$$

$$Sadness = 0.2168 - 0.0031x_1 + 0.0004x_2 + 0.0104x_3 + 0.0087x_4 - 0.0025x_5 - 0.0006x_6 - 0.001x_7 + 0.0066x_8 + 0.037x_9 - 0.029x_{10} \quad (21)$$

$$Surprise = 0.0324 + 0.0002x_1 - 0.0008x_2 - 0.0001x_3 + 0.0004x_4 - 0.002x_5 - 0.0009x_6 + 0.0011x_7 + 0.0022x_8 - 0.007x_9 + 0.0074x_{10} \quad (22)$$

From these equations, it is observed that metaphors contribute the most to identifying emotions like love, fear, and surprise. Further, it is noted that similes portray sadness, verbs display anger, and negation words reveal courage. Nouns illustrate hate, last-line words depict joy and adjectives evoke peace. The result of this prediction experiment is given below. For a given test poem, if the emotion to be predicted is love, then, of the results obtained by Equations (14) to (22), the output of Equation (14) should be greater than all the other values.

4 Results and Discussions

The main objective is to evaluate the importance of the features described in Section 3.2 for recognizing the emotion of a poem. To evaluate these features, the experiments are carried out using the PERC, by considering features in single and combined mode. These methods are experimented with in 10-fold cross-validation. In 10-fold cross-validation [27], the data set is split into ten sets of size n=10; train on nine datasets and test on 1. This process is repeated ten times to achieve better accuracy. Another key reason for using ten-fold cross-validation instead of conventional validation is if there is insufficient data to be partitioned into two distinct training and testing sets without missing out on unique information. In such cases, cross-validation is a suitable technique.

Table 3 shows that 443 poems are correctly classified using linguistic features alone. Nouns, verbs, adjectives, and adverbs are the linguistic features included in the model. Table 4 shows that 484 poems are correctly classified using only poetic features, similes, and metaphors.

Table 5 shows that 503 poems are classified correctly when a combination of poetic and linguistic features is used. Table 6 shows that 535 poems are identified correctly when all the features are used for classification. A consistent improvement in results is observed, and Tables 3, 4, 5, and 6 show that the classification carried out with all the features (linguistic, semantic, orthographic, poetic and statistical) gives better results than the other methods. Table 7 shows that 560 poems are identified correctly by this method.

Table 3. Confusion Matrix of 10-fold Cross Naive Bayes Classifier with Linguistic Features.

Emotion	No. of Poems in PERC	Anger	Courage	Fear	Hate	Joy	Love	Peace	Sad	Surprise
Anger	52	30	8	0	7	0	0	0	7	0
Courage	64	7	32	0	5	8	5	0	3	4
Fear	54	3	0	20	9	2	9	1	10	0
Hate	54	13	5	3	22	0	1	0	9	1
Joy	100	1	9	0	1	68	14	2	0	5
Love	154	2	4	0	2	8	115	3	18	2
Peace	82	1	4	3	0	8	7	50	4	5
Sad	126	4	0	5	7	0	15	3	92	0
Surprise	50	3	4	4	1	13	5	6	0	14

Table 4. Confusion Matrix of 10-fold Cross Naive Bayes Classifier with Poetic Features.

Emotion	No.of Poems in PERC	Anger	courage	Fear	Hate	Joy	Love	Peace	Sad	Surprise
Anger	52	32	7	0	7	0	0	0	6	0
Courage	64	5	34	1	5	6	5	1	3	4
Fear	54	3	0	24	9	2	9	0	7	0
Hate	54	10	4	5	20	0	4	0	8	3
Joy	100	1	8	1	1	65	16	2	1	5
Love	154	2	3	0	1	5	120	1	21	1
Peace	82	3	4	2	0	5	7	52	4	5
Sad	126	2	0	1	1	0	12	0	110	0
Surprise	50	1	2	2	1	8	4	5	0	27

Table 5. Confusion Matrix of 10-fold Cross Naive Bayes Classifier with Linguistic and Poetic Features.

Emotion	No. of Poems in PERC	Anger	Courage	Fear	Hate	Joy	Love	Peace	Sad	Surprise
Anger	52	34	4	0	6	0	1	0	7	0
Courage	64	4	38	0	4	5	4	0	4	5
Fear	54	2	0	21	10	2	11	0	8	0
Hate	54	7	4	6	25	0	2	0	7	3
Joy	100	1	7	0	0	67	18	2	1	4
Love	154	2	5	0	2	3	120	3	18	1
Peace	82	3	4	2	0	4	5	57	4	3
Sad	126	2	0	1	2	0	9	0	112	0
Surprise	50	1	3	2	2	10	5	4	0	23

Table 6. Confusion Matrix of 10-fold Cross Naive Bayes Classifier with Linguistic, Orthographic, Poetic, Statistical and Semantic Features.

Emotion	No. of Poems in PERC	Anger	Courage	Fear	Hate	Joy	Love	Peace	Sad	Surprise
Anger	52	40	6	0	3	0	0	0	3	0
Courage	64	6	42	0	3	7	3	0	1	2
Fear	54	1	0	30	7	0	7	0	9	0
Hate	54	11	3	2	32	0	0	0	6	0
Joy	100	0	8	0	0	78	11	0	0	3
Love	154	0	1	0	0	6	130	1	16	0
Peace	82	1	3	1	0	7	5	60	2	3
Sad	126	2	0	2	5	0	12	3	102	0
Surprise	50	2	3	5	0	9	7	3	0	21

Table 7. Confusion Matrix of 10-fold Cross Logistic Regression.

Emotion	No.of Poems	Anger	courage	Fear	Hate	Joy	Love	Peace	Sad	Surprise
Anger	52	42	1	4	3	0	1	0	1	0
Courage	64	4	36	0	4	6	5	2	2	5
Fear	54	0	0	42	2	0	4	0	4	2
Hate	54	7	4	3	34	0	2	1	6	1
Joy	100	3	5	0	0	78	6	5	1	2
Love	154	0	0	0	0	5	137	3	7	2
Peace	82	1	1	3	0	7	7	60	1	2
Sad	126	2	1	2	5	0	8	2	105	1
Surprise	50	1	2	4	0	10	4	3	0	26

Table 8. Comparison of Precision Measures.

Emotion	All Features	Linguistic Features	Poetic Features	Linguistic and Poetic Features	Logistic Regression
Anger	0.635	0.469	0.542	0.607	0.69
Courage	0.636	0.485	0.548	0.585	0.75
Fear	0.75	0.571	0.667	0.711	0.652
Hate	0.64	0.407	0.444	0.49	0.696
Joy	0.729	0.636	0.714	0.736	0.736
Love	0.743	0.673	0.678	0.686	0.778
Peace	0.896	0.769	0.852	0.864	0.789
Sad	0.734	0.643	0.688	0.696	0.823
Surprise	0.724	0.452	0.6	0.59	0.583

Table 8 gives detailed information on the precision measures obtained in the four experiments. The experiment with all the features gives a more promising precision measure when compared to the other three methods. The emotion category, peace, has a higher precision measure compared to the other classes. Table 9 details the recall measures obtained in the four experiments. The experiment with all the features gives a more promising recall measure, except in the case of the emotion category, sadness. The experiment with linguistic and poetic features gives better results in the classification of sad poems. Table 10 shows that the experiment with all the features offers better results compared to that of linguistics and poetic features. However, in the classification of sad poems, the experiment with all the features as well as the experiment with the linguistic and poetic features combined give, more or less, the same F-measure.

Table 9. Comparison of Recall Measures.

Emotion	All Features	Linguistic Features	Poetic Features	Linguistic and Poetic Features	Logistic Regression
Anger	0.769	0.577	0.615	0.654	0.8
Courage	0.656	0.5	0.531	0.594	0.6
Fear	0.556	0.37	0.444	0.5	0.714
Hate	0.593	0.407	0.37	0.463	0.615
Joy	0.78	0.68	0.65	0.67	0.78
Love	0.844	0.747	0.779	0.779	0.884
Peace	0.732	0.61	0.634	0.695	0.732
Sad	0.81	0.73	0.873	0.889	0.829
Surprise	0.42	0.28	0.54	0.46	0.467

Table 10. Comparison of F Measures.

Emotion	All Features	Linguistic Features	Poetic Features	Linguistic and Poetic Features	Logistic Regression
Anger	0.696	0.517	0.577	0.63	0.741
Courage	0.646	0.492	0.54	0.589	0.667
Fear	0.638	0.449	0.533	0.587	0.682
Hate	0.615	0.407	0.404	0.476	0.653
Joy	0.754	0.657	0.681	0.702	0.757
Love	0.79	0.708	0.725	0.729	0.828
Peace	0.805	0.68	0.727	0.77	0.759
Sad	0.77	0.684	0.769	0.78	0.826
Surprise	0.532	0.346	0.568	0.517	0.519

5 Conclusion

In this research, the importance of features in emotion identification is studied. The proposed work, also groups the features and evaluated the significance of each group of features in terms of the results. It is found that certain features - linguistic, poetic, statistical, semantic and orthographic - influence the results significantly. Accuracy was somewhat diminished, given the shortcomings of the Naive Bayes classifier in terms of conditional independence, data scarcity, and imbalanced classes. Laplace smoothing solved information scarcity. Experiments with the Logistic Regression

Method helped determine the contribution of each feature in the process of prediction. As future work, we intend to increase the corpus size and number of poetic features such as meter, sarcasm, and other poetic devices. To improve the accuracy some ensemble of base classifier can be applied.

References

1. Rangacharya, A.: Bharatha Natyshastra. Munshiram Manoharlal Publishers (1996)
2. Ghosh, M.: The natyashastra (English translation) volume i (chapters i-xxvii). Calcutta: The Royal Asiatic Society of Bengal (1950)
3. Alm, E.C.O.: Affect in text and speech. University of Illinois at Urbana Champaign (2008)
4. Wu, C.H., Chuang, Z.J., Lin, Y.C.: Emotion recognition from text using semantic labels and separable mixture models. ACM transactions on Asian language information processing (TALIP), 5(2), 165–183 (2006)
5. Strapparava, C., Mihalcea, R.: Learning to identify emotions in text. In: Proceedings of the 2008 ACM symposium on Applied computing, ACM, 1556–1560 (2008)
6. Ekman, P.: An argument for basic emotions. Cognition & emotion 6(3-4), 169–200 (1992)
7. Minato, J., Bracewell, D.B., Ren, F., Kuroiwa, S.: Japanese emotion corpus analysis and its use for automatic emotion word identification. Engineering Letters, 16(1) (2008)
8. Das, D., Bandyopadhyay, S.: Word to sentence level emotion tagging for bengali blogs. In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Association for Computational Linguistics, 149–152 (2009)
9. Perfors, A., Tenenbaum, J.B., Griffiths, T.L., Xu, F.: A tutorial introduction to bayesian models of cognitive development. Cognition, 120(3), 302–321 (2011)
10. Bielza, C., Li, G., Larranaga, P.: Multi-dimensional classification with Bayesian networks. International Journal of Approximate Reasoning, 52(6), 705–727 (2011)
11. Yoon, H.J., Chung, S.Y.: Eeg-based emotion estimation using bayesian weightedlog-posterior function and perceptron convergence algorithm. Computers in biology and medicine 43(12), 2230–2237 (2013)
12. Lei, J., Rao, Y., Li, Q., Quan, X., Wenyin, L.: Towards building a social emotion detection system for online news. Future Generation Computer Systems, 37, 438–448 (2014)
13. Lee, S.H., Yang, K.M., Cho, S.B.: Integrated modular bayesian networks with selective inference for context-aware decision making. Neurocomputing, 163, 38–46 (2015)
14. Steyerberg, E.W., Eijkemans, M.J., Harrell Jr, F.E., Habbema, J.D.F.: Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Medical Decision Making, 21(1), 45–56 (2001)
15. Dmello, S.K., Craig, S.D., Witherspoon, A., Mcdaniel, B., Graesser, A.: Automatic detection of learners affects from conversational cues. User modeling and user adapted interaction, 18(1), 45–80 (2008)
16. Fleiss, J.L.: Measuring nominal scale agreement among many raters. Psychological bulletin, 76(5), 378 (1971)
17. Lubis, N., Lestari, D., Purwarianti, A., Sakti, S., Nakamura, S.: Construction and analysis of Indonesian emotional speech corpus. In: Coordination and Standardization of Speech Databases and Assessment Techniques (COCOSDA), 2014 17th Oriental Chapter of the International Committee for the IEEE, 1–5 (2014)
18. Toutanova, K., Klein, D., Manning, C.D., Singer, Y.: Feature-rich part-of-speech tagging with a cyclic dependency network. In: Proceedings of the 2003 Conference of the North

American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, Association for Computational Linguistics, 173–180 (2003)
19. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Communications of the ACM, 18(11), 613–620 (1975)
20. Liu, H., Singh, P.: Conceptneta practical commonsense reasoning tool-kit. BT technology journal, 22(4), 211–226 (2004)
21. Liu, H., Singh, P.: Commonsense reasoning in and over natural language. In: Knowledge-based intelligent information and engineering systems, Springer, 293–306 (2004)
22. Sreeja, P.S., Mahalakshmi, G.S.: Concept identification from poems. In: Recent Trends and Challenges in Computational Models (ICRTCCM), 2017 Second International Conference on IEEE, 211–216 (2017)
23. Tan, P.N., et al.: Introduction to data mining. Pearson Education India (2006)
24. Sreeja, P.S., Mahalakshmi, G.S.: Emotion recognition from poems by maximum posterior probability. International Journal of Computer Science and Information Security 14, (CIC 2016), 36–43 (2016)
25. Domingos, P., Pazzani, M.: On the optimality of the simple bayesian classifier under zero one loss. Machine learning, 29(2), 103–130 (1997)
26. Rennie, J.D., Shih, L., Teevan, J., Karger, D.R.: Tackling the poor assumptions of naive bayes text classifiers. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), 616–623 (2003)
27. Kohavi, R., et al.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai, Volume 14, Stanford, CA, 113–1145 (1995)