ORIGINAL RESEARCH

BEmoC: A Corpus for Identifying Emotion in Bengali Texts

MD. Asif Iqbal[1] · Avishek Das[1] · Omar Sharif[1] · Mohammed Moshiul Hoque[1] · Iqbal H. Sarker[1]

Abstract

Emotion classification in text has growing interest among NLP experts due to the enormous availability of people's emotions and its emergence on various Web 2.0 applications/services. Emotion classification in the Bengali texts is also gradually being considered as an important task for sports, e-commerce, entertainments, and security applications. However, It is a very critical task to develop an automatic emotion classification system for low-resource languages such as, Bengali. Scarcity of resources and deficiency of benchmark corpora make the task more complicated. Thus, the development of a benchmark corpus is the prerequisite to develop an emotion classifier for Bengali texts. This paper describes the development of an emotional corpus (hereafter called 'BEmoC') for classifying six emotions in Bengali texts. The corpus development process consists of four key steps: data crawling, pre-processing, labelling, and verification. A total of 7000 texts are labelled into six basic emotion categories such as anger, fear, surprise, sadness, joy, and disgust, respectively. Dataset evaluation with 0.969 Cohen's κ score indicates the close agreement between the corpus annotators and the expert. The analysis of evaluation also represents that the distribution of emotion words obeys Zipf's law. Moreover, the results of BEmoC analysis shown in terms of coding reliability, emotion density, and most frequent emotion words, respectively.

Keywords Natural language processing · Bengali language processing · Emotion corpus · Emotion classification · Emotion text · Evaluation

Introduction

Classification of emotion in text implies to the task of automatically attributing an emotion class to a script chosen from a set of predetermined emotion classes. With a growing amount of users in virtual platforms producing online contents steadily as a fast-paced, interpreting emotion or sentiment in online contents is great value for consumers, business leaders, and other parties concerned. Classifying emotion in text plays a vital role in several HCI applications

where the text used as the key means of communication such as, emails, instant messages, chatting, forums, reviews, blogs, and other Web 2.0 platforms (Twitter, YouTube, and Facebook). There are several applications areas where there is a need to understand and interpret emotions in text contents such as education, business, sports, politics, psychology, and entertainment. Classifying emotion recognition in text is one of the complicated tasks in NLP, which demands the understanding of natural language [1]. The hindrance commences at sentence level in which an emotion stated through the semantics of words and their connections; as the level enhances, the difficulty of the problem grows. Nevertheless, not all opinions are stated explicitly; there are metaphors, mockery, and irony [2].

Sentiment classification from texts can be divided into two categories: opinion-based and emotion-based. Opinion classification is based on text polarity, which classifies text/sentences into positive, negative, or neutral sentiments [3, 4]. Emotion classification deals with classifying sentences according to their emotions [5]. Bengali is the fifth most-spoken native language in the world. Approximately 228 million people all over the world speak Bengali as their first

✉ Mohammed Moshiul Hoque
moshiul_240@cuet.ac.bd

MD. Asif Iqbal
asifiqbalsagor123@gmail.com

Avishek Das
avishek.das.ayan@gmail.com

Omar Sharif
omar.sharif@cuet.ac.bd

Iqbal H. Sarker
iqbal@cuet.ac.bd

[1] Department of CSE, CUET, Chittagong 4349, Bangladesh

language, and around 37 million people speak it as a second language [6]. In recent years, data storage on the web increased exponentially due to the emergence of Web 2.0 applications and its related services in the Bengali language. Most of these data are available in textual forms such as reviews, opinions, recommendations, ratings, comments and feedback which are mostly in unstructured form. Due to the supervised characteristics of the classification approaches, a comparatively larger amount of labelled corpus is prerequisite to train the machine learning models to acquire reasonable performances. Nevertheless, quality labelled data is inadequate in various domains. The analysis of these enormous amounts of data to extract underlying sentiment or emotions is a challenging research problem for the resource-constrained language, especially in Bengali [7]. NLP technologies have failed to achieve reasonable performance when dealing with languages with limited resources, such as Bengali. The complexity arises due to various limitations such as the lack of tools, scarcity of benchmark corpus, and learning techniques.

Because Bengali is a resource-scarce language, emotion based classification based on six emotion classes have not yet been performed, to the best of our knowledge. Emotion corpus is the key requirement and hence prerequisite to develop the automatic emotion classifier for Bengali text. Thus, the question becomes: how to develop an emotion corpus for the purpose of emotion classification in Bengali text? To address this question, this paper outlines the development process of a corpus (we called it *BEmoC-Bengali Emotion Corpus*) that can be used for classifying emotions in Bengali texts. We consider six types of textual emotions such as joy, sadness, anger, fear, surprise, and disgust based on Ekman's basic classification of emotion [8]. The major contributions of our work illustrates in the following:

- Develop an emotion corpus for classifying Bengali texts into one of the six basic emotions including detailed description of development phases.
- Identify the characteristics of data in each emotion class.
- Labelled the 7000 text data into six emotion classes to develop BEmoC.
- Evaluate the developed BEmoC with several metrics including Cohen's Kappa, Zipf's law, coding reliability, most frequent emotion word and density of emotions.

The remaining of the paper is arranged as the following: Section "Related Work" represents the related work. The properties of emotion in Bengali text is explained in Section "Properties of Emotion in Bengali Texts". Section "BEmoC Development" describes the development process of the BEmoC including its major steps. Various analysis of BEmoC is presented in Section "Analysis of BEmoC". Section "Metrics for Evaluation" explains several performance matrices used to evaluate the developed BEmoC. The details analysis of evaluation is presented in Section "Evaluation of BEmoC". Section "Discussion" discussed some critical issues during the development of BEmoC. Finally, the paper is concluded with a summary in Section "Conclusion".

Related Work

Emotions are the long studied topic in psychology. An emotion can be described as a particular empathy that denotes a person's state of mind, for example, love, joy, anger, fear, and disgust [9]. Classification of emotion can be divided into three major classes: emotion classification from facial expressions, emotion classification from voice, and emotion classification in text. A dataset consisted of 3018 documents was developed which consider both motion captures and audio data annotated with 4 emotion classes (angry, joy, neutral, and sad) [10]. Two very popular datasets AIBO [11] and IEMOCAP [12] were developed for analyzing emotions. The AIBO dataset contained four emotion classes (angry, emphatic, neutral, and positive) with 18,216 audio data from the interaction between children and robots. The IEMOCAP dataset included 840 audio/video utterances with four annotation classes (angry, joy, sad, and neutral). The key focus of our work is to classify the emotion from the text. Various corpus are available to classify emotion in text for different languages which are mainly on sentiment polarity based or emotion class based. In the next section, we review the recent developments on corpus for emotion classification in text in terms of two aspects: corpus for sentiment classification and corpus for emotion classification.

Corpus for Sentiment Classification

Sentiment classification refers to the technique of retrieving implicit or explicit polarity of expressions in textual contents [13]. Datasets of several languages were created to perform the sentiment classification in text. Smadi et al. [14] developed a corpus consisting of 2838 Arabic book reviews which are labelled in positive, negative, conflict, and neutral categories. A dataset developed for analyzing sentiment in Czech text which contains 2200 IT product reviews in two sentiment polarities: positive and negative [15]. Apidianaki et al. [16] developed a corpus which consists of 457 restaurant reviews in French language and reviews are labelled into positive, negative and neutral sentiment classes. Semantic Evaluation (SemEval) introduced a complete dataset [17] in English for sentiment classification where restaurant (3841 reviews) and laptop (3845 reviews) texts annotated with four sentiment categories: positive, negative, conflict, and neutral. Later this was expanded in SemEval-2016 [18] to the sentiment analysis by adding multi-lingual datasets in

which eight languages (English, French, Dutch, Russian, Spanish, Arabic, Chinese, and Turkish) over six domains (such as restaurants, laptops, hotels, consumer electronics, telecom, and museum reviews) were incorporated. A corpus developed on analyzing sentiment in Italian language which contains 3288 twitter messages into five categories: positive, negative, ironic, mixed, and objective [19]. Few datasets were developed for analyzing sentiment in different domains such as, product data [20], multi-domain sentiment dataset [21], IMDB movie reviews dataset [22], and Stanford sentiment treebank [23]. These datasets are mainly used for research. An Arabic sentiment dataset containing 10,000 twitter messages was developed to classify sentiment into positive, negative or neutral polarities [24]. A recent dataset is developed for sentiment classification (into positive, negative, and neutral) in Chinese text which contained 9522 data from microblogs, travel websites, product forums, and other fields [25]. Three datasets (Yelp [26], WE [27], and RT [28]) are developed for sentiment analysis in English text. Yelp corpus contained the restaurant reviews annotated with positive, negative, neutral and conflict classes. The WE dataset included the wine reviews which are labelled into positive, negative and neutral sentiment classes. The RT dataset consisted of rotten tomatoes movie reviews labelled into positive and negative polarity. Mamta et al. [29] developed a multi-domain tweet corpus where 12,737 tweets are tagged into three classes, namely negative, positive, and neutral.

Although Bengali is a resource-poor language, few small datasets are available for sentiment analysis in Bengali text. Hasan et al. [30] developed a corpus which consider 10,000 Romanized Bangla text labelled into positive, negative, and ambiguous categories. A small dataset developed for sentiment classification where 1400 Bengali tweets were labelled into positive, and negative categories [31]. Few corpus have developed to classify sentiment in Bengali texts into positive or negative polarities in different domains such as, blog comments [32], and translated review datasets [33]. Rahman et al. [34] used cricket Dataset for sentiment analysis in Bengali text where 2900 online comments are labelled into positive, negative, and neutral sentiment. Some small corpus developed for sentiment classification in Bengali text. For example, restaurant review corpus contained 1000 reviews [35], and book review corpus included 2000 reviews [36]. Both corpus labelled into positive and negative sentiment polarities. A recent study developed a dataset for sentiment analysis which contained 2979 Bengali reviews and comments [37]. This dataset was annotated with positive, negative, and neutral polarities.

Corpus for Emotion Classification

Different corpora have been developed to classify emotions in text for various languages. Alm et al. [38] developed a corpus consists of approximately 185 stories which were annotation is performed at the sentence level with neutral, anger-disgust, sadness, fear, happiness, positive surprise, and negative surprise classes. A corpus consists of blog posts that represent 8 emotions label including Ekman's six basic emotions [39]. ISEAR2 [40] is a corpus with joy, fear, anger, sadness, disgust, shame, and guilt classes. SemEval-2007 [41] corpus consists of news headlines in which headlines are annotated with one or more of the following emotions: anger, disgust, fear, joy, sadness, and surprise. Another corpus proposed by SemEval [42] consisting of English, Arabic, and Spanish tweets and these are labelled eleven emotions. A recently developed corpus consists of textual dialogues where each conversation is either labelled as joy, anger, sadness, or others [43]. Oramas et al. [44] developed an English emotion corpus using feedback from students during learning process in which the feedbacks were tagged with emoticons by the students themselves. Data from an English TV series and chatlogs are annotated with emotion tags using a web interface [45]. A tweeter corpus containing Hindi-English code mixed text was annotated with Ekman's six basic emotions by two language proficient and the quality was validated using Cohen's Kappa coefficient in [46]. Two datasets (deISEAR and enISEAR) are used by Enrica et al. [47] where 1001 event descriptions are annotated into seven emotion classes such as, sad, anger, fear, disgust, joy, guilt, and shame. An Arabic tweets dataset developed with 5600 tweets for classification emotions into sad, fear, joy and anger [48].

Although the analysis of emotion in Bengali text is in its infancy, some small datasets are developed to classify emotions into different classes. Das et al. [49] approached with semi-automated word level annotation using Bangla Emotion word list, which was translated from English Word-Net Affect lists [50]. The authors focused heavily on words rather than the context of the sentences and build an emotion dataset of 1300 sentences with six classes. Das et al. [51] tracked emotions Bangla blogs using Bangla wordnet affect list. Prasad et al. [52] tagged sentiments of tweet data based on the emoticon and hashtags. They build a sentiment dataset having 999 Bengali tweets. Nafis et al. [53] developed a Bangla emotion corpus containing 2000 Youtube comments which annotated with Ekman's six basic emotions. They measured majority vote for final labelling and came up tagging with four emotion labels as some of them are ambiguous. Ruposh et al. [54] developed a dataset for analyzing six emotions (such as, joy, sad, surprise, angry, disgust, and fear) in Bengali text. This dataset contained 1200 text data from online sources.

Most of the previous corpora were developed to classify Bengali text into three sentiment polarities: positive, negative, and neutral. Moreover, these datasets are small and hence not suitable to use in machine learning or deep

Table 1 Commonly used keywords in Bengali emotion expression

Emotion class	Keywords	English Equivalent
Anger	রাগ, মেজাজ, শালা	Anger, Mood, Damn it
Fear	ভয়, নিরিজন, আঁতকে	Fear, Lonely, Frightened
Surprise	অবাক, আশ্চর্য, হঠাৎ	Surprised, Wondered, Suddenly
Sadness	কষ্ট, কান্না, খারাপ	Pain, Cry, Bad
Joy	সুন্দর, ভালোবাসা,আনন্দ	Beautiful, Love, joy
Disgust	বিরক্ত, বাজে, ফালতু	Annoyed, Bogus, Nonsense

learning based emotion classification model with reasonable accuracy. To the best of our knowledge, there is no publicly accessible corpus in Bengali language dedicated for analyzing emotion. In our work, we introduce a larger corpus which can be used by machine learning model for classifying emotion in Bengali text to tag each text into six basic emotion classes, namely sadness, joy, anger, disgust, surprise, and fear. This is the very first endeavour towards developing a benchmark corpus for emotion analysis in Bengali text.

## Properties of Emotion in Bengali Texts

We investigated several text expressions in the Bengali language to identify the distinguishing characteristics of emotion classes defined by Ekman [55]. In this work the considered classes are: anger, fear, surprise, sadness, joy and disgust. To identify the distinct characteristics of each emotion in Bengali texts, several factors are considered such as keywords, intensity of emotion word, semantic of sentence, emotion engagement, think like the person.

- *Emotion Keywords:* we identified words commonly used in the context of a particular emotion. For example, the words, "joy", "enjoy", "pleased" are considered as seed words for the joy category. Thus, some specific seed words are stored for a specific emotion in Bengali. For example, "রাগান্বিত" (Angry) or "ক্রোধ" (Anger) is usually used for expressing "Anger" emotion. Likewise, "খুশি" (joy) or "মন ভাল" (Good mood) are usually used for expression "joy" emotion. Table 1 shows some commonly used emotion keywords in the context of a particular emotion.
- *Intensity of Emotion Word* In Bengali, different seed words express different emotions in a particular context. In such cases, seed words are compared in terms of intensity and choose the highest intensity seed word, including its emotion class is assigned for the emotion of that context. Consider the following example, "আলেকজান্ডারের মৃত্যুর সংবাদ যখন অ্যাথেন্সে পৌঁছুলে, তখন একজন অবাক হয়ে প্রশ্ন করলে,"আলেকজান্ডার মৃত! অসম্ভব! তিনি মারা গেলে পৃথিবীর প্রতিটা কোণা থেকে তার মৃত লাশের গন্ধ ভেসে আসত।" (*English translation*: when the news of Alexander's death reached at Athens, someone was surprised and asked, "Alexander is dead! Impossible! If he's dead, the smell of his dead body would waft from every corner of the earth.") In these texts, several key words are existed such as, "মৃত্যুর সংবাদ" (death news), "অবাক" (surprised) and "মৃত! অসম্ভব!" (dead! Impossible!) Here the words, "অবাক" (surprised) and "মৃত! অসম্ভব!" (dead! Impossible!) have more weight than "মৃত্যুর সংবাদ" (death news). Thus this type of texts can be considered as "surprise" because the intensity of this emotion is higher than the intensity of sadness emotion.

- *Semantic of Sentence:* observing the semantic meaning of the texts is one of the prominent characteristics of ascertaining emotion class. In the previous example, though the sentence started with death news of Alexander, this sentence turns into astonishment of a regular person in Athens. So, sentence semantics make an essential parameter in designating emotion expression.

- *Emotion Engagement:* it is imperative to involve the annotation actively while reading the text for understanding the semantic and context of the emotion expression explicitly. For example, "সেন্টমার্টিনে কাটানো প্রতিটি মুহূর্ত অসাধারণ ছিলো। অসংখ্য সে মুহূর্ত থেকে ক্যামেরা বন্দী কিছু মুহূর্ত" (*English translation*: Every moment spent in St. Martin was awesome. Here are some from those countless moments captured on camera). In this particular expression, annotators can feel some happiness as it describes an original moment of someone's experience. This feeling causes annotators engaged with happiness, and the expression designated as "joy".

- *Syntactic Structure:* sometimes, a syntactic structure plays a vital role during annotation. Let us consider two examples, "কে বলেছে তোমাকে এই কাজ করতে? অনেক কাবিল হয়ে গিয়েছ তাই না?" (English translation: Who the hell told you to do this? You have become very expert or what? Apologize to him now.) Another one, "তোমাকে এই কাজ কে করতে বললো? এমন কাজ তুমি করতে পারলে!" (*English translation*: Who told you to do this? How could you do that!) By investigating these sentences, it is found that both the sentences consist of similar words, but their syntactic structures are different. The first example is like, someone encountered with rage but the second one with astonishment. Thus, annotators label the first sentence as "Anger" and the second as "Surprise".

- *Think Like The Person (TLTP):* usually, an emotion expression is a type of expression of someone's emotion in a particular context. By TLTP, an annotator imagines

Fig. 1 Schematic process of BEmoC development
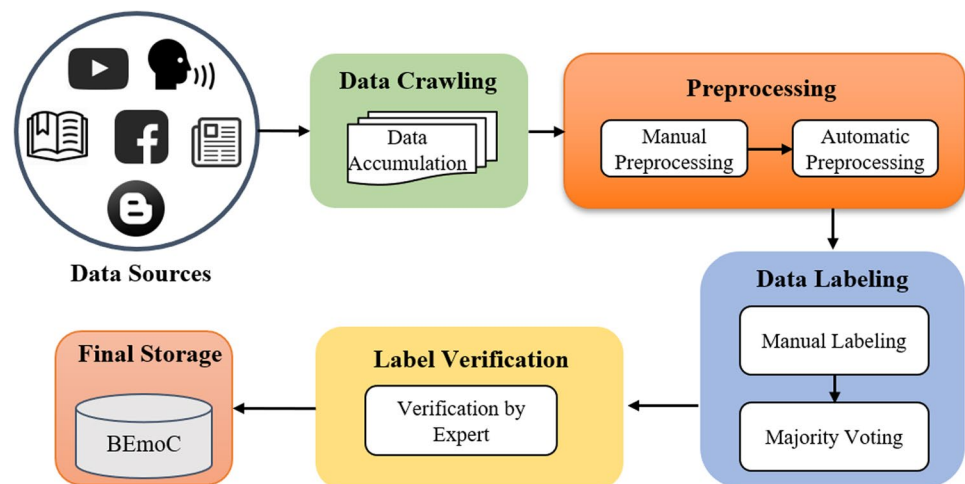
him/her in the same context where the emotion expression displayed. By repeatedly uttering, an annotator tried to imagine the situation and annotated the emotion class.

By taking into consideration the above characteristics, each text will be annotated into one of the six emotion classes: joy, sadness, anger, disgust, surprise, and fear. Table 1 illustrates the commonly used emotion keywords in Bengali text expression.

BEmoC Development

The key objective of our work is to develop a Bengali emotion corpus which is described in the following sub- sections. It can be further used for emotion analysis and classification purpose. Figure 1 shows the overview of the development process of BEmoC which consists of four major phases: data crawling, prepossessing, data labelling, and label verification, respectively.

Data Crawling

Bengali text data were harvested from several sources such as Facebook comments/posts, YouTube comments, online blog posts, Bengali story books [56–59], Bengali novels [60–62], text conversations, and newspapers. Five participants were assigned to accumulate data. They manually collected 7125 text expressions over a three-month period (September 10, 2019–December 12, 2019).

Preprocessing

Preprocessing performed in two phases: manual and automatic. In the manual phase, "typo" errors were eliminated from the collected data. We took Bangla academy supported accessible dictionary (AD) database [63] to find the appropriate form of a word. If a word existed in the input texts but not in AD, then this word considered to be a typo word. The appropriate word searched in AD and the typo word replaced with this corrected word. For example, the text, "জাহাজ এই প্রথমবারের মতনে ওঠা এবং সাগরের মাজখানে দিয়ে যাওয়ার সময়গুলনে ম্যজকিাল ছলিনে একদম। আহা সৌন্দর্য ❤❤❤". In this example, the bold words indicates the typo errors that need to be corrected by using AD. After replacing, the sentence turned into "জাহাজ এই প্রথমবারের মতনে উঠা এবং সাগরের মাঝখানে দিয়ে যাওয়ার সময়গুলনে ম্যাজিকাল ছলি একদম। আহা সৌন্দর্য❤❤❤"

It has been observed that emojis and punctuation marks sometimes create perplexity about the emotional level of the data. That's why in the automatic phase, the emojis were eliminated from the manually processed data. We made an emoji to the hex (E2H) dictionary from [64]. Further, all the elements of E2H were converted to Unicode to cross-check them with our corpus text elements. A dictionary introduced which contains punctuation marks and special symbols (PSD). Assume any text element matched with elements in E2H or PSD substituted with blank space. All the automatic pre- processing was done with a python-made script. After automatic preprocessing, the above example comes into "জাহাজ এই প্রথমবারের মতনে উঠা এবং সাগরের মাঝখানে দিয়ে যাওয়ার সময়গুলনে ম্যাজিকাল ছলি একদম আহা সৌন্দর্য" Although most of the data collected from on-line sources, few data also created by observing people's conversations. In social media, many Bengali native talkers wrote their comments or posts in the form of transliterated Bengali. For example, a transliterated sentence, "muvita dekhe amar khub valo legeche. ei rokom movi socharacor dekha hoy na." (Bangla: "মুভিটা দেখে আমার খুব ভাল লেগেছে। এই রকম মুভি সচারাচর দেখা হয় না" [English translation: I really enjoyed watching this movie. Such movies are not

Table 2 Discarded sentences during preprocessing

No	Sentence	Reason to discard
1	তিনি দ্রুত এসে গেলেন (He came quickly)	Short in length
2	সরি দোস্ত, আজকে একটা impor-tant meeting ছিল (Sorry my friend, today I had a very important meeting)	Mixed language (Bengali-English)
3	ট্রেনে খুব দ্রুত চলাচল করে কিন্তু বাস অনেক ধীরে চলে (Trains run fast but buses run slower)	Neutral emotion

commonly seen]. This type of texts demands to be converted phonetically by the phonetic conversion. However, errors may take place during phonetic conversion. For instance, in the above texts, the word "socharacor (English: usually) could be translated in Bengali as, "সছারাচর" after phonetic conversion whereas the accurate word should be "সচরাচর" Therefore, correction should handle because there is no such word like, "সছারাচর" in Bengali Dictionary [63].

To convert the numeric values into the text form, we designed a converter. For example, the sentence "আমি তাকে ১০ বার কল দিলাম একবারও রিসিভ করল না" (I have called him **10** times but he did not receive for once) has numerical values. So after conversion, it became "আমি তাকে দশ বার কল দিলাম একবারও রিসিভ করল না" (I have called him **ten** times but he did not receive for once). Raw crawled dataset includes numerous unrelated texts. To decrease annotation efforts for such unrelated texts, we developed a program to obtain only related texts. As data containing less number of words has less chance to contain emotion; thus, data with less than five words are alleviated automatically. The dataset collected from the numerous sources which can contain duplicate data. Data duplication avoided by inspecting the dataset automatically. The resources are available from here.[1] Texts containing mixed language and neutral emotion is discarded from the dataset by hand. Table 2 shows the few examples of the texts along with their reason to discard.

Data Labelling

The whole corpus was labelled manually followed by majority voting to assign the suitable label. The initial labelling or annotation tasks performed by 5

undergraduate students having a Computer Engineering background. Group members were instructed to label the text without being prejudiced towards any specific demographic region, customs, and religion. Initial label of the 7125 processed texts are determined by following Algorithm 1.

Algorithm 1: Majority Voting and Initial Label

T ← Text corpus;
AL ← Label of annotators;
IL ← Initial labels;
CL ← Label counter for each class;
for $t_j \in T$ **do**
  CL = {0} // Initialization of counter array
  for $a_i \in AL$ **do**
    CL[a_i]++;
    i++;
  end
  idx=0, max_val = CL[idx];
  for $(k = 0; k \leqslant 5; k + +)$ **do**
    if $(CL[k]>max\_val)$ **then**
      idx = k;
      max_val=CL[k];
    end
  end
  IL[j] = idx; //idx holds the label which got maximum count
  j++;
end

For each text t_j in the corpus annotators label are counted. a_i indicates the ith annotator label for jth text in the corpus. An annotator label a text with an integer from 0 to 5. Here, 0, 1, 2, 3, 4, 5 represents the anger, fear, surprise, sadness, joy and disgust classes, respectively. Label of a text is decided by majority voting [65]. The counter array (CL) holds the annotators vote count of each class for jth text. By using a loop we check the count for all labels. The label which got maximum vote from the annotators is selected as initial label for the jth text. Table 3 shows the annotation example of each emotion classes with their meaning.

Label Verification

An expert who is an academician and working on NLP for several years manually verified each labelling done by the annotators. Initial label of a text assigned by the annotators considers as the ultimate if its label matches with the expert's label. Cross validation was performed by following Algorithm 2.

---

[1] https://github.com/avishek-018/BEmoC-Bengali-Emotion-Courpus/tree/main/preprocessing.

Table 3 Sample annotated texts in BEmoC

Emotion Class	Meaning	Text Example
Anger	A text is consider as anger, if it express unfairness that is happening to us or others [55]	আকাশ থেকে পড়ার ভান ধরনে না! তোমার চাহনির মধ্যেই সব জানা যায়! তুমি ওর গায়ে কয়বার হাত দিছি? (Do not pretend to fall from the sky! Everything is known in your gaze! How many times have you touched her?)
Fear	A text express fear, when it shows some threats or danger to our safety or existence [55]	বোন করোনা ভাইরাস ধরেছে আজরাইল যখন আসবে তখন কি করবেন? আল্লাহকে ভয় করুন আমিন (Sister, Corona virus has been detected. What will you do when Azrael comes? Fear Allah. Amen)
Surprise	A text consider as surprise if it express an unexpected or astonishing event, happened suddenly [55]	আমরা ড্রয়িং রুমে বসে আছি মেয়ে দেখার জন্য। কিছুক্ষণ পর মেয়ে আসলো। মেয়েকে দেখেই কয়েক মিনিটের জন্য আমি হতভম্ব হয়ে গেলাম। একটা মেয়ে কি করে এত সুন্দর হতে পারে। (We are sitting in the drawing-room to see the girl. After a while, the girl came. I was stunned for a few minutes to see the girl. How can a girl so beautiful.)
Sadness	Text can consider as sadness category if it expresses a response to loss, and feeling sad helps us to convince others that we need help [55]	একটা ব্যাপার খারাপ লাগছে। সবাই তো কান্নাকাটি করতেছে। ছুটির দিনে এমন বীভৎস খবর শুনতে কারোরই ভালো লাগার কথা না (One thing seems bad. Everyone is crying. Nobody likes to hear such horrible news on holidays
Joy	A text categorize as joy when it express positive feeling which derives from both novel and familiar experiences [55]	শুভ জন্মদিন সাউথের অন্যতম সুন্দরী অভিনেত্রী অনুপমা পরমেশ্বর (joy Birthday Anupama Parameshwar, one of the most beautiful actresses in the South)
Disgust	A text can consider as Disgust category while it contains of disappointed feeling about something terrible or unpleasant happens, physically or mentally [55]	দাঁত থাকতে দাঁতের মর্যাদা দিতে জানিনা আমরা হুমায়ূন আহমেদ স্যারকে হারিয়ে বুঝতে পারছি উনি কি ছিলি। ওপারে ভাল থাকবেন স্যার, আল্লাহ আপনাকে জান্নাতে নসিব করুক (Blessings are not valued until they are gone. We have lost Humayun Ahmed. Now we can understand what he was. Sir, stay well on the other side, may Allah bless you with Paradise)

Algorithm 2: Cross validation between expert and annotators

```
T ← Text corpus;
A ← Initial annotations;
E ← Expert's annotations;
Sz = Size of the corpus;
for tᵢ ∈ T do
    if Aᵢ = Eᵢ then
        Agree++;
    end
    i++;
end
Similarity = (Agree/Sz)*100%;
```

For the purpose of cross validation, initial annotation (A_i) matched with expert annotation (E_i) for the text t_i. If both annotation matches then total agreement is increased. Finally, we get the similarity score by taking the ratio between number of agreements and size of the corpus. After the cross-validation, it is observed that among 7125 data labels a total of 6945 data labels are directly matched with the expert labels. From the remaining 180 data labels, the expert rectify 55 data labels due to the omission occurred during initial labelling of annotators. Rest of 125 data labels are precluded as the experts did not find the appropriate labels. This exclusion happened due to the texts with neutral emotion, implicit emotion and ill-formatted. Finally, BEmoC included verified 7000 data, including their labels and saved in *.xlsx format. The errors occurred in the dataset development mainly due to the annotator's mislabelling. The expert corrected the mislabelled data annotation during the verification phase. Some examples of the discarded data and rectified labels with proper reasoning are mentioned in Table 4.

Analysis of BEmoC

Corpus analysis was performed by determining the data distributions concerning source and emotion classes. Several statistics of BEmoC also presented.

Table 4 Discarded sentences after verification

Discarded Samples	Initial Label	Expert Label	Reason
একটা নীতিতে থাকনে ভাই, শুধু বর্তমান টাই দেখেনে দুইদনি আগরে কথাই এত দ্রুত কীভাবে ভুলে যান!! (You have only one policy, brother, just look at the present, how can you forget the words of two days ago so fast!!)	Anger	Accepted	–
রাস্তার ট্রাফকি লাইটটি আবার লাল থেকে নীল হলো। (The street traffic light went from red to blue again.)	Surprise	Discarded	Neutral emotion
সে হাসতে হাসতে কেঁদে ফেলল, এ যেন সুখরেই কান্না। (She laughed and cried, as if it were a cry of happiness.)	Sadness	Discarded	Implicit emotion
শুনে খুব ভাল লাগলো মহামান্য রাষ্ট্রপতি। ধন্যবাদ (It was very good to hear from you my Excellency the President. Thank you.)	Joy	Accepted	-
আজাইরা পেছাল পারতে ভাল লাগেনা যতসব বাজে খবর সময়রে কন্ঠস্বর। (I don't like to talk at random, all the stupid news of Somoyer Konthoshor)	Disgust	Anger	Semantic of sentence
আরে এইত সেই মহলিা যে বলছলিনে খুশিতে ঠেলায় ভেণ্টে দতিে আসছে! (Isn't she is that woman who came to vote for happiness!)	Joy	Surprise	Intensity of emotion word

Fig. 2 Data collection sources
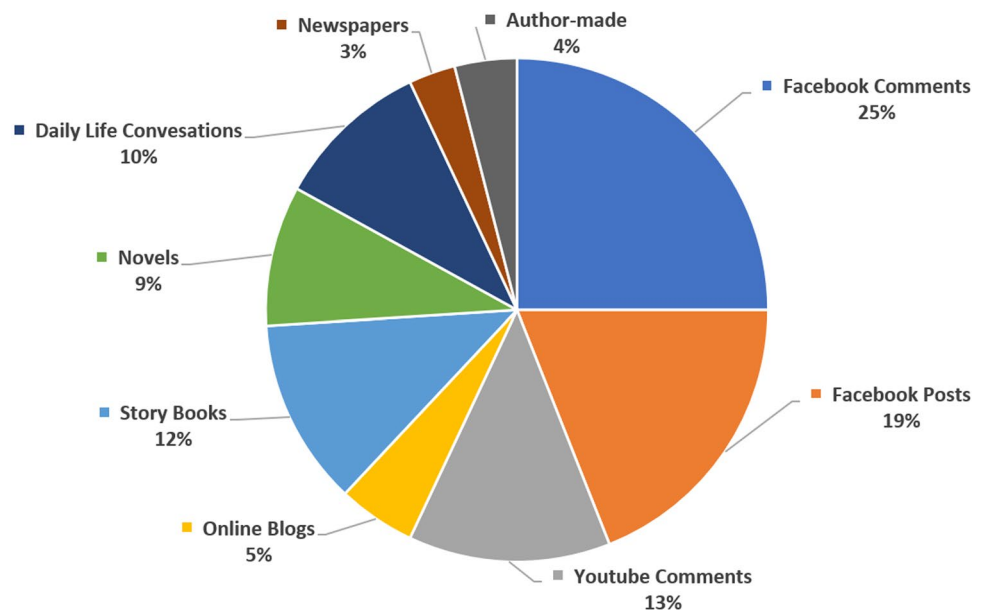
Source-wise Distribution of Emotion Text

Bengali text data are collected from several sources. Figure 2 represents the proportion of collected data as per their sources. Majority amount of the data (about 65%) collected from the online sources. For example,

25% of the data were collected from the Facebook comments and 19% data were accumulated from the Facebook posts. Youtube comments, online newspapers and blogs contributed 13%, 3% and 5% respectively. On the other hand, the offline sources contained 35% amount of data

into the corpus. A significant proportion of data comes from the story books and novels (12% & 10%) whereas the authors contributed only 4% data and 9% accumulated from the daily conversation.

Statistics of BEomC

Table 5 illustrates the critical characteristics of the developed BEmoC, which consists of 148491 words in total and 25458 unique words under 11973 sentences.

Table 5 Data statistics of BEmoC

Corpus Attributes	Attributes Value
Total data in BEmoC	7000
Size on disk	2340 KB
Total number of sentences	11,973
Total number of words	148,491
Unique words	25,458
Maximum words in a single Data	213
Minimum words in a single Data	6
Average words in a single Data	22 (approximately)

Table 6 Data statistics by categories

Category	No. of Data	No. of Sentences	No. of Words	No. of Unique Words
Anger	1052	1588	21,988	7418
Fear	962	1718	20,319	7079
Surprise	779	1533	17,926	7222
Sadness	1327	2346	30,603	9086
Joy	1524	2534	30,298	9102
Disgust	1356	2254	27,357	8421
Total	7000	11,973	148,491	48,328

Categorical Distribution of Emotion Text

The expression is labelled one of the six basic emotion categories after the validation process. Table 6 shows the categorical summary in BEmoC. It is observed that the highest number of data points belong to the Joy, whereas the lowest number of data belong to the surprise category. The entire corpus is publicly available on GitHub.[2]

Metrics for Evaluation

The rendition of emotion clue in the text is highly subjective, which directs to a discrepancy in the annotations by distinct annotators. The difference in experiences, the annotation task itself, and focus on the annotators contribute to a disagreement between the annotators [66]. Thus, we investigate to find how much the annotators agree in assigning emotion classes by using Cohen's kappa [67]. We also measure the coding intervals, the density of emotion words, high-frequency emotion words, and distribution of emotion words with Zipf's law, respectively.

- *Cohen's kappa* (κ): Cohen's kappa measures the agreement between two annotators, each classifying N items into C categories which are mutually exclusive. This coefficient determines how well one annotator agrees with another annotator. If the annotators are in complete agreement then $\kappa = 1$ but if they disagree, $\kappa \leq 0$. The κ coefficient defined by the Eq. 1.

$$\kappa = \left(p_o - p_e\right) / \left(1 - p_e\right), \tag{1}$$

where p_o represents relative observed agreement which measured by the ratio of number in agreement and the Eq. 2 determines the hypothetical probability of chance agreement.

$$p_e = 1/N^2 \sum_k n_{k1} n_{k2}, \tag{2}$$

k is the number of categories; N is the number of observations to categorize and n_{ki} is the number of times that the annotator i predicted category k.

- *Coding reliability* As BEmoC a newly proposed corpus, it required to ensure the quality. To ensure the validity of the coding scheme intercoder reliability is calculated using Eq. 3.

$$A = \frac{M}{\frac{\sum_{i=1}^{n} N_i}{n}}, \tag{3}$$

where A denotes an agreement; M indicates the number of coding events agreed by all coders; N_i denotes the number of coding events assigned by ith coder, and n represents the number of coders.

- *Parts of speech analysis:* in linguistic corpus, Parts of Speech (POS) Tagging, also known as grammatical tagging, is a process to label tokens with their lexical categories [68]. POS tagging is different in different languages. Many POS taggers used in the corpus named N-gram, Hidden Markov Model (HMM), Conditional Random Fields (CRF), and Brill's Tagger. From these taggers, Brill's Tagger is most reliable with high accuracy and is used to tag our Bemoc.

- *Density of emotion words:* the utilization of emotion words could reflect the degree to which the writer released emotions. To measure the influences of emotion words in various classes, we incorporate the density of negative words [69]. Density computed by the ratio of the number of emotion words to the total number of words in each class Eq. 4.

$$\text{Density } D = \frac{\text{No. of emotion words}\,(N)}{\text{Total words in each class}\,(T)}. \tag{4}$$

Table 7 Kappa statistics

Kappa Metrics	Annotators vs. Expert
Number of observed agreements (p_o)	97.47% of the observations
Number of agreements expected by chance (p_e)	16.48% of the observations
Average Kappa (κ)	0.969
SE of average kappa	0.002
95% confidence interval	0.965–0.978

Table 8 Pair-wise Kappa score for each emotion class

Pairs	Emotion categories					
	Anger	Disgust	Sadness	Fear	Joy	Surprise
E-A1	0.951	0.963	0.961	0.979	0.985	0.975
E-A2	0.953	0.974	0.953	0.987	0.989	0.985
E-A3	0.950	0.969	0.964	0.965	0.982	0.975
E-A4	0.952	0.963	0.972	0.963	0.981	0.968
E-A5	0.961	0.966	0.967	0.960	0.987	0.971
Average on each emotion category	0.953	0.967	0.963	0.971	0.985	0.974

- *The high-frequency emotion words:* high frequency items are those which occurs the most in the corpus. In language processing field, typically high frequent words are derived from each class rather than the whole corpus.
- *Distribution of emotion words obeys Zipf's law:* Zipf's law reveals an empirical observation that states that the frequency of a given the word should be inversely proportional to its rank in the corpus. Zipf curve is considered to be a histogram sorted by word rank, with the most frequent words first [70]. Equation 5 can be used to represent the Zipf's law.

$$f(k : S, N) = \frac{1/k^s}{\sum_{n=1}^{N} (1/n^s)} \tag{5}$$

Evaluation of BEmoC

Out of 7125 data, a total of 7000 data are finally included in BEmoC after the verification process. We performed the cross-validation between the annotators and the expert using Algorithm 2 which achieved a similarity scores 97.47\%. This indicates that there is a high similarity to assign the label of the data by the both annotators and the expert.

To investigate the standard inter-annotator agreement, a pairwise kappa coefficient is calculated using Eq. 1 in order to check the goodness of annotations given by different annotators and the expert. Table 7 shows the average result of Kappa statistics where every row denoted as κ agreement between expert and annotators and every column denoted

as κ agreement on individual emotion categories. It also shows the Kappa statistics between the labelling of expert and majority voting of annotators. According to Landis et al. [71], a kappa score between 0.81 and 1.00 is almost perfect agreement. We achieved an average kappa score of 0.969, which indicates an almost perfect agreement between annotators. Table 7 also shows standard error and 95% confidence interval between the annotators and the expert.

The pairwise kappa scores are summarized in Table 8. Five pairs were constructed having one annotator and the expert in each pair. The table shows the pairwise kappa value for each class as well as the average.

A category wise Kappa is shown in Figure 3, which explicitly shows the joy category data has a more accurate kappa agreement score of 0.985. This category of emotions could easily be identified as it is distinguishable from other emotion categories. The anger category achieved minimum kappa score (0.953) due to its conflict with the category disgust.

Figure 4 shows the inter-coder reliability between G1 and G2 for each emotion category. It is observed that the category joy gained the more accurate inter-coder reliability with 97.4% whereas the anger category achieved the less accurate inter-coder reliability (94.7%). The average coding reliability is 95.91% and total coding reliability of whole BEmoC is 97.8%.

A tagged corpus is consisting of 7168 sentences and with almost 103k tokens used for parts of speech tagset analysis. This corpus named "Indian Language Part- of-Speech Tagset: Bengali" whose primary data source are weblogs, and web collection [72]. There is a tagset with 11 tags used to

Fig. 3 Kappa agreement in each emotion categories
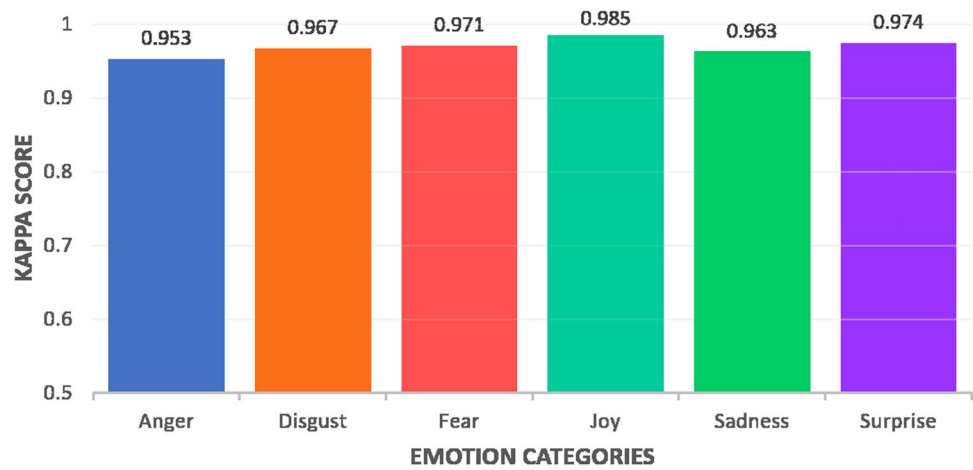
Fig. 4 Coding reliability of BEmoC
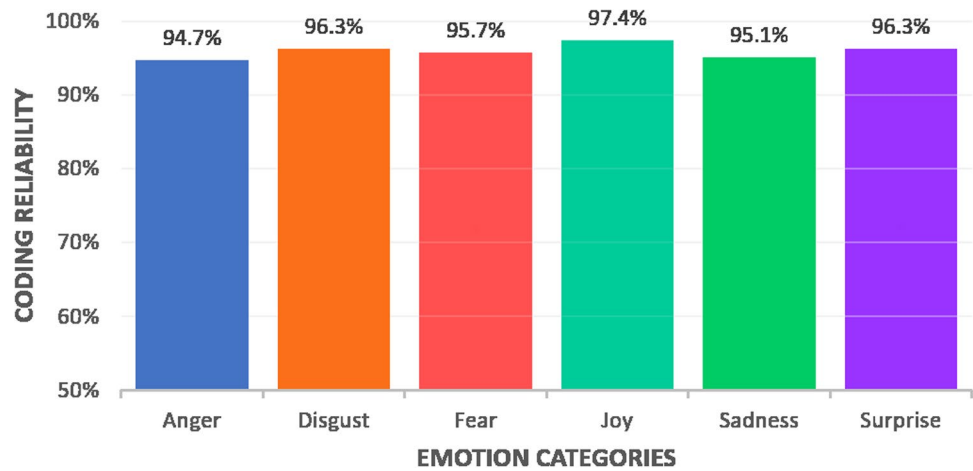
Table 9 POS tagset analysis in terms of number of tokens

Parts of speech	Anger	Fear	Surprise	Sadness	Joy	Disgust
Noun	10,825	8919	8484	13,804	15,190	13,439
Pronoun	1935	1733	1333	2761	2325	2190
Verb	3728	4345	3230	5736	4510	4322
Nominal modifier	2470	2341	2351	3444	4223	2982
Demonstrative	578	490	416	700	679	767
Adverb	686	601	585	1070	665	685
Participle	206	130	92	267	208	231
Post position	661	578	483	831	1022	812
Particle	1376	1076	851	1871	1220	1798
Residual	123	106	101	119	256	131
Punctuation	1588	1718	1533	2346	2534	2254

tag the sentences, e.g., noun, verb, nominal modifier, adverb, and etc. We used this corpus to train Brill's Tagger model [73]. After training the model with the Bengali POS tagset, we prepare the model for evaluation. For analysis, we tagged every sentence and counted the tags of each category of BEmoC. The distribution of number of tokens in POS class is computed and reported in Table 9. Every category roughly has a 40% "Noun" of total tokens. The second highest tagset is "Verb" with roughly 12% of total tokens. The

Table 10 Density of emotion words in each class

Emotion Class	Unique Words (T)	Emotion Words (N)	Density (D)	D-0.2433
Anger	7418	1295	0.1745	− 0.0877
Fear	7079	1360	0.1921	− 0.0701
Surprise	7222	1918	0.2655	− 0.0033
Sadness	9086	2959	0.3256	0.0634
Joy	9102	2741	0.3011	0.0389
Disgust	8421	2652	0.3149	0.0527
Total/ average	48,328	12,925	0.2622	0.0000

third most tagset is "Nominal Modifier" with roughly 11% of total tokens.

The density of emotion words per class shown in Table 10. The overall density in the whole corpus is 0.2622, which represents the average density of emotion words. The average density can be brought into comparison with the density of emotion words for each class. If the density for one class is more significant than 0.2622, it signifies that the writer conveys enhanced emotion concerning this class; it also indicates that emotions focused within this class.

Figure 5 shows the variance of each class density with the average density. The figure indicates that the densities of sadness, joy, and disgust classes are higher than the average density of 0.2622, which reveals that people are more responsive in the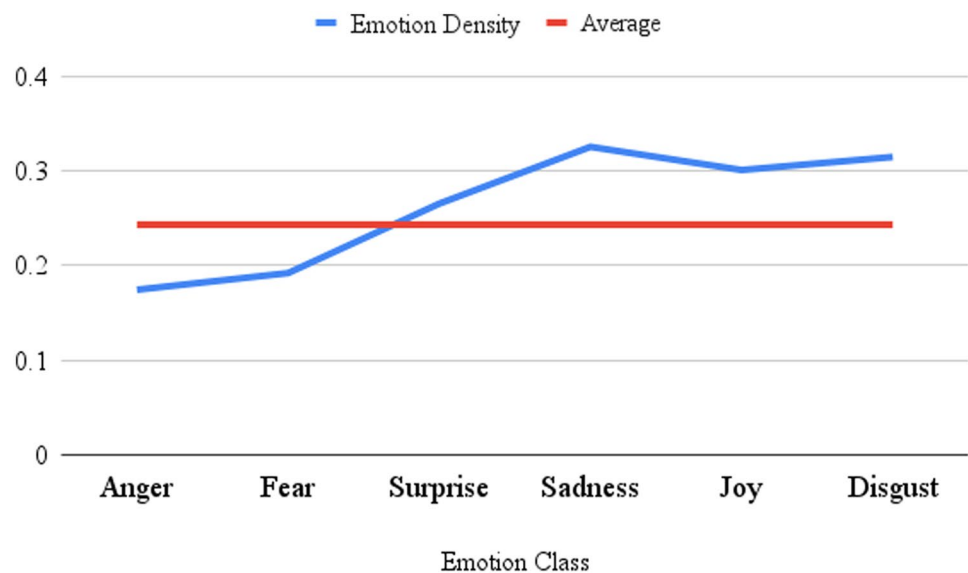se classes and use more emotional words. However, the emotion word densities for some classes are lower than average, such as anger and fear. One of the possible causes is that people utilize more impartial words to express events rather than describing emotion by using emotion words.

The frequency of emotion words was counted in BEmoC. This frequency of emotion words brings to the conclusion that some specific words always meant to express specific emotions of humans. Figure 6 depicts the most frequent emotion words on each emotion category.

We used a data analytic tool (Tableau desktop version) to visualize frequency bubbles where the large bubbles denotes the high frequency words and the small bubbles indicates the relatively low frequency. For example, in Figure 6 the words, "রাগ", "শালা" belongs to Anger class where, "শালা" is a slang word. Likewise, "ভয়", "অবাক", "কষ্ট", "সুন্দর", "বিরক্ত" are the most common and frequent word in Fear, Surprise, Sadness, Joy and Disgust classes, respectively. It is noticed that "ভয়", "অবাক" are the most frequent words over all the classes.

Zipf's law states that if the Zipf curve is plotted on a log-log scale, a straight line with a slope of − 1 must be obtained. Figure 7 shows the resultant graph for each classes. It is observed that the curves obey the Zipf's law as the curve follows a slope of − 1.

Considering all the evaluation measures along with 97.47% similarity score, average κ score of 0.969, and the obeying the Zipf's law, the developed corpus (i.e., BEmoC) can be used to classify basic emotions from Bengali text expressions. This work provides a primary foundation to detect human's six basic emotions by relating Bengali text expressions.

Discussion

While developing the Bengali emotion corpus, we faced several challenges. Few of them are described in the following:

Fig. 5 Emotion word density vs. average density

Fig. 6 Highest frequency emotion words in BEmoC for each emotion: **a** anger, **b** fear, **c** surprise, **d** sadness, **e** joy, and **f** disgust

Perplexity to Assign Emotion Classes

The initial annotation effort suggested that in many instances a sentence was found to exhibit more than one emotion. Moreover, the emotion conveyed in some sentences could not be attributed to any basic category. Based on the observation during development phase, we illustrates the following remarks:

- There is a good confusion between "anger" and "disgust" data. A particular data may represents both anger and disgust at the same time in some cases. Consider an example, "কিছুই হয়নি, ফালতু একটা জিনিস। দুনিয়ার অখাদ্য কুখাদ্য হইছে।। কুষ্টিয়ার ভাষাটাও হয়নি, কিচ্ছু হয়নি, কিচ্ছু না।" (English translation: Nothing happened, such a stupid thing. The food is rubbish. Even the language of Kushtia is not right. Nothing happened. Nothing"). This data was annotated as "Disgust". During annotation, among 6, 4 annotators supported this expression as the disgust category but at the same time, 2 of them annotated as the anger. Therefore, confusion may arise on labelling the data with anger or disgust. Special attention is required to label this types of expression. There is also a logical confusion between the classes surprise, and

joy. Assume an example, "মন ছুঁয়ে গেলো তোর লেখাটা, মৃণালিনীর পাশেই সন্তোষালয়ের বাচ্চা গুলেনে কে দেখে কি যে ভালো লাগলো,তুই বোধহয় ততদিনে পাঠভবনে উত্তীর্ণ হয়েছিস।" (English translation: Your writing touched my heart, It felt so nice to see the children of Santoshalay next to Mrinalini, you probably have passed Pathabhavan by that time). This expression was annotated as the joy. At the time of annotation, 2 out of 6 annotators supported this expression as the "Joy" category whereas 4 participants annotated as the surprise. Therefore, careful attention also required in these instances.

- During development phase, confusion arises for surprise and fear classes. Consider an example, "হিন্দুকুশ পর্বত কিংবা সাহারা মরুভূমি, সবকিছুর চেয়ে এই ইন্দু নদীই তাদের কাছে বেশি ভয়ঙ্কর মনে হচ্ছিলো।" (English translation: The Hindu Kush mountains or the Sahara desert, the Indus river is more dangerous to them than anything else). This expression was annotated as the fear class. Among annotators, 4 were supported this expression as the fear class and 2 annotators were labelled it as the surprise category. Thus, special attention should be provided to annotated this type of expression.

Local and Global Emotional Thesauri

According to [74] local analysis by using context and phrase structure is more effective. In our study, we found that some words such as খারাপ (Bad), সমস্যা (Problem), অসাধারণ (Amazing) are used in almost all the emotion classes, so they are considered as global emotional thesauri. Some words like কষ্ট (Trouble), কান্না (Crying), অসহায় (Helpless) are mentioned explicitly in the sadness class. These are referred to as local emotional thesauri. Thus, further research should conduct to construct emotion thesauri, which helps to understand a person's affectation in by identifying emotion words.

The developed corpus should enlarge with more data to apply in sophisticated classification technique such as deep learning. Data which are more prone to confusion among team members should be removed for suppressing the error rate. Classification of implicit emotion is the most critical problem due to this emotion is typically unapparent within the expression, and thus, its solution demands to interpret the context. Even when a sentence includes an emotion keyword, it is not assured to express the same emotion because the word semantic can alter according to the context. Emotions are complicated; humans often suffered problems to express and understand emotions. Classifying or detecting emotions in the text increases the complicity of interpreting emotion because of the lack of visible facial expressions, body gestures, and voice [2]. Automating emotion recognition is a difficult task. A machine needs to deal with the complexity of linguistics and the context of the written text.

Fig. 7 Zipf's curve plotted on a log–log scale for each emotion: **a** anger, **b** fear, **c** surprise, **d** sadness, **e** joy, and **f** disgust

Other Factors in Developing Emotion Corpus

Several other factors should consider to assign a label of data and include it into an appropriate category of emotion/sentiment corpus. The exposure of subjective and objective texts with their appropriate tone is significant to interpret emotion. All objective texts do not include explicit emotions. Some predicates (i.e., verbs, adjectives, and few nouns) cannot be prescribed the same concerning how they generate emotion. Context of an utterance is a very vital factor in determining its emotion class. Expressions related to irony and sarcasm (where human express their negative emotions

using positive words), can be complicated to determine emotion class without having an absolute perception of the context of the circumstances in which a feeling represented. Text express comparisons in emotion analysis is another complicated difficulty that needs to handle attentively before settling the emotion class. Emojis also play a vital role in the emotion of texts that require to convert into tokens (of appropriate category) in order to enhance the performance of emotion analysis.

Conclusion

Emotion classification is still developing areas of research, and challenges for low-resource languages are daunting. Scarcity of benchmark dataset is one of the vital challenges to perform the emotion classification task in the Bengali language. Thus, in this work, we presented a new corpus (called BEmoC) for classifying the expression of emotion in Bengali texts and explained its development processes in details. We have crawled texts from online and offline sources, applied several prepossessing techniques to clean the data, and annotated the corpus with basic six emotions including joy, fear, anger, sadness, surprise and disgust. To best of our knowledge, this is the first Bengali corpus annotated with six basic emotion classes. The paper revealed several features of emotion texts, especially concerning each different class, exploring what kinds of emotion words human use and why they choose those words to express their particular emotions. The evaluation of BEmoC shows that the developed dataset followed the distribution of Zipf's law and maintained an agreement among annotators with perfect κ score. More data sample can be included in the current BEmoC for its effective use in sophisticated classification techniques. In further research, BEmoC can be considered to annotate emotion expressions in terms of various domains including emotions in emojis, text containing sarcasm/irony or comparative expressions.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Anusha V, Sandhya B. A learning based emotion classifier with semantic text processing. Adv Intell Inf. 2015;320:371–82.
2. Alswaidan N, Menai MB. A survey of state- of-the-art approaches for emotion recognition in text. Knowl Inf Syst. 2020;62:2937–87.
3. Hossain E, Sharif O, Hoque MM, Sarker IH. Sentilstm: a deep learning approach for sentiment analysis of restaurant reviews. In: hybrid intelligent systems: 20th international conference on hybrid intelligent systems (HIS 2020), springer nature, 2020;1375:193–203.
4. Liu B, et al. Sentiment analysis and subjectivity. Handb Nat Lang Process. 2010;2(2010):627–66.
5. Garg K, Lobiyal DK. Hindi emotionnet: a scalable emotion lexicon for sentiment classification of Hindi text. ACM Trans Asian Low-Resour Lang Inf Process. 2020;19(4):1–35.
6. Bengali language. 202. https://en.wikipedia.org/wiki/Bengali_language. Accessed 2 Jan 2020.
7. Das A, Iqbal MA, Sharif O, Hoque MM. Bemod: development of Bengali emotion dataset for classifying expressions of emotion in texts. In: Vasant P, Zelinka I, Weber GW, editors. Intelligent computing and optimization. Cham: Springer International Publishing; 2021. p. 1124–36.
8. Eckman P. Universal and cultural differences in facial expression of emotion. Nebraska Symp Motiv. 1972;19:207–84.
9. Agrawal A, An A. Unsupervised emotion de-tection from text using semantic and syntactic relations. In: IEEE/WIC/ACM Int. J. Conf. on Web Int. & Int. A. Techno., IEEE, Macau, China, 2012;1:346–53.
10. Kim Y, Lee H, Provost EM. Deep learning for robust feature generation in audiovisual emotion recognition. In: 2013 IEEE International Conference on acoustics, speech and signal processing, IEEE, 2013; p. 3687–691.
11. Steidl S. Automatic classification of emotion related user states in spontaneous children's speech. University of Erlangen-Nuremberg Erlangen; 2009.
12. Busso C, Bulut M, Lee CC, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS. Iemocap: Interactive emotional dyadic motion cap- ture database. Lang Resour Eval. 2008;42(4):335.
13. Tao J, Fang X. Toward multi-label sentiment analysis: a transfer learning based approach. J Big Data. 2020;7:1. https://doi.org/10.1186/s40537-019-0278-0.
14. Al-Smadi M, Qawasmeh O, Talafha B, Quwaider M. Human annotated arabic dataset of book reviews for aspect based sentiment analysis. In: Int. Con. on Fu. IoT & Clo., IEEE, Rome, Italy, 2015; p. 726–30.
15. Ales T, Ondrej F, Katerina V. Czech aspect- based sentiment analysis: A new dataset and preliminary results. In: ITAT, 2015; p. 95–99/
16. Apidianaki M, Tannier X, Richart C. Datasets for aspect-based sentiment analysis in French. In: Int. Con. on Lan. Res. & Eva., ELRA, Portorož, Slovenia, 2016; p. 1122–126.
17. Pontiki M, Galanis D, Pavlopoulos J, Papageor- giou H, Androutsopoulos I, Manandharet S. Semeval-2014 task 4: aspect based sentiment analysis. In: Int. Work. on Sem. Eva., ACL, Dublin, Ireland, 2014; p. 27–35.
18. Pontiki M, Galanis D, , Papageorgiou H, An- droutsopoulos I, Manandharet S, AL-Smadi M, Al- Ayyoub M, Zhao Y, Qin B, Clercq OD, Hoste V, Apidianaki M, Tannier X, Loukachevitch N, Kotel- nikov E, Bel N, , Jimenez-Zafra SM, , Eryiğit G. Semeval-2016 task 5: Aspect based sentiment analysis. In: Int. Work. on Sem. Eva., ACL, San Diego, California, 2016; p. 19–30.
19. Bosco C, Patti V, Bolioli A. Developing corpora for sentiment analysis: the case of irony and SentiTUT. IEEE Intell Syst. 2013;28(2):55–63.
20. He R, McAuley J. Ups and downs: Modeling the visual evolution of fashion trends with one- class collaborative filtering. In: Proceedings of the 25th International Conference on world wide web, 2016; p. 507–17.
21. Blitzer J, Dredze M, Pereira F (2007) Biographies, Bollywood, boom-boxes and blenders: Do- main adaptation for sentiment

classification. In: Proceedings of the 45th Annual Meeting of the association of computational linguistics, 2007; p. 440–47.

22. Imdb. 2020. http://www.imdb.com/. Accessed 5 Mar 2020.

23. Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng AY, Potts C. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 conference on empirical methods in natural language processing, 2013; p. 1631–642.

24. Nabil M, Aly M, Atiya A. Astd: Arabic sentiment tweets dataset. In: Proceedings of the 2015 Conference on empirical methods in natural language processing, 2015; p. 2515–519.

25. Donghang P, Jingling Y, Lin L, Deming S. Deep neural network-based classification model for sentiment analysis. In: Int. Con. on Behav., Eco. & Socio-Cult. Com., IEEE, Beijing, China, 2019; p. 1–4.

26. The yelp restaurant reviews. 2004. https://www.yelp.com/dataset/. Accessed 5 Mar 2020.

27. Wine reviews. 2017. https://www.kaggle.com/zynicide/wine-reviews. Accessed 5 Mar 2020.

28. Movie reviews. 2018. https://www.kaggle.com/ayushkalla1/rotten-tomatoes-movie- database/. Accessed 5 Mar 2020.

29. Mamta, Ekbal A, Bhattacharyya P, Srivastava S, Kumar A, Saha T. Multi-domain tweet corpora for sentiment analysis: resource creation and evaluation. In: Proc. 12th Lang. Res. & Eva. Conf., ACL, Marseille, France, 2020; p. 5046–54.

30. Hassan A, Amin MR, Al Azad AK, Mohammed N. Sentiment analysis on Bangla and Romanized Bangla text using deep recurrent models. In: 2016 International Workshop on Computational Intelligence (IWCI), IEEE, 2016; p. 51–56.

31. Chowdhury S, Chowdhury W. Performing sentiment analysis in bangla microblog posts. In: Int. Con. on Inf., Elec. & Vis., IEEE, Dhaka, Bangladesh, 2014; p. 1–6.

32. Islam S, Islam A, Hossain A, Dey JJ. Super- vised approach of sentimentality extraction from Bengali Facebook status. In: Int. Con. on Com. & Info. Tech., IEEE, Dhaka, Bangladesh, 2016; p. 383–87.

33. Paul AK, Shill PC. Sentiment mining from Bangla data using mutual information. In: 2016 2nd International Conference on electrical, computer telecommunication engineering (ICECTE), 2016; p. 1–4, https://doi.org/10.1109/ICECTE.2016.7879569

34. Rahman A, Dey EK. Datasets for aspect- based sentiment analysis in Bangla and its baseline evaluation. Data. 2018;3(2):15–25.

35. Sharif O, Hoque MM, Hossain E. Sentiment analysis of Bengali texts on online restaurant reviews using multinomial Naïve Bayes. In: Int. Con. on Adv. in Sci., Eng. & Rob. Tech., IEEE, Dhaka, Bangladesh, 2019; p. 1–6.

36. Hossain E, Sharif O, Hoque MM (2020) Sentiment polarity detection on Bengali book reviews using multinomial naive Bayes. arXiv preprint arXiv: 200702758

37. Sharmin S, Chakma D. Attention-based convolutional neural network for Bangla sentiment analysis. AI & Soc. 2020;36:381–96.

38. Alm OC, Roth D, Richard S. Emotions from text: Machine learning for text-based emotion pre- diction. In: Conference on human language technology and empirical methods in natural language processing, ACL, Vancouver, British Columbia, Canada, 2005; p. 579–86.

39. Aman S, Szpakowicz S. Identifying expressions of emotion in text. In: Matoušek V, Mautner P, editors. Text Speech and Dialogue. Berlin: Springer; 2007. p. 196–205.

40. Scherer KR, Wallbott HG. Evidence for universality and cultural variation of differential emotion response patterning. J Pers Soc Psychol. 1994;66(2):310–28.

41. Strapparava C, Mihalcea R. Semeval-2007 task 14: affective text. In: Int. Work. on Sem. Eva., ACL, Prague, Czech Republic, 2007; p. 70–7.

42. Mohammad S, Bravo-Marquez F, Salameh M, Kiritchenko S. Semeval-2018 task 1: affect in tweets. In: Int. Work. on Sem. Eva., ACL, New Orleans, Louisiana, 2018; p. 1–17.

43. Chatterjee A, Narahari KN, Joshi M, Agrawal P. SemEval-2019 task 3: EmoContext con- textual emotion detection in text. In: Proceed- ings of the 13th International Workshop on Se- mantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019; p. 39–48, DOI https://doi.org/10.18653/v1/S19-2005, https://aclanthology.org/S19–2005.

44. Oramas-Bustillos R, Barron-Estrada ML, Zatarain-Cabada R, Ramírez-Ávila SL. A corpus for sentiment analysis and emotion recognition for a learning environment. In: 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), IEEE, 2018; p. 431–35.

45. Chen SY, Hsu CC, Kuo CC, Ku LW, et al. Emotionlines: an emotion corpus of multi-party conversations. arXiv preprint arXiv: 180208379. 2018.

46. Vijay D, Bohra A, Singh V, Akhtar SS, Shrivastava M. Corpus creation and emotion prediction for Hindi-English code-mixed social media text. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop, 2018; p. 128–35.

47. Troiano E, Padó S, Klinger R. Crowdsourcing and validating event-focused emotion corpora for German and English. In: 57th Annual Meeting of the Asso. for Com. Ling., ACL, 2019; p. 4005–11.

48. Baali M, Ghneim N. Emotion analysis of Arabic tweets using deep learning approach. J Big Data. 2019;6:1–2.

49. Das D, Bandyopadhyay S. Word to sen- tence level emotion tagging for bengali blogs. In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, 2009; p. 149–152

50. Strapparava C, Valitutti A, et al. Wordnet affect: an affective extension of wordnet. In: Lrec, Citeseer, 2004; 4: 40.

51. Das D, Roy S, Bandyopadhyay S. Emotion tracking on blogs-a case study for Bengali. In: International Conference on industrial, engineering and other applications of applied intelligent systems, Springer, 2012;7345:447–56.

52. Prasad SS, Kumar J, Prabhakar DK, Tripathi S. Sentiment mining: An approach for Bengali and Tamil tweets. In: 2016 Ninth International Conference on Contemporary Computing (IC3), IEEE, 2016; p. 1–4.

53. Tripto NI, Ali ME. Detecting multilabel sentiment and emotions from Bangla Youtube comments. In: 2018 International Conference on Bangla Speech and Language Processing (ICB- SLP), IEEE, 2018; p. 1–6.

54. Ruposh HA, Hoque MM. A computational approach of recognizing emotion from Bengali texts. In: Int. Con. on Adv. in Ele. Eng., IEEE, Dhaka, Bangladesh, 2019; p. 570–74.

55. Ekman P. Basic emotions. Handbook of cog- nition and emotion. 1999;98(45–60):16.

56. Bandyopadhyay B. Aam Atir Vepu. Signet Press; 2017.

57. Zaman S. Mamlar Shakkhi Moyna Pakhi. Prothoma Prokashon, 1st Printed, 2019.

58. Bandyopadhaya M. Manik Bandyopadhaya Sherstha Galpo. ABOSOR; 2011.

59. Ray S. Feluda Samagra Vol. I & II. Ananda Publishers; 2014.

60. Chattopadhyay BC. Bankim Rachanabali (Upanyas Samagra). Tuli-Kolom; 1986.

61. Basu B. Tithidor. Ajkal Prokashoni. Reprinted, 2016.

62. Bandopadhyay M. Putul Nacher Itikatha. CreateSpace Independent Publishing Platform; 2018.

63. Accessible dictionary. 2020. https://accessibledictionary.gov.bd/. Accessed 5 Mar 2020.

64. Full emoji list. 2020. https://unicode.org/emoji/charts/full-emoji-list.html. Accessed 5 Mar 2020.

65. Magatti D, Calegari S, Ciucci D, Stella F. Automatic labeling of topics. In: 2009 Ninth International Conference on intelligent systems design and applications, IEEE, 2009; p. 1227–232.

66. Passonneau R. (2006) Measuring agreement on set-valued items (masi) for semantic and pragmatic annotation. In: Int'l Conf. on Lan. Res. & Eval., ELRA, Genoa, Italy (2006)

67. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46.

68. Jahara F, Barua A, Iqbal MA, Das A, Sharif O, Hoque MM, Sarker IH. Towards POS tagging methods for Bengali language: a comparative analysis. In: Vasant P, Zelinka I, Weber GW, editors. Intelligent computing and optimization. Cham: Springer International Publishing; 2021. p. 1111–23.

69. Zhang L, Dong W, Mu X. Analysing the features of negative sentiment tweets. Electron Libr. 2018;36(5):782–92.

70. Manning C, Schutze H. Foundations of statistical natural language processing. MIT Press; 1999.

71. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–174.

72. Bali K, Choudhury M, Biswas P, Jha GN, ChoudharySharma NM. Indian language part-of-speech tagset: Hindi. Philadelphia: Linguistic Data Consortium; 2010.

73. Hasan MF, UzZaman N, Khan M (2007) Comparison of unigram, bigram, hmm and brill's postagging approaches for some south Asian languages. Brac J. 2007.

74. Xu J, Croft WB. Quary expansion using local and global document analysis. In: Acm Sigir Forum, ACM New York, NY, USA, 2017; 51: 168–175.