Text Emotion Recognition Model for Human-Robot Interaction using Machine Learning and Transfer Learning

Bolsista Neelakshi Joshi (CTI) njoshi@cti.gov.br
Orientador: Dr. Josué J. G. Ramos

Abstract

Among all the modalities used to study underlying emotions, text is widely available. Text Emotion Recognition (TER) aims to identify underlying emotions expressed in written text and has broad applications across customer satisfaction, health analysis, and digital learning platforms. Various large models based on Transformer architecture are advancing in NLP tasks and for those TER is of keen interest. TER modeling being a supervised task, it is data-incentive. Online text contains many non standard formats, emojis, emoticons, and can contain words or sentences in multiple languages. Classifying available text in emotions and constructing a corpus is a challenging task. This work presents analysis of Brazilian Portuguese (BP) TER corpora with traditional machine learning algorithms and transfer learning. The findings confirm the effectiveness of pre-trained Transformer architectures for emotion recognition in BP text and highlight the potential for improved emotion-aware applications in Portuguese NLP.

Key-words: Text Emotion Recognition, Transfer Learning, Machine Learning, HRI, TER.

1. Introduction

Among all the modalities used to study underlying emotions, text is widely available. In natural language processing (NLP), text emotion recognition (TER) or emotion mining discover and classify writers' emotions towards the topic or event described, and is different from sentiment analysis which focuses on knowing the polarity (NANDWANI & VERMA, 2021). TER has many real-world applications such as understanding customer satisfaction, product recommendation based on customer emotions, in e-learning and health care applications.

With the increased use of the internet, various online shopping, media, websites, and channels are providing a platform to buy, express and entertain. Different user contents are available in terms of reviews, conversations, reactions, and expressions and that with many freedoms like use of emojis, non-standardize words, mixing different languages together, and using undefined, self created shortforms of the words. Constructing TER corpus from such a text and further identifying emotions is challenging. Various research addresses corpus building and validation processes (BRITO et al., 2025).

TER analysis involves processing, feature extraction, and modeling steps:

1. To prepare the text corpus for the analysis, *preprocessing* is the crucial step which involves

- ➢ *cleaning* irrelevant elements such as html tags, names, white spaces; *tokenization* splits text into individual words, known as tokens;
- ➢ *normalization* includes converting all text to lower case, remove stop words, apply lemmatization, i.e., using a base form of words;
- ➢ *label encoding* that converts categorical labels into numerical format.

2. The next step is to *extract* relevant *features* from processed text. few such methods are
   - ➢ *Term Frequency-Inverse Document Frequency* (TF-IDF): this statistical measure identifies most significant words and filters out common or stop words. Thus, reduces the dimensionality.
   - ➢ *Word Embeddings* captures the semantic meaning of words in the vector space as it encodes the contextual relationship between words.
3. *Feature modeling* can be performed with various machine learning algorithms like support vector machines, random forest, logistic regression, or naive bayes. or can implement different deep learning models.

Following the introduction of the Transformer model by Vaswani et al. (2017), which incorporated self-attention mechanisms and enabled the parallelization of sequential processes, the advancements in NLP models and research have progressed very rapidly. The state-of-the-art Bidirectional Encoder Representations from Transformers (BERT) models established pivotal in many fields.

The *bert-base-portuguese-cased* model, also known as *BERTimbau Base* (SOUZA et al., 2020) is developed by neuralmind for Brazilian Portuguese (BP) language. The cased model is sensitive to the cases thus can preserve proper nouns and grammatical context. It is available in two sizes, Base and Large with 110M and 355M parameters respectively on HuggingFace[1]. This model achieved state-of-the-art performance for named entity recognition (NER), sentence textual similarity (SNS) and recognizing textual entailment (RTE) NLP tasks. NER identifies named entities like location, people, organizations. SNS assesses similarity between sentences and discerns the relationship between them as an RTE task. Also it is capable of doing many other NLP tasks such as translation, summarization, and text generation. Though it is not trained to identify emotions, it looks promising for transfer learning.

This work presents TER analysis performed with four BP TER corpora. Section 2 details the databases and method of analysis. Section 3 describes the analysis and corresponding results. Section 4 concludes the article by discussing findings.

2. Data and Methodology

This section describes TER datasets used in the analysis and the methodology.

- Nascimento et al. (2018) created the MQDEmotion2018[2] corpus by extracting posts from social network 'Meu Querido Diário' (MQD). On MQD, users share their daily

[1] https://huggingface.co/neuralmind/bert-base-portuguese-cased
[2] https://github.com/LaCAfe/

lives expressing sentiments and emotions; even users can label their posts with the basic emotions. This dataset contains a total of 18,601,010 words and a vocabulary of 265,741 distinct terms. The authors selected a total of 79,523 entries categorizing among anger (6,323), disgust (1,512), fear (6,112), happiness (32,672), sadness (27,642), and surprise (5,262) emotions.

- Siqueira et al. (2024) developed a dataset[3] of polarities and emotions by collecting reviews in BP from the first top ten most downloaded applications on Google Play Store in May 2023. A total of 300 reviews were manually classified and annotated carefully for polarity among positive, negative and neutral, and for emotions among anger, disgust, fear, happiness, sadness, and surprise. All the annotations are finalized by three annotators unanimously for accuracy and reliability.

- Cortiz et al. (2021) created the first TER corpus[4] classifying texts in 28 fine-grained categories in BP by reviewing the GoEmotions database emotion categories. GoEmotions corpus[5] (DEMSZKY et al., 2020) contains 58,000 carefully curated Reddit comments in English, which are manually annotated in 27 different fine-grained emotions. During the first stage of review, seven researchers from different backgrounds suggested BP translation of emotions from English based on the definition of each emotion and created an emotion list. These emotions were reviewed for consistency in the second stage, resulting in 28 distinct emotions (see Table I in CORTIZ et al., 2021). This is also available on Kaggle as Go Emotions PT-BR dataset[6]. In this work, only basic emotions were used.

- With an effort of many large groups of psychologists all over the world and student respondents, the international survey on emotion antecedents and reactions (ISEAR) dataset[7] was developed under the ISEAR project (SCHERER et al., 2001). In a survey, respondents addressed the situation they appraised and reacted with the following emotions: anger, disgust, fear, guilt, happiness, sadness, and shame. Machine translation of the ISEAR dataset in BP is made available as ISEAR Corpus Translated to Portuguese BR[8] by the Kaggle user. Additionally, the translated dataset was augmented using a synonym dataset[9].

These datasets were preprocessed individually to remove null and duplicated texts. The GoEmotion dataset contains raters evaluation where the same text is annotated with different emotions by different raters. To remove duplicates, a majority vote was taken in such a dilemma. Additionally, MQD dataset is available in json format, hence it is also processed to remove html tags and line breaks. Then all the datasets were combined together. It is highly imbalanced. Figure 1 presents it graphically.

The analysis is performed with traditional machine learning algorithms support vector machines, random forests, logistic regression, and multinomial naive bayes. Also transfer

[3] https://zenodo.org/records/10823148
[4] https://github.com/diogocortiz/PortugueseEmotionRecognitionWeakSupervision
[5] https://github.com/google-research/google-research/tree/master/goemotions
[6] https://www.kaggle.com/datasets/antoniomenezes/go-emotions-ptbr
[7] https://www.unige.ch/cisa/index.php/download_file/view/395/296/
[8] https://www.kaggle.com/datasets/antoniomenezes/isear-corpus-translated-to-portuguese-br
[9] https://github.com/fititnt/DicSin-dicionario-sinonimos-portugues-brasileiro

learning is experimented with the Transformer based *bert-base-portuguese-cased* model that has trained for Brazilian Portuguese language.

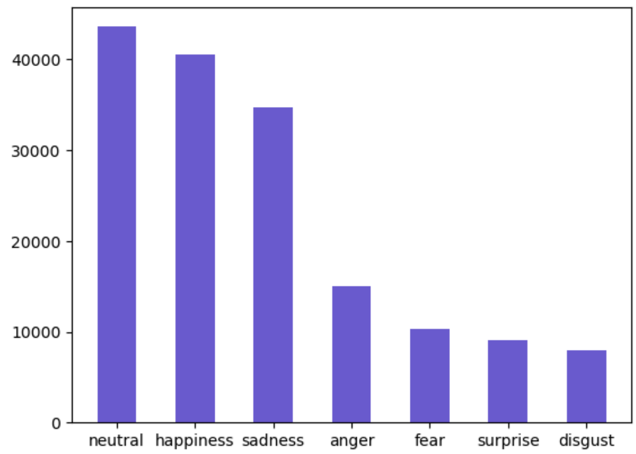

Figure 1 – Distribution of text instances per emotion in the combined dataset

3. Analysis & Results

For analysis, first of all, text data was carefully preprocessed to preserve special characters such as accents, punctuation, and emoji/emoticon tokens; unicode emojis converted to text tokens; mapped emoticons to text tokens using emoji and re library. Text is further processed to remove white spaces and control characters; then converted to lower case. Categorical emotions were mapped to numerical emotions. This processed text is considered for feature extraction and used with numerical labels.

As the combined dataset is imbalanced, we created a new version of the dataset by downsampling it, by selecting 7925 instances per emotion randomly as this is the number of maximum samples for disgust emotion, which is the least in the combined dataset. Both the datasets, combined & balnced but downsampled, were splitted among 80% train set and 20% test set. For machine learning algorithms, IF-IDF vectorization is obtained with scikit-learn library. For transfer learning, *bert-base-portuguese-cased* model specific tokenizer is used to take advantage of model' vocabulary. The analysis was carried out on a server with Ubuntu 24.04.2 LTS OS and GPU (cuda 12.9). The environment was built on Python 3.11.

Table 1 presents the result. Among traditional ML models, the LinearSVC algorithm provided better classification. Though the combined dataset is imbalanced the obtained measures are much higher than the balanced dataset. Results with the bert based model are even higher than the traditional algorithms even after just 2 epochs.

dataset	model	accuracy	F1 score	MCC
Combined	Linear SVC	0.66	0.64	0.56
		0.72	0.71	0.6
balanced downsampled	Linear SVC	0.51	0.5	0.42
		0.58	0.57	0.5

Source: Author's production

Table 1 – results of TER analysis

4. Discussion / Conclusion

This work explores text emotion corpora available in Brazilian Portuguese. With the increased use of the internet and social platforms, though available text is ample, classifying those in emotions and constructing the corpus is challenging which is evident with few available corpora in BP.

Processing texts for analysis is crucial. As analyzed corpora are constructed from the internet, have many non standard abbreviations, emoji and emoticons. Section 3 detailed the preprocess steps we followed before combining datasets together and also after combining, to prepare for the analysis. The preprocessing pipeline, especially the careful treatment of emojis and emoticons proved critical for preserving affective cues in text.

The results demonstrate that transfer learning with a pre-trained BERT model bert_base_Portuguese_cased is effective and outperforms traditional ML algorithms with very less training. This model is trained on huge data hence it has very vast vocabulary. Embeddings created by the model contain more contextual information compared to statistical TF-IDF methods. Among traditional models, linear support vector classification outperformed logistic regression, random forest and multinomial Naive Bayes. This outcome is consistent with previous studies showing that SVMs handle high-dimensional sparse text representations efficiently (NANDWANI & VERMA, 2021).

Results on both datasets are intriguing. Training on larger data seems beneficial as the model can learn more nuances in spite of having class imbalance bias. Although downsampling achieved class balance, it reduced the number of training samples, affecting model' performance.

This study provides a baseline for TER in Brazilian Portuguese using both classical and Transformer-based approaches. The performance of BERTimbau suggests that large-scale, language-specific pretraining is essential for advancing in NLP tasks and can adapt to non-trained tasks like emotion recognition. Future work could explore more advanced rebalancing and data augmentation strategies, exploring large pretrained models for emotion learning and extending to more fine grained emotions.

Referências

BRITO, I. FARBER, F. DOLLIS, J. PEDROZO, D. NOVAIS, A. SILVA, D. & FILHO, AG. *AKCIT at SemEval-2025 Task 11: Investigating Data Quality in Portuguese Emotion Recognition*. In Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), pp. 2305-2310. 2025.

CORTIZ, D. SILVA, JO. CALEGARI, N. FREITAS, AL. SOARES, AA. BPTELHO, C. RÊGO, GG. SAMPAIO, W. & BOGGIO, PS. A Weakly Supervised Dataset of Fine-Graine*d Emotions in Portuguese*. In Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana, pp. 73–81, 2021, doi:10.5753/stil.2021.17786.

DEMSZKY, D. MOVSHOVITZ-ATTIAS, D. KO, J. COWEN, A. NEMADE, G. & RAVI, S. *GoEmotions: A Dataset of Fine-Grained Emotions*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 4040–4054. 2020. doi:10.18653/v1/2020.acl-main.372.

NANDWANI, P. & VERMA, R. *A review on sentiment analysis and emotion detection from text*. Social Network Analysis and Mining, 11:81, 2021.https://doi.org/10.1007/s13278-021-00776-6

NASCIMENTO, G. DUARTE, F. & GUEDES, GP. *Emoções em português do Brasil: um conjunto de dados e resultados de base*. In: Anais do VII Brazilian Workshop on Social Network Analysis and Mining, SBC, 2018. doi: 10.5753/brasnam.2018.3595

SCHERER, KR. BANSE, R. & WALLBOTT, HG. *Emotion inferences from vocal expression correlate across languages and cultures.* Journal of Cross-cultural psychology, vol. 32, no. 1, pp. 76–92, 2001. doi:10.1177/0022022101032001009

SIQUEIRA, V. LUNARDI, MG. SILVA, W. COSTA, RLH. & TALES, SS. *A Dataset of Polarities and Emotions from Brazilian Portuguese Play Store Reviews* [Data set], 2024, Zenodo. doi:10.5281/zenodo.10823148

SOUZA, F. NOGUEIRA, R. & LOTUFO, R. *BERTimbau: pretrained BERT models for Brazilian Portuguese*. In 9th Brazilian Conference on Intelligent Systems, BRACIS, pp. 403-417, 2020.

VASWANI, A. SHAZEER, N. PARMAR, N. UsSZKOREIT, J. JONES, L. GOMEZ, AN. KAISER, Ł. & POLOSUKHIN, I. *Attention is all you need*. Advances in neural information processing systems, 30, 2017.