A Weakly Supervised Dataset of Fine-Grained Emotions in Portuguese

Diogo Cortiz[1,2], Jefferson O. Silva[2,4], Newton Calegari[2], Ana Luísa Freitas[3], Ana Angélica Soares[3], Carolina Botelho[3], Gabriel Gaudencio Rêgo[3], Waldir Sampaio[3], Paulo Sergio Boggio[3]

[1]Brazilian Network Information Center (NIC.br)
São Paulo, SP – Brazil

diogo@nic.br

[2]Pontifical Catholic University of São Paulo (PUC-SP)
São Paulo, SP – Brazil

{dcortiz, silvajo, njcalegari}@pucsp.br

[3]Mackenzie Presbyterian University
São Paulo, SP – Brazil

{paulo.boggio}@mackenzie.br

[4]Jusbrasil
São Paulo, SP – Brazil

Abstract. *Affective Computing is the study of how computers can recognize, interpret and simulate human affects. Sentiment Analysis is a common task in NLP related to this topic, but it focuses only on emotion valence (positive, negative, neutral). An emerging approach in NLP is Emotion Recognition, which relies on fined-grained classification. This research describes an approach to create a lexical-based weakly supervised corpus for fine-grained emotion in Portuguese. We evaluate our dataset by fine-tuning a transformer-based language model (BERT) and validating it on a Gold Standard annotated validation set. Our results (F1-score= .64) suggest lexical-based weak supervision as an appropriate strategy for initial work in low resourced environment.*

Resumo. *A Computação Afetiva é o estudo de como os computadores podem reconhecer, interpretar e simular os afetos humanos. A Análise de Sentimento é uma tarefa comum em PLN, mas se concentra apenas na valência da emoção (positiva, negativa, neutra). Uma abordagem emergente é o Reconhecimento de Emoção, que depende de uma classificação refinada. Nesta pesquisa, descrevemos uma abordagem de supervisão fraca baseada em Itens Lexicais para criar um corpus de emoções refinadas em português. Avaliamos nosso corpus fazendo o ajuste fino de um modelo de linguagem baseado em Transformer (BERT) e avaliando-o em um conjunto de validação anotado. Nossos resultados (F1-score= .64) sugerem que a supervisão fraca baseada em Itens Lexicais pode ser uma estratégia apropriada para o trabalho inicial em ambiente de poucos recursos.*

1. Introduction

Affective Computing comprises the study of how computers can recognize, interpret and simulate human affects. According to Picard [Rosalind 2000], a field pioneer, it is imperative to develop ways for computers to be able to recognize, understand and express emotions for intelligent and natural interaction between humans and machines. Although Affective Computing may employ several input types such as facial expression images, voice, or physiological data, our research focuses on the written language. Thus, our scope lies in the area of Natural Language Processing (NLP).

A common task in NLP is Sentiment Analysis which classifies a text into three different categories: positive, negative, and neutral [Drus and Khalid 2019]. The Emotion Recognition task, however, is a more detailed NLP task. Rather than classifying text only into valence categories (positive, negative, or neutral), it classifies text into more detailed emotional categories.

The delimitation of this research is concentrated in the area of fine-grained Emotion Recognition. We studied an approach to create a corpus in Portuguese for this task using a weak supervision approach.

2. Related Work

Several works in NLP are based on the theory of basic emotions [Ekman 1992] to classify texts into defined categories of emotions, the number of categories ranging from 4 (four) to 8 (eight). One of the studies adopted the 6 (six) basic emotions proposed by Ekman (joy, fear, anger, sadness, surprise, and disgust) to train an Emotion Recognition model [Batbaatar et al. 2019]. Another research added trust and anticipation to the basic emotions, working with 8 (eight) basic emotion categories [Sosea and Caragea 2020].

In the realm of emotion study, other relevant theory is the Theory of Constructed Emotion [Barrett 2016], which assumes that emotions are not universal, but idiosyncratic. This theoretical debate imposes methodological limitations that a different computational approach may help to solve. The Semantic Space Theory [Cowen and Keltner 2021] let us recognize and analyze emotional content of naturalistic stimuli, using open-ended statistical techniques to capture emotional variations in behavior. The results suggest that more than 25 emotional classes have distinct profiles of previous expressions and events. The authors argue that these emotions are high-dimensional, categorical, and often blended.

GoEmotions is a dataset with more than 58,000 English Reddit comments for training NLP models in the Emotion Recognition task. It was annotated for 27 categories of emotions and Neutral based on the Semantic Space Theory. The authors fine-tuned a BERT language model and achieved an average F_1-score of $.46$ [Demszky et al. 2020].

Despite the average F_1-score below $.50$, some classes scored above $.70$. It is a complex dataset with many categories that often have fuzzy boundaries between them. It is essential to discuss the importance of creating a fine-grained dataset with more emotional categories. Research in the Affective Computing area is not limited to Sentiment Analysis or categories proposed by the basic theory.

Although the GoEmotion dataset was released according to open data standards [Demszky et al. 2020], the scope of the corpus is limited to English, which makes it difficult to use in applications in other languages. One of the challenges that Machine Learn-

ing faces is dealing with a low-resourced environment (when the data available is not enough to train the models). This phenomenon can happen in specific domains of applications but also in specific geographic regions.

In the area of NLP, there is a lack of datasets and corpus available in many languages. It is the case of Portuguese, which has a small amount of Sentiment Analysis datasets when compared to English [Pereira 2021]. It is worth noting that we searched for a dataset of fine-grained emotion, but we did not find any in Portuguese.

3. Objectives, Research Questions and Hypothesis

This research aims to study the creation of a corpus of fine-grained emotions for low resourced languages, specifically Portuguese. Due to limited financial resources, a specific objective of this work is to study the use of the weak supervision strategy to construct our corpus. Weak supervision is a strategy when there is no human annotation of each data point, but the labels are attributed using noisy and limited sources or specific rules. We proposed the following research questions (RQ) to guide our work:

RQ1: Is the weak supervision strategy suitable for building an NLP corpus for the fine-grained Emotion Recognition task in a low resourced environment?

RQ2: What is a proper weak supervision approach to construct a corpus for fine-grained Emotion Recognition tasks in NLP?

Our first hypotheses (H1) is that weak supervision could be a suitable strategy to build NLP corpus for emotion recognition. Our second hypotheses (H2) is that lexical-based approach can be an adequate strategy to collect samples for each of the categories of our dataset, using the Lexical Items (LI) as a criterion for defining the label in an adequate way for Portuguese. A third hypothesis (H3) is that using SOTA Machine Learning techniques (specifically Transformers-based language models), combined with masking techniques in the LI presented in the weakly supervised corpus, can avoid the model overfit the learning phase.

To answer RQ1 and RQ2 and validate our hypotheses, we prepared an experiment to create a weakly supervised corpus in Portuguese and measure its performance by training a classification model. The following sections will describe our experimental protocol, including how we collected and weakly annotated the data, our model architecture, metrics, and results.

4. Experimental Protocol

Our experiment is composed of the following pipeline: defining emotion categories based on semantic space theory for Portuguese; selecting the lexical items related to each emotion category based on its definition; collecting the data; manually annotating a test dataset to create a gold standard; defining the model architecture; training the model and evaluating it on the gold standard. Each of them is described in detail in this paper.

4.1. Defining Emotion Categories

The emotion categories for this research were defined from a review of the GoEmotion work [Demszky et al. 2020]. The review process had two stages and the participation of a

group of 7 (seven) researchers with different backgrounds (psychology, neuroscience, sociology, communications, cognitive science, and computer science). In the first stage, the researchers discussed and reviewed each emotion in English during a working meeting. They proposed a translation into Portuguese based on the definitions of each emotion. The result of this first stage was a translated list of terms with consensus among the reviewers.

The second stage was reviewing the categories' definitions in Portuguese to check if they were consistent with the language. The reviewers suggested changing the emotional category *cuidado*, translated from *caring* to *compaixão*, as it is a more broad and blended category in the Portuguese language. The second proposal was the removal of the emotion *realization*, in the sense of perceiving something, as it is not a much prevalent emotional category in the Portuguese language. Finally, there was a consensus among researchers to add the categories *saudade* and *inveja* to the list. We also removed neutral to focus on emotions. The final list consists of 28 emotional categories in total. All emotions and their definitions are presented in Table 1.

Table 1. Portuguese Emotion Categories

CATEGORY IN PORTUGUESE	CORRESPONDING IN GOEMOTIONS	DEFINITION IN PORTUGUESE
ADMIRAÇÃO	ADMIRATION	Achar algo impressionante ou digno de respeito.
DIVERSÃO	AMUSEMENT	Achar algo engraçado ou se divertir.
RAIVA	ANGER	Forte sentimento de desprazer ou antagonismo.
ABORRECIMENTO	ANNOYANCE	Raiva leve, irritação.
APROVAÇÃO	APPROVAL	Ter ou expressar uma opinião favorável.
CONFUSÃO	CONFUSION	Falta de compreensão, incerteza.
CURIOSIDADE	CURIOSITY	Forte desejo de saber ou aprender algo.
DESEJO	DESIRE	Forte sentimento de querer algo ou desejar que algo aconteça.
DECEPÇÃO	DISAPPOINTMENT	desprazer causado pelo não cumprimento de expectativas.
DESAPROVAÇÃO	DISAPPROVAL	Ter ou expressar opinião desfavorável.
NOJO	DISGUST	Repulsa despertada por algo desagradável ou ofensivo.
VERGONHA	EMBARRASSMENT	Vergonha ou constrangimento.
ENTUSIASMO	EXCITEMENT	Sensação de grande empolgação e ansiedade.
MEDO	FEAR	Estar com medo ou preocupado.
GRATIDÃO	GRATITUDE	Sentimento de gratidão e apreciação.
LUTO	GRIEF	Tristeza intensa, especialmente causada pela morte de alguém.
ALEGRIA	JOY	Sensação de prazer e felicidade.
AMOR	LOVE	Forte emoção positiva de consideração e carinho.
NERVOSISMO	NERVOUSNESS	Apreensão, preocupação, ansiedade.
OTIMISMO	OPTIMISM	Esperança sobre o futuro ou sobre o sucesso de algo.
ORGULHO	PRIDE	Prazer devido às próprias conquistas ou de alguém
ALÍVIO	RELIEF	Tranquilidade e relaxamento após ansiedade ou angústia.
REMORSO	REMORSE	Arrependimento ou sentimento de culpa.
TRISTEZA	SADNESS	Dor emocional, tristeza.
SURPRESA	SURPRISE	Sentir-se surpreso, assustado com algo inesperado.
INVEJA	-	Desgosto provocado pela felicidade ou prosperidade alheia
SAUDADE	-	Lembrança grata de pessoa ausente ou um momento passado.
COMPAIXÃO	-	Sentimento piedoso de simpatia e de ajuda

4.2. Selecting Lexical Items for weak supervision

After translating and defining emotions into Portuguese, the next step was to select the Lexical Items that would serve as a filter to search for examples and label assignment rules

(weak supervision). For each of the emotions on the list, we initially look for related Lexical Items by synonyms. For this, we use the database available at www.sinonimos.com.br, which has more than 30 thousand synonyms of words and expressions for Portuguese.

Because some words of emotion presented polysemic behavior, we opted for human curation to select the proper Lexical Items. Only synonyms with a semantic relationship with the definition of emotion were considered. For each Verbal Lexical Item, we collect the different conjugations in the repository www.conjugacao.com.br to cover all tenses and moods in Portuguese. To avoid the negation effect, we manipulated the data as follows: we searched for the combination of the word "não" (no/not in Portuguese) or "nem" (neither in Portuguese) followed by a Lexical Item in our list. If an example was found, we removed it from our dataset. We also added slang and terms related to emotions that were known to the authors. The result of this step was a list in which each emotion was associated with a set of lexical items, which were later used as a data collection filter and label assignment rule.

4.3. Data

We use Twitter as a data source. The collection was made between the 23rd and 24th of June (2021) using the platform's official API. The filters used were the list of terms associated with each emotion. Retweets and replies were not considered, keeping only original tweets. Hashtags were removed, but emojis were kept.

In total, 49179 tweets were collected using a weak supervision approach. Each example received the category label according to the Lexical Item used in the collection. For example, if a tweet was collected because it was filtered by a term associated with the emotion *amor*, it would be labeled to the *amor* category.

We tried to maintain a balanced distribution of examples among the classes, but the results of our collection process suggest that some emotional categories are more prevalent than others. We intend to focus on additional data collection for the categories with the smallest number of examples to achieve a better balance distribution in future work. For the training set, we had a total of 47405 examples. We present in Figure **??** the total number of examples by category and the descriptive statistics of our dataset [1].

4.3.1. Masking Lexical Itens

A hypothesis that appeared during the execution of this research was that the models could memorize the Lexical Items (LI) associated with each emotion, reducing generalization properties and causing the model to overfit. We chose to apply a masking technique to the Lexical Items used for collection and label assignment to investigate this phenomenon. The masking technique consisted of replacing an LI by [MASK], as can be seen in the examples in Table 2.

We ended up with three datasets for training three different models. The first is the original dataset that we created using the weakly supervised approach without any masking technique. We identified this dataset as NoMask. The second dataset is the result of applying the masking technique to 30% of examples for each category. We identified

[1]Data available at: https://github.com/diogocortiz/PortugueseEmotionRecognitionWeakSupervision

Figure 1. Examples per categories.

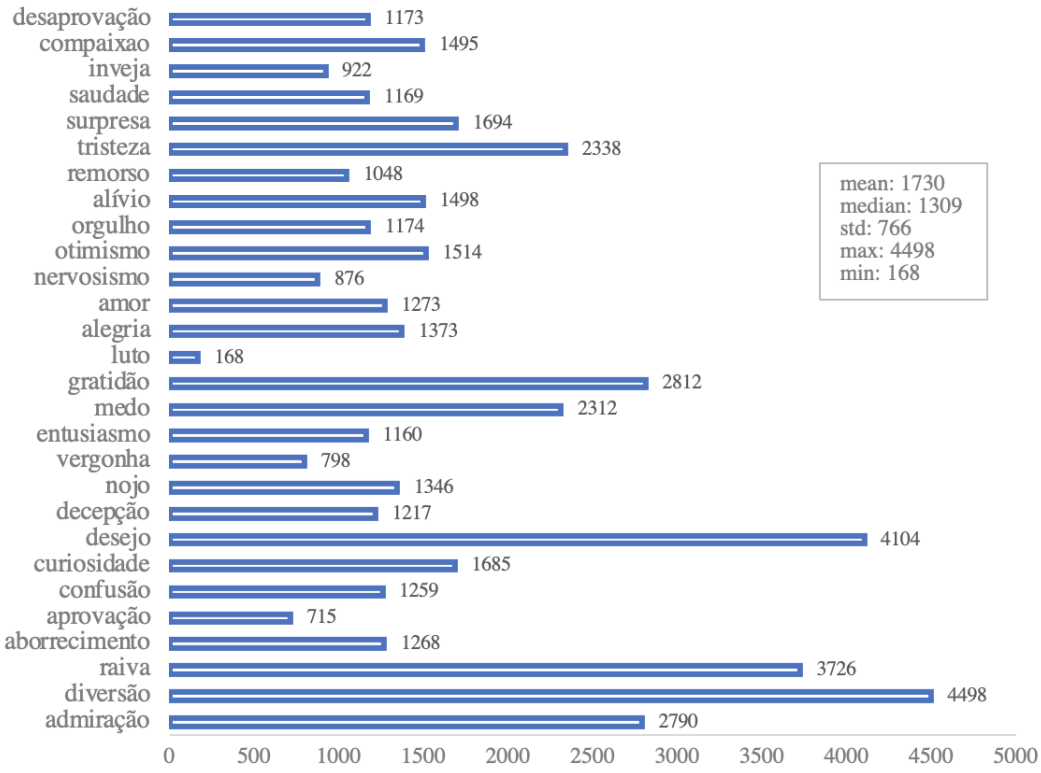

Table 2. Masking technique

Original	tô indignada e não é pouco!
Masked	tô [MASK] e não é pouco!

this dataset as 30Mask. The third dataset is the result of masking all Lexical Items. We identified this dataset as FullMasked.

4.3.2. Gold standard for validation

Despite this research studying the feasibility of the weak supervision approach for the Emotion Recognition task in NLP, it is worth noting the importance of building a dataset with human curation to evaluate the performance of a trained model with the created dataset.

To meet this requirement, we separated a set composed of 1773 examples from the dataset created earlier; we removed the labels assigned by the weak supervision approach so that a human could manually annotate them. We did not apply any Masking technique to this set. Due to limited resources, it was not possible to cross-annotate the validation dataset. Only one annotator annotated each example. For this reason, it is not possible to present any measure of agreement between the annotators. We recognize the limitations of this procedure, which can reduce the quality of supervision and introduce bias.

4.4. Models

To study the performance of our dataset, we needed to fine-tune the BERT language model to the Emotion Recognition task using our weakly supervised dataset. The Bidirectional Encoder Representations from Transformers (BERT) pre-trained language model [Devlin et al. 2018] was released by Google in 2018. Since then, the use of this architecture has improved the performance in different natural language tasks. In our research, we used BERTimbau [Souza et al. 2020], a pre-trained BERT model for Brazilian Portuguese. We fine-tuned three different models using our three different datasets (NoMask, 30Mask, and FullMask).

4.4.1. Parameters Settings

When finetuning the BERT language model, we keep most of the hyperparameters set in the original paper [Devlin et al. 2018]. We changed only batch size and learning rate as proposed by [Demszky et al. 2020]. We trained each model for 4 (four) epochs. The threshold to set a classification as positive was .30 (the same used by [Demszky et al. 2020]). All models were implemented using the huggingface library. The training process used the same computing environment (Quadro RTX 6000).

4.5. Results

The results in Table 3 show the performance of our three models. As we can observe, the model trained with 30% of LI masked (30Mask) had a similar performance to the model trained with the original dataset with no intervention (NoMask).

Those results suggest that masking a certain amount of Lexical items used as a label rule (weak supervision) could be an appropriate strategy to stimulate the model to learn by context and not only by memorizing LI. However, masking all LI introduces much noise to the training dataset, significantly impacting the model performance.

5. Conclusions

According to the results presented, we argue that the adoption of Weak Supervision may be an appropriate strategy for some NLP activities in low-resource scenarios. The creation of datasets is costly and often prohibitive for some economies, making weak supervision an initial alternative for projects when there are insufficient resources to adopt a human supervision methodology. Our RQ1 inquired whether weak supervision is a proper approach to construct a corpus for fined-grained Emotion Recognition in low resourced environment. We found consistent results when evaluating our models, suggesting that weak supervision is an appropriate approach for initial work in the Emotion Recognition NLP task in Portuguese. The results supports our first hypotheses (H1).

This research used a Lexical-based approach to collect, and weak supervise the dataset. According to the results achieved and based on our empirical experience during its execution, we argue that this approach can be appropriate for collecting and annotating data in tasks involving narrow scenarios and well-defined problems. The results help us to answer our RQ2 and validate the H2. However, our experiment has some limitations, such as the validation dataset created from the initial collection of Lexical Items, which

Table 3. Results based on weak supervision

	NoMask			30Mask			FullMask		
Emotion	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
admiração	.67	.44	.53	.71	.44	.54	.38	.39	.39
diversão	.49	.50	.49	.54	.50	.52	.23	.33	.27
raiva	.84	.50	.63	.80	.50	.62	.45	.15	.22
aborrecimento	.82	.74	.78	.81	.75	.78	.47	.26	.34
aprovação	.58	.38	.46	.60	.36	.45	.26	.11	.16
confusão	.68	.66	.67	.66	.63	.65	.42	.22	.29
curiosidade	.71	.61	.66	.71	.61	.66	.37	.15	.21
desejo	.52	.54	.53	.49	.51	.50	.15	.12	.13
decepção	.69	.40	.51	.71	.40	.51	.43	.05	.10
nojo	.81	.95	.87	.83	.97	.89	.53	.15	.24
vergonha	.88	.89	.88	.87	.86	.87	.29	.06	.10
entusiasmo	.74	.91	.81	.72	.89	.80	.30	.16	.21
medo	.75	.93	.83	.76	.87	.81	.43	.35	.38
gratidão	.31	.57	.40	.32	.57	.41	.09	.32	.14
luto	.91	.56	.69	.83	.56	.67	.00	.00	.00
alegria	.68	.66	.67	.69	.65	.67	.27	.24	.25
amor	.88	.50	.64	.80	.50	.61	.39	.41	.40
nervosismo	.94	.64	.76	.86	.66	.74	.50	.03	.06
otimismo	.49	.42	.45	.50	.45	.47	.20	.05	.09
orgulho	.70	.59	.64	.70	.59	.64	.10	.03	.04
alívio	.55	.89	.68	.54	.86	.67	.15	.22	.18
remorso	.64	.71	.67	.62	.71	.66	.40	.08	.13
tristeza	.71	.54	.62	.76	.48	.58	.47	.19	.27
surpresa	.61	.91	.73	.61	.85	.71	.48	.44	.46
saudade	.75	.68	.72	.75	.72	.74	.52	.32	.40
inveja	.93	.93	.93	.93	.93	.93	.88	.34	.49
compaixão	.58	.88	.70	.59	.88	.71	.25	.12	.16
desaprovação	.60	.02	.03	.50	.02	.03	.26	.03	.06
macro avg	**.70**	**.64**	**.64**	**.69**	**.63**	**.64**	**.35**	**.19**	**.22**

makes it difficult to assess the generalization performance of the models. In this sense, we can neither validate nor refute our third hypothesis (H3). We plan to build a new dataset with human supervision in future work without using the Lexical Items list in the filter during data collection. It will be possible to validate and compare the generalization performance of models using different datasets.

References

Barrett, L. F. (2016). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, page nsw154.

Batbaatar, E., Li, M., and Ryu, K. H. (2019). Semantic-Emotion Neural Network for Emotion Recognition From Text. *IEEE Access*, 7:111866–111878.

Cowen, A. S. and Keltner, D. (2021). Semantic Space Theory: A Computational Approach to Emotion. *Trends in Cognitive Sciences*, 25(2):124–136.

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., and Ravi, S. (2020). GoEmotions: A Dataset of Fine-Grained Emotions. In *Proceedings of the 58th Annual*

Meeting of the Association for Computational Linguistics, pages 4040–4054, Stroudsburg, PA, USA. Association for Computational Linguistics.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Drus, Z. and Khalid, H. (2019). Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. *Procedia Computer Science*, 161:707–714.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200.

Pereira, D. A. (2021). A survey of sentiment analysis in the Portuguese language. *Artificial Intelligence Review*, 54(2):1087–1115.

Rosalind, P. (2000). *Affective Computing*. MIT Press, Cambridge.

Souza, F., Nogueira, R., and Lotufo, R. (2020). BERTimbau: Pretrained BERT Models for Brazilian Portuguese. pages 403–417.