

# Emotion Detection for Spanish by Combining LASER Embeddings, Topic Information, and Offense Features

Fedor Vitiugin and Giorgio Barnabò

Universitat Pompeu Fabra, Barcelona, Spain  
fedor.vitiugin@upf.edu

**Abstract.** This paper describes the system submitted by WSSC Team to the EmoEvalEs@IberLEF 2021 emotions detection competition. We propose a novel model for Emotion Detection that combines transformers embeddings with topic information and offense features. The system classifies social media text emotions leveraging its context representations. Our results show that, for this kind of task, our model outperforms baselines and state-of-the-art text classification methods. As for the leader-board, our classification model achieved a macro weighted averaged F1 score of 0.661427, and a overall accuracy of 0.675725, reaching the 9th and 10th place respectively.

**Keywords:** Natural language processing · Emotion detection · Deep learning.

## 1 Introduction

Emotion Detection is a branch of sentiment analysis that seeks to extract fine-grained emotions from either speech/voice, image, or text data. Detecting emotions from texts has proven to be quite a challenging task, regardless of the quantity of available data [1]. Understanding emotions expressed by users on social media is particularly hard due to the absence of voice modulation, facial expressions, and other features that may work as clues during the context and relation extraction process.

Besides that, the need for disambiguating emotion-conveying words in order to verify classified emotions as real emotions still represents a significant hitch, since texts often contain expressions that could refer to different emotions. For example, a phrase like “I can’t stand it” could convey anger and disgust depending on the context. Nonetheless, recently, state-of-the-art results were obtained by using pre-trained transformer-based models. Needless to say, in the past three

---

*IberLEF 2021, September 2021, Málaga, Spain.*

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

years, pre-trained language models such as BERT [6] revolutionized the NLP world allowing to achieve extraordinary results in almost any known task. These models are particularly effective because they generate word embeddings that capture the semantic and contextual information of texts.

The existing state-of-the-art emotion detection models usually only extract context features from texts and pay less attention to external features like the kind of event these messages were posted for. In our work, we tried to fill this gap by including additional context information and by also considering the presence of offenses inside these messages. LASER [4] embeddings were used to encode the social media texts and were then combined with topic features and offense features.

The main contribution of this study is an approach based on a combination of contextualized word embeddings, topic information, and offense features specifically tailored for improving the emotion detection process. We evaluated our methodology on the EmoEvalEs@IberLEF 2021 [11] competition dataset, showing that our model outperforms baselines. We also analyzed the most frequent mistakes that our model made.

The remainder of this paper is organized as follows. We first present the related work, then we introduce our approach, and finally we show the experiment results and the error analysis.

## 2 Related work

There are five classes of approaches for recognizing emotions in texts: keyword-based approaches, rule-based approaches, classical learning-based approaches, hybrid approaches, and deep learning approaches[3]. Recent approaches for emotions detection propose solutions that use deep learning techniques to classify emotions in texts.

### 2.1 LSTM

Deep learning is a branch of machine learning in which deep neural network architectures learn from experience and understand the world in terms of a hierarchy of concepts, where each concept is defined in terms of its relation to simpler concepts. This approach allows a model to incrementally learn complex concepts putting together simpler ones [7]. In this context, the long short-term memory (LSTM) architecture proved to be particularly effective. LSTM is a special form of recurrent neural network (RNN) with the capability of handling long-term dependencies. LSTM overcomes the vanishing or exploding gradient problem common in other type of RNNs.

Here's a list of the main steps to take when using LSTM for emotion recognition in texts:

1. text preprocessing, that is tokenization, stopwords removal, and lemmatization;

2. encode texts through an embedding layer and then use these embeddings to feed one or more LSTM layers.
3. delivering outputs into a dense neural network (DNN) with units equal to the number of emotion labels and a sigmoid activation function to perform classification.

## 2.2 Transformers

The encoder block of transformers, initially designed for machine translation, has become the de-facto standard pre-trained language modeling architecture for solving most NLP tasks such as text classification, text generation, document summarization, question answering just to name a few [2]. Up to now, several state-of-the-art model for detecting text-based emotions already use BERT and its variants.

One of the ways to improve the performance of emotion classification is the extension of BERT model by a linear transformation layer with sigmoid activation. The proposed model was evaluated using the EmoBank data and obtained a micro F1 score of 0.688 and 0.695 when fine-tuned on the ISEAR and SemEval datasets, respectively [12]. Another way of using BERT for emotion classification is to use a two-step approach that first encode texts into vectors and then classify them into emotions using the soft max classifier [10]. One more way of using BERT is extracting contextualized word embeddings from text data and subsequently use SVM to perform classification. Authors of this approach[8] feed the model with text passages of an average length of 650 tokens. Since BERT can only process 512 input tokens, the essays were divided into sub-documents. The sub-documents were pre-processed and fed into the BERT base model. Feature vectors for the document were obtained by computing the mean of each of the 12 BERT layers' contextual token representations. The last four-layer representations were then concatenated with the corresponding 84 Mairesse features for the essay. The feature vector was then fed into the SVM classifier, producing a prediction. The final prediction was obtained through majority voting.

## 3 Model

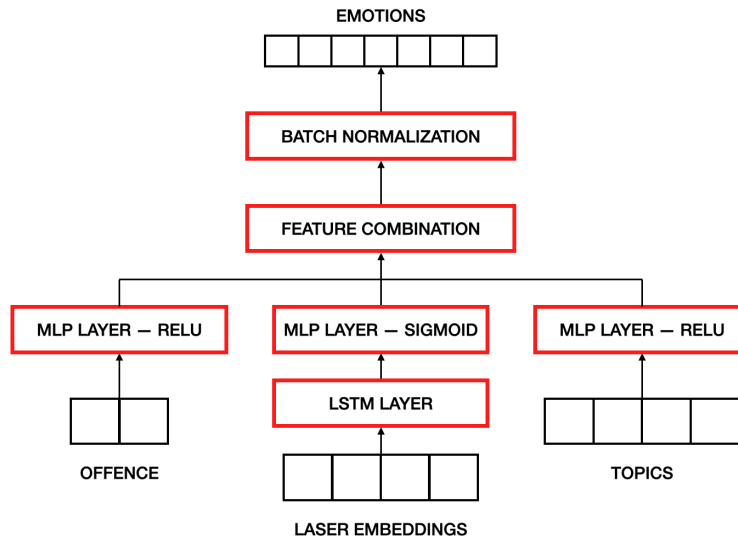
### 3.1 Pre-processing

During the pre-processing step, we only detected and replaced all emojis with their respective short-codes using a freely available Python library *emoji*. Since data provided by organizers of the competition were polarized, they replaced all the hashtags with the keyword "HASHTAG" in order to prevent the automatic classifier from relying on hashtags to categorize the emotion associated with a tweet. Moreover, the user mentions were replaced by "@USER".

### 3.2 LASER Embeddings

For representing the input data, we used embeddings generated by two pre-trained transformer-based models: DistilBERT and Language-Agnostic SEntence Representations (LASER) [4]. The main difference of LASER from other transformers is generating sentence-level embeddings instead of word/token-level embeddings.

Given an input sentence, LASER provides sentence embeddings which are obtained by applying max-pooling operation over the output of a Bidirectional LSTM (BiLSTM) encoder. BiLSTM output is constructed by concatenating outputs of two individual LSTMs working in opposite directions (forward and backward). This way more contextual information is included in the output with respect to a single LSTM reading text from left to right. In our experiments, we used LASER to embed all tweet sentences into 1024-dimension fixed-size vectors.



**Fig. 1.** Combining the transformer embeddings, topic information and offense features using deep MLP.

### 3.3 Proposed Model

As additional features, we detected offenses and we extracted topics of tweets. The both types of features were provided in the EmoEvent corpus. LASER embeddings are passed as input to a Long Short Term Memory Network model to encode the social media texts. Finally, we combined all features through the architecture originally proposed for the detection of fake news articles [5]. The full architecture is shown in Figure 1.

## 4 Experiment

### 4.1 Dataset Description

We use the dataset released for the EmoEvalEs@IberLEF 2021 competition [14] — shared task “Emotion detection and Evaluation for Spanish”. The task consists of classifying the emotion expressed in a tweet as one of the following emotion classes:

- anger (also includes annoyance and rage);
- disgust (also includes disinterest, dislike, and loathing);
- fear (also includes apprehension, anxiety, concern, and terror);
- joy (also includes serenity and ecstasy);
- sadness (also includes pensiveness and grief);
- surprise (also includes distraction and amazement);
- others: the emotion expressed in a tweet as ‘neutral or no emotion’.

The dataset is based on events that took place in April 2019 related to different domains: entertainment, catastrophe, political, global commemoration, and global strike. There are messages in a total of 8 different topics. For the task dataset was split into dev, training, and testing partitions. The distribution of EmoEvalEs@IberLEF 2021 dataset is shown in Table 1.

**Table 1.** EmoEvalEs@IberLEF 2021 dataset description.

	anger	disgust	fear	joy	sadness	surprise	others	total
train	589	111	65	1227	693	238	2800	5723
dev	85	16	9	181	104	35	414	844
test	168	33	21	354	199	67	814	1657

### 4.2 Training parameters

The proposed model computes the feature vectors separately and then combines these with the help of an MLP layer. We use categorical cross-entropy as the loss function to optimize our architecture with a soft-max layer that tries to classify any given social media text into one of seven emotion classes. The hyperparameter setting is shown in Table 2. The full code is provided in the project repository <https://github.com/vitiugin/ComboLASER>.

### 4.3 Baselines and compared methods

In the current work, we also used schemes with a combination of DistilBERT embeddings. The concept of distillation in neural networks aims at speeding up models. The key idea is to replace massive architectures with countless parameters with a lightweight version of the same architecture that possesses fewer

**Table 2.** Values of hyper-parameters. The first row of the table describes the parameters for extracting individual features. The second row shows the parameter setting of the feature combination layer.

Hyperparameter	Offense Features	LASER Embeddings	Topic features
MLP layers	2	1	2
MLP neurons	128;24	256;128	128;24
Dropout	-	0.5	-
Activation	relu	sigmoid	relu
MLP layers		1	
MLP neurons		7	
Activation		softmax	
Optimizer		Adam	
Learning rate		0.001	
Batch size		100	
Loss		Categorical Crossentropy	

parameters [15]. The DistilBERT takes the architecture of the initial version of BERT, reduces the number of layers in the BERT-base model by a factor of 2, removes token embeddings and poolers to yield a much smaller and faster version of BERT for general-purpose use. It applies dynamic masking and ignores the next sentence predictions for better inference [9].

According to recent surveys SVM is the most popular machine learning scheme for emotion detection from text [3]. Subsequently, one of our baselines is a model that concatenates vectors of transformer embeddings (LASER and DistilBERT), topics and offense features which passes them to an SVM classifier.

To prove the need of using additional topic and offense feature vectors we also used only transformer embeddings as input to a LSTM model.

#### 4.4 Results

As evaluation measures we used two multi-class classification metrics: accuracy and macro weighted averaged F1 score. The full results on development and test splits are shown in Table 3.

We can observe that the SVM-based models with concatenated feature vectors have high performance even compared with LSTM-based networks trained only on transformers embeddings. Further LASER embeddings demonstrate a higher performance compared with DistilBERT embeddings. The proposed *Combo LASER* model shows the highest performance, which is perhaps due to the fact that the model takes into consideration the sentence-level context encoded in the LASER embeddings. In terms of performance, the proposed solution is worse than the solution that took the first place by 4.5%.

Analysing our model mistakes, we found that our model often (more than 50% compared to the volume of class in tested data) mis-classified *Disgust* as *Anger* and *Fear* as *Sadness*. On the other hand, the best results were achieved for *Sadness*, *Surprise*, and *Others* (less than 25% of mistakes).

**Table 3.** Comparison with baselines. Results multiclass classification. Best performances are in bold. (5 fold CV). \* denotes the proposed model.

Model Scheme	ACC	F1
<i>SVM+DistilBERT</i>		
dev	66.89±0.17	65.57±0.14
test	66.99±0.12	65.34±0.14
<i>SVM+LASER</i>		
dev	67.48±0.16	65.62±0.12
test	66.49±0.11	64.76±0.12
<i>LSTM+DistilBERT</i>		
dev	67.63±0.52	58.63±0.19
test	64.76±1.13	60.84±0.36
<i>LSTM+LASER</i>		
dev	67.84±0.84	60.61±0.46
test	66.86±0.49	61.82±0.58
<i>Combo DistilBERT</i>		
dev	64.00±1.38	61.63±1.17
test	62.68±0.49	62.68±0.46
<i>*Combo LASER</i>		
dev	<b>68.10±1.68</b>	<b>66.16±0.67</b>
test	<b>67.54±0.78</b>	<b>66.32±0.76</b>

We also found out that two pairs of emotions detected with mistakes on the both sides: of *Anger-Disgust* and *Joy-Other*. While the similarity of the first pair could be explained by close nature of this emotions, the second pair could be explained only by the size of trained and test data. *Joy* and *Others* classes are over-represented in the train and test data.

## 5 Conclusion

In this paper, we explored the benefit of incorporating transformers, topic information, and offense features to deep neural networks on the task of multi-class emotions detection. We also presented our model based on extracted pre-trained LASER embeddings. Experiments on the dataset released during EmoEvalEs@IberLEF 2021 competition demonstrate that our *Combo LASER* model performs better than several baselines and additional features improves the model performance compared with models based only on transformer embeddings[13]. We presented analyses of mistakes that our model made at classification time which can inform future studies for emotion detection.

## References

1. Acheampong, F.A., Nunoo-Mensah, H., Chen, W.: Transformer models for text-based emotion detection: a review of bert-based approaches. Artificial Intelligence Review pp. 1–41 (2021)

2. Al-Rfou, R., Choe, D., Constant, N., Guo, M., Jones, L.: Character-level language modeling with deeper self-attention. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 3159–3166 (2019)
3. Alswaidan, N., Menai, M.E.B.: A survey of state-of-the-art approaches for emotion recognition in text. Knowledge and Information Systems pp. 1–51 (2020)
4. Artetxe, M., Schwenk, H.: Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. Transactions of the Association for Computational Linguistics **7**, 597–610 (2019)
5. Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B., Mittal, A.: On the benefit of combining neural, statistical and external features for fake news identification. arXiv preprint arXiv:1712.03935 (2017)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
7. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep learning, vol. 1. MIT press Cambridge (2016)
8. Kazameini, A., Fatehi, S., Mehta, Y., Eetemadi, S., Cambria, E.: Personality trait detection using bagged svm over bert word embedding ensembles. arXiv preprint arXiv:2010.01309 (2020)
9. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
10. Luo, L., Wang, Y.: Emotionx-hsu: Adopting pre-trained bert for emotion classification. arXiv preprint arXiv:1907.09669 (2019)
11. Montes, M., Rosso, P., Gonzalo, J., Aragón, E., Agerri, R., Álvarez-Carmona, M.Á., Álvarez Mellado, E., Carrillo-de Albornoz, J., Chiruzzo, L., Freitas, L., Gómez Adorno, H., Gutiérrez, Y., Jiménez-Zafra, S.M., Lima, S., Plaza-de Arco, F.M., Taulé, M. (eds.): Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021) (2021)
12. Park, S., Kim, J., Jeon, J., Park, H., Oh, A.: Toward dimensional emotion detection from categorical emotion annotations. arXiv preprint arXiv:1911.02499 (2019)
13. Plaza-del-Arco, F.M., Jiménez-Zafra, S.M., Montejo-Ráez, A., Molina-González, M.D., Ureña-López, L.A., Martín-Valdivia, M.T.: Overview of the EmoEvalEs task on emotion detection for Spanish at IberLEF 2021. Procesamiento del Lenguaje Natural **67**(0) (2021)
14. Plaza-del-Arco, F., Strapparava, C., Ureña-Lopez, L.A., Martin-Valdivia, M.T.: EmoEvent: A Multilingual Emotion Corpus based on different Events. In: Proceedings of the 12th Language Resources and Evaluation Conference. pp. 1492–1498. European Language Resources Association, Marseille, France (May 2020), <https://www.aclweb.org/anthology/2020.lrec-1.186>
15. Tang, R., Lu, Y., Liu, L., Mou, L., Vechtomova, O., Lin, J.: Distilling task-specific knowledge from bert into simple neural networks. arXiv preprint arXiv:1903.12136 (2019)