SINAI at EmoSPeech-IberLEF2024: Evaluating Popular Tools and Transformers Models for Multimodal Speech-Text Emotion Recognition in Spanish

Daniel García-Baena[1,*], Miguel Ángel García-Cumbreras[1] and Salud María Jiménez-Zafra[1]

[1] *Computer Science Department, SINAI, CEATIC, Universidad de Jaén, 23071, Spain*

Abstract

This work presents the participation of the SINAI team at EmoSPeech-IberLEF2024 shared task, Multimodal Speech-Text Emotion Recognition in Spanish. We have addressed the first of the proposed tasks, focused on extracting features and identifying the most representative ones of each emotion in a dataset created from real-life situations compiled from YouTube videos. For emotion analysis, we have evaluated some of the most popular transformers models and specific emotion analysis open source transformers publicly available on Hugging Face. In total, 14 systems have participated (including the baseline provided by the organizers). The best run sent by our team have been placed in position 8th with an F1-score of 0.5200, being 0.6719 the best result obtained in the first task ranking.

Keywords

emotion analysis, text emotion recognition, transformers, natural language processing

1. Introduction

IberLEF is a shared evaluation campaign for Natural Language Processing (NLP) systems in Spanish and other Iberian languages [1]. In an annual cycle that starts in December (with the call for task proposals) and ends in September (with an IberLEF meeting collocated with SEPLN), several challenges are run with large international participation from research groups in academia and industry. Specifically, this shared task was titled *EmoSPeech 2024 Task - Multimodal Speech-Text Emotion Recognition in Spanish* [2], and aims to explore multimodal speech-text emotion recognition for texts written in Spanish [3].

We found interesting to take part into this shared task specially because being able to recognize human emotions is crucial for building positive relationships, whether it is in person or through interactions with computers [4]. Automatic Emotion Recognition (AER) has a growing importance due to its impact on various fields such as healthcare, psychology, social sciences and marketing [5]. AER software can help providing personalized responses and recommendations,

IberLEF 2024, September 2024, Valladolid, Spain

[*] Corresponding author.

daniel.gbaena@gmail.com (D. García-Baena); magc@ujaen.es (M. García-Cumbreras); sjzafra@ujaen.es (S. M. Jiménez-Zafra)

0000-0002-3334-8447 (D. García-Baena); 0000-0003-1867-9587 (M. García-Cumbreras); 0000-0003-3274-8825 (S. M. Jiménez-Zafra)

leading to improved user engagement and satisfaction. The process of AER can be addressed using several taxonomies and is focused on recognizing six basic emotional expressions as they are: anger, disgust, fear, happiness, sadness and surprise [6]. By automatically recognizing emotions, a system could identify, interpret and respond to different human ways of communication, such as text, facial expressions, voice tones or even body language. Different features can be used to identify emotions [7] but, in this work, our team focused exclusively in text for developing automatic emotion analysis systems. Competition is available in CodaLab: https://codalab.lisn.upsaclay.fr/competitions/17647

2. Task description

As we previously noted, our team concentrated on tackling the first subtask from this challenge, analyze a given text and identify the emotion it conveys based on five of Ekman's six basic emotions: anger, disgust, fear, joy, and sadness, as well as one neutral emotion. Therefore, we developed several AER systems that worked exclusively with text written in Spanish. We took all the different texts collected from YouTube by the shared task organizers, about 3500-4000 transcripts of audio segments divided into training and test in an 80%-20% split (see Table 1), and compiled into the available dataset and extracted different features in order to identify the most representative emotions present in each of the cited texts that constitute a corpus created from real-life situations.

The aim of the first task was to analyze texts and identify the emotion that their convey based on the popular Ekman's six basic emotions: anger, disgust, fear, joy, sadness and surprise. It is important to notice that the organizers removed surprise from the list, due to the small amount of examples, and added as well one neutral emotion in order to classify those texts that do not relate with any of Ekman's list of emotions. The evaluation measures for this subtask were: precision, recall and F1-score. Consequently, emotion classification systems in this work were ranked attending to their macro-F1 scores.

3. Methodology

We evaluated several public available models from Hugging Face for task 1. With this task, organizers pretended to classify texts, extracted from YouTube, and classify them according to Ekman's five basic emotions plus one neutral additional option. With this purpose, we chose ten of the most popular source code transformer models from Hugging Face and evaluated their results when they were trained with the shared dataset.

For selecting all the ten LLM from Hugging Face, we used the website public filter to choose those especially indicated for emotion analysis, while working with texts written in Spanish. As we pretended to discover which of the more popular were giving the best results for this first task, we sorted the results from the Hugging Face filter in order to show the most downloaded transformers models on top.

As it can be seen in Table 2, we worked with several types of models while trying to perform emotion analysis over the shared dataset. Thus, we did not evaluate LLM pretrained just texts written only in Spanish, but we did evaluate too those pretrained only with text written in

Table 1
Dataset distribution

Set	Sentiment	Total headlines
dev-train	anger	51
	disgust	91
	fear	3
	joy	46
	neutral	149
	sadness	44
dev-test	anger	13
	disgust	22
	fear	1
	joy	12
	neutral	37
	sadness	11
train	anger	399
	disgust	705
	fear	23
	joy	362
	neutral	1166
	sadness	345
test	anger	100
	disgust	177
	fear	6
	joy	90
	neutral	291
	sadness	86

English and the ones developed using several different languages, in this last case, always including Spanish. In addition, we did not focused exclusively in models that were made precisely to perform emotion analysis but also the most popular options available for general purpose and LLM made to work in similar areas as sentiment analysis.

4. Experimental setup

It is important to note that we did not perform any prior data pre-processing on the shared dataset before of performing all of the experiments.

With respect to the models, all were downloaded from their public profiles in Hugging Face. During the finetuning process we always used Google Colab for coding under a Pro configuration for being able to use their GPU based hardware options.

Finally, concerning the hyperparameters, we did not performed any hyperparameters search so all model configurations are the default ones.

Table 2
Rank list for the training phase

Model	Main language	F1-score
finiteautomata/beto-emotion-analysis	Spanish	0.8172
pysentimiento/robertuito-emotion-analysis	Spanish	0.5486
finiteautomata/beto-sentiment-analysis	Spanish	0.5064
lxyuan/distilbert-base-multilingual-cased-sentiments-student	Multilingual	0.4634
nlptown/bert-base-multilingual-uncased-sentiment	Multilingual	0.4342
somosnlp/bertin_base_climate_detection_spa_v2	Spanish	0.4063
mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis	English	0.3878
distilbert/distilbert-base-uncased-finetuned-sst-2-english	English	0.3535
distilbert-base-uncased-finetuned-sst-2-english	English	0.3104
papluca/xlm-roberta-base-language-detection	Multilingual	0.2408

Table 3
Rank list for the test phase

Model	Main language	F1-score
finiteautomata/beto-emotion-analysis	Spanish	0.5200
finiteautomata/beto-sentiment-analysis	Spanish	0.4919
pysentimiento/robertuito-emotion-analysis	Spanish	0.4704
somosnlp/bertin_base_climate_detection_spa_v2	Spanish	0.4321
nlptown/bert-base-multilingual-uncased-sentiment	Multilingual	0.4074
lxyuan/distilbert-base-multilingual-cased-sentiments-student	Multilingual	0.3931

5. Results and discussion

This section presents the results obtained in the evaluation phase of the shared task EmoSPeech [2], Multimodal Speech-Text Emotion Recognition in Spanish, at IberLEF 2024. The organizers selected target F1-score for ranking the systems from task 1. Each participating team could submit a maximum of ten runs through CodaLab, from which each team had to select the best one for the ranking. We selected our top runs based on the experiments carried out on the training phase. The models and their results for each of the test runs are shown, sorted by their F1-score, in Table 3.

Firstly, we would like to highlight how the Spanish based systems outperformed the multilingual ones. Results from training for English only LLM were specially disappointing so we focused on those models that achieved over 0.4000 F1-score in the training phase (see Table 2).

On the other hand, those models precisely made for emotion analysis generally outperformed those with general or not exactly the same purpose (sentiment analysis). Nevertheless, one sentiment analysis focused model, finiteautomata/beto-sentiment-analysis [8, 9], was able to score a better F1-score than the emotion analysis focused LLM pysentimiento/robertuito-emotion-analysis [10, 9, 11].

Now, paying attention to Table 2, the best result from finiteautomata/beto-emotion-analysis [10, 9] from training, it is surprisingly better than the one from Table 3 (test phase), 0.8172 VS

0.5200, respectively. We cannot categorically confirm this but it seems like the training phase subset that was distributed by the shared task organizers contained several texts that were similar or directly extracted from the same dataset that was used for training the finiteautomata/beto-emotion-analysis model. On the contrary, test subset from this shared task should not contain so similar texts to finiteautomata/beto-emotion-analysis training dataset.

On the other hand, as we were expecting to happened, both finiteautomata/beto-emotion-analysis and pysentimiento/robertuito-emotion-analysis models, that were trained with all the six Ekman's basic emotions, achieved the top positions in Table 3. However, models as finiteautomata/beto-sentiment-analysis, somosnlp/bertin_base_climate_detection_spa_v2 and lxyuan/distilbert-base-multilingual-cased-sentiments-student achieved similar F1-score with a positive, negative and neutral configuration.

In addition, we relate the low F1-scores to the big amount of tags that were expected to be taken into account during the classification process. Systems needed to distinguish between six different options as they were: anger, disgust, fear, joy, sadness and neutral; and this level of exigence make this task way harder than developing a simple binary classifier.

In relation to the last, we find important to highlight that even with a presumably low top F1-score of 0.5200, we were just 0.1519 points behind of the best performer team of the first task, classifying on 8th position for this task 1. This small difference reassures our thinking about how classifying with six different categories, is a hard task for current open source most popular LLM.

6. Conclusions and future work

In this paper we have presented the participation of the SINAI team in the shared task EmoSPeech, Multimodal Speech-text Emotion Recognition in Spanish, at IberLEF 2024. The objective of our experiments, for the emotion analysis task, was to test the performance of the most popular emotion analysis transformer-based models. The main conclusion is that most popular transformers-based solutions are not precise when they have to take into account six different options.

In the future, we want to continue evaluating more different resources in order to further improve our systems by analyzing the contribution of each LLM, testing different transfer learning systems and using different data preprocessing systems to generate new datasets and/or augment existing ones.

Acknowledgments

This work has been partially supported by Project CONSENSO (PID2021-122263OB-C21), Project MODERATES (TED2021-130145B-I00) and Project SocialTox (PDC2022-133146-C21) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, Project PRECOM (SUBV-00016) funded by the Ministry of Consumer Affairs of the Spanish Government, Project FedDAP (PID2020-116118GA-I00) supported by MICINN/AEI/10.13039/501100011033, Project PID2020-119478GB-I00 supported by MICINN/AEI/10.13039/501100011033, and WeLee project (1380939, FEDER Andalucía 2014-2020) funded by the Andalusian Regional Government. The research work conducted by Salud María Jiménez-Zafra has been supported by Action 7 from Universidad de Jaén under the Operational Plan for Research Support 2023-2024.

References

[1] L. Chiruzzo, S. M. Jiménez-Zafra, F. Rangel, Overview of IberLEF 2024: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS.org, 2024.

[2] R. Pan, J. A. García-Díaz, M. Á. Rodríguez-García, F. García-Sánchez, R. Valencia-García, Overview of EmoSPeech at IberLEF 2024: Multimodal Speech-text Emotion Recognition in Spanish, Procesamiento del Lenguaje Natural 73 (2024).

[3] R. Pan, J. A. García-Díaz, M. Á. Rodríguez-García, R. Valencia-García, Spanish meacorpus 2023: A multimodal speech-text corpus for emotion analysis in spanish from natural environments, Computer Standards & Interfaces (2024) 103856.

[4] A. Varghese, J. Cherian, J. Kizhakkethottam, Overview on emotion recognition system, 2015, pp. 1–5. doi:10.1109/ICSNS.2015.7292443.

[5] F. Chenchah, Z. Lachiri, Speech emotion recognition in noisy environment, 2016, pp. 788–792. doi:10.1109/ATSIP.2016.7523189.

[6] E. Rolls, P. Ekman, D. Perrett, H. Ellis, Facial expressions of emotion: An old controversy and new findings: Discussion, Royal Society of London Philosophical Transactions Series B 335 (1992) 69–. doi:10.1098/rstb.1992.0008.

[7] M. S. Fahad, A. Ranjan, J. Yadav, A. Deepak, A survey of speech emotion recognition in natural environment, Digital Signal Processing 110 (2020) 102951. doi:10.1016/j.dsp.2020.102951.

[8] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, Pml4dc at iclr 2020 (2020) 1–10.

[9] J. M. Pérez, J. C. Giudici, F. Luque, pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks, 2021. arXiv:2106.09462.

[10] F. M. P. del Arco, C. Strapparava, L. A. Ureña-López, M. T. Martín-Valdivia, Emoevent: A multilingual emotion corpus based on different events, in: Proceedings of the 12th Language Resources and Evaluation Conference, 2020, pp. 1492–1498.

[11] J. M. Pérez, D. A. Furman, L. Alonso Alemany, F. M. Luque, RoBERTuito: a pre-trained language model for social media text in Spanish, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 7235–7243. URL: https://aclanthology.org/2022.lrec-1.785.