2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Classification of Songs in Spanish with LLMs: An analysis of the construction of a dataset, through classification

Tania Alcántara
Computational Cognitive Sciences Lab. – CIC
Instituto Politécnico Nacional
México, Ciudad de México
0009-0001-4391-6225

Omar Garcia-Vazquez
Computational Cognitive Sciences Lab. – CIC
Instituto Politécnico Nacional
México, Ciudad de México
omar.gava@hotmail.com

Marco A. Cardoso-Moreno
Computational Cognitive Sciences Lab. – CIC
Instituto Politécnico Nacional
México, Ciudad de México
0009-0001-1072-2985

Hiram Calvo
Computational Cognitive Sciences Lab. – CIC
Instituto Politécnico Nacional
México, Ciudad de México
0000-0003-2836-2102 hcalvo@cic.ipn.mx

***Abstract*—Songs convey emotions through melody and lyrics. They capture feelings in small text fragments. Emotions within songs vary: positive, negative, or neutral. This study merged two datasets to create a third, leveraging LLMs for competitive song text classification results.**

***Index Terms*—Machine learning, classification, natural language processing, LLMS, Songs**

I. Introduction

Emotions are human, arising from biology, culture, and neural interactions. Neurobiology links brain reactions to stimuli, upbringing, and cognition. Music and words evoke emotions. NLP aids emotion classification [1]. Emotion identification in sentences is complex due to subjectivity, culture, and language. Spanish differs across regions. Text length affects accuracy. A project compared general and Mexican Spanish song lyrics, unveiling emotional linguistic subtleties. Merging insights established an emotion classification framework embracing diversity and universal emotions.

II. State of the Art

Efforts in Emotion Classification are notable, particularly in analyzing emotions in Thai songs through polar classification using lyrics [5]. This approach combines a lexicon with traditional machine learning, focusing on emotional features in chorus and verses. [7] employs *SentiWordNet* ontology for extraction, using Naive Bayes, *K-Nearest Neighbor*, and SVM for classification. This method explores *SentiWordNet* ontology's role, emphasizing algorithmic emotion classification in songs. It combines ontology and machine learning to enhance emotional understanding, with *Tweet*-trained models excelling in song analysis, while Google *News* and *Common Crawl* models perform better for movies.

Skip-grams outperform GLoVe, highlighting context's importance. Unconventional model blends like CNN and bidirectional LSTMs achieve 90.66% accuracy [9], enhancing text classification for artistic content.

III. Datasets

The first dataset, **Texts of Songs in Spanish**, is a private collection by the Natural Language Processing Lab at Instituto Politécnico Nacional [10]. It contains 91 Spanish songs from Latin America and Spain, spanning genres like bachata, pop, and ballad. Each song's paragraphs, preserving emotional flow, were divided for coherence, yielding 1,477 items without truncation. ”S” stands for neutrality, ”P” for positivity, and ”N” for negativity, each category a determinate number of examples: S, 97; P, 780; N, 600. The second dataset, **Texts of Songs in Spanish from Mexico**, is from the Computational Cognitive Sciences Lab at Instituto Politécnico Nacional [12]. This set includes 200 Mexican Spanish songs across genres like banda, pop, regional, rock, and cumbia. Each song is meticulously segmented into concise paragraphs, maintaining flow and avoiding incomplete ideas. This process resulted in 4,555 fragments. Labels are S for neutral, P for positive, and N for negative emotions. The nomenclature to identify each category is the same. The dataset 2 has the following distribution: S, 1574; P, 1368; N, 1613.

In previous experiments, the dataset 1 was a better result than in dataset 2, giving the following hypotheses: **Short Phrases:** Dataset two was algorithmically created, with each row representing a sentence from the song. In contrast, dataset 1 was manually compiled, with ideas completed and sentences combined. Merging the two datasets is proposed to maximize their benefits. Alternating the data creates an organic set, ensuring no bias. The datasets have matching columns and labels, with dataset 2's phrases largely aligned with the new dataset after a 30% length increase. To address accentuated characters in dataset 1's CSV format, UTF-8 encoding was adopted. The new distribution after the union was: S, 2211 (33.11%); P, 2140 (35.90%); N, 1606 (26.95%).

IV. Solution Proposal

A. Preprocessing

For optimal results, preprocess text using conventional techniques: Apply classic data preprocessing – remove, spell check, eliminate stopwords, diagonals, inverted diagonals, numbers, line breaks, parentheses, and double spaces. After initial preprocessing, adapt to Large Scale Language Models (LLM) characteristics. The process consists of two phases. First, models like BERT tokenize text, adding necessary tokens for processing song fragments to capture relationships and contexts. The second phase converts numeric labels to a model-understood format, representing musical traits. Note that LLMs need no extra preprocessing; the aim is accurate context comprehension in song snippets.

B. Classification

It was decided to use the same models as in the first experiments to verify and compare the results. The data set was divided into 80% for training and 20% for validation. The models used for this task were the following:

1) BERT-base-multilingual-cased
2) RoBERTa-base
3) DistilBERT-base-multilingual -cased

Since 2 different data sets were joined, a maximum length of 300 characters was considered, since that was the maximum length used for data set 1.

Table I shows the new configuration for each model:

	Values		
Hyperparameters	**RoBERTa**	**BERT**	**DistilBERT**
Maximum sentence length	300	300	300
Minimum learning rate	0.00004	0.00004	0.00004
Batch size	24	36	48
Evaluation batch size	24	36	48

TABLE I: Modified hyperparameters proposed for the BERT, RoBERTa and DistilBERT

V. Results

The results derived from the datasets and their evaluation through the BERT, RoBERTa and DistilBERT models are presented in table II, making an explicit comparison between each individual dataset, as well as the merged one from our proposal.

It is important to note that a direct comparison with the state of the art is not feasible, since different datasets, preprocessing pipelines and models are involved; therefore, our comparison is directly with our results.

	Accuracy Metric		
Model	**Dataset 1**	**Dataset 2**	**Dataset 3**
RoBERTa	95.66%	91.29%	**96.34%**
BERT	95.58%	91.35%	**96.34%**
DistilBERT	95.83%	92.63%	**95.88%**

TABLE II: Comparative table of the previous experiments and results obtained.

VI. Conclusions and future work

This study focuses on automating categorization of Spanish song texts using models like BERT Multilanguage, RoBERTa, and DistilBERT Multilanguage. These models share similarities in approach. Combining datasets enhances performance, as depicted in Table II.

Consolidating datasets mitigates individual drawbacks, improving overall model performance. Dataset 1 could still be enriched, adding information to enhance language models for better results. In summary, this study illuminates Spanish song text categorization through diverse models, paving the way for refining this technique in future research.

References

[1] Poria, S., Cambria, E., Hazarika, D., Majumder, N., Zadeh, A., & Morency, L. P. (2017). Context-dependent sentiment analysis in user-generated videos. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 873-883). Association for Computational Linguistics. doi: 10.18653/v1/P17-1081.

[2] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[3] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, I. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing System, 6000–6010, 2017.

[5] Srinilta, C., Sunhem, W., Tungjitnob, S., & Thasanthiah, S. (2017). Lyric-based Sentiment Polarity Classification of Thai Songs. Proceedings of the International MultiConference of Engineers and Computer Scientists, 1.

[6] Giras, A., Advirkar, A., Patil, C., Khadpe, D., & Pokhare, A. (2014). Lyrics Based Song Genre Classification. Journal of Computing Technologies, 3. Recuperado de http://jctjournals.com/feb2014/v4.pdf.

[7] Kumar, V., & Minz, S. (2013). Mood classification of lyrics using SentiWordNet. En 2013 International Conference on Computer Communication and Informatics (pp. 1-5). doi: 10.1109/ICCCI.2013.6466307.

[8] Çano, E., & Morisio, M. (2020). Quality of Word Embeddings on Sentiment Analysis Tasks. CoRR, abs/2003.03264. Recuperado de https://arxiv.org/abs/2003.03264.

[9] Pham, V. Q., Nguyen, D. T., & Nguyen, T. H. (2020). CNN-BiLSTM model for document-level sentiment analysis. Journal of Information and Telecommunication.

[10] Sidorov, G., (2019) Corpus Textos de Canciones en español. Laboratorio de Procesamiento de Lenguaje Natural, Centro de Investigación en Computación.

[11] Garcia-Vazquez, O, Alcantara, T, Sidorov, G and Calvo H, A. (2023) Análisis de Sentimientos en textos de canciones en idioma español, Research in Computing Science Issue 152(7) (2023)

[12] Alcántara, T., Garcia-Vazquez, O., Calvo, H. and Desiderio, A. (2023) Corpus Textos de Canciones en español mexicano. Laboratorio de Ciencias Cognitivas Computacionales, Centro de Investigación en Computación.