N&&N at HAHA@IberLEF2021: Determining the Mechanism of Spanish Tweets using ColBERT

Nedaa Alsalman[1] and **Noor Ennab**[2]

[1] **Jordan University of Science and Technology, Irbid, Jordan**
[2] `nmalsalman19, nsabuennab20 @cit.just.edu.jo`

Abstract. In modern technologies humor determination has value for the evaluation of chatbots and personal assistants, the purpose of humor detection is to detect whether a tweet, a sentence or even a word have humor or not. We have participated in HAHA@IberLEF2021 competition for detecting humor in Spanish tweets based on the general linguistic structure of humor task1 and determining the mechanism of these tweets for task 3. So, we used the dataset available on this competition, about 24,000 tweets. Our approach was using ColBERT to generate embeddings for each tweet and then use these embeddings as inputs of hidden layers in a neural network, to helping measure if the tweet is humor or not for task 1. Also, we used sequential Bert for detecting the mechanism of humor for task 3. Our results in evaluation phase for competition given 0.7693 for task 1 (position 12), and 0.0404 for task 3 (position 9).

Keywords: ColBERT· Humor· Mechanism· Spanish· Tweets.

1 Introduction

Humor is considered as the experience to provoke laughter and provides entertainment, it is a good, important aspect of human behavior, which can be expressed in several ways, including text, gestures, and voice signals. It has received a lot of attention as a research study in several areas such as philosophy, psychology, sociology, and linguistics. For computational texts, several methodologies have been developed and studied for classifying them including social media text to identify the meanings of several words, learning about cultures, and the ways of expressing them based on machine learning or deep learning as a new research topic for natural language processing. From the previous studies on computational text, analysis feelings like irony through figurative language[7], classifying tweets depending on the most humorous tweets[9], detecting humor's tweets automatically, and rating them depending on the most humorous one[18][6].
The subject of humor has an ancient history in the human archives. when saying

that humor is the remarkable characteristic of mankind is no exaggeration, It has been the subject of debate for at least two or three thousand years. Humor, considered an umbrella term that captures the irony of laughter with sarcastic humor and comedy[17].

In this we are paper targeting task 1 and task 3 in HAHA@IberLEF2021, task 1 humor detection: deciding whether a tweet has humor or not, second prediction score: the funniness class: calculate a funniness score for a tweet, these tasks are similar to the tasks in HAHA@IberLEF 2018 and 2019, while the other tasks are new, as the main goal in this paper is to classify the mechanism of humor: after classifying a tweet to be funny then predict the mechanism from a set of classes, such as irony, pun, or exaggeration, only one tweet per class is allowed in this task[5].

Figure 1 shows the task objective of our work.
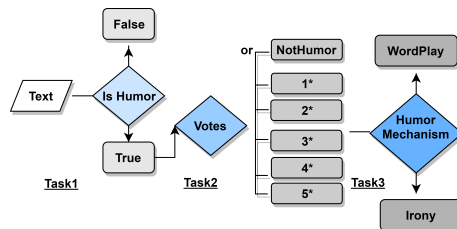
Figure 1: Tasks Workflow

Our approach used sequential Bert model and some prepossessing technique to encode text into word embedding represent as layer and hidden layer that help to understand text constant and feature related to it, then combine layer to predict the correct output. In addition to obtaining an accurate model, to evaluate model performance in detection humor mechanism through F1-score and Macro-F1 score. The structure of the paper is as follows: Section 2 Related work. Section 3 Methodology and that describe works step with details. Section 4 Result of experimental methodology 5 presents our experimental results. Section 6 Conclusion.

2   Related work

(Aniruddha Ghosh et al). 2015, they participated in task 11 for SemEval 2015 about sentiment analysis of figurative language tweet to detected sarcasm and metaphor and evaluated their model worked accuracy using Cosine for similarity and a Mean-Squared-Error(RMSE) measure[7]. (Han et al). 2017, they participated in semeval-2017 task 6 which to understand a sense of humor in text and help to classified it to humor or not, using Naïve Bayes model Multinomial and measure training performance using macro-average recall [9], while (Yan et al.). 2017, they participated on same task but using big ram language models, were better than trigram models on the evaluation data [18]. On the other-hand (Fleşcan Lovin Arseni et al). 2017, they participated on task to find humor tweets through comparing and ordering them debend on which is most

humor using Naïve Bayes and neural networks algorithms[6]., while (Potash et al). 2017, they participated for same task but worked on extract features of these tweet to determine humorous using neural network model [14].

(Ahuja et al). 2018, they suggested depending on classical approaches to classify the text is humor or not and dark humor, using large data set from jokes, machine learning techniques to classified, then distinguish between models accuracy that were achieves, so SVM is best model[1]. While (Liu et al). 2018, they suggested to worked on exploiting the syntactic infrastructure to support humor recognition, have made significant improvements and some features of the syntactic structure are constantly linked to humor, which indicates importance the linguistic phenomena[12].(Ankush Khandelwal et al). 2018, they suggested detecting humor tweets on different languages such as Indian, English with symbols and classified them humorous(H) and non-humorous(N), using n-grams with support vector machine model and bag-of-Words[11].While(Castro et al). 2018, the organizers of HAHA IberLEF 2018, they provide an overview of the competition goals and results, so, it was to detection humor for Spanish tweets through automatic detection and automatic rating them [3].

In (Sushmitha Reddy Sane et al). 2019, they suggested same Ankush Khandelwal et al) but using deep learning model CNN, biLSTM and word2vec, use of bilingual motifs as input and representation it[15].While (Chiruzzo et al). 2019, the organizers of HAHA at IberLEF 2019, they provide an overview of the competition goals and results, so, it was to recognition if a tweet considers humorous or not and giving the result is a cheerful sense of humor or not, [4]. In (Kamrul Hasan et al). 2019, they proposed new database to detect humor through multimedia, vision and audio, depend on natural language processing approach[10].

In (Miller et al).2020, they participated on SemEval-2021 Task 7 to detection and rank humor for short tweets by automatically exploiting judgments of human preference and linguistic annotations using GCPL that was good at English language and applied it in Spanish language in the work[13].While (Winters et al).2020, they suggested making the dataset larger and more complex than prior works especially for Dutch language, so, worked on three set using several algorithms to created texts as pattern or similar jokes in the dataset, then detection the humor text using several languages through models and comparing results in what model was could be able to differentiate the jokes or not, so, the RobBERT was best[16].

(Gupta et al). 2021, they participated in SemEval-2021 Task 7 to show that could develop reasonable humor and detection it was using some model like NLP, BERT, RoBERTa, ERNIE2.0, DeBERTa and XLNet, because there is no widespread as a dis aggregated dataset, most previous work in the field has not explored large neural models for personal understanding of humor[8].

3 Overview

In this section, we summarize our participation as follows:
• Working on dataset available in HAHA@IberLEF2021, then using some pre-processing techniques on dataset to be more fixable to train model.
• Propose an automated approach for humor mechanism detection, then introduce the model architecture and components in detail.
• Evaluate our model and compare its performance with previous worked.
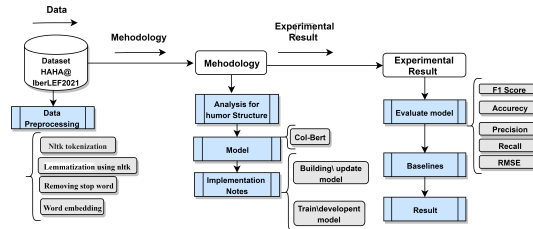Figure 2 Show: overall workflow structure in this paper.

Figure 2: Paper workflow.

4 Data

The dataset CSV files contain tweets in the Spanish language, 24,000 tweets for train, 6,000 tweets for development and others 6,000 for the test. With features ID, tweets is humor or not classifying as 0 or 1, score from one (not funny) to five (excellent), mechanism of humor to conveys and classify humor from a set of classes, and predict the target of the joke.
• Table 1 Numerical analysis of data.
• Table 2 Show the number of tweets for each mechanism category.

count	isHumor	votesNo	votes1	votes2	votes3	votes4	votes5
mean	0.38	2.47	0.79	0.61	0.42	0.18	0.05
std	0.48	11.98	1.80	1.61	1.70	1.00	0.39
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	0.00	1.00	0.00	0.00	0.00	0.00	0.00
50%	0.00	3.00	0.00	0.00	0.00	0.00	0.00
75%	1.00	3.00	1.00	1.00	0.00	0.00	0.00
max	1.00	1290.0	183.0	161.0	208.0	129.0	46.0

Table 1. *Numerical analysis of data*

mechanism type	Number of it in dataset
wordplay	28
unmasking	25
misunderstanding	25
reference	19
absurd	17
exaggeration	16
analogy	14
irony	13
stereotype	9
parody	9
embarrassment	9
insults	7

Table 2. *Number of tweets for each mechanism category.*

4.1 Data preprocessing

Preprocessing techniques are used based on the text data type, start by dividing the given content into smaller parts called token, each word, number, and accentuation marks can be considered as tokens, it is exceptionally valuable for finding patterns and considered as a base step for stemming and lemmatization, the tokenization process make a difference in changing sensitive information to non-sensitive components. Then removing stop words that most common words in a regional language that represent meaningless words or derivation from words to reduced their base or root form. Word embedding algorithms utilizing the pre-trained GloVe Global Vectors for Word Representation , an unsupervised learning process for getting vector representations of words, it permits us to require words from the content, and change each word in that corpus into a sample in high-dimensional space and implies that each comparable word will be assembled. More precisely, to get all the word with the same meaning together.

5 Methodology

This section will target our approach method to determine whether tweets are humorous or not, then find its mechanism. So, in a technical viewpoint, using ColBERT modeling algorithm that takes the input to find the target value.

5.1 Analysis for humor Structure

Starting to view the structure of the tweets, for understanding them, finding words that relate to determining the text humorous or not and helps identify a mechanism for that. For example "Necesito con suma urgencia un calentamiento global para mi casa" this tweet in Spanish language which mean: "I urgently need global warming for my home". There are several approaches that have been taken to understand humor structured and classified it as humor and what its mechanism, our approach in next subsection.

5.2 Models

As following to analysis Structure, the understanding the structure of the tweets to find the meaning that can helping in determining which tweet is humor, by extracting features of it, then take this features as inputs for the model, process them sequentially and represent them as layers, after that using the activation function to design a neural network that helping control all the layers to make a correct prediction that represent as a target output layer by using the classification model that using several paths to extract features for the hidden layers to get final output consider statues of tweets as we need.

Figure 3 displays the architecture of ColBERT model, which contains a set of steps:

• First, take the sentence into two parts left and right in sequential way word by word to find contextual relations between them then tokenize and encode it using Bert, which working to training the model language on input that selecting randomly to predicting 20% or fewer of tokens then split it almost 80% of word that selecting randomly replacing with [MASK] token with 10% of word that selecting randomly and 10% of the input word.

• Taking the tokens as an input to accumulate it with word embedding, and we used other layer type as flatten, dense for the hidden layer in the neural network to sequentially extract the features using the activation function to activate the work of the layer in neural network.

• To predict the output of each sentence after combining all layer, represent it as vector of numbers. Therefore, it will look at the goal to determine the sentence is humor or not and find the mechanism for it.
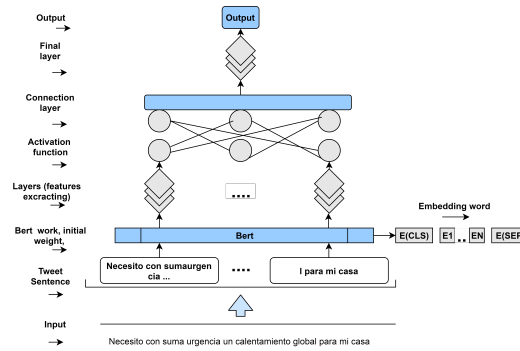

Figure 3: ColBert Work[2]

5.3 Implementation Notes

To deal with clear data and impurities free text, applying some preprocessing techniques from Keras for textual data:

• Convert the text to lowercase then remove stop word such as "is, a, the", after that replace the symbols such as (/()[]—@,;) in the text to space and remove

bad symbols such as (0-9a-z +-)from the text.

• Tokenize the data for top 10000 common words, and cut off sentences after 50 words, then padding the sequences, after that converting categorical labels in humor mechanism to a dummy values such as irony that converted to 01011, finally, shuffle the data.

• Using Glove unsupervised algorithm for embedding word, that allow to take subset of data to put the similar word together.

• Model type 'Bert base uncased' to determine whether the text is humor or not, after tokenizing the data, then fit model using epochs=50, batch-size=32, validation-split=0.1 we have a score of accuracy about 65% therefor, we try different number to get best result. • Bert sequential: starting by determining the number of dense layers, selecting the activation function type, and the dropout(adding regularization) to avoid overfitting in the train, loading the pre-trained word embedding into the embedding layer and freezing other layers.

• Fit the model using learning rate =1e-4,loss= 'categorical cross entropy' for epochs=200, batch-size=1000, validation-split=0.1.

• The evaluation metrics for the model in our colab using TP, TN, FP, FN, F1 Score and MSR from sklearn to find accuracy, then evaluate accuracy and loss function for score using batch-size=32,and verbose=1.

5.4   Experimental Results

As follows to implementation notes subsection, our experiments result in development stage for task 1 using ColBERT model with model type 'bert base uncased'. F1 metric for task 1 in the evaluation phase, which was 0.7693, in 12th place, then comparing our result in HAHAIberLEF-2021 with F1 metrics for 2018 which was 0.7972 and for 2019 which was 0.821. Task 3 which new ideas and main goal for this paper F1 metrics was 0.0404 in 9th place.

6   Conclusion

When saying that humor is the remarkable characteristic of mankind is no exaggeration. It has been the subject of debate for at least two or three thousand years. We used sequential Bert for detecting the mechanism of humor base on the given data set on HAHA@IberLEF2021 competition for detecting humor in Spanish tweets based on the general linguistic structure of humor, and we proposed Col-BERT to generate embeddings word from these tweets and using these embeddings as inputs of hidden layers in a neural network, to find our target output.

References

1. Vikram Ahuja, Taradheesh Bali, and Navjyoti Singh. What makes us laugh? investigations into automatic humor classification. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 1–9, 2018.

2. Issa Annamoradnejad and Gohar Zoghi. Colbert: Using bert sentence embedding for humor detection. *arXiv preprint arXiv:2004.12765*, 2020.

3. Santiago Castro, Luis Chiruzzo, and Aiala Rosá. Overview of the haha task: Humor analysis based on human annotation at ibereval 2018. In *IberEval@ SEPLN*, pages 187–194, 2018.

4. Luis Chiruzzo, Santiago Castro, Mathias Etcheverry, Diego Garat, Juan José Prada, and Aiala Rosá. Overview of haha at iberlef 2019: Humor analysis based on human annotation. In *IberLEF@ SEPLN*, pages 132–144, 2019.

5. Luis Chiruzzo, Santiago Castro, Santiago Góngora, Aiala Rosá, J. A. Meaney, and Rada Mihalcea. Overview of HAHA at IberLEF 2021: Detecting, Rating and Analyzing Humor in Spanish. *Procesamiento del Lenguaje Natural*, 67(0), 2021.

6. Iuliana Alexandra Fleşcan-Lovin-Arseni, Ramona Andreea Turcu, Cristina Sirbu, Larisa Alexa, Sandra Maria Amarandei, Nichita Herciu, Constantin Scutaru, Diana Trandabat, and Adrian Iftene. # warteam at semeval-2017 task 6: Using neural networks for discovering humorous tweets. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 407–410, 2017.

7. Aniruddha Ghosh, Guofu Li, Tony Veale, Paolo Rosso, Ekaterina Shutova, John Barnden, and Antonio Reyes. Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 470–478, 2015.

8. Aishwarya Gupta, Avik Pal, Bholeshwar Khurana, Lakshay Tyasgi, and Ashutosh Modi. Humor@ iitk at semeval-2021 task 7: Large language models for quantifying humor and offensiveness. *arXiv preprint arXiv:2104.00933*, 2021.

9. Xiwu Han and Gregory Toner. Qub at semeval-2017 task 6: Cascaded imbalanced classification for humor analysis in twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 380–384, 2017.

10. Md Kamrul Hasan, Wasifur Rahman, Amir Zadeh, Jianyuan Zhong, Md Iftekhar Tanveer, Louis-Philippe Morency, et al. Ur-funny: A multimodal language dataset for understanding humor. *arXiv e-prints*, pages arXiv–1904, 2019.

11. Ankush Khandelwal, Sahil Swami, Syed S Akhtar, and Manish Shrivastava. Humor detection in english-hindi code-mixed social media content: Corpus and baseline system. *arXiv preprint arXiv:1806.05513*, 2018.

12. Lizhen Liu, Donghai Zhang, and Wei Song. Exploiting syntactic structures for humor recognition. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1875–1883, 2018.

13. Tristan Miller, Erik-Lân Do Dinh, Edwin Simpson, and Iryna Gurevych. Predicting the humorousness of tweets using gaussian process preference learning. *Procesamiento del Lenguaje Natural*, 64:37–44, 2020.

14. Peter Potash, Alexey Romanov, and Anna Rumshisky. Semeval-2017 task 6:# hashtagwars: Learning a sense of humor. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 49–57, 2017.

15. Sushmitha Reddy Sane, Suraj Tripathi, Koushik Reddy Sane, and Radhika Mamidi. Deep learning techniques for humor detection in hindi-english code-mixed tweets. In *Proceedings of the Tenth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 57–61, 2019.

16. Thomas Winters and Pieter Delobelle. Dutch humor detection by generating negative examples. *arXiv preprint arXiv:2010.13652*, 2020.

17. Zhihui Wu. The laughter-eliciting mechanism of humor. *English Linguistics Research2*, 1:52–63, 2013.

18. Xinru Yan and Ted Pedersen. Duluth at semeval-2017 task 6: Language models in humor detection. *arXiv preprint arXiv:1704.08390*, 2017.