

# Humor Detection in Spanish Tweets Using Neural Network

Rida Miraj<sup>1,2</sup>[0000-0002-1605-0403] and Masaki Aono<sup>1,3</sup>[0000-0003-1383-1076]

<sup>1</sup> Toyohashi University of Technology, Toyohashi, Japan

<sup>2</sup> ridamiraj974@gmail.com

<sup>3</sup> aono@tut.jp

**Abstract.** With its linguistic, social, and psychological components, humor has long been a part of human life. Allowing computers to interpret humor has become extremely important due to its wide range of uses and growing popularity on social media platforms. While humor has long been researched from psychological, cognitive, and linguistic perspectives, computational linguistics has yet to investigate it. Our contribution to HAHA@IberLEF2021: Humor Analysis based on Human Annotation is described in this paper. We offer a deep neural network-based approach in this paper. Our research team uses multi-kernel convolution recurrent neural network model for the humor detection in tweets. We examine our method's performance and show how each component of our design contributes to its overall success.

**Keywords:** Humor · Recurrent Neural Network · Text · CNN · tweets.

## 1 Introduction

Humor is a universal and subtle emotion that can be found all over the world. The majority of prior research and studies on humor difficulties were focused on binary categorization or the identification of linguistic features. Purandare and Litman utilized typical supervised learning classifiers to recognize hilarious speech in a hilarious spoken dialogue as data from a famous comedy television show [18]. Taylor and Mazlack used the methodology that was based on the extraction of structural patterns and peculiar structure of jokes newcite [20]. Luke de Oliveira and Láinez applied recurrent neural network (RNN) and convolutional neural networks (CNNs) to humor detection from reviews in Yelp dataset [9].

Because tweets are brief and informal user-generated text that typically do not follow grammatical standards, detecting humor in them presents particular problems to the research community. Furthermore, tweets feature a plethora of

---

*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).

unique abbreviations as well as Twitter-specific syntaxes like hashtags and emojis. To address the challenge of humor detection, Chiruzzo et al. [8] presented a task that focuses on detecting humor in Spanish tweets. Various subtasks related to automated humor detection are proposed in the present HAHA assessment campaign. In this paper, we propose a neural network method that combines the multi-kernel convolution with the Bi-LSTM. Experimental results on given Spanish tweets demonstrate the submitted results of our framework.

The rest of the paper is structured as follows: **Section 2** presents a summary of previous studies. In **Section 3**, we introduce our proposed humor detection framework. **Section 4** includes experiments and evaluations. Some concluded remarks of our work are described in **Section 5**.

## 2 Related Research

In the related work of humor identification, there are a lot of work that is done over the year which includes statistical and N-gram analysis [20], Regression Trees [18], Word2Vec combined with K-NN Human Centric Features, and Convolutional Neural Networks [6]. When working with a limited number of characteristics, neural networks function exceptionally effectively. When dealing with changing length sequences, sequence variants to prior states, as in recurrent neural networks, can be introduced to the network. To identify jokes from non-jokes, several humor detection algorithms include hand-crafted (typically word-based) characteristics [21, 12, 3, 15]. Such word-based features work well when the non-joke dataset contains terms that are entirely distinct from the humor dataset. According to humor theory, the sequence of words matters, because announcing the punchline before the setup would merely lead to the discovery of the second interpretation of the joke, causing the joke to lose its humorous component [19]. For some years, figurative language, particularly humor, has been a fruitful field of research in the domain of shared tasks. One of the problems provided by figurative language, such as metaphors and irony, was discussed in Semeval-2015 Task 11 [10]: its influence on Sentiment Analysis. Semeval-2017 Task 6 [17] supplied participants with hilarious tweets sent to a comedy show, and asked them to guess how the audience and producers of the show would rate the tweets. The HAHA task was arranged by Grupo PLN-InCo at two different conferences: IberEVAL 2018 [5]. There was two subtasks: humor detection and prediction of funniness score. SemEval-2021 Task 7 [11] is a newer task that combines humor detection with offense detection. It includes all of the subtasks from HAHA 2018 and 2019[7], plus two new ones: Offense Score Prediction and Controversial Humor Classification. In [8], there were four subtasks, from which we participated in two of them.

### 3 Framework

In this section, we describe the details of our proposed framework for humor detection in Spanish tweets. Figure 1 depicts an overview of our proposed framework.

We utilize the pre-trained-word-vectors-for-spanish word embedding purposes. The embedding matrix is fed into the embedding layer of our neural network. We start by extracting higher-level feature sequences from the target added tweet embeddings using multi-kernel convolution filters. These feature sequences are supplied into the Bi-LSTM that is linked to it. Following that, we go through each component in detail.

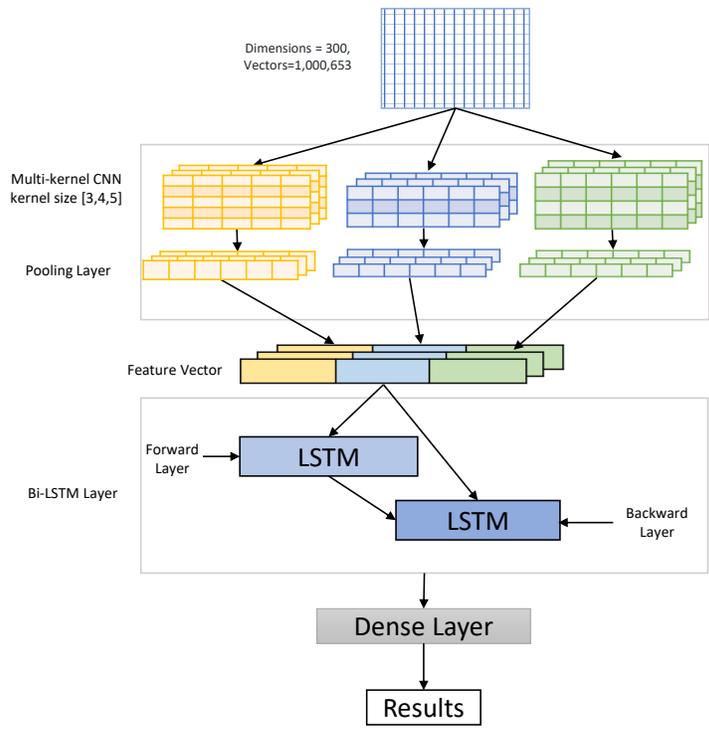


Fig. 1. Proposed framework.

#### 3.1 Embedding

Starting with random weights, the Embedding layer in Figure 1 will learn an embedding for all of the words in the training dataset. The initial hidden layer

of a network is defined as this flexible layer. A pre-trained model used in the embedding layer simply required a file containing tokens and their associated word vectors. The pre-trained word vectors for Spanish model was built using 300-dimensional word vectors. Dimensionality is a term that refers to how many dimensions there are in anything. The Embedding matrix will have a dimensionality of  $L \times D$ , where  $L$  is the sentence length and  $D$  is the word-vector dimension.

### 3.2 Convolution Neural System

We use the technique given by [13] to extract higher-level features in our multi-kernel convolution. The embedding matrix created in the embedding layer is the module’s input. Then, using a filter, we apply convolution on it. We use three distinct kernel sizes, or the size of the convolution filters, to apply multiple convolutions: 3, 4, and 5. Each filter creates the matching feature maps after performing convolutions, and then a max-pooling function is used to build a univariate feature vector. Finally, each kernel’s feature vectors are concatenated to create a single high-level feature vector.

### 3.3 Bi-LSTM

Bidirectional Long Short Term Memory (Bi-LSTM) is a bidirectional variant of LSTM seen in the center of Figure 1. Bi-LSTM combines the forward and backward hidden layers, allowing access to both the previous and subsequent contexts. The Bi-LSTM neural network is used to obtain a vector representation of the input sentence that captures the semantics of the phrase effectively. The final result from Bi-LSTM’s output layer is formed by merging the results from both RNN hidden layers, namely the forward and backward layers.

### 3.4 Humor classification and Prediction

We get our results from the last linear layer of the model. We consider binary cross-entropy and mean square error (mse) as the loss function in sub-task1 and sub-task2, respectively. We use the stochastic gradient descent (SGD) to learn the model parameter and adopt the Adam optimizer [14].

**Table 1.** Results of Sub-task1 and Sub-task2 with other teams results.

Team	F1	<i>RMSE</i>
<b>Baseline</b>	0.6493	0.6532
<b>Our Framework</b>	0.7441	1.5164
Jocoso	0.8850	0.6296
icc	0.8716	0.6853
RoBERToCarlos	0.7961	0.8602

## 4 Evaluation

### 4.1 Dataset

The organizer provide a corpus of crowd-annotated tweets divided into three subsets for tasks 1 and 2: training (24,000 tweets), development (4,000 tweets), and testing (6,000 tweets). The annotation has a voting system in which users can choose from six different choices. The tweet is either not funny or funny, with a score ranging from one (not funny) to five (very hilarious) (excellent).

To prepare the data, we eliminated stop words using NLTK’s standard stoplist, eliminated special characters, and performed hashtag segmentation using the hashtag segmentation tool [2]. The fixed length of a sentence was set in the beginning of embedding technique.

### 4.2 Model Configuration

In the following, we describe the set of parameters that we have used in our framework during experiments. We used one embedding model to initialize the word embeddings in the embedding layer. The embedding model has 300-dimensional with 1,000,653 vectors. It is trained on Spanish Billion Word Corpus which has the size of 1.4 billion words [4]. For the multi-kernel convolution, we employed 3 kernel sizes (3,4,5), and the number of filters was set to 36. The framework which we used to design our model was based on TensorFlow [1] and training of our model is done on a GPU [16] to capture the benefit from the efficiency of parallel computation of tensors. We trained our model for a max of 50 epochs with a batch size of 64 and an initial learning rate of 0.001 by Adam optimizer. In this paper, we reported the results based on these settings. Unless otherwise stated, default settings were used for the other parameters.

### 4.3 Results and Analysis

Our target is to classify the tweets into humorous or not humorous from Spanish tweets. In Table 1, at first we reported the results of sub-task1 and sub-task2 based on a Naive Bayes with tfidf features and SVM regression with tfidf features, respectively. Next, we reported the results of our proposed framework that were submitted in the competition. After the competition, some parameters were changed to check the betterment of results, and we found that we could improve our results so far to 77.453 percent F1 for sub-task1, and 0.6977 RMSE error for sub-task2, although these results could not be submitted to the competition.

## 5 Conclusion

Our technique for HAHA@IberLEF2021: Humor Analysis based on Human Annotation Forum was described in this article. Humor detection is a difficult process. We used deep learning techniques to try to solve the problem. We ran

some tests with other models, such as a basic regression model and a multi-layer perceptron model, however the model described in this paper was the one that produced the best results. In a summary, our unified framework’s key contribution is that it successfully learns contextual information, which improves comedy detection performance. We want to leverage external data to generalize our model for comedy identification in the same region in the future.

## Acknowledgments

This research was supported by the Japan International Cooperation Agency – JICA under Innovative Asia program.

## References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). pp. 265–283 (2016)
2. Baziotis, C., Pelekis, N., Doukeridis, C.: DataStories at SemEval-2017 task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). pp. 747–754. Association for Computational Linguistics, Vancouver, Canada (Aug 2017). <https://doi.org/10.18653/v1/S17-2126>, [bluehttps://www.aclweb.org/anthology/S17-2126](https://www.aclweb.org/anthology/S17-2126)
3. van den Beukel, S., Aroyo, L.: Homonym detection for humor recognition in short text. In: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. pp. 286–291 (2018)
4. Cardellino, C.: Spanish Billion Words Corpus and Embeddings (August 2019), [bluehttps://crscardellino.github.io/SBWCE/](https://crscardellino.github.io/SBWCE/)
5. Castro, S., Chiruzzo, L., Rosá, A.: Overview of the haha task: Humor analysis based on human annotation at ibereval 2018. In: IberEval@ SEPLN. pp. 187–194 (2018)
6. Chen, P.Y., Soo, V.W.: Humor Recognition Using Deep Learning pp. 113–117 (2018). <https://doi.org/10.18653/v1/n18-2018>
7. Chiruzzo, L., Castro, S., Etcheverry, M., Garat, D., Prada, J.J., Rosá, A.: Overview of haha at iberlef 2019: Humor analysis based on human annotation. In: IberLEF@ SEPLN. pp. 132–144 (2019)
8. Chiruzzo, L., Castro, S., Góngora, S., Rosá, A., Meaney, J.A., Mihalcea, R.: Overview of HAHA at IberLEF 2021: Detecting, Rating and Analyzing Humor in Spanish. *Procesamiento del Lenguaje Natural* **67**(0) (2021)
9. De Oliveira, L., Rodrigo, A.L.: Humor detection in yelp reviews. Retrieved on December **15**, 2019 (2015)
10. Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., Reyes, A.: SemEval-2015 task 11: Sentiment analysis of figurative language in Twitter. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). pp. 470–478. Association for Computational Linguistics, Denver, Colorado (Jun 2015). <https://doi.org/10.18653/v1/S15-2080>, [bluehttps://www.aclweb.org/anthology/S15-2080](https://www.aclweb.org/anthology/S15-2080)

11. Gupta, A., Pal, A., Khurana, B., Tyagi, L., Modi, A.: Humor@iitk at semeval-2021 task 7: Large language models for quantifying humor and offensiveness (04 2021)
12. Kiddon, C., Brun, Y.: That’s what she said: double entendre identification. In: Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies. pp. 89–94 (2011)
13. Kim, Y.: Convolutional Neural Networks for Sentence Classification pp. 1746–1751 (2014)
14. Kingma, D.P., Ba, J.L.: A : a m s o pp. 1–15 (2015)
15. Mihalcea, R., Strapparava, C.: Making computers laugh: Investigations in automatic humor recognition. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. pp. 531–538 (2005)
16. Owens, J.D., Houston, M., Luebke, D., Green, S., Stone, J.E., Phillips, J.C.: Gpu computing. Proceedings of the IEEE **96**(5), 879–899 (2008)
17. Potash, P., Romanov, A., Rumshisky, A.: SemEval-2017 task 6: #HashtagWars: Learning a sense of humor. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). pp. 49–57. Association for Computational Linguistics, Vancouver, Canada (Aug 2017). <https://doi.org/10.18653/v1/S17-2004>, [bluehttps://www.aclweb.org/anthology/S17-2004](https://www.aclweb.org/anthology/S17-2004)
18. Purandare, A., Litman, D.: Humor: Prosody analysis and automatic recognition for f\* r\* i\* e\* n\* d\* s. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. pp. 208–215 (2006)
19. Ritchie, G.: Developing the incongruity-resolution theory. Tech. rep. (1999)
20. Taylor, J.M., Mazlack, L.J.: Computationally Recognizing Wordplay in Jokes Theories of Humor (1991) (2000)
21. Taylor, J.M., Mazlack, L.J.: Computationally recognizing wordplay in jokes. In: Proceedings of the Annual Meeting of the Cognitive Science Society. vol. 26 (2004)