UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

TESIS FORMATO ARTICULO TITULADA:
**DEEP NEURAL NETWORK APPROACHES FOR SPANISH
SENTIMENT ANALYSIS OF SHORT TEXTS**

Tesis presentada por el bachiller:

DISRAELI FAUSTO ARI MAMANI

Para optar el Título profesional de

INGENIERO DE SISTEMAS.

ASESOR: DR. JOSÉ EDUARDO
OCHOA LUNA.

AREQUIPA-PERÚ

2019

Agradecimientos

En primer lugar, agradecer a Dios por haberme guiado a lo largo de estos años de estudio y darme la fortaleza para continuar a pesar de las dificultades, darme la gracia de tener a mis seres queridos cerca brindándome su apoyo, también por permitirme conocer personas en el camino que hicieron posible mi progreso y crecimiento en mi camino como profesional.

También agradezco de forma muy especial a mi asesor Dr. José Eduardo Ochoa Luna por su apoyo, enseñanzas y guía para poder completar esta investigación, sus conocimientos, orientación y paciencia han sido fundamentales para mi formación como investigador.

Gracias a mi enamorada Lesli, por tanto apoyo, gracias a ella porque en todo momento fue un apoyo incondicional en mi vida, brindándome esa fortaleza necesaria para no rendirme y poder cumplir mis metas, nunca podre terminar de agradecerle por todo su apoyo. Gracias también a mis padres, que fueron mi motivación para continuar, brindándome ese apoyo tan importante como familia.

Gracias a mi universidad, por ser mi hogar de formación, gracias a todas las personas que participaron en mi formación y que me dieron ese aporte para poder culminar mi paso por la universidad, mi alma máter a la que siempre guardare cariño y agradecimiento.

RESUMEN

DEEP NEURAL NETWORK APPROACHES FOR SPANISH SENTIMENT
ANALYSIS OF SHORT TEXTS

Disraeli Fausto Ari Mamani

Arequipa-Perú

Asesor: José Eduardo Ochoa Luna.

El análisis de sentimientos se ha investigado ampliamente en los últimos años. Si bien se han obtenido importantes resultados teóricos y prácticos, todavía hay margen de mejora. En particular, cuando se considera oraciones cortas y lenguas de bajos recursos. Por lo tanto, en este trabajo nos centramos en el análisis de sentimientos para los mensajes de Twitter en español. Exploramos la combinación de varias representaciones de palabras (Word2Vec, Glove, Fas-text) y modelos de redes neuronales profundas para clasificar textos cortos.

Los enfoques anteriores de Deep Learning no pudieron obtener resultados óptimos para la clasificación de frases en español en Twitter. Por el contrario, mostramos resultados prometedores en esa dirección. Nuestra mejor configuración combina aumento de datos, representaciones de inserción de tres palabras, redes neuronales convolucionales y redes neuronales recurrentes. Esta configuración nos permite obtener resultados de vanguardia en el conjunto de datos de referencia español TAS-S/SEPLN, en términos de precisión.

Palabras clave

Redes neuronales profundas, Análisis de sentimientos, Oraciones de Twitter

ABSTRACT

DEEP NEURAL NETWORK APPROACHES FOR SPANISH SENTIMENT
ANALYSIS OF SHORT TEXTS

Disraeli Fausto Ari Mamani
Arequipa-Perú

Advisor: José Eduardo Ochoa Luna.

Sentiment Analysis has been extensively researched in the last years. While important theoretical and practical results have been obtained, there is still room for improvement. In particular, when short sentences and low resources languages are considered. Thus, in this work we focus on sentiment analysis for Spanish Twitter messages. We explore the combination of several word representations (Word2Vec, Glove, FastText) and Deep Neural Networks models in order to classify short texts.

Previous Deep Learning approaches were unable to obtain optimal results for Spanish Twitter sentence classification. Conversely, we show promising results in that direction. Our best setting combines data augmentation, three word embeddings representations, Convolutional Neural Networks and Recurrent Neural Networks. This setup allows us to obtain state-of-the-art results on the TASS/SEPLN Spanish benchmark dataset, in terms of accuracy.

Keywords

Deep neural networks, Sentiment analysis, Twitter sentences

Índice general

Agradecimientos	i
Resumen	ii
Abstract	iii
1 Deep neural network approaches for spanish sentiment analysis of short texts	1
1.1 Autor	1
1.2 Asesor	1
1.3 Planteamiento del problema	2
1.4 Objetivos	3
1.4.1 Objetivo General	3
1.4.2 Objetivos Específicos	3
1.5 Descripción del proyecto	4
1.5.1 Análisis de Sentimientos	4
1.5.2 Deep Learning	4
1.5.3 Word Embeddings	4
1.5.4 Procedimiento	5

1.5.4.1	Pre-Procesamiento	5
1.5.4.2	Características no supervisadas	6
1.5.4.3	Clasificador CNN(Convolutional Neuronal Network)	
	– RNN(Recurrent Neuronal Network)	6
1.6	Justificación	6
1.7	Delimitaciones	8
	Referencias	9

Capítulo 1

Deep neural network approaches for spanish sentiment analysis of short texts

1.1 Autor

- **Apellidos y Nombres:** Ari Mamani Disraeli Fausto

1.2 Asesor

Asesor:	Dr. José Eduardo Ochoa Luna.
Grado Académico:	Doctor en Ciencias por la Universidad de Sao Paulo – Brasil
Reseña del autor:	Doctor en Ciencias por la Universidad de Sao Paulo – Brasil, Magister en Ciencia de la Computación por la Universidade Federal do Mato Grosso do Sul - Brasil, Bachiller en Ingeniería de Sistemas por la Universidad Nacional de San Agustín - Perú. Actualmente se dedica a la investigación en Ciencia de la Computación en el Centro de Investigación e Innovación en Ciencia de la Computación (RICS) de la Universidad Católica San Pablo (UCSP) - Perú.

Reseña del autor:	Es autor de más de 20 artículos de conferencias y journals. Ha participado en 4 proyectos de investigación en Brasil y Perú. Es docente a nivel de pregrado y posgrado en la UCSP, y ha asesorado tesis a nivel de pregrado, maestría y doctorado. Índice H = 7.
-------------------	--

1.3 Planteamiento del problema

El análisis de sentimientos es un área muy importante en el procesamiento de lenguaje natural, consiste en analizar un texto y asignarle una etiqueta que indique si un comentario es positivo o negativo, en ocasiones también puede indicar si es neutro o que no expresa sentimiento. Esta tarea podría efectuarse manualmente, etiquetando un texto luego de interpretarlo, sin embargo, en los últimos años la cantidad de texto a interpretar se ha incrementado por las fuentes de información como Facebook^[1], Twitter^[2], foros y otras fuentes que hacen casi imposible la tarea de clasificar texto manualmente, ya que cada segundo se va generando más información [10].

La tarea de clasificación de sentimientos se ha intentado resolver con menor precisión, por ejemplo, usando localización de palabras clave, que consiste en seleccionar palabras que por sí solas ya expresan un sentimiento como feliz o triste y que directamente indican la polaridad de un texto. También se usó afinidad léxica que como el anterior intenta asociar palabras a una afinidad de emociones para poder indicar un sentimiento, luego se emplearon métodos estadísticos que son técnicas como SVM[17], NN[13], BoW[11] y orientaciones semánticas. Existen también herramientas que procesan texto para generar análisis de sentimiento, la mayoría de estas herramientas solo esta disponible para el idioma inglés, algunas de estas herramientas son SentiWordNet[3], SenticNet[2], que son modelos pre entrenados con métodos estadísticos y que otorgan una clasificación a un texto de entrada.

Deep Learning ha solucionado muchos proceso de automatización y se ha apli-

¹<https://es-la.facebook.com/>

²<https://twitter.com/?lang=es>

cado en diversos campos, principalmente en el procesamiento de imágenes, consiste en extraer características simples en una capa y luego combinarlas con otras características de una siguiente capa, de forma que se pueda extraer información más compleja con cada nivel[9]. Este concepto es aplicado en texto, donde cada entrada es transformada a un vector característico usando Word Embeddings que consiste en representar cada palabra por un vector[8], combinando las dimensiones de la longitud del vector por palabra y la longitud de la entrada de texto, se puede construir una matriz donde se puede trabajar con Deep Learning para descubrir características complejas para realizar una clasificación de sentimientos.

1.4 Objetivos

1.4.1 Objetivo General

Proponer un clasificador eficiente usando y combinando técnicas de Deep Learning que permita clasificar sentimientos en textos cortos para el idioma español, con resultados comparables al estado del arte.

1.4.2 Objetivos Específicos

- Seleccionar las características principales de un texto en el idioma español.
- Proponer el uso de Word Embeddings para extraer características complejas de un texto.
- Encontrar la mejor configuración usando Deep Learning para obtener buenos resultados en la clasificación de textos cortos en español.
- Comparar el desempeño del clasificador con respecto a técnicas similares que usan Deep Learning.

1.5 Descripción del proyecto

1.5.1 Análisis de Sentimientos

El Análisis de sentimientos es el uso de técnicas para identificar información subjetiva de un texto, es decir polaridades, como positiva o negativa. Algunas de las tareas de análisis de sentimientos incluyen el estado neutro y el estado que no indica sentimiento, este sentimiento se determina en relación con un tema y un lenguaje.

Figura 1.1: Interpretación de texto [5]: Clasificación de sentimiento de un texto.

1.5.2 Deep Learning

Deep Learning intenta representar el funcionamiento del sistema nervioso humano. Busca encontrar características de un modo jerárquico anidando conceptos, seleccionando características simples en una capa inicial que sirven para calcular características más abstractas en una siguiente capa, con el número exacto de capas se puede encontrar representaciones más abstractas y complejas de información.

1.5.3 Word Embeddings

Word Embeddings consiste en la transformación de una palabra en un vector de n -dimensiones, esto se hace con el fin de que el vector generado pueda representar un modelo matemático de la palabra y que podamos usarlo como un modelo matemático que conserve sus características. El primer modelo de Word Embedding, se podría decir, fue el modelo de bolsa de palabras, donde un diccionario puede ser

Figura 1.2: Deep Learning[1]: Deep Learning es un subconjunto de Machine Learning que logra un gran poder y flexibilidad al aprender a representar el mundo como una jerarquía anidada de conceptos

representado como un vector binario de forma hot-encoding y los actuales como Word2Vec[4] y glove[6] que usan relaciones de coocurencia y skipgrams.

Figura 1.3: Palabras en campos vectoriales

1.5.4 Procedimiento

1.5.4.1 Pre-Procesamiento

- Se hace uso de un tokenizador especial para texto de Twitter (NLTK Tweet-Tokenizer)
- Divide las oraciones en tokens, de esta forma que se eliminan sectores de la entrada que no aportan información al clasificador.

- Se eliminan URLs y links, también se extraen palabras propias de twitter como el # y @ y son transformadas en entidades.
- Los emoticones son sustituidos por palabras clave.
- También se eliminan palabras y símbolos repetidos usando expresiones regulares.
- Palabras de alta y baja frecuencia son eliminados del clasificador.

1.5.4.2 Características no supervisadas

Las oraciones luego de pasar por el procedimiento de pre procesamiento son usadas para encontrar una representación de palabra asociada, se usan 3 técnicas de representación de palabras, que posteriormente se convertirán en canales de una red neuronal convolucional, los 3 canales de la red estarán conformados por los vectores de Word2Vec, Glove y FastText[14].

1.5.4.3 Clasificador CNN(Convolutional Neuronal Network) – RNN(Recurrent Neuronal Network)

el clasificador usa un modelo en serie de una Red Neuronal Convolucional[16] y una Red Neuronal Recurrente[12] como se puede ver en la figura.

La Entrada de la CNN esta conformada por 3 canales (3D CNN), cada canal corresponde a una representación de palabras distinta. La capa de CNN es conectada secuencialmente a un RNN, luego de una capa de max pooling. La capa final de la RNN es la encargada de clasificar el texto en sus 4 salidas (polaridades.)

1.6 Justificación

En las redes sociales, foros y otros medios de información se genera información a cada hora, en estos datos se pueden realizar tareas como análisis de datos, tendencias, temas de interés y otros temas relacionados a la minería de estos datos. La toma de

Figura 1.4: Clasificador para textos cortos en español usando un combinación CNN-RNN[7]

decisiones es una de las principales tareas a las que apoya el análisis de sentimientos, otorgando luego de un entrenamiento con datos correctos, asistencia para reducir riesgos en la toma de decisiones[15].

El análisis de tendencias también puede ser trabajado por el análisis de sentimientos a fin de conocer tendencias que ayuden, por ejemplo, a empresas tomar decisiones de lanzar sus productos al mercado. Para procesar esta gran cantidad de datos se requiere tecnologías recientes que analicen y clasifiquen con un grado de exactitud alta los textos a fin de asignarle una categoría de sentimiento correcto o muy cercana a la de una interpretación humana.

Existen más documentos y propuestas para el idioma Inglés que para el idioma español, los modelos creados para el análisis de sentimientos cambian de idioma a idioma, por las características semánticas y sintácticas propias de cada idioma. Por lo que es necesario la creación de un modelo que seleccione características propias del idioma español.

1.7 Delimitaciones

Word Embeddings es una potente técnica de descubrimiento de características en texto, pero no existe a la fecha de la publicación de este documento, vectores preparados para el idioma español.

Muchas investigaciones se han centrado en el idioma inglés, por lo que los conjuntos de datos para hacer comparaciones y entrenamiento son limitadas.

Referencias

- [1] deep-learning, <https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063>.
- [2] BIAGIONI, R. Senticnet. *The SenticNet Sentiment Lexicon: Exploring Semantic Richness in Multi-Word Concepts SpringerBriefs in Cognitive Computation* (2016).
- [3] DENECKE, K. Using sentiwordnet for multilingual sentiment analysis. *2008 IEEE 24th International Conference on Data Engineering Workshop* (2008).
- [4] [HTTPS://CODE.GOOGLE.COM/ARCHIVE/P/WORD2VEC/](https://code.google.com/archive/p/word2vec/). word2vec.
- [5] [HTTPS://WWW.LEXALYTICS.COM/TECHNOLOGY/SENTIMENT](https://www.lexalytics.com/technology/sentiment). lexalytics.
- [6] JEFFREY PENNINGTON, RICHARD SOCHER, C. D. M. Glove: Global vectors for word representation.
- [7] JOSÉ OCHOA-LUNA, D. A. Deep neural network approaches for spanish sentiment analysis of short texts. *Advances in Artificial Intelligence - IBERAMIA 2018* (2018).
- [8] JURDZINSKI, G. Word embeddings for morphologically complex languages. *Schedae Informaticae* (2017).
- [9] KETKAR, N. Introduction to deep learning. *Deep Learning with Python* (2017).
- [10] MASSUNG, C. Z.-S. *Book Title Text data management and analysis: a practical introduction to information retrieval and text mining*. Acm Books, 2016.

-
- [11] MIAOULIS, D. P.-G. *Intelligent computer graphics 2012*. Springer, 2013.
 - [12] OSCAR ARAQUE, RODRIGO BARBADO, J. F. S.-R. C. A. I. Applying recurrent neural networks to sentiment analysis of spanish tweets. *TASS 2017* (2017).
 - [13] PARKER, D. *Neuronal network analyses: progress, problems, and uncertainties*. Royal Society, 2010.
 - [14] ROSA MARÍA MONTAÑÉS SALAS, RAFAEL DEL-HOYO ALONSO, J. V.-M. M. R. A. G. F. J. L.-P. Fasttext como alternativa a la utilización de deep learning en corpus pequeños. *TASS 2017* (2017).
 - [15] SALLES, M. Decision-making. *Decision-Making and the Information System* (2015).
 - [16] SEGURA-BEDMAR, I., Q. A. M.-P. Exploring convolutional neural networks for sentiment analysis of spanish tweets. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. pp. 1014–1022. Association for Computational Linguistics* (2017).
 - [17] XUE, Z. W.-X. Multi-class support vector machine. *Support Vector Machines Applications* (2014).

Deep Neural Network Approaches for Spanish Sentiment Analysis of Short Texts

José Ochoa-Luna¹^[0000–0002–8979–3785] and Disraeli Ari²

¹ Department of Computer Science
Universidad Católica San Pablo, Arequipa, Perú
jechoa@ucsp.edu.pe

² Universidad Nacional de San Agustín, Arequipa, Perú
darim@unsa.edu.pe

Abstract. Sentiment Analysis has been extensively researched in the last years. While important theoretical and practical results have been obtained, there is still room for improvement. In particular, when short sentences and low resources languages are considered. Thus, in this work we focus on sentiment analysis for Spanish Twitter messages. We explore the combination of several word representations (Word2Vec, Glove, Fastext) and Deep Neural Networks models in order to classify short texts. Previous Deep Learning approaches were unable to obtain optimal results for Spanish Twitter sentence classification. Conversely, we show promising results in that direction. Our best setting combines data augmentation, three word embeddings representations, Convolutional Neural Networks and Recurrent Neural Networks. This setup allows us to obtain state-of-the-art results on the TASS / SEPLN Spanish benchmark dataset, in terms of accuracy.

Keywords: Deep Neural Networks · Sentiment Analysis · Twitter sentences.

1 Introduction

Spanish is the third language most used on the Internet³. However, the development of Natural Language Processing (NLP) techniques for this language did not follow the same trend. In particular, this research gap can be observed in Spanish *sentiment analysis*. Sentiment analysis allows us to perform an automated analysis of millions of reviews. Its basic task, called polarity detection, targets at determining whether a given opinion is positive, negative or neutral. This area has been widely researched since 2002 [16]. In fact, it is one of the most active research areas in NLP, data mining and social media analytics [27].

Polarity detection has been addressed as a text classification problem thus, can be approached by supervised and unsupervised learning methods [29]. In the unsupervised approach, a vocabulary of positive and negative words is constructed so as to polarity is inferred according to the similarity between vocabulary and opinionated words. The second approach is based on machine learning,

³ <http://www.internetworldstats.com/stats7.htm>

training data and labelled reviews are used to define a classifier [16]. This last approach relies heavily on feature engineering. However, recent learning representation paradigms perform these tasks automatically [15]. In this context, Machine Learning has recently become the dominant approach for sentiment analysis, due to availability of data, better models and hardware resources [28].

In this paper we adopt a Deep Learning approach for sentiment analysis. In particular we aim at performing automated classification of short texts in Spanish. This is challenging because of the limited contextual information that they normally contain.

To do so, sentence words are mapped to word embeddings. Distributional approaches such as word embeddings have proven useful to model context in several NLP tasks [18]. Three kinds of word representations (Word2vec [18], Glove [24], Fastext [3]) have been considered. This setting, which is novel for Spanish sentiment analysis, can be useful in several domains.

The Deep Learning architecture proposed is composed by a Convolutional Neural Network [14], a Recurrent Neural Network [12] and a final dense layer. In order to avoid overfitting, besides traditional dropout schemes, we rely on data augmentation. Data augmentation is useful for low resources languages such as Spanish.

Those design choices allow us to obtain results comparable to state-of-the-art approaches over the InterTASS 2017 dataset, in terms of accuracy. The dataset was proposed in the TASS workshop at SEPLN. In the last six years, this workshop has been the main source for Spanish sentiment analysis datasets and proposals [17].

The remainder of the paper is organized as follows. Section 2 reviews preliminaries on sentiment analysis and neural networks. Our proposal is presented in Section 3. Results are described in Section 4. Related work is presented in Section 5. Finally, Section 6 concludes the paper.

2 Preliminary

2.1 Sentiment Analysis

Sentiment analysis (also known as opinion mining) is an active research area in natural language processing [28]. Sentiment classification is a fundamental and extensively studied area in sentiment analysis. It targets at determining the sentiment polarity (positive or negative) of a sentence (or a document) based on its textual content [27]. Polarity classification tasks have usually based on two main approaches [4]: a supervised approach, which applies machine learning algorithms in order to train a polarity classifier using a labelled corpus; an unsupervised approach, semantic lexicon-based, which integrates linguistic resources in a model in order to identify the polarity of the opinions.

Since the performance of a machine learner heavily depends on the choices of data representation, many studies devote to building powerful feature extractor with domain expert and careful engineering [20].

As stated by Liu [16], sentiment analysis has been researched at three levels: i) Document level: The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment [22]; ii) Sentence level: The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion; iii) Entity and Aspect level [22]: Both the document level and the sentence level analyses do not discover what exactly people liked and did not like. Aspect level performs finer-grained analysis.

2.2 Deep Neural Networks

Several deep neural network approaches have been successfully applied to sentiment analysis in the last years [31]. However, these results have been mostly obtained for English Language [17]. The related work section further describes, several attempts to apply deep learning algorithms for Spanish sentiment analysis. In this section we only focus on word representations, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which are the main building blocks of our proposal.

Word Representations (Word2vec, Glove, Fasttext) Nowadays, word representations are paramount for sentiment analysis [31]. In order to model text words as features within a machine learning framework, a common approach is to encode words as discrete atomic symbols. These encodings are arbitrary and provide no useful information to the system regarding the relationships that may exist between the individual symbols [28]. The discrete representation has some problems such as missing new words. This representation also requires human labor to create and adapt. It is also hard to compute accurate word similarity and is quite subjective. To cope with these problems, the distributional similarity based representations propose to represent a word by means of its neighbors, its context [27].

Word2vec [18] is a particularly computationally-efficient predictive model for learning word embeddings from raw text. Take a vector with several hundred dimensions where each word is represented by a distribution of weights across those elements [7, 2]. Thus, instead of a one-to-one mapping between an element in the vector and a word, the representation of a word is spread across all the elements in the vector.

In contrast to Word2vec, Glove [24] seeks to make explicit what Word2vec does implicitly: encoding meaning as vector offsets in an embedding space. In Glove, it is stated that the ratio of the co-occurrence probabilities of two words (rather than their co-occurrence probabilities themselves) is what contains information and so look to encode this information as vector differences.

In Fasttext [3] instead of directly learning a vector representation for a word, a representation for each character n-gram is learned. In this sense, each word is represented as a bag of character n-grams, thus the overall word embedding is a sum of these characters n-grams. The advantage of Fasttext is that generates

better embeddings for rare and out-of-corpus words. By using different n-grams Fastext explores key structural components of words.

Convolutional Neural Networks While Convolutional Neural Networks (CNN) have been primarily applied to image processing, they have also been used for NLP tasks [14].

In the image context [15], given a raw input (2D arrays of pixel intensities) several *convolutional* layers allow us to capture features images at several abstraction levels. In this context, a discrete convolution takes a filter matrix and multiply its values element-wise with the original matrix, then sum them up. To get the full convolution we do this for each element by sliding the filter over the whole matrix.

The convolved map feature denotes a level of abstraction obtained after the convolution operations (there are also ReLU activation, Pooling and Softmax layers). CNN exploits the property that many natural signals are compositional hierarchies: higher-level features are obtained by composing lower-level ones. In images, local combinations of edges form motifs, motifs assemble into parts, and parts from objects [15]. All this learning representation is performed in an unsupervised manner. The amount of filters and convolutional layers denote how rich features and abstraction levels we wish to obtain from images.

Conversely, if we wish to apply CNNs in natural language tasks several changes are needed [14]. Texts are tokenized and must be encoded as numbers — input numerical variables are usual in neural networks algorithms. In the last five years, word embeddings representations (but also character and paragraph) have been preferred. This is due to semantical/syntactical similarity is better expressed in a distributed manner [18].

A sentence can be represented as a matrix. Thus, the sentence length denotes the number of rows and the word embedding dimension denotes the number of columns. This allows us to perform discrete convolutions as in the image case (2D input matrix). However, one must be careful when defining filter sizes, which usually have the same width as word embeddings [14].

Instead of working with 2D representation, we may also work with 1D representation, i.e., to concatenate several word embeddings in a long vector and then apply several convolution layers.

Recurrent Neural Networks Recurrent Neural Networks (RNN) [8] are a kind of neural network that makes it possible to model long-distance dependencies among variables. Therefore, RNN are best suited for tasks that involve sequential inputs such as speech and language [15]. RNNs process an input sequence one element at a time, maintaining in their hidden units a state vector that implicitly contains information about the history of all the past elements of the sequence. To do so, a connection is added that references the previously hidden states h_{t-1} when computing hidden state h , formally [21]:

$$h_t = \tanh(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$

$h_t = 0$ when the initial step is $t = 0$. The only difference from the hidden layer in a standard neural network is the addition of the connection $W_{hh}h_{t-1}$ from the hidden state at time step $t - 1$ connecting to that at time step t . As this is a recursive equation that use h_{t-1} from the previous time step.

In the context of Sentiment Analysis, an opinionated sentence is a sequence of words. Thus, RNNs are suitable for modeling this input [12]. Similar to CNNs, input is given as words (character) embeddings which can be learned during training or may also be pre-trained (Glove, Word2vec, Fasttext).

Each word is mapped to a word embedding which is the input at every time step of the RNN. The maximum sequence length denotes the length of the recurrent neural network. Each hidden state models the dependence among a current word and all the precedent words. Usually, the final hidden state, which ideally denotes all the encoded sentence, is connected to a dense layer so as to perform sentiment classification [12].

RNNs are very powerful dynamic systems, but training them has proved to be problematic because the backpropagated gradients either grow or shrink at each time step. Thus, over many time steps they typically explode or vanish. A sequence of words comprise a sequence of RNNs cells. These cells can have some gate mechanism in order to avoid gradient vanishing longer sequences. In this setting Long Short Term Memory Cells (LSTM) or Gated Recurrent Units (GRU) are common choices [21].

3 Proposal

The aim of this paper is to explore several Deep Learning algorithms possibilities in order to perform sentiment analysis. The focus is to tackle polarity detection in Spanish Tweets. In this sense, some models were tested. Details of these experiments are given in Section 4.

In this section, we present our best pipeline for Spanish sentiment analysis of short texts. Basically, it is composed by Word embeddings, CNN and RNN models. The pipeline is showed in Figure 1. A concise description is given as follows.

i) Basic pre-processing is performed as the focus is given to data augmentation; ii) The input is a sequence of words — a short opinionated sentence. These words are mapped to three pre-trained Spanish word embeddings (Word2vec, Glove, Fasttext); iii) The three channels are the input to a 3D Convolutional Neural Network. After several convolutional and max pooling layers we obtain a feature vector of a given length; iv) The feature vector obtained from the CNN is mapped to a sequence and passed to a RNN. It is a simple RNN model, with LSTM cells; v) The final hidden state of the RNN is completely connected to a dense layer.

Further details about these design choices are given as follows.

Fig. 1. Pipeline of our proposal: Word Embeddings+CNN+RNN.

3.1 Data Augmentation

In general a few pre-processing steps are performed over raw data. Since we have few training examples in Spanish and Deep Learning techniques are susceptible to overfitting, we would rather focus on data augmentation. We propose a novel approach for data augmentation. Basically, we identify nouns, adjectives and verbs on sentences by performing Part-Of-Speech tagging⁴. By doing so, we emphasize tokens that are prone to be opinionated words. Then, more examples are created by combining bigrams and trigrams from the former tokens. In addition, we augment data based on word synonyms [31]. Opinionated words are replaced by synonyms. Overall, this process allowed us to obtain better generalization results.

3.2 Word Embeddings Choice

One of the main contributions of this paper was to find the best word embedding setting. We have trained Word2vec and Glove embedding on Spanish corpus and we have used a pre-trained Fasttext embedding. At the end, empirical tests allowed us to decide for using these three mappings as channels in our CNN building block. None of the previous works for Spanish Sentiment Analysis had used three embedding channels in CNNs before.

⁴ The following tool was used to perform POS tagging: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

3.3 CNN Architecture

Our CNN architecture is based on Kim’s work [14]. Since three word embeddings are used, then the first convolutional layer receives a 3D input. Filters have the same width as embeddings dimension, and we perform convolutions from 1 to 5 words. The pooling layer allows us control the desired feature vector obtained.

3.4 RNN Architecture

The RNN receives a CNN vector as input, and LSTMs cells are defined accordingly. The last hidden state is fully connected to a dense layer which allows us to define a classifier [12].

4 Experiments

Experiments were performed using Deep Learning algorithms. CNNs and RNNs were tested separately. Our best result was obtained by composing word embeddings, CNNs and RNNs. We first describe the benchmark dataset used. Then, accuracy results are showed.

4.1 Dataset

The dataset used to perform comparisons was InterTASS, which is a collection of Spanish Tweets, used in TASS at SEPLN workshop in 2017 [17]. We have used this dataset since it is the most recent benchmark that allows us to compare among Deep Learning approaches for Spanish sentiment analysis. The dataset is further detailed in Table 1.

Table 1. InterTASS dataset (TASS 2017)

Corpus	Tweets
Training	1,008
Development	506
Test	1,899
Total	3,413

4.2 Results

We have implemented several deep neural networks models and the dataset InterTASS 2017 was used for training. For this implementation we use Tensorflow⁵.

⁵ <https://www.tensorflow.org/>

In order to find the best hyper parameters, we have used a ten-fold cross validation process. The test set has only been used to report results. In Table 2 we report results in terms of accuracy.

A first attempt was to test several RNNs models (*many-to-one architecture*, single layer, multilayer, bidirectional). The reported model, RNN in Table 2, has a many-to-one architecture. The input is a sequence of words and the output is the resulting polarity. There is only a hidden layer, and the input is a pre-trained sequence of Word2vec embeddings. A second attempt was to test several CNN models, i.e., 1D CNN, 2D CNN and 3D CNNs, until 4 convolutional / pooling layers. The reported model, CNN in Table 2, is a 3D CNN. Thus, the input received three channels of pre-trained word embeddings. It had only three layers: a convolutional, a pooling and a dense layer. It is worth noting that our best result was obtained by the model described in Section 3 (CNN+RNN in Table 2). This is a combination of a 3D CNN and a many-to-one RNN. A 3D CNN architecture whose outputs were mapped to a sequence of LSTM cells. Our data augmentation scheme was also used in order to avoid overfitting.

Table 2. Deep Learning approaches results on InterTASS dataset (TASS 2017)

Our DL Attempts Accuracy	
CNN+RNN	0.609
CNN	0.5552
RNN	0.4972

In Table 3, we compare our best model (CNN+RNN) with the state-of-the-art InterTASS 2017 results, in terms of accuracy. It is worth noting that our approach is comparable to the other approaches. In addition, our proposal is the only top result using a Deep Learning approach.

Table 3. State-of-the-art results on InterTASS dataset (TASS 2017)

System	Accuracy
CNN+RNN (our approach)	0.609
jacerong-run1 [6]	0.608
ELiRF-UPV-run1 [13]	0.607
RETUYT-svm cnn [25]	0.596
tecnolenguasent [19]	0.595

5 Related Work

There is a plethora of related works for sentiment analysis but, we are only interested in contributions for the Spanish language. Arguably one of the most complete Spanish sentiment analysis systems was proposed by Brooke et al [5], which had a linguistically approach. That approach integrated linguistic resources in a model to decide about polarity opinions [29]. However, recent successful approaches for Spanish polarity classification have been mostly based on machine learning [9].

In the last six years, the TASS at SEPLN Workshop has been the main source for Spanish sentiment analysis datasets and proposals [10, 17]. Benchmarks for both the polarity detection task and aspect-based sentiment analysis task have been proposed in several editions of this Workshop (Spanish Tweets have been emphasized).

Recently, deep learning approaches emerge as powerful computational models that discover intricate semantic representations of texts automatically from data without feature engineering. These approaches have improved the state-of-the-art in many sentiment analysis tasks including sentiment classification of sentences/documents, sentiment extraction and sentiment lexicon learning [27]. However, these results have been mostly obtained for English Language. Due to our proposal is based on Deep Learning, the related work that follows emphasizes these kinds of algorithms.

Arguably, the first approach using Deep Learning techniques for Spanish Sentiment Analysis was proposed in the TASS at SEPLN workshop in 2015 [30]. The authors presented one architecture that was composed by a RNN layer (LSTMs cells), a dense layer and a Sigmoid function as output. The performance over the general dataset was poor, 0.60 in terms of accuracy (the best result was 0.69 in TASS 2015).

The first Convolutional Neural Network approach for Spanish Sentiment Analysis was described in [26]. However, the CNN model proposed for sentiment analysis was mostly based on Kim’s work [14]. It was comprised by only a single convolutional layer, followed by a max-pooling layer and a Softmax classifier as final layer. Word embeddings were used in three ways: a learned word embedding from scratch, and two pre-trained Word2vec models. In terms of accuracy they obtained 0.64, which was far from the best result (0.72 was the best result in TASS 2016 [10]).

Another CNN approach for Spanish Sentiment Analysis was presented by Paredes et al [23]. First, a preprocessing step (tokenization and normalization) was performed which was followed by a Word2vec embedding. Then, the model was comprised of a 2D convolutional layer, a max pooling and a final Softmax layer, i.e., it is also similar to Kim’s work [14]. It was reported an F-measure of 0.887 over a non public Twitter corpus of 10000 tweets.

Most of the Deep Learning approaches for Spanish sentiment analysis have been presented in TASS 2017 [17]. For instance, Rosa et al [25] used word embeddings within two approaches, SVM (with manually crafted features) and Convolutional Neural Networks. Pre-trained Word2vec, Glove and fasttext embeddings

were used. Unlike our approach, these embeddings were used separately. In fact, the best results of this paper were obtained using Word2vec. When CNN was employed, unidimensional convolutions were performed. Several convolutional layers were tested. The best model had three convolutional layers, using 2, 3 and 4 word filters. However, their best results were obtained when combined with SVM and CNN, using simply a decision rule based on both probability results. Interesting results were obtained, 0.596 in terms of accuracy, for the InterTASS dataset (the best accuracy result was 0.608 for TASS 2017 [17]).

Garcia-Vega et al [11] used word embeddings with shallow classifiers. Recurrent neural networks with LSTM nodes and a dense layer were also tested. Two kinds of experiments were performed using word embeddings and TFIDF values as inputs. Both experiments obtained poor results (0.333 and 0.404 in terms of accuracy for the InterTASS dataset in 2017).

Araque et al [1] explored recurrent neural networks in two ways i) a set of LSTM cells whose input were word embeddings, ii) a combination of input word vector and polarity values obtained from a sentiment lexicon. As usual, a last dense layer with a Softmax function was used as final output. While interesting, experimental results showed that the best performance was obtained by the second model, LSTM +Lexicon + dense. In terms of accuracy they obtained 0.562. This value is far from the TASS 2017 top results.

In the last years, the best results were obtained for the group ELiRF [13]. In TASS 2017, they obtained the second best result for the InterTSS task, 0.607, in terms of accuracy (The first place presented an ensemble approach [6]). It is worth noting that ELiRF best results were obtained using a Multilayer perceptron with word embeddings as inputs. This MLP had two layers with ReLU activation functions. A Second approach used a stack of CNN and LSTM models, using pre-trained word embeddings. The architecture was composed by one convolutional layer, 64 LSTM cel and a fully connected MLP, with ReLU activation functions. This last architecture had a poor performance (0.436 in terms of Accuracy).

6 Conclusion

Despite being one of the three most used languages at Internet, Spanish has had few resources developed for natural language processing tasks. For instance, unlike English sentiment analysis, Deep Learning approaches were unable to obtain state-of-the-art results on Spanish benchmark datasets in the past. The aim of this work was to demonstrate that Deep Learning is the best choice for Spanish Twitter sentiment analysis. Our experimental results have showed that a combination of data augmentation, at least three kinds of word embeddings, a 3D Convolutional Neural Network, followed by a Recurrent Neural Network allows us to obtain results comparable to state-of-the-art approaches over the InterTASS 2017 benchmark. In addition, this setup could be easily adapted to other domains.

References

1. Araque, O., Barbado, R., Sanchez-Rada, J.F., Iglesias, C.A.: Applying recurrent neural networks to sentiment analysis of spanish tweets. In: Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN. pp. 71–76 (2017)
2. Bengio, Y., Ducharme, R., Vincent, P., Janvin, C.: A neural probabilistic language model. *J. Mach. Learn. Res.* **3**, 1137–1155 (2003)
3. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* **5**, 135–146 (2017)
4. Brody, S., Elhadad, N.: An unsupervised aspect-sentiment model for online reviews. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. pp. 804–812. HLT '10, Association for Computational Linguistics, Stroudsburg, PA, USA (2010)
5. Brooke, J., Tofiloski, M., Taboada, M.: Cross-linguistic sentiment analysis: From english to spanish. In: Proceedings of RANLP 2009. pp. 50–54 (2009)
6. Ceron-Guzman, J.A.: Classier ensembles that push the state-of-the-art in sentiment analysis of spanish tweets. In: Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN. pp. 59–64 (2017)
7. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.P.: Natural language processing (almost) from scratch. *CoRR* **abs/1103.0398** (2011)
8. Elman, J.L.: Finding structure in time. *Cognitive science* **14(2)**, 179–211 (1990)
9. Garcia, M., Martinez, E., Villena, J., Garcia, J.: Tass 2015 the evolution of the spanish opinion mining systems. *Procesamiento de Lenguaje Natural* **56**, 33–40 (2016)
10. Garcia-Cumbreras, M.A., Villena-Roman, J., Martinez-Camara, E., Diaz-Galiano, M., Martin-Valdivia, T., Ureña Lopez, A.: Overview of tass 2016. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN. pp. 13–21 (2016)
11. Garcia-Vega, M., Montejo-Raez, A., Diaz-Galiano, M.C., Jimenez-Zafra, S.M.: Sinai in tass 2017: Tweet polarity classification integrating user information. In: Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN. pp. 91–96 (2017)
12. Graves, A.: Supervised Sequence Labelling with Recurrent Neural Networks. *Studies in Computational Intelligence*, Springer, Berlin (2012), <https://cds.cern.ch/record/1503877>
13. Hurtado, L.F., Pla, F., Gonzalez, J.A.: Elirf-upv at tass 2017: Sentiment analysis in twitter based on deep learning. In: Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN. pp. 29–34 (2017)
14. Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in NaturalLanguage Processing, EMNLP 2014, October 25–29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL. pp. 1746–1751 (2014), <http://aclweb.org/anthology/D/D14/D14-1181.pdf>
15. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521(7553)**, 436–444 (2015)
16. Liu, B.: Sentiment Analysis and Opinion Mining. Morgan and Claypool Publishers (2012)
17. Martinez-Camara, E., Diaz-Galiano, M., Garcia-Cumbreras, M.A., Garcia-Vega, M., Villena-Roman, J.: Overview of tass 2017. In: Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN. pp. 13–21 (2017)

18. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems 26*, pp. 3111–3119. Curran Associates, Inc. (2013), <http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>
19. Moreno-Ortiz, A., Perez-Hernandez, C.: Tecnolengua lingmotif at tass 2017: Spanish twitter dataset classification combining wide-coverage lexical resources and text features. In: *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SE-PLN*. pp. 35–42 (2017)
20. Narayanan, V., Arora, I., Bhatia, A.: Fast and accurate sentiment classification using an enhanced naive bayes model. In: *Proceedings of the 14th International Conference on Intelligent Data Engineering and Automated Learning IDEAL 2013*. pp. 194–201. IDEAL 2013, Springer-Verlag New York, Inc., New York, NY, USA (2013)
21. Neubig, G.: Neural machine translation and sequence-to-sequence models: A tutorial. *CoRR* **abs/1703.01619** (2017), <http://arxiv.org/abs/1703.01619>
22. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2**(1-2), 1–135 (Jan 2008). <https://doi.org/10.1561/15000000011>, <http://dx.doi.org/10.1561/15000000011>
23. Paredes-Valverde, M.A., Colomo-Palacios, R., Salas-Zarate, M.d.P., Valencia-Garcia, R.: Sentiment analysis in spanish for improvement of products and services: A deep learning approach. *Scientific Programming* **6**, 1–6 (2017)
24. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: *Empirical Methods in Natural Language Processing (EMNLP)*. pp. 1532–1543 (2014)
25. Rosa, A., Chiruzzo, L., Etcheverry, M., Castro, S.: Retuyt in tass 2017: Sentiment analysis for spanish tweets using svm and cnn. In: *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN*. pp. 77–83 (2017)
26. Segura-Bedmar, I., Quiros, A., Martínez, P.: Exploring convolutional neural networks for sentiment analysis of spanish tweets. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. pp. 1014–1022. Association for Computational Linguistics (2017), <http://aclweb.org/anthology/E17-1095>
27. Tang, D., Wei, F., Qin, B., Yang, N., Liu, T., Zhou, M.: Sentiment embeddings with applications to sentiment analysis. *Knowledge and Data Engineering, IEEE Transactions on* **28**(2), 496–509 (Feb 2016)
28. Tang, D., Qin, B., Liu, T.: Deep learning for sentiment analysis: successful approaches and future challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **5**(6), 292–303 (2015)
29. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. pp. 417–424. ACL '02, Association for Computational Linguistics, Stroudsburg, PA, USA (2002)
30. Vilares, D., Doval, Y., Alonso, M.A., Gomez-Rodriguez, C.: Lys at tass 2015: Deep learning experiments for sentiment analysis on spanish tweets. In: *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN*. pp. 47–52 (2015)
31. Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis : A survey. *CoRR* **abs/1801.07883** (2018), <http://arxiv.org/abs/1801.07883>