

# LaSTUS-TALN at IberLEF 2019 eHealth-KD Challenge

## Deep Learning Approaches to Information Extraction in Biomedical Texts

Àlex Bravo, Pablo Accuosto, and Horacio Saggion

LaSTUS/TALN Research Group, DTIC  
Universitat Pompeu Fabra, Spain  
C/Tànger 122-140, 08018 Barcelona, Spain  
{name.surname}@upf.edu

**Abstract.** This paper presents the participation of the LASTUS-TALN team in the *IberLEF eHealth-KD 2019* challenge, which proposes 2 sub-tasks in the context of biomedical text processing in Spanish: i) the detection and classification of key phrases and ii) the identification of the semantic relationships between them. We propose an architecture based on a bidirectional long short-term memory (BiLSTM) with a conditional random field (CRF) classifier as the last layer of the network to find and classify the relevant key phrases. Concerning relation extraction problem, for each candidate relationship, we describe a global and local context representing the supposed relationship and the context of the candidate key phrases, respectively and divided the problem into three simpler classification tasks: i) decide if the entities are related, ii) identify the type of relationship and iii) obtain the correct direction. In our model, these three classification tasks were trained at the same time. When key phrase extraction and relation extraction were run in sequence, our system achieved the third highest F1 score in the main evaluation.

**Keywords:** Information Extraction · Deep Learning · Biomedical Text · Natural Language Processing

## 1 Introduction

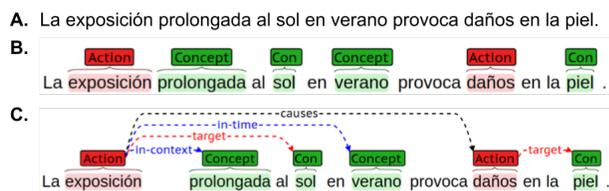
Information Extraction (IE) is the process of finding relevant entities and their relationships in text [14, 13]. This process is a crucial step to structure the valuable knowledge locked in biomedical literature for a variety of purposes (e.g. information retrieval, knowledge discovery and documents recommendation).

---

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). IberLEF 2019, 24 September 2019, Bilbao, Spain.

Many biomedical challenges have been proposed to promote the development of systems to extract, classify and index biomedical knowledge, such as SemEval<sup>1</sup>, CLEF<sup>2</sup> campaigns and others [7].

Previously, the *eHealth-KD 2018* challenge [9], held at the Workshop on Semantic Analysis at the SEPLN (TASS)<sup>3</sup>, promoted the development and evaluation of systems that are able to automatically extract a large variety of knowledge from biomedical documents written in Spanish, including the extraction and classification of key phrases and semantic relations between them<sup>4</sup>. The challenge was organized in two subtasks: i) the detection of entities and ii) classification and the recognition of semantic relationships. Six teams successfully concluded their participation with a great variety of proposed systems [9]. In general, the most competitive approaches in individual subtasks were led by state-of-the-art machine learning approaches. In particular, for the detection of semantic relations, deep learning architectures seemed to outperform more classic techniques. In addition, including domain-specific knowledge (e.g. UMLS) provided significant boost to the results. The best results in the detection and classification of entities were of 0.87 and 0.96 of F-score, respectively. Regarding the detection of semantic relations, the best scores were around 0.45 in F-score. This reinforces the belief that relation extraction is still a challenging task.



**Fig. 1.** Example of the *IberLEF eHealth-KD 2019* challenge subtasks. **A.** Example of input sentence. **B.** Sentence with key phrases annotations (results of the subtask A). **C.** Sentence with annotations of semantic relationships (results of the subtask B) .

Recently, in the *IberLEF eHealth-KD 2019* challenge, the organizers also proposed two subtasks (see Fig. 1<sup>5</sup>): i) the recognition of key phrases (subtask A, covering both, identification and classification) and ii) the detection of semantic relationships between them (subtask B). In this paper, we present our approaches and results for the participation in this challenge. From the previous challenge, we could observe that the best systems in the recognition of key phrases do not correlate with the best systems in relation extraction. Under this assumption, we propose a different deep learning approach for each subtask.

<sup>1</sup> International Workshop on Semantic Evaluation

<sup>2</sup> Conference and Labs of the Evaluation Forum

<sup>3</sup> <http://www.sepln.org/workshops/tass>

<sup>4</sup> <http://www.sepln.org/workshops/tass/2018/task-3/index.html>

<sup>5</sup> <https://knowledge-learning.github.io/ehealthkd-2019/tasks>

## 2 Methods

### 2.1 Subtask A: Identification and classification of key phrases

In this section, we describe our proposal for identifying and classifying key phrases in biomedical texts. Key phrases are considered to be all the entities (single word or multiple words) that represent semantically relevant elements in a sentence.

There are four potential classes for key phrases, as described in [12]: *Concept*, *Action*, *Predicate* and *Reference*. The input is a tokenized text document with a sentence per line. The output consists of a plain text file where each line represents a key phrase with its unique ID, positions of the starting and ending character of the text span, the assigned category and the full span of text containing the key phrase.

For this task we propose an architecture based on a BiLSTM with a CRF classifier as the last layer of the network. We based our implementation on the one made available by the Ubiquitous Knowledge Processing Lab of the Technische Universität Darmstadt [15]<sup>6</sup>. We use two BiLSTM layers with 100 recurrent units with Adam optimizer and naive dropout probability of 0.25.

In order to make them compatible with the proposed architecture, we transform the format of the provided training, development and test sets into text files containing one token per line. The corresponding classes are encoded in the standard beginning-inside-outside (BIO) sequence tagging scheme.

The tokens are fed into the network as 1024-dimensional embeddings obtained by averaging the three output layers of a deep contextualized ELMo model [11]. As the official Allen NLP ELMo models<sup>7</sup> are only available for English, we used Spanish pre-trained ELMo models made available by [4]<sup>8</sup>. We made the necessary modifications to the UKP sequence tagger in order to make it compatible with these representations, as they are not directly pluggable into the Allen NLP API used in the original implementation.

### 2.2 Subtask B: Detection of semantic relations

The goal of this subtask is to recognize the thirteen semantic relationships between the key phrases detected and labelled in each sentence. In addition, every semantic relation is directed, that is, the involved entities must match the correct direction.

In this subtask, we implemented a multi-task learning approach. First, we have broken down the subtask B into three simpler classifications: i) decide if the entities are related, ii) identify the type of relationship and iii) decide the correct direction. Then, in our model, these three classification tasks were trained at the same time.

<sup>6</sup> <https://github.com/UKPLab/elmo-bilstm-cnn-crf>

<sup>7</sup> <https://allennlp.org/elmo>

<sup>8</sup> <https://github.com/HIT-SCIR/ELMoForManyLangs>

Before the relation extraction process, we considered as candidates all entity pairs detected in the same sentence. For each entity pair, we design a *global context* representing the supposed relationship and a *local context* representing the environment of each candidate entity.

**Global and Local contexts from a relationship** Following the philosophy of [6] and [3], we have organized the information of each candidate relationship into two scenarios: *global* and *local contexts*. In detail, the *global context* is based on the assumption that an association between two entities is more likely to be expressed within one of three sequences [10]:

- *Fore-Between*: from the words before and between the two candidates.
- *Between*: from the words between the two candidate entities.
- *Between-After*: from the words between and after the two candidates.

On the other hand, we also defined the *local context* of each candidate entity. The *local context* can provide useful information for detection of the type and direction of the relationship, as well as the presence of the relation itself. This context is also based on a sequence of words (*EntityA-Context* and *EntityB-Context*), which contains the information from the words located at the left and right of the candidate entities (with a window size of 2).

In both contexts, each sequence is represented using the concatenation of the following embeddings: tokens, PoS tags, entity types and dependencies.

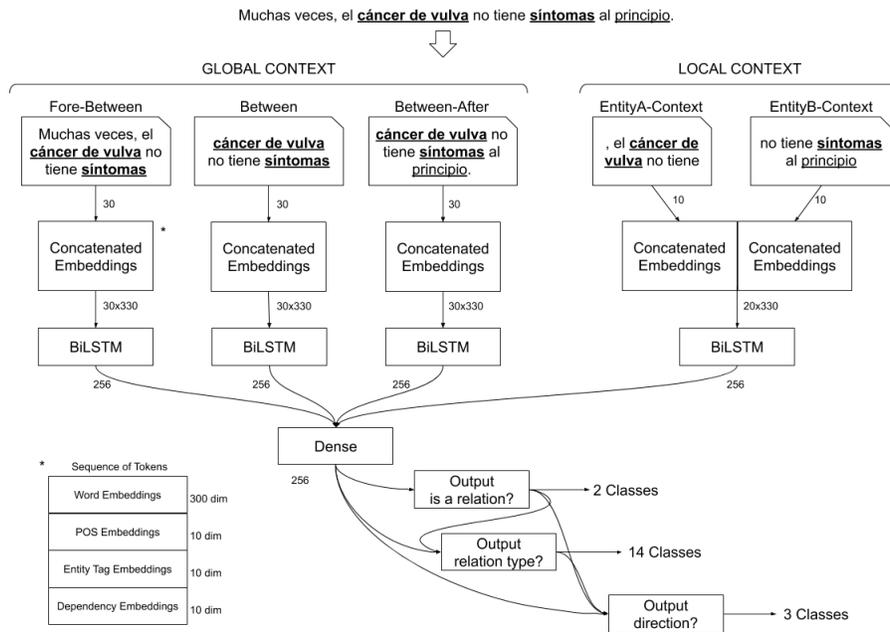
**Model** The model consists of a BiLSTM model with an attention layer on top for each concatenated embedding. The model captures the most important syntactic and semantic information from each sequence (*Fore-Between*, *Between*, *Between-After*, *EntityA-Context* and *EntityB-Context*) to face the three tasks: to detect if the key phrases are related, the type of relationship and its direction. In Fig. 2 a simplified schema of our model can be seen. In the following we explain how the model works.

First, the sentences were processed with Spacy<sup>9</sup> to obtain the tokens, PoS tags and dependencies. From the previous subtask A, we also included the entity type information. Then, the five sequences are generated from the tokenized sentence.

Second, the embedding layers transform each token in the sequence into a set of low-dimensions related to the token itself, the POS tag, the dependency and the entity type.

Specifically, in the case of the tokens, an embedding layer was randomly initialized from a uniform distribution (between -0.8 and 0.8 values and with 300 dimensions) and then, the embedding layer was updated with the word vectors computed from the sentences in Spanish on MedlinePlus by means of fastText [2]. Similarly, the rest of the embedding layers were also randomly initialized from a uniform distribution, but with only 10 dimensions and without pre-trained

<sup>9</sup> <https://spacy.io/>



**Fig. 2.** Simplified schema of the model to detect semantic relationships. From top to bottom, a example of sentence with two candidate (in bold) of three (underlined) key phrases. After sentence processing, five different sequences are created based on their *global* and *local context*. Each sequence is transformed in their concatenated embeddings, and processed by a BiLSTM layer. The output from each BiLSTM layer feed the following dense layer, which directs its output to three parallel fully-connected layers to classify each task.

embeddings. As shown in Fig. 2, for each token of the sequence, the related embeddings are concatenated.

Next, for each sequence, a BiLSTM layer gets high-level features from its corresponding concatenated embeddings. The BiLSTM gets a word embedding sequentially, left-to-right as well as right-to-left order in parallel, producing a hidden step and keeps its hidden state through time. Therefore, it gives two hidden states as output at each step and is able to capture backwards and long-range dependencies.

A critical and apparent disadvantage of LSTM models is that they compress all information into a fixed-length vector, causing the incapability of remembering long sequences. Attention mechanism aims to overcome the limitation of fixed-length vector keeping relevant information from long sequences. Attention techniques have been recently demonstrated success in multiple areas of the NLP such as question answering, machine translations, speech recognition and relation extraction [1, 8, 5, 16]. For that reason, we added an attention layer (after each BiLSTM layer), which produces a weight vector and merge word-level features from each time step into a sequence-level feature vector, by multiplying the weight vector [16]. Furthermore, to alleviate overfitting during the training, we applied dropout regularization, which sets randomly to zero a proportion of the hidden units during forward propagation, creating more generalizable representations of data. In the model, we employ dropout on the embeddings and BiLSTM layers. The dropout rate was set to 0.5 in all cases.

Then, the final relation-level feature vector produced by the previous BiLSTM layers feed the following dense layer, which directs its output to three parallel fully-connected layers to classify each task. Note that the three final output layers are connected in cascade, that is, the output of the first classification (are these entities related?) also feeds the second classification task (type of semantic relationship?), and the last one (direction of the relationship?) is feed by the outputs from the previous dense layer, the first and the second classification task.

Finally, we consider that a relationship has been detected when our model predicts a positive value in the three classification tasks.

**Table 1.** Scores for Scenario 1 (subtasks A and B).

Team	F-Score	Precision	Recall
LASTUS-TALN	0.5816	0.7740	0.4658
LASTUS-TALN (fixed)	0.5903	0.5945	0.5862
Highest Score	0.6394	0.6506	0.6286
Average Score	0.5538	0.6991	0.4675
Baseline	0.4309	0.5204	0.3677

**Table 2.** Scores for Scenario 2 (subtask A).

Team	F-Score	Precision	Recall
LASTUS-TALN	0.8167	0.7997	0.8344
Highest Score	0.8203	0.8073	0.8336
Average Score	0.7749	0.7746	0.7774
Baseline	0.5466	0.5129	0.5851

**Table 3.** Scores for Scenario 3 (subtask B).

Team	F-Score	Precision	Recall
LASTUS-TALN	0.2298	0.1705	0.3521
LASTUS-TALN (fixed)	0.3918	0.3964	0.3873
Highest Score	0.6269	0.6667	0.5915
Average Score	0.4637	0.5305	0.4318
Baseline	0.1231	0.4878	0.0704

### 3 Results

The organizers proposed a main evaluation scenario (Scenario 1) where subtasks A and B are performed in sequence. Additionally, two optional scenarios were considered in order to evaluate each subtask independently of the other (Scenario 2 for subtask A and Scenario 3 for subtask B).

Results of our system for each scenario are given in Tables 1, 2 and 3. In addition to our results, we include the highest and average scores obtained for all the participants in each scenario. Please note that, as there was a bug in our submitted implementation for subtask B, which was subsequently fixed, Tables 1, 2 and 3 show the official results achieved in the challenge as well as the fixed results, which we comment below.

For subtask A, our system achieved one of the best results in the challenge (see Table 2). In contrast, for subtask B, our system obtained one of the lowest scores (see Table 3). When the tasks are performed sequentially (see Table 1) the errors in subtask B are mitigated by the good performance obtained for subtask A, leaving our system in the third position of the challenge.

### 4 Conclusions

In this paper, we presented the participation of the LASTUS-TALN team in the *IberLEF eHealth-KD 2019* challenge, which proposed 2 sub-tasks, the detection and classification of key phrases and the identification of the semantic relationships between them. In the first scenario, we proposed a BiLSTM network for sequence tagging, based on the architecture made available by the Ubiquitous Knowledge Processing Lab of the Technische Universität Darmstadt [15]. We adapted this architecture in order to use an alternative version of ELMo deep contextualized embeddings in Spanish.

On the other hand, we followed a daring philosophy to represent relationships in multiple context. Although our results for task 2 were not as expected, we think that this representation has a lot of potential. For that reason, our future work will focus on the study of this representation and how it behaves in neural network. In this sense, we want to achieve a performance close to the state-of-the-art in this challenge.

## Acknowledgments

**Funding:** This work is partly supported by the Spanish Government under the María de Maeztu Units of Excellence Programme (MDM-2015-0502).

## References

1. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
2. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606 (2016)
3. Bravo, À., Piñero, J., Queralt-Rosinach, N., Rautschka, M., Furlong, L.I.: Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. *BMC bioinformatics* **16**(1), 55 (2015)
4. Che, W., Liu, Y., Wang, Y., Zheng, B., Liu, T.: Towards Better UD Parsing: Deep Contextualized Word Embeddings, Ensemble, and Treebank Concatenation. In: Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. pp. 55–64. Association for Computational Linguistics, Brussels, Belgium (October 2018), <http://www.aclweb.org/anthology/K18-2005>
5. Chorowski, J.K., Bahdanau, D., Serdyuk, D., Cho, K., Bengio, Y.: Attention-based models for speech recognition. In: Advances in neural information processing systems. pp. 577–585 (2015)
6. Giuliano, C., Lavelli, A., Romano, L.: Exploiting shallow linguistic information for relation extraction from biomedical literature. In: 11th Conference of the European Chapter of the Association for Computational Linguistics (2006)
7. Gonzalez-Hernandez, G., Sarker, A., OConnor, K., Savova, G.: Capturing the patients perspective: a review of advances in natural language processing of health-related text. *Yearbook of medical informatics* **26**(01), 214–227 (2017)
8. Hermann, K.M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., Blunsom, P.: Teaching machines to read and comprehend. In: Advances in Neural Information Processing Systems. pp. 1693–1701 (2015)
9. Martínez Cámara, E., Almeida Cruz, Y., Díaz Galiano, M.C., Estévez-Velarde, S., García Cumbreras, M.Á., García Vega, M., Gutiérrez, Y., Montejo Ráez, A., Montoyo, A., Muñoz, R., et al.: Overview of tass 2018: Opinions, health and emotions (2018)
10. Mooney, R.J., Bunescu, R.C.: Subsequence kernels for relation extraction. In: Advances in neural information processing systems. pp. 171–178 (2006)
11. Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep Contextualized Word Representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). vol. 1, pp. 2227–2237 (2018)

12. Piad-Morffis, A., Gutiérrez, Y., Consuegra-Ayala, J.P., Estevez-Velarde, S., Almeida-Cruz, Y., Muñoz, R., Montoyo, A.: Overview of the ehealth knowledge discovery challenge at iberlef 2019 (2019)
13. Piskorski, J., Yangarber, R.: Information extraction: Past, present and future. In: Poibeau, T., Saggion, H., Piskorski, J., Yangarber, R. (eds.) Multi-source, Multilingual Information Extraction and Summarization, pp. 23–49. Theory and Applications of Natural Language Processing, Springer (2013)
14. Poibeau, T., Saggion, H., Piskorski, J., Yangarber, R.: Multi-source, Multilingual Information Extraction and Summarization. Springer Publishing Company, Incorporated (2013)
15. Reimers, N., Gurevych, I.: Reporting score distributions makes a difference: Performance study of LSTM-networks for sequence tagging. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 338–348 (2017)
16. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). vol. 2, pp. 207–212 (2016)