IxaMed at eHealth-KD Challenge 2019

Using Different Paradigms to Solve Clinical Relation Extraction

Iakes Goenaga[1[0000−0002−0946−5782]], Sergio Santana[2], Sara Santiso[1[0000−0002−2314−3018]], Koldo Gojenola[1[0000−0002−2116−6611]], Alicia Pérez[1[0000−0003−2638−9598]], and Arantza Casillas[1[0000−0226−3352−8200]]

[1] University of the Basque Country (UPV/EHU), Basque Country, Spain
{iakes.goenaga,sara.santiso,koldo.gojenola, alicia.perez,arantza.casillas}@ehu.eus

[2] University of the Basque Country (UPV/EHU), Basque Country, Spain
ssantana005@ikasle.ehu.eus

Abstract. The aim of this paper is to present our approach IxaMed in the eHeath-KD 2019 task. The task consists of identifying different types of entities and the relations between them in the clinical domain in Spanish. The evaluation of the tasks is divided in three scenarios: one corresponding to the detection of medical entities, one corresponding to the detection of the relations between gold standard entities and the third one corresponding to the entire automatic pipeline, that is, entities and relations. In order to carry out the task, we have made use of a Recurrent Neural Network (RNN) to identify medical entities and three different approaches to detect relations between them: a Bi-LSTM with a CRF, a joint AB-LSTM, and a dependency parser. We have achieved a F-score of 0.68 identifying named entities and 0.43 in relation extraction with our best proposal for this task.

Keywords: Medical Entity Recognition · Relation Extraction · Joint AB-LSTM · Bi-LSTM · Dependency parser.

1 Introduction

The aim of this paper is to present the work pursued by the IXAMed team in the eHealth-KD Challenge 2019 [16]. The task consists of identification and classification of terms and relations in medical texts. The evaluation of the task is divided in two sub-tasks: a) identification and classification of key phrases, the goal of this sub-task is to identify all the key phrases per document and their classes, where the key phrases are all the relevant terms (single word or multiple words) that represent semantically important elements in a sentence, and b) the detection of semantic relations between the entities detected in sub-task 1. In order to carry out the task, we have used a Recurrent Neural Network (RNN) for sequence to sequence tagging for the first task, and different approaches for the relation extraction task.

2 Related work

The Clinical E-Science Framework (CLEF) project [17] developed a semantically annotated corpus that consists of both the structured records and free text documents. Two main elements were annotated entities and semantic relations between them. [7] used this corpus to evaluate methodologies on semantic relation extraction. In last years in the different areas various shared tasks have focused on relation classification. In 2010 i2b2/VA Workshop on Natural Language Processing Challenges for Clinical Records [22] organized a task for assigning relation types that hold between medical problems, tests, and treatments, the objective was extracting semantic relations from medical documents. Classification systems that used support vector machines were the most common.

SemEval 2017 Task 10 [1] tackled with extraction of semantic relations between keywords from scientific documents. Some of the participants used SVM classifiers while others released on methods based on neural networks.

In 2018 the Workshop on Semantic Analysis at the SEPLN congress organized Task 3 [13] with the objective of discovering semantic relations between concepts from Spanish electronic health documents. Two methodologies were the most frequent: those that use shallow supervised models (CRF, SVM, etc.) and those that are based on deep learning models.

In [19] the authors experimented with a corpus of Spanish Electronic Health Records to extract semantic relations, in particular adverse drug reactions. In this case they explore different deep learning models. In this paper we are going to evaluate different deep learning methodologies with the aim to extract semantic relations from Spanish electronic health documents.

3 Resources

Apart from the tools we will present in the following subsections, we made use of external data with the intention of completing the information the system extracts from the corpus provided by organization. For this purpose we employed word-embeddings [14] we have calculated (window length = 1, dimensions = 300, algorithm = SkipNgram) from Electronic Health Records (50M words).

4 SubTask A: Medical Entity Recognition

In this section we present our approach in order to identify clinical entities which corresponds to the system we have employed in the Subtask A of the shared task[3].

[3] https://knowledge-learning.github.io/ehealthkd-2019

4.1 Bi-LSTM and CRF

We employed neural network based architecture, more precisely an specific Bi-LSTM (a RNN subclass, [8]) with a CRF on top of it [11, 12] using as input raw text and the word-embeddings we have mentioned in section 3. This kind of neural network is widely used to pursue sequence to sequence tagging [12, 9]. One of the advantages of using Bi-LSTM in contrast to other machine learning techniques such as SVM, Perceptron or CRFs is that the size of the context is automatically learned by the LSTM and there is no need to perform any complicated text preprocessing to obtain features to feed the tool.

5 SubTask B: Relation extraction

In this section we present our three approaches in order to extract relations between entities which they correspond to the systems we have employed in the Subtask B of the shared task.

5.1 Joint AB-LSTM

Contextual information resulted important in related tasks such as Drug-Drug Interaction [18]. Bi-LSTM networks, a type of [20], were demonstrated successful to cope with the context in related works [21, 6, 23]. In particular, we opted for a Joint AB-LSTM network, a particular case of the Bi-LSTM with attention mechanisms. Attention aims at capturing relevant evidences (e.g. phrases around the medical entities) to the relation extraction.

For this work we made use of the implementation provided by [18] adapted, slightly, to this task. In what follows, we summarize the architecture and the adaptations made to cope with this task. The architecture comprises these layers:

- Feature Layer: even though the original implementation just considered the word-features, in this implementation we incorporated, as well, the distance from source to target entity since previous works demonstrated that relations tend to occur within a distance-scope mainly (i.e. the distance is not homogeneously distributed). Admittedly, we did not corroborate this fact for the corpus provided to this task but still used the distance. This is a point that should be explored for future works.
- Embedding Layer: the embeddings corresponding to the aforementioned features are obtained and concatenated. To be precise, the word-embeddings were trained as in [19] while the distance was shelf-trained as an extension to the original architecture. From this layer, the information flows to two Bi-LSTMs. Both Bi-LSTMs are equal (and described below) just differ on the pooling strategy (as stated next). Turning to practical details, we chose a vector of dimension 300 and 50 respectively to embed words and distances.
- Bi-LSTM Layer: this layer comprises two LSTMs devoted to capture, respectively, forward and backward information.

- Pooling Layer: one Bi-LSTM employs attentive pooling and he other one employs max pooling. Next, the features obtained with each pooling stage are concatenated.
- Softmax layer: to the output of the pooling layer the tanh activation function is applied in order to obtain the input of the fully connected layer. Next, the predicted relation is attained by means of the Softmax function. Even though the original architecture was conceived for binary (yes/no) relations, for this task a small adaptation was incorporated to enable multi-class classification (there are a number of available relations).

In the training stage, cross-entropy loss function was employed and Adam algorithm employed [10] for the optimization. The training procedure incorporated two regularization approaches: L2-regularization and Dropout.

This work shows several gaps that we mean to explore for future work. Particularly, we find that the main weakness rests on the superficial experimental framework carried out. For example we should have fine tuned sensitive parameters (e.g. the dimensions were arbitrarily chosen). Besides, in an attempt to enhance contextual information, ad-hoc contextual-embeddings derived from either Elmo [15] or Bert [5] seem of interest to this particular task. Regarding the classification approach, we find important to explore alternative variants: binary (with a specialized classifier per relation type) and multi-class (as we did in this case). Finally, we did not cope, either, with the skewed class distribution even though we realized that some classes are much more frequent than others. To sum up, we are aware of the fact that there is room for improvement.

5.2 Bi-LSTM and CRF

In order to extract the relations that occur between entities, we have made use of the same system that we have used to identify entities. The only difference is that this last system is passed as a parameter the previously detected entities and that it has to predict both the label of the relation between entities and the distance between the entity with which a given entity is linked. Let's take as an example the following phrase: *No existe un tratamiento que restablezca la función ovárica normal.* In this case, as the objective is to identify the relations between entities, the entities will be previously identified as follows: *No existe[B-Action] un tratamiento[B-Concept] que restablezca[B-Action] la función[B-Concept] ovárica[I-Concept] normal[B-Concept].* Therefore, what the system that extracts the different types of relations would show as a result the following: *No existe[TwoTarget] un tratamiento que restablezca[MOneSubject] la función[OneIn-context] ovárica normal.*

In the labels returned by the system we can clearly see two parts, the first part corresponds to the distance from the other entity to which it is attached and the second part to the type of relation they share. If the entity to which an entity is attached occurs earlier in the text, an M is placed in front of the distance between entities (e.g MOneSubject). Conversely, if the entity occurs later in the text, only the distance is set (e.g TwoTarget).

Having said that, in the previous example our system would predict that the entity *existe* is linked to the entity *restablezca* through the *target* relation, the entity *restablezca* with the entity *tratamiento* through the *subject* relation and the entity *función ovárica* with the entity *normal* through the *in-context* relation.

5.3 Mate Parser

Relation extraction can be viewed as the process of obtaining a hierarchical structure where some words are related to others by means of binary relations. One of the words can be considered as the head and the other one as the dependent (e.g. "$afecta \rightarrow asma$" by means of the *subject* dependency relation), and this approach allows the use of standard parsing algorithms. The Mate parser [3] is a development of the algorithms described in [4], that basically adopts the second order maximum spanning tree dependency parsing algorithm. In particular, this parser exploits a hash kernel, a new parallel parsing and feature extraction algorithm that improves accuracy as well as parsing speed [2].

In this particular case, what we have done so that the dependency parser solves the relations between entities has been to train a model of dependencies passing as input each word with its features. These features are the word itself, its entity type and if it has any relation with another word in the sentence, the number of that word in the sentence and the type of relation they share. If a particular word is not linked by any relation with another entity, in the column where the information about the type of relation goes is put a NULL and in the column where the information about the head of the relation goes is put a zero. In figure 1 we can clearly see what we have commented.

```
1	No	No	_	_	O	_	0	NULL	_	_
2	existe	existe	_	_	B-Action	_	0	NULL	_	_
3	un	un	_	_	O	_	0	NULL	_	_
4	tratamiento	tratamiento	_	_	B-Concept	_	6	subject	_	_
5	que	que	_	_	O	_	0	NULL	_	_
6	restablezca	restablezca	_	_	B-Action	_	2	target	_	_
7	la	la	_	_	O	_	0	NULL	_	_
8	función	función	_	_	B-Concept	_	6	target	_	_
9	ovárica	ovárica	_	_	I-Concept	_	0	NULL	_	_
10	normal	normal	_	_	B-Concept	_	8	in-context	_	_
11	.	.	_	_	O	_	0	NULL	_	_
```

Fig. 1. Example of a sentence received as input by the dependency parser in order to learn identifying relations between entities.

6 Results

We present the results we have obtained in all the scenarios in table 1. If we analyze the results we can observe that all the systems have obtained the same results in scenario 2 identifying entities due to we have used the same system (a Bi-LSTM with a CRF) in the three approaches. This result is quite low especially if we compare it with the result that we have achieved in exact match (74.4) in the same test dataset. This difference is due to the penalty received by incorrect matches using the proposed evaluation method of the shared task which combines exact match and partial match results in the same evaluation.

On the other hand, in scenario 3 we have obtained very disparate results. It is curious how the only system designed to extract relations (Joint AB-LSTM) has been the one that has obtained the worst results. Mate dependency analyzer has obtained better results taking into account that it has not been designed for this task and surprisingly we have obtained a good result (43.56) using the same neural tagger we have used to identify entities (Bi-LSTM + CRF).

Finally, in scenario 1 we have obtained very similar results for the three systems. This fact may seem difficult to understand given the different results obtained in scenario 3. The main reason for these results is that the Joint AB-LSTM system did not give any answer (null system) while the other two systems succeeded much more but fail even more in a evaluation method that severely penalizes incorrect and spurious cases.

Table 1. All the results obtained by our different systems for all the scenarios.

	BI-LSTM+CRF	**Joint AB-LSTM**	**Mate**
Scenario 1 (all pipeline)	47.13	48.69	46.09
Scenario 2 (entities)	68.25	68.25	68.25
Scenario 3 (relations)	43.56	17.74	21.94

7 Conclusions

The purpose of this work was to evaluate the feasibility of different approaches to medical entity detection and relation extraction. Entity detection was dealt with a sequential tagger that uses word embeddings acquired from electronic health records. The relation extraction task was approached in three different ways:

- A sequential tagger where, for each dependent, a tag indicates the type and position of its head.
- A neural system that, given two entities, decides if they are related or not, also giving the relation type. This approach can be seen as a classification task, after a previous stage of entity detection.
- Construction of a partial dependency tree where the head of each relation is connected to its dependent.

The different approaches range from the simplest one (a tagger) to the more sophisticate parsing algorithm. Surprisingly, the tagger obtained the best results by an important margin, and this aspect deserves a further study of the strengths and weaknesses of each approach.

Acknowledgments

Funding: This work has been partially funded by Spanish ministry project PROSA-MED(TIN2016-77820-C3-1-R) and a Basque Government Grant for the student Sergio Santana.

References

1. Augenstein, I., Das, M., Riedel, S., Vikraman, L., McCallum, A.: Semeval 2017 task 10: Scienceie-extracting keyphrases and relations from scientific publications. arXiv preprint arXiv:1704.02853 (2017)
2. Bohnet, B.: Very high accuracy and fast dependency parsing is not a contradiction. In: Proceedings of the 23rd International Conference on Computational Linguistics. pp. 89–97 (2010)
3. Bohnet, B., Nivre, J.: A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. pp. 1455–1465 (2012)
4. Carreras, X.: Experiments with a higher-order projective dependency parser. In: Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007. Prague, Czech Republic (June 2007)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
6. Gupta, S., Pawar, S., Ramrakhiyani, N., Palshikar, G.K., Varma, V.: Semi-supervised recurrent neural network for adverse drug reaction mention extraction. BMC bioinformatics **19**(8), 1–7 (2018)
7. Gurulingappa, H., Mateen-Rajpu, A., Toldo, L.: Extraction of potential adverse drug events from medical case reports. Journal of biomedical semantics **3**(1), 15 (2012)
8. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (Nov 1997). https://doi.org/10.1162/neco.1997.9.8.1735, http://dx.doi.org/10.1162/neco.1997.9.8.1735
9. Jagannatha, A.N., Yu, H.: Structured prediction models for rnn based sequence labeling in clinical text. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing. vol. 2016, p. 856 (2016)
10. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations. vol. 5 (2015)
11. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360 (2016)
12. Ma, X., Hovy, E.H.: End-to-end sequence labeling via bi-directional lstm-cnns-crf. In: ACL (1). The Association for Computer Linguistics (2016)
13. 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)
14. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119 (2013)

15. 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). pp. 2227–2237 (2018)
16. 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)
17. Roberts, A., Gaizauskas, R., Hepple, M., Davis, N., Demetriou, G., Guo, Y., Kola, J.S., Roberts, I., Setzer, A., Tapuria, A., et al.: The clef corpus: semantic annotation of clinical text. In: AMIA Annual Symposium Proceedings. vol. 2007, p. 625. American Medical Informatics Association (2007)
18. Sahu, S.K., Anand, A.: Drug-drug interaction extraction from biomedical texts using long short-term memory network. Journal of biomedical informatics **86**, 15–24 (2018)
19. Santiso, S., Pérez, A., Casillas, A.: Exploring joint ab-lstm with embedded lemmas for adverse drug reaction discovery. IEEE journal of Biomedical and Health Informatics (2018)
20. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing **45**(11), 2673–2681 (1997)
21. Stanovsky, G., Gruhl, D., Mendes, P.: Recognizing mentions of adverse drug reaction in social media using knowledge-infused recurrent models. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. pp. 142–151 (2017)
22. Uzuner, Ö., South, B.R., Shen, S., DuVall, S.L.: 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association **18**(5), 552–556 (2011)
23. Wunnava, S., Qin, X., Kakar, T., Rundensteiner, E.A., Kong, X.: Bidirectional lstm-crf for adverse drug event tagging in electronic health records. In: International Workshop on Medication and Adverse Drug Event Detection. vol. 90, pp. 48–56 (2018)