A Hybrid Approach For Automatic Disability Annotation

I. Goenaga, A. Atutxa, K. Gojenola, A. Casillas, A. Díaz de Ilarraza, N. Ezeiza, M. Oronoz, A. Pérez and O. Perez de Viñaspre

IXA Group - University of the Basque Country (EHU/UPV)
Ixa Taldea EHU/UPV Informatika Fakultatea M. Lardizabal 1 20008 Donostia
http://ixa.si.ehu.es/
iakesg@gmail.com

Abstract. The aim of this paper is to present the work pursued by the IXA group in the DIANN-Ibereval 2018 task. The task consists of identifying disabilities within a collection of several abstracts from Elsevier journal papers related to the biomedical domain. These corpora include the annotation of negation when it applies to a disability. The evaluation of the task is divided in two sub-tasks; one corresponding to the detection of English entities and the other to Spanish entities. In order to carry out the task, we have created a pipeline combining a Recurrent Neural Network (RNN) used for sequence to sequence tagging with simple rules to boost the coverage. Our system achieves the best task F-score for both English and Spanish disability identification, showing the suitability of our approach even with quite small training corpus. Our F-score is 0.821 for English and of 0.786 for Spanish.

Keywords: Shared task · Disability identification · Neural Networks.

1 Introduction

The International Classification of Functioning, Disability and Health (ICF) is in charge of providing standards for describing health and health-related states. The ICF defines disability as a term comprising itself several terms such as impairments, activity limitations and participation restrictions. So disabilities although related to diseases are not synonyms of the latter ones. In ICF words **disability and functioning are viewed as outcomes of interactions between health conditions (diseases, disorders and injuries) and contextual factors**.

To our knowledge, DIANN-IberEval 2018 [2] represents the first attempt towards automatic tagging of disabilities in Spanish and English texts. Turning to the literature, there has been some work devoted to Medical Entity Recognition (MER) in both Spanish and English. In Spanish, [11] pursued a MER task where the goal consisted in automatically identifying disorders and drug mentions on a corpus conformed of clinical reports. The authors employed supervised machine learning algorithms (SVM, Perceptron and CRF) in conjunctions with multiple features (POS, embedding clusters, Brown clusters among others). The results showed that the combination of supervised algorithms and no-supervised features significantly improved standard supervised techniques.

[9] present a hybrid system based on both the use of gazetteers (more precisely SnomedCT) and embeddings learned in an unsupervised manner. Basically, for each paragraph (P) in a clinical report they modeled through a function they called DNER(t,P), the degree in which the term t is named in the phrase P, where the term t corresponds every noun-phrase obtained from a SnomedCT description. DNER(t,P) uses both n-gram distance and similarity between the embeddings of the terms from the document and the distance of the "word vector" with the SnomedCT description term. They did not report any results what so ever.

It is also worth mentioning two other well known task related to the actual DIANN-IberEval task, namely SemEval-2014 task 7 [12] and SemEVal-2015 Task 14 [13]. In both competitions, there were two subtasks. One consisted in performing the identification of diseases and certain other medical terms that the human taggers considered to be relevant for some reason. The other was a normalization task where each identified term had to get a unique UMLS/SNOMED-CT ontology CUI (Concept Unique Identifier) assigned. In SemEval tasks, disorders could be either continuous (*lower extremity DVT*) or discontinuous spans (*tumor ... ovary*). Both DIANN-IberEVal and SemEval task evaluations are similar. In both partial and exact recognition are measured in order to compare the performance of the different systems.

UTH-CCB team was the winner of the identification task in SemEval-2014 with a 0,81 and a 0,90 F-score on exact and partial evaluations respectively. They used two machine learning algorithms based on CRF and SVM in addition to MetaMap. They employed several features such as Bag Of Words, POS (from Stanford tagger), type of notes, EHR section information, word representations (Brown clustering), random indexing and semantic categories (UMLS lookup, MetaMap and cTAKES). They also applied three types of ensemble on these basic algorithms: a machine learning ensemble, a majority vote and a direct merging of all.

In SemEval-2015 the winner was the ezDi team with a 0.75 and 0.78 F-score on exact and partial evaluations respectively. SemEval-2015 evaluation is not directly comparable with either SemEval-2014 nor DIANN-IberEval because it did not only consist in correctly identifying the term. Correct CUI assignment was included in the evaluations as well, and therefore results are lower than those of the previous year. Exact identification comprised perfect CUI assignment and whole entity identification. Partial match comprised perfect CUI assignment and partial entity identification. EzDi team used both CRF for simple entity recognition with the following features: BOW (window +/− 2),word stemmer, prefix-suffix length 1-5, orthographic features (word contains digit, slash, special char, word shape-kind of reg. exp.), grammatical features (POS, chunk, head of NP/VP), dictionary look-up matches (window +/− 2), section header, document type and sentence cluster id. And an SVM for discontinuous entities with syntactic features like POS, chunk labels between candidates, rules based on the

Charniak parser to find relations, position of prepositions, conjunctions, main verb in context of first candidate, a binary feature indicating which candidates contain NP head and finally lexical features like Bag of Words.

Lately, Jagannatha et al. [4, 5] used different Recurrent Neural Networks to pursue MER. They obtained their embeddings applying Word2Vec's skip-gram algorithm (SkipG henceforth) over PubMed articles, English Wikipedia and 100,000 EHRs. The results show that all RNN models significantly outperform the baseline while their best system improved the recall, precision and F-score of the baseline by 19%, 2% and 11% respectively.

There are also several well known ruled-based or mapping tools for Named Entity Recognition. MedLEE: The Medical Language Extraction and Encoding System (MedLEE) is a rule-based tool. MetaMap: MetaMap maps any texts to the UMLS Metathesaurus. Recognizing terms contained in UMLS and certain variations of those terms. To finish, cTAKES is an information extraction system from clinical narratives. It is an open source project to perform clinical information extraction task, mapping terms to UMLS concepts, and accomplish syntactic and semantic parsing.

2 System Description

We have divided this section in two parts. In 2.1 we are going to explain the external data we used, while in 2.2 we will focus on the pipeline that combines neural networks and rules.

2.1 External Data

We made use of external data with the intention of completing the information the system extracts from the corpus provided by organization. We employed Brown clusters [1] and word-embeddings [10]. For English, we extracted Brown clusters and word-embeddings from MIMIC-III corpus [6]. For Spanish, we calculated the word-embeddings from Electronic Health Records and we did not include any Brown cluster.

2.2 Hybrid Disability Detection Pipeline

In order to identify disabilities, negated disabilities and the negation scope we designed an hybrid system that combines neural networks and rules. First of all we present our module to identify disabilities, negated or not, using neural network based architecture. Then, we focus on the rule-based module to identify negation triggers. On the top of that, we also included a rule-based module that helps identifying those disabilities the neural network can not identify. Finally we explain how to detect negation scopes with a neural network.

Neural Network-Based Disability Identification We employed neural network based architecture, more precisely an specific Bi-LSTM (a RNN subclass, [3]) with a CRF on top of it [7, 8] using as input raw text, word-embeddings and Brown clusters. This kind of neural network is widely used to pursue sequence to sequence tagging [8, 5]. 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.

Rule-Based Negation Trigger Identification In order to identify negated disabilities we created a list of negation triggers for each language and this module labels them as negation if they are close (maximum one word distance) to disabilities identified by the neural network. The list of negation triggers for English is *lack of*, *without*, *with no*, *no*, *not*, *no signs* and *no evidence of*, and for Spanish is *sin*, *ausencia*, *no*, *falta de*, *no hay evidencia* and *sin evidencia*.

Rule-Based Disability Identification This module is responsible of detecting the acronyms of disabilities that are close to the disabilities (maximum one word distance) identified by the neural network. Once the acronyms are detected the module labels them as disabilities in the entire text.

Neural Network-Based Negation Scope Identification The main objective of this last module is to identify the negation scope. Although the used neural network is the same we use to identify disabilities, this time we use the output of the previous modules as features instead of using Brown clusters.

3 Results and Discussion

Three tracks have been evaluated in the present shared task: Disabilities included or not in negation, non-negated disabilities and negated disabilities, and disabilities, negation triggers and the scope of negations. In addition, two types of matching have been used for the evaluation: partial and exact. We show the results obtained by all the systems for English and Spanish in table 1.

If we analyze the results, the good performance of the system in almost all the tracks is notorious both in exact match evaluation as in partial match evaluation. These results have helped us to beat the results of the rest of the systems in the shared task in almost all the tracks. However, there has been a track where the systems results have fallen short, specifically in Negated Disability track for English (exact match). For the evaluation of this track, it is considered as negated disability the set of annotations disability, negation trigger and scope of the negation. With the intention of clarifying the reasons for these low results we have carried out an error analysis and we have concluded that there are three main reasons for these results:

	Disability				**Neg Disability**				**NN + Neg Disability**			
	English		**Spanish**		**English**		**Spanish**		**English**		**Spanish**	
Team	**E**	**P**	**E**	**P**	**E**	**P**	**E**	**P**	**E**	**P**	**E**	**P**
Ours	**0.82**	**0.88**	**0.78**	**0.85**	0.45	**0.95**	**0.80**	**0.90**	**0.77**	**0.87**	**0.77**	**0.84**
UC3M	0.74	0.79	0.72	0.79	0.00	0.75	0.00	0.81	0.68	0.76	0.65	0.76
UPC3	0.68	0.75	0.69	0.76	**0.75**	0.93	0.57	0.84	0.66	0.74	0.65	0.74
UPC2	0.64	0.70	0.59	0.67	0.55	0.80	0.68	0.82	0.60	0.68	0.57	0.66
LSI	0.63	0.80	0.31	0.65	0.15	0.71	0.00	0.23	0.60	0.78	0.30	0.60
IXA	0.60	0.65	0.64	0.72	0.52	0.78	0.72	0.74	0.56	0.63	0.62	0.70
SINAI	0.46	0.51	0.39	0.43	0.47	0.90	0.16	0.24	0.42	0.50	0.33	0.38
GPLSIUA	0.39	0.40	0.28	0.33	0.55	0.80	0.00	0.15	0.36	0.41	0.20	0.26

Table 1. The best F-score results achieved by the participants for each track and each evaluation. The best results among all participants in bold. Neg = Negated, NN = Non-Negated, E = Exact and P = Partial.

- The tendency of the system to include the previous verb to the negation trigger in the negation scope:
  - **showed** no cognitive impairment.
  - **had** no cognitive deterioration.
- The tendency of the system to finish the negation scope with the last annotated disability:
  - No **cognitive deterioration** was found.
  - no **cognitive impairment** according to Reisberg's global deterioration scale (GDS).
- The difficulty of catching all the disabilities and negation triggers. In that way, without correctly annotated disabilities or negation triggers is really difficult for the neural network to catch negation scopes and is not possible to perform well in exact match evaluation.

Taking that into account we realize there are some aspects of the system that can be improved. Nevertheless, we consider the performance of the system proposed by us has been really good and the suitability of it for this task is undeniable.

4 Conclusions

This paper presents a hybrid pipeline combining neural networks and rules for disability identification in clinical texts that provide the best results among the systems presented in the shared task.

A key aspect in order to achieve this performance is the complementarity of neural network-based modules and rule-based modules. To conclude, although it is clear the most important modules are neural network-based ones, those that make the difference are the rule-based modules.

Acknowledgements

This work has been partially funded by:

- The Spanish ministry (projects TADEEP: TIN2015-70214-P, PROSA-MED: TIN2016-77820-C3-1-R).
- The Basque Government (projects DETEAMI: 2014111003, ELKAROLA:KK-2015/00098).

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

References

1. Brown, P.F., Desouza, P.V., Mercer, R.L., Pietra, V.J.D., Lai, J.C.: Class-based n-gram models of natural language. Computational linguistics **18**(4), 467–479 (1992)
2. Fabregat, H., Martinez-Romo, J., Araujo, L.: Overview of the diann task: Disability annotation task at ibereval 2018. In: Proceedings of the Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018) (2018)
3. 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
4. Jagannatha, A.N., Yu, H.: Bidirectional rnn for medical event detection in electronic health records. In: Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting. vol. 2016, p. 473. NIH Public Access (2016)
5. 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)
6. Johnson, A.E., Pollard, T.J., Shen, L., Lehman, L.w.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., Mark, R.G.: Mimic-iii, a freely accessible critical care database. Scientific Data **3** (May 2016)
7. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360 (2016)
8. 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)
9. Martinez Soriano, I., Castro, J.L.: Dner clinical (named entity recognition) from free clinical text to snomed-ct concept. WSEAS Transactions on Computers **16** (2017)
10. 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)
11. Pérez, A., Weegar, R., Casillas, A., Gojenola, K., Oronoz, M., Dalianis, H.: Semi-supervised medical entity recognition: A study on spanish and swedish clinical corpora. Journal of Biomedical Informatics **71** (2017)
12. Pradhan, S., Elhadad, N., Chapman, W.W., Manandhar, S., Savova, G.: SemEval-2014 Task 7: Analysis of Clinical Text, pp. 54–62 (2014)
13. Pradhan, S., Elhadad, N., Chapman, W.W., Manandhar, S., Savova, G.: Semeval-2015 task 14: Analysis of clinical text. In: SemEval@NAACL-HLT. pp. 303–310. The Association for Computer Linguistics (2015)