Improving Named Entity Recognition for Biomedical and Patent Data Using Bi-LSTM Deep Neural Network Models

Farag Saad(✉), Hidir Aras, and René Hackl-Sommer

FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Karlsruhe, Germany
{farag.saad,hidir.aras,rene.hackl-sommer}@fiz-karlsruhe.de

Abstract. The daily exponential increase of biomedical information in scientific literature and patents is a main obstacle to foster advances in biomedical research. A fundamental step hereby is to find key information (named entities) inside these publications applying Biomedical Named Entities Recognition (BNER). However, BNER is a complex task compared to traditional NER as biomedical named entities often have irregular expressions, employ complex entity structures, and don't consider well-defined entity boundaries, etc. In this paper, we propose a deep neural network (NN) architecture, namely the bidirectional Long-Short Term Memory (Bi-LSTM) based model for BNER. We present a detailed neural network architecture showing the different NN layers, their interconnections and transformations. Based on existing gold standard datasets, we evaluated and compared several models for identifying biomedical named entities such as chemicals, diseases, drugs, species and genes/proteins. Our deep NN based Bi-LSTM model using word and character level embeddings outperforms CRF and Bi-LSTM using only word level embeddings significantly.

Keywords: Biomedical · NER · Deep neural network · Bi-LSTM · CRF · Patent

1 Introduction

We have witnessed a massive growth of information in the biomedical domain in the last few decades due to the abundant research on various diseases, drug development, and gene/protein identification etc. However, a large percentage of the information related to the biomedical domain is available as unstructured document publications such as scientific articles, patents etc. In order to effectively exploit such unstructured resources, research in biomedical named entity recognition (BNER) is one of the most promising techniques for automating the utilization of biomedical data. Furthermore, BNER is considered an initial step for many downstream tasks, such as relation extraction, question answering, knowledge base completion, etc. [2].

E. Métais et al. (Eds.): NLDB 2020, LNCS 12089, pp. 25–36, 2020.
https://doi.org/10.1007/978-3-030-51310-8_3

Identifying biomedical entities is not a trivial task due to many factors such as complex entity structures, fuzzy entity boundaries, abundant use of synonyms, hyphens, digits, characters, and ambiguous abbreviations, etc. Despite significant efforts for building benchmark datasets to develop BNER, these datasets are still far from being optimal in quality and in size to speed up the development of BNER tools. For patents, the problem is even more complex as it is not easy to process the patent text due to peculiarities such as usage of generic terms, paraphrasing, and vague expressions, which makes it harder to narrow down the scope of the invention. This causes important contextual information to be lost, which has a negative effect on the performance of BNER tools [18]. A patent is a very important resource to consider as new chemical or biomedical entities are often shown in patent documents before they are even mentioned in the chemical or biomedical literature making patents a valuable, but often not a fully discovered resource. Furthermore, it is estimated that a significant portion of all technical knowledge is exclusively published in patents. For example, two-thirds of technical information related to the medical domain did not appear in non-patent literature [15].

On the basis of the encouraging results we have achieved in our ongoing work for using deep neural network models for the BNER task [17], in this paper, we show our improved deep learning approach that we evaluated on large biomedical datasets for the following biomedical entity types: chemical, disease, drug, gene/protein, and species. Moreover, we show the specific details of the developed neural network architecture and how the various neural network layers are designed to transform an input to a desired output – enabling the neural network to reduce the error/loss and optimize the learning task.

In the following, we firstly review the related work in Sect. 2, followed by a presentation of the proposed approach in Sect. 3. In Sect. 4 an empirical evaluation is presented and discussed. A conclusion is given in Sect. 5.

2 Related Work

In the literature, NER approaches are generally classified into hybrid (rule-based and statistical), supervised (feature-based) and unsupervised learning approaches [21]. In the biomedical domain, for example with regard to the hybrid approach, a two-fold method for Biomedical NER was proposed in which dictionary-based NER was combined with corpus-based disambiguation [1]. Due to the fact that a biomedical named entity can exist in different written forms, e.g., "*SRC1*", "*SRC 1*", and "*SRC-1*", performing the exact match of the biomedical named entity in a given text with the dictionary terms can result in very low coverage. Therefore, different forms of the same named entity were normalized and transformed into a unified representation. However, words from a common vocabulary may be mistakenly recognized as biomedical named entities. In order to tackle this issue a corpus-based disambiguation approach to filter out mistakenly recognized biomedical named entities was applied. The disambiguation process was accomplished based on a machine-learning classifier trained on an annotated corpus.

Tanabe and Wilbur used a combination of a statistical and a knowledge-based approach to extract gene and protein named entities from biomedical text [20]. First, a Brill POS tagger[1] was applied to extract candidates. These were then filtered based on manually curated rules, e.g., morphological clues, to improve the extraction accuracy. Furthermore, a Bayesian classifier was applied to rank the documents by similarity according to documents with known genes and proteins in advance. Hanisch et al. proposed *ProMiner*, which used a pre-processed synonym dictionary to extract gene and protein named entities from biomedical text [10]. ProMiner is composed of three parts, gene and protein named entity dictionary generation, gene/protein occurrence detection, and filtering of matched entities. Rule-based approaches are expensive and time consuming as rules need to be modified each time the data changes. Furthermore, rule-based approaches are usually domain dependent and cannot be smoothly adapted to a new domain.

In recent years, with the availability of the annotated biomedical corpora, several supervised approaches have been developed. For example, *GENIA*[2] and *BioCreative*[3] corpora were intensively used in supervising learning approaches such as Support Vector Machines (SVMs) [22], Conditional Random Fields [19] etc. In [22] Yang and Li proposed a SVM-based system, named *BioPPISVMExtractor*, to identify protein-protein interactions in biomedical text. Features that were set, such as word feature, protein names distance feature, link-path feature etc. were used for SVM classification. Based on these rich features, the SVM classifier was trained to identify which protein pairs have a biological relationship among them. In [19], Settles used the CRF (Conditional Random Field Approach) approach to recognize genes and proteins named entities with a variety of rich features set such as orthographic and semantic features obtained from a lexicon, etc. Based on the fact that a contextual feature is very important for the performance of the CRF approach, Settles models the local context feature by considering neighboring words, one word before and one word after the word in focus, besides other features for improving the sequence labeling task.

The clustering approach is considered as a standard unsupervised approach for biomedical NER. The assumption behind this unsupervised approach is that the named entities in the biomedical text can be clustered based on their contextual similarity. For example, in [23] Zhang and Elhadad proposed an unsupervised approach to extract named entities from biomedical text. The classification approach does not rely on any handcrafted rules, heuristics, or use of annotated data. It depends on corpus statistics and shallow syntactic knowledge, e.g., noun phrase extraction. Han et al. proposed a novel clustering based active learning method for the biomedical NER task [9]. They compared different variations of the proposed approach and discovered the optimal design of the active learning method. This optimal design employs the use of the vector representation

[1] https://www.npmjs.com/package/brill-pos-tagger.
[2] http://www.geniaproject.org/.
[3] https://biocreative.bioinformatics.udel.edu/.

of named entities, and the selection of documents that are representative and informative.

In the past few years, Deep Learning approaches for the NER task (mainly LSTM = Long Short-Term Memory) became dominant as they outperformed the state-of-the-art approaches significantly [5]. In contrast to feature-based approaches, where features are designed and prepared through human effort, deep learning is able to automatically discover hidden features from unlabelled data[4]. The first application for NER using a neural network (NN) was proposed in [3]. In this work, the authors used feature vectors generated from all words in an unlabelled corpora. A separate feature (orthographic) is included based on the assumption that a capital letter at the beginning of a word is a strong indication that the word is a named entity. The proposed controlled features were later replaced with word embeddings [4]. Word embeddings, which are a representation of word meanings in n-dimensional space, were learned from unlabelled data. A major strength of these approaches is that they allow the design of training algorithms that avoid task-specific engineering and instead rely on large, unlabelled data to discover internal word representations that are useful for the NER task.

The prowess of such approaches has since been observed many times. In the BioCreative V CEMP[5] and GPRO[6] tasks, the best algorithm combined deep learning and CRF [14]. Finally, in a study covering 33 datasets, an approach combining deep learning and CRF outperformed not only a plain CRF-based approach, but also entity-specific NER methods (e.g., a dictionary) in recall and F1-score [8]. In recent work, the original BERT (Bidirectional Encoder Representations from Transformers) model [6] was applied for the BNER task, e.g., to train models with biomedical text (BioBERT) [13]. Based on the achieved experimental results, BioBERT, e.g., slightly outperformed the BNER state-of-the-art approaches with 0.62% F-measure improvement and gained a significant improvement (12.24% MRR -Mean Reciprocal Rank-) for the biomedical QA task. The advantage of using the BERT model over other models is that it takes into account polysemous words. For example, Word2Vec produces only one embeddings vector for the polysemous word “apple”, while the BERT model produces different embeddings: one embedding for the fruit and another one for the smart phone brand, etc.

3 Bidirectional Long-Short Term Memory (Bi-LSTM)

LSTM is a special case of Recurrent Neural Network (RNN) which is capable of remembering information of larger contexts. RNN is the most used approach for sequence labelling tasks due to its ability to consider a richer context compared to the standard Feed Forward Neural Network (FFNN). The main fundamental

[4] Readers interested in a more introductory text on the topic may wish to refer to [7].
[5] Chemical Entity Mention in Patents.
[6] Gene and Protein Related Object.

difference between the architecture of a RNN and a FFNN is that the information flow in the RNN is cyclic while the information flow in the FFNN only moves in one direction (feed and then forward). Each node in a RNN is making the prediction based on the current input into the RNN node and the past output of the same node. This mechanism makes RNN ideal for learning time-sensitive information as it doesn't neglect previous input. However, RNN suffers from the vanishing gradient problem which hinders handling wider contexts [11]. The reason therefore is that when fine-tuning the weights during the back propagation, the weight values update of the early layers will be strongly dependent on the weight values of the later layers. When the weight values of the later layers are very small (closer to zero), the weight values of the early layer will vanish very quickly, making it impossible for the RNN to learn the task effectively.

In LSTM instead of having a node with a single activation function as it is the case in RNN, the LSTM nodes can act as a memory cell which is able to store different types of information using a gate mechanism. Gates in LSTM regulate the flow of information, e.g., forget gates do not allow irrelevant information to pass through. There are two types of LSTM, unidirectional LSTM, which can handle information from the past, and bidirectional LSTM (Bi-LSTM), which can handle information from the past and from the future. One LSTM performs a forward operation so it can handle the information from the past and the second LSTM performs the backward operation so it can handle the information from the future and hence consider a wider context which can help with the predicting task. For more detailed information about the conceptual idea of the LSTM approach we refer the reader to the work proposed in [12].

The architecture of the Bi-LSTM deep neural network model is illustrated in Fig. 1. The input to the model is the *"mutant superoxide dismutase-1 gene"* sequence. The word embeddings is learned using unlabelled datasets, e.g., for chemical, drug, disease, gene, protein, etc. We used the Bi-LSTM to encode character-level information of a given word into its character-level representation (embeddings). If we consider the *"mutant"* token as an example, its characters *"M U T A N T"* will be used as input into the Bi-LSTM model and hence its character-based representation is generated. A combination of the character-based embeddings and the corresponding word embeddings which were generated using an unlabelled dataset will be the input to the Bi-LSTM layer. The result of this step is a richer contextual representation (vector representation) for the input sequence, which will be the input to the CRF model layer for the best label sequence tagging generation. The tagging layer (the CRF model) uses a probabilistic sequence-labelling model for sequence tagging. The CRF model takes as input a token sequence and assigns the most related label to each token based on the training dataset (see Sect. 4.1). As it is possible that a named entity spans over multiple tokens, and in order to tackle this issue, we used the *IOB* format scheme to define the entity boundaries. The training dataset represents the corresponding IOB-tags where IOB refers to *Inside*, *Outside* and *Beginning* and it is widely used as an encoding scheme for the NER task. Words tagged with *"O"* are outside of named entity, whereas words tagged with *"I"* lie inside

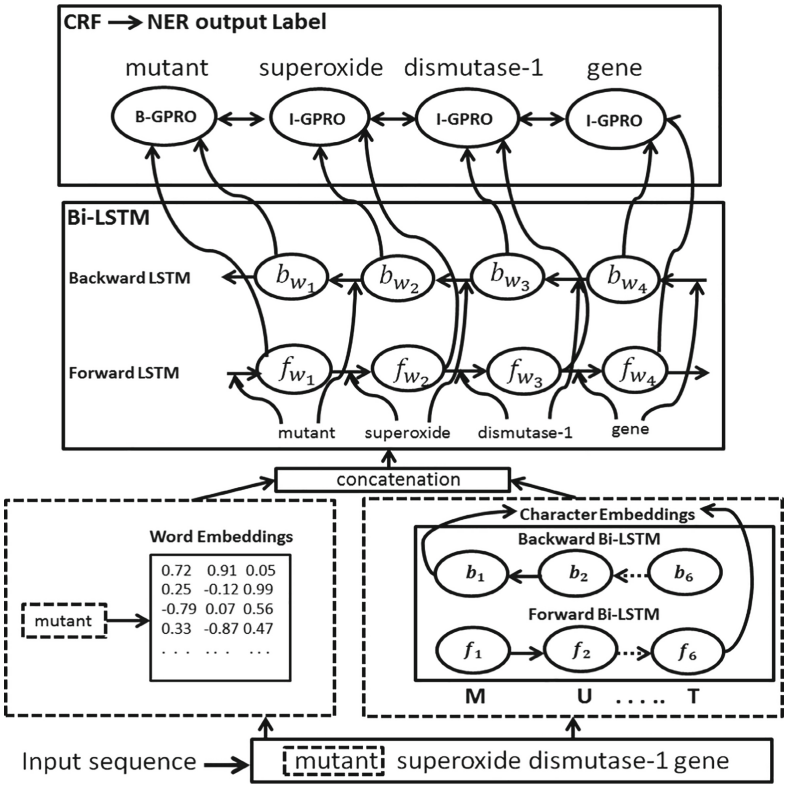

Fig. 1. The overall architecture of the Bi-LSTM-CRF Model for BNER.

of a named entity. *"B"* refers to words that represent the beginning of a named entity. To tag each token in the given sequence, the CRF model builds a set of inference rules based on the training corpus and the refined context obtained by the Bi-LSTM model. For the algorithm implementation we used the default value settings, e.g., embeddings dimensions of value 300, dropout of value 0.5, epochs of value 25, batch size of value 20. lstm size of value 100 etc.

4 Evaluation

In this section we present our empirical study of biomedical NER applied on various biomedical datasets obtained from biomedical literature and patents. Next, we briefly describe the datasets we used for training, word embedding generation, and evaluation of the proposed algorithm variants.

4.1 Dataset

We evaluated and trained several models on six different datasets employing five entity types: chemical, gene/protein, disease, drug, and species. Four datasets

(chemical/drug, gene, disease and species) were acquired from biomedical literature while two datasets (chemical and gene/protein) were acquired from patents belonging to various patent offices (cf. Table 1). All datasets were manually annotated by domain experts. The BC4CHEMD, BC2GM, CEMP and GPRO datasets were obtained from BioCreative[7], the NCBI-disease dataset from the National Center for Biotechnology Information[8], while the Linnaeus dataset was obtained from the Linnaeus website[9].

Table 1. The number of training and test instances for each dataset.

Dataset	Type	Training instances	Test instances
BC4CHEMD	The BioCreative IV chemical and drug	30682	26364
BC2GM	The BioCreative II gene	12574	5038
NCBI-disease	Diseases	4560	4797
Linnaeus	Species	11935	7142
CEMP chemical patent	Chemical	43307	19274
Chemdner GPRO patent	Gene/protein	10249	5652

Word embeddings are usually represented by lower-dimensional vectors of mostly up to 300 words in length, e.g., the vector of the word disease is very close to the vector's representation of, chronic, disorder, treatment, drugs etc. These relationships between vectors are not explicitly enforced by humans during training instead they are learnt by the training algorithm in an unsupervised manner based on large unlabelled datasets. The unlabelled datasets which we used to generate the word embeddings model are obtained from PubMedCentral (PMC)[10] full-text articles, English Wikipedia[11] full-text articles, and a combination of them. We downloaded this data and performed basic cleansing steps such as removing unnecessary tags, references, authors sections etc. We then built the word embeddings models using the GloVe algorithm [16] based on a vector size of 300 and a contextual window size of 15.

Character embeddings can be used to improve the semantic representation of some words. Using word embeddings, we obtained the vector representations of most of the words included in the unlabelled dataset. However, in some cases word embeddings are not enough and won't capture all words such as out-of-vocabulary (OOV) words, different written forms of the same entity, misspelled words, etc. To identify such words, character embeddings are used to generate vector representations of words by considering their character-level structure, e.g., "alpha-lipoic-acid" and "α-lipoic-acid" will be considered as the same even though they are not orthographically similar.

[7] https://biocreative.bioinformatics.udel.edu/resources/.
[8] https://www.ncbi.nlm.nih.gov/CBBresearch/Dogan/DISEASE/.
[9] http://linnaeus.sourceforge.net/.
[10] ftp://ftp.ncbi.nlm.nih.gov/pub/pmc/oa_bulk/ (until 12/2019).
[11] https://dumps.wikimedia.org/enwiki/latest/ (until 12/2019).

4.2 Experiments

We conducted five experiments as the baseline using the CRF approach. We then compared the results of the CRF approach with two variants of the Bi-LSTM model based on word- and character-level representations. We compared all methods in terms of precision, recall, and F1 score over the six test datasets. For the biomedical literature: Chemical and Drug (BC4CHEMD) with 26,364 test instances, Genes (BC2GM) with 5038 test instances, Disease (NCBI-Disease) with 4797 test instances and Species (Linnaeus) with 7142 test instances (See Table 1). For patent test datasets: Chemical (CEMP) with 19,274 test instances and Gene/Protein (GPRO) with 5652. The evaluation for the deep learning approach variants Bi-LSTM and CHARS-Bi-LSTM was performed based on the embedding vectors described in Sect. 4.1.

Table 2 shows the evaluation results of comparing both the Bi-LSTM and CHARS-Bi-LSTM models. The first embeddings model is learned based on the unlabelled dataset of PMC while the second embeddings model is learned using a combination of PMC and Wikipedia. The second word embeddings model was used to evaluate whether the combined embeddings model will have a significant impact on the Bi-LSTM model's performance. As shown in Table 2, for the BC4CHEMD test dataset (Chemical & Drug), the CHARS-Bi-LSTM model trained on the PMC unlabelled dataset achieved a significantly higher precision, recall, and F-measure with 0.90, 0.93, and 0.91, respectively, compared to the results of the CRF model (e.g., recall was improved by 15%) and compared to the Bi-LSTM model (e.g., recall was improved by 8%). However, the CHARS-Bi-LSTM model using word embedding trained on a combination of PMC and Wikipedia achieved a minor precision improvement by 1%, and has a drop of recall by 2% while the F-measure remains the same. The same applies for the BC2GM dataset where the CHARS-Bi-LSTM model using word embeddings trained on PubMed achieved a significant improvement over CRF (e.g., recall by 16%) and over Bi-LSTM (e.g., recall by 6%).

Using a word embedding trained on PMC and Wikipedia leads to a minor decrease in recall and F-measure by 1%. For the other test datasets (disease and species), the CHARS-Bi-LSTM trained on a combination of PMC and Wikipedia achieved a better improvement over the CHARS Bi-LSTM model using word embeddings trained only on a PMC dataset. For the disease dataset, precision improved by 10% while for species remains the same and F-measure improved by 4% while for species improved by 5%. Recall decreased by 1% while for species improved by 8%. The improvement can be interpreted as that Wikipedia is a significant resource for diseases and species, providing a richer data resource for the word embeddings learning task. For chemical and gene/protein patent test datasets, adding the Wikipedia data had almost no impact on the Bi-LSTM performance. This is due to the nature of the patent text since newly invented entities usually do not show up immediately in other resources, e.g., Wikipedia. To improve the BNER for biomedical data in patent resources, we built a new word embedding model trained in patent text to evaluate whether the developed patent model can raise the Bi-LSTM model's performance. We collected 1.5

million titles and abstracts obtained from EPO[12], USFULL[13] and PCTFULL[14] patent databases. We then kept only documents that belong to the life science domain by filtering over the International Patent Classification (IPC) code. Next, we combined the patents with PMC and applied the GloVe algorithm on the combined unlabelled dataset to build the word embedding model.

Table 2. Precision, Recall and F-measure of CRF and Bi-LSTM variants using various Word and Character level embeddings

Method	Word embeddings	Metrics	Test datasets					
			D1	D2	D3	D4	D5	D6
CRF	-	Precision	0.89	0.80	0.89	0.95	0.92	0.82
		Recall	0.78	0.72	0.76	0.49	0.87	0.74
		F-measure	0.83	0.76	0.81	0.62	0.90	0.79
Bi-LSTM-CRF	PubMed	Precision	0.87	0.78	0.87	**0.98**	0.92	0.81
		Recall	0.85	0.78	0.77	0.76	0.89	0.82
		F-measure	0.86	0.78	0.82	0.85	0.90	0.81
Bi-LSTM-CRF	PubMed + Wikipedia	Precision	0.85	0.79	0.97	**0.98**	0.92	0.80
		Recall	0.86	0.78	0.84	0.84	0.90	0.82
		F-measure	0.86	0.78	0.90	0.91	0.91	0.81
CHARS-Bi-LSTM-CRF	PubMed	Precision	0.90	**0.83**	0.88	**0.98**	0.93	0.82
		Recall	**0.93**	**0.84**	0.85	0.82	0.94	0.88
		F-measure	**0.91**	**0.84**	0.87	0.89	0.93	0.85
CHARS-Bi-LSTM-CRF	PubMed + Wikipedia	Precision	**0.91**	**0.83**	**0.98**	**0.98**	**0.94**	**0.84**
		Recall	0.91	0.83	**0.86**	**0.90**	0.95	0.87
		F-measure	0.90	0.83	**0.92**	**0.94**	**0.94**	0.85
CHARS-Bi-LSTM-CRF	PubMed + Patent	Precision	–	–	–	–	0.91	0.83
		Recall	–	–	–	–	**0.97**	**0.90**
		F-measure	–	–	–	–	**0.94**	**0.86**

Remarks: D1 refers to the BC4CHEMD, D2 refers to the BC2GM, D3 refers to the NCBI-Disease, D4 refers to the Linnaeus, D5 refers to the CEMP and D6 refers to the GPRO datasets

Using this combined word embeddings model applied on the patent chemical and gene/protein test datasets leads to a minor improvement of the CHARS-Bi-LSTM model (average recall improvement of 2%, see Table 2). This is an indication that patent data word embeddings models could slightly help to recognize more entities. For future evaluation, we will further increase the size and the focus of the patent data to include more chemical genes/proteins so a significant assessment can be performed.

Overall, the CHARS-Bi-LSTM model trained using character and word level embeddings achieved superior performance compared to the CRF and Bi-LSTM using only word embeddings. This indicates that character-level embeddings can be useful in handling out-of-vocabulary words, misspelled words, different forms of the same entity, etc., and hence the character-level representation is

[12] https://publication.epo.org/raw-data/product-list.
[13] http://patft.uspto.gov/.
[14] https://stn.products.fiz-karlsruhe.de/sites/default/files/STN/summary-sheets/PCTFULL.pdf.

significantly able to infer a representation of unseen words in the training data and increase the Bi-LSTM model performance.

4.3 Application

In the following patent use case, we illustrate how BNER can be used for improving patent retrieval for discovering relevant inventions, technologies, and detailed information from text. As an example, a key term search for finding biotechnologies related to *biosensor devices* in medicine could be initiated using the key term "biosensor device". As biosensors are devices which have a broad range of applications such as in medicine, environmental research, agriculture, etc. a more-fine grained (entity-based) retrieval is required for finding more precise results. In our example, we can utilize biomedical annotations in order to narrow down our search to focus the domain of interest like biosensor device usage in medicine, e.g., DNA hybridization detection, glucose measurement, antibody detection, etc. In a different example, in case of the "*biosensor device*" query, the patent retrieval system will respond by suggesting specific biomedical terms, e.g., "*miRNA*", which are related to the usage of biosensor devices in the biomedical domain. As a result, specific patents related to "miRNA" and biosensors can be retrieved more efficiently, e.g., "Method for preparing self-energized miRNA biosensor", "Biological probe and detection method for detecting miRNA and application", etc.

5 Conclusion

We have presented a deep neural network architecture based on Bi-LSTM and a setting for the efficient recognition of different classes of biomedical named entities. To achieve that goal, we have built and utilized several pre-trained embeddings models based on word and character level embeddings. Our experiments show that combining heterogeneous pre-trained word embedding models allows us to achieve better results in recognizing various types of biomedical named entities. For example, a small pre-trained patent word embeddings model combined with the PMC model has shown an improvement in the patent BNER task. Overall, the CHARS-Bi-LSTM model, trained using character and word level embeddings, achieved superior performance compared to the traditional CRF and Bi-LSTM approach using only word embeddings. This indicates that character-level embeddings seem to be very useful in handling out-of-vocabulary words, misspelled words, different forms of the same entity, etc.

References

1. Basaldella, M., Furrer, L., Tasso, C., Rinaldi, F.: Entity recognition in the biomedical domain using a hybrid approach. J. Biomed. Semant. **8**, 51 (2017)
2. Cokol, M., Iossifov, I., Weinreb, C., Rzhetsky, A.: Emergent behavior of growing knowledge about molecular interactions. Nat. Biotechnol. **23**(10), 1243–1247 (2005)

3. Collobert, R., Weston, J.: A unified architecture for natural language processing: deep neural networks with multitask learning, pp. 160–167 (2008)
4. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.P.: Natural language processing (almost) from scratch. Computing Research Repository - CORR abs/1103.0398 (2011)
5. Dang, T.H., Le, H.Q., Nguyen, T.M., Vu, S.T.: D3NER: biomedical named entity recognition using CRF-BiLSTM improved with fine-tuned embeddings of various linguistic information. Bioinformatics **34**(20), 3539–3546 (2018)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the ACL: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186 (2019)
7. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
8. Habibi, M., Weber, L., Neves, M., Wiegandt, D.L., Leser, U.: Deep learning with word embeddings improves biomedical named entity recognition. Bioinformatics **33**(14), i37–i48 (2017)
9. Han, X., Kwoh, C.K., Kim, J.: Clustering based active learning for biomedical named entity recognition. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 1253–1260 (2016)
10. Hanisch, D., Fundel-Clemens, K., Mevissen, H.T., Zimmer, R., Fluck, J.: Prominer: rule-based protein and gene entity recognition. BMC Bioinform. **6**, S14 (2005)
11. Hochreiter, S.: The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int. J. Uncertainty Fuzziness Knowl.-Based Syst. **6**, 107–116 (1998)
12. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
13. Lee, J., et al.: BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics **36**, 1234–1240 (2019)
14. Luo, L., et al.: A neural network approach to chemical and gene/protein entity recognition in patents. J. Cheminform. **10**(1), 1–10 (2018). https://doi.org/10.1186/s13321-018-0318-3
15. Mucke, H.: Relating patenting and peer-review publications: an extended perspective on the vascular health and risk management literature. Vasc. Health Risk Manag. **7**, 265–272 (2011)
16. Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
17. Saad, F.: Named entity recognition for biomedical patent text using Bi-LSTM variants. In: The 21st International Conference on Information Integration and Web-based Applications & Services (iiWAS 2019) (2019, to appear)
18. Saad, F., Nürnberger, A.: Overview of prior-art cross-lingual information retrieval approaches. World Patent Inf. **34**, 304–314 (2012)
19. Settles, B.: Biomedical named entity recognition using conditional random fields and rich feature sets. In: Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and Its Applications, JNLPBA 2004, pp. 104–107 (2004)
20. Tanabe, L., Wilbur, W.J.: Tagging gene and protein names in biomedical text. Bioinformatics **18**(8), 1124–1132 (2002)

21. Yadav, V., Bethard, S.: A survey on recent advances in named entity recognition from deep learning models. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 2145–2158, August 2018
22. Yang, Z., Lin, H., Li, Y.: BioPPISVMExtractor: a protein-protein interaction extractor for biomedical literature using SVM and rich feature sets. J. Biomed. Inform. **43**, 88–96 (2009)
23. Zhang, S., Elhadad, N.: Unsupervised biomedical named entity recognition: experiments with clinical and biological texts. J. Biomed. Inform. **46**(6), 1088–1098 (2013)