EPiC Series in Language and Linguistics
Volume 4, 2019, Pages 50–60
Proceedings of Third Workshop "Computational linguistics and language science"

An Experimental Study of Hybrid Machine Learning Models for Extracting Named Entities

Lei Jiang[1] and Elena I. Bolshakova[23]

[1] Lomonosov Moscow State University, Moscow, Russia
kiwee@outlook.com
[2] Lomonosov Moscow State University, Moscow, Russia
eibolshakova@gmail.com
[3] National Research University Higher School of Economics, Moscow, Russia

Abstract

The paper describes two hybrid neural network models for named entity recognition (NER) in texts, as well as results of experiments with them. The first model, namely Bi-LSTM-CRF, is known and used for NER, while the other model named Gated-CNN-CRF is proposed in this work. It combines convolutional neural network (CNN), gated linear units, and conditional random fields (CRF). Both models were tested for NER on three different language datasets, for English, Russian, and Chinese. All resulted scores of precision, recall and F1-measure for both models are close to the state-of-the-art for NER, and for the English dataset CoNLL-2003, Gated-CNN-CRF model achieves 92.66 of F1-measure, outperforming the known result.

1 Introduction

Named Entity Recognition (NER) involves identifying names of people, organizations, locations, or other entities in natural language texts, which is the necessary subtask in many natural language processing applications, such as information retrieval, question answering, extraction of information about events, machine translation, semantic parsing and so on. NER is also required for auxiliary natural language processing tasks, in particular, reference resolution. Besides recognition of names, their classification is usually performed by annotating names with corresponding categorization tags. Below we give examples of NER:

- *"Lucy and Jiang Lei are the students of Moscow State University."* In this sentence, *"Lucy"* and *"Jiang Lei"* are the named entities categorized as ***PERSON****(PER)*, whereas *"Moscow State University"* is a named entity ***ORGANIZATION****(ORG)*.
- In the sentence *"There are many universities in Moscow."*, *"Moscow"* is a named entity ***LOCATION****(LOC)*.

- *"iPhone7 and Samsung Galaxy S7 both are very good phones."*. In this sentence *"iPhone7"* and *"Samsung Galaxy S7"* are mobile phone brands and they may be annotated as ***BRAND***.

There are mainly two approaches for NER task [24, 22], the former is based on linguistics rules (for example [19]), the latter makes use of statistical machine learning methods.

Rule-based NER is almost the only choice we have when tagged corpora for training are absent. One problem of rule-based methods is bad portability, since when we change problem domain of texts, it is necessary to modify developed rules or to add new rules, in order to maintain adequate performance.

In machine-learning NER, various traditional machine learning algorithms were applied and tested, such as Support Vector Machine (SVM) [12], Hidden Markov Model (HMM) [29], Conditional Random Fields (CRF) [18, 27]. For these machine learning methods, the quality of recognition vary, giving approximately 84.04% - 96.6% of F1-measure (combined measure of precision and recall), depending on types of recognized entities and corpora used for learning. CRF method is usually frequent choice for NER, since it gives the best results, because it captures correlations between neighbor words and their tags and so decodes the best chain of tags for a given sentence.

In recent years, neural networks [20, 13] are increasingly applied in the field of machine learning and data analysis because of their strong feature extraction capabilities. In image recognition, individual pixels are difficult to classify, so we use neural networks to extract features and further to classify them. Similarly, we can also use neural networks to extract text features to solve the problem of named entity recognition.

Neural networks have become very popular within a short period of time and have proved their applicability in many natural language processing tasks. Some recent papers [1, 4, 6, 15] propose several neural network models for NER and demonstrate good results. Nevertheless, not all neural network architectures are applicable for NER and not all possible architectures are studied by now.

Among various neural network architectures, Bidirectional Long Short-Term Memory (Bi-LSTM) model as a kind of recurrent neural networks (RNNs) [20], has proved to be successful for processing sequential data (in particular, for texts). At the same time, convolutional neural networks (CNN) [13] are not usually applied for handling text data, they have proved to be successful for processing visual object. Thus, it is reasonable to explore some CNN architectures for NER task. In our work we decide to combine the capabilities of powerful feature extraction provided by neural networks with the capabilities of CRF's decoding. Two particular hybrid architectures: Bi-LSTM-CRF primarily proposed in the work [15] and Gated-CNN-CRF proposed in our work were chosen for experiments in NER task.

The first architecture sequentially includes word-embedding layer, Bi-LSTM layer and CRFs layer. In word-embedding layer, the trained word vector model (Word2Vec) [21] is used to generate the input for the neural network layer. Then Bi-LSTM neural network provides a more abstract contextual information as the node of CRFs model. In the proposed Gated-CNN-CRF model (inspired by the new work of Facebook AI research [8]) CNN is used to get the contextual representations, instead of Bi-LSTM. We also introduce a gating mechanism to filter information, which works like "forget gate" of LSTM.

At present, for English texts, the various machine learning NER models are well investigated, however, for Russian and Chinese, various NER methods are less studied, and for Chinese texts, NER is more difficult because of absence of word boundaries and the ambiguity of words. So we have decided to experimentally study two hybrid neural network architectures: the Bi-LSTM-CRF model and the new Gated-CNN-CRF model, aiming to reveal and compare their quality in NER task for these different languages.

Both models were trained and tested on three language datasets: for English CoNLL-2003 [26], Russian Named-Entities-3 [1], and Chinese SIGHAN2006 [2]. The Gated-CNN-CRF model was trained very efficiently and six times faster than the Bi-LSTM-CRF. Our experiments have showed that precision, recall and F1-measure for both models are close to the state-of-the-art for NER, and on English dataset CoNLL-2003, the proposed Gated-CNN-CRF model outperforms the other machine-learning NER models reported in [17].

2 Related Work

By now, automatic NER task is a relatively well investigated area [22], many various statistical machine learning methods were experimentally studied. Primarily, well known Hidden Markov Model (HMM), Support Vector Machine (SVM), and Conditional Random Fields (CRF) methods were applied to the task.

In the work [29] the author proposes a HMM-based chunk tagger, which effectively integrates and applies four different kinds of named entity features, ranging from internal word information to semantic information from gazetteers and macro context of the document. Evaluation of the system at MUC-6 [3] and MUC-7 [4] conferences for English NER tasks has shown F1-measure of 96.6% and 94.1%, respectively.

In the paper [12] an SVM-based method for NER task was presented, which achieves 90.03% of F1-measure on the data set CoNLL-2000 [25]. The goal of SVM is to find a decision function that accurately predict the correct named entities for given sentences.

A large amount of papers are devoted to investigation of Conditional Random Fields (CRF) for NER task, in particular, [18, 27]. In [18] the author has presented a framework for CoNLL-2003 NER task [26], and with just a few person-weeks of effort the developed method has achieved F1-measure of 84.04% on the test set.

In recent years, recurrent neural networks (RNN) [20] have become very popular in the field of natural language processing (NLP) for various tasks, from text classification to machine translation. RNN can capture long-term dependencies and so can model "understanding" sentences.

For NER task, neural network model was applied in a hybrid architecture combining RNN and CRF [15]. Since CRF method can reveal contextual correlations between neighbor words and their tags, such hybrid architecture combines advantages of two models. Output contextual representations of RNN are used as input to CRF model, instead of the original sentences. This hybrid model obtained 90.94% of F1-measure on the English dataset CoNLL-2003 [26].

For Russian language, the work [1] has studied a neural network model, namely Bi-LSTM-CRF, and has proved that encoding input tokens with external word embeddings allowed to achieve the state-of-the-art scores for the Russian NER task. The model was evaluated across three datasets: Gareev's dataset [7], Person-1000 [5], and FactRuEval-2016 [6], it has achieved excellent results: F1-measure of 87.17%, 99.26% and 82.10%, respectively.

For Chinese language, the work [6] has investigated radical-level representations in Bi-LSTM-CRF architecture and has achieved F1-measure of 90.95% on the dataset SIGHAN2006

[1] http://labinform.ru/pub/named_entities/descrne.htm
[2] http://sighan.cs.uchicago.edu/bakeoff2006/registration.html
[3] https://cs.nyu.edu/faculty/grishman/muc6.html
[4] http://www-nlpir.nist.gov/related_projects/muc/proceedings/ne_task.html
[5] http://ai-center.botik.ru/Airec/index.php/ru/collections/28-persons-1000
[6] https://github.com/dialogue-evaluation/factRuEval-2016

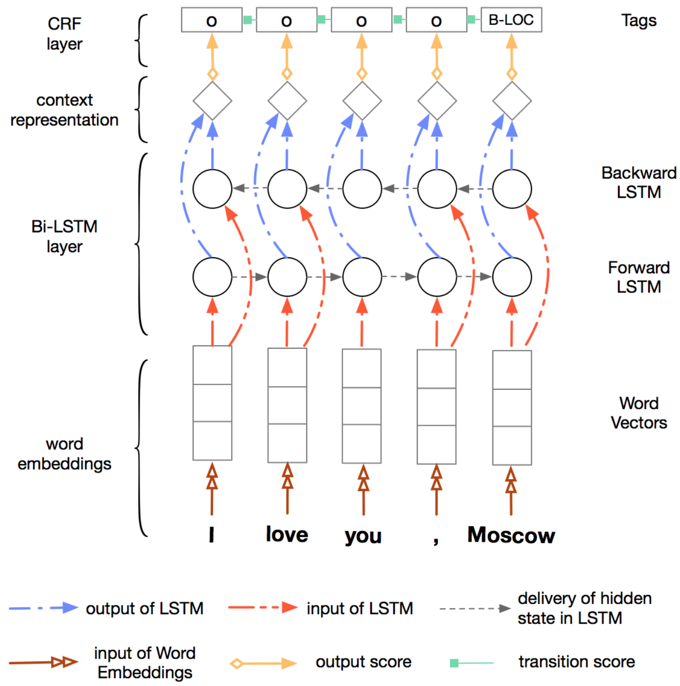

Figure 1: Hybrid Bi-LSTM-CRF Model for NER

NER task for Chinese [7].

3 Hybrid Neural Network Models

In this section, we provide descriptions of two hybrid architectures for NER task: Bi-LSTM-CRF model and Gated-CNN-CRF model, the latter is proposed in our work. Although they have different architectures, their neural network parts are to extract contextual representations. Since the excellent performance of CRF in NER task, we use CRF layer on the top of both architectures to identify named entities.

3.1 Bi-LSTM-CRF Model

The hybrid Bi-LSTM-CRF model is a combination of Bidirectional Long Short-Term Memory [11] and Conditional Random Fields. In Figure 1 we show the whole procedure for NER task by using the model.

In the first layer, we use our trained word embeddings to represent the words in the sentences. The word vectors are given to the Bi-LSTM layer to get the contextual representations, which are the input of CRF layer. CRF can be trained not only to obtain the correspondence between the contextual representations and named entities, but also to capture the internal correlations between neighbor named entities.

[7]http://sighan.cs.uchicago.edu/bakeoff2006/registration.html

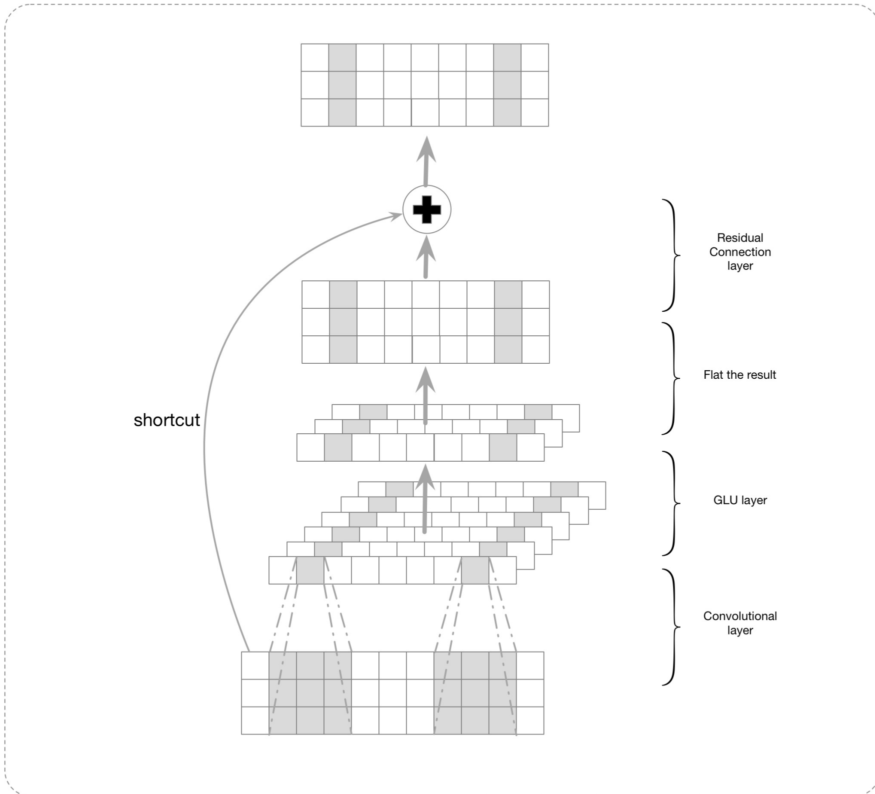

Figure 2: Gated-CNN block

3.2 Gated-CNN-CRF Model

Convolutional neural networks (CNN) have proven to be successful for handling computer vision problems, such as image recognition [13], while recursive neural networks are good for time series tasks. However, some works [28, 8] show that we can also use CNN for text processing tasks.

Unlike the Bi-LSTM-CRF model, in Gated-CNN-CRF model, we use convolutional neural networks to get the contextual representations.

As shown in Figure 2, we construct a Gated-CNN block by stacking the convolutional layer, gated liner units (GLU) [5], and residual connection layer[10]. We use GLU layer to filter the result of the convolutional layer and then flat the result. Then we sum the input and the result to get the contextual representations (residual connection layer). We stack Gated-CNN blocks layer by layer to capture the long-term dependencies like LSTM. It should be noted, that Gated-CNN's ability to capture dependencies between adjacent words is even stronger than RNN and it can be trained faster, because the convolution operations are performed in parallel in GPU.

Figure 3 shows the whole Gated-CNN-CRF architecture for NER task. In this hybrid neural network, we use stacked Gated-CNN blocks to get the contextual representations from the embedding layer. For each Gated-CNN block, we ensure that the convolutional layer matches its input length by padding the input of each layer. So that we can get a contextual representation matrix with the the same "width" and "height" as the embedding matrix. Like in the Bi-

LSTM-CRF model described above, we stack a CRF layer at the top of Gated-CNN blocks to construct the hybrid Gated-CNN-CRF model.

In the first embedding layer, we use the sum of word embeddings and position embeddings to represent the words in the sentence. These new embeddings (word-embeddings + position-embeddings) will be input into stacked Gated-CNN blocks to get the contextual representations. After feeding these representations into a CRF layer, we get a prediction of named entities for the input text.

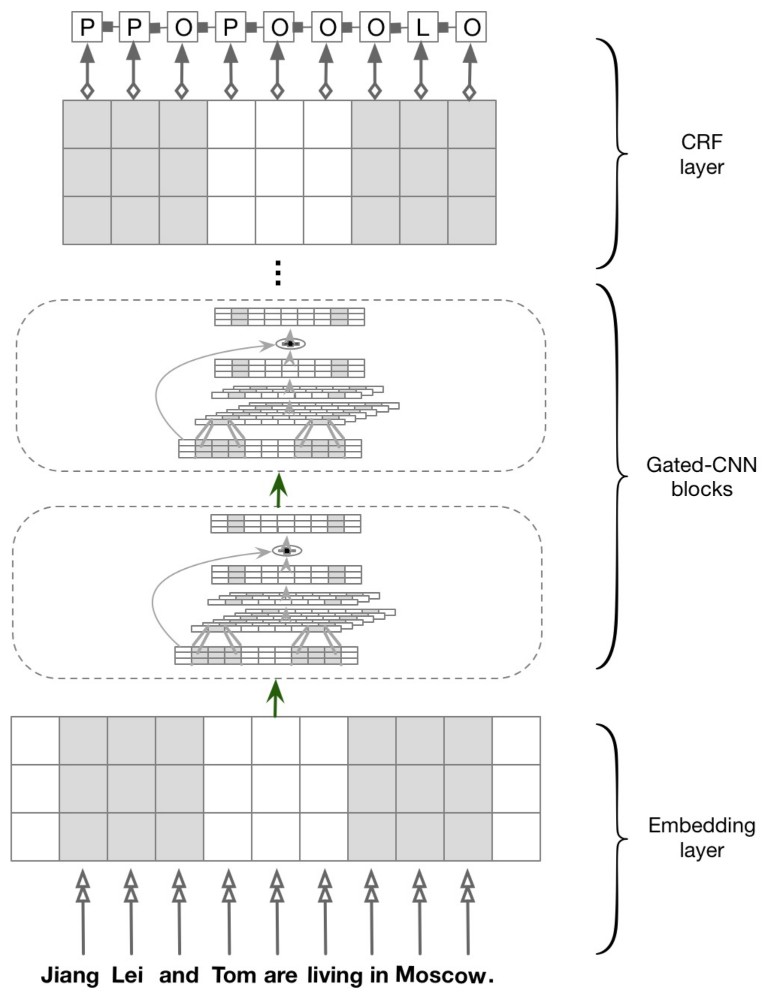

Figure 3: Hybrid Gated-CNN-CRF Architecture

4 Experiments with the Hybrid Models

4.1 Data Sets

To test and evaluate the described hybrid models for various languages, we use three corpora: SIGHAN2006 NER task for Chinese [8], Named-Entities-3 for Russian [9], and CoNLL-2003 for English [26]. Our aim is to test and compare the performance of the models on various corpora, so in experiments we consider the most popular types of named entities: locations (LOC), persons (PER), organizations (ORG).

4.1.1 Chinese Text Corpus

We train and test our models on Microsoft Research (MSR) dataset of The Third SIGHAN Chinese Language Processing Bakeoff named entity recognition task. This dataset is divided into three parts: training data, validation data and testing data. It contains three types of named entities: locations (LOC), persons (PER), organizations (ORG), the number of these entities is: 16571, 8144, 9277, respectively.

4.1.2 Russian Text Corpus

For Russian named entity recognition, we use the Named-Entities-3 corpus, which is created for the estimation of NER quality for texts in Russian. This dataset is an extension of tagged data Persons-1000, prepared by the Artificial Intelligence Research Center of Russian Academy of Sciences. Three types of named entities are included in Named-Entities-3 corpus: LOC, PER, ORG, the corresponding number of them is 3141 for LOC, 10623 for PER, 7032 for ORG.

4.1.3 English Text Corpus

There are many datasets for English NER task and we choose CoNLL-2003 [26]. This dataset includes four types of named entities: LOC, PER, ORG, and miscellaneous names (MISC, but we ignore them in our experiments). The number of entities for each type is: 7140 for LOC , 6600 for PER, and 6231 for ORG.

4.2 Tagging Schemes for Machine Learning

Named entity could span several tokens within a sentence. For example, the named entity "Moscow State University" spans three tokens. There are several tagging schemes for machine learning. Sentences are usually represented in the **BIO** [22] format (Beginning, Inside, Outside) where a particular token is labeled as **B-label**, if it is at the beginning of the named entity, whereas **I-label** corresponds to a token within the named entity, that is not the first one. **O-label** is used for tokens that are not named entities.

Sometimes, **BIOES** format is used, which encodes information about singleton entities (**S-label**) and explicitly marks the end of named entities (**E-label**). Using this scheme, tagging a word as **I-label** with high-confidence narrows down the choices for the subsequent word to **I-label** or **E-label**.

For example, we have a sentence "*Tom and Jiang Lei are in New York*". It is tagged as "*B-PER O B-PER I-PER O O B-LOC I-LOC*" in **BIO** format, while "*S-PER O B-PER E-PER O O B-LOC E-LOC*" in **BIOES** format. In our work, we use the **BIOES** tagging scheme.

[8] http://sighan.cs.uchicago.edu/bakeoff2006/registration.html
[9] http://labinform.ru/pub/named_entities/descrne.htm

	Bi-LSTM-CRF			Gated-CNN-CRF		
	Precision	Recall	F_1	Precision	Recall	F_1
LOC	92.80	91.90	92.35	92.33	92.18	92.25
PER	93.28	92.30	92.79	93.96	92.59	93.27
ORG	90.27	83.99	87.02	89.39	84.39	86.82
Micro	**92.26**	89.86	91.05	91.95	**90.18**	**91.06**

Table 1: Evaluation scores of Bi-LSTM-CRF and Gated-CNN-CRF for Chinese

	Bi-LSTM-CRF			Gated-CNN-CRF		
	Precision	Recall	F_1	Precision	Recall	F_1
LOC	92.79	91.79	92.29	92.85	92.20	92.52
PER	93.30	92.59	92.95	93.40	92.43	92.91
ORG	88.36	84.54	86.41	89.27	83.78	86.44
Micro	**90.03**	91.75	90.88	89.98	**92.05**	**91.01**

Table 2: Evaluation scores of Bi-LSTM-CRF and Gated-CNN-CRF for Russian

4.3 Pre-trained Word Vectors

For the word embedding layer in our models we use three pre-trained word vectors for Chinese, Russian and English respectively. For English NER task, we use GloVe [23] word representations computed by the GloVe algorithm. For Chinese and Russian NER task, we use fastText [2] word representations, which are published by Facebook [10]. All words are represented as vectors in 300 dimension space.

4.4 Training the Hybrid Models

We have implemented both the hybrid models by using PyTorch framework [11]. The computations for our models are run on a GeForce GTX 1080 Ti GPU. For training both hybrid models, we used stochastic gradient descent (SGD) [3] with a learning rate of 0.01 and momentum 0.9 to optimize the parameters within models.

The parameters in our hybrid models are trained to maximize the log-probability [9] of observed sequences of NER tags in an annotated corpus. In the forward propagation process, we get contextual representations from Bi-LSTM or Gated-CNN layer, then we use Viterbe decode [14] to find the optimal tagging sequence.

4.5 Evaluation and Results

To evaluate the quality of NER for the described hybrid models, we compute precision, recall and F1-measure.

Tables 1, 2, 3 show the results of Named Entity Recognition on three different language corpora with the Bi-LSTM-CRF and Gated-CNN-CRF models. One can notice that organization names are always the most difficult to recognize in all languages. For both models, the results on three datasets are close to the state-of-the-art for NER. In particular, for Gated-CNN-CRF model micro F1-measure is 91.06% for Chinese, 91.01% for Russian, and 92.66% for English dataset respectively.

[10] https://fasttext.cc/docs/en/pretrained-vectors.html
[11] http://pytorch.org

	Bi-LSTM-CRF			Gated-CNN-CRF		
	Precision	Recall	F_1	Precision	Recall	F_1
LOC	93.98	93.47	93.72	95.88	95.10	95.49
PER	94.79	96.74	95.75	94.91	97.18	96.03
ORG	80.54	88.59	84.38	87.21	88.96	88.08
MISC	89.89	78.98	84.08	90.38	83.42	86.76
Micro	90.37	91.13	90.75	**92.78**	**92.55**	**92.66**

Table 3: Evaluation scores of Bi-LSTM-CRF and Gated-CNN-CRF for English

Model	F_1
1) CRF [18]	84.04
2) K-Means Clustering [16]	90.90
3) Joint Entity Recognition and Linking [17]	91.2
4) LSTM-CNN model [4]	90.77
5) LSTM-CRF [15]	90.94
6) Bi-LSTM-CRF	90.75
7) Gated-CNN-CRF	**92.66**

Table 4: F1-measure of various models on CoNLL-2003 data

Comparing two models, Gated-CNN-CRF model gives the better results of micro F1-measure than the Bi-LSTM-CRF model, slightly better for Chinese texts, minor for Russian and more essential for English texts. The improvement is achieved mainly due to the recall for tags LOC and PER.

Table 4 presents the comparisons with other NER machine learning models described in the known works and trained on the same English corpus CoNLL-2003. Our Gated-CNN-CRF model outperforms all the systems, achieving F1-measure of 92.66%.

Thus, experimental results show that CNN can also capture the dependencies within sentences. As for the Bi-LSTM-CRF model, it achieves F1-measure of 90.75% and outperforms only the CRF model [18]. We can achieve higher scores by stacking Bi-LSTM layers, but that also means we need to spend more time for training.

5 Conclusions and Future Work

The results of our experimental evaluation of the proposed hybrid neural network Gated-CNN-CRF model proposed for NER task are described in the paper. For comparison, the experiments were also performed for Bi-LSTM-CRF model, and it was shown that the quality of these models measured in precision, recall and F-measure is close to the state-of-the-art scores on three datasets for different languages. At the same time, F1-measure of our Gated-CNN-CRF model is better than the Bi-LSTM-CRF on all the datasets. Moreover, on the English dataset CoNLL-2003, Gated-CNN-CRF model achieves 92.66% of F1-measure and overperforms the known results for this dataset.

Thus, Gated-CNN-CRF model is proved to be applicable to NER task, and not only RNN architecture, but also CNN can extract dependencies between tags in texts. In our work we have proposed Gated-CNN model to extract contextual representations for better performance in NER task.

On the way to improve NER results, we plan further investigations:

- to adjust hyperparameters of the hybrid models;
- to experiment with various word vector spaces;
- to add character-level layer into our models for Russian and English (Chinese NER can not benefit of this level).

References

[1] Mikhail Y Arkhipov, Mikhail S Burtsev, et al. Application of a hybrid bi-lstm-crf model to the task of russian named entity recognition. In *Conference on Artificial Intelligence and Natural Language*, pages 91–103. Springer, 2017.

[2] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*, 2016.

[3] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[4] Jason PC Chiu and Eric Nichols. Named entity recognition with bidirectional lstm-cnns. *arXiv preprint arXiv:1511.08308*, 2015.

[5] Yann N Dauphin, Angela Fan, Michael Auli, and David Grangier. Language modeling with gated convolutional networks. *arXiv preprint arXiv:1612.08083*, 2016.

[6] Chuanhai Dong, Jiajun Zhang, Chengqing Zong, Masanori Hattori, and Hui Di. Character-based lstm-crf with radical-level features for chinese named entity recognition. In *Natural Language Understanding and Intelligent Applications*, pages 239–250. Springer, 2016.

[7] Rinat Gareev, Maksim Tkachenko, Valery Solovyev, Andrey Simanovsky, and Vladimir Ivanov. Introducing baselines for russian named entity recognition. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, pages 329–342, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[8] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. Convolutional sequence to sequence learning. *arXiv preprint arXiv:1705.03122*, 2017.

[9] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, pages 6645–6649. IEEE, 2013.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[12] Hideki Isozaki and Hideto Kazawa. Efficient support vector classifiers for named entity recognition. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, pages 1–7. Association for Computational Linguistics, 2002.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[14] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[15] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*, 2016.

[16] Dekang Lin and Xiaoyun Wu. Phrase clustering for discriminative learning. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 1030–1038. Association for Computational Linguistics, 2009.

[17] Gang Luo, Xiaojiang Huang, Chin-Yew Lin, and Zaiqing Nie. Joint named entity recognition and disambiguation. In *Proc. EMNLP*, pages 879–880, 2015.

[18] Andrew McCallum and Wei Li. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, pages 188–191. Association for Computational Linguistics, 2003.

[19] Slim Mesfar. Named entity recognition for arabic using syntactic grammars. In *NLDB*, pages 305–316. Springer, 2007.

[20] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Interspeech*, volume 2, page 3, 2010.

[21] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[22] David Nadeau and Satoshi Sekine. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26, 2007.

[23] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[24] AS Starostin, VV Bocharov, SV Alexeeva, AA Bodrova, AS Chuchunkov, SS Dzhumaev, IV Efimenko, DV Granovsky, VF Khoroshevsky, IV Krylova, et al. Factrueval 2016: evaluation of named entity recognition and fact extraction systems for russian. 2016.

[25] Erik F Tjong Kim Sang and Sabine Buchholz. Introduction to the conll-2000 shared task: Chunking. In *Proceedings of the 2nd workshop on Learning language in logic and the 4th conference on Computational natural language learning-Volume 7*, pages 127–132. Association for Computational Linguistics, 2000.

[26] Erik F Tjong Kim Sang and Fien De Meulder. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, pages 142–147. Association for Computational Linguistics, 2003.

[27] Natalia Loukachevitch Valeria Mozharova. Combining knowledge and crf-based approach to named entity recognition in russian. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 185–195. Springer, 2016.

[28] Tao Wang, David J Wu, Adam Coates, and Andrew Y Ng. End-to-end text recognition with convolutional neural networks. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3304–3308. IEEE, 2012.

[29] GuoDong Zhou and Jian Su. Named entity recognition using an hmm-based chunk tagger. In *proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 473–480. Association for Computational Linguistics, 2002.