

DUTIR in BioNLP-ST 2016: Utilizing Convolutional Network and Distributed Representation to Extract Complicate Relations

Honglei Li, Jianhai Zhang, Jian Wang, Hongfei Lin, Zhihao Yang

School of Computer Science and Technology

Dalian University of Technology

116024 Dalian, China

201081023@mail.dlut.edu.cn

jianhai0527@mail.dlut.edu.cn

wangjian@dlut.edu.cn

hflin@dlut.edu.cn

yangzh@dlut.edu.cn

Abstract

We participate in the two event extraction tasks of BioNLP 2016 Shared Task: binary relation extraction of SeeDev task and localization relations extraction of Bacteria Biotope task. Convolutional neural network (CNN) is employed to model the sentences by convolution and max-pooling operation from raw input with word embedding. Then, full connected neural network is used to learn senior and significant features automatically. The proposed model mainly contains two modules: distributive semantic representation building, such as word embedding, POS embedding, distance embedding and entity type embedding, and CNN model training. The results with F-score of 0.370 and 0.478 in our participant tasks, which were evaluated on the test data set, show that our proposed method contributes to binary relation extraction effectively and can reduce the impact of artificial feature engineering through automatically feature learning.

1 Introduction

Information extraction devotes to finding useful data and hidden knowledge for researchers from amounts of texts. With the demands of rapidly and accurately locating key issues about life and biology increasing, bio-IE appears timely and has attracted more and more researchers to address this question (Krallinger et al., 2005; Zweigenbaum et al., 2007). Much progress has been made in named entity identification, protein-protein relations classification (Blaschke et al., 1999) and drug-drug interaction extraction (Rodrigues et al., 2008). Furthermore, fine-

grained information extraction in biology, in particular event extraction has entered the spotlight of people and, appeared many meaningful and challenge tasks for event extraction, which can gather the community-wide efforts and contribute to the development of biology information extraction (Kim et al., 2009; Kim et al., 2011; Nédellec et al., 2013).

The BioNLP Shared Task series (Kim et al., 2009; Kim et al., 2011; Nédellec et al., 2013) is a representative for biomolecular event extraction, which has been held four times including this year. The topics of the series range from fine-grained extraction, generalization to knowledge base construction. In addition, the scope that this task involved has become much broader at each edition. For example, BioNLP-ST 2013 (Nédellec et al., 2013) covers many new hot topics compared to the previous editions, such as Cancer Genetics, Pathway Curation and Gene Regulation Network in Bacteria.

BioNLP-ST 2016 further broadens the scope of the text-mining application domains in biology by introducing a new issue on seed development, named the issue as the SeeDev task. The development of the seed is a critical issue in agriculture and presents an opportunity for the community to contribute the common efforts in bio-IE. The other task, Bacteria Biotope of the BioNLP-St'13 expands on the previous editions by replacing the Web Pages with scientist papers abstracts to organize the corpus, which is much closer to the actual needs of detailed and scientific information for biologists. The third task focuses on the Genia corpus as previous edition, but gives more emphasis in the contribution from any aspect of knowledge extraction, which is an open question to participants.

We focus on the two events extraction sub-tasks of BioNLP 2016 Shared Task: binary rela-

tion extraction of SeeDev task and localization relations extraction of Bacteria Biotope task. Both tasks broaden the scope of fine-grained information extraction in biology, and contribute to the development of the actual application in text mining.

The SeeDev task has not been introduced in the previous BioNLP-ST and aims at exploring the knowledge of the molecular network underlying the regulation of seed development. The SeeDev task is similar to the GRN (Gene Regulation Network in Bacteria) task in BioNLP'13, aiming at extracting a regulation network that links and integrates a variety of molecular (Bossy et al., 2013) or processes interactions between entities. Therefore, the superior systems from the GRN can give us some useful heuristics. Five systems participated in GRN and all systems applied machine learning algorithms with many different resources of information and pre-processing in BioNLP'13. Lots of features, such as linguistic features, semantic and syntactic information between two entities, were added into these systems. However, they implemented different ML algorithms, including SVM, CRF and KNN (Bossy et al., 2013). For example, Provost (2013) employed a basic Support Vector Machine framework and focused more on the domain of feature definition and exploration. They achieved an F-score of 0.313, standing on second place in GRN task of BioNLP'13. IRISA system (Claveau, 2013) emphasized the similarity between the known instances and the closest known examples based on K-Nearest Neighbor algorithm.

Bacteria Biotope task in the BioNLP-ST 2016, our second participation, was the third edition that focuses on extracting localization relations between bacteria and their habitats from scientific papers abstracts. Many systems had contributed their efforts to the task in the previous editions. Boun system (Karadeniz et al., 2013) used the shallow linguistic knowledge of the corpus to implement the prediction based on previously defined syntactic rules and discourse-based rules, coming the F-score of 0.27. The Alvis system (Ratkovic et al., 2011) also employed hand-designed patterns to detect the relations between bacteria and habitat with the linguistic and lexical knowledge. UTurku and JAIST (Karadeniz et al., 2013) systems in BioNLP'11 explored different approaches from the above mentioned and regarded the binary event extraction as a classification problem, thus applying machine learning methods. In BioNLP'13, TEES-

2.1 and IRISA (Bossy et al., 2013) also employed the same idea to this question, and achieved the state-of-the-art results with F-score of 0.42 and 0.40, respectively, which were much higher than the two hand-designed rules methods: LIMSI and Boun.

Most of systems delivered their good ideas and achieved the better results for these tasks in BioNLP-ST, which have positively promoted the development of biology information extraction. So, it is an opportunity for researchers to apply various approaches and new ideas to these tasks. Over recent years, the landscape of Convolutional Neural Network (CNN) has been obviously prosperous and pushed forward through the expansion of actual application of various fields. The introduction of convolution layers and pooling layers in CNN has helped to improve the performance of features automatically learnt in networks. Therefore, in our work, we explore the CNN to learn features automatically for the two binary relations extraction tasks, significantly differed from previous systems in BioNLP-ST.

2 Method

The tasks of SeeDev-binary and BB-event both can be treated as binary relation extraction which specifies whether there is interaction between two entities. In relation extraction, the semantic and syntactic information for sentence act as a significant role. Traditional method usually need to design and extract complex features from sentence based on domain-specific knowledge, such as tree kernel and graph kernel, to model the sentences. As a result, this will lead to much lower ability of generation for corpus dependent. Consequently, instead of complicate hand-designed feature engineering, we employ convolutional neural network, also called CNN, to model the sentences by convolution and max-pooling operation from raw input with word embedding and full connected neural network to learn senior features automatically. Furthermore, we employ POS embedding to enrich the semantic information of words, distance embedding to capture the information of relative distance between the entities and entity type embedding as the supplement features of the sentence. All the feature embedding is combined to build final distributive semantic representation which is fed to convolutional neural network.

As described in Fig.1, the proposed model mainly contains two modules: distributive se-

mantic representation building, such as word embedding, POS embedding, distance embedding and entity type embedding, and CNN model training. In the next parts, we will introduce more details.

2.1 Build Distributive Semantic Representation

Traditional one-hot representation, which is employed by mostly machine learning methods, can vectorize the text and plays an important role. However, it can result in the problems of semantic gap and dimension disaster which restrict its application. Consequently, in our proposed

method, we employ distributive semantic representation, proposed by Hinton (1986) at first, as the feature representation of the model. And then, we exploit the advantage of convolutional neural network at modeling the sentences to learn sentence-level representation from raw input. The distributive semantic representation is built as follows. For simply definition, we assume $S = E_1 W_1 W_2 W_3 \dots W_n E_2$ as the word sequence between two entities in one sentence, where E_1, E_2 stand for the entities and $W_1 \dots W_n$ stand for the words between two entities.

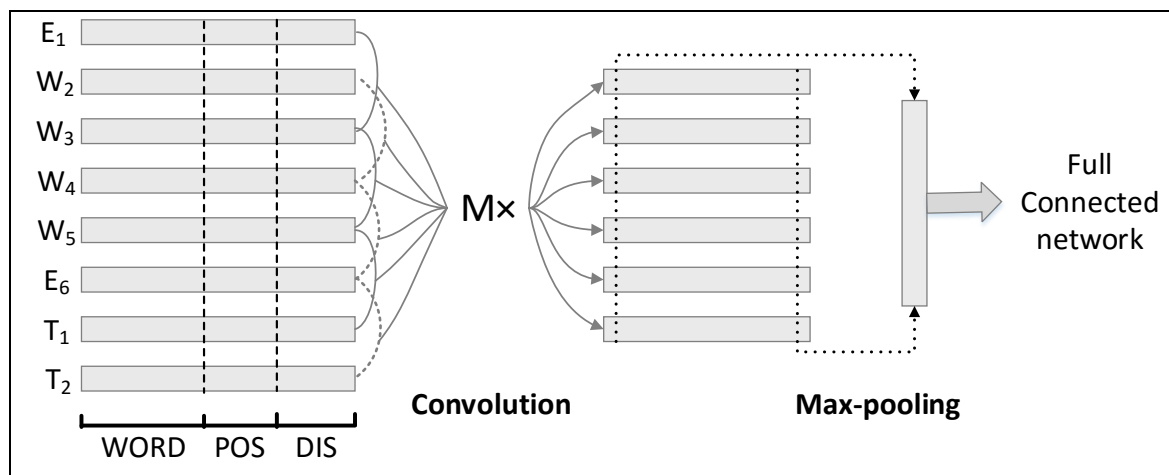


Figure 1: The model of convolutional neural network with distributive vector

2.1.1 Word Embedding

Instead of traditional one-hot representation, we utilize the distributive semantic representation of words for solving the problem of dimension disaster and semantic gap. Firstly, we employ word2vec tool, which can effectively learn distributive representation of words from massive and unlabeled data, to train word embedding from massive available Pubmed abstracts. The embedding with low dimension and realistic value contains rich semantic information and can be treated as feature representation of words instead of one-hot.

Inspired by language model, we employ the contexts of two entities to predict the relation type. In our experiments, the contexts are expressed by the words between two entities in one sentence. Then, the word sequence is transformed into word embedding matrix by looking up the word embedding table. The word embedding matrix can be treated as local feature of the sentence and fed to CNN model to learn global

feature which can contribute to the relation identification. The word embedding matrix is represented as follows:

$$LT_W(S) = [\langle W \rangle_{E_1}, \langle W \rangle_{W_1}, \langle W \rangle_{W_2}, \dots, \langle W \rangle_{W_n}, \langle W \rangle_{E_2}]$$

Where $W \in \mathbb{R}^{|\mathcal{D}| \times dim}$ ($|\mathcal{D}|$ is the size of dictionary and dim is the dimension of word embedding) is the word embedding table trained by word2vec with Pubmed abstracts and fine-tuned while training.

2.1.2 Entity Type Embedding

Through analyzing the dataset, we observe that different entities with different types have different probability to interact with each other if the entity type satisfies the relation constraints. Consequently, entity type of two entities is an import factor for predicting the relation type. In our model, entity types are treated as the extra features of the relation and the supplement of word

sequence. $\langle W^T \rangle_{type(E_1)}, \langle W^T \rangle_{type(E_2)}$ are added as the extra features of the relation:

$$LT_{W,W^T}(S) = [\langle W \rangle_{E_1}, \langle W \rangle_{W_1}, \dots, \langle W \rangle_{E_2}, \langle W^T \rangle_{type(E_1)}, \langle W^T \rangle_{type(E_2)}]$$

Where $W^T \in \mathbb{R}^{|\mathcal{D}_T| \times dim}$ is type embedding which is randomly initialized by random sampling from the uniform distribution $([-0.25, 0.25])$. $type(\cdot)$ stands for the entity type. \mathcal{D}_T is the dictionary of entity types.

2.1.3 POS Embedding

Word semantics usually have several aspects containing similarity, POS (part-of-speech) and so on. For enriching the semantic representation of each word, POS embedding is introduced as the supplement of word embedding:

$$LT_{W^p}(S) = [\langle W^p \rangle_{p(E_1)}, \langle W^p \rangle_{p(W_1)}, \dots, \langle W^p \rangle_{p(W_n)}, \langle W^p \rangle_{p(E_2)}, \mathbb{O}, \mathbb{O}]$$

We denote $W^p \in \mathbb{R}^{|\mathcal{D}_p| \times dim_p}$ as the POS embedding which is randomly initialized as well as type embedding, where \mathcal{D}_p is the size of POS dictionary and, dim_p , a hyper-parameter, is the dimension of POS embedding. We set $dim_p = 5$ through trying different configuration. Zero vector (\mathbb{O}) is used to pad the sentence.

2.1.4 Distance Embedding

In relation classification tasks, distance information usually plays an important role. Distance can capture the relative position between two entities. As shown in followed formulas, $LT_{W^d}(S)_1$ stands for the relative distance between words and the first entity, and $LT_{W^d}(S)_2$ stands for the relative distance between words and the second entity.

$$LT_{W^d}(S)_1 = [\langle W^d \rangle_{d(E_1, E_1)}, \dots, \langle W^d \rangle_{d(W_n, E_1)}, \langle W^d \rangle_{d(E_2, E_1)}, \mathbb{O}, \mathbb{O}]$$

$$LT_{W^d}(S)_2 = [\langle W^d \rangle_{d(E_1, E_2)}, \dots, \langle W^d \rangle_{d(W_n, E_2)}, \langle W^d \rangle_{d(E_2, E_2)}, \mathbb{O}, \mathbb{O}]$$

Where $W^d \in \mathbb{R}^{|\mathcal{D}_d| \times dim_d}$ stands for the distance embedding and $|\mathcal{D}_d|$ is the number of different distances. The embedding is randomly initialized and fine-tuned while training. We set $dim_d = 5$ through trying different configuration. Zero vector (\mathbb{O}) is also used to pad the sentence.

As shown in followed formula, the final distributive semantic representation is acquired by joining the word embedding, type embedding, POS embedding and distance embedding.

$$\varphi(S) = \begin{bmatrix} LT_{W,W^T}(S) \\ LT_{W^p}(S) \\ LT_{W^d}(S)_1 \\ LT_{W^d}(S)_2 \end{bmatrix}$$

2.2 Model Training and Parameters Tuning

After building the distributive semantic representation of relation, we employ convolution and max-pooling to learn the global feature representation from raw input. The detailed computation procedure is described as follows.

$$\langle f \rangle_t = f(W \cdot \varphi(S) + b)$$

$$\langle h \rangle = \max_t \langle f \rangle_t$$

Where W is the convolution filter, it extracts local features from given window of word sequence. $\langle h \rangle$ can be treated as the global feature representation learned from raw distributive representation $\varphi(S)$ and be fed to the full connection layer to learn hidden and senior features.

As we all know, convolutional neural network is a model with vast computation cost. Consequently, we implement the CNN model with theano (Bergstra et al., 2010; Bastien et al., 2012) and run in GPU kernels for accelerating the training procedure. As a result, it takes about half hour to train a CNN model. Meanwhile, we make some modifications in our model for achieving more significant experiment results. In the convolutional layer, we make use of multiple convolution kernels with different window size for capturing sentence features from different views. In the full connection layer, we modified the network with dropout (Srivastava et al., 2014) which is a much simple and efficient method to prevent the problem of overfitting. The dropout network can prevent the co-adaption between the nodes through randomly dropping some nodes or make them not work. Learning rate is the most important hyper-parameter in deep learning. Consequently, we employ Adadelta (Zeiler, 2012) an adaptive learning rate method, to automatically adapt the learning rate instead of configuring it manually. Finally, we empirically search for the reasonable combination of all the hyper-parameters and tune in development dataset. The

optimal parameters of CNN model are described in Table 1.

hyper-parameter	value
Word embedding	50
filter	1800
window	[3,5,7]
layer	3
dropout	0.3
batch	128

Table 1: The parameters of CNN model

3 Results and Discussions

This section presents our results on the SeeDev-binary and BB-event tasks respectively.

3.1 The results of SeeDev-binary task

The SeeDev-binary task datasets contains three parts, namely the training set, development set and test set respectively, which are totally 87 segments from 20 full articles on seed development of *Arabidopsis thaliana*. The task defines 17 different types of entities and 22 different types of binary relations. Table 2 shows the detailed distribution of data.

	#	Train	Dev	Test
Segments	87	39	19	29
Entities	7082	3259	1607	2216
Binary relations	3575	1628	819	1128

Table 2: Detailed statistics of SeeDev-binary task corpus

We aim at extracting the relation between the two target entities and reducing the participa-

tion of hand-designed feature engineering by using our proposed model. Table 3 lists the results of our method on the development and test datasets for SeeDev-binary task. The first two lines are the systems with the two best F-score in official results in BioNLP-ST 2016.

Our method achieved the F-score of 0.368 and 0.370 on the development set and test set, respectively. Compared to the official results from different systems, we stood the similar place with the second best system UniMelb which achieved the F-score of 0.364. It demonstrates that our proposed method has a good performance on binary relations extraction.

In previous methods to binary relations classification, more systems prefer to rules-based or feature engineering methods. However, we employ a different idea, which utilizes the advantages of distributive semantic representation and the CNN model. From the detailed results in Table 4, we can find that the proposed model is of benefit to SeeDev binary task. Moreover, the better recall than precision is achieved on the test datasets. In Table 4, four relations, such as “Occurs_In_Genotype”, and “Regulates_Molecule_Activity”, are not identified by the system, which may be a reason that the size of these relations in corpus is very small.

Methods	Recall	Precision	F-score
LitWay	0.448	0.417	0.432
UniMelb	0.386	0.345	0.364
Our method (on dev set)	0.396	0.344	0.368
Our method (on test set)	0.417	0.333	0.370

Table 3: Results of our method on the development and test data sets for SeeDev-binary task

Binary relation type		Dev data set	Test data set
		R/P/F-score	R/P/F-score
When and Where	Exists_In_Genotype	0.506/0.273/0.355	0.520/0.361/0.426
	Occurs_In_Genotype	0.000/0.000/0.000	0.000/0.000/0.000
	Exists_At_Stage	0.125/0.100/0.111	0.100/0.045/0.063
	Occurs_During	0.200/0.333/0.250	0.083/0.143/0.105
Function	Is_Localized_In	0.426/0.253/0.318	0.290/0.231/0.257
	Is_Involved_In_Process	0.000/0.000/0.000	0.000/0.000/0.000
	Transcribes_Or_Translates_To	0.154/0.286/0.200	0.313/0.208/0.250
Regulation	Is_Functionally_Equivalent_To	0.575/0.821/0.677	0.636/0.745/0.686
	Regulates_Accumulation	0.103/0.231/0.143	0.125/0.100/0.111
	Regulates_Development_Phase	0.119/0.206/0.151	0.221/0.218/0.219
	Regulates_Expression	0.451/0.485/0.467	0.370/0.307/0.336
	Regulates_Molecule_Activity	0.000/0.000/0.000	0.000/0.000/0.000
	Regulates_Process	0.693/0.363/0.476	0.613/0.357/0.451

	Regulates_Tissue_Development	0.000/0.000/0.000	0.000/0.000/0.000
Composition and Membership	Composes_Primary_Structure	0.200/0.500/0.286	0.563/0.750/0.643
	Composes_Protein_Complex	0.000/0.000/0.000	0.667/0.067/0.121
	Is_Protein_Domain_Of	0.172/0.278/0.213	0.129/0.400/0.195
	Is_Member_Of_Family	0.364/0.308/0.333	0.547/0.338/0.418
	Has_Sequence_Identical_To	0.613/0.905/0.731	0.730/0.852/0.786
Interaction	Interacts_With	0.281/0.500/0.360	0.019/0.500/0.036
	Binds_To	0.208/0.227/0.217	0.188/0.240/0.211
Other	Is_Linked_To	0.087/0.133/0.105	0.350/0.350/0.350
	=[ALL RELATIONS]=	0.396/0.344/0.368	0.417/0.333/0.370

Table 4: Detailed results of our method on the development and test data sets for SeeDev-binary task

3.2 The results of BB-event task

For localization relations extraction of Bacteria Biotope task, we also use our proposed system to evaluate the performance. Table 5 shows the results on the development and test datasets. The F-score of 0.478 on test dataset suggest that the proposed method has positive effects on identifying the binary relation. However, the recall on the test dataset is lower than the precision, which may be overfitting on training data. The F-score of 0.499 on the development data set achieve better performance than that on test data set.

The prediction of location relations remains many challenges. First, high diversity of bacteria and locations increases the difficult of the correct pairing. Second, cross-sentences relations caused by coreferences usually are ignored by most system due to complexity and difficulties. In our system, we only considered the relations in one sentence, which many relations in cross sentences were ignored and might cause some reduce on the performance.

Data set	Recall	Precision	F-score
Dev	0.561	0.449	0.499
Test	0.397	0.600	0.478

Table 5: Results of our method on the development and test data sets for BB-event task

From above analysis, the cross-sentences relations extraction is a big challenge, due to much coreferences relations and increasing negative examples. We conduct another experiment to extract relations at the documental level, but not considering the coreferences resolution. Table 6 shows the evaluated results of our method on the development set and test sets at the documental level and sentence level.

At the documental level, the F-score has an about 2% increase on development dataset, while the F-score increases by 6% on test dataset. It

may be because the distribution of relations on two datasets has large different, which there are more cross-sentence relations on test dataset than development dataset. Furthermore, Table 7 shows the statistics of positive and negative examples on training data and development data at the two levels. (It is not nearly possible to have relations between two candidate entities if their distance is too large. Therefore, we remove the candidate examples if the distance is larger than 60.) We can find that, the ratio between positive and negative examples at the documental level is significantly higher than that at the sentence level. The imbalance between positive and negative examples can significantly influence the performance of models. Therefore, we should devote more techniques and good designs to cross-sentences relation extraction.

Models	Recall	Precision	F-score
CNN-Doc (on dev set)	0.552	0.496	0.523
CNN-Sen (on dev set)	0.561	0.449	0.499
CNN-Doc (on test set)	0.563	0.515	0.538
CNN-Sen (on test set)	0.397	0.600	0.478

Table 6: Results of our method on the development and test data sets for BB-event task

Models	#Positive examples	#Negative examples	Ratio
Doc-level (on train set)	16%(298)	84%(1525)	5.1
Sen-level (on train set)	45%(227)	55%(275)	1.2
Doc-level (on dev set)	13%(210)	87%(1462)	6.9
Sen-level (on dev set)	32%(165)	68%(348)	2.1

Table 7: Statistics of positive and negative examples on training data and development data at the documental and sentence levels for BB-event task (ratio = #negative examples / #positive examples).

We conduct another experiment on SVM¹ to analysis the superiority of CNN model compared with SVM model. Each raw input into the SVM and CNN models is same, which contains words between two candidate entities, distance between two candidate entities, and the types of two candidate entities. Then, the raw input for SVM is represented traditional one-hot features, and the raw input for CNN is represented by distributed representation. In Table 8, we compared the two models. F-score of using CNN model is higher than that using SVM model on two data sets, which shows that the effectiveness of using CNN model and distributed representation.

Models	Recall	Precision	F-score
SVM (on dev set)	0.459	0.490	0.474
CNN (on dev set)	0.561	0.449	0.499
SVM (on test set)	0.336	0.594	0.429
CNN (on test set)	0.397	0.600	0.478

Table 8: Results of using SVM and CNN models on the development and test data sets for BB-event task

4 Conclusions

Instead of complicate hand-designed feature engineering, we employed the distributed semantic representation and CNN model to extract binary relations between entities. SeeDev-binary task and BB-event task are regarded as classification problems. And then, Word embedding, POS embedding, distance embedding and entity type embedding, which contain rich semantic knowledge, are built to be fed into Convolutional neural network and to learn the inner relationship between candidate entities. The results with F-score of 0.370 and 0.478 in our participant tasks, which were evaluated on the test data set with online evaluation² show that our proposed method has been contributed to binary relation extraction.

¹ http://www.cs.cornell.edu/People/tj/svm_light/

² <http://2016.bionlp-st.org/tasks/seedeval/seedeval-evaluation>

Only using embedding of original words fed into CNN, may be not sufficient for understanding the hidden information among words. Therefore, in our future work, we will still concentrate more on the building of rich distributed semantic embedding and construct a better representation with human knowledge for CNN model. Furthermore, we will also explore various neural networks with multi-layer architectures, such as RNN, to address binary relation or event extraction.

5 Acknowledge

This work is supported by the grants from the National Natural Science Foundation of China (No. 61572098, 61572102, 61272373 and 61300088), Trans-Century Training Program Foundation for the Talents by the Ministry of Education of China (NCET-13-0084) and the Fundamental Research Funds for the Central Universities (No. DUT13JB09).

References

- Blaschke C, Andrade M A, Ouzounis C A, et al. Automatic extraction of biological information from scientific text: protein-protein interactions[C]//Ismb. 1999, 7: 60-67.
- Bergstra J, Breuleux O, Bastien F, et al. Theano: a CPU and GPU math expression compiler[C]//Proceedings of the Python for scientific computing conference (SciPy). 2010, 4: 3.
- Bastien F, Lamblin P, Pascanu R, et al. Theano: new features and speed improvements[J]. arXiv preprint arXiv:1211.5590, 2012.
- Bossy R, Bessières P, Nédellec C. Bionlp shared task 2013—an overview of the genic regulation network task[J]. ACL 2013, 2013: 153.
- Bossy R, Golik W, Ratkovic Z, et al. BioNLP Shared Task 2013—an overview of the bacteria biotope task[C]//Proceedings of the BioNLP Shared Task 2013 Workshop. 2013: 161-169.
- Claveau V. IRISA participation to BioNLP-ST 2013: lazy-learning and information retrieval for information extraction tasks[C]//BioNLP Workshop, colocated with ACL 2013. 2013: 188-196.
- Hinton G E. Learning distributed representations of concepts[C]//Proceedings of the eighth annual conference of the cognitive science society. 1986, 1: 12.
- Krallinger M, Erhardt R A A, Valencia A. Text-mining approaches in molecular biology and bi-

omedicine[J]. Drug discovery today, 2005, 10(6): 439-445.

Kim J D, Ohta T, Pyysalo S, et al. Overview of BioNLP'09 shared task on event extraction[C]//Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task. Association for Computational Linguistics, 2009: 1-9.

Kim J D, Pyysalo S, Ohta T, et al. Overview of BioNLP shared task 2011[C]//Proceedings of the BioNLP Shared Task 2011 Workshop. Association for Computational Linguistics, 2011: 1-6.

Karadeniz I, Ozgür A. Bacteria biotope detection, ontology-based normalization, and relation extraction using syntactic rules[C]//Proceedings of the BioNLP Shared Task 2013 Workshop. 2013: 170-177.

Nédellec C, Bossy R, Kim J D, et al. Overview of BioNLP shared task 2013[C]//Proceedings of the BioNLP Shared Task 2013 Workshop. 2013: 1-7.

Provoost T, Moens M F. Detecting relations in the gene regulation network[C]//Proceedings of BioNLP shared task 2013 workshop: the Genia event extraction shared task. 2013: 135-138.

Rodrigues A D (Ed). Drug-drug interactions. CRC Press. 2008.

Ratkovic Z, Golik W, Warnier P, et al. BioNLP 2011 task bacteria biotope: the Alvis system[C]//Proceedings of the BioNLP Shared Task 2011 Workshop. Association for Computational Linguistics, 2011: 102-111.

Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

Zweigenbaum P, Demner-Fushman D, Yu H, Cohen KB. Frontiers of biomedical text mining: current progress. Brief Bioinform. 2007, 8:358-375.

Zeiler M D. ADADELTA: an adaptive learning rate method[J]. arXiv preprint arXiv:1212.5701, 2012.