Trigger Words Detection by Integrating Attention Mechanism into Bi-LSTM Neural Network — A Case study in PubMED-wide Trigger Words Detection for Pancreatic Cancer

Kaiyin Zhou[12], Xinzhi Yao[1], Shuguang Wang[1], Jin-Dong Kim[3], Kevin Bretonnel Cohen[4], Ruiying Chen[1], Yuxing Wang[12], and Jingbo Xia[12*]

[1] College of Informatics, Huazhong Agricultural University, Wuhan, China
[2] Hubei Key Laboratory of Agricultural Bioinformatics, Wuhan, China)
[3] Database Center for Life Science (DBCLS), Research Organization of Information and Systems (ROIS), Tokyo, Japan
[4] School of Medicine, University of Colorado Denver, Anschutz Medical Campus, U.S
[*]Correspondence: xiajingbo.math@gmail.com, xjb@mail.hzau.edu.cn

Abstract. A Bi-LSTM based encode/decode mechanism for named entity recognition was studied in this research. In the proposed mechanism, Bi-LSTM was used for encoding, an Attention method was used in the intermediate layers, and an unidirectional LSTM was used as decoder layer. By using element wise product to modify the conventional decoder layers, the proposed model achieved better F-score, compared with other three baseline LSTM-based models. For the purpose of algorithm application, a case study of causal gene discovery in terms of disease pathway enrichment was designed. In addition, the causal gene discovery rate of our proposed method was compared with another baseline methods. The result showed that trigger genes detection effectively increase the performance of a text mining system for causal gene discovery.

Keywords: natural language processing · LSTM · encoder/decoder model · trigger words.

1 Introduction

Named Entity Recognition (NER) is to detect mentions that we concerned from text [1], and NER is generally the first stage for complex natural language processing tasks (NLP). In tradition, most sequence labeling models were linear statistical models. However, these models usually heavily depended on specific feature engineering and high-quality labeled data. In recent years, the development of deep learning has broke this limitation dramatically. Performance of natural language processing tasks, such as machine translation, semantic relation extraction, automatic summarization, and so on, have successfully outperformed conventional machine learning methods, and NER tasks are no exception.

Recently, attention mechanism was successfully applied to the machine translation model and achieved state-of-art result in many public data sets [2]. This

made us recognize the importance of the attention mechanism. In the machine translation model, query vector was used to perform alignment with each encoding results, which was considered as an effective simulation of human reading and translating process. This method hinted that the attention mechanism would also be suitable for the sequence labeling model.

In a cross-disciplinary field of Biomedical Natural Language Processing (BioNLP), the mainstream text data comes from a huge publication repository PubMED (https://www.ncbi.nlm.nih.gov/pubmed/), run by U.S national library of medicine and national institutes of health. The dataset has been increasing dramatically, and the amount reached 28 million in 2018. PubMED was treated as a useful resource for Bioinformatics research to curate bio-related knowledge [3, 4]. For example, 93,096 PubMED abstracts entries were found for pancreatic cancer at 6th, June, 2018, and the amount was still increasing. Since straightforward curation of mass data was hard for knowledge discovery, automatic NER of trigger words made it possible to increase the relevant entries filtering and knowledge inference [5–7].

In this paper, we modified the classic encoder/decoder model [8] for NER of trigger words, incorporated with Bengio's attention mechanisms [9]. In the encoding layer we used the usual Bidirectional Long Short Term Memory (Bi-LSTM) structure to capture the context information. In the decoding layer we used an unidirectional LSTM. For the LSTM unit, the output of encoding layer and the result of the attention mechanism were combined in each time step appropriately. Unlike the usual case that these two elements were connected in a straightforward manner, this is an effective modification taken in our proposed scheme. Actually, an illuminative trick-playing were carried on by replacing concatenation to element-wise product. Henceforth, a Bi-LSTM-Attention-ElmentwiseProduct (Bi-LSTM-AEP) algorithm was achieved. By applying this new model to a manually labeled dataset, the highest F-score was obtained by the proposed algorithm after comparing to other three baseline popular sequence labeling methods — two without attention [10, 11], and one with attention [12]. As an application, a bioinformatics case study was carried on by using the proposed trigger word NER algorithm to increase the discovery of causal genes that affect pancreatic cancer.

2 Material and Method

2.1 Dataset

An under-developed corpus, Active Gene Annotation Corpus (AGAC) [13], was chosen as the training set in this research. AGAC corpus contained structured texts with semi-manual annotations, which included five trigger labels containing the biological concepts from molecular level to cell level, three regulatory labels representing the mutant directions and two kinds of semantic relations between trigger words.

To better understand the trigger word setting in this corpus, a snapshot of annotation example in PubAnnotation platform [14] for trigger words is shown at figure 1.

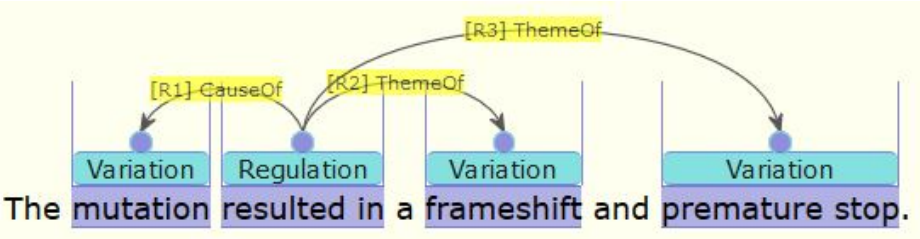

Fig. 1. An example for AGAC

In this work, we focused on 2 trigger labels: *Variation* and *Regulation*. Both of them are important for representing the function of genes, and these trigger labels were treated as the essential elements to present the causal gene information.

As a typical example shown in Figure 1, a sentence "The mutation resulted in a frame shift and premature stop" was annotated. In this annotation manner, *Variation* label included the origin causality of all the other biological concept labels, and *Regulation* label was the regulatory label which was the center of a relation. In this sentence, "mutation", "result in", "frameshift", and "premature stop", were all treated as trigger words. Therefore, if a gene co-occurred with the above trigger words, this gene actually would play a so-called Loss-of-function role [15] in the molecular level, and that made this gene a causal gene in the context.

Since the purpose of this research is to compare the Bi-LSTM-AEP with other popular encode/decode-based mechanisms, AGAC corpus is an acceptable data resource for training and testing. In addition, it was assumed that trigger words in AGAC corpus represented the functioning effectiveness of curated genes, and this assumption made it possible to discriminate causal genes in a mass text data, i.e., PubMED.

2.2 Word embeddings library

It is widely accepted that the quality of the pre-trained word embeddings is a key factor for various NLP tasks, including NER [16]. Due to the differences in writing style and huge variation of terminology in biology, it is still a challenge to make a domain-free word embedding library to suit the biology-related application [17]. Therefore, a domain-specific word embeddings, BioASQ [18], was selected as our pre-trained embedding library. BioASQ was trained from a corpus of 10,876,004 English abstracts of biomedical articles from PubMed and

contained 1,701,632 distinct words, thus that made it a proper one for embedding preprocessing in our method.

2.3 General encode/decode framework based on recurrent neural network, and two LSTM-based baseline methods without Attention mechanism

As a general encode/decode framework for tackling NER or machine translation task [8], the input is a sequence of vectors $x = (x_1, \cdots, x_{T_x})$, and output is a sequence of label vectors or word vector $y = y_1, \cdots, y_{T_y}$. For tacking NER, in a general Encoder-Decoder framework, there is an encoder which reads the input sentence into a context vector c, and a decoder which predicts the next word y'_t given c and all the previously predicted words $\{y_1, \cdots, y_{t'-1}\}$.

In recent years, the most common approach is to use a recurrent neural network (RNN) and make decoder represent a probability over y by decomposing the joint probability into the ordered conditionals,

$$P(\{y_1, \cdots, y_{T_y}\}) = \prod_{t=1}^{T_y} P(\ y_t | \{y_1, \cdots, y_{t-1}\}, c\ ), \tag{1}$$

and models each conditional probability as

$$P(y_t | \{y_1, \cdots, y_{t-1}\}, c) = g(y_{t-1}, s_t, c), \tag{2}$$

where the context vector c is computed by using information of hidden layer of RNN, i.e., $h_t = RNN(x_t, h_{t-1})$.

The reason for using RNN in encode/decode frame work stems from the powerful modeling structure of recurrent network, generally speaking, is that RNN is regarded as being capable of capturing time dynamics via cycles in the graphs. Though RNN usually leads to gradient vanishing/exploding problems(Bengio et al.,1994;Pascanu et al.,2012), some RNN variants perform better in practice, e.g. LSTM neural network.

LSTM is a set of special Recurrent neural networks, which could especially capture long-distance dependencies with the appropriate employment of gating functions at each time step i.e. input gate, forget gate, and output gate. These tricks are good solutions to the gradient vanishing for conventional RNNs. Formally, the formulas to update an LSTM unit at time t are:

$$\begin{cases} i_i = \sigma(W_{Ei} E_{xi} + U_{hi} h_{i-1} + b_i) \\ f_i = \sigma(W_{Ef} E_{xi} + U_{hf} h_{i-1} + b_f) \\ z_i = \tanh(W_{Ez} E_{xi} + U_{hz} h_{i-1} + b_z) \\ c_i = f_i * c_{i-1} + i_i * z_i \\ o_i = \sigma(W_{Eo} E_{xi} + U_{ho} h_{i-1} + b_o) \\ h_i = o_i * \tanh(c_i) \end{cases}, \tag{3}$$

where σ is the element-wise sigmoid function and * is the element-wise product.Here E_{xi} is the input vector at time i, usually represented by word embedding. h_{i-1} is the hidden state vector of last time step $i-1$. $W_{Ei}, W_{Ef}, W_{Ez}, W_{Eo}$ are the weight matrices of different gates for input E_{xi}, and $U_{hi}, U_{hf}, U_{hz}, U_{ho}$ are the weight matrices for hidden state h_t.

2.3.1 Baseline method 1: Bi-LSTM-CRF Model

Bi-LSTM-CRF [10] is a typical neural network model used in sequence labeling tasks. It carried on LSTM training with the data for two times, and the only difference for each time was that the order of the two times input data was completely reversed, then the results of each LSTM layer were concatenated as an output of words encoding results. Thus, Bi-LSTM model captured both the past and the future information respectively. Then, the output vectors of Bi-LSTM were fed to the CRF layer to jointly decode the best label sequence. This model has been proved to be reasonable and has achieved state-of-art scores on many sequence labeling tasks.

2.3.2 Baseline method 2: Bi-LSTM-ED Model

Bi-LSTM-ED [11] is another model used to carry on sequence labeling tasks. This model still used Bi-LSTM as encoding layers. Being different from other models, the decode layer in this model was a Variant LSTM. The units of the decoding LSTM were the same as the encoding LSTM except for the input gate, which was replaced by

$$\begin{cases} i_i = \sigma(W_{ii}E_{xi} + U_{ii}h_{i-1} + V_{ii}T_{i-1} + b_{ii}) \\ T_i = W_{is}h_i + b_{is} \end{cases}. \tag{4}$$

This model obtained good improvement in several sequence labeling tasks.

2.4 Proposed LSTM and Attention neural network for trigger words recognition

In this section, Liu's model [12], a similar structure of Bi-LSTM-Attention [9], is listed as the 3rd baseline method. Subsequently, we propose an updated encode/decode scheme for NER by using Bi-LSTM and attention mechanism, while element product is used in decoding layer, and the abbreviation of this structure is Bi-LSTM-AEP.

In the conventional encode/decode attention-based models, encode layer is a bidirectional recurrent neural network (Bi-LSTM), which reads the words one by one in a sentence and then outputs metrics containing the forward and backward hidden states of each words. Attention mechanism acts as a weight calculator to change the importance of different inputs by taking both inputs and outputs into consideration. Decode layer is a unidirectional LSTM.There are many variants about this model. We notice that all of the variants only concerned how to optimize attention mechanism and design a appropriate score function,for instance, Luong et.al (2015) [19] presented a contend based and location based

score functions, and Jean et al. (2014) [20] added the target word embedding as input for the score function.

2.4.1 Baseline method 3: Bi-LSTM-Attention Model

In past few years, Bi-LSTM-Attention model [9] has been widely used in speech recognition, image caption generation, visual question answering, machine translation and other fields, while few people applied it to sequence labeling tasks. In 2016, Liu [12] converted Bi-LSTM-Attention model into a NER-purposed one, where s_i in encoder layer was computed by a GRU unit,

$$s_i = GRU(h_i, s_{i-1}, c_i), \tag{5}$$

where the detailed formulas are below:

$$\begin{cases} s_i = GRU(h_i, s_{i-1}, c_i) = (1 - z_i) \circ s_{i-1} + z_i \circ \tilde{s_i} \\ \tilde{s_i} = \tanh(W h_{i-1} + U[r_i \circ s_{i-1}] + C c_i) \\ z_i = \sigma(W_Z h_i + U_z s_{i-1} + C_z c_i) \\ r_i = \sigma(W_r h_i + U_r s_{i-1} + C_r c_i) \end{cases} . \tag{6}$$

2.4.2 Proposed mechanism: Bi-LSTM-AEP neural network

In our work, we also applied Bahdanau et al's model [9] in NER and sequence labeling tasks. Here attention mechanism was used to adjust the weight of input information. As introduced in the above section, the difference of Bi-LSTM-Attention model and our proposed model mainly exist at the Decoder-layer, see equation (6) and (9), where GRU and element-wise LSTM were used separately.

In addition, we focus on how to combine the output of attention mechanism with the input of decoding layer. Here we propose an element-wise multiplication rather than conventional weight sum. The structure of the neural network is shown in Figure 2, while the complete description of the framework is shown as below, which is a modification of Bi-LSTM-Attention model [9].

Algorithm of Bi-LSTM-AEP:

- Encode layer:
  Since we were desired to take both the last word and the next word into consideration, Bi-LSTM was employed as the encode layer, which contains a forward LSTM and a backward LSTM. At first, forward LSTM and backward LSTM read the words in a sentence x=$(x_1, x_2, ..., x_{Tx})$ respectively, and calculates the hidden states of each word: ($\overrightarrow{h_1}$,$\overrightarrow{h_2}$,..., $\overrightarrow{h_{Tx}}$) and ($\overleftarrow{h_1}$,$\overleftarrow{h_2}$,..., $\overleftarrow{h_{Tx}}$) [9].
  Then the forward hidden states and the backward hidden states were combined in the third dimension as the annotation for each word: $h_j = [\overrightarrow{h_j}, \overleftarrow{h_j}]$. Therefore, annotation h_j represents not only the x_j itself but also the context information around it.
- Attention-Mechanism:
  After encoding, the input words were transformed to annotation h. If attention mechanism was not taken into consideration, the annotation h for each

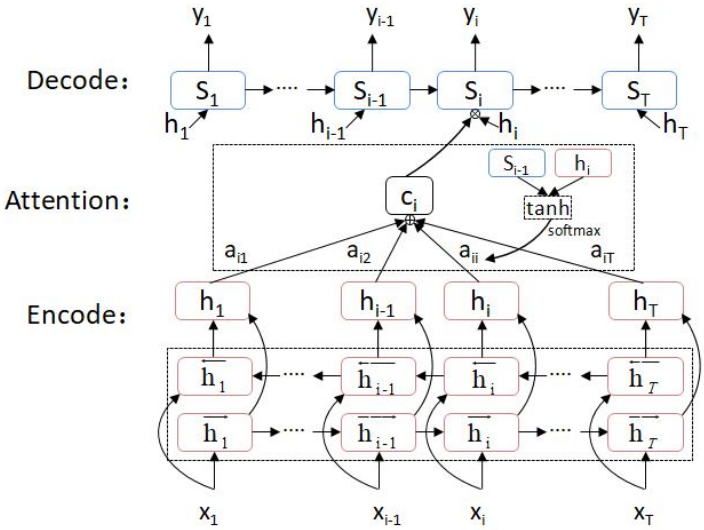

Fig. 2. Architecture of Bi-LSTM-AEP neural network

input words (from h_1 to h_j) would be combined directly as a context vector which contained all the information in the sentence equally, and then the context vector was one of the inputs to decode layer. However, the importance of each annotation h should be different, so we introduced attention mechanism to calculate the different weight for each input by scoring to the alignment of input at position j and output at i [9].
At this part, the inputs is annotation h for each words in the sentence and the hidden state at last position of output sequence s_{i-1}, and the output is the context vector after considering the weight for each annotation h. The formulas of attention mechanism are shown below:

$$\begin{cases} c_i = \sum_{j=1}^{T_x} a_{ij} h_j \\ a_{ij} = \frac{exp(e_{ij})}{\sum_{k=1}^{T_x} exp(e_{ik})} \\ e_{ij} = V_a^T tanh(W_a s_{i-1} + U_a h_j) \end{cases}, \tag{7}$$

where a_{ij} is the weight for different annotation h, c_i is the context vector, and e_{ij} is the alignment scores.

- Decode-layer:
The decode layer is a unidirectional LSTM. In previous models, the output of encode results $h = (h_1, h_2 ... hi, ... h_{Tx})$ and the context vector c_i from attention mechanism were added at each time step $s_i = GRU(h_i, s_{i-1}, c_i)$[12]. However, we proposed our decode-layer

$$s_i = LSTM(h_i * c_i, s_{i-1}), \tag{8}$$

where * is the element-wise product. The complete specific formulas are:

$$\begin{cases} i_i = \sigma(W_{ei}[h_i * c_i] + U_{si}s_{i-1}) \\ f_i = \sigma(W_{ef}[h_i * c_i] + U_{sf}s_{i-1}) \\ z_i = tanh(W_{ez}[h_i * c_i] + U_{sz}s_{i-1}) \\ \widetilde{z}_i = f_i * \widetilde{z}_{i-1} + i_i * z_i \\ o_i = \sigma(W_{eo}[h_i * c_i] + U_{so}s_{i-1}) \\ s_i = LSTM(h_i * c_i, s_{i-1}) = o_i tanh(\widetilde{z}_i) \end{cases} . \tag{9}$$

For simplicity, the bias terms were omitted in the above formulas.

3 Experimental settings

The experiments are designed for two purposes: 1) to preliminary evaluate the significance of the database, which we designed for large-scale biological literature mining; 2) to test the performance of our newly designed model.

In this work, 28 labeled texts were selected from the AGAC data set. BioASQ was employed as pre-trained embeddings, and the dimension of words vector is 200. The hidden unites of Encode/Decode layer are set as 100. The model is trained with the RMSProp algorithm.

Precision, recall and F1-measure are taken as the evaluation criteria. The finally evaluation scores are computed by averaging all tags scores. Weights in all formulas are randomly initialized as uniform distribution with support $[-0.01, 0.01]$.

3.1 Models

In order to better evaluate the performance of the proposed models, we compare it with four different kinds of model. Bi-LSTM-CRF is a usual used model in sequence labeling tasks, while Bi-LSTM-ED performed better in some tasks. Besides, Bi-LSTM-Attention was also provided to evaluate the importance of attention mechanism in such tasks.

3.2 Case study design for trigger words detection in terms of pancreatic cancer pathway enrichment

In order to evaluate the application of our trigger word detection in mass text. A case study was performed.

As a target disease, Pancreatic cancer is a disease that attracts much attention in academia. Through keywords searching in the PubMed-wide scale, ninety hundreds abstracts were downloaded and all of the bio-entities, including gene mentions,and were retrieved by using Pubtator [21]. By counting the occurrence of the gene mentions, the rank of active gene related to pancreatic cancer was obtained. According to a practice of co-occurrence manner, we obtained a knowledge entry rank list for causal gene discrimination in terms of pancreatic cancer.

In another words, the higher a gene ranked in the list, the higher chance the gene had to be relevant with the pancreatic cancer.

For the sake of performance evaluation for the application of trigger word detection in Re-ranking. Gene mentions were filtered by using Bi-LSTM-AEP Model. Here, a medium-sized training model of AGAC corpus were used to handle all of the Pancreatic-related PubMED abstracts. Subsequently, trigger words were detected for each abstract. In the re-ranked gene list, only genes co-occurred with these trigger words were treated causal and kept in the list. Thus a new gene list was obtained.

After using trigger words detection, the updated gene rank list were compared with the previous gene rank according its enrichment in pathway. Within domain knowledge, pathway is a directional map connected genes, proteins and metabolites, and the pathway enrichment analysis was generally an effective method to evaluate the accuracy of gene relevance. The design of the case study was shown in Figure 3.

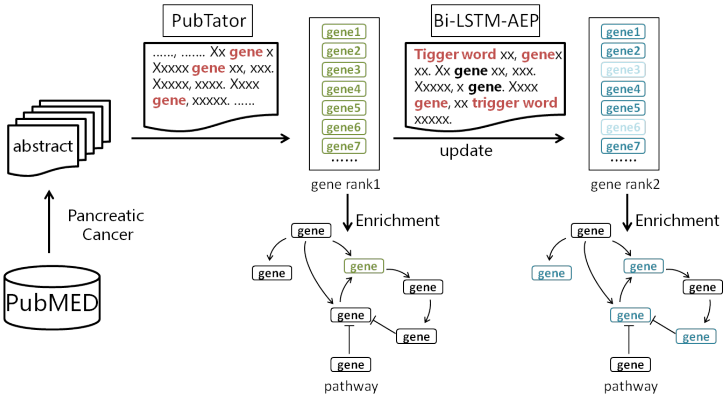

Fig. 3. A Case study in trigger words detection in terms of pancreatic cancer.

4 Result

4.1 Performance of trigger words extraction algorithm

The experimental results were based on a five-fold cross-validation and three replicates were performed. Finally, the results of the three experiments were averaged as the evaluation criteria.

Table 1 showed the results of the four different models: Bi-LSTM-Attention, Bi-LSTM-AEP, Bi-LSTM-CRF, Bi-LSTM-DE. In these four models, the first two were ones without attention mechanism, and the latter two were attention-based. A dramatic improvement in F value was achieved when using the attention mechanism.

In model with attention mechanism proposed by Liu [12], the F value is 0.4382. When using the model Bi-LSTM-AEP proposed by us, the F value further improved to 0.4951. As a result,the result showed that our proposed Bi-LSTM-AEP model outperformed other lSTM-based encoder/decoder models, and it was most suitable for our tasks.

Method	**Precision**	**Recall rate**	**F1-measure**
Bi-LSTM-AEP (Ours)	0.7576	0.4171	0.5160
Bi-LSTM-Attention [12]	0.7092	0.3286	0.4368
Bi-LSTM-ED [11]	0.6604	0.3249	0.4263
Bi-LSTM-CRF [10]	0.5849	0.3051	0.3947

Table 1. Experimental results on four different models.

4.2 Case study in PubMED-wide trigger words detection in Pancreatic cancer

Disease-related pathways are always the focus of attention in disease and drug research. Here, Kyoto Encyclopedia of Genes and Genomes (KEGG) is an authoritative and commonly used database in biological research. KEGG pathway database contains a large number of manually curated pathway maps focusing on intermolecular interaction networks. A biological pathway is a series of molecular actions in a cell, which produces metabolites or generate changes in the cell. For example, Inactivation of the SMAD4 tumour suppressor gene leads to a loss of the inhibitory influence of the transforming growth factor-beta (TGF-Beta) signaling pathway and henceforth promotes the occurrence of cancer.

To compare the accuracy of the above mentioned two methods, Pubtator and Bi-LSTM-AEP. The result of the comparison is shown in Table 2.

	Extracted terms	Extracted pathway genes	Accuracy
Method 1(Pubtator)	28336	54	0.19%
Method 2(Bi-LSTM-AEP)	11675	52	0.45%

Table 2. The result of the comparison

Due to the rigorousness of KEGG database, the number of pathway genes in pancreatic cancer is very small compared to the results of text mining. Even so, Bi-LSTM-AEP extracted 52 key genes out of 11,675 extracted ones, while the ratio for Pubtator was 54/28,336. The result showed that Bi-LSTM-AEP narrow down the gene searching scale from 28,336 to 11,675, which is a 58.80 % reduce in amount, and the target genes remained in the shorter list.

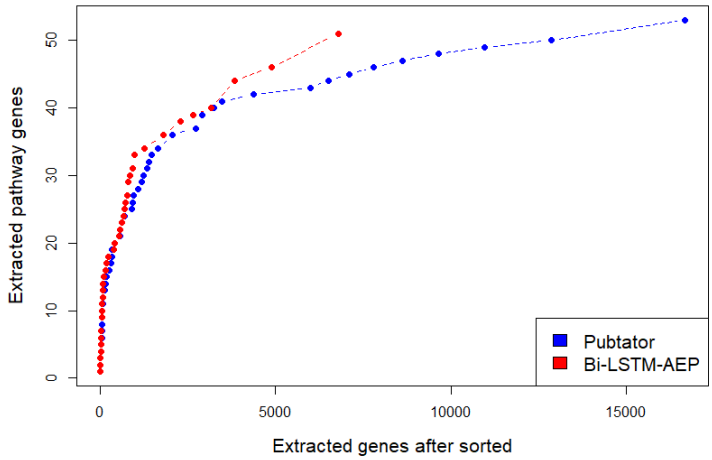

Fig. 4. Distribution of KEGG Pathway Genes in the Results of Two Methods

It's noted that the accuracy of both methods were lower than 1 percent, i.e., 0.19% and 0.45%, respectively. However, the discovery of causal gene was actually an unsolved challenge. Fortunately, using a better text mining tool was capable of achieve a higher knowledge discovery rate.

As shown in Figure. 4, in order to explore the distribution of the KEGG pathway genes in two gene lists, we visualized the causal gene discovery rate by representing the appearing order of relevant genes. In detail, extracted genes were sorted by frequency, and the amount of accumulated target genes was marked with round dot at the corresponding sorted positions of the KEGG genes. In the figure 4, it clearly showed that the red line is at the top left of the blue one, which was sufficed to discover more target genes within less amount.

5 Acknowledgement

This work is funded by the Fundamental Research Funds for the Central Universities of China (Project No. 2662018PY096).

References

1. David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. Lingvisticae Investigationes, 30(1):326.
2. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[J]. 2017.
3. Sintchenko V, Anthony S, Phan XH, Lin F, Coiera EW. A PubMed-wide associational study of infectious diseases. PLoS One. 2010 Mar 10;5(3):e9535.

4. Allot A, Peng Y, Wei CH, Lee K, Phan L, Lu Z. LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. Nucleic acids research. 2018 May 14.
5. Cohen KB, Verspoor K, Johnson HL, Roeder C, Ogren PV, Baumgartner Jr WA, White E, Tipney H, Hunter L. High-precision biological event extraction with a concept recognizer. InProceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task 2009 Jun 5 (pp. 50-58). Association for Computational Linguistics.
6. Song M, Kim M, Kang K, Kim YH, Jeon S. Application of Public Knowledge Discovery Tool (PKDE4J) to Represent Biomedical Scientific Knowledge. Frontiers in Research Metrics and Analytics. 2018 Feb 26;3:7.
7. Zhou H, Yang Y, Ning S, Liu Z, Lang C, Lin Y, Huang D. Combining Context and Knowledge Representations for Chemical-disease Relation Extraction. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2018 May 21.
8. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, and Bengio Y. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
9. Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science, 2014.
10. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv:1508.01991, 2015.
11. Zheng S, Hao Y, Lu D, et al. Joint Entity and Relation Extraction Based on A Hybrid Neural Network[J]. Neurocomputing, 2017, 257(000):1-8.
12. Liu B, Lane I. Attention-based recurrent neural network models for joint intent detection and slot filling. arXiv preprint arXiv:1609.01454. 2016 Sep 6.
13. Wang Y, Yao X, Zhou K, Qin X, Kim J D, Cohen K B, Xia J*. Guideline Design of An Active Gene Annotation Corpus for the Purpose of Drug Repurposing. OHDSI 2018 workshop, July, Guangzhou. (Submitted)
14. Kim JD, Wang Y. PubAnnotation: a persistent and sharable corpus and annotation repository. InProceedings of the 2012 Workshop on Biomedical Natural Language Processing 2012 Jun 8 (pp. 202-205). Association for Computational Linguistics.
15. Wang Z Y, Zhang H Y. Rational drug repositioning by medical genetics. Nature Biotechnology, 2013, 31(12):1080-2.
16. Ma X, Hovy E. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. 2016.
17. Huang EH, Socher R, Manning CD, Ng AY. Improving word representations via global context and multiple word prototypes. InProceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1 2012 Jul 8 (pp. 873-882). Association for Computational Linguistics.
18. Ioannis P, Aris K, Ion A. Continuous Space Word Vectors Obtained by Applying Word2Vec to Abstracts of Biomedical Articles. 2014.
19. Luong M T, Pham H, Manning C D. Effective Approaches to Attention-based Neural Machine Translation[J]. Computer Science, 2015.
20. Jean S, Cho K, Memisevic R, et al. On Using Very Large Target Vocabulary for Neural Machine Translation[J]. Computer Science, 2014.
21. Wei CH, Kao HY, Lu Z. PubTator: a web-based text mining tool for assisting biocuration. Nucleic acids research. 2013 May 22;41(W1):W518-22.