


RESEARCH

Open Access



Document-level medical relation extraction via edge-oriented graph neural network based on document structure and external knowledge

Tao Li¹, Ying Xiong¹, Xiaolong Wang¹, Qingcai Chen^{1,2} and Buzhou Tang^{1,2*} 

From The 5th International Workshop on Semantics-Powered Data Mining and Analytics (SEPDA 2020) Virtual. 16-19 December 2020

Abstract

Objective: Relation extraction (RE) is a fundamental task of natural language processing, which always draws plenty of attention from researchers, especially RE at the document-level. We aim to explore an effective novel method for document-level medical relation extraction.

Methods: We propose a novel edge-oriented graph neural network based on document structure and external knowledge for document-level medical RE, called SKEoG. This network has the ability to take full advantage of document structure and external knowledge.

Results: We evaluate SKEoG on two public datasets, that is, Chemical-Disease Relation (CDR) dataset and Chemical Reactions dataset (CHR) dataset, by comparing it with other state-of-the-art methods. SKEoG achieves the highest F1-score of 70.7 on the CDR dataset and F1-score of 91.4 on the CHR dataset.

Conclusion: The proposed SKEoG method achieves new state-of-the-art performance. Both document structure and external knowledge can bring performance improvement in the EoG framework. Selecting proper methods for knowledge node representation is also very important.

Keywords: Medical relation extraction, Graph neural network, Document structure, External knowledge

Background

Relation extraction (RE) that extracts relations among entities in the text is a fundamental task of natural language processing (NLP). There may be two kinds of RE: (1) sentence-level RE that extracts relations in the same sentence, called intra-sentence relations; (2) document-level RE that extracts relations in the same sentence and cross sentences, and the relations cross sentences are called inter-sentence relations. Compared with

sentence-level RE, document-level RE is more challenging as document-level RE needs to consider both intra-sentence relations and inter-sentence relations as a whole, as shown in Fig. 1.

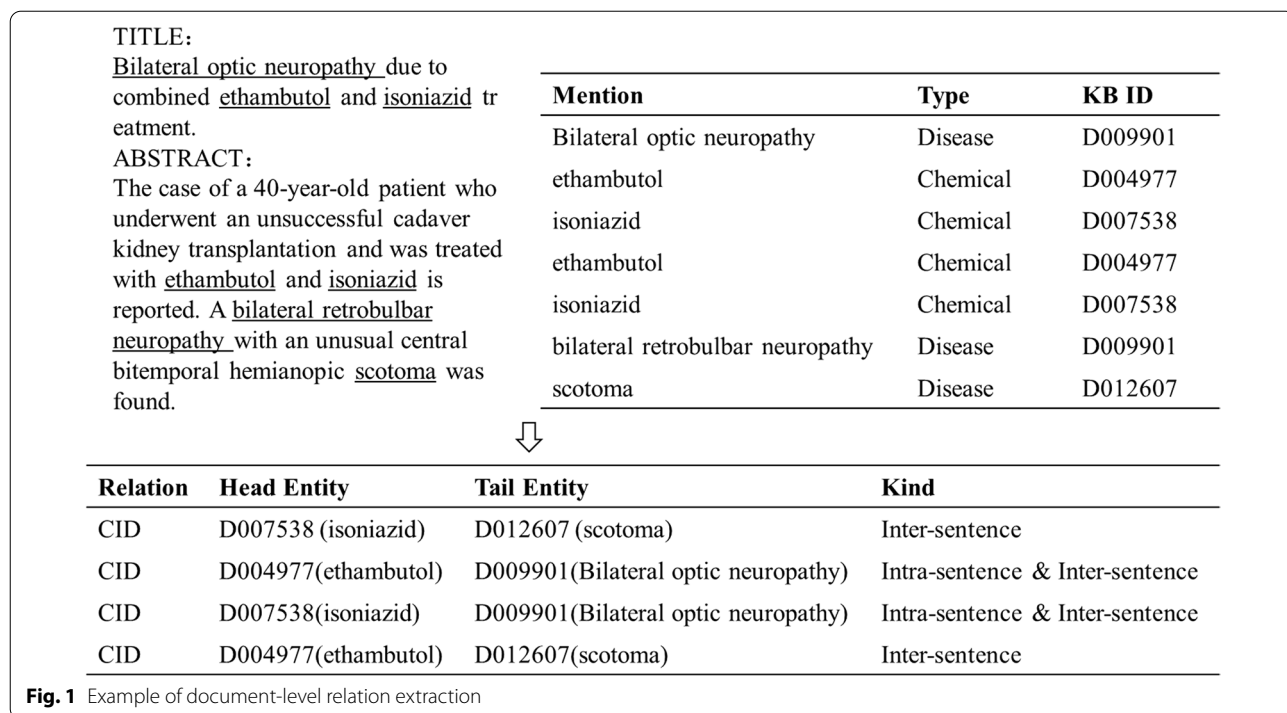
In recent years, document-level RE has attracted more and more attention from researchers, and various kinds of machine learning methods have been proposed. Among these methods, multi-instance learning (MIL) first introduced by Riedel et al. [1] for document-level RE is one of the most popular. MIL models multiple entity mention pairs of the same two given entities over a document and has the ability to reduce noise in distant supervised learning [2, 3]. Although the existing MIL methods

*Correspondence: tangbuzhou@gmail.com

¹ Harbin Institute of Technology, Shenzhen, China

Full list of author information is available at the end of the article





achieve considerable results, they also have some disadvantages. One disadvantage of these methods is that all entity pairs are considered individually and the implicit correlations among entities in different pairs in a document are ignored.

Graph neural networks (GNNs) that can represent the whole document and consider implicit correlations among entities in different pairs have shown great potential for document-level RE [4–10]. They may fall into two categories: (1) node-oriented GNNs [4–9]; (2) edge-oriented GNNs (denoted by EoG) [10]. Node-oriented GNNs mainly focus on node representation, while edge-oriented GNNs mainly focus on edge representation. As a relation between two entities is an instinctive edge in GNNs, edge-oriented GNNs outperformed node-oriented GNNs on document-level RE in some studies [10, 11]. In the case of EoG, document structure and external knowledge have been proved meaningful. However, there is no study to investigate them comprehensively.

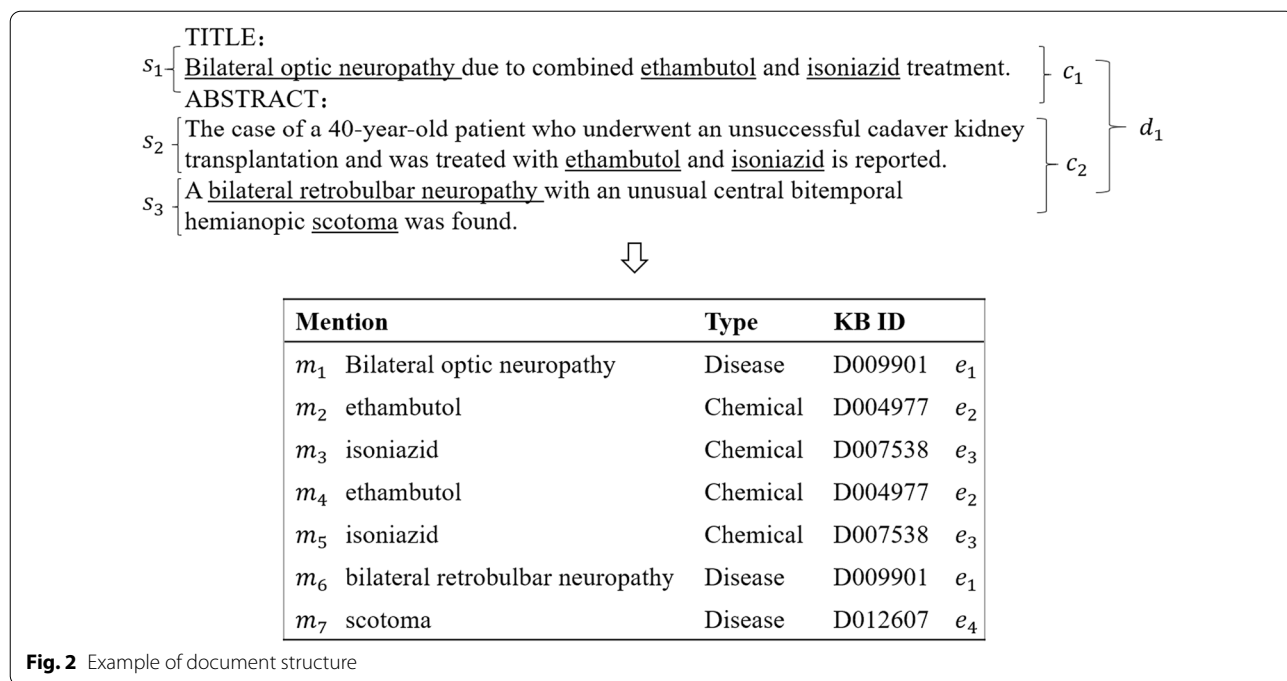
In this study, based on the backbone of EoG, we propose a novel GNN to consider document structure and external knowledge for document-level RE comprehensively, called SKEoG, which is an extension of KEoG proposed in our previous study [11]. To evaluate SKEoG, we conduct experiments on two public medical datasets. Experiment results show that both document structure and external knowledge are beneficial to document-level medical RE in the backbone of EoG, and the

proposed SKEoG model achieves new state-of-the-art performance, outperforming KEoG.

Related work

The studies most related to our work are EoG [10] and KEoG [11]. EoG is the first edge-oriented GNN for document-level RE proposed by Christopoulou et al. [10]. In EoG, information at different levels, including mention, entity and sentence, are regarded as nodes connected by five types of edges. EoG models document-level relations between entities directly and achieves much better results than node-oriented GNNs [10]. KEoG is an extension of EoG by introducing two new types of nodes regarding the document itself and knowledge concept and two new types of edges to connect the two new types of nodes. KEoG shows much better performance than EoG [11].

Inspired by KEoG [11], we propose a novel EoG that considers document structure and external knowledge comprehensively, that is SKEoG. Based on KEoG, SKEoG further introduces two new types of nodes, one regarding document structure and the other regarding external knowledge such as entity description. In this study, we use three models, that is, TransE [12], RESCAL [13] and GAT [14] to represent knowledge node based on knowledge graph respectively, and use two models, that is Doc2vec [15] and an end-to-end neural network, to represent knowledge node based on entity description.



Methods

In this section, we first introduce RE based on document structure, and then RE based on external knowledge from two aspects: knowledge graph and entity description.

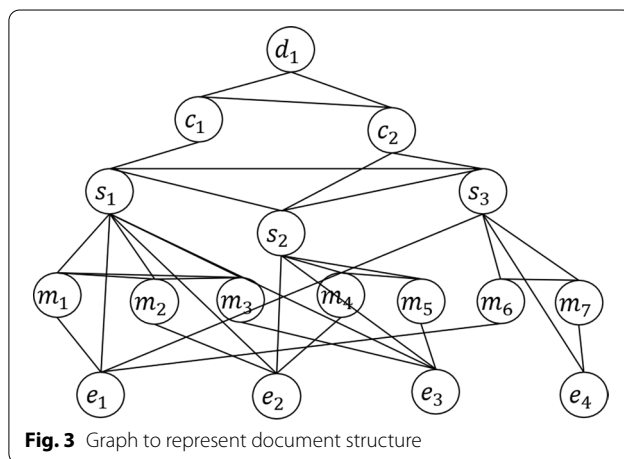
Relation extraction based on document structure

A document usually has a hierarchical structure like an example, as shown in Fig. 2, where a document d_1 consists of two chapters c_1 and c_2 , and each chapter contains some sentences with many entity mentions. Suppose that a sentence $s = w_1 w_2 \dots w_{|s|}$, it can be represented as $H_s^{local} = [h_1^{local}, h_2^{local}, \dots, h_{|s|}^{local}]$ via an encoding layer.

In a document with $|d|$ sentences $d = s_1, s_2, \dots, s_{|d|}$, there are five kinds of nodes corresponding to document structure as follows:

- Mention Node (M). Each mention node m is represented as $n_m = [avg_{w_i \in m}(h_i^{local}); t_m]$, where ‘;’ denotes concatenation operation, and t_m is an embedding to represent the node type of mention node.
- Entity Node (E). An entity e is represented as $n_e = [avg_{m \in e}(n_m); t_e]$, where $avg_{m \in e}(n_m)$ is the average representation of all mentions corresponding to e , and t_e is an embedding to represent the node type of entity node.

- Sentence Node (S). Each sentence node s is represented as $n_s = [avg_{w_i \in s}(h_i^{local}); t_s]$, where t_s is an embedding to represent the node type of sentence.
- Chapter Node (C). A chapter node c is represented by the average representation of all sentence nodes it contains and the embedding of the node type of chapter, that is, $n_c = [avg_{s \in c}(n_s^{global}); t_c]$;
- Document Node (D). A document node d is represented by the average representation of all chapter nodes and the embedding of the node type of document $n_d = [avg_{c \in d}(n_c); t_d]$.



Given the five kinds of nodes above, we connect them with the following six kinds of edges, as shown in Fig. 3:

- Mention-Sentence (MS). When an entity mention m appears in a sentence, there is an edge between the corresponding entity mention node and the sentence node s , and the edge is represented as $e_{MS} = [n_m; n_s]$
- Mention-Mention (MM). When two entity mentions m_1 and m_2 appear in the same sentence s , there is an edge between the two corresponding entity mention nodes n_{m_1} and n_{m_2} . The edge can be represented as $e_{MM} = [n_{m_1}; n_{m_2}; c_{m_1 m_2}; d(s_1, s_2)]$, where $d(m_1, m_2)$ is the representation of the relative distance between the two entity mentions in the sentence, and $c_{m_1 m_2}$ is the attention vector between the two entity mentions calculated by the following equations:

$$\alpha_{k,i} = n_{m_k}^T w_i, \tag{1}$$

$$a_{k,i} = \frac{\exp(\alpha_{k,i})}{\sum_{j \in [1, n], j \neq m_k} \exp(\alpha_{k,j})}, \tag{2}$$

$$a_i = \frac{a_{1,i} + a_{2,i}}{2}, \tag{3}$$

$$c_{m_1, m_2} = H^T a, \tag{4}$$

where $k \in \{1, 2\}$, a_i is the attention weight of the i th word in the entity mention pair $\langle m_1, m_2 \rangle$, and $H \in R^{hidden_dim \times |s|}$ is the representation of sentence s ;

- Entity-Mention (ME). There is an edge between an entity mention node m and the corresponding entity node e , that is, $e_{ME} = [n_m; n_e]$
- Sentence-Sentence (SS). For all sentence nodes in a document, there are edges between any two sentence nodes. An SS edge is represented by $e_{SS} = [n_{s_i}; n_{s_j}; d(s_i, s_j); |n_{s_i} - n_{s_j}|]$ ($i \neq j$), where n_{s_i} and n_{s_j} are the representation of s_i and the representation of s_j , and $d(s_i, s_j)$ is the representation of the relative distance between s_i and s_j measured by the number of sentences between them;
- Entity-Sentence (ES). When there is an entity mention node m corresponding to an entity node e in a sentence s , there is an edge between e and s . The edge is represented as $e_{ES} = [n_e; n_s]$
- Sentence-Chapter (SC). There is an edge between a sentence node s and a chapter node c , and it is represented as $e_{SC} = [n_s; n_c]$
- Chapter-Chapter (CC). There is an edge between two chapter nodes c_1 and c_2 in a document, and it is represented as $e_{CC} = [n_{c_1}; n_{c_2}]$

- Chapter-Document (CD). There is an edge between a chapter node c and a document node d , and it is represented as $e_{DC} = [n_d; n_c]$.

We further apply a linear transformation to all edge representations using the following equation:

$$v_z^{(1)} = W_z e_z, \tag{5}$$

where $z \in \{MS, MM, ME, SS, ES, SC, CC, CD\}$ and W_z is a learnable parameter matrix.

Relation extraction based on external knowledge

To utilize external knowledge, we regard any entity in external knowledge that also appears in text as an additional node and connect it to the corresponding entity node in text. In this paper, we introduce two kinds of knowledge nodes according to the forms of external knowledge of entities: (1) entity description and (2) knowledge graph.

Suppose that e_1 , e_2 and e_3 have their external description, e_1 and e_3 exist in an external knowledge graph, the graph based on document structure as shown in Fig. 3 can be extended to the graph as shown in Fig. 4 after adding knowledge nodes, where kd_i and ks_j denote knowledge node based on entity description and knowledge node based on knowledge graph, respectively. In this way, we can obtain a graph that takes full advantage of external knowledge as much as possible.

Knowledge node representation based on knowledge graph

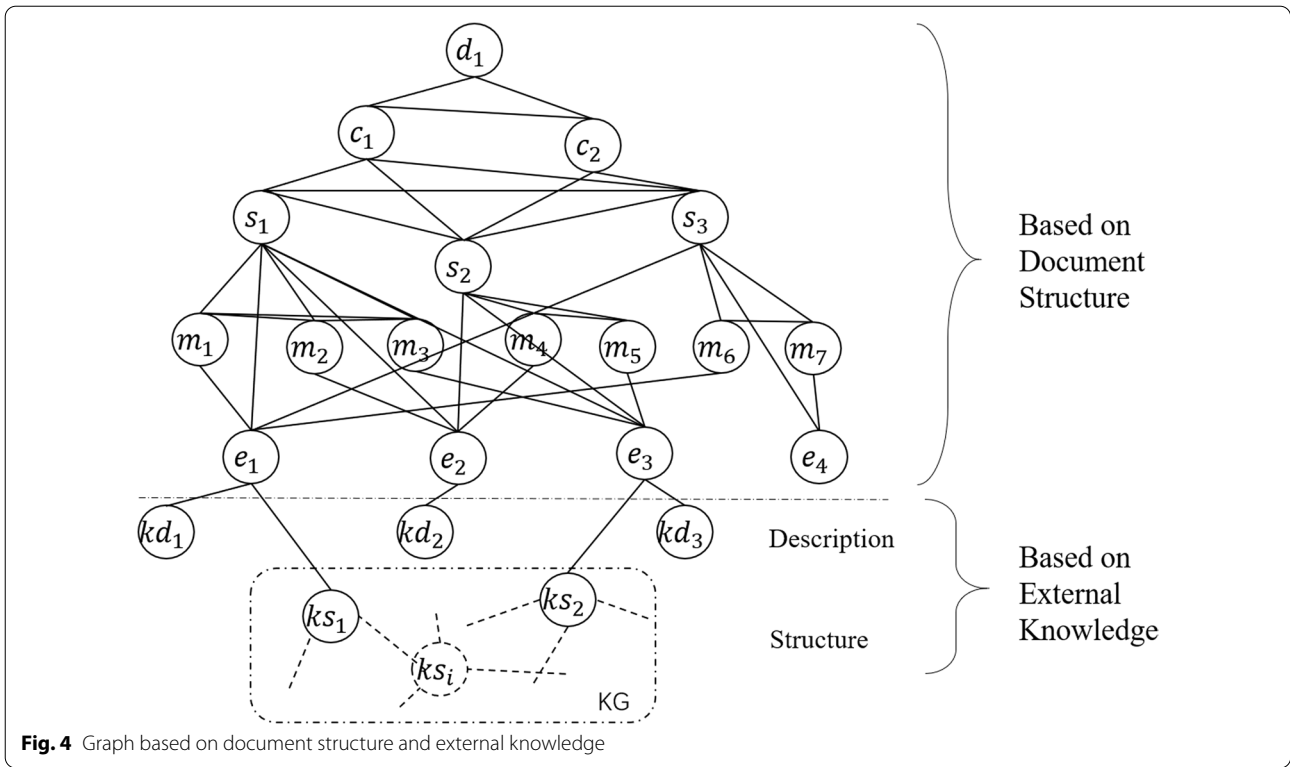
We deploy a translation distance model, a semantic matching model and a graph model, that is, TransE [12], RESCAL [13] and GAT [14], to represent knowledge nodes based on knowledge graph respectively.

TransE assumes that any triple $\langle h, r, t \rangle$, where h is a head entity node, r is a relation, and t is a tail entity node, satisfies the hypothesis of $h + r \approx t$, so as to ensure that the distance between two entity nodes is close to the representation of the relation between the two nodes. In this way, the multi-hop relation between two entities can be represented by additive transitivity, that is, if there is a relation r_1 between h_1 and t_1 , a relation r_2 between t_1 and t_2 , ..., and a relation r_K between t_{K-1} and t_K , there is an implicit relation between h_1 and t_K as follows:

$$h_1 + r_1 + r_2 + \dots + r_K \approx t_K, \tag{6}$$

The max-margin function of negative sampling is used as the objective function of TransE:

$$L = \sum_{(h,r,t) \in \Delta} \sum_{(h',r',t') \in \Delta'} \max(f_r(h,t) + \gamma - f_{r'}(h',t'), 0), \tag{7}$$



where $(h, r, t) \in \Delta$ is a true triplet, while $(h', r', t') \in \Delta'$ is a negative triplet obtained by sampling, $f_r(h, t)$ is the score of (h, r, t) , and $\gamma > 0$ denotes the margin usually set to 1. Finally, the learned h is regarded as h_{ks} , the knowledge node representation corresponding to node ks without considering its type.

RESCAL captures the potential semantics between two entities through the bilinear function as follows:

$$f_r(h, t) = h^T M_r t, \tag{8}$$

As shown in Fig. 5, RESCAL represents relation triples as a three-dimensional tensor \mathcal{X} , where $\mathcal{X}_{ijk} = 1$ indicates that there is a true triplet $\langle e_i, r_k, e_j \rangle$. The tensor decomposition model is used to model the relationship implicitly:

$$\mathcal{X}_k \approx AR_kA^T, \text{ for } k = 1, \dots, m, \tag{9}$$

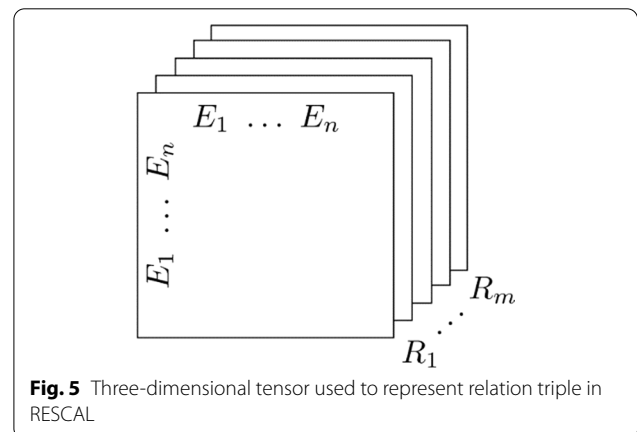
where \mathcal{X}_k is the k th component of \mathcal{X} , $A \in R^{n \times r}$ contains the potential representations of entities, $R_k \in R^{r \times r}$ is a symmetric matrix used to model the potential interactions in the k th relation:

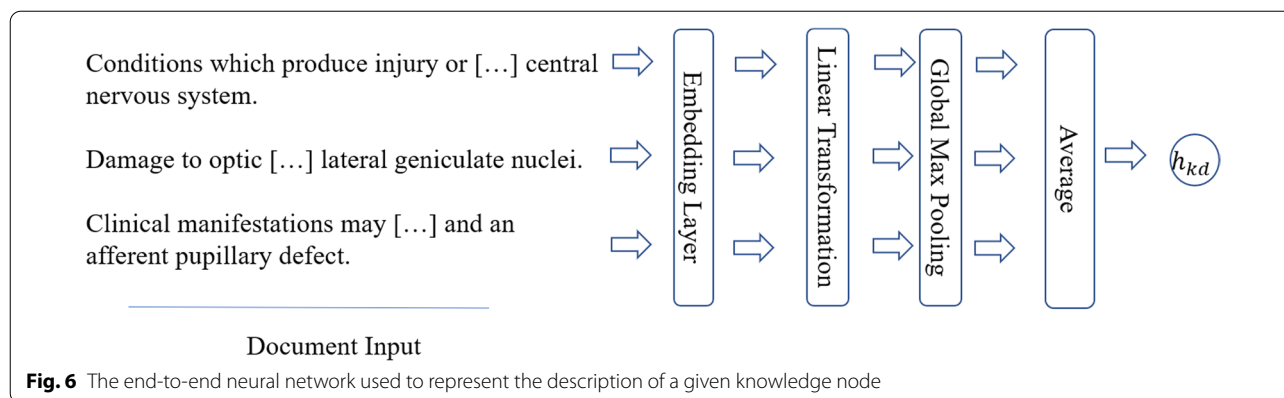
$$f(A, R_k) = \frac{1}{2} \sum_{i,j,k} (\mathcal{X}_{ijk} - \mathbf{a}_i^T R_k \mathbf{a}_j)^2, \tag{10}$$

where h_{ks} is the component of A corresponding to node ks .

In addition, we also represent the knowledge node ks by the subgraph centered on the node using GAT.

Based on knowledge graph, a node ks is represented by $n_{ks} = [h_{ks}; t_{ks}]$, where h_{ks} is the representation obtained from TransE, RESCAL or GAT, and t_{ks} is the embedding of the node type of knowledge graph node. The edge between an entity node e and the corresponding knowledge node ks is represented as $e_{EKS} = [n_e; n_{ks}]$, and it is also further transformed into $v_{EKS}^{(1)}$ via a linear transformation function:





$$v_{EKS}^{(1)} = \mathbf{W}_{EKS} e_{EKS}, \tag{11}$$

where \mathbf{W}_{EKS} is a learnable parameter matrix.

Knowledge node representation based on description

In this paper, we use the following two methods to obtain knowledge node representation based on the entity description:

1. Doc2vec [15] (also called paragraph2vec), inspired by word2vec [16] proposed by Tomas Mikolov, which can transform a sentence or a short text into a corresponding low dimensional vector representation of fixed length.
2. An end-to-end neural network, as shown in Fig. 6, which are used to encode the description text of a given knowledge node, called EMB.

Similar to knowledge node ks , knowledge node kd based on description is represented as $n_{kd} = [h_{kd}; t_{kd}]$. The edge between kd and the corresponding entity node e is represented as $e_{EKD} = [n_e; n_{EKD}]$ and is further transformed by

$$v_{EKD}^{(1)} = \mathbf{W}_{EKD} e_{EKD}, \tag{12}$$

where \mathbf{W}_{EKD} is a learnable parameter matrix.

Inference

Following KEoG, with the help of the walk aggregation layer [17], a path between two entity nodes i and k of length $2l$ can be represented as

$$f(v_{ik}^{(l)}, v_{kj}^{(l)}) = \sigma(v_{ik}^{(l)} \odot (\mathbf{W} v_{kj}^{(l)})), \tag{13}$$

where σ is the sigmoid activation function, \odot is the element-wise multiplication, and $\mathbf{W} \in \mathbb{R}^{d_z \times d_z}$ is a learnable parameter matrix used to combine two short paths of

length l (path between i and j , and path between j and k) to generate one long path of length $2l$.

All paths from node i to node k are aggregated to form the representation of the edge from node i to node j of length $2l$ as follows:

$$v_{ij}^{(2l)} = \alpha v_{ij}^{(l)} + (1 - \alpha) \sum_{k \neq i, j} f(v_{ik}^{(l)}, v_{kj}^{(l)}), \tag{14}$$

where $\alpha \in [0, 1]$ is a linear interpolation scalar to control the contribution of edges of length l .

After obtaining the path representation of any entity pair of interest, we adopt the softmax function as classifier. Like in KEoG, both cross-entropy loss function and soft F-measure loss function are used as a part of the total loss function.

Experiments

Datasets

We conduct all experiments on the following two datasets:

- *Chemical-Disease Relation (CDR) dataset* is a dataset for document-level chemical-induced disease (CID) relation extraction, which is provided for the Bio-Creative V challenge [18]. It contains a training set of 500 abstracts, a development set of 500 abstracts and a test set of 500 abstracts from PubMed.
- *Chemical Reactions dataset (CHR) dataset* [9] is a dataset provided by the national text mining center (NaCTeM) of the school of computer science, University of Manchester. It contains 12,094 PubMed abstracts with their titles. Following Li et al. [11], we split the CHR dataset into a training set of 7,298 PubMed abstracts, a development set of 1,182 PubMed abstracts and a test set of 3,614 PubMed abstracts.

Table 1 Statistics of the CDR and CHR datasets

Dataset	CDR			CHR		
	#doc	#positive	#negative	#doc	#positive	#negative
Train	500	1038/284	4202/2746	7298	19,644/6438	33,860/20816
Dev	500	1012/246	4075/2478	1182	3186/1051	5535/3425
Test	500	1066/319	4138/2593	3614	9578/2962	16,151/9708

Table 2 Comparison results of SEoG and other different methods on the CDR and CHR test sets (%)

Dataset	Method	Overall			Intra	Inter
		P	R	F1	F1	F1
CDR	Gu et al. [3]	55.7	68.1	61.3	57.2	11.7
	Verga et al. [20]	55.6	70.8	62.1	–	–
	Nguyen and Verspoor [5]	57.0	68.6	62.3	–	–
	Sahu et al. [9]	52.8	66.0	58.6	–	–
	Christopoulou et al. [10]	62.1	65.2	63.6	68.2	50.9
	KEoG (node) [11]	65.4	71.2	68.2	71.8	58.3
	SEoG	64.5	75.5	69.6	73.4	59.9
CHR	CNN-RE [9]	81.2	87.3	84.1	–	–
	RNN-RE [9]	83.0	90.1	86.4	–	–
	Sahu et al. [9]	84.7	90.5	87.5	–	–
	KEoG (node) [11]	89.9	92.6	91.2	93.4	86.3
	SEoG	88.3	92.6	90.4	93.1	84.4

Bold highlight the highest result on a given dataset in our experiments

In this paper, MeSH¹ and Biochem4j² are used as the external knowledge of the CDR dataset and CHR dataset, respectively.

Experimental settings

Following our previous work, we first train all models on the training set, select the best hyper-parameters on the development set, then use the same hyper-parameters retrain on the combined set of the training set and development set, and finally report the results on the test set.

For the CDR dataset, hypernym filtering is also used to ensure that only relations between hyponym entities are kept, rather than relations between rough hypernym entities. For the CHR dataset, the entities that are not in Biochem4j are removed, and the self-relations, whose head entity and tail entity are same, are removed. The statistics of the two datasets are listed in Table 1, where “#” denotes the number of “*”, the numbers split by “/” are the total number of pairs and the number of inter-sentence pairs.

We start with EoG that only considers document structure, called SEoG, then investigate EoG that considers both document structure and external knowledge, i.e., SKEoG, and finally compare them with other state-of-the-art methods. For convenience, we use “SKEoG(KG + KD)”, such as “SKEoG(TransE + Doc2vec)”, to denote the SKEoG model using “KG” to obtain knowledge node representation based on knowledge graph and “KD” to obtain knowledge node representation based on entity description. All word embeddings are initialized by the pre-trained PubMed word embeddings [19]. Precision (P), recall (R) and F1-score (F1) are used as measures for model performance evaluation.

Results

We compare SEoG with other state-of-the-art methods, and the results are shown in Table 2. SEoG outperforms all other methods on the CDR and CHR test sets except KEoG(node) on the CHR dataset, which considers knowledge nodes based on the knowledge graph. To investigate the effect of the document structure presented in this paper, we further compare SEoG with its variant that does not consider chapter node, i.e., KEoG(node) without

¹ <ftp://ftp.nlm.nih.gov/online/mesh/2017/mesh2017.nt>.

² <http://biochem4j.org/>.

Table 3 Effect of different knowledge node representations base on knowledge graph on the CDR and CHR test sets (%)

Dataset	Method	Overall			Intra	Inter
		P	R	F1	F1	F1
CDR	SEoG	64.5	75.5	69.6	73.4	59.9
	SKEoG (TransE)	65.7	75.7	70.4	74.0	60.5
	SKEoG (RESCAL)	67.7	70.4	69.0	73.0	58.4
	SKEoG (GAT)	64.7	72.5	68.4	72.6	57.6
CHR	SEoG	88.3	92.6	90.4	93.1	84.4
	SKEoG (TransE)	89.9	91.4	90.6	93.2	85.0
	SKEoG (RESCAL)	90.6	92.1	91.4	93.6	86.4
	SKEoG (GAT)	87.5	93.6	90.5	93.1	84.6

Bold highlight the highest result on a given dataset in our experiments

using knowledge nodes. SEoG shows much better performance than that variant (overall F1-score: 69.6 vs 67.9 on the CDR dataset, 90.4 vs 89.0 on the CHR dataset). This result indicates that the introduced chapter node is effective.

Discussion

To investigate the effect of different knowledge node representations base on the knowledge graph for SKEoG, we compare SKEoG using different knowledge node representations with SEoG and present the results in Table 3. SKEoG using a specific knowledge node representation can achieve better performance than SEoG on both the CDR and CHR datasets. On the CDR dataset, SKEoG(TransE) achieves the highest F1-score of 70.4, while SKEoG(RESCAL) achieves the highest F1-score of 91.4 on the CHR dataset. It is a little strange that SKEoG(RESCAL) and SKEoG(GAT) even perform worse than SEoG in some cases. These results may be caused by the characteristics of different knowledge graphs used for knowledge node representation. There are only three types of relations in MeSH, and each entity has only one neighbor on average, that is, most of the existing entities and relations in MeSH can be effectively modeled by TransE, rather than RESCAL and GAT. In Biochem4j, there are nine types of relations, and each entity has three neighbors on average. The one-to-many, and many-to-one complex relations in biochem4j cannot be completely modeled by TransE, but can be modeled well by RESCAL.

Moreover, we also investigate the effect of different knowledge node representations base on entity description for SKEoG by comparing SKEoG(TransE) using different knowledge node representations with SKEoG(TransE) on the CDR dataset. Table 4 shows the comparison results. We can see that the knowledge node representation learned by EMB can bring performance improvement by an F1-score of 0.3%. However, the knowledge node representation learned by Doc2vec

Table 4 Effect of different knowledge node representations base on description on the CDR test sets (%)

Method	Overall			Intra	Inter
	P	R	F1	F1	F1
SKEoG (TransE)	65.7	75.7	70.4	74.0	60.5
SKEoG (TransE + Doc2vec)	66.2	72.2	69.1	73.7	56.9
SKEoG (TransE + EMB)	65.5	76.6	70.7	74.3	61.3

Bold highlight the highest result on a given dataset in our experiments

hurts the performance of SKEoG(TransE). The results indicate that we should be careful to utilize the knowledge based on entity description.

Conclusion

We extend our previous work KEoG to SKEoG, which takes full advantage of both hierarchical document structure and external knowledge for document-level medical RE. In this study, we comprehensively investigate different methods to obtain knowledge node representation based on knowledge graph and entity description. Experimental results on two public datasets show that both document structure and external knowledge are beneficial to medical RE in the EoG framework. In the case of external knowledge, selecting proper methods for knowledge node representation is also very important.

Abbreviations

RE: Relation extraction; CDR: Chemical-disease relation; CHR: Chemical reactions dataset; NLP: Natural language processing; MIL: Multi-instance learning; GNNs: Graph neural networks; EoG: Edge-oriented GNNs.

Acknowledgements

Not applicable.

About this supplement

This article has been published as part of BMC Medical Informatics and Decision Making Volume 21 Supplement 7 2021: Selected articles from the Fifth International Workshop on Semantics-Powered Data Mining and Analytics (SEFDA 2020). The full contents of the supplement are available at <https://bmcmmedinformdecismak.biomedcentral.com/articles/supplements/volume-21-supplement-7>.

Authors' contributions

TL, YX and BT design the experiments, TL and BT write the manuscript, and TL, YX, XW, QC and BT revised the manuscript. All authors read and approved the final manuscript.

Funding

Publication costs are funded by NSFCs (National Natural Science Foundations of China) (U1813215 and 61876052), National Key Research and Development Program of China (2017YFB0802204), Special Foundation for Technology Research Program of Guangdong Province (2015B010131010), Natural Science Foundation of Guangdong Province (2019A1515011158), Strategic Emerging Industry Development Special Funds of Shenzhen (JCYJ20180306172232154) and Innovation Fund of Harbin Institute of Technology (HIT.NSRIF.2017052). All the funders do not play any role in the design of the study, the collection, analysis, and interpretation of data, or in writing of the manuscript.

Availability of data and materials

The datasets analyzed during the current study are available in the <https://biocreative.bioinformatics.udel.edu/tasks/biocreative-v/track-3-cdr/> (accessed: Sep. 17, 2021) website and in the <http://www.nactem.ac.uk/CHR/> (Sep. 17, 2021) website.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent to publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Harbin Institute of Technology, Shenzhen, China. ²Peng Cheng Laboratory, Shenzhen, China.

Received: 14 December 2021 Accepted: 21 December 2021

Published: 30 December 2021

References

1. Socher R, Huval B, Manning CD, Ng AY. Semantic compositionality through recursive matrix-vector spaces. In: Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, Jeju Island, Korea, July 2012. p. 1201–11. Accessed 16 Oct 2020 [Online].
2. Zeng D, Liu K, Lai S, Zhou G, Zhao J. Relation classification via convolutional deep neural network. In: Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers, Dublin, Ireland, Aug 2014. p. 2335–44. Accessed 16 Oct 2020 [Online].
3. Gu J, Sun F, Qian L, Zhou G. Chemical-induced disease relation extraction via convolutional neural network. Database. 2017. <https://doi.org/10.1093/database/bax024>.
4. Verga P, Strubell E, McCallum A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. In: Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies, New Orleans, Louisiana, Jun 2018, vol. 1. p. 872–84. Accessed 26 Oct 2020 [Online].
5. Nguyen DQ, Verspoor K. Convolutional neural networks for chemical-disease relation extraction are improved with character-based word

6. embeddings. In: Proceedings of the BioNLP 2018 workshop, Melbourne, Australia, 2018. p. 129–36.
7. Zeng D, Liu K, Chen Y, Zhao J. Distant supervision for relation extraction via piecewise convolutional neural networks. In: Proceedings of the 2015 conference on empirical methods in natural language processing, Lisbon, Portugal, Sept 2015. p. 1753–62. Accessed 16 Oct 2020 [Online].
8. Jiang X, Wang Q, Li P, Wang B. Relation extraction with multi-instance multi-label convolutional neural networks. In: Proceedings of the 26th international conference on computational linguistics: technical papers, Osaka, Japan, Dec 2016. p. 1471–80. Accessed 16 Oct 2020 [Online].
9. Kipf T, Welling M. Semi-supervised classification with graph convolutional networks. Presented at the international conference on learning representations, 2017.
10. Sahu SK, Christopoulou F, Miwa M, Ananiadou S. Inter-sentence relation extraction with document-level graph convolutional neural network. In: Proceedings of the 57th annual meeting of the association for computational linguistics, Florence, Italy, 2019. p. 4309–16.
11. Christopoulou F, Miwa M, Ananiadou S. Connecting the dots: document-level neural relation extraction with edge-oriented graphs. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing, 2019. p. 4927–38.
12. Li T, Peng W, Chen Q, et al. KEoG: a knowledge-aware edge-oriented graph neural network for document-level relation extraction. In: 2020 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE, 2020. p. 1740–7.
13. Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data. In: Neural information processing systems (NIPS); 2013. p. 1–9.
14. Nickel M, Tresp V, Kriegel H-P. A three-way model for collective learning on multi-relational data. In: International conference on machine learning, 2011, vol. 11. p. 809–16.
15. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. Presented at the international conference on learning representations, 2018.
16. Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proceedings of the 31st international conference on machine learning, Beijing, China, 2014, vol. 32. p. II-1188–96.
17. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space, 2013.
18. Christopoulou F, Miwa M, Ananiadou S. A walk-based model on entity graphs for relation extraction. In: Proceedings of the 56th annual meeting of the association for computational linguistics, Melbourne, Australia, 2018. p. 81–8. Accessed 16 Oct 2020.
19. Li J, et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction. Database. 2016;2016:2016.
20. Chiu B, Crichton G, Korhonen A, Pyysalo S. How to train good word embeddings for biomedical NLP. In: Proceedings of the 15th workshop on biomedical natural language processing, Berlin, Germany, 2016. p. 166–74. <https://doi.org/10.18653/v1/W16-2922>.
21. Verga P, Strubell E, McCallum A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. In: Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies, vol. 1 (long papers), New Orleans, Louisiana, 2018. p. 872–84. <https://doi.org/10.18653/v1/N18-1080>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.