

RESEARCH

Open Access



Attention-based deep residual learning network for entity relation extraction in Chinese EMRs

Zhichang Zhang*, Tong Zhou, Yu Zhang and Yali Pang

From 4th China Health Information Processing Conference
Shenzhen, China. 1-2 December 2018

Abstract

Background: Electronic medical records (EMRs) contain a variety of valuable medical concepts and relations. The ability to recognize relations between medical concepts described in EMRs enables the automatic processing of clinical texts, resulting in an improved quality of health-related data analysis. Driven by the 2010 i2b2/VA Challenge Evaluation, the relation recognition problem in EMRs has been studied by many researchers to address this important aspect of EMR information extraction.

Methods: This paper proposes an Attention-Based Deep Residual Network (ResNet) model to recognize medical concept relations in Chinese EMRs.

Results: Our model achieves F_1 -score of 77.80% on the manually annotated Chinese EMRs corpus and outperforms the state-of-the-art approaches.

Conclusion: The residual network-based model can reduce the negative impact of corpus noise to parameter learning, and the combination of character position attention mechanism will enhance the identification features of different type of entities.

Keywords: Chinese electronic medical record, Entity relation extraction, Deep residual learning network, Attention mechanism

Background

EMR is used by medical staff to record texts, symbols, charts, graphics, data, and other digital information generated by HIS (hospital information system). With the tremendous growth of the adoption of EMR, various sources of clinical information (including demographics, diagnostic history, medications, laboratory test results, and vital signs) are becoming available, which has established EMR as a treasure trove for large-scale analysis of health data. Unstructured medical text in EMR is one kind of narrative data, including clinical notes, surgical records, discharge records, radiology reports, and pathology reports. For the convenience of narration, we

use EMR to represent unstructured EMR text in the following.

Identifying semantic relations existing among medical concepts in EMRs is of great importance to health-related various applications. These relations are hold between medical problems, tests, and treatments. Table 1 presents two examples of semantic relation, one of which is between medical concept e_1 ="cold" and e_2 ="fever" in sentence S_1 , and the other is between e_1 ="Head MRI" and e_2 ="lacunar infarction" in sentence S_2 .

On account of the importance of this subject, the 2010 i2b2/VA NLP challenge for clinical Records presented a relation classification task focused on assigning relation types between medical concepts in EMRs. Since then medical concept relation classification has being paid attention by more and more researchers.

*Correspondence: zhichangzhang@qq.com
College of Computer Science and Engineering, Northwest Normal University,
967 Anning East Road, 730070 Lanzhou, China



Table 1 Examples of the relations between medical entities

Sentence	Relation
S_1 : The patient has a cold, feels a fever and headache.	Disease causes symptoms (DCS)
S_2 : Head MRI shows lacunar infarction.	Test reveals the disease (TeRD)

In the traditional natural language processing (NLP) research, semantic relations between named entities can be used for many applications including knowledge graph construction, sentiment analysis, question answering, etc. [1], relation extraction or classification therefore has always been an important issue [2]. In previous open-domain entity relation extraction studies, researchers applied many different traditional machine learning models include Logistic Regression, SVM and CRF to recognize relations [3–7]. Li et al. used CRF model to reduce the space of possible label sequences and introducing long range features for relation recognition [8]. Mintz et al. put forward a remote monitoring relation classification method which could generate adequate training data by aligning text and knowledge base to solve the problem of lack of enough training data [9]. Socher et al. firstly employed recurrent neural network (RNN) on the task of relation extraction, while utilizing the syntactic structure information of sentences [10]. Miwa et al. proposed a neural network relation extraction architecture based on bidirectional LSTM and tree LSTM to encode entities and sentences simultaneously [11].

Drawing on these studies on open-domain relation extraction, similar task on EMRs was formally defined

in the 2010 i2b2/VA Challenge Evaluation [12]. Some researchers proposed various models for relation classification of EMRs. Bruijn et al. used SVM to train multiple classifiers to deal with different relation categories, and improved the effect of classification [13]. Rink et al. use external dictionaries to increase the effect of entity relationship recognition [14]. Fang et al. extracted the relations from relevant articles of Chinese herbal medicine based on manually designed rules and created a relation database [15]. Zhou et al. utilized a bootstrapping framework to extract relations from the medical articles and created a knowledge base [16]. Li et al. raised an electronic health records relation classification model based on CNN-LSTM [17]. Overall, the existing models mainly focus on English EMR texts, and on the other hand it still cannot deliver satisfactory recognition performance. Concerning the increasing availability of digitalized Chinese EMRs, this paper addresses the semantic relation identification problem among medical concepts in Chinese EMRs. We propose an attention mechanism based deep residual network model to classify the medical entity relations in Chinese EMRs. Experimental results performed on a manually labeled Chinese EMR corpus show that our model achieved better performance with F_1 -score of 77.80% compared with other methods.

Methods

Our model is based on a CNN architecture as shown Fig. 1. The model consists of five parts: vector representation layer, convolution layer, residual networks layer, position attention layer and output layer.

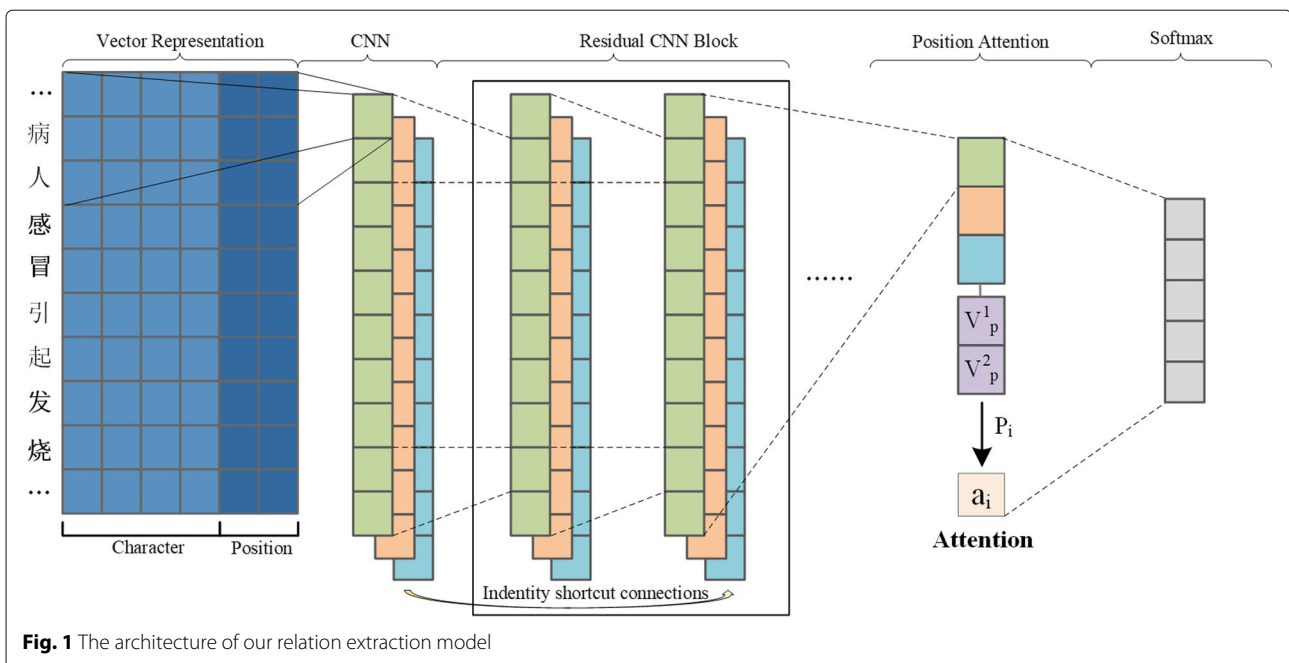


Fig. 1 The architecture of our relation extraction model



Fig. 2 An example of the relative distance between an entity and a character. The relative distance of a character to medical entity “感冒(cold)” and “发烧(fever)” are 2 and -2 respectively

Character embedding

Given a Chinese sentence $S = (c_1, c_2, \dots, c_n)$ which contains two entities e_1 and e_2 . Each character c_i will be mapped to a low-dimensional dense vector $V_i = (V_w^i, V_p^i)$, in which V_w^i represents the character vector and V_p^i is the vector of character position in the sentence. The character embedding initialized with vector which is pre-trained by word2vec, and d_w is the dimension of character vector.

Position embedding

Position embedding V_p^i is also a low-dimensional vector of character position in the sentence, which can combine the relative positions (see Fig. 2) of the current character to the first entity e_1 as well as the second entity e_2 . Each relative position corresponds to a position embedding $V_p^i \in R^{d_p}$, d_p is the dimension of position embedding.

The vector $V_i \in R^{d_v}$ is concatenation of character vector V_w^i and two position vectors, where $d_v = d_w + 2d_p$.

Convolution

Convolution is to extract the effective local feature information from characters and their corresponding contexts. The V_j is a vector which corresponds the j -th character in the sentence $S = (V_1, V_2, \dots, V_n)$, here n is the sentence length. We use filter $W \in R^{h \times d_v}$ to extract local features from the sentence S . A feature c_j is generated from a window of character V_{j+h-1} by

$$c_j = f(W \cdot V_{j+h-1} + b), \tag{1}$$

where b is a bias terms and f is a non-linear function. We apply dropout layer in convolution to prevent data from overfitting.

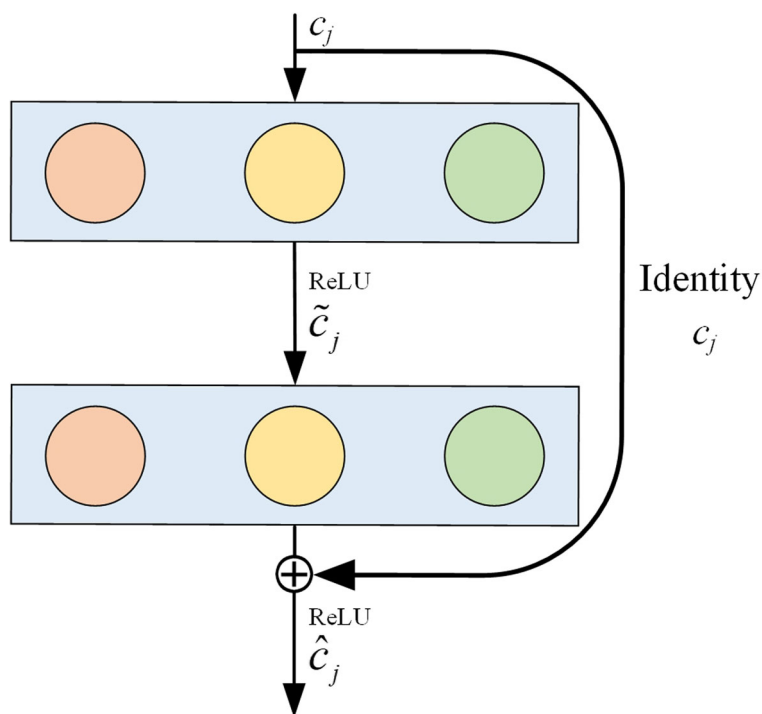


Fig. 3 The residual convolution block

Table 2 The relation number of every fine-grained category in the corpus

Fine-grained category	Training	Develop	Test
TrID	368	260	193
TrWD	229	149	102
TrCD	423	284	265
TrAD	4706	2096	1581
TrNAD	110	35	41
TrIS	1351	371	427
TrWS	598	152	163
TrCS	118	41	57
TrAS	2093	1083	1154
TrNAS	98	36	21
TeRD	1770	498	603
TeCD	85	23	27
TeRS	13963	8180	4998
TeBS	1492	214	388
DCS	5465	2677	3251

Residual networks

Residual learning connects low-level to high-level representations directly and solves the vanishing gradient problem, we superimposed the identity mapping function on a network. In our model, each residual convolution block (see Fig. 3) has two convolutional layers, each one followed by a ReLU activation, we use shortcut connection between each of the residue convolution block $W_1, W_2 \in R^{h \times 1}$ are two convolution filters, where h is convolution kernel size. The first convolutional layer is

$$\tilde{c}_j = f(W_1 \cdot c_{j:j+h-1} + b_1), \quad (2)$$

and the second is

$$\hat{c}_j = f(W_2 \cdot \tilde{c}_{j:j+h-1} + b_2 + c_j), \quad (3)$$

here b_1, b_2 are bias terms. The residual convolution block output is the vector \hat{c}_j . This block will be multiply concatenated in our architecture by a shortcut connection.

Table 3 Hyper parameters of the residual neural network

Parameter	Description	Value
d_w	Dimension of word embedding	100
d_p	Dimension of position embedding	5
k	Window size	3
m	Number of filters	128
B	Batch size	50
λ	Learning rate	0.01
p	The ration of dropout	0.3

Table 4 Comparison of overall relation classification result of different model

Model	Precision	Recall	F_1 -score
SVM	65.24	55.26	59.84
CNN-Max [18]	55.34	50.84	52.99
LSTM-Max	69.27	70.51	69.88
BLSTM-Attention [19]	74.12	66.95	70.35
ResNet-BLSTM	78.81	71.24	74.83
ResNet-Max [20]	65.24	67.45	66.33
ResNet-PAtt	76.48	79.16	77.80

Position attention

Recently attention mechanism has been widely used in machine learning, and great achievements have been made in various NLP problems. In this paper, we use the position attention to enhance relation extraction ability. Firstly, we carry the max-pooling operation on the residual learning result. Secondly, as shown in Fig. 1, we concatenate the max-pooling results with the position embedding of entity. Finally, we use the attention mechanism to balance the weight to the sentence.

$$S_i = \sum_i \alpha_i \times P_i, \quad (4)$$

where α_i represents the attention weight. P_i is a result which concatenates the max-pooling results with the position embedding of entity. Finally, we use the softmax function to normalize and output entity relation probability.

Table 5 Classification performance of our model on every fine-grained relation category.

Relation	Precision	Recall	F_1 -score
TrID	46.84	42.61	44.62
TrWD	41.35	40.12	40.73
TrCD	47.73	45.33	46.50
TrAD	72.42	68.48	70.39
TrNAD	45.88	46.18	46.03
TrIS	57.42	55.67	56.53
TrWS	50.21	48.28	49.23
TrCS	38.36	42.64	40.39
TrAS	61.38	80.55	69.67
TrNAS	35.76	36.51	36.13
TeRD	74.81	72.51	73.64
TeCD	41.55	39.35	40.42
TeRS	83.57	81.68	83.61
TeBS	56.72	58.31	57.50
DCS	76.86	74.53	75.68

$$\alpha_i = \frac{\exp(e_j^i)}{\sum_k \exp(e_j^k)} \tag{5}$$

Results

Dataset and evaluation metrics

On the basis of reference to medical semantic relation annotation specification of 2010 i2b2/VA Challenge, we established our own relation annotation specification of Chinese EMRs, in which semantic relations between medical concepts fall into five coarse-grained categories and fifteen fine-grained categories. All of relation category are detailed as follows.

Coarse-grained category 1: Treatment -Disease Relation. This category contains five fine-grained categories, including TrID (Treatment improves the disease), TrWD (Treatment worsens the disease), TrCD (Treatment causes the disease), TrAD (Treatment is administered for the disease), and TrNAD (Treatment is not administered because of the disease).

Coarse-grained category 2: Treatment -Symptoms Relation. This category also contains five fine-grained categories, including TrIS (Treatment improves the symptoms), TrWS (Treatment worsens the symptoms), TrCS (Treatment causes the symptoms), TrAS (Treatment is administered for the symptoms), and TrNAS (Treatment is not administered because of the symptoms).

Coarse-grained category 3: Test-Disease Relation. This category contains two fine-grained categories, including TeRD (Test reveals the disease) and TeCD (Test conducted to investigate the disease).

Coarse-grained category 4: Test-Symptoms Relation. This category also contains two fine-grained categories, including TeRS (Test reveals the symptoms) and TeBS (Test based on symptoms).

Coarse-grained category 5: Disease-Symptoms Relation. This category contains only one fine-grained category named as DCS (Disease causes symptoms).

According to our specification, we manually annotated 3000 de-identified Chinese EMR texts from different clinical departments of a grade-A hospital of second class in Gansu Province, China. 2000 medical texts are selected as training data, 500 medical texts as develop data, and 500 medical texts for test while evaluating our method on this dataset. The relation numbers of every fine-grained category in this dataset are given in Table 2. *Precision, Recall* and *F₁-score* are used as evaluation metrics.

Models and parameters

We carry out the experiments to compare the performance of our model with others described in the following.

CNN-Max: This model was used by Sahu, et al. [18], which encoded the sentence vectors with CNN, and outputted the results after max-pooling and softmax function.

BLSTM-Attention: This model was proposed by Li, et al. It mainly consists of bidirectional LSTM and attention mechanism [19].

ResNet-Max: This model was proposed by Huang, et al. Compared with our model, this model did not combined attention mechanism [20].

ResNet-BLSTM: The basic framework of the method is close to our model. The difference between this one with ours is that this model combine the residual network with Bi-LSTM.

ResNet-Patt: This is the model presented in this paper. Table 3 gives the chosen hyper-parameters for all experiments. We tune the hyper-parameters on the development set by random search. We try to share as many hyper-parameters as possible in experiments.

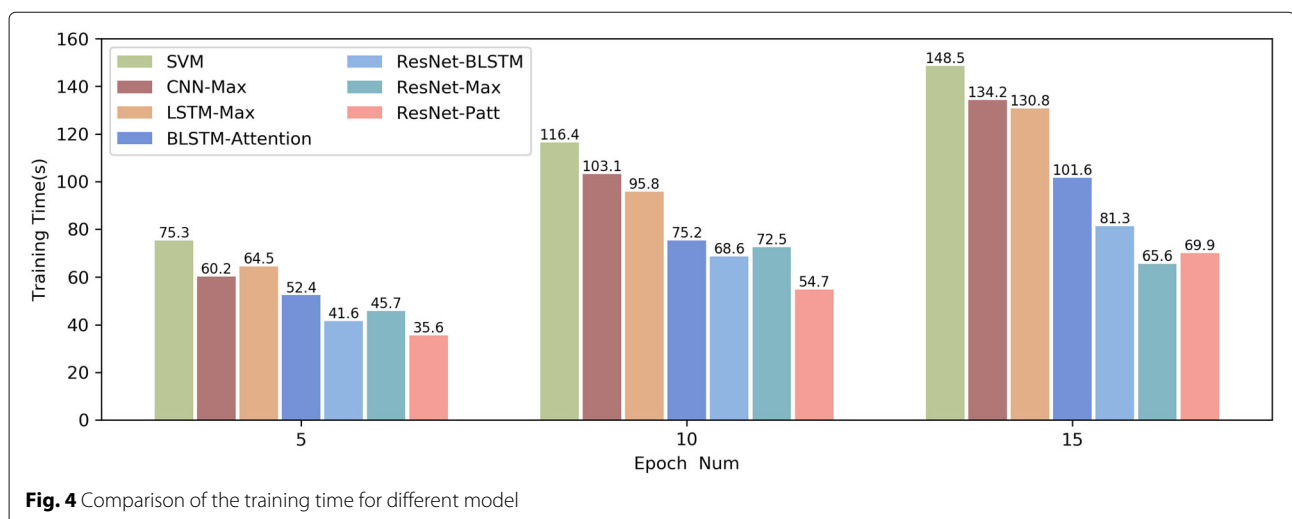


Fig. 4 Comparison of the training time for different model

Table 6 Comparison of F_1 -score for each model on every fine-grained relation category

Category	SVM	CNN-Max	LSTM-Max	BLSTM-Attention	ResNet-BLSTM	ResNet-Max	ResNet-PAtt
TrID	20.06	29.68	36.78	40.38	42.67	35.42	44.62
TrWD	19.35	28.34	25.35	35.21	33.21	30.43	40.73
TrCD	28.52	28.02	39.41	46.32	48.57	42.6	46.50
TrAD	63.21	43.52	58.31	71.65	68.33	64.54	70.39
TrNAD	12.36	22.46	18.24	36.69	37.26	35.42	46.03
TrIS	57.24	48.52	49.31	54.37	52.44	52.31	56.53
TrWS	36.41	49.51	37.53	46.21	48.18	42.43	49.23
TrCS	39.04	39.53	41.52	39.46	40.93	39.5	40.39
TrAS	60.26	58.33	62.34	66.83	72.36	61.36	69.67
TrNAS	13.54	14.39	14.56	28.67	30.31	24.67	36.13
TeRD	62.35	60.27	62.24	71.36	74.22	69.96	73.64
TeCD	12.34	16.52	18.36	37.23	32.88	31.48	40.42
TeRS	82.53	71.26	75.34	80.44	81.63	78.45	83.61
TeBS	48.42	46.34	47.21	58.64	57.94	51.20	57.50
DCS	64.28	65.31	65.67	74.24	73.55	70.69	75.68

Experimental results

Table 4 shows the overall classification performance of different models on our evaluation corpus. It can be seen that our method ResNet-PAtt is better than other methods in F_1 -score while precision, recall and F_1 -score reaches 79.16 and 77.80% respectively. Of all other methods, the model ResNet-BLSTM achieves the best performance on F_1 -score, and our model improves 2.97% F_1 -score compared with it, then our method is more effective. In addition, we can find that overall the residual network based methods are better than other relation extraction methods.

Discussion

The reasons our model achieves best performance maybe owe to that the residual network-based model could reduce the negative impact of corpus noise to parameter learning, and the combination of character position attention mechanism could enhance the identification information of different type of entities. Table 5 gives the classification performance of our model on every fine-grained relation category. As can be seen from these data, our model performs best on relation category TeRS and worst on category TrNAS, which shows that it is more difficult to recognize category TrNAS correctly. We also evaluate the training time of different models. Figure 4 shows that the consumed times by these models while epoch is set as 5, 10 and 20 respectively. Overall, our model takes the shortest time to complete parameter training, and the traditional machine learning method SVM takes the longest time to train.

Table 6 is comparison of F_1 -score for each model on every fine-grained relation category. The model has better classification performance and faster response speed.

Conclusions

In this paper, we propose a deep residual network model based on the attention mechanism to classify the relation of entity pairs in Chinese EMRs. The method reduced the influence of data noise on the model training, and enhance entity discrimination feature with position attention mechanism so that the entity information can be combined effectively in the relation extraction. Experimental results show that the model reached 77.80% F_1 -score value, and significantly improved the classification performance of the few instance categories. At present, most relation classifications are based on entity recognition tasks and need to specify the entity in the sentence. In the future, we will study the joint extraction of entity and entity relation to further improve the efficiency of entity and entity relation recognition simultaneously.

Acknowledgements

We thank the anonymous reviewers for their insightful comments.

Funding

Publication of the article is supported by the National Natural Science Foundation of China (NO. 61762081, No. 61662067, No. 61662068) and the Key Research and Development Project of Gansu Provincial (No. 2017GS10781).

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable requests.

About this supplement

This article has been published as part of *BMC Medical Informatics and Decision Making Volume 19 Supplement 2, 2019: Proceedings from the 4th China Health Information Processing Conference (CHIP 2018)*. The full contents of the supplement are available online at URL: <https://bmcmidinformatdecismak.biomedcentral.com/articles/supplements/volume-19-supplement-2>.

Authors' contributions

ZZ and TZ had part in conceiving the study, and have substantially contributed in writing and revising the manuscript. YZ and YP analyzed the

entity relation data regarding the Chinese electronic medical records, and was a major contributor in establishing annotation specification. All authors read and reviewed the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Published: 9 April 2019

References

- Lin Y, Liu Z, Sun M. Neural relation extraction with multi-lingual attention. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Vancouver: Association for Computational Linguistics; 2017. p. 34–43.
- Zheng S, Wang F, Bao H, et al. Joint extraction of entities and relations based on a novel tagging scheme. 2017. arXiv preprint arXiv:1706.05075 [cs.CL]. <https://arxiv.org/abs/1706.05075>.
- Takamatsu S, Sato I, Nakagawa H. Reducing wrong labels in distant supervision for relation extraction. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Jeju Island: Association for Computational Linguistics; 2012. p. 721–9.
- Miller S, Fox H, Ramshaw L, Weischedel R. A novel use of statistical parsing to extract information from text. In: Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference. Seattle: Association for Computational Linguistics; 2000. p. 226–33.
- Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In: Proceedings of the ACL 2004 on Interactive poster and demonstration sessions. Barcelona: Association for Computational Linguistics; 2004. p. 22–25.
- Culotta A, McCallum A, Betz J. Integrating probabilistic extraction models and data mining to discover relations and patterns in text. In: Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics; 2006. p. 296–303.
- Wang T, Li Y, Bontcheva K, Cunningham H, Wang J. Automatic extraction of hierarchical relations from text. Eur Semantic Web Conf. 2006;4011: 215–29. Lecture Notes in Computer Science.
- Li Y, Jiang J, Chieu H, Chai K. Extracting relation descriptors with conditional random fields. In: Proceedings of 5th International Joint Conference on Natural Language Processing. Chiang Mai: Asian Federation of Natural Language Processing; 2011. p. 392–400.
- Mintz M, Bills S, Snow R, Jurafsky D. Distant supervision for relation extraction without labeled data. 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. Suntec: Association for Computational Linguistics; 2009. p. 1003–11.
- Socher R, Huval B, Manning C, Ng A. 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: Association for Computational Linguistics; 2012. p. 1201–11.
- Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures. 2016. arXiv preprint arXiv:1601.00770 [cs.CL]. <https://arxiv.org/abs/1601.00770>.
- Uzuner Ö, South B, Shen S, DuVall S. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc. 2011;18:552–6.
- De B, Cherry C, Kiritchenko S, Martin J, Zhu X. Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010. J Am Med Inform Assoc. 2011;18:557–62.
- Rink B, Harabagiu S, Roberts K. Automatic extraction of relations between medical concepts in clinical texts. J Am Med Inform Assoc. 2011;18:594–600.
- Fang Y, Huang H, Chen H, Juan H. TCMGeneDIT: a database for associated traditional Chinese medicine, gene and disease information using text mining. BMC Complement Alternat Med. 2008;8:58–9.
- Zhou X, Liu B, Wu Z, Feng Y. Integrative mining of traditional Chinese medicine literature and MEDLINE for functional gene networks. Artif Intell Med. 2007;41:87–104.
- Li L, Zhao S. Relation Classification in Electronic Health Records via CNN-LSTM Model. In: Proceedings of the 3th China Health Information Processing Conference. Shenzhen: Chinese Information Processing Society of China; 2017. p. 40–6.
- Sahu S, Anand A, Oruganty K, Gattu M. Relation extraction from clinical texts using domain invariant convolutional neural network. 2016. arXiv preprint arXiv:1606.09370 [cs.CL]. <https://arxiv.org/abs/1606.09370>.
- Li L, Nie Y, Han W, Huang J. A Multi-attention-Based Bidirectional Long Short-Term Memory Network for Relation Extraction. Int Conf Neural Inf Process. 2017;10638:216–27. Lecture Notes in Computer Science.
- Huang Y, Wang W. Deep Residual Learning for Weakly-Supervised Relation Extraction. 2017. arXiv preprint arXiv:1707.08866 [cs.CL]. <https://arxiv.org/abs/1707.08866>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

