Proceedings of the 4th International Conference on Computing and Data Science (CONF-CDS 2022)
DOI: 10.54254/2755-2721/2/20220560

A novel information extraction model integrating multi-granularity global information for medical dialogue

Kehang Song

National University of Singapore, Singapore

kehangsong@gmail.com

Abstract. Electronic medical records (EMRs) are one of the ways to help doctors effectively manage and analyze patient medical records. These EMRs can not only help doctors save a lot of time in analyzing medical records, but also reduce the hospital's demand for doctors and reduce the hospital's expenditure cost. Therefore, we propose a novel information extraction model integrating multi-granularity global information to efficiently extract information about a ptient's physical condition in a doctor-patient dialogue. Experimental results show that our model achieves better results compared to the baseline model, indicating the effectiveness of the model.

Keywords: information extraction, medical dialogue, deep learning.

1. Introduction

EMRs (Electric medical records) are commonly used in modern medical system as a replacement of traditional way that is used to record the patients' information. Doctors are getting used to EMRs as they help them to easily navigate through the medical history, the treatments and the medicine that patients used before. But it is a heavy work for doctors to fill in the EMRs when it comes to a large amount of patients. According to statistics [1], it takes roughly 2 hours to finish the EMRs writing work for each doctor-patient interrogation, which turns out to be a heavy burden for the medical experts, and our goal in this paper is proposing a novel method to automate the process mentioned above.

Automatic generation of EMRs has two technical directions. One is end to end method [2,3,4] which is to generate EMRs given the dialogue between doctors and patients. But this method is data-greedy and the error and loss is uncontrollable during the generation of models. The other direction [5,6,7] is based on a pipeline framework which is to extract useful information from conversation between doctors and patients and transfer it into EMRs according to natural language generation techniques.

In this paper, we primarily focus on the second method. There are several representative works in recent years for doctor-patient information extraction. [8] defined medical information extraction task. This article open-sourced a dataset that includes 186 symptoms and their related states, along with a baseline model for the extraction of related information. [9] modified the above model, they proposed a sequence labeling method on medical data to perform named entity recognition. This model brought in global attention block and a statistic graph to improve the model's performance. [5] introduced a new coarse-grained dataset that contains 4 major groups which are symptom, surgery, test and others and each group also includes its corresponding states. The author proposed a MIE model structure to extract

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2/20220560

useful medical information. MIE model used attention mechanism [10,11] and considered the information between multiple rounds of dialogue which increases the accuracy of dialogue state extraction.

However, MIE model separates the whole dialogue into different windows with fixed size and the prediction only considered the contextual information within current window. This makes the model unable to utilize global conversation information, which is vital for the model to predict the state from the whole dialogue. Therefore, in this paper, we present a new model based on MIE model. Our model also uses a sliding window dialogue label prediction method. But different with MIE, the global information is incorporated into our model from different aspects, including fine-grained co-occurring word information and coarse-grained lower window information. As a result of involving several extra information, we designed a different attention mechanism to calculate the weight of information to filter out noise information. This makes our model reached out 59.85% on precision, 63.60% on recall and 61.67% on f1 score in the full-labels mode. Compared to the best performance of MIE model, our model increased 2.30% on recall and 0.58% on f1 score. This verifies the effectiveness of our model in the full labels mode.

In section 2, we elaborate on the representative works in the field of dialogue information extraction. In section 3, we introduce our model structure. In section 4, we first explain the dataset we use, as well as the settings and parameter details during the experiment. After that, we perform detailed analysis on our results. In section 5, we perform some auxiliary experiments to analyze the performance and characteristics of the model from different perspectives. We summarize and conclude the whole article in section 6.

2. Related work

Medical information extraction is a long-term task for NLP community. The earliest dataset that has huge influence on this task is the dataset provided in 2010 i2b2 challenge [12]. But the direction of extracting medical information from dialogues has only emerged in recent years. In 2018, Finley [7] introduced a five-step approach including knowledge extraction, which uses dictionary, regular expression and other supervised machine learning techniques.

The most similar work of ours is Zhang's MIE model [5] which extracts the medical information from doctor-patient dialogues that are annotated in a window-sliding style. MIE model extracts medical information from dialogues in a single window instead of whole dialogues and uses deep matching architecture, taking dialogue turn-interaction into account. Our model differs from MIE in one aspect: we extract the medical information not only from the current window but also from the whole dialogues.

3. Model design

In this part, we will elaborate on our proposed model. Specifically, our model has 4 blocks, (1) encoder block converts raw input into corresponding contextual representation, (2) fine-grained global information aggregation block fuses all cooccurring word information for the current word in the full text, (3) coarse-grained global information aggregation block fuses the most informative window information into the current window, (4) MIE predictor uses a binary classifier to make a prediction on the current label.

3.1. Encoder

For a given window $X = \{x_1, x_2, \dots, x_T\}$ and label L, we first use GloVe word embeddings to pretrain the model and convert certain window and the current label into semantic vector representation.

$$X_E = \{e_1, e_2, \dots, e_T\} = \mathrm{GloVe}\{x_1, x_2, \dots, x_T\}$$
$$L_E = \mathrm{GloVe}\{L\} \quad (1)$$

where T represents the total number of words in window, in other words, length of sentence. After that, we use a BiLSTM [13] network to encode every token in X_E and L_E. Thus, we get the contextual representation of windows and labels.

$$h_i = BiLSTM(e_i)$$
$$g_i = BiLSTM(l_i) \quad (2)$$

DOI: 10.54254/2755-2721/2/20220560

where h_i represents the ith word's contextual representation in the window and g_i represents the ith word' representation in the labels

3.2. Fine-grained global information aggregation

As every label's state is directly related to all the co-occurrence words in the whole dialogue, it is necessary for us to utilize all directly related word information in the global dialogue, which is co-occurrence word information, to give some global word-level extra features for each word in the current window.

We bring in MemoryNet to achieve it [14,15]. Specifically, for every word h_i in the current window, we define a Memory set $M = \{(k_1, v_1), (k_2, v_2), \dots, (k_p, v_p)\}$ to store its corresponding co-occurrence word information where k_i represents the ith co-occurrence word's word vector matrix and v_i represents the hidden layer result h_i of ith co-occurrence word's LSTM.

For every token h_i in the current window, to aggregate Fine-grained Global information, we use h_i as the query of attention to calculate every pair's attention matrix in the Memory Set as follows:

$$a_{ij} = \sigma(h_i \cdot k_j) \tag{3}$$

After that, we apply softmax [16] on it and get the weight w_{ij}:

$$w_{ij} = \frac{\exp(a_{ij})}{\sum_{k=1}^{p} a_{ik}} \tag{4}$$

Accordingly, we calculate Fine-grained Global information as follows:

$$g_i^{fine} = \sum_{j=1}^{p} a_{ij} \cdot w_{ij} \tag{5}$$

Then we will fuse the initial token hidden representation h_I and Fine-grained Global information g_i^{fine}:

$$G_i = \lambda \cdot h_i + (1 - \lambda) \cdot g_i^{fine} \tag{6}$$

where λ is a super parameter used to balance hidden state $h_i$and global information g_i^{fine}.

3.3. Coarse-grained global information aggregation

Because state is not only determined by the current dialog window, but also often affected by the information of the window below, intuitively, we introduce the lower window information to improve the accuracy of information prediction in the current window. We used a dynamic Attention mechanism to achieve it. The reason why it is called dynamic is because number of windows below will change with the current window. Specifically, suppose the current window is represented as g_i and the lower window is represented as $\{g_{i+1}, \dots g_t\}$ where t is the number of windows. We first calculate the attention weight of each lower window to the current window as follows

$$g_{ij} = \sigma(g_i \cdot g_j)$$
$$p_{ij} = \frac{exp(g_{ij})}{\sum_{k=i+1}^{t} g_{ik}} \tag{7}$$

where p_{ij} refers to the attention weight score of the j-th window to the current window g_i. After that, we select the lower window with the highest attention weight as our global window information and splice it into the representation of the current window, thus forming a new current window that incorporates the global window information as $\tilde{g}_i$:

$$g_{global} = \max \arg \{p_{ij}\}$$
$$\tilde{g}_i = [g_I; g_{global}] \tag{8}$$

where ; represents concatenation.

3.4. Classifier

We use the output $\tilde{g}_I$ in the previous section as the input feature of the classifier for binary classification. Specifically, given a label I, we use a sigmoid classifier to classify it:

$$Y^{pred} = sigmoid(\tilde{g}) \tag{9}$$

DOI: 10.54254/2755-2721/2/20220560

If y^{pred} is larger than 0.5, it indicates that the label I is expressed in the current dialog window. If y^{pred} is smaller than or equal to 0.5, it indicates that the label I is not expressed in the current dialog window.

4. Experiment

4.1. Dataset

The dataset we used is same with MIE. This dataset is doctor-patient dialogues collected from a Chinese medical consultation website, Chunyu-Doctor and all the dialogues are selected from cardiology topic consultations as a result of more inquiries than other topics. This dataset uses a window-to-information annotation method which defines four main categories and frequent items for each category. As finishing reading the whole dialogues at once can't give correct labels for all the information mentioned in the dialogues, dividing them into pieces which consist of several turns of dialogue is helpful for the labeling. The sliding window size is set to 5 which is a proper size to include appropriate amount of information.

4.2. Model details

For the fairness of the experiment, our experimental setup is basically consistent with MIE. We use Adam [17] to optimize our model during training process as other work did [18,19]. We use dropout [20] and L2 weight regularization [21] to alleviate the overfitting problem. The specific parameters can be found in Table 1.

Table 1. Model parameters.

Parameters	Quantity
Word embedding dimension	300
LSTM hidden size	400
Drop out	0.2
MemoryNet maximum size	10

4.3. Results

As shown in Table 2, medical label only represents we only test the results in the scope of medical labels, without considering their states. Full-labels denotes we both take medical label and their states into account. For example, if a ground truth is "heart failure: positive", while the predicted label is "heart failure: negative". It will be counted as correct in medical label only, as incorrect in the full-labels.

Among all the results, Plain-classifier undoubtedly achieved the lowest model effect. This proved the fact that simple LSTM network has poor fitting ability for the extraction of medical dialogue information. But in MIE model, MIE-single representation model only considers the semantic information between a dialogue sentence, in which case it is hard to capture the state information that rely on multiple sentences and therefore we only get 83.46% and 61.09% F1 score, respectively. But for our model, we achieved the result of 86.24% and 61.67% F1 score which is an improvement of 2.78% and 0.58% compared to the baseline. This also verifies that in MIE dataset, a large part of the label information depends on the information of the lower window, and our model can introduce the information of these lower windows into the current window multiple dimensions. This is the key to the performance improvement of our model.

Table 2. Experimental results.

	Medical labels only			Full-labels (with state)		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Plain-classifer	81.75	73.76	77.29	59.98	52.65	56.08
MIE-single	87.02	80.02	83.46	**61.15**	61.30	61.09
MIE-multi	85.32	80.48	82.83	60.30	60.78	60.54

DOI: 10.54254/2755-2721/2/20220560

Table 2. (continued).

The proposed model	**88.48**	**84.10**	**86.24**	59.85	**63.60**	**61.67**

5. Discussion

5.1. Ablation study

Because our model includes several key modules, including coarse-grained global information aggregation, Fine-grained global information aggregation, to verify the importance of each module, we test its performance separately. Raw model represents that we only use a simple LSTM-based encoder to perform our task. According to Table 3, it can be seen that by adding coarse-grained global information and Fine-grained global information, our model improves the F1 score of 9.62% and 15.39% F1 on medical labels only, and 5.36%, 6.91% on full labels, which proves that our proposed module can improve the model performance for MIE tasks. Take an example for illustration, under the medical labels only settings, coarse-grained global information improves the recall of 5.4%, the precision of 13.17%, and the f1 of 9.62%. This is because it introduces the most valuable window information in the lower window into the judgment of the current window, which is equivalent to providing global sentence-level semantic information for the classification of the current medical entity, and this semantic information is very important for state judgment. At the same time, Fine-grained global information improves the recall of 12.57%, the precision of 18.30%, and the f1 of 15.39%. This is because the co-occurrence words in the full dialogue have a certain auxiliary role in the extraction of the current word and the negative and positive judgement. When the positiveness changes in the windows below, the co-occurring words of this window are able to capture this change in semantic information through the context, so as to reflect this semantic change in the current window to correctly judge the state.

Table 3. Ablation study results.

	Medical labels only			**Full-labels (with state)**		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Raw model	57.01	66.13	61.23	48.49	50.34	49.40
Add coarse-grained info	70.18	71.53	70.85	54.51	55.02	54.76
Add fine-grained info	88.48	84.10	86.24	59.85	63.60	61.67

5.2. Case study

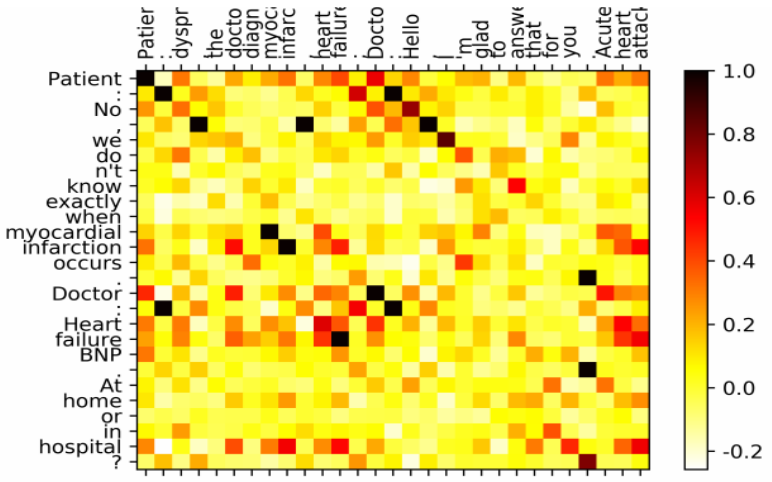

Figure 1. Case study: attention heatmap o between current window and the global window.

DOI: 10.54254/2755-2721/2/20220560

Figure 1 indicates the cross-window attention heatmap between the current window and the global window gained from the coarse-grained global information aggregation module. This attention heatmap can describe the importance of a word by the depth of colour, that is, given the words in the window below, the importance scores of these words for each word in the current window. In this figure, our vertical axis represents the dialog information of the current window, and the horizontal axis represents the dialog information of the lower window. We can see that for the medical entity information "heart failure" in the original text, words such as "myocardial", "infarction", and "occur" in the corresponding words in the lower window have higher weights. From the medical meaning of these words, "myocardial", "infarction" and "heart failure" are all positively related and from purely semantic point of view, "occur" also verifies its positive characteristics. These words with higher weights will also occupy more feature space and dimensions in subsequent classifiers, thus assisting the final decision.

6. Conclusion

In this work, we propose a novel information extraction model for the MIE task. Our model can be divided into four parts, the encoder is used to encode the original input, the fine-grained information aggregation introduces the co-occurrence word information, and the coarse-grained information aggregation introduces the lower window information, which all have positive effect towards the judgement of final classifier. The experimental results also verify the performance of our model. Specifically, compared to the baseline models, our model improves the precision to 88.48%, recall to 84.10% and F1 score to 86.24%. At the same time, we conducted ablation study and case study on our model, which shows the reason why the model is effective from another perspective. But our model also has some problems, such as poor performance in the judgment of rare categories, which will be improved in future work.

References

[1] Christine Sinsky, Lacey Colligan, Ling Li, Mirela Prgomet, Sam Reynolds, Lindsey Goeders, Johanna Westbrook, Michael Tutty, and George Blike. 2016. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Annals of internal medicine, 165(11):753–760.

[2] Cao J, Ananiadou S. GenerativeRE: Incorporating a Novel Copy Mechanism and Pretrained Model for Joint Entity and Relation Extraction [C]//Findings of the Association for Computational Linguistics: EMNLP 2021. 2021: 2119-2126.

[3] Cao J, van Veen E M, Peek N, et al. EPICURE: Ensemble Pretrained Models for Extracting Cancer Mutations from Literature [C]//2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2021: 461-467.

[4] Tran D, Bourdev L, Fergus R, et al. Deep end2end voxel2voxel prediction [C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2016: 17-24.

[5] Zhang Y, Jiang Z, Zhang T, et al. MIE: A medical information extractor towards medical dialogues [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 6460-6469.

[6] Cao J, Xiang Y, Zhang Y, et al. CONNER: A Cascade Count and Measurement Extraction Tool for Scientific Discourse [C]//Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021). 2021: 1239-1244.

[7] Gregory Finley, Erik Edwards, Amanda Robinson, Michael Brenndoerfer, Najmeh Sadoughi, James Fone, Nico Axtmann, Mark Miller, and David Suendermann-Oeft. 2018. An automated medical scribe for documenting clinical encounters. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pages 11–15.

[8] Nan Du, Kai Chen, Anjuli Kannan, Linh Tran, Yuhui Chen, and Izhak Shafran. 2019. Extracting symptoms and their status from clinical conversations. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 915– 925.

DOI: 10.54254/2755-2721/2/20220560

[9] Lin X, He X, Chen Q, et al. Enhancing dialogue symptom diagnosis with global attention and symptom graph [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 5033-5042.
[10] Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network [C]//The world wide web conference. 2019: 2022-2032.
[11] Cao J, Wang C. Social media text generation based on neural network model [C]//Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence. 2018: 58-61.
[12] Uzuner Ö, South B R, Shen S, et al. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text [J]. Journal of the American Medical Informatics Association, 2011, 18(5): 552-556.
[13] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural computation, 1997, 9(8): 1735-1780.
[14] Tai Y, Yang J, Liu X, et al. Memnet: A persistent memory network for image restoration [C]//Proceedings of the IEEE international conference on computer vision. 2017: 4539-4547.
[15] Cheng J, Dong L, Lapata M. Long short-term memory-networks for machine reading [J]. arXiv preprint arXiv:1601.06733, 2016.
[16] Sun S, Zhang Z, Huang B C, et al. Sparse-softmax: A Simpler and Faster Alternative Softmax Transformation [J]. arXiv preprint arXiv:2112.12433, 2021.
[17] Kingma D P, Ba J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv:1412.6980, 2014.
[18] Nie Y, Tian Y, Song Y, et al. Improving named entity recognition with attentive ensemble of syntactic information [J]. arXiv preprint arXiv:2010.15466, 2020.
[19] Nie Y, Tian Y, Wan X, et al. Named entity recognition for social media texts with semantic augmentation [J]. arXiv preprint arXiv:2010.15458, 2020.
[20] 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.
[21] Van Laarhoven T. L2 regularization versus batch and weight normalization [J]. arXiv preprint arXiv:1706.05350, 2017.