



Constructing a Multi-scale Medical Knowledge Graph from Electronic Medical Records

Yikai Zhou, Ziyi Wang, Miao Li^(✉), and Ji Wu

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
{zyj21,wangziyi21}@mails.tsinghua.edu.cn,
{miao-li,wuji_ee}@mail.tsinghua.edu.cn

Abstract. Knowledge graphs play a crucial role in the medical field. Most existing knowledge graphs are manually created by experts or extracted from medical encyclopedias, resulting in the omission of valuable knowledge from medical clinical practice. Entities like diseases and symptoms in medicine exist at different levels, but current knowledge graphs fail to handle the induction and integration of this multi-scale information effectively. In our study, we constructed a knowledge graph that better aligns with real clinical data and effectively integrates multi-scale medical information by performing data preparation, medical entity extraction, negation handling, relation extraction, and graph cleaning. The reliability and rationality of the knowledge graph have been verified through subjective and objective assessments.

Keywords: multi-scale knowledge · medical knowledge graph · knowledge graph construction

1 Introduction

Knowledge graphs express and store a large number of knowledge elements such as entities, concepts, properties, and relationships in a structured way, forming a knowledge network with associative relationships. It is widely used in various fields such as healthcare [9], military [8], and finance [2]. With the development of neural networks, knowledge graphs can provide support for other deep learning tasks more effectively [11] and can also expand the scale of knowledge graphs through deep learning [16], by which research on knowledge graphs is receiving more and more attention from researchers.

Unlike other fields, medical knowledge has stronger specialization and higher complexity. As shown in Fig. 1, this complexity is reflected in the multi-scale characteristics of various types of medical knowledge such as symptoms, diseases, signs, and locations, for example, “high fever” and “low fever” are both finer-scale symptoms of “fever”. In many tasks in the medical field, such as assisted diagnosis, medical Q&A, and ICD coding, information at different scales plays an important role in refining task results and improving method robustness.

For example, in an intelligent assisted diagnosis task, if an intelligent diagnosis system diagnoses a patient with “acute pharyngitis” as “acute upper respiratory infection”, it may seem like an incorrect diagnosis. However, in reality, “acute pharyngitis” is a subclass of “acute upper respiratory infection”, so such an error may not be strictly considered a mistake. Perhaps if the patient provides more detailed information at a finer scale, the model can correct the result to “acute pharyngitis”. However, current medical knowledge graphs generally ignore the multi-scale characteristics of medical knowledge and do not effectively integrate the hierarchical relationships between multi-scale medical knowledge into the knowledge graph.

In terms of knowledge storage, most knowledge graphs store knowledge in the form of triples (head entity, relationship, tail entity). The medical knowledge described by each triplet is absolutely relevant (it can be understood that the confidence level of each triplet is 1). However, in clinical practice, diseases are often not absolutely related to other medical entities. For a certain disease, some symptoms may occur in almost all patients with the disease, while others may only appear in some patients. Whether the symptom appears in the patient will be affected by other symptoms and signs of the patient. In knowledge graphs stored in triple form, the relationships between these symptoms and diseases are often blurred into the deterministic relationship of “disease-related symptoms”, ignoring the modeling of the degree of correlation between diseases and symptoms, resulting in information loss.

For medical knowledge, many connections between medical entities exist implicitly in clinical text data, rather than explicitly in the cognition of medical experts or medical encyclopedias. Currently, some researchers are constructing medical knowledge graphs based on electronic medical record data [5, 6, 12, 18]. However, in the process of extracting knowledge, they overlook the negation expressions in medical record texts. For a patient, the descriptions of “having a fever” and “not having a fever” have completely different meanings. If the content related to negation words such as “no” and “deny” is not handled separately, it will inevitably lead to errors in the information contained in the constructed knowledge graph.

To address the shortcomings of current medical knowledge graphs, we have constructed a knowledge graph that is more in line with real clinical data and incorporates hierarchical relationships between multi-scale knowledge based on electronic medical record data, using a combination of manual construction and automatic construction. We have verified the reliability of the constructed knowledge graph and the rationality of the graph construction method through subjective evaluation by medical experts and objective evaluation through experiments.

The main contributions of this article are as follows: 1. Believing that information at different scales is crucial for various applications in the medical field, we effectively organize and integrate the hierarchical relationships between medical entities of different scales into the constructed knowledge graph. These entities with hierarchical relationships include diseases, symptoms, signs, and locations. 2. In the process of mining implicit connections between medical entities

from electronic medical record data, this paper proposes a new method to model the correlation degree between medical entities, which ensures the rationality and reliability of the medical knowledge graph. 3. When mining the relationships between medical entities through electronic medical record texts, we handle the entities related to negation words in the medical record texts separately, further improving the quality of the knowledge graph. This is a feature that other related works do not possess.

The remaining parts of the article will be presented in the following order. In Sect. 2, we will review related work on the construction of medical knowledge graphs. Section 3 will introduce the methods used to construct the knowledge graph in this article. Section 4 will describe how we evaluate the quality of the constructed knowledge graph. Section 5 will provide a summary of the work done in this article.

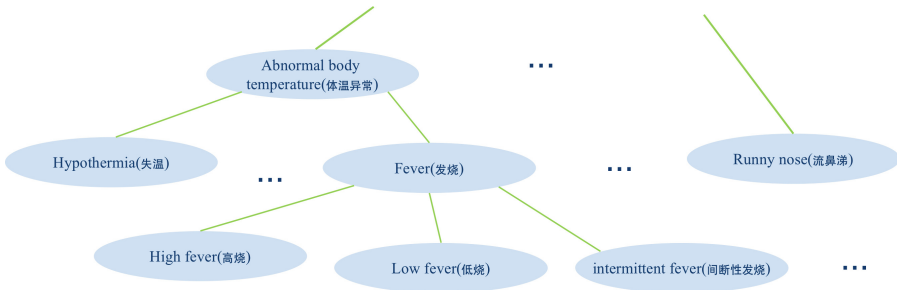


Fig. 1. In the medical field, entities have multi-scale characteristics.

2 Related Work

With the advancement of technology in the field of knowledge graphs, the construction of specialized and comprehensive medical knowledge graphs has become a hot research topic. The construction of medical knowledge graphs is driven by both medical knowledge resources from professional institutions or open-source, such as UMLS [1] and SNOMED CT [3], and real world clinical medical data.

There are some works constructing graphs based on co-occurrence relationships between entities. Finlayson et al. [4] analyzed and merged medical terms using over 20 million clinical medical data spanning over 19 years, constructing a co-occurrence matrix of 1 million medical clinical concepts to quantify the relationships between medical terms. Some works focus on designing a knowledge graph construction system for specific data, which are mainly based on rules. Lin et al. [7] proposed the MEDLedge model, which employs a hierarchical segmentation approach and a voting algorithm to extract entities and relationships from

clinical data and construct a knowledge graph. Shi et al. [13] utilized data from the Health Information System from Zhejiang, China, to propose a medical information integration model that standardizes heterogeneous medical information into a shareable and consistent format. Additionally, some constructed knowledge bases are disease-centered, without relation exploitation between other types of entities. Rotmensch et al. [10] extracted medical concepts from over 270 thousand patient records, utilizing probability models like Bayesian models to automatically construct a knowledge graph that links diseases and symptoms, creating a high-quality knowledge base from medical records. Zhao et al. [18] sampled 992 medical records, representing medical entities as nodes and co-occurrence relationships as edges, to establish an EMR-based medical knowledge network (EMKN). Furthermore, Zhao et al. [17] integrated EMKN with Markov Random Fields (MRF) for general medical knowledge representation, including five types of medical entities, and designed different energy functions based on inference scenarios. Li et al. [6] constructed a knowledge graph from 16 million clinical records and comprehensively described the entire graph construction process. In contrast to traditional triplets, they proposed a new quadruplet structure that leverages some attributes, including co-occurrence probability, reliability, and specificity, to better express the relationship between entities. However, their approach to constructing knowledge graphs only considers the relationships between diseases and other types of medical entities, while neglecting the interconnections between entities beyond diseases. Considering the large-scale knowledge graph, Yu et al. [15] built the first large-scale publicly available biomedical knowledge graph, containing millions of bilingual concepts and terms and 7.3 million relation triplets, which are all generated algorithmically without human participation.

To the best of our knowledge, there is no work that has taken advantage of the multi-scale hierarchical relationships between entities in building a knowledge graph based on electronic medical record texts. Additionally, most of them only consider the correlation information between diseases and other types of entities, while ignoring the correlation information between entities other than diseases. Furthermore, no researchers have performed additional effective handling of negative expressions in the process of constructing knowledge graphs from electronic medical records, which greatly affects the quality of the knowledge graphs.

3 Method

In this study, we comprehensively consider the professionalism of medical knowledge and the authenticity of clinical knowledge. We determine the scale information of medical entities based on expert experience and obtain the entity relationships from medical records through automated methods. After five steps including data preparation, medical entity extraction, negation handling, relationship extraction, and graph cleaning, we construct a high-quality multi-scale medical knowledge graph. The overall construction process is shown in Fig. 2.

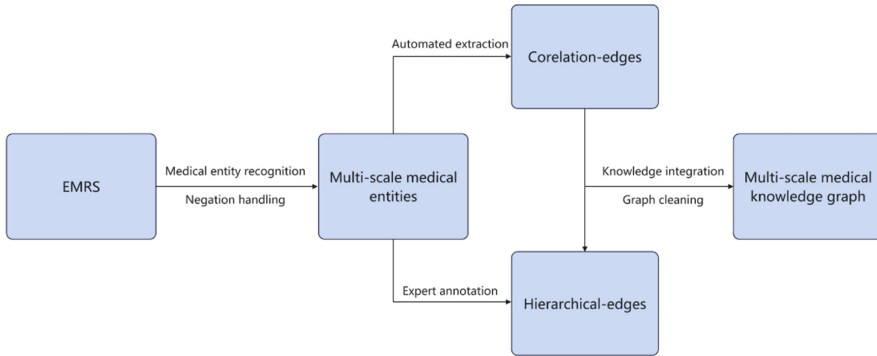


Fig. 2. Framework of our method to build a multi-scale medical knowledge graph.

3.1 Data Preparation

We aimed to construct a knowledge graph focused on lung diseases with a hierarchical structure. For these lung diseases, we sampled some medical records from the electronic medical record database, ensuring the distribution of medical records related to different diagnoses was as uniform as possible. These records will be used for both the construction of a multi-scale medical knowledge graph and as a dataset for subsequent validation of the quality of the knowledge graph.

Each medical record includes multiple sections such as “admission record”, “initial course record”, “examination report”, and “discharge record”. Each section contains multiple fields. To ensure comprehensive coverage of knowledge, we will use a total of 7 sections including “admission record”, “examination report”, and “discharge record”, and 24 fields including “chief complaint”, “present illness history”, and “main symptoms” for the construction of this knowledge graph.

In addition to medical record data, the hierarchical relationships between diseases, symptoms, signs, and locations are essential for constructing a multi-scale medical knowledge graph. They organize the hierarchical information between medical entities and establish connections between the information hierarchy and diagnostic hierarchy. This hierarchical knowledge is annotated by medical experts.

3.2 Medical Entity Extraction

After obtaining the medical records required to construct the knowledge graph, we need to extract medical entities of different scales from the unstructured medical record texts. In the medical record texts, there are medical entities such as diseases, symptoms, signs, and medicine which are important for accurate diagnosis. It is crucial to extract these entities effectively, accurately, and comprehensively for the high-quality construction of the knowledge graph.

First, we remove irrelevant information such as symbols and stop words from the medical record texts. Then we perform medical entity recognition. Regarding the methods for entity recognition, we compared statistical-based recognition methods with dictionary-based bidirectional maximum matching recognition methods. It was observed that the statistical-based recognition method tends to identify entities at a very rough level. Entities with finer granularity often have longer text representations, and the statistical-based method often recognizes a fine-grained entity as multiple coarse-grained entities. For instance, the term “ANCA-associated vasculitis” is identified as three separate entities: “ANCA”, “associated”, and “vasculitis”. “Chronic kidney disease stage 5” is recognized as “chronic” and “kidney disease”. Obviously, this is very disadvantageous for constructing a medical knowledge graph with multiple scales of information. In contrast, the dictionary-based bidirectional maximum matching method effectively resolves this issue. Consequently, the dictionary-based bidirectional maximum matching algorithm is employed in this study for extracting medical entities.

3.3 Negation Handling

During the entity extraction process, we have observed a significant presence of disease denials and negative symptoms in the medical records. For instance, in the phrase “no fever symptoms, without vomiting”, both “fever” and “vomiting” are negative symptoms. If these negative symptoms are not appropriately addressed, they may be mistakenly identified as positive symptoms, leading to confusion and compromising the quality of the knowledge graph.

To mitigate this issue, we employ text understanding techniques to identify entities associated with disease denials and negative symptoms in the medical records. Subsequently, we process these entities separately and incorporate the negation semantic information into the knowledge graph.

3.4 Relation Extraction

Furthermore, we need to extract the relationships between entities from medical records. Unlike conventional knowledge graphs that store entity relationships using triplets, we calculate a weight for each possible triplet to measure its confidence.

We refer to the graph construction method of TextGCN [14] and model the medical record text as a disease document node. By calculating the TF-IDF weights between the document node and various medical entities in the medical record, we model the relevance between disease and non-disease entities. For the relevance between non-disease medical entities in the medical record text, we use PMI weights for modeling.

When calculating the weights between nodes in the knowledge graph, we take into account that the number of medical records for each diagnosis varies and the length of each medical record text is also different. Therefore, we consider both factors and normalize the weights to a range between 0 and 1.

In addition, for entities with hierarchical relationships in the hierarchy system, considering that the weight reflects the relevance between two entities and the hierarchical relationship is annotated by medical experts, we believe that medical entities with hierarchical relationships have a strong correlation, so we directly set their weight to 1.

3.5 Graph Cleaning

After completing the preliminary construction of the knowledge graph, there remains a significant amount of redundant information that needs to be processed. Whether it is introducing the relationship between disease and non-disease entities through TF-IDF or introducing the relationship between non-disease entities through PMI, a considerable amount of noise is introduced into the constructed knowledge graph. The abundance of noise makes the connection between different diseases relatively similar and difficult to distinguish.

Therefore, to enhance the diversity of connections between different diseases in the graph, we undertake a cleaning process on the preliminary constructed knowledge graph. Since the weights between entities directly reflect their relevance, we establish thresholds for both PMI weights and TF-IDF weights. By setting these thresholds, we delete edges with lower relevance, thereby further improving the quality of the knowledge graph.

4 Knowledge Graph Quality Assessment

In order to validate our method, we selected 28 lung diseases that have hierarchical relationships with each other as the disease system within the knowledge graph. We sampled a total of 4548 medical records with main diagnoses that fall within the aforementioned 28 diseases from a large electronic medical record database; and ensured that the distribution of medical records related to different diagnoses was as uniform as possible. Based on the above process, we constructed a multi-scale medical knowledge graph containing 21,950 entities and 1,540,375 edges, in which 3346 multi-scale entities are included, while the numbers of each type of edge are illustrated in Table 1. The quality of this knowledge graph was then verified through subjective and objective evaluations.

During the subjective evaluation, we primarily focused on two questions: 1. For sibling disease pairs, whether the differences reflected in the knowledge graph support the differentiation of the two diseases? 2. For parent-child disease pairs, whether there is consistency at the scale level between the hierarchy of diseases and the hierarchy of related entities? In objective evaluation, we verified the quality of the constructed knowledge graph and the effectiveness of the knowledge graph construction method through a disease classification task.

4.1 Subjective Assessment

After completing the construction of the knowledge graph, we conducted a preliminary evaluation to assess the quality of the knowledge graph, verifying

Table 1. Number of Different Types of Edges in our Knowledge Graph.

hierarchical edge	TF-IDF edge	PMI edge
3346	127403	1409626

whether the information reflected by different diseases in the knowledge graph has discriminative significance. We selected some representative sibling disease pairs and parent-child disease pairs from the disease tree, extracted disease sub-graphs from the knowledge graph, compared the disease sub-graphs, and counted the discriminative entities between disease pairs. Medical experts then evaluated the discriminative entities extracted from the knowledge graph to verify whether they conform to the clinical significance of discrimination. Our method for extracting discriminative entities is illustrated in Fig. 3.

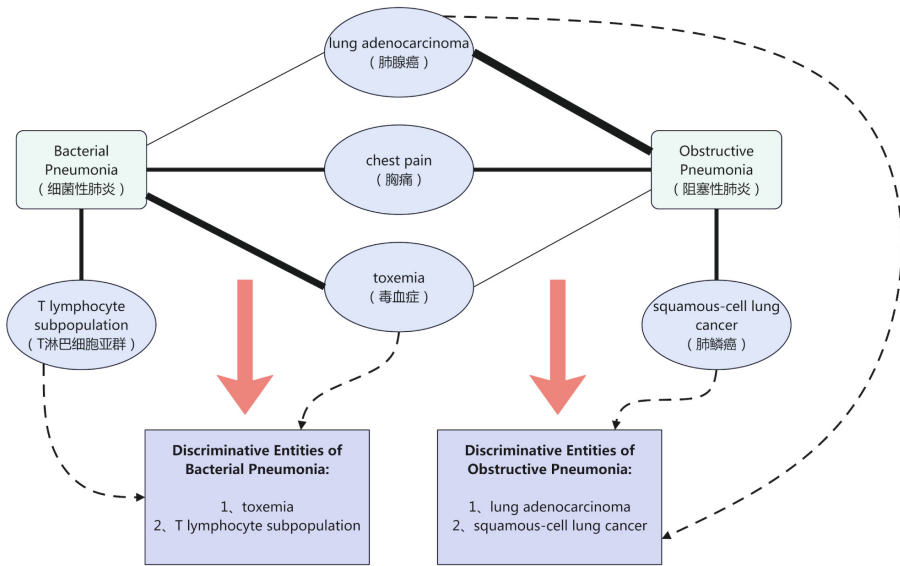


Fig. 3. This figure illustrates our method for extracting discriminative entities. The thickness of the lines connecting diseases and their related entities represents the degree of correlation between them. In two scenarios, the associated entities of a disease are extracted as discriminative entities. 1) The entity is only associated with that disease and not with its sibling diseases. 2) The degree of correlation between the entity and the disease is much higher compared to its sibling diseases (shown as a significant difference in line thickness in the figure).

When verifying the information difference between sibling disease pairs, we took “obstructive pneumonia” and “bacterial pneumonia” as examples. We sorted the discriminative entities reflected by these two diseases in the knowledge

graph based on their weights. We found that in the knowledge graph, the discriminative entities of “obstructive pneumonia” include “pulmonary squamous cell carcinoma”, “malignant tumor immunotherapy”, “small cell lung cancer”, etc., while the discriminative entities of “bacterial pneumonia” include “T lymphocyte subset”, “acid-fast staining”, “serum myoglobin”, etc. We asked doctors to label these high-confidence discriminating entities extracted from the knowledge graph. The results of the doctor’s labeling are shown in Table 2. (where ✓ represents approval, × represents objection, ★ represents uncertainty), it is evident that doctors also believe these entities have significant discriminatory value for these two diseases in clinical practice. In addition, relevant medical knowledge further demonstrates the reliability of the knowledge graph. Medical knowledge shows that the cause of bacterial pneumonia is bacterial infection, and common pathogens include *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Staphylococcus aureus*, etc.; while the cause of obstructive pneumonia is chronic obstructive pulmonary disease (COPD) and other respiratory system diseases, leading to airway narrowing, gas exchange disorders, etc. We observed that the discriminative entities of bacterial pneumonia primarily consist of examination entities, while the discriminative entities of obstructive pneumonia mainly include some obstructive lung diseases, which are consistent with the clinical etiology of the two diseases. Therefore, we believe that the associated entities of sibling diseases in the constructed knowledge graph have differential information for discrimination.

Table 2. Discriminative Entities of Sibling Disease Pair(Bacterial Pneumonia and Obstructive Pneumonia), in which ✓ represents APPROVAL, × represents OBJECTION, ★ represents UNCERTAINTY.

Disease	Discriminative Entities from Knowledge Graph	Expert Annotation
Bacterial Pneumonia (细菌性肺炎)	T lymphocyte subpopulation(T淋巴细胞亚群)	✓
	acid-fast stain(抗酸染色)	✓
	acute upper respiratory tract infection(急性上呼吸道感染)	✓
	absolute count(绝对计数)	★
	serum myoglobin(血清肌红蛋白)	✓
	vena epigastrica(腹壁静脉)	×
	Serum troponin T(血清肌钙蛋白T)	★
	gram stain(革兰染色)	✓
Obstructive Pneumonia (阻塞性肺炎)	squamous-cell lung cancer(肺鳞癌)	✓
	Castleman disease(castleman病)	✓
	aversion to cold with fever(恶寒发热)	★
	immunotherapy for malignant tumors(恶性肿瘤免疫治疗)	✓
	lung-distension(肺胀)	★
	lung adenocarcinoma(肺腺癌)	✓
	small cell lung cancer(小细胞肺癌)	✓
	liver puncture(肝脏穿刺)	×

In addition, we also observed the consistency of discriminative entities between parent-child disease pairs, confirming the consistency of the scale level

between superordinate and subordinate disease entities and other types of superordinate and subordinate entities. We took the three diseases “respiratory tract infection”, “acute upper respiratory tract infection” and “pulmonary infection” as examples. “Acute upper respiratory tract infection” and “pulmonary infection” are both sub-disease nodes of “respiratory tract infection”. We found that in the knowledge graph, “neck” is the discriminative entity for “respiratory tract infection”, while the sub-entities “pharynx” and “tonsil” of “neck” are the discriminative entities for “acute upper respiratory tract infection”. “Pain” and “enlargement” are the discriminative entities for “respiratory tract infection”, while the sub-entities “chest pain” and “enlarged cardiac silhouette” of “pain” and “enlargement” are the discriminative entities for “pulmonary infection”. Figure 4 shows the consistency of diseases and symptoms in terms of scale in the knowledge graph. It can be seen that the constructed knowledge graph not only reflects the differences in related entities between sibling diseases but also the associated entities of parent-child diseases have consistency at the scale level, which basically meets our expectations for the quality of the knowledge graph.

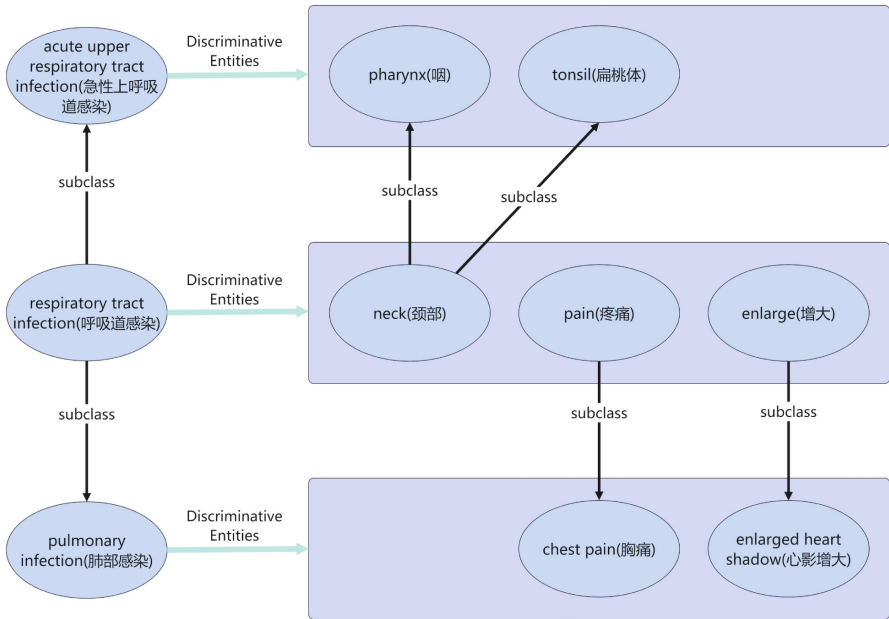


Fig. 4. For parent-child disease pairs (take “Respiratory Tract Infection”, “Acute Upper Respiratory Tract Infection” and “Pulmonary Infection” as examples), there is consistency at the scale level between the hierarchy of diseases and the hierarchy of related entities.

4.2 Objective Assessment

In order to further verify the quality of the constructed knowledge graph, we designed a simple disease classification task. We used medical records as input, extracted all medical entities from the records, and connected them to the knowledge graph. Based on these entities, we extracted a subgraph from the knowledge graph and directly classified it into 28 disease categories.

It is worth mentioning that in this validation task, we did not have any model training process. For each subgraph of medical records, we set the node features as 28 dimensions, with each dimension representing the TF-IDF weight value of the node with respect to the 28 diseases. The PMI weights between nodes in the subgraph were stored using an adjacency matrix. Then, we read out each subgraph node into a 1*28 dimensional vector, and after applying softmax operation, we unexpectedly achieved 44.0% accuracy in the 28-class classification of the medical records. It should be noted that if these test data are randomly classified into 28 categories, the performance can only achieve an accuracy of 5.1%. Based on this, we constructed a knowledge graph without considering the handling of negation words; and applied this knowledge graph to the same classification task. We found that the accuracy of the classification significantly declined to 41.5%, indicating the importance of negation word handling for the quality of the knowledge graph.

We also compared the performance of Bert on our dataset. Clearly, without any training data, Bert is unable to classify these diseases. However, our knowledge graph method can achieve a classification accuracy of 44.0% in the case of zero-shot. This suggests that the quality of the knowledge graph we constructed has a certain guarantee. In Table 3, we present the performance of different methods on the 28-classification task, in which methods with “(*)” mean that they are zero-shot predictions.

Table 3. Performance of Different Methods on the 28-classification Task(Among them, methods with * mean that they are zero-shot predictions.)

Method	accuracy	macro-precision	macro-recall	macro-f1
random(*)	0.051	0.028	0.033	0.030
KG-based w/o negation handling(*)	0.415	0.519	0.419	0.464
KG-based w/ negation handling(*)	0.440	0.541	0.443	0.487
Bert	0.521	0.492	0.510	0.501

5 Conclusion

In this paper, we constructed a multi-scale medical knowledge graph by mining the hidden connections between medical entities in electronic medical record

data and introducing medical multi-scale information through expert annotation. It is worth noting that we also considered the impact of negation words in medical record texts on the quality of the knowledge graph and made additional processing during the construction of the knowledge graph. Subsequently, we preliminarily confirmed the quality of the constructed knowledge graph through subjective evaluation by medical experts, and further confirmed the effectiveness of our multi-scale knowledge graph construction method and the importance of handling negation through objective evaluation in experiments.

Acknowledgments. The work is supported by National Key R&D Program of China (2021ZD0113404).

References

1. Bodenreider, O.: The unified medical language system (UMLs): integrating biomedical terminology. *Nucleic Acids Res.* **32**(suppl.1), D267–D270 (2004)
2. Cheng, D., Yang, F., Wang, X., Zhang, Y., Zhang, L.: Knowledge graph-based event embedding framework for financial quantitative investments. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2221–2230 (2020)
3. Donnelly, K., et al.: SNOMED-CT: the advanced terminology and coding system for ehealth. *Stud. Health Technol. Inform.* **121**, 279 (2006)
4. Finlayson, S.G., LePendou, P., Shah, N.H.: Building the graph of medicine from millions of clinical narratives. *Sci. Data* **1**(1), 1–9 (2014)
5. Li, L., et al.: A method to learn embedding of a probabilistic medical knowledge graph: algorithm development. *JMIR Med. Inform.* **8**(5), e17645 (2020)
6. Li, L., et al.: Real-world data medical knowledge graph: construction and applications. *Artif. Intell. Med.* **103**, 101817 (2020)
7. Lin, K., Wu, M., Wang, X., Pan, Y.: MEDLedge: a Q&A based system for constructing medical knowledge base. In: *2016 11th International Conference on Computer Science & Education (ICCSE)*, pp. 485–489. IEEE (2016)
8. Liu, C., Yu, Y., Li, X., Wang, P.: Application of entity relation extraction method under CRF and syntax analysis tree in the construction of military equipment knowledge graph. *IEEE Access* **8**, 200581–200588 (2020)
9. Liu, W., Yin, L., Wang, C., Liu, F., Ni, Z., et al.: Multitask healthcare management recommendation system leveraging knowledge graph. *J. Healthcare Eng.* **2021** (2021)
10. Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S., Sontag, D.: Learning a health knowledge graph from electronic medical records. *Sci. Rep.* **7**(1), 5994 (2017)
11. Sang, L., Xu, M., Qian, S., Wu, X.: Knowledge graph enhanced neural collaborative recommendation. *Expert Syst. Appl.* **164**, 113992 (2021)
12. Shen, Y., et al.: CBN: constructing a clinical Bayesian network based on data from the electronic medical record. *J. Biomed. Inform.* **88**, 1–10 (2018)
13. Shi, L., et al.: Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services. *BioMed Res. Int.* **2017** (2017)
14. Yao, L., Mao, C., Luo, Y.: Graph convolutional networks for text classification. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 7370–7377 (2019)

15. Yu, S., et al.: Bios: an algorithmically generated biomedical knowledge graph. arXiv preprint [arXiv:2203.09975](https://arxiv.org/abs/2203.09975) (2022)
16. Zhang, Z., Zhuang, F., Zhu, H., Shi, Z., Xiong, H., He, Q.: Relational graph neural network with hierarchical attention for knowledge graph completion. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 9612–9619 (2020)
17. Zhao, C., Jiang, J., Guan, Y., Guo, X., He, B.: EMR-based medical knowledge representation and inference via Markov random fields and distributed representation learning. *Artif. Intell. Med.* **87**, 49–59 (2018)
18. Zhao, C., Jiang, J., Xu, Z., Guan, Y.: A study of EMR-based medical knowledge network and its applications. *Comput. Methods Programs Biomed.* **143**, 13–23 (2017)