ISSN 2616-5775 Vol. 6, Issue 8: 75-81, DOI: 10.25236/AJCIS.2023.060809

A Modeling Study of a Multi-modal Knowledge Graph of Children's Medical Information

Wanting Wang, Shishu Yin*

School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, 233000, Anhui, China
yin_shishu@163.com
**Corresponding author*

Abstract: *With the rapid development of the application of information technology in the medical field and the gradual improvement of medical information storage standards, medical data presents a multi-modal form while growing rapidly. For managing, organizing and analyzing multi-modal medical data effectively, this paper takes children's medical data as an example, and uses the computer vision processing technology to realize knowledge acquisition, knowledge extraction, entity linking, knowledge storage of multi-modal children's medical data. The structured and unstructured medical data are organized together to achieve the multi-modal children's medical information knowledge graph.*

Keywords: *Multi-modality; Children's healthcare; Knowledge graph; Modeling; Evaluation*

1. Introduction

Recently, medical data has shown a multi-modal form while growing rapidly. This depends on the development of information technology in the medical field and the improvement of medical information storage standards. Diagnostic reports, multiple medical imaging devices, computed tomography scans and other medical images are all called multi-modal data. These multi-modal data mix and coexist, forming a semantically similar and interrelated complex features. How to construct a large-scale multi-modal and multi-source medical knowledge base will be a difficult and hotspot of research in the medical field.

Multi-modal knowledge graph, shown as one of hotspots in the field of artificial intelligence, is a product of the organic combination of knowledge graph and multi-modal learning. Different from the early single-graph database, the multi-modal knowledge graph mainly focuses on the two modalities of image and text. Wang[1] et al. think that multi-modal knowledge graph is constrcuted on the basis of the traditional knowledge graph with entities in multiple modalities and contain semantic relationships between entities in multiple modalities. Sun[2] et al. consider that a multi-modal knowledge graph is a knowledge graph containing multiple data types, such as text and image. Here, text or image is an entity or an attribute of an entity. Now, researches on multi-modal learning and knowledge graph have made great progress in the medical field respectively, but the study of multi-modal knowledge graph in the same field is still in its infancy. A model is proposed in[3] to build a multi-modal knowledge graph which is based on epilepsy field papers and to support downstream applications in an efficient knowledge organization way. In this work, we take children's medical data as an example to study the method of constructing multi-modal knowledge graph. Succinctly, computer vision processing technology is uesed to realize knowledge acquisition, knowledge extraction, entity linking, knowledge storage of multi-modal children's medical data. Integrate multiple medical data sources together to form a graph containing both structured and unstructured medical data. When the multi-modal knowledge graph is achieved, some metrics, such as precision rate and recall rate, can be used to evaluated its accuracy and effectiveness.

The structure of this paper is briefly introduced as follows: Firstly, the esstential elements of multi-modal children's medical information knowledge graph are studied in part II. Then, a detailed construction process of multi-modal children's medical information knowledge graph is proposed in part III. In the fourth part, some metrics are suggested to evaluate the constructed knowledge graph. Finally, a conclusion is drawn.

ISSN 2616-5775 Vol. 6, Issue 8: 75-81, DOI: 10.25236/AJCIS.2023.060809

2. Elemental Components of a Multi-modal Knowledge Graph for Children's Medical Information

2.1. Knowledge Graph

Google Inc. proposed the concept of knowledge graph in 2012[4], which has gained widespread attention in academia and industry. Knowledge graph describes concepts, entities, and relationships between them in the objective world in the form of factual triples. Formally, a triple can be represented as "(head entity, relationship, tail entity)", e.g., "(autism, clinical manifestation, inattention)" constitutes a triple instance. Currently, common large-scale knowledge graphs include Freebase[5], DBpedia[6], YAGO[7], etc. Knowledge graph is essentially a graph model-based knowledge representation of associative networks, aiming to adopt the structure of graphs to model and record the associative relationships and knowledge between everything in the world for more accurate object-level search[8]. Knowledge graph has experienced the process of being constructed by artificial intelligence and group intelligence in the early stage to automatic acquisition by using technologies such as machine learning and information extraction, and gradually expanded from a single text mode to a huge multi-modal coexistence.

2.2. Multi-modal Knowledge Graph for Children's Medical Information

Multi-modal knowledge graph takes text, picture, video, and audio and other multi-modal data as elements, which is a network of association paths with cross-modal, class relationship, directed, non-crossing, and is essentially a kind of relational data connectivity in the form of semantic network[9]. Multi-modal knowledge graph for children's medical information field mainly focuses on children's medical data distribution, characteristics, connections, scale, etc. to carry out graph construction technology research.

The elements of multi-modal children's medical information knowledge graph are composed of nodes and edges, with nodes corresponding to entities and edges corresponding to relationships. According to the data sources, entities are mainly categorized into 5 types: concepts, text, images, video, and audio. Figure 1 represents the composition of different entities corresponding to different attributes in the field of children's medical information, where conceptual entities include inherent and environmental attributes. For example, autism is a conceptual entity whose inherent attributes include: etiology, clinical manifestations, differential diagnosis, etc. Its environmental attributes refer to the attributes in a specific scenario, e.g., the child is subjected to environmental pollution for a long period of time, which causes damage to the nerves of the brain and thus suffers from autism, and at this time the induced cause is due to the environmental attributes.

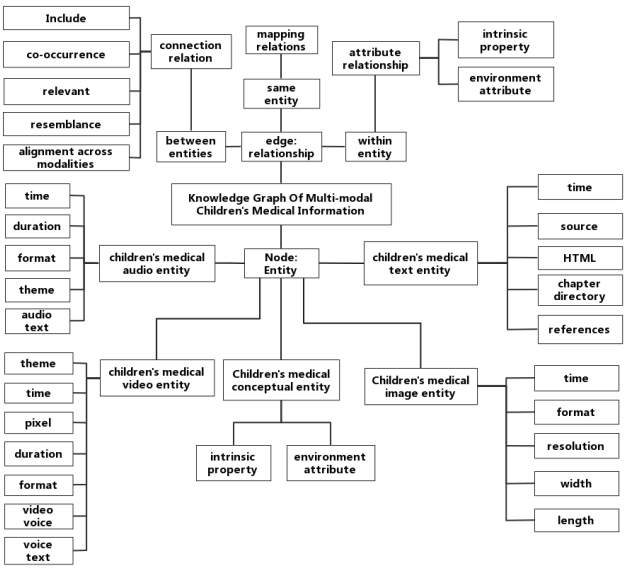

Figure 1: Schematic diagram of the composition of the elements of the multi-modal children's medical information knowledge graph

ISSN 2616-5775 Vol. 6, Issue 8: 75-81, DOI: 10.25236/AJCIS.2023.060809

3. Construction of a Multi-modal Knowledge Graph for Children's Medical Information

3.1. Construction of Knowledge Graph

The construction model of the knowledge graph consists of named entity recognition and relationship extraction, as shown in Figure 2. To construct the knowledge graph of children's medical information, the preprocessed children's medical data are firstly used to find out the medical-related entities in the text through the named entity recognition model. Then to find out the relationships between the named entities through the relationship extraction model, and use these entities and relationships to construct the knowledge graph.

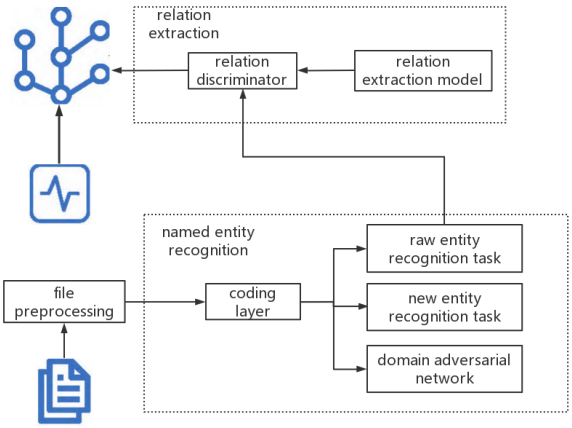

Figure 2: Knowledge graph construction model

3.2. Construction of a Multi-modal Knowledge Graph

Multi-modal knowledge graph embeds multi-modal entities on the basis of traditional knowledge graph and constructs semantic relationships between multi-modal entities, which further helps to understand multi-modal data and extends the function of knowledge graph in search and visualization. According to the different focuses of research directions, there are currently two typical ideas for constructing multi-modal knowledge graphs, as shown in Figure 3.

a) It is relatively mainstream to construct multi-modal knowledge graph from the perspective of natural language processing, but it has not yet got rid of the dependence on traditional textual knowledge graph. On the basis of the construction of textual knowledge graph, the work of supplementing visual information to the entities is essentially the knowledge graph complementation and doing the discovery of visual relationships and cross-modal entity linking among images. For the expansion of image entities and the determination of linking relationships mainly rely on the metadata of multi-modal data. However, this approach is still coarse-grained for visual feature extraction and multi-modal relationship mining, and some entities such as concepts do not contain image information in real scenes.

b) The multi-modal knowledge graph construction work from the perspective of computer vision is built on the basis of scene atlas generation, which is a distributed construction method that bridges visual knowledge with external textual knowledge graph. Compared with the general knowledge knowledge graph in which a node represents an entity or predicate class, the graph node of the scene graph represents an entity or predicate instance in a specific image, which needs to be linked to the corresponding entity or predicate class in the text knowledge graph to complete the construction. As computer vision for image recognition effect is biased towards abstract conceptual level recognition, and the visual relationship is relatively single, there are more noise effects[9]. To achieve fine-grained entity alignment and visual relationship discovery under this construction idea still requires the assistance of more new technologies.

ISSN 2616-5775 Vol. 6, Issue 8: 75-81, DOI: 10.25236/AJCIS.2023.060809

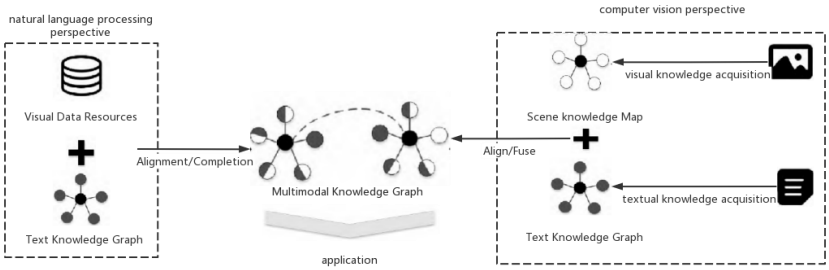

Figure 3: Two typical methods for multi-modal knowledge graph construction

3.3. A Multi-modal Approach to Constructing a Knowledge Graph of Children's Medical Information

3.3.1. Build Process

In this paper, multi-modal children's medical information knowledge graph is constructed from the perspective of computer vision, and the construction steps include knowledge acquisition, knowledge extraction, entity linking, and knowledge storage. The construction of children's medical information knowledge graph requires a large number of datasets. This paper uses Python web crawlers to crawl children's medical data. After data cleaning and preprocessing, the data is divided into training set, test set and verification set for subsequent knowledge graph construction. Knowledge extraction is knowledge extraction from data of different sources and structures, usually including entity extraction, relation extraction, attribute extraction, etc. The knowledge extraction results are stored in the form of relation triples or attribute triples in the database. Entity linking includes entity disambiguation and co-reference disambiguation, which is used to solve the problem of multiple meanings of one word and multiple words with the same meaning in the knowledge graph, and is an indispensable step in the process of knowledge graph construction. The specific technical route is shown in Figure 4.

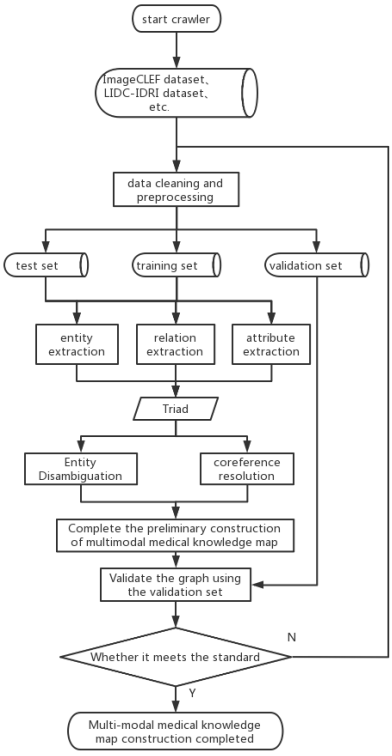

Figure 4: Technology roadmap

ISSN 2616-5775 Vol. 6, Issue 8: 75-81, DOI: 10.25236/AJCIS.2023.060809

3.3.2. Knowledge Acquisition and Preprocessing

The dataset is crucial to carry out the research on the construction of multi-modal children's medical information knowledge graph, and in this paper, we choose Medline dataset, CDC dataset, MIMIC dataset, and Baidu dataset. In the defined ontology of children's medical information domain, the categories with more accurate and specific image and video information are selected, so it is necessary to obtain the relevant image and video information while crawling the text information of these categories. The URLs of web pages containing information about these categories are first collected, they are stored in a txt document, and the crawler is used to crawl the images, videos and the surrounding text.

Data preprocessing of children's medical information includes text data preprocessing and picture and video data preprocessing. Text data preprocessing mainly includes corpus cleaning, Chinese word segmentation, lexical labeling and other steps; the preprocessing of picture and video data is mainly to label them with appropriate text labels, i.e., the corresponding entity names, so as to facilitate the subsequent addition of picture and video attributes to the entities. When crawling, data links and information with related fields are stored in different txt files according to categories.

3.3.3. Joint Entity-Relationship Extraction

In this paper, we perform sentence-level joint entity-relationship extraction based on a deep learning named entity recognition model. Named entity recognition, as a fundamental task in natural language processing, aims to identify entities with specific meanings, such as patients, diseases, and healthcare organizations, from unstructured text[10]. The input to the named entity recognition task is a sequence and the output of the model is a sequence of labels of the input sequence. A relationship is defined as some kind of connection between two or more entities, the input to the relationship extraction task is a piece of text, and the output is usually a triple. The output of joint entity-relationship extraction consists of the entities in each sentence, the type of entity, and the triad of relationships extracted from the sentence.

3.3.4. Entity Links

The purpose of entity linking is to link new entities with their counterparts in the knowledge base to supplement the content of the knowledge graph, which is used to solve the problem of entity ambiguity and diversity[11].

In this paper, a new Multi-Modality Cross Attention (MMCA) model is used for image and sentence matching by jointly modeling the internal and intermediate modality relationships of image regions and sentence words in a unified deep model[12]. The model, in order to achieve robust cross-modal matching, is designed with two effective attentional modules, including the self-attention module and the cross-modal module, which play an important role in modeling the intra-modal and inter-modal relations.

In the self-attention module, a bottom-up model is used to extract features of salient image regions. At the same time, word token embeddings are used as linguistic elements. Then, independently, the image regions are fed into the Transformer unit and the word tokens are fed into the supervised pretrained language model to model the relationships within the modality. The global representation can then be obtained by aggregating these fragment features.

In the cross-attention module, representations from image regions and sentence words are stacked and then passed into another Transformer unit, followed by one-dimensional convolution (1d-CNN) and pooling operations to fuse inter-modal and intra-modal information. Then, based on the updated features of the visual and textual data, we can predict the similarity scores of the input images and sentences.

For a given image-text pair, the MMCA model first uses a pretrained model on the dataset to extract the image features, WordPiece to process the text sentences, and then a self-attention module, a cross-attention module, and then 1d-CNN and pooling for fusion, respectively. The final model yields two pairs of graphic features for performing image and text matching. After entity linking using the MMCA model, if the entity is found to exist in the knowledge graph, its new attribute information is added to the knowledge base; if the entity does not exist in the knowledge graph, both the entity and its attribute information are added together to the knowledge base. The MMCA model is also able to minimize the loss of the hard-negative based triples by updating both visual and textual features into a common embedding space.

In what follows, we present the metrics criteria used to evaluate the multi-modal children's medical information knowledge graph.

ISSN 2616-5775 Vol. 6, Issue 8: 75-81, DOI: 10.25236/AJCIS.2023.060809

4. Indicators for Graph Assessment

After the construction of the multi-modal children's medical information knowledge graph is completed, it is necessary to validate and evaluate the multi-modal children's medical information knowledge graph to verify the accuracy and effectiveness of the model when constructing and adding entities.

The evaluation indexes are: Precision, Recall, and F1 value (F1) are commonly used in the field of named entity recognition as the indexes for evaluating the recognition performance of the model, and the formula is as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{2}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision \times Recall} \times 100\% \tag{3}$$

Among them: TP means named in the sample, i.e., the sample value for which the named entity in the validation set is correctly linked; FP means the negative class is judged to be positive, i.e., the sample value for which the named entity in the validation set is incorrectly linked and the model result is shown to be correct; and FN means the positive class is judged to be negative, i.e., the sample value for which the named entity in the validation set is correctly linked and the model result is shown to be incorrect.

5. Conclusion

This paper provides a method to effectively model and manage structured and unstructured multi-modal medical data in the form of knowledge graph. Taking children's medical data as an example, knowledge acquisition, knowledge extraction, entity linking and knowledge storage by making use of MMCA are realized using computer vision processing technology to achieve the multi-modal children's medical information knowledge graph. Finally, the indicators of precision and recall are suggested to evaluate the accuracy and validity of the constructed knowledge graph.

References

[1] Wang Meng, Qi Guilin, Wang Haofen, et al. Richpedia: a compre-hensive multi-modal knowledge graph[C]//Proc of Joint International Semantic Technology Conference. Cham: Springer, 2019:130-145.
[2] Sun Rui, Cao Xuezhi, Zhao Yan, et al. Multi-modal knowledge graphsfor recommender systems[C]//Proc of the 29th ACM International Con-ference on Information & Knowledge Management. 2020:1405-1414.
[3] Li Xingyuan, Wang Peng, Shen Mu, etc. A Preliminary Study on the Construction of multi-modal Knowledge Graph for Epilepsy Related Papers[J]. Journal of Beijing University of Posts and Telecommunications, 2022, 45(04):19-24.DOI:10.13190/j.jbupt. 2021-187.
[4] A. Singhal, Introducing the knowledge graph: Things, not strings, May 2012. [Online]. Available from: http://googleblog.blogspot.com/2012/05/introducingknowledge-graph-things-not.html.
[5] Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//27th ACM SIGMOD International Conference on Management of Data, 2008:1247-1250.
[6] Auer S, Bizer C, Kobilarov G, et al. Dbpedia: a nuc-leus for a web of open data[M]. Springer, Berlin, Heidel-Berg, 2007:722-735.
[7] Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge[C]//16th International Conference on World Wide Web (WWW), 2007:697-706.
[8] Chen Ye, Zhou Gang, Lu Jicang. A review of research on multi-modal knowledge graph construction and application[J]. Computer Application Research, 2021, 38(12):3535-3543. DOI:10.19734/j.issn. 1001-3695.2021.05.0156.
[9] Peng Jinghui, Wang Zhen, Li Yue, Hou Ping. Research on multi-modal knowledge graph technology in equipment field[J]. Journal of Ordnance Equipment Engineering, 2022, 43(11):136-140+153.
[10] Wen Y, Fan C, Chen G, et al. A survey on named entityrecognition. Proceedings of the 8th

ISSN 2616-5775 Vol. 6, Issue 8: 75-81, DOI: 10.25236/AJCIS.2023.060809

International Conferenceon Communications, Signal Processing, and Systems. Urumqi: Springer, 2019. 1803–1810.
[11] Li Huayu, Fu Yafeng, Yan Yang, et al. Construction of multi-modal domain knowledge graph based on LEBERT[J]. Computer System Applications, 2022, 31(11):79-90. DOI:10.15888/j.cnki.csa. 008799.
[12] Wei Xi, Zhang Tianzhu, Li Yan, et al. Multi-modality cross attentionnetwork for image and sentence matching[C]//Proc of IEEE/CVF Con-ference on Computer Vision and Pattern Recognition. 2020:10941-10950.