Towards using Automatically Enhanced Knowledge Graphs to Aid Temporal Relation Extraction

Timotej Knez, Slavko Žitnik
University of Ljubljana, Faculty of Computer and Information Science
Večna pot 113
{timotej.knez, slavko.zitnik}@fri.uni-lj.si

Abstract

Temporal relation extraction in medical document analysis is crucial for understanding patient histories and treatment outcomes. This paper introduces a novel approach leveraging a bimodal model integrating textual content and a knowledge graph to enhance temporal relation extraction. The paper presents ongoing research on constructing an optimal knowledge graph by augmenting PrimeKG with dynamically expanded information using a language model-generated knowledge graph. It also further personalizes the information with patient-specific graphs tailored for relation prediction. The pipeline for constructing this enriched knowledge graph is detailed, aiming to improve the capabilities of temporal relation extraction models. The preliminary results show that adding a simple knowledge graph to the temporal relation extraction model can significantly increase the performance, achieving new state-of-the-art results. While research on enhanced knowledge graphs is ongoing, this paper lays the groundwork for leveraging common knowledge to advance temporal relation extraction in medical contexts. This approach holds promise for enhancing the understanding of patient histories and treatment outcomes, potentially leading to improved healthcare decision-making and patient care.

Keywords: Information extraction, Knowledge graph building, Large language models

1. Introduction

In medical document analysis, extracting temporal relations is pivotal in enhancing our understanding of patient histories and treatment outcomes. In our preliminary research, we showed that we can improve temporal relation extraction by using a bimodal model that integrates information not only from the textual content of medical documents but also from a knowledge graph containing information about the patient's treatment from a discharge summary. We suggest further improving the results by introducing common knowledge from multiple sources to the model. Our experiments use general medical knowledge from a knowledge graph named PrimeKG (Chandak et al., 2023), providing a broad information foundation. As this is a broad human-curated graph, it does not contain information about all the concepts appearing in the analyzed documents. Therefore, augmenting the information and adding some common sense knowledge about the missing concepts is important. To address this, we use a knowledge graph automatically constructed using a large language model (Wang et al., 2023), allowing for dynamic expansion to include details about each entity in the dataset. Lastly, we introduce a patient-specific graph, individually tailored for each relation prediction. This graph contains other events from the document we are analyzing and the relations between them. Such information can give the relation prediction model temporal context about the events it analyzes. We test the proposed methods on the i2b2 2012 dataset (Sun et al., 2013) containing medical discharge summaries.

2. Related Work

Temporal relation extraction aims to discern the temporal association between events within a text document, offering insights into the narrative behind the document. This paper focuses on extracting such relations from medical documents. Early approaches relied on rules (Gaizauskas et al., 2006; Dorr and Gaasterland, 2007) and traditional feature-based machine learning(Mani et al., 2006; Bethard, 2013; Chang et al., 2013), while recent years have seen a shift towards deep neural network-based models. Researchers have used two main architectures of deep neural networks. The first are long short-term memory networks (LSTM) exemplified by Tourille et al. (2017), Cheng and Miyao (2017), and Leeuwenberg and Moens (2018). Such networks build sentence embeddings by applying the LSTM layer over the tokens to generate token and sentence embeddings. They predict the relations based on the embeddings of the tokens corresponding to the events. Another approach is the use of pre-trained language models (PLM). Lin et al. (2019) mark events with special tokens and use a BERT network to encode the text into a vector representation. They classify the relation based on the sentence embedding that the PLM produces. Zhou et al. (2021) use a similar approach but enhance the results with soft logic regularization.

Proceedings of CL4Health 2024 @LREC-COLING-2024, pages 131–136

2.1. Use of Common Knowledge

One of the improvements for temporal relation extraction that has been proposed recently is the introduction of common sense knowledge. Such knowledge can help the model in cases where relations are not explicitly mentioned in the text, as it can be used to reason about how events might be related. Ning et al. (2018) developed a statistical resource for temporal relation extraction, which includes statistics about common relations between events. These statistics are then used in their model to improve predictions. They managed to improve the prediction accuracy of their model by 3% when introducing the resource (Ning et al., 2019). The idea was later expanded by Han et al. (2020). They enhance predictions using domain knowledge, which includes statistics, event types, and structured constraints based on the relationships between events.

Some researchers propose using knowledge graphs as an additional source of common knowledge to help with tasks similar to temporal relation extraction. Lin et al. (2023) propose a model for extracting disease relations from text using a bimodal architecture where a SciBERT model encodes text. In contrast, a heterogeneous graph attention network encodes additional information from a knowledge graph. They make the final relation prediction based on both sets of encodings. Similarly, Yasunaga et al. (2022) propose a pretrained DRAGON model, which simultaneously encodes and combines text and graph information. Such a network can be used to improve language processing using a knowledge graph.

2.2. Automatic Knowledge Graph Construction

When using knowledge graphs as a source of common knowledge to help information extraction models, an important problem that needs to be addressed is where to get a knowledge graph that contains all the necessary information. As the abilities of large language models to express common sense advanced, the idea of automatically constructing knowledge graphs was presented. Jiang et al. (2023) propose to use ChatGPT to generate knowledge graph triplets about drugs, conditions, and procedures. Such an approach enables the creation of specialized knowledge graphs containing much information about the relevant concepts.

3. Using Common Knowledge in Temporal Relation Extraction

Significant advancements in natural language processing (NLP) tasks have been observed in recent years, driven by the introduction of large pretrained language models. These models achieve high performance due to their extensive training on vast datasets, providing them with common sense knowledge. However, employing such models has drawbacks, including the need for large computational resources and lengthy training times. We propose that equipping models with specialized common sense knowledge tailored to specific tasks could yield comparable or superior performance using simpler models. To realize this concept, we present a bimodal model for temporal relation extraction. This model operates on two inputs: the text describing events and a knowledge graph offering general knowledge about these events. By encoding general knowledge in the knowledge graph, we provide essential information aiding the model in determining temporal relations. This knowledge inclusion is crucial, especially in medical contexts, where predictions often demand specialized knowledge that is not adequately represented in general datasets. Furthermore, predictions in the medical domain may rely on patient-specific histories, which must be incorporated at inference time.

Our preliminary experiments (Knez and Žitnik, 2024) show that introducing general knowledge to such a model can improve temporal relation prediction results. We used a knowledge graph containing automatically predicted temporal relations between other events in a document. Such relations give the model temporal context about the events we are observing. In our tests, the model managed to achieve new state-of-the-art results on the i2b2 2012 temporal relation extraction dataset (Sun et al., 2013) using this method (see Table 1). Based on this result, we believe that introducing additional general knowledge would improve the results even further.

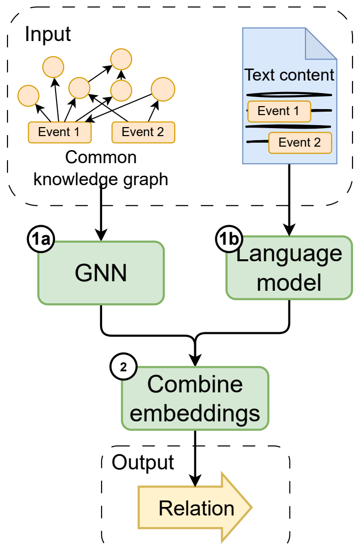

Figure 1: The architecture of the temporal relation extraction model.

3.1. Model Architecture

The temporal relation extraction model comprises two event encoders labeled 1a and 1b in Figure 1. The graph encoder (1a) utilizes a convolutional graph neural network to aggregate information from the local knowledge graph centered around a node representing an event from the text, thus computing an event embedding. Similarly, the text encoder (1b) employs the EntityBERT model to generate embeddings for tokens corresponding to the events in the text, capturing sentence meaning. In the final phase of the model (2 in Figure 1), both types of embeddings are merged to compute the final temporal relation prediction.

4. Building Knowledge Graphs

The proposed approach's main advantage is the model's ability to use common knowledge from a knowledge graph. We must determine how to get the knowledge graph to include such knowledge. In our research, we combine three separate kinds of knowledge graphs to provide as much relevant information as possible, shown in Figure 2. While PrimeKG and automatically generated graphs contain general relations between medical concepts, the patient-specific temporal knowledge graph contains temporal relations that determine in what order the medical events occurred to the patient. We can combine the graphs by linking events from the text to all three knowledge graphs.

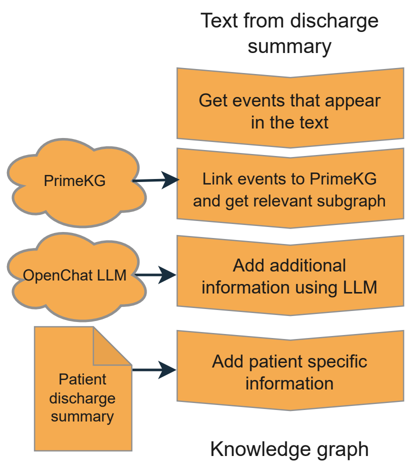

Figure 2: The proposed process of constructing a knowledge graph.

4.1. PrimeKG Knowledge Graph

Our first source of common knowledge is PrimeKG, a knowledge graph developed by Chandak et al. (2023). PrimeKG focuses on precision medicine and integrates data from 14 sources into a structured graph format. To leverage PrimeKG, we employ the SciSpacy library to identify UMLS concepts within event mentions. 85% of the nodes in PrimeKG are linked to their corresponding UMLS concepts. We utilize these connections to determine the PrimeKG node that best represents the medical event mentioned in the text using the UMLS identifiers that the SciSpacy library provides. We use the most likely identifier that appears in the PrimeKG knowledge graph. Subsequently, we extract the neighborhood of this node from PrimeKG and incorporate it as a subgraph in our model.

4.2. Automatically Generated Knowledge Graph

A problem with using the PrimeKG as a source of common knowledge is that a limited number of medical concepts are present in such graphs. We observed that 78% of the events described in clinical discharge summaries do not correspond to any node in PrimeKG. We propose generating larger common knowledge graphs using large language models to improve that. Our method is based on the one used by Jiang et al. (2023).

We used the OpenChat pre-trained large language model (Wang et al., 2023), which we prompt to generate knowledge graph triplets related to the concept in question. We generate prompts detailing that the model should prepare as many relations as possible about a medical concept in the format [entity 1, relation, entity 2]. We found that the model performs better if we include a small number of examples of responses that we expect in the prompt. In this way, we were able to create a knowledge graph that contains all of the events that occur in the dataset. While generating such a graph requires a large amount of resources, making it unpractical for some applications, it is possible to generate the graph in advance and use it at inference time.

4.3. Patient-Specific Knowledge Graph

In understanding medical notes concerning patients, their medical history holds significant importance. We aim to incorporate this information into our knowledge graph by automatically generating an additional graph based on patient data extracted from clinical discharge summaries. Utilizing a discharge summary as a base, our model predicts relations between different events within the document. These predicted relations serve as the basis for constructing a knowledge graph, providing temporal context for the events under consideration. This approach enhances our model's ability to comprehend the chronological sequence of events. We construct the final graph to be used by the model by combining nodes from all three graphs.

Table 1: Comparison of temporal relation extraction models on the i2b2 dataset without any common sense knowledge graph to the model with information from a patient-specific knowledge graph.

Model	Text
EntityBERT encoder	73.11%
EntityBERT + patient-specific KG	**74.16**%
BERT base (Ul Haq et al., 2022)	72.41%
BioBERT (Ul Haq et al., 2022)	73.60%
Alpaca model	35.01%

5. Results

We tested our approaches on the i2b2 2012 (Sun et al., 2013) dataset for temporal relation extraction from clinical discharge summaries. The dataset contains 14,256 training relations and 11,752 testing relations.

5.1. Improving Temporal Relation Prediction

When predicting a temporal relation in our preliminary tests, we used knowledge graphs containing predicted temporal relations between other events to give the model some temporal context as described in Section 4.3. While constructing such graphs increases the computational requirements of the model, it shows how the prediction accuracy can be improved using additional knowledge.

The F1 scores of the predictions in our experiments are shown in Table 1. We can see that adding a patient-specific knowledge graph improves prediction results by 1 pt. We rerun the training 20 times and found that the improvement is statistically significant, with a P-value of 0.039. We compare our model to two models proposed by Ul Haq et al. (2022) and a prompt-based approach using the Alpaca large language model. Our proposed bimodal model surpasses other state-of-the-art models on the same dataset.

5.2. Knowledge Graph Construction

Table 2: Comparing knowledge graphs based on PrimeKG with knowledge graphs generated using large language models.

	PrimeKG	LLM
i2b2 events present	22%	100%
Average number of nodes	133.7	27.6
Average node degree	2.00	1.78

We analyzed the graphs constructed using the procedure described in Section 4. We found that only 22% of the concepts from the i2b2 dataset are successfully linked to PrimeKG. This happens because we need a very broad knowledge graph that contains concepts from a large variety of areas. As a result, such a knowledge graph cannot contain all of the concepts from each area. This represents a large problem for our relation extraction model, as it has no common sense information for most of the events it encounters in a clinical document. For this reason, we enrich the knowledge graph using nodes and relations generated using a large language model.

Using a large language model to construct the knowledge graphs, we generated a knowledge graph about each concept from the i2b2 dataset. While generated graphs can contain errors as humans did not curate them, we believe they can be much more useful as they contain information about each concept.

We compare the knowledge graphs extracted from PrimeKG to those automatically generated using a large language model in Table 2. We found that the subgraphs we extracted from PrimeKG are generally larger than those created by a large language model. While a larger graph is generally beneficial, the graphs created using LLMs might be more useful for helping a machine learning model as they contain only the highly relevant information. The average degree of nodes in both graphs is quite similar at around two. Based on the analysis results, we recommend using a manually curated knowledge graph like PrimeKG enriched using automatically determined relations.

6. Conclusion

In our study, we demonstrated that adding additional knowledge to the model for temporal relations extraction can improve its performance, allowing it to achieve state-of-the-art results. While we only tested the performance gains when using knowledge graphs containing information from the active document, we believe including additional information could improve the results even further. When constructing knowledge graphs, we propose to enrich information from a large knowledge graph like PrimeKG by automatically generating relations using a large language model. Our results show that adding such relations greatly increases the coverage of a knowledge graph for the relation extraction task. The research is ongoing, and we expect to evaluate the model's performance using the proposed knowledge graph in future work.

7. Acknowledgements

This work was supported by Slovenian Research Agency grant number H017002 (Young Researchers grant).

8. Bibliographical References

Steven Bethard. 2013. Cleartk-timeml: A minimalist approach to tempeval 2013. In *Second joint conference on lexical and computational semantics (* SEM), volume 2: proceedings of the seventh international workshop on semantic evaluation (SemEval 2013)*, pages 10–14.

Payal Chandak, Kexin Huang, and Marinka Zitnik. 2023. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1):67.

Yung-Chun Chang, Hong-Jie Dai, Johnny Chi-Yang Wu, Jian-Ming Chen, Richard Tzong-Han Tsai, and Wen-Lian Hsu. 2013. Tempting system: a hybrid method of rule and machine learning for temporal relation extraction in patient discharge summaries. *Journal of Biomedical Informatics*, 46:S54–S62.

Fei Cheng and Yusuke Miyao. 2017. Classifying temporal relations by bidirectional lstm over dependency paths. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1–6.

Bonnie J Dorr and Terry Gaasterland. 2007. Exploiting aspectual features and connecting words for summarization-inspired temporal-relation extraction. *Information processing & management*, 43(6):1681–1704.

Rob Gaizauskas, Henk Harkema, Mark Hepple, and Andrea Setzer. 2006. Task-oriented extraction of temporal information: The case of clinical narratives. In *Thirteenth International Symposium On Temporal Representation And Reasoning (time'06)*, pages 188–195. IEEE.

Rujun Han, Yichao Zhou, and Nanyun Peng. 2020. Domain knowledge empowered structured neural net for end-to-end event temporal relation extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5717–5729.

Pengcheng Jiang, Cao Xiao, Adam Cross, and Jimeng Sun. 2023. Graphcare: Enhancing healthcare predictions with open-world personalized knowledge graphs. *arXiv preprint arXiv:2305.12788*.

Timotej Knez and Slavko Žitnik. 2024. Multi-modal learning for temporal relation extraction in clinical texts. *Journal of the American Medical Informatics Association*.

Artuur Leeuwenberg and Marie Francine Moens. 2018. Temporal information extraction by predicting relative time-lines. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1237–1246.

Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard, and Guergana Savova. 2019. A bert-based universal model for both within-and cross-sentence clinical temporal relation extraction. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 65–71.

Yucong Lin, Keming Lu, Sheng Yu, Tianxi Cai, and Marinka Zitnik. 2023. Multimodal learning on graphs for disease relation extraction. *Journal of Biomedical Informatics*, page 104415.

Inderjeet Mani, Marc Verhagen, Ben Wellner, Chungmin Lee, and James Pustejovsky. 2006. Machine learning of temporal relations. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 753–760.

Qiang Ning, Sanjay Subramanian, and Dan Roth. 2019. An improved neural baseline for temporal relation extraction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6203–6209.

Qiang Ning, Hao Wu, Haoruo Peng, and Dan Roth. 2018. Improving temporal relation extraction with a globally acquired statistical resource. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 841–851.

Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *Journal of the American Medical Informatics Association*, 20(5):806–813.

Julien Tourille, Olivier Ferret, Aurelie Neveol, and Xavier Tannier. 2017. Neural architecture for temporal relation extraction: A bi-lstm approach for detecting narrative containers. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 224–230.

H Ul Haq, V Kocaman, D Talby, APB Veyseh, F Dernoncourt, TH Nguyen, and VD Lai. 2022. Deeper clinical document understanding using relation extraction. In *Proceedings of the workshop on scientific document understanding*. CEUR.

Guan Wang, Sijie Cheng, Xianyuan Zhan, Xiangang Li, Sen Song, and Yang Liu. 2023. Openchat: Advancing open-source language models with mixed-quality data. *arXiv preprint arXiv:2309.11235*.

Michihiro Yasunaga, Antoine Bosselut, Hongyu Ren, Xikun Zhang, Christopher D Manning, Percy S Liang, and Jure Leskovec. 2022. Deep bidirectional language-knowledge graph pretraining. *Advances in Neural Information Processing Systems*, 35:37309–37323.

Yichao Zhou, Yu Yan, Rujun Han, Harry J Caufield, Kai-Wei Chang, Yizhou Sun, Peipei Ping, and Wei Wang. 2021. Clinical temporal relation extraction with probabilistic soft logic regularization and global inference. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'21)*.