Answering Causal Questions with Reinforcement Learning

Lukas Blübaum

July 1, 2024

Version: Proposal

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Faculty for Computer Science, Electrical Engineering and Mathematics
Department of Computer Science
DICE Research Group

Proposal

Answering Causal Questions with Reinforcement Learning

Lukas Blübaum

1. Reviewer Prof. Dr. Axel-Cyrille Ngonga Ngomo
Department of Computer Science
Paderborn University

2. Reviewer Jun.-Prof. Dr. Sebastian Peitz
Department of Computer Science
Paderborn University

Supervisor Dr. Stefan Heindorf

July 1, 2024

Lukas Blübaum
Answering Causal Questions with Reinforcement Learning
Proposal, July 1, 2024
Reviewers: Prof. Dr. Axel-Cyrille Ngonga Ngomo and Jun.-Prof. Dr. Sebastian Peitz
Supervisor: Dr. Stefan Heindorf

Paderborn University
DICE Research Group
Department of Computer Science
Faculty for Computer Science, Electrical Engineering and Mathematics
Warburger Straße 100
33098 Paderborn

Contents

1 Introduction

Causal question answering addresses the problem of determining the causal relations between given causes and effects [Kay+20; DAB21]. This could involve examining whether a causal relation exists or how a causal relation can be explained. For example, causal questions like *"What causes pneumonia?"* search for the cause of a given effect. Instead, more involved questions like *"How does pneumonia cause death?"* might ask for an explanation of a causal relation. Nowadays, the necessity to answer causal questions arises in various domains, for example, users often seek answers to causal questions from virtual assistants like Alexa or from search engines [Ngu+16; Hei+20]. Similarly, causal questions and explanations for causal relations are important for argumentation [Wal07; Hab+18] and automated decision-making [Has+19; Kay+20; Hei+20]. When possible, following the causality chains that led to an answer can provide further explanations and allow for a deeper understanding of the causal relations in question.

Therefore, the literature started to introduce approaches for causal question answering [Sha+16; Has+19; Kay+20; Dal21]. However, they only consider simple questions and are typically focused on narrow types of questions. For instance, some only focus on binary questions [Has+19; Kay+20] while others focus on multiple-choice questions [Sha+16; Dal21]. Hence, we plan to gradually consider more complex questions and build a unified causal question answering system that can answer questions of different types. Additionally, existing approaches identified the problem of a lack of large, high quality datasets for causal relations [Sha+16; Has+19; Kay+20]. Consequently, the authors resorted to creating their own datasets from sources like news articles or Wikipedia to support the identification of causal relations to answer causal questions. Thus, the introduction of CauseNet [Hei+20], a large-scale knowledge graph consisting of causal relations with context information, provides new opportunities to build effective causal question answering systems. Accordingly, we can apply various approaches from the knowledge graph literature that utilize the structure and rich semantics.

In particular, reinforcement learning was successfully applied to knowledge graphs on different tasks such as link prediction [XHW17], fact checking [Das+18], or conversational question answering [KSW21]. Hence, we propose to model the causal question answering task as a sequential decision problem similar to the aforementioned approaches. Therefore, we train a reinforcement learning agent that learns to walk over the graph to find good inference paths to answer causal questions. Starting at entities found in the question, the agent searches through the graph until it finds an answer entity. The resulting reasoning capabilities are beneficial for questions whose answers require a long chain of multiple hops [XHW17; Das+18; WD21]. Additionally, the agent prunes the search space since it only considers small neighborhoods around entities of the question [Das+18; Qiu+20].

Using this approach, the agent is able to answer binary questions or questions that ask for a simple cause or effect. However, it might be hard to answer complex questions which expect more detailed answers. Consequently, for these question types, we propose to take the paths extracted by the agent and feed them into a pre-trained language model [Vas+17; Kha+20] as additional context. Dalal et al. [DAB21; Dal21] already consider a similar setup applied on CauseNet. However, they extract relevant causal relations via string matching while we use the paths found by the RL agent. Notably, the paths that led to an answer allow for post-hoc explanations and enable the user to follow the reasoning chain. In case of CauseNet [Hei+20], we can also take the context information into account. Thus, for each relation on a path, we can provide supplemental explanations such as the original sentences and source URLs.

Overall, this proposal is structured as follows: First, we introduce related approaches on causal question answering, causal knowledge graphs, and reinforcement learning on knowledge graphs. Afterward, we give the concrete problem definition, including a breakdown of successive steps to answer causal questions of increasing difficulty. Besides, we discuss potential improvements and solutions and briefly describe the evaluation setup. Finally, we provide a preliminary thesis structure and a Gantt chart detailing the time schedule.

2 Related Work

In the following, we summarize related work. First, we provide an overview of approaches for causal question answering. These approaches mainly employ pattern-based searches, embeddings, and language models. To the best of our knowledge, we are the first to use reinforcement learning for causal question answering. Afterward, we present several knowledge graphs that focus on causal and commonsense knowledge. Lastly, we discuss approaches that apply reinforcement learning to reasoning and question answering tasks on knowledge graphs.

Causal Question Answering Currently, only few approaches directly focus on the causal question answering task. Most of these approaches focus on binary questions, i.e., questions such as "*Does X cause Y?*" which expect a yes or no answer. Kayesh et al. [Kay+20] model the task as a transfer learning approach. Therefore, they extract cause-effect pairs from news articles via causal cue words. Subsequently, they transform the pairs into sentences of the form "*X may cause Y*" and use them to finetune BERT [Dev+19]. Similarly, Hassanzadesh et al. [Has+19] employ large-scale text mining to answer binary causal questions. They introduce several unsupervised approaches ranging from simple string matching to embeddings computed via BERT. Each approach uses the text corpora to find evidence for a potential causal relation. Afterward, the evidence is used to compute a score that results in a yes or no answer depending on a threshold value.

Instead, Sharp et al. [Sha+16] consider multiple-choice questions of the form "*What causes X?*". First, they mine cause-effect pairs from Wikipedia via syntactic patterns and train an embedding model to capture the semantics between the pairs. At inference time, they compute the embedding similarity between the question and each answer candidate. The approach by Dalal et al. [DAB21; Dal21] is most similar to ours. Given a question, they apply string matching to extract relevant causal relations from CauseNet [Hei+20]. Subsequently, they provide the question with the causal relations to a language model as additional context. In our work, we consider a similar idea for complex questions by selecting relevant paths via a reinforcement learning agent.

Another approach that combines knowledge graphs and language models is QA-GNN [Yas+21]. Specifically, QA-GNN creates an augmented graph by scoring entities and relations with a pre-trained language model according to their relevance to the question. Then, QA-GNN applies a graph attention network (GAT) [Vel+18] on the augmented graph to answer the question. Likewise, we apply the RL agent on an augmented graph to provide a better reward signal, similar to the reward shaping strategies by prior RL approaches on knowledge graphs [LSX18; Qiu+20].

Causal Knowledge Graphs ConceptNet [SCH17] is a general knowledge graph consisting of relations between natural language terms, which contains 36 relations, including a *Causes* relation. CauseNet [Hei+20] and Cause Effect Graph [Li+20] take a narrower view and specifically focus on causal relations extracted via linguistic patterns from web sources like Wikipedia and ClueWeb12 [1].

In contrast, ATOMIC [Sap+19] focuses on inferential knowledge. Specifically, ATOMIC consists of *"If-Event-Then-X"* relations that are based on social interactions or events in the real world. ATOMIC$^{20}_{20}$ [Hwa+21] selects relations from ATOMIC and ConceptNet to create an improved graph while adding more relations via crowdsourcing. Instead, West et al. [Wes+21] automate the curation of inferential relations by clever prompting of a language model. Finally, CSKG [ISZ21] builds a consolidated graph combining seven knowledge graphs, including ConceptNet and ATOMIC. For this thesis, we focus on CauseNet and potentially Cause Effect Graph because most of the other knowledge graphs either focus on inferential knowledge or are not limited to causal relations. Although, if time permits, making use of a combination of multiple graphs might be a potential improvement.

Knowledge Graph Reasoning with Reinforcement Learning In recent years, reinforcement learning on knowledge graphs has been applied in link prediction [Das+18], fact checking [XHW17], or question answering [Qiu+20]. Given a source and a target entity, DeepPath [XHW17] learns to find paths between them. In the first step, DeepPath is trained in a supervised manner on paths found by BFS. Afterward, DeepPath applies REINFORCE [Wil92] policy gradients to improve the policy further. During inference time, the paths are used to predict links between entities or check the validity of triples. Subsequently, MINERVA [Das+18] improves on DeepPath by introducing an LSTM [HS97] into the policy network to account for the path history. Moreover, MINERVA does not require knowledge of the target entity and is trained end-to-end without supervision at the start. Lin et al. [LSX18] propose two

[1] https://lemurproject.org/clueweb12/

improvements for MINERVA. First, they apply reward shaping by scoring the paths with a pre-trained KG embedding model [Det+18] to reduce the problem of sparse rewards. Second, they introduce a technique called action dropout, which randomly disables edges at each step. Action dropout serves as additional regularization and helps the agent learn diverse paths.

Contrary to these approaches, M-Walk [She+18] does not rely on REINFORCE but instead looks at the problem from a model-based viewpoint. Like AlphaZero [Sil+18], M-Walk applies Monte Carlo Tree Search (MCTS) as a policy improvement operator. Thus, at each step, M-Walk applies MCTS to produce trajectories of an improved policy and subsequently trains the current policy to imitate the improved one. GaussianPath [WD21] takes a Bayesian view of the problem and represents each entity by a gaussian distribution to better model uncertainty.

While all these approaches focus on link prediction or fact checking, there has been some work on natural language question answering. Qiu et al. [Qiu+20] introduce the Stepwise Reasoning Network (SRN). They observe that for questions requiring multiple hops, different parts of the question should be important at each time step. Therefore, they introduce an attention mechanism into the policy network to enable the agent to attend to different parts of the question at every step. In turn, this should help the agent to focus on the aspect of the question that is important for the current decision. Additionally, SRN utilizes extra reward signals by considering the semantic similarities between the question and the action history via a potential-based reward function [NHR99]. CONQUER [KSW21] employs REINFORCE policy gradients for conversational question answering. Multiple agents search through the graph in parallel and aggregate their answers at the end. Importantly, the reward depends on user feedback. A positive reward corresponds to a new question, and a negative reward to a reformulation of the last question.

Approach

3

3.1 Overview

First, we present a general overview of the approach based on related work. This serves as a starting point for our work to build and improve upon while acting as an initial baseline. Following previous approaches for RL on KGs [Das+18; Qiu+20; KSW21], we model the causal question answering task as a sequential decision problem. Therefore, we define a Markov Decision Process (MDP) as a 4-tuple $(\mathcal{S}, \mathcal{A}, \delta, \mathcal{R})$ on the knowledge graph, where $\mathcal{S}$ represents the states (nodes), $\mathcal{A}$ the actions (edges), δ the transition dynamics, and $\mathcal{R}$ the reward. Additionally, we introduce a policy network $\pi_\theta(a|s)$ which generates a distribution over actions given a state s, i.e., the probability for each outgoing edge at the current node. For the initial baseline, we model π_θ as an MLP and later extend this to use LSTMs [HS97] or GRUs [Cho+14] as done in MINERVA [Das+18] and SRN [Qiu+20]. Currently, most approaches apply REINFORCE [Wil92] policy gradients [WD21]:

$$\mathbb{E}_{\pi_\theta}[\nabla_\theta \log \pi_\theta(a|s)\ \mathcal{R}] \tag{3.1}$$

During training, we run multiple rollouts for each example and update π_θ via equation (3.1). For each rollout, the agent receives a terminal reward $\mathcal{R}$ of 1 if the answer is correct and 0 otherwise. Moreover, we add inverse relations to the graph so the agent can undo wrong decisions [XHW17; Das+18]. In the context of CauseNet, this also implies that the agent can move from an effect to its cause.

Causal Questions Following related work, we categorize causal questions into different types. This allows us to split the implementation into successive phases where we consider questions of increasing difficulty in each one. First, we start with binary causal questions of the form *"Does X cause Y?"* which expect a yes or no answer. Afterward, we move to Cause/Effect questions like *"What causes X?"*, which ask for a cause of an effect or vice versa. Multiple choice questions can also be grouped into this category. Finally, we consider more complex causal questions like *"How does pneumonia cause death?"*. In this example, the question asks for an explanation of

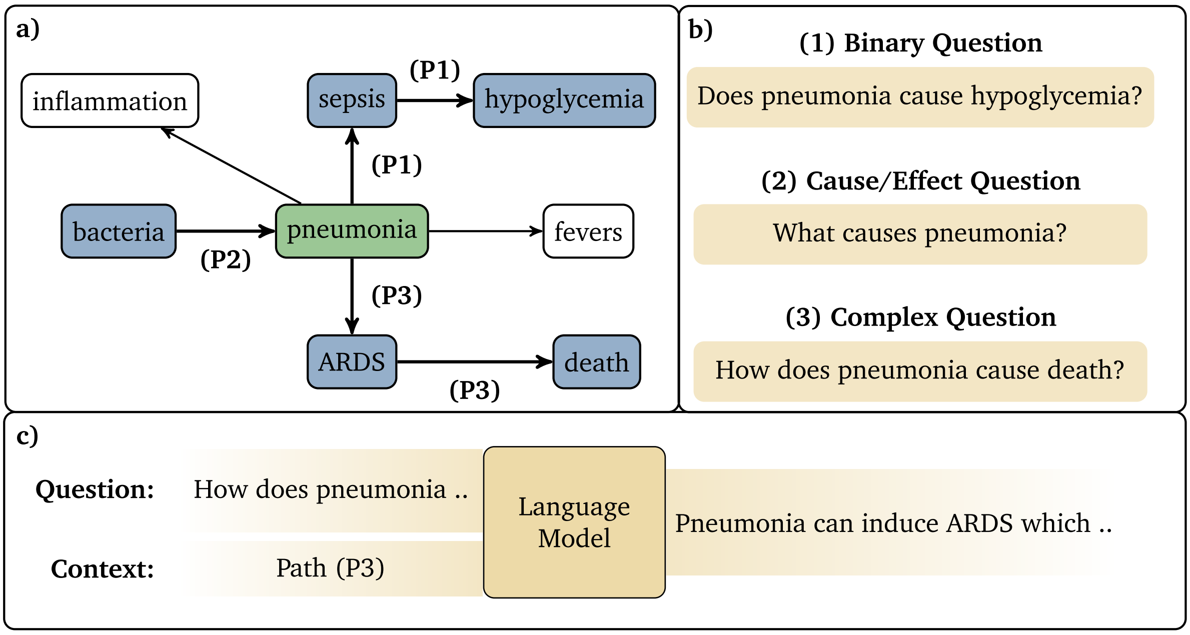

Fig. 3.1: A snapshot from CauseNet **a)** together with concrete examples for each question type **b)**. The annotated paths in the graph correspond to possible answers for each question, respectively. Note how question 3) asks for an explanation of a causal relation. In this case, that pneumonia causes death by inducing acute respiratory distress syndrome (ARDS). For this question type, we take the path(s) found by the agent and feed them into a language model as additional context **c)**.

how a cause and effect relate to each other. Figure 3.1 illustrates a concrete example for each question type together with a snapshot from CauseNet [Hei+20] which is able to answer the questions.

3.2 Potential Directions and Improvements

Representation of States and Actions Finding suitable representations for the states (nodes) and actions (edges) is one of the essential aspects to consider. Previous approaches have already explored various directions in that regard. DeepPath [XHW17] encoded the states and actions via pre-trained KG embeddings models. In contrast, approaches like MINERVA [Das+18] optimized them during training, while other approaches embedded the labels of nodes and edges through word embeddings, e.g., CONQUER [KSW21]. Since CauseNet also contains labels for each entity, we plan to follow approaches like CONQUER and use word embeddings to represent them. Starting with simple GloVe [PSM14] embeddings for the first baseline and subsequently moving to encoders like BERT [Dev+19]. Additionally, we plan to experiment with initialization from word embeddings and further optimization during training, which was not considered by related approaches yet.

As described in Section 3.1, CauseNet contains only one edge type. Therefore, each action is entirely defined by the target node. Thus, we can decide whether to use one or two representations for each node, i.e., one for the node as state and action, respectively. Instead, another idea would be to use the provenance information of each relation [Hei+20] to define the actions. Here, we could use the patterns or original sentences through which the relation was found to create action representations.

Reward Shaping Due to the terminal reward at the end, the agent only receives feedback about the final answer. Specifically, there is no feedback about the quality of the path itself or the actions taken. To mitigate this problem, Lin et al. [LSX18] introduced additional reward signals by scoring the final relation on a path with a pre-trained KG embedding model. Likewise, SRN [Qiu+20] added a potential-based reward function that scores the action history at each step according to its semantic similarity with the question.

Similarly, we propose to apply the agent on an augmented graph. As done by QA-GNN [Yas+21], we use a language model to score each node according to its relevance to the question, including context information if available. Additionally, we can include the provenance information of each relation into the score computation. Due to runtime constraints, the score computation can be limited to a subgraph around the entities found in the question. Accordingly, the score serves as additional reward signal to provide feedback about the quality of the actions on a path.

Answering Complex Causal Questions In Section 3.1, we introduced three types of causal questions. Generally, binary questions can be formulated as a query of the form $(cause, mayCause, effect)$? while Cause/Effect questions can be formulated as queries like $(cause, mayCause, ?)$ or $(?, mayCause, effect)$. Therefore, to provide an answer, the agent can walk over the graph and search for a suitable target node. However, we have to extend this setup for more challenging questions that ask for complex causal relations or an explanation of a causal relation. Consequently, we plan to introduce a language model into our approach, which receives paths from the agent as additional context. Given a question, the agent walks over the graph as before. Afterward, we take the path, including the provenance information for each relation, and feed it into the language model together with the original question. Figure 3.1 c) shows an example. A possible choice for the language model is UnifiedQA [Kha+20; KKH22], which was trained on various question answering datasets of different domains.

Algorithm Most existing approaches apply REINFORCE [Wil92] policy gradients to train the agent. Thus, one straightforward direction is the application of more sophisticated policy gradient methods. Therefore, we plan to experiment with different methods, e.g., Actor-Critic algorithms [Mni+16] for further variance reduction. Existing approaches argued in favor of policy gradients methods because of the huge action spaces in these problems [XHW17; KSW21]. However, due to the unique characteristic that CauseNet contains only one edge type, it might be interesting to revisit ideas from value-based methods. Particularly, learning a state-state value function $Q(s, s')$ [Edw+20] might be interesting for this setting. If time permits, we also plan to look into model-based methods, as done in M-Walk [She+18].

3.3 Evaluation Setup

First, we have to create datasets for the three question types. As a starting point, we take Webis-CausalQA-22[1], a collection of 1.1M causal questions extracted from 10 popular question answering datasets, including MS MARCO [Ngu+16] and SQuAD v2.0 [RJL18].

Next, we compare our approach to different baselines, starting with a simple approach that searches all hops up to length two on CauseNet. Furthermore, we consider different approaches from related work, particularly the approach by Dalal et al. [DAB21; Dal21], which also uses CauseNet. Likewise, we compare our approach to UnifiedQA without the additional context provided by our agent. The selection of baselines also depends on the availability of their code.

To better understand our approach, we will run several ablations. These include different graph augmentations, reinforcement learning algorithms, state-action representations, and network architectures. Possibly also scaling experiments on the number of parameters, including UnifiedQA models of different sizes. Similarly, we will compare performance on CauseNet-Precision and CauseNet-Full [Hei+20] and potentially other graphs as well. Additionally, we will conduct runtime experiments for inference and training. Finally, we will analyze the agents' paths concerning their reasonability and interpretability.

[1] https://webis.de/data/webis-causalqa-22

4 Work Plan

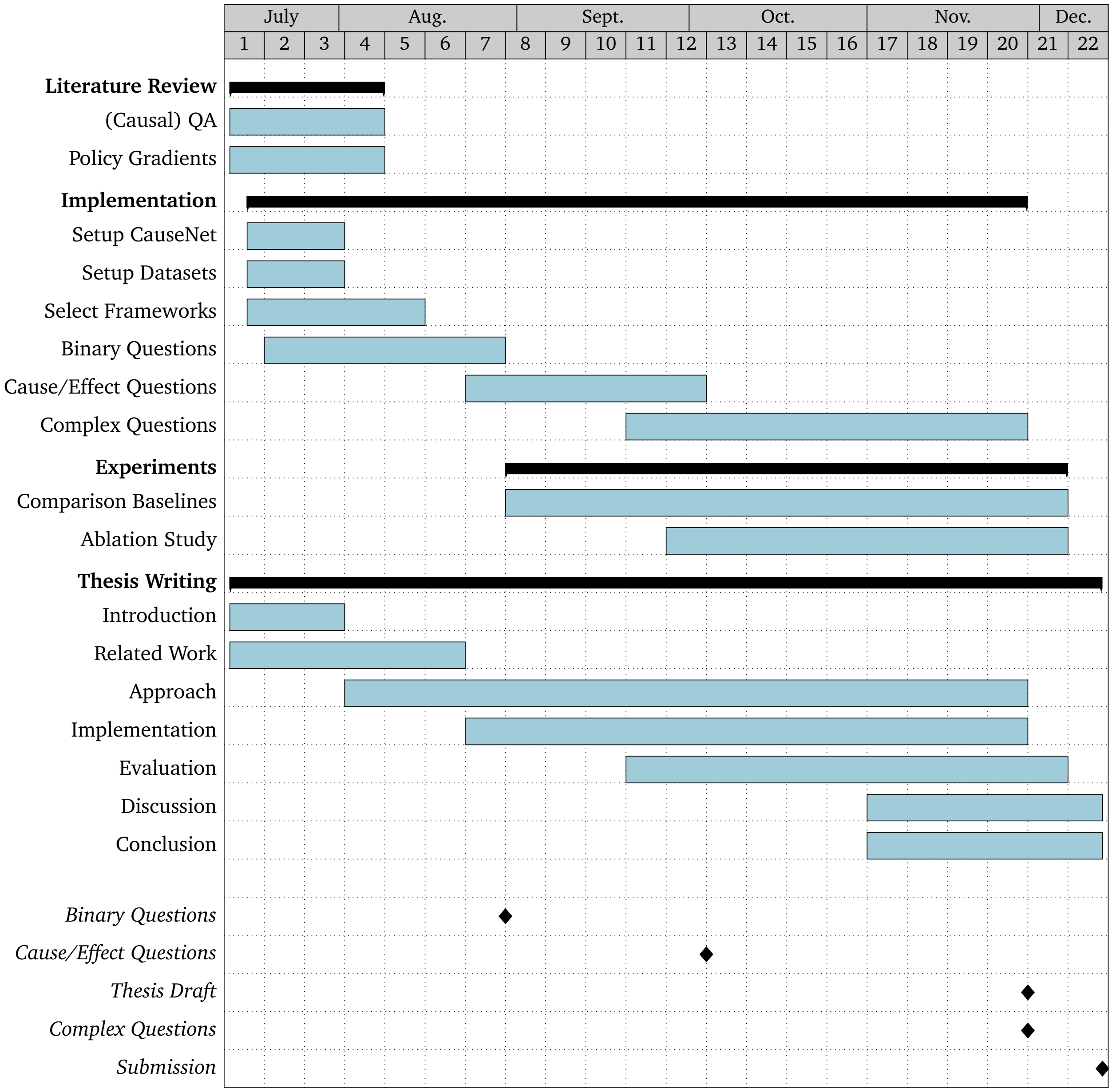

Fig. 4.1: Work plan of the thesis.

5 Thesis Structure

1. Introduction
2. Related Work
    2.1 Causal Question Answering
    2.2 Causal Knowledge Graphs
    2.3 Knowledge Graph Reasoning with Reinforcement Learning
3. Approach
    3.1 Preliminaries
        3.1.1 CauseNet
        3.1.2 Reinforcement Learning
        3.1.3 Language Models
    3.2 Problem Definition – Environment
    3.3 Approach Description
4. Implementation
    4.1 Project Structure
    4.2 Frameworks
5. Evaluation
    5.1 Datasets
    5.2 Comparison to Baselines
    5.3 Ablation Study
6. Discussion
7. Conclusion and Future Work

Bibliography

[Cho+14] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, et al. “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*. Ed. by Alessandro Moschitti, Bo Pang, and Walter Daelemans. ACL, 2014, pp. 1724–1734 (cit. on p. 7).

[Dal21] Dhairya Dalal. “Knowledge augmented language models for causal question answering”. In: *CEUR Workshop Proceedings* 3005 (2021), pp. 17–24 (cit. on pp. 1–3, 10).

[DAB21] Dhairya Dalal, Mihael Arcan, and Paul Buitelaar. “Enhancing Multiple-Choice Question Answering with Causal Knowledge”. In: *Proceedings of Deep Learning Inside Out (DeeLIO): The 2nd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*. Online: Association for Computational Linguistics, June 2021, pp. 70–80 (cit. on pp. 1–3, 10).

[Das+18] Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, et al. “Go for a Walk and Arrive at the Answer: Reasoning Over Paths in Knowledge Bases using Reinforcement Learning”. In: *ICLR*. 2018 (cit. on pp. 1, 2, 4, 7, 8).

[Det+18] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. “Convolutional 2D Knowledge Graph Embeddings”. In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*. AAAI’18/IAAI’18/EAAI’18. New Orleans, Louisiana, USA: AAAI Press, 2018 (cit. on p. 5).

[Dev+19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 4171–4186 (cit. on pp. 3, 8).

[Edw+20] Ashley D. Edwards, Himanshu Sahni, Rosanne Liu, et al. “Estimating Q(s,s’) with Deep Deterministic Dynamics Gradients”. In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 2825–2835 (cit. on p. 10).

[Hab+18] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. "The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*. Ed. by Marilyn A. Walker, Heng Ji, and Amanda Stent. Association for Computational Linguistics, 2018, pp. 1930–1940 (cit. on p. 1).

[Has+19] Oktie Hassanzadeh, Debarun Bhattacharjya, Mark Feblowitz, et al. "Answering Binary Causal Questions Through Large-Scale Text Mining: An Evaluation Using Cause-Effect Pairs from Human Experts". In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, July 2019, pp. 5003–5009 (cit. on pp. 1, 3).

[Hei+20] Stefan Heindorf, Yan Scholten, Henning Wachsmuth, Axel-Cyrille Ngonga Ngomo, and Martin Potthast. "CauseNet: Towards a Causality Graph Extracted from the Web". In: *CIKM*. ACM, 2020 (cit. on pp. 1–4, 8–10).

[HS97] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Comput.* 9.8 (Nov. 1997), pp. 1735–1780 (cit. on pp. 4, 7).

[Hwa+21] Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, et al. "(Comet-) Atomic 2020: On Symbolic and Neural Commonsense Knowledge Graphs". In: *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 2021, pp. 6384–6392 (cit. on p. 4).

[ISZ21] Filip Ilievski, Pedro A. Szekely, and Bin Zhang. "CSKG: The CommonSense Knowledge Graph". In: *The Semantic Web - 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings*. Ed. by Ruben Verborgh, Katja Hose, Heiko Paulheim, et al. Vol. 12731. Lecture Notes in Computer Science. Springer, 2021, pp. 680–696 (cit. on p. 4).

[KSW21] Magdalena Kaiser, Rishiraj Saha Roy, and Gerhard Weikum. "Reinforcement Learning from Reformulations in Conversational Question Answering over Knowledge Graphs". In: *SIGIR*. 2021 (cit. on pp. 1, 5, 7, 8, 10).

[Kay+20] Humayun Kayesh, Md. Saiful Islam, Junhu Wang, et al. "Answering Binary Causal Questions: A Transfer Learning Based Approach". In: *2020 International Joint Conference on Neural Networks (IJCNN)*. 2020, pp. 1–9 (cit. on pp. 1, 3).

[KKH22] Daniel Khashabi, Yeganeh Kordi, and Hannaneh Hajishirzi. "UnifiedQA-v2: Stronger Generalization via Broader Cross-Format Training". In: *CoRR* abs/2202.12359 (2022). arXiv: `2202.12359` (cit. on p. 10).

[Kha+20] Daniel Khashabi, Sewon Min, Tushar Khot, et al. “UnifiedQA: Crossing Format Boundaries With a Single QA System”. In: *Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020*. Ed. by Trevor Cohn, Yulan He, and Yang Liu. Vol. EMNLP 2020. Findings of ACL. Association for Computational Linguistics, 2020, pp. 1896–1907 (cit. on pp. 2, 10).

[Li+20] Zhongyang Li, Xiao Ding, Ting Liu, J. Edward Hu, and Benjamin Van Durme. “Guided Generation of Cause and Effect”. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*. Ed. by Christian Bessiere. ijcai.org, 2020, pp. 3629–3636 (cit. on p. 4).

[LSX18] Xi Victoria Lin, Richard Socher, and Caiming Xiong. “Multi-Hop Knowledge Graph Reasoning with Reward Shaping.” In: *EMNLP*. Ed. by Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun’ichi Tsujii. Association for Computational Linguistics, 2018, pp. 3243–3253 (cit. on pp. 4, 9).

[Mni+16] Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, et al. “Asynchronous Methods for Deep Reinforcement Learning”. In: *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*. Ed. by Maria-Florina Balcan and Kilian Q. Weinberger. Vol. 48. JMLR Workshop and Conference Proceedings. JMLR.org, 2016, pp. 1928–1937 (cit. on p. 10).

[NHR99] Andrew Y. Ng, Daishi Harada, and Stuart J. Russell. “Policy Invariance Under Reward Transformations: Theory and Application to Reward Shaping”. In: *Proceedings of the Sixteenth International Conference on Machine Learning*. ICML ’99. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999, pp. 278–287 (cit. on p. 5).

[Ngu+16] Tri Nguyen, Mir Rosenberg, Xia Song, et al. “MS MARCO: A Human Generated MAchine Reading COmprehension Dataset”. In: *Proceedings of the Workshop on Cognitive Computation: Integrating neural and symbolic approaches 2016 co-located with the 30th Annual Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, December 9, 2016*. Ed. by Tarek Richard Besold, Antoine Bordes, Artur S. d’Avila Garcez, and Greg Wayne. Vol. 1773. CEUR Workshop Proceedings. CEUR-WS.org, 2016 (cit. on pp. 1, 10).

[PSM14] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. “Glove: Global Vectors for Word Representation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*. Ed. by Alessandro Moschitti, Bo Pang, and Walter Daelemans. ACL, 2014, pp. 1532–1543 (cit. on p. 8).

[Qiu+20] Yunqi Qiu, Yuanzhuo Wang, Xiaolong Jin, and Kun Zhang. “Stepwise Reasoning for Multi-Relation Question Answering over Knowledge Graph with Weak Supervision”. In: WSDM ’20. Houston, TX, USA: Association for Computing Machinery, 2020, pp. 474–482 (cit. on pp. 2, 4, 5, 7, 9).

[RJL18] Pranav Rajpurkar, Robin Jia, and Percy Liang. “Know What You Don’t Know: Unanswerable Questions for SQuAD”. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*. Ed. by Iryna Gurevych and Yusuke Miyao. Association for Computational Linguistics, 2018, pp. 784–789 (cit. on p. 10).

[Sap+19] Maarten Sap, Ronan Le Bras, Emily Allaway, et al. “ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning”. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 3027–3035 (cit. on p. 4).

[Sha+16] Rebecca Sharp, Mihai Surdeanu, Peter Jansen, Peter Clark, and Michael Hammond. “Creating Causal Embeddings for Question Answering with Minimal Supervision”. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 138–148 (cit. on pp. 1, 3).

[She+18] Yelong Shen, Jianshu Chen, Po-Sen Huang, Yuqing Guo, and Jianfeng Gao. “M-Walk: Learning to Walk over Graphs using Monte Carlo Tree Search”. In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, et al. Vol. 31. Curran Associates, Inc., 2018 (cit. on pp. 5, 10).

[Sil+18] David Silver, Thomas Hubert, Julian Schrittwieser, et al. “A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play”. In: *Science* 362.6419 (2018), pp. 1140–1144. eprint: `https://www.science.org/doi/pdf/10.1126/science.aar6404` (cit. on p. 5).

[SCH17] Robyn Speer, Joshua Chin, and Catherine Havasi. “ConceptNet 5.5: An Open Multilingual Graph of General Knowledge”. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*. Ed. by Satinder Singh and Shaul Markovitch. AAAI Press, 2017, pp. 4444–4451 (cit. on p. 4).

[Vas+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. “Attention is All you Need”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. Von Luxburg, S. Bengio, et al. Vol. 30. Curran Associates, Inc., 2017 (cit. on p. 2).

[Vel+18] Petar Veličković, Guillem Cucurull, Arantxa Casanova, et al. “Graph Attention Networks”. In: *International Conference on Learning Representations* (2018) (cit. on p. 4).

[Wal07] Douglas N. Walton. *Dialog theory for critical argumentation*. Vol. 5. Controversis. Benjamin/Cummings, 2007 (cit. on p. 1).

[WD21] Guojia Wan and Bo Du. “GaussianPath:A Bayesian Multi-Hop Reasoning Framework for Knowledge Graph Reasoning”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.5 (May 2021), pp. 4393–4401 (cit. on pp. 2, 5, 7).

[Wes+21] Peter West, Chandra Bhagavatula, Jack Hessel, et al. “Symbolic Knowledge Distillation: from General Language Models to Commonsense Models”. In: *CoRR* abs/2110.07178 (2021). arXiv: `2110.07178` (cit. on p. 4).

[Wil92] Ronald J. Williams. “Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning”. In: *Mach. Learn.* 8.3–4 (May 1992), pp. 229–256 (cit. on pp. 4, 7, 10).

[XHW17] Wenhan Xiong, Thien Hoang, and William Yang Wang. “DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, Sept. 2017, pp. 564–573 (cit. on pp. 1, 2, 4, 7, 8, 10).

[Yas+21] Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. “QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering”. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, June 2021, pp. 535–546 (cit. on pp. 4, 9).