Visual Causal Question and Answering with Knowledge Graph Link Prediction*

Utkarshani Jaimini[1,*], Cory Henson[2] and Amit Sheth[1]

[1]Artificial Intelligence Institute, University of South Carolina, Columbia, SC, USA
[2]Bosch Center for Artificial Intelligence, Pittsburgh, PA, USA

Abstract

The ability to answer causal questions is important for any system that requires robust scene understanding. In this demonstration, we develop a prototype system that leverages our causal link prediction framework, CausalLP. CausalLP framework uses a visual causal knowledge graph and associated knowledge graph embedding for two visual causal question and answering tasks- (i) causal explanation and (ii) causal prediction. In the live demonstration sessions, the participants will be invited to test the efficiency and effectiveness of the system for visual causal question and answering.

Keywords

Visual causal knowledge graph, causal explanation, causal prediction, causal link prediction

1. Introduction

Answering questions about scenes often requires knowledge of the causal relations between events. As an example, consider a scene in which a yellow ball collides with a blue cylinder, as depicted in Figure 1. Several questions may be asked about this collision event, including:

- **Question**: What is the cause of the collision? **Answer**: The red cube collides with the yellow ball.
- **Question**: What is the effect of the collision? **Answer**: The blue cylinder moves.

The first question type is referred to as a causal explanation; i.e. what is the cause of an event. The second question type is referred to as a causal prediction; i.e. what is the effect of an event. The ability to answer these types of causal questions is important for any system that requires robust scene understanding. In this demo, we will show how these types of questions are answered with the Causal Link Prediction (CausalLP) framework [1]. The information about objects and events occurring in the scene are represented in a knowledge graph (KG) along with the their associated causal relation. The link prediction techniques are used to infer new causal relations between events. These newly inferred causal links serve as answers to the explanation and prediction questions.

Posters, Demos, and Industry Tracks at ISWC 2024, November 13–15, 2024, Baltimore, USA

*You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

*Corresponding author.
ujaimini@email.sc.edu (U. Jaimini); cory.henson@us.bosch.com (C. Henson); amit@sc.edu (A. Sheth)
0000-0002-1168-0684 (U. Jaimini); 0000-0003-3875-3705 (C. Henson); 0000-0002-0021-5293 (A. Sheth)

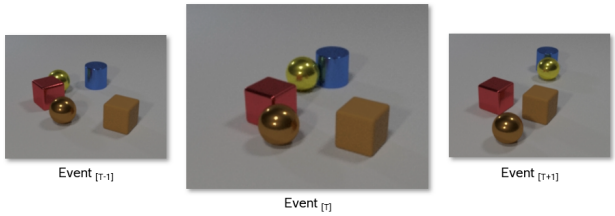

Figure 1: The center frame shows a video scene of a collision event, Event[T], between the yellow ball and blue cylinder. To the left, Event[T-1], shows a prior collision event between the red cube and yellow ball that caused Event[T]. To the right, Event[T+1], shows a subsequent event of the blue cylinder moving that is caused by Event[T].

The recent work in event level visual causal questions and answering focuses on the task of causal reasoning by discovering visual-linguistic causal patterns, temporal causal structures, and object-level causal relationship between object and language semantics [2, 3, 4]. To the best of our knowledge, the proposed CausalLP framework is the first attempt towards incorporating weights between the events (i.e. weighted causal relations) with the knowledge graph embedding (KGE) for visual causal question and answering.

2. Demonstration

The demonstration[1] of CausalLP focuses on showcasing key functionalities along with the benefits of using KG link prediction for the visual causal question and answering task [1]. This approach is applied to the CLEVRER [5] and CLEVRER-Humans [6], visual causal reasoning benchmark datasets to answer questions about video scenes with objects moving and interacting in a simulated environment. These datasets contains over 1000 simulated video scenes, annotated with information about the events, the participating objects, the causal relations between events, and the weights for each relation (i.e. weighted causal relation). The CLEVRER-Humans dataset provides information about the causal relations between events in the form of a Causal Event Graph (CEG). A CEG is constructed for each video through human annotators working with Mechanical Turk. For more information about the CLEVRER-Humans dataset, see [6].

Figure 2 shows an example with the interactive Python interface, where CausalLP is able to answer causal explanation and causal prediction questions about an event in the video scene. As shown in Figure 2 (A), the user can choose a target video in order to ask causal explanation and causal prediction question. Figure 2 (B) lists the events that occur in the video. Figure 2 (C) shows how a user can ask an explanation question about an event and display the result, such as *What is the cause of the yellow ball hits the light blue cylinder.* The event is caused by a *comeFrom* event. Figure 2 (D) shows how a user can ask a prediction question for an event and display the result, such as *What is the effect of the gray ball enter from the left?.* This event causes a *Hit* event in subsequent frames.

To perform the question and answering task with CausalLP, two models were trained for the explanation and prediction questions. The training and testing data were selected by splitting

[1]https://drive.google.com/file/d/1P3D3HIppZFsabsknLVq-4GwqLUciCcWQ/view?usp=sharing

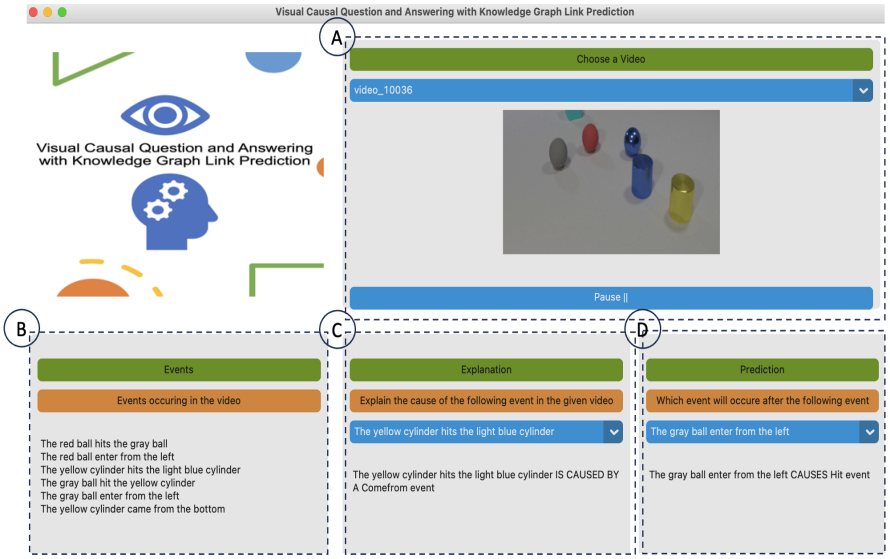

Figure 2: Visual Causal Question and Answering demonstration system. In the window (A) the user can choose a target video for causal explanation and causal prediction question. (B) lists the events that occur in the video. (C), and (D) are causal explanation and causal prediction questions and answering windows respectively.

the causal relations for each video scene based on their temporal positioning [1]. For the explanation model, the first few events in each scene are removed from the training data and only used for testing. For the prediction model, on the other hand, the final few events in each scene are removed from the training data and used for testing. With this setup, the initial events in each scene serve as answers to explanation questions while the final events serve as answers to prediction questions. Evaluation results of the CausalLP approach with the CLEVRER and CLEVRER-Humans datasets, as used in this demonstration, are promising. Using DistMult alone to train the KGE, i.e. without weights, results in an MRR score of 0.37. On the other hand, using DistMult together with FocusE, i.e. with weights, results in an MRR score of 0.56. On an average across all the models (i.e., TransE, DistMult, HolE, ComplEx), integrating weights (i.e. weighted causal relations) leads to a +75% MRR score improvement. Additionally, adding knowledge about the types of events and participating objects improves MRR score by +31%.

3. Conclusion and future work

In this paper, we present the CausalLP framework and demonstrate its use for a visual question and answering task. Specifically, causal explanation and prediction questions are answered based on video scenes from the CLEVRER and CLEVRER-Humans benchmark datasets. The proposed framework can be used for problems which involve cause and effect associations such as root cause analysis at time of system failure, cause and effect of a collision understanding in the autonomous driving systems, and trajectory prediction of a vehicle after a collision. In the future, we aim to extend the CausalLP for answering counterfactual "What if" questions.

Acknowledgments

This work is supported in part by NSF grants #2133842, "EAGER: Advancing Neuro-symbolic AI with Deep Knowledge Infused Learning", and #2119654, "RII Track 2 FEC: Enabling Factory to Factory (F2F) Networking for Future Manufacturing". Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] U. Jaimini, C. Henson, A. P. Sheth, Causallp: Learning causal relations with weighted knowledge graph link prediction, arXiv preprint arXiv:2405.02327 (2024).

[2] J. Xiao, X. Shang, A. Yao, T.-S. Chua, Next-qa: Next phase of question-answering to explaining temporal actions, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 9777–9786.

[3] C. Zang, H. Wang, M. Pei, W. Liang, Discovering the real association: Multimodal causal reasoning in video question answering, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 19027–19036.

[4] Y. Liu, G. Li, L. Lin, Cross-modal causal relational reasoning for event-level visual question answering, IEEE Transactions on Pattern Analysis and Machine Intelligence (2023).

[5] K. Yi, C. Gan, Y. Li, P. Kohli, J. Wu, A. Torralba, J. B. Tenenbaum, Clevrer: Collision events for video representation and reasoning, in: International Conference on Learning Representations, 2019.

[6] J. Mao, X. Yang, X. Zhang, N. Goodman, J. Wu, Clevrer-humans: Describing physical and causal events the human way, Advances in Neural Information Processing Systems 35 (2022) 7755–7768.