A Case-Based Reasoning Approach to Knowledge-Based Event Prediction

Sola S. Shirai[1,2,*], Debarun Bhattacharjya[2] and Oktie Hassanzadeh[2]

[1]Rensselaer Polytechnic Institute, Troy, NY, USA
[2]IBM Research, Yorktown Heights, NY, USA

Abstract
Forecasting the potential effects of significant events is an important task that can help inform proactive decision making. Using interpretable and explainable models for this challenging task could be greatly beneficial for developing trust in the predictions made by the system. In this demo, we present a prototype event prediction system using the EvCBR model, which uses a case-based reasoning approach to make predictions about cause-effect event pairs in a Knowledge Graph (KG). By performing reasoning over past cases of causal event relations present in the KG, EvCBR naturally provides explainability to its prediction results while also enabling easy integration of existing open-source KGs. Importantly, EvCBR is designed to perform inductive link prediction on unseen events and requires no training, allowing it to overcome the common limitations of other KG completion models. We demonstrate how EvCBR can be used together with a portion of the Wikidata KG to query for relevant entities, make predictions about ongoing events, and visualize past event cases and reasoning paths through the KG that were used to make predictions for the ongoing events. Our code is available at https://doi.org/10.5281/zenodo.8341874.

Keywords
Event Prediction, Case-Based Reasoning, Knowledge Graphs

1. Introduction

Being able to predict the effects of ongoing events is an important endeavor, enabling key decision-makers to gain greater insights about potential future outcomes and thereby proactively take measures to avail of opportunities or mitigate damage. As a simple example, one may predict that an *earthquake* event may cause a *tsunami* event, which subsequently may be used to issue evacuation warnings to local residents. As more variables become involved with the events, it could also become more difficult to produce effective predictions – for example, whether an *earthquake* leads to a *tsunami* will depend on the location where the event actually took place. Making such predictions about the consequences of events is of interest to various domains and use cases, such as disaster relief planning and financial analysis.

While making predictions about events with less straightforward causes and consequences, it is important for a model to be able to explain why it made certain predictions. This is valuable

ISWC'23: International Semantic Web Conference, November 06–10, 2023, Athens, Greece
[*]Part of this work was done while working at IBM Research.
shiras2@rpi.edu (S. S. Shirai); debarunb@us.ibm.com (D. Bhattacharjya); hassanzadeh@us.ibm.com (O. Hassanzadeh)
0000-0001-6913-3598 (S. S. Shirai); 0000-0002-9125-1336 (D. Bhattacharjya); 0000-0001-5307-9857 (O. Hassanzadeh)

CEUR Workshop Proceedings (CEUR-WS.org)

to convince users why a prediction may be reasonable, as well as to understand situations under which the predictions made by the model should *not* be taken at face value. A knowledge-driven approach to event prediction could therefore be beneficial, since methods leveraging background knowledge often tend to be more accurate but also more interpretable than many black-box models. While the task of event prediction has a rich history across many domains [1], work exploring the use of Knowledge Graphs (KGs) has been limited.

To tackle this challenge, we utilize a case-based reasoning model, EvCBR [2], to frame event prediction as a link-prediction task in a KG. EvCBR uses a KG containing examples of past cause-effect event pairs to make predictions about the effect of a new event. The rich semantic information captured within KGs allows us to leverage unambiguous and interoperable representations of events, rather than relying solely on textual sources or time-series data. By drawing upon cases of past events to make predictions, EvCBR also supports producing more interpretable results with interesting connections to the past. Furthermore, EvCBR requires no training, making it well suited to KGs which are constantly undergoing changes and additions such as Wikidata [3].

Compared to past work, EvCBR is not limited to only making predictions about entities seen during training, as is the case for many embedding-based prediction methods over KGs [4, 5]. Compared to similar case-based reasoning methods introduced recently [6, 7], EvCBR introduces improved methods to retrieve relevant cases as well as a novel framing of the link-prediction problem to make our methods more suitable for the specific task of event prediction.

2. 2-Hop Link Prediction for Event Prediction

The core idea in our work is to frame event prediction using a KG as a 2-hop link prediction task, where we aim to predict properties about some unseen *effect* event that is caused by a *cause* event. To provide greater explainability and to support predictions for new cause events, we develop a case-based reasoning approach to find similar past events and retrieve relevant patterns to make predictions.

Intuitively, case-based reasoning techniques aim to solve new problems using past cases of similar problems and their solutions. Extending this idea to event prediction and data present in KGs, we treat events in the KG connected by causal relations [8] as cases, and their "solutions" consist of paths through the KG that can be used to connect the cause event to its effect's properties. EvCBR follows a similar set of steps, first retrieving cases of cause-effect event pairs in the KG that are similar to a new prediction query, identifying reasoning paths that are present in those cases, then applying them to make predictions about the effects of the query event.

A high-level overview of EvCBR's methodology is shown in Figure 1. Starting from a new event, such as a Protest in Iran, we proceed with making predictions about a new effect event that may be caused by it. This example is an illustration of our attempt to predict consequences of the protests in Iran [9] on a KG curated before the protests started, therefore the KG does not contain this cause event. We will use similar examples in our demonstration.

Retrieving Cause-Effect Event Cases: For a given input event for which we wish to make predictions, we first identify cases of cause-effect event pairs in the KG that are similar to our new input event. We select cases based on the similarity of the case's cause to our new query

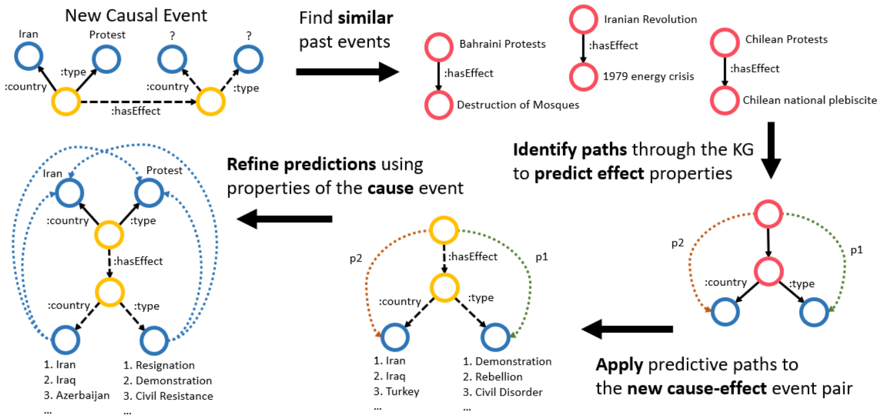

Figure 1: A high-level overview of EvCBR's methodology, using an example about a new Protest event.

event as well as the similarity of the case's effect's properties to the properties about our new, unseen effect event for which we wish to make predictions.

The similarity of individual entities in the KG are computed based on outgoing edges as well as the similarity of their subclass hierarchy. For case similarity, we compute a measure of importance for each relation based on how common the relation is, allowing us to weight rarer relations or entities as being more important when selecting which cases to retrieve.

Identifying and Applying Prediction Paths: After retrieving cases of past events in the KG, EvCBR next begins to identify paths through the KG which can be used to make predictions. Given that the retrieved cases have cause and effect event pairs, we aim to discover prediction paths of relations in the KG that can be followed from the cause event entity to the effect's properties for each case. The confidence of each of these prediction paths is then scored, with higher confidence being assigned to paths that lead to the correct entity more often across all the retrieved cases.

We next apply these learned prediction paths to make predictions about our new query event. From the new input event, we follow paths to make predictions about the unseen effect event's properties. The confidence of each path, calculated based on the precision of following the paths in past cause-effect event pairs, is used to compute an aggregated score of predictions for each property of the effect event.

Using this path-based prediction method allows EvCBR to inherently provide more explainability than many embedding-based models. Additionally, our method of identifying and scoring prediction paths is performed on-the-fly rather than requiring any pretraining process.

Refining Predictions: Lastly, we introduce an additional step to refine our prediction results. The intuition of our refinement method is that if EvCBR's predictions about the effect event are correct, we should be able to follow the procedure in the reverse direction to accurately "predict" the input cause event's properties starting from the effect event. Performing this process in reverse gives us a way to ground the confidence of our predictions in terms of the "ground-truth" properties of the input cause event.

Evaluation: We evaluated the performance of EvCBR against baseline KG completion methods, encompassing a number of notable models using methods such as distance based embeddings, graph neural networks, rule-learning, and case-based reasoning. We find EvCBR to have superior performance for event prediction as well as competitive performance for more general 2-hop link prediction. In addition to showing strong performance in our evaluations, the fact that EvCBR requires no training makes it well suited to our event prediction use case, and more generally when applied to a KG like Wikidata which is frequently updated.

3. Demonstration Plan

We plan to demonstrate our event prediction system using a Jupyter Notebook environment along with visualizations of the graph and predictions. Our target audiences for this demonstration are researchers and developers of decision support solutions, e.g., in finance [10] or enterprise risk management [11]. The functions in our approach could be used in such solutions through API calls and shown to end users on custom interfaces designed for analysts and non-expert users. During the demonstration, we first discuss the details of each of the steps in EvCBR using a fictional Megathrust Earthquake event in Japan. We then showcase the impressive predictions around the recent protests in Iran. We highlight the differences in our predictions with those returned by the highly-popular ChatGPT. In our examples, ChatGPT returns either very accurate but not very interesting results, or results that are partially incorrect as a result of the so-called "hallucination" of such large language models. This further supports the benefit of an explainable knowledge-based solution for the challenging task of event forecasting.

References

[1] L. Zhao, Event prediction in the big data era: A systematic survey, ACM Computing Surveys 54 (2022).

[2] S. Shirai, D. Bhattacharjya, O. Hassanzadeh, Event prediction using case-based reasoning over knowledge graphs, in: WWW, 2023.

[3] D. Vrandečić, M. Krötzsch, Wikidata: A free collaborative knowledgebase, Communications of the ACM 57 (2014).

[4] Q. Wang, Z. Mao, B. Wang, L. Guo, Knowledge graph embedding: A survey of approaches and applications, IEEE Transactions on Knowledge and Data Engineering 29 (2017).

[5] M. Wang, L. Qiu, X. Wang, A survey on knowledge graph embeddings for link prediction, Symmetry 13 (2021).

[6] R. Das, A. Godbole, N. Monath, M. Zaheer, A. McCallum, Probabilistic case-based reasoning for open-world knowledge graph completion, in: Findings of ACL: EMNLP, 2020.

[7] R. Das, A. Godbole, S. Dhuliawala, M. Zaheer, A. McCallum, A simple approach to case-based reasoning in knowledge bases, in: AKBC, 2020.

[8] S. S. Shirai, A. Khatiwada, O. Hassanzadeh, D. Bhattacharjya, Rule-based link prediction over event-related causal knowledge in wikidata, in: Wikidata@ISWC'22, 2022.

[9] Wikipedia contributors, Mahsa Amini protests, 2023. URL: https://en.wikipedia.org/wiki/Mahsa_Amini_protests, [Online; accessed 8-September-2023].

[10] P. Yoo, M. Kim, T. Jan, Machine learning techniques and use of event information for stock market prediction: A survey and evaluation, in: CIMCA-IAWTIC'06, 2005.

[11] S. Sohrabi, A. V. Riabov, M. Katz, O. Udrea, An AI planning solution to scenario generation for enterprise risk management, in: AAAI, 2018.