Present Causal Relationship Retrieval for Historical Analogy

Kenta Matsumaru[1], Ryohei Ikejiri[2], and Yasunobu Sumikawa[1]

[1] Depat. of Computer Science, Takushoku University, 815-1 Tatemachi, Hachioji-shi, Tokyo, Japan
198466@st.takushoku-u.ac.jp, ysumikaw@cs.takushoku-u.ac.jp
[2] Interfaculty Initiative in Information Studies, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
ikejiri@iii.u-tokyo.ac.jp

1 Extended Abstract

The importance of studying history and applying that knowledge in the modern world is widely recognized. The widely recognized benefits of the activity is that it enables us to understand the formation process of modern societyand to analogously use that knowledge to present-day events. Indeed, research is being conducted on how to effectively develop the ability to use this knowledge to think of ways to solve problems that arise in modern society. This ability is called historical analogy.

Contributions. In this study, we propose a search algorithm to promote historical analogy by immediately exploiting knowledge of history just learned. When texts expressing the causal relationship of events that occurred in the past are input, our algorithm outputs ranked present causal relationships according to their similarity to the input. Previous studies investigated that a definition of the conditions necessary to facilitate historical analogies [4] and proposed an algorithm to search for past causal relationships that are similar to present causal relationships [6]. In particular, the algorithm for searching past causal relationships that is the closest past study to this study provided an opportunity to make practical use of learned history by making the input present causal relationships. However, the past causal relationships in the output may include those that have not yet been learned by the users. This study provides an immediate connection between the past and the present for many learners, regardless of age or country.

Approach. We represent causal relationships by combining related events. As the purpose of this study is to facilitate historical analogies, we arrange the events in chronological order and combine them to make the causal relationships. We then calculate the rank of each causal relationship by Event Causality relationship similarity Measurement (ECM) [5].

In this study, it is assumed that the input is past causal relationship. In texts of past events, named of people, organizations, or events are often used. On the other hand, these entities are not always used in present causal descriptions.

Therefore, our algorithm defines *temporal entity* as an entity that is used only during a specific period of time to replace them with its type.

Algorithm. We now describe entity extraction, temporal entity replacement, ECM application, and ranking of the results. We first extract entities from the input text using TagMe [1]. As TagMe extracts named entities by assigning a link to a Wikipedia article to a word in the text, we collect these words and the Wikipedia articles. We then analyze the years of existence of the entities by checking the collected Wikipedia articles in order to detect temporal entity. For the existence years analysis, we used the birth/establishment and death/destruction years specified in the Wikipedia categories to determine whether the distribution of the years of difference between them was multimodal or not. The results of this analysis revealed that country, city, organization, person, and event are the five entity types we can use as temporal entities. In particular, we found that we should set 120 years as a valid criterion for countries and cities. All entities of the other three types are considered as temporal entities. After converting the temporal entities to their type names, we create feature vectors using the topic distribution obtained by applying latent semantic analysis to the resulting sequence of tokens. After creating a feature vector for each event, we create a bipartite graph for the two causal relationships. We then apply ECM on the graph to determine the similarity between causal relationships. This is done for all combinations of the input past causal relationships and present ones stored in the database. Finally, we output the top k with the highest score.

Evaluations. In order to evaluate the effectiveness of our algorithm, we used the past causality data [3], which includes past causal relationships, and W2E [2], which includes present causal relationships, as the datasets for evaluation. These datasets allocate categories for all causal relationships.

For these datasets, we use three methods as baselines: a method for finding the similarity of a set of sentences using the Jaccard coefficient, a method for sentence analysis using cosine similarity, and a method using ECM without extraction of temporal entity. We checked whether the input and output categories of results of the 3 baselines and our algorithm matched to calculate F1 scores. We confirmed that the Jaccard coefficient was 37.7%, the cosine similarity was 40.7%, and the ECM was 46.2%. The proposed method, on the other hand, was 53.1%. We can say that our algorithm is the best.

Future Work. Future work will include *proposing an effective learning curriculum for using this learning environment into history classes.* As this study evaluated the accuracy of the search algorithm, it is important to evaluate whether it actually facilitates historical analogies.

References

1. Ferragina, P., Scaiella, U.: Tagme: On-the-fly annotation of short text fragments (by wikipedia entities). In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. pp. 1625–1628. CIKM '10, Association for Computing Machinery, New York, NY, USA (2010)

2. Hoang, T.A., Vo, K.D., Nejdl, W.: W2e: A worldwide-event benchmark dataset for topic detection and tracking. In: CIKM '18. pp. 1847–1850. Association for Computing Machinery (2018)
3. Ikejiri, R., Sumikawa, Y.: Past causalities and event categories for connecting similar past and present causalities. https://doi.org/10.5281/zenodo.3601707 (jun 2019). https://doi.org/10.5281/zenodo.3601707
4. Lee, P.: Historical literacy: Theory and research. International Journal of Historical Learning, Teaching and Research **5**(1), 25–40 (2005)
5. Sumikawa, Y.: Event causal relationship retrieval. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. pp. 318–325. WI-IAT '21, Association for Computing Machinery, New York, NY, USA (2021)
6. Sumikawa, Y., Ikejiri, R.: Mining historical social issues. In: IDT. pp. 587–597 (2015)