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

Abstract. The importance of studying history and applying this knowledge is widely recognized. In this study, we propose a search algorithm to promote historical analogy by immediately exploiting historical knowledge. Our algorithm outputs present causal relationships similar to the input past causal relationship. It first replaces temporal entities that are limited to certain periods with their entity types to enhance the connection between the causal relationships of the past and present. Subsequently, it measures causal similarities on a bipartite graph of the past and present causal relationships. Finally, the accuracy of the algorithm is evaluated on two datasets comprising present and past causal relationships, respectively. We found that the temporal entity replacement is effective in searching for causal relationships.

Keywords: Causal relationship retrieval · Historical analogy · Temporal entity

1 Introduction

Historical knowledge can be beneficially applied to present scenarios. For instance, it helps us understand the formation processes of modern society and gain a deeper understanding of the identity of different countries or regions. Indeed, many countries have history classes from elementary school onward.

Previous studies have defined the conditions necessary to facilitate historical analogies [16], proposed an algorithm to search for past causal relationships that are similar to the present causal relationships [22], and designed history classes using a search algorithm [13]. In one study on designing classes, students were motivated to connect acquired historical knowledge with the present and use to develop solutions to contemporary problems, called historical analogy, rather than merely memorizing past events. When such classes are regularly incorporated into the curriculum, learners gain a deeper understanding of history and develop ways to apply it. However, it requires the historical DB to be prepared in advance. Thus, the teacher must pre-select past events that can facilitate analogies and describe causal relationships in appropriate sentences. Furthermore, utilizing freshly acquired knowledge is difficult for learners because they lack a comprehensive understanding of the past causal relationships output in the search results.

Contributions. In this study, we propose a search algorithm to promote historical analogy using freshly acquired historical knowledge. When texts expressing the causal relationship of past events are input, our algorithm outputs present causal relationships that are ranked according to their similarity to the input. Compared with past studies that promote historical analogies [22, 13], the novelty and main contribution of this study is that it develops an environment conducive to *immediately* facilitates historical analogies. This novelty provides two contributions compared to this previous study. The first one is that the past causal relationship is usable for the historical analogy. The second one is to eliminate need to prepare a historical database in advance. These contributions provide an immediate connection between the past and the present for many learners, regardless of their age or country.

We represent causal relationships by combining related events. As the purpose of this study is to facilitate historical analogies, we arranged the events in chronological order and combined them to create causal relationship. This definition of causal relationship is typically employed in research on topic detection and tracking (TDT) [19]. To calculate the rank of each causal relationship, we assume that events are in similar chronological order are more similar than others. The calculation is performed using the event causality relationship similarity measurement (ECM) [21].

In texts of past events, person or organizations are named as perpetrators of events in a certain time period. However, these entities are not always used in present causal relationship descriptions. Moreover, revealing the type of entity, which is a higher-order expression, is more effective than revealing their names in promoting analogy [10]. Therefore, our algorithm defines a *temporal entity* as an entity that is used only during a specific period and replaces it with its type.

The remainder of this paper is organized as follows. Section 2 provides summaries of several related works. The algorithm proposed in this study is described in Section 3. The results of the experimental evaluation of the proposed algorithm are discussed in Section 4, and the conclusions are presented in Section 5.

2 Related Works

2.1 History Learning Support

In the research area of history study and in the curriculum guidelines in Japanese high schools, developing the ability to apply acquired knowledge has become one of the most important goals. Reviewing the results of research on history study, the following achievements have been identified: the discovery of the basic elements for facilitating historical reasoning [7], in depth analysis of successful facilitation of historical analogies [4], and the definition of a framework of events that can be connected in the past and present [16].

Learning environments have also been proposed based on the above research findings. Ikejiri *et al.* proposed a competitive card game, in which players must construct causal relationships for present social problems that are structurally identical to certain past causal relationships. The game helps students in identifying and comparing two causal relationships and stimulating historical analogy. Ikejiri *et al.* also proposed another competitive card game that requires players to create new policies that would revitalize Japan's economy by learning about past economic policies [12]. A search engine has also been proposed to facilitate historical analogies from information in news articles reporting present events [22]. Researchers also proposed using this search engine to design an actual history class [13].

Previous studies have identified the mechanisms that facilitate historical analogy and proposed learning environments that use these mechanisms. However, they have failed to propose methods by which students can immediately utilize freshly acquired historical knowledge. This study is orthogonal to previous studies; while the search engine realized in this study can be used to promote historical analogies in real-time, the learning system proposed in the previous study can be used to promote historical analogies over multiple reviews.

2.2 Temporal Information Retrieval

The proposed algorithm measures the similarity of causal relationships occurring at different points on a time axis. Therefore, we survey previous studies on temporal information retrieval (T-IR). Previous studies on T-IR include the detection of temporal representations [11], timeline generation [2], historical image retrieval [6], and future forecasting [15, 18]. This study focuses on historical information retrieval and analogy retrieval. See survey paper [5] for more information on other previous studies.

Most algorithms search for information that users are seeking [3, 17, 20]. For example, a search algorithm specific to the literature review methodology of historians has been proposed [20]. According to this study, a historian first finds an overview of the subject and then searches for a detailed analysis. The algorithm proposed in that study not only outputs a variety of content related to a single query but also emphasizes the diversity of aspects in the output. A method has also been proposed to output past as well as possible future events [1].

Our algorithm does not output the past events that are directly related to the causal relationship provided as the input; however, it outputs similar events in the form of a causal structure in which similar events occur in the same order.

Several analogous search algorithms have been proposed to identify structural similarities [23, 24]. These algorithms are designed to detect entities that are past counterparts of present entity. For this purpose, they calculate the similarity of the entities using words used to describe each entity. Therefore, the target of the search is different from that of this study. The targets of these previous studies are entity whereas this study focuses on causal relationships that brings together multiple events.

3 Algorithm

Our algorithm outputs ranked present causal relationships by performing temporal entity replacement, measuring the similarity between past and present causal relationships, and ranking the results. We assume that the present causal relationships are stored in a database. As the purpose of this study is to output ranked present causal relationships, we do not place any restrictions on how the present causal relationships are constructed. In other words, it can be generated automatically using algorithms designed for TDT [19]. The appropriate relationships can also be manually selected and stored in the database in advance.

3.1 Temporal Entity Replacement

This study defines a *temporal entity* as an entity that occurs in a certain period (e.g., persons or organizations). In contrast, non-temporal entities are entities that occur over long periods (e.g., France). The purpose of this temporal entity replacement is to enhance the similarity between causal relationships of the past and present by increasing the number of structural matches. For instance, one type of causal structure that pertains to both the past and present is: this event is a result of the actions of certain persons.

To perform temporal entity replacement, we first extract entities from the input text using TagMe [8]. To distinguish between temporal and non-temporal entities, we set a threshold for the number of years for which they have existence. We obtained the Wikipedia articles on all the years from 1 to 2020, which recorded important events from those years and applied TagMe to them. Because TagMe extracts named entities by assigning a link to a Wikipedia article to a word in the text, we collected these words and the corresponding Wikipedia articles. For the analysis of the number of years of existence, we used the year of birth, establishment, and death or destruction specified in Wikipedia categories to obtain a distribution for each entity type. If the shape of the distribution for an entity type is unimodal and its mode is not over a few hundred years, then all entities of that type are immediately considered to be temporal. As expected, all entities of organizations, persons, and events fell into this category. In addition to these 3 types, countries are revealed as an entity type that can be used as a temporal entity.

We manually checked the countries collected using the above methods and found that some of them, such as Ottoman Empire, no longer exist. To ascertain the number of years of existence of a country, we collected the year of establishment and disestablishment of each country from the Wikipedia categories and infoboxes. Note that since the year of disestablishments is not listed for existing countries such as Japan and the U.S., we consider 2022 as the year of disestablishment for those countries. Fig. 1 shows the distribution of years of existence.

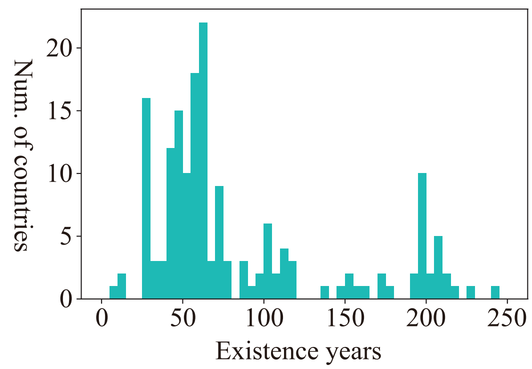

Fig. 1. Existence year distribution for countries

The vertical axis represents the number of countries, whereas the horizontal axis represents the number of years a country has existed. This result indicates that 120 years should be set as a criterion for a country to be considered as a temporal entity.

After identifying the temporal entities in the text of the event, our algorithm replaces them with their types. For example, if the temporal entity is a country, we replace the name of the country with "country."

3.2 Past and Present Causal Relationship Similarity Measurement

After converting the temporal entities to their type names, we perform tokenizing, lemmatizing, normalizing, and removing stopwords. Subsequently, we create feature vectors such as using the topic distribution obtained by applying latent semantic analysis (LSA) to the resulting sequence of tokens. After creating a feature vector for each event, we create a bipartite graph of the two causal relationships. In the graph, we use events as nodes and the similarities between the events as edges. Event similarity is calculated using the cosine similarity for the two feature vectors. We then apply ECM on the graph to determine the similarity. These procedures are performed for all combinations of the input past causal relationships and the present ones stored in the database. Finally, we output the top k with the highest scores.

3.3 Algorithm Overview

Algorithm 1 is an overview of the proposed algorithm. First, in lines 2∼4, the temporal entity replacement described in Sec.3.1 is performed on the input past causal relationship texts. In lines 2∼3 TagMe is applied to the past texts to extract entities, and subsequently filter out temporal entities using the results of the existence year analysis. The temporal entities are replaced with their type names in line 4. In lines 5∼7, feature vectors are created. In line 6,

Algorithm 1 Algorithm overview

Input: A past causal relationship: *HCausalityRel*, A list of present causal relationships *Presents*

Output: A present causal relationship

```
1: Function HistoryToPresent(HCausalityRel, Presents)
2:   Entities ← getEntities(HCausalityRel) // Applying TagMe
3:   TempEntity ← getTempEntity(Entities) // Temporal entity detection
4:   Texts ← replaceTempEntity(HCausalityRel, TempEntity)
5:   // Normalizing, tokenizing, lemmatization, removing stop words from the texts of past causal relationship
6:   Tokens ← preprocess(Texts)
7:   PastFVecs ← makeFVec(Tokens) // Feature vector creation
8:   Scores ← []
9:   for PrstFVec ∈ Presents
10:    BGraph ← createBipartitleGraph(PastFVecs, PrstFVec) // Bipartite graph construction
11:    // Measuring causal relationship similarity
12:    Val ← ECM(BGraph)
13:    Scores.append(Val)
14:  // Return the highest sum of similarity between causal relationships.
15:  Top ← top(Scores, Presents)
16:  return Top
```

preprocessing, which includes tokenization, normalization, lemmatization, and the removal of stopwords, is performed. In line 7, we create the feature vector for the preprocessed text. In this evaluation, we use the topic vector of the LSA as the feature vector for each event. Subsequently, we measure the similarity of each combination of present and past causal relationships in lines 8∼13. The algorithm functions under the assumption that feature vectors for present causal relationships have already been created using the same procedure as the past causal relationship. The feature vectors for the present causal relationships are presented as arguments. In line 9, the present causal relationships are individually selected, and a bipartite graph is constructed from the feature vectors of the present and past causal relationships in line 10. Line 12 measures the similarity between the two causal relationships on the bipartite graph. Line 13 stores the result in a list *Scores*. Finally, after measuring the similarity of all the causal relationships, line 15 returns the results converted into a form that can actually be output.

4 Experimental Evaluation

4.1 Setting

Dataset. To evaluate the effectiveness of our algorithm, we used W2E [9], which comprises present causal relationships, and past causality data [14], which comprises past causal relationships. The W2E dataset is a large dataset comprising news articles published in more than 50 prominent mass media worldwide. This dataset includes lists of multiple events that were manually grouped together as topics. These events were recorded in the current events portal of the English Wikipedia. Each past causal relationship contains texts on cause, action, and effect. To create a bipartite graph, this study considered the 3 types of text as separate events.

These datasets allocated categories to all causal relationships. All data in the W2E dataset were also assigned to one of the following 9 categories: `Armed Conflicts` and `Attacks` (`AA`), `Arts` and `Culture` (`AC`), `Business` and `Economy` (`BE`), `Disasters` and `Accidents` (`DA`), `Health` and `Environment` (`HE`), `Law` and `Crime` (`LC`), `Politics` and `Elections` (`PE`), `Science` and `Technology` (`ST`) and `Sport` (`S`). In this study, we extracted 3,038 items from the dataset as output candidates. All past causal relationships were assigned one of the following 13 categories; `Reign`, `Diplomacy`, `Production`, `Commerce`, `Religion`, `Literature and Thought`, `Technology`, `Popular Movement`, `Community`, `Environment`, `Study`, `War`, and `Disparity`. In this study, 138 items from this dataset were used as inputs.

In this study, we evaluated the output results based on whether the category names matched as described below. However, the above two datasets were constructed separately and therefore had different categories. For the evaluation, we used only those categories that commonly existed in the two datasets. Specifically, we use the following 5 categories: `Diplomacy`, `Reign`, `Literature`, `Environment`, and `Technology`. As `HE` and `ST` in the W2E dataset correspond to `Environment` and `Technology`, respectively, they were directly replaced with those in the past causality database. The 3 categories of the W2E dataset, `BE`, `PE`, and `AC`, are named differently from the above 5 categories; however, the events in each category were identified and rewritten as follows. As the events in `BE`, especially those related to the economy, involved more than one country, they were designated as `Diplomacy`. The political events in `PE` were designated as `Reign` because they were related to national policies. Finally, numerous works of literature are described in the text categorized as `AC`. Thus, this category was designated as `Literature`.

Evaluation Criteria. We evaluated retrieval results by whether the input and output categories were consistent. The actual evaluation criteria used were macro-averaged precision (P), recall (R), and F1 value ($F1$) for the top case and mean average precision (MAP) for the top 10 cases.

Baselines. For these datasets, we used three methods as baselines: a method for finding the similarity of a set of words using the Jaccard coefficient (Jaccard), a method using LSA and cosine similarity (Cos), and a method using ECM without temporal entity replacement (ECM). We refer to our algorithm as temp+ECM in the evaluation. For Jaccard and Cos, we combined all event texts in a single past causal relationship or topic into a single document.

Table 1. Scores of macro-averaged precision (P), recall (R), F1 value ($F1$), and mean average precision (MAP).

	P	R	$F1$	MAP
Jaccard	39.1%	37.5%	37.7%	64.7%
Cos	41.7%	39.4%	40.7%	58.8%
ECM	48.9%	45.0%	46.2%	67.2%
temp+ECM	*54.4%*	*51.4%*	*53.1%*	*71.3%*

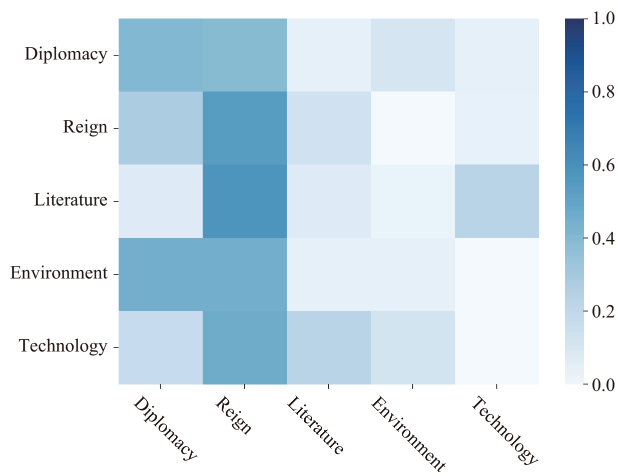

Fig. 2. Matched category ratios resulted by our algorithm.

4.2 Results

Table 1 shows all the scores for the 3 baselines and our algorithm. We confirmed that the $F1$ scores for Jaccard, Cos, ECM, and temp+ECM were 37.7%, 40.7%, 46.2%, and 53.1%, respectively, thereby indicating that our algorithm exhibits the best performance. In particular, it performed better than the 3 baselines in terms of P, R, and $F1$ scores by 15∼5.5% points. These results indicate that when searching for causal relationships, the similarity of events in chronological order and temporal entity replacement significantly impact the result.

To analyze the results of the proposed algorithm, we studied the percentage of category matches between the top causal relationship and the input. Fig. 2 shows this result. The vertical and horizontal axes represent the categories of past and present causal relationships, respectively. This result indicates that a high accuracy was obtained for the categories `Diplomacy` and `Reign`.

To analyze the reasons for the low scores in several categories, we measured the similarity and dependence between categories using the Jaccard coefficient

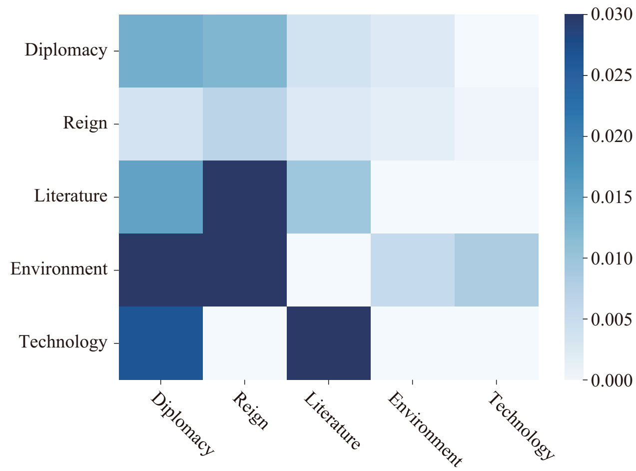

Fig. 3. Category similarities by Jaccard

and mutual information (MI) defined as follows:

$$Jaccard(A, B) = \frac{\mid T_A \cap T_B \mid}{\mid T_A \cup T_B \mid} \tag{1}$$

$$MI(A, B) = \sum_{a \in A} \sum_{b \in B} p(a, b) \log \left(\frac{p(a, b)}{p(a)p(b)} \right) \tag{2}$$

where $\mid \cdot \mid$ represents the size of the set "$\cdot$". T_A and T_B are tokens used in categories A and B, respectively. The higher the Jaccard coefficient score, the more similar the two arguments are. The higher the MI score, the more dependent the two arguments are.

Figs. 3 and 4 show these scores. For the 3 categories of `Literature`, `Technology`, and `Environment`, the scores between sentences in the same category were lower than those for `Diplomacy` and `Reign`. In addition, the MI scores of `Literature`, `Environment`, and `Technology` between `Diplomacy` and `Reign` tend to be relatively high. We manually checked the incorrectly retrieved results and found that the range of events covered by categories was not strictly consistent between the past and present datasets, even if the categories had the same name. For example, the `Diplomacy` in the present dataset included news related to international relations and economics; however, these news events also involved domestic governance and foreign invasions. Table 2 shows an example of category mismatch. This example shows that both past and present events are related to economics; however, they were assigned the categories of `Diplomacy` and `Reign`, respectively. Such discrepancies could be reduced if potential semantic matches between categories of different datasets could be detected and converted into multi-label datasets.

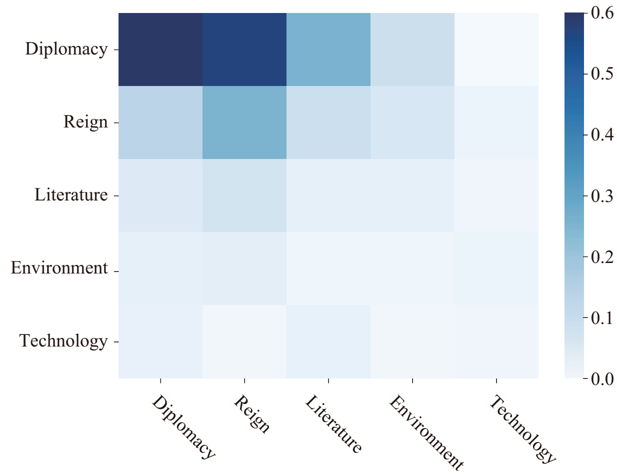

Fig. 4. Category dependency by MI

Finally, we checked the MAP scores for the 4 algorithms. The results are presented in Table 1. We can see that the MAP score of the proposed algorithm was also higher than those of the baselines. In particular, a comparison of the proposed algorithm and ECM shows that the accuracy is increased by approximately 4% points due to temporal entity replacement.

5 Conclusions & Future Work

In this study, we proposed an algorithm that retrieves present causal relationships that are similar to the past causal relationships that are input. This algorithm replaces the entities limited to specific periods with their types. Subsequently, it measures the similarity of causal relationships that include multiple events sorted in chronological order. To confirm the effectiveness of the proposed algorithm, we compared the $F1$ and MAP scores for our algorithm against the baselines and found that our algorithm obtained the best scores.

In future work, we hope to *propose a method to gauge the potential meanings of categories in different datasets,* as discussed in the evaluation. This approach could improve search accuracy in situations where different categories should be defined for different years; however, the search would be conducted using all the datasets to search over a broad range of time. Furthermore, we hope to *develop a learning environment that incorporates this search algorithm.* We also envision creating a UI that aids students in inputting sentences that consider causality. In this manner, our proposed algorithm can be deployed in actual classes. Finally, we hope to *propose an effective curriculum incorporating this learning environment into history classes.* Further research is also required to confirm whether

Table 2. Example of mis prediction.

	Past (`Diplomacy`)
1.	Sparta comprised Spartan citizens (a minority) and slaves (the majority).
2.	The former were always fearful of a revolt by the latter. Efforts were made to maintain unity among the citizens: the use of currency was banned so that wealth disparities would not arise among them; land was distributed equally to make all entirely equal; and the country was closed off to prevent outside influences.
3.	Thus, the unity of Spartan citizens was heightened, and Sparta became the strongest Greek army state.
	Present (`Reign`)
1.	The European Central Bank reinstates the waiver that enables Greek banks to use the country's bonds as collateral for regular loans, for the first time in more than a year.

the proposed search algorithm facilitates the formation of historical analogies in actual classrooms.

Acknowledgments. This work was supported in part by MEXT Grant-in-Aids (#22K02884).

References

1. Abujabal, A., Berberich, K.: Important events in the past, present, and future. pp. 1315–1320. WWW'15, ACM, New York, NY, USA (2015)
2. Althoff, T., Dong, X.L., Murphy, K., Alai, S., Dang, V., Zhang, W.: Timemachine: Timeline generation for knowledge-base entities. pp. 19–28. KDD'15, ACM, New York, NY, USA (2015)
3. Bogaard, T., Hollink, L., Wielemaker, J., Hardman, L., van Ossenbruggen, J.: Searching for old news: User interests and behavior within a national collection. pp. 113–121. CHIIR'19, ACM, New York, NY, USA (2019)
4. Boix-Mansilla, V.: Historical understanding: Beyond the past and into the present. Knowing, Teaching, and Learning History: National and International Perspectives pp. 390–418 (2000)
5. Campos, R., Dias, G., Jorge, A.M., Jatowt, A.: Survey of temporal information retrieval and related applications. ACM Computing Surveys (CSUR) **47**(2), 15 (2015)
6. Chew, M.M., Bhowmick, S.S., Jatowt, A.: Ranking without learning: Towards historical relevance-based ranking of social images. pp. 1133–1136. SIGIR'18, ACM, New York, NY, USA (2018)
7. van Drie, J., van Boxtel, C.: Historical reasoning: Towards a framework for analyzing students' reasoning about the past. Educational Psychology Review **20**(2), 87–110 (2008)
8. 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)

9. 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)
10. Holyoak, K.J., Thagard, P.: Mental Leaps: Analogy in Creative Thought. MIT Press (1980)
11. Holzmann, H., Risse, T.: Named entity evolution analysis on wikipedia. pp. 241–242. WebSci'14, ACM, New York, NY, USA (2014)
12. Ikejiri, R., Fujimoto, T., Tsubakimoto, M., Yamauchi, Y.: Designing and evaluating a card game to support high school students in applying their knowledge of world history to solve modern political issues. ICoME '12, Beijing Normal University (2012)
13. Ikejiri, R., Sumikawa, Y.: Developing world history lessons to foster authentic social participation by searching for historical causation in relation to current issues dominating the news. Journal of Educational Research on Social Studies **84**, 37–48 (2016), (in Japanese)
14. 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
15. Jatowt, A., Kanazawa, K., Oyama, S., Tanaka, K.: Supporting analysis of future-related information in news archives and the web. pp. 115–124. JCDL'09, ACM, New York, NY, USA (2009)
16. Lee, P.: Historical literacy: Theory and research. International Journal of Historical Learning, Teaching and Research **5**(1), 25–40 (2005)
17. Odijk, D., de Rooij, O., Peetz, M.H., Pieters, T., de Rijke, M., Snelders, S.: Semantic document selection. pp. 215–221. TPDL'12, Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
18. Radinsky, K., Horvitz, E.: Mining the web to predict future events. pp. 255–264. WSDM'13, ACM, New York, NY, USA (2013)
19. Radinsky, K., Davidovich, S.: Learning to predict from textual data. J. Artif. Int. Res. **45**(1), 641–684 (2012)
20. Singh, J., Nejdl, W., Anand, A.: History by diversity: Helping historians search news archives. pp. 183–192. CHIIR'16, ACM, New York, NY, USA (2016)
21. 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)
22. Sumikawa, Y., Ikejiri, R.: Mining historical social issues. In: IDT. pp. 587–597 (2015)
23. Zhang, Y., Jatowt, A., Bhowmick, S., Tanaka, K.: Omnia mutantur, nihil interit: Connecting past with present by finding corresponding terms across time. pp. 645–655. ACL/IJCNLP'15, ACL (2015)
24. Zhang, Y., Jatowt, A., Tanaka, K.: Temporal analog retrieval using transformation over dual hierarchical structures. pp. 717–726. CIKM'17, ACM, New York, NY, USA (2017)