2021/12/17 @ WI-IAT'21

Event Causal Relationship Retrieval

Takushoku Univ.
Yasunobu Sumikawa

Background | The importance of history

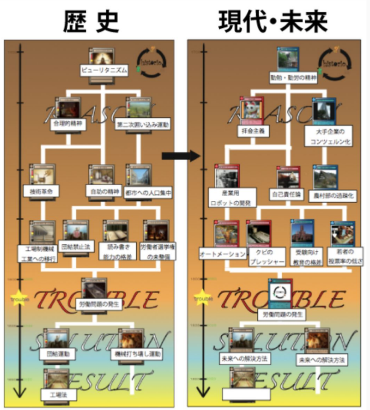
- *History does not repeat itself, but rhythms repeat themselves*
- Analyzing and knowing history is important
  - Understanding how the present shape
  - Using historical knowledge in the present
  - Many countries have classes for learning history from elementary school
- Practice and research in history learning supporting
  - Understanding history → Developing Thinking Skills → Utilization of the ability of using historical knowledge [1]
  - Research on learning environments that support the ability to use history in school education[2]

[1]: MEXT (2018): government course (curriculum) guidelines
[2]: P., Lee (2005) : Historical Literacy: Theory and Research, History Education Research Journal 5(1)

Related works | Focusing on Causal Relationships

- We can use the knowledge of the past for the present and future.

- Promoting historical analogies [3]
  - Card-game-based learning material for finding similar causal relationships between past and present
  - Learners manually constructs the causal relationship
  - Theme is fixed by the research



[3]: Ryohei Ikejiri. 2011. Designing and Evaluating the Card Game which Fosters the Ability to Apply the Historical Causal Relation to the Modern Problems. Japan Society for Educational Technology 34, 4 (april 2011), 375–386. (in Japanese).

Related works | Focusing on Causal Relationships

- We can use the knowledge of the past for the present and future.

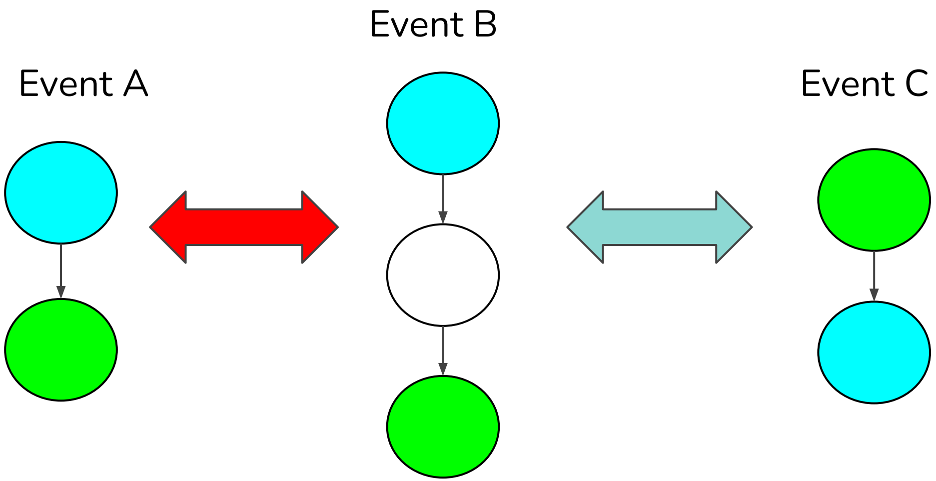
- Predicting future events [4~5]
  - Learning relationship A→B from texts
  - Predicting ○ of A' → ○ using the learned relationships

Earthquake → Tsunami

[4]: K., Radinsky, S., Davidovich, S., Markovitch (2012): Learning to Predict from Textual Data. J. Artif. Intell. Res. 45: 641-684.)
[5]: A., Jatowt, C.-m. A., Yeung (2011): Extracting collective expectations about the future from large text collections. CIKM: 1259-1265

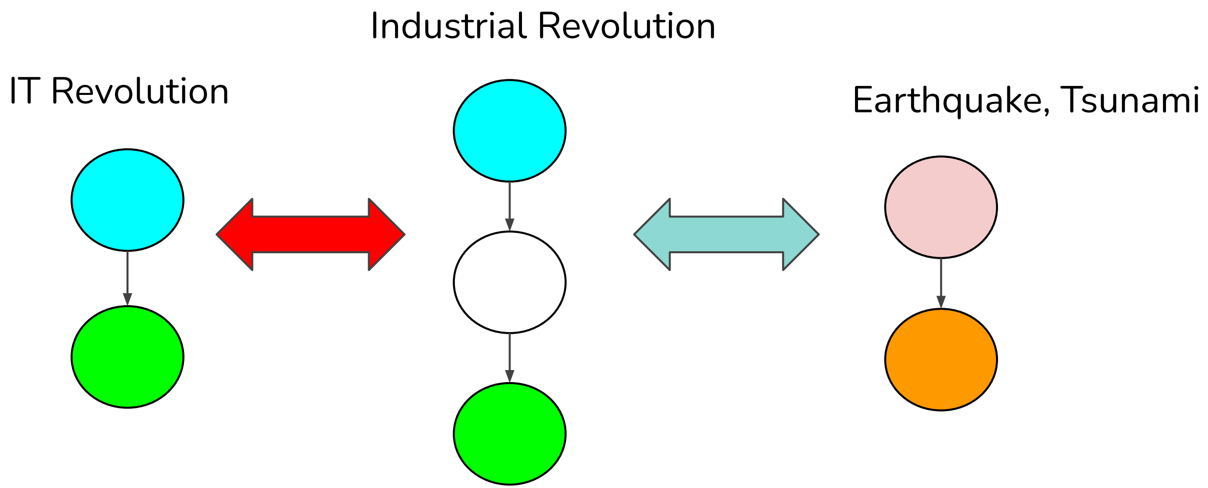
Objective of this study

- Measure similarity between causal relationships
    - Assumption : Event graph is given



Usage | Future work

- Searching past/present events with the same/similar cause
  - for history education

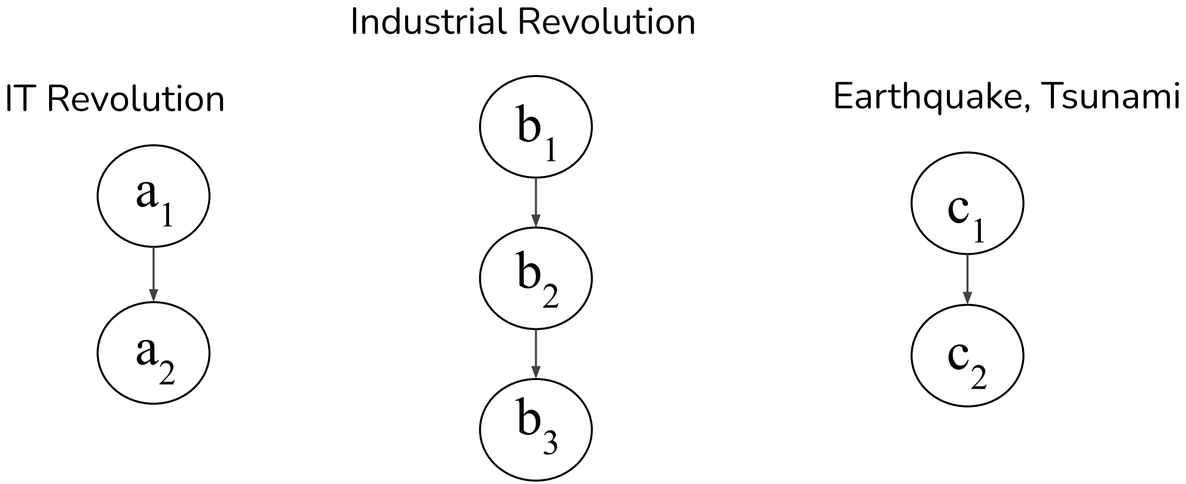


Outline

- Definition
- Proposed algorithm (ECM)
  - Event Causal relationship Measurement
- Experimental evaluation
- Conclution

Data representation

A causal relationship is a list of events

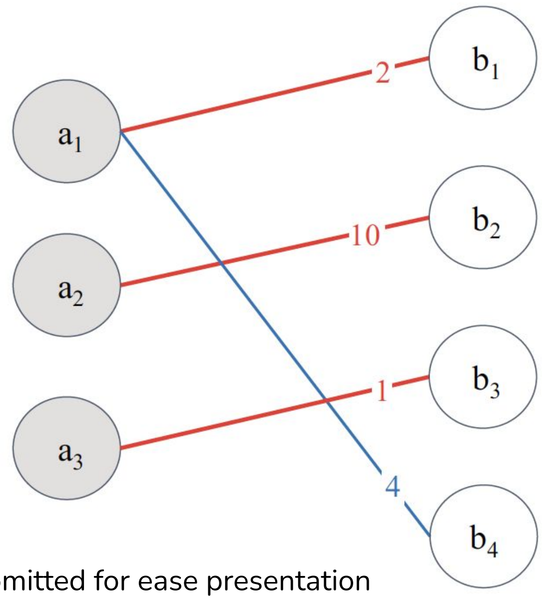


Bipartite graph construction

To compare two causal relationships A and B

ECM constructs a bipartite graph G = (A, B)

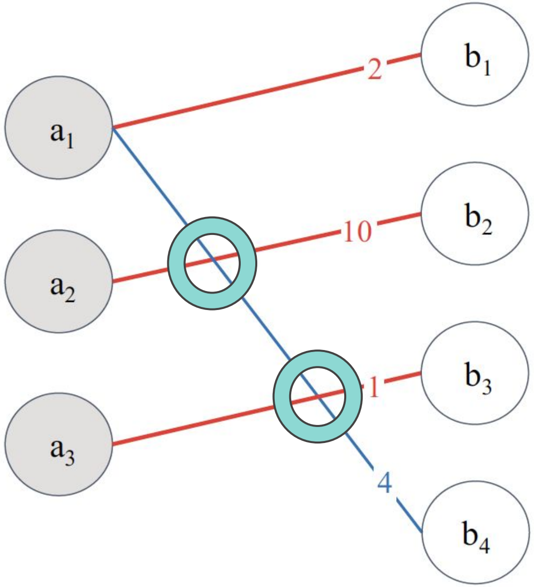
- A, B: lists of events
- A weight w of e = (a_i, b_j): similarity between two events a_i and b_j



※ some edges are omitted for ease presentation

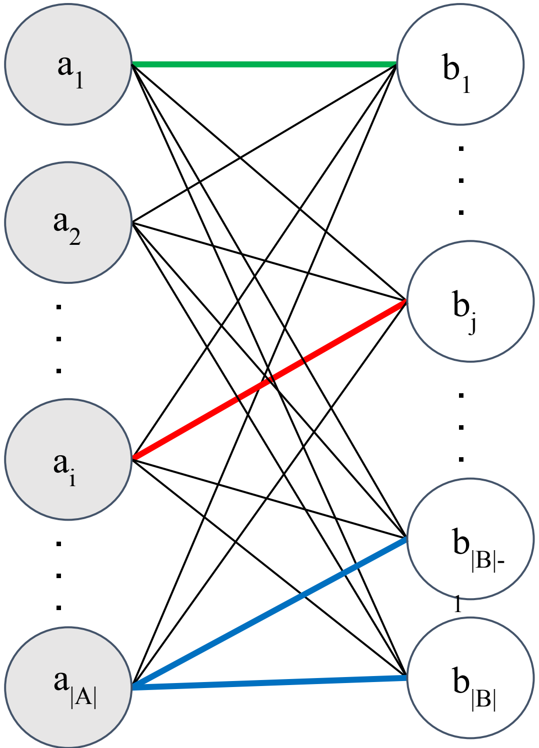
Algorithm | Theory

- Maximum weight matching problem on Bipartite graph
- We extend the general problem by adding
  - *no intersection points on the edges that are the solutions to the problem*



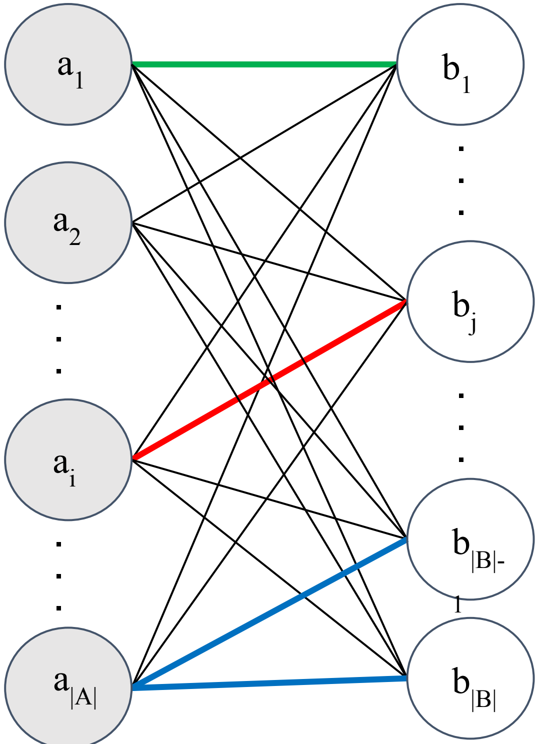
Red : solutions of the extended problem
Blue : solutions of the general problem

Algorithm | Implementation as a dynamic programming



Green: Selected edges as solutions
Red: An edge ECM is analyzing
Blue: Edges ECM will analyze future

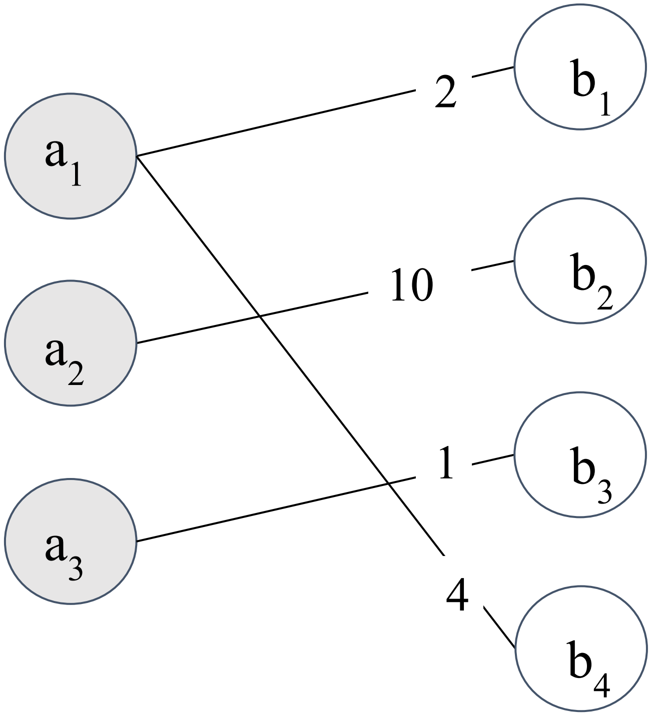
Algorithm | Implementation as a dynamic programming



	1	2	3	…	j	…	|B|
1	70	66	99	57	56	76	94
2	2	18	73	10	82	69	3
3	27	26	13	96	79	89	22
…	58	85	54	38	46	67	30
i	8	55	14	78			
…							
|A|							

Green : Selected edges as solutions
Red : An edge ECM is analyzing
Blue : Edges ECM will analyze future

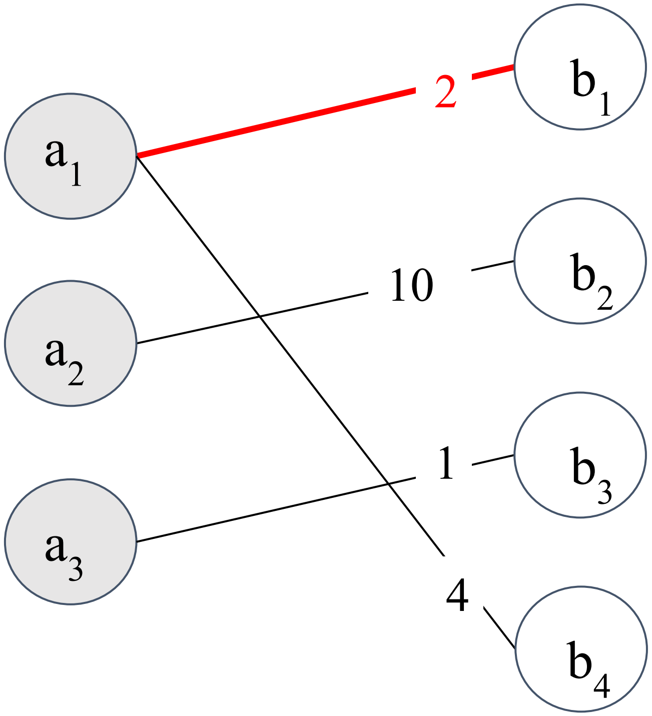
Tables used in DP



W	b_1	b_2	b_3	b_4
a_1	2	0	0	4
a_2	0	10	0	0
a_3	0	0	1	0

DP	0	0	0	0
0				
0				
0				

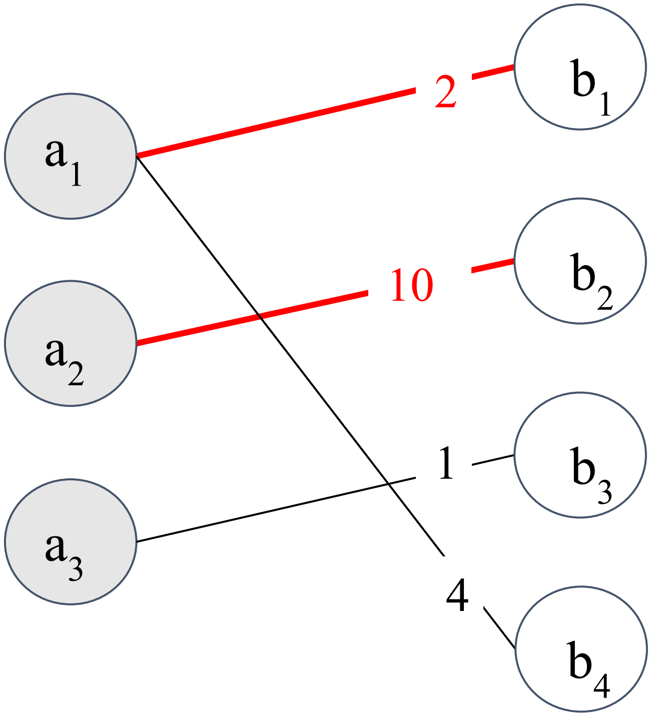
Tables used in DP



W	b_1	b_2	b_3	b_4
a_1	2	0	0	4
a_2	0	10	0	0
a_3	0	0	1	0

DP	0	0	0	0
0	2			
0				
0				

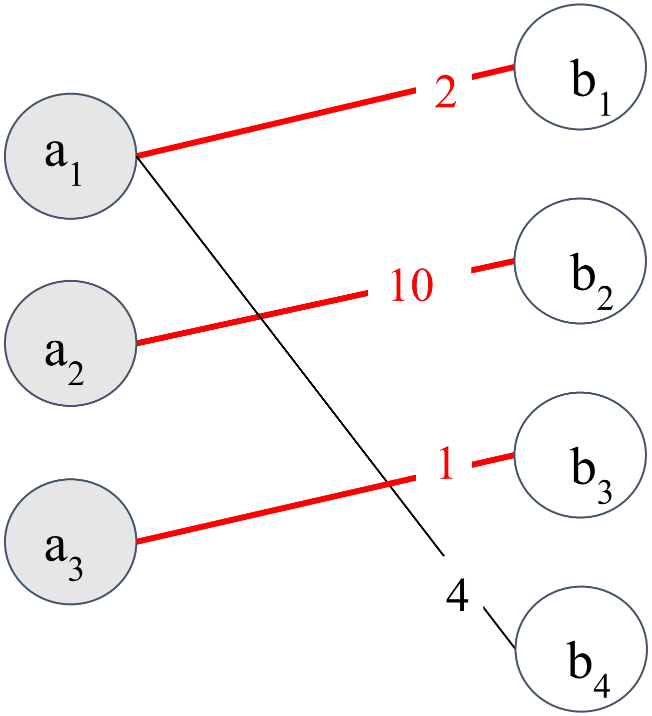
Tables used in DP



W	b_1	b_2	b_3	b_4
a_1	2	0	0	4
a_2	0	10	0	0
a_3	0	0	1	0

DP	0	0	0	0
0	2			
0		12		
0				

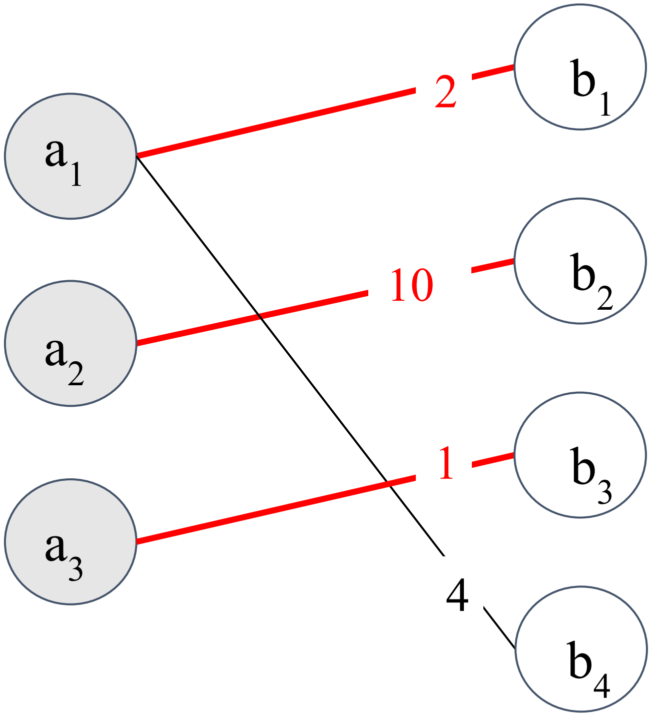
Tables used in DP



W	b_1	b_2	b_3	b_4
a_1	2	0	0	4
a_2	0	10	0	0
a_3	0	0	1	0

DP	0	0	0	0
0	2			
0		12		
0			13	

Tables used in DP



W	b_1	b_2	b_3	b_4
a_1	2	0	0	4
a_2	0	10	0	0
a_3	0	0	1	0

DP	0	0	0	0
0	2	2	2	6
0	2	12	12	12
0	2	12	13	13

Experimental evalution | Dataset

- W2E[6]: Wikipedia-based causal relationship dataset
    a. Every causal relationship has a category
    b. Num. of causal relationship: 322
    c. Num. of events: 1,041
    d. Num. of categories: 10
    e. Ave. num. of events per causal relationship: 3.23

[6]: Tuan-Anh Hoang, Khoi Duy Vo, and Wolfgang Nejdl. 2018. W2E: A WorldwideEvent Benchmark Dataset for Topic Detection and Tracking. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (Torino, Italy) (CIKM '18). Association for Computing Machinery, New York, NY, USA, 1847–1850.

Experimental evalution | Dataset

	S	AA	BE	ST	AC
Ave. Num. of events	2.92	3.26	3.85	2.0	2.66
Num. of topics	13	73	7	2	3
Ave. Num. of tokens	29.44	27.15	34.59	23.25	34.37
	LC	**PE**	**IR**	**DA**	**HM**
Ave. Num. of events	3.06	3.14	3.43	3.11	4.66
Num. of topics	32	84	57	45	6
Ave. Num. of tokens	34.53	31.87	38.20	29.54	28.25

S: Sports

AA: Armed conflict and Attack

BE: Business and Economy

ST: Science and Technology

AC: Arts and Culture

LC: Law and Crime

PE: Politics and Election

IR: International relations

DA: Disaster and Accident

HM: Health and medicine

Examples of W2E

Category	TOPIC ID Event texts	TOPIC ID Event texts
Politics and elections	2016-07-13 The new Prime Minister of the United Kingdom Theresa May begins forming her ministry following the end of the Second Cameron ministry. 2016-07-14 Elizabeth Truss is named Secretary of State for Justice and first ever female Lord Chancellor of the United Kingdom as former chancellor Michael Gove is ousted from the cabinet.	TOPIC-1780 2016-06-29 The process to elect a new leader of the Conservative Party to replace outgoing Prime Minister David Cameron begins in the United Kingdom. 2016-07-05 Home Secretary Theresa May gets 165 votes after the first ballot of Conservative members of parliament to select a new Leader and the next Prime Minister. 2016-07-11 Prime Minister David Cameron announces he will step down on Wednesday, July 13.

Baselines

1. BM25:

   Bag-of-words-based ranking function

2. TF-IDF + dynamic time warping（DTW）

   DTW evaluates the similarity between data considering time series

Bipartite graph construction

Each node is constructed by TF-IDF vector

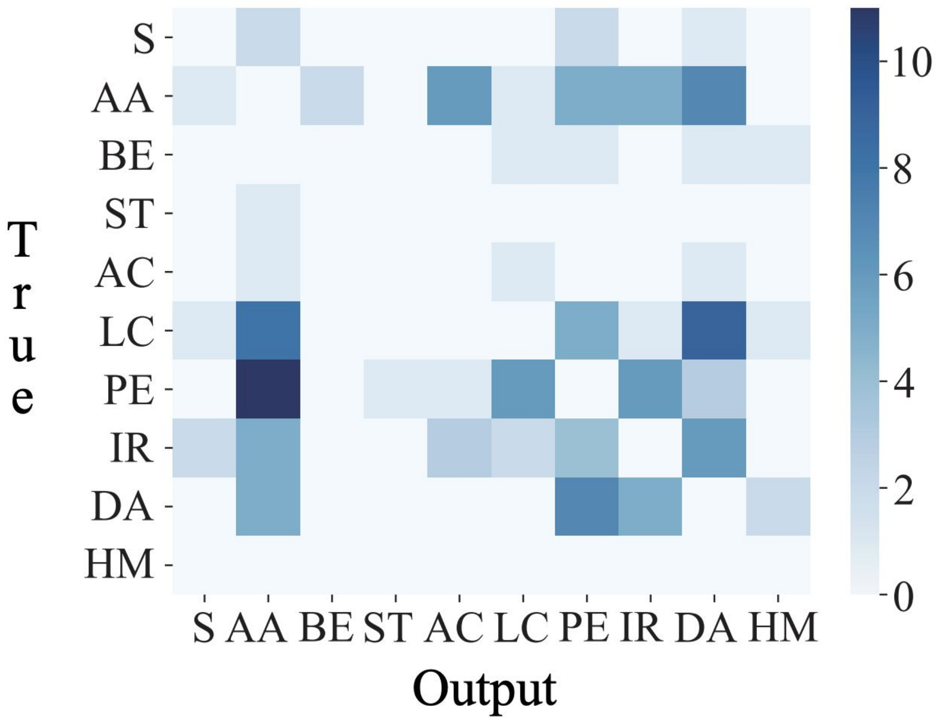
Each weight of an edge is a result of cosine similarity

	TF-IDF	LSA	LDA	Doc2Vec
CosSim	**59.8%**	48.8%	26.4%	14.3%
Euclid	17.7%	11.2%	20.2%	15.0%
JS	14.0%	-	17.7%	-

Results

	MAP	P	R	F_1
BM25	33.1%	25.8%	25.8%	25.8%
DTW	54.4%	49.4%	49.4%	49.4%
ECM	**62.5%**	**59.8%**	**59.8%**	**59.8%**

Results | Prediction distribution



S: Sports

AA: Armed conflict and Attack

BE: Business and Economy

ST: Science and Technology

AC: Arts and Culture

LC: Law and Crime

PE: Politics and Election

IR: International relations

DA: Disaster and Accident

HM: Health and medicine

Error Analysis | Correct ratio per category

Sport	Armed conflicts	Business	Science	Arts	Law	Politics	International	Disaster	Health
61.5%	63.0%	42.8%	50.0%	0.0%	21.8%	66.6%	61.4%	57.7%	100.0%
13	73	7	3	3	32	84	57	45	6

Error Analysis | Jaccard coefficient and mutual information

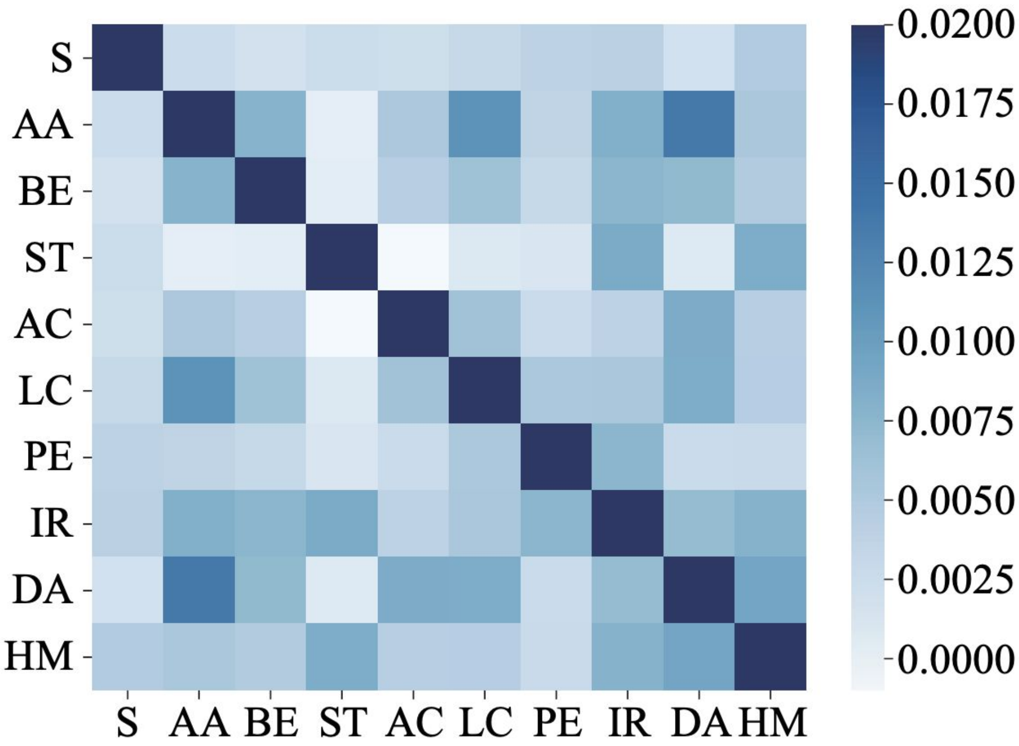
Sport	Armed conflicts	Business	Science	Arts	Law	Politics	International	Disaster	Health
0.11	0.11	0.09	0.19	0.08	0.10	0.12	0.10	0.13	0.13
0.10	0.10	0.11	0.27	0.05	0.09	0.13	0.10	0.13	0.10

Whole average

Jaccard: 0.12

Mutual information: 0.12

Error Analysis | Mutual information



S: Sports

AA: Armed conflict and Attack

BE: Business and Economy

ST: Science and Technology

AC: Arts and Culture

LC: Law and Crime

PE: Politics and Election

IR: International relations

DA: Disaster and Accident

HM: Health and medicine

Conclusion

We proposed an algorithm to evaluate the similarity of causal relationships.

- Goal：measuring the similarity of events with a focus on causality to facilitate historical analogies
- Proposed algorithm: using DP to calculate similarity between two relationships

Future works

- Using the proposed algorithm in history learning tool to evaluate how the algorithm is useful to enhance historical analogy