

2021/3/3 @ オンライン, F31-4 (10:00~)

因果関係に基づく類似出来事の検索

東京都立大学

澄川 靖信



背景 | 歴史の重要性

- 歴史を知ることが重要
 - 現代の形成過程の理解
 - 過去の知見を現代で活用
 - 小学校から開講されている基礎科目の一つ



背景 | 歴史の重要性

- 歴史を知ることが重要
 - 現代の形成過程の理解
 - **過去の知見を現代で活用**
 - 小学校から開講されている基礎科目の一つ

- 歴史学習支援の実際・研究
 - 内容理解 → 思考力育成 → **歴史を活用できることが目標**[1]
 - 学校教育における**歴史活用能力を支援する学習環境の研究**[2]

[1]: 文部科学省 (2018): 学習指導要領

[2]: P., Lee (2005) : Historical Literacy: Theory and Research, History Education Research Journal 5(1)

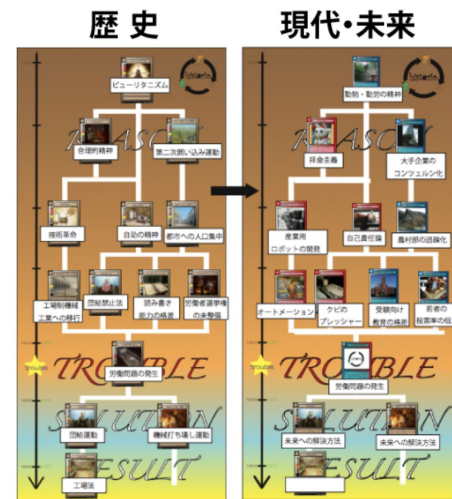


背景 | 因果関係に着目する意義

- 過去の知見を現代・未来に活用出来る

背景 | 因果関係に着目する意義

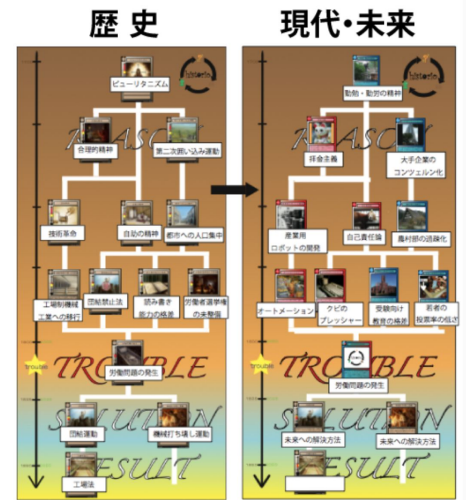
- 過去の知見を現代・未来に活用出来る
- 歴史的類推の促進[3]
 - カードゲームで過去と現代の出来事の因果関係の類似性を見出す教材
 - 題材はゲームの作者が決め、手動で構造を決定



背景 | 因果関係に着目する意義



- 過去の知見を現代・未来に活用出来る
- 歴史的類推の促進[3]
 - カードゲームで過去と現代の出来事の因果関係の類似性を見出す教材
 - 題材はゲームの作者が決め、手動で構造を決定
- **将来**起こりうる出来事を予測出来る[4~5]
 - A→Bの関係性を学習
 - A'→○の○を**予測**



[3]: 池尻良平 (2011): 歴史の因果関係を現代に応用する力を育成するカードゲーム教材のデザインと評価 . 日本教育工学会論文誌 34巻4号, 375-386.

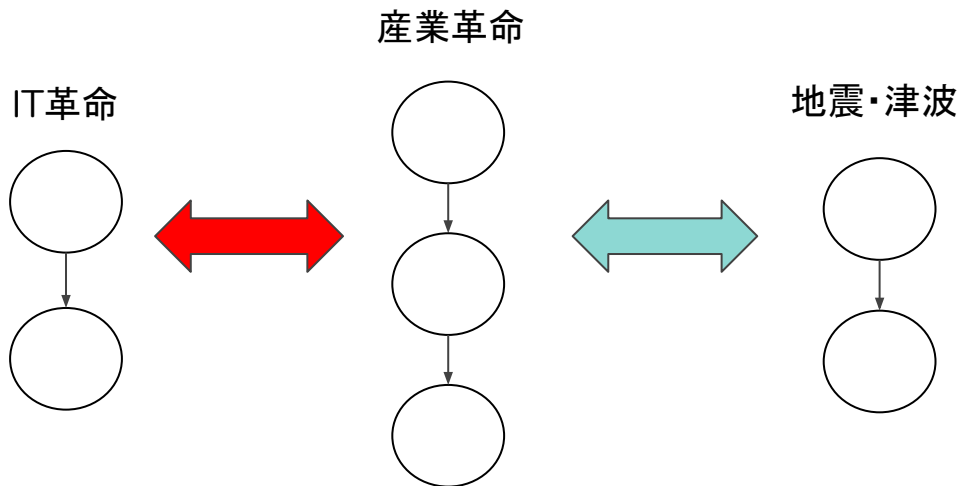
[4]: K., Radinsky, S., Davidovich, S., Markovitch (2012): Learning to Predict from Textual Data. J. Artif. Intell. Res. 45: 641-684.)

[5]: A., Jatowt, C.-m. A., Yeung (2011): Extracting collective expectations about the future from large text collections. CIKM: 1259-1265



目的

- 因果関係を表す**出来事集合の類似度を評価**
 - 前提: 出来事のグラフは予め定義されている





本発表での貢献

- 主な貢献は以下の通り



本発表での貢献

- 主な貢献は以下の通り
 - a. 因果関係を表す出来事間の類似度を求める**アルゴリズムを提案**



本発表での貢献

- 主な貢献は以下の通り
 - a. 因果関係を表す出来事間の類似度を求める**アルゴリズムを提案**
 - b. **トイデータセット**を用いた評価
 - c. **広く使われている手法**との比較



本発表での貢献

- 主な貢献は以下の通り
 - a. 因果関係を表す出来事間の類似度を求める**アルゴリズムを提案**
 - b. **トイデータセット**を用いた評価
 - c. **広く使われている手法**との比較

- 今後の課題
 - a. 因果関係を木構造で表現できる一般化
 - b. ground truthとなるデータセットの構築とその上での評価
 - c. 文書検索アルゴリズムのstate-of-the-artな手法との比較
 - d. 歴史的類推を促進させる**学習環境としての検索エンジン**の実現



目次

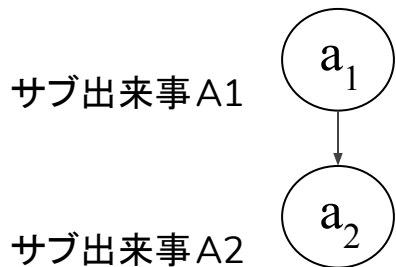
- 提案手法を適用するためのデータ表現の定義
- アルゴリズム
- 実験
- まとめ



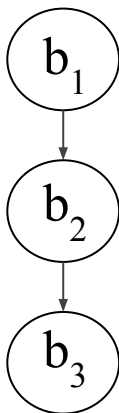
データ表現

比較する因果関係はサブ出来事を節とする線形リストで表す

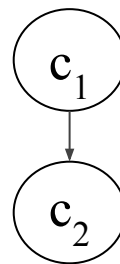
出来事A: IT革命



出来事B: 産業革命



出来事C: 地震・津波

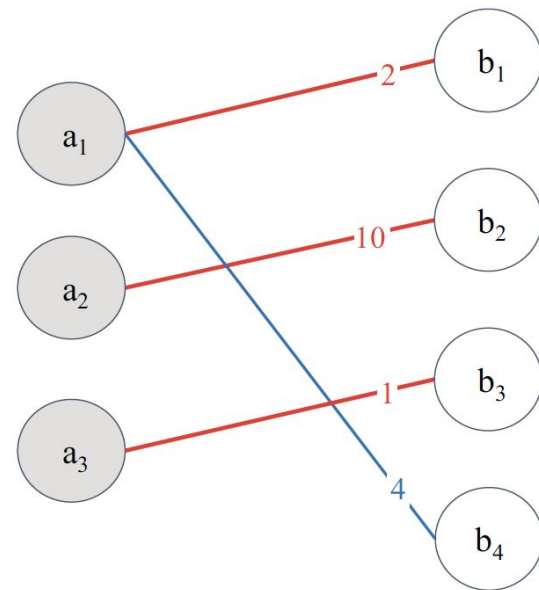




データ表現

比較する因果関係はサブ出来事を節とする線形リストで表す

因果関係の比較のために二部グラフを構築する



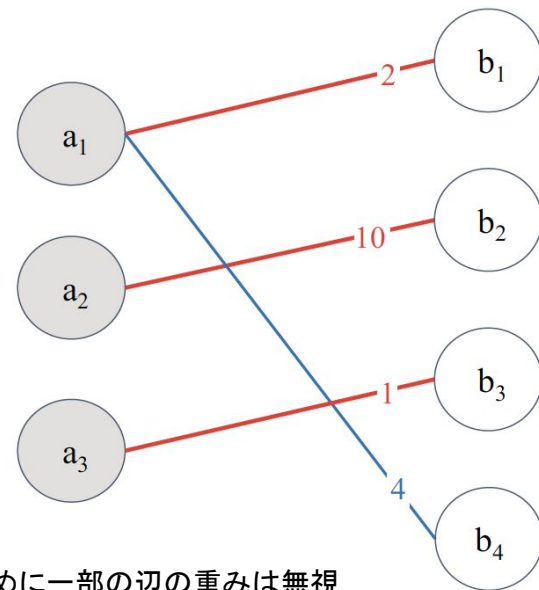


データ表現

比較する因果関係はサブ出来事を節とする線形リストで表す

因果関係の比較のために二部グラフを構築する

- サブ出来事同士の類似度を辺の重みとする



※プレゼンのために一部の辺の重みは無視

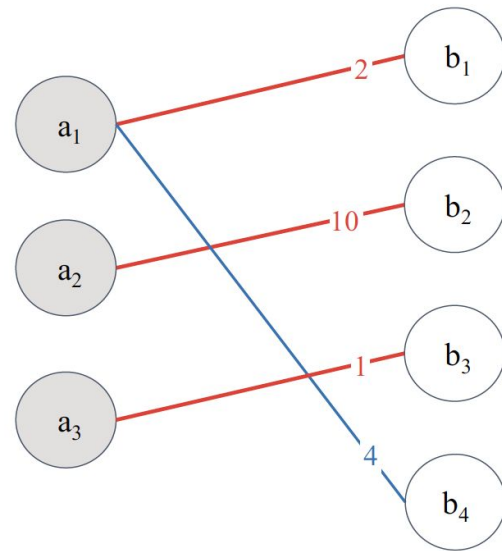


アルゴリズム | 理論

- 二部グラフ上での**最大重みマッチング**を拡張

赤線: 本研究で解く最大重みマッチング

青線: 一般的な最大重みマッチング





アルゴリズム | 理論

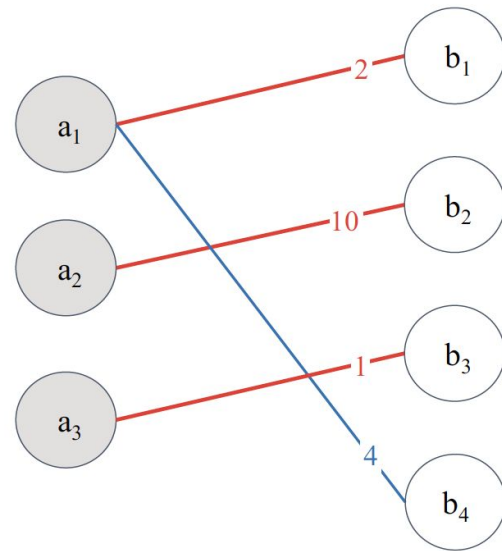
- 二部グラフ上での**最大重みマッチング**を拡張

赤線: 本研究で解く最大重みマッチング

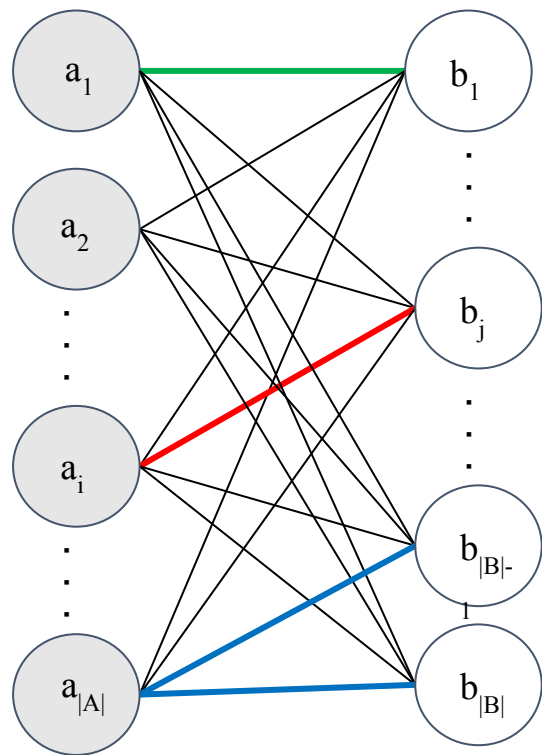
青線: 一般的な最大重みマッチング

- 拡張方法: 以下の制約を追加

解となる辺集合で交点は無い



アルゴリズム | 動的計画法

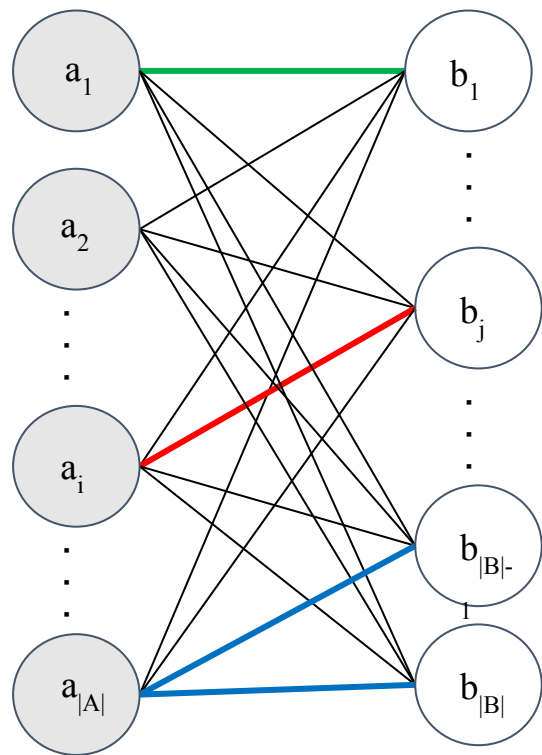


緑線: 解として選択済みの辺

赤線: 解に含めるか分析中の辺

青線: 未だ分析していない辺

アルゴリズム | 動的計画法



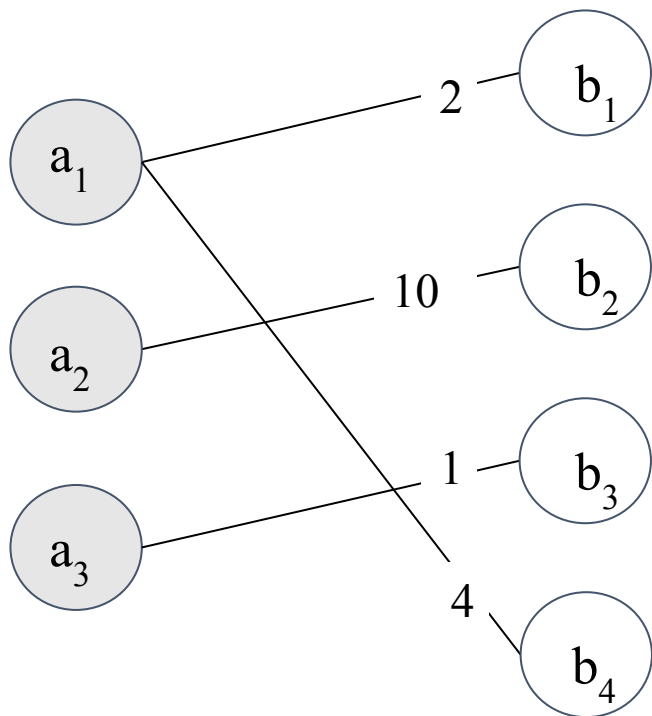
	1	2	3	...	j	...	B
1	70	66	99	57	56	76	94
2	2	18	73	10	82	69	3
3	27	26	13	96	79	89	22
...	58	85	54	38	46	67	30
i	8	55	14	78			
...							
A							

緑線: 解として選択済みの辺

赤線: 解に含めるか分析中の辺

青線: 未だ分析していない辺

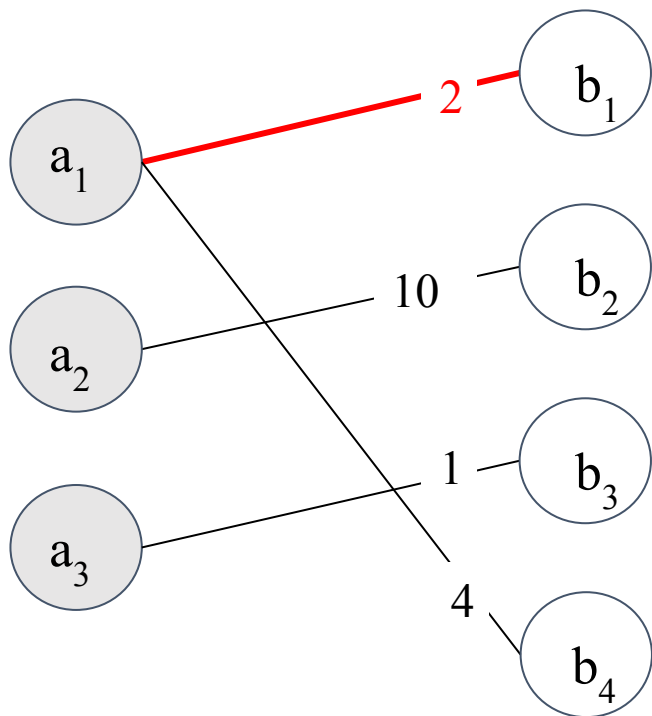
アルゴリズム | 動的計画法



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				

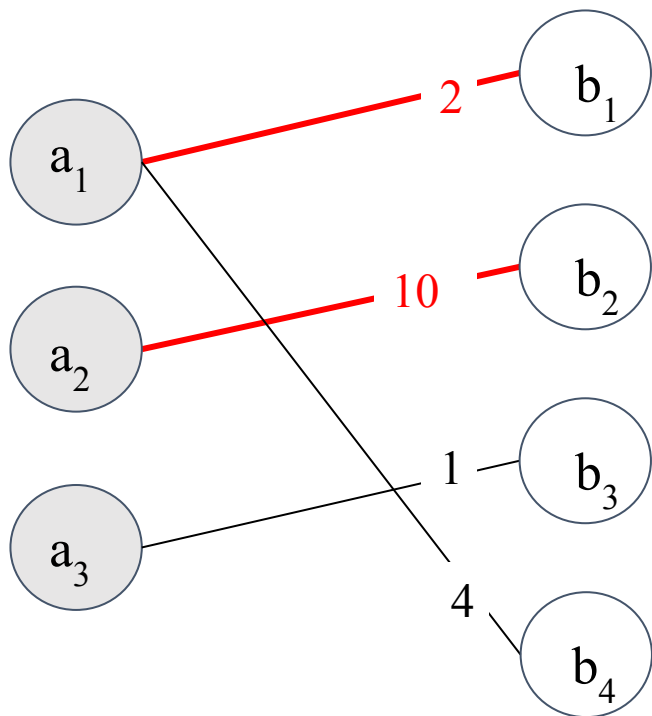
アルゴリズム | 動的計画法



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				

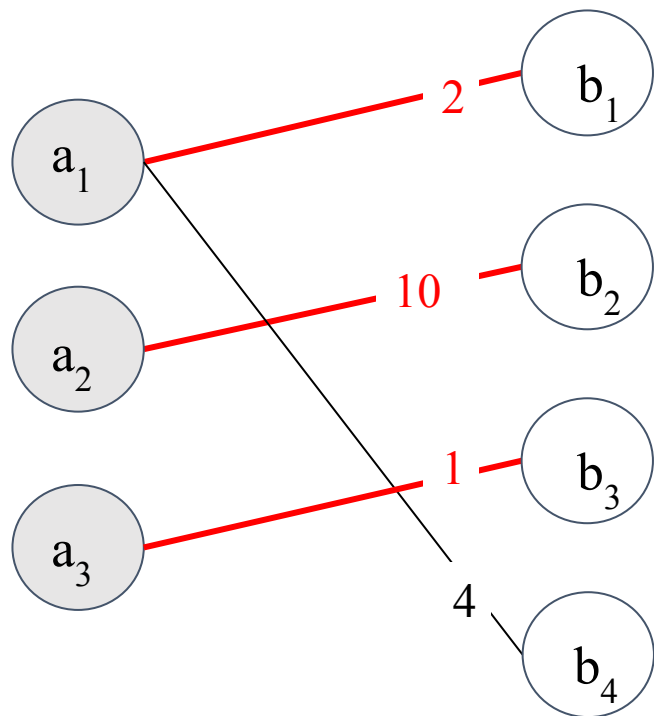
アルゴリズム | 動的計画法



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				

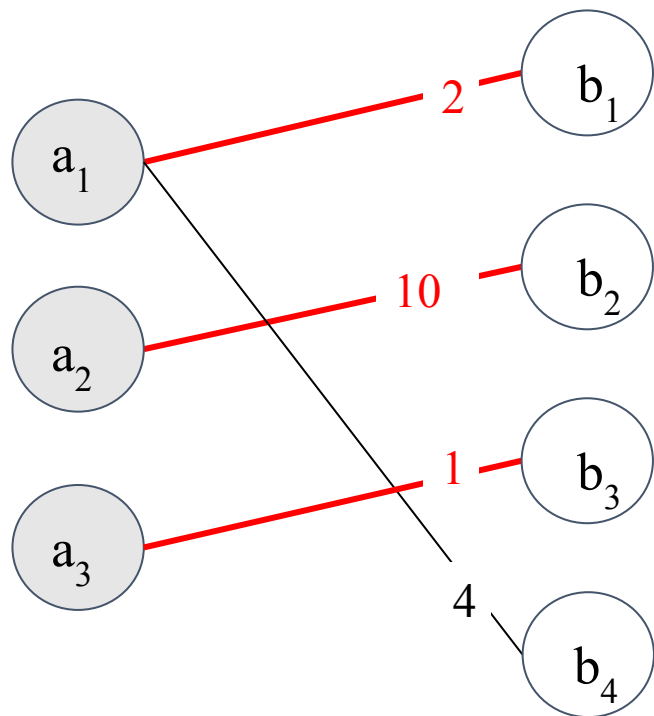
アルゴリズム | 動的計画法



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	

アルゴリズム | 動的計画法



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



実験 | データセット

- Wikipediaから記事を収集
 - a. WikipediaでEventとして定義されているカテゴリを参考に記事を収集した。
 - b. セクションごとにサブ出来事が記述されているものを対象

- 因果関係として表1のものを収集

表1 評価用データセットの統計情報.

地震	222
地震と津波	57
地震と地崩れ	3
ウィルスの発生とワクチン開発	5
ウィルスの発生と治療	6
計	293



実験 | 比較対象(先行研究2+提案手法)

1. コサイン類似度

Wikipedia記事の本文を抽出して純粋な類似文章の検索



実験 | 比較対象(先行研究2+提案手法)

1. コサイン類似度

Wikipedia記事の本文を抽出して純粋な類似文章の検索

2. 動的時間伸縮法(DTW)

時系列性を考慮したデータ間の類似度を評価(信号処理でよく使われている)



実験 | 比較対象(先行研究2+提案手法)

1. コサイン類似度

Wikipedia記事の本文を抽出して純粋な類似文章の検索

2. 動的時間伸縮法(DTW)

時系列性を考慮したデータ間の類似度を評価(信号処理でよく使われている)

3. 提案手法

※ 特徴ベクトルはTF-IDFのみで生成

※提案手法ではTF-IDF+コサイン類似度で辺の重みを計算



実験 | 評価($p@1$)

- 交差検定(分割数: 10)で訓練/テストデータに分割
- 次の表は結果の平均

コサイン類似度	DTW	提案手法
0.567	0.753	0.858



まとめ

本発表: 因果関係の類似度を評価するアルゴリズムを提案した。

- 研究のゴール: 歴史的類推を促進するための因果関係に着目した類似出来事を 検索できる学習環境の実現
- 本研究の提案手法: 2つの出来事の類似度をDPで評価
- 実験: Wikipediaにある因果関係が記載されている記事+一般的な手法

今後の課題

1. アルゴリズム: 因果関係の表現形式をグラフ構造に一般化
2. 実験: 専用のデータセットを構築して本格的な評価の実施