

2023/03/02@IPSJ 全国大会

歴史的類推を促すための 現代の因果関係検索学習支援システム

松丸 健太¹ 池尻 良平² 澄川 靖信¹

1.拓殖大学 2.東京大学大学院

本研究の目的と背景

・ 目的

- ・ 学んだばかりの高校の歴史の授業の内容を、現代の諸問題に活かすためのシステムを開発する

・ 背景：高等学校における歴史教育

- ・ 暗記や歴史的思考力の育成を目指すだけでなく、
歴史的思考力を現代に応用する段階まで考慮するべき¹

[1]文部科学省（2018）・高等学校学習指導要領 解説地理歴史編（平成 30 年 8 月改訂版）

・ 歴史的思考力²

- ・ 批判的に読む力 ・ 因果関係を理解し、結びつける力
- ・ 現代の諸問題と結びつけて考えたりする力

[2]池尻良平, 澄川靖信：“真正な社会参画を促す世界史の授業開発—その日のニュースと関連した歴史を検索できるシステムを用いて—”, 社会科研究, 84, pp.37–48 (2016).

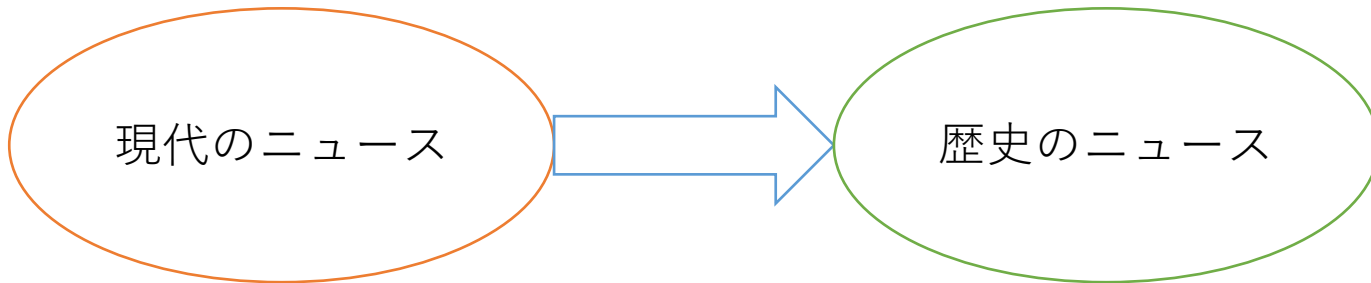
先行研究

・ 目的

- ・ 歴史的思考力 の育成
- ・ 現代のニュースを入力として、類似した歴史のニュースを出力するシステムを開発する
- ・ 現代のニュースの解決方略を考える際に有効なシステム

・ 問題点

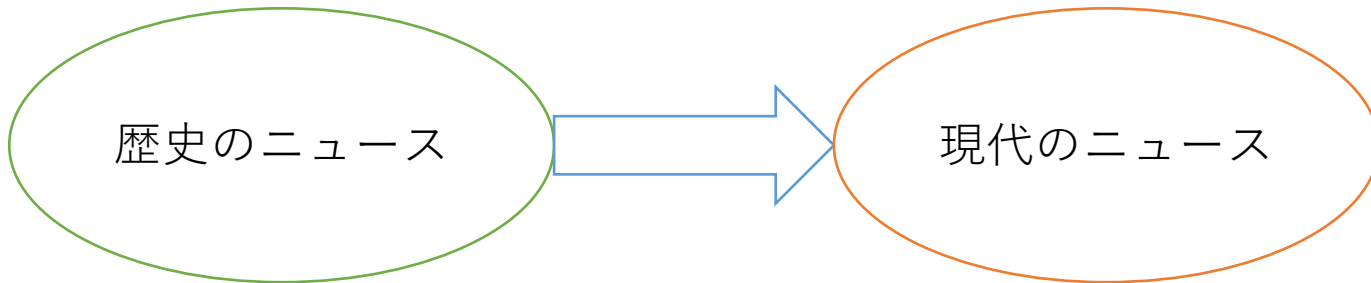
- ・ 検索結果が歴史の為、利用前から歴史についての知識が必要
- ・ 授業で利用するには、リアルタイム性に欠けるため難しい



本研究の課題

・課題

- ・ 授業で学んだばかりの歴史のニュースをを入力に用いる事で、リアルタイムに利用可能なシステムを作成
- ・ 歴史についての知識が充分でない高校生でも容易に学習が可能になる
- ・ 歴史のニュースから類似した現代のニュースを検索するシステムを開発する



本研究の目的

- ・ 歴史の因果関係から類似した現代の因果関係を
検索するシステムを作る事で歴史的思考力を育成し、
現代で生じている歴史教育における
諸問題の解決方略を考える足場かけとする。

本システムの実行例 | 入力と出力の例

入力例: 【歴史の因果関係】

1. Due to the mass movement of the Carolingians and the **invasions** by the Islamic powers, the Mediterranean culture that had continued since ancient times lost its unity.
2. In response, Charles the Great, the Frank king, invited many scholars to his court and promoted cultural activities.
3. As a result, there was a Latin-based literature re-vival, which became known as the Carolingian Renaissance, after which there was considerable cultural progress

出力例: 【現代の因果関係】

1. 2021-9-26 The Taliban calls for international commercial flights at the Kabul International Airport to resume after the airport was closed for **evacuation**.
2. 2021-9-27 ISIL – K leaders Farooq Bengalzai and Mawlawi Ziya ul-Haq are killed and more than 80 more fighters are arrested during Taliban raids in Nangarhar. The raids are carried out as ISIL has killed several people during attacks in Jalalabad in recent days.

本研究の実行例 | 入力と出力の類似点

- キーワード
 - 侵攻を意味する **invasions**
 - 避難を意味する **evacuation**
- 類似点
 - イスラム系のニュース
 - 外敵からの進行によって起こった出来事

開発システム

- 文章の正規化

トークン化

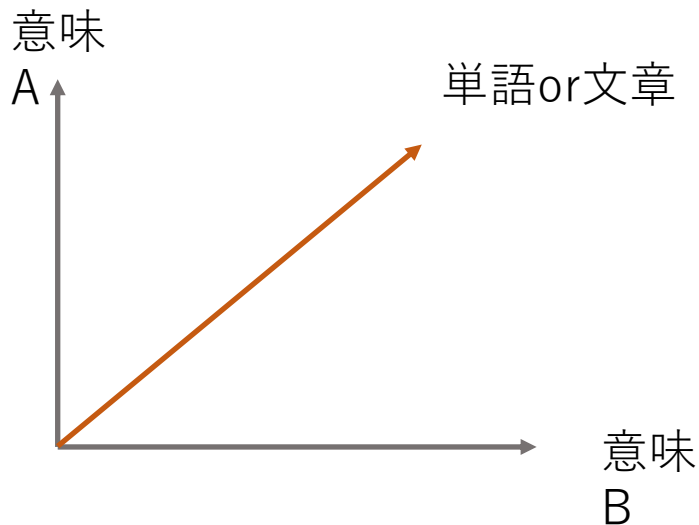
レンマ化

ストップワードの除去

- 時間的固有表現の抽出

TagMe³を用いてWikipediaの
情報から変換

- LSAを用いて文章をベクトル化
- ECMで類似度を測定



[3]Ferragina, P. and Scaiella, U.: Fast and Accurate Annotation of Short Texts with Wikipedia Pages, IEEE Software, Vol. 29, No. 1, pp. 70–75 (2011).

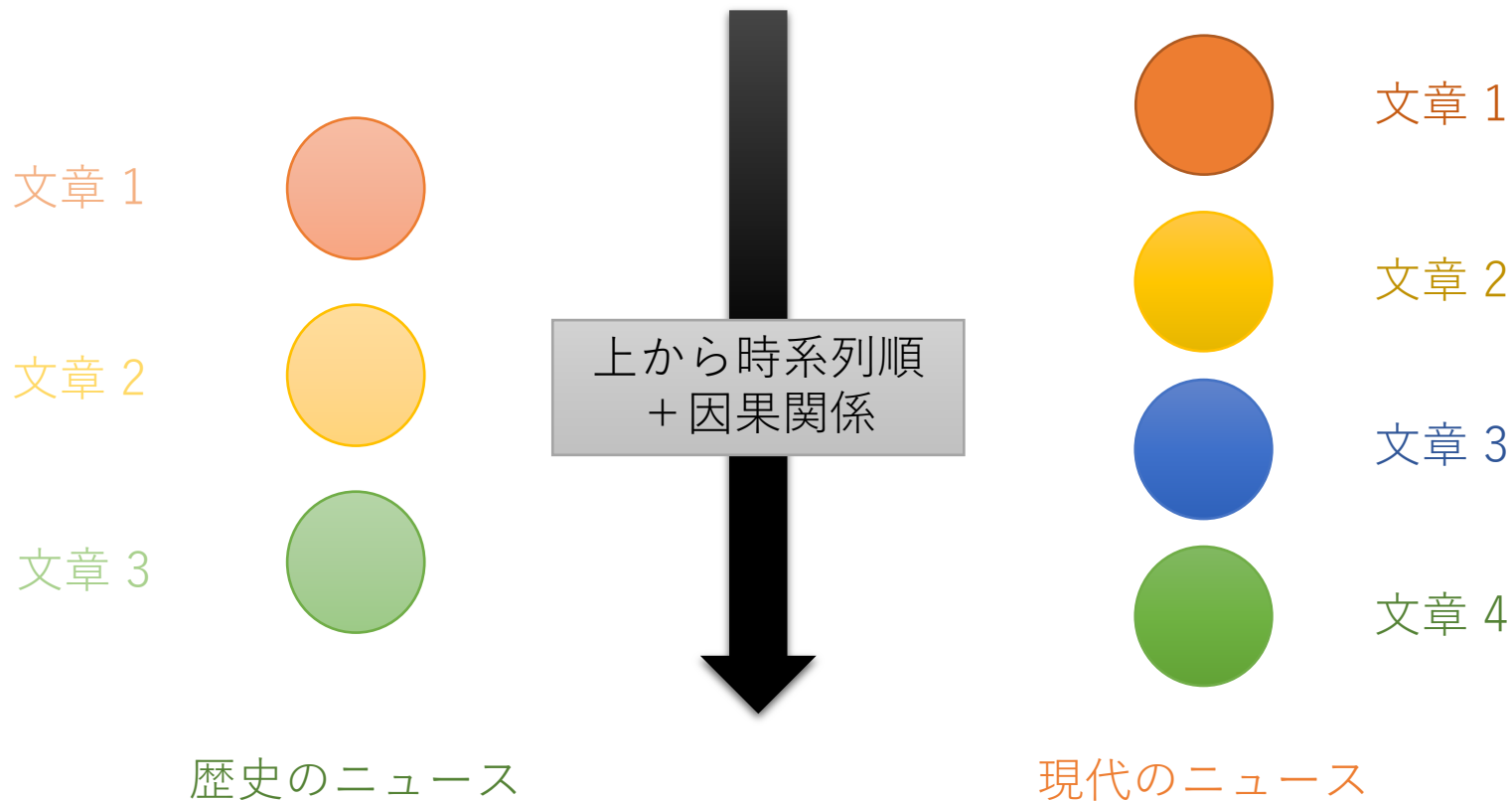
提案手法① | ECM⁴

- ・ Event Causal Relationship Measurement(ECM)とは

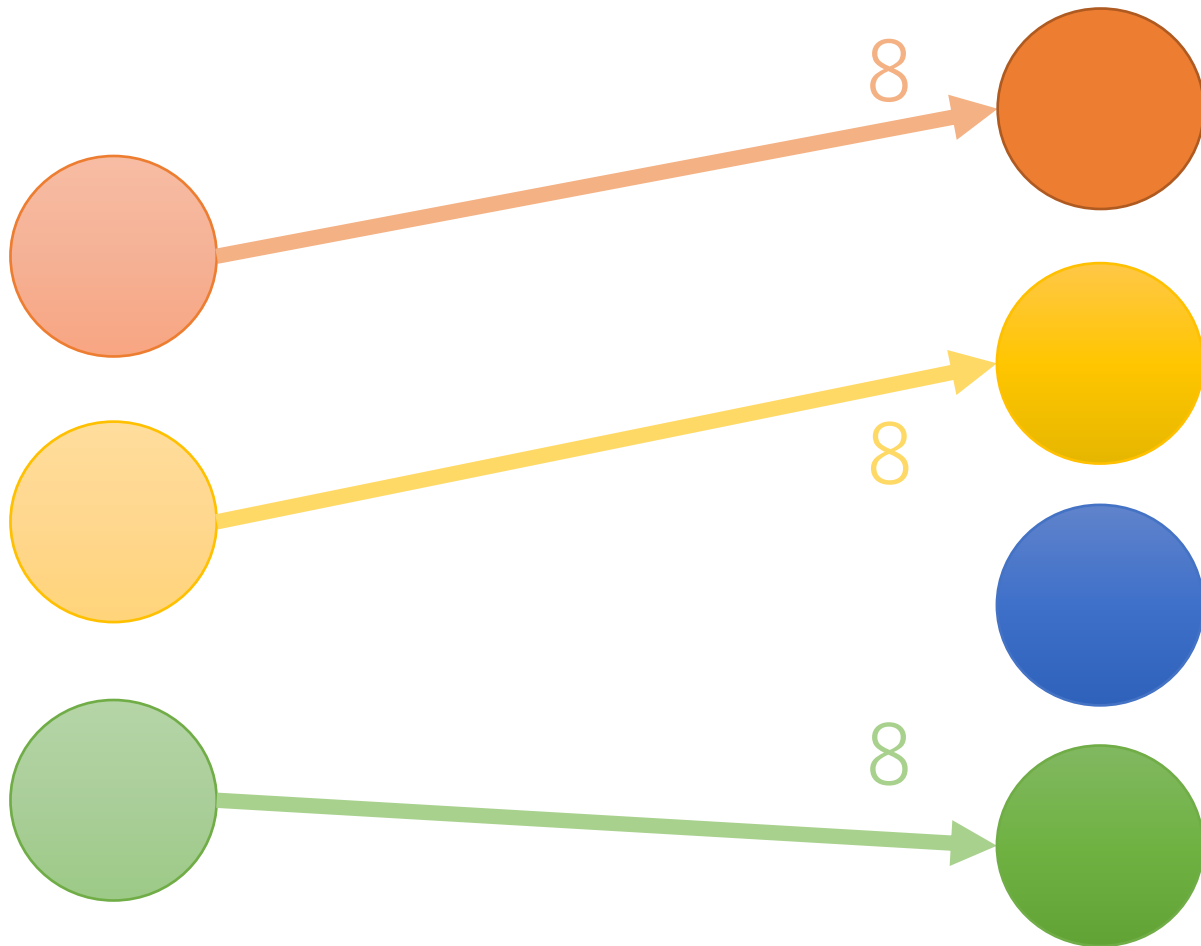
ノード間の交点を無くす事で
時系列を考慮した最大重みマッチングから
因果関係の類似度を算出するマッチング手法

[4]Sumikawa, Y.: Event Causal Relationship Retrieval. WI-IAT '21. pp. 318–325, 2021.

最大重みマッチング

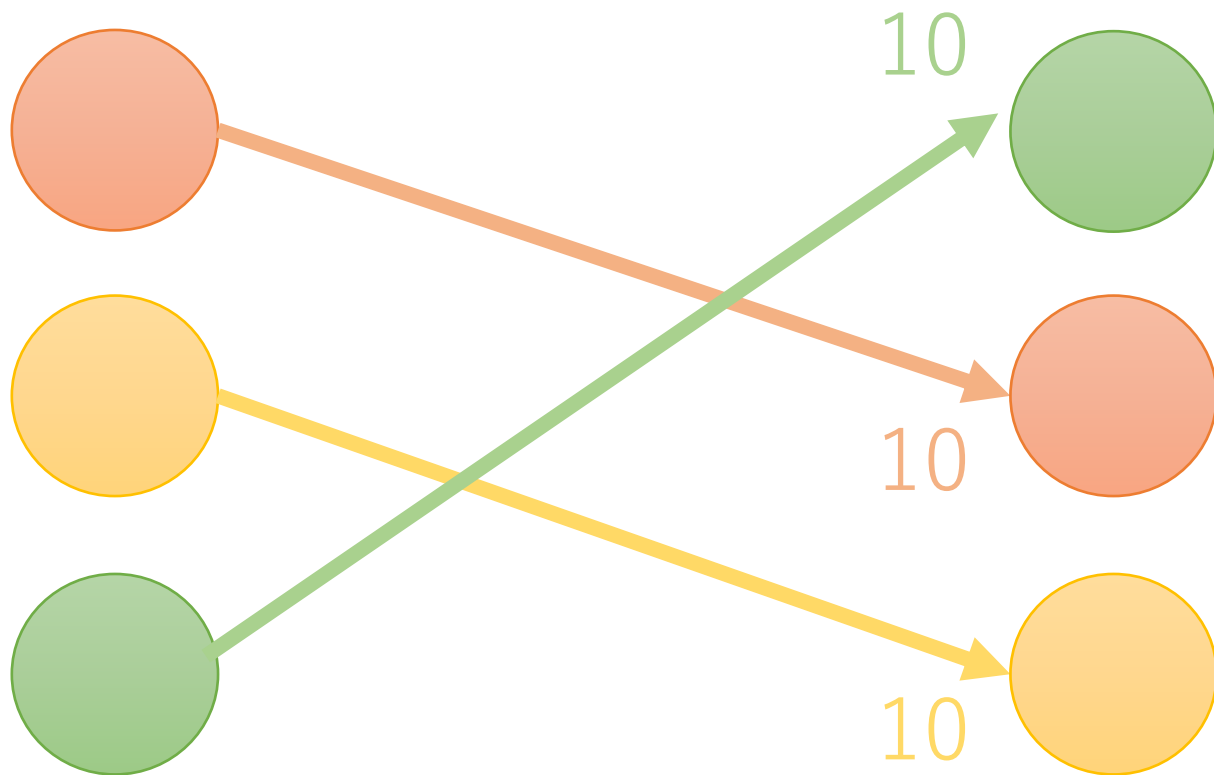


最大重みマッチング | 例①



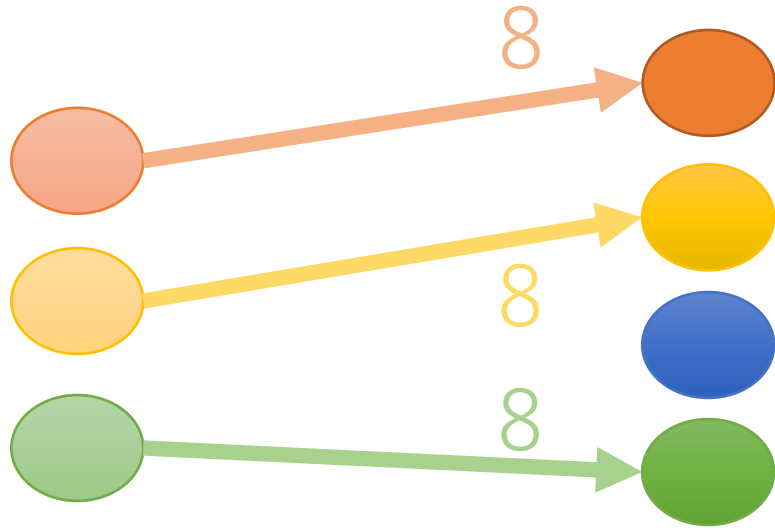
合計 24

最大重みマッチング | 例②

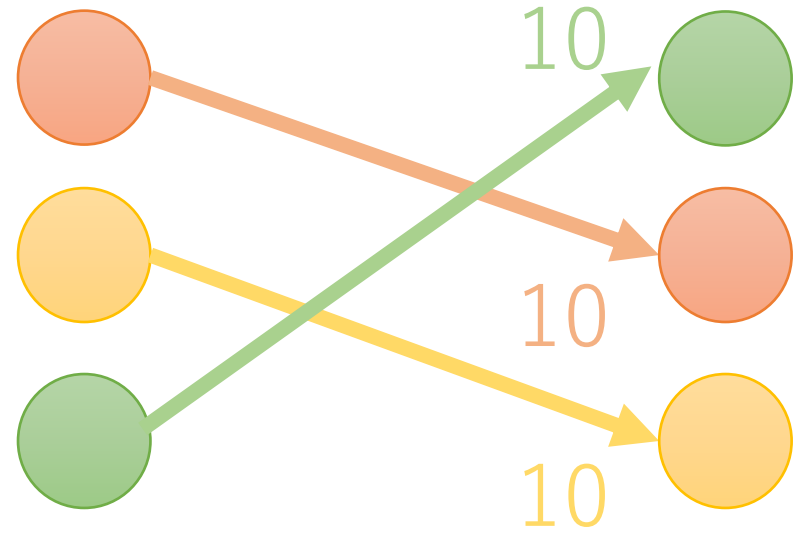


合計 30

最大重みマッチング | 比較



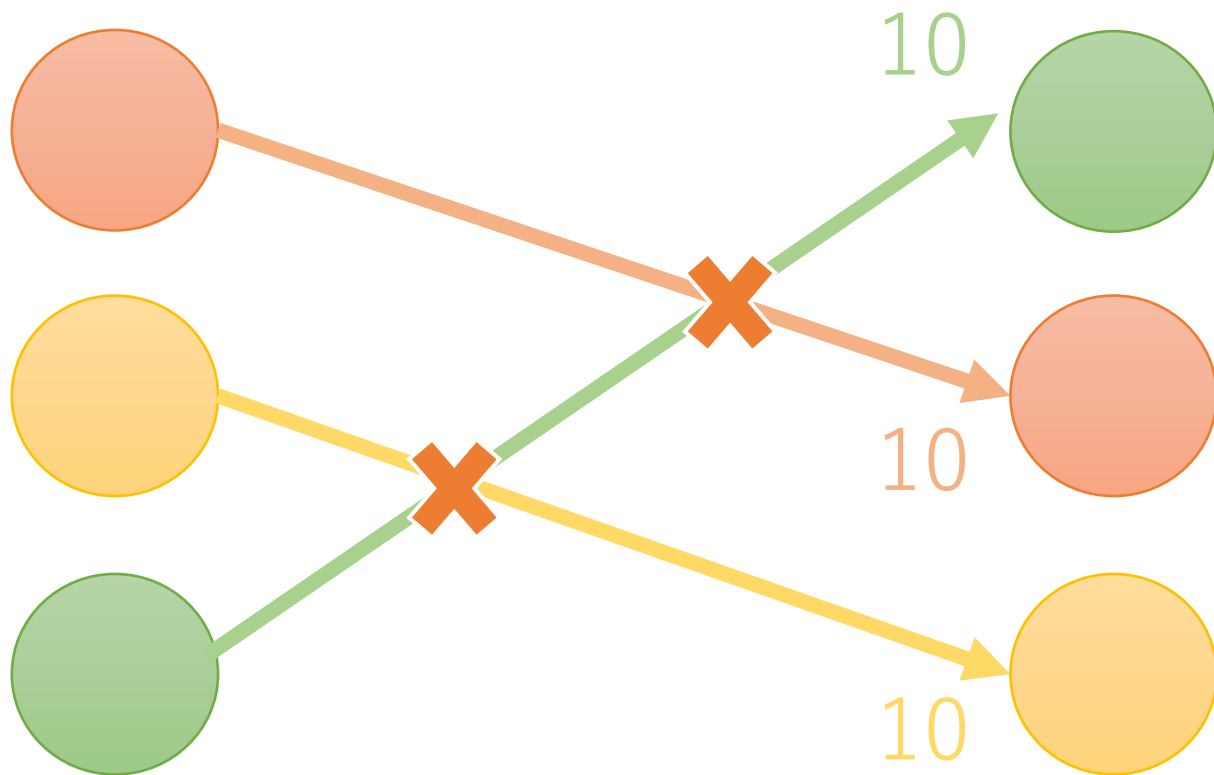
合計 24



合計 30

こちらの方が類似していると判断

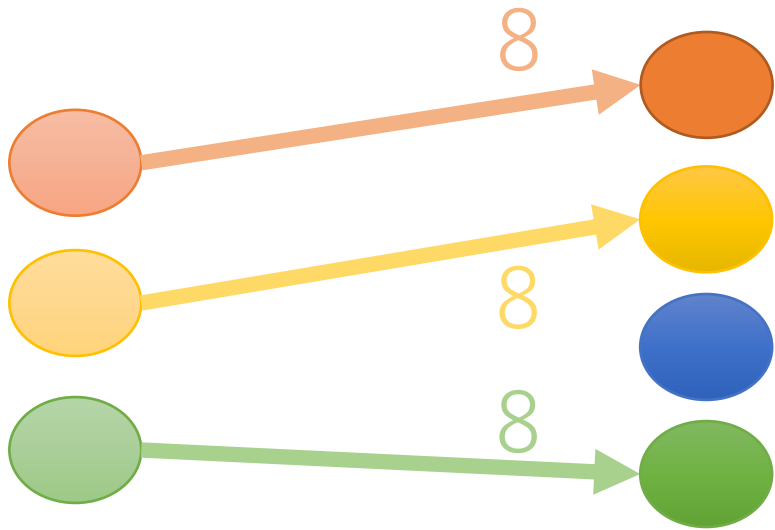
ECM | 交点を作らない



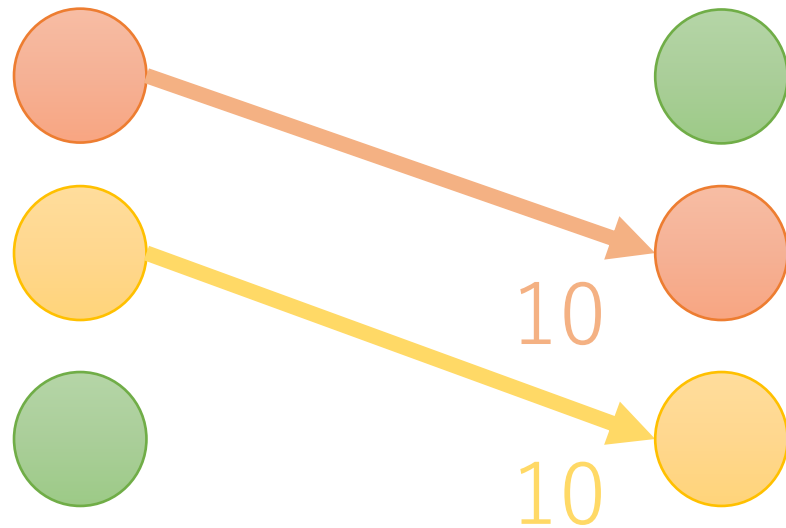
合計 20

- 文章同士を時系列に沿って比較
- 文章間の因果関係を考慮した測定が可能になる

ECM | 比較



合計 24



合計 20

交点を無くし、時系列に即した算出をすると、
こちらの方が似ていると判断

提案手法② | 時間的固有表現

- 時間的固有表現

人・組織などのような、ある一定期間に頻出する固有名詞の事を指す (例: 織田信長)

- 非時間的固有表現

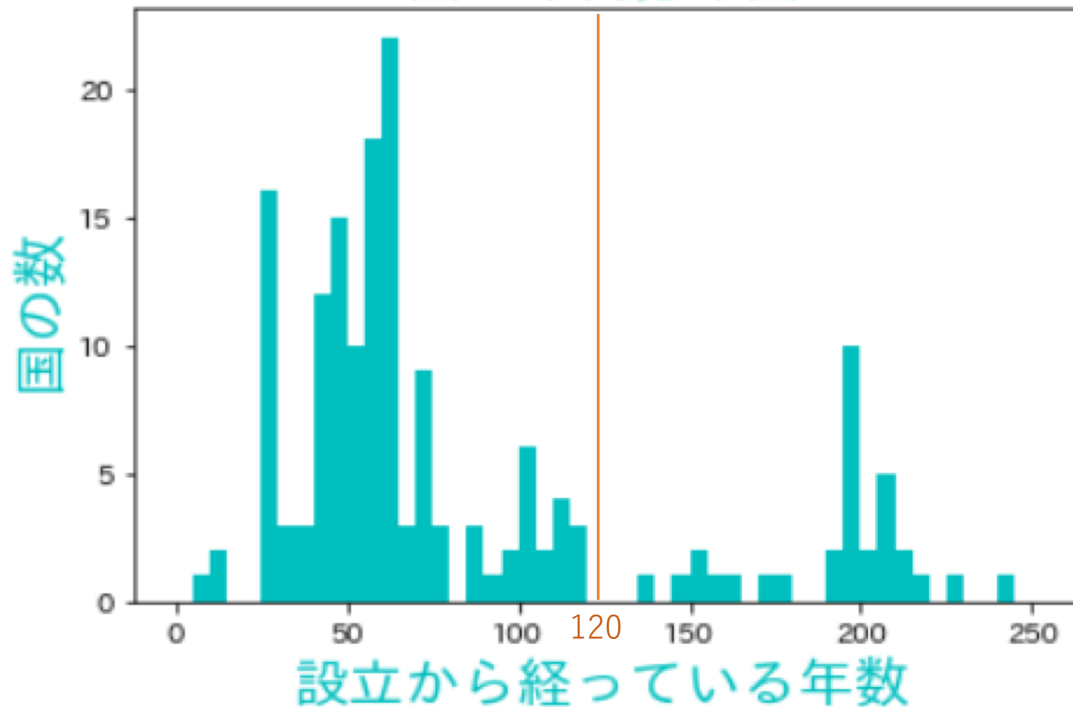
長い期間存在している固有名詞 (例: フランス)

提案手法② | 時間的固有表現

- 時間的固有表現の変換
 - 時間的固有表現を文章中から抽出し、より抽象的な固有表現に変換
- Tagme
 - 固有表現の抽出が可能
- 一定期間の設定
 - 抽出した固有表現からwikipediaのカテゴリを抽出
 - 付与されたカテゴリを単語から期間の閾値を設定

提案手法② | 閾値の例

国の年代分布図



時間的

非時間的

提案手法② | カテゴリ変換

カテゴリ	変換後	年数
国	Country	120年以下
都市	City	120年以下
組織	Organization	年数制限無し
人	People	年数制限無し
出来事	Event	年数制限無し

時間的固有表現の変換例

変換前

06-09 United Nations Secretary-General Ban Ki-moon temporarily
remove Saudi Arabian-led intervention Yemen UN blacklist violating child`s
right supporter threatened stop funding many UN program
06-12 Missiles fired US drone strike four-wheel drive vehicle southern
Yemen killing two injuring another suspected militant

変換後

06-09 United Nations Secretary-General Ban Ki-moon temporarily
remove country Arabian-led intervention country organization blacklist violating child`s
right supporter threatened stop funding many organization program
06-12 Missiles fired US drone strike four-wheel drive vehicle southern
country killing two injuring another suspected militant

実験 | 評価

・ 評価基準

- ・ 歴史の因果関係を入力とし、出力が類似した現代の因果関係として適切であるかどうかを判断する

① 分類の評価に広く用いられている適合率 (P), 再現率 (R), F1 値 (F1), MAP 値を用いて評価する

② 混合行列で表し, ヒートマップで結果を可視化

・ 比較対象

1. Jaccard 係数を用いた文章の集合の類似度を求める手法
2. 文を分割せず コサイン類似度を用いた潜在意味解析の手法
3. 時間的固有表現を用いない ECM のみの手法

実験 | 評価用データセット

- 歴史のニュースのデータセット

- 歴史の教科書を参考に専門家が手作業で作成した過去の因果関係データセット⁵

[5] Ikejiri Ryohei, & Sumikawa Yasunobu. (2019). Past Causalities and Event Categories for Connecting Similar Past and Present Causalities [Data set]. Zenodo.

- 現代のニュースのデータセット

- 2016年に起きた出来事としてWikipediaに記録されたものを手作業で時系列順にトピックとしてまとめた現代の因果関係データセット⁶

[6] T.-A. Hoang, K. D. Vo, and W. Nejdl.

W2e: A worldwide-event benchmark dataset for topic de-tection and tracking.

実験 | 準備

- 判定方法

- 歴史と現代のデータセットは文章毎にカテゴリが付与
- 入出力の文章同士でカテゴリ一致度を判定

- 問題点

- 歴史と現代でカテゴリの名称が異なっている

- 解決策

- 現代のカテゴリは歴史のカテゴリと類似したものを選定し、名称を統一した

実験 | 準備

歴史カテゴリ		類似する現代カテゴリ1	類似する現代カテゴリ2
Diplomacy	→	International relations	Business and economics
Reign	→	Politics and elections	
Literature	→	Arts and culture	
Environment	→	Health and environment	
Technology	→	Science and technology	

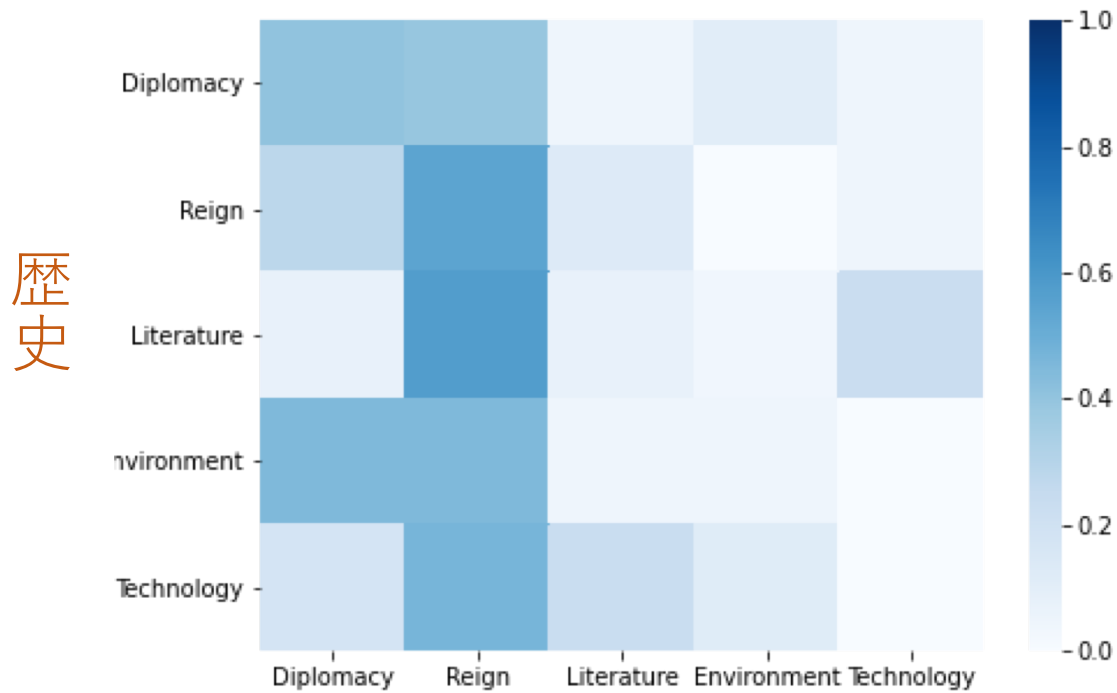
結果 | 各値の比較

表 1 適合率, 再現率, F1 値, MAP 値の比較

	<i>P</i>	<i>R</i>	<i>F1</i>	<i>MAP</i>
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%
提案手法	54.4%	51.4%	53.1%	71.3%

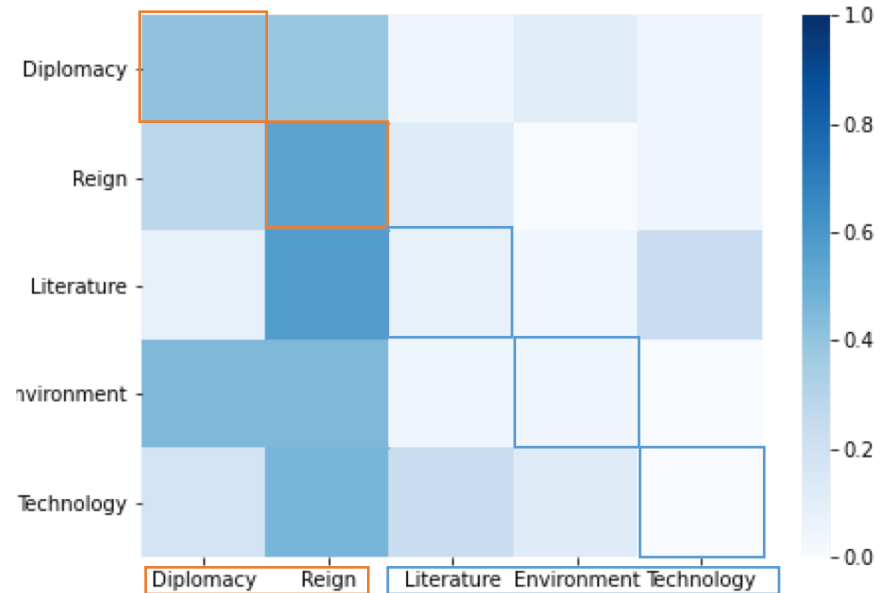
ECM と時間的固有表現の抽出を組み合わせた本手法が有効である

結果 | ヒートマップ



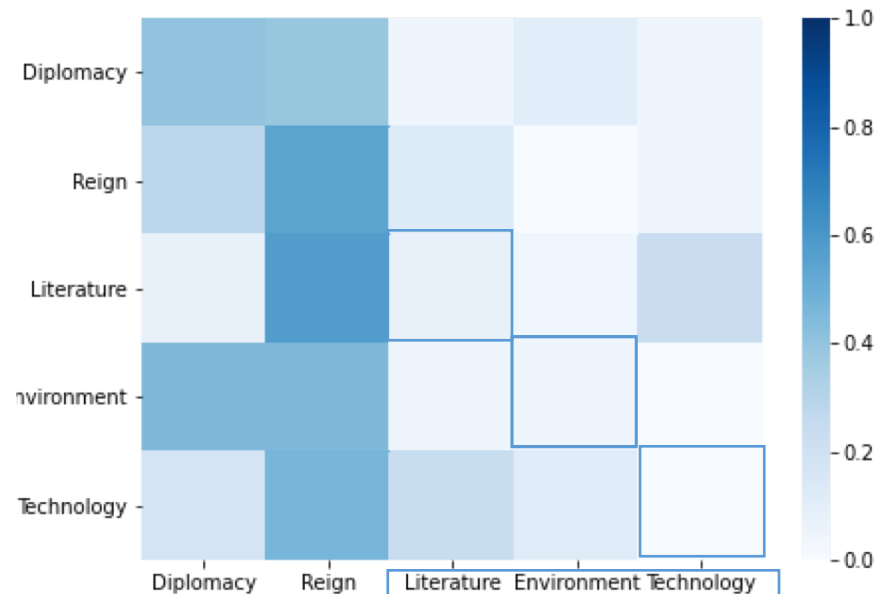
分析 | ヒートマップ

- Diplomacy Reignのカテゴリで強い相関が見られた
- Literature Environment Technologyのカテゴリではあまり相関が見られなかった



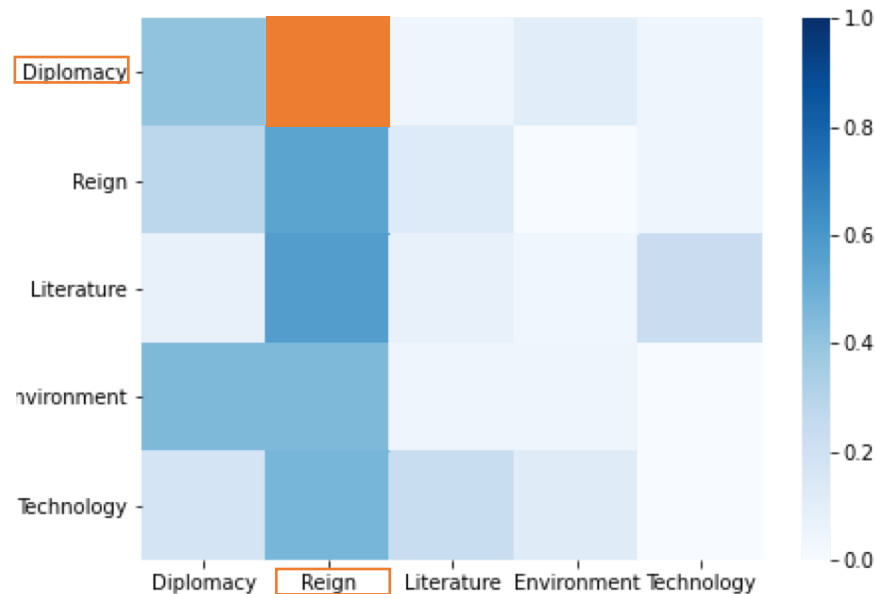
分析 | ヒートマップ

- 強い相関がみられなかったのカテゴリの歴史のデータセットは他二つに比べてデータ数が少ない
- 適切なサンプル数でのデータ抽出が行えなかった



分析 | カテゴリ不一致例

- DiplomacyとReignの相関関係が高い
- 現代の国際関係や経済カテゴリをDiplomacyのカテゴリとして変換
- 国内統治や海外侵略等の出来事がReignカテゴリの文脈と一致する部分が見られた



分析 | カテゴリ不一致 文章例

歴史 Diplomacy

- スパルタはスパルタ市民奴隷で構成されていた 前者は常に反乱を恐れていた 後者は市民の団結を維持するために努力した：通貨の使用は禁止されていた 富の格差が生じる
- 土地は平等に分配されていた 完全に平等にする
- 国は閉じていた 外部からの影響を防ぐ こうしてスパルタ市民の結束は高まった スパルタはギリシャ最強の軍隊国家となった。

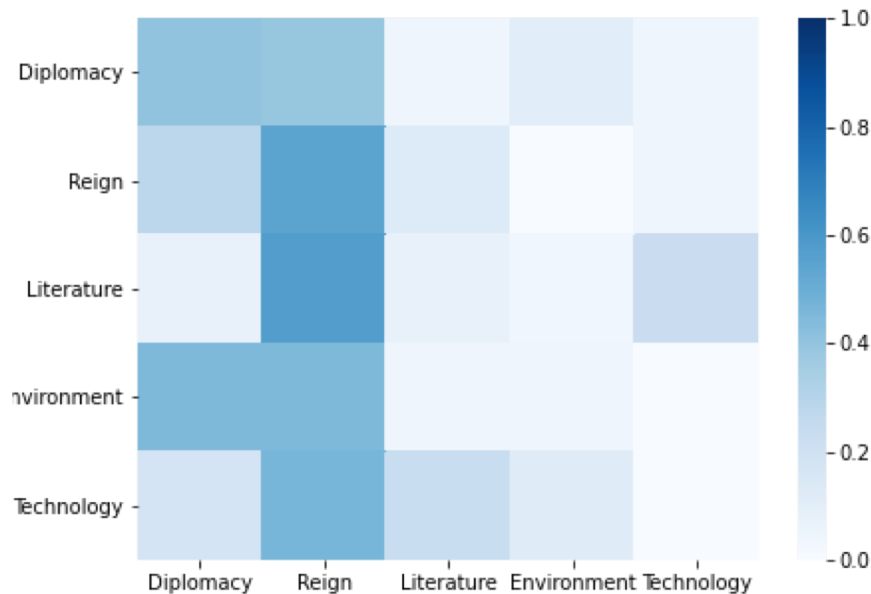
現代 Reign

- 2016-06-22 欧州中央銀行が権利放棄を復活させ、ギリシャの銀行が同国の国債を担保にした通常融資を初めて可能にした。

経済関連のニュースが類似した結果、カテゴリの不一致が起きた

分析 | カテゴリ不一致

- 変換前カテゴリに含まれる意味的なニュアンスの違い
- カテゴリ一致面でより高い精度での検証が行える余地がある



議論

- より多くのサンプル数での検証が必要
- より検索の精度を上げる為に、統一したお互いのカテゴリ内の文章を再検討
- 適切なカテゴリの付与をされた文章を抽出する必要がある

本研究のまとめ

- 高等学校における歴史教育の支援システムとして、歴史の因果関係から類似した現代の因果関係を出力するプログラムを開発
- ECMと時間的固有表現の抽出を行い、カテゴリの一致度を評価
- 実験では提案手法が他の手法よりも5.5~15%程度、他手法より効果的な結果が得られた

今後の課題

- 本システムを用いて歴史の授業に組み込むための効果的な学習方法の提案
 - 授業で学んだばかりの歴史の出来事から現代の出来事について考察させる
 - 歴史思考力を現代に応用する段階まで考慮
- 本システムを適切に利用できるUIの開発
 - 高校生でも簡易に利用可能なものにする

