

歴史マイニング：現代社会の問題状況に類似した歴史の発見

澄川 靖信 *
東京理科大学 *

池尻 良平 **
東京大学 **

1 はじめに

近年、国内外の歴史教育では、過去に起きた問題の解決方略を学び、その知見を現代社会で起きている問題の解決に用いることが重要だと考えられている [2, 4]. ある文脈で得た知識を異なる文脈で使用することは**転移** (transfer) と呼ばれ、1つの題目に沿った歴史を学習することによって得られた知見を、現代社会の問題解決に転移させる効果的な学習方法が明らかになっている [6]. しかしながら、教育に関する研究分野では、学習内容と現実世界で生じている問題を適切に関連付けることによって学習効果が高まることが知られており、この学習方法は**真正な学習** (authentic learning) と呼ばれる [5].

本研究では、学習者が現代社会で実際に生じている問題の特性と説明文を入力すると、クラスタリングと半教師あり学習に基づく文章分類の2つの手法を基に類似した歴史を学習者に提示する学習環境（以降、歴史マイナーと呼ぶ）を実現する。歴史教育で使用可能な真正な学習を提供する学習環境を実現したのは本研究が初めてである。

以降の構成は次のとおりである。第2節で本研究で使用する文章データの表現形式を定義する。第3節で類似度を求める2つの手法について説明した後、第4節でまとめる。

2 文章ファイルの表現形式

Lee は、歴史と現代は社会的な変化のテーマによって接続しうることを指摘している [2]. そこで、本研究で使用する歴史は社会的な変化に関するものに限定し、著者らが教科書などをもとに、原因、問題、解決方法、結果の順で記述したデータを用いる。この順序で記述した文章は、転移を促進させる効果があることが知られている [1]. また、現代社会の問題状況については、実際の新

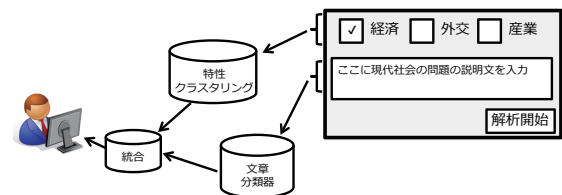


図1 システム全体像

聞記事を用いる。

3 類似度を求めるアルゴリズム

現代社会の問題と歴史の類似度を2種類の手法によって求める。1つは、問題の特性に着目し、より多くの特性が一致する歴史をクラスタリングする。もう1つは、各歴史をカテゴリと考慮して新聞記事を分類する。この2つの手法を組み合わせることで、それぞれの利点を活かしながら類似した歴史を求める。前者の手法の利点は、歴史マイナーの利用者が調べたい特性を選択するだけで関連する歴史を閲覧できることである。一方、後者の手法は、現代の問題を記述した新聞記事を基に分類するので、利用者の先入観に依存せずに分類できる利点がある。

本節では、まず、歴史マイナーの全体的なアルゴリズムについて述べる。次に、2つの手法の詳細について述べる。

3.1 システム設計

歴史マイナーの全体像を図1に示す。学習者は、現代社会の問題の特性を選択し、その問題を記述している新聞記事の文章を入力する。これらの情報を使用して前述した2種類の手法を並列に実行する。最後に、それぞれの結果を基に歴史を提示する。

3.2 問題の特性に基づくクラスタリング

社会的な問題は、その特性に着目すると、経済、外交、産業などに属すると考えられる。そこで、Lee が定義する社会的変化の枠組み [2] を参考に、どの時代にも存在する類似した因果関係を抽出し、それらを問題の特性とする。

History Mining: Discovering History Based on Similarity Between History and Modern Problems

*Sumikawa Yasunobu, Tokyo University of Science

**Ryohei Ikejiri, The University of Tokyo

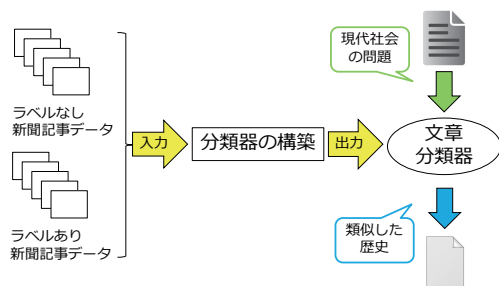


図2 使用する歴史をカテゴリとする分類器の全体像

この結果を用いて、すべての歴史と現代社会の問題を、それぞれ、該当する特性の特徴ベクトルとして表現する。これらの特徴ベクトルから行列を作成し、その転置行列との積を取ることで、各データと一致する特性の数を成分とする行列を求める。更に、この行列を用いてデータ間の類似度を距離で表し、現代社会の問題との距離が短い上位 k 個の歴史を結果とする。

3.3 使用する歴史をカテゴリと考慮した分類

本研究で扱う歴史は、ある一定期間で繰り返し起こる社会的な問題なので、ある歴史に似た現代の問題を新聞記事に記述されていることがある。そこで、歴史の説明文と新聞記事を用いた文章分類によって、類似した歴史を求めることができる。しかしながら、本研究で使用する歴史データは第2節で述べた順序で記述しなければならないので、歴史データだけを使用して分類器を学習するために必要なデータ数を作成するためには高いコストが必要となる。一方、新聞記事は大量のデータを容易に取得することができる。したがって、歴史をカテゴリと考慮して少数の新聞記事に類似している歴史名をラベル付けすることによって、大量のラベルなし新聞記事を用いて分類器を学習する半教師あり学習によって分類器を構築することが可能となる [3]。

本研究では、CD-毎日新聞データ集 2012^{*1}に収録されている少数の新聞記事に対して類似している歴史名をそれぞれラベル付けする。このデータを基に分類器を学習し、ラベルなし新聞記事へのラベル付けと分類器の再学習を繰り返す。学習する分類器は、単純ベイズやサポートベクターマシン、決定木といった、文章分類で一般的に用いられているものを使用する。

半教師あり学習による分類器の構築は、図2に示すように、学習者が歴史マイナーを使用する前に行う。この設計によって、現代社会の問題を入力すると、直ちに分

類結果を得ることができる。

4 まとめと今後の課題

本研究では、歴史学習における転移の促進と真正な学習を可能とする学習環境を実現するために、問題の特性を利用したクラスタリングと、歴史をカテゴリと考慮する文章分類の2種類の手法を用いて現代社会の問題に類似した歴史を求めるアルゴリズムを提案した。

今後の課題としては、次の3つが考えられる。1) 歴史マイナーの出力結果の精度の評価。2) 出力する歴史の順序の決定とインターフェースの開発。3) 歴史マイナーを実際に高校生に使用してもらい、どのような転移が起こるかの学習評価。

謝辞 本研究は、科学研究費助成事業平成26年度若手研究(B) 26750076「歴史を現代社会の問題解決方略に転移させる学習メディアの開発と評価」の助成を受けて、実施している。

参考文献

- [1] L. M. Gick and J. K. Holyoak. Analogical problem solving. *Cognitive Psychology*, Vol. 12, pp. 306–355, 1980.
- [2] P. Lee. Historical literacy: Theory and research. *International Journal of Historical Learning, Teaching and Research*, Vol. 5, No. 1, pp. 25–40, 2005.
- [3] Kamal Nigam, Andrew Kachites McCallum, Sebastian Thrun, and Tom Mitchell. Text classification from labeled and unlabeled documents using em. *Mach. Learn.*, Vol. 39, No. 2-3, pp. 103–134, May 2000.
- [4] 文部科学省. 高等学校学習指導要領, 2009.
- [5] 米国学術研究推進会議編著. 授業を変える - 認知心理学のさらなる挑戦. 北大路書房. , 2002.
- [6] 池尻良平, 藤本徹, 椿本弥生, 山内祐平. 歴史的事象を現代の問題解決に応用する力を育成する教材のデザインと評価. *教育メディア研究*, Vol. 19, No. 1, pp. 1–12, 2012.

^{*1} CD-毎日新聞データ集 2012 の利用については、毎日新聞社、日外アソシエーツとの契約の下行っている。