

# 半教師あり NMF を用いた裁判事例に関連する憲法条項の推定

山本 京佳<sup>†</sup>    山西 良典<sup>††</sup>    松下 光範<sup>††</sup>

<sup>†</sup> 関西大学大学院総合情報学研究科    〒 569-1095 大阪府高槻市霊仙寺町 2-1-1

<sup>††</sup> 関西大学総合情報学部    〒 569-1095 大阪府高槻市霊仙寺町 2-1-1

E-mail: <sup>†</sup>{k166981,ryama,t080164}@kansai-u.ac.jp

**あらまし** 特定の専門分野における専門知識をもとにして様々な事例にアドバイスや解決方法を提案する職業（例えば、弁護士や医師など）では、専門分野の知識体系からそれぞれの事例に関連する知識を論理的に導き出している。事例と専門知識を関連付ける思考過程は暗黙知とされることが多く、特定の分野での専門家になるためには、専門知識を網羅した学習を行ったうえで、様々な状況に応じた知識の適用例を見聞きすることで、適切に紐つける方法を習得する必要がある。本稿では、判例を対象として、事例（裁判に関わる事実）と専門知識（憲法条項）を自動的に紐付ける技術の有用性を検証する。提案手法では、憲法条文を半教師あり非負値行列因子分解のテンプレートとして、裁判に関わる事実を示す文書集合を因子分解することで、裁判事例に関連する憲法条項の関連率を分析した。判例ごとに判決文を参考に検証した結果、裁判事例に関連する憲法条項に適当な推定が行われていることを確認した。

**キーワード** 知識の構造化, データマイニング, 関係推定, 判例分析, 半教師あり非負値行列因子分解

## 1 はじめに

特定の専門分野における専門知識をもとにして様々な事例にアドバイスや解決方法を提案する職業（例えば、弁護士や医師など）では、専門分野の知識体系からそれぞれの事例に関連する知識を論理的に導き出している。事例と専門知識を関連付ける思考過程は暗黙知とされることが多く、特定の分野での専門家になるためには、専門知識を網羅した学習を行ったうえで、様々な状況に応じた知識の適用例を見聞きすることで、適切に紐つける方法を習得する必要がある。この問題を計算機科学によって解決する手段としては、複数の情報ソース間での潜在的な関係性を抽出して可視化する方法が考えられる。

専門知識の習得を必要とする知識の分野のひとつとして、法律関連の知識がある。自然言語処理分野でも、法律文書を扱った技術や分析結果に関する研究がいくつか報告されており、コンテスト型ワークショップ [1] が開催されるなど、注目を集めている。このワークショップでは、以下のタスクが設定され様々な研究チームが各タスクに対しての解決手法を提案し、その有用性を競っている。

**Task1:** 200 件の判例の中から、与えられた判決に対して必要とされた複数の判例を答える設定とした情報検索タスク。

**Task2:** 判決と必要な判例を既知として、判例内のどの段落が判決を含意するかを答える設定とした含意関係検索タスク。

**Task3:** 司法試験から抽出された問題文が与えられたときに、民法条文から（場合によって複数の）関連する条文を返す情報検索タスク。

**Task4:** 司法試験から抽出された問題文が与えられたときに、Yes/No の二値解答を行う質問応答タスク。

これらのタスクの他にも、異なる判例同士の類似度を計算的に

測る技術も研究されている。高野ら [2] は株主収集通知など形式的な特徴を持つテキストを対象に研究を行っており、法律文書も対象として判例の文末手がかり表現によって入力となる判例の重要文を抽出し判例文と判例文の類似度をはかる手法を提案している。關 [3] や野坂ら [4] は、判例文における文章間の類似性を捉えるために判例文の文章を隣接グラフ形式で表現し、グラフ間の類似度によって判例文書間の類似度を評価する手法を提案している。

本稿では、判例と判例を結びつける過程ではなく、事例の状況から数ある法律文書の中から関連のある専門知識に目星を付ける過程を対象とした分析を行う。裁判に関わる事実と事実に関わる専門知識である法律との紐づけの自動化を目標課題とし、複数の判例の事実に関わる文書集合に対して参照候補となる法律知識の関連度を示す手法を提案する。提案手法では、裁判に関わる事実と法律知識の関連性は法律知識を表現した特徴ベクトルが各事例を表現する文書ベクトルに因子として含まれる割合として表現できると考え、因子分解の技術を応用する。法律知識に関する知識の構造化における提案手法の有用性を考察する。

## 2 提案手法

本稿では、先行研究で提案した大学教育における講義シラバスと各研究室に所属する学生が執筆した卒業論文との関連性を分析した手法 [5] を本課題に応用する。先行研究では研究室と講義の知識の共通因子を顕在化させ両者の関係性を推定しているのに対して、本稿では判決書に記載されている裁判に関わる事実（以下、事案）と日本国憲法（以下、憲法。）を対象として因子分解を行うことで、事案と憲法の共通因子を顕在化を試みる。

因子分解の基本的な考え方は、下式の様に表現される。

観測変数集合 = 共通因子 × 独自因子。

ここで、異なる観測変数に共通して見られる特徴が共通因子、共通因子では説明できない各観測変数固有の要素が独自因子と、それぞれ定義される。提案手法では、因子分解技術の中でも、非負値行列因子分解（Non-negative Matrix Factorization：以下、NMF）[6] の拡張である半教師あり NMF（Semi Supervised NMF：以下、SSNMF）[7] を用いる。SSNMF では、観測変数集合に対して分析対象集合のテンプレートパターン（音源の周波数や、知識）を共通因子として与えることで、それぞれの分析対象がどの観測変数に活性化するかを独自因子として獲得できる。このとき、観測対象及び分析対象に集合を用いるため、集合に含まれる対象同士の相対的な関係性を踏まえて、各分析対象のそれぞれの観測変数における活性化を表現できる。

本稿の課題では、観測変数である事案に対して、分析対象となる憲法をテンプレートパターンとして与えることで、事案に対して相対的に関わりの深い憲法に記載された専門知識の顕在化を図る。一般的には暗黙知とされることの多い事案と憲法の潜在的な関係性を計算的に紐づけることで、専門家が有する暗黙知を習得する一助になると考えた。

## 2.1 NMF

NMF における因子分解は、式 (1) によって定義される。

$$y_{ij} \approx y'_{ij} = \sum_{k=1}^K h_{ik} u_{kj}, \quad (1)$$

ここで、 $y \in Y$ ,  $h \in H$ ,  $u \in U$  はそれぞれ観測変数行列、基底行列、活性化傾向行列の各要素を示す。また、 $i$  と  $j$  は行列内の要素のインデックスを示し、 $K$  は基底行列  $H$  の次元数となる基底数を示す。与えられた  $Y$  を、 $H$  と  $U$  の積である  $y'_{ij}$  によって近似することによって、観測行列  $Y$  を基底行列  $H$  と活性化傾向行列  $U$  に分解する。

NMF において、 $H$  と  $U$  の初期値はランダムな値を与えることで作成される。行列  $Y$  と行列  $HU$  の積である  $Y'$  の誤差を最小化することで、行列  $H$  と  $U$  をそれぞれ更新していく。このとき、行列  $H$  は観測変数行列  $Y$  をあらかじめ指定した基底数で表現可能な特徴パターンの集合となり、行列  $U$  は特徴パターンである  $h \in H$  の観測変数行列  $Y$  内での活性化傾向を表現している。誤差最小化が十分に行われたとき、2 種類の行列の掛け合わせによって観測変数行列を表現可能な特徴パターンの集合とその活性化傾向の集合がそれぞれ計算的に求められる。行列  $Y'$  と行列  $HU$  の誤差の最小化で用いられる誤差関数にはいくつかの種類が考えられるが、本稿では、下式で表される二乗 Euclid 距離  $D_{Euclid}$  を用いた。

$$D_{Euclid}(Y, HU) = \|Y - HU\|^2. \quad (2)$$

## 2.2 SSNMF

NMF において、基底行列と活性化傾向行列の初期値がラン

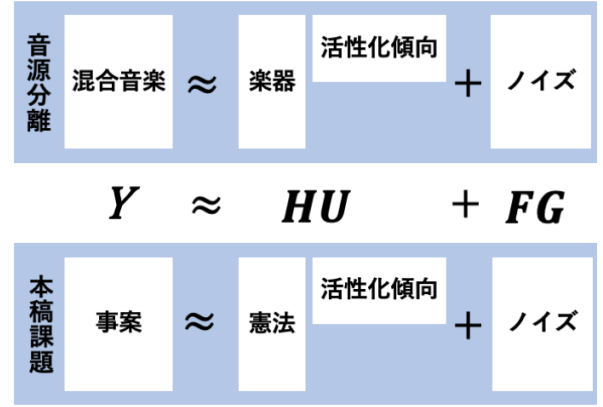


図 1 音源分離と本稿で扱う課題の SSNMF 適用の対応

ダムな値として作成され、両者の積を観測行列に近似するように基底行列と活性化傾向行列の両者の値をそれぞれ更新していくが、SSNMF では基底行列としてあらかじめ用意したテンプレートパターン行列を定め、活性化傾向行列の値のみ初期値をランダムで与え誤差最小化によって更新していく手法である。

SSNMF は、音源分離で利用される事例 [8], [9] が多く報告されている。これらの研究では、音源のスペクトル構造をテンプレートパターンとして与えることで各音源の活性化傾向を獲得し、音源からの自動採譜などに応用されている。SSNMF は下式で定義される。

$$Y \approx HU + FG, \quad (3)$$

ここで、 $H$  と  $U$  はテンプレートパターンベクトルとその活性化傾向行列、 $FG$  はノイズ項を示す。このとき、 $Y$ ,  $H$ ,  $U$ ,  $F$ ,  $G$  はそれぞれ、 $l \times n$ ,  $c \times n$ ,  $l \times c$ ,  $r \times l$ ,  $n \times r$  で定義されるサイズの行列を示す。 $l$  は事案数、 $c$  は憲法の章数、 $n$  は事案数および憲法の章数の次元数、 $r$  はノイズ項の次元数を示す。

## 2.3 SSNMF の知識関係顕在化課題への適用

SSNMF の本課題への適用を示す。本課題は、事案に対して相対的に関わりの深い憲法に記載された専門知識の顕在化である。SSNMF を適用し、事案集合に対して、専門知識のテンプレートパターンとして憲法集合を与えることで、それぞれの憲法がどの事案に活性化するかを数値行列で獲得する。このとき、事案及び憲法に集合を用いるため、それぞれに含まれる対象同士の相対的な関係性を踏まえて、憲法それぞれの各事案における活性化を表現できる。

SSNMF を用いた他の適用事例と比較して本稿での問題構造を説明する。図 1 に、音源分離と本課題の SSNMF の適用事例の対応を示す。SSNMF が音源分離で利用される場合、複数の音源が多重に混合された音声を示す周波数と時刻で構成される行列  $Y$  を、抽出したい楽器の周波数構造を示すテンプレートパターン  $H$  で分解する。結果として、多重に音源が混合された音楽における楽器の各時刻の活性化を示す行列  $U$  が抽出される。一方で、本稿は専門知識と対応する事象が混在した事案の中で、憲法で扱う知識がどの事案に関連するのかの顕在化を目的にしている。本稿の問題構造は、混在した集合の中から

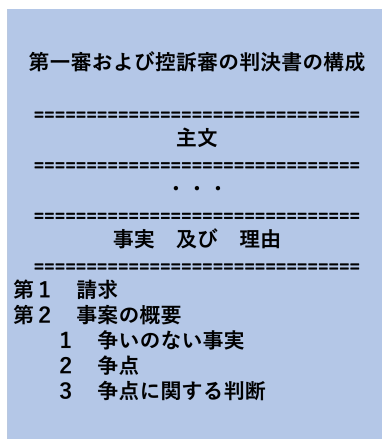


図 2 第一審および控訴審の判決書の構成

特定のテンプレートパターンがどこで活性化しているのかを求めている点において音源分離の問題構造と類似しており、同様に SSNMF が適用できると考えた。つまり、様々な専門知識と対応する事象が混在した事案を示す行列  $\mathbf{Y}$  を、憲法で扱う知識を示す行列  $\mathbf{H}$  で分解することで、事案と憲法で扱う知識の関係を示す活性化行列  $\mathbf{U}$  が得られると期待される。このとき、事案と憲法のリソースとして、非負値で表現可能な行列情報を扱う。この処理については、3 章で詳細に説明する。

### 3 提案手法による事案と専門知識の関係推定

提案手法を用いて、判決文書を対象として、各事案に関連する憲法の知識の関連率推定における有用性を検証した。以下、使用したデータと提案手法におけるパラメータ設定等について説明する。

### 3.1 データの準備

### 3.1.1 事案の情報

本稿では、法律分野に関わる専門家が憲法を学習する際に重要とされる判例を推定対象とした<sup>1</sup>。判例とは、裁判において実際に起こった事件に対する法律的判断が記録されたものである。判例ではどのような事件において、何を根拠に、何の判例が使用され、どんな判決が下されたのかといったデータが記載されている[3]。主題別判例一覧から取得できた客観的事実が記載されている第一審判決の判例 23 件を使用する。

裁判において、第一審や控訴審は事実認定の判断を行うのに対して、上告審では原則として事実認定を行わない。そのため、第一審・控訴審と上告審とでは、判決書の書き方が異なる。控訴審は、第一審を基礎として行われることから、判決書は第一審の事実認定を引用する形式をとる。本稿では、判例文書の中から、法的解釈等をふくまない客観的事実のみの使用を目的としているため、第一審のみを対象とした。判決書は、「新様式」（最高裁判所事務総局編「民事判決書の新しい様式について」〔平成2年法曹会〕）と呼ばれる記載方式で書かれたもののみを対象とした。

### a) 判決書（新様式）の構成

第一審および控訴審の判決書の構成は図 2 のとおりである。以下、各項目について述べる<sup>2</sup>。

*i)* 主 文

第一審判決の主文は、原告の請求に理由がある場合は、原告の請求内容に相応した裁判所の結論が記載される。原告の請求に理由がない場合には請求は棄却され、本案判決（請求認容または請求棄却の判決）をするために必要な訴訟要件（訴えの適法性）が欠けている場合は、却下の判決が記される。

ii) 事実及び理由

第一審の事実及び理由には、主に請求内容と事案の概要が記される。請求内容には原告の主張となる部分が示され、事案の概要には事件がどのような種類の争いであるかを簡潔に表示し、かつ、本件での争点を示している。「争点」は原告および被告（控訴人および被控訴人）が主張したことをまとめたものである。裁判所が争点についてどのような判断をしたかは、「第3争点に対する判断」に記載されている。

## b) 本稿における事案抽出

本稿では、事案として、事件の状況を客観的に説明している箇所を抽出した。判決書の構成に着目し、主張者によって解釈の違いが起こらない状況説明のみを客観的に説明している箇所とすることとし、対象とした判例データに、人手で下記の項目ごとにタグを付与した。

- Tag0  
項目箇所が客観的事実に基づいている。
- Tag1  
項目内容が記載されている。
- Tag2  
項目内容が記載されていない。

タグ付与の結果は表 1 に示す。この結果より、「事件の概要」、「争いのない事実」もしくは「前提のなる事実」に、解釈の違いが起これないと判断できる客観的状況説明が記載されていたことから、この 2 項目から事案の該当箇所を抜き出した。

### 3.1.2 専門知識の情報

専門知識のリソースとして、憲法を用いた。憲法とは、国の最高法規であり、国のあり方を示すものである。すべての法律は憲法の方針に添い作成されていることから、民法や刑法といった特定の事案のみに関わりがあるものと異なり、憲法内のいずれかの条文は、必ず既存の判例および法律と包含関係にあるため、法律文書を扱う上で避けては通れないものである。

本項では憲法の条項を、章ごとにまとめ全 11 章抽出した。憲

1 : <https://www.cc.kyoto-su.ac.jp/~suga/hajimeni.html>

2 : <https://www.rosei.jp/lawdb/static.php?p=judge003>

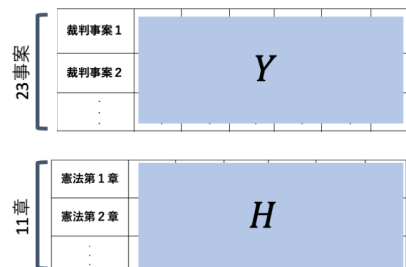
Tag0: 項目箇所が事実について言及している。  
Tag1: 項目が記載されている。  
Tag2: 項目が記載されていない

### 3.1.3 テキストの正規化

### 3.2 データの数値表現化と分散表現化

得られたベクトル表現には、全文書で出現する単語すべてを扱っているため、文書集合中の1文書に1度しか出現しない単語も含まれており、これらの単語は文書の特徴表現としては不適切である可能性が高い。そこで、単語の重み付けのために、全文書での合計出現頻度が2以上の単語のみを採用し、事案と憲法を857単語の出現頻度傾向によってベクトル表現した。

これらの処理を経て得られた (23 事案 + 11 章) × 857 単語



め、2 種類の異なる文書集合に共通したモデルでの分散表現が得られる。

3.3 SSNMF の適用

3.2 節で得られた分散表現に対して、SSNMF を適用することで、事案と憲法の潜在的な関係性を顕在化する。図 4 に示すように、事案と憲法の集合を示す 2 種類の分散表現の集合へ分割した。この 2 種類の分散表現の集合は、どちらも 1) 非負値、2) 行列、3) 各集合の知識内容を表現可能という SSNMF を適用するための 3 つの条件を満たしている。

SSNMF の適用では、近似目標となる行列  $Y$  として事案についての分散表現の集合を、因子分解を実施するうえでのテンプレートパターンである行列  $H$  として憲法についての分散表現の集合をそれぞれ扱う。ここで、SSNMF の近似におけるイテレーション数は、山本らの先行研究と同じく 80,000 回とした [5]。

3.4 評価方法

事案と憲法の関係性の顕在化において、提案手法の有用性を定量的に評価する。評価には、事案に関連する憲法の章の正答と Mean Reciprocal Rank (以下、MRR) を用いる。

本稿では事案に関わる憲法の章の正解ラベルは、判決文に記載されている条文の出現頻度をもとに定める。各事案についての判決文において、最も出現頻度が高い憲法の章を事案と関連深い憲法が記載されている章とした。SSNMF を適用した因子分解で得られる活性化傾向行列において、各事案に対して憲法の章が相対的にどの程度活性化しているのかによって、事案への憲法の章の関連度の順位付けを行う。MRR の評価では、事案ごとに正解ラベルの憲法の章が初めて現れる順位の逆数 (Reciprocal Rank) を計算し、事案の RR を平均した値を算出した。RR の値が 1 に近いほど、全体的に事案に関わりの深い憲法の章を高い順位で予測できていることを示す指標となる。本稿では分析対象とした全事案について算出した MRR、同一の正解ラベルが付与された章ごとの事案に対する MRR の 2 つの指標をそれぞれ算出した。

4 結果と考察

表 2 に、正答数を示した結果を示す。多くの事案において、

表 2 提案手法における正当数	
正解ラベルの章が初めて現れる順位	事案数
1 位	6
2 位	7
3 位	3
4 位	3
5 位	0
6 位	0
7 位	1
8 位	1
9 位	0
10 位	0
11 位	2

表 3 同一正解ラベルが付与された章ごと MRR による評価

正解ラベル	MRR
第 3 章国民の権利	0.46
第 4 章国会	1.00
第 5 章内閣	1.00
第 10 章最高法規	0.50

推定上位に正解ラベルの章が出現しているものの、いくつかの事案においては良好とは言えない結果となった。提案手法では対象データ 23 件の事案中、16 件 (69%) が 3 位以内に正解ラベルを推定しており、概ね良好な結果であるといえる。本実験での問題は 11 章の中から 1 つの正解を推定するため、チャンスレベルは 9 % となる。このチャンスレベルを考慮すると、事案に関連する憲法の章の推定において、提案手法の有用性が示唆された。ただし、他の手法との比較については今後の課題となる。

4.1 良好な結果に関する考察

表 5 は、「幸福追求権」に関係の深いとされる「エホバの証人輸血拒否事件」の判決文書の結果である。「幸福追求権」は憲法第 3 章第 13 条に明記されており、本稿の正解ラベルも「第 3 章国民の権利及び義務」に付与されている。活性化傾向値は、正解ラベルが付与されている「第 3 章国民の権利及び義務」が極端に高い値を示している。

表 6 「参政権、選挙権」に関係の深いとされる「議員定数訴訟衆議院小選挙区違憲状態判決 (平成 23 年)」の判決文書の結果である。「参政権」及び「選挙権」は憲法第 3 章第 15 条に主に明記されているが、本稿の正解ラベルは第 4 章の国会に付与されている。「議員定数訴訟衆議院小選挙区違憲状態判決 (平成 23 年)」は一票の較差に関する訴訟である。一票の較差とは、形式的には一人一票の選挙権が保障されているけれども、有権者と選出される議員の数の割合を選挙区同士で比較すると差が生じていることを示す表現である。「議員定数訴訟衆議院小選挙区違憲状態判決 (平成 23 年)」では一票の較差により投票価値に実質的な差が生じていることを理由に選挙を無効とすべきかどうかを判断している。活性化傾向は、正解ラベルが付与された「第 4 章の国会」に加え、「第 11 章補則」「第 6 章司法」の値が極端に高く、「第 9 章改正」「第 8 章地方自治」「第 3 章国民の権利及び義務」に関しても相対的に見ると高い値となっている。これは、本訴票が選挙の問題を取り扱っていることに所以する結果であると考察できる。「第 11 章補則」には参議院成立前の国会についてや、参議院議員の任期について明記されている。「第 6 章司法」には、日本の司法について規定されており、最高裁判所についてや下級裁判所についても述べられている。裁判において、似た事案がどのような判決を示したかは重要な事案であり、今回の判決文にも多く登場している。特に本件の一票の較差についての問題は、衆議院選挙に限らず、参議院選挙や地方選挙においての最高裁判所や下級裁判所の判決が関わってくる事案である。そのため、活性化傾向も広く多くの項目で高い値を示したと考えられる。

表 7 「思想・良心の自由」に関係が深いとされる『『君が代』



表 4 提案手法によって良好な結果を示した事案. 活性化傾向行列において, 各事案に対して憲法の章が相対的にどの程度活性化しているのかを関連度順位でソートして示す. 表中の赤字は正解ラベルが付与された章を示す.

正解ラベルの章	活性化傾向値
<b>第 3 章国民の権利及び義務</b>	0.097919
第 1 章天皇	1.627e-12
第 11 章補則	5.768e-44
第 9 章改正	2.827e-75
第 8 章地方自治	1.40e-100
第 5 章内閣	1.43e-117
第 4 章国会	3.08e-169
第 7 章財政	7.27e-177
第 10 章最高法規	7.42e-184
第 6 章司法	5.89e-229
第 2 章戦争の放棄	0.0

正解ラベルの章	活性化傾向値
<b>第 4 章国会</b>	0.999883
第 11 章補則	0.978879
第 6 章司法	0.823060
第 9 章改正	0.514399
第 8 章地方自治	0.198834
第 3 章国民の権利及び義務	0.019298
第 10 章最高法規	5.109e-13
第 7 章財政	3.475e-13
第 5 章内閣	1.760e-14
第 1 章天皇	1.767e-74
第 2 章戦争の放棄	6.007e-78

正解ラベルの章	活性化傾向値
<b>第 3 章国民の権利及び義務</b>	0.481428
第 10 章最高法規	0.349606
第 1 章天皇	0.000909
第 7 章財政	1.897e-12
第 9 章改正	6.633e-23
第 8 章地方自治	4.419e-84
第 5 章内閣	1.539e-98
第 6 章司法	1.52e-114
第 11 章補則	4.21e-117
第 4 章国会	1.89e-277
第 2 章戦争の放棄	0.0

表 8 提案手法によって良好ではない結果を示した事案. 活性化傾向行列において, 各事案に対して憲法の章が相対的にどの程度活性化しているのかを関連度順位でソートして示す. 表中の赤字は正解ラベルが付与された章を示す.

正解ラベルの章	活性化傾向値
第 9 章改正	0.306758
第 5 章内閣	0.189350
第 11 章補則	0.017433
第 2 章戦争の放棄	0.005006
第 4 章国会	0.000116
第 8 章地方自治	3.236e-09
第 7 章財政	1.446e-19
第 1 章天皇	1.728e-22
第 10 章最高法規	4.460e-25
第 6 章司法	1.413e-41
<b>第 3 章国民の権利及び義務</b>	2.179e-47

起立斉唱拒否事件（東京都立高校再雇用拒否）」の結果である。「思想・良心の自由」は第 3 章第 19 条に明記されており, 本稿の正解ラベルも「第 3 章国民の権利及び義務」に付与されている. 活性化傾向は, 正解ラベルが付与されている「第 3 章国民の権利及び義務」が最も高い値を示している. また, 「第 3 章国民の権利及び義務」の次に高い値となる「第 10 章最高法規」は, 判決文でも触れられており事案と関係の深い憲法といえる.

#### 4.2 良好とはいえない結果に関する考察

全事案を対象とした MRR は, 0.505 と低い値を示した. これは, 極端に良好ではない結果を示す事案が存在したことが原因であると考えられる. 表 8 は, 「法の下での平等」に関係深い「再婚禁止期間違憲判決」の判決文書の結果である. 正解ラベルである「法の下での平等」について記載されている「国民の権利及び義務」の活性化傾向値は低く, 「改正」の章の値が高くなっている. この判例の争点としては, 『民法 733 条 1 項の再婚禁止期間の規定のため婚姻が遅れ, これによって精神的損害を被ったと主張しているが, 立法不作為が国家賠償法 1 条 1 項の規定の適用上違法となるか』が挙げられており, 民法 733 条を改定するか否かについても述べられている. しかし, この裁判の本質は, 法の下での平等（憲法 14 条: すべて国民は, 法の下に平等であつて, 人種, 信条, 性別, 社会的身分又は門地により,

政治的, 経済的又は社会的関係において, 差別されない. 憲法 24 条配偶者の選択, 財産権, 相続, 住居の選定, 離婚並びに婚姻及び家族に関するその他の事項に関しては, 法律は, 個人の尊厳と両性の本質的平等に立脚して, 制定されなければならない.) の観点から, 民法 733 条は違憲であるかどうかで争うことであり, 法の改定についての判決ではない. 本来は, 法の下での平等のついて争そわれた事案であるが, 法改正についても詳細に述べられていたため, 提案手法の適用では「改正」の章が高くなり, 良好な結果は得られなかった. 表 3 に示した同一の正解ラベルが付与された章ごとの事案に対する MRR の結果からも, 憲法の章ごとに MRR の値の分散が大きいことが読み取れる. これらの問題は, BoW 法によって文書表現しており, 単語の意味的な重み付けが行われていないことによるものであると考える. この問題の解決には, 単語の文脈を考慮した word2vec [12] で得られた分散表現を用いるアプローチが考えられる. 提案手法で用いた SSNMF を適用するためには分散表現のベクトル値の非負化が必要となるため, 非負値で単語の文脈を表現可能な分散表現の獲得が課題となる.

以上より, 極端に不適当な結果を示した事案が存在したことで MMR による全体的な傾向は良好とは言えないものの, 正答数をもとにした評価では概ね良好な結果が得られたと考察する. 不適当な結果を示した事案についても問題の原因は明らかであり, 解決方法についても今後の展望が得られた.

## 5 おわりに

本稿では, 判決文書に記載されている客観的事実と国の最高法規である憲法の関係を関連付ける手法を提案した. 提案手法では, SSNMF を適用することで, 事案と憲法という 2 種類の異なる文書間の関連性を顕在化した. 定量評価による考察の結果, 提案手法によって概括的に事案と憲法を結びつけられる有用性が示唆された.

今後は, 事案と憲法の数値表現に単語の局所的なコンテキスト情報を考慮可能な分散表現を適用する. また, キャリア形成に関わる大学の学部情報と企業情報の関係推定などにも応用し, 提案手法の適用範囲を検討する.

## 謝 辞

この研究は 2021 年度国立情報学研究所公募型共同研究 (21S0501) の助成を受けています。

## 文 献

- [1] 狩野芳伸, 吉岡真治, ラベロジュリアーノ, キムミユン, ゴーベルランディ, 佐藤健. Colicee: 法律文書の情報抽出および含意関係認識を行う国際コンペティション. 人工知能, Vol. 35, No. 3, pp. 377–384, 2020.
- [2] 高野海斗, 酒井浩之, 中川慧. 学習データの自動生成による深層学習を用いた株主招集通知の重要ページ抽出. 人工知能学会論文誌, Vol. 36, No. 1, pp. WI2–G\_1–19, 2021.
- [3] 涼介關, 海斗高野, 浩之酒井, 良三北島. 判例テキストデータを用いた類似判例の自動抽出. Technical Report 3, 成践大学, 成践大学, 成践大学, 成践大学, sep 2019.
- [4] 野坂卓矢, 原口誠. 文章の隣接グラフ化とグラフマッチングに基づく判例文の類似度計算. 人工知能学会全国大会論文集, Vol. JSAI2011, pp. 3H2OS38–3H2OS38, 2011.
- [5] 山本京佳, 山西良典, 松下光範. 半教師あり NMF を用いた専門分野と講義の関係推定. 人工知能学会全国大会論文集, Vol. JSAI2021, pp. 1I2GS4a01–1I2GS4a01, 2021.
- [6] Daniel D. Lee and Hyunjune Sebastian Seung. Algorithms for non-negative matrix factorization. In *Proc. Advances In Neural Information Processing Systems*, pp. 556–562, 2001.
- [7] Hyekyoung Lee, Jiho Yoo, and Seungjin Choi. Semi-supervised nonnegative matrix factorization. *IEEE Signal Processing Letters*, Vol. 17, No. 1, pp. 4–7, 2010.
- [8] 林亜紀, 亀岡弘和, 松林達史, 澤田宏. 非負値周期成分分析手法による音楽音響信号の音源分離. 日本音響学会 2016 年春季研究発表会講演論文集, pp. 639–642, 2016.
- [9] 北村大地, 猿渡洋, 鹿野清宏, 近藤多伸, 高橋祐. 基底変形型教師あり NMF による実楽器信号分離. 電子情報通信学会技術研究報告, Vol. 112, No. 388, pp. 13–18, 2013.
- [10] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *Proc. 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 230–237, 2004.
- [11] 大島裕明, 中村聡史, 田中克己. Slothlib web サーチ研究のためのプログラミングライブラリ. *DBSJ letters*, Vol. 6, No. 1, pp. 113–116, 2007.
- [12] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Proc. of the 26th International Conference on Neural Information Processing Systems*, p. 3111–3119, 2013.