

地方自治体の例規比較に用いる条文対応表の自動生成

竹中 要一

大阪大学大学院情報科学研究科
takenaka at ist.osaka-u.ac.jp

若尾 岳志

獨協大学法学部
wakao2 at dokkyo.ac.jp

概要 地方自治体が制定する法を例規という。二つの自治体に存在する例規の違いを明確にするための例規比較は、法学教育や法学研究、地方自治体法務、企業法務において頻繁に行われている。本研究は、現在手作業で実施されている例規比較を計算機を用いて自動化する事を目的とする。この目的を達成するため、最長共通部分文字列、ベクトル空間モデル、及び文字列アライメントを基本とする96種類の手法を比較検討した。その結果、例規の漢字のみを抽出した文字列に対するアライメントが、法学者の結果と最も近かったことを示す。

1 はじめに

法は国の制定する法律と規則、地方自治体が制定する条例と規則の二つに分類される。前者は法規、後者は例規と総称される。法規を制定する日本国は一つだが、例規を制定する地方自治体は多数存在するため、同一項目を規程する多数の例規が地方自治体ごとに存在することになる。具体的には、青少年の保護育成を目的とする条例や、各県の象徴である「県章」を定めた条例等があげられる。これら同一項目に関する条例は相互に類似しているものの、地方自治体の置かれた状況が異なるため、随所に相違点が存在している。一例として、青少年の保護育成を目的とした条例では、青少年の夜間外出を制限しているが、その制限される時間が異なっている事が挙げられる。愛媛県では午後1時から午前4時を夜間と定義している一方、高知県では午後10時から午前4時を夜間としている。このような違いを明らかにするため例規比較が行われる。

例規比較は、地方自治体間の違いを明らかにする教育・研究活動以外にも、企業法務や自治体法務においても発生する作業となっている。自治体法務における例としては、例規の制定・改訂する際の参考資料作成、さらには自治体合併時に全例規を擦り合わせる行程が挙げられる。特に自治体合併時には、対象となる全自治体の全例規に対する例規比較を短時間に行う必要がある仕事量の多い法務となっている

[7, 8, 9]。現在、この例規比較は専門家が手作業で実施しているため、計算機を利用した作業の省力化が望まれている。そこで本研究では、与えられた2つの例規比較を計算機で実施する有効な手法について検討を行う。

法を計算機で扱う研究は、平成5年度から9年度の文部省化学研究費重点領域研究「法律エキスパートシステムの開発研究」において促進され [5]、現在では多くの自治体が例規をインターネット上に公開する所まできている。しかし例規を対象とした研究は少なく、ある自治体の例規を分類する研究 [6] が存在するに留まっている。

2 問題定義

2.1 例規の構造

本研究で対象とする例規とは、法特有の構造を有する文章である。典型的な構造としては、例規名を表す表題、効力を発する日を記した発令、公布を宣言する公布分、例規の内容に当たる本則、そして制定及び改正附則が挙げられる。また、本則は章条項号といった階層を有する文章の集合となっている。図1に例規の共通構造を示す。ただし、実際の例規構造には例外が多く存在し、特に古い例規を扱う際には注意が必要な事を付記しておく。

2.2 条文対応表

同一項目規程する例規の比較は、例規構造のうちの条を単位とした対応表を作成することが基本にな

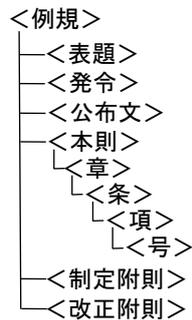


図1 例規の構造（主要な共通部分を抜粋）

る。例えば愛媛県の条例の2条と香川県の条例の4条が対応する、といった関係を一覧表示したものになる。この対応表をもとに夜間の定義等の細かい文章比較を行う事になる。図2に対応表の例を挙げる。本研究では与えられた2条例より条文比較表を自動生成する事を目的とする。



図2 例規の対応表の例（一部抜粋）

2.3 条文対応表生成問題

条の対応関係は多対多であり、対応する条がない場合も存在する。そのため条文対応表は、2つの各例規の条を独立頂点集合とした二部グラフで表現可能である。ここで条文対応表生成問題を次のように定義する。

入力 2つの例規 A, B

出力 二部グラフ $G = (V_A, V_B, E)$. ただし A の条

がノード V_A に、 B の条が V_B に対応する。

目的関数 不明

目的関数を「不明」としているのは、現時点では法学者の要求を数式化することが不能であるためである。そのため本研究は、3種類の文字列比較手法に対してパラメータ変更する事で、現時点で不明である目的関数に沿う手法を発見することを試みる。

3 文字列比較手法

条文比較表を自動生成するためには、条文を比較する手法が必要となる。条文は文章、すなわち文字列で構成されているため、以下の3種類の手法を用いた。

- 最長共通部分列 [2]
- ベクトル空間モデル [3]
- 文字列アライメント [4]

最長共通部分文字列とは、入力として与えられた2つの文字列における最長の共通部分文字列をいう。共通部分文字列とは、もとの文字列から文字を出現順序をかえずに取り出したものとなる。ベクトル空間モデルでは、文章を各文字や単語の出現頻度を表したベクトルで文章を表現する。二つの文章に対応するベクトル間の距離を計算することによって文章間の関連度を求める方法である。

文字列アライメントは、生物情報学において遺伝子の類似性を見る際に用いられる方法である。最長共通部分文字列と同様に、動的計画法により計算される。その違いは、共通部分文字列の間に存在する非共通部分にペナルティが課される事が挙げられる。一般に最長共通部分文字列や文字列アライメントでは、共通部分文字列の順序は保存される。しかし例規比較において対応させるべき条文は、必ずしも順序が保存されているとはいえない。そこで、本研究では順序が保存されていない部分文字列を考慮可能なアライメント手法も比較対象とした。

4 条文対応表生成法

上記のアルゴリズムを用いた条文対応表の生成法は以下の通りである。入力として2つの例規 A,B が与えられたとする。A の各条に対して、

- 1 B の全条との間に類似度を計算する。
- 2 類似度の最も高かった条間に辺を引く。

5 評価実験

5.1 実験条件と評価項目

青少年保護に関する愛媛県と香川県の条例を対象とし、全96手法の性能評価を行った。

類似スコアを計算する対象文は条文とし、最長共通部分文字列のみ条題も対象とした。ベクトル空間モデルにおける距離にはコサインを用いた。最大共通部分文字列及びアライメントにおいて一致した場合のスコアとして定数及び、Mecab[1]により解析した品詞の tf-idf 値を用いた。

手法で得られるスコアは条文の文字数に依存する。そこで得られたスコアを短い方の条文の文字数で割った値を相対スコアとして類似尺度に用いた。ベクトル空間モデルにおけるベクトル長は、出現頻度順に何位の品詞までをベクトルの要素として利用したかを表している。

5.2 結果と考察

表 1,2,3 に最長共通部分文字列、アライメント、ベクトル空間モデルに基づく手法の正解率を示す。表中の「五詞」は名詞、副詞、形容詞、動詞、連体詞を計算対象として用いた事を表している。

これらの表より文字列アライメントに基づく手法とベクトル空間モデルに基づく手法が優位であることがわかる。また、相対スコアよりも絶対スコアの、tf-idf スコアよりも定数スコアの正解率が高く、定数スコアの場合、漢字のみを用いた方がよい結果が得られている。ベクトル空間モデルの場合、次元が大きい方が正解率が高く、名詞のみの利用が他品詞を利用した場合よりも若干よい事がみとれる。

実務では条文比較表を作成した後、対応する条文において異なる部分を出す必要が生じる。その

際、ベクトル空間モデルでは、対応する条文を決定するだけで条文内の異なる部分を見つける事の役には立たない。一方、文字列アライメントに基づく手法では、異なる場所が既に求められているため、例規比較としてより有効であると考えられる。

6 まとめ

地方自治体の法である例規を比較するための条文対応表の自動生成問題を定義し、目的に沿う手法を探索した。青少年保護に関する条例を対象として行った実験により、文字列アライメントに基づき、1) 比較対象を漢字に限定し、2) 部分文字列の順序を保存し、3) 条文の長さに依存しない絶対スコアを用いた手法が最も有効であることを示した。本論文に記した手法は、全ての条に対応する条が存在することを前提にしている。そのため今後は対応関係が存在しない場合を認めるアルゴリズムを考案する必要がある。それ以外にも手法の組合せ等による正解率向上、結果の尤もらしさの提示、法学者が利用可能なユーザインターフェースの構築などの課題が挙げられる。

7 謝辞

本研究の一部は科研費(21500253)によった。

参考文献

- [1] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 230–237, 2004.
- [2] David Maier. The complexity of some problems on subsequences and supersequences. *J. ACM*, Vol. 25, pp. 322–336, April 1978.
- [3] Gerard Salton, Anita Wong, and Chung-Shu Yang. A vector space model for automatic indexing. *Communications of the ACM*, Vol. 18, No. 11, pp. 613–620, 1975.
- [4] T. Smith and M. Waterman. Identification

表1 最長共通部分文字列の正解率

定数スコア			
	対象	絶対スコア	相対スコア
条題	全語	51%	55%
	漢字	51%	50%
条文	全語	54%	46%
	漢字	70%	58%

tf-idf スコア			
	品詞	絶対スコア	相対スコア
条文	全品詞	58%	42%
	原形	46%	41%
	名詞	50%	40%
	五詞	48%	40%

表2 アライメントの正解率

定数スコア			
順序	対象	絶対スコア	相対スコア
保存	全語	75%	60%
	漢字	80%	58%
非保存	全語	76%	58%
	漢字	78%	58%

tf-idf スコア			
順序	品詞	絶対スコア	相対スコア
非保存	全品詞	56%	49%
	原形	58%	49%
	名詞	52%	34%
	五詞	45%	34%

of common molecular subsequences. *Journal of Molecular Biology*, Vol. 147, pp. 195–197, 1981.

- [5] 吉野一（編）. 法律人工知能 — 法的知識の解明と法的推論の実現. 創成社, 2000.
- [6] 原田隆史, 青木淳一, 真島由里香. クラスタリング手法に基づく条例の自動分類. 情報ネットワーク法学会第9回研究大会予稿集, pp. 65–68, 2009.
- [7] 加藤幸嗣. 比較分析 市町村合併と条例制定—福知山市の公の施設条例等を題材として (自治体情報 条例 制定の動向). 法令解説資料総覧, No. 292, pp. 76–78, 2006-05.
- [8] 伊佐美浩一. 市町村合併調整のポイント (1) 合併に関する法的問題 (1) 条例・規則の調整西

表3 ベクトル空間モデルの正解率

定数スコア			
ベクトル長	対象	絶対スコア	相対スコア
10	全品詞	24%	5%
	原形	25%	5%
	名詞	26%	5%
	五詞	26%	5%
50	全品詞	56%	14%
	原形	58%	14%
	名詞	57%	16%
	五詞	55%	16%
100	全品詞	61%	16%
	原形	62%	16%
	名詞	72%	21%
	五詞	70%	22%
最大長	全品詞	75%	13%
	原形	70%	10%
	名詞	75%	31%
	五詞	76%	19%

tf-idf スコア			
ベクトル長	対象	絶対スコア	相対スコア
10	全品詞	22%	5%
	原形	24%	5%
	名詞	21%	4%
	五詞	21%	4%
50	全品詞	53%	18%
	原形	60%	17%
	名詞	53%	16%
	五詞	53%	16%
100	全品詞	57%	16%
	原形	64%	17%
	名詞	57%	18%
	五詞	59%	17%
最大長	全品詞	61%	18%
	原形	68%	19%
	名詞	57%	19%
	五詞	60%	14%

東京市. 自治体法務研究, No. 1, pp. 108–114, 2005-05.

- [9] 藤井真知子. 市町村合併における自治体法務の現状と課題：甲賀市の条例整備を手がかりとして. 龍谷大学大学院法学研究, Vol. 9, pp. 181–214, 2007-07-31.