

都道府県に共通する例規の抽出と応用

竹中 要一^{1,a)} 若尾 岳志^{2,b)}

概要：

日本の地方自治体が各議会において国の法律の範囲内で制定する法を条例といい、憲法を頂点とする国内法体系の一部を担っている。この条例に地方自治体が定める規則を合わせて例規と呼ぶ。地方自治体はそれぞれが独立した議会有しているものの、同一の国に属している事及び地域的、環境的な類似性により、同一事項について規定している例規が多数の地方自治体に存在している。以降、同一事項について規定している例規を共通例規と呼ぶ。地方自治体において、近隣自治体の例規を比較する法務がしばしば発生するが、共通例規の探索がその端緒となる。本研究では計算機による共通例規の自動抽出法を提案し、47都道府県の全例規を対象とした性能評価を行った。提案手法の出力結果が共通例規であるか否かの判定を手で行った結果、受信者操作性曲線下面積 0.929 と良好な結果である事を明らかにした。また、抽出された共通例規の例規解析への応用例を提示する事により、本研究が法学の発展にも貢献しうる事を示した。

キーワード： 最長共通部分列, 地方自治体, 例規, 法情報学

Automatic Extraction of Common Local Government Laws among Prefectures and its Applications

TAKENAKA YOICHI^{1,a)} WAKAO TAKESHI^{2,b)}

Abstract: Local governments establish their ordinances in each parliament with in the laws of the country. Ordinances are responsible for the part of the domestic legal system whose summit is the Constitution. The ordinances and regulations that local government enacts are collectively referred to as statutes. Although local governments establish each statute independently, many of them have some similarities with other local governments' statutes because of the regional and environmental similarity. Finding and comparing similar statutes from neighbor local governments is one of typical legal affairs. But it is not automatically yet. We propose an automatic extraction method of similar statutes and apply it to all statutes of 47 prefectures. Human annotations were used to evaluate our method and it shows that the performance were well. Using the results of our method, we report example applications of statute analyses.

Keywords: Longest Common Subsequence, Local Government, Statutes, Legal Informatics

1. はじめに

制定法は基本的に、章節及び条項号という階層を有する構造化された文章である。日本の国会で制定される法を法

律といい、日本の地方自治体が各議会において国の法律の範囲内で制定する法を条例という。法律と条例等を含めて制定法と呼び、憲法を頂点とする国内法体系を形作っている。法律に国の行政機関が制定した法規範である命令を合わせたものを法令、条例に地方自治体が制定した法規範である規則を合わせたものを例規と呼ぶ。日本の立法府は一つしかないため、同一事項、例えばある種の犯罪について規定した法令は一つしか存在しない。そのため、ある事項についての法務調査としては、関連法規の検索と時系列に

¹ 大阪大学
Osaka University, Suita, Osaka 565-0871, Japan

² 獨協大学
Dokkyo University, Souka, Saitama 340-0042, Japan

a) takenaka@ist.osaka-u.ac.jp

b) wakao2@dokkyo.ac.jp

沿った変遷や、他国との比較が主たるものとして挙げられる。一方、同一事項について規定している例規は、都道府県や市町村といった同じ階層にある各自治体ごとに存在することが想定できる。例えば、各県の象徴であり旗に用いられる県章を定めた条例は全都道府県で制定されており、青少年の保護育成を目的とする条例は長野県を除く46都道府県で制定されている。

これら同一事項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、青少年の深夜外出を制限していることが多いが、制限規定の有無やその制限される時間が異なっている事が挙げられる。具体的には、東京都や愛媛県では午後11時から午前4時を深夜と定義している一方、高知県では午後10時から午前4時を深夜としている。また、大阪府では外出を制限する時間帯を年齢によって変えており、16歳未満の場合は午後8時から午前4時まで外出を制限される。このように例規比較をすることにより、相違点が明確化される。例規比較は、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業務である。自治体法務における例としては、例規を制定・改正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せて一つに纏めるための準備作業が挙げられる。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日に行う必要がある仕事量の多い法務となっている [1], [2], [3], [4]。この法務、特に合併時の例規比較については弁護士を核とする法律の専門家が手作業で行っておりコストの高い作業となっており、計算機支援に基づく省力化が求められている。理由の一つとしては、比較対象となる地方自治体数、そして例規数が多い事が挙げられる。例規比較を行うためには、比較対象となる例規を各自治体の例規集から探し出す必要がある。しかしながら、都道府県数は47、市町村数に至っては1700以上もあるため、手作業では近隣自治体との比較が限界であり、網羅的な例規比較は容易でない。また、例規数も都道府県を例に挙げると、最少が山口県の974件、最多が北海道の4156件、全都道府県の例規数の合計は74,880件(2012年9月現在)もあり、計算機支援は不可欠となっている。

この流れを受け、近年計算機を利用した作業の省力化法についての報告がなされている。竹中らによる二つの例規の条文対応表の自動作成 [5], [6] や、類似手法に基づく角田による例規分類 [7] がこれに該当する。しかしながら前者の竹中らの研究では、比較対象となる二つの例規の抽出法について述べられていない。また後者の角田の研究では例規構造に内在する共通パターン抽出を目的としているため、分類群が例規比較の対象となりうる例規の集合か否かの評価はなされていない。そして分類群の作成にスーパー

コンピュータを用いる必要がある程計算量の多い手法であり、且つ研究分野が法学であるためか研究成果を査読論文として出版していないため再現が困難である。そこで本研究では、一つの例規を入力として与えた時、対象となる階層の全地方自治体から類似例規を網羅的に抽出するための計算量の少ない手法を提案する。提案手法の出力結果に基づく例規解析例を示す事により、提案手法の有効性及び実用性を明らかにする。

本節の最後に、法を計算機で扱う研究についての概況を述べる。法を計算機で扱う研究は法情報学 (Legal Informatics) と呼ばれ、法律の専門家を模倣するエキスパートシステムに関する研究として、人工知能研究の派生領域として発達してきた。本分野初期の国際会議として、1987年より隔年開催されている International Conference on Artificial Intelligence and Law (ICAIL) [8] と、1988年より毎年開催されている International Conference on Legal Knowledge and Information System (JURIX) [9] がある。日本では平成5年度から9年度にかけて文部省科学研究費補助金の重点領域研究「法律エキスパートシステムの開発研究」として採択された事により、特に判例を計算機で利用する知的システムに関する多数の研究が実施された [10]。また研究期間中にインターネット上の法律情報の整備が進み、各種法律情報の閲覧が容易になった事が以降の研究を促進させた。これが、2007年より日本にて毎年開催されている法情報学の国際ワークショップ International Workshop on Juris-informatics (JURISIN) [11] の開催へと繋がっていった。法律以外の法関係文書に対する法情報学研究としては、特許における公開特許公報中の請求項と発明の詳細な説明文との対応付け [13] や、法律用語の語彙体系オントロジー構築の研究 [14], [15] がある。また例規を対象とした法務の計算機支援に関する研究として他に例規種別の自動分類 [16] や例規改正の編集過程に基づいて改正前後の差異を表現した新旧対照表の自動作成 [17] がある。

2. 類似例規

法令や例規は法特有の階層構造を有する文章である。本研究で対象としている例規について、典型的な階層構造を以下に述べる。まず、例規名を表す「題名」が最初に位置し、続いて効力を発する日を記した「発令」、公布を宣言する「公布文」、例規の本来の目的とする事項を記した「本則」、そして「制定附則」や「改正附則」が続く。これらが第一階層を構成する。本則には、「章」「節」「条」「項」「号」といった下位階層が構成されていく。図1に例規構造の典型例を記す。ただし、実際の例規では「章」や「節」が存在せず「条」の階層から始まるものも多い事や、制定時期の古い例規では、階層構造に従わない場合もある。すなわち法は、厳密に定義された構造を有さない半構造データであると言える。

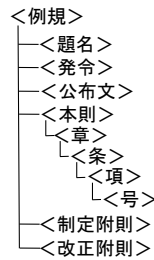


図 1 法の典型的な階層構造
Fig. 1 Typical Hierarchical Structure of Laws

各自治体はその議会において個別に例規を制定している。各自治体が制定する例規の内容は基本的に異なっており、制定された例規の数も異なっている。ただし、各自治体の全ての例規が独自というわけではなく、多くの自治体に共通する事項について制定された例規も多く存在する。前節で述べたように県章を定めた条例や青少年の保護育成を目的とする条例がほぼ全ての都道府県で制定されている事は法学者の中でよく知られている。しかし、全都道府県の類似例規について網羅的に調べた研究は存在しない。これは、そもそも計算機支援なしに類似例規を網羅的に調べる事には現実的でない作業量を要する事と、網羅的に調べる事により得る事ができる知見の価値あるいは、法学分野への貢献度が不明であったためである。そこで本研究では多くの地方自治体に共通して存在する類似例規の抽出を行う軽便な計算機科学的手法を提案するとともに、抽出された結果を用いる事で得られる知見、すなわち例規の法学的な解析例を述べる。

3. 類似例規の自動抽出法

本研究では、ある自治体の例規と類似した例規を他自治体の例規集合から計算機で自動抽出する方法を提案する。この目的を達成するため、二つの例規間の類似度スコアを定める。適切な類似度スコアを定める事ができれば、他自治体の例規集合から最も類似度スコアの高い例規を類似例規として出力する事で類似例規の自動抽出を達成する事ができる。ただし、ある自治体の例規と類似した例規が他自治体に必ず存在するわけではない。類似例規が存在しない事態に対応するため閾値を導入する。ある自治体の例規集合で最も高い類似度スコアが閾値未満の場合、その自治体に類似例規は存在しないと判定する。

いま、二つの例規 x, y の類似度スコアを $S(x, y)$ とし、類似度スコアの閾値を T とする。ある二つの地方自治体 X, Y の例規集合をそれぞれ X, Y とした時、地方自治体 X の例規 $x \in X$ に対する地方自治体 Y の類似例規を以下のように定義する。

地方自治体 Y の例規集合 Y のうち x との間で最も高い類似度スコアを持つ例規を $y \in Y$ とする。例規 $y \in Y$ が $S(x, y) \geq T$ の時、 y を都道府県 Y における類似例規とす

る。例規 $y \in Y$ が $S(x, y) < T$ の時、 x の類似例規が都道府県 Y はないとする。

この定義に適合する計算可能な類似度スコア $S(x, y)$ を決める事ができれば、計算機による類似例規の抽出が可能になる。すなわち、地方自治体 X の例規 x に対応する地方自治体 Y の類似例規の自動抽出が可能となる。また、この自動抽出を地方自治体 X と同階層の全自治体に適用する事により、全地方自治体の網羅的な類似例規抽出が実現できる事は明らかである。

次に、本研究で用いた上記定義に適合する類似度スコアについて述べる。例規は半構造化された自然言語データであるため、自然言語処理で用いられる二文書間の類似度を利用する事が可能である。例規に対して計算機科学的処理を行った先行研究 [6], [7], [16] では、ベクトル空間モデルの相関係数、TF-IDF 重み付きベクトル空間モデルの相関係数、最長共通部分列 (Longest Common Subsequence)[18], 文字列アライメント等が用いられてきた。このうち本研究で着目したのは、竹中らによる先行研究 [6] である。この研究は例規の階層構造のうち<条>に着目した解析を行う事で、二例規間の条文対応表(条をノードとし、類似した条の間に辺を引いた二部グラフ)を自動作成するにあたり、96種類の条間類似度スコアの検討を行っている。そして、条の全文を用いなくとも、条の見出しの最長共通部分列長が優秀な条間類似度スコアとなりうることを明らかにしている。前節でも述べた通り、本研究ではスーパーコンピュータを用いずに解析可能な軽便な解析手法を必要としている。そこで本研究では条文を利用せず、題名の最長共通部分列長に基づく類似スコアを提案する。

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最長共通部分列という。いま文字列 s, t の最長共通部分列を $LCS(s, t)$ とし、各文字列の長さを $|s|, |t|, |LCS(s, t)|$ とする。例規 x, y の題名を $x.title, y.title$ とした時、類似度スコアを以下のように定義する。

$$S(x, y) = \frac{|LCS(x.title, y.title)| * 2}{|x.title| + |y.title|} \quad (1)$$

最長共通部分列の最大長は、入力される文字列長を超えないため、類似度スコア $S(x, y)$ のとりうる値は0以上1以下である。

4. 評価実験

前節で提案した類似度スコアの有効性を検証するため、共通類似例規の推定精度を評価する実験を行う。この実験の結果を用いて適切な閾値 T の値を決める。そして提案手法で得られた網羅的な解析結果から得られる知見、すなわち例規の法学的な解析への応用例を記す。

4.1 実験条件

本研究では、地方自治体の階層として都道府県レベルを

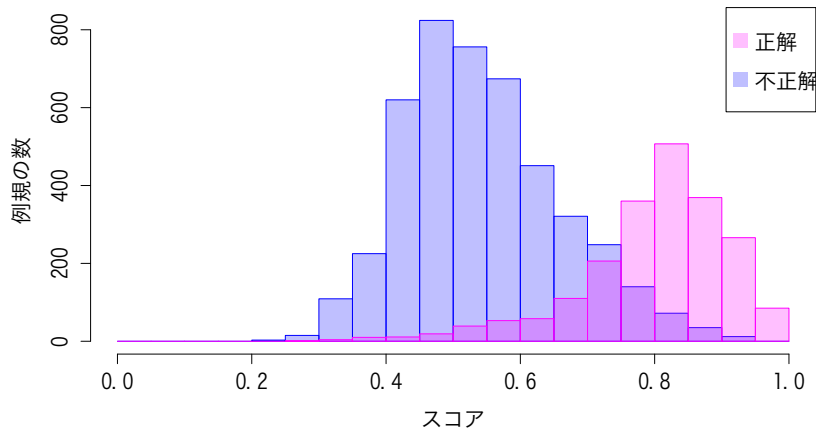


図 2 例規のスコア分布

Fig. 2 Score Distribution of selected Local Government Laws

表 1 類似度スコア最大の例規に対する人手による評価

Table 1 Human evaluation against statutes with the highest similarity

評価値	例規数
一致	2098
種別違い一致	432
高い類似性	641
低い類似性	153
不一致	4505
判別不能	4
合計	7833

対象とする。全国 47 都道府県より選択した 205 個の例規に対して、他 46 都道府県で類似度スコア最大となる例規を抽出した。実験データとして用いた例規数は、47 都道府県合計で 74,880 件であり、抽出された例規の数は 7833 個となった。抽出された例規を類似例規の推定結果とみなし、その推定結果が正しいか否かを人手により判定した。判定は第二著者監督の下で法学部学生が行った。この判定結果を用い、閾値 T を操作変数とする受信者操作性曲線を描く事で有効性を評価するとともに適切な閾値 T の値を決定する。

4.2 評価

類似度スコア最大の例規 7833 個の人手による評価結果を表 1 に記す。本研究では例規を 6 段階で評価した。ここで種別違い一致とは、例規が規定している内容は類似しているものの、条例か規則という種別すなわち、制定主体が議会か行政機関かという違いが存在している事を示す。6 段階の評価値より上位 4 段階を類似例規である、すなわち正解とラベル付けし、同様に下位 2 段階を類似例規でない、すなわち不正解とラベル付けをする。このラベル付け

で分類した類似度スコア最大の例規のスコア分布を図 2 に記す。横軸に類似度スコアを、縦軸に例規数を配したヒストグラムである。図より 2 つの集合のヒストグラムが 2 つの山に分かれている事がわかる。これは適切な閾値 T を決める事により、類似度スコアによって高精度で正しく推定する事が可能である事を示唆している。

高精度での推定が可能である事を示すため、閾値 T を操作した場合の受信者操作特性曲線 (ROC) を図 3 に描いた。図より曲線下面積 (AOC) の面積は 0.929 と高い判別能を有している事がわかる。グラフ左上座標 (0,1) に最も近い ROC 上の点における閾値 T の値は 0.7037 である。この値を閾値として採用する事により推定結果における第一種誤りと第二種誤りバランスをとる事ができる。また、より高い閾値を用いることによって類似例規であると推定した結果中の誤りを少なくし、正解率を高くする事ができる。類似度スコアの閾値 T を変化させた時、 T 以上の類似度スコアを持つ例規の正解、不正解の内訳、及び正解率を図 4 に示す。この図より閾値 T が大きくなるに従い、不正解である例規数がまず減り始め、次いで正解である例規数が減っていく傾向が見て取れる。以上の結果より、最長共通部分列長に基づく類似度スコアは計算機科学的には高い判別能を提供しうる優れた指標である事が明らかとなった。この結果を受け、次節以降に提案手法に基づく例規解析を行う。

5. 法学的解析への応用

前節で定義した類似度スコアを用いた類似例規抽出手法によって可能となる法学的解析への応用例を挙げる。解析には全都道府県の全例規、74,880 個を用いた。他都道府県の全例規との類似度スコアを計算した結果に基づいた。類似度スコア計算を行った例規の組合せ数は 5,473,310,508 個であった。

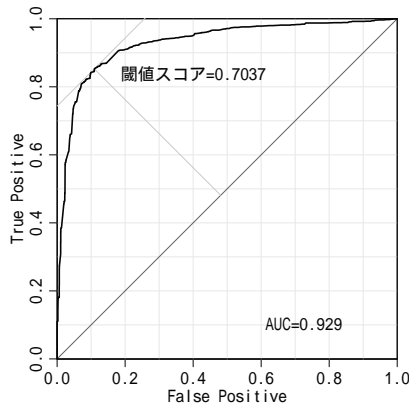


図 3 スコア閾値を操作した場合の受信者操作特性曲線

Fig. 3 Receiver Operating Characteristic Curve when Operating Score Threshold

表 2 類似例規の存在する例規数

Table 2 Number of local movement laws with similar ones

スコア 閾値	都道府県数					合計
	42	43	44	45	46	
1	0	0	0	0	0	0
0.95	43	0	0	0	0	43
0.9	137	82	136	1	0	356
0.85	356	166	237	186	74	1019
0.8	778	431	565	335	187	2296
0.75	1069	1163	903	814	494	4443
0.7	1655	1778	1499	1434	1397	7763

5.1 他都道府県に存在する類似例規数

提案手法を全都道府県の全例規に適用する事により、ある例規と類似した例規が存在する他都道府県を見つける事が可能となる。そこで、まず始めに類似した例規が存在する他都道府県の数に着目した解析例を示す。表 2 に 42 以上の都道府県に類似例規が存在する例規数をカウントした結果を記す。なお、都道府県数はスコアの閾値によって異なるため、前節の ROC 解析で得られた 0.7037 という閾値を利用し、スコアの閾値 T を 0.7 から 1 まで 0.5 刻みで変化させた。ROC 解析で得られたに近い閾値 $T = 0.7$ の場合、42 以上の都道府県に類似した例規が存在する例規が 1 都道府県あたり、 $7763/47 = 165.17$ 個存在する事を表している。この数字を多いととるか少ないととるかは別にしても、都道府県横断的に類似例規の解析を行うには 1 都道府県あたり 165 例規は調べる必要がある事を示唆する値だと言える。

スコアの閾値 $T = 0.7$ の結果を用いて、都道府県ごとの類似例規の制定時期の遅速解析を行った結果を図 5 に示す。各都道府県ごとに、類似した例規が 30 以上の都道府県にある例規のうち、制定年が類似例規中で最も早かった例規数及び、最も遅かった例規数を、それぞれ横軸、縦軸

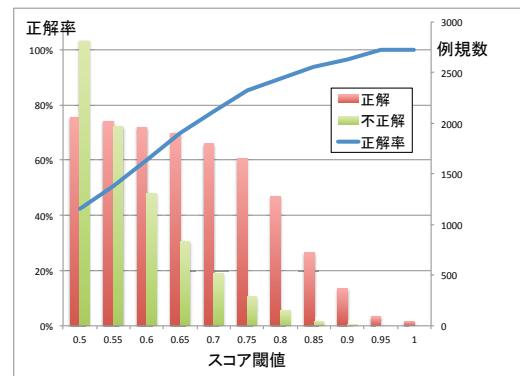


図 4 スコア閾値と正解率

Fig. 4 Score Threshold and Accuracy Rate

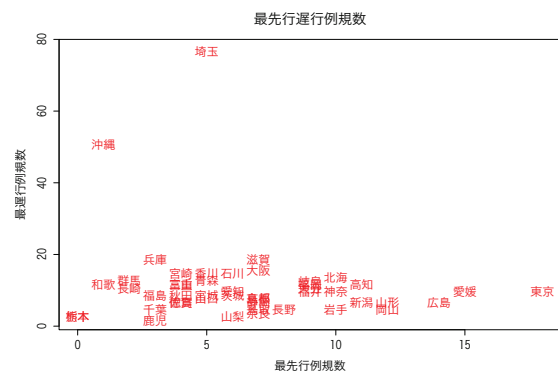


図 5 例規制定時期の遅速

Fig. 5 Establish timing of Local Government Laws

とする散布図を描いた。なお最も早い、又は最も遅い制定年の都道府県が複数ある場合は数え上げ対象から外した。図より、東京都が最初に制定した例規を他道府県が後に制定している傾向や、埼玉県や沖縄県は他の都道府県が制定した後、最も遅れて例規を制定する傾向にある事がわかる。首都が位置する東京都の例規制定が早い傾向にある事や、最南端かつ 1972 年に本土復帰を果たした沖縄県が遅い傾向にある事は尤もらしいと感じる。他方、東京都の隣に位置する埼玉県の制定時期が最も遅いという結果は法学者からみても意外であり、埼玉県の行政方針について研究する切り口となりうる結果であると言える。

5.2 個々の例規解析事例

他都道府県に多くの類似例規が存在する例として、大分県屋外広告条例を挙げる。この例規は、大分県を含めて合計 45 都道府県に類似した例規が存在すると推定された。類似例規が存在しないと推定された都道府県は栃木県と石川県の 2 県であった。この 2 県において類似度スコアが最も高い例規は、栃木県屋外広告物条例〔県土整備部都市計

Vol.12, pp.111--128 (2005)

- [14] 山口高平, 樽松理樹:法律オントロジー, 人工知能学会誌 (小特集 法律と人工知能) Vol.13, No.2 pp.189-196 (1998)
- [15] 樽松理樹, 山口高平:法律知識の体系的定義としての法律オントロジー, 人工知能学会誌 Vol.19, No.2 pp.144-150 (2004)
- [16] 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).
- [17] 角田 篤泰:ソフトウェア工学との類似性に着目した立法支援方法 (三), 名古屋大學法政論集, 名古屋大學法學部 Vol.237, pp.191--252 (2010)
- [18] Maier, David:The Complexity of Some Problems on Subsequences and Supersequences, J. ACM, Vol.25, pp.322--336 (1978).