

木構造編集距離に基づく都道府県例規集構造の類似性評価

竹中 要一

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

若尾 岳志

獨協大学 法学部
wakao2@dokkyo.ac.jp

1 はじめに

都道府県を始めとする地方自治体が各々の議会にて定める法形式に条例がある。この条例に地方自治体が定める規則等を併せて例規と呼んでおり、ある地方自治体の全例規の集合を例規集と呼ぶ。例規は都道府県ごとに階層構造を有する項目に分類されており、これを例規集の構造と呼ぶこととする。本研究では、都道府県の例規集構造の類似性を明らかにすることを目的に、順序木の編集距離に基づく類似度の提案を行う。また、類似度に基づき多次元尺度構成法により都道府県を2次元平面上に描画することで、都道府県例規集構造の類似性の視覚的な把握を容易とした。

例規は構造を有する日本語文書であるため、法務や法学教育支援を目的とした計算機による言語処理の研究対象となってきた。例規を対象とした研究としては、一つの例規の改正過程を明示化する新旧対照表の自動作成 [1]、二つの類似例規の条文対応表の自動作成 [2] がある。また、地方自治体の例規全体を対象とする研究としては、記載内容に基づく例規の自動種別分類 [3] や、ある地方自治体の例規をクエリとし他自治体で類似した例規を発見する研究 [4] や、多くの自治体に存在する類似例規のクラスタを作成する研究が存在する [5]。しかし現在まで本研究以外に例規集構造に着目した研究は行われていない。

本研究対象とする例規集構造について説明する。例規集の階層構造は基本的に上位より、編章節款項と呼ばれる分類単位で構成されている。例えば東京都例規集の最上位階層は以下のような17編で構成されている。第1編:総規、第2編:人事、第3編:財務、…、第17編:消防。また、第1編:総規は第1章:総則から始まる計10章で、同様に第1編第1章:総規は4つの節で構成されている。そして、第1編第1章第1節:通則には「東京都庁の位置を定める条例」をはじめに22個の例規が分類されている。図1に東京都の例規集構造の一部を記す。なお各階層には通し番号が付与されているが図では省略している。

例規は地方自治体が独自に制定しているため、例規

を分類する例規集構造も地方自治体毎にそれぞれ異なっている。表1に例規構造の最上位階層である編の名称一覧を記す。表より都道府県によって編の数や名称が異なる事がわかる。そして、第一編の名称は「総規」または「総務」と類似しているものの、他の名称は微妙に異なり、かつ対応関係が分からないものが多数存在している。また、4都府県に共通する編名が存在しない事も明らかである。このように、例規集構造にも地方自治体の独自性が反映されている事がわかる。

地方自治体によって階層構造が異なることを示す例として、石川県を除く46都道府県で制定されている「屋外広告物条例」が各都道府県において例規集構造のどこに属しているかを表2に示す。表より編名に「建設」「建築」「土木」といった語が、章名に「都市計画」「建築」「計画」「広告」といった語が頻出している事がわかる。前段落の表1において、都道府県の例規集構造が異なると記したが、表2から何らかの類似性が存在することが伺える。本研究では、都道府県の例規集構造の類似度を情報科学的に定義する事で、その類似性を明らかにする事を目的とする。

2 例規集構造の類似度の定義

例規集構造は、情報科学的には木構造で表現する事ができる。編章節等の各階層の項目を頂点とし、項目が所属する上位階層の項目との関係を辺として表す。ただし、例規集構造では編が最上位階層であり根が存在しないため、データ構造上、全ての編を子とする根を置く事とする。なお、頂点是对応する項目名がラベルとして与えられてるものとする。また、例規集構造では、同一の親を持つ項目毎に、第一章、第二章といった通し番号が割り振られている。そのため、この木構造は兄弟ノード間に順序のある順序木であると言える。

今、ある地方自治体 A, B の例規集構造を表現した順序木を T_A, T_B とする。本研究では、地方自治体 A, B の例規集構造の類似度を順序木間の編集距離 $TED(T_A, T_B)$ で定義する。

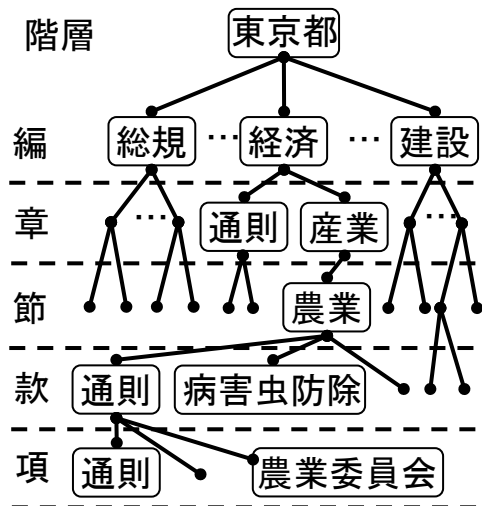


図 1: 東京都の例規集構造 (一部抜粋)

表 1: 最上位の例規構造「編」の名称

東京都	大阪府	長野県	福井県
総規 人事 財務	総規 職員 財務	総規 財務 社会・衛生 ・生活環境	総務 健康福祉 県民生活
福祉 労働 衛生 経済 都市計画 環境保全 住宅 建設 港湾 交通 水道・下水道	環境保全 民政 衛生 商工 農林 労働 土木 建築 公営企業 教育文化 警察	経済 土木・建築 教育・文化 警察・消防 公営企業 会計関係例規 - - - - -	商工労働 農林水産 土木 公営企業 教育 警察

表 2: 屋外広告物条例の例規集構造における位置

階層	編	章	節
北海道	建設	都市計画	景観・美観風致
秋田県	建設	都市計画	-
東京都	都市計画	通則	-
神奈川	都市	屋外広告物-	-
長野県	土木・建築	建築	-
福井県	土木・建築	建築	-
滋賀県	土木・建築・観光	計画・観光	-
奈良県	土木	建築 広告	-
大阪府	土木	屋外広告物	-
岡山県	土木	都市計画	-
広島県	土木建築	計画	-
愛媛県	土木 (2)	広告物	-
香川県	土木	屋外広告物	-
福岡県	土木	計画	広告物

順序木の編集距離 $TED(T_A, T_B)$ は, T_A を T_B に変換するのに必要な編集操作の最小回数, または, 編集操作の操作重み合計の最小値として定義される. 編集操作は, [1. 削除] T_A の頂点 v を削除し, v の子を v の親の子とする, [2. 挿入] 削除の逆操作, [3. 置換] T_A の頂点 v のラベルを異なるラベルへと置き換える, の 3 種類で構成されている.

順序木の編集距離の計算法は, 1979 年 Tai によって $O(n^6)$ 時間アルゴリズムが提案された [6]. その後, 1989 年には Zhang らにより動的計画法を用いた $O(n^4)$ 時間アルゴリズム [7] が, 2009 年には E. Edemaine らにより $O(n^3)$ 時間アルゴリズム [8] が提案されている.

3 評価実験

全国 47 都道府県の例規集構造を対象に前節に掲げた木構造編集距離の計算を行った. 例規集構造の類似性を俯瞰するため, 多次元尺度構成法を用いて 2 次元平面上に描画した. 順序木の木構造編集距離の計算にあたり, 編集操作の重みとして以下 2 種類の距離を用いた.

- A) ラベルが一致した場合 0, 一致しない場合 1.
- B) ラベルの文字列編集距離

木構造編集距離の計算には Zhang-Shasha のアルゴリズム [7] を実装した Python の zss ライブラリを用いた. 多次元尺度構成法には, R の Classical Metric Multidimensional Scaling 関数 (cmdscale) を用いた.

47 都道府県の全組合せに対する木構造編集距離の分布を図 2,3 に, 各種統計量を表 3 に示す. 同表には最小距離及び最大距離を示した都道府県の組合せを併記した. 図 4,5 には多次元尺度構成法に基づき都道府県を二次元平面上に描画した図を示す.

図 2,3 より, 2 種類の編集操作重みそれぞれの分布が類似している事がわかる. また, 最大距離をとる都道府県の組合せは, いずれも (北海道, 広島県) である. 最小距離の組合せは, A) 階層名の一致で (秋田, 長野), B) 階層名の編集距離で (長野, 福井) と異なる. しかし, (秋田, 長野) 間の B) による距離は 397, (長野, 福井) 間の A) による距離は 101 と, それぞれの最小値に近い値を示している.

最大距離を取る組合せとなった (北海道, 広島県) は, それぞれ図 4,5 においてはずれ値となっていることがわかる. 東京都, 神奈川県は両方の図において外れ値になっている事から, 両県の例規集構造は他道府県と異なる度合いが高いことがわかる. 他方, 愛知, 富山, 福井, 滋賀, 奈良, 徳島, 岡山, 福岡の各県は

表 3: 木構造編集距離の統計値

	A:階層名の一致	B:階層名の編集距離
最小距離	96 (秋田, 長野)	378 (長野, 福井)
最大距離	379 (北海道, 広島)	1794 (北海道, 広島)
平均	239	1007
標準偏差	49	247

両図において近隣に位置しており、類似性が高いクラスタを形成している事がわかる。

4 まとめと今後

例規集構造の木構造距離に基づき都道府県の類似度を計算し、その結果を多次元尺度構成法にて平面上に写像した。二種類の編集操作重みの結果は、おおよそ一致している事が図より見て取れる。本研究の結果により、都道府県の例規集構造の類似性の一端を明らかにできた。提案手法は、例規を制定、改訂する際に参考とする自治体を探す際に役立つと考えている。特に町村等の小規模自治体の例規集に対して提案手法を適用する事が有意義であると思われる。

日本には包括的地方自治体である都道府県以外にも、基礎的地方自治体である市町村が存在する。そして市町村もそれぞれの議会にて例規を制定しており、都道府県同様に階層構造を持つ例規集を有している。そのため本研究の手法を用いる事により市町村間の類似性を評価する事が可能となる。2014年1月1日現在、日本には1742の基礎的地方自治体が存在する。その中には人口1万人以下の村も多く含まれる。このような自治体は職員数が少なく、例規を制定するにあたり、他自治体で制定された例規を参考とする必要性の度合いが高いと思われる。参考とする自治体を見つけるにあたり、本研究の手法を基礎的自治体に対して適用した結果は有効であると思われる。

また、本研究は各自治体に存在する類似例規の発見や類似例規クラスタの作成に関する研究[4, 5]にも利用可能な情報だと考えている。これらの研究では例規の見出しや本文の類似性に基づいた計算処理を行っているが、例規が属する例規集の階層についての知見を用いていない。しかし表2に示したように、類似した例規の属する階層は類似している事が予想される。従って、木構造編集距離の計算過程で得られる階層構造の対応関係の利用が上記研究[4, 5]の高精度化に資すると想定している。

本研究では、例規集構造の各項目の番号を鑑み、順

序木とみなして距離を計算した。しかし表2より、階層名の出現順を考慮しない方が望ましい可能性がある。これは、例規集構造を無順序木と見做す事と同一である。無順序木の木構造距離を計算するアルゴリズムはNP困難であるため[9]近似解法を用いる必要があるが、有意義な結果が得られる可能性があるため今後の課題としたい。

参考文献

- [1] 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), 第237巻, 第二節, pp. 191–252. 名古屋大学法学部, 2010.
- [2] 竹中要一, 若尾岳志. 地方自治体の例規比較に用いる条文対応表の作成支援. 自然言語処理, Vol. 19, No. 3, pp. 193–212, sep 2012.
- [3] 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.
- [4] 竹中要一, 若尾岳志. 都道府県に共通する例規の抽出と応用. 情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告, Vol. 2014, No. 5, pp. 1–7, sep 2014.
- [5] 角田篤泰. スーパーコンピュータを用いた自治体例規の類似度分析と例規データベースへの応用. 名古屋大学法政論集, Vol. 246, pp. 91–69, sep 2012.
- [6] Kuo-Chung Tai. The tree-to-tree correction problem. *J. ACM*, Vol. 26, No. 3, pp. 422–433, July 1979.
- [7] K. ZHANG. Simple fast algorithms for the editing distance between trees and related problems. *SIAM Journal on Computing*, Vol. 18, No. 6, pp. 1245–1262, 1989.
- [8] Erik D. Demaine, Shay Mozes, Benjamin Rossman, and Oren Weimann. An optimal decomposition algorithm for tree edit distance. *ACM Trans. Algorithms*, Vol. 6, No. 1, pp. 2:1–2:19, December 2009.
- [9] Kaizhong Zhang, Rick Statman, and Dennis Shasha. On the editing distance between unordered labeled trees. *Inf. Process. Lett.*, Vol. 42, No. 3, pp. 133–139, May 1992.

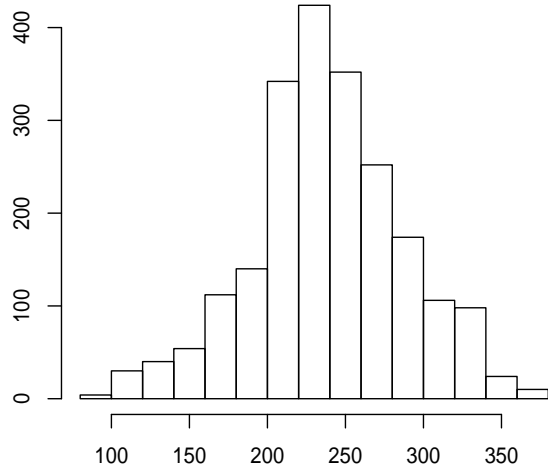


図 2: 木構造編集距離の分布 (A:階層名の一致)

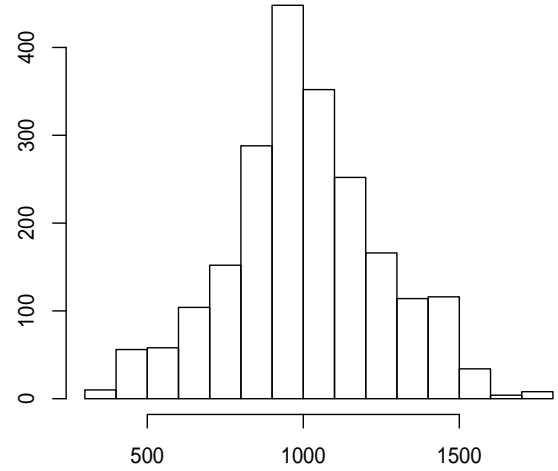


図 3: 木構造編集距離の分布 (B:階層名の編集距離)

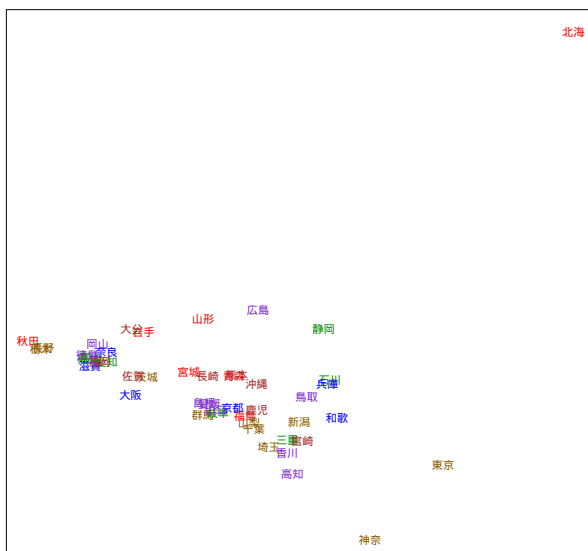


図 4: 例規集構造の類似性 (A:階層名の一致)

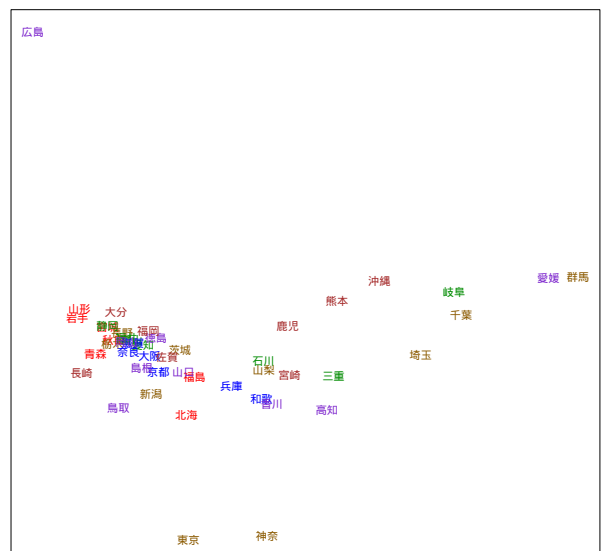


図 5: 例規集構造の類似性 (B:階層名の編集距離)