

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 20 日現在

機関番号：32406

研究種目：基盤研究(C)

研究期間：2009～2013

課題番号：21500253

研究課題名(和文) 例規条項の自治体間対応関係と差異の網羅的な自動抽出—道州制への円滑な移行に向けて

研究課題名(英文) Computer based methods for automatic sampling and correspondence table generation for similar regulations among multiple municipalities

研究代表者

若尾 岳志 (Wakao, Takeshi)

獨協大学・法学部・准教授

研究者番号：50388328

交付決定額(研究期間全体)：(直接経費) 3,400,000円、(間接経費) 1,020,000円

研究成果の概要(和文)：例規集の自治体間、特に都道府県間の差異の明確化を目的に、計算機による類似例規の自動抽出法及び例規の条文対応表の自動作成法の研究を行った。まず、全都道府県の例規集のXML文書化を行った。また、条文対応表の自動作成法として、条文に対するTF-IDF法の有効性を示した。この手法は長い計算時間がかかるため、別の手法を模索した所、評価尺度として条文見出しに対する最大共通文字列長が計算時間が短く判別能力も高かった。他、都道府県の類似例規の自動抽出にも、条例名に対する最大共通文字列長を評価尺度にする手法が有効であった。以上の成果を利用する事で、自治体合併や道州制移行時に発生する法務作業量の削減が可能になった。

研究成果の概要(英文)：For the purpose of clarifying differences in regulations among municipalities, we studied methods to automatically sample similar rules by computer and to automatically generate the clause-correspondence table between two municipalities. We initially converted codes of all Japanese prefectural governments, which are open on Website, into XML documents and demonstrated that TF-IDF method is effective to automatically generate the clause-correspondence table. However, the TF-IDF method requires too long. We investigated alternative methods with shorter computational time and identified that the length of longest common subsequence (LCS) among clause-headings is the most effective with the shortest time. We also clarified that the length of LCS is effective to automatically extract similar regulations from other prefectural governments. Thus, we concluded that our study should contribute to reducing workloads of legal sections of municipalities, especially when they would merge.

研究分野：総合領域

科研費の分科・細目：情報学、図書館情報学・人文社会情報学

キーワード：言語処理 都道府県 道州制 条例 計算機 自動抽出 アライメント 例規

## 1. 研究開始当初の背景

比較法学という分野が既に確立していることから明らかなように、各国の法律の比較研究は盛んに行われている。一方、国内の地方自治体の例規全体を自治体間で比較することは、市町村合併の協議会等の実務としては行われているものの、研究として扱ったものは皆無と言ってよい程存在しなかった。これは、一つには、国の権限が大きいこと、比較対象となる例規間で差が少ないことがあり、もう一つには、例規集の基本構造である体系が整っており、比較対象を探しやすいことがあると考えられる。

日本の地方自治体は「3割自治」と呼ばれるように、国から地方への補助が歳入の大きな割合を占めており、国によって地方の財政が縛られているといえる。それと同様に地方自治体の活動を示す例規も、国の強い権限により陰に陽に監督されているため、地方自治体間で例規に大きな差が生じにくい状況がある。そのため、あえて例規を比較する研究が発生しなかったと考えられる。

また、このことに付随する事柄であるが、各自治体で制定される例規の体系も類似しており、その対応関係を人が容易に理解しやすいものとなっている。そのため、ある例規に類似した例規を他の地方自治体の例規集から探すのは比較的容易であるため、比較すること自体が学術的な研究にそぐわないと看做されてきたと考えられる。

さらに付言するならば、例規が「日本語で書かれている」という所にも学術的な分野と看做されてこなかったという要因があると考えられる。この点は、詳しく述べない。

法務の背景としては、地方自治体への権限委譲の一方法として、道州制の導入に向けた動きが活発化していた。2006年2月28日には内閣総理大臣の諮問機関である地方制度調査会が道州制のあり方に関する答申を発

表し、2008年5月29日には、政権与党の自由民主党の道州制推進本部が道州制の区割り案を報告している。道州制の導入あたり、名称や州都の位置、財源と権限の移譲など様々な問題点が指摘されているが、その導入にむけて着実にことが進んでいる。その際、都府県の例規を道州ごとに一つの例規へと纏め上げなければならないが、それに伴う困難な点や作業量に対する問題点の提起はこれまでになされていない。問題とされてこなかった理由の一つとして挙げられるのは、全国における市町村合併に伴い例規集の統一作業が行われてきた実績があるからである。しかしながら、以下に示す4項目の理由により、市町村例規集の統一と都府県例規集の統一とは全く条件が異なり大きな問題を孕んでいると言える。

- (1) 市町村に比し、都道府県の例規はその件数が多いこと
- (2) 市町村合併と異なり、都府県の合併は全て対等に扱われなければならないこと
- (3) 同時に合併する自治体数の最大値が9つ(9道州制の中国・四国州における、鳥取県、島根県、岡山県、広島県、山口県、徳島県、香川県、愛媛県、高知県)もあると想定されること
- (4) 統一作業を同時期に実施する必要があるため、人的資源が不足するであろうこと

以下、各項目について少し詳しく述べる。

まず、(1)について、東京都例規は2296件、埼玉県例規は3435件であるのに対し、埼玉県の県庁所在地であるさいたま市の例規は982件、埼玉県狭山市は731件、埼玉県嵐山町で591件である。このように都道府県の例規数は、市町村の2倍から3倍程度の大きさとなっている。

また、(2)及び(3)に関連することであるが、市町村合併では1乃至2自治体が主導権を握

って合併を推進するため、これら主導権を握る自治体の例規を中心として協議を行うことは容易である。しかしながら、最大で9県の例規を対等な形で行うのは容易ではなく、客観的な基準が不可欠となるであろう。

さらに(3)について、ある例規に対応する他自治体の例規数が常に1件である場合、その対応関係の数はある自治体の例規件数と等しくなる。しかし、実際にはある例規に対応する他自治体の例規数は1件とは限らず、0件であったり2件であったりする。この場合、対応関係は自治体の組合せ数と同様に大きくなっていき、その最大数は、統一する地方自治体の指数乗に上ることになる(具体的には6)。

自治体の例規の対応関係数に比べて、9自治体の例規の対応関係数は、 $9/6=1.5$ 倍ではなく、8倍(2の(9-6)条=2の3条)となる。実際にはそこまで多くなるとしても、自治体の例規件数を合計した数よりもはるかに多くの対応関係を調べなければならなくなる。

最後に(4)についてであるが、一時期に同類の専門家が必要となるような施策において、専門家の人的資源が不足するのは常のことであるため、事前に準備しておく必要が発生する。

以上の理由により、例規の対応関係を明らかにし、その差異を明らかにする研究は、必要であると判断される。また、本研究の成果は、道州制の実現のみに資するのではなく、実現化されることにより個性化するであろう道州間の比較研究を行うための手法としても有効に機能すると考えられた。

## 2. 研究の目的

本研究では、例規集の自治体間、特に都道府県間の差異を明確化させることを目的とする。この目的を達成させるため、まず、自治体間において対応する例規条項を網羅的に抽出し、且つその差異の明示を自動的に行う計

算機科学的手法の開発し、次に、その結果の検証及び公表を行うこととした。例規の網羅的な比較は、市町村レベルにおいて実務として行われてきたが、都道府県レベルで行われたことはない。しかし道州制の導入が決定された暁には短時日で行わなければならない。現状ではある例規に対応する他自治体の例規を探すことは比較的容易である。しかしながら、この対応関係は必ずしも一対一に対応しておらず、また全射や単射の関係も保持されない。そのため、全例規の対応関係を見ていくことは複雑な作業となり人手で行うのは困難である。そこで本研究ではこの作業を計算機で自動的に行うための方法の開発、及び結果の検証及び公開を行う。本研究の成果は、道州制施行時に役立つのみではなく、今後自治体への権限委譲が進むに従って複雑化するであろう例規の比較にも有効である。

## 3. 研究の方法

本研究の目的を達成するため、コンピュータを用いて例規・条項の類似性を判定し、類似した例規・条項における相違点を明確化する客観的な手法を構築する。これを実現するため、計算機科学分野における動的計画法と呼ばれるアルゴリズムを用いる。

具体的には、コンピュータを用いた例規の類似性及び、類似した例規における条項レベルの相違点を明確化する方法を確立させる。類似性の判定として文字レベル、単語レベル、法令用語語彙体系レベル(法律オントロジーレベル)の3レベルを試み、精度の高い計算手法を提案する。このコンピュータによる結果は、法学者の検証を経ることにより実用性の担保をはかる。

上記の実現化のため、全都道府県の例規収集をするとともに、コンピュータが容易に例規を扱うことを可能とするために例規のXML化を行う。各都道府県のウェブページにおいて例規データベースと称するものが公

開されている。これらのデータベースは人が閲覧するのに適した形式で供されているが、計算機が扱うのには適していない。具体的には現状のデータベースでは、計算機は例規の条文を一単位として扱っても、条や項を単位として扱うことはできない。より端的にいえば、現状公開されている例規データベースは、計算機科学分野でデータベースと称するものとは異なるものである。そのため、計算機の扱いを可能とする例規データベースを構築する必要がある。本作業は計算機プログラムによる自動構築を基本とするが、各都道府県で異なるプログラムを作成する必要がある。

例規データベース作成の後、都府県間で類似した例規の対応関係を抽出し、且つ類似した例規中の相違点を明示するための計算機科学的手法の開発と実装、及び結果の検証を行う。類似する対応関係を抽出するためには、動的計画法を用いた整列アルゴリズムを用いる。このアルゴリズムは、類似遺伝子の抽出、計算機プログラムの冗長部位の発見、盗作レポートの検出等で実用化されている。本アルゴリズムを例規に対して適用することで、目的とするプログラムの開発と実装を行う。なお適応する際に例規を分割する単位として、文字レベル、単語レベル、意味レベル（法令用語語彙体系レベル）の3種類が考えられる。いずれも頻度統計及び情報量に基づいた類似度判定を行うため、文字レベルよりも単語レベル、意味レベルの方が優れているとは限らない。そのため3種類すべてを実装し、その結果を人が検証することで最も優れたものを1つ決定する。

#### 4. 研究成果

研究成果として、(1)例規の条文対応表の自動生成と(2)全都道府県の類似例規自動抽出の2項目に分けて記述する。

#### (1) 例規の条文対応表の自動生成

条文対応表を計算機によって自動作成することを目的とし、その目的を達成するため、条文対応表のうち条文の対応関係に着目し、各条をノードとする有向二部グラフとしてモデル化した(図1)。

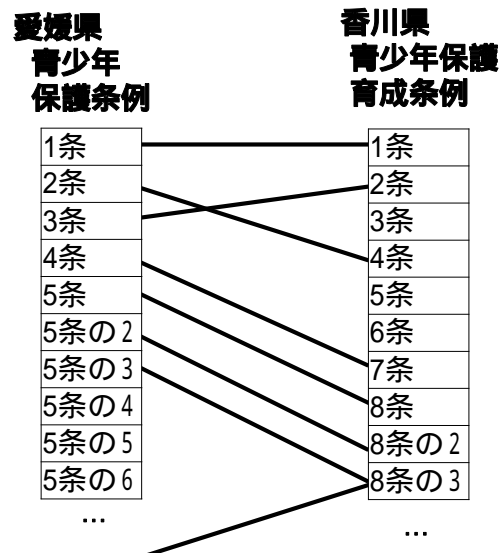


図1 例規対応表の二部グラフモデル

そして、対応関係となる条の適切な探索に有効な方法を明らかにするため、ベクトル空間モデル、最長共通部分文字列、及び文字列アライメントに基づく96種類の手法に対する検討を行った。検討にあたり、愛媛県と香川県の22条例に対して条文対応表を自動生成し、法学者が作成した条文対応表に基づき正解率を求めた。その結果、名詞、副詞、形容詞、動詞、連体詞を対象としたベクトル空間モデルに基づく手法の正解率が85%と最も高かった。また、文字列アライメントに基づく手法の正解率は最高で81%、最長共通部分文字列は75%であった。

#### (2) 全都道府県の類似例規自動抽出

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、都道府県が制定した例規の数は、最少の山口県でも974件、最多の北海道では4156件あり、全都道府県の例規数の合計は74,880件(2012年9月現在)

である。全都道府県における網羅的な比較対象例規の抽出は困難である。そこで、本研究では、比較対象となる例規を全都道府県から抽出するための手法の提案を行った。

提案手法は以下の通りである。二つの例規間に類似度スコアを定め、他都道府県で最も類似度スコアの高い例規を類似例規候補とする。類似例規候補のうち、類似度スコアがある一定値以上の場合、類似例規と見做す。この最長共通部分列 (Longest Common Subsequence) に基づき定義した類似度スコアが、受信者操作特性曲線の曲線下面積が 0.929 と大変優れた性能を有することを明らかにした (図 2)。

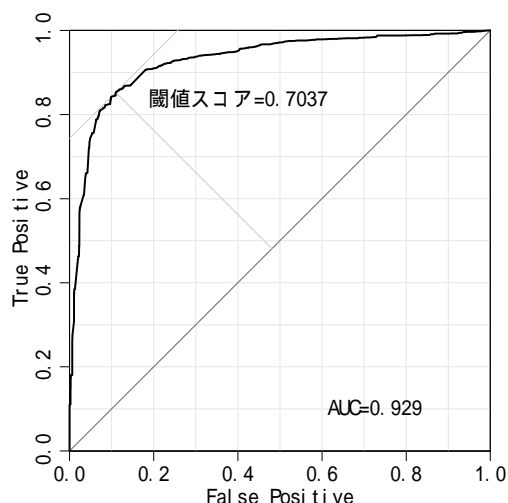


図 2 受信者操作特性曲線

本成果を用いた解析例を挙げる。大分県屋外広告条例と類似した例規を持つ都道府県名を、例規の制定年を横軸としてプロットした図を以下に記す (図 3)。

図より、東京、大阪といった大都市で 1950 年代という早期に広告条例が制定され、それが地方に波及していったという傾向が見て取れる。本研究の成果は、類似例規の自動抽出に留まらず、計算機を用いた新たな例規解析が可能となることを明らかにした。

大分県屋外広告条例

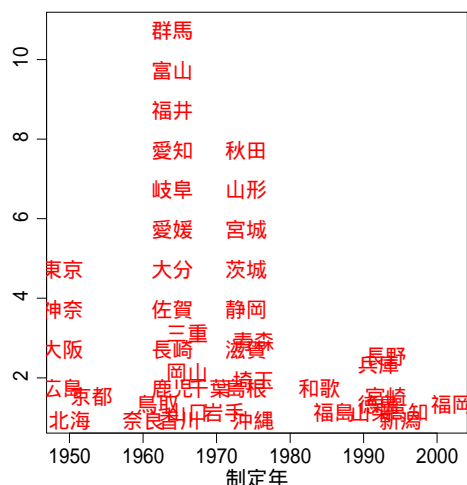


図 3 大分県屋外広告条例に類似した例規の存在する都道府県と制定年

## 5. 主な発表・論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

- (1) 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 Num.3 pp193-212 (2012年 09 月)

〔学会発表〕(計 2 件)

- (1) 竹中要一, 若尾岳志, 「都道府県に共通する例規の自動抽出」言語処理学会第 20 回年次大会, B6-4, 北海道大学, 2014 年 3 月 20 日
- (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自動生成」, 言語処理学会 第 17 回年次大会 発表論文集 pp.520-523 (D2-5), 豊橋技術科学大学 (2011 年 3 月 9 日).

〔図書〕(計 0 件)

〔産業財産権〕  
出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

ホームページ等

<http://www-mats.ist.osaka-u.ac.jp/~takenaka/research/legalinfo/>

## 6. 研究組織

- (1) 研究代表者

若尾 岳志 (WAKAO, Takeshi)

獨協大学・法学部・准教授

研究者番号：50388328

(2)研究分担者

竹中 要一 (TAKENAKA, Yoichi)

大阪大学・大学院情報科学研究科・准教授

研究者番号：00324830

(3)連携研究者

なし