

都道府県に共通する例規の自動抽出

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

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。制定法の様々な法形式のうち、地方自治体が定めているのは条例である。この条例に、地方自治体が定める規則を合わせて、例規と呼んでいる。国会が定める法律は、国際的な比較をする意味はあっても、国内的には、同一事項については一つしか存在しない。そのため、時系列に添った比較ならいざ知らず、並列的な比較の意味はあまりない。これに対して、同一事項について規定している例規は、同じ地位にある、各地方自治体ごとに存在することが想定できる。例えば、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が、その制限される時間が異なっている事が挙げられる。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。このような違いを明確化するため例規比較が行われる。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 [4, 5, 6]。

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、各都道府県が制定した例規の数は最少が山口県の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)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うとともに、その結果の利用例を示すにより、手法の有効性及び実用性を明らかにする。

法を計算機で扱う研究は、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。日本では平成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」において促進された [2]。この期間を通してインターネット上における法律の閲覧が可能となり、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。例規に関する研究としては、例規種別の自動分類 [3]、例規改正の編集過程に基づき、改正前後の差異を表現した新旧対照表の自動作成 [8] がある。また、例規比較を支援する研究として、二例規間の条文対応表の自動作成 [7] が挙げられる。しかしながら、比較を行う対象となる、複数自治体で共通する例規を抽出する研究はこれまでに存在しておらず、本研究が嚆矢である。

2 都道府県の類似例規

本研究で対象とする例規とは、法特有の階層構造を有する文章である。典型的には、例規名を表す「題名」、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」が第一番目の階層を構成する。このうち本則は、「章」「節」「条」「項」「号」といった副階層を有している。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く、特に制定時期が古い例規では階層構造に従わない場合もある。章と一部の条にはその内容を記し

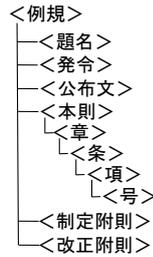


図 1: 例規の典型的な階層構造

た見出しがついている。

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。前述の通り、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ものの、全都道府県の類似例規について網羅的に調べた研究や、網羅的に調べる手法に関するは現時点で存在しない。そこで本研究では多くの都道府県に共通して存在する類似例規の抽出を計算機により行う手法を提案する。

3 類似例規の自動抽出法

多くの都道府県に共通して存在する類似例規を抽出するため、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、他都道府県で最も類似度スコアの高い例規を見つける事により類似例規の候補を抽出する事ができる。ただし、必ずしも類似例規が存在するとは限らない。そのため共通する類似例規が存在しない状況に対しては、ある閾値以上の類似度スコアを持つ場合のみ類似例規と判定する事により対応する。

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 と設定する。ある都道府県 A の例規 x に類似する他の都道府県 B の類似例規を以下のように定義する。

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

これを A 以外の 47 都道府県を B として適用する事により全都道府県における x の類似例規を抽出する事が可能となる。すなわち、上記方法で正しく類似例規

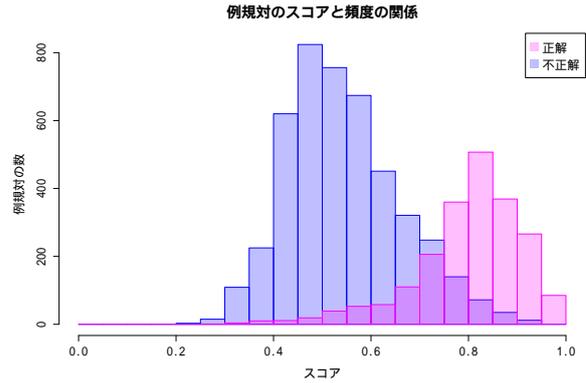


図 2: 例規対のスコア分布

を抽出する事ができるような類似度スコア $S(x, y)$ を定義する事ができれば、共通類似例規の自動抽出が可能となる。

本研究では、類似度スコアを、最長共通部分列 (Longest Common Subsequence) に基づき定義した。最長共通部分列は、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い [1]、例規の条文対応表を自動生成する手法としても有効である事が明らかとなっている [7]。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

前節で定義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った。

4.1 実験条件

全国 47 都道府県より抽出した 205 個の例規に対して、各都道府県で類似度スコア最大となる例規を抽出した。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判定した。判定は第二著者である若尾による監督の下で法学部学生が行った。また、評価に用いた (例規, 都道府県) 対の数は 7833

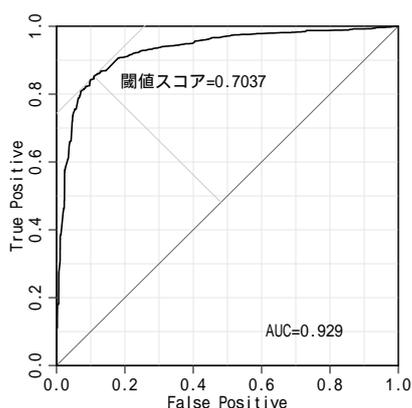


図 3: 閾値を操作した場合の受信者操作特性曲線

個である。なお、使用した例規は各都道府県が2012年9月時点にてインターネットで公開している例規集に基づいた。

4.2 評価

類似例規と類似度スコアとの関係を調べた結果を図2に記す。類似例規だと正しく推定された例規集合及び、誤って類似例規だと推定された例規集合のそれぞれについて、類似度スコアを横軸として描い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ことを示唆している。この事を明らかにするため、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と類似例規だと正しく推定できた例規対の数及び、誤って類似例規だと推定された例規対の数及び正解率を表したグラフを図4に示す。

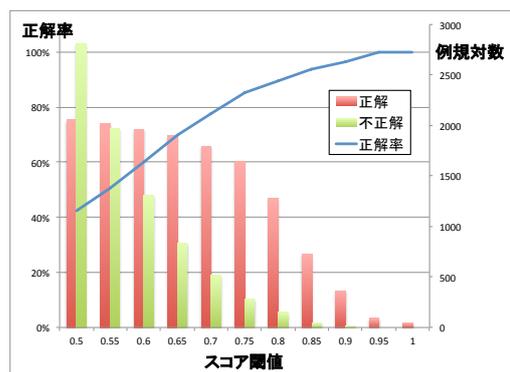


図 4: スコア閾値と正解率

5 応用

前節で定義した類似スコアを用いる事によって可能となる応用例として2件の都道府県の例規解析を挙げる。解析元となったデータは、全都道府県の全例規74880に対して、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対数は5,473,310,508であった。

5.1 多くの都道府県に存在する類似例規

解析を行う事により、多くの都道府県に存在する類似例規を見つける事ができる。その一例として、大分県屋外広告条例に類似した例規の結果を挙げる。本例規は、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〔県土整備部都市計画課〕(類似度スコア0.5)と石川県公告式条例(類似度スコア0.518)である。栃木県で最も類似度の高い例規は、大分県屋外広告条例と類似しているものの、類似度スコアが閾値を下回ったために類似していないと推定された、偽陰性(第二種過誤)である。一方、石川県に類似した題名の条例は存在しない事がわかる。この情報に基づき石川県の条例集を調べた結果、平成20年7月に屋外広告物条例と景観条例が一本化された「いしかわ景観総合条例」が制定された事がわかった。すなわち、この解析結果は石川県の景観に関する取り組みに関する特長を表すものである。このような調査を行う端緒となる本解析手法は有用である事がわかる。

