

平成23年度 情報工学コース卒業研究報告要旨

外山 研究室	氏 名	今 田 悠 幾
卒業研究題目	地方自治体の条例に対するクラスタリング手法の比較	
<p>条例とは、国が定める法令とは異なり、地方自治体が独自に定める自主法のことである。各地方自治体が独自に定めるので、同じタイトルの条例でも各自治体によって内容や各用語の定義が異なる。例えば、長野県を除く全ての都道府県で制定されている青少年保護育成条例では、「有害薬品の販売等制限」は茨城県や群馬県など一部の自治体が採用しているのみであり、深夜外出の制限に関しては「深夜」と定義する時間が自治体によって異なっている。これは、各自治体がその地域の実情に合わせた形で条例を制定すること起因する。このため自治体が新たに条例を作成する際には、国が提示する条例準則や他の自治体の条例を職員の手作業によって比較した上で、その自治体に合わせて内容を取捨選択、あるいは変更してきた。しかし、近年公務員が削減されてきたり、景気が低迷して自治体の収入が減少していたりすることから、条例作成事務に回せる職員や時間が少なくなっている。以上の背景から、条例の作成事務におけるコストダウンが必要である。そこで本研究では条例の比較作業に着目し、コンピュータによって支援することを目的とする。</p> <p>条例を比較する際には、それぞれの内容について各条例に採用されているかどうかを調べたり、同じ用語に対して異なる定義をしていないかどうかを調べたりする。そのため、条例を構成する各条を同じ内容ごとにまとめていく必要がある。そこで本研究は、条例どうしの比較支援を各条のクラスタリングによって行う。ところが、クラスタリングには様々な手法が知られている。そこで本研究では、単語や文書の類義性を考慮する際の閾値となるLSIの累積寄与率、要素間の距離の計算方法である距離関数、要素をまとめる順序を決定するクラスタリング方法、およびクラスタリングの終了距離の計4種類のパラメータを設定し、各都道府県の青少年保護育成条例を用いて、条のクラスタリングに適したパラメータの組み合わせを実験的に求めた。なお、距離関数はピアソン相関距離、コサイン距離、ユークリッド距離、マンハッタン距離、キャンベラ距離を、クラスタリング手法は最近隣法、最遠隣法、群平均法、重心法、中央値法、ワード法、マッククイティ法、およびk平均法を対象にした。</p> <p>上記の実験で求めたパラメータの有効性を検証するために、各都道府県の暴力団排除条例47条例にも適用してクラスタリングが直観と一致するかどうかを評価した。手作業で作った正解セットを用いて評価すると、F値は0.7737となった。クラスタリング結果を見ると、各クラスタに属する条の平均82.9%が同じ内容であった。</p> <p>また、ほぼ全ての条例の各条には、その条の内容を端的に表す条見出しがつけられている。そこで、各クラスタに属する条の条見出しについて多数決を用い、そのクラスタの名前を決定した。暴力団排除条例に対する各クラスタを対象として名前付けを行ったところ、付けられた名前のうち76.9%が正解セットにおいて付けられた名前に一致していた。</p> <p>以上から、実験で求めたパラメータを用いた条例のクラスタリングは有効であることが明らかになった。</p>		