

[研究課題名]

AI技術を用いた法的文書作成支援

(JPMJCR1782)

2019年6月24日

角田篤泰

中央大学・国際情報学部・教授

背景

法的文書

法令

契約・
約款

条例・
規則

規程

法的文書作りは大変！

法令作りも大変。
条例作りも大変。
規程作りも大変。
契約書作りも大変。



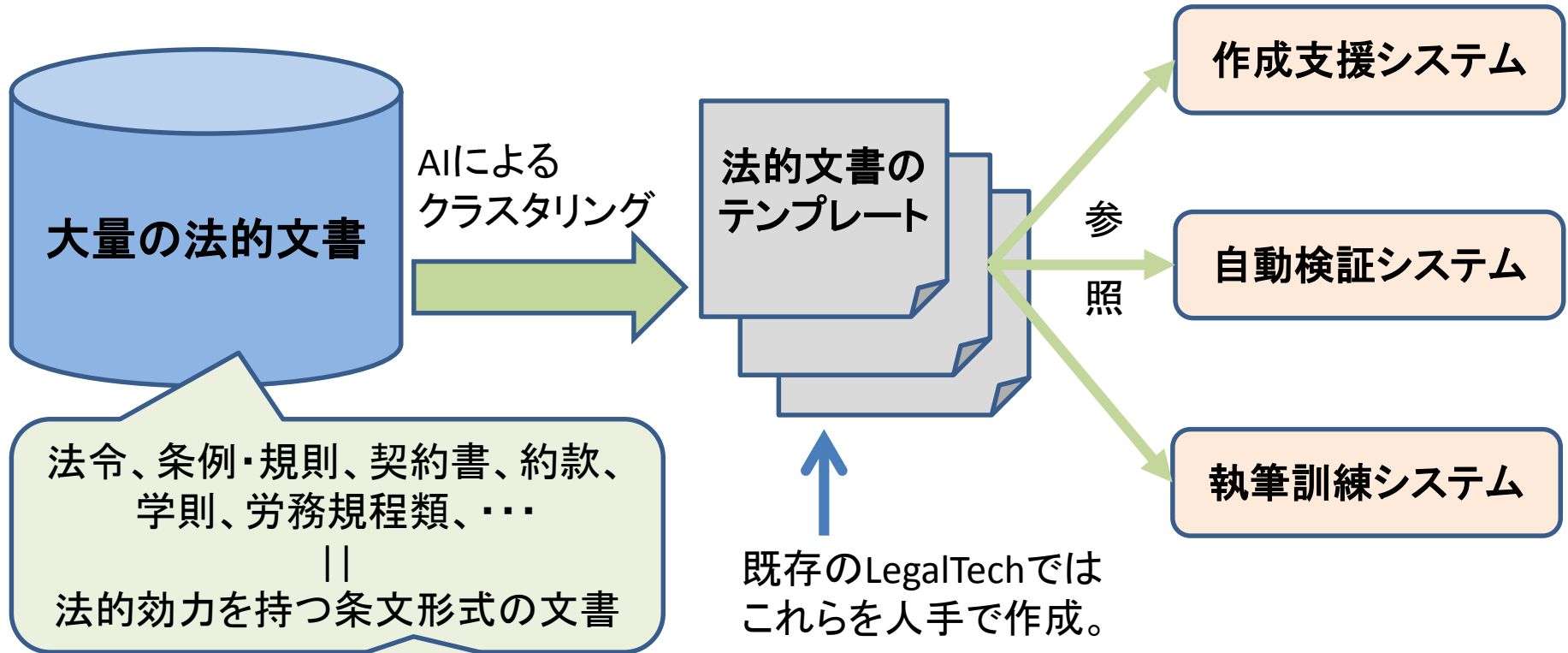
通常は...

テンプレートを利用する。

テンプレート作りが
そもそも大変！

テンプレートがあっても、
カスタマイズが大変！

研究概要



スモールフェーズでは、全国自治体の約9割の条例・規則(＝例規)を用いた自動テンプレート合成のしくみの開発を行う。約120万件(1400万条)。
先行研究で、既に、すべて政府の法令標準形式と同じXML形式に加工したり、形態素解析もできている。(本研究者と共同研究者だけが保持している)

※既に名古屋大学や鹿児島大学から全国の自治体に、立法支援機能付きのWebデータベースの形にして提供され、全国の約半数の770団体の利用実績。

テンプレート作成の方式のイメージ

例規のテンプレート

空欄

施設利用者は
の申請様式より
に届け出る

[候補一覧]

管財課
施設課
観光課
:

いずれかを選べば適切なものになる。
可変部分の可能範囲

[候補一覧]

指定管理者
市長
町長
:

共通部分を抽出す

(抽象化操作)

施設利用者は
管財課
の申請様式より
指定管理者
に届け出る

施設利用者は
施設課
の申請様式より
市長
に届け出る

施設利用者は
観光課
の申請様式より
町長に
に届け出る

クラスタを
変化させると
多様なテンプレート
が合成可能

AIによってクラスタリングされた類似文書群

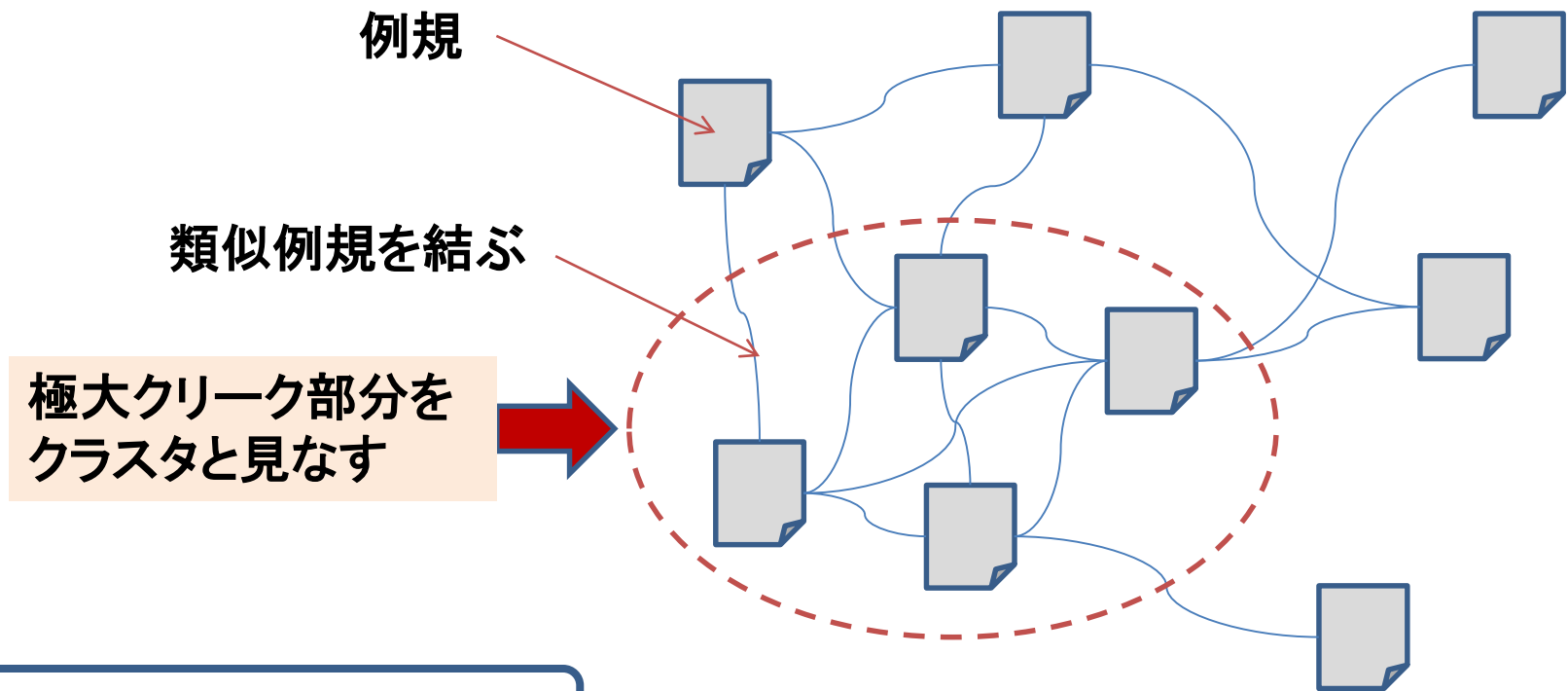
先行研究成果の比較表自動作成機能

	平塚市 環境審議会規則	茅ヶ崎市 環境審議会規則	名古屋市 環境審議会規則
会長及び副会長	第2条 審議会に会長及び副会長1人を置き、委員の互選により定める。2 会長は、会務を総理し、審議会を代表する。3 副会長は、会長を補佐し、会長に事故あるとき又は会長が欠けたときは、その職務を代理する。	第4条 審議会に会長及び副会長1人を置き、委員の互選により定める。2 会長は、審議会の会務を総理し、審議会を代表する。3 副会長は、会長を補佐し、会長に事故があるとき又は会長が欠けたときは、その職務を代理する。	第2条 審議会に会長及び副会長各1人を置く。2 会長及び副会長は、委員の互選によって定める。3 会長は、審議会を代表し、会務を総理する。4 副会長は、会長を補佐し、会長に事故があるとき又は会長が欠けたときは、その職務を代理する。
意見の聴取等	第4条 審議会は、その審議事項について必要があると認めるときは、委員以外の者の出席を求め、その意見を聴くことができる。	第7条 審議会は、その任務を行うため必要があると認めるときは、会議に関係者の出席を求め、その意見又は説明を聴くことができる。	—————
庶務	第6条 審議会の庶務は、環境部環境政策課で処理する。	第10条 審議会の庶務は、環境部環境政策課において処理する。	第7条 審議会の庶務は、環境局において処理する。



運用開始すると、文面まで似ている例規が沢山あった！

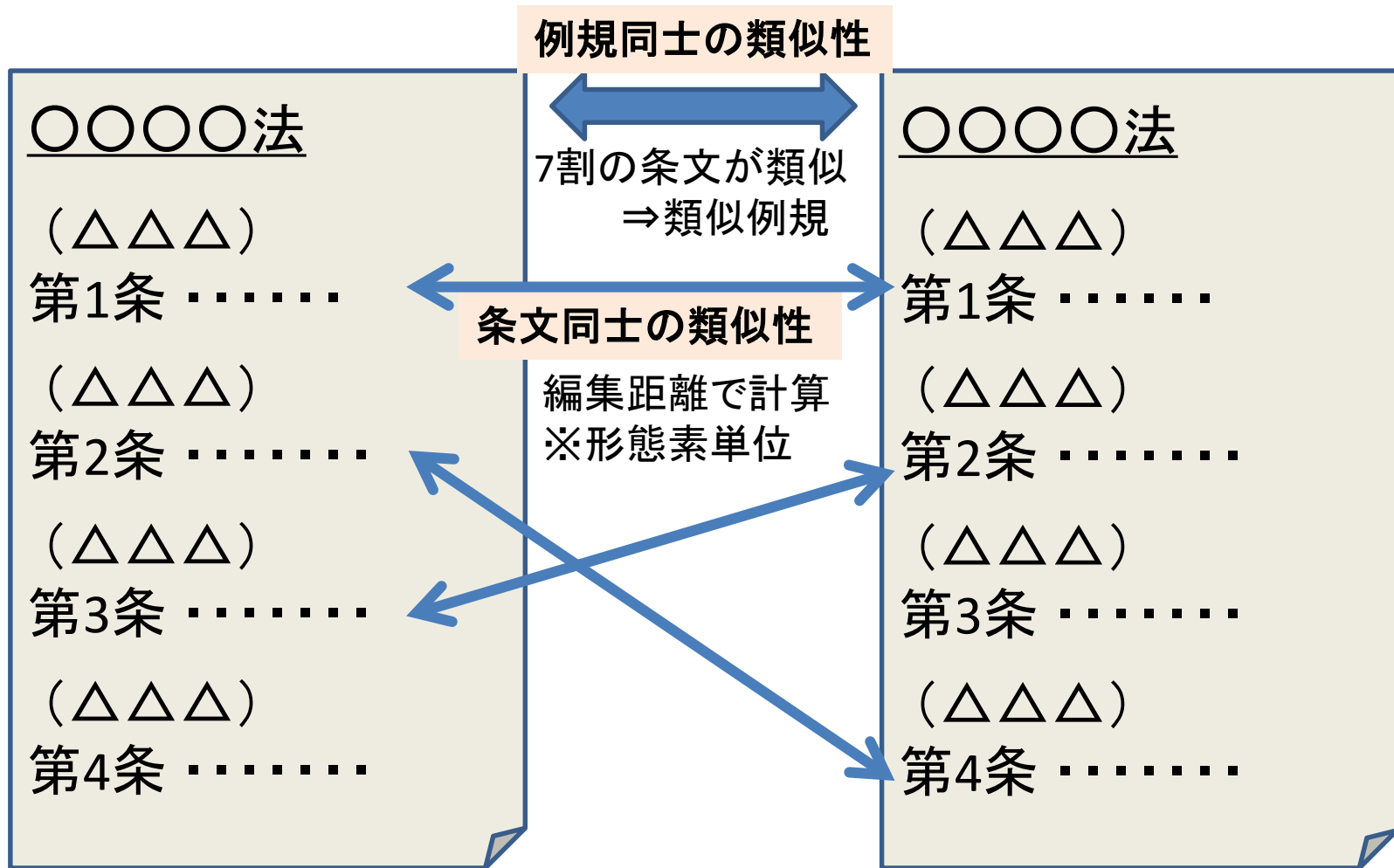
クラスタリング方式: 類似例規のグラフ上のクリーク計算



「クリーク」=
部分グラフが完全グラフ。

その集合内の各要素が他の全ての要素とリンクされている状態。
リンクを2項関係と見れば「同値関係」となる。

類似例規の判定方法



※各自治体の情報なども持っており、さらに、法制度の構造情報(条文相互の関係や位置づけ)も付与して、それらを類似性検出に用いる。

簡単な比較実験

サンプルの23自治体の全例規を使って、
doc2vec方式と本方式(編集距離)とのクラスタリング結果を比較した。

実験データの規模:

全例規数 12,907 件 (全条文数 157,086 条)

例規テキストサイズ合計 47.97 MB

doc2vec方式

計算時間: 約1時間半(ベクトル化含)

結果:

テンプレートにならないクラスタばかり。

本方式(編集距離+クリーク計算)

計算時間: 約2時間(類似度計算含)

結果:

テンプレートになるクラスタばかり。

Python のトピックモデリングのためのライブラリ gensim を利用。

(「Paragraph Vector」モデルに基づく)

分散表現ベクトルの学習パラメタは、ベクトル長=300、エポック数=1000、

ウィンドウサイズ=30、負例サンプル数=5。

各例規の分散表現ベクトルに対して、凝集型階層的クラスタリングを実施。

距離計算手法としてはward法、メトリックとしてはユークリッド距離採用。



やはり、トピックや意味の類似性より、字面の一致度合が大きく効いてくる。

スモールフェーズにおける達成目標

作業

最新の約100万件以上の例規(約1200万条)のクラスタリング方式を検討・実施し、先行研究で予想されていた約600件の大きなクラスタに対して、専門的チェックを行い、法的な典型クラスタを特定し、その特性を形成する属性やその組合せを特定する。属性とは、制度タイプ、地域性などの情報である。

方式の検討

成果物

● 上記クラスタから例規テンプレートを自動作成するシステムを開発する。

例規テンプレートを入力とした3システムの自治体提供:

- 例規の作成を支援するシステム
- 執筆練習のシステム
- 検査システム

法律・自治体関係者による
人海戦術で対応

加速フェーズでの課題

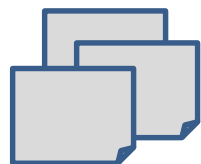
ハイブリッドAI的手法による法令作り支援方式の確立

条例・規則以外にも適用する。

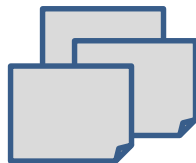
本課題で得られた方式・知見

海外展開

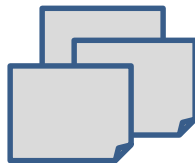
■ ■ ■ ■



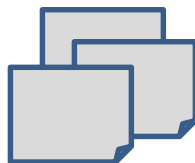
法令



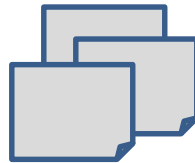
契約書・
約款



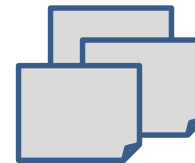
規程類：
・労働規程
・マンション組
合規定



学則



仕様書類：
・関数
・プロトコル
・コマンド



箇条書き形式で1つ
のグループを形成
する文書など