

# 裁判例データ・データベースを用いた 新たな研究手法の開発

法学部法学科 山下 純 司  
計算機センター 久保山 哲 二

## 1 研究の目的

本研究では、①裁判例データをデータベースから効率的に取得する手法の開発、②取得した判決文データを電子的に処理しやすいよう、また人間がより読みやすくなるような工夫、③テキストマイニングや機械学習等の手法による新たな法学研究の可能性を探ることの3点を目的としている。

まず、本研究の背景にある問題意識について説明し(→2)、その後、①の点についての本研究での作業と成果(→3)、次いで②③の部分について本研究で行った作業とその成果、そこで見出された問題点について説明する(→4)。

## 2 研究の背景にある問題意識

### (1) 法学における判例研究

法学の分野では、裁判所の判決を分析する判例研究と呼ばれる研究手法が重要な位置を占める。

裁判所によってくだされる判決には、主文と呼ばれる裁判所が紛争を解決するために出した命令文(例えば、「被告は、原告に金100万円を払え。」)の他に、その主文を導くための理由付けが示される。判決の理由付けには、紛争の事実経緯、原告・被告それぞれの主張、裁判所の判断理由などが記載されており、それを読めばどのような紛争について、裁判所でどのような議論がなされて、どのようなルールを根拠に問題が解決されたかが分かるようになっている。

法律家にとっては、このうち、特にルールの部分が重要である。紛争解決のルールは、法律の条文から解釈によって導かれるが、条文自体は抽象的な文言で書かれているため、具体的な場面でどのように適用するかは、個々の裁判官の判断に任されている。判決文を分析することで、ある法律の条文の、実際の用いられ方が分かるのであり、現在の社会に通用しているルールはどのようなものか、そのルールには問題はないのか、問題があるとすればどのように対処するべきか、といった議論が可能になる

のである。

こうした判例研究は、法学の研究者だけでなく、裁判官や弁護士といった実務家にとっても、新たな紛争を解決するために参考になる先例を発見するという点で、重要な意味を持つものである。

## (2) 裁判例データの電子データベース化

判例研究の重要性から、法学の分野では、判例集と呼ばれる資料が古くから存在している。裁判所によってくだされた重要な判決を収録して定期的に発行されるものであり、最高裁判所が出す公的な判例集のほか、民間業者によって発行される複数の判例集が存在する。さらに、こうした判例集によって提供される裁判例データを、体系的に整理し、データベース化する民間業者が古くから存在する。第一法規社が発行する「判例体系」はその代表的なものであり、昭和39年から判例データベースを提供し判例研究にとって必須のデータベースでありつづけた。当時のデータベースは(当然のことながら)紙製で、紐とじのファイルとなっており、定期的に業者によって入れ替え作業が行われていた。

その後、裁判例データについて、インターネット上のデータベースによる提供が普及し始めた。第一法規社の場合、平成17年から「第一法規法情報総合データベース D1-Law<sup>1</sup>」の提供が開始している。現在では、この他にもエル・アイ・シー社の提供する「LLI 判例秘書<sup>2</sup>」や、TKC 社の提供する「LEX/DB インターネット<sup>3</sup>」など、複数の民間業者による有料データベースの提供がなされている。また、国も、裁判所のホームページを通じて、「裁判例情報<sup>4</sup>」という無料の検索システムを提供している。

このように現在では、判決文を取得するデータベースが整備され、キーワード検索や判例年月日の指定により、関連する裁判例を集めることが容易になった。また、データベースの利用者は、判決文をコンピュータ画面上で閲覧するだけでなく、必要に応じてダウンロードやプリントアウトするなどが可能になっている。

## (3) 電子化のメリットを活かせていない現状

しかし、データベースの電子化が進んだにも関わらず、法学の分野における判例研究は、そのメリットを十分に活かせていないのが現状である。

<sup>1</sup> <https://www.d1-law.com/>

<sup>2</sup> <https://www.lli-hanrei.com/>

<sup>3</sup> <https://lex.lawlibrary.jp/>

<sup>4</sup> [http://www.courts.go.jp/app/hanrei\\_jp/search1](http://www.courts.go.jp/app/hanrei_jp/search1)

キーワード検索を利用すれば、同種の紛争の裁判について、判決文をテキストデータとして大量に入手することは容易なはずである。こうした電子化された裁判例データを、例えば、テキストマイニングの手法で分析することができれば、人の手によって判決文を読み込むのとは異なる新たな視点が見いだされる可能性がある。ところが、こうした研究の試みは、わが国ではほとんど行われていない。

裁判例データの活用が十分に行われない原因の一つは、上記のデータベースによる裁判例データの提供の仕方にあると考えられる。現在提供されている裁判例データのデータベースの多くは、検索してヒットした複数の判決文を、一括ダウンロードできないように作られている。利用者は、一つ一つの判決文にアクセスし、1判決ずつダウンロードをしてくる必要がある。このため、大量の裁判例データを入手することに手間がかかり、機械学習のような手法に必要なだけのデータを集めるのが困難なのである。

また、機械学習を使った裁判例データの分析にとってもう一つの障害となるのが、判決文自体の形式である。判決文には、裁判所によりまとめられた事実概要、原告・被告の主張、裁判所による判断など、さまざまな情報が記載されている。ところが、その記載の仕方は判決文を書く裁判官ごとに異なっており、見出しなども統一されていない。このため、判決文のどこにどのようなことが書いてあるのかは、人間が読んで文脈を理解しないと判断できないのである。このため、判決文から裁判官の主張だけを取り出して分析をするといったことが困難なのである。

本研究の目的として、①裁判例データをデータベースから効率的に取得する手法の開発、②取得した判決文データをテキストマイニングや機械学習で処理しやすいよう、また人間がより読みやすくなるよう整理、③テキストマイニング等の手法による新たな法学研究の可能性を探ることの3点を掲げたのは、以上のような法学の状況があるからである。

### 3 データベースからの裁判例データの一括取得

#### (1) 用いた手法

本研究の第一の目標は、裁判例データをデータベースから効率的に取得する手法を開発することである。このため、本研究では、データベースから必要な裁判例データを自動でダウンロードするプログラムの開発を試みた。

使用言語は、Ruby を使い、watir-webdriver というライブラリを利用することで、裁判所の提供する「裁判例情報」のホームページに自動的にアクセスし、予め指定したキーワードや判例年月日に対応する複数の裁判例データを、ダウンロードしてくるプロ

グラムを作成することができた。なお、プログラムの作成にあたっては、使用の際に裁判所のサーバに過大な負荷をかけないよう、1判決をダウンロードするごとに十分な時間的間隔を置くことを心がけた。

## (2) 結果

実際に使用してみたところ、プログラムは正常に動き、手作業で行うよりもはるかに効率よく裁判例データをダウンロードすることが可能であった。なお、ダウンロードされるデータは PDF 形式であるが、テキストデータが含まれているため、いくつかの処理を行うことで簡単に裁判例データのテキストファイルを入手することができた。

## 4 取得した判決文データの処理

### (1) 分かち書きの必要と問題点

上記のような手法によって取得した判決文データは、そのままではテキストマイニングや機械学習の手法を適用することができない。まず、日本語は英語と異なり文章が単語ごとに区切られていないため、単語数をカウントするなどテキストマイニングで用いられる手法を行うには、いわゆる「分かち書き」を行う必要がある。

そこで、入手した判決文データについて、Mecab や Juman++といった形態素解析プログラムを使って分かち書きを行った。しかし、分かち書きされたデータを用いて、実際にテキストマイニングを行ってみても、法律家の目から見て意味のある分析結果はほとんど得られなかった。

そこで、なぜ分析が上手くいかないのかという観点から判決文データを見直してみたところ、分かち書きに問題があるということが原因の一つと推測された。法学の分野では、日常は使われない専門用語が多い。このため、形態素解析プログラムが元々持っている辞書では、適切に単語が区切られないのである。

たとえば、「権利能力なき社団」という概念が使われる判決文を分析の対象とした場合、形態素解析プログラムは「権利／能力／ない／社団」という4つの単語に分かち書きしてしまう。このため、権利能力なき社団という法律上の概念がどのような文脈で用いられるかといった問題意識で分析をしようとしても、思うような結果が得られないわけである。

このため、法学専門用語を認識できるように、形態素解析プログラムの辞書を充実させる必要がある。手作業による打ち込みで専門用語の辞書づくりを一部実施したが、2016年度は十分な成果が得られず、今後どのような形で辞書の充実を図るかが課題となった。

## (2) 判決文の内容による区分け作業と問題点

すでに述べたように、裁判所の判決文は、裁判所の命令、裁判所によりまとめられた事実概要、原告・被告の主張、裁判所の判断理由など、複数の立場からの意見が混在している。このため、判決文は人間にとっても決して読みやすいものではなく、訓練を受けていないと要旨をとらえ間違える可能性がある。まして、テキストマイニングや機械学習の手法を適用する場合に、裁判所の主張のみを取り上げて分析の対象とするといったことは、非常に困難である。

そこで、本研究では、取得した判決文データについて、教師つき機械学習の手法を適用できるように、正解ラベル付きのモデルを作ることを試みた。これらのデータを正解データとして、機械学習により判決文の分析を行う可能性を探るためである。

まず法学部の学生によって、判決文の文章を①原告の主張、②被告の主張、③裁判官の主張、④前の裁判所(原審)の判断、⑤その他、に色分けしてもらい、人間にとって読みやすく整理されたデータとして作成しなおすことを行った。さらに、これらの色分けされた判決文データを、機械学習で処理しやすいラベル付けされたデータに変換するプログラムを業者に作成してもらい、実際にラベル付きデータを作成した。

もっとも、作業を行って見て、いくつかの課題も判明した。

第一に、学生が作成した色分けデータの精度が、あまり高くないと思われることである。判決文を正確に読むのは訓練が必要であり、法学部の学生といえども容易ではない。また、判決文は裁判官が原告の主張、被告の主張を自分の視点からまとめなおしているため、学生によってそれが裁判官自身の主張なのか、それとも原告や被告の主張なのかの理解にブレが生じてしまった。この点は、色分けの基準をより明確にするなど、工夫が必要であることがわかった。

第二に、上記の方法で作成できたラベル付きデータの数は 500 件程度であるが、機械学習を行うためには、もっと多くのデータが必要だという点がある。どのようにして効率的にデータを集めるかが課題である。

## 5 おわりに

以上のように、いくつかの課題は残ったものの、本研究では裁判例データのデータベースから判決文を一括ダウンロードする手法を開発することができ、また入手したデータを今後どのように活用していくかについての方向性のある程度示すことができたと考えている。わが国の法学研究ではこうした研究手法それ自体が初めての試みであるから、不十分な部分は多々あるが、そうした問題点も含めて今後の課題としたい。