

新聞記事を対象とした刑事事件の性質と量刑との関連性の可視化

Relevance Visualization between Sentencing and Features of Criminal Cases
Retrieved from News Articles

松原 雅文*1
Masafumi Matsuhara

木村 泰知*2
Yasutomo Kimura

渋谷 英潔*3
Hideyuki Shibuki

高丸 圭一*4
Keiichi Takamaru

RZEPKA Rafal*5

村上 浩司*6
Koji Murakami

*1 岩手県立大学

Iwate Prefectural University

*2 小樽商科大学

Otaru University of Commerce

*3 北海学園大学

Hokkai-Gakuen University

*4 宇都宮共和大学

Utusnomiya Kyowa University

*5 北海道大学

Hokkaido University

*6 奈良先端科学技術大学院大学

Nara Institute of Science and Technology

We propose a system visualizing the relevance between sentencing and features of criminal cases retrieved from news articles. In May 2009, so called " lay judge system " will be introduced in Japan. Citizens will join professional judges in sentencing serious criminal cases. Our goal is to construct a system that could support a lay judge helping him or her to acquire knowledge on judicial precedents of similar cases. A person without any law experience chosen for a lay judge will need a lot of time to prepare for a trial; therefore our system could decrease the costs. A user inputs a criminal case into our system. Our system shows a scatter plot of similar criminal cases based on features and sentences. Therefore, it should be easier for a user to analyze the case and help to decide the sentence.

1. はじめに

わが国においては、司法制度改革により「裁判員の参加する刑事裁判に関する法律」が成立し、遅くとも、2009年5月までに裁判員制度が実施される予定である*1。従来の裁判制度において、その審議は職業裁判官のみにより行われていた。しかしながら、裁判員制度においては、一般の国民が裁判員として無作為に選ばれ、裁判官とともに審議に参加することになる。

適切に量刑を判断するためには、事件の性質を正確に把握し、関連する事件との類似性を広く考慮する必要があるものと考えられる。関連する事件を考慮するためには、新聞記事や過去の判例など、法律関係の文書を大量に読む必要がある。しかしながら、このような法律関係の文書は、その内容の理解に向けての研究[石塚 06][山田 08]が行われてはいるものの、一般に専門用語が多く、文法や意味解釈の点において、一般の国民にとっては、内容理解が困難である場合が多い。参審制のように任期制であれば、一般の国民においても、複数回の審議を経験することにより、適格な判断をより迅速に行うことができるようになる可能性がある。しかしながら、裁判員制度においては、事件ごとに人員が新しく選出されるため、ほとんどの場合、選ばれた国民は、量刑判断に伴う作業の経験を持ち合わせていないことになる。よって、一般の国民である裁判員が、適切かつ迅速な量刑判断を行うためには多大な労力を要することが想像され、これを手助けするようなシステムが望まれる。

そのためにも、関連する事件の検索は必須であり、判例文を対象として検索を行う研究[原田 01][江越 05]が行われている。しかしながら、各裁判員が独自にそれぞれの判断を下せるようにするためには、関連する事件の検索の際、一般の国民が容易に入手可能な情報を用いる必要があるものと考えられる。

そこで、われわれは、一般の国民が容易に入手可能な情報である新聞記事を対象として、ある事件に対して量刑を判断する場合を想定し、当該事件を基準として、これと類似する事件を検索し、その関連性を可視化する手法を提案する。量刑判断の一助とするため、可視化の際には、類似度と同時に量刑も考慮し、関連する各事件を出力することとしている。さらに、事件を表現している単語などの素性情報の有無が量刑に与える影響を識別可能とすることを目指している。

連絡先: 松原 雅文, 岩手県立大学ソフトウェア情報学部,
masafumi@soft.iwate-pu.ac.jp

*1 裁判員制度, <http://www.saibanin.courts.go.jp/>

2. 可視化システムの概要

ユーザにより対象となる事件が入力されると、本システムでは、この入力された事件を基準とし、これと関連する事件を新聞記事中より検索し、その類似度および量刑に基づき、各事件を2次元座標上にプロットして出力する。事件における素性を選択し、その有無を区別して表示することにより、関連する事件の考慮を促し、ユーザが行う量刑判断をサポートすることを、本手法では目指している。

3. 処理過程

本手法の処理過程を図1に示す。

3.1 素性の抽出処理

対象となる事件は、自然言語文により入力される。この自然言語文を解析し、これから素性を抽出する。ここでは、素性として名詞単語に着目し、これを形態素解析器「茶筌」[松本 03]を利用して抽出するものとしている。なお、複合名詞は事件の性質を判別する上で、重要な意味を持つものと考えられる。よって、名詞単語の連続はすべてつなげて、1つの素性として扱うこととした。

3.2 関連事件の検索処理

得られた素性情報をもとに、関連する事件を検索する。ここでは、抽出したすべての複合名詞を用い、OR検索を行う。一般の国民が容易に利用できるよう、新聞記事を対象として、事件の検索を行う。具体的には、新聞記事コーパスとして毎日新聞コーパスを用い、検索には、Hyper Estraier*2を利用する。

3.3 類似度の算出

検索された各事件について、基準となる入力事件との関連性を求める。関連性は、各事件における素性ベクトルを用い、これらの類似度として算出する。

入力された対象となる事件に対して行った、素性の抽出処理と同様に、検索された各事件記事において、すべての素性を抽出する。それぞれの記事において、素性の出現頻度に対してtf-idfによる重み付けを施す。そして、これを各事件における素性ベクトルとし、基準となる事件とのコサイン距離を計算する。このようにして、基準となる事件と各事件との類似度を算出する。

*2 Hyper Estraier, <http://hyperestraier.sourceforge.net/>

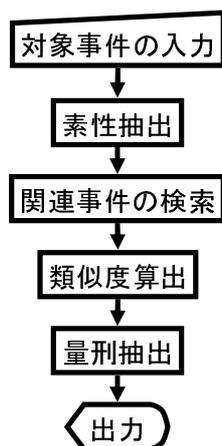


図 1: 処理過程

3.4 量刑の抽出処理

各事件記事から量刑を抽出する。ここでは、単純に「懲役」をキーワードとして、この前後に出現する年月を量刑として、抽出している。例えば、事件記事中の記述が「懲役 15 年を言い渡した」であれば量刑として「15.0」を抽出し、「懲役 2 年 6 月の判決」であれば量刑として「2.5」を抽出する。

3.5 出力処理

求められた類似度と量刑を 2 次元座標上にプロットすることにより、事件の性質と量刑の関連性を可視化する。各事件の量刑を横軸、基準となる入力事件に対する各事件の類似度を縦軸とし、それぞれの事件をプロットする。この際に、素性を選択し、これを区別して表示する。お互いに関連性の強い、十分な数の事件集合であれば、これは、平均的な量刑を中心として正規分布状に事件がプロットされることが期待される。よって、ある素性を選択した場合、それを含む事件集合の分布を、それを含まない事件集合の分布とは区別して認識することが可能であるものと考えられる。これにより、選択した素性が量刑に与える影響を知ることができる。

このようにして、一般の国民が入手可能な情報である新聞記事を対象として、類似事件の関連性を可視化し、ユーザが行う量刑判断をサポートするシステムの実現を目指している。

4. 評価実験

本システムの有効性を確認するため、評価実験を行った。

4.1 実験方法

検索対象となるデータには、毎日新聞 98 年の事件記事を使用した。本システムの入力としては、事件を表す自然言語文を想定しているため、これに相当するものとして、新聞記事コーパス中の事件記事本文をそのまま用いることとした。ここでは、「保険金殺人・詐欺」事件に関する、ある 1 つの事件記事を入力とした。これに対して、素性の有無が量刑に与える影響を確認するため、各素性の有無による、それぞれの事件集合において量刑の平均値を求め、その差を評価した。なお、入力とした事件における、実際の量刑は 15.0 であった。

4.2 実験結果および考察

各素性、当該素性が出現する記事数、当該素性が含まれる事件集合の平均量刑、当該素性が含まれない事件集合の平均量刑、および、その差を評価した結果を表 1 に示す。入力記事に含まれる全素性 67 個のうち、検索された記事数が多いものと少ないものから合計 10 個を抜粋したものである。

表 1 から分かるとおり、「判決」や「求刑」など、量刑に直接影響しない、あるいは、ほとんどの記事において出現し、重要性が低いと思われる素性については、その有無による平均量刑の差は小さいことが分かる。これに対して、素性「保険金目当て」や「保険金殺人」などについては、その有無による量刑の差が大きくなっており、これらが量刑に及ぼす影響が大きい

表 1: 素性の有無による平均量刑の差

素性	記事数	有	無	差
判決	375	6.3	5.9	0.4
求刑	372	5.8	7.9	-2.1
東京地裁	174	4.9	6.9	-2.0
犯行	165	6.9	5.8	1.1
殺害	106	10.0	5.1	4.9
保険金目当て	7	11.8	6.1	5.7
暮らし	6	6.1	6.2	-0.1
保険金殺人	5	14.5	6.1	8.4
杉並区	5	8.6	6.2	2.4
事故死	4	13.2	6.1	7.1

ことが分かる。また、これらは、実際の量刑である 15.0 にも近い値となっていることが確認された。

なお、素性「事故死」については、その有無による差は 7.1 と大きな値となっている。ここで実際に検索された事件は「事故死を偽った」事件であり、素性「事故死」の有無は、その量刑に大きな影響を与えているものと考えられる。しかしながら、本手法においては、素性を画一的に評価していることから、量刑が小さくなると考えられる「誤って事故死させた」事件が存在した場合であっても、その素性は同一とみなされ、同様の結果となってしまう可能性がある。よって、この差別化を図るために、同一の素性であってもその状況を考慮できるよう、タグ情報の付与などを検討する必要がある。

5. おわりに

本稿では、一般の国民が容易に入手可能である、新聞記事データを対象として、ある事件を入力すると、これに関連する事件を検索し、それらの類似度および量刑を可視化して出力する手法を提案した。評価実験の結果、量刑に大きく影響を及ぼし、その重要性が高いと考えられる素性については、その有無により、平均量刑に大きな差が生じることが確認され、これを 2 次元座標上にプロットすることで、人間が行う量刑判断をサポートできる可能性が示唆された。

今後は、実際に 2 次元座標上へのプロットを行い、人間が行う量刑判断に対して与える影響を確認する必要がある。また、動機や計画性の有無などのタグ情報を素性に付加することにより、さらなる精度の向上を目指す予定である。

備考

本研究では、国立情報学研究所主導による NTCIR-7^{*3} の「動向情報の要約と可視化に関するワークショップ」^{*4} [加藤 04] における研究用データセットの一部である、毎日新聞 98 年の記事を利用している。

参考文献

- [石塚 06] 石塚隆男: 法律条文の自動図解化に向けての研究, 電子情報通信学会研究報告 (NLC2005-122), pp.49-54, February 2006.
- [山田 08] 山田将之, 小川泰弘, 外山勝彦: 構文情報付き法律文コーパスの設計と構築, 言語処理学会第 14 回年次大会 (NLP2008), PA2-7, pp.604-607, March 2008.
- [原田 01] 原田実, 鈴木亮: 意味グラフのマッチングによる事故問い合わせ文からの判例検索システム JCare, 情報処理学会研究報告 (2000-FI-061), pp.15-22, March 2001.
- [江越 05] 江越 裕紀, 片上 大輔, 新田 克己: 判例の構造を利用した判例文書検索, 情報処理学会研究報告 (2004-DD-048), pp.1-8, January 2005.
- [松本 03] 松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 高岡 一馬, 浅原 正幸: 日本語形態素解析システム『茶釜』 version 2.3.3 使用説明書, 奈良先端科学技術大学院大学, August 2003.
- [加藤 04] 加藤恒昭, 松下光範, 平尾努: 動向情報の要約と可視化に関するワークショップの提案, 情報処理学会研究報告 (2004-NL-164), pp.89-94, November 2004.

^{*3} NTCIR, <http://ntcir.nii.ac.jp/>

^{*4} MuST, <http://must.c.u-tokyo.ac.jp/>