HOKKAIDO UNIVERSITY

Title	E-042 裁判員支援に向けた新聞記事を対象とした刑事事件の性質に基づく量刑推定手法の提案(自然言語・音声・音楽,一般論文)
Author(s)	松原, 雅文; 木村, 泰知; 渋木, 英潔 他
Description	第7回情報科学技術フォーラム(FIT2008). 2008年9月2日～4日. 慶應義塾大学 湘南藤沢キャンパス,藤沢市.
Citation	情報科学技術フォーラム講演論文集, 7(2), 235-236
Issue Date	2008-08-20
Doc URL	https://hdl.handle.net/2115/63785
Rights	ここに掲載した著作物の利用に関する注意 本著作物の著作権は電子情報通信学会および情報処理学会に帰属します。本著作物は著作権者である電子情報通信学会および情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。
Type	conference paper
File Information	KJ00006439197.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

E-042

裁判員支援に向けた新聞記事を対象とした刑事事件の性質に基づく量刑推定手法の提案

Sentence Estimation based on Features of Criminal Cases Retrieved from News Articles for Supporting Lay Judge

松原 雅文 † 木村 泰知 ‡ 渋木 英潔 *
Masafumi MATSUHARA Yasutomo KIMURA Hideyuki SHIBUKI
高丸 圭一 § ラファウ ジェプカ ¶ 村上 浩司 ‖
Keiichi TAKAMARU Rafal RZEPKA Koji MURAKAMI

1. はじめに

2009 年 5 月に施行される裁判員制度では，裁判員として適切な量刑を判断することが一般の国民に求められる．しかしながら，法的な知識を持たない一般の国民にとって，犯罪行為の反社会性や悪質性，あるいは，犯人への酌量などを判断することはできても，それらの判断を量刑という客観的な値に対応付けることは困難であり，妥当な量刑の見当をつけることすら容易ではない．それゆえ，目安となる量刑の範囲や判断材料となるポイントを示すことは裁判員にとって有益であると考えられる．このような背景から，我々は裁判員の量刑判断を支援するシステムの構築を目指している [1][2].

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，裁判員自身が，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，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．本研究では，当該事件が与えられた場合に，これと関連する事件を検索し，平均的な量刑を推定し，その類似性とともに出力することとしている．そのため，内容理解が困難である文書を対象とした場合には，その提示方法に関しても，さまざまな工夫が必要になると考えられる．

そこで，本研究では，関連する事件の考慮方法に関して焦点を絞り，一般の国民が内容を比較的理解しやすく，かつ，大量のデータ入手が容易である新聞記事を対象としている．類似性を広く考慮するために，関連する事件の検索は必須である．判例文を対象として検索を行う研究 [3][4] が行われているが，本研究では，関連する事件を検索するだけでなく，これらと当該事件との類似度を算出し，これに基づき当該事件の量刑を推定する．この際，量刑の範囲や判断材料となるポイントを示すため，事件の性質を表す素性の有無により，その平均量刑の差も出力するものとしている．このようにして，本手法においては，一般の国民が行う量刑の判断をサポートすることを目指している．

2. 量刑推定手法の概要

ユーザにより対象となる事件が入力されると，本手法では，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する．各事件の量刑を横軸，類似度を縦軸とし，各事件を 2 次元座標上にプロットすることを考える．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，これに基づき，当該事件における平均的な量刑を推定するものとしている．さらに，事件の性質を表す素性を選択し，その有無により，平均量刑の差を提示することにより，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

3.1 素性の抽出

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」[5] 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

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 [1]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．なお，類似性が低い事件はその重要性も低いと考えられることから，ここでは，類似度の値が 0.005 未満のものは，推定量刑算出の対象から除外している．

3.4 量刑の抽出

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．なお，無罪や無期懲役なども含め，ここでの量刑抽出に失敗したものは，その量刑を「0」とし，推定量刑算出の対象から除外している．

†岩手県立大学, Iwate Prefectural University
‡小樽商科大学, Otaru University of Commerce
*横浜国立大学, Yokohama National University
§宇都宮共和大学, Utsunomiya Kyowa University
¶北海道大学, Hokkaido University
‖奈良先端科学技術大学院大学, NAIST

[1]Hyper Estraier, http://hyperestraier.sourceforge.net/

3.5 推定量刑の出力

求められた類似度と量刑に基づき，平均的な量刑を算出し，これを推定量刑として，ユーザに提示する．さらに，各素性の有無により，その平均量刑の差も出力する．

このようにして，類似事件の関連性を考慮して，当該事件の推定量刑を出力し，さらに，素性の有無が量刑に与える影響を提示することにより，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

本手法の有効性を確認するため，処理過程に基づいたシステムを作成し，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，99 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ので，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 7 個を抜粋したものである．表 1 から，素性「保険金目当て」や「事故死」のように，事件を特徴づけ，重要と考えられる素性については，その有無により，平均量刑の差が大きくなっていることが分かる．このことから，重要性が高い素性が量刑に大きな影響を与えていることが確認された．また，これらの平均量刑の値は，他のものに比べて，実際の量刑である 15.0 に近い値となっている．

素性「保険金目当て」を含む事件記事は 7 件が検索されていた．これらの平均量刑は 11.5 となっており，その標準偏差の値は 6.8 となっていた．実際に検索された 7 件の事件記事を表 2 に示す．検索された事件記事それぞれにおける，入力した該当事件との類似度，検索された全事件記事中での順位，および，その量刑を表している．上位 5 位までの事件記事は，入力した当該事件と内容的にも非常に類似したものであり，その量刑も実際の量刑である 15.0 に近い値となっている．しかしながら，6 位以下の事件記事は，内容的にも類似度が低く，その量刑も 15.0 とは大きく離れた値となっている．このため，標準偏差の値も 6.8 と大きな値となった．6 位以下の事件記事については，特に順位において，上位 5 位までの事件記事とは，その差が大きいことが分かる．そこで，この差異を利用し，閾値を設けるなどして，下位のデータを排除することにより，平均量刑の値を実際の量刑の値に近づけることができるものと考えられる．以上のように，類似度の高い事件記事を対象として平均量刑を算出することにより，高い精度での量刑推定を行える可能性が示唆された．

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	744	6.1	6.1	0.0
求刑	702	5.9	6.8	-0.9
東京地裁	314	5.2	6.6	-1.4
殺人	257	9.7	4.7	5.0
保険金目当て	7	11.5	6.1	5.4
事故死	6	10.1	6.1	4.0
不倫相手	4	8.8	6.1	2.7

表 2: 素性「保険金目当て」を含む事件記事

	類似度	順位	量刑
記事 1	0.761	1	20.0
記事 2	0.076	17	15.0
記事 3	0.074	19	15.0
記事 4	0.047	61	13.0
記事 5	0.043	84	15.0
記事 6	0.033	144	1.5
記事 7	0.030	180	1.0

5. おわりに

本稿では，裁判員として一般の国民が行う量刑判断を支援するため，関連する事件を広く考慮して，当該事件の推定量刑を出力するとともに，事件の性質を表す素性の有無が量刑に与える影響を提示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た．また，類似度が高い事件記事を利用することにより，精度の高い量刑推定を行える可能性が示唆された．

しかしながら，事件の性質などを考慮せず，純粋に量刑推定のみを行うのであれば，ある事件集合に対して，各種クラスタリング手法を適用するなどの方法も考えられる．そして，ここで得られた推定量刑を本手法において利用することにより，さらに高い精度で関連する事件を検索できる可能性がある．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，さらなる検索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[2] の「動向情報の要約と可視化に関するワークショップ」[3][6] 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，99 年の記事を利用している．

参考文献

[1] Rafal Rzepka, Hideyuki Shibuki, Yasutomo Kimura, Keiichi Takamaru, Masafumi Matsuhara and Koji Murakami: Judicial Precedents Processing Project for Supporting Japanese Lay Judge System, Workshop on Semantic Processing of Legal Texts, LREC2008, pp.33-41, May 2008.

[2] 松原 雅文, 木村 泰知, 渋木 英潔, 高丸 圭一, Rzepka Rafal, 村上 浩司: 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, 第 22 回人工知能学会全国大会, 3K3-05, June 2008.

[3] 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.

[4] 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.

[5] 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『茶筌』 version 2.3.3 使用説明書, 奈良先端科学技術大学院大学, August 2003.

[6] 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.

[2] NTCIR, http://ntcir.nii.ac.jp/

[3] MuST, http://must.c.u-tokyo.ac.jp/