



HOKKAIDO UNIVERSITY

Title	複数文からなる質問を想定した交通事故判例検索の提案
Author(s)	山田, 浩之; 木村, 泰知; ジェプカ, ラファウ 他
Description	情報処理北海道シンポジウム2009. 2009年10月3日. 旭川市大雪クリスタルホール, 旭川市.
Citation	情報処理北海道シンポジウム講演論文集, 2009, 172-177
Issue Date	2009-10-03
Doc URL	https://hdl.handle.net/2115/63791
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情報処理学会に帰属します。本著作物は著作権者である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	2009-D-6.pdf



複数文からなる質問を想定した交通事故判例検索の提案

山田浩之^{*†} 木村泰知[‡] ジェプカ・ラファウ[†] 荒木健治[†]
(北大情報科学)[†] (小樽商科大学商学部)[‡]

1 はじめに

現在、ウェブ上で判例を検索するデータベースは数多く存在する。例えば、レクシスネクシス・ジャパン[1]やLEX/DB インターネット, TKC 法律情報データベース[2], LLI 判例検索システム[3], そして裁判所のウェブページ[4]などがある。これらのシステムに含まれる判例は主に専門家に利用されているが、一般人が必要とする情報も含まれており、有効な利用が望まれる。

判例とは、過去の裁判の内容を文書としてまとめたものである。判例を参照することにより、今後行われる裁判に対してどのような判決が導かれるかということをおおまかに推測する事が可能となる。そのため、発生した事案に対して類似する過去の判例を検索する事は大変有益な事である。

特に、交通事故に関する判例は交通事故自体がとても身近な事件であり、一般の人々にとっても有益である。具体的に有益な場面としては、交通事故発生後の保険会社との損害額に関する交渉を行う際の妥当性の検証や、請求額の根拠としてなどである。

既存の判例検索システムでは、索引語を用いたキーワード検索が主流である。しかし、キーワードだけの検索では要求した文と合致した検索結果を得ることは難しい。例えば「右折」や「衝突」といった単語を検索のクエリとして用いたとする。その場合、「衝突」という単語は本実験の対象とした判例 1,438 件の判例の内、1,295 件の判例に出現するため判例を絞り込むための有効なクエリとはならない。また、「右折」という単語による検索結果を見てみると、「右折する」という行為以外でも「右折専用レーン」や「右折禁止」、「右折青矢印」等の部分でも幅広く「右折」という単語は用いられているため、そのままクエリを用いたところでは状況に関する絞り込みは望めない

い。

さらに、判例に用いられる単語の中には日常的にはあまり用いられない単語も多い。「原動機付自転車」について日常的には「原付」と省略して使用することが多い。

「原付」での検索結果は 37 件であるのに対して、「原動機付」での検索結果は 216 件にも及んだ。そのため、判例全体から漏れなく検索を行うためには判例で用いられている単語について知ること、すなわち判例に対する知識が必要となる。

このような問題がある一方で、一般ユーザの判例に対する関心は高く、教えて!goo[5]や Yahoo!知恵袋[6]等の Q&A コンテンツへの書き込みが多い。中でも Yahoo!知恵袋では「法律、消費者問題」というカテゴリ内に個別に「交通事故」というカテゴリが用意されており、交通事故に対してユーザの関心が高いことがわかる。しかしながら、これらの Q&A サイトでは質問をしたとしても、すぐに回答が得られないことがある他、専門家からのアドバイスがもらえるとは限らないという問題もある。したがって、自ら検索を行い、即座に求める判例が見つかる判例検索システムが必要となる。

このような点から、本研究では Q&A コンテンツで質問文として投稿されるような複数文を入力として、入力された文と最も合致した交通事故判例を検索するための手法についての提案を行う。

以下、2 節で関連研究と本研究とを比較する。3 節では、類似判例検索実験について述べる。4 節では、実験結果の考察を行う。最後に、5 節では本稿のまとめと今後の課題について述べる。

2 関連研究

現在、判例検索に関する研究は様々なアプローチによって行われている。原田ら[7]の研究では、文章が表す内

^{*}yamada_h@media.cng.hokudai.ac.jp

[†]札幌市北区北 14 条西 9 丁目北海道大学大学院情報科学研究科

容を語が格納されたノードと語間の関係を表すアークからなる意味グラフとして捉え、判例文と問い合わせ文の文章間の内容類似度を算出することで検索を行っている。しかし、これらのシステムでは入力時に詳細な事故の状況を記述する必要があり、ユーザにとっての負担は小さくない。

その他の判例に関する研究としては、江越ら[8]の研究がある。江越らの研究では判例が一つの文書ではなく、役割の異なる複数の文書から構成されていることに着目し、語の偏りによる $tf \cdot idf$ 値の修正等が行われている。しかし、この手法では交通事故判例のような、判例のどの部分に対しても「右折」や「直進」といった事故に関する単語が出現する文書においては語の偏りによって事故の状況を特定するのは困難であると考えられる。また、法律に対しての知識が無いユーザの入力には、判例内で使用される語と同様の語が必ずしも用いられるわけではなく、そういった判例内にはあまり登場しない語を検索対象の判例の単語といかにしてマッチングさせるかという課題に対する考慮がなされていない。

また、自然言語文の入力による検索の研究としては黒橋ら[9]の研究が挙げられる。この研究では、知識ベースを自然言語によって記述し、ユーザの入力質問と知識ベースとの柔軟なマッチングを行うことを処理の基本としている。しかし、この手法では単文による入力しか想定されておらず、またあらかじめ自然言語によって記述された検索対象が用意されているため構文木によるマッチングが容易になっている。

こういった問題から、本研究では交通事故判例内にある「事故態様」という部分に着目する。事故態様とは、どのように事故が発生したのかという客観的な事実を記述した部分である。この事故態様部分を用いて自動的に効率的な交通事故判例検索手法を提案する。

また、本研究では入力文を判例とマッチングさせるために WordNet[10]を用いたクエリ拡張を提案する。WordNet は意味辞書を開発する目的で研究が進められているため、単語の検索を行うことによりその単語の同義語が得られる。その働きによって、入力されたクエリによる検索対象の網羅性をより高めることができる。

そして、本研究は複数文による入力を想定しての実験を行う。また事前にどのような質問が多いといったことが想定できないものを検索対象としているため、人手に

よる正解の準備などの処理は行わずに有効な検索範囲を特定した後に検索を行う。

3 実験

獲得した判例データの名詞に対して3つの手法によって $tf \cdot idf$ による重み付けを行い、入力文から抽出された名詞に対してコサイン類似度を計算して類似判例検索実験を行う。

なお、 $tf \cdot idf$ とコサイン類似度については以下のように定義する。

$$tf = \frac{\text{対象名詞の出現数}}{\text{文書中の総名詞数}}$$

$$idf = \log \frac{\text{総ドキュメント数(1,438件)}}{\text{対象名詞を含むドキュメント数}}$$

$$tf \cdot idf = tf \times idf$$

$$\cos \theta = \frac{w_{q1}w_{i1} + \dots + w_{qn}w_{in}}{\sqrt{w_{q1}^2 + \dots + w_{qn}^2} \times \sqrt{w_{i1}^2 + \dots + w_{in}^2}}$$

w_{qn} : 単語 n を含めば1,

含まなければ0となるベクトル

w_{in} : 単語 n の $tf \cdot idf$ 値

3.1 実験目的

判例検索において最も大きな障害となるのは、文章中に同じ単語が繰り返し記述されるが、それらの単語の役割は判例内の記述されている部分によって異なるという点にあると考える。そのため、判例全体に対する検索ではなく、特定の範囲に対する検索を行うことが判例検索では有効であると仮定した。本手法では、判例全体での検索と事故態様に対する検索という形でこの仮定について検証する。

また、意味的には同様の内容が記述されていたとしても、入力文と判例で使用されている単語が表記的には異なってしまうとマッチングが行われないという問題も考えられる。そのため、本実験では WordNet によるクエリ拡張を行う。特に、一般ユーザの入力と判例との間で顕著であったのが、乗り物についての表記のゆれである。例えば、一般ユーザが「バイク」と記すものは判例内では「自動二輪車」とされることが多く、バイクのままでは大半の判例とマッチングされないことになってしまう。そこで、入力文の乗り物に対する部分を形態素解析で抽出した後、その単語に対して WordNet によるクエリ拡張を行った。上で述べた「バイク」に対しては、「バイク、自

(1) 事故の発生
 以下の内容の交通事故（以下「本件事故」という。）が発生した。
 ア 発生日時 平成一六年一月八日午前八時九分ころ
 イ 発生場所（以下「本件事故現場」という。）
 名古屋市中村区東宿町二丁目一四番地先路線上
 ウ 車両一（以下「被告車両」という。）
 車種 普通乗用自動車
 車両番号 [ナンバー略]
 運転者 被告
 所有者 被告
 エ 車両二（以下「原告車両」という。）
 車種 原動機付自転車
 車両番号 [ナンバー略]
 運転者 原告
 オ 事故態様 被告は、被告車両を運転して北進し、
 本件事故現場である信号機の設置された交差点を右折しようとしたが、
 その際、
 対向方向から南進して同交差点を直進しようとした原告車両の右側面に、
 被告車両を衝突させ、
 原告を負傷させた。

図1 判例と事故態様の例

自分は原付バイク、相手は自動車です。
 信号のないきつい左カーブの所で右折して小
 道に入ろうとしたところ、
 反対車線の前方から来た自動車にぶつかり、倒
 れて軽い怪我をしました。
 双方に前方不注意があったようです。
 このような事故について、過失割合がどのぐら
 いなのか教えていただけないでしょうか。
 よろしくをお願いします。

図2 教えて!gooでの相談文の例

動二輪車、モーターサイクル、オートバイ、単車、自動
 自転車、モーターバイク」が WordNet の検索結果として
 出力されたためそれらを判例検索に用いる。

さらに、事故態様部分を判例から抽出して検索に用い
 る上での問題としては一部の情報が欠落してしまうとい
 うものがある。例えば、事故態様内で「被告車」と記述
 されている部分ではどのような車の種類であったかの情
 報が欠落している。車の種類に関する情報は、事故の状
 況や過失割合を決める上で大変重要である。例えば直進
 車と右折車との事故においても、乗用車同士の事故と、
 乗用車と自動二輪車との事故では過失割合は大きく異な
 る。

そこで、本実験では事故態様中に書かれている車につ
 いて示す部分に対して、判例内の車に関する情報が書か
 れている部分を自動的に照応することで車の情報を捕捉
 する。

上で述べた車の情報付与とクエリ拡張によって、入力
 文が判例の文章と正しくマッチングされるかを検証する。

3.2 実験に用いるデータ

検索対象とする判例は LEX/DB インターネットの「交

表1 実験に用いた教えて!gooの相談文

	質問番号と主な出現名詞
(a)	qa147921(バイク, 右折, 車)
(b)	qa259353(T 字路, バイク, 自転車, 右折)
(c)	qa259357(交差点, 自転車, 左折)
(d)	qa1401434(信号, 交差点, 自転車, 転倒)
(e)	qa1591856(自転車, 左折, 転倒)
(f)	qa2692085(T路地, 右折, 停止, スピード)
(g)	qa674677(バイク, 停止, タクシー, 衝突)
(h)	qa1764342(子供, 自転車, 信号, 無視, 事故)
(i)	qa3110470(直線, 道路, 停車, 追突)
(j)	qa1657503(前方, 救急車, 停車, 後方, 追突)

表2 入力文と正解判例数

	正解判例数
(a)	10
(b)	1
(c)	7
(d)	4
(e)	2
(f)	3
(g)	15
(h)	5
(i)	1
(j)	1

表3 比較実験方法

手法	検索範囲	クエリ拡張
①	判例全体	無し
②	事故態様 or 判例全体	無し
③(提案)	事故態様 or 判例全体	あり (車の情報付与)

通事故判例検索」に含まれる判例のうち、死亡事故を含
 まず、事故に対する明確な答えとなる「過失割合」を含
 むものとする。

過失割合とは、交通事故におけるお互いの過失(不注
 意)の度合いを割合で表したものである。そのため、類似
 した事案に関しては近い割合の過失割合が導き出される。
 この過失割合を含むものを検索対象とした理由としては、
 入力文に対しての答えとなる判例を設定する際の手がか

りとするためである。また、今回死亡事故を対象外とした理由は、Q&A サイトにおける死亡事故に関する相談が極端に少なかったことに起因する。このようにして獲得された判例は 1,438 件となった。

また、実験に用いるため上述のようにして獲得した 1,438 件の判例から「事故態様」部分の自動抽出を行った。事故態様を含む判例は 926 件であった。なお、実際の判例とその中に含まれる事故態様の例を図 1 に示す。

また、入力として用いる文としては教えて!goo の「交通事故」カテゴリ内にあり、「過失割合」について質問しており、かつ専門家と名乗る回答者からのレスポンスが良回答に含まれ、具体的な過失割合について記述されているものを対象とした。

教えて!goo を入力文に選んだ理由は、使用しているユーザは法律に精通していないため Q&A サイトに質問しているのであるから、本研究で想定している判例検索を行うための法律知識を有していないユーザと一致するためである。

「過失割合」についての質問がされているものを対象とした理由は、過失割合によって正解となる判例とのデータセットの作成を行うためである。

また、専門家と名乗る回答者による回答であり、良回答として選ばれた回答を含むものを対象とした理由は、データの信憑性を高めるためである。良回答とは質問者からポイントによるお礼がされているもので、実際に質問者の役に立った回答である。

入力文としては 2008 年 10 月 24 日時点における教えて!goo の「法律」カテゴリに対して、クエリ「交通事故」「過失割合」を用いて検索された 1,323 件のうち、最新の出力結果 1,000 件を収集した。その後、専門家による回答を良回答に含み、かつ具体的に過失割合について記述されているものを抽出した結果、抽出された入力文は 10 件であった。その後、抽出された入力文に対して 1,438 件の判例の中から正解と考えられる判例を手により抽出した。

今回の実験で用いた教えて!goo 入力文の質問番号を表 1 に示す。また、それぞれの入力文に対して正解とした判例の数を表 2 に示す。また、実験に用いた教えて!goo の相談文の例を図 2 に示す。

3.3 実験の前処理

検索対象とする判例に対して形態素解析を行い、名詞

を抽出する。

教えて!goo のデータに対しては、事故に関する客観的な状況の記述がなされている部分を人手で抽出し、その文に対して形態素解析ツール MeCab[11]を用いて名詞を抽出する。

3.4 比較実験方法

提案手法の有効性を確認するため、以下のような実験を行った。

- ①全ての判例の全ての名詞に対して $tf \cdot idf$ を計算し、入力には収集した入力文をそのまま用いる。その後、入力文と判例のコサイン類似度を計算し、上位から出力する。
- ②事故態様を含む判例では事故態様に対して、事故態様を含まない判例に対しては判例全体に対して抽出した名詞の $tf \cdot idf$ を計算し、入力には収集した入力文をそのまま用いる。その後、入力文と判例のコサイン類似度を計算し、上位から出力する。
- ③事故態様を含む判例では事故態様に対して、事故態様を含まない判例に対しては判例全体に対して抽出した名詞の $tf \cdot idf$ を計算し、かつ事故態様を含む判例に対しては車の情報を自動的に付与する。そして、入力文の車に対して WordNet によるクエリ拡張を行う。その後入力文と判例のコサイン類似度を計算し、上位から出力する。

この比較実験については表 3 に簡略化して記述する。

評価式には最高順位と平均順位を用いる。そして最高順位については平均逆数順位も算出する。

最高順位は入力文に対して、複数ある正解として設定した判例のうち最も上位に出力されたものの順位であり、平均順位は複数ある正解として設定した判例全ての出力された順位を平均したものである。

平均逆数順位は、最も上位に出力された正解判例の順位の逆数を 10 組の入力に対して平均した値であり、正解である判例が上位に出力されたシステムほど値が高くなる。そのため、システムの総合的な性能を評価することが出来る。

3.5 実験結果

10 件の入力文(a)~(j)に対して実験を行った結果を表 4、表 5 に示す。なお、縦軸が出力された順位であり、横軸

表 4 それぞれの入力における最高順位

	手法① 判例全体	手法② 事故態様	手法③ 事故態様 &クエリ拡張
(a)	309	36	<u>27</u>
(b)	197	61	<u>49</u>
(c)	<u>183</u>	291	188
(d)	75	24	<u>12</u>
(e)	227	157	<u>146</u>
(f)	<u>11</u>	1046	514
(g)	37	11	<u>5</u>
(h)	12	5	<u>3</u>
(i)	<u>132</u>	702	934
(j)	<u>159</u>	280	199

表 5 それぞれの入力における平均順位

	手法① 判例全体	手法② 事故態様	手法③ 事故態様 &クエリ拡張
(a)	578	426	<u>250</u>
(b)	197	61	<u>49</u>
(c)	<u>545</u>	659	650
(d)	236	<u>191</u>	219
(e)	<u>358</u>	559	625
(f)	<u>530</u>	1166	865
(g)	537	588	<u>495</u>
(h)	<u>18</u>	596	491
(i)	<u>132</u>	702	934
(j)	<u>159</u>	280	199

表 6 最高順位の平均逆数順位

手法	①	②	③
平均逆数順位	0.025	0.039	<u>0.069</u>

が入力した教えて!goo の相談文に割り当てた記号である。

そして、最高順位の平均逆数順位を表 6 に示す。

4 考察

4.1 最高順位について

最高順位の結果を表 4 に示し、最も高い値となったものを太字とアンダーラインで表す。また、平均逆数順位を表 6 に示す。

最高順位については、本稿で提案した事故態様を抽出してクエリ拡張を行う手法が、10 件の入力文のうち 6 件に対して他手法よりもより上位に正解判例が出力された。この 6 件について見てみると、入力文に「バイク」、「自転車」などの表記が見られるものが多かった。そのため、車の情報を事故態様に付与してからクエリ拡張を行う手法は有効であったと考えられる。

しかし、判例全体に対して検索を行ったものよりもかなり低い値を示すものが 2 件ほど存在した。この入力について見てみると、事故の発生した場所が交差点ではあるが、その交差点は一般的な十字路交差点ではなく、T 字路交差点であったことや、訴訟の対象外となっている他の車を避けたことによる事故の発生、そして自動車同士の事故であったことなどが観察された。本実験では交

差点の情報などに関しては事故態様に対して情報の付与を行っておらず、また車同士の事故である場合はクエリの拡張が有効に働かないため最高順位において低い値を出力したと考えられる。また、訴訟外の他の車を避けたことなどの事情についてであるが、そもそも事故態様は客観的な事実のみが記述されるという性質を有している。そのため、事故の当事者間の主張によって争われる論点について書かれることは無い。そのため、そういった情報が述べられている部分が欠落してしまい最高順位の値が低くなったと考えられる。

そして、入力文(i)では事故態様に対して車の情報を付与してクエリ拡張を行った提案手法が事故態様のみを用いて検索を行ったものよりも大幅に低い値を示す結果となった。この事故についてであるが、事故の当事者が両者とも普通の乗用車であったこと、また事故の客観的な状況よりも、どのような状態で停車していたかなど主観的な部分で争われている点から、車の情報付与が逆に不必要な情報となり、値を下げたものと考えられる。

そして、表 6 の平均逆数順位についてであるが、提案手法である手法 3 が全手法の中で最も大きな値を示している。このことによって、全実験結果を考慮して提案手法は 3 つの手法の中で最も正解判例を上位に示しやすいシステムであると考えられる。

4.2 平均順位について

平均順位であるが、判例全体に対して検索を行ったも

のと事故態様を検索範囲としてクエリ拡張を行った提案手法によるものを比べると、提案手法は4件で上回るのみとなった。このことについて、まず事故態様が判例中に含まれていないものが全判例1,438件中に512件ある。提案手法では事故態様の含まれる判例については事故態様に対して、事故態様の含まれていない判例については判例全体に対して $tf \cdot idf$ 値を計算して検索を行う。そのため、入力文に事故の客観的な状況を示す単語が有効に用いられている場合は高い順位となるが、入力文にそれ以外のノイズとなる単語があり、それが $tf \cdot idf$ について大きな値をとっており、事故態様以外の部分にマッチングした場合には提案手法は低い値をとるといった結果となった。

また、クエリ拡張を行わず事故態様を検索範囲とした手法と提案手法での比較では提案手法が7件の値が改善された。値を下げた3件についてであるが、正解判例が1件、2件、4件と少ないものであることを考えると、事故態様と上手くマッチングしなかった場合に、車の情報付与とクエリ拡張がノイズとして働いたということが考えられる。

ただ、平均順位についてであるが、文書検索においてユーザは検索結果全てに目を通すわけではなく、平均順位の若干の違いよりは正解が上位に現れる最高順位を重視するべきものと考えられるため、提案手法の問題点を浮き彫りにするために用いれば十分なものと考えられる。

5 まとめ

本稿では、複数文で記述された事故状況に対して、クエリ拡張を行うことで判例中の単語とマッチングさせ、交通事故判例中の事故態様部分に着目して素性に対する重み付けを行い、事故態様のみでは欠落する情報の補充による類似判例検索手法について述べた。

本論文の提案手法は最高順位において、判例全体に対する $tf \cdot idf$ によるコサイン類似度計算を行う手法よりも10件中6件で優れた値を示し、また平均逆数順位においても他の手法より高い値を示し、その有効性を確認した。

今後の課題としては、入力文からのノイズとなる単語の特定と除去、事故態様で欠落する車に関する情報以外の情報の補充、または事故態様に代わる部分を判例から探し出す手法についての検討、入力された文章の係り受けを考慮した検索、そして、WordNetのみでは行えない判例に特化したクエリ拡張等が必要であると考えられる。

参考文献

- [1] レクシスネクシス・ジャパン,
<http://www.lexisnexis.jp/>
- [2] LEX/DB インターネット TKC 法律情報データベース, <http://www.tkclex.ne.jp/>
- [3] LLI 判例検索システム, <http://www.lli-hanrei.com/>
- [4] 裁判所, <http://www.courts.go.jp/>
- [5] 質問&回答 (Q&A) コミュニティ - 教えて goo,
<http://oshiete.goo.ne.jp/>
- [6] Yahoo!知恵袋, <http://chiebukuro.yahoo.co.jp/>
- [7] 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. (2001)
- [8] 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(セッション1:構造化文書による検索応用)”, 情報処理学会研究報告. (2005)
- [9] 黒橋禎夫, 日笠亘, 藤井綱貴: “入力質問と知識ベースとの柔軟なマッチングに基づく対話的ヘルプシステム” 情報学シンポジウム論文集. (2001)
- [10] Francis Bond, Hitoshi Isahara, Kiyotaka Uchimoto, Takayuki Kuribayashi and Kyoko Kanzaki (2009), “Extending the Japanese WordNet”, In 15th Annual Meeting of The Association for Natural Language Processing, Tottori, C1-4, (2009)
- [11] 工藤拓. MeCab. <http://mecab.sourceforge.jp/>.