

法的判例検索における書誌結合を用いた評価指標の提案

関口 敦仁¹ 吉岡 真治^{2,3}

¹ 北海道大学 工学部

² 北海道大学大学院情報科学院

³ 北海道大学大学院情報科学研究院

sekiguchi.atsuhito.v8@elms.hokudai.ac.jp

yoshioka@ist.hokudai.ac.jp

概要

判例検索タスク COLIEE では、クエリ判例が引用した判例のみを正解とするが、未引用でも有用な判例の存在が指摘されている。そこで本研究では、正解判例を引用し、クエリと書誌結合の関係にある判例に着目する。これらは読むことで正解判例への到達を可能にするため、有用とみなして評価に組み込む手法を提案する。さらに、提案指標を用いて既存の検索結果を再分析し、書誌結合判例の挙動や本提案手法の性質について議論する。

1 はじめに

判例法主義を採用する国々において、過去の判例（先例）に基づく法的判断は不可欠であり、先行判例検索は極めて重要なタスクである。本研究では、COLIEE (Competition on Legal Information Extraction/Entailment) [1] Task 1 を対象とする。本タスクの目的は、与えられたクエリ判例が引用すべき過去の判例を特定することにある。現行のデータセットでは、クエリ判例によって実際に引用された判例（以下、Gold 判例もしくは G ）のみを正解としている。しかし、参加者からは「直接引用されていない判例の中にも、法的判断において有用な判例が存在するのではないか」という指摘がなされている。

そこで本研究では、クエリ判例と「書誌結合 (Bibliographic Coupling)」の関係にある判例に着目し、正解データの拡張を試みる。書誌結合とは、2つの文献が共通の第三文献を引用している状態を指す [2]。先行研究において、書誌結合を用いた類似度は専門家の判断と高い相関を持つことが示されており [3, 4]、クエリ判例と書誌結合関係にある判例（以下、書誌結合判例もしくは BC ）は、クエリ判例と同一の法的論点を共有している蓋然性が高い。

書誌結合判例を利用する利点は、引用関係に基づき機械的に抽出可能であるため、新たな正解データの作成にかかる専門家の判定コストを低く抑えられる点にある。また実用面においては、書誌結合判例を経由することで、その中で引用されている本来参照すべき Gold 判例へ到達可能である。この「ある文書を読むことで、引用先の正解文書にたどり着ける文書もまた正解とみなす」というアプローチは、NTCIR Web 検索タスクにおける「one click distance document model」 [5] と共通する概念である。

本研究では、書誌結合判例を有用な判例とみなすことで正解データを拡張し、それを用いた新しい評価指標を検討する。さらに、既存の COLIEE 参加システムの提出結果を分析することで、現状のシステムにおける書誌結合判例の振る舞いを明らかにし、提案する評価手法の意義について考察する。

2 書誌結合判例を用いた評価指標の提案

書誌結合判例の有用性は、本来引用すべき判例 (Gold 判例) への到達可能性にある。すなわち、同一の Gold 判例への重複到達よりも、各 Gold 判例へ至る経路が少なくとも1つ確保されることが重要である。その点を考慮し、以下の指標を導入する。

クエリ q の引用判例集合 (Gold) を $G_q = \text{Cite}(q)$ 、検索結果を R_q 、判例 d の引用判例集合を $\text{Cite}(d)$ とし、2 指標を定義する。

Extended Precision (拡張適合率)

検索結果 R_q に含まれる判例のうち、Gold 判例自身、あるいは書誌結合判例 (Gold 判例を引用する判例) の割合である。クエリと意味的類似度が高い書誌結合判例を検索することにペナルティを与えないように設計している。

$$\text{Precision}_{\text{ext}} = \frac{|\{d \mid d \in R_q \wedge (d \in G_q \vee (\text{Cite}(d) \cap G_q \neq \emptyset))\}|}{|R_q|} \quad (1)$$

Coverage (カバー率)

Gold判例 $g \in G_q$ のうち、検索結果 R_q から直接あるいは引用経由で到達可能な判例の割合である。単に書誌結合判例の数を問うのではなく、多様な Gold判例への到達ルートをどれだけ網羅的に検索できたかを評価する。

$$\text{Coverage} = \frac{|\{g \mid g \in G_q \wedge (g \in R_q \vee (\exists d \in R_q, g \in \text{Cite}(d)))\}|}{|G_q|} \quad (2)$$

3 既存システムにおける書誌結合判例の分析

本章では、COLIEE2025 Task 1 の上位成績チームの提出結果に基づき、書誌結合判例の挙動を分析する。また、それを通じ本提案手法の意義について考察する。実験データは全判例 1,870 件、クエリ 400 件で構成される。本タスクの提出形式は順位を含まない集合であるため、ある判例が提出データに含まれた回数を「投票数 (Voting Score)」と定義し、システム群の総意としての選好を定量化した。分析対象には、出力の信頼性を考慮して成績上位 4 チーム (計 12Run) を選定し、以下の 2 種類のリストを作成した。

- **統合候補リスト (Integrated Candidates):** 各クエリに対し 1 回以上投票された判例の集合。検索された全候補のプールとする。
- **統合ランキング Top5 (Integrated Ranking Top5):** 投票数が上位 5 件の判例集合。システム群によって関連性が高いと合意された最終出力とみなす。

3.1 提出結果に含まれる書誌結合判例の影響の分析

1 節で述べたように書誌結合判例は引用判例と同程度に意味的に類似している可能性がある。そのため、多くのシステムにおいて、書誌結合判例を引用判例と同様に提出データに含まれる可能性があるのではないかと考え、上位の検索結果である統合ランキング Top5 に含まれる書誌結合判例の数について

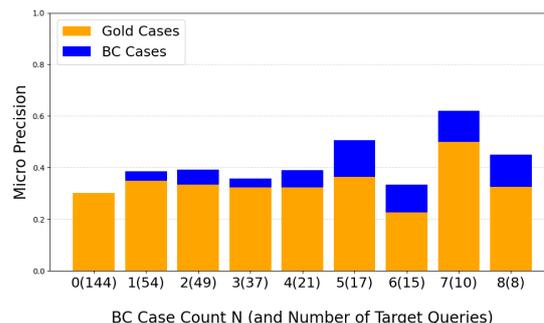


図 1 書誌結合判例 (BC) の保有数に基づくクエリ層別化における適合率の推移。

分析を行った。

まず、統合ランキング Top5 において書誌結合判例が占める割合を Micro Precision を用いて分析した。クエリによって書誌結合判例の保有数は異なる。そこで、書誌結合判例の多寡が評価に与える影響を詳細に分析するため、その保有数 (N) でクエリを分類した。

結果は図 1 が示す通りである。書誌結合判例が多く存在するクエリほど、Top5 内の書誌結合判例占有率が增加する傾向が確認された。なお、この占有率は同数のランダムな判例集合と比較しても有意な差が確認された (同数のランダムな判例集合の場合、Micro Precision は常に 0 付近であった)。

続いて、個々の Gold 判例 g と、それを参照する書誌結合判例集合 $BC(g)$ との競合関係を検証した。分析対象は、対応する書誌結合判例が存在する Gold 判例 641 件である。具体的には、集合 $BC(g)$ に含まれる判例 b のうちいくつかは統合候補リスト (何らかのシステムにより検索された集合) に入っているかを数え、その数で Gold 判例を仕分けた。

分析の結果 (図 2 参照)、書誌結合判例が統合候補リストに含まれるケースは約半数に上った。これは、Gold 判例が検索されるべき場面の多くで、書誌結合判例も有力な候補として挙がっていることを意味する。さらに、中には複数の競合する書誌結合判例を持つ Gold 判例も見られた。すなわち、この結果は、書誌結合判例の全てではないものの、その一部がシステムにとって Gold 判例と区別がつかない判例として存在し、現状のシステムにおいて Gold 判例と同等又はそれ以上の関連性の高い判例であると認識されている実態を表す。

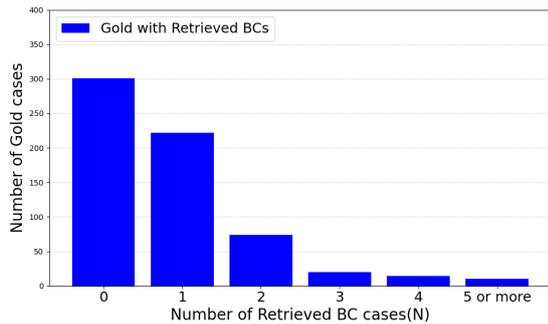


図2 競合する書誌結合判例 (BC) の数による Gold 判例の分布

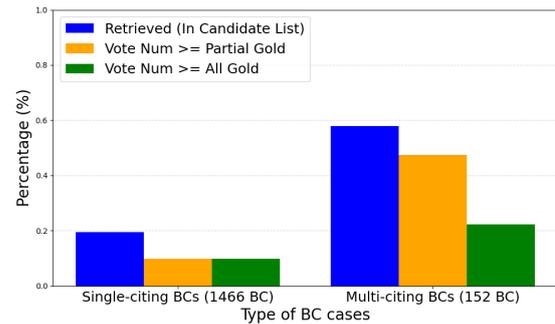


図3 単一参照書誌結合判例と複数参照書誌結合判例における評価指標の比較

3.2 書誌結合判例の共有判例数が与える影響の分析

さらに、書誌結合判例の中には、一つの判例を共有している判例だけではなく、複数の判例を共有している判例が存在する。このような判例は、複数の観点からの類似性を有しているため、さらにクエリとなる判例との類似性が高いことが想定される。この影響を分析するため、Gold判例を1件のみ参照している「単一参照書誌結合判例」と、2件以上参照している「複数参照書誌結合判例」の2グループに分類し比較を行った。

各グループの評価指標として、統合候補リストに含まれた判例の割合を示す「検索候補率 (Retrieved)」, 参照先の Gold判例の少なくとも1件と同等以上の投票数を得た判例の割合を示す「部分的優位率」, および参照しているすべての Gold判例と同等以上の投票数を得た判例の割合を示す「完全優位率」を定義した。

分析の結果 (図3参照), 複数参照書誌結合判例は単一参照書誌結合判例に比べ、検索候補率において約3倍の値を示した。また、完全優位率においても複数参照書誌結合判例は20%を超えており、複数の Gold判例を参照して論じている書誌結合判例は、個々の Gold判例以上にクエリの論点を包括しているとシステムに判断されやすいことが示された。

3.3 考察

前節の分析結果より、テキスト類似度を重視する今回のタスク設定において、書誌結合判例 (とりわけクエリ判例と複数の引用先を共有する判例) は、Gold判例と同等、あるいはそれ以上に重要視される傾向が確認された。この主たる要因は、判例法主義における引用の性質に起因すると考えられる。

「先例拘束性 (Stare Decisis)」 [6] の原則下において、引用は単なる関連情報の提示ではなく、係争中の事案と過去の判例との間に「重要な事実関係 (Material Facts) の共通性」が存在することを示す証拠として機能する [7]。すなわち、共通の先例 (Gold判例) を参照して議論を展開するクエリ判例と書誌結合判例は、必然的に類似した事実や論点を扱っている可能性が高い。一方、引用する判例を選ぶ場合には、より類似度が高い判例が選ばれるのではなく、先例が優先される。しかし、現状のタスク設定では、どちらが先例かを判断する手がかりは提供されていない。したがって、テキスト類似度を重視する現行の検索モデルがこれらを高く評価することは、極めて自然かつ妥当な挙動といえる。

しかし、既存の評価指標では、これらの書誌結合判例はすべて「不正解」として扱われるため、実質的に有益な判例を取得しているにもかかわらず、システムは不当なペナルティを受けていることになる。また、文脈上非常に類似した判例が多数存在する中で、テキストの「文脈的な類似度」のみを頼りに、直接引用した判例 (Gold判例) のみをピンポイントで識別し抽出することは極めて困難である。その結果、既存システムは適合率を維持しようとするあまり、同一の論点を持つ類似判例 (書誌結合判例) を過剰に検索結果に含めてしまう傾向があり、ユーザビリティを損なう恐れがある。

これに対し、本研究で提案した評価手法は、Gold判例への「到達可能性」を考慮し、書誌結合判例を有用とみなすことで、上述の不当なペナルティを解消するものであるといえる。とりわけ「Coverage (カバー率)」の導入は、単一の正解判例をピンポイントで特定することの困難さとリスクを低減させる。これにより、システムは特定の論点 (同一の Gold判例) に関連する書誌結合判例ばかりを列挙する必要

がなくなり、より多様な Gold 判例への到達ルート
を確保すること、すなわち多様な観点を網羅的に取
得するシステム開発を促進する効果が期待される。

4 提案指標による最良提出データの 詳細分析

本節では、COLIEE 2025 Task 1 の最高成績モデル
を対象に、書誌結合判例数でクエリを層別化し、既
存指標と提案指標 (Extended Precision, Coverage) の
変化を検証する。本分析は、新定義下での最良モデ
ルの挙動解明を目的とする。分析の結果、Precision
(Extended Precision) は書誌結合判例数に伴い増加し
た。これは前節までの分析で示唆された通り、当該
モデルが既に多くの書誌結合判例を検索結果に含
んでいたことを裏付けるものであり、正解集合の拡
張が適合数増加に直結したといえる。一方、Recall
(Coverage) の上昇幅は限定的であった。これは定義
を「到達可能性」まで拡張しても、多様な観点を持
つ判例群を十分には捕捉できていないことを示し、
網羅性 (Coverage) に改善の余地が大きいことを示
唆する。

これらの現状を踏まえ、本指標を導入した際の影
響について考察する。既存のシステムでは、類似度
が高く有用な書誌結合判例であっても不正解と判定
されるため、これらを多く検索することは適合率の
低下に直結する。そのため、システムは検索結果の
出力を躊躇 (抑制) する傾向にあり、結果として類
似度の低い正解判例の取りこぼしを招いている可
能性がある。本評価手法によって書誌結合判例の抽
出が精度の向上につながるようになれば、システム
はより積極的に検索候補を広げることができ、類似
度の低い Gold 判例も正解として捕捉できるよう
になると考えられる。さらに訓練時においても、現
状では負例とされている意味的に類似した書誌結
合判例を正例として扱えるようになるため、より一
貫した類似度判定の学習が可能になると期待される。

5 おわりに

本研究では、法的判例検索タスクにおいて、参加
者から指摘されていた「直接引用でなくても有用な
判例が存在する」という課題に対し、書誌結合判
例を正解データとして拡張する新たな評価方法を
提案した。また、単に適合判例を増やすだけでな
く、本来参照すべき判例への到達可能性を評価す
る「Coverage」指標を導入した。本提案手法は、従来の

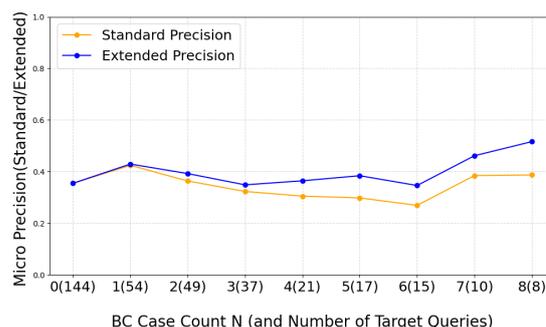


図4 書誌結合判例 (BC) 数ごとの Micro Precision/Extended Precision の推移

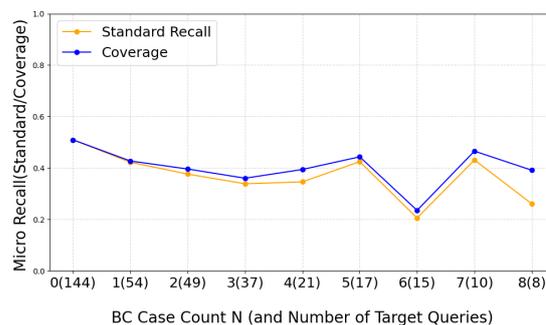


図5 書誌結合判例 (BC) 数ごとの Micro Recall/Coverage の推移

評価指標下において既に多くの書誌結合判例が検索
されていた事実を踏まえると、「良い」とされるシス
テムの定義を根本から覆すものではない。むしろ、
現状の優秀なシステムが文脈から有益とみなして取
得した判例を正當に評価し、その能力を「Coverage」
という新たな観点から包含・拡張する評価枠組みで
あるといえる。今後は、この Coverage を直接的に最
適化するような検索モデルの構築や、書誌結合判例
以外の有用な判例の定義についても検討を進める必
要がある。

謝辞

本研究で用いたデータは COLIEE の Task1 のオーガナイザーから提供されたものである。データの処理に協力して頂いたアルバータ大学の Calum Kwan 氏には特に感謝の意を表します。

参考文献

- [1] Randy Goebel, Yoshinobu Kano, Mi-Young Kim, Juliano Rabelo, Ken Satoh, and Masaharu Yoshioka. Summary of the competition on legal information, extraction/entailment (coliee) 2023. In **Proceedings of the nineteenth international conference on artificial intelligence and law**, pp. 472–480, 2023.
- [2] Maxwell Mirton Kessler. Bibliographic coupling between scientific papers. **American documentation**, Vol. 14, No. 1, pp. 10–25, 1963.
- [3] Sushanta Kumar, P Krishna Reddy, V Balakista Reddy, and Aditya Singh. Similarity analysis of legal judgments. In **Proceedings of the fourth annual ACM Bangalore conference**, pp. 1–4, 2011.
- [4] Paheli Bhattacharya, Kripabandhu Ghosh, Arindam Pal, and Saptarshi Ghosh. Methods for computing legal document similarity: A comparative study. **arXiv preprint arXiv:2004.12307**, 2020.
- [5] Koji Eguchi, Keizo Oyama, Akiko Aizawa, and Haruko Ishikawa. Overview of the informational retrieval task at ntcir-4 web. In **NTCIR**, 2004.
- [6] William O Douglas. Stare decisis. **Columbia Law Review**, Vol. 49, No. 6, pp. 735–758, 1949.
- [7] Arthur L Goodhart. Determining the ratio decidendi of a case. In **Precedents, Statutes, and Analysis of Legal Concepts**, pp. 21–43. Routledge, 2013.