

反証センテンスの提示による信憑性指向のウェブ検索支援

山本 祐輔^{1,a)} 田中 克己^{1,b)}

受付日 2012年9月20日, 採録日 2012年12月18日

概要: 本稿では, 信憑性指向のウェブ検索を支援するために, 新しいタイプのクエリ推薦手法を提案する. 提案システムはユーザからクエリが入力されると, クエリに関するセンテンスのうち, ウェブ上で反証されているセンテンスをリアルタイムに抽出・収集する. その後, 収集された被反証センテンスのウェブ上における典型度, およびクエリとの関連度を計算する. 最終的に, 提案システムは典型度および関連度が高い反証センテンスの上位 N 件を, 検索中のユーザに提示する. 既存のクエリ推薦手法は, ユーザの検索意図に合致するウェブページの検索を効率化することに焦点が当てられている. それゆえ, ユーザが特定の意見・記述の信憑性を検証するためにウェブ検索を行う場合, 既存のクエリ推薦手法によって提示されたキーワードが, 信憑性判断の手がかりを探すために役に立つとは限らない. また, たとえ検索中のウェブ情報の中に信憑性が疑わしいものが存在しても, ユーザが情報の信憑性についてあまり注意を払っていないければ, それに気付くことは難しい. 本稿で提案する反証センテンス手法により, 信憑性が疑われるセンテンスへユーザの注意を喚起すること, 信憑性を意識しながらウェブ検索を行うことが可能となる.

キーワード: 情報の信憑性, ウェブ検索, クリティカルシンキング, クエリ推薦, 反証提示

Disputed Sentence Suggestion for Credibility-oriented Web Search

YUSUKE YAMAMOTO^{1,a)} KATSUMI TANAKA^{1,b)}

Received: September 20, 2012, Accepted: December 18, 2012

Abstract: We propose a novel type of query suggestion to support credibility-oriented Web search. When users issue queries to search for Web pages, our system collects disputed sentences about queries from the Web. Then, the system measures how typical and relevant each of the collected disputed sentences are to the given queries. Finally, the system suggests some of the most typical and relevant disputed sentences to the users. Conventional query suggestion techniques focus only on making it easy for users to search for Web pages matching their intent. Therefore, when users search for Web pages to check the credibility of specific opinions or statements, queries suggested by conventional techniques are not always useful in searching for evidence for credibility judgments. In addition, if users are not careful about the credibility of information in the Web search process, it is difficult to be aware of the existence of suspicious Web information through conventional query suggestions. Our disputed sentence suggestion enhances users' awareness of suspicious statements so that they can search for Web pages with careful attention to them.

Keywords: information credibility, Web search, critical thinking, query suggestion, dispute suggestion

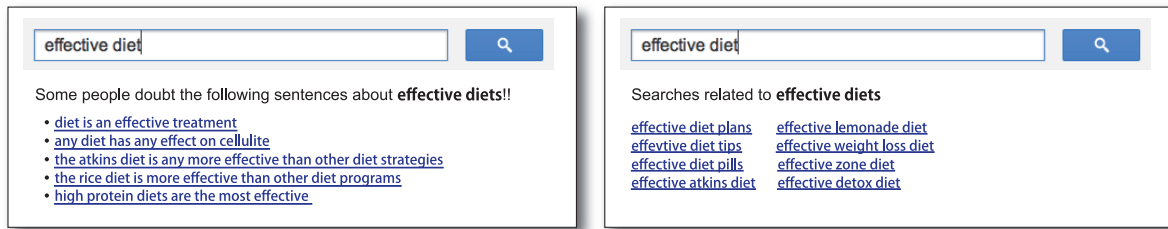
1. はじめに

今日, 検索エンジンや様々なウェブサイトを通じて, 欲しい情報をいつでも簡単に取得することができる. しか

し, ウェブ情報の信憑性が問題になりつつある [1], [2].

ウェブ情報の生成・公開プロセスは既存のマスメディア情報と異なり, 情報の質は保証されていない. それゆえ, ウェブ上には, ユーザに不利益を被りかねない不正確で信憑性の低い情報が多数存在する. たとえば, Sillence らの報告によると, ウェブ上には 20,000 以上の医療関係のウェブサイトが存在するが, その大半は医療専門家のチェックを受けていない. その一方で, ユーザの多くはウェブ情報を

¹ 京都大学
Kyoto University, Kyoto 606-8501, Japan
a) yusuke@hontolab.org
b) tanaka@dl.kuis.kyoto-u.ac.jp



(a) 反証センテンスの提示

(b) 従来型のキーワード推薦

図 1 (a) 反証センテンスの提示と (b) 従来型のキーワード推薦

Fig. 1 (a) Disputed sentence suggestion and (b) conventional keyword suggestion.

ある程度信用しているとの報告もなされている [2]. 特に, ウェブ検索エンジンに関しては, それを利用する 50%以上のユーザは検索エンジンの検索結果に含まれるウェブページはある程度信用できると考えている, という調査結果が報告されている.

ウェブ情報の信憑性が問題になりつつある中, 既存のウェブ検索エンジンは, 依然としてユーザの検索意図に合致する情報の検索に焦点が当てられている [3], [4]. また, ウェブ検索エンジンの返す検索結果には, 情報の信憑性を判断するための手がかりはほとんど示されていない. たとえば, あるユーザが「効果的なダイエット」に関するウェブページを検索しているとする. 今日, ウェブ上にはダイエットに関するページが多数存在するが, アトキンスダイエット*1の例のように, その信憑性が疑われているにもかかわらず, 効果的なダイエットとして検索結果の上位のウェブページに記載されていることが多々ある. 通常, ウェブ検索エンジンが返す結果には, タイトルとスニペット, URL しか含まれていない. この程度の情報量では, 疑わしいウェブ情報の信憑性を確認したり, 信憑性の高いウェブページを探し出したりすることが難しい. ユーザがウェブ情報の信憑性を意識していなければ, 最悪の場合, 誤った情報を鵜呑みにしてしまう可能性もある. このような事態を防ぐためにも, ウェブ情報の信憑性に対するユーザの意識を高め, ウェブ情報の信憑性判断を支援するツールが必要となる.

本稿では, 信憑性指向のウェブ検索を支援するために, ウェブ検索中のユーザに対して, 疑わしいウェブ情報に警鐘を促す新しいタイプのクエリ推薦システムを提案する. 提案システムは, ユーザから検索クエリを受け取ると, クエリに関する情報に対する反証センテンスを提示する. 図 1 (a), (b) に, ユーザが *effective diet* というクエリを入力した際の, 提案システムと既存のウェブ検索エンジンの振舞いを示す. 既存のウェブ検索エンジンでは, 検索クエリに対して, *effective diet plans* や *effective atkins diet* といったキーワードが推薦されている (図 1 (b)). この例のように, 既存のクエリ推薦手法では, クエリの背後にあ

る検索意図を明確化し, 検索意図に合致するウェブページの検索を支援することに焦点が当てられている. 提示されるクエリがキーワードであることも多い. 一方, 提案システムでは, “*diet is an effective treatment*” や “*the atkins diet is not any more effective than other diet strategies*” といったクエリに関連するセンテンスに対して, 反証がなされていることを示唆している (図 1 (a)).

ウェブ検索プロセスにおいて, 反証センテンスの提示は既存のキーワードクエリの提示に比べ, 以下のようなメリットがある:

- ウェブ検索時にユーザが情報の信憑性をほとんど意識していない場合でも, 信憑性が疑わしいウェブ情報の存在をはっきりと認識することができる.
- ウェブ情報の信憑性を判断する手がかりとして, 反証情報を検索するためのクエリを知ることができる.
- 反証センテンスはキーワードの羅列ではなく自然言語で表現されているので, その意味を直感的に理解しやすい.

本稿の構成を以下に記す. 次章では, 関連研究について述べる. 3 章では, ウェブから反証センテンスを収集しランキングする方法を示す. 4 章では, ウェブ検索時におけるユーザの信憑性判断支援に反証センテンスの提示がどの程度有効かを, いくつかのクエリ推薦手法と比較しながら評価する. 最後に, まとめと今後の課題を述べる.

2. 関連研究

2.1 クエリ推薦

クエリ拡張 [5] やクエリ置換 [6], [7] を含め, クエリ推薦に関する研究はこれまで多数行われてきた. Cui らは, 文書空間とクエリ空間の違いを考慮してクエリ拡張を行う手法を提案している [5]. Boldi らは, ユーザのクエリ修正のパターンを一般化 (generalization), 詳細化 (specialization), 平行移動 (parallel movement), エラー修正 (error correction) の 4 種類に分類する手法を提案している [6]. Kotov らは, 大量のウェブページに自然言語処理を適用し, それぞれの内容を問う疑問文を生成し, それらをウェブ検索中のユーザに提示することで, 対話形式でユーザを目的のページへと導くシステムを提案している [7]. これらの手

*1 アトキンスダイエット: <http://ja.wikipedia.org/wiki/アトキンスダイエット>

法は、ユーザの検索意図とクエリ間のギャップを解消し、ユーザに満足のいく検索結果を提供することを目指している。それゆえ、これらの手法を用いて何らかのクエリが提示されても、信憑性が疑わしいウェブ情報に対して注意を促したり、信憑性判断を行うための手がかり情報を収集したりするのを支援をすることは難しい。

2.2 ウェブ情報の信憑性評価

特定のウェブコンテンツの信憑性を何らかの観点から評価する手法は、いくつか提案されている。特に、Yahoo! Answer^{*2} や Twitter^{*3}といったソーシャルネットワーク上のコンテンツの信憑性を評価する研究はさかんになりつつある。Suryantoらは、Q&A サイト上の回答者の専門性を分析し、投稿された回答の信憑性を評価するための手法を提案している [8]。Castilloらは、Twitter 上でのニュースに関するつぶやきを解析し、拡散中のニュース情報の信憑性レベルを自動的に判定する手法を提案している [9]。Chiaらは、ウェブアプリケーションの安全性を調べるうえでウェブ上の集合知がどの程度信頼できるかについて検証を行った [10]。ウェブページやウェブ検索結果の信憑性を、様々な観点から評価するためのシステムもいくつか提案されている [11], [12]。これらのシステムでは、人気度、コンテンツの典型度、コンテンツの新鮮さ等、信憑性判断時に重要となる尺度に沿ってウェブ検索結果がスコアリングされ、各検索結果に可視化されたスコア情報が提示される。信憑性の高いウェブ情報をユーザ自らが取捨選択して検索したい場合、ウェブ情報の信憑性をスコア化することは有益であろう。

2.3 ウェブ情報の信憑性判断支援

ユーザ自ら信憑性判断を行うことを支援するためのシステムに関する研究も少なからず行われている。PirolliらはWikipediaの記事の編集履歴を可視化するシステム WIKI-DASHBOARDを開発し、システムがユーザの信憑性判断にどのような影響を与えるかを分析している [13]。Ennalsらは、閲覧中のウェブページ中に信憑性が疑わしいセンテンスがあったときに、それにハイライトをするシステム DISPUTE FINDERを提案している [14]。DISPUTE FINDERによってハイライトされる信憑性が疑わしいセンテンスは、あらかじめデータベースに登録されたものに限定されているため、クエリによって異なる検索結果が提示されるウェブ検索システムではうまく機能しない。一方、我々が提案するシステムは、いったんクエリが入力されると反証センテンスをオンデマンドに取得するため、ロバストである。また、提案システムは入力されたクエリに関する反証センテンスの一覧を、網羅的に確認することができる。Kawahara

らや Yamamotoらは、信憑性が疑われるセンテンスを入力すると、その信憑性を調べるために有用な比較センテンスをウェブから抽出する手法を提案している [15], [16]。

3. 反証センテンスの検索

本章では、入力されたクエリに関する反証センテンスをウェブから検索し、それをユーザに提示する手法について述べる。本稿では、センテンス d についてウェブ上で反証がなされている場合、 d を被反証センテンス (例: “*Atkins diet is effective*”), 反証を示唆する接頭句を d に付けたものを反証センテンス (例: “*Some people doubt that Atkins diet is effective*”) と呼ぶことにする。以下に提案システムの動作フローを記す:

- (1) 入力されたクエリに関する被反証センテンスの候補をウェブから収集する。
- (2) 被反証センテンスの典型度とクエリに関する関連性を考慮することで、被反証センテンスの候補をスコアリングする。
- (3) ランキングされた被反証センテンスの上位 k 件に反証の存在を示す接頭句を付け、それらを反証センテンスとしてユーザに提示する。

3.1 ウェブからの反証センテンス抽出

ある内容について反証しているセンテンスを網羅的に収集する方法の1つとして、大規模な文書群に対して深い自然言語処理を適用することが考えられる。しかし、この種の手法は計算コスト、時間コストがかかる。この問題を解決するために、Ennalsらは言語パターンを用いてウェブから反証センテンスを収集する手法を提案している [17]。本項では、Ennalsらの手法を応用し、クエリに関する反証センテンスをオンデマンドに収集する手法について述べる。

Ennalsらの手法では、まず、センテンス d が反証されている可能性を示す手がかり表現 (以下、反証手がかりと呼ぶ) (例: “*the misconception that d*”, “*it is not true that d*”) の集合を事前に用意する。その後、図 2 のように、反証手がかりをフレーズクエリとしてウェブ検索エンジンに入力し、得られたウェブ検索結果から手がかり表現の直後に出現するセンテンスを抽出する。

Ennalsらの研究では、(被)反証センテンスをオフラインで網羅的に収集することを目的としていた。一方、本研究ではユーザが入力したクエリに関連する(被)反証センテンスを、リアルタイムで収集することを目的としている。この目的を達成するために、提案システムではEnnalsらが利用した54の反証手がかり表現のうち15個に限定して利用する。利用する反証手がかり表現を表 1 に記す。本章では、事前に選んだ15の反証手がかり表現を用いてウェブから被反証センテンスを抽出する手法を提案する。手順は以下のとおりである:

*2 Yahoo Answers: <http://answers.yahoo.com/>

*3 Twitter: <http://twitter.com/>

表 1 ウェブから被反証センテンスを収集するための手がかり表現
Table 1 Linguistic patterns to collect disputed sentences from the Web.

反証手がかり表現
“no proof that”, “mistakenly believe that”, “no evidence that”, “lie that”
“it is not true that”, “it is not clear that”, “it is unlikely that”, “it is wrong that”
“it is not to say that”, “it is denied that”, “into believing that”
“misconception that”, “myth that”, “it is speculated that”, “it is delusion that”

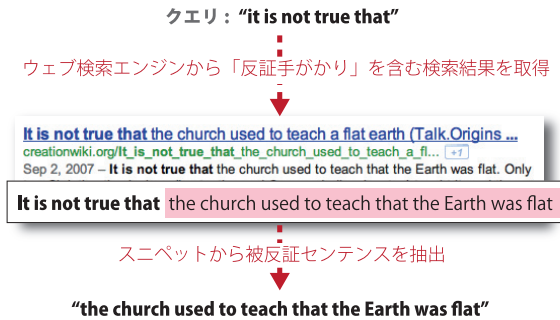


図 2 ウェブ検索結果のスニペットから被反証センテンスを抽出する例

Fig. 2 Example of disputed sentence extraction from snippets of Web search results.

- (1) ユーザから入力されたクエリと反証手がかり表現のペアを Yahoo! Search Web API*4 に入力する*5.
- (2) 区切り文字を用いて、ウェブ検索結果中に含まれるスニペットをセンテンスに分割する (区切り文字の例: “.”, “,”, “!”, “?”, “and”, “but”).
- (3) 分割されたセンテンスから反証手がかり表現の直後に出現するセンテンスを、被反証センテンスの候補として抽出する.
- (4) 被反証センテンス候補が代名詞, 接続詞, 前置詞のいずれかで始まる場合, 被反証センテンスの候補から取り除く.

3.2 被反証センテンスのランキング

抽出した被反証センテンス候補の中から、ウェブ情報の信頼性判断支援に有用な反証センテンスを選択・提示するために、被反証センテンス候補をランキングする.

今、ユーザが入力したクエリを q , 抽出された被反証センテンスを d とする. 提案システムでは、クエリ q との関連性を考慮して、収集された被反証センテンス $D = \{d_1, d_2, \dots, d_n\}$ のランキングを行う. クエリ q と被反証センテンス d との関連性の評価には、以下の式で表されるクエリ尤度モデルを用いる [3]:

*4 Yahoo! Search Web API:

<http://developer.yahoo.co.jp/webapi/search/>

*5 たとえば、クエリとして *global AND warming AND cause* が入力された場合、提案システムは拡張されたクエリ *global AND warming AND cause AND “it is not true that”* を API に入力する.

$$p(d|q) \propto p(d) \prod_{t \in q} ((1 - \lambda)p(t|M_c) + \lambda p(t|M_d)) \quad (1)$$

$$p(q|M_d) = \prod_{t \in q} \frac{tf_{t,d}}{L_d} \quad (2)$$

ここで M_d は被反証センテンス d の言語モデル, $tf_{t,d}$ は d における語 t の出現頻度, L_d は d に含まれる語の数を意味する. λ は重みパラメータで, $0 < \lambda < 1$ の値をとる. M_c は被反証センテンスの集合 D から構築される言語モデルである.

クエリ尤度モデルでは、文書の事前確率 $p(d)$ はすべての文書 d において一様であると扱われることが多い. しかし、提案システムでは、収集された被反証センテンスがウェブ上でどの程度典型的なものかあるかを評価することで $p(d)$ を計算する.

ウェブ上における被反証センテンスの典型度を評価する方法の1つとして、センテンスが何回出現するかを数え上げることが考えられる. しかし、“a mobile phone is bad for your health” や “mobile phones are a health risk” のように、同じ内容が異なるセンテンスで表現されることも多い. それゆえ、単にセンテンスの出現頻度を調べるだけでは、被反証センテンスの典型度を正確に評価することは難しい. この問題を解決するために、本稿では *LexRank* アルゴリズムを用いる [18]. *LexRank* アルゴリズムでは、テキストコンテンツをノードとし、テキスト間の類似度を枝の重みとするグラフを生成する. その後、グラフ中におけるテキストノードの中心性をテキストの典型度として計算する. 提案システムでは、*LexRank* アルゴリズムを用いて各被反証センテンス $d \in D = \{d_1, d_2, \dots, d_n\}$ の典型度を計算し、それを $p(d)$ として評価する. 具体的な計算は以下のように定式化される:

$$\mathbf{t} = \alpha \mathbf{S}^* \times \mathbf{t} + (1 - \alpha) \mathbf{p}, \quad \text{where } \mathbf{p} = \left[\frac{1}{n} \right]_{n \times 1} \quad (3)$$

ここで、 \mathbf{S}^* は、 i 行 j 列目の要素 $S_{i,j}$ が被反証センテンス d_i, d_j 間の類似度 $\text{sim}(d_i, d_j)$ を表す行列 \mathbf{S} を各列ごとに正規化した行列を意味する. \mathbf{t} は、被反証センテンス $d \in D = \{d_1, d_2, \dots, d_n\}$ の典型度 $t(d)$ を要素とするベクトルを意味する. 提案システムでは、 $t(d)$ を $p(d)$ として計算する.

実際には、被反証センテンス $d \in D$ の $p(d|q)$ は、以下の手順で算出する:

- (1) ウェブから被反証センテンス群 $D = \{d_1, d_2, \dots, d_n\}$ を収集する (3.1 節参照).
- (2) $d \in D$ について, 式 (2) に従い $p(d|M_c)$ と $p(q|M_d)$ を計算する. なお, 計算の過程でストップワードは無視する.
- (3) 被反証センテンスの特徴ベクトルを生成する. ベクトルの要素はセンテンスを構成する語であり, 語の重み付けには $tf-idf$ 値を用いる.
- (4) $d \in D$ について典型度 $t(d)$ を計算し, それを $p(d)$ とする. 典型度の計算では, スコアが収束するまで, 式 (3) を用いて \mathbf{t} を再帰的に計算する.
- (5) $d \in D$ について, ステップ (2), (4) の結果と式 (1) を用いて $p(d|q)$ を計算する.

4. 評価実験

ウェブ検索プロセスにおけるユーザの信憑性判断支援に, 提案手法がどの程度有効かを評価するために評価実験を行った. 実験では, 以下の3つの観点から反証センテンスの提示の有効性を評価した.

- **関連度**: 入力されたクエリに対して, 反証センテンスはどの程度関連があるか?
- **警鐘度**: 反証センテンスが提示された後, ユーザはクエリに関する検索結果にどの程度注意を払うようになるか?
- **有用度**: 検索結果の信憑性を判断する手がかりを検索するのに, 提示された反証センテンスがどの程度有用であるか?

実験では, 提案手法 **Ours** と2つのベースライン手法 **CKS**, **TDS** を比較した. ベースライン手法 **CKS** は, クエリログを基にクエリに関連するキーワードを提示する従来のクエリ推薦手法である. 従来型のクエリ推薦を模倣するために, 実験では, あらかじめ用意したクエリを Yahoo! Web Search^{*6}に入力することで推薦されたキーワードを収集し, それを提示という操作を行った. 提示するキーワードのランキングは, Yahoo! Web Search が提示したキーワードの順序をそのまま用いた. ベースライン手法 **TDS** は, 3章で説明した被反証センテンスのランキング過程で $p(d)$ のみを用いてランキングを行う, 典型度ベースの反証センテンス提示手法である. **Ours** と **TDS** では, (被) 反証センテンスの収集を行うために, 1クエリあたり検索結果を40件取得するよう設定した. また, **Ours** のランキングでは, 式 (1) における λ の値を0.5に設定した.

4.1 評価尺度

実験では, 提案手法およびベースライン手法の出力結果の上位 k 件に, 関連度の高い/警鐘度が高い/有用度が高い

反証センテンス/キーワードがどの程度含まれているかを評価した. 評価式は以下のとおりである:

$$r@k = \frac{1}{|Q|} \sum_{q \in Q} \frac{|Relevant(S_q(k))|}{k} \quad (4)$$

$$a@k = \frac{1}{|Q|} \sum_{q \in Q} \frac{|Alerting(S_q(k))|}{k} \quad (5)$$

$$u@k = \frac{1}{|Q|} \sum_{q \in Q} \frac{|Useful(S_q(k))|}{k} \quad (6)$$

ここで, Q は実験に用いたクエリセット, $S_q(k)$ は各手法がクエリ q に対して出力した反証センテンス/反証センテンスの上位 k 件を表す. $|S|$ は集合 S の要素数を表す. $Relevant(S)$, $Alerting(S)$, $Useful(S)$ は, 2人以上の評価者が「クエリと関連がある」「検索結果に注意を払いたくなる」「検索結果の信憑性を判断するための証拠を探すのに有用なクエリである」と判断した反証センテンス/キーワードを意味する.

4.2 評価者と実験データ

実験には, 3人の評価者が参加した. 評価者はウェブ検索エンジンの利用経験がある者のみを選んだ. 実験に際し, 5つのカテゴリからなる計20個のクエリを事前に用意した. 用意したクエリを表2に示す.

4.3 実験手順

実験の手順は以下のとおりである. まず, 表2の各クエリを3つの手法 (**Ours**, **TDS**, **CKS**) に入力し, 出力された反証センテンス/キーワードを評価者に提示した. 評価者には, 以下のようなタスクに関する説明を提示した後, 提示された反証センテンス/キーワードの関連度, 警鐘度, 有用度を評価してもらった.

“mobile phone health” というキーワードでウェブ検索しているケースを想定してください. 今, あなたは, ウェブ検索エンジンから検索結果に加え, 以下のようなセンテンス/キーワードを提示されたとします. 提示されたセンテンス/キーワードに関して, 3つの質問に答えてください.

CKS に関する評価タスクでは, Yahoo! Web Search が推薦したキーワードをそのまま評価者に提示した. **Ours**

表2 実験に用いたクエリセット

Table 2 Query set.

Category	Query
論争	dinosaurs extinction, global warming, earthquake cause plastic recycling, renewable energy
健康	effective diet, cancer treatment, mobile phone health vaccine side effects, caffeine overdose
勘違い	light bulb inventor, telephone inventor civil war reason, caesar salad name
食べ物	coffee health, grapefruit seed extract benefit, potato poison
金融	forex risk, mortgage refinancing, personal debt reduction

*6 Yahoo Web Search: <http://search.yahoo.com/>

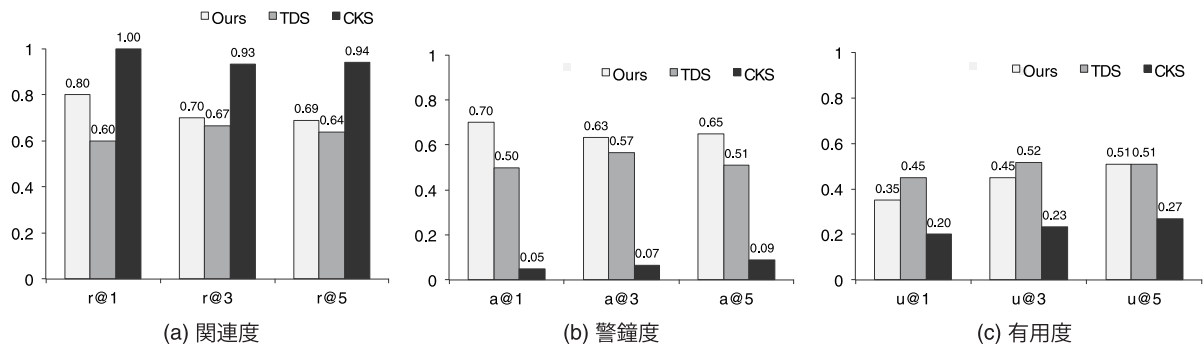


図 3 関連度, 警鐘度, 有用度の比較

Fig. 3 Comparison among relevance, alertingness, and usefulness of three methods.

および TDS に関する評価タスクでは, 各手法で検索された被反証センテンスに接頭句 “Some people doubt:” 付加した反証センテンスを評価者に提示した*7.

各手法によって提示された反証センテンス/キーワードの関連度, 警鐘度, 有用度を評価するために, 評価者には以下の質問に YES/NO で答えてもらった:

- 関連度: 提示されたセンテンス/キーワードはクエリと関連がありますか?
- 警鐘度: 提示されたセンテンス/キーワードを見ると, ウェブ検索結果の信憑性を慎重に吟味しようという気になりますか?
- 有用度: 提示されたセンテンス/キーワードは, クエリに関するウェブ検索結果の信憑性を判断する手がかりを検索するのに, 役に立ちそうですか?

4.4 結果

図 3 に, 各手法の関連度 $r@k$, 警鐘度 $a@k$, 有用度 $u@k$ の平均値を示す. 図 3(a) によると, 関連度の観点では CKS は Ours と TDS に比べ, かなり性能が高かったことが分かる. しかし, Ours の性能はそれほど悪くはなかった ($r@1 = 0.80$, $r@3 = 0.70$, $r@5 = 0.69$). CKS は, 実際にウェブ検索で発行されたクエリを基にキーワードを提示する手法である. それゆえ, 文書を解析して反証センテンスを提示する Ours や TDS に比べ, CKS はクエリにきわめて関連するキーワードを提示することが可能である. $r@1$, $r@3$, $r@5$ を見ると, Ours は TDS よりもつねに優れていた. 特に $r@1$ に関しては, Ours は TDS に比べ 33% も性能が高かった. これは, 式 (1) のクエリ尤度 $\prod_{t \in q} ((1-\lambda)p(t|M_c) + \lambda p(t|M_d))$ が関連度の評価に寄与したと考えられる.

警鐘度に関しては, 図 3(b) が示しているとおり, Ours と TDS が上位 5 件に提示したセンテンスのうち, 少なくとも 50% がユーザの信憑性判断への喚起に成功した. Ours

は $a@1$, $a@3$, $a@5$ のすべてにおいて, TDS よりも高い性能を示した. 一方, CKS に関しては, 評価者は提示されたキーワードを見ても, 信憑性判断への意識は高まらなかった ($a@1 = 0.05$, $a@3 = 0.07$, $a@5 = 0.09$). 既存のキーワードクエリ推薦の目的は, 信憑性が疑われる情報への注意喚起を目的としていない. それゆえ, CKS の警鐘度が低いことは自然である.

図 3(c) によると, Ours と TDS は CKS よりも有用度が高かったことが分かる. この結果から, クエリに関するウェブページの信憑性判断を行うための手がかり情報を検索するには, 反証センテンスの提示は既存のキーワード推薦手法よりも有用であることがうかがえる. しかし, 反証センテンス提示の有用度スコアは当初期待していた値よりも低かった. また, $u@1$ と $u@3$ に関しては, Ours は TDS よりも低いスコアを示した.

これらの理由を調べるために, 実験後, 評価者にインタビューを行った. 以下は, 評価者の回答の一部である:

提示された反証センテンスの大半はそれを知るだけで, 検索ワードに関するウェブページの信憑性を判断する証拠や手がかりとして十分でした. だから, わざわざ提示されたセンテンスを使って証拠情報を検索する必要はないと判断しました.

図 3(c) を見る限りでは, 反証センテンスは信憑性判断の手がかりを検索するのに有用ではないと思われた. しかし, 上のインタビュー回答は, 反証センテンスが役に立たないというわけではないこと示している. このことから, クエリに関連する反証センテンスを提示することができれば, クエリに関連するウェブページの信憑性を判断するうえで, 反証センテンスそのものが有用な手がかりになりえると考えられる.

以上の結果から, クエリに対してある程度の関連する反証センテンスを提示できれば, 提案手法はベースライン手法よりも, ウェブ検索時におけるユーザの信憑性判断支援に有用であることが示された.

*7 クエリ “mobile phone health” に対して, システムが “mobile phone has a health risk” を被反証センテンスとして抽出した場合, “Some people doubt: mobile phone has a health risk” を反証センテンスとして評価者に提示した.

表 3 反証センテンス提示の成功例. 括弧内の数字は提示された反証センテンスの順位を表す
 Table 3 Good examples of suggested disputed sentences. Numbers in parentheses are ranking orders of suggested disputed sentences.

クエリ	提示された(被)反証センテンス
potato poison	"green potatoes are poisonous" (1)
vaccine side effects	"tamiflu does not have potential side effects" (5)
caesar salad name	"the salad takes its name from julius caesar" (1)
civil war reason	"slavery was the cause of the civil war" (1)
earthquake cause	"fracking causes earthquakes" (3)

表 4 反証センテンス提示の失敗例とその種類. 括弧内の数字は提示された反証センテンスの順位を表す

Table 4 Bad examples of suggested disputed sentences and error types. Numbers in parentheses are ranking orders of suggested disputed sentences.

クエリ	(被)反証センテンス	エラーの種類
potato poison	"potatoes are fattening" (5)	クエリと無関係
global warming	"global warming was the cause" (1)	意味が曖昧
telephone inventor	"agb was not the inventor" (3)	ユーザが知らない
plastic recycling	"shipping plastics for recycling" (2)	文法ミス

4.5 ケーススタディ

表 3 と表 4 に, 提案手法による反証センテンス提示の成功例, 失敗例を記す. 評価者が反証センテンスとして有用でないと判断したものを確認したところ, 「クエリと無関係なもの」「意味が曖昧なもの」「ユーザが知らないもの」「文法ミスがあるもの」に分けられた.

クエリ *potato poison* に対してセンテンス "*potatoes are fattening*" が提示された例のように, 提案システムはときどきクエリに関係のないセンテンスを提示することがあった. これは, 抽出された被反証センテンスがクエリとほとんど無関係でも, それがウェブ上で典型的なものであれば, 提案システムは高いスコアを与えてしまうことが原因として考えられる. この問題に対応するためにも, 式 (1) におけるクエリ尤度スコアと典型度スコアの計算のバランスを調整する必要がある.

提案システムがクエリと関連度の高い被反証センテンスを抽出できたとしても, クエリ *global warming* に対するセンテンス "*global warming was the cause*" のように, 意味が曖昧なセンテンスを提示してしまうことがあった. 提案手法では, "*it is not true that*" や "*no proof that*" のような言語パターンの直後に出現するセンテンスを単純に抽出して, それを被反証センテンスとしている. それゆえ, 文脈が分からなければ, 意味の分からないセンテンスが抽出されてしまったと考えられる.

提示された反証センテンスがクエリと関連があり, ウェブ情報の信憑性への警戒を促すものであっても, 評価者がそれを有用でないと判断してしまうケースもあった. これは, 評価者が提示されたセンテンスに含まれる語について知識がなかったために, センテンスの意味するところを理解できず, 実際には反証センテンスとして有用であるに

もかわらず「有用でない」と判断してしまったと考えられる. たとえば, *telephone inventor* というクエリに対し, 提案手法は "*agb was not the inventor*" というセンテンスは, ウェブ上で反証されていると示唆した. もし, ユーザが *agb* という語が *Alexander Graham Bell* を意味することを知らなければ, 提示された反証センテンスは役に立たないと考え, 表示されたウェブ検索結果の信憑性に注意を払わないだろう. それゆえ, ユーザに反証センテンスを提示する際には, センテンスの理解容易性も考慮することも重要である.

クエリ *plastic recycling* に対して提示された反証センテンス "*someone doubts: shipping plastics for recycling*" のように, 提案手法によって提示された反証センテンスの中には, 文法的に読みにくいものが含まれていることがあった. 提案手法では, 句点 (.) で文書を分割し, 単純に that 節に着目して被反証センテンスを抽出している. それゆえ, 文法的に完全なセンテンスを抽出できないことがしばしば起こりうる. この問題に対処するには, より深い自然言語処理を行う必要がある.

5. おわりに

本稿では, ウェブ検索プロセスにおいて検索結果の信憑性判断への注意喚起を目的とする, 新しいタイプのクエリ推薦手法を提案した. 提案システムでは, *mobile phone health* といったキーワードクエリが入力されると, ウェブ上で反証されているセンテンスを収集し, それを "*Some people doubt: mobile phone are bad for your health*" のような形式でユーザに提示する. 既存のキーワードクエリ推薦は, ユーザの検索意図に合致するウェブページを検索しやすくすることを目的としているが, 本稿で提案した反証

センテンス推薦は、ユーザに入力されたクエリに関して疑わしい情報の存在を気付かせたり、ウェブページの信憑性判断を行うための手がかり情報を検索したりしやすくすることを目的としている。ユーザが、ウェブ情報の信憑性に注意を払っていなかったり、疑わしいウェブ情報の存在に気付いていなかったりしたとしても、提案システムによって、ユーザは信憑性を意識しながらウェブ検索を行うことが可能となる。

実験結果によって、提案手法によって提示された反証センテンスは、ウェブ検索時に情報の信憑性をユーザに意識させるのに効果があるということが分かった。しかし、クエリに対する提示されたセンテンスの関連度という観点では、提案手法は既存のキーワードクエリ推薦手法よりも性能が悪かった。より良い反証センテンスを提示するには、クエリ尤度スコアと典型度スコアの計算を最適化し、反証センテンスの関連度を改善する必要がある。実験では、提示された反証センテンスの関連度が高かったとしても、ユーザがセンテンスの意味を理解できないケースもあった。それゆえ、理解しやすい反証センテンスを提示する必要がある。他の課題としては、信憑性の高いウェブページの検索を支援するために、提示された反証センテンスをどのように活用するか、ということがあげられる。

膨大なウェブ情報の中から必要な情報を効率良く安全に取得するには、信憑性に着目した検索システムが重要となる。提案システムは、信憑性指向のウェブ検索に寄与できると信じている。

謝辞 本研究の一部は、文部科学省科学研究費補助研究基盤研究(A)「ウェブ検索の意図検出と多面的検索意図指標に基づく検索方式の研究」(研究代表者: 田中克己, 課題番号: 24240013)によるものです。ここに記して謝意を表します。

参考文献

- [1] Sillence, E., Briggs, P., Fishwick, L. and Harris, P.: Trust and Mistrust of Online Health Sites, *Proc. SIGCHI conference on Human factors in computing systems (CHI 2004)*, pp.663-670 (2004).
- [2] Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., Oyama, S. and Tanaka, K.: Trustworthiness Analysis of Web Search Results, *Proc. 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2007)*, pp.38-49 (2007).
- [3] Ponte, J.M. and Croft, W.B.: A language modeling approach to information retrieval, *Proc. 21st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR 1998)*, pp.275-281 (1998).
- [4] Liu, Y., Gao, B., Liu, T.-Y., Zhang, Y., Ma, Z., He, S. and Li, H.: BrowseRank: Letting Web Users Vote for Page Importance, *Proc. 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR 2008)*, pp.451-458 (2008).
- [5] Cui, H., Wen, J.-R., Nie, J.-Y. and Ma, W.-Y.: Probabilistic Query Expansion Using Query Logs, *Proc. 11th international conference on World Wide Web (WWW 2002)*, pp.325-332 (2002).
- [6] Boldi, P., Bonchi, F., Castillo, C. and Vigna, S.: From "Dango" to "Japanese Cakes": Query Reformulation Models and Patterns, *Proc. 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI 2009)*, pp.183-190 (2009).
- [7] Kotov, A. and Zhai, C.: Towards natural question guided search, *Proc. 19th international conference on World Wide Web (WWW 2010)*, pp.541-550 (2010).
- [8] Suryanto, M.A., Lim, E.P., Sun, A. and Chiang, R.H.L.: Quality-Aware Collaborative Question Answering: Methods and Evaluation, *Proc. 2nd ACM International Conference on Web Search and Data Mining (WSDM 2009)*, pp.142-151 (2009).
- [9] Castillo, C., Mendoza, M. and Poblete, B.: Information credibility on twitter, *Proc. 20th international conference on World Wide Web (WWW 2011)*, pp.675-684 (2011).
- [10] Chia, P.H., Yamamoto, Y. and Asokan, N.: Is This App Safe?: A Large Scale Study on Application Permissions and Risk Signals, *Proc. 21st international conference on World Wide Web (WWW 2012)*, pp.311-320 (2012).
- [11] Schwarz, J. and Morris, M.: Augmenting web pages and search results to support credibility assessment, *Proc. 2011 annual conference on Human factors in computing systems (CHI 2011)*, pp.1245-1254 (2011).
- [12] Yamamoto, Y. and Tanaka, K.: Enhancing Credibility Judgment of Web Search Results, *Proc. 2011 annual conference on Human factors in computing systems (CHI 2011)*, pp.1235-1244 (2011).
- [13] Pirolli, P., Wollny, E. and Suh, B.: So You Know You're Getting the Best Possible Information: A Tool that Increases Wikipedia Credibility, *Proc. 27th international conference on Human factors in computing systems (CHI 2009)*, pp.1505-1508 (2009).
- [14] Ennals, R., Trushkowsky, B. and Agosta, J.M.: Highlighting Disputed Claims on the Web, *Proc. 19th international conference on World Wide Web (WWW 2010)*, pp.341-350 (2010).
- [15] Kawahara, D., Inui, K. and Kurohashi, S.: Identifying contradictory and contrastive relations between statements to outline web information on a given topic, *Proc. 23rd international conference on Computational Linguistics (COLING 2010)*, pp.534-542 (2010).
- [16] Yamamoto, Y., Tezuka, T., Jatowt, A. and Tanaka, K.: Supporting Judgment of Fact Trustworthiness Considering Temporal and Sentimental Aspects, *Proc. 9th international conference on Web Information Systems Engineering (WISE2008)*, pp.206-220 (2008).
- [17] Ennals, R., Byler, D., Agosta, J.M. and Rosario, B.: What is disputed on the web?, *Proc. 4th workshop on Information credibility (WICOW 2010)*, pp.67-74 (2010).
- [18] Erkan, G. and Radev, D.: LexRank: Graph-based lexical centrality as salience in text summarization, *Journal of Artificial Intelligence Research*, Vol.22, pp.457-479 (2004).



山本 祐輔

京都大学学術研究支援室特定専門業務職員。2011年京都大学大学院情報学研究科博士後期課程修了。情報の信憑性，ウェブマイニング，HCI，研究室運営，研究支援に関する研究に従事。ACM，人工知能学会，日本データベース学会，研究・技術計画学会等各会員。



田中 克己 (正会員)

京都大学大学院情報学研究科社会情報学専攻教授。1976年京都大学大学院修士課程修了。博士(工学)。主にデータベース，マルチメディアコンテンツ処理の研究に従事。IEEE Computer Society，ACM，人工知能学会，日本ソフトウェア科学会，日本データベース学会等各会員。

(担当編集委員 中辻 真)