

指導教員（主査）：山本祐輔 講師

副査：大島純 教授

2018 年度 静岡大学情報学部 卒業論文

文章表現の曖昧さ指摘による ウェブ情報精査の態度・行動促進

静岡大学 情報学部 IS プログラム 所属

学籍番号 70510038

齊藤 史明

2019 年 2 月 14 日

概要

本研究では、ブラウジング中のウェブページにある表現が曖昧な文をハイライトするシステムを提案する。提案システムは、文章表現の 5W1H が明確になっていない文を「曖昧な文」と定義し、閲覧中のウェブページから曖昧な文を検知する。本研究では曖昧な文を自動検出するために、Wikipedia で用いられている「言葉を濁さない」ガイドラインに着目する。Wikipedia 記事から「言葉を濁さない」とラベリングされた文を曖昧な文として収集し、SVM を用いて曖昧文の分類器の作成を行った。また、ウェブ情報の曖昧な文をハイライト表示することが、ウェブ情報の精査態度、行動に与える影響を分析するためのユーザ実験を行った。ユーザ実験の結果、提案システムを用いた被験者はハイライト表示の有無を信憑性判断の手がかりとし、効率化された探索行動を行うようになった。

目次

第 1 章	はじめに	5
第 2 章	関連研究	8
2.1	信憑性判断の能力	8
2.2	テキスト情報の信憑性判断	9
第 3 章	提案内容	11
3.1	提案システムの概要	11
3.2	曖昧な文の検出	11
第 4 章	評価実験	15
4.1	被験者	15
4.2	実験条件	15
4.3	タスク	16
4.4	手順	16
第 5 章	結果	18
5.1	情報精査行動	18
5.2	情報精査態度	19
5.3	回答への自信と精査行動	20
第 6 章	考察	26
6.1	曖昧な文の検出性能	26
6.2	実験結果	28
第 7 章	まとめ	30

目次

1.1	実際に Wikipedia で使われている「誰？」の例	6
3.1	システム稼動例	12
5.1	タスク, 条件別のタスク所要時間 (秒)	19
5.2	タスク, 条件別の記事ページ総閲覧時間 (秒)	20
5.3	タスク, 条件別の検索結果表示ページの閲覧時間 (秒)	20
5.4	タスク, 条件別の記事ページ閲覧件数	21
5.5	タスク, 条件別の各記事ページの平均閲覧時間 (秒)	21
5.6	タスク, 条件, 回答への自信別の記事ページ閲覧件数	23
5.7	タスク, 条件, 回答への自信別の記事ページ総閲覧時間 (秒)	23
5.8	タスク, 条件, 回答への自信別の平均記事ページ閲覧時間 (秒)	24
5.9	タスク, 条件, 回答への自信別の検索結果表示ページの閲覧時間 (秒)	24
5.10	タスク, 条件, 回答への自信別のタスク所要時間 (秒)	25

表目次

3.1	対象としたテンプレートの一覧	12
3.2	言葉を濁した表現の典型例の例	13
3.3	性能評価	14
5.1	タスク，条件別の回答に対する自信の人数分布とスコア平均	22
5.2	条件別の信憑性判断で考慮した要素	22
6.1	良い検知例	27
6.2	悪い検知例	27

第 1 章

はじめに

ウェブの情報の中には、フェイクニュースのような嘘の情報や信憑性の低い情報が相当数存在しており、その影響が社会問題となっている。信憑性が低いウェブ情報には様々なタイプがある。その 1 つに、曖昧な表現で文章が書かれているために、内容の正確さに疑問が持たれる情報がある。

ウェブ上のテキスト情報の中には、「業界関係者は～と語る」、「有識者は～と見る」、「ネット上では～と言われている」など、情報の提供者が信頼できる人物なのかが確認できないものがある。そのような情報は読み手が情報の信憑性について注意すべきものの一つとして考えられる。また、ある事象への評価について「賛否両論」「～という意見が殺到する」「物議を醸す」のようにコンテンツ作成者の主観が含まれる表現が使われている文も散見される。このような文章表現は、意見が誰にどれ程の規模で支持されているか曖昧で、事象への評価が読み手の捉え方次第となり、正確な情報を受け取った場合とは違う印象を持ってしまう危険性があるそのため、この種のテキスト情報に対しても信憑性について注意を払うべきだと考えられる。

多くのウェブユーザが情報源として利用している Wikipedia にも、曖昧な文章表現を含む記事が相当数存在することが明らかになっている。日本語版の Wikipedia には、ユーザの善意によって文章表現が曖昧で問題があるとチェックされた記事が 6430 件存在することが報告されており、内容改善が求められている^{*1}。それゆえ、曖昧な表現を含むウェブ文書の信憑性を判断するための支援が必要となる。

これまで情報アクセスシステムの研究分野では、信憑性判断支援として、閲覧中のウェブページの内容に反証や対立意見が存在する場合にその旨を警告するためのシステムやア

^{*1} 2018 年 4 月 19 日時点

岐阜の名の由来については「[岐阜県#県名の由来](#)」を参照

織田信長が、美濃国を攻略した際に、稲葉山の城下の井ノ口を岐阜と改めたとするのが定説だが、異論も一部にある^[誰?]。

図 1.1 実際に Wikipedia で使われている「誰？」の例

ルゴリズムが研究開発されてきた [3][4]。これらの研究では、信憑性の検証対象に対して反証や対立意見が存在しているかを判断基準にしている。しかし、反証や対立意見が存在するか否かは、信憑性判断基準の一つでしかない。反証や対立意見が存在しなくても、信憑性に注意を払うべき情報は多数存在する。

本稿では、閲覧中のウェブページ内に存在する文章表現が曖昧な文をハイライトすることで、慎重な情報精査を促すシステムを提案する。Wikipedia には、記事を書く際のガイドラインとして「言葉を濁さない」がある。言葉を濁すとは、ガイドラインには「意見や主張を記事中に書くときにその意見の論者を明確にせず一般的なことであるかのように表現にすること」*2とある。Wikipedia ではこのガイドラインに則り、文の 5W1H について記述が十分でないことを発見した場合に「誰？」や「いつ？」などのテンプレートが文に付け、改善が求められることができる。例として図 1.1 を挙げる。この例では岐阜県の県名の由来について「稲葉城の城下の井ノ口を岐阜と改めたとするのが定説だが、異論も一部にある。」という文に対して、「誰？」とテンプレートが付けられている。この例では、「異論」を唱えた人物は誰なのかを補完する必要がある。本稿では、このような正確性の不足を指摘するテンプレートが付けられた文をコーパスとし、機械学習の教師あり学習を行うことで、5W1H の「誰が」について曖昧な文の検出器を構築する。文章表現の曖昧さに着目することで、従来研究では対応できていない観点からウェブ情報の信憑性判断の支援を行うことが可能となる。

本研究では、提案システムがウェブ情報探索中のユーザの情報精査の態度、行動に与える影響を分析するために、以下の仮説を設定し、提案システムを用いたユーザ実験を行う。

- H1** 正確性について曖昧な文をハイライトすることで、ウェブ探索に掛ける時間が増えるなど、情報精査の行動が促される。
- H2** 正確性について曖昧な文をハイライトすることで、情報精査の態度が促進され、文章表現やデータの有無などの信憑性判断に必要な要素に注目するようになる。
- H3** 正確性について曖昧な文をハイライトすることで、意思決定の信念の強さに依存

*2 <https://ja.wikipedia.org/wiki/Wikipedia:言葉を濁さない>

する.

本論文の構成は以下の通りである．次章で関連研究について述べる．3 章では提案システムの概要，表現が曖昧な文の検出方法および検出性能について述べる．4 章では提案システムを用いたユーザ実験の方法について述べる．5 章ではユーザ実験の結果について述べる．6 章では曖昧な文の検出方法，検出性能の考察，及びユーザ実験結果の考察を述べる．最終章では本稿のまとめを行う．

第 2 章

関連研究

本章では「信憑性判断の能力」と「テキスト情報の信憑性判断支援」に関する関連研究について記す。

2.1 信憑性判断の能力

田中 [1] は、批判的思考の促進・抑制を行う要因として、論法の誤りのタイプ、批判的思考の外的要求レベル、情報ソースの信憑性に注目し、その影響を分析する実験を行った。被験者にはそれぞれタイプの違う論理的に誤りのある文章を 3 つ読ませ、その誤りを指摘できるかテストを行った。その結果、論法の誤りのタイプについて、読み手が論法の誤りに気づきやすいものと気づきにくいものがあることが分かった。また、外的要求について、文章に対して批判をするように要求することで批判的思考が促進された。また、情報ソースの種類によって読み手の批判的態度が変化した。新聞記事や著者が大学教授である文章を読ませる場合と、匿名の大学生が書いたレポートとして読ませる場合では、信憑性が高いと読み手が感じる情報ソースの文章に対しては批判的思考が抑制される事が分かった。しかし、情報ソースの信憑性が批判的思考に及ぼす抑制効果は、論法の誤りのタイプや外的要求の程度で変化すると分析している。

この実験では、批判的思考の要求のレベル設定として、文章の感想を書かせる設定、文章中で間違いがある箇所をいくつか指摘させる設定、文章中で間違いがある箇所をすべて指摘させる設定を設けている。後者 2 つの設定では、論理的に間違いがある文章が存在していることを被験者に知らせている。実験の結果、後者 2 つの設定によって批判的思考が促進されたことが報告されている。この結果を踏まえて、本稿では、文の間違いや曖昧さを読み手に知らせることで、批判的なウェブページの閲覧行動が促進されるという着想に

至った。

後藤 [2] は、ウェブ情報への批判的思考能力の発達傾向を検討するために、小学生から大学生を対象に批判的思考のスキルを測定した。ある情報の信頼性を判断するための方法と必要となる情報を被験者に自由記述させ、その回答内容から批判的思考のレベルを測定した。その結果、教育段階に応じて批判的思考スキルが発達する傾向があると分析した。しかし、情報信頼性を判断するための方法を回答することができない大学生もあり、大学生でもスキルが不足している層が一定数存在することが明らかになった。

2.2 テキスト情報の信憑性判断

Liao[4] らは、ユーザが自分にとって都合の良い情報のみを集めて偏った意見が強化されるエコーチェンバーへの対策として、意見の立場や情報の発信者の専門性を表示する手法を提案した。ユーザ実験の結果、情報発信者の専門性のみを表示した場合、ユーザは専門性の高い意見をより信頼する傾向にあることを明らかにした。一方で、情報発信者の専門性と同時に発信者の立場を表示すると、ユーザは自身が賛同している立場の意見を選好する傾向が弱まり、このような情報提示手法がエコーチェンバー対策に一定の効果があることを明らかにした。

Ennals[3] らは、ウェブページ上の情報の中で、他のウェブページで反証が行われている文をハイライトすることで、信憑性が疑わしい情報に対して警告をおこなうシステムを提案した。ハイライトされた箇所をクリックすることで、対立した意見の書かれた記事が表示される。この研究では、真偽や専門性をウェブ上の情報探索で確かめることができる情報であることが前提となっている。ウェブ上に存在する情報の中には、ニュース記事のような 1 次情報を探すことが困難で、その真偽を確かめる方法がない情報も存在しており、ウェブ探索時の信憑性判断支援を行う為にはそのような情報の信憑性についても注意させる必要があると考える。

Yang[5] は、政治に関する発言や情報について真偽を検証するファクトチェックを行うウェブサイトの「PolitiFact」のデータを利用し、ニュース情報の信憑性を自動判定するアルゴリズムを提案している。PolitiFact は政治に関する発言や情報の信憑性について、「True (正確な情報である)」から「Pants on Fire (馬鹿げている)」までの 6 段階の評価が付与されている。Yang らの提案手法はニュース情報に登場する政党や職業、話者の信頼性などを特徴量として、ニュース記事の信憑性を先の 6 段階で評価するための分類器を構築した。しかし、提案システムが考慮している要素や学習器の特徴量を含まれない、あるいはシステムの穴をつくような低信憑性情報は、提案システムを用いても検出が難しい。

ユーザは情報の真偽を確かめることができない情報についても信憑性を疑う必要がある。このような問題に対応するために、本稿ではユーザの情報精査行動を促進することに焦点を当てる。

第 3 章

提案内容

本章では，閲覧中のウェブページ内の曖昧な文のハイライトするブラウザシステム，および曖昧文の検出方法について述べる．

3.1 提案システムの概要

提案システムは，ウェブブラウザの拡張システムであり，閲覧中のウェブページ内に正確性が曖昧な文が存在する場合，それをハイライト表示する．提案システムは，ユーザがウェブページにアクセスした際に，ページの URL を API サーバに送信する．ページの URL を受け取った API サーバは，ウェブページ内のテキストを取得し文に分解する．その後，API サーバは事前に構築した文の曖昧さ判定器を用いて各文の曖昧さを判定する．最終的に，曖昧と判定された文がブラウザに送信され，ブラウザ拡張は閲覧中のウェブページに存在する曖昧な文をハイライトする．曖昧文のハイライトは，該当箇所の背景色を変更し，文字を太字にして表示することによって行われる．システム稼動例を図 3.1 に示す．

3.2 曖昧な文の検出

提案システムを実現するために，教師あり学習を用いた曖昧な文の検出器を構築する．曖昧文検出器は文を入力として受け取り，その文が 5W1H の「誰」の部分に関して「曖昧か」「曖昧でないか」の 2 値分類を行う．

分布 編集

南アメリカ原産で、世界中に帰化している^[2]。日本では沖縄で帰化植物として野外で繁殖している^[2]。日本へは江戸時代後期にオランダ船によって持ち込まれたといわれている。



分布 編集

南アメリカ原産で、世界中に帰化している^[2]。日本では沖縄で帰化植物として野外で繁殖している^[2]。日本へは江戸時代後期にオランダ船によって持ち込まれたといわれている。

図 3.1 システム稼動例

表 3.1 対象としたテンプレートの一覧

テンプレート名	記事への表示方法	Wikipedia のテンプレート
誰	誰？	{{ 誰 date=yyyy 年 m 月 }}
誰範囲	誰？（該当箇所に下線）	{{ 誰範囲 date=yyyy 年 m 月 該当箇所 }}
誰 2	誰によって？	{{ 誰 2 date=yyyy 年 m 月 }}
誰範囲 2	誰によって？（該当箇所に下線）	{{ 誰範囲 2 date=yyyy 年 m 月 該当箇所 }}

3.2.1 教師データ

本稿では教師データの作成のために、Wikipedia の「言葉を濁さない」ガイドラインに着目した。Wikipedia の日本語版記事全文データ^{*1}から当該ガイドラインに抵触していることを示すテンプレートが用いられた文を抽出し、その中で 5W1H の「誰が」が曖昧であるとラベリングされた文を、正解データ（曖昧である文）とした。表 3.1 は対象となるテンプレートの一覧と、そのテンプレートが付けられた際に実際に文末に表示される言葉の一覧である。取得した記事の文章の中から、テンプレートの付けられた文を正解データとし、残りのテンプレートの付いていない文の中で無作為に抽出した文を不正解データとした。取得した Wikipedia のデータは、2018 年 4 月 19 日時点のデータである。教師データの件数は正解データ、不正解データ共に 2236 件で合計 4472 件である。

^{*1} <https://dumps.wikimedia.org/jawiki/>

表 3.2 言葉を濁した表現の典型例の例

～と言っている，～という話もある
～と信じている人もいる，～という批判もある
なぜか～である，～と思われる，～ようである
言わずと知れた～である，伝えられるところでは～
評論家は～と論じている，学者は～と論じている

3.2.2 素性

提案手法では，文の曖昧性の分類に用いる素性として，「Bag-of-Words」および「言葉を濁す特有表現」の 2 つの素性を使用する．

Bag-of-Words 文中における特定の名詞，形容詞，動詞の出現の有無を表す素性．教師データ中の文に対して形態素解析を行い，名詞，形容詞，動詞を抽出し，特徴語辞書を作成し，文中の単語が作成した辞書に含まれているかどうかで 2 値ベクトルで表す．形態素解析には MeCab^{*2}を利用した．

言葉を濁す特有表現の有無 表 3.2 の 27 例のそれぞれについて，文章中に存在しているかどうかを表す素性．表 3.2 は Wikipedia の「言葉を濁さない」ガイドラインの「言葉を濁した表現の典型例」であり 27 例の特有表現が問題とされている．これら表現を様々な活用形に対応できるよう短縮し，それらが文中に存在しているか否かを 2 値で表現したものを特徴量とした．

3.2.3 学習器

学習器には，サポートベクターマシン（SVM）を用いた [7]^{*3}．

3.2.4 評価

学習器の評価結果を表 3.3 に示す．SVM 学習では RBF カーネルを使用した．グリッドサーチによって最適なパラメータを $C=10$ ， $\gamma=0.01$ と設定した．その後，3.2.1

^{*2} <http://taku910.github.io/mecab/>

^{*3} 実装には，LIBSVM を用いた：<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

にて用意した教師データを用いて、曖昧な文を検出できるか 5 分割交差検定を行った。

評価指標として、精度 (accuracy), 適合率 (precision), 再現率 (recall), F 値 (f1) を用いた。それぞれの定義を下記に記す。

$$accuracy = \frac{\text{予測が正解だった件数}}{\text{テストデータの件数}} \quad (3.1)$$

$$precision = \frac{\text{曖昧な文と判定された正解データの件数}}{\text{曖昧な文と判定した件数}} \quad (3.2)$$

$$recall = \frac{\text{曖昧な文と判定された正解データの件数}}{\text{テストデータ中の正解データの件数}} \quad (3.3)$$

$$f1 = \frac{2 \cdot recall \cdot precision}{recall + precision} \quad (3.4)$$

スコアは, accuracy が 0.764, precision が 0.772, recall が 0.748, f1 が 0.760 であった。

表 3.3 性能評価

評価指標	accuracy	precision	recall	F1
スコア	0.764	0.772	0.748	0.760

第 4 章

評価実験

本章では，ウェブ情報の曖昧な文をハイライト表示することが，ウェブ情報の精査態度・行動に与える影響を分析するためのユーザ実験について述べる．

4.1 被験者

クラウドソーシングサービスの Lancers.jp を用いて，100 名の被験者を募集した．被験者 100 名のうち，回答に不備のあった被験者，あるいはタスク中に退席したと思われる被験者の回答を削除した結果，96 名の被験者の回答を分析に用いた．各被験者には実験参加の報酬として 100 円を支払った．

4.2 実験条件

実験は，ブラウジングインタフェースを要因とする 1 要因 2 水準の実験計画で実施する．被験者を，提案システムを使用する proposed 群と使用しない control 群の 2 つに分ける．実験群が訪れるウェブページには，文が曖昧であるかチェックされており，曖昧な文にはハイライト表示がされている．統制群にはシステムを使用せず，ハイライト表示のないウェブページを見て情報探索を行ってもらう．各実験条件への被験者の割り当てはランダムに行った．

4.3 タスク

テーマについて様々な意見が掲載されている分野で、特に信憑性の低い情報が含まれる分野を情報探索の対象とするために、医療に関するタスクを2件用意した。

タスクで被験者に問う質問は「シナモンは糖尿病の症状改善に有効でしょうか?」「ビタミンCは肺炎の予防に有効でしょうか?」である。タスクの選定には、Yamamoto が行った検索タスクを参考にした [8]。

4.4 手順

本ユーザ実験は「事前説明」、「情報探索タスク」、「事後アンケート」の3ステップから成る。

事前ステップでは、タスクの大まかな流れを説明を行う。また、提案システムを利用する利用者には、「ブラウザ上で5W1Hの「誰」に関して明確になっていない文には、ハイライト表示がされている」旨を伝える。

情報探索ステップでは、被験者は後述する情報探索タスクに取り組む。タスクの開始時に、被験者には以下のようなシナリオを提示する。

「シナモンは糖尿病の症状改善に有効でしょうか? あなたはこの質問への回答を出すために、たった今ウェブ検索を行い、検索結果として以下のウェブページのリストが表示されたとします。表示されたウェブページのリストを参考に質問に対する回答を考えてください。あなたなりの答えが出たら、情報探索を終了し、質問に対する回答を入力してください。」

被験者がタスクシナリオを読んだ後、検索タスクに関連する検索結果リストが表示される。被験者は検索結果リストに掲載された記事ページを訪問し、質問への回答を考える。検索結果リストに表示される記事ページは、それぞれ Google の検索エンジンにて「糖尿病 シナモン」「肺炎 ビタミン c」と検索して表示される上位10件とした。タスク中は、ブラウザの拡張機能を用いて、訪問ページのURL、ページの訪問回数、各ウェブページの滞在時間、ブラウザのタブの切り替え回数などを測定する。

被験者は、各情報探索タスクが終了後に回答に対する自信について、全情報検索タスクが終了後に信憑性判断で考慮した要素について事後アンケートに回答する。

回答に対する自信 自分の回答にどれ位の自信があるか、タスクごとに「とても自信があ

る」「自信がある」「少し自信がある」「少し自信が無い」「自信が無い」「とても自信が無い」の6段階の尺度で回答してもらう。

信憑性判断で考慮した要素 あらかじめ用意した7項目から、情報探索時に情報が信頼できるか判断するために気をつけた要素を選択してもらう。選択肢は、「ウェブページのデザイン」「記事の新しさ」「情報の提供者が明記されているか」「情報の提供者の専門性」「情報が客観的に書かれているか」「証拠情報の有無」「証拠情報の信憑性」の7つである。これらの項目の選定には、山本 [9] らの研究と、平山 [10] らの研究と、四谷 [11] らの研究を参考にした。

第 5 章

結果

本章では，4 で述べた実験の結果について記す．被験者 100 名のうち，不備のあった被験者の回答を削除した結果，controlled 群が 50 名，proposed 群が 46 名の被験者を分析対象とした．

なお，収集データに正規性が確認できなかったため，データに整列ランク変換を適用し，ノンパラメトリック分散分析を行った [13]．分散分析には ARTool 統計分析パッケージを用いた^{*1}．

5.1 情報精査行動

各タスクに掛けた時間（図 5.1）は，ハイライト表示の有無によって統計的有意差は確認されなかった（UI: $F(1,182) = 1.34$, $p = 0.249$)．各群の被験者が検索結果表示ページを閲覧した総時間については，図 5.3 が示しているとおり，ハイライト表示を行った proposed 群の方が controlled 群に比べて有意に短くなることが確認された（ $F(1,182) = 4.06$, $p < .05$)．記事ページを閲覧した総時間（図 5.2）については，タスク間で統計的有意差が確認された（ $F(1,182) = 7.71$, $p < .01$)．また，ハイライト表示を行った場合，ハイライト表示をしなかった場合に比べて，記事ページ総閲覧時間が短くなる傾向が見られた（ $F(1,182) = 2.89$, $p < .1$)．

ハイライト表示をすることでより慎重な精査行動が促進され，各ウェブページの閲覧時間が長くなると予想していたが，結果として閲覧時間は短くなる傾向にあることが明らかになった．また，各タスクごとの記事ページの平均閲覧時間（図 5.5）も proposed 群では短

^{*1} <http://depts.washington.edu/madlab/project/art/>

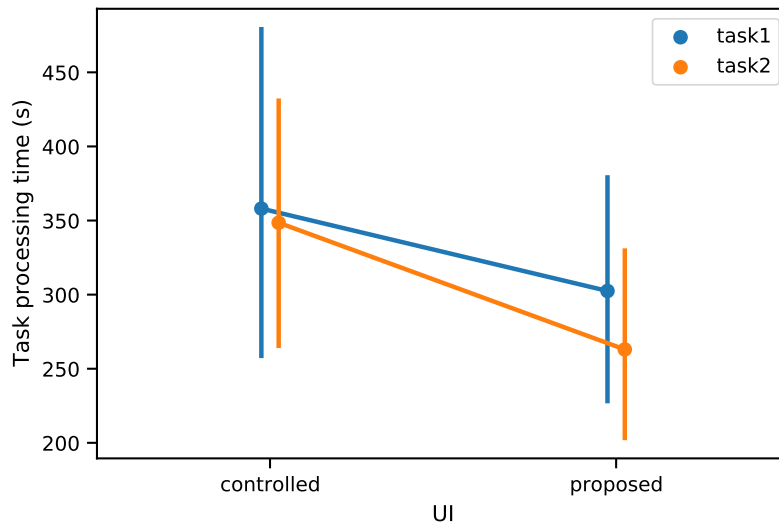


図 5.1 タスク，条件別のタスク所要時間（秒）

縮される傾向にあり，短時間で情報探索行動を終えるようになった（UI: $F(1, 182) = 5.72$, $p < .05$ ）。

各群の被験者が検索結果表示ページに表示された 10 件のうち，閲覧した記事ページの種類の数の平均値を図 5.4 に示す．両タスクで平均値は proposed 群の方が減少しているように見えるが，統計的検定の結果，ハイライト表示の有無による統計的有意差は確認されなかった（ $F(1, 182) = 0.07$, $p = .79$ ）

5.2 情報精査態度

各タスクで，自分の回答にどの程度自信があるかアンケート調査を行った．その結果を表 5.1 に示す．ハイライト表示を行い，ウェブ上の文章について曖昧な点が存在していることを警告することで，情報探索タスクの回答に対する確信が弱まると予想した．結果は，スコア平均で比較すると，両タスクで controlled 群に比べ proposed 群の平均値が 0.2 程度上昇し，予想とは反対の結果となった．この結果から，proposed 群はウェブページ上のハイライト表示の有無を信憑性判断の 1 つの目安として頼っていることが考えられる．

全タスク終了後に，信憑性判断で考慮した要素についてアンケート調査を行った．その結果を表 5.2 に示す．どの項目でも各群の回答に差は見られなかった．

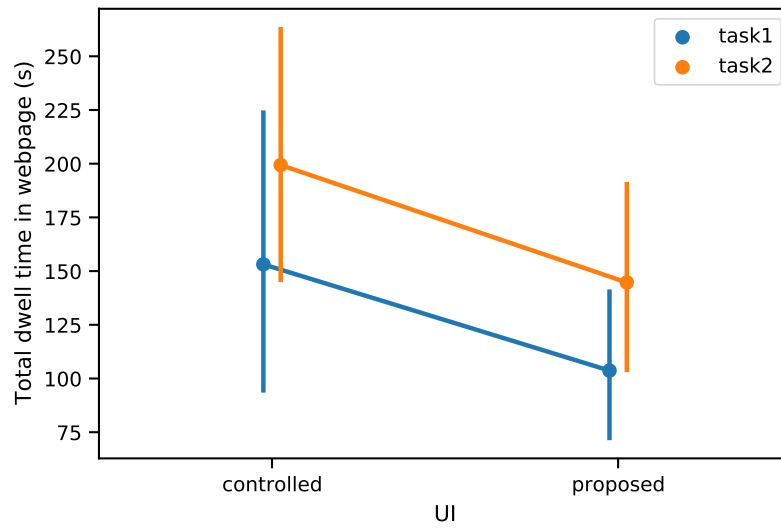


図 5.2 タスク，条件別の記事ページ総閲覧時間（秒）

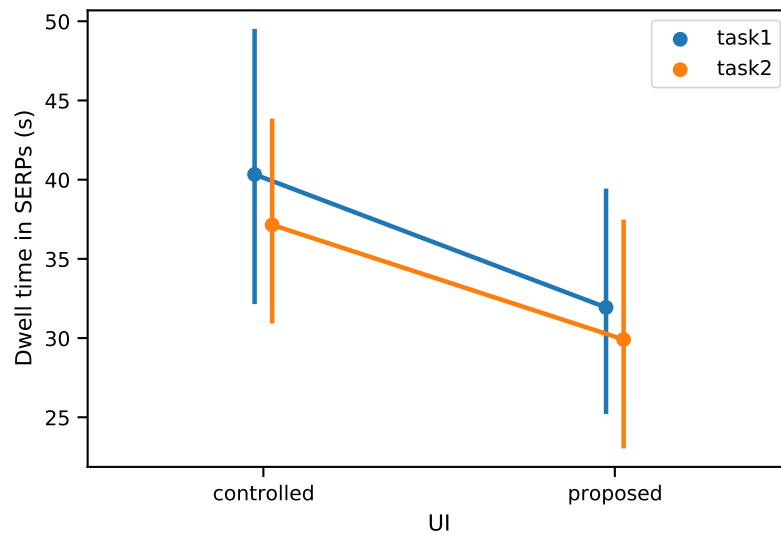


図 5.3 タスク，条件別の検索結果表示ページの閲覧時間（秒）

5.3 回答への自信と精査行動

回答への自信と行動の関係について，タスク後のアンケートで自分の回答に自信があると回答した被験者群と自信がないと回答した被験者群に分け，分析を行った．図 5.6 か

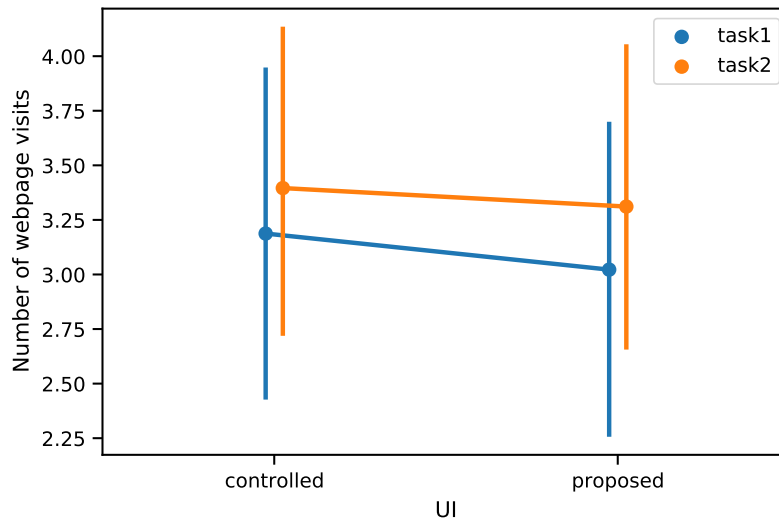


図 5.4 タスク，条件別の記事ページ閲覧件数

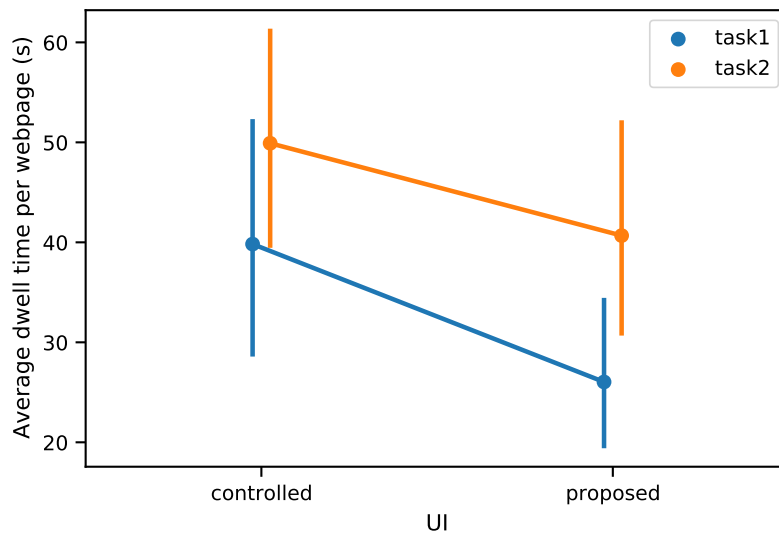


図 5.5 タスク，条件別の各記事ページの平均閲覧時間（秒）

ら，タスク 1 については自信がないと回答した proposed 群は，自信がないと回答した controlled 群に比べて多くの記事ページを訪れる傾向が見られた。しかし，記事ページの総閲覧時間や平均閲覧時間は自信に関わらず減少傾向にあり（図 5.7, 5.8），ハイライト表示によって文章を注意深く読むといった精査行動には結びつかなかったと考える。

検索結果表示ページの閲覧時間について（図 5.9），タスク 1 では自信が無いと回答した

表 5.1 タスク，条件別の回答に対する自信の人数分布とスコア平均

回答	task1		task2	
	controlled (50 人)	proposed (46 人)	controlled (50 人)	proposed (46 人)
とても自信がある (+3)	1	1	2	4
自信がある (+2)	9	10	10	5
少し自信がある (+1)	23	22	22	28
あまり自信が無い (-1)	14	12	14	8
自信が無い (-2)	3	1	2	1
とても自信が無い (-3)	0	0	0	0
スコア平均	0.480	0.674	0.600	0.870
合計人数	50	46	50	46

表 5.2 条件別の信憑性判断で考慮した要素

質問	UI	
	controlled(50 人)	proposed(46 人)
ウェブページのデザイン	0.36	0.35
情報の新しさ	0.48	0.48
情報の提供者が明記されているか	0.76	0.78
情報の提供者の専門性	0.68	0.67
情報が客観的に書かれているか	0.84	0.78
証拠情報の有無	0.76	0.76
証拠情報の信憑性	0.66	0.63

群の方が長く閲覧しているのに対し，タスク 2 では自信があると回答した群の方が長く閲覧している．また，タスクの所要時間については（図 5.10），タスク 1 では自信が無いと回答した群ではハイライトの有無による差が見られないのに対し，タスク 2 では差が見られた．被験者の回答への自信と共に，各タスクで扱った話題に影響を受けることが考えられる．

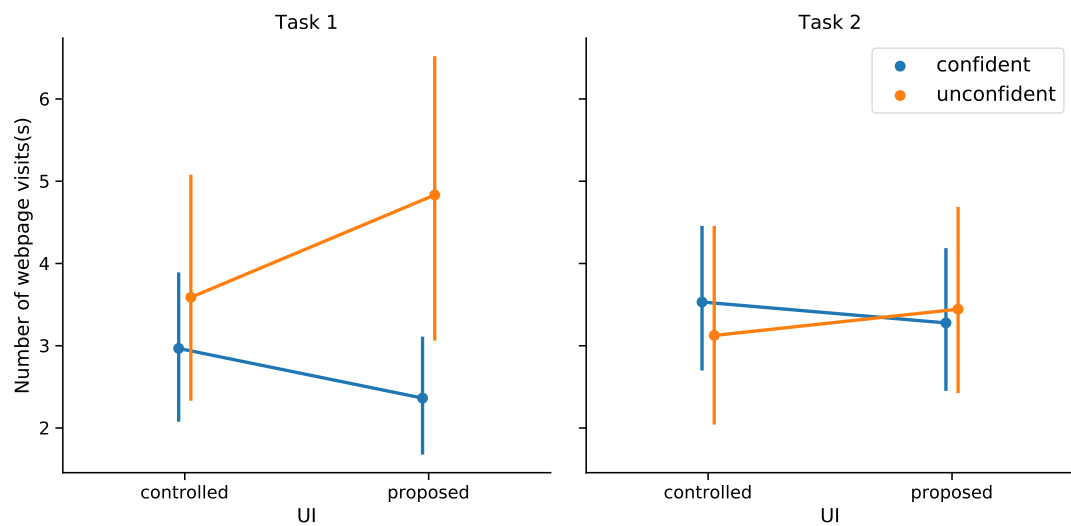


図 5.6 タスク，条件，回答への自信別の記事ページ閲覧件数

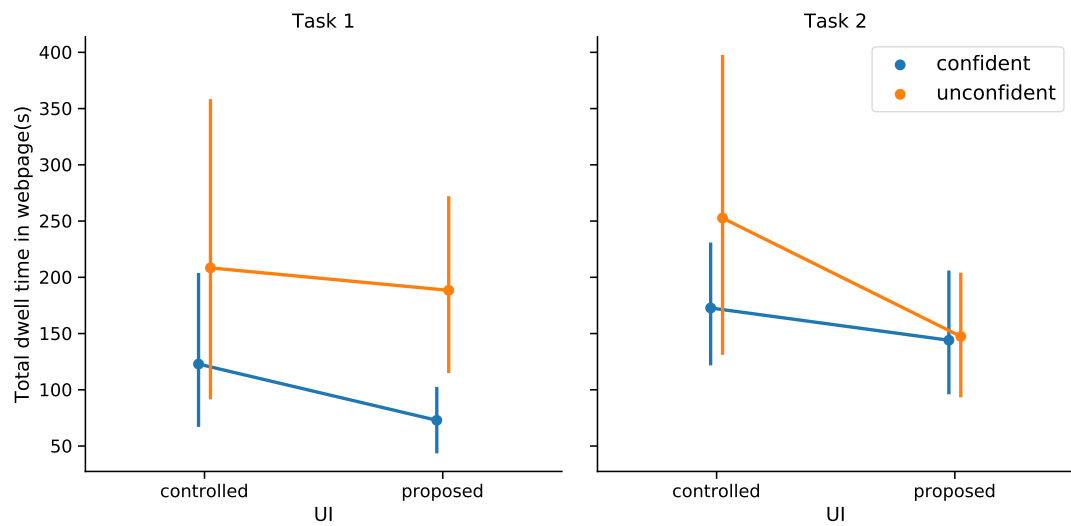


図 5.7 タスク，条件，回答への自信別の記事ページ総閲覧時間（秒）

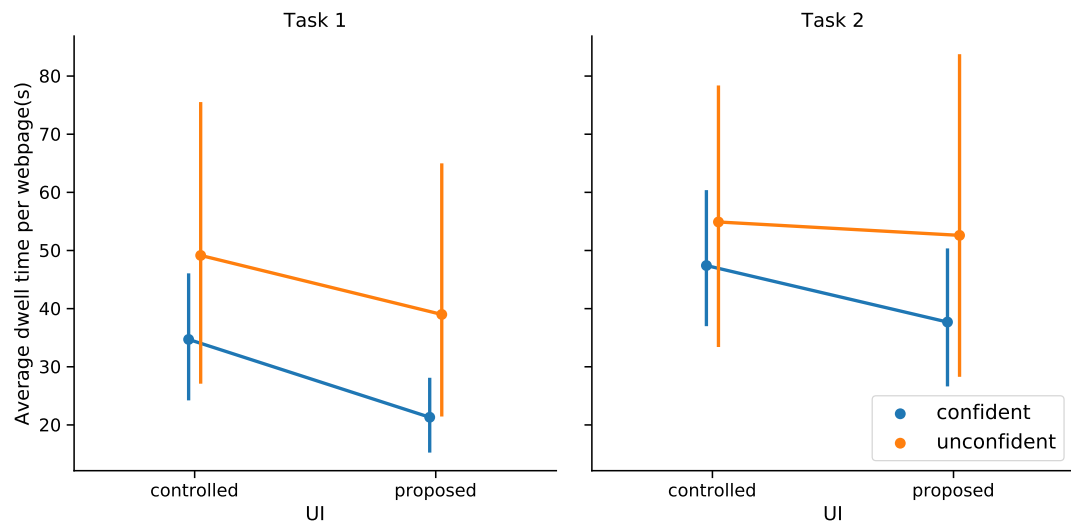


図 5.8 タスク，条件，回答への自信別の平均記事ページ閲覧時間（秒）

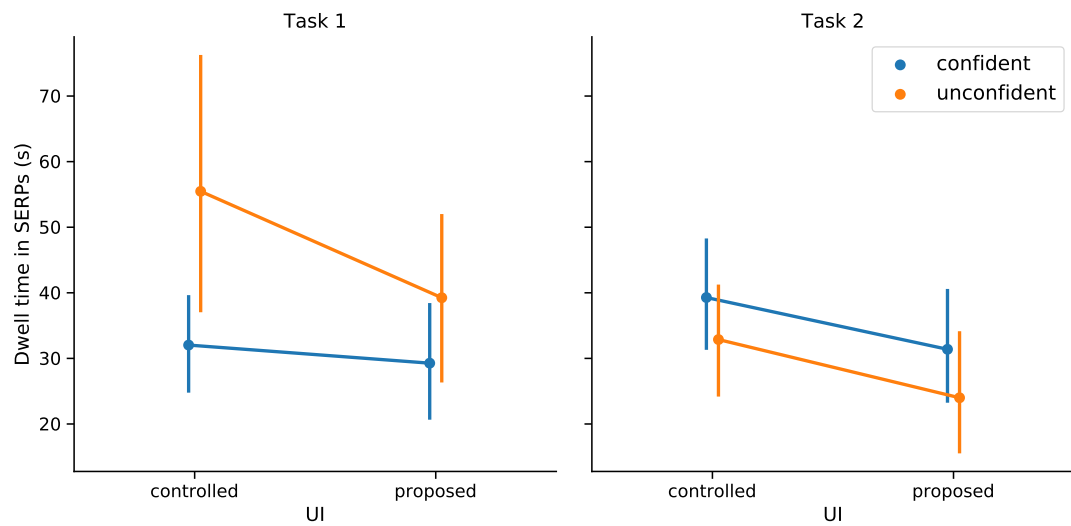


図 5.9 タスク，条件，回答への自信別の検索結果表示ページの閲覧時間（秒）

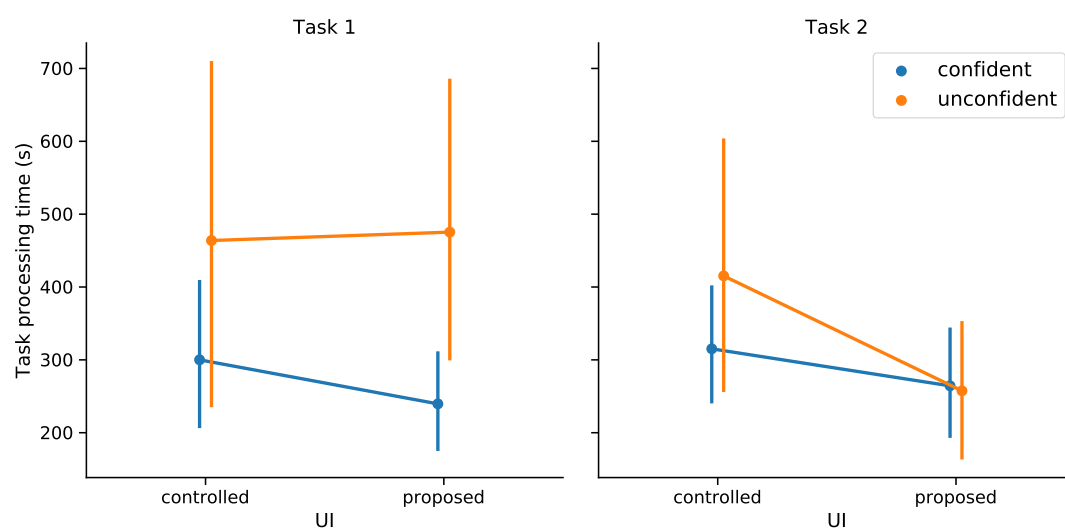


図 5.10 タスク，条件，回答への自信別のタスク所要時間（秒）

第 6 章

考察

6.1 曖昧な文の検出性能

精度，適合率，再現率，F 値全ての指標でスコアは 0.7 を超えた．正解・不正解データの割合は 1 : 1 なので，ランダムに分類を行う場合よりも提案手法の方が性能的に優れていることは明らかになった．しかし，性能改善の余地は大いにあると考える．

実際に Wikipedia の記事で学習器を試し，学習器が曖昧と判断した文の中で，実際に上手く稼動した例と，誤検出してしまった例を表 6.1 と表 6.2 に示す．良い例の 1 つは，Wikipedia の「言葉を濁した表現の典型例」に近い表現が含まれた文を検出できた点である．2 つ目は，「要出典」のテンプレートが付けられた文を取得した点である．Wikipedia には情報について出典を求める「要出典」^{*1}というテンプレートが用意されている．出典を求める理由は，5W1H の「誰が」が不足していること以外にもある．「要出典」テンプレートは今回は教師データの中で明示的に利用しなかったが，そのテンプレートの付いた文を検出した例があったことは良い点である．悪い例の 1 つは，対象として明らかに曖昧でない文を曖昧な文として検出した事である．Wikipedia のカテゴリ名の他にも，記事内の URL や出典の本のタイトルなどを曖昧な文として検出したケースがあった．2 つ目は，その前の文に出典や誰の意見かなどが書かれ，その続きとして展開している文を曖昧な文として検出したケースである．文章として読むと正確性に問題のないものでも，1 文それぞれでチェックをさせることで曖昧な文と検出してしまった．周辺の文や語順についても考慮した特徴量を追加する必要がある．

分類器の性能改善に向けての課題としては，教師データの問題が上げられる．教師デー

^{*1} <https://ja.wikipedia.org/wiki/Wikipedia:出典を明記する>

表 6.1 良い検知例

絵の筆者は不明であるが、一之間の竹虎図は狩野長信の作と推定されている
近年の研究結果が反映された歴史教科書では～（中略）～が記述されている [要出典]

表 6.2 悪い検知例

イギリスの計算機科学者
『復元の名古屋城本丸御殿 将軍御座所に井波彫刻欄間 7 年かけ
忠実再現』北日本新聞 2018 年 6 月 6 日 26 面
言葉を濁した記述のある記事/2018 年 6 月
井伊家の菩提寺・豪徳寺にある墓碑に、命日が「三月二十八日」
と刻まれているのはこのためである

タは Wikipedia で文章表現に問題のある記事から取得した文であるが、文章表現に問題のある記事のカテゴリに偏りがある可能性が考えられる。Wikipedia の記事の編集は全ユーザが可能である*²。記事の編集回数や閲覧回数は記事によって差があり*³、閲覧回数が多い記事は、そのときに話題となっているトピックや人物を扱う記事が多い*⁴。このことから、文章の正確性についてチェックが行われる頻度は記事によって違うことが予想され、言葉を濁した表現であることを示すテンプレートが付けられた文が存在する記事が多く存在しているカテゴリと、そうでないカテゴリがある事が考えられる。分類器の改善に向けてデータの取得の際には、幅広い分野の記事から文を取得することを考慮する必要があると考える。今回は、2018 年 4 月 19 日時点に Wikipedia の記事内に存在していた問題のある文を取得したため、記事の編集履歴から、テンプレートの付いていた文を取得することで、学習データの増加と共にカテゴリの網羅性を改善することが可能になる。また、学習データの中には「曖昧である」というラベリングが誤っているものも少なからずあった。

今後は学習モデルや素性を再検討することで、分類性能のさらなる改善を図る計画である。本研究では、分類器の素性として単純な Bag-of-Words や、Wikipedia の言葉を濁した表現の典型例の有無を入れたのみなので、分類性能の向上には、その他の特徴量についても改善が必要であると考えられる。

*² <https://ja.wikipedia.org/wiki/Wikipedia:記事を執筆する>

*³ <https://ja.wikipedia.org/wiki/Wikipedia:編集回数の多いページの一覧>

*⁴ <https://tools.wmflabs.org/topviews/>

6.2 実験結果

仮説 H1 について、ユーザ実験の結果、ウェブ情報の曖昧な文をハイライトすることで、タスクの所要時間、ウェブページの閲覧時間などが短縮され、曖昧な文のハイライトは仮説と逆の効果があることが示唆された。ハイライト表示を行うことで、被験者はウェブページの情報を取捨選択する際にハイライトの有無を利用したことが考えられる。アンケートによる精査態度に関する分析でも、ハイライトの有無による情報精査態度の差は見られなかったことから、ハイライト表示された文に注目するのではなく、ハイライト表示自体を信憑性判断の目安にされてしまったと考える。被験者は、ハイライト表示の少ないウェブページに信頼できる情報が存在していると考え、そのようなウェブページから情報を取得しようと効率化が行われたと考える。

一方で、ハイライト表示の有無と回答の自信の有無で被験者を分けて分析すると、自信が無いと回答した被験者群には、ハイライト表示を行うことで多くのウェブページを閲覧する傾向が見られた。H3 は部分的に支持されたことになる。しかし、記事ページの平均閲覧時間や結果結果表示ページの閲覧時間を見ると、ハイライト表示をすることで減少傾向にあったことから、回答に自信の無い proposed 群の被験者も、回答に自信のある proposed 群と同じ様にハイライトの有無を信憑性判断の目安として利用したことが考えられる。

仮説 H2 について事後アンケートから、信憑性判断で考慮した要素について、どの項目にも各群の回答に差は見られず、ハイライト表示による情報精査の態度促進は見られなかった。

本稿の実験では、ハイライト表示を行う基準が文章表現の曖昧さのみであり、その情報に信憑性があるかどうかを表すものではない。慎重な情報精査を促す目的でハイライト表示を行ったが、簡単に情報精査行動を終えてしまったことから、被験者は文章内容の精査をハイライト表示に任せてしまったと考える。ハイライト表示の基準や役割をより詳細に知らせておくことで行動、態度が変化すると考える。山本ら [12] は、注意深い情報探索を促す為に、検索時に「根拠」、「比較」などの批判的思考を促す単語を検索クエリに補完、推薦する手法を提案し、ユーザ実験にて効果を測定した。その結果、検索回数や検索結果表示ページの閲覧数が増加し、大卒経験のある被験者は参照情報が書かれていることに注目するようになった。検索ワードを入力する際に、情報探索にどのような観点が必要なのかを具体的に示している。このように、ユーザに情報精査をさせるためには、情報精査をするための具体的な方法や詳細な問題点を情報探索中に見せる必要があると考えた。本稿

のユーザ実験では proposed 群の被験者に、タスク開始前にハイライト表示の説明を行ったのみで、情報探索時には正確性が曖昧な文に太字で背景を黄色にする以外に表示を行わなかった。情報探索時に、触れている情報について何が問題であるかを問題のある箇所と共に見せることで、行動、態度変容が起きるのではないかと考える。ユーザにハイライトの有無のみを当てにするのではなく、文を読んで信憑性判断をしてもらうために、ハイライト表示方法の検討が必要である。

実験の結果、図 5.8 や図 5.10 のようにタスクで扱ったテーマによる影響を受けていると考えられる指標があったことから、タスクで調べる必要のあるテーマの種類や、被験者の事前知識を考慮した実験が必要である。

第 7 章

まとめ

本稿では，ウェブ上で情報探索をするユーザにより注意深い信憑性判断と批判的な情報閲覧を促進するために，ユーザが閲覧中のウェブサイトから文の正確性に不足のある文章を検出してハイライトを行うシステムを提案した．提案システムは，正確性の問題のある曖昧な文を検出するために，Wikipedia を教師データとし単語と特有表現の有無を特徴量とした分類器を用いて，閲覧中のウェブページに掲載された曖昧な文を検出する．

評価の結果，提案手法の文の曖昧さを判定性能は F 値で 75% 程度であった．曖昧さの検取性能をさらに高めるために，今後は様々なサイトの文章情報に対応し，提案システムの実用性を高めるために，学習データの文のカテゴリの網羅性を改善する必要がある．また，分類性能の向上のために学習器の特徴量についても再考する必要がある．本研究では，分類器の素性として単純な Bag-of-Words や，Wikipedia の言葉を濁した表現の典型例の有無を入れたのみなので，分類性能の向上には，その他の特徴量についても改善が必要であると考ええる．

ユーザ実験を行った結果，正確さに不足のある曖昧な文にハイライト表示を行うことでウェブページの閲覧時間が短縮され，ユーザはハイライトの有無を基準に情報探索を行うようになった．また，情報精査の態度についてもアンケート調査を行ったが，ハイライトの有無による差は見られず，仮説通りの結果にはならなかった．今後は，ハイライトの表示方法の検討や検索するテーマへの関心，事前知識にも注目した実験を行う予定である．

参考文献

- [1] 田中優子 (2009) 「批判的思考の促進・抑制に及ぼす論法のタイプ, 外的要求, 情報ソースの信憑性の効果」, 『日本教育工学会論文誌』 33(1), pp.63 – 70, 日本教育工学会.
- [2] 後藤康志 (2006) 「学習者の Web 情報に対する批判的思考の発達」, 『日本教育工学会論文誌』 30, pp.13-16, 日本教育工学会.
- [3] Rob Ennals, Beth Trushkowsky, John Mark Agosta (2010), “Highlighting Disputed Claims on the Web”, *Proceedings of the 19th international conference on World wide web (WWW 2010)*, pp.341-350, ACM.
- [4] Q.Vera Liao, Wai-Tat Fu(2014), “Expert Voices in Echo Chambers: Effects of Source Expertise Indicators on Exposure to Diverse Opinions”, *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2014)*, pp.2745-2754, ACM..
- [5] William Yang Wang(2017), “ “Liar Liar Pants on Fire” :A New Benchmark Dataset for Fake News Detection”, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*(Volume 2: Short Papers), pp.422-426L.
- [6] Chih-Chung Chang, Chih-Jen Lin (2011), “LIBSVM: A library for support vector machines”, *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol.2, Issue 3, No. 27, ACM.
- [7] Vladimir Vapnik (1995), “Support-Vector Networks”, *Machine Learning*, Vol.20, Issue 3, pp.273-297 , Kluwer Academic Publishers.
- [8] Yamamoto, Y. and Yamamoto, T. (2018) “Query Priming for Promoting Critical Thinking in Web Search”, *Proceedings of the 3rd ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2018)*, pp.12-21, ACM.

- [9] 山本祐輔, 山本岳洋, 大島裕明, 川上浩司 (2019), 「ウェブアクセスリテラシー尺度の開発」『情報処理学会論文誌：データベース (TOD80)』, Vol.12, No.1, pp.24-37.
- [10] 平山るみ, 楠見孝 (2004), 「批判的思考態度が結論導出プロセスに及ぼす影響-証拠評価と結論生成課題を用いての検討」, 『教育心理学研究』第 52 巻 2 号, pp.186-198, 日本教育心理学会.
- [11] 四谷あさみ, 鷺見克典 (2002), 「ウェブサイトの質に対する評価尺度の開発における予備的検討」, 『情報処理学会研究報告コンピュータと教育 (CE)』62(2002-CE-065), pp.25-32, 2002.
- [12] 山本祐輔, 山本岳洋 (2019), 「批判的なウェブ検索を促進するクエリプライミング」『情報処理学会論文誌：データベース (TOD80)』, Vol.12, No.1, pp.38-52.
- [13] Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J.: “The Aligned Rank Transform for Non-parametric Factorial Analyses Using Only Anova Procedures”, *In Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2011)*, pp.143146, ACM (2011).