

Web 文書の時間・論理関係分析に基づく 情報信頼性判断支援システムの開発と実証実験

岡嶋穰*1, 河合剛巨*1, 中澤聡*1, 村上浩司*2, 松吉俊*3, 水野淳太*3*4, エリック・ニコルズ*4,
渡邊陽太郎*4, 乾健太郎*3*4, 渋谷英潔*5, 中野正寛*5, 宮崎林太郎*5, 石下円香*5, 森辰則*5

*1 NEC *2 楽天技術研究所 *3 奈良先端科学技術大学院大学 *4 東北大学 *5 横浜国立大学
ic_renraku@cps.jp.nec.com

1 はじめに

インターネット上の情報が爆発的に増加し多様化するに伴い、個々のトピックに関連する情報の全体像を把握することが困難になっている。検索エンジンで上位に表示される文書はインターネット上の文書の一部である。一部の文書に記載された情報だけを判断の手がかりにする場合、たまたま閲覧した信頼性の低い情報を信じてしまう恐れが大きくなる。

インターネット上の情報の信頼性を正しく判断するためには、多様な観点の情報を俯瞰して見比べる必要がある。一部の立場の意見だけに注目するだけでなく、賛成意見と否定意見の双方を調べ、それぞれの意見の根拠や影響、互いの対立点などを考え総合的に判断することで、誤った情報を避け、信頼できる情報を見分けることができると考えられる。

我々は、インターネット上のテキストを分析して多様な観点の情報を俯瞰することを可能にする技術を開発するための研究プロジェクトを実施した。その研究成果として情報信頼性判断支援システムを開発した。このシステムは、言論マップ、調停要約、意見変化イベント抽出の3つの機能によってテキストを分析することで、多様な観点からの情報閲覧をサポートし、ユーザの情報信頼性判断を支援する。

アンケート式の評価実験によって一般の評価者にシステムを試用してもらい、多様な観点から判断を支援できる点において既存のキーワード検索による調査方法を上回る評価を得た。

本稿では、まず2章で情報信頼性分析プロジェクトの全体像、およびその中での本研究の位置づけについて説明する。3章で本研究の概要と開発した情報信頼性判断支援システムについて説明する。そして、4章で信頼性判断支援におけるシステムの効果を評価するために実施した実証実験とその成果について説明する。5章でまとめと今後の予定を述べる。

2 Web コンテンツの信頼性分析プロジェクト

情報通信研究機構(NICT)ではWebコンテンツの信頼性分析に関する研究開発プロジェクトを2006年4月から開始した。NICTの研究成果は、情報分析システム WISDOM[1]として公開されている。WISDOMは、ユーザが入力したトピックに関して、情報の内容、発信者、概観等の観点から、関連する情報の抽出・分類等を行うことで信頼性判断を支援する。

信頼性判断を支援する目的においては、さらに、インターネット上の関連するテキストの意味を分析することで多様な観点の情報を俯瞰する技術が必要となる。賛成意見と否定意見の双方を調べ、それぞれの意見の根拠や影響、互いの対立点などを明らかにし総合的な判断を支援するための技術である。

このような技術の実現を目指し、日本電気、奈良先端科学技術大学院大学、東北大学および横浜国立大学が共同で、テキストのより深い意味解析を用いた情報信頼性判断支援技術を研究している。本研究はNICTからの委託研究である。2008年からの3年間に渡り研究開発を行い、その成果として、各共同研究者が開発した技術を統合した情報信頼性判断支援システムを構築した。本稿ではこの情報信頼性判断支援システムについて述べる。

3 情報信頼性判断支援システム

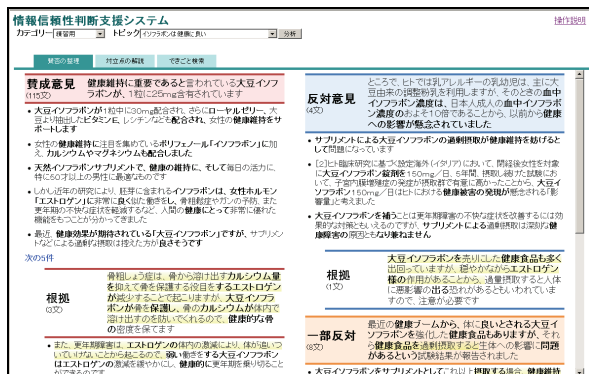
本研究で提案する情報信頼性判断支援システムは、テキスト間の論理的な関係・時間的な関係を分析することで、インターネット上のテキストから信頼性判断に役立つ情報を抽出するシステムである。ユーザは自分が分析したいトピックを短い文の形で入力する。提案システムは、ユーザが入力したトピックに関連するテキストをインターネット上から探し出し、言論マップ、調停要約、意見変化イベント抽出の3つの機能によってテキストを分析した結果をユーザに提示する。

3.1 言論マップ(奈良先端科学技術大学院大学、東北大学)

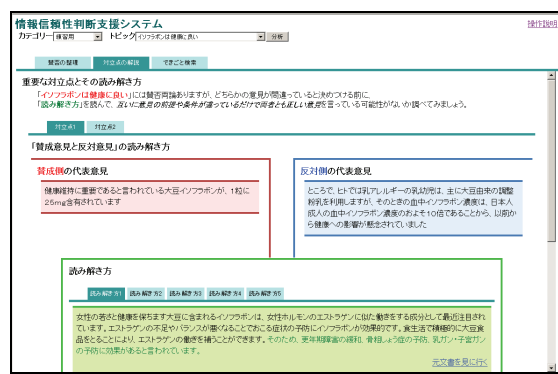
本機能では、ユーザが入力したトピックに関連するインターネット上のテキストを**言明**として抽出し、以下の6種類に分類して表示する(図1(a)) [2]。

- 同意** トピックとおおよそ同じ意味を持つ言明
- 同意の根拠** 同意言明を支持する根拠を述べる言明
- 対立** トピックと同時に成り立たない言明
- 対立の根拠** 対立する言明を支持する根拠を述べる言明
- 弱対立** トピックの範囲や程度を制限する、弱い対立を表す言明
- 弱対立の根拠** 弱対立言明を支持する根拠を述べる言明

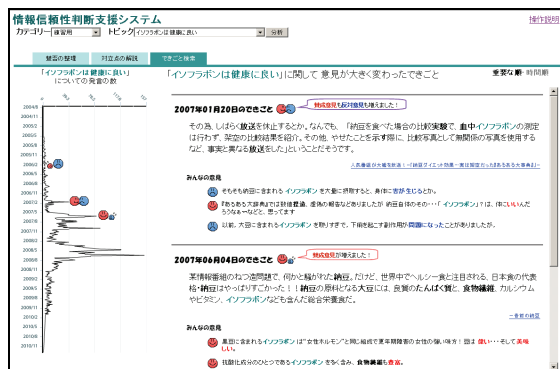
言論マップの特長は、個々の言明を、関連する他の言明との意味的關係性の中に位置づけられる点である。ユーザは個々の言明の根拠となる言明や対立している



(a) 議論マップ



(b) 調停要約



(c) 意見変化イベント抽出

図 1 情報信頼性判断支援システム

言明を知ること、その言明が信頼できるかどうかを他の言明と比較して判断することができる。たとえば「コラーゲンでシワが取れる」のようなトピックを与えると、それに同意する言明や対立する言明、さらには「コラーゲンは肌から浸透しないので、出来てしまったしわに対しての効果は期待できない」のような根拠情報まで、Web から広く収集する。こうした多様な立場からの言明を整理して俯瞰することにより、ユーザは個々の言明の信頼性を相対的な比較によって判断することができる。

3.2 調停要約 (横浜国立大学)

調停要約とは、対立しているようにみえる二つの言明が、他の文章に書かれた条件等を考慮すると実際には共存可能である場合において、共存可能となる状況を簡潔に説明している文章をインターネット上から見つける要約である[3]。調停要約を読むことで、ユーザは賛成意見と反対意見の対立点の読み解き方を理解し、情報の信頼性を判断することができる。

たとえば、「ディーゼル車は環境に良い」というトピックをユーザが入力したときに、「昔のディーゼル車は環境に悪かったが、技術の発達によって最近のディーゼル車はガソリン車よりも環境に良い」といった、対立する意見を読み解くための文章を見つけ、ユーザに提示する。

図 1(b)が、提案システム上に表示された調停要約で

ある。対立している賛成意見と否定意見を表示して、その下に対立点を読み解くための文章を表示している。

3.3 意見変化イベント抽出 (NEC)

本機能では、ユーザが入力したトピックに関して、人々の意見を大きく変えた出来事をインターネット上のテキストから見つけ出して提示する[4]。

人々の意見を大きく変えた出来事は、ユーザの意見にも影響する可能性があり、信頼性判断をする上で知っておくべき出来事だと考えられる。

たとえば「イソフラボンは健康に良い」というトピックが与えられたとすると、インターネット上のテキストを分析し、「イソフラボンの過剰摂取でホルモンバランスが崩れる恐れがあると食品安全委員会が報告した」というような、トピックに関して人々の意見を大きく変えた出来事を見つけ出して提示する。

図 1(c)が、提案システム上に表示された意見変化イベント抽出機能である。人々の意見を変えた出来事を説明する文章と、人々の意見が書かれた文章を抽出して提示している。

4 実証実験

提案システムが情報信頼性判断支援に有用かどうかを確かめるため、一般の評価者にシステムを試用してアンケートに回答してもらった実証実験を行った。

表 1 システム評価に使用したトピックおよび各トピックを選択した評価者の延べ人数

「4トピック」	人数
こんにゃくゼリーは安全だ	68
バナナダイエットは効果がある	64
ディーゼル車は環境に良い	58
アスベストには毒性がある	34
合計	224

「50トピック」(上位4つ)	人数
血液型で性格判断できる	21
アルツハイマーは予防できる	18
レーシックは安全だ	12
飲酒は健康に良い	10
その他(46トピック)	163
合計	224

尚、現在の提案システムは1回の入力に処理結果を返すのに平均5分ほどの処理時間がかかるため、実証実験では分析済みデータを用いて評価している(今後並列処理等の導入により高速化を図る予定である)。

4.1 実験方法

実験対象として112名の評価者を集めた。年齢や性別が偏らないように広く一般から集めた結果、年齢は20代から70代まで分布し、男女はほぼ同数となった。職業は会社員、主婦、自営業の順に多かった。ほぼ全員が毎日インターネットを利用していた。なお、評価者の募集や実験の運営は企業システムのUI評価を専門に扱う企業に委託し、研究者は直接評価者に接触しないことで、評価にバイアスがかからないようにした。

これらの評価者に対し、複数のトピックについて提案システムを用いた調査を体験してもらい、システムの表示結果に関する設問に回答してもらった。

調査対象のトピックとして、4トピックと50トピックの2つのセットを用意した。そして、「4トピック」のセットから2つ、「50トピック」のセットから2つ、合計4つのトピックを評価者自身の興味に従い選択して調査してもらった。「4トピック」と「50トピック」に含まれるトピックの例が表1に示されている。「4トピック」は、同一のトピックを多数の人が評価した場合に評価がどのように分散するかを調べるために少数のトピックを選定したセットである。一方、「50トピック」は、多数の幅広いトピックについての分析結果を評価するため、「健康」「社会・生活」「環境」の3つの分野からトピックを選定したセットである。

比較対象として、提案システムで調査するのと同じトピックについて、既存のキーワード検索サービスによる調査も行うようにした。この調査を**既存検索サービスによる調査**と呼ぶ。使用する検索サービスとしては広く普及している標準的な検索サービスを用いるよう指示した。評価者は、前述のように選んだ4つのトピックについて、提案システムによる調査と既存検索

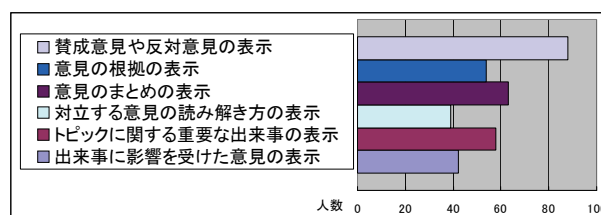


図 2 便利だと思った機能

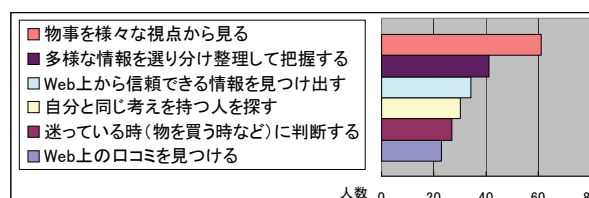


図 3 提案システムがどのようなとき役に立つか

サービスによる調査を交互に行う。交互に行うのは調査順序によるバイアスを防ぐためである。評価者は、提案システムによる調査と既存検索サービスによる調査が終わるたびに、どのような表示結果が得られたかを問う9つの設問に対して、5段階評価で回答する。

4.2 実験結果と考察

評価者が選択したトピックと、各トピックを選択した延べ人数を表1に示す。健康問題など身近なトピックが選ばれやすく、環境問題など身近でないトピックは選ばれにくい傾向がみられた。

調査結果に関する設問内容および結果を表2に示す。設問1,2は言論マップ、3,4は調停要約、5,6,7は意見変化イベント抽出の評価を、8,9はシステム全体の評価を想定している。前半(4トピック)のみと、後半(50トピック)のみと、前半と後半を合わせた総合結果の3通りについて、評価者が5段階で評価した値の平均得点を示している。また、参考までに総合結果のみ得点の分散を示す。表2から、全ての設問において、提案システムが既存検索サービスを上回る評価を得ていることが分る。なかでも設問3や4のような、従来の検索サービスでは見つけにくい情報については、特に大きく評価が上回っている。一方、見つけるのが困難な情報の有無を問う設問であるため、平均得点が低い、得点の分散が大きいなどの特徴が見られる。設問8や9も得点の分散が大きい、これらの設問は評価者の主観的な判断を問うため、個人によって揺れが大きいことを示していると考えられる。

トピックごとの標本数を多く確保できている前半4トピックについては、トピックごとに全設問に関する得点平均と分散を計算し表3にまとめた。「バナナダイエットは効果がある」の分散が比較的大きい。ダイエットのような、個人によって効果が変わる問題については、明確な判断が下しにくく、提示された情報に対する価値判断も多様化するためかと考えられる。

評価者が便利だと思った機能を、複数回答可で尋ね

表 2 評価者によるシステム評価結果

	平均得点						分散	
	前半		後半		総合		総合	
	提案	既存	提案	既存	提案	既存	提案	既存
1 賛成意見や反対意見など、賛否両論を見つげられた	4.17	3.60	3.94	3.47	4.06	3.54	0.71	1.26
2 賛成意見や反対意見の根拠となる情報を見つげられた	3.94	3.29	3.76	3.19	3.85	3.24	0.78	1.19
3 賛成意見や反対意見の両方をまとめている文章を見つげられた	3.77	2.58	3.53	2.71	3.65	2.65	0.97	1.25
4 両方の意見とも正しいという説明がされている文章を見つげられた	3.65	2.58	3.53	2.71	3.51	2.66	1.05	1.25
5 どんな出来事があったか分った	4.17	3.53	3.80	3.32	3.99	3.42	0.85	1.25
6 トピックについて考える上で知っておくべき出来事が見つかった	3.67	3.17	3.39	3.20	3.53	3.18	0.88	1.17
7 出来事が人々の意見にどんな影響を与えているか分かった	3.89	3.08	3.61	2.97	3.75	3.03	0.94	0.99
8 これまで知らなかった意外な情報が見つげられた	3.69	3.17	3.50	3.27	3.59	3.22	1.20	1.28
9 各意見の正しさを判断する上で役に立った	3.86	3.30	3.60	3.39	3.73	3.35	0.97	1.15
設問全体での平均	3.87	3.15	3.61	3.14	3.74	3.14	0.93	0.99

た結果が図 2 である。「賛成意見や反対意見の表示」が最も評価が高く、他の機能も一定の評価を得た。

また、自由回答で、提案システムが特に役に立った点と、良くなかった点について尋ねた。役に立った点としては、多数の意見をまとめて整理して閲覧できることが挙げられた。従来の検索結果では上位に表示されない情報についても短時間で閲覧でき、より公平な判断ができるとの評価が得られた。

良くなかった点としては、賛成・反対の分類誤りなどの分析精度に関する問題や、追加で情報を調べられないなど実装上の問題が挙げられた。情報が多すぎて読みづらいなどインタフェースの問題も指摘された。

図 3 は提案システムがどのような時に役に立つか尋ねた結果である。提案システムは、信頼できる情報を見つけ出すためというより、様々な観点から多様な情報を相対的に位置づけるためのものとして評価されている。本システムを使って様々な観点の情報を調べる場合、ひとつの情報が絶対的に信頼できると考えることはむしろ少なくなり、それぞれの主張の長所や欠点を相対的に把握できるようになるためと考えられる。

アンケート全体への感想としては、「今まで知らなかった知識が得られ有意義だった」等の肯定的評価が多く得られた。また、「今後は賛成意見と否定意見を両方確認したい」「根拠まで調べるようにしたい」などの感想がみられた。このことは、多様な情報を比較した上での判断を行えるように、提案システムが評価者を支援できた結果と考えられる。

支援システムを今後も使用したいか尋ねたところ、他の検索エンジンとの併用を含め、71%が今後も使用したいと回答した。14%は、分析精度やインタフェースが改善されれば使用したいと回答した。

5 まとめと今後の予定

インターネット上の情報の信頼性判断を支援する技術の研究プロジェクトにおいて情報信頼性判断支援のためのシステムを開発した。本システムは、言論マッ

表 3 トピックごとの評価 (前半 4 トピック)

「4 トピック」	平均得点		分散	
	提案	既存	提案	既存
こんにやくゼリーは安全だ	3.94	3.16	0.68	1.12
バナナダイエットは効果がある	3.74	3.12	1.22	2.18
ディーゼル車は環境に良い	3.94	3.13	0.78	1.16
アスベストには毒性がある	3.77	3.15	0.95	0.99
4 トピック全体での平均	3.87	3.15	0.76	1.12

プ、調停要約、意見変化イベント抽出の3つの観点からインターネット上のテキストを分析した結果を提示することで信頼性判断を支援する。アンケート式評価実験において一般の評価者に試用してもらい、多様な観点から判断を支援できる点において既存のキーワード検索による調査方法を上回る評価を得た。

提案システムは、今後、インターネット上で試験的に期間を定めて一般公開する。実証実験よりもさらに広い層に試用して頂き、今後の研究開発の参考とするためである。今年(2011年)2月に、分析済みトピックから選択入力する形式での一般公開を行う。その後、段階を踏んで、ユーザが自由に分析したいトピックを入力できる形式で公開することを検討している。

謝辞

本研究は、独立行政法人情報通信研究機構(NICT)の委託研究「電気通信サービスにおける情報信頼性検証技術に関する研究開発」の成果である。

参考文献

- [1] 黒橋禎夫: 情報の信頼性評価に関する基盤技術の研究開発, 人工知能学会誌, Vol.23, No.6, pp.783-790, 2008.
- [2] 水野淳太, Eric Nichols, 渡邊陽太郎, 村上浩司, 松吉俊, 大木環美, 乾健太郎, 松本裕治: 言論マップ生成技術の現状と課題, 言語処理学会第17回年次大会, 2011.
- [3] 中野正寛, 渋谷英潔, 宮崎林太郎, 石下円香, 永井隆広, 森辰則: 直接調停要約自動生成システム HERMeS の実装と性能評価, 言語処理学会第17回年次大会, 2011.
- [4] 河合剛巨, 岡嶋穰, 中澤聡: Web 文書の時系列分析に基づく意見変化イベントの抽出, 言語処理学会第17回年次大会, 2011.