

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年3月31日現在

機関番号：34304

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21700120

研究課題名（和文） 情報発信者および受信者の信頼性評価尺度分析に基づく情報推薦技術に関する研究

研究課題名（英文） A Development of Information Recommendation System based on Credibility Analysis of Writer and Reader

研究代表者

河合 由起子 (KAWAI YUKIKO)

京都産業大学・コンピュータ理工学部・准教授

研究者番号：90399543

研究成果の概要（和文）：本研究課題では、情報発信者が話題に対して抱く感情値を抽出し、各発信者の立場の違いによる書き方の相違を検出し、さらに受信者の評価尺度を決定することで、その話題に対する新たな記事の中からユーザにとって信頼性の高い記事を提供できる情報推薦技術の研究開発を行った。

研究成果の概要（英文）：This research is focusing on a development of information recommendation system based on credibility analysis by using sentiment feature of writer and reader. Furthermore, this system can provide credible articles to each user based on different opinion between news site sentiments.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	900,000	270,000	1,170,000
2010年度	1,100,000	330,000	1,430,000
2011年度	1,200,000	360,000	1,560,000
年度			
年度			
総計	3,200,000	930,000	4,160,000

研究分野：総合領域

科研費の分科・細目：メディア情報学・データベース

キーワード：Web, 情報推薦, 情報検索, パーソナライゼーション, 感情分析

1. 研究開始当初の背景

ユーザに有益な情報を推薦する技術開発は、Webにおいて日進月歩で行われている。情報推薦は大きく協調フィルタリングとベクトル空間モデルに分類することができる。書籍販売サイト等で採用されている協調フィルタリングは、情報の内容に関係なくアイテムの評価値のみで類似性を判断するため低コストであるが、半面、評価数が少ないと精度が低いため速報性の高い記事配信では用いられることはあまりない。一方でベクトル空間モデルは、情報の内容から特徴量を抽出してベクトルの距離が近い情報を推薦する手法であり、新たな情報も効率よく評価で

きることから、MSN Newsbot等にみられる記事推薦システムに広く利用されている。ところが、記事のテキストで抽出される特徴量は単語の出現頻度が用いられるため、出現しない単語に対する評価ができず、話題に対して抱く感情のような深層の特徴量を発見することができないという問題があった。そこで、応募者は、近年研究が盛んになりつつある感情抽出に関する研究（文献[1]～[3]）に取り組んできた（文献[1]）。記事から書き手の感情を特徴量とする感情値を抽出する手法は、辞書を用いて記事を構成する各単語に対する感情値を決定し、それらの統計情報を基に記事単位に算出する。用いられる辞書

は、シードとなる単語に感情値を定義し、過去200万件強の記事からシードと共起する単語に対してシードの単語と同じ感情値を算出することで構築した。

一方、発信される情報の信頼性判定では、従来、スパムサイトや成人サイトといったいわゆる有害サイトの発見が中心であったのに対し、これまであまり重要視されなかった新聞社や官公庁等の公共性の高いサイトの情報にも強く求められるようになってきている。従来、信頼性の高い情報の発見手法は、ページの内容の更新頻度やページ間の相違に基づいた特徴量抽出手法が一般的である。しかしながら、ベクトル空間モデルを用いた従来の情報推薦と同様に、各ページを発信する書き手や読み手の立場の違いによって情報の書かれ方や読まれ方が異なるため、書き手の立場の違いに基づいた内容の相違といった深層的特徴量抽出はできなかつた。また、これまでの手法では内容の相違を提示することで信頼性の支援は可能であったが、読み手となるユーザの判断基準に基づいて信頼性の高い情報を推薦することはできないという問題があった。

2. 研究の目的

本研究課題では、情報発信者が話題に対して抱く感情値を抽出し、各発信者の立場の違いによる書き方の相違を検出し、さらに受信者の評価尺度を決定することで、その話題に対する新たな記事の中からユーザにとって信頼性の高い記事を提供できる情報推薦技術の研究開発を目標とする。本研究課題を達成するための具体的な要件は以下の通りである。

- (1) 各記事から話題に対する発信者の意図（話題に対して抱く感情）抽出技術の開発
- (2) 発信者の立場の違いに基づく意図提示技術および信頼性判断支援システムの開発
- (3) 閲覧時操作に基づいた話題に対する受信者意図抽出技術と高信頼性記事推薦システムの開発

以下では、信頼性判定のための感情抽出ならびに評価尺度抽出法について述べる。まず、応募者がこれまで取組んできた文章からの感情値抽出技術を向上させることで、記事ごとの書かれ方となる8つの感情値（明るい⇔暗い、嬉しい⇔悲しい、承認⇔拒否、怒り⇔恐れ）を抽出する（要件①）。次に抽出した感情値を用いて、立場となる記事集合（ドメイン）をユーザが動的に設定できる機構を構築し、設定されたドメインごとに話題に対する感情値を算出することで、立場の違いによる発信者の書き方の違いを提示し、ドメインごとに関連する記事を推薦するシステムを開発する（要件②）。要件①と②の達成により、例えば「イラク戦争」に対して

抱く感情は新聞社あるいは各国といった発信者の立場によって異なるが、この相違が分かることで、話題に対する記事の信頼性の判断ができると考えられる。つまり、任意の話題に対して常に感情が異なる立場の発信者が、その話題の新着情報に対して①同じ内容の記事を発信した場合と②異なる内容の記事を発信した場合では、前者①の内容の記事が「信頼性が高い」と判断する。また、読み手となるユーザごとに話題に対する捉え方も異なるため、各ユーザの話題に対する感情値も要件①と同様に抽出することで、各ユーザの信頼性判断の基準値を決定する。さらに、ベクトル空間モデルに基づき抽出した評価尺度とドメインごとの感情値との相関性を判定し、そのユーザの評価尺度にあった信頼性の高い記事を推薦することも今回の開発目標の一つである（要件③）。

これまで応募者は記事のテキストから感情値ならびにユーザの嗜好性に合わせた情報推薦技術を実証しており（業績参照）、以上の要件①～③を全て達成することで、情報発信者と受信者の立場の違いに基づいた信頼性判定ならびに高信頼性記事推薦システム構築の実現を目指す。

3. 研究の方法

本研究では、情報発信者・受信者が話題に抱く感情を抽出することで、各発信者の話題に対する捉え方の違いを検出し、高信頼性情報を提示できる情報推薦技術の研究開発を目的とする。研究計画の概要を図1に示す。

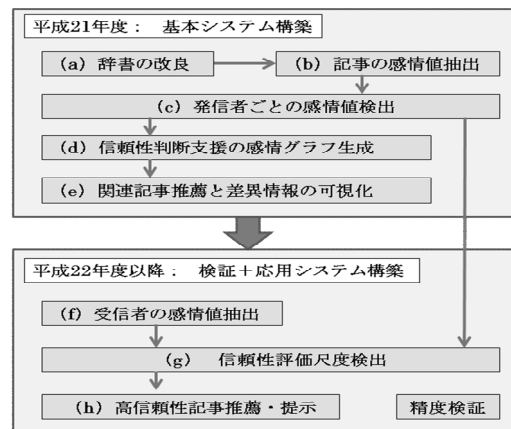


図1：研究計画概要

平成21年度は、提案するシステムの基本的な実装を行うと同時に、各情報発信者の感情値抽出ならびに信頼性判断基準となる感情グラフを立場（ドメイン）ごとに提示する差異情報可視化機構を完了する。

平成22年度以降は、受信者の感情値を抽出し、平成21年度に抽出した発信者の感情値との相関関係から信頼性評価尺度を検出する。検出した評価尺度を用いることで高信

頼性記事を提示する推薦機構を実装する。

また、平成 22 年度以降は、高信頼性記事推薦の精度実験を行い提案手法の検証も同時に行う。

4. 研究成果

初年度は提案するシステムの基本的な実装を行った。これまでに記事の感情値を抽出する際に辞書を利用してきたが、この辞書は共起関係にのみ基づいて構築されていたため、実際に信頼性に基づく記事推薦システムに用いるためには精度の高い辞書構築を行う必要がある。また、本研究では、発信者の立場（ドメイン）ごとに感情値を算出する必要があり、各記事の感情値を用いてドメインごとの感情グラフを動的に生成し、効果的にその差異を提示する必要がある。以下に研究開発した各詳細について述べる。

(a) 辞書の改良

研究代表者がこれまでにを行った基礎研究により、記事のテキストから感情値を抽出できることが確認されている。しかしながら、用いた辞書は 200 万記事から共起関係に基づき構築されているが、精度が高くなく、被験者評価実験と比べて感情値の誤差が大きかった。これは、各記事内の共起関係だけでは、対極の感情値をもつ単語が連結している場合を考慮できないためと考えられる。そこで、これまでの共起関係に基づく感情値算出手法だけでなく、単語 bigram を生成することで、前出の単語と後出の単語との関係を考慮し、誤差の少ない感情値の算出を行った。具体的には、前出する動詞／形容詞／判定詞の単語に関しては感情値判定は行わないことで生成した。ただし、前出するサ変名詞までは含むものとする。これにより、高精度の辞書構築を効率よく行えた。

(b) 記事の感情値抽出

これまでに行った基礎研究より、5 社以上の Web ニュースサイトから国内・国外のスポーツ/時事/政治のカテゴリ内の記事（1 サイト 200 記事程度）を 1 日 1 回収集し、効率よくデータベースで管理できた。また、収集した記事のみを抽出して形態素解析し、tf・idf 値を算出し、tf・idf 値より各記事の特徴語を抽出し、(e)の記事推薦のランキングに利用した。同時に、(a)で構築した高精度の辞書を用いて、各記事の感情値を算出した。この際に、(a)と同様に、単語 bigram を生成することで記事に対する高精度の感情値を効率良く抽出できた。

(c) 発信者ごとの感情値検出

発信者の立場の違いによる内容の違いを可視化することで、信頼性判断支援を行った。そのために、発信者の立場となるドメインをユーザが柔軟に設定できる地図インタフェースを構築した（図 2）。ユーザは地図の拡

大/縮小により、話題に対して、各 Web サイト（県）の感情値、地域ごとの感情値、国ごとの感情値を確認できる。これにより、話題に対する発信者ごとの感情値の相違を容易に把握できるようになった。一般的に地図の拡大/縮小レンジは固定であるので、オフラインで各ドメイン（県、地域、国）のレンジを定義し、感情値算出のための記事集合範囲を決定する。システムはユーザが入力した話題語と地図のレンジ情報の両方を受け取ると、各ドメイン内の記事集合から (b)の各記事の感情値の平均値を算出し、発信者ごとの感情値として検出する。なお、地図の API として google と yahoo!Map の多様な仕様を用いている。

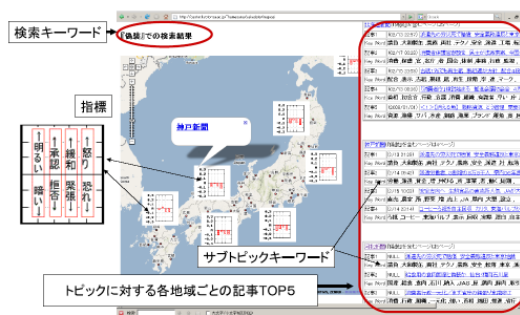


図 2：地理的感性値マップ

(d) 信頼性判断支援のための感情グラフ生成と (e) 関連記事推薦と差異情報の可視化

信頼性判断支援のために (c) で検出したドメインごとの話題に対する感情値をグラフ化し、地図にマッピングする。感情値はこれまでの基礎研究から 4 つの値(明るい⇔暗い、嬉しい⇔悲しい、承認⇔拒否、怒り⇔恐れ)を算出しており（業績 [5]～[7][12][17]～[18]）、これらの値を 4 本のグラフとして提示する。また、2 値化したグラフも作成することで、比較検証も行った。

感情グラフの地図上での提示と同時に、各ドメインごとの記事を推薦する。ここでの記事推薦は、話題語の tf・idf 値の高い順に記事をランキングしてそのタイトルを提示する。同時に関連キーワードとして (b) で抽出した特徴語を提示する。ユーザはタイトルを選択することで記事を閲覧できる。尚、この話題語と記事選択の閲覧履歴はユーザプロフィールとして保存される。

(f) 受信者の感情値抽出

ユーザプロフィールの話題語と記事選択の閲覧履歴に基づき、話題ごとに閲覧した記事集合を特定した。各記事集合から感情値の平均値を算出し、これをユーザの話題に対する感情値とする。ユーザプロフィールの生成ならびに閲覧履歴に基づく選好抽出は行っているため、それらを用いて各記事集合から話

題に対する感情値を算出した。

(g)信頼性評価尺度検出と(h)高信頼性記事推薦・提示

発信者ごとの感情値を基に、話題に関する各記事の信頼性判定機能を実装した(図3)。信頼性の判定は、閾値を基に各ドメインの感情値の偏りを検出し、信頼性を判定し、ユーザへ推薦する。この閾値となる評価尺度は、ユーザの感情値と各ドメインの感情値との相関関係に基づき検出する。また、(f)と同様に、選好に基づく情報推薦を行っている。

なお、(d)~(e)の機能を応用し、(d)の地図上の感情グラフと(g)の高信頼性の記事も提示した。推薦された記事を開覧すると、ユーザプロフィールは再構成され、(f)のユーザの感情値と(g)の評価尺度を再帰的に検出した。



図1：研究計画概要

以上の成果より、話題に対する情報発信者と情報受信者(ユーザ)の意図(話題に抱く感情)を抽出することで、各発信者の話題に対する捉え方の違いを検出でき、受信者にとって信頼性の高い情報を推薦でき、本研究の実用性および有効性を確認できた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計7件)

- ①熊本忠彦, 河合由起子, 田中克己, 新聞記事を対象とするテキスト印象マイニング手法の設計と評価, 電子情報通信学会論文誌(D), Vol. J94-D, No. 3, pp. 540-548, 2011年3月. 査読有
- ②張建偉, 河合由起子, 熊本忠彦, 田中克己, 地域性に基づく発信者の観点差異を可視化するセンチメントマップシステムの提案, 情報処理学会論文誌: データベース, Vol. 3, No. 1 (TOD 45), pp. 38-48, 2010年3月. 査読有
- ③Jianwei Zhang, Yukiko Kawai, Tadahiko

Kumamoto, and Katsumi Tanaka, "A Novel Visualization Method for Distinction of Web News Sentiment", Proc. Tenth International Conference on Web Information Systems Engineering (WISE 2009), LNCS 5802, pp. 181-194, October 2009. 査読有

〔学会発表〕(計4件)

- ①松本好史, 張建偉, 河合由起子, 中島伸介, ニュースサイトにおける話題およびセンチメント差異提示システムの検討, 情報処理学会創立50周年記念全国大会(第72回全国大会), 2010年3月10日.

〔図書〕(計1件)

- ①Yuya Matsui, Yukiko Kawai, and Jianwei Zhang, "Page as a Meeting Place: Web Search Augmented with Social Communication", Xu Huang, Oscar Castillo, and Sio-long Ao (Eds.): Intelligent Control and Innovative Computing, Springer, pp. 303-317, 2012.

6. 研究組織

(1)研究代表者

河合 由起子 (KAWAI YUKIKO)

京都産業大学・コンピュータ理工学部・
准教授

研究者番号: 90399543

(2)研究分担者

なし

(3)連携研究者

なし