

機関番号：34304

研究種目：若手研究（B）

研究期間：2008～2010

課題番号：20700089

研究課題名（和文）

背景データの解析に基づく高信頼性Webコンテンツの取得・呈示に関する研究

研究課題名（英文）

Acquisition and presentation of credible Web contents based on their background data.

研究代表者

中島 伸介（NAKAJIMA SHINSUKE）

京都産業大学・コンピュータ理工学部・准教授

研究者番号：90399535

研究成果の概要（和文）：

ブログ著者コミュニティの発見・解析に基づくブログ関連コンテンツ評価を踏まえて、本手法による高信頼性コンテンツの選択的な取得方法について検討すると共に、その機能開発に取り組んだ。また、Web上で公開されているソーシャルブックマークシステム（Webページ等）より、データを取得し、これらを利用した実験システムの開発に取り組んだ。結果として、ある検索要求に対して切り口の異なる高信頼性情報の呈示が可能なシステムの構築を行った。

研究成果の概要（英文）：

I tried to evaluate blog contents based on detecting and analyzing blogger communities. I proposed a method to obtain highly credible contents and implemented the function as well. Moreover, I tried to develop prototype Web recommender system using social bookmark data. Finally, the system that can represent highly credible and various information with several aspects is implemented.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	900,000	270,000	1,170,000
2009年度	1,100,000	330,000	1,430,000
2010年度	800,000	240,000	1,040,000
年度			
年度			
総計	2,800,000	840,000	3,640,000

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：情報検索，Webマイニング

1. 研究開始当初の背景

大量かつ多様な情報から価値の高い情報のみを効率良く取得しようとする試みは、これまでもWeb検索エンジンの改良を中心に盛んに行われてきた。その中で、最も成功したもの1つが、Googleであるが、多くのユーザはGoogleに満足している訳ではなく、価値の高い情報をよりの確に検索できるような新たな方式が望まれている。Googleをはじめ、従来の検索手法の多くは、検索対象で

あるWebコンテンツを解析対象としている。しかしながら、コンテンツ作成者がある種のテクニックに基づいてコンテンツの内容を編集することができるため、その解析結果は公平かつ公正なものとは言い難い。そこで私は検索対象コンテンツを直接的に解析するのではなく、検索対象コンテンツの背景データを解析することで、公平かつ公正な価値評価を行うことで、信頼性の高いコンテンツの取得を試みることを考えた。

2. 研究の目的

研究代表者は、これまでに重要なブログ著者の発見と信頼性の高い情報の取得を目指した研究を行ってきた。すなわち、ブログ記事を直接解析するのではなく、ブログ記事の背景データの一部であるブログ著者の解析により、元のコンテンツの評価を行うものである。この研究の中で、背景データの解析に基づくコンテンツの信頼性評価手法の有効性を確認したが、まだ多くの課題があり実用システムの構築には至っていない。そこで今回は、実用化を実現するためにも、ブログ検索エンジンを有する、きざしカンパニー社と連携することで、大量のデータに対する実証試験を行いながら、ブログ著者の解析に基づくブログ関連コンテンツの評価を行うことを考えた。

また、Web コンテンツの背景データの利用例として、ソーシャルブックマークにおけるタグシステムへの注目が高まっている。Web ページでは、del.icio.us、動画では YouTube 等が有名である。このようなタグ情報は、コンテンツ作成者ではなく、第三者によって付与されるものであり、Web コンテンツの重要な背景データの1つであると考えられる。既にタグ解析に基づいた信頼性の高い Web コンテンツの推薦方式に関する研究準備も進めている。

したがって、上記のようにコンテンツに密接に関連した背景データを解析することによる、対象コンテンツの信頼性評価に関する研究を行うことを考えた。最終的には、高信頼性 Web コンテンツの取得・呈示システムの開発に取り組む。

3. 研究の方法

(1) ブログ著者コミュニティの発見および解析に基づくブログ関連コンテンツ評価

我々はコンテンツ作成者の特定が容易であり、その作成者による過去のコンテンツの取得が容易であるブログに着目した。初めに、あるトピックに関するブログ記事の投稿数により、ブログ著者の熟知度を算出する。そして、この熟知度に基づいたブログ著者コミュニティの抽出を試みる。また、熟知度の高いブログ著者の多くから、リンク参照などにより注目されているブログ著者は、このコミュニティにおけるオピニオンリーダー的な存在であると考えられ、熟知度の高いブログ著者よりも、さらに重要な存在であると認定する。このようにして、コミュニティにとって重要なブログ著者を特定することにより、その著者が投稿した記事や、そこで言及されている Web コンテンツの評価を行う。さらに、各トピックにおいて信頼性の高いコンテンツを評価し、これを取得する手法の確立を目指す。なお、ブログのみならず、他のコンテ

ンツに対してもその作成者情報を解析することでコンテンツの信頼性評価を行えるような、汎用性および拡張性の高い手法として開発することを心がける。

(2) ソーシャルブックマークにて付与されるタグ情報の解析に基づくコンテンツ評価

ソーシャルブックマークによって扱われるコンテンツにおいて、これらに付与されているタグ情報は、そのコンテンツの特性を表現する背景データとして重要である。しかしながら、あるコンテンツに対して付与されるタグが表す意味は、各々のユーザの感じ方が異なる為に多様であると考えられる。したがって、あるトピックに関して同様な感じ方をするユーザ（トピック依存の類似ユーザ）を、このタグ情報を解析することで効率良く発見する手法を検討する。さらに、この類似ユーザ間で共有するコンテンツを、付与されたタグに基づいて評価することにより、各ユーザにとって信頼性の高いコンテンツを評価し、取得する手法の確立を目指す。

(3) 背景データの解析に基づく高信頼性 Web コンテンツの検索・呈示システムの開発

ここでは、対象コンテンツの背景データ解析に基づいて実現される、前述の2テーマによる Web コンテンツの信頼性評価を効率的に統合する。これにより、ある特定のトピックに関する高信頼性情報を取得したいというユーザの要求に対し、適切なコンテンツの呈示を可能とするシステムの構築を目指す。すなわち、複数の信頼性評価の手法を有機的に利用することにより、ある検索要求に対して切り口の異なる結果の呈示を目指すと共に、Web ページのみ、動画のみとメディアを限定することなく、各種メディアによる結果の呈示を目指す。これにより、ユーザの検索要求に対して、複数の観点から、複数のメディアによって結果を呈示することができ、ユーザ自身が呈示結果の信頼性を認識しつつ、コンテンツを利用することが可能となる。

4. 研究成果

(1) ブログ著者コミュニティの発見および解析に基づくブログ関連コンテンツ評価

ブログ著者のあるトピックに関する熟知度に基づいて、ブログ著者コミュニティを適切に抽出する手法について検討すると共に、このコミュニティにとって価値のあるコンテンツの取得方法を検討した。

ブログ著者のあるトピックに関する熟知度に基づいて、ブログ著者コミュニティを適切に抽出する手法の実験システムを構築し、提案手法の有効性の評価を行った。

ブログ著者コミュニティの発見・解析に基づくブログ関連コンテンツ評価を踏まえて、

本手法による高信頼性コンテンツの選択的な取得方法について検討すると共に、その機能開発に取り組んだ。さらに、ユーザ側でコンテンツの信頼性評価が可能な背景データの提示を実現するための機能拡張を行った。

図1に研究協力者と共に開発したブロガーの特性分析を行うシステムblogram.jpを示す。このサイトでは、ブログURLを登録するとそのブロガーの特性を種々の側面から分析した結果を提示することが可能である。

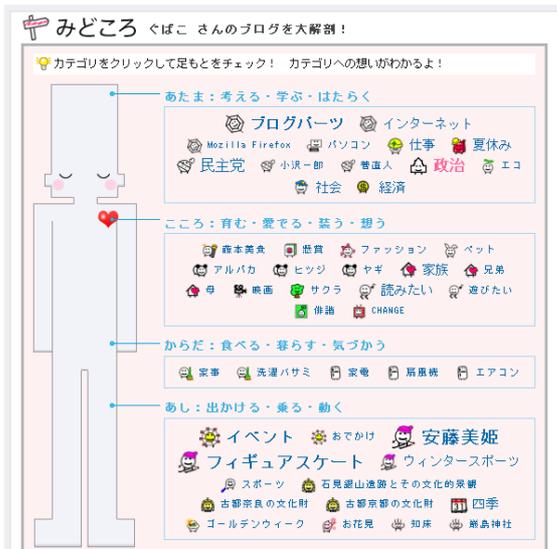


図1 : blogram.jp

(2) ソーシャルブックマークにて付与されるタグ情報の解析に基づくコンテンツ評価

Web上で公開されているソーシャルブックマークシステム (Webページ等) より、データを取得し、これらを利用した実験システムの開発に取り組んだ。この中で、付与されるタグとタグを付与した他のユーザを解析することで、各ユーザの嗜好に合致すると共に価値の高いコンテンツの取得方法について検討した。

Web上で公開されているソーシャルブックマークシステム (Webページ等) より、データを取得し、これらを利用した実験システムを用いた評価方法を検討すると共に、提案手法の有効性に関する評価を行った。

図2に開発したシステムのインタフェース画面を示す。本システムではユーザが閲覧しているWebページ (右部) に対して、ユーザが好むWebページを提示することができる。図3に従来技術との比較実験結果を示す。推薦精度の適合率を比較した結果、提案手法 (RCF) が最も良い結果を示している。



図2 : タグ情報を利用したWebページ推薦システム

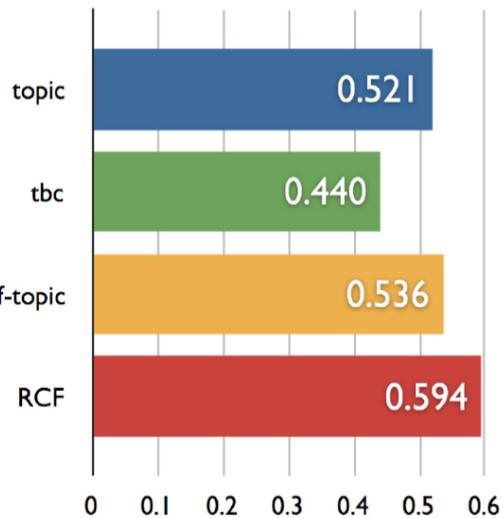


図3 : 従来技術との比較 (適合率)

(3) 背景データの解析に基づく高信頼性Webコンテンツの検索・提示システムの開発

高信頼性 Web コンテンツ検索・提示システムの開発準備を進め、前述の2テーマの状況に応じて、実験システムの開発を行った。

高信頼性 Web コンテンツ検索・提示システムのプロトタイプ構築を行い、ユーザ評価実験を行うと共に、システムの改良に向けた検討を行った。

さらに Web コンテンツの信頼性評価に関する2つの手法を有機的に統合したシステムの開発を継続的に行った。すなわち、ある検索要求に対して切り口の異なる高信頼性情報の提示が可能なシステムの構築を行うと共に、その精度向上に努めた。

図4に高信頼性ブログコンテンツの提示システムを示す。このシステムでは、ある検索キーワードに関して書かれているコンテンツに対して、関連する幾つかの熟知度ランキングを提示することができ、異なる切り口での情報閲覧を実現している。



図4：高信頼性ブログコンテンツの呈示システム

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3件)

- ① 稲垣陽一, 中島伸介, 張建偉, 中本レン, 桑原雄, ブLOGGERの体験熟知度に基づくブログランキングシステムの開発および評価, 情報処理学会論文誌: データベース, Vol. 3 No. 3 (TOD47), pp. 123-134, 2010年9月. (査読有)
- ② 奥健太, 中島伸介, 宮崎純, 植村俊亮, 加藤博一, 「情報推薦におけるユーザの価値判断基準モデルに基づくコンテキスト依存型ランキング方式」, 情報処理学会論文誌: データベース, Vol. 2 No. 1 (TOD41), pp. 57-80, 2009年3月. (査読有)
- ③ 中島伸介, 稲垣陽一, 草野奉章, 高信頼性情報の提示を目指した熟知度に基づくブログランキング方式の提案, 日本データベース学会論文誌 (Journal of the DBSJ), Vol. 7, No. 1, pp. 257-262, 2008年6月. (査読有)

〔学会発表〕(計2件)

- ① 山岡千夏, 中島伸介, 張建偉, 稲垣陽一, 中本レン, ブログ記事の時系列解析に基づく流行語候補「兆し」の早期発見手法, 第3回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2011) A4-1, 静岡県伊豆市, 2011年2月.
- ② Shinsuke NAKAJIMA, Adam JATOWT, Inagaki YOICHI, Reyn NAKAMOTO, Jianwei ZHANG, and Katsumi TANAKA, Finding

Forerunner Bloggers by Temporal Analysis of Blogger Posts, 第3回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2011) A4-6, 静岡県伊豆市, 2011年2月.

- ③ Takumi Shihoya, Shinsuke Nakajima, Kazutoshi Sumiya, and Yoichi Inagaki, Niche Trend Search for Recommender System based on Knowledgeable Blogger Ranking, 2nd Workshop on Recommender Systems and the Social Web (RSWeb2010) in conjunction with ACM Recommender Systems 2010, pp. 56-59, Barcelona, September 2010.
- ④ Kenta Oku, Shinsuke Nakajima, Jun Miyazaki, Shunsuke Uemura, Hirokazu Kato, and Fumio Hattori, A Recommendation System Considering Users' Past / Current / Future Contexts, 2nd Workshop on Context-Aware Recommender Systems (CARS-2010) in conjunction with ACM Recommender Systems 2010, Barcelona, September 2010.
- ⑤ 志甫谷匠, 中島伸介, 角谷和俊トレンド分析および対象グループ粒度推定に基づく情報推薦システム第2回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2010) A3-5, 兵庫県淡路市, 2010年2月.
- ⑥ 桑原雄, 稲垣陽一, 草野奉章, 中島伸介, 張建偉, マイクロブログを対象としたユーザ特性分析に基づく類似ユーザの発見および推薦方式, 情報処理学会: 研究報告-データベースシステム (DBS), Vol. 2009-DBS-149 No. 18, 横浜市, 2009年11月.
- ⑦ Shinsuke Nakajima, Jianwei Zhang, Yoichi Inagaki, Tomoaki Kusano, and Reyn Nakamoto, Blog Ranking Based on Bloggers' Knowledge Level for Providing Credible Information. Proc. of the 10th International Conference on Web Information Systems Engineering. WISE2009, pp. 227-234, Poznan, Poland, October 2009.
- ⑧ 中島伸介, 稲垣陽一, 張建偉, 草野奉章, Reyn NAKAMOTO, ブLOGGERの体験熟知度に基づくブログランキングシステム, 電子情報通信学会 第15回 Web インテリジェンスとインタラクション研究会 WI2-2009-28~45, pp7-12, 広島市, 2009年7月.
- ⑨ 桑原雄, 稲垣陽一, 中島伸介, 張建偉, 重要BLOGGER判定手法を利用した意見分析サイトの構築, 第1回データ工学と情報マネジメントに関するフォーラム

(DEIM2009), i1-35, 静岡県掛川市, 2009年3月.

- ⑩ Kenta OKU, Shinsuke NAKAJIMA, Jun MIYAZAKI, Shunsuke UEMURA, and Hirokazu KATO: A Recommendation Method Considering User's Time Series Contexts, The 3rd International Conference on Ubiquitous Information Management and Communication (ICUIMC2009), pp.495-500, Suwon, Korea, January 2009.
- ⑪ Reyn Nakamoto, Shinsuke Nakajima, Jun Miyazaki, Shunsuke Uemura, Hirokazu Kato, Youichi Inagaki, Reasonable Tag-Based Collaborative Filtering for Social Tagging Systems, Second Workshop on Information Credibility on the Web (WICOW 2008), pp.11-18, Napa Valley, California, USA, October 2008.
- ⑫ 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮, 加藤博一: ``ユーザの時系列コンテキストを考慮した情報推薦方式の提案'', iDB フォーラム 2008: 研究報告-データベースシステム (DBS), Vol.2008 No.88, 福島県福島市, 2008年9月.

6. 研究組織

(1) 研究代表者

中島 伸介 (NAKAJIMA SHINSUKE)
京都産業大学・コンピュータ理工学部・
准教授
研究者番号: 90399535

(2) 研究分担者

なし

(3) 連携研究者

なし