

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：基盤研究（B）

研究期間：2008～2011

課題番号：20300042

研究課題名（和文）放送番組とウェブ情報を用いた低信頼度情報の発見と知識化に関する研究

研究課題名（英文） Research on finding unreliable information and application to knowledge formation using broadcast programs and Web information

研究代表者

宮森 恒（MIYAMORI HISASHI）

京都産業大学・コンピュータ理工学部・准教授

研究者番号：90287988

研究成果の概要（和文）：本研究の目的は、テレビ番組やウェブ情報の中から、信頼性の高い情報を、内容の偏りなく動的に組織化・提示し、知識として活用できる基盤技術を確立することである。主な成果として、低信頼度情報検出のための、信頼性が不確かな情報と基準情報との比較手法、コンテンツの組織化のための、コンテンツ間の関連性発見と情報分布の分析手法、ユーザインタラクションにおける、社会的フィルタリングに着目した信頼性向上手法が挙げられる。

研究成果の概要（英文）：The aim of this research is to develop fundamental technologies to dynamically and impartially organize and provide credible information from TV programs and Web contents for their utilization as knowledge. Its major achievements include a method of comparing uncertain information with the given standard information for efficient detection of unreliable information, a method of detecting relevance between contents and analyzing information distribution for organizing contents, and a method of improving credibility focusing on social filtering in user interactions.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	4,900,000	1,470,000	6,370,000
2009年度	3,900,000	1,170,000	5,070,000
2010年度	3,100,000	930,000	4,030,000
2011年度	2,800,000	840,000	3,640,000
年度			
総計	14,700,000	4,410,000	19,110,000

研究分野：Webマイニング、情報検索、自然言語処理、画像処理

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：情報信頼性

1. 研究開始当初の背景

現在、我々が利用できる情報は増大の一途を辿っており、その中から本当に意味のある情報を探すには多大な労力を要する。特に、機械操作に不慣れな一般利用者にとっては、どの情報をどう見ていけばよいのか途方にくれることも珍しくない。一方、そのような利用者にとっても、信頼性の低い意味不明な

情報は不必要だが、なるべく多様な意見や見方は知りたいというニーズが根強く存在するの事実である。

ウェブコンテンツはその量や多様性といった面では非常に膨大となっているが、その質という意味で、本当に役立つ質の高い情報と何の根拠もない嘘やデマといった質の低い情報が玉石混淆の状態で存在していると

いう問題がある。現在、数々の検索エンジンが利用可能であり、キーワードを含む情報の収集を行うことは容易となったが、一般利用者がそれら検索結果から情報の質を効果的に見分ける手段は存在しているとはいえない。

情報の信頼性については、情報通信機構知識処理グループにおいて、ウェブのテキストデータを対象として、利用者に信頼性判断のベースとなる分析結果を提示することで、利用者自身による信頼性判断を支援する研究が行われている。ここでは、「誰々がどうした」といった述語項構造を処理の基本単位とした自然言語処理に基づく深い意味処理を行うアプローチがとられている。一方、京都大学田中研究室においても同様の研究が行われており、メタサーチエンジン解析やデータ解析といった比較的浅い意味処理によるアプローチがとられている。欧米などでは信頼のモデル提案や社会ネットワークに基づく信頼度計算方法等がいくつか提案されている。

さて、テレビ番組は、従来からプロの番組制作者によって作成された映像コンテンツであり、リアリティや質が極めて高い。しかし、番組制作のためには公共性や視聴率を確保する必要があるため、一定時間内の番組で提供される情報の詳細さや多様さについては限界があり、利用者にとって必ずしも十分な情報が得られないことも多い。

ここでは、テレビ番組の高い品質に着目し、これを一つの既存知識とすることで、ウェブコンテンツ中の高信頼度情報を効率よく分析できるのではないかと考えた。また、ここでは、高信頼度情報を、明らかな誤りや矛盾等を含む情報(低信頼度情報と呼ぶ)以外の情報として広くとらえることで、信頼性の低い情報以外の、多様な意見や見方を含む情報だけをより確実に利用者に情報提供できると考えた。

上記アプローチにより、信頼性の高いウェブコンテンツが利用可能となり、これをさらにテレビ番組と関連付けることで、番組では提供されない詳細かつ多様な情報が補われた、より網羅性の高い知識として利活用できるようになる。これは、健全で安全・安心な情報社会を実現する上で極めて重要な基盤技術になることが期待でき、その影響は計り知れない。

2. 研究の目的

本研究の目的は、テレビ番組やウェブ情報の中から、信頼性の高い情報を、内容の偏りなく動的に組織化・提示し、知識(いわば動的な百科事典)として利活用できる基盤技術を確立することである。具体的には、高信頼度情報を見つけるために、既存知識と整合しな

い明らかな誤りや矛盾等を含む低信頼度情報を検出する技術、内容の偏りを最小限にするために、内容の網羅性が高く重複の少ない最小のコンテンツ群で関連付け、その場で組織化する技術、内容の動的な提示のために、情報の最適な配置やスタイルをユーザインタラクションに応じて変化させ制御する記述言語とその表示技術の確立を目指す。

3. 研究の方法

図1に本研究の概要を示す。図中の「分析・選別」「関連付け、組織化」「柔軟なインタラクション」の各部分について、分担して連携しながら、並行して研究を進める。

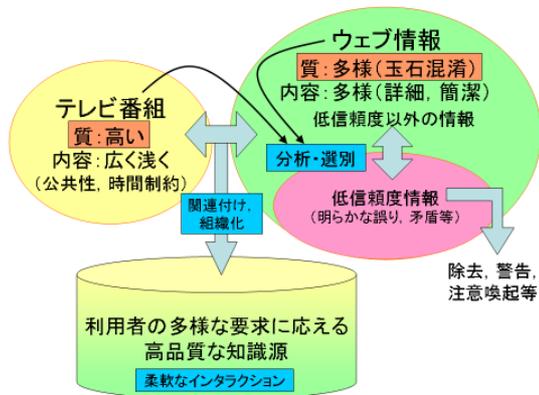


図1. 本研究の概要

(1) 低信頼度情報の検出

テレビ番組やニュース記事に用いられる用語集を作成し、その定義や説明、用語の使い方の実例を収集した基盤データを構築する。書き言葉だけでなく話し言葉での言い換えや、用語に関連して官庁や業界団体等から発表される統計情報、テレビ番組や百科事典等に用いられる図表・解説等についても可能な限り人手あるいは自動で収集する。これらを低信頼度情報検出のための言語的・統計的知識として利用し、該当する用語等の使われ方に誤りがないか、矛盾点がないかを検出する方法を検討する。

(2) 内容のカバレッジ等を考慮したコンテンツの組織化

低信頼度情報が除去されたコンテンツ間の関係を分析し、知識(動的な百科事典)の項目として最適なコンテンツを動的に選別・組織化可能にする手法を開発する。例えば、事象やエンティティなどを対象に、コンテンツの表現モデルとその分析モデルを開発して、事象やエンティティ間の関係を明らかにし、それに基づく知識の動的組織手法について検討する。

(3) 柔軟なユーザインタラクションを提供するブラウザ開発

知識の項目を選択することで、その内容に最適な番組シーンやウェブページを参照したり、利用者の興味に応じて両者の間を自由に行き来したり、関連語句等で両者のコンテンツを適宜絞込み、個々の内容に応じた表示属性や最適なレイアウトを動的に再構成・表示・制御するといった、柔軟で新しいユーザインタラクションを提供するブラウザを開発する。データ記述言語やその自動生成手法を検討し、実装においては既存ブラウザへの組み込みが容易となるように配慮する。

4. 研究成果

(1) 低信頼度情報の検出

信頼性が不確かな情報と基準情報との比較手法、および、テレビ番組を利用した動向情報の信頼性判断支援について主に研究開発を行った。

① 信頼性が不確かな情報と基準情報との比較手法

自分が基準として考える情報(比較基準)に対する比較対象の位置づけを明示する手法を提案した。具体的には、情報を、属性値をもたないラベル集合と重み付き属性値をもつラベル集合から構成される集合と捉え、2つの情報が共通に含む属性と、各情報が固有にもつ属性に基づき定義された非類似度の線形和を用いて比較する手法を提案した。これにより、利用者ごとに異なる観点に沿った比較が実現でき、比較対象が自分の価値基準に合った妥当な情報かどうか判断しやすくなる。

例えば、料理レシピのような情報の場合、値をもたないラベル集合には、名称や著者、日時といった書誌的情報が対応し、値をもつラベル集合には、食材名や分量が対応することになる。2つのレシピが与えられたとき(一方は信頼性が不確かな情報、他方は自分が信頼する基準情報に相当する)、共通して含まれる食材と、それぞれが固有に含む食材に基づき比較が行われることになる。比較対象の固有性や、比較基準に含まれる重要な属性をどの程度含んでいないかといった側面を強調した比較が可能となる。

数十万件の料理レシピを例としたプロトタイプシステムを開発し、基本的な有効性を確認した(図2)。

② テレビ番組を利用した動向情報の信頼性判断支援

Web情報における動向情報を対象に、それと関連するテレビ番組の図表画像を検索することで信頼性判断を支援する手法を提案し、プロトタイプシステムを実装した。ここ

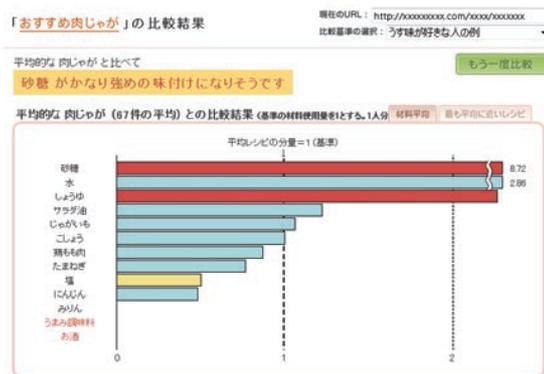


図2. 料理レシピを対象とした比較例

で、動向情報とは、商品の売り上げ推移や内閣支持率の変化等、ある事柄や数値についての時間的変化を表現する言語情報を指し、図表画像とは、その中に折れ線グラフや円グラフ等の図表が主として出現している画像フレームのことである。テレビ番組から図表画像を認識し、その前後の時区間に対応する字幕情報からキーワードを抽出し、索引を構築した。クエリとして、動向情報が入力され、そこから抽出されるキーワードを用いて関連図表画像をランキング提示する。実験の結果、クエリに該当する図表がテレビ番組中に存在する場合、一定の良好な精度で関連図表を提示することを確認でき、動向情報の信頼性判断を効率よく支援することにつながるという見通しを得た。

(2) コンテンツの組織化技術

コンテンツ間の関連性の発見とそれに基づく情報分布の分析手法について研究開発を行った。具体的には、ニュースコンテンツを対象とした話題と視点に基づく情報の分布分析手法、人物に関する記述や評判情報の分布分析手法、ネットワークモデルや構造特徴を利用した因果関係の分析手法を検討し、プロトタイプシステムを通して、提案手法の有効性を確認した。

① 話題と視点に基づく情報の分布分析

ニュース記事に書かれて事象をキーワードの構造体 Content-Structure を用いて表現するモデルを提案した。Content-structure は、subject-term、aspect-term と state-term から構成されるキーワードのタプルであり、語の頻度、品詞、位置および共起関係に基づいて抽出される。直感的に、subject-term は、イベントやアクティビリティの主題を表し、aspect-term と state-term はそのアスペクトと状態を表す。

本研究では、ニュースコンテンツのテキストの全体から抽出される content-structure をそのニュースの話題とし、ニュースの視点

(注目点)を、追加説明や独自素材のよく書かれている最終段落から抽出される content-structure とする。関連ニュースコンテンツを分類して話題や視点の違いを分析し、エントロピーを計算して報道情報の多様性を明らかにしてユーザに提示するシステムを開発した。

②人物に関する記述や評判情報の分析

実世界のエンティティ(人物, 組織, 地域, etc.)がニュースコンテンツの記述対象となることが多いこと。その為、本研究では、ニュースコンテンツにおけるエンティティ、特に人物に対する記述を分析し、情報間の差異関係を明らかにする手法を開発した。

本研究では、人物に関する記述は次の2通りに分けて分析を行っている。

- ・主観的記述：人物に対する直接的な評価, 意見

- ・客観的記述：人物に関する事象, 事実の記述

例えば「北朝鮮が悪い」といった記述は北朝鮮に対する否定的な意見が現れている。一方「北朝鮮がミサイルを発射する」といった記述は単なる事実を述べた記述である。我々は前者を主観的記述、後者を客観的記述と呼ぶ。客観的に記述を事象レベルに分類して比較することによって、客観事実の網羅性を評価することができる。一方、主観的記述を論点とその極性(ポジティブ, ネガティブ)に分類して比較し、論点の網羅性とその極性の差を明らかにすることができる。

③事象間の因果関係分析手法

ニュースコンテンツにおける事象はエンティティと同様、重要な報道対象である。事象間の因果関係の分析は、知識の体系化および情報の由来や背景を理解には重要である。時間の経過に伴う事象と因果関係が変化する。いわゆる、動的である。これに対応するため、因果関係ネットワークを増分的に構築する手法を開発している。記事は日々追加されるので、我々はまず、一定期間に追加された記事から、手がかり表現を用いて因果関係を抽出し、追加された因果関係の中で事象の結合と整理を行い、因果関係ネットワークを構築する。次に、追加された記事から作成した因果関係ネットワークとそれまでに構築された因果関係ネットワーク間で事象ノードのマージとネットワークの整理を行う。これを繰り返すことで、因果関係ネットワークを増分的に構築する手法を開発した。

(3) ユーザインタラクション技術

情報を閲覧する際のインタラクションにおいて重要となる要素を検討し、ユーザの視点(ユーザ基準)に応じた動的な情報閲覧、

および、信頼性などの特性が直感的に理解できる表現方法を中心に研究を行った。

①ユーザの視点(ユーザ基準)に応じた動的な情報閲覧

ユーザ基準に照らした際のデータの合致度の視覚化、および、ユーザ基準に基づく推薦機能を実装した。具体的には、一定の条件のもとで処理対象となるデータと、ユーザが設定した基準に合致するデータの各分布を視覚化し、検出された低信頼度情報の妥当性を俯瞰的に確認できるようにした(図3)。

また、一定の条件のもとでユーザが設定した基準との一致度が高い順に情報を推薦する機能を追加し、ユーザ基準や興味に応じたわかりやすいナビゲーション手段の一部とした。

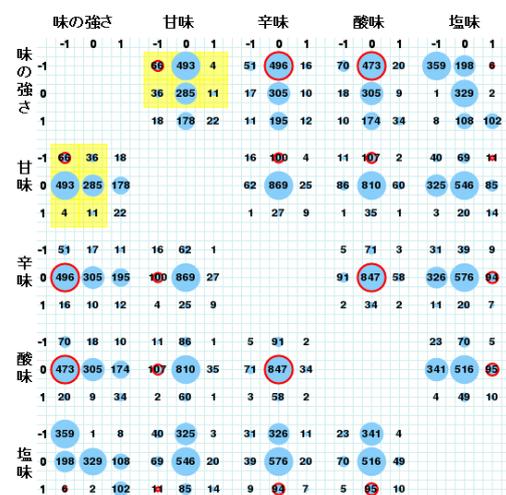


図3. ユーザ基準に照らした際のデータの合致度の視覚化画面(料理レシピの味の強さ, 味覚を各3段階で表現した際のレシピ集合の分布をバブルチャートで可視化。ユーザ基準(黄色部分)と比較対象(赤色部分)の合致度を俯瞰することができる

②信頼性などの特性が直感的に理解できる表現方法

アニメーション表現を利用した情報表示手法を実装した。アニメーションのパターン, 表示面積の変化, 速度などの表示属性に対する認識特性の知見を元に、例えば気づかせたい情報ほど速い動きで動かし、かつ表示面積が大きくなるように変化させて可読性を向上させるといった表現方式を開発した。図4は電子メールの通知アプリに実装した表示例である。表示は独自開発の文字アニメーション記述言語からSVGアニメーションに変換しているため、一般的なWebブラウザで表示可能である。

情報とのさらなるインタラクションとして、社会的フィルタリングに着目した信頼性



図 4: アニメーションを利用した電子メール通知の表示例. 図中, ★印表示は個人の知識による情報補完を用いた公共空間における情報隠匿のためのもの.

向上手段について研究開発した. 具体的には, 簡単な振る操作でユーザの感情を表現・共有可能な機能と, Twitter のエージェントアカウントを介したグループ内情報共有支援システムを実装した. 前者は, 上記のような情報のアニメーション表現において, 動きをユーザ自身が簡単な操作で付与することを可能とするものである. 後者は, 多様な情報共有の操作や閲覧をコミュニケーションと統合することによって, 共有する情報に対する意見交換を活性化するものである. 構成図を図 5 に示す. これらのシステムを実運用し評価した結果, 感情の容易かつ正確な伝達, および, 人による保証を伴った情報流通が可能であることを確認した.

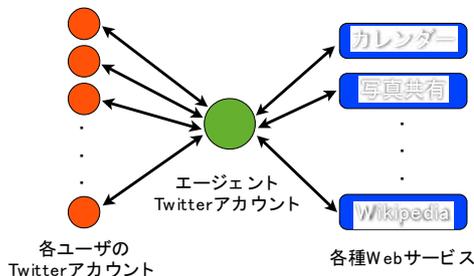


図 5: Twitter を UI とする情報共有とコミュニケーションの統合システム

本研究の成果は, Web 情報のみを対象とする国内外の関連研究の成果に対し, Web 情報だけでなく, 放送番組の特性に着目し, これを情報の信頼性判断支援に活用する基盤技術の確立に貢献している点に特徴がある. 個々の成果技術については, Web や放送番組を用いた検証が行われ, 特定の場面や分野のデータについて一定の有効性が示されたという段階であり, 汎用性に関する検証と改良を進めていく点に課題がある. 今後は, 得られた基盤技術の汎用性の検証と改良, および, 放送番組についてもより大規模化された映像アーカイブを利用した応用技術への発展が期待される.

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) Hiroshi Ishii, Qiang Ma, Masatoshi Yoshikawa, Incremental Construction of Causal Network from News Articles, Journal of Information Processing, Vol.20, pp.207-215, 2012, 査読有, 10.2197/ipsjjip.vol.20.207
- (2) Tatsuya Ogawa, Qiang Ma, Masatoshi Yoshikawa, News Bias Analysis Based on Stakeholder Mining, IEICE TRANS. INF. & SYST., Vol. E94-D, No. 3, pp 578-586, 2011, 査読有, 10.1587/transinf.E94.D.578
- (3) Shin Ishida, Qiang Ma, Masatoshi Yoshikawa, Extraction of Characteristic Description for Analyzing News Agencies, Journal of Digital Information Management, Vol.8, Issue 6, pp 349-355, 2010, 査読有, <http://db/journals/jdim/jdim8.html#IshidaMY10>

[学会発表] (計 39 件)

- (1) 木下翔伍, あら振ったー:振って文字にアニメーションを付加するメッセージングツール, 情報処理学会エンタテインメントコンピューティング研究会 (EC), 2012 年 3 月 27 日, ポートアイランド (兵庫県)
- (2) 長谷川馨亮, 時空間連続性を考慮した共起辞書構築による旅行ツイートの組織化, 第 22 回 Web インテリジェンスとインタラクション研究会 (SIG-WI2), 2012 年 3 月 17 日, 中央大学 (東京都)
- (3) 長谷川馨亮, 行動の時空間連続性を考慮した旅行ツイートの組織化, 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 (DEIM2012), 2012 年 3 月 4 日, シーサイドホテル舞子ビラ神戸 (兵庫県)
- (4) Ling Xu, A Complementary Tree for Recommendation of Multimedia News Items, 第 4 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 2012), 2012 年 3 月 4 日, シーサイドホテル舞子ビラ神戸 (兵庫県)
- (5) 佐伯隆太, テレビ番組に基づく Web コンテンツの信頼性判断支援システムの提案, 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 (DEIM2012), 2012 年 3 月 3 日, シーサイドホテル舞子ビラ神戸 (兵庫県)
- (6) 渡部将太, Twitter User Recommender～

- Twitterのお気に入り機能を用いた有用なフォロー先推薦システム～、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(DEIM2012), 2012年3月3日, シーサイドホテル舞子ビラ神戸(兵庫県)
- (7) 赤澤康幸, 冷蔵庫食材を考慮した料理レシピ検索システムの提案, 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(DEIM2012), 2012年3月3日, シーサイドホテル舞子ビラ神戸(兵庫県)
- (8) 栗納裕貴, 失敗知識データベースを用いた失敗事象の原因分析, 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(DEIM2012), 2012年3月3日, シーサイドホテル舞子ビラ神戸(兵庫県)
- (9) Liang Yan, Where can I Buy iPhone4S Now? :Spatio-Temporal Entity Retrieval on Twitter, 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(DEIM2012), 2012年3月3日, シーサイドホテル舞子ビラ神戸(兵庫県)
- (10) 樋渡達也, Twitterをインタフェースとした情報共有サービスの統合によるグループコミュニケーションの活性化, 情報処理学会ヒューマンコンピュータインタラクション研究会(HCI), 2012年1月19日, 国立民族学博物館(大阪府)
- (11) Ling Xu, Credibility-Oriented Ranking of Multimedia News Based on a Material-Opinion Model, The 12th International Conference on Web-Age Information Management (WAIM 2011), 2011年9月15日, 武漢(中国)
- (12) 赤澤康幸, 冷蔵庫食材を考慮した料理レシピ検索システムの提案, 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(DEIM2011), 2011年2月27日, ラフォーレ修善寺(静岡県)
- (13) 野島裕輔, ニュース記事の内容と構造特徴を考慮した因果関係マイニング, 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(DEIM2011), 2011年2月27日, ラフォーレ修善寺(静岡県)
- (14) Ling Xu, Discovering Inconsistency in Multimedia News Based on a Material-Opinion Model, the 44th Hawaii International Conference on Systems Science, 2011年1月7日, ハワイ(米国)
- (15) 石井裕志, An Incremental Method for Causal Network Construction, The 11th International Conference on Web Age Information Management, 2010年7月17日, 四川省(中国)

- (16) 宮森 恒, 情報の重み付き属性値の比較に基づく情報の妥当性判断支援手法, 情報処理学会創立50周年記念(第72回)全国大会, 2010年3月9日, 東京大学本郷キャンパス(東京都)
- (17) 宮森 恒, Assisting the Validity Assessment of Items based on Composition Similarity, ACM Multimedia 2009 Workshop on Multimedia for Cooking and Eating Activities (CEA2009), 2009年10月23日, Beijing Hotel (Beijing, 中国)
- (18) 水口 充, Richbiff: e-mail message notification with richer clues, HCI International 2009, 2009年7月22日, Town and Country Resort & Convention Center (San Diego, 米国)
- (19) 馬 強, Topic and Viewpoint Extraction for Diversity and Bias Analysis of News Contents, APWeb/WAIM2009, 2009年4月2日, Suzhou International Conference Center (Suzhou, 中国)
- (20) 馬 強, Ranking People Based on Metadata Analysis of Search Results, Workshop on Web Information Systems for Electronic Businesses and Governments (E-BAG 2008), 2008年9月2日, Auckland, New Zealand

[産業財産権]

○出願状況(計1件)

名称: 情報比較プログラム, 情報比較装置

発明者: 宮森 恒

権利者: 学校法人京都産業大学

種類: 特許権

番号: 特願 2009-067471

出願年月日: 2009年3月19日

国内外の別: 国内

6. 研究組織

(1) 研究代表者

宮森 恒 (HISASHI MIYAMORI)

京都産業大学・コンピュータ理工学部・准教授

研究者番号: 90287988

(2) 研究分担者

水口 充 (MITSURU MINAKUCHI)

京都産業大学・コンピュータ理工学部・教授

研究者番号: 60415859

馬 強 (QIANG MA)

京都大学・情報学研究科・准教授

研究者番号: 30415856