

平成 22 年度研究開発成果概要書
「電気通信サービスにおける情報信憑性検証技術に関する研究開発」
課題ア Web コンテンツ分析技術

(1) 研究開発の目的

本研究の最終目標は、本研究課題のアに対応して、Web 検索エンジン等によって得られる画像・音声・映像やテキストといった Web コンテンツの信憑性・信頼性を情報利用者が判断するために資する情報を、画像・音声・映像などの周辺コンテンツやテキストの表層的特徴分析をもとに「現実的な処理時間」で収集・分析・提示できる機能の研究開発、および、開発技術の ISP などの実環境における実証実験を行うことである。

それぞれの課題において各種の評価実験を設定し、課題アー 1、課題アー 2 については、それぞれ、10 以上のトピックについて人間による評価と比較し、精度 80%以上を達成する。なお、処理時間は、テキスト情報の分析技術のリアルタイム型の信憑性判断情報提供は「数秒～10 秒程度以内」を目標とする。

さらに、国内外の学術会議において積極的に発表等を行ったり、特許を取得したりすることにより、本課題の有効性を世の中に広く発信する。

(2) 研究開発期間

平成 19 年度から平成 22 年度（4 年間）

(3) 委託先企業

京都大学、兵庫県立大学、京都産業大学、株式会社きざしカンパニー

(4) 研究開発予算（百万円）

平成 19 年度 119.6

平成 20 年度 125.0

平成 21 年度 105.6

平成 22 年度 99.3

(5) 研究開発課題と担当

課題アー 1 画像・音声・映像情報の分析技術

課題アー 1－1 （京都大学）

画像・音声・映像情報の信憑性検証のためのクロスメディア情報分析・検索技術の開発

課題アー 1－2 （兵庫県立大学）

発信者・ユーザ意図検出と画像等の周辺コンテンツのクロスメディア情報分析による信憑性検証

課題アー 2 テキスト情報の分析技術

課題アー 2－1 （京都大学）

サーチエンジンの索引データを用いたテキストの表層的特徴分析技術

課題アー 2－2 （京都産業大学）

発信履歴情報からの発信者熟知度・センチメント分析に基づく発信者
分析エンジンの開発
課題アー２－３ （株式会社きざしカンパニー）
マニア指向ランキング機能を有するブログサーチエンジン

（６）これまで得られた研究開発成果

		（全体） 186 件	（当該年度） 54 件
特許出願	国内出願	3	1
	外国出願	0	0
外部発表	研究論文	74	31
	報道発表	1	0
	その他研究発表	87	21
	展示会	21	1
	標準化提案	0	0

具体的な成果

（１）Web 上のマルチメディア情報の信憑性

ウェブ上で公開されている、撮影者の意図が強く反映された画像の信憑性を判断するため、関連画像とその周辺テキストを収集し、画像・テキスト対の整合性に基づき信憑性を算出することで、利用者にとって判断できるようなサーチエンジンおよびインタフェースを開発しました。

また、Wikipedia の記事及び画像の信憑性を判断するため、記事に含まれている画像が、その記事およびキャプションにふさわしいかどうかを、記事とキャプションで表される事物の関係およびキャプションと画像の関係に着目して検証するエンジンおよびインタフェースを開発しました。

（京都大学）

ウェブ上の動画・音声コンテンツの内容の信憑性を利用者が判断できるようにするために、動画や音声の類似性と、視聴者の評価（ソーシャルアノテーション）の類似性を考慮することで、動画・音声コンテンツが信頼できる内容かどうか、見るに値する内容かどうかを判断するのに有用な情報を提示するシステムを開発しました。

また、映像ニュースが伝えている内容が、他局ではどのように伝えているか、差異はどの程度あるのかなど、視聴している映像の全体における偏りや説明と主張の不整合を、人物に関する表現と記述極性に着目して分析して提示するエンジンを開発しました。また、開発したエンジンと WISDOM (NICT) との連携も行いました。（京都大学）

また、特定のニュースのトピックに対して主要な論点を抽出し、情報発信者が各論点を肯定的に伝えているか、否定的に伝えているかといった編集者の意図を自動分析し、比較しやすく提示したり、伝えている論点に偏りがないかを数値化したりする技術も開発しました。（兵庫県立大学）

Web 上の電子地図は広く普及していますが、電子地図は頻繁に更新され

ておらず、すでに存在しない地物が地図上に掲載されているということがあります。また、制作者の主観によって表示の制御がおこなわれるなど、電子地図の信憑性は重要な課題となっています。本研究では、電子地図データと関連 Web 情報アーカイブを用いて、このような地物の誤記載や未記載の量、表示基準の不統一性を電子地図の信憑性を量る尺度として計算する手法を開発しています。また、個別の Web サイトに記載されている編集された略地図画像に関して、その空間的配置の整合性を分析し、略地図の信憑性を量る尺度として計算する技術を開発しました。（兵庫県立大学）

(2) Web 上のテキスト情報の信憑性

Web 検索エンジンで検索された Web ページ、質問回答サイトの回答文、レストラン検索で検索されたレストランページなどのテキスト情報を対象とした、テキスト情報の信憑性検証技術について研究を行いました。テキスト情報自身の内容の分析と、テキスト情報の社会的な支持の分析という 2 つの側面からの分析によって、人が信憑性検証を行うことを手助けする技術を開発しました。テキスト情報自身の内容の分析としては、文章の体裁、検索クエリに関する代表的な話題をどの程度網羅しているか、専門的な内容をどの程度含んでいるか、ページの内容が検索結果の中でどの程度メジャーな内容かといった指標を開発しました。また、テキスト情報の社会的な支持の分析としては、ソーシャルブックマークの支持度、被リンク数、リンク元の地理的な一様性や近接度合の指標を開発しました。（京都大学、兵庫県立大学）

Web 情報の時間的分析を行うシステムとして、検索結果の鮮度を評価するシステムを開発しました。ユーザが Web 検索を行った際に、検索結果のコンテンツがどれだけ新鮮であることを示すスコアを提示することで、現時点で正しいコンテンツが提供されているかどうかを判断する支援を行います。このシステムではニュースのアーカイブを用いて、関連するイベント情報を含むニュース記事を取得し、クラスタリングを行い、鮮度を判定したいコンテンツと類似するイベントがいつ頃発生したかを計算します。（京都大学、兵庫県立大学）

ブロガーが投稿した記事の履歴と、世の中で語られた話題の内容を時系列的に分析することによる、鮮度の高い情報を発信するブロガーの発見手法を開発しました。本手法では、1 ヶ月毎に取りまとめられたブログ記事を分析することで、各話題領域に関する関連語辞書の 1 ヶ月毎の履歴を作成します。これにより、各話題に関する世の中での語られ方の時系列変化を把握することができます。この辞書の履歴とブロガーが投稿した記事の類似度を計算し統計的に評価することで、鮮度の高い情報を発信するブロガーかどうかを判定することができます。（京都大学、京都産業大学、きざしカンパニー）

Web コンテンツの発信者がどのような感情（センチメント）で情報を発信しているかは、発信情報の信憑性を判断するために重要な情報です。これまで我々は、信憑性向上のため、各 Web サイトのセンチメントに基づき Web サイト間の排反関係を分析し、分析した排反関係を用いることで、ト

ピックに対する各記事の信憑性評価値となる信頼度を抽出しました。また、センチメント分析技術の改良を行うことで、情報発信者分類機能の精度向上を実現しました。これにより、ユーザは任意のトピックに関して書かれている記事に対する信頼度を得られるため、よりの確かな信憑性判断ができるようになりました。最終年度では、これまで開発してきた情報発信者の熟知度・センチメント分析に基づく発信者分析エンジンを課題アその他機関で開発してきた信憑性検証システムと融合することで、記事に対する信憑性判断を発信者のセンチメント傾向から判定することができるようになりました。

また、きざしカンパニーと連携しつつ、体験熟知度判定技術の精度向上に取り組むと共に、センチメント分析技術の改良を行うことで、情報発信者分類機能の精度向上を実現しました。これにより、信憑性・信頼性の高いブログ情報を発見するための検索機能の技術開発という観点から、10以上のトピックについて精度80%以上を達成しました。さらに、京都大学にて行う画像信憑性検証を目的とした発信者分類に関し、熟知度に基づく発信者分類手法および結果を提供すると共に、画像信憑性検証において効果的な熟知度分類手法に関しても検討を行いました。また、ブロガー以外の情報発信者分析を想定した熟知度判定手法の改良についても検討を行いました。(京都産業大学)

京都産業大学による情報発信者の熟知度分析手法を用いて、体験熟知度判定技術の精度向上に取り組むと共に、センチメント分析技術の改良を行うことで、情報発信者分類機能の精度向上を実現しました。また、熟知度判定辞書の精度を高めるための共起度計測APIの改良および、検索エンジンのリアルタイム性の向上に取り組みました。これら技術を駆使することにより、既に一般公開を行っている実証実験システムの改良を行ないました。なお、実証実験システムに基づいた評価実験を行ったところ、信憑性・信頼性の高いブログ情報を発見するための検索機能の技術開発という観点から、10以上のトピックについて精度80%以上を達成しました。以上より、生活体験性を考慮した、精度の高い情報発信者分析が実現でき、信憑性の高い情報の取得に寄与するシステムの開発が可能となりました。(京都産業大学、きざしカンパニー)

(3) 情報信憑性検証に関する研究開発に関する国際的な情報発信

本プロジェクト期間中に、Web情報の信憑性に関する国際ワークショップを4回開催し、国際会議のミニトラックとしてWeb情報の信憑性に関するミニトラックを1回主催しました。これらの国際ワークショップや国際会議は今後も継続して開催される予定であり、我々がWeb情報の信憑性に関する国際的な情報発信の場を形成したといえます。

前年度までの活動としては、Web情報の信憑性に関する国際ワークショップ Workshop on Information Credibility on the Web (WICOW) を、Webや情報検索などの最難関国際会議と併設して3回開催しました。平成20年10月にアメリカ合衆国で開催された ACM CIKM 2008 国際会議において、WICOW 2008 (<http://www.dl.kuis.kyoto-u.ac.jp/wicow2/>) を開催、平成

21 年 4 月にスペインで開催された World Wide Web 2009 (WWW2009) 国際会議において、WICOW 2009 (<http://www.dl.kuis.kyoto-u.ac.jp/wicow3/>) を開催しました。

今年度は、平成 22 年 4 月にアメリカ合衆国で開催された World Wide Web 2010 (WWW 2010) 国 際 会 議 に お い て 、 WICOW 2010 (<http://www.dl.kuis.kyoto-u.ac.jp/wicow4/>) を開催しました。また、平成 23 年 3 月にインドで開催された World Wide Web 2011 (WWW2011) 国際会議において、WICOW と AIRWeb と連携し、ウェブ情報の品質に関するワークショップ Joint WICOW/AIRWeb Workshop on Web Quality (WebQuality 2011) を開催しました。AIRWeb (International Workshop on Adversarial Information Retrieval on the Web) は過去 5 回開催されている国際ワークショップで、Web におけるスパムなどの発見に関するワークショップです。さらに、平成 23 年 1 月には、アメリカ合衆国で開催された HICSS 44 国際会議において、情報信憑性に関するミニトラックを (<http://www.dl.kuis.kyoto-u.ac.jp/HICSS-2011-WIQA-MINITRACK/>) 主催しました。