

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、画像等マルチメディア情報の内容の特徴量分析、発信者分析等をもとに「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－1、課題ア－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 60%以上を達成する。

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) 研究予算（百万円）

平成 19 年度 119.6 百万円

平成 20 年度 125.0 百万円

平成 21 年度 117.5 百万円（予定）

平成 22 年度 110.4 百万円（予定）

(5) 研究開発課題と担当

課題ア-1 画像・音声・映像情報の分析技術

課題ア-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-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-2 テキスト情報の分析技術

課題ア-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-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-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) 主な研究成果

特許出願： 0 件

外部発表： 17 件

具体的な成果

(1) Web 画像情報の信憑性

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、他の情報源から関連する画像を自動的に検索し比較分析して提示することで、元の画像がどの程度実物と相違するかや、元の画像がどの程度標準的・典型的なものを掲載しているかが利用者にとって判断しやすいサーチエンジンおよびインタフェースを試作開発しました。また、ある人物や組織などに関する Wikipedia ページに含まれている画像が、その人物や組織とどの程度関係が深いのかといった、画像の整合性を検証するエンジンを試作開発しました (京都大学)。

(2) 動画・音声コンテンツの信憑性

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、動画や音声に対して他の多くの視聴者がつけた感想・評価コメント (アノテーション情報) を自動的に取得・分析し、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や、インターネットや TV のニュースなどの映像・音声コンテンツが伝えている内容が、他局や他のメディア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・音声コンテンツの全体における偏り（不整合度）を、字幕データや関連情報をもとに自動分析して見せてくれるシステムを開発しています（京都大学）。また、情報発信者が何を伝え、何を伝えていないかといった編集者の意図を自動分析する技術（兵庫県立大学）も開発しています。

(3) Web 上の電子地図情報の信憑性

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たり、新たに存在する地物が未だ地図には掲載されていなかったりすることがあり、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（兵庫県立大学）。

(4) Web テキスト情報の信憑性

Web 検索エンジンで検索された Web ページのテキスト情報自身の内容がどの程度信頼できるかを、文章の体裁、ページ内容が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、検索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、検索された Web ページが社会的・地理的にどの程度支持されているかといった指標で判断できる技術を開発しています（京都大学、兵庫県立大学）。

(5) Web コンテンツの発信者の信頼度

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。本研究では、読売新聞（2006～2001 年の 6 年分）と朝日新聞（2006～2001 年の 6 年分）の記事データを分析して感情表現語辞書を開発しました。評価実験を通じて、この感情表現語辞書の精度について 10%程度の平均誤差を確認しています。この辞書を用いることで、例えば、同一の事象・事件を報道する各新聞社のセンチメントの相違などを見ることができます。また、発信者が発信している情報の分野にどの程度熟知しているかを、過去の発信情報履歴を分析して推定する手法を開発し、ブログ情報の書き手の熟知度分析などに応用しています（京都産業大学）。

(6) 熟知度分析に基づくブログの書き手（ブロガー）の自動分類とブログランキング手法の開発と実証実験

京都産業大学で開発された情報発信者の熟知度分析手法を用いて、ある話題に関するブログを検索した場合に、その話題に関連する分野に熟知しているブロガーのブログを上位にランキングする手法を開発し、株式会社きざしカンパニーの Web サイトでの実証実験を開始しました。

(7) 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

平成 20 年 10 月に米国で開催される ACM CIKM 国際会議において情報信憑性に関する国際ワークショップ WICOW 2008 (<http://www.dl.kuis.kyoto-u.ac.jp/wicow2/>) の開催を計画しました。

(7) 研究開発イメージ図

**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**
(京都大学, 兵庫県立大学, 京都産業大学, (株)きざしカンパニー)

マルチメディア情報の信憑性検証

(周辺コンテンツや他情報源から収集したテキスト情報の分析・画像特徴量分析による信憑性検証)



Webテキスト情報の信憑性検証

(Webページのメジャー度・話題網羅度・社会的な支持分析、発信者分析に基づく信憑性検証)



ヤフー株式会社の
ポータルサイトや検索エン
ジン連動での実証実験

(株)きざしカンパニーの
ポータルサイトや検索エン
ジン連動での実証実験

**情報信憑性検証基盤技術の確立
国際的な情報発信**

i-Believe