

研究課題別評価

1.研究課題名 情報検索における対象知識獲得支援システムの構築

2.研究者氏名 高間 康史

3.研究の狙い：

本研究の狙いは、Web 上に膨大かつ多様に蓄積された情報源のより高度な活用を実現するための、知的インタフェースの開発である。現在の Web への一般的アクセス手段である情報検索を、(インタフェースを介した)情報空間とユーザとの対話(インタラクション)の観点からとらえ、話題豊富な人との対話「感覚を与える知的インタフェース実現のための要素技術として、話題分布情報などのメタ情報の抽出・可視化手法について研究し、プロトタイプシステムとして実装する。

本研究で目指す「対話感覚」とは、システムの擬人化ではなく「人間同士の対話では主題だけでなく付帯する多様なメタ情報も同時にやりとりすることにより豊かかつ広がりのあるものになっている」ことを指している。

従来の Web 情報検索システムは、データベース検索や文献検索システムの研究の延長として研究・開発が進められてきた経緯があるため、その主眼は「HTML 文書(ページ)=文書」をやりとりの最小単位として、インデキシング(索引付け:キーワードなどの特徴量を文書に付与し管理すること)の効率化や大規模データの処理方法などにあった。しかし Web 上で公開されている文書は教科書や辞書の様な(ある程度)客観的・常識的に整理・体系化されたものではなく情報発信者の解釈により構成された単位として公開されており、その様な性質も踏まえてユーザに情報を提供することができれば、本研究で目的とする「対話感覚」の実現につながると考える。

本研究課題における主な研究項目は次の通りである。

- ・ ユーザが多様な観点・粒度で情報空間・検索結果を眺めることを可能とする情報可視化システム及び知的インタフェースの開発
- ・ 情報空間・検索結果文書集合から多様な話題を抽出する手法の開発
- ・ 情報源(文書、サイトなど)の持つ、各種メタ情報の提示

4.研究成果

(1)免疫ネットワークモデルに基づく話題分布情報の可視化

検索結果やオンラインニュース記事集合など、Web から入手可能な文書集合を対象とした従来の Web 情報可視化システムでは、効率よいブラウジングを支援するために文書クラスタリングを適用する事が多い。一方、発想支援システムやテキストマイニングなどでよく利用されているキーワードマップは、関連するキーワード同士が近くに配置されるようにしたものであり、文書間の関係や文書集合中の話題の分布、流行といった、単独の文書からでは得られない情報を得るのに有効であることが知られている。

本研究では、Web から得られる話題分布情報をユーザが容易に把握するためには、キーワードマップと文書クラスタリングの両方を同時に考慮することが重要と考え、これらを両立するために、キーワードマップ上で話題を構成するキーワードの塊を把握する手がかりとなるキーワード(ランドマーク)を抽出する。本研究では免疫ネットワークモデルの活性伝播機構により、この様な条件

を満たすキーワードの抽出を試みた。

評価実験は定性的・定量的両観点から行い、定量的観点からはオーバーラップのない文書クラスターを高速(90文書で30秒以内)で生成可能であり、クラスター間のサイズのばらつきも少ないことを確認した。定性的観点からは、提案手法により抽出されたランドマークおよび対応する文書クラスターについて、代表的なクラスタリング手法である k-means との比較を、被験者へのアンケートにより行い、良好な結果を得た。

また、免疫記憶細胞の機構を取り入れることにより、コンテキストを考慮したランドマーク抽出を行い、時系列文書集合族からの話題ストリーム抽出にも応用可能であることを示した。

(2) キーワードマップに基づく情報可視化システムの開発

通常反応時の免疫ネットワークでは、侵入抗原を認識可能な抗体が高活性化し、抗原を抑制すると同時に関連抗体も抑制する様に、ランドマークは関連の強いキーワードを抑制する。従って、強接続キーワードを近傍に配置した場合、ランドマークとその周囲に存在する、抑制されたキーワードの塊が、文書集合中の話題の一つに対応する事になるため、これらの関係を強調してキーワードマップ上に表現することにより、可読性の向上が期待できる。また、ランドマーク・関連キーワード間の関係を強調・優先してキーワード配置を行うことにより、実行の度に安定した配置を得ることもつながる。

以上の考えに基づき、ばねモデルを改良したアルゴリズムによりキーワードマップを生成する情報可視化システムを開発した。また、摩擦力の導入により、上述の話題ストリームを可視化する用途へも拡張可能であることを示した。

(3) 既存ブラウザと連携可能なプロトタイプインタフェースの開発

上述の研究成果をもとに、知的な Web インタクションを実現するプロトタイプインタフェース TAIWA-01(Topic Analyzer for Intelligent Web Assistane)を、サーバサイドプログラミングにより開発した。TAIWA-01 は通常のブラウザからアクセスでき、ブラウジング、検索エンジンなどを利用して収集したページ集合を解析し、話題分布情報などのメタ情報をユーザに提示する。この時、キーワードマップに基づく情報可視化システムとも連動して使用可能である。

(4) 画像情報源の発見

趣味のサイトや企業などにおいて、類似画像のコレクションを公開しているサイトが多いとの分析に基づき、サイト単位で類似画像集合を発見し、ユーザに提示するシステムに関する研究を行った。具体的には、サイト単位で画像収集する Web ロボットの開発、画像内容に基づく汎用的画像検索モデルとして LSP(Local Similarity Pattern)の提案、サイト内に存在する画像に共通する特徴の把握を容易にするための共通特徴を強調したサイト集約画像について研究を行った。

結果として、LSP は従来の画像内容に基づく検索方法よりも高精度な検索が可能であること、および抽出された特徴モデルを用いて、サイトに共通する画像特徴を強調したサイト集約画像が生成可能であることを示した。

(5) 情報発信者の主観的評価を推定するための特徴量についての考察

主観的評価推定のためには、形容詞・副詞といった、従来情報検索では利用されることの少な

かった特徴が有効であるとの仮説に基づき、映画批評サイトを用いて予備実験を行った結果、良好な結果を得た。本研究については予備的な知見を得たに過ぎず、この成果に基づく新たな情報検索モデルを提案すべく 継続して研究中である。

5.自己評価：

本研究課題を始める上で重点を置いたのは、システムの見かけやモダリティ(視覚や聴覚などの感覚を用いて外界を知覚したり 情報伝達を行うやり方)ではなく、やりとりする情報を豊かにすることであった。特に、従来の情報検索技術による Web インタクションでは、名詞句を中心としたキーワードを入力し、関係するページを得る事を目的として設計されており、検索精度の向上が主目的であった。これに対し本研究では、名詞句では表せない情報や、個々のページからでは得られない情報をユーザに提供することを目的とした点で大きく異なる。また、HTML で記述されたテキスト情報の収集が主であり、画像などは大量に存在する割にはその収集・利用手段は十分に研究されていない事を踏まえ、画像を効率よく収集するために画像情報源のコンセプトを提案し、効率よい収集・閲覧を可能とする技術についても研究を行った。

3年間の研究期間を通じて、主観的評価推定技術については予備的な知見を得るにとどまるなど達成度は研究項目により異なるが、全般として Web からユーザに提供可能な情報の多様化には貢献できたと考える。

本研究課題では、Web からユーザへの向き(多様な情報の提供)に主眼を置いているが、ユーザが本当に必要な情報を必要なときに入手できるためには、ユーザから Web (システム)への向き、すなわちユーザ意図などのフィードバックの獲得も重要である。この点に関しては、相互作用と賢さ領域の他の研究者との交流を通じて、かなりの知見を得、理解を深めることができた。これは、今後研究をさらに発展させていく上で、形には残らないが研究者にとって非常におおきな財産であると考ええる。

6.研究総括の見解：

Web 上に膨大かつ多様に蓄積された情報源のより高度な活用を実現するための、知的インタフェースの開発を目指し、挑戦的な取り組みとたゆまぬ努力により、顕著な成果を挙げた。具体的には、免疫ネットワークモデルに基づく話題分布情報の可視化、キーワードマップに基づく情報可視化システムの開発、既存ブラウザと連携可能なプロトタイプインタフェースの開発、また、画像情報源の発見などである。また、外部に対する成果の発表にも積極的であった。

本研究の成果の発展に一層努力すれば、情報検索における対象知識獲得支援システムの構築に向けて、将来の大きな貢献が期待できる。

7.主な論文等：

発表論文

1. Z. Stejic, Y. Takama, K. Hirota, "Modified hierarchical genetic algorithm for relevance feedback in image retrieval," Intelligent Data Analysis, Vol. 8, No. 4, 2004(in Press).
2. Z. Stejic, Y. Takama, and K. Hirota, "Relevance Feedback-Based Image Retrieval Interface Incorporating Region and Feature Saliency Patterns as Visualizable Image Similarity Criteria," IEEE Transaction on Industrial Electronics, Vol. 50, No. 5, 2003 (Accepted).

3. Z. Stejic, Y. Takama, and K. Hirota, "Genetic algorithms for a family of image similarity models incorporated in the relevance feedback mechanism," *Applied Soft Computing*, Vol. 2, No. 4, pp. 306-327, 2003.
4. Z. Stejic, Y. Takama, and K. Hirota, "Mathematical aggregation operators in image retrieval: effect on retrieval performance and role in relevance feedback," *Signal Processing, Special Issue on Content-Based Visual Information Retrieval*, 2003. (accepted).
5. Z. Stejic, Y. Takama, and K. Hirota, "Weighted local similarity pattern as image similarity model incorporated in GA-based relevance feedback mechanism," *Intelligent Data Analysis*, Vol. 7, No. 5, 2003 (in press).
6. Y. Takama, K. Hirota, "Web Information Visualization Method Employing Immune Network Model for Finding Topic Stream from Document-Set Sequence," *J. of New Generation Computing*, Vol. 21, No. 1, pp. 49-59, 2003.
7. 高間 康史, 廣田 薫, 免疫ネットワーク・メタファに基づくWeb 情報可視化手法, *日本ファジィ学会誌*, Vol. 14, No. 5, pp.472-481, 2002.
8. Z. Stejic, Y. Takama, K. Hirota, "Genetic Algorithm-based Relevance Feedback for Image Retrieval Using Local Similarity Patterns," *J. of Information Processing and Management*, Vol. 39, No.1, pp. 1-23, 2002.
9. K. Hirota, N. Iwamatsu, Y. Takama, "A Proposal of an Internal-state Inference System Based on Multi-modal Sensory Fusion Method," *IEEE Transactions of Instrumentation and Measurement*, Vol. 51, No. 2, pp. 347-352, 2002.
10. Y. Takama, K. Hirota, "Topic-based Intelligent Support System for Information Retrieval," *Journal of Advanced Computational Intelligence*, Vol. 4, No. 6, pp. 457-463, 2000.

総説 解説

1. 高間 康史, Web 情報ストリーム, *情報処理学会誌*, Vol. 44, No. 7, pp. 720-725, 2003.
2. 高間 康史, SOFT におけるチャンス発見 - 投稿論文データの情報可視化から -, *知能情報ファジィ学会誌*, Vol. 15, No. 3, pp. 270-274, 2003.
3. 大澤 幸生, 高間 康史, WWW に潜む創造的意思決定のチャンス, *人工知能学会誌*, Vol. 16, No.4, pp. 530-534, 2001.
4. 松下 光範, 高間 康史, ネット上の情報を可視化する技術, *日本知能情報ファジィ学会誌*, Vol. 15, No. 5, pp. 506-514, 2003.

著書

1. 高間 康史, 「1.3. 意思決定における人と環境の相互作用」, 大澤 幸生 (監修) 知識マネジメント, オーム社, pp. 9-20, 2003.
2. Y. Takama, "Part 3: Interaction with Environment for Awareness of Chances," in Y. Osawa and P. McBurney, Eds., *Chance Discovery? Foundations and Applications* -, pp. 170-183, Springer, 2003.
3. Z. Stejic, E. M. Iyoda, Y. Takama, K. Hirota, "Improving Image Database Summarization by Automatic Image Feature Selection Using GA," in L. Rutkowski, J. Kacprzyk, Eds., *Neural*

Networks and Soft Computing, Advances in Soft Computing, Physica-Verlag, pp. 674-679, 2003.

4. 高間 康史, 「Web エージェント」, 土屋 俊, 中島 秀之, 中川 裕志, 橋田 浩一, 松原 仁, 大澤 幸生, 高間 康史 (編集), AI 事典, 共立出版, pp.116-117, 2003.
5. 高間 康史, 「情報可視化」, 土屋 俊, 中島 秀之, 中川 裕志, 橋田 浩一, 松原 仁, 大澤 幸生, 高間 康史 (編集), AI 事典, 共立出版, pp.76-77, 2003.
6. Y. Takama, K. Hirota, "Immune Network-based Clustering for WWW Information Gathering/Visualization," in H. Motoda Ed., ACTIVE MINING, New Directions of Data Mining, IOS Press, pp. 21-29, 2002.
7. Y. Takama, K. Hirota, "Visualization of Topic Distribution from Document Sequence on Web," in P. Sincak et al. Ed., Intelligent Technologies ? Theory and Applications ? New Trends in Intelligent Technologies, IOS Press, pp. 189-195, 2002.

口頭発表

・国際会議 : 25 件 (招待講演 1 件)

・国内会議 : 17 件

特許出願

アイコンタクトによるコミュニケーション方式 (2 件)

防災ネットワーク方式

その他

SCIS&ISIS2002 Excellent Presentation Award, 2002.10.

日本ファジィ学会奨励賞 ,2002.8.

日本ファジィ学会論文賞 ,2002.8.

2nd Int. Symposium on Advanced Intelligent System Conference (ISIS2001) 優秀論文賞 ,2001.8.