

論文マッピングによる研究知識の可視化手法の提案*

高沢 健太[†] 矢吹 太郎 佐久田 博司青山学院大学 理工学部 情報テクノロジー学科[‡]

1 序論

1.1 背景

科学技術の発展とともに知識の量は増大し、知識領域は高度専門化され細分化されていったが、その反面知識領域間の相互理解は困難になった。現代医療の発展が医学だけでなく理学、工学、法学など様々な領域の知識によってもたらされた様に、新たな知識の発見や発展のためには知識領域間の連携・融合が必要であり、その為に知識の全体構造を俯瞰するための仕組みが求められている。[1]

本学理工学研究科では研究活動を経て蓄積された論文知識を効率的に運用するための学術論文アーカイブ・AAA (Aoyama Academic Archive) を構築中である。蓄積された論文知識をメタデータをもとに可視化することにより、特定の領域の研究活動の変遷を辿ることや、研究者・研究室間でのコラボレーションの提案など、研究活動におけるブレイクスルーを生み出す一助となることが期待されている。

1.2 関連研究

知識の構造化・可視化の研究は過去にも幾つか例がある。研谷らはトピックマップ [2] を用いた人名典拠情報の構築を行い、人名を中心とした人物情報をグラフによって構造的に可視化することを行った [3]。しかしこの研究では情報間の関連の有無しか記述されておらず距離の概念が存在しないため、知識間の距離が定義できず距離を考慮しない相対的なマッピングしかできない。

伊神ら [4] によって開発されたサイエンスマップでは、様々な分野における被引用回数が上位 1 % にある論文の引用関係をもとにクラスタリングを行い、クラスタ間の類似度を用いた力学モデルで研究領域のマッピングを行っている。この研究では異なる二つの期間でマッピングを行い、その結果を比較することで科学研究の変化を視覚的に表現した。しかしこの研究では相当量の論文データがあり各論文が十分な被引用回数を持つことが前提となっ

ており、被引用回数の少ない論文には適していない。

卒業論文のような参考文献数、被引用回数が少ない論文の多いデータベースでも実現可能な手法を考える必要がある。

1.3 本研究の手法

以上を踏まえ、本研究では本学科の過去の卒業論文・修士論文のメタデータを用いて研究知識をマッピングによって可視化する手法を提案する。可視化の手法としては知識の表現と交換に関する ISO 標準であるトピックマップを利用しつつ、知識間の距離の定義を加える手法をとる。

2 システム概要

データベースにアクセスし、格納されたメタデータをトピックマップにより構造的に記述し、可視化するシステムを構築する。

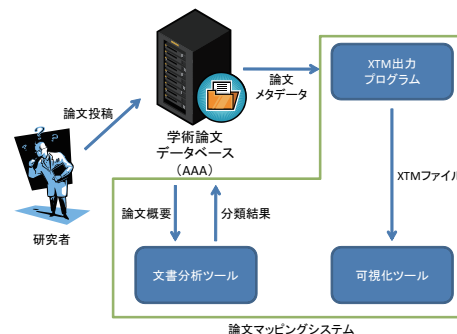


図1 システム概要図

まず、可視化の対象となるデータを格納したデータベースを用意する。本研究では学術論文のメタデータを格納したデータベースを用いた。メタデータには各論文の題名、著者、出典、発表年度、参考文献、分類、論文概要と、著者の所属組織が含まれる。しかし「分類」は明記してある論文が少ない上、学会などを越えた広統一的な分類は存在しないため、論文概要をクラスタ分析した結果をもとに論文を分類した。

クラスタ解析の過程でクラスタ間の類似度も算出し、「分類」のマップ上での距離をこの類似度をもとに定義し、マップ上に配置する。「分類」以外のメタデータは「分類」を中心に2次キーとしてマッピングされる。

クラスタ分析には計量テキスト分析、テキストマイニン

* Research knowledge visualization by means of article mapping.

[†] Kenta TAKAZAWA (a5805041@aoyama.jp)

[‡] Department of Integrated Information and Technology, College of Science and Engineering, Aoyama Gakuin University

グのためのフリーウェアである KH Coder[5] を用いた。クラスタ内の概要文中で出現確率比の高い語からクラスタの名前を決定し、「分類」としてメタデータに加える。

データベース上の論文メタデータは変換プログラムにより XTM(XML Topic Map) ファイルとして出力される。XTM は XML 形式によりトピックマップを記述する構文である。

可視化にはオープンソースのトピックマップツールである Ontopia[6] を使う。Ontopia は XTM によって構造的に記述されたデータを閲覧するインターフェイスを提供し、トピックの関連をグラフを用いて可視化する機能を持つ。しかしこの可視化機能はデータ同士の関連の有無しか表現できず距離の定義が存在しないため、データ間の距離を考慮してマッピングできるように改良した。

3 実験

3.1 目的

実装したシステムの動作テストを行い、生成された論文マップについての考察を行う。

3.2 実験手法

データベース上に青山学院情報テクノロジー学科の 2006 年度～2008 年度の卒業論文・修士論文とその参考文献のデータを用意し、システムに処理させ論文データを可視化する。論文の数は約 900、うち日本語の概要が公開されている論文が約 400 あり、クラスタ解析を行った。概要が公開されていない論文はクラスタ解析には用いず、メタデータの関連のみでマッピングを行った。

3.3 結果

システムは正常に動作し、図 2 のような論文マップが生成された。このマップは「佐久田博司」という人物トピックを中心に表示した状態で、薄いピンク色のノードが論文、茶色のノードが人物（おもに著者）、濃いピンクのノードが分類を示し、それぞれが著作、参考文献などの関係で繋がれている。



図 2 トピックマップ可視化結果

4 考察

まずはクラスタ解析によりマッピングされた分類ノードについて見てみると教育を中心に文書推敲や学習支援

などの分類が配置され、学習支援の両脇に設計とプログラミングが配置された。これは学習支援の対象としてプログラミングや設計を扱った論文が何本かあったことに起因する。その為か、プログラミング、教育など一般的な分野は集中する傾向にある。逆に放電加工などの専門的な分野は固まりの外へ分散する傾向にあり、専門度に応じて知識は放射状に広がっていくと考えられ、高度専門化するにつれ知識領域間の相互理解が難しくなることと一致することから、クラスタ解析による分類結果は妥当なものだと考えられる。

また、このような可視化を行うことにより、既存の論文データベースでは見えなかった共通の参考文献をもつ論文を探すなど、ある論文の基礎となった知識やその派生知識へのアクセスが容易になり、研究者間の交流や複雑系の理解、新知識の生成などに役立てられるだろう。

5 結論

本研究ではトピックマップと文書クラスタ解析を用いた論文マッピングによる研究知識の可視化手法の提案を行った。

今後の課題として、本論文では AAA から抽出する論文パーサを用意する形をとったので、AAA のデータベース構造に合わせてトピックマップ変換プログラムを書き換える必要があること、可視化手法において必要な処理ごとに別々のツールを使っていたので、AAA と連携し、一連の処理を一括でできるようなサーバーアプリケーションとしての実装や、表示情報が多くなった際にマップを適宜見やすい形に調整するようなインターフェイスが必要になる。また、使用した論文のメタデータ記述書式が統一されておらず、クラスタ解析の際少なからず影響が出るため書式を統一したメタデータの管理が望まれる。

本研究では情報テクノロジー学科の卒業論文を対象に実験を行ったが、それ以外にも学科のカリキュラムや研究成果をもとに同様のマッピングを行い大学への受験生の進路選びなどに役立てたり、医療機関で扱った難病などの症例をもとに専門医のネットワークを構築するなどの応用が期待される。

参考文献

- [1] 小宮山宏：知識の構造化，オープンナレッジ（2004）。
- [2] 内藤求，加藤弘之，桐山孝司，小町祐史，瀬戸川教彦，中林啓司，吉田光男：トピックマップ入門，東京電機大学出版局（2006）。
- [3] 研谷紀生，内藤求：トピックマップを用いた人名典拠情報の構築，情報知識学会誌，Vol. 19, No. 2, pp. 57-62（2009）。
- [4] 伊神正貴，阪彩香：サイエスマップによる科学研究の動的变化の観測，情報管理，Vol. 52, No. 5, pp. 255-266（2009）。
- [5] KH Coder，<http://khc.sourceforge.net/>。
- [6] Ontopia，<http://code.google.com/p/ontopia/>。