

研究種目：若手研究 (B)

研究期間：2007～2008

課題番号：19700132

研究課題名 (和文) 文献情報に基づく研究分野間ネットワーク分析

研究課題名 (英文) Research Field Network Analysis based on Repository

研究代表者

片上 大輔 (KATAGAMI DAISUKE)

東京工業大学・大学院総合理工学研究科・助教

研究者番号：90345372

研究成果の概要：

本研究では、研究領域の垣根を越えた統一的な設計原理構築を目指し、特に、様々な分野にまたがり関連する研究を、分類軸を参照しつつ、分野間の関係を明示的かつ多角的に捉え、整理し考察することによって、これまで曖昧であった一連の研究分野間の全体像を明確にすることを目指した。今年度は、(カンファレンスや学会誌単位などの) テーマに沿って集められた文献情報に基づき、文献と研究内容に関する関係を研究分野間の関係性としてネットワーク的に捉えることで、各研究分野を中心とし様々な視点から視覚化・分析・比較を行うことに重点をおいて研究を行った。ツール開発と実験を通して、様々な分野にまたがり関連する研究分野において、研究分野間の関係を明示的かつ階層的に捉え整理することによって全体像を明確にし、初心者にも専門家と同等以上の研究分野把握を支援した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	2,300,000	0	2,300,000
2008年度	1,000,000	300,000	1,300,000
年度			
年度			
年度			
総計	3,300,000	300,000	3,600,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：知能情報処理、視覚化、マイニング、文献情報

1. 研究開始当初の背景

近年ユビキタス社会・ロボット・ネットワーク社会など、社会がますます高度化、複雑化、多様化するなかで、各分野の研究も細分化された知を有機的に統合するグローバルな、横断型科学技術がその重要性を高めている。昨年度は文理にまたがる43の学会が、自然科学とならぶ技術の基礎である「基幹科学」の発展と振興をめざして横幹連合が設立

され、知の横断型会議である「第1回横幹コンファレンス」が盛況のうちに開催された。しかし、このような試みは近年始まったばかりであり、特に各分野の知見をどのように活用するかが重要かつ必須の問題とされている。現状では、各分野における情報交換にとどまっているといっても過言ではない。そこには、各分野に共通するような横断的かつ包含的な事象に基づいた解析とその情報提供

が不可欠である。

一方、これまで各研究分野における文献探索を行なう場合、検索エンジンや文献の索引などを駆使して行なってきた。CiteSeer等、幾つかの特殊なサイトには詳細な文献情報があり、参照情報や文献間の類似情報により必要な文献を見つけることが可能である。しかし、各研究分野において実際にどのような研究がどのように行われているか、ある会議の今年と昨年の違いは？この10年でどのように研究テーマは変遷してきたのか？など、全体像の把握を行なうことは難しかった。

最近ではネットワーク分析に関する研究が盛んになってきており、松尾ら[松尾 05]や、友部ら[友部 05]はWebからの研究者関係を抽出しそのネットワークを構築、中心的な研究者を分析する研究を行っている。また、難波らは、文献間の参照関係から関連論文の組織化を行っている[難波 01]。しかし、これまで研究者や文献という観点でネットワークを構築することは多かったが、研究分野という観点でネットワークを構築する研究はなく、学会やジャーナルごとの分野の分布や偏り、時系列における研究分野の変遷や鳥瞰図的な視点などを知ることはできなかった。

2. 研究の目的

本研究では、研究領域の垣根を越えた統一的な設計原理構築を目指して、特に、様々な分野にまたがり関連する研究を、分類軸を参照しつつ、分野間の関係を明示的かつ多角的に捉え、整理し考察することによって、これまで曖昧であった一連の研究分野間の全体像を明確にすることを目的とする。そこで、本研究では、(カンファレンスや学会誌単位などの)テーマに沿って集められた文献情報に基づき、文献と研究内容に関する関係を研究分野間の関係性としてネットワーク的に捉えることで、各研究分野を中心とし様々な視点から視覚化・分析・比較を行う。

3. 研究の方法

前述の開発と分析を実現するには、図1のように(1)論文集合収集、(2)書誌情報抽出、(3)研究分野の知識構造化、(4)研究分野ネットワーク生成、(5)ネットワーク可視化、(6)ユーザによる分析とパラメータ調整、の6つのステップを必要とする。これらのステップを通して、たとえば、「文献セットはどのような研究分野の研究が集まっているのか?」「それらの関係はどうなっているのか?」「その関係はどのように変遷しているのか?」など、これまで得ることが難しかった関係を明らかにしユーザに提供する。ネットワークの定義に関しては、研究分野をノードとしたネットワークを構築する。一般的に1つの文献は複数の研究分野(キーワード)に

関わる。この関係性を利用し、カンファレンスや学会誌単位などの関連する複数の文献セットからネットワークを構築する。

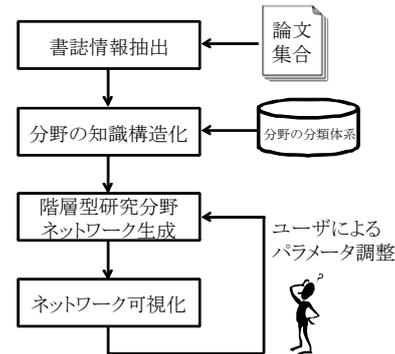


図1 専門性把握支援の手続きの流れ

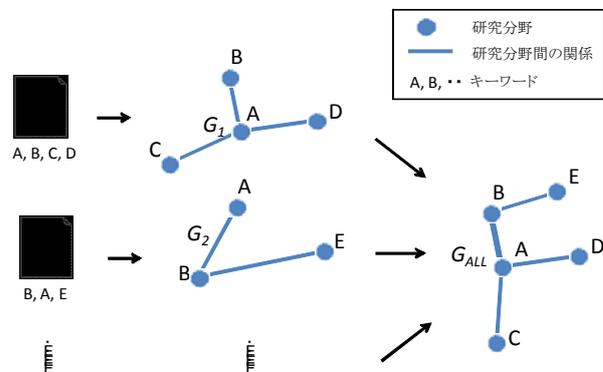


図2 論文集合ネットワークの生成

知識構造化では、まずそれぞれの論文に付加されたコードの情報を用いてネットワークを作成する。それぞれの論文に付加された複数のコードを元に論文ごとに論文ネットワークを作り、それを統合したものが学術論文誌全体のネットワークになる。コードにおいては、論文に関連の強い研究分野から順番に記入する慣習があることが多い。その場合は、付加されているコードにおいて先頭にあるコードを主分野と仮定し、その他のコードは関連分野とすることで半順序関係を作成する。これによって主分野から関連分野へと向けてエッジを作成することが可能になる。構築したネットワークは、頂点集合(vertex set) $V(G)=\{u,v,w,z\}$ と辺集合(edge set) $E(G)=\{uv,uw,uz\}$ からなる一般グラフ $G=\{V,E\}$ となる。ここで、多重辺は分野間のつながりの強さを表すが、同じコードを1つの論文に2つ以上つけることは一般的にないためループは存在しない。

たとえば、学術論文誌や学術会議などを単位に論文集合を作成する場合、論文集合に含まれる各論文からまず各論文が関わるネットワークを作成する。図2のように、論文d₁にA、B、C、Dの分野コードが付けられている場合はメイン分野となるAとサブ分野B、

C、D をエッジで結合し、頂点集合 $V_1(G_1) = \{A, B, C, D\}$ と辺集合 $E_1(G_1) = \{AB, AC, AD\}$ からなる論文ネットワーク $G_1 = \{V_1, E_1\}$ を作成する。同様に、論文 d_2 について、 $V_2(G_2) = \{A, B, E\}$ 、 $E_2(G_2) = \{BA, BE\}$ からなる $G_2 = \{V_2, E_2\}$ を作成する。

構築した論文ネットワークを統合することにより論文集合の特性を表すネットワークが作成される。 G_1 、 G_2 、 \dots を統合すると、 $V_{ALL}(G_{ALL}) = \{A, B, C, D, E\}$ 、 $E_{ALL}(G_{ALL}) = \{AB, AC, AD, BE\}$ 、 $W_{ALL} = \{2, 1, 1, 1\}$ からなる重み付き一般グラフ $G_{ALL} = \{V_{ALL}, E_{ALL}, W_{ALL}\}$ が作成される。このようにして作成された G_{ALL} を本研究では研究分野ネットワークと呼ぶ。

これらの研究分野ネットワークを研究分野の階層構造に対応させるために、階層型の研究分野ネットワークを開発した。本研究では、分野階層として図 3 のような、3 階層構造からなる経済学分野の JEL 分類コードを分野階層として考えた。たとえば、分野 C (数学的定量的手法)、分野 D (ミクロ経済学) における研究分野体系は、JEL 分類コードを用いて図 3 の様に大分類、中分類、小分類の 3 つの階層構造によって表現することができる。

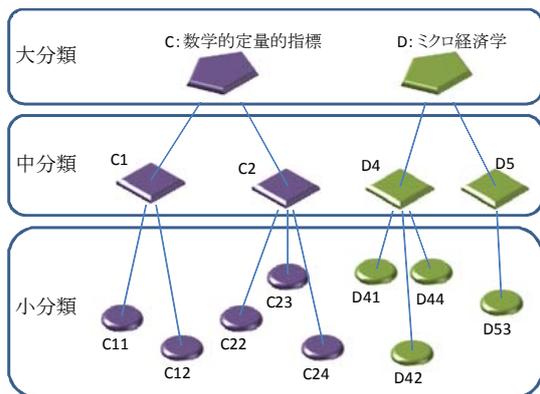


図 3 研究分野の階層構造

論文に付加された分野がすべて同じ階層レベルにおける関係性であった場合(図 4(a))、たとえば C12 から D41 に関係性があった場合(青点線)は、分野間関係性が同レベルの階層なので、ノード間をエッジ(赤太線)で結合し 4.2 節の手続きにて各論文ネットワークを作成することができる。作成されたネットワークの上位階層の研究分野間、C1-D4 間、C-D 間にも同様にエッジが張られ、各階層ごとに目的の関係性を可視化することができる。

一方、研究分野のキーワードの選択として、異なる階層のコードが一緒に選ばれる場合も存在する。この場合、異なる階層のコード間関係性が存在することになる。たとえば図 4(b) のような場合、C12 から D4 に関係性

があるが、D4 という分野は小分類レベルのネットワークとしては D4 に属する分野を特定できず表示することができない。しかし、中分類レベルでの表示を指定した場合、D4 は中分類に存在するが、C12 は中分類レベルでは C1 となり D4 と C1 の間で一本のエッジを作成することができる。大分類における表示も同様の手続きで C、D 間にエッジが作成される。

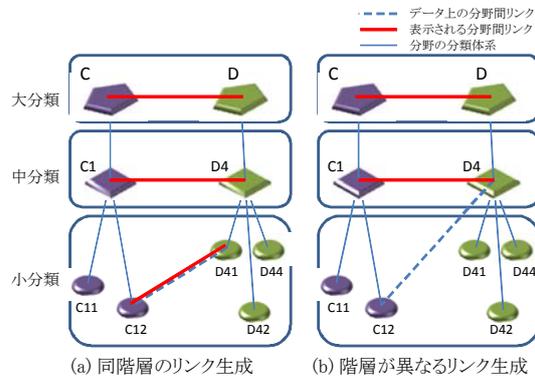


図 4 異なる階層レベルでのエッジ生成

以上のように、異なる階層レベル間の関係性を構築する際に、下位に複数存在する分野ノードを統合した一つの上位分野ノードとして扱いエッジ生成を行なうことで、階層関係をもつ研究分野間の関係性を各階層ごとに可視化・比較を行なうことができる。

本研究では、このようにして作成されたネットワークを階層型研究分野ネットワークと定義する。このネットワークは、各論文の内容に基づき論文著者が付加した情報を用いており、学術論文誌・学術会議等の分野分布を適切に表すことができる。開発した情報可視化ツールを用いて表示した例を図 5 に示す。ノードの大きさが異なるのは、キーワードの共起頻度に対応してノードの大きさを変更させているためである。

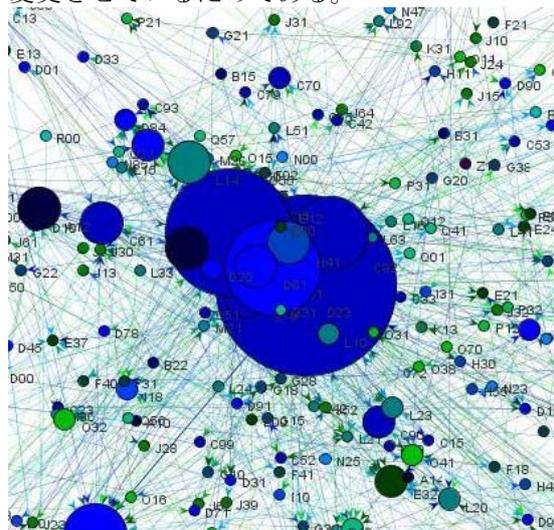


図 5 研究分野間の関係性を情報可視化ツールを用いて可視化した例(経済学の学術論文

誌において 10 年間に掲載された全論文における JEL 分類コード小分類の全関係性)

また、2 つのグラフを比較するとき有用な手法として差分グラフを実装した。これは 2 つのグラフを指定してどちらの特徴が強いかわかりやすく表示する機能である。またその差分グラフについての統計情報を出力する機能を持つ。差分グラフの作成方法は、図 6 の通りである。

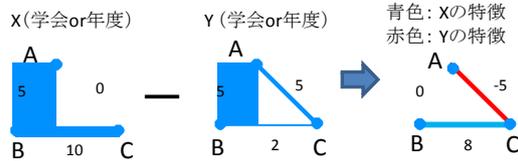


図 6 研究 2 つの比較対象となるネットワークによる差分研究分野ネットワークの作成

ある 2 つの論文集合ネットワーク間の中分類における差分グラフを図 7 に示す。どちらかに特徴の強い研究分野間の関係性が青もしくは赤のネットワークによって表現される。

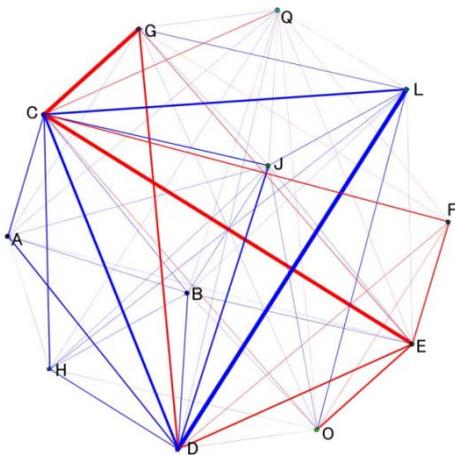


図 7 2 つの学術論文誌の階層レベル大分類で度数 50 以上の論文集合ネットワーク間の差分グラフ

開発したこれらのツールを用いて、経済学における 2 学術論文誌の公表論文の分類実験を行なった。この実験により、専門家と対象領域外の初心者ユーザが実際にどの程度対象の専門領域を理解しているかを調査し、そして前述の可視化手法を使用することで専門性の違いを初心者ユーザに対して明確に捉えることを支援できるかを調べた

4. 研究成果

本研究の成果として、前述の 2 学術論文誌の公表論文の分類実験の結果における研究

分野把握の精度比較により、専門家と初心者の把握能力に有意差はないこと。しかしながら、専門家と初心者では把握の内容が大きく異なることがわかった。

提案手法により研究分野間の階層的関係性をネットワーク化し可視化を行なうことで、初心者の研究分野把握を、専門家よりも高く、かつ安定して提供できることを確認したといえる。さらに専門家の分野把握においても専門分野の知識に基づき分類することに焦点が絞られていることで逆にトピック間の関係性や研究領域がカバーする範囲の差異などにはほとんど考慮がされないという危うさを示唆した。

本研究では、ツール開発と実験を通して、様々な分野にまたがり関連する研究分野において、研究分野間の関係を明示的かつ階層的に捉え整理することによって全体像を明確にし、初心者にも専門家と同等以上の研究分野把握を支援したといえる。

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)

- [学会発表] (計 4 件)
- ① 奥岡晋大, 片上大輔, 新田克己, 研究分野ネットワークによる学術分野動向の解析と評価, 人工知能学会研究会 SIG-KABS-A803, 2009. 01. 29, 武蔵野市, 査読無
 - ② D. Katagami, T. Yamada, K. Nitta, Comprehension of Invisible Specialty in Research Field through Interaction with Novice Researchers, International Workshop on Intelligent Web Interaction on IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 2008. 12. 09, Sydney, Australia, 査読有
 - ③ 奥岡晋大, 片上大輔, 新田克己, 投稿論文に基づく研究動向の変遷の解析と検索支援, 第 22 回人工知能学会全国大会 1G1-02, 2008. 06. 11, 旭川市, 査読無
 - ④ 片上大輔, 清水英明, 田中貴紘, 新田克己, 山田隆志: 文献情報に基づく学際的分野間ネットワーク分析, 第 21 回人工知能学会全国大会, 1B2-7, 2007. 06. 20, 宮崎, 査読無

〔その他〕

- ① 現在、日本知能情報ファジィ学会誌「知と情報」に投稿中。
- ② 現在、日本知能情報ファジィ学会誌「知と情報」提言論文にシステムを利用した結果を分析として用いた内容を投稿中。

6. 研究組織

(1) 研究代表者

片上 大輔 (KATAGAMI DAISUKE)
東京工業大学・大学院総合理工学研究科・助教
研究者番号：90345372

(2) 研究分担者

なし

(3) 連携研究者

なし