

萌芽期の研究に着目した学術動向予測モデルの評価

山下 長義^{†1} 沼尾 正行^{†2} 市瀬 龍太郎^{†3}

萌芽期の研究に着目した学術動向を予測するモデルを提案する。そのために、現在の専門分野と萌芽期の研究における専門分野を研究者ごとに調査する。萌芽期の研究において大きな成果が得られれば、現在の専門分野から萌芽期の研究における専門分野に研究基盤が移行する可能性が高いからである。このような研究者が同じ時期に同じ研究分野間で多数存在すれば、学術動向の先行指標となると考えられる。そこで本論文では、科学研究費補助金（以下科研費とする）の応募データにおいて、現在の専門分野に応募している基盤研究・若手研究での応募細目と、萌芽期の研究で応募している挑戦的萌芽研究での応募細目との重複応募状況を用いて予測モデルを作成する。実験の結果、情報学と社会科学にて予測モデルが有効であることを示した。これらの分野では、研究者の研究経歴によって細目の関係を表す研究領域ネットワークにおいて、クラスター係数が大きな値を示すことがわかった。

Evaluations of A Model for Predicting Science Trends Focusing on Breakthrough researches

NAGAYOSHI YAMASHITA,^{†1} MASAYUKI NUMAO^{†2}
and RYUTARO ICHISE^{†3}

This paper proposes a method for predicting science trends focusing on breakthrough research. Current research field and research field in breakthrough research are investigated in each researcher. A large research outcome of breakthrough research makes researchers move from current research field to new research field of breakthrough research. If a lot of these researchers are existing in the same time and the same pair of research fields, this pair of research fields becomes an advance indicator of science trends. Thus, in this paper we introduce a prediction model using pairs of research fields which are proposed by the same researchers in “Scientific Research” and “Challenging Exploratory Research” of Grants-in-Aid for Scientific Research. The experimental results showed that our proposed method is effective especially in Areas whose clustering coefficients are high in the research fields network consisting of research

1. はじめに

数学や哲学などの古典的な学問から環境学やゲノム科学などの幅広い分野をカバーする学問が存在している現在、学術の動向を把握することは重要になりつつある。科学における知の発見は、既存の研究の新たな組み合わせから生み出されることが多いからである。しかし、学問が細分化されそれぞれの研究分野が高度化されているため、自らの研究を遂行しながら幅広く学術の動向を把握することは難しいのが実状である。そこで、我々は研究者の研究履歴を用いた学術動向を把握するモデルを提案した¹⁾。科学研究費補助金（以下科研費とする）の応募データから、各研究者が採択されていた細目と次に応募する細目を抽出することで研究履歴を調査し、これらが学術の動向と一致していることを示した。さらに、学術動向の変化を予測できれば研究方針の決定に役立つことが期待される。

本論文では、萌芽期の研究に着目した学術動向を予測するモデルを提案する。そのために、研究者ごとに現在の専門分野と萌芽期の研究における専門分野を調査する。萌芽期の研究において大きな成果が得られれば、現在の専門分野から萌芽期の研究における専門分野に研究基盤が移行する可能性が高いからである。このような研究者が同じ時期に同じ研究分野間で多数存在すれば、学術動向の先行指標となると考えられる。そこで科研費において、現在の専門分野に応募している基盤研究・若手研究での応募細目と、萌芽期の研究で応募している挑戦的萌芽研究（萌芽研究を含む）での応募細目との重複応募状況を用いて予測モデルを作成する。

本論文の構成を以下に示す。2章では関連研究、3章では科研費制度の概要、4章では重複応募状況による学術動向の予測と評価について述べる。5章にて考察を述べ、6章ではまとめとこれからの課題を述べる。

^{†1} 日本学術振興会 研究事業部

Research Program Department, Japan Society for the Promotion of Science

^{†2} 大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

^{†3} 国立情報学研究所 情報学プリンシプル研究系

Principles of Informatics Research Division, National Institute of Informatics

2. 関連研究

これまで学術の動向を調査する研究のほとんどは、論文における情報を用いたものであった。代表的なものに論文の共著関係を用いた研究²⁾、論文の文献情報を用いた研究³⁾や文獻キーワードを利用した研究⁴⁾などがある。しかし、論文のデータは様式が一樣ではなく、さまざまな分野の論文を揃えることが難しいといった問題がある。

学術に関するデータは論文の他に科研費の応募データがある。科研費の応募データの様式は一樣で、人文学・社会科学から医学まで幅広い分野の研究者からの応募があり、多くの研究者が毎年継続して応募しているため年度ごとの変化を調査するのに適している。また、それぞれの応募課題に対して分野、分科、細目などの研究領域やキーワード、研究者の所属機関などの属性が付加されている。科研費などのグラントの応募データを用いた研究は、科研費における研究代表者と研究分担者の関係によって研究領域ネットワークを抽出しその経年変化の予測する研究⁵⁾や、NIH（アメリカ国立衛生研究所）の採択課題のタイトルやアブストラクトから抽出した関係を可視化しインタラクティブに検索できるシステムを提案している論文⁶⁾がある。著者らは、科研費の応募データから、各研究者が採択されていた細目と次に応募する細目を抽出することで研究履歴を調査し、これらが学術の動向と一致していることを示した⁷⁾。本論文では重複応募状況によって学術動向を予測するモデルを提案し評価を行う。

3. 科研費制度の概要

文部科学省および日本学術振興会によって公募、審査、交付される科研費は、日本で提供される研究助成金の中でもっとも一般的なものである。「人文・社会科学から自然科学までの全ての分野にわたり、基礎から応用まであらゆる「学術研究」（研究者の自由な発想に基づく研究）を格段に発展させることを目的とする「競争的資金」であり、ピア・レビュー（専門分野に近い複数の研究者による審査）により、豊かな社会発展の基盤となる独創的・先駆的な研究に対する助成を行うものです⁸⁾。

科研費の応募データは、応募課題の研究領域を表す細目やキーワード、応募研究者の所属や研究分担者、研究種目などの様々な属性が付加されている。応募課題の研究領域は、「応募に際しては、研究計画の内容に照らし、審査希望分野を示す分類表である「科学研究費補助金 系・分野・分科・細目表」から適切な細目を応募研究者によって選定されるとともに、最も関連が深いと思われるキーワードを「キーワード一覧」より選定⁸⁾される。キー

ワードは、細目よりも詳細な研究領域を表す。主な研究種目には基盤研究 (S)・(A)・(B)・(C)、挑戦的萌芽研究、若手研究 (A)・(B) があり、基盤研究は独創的・先駆的な研究を助成の対象に、挑戦的萌芽研究は独創的な発想に基づく挑戦的で高い目標設定を掲げた芽生え期の研究を助成対象に、若手研究は 39 歳以下の研究者が 1 人で行う研究を対象に助成を行っている。ただし、挑戦的萌芽研究は基盤研究 (S)・(A)・(B)、若手研究 (A) と重複して助成を受けることが可能である⁸⁾。

4. 重複応募状況による学術動向の予測と評価

基盤研究・若手研究と挑戦的萌芽研究との重複応募状況を用いて学術動向を予測するモデルを提案し、評価を行う。4.1 節では重複応募状況と細目間の移動研究者数の抽出法について、4.2 節では予測モデルを作成し予測実験を行う。4.3 節にて予測モデルの評価を行い、4.4 節にて予測結果を評価する。

4.1 重複応募状況と細目間の移動研究者数の抽出法

重複応募状況を抽出するために重複応募数と累積重複応募数を定義する。平成 n 年度における細目 x から細目 y への重複応募数（採択重複応募数）は、平成 n 年度に基盤・若手研究において細目 x に応募し（採択され）、挑戦的萌芽研究において細目 y に応募した研究者数とする。平成 n 年度において基盤・若手研究の細目 x に応募した（採択された）研究者集合を $S_x(n)$ ($S_x^a(n)$)、挑戦的萌芽研究の細目 y に応募した研究者集合を $C_y(n)$ とすると重複応募数（採択重複応募数）は以下の式により求まる。

$$|S_x(n) \cap C_y(n)|, \quad (|S_x^a(n) \cap C_y(n)|) \quad (1)$$

ただし $S_x(n)$, $C_y(n)$, $S_x^a(n)$ は、新規・継続課題を用いる。たとえば、平成 22 年度における知能情報学からメディア情報学・データベース A への重複応募数（採択重複応募数）は、平成 22 年度において基盤・若手研究の知能情報学に応募し（採択され）、挑戦的萌芽研究のメディア情報学・データベース A へ応募している研究者数となる。さらに累積重複数（採択累積重複数）は、平成 $n-2$ 年度から平成 n 年度までの同一細目対における重複応募数（採択重複応募数）の合計とする。

採択重複応募数は、基盤研究・若手研究における継続課題が終了するまで応募細目が変更されないため、重複応募数よりも応募細目の更新頻度は低い。しかし採択重複応募数では、継続課題終了後に新規で応募するため、前年度までの科研費による研究成果を基に次に応募する細目が決定されることが期待される。

次に、移動数（採択移動数）を定義する。平成 n 年度における細目 x から細目 y への移

動数（採択移動数）は、平成 n 年度に細目 x に応募した（継続終了した）研究者のうち、平成 $n+1$ 年度に細目 y に応募した研究者数とする。平成 n 年度に細目 x に応募した（継続終了した）研究者集合を $S_x(n)$ ($S_x^a(n)$)、平成 $n+1$ 年度に細目 y に応募した研究者集合を $S_y(n+1)$ として移動数（採択移動数）は、以下の式により求まる。

$$|S_x(n) \cap S_y(n+1)| \quad (|S_x^a(n) \cap S_y(n+1)|) \quad (2)$$

ただし同一課題の継続分を除外するために、平成 n 年度における研究者集合は新規と継続課題から抽出し、平成 $n+1$ 年度における研究者集合は新規課題から抽出する。たとえば、平成 22 年度におけるメディア情報学・データベース A から知能情報学への採択移動数は、平成 22 年度にメディア情報学・データベースにおいて継続課題が終了した研究者のうち、平成 23 年度に知能情報学で応募した研究者数となる。さらに、累積移動数（採択累積移動数）は、平成 $n-2$ 年度から平成 n 年度までの同一細目対における移動数（採択移動数）の合計とする。

また、平成 n 年度における細目 x から細目 y への採択移動率は、以下の式により定義する。

$$\frac{|S_x^a(n) \cap S_y(n+1)|}{|S_x^a(n)|} \times 100 \quad (3)$$

4.2 予測実験

平成 17 年度における応募データを基に予測モデルを作成し、この予測モデルを平成 19 年度の応募データに適用することで平成 20 年度から平成 21 年度において採択移動数が増加する細目対を予測する。属性は、細目対ごとの累積重複数、累積移動数、採択累積重複数、採択累積移動数を用いる。2 年間の採択移動数が 2 以上の細目対を正例、1 以下の細目対を負例とする。訓練データは、平成 18 年度から平成 19 年度における正例、負例のラベルを平成 17 年度の実績データに付して作成し、この訓練データに対してロジスティック回帰を適用することによって予測モデルを作成する。予測モデルは、属性ごとのロジスティック回帰係数によって得られる。そして、この予測モデルを平成 19 年度の実績データに適用することで平成 20 年度から平成 21 年度において正例となる細目対を予測する。

分野別に予測モデルを作成し予測を行う。ここで分野とは、人文学・社会科学・数物系科学・化学・工学・生物学・農学・医歯薬学に情報学・環境学を加えたものとする。分野別の分析対象となるデータは、各分野の細目とその分野の細目との間で採択移動率が 1.0 以上である他分野の細目を合わせたものとし、いずれか一方がその分野内の細目である細目対を対象とする。それぞれの分野において分析対象になった細目数は表 1 に示す。

分野名	情報	環境	人文学	社会	数物	化学	工学	生物	農学	医歯薬
分野内細目数	17	7	20	29	20	13	46	15	25	62
対象細目数	228	246	104	156	156	143	216	197	220	228
細目対数	7,446	3,388	3,740	8,178	5,820	3,536	17,710	5,670	10,350	24,366

表 1 分析の対象となる細目数

Table 1 The Number of Research Field in Each Area

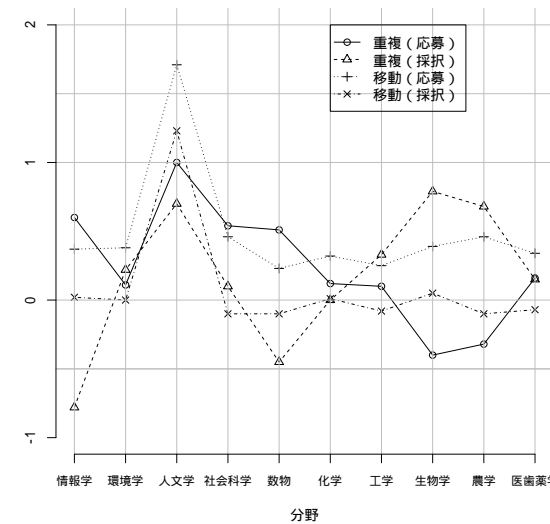


図 1 予測モデルにおける属性別の係数の大きさ

Fig. 1 Coefficients of Prediction Model in Each Area

4.3 予測モデルの評価

予測モデルの評価を分野別に行う。図 1 は、予測モデルにおける属性別の回帰係数の大きさを示し、重複とは累積重複数を、移動とは累積移動数のことを表す。横軸は分野を、縦軸は予測モデルにおける属性ごとの回帰係数の値を示している。

累積重複数は、情報学・人文学・社会科学・数物系科学で 0.5 を上回る大きい値を示し、採択累積重複数は、工学・生物学・農学において 4 つの係数の中で最も高い値を示してい

る。よって、以上の分野における予測モデルでは、累積重複数が大きな影響を与えていることがわかる。環境学・人文学・化学・医歯薬学では累積移動数の値が最も大きく、過去の移動研究者数の影響を大きく受けていることがわかる。また、採択累積移動数は、人文学を除いて0前後の値を示していることから、人文学以外では予測モデルに対して大きな影響を及ぼさないことがわかる。

4.4 予測結果の評価

適合率、再現率、F 値によって予測結果の評価を行う。そのために、予測結果と、平成 20 年度から平成 21 年度の応募データにおける実際の状況を用いる。適合率とは予測によって正例であると評価された細目対の中に含まれる正例の割合とし、再現率とは全正例のうち予測によって正例と評価された細目対の割合とし、F 値は、適合率と再現率の調和平均とする。図 2 は、分野別の予測結果を示し、横軸は分野、縦軸は適合率・再現率・F 値の値を示している。

F 値は、社会科学・化学・情報学の順に高く、これらの分野では提案手法が有効であることがわかる。ただし、化学の予測モデルでは、累積移動研究者数の影響が最も大きく（4.3 節、図 1）、過去の移動研究者の傾向から強い影響を受けていることがわかる。

5. 考 察

予測結果の評価にて、分野によって大きな違いが見られた（4.4 節）。その要因を調査するために、研究領域ネットワークにおけるクラスタ係数と、研究者の移動状況・重複応募状況を分野別に調査する。5.1 節では研究領域ネットワークの作成法について示し、5.2 節では研究領域ネットワークにおけるクラスタ係数による評価、5.3 節では研究者の移動状況・重複応募状況による評価を行う。

5.1 研究領域ネットワークの作成法

研究領域ネットワークは、ノードを各細目とし、 α 以上の採択移動率（4.2 節）があるノード間をリンクでつなぐことにより形成される。

α を決定するために、 α の値別に平成 17 年度の全 358 細目による研究領域ネットワーク（有向グラフ）を構築し平均経路長とクラスタ係数を求める。ただし平均経路長の計算では、到達が不可能なノード対の平均経路長は除外し、クラスタ係数は研究領域ネットワークを無向グラフとみなし求める。平均経路長は α が 4 のときに最大となり、 α が 5 以上のとき急激に減少する。クラスタ係数は α が増加するにつれ減少する。このようなことから以下では、少数のリンクにより多数のノードが連結している α が 5 の場合を採用する。ノードの

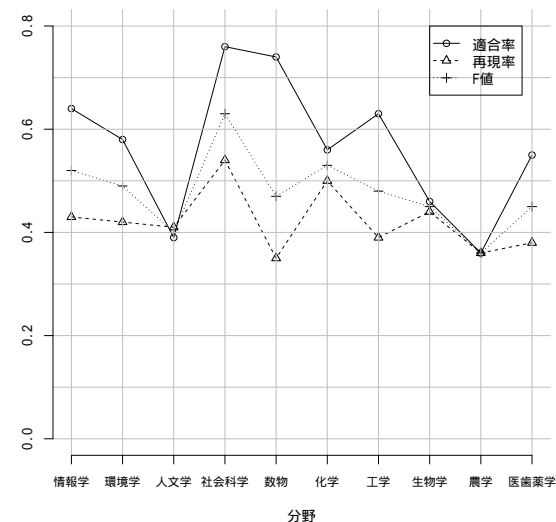


図 2 分野別の適合率・再現率・F 値

Fig. 2 Precisions, Recalls and F Measures in Each Area

色は分野を、大きさは応募件数に比例している（図 3）。

5.2 研究領域ネットワークにおけるクラスタ係数による評価

各分野の研究領域ネットワークにおけるクラスタ係数（図 4）による評価を行う。クラスタ係数が高いネットワークは、密結合している部分が多く存在している。

情報学・人文学・社会科学の研究領域ネットワークでは、クラスタ係数が 0.1 以上であることから、これらの分野の研究領域ネットワークでは、他の分野よりも密結合している部分が多く存在していることがわかる。

5.3 細目別の研究者の移動・重複応募状況による評価

基盤研究・若手研究において継続課題終了後に異なる細目へ応募する研究者の割合（図 5 上）と、基盤研究・若手研究で採択されている研究者が挑戦的萌芽研究へ応募する研究者の割合（図 5 下）を分野別に集計した。図 5 下における「同一の細目」とは、基盤・若手研究と挑戦的萌芽研究で同一の細目に重複して応募している割合を示し、「異なる細目」とは、

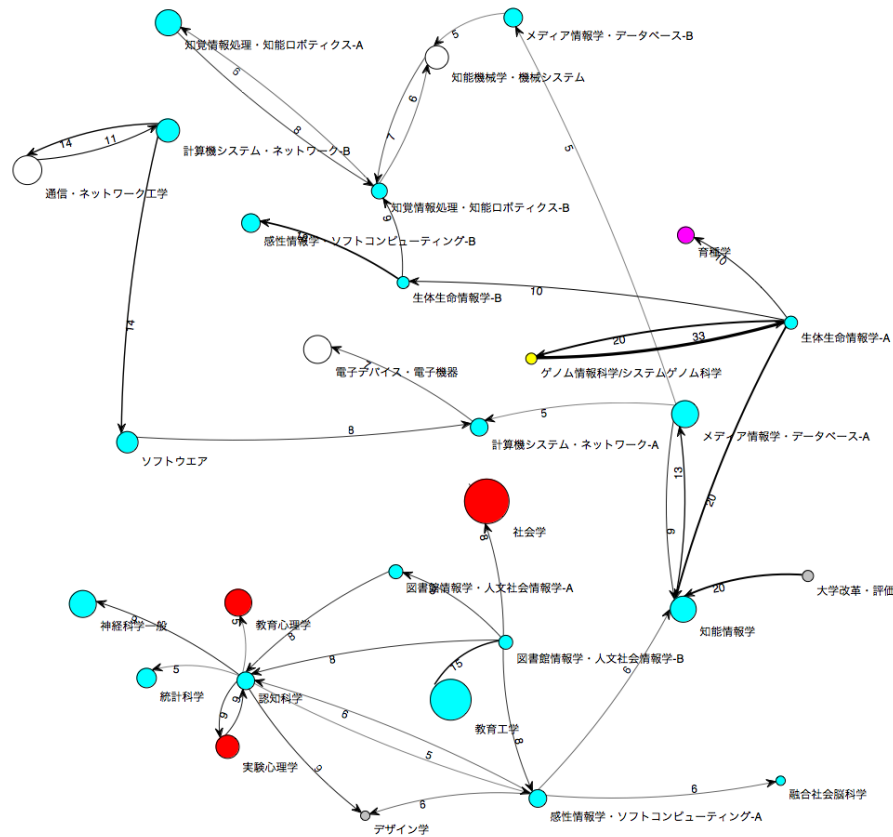


図3 情報学における研究領域ネットワーク (平成21年度)
Fig.3 Research Field Network in Informatics in 2009

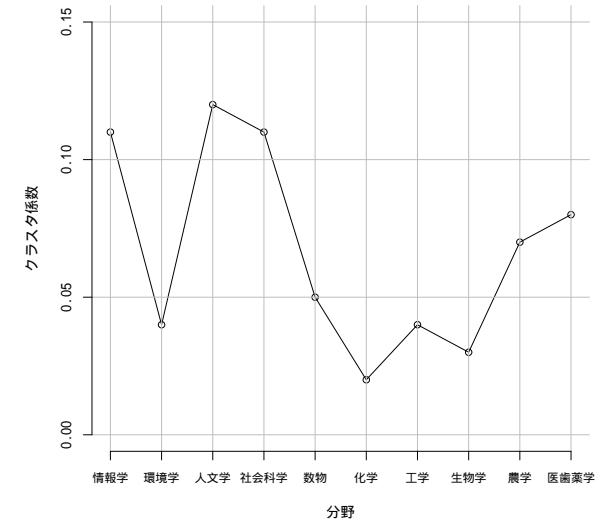


図4 研究領域ネットワークにおけるクラスタ係数
Fig.4 Clustering Coefficients of Research Field Networks

基盤・若手研究と挑戦的萌芽研究で同一の細目に重複して応募している割合を示している。
 基盤研究・若手研究において継続課題終了後に次年度にて新規で応募した研究者のうち、異なる細目へ応募する研究者の割合は、環境学で35.8%と最も高く、数物で10.7%と最も低いことがわかる。また、基盤・若手研究において採択されている研究者のうち挑戦的萌芽研究で異なる細目に応募している研究者の割合は、人文学では1.9%、社会科学では2.9%と低い。細目別に調査すると、社会科学では、経済学と心理学で重複応募数が多い一方、人文学では重複応募数が4名を超える細目はない。これは、人文学における予測結果においてF値が低い(図2)原因のひとつであると考えられる。以上から提案手法が有効である分野(情報学・社会科学)では、クラスタ係数が大きく、一定数以上の重複応募数が存在していることがわかる。

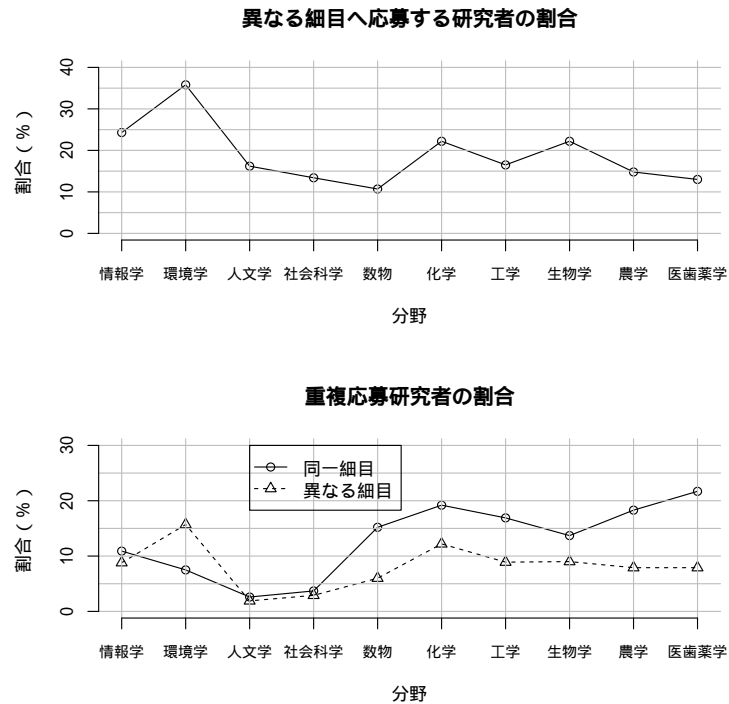


図 5 異なる細目へ移動する割合と異なる細目へ重複応募する割合
Fig. 5 The Ratio of Researchers who Change Research Fields and Researchers who Propose on Duplication

6. まとめとこれからの課題

研究者の研究履歴の経年変化を基盤研究・若手研究と挑戦的萌芽研究における重複応募状況によって予測する手法を提案した。その結果、提案手法が有効である分野では、研究領域ネットワークにおけるクラスタ係数が大きく、一定数以上の重複応募数が存在していることがわかった。

これからの課題は、細目対を移動する研究者数が急激に増加しているなど大きな変化が起こっている細目対において研究代表者と研究分担者による研究者ネットワークの経年変化を調査し、これらにおける共通点を見つけ出すことである。これらにおいて共通点が見つかれば、研究領域ネットワークの変化を予測する精度を向上させることが期待される。

謝辞 本研究を進めるにあたり貴重なご意見をいただいた日本学術振興会 学術システム研究センターの研究員の先生方に感謝いたします。

参考文献

- 1) 山下長義, 沼尾正行, 市瀬龍太郎: 学術動向を把握するための研究経歴抽出法と予測モデルの提案, 第 52 回数理解社会学会大会 (JAMS52) 第 7 回ネットワークが創発する知能研究会 (JWEIN'11) 論文集 (2011).
- 2) Newman, M. E. J.: Coauthorship Networks and Patterns of Scientific Collaboration, *Proceedings of the National Academy of Sciences*, pp.5200-5205 (2004).
- 3) 阪 彩香, 伊神正貫, 桑原輝隆: サイエンスマップ 2008 論文データベース分析 (2003 年から 2008 年) による注目される研究領域の動向調査, *NISTEP REPORT No.110* (2010).
- 4) 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, *人工知能学会知識ベースシステム研究会資料*, Vol.84, pp.21-26 (2009).
- 5) 佐藤和宏, 市瀬龍太郎, 栗原 聡, 相澤彰子, 沼尾正行: 科学研究費申請データの解析, 第 25 回ファジィシステムシンポジウム講演論文集 (CD-ROM) (2009).
- 6) Bruce William Herr II, Edmund M. Talley, Gully A. P. C. Burns, David Newman, and Gavin Larowe.: The NIH Visual Browser: An Interactive Visualization of Biomedical Research, *International Conference on Information Visualisation*, pp. 505-509 (2009).
- 7) 山下長義, 沼尾正行, 市瀬龍太郎: 科研費における応募細目の変遷による細目間の関係抽出とその予測, *情報処理学会研究報告*, Vol.2010-CS-161, No.2 (2010).
- 8) 独立行政法人日本学術振興会: 平成 23 年度科学研究費補助金公募要領 (特別推進研究, 基盤研究, 挑戦的萌芽研究, 若手研究 (A・B)), http://www.jsps.go.jp/j-grantsinaid/03_keikaku/h23_download.html (2010). (参照 2011-7-14).