「第19回年次大会予稿」

キーワード分析による基本計画重点分野の動向調査

A trend investigation of the Science and Technology Basic Plan important fields by keyword Analysis

西澤正己*, 孫媛

Masaki NISHIZAWA*, Yuan SUN

国立情報学研究所 情報社会相関研究系

Information and Society Research Division, National Institute of Informatics

〒101-8430 千代田区一ツ橋2-1-2

Email: { nisizawa, yuan}@nii.ac.jp

*連絡先著者 Corresponding Author

我々はこれまでにキーワード分析により、ゲノム関連、ナノテクノロジー関連、情報関連等、科学技術基本計画の重点領域研究分野の動向とその広がりについて調査している。ここでは科学研究費補助金における細目分野相互の関係を各細目分野に含まれる研究課題名から抽出したキーワードの関連度を用いて多次元尺度法により図示し、その中での科学技術基本法の重点分野関連の分野の位置関係を見た。さらに、これらの重点領域を中心に研究している研究所や研究助成金の研究課題名から、科学研究費補助金の細目分野との関係の分析をおこなった。

We have been carrying out quantitative analysis of the trends and extent of important territory research fields of the Science and Technology Basic Plan by keyword analysis. In this study, we show the relationship of research items of Grants-in-Aid for Scientific Research among each other by using Multidimensional Scaling method. Furthermore, relations among the research items are also analyzed based on research subjects and grant themes of those laboratories focusing on researches around these important areas.

キーワード: 科学技術基本計画, 環境研究, キーワード分析, 関連分析, 研究動向

Keyword1, Science and Technology Basic Plan, environmental field-related research, keyword analysis, relationship, research trend

1 はじめに

政府および総合科学技術会議は科学技術基本法のもと、1996年から第1期、第2期科学技術基本計画[1]を進めてきた。第2期科学技術基本計画（平成13～17年度）では重点分野として「ライフサイエンス分野」、「情報通信分野」、「環境分野」、「ナノテクノロジー・材料分野」が特に重点を置かれ、資源配分が重点的に行われてきた。これらの分野は第3期計画（平成18～22年度）においても、引き続き重点分野とされている。

文部科学省科学研究費補助金(科研費)[2]は大学等の研究者に対しての個人申請・ピアレビュー審査による研究費配分制度である。この制度は1939年に始まり、古くから基礎研究の下支え、底上げの役目を果たしてきたと言われている。現在では、科研費以外にも多くの競争的資金が配分されているものの、研究費総額では、科研費が約半額を占めている。この科研費において、科学技術基本計画の重点分野に対してどのように研究費が配分され、どのように変化してきたかは興味を持たれるところである。

我々は、科学研究費補助金データベース(KAKEN, KAKENK)[3]からこれらの重点領域に対するキーワード(トピック関連キーワード)[4][5]を抽出したうえで、科研費のすべての採択課題の課題名からこれら重点領域(特に、ライフサイエンス分野のゲノム関連分野とナノテクノロジー・材料分野の2分野)に関連する課題を選び出し、配分額の変遷を調べた。また、科研費の分科細目表に対しても同様に特徴的なキーワードを抽出し、それらの一致度を見ることにより、重点領域との関連分野を対応分析により示した[6]。さらに、キーワード群間の近似度を測ることにより、環境関連分野を特に国立環境研究所の研

究成果データを用いて科学研究費補助金のデータとの比較を行った[7]。

ここでは、科学研究費補助金における細目分野相互の関係を各細目分野に含まれる研究課題名から抽出したキーワードの関連度を用いて多次元尺度法により図示し、その中での科学技術基本法の重点分野関連の分野の位置関係を見る。さらに、これらの重点領域を中心に研究している研究所や研究助成金の研究課題名から、科学研究費補助金の細目分野との関係の分析をおこなった結果を述べる。

2 データ

2.1 科学研究費補助金データ

国立情報学研究所では、科学研究費補助金の採択課題の速報を収録した「科学研究費補助金採択課題データベース(KAKENK)」[3]およびその成果概要を収録したデータベースである「科学研究費補助金成果概要データベース(KAKEN)」[3]が作成されている。KAKENKにおいては1996年度から最新の2010年度までの採択課題(現在では遡及入力が進み、一部の項目を除いては1965年からのデータが収められている)、KAKENにおいては1985年度から2008年度までの成果概要が収録されている。KAKENKデータベースには採択課題名、研究種目、分科細目(大型研究費、奨励研究を除く)、新規・継続の別、配分額等が収録されており、KAKENについてはKAKENKの情報に加え、成果概要、キーワード、発表文献リスト等が収録される。なお、KAKENデータベースは成果概要の報告書を収録するため、採択後少なくとも約1年半後の課題が収録されていることになる。今回の分析には1976 - 2009年度のKAKENKデータベースを使用した。今回の報告では連続的な経年変化

表 1 : 科学研究費補助金分科細目表(2009 年度)

[総合・新領域系] [Synthesis/New Area]			[人文社会系] [Humanities/Social Science]			[理工系] [Science/Engineering]		
分野	分科	細目番号	分野	分科	細目番号	分野	分科	細目番号
総合領域	情報学	情報学基礎	文学	哲学	哲学・倫理学	理学	数学	代数学
		1001			2801			4001
		ソフトウェア			2802			4002
		1002			2803			4003
		計算機システム・ネットワーク			2804			4004
		1003			2805			4005
		メディア情報学・データベース			2806			4006
		1004			2807			4007
		知能情報学			2808			4008
		1005			2809			4009
総合領域	情報学	知能情報学・知能ロボティクス	文学	言語学	言語学・言語学・言語学	理学	物理学	物理学・物理学・物理学
		1006			2810			4010
		知能情報学・知能ロボティクス			2811			4011
		1007			2812			4012
		情報情報学・人文社会情報学			2813			4013
		1008			2814			4014
		情報情報学・ソフトウェア			2815			4015
		1009			2816			4016
		情報情報学・ソフトウェア			2817			4017
		1010			2818			4018
総合領域	情報学	情報情報学・ソフトウェア	文学	言語学	言語学・言語学・言語学	理学	物理学	物理学・物理学・物理学
		1011			2819			4019
		情報情報学・ソフトウェア			2820			4020
		1012			2821			4021
		情報情報学・ソフトウェア			2822			4022
		1013			2823			4023
		情報情報学・ソフトウェア			2824			4024
		1014			2825			4025
		情報情報学・ソフトウェア			2826			4026
		1015			2827			4027
総合領域	情報学	情報情報学・ソフトウェア	文学	言語学	言語学・言語学・言語学	理学	物理学	物理学・物理学・物理学
		1016			2828			4028
		情報情報学・ソフトウェア			2829			4029
		1017			2830			4030
		情報情報学・ソフトウェア			2831			4031
		1018			2832			4032
		情報情報学・ソフトウェア			2833			4033
		1019			2834			4034
		情報情報学・ソフトウェア			2835			4035
		1020			2836			4036

の分析が終わっていないので、2009年度の結果を中心に報告する。2009年度の科学研究費補助金分科細目表を表1に示す。

2.2 研究所および助成金の研究課題データ

本分析では、研究所の研究成果データとして、重点領域の4分野（環境、情報、ライフサイエンス、ナノテクノロジー）をおこなっている産業技術総合研究所(AIST)[8]、環境関連の研究を行っている国立環境研究所(NIES)[9]、気象庁気象研究所の環境・応用気象研究部(MRI)[10]の研究成果リストから研究課題を抽出した。また、環境研究の助成金として環境研究総合推進費(旧地球環境研究総合推進費)[11]、および環境技術開発等推進費[12]の研究課題を用いた。

3 分析手法および結果

3.1 キーワード分析

キーワード分析の手法としては、前回の分析[7]と同様に、課題名から形態素解析システム茶筌[13]を用いてできるだけ短い複合語になるようにキーワード抽出をおこなった。ただし、科学研究費補助金の細目分野で広く共通に使われる、「研究」、「開発」、「解明」、「構築」、「システム」、「機能」、「構造」、「応用」、「形成」、「解析」、「利用」、「モデル」の12キーワードは全細目の80%以上で使用されていたので、今回は分析対象から外した。本来は共通性のみだけでなくその集中度も考慮すべきであるが、今回は間に合っていない。

このようにして抽出したキーワードを細目分野、あるいは研究所、助成金の研究分野ごとにグループ化した上で各キーワード群間の近似度を以下の余弦公式を用いて計算した。重み W は今回の分析では1として計算している。

$$Sim(A,B) = \frac{\sum_{i=1}^n W(A_i) \times W(B_i)}{\sqrt{\sum_{k=1}^n (W(A_k))^2} \times \sqrt{\sum_{j=1}^n (W(B_j))^2}}$$

3.2 科学研究費補助金の細目分野間の関連分析

上の近似度計算式を用い、科学研究費補助

金の細目分野（2009年度では時限付き細目分野を除いて284細目分野）に対して、各細目分野に含まれる研究課題名から抽出したキーワード群間の（2009年度では284×284）近似度を計算した。この結果を用いて、多次元尺度法を用いて2次元の各細目分野の位置関係を示したのが図1である。図中の各マーカーに付けられたラベルは細目表の細目番号（表1）を表している。図中の上方には理工系分野、右方には人文系分野、左下方には生物系分野がほぼグループ分けして配置されているのがわかる。また、中央寄りに複合分野、中央から遠くに行くほど専門性が高い分野が配置されているものと思われる。この中で、科学技術基本計画の重点領域である4分

野（環境、情報、ライフサイエンス中のゲノム関連、ナノテクノロジー）の状況を見る。

環境関連では環境学として、環境動態解析(2001)、環境影響評価・環境政策(2002)、放射線・化学物質影響科学(2003)、環境技術・環境材料(2004)の4細目がある。これらの細目は図中のほぼ中央に位置し、複合的な分野であることを良く示している。この中では環境動態解析(2001)が中央にあり、環境影響評価・環境政策(2002)が社会系分野寄り、放射線・化学物質影響科学(2003)が生物系寄り、環境技術・環境材料(2004)が工学分野寄りの傾向を示していることがわかる。

情報関連分野は情報学として11細目(表1中の細目番号1001から1011まで)ある。これらの分野はほぼ図中の中央から右上方向に分布している。この中では情報学基礎(1001)が数学関連分野の近く、情報図書館学・人文社会情報学(1008)が人文系寄り、生体生命情報学(1011)が生物系分野寄りに位置している。

ゲノム関連分野はゲノム科学として基礎ゲノム科学(2301)、応用ゲノム科学(2302)、ゲノム情報科学(2303)の3分野がある。基礎ゲノム科学(2301)は遺伝・ゲノム動態(5701)等の基礎生物学(5701-5706)の近くに位置している。応用ゲノム科学(2302)は農学や医学寄りの中央に、ゲノム情報科学(2303)はゲノム系と情報系の中間に位置している。

ナノテクノロジー関連分野はナノ・マイクロ科学としてナノ構造科学(2102)、ナノ材料・ナノバイオサイエンス(2102)、マイクロ・ナノデバイス(2103)の3分野がある。これらの3分野は左上の部分に集中しており、応用物理、化学系、材料工学系の間に位置している。

このようにこの分析では研究分野の広さ

や特徴を良く表すことができていると思われる。現時点では時系列分析が途上であるが、この分析から研究分野の動向も分析できるのではないかとと思われる。

3.3 研究所および助成金の研究課題と科研費の関連分析

3.2章の分析と同様に、産業技術総合研究所(AIST)、国立環境研究所(NIES)、気象庁気象研究所の環境・応用気象研究部(MRI)の研究成果リストから研究課題を抽出して、科学研究費補助金の細目分野との近接度を計算した。また、環境研究の助成金として環境研究総合推進費(旧地球環境研究総合推進費)、および環境技術開発等推進費の研究課題についても同様の計算をおこなった。

産業技術総合研究所(AIST)については、研究成果に分野が付与されており、この内の環境・エネルギーをAIST-Env、情報通信・エレクトロニクスをAIST-Info、ナノテクノロジー・材料・製造をAIST-Nano、ライフサイエンス分野をAIST-Lifeと分類してキーワード群を作成した。

国立環境研究所(NIST)は前回の分析[7]と同様に研究グループ、研究プロジェクトごとに17の分類:Aerosphere(大気圏)、AreaEnv(地域環境)、BioEnv(生物環境)、ChemEnv(化学環境)、EarthEnv(地球環境)、Env-risk(環境リスク)、HealthEnv(健康環境)、Infra-Info(技術・情報)、Recycle(循環型社会)、SocEnv(社会環境)、WaterClod(水土圏)、の11の研究領域とBasinZone(流域圏環境管理プロジェクト)、Bio-diversity(生物多様性プロジェクト)、dyoxin-proj(環境ホルモン・ダイオキシンプロジェクト)、ozone(成層圏オゾン変動プロジェクト)、PM-DEP(PM2.5-DEPプロジェクト)、

WarmerClimate（地球温暖化プロジェクト）の6つの研究プロジェクトに分類し、研究成果の課題名よりキーワード群を作成した。

気象庁気象研究所の環境・応用気象研究部(MRI)については日本語の研究成果からキーワード群を作成し、JMA-Jとした。

環境研究総合推進費（旧地球環境研究総合推進費）については、研究対象分野としてⅠ．全球システム変動(EEF-1_WholeEarth)、Ⅱ．越境汚染(EEF-2_CrossBorderPoll)、Ⅲ．広域的な生態系保全・再生(EEF-3_WideBiogeocen)、Ⅳ．持続可能な社会・政策研究(EEF-4_Sustainable)、Ⅴ．地球環境研究革新型研究領域(EEF-5_InnovativeRes)があり、実施対象課

題名からそれぞれキーワード群を作成した。環境技術開発等推進費（EnvTech-all）についても同様にキーワード群を作成した。

以上の28の研究所・助成金の研究分野のキーワード群と3.2章と同様の284の科研費分科細目のキーワード群について近接度 28×284 の近接度を計算し、関連度が0.2以上になった組み合わせを選んだ上、さらに研究所・助成金分野と関連度が高いそれぞれ上位10細目を選んだ。この結果77の細目分野が選ばれ、この 28×77 のデータについて、距離に換算した上で、対応分析をおこなった。その結果を図2に示す。

図1と分析手法は違っており、また、一部だけの表示であるので、科学研究費野細目分

図2：28の研究所・助成金分野と関連の強い科学研究費補助金の77分野との対応分析。

野のみの多次元尺度法による分布と細かい点での違いはあるが、概ね同様の分布が得られている。この細目分野と研究所・助成金の関連度については、特徴的な結果が得られている。

産業技術総合研究所(AIST)の4領域については、AIST-Nano, AIST-Info, AIST-Envの領域が左方向に集まり、工学、化学的要素が強い結果となっている。AIST-Lifeについては、選ばれた中で唯一のライフサイエンス関連の研究であったので、左下方向の生物系方向に位置する結果となった。象研究所の環境・応用気象研究部(JMA-J)については、右上に位置し、個体地球惑星物理学(4001)や超高層物理学(4003)と特に関連が強い。他の国立環境研究所(NIST)の研究領域や助成金の分布は概ね中央付近に集まって分布している。中でも、NAISTのAerosphere(大気圏)やozone(成層圏オゾン変動プロジェクト)は気象研究所(JAM-J)の近くに示され、WarmerClimate(地球温暖化プロジェクト)はAerosphereやJMA-Jよりも国際関係論(3502)や国際法学(3403)、気象・海洋物理・陸水学(4402)に近い位置にあることがわかる。また、これらの環境関連分野においては、全般に上方寄りに気象系、右寄りに社会系、下方寄りに生物系、右寄りに工学系が分布しているように思われる。

4 まとめ

本論文ではキーワード分析の手法を用いて、科学研究費補助金の細目分野の関連度を計算し、相互の位置関係を多次元尺度法を用いて図示した。また、今回は環境研究部分が主であるが、重点領域の4分野(環境、情報、ライフサイエンス、ナノテクノロジー)に近

い領域の研究動向として、研究所や助成金の研究課題と科学研究費補助金の細目分野との関係について対応分析を用いて調査した結果を示した。まだ、暫定結果ではあるが、キーワード群の相互の類似度と多次元尺度法や対応分析により、科学研究費補助金の細目分野の相互関係や、それと研究所や助成金の研究課題郡との位置関係がよく示されることがわかった。

今後は不要な広い意味のキーワードの除外や、類似度計算の重みの最適化、キーワードの組み合わせによる個別の関連課題の抽出方法等についても考察していく予定である。また、他の研究所の研究課題や助成金の研究課題との関連、時系列変化を追った動向分析も行う予定でいる。

参考文献

- [1] 総合科学技術会議、科学技術基本計画, <http://www8.cao.go.jp/cstp/kihonkeikaku/index3.html>
- [2] 科学研究費補助金, <http://www.jsps.go.jp/j-grantsinaid/index.html>
- [3] 2005年4月からKAKENK, KAKENデータベースを統合して科学研究費補助金データベース(KAKEN)としてサービスされている。: <http://seika.nii.ac.jp/>
- [4] 西澤 正己、孫 媛, 「科研費分野変更に伴う情報学分野の変遷マッピング」, 情報知識学会 第12回研究報告会講演論文集, 2004年5月, pp. 9-12
- [5] Masaki Nishizawa, Yuan Sun, "Research Fields related to Information Science under a new classification in Japan", Proceedings on 4th International

Conference on University Evaluation and Research Evaluation, pp 67-73, Whan University, China, 27 September (2004)

[6] 西澤 正己、孫 媛, 「キーワード分析による科研費におけるゲノムおよびナノテクノロジー関連研究の動向調査」, 情報知識学会誌, Vol. 17, No. 2, pp. 117-122, 2007

[7] 西澤 正己、孫 媛, 「キーワード分析による環境関連研究の動向調査」, 情報知識学会誌, Vol. 20, No. 2, pp. 155-162, 2010

[8] 産業総合研究所研究成果 : http://www.aist.go.jp/aist_j/research/index.html

[9] 国立環境研究所研究成果 : <http://www.nies.go.jp/db/index.html#seika>

[10] 気象庁気象研究所研究成果 : <http://www.mri-jma.go.jp/Publish/publish-sjis.html>

[11] 環境研究総合推進費 : <http://www.env.go.jp/earth/suishinhi/index.html>

[12] 環境技術開発等推進費 : <http://www.env.go.jp/policy/tech/suishin.html>

[13] 形態素解析器: 茶釜システム, <http://chasen-legacy.sourceforge.jp/>