

KAKEN を用いた研究動向の把握支援

古城 隆彦[†] 佐野 雅彦[†]

徳島大学[†]

1. はじめに

大学等の研究組織や研究者において、自身の研究に関連する研究分野のテーマや研究内容（以下、研究動向）を把握することは重要である。しかし、それには多大な労力を必要とする。本研究は、その把握を容易にし、その活動を支援することを目的とする。

先行研究[1]では、科学研究費助成事業（以下、科研費）のデータベース(KAKEN)[2]を基に研究組織の研究を分類し、科研費と所属組織の各研究分野への注力度合いとその推移を可視化している。しかし、2018年度からは、それ以前の「系・分野・分科・細目表」で構成される「研究分野」（以下、旧分類）から「大区分・中区分・小区分」で構成される「審査区分」（以下、新分類）に変更されたことから、研究分野の分類方法が異なる年度間の研究動向を把握することが困難となった。

そこで本研究では、旧分類における採択研究課題（以下、研究課題）を新分類に対応させることにより、新旧分類が混在する長期間の研究動向を把握することを目標とする。

2. 提案手法

KAKEN 情報から旧分類と新分類の研究分野の関連度を計算し、旧分類における研究課題を新分類に対応させることで分類の方法を統一した後、全体の研究動向の分析結果を可視化する。そのシステムの概要を図1に示す。

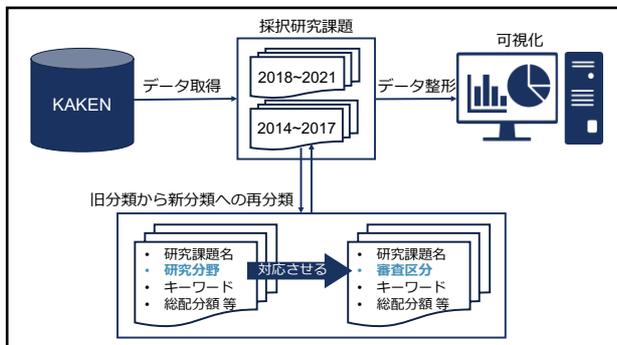


図1 提案システムの概要

Support for understanding research trends using KAKEN
Takahiko Kojot, Masahiko Sano†, †Tokushima University

2.1. 情報取得と前処理

本研究では KAKEN から研究課題の研究課題名、採択年度、キーワード、研究分野、審査区分、総配分額の情報を取得する。データの前処理として、旧分類及び新分類における研究分野毎のキーワードを抽出し、それらを正規化したものをデータセットとした。このデータセットを用いて scikit-learn の TfidfVectorizer によって tf-idf モデルの作成を行い、各分野のキーワード群をベクトル化して特徴量とした。

2.2. 研究分野関連度

文書間の類似度の算出は様々な手法[3][4]があるが、本研究では旧分類と新分類の研究分野関連度を、前節で求めた各特徴量間のコサイン類似度とし、旧分類に対する関連度上位の新分類を出力結果とした。出力例を表1に示す。

表1 旧分類に対する関連度上位の新分類

| 旧分類：ソフトウェア/計算基盤/情報学/総合系 | | |
|-------------------------|----------------------|---------|
| 順位 | 新分類 | 関連度 |
| 1 | 小区分 60050:ソフトウェア関連 | 0.57816 |
| 2 | 小区分 60010:情報学基礎論関連 | 0.10967 |
| 3 | 小区分 60070:情報セキュリティ関連 | 0.10644 |
| 4 | 小区分 60040:計算機システム関連 | 0.09666 |
| 5 | 小区分 60090:高性能計算関連 | 0.09521 |

2.3. 採択研究課題の再分類

旧分類で登録されている研究課題を新分類に対応させるため、以下2つの手法で再分類を行う。

- 2.2 で求めた関連度計算結果を基に再分類
- 2.1 で求めた特徴量を用いて、研究課題名及び付与されているキーワードを基に再分類

手法 2) は研究分野が登録されていない研究課題に対しても分類を可能にするために検証として実験を行う。再分類結果を研究課題の新分類として登録する。

2.4. 研究動向の可視化

研究課題の分類を統一した後に、KAKEN に登録されたデータを、Python を用いて分析し、研究動向を可視化する。可視化の概要としては主に以下2つの機能で構成する。

・全体の動向

全体の動向は、各年度における全分野の動向を可視化する機能である。ユーザが一目見て、どの分野が盛んであるかが把握できるグラフを想定する。先行研究[1]と同様の各年度の課題数

や総配分額を表示する要素に加えて、新たに追加する機能を検討する。表示例を図2に示す。



図2 全体図 (2020年度における各区分の課題数)

・各区分の動向

各区分の詳細図は、新分類の最小分野単位である小区分ごとの動向を可視化する。採択年度ごとの頻出キーワードや課題数、補助金額の時系列変化や、それらの前年度と比較した増減傾向も表示する。表示例を図3に示す。

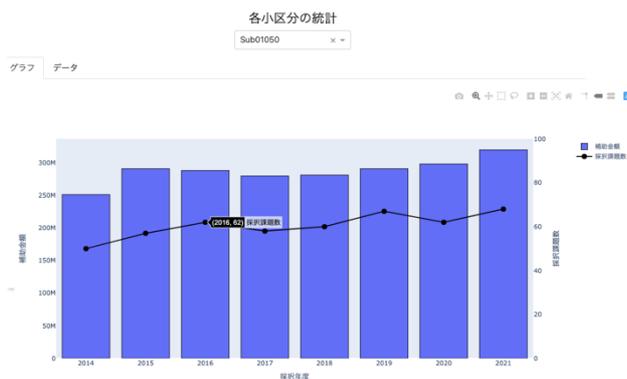


図3 小区分 01050 の課題数・補助金額の推移

3. 評価実験

旧分類から新分類への対応及び再分類結果の妥当性の評価を目的とし、大学の研究室に配属されている学生 38 名を対象にアンケート回答を 2021 年 2 月 2 日～2021 年 2 月 7 日に実施した。

3.1. 実験概要

評価アンケートの構成は以下の通りである。

① 手法 1) の妥当性の評価

各旧分類に対して関連度の高い新分類の上位 1, 2 位を表示し、その対応の妥当性の評価。

② 被験者による研究課題の分類結果との比較

ランダムに抽出された旧研究課題に対して、被験者に適切な新分類を選択してもらい、上記①の結果 (上位 5 件) に含まれる割合の評価。

③ 手法 1) と手法 2) の比較及び評価

ランダムに抽出した研究課題 5 つに対して下記 3 つの分類方法における再分類結果より、被験者に研究課題の課題名・キーワード・旧分類名の

情報を基に、どの分類方法が適切であるか評価。

- ・分類方法 1: 手法 1)
- ・分類方法 2: 手法 2) のキーワードによる分類
- ・分類方法 3: 手法 2) の研究課題名による分類

3.2. 実験結果

①より、約 95% (評価*が 3, 4 の合計) の人が概ね妥当であると判断した (表 2)。②より、被験者の選択した新分類が、①の関連度上位 5 つに含まれる割合は 55%であった。③より、分類方法 1 が最も平均値が高い結果であった (表 3)。

表 2 ①の結果

| 評価* | 1 | 2 | 3 | 4 |
|------|------|------|-------|-------|
| 回答数 | 0 | 2 | 14 | 22 |
| 回答割合 | 0.0% | 5.3% | 36.8% | 57.9% |

評価*: {妥当でない [1・2・3・4] 妥当である}

表 3 ③の結果 (評価**の平均値)

| | 分類方法 1 | 分類方法 2 | 分類方法 3 |
|----|--------|--------|--------|
| 平均 | 3.34 | 3.14 | 2.97 |
| 分散 | 0.68 | 0.61 | 0.88 |

評価**: {適切でない [1・2・3・4] 適切である}

3.3. 考察

手法 1) による旧分類から新分類への対応の評価は妥当な結果であることが示されたが、異なる被験者層を加えた評価が必要である。手法 2) については、その手法を含め改善の余地がある。特に分類方法 2 は研究課題による登録されたキーワード数のばらつきが原因であると考えられる。

4. おわりに

本研究は長期間の研究動向把握の支援のため KAKEN を用いて新旧分類を統一した後に、研究動向を把握する手法を提案した。

今後の展望として、研究動向の把握支援だけでなく、今後伸びる分野の推測や、ある分野における研究題目の重複確認や推奨テーマの表示等の機能追加を検討している。それにより動向を知りたい研究者だけでなく、新たに研究を始める学生への研究テーマ設定の支援にも役立つシステムになると考えられる。

謝辞 本研究は JSPS 科研費 JP18K11572 の助成を受けたものです。

参考文献

[1] 竹野凌ほか, “研究組織における研究分野の可視化”, AXIES 2017 年度年次大会論文集, FA3-5, 2017.
 [2] KAKEN 科学研究費助成事業データベース, <https://kaken.nii.ac.jp> (2022 年 1 月 5 日参照)
 [3] 野澤孝之ほか, “シラバスの文書クラスタリングに基づくカリキュラム分析システムの構築”, 情報処理学会論文誌, vol. 46, No. 1, pp. 289-300, 2005.
 [4] 川谷隆彦, “多文書間の共通性分析による文書クラスタリング”, 情報処理学会研究報告, No. 154, pp. 93-100, 2003