

Content Analysis of the Bulletin of Musashino University Institute of Environmental Sciences Using Text Mining

メタデータ	言語: jpn 出版者: 公開日: 2023-04-20 キーワード (Ja): キーワード (En): 作成者: 伊藤, 伸也 メールアドレス: 所属:
URL	https://mu.repo.nii.ac.jp/records/2100

テキストマイニングを用いた環境研究所紀要の内容分析

Content Analysis of the Bulletin of Musashino University Institute of Environmental Sciences Using Text Mining

伊藤 伸也*
Shinya Ito

1. はじめに

本稿は、武蔵野大学学術機関リポジトリ¹⁾に登録されている、武蔵野大学環境研究所紀要（以下、「論文」とする。）を、テキストマイニングを用いて内容分析を試みたものである。

論文は2012年3月1日に第1号が発行されて以降毎年度発行が継続し、直近では2022年2月28日に第11号が発行した。この11年間の論文には、論文数として計94報の知が蓄積されている。著者が本校大学院博士後期課程在籍時に、次の2点について考える場面が多々あったのだが、残念ながら答えを見出せずに課程を修了することとなった。

環境研究所は

- ① どのような視点から環境にアプローチをしているのか
- ② どのような環境研究領域に強みを有しているのか

今般、論文のタイトルおよび本文に出現する語を抽出し、その語のつながりを視覚的に表示することで、上記①②に対する自身の解を見つけるヒントを得ることができるのではないかと考えた。語と語を視覚的に結びつけるツールであるテキストマイニングを活用し、論文の内容分析を試みたので報告をする。

2. 分析方法

①分析ツールと対象論文

分析ツールには、「KH Coder²⁾」を用いた。また分析の対象論文は、武蔵野大学学術機関リポジトリに登録されている紀要で、日本語で書かれた計90報^{*1)}とした。

* 環境研究所 客員研究員

受理日：(2022年10月29日)
発行日：(2023年3月1日)

②分析手順

分析は以下の手順で実施した (図1)。

手順1：分析対象の論文の「発行号、タイトル (日本語)」を Excel ファイルに抽出し、分析の基礎データ (以下、「基礎データ」とする。) を整理した (図2)。

手順2：KH Coder を起動し、ツールバーにある「前処理 (R)」の“テキストのチェック”および“前処理の実行”の順に実行した。

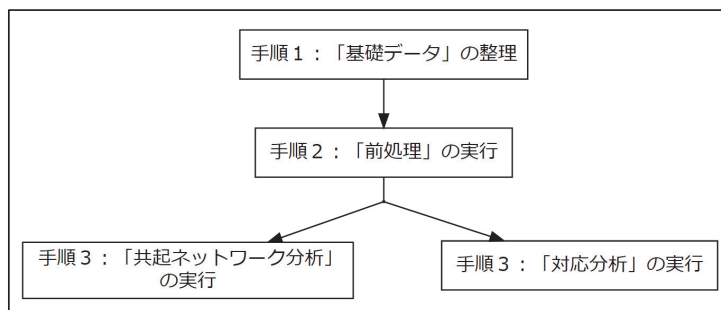
手順3：手順1, 2の基礎データを用いて、“共起ネットワーク^{*2)}分析”および“対応分析^{*3)}”を実行した。

*1) 各号の目次では計94報が掲載 (2022年8月1日現在) されていたが、実際に閲覧可能な論文は91報、そのうち1報は英語論文であり、差し引き計90報を分析対象と設定した。

*2) 共起ネットワークとは・・・一つ一つの文書で出現する単語「抽出語」のうち、「距離」が近いか遠いかを計算し、図示したもの³⁾。

*3) 対応分析とは・・・対応分析では原点 (0,0) 付近に特徴のない、どこにも出現する語が集まるのに対して、原点から離れた語からはデータの特徴が読み取れるもの⁴⁾。

(図1) 分析手順



(図2) 基礎データの Excel サンプル

1	号	タイトル
2	1	2011年夏 武蔵野大学本部校地における放射線量測定
3	1	ポスト京都議定書における「数値目標」の考え方:過度の予防原則からの脱却
4	1	メリトの鑑賞 (Inv.no.S.8493、トリノ博物館蔵) について
5	1	国立大学法人発行環境報告書の現状分析
6	1	持続可能な社会を構築するための教科「家庭」に求められる階層と連結概念:改正学習指導要領を有効に機能させるために
7	1	武蔵野大学における防災情報に関する研究 その5 自治体が提供する外国人向け防災情報の分析をふまえた留学生向け防災情報リーフレットの作成
8	2	環境学を専攻する学生を対象としたライフヒストリー研究
9	2	環境文化によって解釈された鎌倉と地中海周辺の地域風
10	2	企業における温室効果ガス削減費用の算出方法に関するアンケート調査
11	2	旧西本組本社ビル
12	2	暑害による受信障害の緩和対策を施したパラポラアンテナのライフサイクルインベントリ分析
13	2	武蔵野大学キャンパスおよび周辺地域における放射線量測定と除染の展望
14	2	和歌浦「あしべ屋別荘」と夏目漱石
15	3	温室効果ガスの削減費用の計算方法
16	3	価値共創の場が涵養する学生の主体的な学習意欲について:体験学習型授業「自然環境教育演習1」のアプローチ
17	3	旧西本組本社ビルの建造年代と平面の分析
18	3	時間的展望と環境意識や行動との関係
19	3	地域自然エネルギーの導入がもたらす社会への影響:相乗効果の重要性
20	3	東京湾岸・主要河川河口地域における放射線量測定
21	3	武蔵野大学における防災情報に関する研究 その6 東日本大震災発生直後の学生の状況

③ 分析条件

前記分析手順で示した基礎データを加工・整理するにあたり、特記すべき条件を表1に示す。

(表1) 分析実行における特記事項等

項目	条件等
分析対象とした論文の箇所	<ul style="list-style-type: none"> 共起ネットワーク分析では、論文の日本語タイトルを分析対象とした。 対応分析では、「抽出語：タイトル（日本語）と外部変数：号」を分析対象とした。
分析の対象とした語	品詞による語の取捨選択で、「名詞」および「サ変名詞」を分析対象とした。
共起分析の条件（図3-1）	<ul style="list-style-type: none"> 最小出現数：2 現在の設定で利用される語の数：113 描画する共起関係(edge)の選択：上位150
対応分析の条件（図3-2）	<ul style="list-style-type: none"> 最小出現数：3 現在の設定で利用される語の数：53 バブルプロット：バブルの大きさ100% 原点を表示

図3-1 共起分析の条件設定

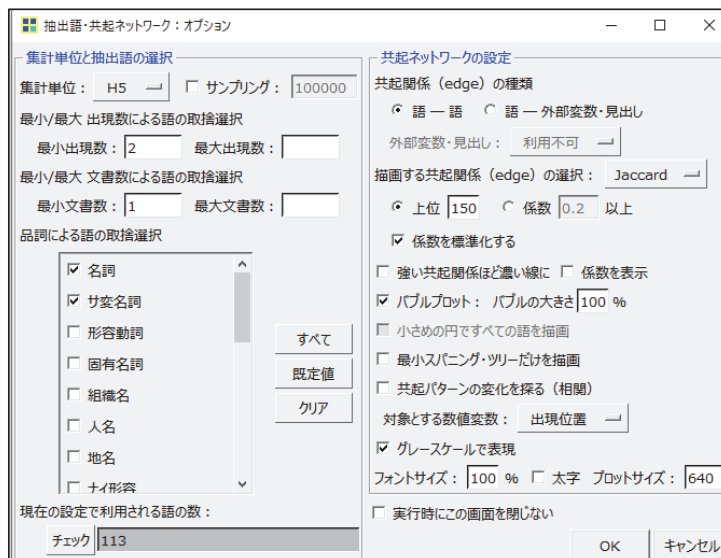
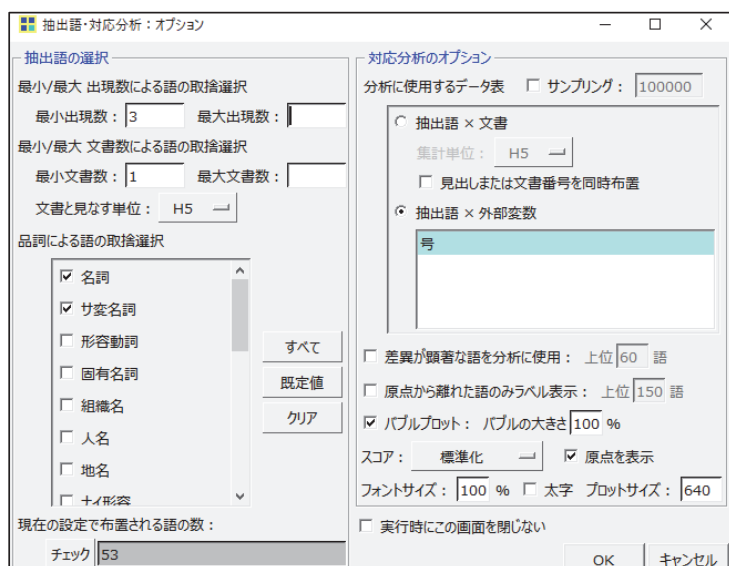


図 3-2 対応分析の条件設定



3. 分析結果

① 共起分析

前章の分析方法に基づき実行した結果を図 4 に示す。本結果より、次の領域での研究報告が多く見られた。

- ・ オゾン、水銀、大気、放射線などの自然科学の領域
- ・ 学生の教育や学習または行動や意識のアンケートや調査に関する領域
- ・ ライフサイクルや企業の温室効果ガス削減費用などの環境経済に関する領域
- ・ 防災に関する領域
- ・ 史跡や建築に関する領域

など

② 対応分析

前章の分析方法に基づき実行した結果を図 5 に示す。本結果より、発行号と出現語に次の特徴が見られた。

- ・ 第 1 号、第 2 号、第 3 号、第 4 号で比較的近接した位置でまとまりが見られた。
- ・ 第 5 号が原点近くに位置していた。
- ・ 第 6 号、第 7 号、第 8 号で比較的近接した位置でまとまりが見られた。
- ・ “環境”という語が原点付近に位置し、かつ頻出語として確認された。
- ・ 第 9 号、第 10 号、第 11 号はそれ以外の号と離れた箇所に位置していた。また、共通的な出現語も少なく、多様な領域でのテーマでの論文が発表される傾向が確認された。

4. まとめ

本稿では、武蔵野大学学術機関リポジトリに登録されている論文について、テキストマイニングを用いて共起分析および対応分析を試みた。以下、本稿の研究目的である「環境研究所は、どのような視点から環境にアプローチをしているのか」および「環境研究所は、どのような環境研究領域に強みを有しているのか」について考察した。

共起分析から見てきたこととして、自然科学の領域、学生へのアンケート調査に関する領域、環境経済に関する領域、防災に関する領域、史跡や建築に関する領域などを主な研究分野とするテーマが多く確認できた。研究目的の一つ目である“環境研究所は、どのような視点から環境にアプローチをしているのか”については、この共起分析の結果が一つの解と考える。次に、対応分析から見てきたこととして、研究目的の二つ目である“環境研究所は、どのような環境研究領域に強みを有しているのか”について整理した。まず、図5に示されている通り、原点付近に“環境”という語が大きなバブルチャートで出現している。当たり前のことかもしれないが、“環境”を中心に据えた研究者集団ということが確認できた。さらに、第1号、第2号、第3号、第4号が図5の左上箇所にとままりが確認できた。このまとまりの特徴としては、放射線や防災に関するテーマや語の繋がりが認められた。論文は2012年3月1日に第1号が発行されたが、この前年の2011年3月11日に東日本大震災が発生したことが、放射線や防災をテーマとした論文が多く出されたと推察する。第6号、第7号、第8号で図5の左下箇所にとままりが確認できた。このまとまりの特徴としては、建築領域や環境経済領域のテーマが以前よりも多く出されてきたことが言える。そして、第9号、第10号、第11号はそれぞれ独立した位置で存在していることが確認できた。このことは、環境学が如何に広域におよぶ学問領域を扱っているかという点と、環境学研究所の専門領域の多様性、つまり強みを有しているという点を示していると考ええる。

今回の紀要にて、課程在籍中にやり残したことを整理し、自身の解を示せたことは一定の成果と考える。同時に、一定期間ごとに、一定量の論文をテキストマイニング分析することで、過去から現在までの傾向や強み、または強化が求められる分野の特定などを戦略的に考える一つの材料として十分活用できるのではないか、ということも感じた。

参考文献

- 1) 武蔵野大学学術機関リポジトリ「環境研究所紀要」,
https://mu.repo.nii.ac.jp/index.php?action=pages_view_main&active_action=repository_view_main_item_snippet&index_id=45&pn=1&count=20&order=17&lang=japanese&page_id=13&block_id=21 (2022年8月10日参照)
- 2) KH Coder, <http://khcoder.net/dl3.html> (2022年8月10日参照)
- 3) 統計ER「共起ネットワーク - KH Coderによるテキストマイニング」,
<https://toukeier.hatenablog.com/entry/2019/08/19/203408#%E5%85%B1%E8%B5%B7%E3%83%8D%E3%83%83%E3%83%88%E3%83%AF%E3%83%BC%E3%82%AF%E3%81%A8%E3%81%AF> (2022年8月10日参照)

- 4) 樋口耕一「言語研究の分野における KH Coder 活用の可能性」,
計量国語学 31 卷 1 号（2017 年 6 月） pp.36-45