J. Jpn. Soc. Soil Phys.
土壌の物理性
No. 150, p.85 ～ 91 (2022)



# みんなのミニレビューを分析してみると

溝口 勝 [1]・杉野弘明 [1]

## 1. はじめに

1959 年に創刊した土壌の物理性の 150 号発刊の記念に「みんなのミニレビュー」が企画された．今回の企画では土壌物理を志したきっかけや取り上げた研究テーマに興味を持った経緯等を 140 字程度 （ツイッターと同じ）で書くことが求められた．私（溝口, 2015）は国際土壌年特集で「私の土壌物理履歴書」を書いてしまっていたので今更ミニレビューを書くこともないと思っていたが，記念号に関与しないのも寂しいので最近話題のテキストマイニングの手法で「みんなのミニレビュー」を分析することにした．今回集まったレビューは必ずしも土壌物理学会員全員ではないものの 72 件のレビューを分析することで見えてきたことがある．本論文では分析の方法と結果について紹介しながら次の 50 号に向けた期待について述べたい．

## 2. 分析手法

まず 2022 年 1 月までに「土壌の物理性」に投稿されたミニレビュー 72 件の電子データを分析可能な形にするため，いくつかの事前処理を施した．すなわち，1）タイトル，著者名，所属，キーワード，本文，引用文献の各セクションへの分解，2）所属について機構名以降の細かい所属を切ったデータの準備，3）データ全体における重複の無いキーワードのリスト化，4）プレーンテキスト化に伴い生じた本文データの修正したものを分析データとした．

本研究で扱うテキストデータの分析方法としては KJ 法（川喜多, 2006）や GTA（Glaser and Strauss, 1967）といった質的アプローチが広く知られている．しかし，手作業での分類はデータ量の多さに比例して効率が低下するため，多量テキストデータに対しては量的アプローチも望まれる．近年ではコンピュータの処理能力の向上と分析に用いるためのソフトウェアの開発・発展に伴い，分析の客観性保持と恣意性排除が叶う定量的な分析手法としてテキストマイニングが注目されている．テキストマイニングとは，テキストと呼ばれる言語データを対象とするデータマイニングの理論および技術の総称であり（小林, 2017），心理学や医療，マーケティングなど幅広い分野において研究が蓄積されてきている（例えば、木村ら, 2005; 磯島, 2006）．本研究では全 72 件，本文内容は 61,026 語が構成する大量の言語データであるため，テキストマイングが適していると考えられる．

本研究では，「内容」，「タイトル」および「キーワード」を分析対象とした．「内容」データは頻度が 25 以上の頻出語を，「タイトル」データは頻度が 2 以上の頻出語を，「キーワード」データは重複の無い 254 語を利用した．

まず「内容」データを対象に頻出語同士の共起度合いを隣接行列（Table 1）により表現し，それを基にグラフ理論により算出した共起ネットワークを用いて，単語のグルーピングを行うことで，頻出単語の共起の塊を視覚化し，トピック構造を捉えることを試みた．また，同様の分析を「タイトル」データを対象に実施した．

次に，「内容」データの頻出語を利用した文書（各レビュー）× 頻出語の頻度行列（須賀ら, 1993; 大井, 1994）を作成し，各レビューに紐づいた「所属」のデータを結

**Table 1**　単語同士の隣接行列の例 （表中の数字は共起数）.

| | 土壌 | 研究 | 土 | 行う | 考える | 水 |
|---|---|---|---|---|---|---|
| 土壌 | – | 124 | 50 | 27 | 85 | 139 |
| 研究 | 124 | – | 67 | 24 | 38 | 98 |
| 土 | 50 | 67 | – | 13 | 22 | 51 |
| 行う | 27 | 24 | 13 | – | 77 | 85 |
| 考える | 85 | 38 | 22 | 77 | – | 34 |
| 水 | 139 | 98 | 51 | 85 | 34 | – |

**Table 2**　文書 × 頻出語の頻度行列の例.

| ID | 所属 | 土壌 | 研究 | 土 | 行う | 水 |
|---|---|---|---|---|---|---|
| 1 | 新潟大学 | 2 | 3 | 2 | 0 | 1 |
| 2 | 農研機構 | 4 | 6 | 2 | 3 | 3 |
| 3 | 筑波大学 | 0 | 0 | 1 | 0 | 5 |
| 4 | 宇都宮大学 | 1 | 0 | 1 | 1 | 0 |
| 5 | 岩手大学 | 3 | 2 | 0 | 0 | 3 |
| 6 | 北海道大学 | 5 | 1 | 0 | 0 | 2 |

[1]Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, JAPAN. Corresponding Author: 溝口 勝, 東京大学大学院農学生命科学研究科.

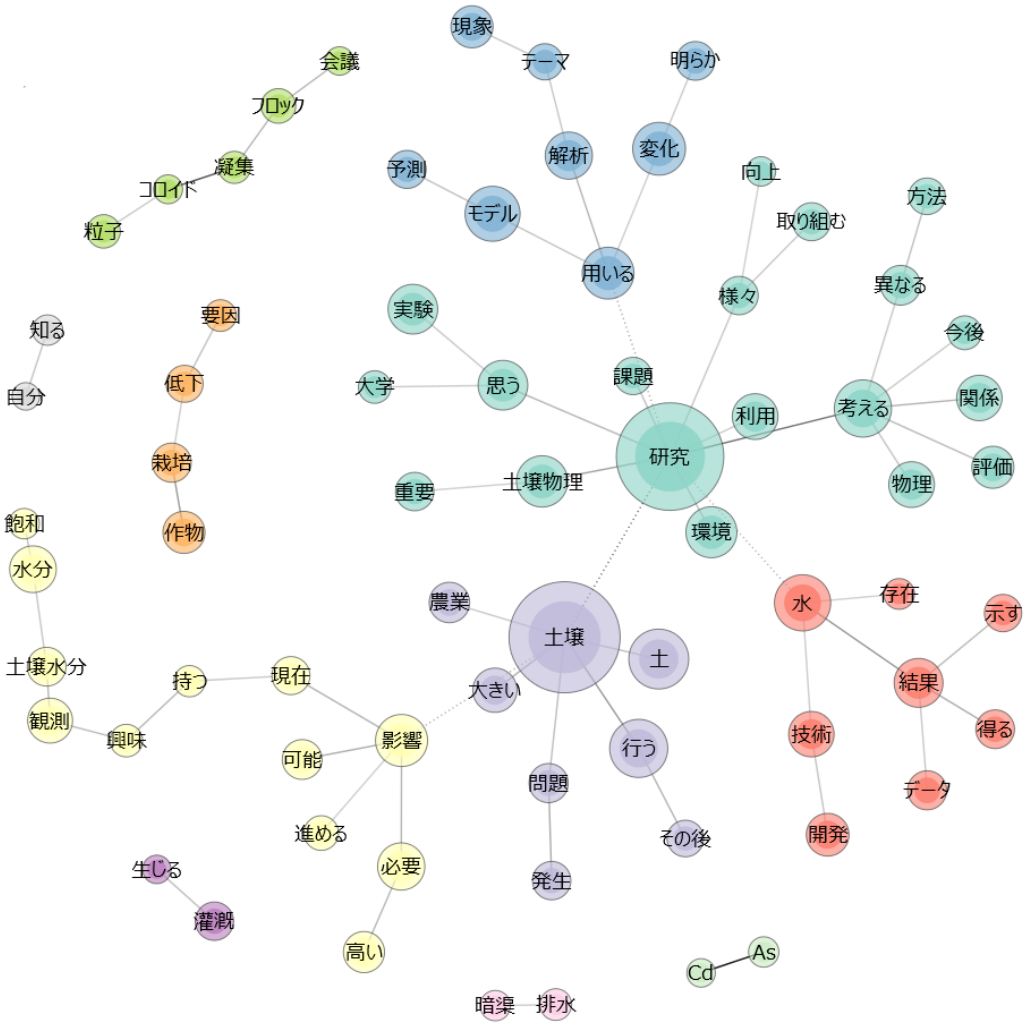



**Fig. 1**　レビュー内容における頻出語の共起ネットワーク．

合させた上で（Table 2），所属と頻出単語の 2 種類のノードから成る 2 部グラフを描画した．2 部グラフとは，グラフ内の集合を 2 つに分割し，全ての辺が各集合に所属するノードを結ぶようにして作成されるグラフ（ネットワーク）である（Krebs, 2002）．これにより，キーワードを介した所属組織同士の繋がりを見ることができる．

最後に，上記で使用した「内容」データの文書 × 頻出語の頻度行列に，タイトルとキーワードの頻度行列を結合させたものを利用し，クラスター分析を行うことによってレビュー同士の類似度を算出し，可視化した．分析を実施する際には，各データに含まれる共通単語が占める割合から算出する Jaccard 係数を距離として用い，また結合法にはグループ内の分散に対するグループ間の分散を最大化する Ward 法（金, 2007）を利用した．

なお，分析には，統計解析ソフトウェアである R Ver. 4.1.0 を使用し，TM が可能なパッケージ「RMeCab」とネットワーク分析用パッケージ「igraph」を主に用いた．

## 3.　分析結果

レビュー内容のデータから作成した頻出語の上位 150 語を Table 3 に示す．「土壌」や「研究」，「土」，「水」といったどのレビュー文章にも入る単語が上位に入った上で，「モデル」や「実験」，「水田」，「土壌水分」など多く扱われたテーマに関連する単語が続く形となった．また，レビュー内容の頻出語を使った隣接行列から作成した共起ネットワークを Fig. 1 に示す．描画されたネットワーク中，複数の頻出語によって形成された下位グループは 11 個見られた．「研究」という単語を中心に形成された一番大きなグループには，「土壌物理」，「実験」，「物理」，「評価」，「課題」などの単語が含まれた．「研究」を中心としたグループに繋がる形で，「水」，「技術」，「開発」，「結果」，「データ」といった単語群で形成されたグループが示された．また再頻出語の「土壌」を中心に作られたグループでは，「農業」や「問題」といった単語が

**Table 3** レビュー内容の頻出上位 150 語.

| 抽出語 | 出現回数 | 抽出語 | 出現回数 | 抽出語 | 出現回数 |
|---|---|---|---|---|---|
| 土壌 | 464 | 特性 | 52 | 地域 | 36 |
| 研究 | 434 | 明らか | 52 | 地表 | 36 |
| 土 | 130 | 重要 | 51 | 有機物 | 36 |
| 行う | 122 | 条件 | 51 | 解明 | 35 |
| 考える | 118 | 畑 | 51 | 凝集 | 35 |
| 水 | 114 | 発生 | 51 | 透水性 | 35 |
| モデル | 109 | 様々 | 51 | 理解 | 35 |
| 変化 | 100 | 灌漑 | 51 | 稲 | 34 |
| 影響 | 97 | 現在 | 50 | 地下 | 34 |
| 用いる | 96 | 取り組む | 50 | 排水 | 34 |
| 土壌物理 | 94 | 予測 | 50 | 与える | 34 |
| 環境 | 93 | 多く | 49 | 要因 | 34 |
| 窒素 | 92 | 課題 | 48 | 持つ | 33 |
| 思う | 88 | 効果 | 48 | 受ける | 33 |
| 実験 | 88 | 示す | 48 | 深い | 33 |
| 測定 | 88 | 過程 | 47 | 分布 | 33 |
| 根 | 87 | 推定 | 47 | 実施 | 32 |
| 結果 | 85 | 低下 | 47 | 場合 | 32 |
| 解析 | 79 | 日本 | 47 | 状態 | 32 |
| 必要 | 79 | 伴う | 47 | 植物 | 32 |
| 濃度 | 78 | その後 | 46 | 浸潤 | 32 |
| 水分 | 77 | 方法 | 46 | 存在 | 32 |
| 圃場 | 77 | テーマ | 45 | As | 31 |
| 物理 | 74 | 移動 | 44 | 暗渠 | 31 |
| 利用 | 74 | 水管理 | 44 | 知る | 31 |
| 関係 | 73 | 生育 | 44 | 飽和 | 31 |
| 技術 | 73 | 強度 | 42 | 凍土 | 30 |
| 観測 | 71 | 現場 | 42 | システム | 29 |
| 水田 | 71 | フロック | 41 | 感じる | 29 |
| 論文 | 71 | 進める | 41 | 挙動 | 29 |
| 大きい | 69 | 速度 | 41 | 向上 | 29 |
| 調査 | 66 | 多い | 41 | コロイド | 28 |
| 開発 | 64 | 化学 | 40 | 乾燥 | 28 |
| 地下水 | 63 | 間隙 | 40 | 機能 | 28 |
| 現象 | 62 | 手法 | 40 | 吸着 | 28 |
| 作物 | 62 | 森林 | 40 | Cd | 27 |
| 評価 | 62 | 凍結 | 40 | パラメータ | 27 |
| 高い | 59 | 農地 | 40 | 会議 | 27 |
| 降雨 | 58 | 理論 | 40 | 改善 | 27 |
| 農業 | 57 | 今後 | 39 | 経験 | 27 |
| 構造 | 56 | 生産 | 39 | 使う | 27 |
| 試験 | 56 | 改良 | 38 | 生じる | 27 |
| 異なる | 55 | 管理 | 38 | 特に | 27 |
| 可能 | 55 | 対象 | 38 | 内容 | 27 |
| 先生 | 54 | 分野 | 38 | 流出 | 27 |
| データ | 53 | 粒子 | 38 | FB | 26 |
| 栽培 | 53 | 汚染 | 37 | 温度 | 26 |
| 問題 | 53 | 興味 | 36 | 及ぼす | 26 |
| 土壌水分 | 52 | 検討 | 36 | 求める | 26 |
| 得る | 52 | 大学 | 36 | 植生 | 26 |

**Table 4**　タイトルの頻出上位 30 語.

| 抽出語 | 出現回数 | 抽出語 | 出現回数 | 抽出語 | 出現回数 |
|---|---|---|---|---|---|
| 土壌 | 14 | 関係 | 3 | 向ける | 2 |
| 研究 | 7 | 技術 | 3 | 構造 | 2 |
| 土壌物理 | 7 | 水分移動 | 3 | 出会う | 2 |
| 畑 | 5 | 窒素 | 3 | 消費 | 2 |
| 現象 | 4 | 土 | 3 | 振り返る | 2 |
| 森林 | 4 | 物質循環 | 3 | 水移動 | 2 |
| 水田 | 4 | 予測 | 3 | 水管理 | 2 |
| 測定 | 4 | 溶質移動 | 3 | 水分 | 2 |
| 土中 | 4 | 汚染 | 2 | 制御 | 2 |
| 環境 | 3 | 活動 | 2 | 土壌水分 | 2 |

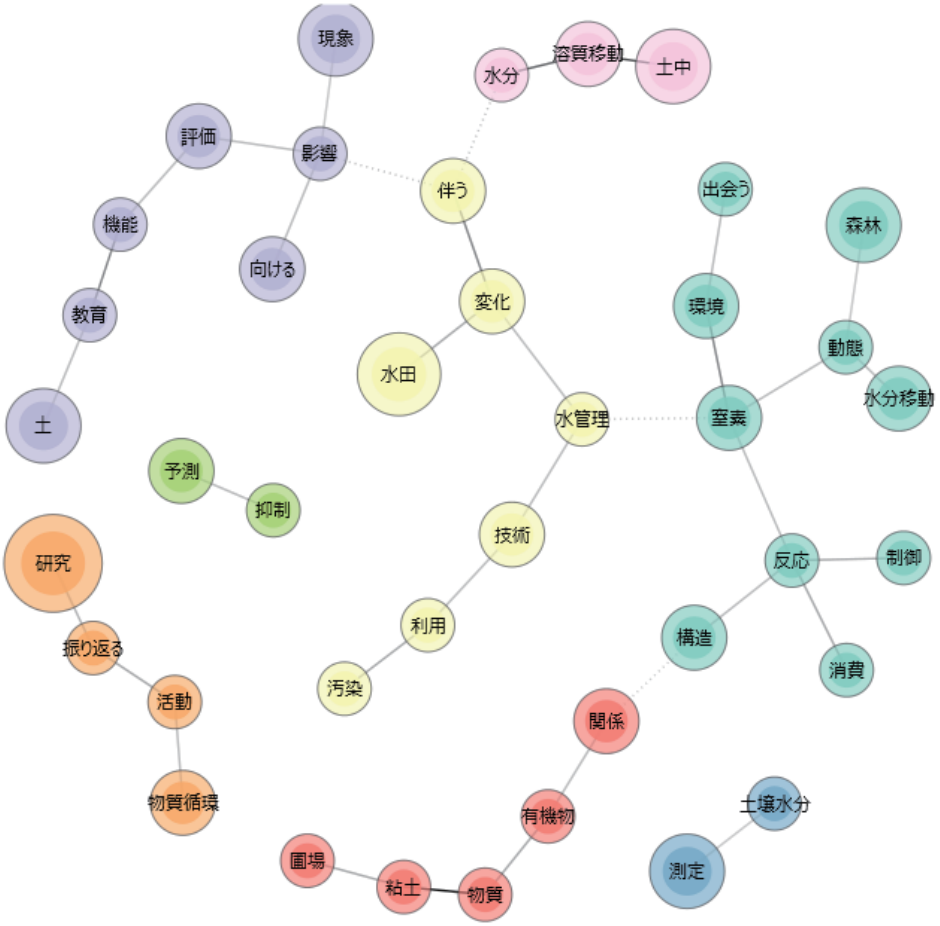



**Fig. 2**　タイトルにおける頻出語の共起ネットワーク.

含まれた．また，「水分」，「飽和」，「観測」，「影響」などといった単語が含まれたグループも「土壌」を中心としたグループからの繋がりを持つ形で示された．他にも「暗渠」と「排水」の 2 語で形成されたグループや，「フロック」，「凝集」，「コロイド」といった語で形成されたグループ，そして「栽培」，「作物」，「低下」，「要因」などの語で形成されたグループも存在した．

同様に，タイトルのデータにおける頻出語の上位 30 語を Table 4 に示す．表を見ると，レビュー内容のデータと同様に「土壌」や「研究」，「土壌物理」といった単語が頻出しており，「畑」や「森林」，「水田」，「土中」といった研究対象を示す単語が続く結果が見て取れる．また，「向ける」（「○○に向けて」等），「出会う」（「○○との出会い」等），「振り返る」（「○○を振り返る」等）といった単語もタイトルのデータを特徴づけるものであった．タイトルの頻出語を使った共起ネットワークを Fig. 2 に示す．ネットワークには 7 個の下位グループが示された．一番大きい下位グループ（図中右）には「窒素」，

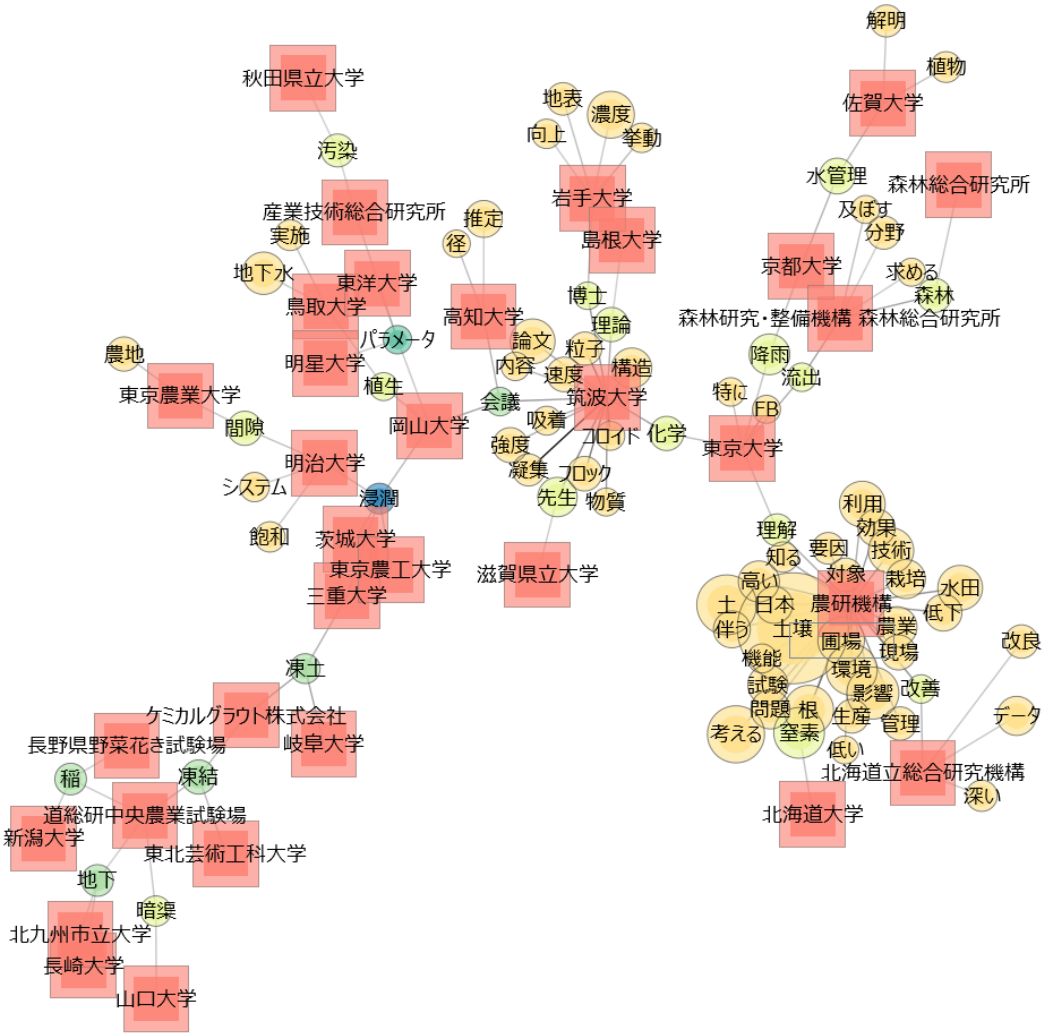



**Fig. 3**　レビュー内容における頻出語と所属の 2 部グラフ．

「動態」，「水分移動」，「環境」，「反応」といった単語が含まれた．「窒素」を軸として「水管理」，「技術」，「変化」といった語で構成されたグループが繋がっており，さらに「影響」や「評価」，「現象」，「機能」で構成されたグループへと接続していた．また，周辺の下位グループとして「有機物」，「物質」，「粘土」が構成するもの，「水分」，「溶質移動」，「土中」が構成するもの等が示された．

次に，レビュー内容の頻出語を所属のデータと結合させた頻度行列から，所属と頻出単語の 2 種類のノードから成る 2 部グラフを描画した（Fig. 3）．図中右下を見ると，「農研機構」を中心とし，「土壌」，「水田」，「圃場」，「環境」，「技術」などのキーワードを含んだ大きなグループが形成されていることが分かる．また，「農研機構」に繋がった単語群に対して「北海道大学」と「北海道立総合研究機構」が繋がっていた．この大きなグループに対して，「理解」という単語で「東京大学」に繋がっており，そこから「降雨」や「流出」で「京都大学」や「森林研究・整備機構森林総合研究所」，「佐賀大学」に繋がり，周辺には「水管理」や「森林」，「植物」といった単語が付置された．また，図中中央には「筑波大学」を中心としたグループが形成されており，「博士」や「理論」といった単語を介して「岩手大学」や「島根大学」に対して繋がり，その周辺には「地表」や「濃度」，「挙動」といった単語が繋がっていた．また「岡山大学」，「東洋大学」，「高知大学」，「明治大学」などには「会議」や「パラメータ」，「植生」，「浸潤」といった単語を媒介してグループが広がっていた．「高知大学」の先には「推定」や「径」が，「東洋大学」の先には「汚染」や「秋田県立大学」が，「鳥取大学」の先に「地下水」や「実施」が，「間隙」や「農地」というキーワードの周りに「東京農業大学」，「浸潤」の先に「明治大学」や「茨城大学」，「東京農工大学」，「三重大学」が繋がっていた．また「凍土」がハブのキーワードとなって「三重大学」，「ケミカルグラウト株式会社」，「岐阜大学」と繋いでおり，その先には「凍結」や「稲」，「地下」，「暗渠」といったキーワードにより「長野県野菜花き試験場」，「道総研中央農業試験場」，「東北芸術工科大学」，「新潟大学」や「長崎大学」，「北九州市立大学」，「山口大学」を繋げたグループが存在した．

最後に，レビュー内容から作成した頻度行列を基に実施したクラスター分析の結果として，併合過程を樹形図

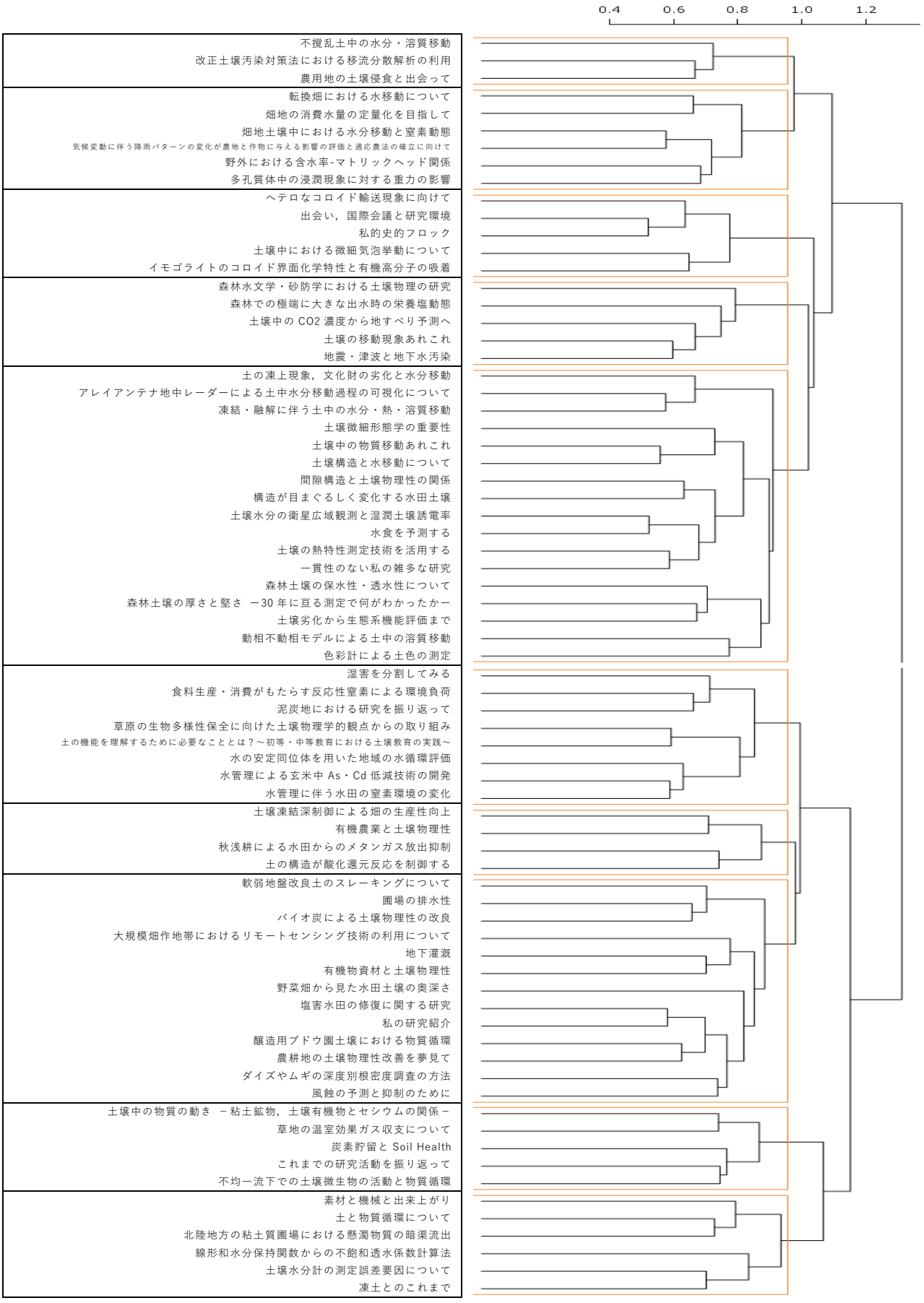



**Fig. 4** レビューのデンドログラム（クラスター分析）.

（デンドログラム）状に描画したものを Fig. 4 に示す．図中，タイトル同士の距離はそれらから延びている線が連結するまでの高さで示されており，短い方が似ていると解釈する（金, 2007）．全データは 10 個のクラスターに分けられ，一番大きなクラスターには 17 のレビューが，最小のクラスターには 3 つのレビューが所属する結果となった．最大のクラスターには，「土の凍上現象，文化財の劣化と水分移動」や「凍結 · 融解に伴う土中の水分 · 熱 · 溶質移動」，「土壌水分の衛星広域観測と湿潤土壌誘電率」，「森林土壌の厚さと堅さ — 30 年に亘る測定で何がわかったか —」といったタイトルのレビューが含まれた．また，次に大きなクラスターには，「大規模畑作地帯におけるリモートセンシング技術の利用について」，「圃場の排水性」，「地下灌漑」，「農耕地の土壌物理性改善を夢見て」といったレビューが含まれた．他にも，「湿害を分割してみる」，「草原の生物多様性保全に向けた土壌物理学的観点からの取り組み」，「水の安定同位体を用いた地域の水循環評価」，「水管理に伴う水田の窒素環境の変化」といったタイトルのレビューが形成するクラスターや，「有機農業と土壌物理性」，「秋浅耕による水田からのメタンガス放出抑制」，「土の構造が酸化還元反応を制御する」といったレビューのクラスターが確認された．

## 4. 考察

レビュー内容を対象とした頻出語の分析結果を概観すると，ミニレビューは大きく 1）土壌物理にまつわる諸課題を紹介したもの，2）各課題に対して詳しい分析手法などを紹介したもの，3）各テーマの歴史的背景を探ったもの，4）諸テーマに関する現状紹介と将来的課題を示したもの，に分類することができる．また扱われる課題やテーマに関しては，灌漑や暗渠排水，作物栽培，土壌中の水分移動，乾燥問題など多岐に渡っており，関連するキーワードが頻出語として挙げられている．加えて，タイトル中に含まれる頻出語の分析結果を見ると，物質循環や水管理，溶質移動や粘土，窒素，動態といったより具体的なキーワードを確認することができる．

2 部グラフの結果からは，農研機構や筑波大学のミニレビュー執筆における中心的な役割がネットワーク上にも表れていると言える．また，キーワードを介した所属機関の繋がりは，表面上はレビュワー同士の繋がりを意味しないが，例えば三重大学を中心として広がる凍土に関する研究など，研究テーマの系譜や機関間における伝播の足跡を知ることができる．

クラスター分析の結果では，頻度行列から自動的に分類を行っていることから必ずしも綺麗な分類となっていないところもあるが，概ね下記のような分類を見出すことができる：①土壌汚染 · 土壌侵食，②畑地と作物，③コロイドやフロック，④森林での土壌物理と移動現象，⑤土壌と水，⑥湿害と水管理，⑦有機と反応，⑧研究紹介と将来展望，⑨土壌と物質循環，⑩土壌水分と計算．これらは必ずしも学会の論文や研究発表会におけるテーマ別と符合するものではないが，各レビュワーが自身の専門分野からあるテーマをその歴史から将来展望にかけて広く紹介している傾向が見て取れる．

## 5. おわりに

テキストマイニングはデータサイエンスの分野で定性的なビックデータを可視化する手法として使われ始めている．今回は 72 件のレビューをこの手法を使って分析してみた．72 件は全正会員の 35 % に過ぎずとてもビックデータとは言えないが，現役で活躍する土壌物理学会員の傾向をそれなりに可視化できたように思う．

土壌の物理性 50 号にはテーマを予め設定した「土壌物理研究レビュー」と座談会「土壌物理の歴史とその展望」が掲載されている．私（溝口）は 50 号が発行された 1 か月後に博士課程を中退して三重大学の助手になった．その時から 100 号分の歳月を経て 50 号で議論された展望がどれくらい開けたのだろうか？ 結局のところ土壌物理が扱う領域は何も変わっていないような気もする．

## 引用文献


Glaser, B. and Strauss, A. (1967): The discovery of grounded theory: strategies for qualitative research. Aldine Pub. Co., Chicago.

磯島昭代 (2006): テキストマイニングを用いた米に関する消費者アンケートの解析, 農業情報研究, 15(1): 49–60.

川喜田二郎 (2006): 発想法, 中央公論新社, 東京.

木村昌臣, 古川裕之, 塚本 均, 田崎久夫, 空閑正浩, 大倉典子, 土屋文人 (2005): 医薬品使用の安全性に関するアンケートの解析 テキストマイニング手法の適用, 人間工学, 41(5): 297–305.

金 明哲 (2007): R によるデータサイエンス. 森北出版, 東京

小林雄一郎 (2017): R によるやさしいテキストマイニング, オーム社, 東京.

Krebs, V. (2002): Uncloaking terrorist networks. First Monday, 7(4). https://doi.org/10.5210/fm.v7i4.941

溝口 勝 (2015): 土壌の物理性, 130: 35–37.

大井 紘（編）(1994): 自由記述法による生活環境に関する地域住民の意識の調査と分析, 国立環境研究所研究報告 第 132 号.

須賀伸介, 大井 紘, 原沢英夫 (1993): 自由連想調査とクラスター分析による水辺に対する住民意識の研究, 土木学会論文集 No. 458/IV-18: 91–100.