

自然言語処理と深層学習の分野は密接に関連しており、自然言語処理(図中緑色)を学ぶ際には、学習者が機械学習(図中緑色)や、確率統計(図中紫色)についての知識を有していると学びやすくなると考えられる。また広く分布しているオブジェクト指向(図中赤色)や情報通信(図中黄色)、深層学習といった分野は横断的分野の知識を必要とする、または必要とされる内容のプログラミング分野であると考えられる。さらにプログラミング科目間の関連性を明らかにすることで学習計画の具体的な支援が可能である。

3.2 Doc2vec を用いた文書ベクトル化

Word2vec に続き、文書全体の意味を捉える Doc2vec 技術を用いて、シラバスの講義内容をベクトル化しクラスタリングすることで、科目間の関連性を明瞭にした。このアプローチにより、似たテーマの講義は空間上で近接し、異なるテーマは遠ざかるため、講義同士の関連性を視覚的に提示できる。しかし、同様の講義が多く挙げられたため、単純な類似講義の推薦としては限界があることも示された。これは、より多様で異なる関連講義を推薦する新たなアプローチが必要であることを示唆している。全講義の文書ベクトルを2次元に削減し、可視化した結果を図 2 に示す。“オブジェクト指向”を含む講義名を赤色のノードで、その類似講義は緑色のノードで表し、それらを緑のエッジでつないでいる。

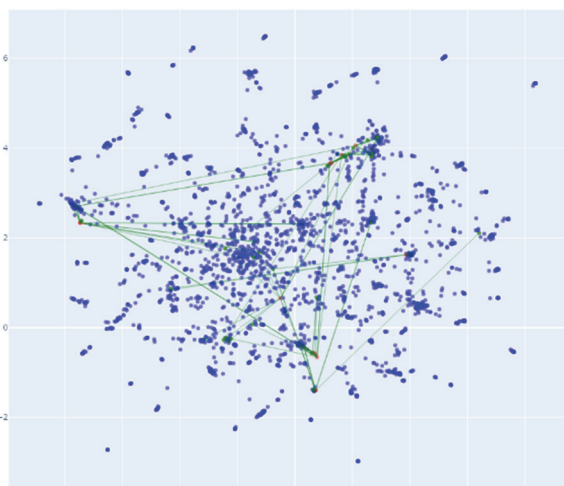


図 2 可視化された全講義の文書ベクトル

本研究で行われたシラバスデータの可視化により、収集されたデータの洞察が得られた。プロットされたノードには、一定の疎密が観察された。これは、講義内容やテーマの関連度に基づいて、講義が空間上でどのように配置されているかを示している。特に、プロットの中心部分には、いくつかの講義が密集して集合していることが観察された。これらの講義が類似したテーマや内容を持ち、教育プログラム内で核となるトピックを反映している可能性があり一方で空間上の外側に位置するノードは、より独特または特定のニッチなテーマを扱っている講義を表している可能性がある。内容を実際に確認するため、特定のワードを検索キーワードとしてそれらを含む内容を図 3 に色分けした。可視化データから、“オブジェクト指向”に関連する Java や C 言語、FORTRAN などの講義を含む集合を特定し、これらの言語がプログラミング教育の中核となっていることを示した。特に Java と C 言語のオブジェクト指向との関連性は、プログラミングとソフトウェア開発の基本概念への理解を深める重要なステップを強調している。この分析を通じて、学習計画支援システムは中核的な言語や概念を踏まえた

効果的な学習経路を学習者に提供する可能性を示している。

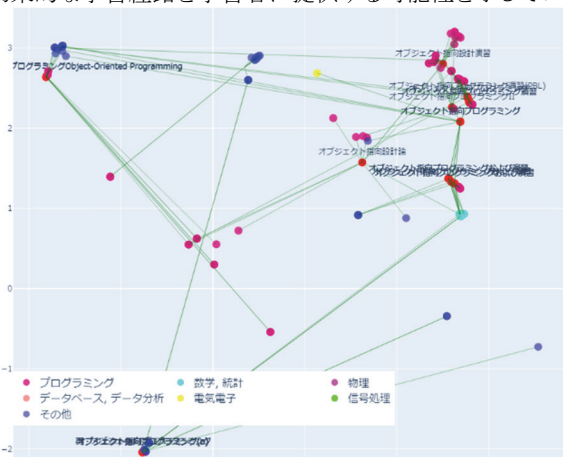


図 3 色分けした“オブジェクト指向”関連講義ベクトル学習計画支援システムは、これらの中核的な言語や概念に基づいた学習経路をさらに分かりやすい形で学習者に提案することにより効果的な学習体験を提供することが可能となる。

3.3 BERTopic を用いたクラスタリング

BERTopic は BERT と c-TF-IDF を用いた新たなトピックモデリング手法である。事前学習を用いた言語モデルを利用し、文書を埋め込みベクトルへ変換し、次元削減を行って最適化されたクラスタを作成する。その後クラスタベースの TF-IDF を用いて特徴語を抽出することでトピック表現を生成する[4]。講義シラバス用にカスタマイズされた BERTopic を用いてクラスタリングを含むトピックモデリングを行った。結果については発表にて述べる。

4. まとめと今後の課題

本研究では、プログラミング教育における講義間関連性の解析と論理的思考力の向上を目指す学習支援システムの開発に焦点をあて、分析を行った。Word2vec と Doc2vec を用いたコーパスの評価と BERTopic によるトピックモデリングを通じて、これに基づき教育プログラムと学習者のニーズに適した学習経路の提案を行った。今後の研究では、学習者が段階的にスキルを習得し次のトピックへ進むための明確な指針を提供することに注力する。学習者の能力レベルや興味に応じた個別化された学習経路を提案し、より効果的な学習体験を実現することが目標である。また、講義間関係の方向性付与は、学習者自身が自分の学習進捗を追跡し、次のステップを自律的に選択するための重要なツールとなる。これらの研究成果を活用して、個別化された学習計画を提案する学習支援システムの開発を目指す。

参考文献

- [1] 北村利夫, 土肥紳一, プログラミング学習計画補助システムに向けた用語関係可視化の試み, 情報処理学会第 85 回全国大会論文集(4), pp691-692, 2023.
- [2] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, Efficient Estimation of Word Representations in Vector Space, arXiv:1301.3781, 2013.
- [3] Wikipedia 財団, ウィキペディアフリー百科事典, <https://dumps.wikimedia.org/jawiki/latest/197-208/>, (参照 2022-10).
- [4] Maarten Grootendorst, BERTopic: Neural topic modeling with a class-based TF-IDF procedure, arXiv:2203.05794, 2022.