

情報処理能力分野のクラスタリングによるプログラミング学習 計画支援システムの提案

北村利夫[†] 土肥紳一[‡]

東京電機大学システムデザイン工学研究所[†] 東京電機大学システムデザイン工学部[‡]

1. はじめに

学習プロセスを重視した教育アプローチは、知識分野の理解深化や学習目標の明確化、さらには将来の学習設計への適用可能性といった観点から教育への有効性が認められている。本研究は、このアプローチをプログラミング教育に適用し、プログラミング教育上の課題解消を目指す。大学の講義を通じて学生が情報技術をどのように習得しているかを講義ごとのシラバスデータを用いて自然言語処理を用いて分析する。さらに、プログラミング学習における情報処理能力分野を同定し、その知識分野を自然言語単位でクラスタリングする。これにより、学習者は自身の学習進捗や興味に基づき、特定の情報処理分野に属する授業や教材を選択することが可能となる。本発表では、これらの手法による情報処理能力分野の同定とクラスタリングの手法について述べる。

2. データ分析

2023年3月に行われた情報処理学会全国大会の発表では、Word2Vecを用いてシラバス+補完語コーパスからプログラミング教育による専門的な言語関係を学習し、プログラミングに関連した特徴語の類似語を可視化した。プロットされた結果からは、各分野間の関連性が確認できた。図1は、特徴語群の外側を結び学習分野の一例を示している。情報通信工学(図中緑色)を学ぶ際には、学習者が演算処理装置(図中赤色)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。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。シラバスの講義内容に基づくクラスタリングを行い、

科目間の関連性を明瞭にする。

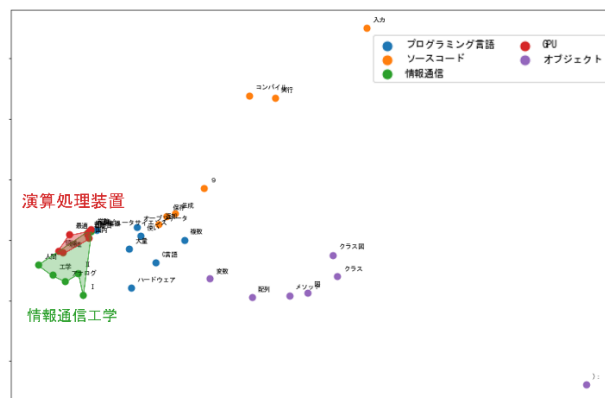


図1 プログラミング特徴語から識別される学習分野

2.1 対象データ

本研究は、教育現場における学習内容から、学習者がどのような順序で学習を進めているかを把握することを目的としている。そのため、各大学のシラバスデータを収集しており、前回の発表では東京電機大学千住キャンパスで提供されているプログラミング関連の講義を対象としたが、本研究ではその範囲を拡大し、他大学で提供されているプログラミング関連講義のシラバスデータも収集し、分析を行う。

2.2 情報処理能力分野

情報処理能力分野として様々な要因が考えられ、その背景には論理的思考力として、必要な素養が段階的に存在する。初等中等教育からプログラム教育が必修化され、盛んに教育が行われている英国のコンピューティング学習では、ナショナルカリキュラムをより詳細化した Computing Processing Pathways を採用している^[1]。アルゴリズム、プログラミングと開発、データとデータ構造、ハードウェアとソフトウェア、コミュニケーションとネットワーク、情報技術の6トピックに分け必要な技能を段階的に示しており、さらに論理的思考力として内容を抽象化、デコンポジション、アルゴリズム的思考、評価、一般化の5つの能力に分けている。 Computing Processing Pathways をフレームワ

Proposal for a Programming Learning Plan Support System
by Clustering of Information Processing^{†‡}

[†] Rio Kitamura, Graduate School of System Design
Engineering, Tokyo Denki University.[†]

[‡] Shinichi Dohi, School of System Design Engineering,
Tokyo Denki University.[‡]

ークに対応させ情報処理能力分野を考えることによって、基礎から高度な情報能力への学習経路が整理されること、また論理的思考力を基底として考えられているプロセスをもとに考えることで論理的思考力の醸成を促進することが望めることから、情報処理能力分野として採用した。

2.3 BERT

学習に利用する言語モデルとして、BERT を利用する。BERT は 2018 年に Google によって発表されたモデルで、他の言語モデルと大きく異なる点は事前学習とファインチューニングの 2 つから成る学習を行う点にある。従来自然言語処理には大量の文書データが必要となるが、大量の穴埋め問題を行う事前学習によってある程度言語の理解を行ったモデルを利用しファインチューニングを行うことで、比較的少量の文書データからも目的の学習を行うことができる。これらを含めた転移学習によって文脈を理解し高い性能を誇るモデルを利用し、シラバスデータをファインチューニングに利用し学習することで、プログラミング教育を理解したモデルを作成し、情報処理能力に照らし合わせた講義間の関連性を理解することができる。

2.4 学習手順

クラスタリングにおける一連の手順を図 2 に示す。

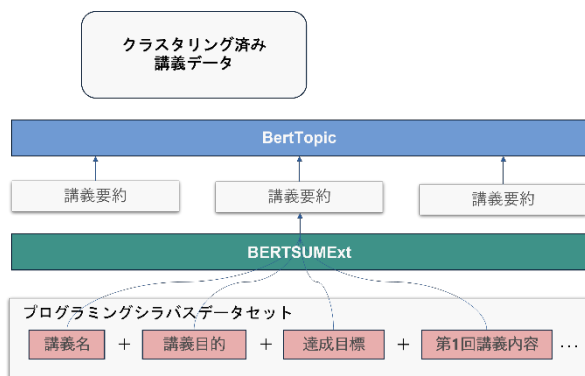


図 2 クラスタリング手順

初めに、講義シラバスデータを入力文書とし、それに対して講義文書の要約を行う。要約手法には抽出型要約と抽象型要約が存在する。抽出型要約は、入力文書から重要な内容を選択し、原意を保持する形で要約を作成する。一方、抽象型要約は、文書全体の内容を理解し、それに基づいて新たな文書を生成する。抽象型要約では出力文書が自然な形となる一方で、時として内容が損なわれることがある。対照的に、抽出型要約では出力文書が不自然な形となることがあるが、元の内容は保持

される。講義データの要約においては、自然な文書であることと同時に元の内容を保持することが重要であると考えられる。したがって、BERTSUMExt モデル^[2]を用いて元の内容を保持する抽出型要約を採用する。要約された内容を、講義間の関連性を抽出するためにクラスタリングし、情報処理能力分野に該当するデータセットとして追加準備する。クラスタリングには非階層型クラスタリング手法である k-means 法を利用する。BERT による埋め込み表現が従来の特徴ベクトル手法よりも優れている点を考慮し^[3]、BERT によってトークン化を行う。非階層型のクラスタリングと階層型のクラスタリングのどちらが高精度に学習できるかという点を考慮し、評価を行い必要に応じて非階層型クラスタリング手法を採用する。情報処理能力分野に応じた 6 つのクラスタにそれぞれ分類する。グループ化された内容がどの情報処理能力分野に当てはまるかを推論し、情報処理能力分野とプログラミング講義の関連性を明らかにする。

3. まとめと今後の課題

本発表ではプログラミング学習計画支援システムのためのモデル構築手法について提案した。本提案に基づき、BERT モデルを利用した講義内容の要約、クラスタリングおよび情報処理分野の同定を行う。今後まとめられた内容を用いて学習者の持つ知識から学ぶべき情報処理能力分野と、その分野に属する授業を提示する。また提示された授業を元に学習者のレベルや目的に応じて学習計画を個別に提案するシステムを構築する。さらに学習者の学習状況を追跡し、フィードバックを提供することで、学習者の学習状況に応じた学習計画の修正を行う。

参考文献

- [1] 太田 剛, 森本容介, 加藤 浩, 諸外国のプログラミング教育を含む情報教育カリキュラムに関する調査—英国, オーストラリア, 米国を中心として—日本教育工学会論文誌 40(3), 197-208, 2016.
- [2] Yang Liu, Fine-tune BERT for Extractive Summarization, arXiv:1903.10318, 2019.
- [3] 馬 ブン, 田中 裕隆, 曹 鋭, 白 静, 新納 浩幸, BERT を利用した単語用例のクラスタリング, 言語資源活用ワークショップ発表論文集, pp 343 - 350, 2019.