

科目順序推定のためのボックス埋め込み表現の活用 Utilization of Box Embedding Representation for Subject Sequence Estimation

橋本 菜々子^{*1}, 越智 洋司^{*2}

Nanako HASHIMOTO^{*1}, Youji OCHI^{*2}

^{*1} 近畿大学大学院総合理工学研究科

^{*1} Graduate School of Science and Engineering, Kindai University

^{*2} 近畿大学情報学部, 近畿大学情報学研究所

^{*2} Faculty of Informatics, Kindai University

^{*2} Cyber Informatics Research Institute, Kindai University

Email: 2233340409v@kindai.ac.jp

あらまし: 学生が未知の分野を学ぶ際に, 基礎から発展と段階を踏んで学習することが知識習得の近道となる。しかし, 初学者が学習順序を理解して適切に選択することは難しい。そこで本研究では, ボックス型の埋め込み表現を用いて科目順序を決定する手法について検討する。

キーワード: 情報推薦, 自然言語処理, 順序性, Word2Box, シラバス

1. 背景

近年, 機械学習や深層学習を用いたレコメンドシステムの開発がさかんに行われており, 教育の分野でも科目推薦などに活用されている。また, より良い科目推薦を行うためには, 学ぶ順序を考慮した推薦が必要不可欠であると考えた。本研究では学習順序という点に着目し, 科目間の関係性と順序性をボックス埋め込み表現を用いて可視化する手法について検討する。

2. 本研究の位置付けと目的

先行研究でもさまざまな手法で科目推薦システムを提案している。竹森ら⁽¹⁾は Doc2Vec を用いてシラバス文章を分散表現に変換し, ユーザーの好みとマッチする科目を推薦するシステムを実装した。また斉藤⁽²⁾は過去の学生の履修データとユーザの履修データをもとに協調フィルタリングによって科目を推薦するシステムを実装した。これらの研究では推薦科目をランキング形式で提案しており, 学習順序を考慮した科目推薦がされていない。また, 科目間の関連性について着目した益子ら⁽³⁾の研究では, 科目を「点」で可視化し関連性を明らかにするシステムを実装した。しかし履修支援を行うにあたって, なぜそれらの科目間に関連性があるのか, 根拠を明確にすることが必要であり, 関連性のみならず順序性を考慮することも必要であると考えた。

そこで本研究では, Word2Box⁽⁴⁾を用いて科目間の関係性を「箱」で可視化することにより順序性を決定する手法について提案する。

3. 順序性の決定

シラバス文書から科目の順序を決定するアプローチを示す。3.1 節ではボックス埋め込み手法について, 3.2 節は科目順序の決定手順について示す。

3.1 ボックス埋め込み手法

本研究では Word2Box を用いてボックス埋め込み

を生成する。Word2Box とは, Word2Vec のアルゴリズムを活用して単語を「ボックス」で表現する学習方法である。ボックス表現では, 体積が大きいほど単語の意味も大きくなるという性質をもつ。このモデルは単語に対する生成を行うため, 科目をボックス表現するために, 科目の中でも TF-IDF により上位 20 単語を特徴語として抽出し, それらの総称を「科目のボックス」とみなす (図 1)。

生成されたボックスは 128 次元と設定している。これは Word2Box の論文にてベクトルモデルと比較検証を行った際に設定されていた次元数であり, ベクトルモデルを上回る結果を得られたことから, 本研究でも次元数を 128 次元とした。しかし, 高次元データの重なりを人間が認識し, 順序を決定することは難しい。そのため UMAP⁽⁵⁾を用いて次元圧縮を行い可視化する。UMAP は高次元データの局所的な構造を保持しながら次元を圧縮することが可能なため, 科目内の情報を保持することができると考えた。

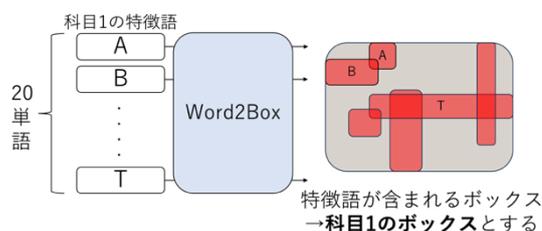


図 1 科目のボックスの定義

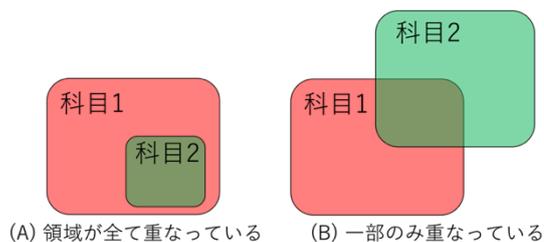


図 2 順序性の決定

3.2 可視化による科目順序の決定手順

科目順序はボックスの大きさと科目間重なりで決定する。図2の例を示す。(A)のように全て収まっている場合のみ順序が決定可能で、科目1→科目2となる。(B)の場合には、2つの科目間には関連性はあるものの順序性を決定づけるのは難しいため、科目内の単語にも着目する必要がある。

4. 実験

本実験では、近畿大学理工学部情報学科および電気電子工学科のシラバスを学習データとして用いている。本研究で活用したボックス埋め込み表現が科目順序の決定に対して有効であるかを検証する。図3に4科目を可視化し、科目ボックスの大きさ比較を表1に示す。その結果、順序性が明示的な科目である、「微分積分学 I」「微分積分学 II」、「線形代数学 I」「線形代数学 II」において応用科目が基礎科目内に含まれており、順序を反映した結果となっている。一方、順序性が明示的でない科目については、「コンピュータ概論」と「ネットワーク工学」では、順序を反映しているが、「プログラミング実習 I」と「エネルギー変換工学」では、重なりが生じている。

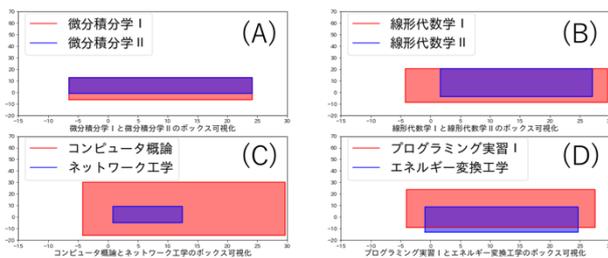


図3 科目ボックスの可視化例

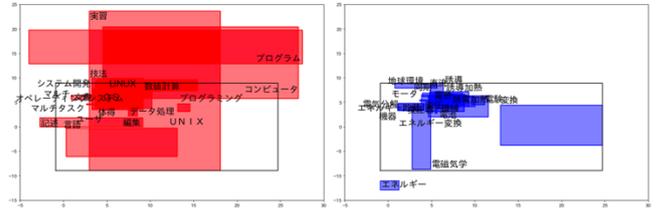
表1 ボックスの大きさ比較

科目名	領域
微分積分学 I	585
微分積分学 II	424
線形代数学 II	985
線形代数学 I	616
コンピュータ概論	1568
ネットワーク工学	163
プログラミング実習 I	1026
エネルギー変換工学	558

5. 考察

図3(A)(B)(C)の結果から、Word2Boxによるボックス埋め込みおよびUMAPによる可視化手法は、科目順序を決定するための手法として有用であることが示唆された。しかし、図3の(D)のように、科目の分野が異なる場合でも、ある程度の重なりがあったことから、科目内の特徴語単位での可視化を行い、その原因を考察した。図4では、特徴語単位で可視化し、2つの科目が重なり合っている領域を黒で囲んでいる。2つの科目の分野が異なっているにもかかわらず、

ならず、(B)エネルギー変換工学が、(A)プログラミング実習 I の領域にほとんど含まれてしまうという結果となった。本可視化では、これらの特徴語の領域を2次元に写像しているため、異なるはずの領域が重なって可視化されている可能性がある。



(A) プログラミング実習 I (B) エネルギー変換工学

図4 科目内の単語可視化

6. まとめと今後の課題

本研究では、科目推薦に順序性を考慮した手法について、Word2Boxを用いてボックス埋め込みを生成し、UMAPによって次元圧縮を行うことを提案した。その結果、これらの手法は順序性について妥当性があることが示唆された。

一方、順序性がない科目についても、領域の重なりを示す結果となった。そこで科目の領域を可視化する際に、一様なボックス表現ではなく、科目内の特徴語の分布を反映させたり、ボックス表現を2次元ではなく3次元に拡張するといった新たな可視化手法について検討することが今後の課題となる。

また、科目の順序性をもとにした学習支援についても検討する。例えば、科目領域の重なりとの関係とシラバスの情報を連携させ、科目の順序性から学習内容の順序性を可視化し、学生自身が知識構造を把握して、適切な学習プランを立てられるような学習支援システムの開発について検討する。

謝辞

本研究は JSPS 科研費 22K12304 の助成を受けた。

参考文献

- (1) 竹森汰智, 亀井清華: “履修支援のための Doc2Vec を用いた科目推薦システム”, 情報処理学会誌, Vol.12, No.4, pp.9-14 (2018)
- (2) 斉藤正武: “大学生の履修科目登録のためのリコメンドシステムの開発および有効性の検証”, 日本情報経営学会第 81 回全国大会, pp.61-64 (2021)
- (3) 益子英俊, 牧野光則: “科目間の連携を考慮したカリキュラムの可視化”, 電気情報通信学会, 第 19 回データ工学ワークショップ (2008)
- (4) Shib Sankar Dasgupta et al.: “Word2Box: Capturing Set-Theoretic Semantics of Words using Box Embeddings”, Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Vol1, pp.2263-2276 (2022)
- (5) Leland McInnes, John Healy, James Melville: “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”, arXiv: 1802.03426 (2018)