

シラバスと履修登録情報を用いた大学院科目間関係の可視化

新美 礼彦^{1,a)} 小澤 海樹¹

概要: 科目でなにを学ぶのか、その科目がどの科目と関係があるのかはシラバスに記述されているはずである。しかし、一見しただけでは科目間の関係を読み解くのは困難である。本研究では、2つの手法により科目間の関係を可視化する方法を提案する。1つめの手法は、シラバスで同じキーワードを用いている科目間の関係を可視化する手法である。2つめの手法は、履修登録情報から履修科目の共起頻度を求め、科目間の関係を可視化する手法である。提案手法を大学院のシラバス、履修登録データに適用し、結果を考察した。

Visualization of Relationships Between Graduate School Subjects Using Syllabus and Course Registration Information

Abstract: The syllabus should describe what student will learn in the subject and which subject the subject is related to. However, it is difficult to understand the relationships between subjects by the syllabus. In this research, we propose a method to visualize the relationship between subjects by two methods. The first method is to visualize the relationship between subjects that use the same keywords in the syllabus. The second method is to find the co-occurrence pairs of the courses taken from the course registration information and visualize the relationship between the courses. The proposed method was applied to the syllabus and course registration data of the graduate school, and the results were considered.

1. はじめに

多くの学生は、教育機関から指定された必修科目と自由に選択することのできる選択科目を組み合わせ、決まった期間内において学期中に履修する科目を決定し履修登録を行う。

学生にとって、自由に履修することのできる選択科目は、学術的な知識と興味を満たすことができる重要な学びの機会である。その内容は、学部や研究科、専門という枠を超え学生の知見をより広めるためのものがあれば、学生が必ず履修しなければならない必修科目の枠を超え学生の専門性の向上を助けるものなど、多岐にわたる。学生がこのような大学の授業を履修するにあたって、授業の大まかな内容や特性、分野などを把握しておくことは、授業内でのふるまいや評価に影響するとともに、選択科目であれば、学生の興味や嗜好に合致する科目を見つける手掛かりとなる可能性がある。

科目履修を科目選択の問題と捉えれば、履修登録では取

るべき科目を推薦する推薦システムと考えることもできる。履修科目推薦システムとして考えると、内容ベースフィルタリング、協調フィルタリングなどが考えられる。内容ベースフィルタリングと考えるとシラバスを利用するところが考えられる。協調ベースフィルタリングと考えると学生の履修登録情報を利用することが考えられる。

本研究では、履修登録支援として、シラバスからの科目間の関係の可視化と履修登録情報を用いた履修共起頻度の高い科目の提示を行うことを提案する。本論文では、大学院のカリキュラムを対象にしているが、提案手法を用いた学部シラバスの分析については小澤の卒業論文 [1] で取りあげている。

2. 科目履修時の情報について

科目履修時に、科目内容や他の人がどのように履修登録しているかは参考になる情報である。ここでは、シラバスの記述の難しさと履修登録情報の取り扱いの難しさについて議論する。

学生は科目の情報を手に入れるために、講義の内容や進め方が記載された授業計画書であるシラバスを参照することができる。図1は公立はこだて未来大学の大学院のシラ

¹ 公立はこだて未来大学
Future University Hakodate
^{a)} niimi@fun.ac.jp

パスの一部である。シラバスには様々な情報が書かれている。授業概要、キーワード、授業計画のような講義内容に関すること、授業・試験の形式（対面、オンライン、ハイブリッドなど）や教授言語（日本語、英語）などの授業スタイルに関する情報、事前・事後の学習や履修上の注意など受講に関する情報、教科書・参考書のような講義資料に関する情報、到達目標、成績の評価方法など成績に関する情報が記述されている。大学院のシラバスの場合、さらに英語でも同じ情報が書かれている。大学によっては前提科目や関連科目が書かれていることもあるが、低学年の基礎科目が書かれていることが多く、どの科目にもよく書かれる科目（数学やプログラミングの基礎科目など）が書かれる一方、高学年での専門科目が書かれることは少なく、専門科目間の関係性は分かりにくい。また、シラバスの科目についての文章は、その科目や分野の専門家である担当教員によって考えられるのが一般的である。しかし、専門家である科目の管理者が考えた文章には当然、専門的な用語や言い回し等が含まれている可能性がある。学術的な知識や判断力の乏しい教育課程にある学生がそのような情報を基に科目の内容を理解することは難しいと考えられ、学生にとって重要な学びの機会や得られる知識が失われる可能性がある。このように、シラバスの情報量は膨大になる傾向があり、学生が履修したい科目の情報を得るために、逐一シラバスに記載されている科目の内容に目を通し、科目選択の手がかりとすることは学生にとってかなりの負担となる可能性がある。

学生の履修登録情報をもとに選択科目の支援を行う方法を考える。履修登録情報をもとに、実際に履修された科目を分析し、一緒に履修されやすい科目を抽出することにより、学生の科目選択の支援が行える。しかし、学生の履修登録情報は個人情報であり、取り扱いに気をつける必要がある。本研究では、公立はこだて未来大学の大学院教務委員会の協力の下、2019年から2021年の匿名化した履修登録情報を分析対象とした。また、分析に当たっては必須科目と選択科目の関係も考慮する必要がある。必須科目は該当学生すべてが履修すべき科目であるので、選択科目にくらべて共起する科目が多くなる。後述する実験では、全科目を対象にした分析を行ったところ、必須科目同士の履修組み合わせが頻度上位を占めたため、必須科目を外した分析も行った。

3. 関連研究

この節では本研究を行うにあたって参考にした既存研究、また本研究の立ち位置について述べる。共起頻度の高いアイテムを推薦する協調フィルタリングに関しては、本研究では特別な方法を提案しないため、関連研究は割愛する。

3.1 iFUN の開発

Vallance ら [2] は、公立はこだて未来大学のカリキュラムをマッピングしたアプリケーション「iFUN」を開発した。この研究では、「カリキュラム・マッピング」によって俯瞰的な視点を生み出し、科目の管理者や教員がカリキュラムの長所・短所を把握できるよう改善することを目的としていた。iFUN の開発では特に、ユーザがカリキュラムにアクセスしやすいようにマップ上で様々な科目へのリンクを表示するデザイン面での工夫や、安全なサーバから通常は紙面で提供されるシラバスをデジタルデータで提供する技術面での工夫がこらされた。iFUN の主な機能としては、科目どうしをネットワークでつなぎキーワード検索を行うことができる「検索と地図」機能がある。加えて、直接的なユーザ側の機能ではないが、開発側やカリキュラム提供側がリアルタイムにシラバスデータを更新することも可能になり、ユーザは最新のコース情報にアクセスすることが可能になった。これらの工夫や機能は、科目の管理者やスタッフがより良い情報を得るための支援になり、授業の方式や内容、評価方法を調整することが可能になる。

しかし、大学の教員と学生の科目探索における視点は大きく異なり、学生であればカリキュラムやシラバスの説明に使われている単語や語句が分からない場合や、キーワード検索を行う知識の不足などが考えられる。そこで本研究では科目の管理者ではなく学生を対象にし、自然言語処理を用いることで科目の知識が不足している学生であっても、科目の内容を把握・理解できるよう支援を行う。

3.2 大学教育の改革前後でのシラバス分析

渡辺ら [3] は、学習成果に基づく授業設計の視点からある大学のシラバスの内容分析を行った。具体的にはシラバスの各項目について「文字数のカウント」「記述方法の確認」「頻出後の抽出」などを行っている。

「文字数のカウント」では、シラバス全体及び記載項目ごとに文字数を算出し、1科目あたりの平均値を比較した。その結果、2005年度と比べて2011年度ではその大学のシラバスの全ての項目において文字数が優位に増加していることが分かった。「記述方法の確認」では、2005年度と2011年度を比べて各回の授業の説明について記述がなされている科目が明らかに増加していることが分かった。また、授業の説明の項目だけではなく評価方法についても、記述がない科目が減り、さらには評価の比重まで細かく記されている科目もあった。「頻出語の抽出」では、2005年度と2011年度を比較し2011年度では、「教示」の意味が強い語ではなく学生主体であることを表す「考える」「学ぶ」という語の増加や、アクティブラーニングに関連のある「理解」「深める」などの語がシラバスの項目の「習得できる知識・能力や授業の目的・ねらい」と関連が強くなっていることから各科目の学生の主体性が向上していること

2021/19		STUDENT	
2021/02/15(日)09:36			
科目名 / Subject	データ科学特論(Advanced Topics in Data Science)	履修年次 / School Year	履修学期 / Semester
単位数 / Credits	2	履修科目 / Subject	特論
講義形態区分 / S-Categorized Subject	対面型(オンキャンパス) / 対面型(オンキャンパス)	履修科目 / Subject	特論
科目区分 / Category	工学系 / 工学系	履修科目 / Subject	特論
担当教員	田中 太郎	履修科目 / Subject	特論
授業形態	対面型(オンキャンパス) / 対面型(オンキャンパス)	履修科目 / Subject	特論
授業内容	本講義は、データベースシステムの理論、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論
授業目標	データベースシステムの大規模データベース構築、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論
授業計画	Part A 1. データモデル 2. Relational data model 3. SQL 4. Database design 5. Transaction processing 6. Database and programming 7. Massive data processing Part B 1. Maximum Likelihood Estimation(MLE) and Bayesian Inference 2. Model selection 3. Classification method 4. Unsupervised learning 5. Non-parametric method 6. Ensemble method	履修科目 / Subject	特論
評価方法	Part A 1. 最終レポート作成 2. 中間試験 3. 課題 4. 模試 5. レポート 6. テスト 7. プレゼンテーション	履修科目 / Subject	特論
事前準備	データベースシステムの基礎知識を事前に学ぶこと。データベースシステムの基礎知識を事前に学ぶこと。	履修科目 / Subject	特論
成績評価	本講義は、データベースシステムの理論、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論
履修上の注意	本講義は、データベースシステムの理論、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論

2021/19		STUDENT	
2021/02/15(日)09:36			
科目名 / Subject	データベース特論(Advanced Topics in Database Science)	履修年次 / School Year	履修学期 / Semester
単位数 / Credits	2	履修科目 / Subject	特論
講義形態区分 / S-Categorized Subject	対面型(オンキャンパス) / 対面型(オンキャンパス)	履修科目 / Subject	特論
科目区分 / Category	工学系 / 工学系	履修科目 / Subject	特論
担当教員	田中 太郎	履修科目 / Subject	特論
授業形態	対面型(オンキャンパス) / 対面型(オンキャンパス)	履修科目 / Subject	特論
授業内容	本講義は、データベースシステムの理論、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論
授業目標	データベースシステムの大規模データベース構築、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論
授業計画	Part A 1. Data model 2. Relational data model 3. SQL 4. Database design 5. Transaction processing 6. Database and programming 7. Massive data processing Part B 1. Maximum Likelihood Estimation(MLE) and Bayesian Inference 2. Model selection 3. Classification method 4. Unsupervised learning 5. Non-parametric method 6. Ensemble method	履修科目 / Subject	特論
評価方法	Part A 1. 最終レポート作成 2. 中間試験 3. 課題 4. 模試 5. レポート 6. テスト 7. プレゼンテーション	履修科目 / Subject	特論
事前準備	データベースシステムの基礎知識を事前に学ぶこと。データベースシステムの基礎知識を事前に学ぶこと。	履修科目 / Subject	特論
成績評価	本講義は、データベースシステムの理論、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論
履修上の注意	本講義は、データベースシステムの理論、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論

2021/19		STUDENT	
2021/02/15(日)09:36			
科目名 / Subject	データベース特論(Advanced Topics in Database Science)	履修年次 / School Year	履修学期 / Semester
単位数 / Credits	2	履修科目 / Subject	特論
講義形態区分 / S-Categorized Subject	対面型(オンキャンパス) / 対面型(オンキャンパス)	履修科目 / Subject	特論
科目区分 / Category	工学系 / 工学系	履修科目 / Subject	特論
担当教員	田中 太郎	履修科目 / Subject	特論
授業形態	対面型(オンキャンパス) / 対面型(オンキャンパス)	履修科目 / Subject	特論
授業内容	本講義は、データベースシステムの理論、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論
授業目標	データベースシステムの大規模データベース構築、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論
授業計画	Part A 1. Data model 2. Relational data model 3. SQL 4. Database design 5. Transaction processing 6. Database and programming 7. Massive data processing Part B 1. Maximum Likelihood Estimation(MLE) and Bayesian Inference 2. Model selection 3. Classification method 4. Unsupervised learning 5. Non-parametric method 6. Ensemble method	履修科目 / Subject	特論
評価方法	Part A 1. 最終レポート作成 2. 中間試験 3. 課題 4. 模試 5. レポート 6. テスト 7. プレゼンテーション	履修科目 / Subject	特論
事前準備	データベースシステムの基礎知識を事前に学ぶこと。データベースシステムの基礎知識を事前に学ぶこと。	履修科目 / Subject	特論
成績評価	本講義は、データベースシステムの理論、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論
履修上の注意	本講義は、データベースシステムの理論、モジュール設計について概論を、それらに基づき大規模データベースの設計について学ぶ。データベースから得られる情報性をもとに、様々な手法を用いて、様々な課題を解決する。色々と進んで、大規模データベースの構築、利用、発展の最新動向を学ぶ。	履修科目 / Subject	特論

図 1 シラバスの例
Fig. 1 Example of Syllabus

が示唆された。

これらは、背景でも述べた 2007 年(平成 19 年)に改正された大学教育の改革理念の影響を受けていると考えられ、学生に提示される科目の情報量は増加したが、それを把握・理解するのに伴うコストも同時に増加傾向にあると思われる。改正された大学教育の改革理念は現在も継続して掲げられており、これからのシラバスの情報量の減少やカリキュラムが単純化の方向に進むことはないと考えられる。

これらのシラバス分析を踏まえ、本研究では学生に提示される情報の増加に対応するため、その情報の理解を支援する方法を提案する。

3.3 専門分野の文書とキーワード

文書の探索や検索を行う際、文書に紐づけられているキーワードは大きな役割を果たす。キーワードの多くは、文書の著者が自ら付けたものであり、単語レベルでありながらも文書の内容を大まかに把握する重要な手掛かりとなる。

しかし、キーワードの付け方は文書の著者や掲載されるプラットフォームなどによって異なり、固有名詞や他の文書の内容に含まれないようなユニークな単語などで文書の特徴づけるキーワード、その文書がどのようなカテゴリに含まれるかを示す文書を分類するためのキーワード、複数の単語を掛け合わせることで意味を成すキーワード、それぞれが独立して文書の特徴となるキーワードなど、一貫しているわけではない。

このキーワードの曖昧さを軽減することができれば、文書の内容把握や探索に改善が望めると考えられる。

また文書のキーワードは、ほとんどの場合その文書の作

成者本人によって手動で付けられるが、自然言語処理を用いることで文書から動的にキーワードを生成する研究がしばしばなされている。

吉田ら [4] は、医療分野の文献を対象として類似文書の検索・キーワード抽出を行った。具体的には、まず医療分野の論文集の各文書から形態素解析により単語を抽出、さらに頻出回数の重みづけによってキーワードの抽出を行った。このキーワード抽出に関して、吉田らは形態素解析の段階において辞書に登録されていない語句に着目し、未知語を複合語化することで有用な語への変換を行った。また、分野に偏りのない語句は文書の特徴づけるキーワードとして成り立たないと仮定し、不要語の選定も行った。

専門分野の文書では、通常語彙に含まれない語句が重要な意味を持つことが多い。そのため、「どんな文書にも登場する語は文書の内容を表す語句としてふさわしくない」とされる。本研究においては、文書から語句の抽出を行うのではなく、既に文書の特徴として文書の作成者自身に提示されたキーワードから新たに語句を抽出することによって、そのキーワードの属する文書及びその科目を特徴付ける手掛かりとする。キーワードから抽出した関連語については既に、キーワードから抽出した語であるという関連をもつと考えられ、文書に登場する語の偏りなどは考慮しないが、少なくともキーワードの属する文書内に抽出した語句が含まれているかのフィルタリングを行うことで、科目との関連性を保証する。

3.4 自然言語処理と Word2vec

近年、人の話し言葉や書き言葉が持つ意味を解析する自然言語処理が盛んに行われ、キーワードを用いてインターネット上の情報探索を行う検索エンジンや言語を他の国

の言語に変換する翻訳ツールなどに活用されている。

自然言語処理は、画像認識や音声認識などの分野と異なり、連続的な物理的現象を処理するものではない。自然言語はあくまで人間が文字や音声に意味を付加したものであり、機械がその複雑な法則や使われ方を理解するためには、実際に使われている自然言語から特徴を抽出する方法が必要であった。それらは今では自然言語処理モデルとして体系化され、文書を機械的に大量かつ高速に処理することが可能になってきた。

Mikolov ら [5] が 2013 年に発表した自然言語処理モデルの一つである「Word2vec」は 2 層のニューラルネットワークを用いて学習させることで単語を分散表現として扱い、その意味を定義することができる。2 層のニューラルネットワークは、それぞれ skip-gram 法と CBOW (Continue Bag-of-Words) と呼ばれるモデルで構成される。skip-gram 法では、入力として中心となる語を与え、その周辺に出現する語を予測するというタスクの下に、予測した周辺語を出力し、出力と実際の文の周辺語とを比較することで単語に重みづけをするという教師あり学習を行う。CBOW では逆に、入力として周辺語を伴い、その中心語を予測し、skip-gram 法と同じく教師あり学習を行う。これらのモデルは、それまでの提案手法より計算の複雑さが排除され、高速に学習を行うことができるようになり、高次元のベクトルを用いて単語の意味を表現することができるようになった。

現在では、主に Python のオープンソースライブラリである「Gensim」[6] などにより、簡単な準備と操作で Word2vec を扱えるようになってきている。加えて、Word2vec の学習済みモデルは、オンライン上に数多く公開されている。学習済みモデルは学習させるデータ、適用するアルゴリズム (skip-gram 法や CBOW)、生成するベクトルの次元数などがそれぞれのモデルで異なり、用途を考慮し学習済みモデルを活用することで、より精度の高い自然言語処理を行うことができる。

よく Word2vec が用いられる事例として、ベクトルどうしのコサイン類似度やユークリッド距離を計算することで単語間や文書間の類似度を算出することや、主成分分析により次元を圧縮し、単語をグラフに出力することによる視覚的な分析、文章の内容から読み取れる喜怒哀楽などの感情を数値で出力する感情分析等がある。

3.5 シラバスの項目を用いた分析手法

吉崎ら [7] は、科目の講義内容の理解を支援するため、シラバスの項目にある「授業計画」に着目し、科目間でその類似度を算出・比較することで関連する科目の導出を行った。

具体的には、授業計画にある文章を形態素解析し、単一文書につき単語の集合を作成、これを入力とすることで、文書をベクトル化することのできる「Doc2Vec」[8] を用い

て授業計画の文章間の類似度を求めた。関連科目の導出の精度を検証するための実験としては、各科目のシラバスに記載されている関連科目と手法によって導出された科目が対応しているかの確認を佐々木ら [9] が提案した「期待適合最大化モデル」によって行った。

その結果、「シラバスの関連科目に対する、シラバスと手法で一致した科目の割合 (再現率)」、「手法によって導出できた科目に対する、シラバスと手法で一致した科目の割合 (適合率)」は次元数などにより変動はするが、共に安定して 3~4 割を示した。このことから、部分的ではあるが科目の特徴を授業計画の文章から抽出できたとと言える。

本研究では、科目の特徴として授業計画ではなく「キーワード」を用いる。授業計画は文章であったが、キーワードは単語レベルの情報であるため単語をベクトルに変換できる Word2vec を用いて、キーワードを説明するような語群を生成する。しかし、これだけでは科目の特徴ではなく、あくまでキーワードの特徴を抽出することになる。よって、吉崎らが示したように授業計画の文章には科目を特徴付ける情報が含まれていると考えられるため、本研究では「授業概要」及び「授業計画」の文章内を探索することで科目に関連のある語を収集する際のフィルタリングとして用いる。

4. 提案手法

4.1 シラバスからの科目間関係の可視化手法

本研究では科目理解の支援のために、科目のキーワードを説明し得るような語群を生成する手法を提案する。提案手法である語群の生成を行うために、(1) 関連語収集、(2) 関連語群生成、という 2 つの手順を行う。

(1) 関連語収集について、この手順はキーワードについての関連語群生成に用いるための語句を集める作業である。集める語句は無作為に選ばれるのではなく、キーワードとの関連があるものを、個数を効率よく増幅させながら抽出するため、Word2vec [5] の学習済みモデルからキーワードとのコサイン類似度が高い関連語を抽出した後、それを新たな入力として関連語を抽出、これを繰り返すことで関連語を収集する。

(2) 関連語群生成について、各繰り返しで得られた関連語から任意の個数を組み合わせ、Word2vec を用いて個別にベクトルへ変換、これらを合成したベクトルと生成元であるキーワードのベクトルとのコサイン類似度を、キーワードに対する説明性の指標と仮定し、最も類似度の高い組み合わせを関連語群の候補とする。最後に各繰り返しで生成された関連語群候補を比較し、最もコサイン類似度が高くなった関連語群の候補を、キーワードの意味を説明し得るような関連語群とする。(1) 関連語収集について、科目のキーワードを説明し得るような語群を生成するために、語群の要素となる科目のキーワードに関連する語句を

収集するが、関連語からまた新たに関連語を抽出する処理を繰り返すことで、科目のキーワードとの関連を失わずに効率よく関連語の個数を充実させることができると考えられる。

(2) 関連語群生成について、本研究では「特定の単語と語群のコサイン類似度が高ければ、その語群は単語の意味の説明や補完になる」と仮定し、自然言語処理モデルである Word2vec を用いてキーワードを説明するような関連語群を生成する。Word2vec は単語をベクトルとして扱うことができるため、単語どうしが加法性を有する。代表的な例として「王様 - 男 + 女 = 女王」という演算があるが、必ずしもこのように演算が成り立つ訳ではなく、その精度はモデルを学習させる文書の質や量に大きく依存する。しかしながら、大量の文書から単語を学習させることで、分散表現であっても大まかな意味を定義することが可能であり、それは演算においても大体的意味が成り立つことに繋がる。この性質を利用し、本研究では「キーワード = 関連語 + 関連語 + ...」という演算が成り立つように関連語群を生成する。

以下の手順を任意に n 回繰り返すことで、ある科目の単一のキーワードについて関連語群の要素の候補となる語句を関連語リストに格納し最大 n 個収集する。

手順 1.1 キーワードのリストから要素の一つ取り出す。

手順 1.2 Word2vec の学習済みモデルから、キーワードに関連するキーワード以外の語句をコサイン類似度の高い順に任意個数抽出、キーワードの関連語候補として関連語リストに加える。

手順 1.3 Wikipedia [10] の関連語候補に紐づけられているキーワードのページの本文と、シラバスの授業や講義の具体的な説明のテキスト中に関連語候補が含まれているかどうかの探索を行い、関連語候補に見当違いな語句が混じっていないか調べる。いずれかのテキスト内に語句が含まれていれば、キーワードに関連があるとみなし、そうでなければ関連語リストから外す。関連語リストの全ての語句を調べ終わった後、次の手順へと進む。

手順 1.4 関連語群の組み合わせを見つける作業（手順 2.1 以降）のために、収集した関連語のリストを記憶しておく。そして、キーワードのリストを関連語リストで上書きし、関連語リストを初期化、手順 1.1 に戻り抽出した関連語に関する語句を新たに集める。

次に、関連語収集の手順において得られた各繰り返しの回の関連語リストを用いて、各回ごとにキーワードについて最も類似度が高くなる関連語の組み合わせを生成する。

以下の手順を行うことで、ある科目の単一のキーワードについて最も類似度の高い関連語群を作成するが、 k は、その手順で何回目の繰り返しで得られた関連語リストに注目しているかを表し、初期値は 1 である。

手順 2.1 単一のキーワードについて最大 n 個ある関連語リストのうち k 回目に生成されたものを取り出す。

手順 2.2 取り出した関連語リストの生成元であるキーワードを取り出し、Word2vec の学習済みモデルを用いてキーワードのベクトルを取得する。次に、キーワードについて k 回目の繰り返しで生成された関連語リストから任意に $1 \sim m$ 個までの語句の組み合わせを求めたリストを作成する。作成したリストから関連語の組み合わせを取り出し、こちらも全ての語句のベクトルを取得する。これらを合成した関連語の組み合わせのベクトルとキーワードのベクトルとの類似度を求める。全ての組み合わせとキーワードとのベクトルの類似度を求め終わったら、最も類似度が高くなった関連語の組み合わせを関連語群候補とし、類似度の値と共にキーワードに関連付け記憶しておく。

手順 2.3 $k = n$ であれば手順を終了し、でなければ $k = k + 1$ として、手順 2.1 に戻る。

これらの手順を行うことである科目の単一のキーワードについての関連語群候補が最大 n 個生成されているが、この中で最も類似度の高い組み合わせのものを、提案手法で求める科目のキーワードを説明するような関連語群とする。

4.2 履修登録情報からの共起頻度が高い科目群の抽出

履修登録情報から一緒に履修されている科目の組み合わせをカウントし、高頻度の科目の組み合わせを抽出する。

例えば、学生 a さん、b さん、c さんが次のように科目を履修したとする時、高頻度の科目の組み合わせとして A と B の組み合わせが 3 回、A と B と E の組み合わせが 2 回出現する。そのため、抽出結果として $\{ 'A', 'B' \}, 3, \{ 'A', 'B', 'E' \}, 2$ を得る。

- a さん: A B C E
- b さん: A B D E
- c さん: A B F G

5. 実験

シラバスからの科目間関係の可視化手法と履修登録情報からの共起頻度が高い科目群の抽出に分けて、実験と結果を示す。

5.1 シラバスからの科目間関係の可視化手法

提案手法を用いて公立はこだて未来大学の大学院のシラバスの科目間関係の可視化を行った。

本研究で用いた Word2vec の学習済みモデルのデータセットについて、日本語用と英語用の二つを用意した。日本語の学習済みモデルには、鈴木 [11] が作成した「日本語 Wikipedia エンティティベクトル」で、2019 年に公開された最新のものを採用し、英語の学習済みモデルには、Studio Ousia が開発・保守している「Wikipedia2Vec」[12] の英語

版の学習済みモデルを採用した。これらは共に Wikipedia 本文のテキストを学習データとしてニューラルネットワークを用いた教師あり学習によって単語を分散表現にしたものであり、各ベクトルの次元数は 300 次元となっている。また、それぞれの学習済みモデルで単語数は異なり、日本語用は 760,421 件、英語用は 4,530,030 件の単語及び分散表現が収録されている。

結果を図 2 に示す。科目間の関係を可視化することはできたが、キーワードが専門的すぎる場合は語群を作成できない場合が多い結果となった。

5.2 履修登録情報からの共起頻度が高い科目群の抽出

実験では、公立ほこだて未来大学の大学院教務委員会から提供された、2019 年から 2021 年の大学院の履修登録情報を用いた。履修データを前期後期に分けて、各学生ごとにまとめて履修されている科目の組み合わせをカウントした。実験結果の一部を図 3, 4 に示す。図 3 は前期科目、図 4 は後期科目の結果である。科目名のあとの数字は共起頻度を表している。システム情報科学におけるアカデミックリテラシー I、課題研究 I,II,III、システム情報科学研究は必須科目である。

共起科目の内、必須と思われる科目に関する組み合わせが高頻度で抽出されたため、必須と思われる科目を除いて実験を行った。実験結果の一部を図 5, 6 に示す。図 5 は前期科目、図 6 は後期科目の結果である。

シラバスから可視化した科目関係とくらべても概ね妥当な関係性が抽出できた。シラバスの可視化からは分からなかったインターンシップ I のほかの科目との関係も捉えることができた。

6. おわりに

本研究では、科目履修を科目選択の問題と捉え、履修登録では取るべき科目を推薦する推薦システムと考え、シラバスのキーワードを元にした科目間の関係の可視化と履修登録情報を用いた履修共起頻度の高い科目の提示を行うことを提案した。提案手法を公立ほこだて未来大学大学院のシラバス、履修登録情報に適用し、提案手法の有効性を検証した。検証結果から、科目間の可視化と履修共起頻度の高い科目を提示することができた。

参考文献

- [1] 小澤 海樹. “シラバスのキーワードに着目して Word2vec を用いることによる科目理解の支援”, 公立ほこだて未来大学 卒業論文, 2022.
- [2] Michael Vallance, Hiroaki Murayama, David Sperling. “A multi-disciplinary approach developing a mobile Curriculum Mapping App called iFUN”. IEEE Transactions on Learning Technologies. 2014, 6, 1, pp. 10–13.
- [3] 渡辺雄貴, 大森不二雄, 永井正洋. “学習成果に基づく授業設計の視点から見たシラバスの内容分析”. 大学基準協

- 会大学評価・研究部. 2014, 13, pp. 113–122.
- [4] 吉田鑑地, 中村明, 川尻光博, 松本忠博, 池田尚志. “専門分野の文書を対象としたキーワード抽出・類似文書検索と評価実験”. 言語処理学会第 12 回年次大会. 2006, 12, pp. 660–663.
- [5] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean. “Efficient Estimation of Word Representations in Vector Space”. 2013, pp. 1–12.
- [6] Radim Rehurek. “Gensim: Topic modelling for humans”. Gensim. <https://radimrehurek.com/gensim/>, (参照 2022-08-06).
- [7] 吉崎辰悟, 越智洋司. “関連科目を検索するためのシラバスデータ分析手法の検討”. Web インテリジェンスとインタラクション研究会予稿集 2. 2017, 11, pp. 29–30.
- [8] Quoc Le, Tomas Mikolov. “Distributed Representations of Sentences and Documents”. 2014, pp. 1–9.
- [9] 佐々木健太郎, 土田正明, 田村晃裕, NEC 情報・ナレッジ研究所. “期待再現率における期待適合率最大化モデルの学習方法”. 言語処理学会第 21 回年次大会. 2015, D3-3, pp. 481–484.
- [10] Larry Sanger, Jimmy Wales. “Wikipedia”. Wikipedia. https://en.wikipedia.org/wiki/Main_Page, (参照 2022-08-06).
- [11] 鈴木正敏. “Wikipedia Entity Vectors”. GitHub. <https://github.com/singleton/WikiEntVec/releases>, (参照 2022-08-06).
- [12] Studio Ousia, “Pretrained Embeddings”, Wikipedia2Vec, <https://wikipedia2vec.github.io/wikipedia2vec/#pretrained-embeddings>, (参照 2022-08-06).

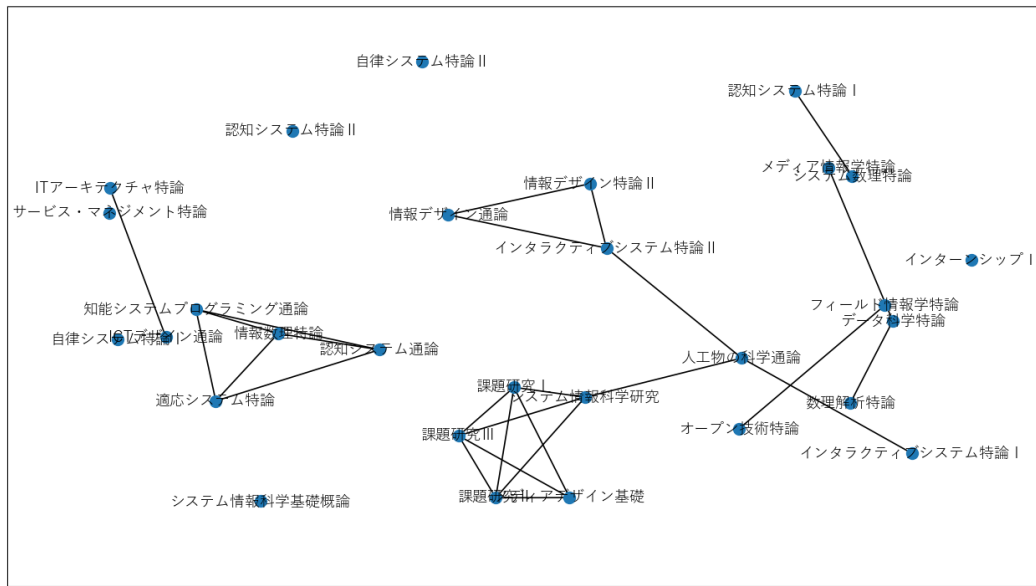


図 2 シラバスからの科目関係の可視化

Fig. 2 Visualization of Subject Relationships from the Syllabus

{ 'システム情報科学におけるアカデミックリテラシー I', 'システム情報科学研究', '課題研究 I' }, 2810
 { '課題研究 III', 'システム情報科学研究' }, 1046
 { 'システム情報科学におけるアカデミックリテラシー I', '情報環境学特論', 'システム情報科学研究', '課題研究 I' }, 1010
 { '情報環境学特論', 'システム情報科学研究' }, 858
 { 'インターンシップ I', 'システム情報科学におけるアカデミックリテラシー I', 'システム情報科学研究', '課題研究 I' }, 844
 { 'システム情報科学におけるアカデミックリテラシー I', '知能情報科学通論', 'システム情報科学研究', '課題研究 I' }, 593
 { '知能情報科学通論', 'システム情報科学研究' }, 429
 { 'システム情報科学におけるアカデミックリテラシー I', 'データ科学特論', 'システム情報科学研究', '課題研究 I' }, 390
 { 'システム情報科学研究', '情報ネットワーク特論 II' }, 336
 { '知能情報科学通論', 'システム情報科学研究', 'システム情報科学におけるアカデミックリテラシー I', '情報環境学特論', '課題研究 I' }, 332

図 3 必須科目を含む高頻度共起科目群 (前期)

Fig. 3 High-frequency Co-occurring Subjects Including Compulsory Subjects (1st Semester)

{ 'システム情報科学研究', '課題研究 II' }, 5753
 { 'システム情報科学研究', 'オープン技術特論', '課題研究 II' }, 711
 { 'システム情報科学研究', 'サービス・マネジメント特論', '課題研究 II' }, 540
 { '人工物の科学通論', 'システム情報科学研究', '課題研究 II' }, 508
 { '知能システムの歴史と未来', 'システム情報科学研究', '課題研究 II' }, 476
 { 'インターンシップ I', 'システム情報科学研究', '課題研究 II' }, 464
 { '知能システムの歴史と未来', 'システム情報科学研究' }, 404
 { '適応システム特論', 'システム情報科学研究', '課題研究 II' }, 387
 { 'メディア情報学特論', 'システム情報科学研究', '課題研究 II' }, 354
 { '適応システム特論', 'システム情報科学研究' }, 288

図 4 必須科目を含む高頻度共起科目群 (後期)

Fig. 4 High-frequency Co-occurring Subjects Including Compulsory Subjects (2nd Semester)

{ '知能情報科学通論', '情報環境学特論' }, 486
{ '情報環境学特論', '情報ネットワーク特論 II' }, 257
{ '情報環境学特論', 'ICT デザイン通論' }, 253
{ '情報環境学特論', 'インターンシップ I' }, 252
{ '知能システムプログラミング通論', '知能情報科学通論' }, 215
{ 'フィールド情報学特論', '情報ネットワーク特論 II' }, 203
{ '知能システムプログラミング通論', '情報環境学特論' }, 183
{ 'フィールド情報学特論', '情報環境学特論' }, 183
{ '情報ネットワーク特論 II', 'インターンシップ I' }, 180
{ '知能情報科学通論', '情報数理特論' }, 172

図 5 必須科目を除いた高頻度共起科目群 (前期)

Fig. 5 High-frequency Co-occurring Subjects Excluding Compulsory Subjects (1st Semester)

{ '知能システムの歴史と未来', '適応システム特論' }, 191
{ '知能システムの歴史と未来', '知能メディア特論' }, 140
{ 'オープン技術特論', '知能メディア特論' }, 107
{ '知能システムの歴史と未来', 'メディア情報学特論' }, 91
{ 'オープン技術特論', 'インターンシップ I' }, 86
{ 'オープン技術特論', 'サービス・マネジメント特論' }, 85
{ 'オープン技術特論', '組込システム特論' }, 79
{ 'インターンシップ I', '情報ネットワーク特論 I' }, 79
{ 'インターンシップ I', '人工物の科学通論' }, 79
{ 'オープン技術特論', 'IT アーキテクチャ特論' }, 73

図 6 必須科目を除いた高頻度共起科目群 (後期)

Fig. 6 High-frequency Co-occurring Subjects Excluding Compulsory Subjects (2nd Semester)