学科の特異性を明らかにするための科目概念の推定手法に関する研究

筑波大学
人間総合科学学術院人間総合科学研究群
情報学学位プログラム

2022年3月

熊田　大雅

学科の特異性を明らかにするための
科目概念の推定手法に関する研究

A Study on the Concept Estimation Method of Subjects for Clarifying Specificities of Academic Departments

氏名：熊田　大雅
Kumada Taiga

大学などの高等教育機関が一般公開しているシラバスは，講義の概要や目的，授業計画といった科目内容をまとめた文書であり，在学生をはじめとした様々な人々が授業の傾向や学科のカリキュラムを把握するために利用している．しかし，現在のシラバスは個々の科目ごとに独立して作成されているため，複数の科目間を比較することは困難を伴い手間がかかる．また，シラバスは科目ごとに断片化しているため，学科全体の傾向を捉えることは容易でない．

本研究では，上述した科目間の比較や学科全体の傾向を比較することを目的として，それぞれの科目に「伝統的科目」と「萌芽的科目」からなる指標を考案し，これらの指標に基づいて科目概念の推定手法を提案する．ここで，異なる学科や学群で開講される名称が一致している科目において類似の内容を履修する科目を伝統的科目と称し，科目名が類似していても異なる内容を履修する科目を萌芽的科目と称する．提案手法では，最初に，科目名を特徴量として名称が一致あるいは類似している科目を類似科目名群として分類する．分類は，科目名に対する正規化編集距離および非階層的クラスタリングを用いて行う．次に，類似科目名群ごとに，分類された科目のシラバスに出現する用語の頻度分布を算出し，科目間で頻度分布の分散や，頻度分布を表す近似曲線の係数，順位相関を用いて科目を重畳した時の変化を明らかにする．科目を重畳した際に，頻度分布の分散が大きくなる場合は，それらの科目は同一の内容を履修する科目であることから伝統的科目の傾向が高く，分散が小さくなる場合は萌芽的科目の傾向が高いと推定する．最後に，シラバスを収集した学科・学類ごとに，伝統的科目と萌芽的科目の構成比率を算出し，学科・学類の特異性を明らかにすることを目指す．

提案手法の有効性を検証するために，情報学を学ぶ5学科のシラバスを収集し，評価実験を行った．その結果，各手法の比較評価により，提案した各手法は学科の特異性を明らかにし，各科目の科目概念を推定することができることを示した．

主研究指導教員：佐藤　哲司
副研究指導教員：高久　雅生

目 次

第 1 章　序章 **1**
1.1 背景 1
1.2 本論文の構成 2

第 2 章　関連研究 **3**
2.1 シラバス分析およびシラバス可視化に関する研究 3
2.2 カリキュラムの全体構成に着目した研究 4
2.3 単語の出現頻度に着目した研究 5
2.4 分散表現を用いた文書類似度の算出に関する研究 5
2.5 本研究の位置づけ 6

第 3 章　科目概念の推定手法の提案 **7**
3.1 全体構成 7
3.2 科目内容の抽出手法 7
3.3 Normalized Levenshtein Distance の算出手法 8
3.4 類似科目名群における科目概念の推定手法 9
3.4.1 出現単語の頻度分散の算出手法 9
3.4.2 近似曲線の係数の算出手法 10
3.4.3 Spearman の順位相関係数の算出手法 10
3.5 孤立科目群における科目概念の推定手法 12
3.6 学科の特異性把握手法 12

第 4 章　評価実験 **14**
4.1 評価に使用するシラバス 14
4.2 類似科目名群における科目概念の推定結果 14
4.3 孤立科目群における科目概念の推定結果 16
4.4 学科の特異性の分析結果 17

第 5 章　考察 **26**
5.1 科目概念の推定に関する考察 26
5.1.1 代表的な科目群の科目概念の推定 26
5.1.2 特異な科目概念の推定 27
5.2 学科の特異性に関する考察 28
5.2.1 各手法における学科の特異性 28
5.2.2 学科の特異性の比較 29
5.3 提案手法の応用例に関する考察 30

第 6 章　まとめ **31**

謝辞 **32**

参考文献 **33**

図 目 次

3.1 提案手法の全体構成 8
3.2 科目の重ね合わせの例 9
3.3 近似曲線係数の考察に関するグラフ 11

4.1 出現頻度による科目概念の推定（データ構造とアルゴリズム） 18
4.2 出現頻度による科目概念の推定（機械学習） 18
4.3 出現頻度による科目概念の推定（最適化） 19
4.4 近似曲線による科目概念の推定（データ構造とアルゴリズム） 19
4.5 片対数近似曲線による科目概念の推定（データ構造とアルゴリズム） 20
4.6 近似曲線による科目概念の推定（機械学習） 20
4.7 片対数近似曲線による科目概念の推定（機械学習） 21
4.8 近似曲線による科目概念の推定（最適化） 21
4.9 片対数近似曲線による科目概念の推定（最適化） 22
4.10 伝統的度合いおよび萌芽的度合い（類似科目名群） 23
4.11 伝統的度合いおよび萌芽的度合い（孤立科目群） 24
4.12 3 手法と科目の類似度を組み合わせた伝統的度合いおよび萌芽的度合い . . . 25

表 目 次

4.1 分析に用いた各学科の科目数 14
4.2 出現単語に関する数値データ 15
4.3 出現単語の頻度分散，近似曲線の係数，順位相関係数の平均 15
4.4 出現単語の頻度分散，近似曲線の係数，順位相関係数の平均（標準化） . . . 16
4.5 類似科目名群における科目の T_{score} と S_{score} 16
4.6 孤立科目群における科目の T_{score} と S_{score} 17
4.7 各学科の T_{score} と S_{score} 17

5.1 各手法において最も特異な科目概念が推定された科目群 27
5.2 最も特異な科目概念が推定された科目群の出現単語に関する数値データ . . . 27
5.3 最も特異な科目概念が推定された科目群の 3 手法の算出値（標準化） 28

第1章　序章

1.1　背景

大学などの高等教育機関では，個々の授業ごとに授業内容をシラバスとして作成している．シラバスとは，各講義科目の詳細な授業計画が示された文書である．一般に，大学の授業名，担当教員名，講義目的，各回の授業内容，成績評価方法・基準，準備学習等についての具体的な指示，教科書・参考文献，履修条件等が記されている [1] ．

シラバスの主な利用者は各講義科目の情報を確認する在学生であるが，大学への入学を考えている受験生や編入生が自身の興味関心と志望学科のカリキュラムポリシーが合っているかを調べる際にも利用される．加えて，シラバスの利用者は学生だけにとどまらず，企業の採用担当者や大学に所属する職員および教員などのカリキュラムの管理者も利用することができる．企業の採用担当者は，学科の Web ページとともに，シラバスから各学科の特徴を理解することで，学科固有の学びや経験による学生の強みを理解し，企業が欲する人材の獲得につなげる．また，大学職員およびカリキュラムの管理者は，個々の科目デザインの更新や改善をシラバスを介して行える．さらに，講義科目を精査することで，カリキュラム全体の一貫性および整合性の検証を可能とする．ここで，学科の課程を終えることで得られる学科固有の学びや経験を本研究では学科の特異性と称する．

現在のシラバスは科目ごとに講義概要や授業計画が記述されているため，個々のシラバスから網羅的に学科の全体像を把握し，その特異性を明らかにすることは難しい．ましてや複数の学科のシラバスを通読し，学科同士を比較することは大変困難である．また，学科の特異性を把握するために，シラバスの代わりに学科名や学科の Web ページを用いることは有効な手段とは言えない．これは，ユーザが複数の Web ページを横断的に読み込む必要があり，学科の本質を理解するのに一定の作業量が必要なためである．また，Web ページに記載されている内容と実際の教育現場で運用されているカリキュラムに齟齬が生まれている可能性もある．

そこで本研究では，学科の特異性を明らかにするために科目概念を推定する手法を提案する．科目の持つ学びの広さと深さを「科目概念」と称し，科目を伝統的科目と萌芽的科目に分類する．本研究では，どの学科においても同様の内容を習得する科目を「伝統的科目」，学科によって異なる内容を習得する科目を「萌芽的科目」と称する．提案手法では，大学ごとに公開されているシラバスを収集し，科目内容を抽出する．抽出した科目内容に対して，形態素解析を行い，名詞におけるサ変接続，一般，固有名詞を分析対象とする．次に，各科目に対して科目概念を推定する．科目概念の推定では，科目名に対して Normalized Levenshtein Distance（以下 NLD とする）を適用することで科目を分類する．NLD が一定値を下回った科目を類似科目名群と称し，それらの科目内容を重ね合わせる．科目内容の重ね合わせとは同一単語の出現頻度を科目群ごとに累積させることである．この科目の重ね合わせに対して出現単語の頻度分散，近似曲線の係数，順位相関を算出し，科目を伝統的科目，萌芽的科目に分類する．一方で，NLD が一定値を上回った科目を孤立科目群と称し，科目名をクラスタリングする．作成されたクラスタ内で，分散表現を用いて科目内容

の類似度を算出することで科目を伝統的科目，萌芽的科目に分類する．それぞれで分類された科目を学科ごとに累積させることで，学科の特異性を明らかにする．

1.2　本論文の構成

本論文では，2 章で本研究に関連する先行研究を概観し，本研究の位置づけを明らかにする．3 章では学科の特異性を明らかにすることを目的とした科目概念の推定手法を提案する．4 章では提案手法の有効性を検証するために行った評価実験の内容と結果を述べる．5 章では評価実験を踏まえて考察を行う．6 章では全体をまとめ，今後の展望を述べる．

第2章　関連研究

本研究では，シラバスに出現した単語とその出現頻度に基づいて科目概念を推定し科目を伝統的科目と萌芽的科目に分類する．この手法と合わせて，シラバスの科目内容から分散表現を得ることで科目概念を推定している．これらの推定された科目概念を可視化することで，学科の特異性を明らかにする．このことから以下では，本研究に関連した先行研究を概観し，本研究の位置づけを示す.

2.1　シラバス分析およびシラバス可視化に関する研究

シラバスのテキストデータを分析し可視化することで，カリキュラムの全体像や科目間の関係性を明らかにする研究は数多くなされている．これらの研究は，抽出する科目内容の違いや分析手法および可視化手法の違いで差別化されている．

田中 [2] は，同一課程の異なる年度のシラバスを収集しコレスポンデンス分析することで，カリキュラムポリシーが科目内容に十分に反映されていることを示している．学科全体の授業科目および専門科目のみを対象としたコレスポンデンス分析を行うことで，専攻の特色を考察している．また，教員を養成するプログラムでは，実習科目が充実していることを科目内容の可視化によって明らかにしている．

また同様の研究として，宮原 [3] は，基礎演習および専門演習と位置付けられた学びの中核となる科目に着目し，同一学部における複数年度のシラバス分析およびシラバス可視化をしている．シラバスのテキストデータから,「授業の概要」,「授業の到達目標」を抽出し，コレスポンデンス分析を行っている．加えて，抽出語を共起ネットワークで可視化することで,「授業の概要」,「授業の到達目標」の双方において，学年が上がるにつれて受講する授業の内容が基本的な内容から専門的な内容に遷移する傾向があることを示している．

中村 [4] らは，自大学の理工学に関する学科のシラバスを収集し，トピックモデルを分析に用いることで，学科間の傾向を明らかにした．シラバスの科目内容から「授業概要」,「目的」,「到達目標」を抽出し，1 科目 1 文書として科目間の関係を LDA（Latent Dirichlet Allocation）を用いて分析している．彼らは，これらの手法が学科名が更新された場合における授業内容の変遷の分析に応用可能であると述べている．

永嶋 [5] は，新たなカリキュラムの運用を開始した自大学の学部のシラバスに対して，形態素解析サービスの WebAPI を用いたシラバス分析システムを構築し分析している．科目間の抽出語の特徴を明らかにするためネットワーク分析を用いて科目間の抽出語の関係性を可視化している．これらの可視化結果から，カリキュラムの特徴を明らかにするとともに，シラバス設計の改善点を指摘している．

特定の科目を学ぶことができる複数学部を横断的に分析した研究として，石井 [6] の研究がある．石井は，複数の学部における特定の共通教育科目に対して，シラバスから出現単語を抽出し共起ネットワークを作成している.「授業概要」,「到達目標」,「授業計画」のそれぞれについて抽出語と共起ネットワークを分析することで，特定の学部を有する大学の特徴や指導方針などを示している．

竹森 [7] らは，doc2vec [8] を用いて科目内容の分散表現を得ることで，階層的クラスタリングを行った．クラスタごとにワードクラウドを適用することで，クラスタの性質を可視化している．可視化されたクラスタをユーザに提示し，ユーザの興味のある候補科目を絞り込むことで，ユーザに科目を推薦する．これにより，複数の科目の中から自分自身の履修方針にあった科目の発見および検討が容易になるとしている．

下司 [9] らは，シラバスの文書群を用いることで，単語間の上位下位関連の抽出手法を提案している．出現した単語の頻度を用いて専門性の高い単語を抽出することで，出現単語をノード，上位の単語と下位の単語の組をエッジとした概念グラフを生成し，上位下位関連を求めている．

米田 [10] らは，シラバスのテキストデータを Linked Open Data 化することで，再利用性や相互運用性を高めている．また米田らは，教育分野において Linked Open Data 化したシラバスの活用の可能性を示唆している．特に，カリキュラム分析に応用できると考えており，実際に検証した情報系の専門教育課程のカリキュラム分析では，共通キャリア・スキルフレームワーク（CCSF）の知識項目との関連を分析し，クロス集計を行うことでカリキュラムの特徴を可視化している．

2.2　カリキュラムの全体構成に着目した研究

学びのグローバル化が進んでいる現代では，高等教育における国際的な人材の育成が重要視されている [11]．これに伴い，シラバスおよびカリキュラムの作成には国際的な視点が不可欠となっている．Fuentes [12] らは，高等教育において公平性，多様性，包括的（EDI）の 3 指標が重要であることを強調し，シラバスの作成および改訂支援を目的とした指針を定めている．シラバスに EDI を導入することで，世界的な問題となっている人種差別などによる疎外感を減少させ，学生間の異文化コミュニケーションを円滑に行えると結論付けている．

また，シラバスのテキストデータからカリキュラムの特徴を抽出することで，簡潔にシラバスやカリキュラムの作成および更新ができる．野澤ら [13] は，シラバスのテキストデータから抽出した専門用語の出現頻度に基づいて，シラバス間の類似度計算およびクラスタリングを行い，クラスタへの帰属分布の分析によって，カリキュラムの特徴理解を支援するシステムを提案している．野澤らは，カリキュラムの設計および分析の重要性を指摘しており，このシステムは，学問や技術，社会のニーズに合った独創的なカリキュラム設計や，教育機関のカリキュラム評価の支援を可能としている．

金城 [14] [15] は，新たに改訂される教育要領の趣旨や内容に則したカリキュラムの構築やシラバスの改定を目的とし，質的および量的の 2 側面からシラバスを分析している．質的な分析では，学科における類似した授業テーマをカテゴリとしてまとめ，カテゴリごとの授業テーマ数（学科数）を算出し，特定の分野における授業計画と授業テーマ数の傾向を調査した．量的な分析では,「授業概要」,「到達目標」,「授業計画」を対象にテキストマイニングを行った．作成した共起ネットワークを分析することで，授業内容の特徴や語句間の関係性の構造を明らかにしている．

齋藤ら [16] は，学生が講義中に得られた学習成果をリアルタイムで大学内外に公開する e シラバスのシステムの構築を目的とした，学習成果可視型のシラバス作成支援システムを開発している．このシステムによって，明らかにされた学習成果をレーダーチャート上に反映することで，機関単位の統一的枠組みを可視化し機関全体の根拠の一つになり得ることが期待されるとしている．

増田 [17] は，カリキュラムの比較を目的とした講義シラバスの自動分類を行っている．Random Forest を用いて科目シラバスに対して自動的に日本十進分類（NDC）を付与している．科目と図書の結びつけを容易にするために日本十進分類を用いている点が特徴的である．

2.3 単語の出現頻度に着目した研究

専門用語の抽出や文書の有用性を検証するために単語の出現頻度を用いる研究は多く存在している．これらの研究は，特定の文書やコーパスから単語を抽出し，その出現頻度に基づいて，単語のバリューを定めている．

和多ら [18] は，文書の表層的な特徴量である単語の出現頻度から文書の有用性を判定する手法を提案している．ある特定のテーマに頻出する特徴的な名詞，動詞，形容詞，副詞を抽出することで，そのテーマを持つ文書群と，一般的な文書群に出現する単語の頻度を比較している．

中川 [19] らは，分野固有の概念への関連性の強さを表すターム性に着目しながら，単語の出現頻度と連接頻度を用いることで，名詞を対象とした専門用語の抽出手法を提案している．具体的には，単語の出現頻度と複合名詞を形成している単名詞に連接している単語頻度を統計量として用いてスコア付けすることで，単語の重要度を判定している．

柿本ら [20] は,「テキスト中のある区間において複数回出現している単語群は，そのテキスト中で話題を形成している単語群である」と定義し，単語の出現頻度を用いて，時間の変化を伴ったテキスト中の話題推移を可視化している．また，出現した単語の頻度傾向の相関や単語の共起情報を分析することで，テキストに出現する単語の話題分割を可能としている．

2.4 分散表現を用いた文書類似度の算出に関する研究

単語や文書から分散表現を得ることで，単語間および文書間の類似度を測定することが可能となる．単語の分散表現を利用して文書間の類似度を算出した研究として，柳本 [21] の研究がある．柳本は，word2vec [22] で単語の分散表現を獲得し，Earth Mover's Distance [23] を用いることで，単語間の類似性を考慮した文書間の類似度を算出する手法を提案している．ニュース記事をコーパスとして用いて提案手法の有効性を検証し，同義語や類義語を配慮した文書間の類似度を求められることを確認している．

新濱 [24] らは，オンラインジャッジシステム[1]の問題文間の類似度を算出することで，学習者のレベルに合わせた問題選択を支援している．word2vec と doc2vec を用いて問題文の分散表現を獲得し評価することで，問題文間の類似度が問題選択における新たな指標になり得ることを示している．

一方で，word2vec のパラメータ変化による影響を検証した研究として，内田 [25] の研究がある．内田は，ベクトルの次元数と適用文脈の幅を段階的に変更することで，類義語の抽出結果に与える影響を分析している．また，モデルの検証を全単語，名詞のみ，形容詞のみ，動詞のみの 4 パターンで行っていることも特徴的である．

[1] プログラミングコンテストに用いられるプログラム問題の自動採点を行うシステム

2.5　本研究の位置づけ

シラバスから有用なテキストデータを抽出し可視化する研究は数多くなされているが，これらの研究は特定の学科，専攻を対象としており，複数の大学あるいは学科のシラバスを用いて分析を行う研究は知られていない．本研究では，科目の持つ学びの広さと深さといった科目概念に着目し，科目を伝統的科目と萌芽的科目に分類することで，複数の大学あるいは学科のシラバスから，学科の特異性を明らかにしていることに特徴がある．複数の学科において似た文字列の科目名をもつ科目に着目し，出現する単語の出現頻度を累積させることで，科目概念を推定する手法を提案している．加えて，似た意味合いの科目名を持つ科目に対してもクラスタリングを行い，分散表現を用いて科目間の類似度を算出することで，科目概念を推定している．これらの推定された科目概念を学科ごとに累積させることで，学科の特異性を明らかにしている．

第3章　科目概念の推定手法の提案

3.1　全体構成

本研究ではシラバスに出現する単語の出現頻度に着目し，科目を伝統的科目と萌芽的科目に分類し，科目概念を推定することで学科の特異性を明らかにする．科目概念を推定する提案手法の全体構成を**図 3.1** に示す．

大学あるいは学科，専攻ごとに異なる科目展開がなされているが，類似した学科等であれば一定程度共通する科目が存在する．これらの共通して展開されている科目は，異なる大学の学科間で比較しても類似した科目名が付けられていることが多い．そこで，これらの科目の科目内容を比較するために，正規化した編集距離を用いて科目を分類する．収集したシラバスの科目名から NLD（Normalized Levenshtein Distance）を算出する．NLD が θ 未満の科目群それぞれに対して，出現単語の頻度を足し合わせることで科目を重ね合わせる．科目を重ね合わせた科目群に対して，出現単語の頻度分散，近似曲線の係数，Spearman の順位相関に基づき，科目を伝統的科目と萌芽的科目に分類する．

一方で，NLD が θ 以上の科目群の科目名に対して，k-means によるクラスタリングを行うことで，意味的に似た科目名のクラスタを生成する．生成したクラスタ内の科目内容の類似度を総当たりで算出することで，科目を伝統的科目と萌芽的科目に分類する．

3 種類の科目の分類結果と類似度を用いた分類結果をそれぞれ組み合わせることで，対象の学科の科目傾向を明確にし特異性を明らかにする．

3.2　科目内容の抽出手法

収集したシラバスから科目内容を抽出する．本研究で抽出する科目内容は以下の 6 項目とする．

- 科目名
- 授業概要
- 授業計画
- 到達目標
- キーワード
- 教材，参考文献

これら 6 種類の科目内容に対して形態素解析を行い形態素に分解する．

シラバスの科目内容を分析する場合は，機能語が不要である．機能語とは，代名詞，前置詞，接続詞，助動詞などの語彙的意味を持たない非自立語である．そこで，抽出した科目内容から接続詞，前置詞，助詞などの科目内容と関係の薄い品詞を除外した．また本研究で

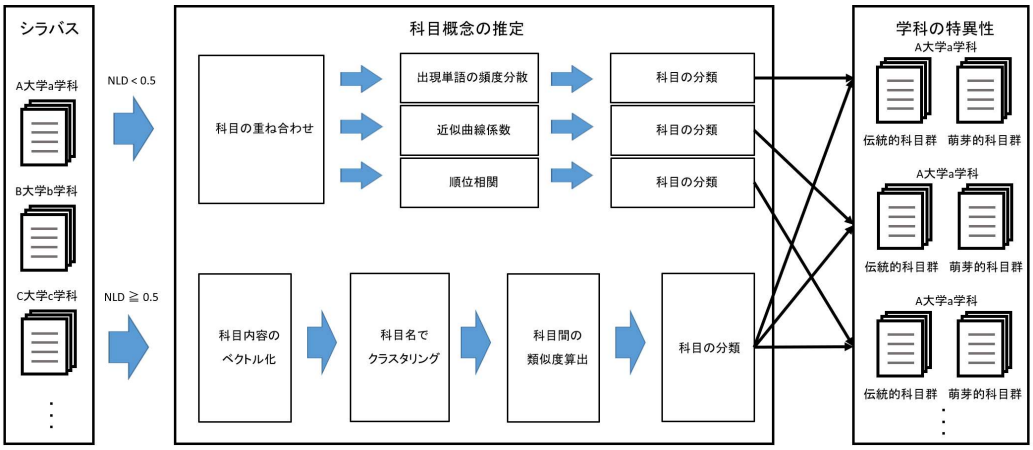

図 3.1: 提案手法の全体構成

は，科目概念の推定を行っている．科目概念の推定では，科目の持つ学びの広さと深さを，単語自体に直接的な意味を持つ科目内容の出現語から推定している．そのため，分析に用いる品詞を名詞と限定し，サ変接続，一般，固有名詞のみ抽出し分析に用いた．このサ変接続，一般，固有名詞は後述する mecab-ipadic-NEologd で用いられている名詞の分類体系である．なお，単一の数字およびアルファベットは分析対象から除いた．

形態素解析器は MeCab[1] を使用し，単語分かち書き用の辞書は mecab-ipadic-NEologd [26–28] を使用する．mecab-ipadic-NEologd は固有名詞や複合名詞を 1 単語として分かち書きする MeCab 用の辞書である．特定の分野に頻出する専門用語はその分野を体現する重要な複合語であるため，形態素に分解せず語の形を維持したまま処理する必要がある．そのため，単語分かち書き用の辞書として mecab-ipadic-NEologd を使用した．

3.3 Normalized Levenshtein Distance の算出手法

収集したすべてのシラバスの科目名に対して，NLD を総当たりで算出する．Levenshtein Distance [29] とは，ある文字列から別の文字列へ変形する際の挿入・削除・置換の最小回数で定義される．

本研究では Levenshtein Distance を正規化した NLD を用いる．ある 2 つの科目名を str_1, str_2 としたとき，NLD は以下の式で与えられる．

$$NLD = \frac{LevenshteinDistance(str_1, str_2)}{max(len(str_1), len(str_2))} \tag{3.1}$$

ここで，$max(len(str_1), len(str_2))$ は，より文字列の長い科目名の文字数を表している．

以下では，NLD が θ 未満の科目群を類似科目名群，NLD が θ 以上の科目群を孤立科目群と称する．ただし，NLD が θ 未満の科目群のうち，科目ペアの組み合わせが同一の学科のみで構成されていた場合は，その科目ペアの組み合わせを孤立科目群とみなす．

[1] https://taku910.github.io/mecab/ （最終閲覧 2021 年 12 月）

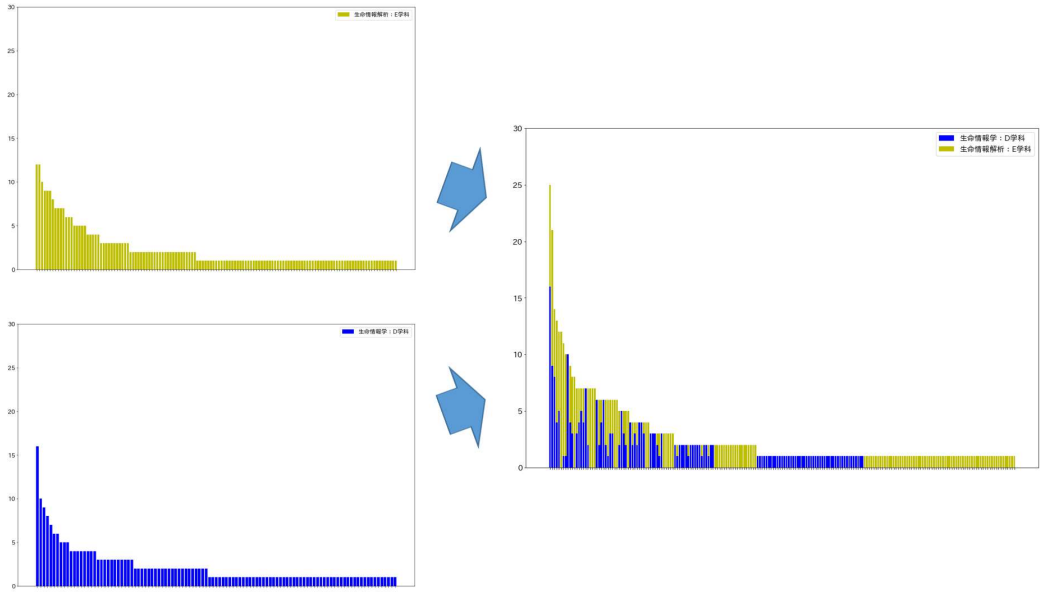

図 3.2: 科目の重ね合わせの例

3.4 類似科目名群における科目概念の推定手法

本節では，NLD によって分類された類似科目名群に対して，科目概念の推定手法を明示する．まず，類似科目名群の中から，NLD が θ 未満となった科目組み合わせの科目を重ね合わせる．この科目の重ね合わせは,「どの学科においても同様の内容を習得する科目」である伝統的科目や,「学科によって異なる内容を習得する科目」である萌芽的科目といった,複数学科間での科目内容の比較および検討を前提としている．そのため，NLD が θ 未満となった科目の組み合わせごとに科目を重ね合わせた．

科目の重ね合わせのイメージを**図 3.2** に示す．科目の重ね合わせでは，科目の組み合わせごとに同一の単語の出現頻度を合算する．その後，出現単語を出現頻度の合算値の大きい順に並び替える．これらの出現単語と出現頻度に対して，以下では 3 種類の科目の分類手法を用いて，科目を伝統的科目および萌芽的科目に分類し科目概念を推定する．

3.4.1 出現単語の頻度分散の算出手法

科目を重ね合わせた科目群それぞれに対して，出現単語の頻度分散を算出する．分散とは，標本や母集団のばらつきの程度を表すための指標である．単語の出現頻度を $x_1, x_2, \cdots, x_n$ としたとき，頻度分散 s^2 は以下の式で算出される．

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{3.2}$$

$\bar{x}$ は単語の出現頻度の平均値を表す．

ここで，科目を重ね合わせた科目群における頻度分散 s^2 について考察する．伝統的科目である同様の内容を持つ科目が複数出現すると，科目の重ね合わせにより特定の単語の出

現頻度が高くなり，出現単語の頻度のばらつきが大きくなる．そのため，伝統的科目の頻度分散は大きい値が算出されると推測できる．

一方で，萌芽的科目である異なる内容をもつ科目が複数出現すると，科目の重ね合わせにより全体的に単語の出現頻度が同程度になり，出現単語の頻度のばらつきが小さくなる．そのため，萌芽的科目の頻度分散は小さい値が算出されると推測できる．

よって，出現単語の頻度分散による科目の分類では，頻度分散が大きいほどその科目群が伝統的科目である度合いが高いとみなす．また，頻度分散が小さいほどその科目群が萌芽的科目である度合いが高いとみなす．

3.4.2 近似曲線の係数の算出手法

科目を重ね合わせた科目群それぞれに対して近似曲線の係数を算出する．科目を重ね合わせた科目群において，出現単語の頻度を高い順に並び替えたものを横軸に，その頻度を縦軸にした際に描画される折れ線グラフを曲線に累乗近似する．

累乗近似に用いる式を以下のように定義する．

$$y = bx^a \tag{3.3}$$

累乗近似では，数値解析を行う Python 用のパッケージである scipy[2]を用いて，近似する a と b を返す非線形回帰を行った．ここで，$b = 1$ として a の値を変化させた近似曲線係数の考察に関するグラフを**図 3.3** に示す．**図 3.3** をもとに，科目を重ね合わせた科目群における近似曲線の係数 a について考察する．本研究では，出現単語を出現頻度の合算値の大きい順に並び替え，折れ線グラフを作成している．この折れ線グラフを累乗近似することで，単語の出現頻度を曲線として表している．この曲線の曲率は，**式 3.3** の a に依存しており，**図 3.3** より，a の値が小さくなるほど $0 < x < 1$ の範囲で y の値が大きくなる．これに対して，a の値が大きくなるほど $0 < x < 1$ の範囲で y の値が小さくなる．

これらを踏まえて近似曲線の係数 a について検討する．伝統的科目である同様の内容を持つ科目が複数出現すると，科目の重ね合わせにより特定の単語の出現頻度が高くなる．そのため，関数 $y = bx^a$ における指数部 a の値が小さくなると推測できる．

一方で，萌芽的科目である異なる内容をもつ科目が複数出現すると，科目の重ね合わせにより全体的に単語の出現頻度が同程度になる．そのため，関数 $y = bx^a$ における指数部 a の値が大きくなると推測できる．

ここで，指数部 a を他の尺度と統一した相対振幅にするために指数部 a に -1 を乗算する．よって，近似曲線の係数による科目の分類では，指数部 a の値が大きいほど，その科目群が伝統的科目である度合いが高いとみなす．また，近似曲線の係数が小さいほど，その科目群が萌芽的科目である度合いが高いとみなす．

3.4.3 Spearman の順位相関係数の算出手法

類似科目名群それぞれに対して Spearman の順位相関係数 [30] を算出する．Spearman の順位相関とは，2 変量の順位変数間の相関関係を評価する指標であり，その相関係数 ρ は，$-1 \leq \rho \leq 1$ をとる．

[2]https://scipy.org/ （最終閲覧 2021 年 12 月）

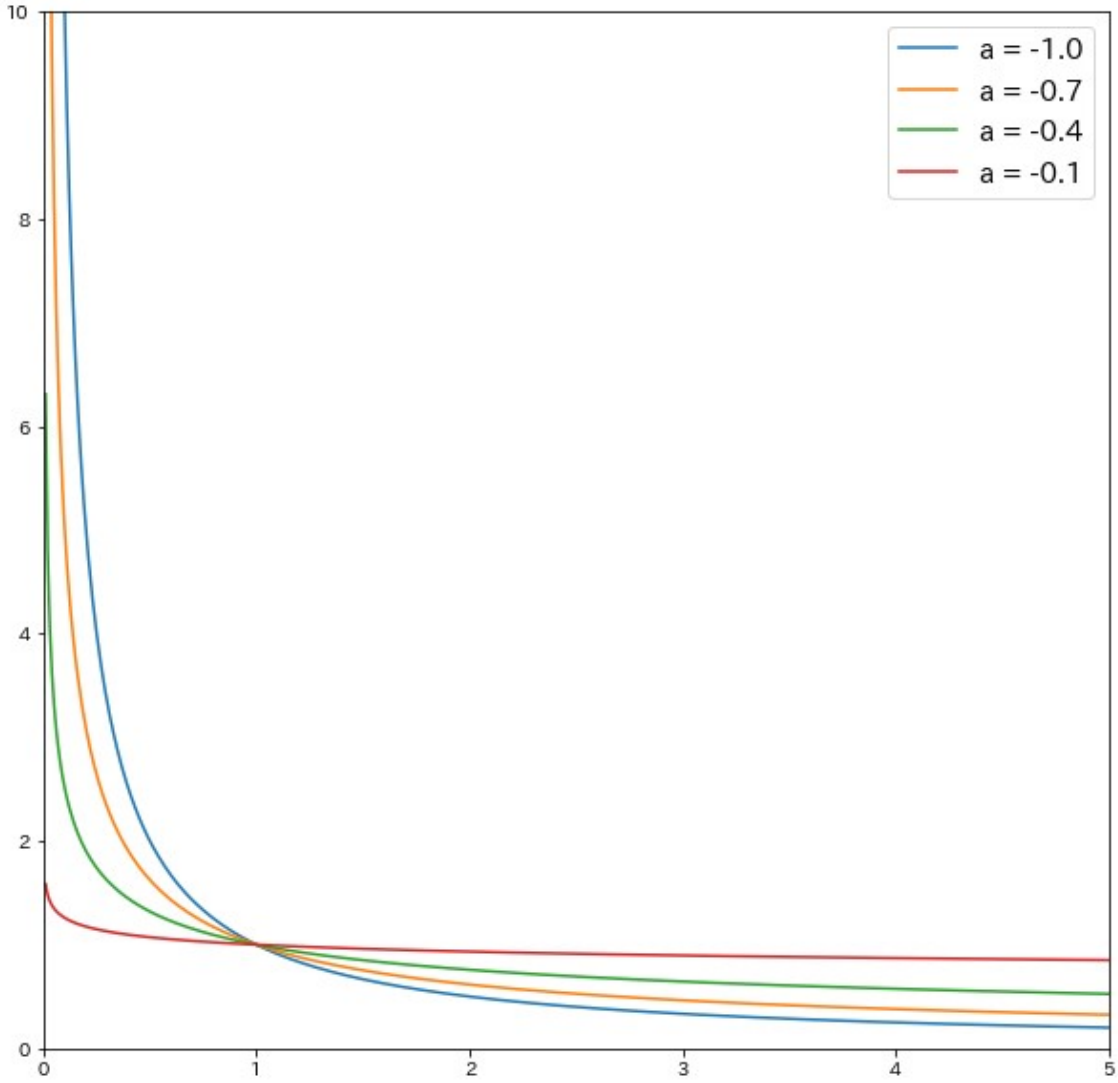

図 3.3: 近似曲線係数の考察に関するグラフ

科目 i と科目 j における同一単語の順位差を D，出現単語ペアの数を N としたとき，Spearman の順位相関係数 ρ_{ij} は以下の式で与えられる．

$$\rho_{ij} = 1 - \frac{6\sum D^2}{N^3 - N} \tag{3.4}$$

ただし，同順位が存在した場合，その順位を最も小さい順位として取り扱う．

また，科目を重ね合わせた科目群内の科目ペアの数を S としたとき，類似科目名群における Spearman の順位相関係数の平均値 $\bar{\rho}$ は以下の式で与えられる．

$$\bar{\rho} = \frac{\sum_{i,j,i<j} \rho_{ij}}{S} \tag{3.5}$$

類似科目名群における Spearman の順位相関係数の平均値 $\bar{\rho}$ について考察する．伝統的科目では，同様の内容を持つ科目間において，同一単語の順位を比較するとその単語の順位差が小さく算出されると推測できる．そのため，Spearman の順位相関係数の平均値が大きくなる．

一方で，萌芽的科目では，異なる内容をもつ科目間において，同一単語の順位を比較するとその単語の順位差が大きく算出されると推測できる．そのため，Spearman の順位相関係数の平均値が小さくなる．

よって，Spearman の順位相関係数の平均値による科目の分類では，Spearman の順位相関係数の平均値が大きいほどその科目群が伝統的科目である度合いが高いとみなす．また，Spearman の順位相関係数の平均値が小さいほどその科目群が萌芽的科目である度合いが高いとみなす．

3.5 孤立科目群における科目概念の推定手法

本節では，NLD によって分類された孤立科目群に対して科目概念を推定する手法を明示する．まず，NLD によって分類された孤立科目群の科目名を抽出し，それぞれを 300 次元のベクトルに変換して分散表現を得る．ベクトル化技術には，doc2vec の PV-DBOW（Paragraph Vector with Distributed Bag of Words) モデル[3]を採用した．

獲得された分散表現を特徴量として，非階層的クラスタリングを行う．なお，非階層的クラスタリング分析には k-means 法を用いた．クラスタリングとは，特徴が類似したデータの集合を部分集合（クラスタ）として扱い，全体が如何なるクラスタで構成されているかを俯瞰する手法である．非階層的クラスタリングは，データが指定された数のクラスタに分類されるように，特徴が類似したデータを切り分けクラスタを作成する．

次に，科目名をベクトル化した同様のモデルを用いて，科目内容を 300 次元のベクトルに変換して分散表現を得る．この分散表現を用いて，作成したクラスタ内で科目間の cos 類似度を総当たりで算出する．a_i，b_i それぞれを科目内容ベクトルとすると，cos 類似度 cos_sim は以下の式で与えられる．

$$cos\_sim = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}} \tag{3.6}$$

ある科目に対して，総当たりで算出された cos 類似度の平均値を科目の類似度 $cos\_sim_{average}$ とする．

ここで，科目の類似度 $cos\_sim_{average}$ について考察する．伝統的科目において，同様の内容を持つ科目間の類似度は高くなる．そのため，伝統的科目の科目の類似度は大きい値をとると推測できる．

一方で，萌芽的科目において，異なる内容をもつ科目間の類似度は低くなる．そのため，萌芽的科目の科目の類似度は小さい値をとると推測できる．

よって，科目の類似度による科目の分類では，科目の類似度が高いほど，その科目群が伝統的科目である度合いが高いとみなす．また，科目の類似度が低いほど，その科目群が萌芽的科目である度合いが高いとみなす．

3.6 学科の特異性把握手法

3.4.1 項，**3.4.2 項**，**3.4.3 項**および **3.5 節**で算出した出現単語の頻度分散，近似曲線の係数，Spearman の順位相関，科目の類似度を，統一的な尺度に変換することを目的として，それぞれ標準化する．

[3]https://github.com/yagays/pretrained_doc2vec_ja （最終閲覧 2021 年 12 月）

標準化後の値 V に対して，$1.5 \leq V$ であれば，$T_{score}2$ を科目に対して付与する．以下，$0.5 \leq V < 1.5$ であれば $T_{score}1$，$-1.5 < V \leq -0.5$ であれば $S_{score}1$，$V \leq -1.5$ であれば $S_{score}2$ を科目に対して付与する．ただし，複数の科目群にわたって出現した科目は付与されたスコアの平均値を与える．

科目に対して付与されたスコアを学科ごとに累積させることで学科の特異性を明らかにする．出現単語の頻度分散と科目の類似度，近似曲線の係数と科目の類似度，Spearman の順位相関と科目の類似度の 3 手法をそれぞれ比較することで，提案手法の有効性を検証する．

第4章　評価実験

提案手法の有効性を検証するために，各手法について比較評価する．また，各学科の科目分布の傾向を示すとともに，最も特徴的な科目概念を持つ科目を明らかにすることに取り組む．

4.1　評価に使用するシラバス

情報学を学ぶことができる日本の 5 学科のシラバスを評価に使用した．これらのシラバスから専門科目および専門基礎科目を抽出し分析に用いる．ただし，科目名の末尾に数字やアルファベットを付与しクラス分けする同一の科目内容をもつ科目は 1 つの科目のみを抽出する．以下では，5 学科をそれぞれ A 学科，B 学科，C 学科，D 学科，E 学科と称する．本研究で分析に用いた各学科の科目数と類似科目名群に属する科目数および孤立科目群に属する科目数を**表 4.1** に示す．

表 4.1: 分析に用いた各学科の科目数

	A 学科	B 学科	C 学科	D 学科	E 学科
類似科目名群科目数	27	25	20	21	20
孤立科目群科目数	79	71	30	37	21
合計科目数	106	96	50	58	41

ここで，Web サイトで公開されているカリキュラムポリシーを要約し，各学科の特徴を記述する．A 学科は，文理をまたいだ広い領域における基礎理論や基盤技術の獲得を教育目標としている．文理融合したカリキュラムに沿って授業を展開することで，知識や情報を適切に構築・管理するため能力および技術力を育成するといった特徴をもつ．B 学科および C 学科は，データサイエンスの名を冠する学科であり，高度なデータ処理能力およびデータ分析力の育成に力点を置いている．データサイエンスの応用領域は，人文・社会科学系分野が多く含まれるため，広い学問領域をおさえている文理融合したカリキュラム構成になっている．D 学科および E 学科は，高度情報化社会における情報の本質を究明し，数理的思考によって高度な実際問題を解決できる人材の育成を目指している．また，情報工学における先進知識や先端技術を学ぶことができるといった特徴をもつ．

4.2　類似科目名群における科目概念の推定結果

ここで，類似科目名群と孤立科目群の分類に用いた NLD（Normalized Levenshtein Distance）の θ を暫定的に 0.5 とする．

類似科目名群には全 351 科目中 113 科目が分類された．NLD の値が 0.5 未満となった科目の組み合わせは 290 組存在し，これらを NLD によって科目名ごとに分類すると 39 の群に分けられる．分類された 39 の科目群それぞれに対して科目を重ね合わせた．

これらの科目群に対して，出現単語の頻度分散，近似曲線の係数，Spearman の順位相関係数を用いて，科目概念を推定する．ここでは，科目を重ね合わせた科目群の一例として，「データ構造とアルゴリズム」,「機械学習」,「最適化」の 3 科目群についての科目概念の推定結果を示す．まず，これらの科目群の出現単語に関する数値データを**表 4.2** に示す．

表 4.2: 出現単語に関する数値データ

	出現単語数	1 科目あたりの出現単語数	出現単語の頻度平均
データ構造とアルゴリズム	302	100.7	2.323
機械学習	430	107.5	1.838
最適化	325	108.3	1.958

「データ構造とアルゴリズム」および「最適化」は 3 科目で構成された科目群であり,「機械学習」は 4 科目で構成された科目群である．この 3 科目群を含めたすべての類似科目名群における 1 科目群当たりの出現単語数の最大値は 223.5 であり，最小値は 77，平均値は 123.25 であった．

科目を重ね合わせた科目群の単語の出現頻度を**図 4.1**，**図 4.2**，**図 4.3** に示す．**図 4.1** より,「データ構造とアルゴリズム」では一部の特定の単語が多く出現していることがわかる．出現頻度の高い単語の出現頻度を学科間で比較すると，学科ごとの出現頻度に大きな差がないことがわかる．一方で**図 4.3** より,「最適化」では様々な種類の単語が出現しており，突出した出現頻度を持つ単語は見られない．

また，科目を重ね合わせた科目群の近似曲線を**図 4.4**，**図 4.5**，**図 4.6**，**図 4.7**，**図 4.8**，**図 4.9** に示す．なお，**図 4.5**，**図 4.7**，**図 4.9** は，縦軸に底を 10 とした片対数をとっている．**図 4.5**，**図 4.7**，**図 4.9** の近似曲線をそれぞれ比較すると,「データ構造とアルゴリズム」は出現頻度 1（$x = 10^0$）の軸を下回るのに対して,「最適化」は出現頻度 1（$x = 10^0$）の軸に漸近している．また「機械学習」は,「データ構造とアルゴリズム」と「最適化」の間をとった曲線の形状をしている．

これらの科目群の出現単語の頻度分散，近似曲線の係数，Spearman の順位相関係数の平均を**表 4.3** に示す．また，それぞれの値を標準化した結果を**表 4.4** に示す．

表 4.3: 出現単語の頻度分散，近似曲線の係数，順位相関係数の平均

	頻度分散	近似曲線係数	順位相関係数
データ構造とアルゴリズム	21.926	0.947	0.597
機械学習	5.751	0.652	0.723
最適化	3.438	0.539	0.558

表 4.4 および **3.6 節**より,「データ構造とアルゴリズム」は，頻度分散，近似曲線係数，順位相関係数のいずれの分類手法においても伝統的科目に分類された.「機械学習」は，頻度分散では伝統的科目に分類されるが，順位相関では萌芽的科目に分類された.「最適化」は，頻度分散と近似曲線係数において萌芽的科目と分類された.

表 4.4: 出現単語の頻度分散，近似曲線の係数，順位相関係数の平均（標準化）

	頻度分散	近似曲線係数	順位相関係数
データ構造とアルゴリズム	1.034	2.743	0.525
機械学習	−0.552	0.047	0.954
最適化	−0.779	−0.983	0.397

次に，これら 3 科目群を含めた類似科目名群における学科ごとの T_{score} と S_{score} を**表 4.5** に示す．ここで，T_{score} および S_{score} を類似科目名群に属する科目数で除算する．算出された商をそれぞれ伝統的度合いおよび萌芽的度合いとする．

表 4.5: 類似科目名群における科目の T_{score} と S_{score}

		頻度分散		近似曲線係数		順位相関係数	
	科目数	T_{score}	S_{score}	T_{score}	S_{score}	T_{score}	S_{score}
A 学科	27	12.0	7.0	8.0	5.5	22.0	3.0
B 学科	25	23.0	9.0	6.0	4.5	22.5	5.0
C 学科	21	6.0	10.0	4.0	10.0	8.0	11.0
D 学科	20	7.5	9.5	3.5	7.5	12.5	7.0
E 学科	20	14.0	2.0	4.0	4.0	17.0	7,0

類似科目名群における学科ごとの伝統的度合いと萌芽的度合いの関係を**図 4.10** に示す．**図 4.10** より，頻度分散は学科によって散らばって分布しているが，近似曲線の係数は伝統的度合いが低く分布している．また，Spearman の順位相関は伝統的度合いがやや高い傾向にあることが見て取れる．

学科ごとに結果を確認すると，A 学科，B 学科，E 学科は全体的に萌芽的度合いが低く算出されている．特に B 学科は，頻度分散および Spearman の順位相関において，伝統的度合いが極めて高い傾向を示しているのに対し，近似曲線係数の伝統的度合いは低い．C 学科は全体的に中央に分布しており，萌芽的度合いはどの手法においても似た値を示しているが，伝統的度合いは手法によって異なった値を示している．D 学科はどの手法においても伝統的度合いと比べて萌芽的度合いが高く算出されている．

4.3 孤立科目群における科目概念の推定結果

孤立科目群には 351 科目中 238 科目が分類された．238 科目に対して，暫定的にクラスタ数を $k = 20$ としたクラスタリングを行い，科目をクラスタごとに分類した．分類されたクラスタごとに科目の類似度を算出し，科目概念を推定した．

孤立科目群における学科ごとの T_{score} と S_{score} を**表 4.6** に示す．孤立科目群における学科ごとの伝統的度合いと萌芽的度合いの関係を**図 4.11** に示す．**図 4.11** の伝統的度合いおよび萌芽的度合いは，類似科目名群における科目の伝統的度合いと萌芽的度合いの算出方法と同様に，各スコアを孤立科目群に属する科目数で除算したものである．

図 4.11 より，A 学科，B 学科は萌芽的度合いが高く，C 学科，D 学科，E 学科は伝統的度合いが高いことがわかる．特に E 学科は，萌芽的度合いと比べて伝統的度合いが高い．

表 4.6: 孤立科目群における科目の T_{score} と S_{score}

	科目数	T_{score}	S_{score}
A 学科	79	3.0	25.0
B 学科	71	16.0	35.0
C 学科	30	14.0	3.0
D 学科	37	14.0	8.0
E 学科	21	17.0	5.0

4.4 学科の特異性の分析結果

科目に対して付与された伝統的度合いおよび萌芽的度合いを学科ごとに累積させる．各学科の T_{score} と S_{score} を**表 4.7** に示す．また，各学科における伝統的度合いと萌芽的度合いの関係を**図 4.12** に示す.

表 4.7: 各学科の T_{score} と S_{score}

		頻度分散		近似曲線係数		順位相関係数	
	科目数	T_{score}	S_{score}	T_{score}	S_{score}	T_{score}	S_{score}
A 学科	106	15.0	32.0	11.0	30.5	25.0	28.0
B 学科	96	39.0	44.0	31.0	39.5	38.5	40.0
C 学科	58	20.0	13.0	18.0	13.0	22.0	14.0
D 学科	50	21.5	17.5	17.5	15.5	26.5	15.0
E 学科	41	31.0	7.0	21.0	9.0	34.0	12.0

図 4.12 より，すべての学科がどの手法においても比較的近い距離に分布していることがわかる．A 学科および B 学科はいずれの手法においても，伝統的度合いと比べて萌芽的度合いが高く算出されている．これに対して，C 学科，D 学科，E 学科では，いずれの手法においても，萌芽的度合いと比べて伝統的度合いが高く算出されている．全体的な傾向として各学科内で，Spearman の順位相関+科目の類似度が一番伝統的度合いが高く，近似曲線の係数+科目の類似度が一番伝統的度合いが低い傾向が見られた.

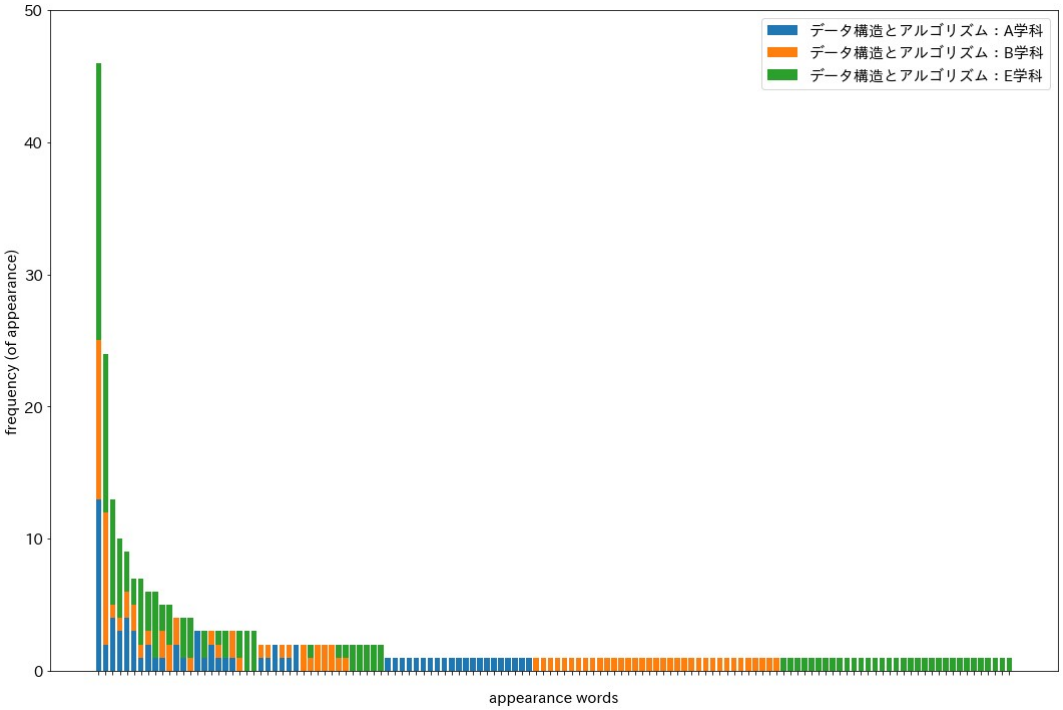

図 4.1: 出現頻度による科目概念の推定（データ構造とアルゴリズム）

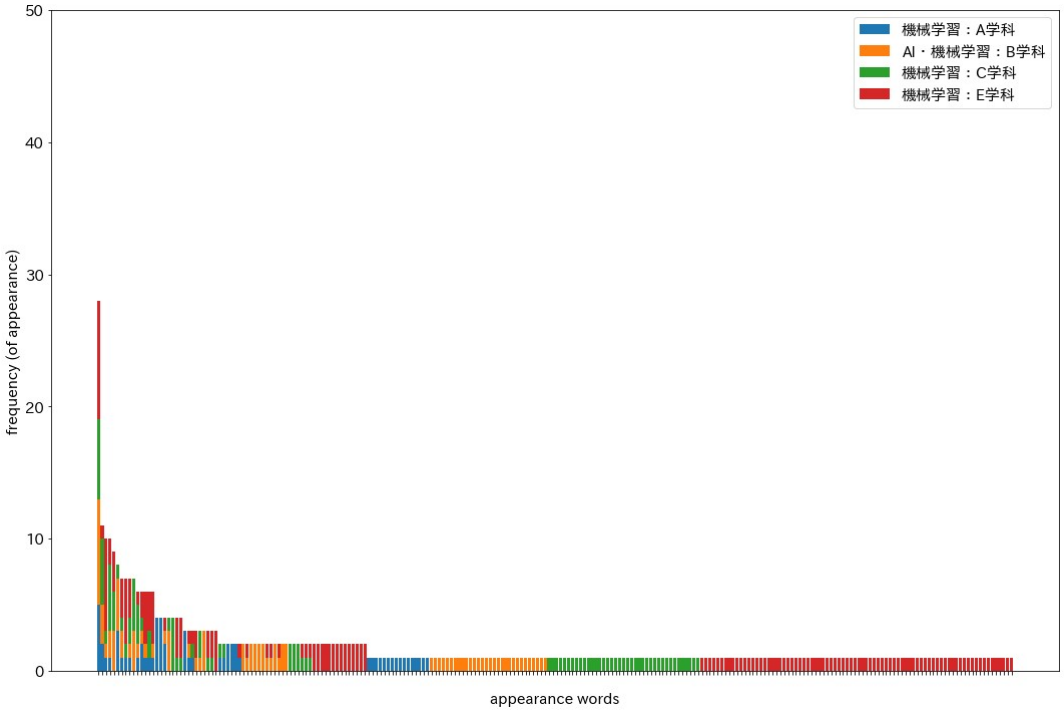

図 4.2: 出現頻度による科目概念の推定（機械学習）

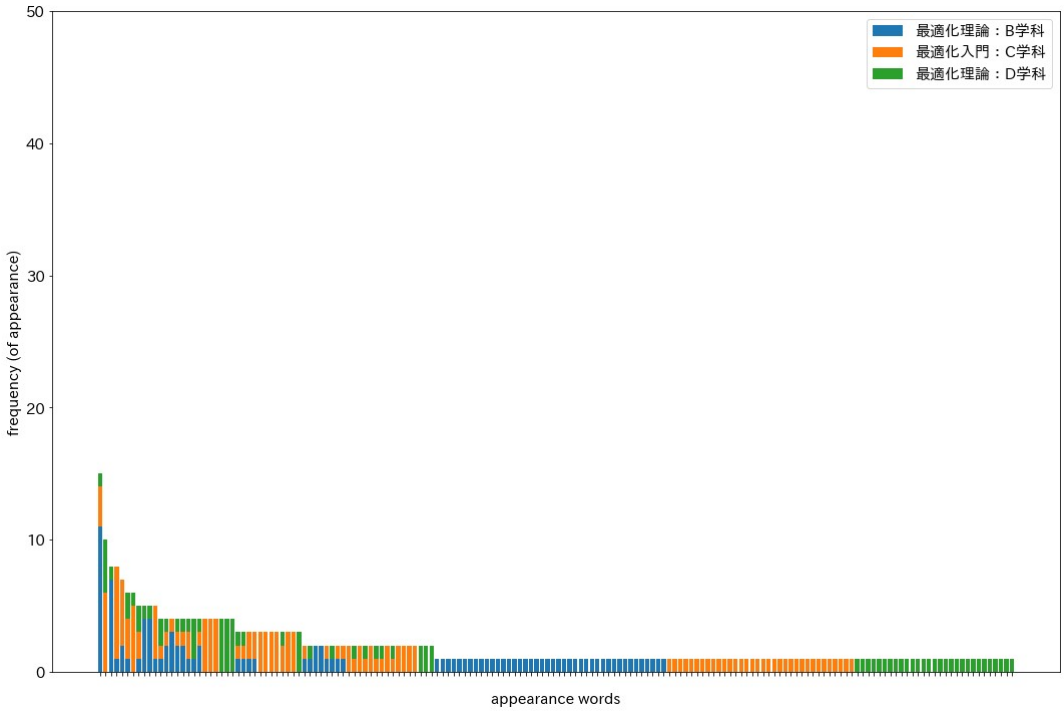

図 4.3: 出現頻度による科目概念の推定（最適化）

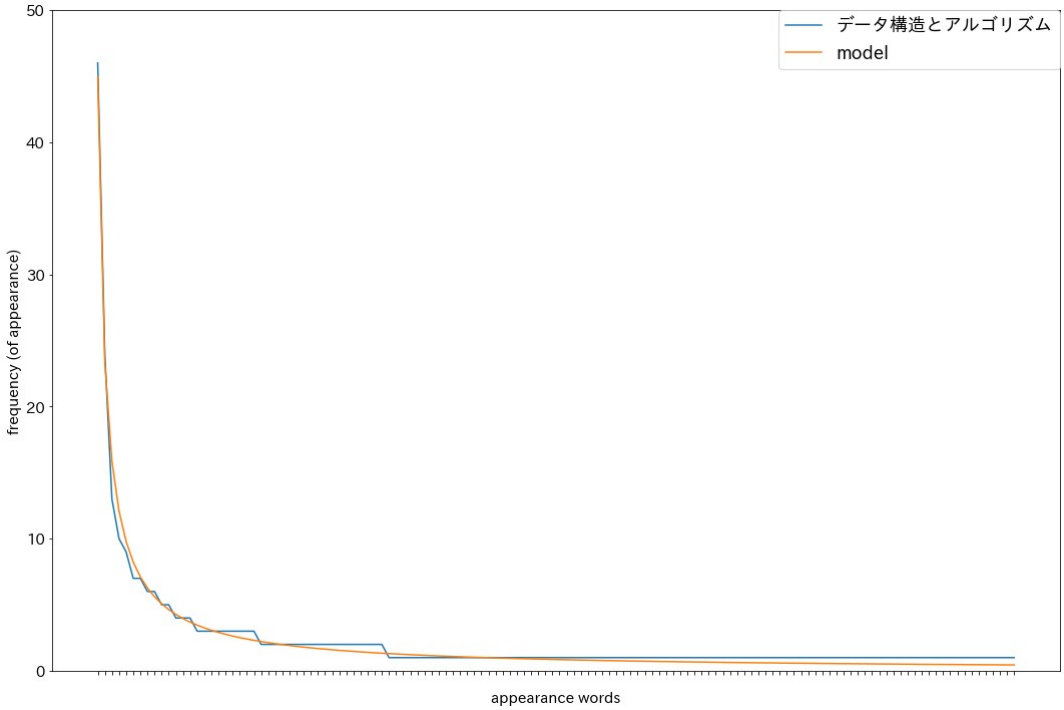

図 4.4: 近似曲線による科目概念の推定（データ構造とアルゴリズム）

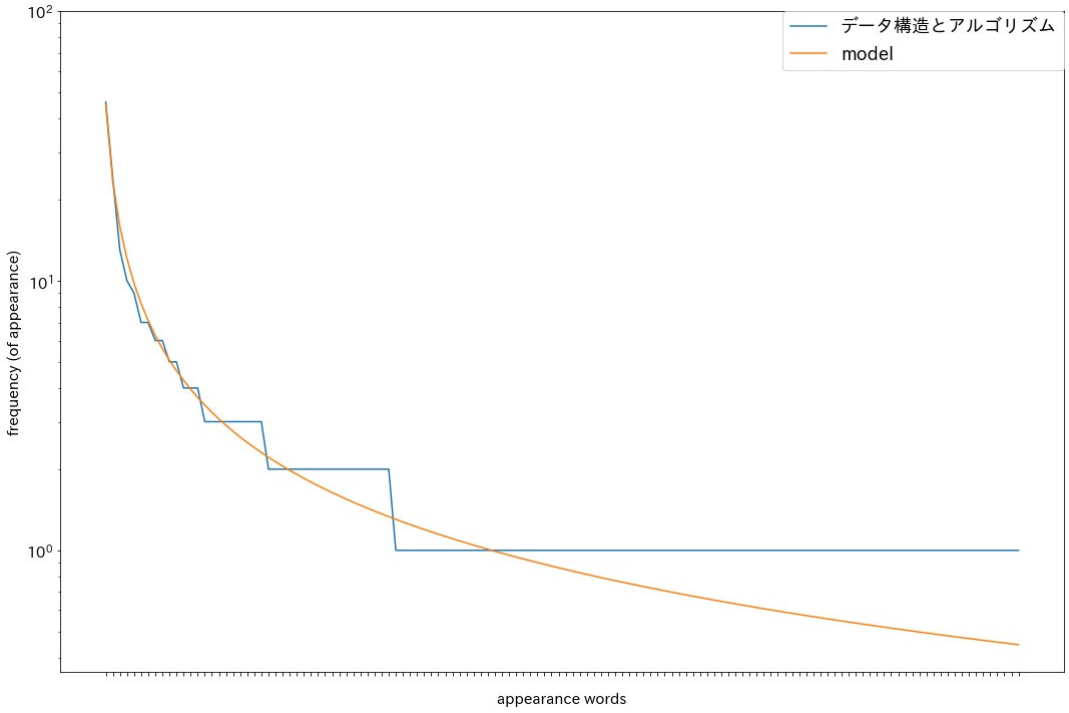

図 4.5: 片対数近似曲線による科目概念の推定（データ構造とアルゴリズム）

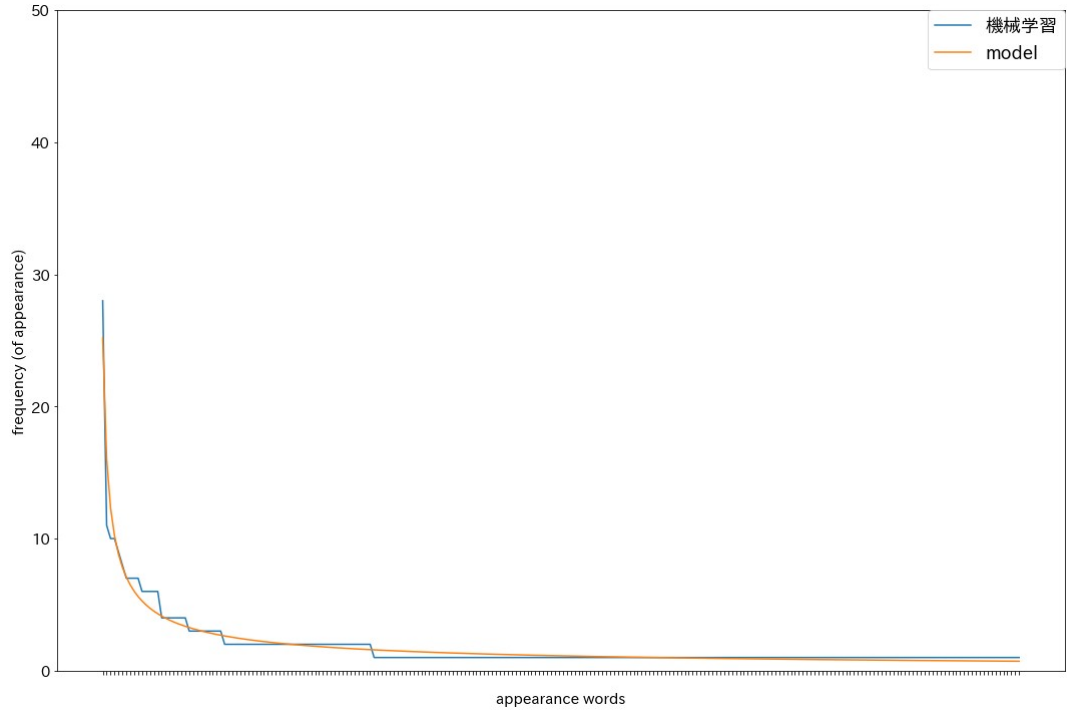

図 4.6: 近似曲線による科目概念の推定（機械学習）

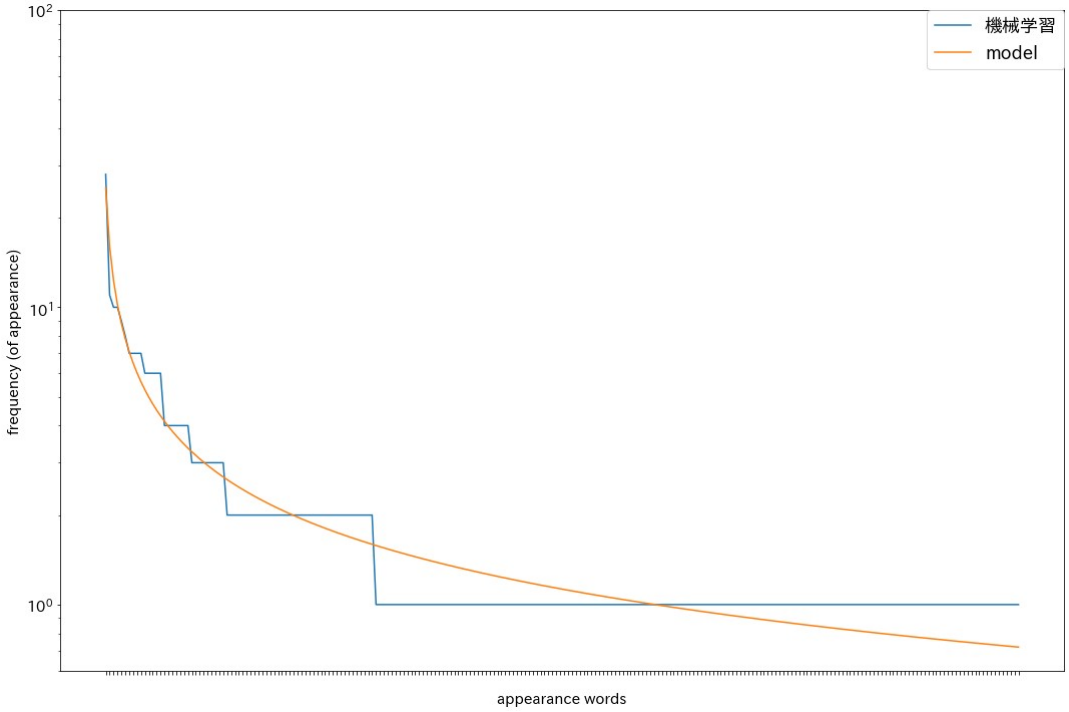

図 4.7: 片対数近似曲線による科目概念の推定（機械学習）

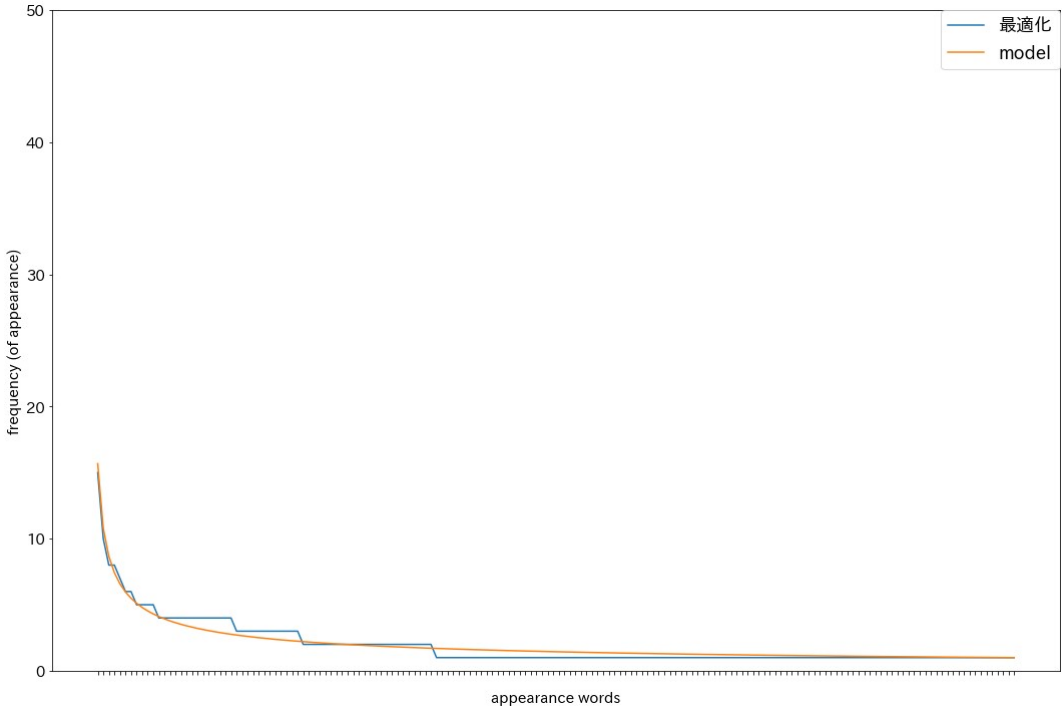

図 4.8: 近似曲線による科目概念の推定（最適化）

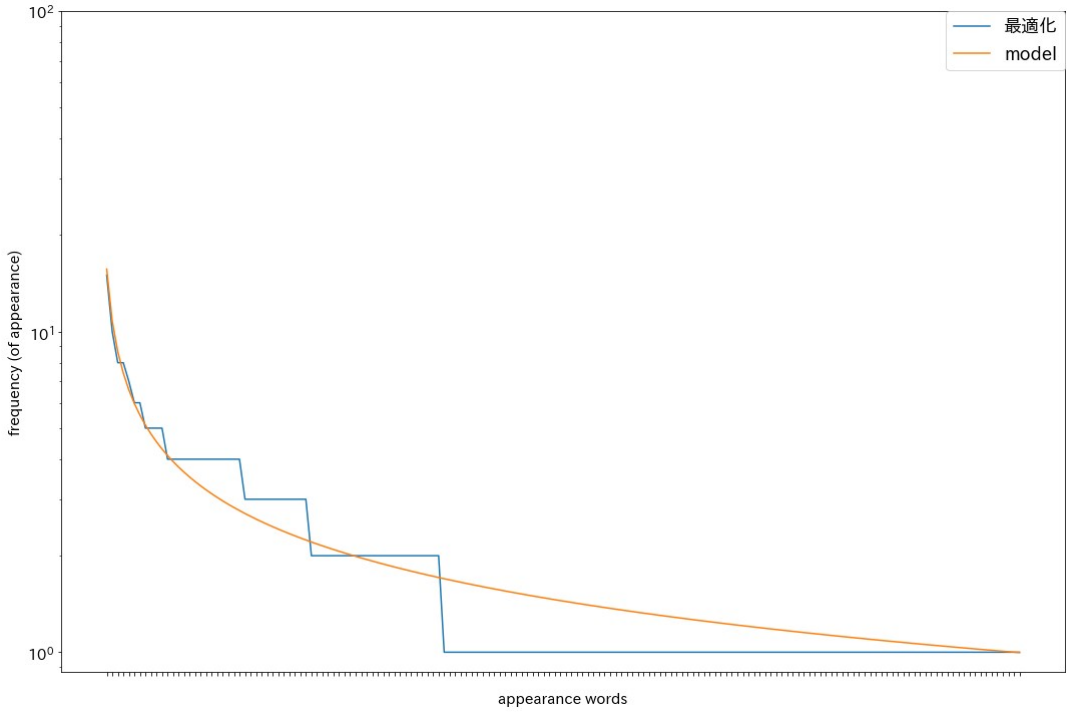

図 4.9: 片対数近似曲線による科目概念の推定（最適化）

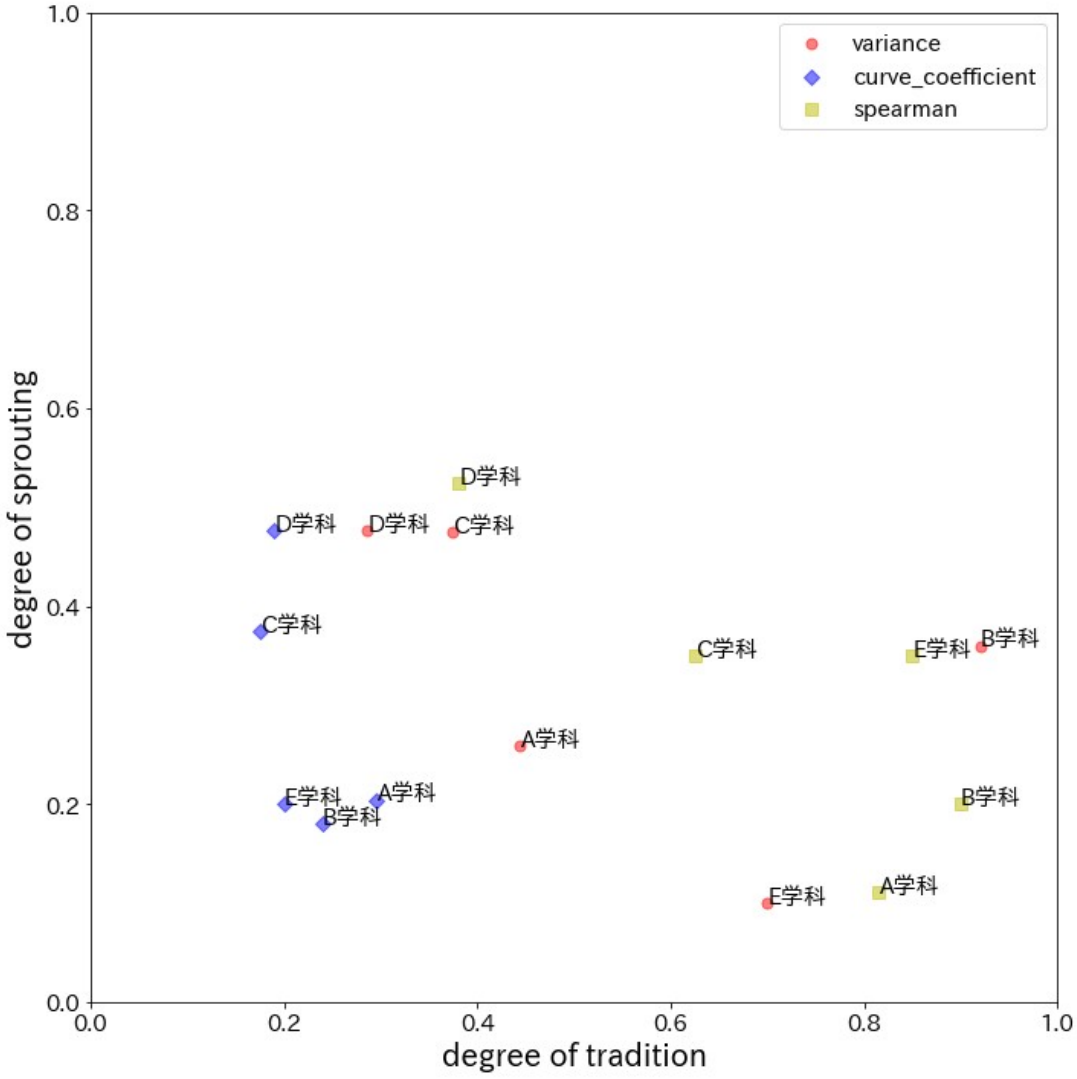

図 4.10: 伝統的度合いおよび萌芽的度合い（類似科目名群）

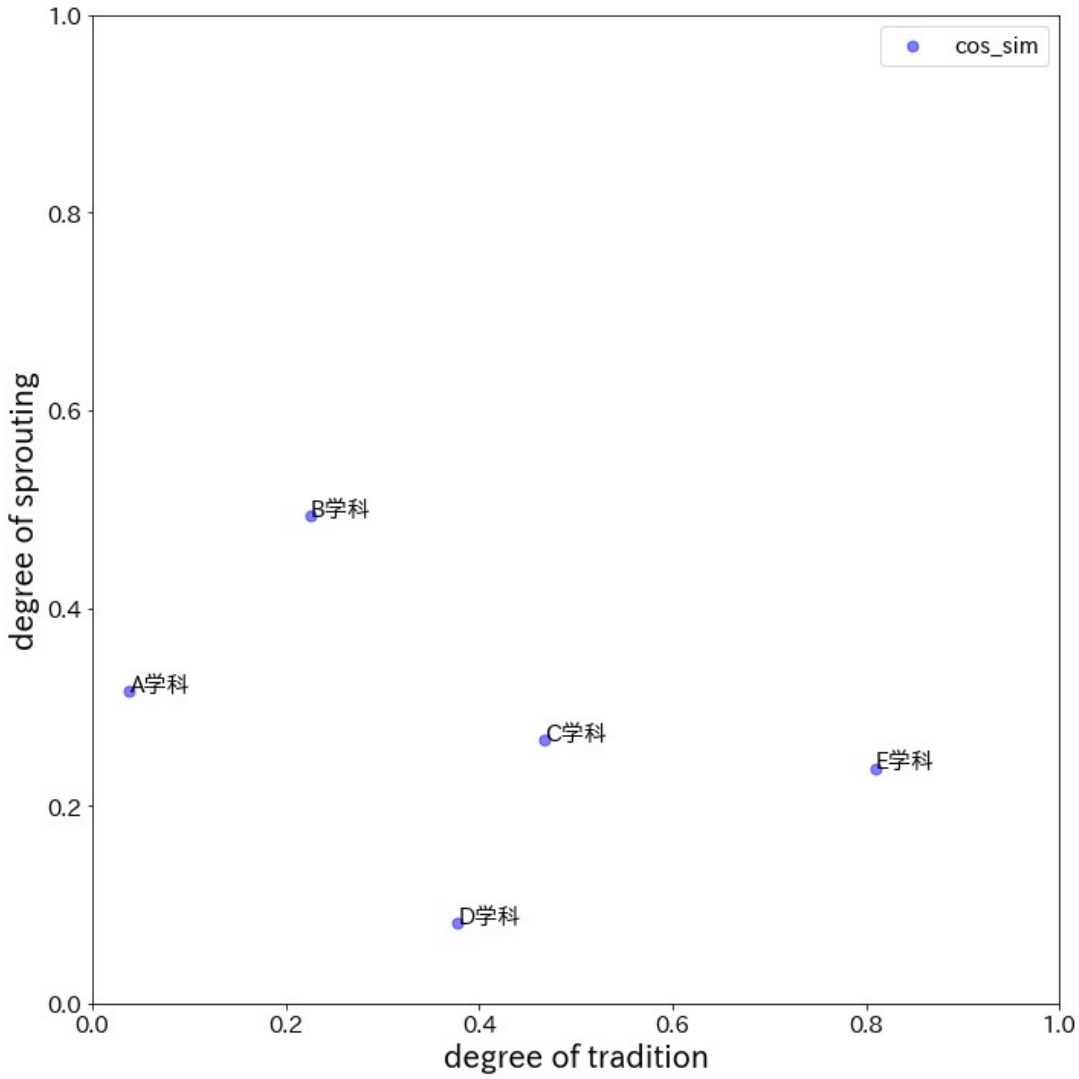

図 4.11: 伝統的度合いおよび萌芽的度合い（孤立科目群）

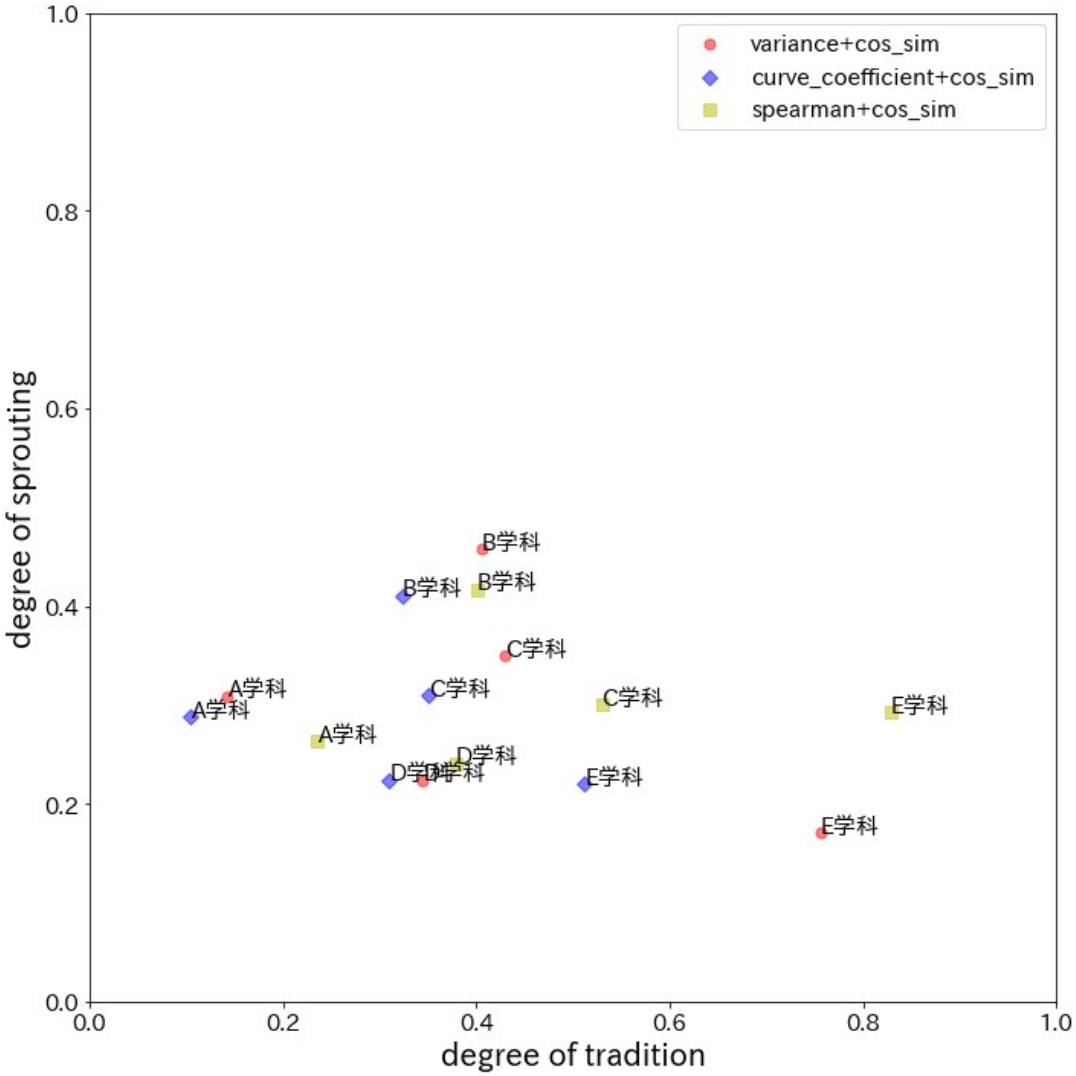

図 4.12: 3 手法と科目の類似度を組み合わせた伝統的度合いおよび萌芽的度合い

第5章　考察

5.1　科目概念の推定に関する考察

5.1.1　代表的な科目群の科目概念の推定

ここでは例として取り上げた「データ構造とアルゴリズム」,「機械学習」,「最適化」の3科目群について考察する．**表 4.2** より,「データ構造とアルゴリズム」および「最適化」は3科目で構成された科目群であり,「機械学習」は4科目で構成された科目群であるため出現単語数に差があるが，1科目当たりの出現単語数で各科目群を比較すると大きな差はないことが言える．

表 4.4 より，各手法で算出された値Vに対して，$1.5 \leq V$であれば$T_{score}2$，$0.5 \leq V < 1.5$であれば$T_{score}1$，$-1.5 < V \leq -0.5$であれば$S_{score}1$，$V \leq -1.5$であれば$S_{score}2$を科目に対して付与している．よって「データ構造とアルゴリズム」は全ての手法において伝統的科目と分類されている．また「機械学習」は頻度分散において萌芽的科目と分類されているが，順位相関係数において伝統的科目と分類されている.「最適化」は頻度分散および近似曲線係数において萌芽的科目と分類されている．

出現単語の頻度分散において伝統的科目と分類された「データ構造とアルゴリズム」は，**図 4.1** より，出現頻度が10回以上の単語数は4単語である．一方で出現頻度が4回以上の単語数は15単語である．これに対して，出現単語の頻度分散において萌芽的科目と分類された「最適化」は，**図 4.3** より，出現頻度が10回以上の単語は2単語である．一方で4回以上出現している単語は25単語である.「最適化」では,「データ構造とアルゴリズム」と比べて，10回以上出現した単語数は減少したが，4回以上出現した単語数は増加した．これらのことから，科目を重ね合わせた科目群の頻度分散について，特定の単語が多く出現すると頻度分散が高くなる傾向がある．また，科目を重ね合わせた結果，科目内容の傾向が異なり，科目内容を特徴づけるような出現単語が科目ごとに違っていると，頻度分散が小さくなる傾向がみられた．一方で**図 4.2** より,「機械学習」は特定の単語が多く出現しているが萌芽的科目と分類された．これは特に多く出現した特定の単語が1種類であり，出現頻度1の単語が多く出現したため，頻度平均に対して偏差の値が大きくならなかったからであると推測できる．

近似曲線の係数による科目の分類では**図 4.4**，**図 4.5** より,「データ構造とアルゴリズム」を伝統的科目に分類している．**式 3.3** および**図 3.3** より，指数部aはaの値が小さくなるほどyの値が大きくなり，aの値が大きくなるほどyの値が小さくなる傾向にある.「データ構造とアルゴリズム」では，特定の単語の出現回数が極端に多かったのに合わせて，出現頻度1の単語も多く出現していたため，aの値が小さくなったと考えられる．**図 4.4**，**図 4.6** を比較すると,「データ構造とアルゴリズム」は，$x = 0$により漸近している．以上より，近似曲線の係数による科目の分類では「データ構造とアルゴリズム」が伝統的科目に分類されたと推測できる．一方で「最適化」は萌芽的科目に分類されている．**図 4.8**，**図 4.9** より,「最適化」では特定の単語が頻出せず，多くの単語が複数回出現した．また，出現頻度1

の単語の全体に占める割合が他の科目群より小さかった．これにより,「最適化」では a の値が大きくなったと考えられる．よって「最適化」は萌芽的科目に分類された．

Spearman の順位相関係数による科目の分類では**表 4.4** より,「データ構造とアルゴリズム」および「機械学習」を伝統的科目と分類している．**図 4.1**，**図 4.2** より，これらの科目に出現する単語は，出現頻度の高い単語において，単語間の順位が似ている傾向が見られた．一方で**図 4.3** より,「最適化」に出現した単語には特定の学科にのみ出現した単語が存在したため，他の 2 科目と比べて相関係数が低く算出されたと推測できる．また,「機械学習」では出現頻度 1 の単語が多く出現したが，これらはそれぞれ出現頻度を 0 とした単語との順位差を算出しているため，相関係数に負の影響を与えなかった．そのため,「機械学習」は Spearman の順位相関係数において伝統的科目に分類されたと推測できる．

5.1.2 特異な科目概念の推定

各手法において最も特異な科目概念が推定された科目群について考察する．各手法において最大となる T_{score} をとった科目群と最大となる S_{score} をとった科目群を**表 5.1** に示す．

表 5.1: 各手法において最も特異な科目概念が推定された科目群

	頻度分散	近似曲線係数	順位相関係数
最も伝統的な科目群	プログラミング	計算機概論	プログラミング
最も萌芽的な科目群	代数学	調査法	アルゴリズム論

また，これらの科目群の出現単語に関する数値データを**表 5.2** に，出現単語の頻度分散，近似曲線の係数，Spearman の順位相関係数の平均をそれぞれ標準化した結果を**表 5.3** に示す．

表 5.2: 最も特異な科目概念が推定された科目群の出現単語に関する数値データ

	出現単語数	1 科目あたりの出現単語数	出現単語の頻度平均
プログラミング	1557	103.8	4.065
計算機概論	243	121.5	1.676
代数学	154	77.0	1.656
調査法	345	115.0	2.091
アルゴリズム論	162	81.0	1.841

出現単語の頻度分散および Spearman の順位相関で最も伝統的な科目群と推定された「プログラミング」は，15 の科目で構成されている科目群である．そのため，出現単語の頻度平均が他の科目群と比べて高く算出されたと推定される．これにより，頻出した単語だけでなく，出現頻度が 1 の単語に対しても偏差が大きく計算された．よって，出現単語の頻度分散が大きくなったと考えられる．また,「プログラミング」に属する各科目間のすべての Spearman の順位相関係数は 0.9 以上であった．そのため,「プログラミング」はどの学科においても，同様の科目内容でシラバスが構成されていると結論付けることができる．

近似曲線の係数で最も伝統的な科目群と推定された科目群は「調査法」であった．この科目群は,「量的調査法」,「質的調査法」,「標本調査法」の 3 科目で構成されている．取り扱

表 5.3: 最も特異な科目概念が推定された科目群の 3 手法の算出値（標準化）

	頻度分散	近似曲線係数	順位相関係数
プログラミング	**3.970**	0.476	**1.753**
調査法	0.709	**3.333**	0.452
代数学	**−0.963**	−1.313	−1.493
計算機概論	−0.958	**−1.813**	−0.865
アルゴリズム論	−0.592	1.203	**−1.541**

う科目内容は 3 科目とも「調査法」で一致しているが，科目名から主要な対象がそれぞれの科目で異なっていることが窺える.「調査法」では,「調査」という単語が最も多く出現し，これは 2 番目に多く出現した単語と比べて約 5 倍も多く出現した．最も出現した単語の出現頻度と 2 番目以降の出現頻度の差による影響により，a の値が最も小さくなったと推測できる．これにより,「調査法」は近似曲線の係数において伝統的な科目群に分類された.

「代数学」は頻度分散において最も萌芽的な科目に分類された.「代数学」は，1 科目あたりの出現単語数が全科目群中で最も少なく 77 だった．また，出現単語の頻度平均も全科目群中で最も低い値が算出された．そのため，出現頻度のばらつきに大きな差が出ずに，頻度分散において最も萌芽的な科目と推定されたと推測できる.

近似曲線の係数で最も萌芽的な科目と推定されて科目群は「計算機概論」である.「計算機概論」は，単語の出現頻度のばらつきが少なく，出現頻度 1 の単語の全体に占める割合が極めて大きかったため，a の値が最も大きくなった．これにより，近似曲線の係数において萌芽的な科目群に分類されたと考えられる.

Spearman の順位相関で最も萌芽的な科目群と推定された科目群は「アルゴリズム論」であり，この科目群は 2 科目で構成されている．算出された Spearman の順位相関係数は −0.019 であり，これらの科目間に相関はないと言える.「アルゴリズム論」といった授業の性質上，様々なアルゴリズムを取り扱うと考えられる．そのため，科目間に順位相関がないと推定された．また，1 科目あたりの出現単語数が少ないことも少なからず Spearman の順位相関係数に影響を与えていると推測できる.

「アルゴリズム論」が Spearman の順位相関において萌芽的科目と分類された一方で**表 4.4** より,「データ構造とアルゴリズム」が Spearman の順位相関において伝統的科目と分類されている．似た科目名を持つ科目において，片方が伝統的科目，片方が萌芽的科目に分類されたことは，カリキュラムの全体像を明らかにするといった点で，今まで考慮されていない新たな観点を持つ分類結果であると言える.

5.2 学科の特異性に関する考察

5.2.1 各手法における学科の特異性

類似科目名群に属する科目について，学科ごとに分析する．**図 4.10** より学科ごとの傾向を分析すると，A 学科および B 学科は各手法において萌芽的度合いよりも伝統的度合いが高く算出されている．特に B 学科は，頻度分散および Spearman の順位相関において，伝統的度合いが極めて高く算出されている．これは，類似科目名群に属した B 学科の科目中にプログラミングと名を冠する科目が 8 科目存在していることに起因している．これらの

科目は，頻度分散および Spearman の順位相関係数において，T_{score} が 2 付与されている．そのため，この 2 手法で伝統的度合いが高い傾向を示した．また同様の理由で，E 学科において，科目中にプログラミングと名を冠する科目が 6 科目存在しているため，頻度分散および Spearman の順位相関係数の 2 手法で伝統的度合いが高い傾向を示した．C 学科は，全体的に中央に分布しており，萌芽的度合いはどの手法においても似た値を示しているが，伝統的度合いは手法によって異なった値を示している．これは，近似曲線の係数の伝統的度合いが全体的に低く算出されてている傾向と，Spearman の順位相関の伝統的度合いが全体的に高く算出されている傾向に則している．D 学科も同様にして，3 手法の伝統的度合いに関する傾向に従っている．

孤立科目群に属する科目について，学科ごとに分析する．**図 4.11** より，C 学科，D 学科，E 学科では，萌芽的度合いと比べて伝統的度合いが高い傾向がみられた．一方で，A 学科および B 学科では，伝統的度合いと比べて萌芽的度合いが高い傾向がみられた．特に A 学科の伝統的度合いは低く，A 学科の孤立科目 79 科目中 3 科目にのみ伝統的スコアが付与されている．次に 71 科目と孤立科目が多かった B 学科も，伝統的度合いが低くなっている．特に A 学科と B 学科は，分析に用いた学科ごとの全科目数に占める孤立科目が多かった．そのため，後述する 3 手法と科目類似度の組み合わせにおいて，A 学科および B 学科の分布は科目の類似度に大きく影響を受けている．

3 手法と科目の類似度をそれぞれ合わせた科目について，学科ごとに分析する．**図 4.12** より，各手法と科目の類似度を合わせた提案手法では，すべての学科が手法ごとに比較的近い距離で分布している．これは類似科目名群に用いたそれぞれの手法と科目類似度の算出手法を組み合わせることで，いずれの組み合わせにおいても同様の結果が得られることを示している．**図 4.10**，**図 4.11** では，学科ごとに散らばった分布を示していたが，**図 4.12** では，各手法の分布において似たような傾向を示した．**図 4.12** において学科ごとに似た伝統的度合いおよび萌芽的度合いの傾向を示しているのにもかかわらず，**図 4.10** において 3 手法が学科ごとに散らばっているのは，類似科目名群に属する科目が全科目数と比べて少数であったからであると推測できる．これらの各手法の比較評価より，提案手法は学科の特異性を明らかにし，各科目の科目概念を推定することのできる手法であると言える．

5.2.2 学科の特異性の比較

4.1 節に示した各学科の特徴と学科の特異性を比較する．A 学科は文理を問わない広い領域での基礎理論や基盤技術の獲得を目標としており，これは学科によって異なる内容を習得するといった萌芽的科目が多く出現した特徴と一致している．よって A 学科は萌芽的な特異性を持った，幅の広い領域を習得することのできる学科であると言える．

B 学科および C 学科は，データサイエンスの領域を始めとした幅広い学問領域をおさえたカリキュラムを構成している．B 学科と C 学科を比較すると，B 学科の方がやや萌芽的度合いが高く算出されており，C 学科の方がやや伝統的度合いが高く算出されている．これにより，B 学科はより萌芽的な科目が多く，C 学科はより伝統的な科目が多く存在していることが分かる．よって B 学科は情報学の分野に留まらず，データサイエンスを含んだ広い学問領域を展開しており，様々な領域の科目を履修することができる学科であると推測できる．一方で C 学科は，幅広いカリキュラムで科目を展開しつつも，情報学を深く学ぶことのできる学科であるといえる．以上より，同じデータサイエンスの名を冠した学科においても，展開されている科目領域が異なっていると言える．

D 学科および E 学科は，高度な知識とスキルを兼ね備えた人材の育成を目指しており，情報工学における深い分野を習得することができるカリキュラムを構成している．D 学科と比べて E 学科は，萌芽的度合いは大きく変わらないものの，伝統的度合いは大きく差が開き，高く算出されている．これは，E 学科がより多くの伝統的な科目で構成されていることを示している．すなわち，E 学科は多くの普遍的な内容を持つ科目でカリキュラムが構成されており，情報工学に特化した科目を多く履修することができる学科であると推測できる．よってこの学科は，学科の課程を終えることで情報工学に関する知識や経験を深く習得できる学科であり，これは E 学科の学科の特異性といえる．一方で，これらの学科は伝統的要素の強い学科であるが，萌芽的な特異性も少なからず兼ね備えている学科であることがわかる．

5.3 提案手法の応用例に関する考察

本研究の応用例として，科目名と科目内容の乖離度の推定が挙げられる．科目内容に対して適切な科目名を名付けることは，学生をはじめとした講義科目に関わる全ての人にとって重要である．科目名と科目内容に齟齬が生じると，科目に対する直感的な理解やイメージが失われるだけでなく，カリキュラム全体に対して一貫性が消失する可能性がある．

本研究では，科目概念を推定することで科目を伝統的科目と萌芽的科目に分類した．また，科目名に着目して科目を分類することで同名科目における科目内容の違いを示した．科目名と科目内容の乖離度の推定では，これらの分析に加えて同様の科目内容を持つ科目について，科目名がどのようにつけられているのかといった傾向を分析する．これにより同名科目における科目内容の違いと，同一科目内容における科目名の違いを示すことができるため，科目名と科目内容の乖離度を明らかにすることができると考えられる．

第6章　まとめ

本研究では，学科の特異性を明らかにすることを目的とした，科目概念の推定手法を提案した．提案手法は類似した科目名を持つ科目内容の重ね合わせに特徴がある．類似科目名群には，出現単語の頻度分散，近似曲線の係数，Spearman の順位相関の 3 つの指標を用いることで科目概念を推定した．また，孤立科目群に対して，科目名でクラスタリングし，生成されたクラスタ内で科目内容の類似度を算出することで科目概念を推定した．推定された科目概念を学科ごとに累積させることで，学科の特異性を明らかにした．

今後の課題は次のとおりである．まず，科目分類の閾値となる NLD の検討および改善である．本研究では，NLD の閾値を 0.5 に設定し科目を類似科目名群と孤立科目群に分類することで分析を行った．しかし似た科目名であるが，全く異なる科目内容をもつ科目を同一の科目群として取り扱ってしまう場合が考えられる．また，本研究の NLD による科目分類手法では，科目名の末尾に数字やアルファベットを付与してクラス分けを行っている同一の科目内容を持つ科目が複数存在した場合，その中から 1 科目のみを抽出して分析している．その一方で，科目名の末尾に数字やアルファベットを付与して異なる科目内容を構成している科目は同一の科目群に分類している．以上を考慮しながら，適切な NLD の閾値を実験的に検証することで，より適切な科目分類を行うことができると考えられる．次に，定量的な評価実験の導入である．本研究では，3 種類の手法を比較することで提案手法の有効性を明らかにした．この比較評価に加えて定量的な評価を追加して行うことで，提案手法の有効性がより一層担保されると考えられる．

謝辞

本研究を進めるにあたり，常に親身な指導，助言をしてくださった佐藤哲司教授，高久雅夫准教授にこの場を借りて，深く感謝をし，厚く御礼を申し上げます.

また，研究を進めていく中で，日常の議論を通じて多くの知識や示唆をいただいたコンテンツ工学研究室の皆様方に深く感謝いたします.

参考文献

[1] 中央教育審議会. 「学士課程教育の構築に向けて」（答申）, 2008.

[2] 田中圭介. 兵庫教育大学修士課程のカリキュラム構造の可視化の試み-シラバスのテキストマイニング. 兵庫教育大学研究紀要, No. 47, pp. 143–151, 2015.

[3] 宮原道子. 研究ノート テキストマイニングを用いたシラバス分析の探索的研究. 大阪観光大学研究論集, No. 21, pp. 95–104, 2021.

[4] 中村修也, 赤倉貴子. 東京理科大学の学部・学科間シラバス分析. 工学教育研究講演会講演論文集 第 66 回年次大会 (平成 30 年度), pp. 240–241. 公益社団法人 日本工学教育協会, 2018.

[5] 永嶋浩. WebAPI を用いた埼学のシラバス分析. 埼玉学園大学紀要. 人間学部篇, Vol. 13, pp. 143–153, 2013.

[6] 石井和也. 地方国立大学における 「地域」 に関する共通教育科目のシラバス分析. 宇都宮大学地域デザイン科学部研究紀要, No. 4, pp. 95–106, 2018.

[7] 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. 情報処理学会論文誌データベース (TOD), Vol. 12, No. 4, pp. 1–14, 2019.

[8] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *International conference on machine learning*, pp. 1188–1196. PMLR, 2014.

[9] 下司義寛, 三輪眞木子, 神門典子, 廣川佐千男. シラバスデータを使った分野ごとの概念マップの生成. 情報処理学会第 68 回全国大会, Vol. 6, p. 5, 2006.

[10] 米田和人, 白井靖人. シラバス情報の Linked Open Data 化とカリキュラム分析への応用. 第 76 回全国大会講演論文集, Vol. 2014, No. 1, pp. 475–476, 2014.

[11] 桑村昭. 日本の大学国際化–課題と展望. 紀要. 地域研究・国際学編, Vol. 45, pp. 191–215, 2013.

[12] Milton A Fuentes, David G Zelaya, and Joshua W Madsen. Rethinking the course syllabus: Considerations for promoting equity, diversity, and inclusion. *Teaching of Psychology*, Vol. 48, No. 1, pp. 69–79, 2021.

[13] 野澤孝之, 井田正明, 芳鐘冬樹, 宮崎和光, 喜多一. シラバスの文書クラスタリングに基づくカリキュラム分析システムの構築. 情報処理学会論文誌, Vol. 46, No. 1, pp. 289–300, 2005.

[14] 金城悟. 保育者養成課程における 「保育内容 (人間関係)」「幼児と人間関係」のシラバス構成に向けた基礎的研究 (1) 授業計画の分析. 東京家政大学教員養成教育推進室年報, Vol. 4, pp. 65–71, 2017.

[15] 金城悟. 保育者養成課程における 『保育内容 (人間関係)』『幼児と人間関係』のシラバス構成に向けた基礎的研究 (2) テキストマイニングによるシラバス分析. 東京家政大学教員養成教育推進室年報, Vol. 5, No. 1, pp. 65–74, 2018.

[16] 齋藤聖子, 中畝菜穂子, 三田地真実. 学習成果可視型シラバス作成支援システムの開発: 学習成果の可視化への試み. 大学評価・学位研究, Vol. 11, pp. 47–61, 2010.

[17] 増田勝也. 日本十進分類を用いたカリキュラム比較のための講義自動分類. 研究報告教育学習支援情報システム (CLE), Vol. 2015, No. 16, pp. 1–5, 2015.

[18] 和多太樹, 関隆宏, 田中省作, 廣川佐千男. 単語の出現頻度に着目した病院評判情報の分析. 情報処理学会研究報告自然言語処理 (NL), Vol. 2005, No. 50 (2005-NL-167), pp. 15–20, 2005.

[19] 中川裕志, 湯本紘彰, 森辰則. 出現頻度と連接頻度に基づく専門用語抽出. 自然言語処理, Vol. 10, No. 1, pp. 27–45, 2003.

[20] 柿本雄輝, 毛利元昭, 打矢隆弘, 船瀬新王, 内匠逸. 単語の出現頻度に基づくテキストの話題分割とラベリング. 第 82 回全国大会講演論文集, Vol. 2020, No. 1, pp. 509–510, 2020.

[21] 柳本豪一. 単語の分散表現を利用した文書類似度. 人工知能学会全国大会論文集 第 29 回全国大会 (2015), pp. 4K11–4K11. 一般社団法人 人工知能学会, 2015.

[22] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[23] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, Vol. 40, No. 2, pp. 99–121, 2000.

[24] 新濱遼大, 槇原絵里奈, 小野景子, 幾島直哉, 山川蒼平. オンラインジャッジシステムにおける問題文の類似度調査. 研究報告ソフトウェア工学 (SE), Vol. 2021, No. 9, pp. 1–7, 2021.

[25] 内田諭. 単語分散表現におけるパラメーター変化の影響: word2vec を用いた事例研究. 統計数理研究所共同研究リポート, Vol. 413, pp. 31–42, 2019.

[26] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き辞書 mecab-ipadic-neologd の実装と情報検索における効果的な使用方法の検討. 言語処理学会第 23 回年次大会 (NLP2017), pp. NLP2017–B6–1. 言語処理学会, 2017.

[27] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き用辞書生成システム neologd の運用 — 文書分類を例にして—. 自然言語処理研究会研究報告, pp. NL–229–15. 情報処理学会, 2016.

[28] Sato Toshinori. Neologism dictionary based on the language resources on the web for mecab. `https://github.com/neologd/mecab-ipadic-neologd`, 2015.

[29] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, Vol. 10, pp. 707–710. Soviet Union, 1966.

[30] Charles E Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, Vol. 15, No. 1, pp. 72–101, 1904.