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

熊田　大雅†　　佐藤　哲司††

† 筑波大学大学院 情報学学位プログラム　〒 305–8550 茨城県つくば市春日 1-2
†† 筑波大学 図書館情報メディア系　〒 305–8550 茨城県つくば市春日 1-2
E-mail: †{kumada,satoh}@ce.slis.tsukuba.ac.jp

あらまし　高等教育機関が公開しているシラバスは，講義概要などの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，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し学科全体の傾向を捉えることには適していない．本研究では，科目間の比較や学科全体の傾向を明らかにすることを目的として，それぞれの科目に対して，多くの大学や学科で類似した内容を教授する「伝統的科目」と，同じ科目名であっても学科ごとに教授内容が異なるとする「萌芽的科目」からなる新たな指標を付与し，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．学科・学類ごとに収集したシラバスのすべての科目に，伝統的科目と萌芽的科目を付与する具体的な手法を考案し，それら科目の構成比率から学科・学類の特異性を明らかにする．提案手法の有効性を検証するために，情報学を標榜している 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顕在化し各科目の科目概念の推定に寄与することを明らかにした．

キーワード　シラバス，科目概念

1　はじめに

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] ．本研究では，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学科の特異性と称する．

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．これに加えて，大学への入学を考えている受験生や編入生も自身の興味関心と志望学科の学問領域および科目展開が一致しているかを調べる際にシラバスの利用が期待できる．受験生や編入生は志望分野に同名の学科や似た名称の学科が複数存在した場合，志望学科を 1 つ決定する必要があるが，シラバスを読むだけでは学科の科目領域を理解することは難しいといえる．これは，現在のシラバスが科目ごとに講義概要や授業計画が記述されているためである．そのため学生は個々のシラバスから学科の全体像を把握し，その特異性を理解することは難しい．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．これに対して，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ことで，学科間の相対的な比較を可能とする．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．提案手法では各科目に対して科目概念を推定する．推定し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する．

2　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．これらの推定された科目概念を学科ごとに可視化することで，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．

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．宮原 [2] 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．抽出語をコレスポンデンス分析し，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 [3] 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ている．シラバスの科目内

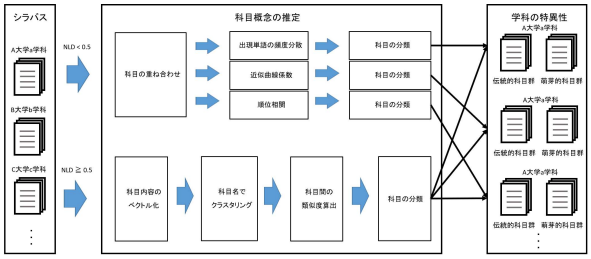

図 1　提案手法の全体構成

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．特定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 [4] の研究がある．石井は，複数学部の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,「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を示している．

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 [5] 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．

以上のように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．

3　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 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 [6] 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

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

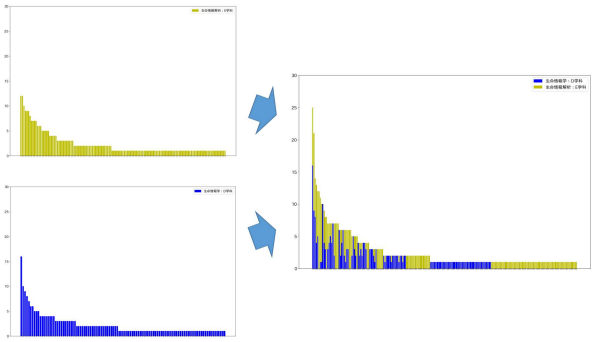

図 2　科目の重ね合わせの例

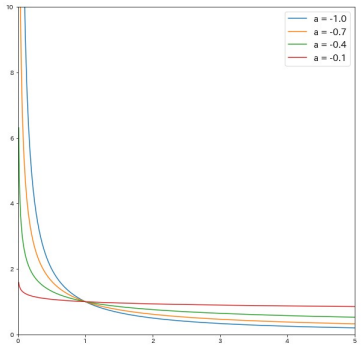

図 3　近似曲線係数の考察に関するグラフ

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 [7] 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．

$$NLD = \frac{LevenshteinDistance(str_1, str_2)}{max(len(str_1), len(str_2))} \tag{1}$$

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．

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例を**図 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．

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

$\bar{x}$ 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．

$$y = bx^a \tag{3}$$

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** に示す．**図 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

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

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** の a に依存しており，**図 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 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 a の値が大きくなると推測できる．

ここで，指数 a を他の尺度と統一した相対振幅にするために指数 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 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 a の値が小さいほど，その科目群が萌芽的科目である度合いが高いとみなす．

3.4.3 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 [8] 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，$-1 \leqq \rho \leqq 1$ をとる．

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．

$$\rho_{ij} = 1 - \frac{6 \sum D^2}{N^3 - N} \tag{4}$$

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．

$$\bar{\rho} = \frac{\sum_{i,j,i<j} \rho_{ij}}{S} \tag{5}$$

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．

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．

$$cos\_sim = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \sqrt{\sum_{i=1}^{n} b_i^2}} \tag{6}$$

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．

標準化後の値 V に対して，$1.5 \leqq V$ であれば，$T_{score}2$ を科目に対して付与する．以下，$0.5 \leqq V < 1.5$ であれば $T_{score}1$，$-1.5 < V \leqq -0.5$ であれば $S_{score}1$，$V \leqq -1.5$ 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．

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

表 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

4　評 価 実 験

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持つ科目は 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 1** に示す．なお，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．

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

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 2** に示す．なお,「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．

科目を重ね合わせた科目群の単語の出現頻度を**図 4**，**図 5**，**図 6** に示す．**図 4** 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 6** 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 7**，**図 8**，**図 9** に示す．なおこれらは，縦軸に底を 10 とした片対数をとっている．**図 7**，**図 8**，**図 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 の軸に漸近している．

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 3** に示す．また，それぞれの値を標準化した結果を**表 4** に示す．

表 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．

次に，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 10** に示す．ここで，T_{score} および S_{score} を科目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．**図 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．

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伝統的度合いと萌芽的度合いの関係を**図 11** に示す．**図 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．

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

表 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

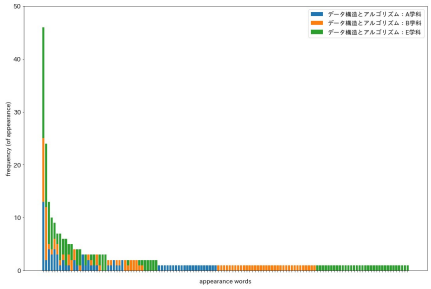

図 4　出現頻度による科目概念の推定
（データ構造とアルゴリズム）

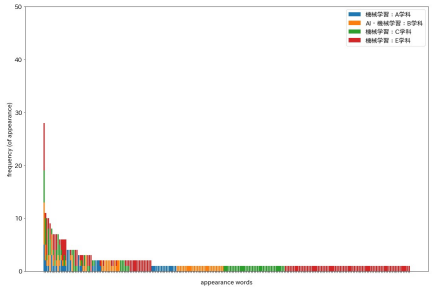

図 5　出現頻度による科目概念の推定
（機械学習）

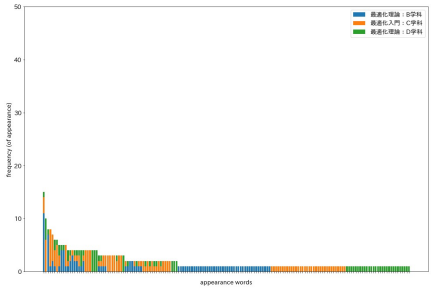

図 6　出現頻度による科目概念の推定
（最適化）

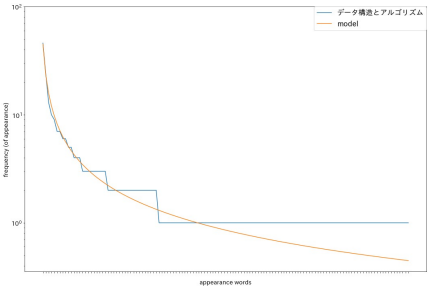

図 7　片対数近似曲線による科目概念の推定
（データ構造とアルゴリズム）

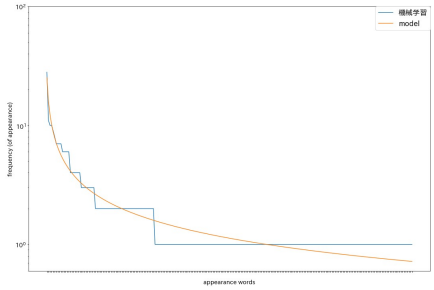

図 8　片対数近似曲線による科目概念の推定
（機械学習）

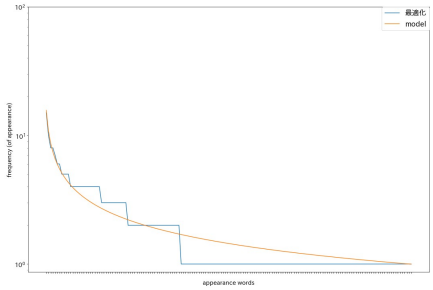

図 9　片対数近似曲線による科目概念の推定
（最適化）

表 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　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

学科ごとに累積させる．各学科の T_{score} と S_{score} を**表 5** 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 12** に示す．

図 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

表 5　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

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．

5　考　　察

5.1　科目概念の推定に関する考察

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．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

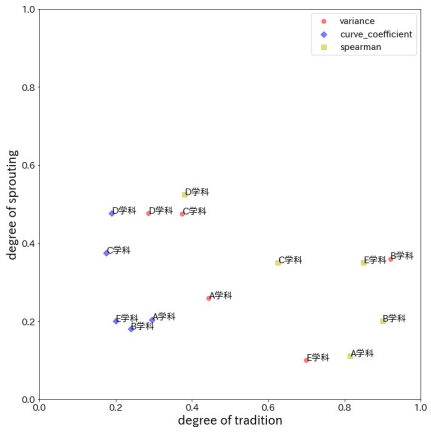

図 10　伝統的度合いおよび萌芽的度合い（類似科目名群）

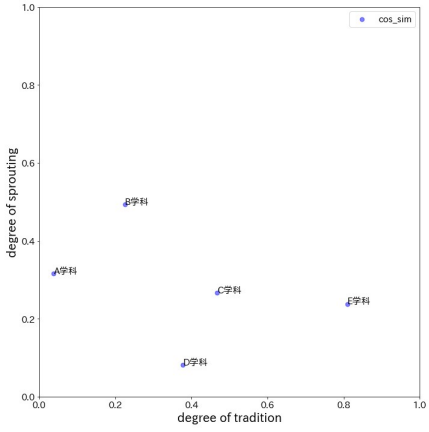

図 11　伝統的度合いおよび萌芽的度合い（孤立科目群）

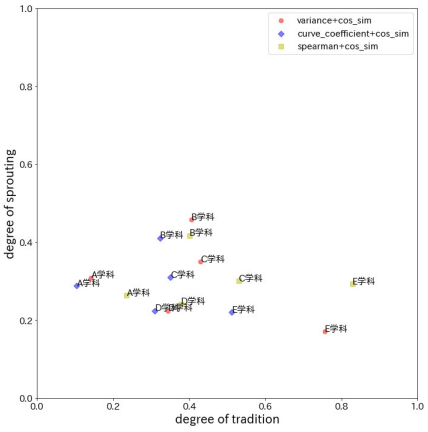

図 12　3 手法と科目の類似度を組み合わせた伝統的度合いおよび萌芽的度合い

アルゴリズム」は，**図 4** 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 6** 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，科目内容を特徴づける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 5** 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 7** 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** および**図 3** より，指数 a は a が小さくなるほど y が大きくなり，a が大きくなるほど y 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 が小さくなったと考えられる．また**図 7** と**図 8**，**図 9** 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ことがわかる．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 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」は指数 a が大きくなり，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** 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**，**図 5** 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 6** 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 とみな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．これにより,「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．

5.2　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 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 学科では，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 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組み合わせた科目について，学科ごとに分析する．**図 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で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 10**，**図 11** では，学科ごとに散らばった分布を示していたが，**図 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 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 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，各科目の科目概念の推定に寄与す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．これは，B 学科の方がより萌芽的な科目が多く，C 学科の方がより伝統的な科目が多く存在していることを示している．これにより，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．

6 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．そこで，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．

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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

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