

LDA と Isomap を用いたカリキュラムの可視化

Visualization of Curriculum on LDA and Isomap

関谷 貴之^{1*} 松田 源立² 山口 和紀²
Takayuki Sekiya¹ Yoshitatsu Matsuda² Kazunori Yamaguchi²

¹ 東京大学情報基盤センター

¹ Information Technology Center, the University of Tokyo

² 東京大学総合文化研究科

² Graduate School of Arts and Sciences, the University of Tokyo

Abstract: A curriculum is crucial to maintaining the integrity of various courses of universities, and an effort to improve a curriculum may include an analysis of the curriculum by comparing the curriculum with previous curricula or curricula of other universities. However, reading syllabi of each curriculum and comparing those curricula is too labor-intensive. We have been developing a method to project syllabi on a map by latent Dirichlet allocation (LDA) and Isomap for the comparison of curricula embodied by the syllabi. In this research, we applied our method to analyzing curricula offered by information science-related departments. The result of analysis of the maps showed that our curriculum maps reflected characteristics of each department.

1 はじめに

情報科学の理論から計算機等のハードウェア、更には各種アプリケーションの応用分野など、大学では「情報系」の学部や学科は多い。しかし、学部や学科の名称は様々で、実際に学科で扱われるカリキュラムの内容也多岐にわたるため、ある学科がどのような教育や研究を行っているのかを知るのは容易でない。

一方我々は、カリキュラムの違いを概観し、これを比較する方法が必要だと考えており、シラバスとして公開されている文書を用いてカリキュラムのマップを作成して可視化する手法を研究している [14]。

そこで我々は、情報系の教育や研究を行う学科のカリキュラムを可視化して、その違いを分析している [15]。まず、国内大学について、情報系の教育や研究を行っているとは推測される学科を選び、各学科のシラバスを Web サイト等から取得した。次に取得したシラバスを用いて、カリキュラムマップを作成した。マップの作成に当っては、情報処理学会によるカリキュラム標準として、コンピュータ科学領域の J07-CS[18] を用いた。その結果、工学の分野で「コンピュータを利用する力を重視する」学科や、理学の分野で「計算機・情報そのものを研究対象とする」学科などの特徴がマップに反映されることが分かった。

本発表では、カリキュラムの可視化方法について述べると共に、情報系学科のカリキュラムの分析結果の一部について報告する。以下、第 2 節では、本研究に関連のあるシラバスを用いた研究やカリキュラム分析の研究を挙げる。第 3 節ではカリキュラムの可視化手法について述べる。第 4 節では、取得した 5 つの学科のカリキュラムの比較結果について述べ、第 5 節で本研究をまとめる。

2 関連研究

高等教育機関において、個々の大学単位でのシラバスのデータベース化やライブラリ化は既に広く実施されているが、単一の大学だけでなく組織を跨がる形で試みも行われている。例えば、Joorabchi らはアイルランド国内の高等教育機関におけるシラバスライブラリの構築について報告している [10]。教員からシステムにアップロードされたシラバス文書を、その内容に応じて分類している。Chatvichienchai は長崎県下の大学における単位互換のために、関連する複数の大学で開講されている講義のシラバスを XML 化した上でこれを管理するリポジトリ構築の試みについて報告している [4]。このような複数組織のシラバスのデータベース化を進めた後には、組織毎の特徴を様々な観点で分析する手法が必要になると考えられる。

カリキュラムを構成する講義間の関係を分析する方法としては、履修者の成績データに対してマイニング

*連絡先：東京大学情報基盤センター
〒153-8902 東京都目黒区駒場 3-8-1 情報教育棟
E-mail: sekiya@ecc.u-tokyo.ac.jp

技術を適用する [17] など様々な手法が考えられるが、講義の内容を要約した情報を含むシラバスを用いることで、講義間の関係をより適切に反映できる。例えば美馬らは、自動的に専門用語を認識してクラスタリングする技術を用いた MIMA search[11] やシラバス検索のための分類システム [12] を開発している。井田らは対話的にカリキュラムを分析するツールを開発している [7, 6]。ここで、複数の組織のカリキュラムを相互に比較する上では、何らかの共通の基準を設定した上で比較することが望ましいが、いずれのツールにも、我々の知る限りはそのような試みは行われていない。一方、本研究ではカリキュラム標準を基準として、複数のカリキュラムを比較可能にする点に特徴がある。

石畑らは J07-CS の知識体系の各ユニットを用いて、理工系の情報学科の授業内容を分析している [8]。分析に当っては、多数の学科のシラバスを取り寄せた上で、各授業がカバーするユニットを複数の専門家が共同で調査している。テキスト処理によって自動的にマップを作る我々の研究は、このような専門家による調査の一助にもなると考えられる。

カリキュラムに限定せず、大量の文書の集合全体を可視化する方法としては、生成モデルに基づいて文書から直接二次元平面にマップするもの [9] がある。しかし、異なる文書集合を同一マップ内で比較するのは困難である。本研究では同一のマップに複数の文書集合を射影して比較できる点に特徴がある。

3 可視化手法

3.1 可視化の要件

本研究では、カリキュラムをシラバスの集合として捉え、そのシラバスを平面上に記号として配置したものをカリキュラムのマップと呼ぶ。そして、カリキュラムの特徴は、このマップ上の記号の分布として視覚的に把握できるものと仮定する。更に、異なる大学のカリキュラムや過去のカリキュラムとの比較を行うことを考えると、複数のカリキュラムを一つのマップの上に載せられることが望ましい。そこで、本研究では、以下の3つの条件を満たす可視化を提案している。

- (A) 与えられたカリキュラムを基準として可視化が行えること
- (B) シラバスの関連度がマップ上の距離として反映されること
- (C) 異なるカリキュラムを同一の基準でマップできること

3.2 LDA

3.1 節で述べた条件 (A) を満たすために、与えられたカリキュラム標準の個々の専門的な領域に合致したト

ピックを LDA を用いて抽出し、そのトピックをカリキュラム分析の基準として用いることにした。LDA とは、確率的文書モデルに基づいて文書の集合から個々の文書の特徴づける潜在的なトピックを抽出する手法である。カリキュラム標準においては、「領域」が予め定義されているが、本研究では領域を LDA の確率的モデルとして再抽出する。これにより、カリキュラム標準に含まれない任意のシラバスが、カリキュラム標準の各領域にどの程度関連するかを計算できるので、各シラバスを共通のトピック空間上に配置することが可能となる。LDA の確率モデルの概要を以下に説明する。詳細は Blei らの論文 [3] を参照して欲しい。

1. 前提として、LDA では文書を確率的に選択された単語の集合 (bag of words) と捉える。そこで、文書 $w = (w_m)(m = 1, \dots, M)$ を、 N 種類の語彙 $v_1, \dots, v_N$ が M 個の単語 $w_1, \dots, w_M$ として現れる多重集合 (同一文書中に単語として同一の語彙が複数回登場することを許す集合) として扱う。すなわち、各 $w_m \in \{v_1, \dots, v_N\}$ である。
2. 最初に、Dirichlet 分布 $\text{Dirichlet}(\alpha)$ に従って、文書の特徴づけるトピックの確率ベクトル $\theta = (\theta_t)(t = 1, \dots, T)$ が決定されると仮定する。ここで、トピックは T 種類、 α は T 次元 Dirichlet パラメータである。 $\sum \theta_t = 1$ である。
3. 次に、文書 w に含まれる m 番目の単語 w_m のトピック $z_m(m = 1, \dots, M)$ が、 θ の多項分布に基づいて決定される。 $z_m \in \{1, \dots, T\}$ である。
4. 最後に単語 w_m が、条件付き多項分布 $p(w_m|z_m, \beta)$ に従って語彙の中から選択される。ここで、 $\beta = (\beta_{nt})$ は $\sum_n \beta_{nt} = 1$ を満たす多項分布パラメータであり、トピック t と語彙 v_n の関連の強さを示す。

文書集合において、個々の文書に含まれる単語の頻度情報を LDA で処理すると、各文書はそのトピックの確率ベクトル θ に基づいてトピック空間内で特徴づけられる。つまり、カリキュラムをシラバスの集合とみなして、LDA で処理する事で、個々のシラバスを θ を用いて T 次元上で特徴づけることができる。

LDA においては、文書 w の集合のみが既知であり、他のパラメータは何らかの基準で決めるか、推定して決める必要がある。推定及び設定手法としては Gibbs サンプルング法 [5] などがあるが、本研究では広く使われている変分ベイズ法 [3] を用いた。

LDA におけるパラメータは、大きく分けて、生成モデルを規定する基準となるパラメータ (α, β) と、ある与えられた文書がどのトピックに関係しているかを示すパラメータ (θ, z_m) の二種類がある。基準となる文

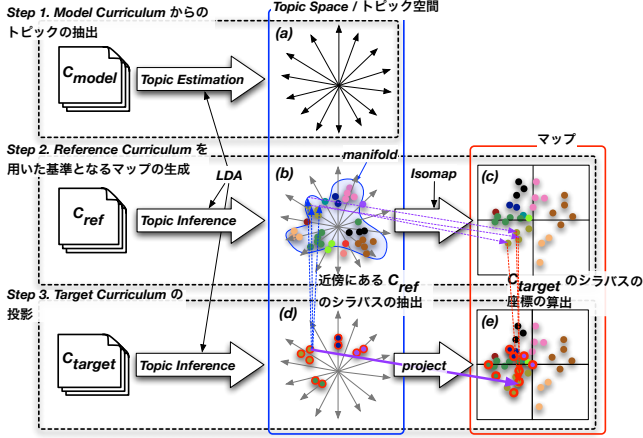


図 1: カリキュラムの可視化の手順の概念図

書集合を元に (α, β) を推定し、それ以外の文書に関しては (α, β) を固定して (θ, z_m) を推定することで、与えられた全ての文書を基準パラメータによって特徴づけることができる。ただし、 (θ, z_m) の推定は膨大な積分計算を必要とするため、一般に困難である。変分ベイズ法では、この推定を近似的に解決している。本稿では、 θ に関してのみ説明する。

θ については、 θ の分布 $P(\theta)$ の推定問題を、 $P(\theta) \simeq \text{Dirichlet}(\theta|\gamma)$ で近似し、近似の最も良くなる T 次元 Dirichlet パラメータ $\gamma = (\gamma_t)$ を推定する問題に帰着させる。 γ は Dirichlet パラメータであり、 θ そのものの値ではないが、Dirichlet 分布の性質により、 θ の期待値 $E(\theta)$ は、合計を正規化した $\bar{\gamma} = \gamma / \sum \gamma_t$ で与えられる。そこで、本研究では、この $\bar{\gamma}$ を、文書のトピック空間での位置と解釈する。ベクトルである $\bar{\gamma}$ のトピック t に対応する座標値が、その文書とトピック t との関連の強さを表す。

本研究では、 $\alpha = (\alpha_t)$ の全ての要素の値は一定値 α であると仮定する。これは、LDA で抽出されるトピックの生成率、つまりトピックに対応するカリキュラム標準の個々の専門的な領域の存在に偏りが無いと仮定することに対応する。J07-CS における与えられた各トピックは代表的なものが偏り無く提案されているので、この仮定は妥当である。

3.3 可視化の手順

本節では、LDA と Isomap を用いたカリキュラムの可視化の手順を説明する。分析には、次の3種類のカリキュラムを用いる。ここでは、カリキュラムはシラバス w_i の集合で表されている。

Target Curriculum, $C_{\text{target}} = \{w_i^{\text{target}}\}$: 本手法を用いて分析するカリキュラム。

Model Curriculum, $C_{\text{model}} = \{w_i^{\text{model}}\}$: C_{target} が対象とする学問分野を広くカバーする標準的、模範

的なカリキュラム。その学問分野のトピックの抽出に用いる。

Reference Curriculum, $C_{\text{ref}} = \{w_i^{\text{ref}}\}$: C_{target} の分析にあたって比較基準となるカリキュラム。

図 1 の可視化の手順の概念図に沿って説明する。

Step 1: Model Curriculum からのトピックの抽出

標準カリキュラム C_{model} を LDA で分析する (Topic Estimation) ことで、分析対象の学問分野におけるトピックの生成モデルのパラメータ β_{model} を推定する (本研究では α は既知としている)。以降の Step では、この β_{model} を固定して、分析対象となるカリキュラムの標準的トピック空間として用いる (図 1(a))。任意のシラバス w_i に同一のトピック生成モデル β_{model} を用いることが出来るので、3.1 節の条件 (A) を満たす。

Step 2: Reference Curriculum を用いた基準となるマップの生成

まず β_{model} を利用して、 C_{ref} の各シラバスの $E(\theta) = \bar{\gamma}(w_i^{\text{ref}})$ を LDA で推定する (Topic Inference)。これにより、 C_{ref} の全シラバスを、基準となるトピック空間上に配置する (図 1(b))。しかし、トピック空間は T 次元空間であり、このままでは、カリキュラムの全体構造を可視化できない。そこで、本研究では、高次元空間内の多様体構造を保持したまま、これを広げるようにして次元を縮退する手法 Isomap[16] を用いる。

$$x(w_i^{\text{ref}}) = \Pi_{\text{Isomap}}(\bar{\gamma}(w_i^{\text{ref}})) \quad (1)$$

但し $\Pi_{\text{Isomap}} : \mathbf{R}^T \rightarrow \mathbf{R}^2$ は Isomap による写像変換。

ここで、 $x(w_i^{\text{ref}})$ は各シラバスの二次元平面上での位置を示す。この結果、3.1 節の条件 (B) を満たすような C_{ref} を二次元平面上に展開したマップが得られる (図 1(c))。Isomap の近傍数 k_{iso} については後述する。

Step 3: Target Curriculum のマップへの射影

まず Step 2 と同様に β_{model} に基づいて、 C_{target} の各シラバスの $\bar{\gamma}(w_i^{\text{target}})$ を LDA で推定する (Topic Inference)。これにより基準トピック空間に C_{target} が配置される (図 1(d))。

次に、各シラバス w_i^{target} を、Step 2 で求めたマップに、以下の式を用いて射影する。

$$x(w_i^{\text{target}}) = \sum_{k=1}^{k_{\text{neigh}}} \eta_k x(w_k^{\text{ref-neigh}}). \quad (2)$$

ここで、 $w_k^{\text{ref-neigh}} \in C_{\text{ref}}$ は、トピック空間上で $\bar{\gamma}(w_i^{\text{target}})$ の近傍となる k_{neigh} 個の比較基準シラバスである。 η_k は $\sum \eta_k = 1$ を満たす加算重みであり、この制約下で $\sum \eta_k \bar{\gamma}(w_k^{\text{ref-neigh}})$ と $\bar{\gamma}(w_i^{\text{target}})$ の距離を最小化することで計算する。この計算は、多様体次元縮退法の一つである LLE[13] で行われるものと同一である。この結

果, C_{target} の全てのシラバスが, Step 2 で求めた C_{ref} のマップと同一の二次元平面上に射影される(図 1(e)). 異なるカリキュラムを同一のマップ上で比較できるので, 3.1 節の条件 (C) を満たす.

4 実験:カリキュラムの可視化

4.1 パラメータの設定

LDA を用いた文書集合のモデル化においては, α の値を 1 より小さくすれば, 各文書は少数のトピックとの関係が強くなり, 1 より大きくすれば多数のトピックとの関係が同程度に強くなる傾向がある. その結果, 本研究のカリキュラムのマップにおいては, α が小さいと同じトピックと関係が強いシラバスがトピック毎に固まって配置され, α が大きいと関係の強いトピックに拘わらずシラバスが一箇所に固まって配置される傾向がある. いずれにしても, シラバス同士の違いや関係を視覚的には把握し辛くなる. このような観点から $\alpha = 1$ とすることが適切であると考えられる.

また, 3.2 節で述べたように, 本研究ではカリキュラム標準における領域を LDA のトピックとして抽出して, これを基準としてトピック空間を形成し, カリキュラムマップを生成することから, トピックの数 T は領域の数と同じ値にすることが適切であると考えられる.

このようなパラメータの設定が妥当か否かを検証するために, Dirichlet パラメータ α とトピック数 T を決める予備実験を行った. 本研究のカリキュラムの分析方法においては, LDA で抽出したトピックと文書との関連の強さが, Isomap を用いて得られるマップ上の文書の座標として適切に表現され, 互いに関係のある文書がマップ上で近くに配置されるかが重要である. そこでこの実験では, 予め分類された文書集合として UCI Machine Learning Repository¹[1] が提供する Twenty Newsgroups Data Set(文書数:2000 個, 分類数:20 個)を用いた. 文書群によって確率モデルが異なるため, 本来はカリキュラム標準である J07-CS 自体をデータセットとして用いるべきである. しかしながら, J07-CS の文書集合のサイズ約 140 個程度では, パラメータの最適値を求めるには十分ではないと考えられる. そこで, 広く用いられている UCI Machine Learning Repository を用いて, あくまで目安ではあるが, 安定的に良い結果を返すと期待される α と T の値を求めている.

3.3 節で述べた LDA と Isomap を使った可視化手法を, パラメータの値を変更しながら適用して, 同じ分類の文書(ネットニュースの記事)がより近くに配置されれば, その時の値が適切だと判断する. 判断指標として, Leave-one-out cross validation²に基づく k 近傍法による分類精度を用いた. 全てのパラメータ値について, 同

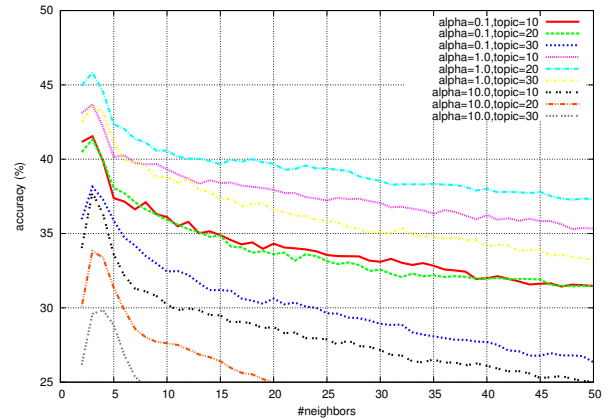


図 2: 異なる k の値に基づく k 近傍の一致率(パラメータ $\alpha = 0.1, 1, 10$, トピック数 $T = 10, 20, 30$)

じ分析処理を 10 回繰り返し, 得られた分類精度の平均値で評価した. $\alpha = \{0.1, 1.0, 10\}$, $T = \{10, 20, 30\}$ の時の精度のグラフを図 2 に示す. この予備実験の結果から, トピック数 T の値を元データの分類数と同じ 20 かつ Dirichlet パラメータ α を 1 とした時の精度が高いことが分かった. そこで, 当初想定した $T =$ カリキュラム標準の領域数及び $\alpha = 1$ を採用することとした.

なお, 3.3 節で説明した分析方法 Step 2 で用いる Isomap では, トピック空間内での近傍数 k_{iso} によって平面上のマップが変化する可能性がある. しかし, 本研究で対象としたカリキュラムを分析した限りでは, ある程度小さい値であれば, その値の違いによってマップ上のシラバスの分布が大きく変わることは無かったため, Reference Curriculum の全てのシラバスを接続可能な(言い換えれば, 全てのシラバス間で有限な距離が定義できる)最も小さい値を用いることとした. 4.2 節で述べる Reference Curriculum では $k_{\text{iso}} = 14$ となった. また, 比較基準シラバス数 k_{neigh} については, 接続可能という条件は必要ないため, 比較的小さい値をとることとし, $k_{\text{neigh}} = 3$ とした.

4.2 Model 及び Reference Curriculum

J07-CS を Model Curriculum として, 情報系学科のカリキュラムを分析するために, Body of Knowledge (BOK)=カリキュラム, 領域=トピック, ユニット=シラバスと解釈した. J07-CS の領域は 15 個, ユニットの総数は 138 個で, 個々の領域のユニットは 4~14 である. ユニットの説明は比較的狭い範囲の事柄を列挙したもので, 同じ領域に属する互いに関係のあるユニットであっても, 共通の単語が無い或いは非常に少なく, LDA で抽出したトピックとの関連の強さが全く異なる場合も有り得る. そこで 3.2 節で述べたように, カリキュラム標準の各領域に合致したトピックを抽出するために, ユニットの説明にそのユニットが属する領域

¹<http://archive.ics.uci.edu/ml/>

²詳しくは, 機械学習の教科書 [2] などを参考にしたい.

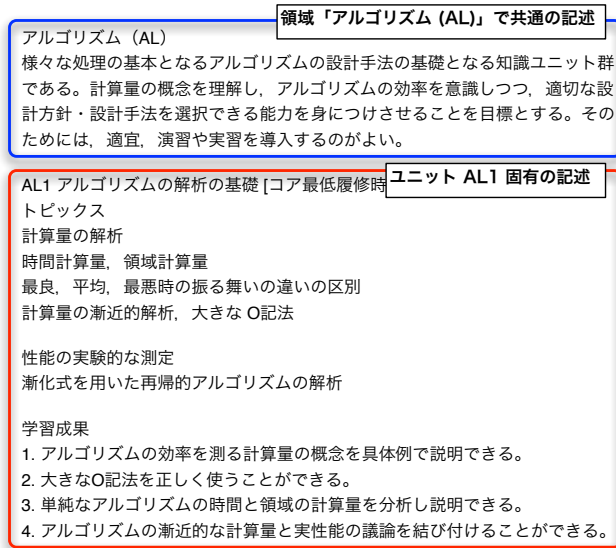


図3: カリキュラム標準 J07-CS のユニットの記述に対する前処理

の記述を加える処理をして、領域を反映したトピックが抽出されやすいようにした。例えば、領域「アルゴリズム (AL)」のユニット「AL1 アルゴリズムの解析の基礎」を処理する際には、領域共通の記述(図3の上部の青い四角に囲まれた部分)とユニット固有の記述(図3の下部の赤い四角に囲まれた部分)を合わせたものを、シラバス相当の文書とした。その結果、同じ領域に属するユニットは、LDA の同じトピックとの関連が強くなり、LDA のトピックと J07-CS の領域とをほぼ一致させることができた。

図1に示すように、本可視化手法において Reference Curriculum はマップ全体の形状や配置の元となることから、分析対象の範囲を広くカバーするカリキュラムが必要である。ここでは、Model Curriculum にも用いた J07-CS の BOK と、情報処理学会が一般情報教育用の教科書として作成した「情報とコンピューティング」³と「情報と社会」⁴を用いた。前者の J07-CS BOK については、Model Curriculum と同様にユニットをシラバス相当の文書としたが、領域の記述をユニットに加える前処理は行わない。後者については、各教科書からテキストを抽出して、図4に示すように、「1.1 符号化の意義」「1.2 符号化の原理と自然数の符号化」のような節の単位で分け、全部で 69 個の節をそれぞれ独立したシラバス文書として扱った。その結果 Reference Curriculum のシラバスの数は合計で 207 個となる。ここで得られた Reference Curriculum のマップを図5-a に示す。

一つのシラバスは様々な内容を含むが、3.2 節で述べ

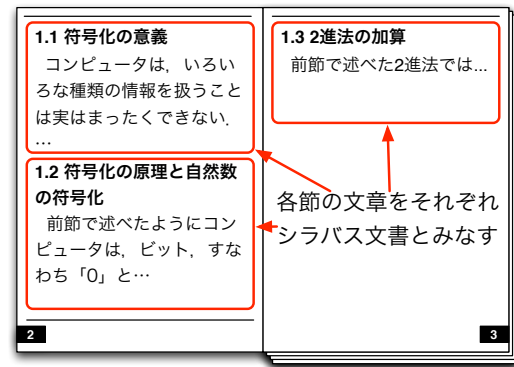


図4: Reference Curriculum 用シラバス文書の抽出方法の概念図

た LDA を用いるとシラバスとトピックとの関係は確率として求められるが、各シラバスに対して最も確率が高いトピックを主たるトピックと呼ぶ。主たるトピックは、そのシラバスの主な内容を示すと考えられる。そこで、表1に示すようにトピックの色を決めて、図5のマップ上では、それぞれのシラバスの主たるトピックに応じた色の丸記号でシラバスを示している。これによって、Reference Curriculum のマップ上で同じ色の丸記号が固まって配置される部分は、表1に示す J07-CS の領域に対応すると考えられる。この丸記号の分布を把握しやすくするために、丸記号の座標値の分散共分散行列から2つの軸を求め、 $2\sigma(\sigma^2 = \text{分散})^5$ の楕円で囲んでいる。

マップは15次元の多様体を Reference Curriculum のシラバスの配置に基づいて2次元に射影しているため、軸の意味は大局的に一定ではないが、図5-aにおいては、左は人や社会に、右は理論に関するものが多い。例えば「SP 社会的視点と情報倫理」の領域は、社会的な視点の内容から左の方に延び、逆に「AL アルゴリズム」や「PL プログラミング言語」の領域は最も右にある。また、「MR マルチメディア表現」は、デジタル表現のような理論的な内容から、国際的な文字コードなどと、理論と社会の両方を含むことから、左右に跨がっていると考えられる。

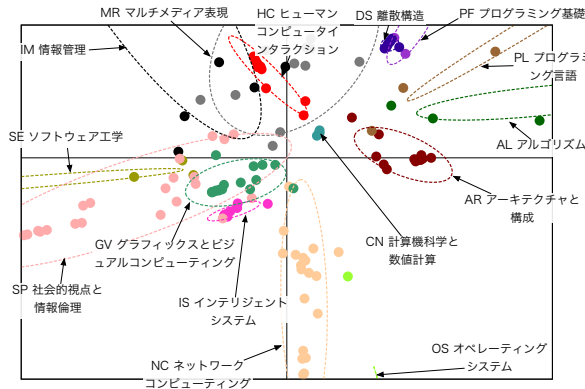
4.3 情報系学科のカリキュラムマップ

本研究では、情報科学や情報工学等の教育・研究を行っている大学の学科を、広く「情報系」の学科として捉え、これらの学科で開講されている講義のシラバスの集合を取得した上で、各学科のカリキュラムを比較分析する。そこで、大学名・学部名・学科名のいずれかに、単語として「情報」「コンピュータ」「ソフト

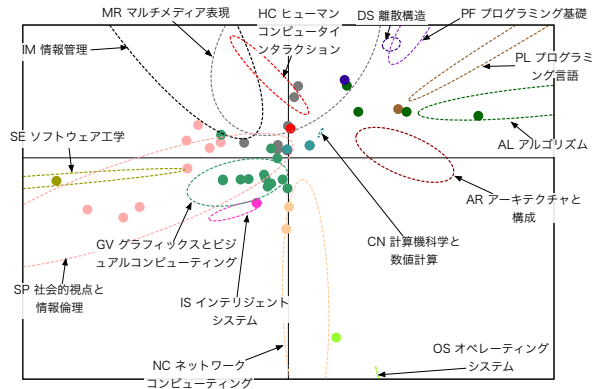
³川合慧監修，河村一樹編著，オーム社，2004 年

⁴川合慧監修，駒谷昇一編著，オーム社，2004 年

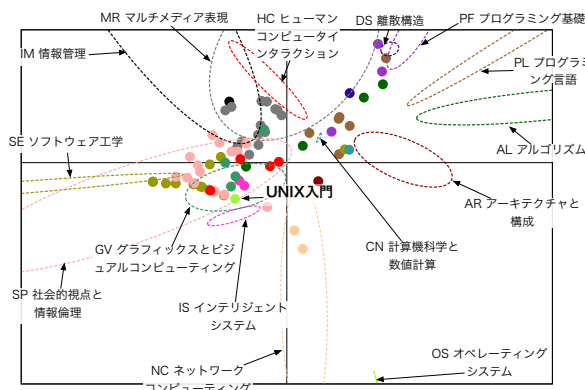
⁵1次元ガウス分布(正規分布)では $\pm 2\sigma$ には 95.4%の要素が含まれるが、2次元ガウス分布では $\pm 2\sigma$ には 86.5%の要素が含まれる。



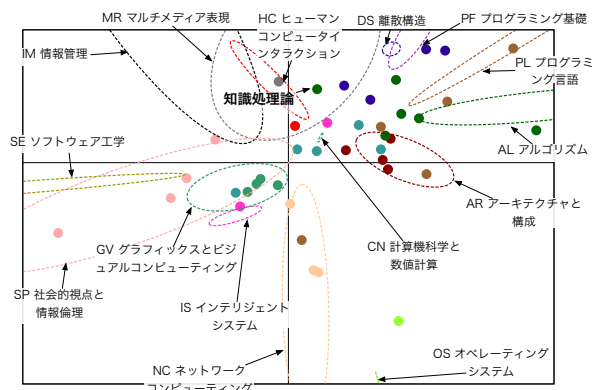
a. J07-CS (Reference Curriculum) のマップ



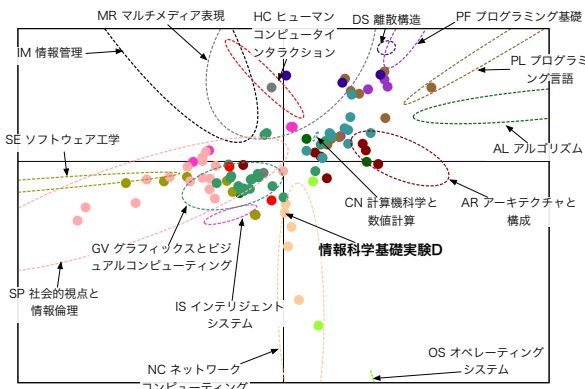
b. A 学科 (国立 工学 その他) のマップ



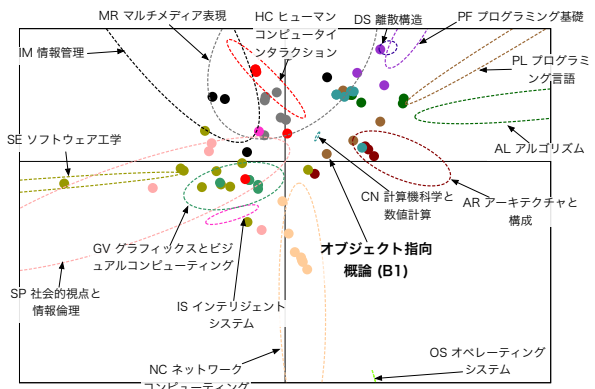
c. B 学科 (私立 工学 その他) のマップ



d. C 学科 (国立 理学 数学関係) のマップ



e. D 学科 (公立 工学 電気通信工学関係) のマップ



f. E 学科 (私立 工学 電気通信工学関係) のマップ

図 5: J07-CS (Reference Curriculum) 及び J07-CS を Reference とする情報系学科のマップ

ウェア」を含む学科を「情報系」の学科とみなし、その中からランダムに学科を9個選択した上で、ツール等を用いて比較的容易にシラバスを取得できた5つの学科を分析対象とした。

表2に5つの学科の概要を示す。A-Eは以後の説明で学科を指すために独自に設定した記号である。分類

は、文部科学省作成の資料に基づいており、対象とした学科が国公立私立のいずれの大学か、またそれぞれの学科の研究及び教育内容がどのように分類されるか(大分類及びその大分類を更に細かく分けた中分類)を示している。学科概要は、各学科のWebサイトで当該学科を説明する文章を抜粋したものである。

表 1: J07-CS の領域とマップ上の色との関係

J07-CS の領域	色
IS インテリジェントシステム	ピンク
PF プログラミングの基礎	紫
AL アルゴリズム	緑
IM 情報管理	黒
PL プログラミング言語	茶
HC ヒューマンコンピュータインタラクション	赤
SP 社会的視点と情報倫理	桃
OS オペレーティングシステム	黄緑
GV グラフィックスとビジュアル・コンピューティング	青緑
NC ネットワークコンピューティング	黄
AR アーキテクチャと構成	茶
SE ソフトウェア工学	黄緑
CN 計算科学と数値計算	青
DS 離散構造	紫
MR マルチメディア表現	灰

表 2: 本研究で分析した情報系の学科
[分類] 及び学科概要

A	[国立 工学 その他] 社会的に新たな価値を創造することのできる想像力と企画力を備えたメディア環境設計の専門家を養成する。
B	[私立 工学 その他] 社会で必要とされるコンピュータを利用する力を重視し、効率的な使い方、便利な使い方を実践的に学ぶ。
C	[国立 理学 数学関係] 情報処理の根本原理の究明・まったく新しい動作原理のコンピュータの設計・コンピュータの新しい使い方の提案というように、計算機・情報そのものを研究対象とする。
D	[公立 工学 電気通信工学関係] システム全体の強調と調和を図る幅広い視野と創造的な技術者、研究者を育成。
E	[私立 工学 電気通信工学関係] 情報ネットワーク技術、情報コミュニケーション技術の先駆的分野を開拓する人材を育成。

図 5 の b-f は、図 5-a の Reference Curriculum を用いて、表 2 の各学科のカリキュラムを Target Curriculum として生成したマップである。

最初に全体的な分析を行う。5 つの学科のマップからは、A 学科、C 学科、D 学科、E 学科において、シラバスが全体的に広く分布しており、J07-CS の領域の上では幅広い内容を扱ったカリキュラムであることが推測

表 3: J07-CS の領域と講義との関係の例

学 科	講義名	関連の強い上位 3 つのトピックと 関連度 (%)
B	UNIX 入門	OS: 14.82 SE: 14.80 PL: 13.05
C	知識処理論	AL: 19.77 IM: 17.22 SE: 14.72

される。A 学科は、全体の中では「PL プログラミング言語」に関わるシラバスが少なく、「GV グラフィックスとビジュアルコンピューティング」や「SP 社会的視点と情報倫理」「MR マルチメディア表現」等に関わる講義が多く、メディア環境設計の専門家を育成するとの特徴を反映している。B 学科は全体的には比較的狭い範囲にシラバスが分布しており、右の理論的な内容や下のアーキテクチャに関する内容は少ない。これは B 学科が「コンピュータの利用」を重視していることが反映されている。C 学科は、「AL アルゴリズム」「PF プログラミング」「AR アーキテクチャと構成」「DS 離散構造」のような、図中では右の方にある領域の内容が多いが、「SE ソフトウェア工学」や「IM 情報管理」のシラバスが少ない。「理学 数学関係」の学科で、情報科学的側面を表している。D 学科はまさに幅広い分野のシラバスを含むし、E 学科も同様である。公立と私立の違いはあるが、いずれも「工学 電気通信工学関係」の学科であることが共通している。

次に一部のマップについて個別的な分析を行う。ここで注目するシラバスを表 3 に示す。B 学科の図 5-c には、最も関連が強いトピックは「OS オペレーティングシステム」だが、OS の領域から大きく外れたシラバスがある。このシラバスは「UNIX 入門」で OS、SE、PL と同程度に関わりが強いので、SE などの領域に近付いたことが分かる。シラバスには「ディレクトリの操作」や「プログラミング入門」などの用語が見られ、オペレーティングシステムの技術的な詳細ではなく、OS の使い方を重視していることが分かる。C 学科の図 5-d には、最も関連が強いトピックは「AL アルゴリズム」だが、AL の領域から外れた中央付近に配置されたシラバスがある。このシラバスは「知識処理論」で、AL の次には IM や SE との関わりが強いので、AL と IM の間に配置されたことが分かる。シラバスには「大規模高次元データから知識抽出を行うための統計的データ解析の理論、および、テキスト処理や知識処理の基本的な技法に関する講義を行う」とあり、純粋なアルゴリズム論ではなくデータ解析など IM の内容があることが分かる。

このようなマップによる全体的及び個別的な分析によって、カリキュラムやシラバスの特徴が分かり、学科概要や分類などの特徴はマップに反映されていると言える。

5 終わりに

本発表では、カリキュラムの可視化方法について述べると共に、情報系の分野の教育や研究を行う5つの学科を対象として、学科のカリキュラムのマップを作成することでカリキュラムの特徴を分析結果について報告した。それぞれの学科のシラバスは、学科のWebサイト等から取得した。マップの作成に当っては、情報処理学会によるカリキュラム標準 コンピュータ科学領域 J07-CS を用い、その結果、工学の分野で「コンピュータを利用する力を重視する」学科や、理学の分野で「計算機・情報そのものを研究対象とする」学科などの特徴がカリキュラムマップに反映されることが分かった。今後は、マップ上にシラバスの特徴を表す専門用語を示すなど、より分かりやすいマップの生成を検討していきたい。

参考文献

- [1] Asuncion, A. and Newman, D.: UCI Machine Learning Repository (2007)
- [2] Bishop, C. M.: *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA (2006)
- [3] Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022 (2003)
- [4] Chatvichienchai, S.: An XML-Based Syllabus Repository System for Inter-university Credit Exchange Systems, in *International Conference on Computer Technology and Development ICCTD 2009*, Vol. 2, pp. 460–464 (2009)
- [5] Griffiths, T. L., Steyvers, M., and Tenenbaum, J. B.: Topics in Semantic Representation, *Psychological Review*, Vol. 114, No. 2, pp. 211–244 (2007)
- [6] Ida, M., Nozawa, T., Yoshikane, F., Miyazaki, K., and Kita, H.: Syllabus database and Web service on higher education, in *7th International Conference on Advanced Communication Technology (ICACT2005)*, Vol. 1, pp. 415–418 (2005)
- [7] 井田 正明, 野澤 孝之, 芳鐘 冬樹, 宮崎 和光, 喜多一: シラバスデータベースシステムの構築と専門教育課程の比較分析への応用, *大学評価・学位研究*, No. 2, pp. 85–97 (2005)
- [8] 石畑 清, 大岩 元, 角田 博保, 清水 謙多郎, 玉井 哲雄, 長崎 等, 中里 秀則, 中谷 多哉子, 疋田 輝雄, 三浦 孝夫, 箕原 辰夫, 和田 耕一, 渡辺 治: 理工系情報学科の授業内容分布のシラバスによる調査 (中間報告), *情報教育シンポジウム SSS2010*, pp. 139–149 (2010)
- [9] 岩田 具治, 山田 武士, 上田 修功: トピックモデルに基づく文書群の可視化, *情報処理学会論文誌*, Vol. 50, No. 6, pp. 1649–1659 (2009)
- [10] Joorabchi, A. and Mahdi, A. E.: An Automated Syllabus Digital Library System for Higher Education in Ireland, *The Electronic Library*, Vol. 27, No. 4, pp. 640–658 (2009)
- [11] Mima, H.: MIMA search: A Structuring Knowledge System towards Innovation for Engineering Education, in *Proceedings of the COLING/ACL on Interactive presentation sessions*, pp. 21–24, Morristown, NJ, USA (2006), Association for Computational Linguistics
- [12] Ota, S. and Mima, H.: Machine Learning-based Syllabus Classification toward Automatic Organization of Issue-oriented Interdisciplinary Curricula, *Procedia - Social and Behavioral Sciences*, Vol. 27, pp. 241–247 (2011)
- [13] Roweis, S. T. and Saul, L. K.: Nonlinear dimensionality reduction by locally linear embedding, *Science*, Vol. 290, No. 5500, pp. 2323–2326 (2000)
- [14] 関谷 貴之, 松田 源立, 山口 和紀: LDA と Isomap を用いた計算機科学関連カリキュラムの分析, *情報処理学会論文誌*, Vol. 54, No. 1, pp. 423–434 (2013)
- [15] 関谷 貴之, 松田 源立, 山口 和紀: 情報系学科のカリキュラムの比較, *情報教育シンポジウム SSS2013*, pp. 33–40 (2013)
- [16] Tenenbaum, J. B., Silva, de V., and Langford, J. C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, Vol. 290, No. 5500, pp. 2319–2323 (2000)
- [17] Tissera, W. M. R., Athauda, R., and Fernando, H. C.: Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining, in *International Conference on Information and Automation, ICIA 2006*, pp. 57–62 (2006)
- [18] 情報処理学会 コンピュータ科学教育委員会: カリキュラム標準 コンピュータ科学 J07-CS 報告書 (2009年1月20日) (2009), http://www.ipsj.or.jp/12kyoiku/J07/20090407/J07_Report-200902/4/J07-CS_report-20090120.pdf