

## カリキュラム標準 J07-CS と CC2001 CS の比較に関する報告

関谷 貴之†, 山口 和紀†

### 概要

我々はシラバスのテキストを分析することで、カリキュラムの特徴や比較を行う方法を研究している。本研究では、計算機科学分野におけるカリキュラム標準として、情報処理学会が提案した「コンピュータ科学カリキュラム標準 J07-CS」と、ACM 及び IEEE Computer Society が提案し、J07-CS も参考にした“Computing Curricula 2001 Computer Science”の2つを、我々の手法を用いて比較分析した結果を報告する。

## Report on Computational Analysis of Difference between J07-CS and CC2001 CS

Takayuki Sekiyama and Kazunori Yamaguchi†

### Abstract

Several curriculum models are proposed as a standard curriculum model of computer science. Information Processing Society of Japan (IPSJ) published J07-CS in 2009. ACM and IEEE Computer Society published Computing Curricula 2001 Computer Science (CC2001CS) in 2001. J07-CS is related to CC2001CS because IPSJ used CC2001CS as a reference. We are studying a curriculum analysis method that uses latent Dirichlet allocation (LDA) and Isomap. To find out that how much J07-CS is similar in content to CC2001CS, we applied our method, and verified that the method is effective.

## 1 はじめに

平成 22 年 6 月 15 日に公布された学校教育法施行規則等の一部を改正する省令(平成 22 年文部科学省令第 15 号)によって、平成 23 年 4 月 1 日より大学等の高等教育機関はシラバスや年間授業計画の概要などを公表することが求められるようになった。従って、カリキュラムを分かりやすく比較可能な形で公開することが以前にも増して重要になっている。

限られた時間の中で学習内容を取捨選択して、系統的に教育するカリキュラムを設計することは当該分野の専門家にとっても難しい。そこで、関連のある学会が、カリキュラムの模範となるカリキュラムモデルを提案することがある。

例えば、ACM と IEEE は計算機科学分野にかかわるカリキュラムの模範的なモデルを古くから提案している。例えば、2001 年には Computing Curricula 2001 Computer Science (CC2001CS)[9] を公開しており、2008 年にも CC2001CS を改訂した Computer Science Curriculum 2008 (CS2008)[8] を発表している。一方、国内では、情報処理学会が CC2001 を参考

にした上で多数の変更と追加を加えた独自のものとして、コンピュータ科学カリキュラム標準 J07-CS[10] を提案している。

情報科学技術の発展の速さを考えると、学会も世の中の変化に追従してなるべく早くカリキュラムのモデルを更新しなければならないが、提案されているカリキュラムモデルは過去のものに比べてどの程度変化しているだろうか。具体的には、CC2001CS を参考に提案された J07-CS との相違点に注目する。

カリキュラムの比較と同様の作業として、文部科学省の学習指導要領の改訂内容の分析が考えられる。例えば中野は高等学校における教科「情報」についてその変化を分析している [3]。分析にあたっては、新旧の指導要領の意味するところや学習項目の順序などを十分に読み解いて分析する必要がある。しかし、分析作業の一部では、例えば特定のキーワードが含まれているか否かなどのテキスト上の機械的な処理も行っている。このような機械的な処理を支援する手法があれば、カリキュラム分析の一助になる筈である。

我々は、講義の内容を説明する材料としてシラバスの文章としての特徴に注目することで、カリキュラムの全体構造を把握するためのマップを作成し、カリキュラムの分析に役立てる方法を研究している [5, 6]。この分析手法は、当該分野における標準的なカリキュラムを用いて、より具体的なカリキュラ

† 東京大学  
The University of Tokyo

ムの特徴を分析するものである。今回我々はこの分析手法を応用することで、CC2001CSとJ07-CSの相違点を分析した。本論文では、カリキュラム標準J07-CSとCC2001CSを比較して、そこから明らかになったことを報告する。

## 2 カリキュラムの分析方法

### 2.1 LDA と Isomap を用いたカリキュラム分析手法

本研究では、カリキュラムやシラバスなどの用語を次の意味で用いる。

**講義** 半年や一年程度を単位として学生に対して実施する教育内容。

**シラバス** 講義で扱われる学習項目や講義の実施方法などを説明する文書。本研究では学習項目を表わす単語を含んだ文章であることに重点を置き、分析にあたっては文書を単語集合 (bag of words) として扱う。

**カリキュラム** 一定の教育目的や修業年限に合わせて総合的に計画された教育内容。本研究では、カリキュラムを講義の集合体として捉え、シラバスを示す文書の集合と見做す。

**トピック** Latent Dirichlet Analysis (LDA)[1]によって抽出される、文書集合を構成する各文書の特徴を示すもの。LDAでは、分野ごとの潜在的なトピックに基づく生成モデルに従って、文書が生成されると仮定する。更に文書と各トピックとの関連の強さを示す値を、文書の座標値と見做すことで、トピック空間内の文書の配置が得られる。

**マップ** シラバスを平面上の点として表現することで、カリキュラムを可視化したもの。シラバスに含まれる単語の数や種類などが、講義内容を示すとの前提の上で、点同士の距離や点の分布から、ある分野に関して類似した内容の講義が多い、この分野の講義は少ないといった、カリキュラム全体の特徴を分析する材料とする。

カリキュラム分析のためのマップは、以下の条件を満たさなければならないと考える。

- 異なるカリキュラムやシラバスを可視化して比較できるように、同じ基底を用いること。
- シラバス同士の関連度が可視化された空間内の距離として反映されること。
- 異なるカリキュラムを比較するために、同一のマップに示すこと。

この条件を満たすカリキュラムの分析手法を以下に説明する。また、その手法の概念図を図1に示す。詳しくは論文[5]を参考にして欲しい。分析にあたっては、次の3種類のカリキュラムを導入する。ここでは、カリキュラムはシラバス  $w_i$  の集合と定義する。ここで  $w_i$  は、シラバスに含まれる単語  $w_1, w_2, \dots$  の出現頻度ベクトルである。

**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,  $C_{\text{model}} = \{w_i^{\text{model}}\}$**  :  $C_{\text{target}}$  が対象とする学問分野を広くカバーする標準的かつ模範的なカリキュラム。前述の条件1を満たす基底となる、当該学問分野のトピックを抽出するために用いる。

**Reference Curriculum,  $C_{\text{ref}} = \{w_i^{\text{ref}}\}$**  :  $C_{\text{target}}$  の分析にあたって比較基準となるカリキュラム。前述の条件3を満たす共通のマップを生成するために用いる。

図1に沿って、分析の手順を説明する。

**Step 1: Model Curriculum からのトピックの抽出**  
標準的なカリキュラム  $C_{\text{model}}$  を LDA で分析する (Topic Estimation) ことで、当該学問分野におけるトピックの生成モデルを推定する。以降の Step では、この生成モデルを固定して、分析対象となるカリキュラムの標準的なトピック空間として用いる (図1(a))。任意のシラバス  $w_i$  に同一のトピック空間を用いることができるので、条件1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 Step 1 の生成モデルを利用して、 $C_{\text{ref}}$  の各シラバスのトピック空間内の位置  $\tilde{\gamma}(w_i^{\text{ref}})$  を LDA で推定する (Topic Inference)。これにより、 $C_{\text{ref}}$  の全シラバスを、基準となるトピック空間に配置する (図1(b))。しかし、トピック空間は  $T$  次元の空間であり、このままでは、カリキュラムの全体構造を把握することが困難である。そこで本研究では、高次元空間内のシラバスの分布がなるべく保たれるように、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である Isomap [7] を用いる。

このようにマップを生成するのは、内容の近いシラバス同士では含まれる単語の種類や数も似ており、その結果、各シラバスを示す LDA のトピック空間内での距離も近いと考えるからである。なお、次元縮退の手法としては、ユークリッド距離に基づく多次元尺度構成法 (MDS) の利用も考えられるが、Isomapの方が同じトピックとの関係が強いシラバス同士が互いに近くに配置されるとの過去の予備的な実験結果から、Isomapを用いている。

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

ここで、 $\Pi_{\text{Isomap}} : R^T \rightarrow R^2$  は Isomap による写像変換を、 $x(w_i^{\text{ref}})$  は各シラバスの二次元平面上

での位置を示す。この結果、条件2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))。

### Step 3: Target Curriculum のマップへの投影

最初に、Step 2 と同様に生成モデルに基づいて、 $C_{target}$  の各シラバスの  $\tilde{w}_i^{target}$  を LDA で推定する (Topic Inference)。これによりトピック空間に  $C_{target}$  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、以下の式を用いて投影する。

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

ここで、 $w_k^{ref-neigh} \in C_{ref}$  は、トピック空間上で  $\tilde{w}_i^{target}$  の近傍となる  $k_{neigh}$  個の比較基準シラバスである。 $\eta_k$  は  $\sum \eta_k = 1$  を満たす加算重みであり、この制約下で  $\sum \eta_k \tilde{w}_k^{ref-neigh}$  と  $\tilde{w}_i^{target}$  の距離を最小化することで計算<sup>†</sup>される。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投影される(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ことで、条件3を満たすことが出来る。

通常カリキュラムでは講義の履修順序などの情報も含まれる。また、個々の講義においても、個別の授業の中である内容を何回目の授業で行うかといった順序も考えられる。しかし、本分析手法では、カリキュラム全体として対象としている学問分野の範囲を示すことを目的としており、順序は考慮しない。従って、文書を単語集合として取り扱って語順の情報は扱えなくても、専門用語のような特徴的な用語を抽出できれば、文書ならびに文書集合の特徴を示すことができると考える。

Step 3 において、Target Curriculum の記述言語が Model Curriculum の言語と異なる場合は、機械翻訳によって Model Curriculum の言語に合わせる。専門用語であれば機械翻訳でもある程度の精度で翻訳が可能で、前述のように特徴的な用語が含まれていれば、文章としての不自然さなどは関係なく処理ができると考えるからである。

## 2.2 カリキュラム標準の比較方法

これまで我々は、ACM と IEEE CS がまとめた Computer Science Curriculum 2008 (CS2008)[8] を Model Curriculum として、2.1 節で述べたカリキュラム分析手法を適用している。2.1 節で定義した本分析手法における各用語と、CS2008 における用語との対応を表1に示す。我々は Unit をシラバスの文書と見なして分析する。LDA のトピックと Area とはそ

表1: 本手法における用語とカリキュラム標準の用語との対応

	カリキュラム標準 (J07-CS/CC2001CS/CS2008)
本手法	BOK
カリキュラム	BOK
トピック	Knowledge Area/エリア
シラバス	Knowledge Unit/ユニット

のままでは一致しないが、Area に関する記述を当該 Area に含まれる個々の Unit の説明に加えることで、同じ Area の Unit であれば、同じ LDA のトピックと関連が強くなるように調整している。これによって、分析対象となる Target Curriculum の個々のシラバスについて、CS2008 の Knowledge Area との関連の強さが得られる。

本手法を応用して、本論文ではカリキュラム標準を比較する。具体的には Knowledge Area やエリア<sup>‡</sup>と呼ばれる学習内容のまとまりに注目して、それらのまとまり同士が似ているか否かを判断する。異なるカリキュラム標準をそれぞれ Model Curriculum として用いて、あるカリキュラムを Target Curriculum として分析するならば、当該シラバスを基準として、カリキュラム標準のエリア 同士を比較できる。

また、トピック空間の基準となるそれぞれのトピックについては、文書に含まれる単語との関連の強さが LDA の処理結果として得られる。ここで、トピックはカリキュラム標準のエリアと一致するように調整していることから、エリアを特徴付ける単語が明らかになる。カリキュラム標準  $C_A$  のエリア  $K_A$  と関係の強い単語の集合と、対応するカリキュラム標準  $C_B$  のエリア  $K_B$  と関係の強い単語の集合は、 $K_A$  と  $K_B$  の内容が近ければ、ほぼ一致することが期待される。つまり、カリキュラム標準の比較手法は以下の通りである。

1. カリキュラム標準  $C_A$  を Model Curriculum として、2.1 節の Step 1 を実施する。これによって、 $C_A$  の各エリアに対応する LDA のトピック、つまり  $C_A$  のエリアに関わる単語の集合を得る。
2. あるカリキュラム  $C$  を Target Curriculum として、2.1 節の Step 3 を実施する。これによって、 $C_A$  のエリアと  $C$  のシラバスとの関係を得る。
3. カリキュラム標準  $C_B$  を Model Curriculum として、また 2 で用いたのと同じカリキュラム  $C$  を Target Curriculum として、上記の 1,2 を実施する。これによって  $C_B$  の各エリアに関わる単語の集合と、 $C_B$  のエリアと  $C$  のシラバスとの関係を得る。

<sup>\*</sup>本研究では、 $k_{neigh} = 3$  とした。

<sup>†</sup>この計算は、多機体次元縮退法の一環である LLE[4] で行われるものと同一である。

<sup>‡</sup>以下では、特に一方のカリキュラム標準について言及する場合を除き、いずれも「エリア」と表記する。

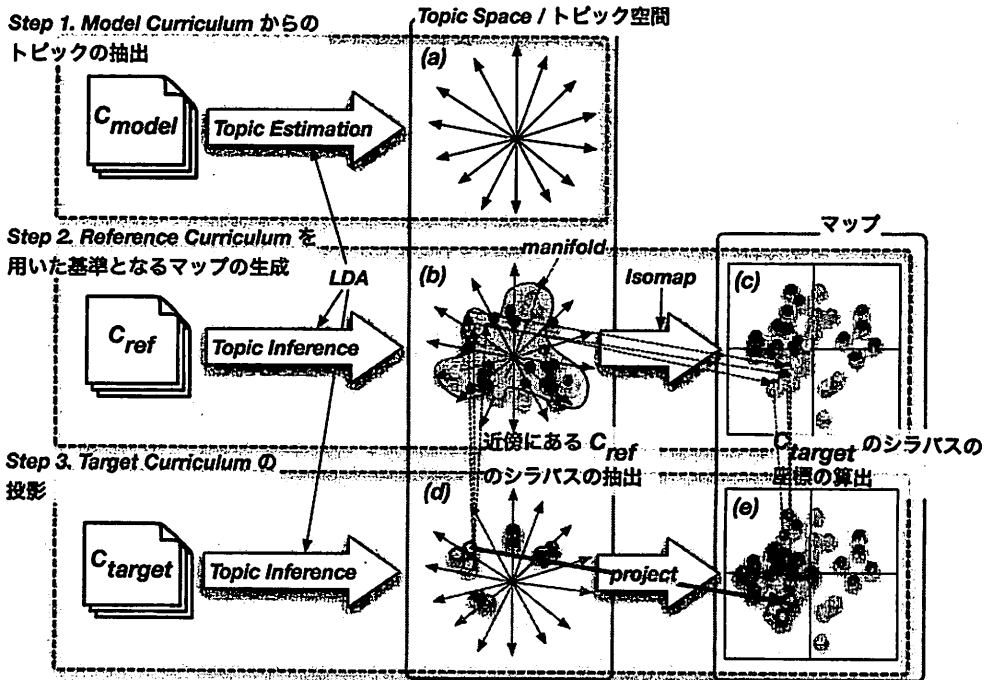


図1: カリキュラム分析手法の概念図

4. 上記1,3の結果から以下の2つについて評価する。

**関連の強い単語に基づくエリアの比較**  $C_A$  と  $C_B$  において、互いに関連があるそれぞれのエリアに対応する単語の集合が一致するか。

**関連の強いシラバスに基づくエリアの比較**  $C$  のシラバスで最も強く関連のあるエリアが、 $C_A$  と  $C_B$  で一致するか。

3節では、2つのカリキュラム標準 J07-CS と CC2001CS を対象として両者を比較する。関連の強い単語に基づくエリアの比較について3.2節で、関連の強いシラバスに基づくエリアの比較について3.3節で述べる。最後に3.4節では、J07-CS と CC2001CS の説明を読んで著者らが判断した結果と本手法による結果との差異について述べる。

### 3 実験

#### 3.1 J07-CS と CC2001CS の比較

CC2001CS は、計算機科学の分野を14個の Knowledge Area に分類した上で、個々の Area には約10個程度の Knowledge Unit を定義している。この Unit が計算機科学分野の大学生が学習すべき内容に相当する。

一方、J07-CS は、それ以前に提案された J97[11] の後継となるカリキュラム標準である。国内の多くの理工系情報学科を想定する一方で、特に米国カリキュラムモデル CC2001CS との整合性を考慮し、かつ日本の科学技術の特長と独自性を活かすべく提案された内容になっている。計算機科学分野を15のエリアに分けた上で、さらにユニットという単位に分ける点は、CC2001CS と同様である。個々のエリアは J07-CS と CC2001CS はほぼ一対一対応するものの、CC2001CS に無い「MR: マルチメディア表現」というエリアが J07-CS では設けられている。エリア毎の対応関係を表2に示す。このようにエリアの対応関係が明らかであることから、2.2節で述べた手法によって、エリアごとの類似性を比較することができる。

なお、J07-CS の分析にあたっては、テキストから単語を抽出する際に形態素解析ツール mecab を用いて、名詞のみを抽出した。形態素解析に用いる辞書は特に変更していない。一方、CC2001CS の分析にあたっては、英文の分かち書きに基づいて単語を抽出し、stop word の除去と stemming を行っている。

#### 3.2 関連の強い単語に基づくエリアの比較

まず、関連の強い単語に基づくエリアの比較結果について評価する。

表 2: J07-CS と CC2001CS との比較

	J07-CS	CC2001CS	J07-CS から見た CC2001CS との違い
DS	離散構造	Discrete Structures	エリアの説明で“set theory,” “graph theory,” “combinatorics” のような専門用語が少ない。CC2001CS DS の 6 つの Unit に加えて, AL の一部 “AL7. Automat theory” が含まれる。
PF	プログラミングの基礎	Programming Fundamentals	ほぼ同内容。
AL	アルゴリズム	Algorithms and Complexity	“Floyd’s Algorithms,” “PRAM (Parallel Random Access Machine)” などの具体的用語が CC2001CS に比べてやや少ない一方で, 「主成分分析」などを説明する「AL9 データ分析アルゴリズム」のような独自のユニットもある。
AR	アーキテクチャと構成	Architecture and Organization	ほぼ同内容。
OS	オペレーティングシステム	Operating Systems	“Scripting (OS12)” が無い一方で, 「分散ファイルシステム」や「セキュアオペレーティングシステム」などの新しい用語が加えられているが, 全体の説明やユニット構成は大体近い。
NC	ネットワークコンピューティング	Net-Centric Computing	全体の説明やユニット構成は大体近い。但し, “Compression and decompression (NC7)” や “Multimedia data technologies (NC8)” のマルチメディア関連の記述が異なる。
PL	プログラミング言語	Programming Languages	「PL8 論理型言語」や「PL9 スクリプト言語」という独自のユニットが含まれる違いがある。
HC	ヒューマンコンピュータインタラクション	Human-Computer Interaction	エリア全体の説明で “interactive software,” や “user interface” などの具体的用語が少ない。個々のユニットの記述はほぼ同じ。
MR	マルチメディア表現	—	
GV	グラフィックスとビジュアル・コンピューティング	Graphics and Visual Computing	全体としては CC2001CS を踏襲しているが, 「メタボール」や「ボクセル」などの用語や, ユニット「GV3 2次元画像の生成と加工」が含まれることが独自。
IS	インテリジェントシステム	Intelligent Systems	「SVM」「PAC 学習」等の用語が含まれるユニット「IS8 機械学習のニューラルネット」が CC2001CS より詳しいが, 全体としては CC2001CS に近い。
IM	情報管理	Information Management	オブジェクト指向モデルの「継承」や「カプセル化」, トランザクションに関する「ACID 特性」のような, 具体的な用語を多く含むが, 全体としては同様の内容。
SP	社会的視点と情報倫理	Social and Professional Issues	ユニットの構成はほぼ同様だが, “International issues” が無い一方で「情報関連法」や「P2P」が加わるなど, 国内事情に合わせて調整されている。また, エリア全体の説明が非常に簡潔。
SE	ソフトウェア工学	Software Engineering	エリア全体の説明が要約されている一方で, 「リファクタリング」や「アジャイル開発」などの用語が追加。ただし, ユニットの構成など全体としてはほぼ同様。
CN	計算科学と数値計算	Computational Science and Numerical Methods	ユニット構成はほぼ同様だが, エリア全体の説明で “Molecular dynamics” など適用範囲の具体例が省略されていたり, 各ユニットで “Markov models and chains,” “Petri nets” などの具体的用語が含まれていない。

表 3: DS: 離散構造 (J07-CS)/Discrete Structure(CC2001CS)に関わる用語の比較

	J07-CS	CC2001CS
1	的	structure
2	数学	discrete
3	形式	computer
4	意味	science
5	記号	theory
6	概念	material
7	具体	set
8	必要	concept
9	論理	graph
10	知識	topic
11	言語	include
12	DS	formal
13	計算	algorithm
14	説明	proof
15	定義	technique
16	性	logic
17	論	example
18	構造	core
19	離散	ability
20	正規	understand

例として、J07-CS のエリア「DS: 離散構造」及び、CC2001CS の Knowledge Area “DS: Discrete Structure” それぞれに強く関係があるとして抽出された上位 20 語を表 3 に示す。

J07-CS の各単語について、CC2001CS の上位 20 語の中に同じ意味を持つと考えられる英単語がある場合はそれを太字で示している。これより J07-CS の単語上位 20 個には CC2001CS の単語 6 個があり、特に J07-CS の上位 10 語のみに絞ると 2 語が含まれている。

同様に他のエリアについても単語の一致を調べた結果と、エリア毎の単語の一致個数の平均を表 4 に示す。数字を太字で示しているのは、全てのエリアの平均よりも一致する語が少ない場合である。あるエリアについて、上位 20 語と 10 語の両方が平均より少ない場合、エリアのタイトルも太字で示している。ここから一致する単語が少ないエリアとして、DS:離散構造, AL:アルゴリズム, HC:ヒューマンコンピュータインタラクション, IM:情報管理, SP:社会的視点と情報倫理は J07-CS と CC2001CS とで内容が異なると考える。

### 3.3 関連の強いシラバスに基づくエリアの比較

国内のある大学の学科の情報系のカリキュラムを Target Curriculum として分析した。2.2 節の分析手法を適用した。

表 4: J07-CS の各エリアの単語集合の内 CC2001CS と一致する個数

エリア	上位 20 語	上位 10 語
<b>DS: 離散構造</b>	6	2
PF: プログラミングの基礎	10	6
<b>AL: アルゴリズム</b>	6	3
AR: アーキテクチャと構成	9	6
OS: オペレーティングシステム	10	6
NC: ネットワークコンピューティング	10	6
PL: プログラミング言語	10	8
<b>HC: ヒューマンコンピュータインタラクション</b>	7	5
MR: マルチメディア表現	-	-
GV: グラフィックスとビジュアル・コンピューティング	11	5
IS: インテリジェントシステム	15	6
<b>IM: 情報管理</b>	9	5
<b>SP: 社会的視点と情報倫理</b>	9	5
SE: ソフトウェア工学	12	7
CN: 計算科学と数値計算	9	6
平均	9.5	5.4

但し、CC2001CS は英語で、分析対象の学科のカリキュラムは日本語で記述されていることから、分析にあたってはカリキュラムの内容を英訳する必要がある。そこで、今回の実験では、Google 翻訳<sup>5</sup>を用いて、学科のカリキュラムを英訳し、その結果を用いて CC2001CS と比較している。表 5 に、各シラバスと最も関係が強いエリアを示す。12 個の内 9 個の講義シラバスで最も関係の強いトピックが一致している。この結果から以下の点が推測される。

- あるシラバスについて、J07-CS と CC2001CS とでエリアが一致しない場合、そのエリアは内容が異なる。→ HC:ヒューマンコンピュータインタラクション, CN:計算科学と数値計算
- あるシラバスについて、J07-CS と CC2001CS とでエリアが一致する場合、そのエリアは内容が類似している。→ DS:離散構造, PF:プログラミングの基礎, OS:オペレーティングシステム, GV:グラフィックスとビジュアル・コンピューティング, SP:社会的視点と情報倫理, SE:ソフトウェア工学

### 3.4 テキストとしての比較

本節では、3.2 節と 3.3 節の分析結果の差異について検討する。

<sup>5</sup><http://translate.google.com/>

表5: 各シラバスと関係が強いエリア

シラバス名	J07-CS	CC2001CS
c01:複合系計画論Ⅱ	SP 0.346	SP 0.250
c02:情報システム科学Ⅰ	DS 0.501	DS 0.258
c03:情報システム科学Ⅱ	SE 0.211	SE 0.273
c04:情報システム科学Ⅲ	OS 0.292	OS 0.264
c05:情報システム科学Ⅳ	HC 0.193	GV 0.179
c06:情報システム科学Ⅴ	GV 0.582	GV 0.701
c07:情報システム科学実習Ⅰ	PF 0.347	PF 0.260
c08:システム基礎科学実験Ⅰ	AL 0.200	IM 0.227
c09:システム基礎数学Ⅲ	DS 0.411	DS 0.232
c10:システム数理演習	CN 0.402	AL 0.187
c11:システム理論Ⅱ	AL 0.231	AL 0.343
c12:システム理論演習	DS 0.522	DS 0.353

まず、表2の最右列には、J07-CSとCC2001CSの各エリア全体を説明する文章と、そのエリアに属するユニットでトピックや学習成果として挙げられている内容とを、著者らが読み取って比較した結果を示している。その結果、全体の説明やユニットの内容が異なるエリアとしては、DS:離散構造、AL:アルゴリズム、PL:プログラミング言語、HC:ヒューマンコンピュータインタラクション、SP:社会的視点と情報倫理、GV:グラフィックスとビジュアル・コンピューティング、CN:計算科学と数値計算が挙げられる。

表6は、J07-CSとCC2001CSとで、対応するエリアがよく一致しているかそれとも異なっているかについて、3.2節と3.3節で示した分析結果と、著者らの判断結果とをまとめたものである。一致すると考えられる場合「○」、異なると考えられる場合は「△」、判断できない場合は「-」としている。

例えばDSは、単語による評価結果よりJ07-CSとCC2001CSとが異なると判断され、テキストを見るとエリア全体の説明やユニットに違いがあることが分かった。CC2001CSDSには無い「文脈自由言語」などを扱うユニットが、J07-CSに含まれており、その結果としてエリアを特徴付ける単語として「言語」などが得られている。

GVは3.3節で用いたTarget Curriculumの講義シラバス「c06:情報システム科学Ⅴ」では、Model CurriculumをJ07-CSとした場合もCC2001CSを

用いた場合もGVに関わる単語が抽出されたことから、J07-CSとCC2001CSに違いが無いと評価したが、テキストで比較すると専門用語が異なっていたり、内容が異なるユニット「GV3次元画像の生成と加工」が含まれていることが分かった。

IMとSPは3.2節の尺度で見ると、単語の一致数がいずれも平均より僅かに劣る程度であるが、テキストを見るとSPではエリアの説明が大きく異なり、対してIMでは用語の違いはあるものの全体として大きな違いは無い。

CNは3.2節の尺度では、単語の一致数がほぼ平均程度で顕著な違いが無いと判断した。しかし、3.3節で用いたTarget Curriculumの講義シラバス「c10:システム数理演習」に最も関係が強いエリアは、Model CurriculumをJ07-CSとするとCN、CC2001CSとするとALとなり、結果が異なった。これはテキストでの比較においてJ07-CSとCC2001CSがやや異なると判断した結果に近い。

単語かシラバスのいずれかかの方法で内容が異なると推測したエリアは、テキストの上でも内容が異なり、2.2節で述べた方法が、カリキュラム標準の相違点の分析において参考にできると考えられる。

一方、次の点については今後更なる検討が必要である。まず、2.2節で述べたトピックをエリアと一致させるために実施した処理では、当該エリアの説明に含まれる用語がエリア全体を特徴付ける用語としても抽出される可能性がある。従って、エリア全体の説明が少ない場合であっても、同一エリア内のユニット同士で関係があることを適切に扱う方法を検討する必要がある。

また、日本語で書かれたJ07-CSから得られた単語も、英語で書かれたCC2001CSから得られた単語もどちらも短いこともあり、個々のユニットの内容を直接指し示す専門的な概念を表す複合語やフレーズは得られていないように見える。特定の一個のシラバスでしか現われない単語では複数のシラバスの比較の基準に用いられず、本分析手法では扱えないが、このように単語を直接示してエリアの評価をするには、専門用語の抽出手法(例えば[2])を適用することも検討した方が良いかも知れない。

次に、3.3節の関連の強いシラバスに基づくエリアの比較では、Target Curriculum全体が対象とする分野や、その中で個々の講義で扱われる内容によって分析結果が変化する可能性がある。今回用いた例(表5)の場合、少なくともAR、NC、PL、ISのエリアと関連の強いシラバスが無いことから、それらのエリアの比較はできない。また、表5の「c05:情報システム科学Ⅳ」のHC(J07-CS)とGV(CCC2001CS)のように、最も関連が強いエリアであっても、他のシラバスと比べると関連度が非常に弱い場合は、そのシラバスに基づくエリアの比較は信頼できない可能性がある。従って、カリキュラム標準と同様の範囲を対象とする他のカリキュラムをTarget Curriculumとして、複数のシラバスに基づいて分析を行うことも考えられる。但し、シラバスの数を増やすと、個々のシラバスの特徴が平均化されて、??節の単語に基

表6: J07-CS と CC2001CS との比較  
 単語 シラバス テキスト  
 (3.2 節) (3.3 節) (3.4 節)

	単語 (3.2 節)	シラバス (3.3 節)	テキスト (3.4 節)
DS	△	○	△
PF	-	○	○
AL	△	-	△
AR	-	-	○
OS	-	○	○
NC	-	-	△
PL	-	-	△
HC	△	△	△
MR	-	-	-
GV	-	○	△
IS	-	-	○
IM	△	-	○
SP	△	○	△
SE	-	○	○
CN	-	△	△

づく分析との違いが無くなる可能性もある。

いずれにしろ、例えば自身が所属する組織のカリキュラムなど、その内容を熟知したカリキュラムを用いることで、シラバスのテキストの分析結果だけでは明らかにならない点を補完したり、逆に予想外の結果が出たことでシラバスの記述を見直すなどの活用も考えられる。

## 4 まとめ

カリキュラム標準 J07-CS と CC2001CS との比較方法として、LDA を用いたカリキュラム分析手法を応用して、その結果がテキストに基づいて比較した結果にある程度一致することを示した。関連の強いシラバスに基づくエリアの比較においては、Target Curriculum に何を用いるかによって結果が変わる可能性があるため、今後他のカリキュラムを用いることで検証を行いたい。

## 参考文献

- [1] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022 (2003).
- [2] 中川裕志, 森 辰則, 湯本紘彰: 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).
- [3] 中野由章: 新しい高等学校学習指導要領 (案) における教科「情報」, *情報処理学会研究報告 2009-CE-98*, pp. 141-148 (2009).
- [4] Roweis, S. T. and Saul, L. K.: Nonlinear dimensionality reduction by locally linear embedding, *Science*, Vol. 290, No. 5500, pp. 2323-2326 (2000).
- [5] Sekiya, T., Matsuda, Y. and Yamaguchi, K.: Analysis of Computer Science Related Curriculum on LDA and Isomap, *ITiCSE'10, Proceedings of the 15th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education*, pp. 48-52 (2010).
- [6] Sekiya, T., Matsuda, Y. and Yamaguchi, K.: Development of a Curriculum Analysis Tool, *ITHET 2010, 9th International Conference on Information Technology Based Higher Education and Training* (2010).
- [7] Tenenbaum, J. B., de Silva, V. and Langford, J. C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, Vol. 290, No. 5500, pp. 2319-2323 (2000).
- [8] The CS2008 Review Taskforce: Computing Science Curriculum 2008: An Interim Revision of CS2001 (2008). <http://www.acm.org/education/curricula/ComputerScience2008.pdf>.
- [9] The Joint Task Force on Computing Curricula IEEE Computer Society/Association for Computing Machinery: Computing Curricula 2001 Computer Science (2001). [http://www.acm.org/education/curric\\_vols/cc2001.pdf/view](http://www.acm.org/education/curric_vols/cc2001.pdf/view).
- [10] 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\\_%xreport-20090120.pdf](http://www.ipsj.or.jp/12kyoiku/J07/20090407/J07_Report-200902/4/J07-CS_%xreport-20090120.pdf).
- [11] 情報処理学会: 大学の理工系学部情報系学科のためのコンピュータサイエンス教育カリキュラム J97 (1999). <http://www.ipsj.or.jp/12kyoiku/J97-v1.1.pdf>.