

4H-6

シラバスデータによる教育課程の分析と視覚化に関する考察

井田 正明[†] 野澤 孝之[†] 芳鐘 冬樹[†] 宮崎 和光[†] 喜多 一[†]
 大学評価・学位授与機構[†] 京都大学 学術情報メディアセンター[†]

1. はじめに

近年における情報ネットワークの整備とともに高等教育においても情報技術の活用が進展しつつある。これまで我々は高等教育における教育課程の分析のためシラバスなど教育課程に関する情報の収集を進めシラバスデータベースを構築してきた。またこのデータベースを活用しシラバス文書に基づいた教育課程の分析システムの構築も進めてきた。本稿においてはより効果的な分析システムの構築のため分析結果の可視化および分析システムの柔軟な活用法について検討する。

2. データベースシステムの構成

現在構築中のデータベースシステムの概略を述べる[1]：①大学など高等教育機関のシラバス等から必要な情報を抽出[2]。②得られた教育課程や授業に関する情報をリレーショナルデータベースに格納（データ構造は XML スキーマにより表現）。③SQL, XPath, Web サービス等をネットワークを通じて提供。④キーワード・専門用語の抽出[3], クラスタリングによる教育課程の分析システムの開発[4]等を進行中。

3. 教育課程の分析手順

開発中のシラバスデータのクラスタリングに基づく教育課程の分析システム[4]における分析手順を述べる。

- (1) 分析対象とする教育課程とシラバスを選択
- (2) 専門用語抽出を適用するデータ項目を選択
- (3) 含まれる専門用語に基づき各シラバス（科目）の内容を縮約・定量化
- (4) シラバス間の類似度を計算
- (5) 類似度に基づきクラスタリングを実施
- (6) クラスタ数を選択（大局および局所情報）
- (7) クラスタ要約の分析と解釈
- (8) コレスポンデンス分析（次元の縮約）
- (9) 3次元表示（アプレットが動作するブラウザにおいてマウスによる3次元回転と拡大縮小）により、近接性（クラスター-クラスター間、クラスター-学科間、学科-学科間）を視覚的に把握
- (10) 分析目的に応じて、(1)-(6)のステップに戻り再分析を実行

4. システム工学系学科の教育課程の分析例

Web 上にシラバスが公開されており学科名が「システム工学」を含んだ（多様な授業科目を有する）13 大学 13 学科を選び、それらの学科が提供している教育課程のシラバス集合（総数 636 科目、2003 年度版）を分析対象とする。システムの各種設定はつぎのとおりである。専門用語抽出の対象項目：科目名, 授業概要, 授業目的, 授業計画, 教科書。専門用語辞書：EDR。用語重み付け関数：TFIDF 値。シラバス間類似度：コサイン類似度。クラスタリング：群平均距離を用いた階層併合的クラスタリング。

まず、クラスタ数 4 でのクラスタ要約（各クラスタの成立に貢献度の高い専門用語のリスト）を Table 1（左列）に示す。また、学科を分類軸にしたクロス表を Table 2 に（`C#` が各クラスタ, `U##` が各学科）、コレスポンデンス分析を用いたクロス表を 3次元空間へマッピングした結果を Fig. 1 に示す。

Table 1 クラスタ要約（クラスタ数4および7）

#C=4	クラスタ要約	#C=7	クラスタ要約
C0	積分, 微分, 展開, 行列, 微分方程式, ベクトル, 関数, 級数, 線形, 変換,	C6	積分, 微分, 展開, 行列, 微分方程式, ベクトル, 関数, 級数, 線形, 変換,
C1	情報, システム, 確率, 管理, 信号, 技術, 通信, 半分, 知識, 分布,	C0	画像, 認識, 音声, パターン, 識別, パターン認識, 画像処理, 特徴抽出, 処理, 特徴,
		C1	情報, システム, 確率, 管理, 信号, 技術, 通信, 半分, 知識, 分布,
C2	回路, 反応, 設計, 運動, 応力, エネルギー, 制御, サイクル, 電気, 流れ,	C2	回路, 電気, 電子, 光, 交流, 半導体, 電気回路, 電子回路, トランジスタ, 電界,
		C3	反応, 設計, 応力, 運動, サイクル, 制御, エネルギー, 流れ, 構造, 振動,
		C5	応用物理学, 報告書, 報告, 測定, 装置, マイク, 仕事当量, 確率誤差, 副尺, 熱の仕事当量,
C3	言語, プログラミング, プログラム, コンピュータ, アルゴリズム, C言語, 処理, 計算機, 論理, 計算,	C4	言語, プログラミング, プログラム, コンピュータ, アルゴリズム, C言語, 処理, 計算機, 論理, 計算,

Table 2 クロス表 (クラスタ数4および7)

#C=4	C0	C1		C2			C3	計
#C=7	C6	C0	C1	C2	C3	C5	C4	計
U07	0	0	3	0	10	0	1	14
U09	10	2	6	2	7	0	10	37
U13	3	2	13	29	17	0	14	78
U15	6	0	8	8	10	0	5	37
U19	19	2	6	4	5	0	8	44
U32	1	0	11	1	4	0	1	18
U35	3	0	2	5	58	0	0	68
U37	5	3	24	16	2	0	10	60
U50	23	1	8	2	46	1	2	83
U61	7	1	16	6	25	0	7	62
U63	8	2	18	9	7	0	9	53
U73	0	0	4	5	31	0	1	41
U78	4	0	3	1	32	0	1	41
#C=7 計	89	13	122	88	254	1	69	636
#C=4 計	89	135		343			69	636

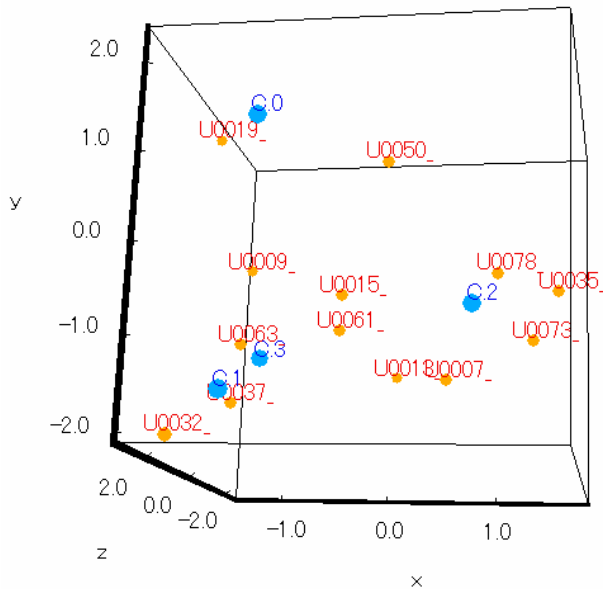


Fig.1 クラスタ数:4

C0 は数学的な用語を多く含むクラスタである。また、C1 は通信工学、C2 は電気・機械・化学工学、C3 は情報工学の分野の用語を多く含むクラスタである。3 軸まで見ることにより、各学科: U07 (機械システム工学科), U09 (情報システム工学科), U19 (情報システム工学科), U13 (応用システム工学科), U15 (システム工学科), U32 (システム工学科), U35 (化学システム工学科), U37 (情報システム工学科), U50 (機械システム工学科), U61 (システム工学科), U63 (知能情報システム工学科), U73 (機械システム工学科), U78 (機械システム工学科) は大まかに関連の深いクラスタの周辺にあることや特徴的な学科 (U09, U19, U32, U50) を見てとることができる。

つぎに、クラスタ数 7 で分析を行った結果を

Table 1 (右列) および Fig.2 に示す。

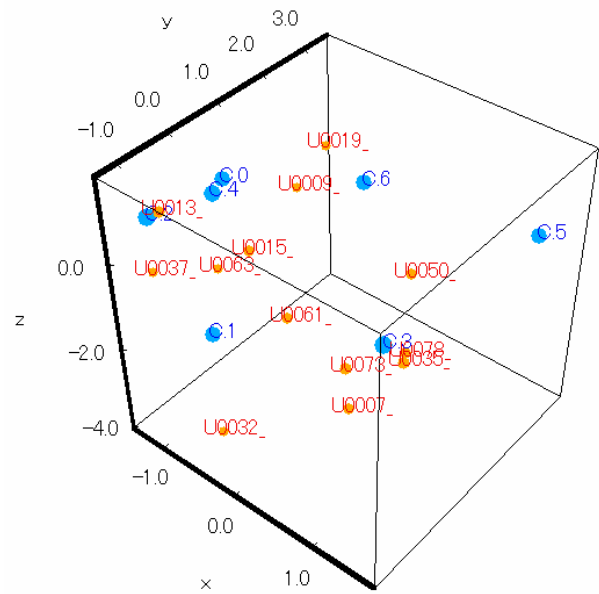


Fig. 2 クラスタ数:7

クラスタ数が 4 の場合と 7 の場合のクラスタの対応関係を Table 2 の上部に示す。これにより、例えば、U07, U35, U73, U78 はクラスタ数を増やし 3 軸までを検討しても互いに近接していることがわかる。他方、Fig.1 内でこれら (U07, U35, U73, U78) に近接している U13 は Fig.2 では分離し、むしろ U15, U37, U63 に近くなる。これは、新たに分離した C2 が電気電子に関連するシラバスから構成されるものであり、機械・化学系のシラバスからなる C3 と分離したためである。

以上のクラスタ数の変更以外にも各種の設定の変更により多様な分析を柔軟に行い、その結果を視覚的にわかりやすく理解することが可能である。

参考文献

- [1] M. Ida, T. Nozawa, F. Yoshikane, K. Miyazaki, H. Kita: Syllabus database and web service on higher education, ICACT2005 (2005)
- [2] 渡辺将尚, 絹川博之, 井田正明, 芳鐘冬樹, 野澤孝之, 喜多一: Web 上のシラバス情報の収集と XML 変換, FIT2004 (2004)
- [3] 芳鐘冬樹, 井田正明, 野澤孝之, 宮崎和光, 喜多一: シラバスからの関連用語収集手法の検討と検索システムへの応用, IPSJ 全国大会 (2005)
- [4] 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 (2005)