

# シラバス内の専門用語間の関連に基づく カリキュラム構造可視化手法の検討

野澤 孝之<sup>†</sup> 渋井 進<sup>†</sup> 芳鐘 冬樹<sup>†</sup> 井田 正明<sup>†</sup> 宮崎 和光<sup>†</sup> 喜多 一<sup>‡</sup>  
(独) 大学評価・学位授与機構<sup>†</sup> 京都大学<sup>‡</sup>

## 1. はじめに

高等教育における教育カリキュラムは、学問の体系性および社会ニーズ等を反映し、同一の分野でも提供する教育機関により多様である。その構造の分かりやすい提示は、教育機関/学習者双方の立場から重要であるが、カリキュラムの全体像を把握し適切に表現・提示するのは判断力と労力を要する困難な仕事でもある[1]。

本研究ではカリキュラムを構成する授業科目のシラバスに注目し、そこから抽出される専門用語間の関連を用いて授業科目間の関係を半自動的に抽出し、その総体であるカリキュラムの構造を可視化する手法とシステムの検討を行う。

## 2. 手法

本研究が提案する手法では、次のような手順で授業科目間の関係を抽出・可視化する：

### 手順 1. 専門用語リストの作成

対象カリキュラムのシラバスの授業内容記述部分をコーパスとして、そのカリキュラムにおける専門用語（複合名詞または単名詞）を抽出する。抽出には、語基の接続頻度に基づく手法[2]を用いる。当該分野における専門用語の辞書がある場合には、代わりにそれを利用できる。

### 手順 2. 用語間関連ネットワークの構成

専門用語を構成する語基集合の包含関係（例：「知的符号化」と「符号化」）に基づき、専門用語間の下位-上位関連を組織する。用語  $t_i$  と  $t_k$  が下位-上位関連にあるとき  $e_T(t_k, t_i) = 1$ 、そうでなければ  $e_T(t_k, t_i) = 0$  という隣接行列  $e_T$  で規定される用語間関連ネットワーク（グラフ）は、シソーラスの原形となる。

### 手順 3. シラバス（科目）内容の定量化

シラバスから専門用語辞書に含まれる用語を抽出し、各科目を文書集合のベクトル空間モデル[3]における用語ベクトルとして定量化する。この際、複合語としての出現は、それを構成するより短い語の出現としてはカウントしない。各科目にお

Study for curriculum visualization based on relation between terms in syllabus data

<sup>†</sup> Takayuki Nozawa, Susumu Shibui, Fuyuki Yoshikane, Masaaki Ida, Kazuteru Miyazaki: National Institution for Academic Degrees and University Evaluation

る用語の重みづけ関数には、TFIDF[4]や  $FFLR=tf \times FLR[2]$  を利用できる。

### 手順 4. 科目間関係ネットワークの導出・可視化

各シラバス対について、シラバスに含まれる用語間の関連を累積して、対応する科目間の下位-上位関係を導出する。すなわち、科目  $s_i$  を上位、 $s_j$  を下位とする関係の強さを

$$e_s(s_i, s_j) = \sum_{t_k \in T} \sum_{t_l \in T} e_T(t_k, t_l) \sqrt{w(s_i; t_k) w(s_j; t_l)} \quad (1)$$

で表す。ただし  $w(s_i; t_k)$  は科目  $s_i$  における用語  $t_k$  の重みを表す。 $e_s$  の値がある閾値以上の科目間関係だけを辺として精選することで、カリキュラムの構造は、科目を頂点、 $e_s$  を辺重みとする有向ネットワークとして表現される。

こうして得られた科目間関係ネットワークは、科目の中心性（隣接辺の重み総和で評価）、専門性（出力辺の重み総和と入力辺の重み総和の差で評価）、および科目間関係の強さ（辺重み）に基づくレイアウト手法を用いて可視化される。

## 3. 試作システム

上記の手法を実行するため、図 1 のようなユーザインターフェースを持つシステムを試作した。このシステムでは、2 節の手順を順番に実行する他に、得られた結果から手順を遡って、手順 1 で抽出された専門用語のうち妥当でないものを削除したり、手順 2 で自動生成された形態的・統語的な特徴に基づく用語間関連に加えて意味的な関連を追加して、シソーラスとカリキュラム構造を同時に改良していくことが可能である。

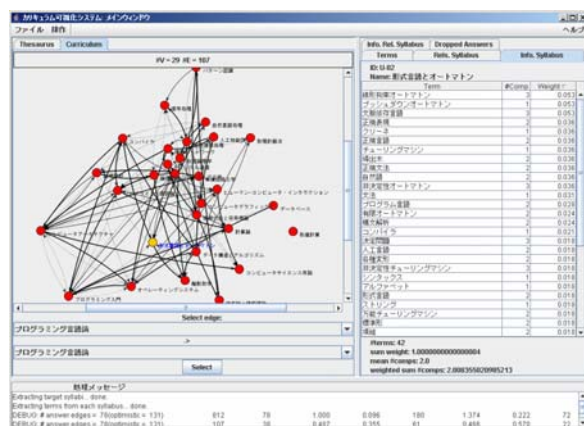


図 1 試作システムのユーザインターフェース

#### 4. 評価実験

情報処理学会による『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』[5] (以下, J97) のシラバスを対象とした分析事例による提案手法の評価を示す. J97 は, 学部教育の 29 科目を提示しており, 科目ごとに「先修科目」「後修科目」「関連科目」を定めている. ここでは科目  $s_j$  が科目  $s_i$  を「先修科目」に指定しているか,  $s_i$  が  $s_j$  を「後修科目」に指定している場合に, 下位科目  $s_j$ -上位科目  $s_i$  の関係があると定め, これを科目間関係の正解集合とする. 正解の科目間関係は 78 あり, これに対する再現率と精度で提案手法を評価する.

分析で用いた用語間関連ネットワークは, 語基の包含関係に基づき自動生成されたものである. 式(1)の  $e_s$  の閾値を徐々に上げて科目間関係を精選していったときの, 再現率, 精度, および  $F$  値 =  $(2 \times (\text{再現率} \times \text{精度}) / (\text{再現率} + \text{精度}))$  [6]の推移を図 2 に示す. なお, 数  $29 \times 28$  の全ての科目対から数  $n$  の科目間関係をランダムに設定したときの, 再現率の期待値 =  $n / (29 \times 28)$ , 精度の期待値 =  $78 / (29 \times 28)$ , および  $F$  値の期待値も比較のために示している. 再現率, 精度は絶対値としては満足な値でないが, 科目間関係をランダムに設定する場合と比べれば, 明らかに優れている.

次に,  $F$  値が最大値をとる閾値の位置 (抽出科目間関係の数 107, うち正解の数 38) で, 科目間関係の抽出の誤りを詳細に見る.

まず抽出されなかった 40 の正解を調べると, これらの多くは「下位科目  $s_j$  が上位科目  $s_i$  の基礎的概念の修得を前提する」というタイプの関係である (例: 「数理計画法の学習にあたり, 離散数学で習う集合, 写像, 証明などの基礎概念は前提」). これらの科目間関係は, 用語間関連ネットワークに「理解の前提となる」という意味的な関連を加えることで捕捉可能となる. また, 正解集合の中に妥当性が不確実な関係も幾つか見つかった (例: 下位科目=コンピュータグラフィックス, 上位科目=データベース).

一方, 抽出された関係のうち不正解の 69 を調べると, 下位-上位を逆向きにして正解になるものが 22, 「関連科目」として定められているものが 23 あった. これらは, 科目間関係の強さは妥当に評価できるが, 下位-上位の向きを正しく捉えるのが困難なケースを示している (たとえば概論的科目は, 高度な概念まで通覧するため, 誤って下位科目に位置付けられ易い). 残り 24 の関係は, 正解集合からは外れるものであるが, 実際の科目内容を見ると, 上記の「抽出されなかった 40 の正解」と同程度に妥当と思われるものが多かった.

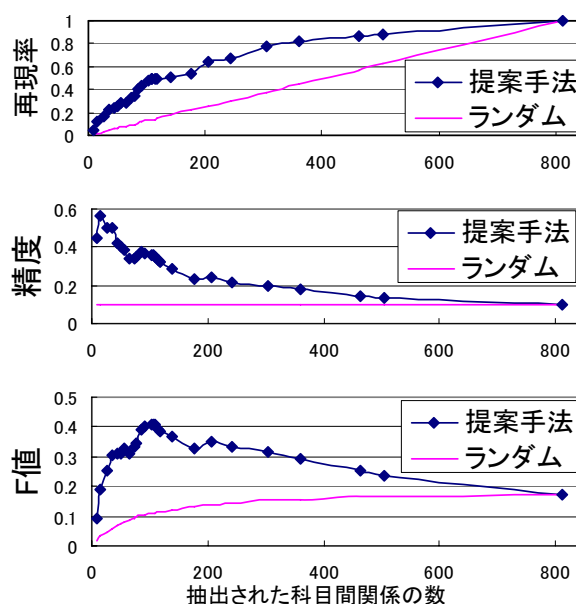


図 2 提案手法の評価結果

#### 5. おわりに

本研究の提案手法は, 『上位概念 (用語) は下位概念 (用語) に先だつて教授されるべきである』という作業仮説を前提している. 一方, J97 はこの前提とは独立に, 『抽象から具象へのトップダウンと具象から抽象へのボトムアップの, 効果的な混在』などのカリキュラム編成の原理を掲げている[5]. しかし 4 節の結果は, 本手法の作業仮説と独立な原理で編成されたカリキュラムであっても, ある程度はその構造を抽出できる見通しを示唆している.

本手法の利点は, 科目間の関係と同時に, その関係を成立させている概念も明示することである. この情報は特に学習途上の学生がカリキュラムの全体像を把握する上で有用であると考えられる. 今後, カリキュラム内の科目群とそこに含まれる概念 (用語) 群という 2 層のネットワーク情報のより分かりやすい提示法を検討していく. また, この 2 層ネットワークの働きの双方向性をより強調し, 強い科目間関係から逆に意味的な概念間関係の候補を積極的に提案する手法も検討中である.

#### 参考文献

- [1] 大場: 「工学知の構造化と可視化」の試み, 大学評価・学位研究, No. 1, pp. 97-109 (2005).
- [2] 中川ら: 出現頻度と接続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] Salton, G.: *Automatic Text Processing*, Addison-Wesley (1989).
- [4] Salton, G.: *Developments in automatic text retrieval*, *Science*, Vol. 253, pp. 974-980 (1991).
- [5] (社) 情報処理学会: 大学の理工系学部情報系学科のための CS 教育カリキュラム J97 第 1.1 版 (1999).
- [6] 徳永: 情報処理と言語処理, 東大出版 (1999).