

Spectral Clustering を用いた 企業アンケートからのモデルコースの抽出

Extraction of Core Course from Survey to Companies by Spectral Clustering

鶴田 直之†
Naoyuki Tsuruta

森元 逞†
Tsuyoshi Morimoto

前田 佐嘉志†
Sakashi Maeda

高橋 伸弥†
Shin-ya Takahashi

1. まえがき

JABEE (Japan Accreditation Board of Engineer Education) [1]の普及により、それまで個別に取り組みれていた各種の教育改善手法が、教育マネジメントサイクルという最適化手法の構成要素として体系的に扱われるようになった。その中で、社会的ニーズと大学教育の一致を図ることの重要性が注目されている。例えば、企業ニーズと大学カリキュラムとのマッチ度を評価し、比較する試みも行われている[2]。

社会的ニーズを調べる最も一般的な方法はアンケートである。近年のホームページ閲覧環境の普及に伴い、企業に限らず、高校教諭や卒業生、父母へのアンケートも容易になっている。特別な費用もかからず、高い頻度で実施することも可能である。アンケート分析に基づいて社会的ニーズを抽出する技術は、教育マネジメントサイクルを実現する上で欠かせないものになると思われ、分析手法の開発が重要な課題となる。

本稿では、2003年に福岡大学電子情報工学科で行った企業アンケート結果の分析に基づき、カリキュラムのモデルコースを自動抽出する試みを報告する。2003年当時は科目の重要度を問う質問の回答から科目間の相関を調べることによって科目を手動で分類し、モデルコースを抽出した。今回は、Spectral Clusteringと呼ばれる分類手法を用いて自動抽出を試みたところ、2003年より優れた結果を得ることができた。これにより、分析の質が向上したのみならず、アンケートの高頻度化・大規模化が容易になったといえる。

以下、アンケート分析の目的と方法、Spectral Clustering、アンケートの実施方法、分析結果の順に述べる。

2. アンケート分析の目的と方法

アンケートを通して知りたい事柄は、得られた回答の表層から直接知りえる顕在要因と、項目間の相関関係などから分析によって把握できる潜在要因に分けることができる。顕在要因は、更に個々の質問で完結する絶対要因と、項目間の差によって把握できる相対要因に分けることができる。2003年に実施したアンケートでは、主要科目の重要度を5段階評価で質問しており、各科目間の平均得点が絶対要因となる。また、同じ質問を「大学一般について」と「福岡大学電子情報工学科について」に分けて行っており、両者の差や科目間の差が相対要因となる。

モデルコースの抽出は、潜在要因として回答における科目間の相関を求め、互いに高い相関を持つ科目群に分類することによって行った。仮に絶対要因として重要度の高い科目が複数あったとしても、それらが相互相関によって二つの科目群に明確に分類される場合、二つの科目群は同時に要求されることが少ないのであって、別の

目的や意義を持って教育されるべきものである可能性が想定し得るからである。

今、科目間の相関係数行列 $\mathbf{R} = [r_{ij}]$ を考える。 r_{ij} は科目 i と j の相関係数で、 \mathbf{R} は対称行列すなわち $r_{ij} = r_{ji}$ を満たしているものとする。このとき、相互に相関が強い科目群が3つ存在していたとする。相関の強い科目どうしが連続するように科目の順番を上手に入れ替えることによって、すなわち \mathbf{R} の行と列の順番を同時に入れ替えることによって、図1のように大きな相関係数を対角上の3つのブロックに寄せ集めることができるはずである。逆に、一つのブロックに集まった相関係数は、元の \mathbf{R} では「閉じた」格子状に分布していたことになる。

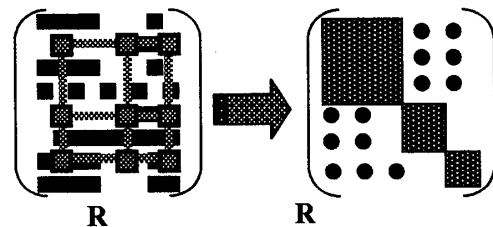


図2は2003年に実施したアンケートの17科目の相関係数行列から、「閉じた」格子状の高い相関係数として科目群(網掛けの部分)を手動で抽出できた様子である。

相関係数 r_{ij} は次式により計算した。 p 番目のアンケート結果における科目 i の重要度を $x_i^{(p)}$ 、全アンケート結果に対する科目 i の重要度の平均を \bar{x}_i と現すと、

$$r_{ij} = \frac{\sum_p (x_i^{(p)} - \bar{x}_i)(x_j^{(p)} - \bar{x}_j)}{\sqrt{\sum_p (x_i^{(p)} - \bar{x}_i)^2 \sum_p (x_j^{(p)} - \bar{x}_j)^2}}$$

である。

分析によって、2つの系列が抽出できた。具体的には、『大学一般および福岡大学に差はなく、「ネットワーク—ソフトウェア設計—プログラミング—コンピュータシステム—計算機ハードウェア/周辺機器—マルチメディア」と「電子・電気回路—LSI設計—半導体・電子材料」という相関の強い2つの科目系列が抽出でき、「ソフトウェアおよび総合的な情報システム」と「電子デバイス」がモデルコースの柱となることが分かった。科目の順番は重要度順である。期待度は時代背景を受け、前者が高かった。』

0.3				-0.02	0.29	0.04	0.08	0.3	0.16	0.42	0.07	0.04	0.06	0.37			
0.3	1	0.18	0.45	0.31	0.54	0.25	0.05	0.52	0.18	0.14	0.57	-0.06	0.23	0.15	0.19	0.15	
		0.18		0.09	0.34	-0.06		0.11	0.3		0.06		0.11	0.05	0.43		
		0.45		0.11	0.36	-0.07	-0.05	0.02	0.23		-0.07		0.04	-0.04	0.52		
-0.02	0.31	0.09	0.11	1	0.37	0.53	0.21	0.3	0.29	0.03	0.26	0.45	0.27	0.14	0.17	0.26	
0.29	0.54	0.34	0.36	0.37	1	0.2	0.31	0.57	0.26	0.25	0.57	0.3	0.24	0.28	0.23	0.31	
0.04	0.25	-0.06	-0.07	0.53	0.2	1	0.17	0.24	0.05	0	0.23	0.34	0.26	0.22	0.47	0.27	
0.08	0.05	0	-0.05	0.21	0.31	0.17	1	0	0.39	0.54	0.24	0.44	0.36	0.62	0.62	0.35	
		0.52		0.3	0.57	0.24	0		0.09	0.16		-0.01		0.18	0.16	0.42	
0.16	0.18	0.11	0.02	0.29	0.26	0.05	0.39	0.09	1	0.3	0.15	0.35	0.26	0.3	0.38	0.22	
0.42	0.14	0.3	0.23	0.03	0.25	0	0.54	0.16	0.3	1	0.32	0.36	0.52	0.32	0.35	0.3	
		0.57		0.26	0.57	0.23	0.24		0.15	0.32		0.1		0.33	0.36	0.5	
0.07	-0.06	0.06	-0.07	0.45	0.3	0.34	0.44	-0.01	0.35	0.36	0.1	1	0.14	0.14	0.28	0.41	0.21
	0.23		0.27	0.24	0.26	0.36		0.26	0.52		0.14		0.2	0.21	0.69		
0.04	0.15	0.11	0.04	0.14	0.28	0.22	0.62	0.18	0.3	0.32	0.33	0.28	0.2	1	0.74	0.38	
0.06	0.19	0.05	-0.04	0.17	0.23	0.47	0.62	0.16	0.38	0.35	0.36	0.41	0.21	0.74	1	0.38	
0.37	0.15	0.43	0.52	0.26	0.31	0.27	0.35	0.42	0.22	0.3	0.5	0.21	0.69	0.38	0.38	1	

図 2. 実際の相関係数行列から科目群を抽出した例

3. Spectral Clustering

Spectral Clustering [3][4]はグラフ理論の最小カット問題の解法として提案された手法であり、その簡潔性から様々な分野で応用されている。以下、Spectral Clusteringの原理を示す。今、無向グラフの最小カット問題を考える。グラフを構成するノードを通し番号で表すことにする。行列 $\mathbf{W}=[w_{ij}]$ はノード i とノード j の結合の有無を表す

もので、結合がある場合が $w_{ij}=1$ 、ない場合が $w_{ij}=0$ である。各ノードが2つのグラフ A, B のいずれに属するかを示す変数として

$$q_i = \begin{cases} 1 & \text{if } i \in A \\ -1 & \text{if } i \in B \end{cases}$$

を導入する。このとき、グラフのカットサイズは

$$\begin{aligned} J = \text{CutSize} &= \frac{1}{4} \sum_{i,j} w_{ij} (q_i - q_j)^2 \\ &= \frac{1}{4} \sum_{i,j} w_{ij} (q_i^2 + q_j^2 - 2q_i q_j) \\ &= \frac{1}{2} \sum_{i,j} q_i (d_i \delta_{ij} - w_{ij}) q_j \\ &= \frac{1}{2} \mathbf{q}^T (\mathbf{D} - \mathbf{W}) \mathbf{q}, \end{aligned}$$

where $d_i = \sum_j w_{ij}$

と表すことができる。ここで、 q_i の値を2値から連続値に緩和するとカットサイズ $\min J(\mathbf{q})$ の解は

$$(\mathbf{D} - \mathbf{W}) \mathbf{q} = \lambda \mathbf{q}$$

の固有ベクトルとして与えられる。 \mathbf{q} の要素は既に連続値に緩和されているので、解では q_i は、 $-1 \leq q_i \leq 1$ の連続値である。そこで、一般にはノードを $\{i | q_i < 0\}$ と $\{j | q_j \geq 0\}$ の2つのクラスに分類する。目的はカットサイズ J を最小にすることであるから、解は小さい固有値に対応した固有ベクトルとなるが、最小固有値は0で、そのときの固有ベクトルは直流成分の $(1, 1, \dots, 1)^T$ であり無意味である。そこで、2番目に小さな固有値に対応した固有ベクトルを選択する。「本稿では」これを第一成分と呼ぶことにする。第二成分以降もクラスタリングにお

ける重要な情報を与えることが多い。第一成分だけでは、全体を2つのクラスにしか分類できないので、より多くのクラスに分類する場合は、第二成分以降も利用する。

解を導く手続き上、 \mathbf{W} は正値対称行列になっていれば十分なので、応用の範囲が極めて広い。クラスタリング問題に適用する場合は、対称性を持つ類似度を用いて項目間の「対戦表」として類似度行列を作ればよい。本稿においては、2003年のアンケート分析に用いた相関係数行列を \mathbf{W} とする。

4. アンケートの実施方法

2003年のアンケートは、質問用紙を各企業の人事課宛に郵送して依頼した。対象企業は、前身の電子工学科を含め、当学科から就職実績および求人実績のあった企業196社とした。会社名は公開しないことをアンケートの条件とした。

質問は、大きく「大学一般について」問う質問群と「福岡大学電子情報工学科について」問う質問群に分かれている。それぞれの質問群の一項目として、別々に次の質問を行った。カリキュラムのモデルコース抽出には、科目の重要度を用いた。

次に挙げる科目を大学で教育することの必要性・重要度はどのくらいですか。また、内容には、どのくらいのレベルを要求されますか。

重要度 (①極めて重要 ②比較的重要 ③さほど重要ではない ④必要ない)

レベル (①最先端レベル ②実践的なレベル ③多少の応用を含む ④基礎だけでよい)

- (1) 一般教養
- (2) 国語力・文章力
- (3) 技術者倫理
- (4) 外国語 (英語)
- (5) 計算機活用能力 (リテラシ)
- (6) 工学基礎科目 (数学, 物理)
- (7) 半導体・電子材料
- (8) 電子・電気回路
- (9) LSI 設計
- (10) 通信・伝送・光
- (11) 計算機ハードウェア/周辺機器
- (12) プログラミング
- (13) ソフトウェア設計
- (14) コンピュータシステム
- (15) ネットワーク
- (16) マルチメディア
- (17) ニューロ・AI

5. アンケートの結果と分析

アンケートの科目重要度から得られた相対要因として、17科目の科目名と重要度の順位を「大学一般について」の場合と「福岡大学電子情報工学科について」の場合について表1に示す。

表1. 科目名と重要度順位

科目名	大学一般	福岡大学
一般教養	5	7
国語力・文章力	3	5
技術者倫理	1	6
外国語(英語)	7	10
計算機活用能力(リテラシ)	2	2
工学基礎科目(数学, 物理)	10	13
半導体・電子材料	15	16
電子・電気回路	9	8
LSI設計	16	15
通信・伝送・光	13	11
計算機ハードウェア/周辺機器	12	12
プログラミング	6	4
ソフトウェア設計	8	3
コンピュータシステム	11	9
ネットワーク	4	1
マルチメディア	14	14
ニューロ・AI	17	17

大学一般には教養を強く望むのに対し、当学科に対しては専門基礎の重要度が高まっている。これは、当学科の特徴的な傾向である。

さて、潜在要因の分析結果について述べる。「大学一般」のデータに Spectral Clustering を適応した結果、第一成分の固有ベクトルの要素 q_i が正だった科目と負だった科目に分類し、要素の絶対値 $|q_i|$ が大きいもの順に並べたものが表2である。数値は要素の絶対値である。2003年の分析と同様に「ソフトウェアおよび総合的な情報システム」と「電子デバイス」に分かれているのが分かる。また、教養科目は双方に混入しているものの、要素の絶対値が小さい。

表2. 第一成分に注目して分類した「大学一般」の結果

要素が負の科目群		要素が正の科目群	
科目名	要素値	科目名	要素値
半導体・電子材料	0.42	コンピュータシステム	0.34
電子・電気回路	0.37	プログラミング	0.33
LSI設計	0.32	ネットワーク	0.33
外国語(英語)	0.23	ソフトウェア設計	0.25
工学基礎科目(数学, 物理)	0.20	マルチメディア	0.18
通信・伝送・光	0.12	計算機ハードウェア/周辺機器	0.13
技術者倫理	0.08	ニューロ・AI	0.11
国語力・文章力	0.06	計算機活用能力(リテラシ)	0.08
		一般教養	0.07

次に、同じ「大学一般」の結果を第2成分に注目して分類した結果を表3に示す。科目が明らかに教養科目と専門科目に分かれたのが分かる。教養科目間の相関は2003年の分析では陽に現れていなかった(見落としていた)。

工学基礎科目や計算機活用能力が教養科目に分類されながらも専門科目に近く位置づけられている点は、我々の直感にも合致し、モデルコースとして妥当な分類であるといえる。

表3. 第二成分に注目して分類した「大学一般」の結果

要素が負の科目群		要素が正の科目群	
科目名	要素値	科目名	要素値
LSI設計	0.22	一般教養	0.71
電子・電気回路	0.19	国語力・文章力	0.54
ソフトウェア設計	0.16	外国語(英語)	0.18
通信・伝送・光	0.12	技術者倫理	0.14
コンピュータシステム	0.09	工学基礎科目(数学, 物理)	0.09
プログラミング	0.08	計算機活用能力(リテラシ)	0.01
ニューロ・AI	0.07		
計算機ハードウェア/周辺機器	0.05		
半導体・電子材料	0.04		
マルチメディア	0.03		
ネットワーク	0.02		

図3に、第一成分と第2成分の固有値の要素が張る2次元平面に各科目をプロットした図を示す。縦軸が第一成分である。

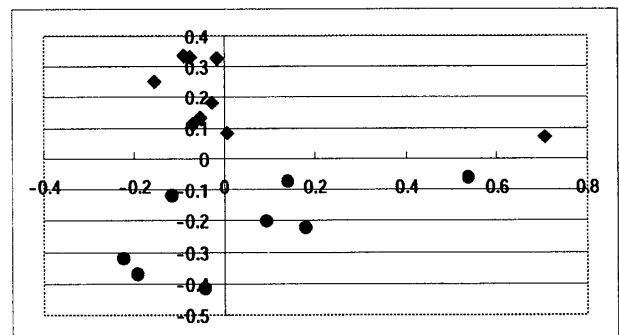


図3. 「大学一般」における科目の分布

表4. 第一成分に注目して分類した「福岡大学」の結果

要素が負の科目群		要素が正の科目群	
科目名	要素値	科目名	要素値
プログラミング	0.39	電子・電気回路	0.35
ソフトウェア設計	0.35	工学基礎科目(数学, 物理)	0.31
ネットワーク	0.34	半導体・電子材料	0.28
コンピュータシステム	0.28	LSI設計	0.25
計算機活用能力(リテラシ)	0.16	外国語(英語)	0.20
計算機ハードウェア/周辺機器	0.16	一般教養	0.20
マルチメディア	0.14	国語力・文章力	0.14
ニューロ・AI	0.08	通信・伝送・光	0.07
技術者倫理	0.05		

同様の分析を「福岡大学電子情報工学科」に行った結果を表4、表5、図4に示す。図4では縦軸が第一成分である。

第一成分による分類結果は、順番に多少の違いはあるものの「大学一般」とほぼ同等の結果が得られ、「ソフトウェアおよび総合的な情報システム」と「電子デバイス」の2コースが抽出できた。一方、第二成分による分類では、「コンピュータシステム」と「計算機ハードウェア/周辺機器」、「技術者倫理」が教養科目側に分類される結果となった。教養科目側に加わった三科目はいずれも他の教養科目に比べて専門科目側に近く、我々の直感とも合致している。

表5. 第二成分に注目して分類した「福岡大学」の結果

要素が負の科目群		要素が正の科目群	
科目名	要素値	科目名	要素値
一般教養	0.58	通信・伝送・光	0.34
国語力・文章力	0.53	電子・電気回路	0.26
計算機活用能力(リテラシ)	0.25	LSI設計	0.20
技術者倫理	0.15	半導体・電子材料	0.11
コンピュータシステム	0.13	外国語(英語)	0.10
工学基礎科目(数学, 物理)	0.08	マルチメディア	0.10
計算機ハードウェア/周辺機器	0.03	ソフトウェア設計	0.10
		ネットワーク	0.09
		ニューロ・AI	0.07
		プログラミング	0.02

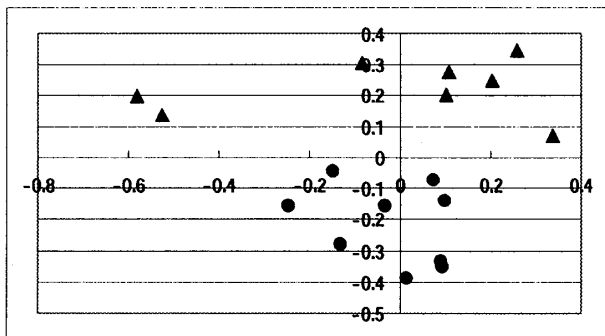


図4. 「福岡大学電子情報工学科」における科目の分布

6. あとがき

本稿では、2003年に福岡大学電子情報工学科で行った企業アンケート結果の分析に基づき、カリキュラムのモデルコースを自動抽出する試みを報告した。2003年当時は科目の重要度を問う質問の回答から科目間の相関を調べることによって科目を手動で分類し、モデルコースを抽出した。今回は、Spectral Clusteringと呼ばれる分類手法を用いて自動抽出を試みたところ、2003年より優れた結果を得ることができた。具体的には、専門科目が2003年の分析と同様に「ソフトウェアおよび総合的な情報システム」と「電子デバイス」の2コースに分類できた。加

えて、専門科目と教養科目の分類に成功した。顕在要因の重要度も加味すると、大学一般には「技術者倫理—計算機活用能力(リテラシ)—国語力・文章力—一般教養—外国語(英語)—工学基礎科目(数学, 物理)」を教養科目系列として望んでいるのに対し、福岡大学電子情報工学科には、「コンピュータシステム」と「計算機ハードウェア/周辺機器」を加えた「計算機活用能力(リテラシ)—国語力・文章力—技術者倫理—一般教養—コンピュータシステム—計算機ハードウェア/周辺機器—工学基礎科目(数学, 物理)」を教養として位置づけ望んでいることが分かった。これは、福岡大学電子情報工学科により実務家教育を求めている顕在要因の裏づけにもなっており、直感的にも納得のいく結果であった。

これにより、分析の質が向上したのみならず、アンケートの高頻度化・大規模化が容易になったといえる。

[参考文献]

- [1] <http://www.jabec.org/>
- [2] 三菱総合研究所：大学活動評価結果公開サイト、<http://www.univinfo.jp/rating/index.php>
- [3] A. Pothén, H. Simon, K. Liou, Partitioning sparse matrices with eigenvectors of graphs, SIAM J. Matrix Anal. 11 (3) (1990) 430–452.
- [4] Chris HQ Ding, <http://crd.lbl.gov/~cding/>