

選択型と記述型の学生アンケートの分析

平澤 茂一†

石田 崇††, 雲居 玄道†††, 後藤 正幸††††

†サイバー大学IT 総合学部

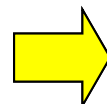
††早稲田大学メディアネットワークセンター

†††早稲田大学理工学術院総合研究所

††††早稲田大学創造理工学部

1 研究背景

大学における授業改善



学生のニーズや不満の抽出

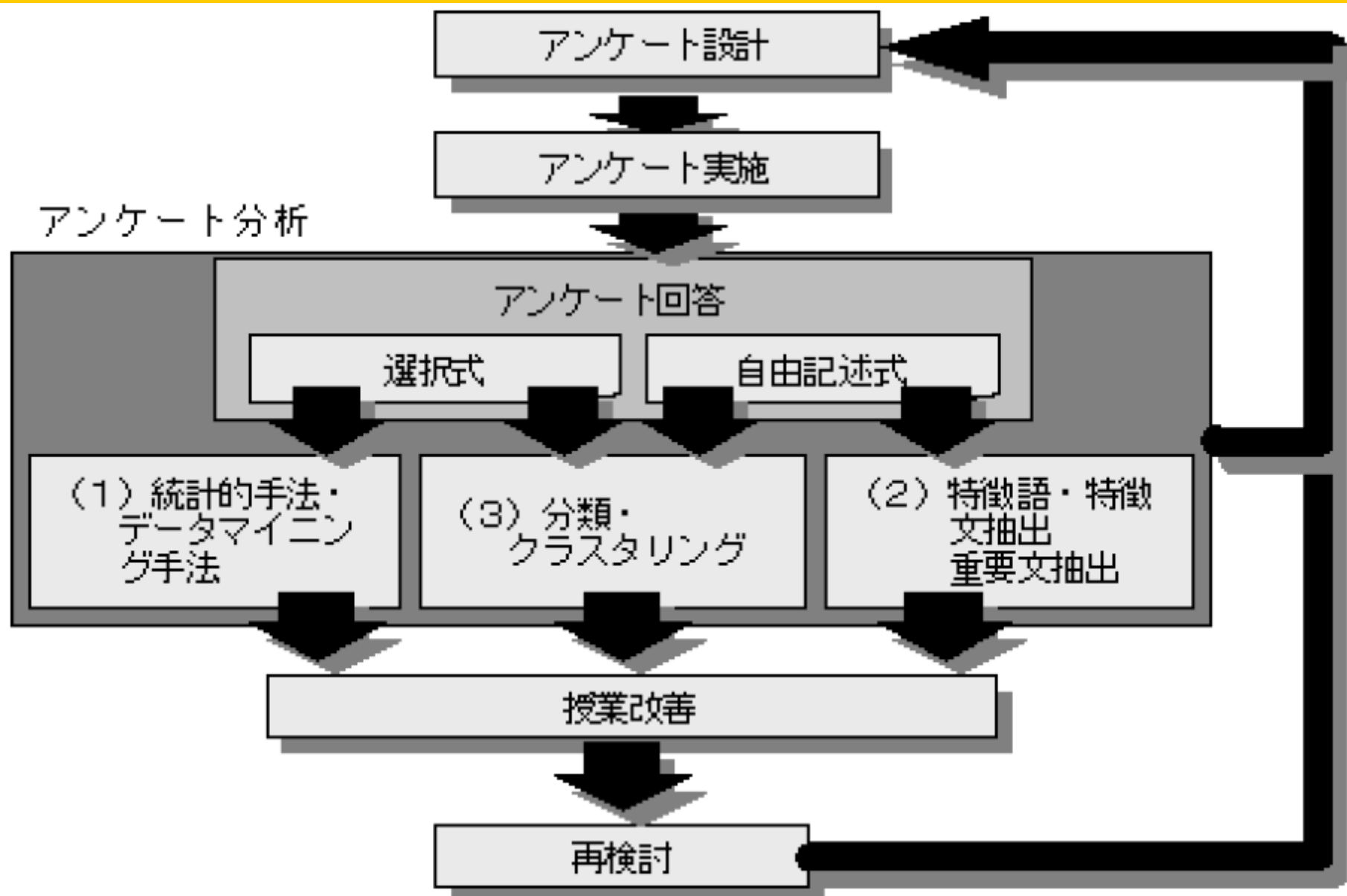
学生アンケート

選択型(二者択一, 5段階評価など)・記述型(与えられたテーマについての自由記述)が混在.

本研究の目的

サイバー大学IT総合学部の専門必修科目「コンピュータ入門」の学生アンケートに適用し, 多角的に分析し授業改善に有効なデータを得る.

2 アンケート分析手法



3 クラスタリング

1. 目的

履修者の集合(学生数: I)の中味(学生の特徴)を明らかにする.

2. 手法

似た特徴の者同士を集め, いくつかのクラスタを作る.

3. 吟味

クラスタを構成する学生の特徴を明らかにする
(特徴語抽出など).

4. 把握

クラスタの特徴を把握し, 履修者の全体像を明らかにする.

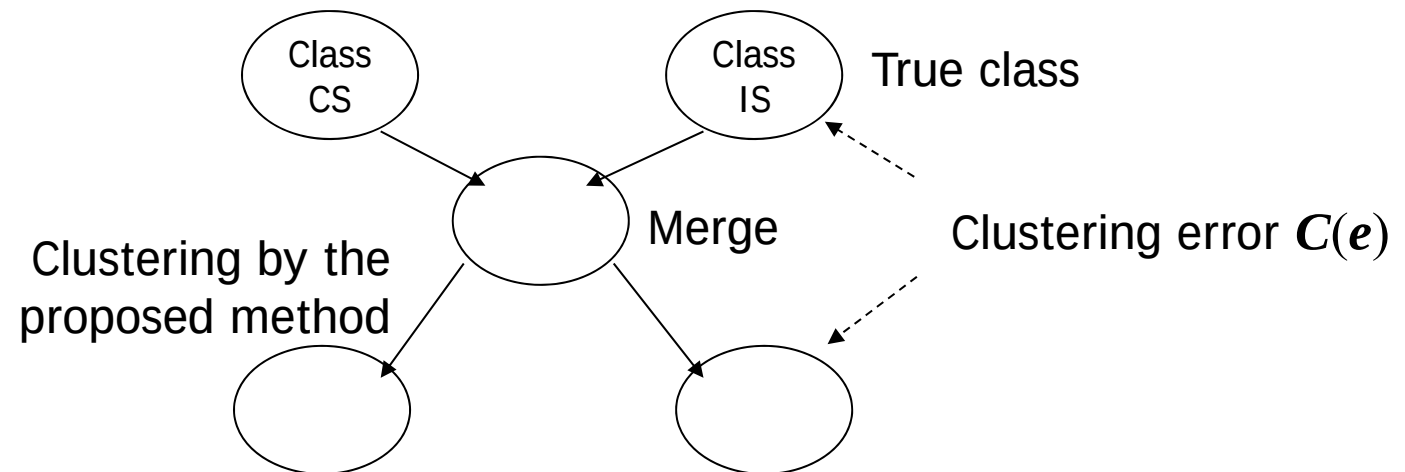
3 従来の適用例 [2]

理工系

「コンピュータ工学」

文系

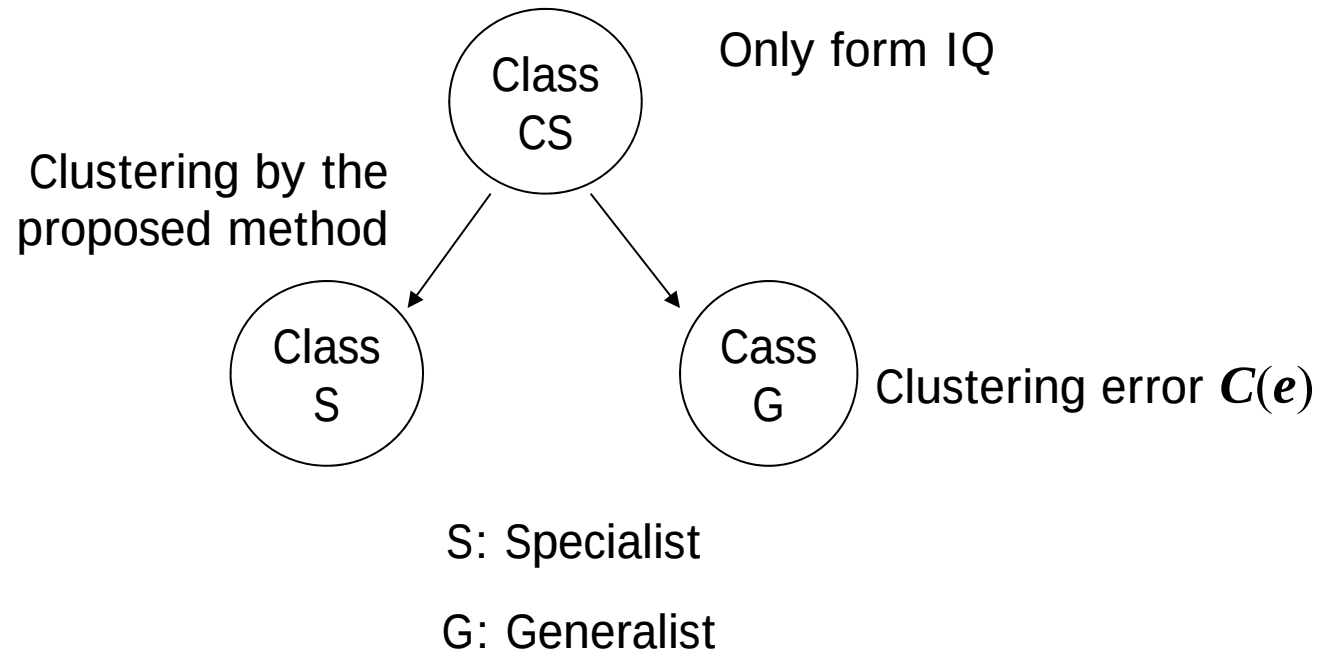
「情報化社会概論」



3 従来の適用例 [2]

クラスタリング結果→分類に使える

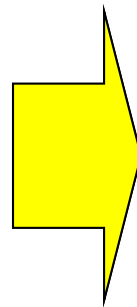
Clustering for class partition problem



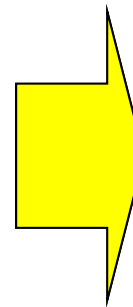
3.1 授業モデル

授業モデル

ITスキル・経験
興味
動機
将来計画
仕事
属性
その他



授業



総合成績
・期末試験
・小テスト
・レポート
満足度

3.2 拡張PLSIモデルとJS Divergence

拡張PLSIモデル

PLSI モデルでは意味的な隠れ属性 $z_k(k=1,2,\dots,K)$ の下で文書 $d_i(i=1,2,\dots,I)$ と単語 $t_j(j=1,2,\dots,J)$ の生起は独立であると仮定する. このとき, d_i と t_j の同時確率 $P(d_i, t_j)$ は

$$P(d_i, t_j) = \sum_k P(d_i|z_k)P(t_j|z_k)P(z_k)$$

で与えられる. 文書 d_i における単語 t_j の出現回数を $n(d_i, t_j)$ とすると, 次式の尤度 L は以下の式となる.

$$L = \sum_i \sum_j n(d_i, t_j) \log P(d_i, t_j)$$

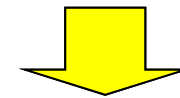
選択型と記述型の文書単語行列をそれぞれ G と H とすると, 文書単語行列 A は, 重み λ ($0 \leq \lambda \leq 1$)により

$$A = [\lambda G \quad , \quad (1 - \lambda)H]$$

JS Divergence

Kullback-Leibler divergence
(非対称)
 $D_{KL}(f(x)||g(x))$

$$= \int f(x) \log \frac{f(x)}{g(x)} dx$$



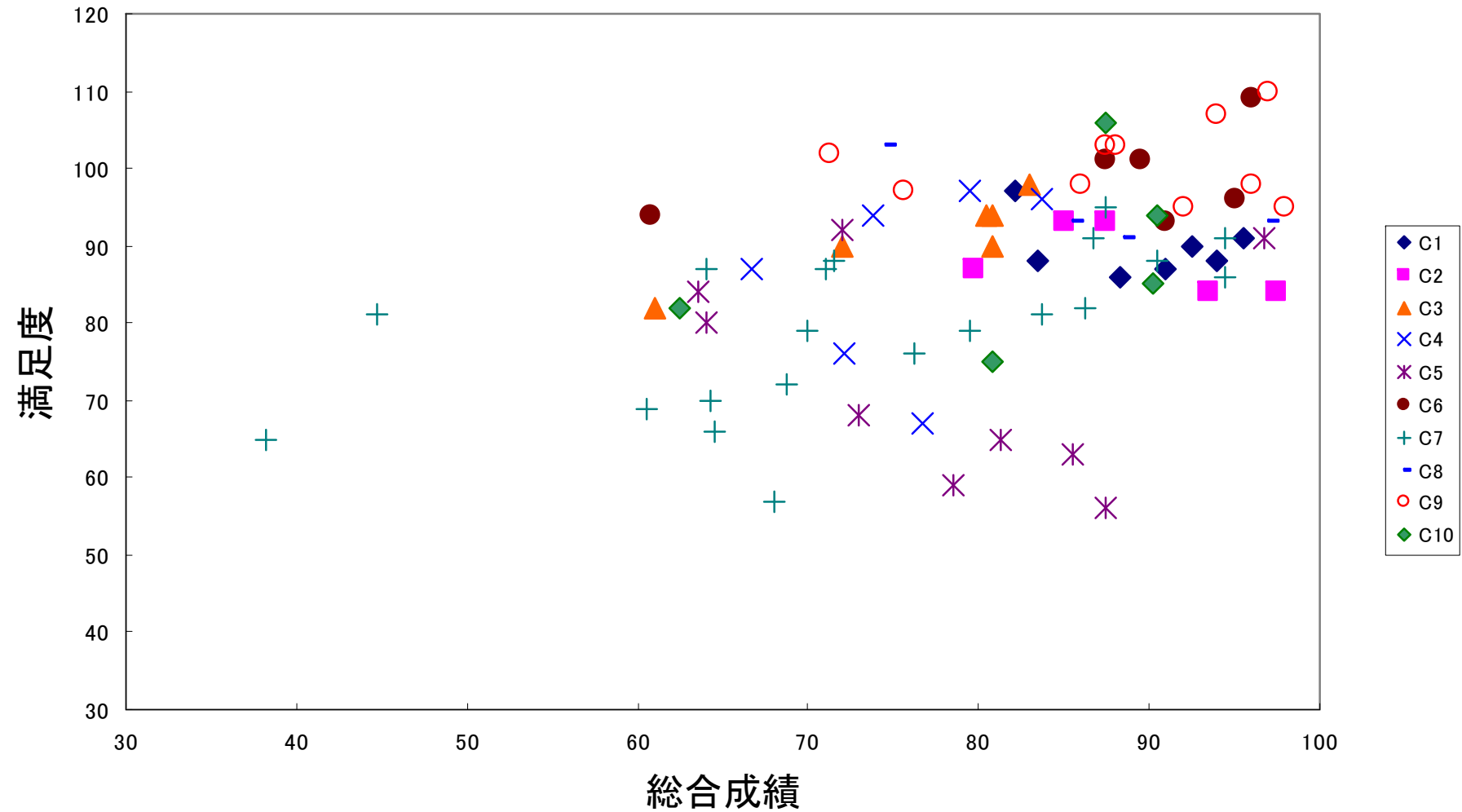
Jensen-Shannon divergence
($\lambda=0.5$ のとき対称)

$$\begin{aligned} D_{JS}(q||p) \\ &= \lambda D_{KL}(q||\lambda q + (1 - \lambda)p) \\ &\quad + (1 - \lambda) D_{KL}(p||\lambda q + (1 - \lambda)p) \end{aligned}$$

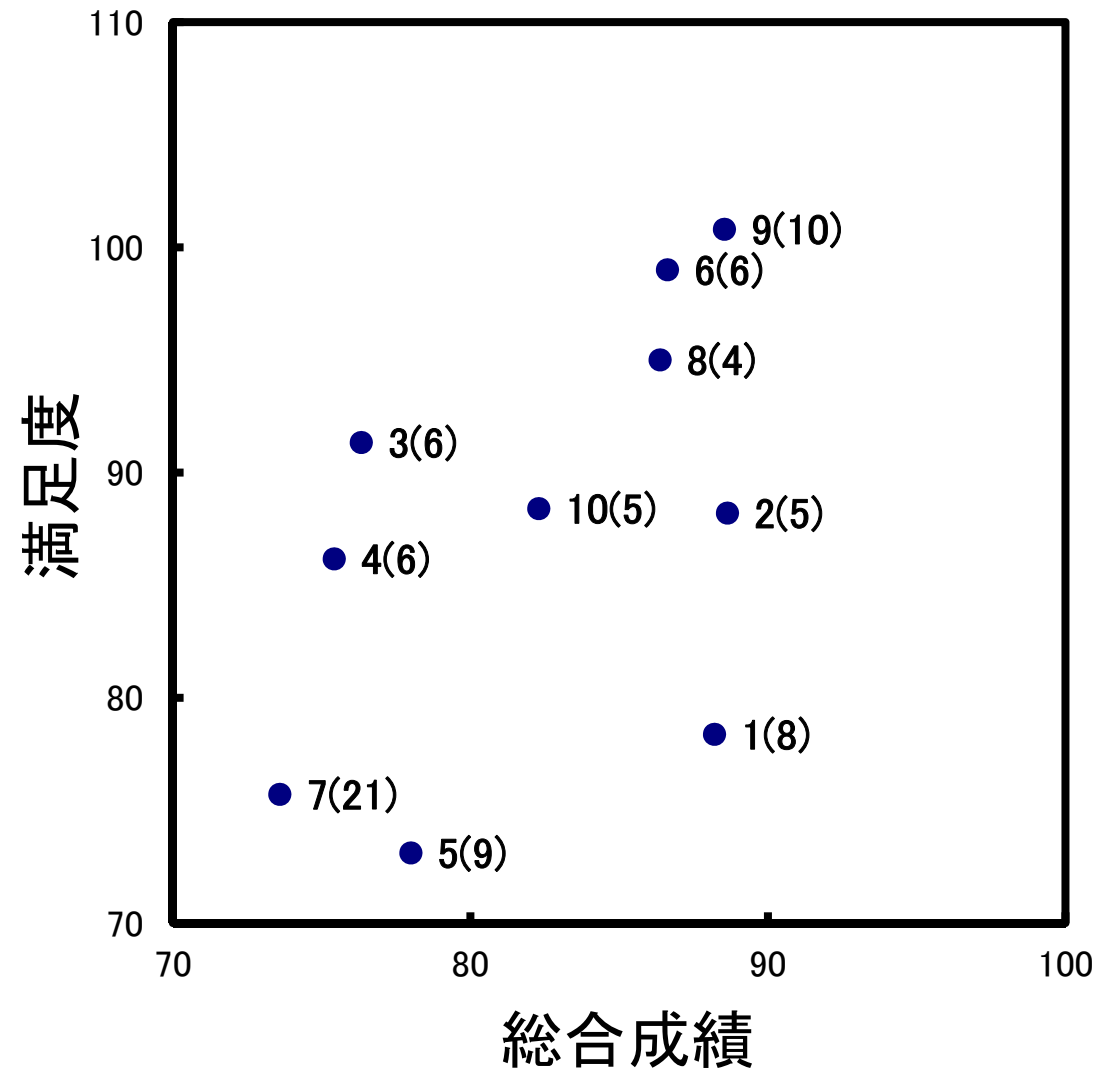
4 アンケート分析結果(分析条件)

- サイバー大学IT総合学部
専門基礎必修科目「コンピュータ入門」の学生アンケート
 - 分析対象件数: 80件(アンケート提出数)
 - アンケート項目
難易度, 進め方, 内容, 要望, 取り組み方, 興味, イメージ,
意見, 将来設計など 86項目
＞うち, 自由記述5項目: 講義への要望, 将来設計, 印象など
-
- 重み $\lambda = 0.5$ (選択式・自由記述式同等)
 - クラスタ数 $K=10$
 - 文書数 $I=80$, 単語数
 $J=2052$ (うち自由記述単語数 1730 : 名詞・動詞のみ)

4.1 クラスタリング結果

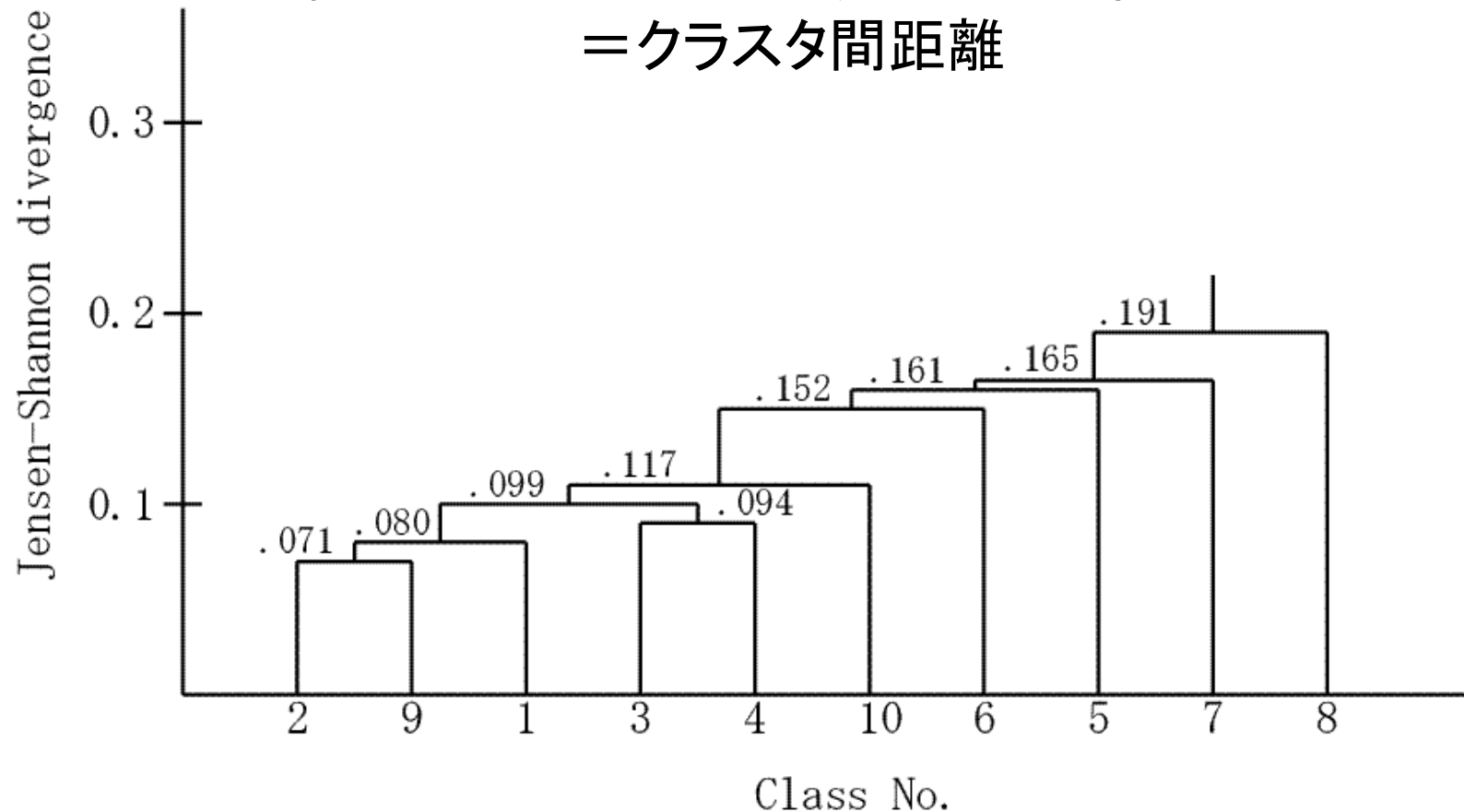


4.1 クラスタリング結果

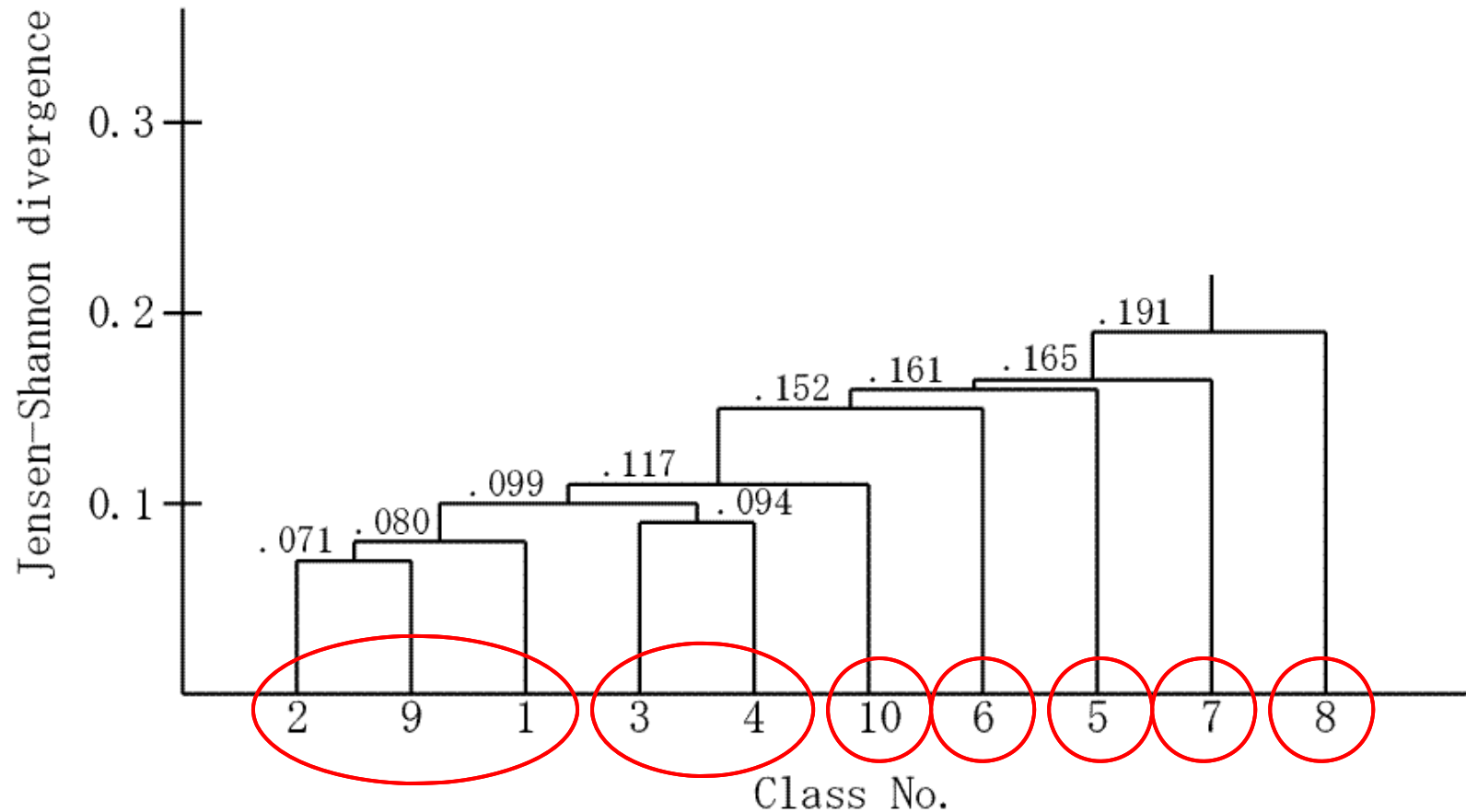


4.1 クラスタリング結果(群平均法)

群平均法: クラスタのすべての個体間の距離の平均
＝クラスタ間距離

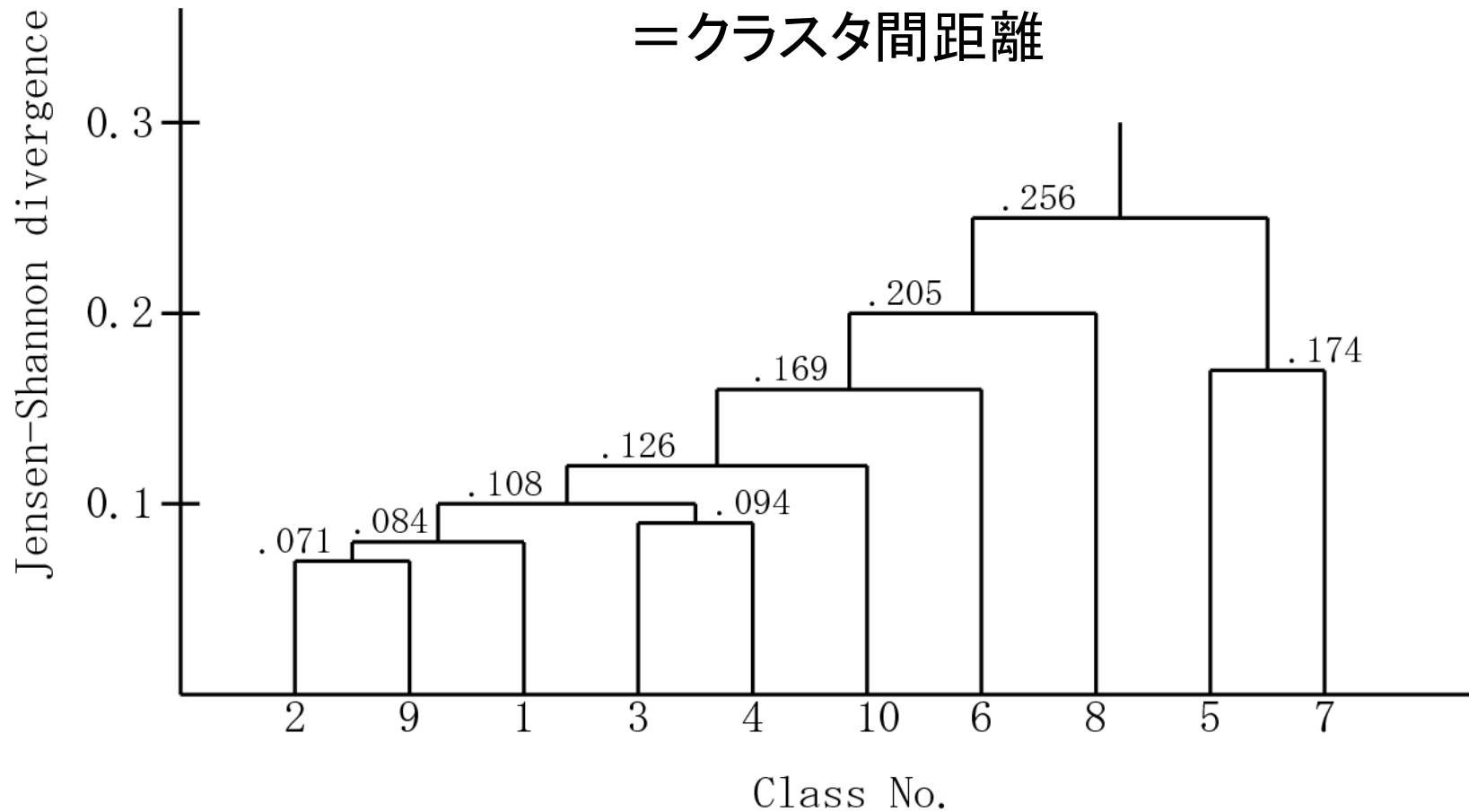


4.1 クラスタリング結果(群平均法)

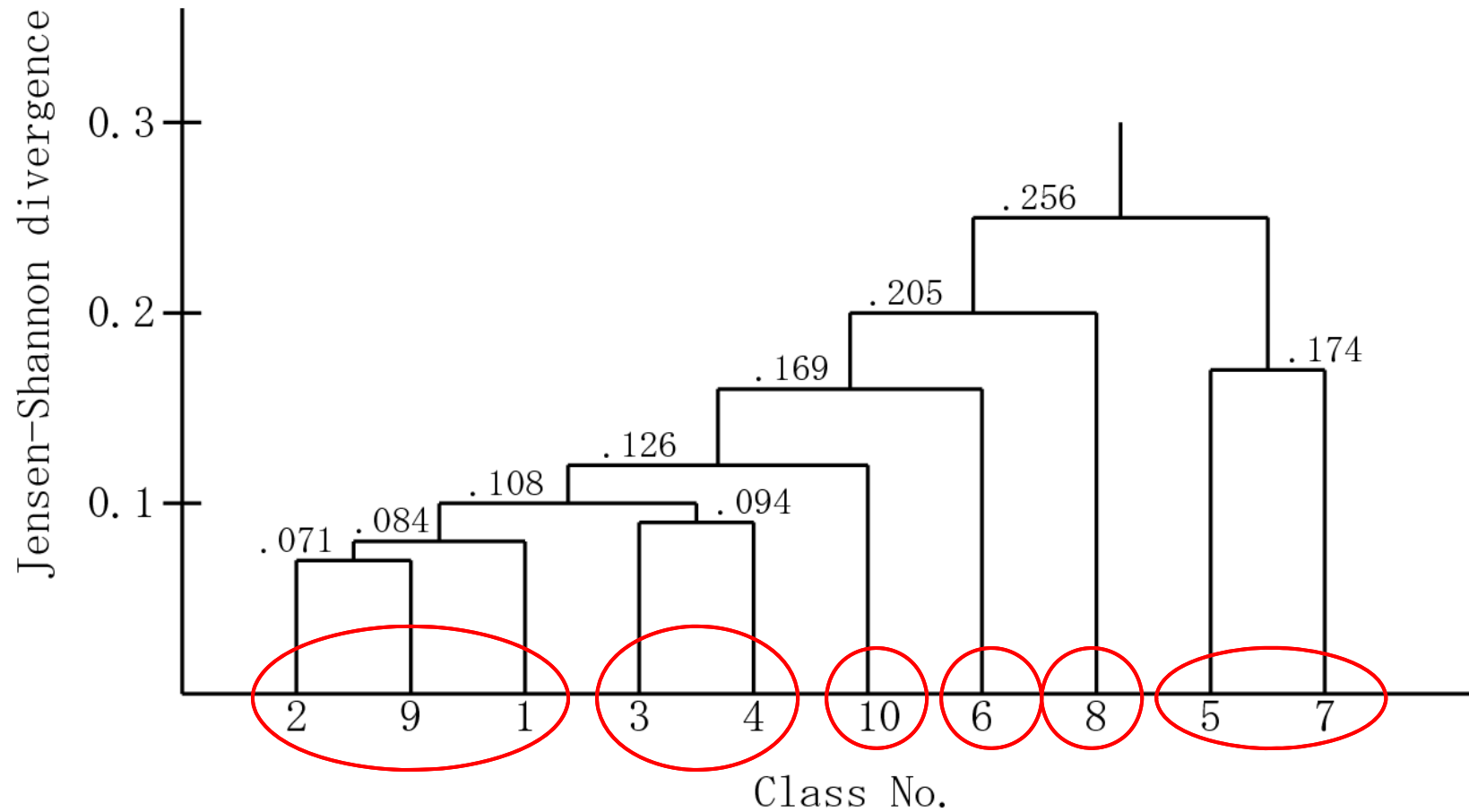


4.1 クラスタリング結果(最長距離法)

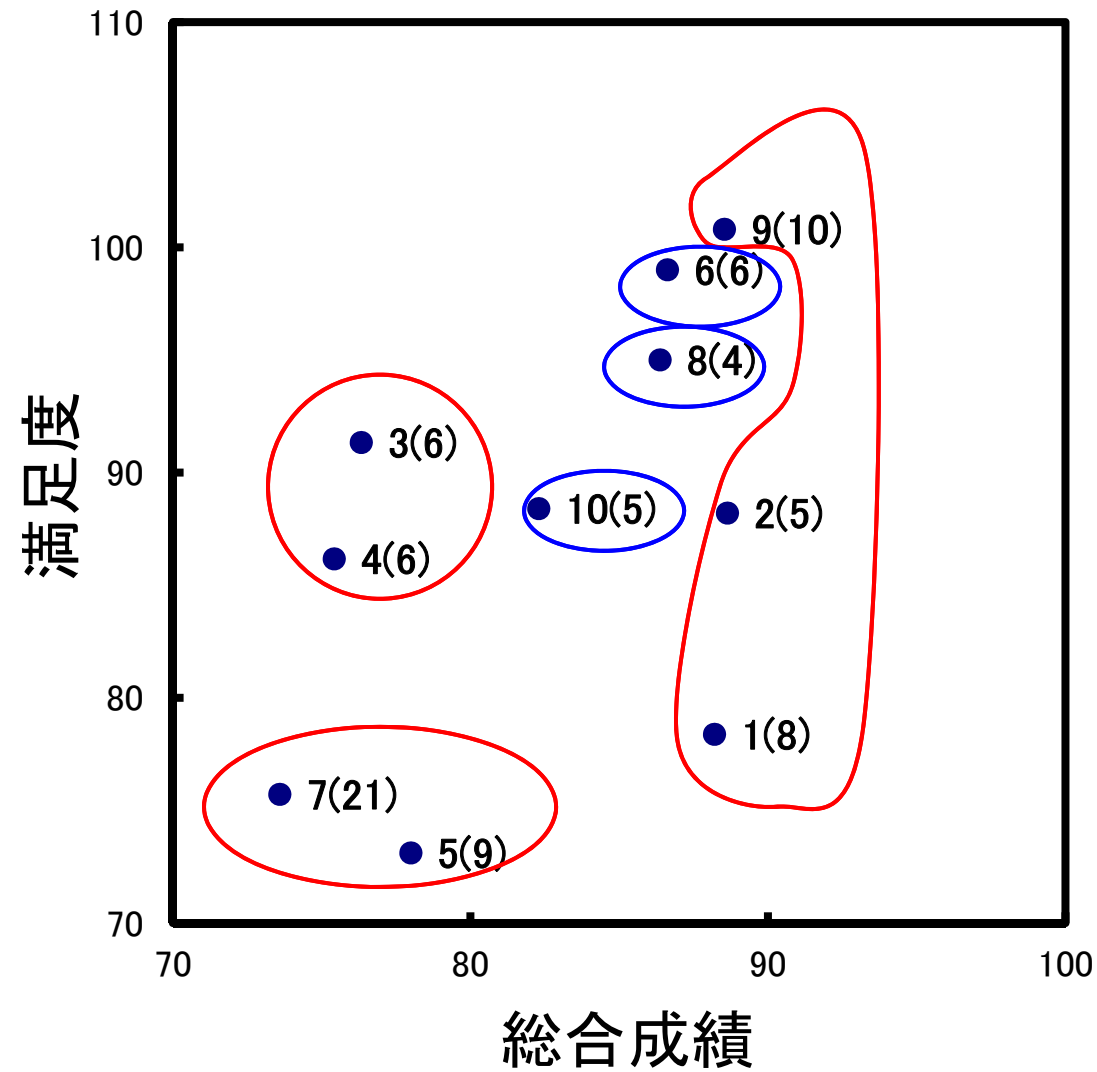
最長距離法: クラスタに属する個体間の距離の最長
＝クラスタ間距離



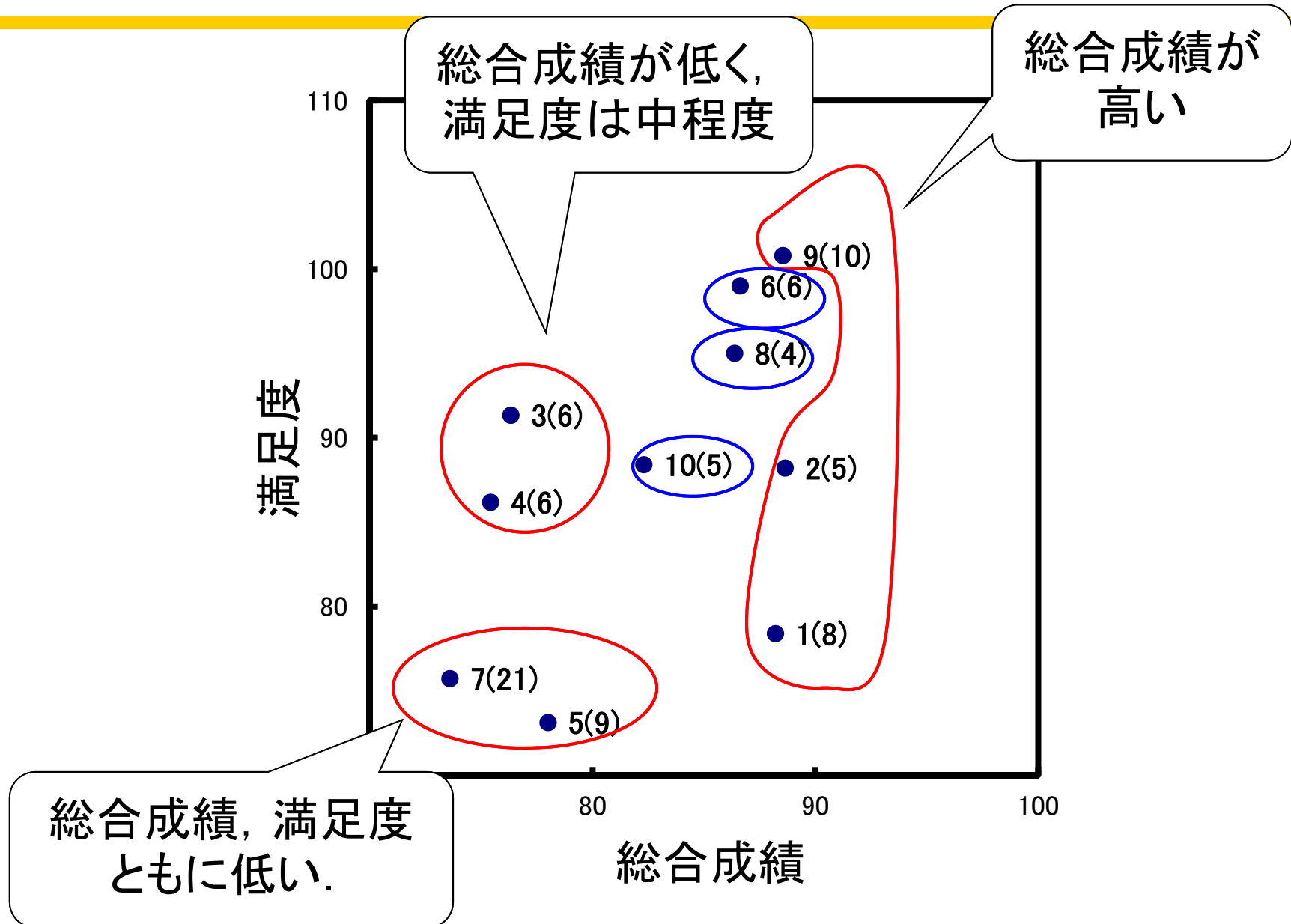
4.1 クラスタリング結果(最長距離法)



4.1 クラスタリング結果



4.1 クラスタリング結果



4.2 クラスタの特徴抽出（総合成績，満足度）

クラスタ	記述式（特徴語）	選択式
1,2,9	現在，取り組む，会社，構造，サーバー，日常，触る，教材，役，形成	▪ 講義内容はだいたい理解できた ▪ スライドはとても見やすかった ▪ レポートの分量はちょうどよかった
3,4	コンピュータ，講義，使う，自分，知識，必要，意味，大学，学ぶ，入門	▪ コンピュータを使いこなせれば原理は知らなくて良い ▪ レポートは難しすぎた
5,7	サイト，2年生，ボード，整理，特，野中，低級，グレード，数式，載せる	▪ 自分にとってこの講義は必要とは思わない ▪ 講義によってコンピュータを活用できるようになるとは全く思わない

4.2 クラスタの特徴抽出

クラスタ	記述式(特徴語)	備考
6	ソフトウェア, 装置, 方, 機器, 店, 接続, メイン, 目標, 転職, 認識	自由記述式より 職業: プログラマ・SE等
8	音楽, 閲覧, インターネット, 関係, 意味, ドライバー, ブレーキ, 制限, サポート	自由記述式より 職業: 映像・音楽関連
10	ソフトウェア, 部分, 基礎, 習得, 部署, 手軽, 集, 面倒, 不足, 様々	選択式より ・ 出席はとらないでほしい ・ スライドは役に立った ・ 毎回レポートにして欲しい

4.2 クラスタの特徴抽出

クラスタ	記述式(特徴語)	備考
6	ソフトウェア, 装置, 方, 機 器, 店, 接続, メイン, 目標,	自由記述式より 職業: プログラマ・SE等
8	職業に基づくコメント が多くそのために, 特殊な学生のクラス タが作成された.	自由記述式より 職業: 映像・音楽関連
10	ソフトウェア, 部分, 基礎, 習得, 部署, 手軽, 集, 面 倒, 不足, 様々	選択式より ・ 出席はとらないでほしい ・ スライドは役に立った ・ 毎回レポートにして欲しい

4.2 クラスタの特徴抽出

クラスタ	記述式(特徴語)	備考
6	ソフトウェア, 装置, 方, 機器, 店, 接続, メイン, 目標, 転職, 認識	自由記述式より 職業: プログラマ・SE等
8	音楽, 閲覧, インター ネット, 関係, 意味, ドライバー, ブレーキ, 制限, サポート	自由記述式より 職業: 映像・音楽関連
10	ソフトウェア 部分 基 足	選択式より 勉強方法がeLearningではなく, 参考資料を 元に自学を好む学生が抽出されている. 毎回レポートにして欲しい

- (1) クラスタ(1,2,9)・・・積極的で成績が良い, (3,4)・・・成績も満足度も中位, (5,7)は批判的で成績も満足度も低い.
- (2) クラスタ(6),(8),は成績・満足度ともに高いが特殊, (10)は中庸.

- (1) 総合成績の高いクラスタは満足度も高い傾向にある
(ただし, クラスタ(1)は常に存在する).
- (2) クラスタ(6),(8),(10)は特徴が把握できていない. 社会人学生の特徴を掴む必要がある.
- (3) 今後, 社会人学生向けのアンケート設計が必要である.

付録 クラスタ数 K の選び方

PLSI $K=7$ との比較 一致割合=0.61

	1	2	3	4	5	6	7
1,2,9	10	1	0	1	0	0	0
3,4	0	0	1	4	0	2	0
5	1	7	0	0	0	1	0
6	2	0	0	0	21	0	0
7	0	2	8	0	0	0	0
8	9	1	0	0	0	0	1
10	1	1	0	1	0	1	4

付録 クラスタ数 K の選び方

PLSI $K=6$ との比較 一致割合=0.70

	1	2	3	4	5	6
1,2,9	4	0	21	0	0	0
3,4	2	1	9	0	0	0
5,7	12	1	0	1	0	0
6	3	3	0	0	0	5
8	1	0	0	5	2	0
10	1	7	0	0	2	0