

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 ††
††† 早稲田大学理工学研究所 ††††

早稲田大学メディアネットワークセンター
早稲田大学創造理工学部

1 はじめに

近年、大学における授業改善を目的とした様々な取り組みがある。学生アンケートの分析により隠された学生のニーズや不満を抽出し、これを授業改善に結びつけることも有効な方法の一つとされている。

一方、数学的モデルを用いた情報検索技術は、文書検索だけではなく文書分類・文書クラスタリングにも応用されている。筆者らは確率的潜在意味インデキシング (Probabilistic Semantic Indexing) モデル [4] を選択型 (2 者択一や 5 段階評価など)・記述型 (与えられたテーマについて自由に記述する) が混在するアンケートの文書に適合するよう拡張し分析を行ってきた [2]。なお、このモデルは少ない文書数でよい分類性能を示すことが分かっている [3]。

本研究では、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する。授業モデル [1][5] を提案し、これに基づく選択型と記述型の学生アンケートを設計・実施し、多角的に分析して有効なデータを得ることを目的としている。主としてクラスタリングと特徴語抽出アルゴリズムを用いて分析した。その結果、総合成績と満足度の関係の一部を明らかにした。

2 アンケート分析手法

アンケートを多角的に分析するため、(1) 統計的手法の他に、(2) 特徴語抽出・特徴文抽出・重要文抽出手法と (3) PLSI モデルを用いた選択型と記述型の文書分類・文書クラスタリングアルゴリズムを開発した。図 1 にアンケート分析手法を示す。ここで、アンケート設計は次節で述べる授業モデルに基づいて作成される。なお、(1) の結果については別に報告する。

(2) は分類 (またはクラスタリング) されたクラス (またはクラスター) に属するメンバーが書いた内容を簡単に把握するために用いる。

3 授業モデルと拡張 PLSI モデル

3.1 授業モデル

アンケート結果の指標 (出力, 目的変数) を (学生の) 総合成績と (学生の授業) 満足度とした。これを説明する項目 (入力, 説明変数) として学生アンケート内容の他、学生の属性を用いる。図 2 に授業モデルを示す。

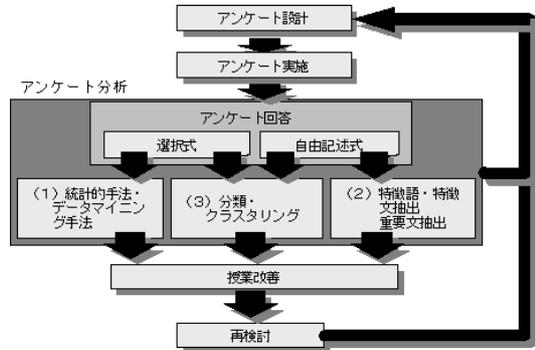


図 1: アンケート分析手法

3.2 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

$$P(d_i, t_j) = \sum_k P(d_i|z_k)P(t_j|z_k)P(z_k) \quad (1)$$

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

$$L = \sum_i \sum_j n(d_i, t_j) \log P(d_i, t_j) \quad (2)$$

を最大にする $P(z_k), P(d_i|z_k), P(t_j|z_k)$ を Tempered EM アルゴリズムを用いて求める [4]。EM アルゴリズムは初期値に依存するため、十分多数回の実行結果から尤度が最大となるものを選ぶ。なお、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、全体の文書単語行列 A は、重み $(0 \leq \lambda \leq 1)$ により

$$A = [\lambda G \quad (1 - \lambda)H] \quad (3)$$

で与えられる。

3.3 特徴語抽出・特徴文抽出・重要文抽出手法

上記 EM アルゴリズムを実行した結果から

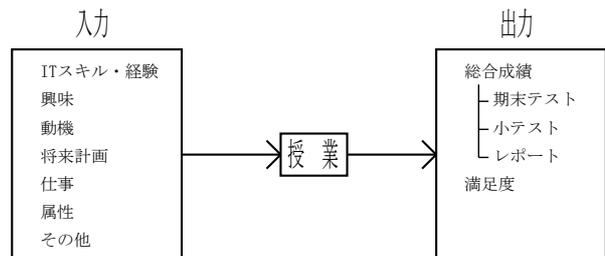


図 2: 授業モデル

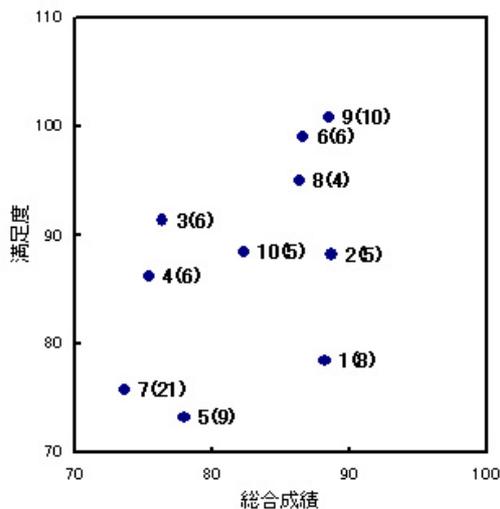
On analysis for student questionnaire with fixed format and free format

† Shigeichi HIRASAWA, Faculty of Information Technology and Business, Cyber University

†† Takashi ISHIDA, Media Network Center, Waseda University

††† Gendo KUMOI, Waseda Research Institute for Science and Engineering

†††† Masayuki GOTO, School of Creative Science and Engineering, Waseda University



・ $X(Y): X$ はクラス番号, Y は人数を示す.

図 3: PLSI クラスタリング

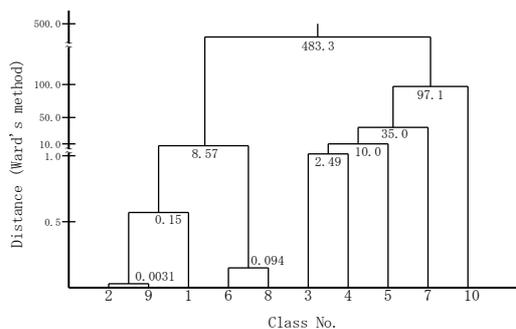


図 4: 総合成績

- $P(t_j|z_k) - P(t_j)$ を t_j のスコアとする. ある文に現れた t_j のスコアの和をその文のスコアとする. このとき, 大きいスコアを持つ t_j を特徴語とし, 大きなスコアを持つ文を特徴文とする.
- 文書中の文と他の文の類似度を測定し, 類似度の和をその文のスコアとする. 最大のスコアを持つ文をその文書の重要文とする.

4 アンケート分析結果

ここでは, $\lambda = 0.5, K = 10$ とする. $I = 80, J = 2052$ である.

4.1 クラスタリング結果

総合成績と満足度の散布図を図 3 に示す. また, 総合成績, 満足度を基準とし, ウォード法によりクラスを併合した. 総合成績のデンドログラムを図 4 に示す.

4.2 クラスタの特徴抽出

表 1 に, PLSI のクラスタリング結果を総合成績-満足度の散布図から特徴的であったクラスの特徴語の一部を示す. また, ウォード法により総合成績, 満足度を基準に, クラスを併合した結果について, 表 2 にそれぞれの平均点を示す.

表 1: クラスタの特徴語

クラス	特徴語
7	複雑, 最初, 演習, 間違い, 落ち込む
9	理論, プログラマ, 現業, 能力, 強まる

表 2: 総合成績, 満足度平均

	クラス	総合成績	満足度
総合成績	1,2,6,8,9	87.86	92.42
	3,4,5,7,10	75.95	79.89
満足度	6,8,9	87.53	99.10
	2,3,4,10	80.24	88.55
	1,5,7	77.72	75.66

5 まとめと考察

- 表 1 の各クラスの特徴語から, クラス 7 では「複雑」や「最初」など授業内容に初めて触れることから苦手意識が現れた単語が多く抽出された. クラス 9 では「プログラマ」や「現業」など職業として授業内容に近いことを行っているなかで, 授業内容に満足をしていることが表われた単語が多く見られた. ここで図 3 を見ると, クラス 9 は成績, 満足度共に高く, クラス 7 は総合成績, 満足度共に低く, これらの特徴に則した単語が抽出されたことがわかる.
- 表 2 より, 総合成績によって併合を行った結果, 総合成績の高いクラス, 低いクラスの 2 クラスとなった. また, 満足度によって併合を行った結果, 総合成績が高く満足度も高いクラス, 平均的なクラス, 総合成績は低く満足度も低いクラスの 3 クラスとなった. これらは, 図 3 よりも明らかであり, 総合成績を入力に持たない PLSI がアンケートデータから授業の理解度や満足度に則したクラスタリングを行えている結果と考えられる.

6 むすび

PLSI モデルでアンケートデータからクラスタリングした結果, 総合成績と満足度には一定の関係が認められた. 総合成績の高いクラスは満足度も高い傾向がある. この結果は, 総合成績と満足度の高低は, アンケート回答にも何らかの形で現れており, 特徴語によって各クラスの特徴を把握することが可能である. 最後に, 日頃ご協力を頂くサイバー大学コンテンツ制作センターの皆様に感謝します.

参考文献

- [1] 後藤正幸, 酒井哲也, 伊藤潤, 石田崇, 平澤茂一, “ 選択式・記述式アンケートからの知識発見,” 2003 年 PC カンファレンス, 鹿児島, 2003 年 PC カンファレンス予稿集, 2003.8.
- [2] Shigeichi Hirasawa and Wesley W. Chu, “ Knowledge acquisition from documents with both fixed and free formats,” Proc. of 2003 IEEE Int. Conf. on System, Man and Cybernetics, Washington, U.S.A., 2003.10.
- [3] 平澤茂一, 石田崇, 足立敏史, 後藤正幸, 酒井哲也, “ 文書分類法とそのアンケート分析への応用,” 経営情報学会, 2005 年度春季全国研究発表大会予稿集, 東京, 2005.6.
- [4] T. Hofmann, “ Probabilistic latent semantic indexing,” Proc. of SIGIR '99, ACM Press, pp.50-57, 1999.
- [5] 伊藤潤, 石田崇, 後藤正幸, 酒井哲也, 平澤茂一, “ PLSI を利用した文書からの知識発見,” 2003 年 FIT, 江別, 2003 年 FIT 論文集, vol. 2, pp.83-84, 2003.9.