

## 台湾における学生アンケート分析の中間報告

## Interim report of student questionnaires analysis in Taiwan

蔡宜静\* 石田崇† 酒井哲也‡ 黄永東§ 平澤茂一†

\*立徳管理学院応用日語学科 †早稲田大学理工学部経営システム工学科

‡(株)東芝研究開発センター §立徳管理学院工業管理学科

要旨：本研究では、台湾における学生アンケートの自由記述に対して情報検索技術を積極的に適用して、その分析手法および分析結果について報告する。このアンケートに対して、情報検索における擬似適合フィードバック技術を用いることによって効率的な分析が可能であり、有用な情報が得られることを示す。

Abstract: This paper analyzes texts with free format from Taiwan student questionnaires. By effectively applying to the information retrieval technology, we will show that application of Pseudo-Relevance Feedback is able to analyze the questionnaires efficiently and obtain the useful information.

## 1 はじめに

蓄積されたテキストデータから知識を発掘する技術としてテキストマイニングが注目されているが、これをアンケート分析に適用することによって、有用な情報を獲得できる可能性がある。日本においてテキスト自動解析手法を教育現場に適用した例として、例えば長坂ら[1]や森田ら[2]が挙げられるが、これらの手法は記述式データのみが解析の対象である。また、平澤らがここ3年間授業改善のための選択式・記述式学生アンケート分析結果を報告したものがあげられる[3]-[10]。

本研究では、台湾の学生を対象とした半年にわたる講義「コンピュータ概論」の初回および最終回に実施した学生アンケートの自由記述式回答を自動分析したのでその分析手法および分析結果について報告する。

筆者らは、個々の学生の自由記述と成績とを対応づけ、これらをトップダウンにグループ化し、[5]の手法に基づき、情報検索における検索条件拡張技術である擬似適合フィードバックを用いて代表語・代表文を抽出した。特に本稿では学生の期末試験の成績をひとつの手掛かりとし、学生アンケートの自由記述から学生の考えや要望を抽出することを目的としており、各学生グループの代表文抽出は、複数文書要約の立場に近い。このような比較的単純な方法により効率的な傾向分析が行えることを示す。本研究における最終目標は台湾の学生アンケートの分析報告から大学教育における授業改善のための知識獲得である。

## 2 データ概要

本研究では、立徳管理学院「工業管理学科」「情報工学科」「日本語学科」および淡江大学「情報管理学科」の一年生の必修科目「コンピュータ概論」を分析の対象として、初回および最終回の選択式・記述式設問を含む記名式アンケート

を実施した。自由記述の長さは100字以上に制限した(日本語に換算すると約250字)。これにより文書長が比較的均一なテキスト集合が得られた。効率化の点からアンケート用のホームページを作成し、学生にはWeb上で回答してもらう。自由記述の設問内容は以下のとおりである。

- ・初回アンケート：Q1 コンピュータ知識、Q2 将来設計、Q3 講義のイメージ、Q4 分野への意欲、Q5 講義への要望、Q6 その他 (6 質問)
- ・最終回アンケート：Q1 将来設計、Q2 コンピュータの印象、Q3 分野への意欲、Q4 講義の感想、Q5 その他 (5 質問)

(このアンケートは下記Web Siteに掲載されている。

<http://www.iim.leader.edu.tw/2006InternationalSeminar/Main.htm>)

このアンケートは日本で実施された学生アンケート[4]-[10]と同様の内容で、人手によって中国語に翻訳したものである。なお、初回・最終回アンケートともに期限内に提出し、かつ期末試験を受験した学生は107人であり、これらを分析対象とした。また、アンケートは記名式であるため、各学生の自由記述と最終成績とを対応づけて分析を行うことができる。

## 3 分析方法

## 3.1 擬似適合フィードバックによる代表語・代表文抽出手法

[5]では、情報検索における検索条件拡張技術である擬似適合フィードバック(Pseudo-Relevance Feedback: PRF)を適用した代表語・代表文抽出手法を提案している。擬似適合フィードバックにおいて、初期検索結果の適合文書中から新たに検索条件(単語)を追加する際に計算される単語スコアの算出式を用いる。これを用いることにより、ある特定の文書集合に対して、スコアの高い順に単語抽出することにより、それをその集合の特徴語とみなすことができる。また、文のスコアを、文を構成する単語のスコアの合計として与える事

により、同様に代表文を抽出することが可能である。

### 3.2 代表語・代表文抽出によるアンケート分析

「概論コース」(CGen)と「専門コース」(CSpe)のどちらの講義内容を希望するかという選択式の設問(最終回アンケート)に対して、「日本語学科」の学生が前者を選ぶ比例は、96%(24/25人)にも達した。「日本語学科」はほぼ「概論コース」を希望し、「工業管理学科」「情報工学科」「情報管理学科」は「概論コース」および「専門コース」希望の二つに分かれる。そこで、前者の文系の特性を考慮に入れながら、後者の理系学部の特徴抽出およびクラス分割問題を考える。

初回(6質問)および最終回(5質問)アンケートのデータを別個のデータセットとして扱い、各データセットに含まれる文系学生24人分と理系学生82人分の{自由記述、最終成績}の対を用いる。カテゴリに基づいてトップダウン的に学生をグループ化し、各グループの代表語・代表文を抽出する事によってその傾向を把握する。

なお、すでに開発されている日本語上のアンケート解析システムを用いるため、学生の回答は中国語から日本語に翻訳して分析を行う。

## 4 分析結果および考察

### 4.1 分析結果

最終成績に基づいて85点以上、75点以上85点以下、65点以上75点以下、65点以下の4つのカテゴリに学生を分類するものとする。本稿では、85点以上のカテゴリと65点以下のカテゴリのみに着目し、両カテゴリに属する文系学生と理系学生のそれぞれについてCgenとCspeと回答した学生のグループに分類し、それぞれの自由記述の比較を初回および最終回アンケートについて行った。代表語・代表文の一部を表1および2に示す。

表1と表2から、文系・理系学生について以下のような大まかな傾向をつかむことができた。

#### (1) 文系学生かつCGenの場合

- ・ 85点以上：ホームページの製作に興味を持った学生が多い。講義ではOfficeのソフトウェアの熟知を将来の仕事に実際に活用する事への期待がある。
- ・ 65点以下：ハードウェアとシステムの基礎構成への理解があった。授業終了後はOfficeの実用性に納得するが、教師の教え方に満足しなく、採点の方法について試験よりレポートのほうが良いと

述べた意見が読み取れる。講義の感想としては自分が文系的な人間だから、やはりコンピュータの学習は難しいという学生の認識がある。

#### (2) 理系学生かつCGenの場合

- ・ 85点以上：ハードウェアを組み立てる経験を持った学生がいる。英語原文と中国語のテキストで概論を教える方法に同感があったが、コンピュータを実際に操作する授業の運営方を望む学生の意見が読み取れる。授業の終了後も理論以外の分野に対する学習意欲がある。また、授業内容は大学院の試験の役に立つという学生の認識がある。
- ・ 65点以下：講義の開始時にはコンピュータと自分の所属と関連があると思い、将来コンピュータにかかわる仕事に就きたいと単純に考えている。ゲームやソフトウェアに興味がある。授業終了後は授業の運営方法に充実感がある。コンピュータの領域は広範だと思い、特にホームページの製作に興味がある。講義を大学院の合格への期待がある。

#### (3) 理系学生(CSpe)の場合

- ・ 85点以上：自分の所属する学部やプログラムの設計に大変興味があり、コンピュータの基礎原理への知識やハードウェアの組み立ての経験を持った学生がいる。講義ではコンピュータを将来の仕事に活用する事への期待がある。同じく英語原文のテキストで授業を教える方法に納得するが、クラスメイトの素質の格差を気にし、授業の運営方式について、クラス分割の方法を提案する意見が読み取れる。

- ・ 65点以下：講義内容への認識があるが、コンピュータはインターネットを使用するための道具としてしか捉えていない。英語原文のテキストで概論を教える方法に同感があったが、授業の内容を「抽象的」だと感じ、自分の授業に臨む姿勢が「受身的」だと言った感想が目立つ。授業終了後は講義の内容を大学院に進学するための授業だったと感じているが、教師の教え方に満足せず、採点の方法について試験よりレポートのほうが良いという声がある。また、ゲームの設計に興味がある。

### 4.2 考察

#### (1) 文系と理系学生の比較

- (a) 65点以下の理系学生(CGenとCSpeの両方)はソフトウェアやホームページの製作に興味を持つ点は、文系学生のそれに似た傾向が読み取れる。
- (b) 教師の教え方に満足せず、採点の方法について試験よりレポートのほうが良いという意見は、

65点以下の文系学生と理系学生(CSpe)に見出せる。

(c) 65点以下の文系学生は授業開始時にハードウェアとシステムの基礎構成への理解があった点は、85点以上の理系学生(CGenとCSpeの両方)と同じである。

(d) 表には掲げていないが、85点以上の文系学生に唯一「専門コース」(CSpe)の講義内容を選択する学生が「もしかするとコンピュータの概論も英文のように、能力でクラスを分けて、それから各クラスの学生と教員が討論して、この学期の授業の内容を決定する」と述べている意見は、85点以上の理系学生(CSpe)の感想と類似している。

#### (2) 理系学生CGenとCSpeの比較

(a) 85点以上のCGenとCSpeの両方とも講義の開始時にはハードウェアを組み立ての経験を持ち、英語原文のテキストで授業を教える方法に同じく納得したが、CSpeの場合は自分の所属する学部やプログラムの設計に大変興味があるという学生の認識がある。

(b) 授業内容は大学院の試験の役に立つという学生の意見は、85点以上の学生(CGen)と65点以下の学生(CGenとCSpe)の代表文に見出せることができる。タームの集合に基づく代表文抽出という我々の単純なアプローチ方法でも違う成績のカテゴリに分類した記述を発見することが可能であることがわかる。なお、自由記述で{テーマ、試験、精密、指導、計算、JAVA}などといった単語を使う学生は相対的に成績が良いようであるのに対して、{概論、レポート、広範、進度、練習、復習、説明、面倒、解決、ゲーム、チャット}などといった単語を使う学生は相対的に成績が悪いようである。

以上、文系学生アンケートの自由記述内容と比較して、理系のCGenとCSpeとで両方に傾向の違いが見られる。よって、テキストのみからの分類も有効であると考えられる。上記のような分析結果を学習データとして用い、来年度の「コンピュータ概論」の講義を対象としてアンケートを実施した時点で、これに応じて授業の内容を変更したり、工夫したりし、「概論コース」と「専門コース」のクラスに分割するための参考資料となるといった応用が考えられる。

## 5 まとめ

本研究では、情報検索における擬似適合フィードバック技術を用いて学生アンケートの記述式デ

ータを分析し、学生のクラス分割問題へのアプローチとした。文系・理系学生のそれぞれについて代表語・代表文抽出を行った結果、授業改善やクラス分割に役立つ情報を得られることが確認できた。

2006年春学期では、予備実験として行ったアンケート回答を用いたが、今後より有効なテキスト処理を可能とするために、

- (1) 長文の自由記述文の収集
- (2) 対象学生の増加
- (3) 中国語上での分析システムの開発

などが必要である。

## 参考文献

- [1] 長坂悦敬, 阿手雅博, “記述問題の自動評価を目指した教育支援情報システムによる Interactive Education,” 情報教育方法研究第3巻, pp. 37-42 (2000).
- [2] 森田直樹, 北英彦, 高瀬治彦, 林照峯, “記述式の解答を即時に講師が把握するためのシステム,” FIT2002 情報技術レターズLN-1, pp. 233-234 (2002).
- [3] 伊藤潤, 石田崇, 後藤正幸, 平澤茂一 “文間の単語共起類似度を用いた重要文抽出手法,” 2002年FIT論文集, pp. 83-84 (2002).
- [4] 石田崇, 伊藤潤, 後藤正幸, 酒井哲也, 平澤茂一 “授業モデルとその検証,” 経営情報学会2003年秋季全国研究発表大会予稿集, pp. 226-229 (2003).
- [5] 酒井哲也, 伊藤潤, 後藤正幸, 石田崇, 平澤茂一 “情報検索技術を用いた効率的な授業アンケートの分析,” 経営情報学会2003年春季全国研究発表大会予稿集, pp. 182-185 (2003).
- [6] 平澤茂一, 石田崇, 伊藤潤, 後藤正幸, 酒井哲也 “授業に関する選択式・記述式アンケートの分析,” 私立大学情報教育協会平成15年度大学情報化全国大会, pp. 144-145 (2003).
- [7] 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,” PCコンファレンス, pp. 43-46 (2003).
- [8] 酒井哲也, 石田崇, 後藤正幸, 平澤茂一 “自然言語表現に基づく学生アンケート分析システム,” FIT 2004 予稿集N-021, pp. 325-328 (2004).
- [9] 石田崇, 足立鉦史, 後藤正幸, 酒井哲也, 平澤茂一 “情報検索技術を用いた選択式・自由記述式の学生アンケート解析,” 経営情報学会2004年秋季全国研究発表大会予稿集, pp. 466-469 (2004).
- [10] 平澤茂一, 石田崇, 足立鉦史, 後藤正幸, 酒井哲也 “文書分類技法とそのアンケート分析への応用,” 経営情報学会2005年春季全国研究発表大会予稿集, pp. 54-57, 東京, 2005年6月.

表1：文系学生と理系学生（CGenとCSpe）の85点以上vs65点以下の比較（初回アンケート）

85点以上（文系学生）代表語：従事 ホームページ スタイル 日本 将来 概論 コンピュータ Word 興味 Excel		
85点以上（理系学生CGen）代表語：テーマ 大学 先生 助け 修理 考え 試験 年生 中国（語） コンピュータ		
85点以上（理系学生CSpe）代表語：理解 設計 興味 最後 多く 学部 勉強 自身 個人 本当		
	順位	代表文
文系	1	私がホームページの設計に対してとても興味があって、自分スタイルのウェブサイトを設計することができるためだ。
	8	コンピュータの概論は、コンピュータの由来、系統的な発展と機能の増加を紹介して、ハードウェアの設備、ソフトウェアの応用を紹介して、計算方法、コンピュータの科学技術、OS、文書処理ソフトウェア、Word、Excel、PowerPoint、ネットワークなどを教える授業だ。
理系CGen	4	私達の先生の授業の方法は悪くなくて、教えたのは非常にきめ細かい。その上中国語と英文の方法を使って教えるから、非常に悪くないと感じるのだ。
	5	興味があるため、ハードウェアの組み立てと修理の試験をしに行く。
	6	コンピュータを利用しながら、この講義の内容をどのように活用できるかを教えてくれれば、より勉強の助けになるだろう。
理系CSpe	1	自身はコンピュータに対してとても興味があり、その上情報工学部も勉強するので、もちろんプログラムを設計するような進路を希望する。コンピュータを組み立てる前に、自分のニーズに合わせて、それぞれの異なっている部品に対応する異なる資料を探す。それぞれのメーカーのブランド、型番の違いを理解し、かつ比較する。最後の選択が決まってから買う。
	3	私は以前から今まで習ったことから、二進法、OSとネットワークbus、およびいくつかの伝送と読み方式について、最も浅い理解を持っていた。
	21	
65点以下（文系学生）代表語：ドライブ ROM CD ボード ディスク CPU インストール 咄雑 語学 XP WIN		
65点以下（理系学生CGen）代表語：確か 情報 ゲーム 同じ 期待 助け 多く 電子 工程 好き 活発		
65点以下（理系学生CSpe）代表語：通信 英文 面倒 クラス 概論 解決 説明 教科 運算 レポート		
文系	2	初めはハードウェアの理解から始めて、CPU、CPUボード、キャプチャカードは何かを知った…ハードディスク、CD-ROMドライブの装置。そしてどのようにコンピュータのコネクタを挿し込むかという問題を知った。
	4	OSはXPでなければ（つまりXP以下の、例えばWIN2000、WIN ME、WIN98など）、システムをインストールすることは比較的にやや面倒だ。
理系CGen	1	でもこの授業は確かに自分の所属する学部に対して多くの助けをしてくれる。
	3	情報工程を専攻して、もちろんコンピュータの方面のものに従事することを望む。
	7	私は小さいプログラムに似ているものを設計するのが好きで、あるいはゲームのあの方面の領域はとても面白いと感じる。
	17	いつも永久に変わらない応用のソフトウェア、一定の規則に従う同じような形式の運用より、この方面が私に新鮮な感じをさせて、私が比較的好きなので、比較的に飽きることはない。
理系CSpe	1	オンライン通信システムを利用すると、多くの問題を解決することができて、そして時間と金を節約できる。
	3	課程はもともとコンピュータの基本の構築と演算を説明している。
	7	コンピュータの概論の授業で自分の英文の能力は更に拔群になることを望んで、第2言語の学方法は、私を更に熟練させる。
10	クラスのいくつかの授業課程は、まったくはつきりなくて、ほぼ受身で聞いているだけだ。	

表2：文系学生と理系学生（CGenとCSpe）の85点以上vs65点以下の比較（最終回アンケート）

85点以上（文系学生）代表語：将来 優位 自然 眼光 熟知 ホームページ アニメーション コンピュータ 操作 方式		
85点以上（理系学生CGen）代表語：点数 先生 年生 精密 指導 計算 教育 助け 変化 JAVA		
85点以上（理系学生CSpe）代表語：基礎 専門 概論 英文 多く 学年 意義 学部 学生 試験 困難 関係		
	順位	代表文
文系	2	就業の方面で、その他のコンピュータを分らない人あるいはコンピュータを熟知しない人より、多くの優位を占める。
	3	活発とアイデアを兼ねるホームページの内容は、一般のものとは異なる特色を現すことができ、1歩進んでネット友達的眼光を捕らえる。
	7	ホームページの設計、Flashアニメーション。
理系CGen	1	後で大学2、3年生の課程の中は基礎を築いて、大学院の試験をうける時に、コンピュータ概論の中のいくつかの知識は多くの試験問題にも出てくるので、やすやすといくつかの基礎の点数を押さえることができる。
	8	授業の方式は非常に私に適すると感じて、そのため先生の授業の時、いつも多くの利益を得ると感じる。教育の内容に、いくつかの基礎理論の以外、実用的なものを学びたい。
	16	JAVAと資料の安全などの部分に興味があるだろう。文書の処理は普通のなくてはならない基礎だ。人工の知恵も試みるかもしれない。
理系CSpe	1	違う程度の学生は、すべて相当する進歩的な空間をさせるため、コンピュータ概論は英文のように、ABクラスに分けて授業をすることができるかどうか。
	11	この授業は私にとってかなり意義がある。
	12	自身は情報工学の学部の学生だから、将来の道もコンピュータとたくさんの関係があるものだ。
	14	授業の方式は大抵悪くないが、もし英文で教えることができれば、もっと良いと思う。
65点以下（文系学生）代表語：レポート 完全 自分 製作 全世界 ステップ PowerPoint 写真 確か 製図		
65点以下（理系学生CGen）代表語：大学院 先生 理解 機会 復習 仲間 広範 興味 状態 収穫		
65点以下（理系学生CSpe）代表語：チャット 明らか 成績 デザイナー 練習 方向 進歩 該当 設計 内容 一部		
文系	2	もした試験で成績を評定するならば、学生はただいくつかのステップを覚えるだけ、どのように他の各種のアプリケーションを使うかを完全に理解しないからだ。
	3	Word、Excel、PowerPointはすべてとても実用的で、平日のレポート或いは宿題について、すべて非常に便利だ。
	22	でも非本科生ではない私たちにとって学習の上で確かにいっぱい骨が折れる感じがする。
理系CGen	2	その上ホームページが小さいときからわりと多く接触したから、興味は相対的にも大きい。先生は授業でまじめな人で、その上小テストの回数はとても多くて、私達は授業の後で復習するよい習慣を身につける。
	4	コンピュータの領域はかなり広範で、3D、アニメーション、音楽、プログラムを書くことなどに対して一定の程度理解ができるように望んで、これらはすべて自分の足りないものだ。
	13	後の期待は、順調にすべての科目をパスできることを望んで、困難な履修単位をもらって、いい大学院に合格することができる。
理系CSpe	1	先生は多くの試験問題を探し集めて、該当する部分の授業をする時、学生に練習させて、学生に普通の試験の方向と変化を理解させることができるし、これは将来の大学院の試験に対しても、少しの助けがあるかも知れなからう。
	3	ゲームの関係のために、私はゲームのデザイナーになりたいくて、ゲームを設計することができるからだ。
	8	試験は決してこの学生のこの課程の能力を明らかに示すのではない。
	12	先生は一回だけ教えてから二度と言わないから、私はまた先生の進歩についていけないのに、先生はすでにまた新しいものを教えた。

\* 台湾における学生アンケートの記述式回答は蔡が北京V5にかけて翻訳したものを点検し、日本語に近い文章表現に再修正したものである。