

選択式・記述式アンケートからの知識発見

後藤正幸[†]，酒井哲也[‡]，伊藤 潤^{*}，石田 崇^{*}，平澤茂一[§]

[†] 武蔵工業大学 環境情報学部，[‡] (株)東芝 研究開発センター，

^{*} 早稲田大学理工学研究科，[§] 早稲田大学 理工学部

[†] e-mail: goto@yc.musashi-tech.ac.jp

1. はじめに

大学における授業改善を目的としたアンケート調査は従来から行われており，実際に授業改善に役立つことが多い事例から明らかになっている．しかしながら，依然として授業評価の平均的な値を求めたり，学生間のばらつきを調べるといった分析手法が多用されているのが現状であろう．自由記述式のアンケートに至っては，多く出された意見を人手で抽出するといった方法が取られるのが一般的である．

一方，近年の情報技術の発展に伴い，アンケート結果を電子ファイルの形で得ることが技術的に容易になっている．例えば，Web ページを作成し，Web 上でアンケートを記述してもらうことにより，容易にアンケートデータをテキストデータとして得ることができる．これらのアンケートには選択式と自由記述式が考えられるが，選択式は分析が容易であるが学生の自由な発想は抽出できず，自由記述式は自由な発想は得られるが，自由記述のテキストを自動的に分析する方法が確立しておらず，人手による分類や意見の列挙などの手間がかかるのが実情である．蓄積されたテキストデータから有用な知識を抽出する技術であるテキストマイニングは，主に膨大なテキストデータを分析することを目的としてさかんに研究がなされている[1]．しかしこのような技術を，授業改善のためのアンケート調査結果からの知識発見問題に適用するためには，対象問題の構造を正しくモデル化し，新たな方法論を構築する必要がある．

ここでは，大学における情報関係の講義を想定して議論を進める．情報関係の講義は，学生の興味や事前知識に大きなばらつきがあり，多人数の学生に対して効果のある講義を行うことが比較的難しい．コンピュータの初心者や応用に興味のある学生とより専門的にコンピュータの中身まで理解したい学生とでは，かなりの温度差がある．とくに，意欲はあっても成績につながらない学生やそもそも興味のない学生，最初からコンピュータを使いこなしている学生を一同に介して講義を行う場合，これらの全ての学生に対し効果的な講義を行うことは無理であろう．著者らはすでに前報において，情報探索技術を

用いた効率的な自由記述式アンケート分析手法の提案を行い，その有効性を示している[2]．ここでは効率的にある種の情報を抽出することに成功しているが，一方でアンケート作成方法については論じておらず，また選択回答式と自由記述式のアンケートを統合して知識発見を行う手法について改善の余地があった．

本稿では，授業改善を目的としたアンケートを Web 上で採取する状況を想定し，1) 効果的なアンケート構成方法，2) 選択式と自由記述式が混在したアンケート結果を分析する手法を提案する．この方法を，大学学部生を対象とした「コンピュータ工学」の授業で実際に適用することにより，その有効性と今後の可能性について考察を与える．その結果，学生に対するアンケート結果を積極的に活用することにより，大学における情報系科目の講義改善に対してかなりの示唆を与えることが可能であることを示す．

2. 状況設定

本研究で対象とするのは，大学学部生対象の情報系科目である．今回はとくに早稲田大学で開講している「コンピュータ工学」の講義を対象とする．本講義は 130 名前後を対象としており，全員に対して効果的な講義を行うことが本質的に困難な状況である．このような大人数の受講生に対する情報系教育の場は，近年どの大学でも生じており，どのような情報教育が有効なのかを議論していくことは重要であろう．本研究では，学生に対するアンケートを Web 上で行うことにより，この結果を積極的に授業改善や学生の成績予測等に用いる方法について検討する．アンケートには選択式と自由記述式が考えられるが，これらを組み合わせてアンケートを構成する方法について述べ，さらにこれらを統合的に分析する方法を提案する．教育分野では，記述問題に対する学生の解答と模範解答との類似度から採点・評価を支援するシステムが提案されているが[3]，本稿ではアンケートから効果的な教育支援を行うことを目的としている．

講義の質に関する評価の視点



図1. 講義の質に関する評価の視点

3. アンケートの作成手順

3-1 前提となる効果的講義モデル

ここでは、特に情報関係の大学講義の質を向上させることを目的とし、情報技術やテキストマイニングの技術を用いて授業改善に結びつける方法について述べる。まずは、アンケート作成にあたり、講義の質の評価について構造を明確にしておく必要がある。講義の質については、図1に示すように、少なくとも「学生」、「教員」、「大学・企業」といった異なる視点が存在する。本稿における講義の質向上は、教員が大学・企業の求める講義の質を満たしつつ、学生にとっても講義の質を最適化することである。ここでは、講義の質の評価項目として、「学生の満足度」と「学生の成績」を採用する。

図2に現状の講義の形態を示す。現状では、全ての学生に対し画一的な講義を進めており、この中で「学生の満足度」と「学生の成績」を向上させるにはおのずと限界がある。そこで、図3のように「各学生の特性に合わせた講義により、講義の評価は向上する」という仮説を立て、このもとで効果的な講義のあり方を考える。この仮説は、教育の経験から一般的に納得のいく仮説であり、きめの細かい教育が有効であるという主張の根拠となっている。

現状

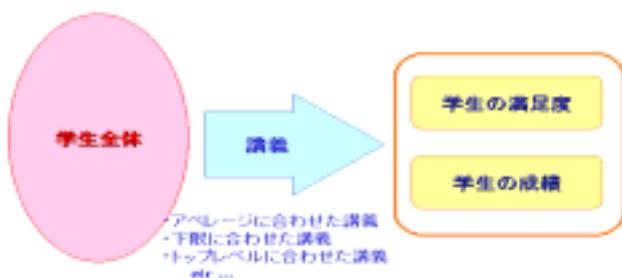


図2. 現状の多くの講義形態

仮説

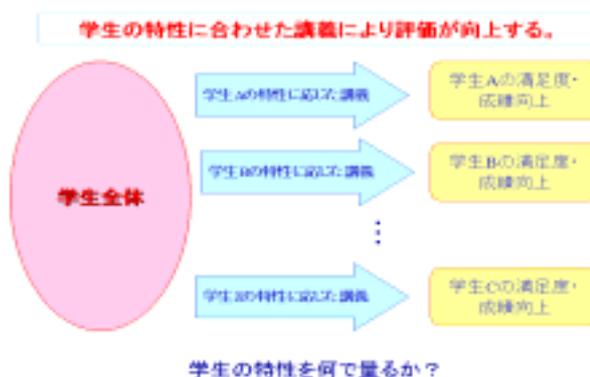


図3. 効果的な教育のモデル(仮説)

図3に示した効果的と考えられる教育モデルにおける一つの課題は「コスト」対「効果」のトレードオフの関係であろう。一人の教員が、130名近い学生の一人一人の特性を把握し、一人一人に最適な教育法を提供するのは不可能である。このようなきめの細かい教育は、非常にコストがかかると言えよう。重要なのは現状のリソースで、どこまで学生を細分化し、適切な教育方法を提示するのが最も効率的で、かつ効果的か?ということである。わかり易く言えば、クラスを二つに分けてそれぞれに異なる方法で講義を行うべきなのか、あるいは三つ以上に分けるべきなのかという問題であり、さらにはどのような視点から学生を分ければいいのか、という問題である。

3-2 アンケート作成のための基礎モデル

本稿では、学生の特性(特徴)によって、講義の評価である「満足度」と「成績」は変わってくるものとする。そして、効果的な講義のあり方を分析するために、第1段階として講義前の学生の特性調査アンケート、及び第2段階として結果である学生の満足度と成績の2つをデータとして採取することとした。まず、第1段階のアンケート項目作成のため、「満足度」と「成績」に影響を与えるであろう学生の特性(特徴)の構造をモデル化した(図4)。ここでは、前年度の講義アンケート結果をもとに議論を重ね、「個性に関する項目」と「講義に対して抱くイメージの項目」に分類し、さらに「個性に関する項目」を「コンピュータに関する事前知識」、「学習意欲」、「興味」、「将来設計」に、「講義に対して抱くイメージの項目」を「講義のイメージ」、「講義方法に対する要望」に分類した。この分類のもとに、それぞれの項目を抽出するための質問項目を設定した。具体的には、例えば学生の「事前知識」を調査するために、すでに知っている専門用語などをあげてもらおうなどの質問項目を用意する。

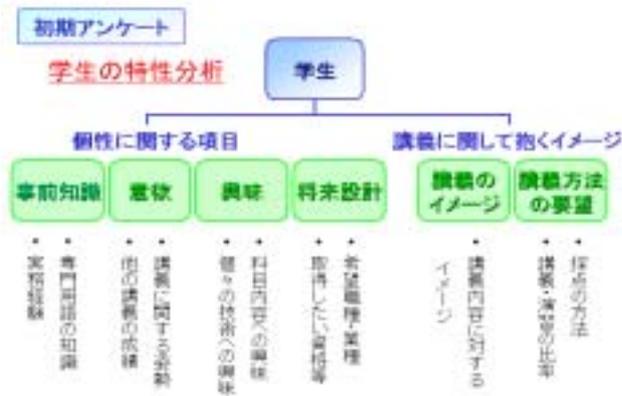


図4．初期アンケートのための学生の特性モデル

3-3 アンケート項目の作成方法

しかしながら、ここで一つの問題が生じる。講義にあたって、関係すると思われる専門用語や技術については、教員側で提示し、学生の事前知識レベルを選択式で回答してもらうのが適切である。しかし、情報技術に関わる専門用語は多岐にわたり、全てを列挙すれば膨大な選択肢となってしまふ。また「興味」や「将来設計」、「講義のイメージ」、「講義方法の容貌」等については多岐に渡るだけでなく、学生によってこちらが思いもつかない回答をする場合がある。これらは、自由回答式のアンケート作成により、情報を抽出できる。そこで、本稿では各設問に対し、選択式と記述式を適切に混在させてアンケートを行うこととした。選択式は、設問内容によってリストから該当する項目を選択する方式とSD法とを併用する。

ここで、大量の自由回答式のアンケートを人手によって読み、分類することは多大な労力を要する。著者らは前報において、自由記述式アンケート結果に対し、情報検索技術を導入することにより、効率的な知識獲得手法を提案している[2]。しかしながら、学生の自由奔放な記述から知識を発見することは容易ではなく、選択式と記述式の間接的なアンケート方法が有用であることが考えられる。そこで、新たに、自由記述式については「項目別」に自由記述する方式を採用し、さらに選択式と自由記述式の間接的な方法として、「単語列挙方式」を導入する。これは、ある設問に対し、想起する単語を順番に並べてもらう方式である。

本研究では、第2段階の学生の「成績」と「満足度」の調査のためにも同様に、成績と満足度の構造モデルを構築し、これに基づいてアンケート項目を設定した。作成したアンケート項目は、Web上で回答できるようにシステム化しており、回答は電子フ

ァイルの形で採取できるようになっている。これは、自動的な分析ツール作成のためにも重要であると考えられる。

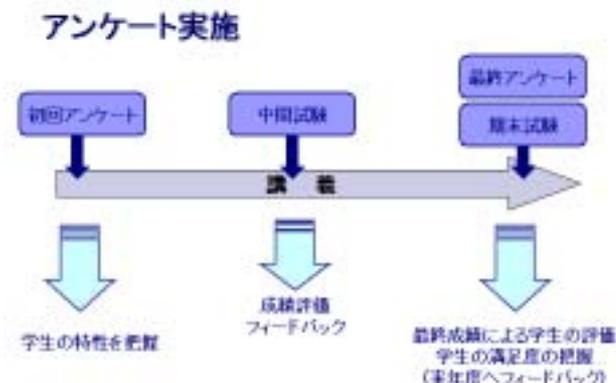


図5．アンケート実施の流れ

4．分析手法の提案

本研究では、学生の特性と講義の結果である「成績」と「満足度」との因果関係を抽出することを目的とし、選択式アンケートと自由記述式アンケートを分析し、それらを統合的に考察する方法を提案する。まず、3章で提示したアンケート作成手順に従って、図4の各項目に対して複数の設問を作成し、全体のアンケートを構成する。アンケート実施には、電子的なデータが得られる様、CGIを用いてWeb上でアンケート結果を得る環境を整備した。アンケートの一部を巻末の付録に示す。

得られたアンケート結果については、選択式アンケートと自由記述式アンケートについて別々に分析を行い、分析結果を統合することで考察を行う。今回提案する分析方法のポイントを以下にまとめる。

[選択式アンケートの分析方法]

- ・ アンケート作成時に構築した「学生の特性モデル」の構造を取り込んだ形で分析を行う。
- ・ 具体的には、図4にある各項目毎について設問項目について回答の重み付け和を取り、各項目の得点とする。
- ・ 上記の重み付け得点に基づき、多変量解析手法を用いて学生の特徴を分析する。その際、結果である成績との関係を分析する。

[自由記述式アンケートの分析方法]

- ・ 学生の自由記述式アンケート結果に対し形態素解を行い、文章中の出現単語と出現回数の表を作成する。
- ・ データマイニング手法であるアプリアリアルゴリズムを適用し、文中で共起関係にある単語セットを抽出する。

- 抽出された単語セットと学生の最終成績データを比較検討し、成績の良い学生と悪い学生の差異について、出現単語の側面から分析を行う。

5. 分析結果

5.1 選択回答式アンケートの分析結果

まず、設問各々に対し授業への積極度にプラス効果のある設問回答にはその重要度に合わせて+1 と +2 を、マイナス効果のある設問回答にはその重要度に合わせて-1 と-2 を重みとし、回答の重み付け和をとり、学生の特性評価モデルを表現する 13 変数へと次元を圧縮した。その後、主成分分析を行うことにより、主成分得点を抽出した。

表 1 主成分固有値と寄与率

| 主成分 | 固有値 | 寄与率 | 累積寄与率 |
|-----|-------|-------|-------|
| 1 | 2.449 | 0.188 | 0.188 |
| 2 | 2.216 | 0.170 | 0.359 |
| 3 | 1.507 | 0.116 | 0.475 |
| 4 | 1.231 | 0.095 | 0.569 |
| 5 | 0.979 | 0.075 | 0.645 |

学生の特性評価モデルでは、なるべく独立な項目を設定するようにモデル化したため、あまり高い寄与率にはならなかった。しかし、一部重み付け和を取るのが困難な項目があったため、多少変数間に相関構造が生じたことから、この主成分分析結果である主成分得点をもとに学生を特性によって、階層クラスタリング手法によりクラスタリングした。クラスタリングした結果をもとに類似した特性をもつ学生をグルーピングすることが可能となったため、これらのグループに対してよりきめ細かい指導を行うことが可能である。しかしながら、これら選択型アンケートの結果について様々な分析を行い、成績との関連をモデル化した結果、選択式アンケートの結果と成績との関連はほとんど抽出できないことが明らかとなった。学生にとって、事前の勉強意欲や興味は成績にはあまり関係がないようである。成績による学生の分類の限界が見えてくる結果となった。

5.2 自由記述式アンケートの分析結果

アプリオリによって共起関係にある単語群を抽出した結果、自由記述式の回答に現れる多頻出単語と成績に強い関係が存在することが明らかとなった。最終成績の良い学生と悪い学生では、自由記述式アンケートで用いる単語に偏りがあることが分かる。次に一例を示す。

[成績高得点群の学生に多出する単語]

LAN モデム, LAN - 知識, プロトコル知識,
LAN - インターネット, モデム-興味
ハードウェア-基礎, 工学 基礎 など

[成績低得点群の学生に多出する単語]

サーバー知識, C 言語 興味, 武器
内容 話, 内容 実際, 分野 専門
機会 専門, 重要 専門 分野 など

5.3 考察

今回、選択式・記述式アンケートからの知識抽出方法を議論した結果、選択式アンケートでは学生を特性によってグルーピングすることは可能であるが、成績予測することが難しいこと、記述式アンケートからは出現単語によって最終成績の差が大きく異なることが明らかとなった。理工系の学生にとって自由記述で「工学」や「基礎」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が、これは多年度に渡って継続的に分析を行い、さらに検討を重ねる必要がある。

6. まとめ

本稿では、授業改善と学生の特性把握を目的としたアンケート作成手順と分析手法について議論した。分析結果から興味深い結果が得られたが、信頼性向上のため継続的にデータ採取と分析を進めていく必要がある。さらに、選択式と自由記述式のアンケート結果を合成し、統合的に分析することも今後の課題である。

参考文献

- [1] M. Hearst: "Untangling Text Data Mining", ACL '99, Proceedings, pp.3-10, (1999)
- [2] 酒井, 伊藤, 後藤, 石田, 平澤: "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)
- [3] 長坂, 阿手: "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, Vol.3, No.1, pp.37-42, (2000)

付録 (作成されたアンケートの一部)

< 講義のイメージに関する質問項目 >

下記の設問について、「全く思わない(-3)」から「強く思う(+3)」までの 5 段階で答えなさい。

- (1) この講義によって、コンピュータを活用できるようになると思いますか？
- (2) この講義はこれからの時代に必要だと思いますか？
- (3) この講義は経営システム工学科にとって必修であるべきだと思いますか？
- (4) 自分にとってこの講義は必要であると思いますか？コンピュータは使いこなせれば原理は知らなくて良いと思いますか？