

< 研究ノート >

正規表現の応用実験

その1 記述式アンケート回答文の解析

江戸 浩幸*・坂本 義行**

Applied Experiments in Perl

— Reply Sentence Analysis of Questionnaires —

Hiroyuki EDO* and Yoshiyuki SAKAMOTO**

概要

これまでに、正規表現形式について学習し、種々のプログラム言語でその機能を利用できることを学んだ。また、その表現の仕方の違い、実装されている機能、さらに、表現形式による効率化の問題等言語による違いを見てきた。この研究は正規表現を十分に利用した応用実験として、大学等で実施されるアンケート調査において、回答の集計で大変な労力を要する記述式のコメント文をコンピュータを利用し、迅速で正確な集計方法を見出すことを目的としている。この論文では、回答文中に出現するキーワードとなる文字列を検索し、その文字列と質問文に含まれるであろう基本概念との相関を分析し、相関関係を集計するプログラムをPerlを用いて実験を進めたので、その解析方法と実験の基本部分について報告する。

キーワード：正規表現、Perl、記述式コメント文、質問文とコメント文の概念相関

0. はしがき

一般に行われるアンケート調査において、選択式と記述式がある。選択式は、結果の集計がコンピュータを用いることにより、迅速かつ正確に結果が得られる。一方、記述式では、人手による集計とその結果をどのような形式で報告すれば、その内容をより正確で迅

速に報告することができるかという点で問題がある。

この研究では、種々のコメント文について、質問のカテゴリ分類を行い。カテゴリ別に、コメント文の集計を行う。この集計でコメント文は、一文からなる日本語文字列からなるものとする。この文字列中には複数のキーワードが含まれるものとする。この

* 情報コミュニケーション学部情報メディア学科、Tsukuba Gakuin University

** 筑波女子大学短期大学部情報処理科、Tokyo Kasei Gakuin Tsukuba Junior College

キーワードを抽出することにより、このキーワードがどのような概念に分類されるかによって、内容の分析を行い。結果を報告するシステムの開発を進める。

システムに使用するプログラム言語として、正規表現の処理機能が強力な Perl を採用した。これは、筆者たちがこれまで研究を進めてきた正規表現の応用実験でもある。

1. 質問のカテゴリー分類

アンケート調査を行うときに、どのような事柄についての質問であるかを明確に決定する必要がある。この実験のサンプルとして、アンケート調査での、学生からの大学での授業に関するコメント文を入力データとして使用した。このため、表 1 に示す 11 の項目からなる質問のカテゴリー分類表を用いることとした。この表は、階層構造でかつ可変長文字列からなっている。

2. カテゴリーの識別と概念

2.1 授業についての質問分類

授業について、満足した点、不満足に感じ

た点についてコメントを収集した。そのうち、満足した点のコメントについて、キーワードを抽出し、概念による分類を行った。一文中のキーワードは、一つとは限らず、複数含まれる場合、またキーワード間に概念として、階層的な相関がある場合がある。

2.2 キーワードと概念の相関

ここでは、「満足している授業について」のコメント文を例に、そこに出現したキーワードと質問文のキーワードとの概念間の相関分析を行なった。質問の概念として、考え方、授業内容、教え方、教室の人数、教員の態度、授業環境、教員の能力、学生の立場について、回答文中にどのようなキーワードが用いられるかを相関分析した表を表 2 に示した。

2.3 キーワード辞書の見出し

回答文中のキーワードはその概念が表している意味の範囲によって階層的な概念分類が必要であると思われるが、ここでは、一次元の文字列として表現したものを「見出し」とした。たとえば、「教え方」という概念についての見出しを、「上手い」、「私語が減

表 1 質問のカテゴリー分類表

	カテゴリー	具体的な要望	細分類
1	授業	満足 不満足	
2	新規開設の授業科目・カリキュラム		
3	授業時間外の学習	図書館	利用できる時間帯 書籍の探しやすさ
4	視聴覚設備の使いやすさ	職員の対応	
5	情報処理演習室	利用できる時間帯	
6	指導体制		
7	学内設備	語学のセルフ・アクセスセンター	
8	授業時間外の学習指導	オフィスアワー その他の要望	
9	クラブ活動	参加しない理由	
10	大学事務局に対する要望		
11	その他		

表2 満足する授業の概念分類表

満足している授業について	考 え 方	おもしろい
	内 容	新鮮
		興味深い
		幅が広い
		外部招請
		種々の理論
		方法論
		生きるための指針
		何気ない話のためになる
		実践的
		将来のためになる
		理解できるから自分のつぎのことが考えられる
		学び始めたら、(非常に)面白かった
		生活の中で役に立つ
		作品づくり
		アイデアを生かせる
	取組む内容が難しくない	
	難しくても、役に立つ	
	教 え 方	上手い
		私語が減る
		話が楽しい
		聞くだけでなく、質問ができる
		ただでなく、テキパキ
		集中して受けられる
		時間短縮 (一時間位)
		学生のリズムに合わせる
		学生の反応をみながら
		学生のペース
		板書が早くノート書きに集中
		講義中に自由にパソコンが使える
実技中心		
パソコンをさわる授業		
授業中に復習が可能		
人 数	少数	
	質問する場が与えられる	
態 度	個人指導	
	権威でなく、個人を人間として	
環 境	私語がないから集中できる	
	うるさくない	
能 力	優秀、教員としての素質	
学生の立場	楽	
	テストがない	

る」...とした辞書を作成した。これを表3に示す。

2.4 見出しと概念による集計表の作成

「教え方」の概念を表現するキーワードとして表3を決定し、この見出しがコメント文中から検出されるとこれを積算することにより、コメント文中に出現したキーワードの頻度集計が行われる。

3. プログラムの作成

3.1 質問カテゴリー表の作成

学生アンケートにおける質問カテゴリーである表1をPerlの配列を用いて表現したプログラムを作成した。ここでは、「授業」のカテゴリーで、「満足である」項目のみについての実験に限った。その変数名を次のように設定した。

```
$jugyou
$manzoku("満足")
```

3.2 キーワード辞書の作成

授業カテゴリー中で「満足する」教え方において出現すると思われるキーワードを表現した表2から、配列で表現した図を図1に示した。

表3 出現するキーワードによる見出し表

上手い
私語が減る
話が楽しい
聞くだけでなく、質問ができる
ただでなく、テキパキ
集中して受けられる
時間短縮 (一時間位)
学生の反応をみながら、学生のリズムに合わせる
学生のペース
板書が早くノート書きに集中
講義中に自由にパソコンが使える
実技中心
パソコンをさわる授業
授業中に復習が可能

```
$oshiekata[0] = "上手い";
$oshiekata[1] = "私語が減る";
$oshiekata[2] = "話が楽しい";
$oshiekata[3] = "聞くだけでなく、質問ができる";
$oshiekata[4] = "ただでなく、テキパキ";
$oshiekata[5] = "集中して受けられる";
$oshiekata[6] = "時間短縮 (一時間位)";
$oshiekata[7] = "学生のリズムに合わせる";
$oshiekata[8] = "学生の反応を見ながら";
$oshiekata[9] = "板書が早くノート書きに集中";
$oshiekata[10] = "講義中に自由にパソコンが使える";
$oshiekata[11] = "実技中心";
$oshiekata[12] = "パソコンをさわる授業";
$oshiekata[13] = "授業中に復習が可能";
```

図1 キーワードを配列で表現

図2 (a) 入力コメント文の領域

BUNIN									
文字位置	1	2	3	4	5	6	7	8	9
文字列	教	え	方	が	上	手	い		

図2 (b) コメント文の照合領域

BUN							
文字位置	1	2	3	4	5	6	7
文字列	教						

図2(c) キーワードの照合領域

KI-WA-DO							
文字位置	1	2	3	4	5	6	7
文字列	上	手	い				

図2 コメント文の読み込みと照合の領域

3.3 コメント文の入力処理

コメント文は、ファイルで入力され、一文が1レコード単位の可変長レコードで読み込まれる。

3.4 キーワードの照合と集計

1) 1レコード文の照合の基本処理手順

文はテキスト形式で1レコード単位で入力される。これを図2のBUNINで示す。この1文の文字列に対して、先頭から1文字づつをBUNに移し、その文字列に対して、図1の各キーワードと照合を行う。BUNの文字列に対して、すべてのキーワードの見出しと一致した文字列が文中で見出されたら、集計表に積算される。BUN中の文字列について、すべてのキーワードについて繰り返し照合が完了すると、同一のBUNINの中で、先頭の文字を1字ずらして、そこを先頭として、BUNINからBUNへ文字を同様に1字づつ移しながら、キーワードとの照合を繰り返す。1文についての照合が完了したら、あらためて、次の1文を読み込み、同様の照合を繰り返す。

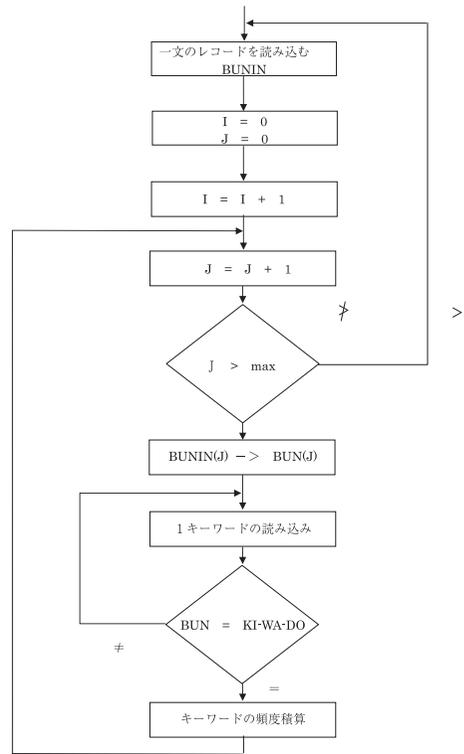


図3 コメント文の読み込みと照合手順

2) すべてのコメント文の照合

すべてのコメント文はテキスト形式のファイルになっており、文単位(1レコード)で読み込まれる。これに対して、繰り返し照合が実施される。

この読み込み照合の基本的な手順を図3に示す。

4. 基本実験プログラム

4.1 コメント文の読み込み実験

1ファイルのコメント文を図4に示す。このコメント文を1文づつ読み込み、文番号を付けて蓄積するプログラムを図5に示す。次にこのプログラムを図6で示すコマンドで実行した結果を図7で示す。

先生の考え方がおもしろい。
 その授業に自分も参加しなければならない。
 毎時間他の学校から先生をお招きして講義を受けられるので新鮮であるし、興味深い話をしてくれるから。
 先生の教え方が上手い授業は非常におもしろいです。私語の多い学生も授業内容次第では私語が減るのではないのでしょうか。
 少人数なので質問しやすい。
 ただ先生の話聞くだけでなく、自分も発言する場がある。
 授業がだらだらとせず、テキパキしている。
 だらだらとした長い授業をしていないから、(約10分~20分早く終わる)→集中して受けられる。
 いろいろな理論や方法論なども教えてくれるのでおもしろいです。
 とらなきゃいけないから。
 ゼミだからがんばるしかない。
 先生としての権威を学生に対して乱用しないで、学生を1人の人間として扱ってくれる。

図4 コメント文

```
$i = 0;
while (<STDIN) {
    print ++$i;
    print " ";
    print;
}
```

図5 コメント文の読み込みプログラム

5. 考察

今回の実験は、Perlの応用実験として、位置づけ、Perlの基本機能である文字列の照合を簡潔で、効率的に行うために、文字列をどのように表現でき、正規表現命令の組み合わせにより、照合が可能なかを検証することからスタートした。

一方、アンケートのコメント文をキーワードという文字列に着目し、これを検出することにより、その頻度から回答者が述べようと

```
C:\Perl>perl numbering.pl < comment.
txt > numbering.txt
```

図6 DOS コマンド

- 1:先生の考え方がおもしろい。
- 2:その授業に自分も参加しなければならない。
- 3:毎時間他の学校から先生をお招きして講義を受けられるので新鮮であるし、興味深い話をしてくれるから。
- 4:先生の教え方が上手い授業は非常におもしろいです。私語の多い学生も授業内容次第では私語が減るのではないのでしょうか。
- 5:少人数なので質問しやすい。
- 6:ただ先生の話聞くだけでなく、自分も発言する場がある。
- 7:授業がだらだらとせず、テキパキしている。
- 8:だらだらとした長い授業をしていないから、(約10分~20分早く終わる)→集中して受けられる。
- 9:いろいろな理論や方法論なども教えてくれるのでおもしろいです。
- 10:とらなきゃいけないから。
- 11:ゼミだからがんばるしかない。
- 12:先生としての権威を学生に対して乱用しないで、学生を1人の人間として扱ってくれる。

図7 実行結果

している意味概念を抽出することができるかの基礎実験である。

したがって、実験に使用したコメント文も簡単な標本を用いたにすぎない。また、プログラムについても、簡単な変数定義と基本的に可能な文字列照合の機能のみを用いている。したがって効率的に見て最適な命令の集合とはいえない。

今後は、実際のコメント文中に表現されている複雑なキーワードをいかにして抽出し、その集計を行うか。さらに、そのキーワードが意味している概念、文が意味している内容をそのキーワードが表現しているかといった問題について分析を進めたいと思っている。