

## 多変量解析を用いたオンライン収集情報評価システムの検討

土屋 雅生      松永 俊雄

matsu@cc.teu.ac.jp

東京工科大学 情報工学科

〒192 東京都八王子市片倉町 1404-1

新商品に関するユーザ満足度やユーザ要望の調査、大学における授業評価などアンケートによる調査は、回答用紙を利用したものが多く、一方、近年では WWW によるホームページの利用が増大し、ユーザは、商品に関する情報の入手やこれらに関する要望の転送を容易に実現できるようになった。

本検討では、WWW を利用したオンラインの情報収集システムを実現し、それにより収集したアンケート結果を多変量解析を用いて統計的に解析するシステムの実現方式の 1 例について述べる。実現システムでは、アンケート対象に合せて、調査項目の作成・編集、回収結果の多変量解析による統計的处理と結果のグラフ表示などを可能にしている。

### Study of an Evaluation System based on Multivariate Analysis of Online Collected Information

by Masao Tsuchiya and Toshio Matsunaga

Department of Information technology, Tokyo Engineering University

1401-1 Katakura-cho Hachioji-shi Tokyo 192 Japan

A questionnaire form is usually used in case of surveying customer satisfaction for new products or in case of investigating the class evaluation in the university. On the other hand, it is recently increased for customer to access homepages publishing new products on WWW, therefore, it is easy for them to procure the information concerning new goods or products and to forward demands for these goods to the information provider.

This paper describes an evaluation system based on multivariate analysis of online collected questionnaires on WWW. This prototype system has some features to make up and edit questions of investigation for objects, collect answers about questions, process them statistically based on multivariate analysis and display the results graphically.

#### 1. はじめに

近年の情報化社会の進展に伴い、ネットワークが一般家庭にまで普及するようになった。特に電子メールやWWWの利用者はここ数年で激増しその存在を知らない人はいないといっても過言ではない。これまで郵便や電話に頼ってい

たアンケートがインターネット上で公開される例も増えてきている。しかしながら、アンケートを自動的に収集し、分析するシステムは一部試みられているが<sup>1)</sup>、アンケート内容やその結果の分析評価するシステムの例は少ない。

そこで、本稿では、「多変量解析を用いたオンライン情報システム」としてアンケートを収集

し、多変量解析を用いて分析することによってアンケートの回答者がどの項目について満足し、また期待しているかを示すことができるシステムについて検討する。本システムは質問をインターネット上に公開し、自動的に分析することによってアンケートを掲載した人に問題点を提示し、改善点の検討に参考となる情報を提供することをねらいとしている。<sup>[2]</sup>

本稿ではこのシステムの設計の考え方と実現した内容について述べ、評価した結果について述べる。

## 2. システムの設計条件

本システムは、WWW上でのアンケートの実施を可能にするため、以下の条件で設計を進めた。

(1) アンケート掲載、回収、分析サーバ。

UNIX系サーバでCGIをサポートしたhttpdが動作していること。またperl ver5、gcc ver2.7、gnuplotが動作すること。

(2) アンケート回収ブラウザ

JavaScriptが動作すること。(Netscape3.0以上、推奨)

(3) システム利用の対象者

HTMLやCGI、多変量解析について高度な知識を有していないユーザ。

## 3. システムの実現

### 3.1 質問掲載ツールの実現

質問掲載ツールとは、本システムを利用するにあたってアンケート項目をWeb上に掲載したいと考えている質問掲載者を支援するためのものである。前項で示した条件からHTMLについて高度な知識を有していないユーザを支援するためのツールである。ここでこのツールは単に支援するためのだけのものであるため、背景色をつけたり、画像

を貼り付けるなどの機能は実現していない。

この質問掲載ツールは図1の画面に示す通り、アンケート項目に必要な質問事項、段階評価例を入力することにより質問ページを作成できる。氏名、ID、および回答者の意見など分析に使用しないアイテムでも質問掲載者にとって付け加えてあった方がよい事項を作成することもできる。このような場合のために、分析に必要な項目の範囲を指定するためのタグを設けている。

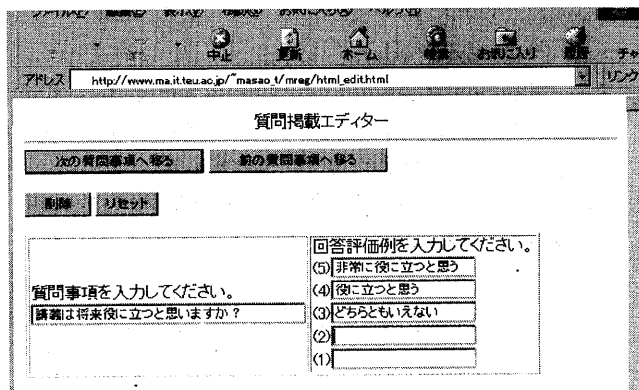


図1 質問掲載エディタ

また、多変量解析を用いたアンケートを実施したことの無いユーザを対象に、多変量解析を行うためにどのような質問事項を用意するとよいのか、また質問内容により回答が変化することがあるなどの例示をヘルプ機能として用意した。この機能はJavaScriptを使用して作成した。必要と思われる項目の上をマウスが通過するとヘルプを表示するようにした。<sup>[3][4]</sup>

このツールを用いて作成されたページはアンケート回答者がマウスのみの利用で回答できるようになっている。このツールによって完成したページの一部分を図2に示す。

### 3.2 回答回収ツールの実現

アンケートの回収は回収ツールを利用して行うことができる。質問回答者が「アンケートに答える」というサブミットボタンを押すとhttpdサーバに決まった文字列が送信されてくるので、そ

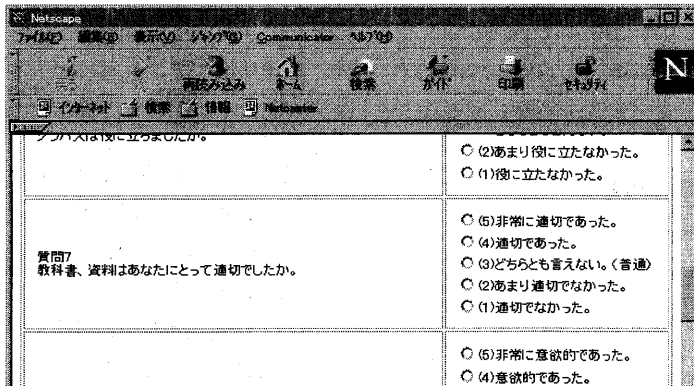


図2 完成したページの一例

の文字列を表計算ソフトウェア等で一般的に利用されている図3のようなCSV形式に変換する。文字列、ファイル操作に関してはその扱いが容易なプログラム言語である perl を使用した。<sup>[5]</sup>

```
3, 3, 2, 4, 1, 5, 3, 2, 4, 1, 2, 4, 1, 4
2, 3, 1, 5, 4, 2, 3, 1, 3, 3, 2, 4, 1, 3
2, 2, 4, 3, 2, 4, 2, 4, 2, 1, 4, 1, 5, 3
```

図3 CSV形式

回収されたデータはそのまま重回帰分析で利用することは可能であるが、数量化 I 類を利用することはできない。数量化 I 類を利用するためのデータに変換する必要がある。そのためのツールも存在し、質問掲載者はアンケートの回収時にどちらを利用するか、選択することができる。

また、回収されたデータはそのまま表計算ソフトウェアで利用することも可能であるがデータベース形式に変換することも可能である。この場合のデータベースの形式は Perl で一般的に使用されている dbm 形式である。また、この変換方法については、Web 上からの直接変換は実現していない。

### 3.3 多変量解析ツールの実現

回収した結果の分析方法は多変量解析の中でも重回帰分析を利用する。回帰式は一般的に

$$Y \leftrightarrow y = a_1x_1 + a_2x_2 + \dots + a_nx_n + a_0$$

という数式で表される。<sup>[6]</sup>この数式は  $y$  (目的数量) と  $x$  (説明変量) との間にはどのような相関があるかを  $a$  (回帰係数) であらわした式である。本研究ではアンケートにおいて総合的な評価 (目的数量) は、個々の機能や要点 (説明変量) からどの位影響を受けたか (回帰係数) を示すものとして使用した。この回収されたデータは CSV 形式で保存されている。この保存データは多変量解析プログラムを

利用して解析を行い、結果を新しいファイルに保存するように設計した。<sup>[6]</sup>

一般的に多変量解析を行うと回帰係数のほかにいろいろな数値が求まる。しかし、本システムは多変量解析について高度な知識を有していないユーザを対象にしていることを考慮し各項目の平均値、回帰係数以外は出力しないようにしている。また平均値、回帰係数をもとに、分析結果を表示する場合は満足度、期待度という用語に変換して表示には多変量解析の実態は見せないような設計をしている。ユーザが見る結果とは、次項で述べるグラフを使った結果表示のみである。

また、多変量解析のプログラムには C 言語を利用した。インタプリタである Perl を利用した場合と比べ実行時間やメモリの消費量の面からみて有利である。<sup>[6]</sup>

なお、本システムが提供するアンケート収集システムを利用していないユーザでも多変量解析を利用できるツールも備えているが、この場合、結果のグラフ化を実行できない。また、数値を表示する形となるので多変量解析についての知識を必要とする。その実行例は図4のようになる。

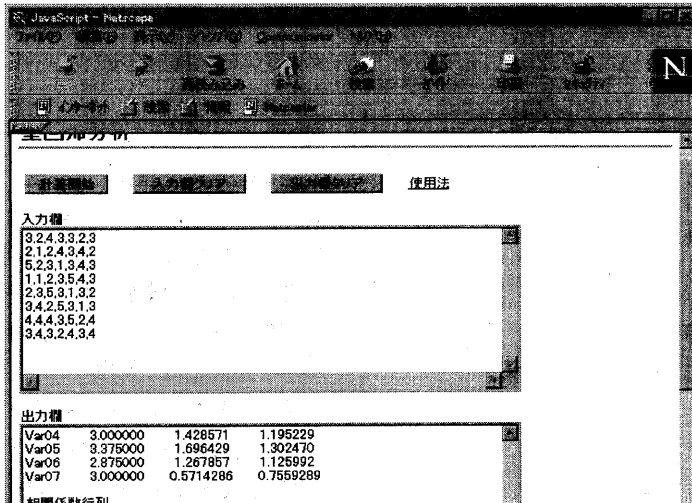


図4 重回帰分析ツール

### 3.4 グラフ化ツールの実現

重回帰分析によって求められたデータは前項で述べたとおり新しいファイルに保存されている。満足度と期待度の関係からつぎの4つの関係を作る。

- ① 問題事項 期待度が高く、満足度が低い。
  - ② 優良事項 期待度が高く、満足度も高い。
  - ③ 改良事項 期待度が低く、満足度が高い。
  - ④ 注意事項 期待度が低く、満足度も低い。
- この4つの関係は図5の通りとなる。

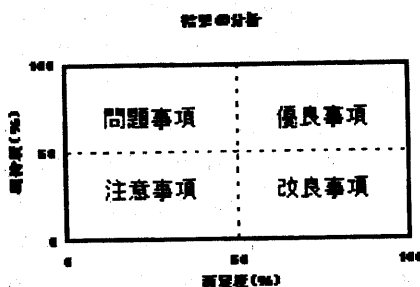


図5 満足度、期待度の関係

分析後のファイルはgnuplotを利用してグラフ化する。一般的にgnuplotはUNIX上で動作するために実行結果の画像形式がxpmである。またブ

ラウザで扱える一般的な画像形式はgif、jpegであるからその変換はまた別のツールを利用することになる。

実行結果は質問掲載者が見たいときに結果を表示する方法と、リアルタイム表示の2通りの表示方法が考えられるが、これについては質問掲載時に選択できるようにしている。しかし、アンケートの回収量が多いとリアルタイム表示はサーバにかなりの負担となる可能性があるため、利用者には特に用途がない場合使用しないようにページに注意書きとして記載した。

### 3.5 本システムの動作

本システムの実行順序は以下の通りである。

- ① 質問掲載者が質問内容を決め、質問掲載エディタを利用して質問事項を書く。
- ② 質問掲載者は多変量解析で重回帰分析を行うか数量化I類を行うかの選択をする。(デフォルトでは重回帰分析のみ)
- ③ 質問掲載者は質問解答者に対し結果を公開するか、しないかを決める。
- ④ 質問掲載エディタによりWeb上で質問が公開される。
- ⑤ 質問解答者がWeb上でのアンケートで回答し、回答をサーバに返す。
- ⑥ 返ってきた回答をCGIを利用してCSV形式で保存する。
- ⑦ CSV形式で保存されたデータを質問掲載者の指示により、回答は多変量解析ツールを使用し分析する。
- ⑧ 結果を期待度、満足度として新しいファイルに保存する。
- ⑨ 期待度、満足度をgnuplotでグラフ化する。
- ⑩ gnuplotで出来上がったグラフを変換

してWeb上に掲載する。

## 4. 評価

### 4.1 システムの利用評価

作成した本システムについて25人に利用してもらい、システムの使いやすさについて評価した。図6は本システムについてアンケートを取った結果である。

本システムの得た評価はほぼ満足できるものであったが、多変量解析への説明が不足していたために利用者からは、質問内容の選択やグラフの見方に関しては低い評価となった。また、本システムはデータの再利用も考えて設計したが利用者がいなかった。これは、ヘルプ機能が少ないという意見が多数あったことに現れていると考えられる。

特にグラフの見方は本システムにおいて重要であるが、評価が低い理由としては、グラフがカラーでないことや、結果表示の方法に問題があることが指摘された。

### 4.2 授業評価アンケートによるシステムの利用

図7は本学の学生によるある科目の授業評価アンケートを行った結果例である。授業評価アンケートはマークシートで行われているものをWeb上に転載したものである。

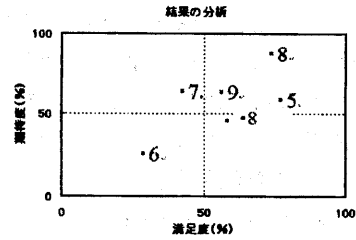


図7 アンケートの結果

このグラフの数字は各質問の番号である。「質問8」は優良事項であるが「質問7」は問題事項である。

利用した結果、アンケートに複数回答をする人や一部の質問に回答のない人がいることが分かった。またグラフ化をすることによって改善点を視覚的に指摘することができたと考えられる。

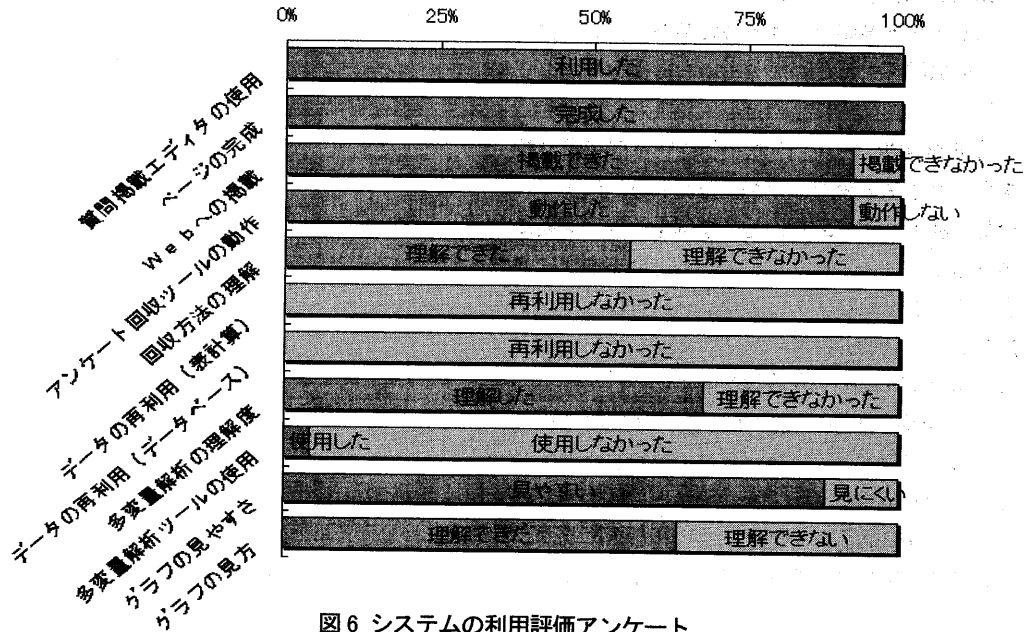


図6 システムの利用評価アンケート

## 5 考察

本システムの利用評価結果から次のことのようなことがいえる。

本システムは郵送を用いたアンケートと比較をすると、回収量や金銭的な面以外でも回収データを分別や分析をする手間が省けるなどの利点があること考えられる。回収データはそのままで活用できるように設計したが、そのデータの再活用やデータベース化まで進めた利用者がいなかった。この点については実際の活用法や活用例を詳しい形でヘルプ機能として表示する必要があると考える。

つぎに、本システムの機能についてみると質問掲載エディタは、回答の評価段階を5段階の10進数表示とし、ほかの段階数を選択する事ができない。また、2進数での5段階以外のカテゴリを要求される数量化I類を利用したい場合、本システムでは対応できない。数量化I類に対応した質問ページは各個人が作成する必要がある。システムの利用者を考える場合は、質問掲載ツールを上記に対応するよう改善する必要がある。

アンケートの回収方法については、アンケートに複数回答する人が数名いたために、これをリジェクトする機能をつける必要がある。アンケートに複数回答すると回帰分析にも影響を与えることになる。

多変量解析については、その中で重回帰分析を用いた手法を検討したが、重回帰分析は多くの分野で利用されており、その事例をヘルプ機能に載せることにより質問事項の内容を選ぶときに便利であると考えた。また、重回帰分析以外にも主成分分析、判別分析等の手法があり、別の分野での分析ができるので今後検討したい。

分析結果の表示については、グラフ内に詳しい内容を掲載すると内容が見にくくなることや、カラーで表示する事ができないなどの問題があるので、gnuplot以外のグラフツールを含めたグラフ化についても今後検討していく予定である。

## 参考文献

- [1] 澤野 貴、柴健一郎他：WWWを利用したイメージアンケートシステム、情報処理学会第54回全国大会、4S-5、1997.3
- [2] 日経ビジネス：「発想法スーパーガイド97」、日経ビジネス、1997
- [3] 古旗一浩：「JavaScript ポケットリファレンス」、技術評論社、1997
- [4] 川崎克巳：「JavaScript マスターバイブル」、スパイク、1997
- [5] 伊藤和人：「入門Perl」、秀和システム、1997
- [6] 有馬哲 石村貞夫：「多変量解析のはなし」、東京図書、1994
- [7] 石村貞夫：「すぐ分かる多変量解析」、東京図書、1994
- [8] 内田治：「すぐわかるEXCELによる多変量解析」東京図書、1995
- [9] 奥村晴彦：「C言語によるアルゴリズム辞典」、技術評論社、1992