

普通紙を用いた調査票データの自動処理

Automatic Data Processing of Questionnaires Using Normal Paper

東京医科歯科大学 若松 秀俊
 横浜国立大学 影井清一郎

1 まえがき

アンケート調査は世論調査、市場調査、疫学的調査など数え上げればきりが無いほど行われている。この調査法は確率・統計学的理論に立脚した学問的にも実用的にも重要な方法である。マーク記入によるアンケートの質問に対する回答は「上質のマークシート」を専用の読取装置を用いてコンピュータへ入力し処理を行うのが普通である。しかし、特殊な紙を用いたマークシートによる場合は、その処理が比較的容易な割に費用がかさむので「普通紙からなる調査票」アンケート調査が各方面から望まれている。ところが「安価な普通紙」による調査では紙の曲がりや歪みから読みとりに誤りが多く、一般に人手に頼るコンピュータ入力に依存しており、調査対象が非常に多い場合はこれらに要する労力・時間は膨大なものになる。ここでは、「普通紙からなる調査票」処理の省力化のために開発したニューラルネットワークを用いた画像読み取りとデータファイルの自動構築のための情報処理システムについて、実際に行った調査の処理を例にとってその機能を説明する。

2 システムの概要

2.1 本システムの構成と処理の概要

本システムはファクシミリ装置とコンピュータを用いた情報処理システムである。調査票の処理に必要なハードウェアはG ファクシミリ装置とG ファックス・アダプタである¹。

アンケート調査のための調査票の回答には正方形のマス目をチェックまたは塗りつぶすマーク記入方式を採用し、その結果を市販のファクシミリ装置とパーソナルコンピュータを使って処理し、後の統計処理が可能なデータに変換する。このシステムは「普通紙を用いた調査票の画像処理によるマーク読み取りとデータファイルの自動構築」のためのものであり、一般のアンケート調査に十分適用可能なものである。図1は実際にアンケート調

査票を部分的に示したもので調査票の回答は鉛筆で記入された回答欄内のマークで与えられる²。調査票をファクシミリ装置で読み取り、回答欄のマークが記入されているかどうかを自動判定し、コンピュータ処理に適したデータに変換する。

2.2 アンケート用紙とデータ形式

調査票の周囲には識別領域を設けてその処理に必要な情報を記載する。回答は調査対象者が簡単に答えられるように白抜きで正方形で与えられる回答欄を選択記入する方式とした。図2はそれを模範的に表したものである。用紙の外側の上端に黒く塗りつぶした識別マークを、左端と下端に基準となる同様の基準マークを描く。基準マークの位置を基にして表せる点を「同期点」と呼ぶ。用紙の下端の横に並んでいる同期点が「横同期点」、用紙の左端の縦に並ぶ同期点が「縦同期点」である。用紙内の文章はどの方向に書かれていてもよいが、用紙は左の方向から縦同期点を先頭にして読み取られる。横同期点には左端を0として右方向に、また縦同期点には下端を0として上方向に一連番号が振られている。回答欄は横と縦の同期点の交差する位置に置かれているので、回答欄の位置は横同期点の番号と縦同期点の番号の組によって表わされる。

用紙のページ識別は上端横同期点で与える識別マークによって行われる。上端の横同期点には基準の位置を示す役割もある。読み取りのときの画像の歪みは同期点の情報を用いて補正される。

調査票への記入は図3に示すように隣の回答欄のマス目に入らない限り、かなり自由に行うことができる。ここで、回答マークが基準の座標軸に関してどこの座標にあって、その回答が個人データファイルのどこにどのような形式で格納するかを記したデータテーブルを作成する。このデータテーブルは回答マークの座標をもとに、それぞれの調査票の質問項目に対応して作成した同期点データよりなる³。回答欄は基準軸の格子点の位置

¹ 実際に用いた装置はファクシミリ装置 (SANYO SFX-20) , G ファックスアダプタ (日本テレコメット, MF-101) , パーソナルコンピュータ (NEC, PC-9801RX, ディスク容量 10MB) で, MS-DOS Ver3.30C以上を装備したものである。現在, Windows 95版のプログラムを作成中である。

² 「生活習慣が子供の健康や心身の発育に及ぼす影響」について, 8分類35I項目にわたる調査をマーク記入による質問形式の調査票を用いて全国規模で行った。調査対象母集団を全国の小・中学生とした。調査対象者の選定方法については調査の客観性と数学的な処理を重視して多段階無作為抽出により12,000人について行った。それらをファクシミリ画像処理し, さらにこれをデータベース化して質的データの数量化評価を用いた統計処理を行った。

³ 回答欄をすべて塗りつぶした調査票をファクシミリ装置に読み込ませ

前述のデータテーブルファイルを参照することにより、調査票から読み込んだ回答データからどのような回答がなされたかを判断できる。マーク判別に最初からニューラルネットを使用すると時間がかかるので、はっきりと分類できないマークデータだけにニューラルネットを応用する。枠外に書かれたマークについては処理しないとして、以下にその方法を順に説明する。

(1) 明確なマークの判別

図5のようなマークの面積算出による判別を行う。これで判別不可能な場合には、次の(2)のステップに進む。

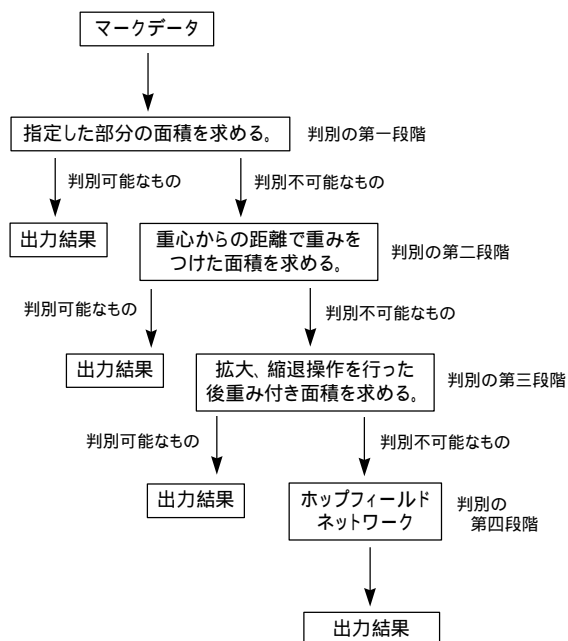
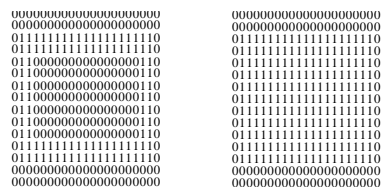
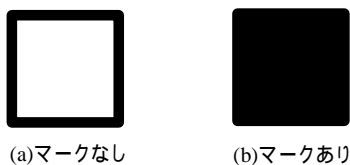


図4 マーク判別の流れ図



マークなしの代表データ マークありの代表データ

図5 明確なマーク

図6のような枠の外にマークが記されていたり僅かに枠にかかったようなものに関してはマークとみなさない。なお、回答者が消しゴムがないなどの理由から一旦マー

クをつけ、例えば×印で消したつものような場合に見られる、回答者の意図通りに判別できないマークも除く。

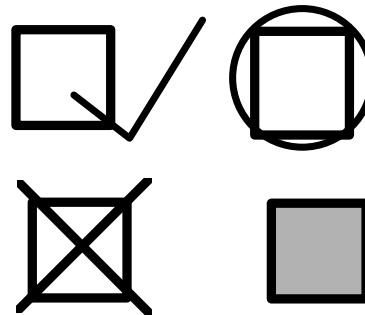


図6 意図しない(取り消し)の例

(2) 曖昧なマークに対する判別

重み付き面積による判別

画像の形状を表す測度として面積、周囲長、分散度、伸張度、オイラー数などがある。ここではマークが枠の中心に記される確率を考慮して、重み付き面積による判別を用いる(画像処理工学 基礎編7章参照)。

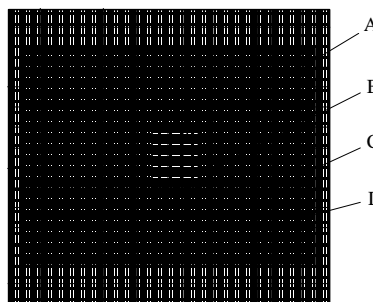


図7 チェック箇所につけた重み

拡大、縮退操作による判別

図形の連結性に基づき2値画像の幾何学的特徴を認識するための方法である。拡大により図形の凹部を、拡大により凸部を滑らかにする効果がある(画像処理工学 基礎編7章参照)。これより回答マークを判別する。

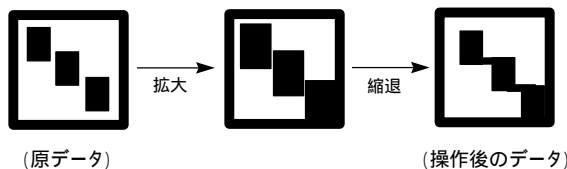


図8 拡大・縮退操作の概念図

ニューラルネットワークによる判別

最後まで判別できなかったマークの第4段階として、ホップフィールドネットワークによる判別を行う(画像処理工学 基礎編8章4節参照)⁴。

⁴ 指定の回答欄がチェックされているかどうか、不明であるかどうか、ま

4 実際の処理

4.1 マーク読み取りの手順

データテーブルの組み込み作業やアンケートの回答用紙からマークを読みとる作業はメニュー画面から機能を選択する形式になっている。またそれぞれの作業に対してサブ画面が用意されている。従ってはじめにメニュー画面を起動すれば、後は画面に表示される指示に従って処理を行うことができる。この様子を図9に示す。

ひとまとまりのアンケート用紙には一つのグループ名をつける。アンケート調査の回答用紙には調査地などを示す識別文字と4桁以内の識別番号が割り当てられる。グループ名や識別番号は、用紙読み取り直前にキーボードから入力され、コンピュータ処理時に調査データの最初に組み込まれる。

マーク読み取りメインメニュー (コマンド選択)	
F1 データテーブルの作成	日付: 時刻 MS-DOS:
F2 用紙の読み込み	
F3 読み取り結果の変換	
F4 終了	
F5	
¥fax¥submenu1.mnu!	
<input type="text"/>	
矢印キーで項目を選択し、リターンキーを押して下さい	

図9 メイン・メニュー画面

4.2 調査データファイルの自動作成

調査票には様々な情報が含まれているので、この調査結果を誰にでも簡単に取り扱えるようする必要がある。そこで得られた2値データから個人データファイルを作成する。まず、調査票の質問項目およびそれに対する回答に番号を付けてコード化しておく。このコード番号がデータとなるので、その扱いが容易になる。これより、調査対象者個人のデータファイルのフォーマットが定まりファイルが構成できる。読み取りデータのうちグループ名と識別番号が個人データファイルの最初に入る。続いて、マークの読みとり結果が指定された順にデータとして記憶される。得られた調査対象全員のデータファイルをまとめる処理によりこれらが連続的に並べられてひとまとまりの調査データファイルが構成され、種々の統計的処理に備えることができる⁵。

た。誤った答え方であるかどうかを統括処理時に検出できる。

⁵ 領収書保と後のデータの読み出し・加工の便を考慮してシーケンシャルファイルとした

5 あとがき

ここでは、ファクシミリ装置とコンピュータを備えれば、若干のプログラムの変更を必要とするだけでアンケートのニューラルネットワークを併用したファクシミリ画像処理が可能なアンケート情報処理システムを紹介した。その処理法は画像処理工学 基礎編(谷口編)に掲載した理論を用いた方法にしたがったものであり、質問票として扱う用紙が「普通紙」で汚れや変形などのために処理しにくい場合に特に有効である。



図A 情報処理システムの外観

参考文献

- 1) G FAXアダプター MULTIFAX MF-101 取り扱い説明書, 日本テレコム社, 東京(1987).
- 2) テレマチック・サービスのための端末装置とプロトコル, 日本ITU協会, 東京 (1988).
- 3) 若松秀俊: 食品および食習慣の子供の健康におよぼす影響に関する調査研究. 浦上財団研究報告書, No.3, 17-29 (1992).
- 4) 谷口慶台編: 画像処理工学 基礎編 共立出版社, 東京(1996).
- 5) 若松秀俊・影井清一郎: 普通紙を用いた調査票データの自動処理. 日本健康科学学会誌, 13-1, 31-38 (1997).