

神経回路網モデルによる虚偽検出検査のコンピュータ判定

足 立 浩 平*

Computer-Aided Detection of Deception Using Neural Network Classifier

Kohei ADACHI*

The purpose of this study was to assess the accuracy of computer-aided detection of guilty knowledge using a feed-forward neural network classifier. A three-layer network classifier using the error back-propagation algorithm was developed to classify polygraph charts on the basis of respiratory, electrodermal and cardiovascular responses. The network classifier was assessed with the 96 charts obtained from the experiment in which the guilty subjects committing a mock theft and innocent ones received guilty knowledge tests. The cross-validation assessment showed that the detection by the network classifier was 77 percent correct. The outcome was not superior to the detection accuracy by linear and quadratic discriminant analyses. Some properties of the neural network classifier were discussed.

Key Words: detection of deception, neural network, discriminant analysis, polygraphs

序 論

虚偽検出検査の主要な質問法である緊張最高点質問法 (Guilty Knowledge Test) では、被験者に犯行内容に関する裁決質問と非裁決質問を提示し、各質問に対して発現する生理反応をみる (Lykken, 1959; 鈴木, 1986; Ben-Shakhar & Furedy, 1990)。例えば、窃盗対象が「万年筆」であり、この犯行内容を無実の者が知らない場合には、「盗まれた物は何でしたか」と教示した後に、「ネクタイピン」、「腕時計」、「万年筆」、「ライター」、「電卓」といった項目を順次提示していく。これらの項目の中で、「万年筆」だけが真の犯行内容を表すが、これを裁決質問と呼び、

他の正しくない項目を非裁決質問と呼ぶ。もし、被験者が犯人であれば、真の犯行内容を知っているので、裁決質問に対して非裁決質問より顕著な生理反応を惹起すると考えられるが、無実であれば、裁決質問と非裁決質問を識別できないので、両種の質問間で有意な反応量の差は見られないだろう。従って、判定は次の方法でなされる。すなわち、裁決質問に非裁決質問より顕著な反応が認められれば、被験者は陽性 (質問された犯行内容の認識あり)、認められなければ陰性 (認識なし) と判定される。

以上が、質問法と判定法の概要であるが、一般に、生理反応は、呼吸運動、皮膚電気反応および脈波等の指標が測定記録され (足立・鈴木, 1991)、現行の

* 科学警察研究所

National Research Institute of Police Science

検査では、これらの反応記録を検査者が視察して判定を行っている。しかしながら、検査の信頼性を考慮すると、視察判定よりも、コンピュータを利用して判定を行う方が望ましいと考えられる。すなわち、コンピュータを用いて、質問に対して生じた反応量を自動計測し、計測した反応量を解析して判定を下すことによって、虚偽検出の客観性を向上させることができよう。

現在まで、幾つかのコンピュータ判定のための手法が考案されているが（足立，1989，1993；徳田，1993；Adachi, 1995），これらのシステムの解析判定法は主に数理統計学の判別分析法に基づくものである（McLachlan, 1992）。しかしながら、判定に適用できる解析法は、統計学的判別分析に限られるわけではなく、他にも有用な方法があろう。特に、近年注目を集めている神経回路網モデル（ニューラルネットワーク；Neural network）は、解析法の重要な候補である。

神経回路網モデルは、神経細胞の回路網を人工的に構築して、脳の情報処理過程をモデル化したものである（Rumelhart, Hinton & PDP Research Group, 1986; 甘利, 1989）。この神経回路網は、脳・神経系のモデルとしてだけでなく、工学や医学等の応用諸領域で、パターン判別の有効な手段として顕著な成果を挙げている（麻生, 1988）。こうした成果を見渡すと、生理反応量のパターン判別である虚偽検出検査の判定に、神経回路網モデルを適用して、その判定精度を評価する必要があると考えられる。そこで、本研究では、神経回路網モデルの中でも特に虚偽検出検査の判定に適合するフィードフォワード（feed-forward）型多層回路網を取り上げ、それに基づくコンピュータ判定手法を提示する。そして、その判定精度を実験データによって評価すると同時に、従来の判別分析法による判定との比較を行う。

神経回路網モデルによる判定

入力層、隠れ層および出力層からなるフィードフォワード型回路網を、判定に用いる。各層は複数のユニット（神経細胞）からなるが、出力層のユニット数は、陰性か陽性かの二群判別が目的であるので1個とする。以上の神経回路網を図示したのが、Figure 1である。図中の矢印で示す結合（シナプス）に沿って、虚偽検出検査で得られた反応量（入力変数）が

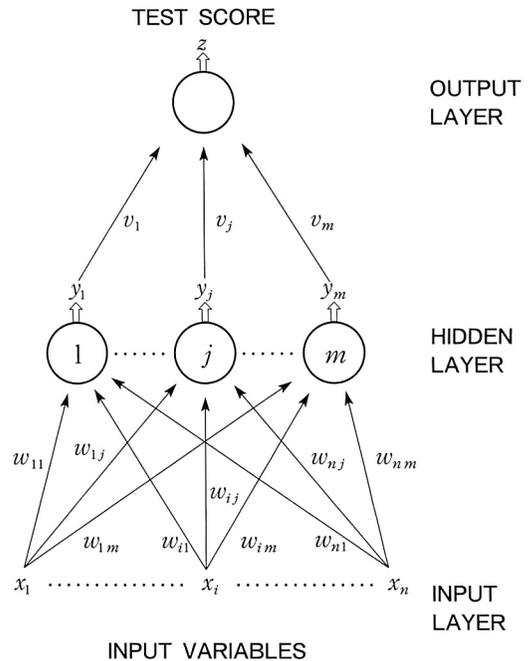


Figure 1 An illustration of the feed-forward neural network

変換処理を受けながら転送される。仮に、入力層を網膜、隠れ層を網膜像から有用な情報を抽出する脳機構、出力層を最終的な視覚判断を行う脳の高次機構とみなせば、この回路網は、視察した対象から判断に至る情報処理過程に対応し、これをプログラム化すれば視覚判断過程の客観的遂行が可能になるといえよう。以下、Figure 1に沿って、判定のための神経回路網の挙動を概説する。

まず、回路網に、被験者が示した呼吸運動・皮膚電気反応・脈波等の n 種の生理指標に関する反応量 x_1, x_2, \dots, x_n が入力される。これらの入力信号は、結合（図中の矢印）を通して隠れ層に送られるが、ここで重みづけを受ける。すなわち、 i ($i = 1, \dots, n$) 番目の入力信号 x_i は、隠れ層のユニット j ($j = 1, \dots, m$) への結合を通る際に w_{ij} で重みづけられ $w_{ij}x_i$ となる。従って、隠れ層のユニット j ($j = 1, \dots, m$) には、重みづけられた n 種の信号 $w_{1j}x_1, w_{2j}x_2, \dots, w_{nj}x_n$ が送られ、それらの合計

$$w_{1j}x_1 + w_{2j}x_2 + \dots + w_{nj}x_n = \sum_{i=1}^n w_{ij}x_i$$

がユニット j の受信信号となる。ユニット j は、受信信号をロジスティック関数 (f と表す) によって、

0以上1以下の値 $y_j = f\left(\sum_{i=1}^n w_{ij}x_i\right)$ に変換する。すなわち、ユニット j は、

$$y_j = f\left(\sum_{i=1}^n w_{ij}x_i\right) = \frac{1}{1 + \exp\left(-\sum_{i=1}^n w_{ij}x_i\right)}$$

を算出し、この y_j を結合を通して出力層に送る。ここでも前層と同様の処理が施され、信号 y_j は v_j で重みづけられ、重みづけ信号の総和 $\sum_{j=1}^m v_j y_j$ が出力ユニットの受信信号となる。出力ユニットは、隠れ層のユニットと同様に、受信信号をロジスティック変換して、

$$z = f\left(\sum_{j=1}^m v_j y_j\right) = \frac{1}{1 + \exp\left(-\sum_{j=1}^m v_j y_j\right)}$$

を出力する。この出力値 z ($1 \leq z \leq 1$) は、被験者が陽性である度合いを表し、これを検査得点と呼ぶことにする。検査得点 z が0と1の中間値0.5以下なら陰性、0.5より大きければ陽性と判定する。

さて、以上の神経回路網モデルは、推定すべき $n \times m + m$ 個のパラメータ、つまり、各結合の重み w_{ij} ($i = 1, \dots, n, j = 1, \dots, m$)、 v_j ($j = 1, \dots, m$) を含む。従って、実際に判定を行うためには、それに先だって、一連のパラメータの値を求めておかなければならない。このために、既に検査を受けて陰性か陽性かが明らかになった複数被験者の反応量データ(学習用データ)を用いて、パラメータを推定(学習)する。パラメータ推定手順の概略は、以下のとおりである。

(1) パラメータ、すなわち、重み w_{ij} 、 v_j に適当な初期値を与える。(2) 学習用データの各被験者の反応量を回路網に入力して、出力された検査得点 z と、「正しい(理想的な)得点」との二乗誤差を求める。すなわち、被験者が陽性のときは検査得点が1となるのが理想であるので、二乗誤差は $(1 - z)^2$ となり、陰性のときは検査得点が0となるのが理想なので二乗誤差は $(0 - z)^2$ となる。この手続きを、学習用データの全被験者のデータについて行い、二乗誤差の合計を求める。(3) 合計二乗誤差が小さくなるように、各重み w_{ij} 、 v_j の値を更新する。すなわち、現在の重みの値から、(合計二乗誤差を重みで微分した量) \times (適当なステップ幅) を減じた値を、新たな重みの値とする。そして、新たな重みを用いて、再び(2)の手続きを行い、合計二乗誤差を求

める。(2)と(3)の手続きを、収束したと判断される(二乗誤差の改善が殆どなくなる)まで、反復する。ここまで概説したパラメータ推定の手続きは、誤差逆伝搬(back propagation) アルゴリズムと呼ばれる(Rumelhart & Hinton, 1986)。

以上の神経回路網による判定およびパラメータ推定の計算プログラムをFORTRANで作成し、ワークステーション(Spark Station 2)によって、以下に記すデータに適用した。

データ

適用対象のデータは、実験によって収集された虚偽検出検査(緊張最高点質問法)の反応データである。以下に、データ収集の手続きと、神経回路網への入力変数とした反応量の算出方法を記す。

1. 実験手続き

大学生48名(男子21名、女子27名)を被験者とし、各被験者ごとに個別に実験を行った。

まず、被験者に、実験者が中を見れない個室に入室させ、個室内の机上に置かれた2つの封筒のいずれか1つを選ばせた後、選んだ封筒中の用紙に記された課題を個室で行わせた。封筒の中身は、陽性群(犯人)の役割あるいは陰性群(無実)の役割を記した用紙からなる。陽性群の役割の用紙には、「個室のレターケース中にある万年筆と現金6万円(現金にはハンカチが被せられている)を盗んだ後、退室して下さい。この課題の後で受ける虚偽検出検査で陰性(無実)と判定されれば、報酬として500円が与えられます」との教示が記されていた。一方、陰性群の役割の用紙には「5分間ほど、個室内で何もせずに休止してから、退室して下さい」との教示が記されていた。なお、幾名かの被験者については、上記の2つの封筒に同じ役割の用紙を挿入し、陽性群と陰性群の被験者が同数になるようにした。以上の手続きにより、模擬窃盗を行う陽性群、および、模擬窃盗の内容を知らされない陰性群の各群に、24名の被験者が割り当てられた。

被験者が個室から退室した後、上記の模擬窃盗に関する質問表を用いた虚偽検出検査を行った。質問表は、(A)「現金の上に置いてあった物」(B)「盗まれた現金の金額」(C)「盗まれた品物」に関する3表を準備し、(A)は「めがね、セロテープ、はがき、ハンカチ、ピンセット」の5項目、(B)は「2

万円, 4万円, 6万円, 8万円, 10万円」の5項目, そして, (C)は「ライター, 腕時計, ネクタイピン, 万年筆, 電卓」の5項目で構成された。ここで, 下線を引いた項目が裁決質問, 他の項目が非裁決質問である。各被験者について, (A), (B), (C)のいずれか2種の質問表を用いて検査を行った。

各質問表について, 検査を次の手続きで行った。5つの質問項目を録音テープの再生によって約20秒間隔で提示して, 各質問項目に対して, 被験者に「いいえ」と返答させた。こうした質問系列を, 項目提示順序を変えながら5回反復した。質問系列中の被験者の呼吸運動, SCR (皮膚コンダクタンス反応), 上腕脈波および指尖脈波をLafayette社製ポリグラフによって測定し, データレコーダ (TEAC社製MR30) で磁気記録した。この記録を, 実験終了後, A/D変換ボード (Micro-science社製ADM1998-BPC) によってサンプリングしてコンピュータに転送し, 次の計測処理を行った。

2. 反応計測

呼吸運動, SCR, 上腕脈波および指尖脈波のそれぞれについて, 以下に記す方法で反応計測を行った。

(1) 呼吸運動: 各質問項目について, 返答後10秒間の呼吸波を0.1秒間隔でサンプリングしたので, 計100個の波高値 (a_1, a_2, \dots, a_{100}) が得られたが, これらの波高値に基づいて, 吸気速度, 呼気速度および自己相関係数を求めた。吸気速度は, 隣接する波高値の中で, 上昇した (吸気方向に移動した) 波高値の差を足し挙げた後に, 吸気時間で除したもので, 厳密には, 次のように表せる。

$$\sum_{i=2}^{100} \delta(a_{i-1}, a_i) |a_{i-1} - a_i| / \sum_{i=2}^{100} \delta(a_{i-1}, a_i)$$

ここで, $a_i > a_{i-1}$ のとき $\delta(a_{i-1}, a_i) = 1$, $a_i \leq a_{i-1}$ のとき $\delta(a_{i-1}, a_i) = 0$ である。呼気速度は, 吸気速度とは反対に, 隣接する波高値の中で, 下降した (呼気方向に移動した) 波高値の差を足し挙げた後, 呼気時間で除したもので,

$$\sum_{i=2}^{100} \{1 - \delta(a_{i-1}, a_i)\} |a_{i-1} - a_i| / \sum_{i=2}^{100} \{1 - \delta(a_{i-1}, a_i)\}$$

と表せる。自己相関係数は, 隣接する波高値の対 (a_1, a_2), (a_2, a_3), \dots , (a_{99}, a_{100}) の積率相関係数

である。

(2) SCR: 質問提示開始0.5秒後から返答後5秒後までの間隔を対象とし, コンダクタンスの最大上昇量を求め, その平方根をSCR振幅とした。

(3) 上腕脈波: 質問提示前3拍の振幅を平均し (質問前平均値), 質問提示開始直後から15秒後までの振幅の平均値 (質問後平均値) から質問前平均値を引いた値を振幅変化量とした。

(4) 指尖脈波: 質問提示前3拍の基線の高さを平均し (質問前平均値), 質問提示開始後15秒間の基線の高さの平均値 (質問後平均値) から質問前平均値を引いた値を基線変化量とした。また, 質問提示開始後15秒間の振幅について, 3次の自己回帰モデル (足立, 1993) で予測される振幅と観測振幅との差の合計を求め, 振幅予測偏差とした。

(5) 心拍率: 質問提示開始後15秒間の指尖脈波から, 各秒ごとの心拍率を求め, 3次の自己回帰モデル (足立, 1993) で予測される心拍率と観測心拍率との差の合計を求め, 予測偏差とした。

3. 反応量の算出

反応計測により, 各質問項目に対して, 吸気速度, 呼気速度, 呼吸自己相関, SCR振幅, 上腕脈波振幅変化, 指尖脈波基線変化, 指尖脈波予測偏差, 心拍率予測偏差の計8指標の計測値が得られた。すなわち, 各質問表について, 8種の計測値がそれぞれ25 (5質問×5系列) 個得られたが, 各指標内で25個の計測値の平均が0, 分散が1となるように標準得点化した。各指標について, 全系列の裁決質問に対する反応計測値 (標準得点) の平均から, 非裁決質問に対する反応計測値 (標準得点) の平均を減じた値を求め, これを反応量 (裁決質問に対する反応の特異性を表す量) とした。以上の手続きより, 各質問表について, 計8種の反応量 (x_1, x_2, \dots, x_8) が得られたが, これら8変数に基づいて, 判定精度の評価試験を行った。

評価試験

各質問表の反応記録について, 陰性・陽性の判定を行った。すなわち, 延べ96個 (被験者48名×2質問表) の記録のそれぞれについて, 8種の反応量 (x_1, x_2, \dots, x_8) を神経回路網への入力データとして, パラメータの推定および判定を行った。ただし, 神経回路網の隠れ層のユニット数 m は, 予備的分析の

結果より8個とした。神経回路網による判定と同時に、その比較対象として、判別分析、すなわち、線形 (linear) 判別分析と二次 (quadratic) 判別分析 (付録参照) による判定精度の評価も行った。手続きと結果は、以下に示すとおりである。

1. 学習用データによる判定

上記の96個の記録を用いて、神経回路網のパラメータ、つまり、重みを誤差逆伝搬アルゴリズムで求めた。ここで、重みの初期値は一様乱数を用いて設定し、重み更新のステップ幅は0.1とした。重み更新の反復は、合計二乗誤差の変化量が0.001以下となった時点で終了した。以上の方法で求めた重みを用いて、96個の各記録を判定した。判定規則は、前述したように、検査得点 z が0.5以下なら陰性、0.5より大きければ陽性とするものである。

さらに、線形および二次判別分析による判定も行った。すなわち、96個の記録からパラメータを推定した後、推定したパラメータに基づいて、各記録の陽性事後確率を算出した。陽性の事後確率を検査得点として、神経回路網の場合と同様に、得点が0.5以下なら陰性、そうでなければ陽性と判定した。

判定結果は、Table 1 に示すとおりである。神経回路網による判定の精度は非常に高率であり、判別分析による判定精度を上回った。しかしながら、以上の判定精度の評価法は、現実の場面には適合しない。すなわち、実際には、陰性・陽性が明らかになった過去の検査記録 (学習用データ) から推定したパ

ラメータを、陰性・陽性が未知の記録に適用することになるが、上記の方法はパラメータ推定に用いた学習用データを、再び判定対象としている。こうした評価法から得られた判定率は、統計学で apparent rate と呼ばれ、真の正判定率を過大評価したものになっている。こうした過大評価を是正して正しい評価値を求める方法として、学習用データと判定用データを分離する交差検証法 (Cross-validation method) があり (McLachlan, 1992)、次にこの方法による手続きを記す。

2. 交差検証法による評価

計48名の被験者のうち、47名の被験者の記録、つまり、94個 (47×2 質問表) の記録を学習用データとして神経回路網および判別分析のパラメータを推定し、これに基づいて、残り1名の被験者の2つの記録を判定した。以上の手続きを、48名の全被験者の記録が判定対象になるように、48回繰り返した。パラメータの推定法と判定規則は、前節の方法と同じである。

判定結果は、Table 2 に示す通りである。陰性・陽性の両群をまとめた全群 (combined group) の正判定率は77%で、前節の学習用データを判定した場合に比べて、神経回路網による判定精度が著しく低下していることが窺える。すなわち、神経回路網の正判定率は、線形判別分析とほぼ同等で、二次判別分析をやや下回った。

Figure 2 は、神経回路網および判別分析の検査得

Table 1 Percentage outcomes of classifications on the learning sample

Classifier	Innocent group		Guilty group		Combined group	
	Correct	Error	Correct	Error	Correct	Error
Neural Network	100.0	0.0	85.4	14.6	92.7	7.3
Linear Discrimination	83.3	16.7	77.1	22.9	80.5	19.5
Quadratic Discrimination	89.6	10.4	79.2	20.8	84.4	15.6

Table 2 Percentage outcomes of classifications by the cross-validation method

Classifier	Innocent group		Guilty group		Combined group	
	Correct	Error	Correct	Error	Correct	Error
Neural Network	87.5	12.5	66.7	33.3	77.1	22.9
Linear Discrimination	81.2	18.8	72.9	27.1	77.1	22.9
Quadratic Discrimination	79.2	20.8	79.2	20.8	79.2	20.8

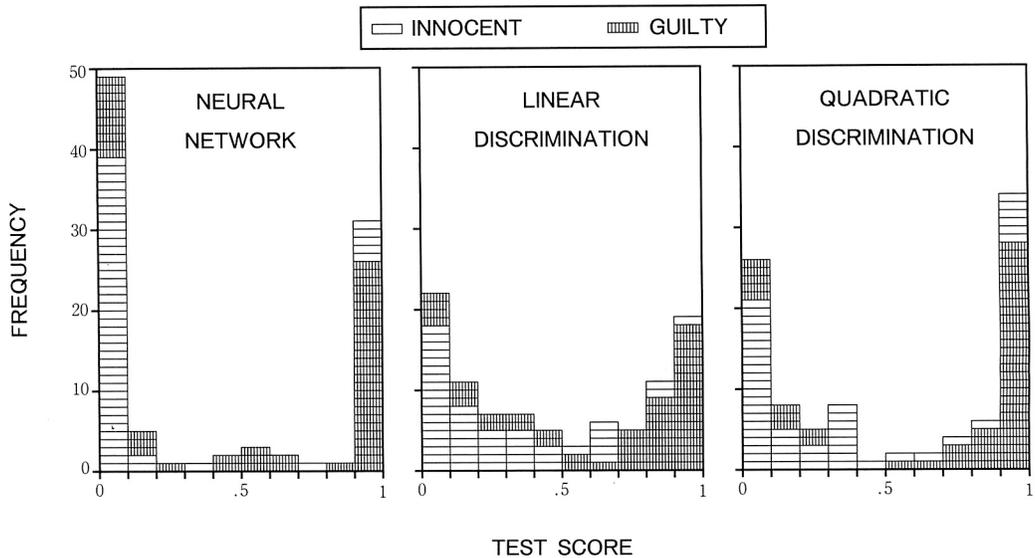


Figure 2 Distributions of test scores for neural network and linear and quadratic discriminant analyses

点の分布を示すが、神経回路網による得点は、判別分析よりやや極端な値を示すことがわかる。すなわち、神経回路網による判定では、得点のほとんどが両極端の0または1に近い値をとり、中間的な値が少ない。さらに、こうした傾向は誤判定した記録にも見られ、得点を判定についての確信の程度とみなせば、確信をもって正判定を下す場合もあれば、確信をもって誤判定を下す場合も少なくないことが窺える。

考 察

神経回路網モデルによる自動判定の精度を交差検証法で評価した結果、正判定率は77%であった。この結果は、線形判別分析による判定と同じで、また、二次判別分析よりわずかに劣った。従って、神経回路網モデルの虚偽検出検査へ適用可能性は示されたものの、従来の統計学的手法に対する優越性は認められなかった。

評価試験から、判定精度だけでなく、神経回路網モデルが持つ幾つかの性質が明らかになった。その1つは、学習用データを判定する精度に比べて、陰性・陽性が未知のデータを判定する精度が著しく低下することである。こうした傾向は、神経回路網モデルが数多くのパラメータを持ち、学習用データに過適合 (over-fitting) することに由来すると考えら

れる。また、正判定だけでなく誤判定の場合にも、検査得点が0または1の両極端に近い値をとる傾向が見られた。検査得点は、陰性・陽性の二群判別に関する確信の度合いを表すべきもので、たとえ誤判定をした場合も中間的な値をとるのであれば、応用上の意味はあるが、上記の傾向はこれに反するものである。すなわち、極端な得点を示すという神経回路網モデルの性質は、応用上不都合な特性であると言えよう。

神経回路網モデルの望ましくない結果を前段で述べたが、このことから直ちに、その有用性が否定できるわけではない。神経回路網が従来の判別分析とほぼ同等の正判定率を示したことから、その虚偽検出検査への適用可能性は確認されたわけであり、この結果を踏まえて、上述したような問題点を克服すべく神経回路網の改良を行うことが重要な課題であると言えよう。

付録. 判別分析の概略

判別分析法の中でも主要な手法である線形判別分析と二次判別分析 (McLachlan, 1992) の概略を、反応量 $X = [X_1, \dots, X_n]$ に基づいて、陰性・陽性の二群判別を行う場面を想定して記す。

両分析法ともに、反応量が、陰性と陽性の両群で異なる多変量正規分布に従うものと仮定する。ここ

で、陰性群の多変量正規分布のパラメータつまり平均ベクトル、共分散行列を、それぞれ μ_1, Σ_1 で表し、分布の密度関数を $\phi(\mathbf{X}; \mu_1, \Sigma_1)$ で表す。同様に、陽性群の多変量正規分布のパラメータを μ_2, Σ_2 で表し、その密度関数を $\phi(\mathbf{X}; \mu_2, \Sigma_2)$ で表す。以上の仮定の上で、両群の共分散行列が等しい ($\Sigma_1 = \Sigma_2$) という制約をおくのが線形判別分析で、この制約を設けないのが二次判別分析である。これが線形および二次判別分析の違いであり、以下に記すことは両者に共通する。

陰性および陽性の事前確率を、それぞれ、 π_1, π_2 とおくと、反応量 \mathbf{x} を示した被験者が陽性である確率 (陽性事後確率) p は、ベイズの定理より、

$$p = \frac{\pi_2 \phi(\mathbf{x}; \mu_2, \Sigma_2)}{\pi_1 \phi(\mathbf{x}; \mu_1, \Sigma_1) + \pi_2 \phi(\mathbf{x}; \mu_2, \Sigma_2)}$$

で与えられ、陰性である確率は、 $1 - p$ で与えられる。判定に先だって、学習用データから、最尤法によって、各パラメータの値を推定して、上式に代入すれば、陽性 (陰性) の確率が得られる。なお、本研究の評価試験では、 $\pi_1 = \pi_2 = 0.5$ とした。

文 献

- 1) 足立浩平：虚偽検出のための乱塊配置モデルに基づく判別関数。科学警察研究所報告法科学編，1989，**42**，160-167.
- 2) 足立浩平：多変量解析による虚偽検出検査の自動判定。応用心理学研究，1993，**18**，55-63.
- 3) Adachi, K.: Statistical classification procedures for polygraph tests of guilty knowledge. *Behaviormetrika*, 1995, **22**, 49-66.
- 4) 足立浩平, 鈴木昭弘：生理指標間の虚偽検出精度の比較。応用心理学研究，1991，**16**，33-43.
- 5) 甘利俊一：神経回路網モデルとコネクショニズム。東京大学出版会，1989.
- 6) 麻生英樹：ニューラルネットワーク情報処理。産業図書，1988.
- 7) Ben-Shakhar, G., & Furedy, J. J.: Theories and applications in the detection of deception: A psychophysiological and international perspective. Springer-Verlag, New York, 1990.
- 8) Lykken, D. T.: The GSR in the detection of guilt. *Journal of Applied Psychology*, 1959, **43** 385-388.
- 9) McLachlan, G. J.: Discriminant analysis and statistical pattern recognition. Wiley, New York, 1992.
- 10) Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representation by back-propagating errors. *Nature*, 1986, **323**, 533-536.
- 11) Rumelhart, D. E., McClelland, J. L., and PDP Research Group: Parallel distributed processing. Vols. I, II. MIT Press, Massachusetts, 1986. 甘利俊一 (監訳) PDP モデル-認知科学とニューロン回路網の探索。産業図書，1989.
- 12) 鈴木昭弘：虚偽検出検査。新美良純，鈴木次郎 (編) 皮膚電気活動，Pp. 192-205，星和書店，1986.
- 13) 徳田 豊：ポリグラフ検査の自動判定システムの実用化の試み。科学警察研究所報告法科学編，1993，**46**，22-26.