

論理回帰分析法による脳機能画像からの知識の発見

Knowledge Discovery from Functional Brain Images by Logical Regression Analysis

森田 千絵
Chie Morita

(株) 東芝
Toshiba Corporation
chie.morita@toshiba.co.jp

柿元 満
Mitsuru Kakimoto

(同 上)
mitsuru.kakimoto@toshiba.co.jp

菊池 吉晃
Yoshiaki Kikuchi

東京都立保健科学大学
Tokyo Metropolitan University of Health Sciences
ykikuchi@post.metro-hs.ac.jp

月本 洋
Hiroshi Tsukimoto

(株) 東芝
Toshiba Corporation
hiroshi.tsukimoto@toshiba.co.jp

keywords: brain, functional brain images, knowledge discovery, fMRI, Logical Regression Analysis, KDD, data mining

Summary

As a result of the ongoing development of non-invasive analysis of brain function, detailed brain images can be obtained, from which the relations between brain areas and brain functions can be understood. The relations between brain areas and brain functions are described by rules. Knowledge discovery from functional brain images is knowledge discovery from pattern data, which is a new field different from knowledge discovery from symbolic data or numerical data. We have been developing a new method called Logical Regression Analysis. The Logical Regression Analysis consists of two steps. The first step is a regression analysis. The second step is rule extraction from the regression formula obtained by the regression analysis. In this paper, we apply the Logical Regression Analysis to functional brain images to discover relations between a brain function and brain areas. We use nonparametric regression analysis as a regression analysis, since there are not sufficient data to obtain linear formulas using conventional linear regression from functional brain images. Experimental results show that the algorithm works well for real data.

1. は じ め に

近年、脳の非侵襲計測技術は急速に発達し、詳細な脳機能画像が得られるようになった。機能的磁気共鳴画像 (functional Magnetic Resonance Imaging : fMRI) は脳の非侵襲計測の代表的な手法の一つで、磁気共鳴現象を利用して脳の神経活動に伴う血流の変化を捉え、画像化するものである。本研究では、fMRI で撮影した脳機能画像から、論理回帰分析法を用いて脳の機能と脳の部位の対応関係に関する知識の発見を行う。

本研究の fMRI 画像からの知識の発見には次のような特徴がある。

- (1) 知識発見の新しい分野であるパターンデータからの知識発見
- (2) fMRI 画像の統計的処理の高度化
- (3) 論理回帰分析法による知識発見

以下にこの 3 点について説明する。

まず、パターンデータからの知識発見について説明する。現在の知識発見 (KDD, データマイニング) は基本的に記号データ、数値データからの知識発見である。し

かし現実には、多くの画像等のパターンデータが存在する。したがって、パターンデータからの知識発見は重要な研究主題である。ところが、パターンデータには、隣接するデータ間 (例えば画像データで隣接する画素間) に高い相関があるという特徴がある。これは記号や数値データとは異なる点であり、知識発見の際に考慮されねばならない。パターンデータと言っても、1 次元では時系列データ、2 次元では画像データ等種々存在する。また画像と言っても、リモートセンシング画像、医用画像等と種々ある。さらに医用画像でも脳、胸、胃等種々あるが、脳が最も機能分担が進んでいて複雑であり、その機能分担を調べるのが現在の脳科学の重要課題であるので、脳を知識発見の対象に選ぶことにする。

次に、fMRI 画像の統計的処理の高度化について述べる。脳の機能と部位の対応関係 (例えば、言語野のある部位が失語症のある症状の責任部位である、というような関係) は、単純な機能の場合には単一部位が責任部位であることが多いが、機能が複雑もしくは高度になれば、単一部位ではなく複数部位が責任部位になる。それら複数部位は、脳が神経回路なので、直列、並列、抑制で接

続されている。直列につながっている (同時に生起する場合) 部位は論理積 AND で、並列につながっている (冗長な場合) 部位は論理和 OR で、抑制として機能している部位は否定 NOT で記述される。したがって機能と部位の対応関係は基本的に AND , OR , NOT で記述されることになる。統計的处理をするアルゴリズムでは関連のありそうな部位を提示するが、それらの部位の結合関係までを表現した知識を提示 (発見) するものではない。本研究は、統計的处理を高度化し、結合関係を表現する知識を発見するものである。

最後に、論理回帰分析法による知識発見について説明する。筆者らは、予測を行い、かつ、人間に理解可能な知識を発見するために、「論理回帰分析法」を開発してきた。論理回帰分析法とは、与えられたデータから統計処理などで回帰式を求め、その回帰式からルール (論理的規則) を抽出することにより、理解可能な知識を発見する手法である。特に、後半のルール抽出の手法をここでは「近似法」と呼ぶことにする。論理回帰分析法は回帰分析を利用しているが、教師あり学習の問題に対する実績も報告されている [月本 96]。論理回帰分析法の前半の回帰式は線形回帰式を基本として、(多変数) 自己回帰式、ノンパラメトリック回帰式、回帰木等あり、それらからルール抽出が可能である。また、それを拡張してニューラルネットワークからもルールを抽出することができる [Tsukimoto 97]。さらに、微分方程式からも、差分方程式に変換することによって、ルールを抽出することができる。本研究ではノンパラメトリック回帰式からルール抽出を行う。本研究で変数に対応するのは各画素であり、変数の数は数千以上になるのに対し、得られるサンプル数はかなり少ない。そのため通常の重回帰分析は実行できない。ノンパラメトリック回帰分析はデータの値の間に連続性や滑らかさ等の制約を入れることにより、少ないサンプル数でも回帰することが可能となっている。

我々は、最終的には新たな知識の発見を目指す。第一段階として論理回帰分析法の fMRI 画像への適用が有効であることを確認する。そのため、本論文では既に得られている知見を論理回帰分析法によって再発見することを目的として、従来手法である z-score [SPM] との比較を行う。

脳の機能と部位の対応関係を調べる問題は、ある機能が働いている/いないをクラスとした教師あり学習の問題と捉えることができる。筆者らは、過去に教師あり学習に対して論理回帰分析法を適用し、良好な結果が得られることを確認している [月本 96]。一方 z-score は基本的に、脳のある機能が働いている/いないに関して変数の雑音を加算平均することで除去し、その二つのクラスでの平均値の差を出している。この処理は、一般に教師あり学習には有効ではない。したがって、脳の機能と部位の対応関係を調べる問題に対して、論理回帰分析法の方が z-score より優れていると考えられる。z-score に対す

るもう一つの優位点は、論理積だけでなく論理和や否定の結合関係を表現できることである。

以下では、2 章で論理回帰分析法による fMRI 画像からの知識発見のアルゴリズムについて説明し、3 章では実際の fMRI 画像からの知識発見の実験と、従来手法の z-score との比較を行う。最後に、4 章でまとめとする。

2. 論理回帰分析法による fMRI 画像からの知識発見のアルゴリズム

2.1 fMRI 画像からの知識発見

例として画像が 64×64 で表されているとすると、図 1 のような画像が一つのサンプルデータになる。図 1 内の数字は画素の番号を表す。この例では $64 \times 64 = 4096$ 個の画素から成り立っている。

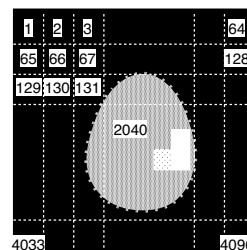


図 1 64×64 の fMRI 画像の例

画素ごとの信号値は数値で得られているので、そのような画像が 10 サンプル得られたとすると、表 1 のようなデータになる。数値が高いほど、画像上で明るく表示されることを表す。クラスの Y, N は何らかの症状が表れているかいないか、機能が働いているかいないかを表す。これより、fMRI 画像からの知識の発見は基本的

表 1 fMRI 画像のデータの例

	1	...	2040	...	4096	クラス
S1	0	...	120	...	2	Y
S2	1	...	142	...	0	N
...
S10	0	...	98	...	1	Y

に教師あり学習の枠組みで処理できることがわかる。しかし fMRI 画像からの知識の発見は、従来の記号や数値データからの知識発見ではなく、パターンデータからの知識の発見という新分野である。パターンデータからの知識の発見が記号や数値データからの知識の発見と違う点は、パターンデータでは隣接する各データ (ある画素に対して、その上下左右の画素) 間に高い相関があることである。

したがって C4.5 等の決定木の学習ではうまくいかない。実際人工データで試してみたがうまくいかなかった [Tsukimoto 98]。これを解決するには、属性間の相関を

考慮した決定木アルゴリズムを新たに開発することも考えられるが、このようなデータに対してノンパラメトリック回帰分析が有効である。ノンパラメトリック回帰分析では、測定値でモデルを構成し、そのモデルを作成する時の評価関数は誤差に滑らかさ等を付加したものを用いる。しかしノンパラメトリック回帰分析で得られるのは線形式であり知識ではない。そこで、近似法によってノンパラメトリック回帰分析からルールを抽出する。

2.2 処理の流れ

図2に論理回帰分析法のfMRI画像への適用の流れを示す。本論文では、抽出されたルールを従来手法のz-scoreと比較する。

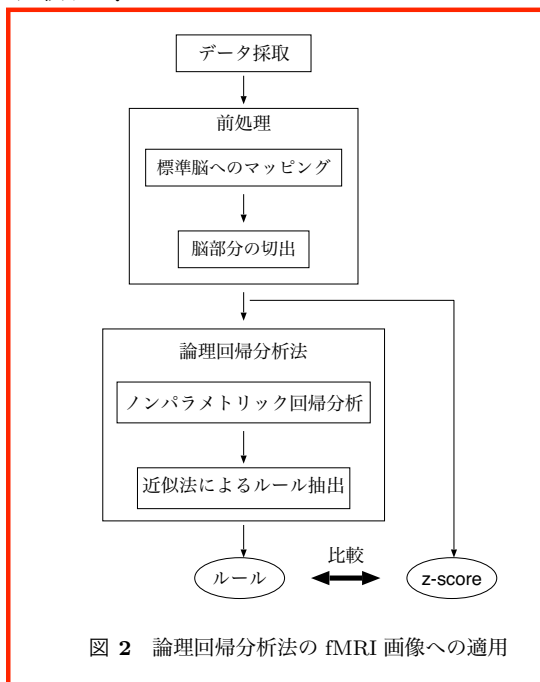


図2 論理回帰分析法のfMRI画像への適用

(1) データ採取

被験者に課題を与え、課題を行っている時(タスク状態と呼ぶ)と休んでいる時(コントロール状態と呼ぶ)のfMRI画像を撮影する。課題とは、たとえば足し算の暗算や指のタッピング(親指と人指し指、中指、薬指、小指を順につけたり離したりする)などである。図3に実際にfMRIで撮影した画像の例を示す。

(2) 前処理

図3に見るように、fMRI画像には脳の外の部分も含まれている。これらの画素を含めると、計算量が増加するだけではなくノイズとなってこの後の分析に悪影響を及ぼす可能性がある。そこで脳の部分のみを切り出す処理を行う。本論文では、SPM(Statistical Parametric Mapping 脳画像分析のためのソフトウェア)[SPM]を用いて脳部分の切出を行う。SPMは、標準脳[Talairach 88]という概念を持っており、fMRI画像を標準脳にマッピングすることによって複数の被験者の脳を扱うことができる。そこでまず、

t



図3 fMRI画像の例

得られたfMRI画像を標準脳にマッピングし、脳部分を切り出す。

(3) ノンパラメトリック回帰分析

脳は3次元であるが、本論文では第一段階として2次元の断面(スライスと呼ぶ)ごとにノンパラメトリック回帰分析を行う(2.3節参照)。

(4) 近似法によるルール抽出

スライスごとに、得られた回帰係数から近似法によってルールを抽出する(2.4節参照)。

2.3 ノンパラメトリック回帰分析

まず、1次元の場合のノンパラメトリック回帰分析を簡単に説明する。

ノンパラメトリック回帰分析は、説明変数の各測定値で回帰式を構成するものである[Eubank 88]。例えば、被説明変数を y とし、説明変数 t の測定値 t_j が m 個ある場合には、

$$y = \sum a_j t_j \quad (j = 1, \dots, m), \quad a_j \text{は実数}$$

となる。被説明変数 y の測定値 y_i が n 個ある場合には

$$y_i = \sum a_j t_{ij} \quad (i = 1, \dots, n)$$

となる。この時、最小にする評価量は誤差だけではなく、連続性もしくは平滑性の項を付加する。連続性の項を付加した場合には以下ようになる。

$$1/n \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^n (\hat{y}_{i+1} - \hat{y}_i)^2$$

となる。 \hat{y}_i は推定される解である。これは隣り合う被説明変数間の1階差分であり被説明変数の連続性を意味する。 λ は連続性の重み係数であり、これが0だと評価量は第1項の誤差だけになり通常の回帰分析になる。逆に λ が非常に大きいと評価量は第2項の連続性の評価量だけになり誤差を無視するので解 \hat{y} は定数になる。

上式は被説明変数に連続性がある場合である。例えば、被説明変数が空間的に隣接しているとか、時間的に隣接している場合である。そうでない場合には、被説明変数の連続性は制約として使えない。その場合には係数 a の連続性を制約として使う。この係数の連続性は隣り合っている説明変数の測定値が被説明変数におよぼす影響に

連続性を仮定するものである [三輪 98]。したがって評価量は以下ようになる。

$$1/n \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^m (\hat{a}_{j+1} - \hat{a}_j)^2$$

この λ の最適値は例えばクロスバリデーションで選ぶ。

上式は 1 次元であるが、2 次元に拡張するには、回帰係数 a_j の連続性に関する評価量を変更すれば良い。1 次元の場合には隣り合う画素は両側の二つしかなかったが、2 次元の場合には隣り合う画素は縦横に計 4 個ある。

したがって回帰係数間の制約は $(\hat{a}_{j+1} - \hat{a}_j)^2$ ではなく、画像で隣接している画素間の制約になる。例えば図 1 の 66 番の画素ではその周りの 2, 65, 67, 130 との制約になり、

$$(\hat{a}_{66} - \hat{a}_2)^2 + (\hat{a}_{66} - \hat{a}_{65})^2 \\ + (\hat{a}_{67} - \hat{a}_{66})^2 + (\hat{a}_{130} - \hat{a}_{66})^2$$

となる。

2.4 ルール抽出

筆者らは近似法と呼ばれるルール抽出アルゴリズムを開発してきた。近似法の基本は、回帰分析で得られた線形式を連続ブール関数 [月本 94] に近似することである。

ブール関数の定義域は $\{0,1\}$ であるが、連続ブール関数とはそれを拡張して定義域を $[0,1]$ としたものである。

線形式から連続ブール関数への近似は、多重線形関数空間に基づいている。多重線形関数は連続ブール関数の拡張であり、多重線形関数の空間はユークリッド空間になる。回帰分析で得られた線形式もこの空間に含まれるので、連続ブール関数に近似することができる。

fMRI 画像からの画素の値は $[0,1]$ ではないが、適当な処理によって $[0,1]$ に正規化することができる。以下では、定義域は $[0,1]$ とする。

詳細は [月本 94, 月本 96] 参照。

§1 連続ブール関数とは

$[0,1]$ での、ブール関数に相当する定性的な表現系を考える。連続領域での定性的表現でもっとも簡単でわれわれが馴染んでいるものは比例、反比例であろう。そこで、ブール関数に相当する定性的な表現系として比例、反比例を定式化することを考える。比例は通常のプロポーション ($y = x$) であるが、反比例は通常のプロポーションと少し違い、 $y = 1 - x$ とする。理由は $y = 1 - x$ がブール関数の否定の自然な拡張であるからである。なお比例は肯定の自然な拡張である。図 4 の左側に $[0,1]$ での比例、右側に $[0,1]$ での反比例を示す。図中で黒丸で示した点は、ブール関数の肯定と否定を表している。ブール関数では定義域が $\{0,1\}$ なので、図 4 の黒丸のように点で表されるが、定義域が $[0,1]$ の場合は黒丸の点以外の領域に関しても扱う。また、ブール関数の論理積と論理和に対しても拡張して $[0,1]$ 上で相当する演算を考え、先の比例、反比例と合わせて、 $[0,1]$ での定性的表現系とする。 $[0,1]$ での論理積、論理和

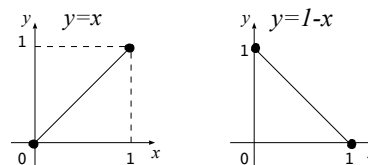


図 4 $[0,1]$ の比例と反比例

もブール代数の論理積、論理和を自然に拡張して得られる。この定性的表現系はブール代数の公理を満たす [月本 94] ため、ここではこれを連続ブール関数と呼ぶ。

連続ブール関数において、肯定は上がることを意味する。例えば「 $x \wedge \bar{y} \rightarrow z$ 」は「 x が上がり、かつ、 y が下がれば、 z が上がる」と読まれる。

§2 多重線形関数とは

多重線形関数の定義は以下の通りである。

$$\sum_{i=1}^{2^n} a_i x_1^{e_{i1}} \cdots x_n^{e_{in}}$$

ただし a_i は実数、 x_i は変数、 e_{ij} は 1 か 0、 n は変数の数である。言い替えれば、多重線形関数は $p_i x_1^{k_1} \cdots x_n^{k_n}$ ($k_j \geq 2$) という項がない多項式関数である。

例えば、2 変数の多重線形関数 $f(x,y)$ は

$$f(x,y) = pxy + qx + ry + s$$

である。

多重線形関数空間は、連続ブール関数のブール代数の原子を基底にする線形空間である。また、適当なノルムを導入することにより、ユークリッド空間になる。

§3 線形式から連続ブール関数への近似

線形式も連続ブール関数も多重線形関数空間に属する。そこで、ユークリッド距離を用いて線形式を連続ブール関数に近似することができる。

多重線形関数空間は連続ブール関数のブール代数の原子を基底にする線形空間なので、多重線形関数はベクトルで表現される。このベクトルを論理ベクトルと呼ぶ。線形式の論理ベクトルを (f_i) 、連続ブール関数の論理ベクトルを (g_i) ($g_i = 0$ or 1) とすると、近似の方法は以下の通りである。

$$g_i = \begin{cases} 1 & (f_i \geq 0.5) \\ 0 & (f_i < 0.5) \end{cases}$$

例

x, y の 2 変数として線形式を

$$0.3x - 0.5y + 0.6$$

とする。2 変数の場合の連続ブール関数のブール代数の原子は $xy, x\bar{y}, \bar{x}y, \bar{x}\bar{y}$ の 4 つなので、この 4 個の定義域

$$(1,1), (1,0), (0,1), (0,0)$$

での線形式の値を求めると

$$(0.4, 0.9, 0.1, 0.6)$$

となる。これが線形式の論理ベクトル (f_i) である。これを 0.5 を閾値として連続ブール関数で近似すると

$$(0, 1, 0, 1)$$

となる。これが連続ブール関数の論理ベクトル (g_i) である。この論理ベクトルは $x\bar{y} \vee \bar{x}y = \bar{y}$ を表している。

ところで、変数の数を n とすると、 $i = 1, \dots, 2^n$ である。この方法の計算量は明らかに変数の数の指数オーダーである。したがって、計算量が多項式オーダーのアルゴリズムが必要になる。筆者らはこの多項式オーダーのアルゴリズムを開発した [月本 96]。

§4 多項式オーダーのアルゴリズム

線形式を

$$p_1x_1 + \dots + p_nx_n + p_{n+1}$$

とする。この線形式に最も近い連続ブール関数を以下のアルゴリズムで求める。

- (1) 近似後の連続ブール関数に、 $x_{i_1} \cdots x_{i_k} \bar{x}_{i_{k+1}} \cdots \bar{x}_{i_l}$ という項が存在するかどうかを

$$p_{n+1} + \sum_{i_1}^{i_k} p_j + \sum_{1 \leq j \leq n, j \neq i_1, \dots, i_l, p_j \leq 0} p_j \geq 0.5 \quad (1)$$

で判定する。第 1 項は定数項、第 2 項は存在を判定する項に肯定で含まれる変数に対応する係数、第 3 項は存在を判定する項に含まれない変数に対応する係数でかつ負のもの、である。

- (2) 存在する項を論理和で接続する。
(3) 上記 1, 2 を最低次の項から判定していき、ある次数で停止する。

例

前述の例

$$0.3x - 0.5y + 0.6$$

で考える。

まず 1 次の項から調べていく。近似後のブール関数に x が存在するかどうかを調べると、第 1 項は $p_{n+1} = 0.6$ 、第 2 項 $\sum_{i_1}^{i_k} p_j = 0.3$ 、第 3 項は判定する項 (今は x) に含まれない変数で負の係数のものなので -0.5 である。すなわち、 $0.6 + 0.3 + (-0.5) = 0.4 < 0.5$ となり、存在しないと判定される。 \bar{x} が存在するかについても $0.6 + 0 + (-0.5) = 0.1 < 0.5$ なので存在しないと判定される。 y と \bar{y} についても同様に調べると、 y は存在しないが \bar{y} は $0.6 + 0 + 0 = 0.6 > 0.5$ となり、判定条件を満たす。したがって、 \bar{y} が近似後のブール関数に存在することがわかる。

次に 2 次の項について調べるが、 \bar{y} が存在することがわかっているので、これに含まれる項、すなわち、 $x\bar{y}$ や $\bar{x}\bar{y}$ は調べる必要がない。よって、 xy と $\bar{x}y$ についてのみ調べるが、それぞれ $0.6 + (0.3 + (-0.5)) = 0.4 < 0.5$ 、 $0.6 + (-0.5) = 0.1 < 0.5$ となり、判定条件を満たさない。この例では 2 変数なので 2 次までで終了し、結局得られたブール関数は \bar{y} である。変数が多い場合にはこの先同様に判定し、適当な次数で停止する。

§5 判定順序の工夫

前項で述べた多項式オーダーのアルゴリズムでは、一つ項を作っては判定するという処理を繰り返す。ところが、fMRI 画像からの知識発見では、扱う変数の数は数千、得られるルールの次数は数十から数百と、非常に大きいので、判定される項の数が膨大になり、考えられる全ての項を判定するのは非現実的である。そこで、項の判定順序を工夫することで、計算速度を向上させる。

ある次数の項を判定する際に、その次数の項の中で判定式 (1) の値の大きい順に項を判定していくことにする。つまり、ある次数の項の中で判定式の値が最大であるような項を最初に判定する。もしこの項が判定式を満たさなければ、この次数では判定式を満たす項は存在しないと言える。なぜなら、最初に判定したのは判定式の値が最大の項であり、それが 0.5 に満たなかったなら、その他の項も判定式を満たさないからである。

ある次数の項の中で判定式が最大である項を求めるには、線形式の係数の絶対値の大きい順に次数個分の係数を選び、選んだ係数が正ならその変数を肯定、係数が負ならその変数を否定とした項を作成すれば良い。

例

x_1, x_2, x_3 の 3 変数とし、

$$0.66x_1 - 0.37x_2 + 0.05x_3$$

を線形式とする。

まず、1 次の項を判定する。係数の絶対値が最大なのは $|0.66|$ 、つまり x_1 に対応する係数でかつ係数が正なのでまず x_1 を判定する。すると判定式は $0 + 0.66 + (-0.37) = 0.29 < 0.5$ となり判定式を満たさない。 x_1 は 1 次の項の中で判定式が最大であったから、これが判定式を満たさない場合、他の 1 次の項 ($x_2, x_3, \bar{x}_1, \bar{x}_2, \bar{x}_3$) は明らかに判定式を満たさない。そこで、すぐに 2 次の項を判定することになる。

2 次の項を判定するため、係数の絶対値の大きい順に二つとると、 $|0.66|$ と $|-0.37|$ である。これは x_1 と x_2 に対応する係数だが、 x_2 に対応する係数は負なので変数を否定して項を作る。つまり、 $x_1\bar{x}_2$ である。これに対する判定式は $0 + (0.66 + 0) + 0 = 0.66 \geq 0.5$ となり、判定式を満たす。そこで残りの 2 次の項について同様に調べると、 \bar{x}_2x_3 が判定式を満たすことがわかる。

次に 3 次の項だが、2 次の項に含まれる項を除くと 3 次で判定式を満たす項はない。結局、判定式を満たす項を論理和で接続すると $x_1\bar{x}_2 \vee \bar{x}_2x_3$ が得られる。

3. 実験

今回の実験では、実データで論理回帰分析法が良好に動作することを確認するため、既に得られている知見に対して論理回帰分析法による再発見を試みる。

論理回帰分析法による結果を、従来手法の一つである z-score と比較する。以下では、まず z-score を紹介し、

指のタッピングと計算のタスクによる実験結果についてそれぞれ述べる。

3.1 z-score

下記の式で画素ごとに z-score を求める。

$$z\text{-score} = \frac{|Mt - Mc|}{\sqrt{\sigma t^2 + \sigma c^2}}$$

ここで、

Mt : タスク画像の信号値の平均値

Mc : コントロール画像の信号値の平均値

σt : タスク画像の信号値の標準偏差

σc : コントロール画像の信号値の標準偏差

である。タスク画像とは、あるタスクを行っている時の画像、コントロール画像とは、あるタスクを行っていない時の画像のことを指す。つまり、z-score が 0 ならば、タスク画像の平均とコントロール画像の平均の差がないと言える。z-score が 1 ならば、タスク画像とコントロール画像の平均の差が画像の信号値のばらつきの分だけ大きいということである。

z-score の値が大きい部位が、そのタスクに関連して動いている部位と考えられる。

3.2 指のタッピング

一つ目の実験は右手の指のタッピングである。指のタッピングとは、人指し指から小指までを順番に親指とつけて離すのを繰り返す動作である。実験条件は以下の通りである。

- 磁場強度: 1.5 テスラ
- 画素数: 64×64
- 被験者数: 1
- タスクサンプル数: 30
- コントロールサンプル数: 30

今回の解像度 (64×64) では 1 つの pixel は 4mm × 4mm となっており、今回の実験で用いた fMRI の撮像には充分と考えられる。しかし、より詳細な解析のために 128×128 など高解像度での実験が望まれる。その場合計算量の増加により現状のままでは困難であるが、計算量を減らした高速版を開発中である。

各スライスごとにノンパラメトリック回帰分析を行った。各スライスごとの誤差を表 2 に示す。スライス番号は頭部の下側から上側への順番である。スライス番号の 0 番から 4 番までが書かれてないが、このデータでは脳の下側が一部欠けていることを意味する。これは fMRI の撮影時の状況と SPM による標準脳へのマップによって発生したものである。

スライス番号 28-33 などの脳の上の方に、回帰誤差が小さいスライスが集中している。そこで、スライス 28 とスライス 29 に対してルールを抽出し、図示したものがそれぞれ図 5 と図 6 である。この時の λ はどちらも 1.0×10^{-4} であった。 λ は必ずしも一意に定まるわけではな

表 2 指のタッピングに関する回帰分析の誤差

スライス	誤差	スライス	誤差
5	0.815	20	0.017
6	0.752	21	0.017
7	0.764	22	0.016
8	0.737	23	0.018
9	0.773	24	0.002
10	0.812	25	0.018
11	0.013	26	0.017
12	0.836	27	0.111
13	0.899	28	0.002
14	0.162	29	0.003
15	0.002	30	0.003
16	0.020	31	0.006
17	0.189	32	0.008
18	0.022	33	0.006
19	0.024		

いが、今回はノンパラメトリック回帰分析の誤差等より適当な値であると考えられる。

ルール画像では、白い部位が肯定 (血流が増加)、濃い灰色の部位が否定 (血流が減少) を表しており、それらの部位は論理積で結合されている。つまり、指のタッピングを行う時、白い部位の活性度が高く、かつ、濃い灰色の部位の活性度が低いというルールである。この実験では、論理和で結合されるようなルールは得られなかった。なお、この図での色の濃さには意味はない。脳の外、脳の内部、肯定、否定の領域を区別するために 4 色に塗り分けた次第である。

図は足元から頭の方を見た断面図なので、図中で向かって左側が人体の右側、向かって右側が人体の左側になる。また、図中の上部が人体の前方 (鼻のある側)、下部が人体の後方 (後頭部側) を表している。

図 5 と図 6 において、図の右側 (人体の左側) に白い部分が見られる。これは、手指の運動野、感覚野が活性化していると思われる。

また、比較として同じデータで z-score を求めたところ、やはり脳の上に z-score の値が大きい部位が多く見られた。図 7 と図 8 は、同じスライスでの z-score の値を表した図である。z-score は各画素に対して値が得られるが、図では値が一定以下である画素は一律に灰色で表示し、一定以上である画素に対して値が大きいほど白く表示してある。図はすべて 256 階調であるが、値の最大値は図によって異なり、図 7 では 4.35、図 8 では 4.54 である。図 7 と図 8 においても、ルールと同様に図の右側 (人体の左側) に特に反応している部位が見られる。

指のタッピングは比較的単純な例であるが、実際の fMRI 画像から論理回帰分析法によって過去に得られた知見を再発見することができた。その他、新たな知識の発見につながるようなルールも得られたが、それについてはさらに実験を重ねた上で別の機会に報告したい。



図 5 ルール画像
(指 スライス 28)



図 6 ルール画像
(指 スライス 29)



図 7 z-score
(指 スライス 28)



図 8 z-score
(指 スライス 29)

3.3 計 算

次に、被験者に、計算をしてもらう実験を行った。これは、頭の中で簡単な足し算を行うもので、タッピングよりやや複雑な実験である。実験条件は以下の通りである。

- 磁場強度: 1.5 テスラ
- 画素数: 64×64
- 被験者数: 2
- タスクサンプル数: 34
- コントロールサンプル数: 36

二人の被験者で同じ実験を行った。それぞれのノンパラメトリック回帰の誤差を表 3 に示す。

表 3 計算に関する回帰分析の誤差

スライス	被験者 1	被験者 2
0	0.924	0.882
1	0.418	0.030
2	0.375	0.538
3	0.016	0.510
4	0.456	0.437
5	0.120	0.469
6	0.965	0.434
7	1.001	0.230
8	1.001	0.388
9	0.968	0.473
10	1.001	0.008
11	1.001	0.066
12	1.001	0.736
13	0.828	0.793
14	0.550	0.822
15	0.528	0.805
16	0.571	0.778
17	0.009	0.007
18	0.089	0.060
19	0.642	0.238
20	0.887	0.514
21	0.282	0.532
22	0.281	0.415
23	0.521	0.422
24	0.814	0.270
25	0.336	0.394
26	0.603	0.008
27	0.535	0.062
28	0.719	0.010
29	0.942	0.310
30	0.898	0.360
31	0.746	0.026

指のタッピングと比較するとノンパラメトリック回帰分析の誤差が大きめだが、スライス 17 の誤差が 2 人とも小さいので、このスライスに関してルール抽出を行った。ちなみにこの時の λ は被験者 1 に対しても、被験者 2 に

対しても $1.0e-4$ であった。図 9、図 10 はスライス 17 に対して抽出されたルールを図示したもので、それぞれ被験者 1、被験者 2 を表す。

ルール画像では、白い部分が肯定、濃い灰色の部分が否定を表しており、それぞれ論理積で結合されていることを意味する。今回も論理和で表現されるルールは得られなかった。なお、この図での色の濃さには意味はない。脳の外、脳の内部、肯定、否定の領域を区別するために 4 色に塗り分けた次第である。

図 9 において、図の右上の白い部分 (人体では左側) はブローカ野に対応するものと思われる。言語的な操作で計算を行ったものかもしれない。また、図の右中央の白い部分は、聴覚野から頭頂連合野で、内的に音韻的な方法で計算を遂行したのかもしれない。左頭頂連合野 (図では右側) は、ここを損傷すると失計算が生じることが多い。図 10 においても、図 9 と同様に図の右側中央から上にかけて白い部分がある。これがブローカ野に対応すると考えられる。また、図の右中央の白い部分も図 9 と同様に聴覚野から頭頂連合野と思われる。

図 11 と図 12 は同じスライスの z-score を表している。どちらも 256 階調だが値の最大値は図 11 は 5.61、図 12 は 2.67 である。図 11 ではルール画像の図 9 と同様に図の右側に反応があることがわかる。図 12 はあまり強い反応ではないが、やはり右側であり、ルール画像と同様の結果が出ている。

計算の課題においても、論理回帰分析法によって脳の機能と部位の対応関係を再発見できた。新たな知識の発見と思われるルールもあったが、慎重に確認した上で別途報告することとする。

3.4 研 究 課 題

今回解像度は 64×64 で行ったが、より詳細な解析のためには 128×128 などに上げることが望まれる。また、今回は脳のスライスごとに処理を行ったが、高度なもしくは複雑な脳の機能に関する脳の部位の同定には 2 次元的な処理では限界があり、脳全体を一度に扱う 3 次元的な分析が必要である。しかし、高解像度や 3 次元での解析では、画素数つまりノンパラメトリック回帰分析の変数の数が増加し、計算量の増加が問題となる。ノンパラメトリック回帰分析の計算量は変数の数の 3 乗であ



図 9 ルール画像
(計算 スライス 17
被験者 1)



図 10 ルール画像
(計算 スライス 17
被験者 2)



図 11 z-score
(計算 スライス 17
被験者 1)



図 12 z-score
(計算 スライス 17
被験者 2)

り、現在は Ultra-SPARC-II 300MHz のマシン十数台を用いた分散処理を行って 2 次元のスライス一枚を処理するのに 3-4 時間要しているが、この問題に対して、今回の問題の特徴を利用して計算量を減らした高速版を開発中である。

また、論理回帰分析法は論理積や論理和、否定など結合関係を表現することができるという特長を持つ。今後はその特長を十分に発揮し、新たな対応関係の発見を目指す。特に否定や論理和についてはこれまでに知見が得られていないため解釈が困難な部分もあるが、慎重な議論の上で本手法の優位性を確認していきたい。

4. お わ り に

fMRI 画像に論理回帰分析法を適用し、過去に得られた知見を再発見することができた。本手法は、従来手法の z-score と比べ、脳の機能に関連する部位の結合関係を発見すること等が特長である。今後、より高度な機能に関する脳の部位の同定には、解像度を上げることや脳全体を一度に扱う 3 次元へ拡張することが必要である。そのため、高速版を開発中である。また、今回の実験で新しい知見らしきものも得られているが、さらに実験を重ね、確認した上で別途報告することとする。

謝 辞

実験に御協力いただきました東北福祉大学感性福祉研究所の畠山英子教授に深く感謝致します。なお、本研究は文部省特定領域研究 (A)「巨大学術社会情報からの知識発見に関する基礎研究 (発見科学)」(課題番号 10143106)の科学研究費の援助を受けて行なわれました。

◇ 参 考 文 献 ◇

- [Eubank 88] R.L.Eubank : *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York, 1988.
- [三輪 98] 三輪 哲久, 高橋 渉, 二宮 正士, Vu Nguyen-Cong, 川口 祐男, 南山 恵: スペクトルデータの解析における PLS 回帰とノンパラメトリック回帰, 第 17 回日本 SAS ユーザ会総会および研究発表会論文集, pp.137-148, 1998.
- [Posner 97] M. I. Posner and M. E. Raichle: *Images of Mind*, W H Freeman & Co, 1997.
- [SPM] <http://www.fil.ion.ucl.ac.uk/spm/>

- [Talairach 88] J. Talairach & Tournoux: *Coplanar Stereotaxic atlas of the human brain*, New York: Thieme Medical, 1988.
- [月本 94] 月本 洋: 古典論理の全ての公理を満たす連続値論理関数について, 電子情報通信学会論文誌, Vol.J77-D-I No.3, pp.247-252, 1994.
- [月本 96] 月本 洋, 森田 千絵, 下郡 信宏: 重回帰分析に基づいた帰納学習アルゴリズム, 電子情報通信学会論文誌, Vol.J79-D-II No.7, pp.1244-1251, 1996.
- [Tsukimoto 97] H. Tsukimoto: Extracting Propositions from Trained Neural Networks, *Proceedings of IJCAI-97*, pp.1098-1105, 1997.
- [Tsukimoto 98] H. Tsukimoto and C. Morita: The discovery of rules from brain images, *The First International Conference on Discovery Science*, pp.198-209, 1998.

〔担当委員：石塚 満〕

2000 年 6 月 5 日 受理

著 者 紹 介



森田 千絵(正会員)

1991 年津田塾大学卒業。同年 (株) 東芝入社。現在研究開発センターで人工知能等の研究に従事。人工知能学会、情報処理学会の会員。



柿元 満

1987 年東京大学大学院理学系研究科物理学専門課程博士課程修了。同年 (株) 東芝入社。現在、研究開発センターにてデータマイニング等の研究に従事。理学博士、情報処理学会の会員。



菊池 吉晃

1978 年東京大学工学部計数工学科卒業。1980 年同大学院修士課程修了。1980-1998 年 東京医科歯科大学難治疾患研究所。1998 年- 東京都立保健科学大学教授。現在、脳機能イメージングによる認知神経科学の研究に従事。工学博士、医学博士、神経心理学会、脳波筋電図学会、音声言語医学会等の会員。



月本 洋(正会員)

1978 年東京大学工学部計数工学科卒業。1980 年同大学院修士課程修了。同年 (株) 東芝入社。現在、研究開発センターにて人工知能等の研究に従事。工学博士、電子情報通信学会、人工知能学会、情報処理学会、AAAI 等の会員。

線形回帰

自己回帰

回帰木

ノンパラメトリック回帰