非加法的測度による非線形積分の理論とその応用
—ゲーム理論, ビッグデータ解析, 機械学習など—

本田　あおい　(九工大情報工)*

1. 非加法的な測度

非加法的な測度とは古典的な測度では当然のように仮定される加法性条件を単調性条件に弱めたものである．このような概念は古くから数学の諸分野に現れ個別に研究されてきたが, 非加法性の重要性が認識されてからは, 主に二つの視点からより精力的に研究がなされるようになった．まず一つは応用的な視点である．工学や経済学における応用, 特に不確実性や曖昧さといった支配原理のわかっていない事柄は加法的な測度ではうまく表現できないものが多い．そのため, これらの分野では, 非加法的な測度(集合関数)を利用した応用研究が数多く行われてきた．相当する応用研究は古くからあちこちに散見するが, 非加法性を明確に意識した研究は1974年の菅野 [1] が端緒である．このあたりの研究が契機となって, 様々な分野や国で精力的な技術開発がなされた．もう一つの視点は理論的な視点である．知られている定理が実は加法性条件を外しても成り立つのか, 成り立つのであれば加法性条件をどこまで緩めることができるのか, 加法性条件を外すと成り立たないのであれば, ではどのくらい弱い定理が成り立つのか, 詳細に考察し測度論を再構築することで測度の本質を追求するという視点である．こちらも相当する研究は古くから存在するが, Choquet [2], あるいは Dobrakov [3] あたりからだんだんと非加法性が意識されている．理論研究は比較的, 進行が穏やかであり, 近年に至るまで応用研究の後塵を拝して来たが, しかしながら徐々に測度論の研究者らが研究に取り組み始めた．特に近年の河邊淳氏による積分収束定理の一連の結果など([4]等)をはじめとする, 特筆すべき成果が数多く出され, 研究者らの興味を喚起している．理論面の整備と, またそれを契機とした応用研究の発展という理想的な関係性がもたらす益々の発展が期待される．

さて, $1+1$ が 2 にならない, いわゆる非加法的な測度の世界では従来のルベーグ式の積分論を適用することができず新しい積分の理論が必要となる．ショケ積分や菅野積分をはじめとする色々な積分が提案されており, どれも線形性を持たず, またそれぞれの積分が長所短所合わせた特有の性質を持っている．包除積分は包除原理に基づく積分で非加法的な測度と多項演算の2つを用いて定義される．測度だけでなく演算も用いて相互作用を表現することができるため柔軟な高い記述力を持つ積分である．また, 非加法的な測度による積分としては比較的扱いやすく, 実問題への応用に適した積分である．本講演ではデータ解析への応用や周辺分野との関連性を示しながら, 包除積分の概略を紹介したい．

本講演の議論は全体集合を一般の集合に拡張できるが, 実問題への応用には有限集合で十分である．そのため, 本講演では全体集合は $\Omega=\{1,2,\ldots,n\}$, つまり n 点からなる有限集合とし, 積分に用いる演算も単純な2項演算などから導入できる多項演算 $\otimes$ とする (Cf. [5, 6]).

* 〒820-8502 福岡県飯塚市川津680-4 九州工業大学 情報工学研究院
e-mail: aoi@ces.kyutech.ac.jp

2. 準備

$\mathcal{P}(\Omega)$ は Ω の部分集合全体とし, $A \in \mathcal{P}(\Omega)$ に対して, $|A|$ は A の要素の個数とする.

定義 1 (単調測度 [7]) 集合関数 $\mu : \mathcal{P}(\Omega) \to [0, +\infty)$ が単調測度であるとは次の (F1), (F2) を満たすことである:

(F1) $\mu(\emptyset) = 0$.

(F2) 任意の $A, B \in \mathcal{P}(\Omega)$ に対して, $A \subset B$ ならば $\mu(A) \le \mu(B)$.

状況に応じて, 連続性等の条件を付加することもある. ファジィ測度 (fuzzy measure), 非加法的測度 (nonadditive measure), 容量 (capacity), submeasure などとも呼ばれ, これらは同一の概念である. ここでは単調性を仮定していることを明示する意味で単調測度と呼ぶことにする. 積分の定義に使用する演算を $\otimes : [0, K]^k \to [0, K], k = 1, 2, \ldots, n$, つまり $[0, K]$ 上の多項演算とする. 被積分関数 f は $[0, K]$ に値をとる Ω 上の関数とする. つまり $f = (x_1, x_2, \ldots, x_n) \in [0, K]^n$ である.

定義 2 (包除積分 [5])
$f = (x_1, x_2, \ldots, x_n) \in [0, K]^n$ の μ と $\otimes$ に関する包除積分は次で定義される:

$$\begin{aligned} \otimes\!\!\int f\, d\mu &:= \sum_{A \in \mathcal{P}(\Omega) \setminus \{\emptyset\}} M^{\otimes}(f \mid A)\mu(A) \\ &:= \sum_{A \in \mathcal{P}(\Omega) \setminus \{\emptyset\}} \left(\sum_{B \supset A} (-1)^{|B \setminus A|} \bigotimes_{i \in B} x_i \right) \mu(A), \end{aligned}$$

ただし $\bigotimes_{i \in \emptyset} x_i = K$.

包除積分の名前は, 包除原理の考え方を利用していることに由来している. $\otimes$ は [5, 6] 等ではより一般な関数とし, interaction operator と総称しているが, 本稿では 2 項演算などから導入できる単純なものだけを考えることとし, また簡単のため単に「演算」と呼ぶことにする.

単調測度による積分で代表的なものが次のショケ積分である.

定義 3 (ショケ積分)
$f = (x_1, x_2, \ldots, x_n) \in [0, K]^n$ の μ に関するショケ積分は次で定義される:

$$(C)\!\int f\, d\mu := \sum_{i=1}^{n} (x_{\sigma(i)} - x_{\sigma(i+1)})\mu(A_i),$$

ただし, σ は $x_{\sigma(i)} \ge x_{\sigma(i+1)}, i = 1, \ldots, n-1$ なる Ω 上の置換, $A_i := \{1, 2, \ldots, i\}, x_{\sigma(n+1)} := 0$.

命題 4 (i) μ が加法性を持つとき, 包除積分はルベーグ積分に一致する. つまり, 任意の $f \in [0, K]^n$, 任意の演算 $\otimes : [0, K]^k \to [0, K]$ に対して,

$$\otimes\!\!\int f d\mu = \int f d\mu := \sum_{i=1}^{n} x_i\ \mu(\{i\}).$$

(ii) 演算に minmum 演算 $\wedge$ を採用した場合, 包除積分はショケ積分に一致する. つまり任意の $f \in [0, K]^n$, 任意の単調測度 μ に対して,

$$\wedge\int f d\mu = (C)\int f d\mu.$$

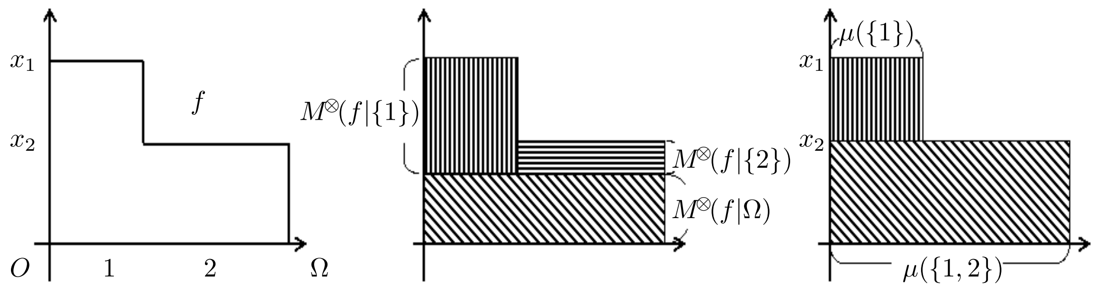
積分と呼ぶからには積分汎関数としての性質は満たしてほしい. そのためには $\otimes$ は多項演算ならどんなものでもよいわけではなく, 条件が必要である [6]. ここでいう積分汎関数とは

(M1) $\mathcal{F}(0) = 0$,
(M2) $f \leq g$ ならば $\mathcal{F}(f) \leq \mathcal{F}(g)$

を満たすもので, さらに χ_A を特性関数とし, $\chi_A^K := K\chi_A$ に対して,

(M3) $\mathcal{F}(\chi_A^K) = K\mu(A)$

も満たしているとなおよい. 命題 4 で示すようにショケ積分は演算に $\wedge$ を採用したもので, この積分は (M1), (M2), (M3) を満たす. この他, 代数積 $x \otimes_a y := xy/K$, や Dubois-Prade の t-norm — これは $x \otimes_{DP} y := xy/(\min\{x, y, \lambda\})$ で定義される $\lambda \in [0, K]$ をパラメータとするパラメトリック t-norm — などの主要な t-ノルムを用いた場合, 包除積分は (M1), (M2), (M3) を満たすことがわかっている. 大雑把にいうと包除積分で用いる演算 $\otimes$ は minimum 演算のような, 最大値 K が単位元となる, かけ算型の演算を想定している. (一番シンプルな演算の例は代数積であるので, $K = 1$ とし $\otimes$ は普通のかけ算とするとイメージが湧きやすい.)



(a) 被積分関数　(b) 包除積分　(c) ショケ積分

図 1: $\Omega = \{1, 2\}$ の場合の包除積分のイメージ

包除積分はメビウスの反転公式を用いて別表現が可能である.

定理 5 包除積分は次のようなメビウス型別表現で表すことができる.

$$\otimes\int f \ d\mu = \sum_{A \in \mathcal{P}(\Omega)\setminus\{\emptyset\}} \left(\bigotimes_{i \in A} x_i \right) \ m^{\mu}(A)$$

ただし m^{μ} は μ のメビウス変換:

$$m^{\mu}(A) := \sum_{B \supset A} (-1)^{|B \setminus A|} \mu(B)$$

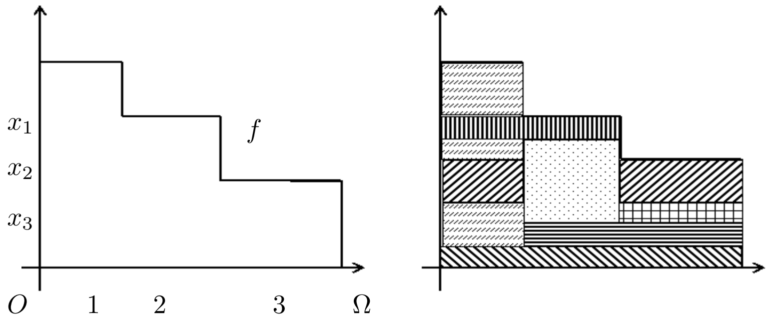


図 2: $\Omega = \{1, 2, 3\}$ の場合の包除積分のイメージ

であり, μ と m^μ は1対1に対応する. 実際

$$\mu(A) = \sum_{B \subset A} m^\mu(B)$$

として逆変換できる.

$\Omega = \{1, 2, 3\}$ のときの包除積分の定義式を書き下すと,

$$\begin{aligned}
\otimes\!\!\int f\ d\mu := &(x_1 - x_1 \otimes x_2 - x_1 \otimes x_3 + x_1 \otimes x_2 \otimes x_3)\ \mu(\{1\}) \\
&+(x_2 - x_1 \otimes x_2 - x_2 \otimes x_3 + x_1 \otimes x_2 \otimes x_3)\ \mu(\{2\}) \\
&+(x_3 - x_1 \otimes x_3 - x_2 \otimes x_3 + x_1 \otimes x_2 \otimes x_3)\ \mu(\{3\}) \\
&+(x_1 \otimes x_2 - x_1 \otimes x_2 \otimes x_3)\ \mu(\{1, 2\}) + (x_1 \otimes x_3 - x_1 \otimes x_2 \otimes x_3)\ \mu(\{1, 3\}) \\
&+(x_2 \otimes x_3 - x_1 \otimes x_2 \otimes x_3)\ \mu(\{2, 3\}) + (x_1 \otimes x_2 \otimes x_3)\ \mu(\{1, 2, 3\}).
\end{aligned}$$

他方, 定理 5 のメビウス型別表現を書き下すと,

$$\begin{aligned}
\otimes\!\!\int f\ d\mu \ \ := \ \ & x_1 m^\mu(\{1\}) + x_2 m^\mu(\{2\}) + x_3 m^\mu(\{3\}) \\
&+(x_1 \otimes x_2) m^\mu(\{1, 2\}) + (x_1 \otimes x_3) m^\mu(\{1, 3\}) + (x_2 \otimes x_3) m^\mu(\{2, 3\}) \\
&+(x_1 \otimes x_2 \otimes x_3) m^\mu(\{1, 2, 3\}).
\end{aligned}$$

メビウス型別表現が定義式に比べて若干すっきりとして見えるのは足したり引いたりの部分がメビウス変換に含まれているためであり, メビウス変換の部分を書き下すと定義式と同程度に煩雑な式となる. 別表現を定義として採用しなかったのは, 定義式の方が測度 μ が直接現れているし, 分割型の積分となっている計算の意味を明示できているからである. とはいえ測度はメビウス変換のままで十分なことも多く, また別表現の方が積分の計算が簡単である他, 様々な利点がある. 言ってみれば, これらの2つの式は統計における分散の定義と分散公式のような関係である.

3. 周辺分野との関連

本節では現在わかっている包除積分と周辺分野の関連性について紹介したい.

3.1. ゲーム理論との関連

積分を被積分関数 $f(x)$ と x 軸で囲まれた領域の面積と捉える場合には, 積分はこの領域を分割した各部分の面積の和の計算であり, 包除積分では前節の図 1, 2 で示すように各分割領域の $M^{\otimes}(f \mid A)$ は縦の長さ, $\mu(A)$ は横の長さとなるので $M^{\otimes}(f \mid A) \geq 0$ と制限するのが自然に思える. その場合, 必要条件の

$$\bigotimes_{i\in A} x_i \leq \min\left\{\bigotimes_{i\in B} x_i \,\middle|\, B \subset A\right\}$$

くらいは $\otimes$ として最低限仮定するのがよいように見える. これまでこの仮定を置いていた. しかし後に示す定理7のような例があることがわかり, 必ずしも正の分割領域に限る必要はないと思うようになった. 定理7 は包除積分の演算に算術平均を用いると, ゲーム理論における重要な解概念, シャープレイ値が現れるというものである.

定義 6 (シャープレイ値) 単調測度 μ に対してシャープレイ値 $\Phi(\mu) := (\phi_1(\mu), \phi_2(\mu), \ldots, \phi_n(\mu))$ は次で定義される:

$$\phi_i(\mu) := \sum_{\substack{A\in\mathcal{P}(\Omega),\\ A\not\ni i}} \frac{|A|!\,(n-|A|-1)!}{n!}\left(\mu(A\cup\{i\}) - \mu(A)\right).$$

定理 7 $\odot$ が算術平均演算:

$$\bigodot_{i\in A} x_i := \frac{1}{|A|}\sum_{i\in A} x_i$$

のとき,

$$\odot\!\int f d\mu = \int f d\phi_i(\mu) \left(= \sum_{i\in\Omega} f(i)\phi_i(\mu)\right),$$

ただし第 2 式はルベーグ積分である. さらに $f_j^1 = (x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n) := (0, \ldots, 0, 1, 0, \ldots, 0)$ とすると

$$\odot\!\int f_j^1 d\mu = \phi_j(\mu).$$

シャープレイ値は協力によって得られた利得を各プレイヤーに公平に配分する方法(協力ゲームの解とよばれる)の1つであり, Ω はプレイヤーの集合, $\mu(A)$ は提携 A で得られる利得, $\phi_i(\mu)$ は各プレイヤー i への利得の分配を表す. シャープレイ値の他にも解概念は提案されているが, これらの中においてシャープレイ値は極めて自然な公理により特徴づけされる合理的な解概念であることが知られている. 演算を算術平均として算出された包除積分と一致することからもシャープレイ値が平均的な, つまり公平性のある解であることがわかるし, 逆に算術平均でシャプレイ値が現れる包除積分は合理的な積算概念の1つであるともいえる. このとき包除積分は分割領域の面積の和ではなくなっている. 実際, $\Omega = \{1, 2\}$ で被積分関数が $f = (x_1, x_2), x_1 > x_2$ のとき積分は

$$\begin{aligned}
&\odot\!\int (x_1, x_2) d\mu \\
&\quad = \frac{x_1+x_2}{2}\mu(\{1,2\}) + \left(x_1 - \frac{x_1+x_2}{2}\right)\mu(\{1\}) + \left(x_2 - \frac{x_1+x_2}{2}\right)\mu(\{2\})
\end{aligned}$$

となり図3のように和だけでなく差も用いた分割領域の面積の計算となる. 負となる第3項はちょうど第1項で足しすぎたところを差し引く計算となっており, やはり関数と横軸に囲まれる領域の面積を計算していることには変わりない. $\odot\int f d\mu$ も (M1), (M2), (M3) を満たし, つまりかけ算型でない演算を用いても意味のある積算概念となり, さらには, 定理 7 が示すように線形な積分となる. この後はこのような方向からも包除積分に用いる演算の条件を考察することができると思う.

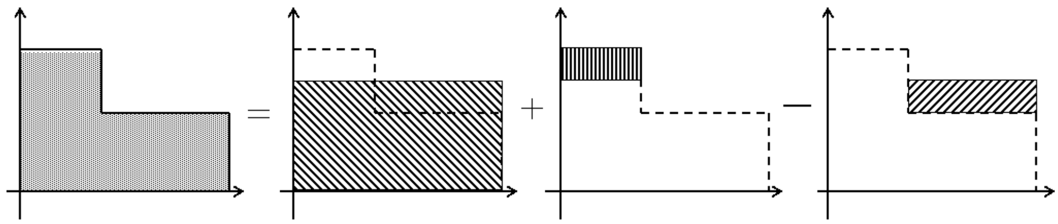


図 3: $\Omega = \{1, 2\}$ で算術平均を用いた場合の包除積分のイメージ

3.2. 離散凸解析との関連

離散的な関数 $f : \{0,1\}^n \to \mathbb{R}$ に対して $F : [0,1]^n \to \mathbb{R}$ が $F(\boldsymbol{x}) = f(\boldsymbol{x}), \boldsymbol{x} \in \{0,1\}^n$ を満たすとき, F を f の拡張と呼ぶ. ここで, $f(\boldsymbol{x}) = \mu(A_{\boldsymbol{x}}), \boldsymbol{x} \in \{0,1\}^n, A_{\boldsymbol{x}} := \{i \in \Omega \mid \boldsymbol{x}_i = 1\}$ として f を単調測度 μ と対応させせると, $F(\boldsymbol{x}), \boldsymbol{x} \in [0,1]^n$ は $\boldsymbol{x}$ の μ による積分に対応できる. というのも拡張であることの条件はちょうど (M3) に対応している. 次のロヴァース拡張は, これはショケ積分 (演算が $\wedge$ の包除積分) に相当する.

定義 8 (ロヴァース拡張) 関数 $f : \{0,1\}^n \to \mathbb{R}$ に対して, 次で定義される $F^L : [0,1]^n \to \mathbb{R}$ をロヴァース拡張と呼ぶ.

$$F^L(\boldsymbol{x}) := \sum_{i=0}^{n} \lambda_i f(S_i),$$

ただし, $\emptyset = S_0 \subset S_1 \subset S_2 \subset \cdots \subset S_n, \sum_{i=0}^{n} \lambda_i \chi_{S_i} = \boldsymbol{x}, \sum_{i=0}^{n} \lambda_i = 1, \lambda_i \geq 0.$

他にも multilinear extension とよばれる拡張は演算を代数積としたときの包除積分に相当する.

定義 9 (multilinear extension) 関数 $f : \{0,1\}^n \to \mathbb{R}$ に対して, 次で定義される $F^M : [0,1]^n \to \mathbb{R}$ を Multilinear extension と呼ぶ.

$$F^M(\boldsymbol{x}) := \sum_{A \in \mathcal{P}(\Omega)} \prod_{i \in A} x_i \prod_{i \notin A} (1 - x_i).$$

これは, 代数積を用いた場合に包除積分が次のように因数分解できることによるものである.

$$\begin{aligned}
\otimes_a \int f d\mu &= \sum_{A \in \mathcal{P}(\Omega)\setminus\{\emptyset\}} \left(\sum_{B \supset A} (-1)^{|B \setminus A|} \bigotimes_{i \in B}{}_a \, x_i \right) \mu(A) \\
&= \sum_{A \in \mathcal{P}(\Omega)\setminus\{\emptyset\}} \left(\bigotimes_{i \in A}{}_a \, x_i \right) \otimes_a \left(\bigotimes_{i \notin A}{}_a \, (K - x_i) \right) \mu(A).
\end{aligned}$$

$K=1$ の場合 $\otimes_a$ は普通のかけ算となるので上式は

$$\otimes_a \int f d\mu \;\; = \sum_{A\in\mathcal{P}(\Omega)\setminus\{\emptyset\}} \prod_{i\in A} x_i \prod_{i\notin A}(1-x_i)\mu(A).$$

となり上記の multilinear extension と同じ形が得られる．まさに包除 (包含と除外) の式になっているが, 他の演算では分配律が成り立たず, このよう表現はできない．

4. 包除積分数理モデルを用いた多変量解析

4.1. 重回帰分析の拡張としての包除積分数理モデル

1つ以上の説明変数の組から目的となる変数値を推定する, いわゆる多変量データ解析問題において包除積分を利用することができる．目的変数を y, 説明変数を $x_1,\ldots,x_n$ とし, $y \sim F(x_1,\ldots,x_n)$ なるモデル式 F を見出すのがこの問題の目的である．古典的に, また現在でもよく使われているのが重回帰分析で, モデル式は説明変数の線形結合:

$$F(x_1,\ldots,x_n) := a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_0,$$

であり, 学習データを用いて最小二乗法で最も当てはまりのよいパラメータの値 $a_1,\ldots,$ a_n, a_0 を算出することでモデル式が決定される．モデル式が決定されると, 目的変数のわからない説明変数のデータの組から, 目的変数を推定することができる．包除積分モデルは包除積分のメビウス型別表現を数理モデルとして用いたもので, モデル式は

$$\begin{aligned} F(x_1,\ldots,x_n) = \;& a_{\{1\}}x_1 + \cdots + a_{\{n\}}x_n \\ & + a_{\{1,2\}}(x_1 \otimes x_2) + \cdots + a_{\{n-1,n\}}(x_{n-1}\otimes x_n) \\ & + \cdots \\ & + a_{\{1,2,\ldots,n\}}(x_1 \otimes x_2 \otimes \cdots \otimes x_n) + a_\emptyset \end{aligned}$$

である．重回帰モデルが n 項と切片項から成るのに対して, 包除積分は重回帰モデルの n 項の後ろに $2^n - n - 1$ 項の相互作用項 を加えたものとみなすことができ, その意味で重回帰モデルの拡張となっている．項数は切片を含めて 2^n 項となる．演算により説明変数を組み合わせて新しい説明変数を作成しているとみなすと重回帰モデルと同じ説明変数の線形和であり, 解析には重回帰モデルの多くの手法を踏襲できる．各パラメータ $a_A, A \in \mathcal{P}(\Omega)$ はメビウス変換 $m^\mu(A)$ に対応する．解析手順は以下の通り．

ステップ 1 パラメータの決定に使用する学習データの前処理を行う．
スケーリング処理の他に, 積分に用いる演算を決定し, 相互作用項のデータを作成する必要がある．

ステップ 2 学習データを用いてパラメータを決定する．
重回帰解析で提案されている色々な手法を利用できる．

ステップ 3 得られた数理モデルを使って, 推定や解釈を行う．
目的変数の値のわからない未知データに対して得られたモデル式に代入して目的変数を推定する他, パラメータから得られる単調測度や演算から, 得られたモデル式を解釈することができる．

ステップ1では, 重回帰モデルの場合と違って, データの前処理により, 説明変数のスケールと大小の方向を揃える必要がある. これは, 説明変数は包除積分の被積分関数に相当するので, 積分の単調性より, 説明変数が大きいほど目的変数が大きくなるように大小の方向が揃っている必要があるためである. また, 説明変数の $\wedge$ などの演算をとることからも説明変数のスケールは揃っている必要がある. 具体的には, 説明変数と目的変数の相関が負であれば大小を逆転し, 最大値 K, 最小値 0 になるように線形あるいは非線形な非減少関数で変換を施せばよい.

ステップ2 は重回帰モデルの場合と同様にしてパラメータを算出することができる. Excel 等の統計ソフトウェアには標準的に重回帰分析ツールが準備されているのでいるので, これを使用することでプログラムや統計の知識がなくても簡単に最小二乗法によるパラメータ決定が可能である [8]. ただし, Excel の場合は説明変数 (包除積分モデルの場合は相互作用項も加えた数) が 16 個までに制限されているので, k-加法的測度などとして制限をもうける必要がある. k-加法的測度とは, $|A| > k$ なる $A \in \mathcal{P}(\Omega)$ に対して $m^{\mu}(A) = 0$ を仮定し若干の加法性条件— k 個以上の要素が集まったことによる相乗・相殺効果はない — を付加した測度である. 例えば, 説明変数が 5 つ, つまり Ω が 5 点集合の場合, 2-加法的測度に制限するとモデル式は次のようになり. 項数を 32 から 16 に減らすことができる:

$$F(x_1, \ldots, x_n) := a_{\{1\}}x_1 + \cdots + a_{\{5\}}x_n + a_{\{1,2\}}(x_1 \otimes x_2) + \cdots + a_{\{4,5\}}(x_4 \otimes x_5) + a_0$$

他の高度なツールを使用する場合にはステップワイズ法等の重回帰分析の変数選択手法を用いて項を取捨選択して項数を減らすこともできる. これらの手法により得られたパラメータが単調測度となる保証はないが, 単調でない場合も包除積分数理モデルといってよいし, 単調性を仮定したい場合には, 何らかの手法で単調性を満たすようにパラメータを決定すればよい.

ステップ3では, 推定に関しては重回帰分析と同じように得られたモデル式に説明変数の値を代入すればよい. 解釈に関しては, 得られたパラメータから, メビウス逆変換で単調測度やシャープレイ値を算出して, 各説明変数の重視度や相互作用を読み取ることができる. また, 積分に用いた演算は相互作用の現れる大きさを表している.

重回帰モデルでも, 説明変数の積を「交互作用項」として新たな変数として追加することは提案されており, 統計ソフトウェアではこの機能が組み込まれたものもある. 包除積分数理モデルはこの重回帰モデルの交互作用項の意味の 1 つの解釈ともいえる. さらに包除積分モデルの場合は演算はかけ算でなくてもよいため重回帰分析の交互作用項より, さらに表現の自由度が高い. 演算には様々な候補があり, Dubois-Prade の t-ノルムのようなパラメトリック t ノルムを用いると, より精密な学習データへのフィッティングが可能となる. このように重回帰モデルでは表現できない説明変数間の複雑な相互作用を表現することができるため, データの支配関係の精密な解析が可能となる. つまり, 現在解析方法が見つかっていないような複雑な構造を持つビッグデータ解析を解析するための有効な手法となることが期待できる. また, 多くの機械学習や AI を用いたデータ解析手法では得られる推論部がブラックボックスとなるのに対して, 包除積分モデルを用いた場合は単調測度と積分の演算を対応させることで得られた推論部の解釈をすることが可能であることも強みである.

4.2. 機械学習への応用

インターネットの発達や大容量ストレージの低価格化に伴い, 大規模データの取得と保存が容易になり, 近年データサイエンスの重要性がますます高まっている. IT 技術の進化で機械学習によるデータ解析手法が発展し, 従来の統計的手法では解析できないようなデータの解析が可能となり今後の発展が期待されている. 中でもディープラーニング (深層学習) はニューラルネットワークの中間層を多層化したもので, これが 2012 年に Google 社の研究チームが, 画像に猫が写っているかどうか膨大なデータを使って自動的に学習することに成功したことで注目を集め, これを契機にニューラルネットワークの再ブームが到来した. 機械学習では IT の腕力を最大限に

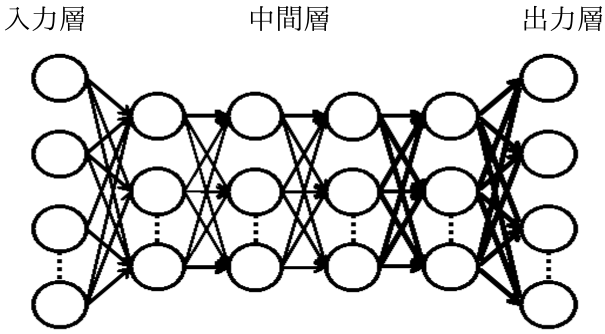


図 4: 深層学習ネットワークのイメージ図

利用して画像データのような高次元データ (説明変数の数が膨大なデータ) の解析を行い, 推定に関して多くの成果を上げている. その反面, 推論の過程, つまり得られたネットワークの意味が解釈できないという欠点である. ここに, 統計的手法を組み入れることで双方の長所を兼ね揃えたデータ解析が実現できないだろうか. 深層学習ネット

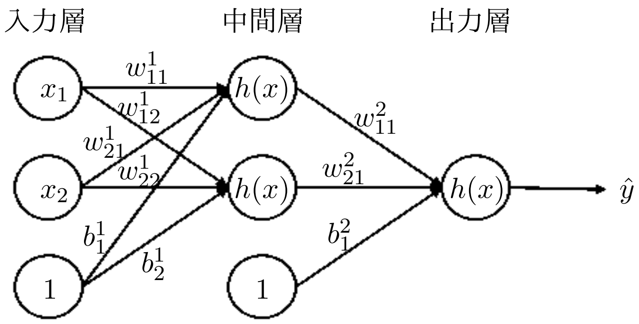


図 5: 簡単な深層学習ネットワーク

ワークは通常, 図 4 のような模式図で表現されるが, これは各出力ユニット毎の数式で表すことができる [9]. 例えば図 5 の深層学習ネットワーク (正確にはニューラルネットワーク. 中間層が 1 層なので深層ではない) を式で表すと

$$\hat{f} = F(x_1, x_2) = h(h(w_{11}^1 x_1 + w_{21}^1 x_2 + b_1^1) w_{11}^2 + h(w_{12}^1 x_1 + w_{22}^1 x_2 + b_2^1) w_{21}^2 + b_1^2)$$

である. $h(x)$ は従来のニューラルネットワークでは主にシグモイド関数:

$$s(x) := \frac{1}{1 + e^{-x}}$$

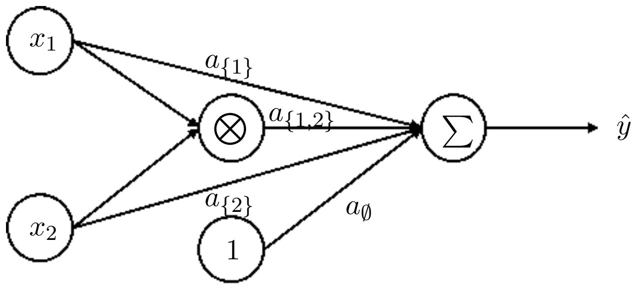


図 6: 簡単な包除積分ネットワーク

が使われていたが, 深層学習ネットワークではランプ関数:

$$r(x) := \max\{0, x\}$$

がよく使われる. ネットワークの深層化が効力を発揮したのはこのランプ関数の利用によるところが大きいようである. このようにネットワークは数式で表現できるので, 逆に解釈可能な数理モデルをネットワークで表現すれば, それは解釈可能なネットワークとなる. 例えば, $\Omega = \{1, 2\}$ とし, 演算を $\otimes$ としたときの包除積分数理モデルを深層学習ネットワークで表現すると図 6 である. ネットワークの構成に制限があるが, 解釈可能なネットワークの構築が可能となる. パラメータの決定などネットワークの構築は深層学習ネットワークの手法を踏襲し, 逆に現在深層学習で解決できていない解釈の部分に対しては包除積分を用いた解釈を行う. このよう IT 的な手法と数理的な手法による両面からのアプローチでビッグデータ解析の諸問題を解決することが今後の目標である.

参考文献

[1] M. Sugeno, Theory of fuzzy integrals and its applications, Ph.D. Dissertation, Tokyo Institute of Technology, Tokyo, (1974).

[2] G. Choquet, Theory of capacities, Ann. Inst. Fourier, **5** (1953), 131–295.

[3] I. Dobrakov, On submeasures I, Dissertationes Math. Rozprawy Mat.), **112** (1974), 1–35.

[4] 河邊淳, 非加法的測度と非線形積分, 数学, **68** (2016), 266-292.

[5] A. Honda, Y. Okazaki, Theory of inclusion-exclusion integral, Information Science, **376** (2017), 136-147.

[6] A. Honda, Y. Okazaki, Generalization of inclusion-exclusion integral for nondiscrete monotone measure space, Fuzzy Sets and Systems, **355** (2019), 42-58.

[7] ファジィ測度, 日本ファジィ学会編, 日刊工業新聞社, (1993).

[8] 本田あおい, 包除積分数理モデルを用いた多変量データ解析, 知能と情報 (日本知能情報ファジィ学会誌), **30** (2018) 183-192.

[9] 深層学習 (機械学習プロフェッショナルシリーズ), 岡谷貴之, 講談社, 2015.