

## スペクトル理論を用いた学習アルゴリズムについて

園原ひかる 月本洋

{ hikaru , tukimoto } @ssel.toshiba.co.jp

(株)東芝 研究開発センター システム・ソフトウェア生産技術研究所

〒210 川崎市幸区柳町 70

従来、数値データからの学習や予測をする方法として、線形回帰分析などが用いられてきた。我々は、少ない学習事例で精度良く予測できる非線形回帰分析を検討してきた。ブール関数の研究において、多重線形関数のフーリエ解析がスペクトル理論と呼ばれて研究されている。この多重線形関数の定義域は  $\{0, 1\}$  で値域は実数である。Linial はそのスペクトル理論に基づく学習アルゴリズムを提案した。我々はそのスペクトル理論の定義域を  $[0, 1]$  に拡張し、それを用いた学習アルゴリズムを考案した。本論文では、この学習アルゴリズムを Linial のアルゴリズムと比較実験し、この学習アルゴリズムが有効であることを確認した。

## A LEARNING ALGORITHM USING SPECTRUM THEORY

Hikaru Sonohara and Hiroshi Tsukimoto

Systems & Software Engineering Laboratory , Research & Development Center, Toshiba Co.  
70 Yanagi-cho, Saiwai-ku, Kawasaki, Kanagawa, 210, Japan

As learning and prediction methods from numerical data, for example, linear regression analysis has been used. We have been studying non-linear regression analysis algorithms. In 1993, Linial presented a learning algorithm based on spectrum theory (Fourier analysis for Boolean functions). In this paper, we extend the spectrum theory and the learning algorithm to continuous domains in order to deal with numerical data. We experimentally confirm that the extended algorithm works well compared with Linial's algorithm.

## 1 はじめに

フーリエ解析は古くは古典数学で、また最近では信号処理をはじめとする多くの理工学分野において用いられている。フーリエ解析とは、関数  $f$  を三角関数を用いて次のように級数展開して解析することである。

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos \frac{n\pi x}{l} + b_n \sin \frac{n\pi x}{l})$$

但し、 $a_0, a_n, b_n, l$  は実数で、 $n$  は自然数である。このフーリエ解析を一般化したものもまたフーリエ解析といい、ある関数空間の任意の関数  $f$  をこの関数空間の正規直交基底の線形結合で表現する。

ブール関数を拡張した多重線形関数  $\{0, 1\}^n \rightarrow R$  においてもこのようなフーリエ解析が可能である。この多重線形関数空間の変数集合の部分集合を  $S$ 、正規直交基底  $\chi_S$ 、係数を  $\alpha_S$  を用いて、任意の関数  $f$  は次のような線形結合で表される。

$$f(x) = \sum_S \alpha_S \chi_S$$

Linial 等 [1],[2] はブール関数の高次数のフーリエ係数は指数関数的に減少することを証明し、この性質を利用して低次数のフーリエ係数のみで学習、予測する予測アルゴリズムを考案した。つまり予測された値を  $\tilde{f}$  とすると、

$$\tilde{f} = \sum_{|S| \leq t} \alpha_S \chi_S$$

但し  $|S|$  は集合  $S$  の元の個数を表す。

また最近ではブール関数  $\{0, 1\}^n \rightarrow \{0, 1\}$  だけではなく、定義域及び値域として  $[0, 1]$  をとる連続ブール関数  $[0, 1]^n \rightarrow [0, 1]$  についても研究がなされている [3]。ブール関数の定義域、値域を  $[0, 1]$  に拡張したものは、以下のように論理演算を定義すれば、ブール代数の全公理を満たすので、このような関数を連続ブール関数と呼ぶ。

$$\begin{aligned} x \wedge y &\rightarrow \tau(xy) \\ x \vee y &\rightarrow \tau(x+y-xy) \\ \bar{x} &\rightarrow \tau(1-x) \end{aligned}$$

但し  $\tau$  は  $\tau(z^2) = z$  を満たすような作用素である。

我々はスペクトル理論を多重線形関数  $[0, 1]^n \rightarrow R$  に拡張し [4]、それを用いた学習アルゴリズムを考案した。またこのアルゴリズムの有効性を検証するために、定義域が  $\{0, 1\}$  の場合との予測精度の比較実験を行なった。この実験を通して定義域

が  $[0, 1]$  の場合の学習及び予測における特徴と有効性を調べることができた。

本論文の構成は、2節ではスペクトル理論を紹介し、3節では定義域  $[0, 1]$  にスペクトル理論を拡張する方法を述べる。4節では定義域  $\{0, 1\}$  と  $[0, 1]$  に共通の学習アルゴリズムと予測アルゴリズムを述べる。5節では学習アルゴリズムの有効性の実験と考察を述べる。

## 2 スペクトル理論

$n$  変数ブール関数は  $f : \{0, 1\}^n \rightarrow \{-1, 1\}$  とみなすことができ、 $n$  変数ブール関数の全体は多重線形関数  $\{0, 1\}^n \rightarrow R$  の空間に埋め込むことができる。多重線形関数の集合は次のように内積を定義すると  $2^n$  次元ユークリッド空間を成す。

$$\langle g, f \rangle = 2^{-n} \sum_{x \in \{0, 1\}^n} f(x)g(x)$$

またノルムは通常のように次で定義する。

$$\|f\| = \sqrt{\langle f, f \rangle}$$

$\{1, \dots, n\}$  の部分集合  $S$  に対して次のように関数  $\chi_S$  を定義する。

$$\chi_S(x_1, \dots, x_n) = \begin{cases} +1 & \sum_{i \in S} x_i \text{ が偶数} \\ -1 & \sum_{i \in S} x_i \text{ が奇数} \end{cases}$$

以上のように定義した  $\chi_S$  はこの関数空間の正規直交基底になっている。多重線形関数  $f$  は一意的に  $\chi_S$  の線形結合で表せる。

$$f = \sum_S C_S \chi_S$$

特に定数  $C_S$  を  $\hat{f}(S)$  と書く。これはフーリエ級数展開である。以下 Linial 等が示した補題を述べる。

### Linial の補題

$f$  を深さ  $d$  で、大きさ  $M$  のブール回路で計算されるブール関数、 $t$  は任意の整数とする。このとき、

$$\sum_{S \in \{1, \dots, n\}, |S| > t} \hat{f}(S)^2 \leq 2M2^{-\frac{t-d}{20}}$$

但し  $|S|$  は集合  $S$  の元の個数を表している。

この補題は、ある次数以上の多重線形関数のフーリエ係数を自乗して加えたものは、次数に関して指数関数的に減少することを述べている。但し次数とは  $|S|$  の値である。

### 3 スペクトル理論の定義域 $[0, 1]$ への拡張

#### 3.1 連続ブール関数

ブール関数の定義域, 値域を  $[0, 1]$  に拡張した関数は, 以下のように論理演算を定義すれば, ブール代数の全公理を満たすので, このような関数を連続ブール関数と呼ぶ.

$$\begin{aligned}x \wedge y &\rightarrow \tau(xy) \\x \vee y &\rightarrow \tau(x + y - xy) \\ \bar{x} &\rightarrow \tau(1 - x)\end{aligned}$$

但し  $\tau$  は  $\tau(z^2) = z$  を満たすような作用素である.

$n$  変数連続ブール関数は  $f: [0, 1]^n \rightarrow [0, 1]$  となる. この連続ブール関数  $f$  を  $2f - 1$  と変換すると, 連続ブール関数は  $[0, 1]^n \rightarrow [-1, 1]$  となる. このような連続ブール関数全体は多重線形関数  $[0, 1] \rightarrow R$  の空間に埋め込むことができる.

#### 3.2 スペクトル理論の多重線形関数空間への拡張

多重線形関数空間は次のように定義される内積によって  $2^n$  次元ユークリッド空間を成す.

$$\langle f, g \rangle = \int_0^1 \tau(fg) dx$$

但し,  $\tau$  は  $\tau(x^2) = x$  を満たすような作用素である. またノルムは通常のように次で定義する.

$$\|f\| = \sqrt{\langle f, f \rangle}$$

この関数空間に次のように基底  $\chi_S$  を定義すると, この基底は正規直交基底になっている [3].

$$\chi_S = \prod_{i \in S} (1 - 2x_i)$$

例えば二変数  $\{x, y\}$  の場合, 基底  $\chi_S$  は以下の四つになる.

$$\begin{aligned}\chi_{\{\emptyset\}} &= 1 \\ \chi_{\{x\}} &= (1 - 2x) \\ \chi_{\{y\}} &= (1 - 2y) \\ \chi_{\{x, y\}} &= (1 - 2x)(1 - 2y)\end{aligned}$$

実はこの基底は定義域  $\{0, 1\}$  の多重線形関数空間においては, 先の節で定義した基底と一致する. この関数空間の任意の関数は一意的に  $\chi_S$  の線形結合で表せる. これはフーリエ級数展開である.

$$f = \sum_S \hat{f}(S) \chi_S$$

$$\text{但し } \hat{f}(S) = \langle f, \chi_S \rangle$$

このとき次のような予想を立てた.

予想

$f$  はゲートの総数  $M$ , 深さ  $d$  の回路で計算されるブール関数とする. また  $t$  は整数とする. このとき次が成り立つ.

$$\sum_{|S| > t} \hat{f}(S)^2 \leq O(M2^{-t})$$

但し  $|S|$  は集合  $S$  の元の個数を表す.

この予想は, ある次数以上のフーリエ係数を自乗して加えたものは次数に関して指数関数的に減少することを述べている. 本論文はこの予想を実験的に検証した.

## 4 学習アルゴリズム

### 4.1 Linial 等の学習アルゴリズム

以下, Linial 等が [1] で提案したスペクトル理論を用いた学習アルゴリズムについて述べる. この学習アルゴリズムに用いる事例  $x$  は, 一様確率分布  $U$  に従って選ばれるものとする. [1] の目的は以下に述べる  $(\epsilon, \delta, U)$  予測アルゴリズムを得ることにある.

スペクトル理論を用いた学習アルゴリズムは, 学習部分と予測部分の二つより成る. 学習部分には属性とそのクラスを 1 組とした事例が与えられ, 学習関数 (スペクトル理論ではフーリエ係数) を求める. 一方予測部分には属性の組だけの事例が与えられ, その属性からクラスを予測する.

#### $(\epsilon, \delta, D)$ 予測アルゴリズム

ある確率分布に従って選ばれたいくつか学習用事例の属性  $x$  とそのクラス  $f(x)$  について学習アルゴリズムを用いて学習関数を求める. その学習関数を用いて確率分布に従って選ばれた予測用事例の属性の組  $x$  から予測したクラスを  $\tilde{f}(x)$  とする. このとき次の関係式が成り立つならば,  $(\epsilon, \delta, D)$  予測アルゴリズムという.

$$Pr[\tilde{f} \text{ は入力 } x \text{ の } \epsilon \text{ 部分に対して } f \text{ と不一致}] \leq \delta$$

但し  $D$  は属性  $x$  を選ぶ確率分布である.

学習部分

一様確率分布に従って選ばれた  $m$  個の事例から次のようにブール関数  $f$  の  $S$  次フーリエ係数  $\alpha_S$  を学習する。  $|S| \leq K$  のとき

$$\alpha_S = \frac{\sum_{i=1}^m f(x_i) \chi_S(x_i)}{m}$$

また  $|S| \geq K$  のときは、 $\alpha_S = 0$  とする。

### 予測部分

学習部分によって得られたフーリエ係数を用いて、新しい事例からそのクラスを予測する。そのときの予測部分は、予測クラスを  $\tilde{f}(x)$ 、 $K$  を整数とし、以下のようにする。

$$\tilde{f}(x) = \sum_{|S| \leq K} \alpha_S \chi_S(x)$$

このように学習部分、予測部分を定義したとき、[1] で次のような定理が証明されている。

### Linial の定理

深さ  $d$ 、サイズ  $M$  の回路を計算するブール関数に対する上のアルゴリズムは、 $m = 4 \left( \frac{2n^K}{\epsilon} \right) \ln \left( \frac{2n^K}{\delta} \right)$ 、 $K = (20 \log \left( \frac{2m}{\epsilon} \right))^d$  とすると、 $(\epsilon, \delta, U)$  予測アルゴリズムになる。

### 4.2 定義域 $[0, 1]$ の学習アルゴリズム

定義域が  $[0, 1]$  の場合の学習アルゴリズムも、上で定義域  $\{0, 1\}$  の場合に定義したものと同じ学習部分と予測部分からなる。以下、この学習アルゴリズムについての予想を述べる。

### 予想

深さ  $d$ 、サイズ  $M$  の回路に対する学習アルゴリズムは、 $m$  及び  $K$  を適当に選べば、 $(\epsilon, \delta, U)$  予測アルゴリズムになる。

本論文はこの予想を実験的に検証した。

## 5 実験と考察

Linial 等が提案した定義域  $\{0, 1\}$  の多重線形関数に対するスペクトル理論を用いた学習アルゴリズムと、我々が考案した定義域  $[0, 1]$  の多重線形関数に拡張したスペクトル理論を用いた学習アルゴリズムの比較実験を行なった。定義域  $[0, 1]$  の学習アルゴリズムは定義域が  $\{0, 1\}$  の拡張であるから、実験の目的は学習事例として数値データを与えたときは離散値データを与えたときに比べてどのような特徴があるのかを検証することである。なお、実験で用いた論理式は以下のものである。

### 論理式

$$(a \wedge b) \vee (c \wedge d) \vee (e \wedge f) \vee (g \wedge h) \\ (a \wedge b \wedge c \wedge d) \vee (e \wedge f \wedge g \wedge h)$$

### 5.1 学習したフーリエ係数に関する実験

学習したフーリエ係数が次数に関してどのように減少しているか定義域  $\{0, 1\}$  の場合と定義域  $[0, 1]$  の場合の比較実験を行なった。実験の条件は以下のものである。

#### 学習した事例

定義域  $\{0, 1\}$  の場合は全 256 事例

定義域  $[0, 1]$  の場合は一様確率分布に従って発生させた 256 事例

図 1 及び図 2 は実験結果のグラフである。グラフの横軸  $K$  はフーリエ係数  $\alpha_S$  の次数である。また縦軸  $Y$  はある次数以上のフーリエ係数について各々の次数のフーリエ係数を自乗して加えた以下の値である。我々はこの値をスペクトルと呼ぶことにする。(一般に言うところのスペクトルとは異なる)

$$y = \left\{ \sum_{|S| > K} \alpha_S^2 \right\} \div 256$$

但し  $|S|$  は集合  $S$  の元の個数を表している。また実線は定義域が  $[0, 1]$  の場合、点線は定義域が  $\{0, 1\}$  の場合を表している。

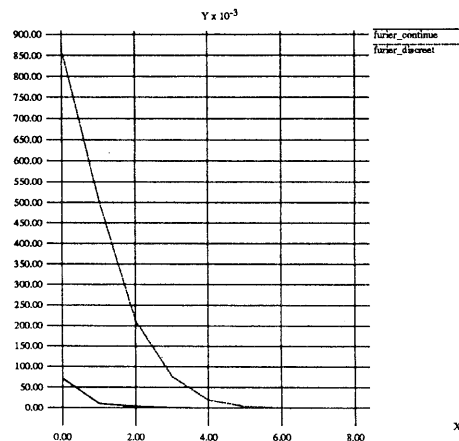


図1  $(a \wedge b) \vee (c \wedge d) \vee (e \wedge f) \vee (g \wedge h)$  の場合

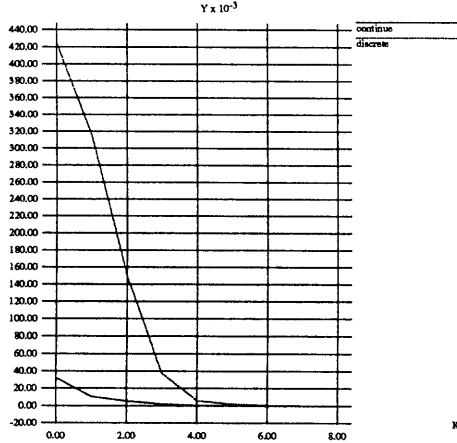


図2  $(a \wedge b \wedge c \wedge d) \vee (e \wedge f \wedge g \wedge h)$  の場合

定義域  $[0, 1]$  の場合は、定義域  $\{0, 1\}$  の場合よりも低い次数でスペクトルは 0 になる。これは学習事例が定義域  $[0, 1]$  の場合は  $[0, 1]$  から一様に選ばれた値が与えられるのに対して、定義域  $\{0, 1\}$  の場合は分散した値が与えられることに因ると思われる。この実験結果より、定義域  $[0, 1]$  の場合の低次数のフーリエ係数のみを用いた予測は、定義域  $\{0, 1\}$  の場合よりも精度が良いであろうと予想できる。

## 5.2 精度に関する実験

定義域  $\{0, 1\}$  の場合と定義域  $[0, 1]$  の場合の精度について比較実験を行なった。学習する事例と予測する事例が同じ場合と異なる場合の二種類の実験を行なった。

まず定義域  $\{0, 1\}$  の場合と定義域  $[0, 1]$  の場合において、学習した事例を予測する実験を行なった。定義域  $\{0, 1\}$  の場合の事例は 256 事例しかなく、256 事例を全て学習したフーリエ係数を用いれば正確な予測をすることができる。一方、定義域  $[0, 1]$  の場合の事例は無限にあり、全てを学習することは不可能である。従ってなんらかの方法で選んだ事例を用いてフーリエ係数を学習させることになる。我々は一様確率分布に従って発生させた 256 事例を用いてフーリエ係数を学習させ、

その同じ事例を用いて予測した場合の予測精度について、定義域  $\{0, 1\}$  の場合と比較実験を行なった。以下実験の条件について述べる。

### 学習した事例

定義域  $\{0, 1\}$  の場合は全 256 事例

定義域  $[0, 1]$  の場合は一様確率分布に従って発生させた 256 事例

### 予測した事例

各々、学習した事例と同じ事例

図3及び図4は実験結果のグラフである。横軸  $X$  は予測のために用いた最高次数である。また縦軸  $Y$  は予測に用いた事例の実際のクラス  $f$  と予測に用いた最高次数が  $X$  のときの予測したクラス  $\tilde{f}(X)$  の誤差の自乗を実験回数 256 回で平均を取ったものである。つまり、次のような値である。

$$Y = \{f - \tilde{f}(X)\}^2 \div 256$$

また実線は定義域が  $[0, 1]$  の場合、点線は定義域が  $\{0, 1\}$  の場合を示している。

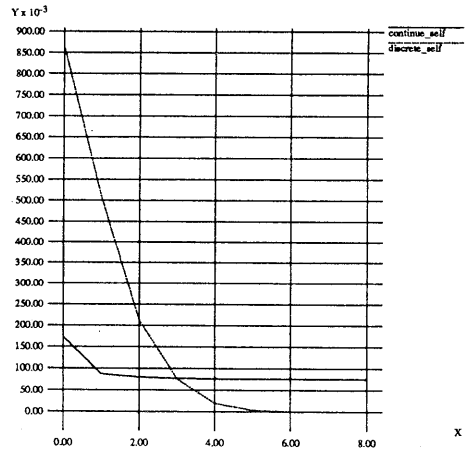


図3  $(a \wedge b) \vee (c \wedge d) \vee (e \wedge f) \vee (g \wedge h)$  の場合

率分布に従って発生させた 128 事例

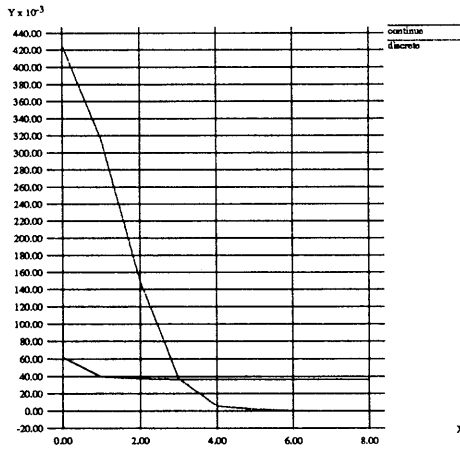


図 4  $(a \wedge b \wedge c \wedge d) \vee (e \wedge f \wedge g \wedge h)$  の場合

定義域  $\{0,1\}$  の場合は予測に用いる最高次数の関して指数関数的に予測精度は良くなっている。一方、定義域  $[0,1]$  の場合は先の実験(学習したフーリエ係数に関する実験)でスペクトルは次数 2 で急激に減少しそれ以上の次数では 0 であったので、予測に用いる最高次数が 2 以上ならば予測精度はほとんど変わらない。また予測に用いる最高次数が小さいときは、定義域が  $\{0,1\}$  の場合と比べて予測精度が良かった。これは、定義域  $[0,1]$  の場合が学習事例として、 $[0,1]$  に一様に分布した事例を用いていることと、学習したフーリエ係数が低い次数でほとんど 0 になることに因ると思われる。

上に述べた実験では、学習した事例と同じ事例で予測を行なったが、次は学習した事例とは異なる事例を与えて予測する実験を行なった。この実験は定義域  $\{0,1\}$  と定義域  $[0,1]$  の予測精度の比較実験である。以下実験の条件について述べる。

学習した事例

定義域  $\{0,1\}$  の場合は 128 事例  
 定義域  $[0,1]$  の場合は一様確率分布に従って発生させた 128 事例

予測した事例

定義域  $\{0,1\}$  の場合は学習事例以外の 128 事例  
 定義域  $[0,1]$  の場合は学習事例とは別に一様確

図 5, 図 6 は実験結果のグラフである。横軸及び縦軸は上述の実験と同様である。また実線は定義域が  $[0,1]$  の場合、点線は定義域が  $\{0,1\}$  の場合を示している。

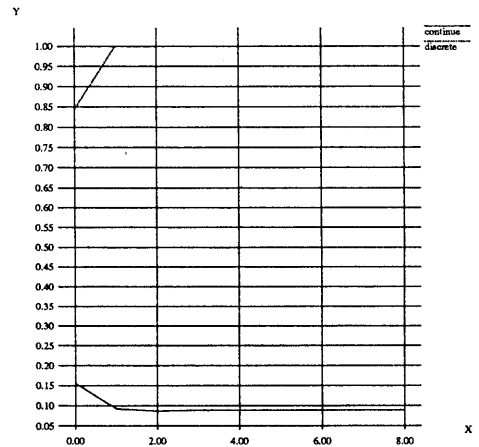


図 5  $(a \wedge b) \vee (c \wedge d) \vee (e \wedge f) \vee (g \wedge h)$  の場合

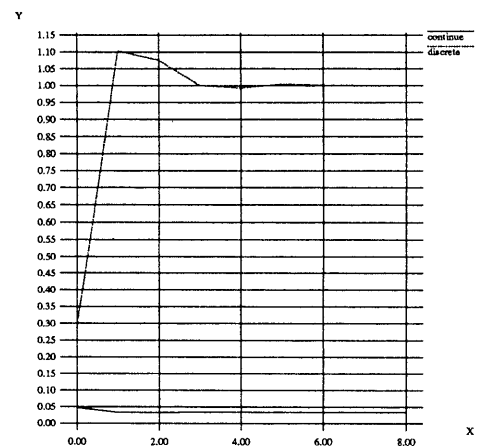


図 6  $(a \wedge b \wedge c \wedge d) \vee (e \wedge f \wedge g \wedge h)$  の場合  
 定義域  $\{0,1\}$  の場合、予測精度はかなり悪かつ

た。一方、定義域  $[0, 1]$  の場合は先の実験 (学習した事例と同じ事例を予測する) の結果とそれほど変わらない精度であった。定義域が  $[0, 1]$  の場合は学習のための事例が一様に与えられるので、事例が分散して与えられる定義域  $\{0, 1\}$  の場合よりも少ない事例で学習効果があることわかる。

### 5.3 学習用事例数と精度に関する実験

学習用事例の数と予測精度の間どのような関係があるか定義域が  $\{0, 1\}$  と定義域が  $[0, 1]$  の場合の比較実験を行なった。以下実験の条件を述べる。

学習した事例

定義域が  $\{0, 1\}$  の場合は 128 事例

定義域が  $[0, 1]$  の場合は一様確率分布に従って発生させた 25, 50, 100, 75, 128 事例

予測した事例

各々, 学習事例とは異なる事例

図 7, 図 8, 図 9, 図 10 は実験結果のグラフである。横軸は学習に用いた事例数  $X$  である。また縦軸  $Y$  は予測に用いた事例の実際のクラス  $f$  と予測したクラスの誤差  $\tilde{f}(X, d)$  の自乗を実験回数 200 で平均を取ったものである。更に 0th, 1st, ..., 8th は各々予測に用いた最高次数  $d$  を表している。つまり, 次のような値を求めている。

$$Y = \{f - \tilde{f}(X, d)\}^2 \div 200$$

また実線は定義域が  $[0, 1]$  の場合, 点線は定義域が  $\{0, 1\}$  の場合を示している。

定義域  $\{0, 1\}$  及び定義域  $[0, 1]$  の場合両方とも学習事例を増やすと予測精度が上がる事が確認できた。特に定義域  $[0, 1]$  の場合は学習事例が少なくても予測精度はかなり良いことがわかった。これは先の実験 (精度についての実験) でもみられたように, 学習事例が一様に分布しているので効率良く事例を学習できるからであろう。

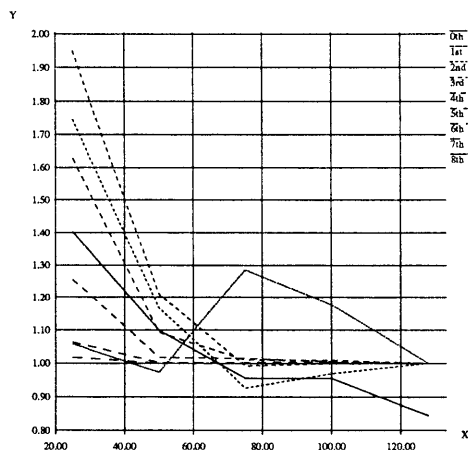


図 7 論理式  $(a \wedge b) \vee (c \wedge d) \vee (e \wedge f) \vee (g \wedge h)$ , 定義域  $\{0, 1\}$  の場合

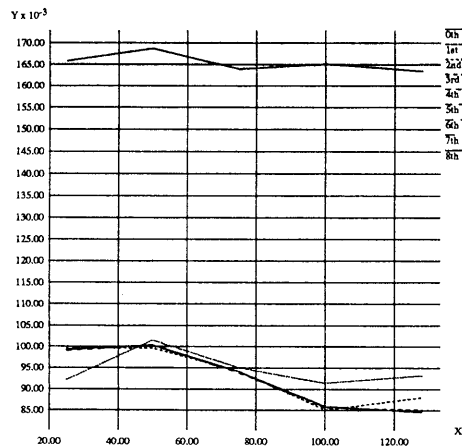


図 8 論理式  $(a \wedge b) \vee (c \wedge d) \vee (e \wedge f) \vee (g \wedge h)$ , 定義域  $[0, 1]$  の場合

例が少なくても良い精度で予測できることが確認できた。

## 6 おわりに

我々は多重線形関数  $[0, 1]^n \rightarrow R$  に拡張したスペクトル理論を用いて、学習アルゴリズムを考案した。本論文ではそのアルゴリズムの精度を、定義域  $\{0, 1\}$  の学習アルゴリズムと比較実験をすることによって検証した。この実験によって、定義域  $[0, 1]$  の学習アルゴリズムは一様確率分布に従って発生した事例を学習することによって、少ない事例でも効率良く学習することができ、予測精度も良いことがわかった。今後はこの実験結果を基に更に本アルゴリズムの改良を行なっていきたい。

## 参考文献

- [1] Nathan Linial, Yishay Mansour and Noam Nisan: Constant Depth Circuits, Fourier Transform and Learnability, *Journal of Association for Computing Machinery*, Vol.40, No.3, pp.607-620, 1993.
- [2] Johan Hastad: *Computational Limitations of Small-Depth Circuits*, Ph.D. dissertation, MIT Press, 1986.
- [3] 月本 洋: 古典論理のすべ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 D-1, Vol. J77-D-1, No.3, pp.247-252, 1994.
- [4] 月本 洋: 論理モデルとしての多重線形関数空間を用いた定性的命題獲得, 学位論文, 東京大学, 1994.

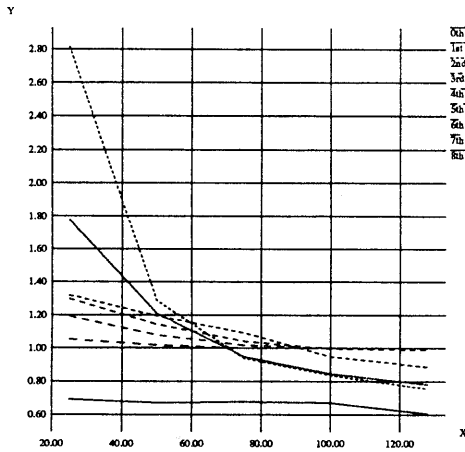


図9 論理式  $(a \wedge b \wedge c \wedge d) \vee (e \wedge f \wedge g \wedge h)$ ,  
定義域  $\{0, 1\}$  の場合

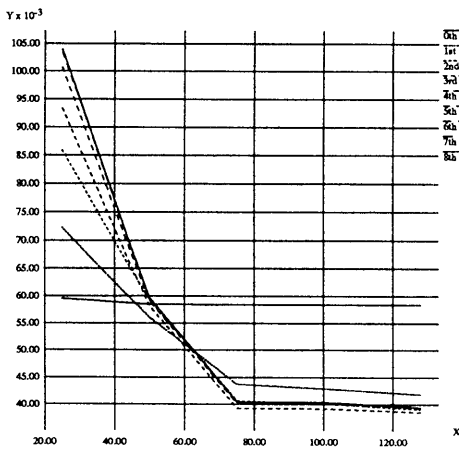


図10 論理式  $(a \wedge b \wedge c \wedge d) \vee (e \wedge f \wedge g \wedge h)$ ,  
定義域  $[0, 1]$  の場合

以上より、多重線形関数  $[0, 1]^n \rightarrow R$  に拡張したスペクトル理論を用いた学習アルゴリズムは、フーリエ係数が急激に減少して低い次数で0になること、また学習事例として  $[0, 1]$  で一様に分布した事例を用いることができることにより、用いるフーリエ係数の最高次数が低くても、また学習事