

修士論文

相互情報量で修飾された自己組織化 マップによる概念構築

指導教員 廣瀬 明 教授

電子工学専攻 66466 北原 邦紀

平成20年2月4日 提出

目次

第1章	序論	1
1.1	研究の背景	1
1.1.1	言語に関する研究	1
1.1.2	機械による言語コミュニケーションの研究	2
1.2	概念構築とは	3
1.3	研究の概要	3
第2章	相互情報量を考慮するSOMによるマッピング	4
2.1	SOMのアルゴリズム	5
2.2	相互情報量を用いたSOMの提案	8
2.3	テストデータでの実験	9
2.3.1	対応一つ	9
2.3.2	対応三つ	14
2.4	実際の画像と音声のデータで実験	20
2.4.1	画像特徴量	21
2.4.2	音声特徴量	22
2.5	実験	26
第3章	SOM of SOMs への拡張	31
3.1	SOM of SOMs の原理	32
3.2	実験	35
3.3	相互情報量を用いたSOM of SOMs	36
3.4	テストデータでの実験結果	36
3.5	実データでの実験結果	41
第4章	まとめ	46
	謝辞	47
	発表文献	52

第1章 序論

1.1 研究の背景

ここ数年、ヒューマノイドロボットの研究が盛んである。二足歩行をはじめ、その研究成果は目を見張るものがある。現在のロボットは工業用のロボットアームの研究からの流れでつくられており、動き等をすべて作りこんでいったものが多い。作りこんだロボットは、ある決まった作業の繰り返し等には非常に強い。しかし、汎用性があまりないというデメリットもある。ヒューマノイドの形状は、あるひとつの作業に従事するためには、あまり向いていない。ヒューマノイドに求められているものは、人間と同程度、もしくはそれ以上に柔軟に様々なことを行えることである。そのため、作りこまなくても学習によって様々なことができるようになっていくロボットが、これから求められていくと思われる。特にコミュニケーションの学習に焦点をあた場合、言語獲得が重要となる。言語獲得のためには、認知言語学の観点から考えると、概念獲得についての研究を行わなくてはならないと考えられる。本研究では、概念獲得に焦点をあて、複数モダリティから概念構築をするための、相互情報量を考慮した自己組織化マップについて研究を行った。

1.1.1 言語に関する研究

言語に関する研究は昔からおこなわれていたが、言語発達研究の詳細な記録は、19世紀の中頃にドイツで始まったとされている。先駆者は哲学者のティーディマンで、彼は実子の誕生から詳細な発達記録を残した。この研究は、「事例研究にもとづく記述研究」とよばれ、今でも行われている。

一方、同時期に、スイスのフェルディナン・ド・ソシュールが構造主義言語学と呼ばれる研究を始めている。一般に、ソシュールは言語の共時的な構造を重視したことで知られる。すなわち、言語の起源や歴史ではなく、ある一時点における言語の内的な構造が、言語を理解する鍵だと考えた。

一方、アメリカでは、ネイティブ・アメリカンの言語を研究していくなかで生まれた構造主義文法から生まれたアメリカ構造主義言語学が主流であった。文字すら存在しない全く未知の言語を記述するには、音の採取から始め、形態の整理、統語の記述、意味の記述という段階を経るのが一般的である。構造主義文法はこの段階を遵守するため、始めに音韻論ありきで、音韻論が一通り整備された段階で形態論、次に統語論とステップアップしていく分析手法を取る。しかし、実際は統語情報が無いと形態論がうまく記述できないなど、ステップは必ずしも一方向的なものでは

ないことが次第に明らかとなった。その結果、ついには形態論の段階から先へはなかなか進展できない事態が生じることとなった。

そんな中、チョムスキーの生成文法論が生まれた。チョムスキーの考え方は、以下のようなものであった。

- 言語は数学モデルのように明確に説明できる
- 言語獲得は、初期状態から定常状態である個別言語への遷移である
- 普遍文法（初期状態）はヒトに固有で、言語獲得は、そのパラメータを変えていくことである

生成文法論の考え方は、コンピュータと非常に相性がよく、文章理解や対話システムで現在よくつかわれている。しかし、言語には例外があまりに多いことや、時代とともに変わっていくことを説明しきれないことなど、批判も多く、アンチチョムスキーともいえる研究も多くなされた。その中のひとつとして、認知言語学が研究されている。認知言語学は、人間の認知活動の一部としての言語を通して、人間と文化の本質を探究する言語学の一分野。主に意味を扱うので認知意味論ともいう。認知言語学は、生成言語学の統語論中心主義への反省から、カリフォルニア大学のレイコフらによって70年代に提唱された。生成言語学が無視したメタファーやレトリックなどの言語現象を、認知心理学のプロトタイプ理論によって読み解き、さらにその言語現象が生成文法のようなメカニズムに起因するのではなく、人間の現象学的・民俗的な身体感覚に起因するとして体系的な記述を試みた。現代言語学・哲学における意味論の伝統を独創的に発展させた認知言語学は、狭隘な客観主義的言語モデルの限界を指摘し、人間心理や身体・空間感覚に即した言語観に、現代思想的特色が見られる。本研究では、この認知言語学の考え方をもとに、概念空間の形成の仕方について研究を行った。

1.1.2 機械による言語コミュニケーションの研究

従来の人と機械の言語によるコミュニケーションの研究は、大きくわけて二種類の流れがある。ELIZA[Weizenbaum 66]は、話題を限定せず、シンプルな規則にしたがって、人に次の発話を促す単純な応答文を生成するプログラムである。人工無能などもこの種類であるとかんがえられる。SHRDLU [Winograd 72]は環境を限定し、「積み木をもちあげて」というと、理解して動く。しかしこれらの対話コミュニケーションは、基本的には作りこみによるものであり、はじめから言葉をしゃべったり理解したりすることができるが、知らない言葉などに弱かったり、人間がすぐに飽きるといった弱点がある。そのため、学習によって言語を獲得していくロボットが近年注目されている。たとえば、infanoidは3、4歳児とほぼ同じ大きさ（座高480mm）の上半身ヒューマノイドが研究されている[1]。頭部には左右の目があり、目から得られる画像は、コンピュータによって処理され、人間の顔やオモチャを検出・追跡する。耳にあたる左右のマイクロフォンから人間の声を聞きとり、その韻律情報や音韻情報を抽出すること、また、それらの情報を音声合成装置に入力

することで、いわゆるオウム返しができるようになっている。人間学習機能は現在も不明な点が多く、新しい発見が非常に多い。ロボット、特にヒューマノイドロボットの研究には、人間の脳や身体の研究が非常に参考になると思われる。今後、より汎用性の強いロボットの実現のため、作りこみ以外の、学習を基本としたロボットがますます重要視されていくとかがえられる。

1.2 概念構築とは

何をもって概念とよび、何をもって概念構築を定義するかは議論の分かれるところではあるが、現在の研究は、おおまかに二種類に分けられる。どのように環境を捉えるのか、というものと、捉えることによってえられた情報をどのように蓄積するか、というものの二種類である。

どのように環境をとらえるのか？という研究は、どのように事物を特徴量に落とすのか、その方法を探すという形で研究が行われている。たとえば、岩橋らは、認知言語学の考え方に基づき、抽象的な概念（飛び越える、のる、あがる）を、物体の相互の関係から特徴に落とし込むことを研究している [2] ~ [4]。

もうひとつの情報をどのように蓄積していくのかという研究は、どう経験を蓄積し、それらの関係をどう保存していくのかという形で行われている。特に概念構築の研究では、得られた情報の蓄え方 + シンボルグラウンディングと考える方法が一般的となっている。情報の蓄積方法は、今後得られる情報にどのようにバイアスをかけるようになるのか、ということにもつながる。また、ラマチャンドランの説明するブーバー・キキ理論 [43] や、大脳のコラム構造等を参考にすると、多モダリティの類似性を考慮したマッピングが創造性につながる可能性があり、非常に重要となる。

シンボルグラウンディングは、基本的には、画像データに音声を文字データにしたものをタグづけしていく形で行われているものが多い。しかし、この方法では、文字データに落とし込んだものしか使えないため、未知言語に対応することがむずかしくなる。

Roy たちの研究では、情報の構築はしていないが、音声と画像の対応付けに相互情報量を用いている [5]。彼らの研究では、文字をつかっておらず、対応付けの正誤を与えなくても良いことから、新しい言葉にも対応できる柔軟な方法であると考えられる。今回の論文では、基本的な対応付けのベースは、この Roy たちの考え方に基づいて、相互情報量を用いて構築を行う方法を提案する。

1.3 研究の概要

本研究では複数モダリティを扱う手法として、相互情報量を考慮する SOM を提案する。第 2 章では、その原理について説明し、テストデータでの実験、および実データでの実験を行った。第 3 章では、非線形データにも応用可能にするために、SOM を SOM^2 に拡張した。第 4 章で、まとめをおこなった。

第2章 相互情報量を考慮するSOM によるマッピング

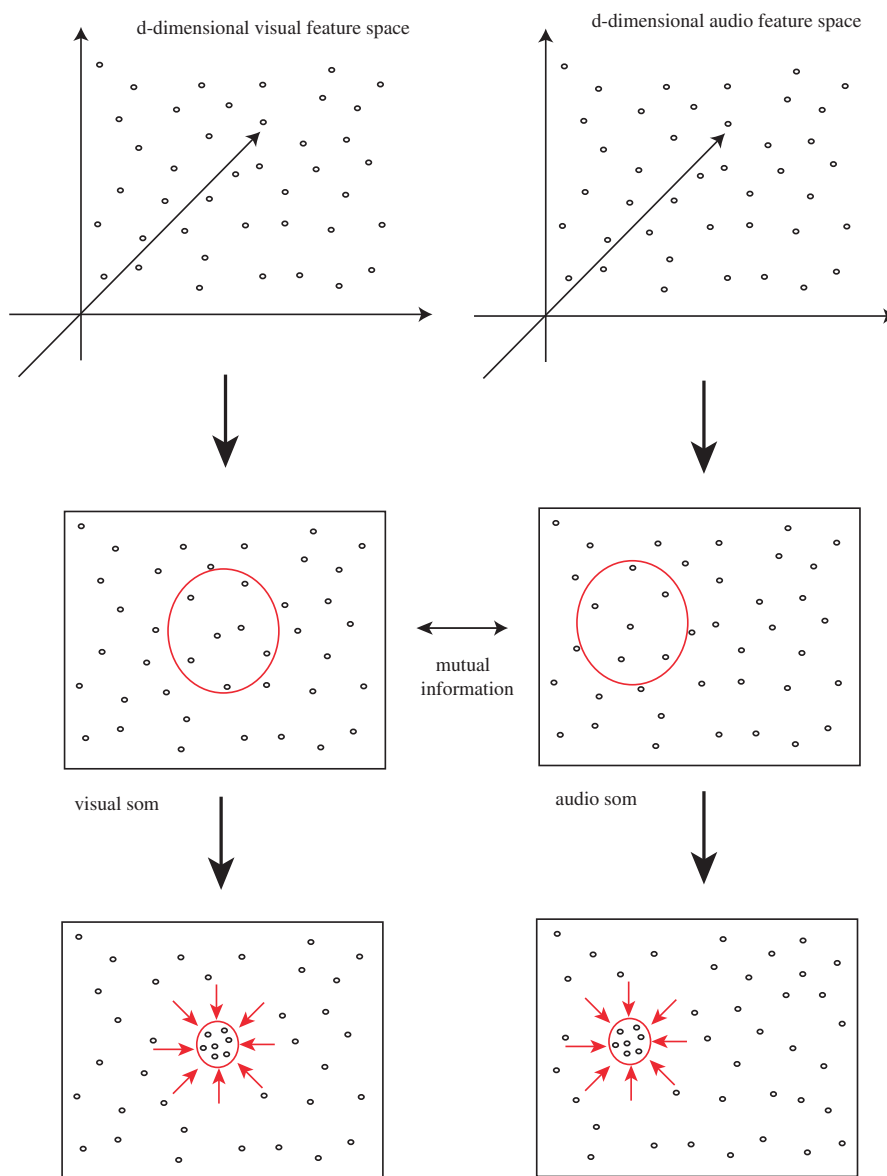


図. 2.1: 全体の流れ

各特徴空間を二次元にマッピング, さらに相互情報量を見ることによって相互の対応を検出. それに応じてまとまりを構築. 実際には, マッピングとまとまりの構築は同時並行で行われる.

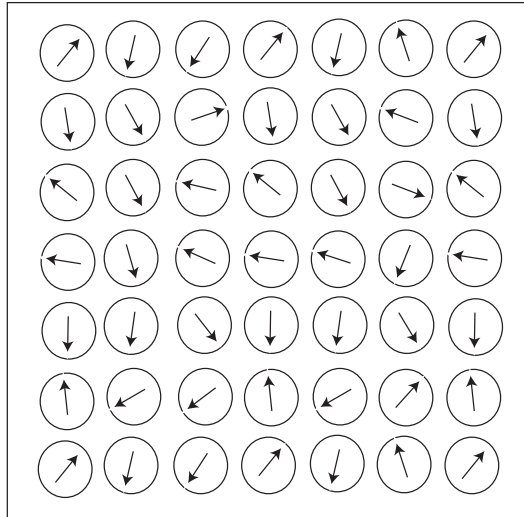
今回提案する相互情報量を考慮する SOM による構築の全体の流れを図 (2.1) に示す。各特徴空間を、いったん SOM によって低次元にマッピングを行う。その後、対応している箇所を相互情報量をみることにより検出する。さらに、対応している箇所が、一箇所に集まってくるようにマッピングをしていく。これにより、一方のモダリティがもう一方の教師の代わりとなるため、教師なしの構築が可能となる。たとえば画像データと音声データの場合、画像データのどこからどこまでを一つの同じカテゴリとしてみるか、ということ、音声データとの相互情報量をみることによって得ることができる。音声側も、どこまでを同一データとしてみるのかを画像データから取得することができる。

2.1 SOM のアルゴリズム

自己組織化マップ (Self-Organizing Map, SOM) は、T.Kohonen により 1981 年頃に発表された教師なし学習ニューラルネットワークで、入力パターン群をその類似度に応じて分類する能力を自律的に獲得していくニューラルネットワークである [7],[8]。n 次元の入力データ群をニューロンの 1 次元配列にマッピングするものを 1 次元 SOM、2 次元配列にマッピングするものを 2 次元 SOM という。SOM では、入力データに対して、そのデータの特徴を最もよく捉えたある 1 つのニューロンが発火する。SOM は、入力データ間の近さが、1 次元または 2 次元配列上で発火するニューロンの近さに反映される「トポロジー保存性」という特徴を持つ。

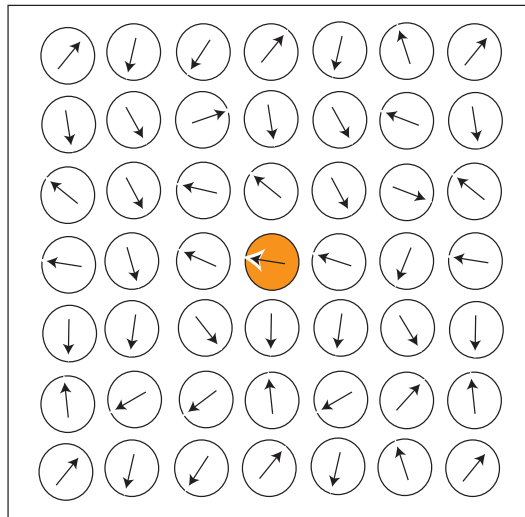
以下にそのアルゴリズムを示す。

The weights of the neurons are initialized to random values



←
input vector

The neuron with weight vector most similar to the input is called the best matching unit (BMU).



The weights of the BMU and neurons close to it in the SOM lattice are adjusted towards the input vector

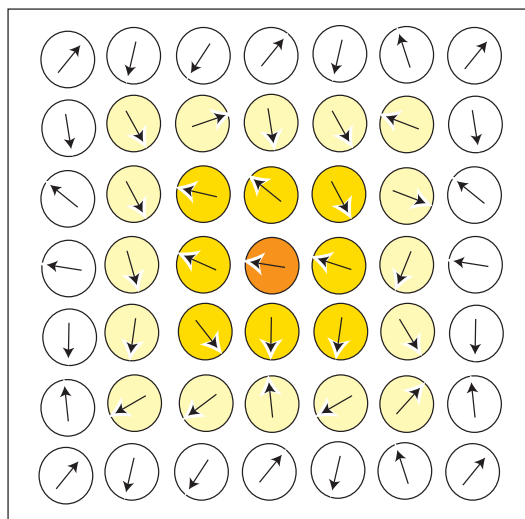


図. 2.2: SOM のアルゴリズム

- 各ノードは, 入力データと同じ次元の重みベクトル W_i をもっている.
- すべてのノードの重みベクトルをランダム化する
- 入力ベクトル x_j を一つ用意する
- マップ上の全てのノード一つ一つに対して, 入力ベクトルと各ノードの重みベクトル間の距離を計算する. 一致値にはユークリッド距離が用いられる.
- 各ノードを検査して, 最も一致値が小さい(ベクトル間の距離が短い=もっとも良く一致した)ノードを見つける. このノードをBMUと呼ぶ (Best Matching Unit).
- BMU の近傍のノードの重みベクトルを次のように変更し, 入力ベクトルに近づける

$$W_i(t+1) = W_i(t) + \Theta(t)\alpha(t)(x_j - W_i(t)) \quad (2.1)$$

ここで, $\alpha(t)$ は時間に対して単調減少する関数で, $\Theta(t)$ は近傍関数とよばれる. $\Theta(t)$ は主に BMU からの距離に比例して減少していく関数を用い, 時間に対しても単調減少していくものを用いる場合が多い. $\alpha(t), \Theta(t)$ の式は, たとえば以下のようなになる.

$$\alpha(t) = \left(1 - \left(\frac{t}{TMAX}\right)\right) * \alpha(0) \quad (2.2)$$

$$\Theta(t) = \exp\left(-\frac{(a-b)^2}{2\delta^2(t)}\right) \quad (2.3)$$

$$\delta(t) = \left(1 - \left(\frac{t}{TMAX}\right)\right) * \delta(0) \quad (2.4)$$

ここで, a は更新する素子の座標, b は BMU の座標である. 以上を繰り返していくことにより, 入力データ間の距離を反映したマッピングができる.

2.2 相互情報量を用いたSOMの提案

複数モダリティの関係性を考慮したマッピングの実現のため、相互情報量を用いたSOMを提案する。

まず、複数（今回は二つ、audio と visual 用に用意）のSOMを用意する。ソムの学習をそれぞれ行うが、学習の際の近傍関数の決定に、相互情報量を用いる。相互情報量は、以下の式によって計算する。

$$I(A; V) = \sum_i \sum_j P(A = i, V = j) \log \left(\frac{P(A = i, V = j)}{P(A = i)P(V = j)} \right) \quad (2.5)$$

$$P(A = i) = \frac{|A = i|}{n} \quad (2.6)$$

$$P(V = j) = \frac{|V = j|}{n} \quad (2.7)$$

$$P(A = i, V = j) = \frac{|A = i, V = j|}{n} \quad (2.8)$$

$$A = \begin{cases} 0, & \text{if } d_A(x, y) > r_A \\ 1, & \text{if } d_A(x, y) \leq r_A \end{cases} \quad (2.9)$$

$$V = \begin{cases} 0, & \text{if } d_V(x, y) > r_V \\ 1, & \text{if } d_V(x, y) \leq r_V \end{cases} \quad (2.10)$$

$I(A, V)$ が相互情報量、 n は入力データ数である。 r_A, r_V は中心データからの距離で、これらを変えていき、 $I(A, V)$ が最大値 I_{max} になるものを探す。

この I_{max} の値によって、近傍関数をかえていく。

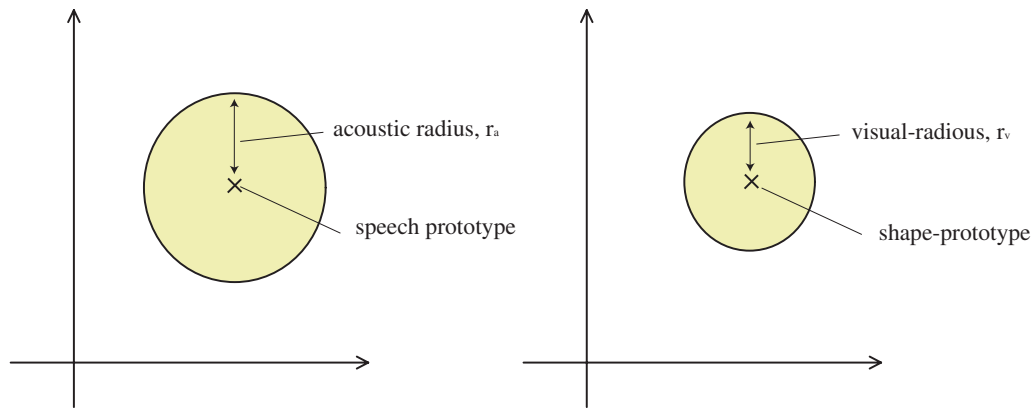


図. 2.3: 相互情報量

本来は周囲との相互情報量をみて、相対的に動作をかえるべきだと思われるが、今回は、相互情報量 I_{max} が閾値 H 以上になったところで、そのときの中心データから r_A, r_V 以内の範囲 ϕ に含まれる入力データ x_i の学習については、近傍関数 $\Theta(t)$ の δ を小さな値にする。

$$\Theta(t) = \exp\left(-\frac{(a-b)^2}{2\delta^2(t)}\right) \quad (2.11)$$

$$\text{if } x_j \subseteq \phi \quad \delta(t) = \left(1 - \left(\frac{t}{TMAX}\right)\right)^2 * \delta(0) \quad (2.12)$$

$$\text{else } \delta(t) = \left(1 - \left(\frac{t}{TMAX}\right)\right) * \delta(0) \quad (2.13)$$

2.3 テストデータでの実験

2.3.1 対応一つ

実際にテストデータを用意し、普通の SOM と提案手法の両方にかけて、その動作を実験した。

テストデータは、図 (2.4),(2.5) のような二次元データを用意した。0 から 1 の範囲のデータを 50 点ずつ用意した。そのうち、はじめの 10 点は input1, input2 共に 0.25 以上 0.75 以下にランダムに分布しており、11 点目以降はそれ以外の範囲にランダムに分布している。

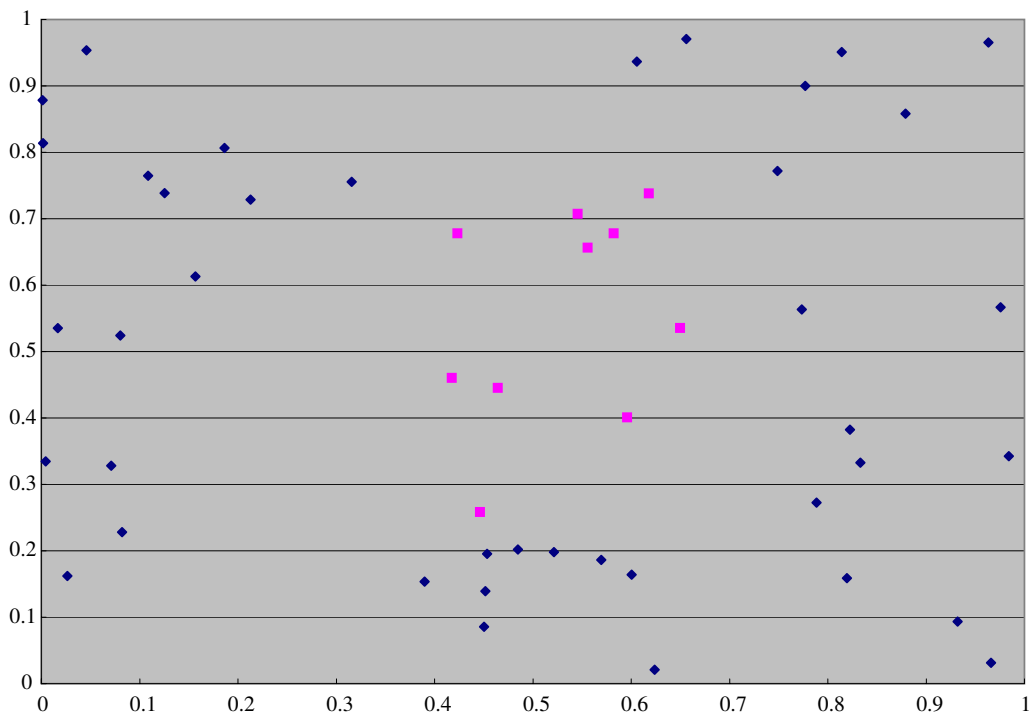


図. 2.4: testdata1

はじめの10点は0.25以上0.75以下にランダムに分布しており、11点目以降はそれ以外の範囲にランダムに分布している。

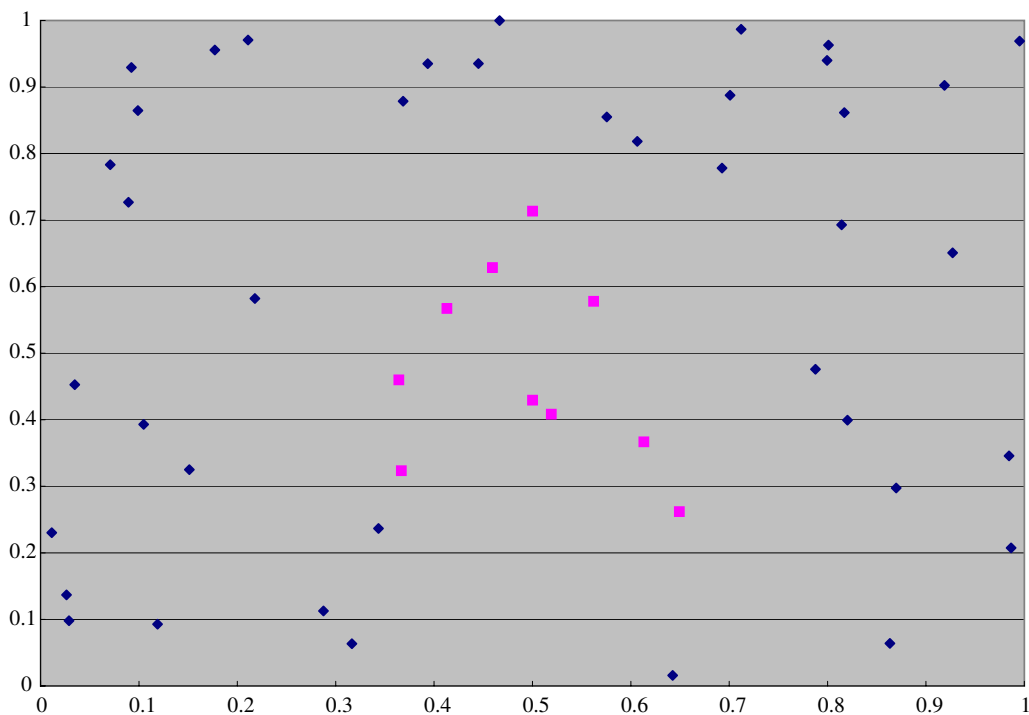


図. 2.5: testdata2

はじめの10点は0.25以上0.75以下にランダムに分布しており、11点目以降はそれ以外の範囲にランダムに分布している。

このテストデータを, 普通の二次元の SOM にかけた. SOM は, 素子を 100×100 平面にならべたものを用意した. 学習の際のパラメータは, $\alpha(0) = 1, \delta(0) = 30$, 学習回数は 200 回とした.

テストデータ 1 とテストデータ 2 を SOM にかけた結果は, それぞれ, 図 (2.6), (2.7) のようになる. また, そのときの相互情報量の値を計算すると, 図 (2.8) のようになる. SOM の結果のみをみるとデータ間の違いはみられないが, 1 番 ~ 10 番のデータは, 他のデータに比べて, 相互情報量が高いことがわかる.

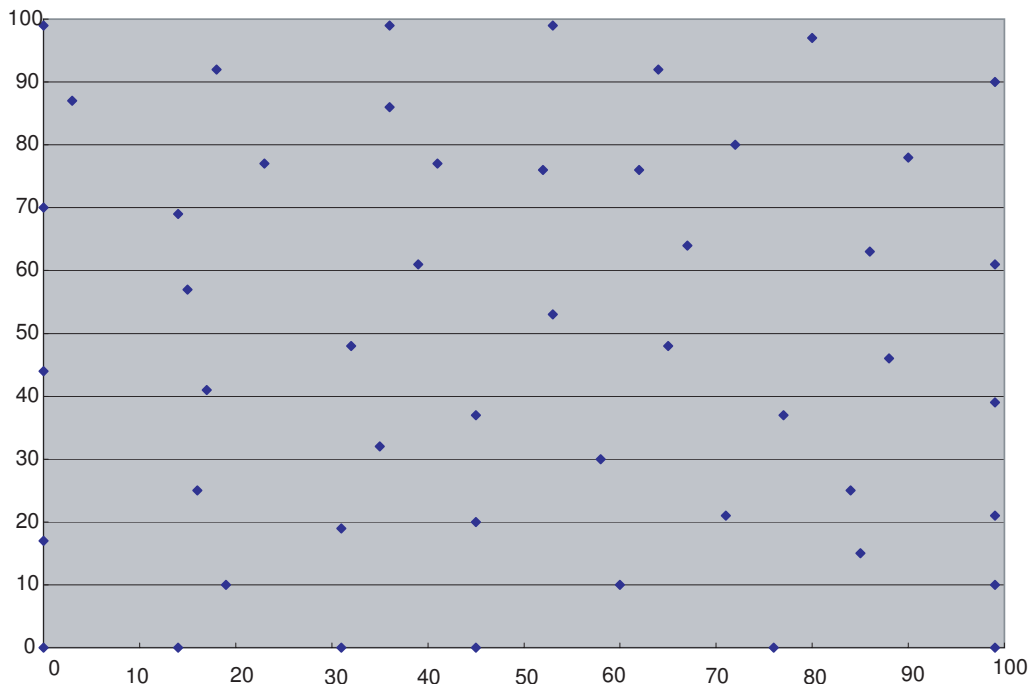


図. 2.6: testdata1 を SOM にかけた結果

縦軸, 横軸はそれぞれ SOM の素子の番号をあらわし, プロットした点は, 最終的な各データに対する BMU (Best Matching Unit) の位置を示している

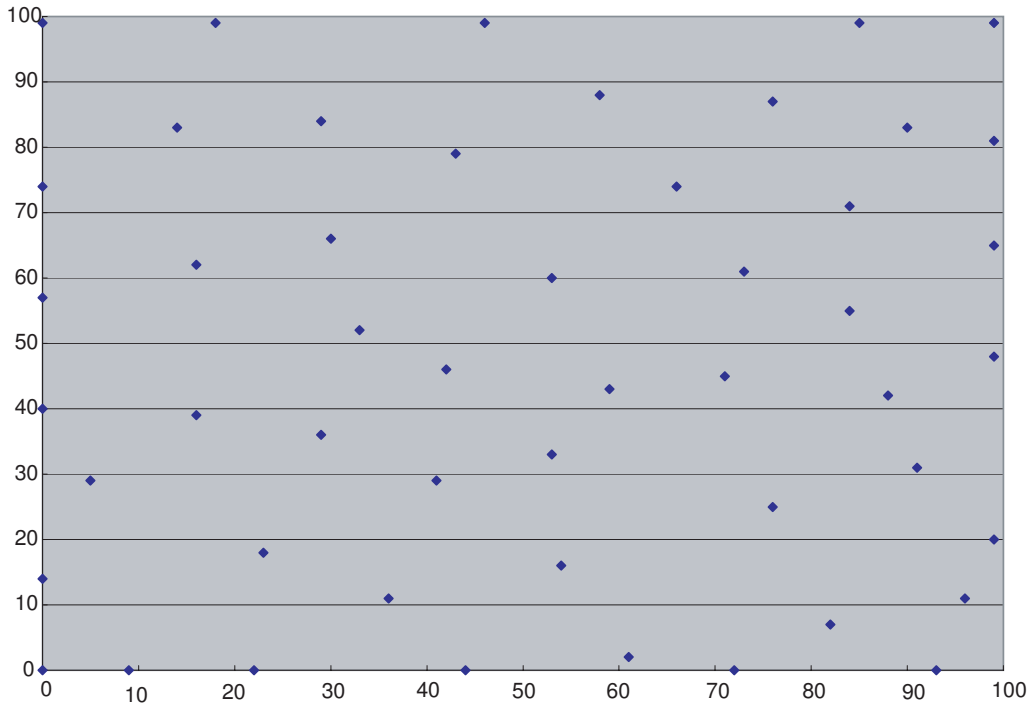


図. 2.7: testdata2 を SOM にかけた結果

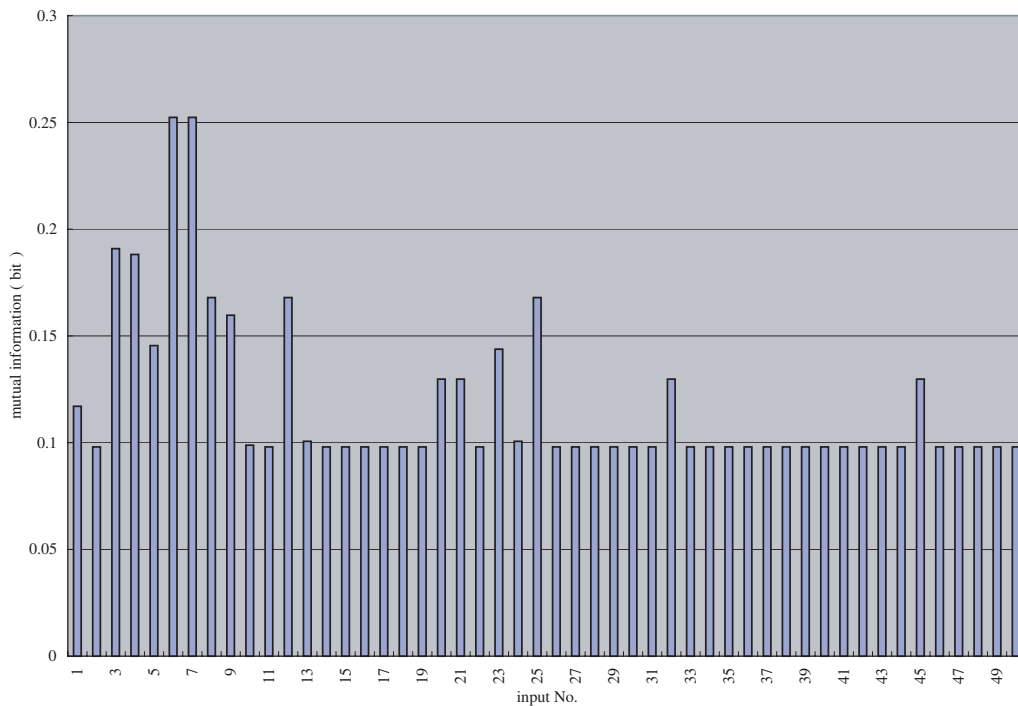


図. 2.8: mutual information

このテストデータを, 今回の提案手法にかける. SOMは, 素子を 100×100 平面にならべたものを二つ用意した. 学習の際のパラメータは, $\alpha(0) = 1, \delta(0) = 30, H = 0.2$, 学習回数は200回とした. 結果は, 図(2.9),(2.10), このときの相互情報量は(2.11)

のようになった。

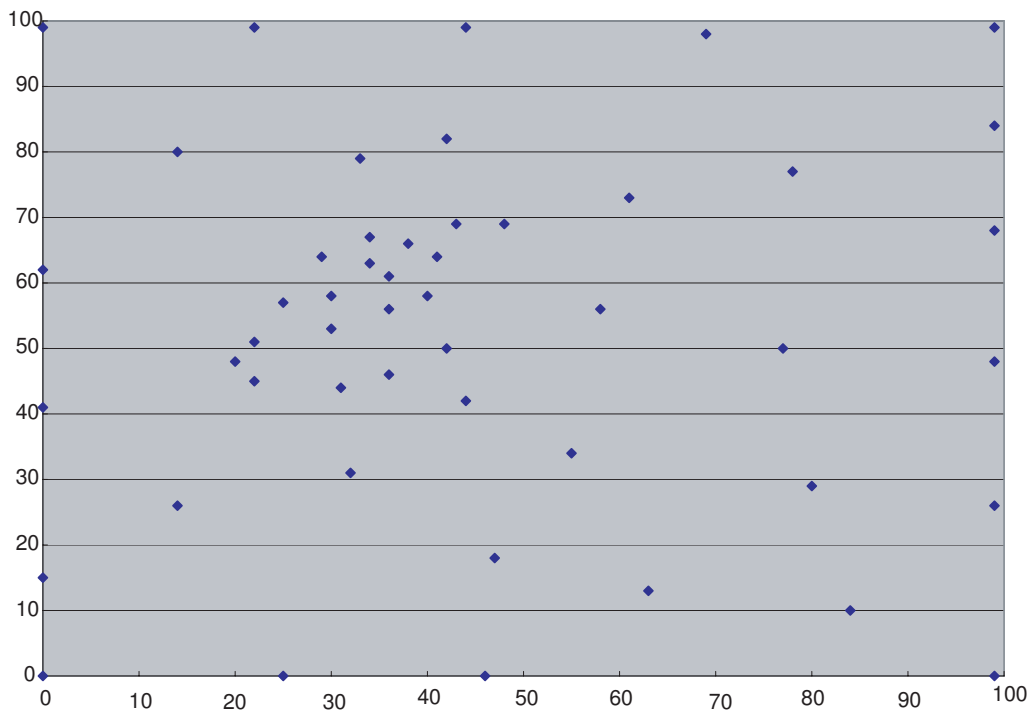


図. 2.9: 提案手法による input1 の結果

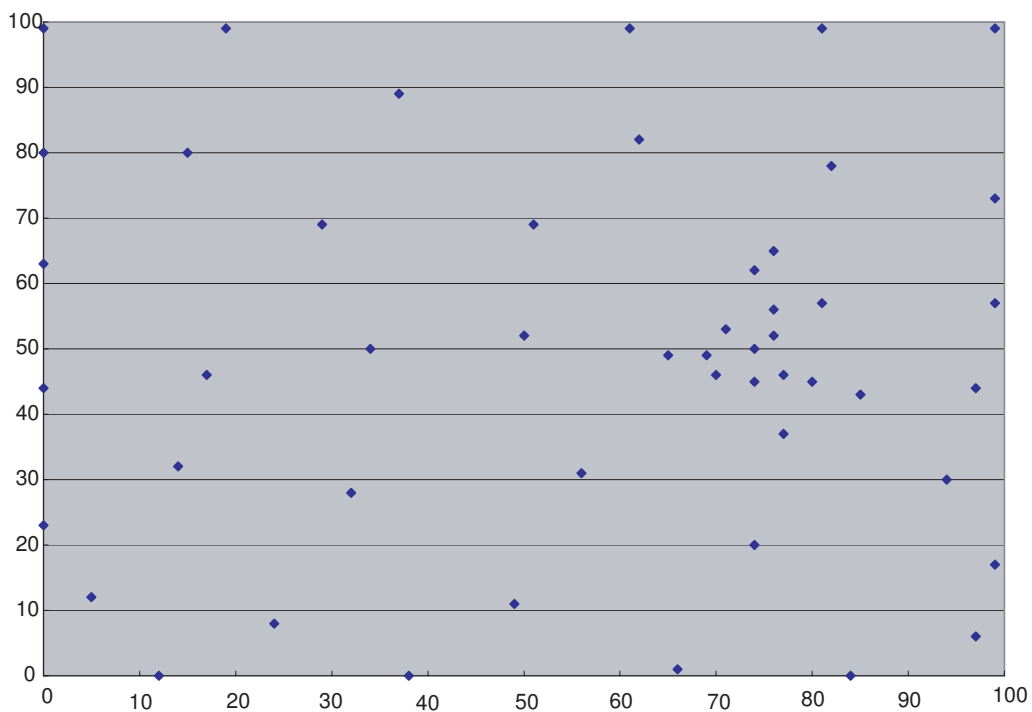


図. 2.10: 提案手法による input2 の結果

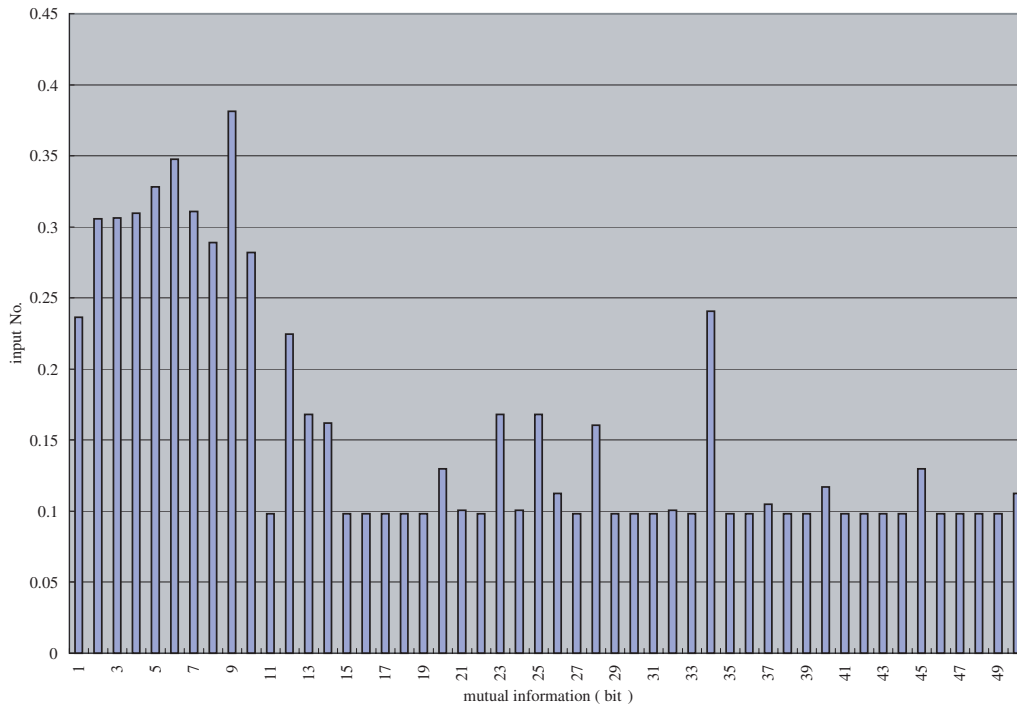


図. 2.11: 提案手法による output の mutual information

この結果から, 今回の提案手法で相互情報量の高いものがまとまっていくマッピングが可能であることがわかる. しかし, 図 (2.9), (2.10) をみると, 一箇所にまとまってきた数が, 10個より多い. これは, 相互情報量をみてまとめる範囲をきめているため, ランダムに配置したデータも範囲に含まれてしまうものがでてきたためだと思われる.

2.3.2 対応三つ

相互情報量が大きくなる箇所が, 一箇所のみではなく, 複数箇所あっても今回の提案手法が有効であるかどうかを調べるため, 図 (2.12)(2.13) のようなテストデータを用意して実験を行った.

それぞれ 250 点のデータを用意し, 最初の 1~10, 11~20, 21~30 番目のデータは特定の場所に分布し, 残りはランダムに分布している.

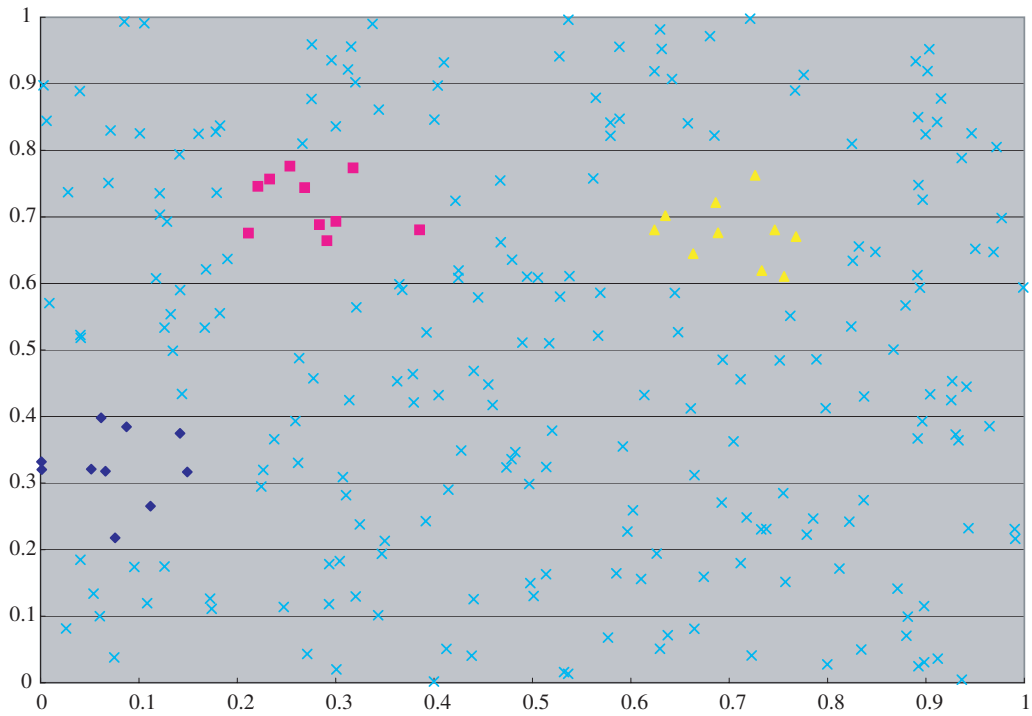


図. 2.12: testdata3

250点のデータ

1~10番目は x は $0 \sim 0.2$, y は $0.2 \sim 0.4$ の範囲

11~20番目は x は $0.2 \sim 0.4$, y は $0.6 \sim 0.8$ の範囲

21~30番目は x は $0.6 \sim 0.8$, y は $0.6 \sim 0.8$ の範囲

残りのデータはそれ以外の場所にランダムに分布している

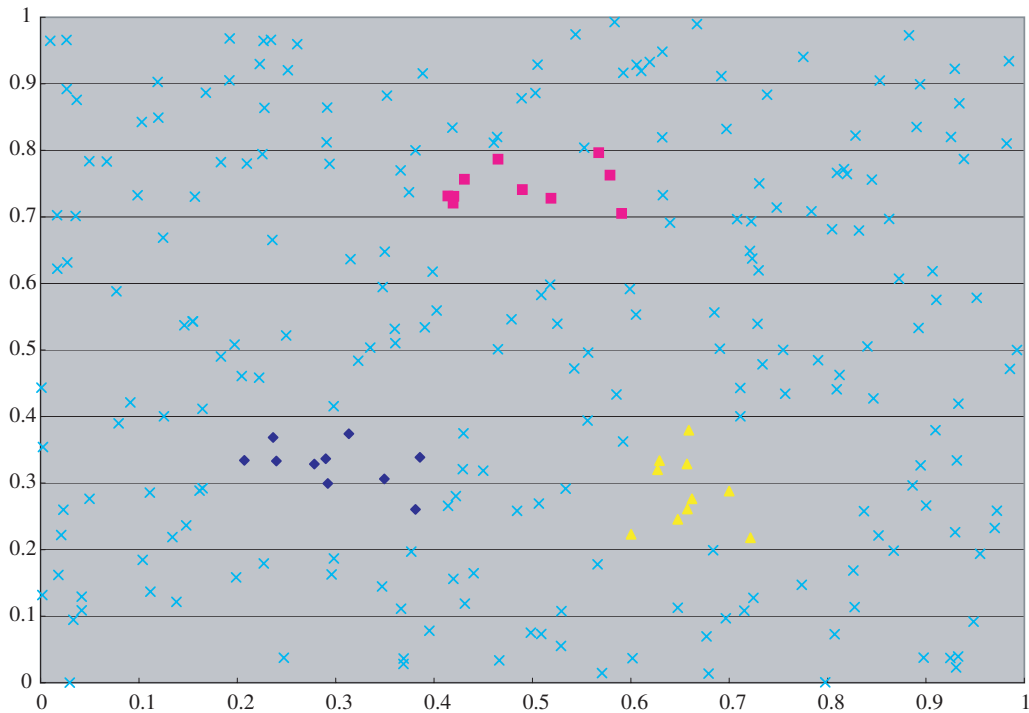


図. 2.13: testdata4

250点のデータ

1 ~ 10 番目は x は 0.2 ~ 0.4, y は 0.2 ~ 0.4 の範囲

11 ~ 20 番目は x は 0.4 ~ 0.6, y は 0.6 ~ 0.8 の範囲

21 ~ 30 番目は x は 0.6 ~ 0.8, y は 0.2 ~ 0.4 の範囲

残りのデータはそれ以外の場所にランダムに分布している

このテストデータを, 普通の二次元の SOM にかけて. SOM は, 素子を 100×100 平面にならべたものを用意した. 学習の際のパラメータは, $\alpha(0) = 1, \delta(0) = 30$, 学習回数は 200 回とした.

テストデータ 1 とテストデータ 2 を SOM にかけての結果は, それぞれ, 図 (2.14), (2.15) のようになる. また, そのときの相互情報量の値を計算すると, 図 (2.16) のようになる.

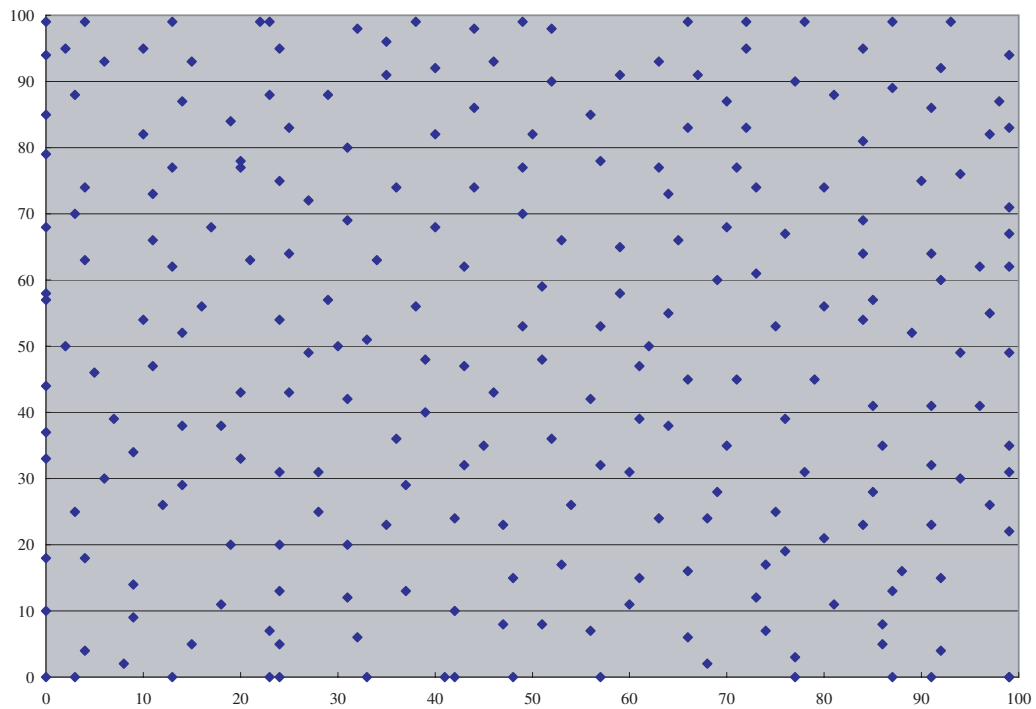


図. 2.14: testdata3 を SOM にかけた結果

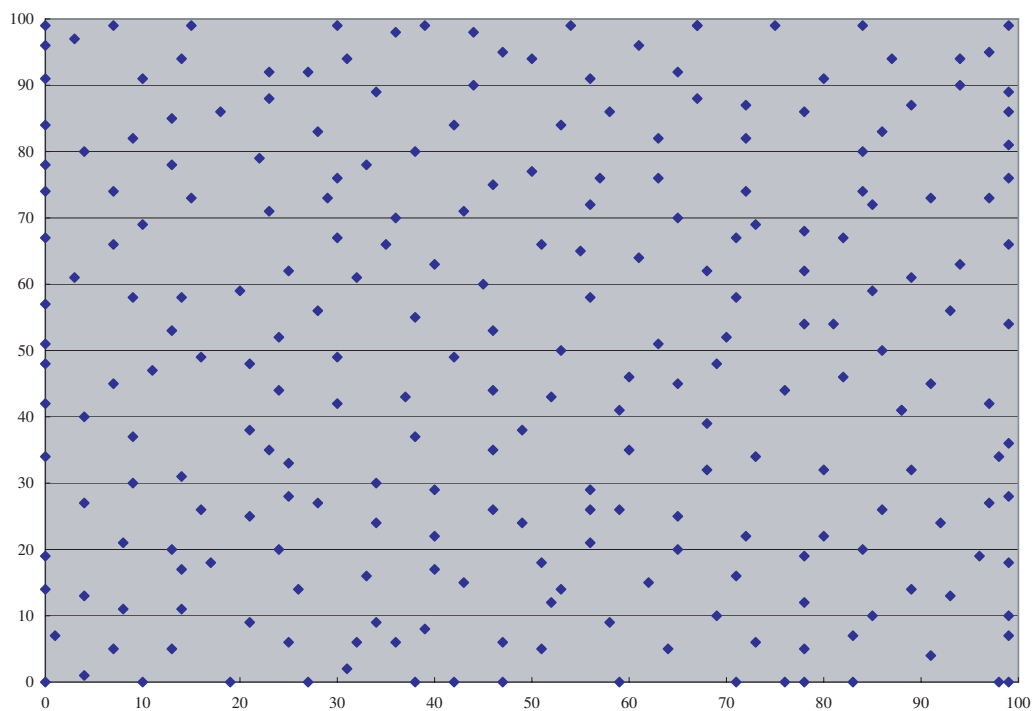


図. 2.15: testdata4 を SOM にかけた結果

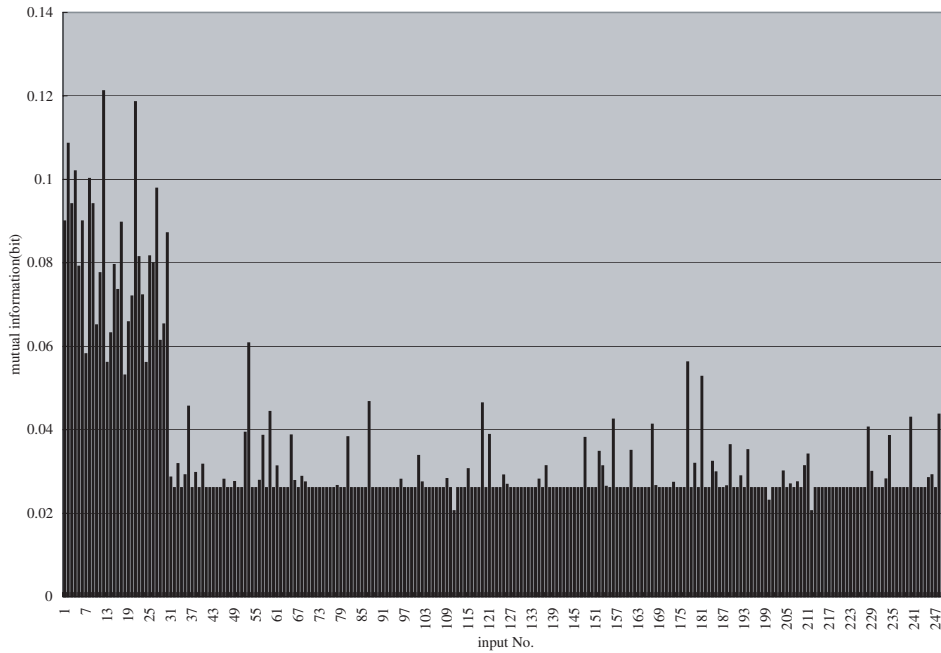


図. 2.16: mutual information

このテストデータを、今回の提案手法にかける。SOMは、素子を 100×100 平面に
 ならべたものを二つ用意した。学習の際のパラメータは、 $\alpha(0) = 1, \delta(0) = 30, H =$
 0.2 , 学習回数は 200 回とした。結果は、図 (2.17),(2.18), このときの相互情報量は
 (2.19) のようになった。図 (2.17),(2.18) それぞれ 3 箇所にとままりができてい
 ることがわかる。この結果から、対応が複数個ある場合にも、今回の手法は有効である
 と考えられる。

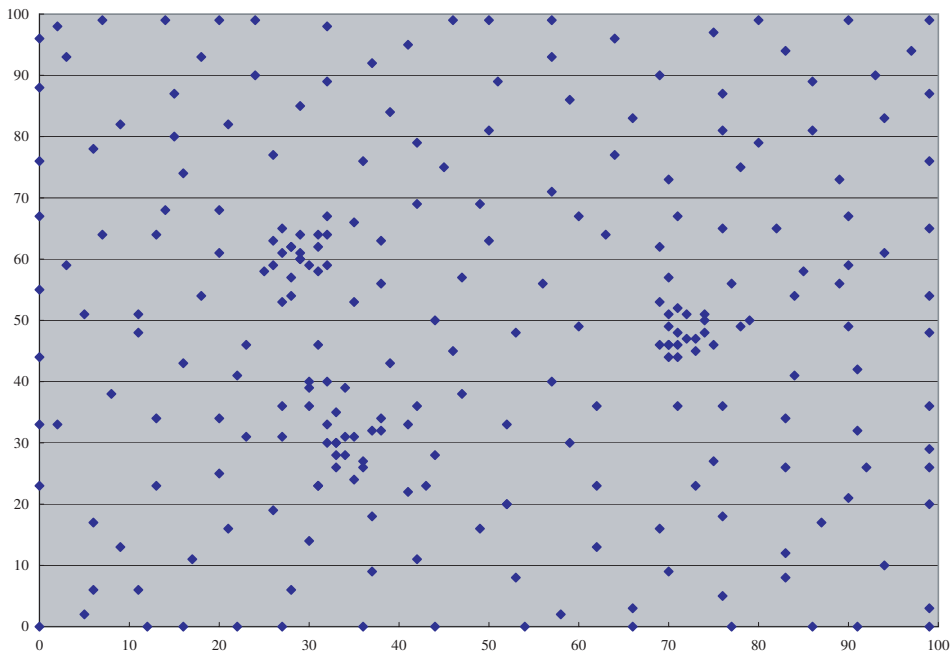


図. 2.17: 提案手法による input3 の結果

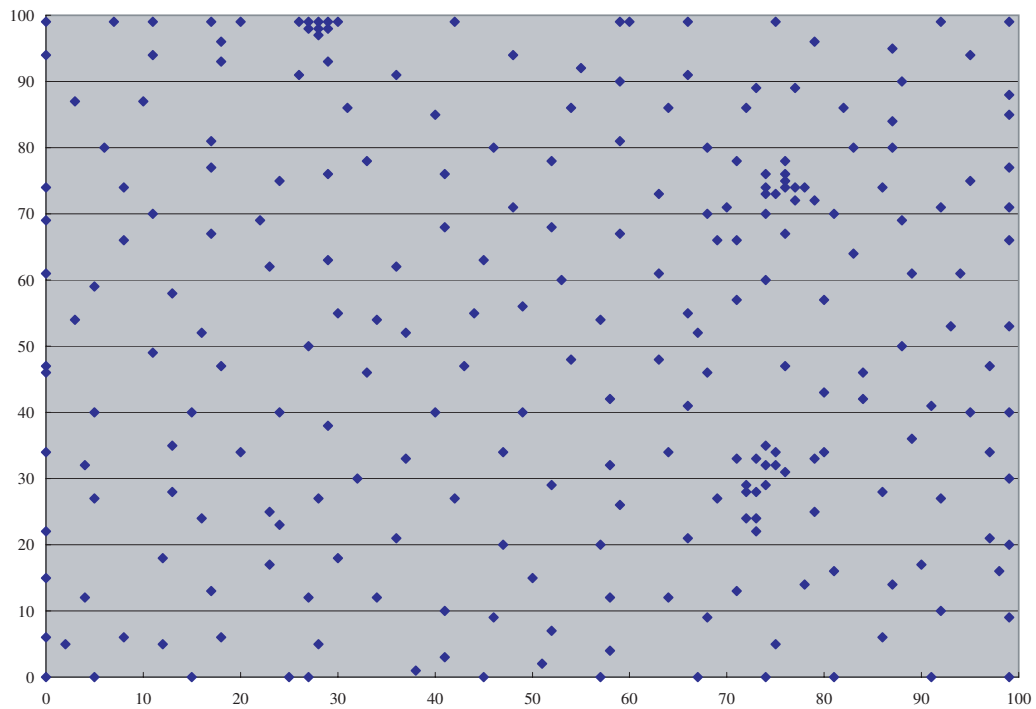


図. 2.18: 提案手法による input4 の結果

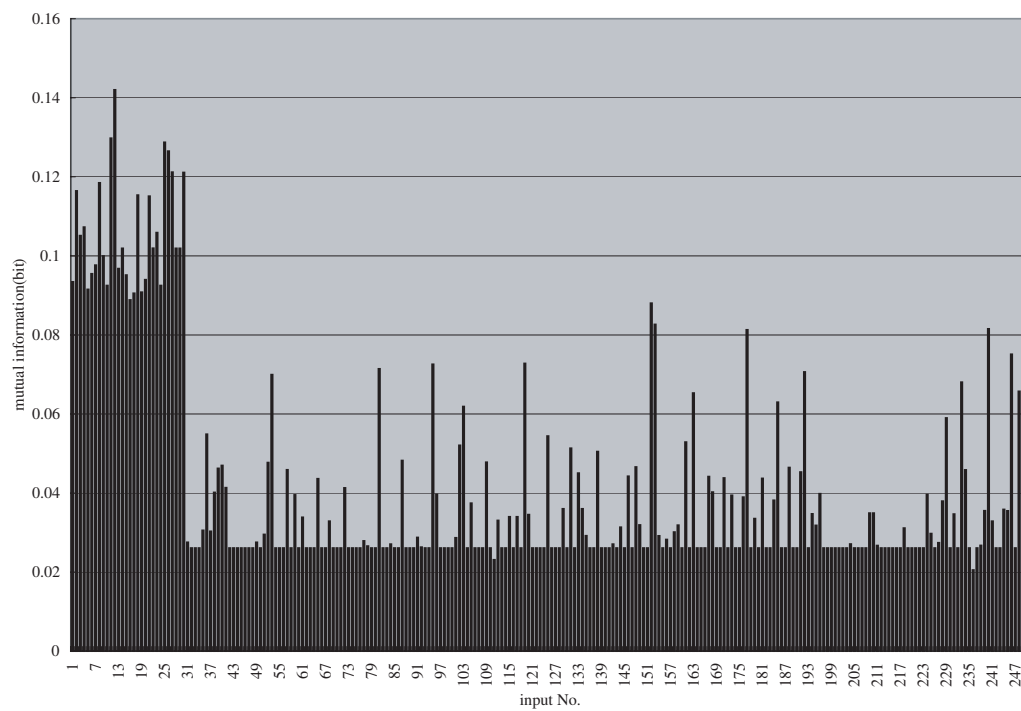


図. 2.19: 提案手法による output の mutual information

2.4 実際の画像と音声のデータで実験

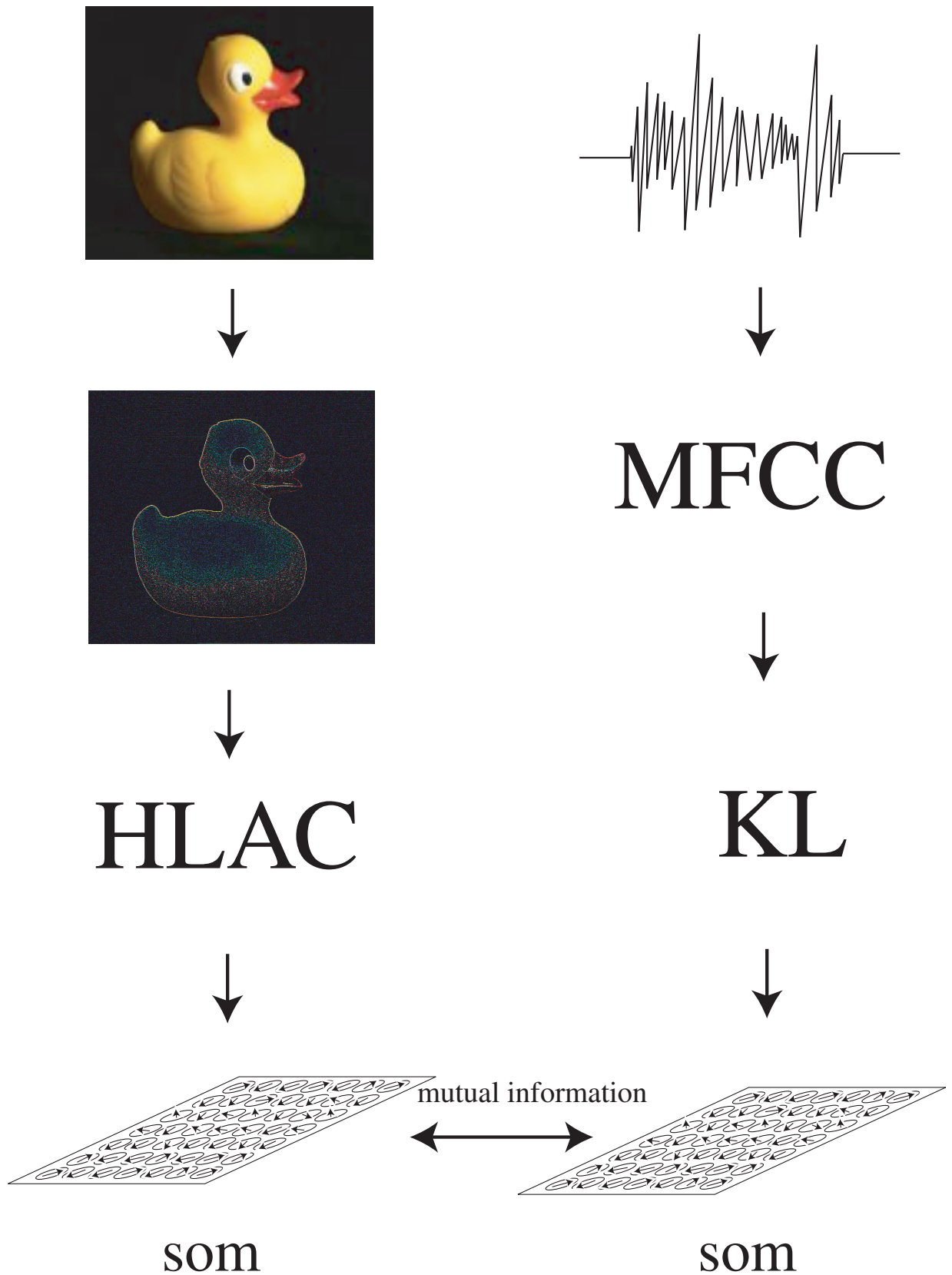


図. 2.20: 画像と音声のデータによる実験の手順

今回の提案手法が実際の画像と音声のデータを用いた場合どうなるかを実験した。手順としては、図(2.20)のようになる。まず画像データは、いったんエッジ処理をほどこし、その画像から高次局所自己相関をとり、35次元の特徴量にする。音声データは、特徴量としてMFCCをとる。このままだと音声データの時間長によって特徴量の次元が異なるため、KL展開によって26次元に落とす。これら二つの特徴量を今回の提案手法にいれる。

2.4.1 画像特徴量

画像にエッジ処理を施した後、高次局所自己相関をとることによって35次元の特徴量を得る。

高次局所自己相関特徴 (HLAC)

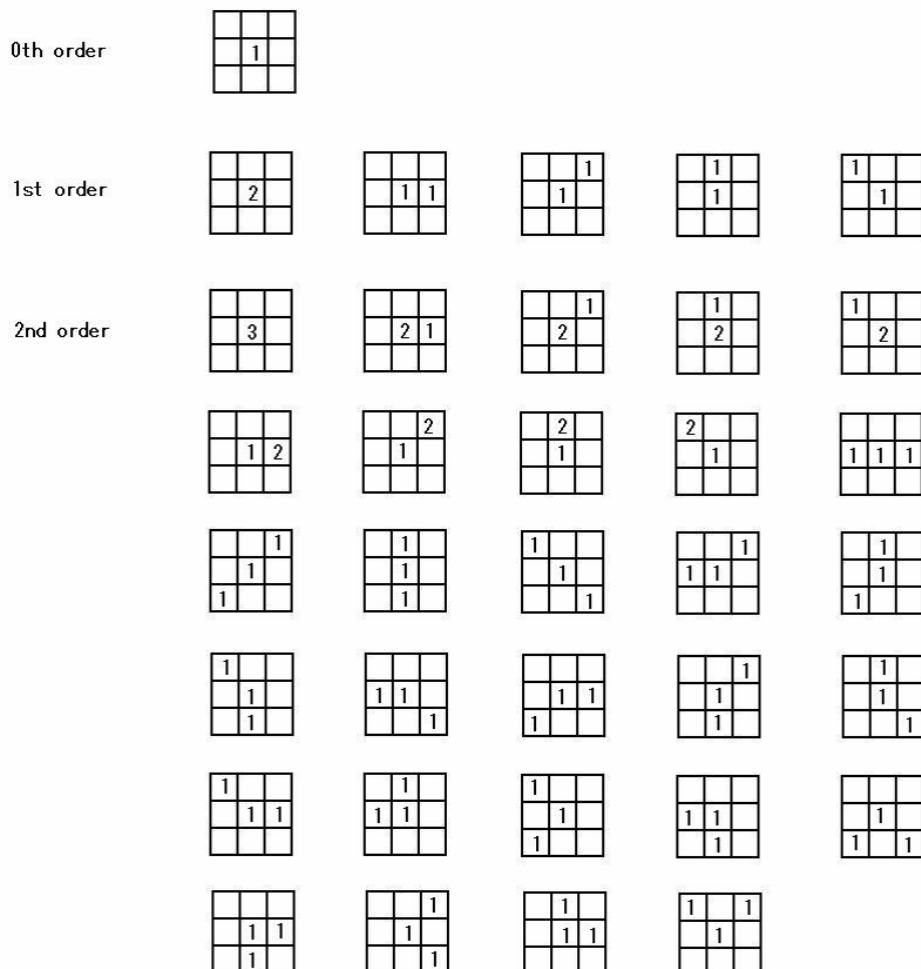


図. 2.21: hlac のとりかた
これにより、35次元の特徴量が得られる

自己相関関数を高次へ拡張したものを, 高次自己相関関数という [13]. 参照点 r の対称の値を $I(r)$ とすると, N 次自己相関関数は, 参照点周りの N 個の変位 (a_1, \dots, a_N) に対して,

$$x(a_1, \dots, a_N) = \int I(r)I(r + a_1)\dots I(r + a_N)dr \quad (2.14)$$

で定義される.

この次数 N や変異 (a_1, \dots, a_N) の取り方は無数に定義できるが, 次数 N を 2 までに, 変異 (a_1, \dots, a_N) を参照点 r の周りの 3×3 の領域に限定したものを, 高次局所自己相関 (high order local autocorrelation) という (図 (2.21)). この特徴量のとり方により, 対象物の位置に不変な特徴量をとることができる.

2.4.2 音声特徴量

音声データは, 特徴量として MFCC をとる. このままだと音声データの時間長によって特徴量の次元が異なるため, KL 展開によって 26 次元に落とす.

MFCC

ケプストラムパラメータには, 多様な計算方法がある. その中の一つとして MFCC (Mel-Frequency Cepstrum Coefficient) がある. MFCC の計算では, スペクトラル分析は周波数軸上に三角窓を配置し, フィルタバンク分析により行う. すなわち, 窓の幅に対応する周波数帯域の信号のパワーを, 単一スペクトルチャンネルの振幅スペクトルの重みづけ和で求める. さらに, 窓はメル周波数軸上に等間隔に配置される. 最終的に, フィルタバンク分析により得られた帯域におけるパワーを離散コサイン変換することで, MFCC が求められる. 現在の音声認識において最も一般的に用いられているスペクトル距離尺度は, LPC ケプストラム, MFCC (Mel Frequency Cepstrum Coefficient) の 2 つ. しかし, いずれもケプストラム距離であることに変わりはなく, 両者の本質的な相違は小さいと考えられている.

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^N m_j \cos\left(\frac{\pi i}{N}(j - 0.5)\right) \quad (2.15)$$

N はフィルタバンクチャンネルの数を表し, m_j は対数フィルタバンクの振幅を表す. 今回は MFCC をとるのに, HTK tool kit を用いた.

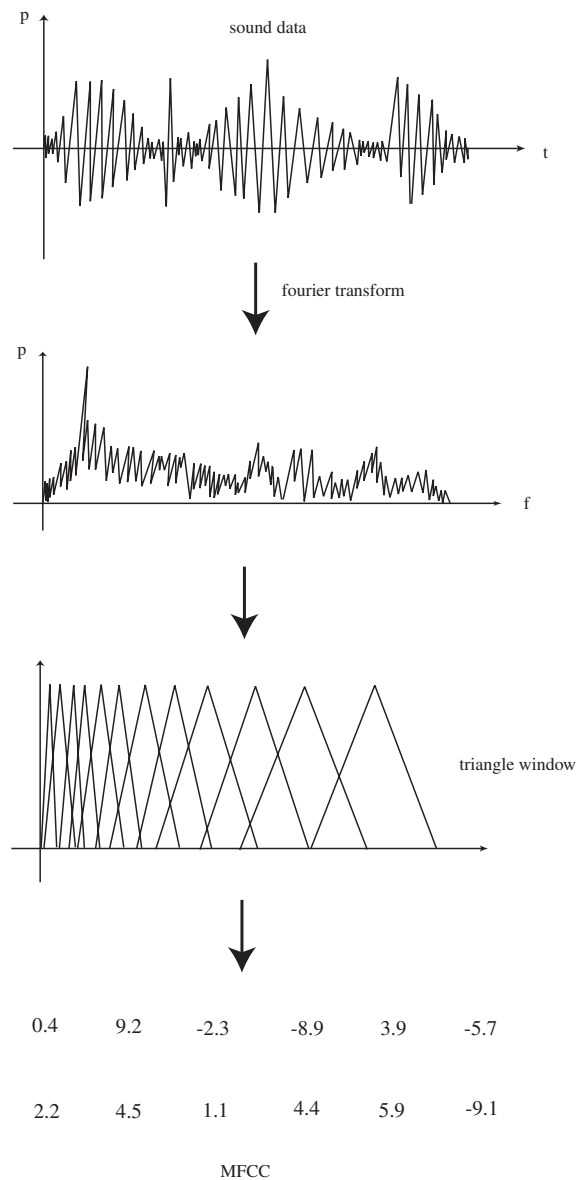


図. 2.22: MFCC の手順

K-L 展開

MFCC をとったままでは音声データの長さによって次元が異なるため、音声同士の比較が難しい。そこで、K-L 展開 (Karhunen-Loeve 展開) によって、次元圧縮を行って、それぞれの音声から得た特徴量を同一次元に落とす。

K-L 展開とは Karhunen-Loeve(カルーネンレーベ) 変換を用いる固有分解の手法である。パターン認識のための次元削減法として用いられる KL 展開は、分散最大基準もしくは原点移動を許した平均二乗誤差最小基準により求まる部分空間を使う。すなわち共分散行列 Σ の上位固有値に対応する固有ベクトルを基底とする部分空間を使う。

KL 展開のアルゴリズム KL 展開は大きく分けて 3 つの処理からなる。まず空間中の全データの共分散行列 C を計算し、次に C の固有値ベクトル行列 E を求め、最後に E を基底として全データを部分空間に射影する。以下にそれぞれの処理の概要を説明する。

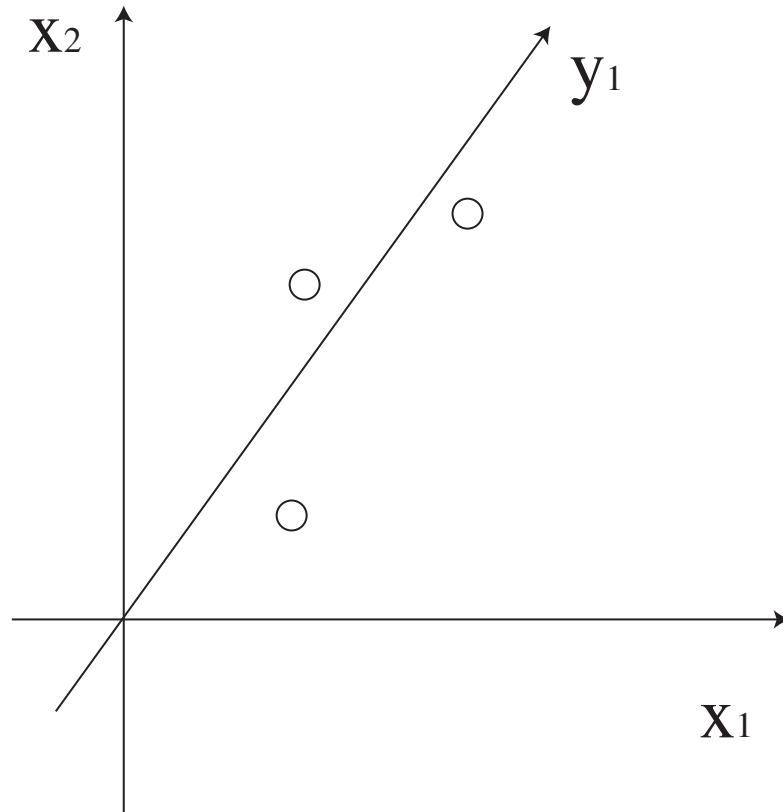


図. 2.23: KL 展開

- 共分散

まず、与えられたデータの原特徴空間における共分散行列 を求める。 C は以下の式で求めることができる。

$$C = \frac{1}{n} \sum (x_i - m)(x_i - m)^t \quad (2.16)$$

ここで m はデータの平均を、 n はデータの総数を、 x_i は i 番目のデータを表す。

- 固有行列

次に、得られた共分散行列 C の固有値ベクトル行列を求める。固有値を求めるためには行列の行数次の多項式を解かねばならないが、数百、数千といった高次の多

項式の解を求めるのは非常に困難であるため通常は近似解析手法を用いる。固有値の近似解析法にはさまざまなものがあるが、ここでは簡便な解法のひとつであるべき乗法を用いる。

べき乗法は逐次固有ベクトル u に初期値 u_0 を与え (通常は単位ベクトルを与える), 近似処理の逐次繰り返しを行って固有値ベクトル行列の近似解を求めるという手法である。べき乗法は固有ベクトルを1つずつ求める手法であるが、これを n 個づつに拡張したものが現在広く用いられている部分空間反復法であり、その基本的な考え方はべき乗法と同じである。

$$u = Au \quad (2.17)$$

を解く場合を考える。ただし u は列ベクトル, A は対称行列である。 u に初期値 u_0 を与え, 以下の式に適用する。

$$u_{k+1} = \frac{Au_k}{c(u_k)} \quad (2.18)$$

ここで, $c(u_k)$ は u_k の絶対値が最大である要素の値を返す関数である。(2.18) 式を繰り返し処理することによって $c(u_k)$ は最大固有値 λ_1 に, また $Vecu_k$ は対応する固有ベクトルに収束する。

行列 A の第2以下の固有値を求めるには,

$$A_{left} = A - \lambda_1 u_1^T u_1 \quad (2.19)$$

この式で求められる A の残差行列 A_{left} に対して同様の処理を行えばよい。

精度および処理速度を改良するための研究が続けられており, 現在ではQR法やQZ法, 修正Newton法などさまざまな手法が提案されている。

- 基底変換

最後に, 得られた固有値ベクトル行列から基底を求め, 次元縮小を行う。原特徴空間を d 次元部分空間に射影するためには, 固有値ベクトル行列から固有値の大きい順に d 個の固有ベクトルを取り出して, それを基底としてデータを部分空間に射影すればよい。

2.5 実験

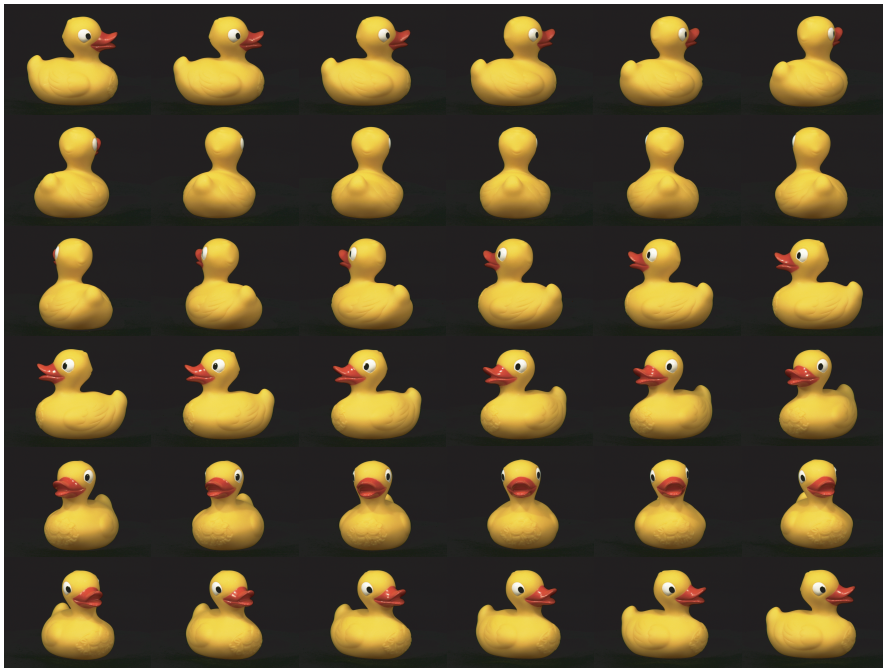


図. 2.24: 画像データ (アヒル)
5度ずつ回転したもの
実験では, この中から無作為に 20 枚を選んだ

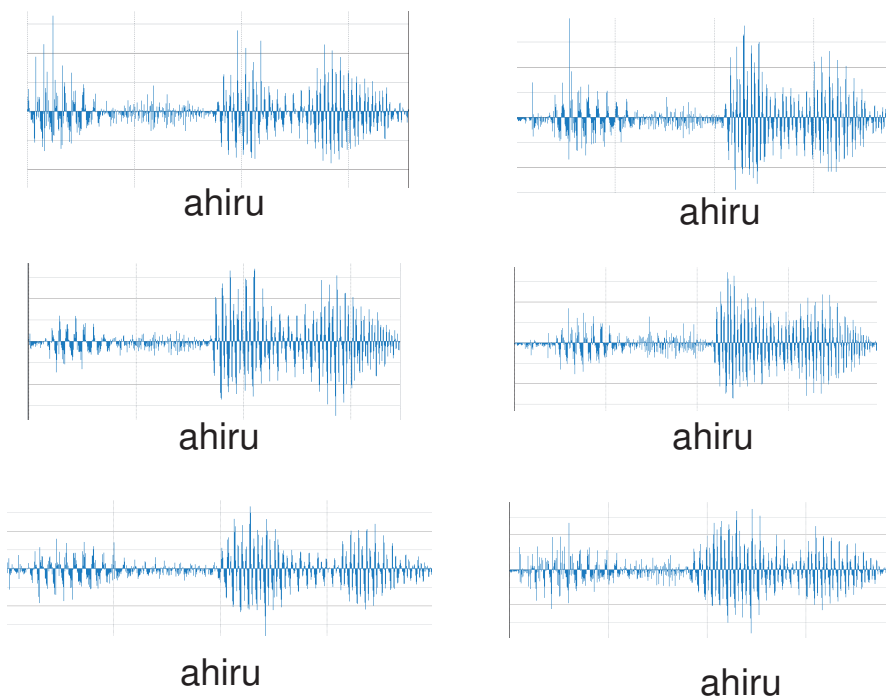


図. 2.25: 音声データ (アヒル)
実験では, 同一話者による音声を 20 回録音したものを用いた.

実データによる実験を行った。画像データは、Amsterdam Library of Object Images (ALOI) のものを用いた。図 (2.25) が使用した画像で、対称を様々な角度からとったデータを使用している。実際に使用したものは、この中から 20 枚である。音声データは、同一話者による音声 (アヒルと発話したもの) を、20 回録音したのものを用いた。サンプリング周波数は 16kHz、16bit LPC で録音した。これらのデータを、入力の 1 番 ~ 20 番とし、そのほかに 80 個のランダムなデータ (画像、音声それぞれ様々なデータ) を用意し、各 100 個のデータから特徴量を抽出した。MFCC は HTK toolkit の HCopy を用いることによって抽出した。MFCC をとる際のフレーム間は 10msec、フレーム幅は 25msec、窓関数はハミング窓を用いて、プリアンファシス係数は 0.95 とした。

まず、このデータを、普通の二次元の SOM にかけた。SOM は、素子を 100×100 平面にならべたものを用意した。学習の際のパラメータは、 $\alpha(0) = 1$ 、 $\delta(0) = 30$ 、学習回数は 200 回とした。

画像データと音声データを SOM にかけた結果は、それぞれ、図 (2.26)、(2.27) のようになる。また、そのときの相互情報量の値を計算すると、図 (2.28) のようになる。

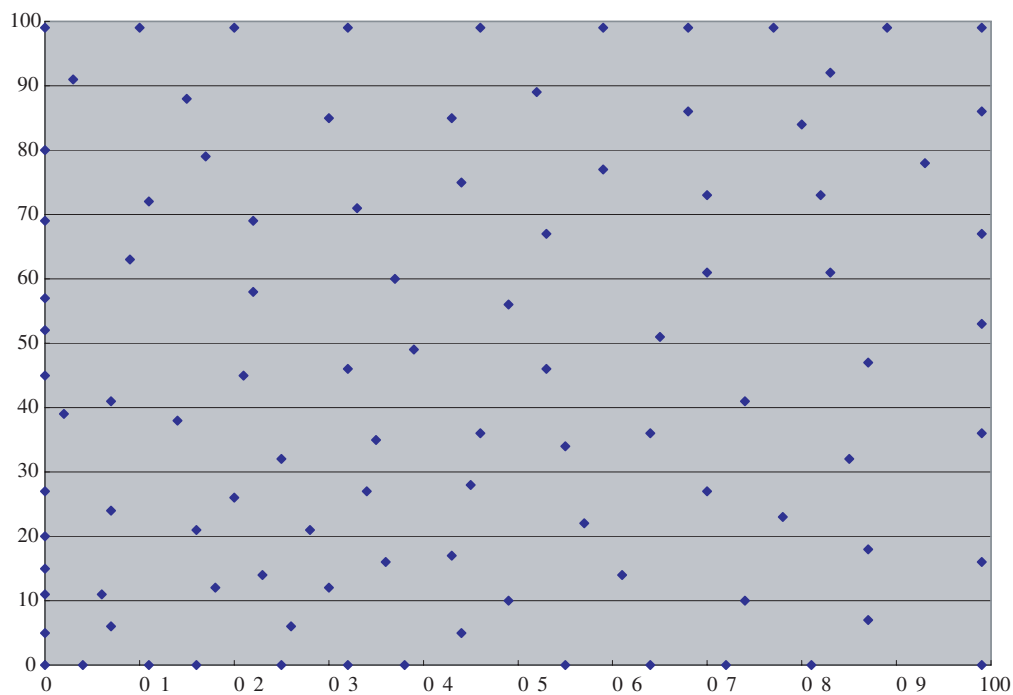


図. 2.26: 画像データを SOM にかけたもの

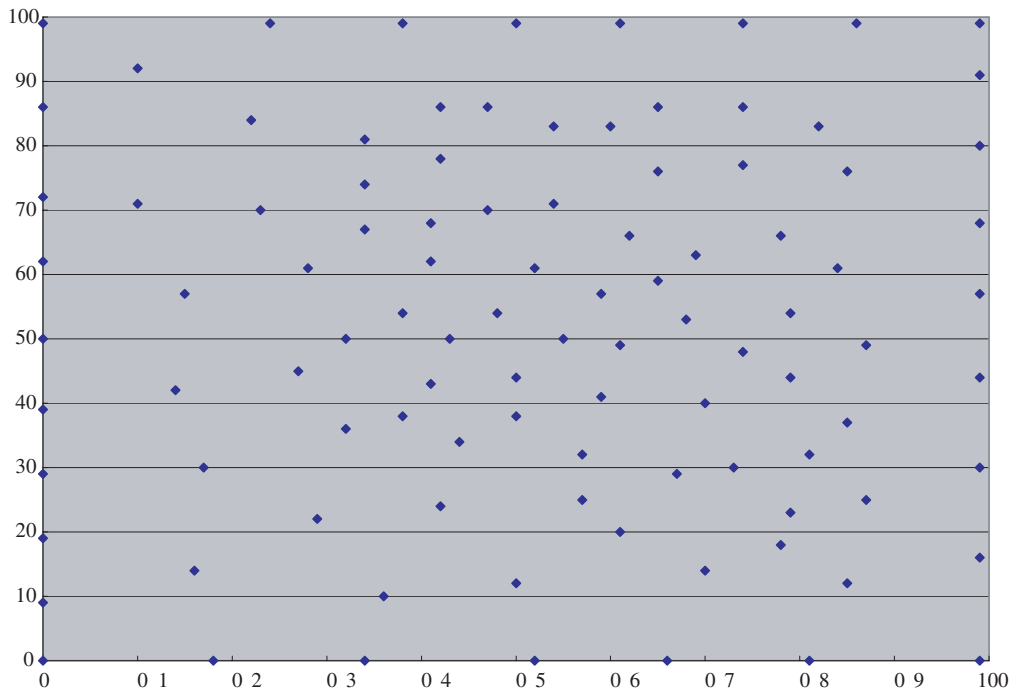


図. 2.27: 音声データを SOM にかけたもの

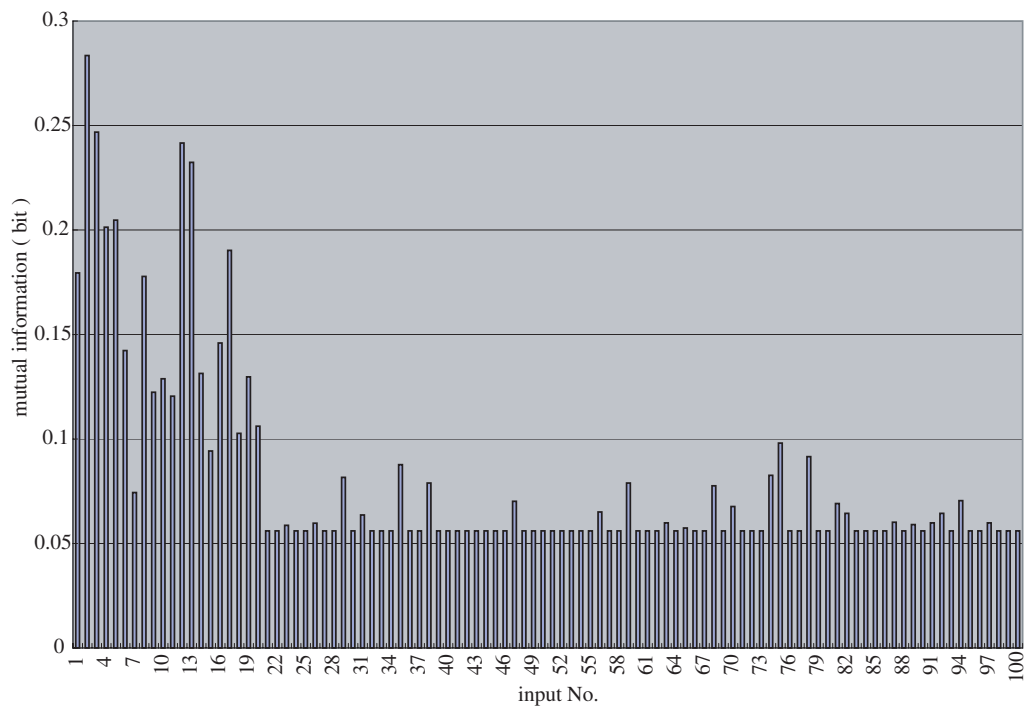


図. 2.28: mutual information

このテストデータを, 今回の提案手法にかける. SOMは, 素子を 100×100 平面にならべたものを二つ用意した. 学習の際のパラメータは, $\alpha(0) = 1, \delta(0) = 30, H = 0.2$, 学習回数は 200 回とした. 結果は, 図 (2.29),(2.30), このときの相互情報量は (2.31) のようになった.

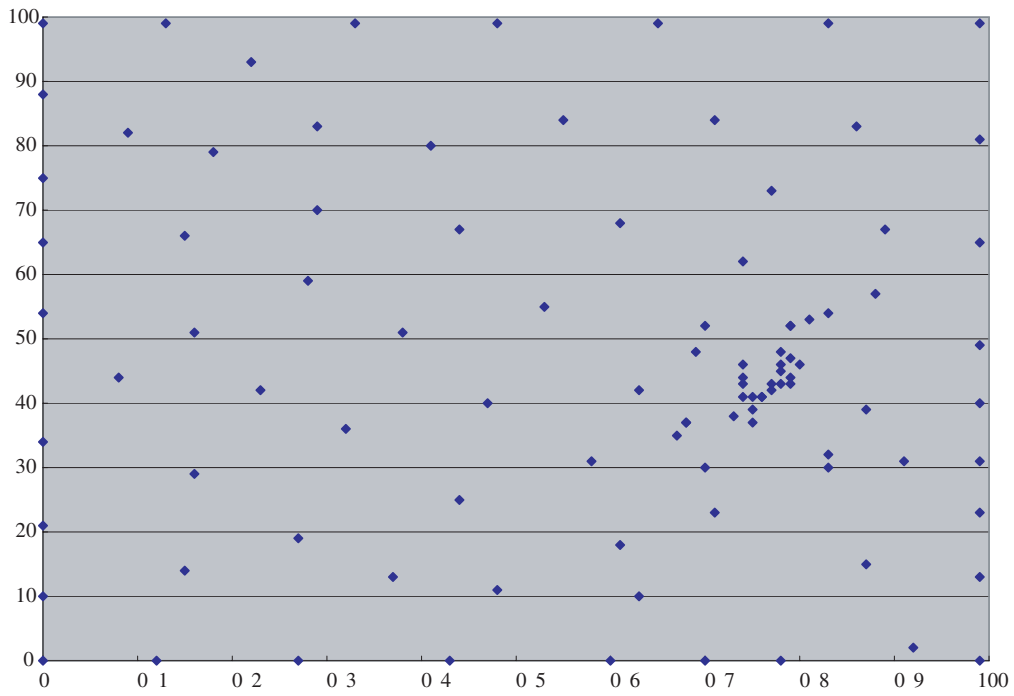


図. 2.29: 画像データを提案手法にかけたもの

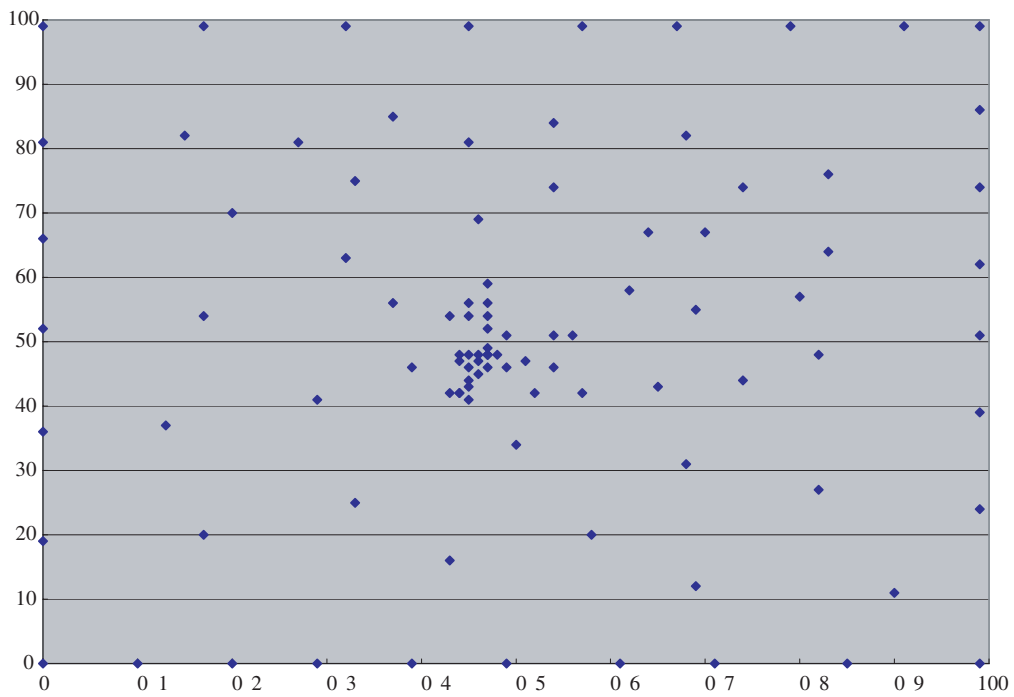
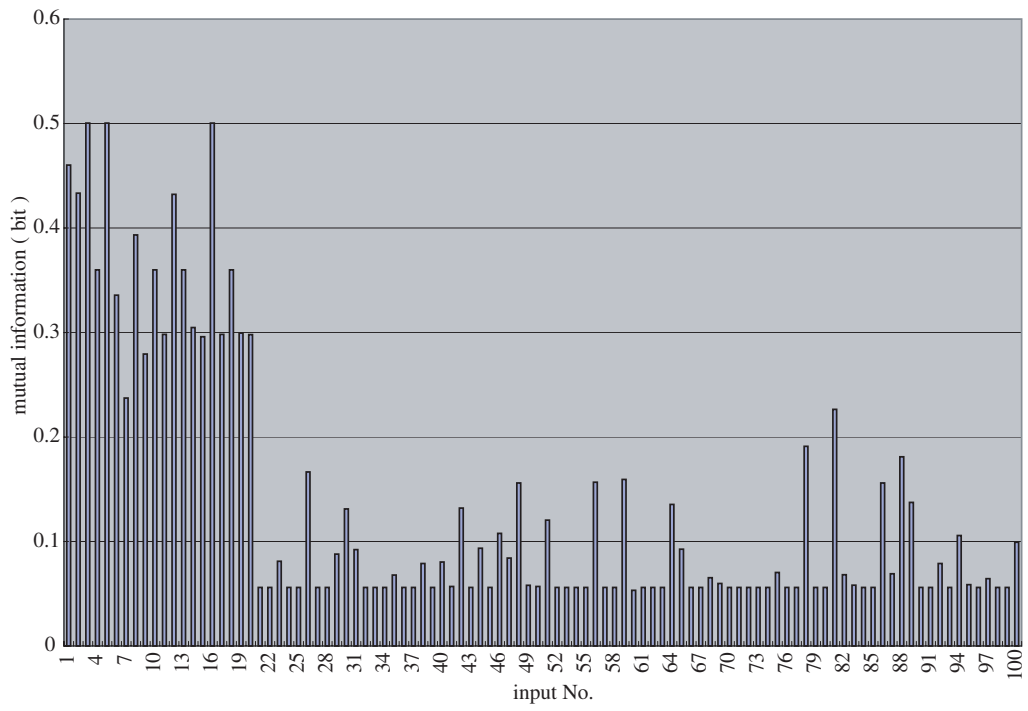


図. 2.30: 音声データを提案手法にかけたもの



☒. 2.31: mutual information

第3章 SOM of SOMs への拡張

相互情報量計算の際, 中心データからの等距離の領域の相互情報量を計算し, 相互情報量の値が最も高くなる範囲を一つのまとまりとしている. これは, 一つのまとまりに対し, 中心データがひとつだけあり, 残りがガウス分布状になっていると仮定している. しかし, 実際には対称物体が回転しているときなどは, 特徴量は必ずしも円形には分布しない(図(3.1)). そのため, 分布が非線形となり, 今回の手法ではうまくいかない場合がある. これを解決するために, SOMをSOM of SOMsへ拡張し, SOM of SOMsに今回の手法を採用することを考える.

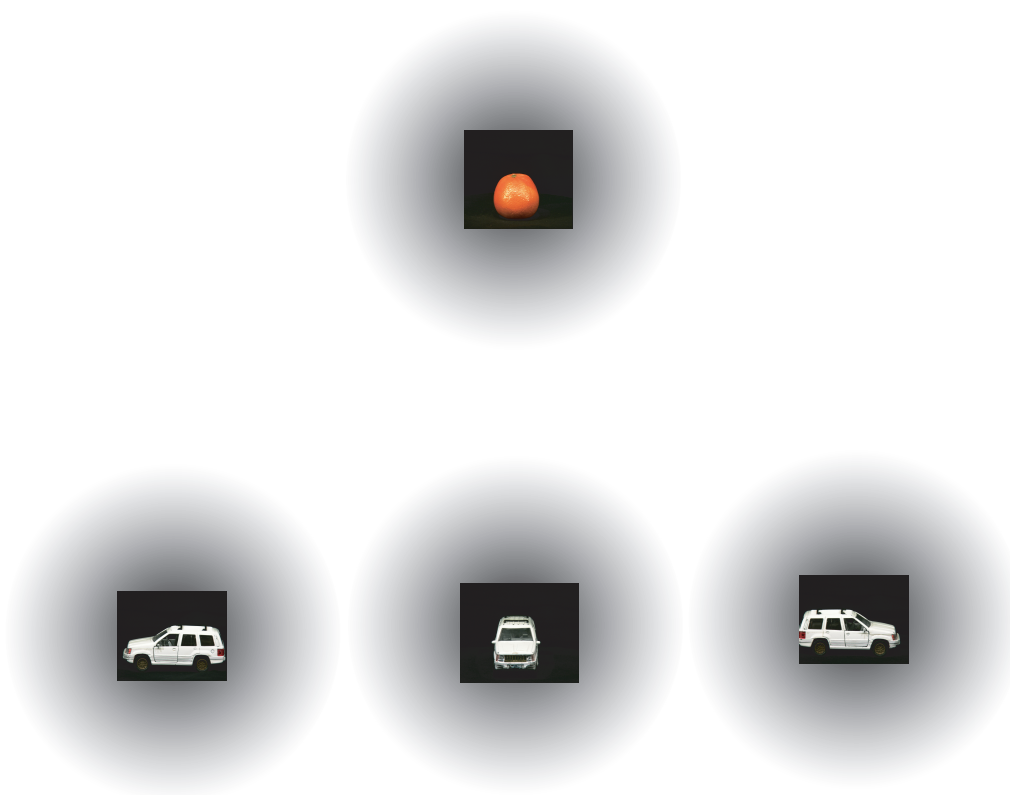


図. 3.1: 中心データが複数ある場合

中心データが一つの場合は, データは円形に分布しているが, 中心データが複数あるような場合には, 1つの円の形には分布しない.

3.1 SOM of SOMs の原理

従来の SOM では入力ベクトル x を入力単位としてマッピングを行っていたが, SOM^2 は, 入力ベクトルの集まりである, エピソード D を入力単位としてマッピングを行う [?] ~ 39.

すなわち, i 番目のエピソードは $D_i = (x_{i,1}, \dots, x_{i,J})$ とあらわされ, エピソード集合 $\{D_1, \dots, D_I\}$ が, 全入力となる.

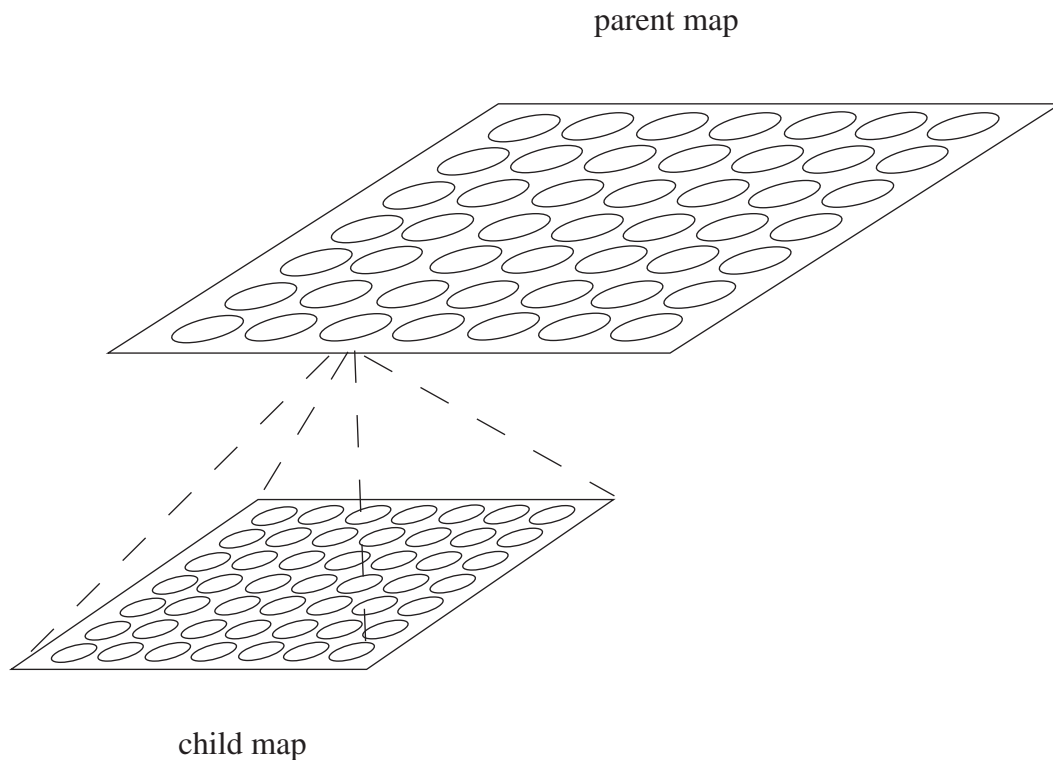


図. 3.2: SOM^2

SOM^2 は従来型の SOM (Basic SOM) が多数並んだ構造を持つ. この並んでいる SOM を, child SOM と呼ぶ. また child SOM の並びを parent map と呼ぶ. . 今, 1 個の SOM^2 が K 個の child map を持ち, 各 child map には各々 L 個の参照ベクトルがあるとするとする.

$w^{k,l}$ を k 番目の child map の l 番目の 参照ベクトルとすると, 参照ベクトルを連結して得られる連結参照ベクトル $W^k = (w^{k,1}, \dots, w^{k,L})$ は k 番目の child map 全体を表現する.

SOM^2 の目的はエピソード集合 $\{D_1, \dots, D_I\}$ を与え, 連結参照ベクトル集合 $\{W_1, \dots, W_K\}$ を自己組織的に (教師なしで) 学習することである. またこれらの他に, episode map と呼ばれる SOM を用意する. これは各 episode ごとのデータ分布を表現するための SOM である. episode map の参照ベクトルおよび連結参照

ベクトルを $V_i = (v_i^1, \dots, v_i^L)$ とする .

episode map はエピソード数 I だけあると考えれば良い . episode map はアルゴリズムを導出するために必要な概念であり , 実際の学習に際しては計算しなくて良い .

SOM^2 のアルゴリズムは次のように記述される .

勝者の決定

k 番目の child map における i 番目の エピソードの j 番目の データ $x_{i,j}$ に対する勝者は次式で定義される .

$$l^*(x_{i,j}, k) = \arg \min \|x_{i,j} - w^{k,l}\| \quad (3.1)$$

これにより量子化誤差は次のように定義される .

$$e^k(x_{i,j}) = \|x_{i,j} - w^{k,l^*(x_{i,j},k)}\|^2 \quad (3.2)$$

エピソードと child map 間の距離は , 平均量子化誤差をもって推定値とする . すなわち

$$\hat{E}^k(D_i) = \frac{1}{J} \sum_{j=1}^J e^k(x_{i,j}) \quad (3.3)$$

である . そして平均量子化誤差を最小にする child map が「勝者マップ」になる .

$$k^*(D_i) = \arg \min \hat{E}^k(D_i) \quad (3.4)$$

また真の勝者ユニットは , 勝者マップ中の勝者ユニット , すなわち

$$l^{**}(x_{i,j}) = l^*(x_{i,j}, k^*(D_i)) \quad (3.5)$$

として定義する . この勝者定義は , SOM^2 が表現するファイバー束に対して , 与えられたエピソードをもっとも良く近似する section を見つけ , かつそのエピソードの各データをもっとも良く近似するファイバーを見つけることに相当する .

episode map の推定

次に，勝者マップを元に episodemap を推定する．すなわち勝者マップの参照ベクトルを初期値とする basic SOM を用意し，エピソードのデータを学習させることでそのエピソードのデータ分布を近似する．SOM の学習回数を多く取ればそれだけ episodemap を正確に表現することになる．もっとも簡単な方法は，SOM のバッチ学習アルゴリズムを 1 回だけ実行することであり，実はこれでも十分な性能が得られる．すなわち

$$\mathbf{v}_i^l = \sum_{j=1}^J \beta_{i,j}^l \mathbf{x}_{i,j} \quad (3.6)$$

とする．ここで $\beta_{i,j}^l$ は規格化された近傍関数によって計算される学習配分率であり，

$$\beta_{i,j}^l = \frac{h_c(d_c(l, l^{**}(\mathbf{x}_{i,j})); \delta_c(T))}{\sum_{j'=1}^J h_c(d_c(l, l^{**}(\mathbf{x}_{i,j'})); \delta_c(T))} \quad (3.7)$$

で与えられる．ここで $h_c(\cdot)$ は child map レベルでの近傍関数で通常はガウス関数を用い，また $\delta_c(T)$ は近傍半径である．近傍半径は学習時間 T に従って狭くする．また $d_c(\cdot)$ は child map 上でのユニット間の距離を与える関数である．episode map の参照ベクトルが求まるので，その連結参照ベクトル \mathbf{V}_i も得られる． \mathbf{V}_i はエピソード D_i のデータ分布をベクトル化したものと見ることができる．

参照ベクトルの更新

続いて parent map レベルでの学習配分率 α_i^k を近傍関数から求める．これは次式で与えられる．

$$\alpha_i^k = \frac{h_p(d_p(l, k^*(D_i)); \delta_p(T))}{\sum_{j'=1}^J h_p(d_p(l, l^{**}(\mathbf{x}_{i,j'})); \delta_p(T))} \quad (3.8)$$

これを用いて SOM^2 の参照ベクトルは次のように更新される．

$$\mathbf{W}^k = \sum_{i=1}^I \alpha_i^k \mathbf{V}_i \quad (3.9)$$

$$\mathbf{w}^{k,l} = \sum_{i=1}^I \sum_{j=1}^J \alpha_i^k \beta_{i,j}^l \mathbf{x}_{i,j} \quad (3.10)$$

(3.10) にはもはや episode map V_i が含まれていない。したがって $\alpha_i^k, \beta_{i,j}^l$ を求めれば SOM^2 の参照ベクトルを更新することが可能になる。上記の3つのステップを、近傍半径を狭めながら繰り返し、定常状態になったところで学習を停止する。

3.2 実験

実際に SOMSOM がどのように動作するのかを実験した。図 (3.3) のようなテストデータを用意した。

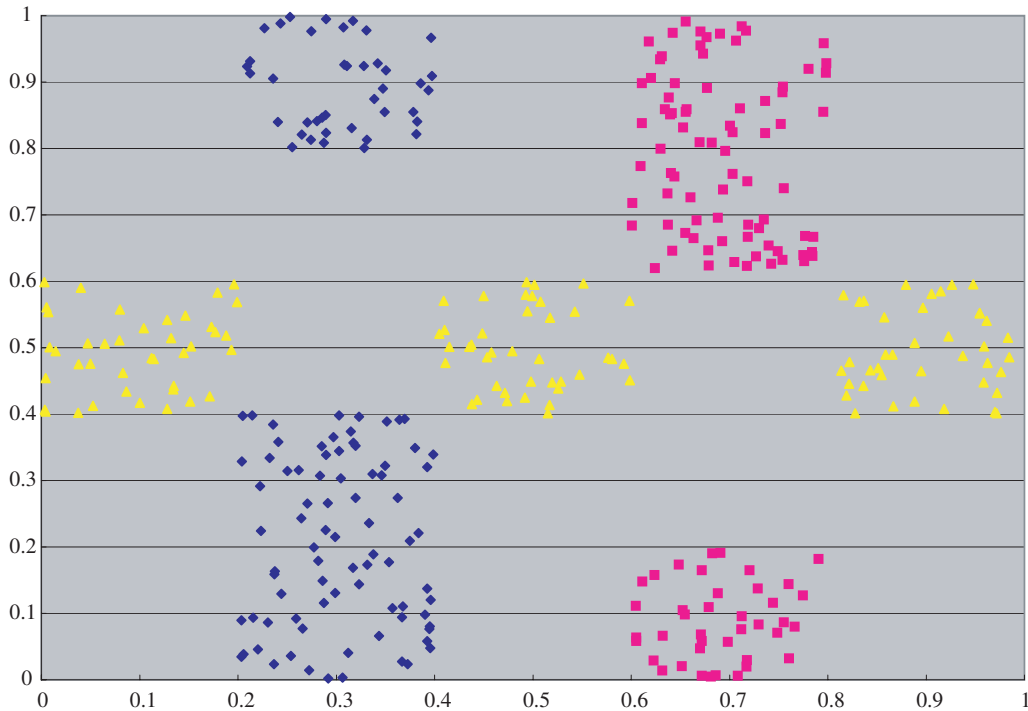


図. 3.3: テストデータ

3種類のエピソードを用意した

それぞれ40個, 全部で120のエピソードからなる

それぞれのエピソードは3つの入力データからなる

- 1種類目のエピソードは, 1つ目のデータが x が $0.2 \sim 0.4, y$ が $0 \sim 0.2$, 2つ目のデータが x が $0.2 \sim 0.4, y$ が $0.2 \sim 0.4$, 3つ目のデータが x が $0.2 \sim 0.4, y$ が $0.8 \sim 1$
- 2種類目のエピソードは, 1つ目のデータが x が $0.6 \sim 0.8, y$ が $0 \sim 0.2$, 2つ目のデータが x が $0.6 \sim 0.8, y$ が $0.6 \sim 0.8$, 3つ目のデータが x が $0.6 \sim 0.8, y$ が $0.8 \sim 1$
- 3種類目のエピソードは, 1つ目のデータが x が $0 \sim 0.2, y$ が $0.4 \sim 0.6$, 2つ目のデータが x が $0.4 \sim 0.6, y$ が $0.4 \sim 0.6$, 3つ目のデータが x が $0.8 \sim 1, y$ が $0.4 \sim 0.6$ の範囲にある

parent SOM は, 素子を 50×50 , child SOM は 5×5 個平面にならべたものを用意した。学習の際のパラメータは, $\alpha(0) = 1, \delta_p(0) = 30, \delta_c(0) = 3$, 学習回数は200回とした。

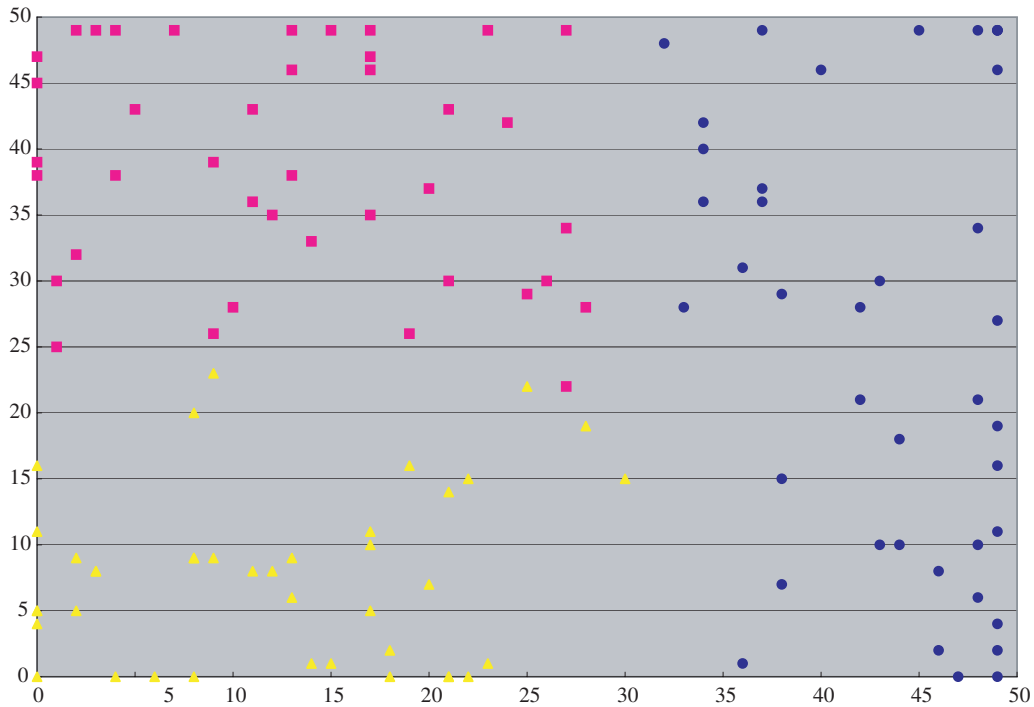


図. 3.4: SOM^2 にかけた結果

parent map のみを表示している

縦軸, 横軸はそれぞれ child SOM の座標をあらわしており, プロットした点は, 勝者マップを表す

3.3 相互情報量を用いた SOM of SOMs

SOM^2 の学習の際, SOM のときと同様に相互情報量によって学習範囲をかえていく. 相互情報量の計算は, child map に採用することもできるが, 今回は parent map のみに採用した. つまり, 学習範囲 $\delta_c(t)$ と $\delta_p(t)$ は以下の式のようなになる.

$$\delta_c(t) = \left(1 - \left(\frac{t}{TMAX}\right)\right) * \delta_c(0) \quad (3.11)$$

$$if I(A, V) > H \quad \delta_p(t) = \left(1 - \left(\frac{t}{TMAX}\right)\right) * \delta_p(0) \quad (3.12)$$

$$else \quad \delta_p(t) = \left(1 - \left(\frac{t}{TMAX}\right)\right) * \delta_p(0) \quad (3.13)$$

3.4 テストデータでの実験結果

図 (3.5)(3.6) のようなテストデータを用意して実験を行った.

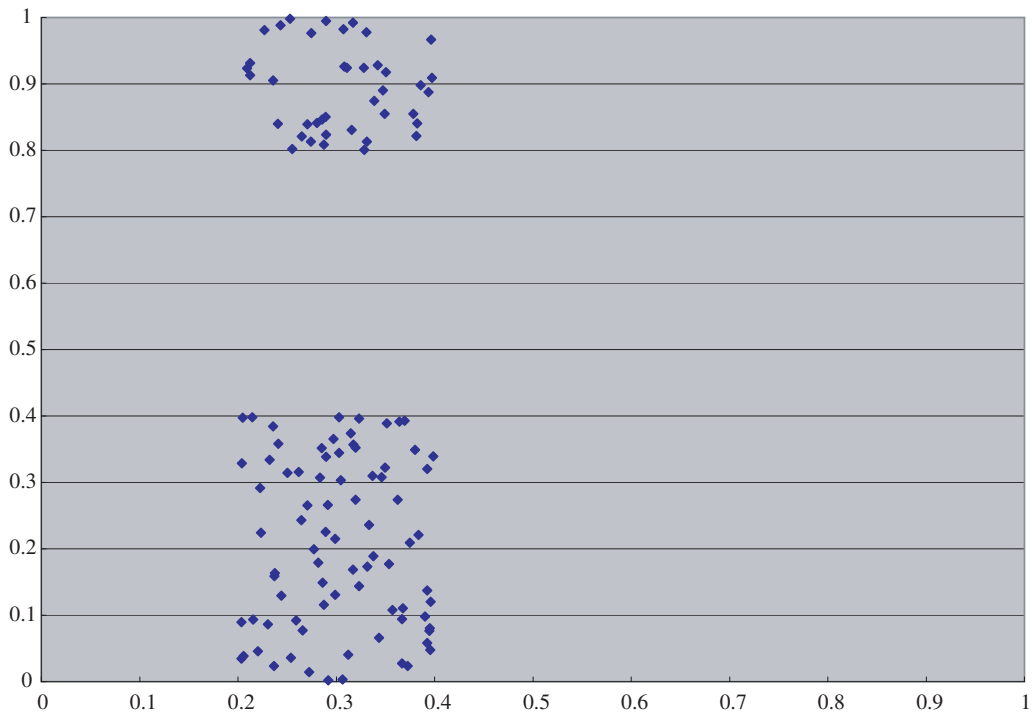


図. 3.5: テストデータ input1

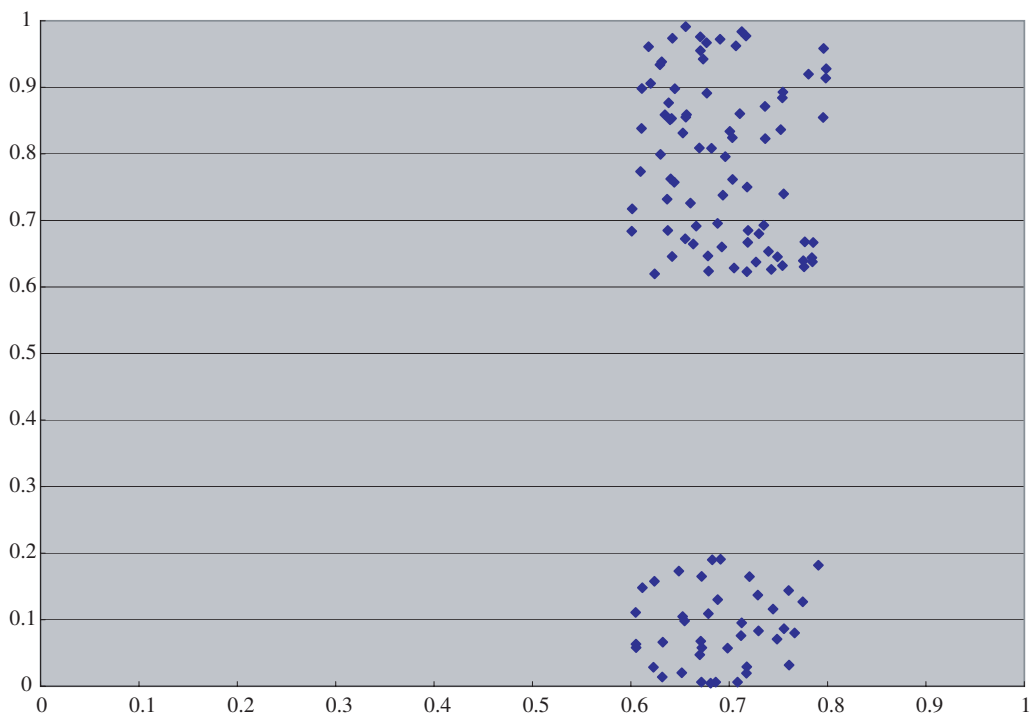


図. 3.6: テストデータ input2

parent SOM は, 素子を 50×50 , child SOM は 5×5 個平面にならべたものを用意した. 学習の際のパラメータは, $\alpha(0) = 1, \delta_p(0) = 30, \delta_c(0) = 3$, 学習回数は 200 回とした. テストデータ 1 とテストデータ 2 を SOM にかけての結果は, それぞれ, 図

(3.10),(3.11) のようになる. また, そのときの相互情報量の値を計算すると, 図 (3.9) のようになる.

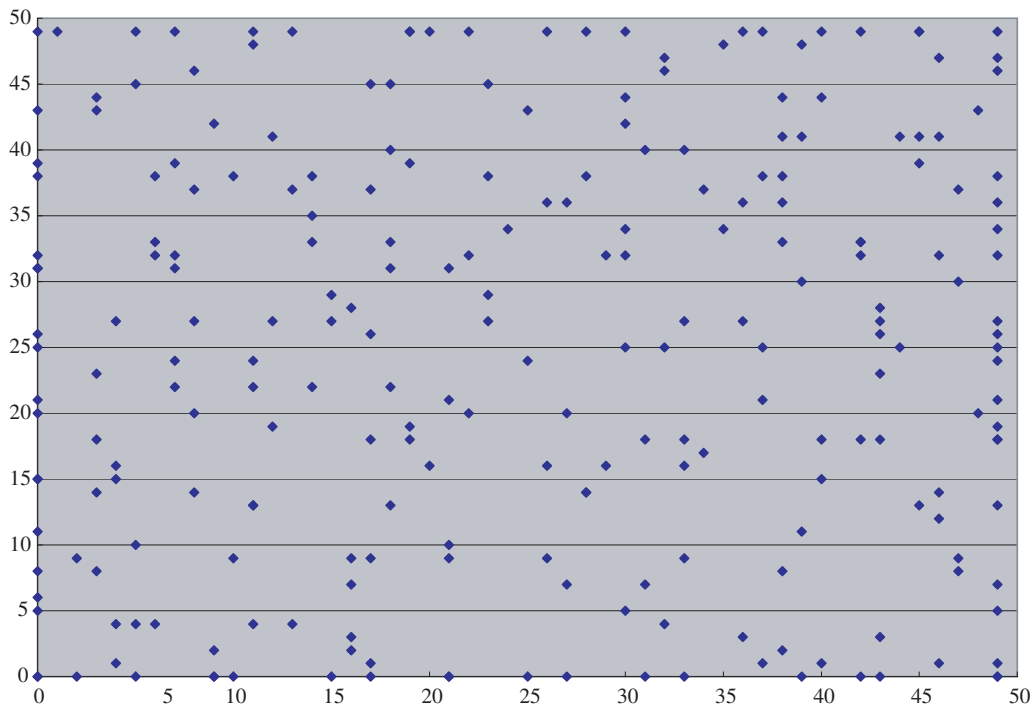


図. 3.7: SOM^2 に input1 を入れた結果

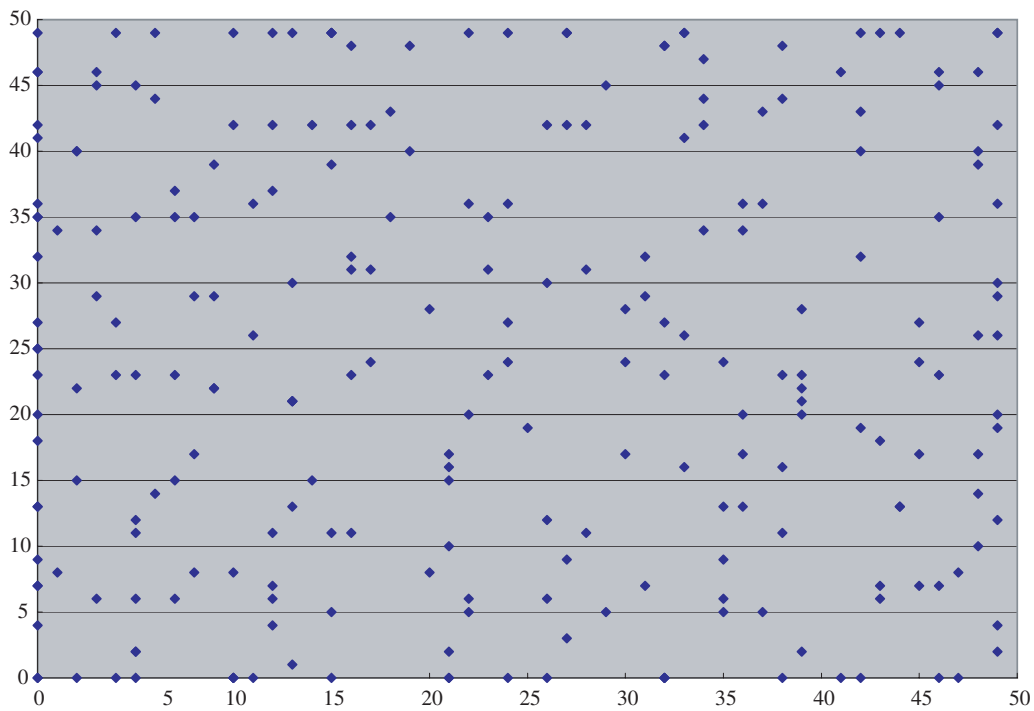


図. 3.8: SOM^2 に input2 を入れた結果

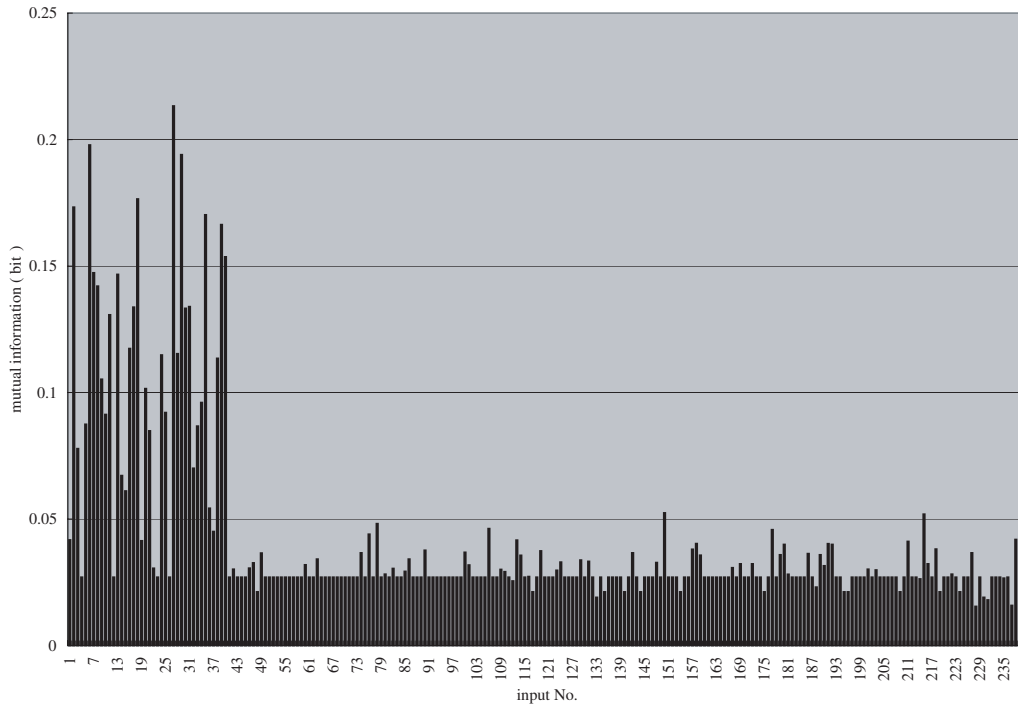


図. 3.9: mutual information

このテストデータを, 今回の提案手法にかける. parent SOM は, 素子を 50×50 , child SOM は 5×5 個平面にならべたものを用意した. 学習の際のパラメータは, $\alpha(0) = 1, \delta_p(0) = 30, \delta_c(0) = 3$, 学習回数は200回とした. 結果は, 図(3.10),(3.11), このときの相互情報量は(3.12)のようになった.

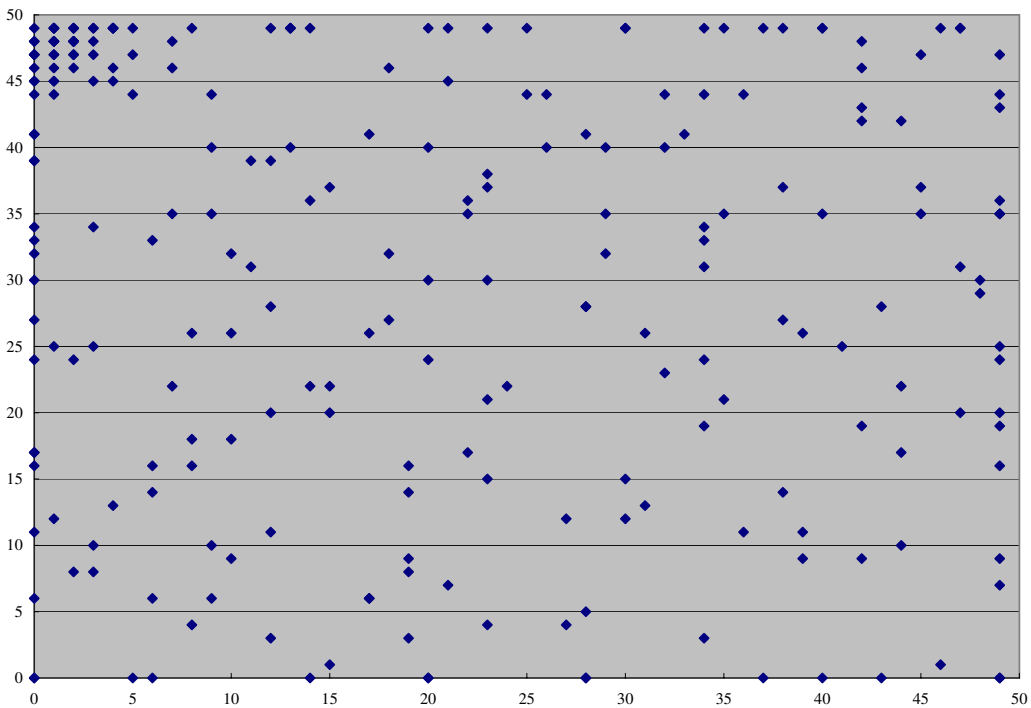


図. 3.10: 提案手法に input1 を入れた結果

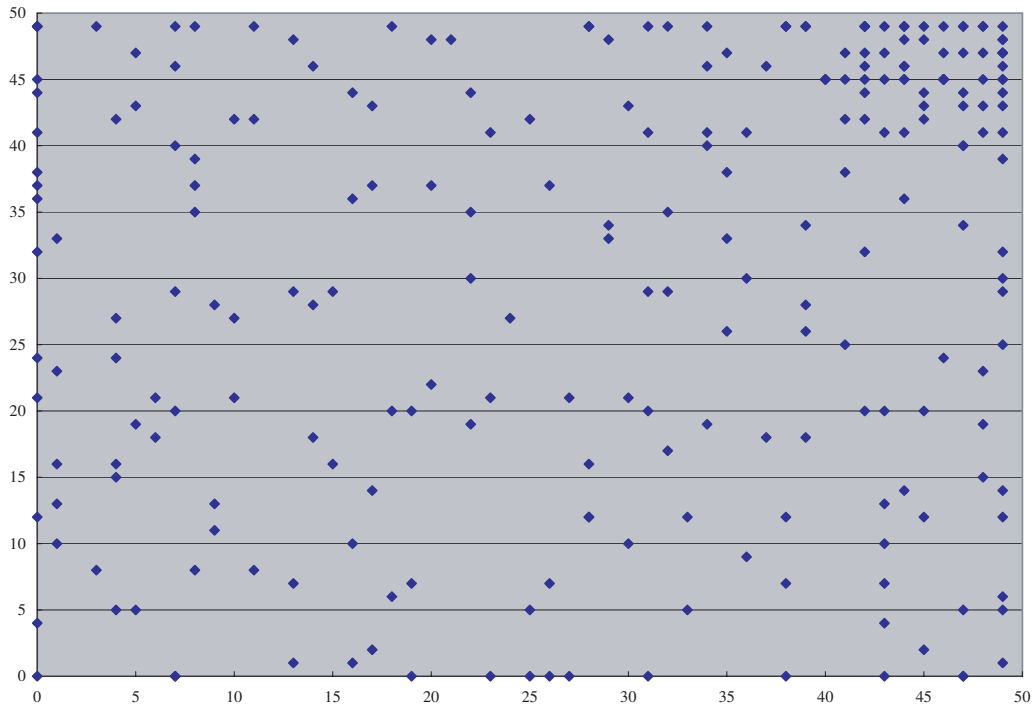


図. 3.11: 提案手法に input2 を入れた結果

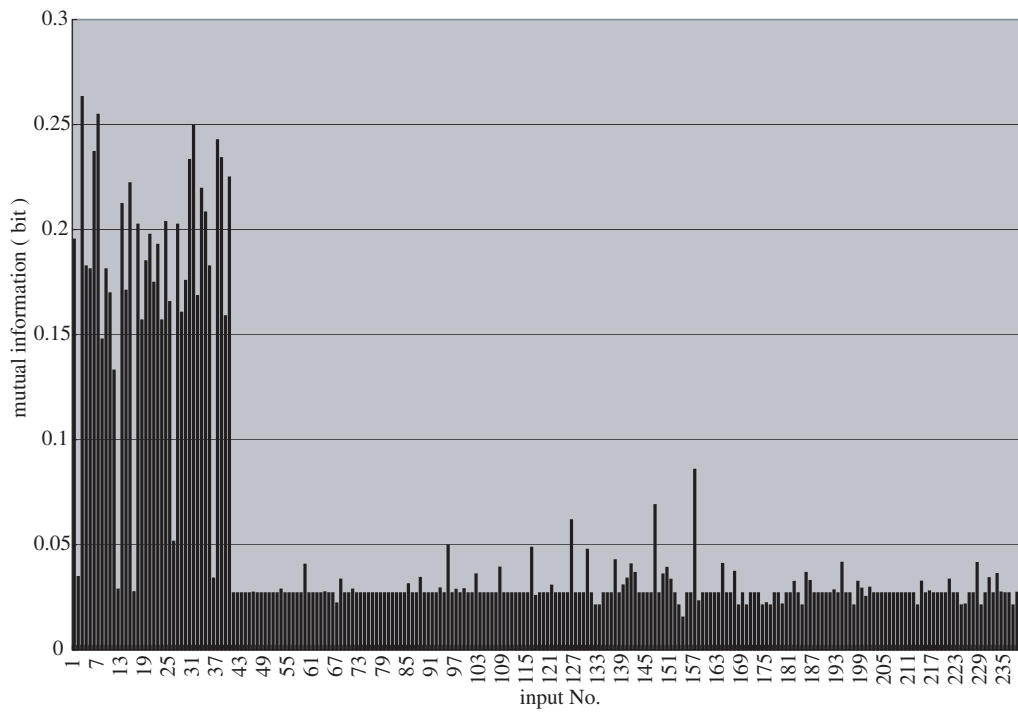


図. 3.12: mutual information

3.5 実データでの実験結果

実データで実験を行った。画像データは、図 (3.13) のように、三方向からとったデータを 1 組のエピソードとする。入力エピソードは、5 °ずつ回転させた、全 7 組を入力データとし、その他ランダムデータな画像のエピソードを 35 組用意した。音声データは、7 つの同内容の音声データ (くるま) と、ランダムな音声データのデータ 35 個を用意した。



図. 3.13: input data のエピソード

画像データは SOM^2 で、音声データは SOM でマッピングを行った。 SOM^2 のほうは、parent SOM は素子を 25×25 , child SOM は 3×3 個平面にならべたものを用意した。 SOM は素子を 25×25 用意した。 学習の際のパラメータは、 SOM^2 は $\alpha(0) = 1, \delta_p(0) = 3, \delta_c(0) = 1$, SOM は $\alpha(0) = 1, \delta(0) = 3$ 学習回数は 200 回とした。

相互情報量を用いない結果は、それぞれ、図 (3.14), (3.15) のようになる。 また、そのときの相互情報量の値を計算すると、図 (3.16) のようになる。

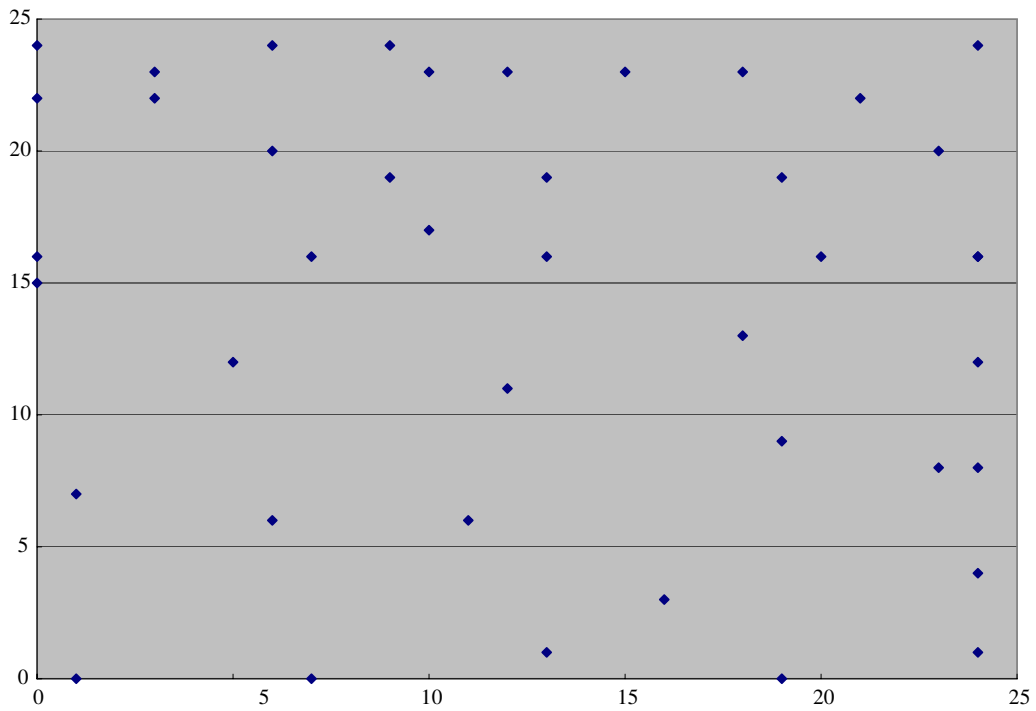


図. 3.14: 画像データを SOM^2 に入れた結果

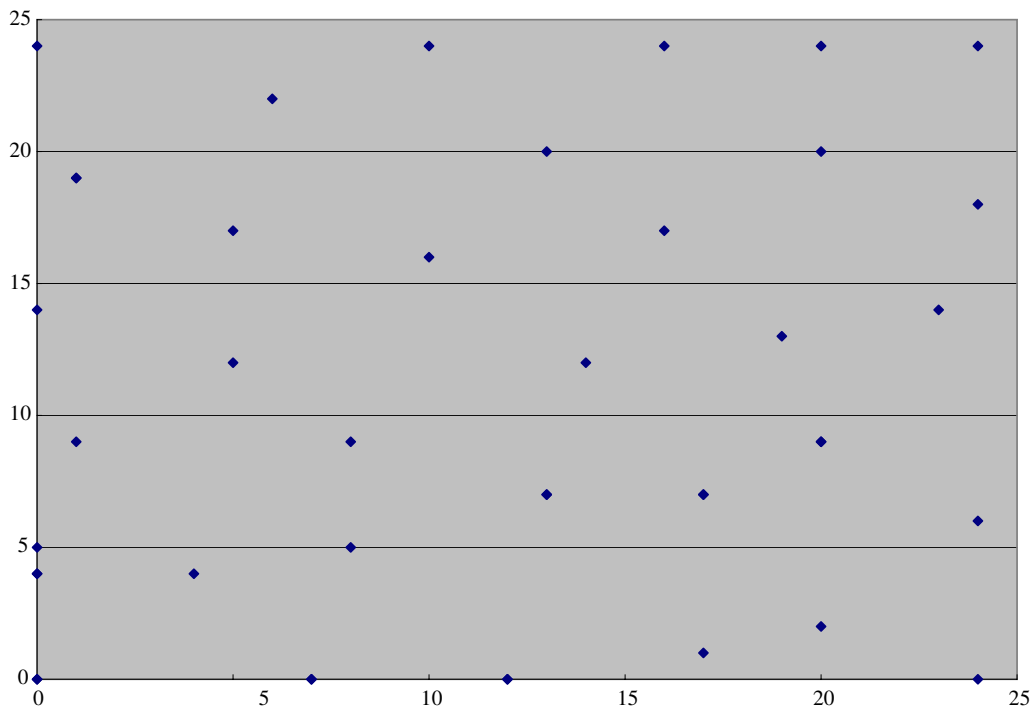


図. 3.15: 音声データを SOM^2 に入れた結果

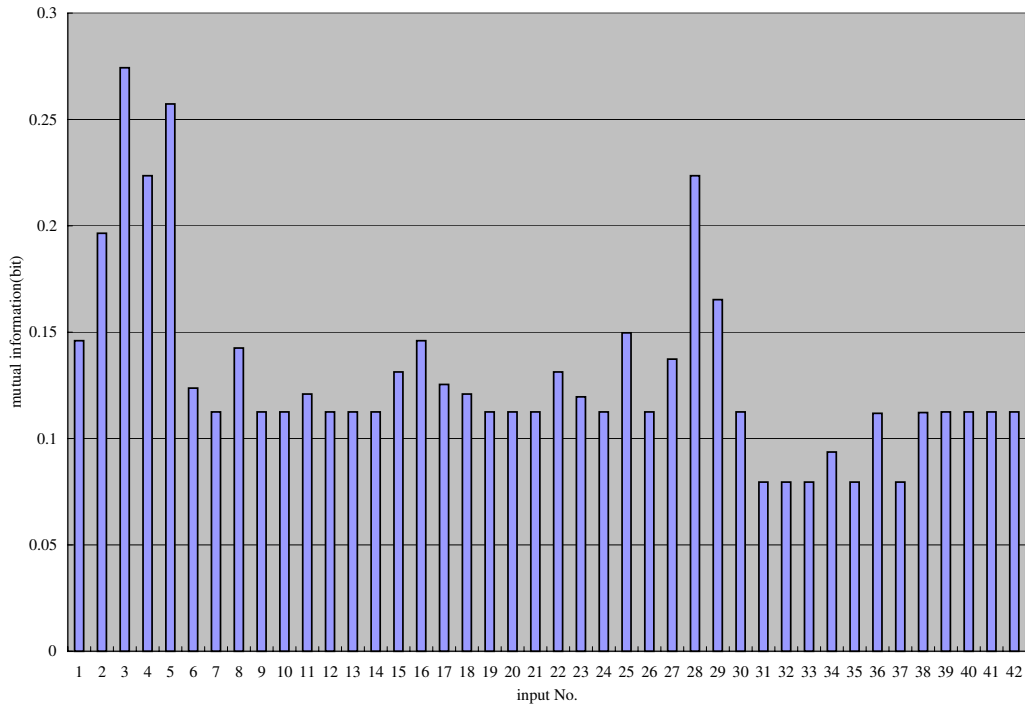


図. 3.16: mutual information

今回の提案手法にかけた. parent SOM は, 素子を 25×25 , child SOM は 5×5 個平面にならべたものを用意した. 学習の際のパラメータは, $\alpha(0) = 1$, $\delta_p(0) = 30$, $\delta_c(0) = 3$, 学習回数は 200 回とした. 結果は, 図 (3.17), (3.18), このときの相互情報量は図 (3.19) のようになった. 今回の実験では入力データ数が少ないため, 分かりづらいが, 図 (3.17) では左下に, 図 (3.18) では左上に集まってきていることが分かる. このことから, 実データでも SOM^2 に拡張した今回の提案手法は有効であると考えられる.

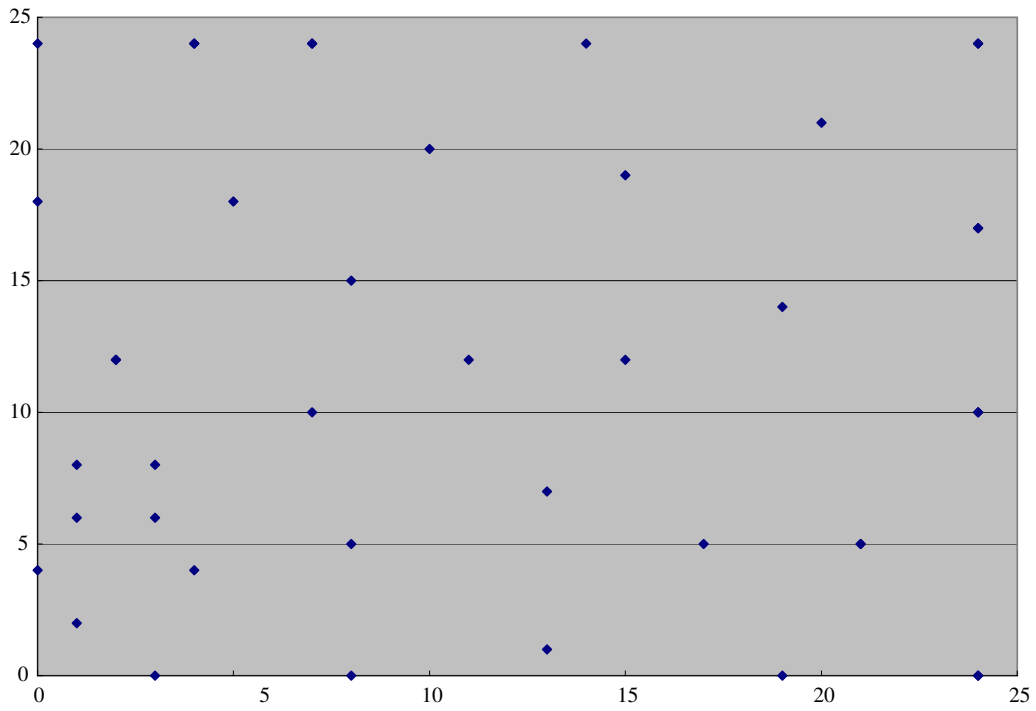


図. 3.17: 画像データの結果

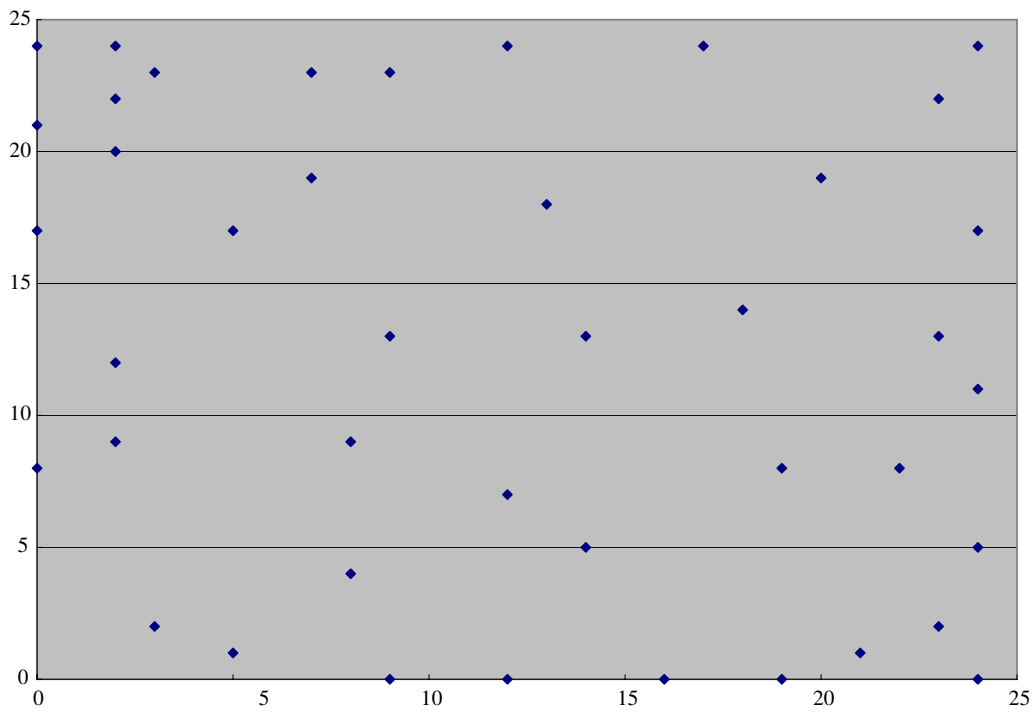


図. 3.18: 音声データの結果

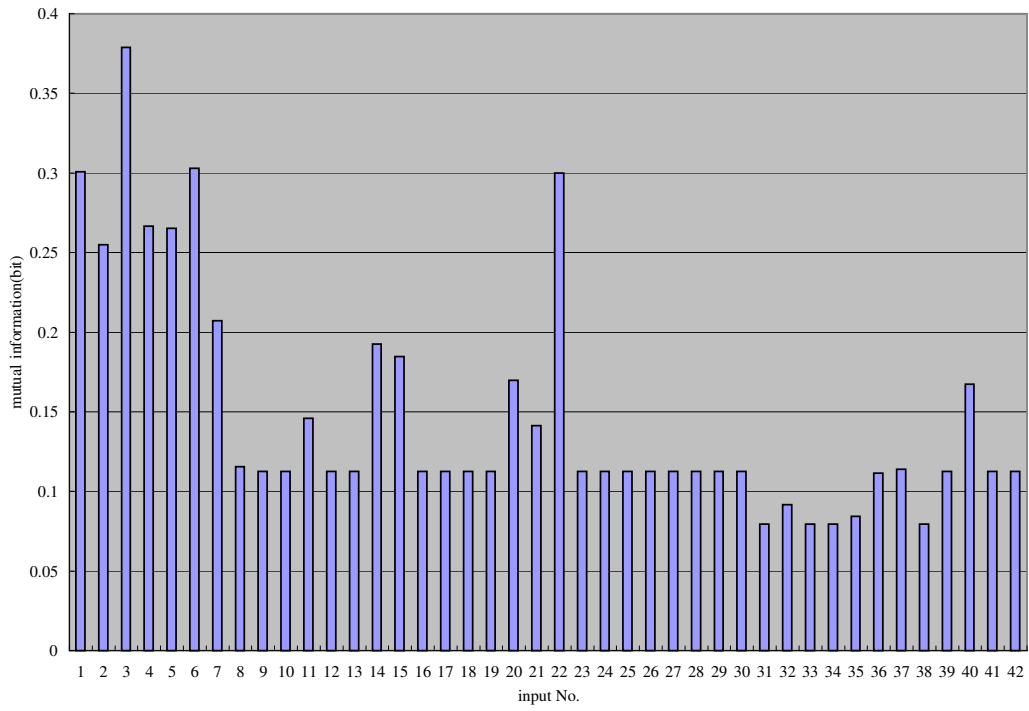


图. 3.19: mutual information

第4章 まとめ

概念構築手法として、相互情報量を用いた SOM を提案し、これにより、教師なしでの概念構築ができることを示した。また、 SOM^2 に拡張することにより非線形データにも対応できることを示した。

現在、SOM をつかってロボットの概念研究している研究は多くあるため、単純な拡張で利用できるのは利点であると考えられる。また、今回は画像と音声の2種類のみを扱ったが、モダリティはなにも画像と音声でなくともよいし、もっと多くの種類を一度に扱うことも可能である。触覚や味覚なども追加でき、その拡張は容易にできる。

また、学習する言語は何語でも良く、言語の変化にも柔軟に対応できる。

将来的に言語をしゃべる方向の研究も行わなくてはならないと考えられるが、パラメトリックな手法との相性は良いと思われる。gnuspeech や、早稲田のロボットやら、実際の口とどの形を扱えるパラメトリックな方法は、かなり実用に近いレベルで研究されている。これらのパラメータを構築していく空間としての応用が期待される。

また、SOM of SOMs に提案手法を拡張したため、この多段型の SOM の使い方を色々考えれば、複雑な対応を学習させることも可能であると考えられる。これは、複雑な対応を必要とする概念構築には不可欠な性質であると考えられる。

また、今回は主に概念構築方法として相互情報量を用いた SOM を提案したが、複数モダリティに対して幅広く応用可能である。たとえば、複数モダリティを用いているものとして、ウェブやテレビがあげられ、それらのマッピングに応用できることも期待される。

謝辞

本研究を行うにあたり、廣瀬明先生をはじめとして、廣瀬研究室の方々にはたいへんお世話になりました。ありがとうございます。

参考文献

- [1] Hideki Kozima: Infanoid: A babybot that explores the social environment, K. Dautenhahn, A. H. Bond, L. Canamero, B. Edmonds (eds.), *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, Amsterdam: Kluwer Academic Publishers, pp.157-164, 2002.
- [2] 岩橋直人 (2005). "人と機械の共有経験を基盤とする言語コミュニケーションの計算機構とシンボルグラウンディングの階層性", *人工知能学会研究会資料*, SIG-FPAI-A503-05, pp.25-32.
- [3] 岩橋直人 (2003). "言語コミュニケーションのための機械学習", 特集「機械学習, それが人に及ばざる理由」, *人工知能学会誌*, Vol.18. No.5, pp.522-525.
- [4] 岩橋直人. "ロボットによる言語獲得: 言語処理の新しいパラダイムを目指して", *人工知能学会誌*, Vol.18, No.1, pp.49-58, 2003.
- [5] D.K.Roy and A.P.Pentland "Learning words from sights and sounds : a computational model" *Cognitive Science*,26,1,pp.113-146(2002)
- [6] D.K. Roy and N. Mukherjee. Towards situated speech understanding: Visual context priming of language models. *Computer Speech and Language*, 19(2):227.248, 2005.
- [7] T. Kohonen, S. Kaski, H. Lappalainen, " Self-organized formation of various invariant-feature in the adaptivesubspace SOM, " *Neural Computation*, 9, 1321-1344, 1997
- [8] T. Kohonen, " Generalization of the Self-organizing map, " *Proc. of IJCNN93*, 457.462, 1993
- [9] S. young, G.Everman, T.Hain, D.Kershaw, G.Moore, J.Odel, D.Ollason, D.Povey, V.Valtchev and P.Woodland: "The HTK Book", Cambridge University Engineering Department 1995.
- [10] T. Furukawa, " SOM of SOMs : Self-Organizing map which maps a group of self-organizing maps, " *Lecture Notes in Computer Science*, 3696, 391-396, 2005
- [11] T. Furukawa, " SOM2 as ' SOM of SOMs ', " *Proc. of WSOM05*, 41-48, 2005

- [12] T. Furukawa, “ SOM of SOMs: An Extension of SOM from ‘ Map ’ to ‘ Homotopy ’, ” Proc. of ICONIP2006
- [13] N.Otsu, and T.Kurita, “A new scheme for practical, flexible and intelligent vision systems,” Proc. IAPR Workshop on Computer Vision, pp.431-435, 1988
- [14] 赤穂, 速水, 吉村, 麻生 ”EM 法を用いた複数情報源からの概念獲得” 電子情報通信学会論文誌, J80A, 9, pp.1546-1553(1997)
- [15] J.Hershey and J.Movellan ”Audio-vision : Using audio-visual synchrony to locate sounds” Advances in Neural Information Processing Systems 12, MIT Press(2000)
- [16] M.Slaney and M.Covell ”Faceync : A linear operator for measuring synchronization of video facial images and audio tracks” Proceedings of Neural Information Processing Society 13 (2001)
- [17] D.Li, N.dimitrova, M.Li and I.Sethi ”Multimedia content processing through cross-modal association” Proceedings of 11th ACM International Conference on Multimedia, pp.604-611(2003)
- [18] J.Fisher, III and T.Darrel ”Speaker association with signal-level audiovisual fusion” IEEE Transaction on Multimedia, 6, 3, pp.406-413(2004)
- [19] D.H.Ballard and C.Yu ”A multimodal learning interface for grounding spoken language in sensory perceptions” Proceedings of the 5th international conference on Multimodal interfaces pp.164-171(2003)
- [20] Rizzolatti, G., Fadiga, L., Fogassi, L., Gallese, V. Premotor cortex and the recognition of motor actions. Cognitive Brain Research 3, 131-141, 1996.
- [21] V. S. Ramachandran and Sandra Blakeslee, Phantoms in the Brain: Probing the Mysteries of the Human Mind, 1998, William Morrow & Co.; 1999, Harper Perennial, ISBN 0688172172
- [22] 正高 信夫 ” 0 歳児が言葉を獲得するとき ~ 行動学からのアプローチ ~ ” 中公新書 (1993)
- [23] David Lee 著, 宮浦 国江 訳, ” 事例で学ぶ認知言語学 ” 大修館書店 (2006)
- [24] D. Hakkani-Tur and G. Riccardi. A general algorithm for word graph matrix decomposition. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003), 2003.
- [25] D. Klein and C. D. Manning. Accurate unlexicalized parsing. In Proceedings of the 41st Meeting of the Association of Computational Linguistics, 2003.

- [26] L. Mangu, E. Brill, and A. Stolcke. Finding consensus among words: Lattice-based word error minimization. In Proceedings of EUROSPEECH '99, volume 1, pages 495-498, Budapest, 1999.
- [27] S. Narayanan. KARMA: Knowledge-based Action Representations for Metaphor and Aspect. PhD thesis, University of California, Berkeley, 1997.
- [28] D. V. Pynadath and M. P. Wellman. Probabilistic state-dependent grammars for plan recognition. In Proceedings of the Conference on Uncertainty in Artificial Intelligence, UAI2000. Morgan Kaufmann Publishers, 2000.
- [29] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 77:257-286, February 1989.
- [30] W. Schuler. Using model-theoretic semantic interpretation to guide statistical parsing and word recognition in a spoken language interface. In Proceedings of the Association for Computational Linguistics, 2003.
- [31] A. Stolcke. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. Computational Linguistics, 21(2):165-201, 1995.
- [32] J. B. Tenenbaum and F. Xu. Word learning as bayesian inference. In Proceedings of the 22nd Annual Conference of the Cognitive Science Society, 2000.
- [33] N. Yoshida. Utterance segmentation for spontaneous speech recognition. Master's thesis, Massachusetts Institute of Technology, 2002.
- [34] 小林洋美, 幸島司郎 "コミュニケーション装置としてのヒトの目の進化" 電子情報通信学会誌 Vol.82 No.6 pp.601-603
- [35] C. Moore and P. J. Dunham, "Joint Attention", Its Origins and Role in Development, Lawrence Erlbaum Associates (1995)
- [36] Yukie Nagai, Koh Hosoda, Akio Morita, and Minoru Asada. "A constructive model for the development of joint attention." Connection Science, Vol. 15, No. 4, pp. 211-229, December 2003.
- [37] 長井志江, 細田耕, 森田章生, 浅田稔. "視覚注視と自己評価型学習の機能に基づくブートストラップ学習を通じた共同注意の創発." 人工知能学会論文誌, Vol. 19, No. 1, pp. 10-19, 2004.
- [38] A. Snyder, H. Bahramali, T. Hawker and D. J. Mitchell "Savant-like numerosity skills revealed in normal people by magnetic pulses" Perception, vol. 35, no. 6 (2006) pp. 837-845
- [39] Kumashiro M, Ishibashi H, Uchiyama Y, Itakura S, Murata A, Iriki A.: Natural imitation induced by joint attention in Japanese monkeys. Int J Psychophysiol. 2003;50(1-2):81-99.

- [40] T. Nitta, "Feature Extraction for Speech Recognition Based on Orthogonal Acoustic- feature Planes and LDA," Proceedings of IEEE ICASSP'1999, pp.421-424, May 1999.
- [41] 新田恒雄, 井上雄, 正井康之, 松浦博, "複合音響特徴平面に基づく音声認識のための局所特徴抽出法," 電子情報通信学会論文誌, D-II, Vol. J83, No.11, pp. 2341-2349, 2000.
- [42] ジュリアン・ポール キーナン うぬぼれる脳 「鏡のなかの顔」と自己意識 日本放送出版協会
- [43] V・S・ラマチャンドラン, サンドラ・ブレイクスリー (山下篤子訳) 脳のなかの幽霊 1999, 角川書店, ISBN 4047913200
- [44] 小林洋美, 幸島司郎 コミュニケーション装置としてのヒトの目の進化 電子情報通信学会誌 Vol.82 No.6 pp.601-603
- [45] Kumashiro M, Ishibashi H, Uchiyama Y, Itakura S, Murata A, Iriki A.: Natural imitation induced by joint attention in Japanese monkeys. *Int J Psychophysiol.* 2003;50(1-2):81-99.
- [46] 小倉 崇, 岡田 慧, 稲葉 雅幸: 注目点を持つ幾何モデルを利用したヒューマノイドの道具利用動作の生成法, in 第23回日本ロボット学会学術講演会講演論文集, 1F15, 2005.
- [47] 森田章生, 吉川雄一郎, 細田耕, 長井志江, 浅田稔. 自己組織化マップによる共同注意の学習速度の向上. 第21回日本ロボット学会学術講演会講演論文集, 3B34, 2003.
- [48] 長井志江, 細田耕, 森田章生, 浅田稔. 視覚注視と自己評価型学習の機能に基づくブートストラップ学習を通じた共同注意の創発. *人工知能学会論文誌*, Vol. 19, No. 1, pp. 10-19, 2004.
- [49] 森田章生, 長井志江, 細田耕, 浅田稔. 視覚フィードバックと自己評価による共同注意の段階的獲得 ロボティクス・メカトロニクス講演会 '03 予稿集, Vol.CD-ROM, 2P1-3F-C1, 2003.

発表文献

国際会議

1. K.Kitahara A.Hirose, "Concept Generation Based on Mutual Information between Multiple Self-Organising Maps," IEEE Transactions on Neural Networks, to be submitted.

国内発表

2. 北原邦紀, 廣瀬明, 「相互情報量で修飾された自己組織化マップによる概念構築」, 電子情報通信学会 ニューロコンピューティング研究会, 発表予定