

階層化された砂時計型ニューラルネットによる自律的な内部表現獲得

Hierarchical Information Integration Model using Unsupervised Autoencoder Neural Networks

山川 宏*
Hiroshi YAMAKAWA

木本 隆**
Takashi KIMOTO

柳沼 義典**
Yoshinori YAGINUMA

*RWC Pつくば研究センタ
Real World Computing Partnership

**¹(株)富士通研究所
FUJITSU LABORATORIES LTD

Abstract – For acquiring abstract internal representation from real world data, gradual learning and hierarchical architecture are essential. We propose hierarchical information integration model using auto-encode neural networks, which can learn high-level internal representation form raw sensor data. We demonstrate that the hierarchical model, which takes two camera images of one target from different view, learns to restore three dimensional potion of that target.

1.はじめに

人間は認識・行動において様々な内部表現を利用しながら情報を段階的に統合する。これは現実世界と直接相互作用する際に取り扱う多様かつ大量の感覚情報や自身の運動情報などに対処する方策であると考えられる。本報告書ではこの内部表現を階層的に自動獲得する方法を提案するが、これに先立ち、内部表現とその自動獲得さらにその階層性の必要性について述べる。

内部表現の重要な役割は、その時に対処している課題に関する情報のみを適切な形で表現することである。もし、コンパクトな内部表現を使わずに解決処理を行うと、課題と無関係な情報を同時に取り扱うことになるため、しばしば考慮すべき組み合わせが膨大になり現実的な速度で実行できない。

内部表現を獲得するには、外部からの設計とシステム自身による学習がある。人間の場合には遺伝という情報伝達手段に制約されるため、多くの内部表現を後天的に獲得せざるを得ないが、人工的なシステムにおいては内部表現を望むだけ設計できるので(設計者の作業量は増大するが)、自動獲得の能力はそれほど重要でなく思える。しかし、設計の段階で将来システムが必要とする全ての内部表現を用意することは不可能なので、システムが外界に対して高度に柔軟な適応能力を持つためには、人工的なシステムにおいても内部表現の自動獲得能力を欠くことはできない。

内部表現を階層的に積み上げる利点として、獲得済の内部表現の流用による学習コストの削減が揚げられる。例えば、すでに"カップ"と"グラス"を区別できる形に関する内部表現と色を区別できる内部表現を持つシステムが、"赤いカップ"と"青いグラス"を区別する内部表現を生成するケースでは、新たにセンサ情報から内部表現を直接生成するよりも、前記二つの内部表現からの出力を統

合した表現を学習する方が容易である。さらに、この様に階層化を行うと、同じような処理システムを複数作ることを避けられるので、処理資源の点からも効率的である。

そこで本報告書では、内部表現を自動的に獲得する能力を有する砂時計型のニューラルネットワークを階層化することでセンサ情報(低次)から内部表現(高次)を自律的に獲得する情報統合モデルを提案する。

2. 砂時計型ニューラルネットワークモデルと階層化

2.1 砂時計型ニューラルネットワークモデル

砂時計型ニューラルネットワークモデル(以下、砂時計ネットワークと呼ぶ)は、入力層と出力層のユニットの数が同じで、しかも最もユニットの数が少ない中間層(圧縮層と呼ぶ)のユニットの数が入出力層のユニット数よりも小さい3層以上のニューラルネットワークである(図1)。

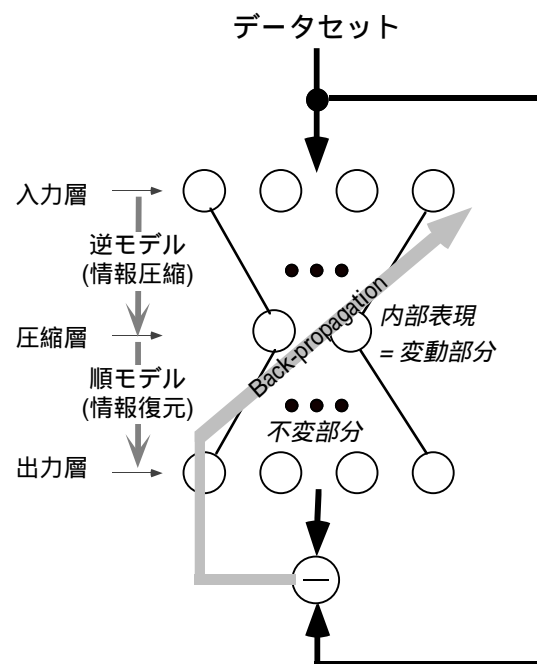


図1 砂時計ネットワークと学習

恒等写像を誤差逆伝搬法で学習したネットワークでは、圧縮層にデータセットの変動部分に応じた情報をマップする内部表現空間が形成され、入力層-圧縮層間に逆モデルが、圧縮層-出力層間にデータセットの不変部分に関する情報をもつ順モデルが、形成される。

砂時計ネットワークの優れた特徴として、少なくとも以下の3つが上げられる。

(1)内部表現の自動獲得と補間機能：

しばしば認識では、変化する部分が重要な意味を持つから、自動獲得される内部表現が変動部分であることは有用である。また、同じデータセットに対して獲得される内部表現には任意性があるが、空間の連続性は保たれる。そして、入出力層、圧縮層ともに補間機能を持つ。

(2)逆モデルの生成が容易：

ある入力に対して一意に出力を決定できない写像(一対多)を要求する内部状態の設計に起因する、学習困難な逆モデルの設計を避けられる。

(3)教師なし学習：

学習は入力層と出力層に同じデータセットを与える恒等写像学習なので、教師が不必要であり自律的システム構築に有利。

砂時計ネットワークの応用例としては、画像圧縮に用いた先駆的な研究[1]や、実行時の圧縮層の出力を特徴量として利用して画像情報から人の性別・表情などの識別する例[2,3]や、実行時の入力層と出力層の誤差量から入力データが学習データセットに含まれるか否か識別する研究[4,5]がある。さらに、圧縮層の最小次元数を自動的に探し出す過負荷学習法[6]や、抜けがある不完全なデータセットにiterative inversion法を組み合わせることで対処する研究[7]もある。

2.2. 階層化された砂時計ネットワーク

階層的に高次の内部表現を獲得するために、各砂時計ネットワークを折り曲げて圧縮層の出力を上位側の入出力層に連結する階層モデルを提案する(図2)。

このモデルでは、各砂時計ネットワークが自律的に学習可能なので、下位の砂時計ネットワークから段階的に上位のネットワークを学習させることで、システム自身により次第に高次の内部表現を獲得できる。

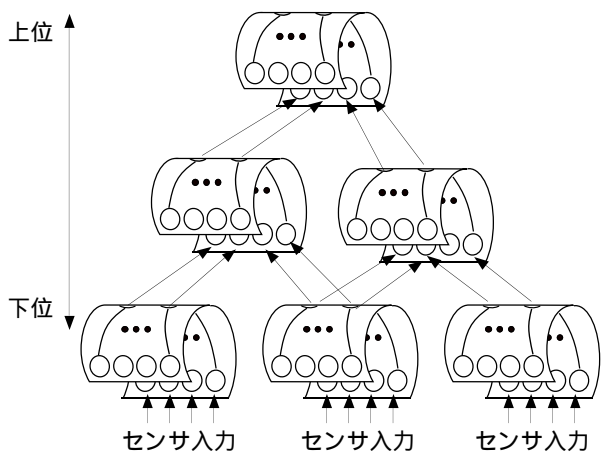


図2 階層化された砂時計ネットワーク

3. 計算機シミュレーション

3.1 課題

提案したモデルの動作確認のために、2台の固定カメラから得られる画像情報から対象物体の三次元情報を復元するシステムを自律的に形成するという、簡単なシミュレーションを行う。

まず、学習後の階層的な情報復元を説明する(図3)。立方体の中を蜂のように飛び回る対象物体を二方向view1、view2から観測し画像1と画像2を得る(数値実験とカメラの位置が違う)。下位ネットワーク1は画像1から対象物の位置(X,Y)を本質的に含む情報を抽出し、下位ネットワーク2についても同様である。上位ネットワークは二つの下位ネットワークからの情報を統合して本質的に位置(X,Y,Z)を含む情報を抽出する。

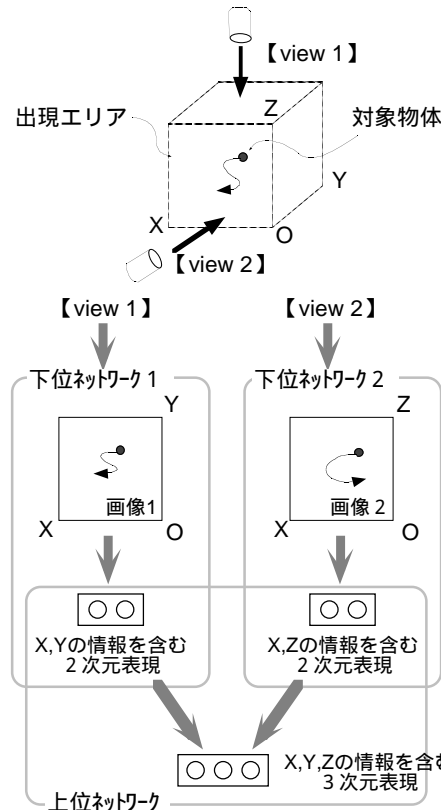


図3 シミュレーションの原理

学習時にはまず、画像を各下位ネットワークに与えて恒等写像の教師無し学習を行い、次に上位ネットワークは下位ネットワークの出力を利用して学習を行う。結果として、2つの画像情報を統合して対象物体の3次元位置を復元する機構を自律的に獲得する。

シミュレーションでは、原点からx軸沿いに左右に6離

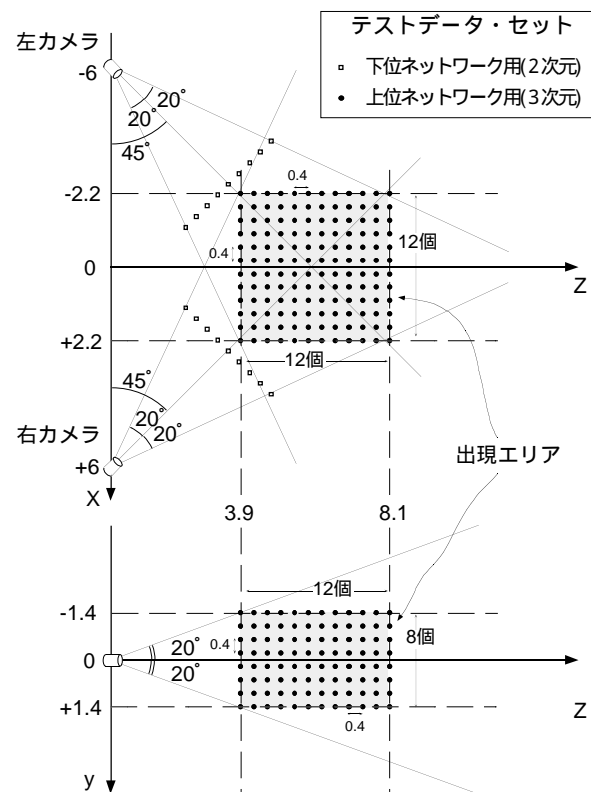


図4 カメラと対象物体の配置

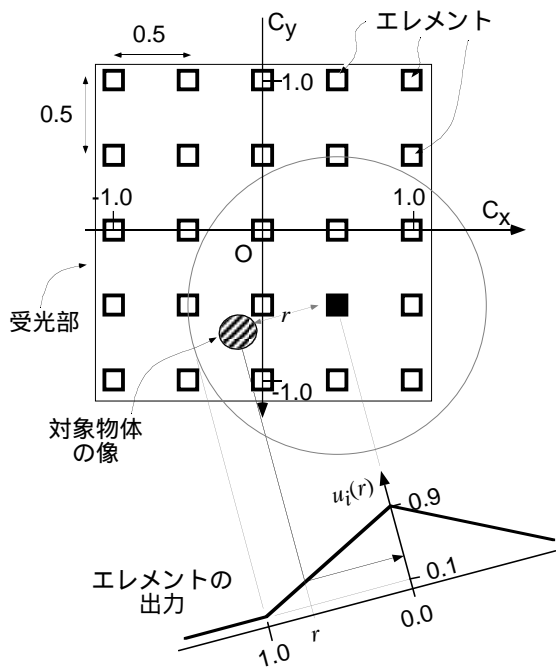


図5 受光部のエレメント配置と出力特性

れた二つのカメラが45度内側を向き、各カメラの視野角は上下左右とも20度である(図4)。なお対象物体の出現エリアは $x = -2.2 \sim +2.2$, $y = -1.4 \sim +1.4$, $z = 3.9 \sim 8.3$ とした。

各カメラ内の受光部上でのCx-Cy座標における $Cx = -1 \sim +1$, $Cy = -1 \sim +1$ が視野角に対応する(図5)。受光部には上記範囲内に等間隔(0.5)で25個(=5*5)の受光エレメントが配置されている。エレメント(i番目)は受光部座標上での、それ自身と対象物体の像との距離(r)に依存した出力特性、

$$u_i(r) = 0.9 - 0.8r \quad \text{if } r < 1.0$$

$$u_i(r) = 0.1 \quad \text{else}$$

を持つ。下位ネットワークの入出力ユニット同士は空間的な隣接関係の情報を持たないため、対象物体の位置をばかすような前処理によりデータセットが隣接関係の情報を持つことは、学習にとって重要である。

両カメラの受光部からの25次元アナログ出力は階層モデルに伝達される(図6)。階層モデルは、対応するカメラから画像情報を受け取り対象物体の画像上での位置に対応する2次元情報を出力する左右2つの下位ネットワー

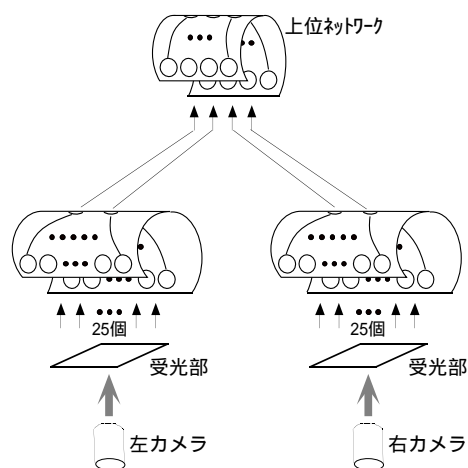


図6 実験に用いた階層モデル

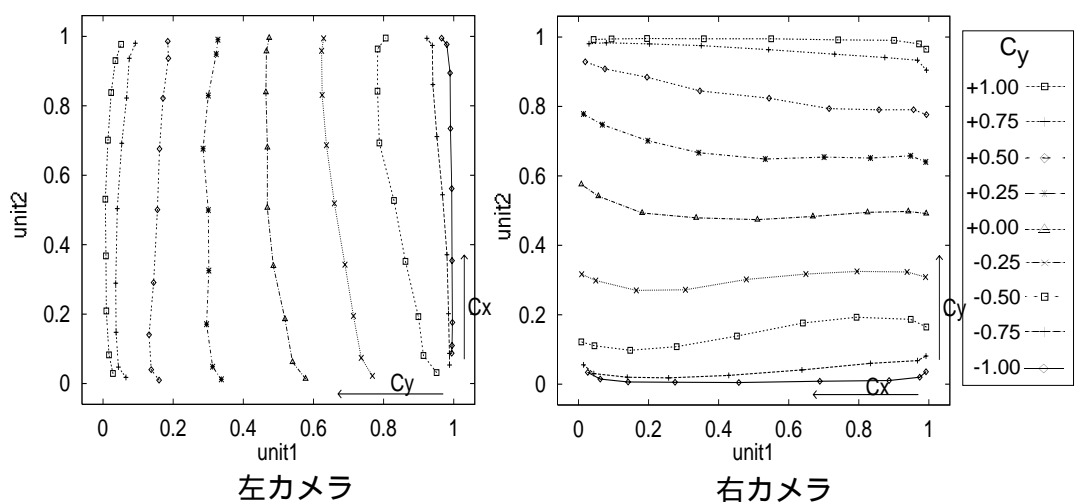


図7 下位ネットワークの内部表現

	下位ネットワーク	上位ネットワーク
レイヤー構成	25-10-2-10-25 (5層)	4-6-3-6-4 (5層)
重みの学習定数	5.0	5.0
重みのモーメント	0.4	0.4
出力の範囲	[0.0, 1.0]	[0.0, 1.0]
学習回数	20000	80000

表1 各砂時計ネットワークの実験条件

クと、下位ネットワークから出力されるの合計4次元の情報から対象物体の3次元的位置情報を復元する上位ネットワークにより構成される。

実験に使用したデータセットは以下である。

(a)学習データセット

対象物体を出現エリア内のランダムな位置にセットした1000個のデータ。

(b)テストデータセット

(b-i)下位ネットワーク用(2次元)

カメラの視野角内をスキャン。つまり、Cx-Cyの各軸に対し0.25刻みで対象物体像を移動した、81個(=9*9)のデータ。

(b-ii)上位ネットワーク用(3次元)

対象物体を出現エリア内でX,Y,Z各方向に0.4刻みでスキャンした、1152個(=12*9*12)のデータ。

3.2 結果

下位ネットワークは学習データセット(a)に対して誤差逆伝搬法で恒等写像を20000イテレーション学習し、その後同じ学習データセットに対する下位ネットワークの圧縮層の出力を利用して同様に上位ネットワークの学習を80000イテレーション行った。

学習した下位ネットワーク(左)にテストデータセット(b-i)を与えた時の圧縮層の出力を示す(図7)。ここではほぼユニット1の出力がCyに、ユニット2の出力がCxに対応している。一方、下位ネットワーク(右)では逆にユニット1の出力がCxに、ユニット2の出力がCyに対応している。よって何れの下位ネットワークとも期待通りに入力情報の変動部分が内部表現に抽出された。

引き続き上位ネットワークの学習では、同じ学習デー

タセット(a)に対する二つの下位ネットワークの圧縮層の出力を利用して80000イテレーションの誤差逆伝搬法を行った。学習させたネットワークにテストデータセット(b-ii)を与えたときの上位ネットワークの圧縮層の出力を図8に示す。このように上位ネットワークは対象物体の3次元位置に対応する内部表現を獲得した。

実験において、下位ネットワークが獲得した内部表現は圧縮層の出力空間の大部分を利用している。これは、入力層と出力層との誤差を小さくする恒等写像の学習が、圧縮層を通過する情報の解像度を落とさないように利用可能な空間をできる限り使用したためではないかと考えられる。これに対して、上位ネットワークの内部表現は圧縮層の出力空間の一部しか使用しておらず、さらに学習を続けても状況の変化はなかった。その点は疑問が残り、今後の検討が必要であると思われる。

4. まとめ

現在のコンピュータなどによる知的システムには、多種多様かつ矛盾や曖昧さを含んだ情報から必要な情報を取捨選択・統融合する高度の能力が必要とされつつある。これらの能力はシステムが、現実世界と直接相互作用して大量かつ多様な情報を扱う場合や、情報のデジタル化・マルチメディア化が進化したネットワーク社会のなかで活躍する場合に重要である。

上記能力の実現にはシステム自身による内部表現の柔軟な操作が一つの重要な基盤技術であると考えられる。そこで本報告書では、砂時計型のニューラルネットワークの階層化により、センサ情報(低次)から内部表現(高次)を自律獲得する情報統合モデルの提案を行った。さら

に、モデルの基本的な動作確認のために、計算機シミュレーションにより提案したモデルが2台の固定カメラの前で動き回る物体の画像情報から3次元情報をコードする内部表現を学習により復元できることを示した。

このように、本報告書では内部表現の自動獲得機構の実現という問題意識に基づき、これを階層化する方法の提案を行った。今回提案したモデルでは、高次の内部表現を獲得するには各ニューラルネットワークの構造やその接続関係をあらかじめ与える必要があり、その点は自律的でなく、さらに研究を進める必要がある。しかし、現状においても表現それ自体を与えるよりも少ない情報からシステムを構築できるので十分に有用である。

参考文献

1. Cottrell, G. W., Muno, P., and Zipser, D., "Image Compression by Back Propagation. An Example of Extensional Programming", ICS Report 8702, Institute for Cognitive Science, UCSD, 1987.
2. Cottrell, G. W. and Metcalfe, J., "EMPATH: Face, Emotion, and Gender Recognition Using Holons", NIPS3, pp.564-572, 1991.
3. DeMers, D. and Cottrell, G. W., "Non-Linear Dimensionality Reduction", NIPS5, pp.580-587, 1993.
4. 鈴木, 安藤, "モジュール学習による3次元物体の認識と類別", 信学技報NC93-62, pp.59-66, 1993.
5. 柳沼義典, 木本隆, 山川宏, "砂時計型ニューラルネットによる物体認識のための複数センサ情報の融合", 第11回ファジィシステムシンポジウム講演予定, 1995.
6. 野田, "過負荷学習を用いた恒等写像学習による内部表現獲得", 日本神経回路学会第5回全国大会講演論文集, pp. 80-81, 1994.
7. 木本隆, 柳沼義典, 山川宏, "不完全なデータに対応する砂時計型ネットワーク", 信学技報, NC94-62, pp.17-24, 1995.

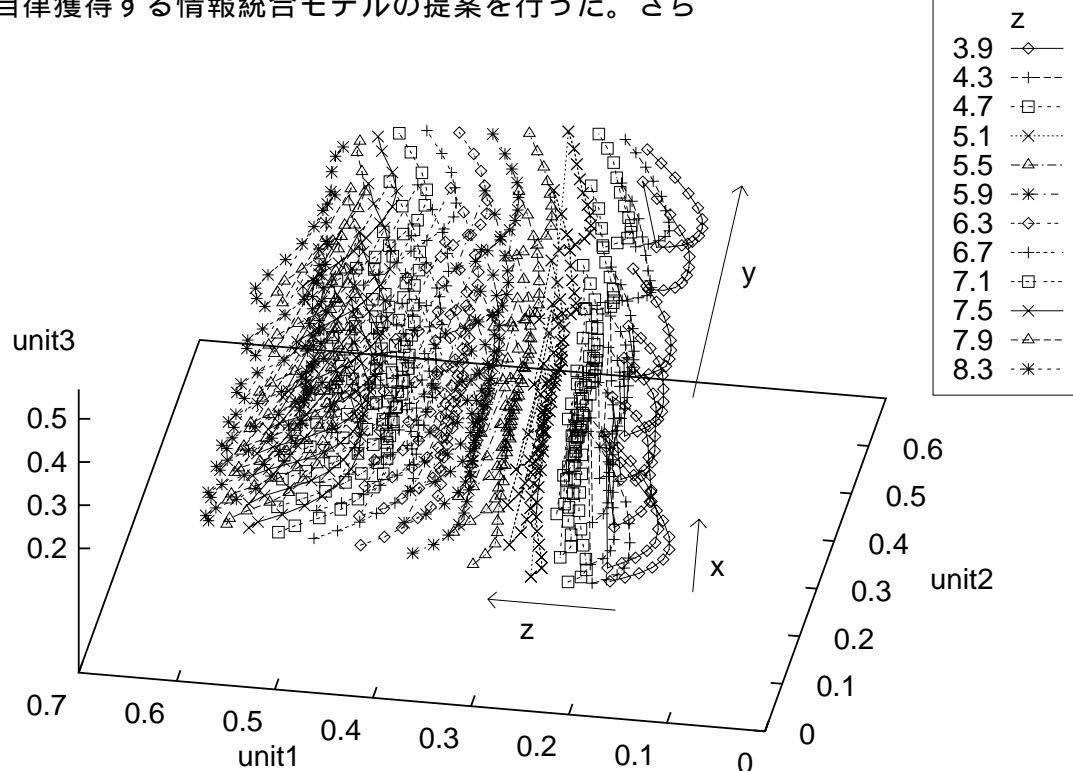


図8 上位ネットワークの内部表現

上位ネットワーク用の12*9*12個のテストデータ(図4参照)の入力に対する上位ネットワーク圧縮層の3ユニットの出力を示した。カメラ座標のx軸がユニット2に、y軸がユニット3に、z軸がユニット1に、ほぼ対応する結果が得られた。

連絡先

山川宏
〒305 茨城県つくば市竹園1-6-1
つくば三井ビル 16F
R W C P つくば研究センタ
能動知能研究室
TEL:0298-53-1664,
FAX: 0298-53-1652
E-mail: yamakawa@rwcp.or.jp

木本隆、柳沼義典
〒211 神奈川県川崎市上小田中1015
(株)富士通研究所
知識処理研究部
TEL:044-754-2667,
FAX: 044-754-2664
E-mail: kimoto@flab.fujitsu.co.jp
yaginuma@flab.fujitsu.co.jp