

動的ネットワーク再構成法による図形特徴抽出の試み 1M-4 —文字認識における大域的的特徴と局所的特徴の利用—

丹羽 建一 小高知宏 小倉久和

福井大学

1 はじめに

現在ニューラルネットワークは多くのところで実際に利用され、様々な応用に用いられている。これらの多くはニューラルネットワークが基本的に画素の有無によって動作していることから、位相幾何的な関係を無視することになっている。しかし、人間の認識過程について考えてみると位相幾何的な関係を無視して認識が成り立っているとは考えにくい。

そこで、これを克服するため人間がそれらの文字を認識する過程について考察してみよう。まず即時には判別不可能な文字があらわれると、その文字の概形からある範囲での候補を選択する。そして次にその候補の中からそれぞれの文字を一意に同定しうる部位に関してより詳しく認識をすすめることで、最終的に該当する文字を決定していると考える。本研究ではこれらの過程をモデル化し、階層型ニューラルネットワークを用いて計算機シミュレーションを試みた。

このモデルでは文字認識過程を前半の「大域的特徴」と後半の「局所的特徴」を捉える2つの過程にわけ、動的ネットワーク再構成法を用いて統合し、人の認識過程をシミュレートすることを目的とする。

2 文字特徴認識モデル

本研究で扱う文字認識過程のモデルは、人間が特に判断を必要とする文字の認識過程について、

- 1 大域的特徴による候補の選択
- 2 局所的特徴による最終的な決定

の2段階認識を仮定している。具体的な処理の流れは、まず判読が困難な「わ」という文字があった場合を考えると、最初に該当する文字と同じ様な概形を持つ幾つかの文字を文字集合全体からピックアップする。選ばれた文字—この場合は「わ、れ、ね」—の中において、さらにそれぞれの文字について、その文字が持つ固有の特徴があらわれる部分を選択し

A research of the classification process in recognition of letters using a dynamic reconstructed neural network system
Ken'iti Niwa, Tomohiro Odaka, Hisakazu Ogura
Fukui Univ

その領域について局所的特徴をとらえて認識をすすめる。

この認識モデルのシミュレーションシステムへの適用は、まず大域的特徴によるグループ分類を G-net(Global-network) 処理部においておこなう。この G-net による出力は、次の L-net(Local-network) 処理部に伝わって、L-net でのネットワークの結合状態を動的に定義することになり、この動的な定義を実現するために、“動的ネットワーク再構成法”をもちいる。

以上のようなモデルの適用によって、人間の認識過程により近い文字認識の実現を目的とする。

2.1 大域的特徴の把握

第一段階における認識のために、文字の大域的特徴を抽出する。認識対象候補の絞り込みを目的とする。平仮名文字の概形についてまとめたものであり、あらかじめ用意したいくつかのパターンを用いた。この処理は、システム内部において G-net 処理部で実現する。この G-net は従来の階層型ニューラルネットワークである。この出力は、対象領域のロケーション情報が 4bit、局所特徴コードが 3bit で一つの L-net を構成する基本的な単位となる。

2.2 局所的特徴の把握

第二段階では、各文字の部分的な特徴を抽出する。前述の大域的特徴によって与えられた認識対象となるグループ分類内において、より詳しい認識をすすめる。この特徴については、大域的特徴によるグループ分類内の文字について、それぞれの文字を一意に特徴づける部分についての特徴の獲得を目的とする。この処理は、システム内部では L-net 処理部で実現する。L-net 処理部については、G-net 処理部の出力によってその内部が構成が動的に定義される。

3 文字認識システム

本システムにおいては、大域的特徴を捉える G-net からの出力に従い、L-net が扱う領域が動的に変化する。“動的ネットワーク再構成法”を用いる。

動的ネットワーク再構成法は、システムの内部動作の結果に従って、新たなネットワークの結合を生成していく方法である。本システムにおいては G-net から出力されたグループ分類から L-net の結合状態をその都度変更していく。

システム全体の構成を図 1 に示す。

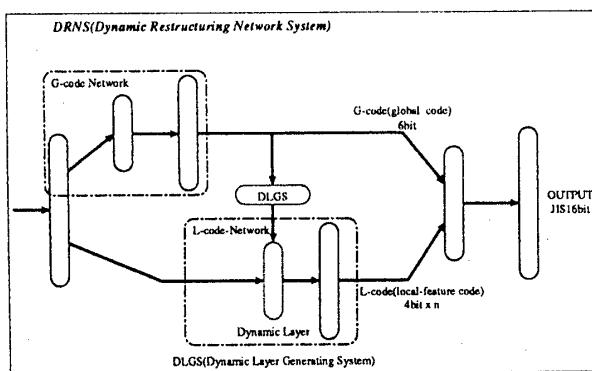


図 1: システム構成

図 1 中の D L G S (Dynamic Layer Generating System) 部において G-net からの出力 (グループ分類) を受け、入力レイヤからの信号のうち特定の領域を選択し学習レイヤを生成して L-net を構成する。

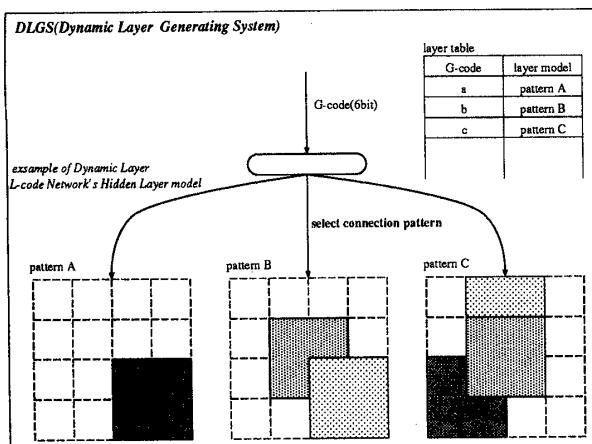


図 2: D L G S の構成

図 2 は D L G S の内部構成の一部を示したものである。ここにおいて、それぞれのグループ分類内の固有領域 (一意に文字を同定しうる部位) について注目することで、最終的な判断をおこなう。L-net 処理部では、 16×16 の領域を基本レイヤとして、 3×3 の入力レイヤからの入力信号をマスクする。入力レイヤに対する基本レイヤの領域指定については、入力レイヤを縦横 8 ブツシフトさせることで 3×3 の格子状に分解したそれを対象とする。

4 システムによる認識結果

実験に用いた文字图形は、五種の文字フォント、

- 1 wadalab-mincho
- 2 pcf(from X11R5)
- 3 wadalab-gothic
- 4 zeit-minchou
- 5 wadalab-maru

である。表 1 は 1、2 を学習データセットとし、4、5 を検査データセットとして用いた本システムの認識結果である。数値は 4、5 の文字に関する認識率である。

表 1: 実行結果 1

グループ	フォント 4	フォント 5
わ、れ、ね	95.4	64.1
け、は、ほ	92.0	56.6
た、な、に、こ	78.3	55.4

表 2 は、1、2、3 を学習データセットに用い、検査データセットとして 4、5 を認識させた同様の認識結果である。

表 2: 実行結果 2

グループ	フォント 4	フォント 5
わ、れ、ね	96.2	68.3
け、は、ほ	91.5	55.9
た、な、に、こ	79.1	58.6

実行結果を検討すると、大域的特徴によるグループ分類によってかなりの差が認められるが、5 のフォントに関しては全般的に従来のニューラルネットワークによる認識と比較して認識率の向上がある。

5まとめ

本研究では、新たな文字認識モデルを検討しシミュレートすることにより、人間の文字認識過程のモデルを検証することを目指した。現状では、通常の 3 層ネットワークによる認識結果と比較すると、丸文字のような文字に関して認識率が向上した。

今後大域的特徴によるグループ分類についての系統的な分析をおこなうことで、漢字の認識について、本モデルの拡張を検討していく。

[参考文献]

1. David Marr, VISION, 産業図書株式会社, 1987
2. 乾 俊郎, 視覚情報処理の基礎, サイエンス社, 1990
3. 安西祐一郎, 認識と学習, 岩波書店, 1989