

# Gradient-Based Learning Applied to Document Recognition の和訳

## 著者

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner

## 概説

逆誤差伝搬法で学習した多層ニューラルネットワークは、勾配に基づく学習法の最たる成功例である。適切なネットワークアーキテクチャがあれば、勾配に基づく学習アルゴリズムは手書き文字などの高次元パターンを分類するような複雑な決定超平面を、最小限の前処理で作成することができる。本論文では、手書き文字の認識に適用される様々な方法を振り返り、標準的な手書き整数認識タスクを用いて比較する。その結果、二次元構造の多様性を取り扱えるように設計された CNN が他の手法に比べて良い性能を示した。

実際の文章認識システムは、フィールド抽出、分割、認識、言語モデリングなど複数のモジュールにより構成されている。Graph Transformer Networks (GTN) と呼ばれる新しい学習の枠組みにより、このような多くのモジュールからなるシステムを勾配に基づく学習アルゴリズムを用いてパフォーマンス指標を最小化するように大域的に学習することが可能となった。

本論文では、オンライン手書き文字認識の 2 つのシステムを紹介する。実験により、大域的な学習の利点と GTN の柔軟性が示された。

また、本論文では銀行小切手を読み取る GTN も紹介する。この手法では、CNN の文字認識と、商業や個人利用の小切手の正確さを記録するための大域的な学習方法を組み合わせている。商業的に利用されており、1 日あたりに数百万の小切手の認識をしている。

キーワード – ニューラルネットワーク, OCR, 文章認識, 機械学習, 勾配に基づく学習, CNN, Graph Transformer Networks, Finite State Transducers.

## 略称

- GT Graph transformer

- GTN Graph transformer network.
- HMM Hidden Markov model.
- HOS Heuristic oversegmentation
- K-NN K-nearest neighbor
- NN Neural network.
- OCR Optical character recognition.
- PCA Principal component analysis.
- RBF Radial basis function.
- RS-SVM Reduced-set support vector method.
- SDNN Space displacement neural network.
- SVM Support vector method.
- TDNN Time delay neural network.
- V-SVM Virtual support vector method.

## 1 序論

ここ数年、機械学習の技術、それらを適用したニューラルネットワークがパターン認識システムの設計において、非常に重要な役割を果たしている。実際に、近年の連続的な音声の認識、手書き文字の認識といったパターン認識アプリケーションの成功にはこのような学習法の利用が重要な要因であると言える。

本論文の主な主張は、人間の経験的な知識に頼らず、自動的な学習に依存することで、より良いパターン認識システムを作成することができるということだ。これは、近年の機械学習や情報工学の進歩により可能となった。文字認識を例として、本論文ではこれまでの手作業の特徴抽出が、ピクセル画像を直接操作する注意深い学習機械に置き換えることができることを示している。また文章認識を例として、本論文では個々に設計されたモジュールを組み合わせることで認識システムを作る従来の方法は、大域的な性能基準を最適化するために全てのモジュールを学習する GTN と呼ばれる統一的で原理的な設計の枠組みに置き換えることができることを示している。

パターン認識の研究初期から、音声、文字パターンなどといった自然界のパターンの豊富さと多様さに

より手作業で完全に正確なパターン認識システムを作成することは難しいことが知られている。そのため多くのパターン認識システムは自動的な学習と人間が考案したアルゴリズムを組み合わせで作られている。個々のパターンを認識するための一般的な方法は、図1に示すようにシステムを2つの主なモジュールに分けて構成する。一方は特徴抽出器と呼ばれるモジュールである。特徴抽出器では入力パターンを、(a) 照合・比較しやすく、(b) 入力パターンの変形や歪みに対して比較的頑強である低次元のベクトルや短い配列で表現できるように変形する。特徴抽出器は予備知識の多くを有しており、タスクに特化している。また、ほとんど手作業で作られるため、設計の際の作業の大半を占める。もう一方は分類機と呼ばれるモジュールであり、汎用的で学習可能である。このアプローチの大きな問題の1つは、設計者の適切な特徴量のセットを考える能力に認識精度が大きく依存することだ。このことは不幸にもなにか問題が発生した時は、気が遠くなるような設計作業を再び行わなければならないことを意味している。パターン認識に関する多くの文献は、特定のタスクに関する異なる特徴量のセットの相対的な長所の比較に費やされている。

歴史的に、分類器が用いる学習方法が要因にクラス分類可能な低次元空間に限定されていたため特徴抽出器が必要とされた[1]。この10年間で3つの要因が重なりこの見解は変化してきた。1つ目の要因は、高速の演算装置を持つ低価格の計算機が登場し理論的な方法より強引な数値解析に頼るようになったこと。2つ目の要因は手書き文字の認識のような市場規模が大きく関心の高い問題に関する大規模なデータベースが使用可能になり、認識システムの設計において手作業での特徴抽出より、実データに依存することができるようになったこと。3つ目の非常に重要な要因は高次元の入力に対応し、大規模なデータを与えられても複雑な決定関数を生成できる高性能な機械学習方法が生み出されたことである。近年の音声認識や手書き文字認識システムの精度向上には、学習技術と大規模なデータセットの進化が大きく起因している。この証拠として最近の商用のOCRシステムの大部分は誤差逆伝搬法で学習された多層ニューラルネットワークを使用している。

本論文では手書き文字認識における課題点を検討し(1節,2節)、手書き数字認識のベンチマークデータセットにおいていくつかの学習方法の性能を比較した(3節)。さらなる自動的な学習は有益であるが、タ

スクに関する最小限の予備知識無しで成功する学習手法は存在しない。多層ニューラルネットワークの場合、予備知識を組み込む良い方法はタスクに応じてアーキテクチャを調整する方法である。2節で紹介するCNN[2]は、局所的な接続パターンを用いたり重みに制約を与えることで2次元形状の不変性に対する予備知識を取り入れた特殊なニューラルネットワークの1つである。分離された手書き整数認識タスクに対するいくつかの方法の比較を3章で示す。4章では個々の文字認識から文書中の単語や文の認識に至るまで、全体の誤差を減らすため学習した多くのモジュールを組み合わせる方法を紹介する。モジュールが有向グラフを操作可能の場合、マルチモジュールシステムを用いた手書きの単語といった可変長のオブジェクトの認識することが最適である。これは4章で述べる学習可能なGTNの概念につながる。5章では、単語や文字列を認識するための古典的なヒューリスティックなオーバーセグメンテーションを紹介する。6章では人力でのセグメンテーションやラベリングを必要としない単語レベルでの識別機の学習を可能にするような識別的、非識別的な勾配に基づく学習法を紹介する。7章では、入力における全ての位置において認識器を作用させることで、分割ヒューリスティックの必要性を排除する有望な空間置換ニューラルネットワークのアプローチを示す。8節では、学習可能なGTNが一般的なグラフ構成アルゴリズムに基づく複数の一般化変換として定式化できることを示している。また、音声認識でよく用いられる隠れマルコフモデルとGTNの関連も示す。第9節は、ペン型コンピュータに入力された手書き文字を認識するための大域的に学習されたGTNシステムについて説明する。ユーザーが書いた文字を計算機が即座に回答を返さなければならないため、この問題は”オンライン”手書き文字認識として知られている。GTNシステムの中核はCNNである。本研究で得られた結果は、認識器を事前に領域分割された手書き文字で学習するのではなく単語レベルで学習することの利点を示している。10節は、手書きと機械印刷の銀行小切手を認識するGTNに基づいた完全なシステムを説明する。このシステムの中核は2節で述べたLeNet-5と呼ばれるCNNである。このシステムは銀行産業向けの小切手認識システムとしてNCR社で商業的に利用されている。全米のいくつかの銀行で一月数百万枚の小切手を読み取っている。

## A. データからの学習

機械学習を自動化するための様々なアプローチがあるが、最も成功したアプローチの1つは、勾配に基づく学習と呼ばれる近年ニューラルネットワークコミュニティで人気のアプローチである。学習する計算機は関数  $Y^p = F(Z^p, W)$  を計算する。ここで、 $Z^p$  は  $p$  番目の入力パターンであり、 $W$  はシステムにおける調整可能なパラメータの集合を表す。パターン認識の設定では、出力  $Y^p$  はパターン  $Z^p$  の認識可能なクラスラベル、あるいは各クラスに関する確率やスコアとして解釈できる。損失関数  $E^p = \mathcal{D}(D^p, F(W, Z^p))$  は、パターン  $Z^p$  における正しいまたは望ましい出力  $D^p$  とシステムの出力との不一致さを測定する。平均損失関数  $E_{train}(W)$  は学習セット  $(Z^1, D^1), \dots, (Z^P, D^P)$  と呼ばれるラベル付きの学習例の集合上の誤差  $E^p$  の平均値である。最も単純な設定において、学習問題は  $E_{train}(W)$  を最小化する  $W$  の値を見つけることである。実際には、学習セットにおける性能はあまり重要ではない。より重要な指標は実際に使用される領域での誤差率である。この指標はテストセットと呼ばれる学習セットから切り離されたサンプル集合に対する正解率を計算することで推定できる。多くの理論的・実験的研究 [3, 4, 5] により、テストセットで予測されるエラー率と学習セットで期待されるエラー率との間の差は学習サンプルの数によっておおよそ式 1 のように減少することが分かっている。

$$E_{test} - E_{train} = k(h/P)^\alpha \quad (1)$$

ここで  $P$  は学習サンプルの数であり、 $h$  は「有効容量」、機械の複雑さの推定値、 $\alpha$  は 0.5–1.0 の値で、 $k$  は定数である。学習サンプルの数が増加した時は差は常に減少する。その上  $h$  が増加すると、 $E_{train}$  が減少する。従って、 $h$  が増えた時、最小の汎化誤差  $E_{test}$  を達成する容量  $h$  の最適な値があれば、 $E_{train}$  の減少量と差の増加はトレードオフの関係となる。多くの学習アルゴリズムは  $E_{train}$  だけでなく差の増加も最小化しようと試みる。この正式な呼び名は構造的リスク最小化と呼ばれ、それぞれのサブネットが以前のサブネットのスーパーセットであるようなパラメータ空間のサブセットの系列に対応する容量に増加した学習機械の系列を定義することに基づいている。実際には、構造的リスク最適化は  $E_{train} + \beta H(W)$  を最小化することを意味している。ここで関数  $H(W)$  は正規化関数と呼ばれ、 $\beta$  は定数である。 $H(W)$  はパラメータ空間の高容量のサブセットに所属するパ

ラメータ  $W$  に関して大きな値を取るよう選択される。 $H(W)$  を最小化するこ

## 参考文献

- [1] Richard O. Duda and Peter E. Hart. Pattern classification and scene analysis. In *A Wiley-Interscience publication*, 1973.
- [2] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, Vol. 1, No. 4, pp. 541–551, 1989.
- [3] H. S. Seung, H. Sompolinsky, and N. Tishby. Statistical mechanics of learning from examples. *Phys. Rev. A*, Vol. 45, pp. 6056–6091, Apr 1992.
- [4] Vladimir Vapnik, Esther Levin, and Yann Lecun. Measuring the vc-dimension of a learning machine. *Neural Computation*, Vol. 6, pp. 851–876, 09 1994.
- [5] Corinna Cortes, L. D. Jackel, Sara Solla, Vladimir Vapnik, and John Denker. Learning curves: Asymptotic values and rate of convergence. In J. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems*, Vol. 6. Morgan-Kaufmann, 1993.