

深層学習のための先進的な学習技術

NTTではAI（人工知能）活用を重要な戦略の1つとして位置付けています。NTTソフトウェアイノベーションセンターではAIの中でも中心的な役割を担う深層学習のアルゴリズムの研究開発に取り組んできました。本稿ではその中でも学習の高速化と安定化を実現するアルゴリズムの概要、および検証活動について紹介します。

井田 安俊 / 金井 関利
 藤原 靖宏 / 八木 哲志
 飯田 恭弘

NTTソフトウェアイノベーションセンター

深層学習の課題

NTTでは「社会的課題の克服」や「産業競争力の強化」といったテーマを掲げるにあたり、AI（人工知能）を活用した取り組みを重要な戦略として位置付けています。機械学習はこのAIの分野において中心的な役割を担う技術の1つです。機械学習ではデータからモデルを学習し、未知のデータに対して何らかの予測を行います。機械学習の中でも深層学習（ディープラーニング：Deep Learning）はさまざまなタスクで実用的な精度を達成したことで注目を浴びています。すでに事業の高度化のために導入が進み、新規事業の創出の観点からも深層学習は期待されています。

深層学習はデータから階層的に特徴抽出を行う仕組みを持った技術です。モデルとしては主に多層のニューラルネットワークが使われます。ニューラルネットワークの中にもさらに細かい分類があり、画像データに対しては畳み込みニューラルネットワーク（CNN：Convolutional Neural Network）、時系列データに対しては再帰型ニューラルネットワーク（RNN：Recurrent Neural Network）というように、タ

スクとデータによって使い分けられます。ニューラルネットワークでは入力データの信号に対して重み付けや非線形変換を階層的に行い、信号を伝播させていくことで予測を行います（図1）。学習とはあらかじめ正解との誤差が小さくなるようにこの重みの値を調整する手続きを指します。この階層的な構造により深層学習はさまざまなタスクで高い精度を達成してきました。

しかし深層学習のモデルが持つ階層的な構造は学習時間の増加や学習自体の不安定化といった問題を引き起こします。したがって深層学習の利活用では学習アルゴリズムを高度化することでこれらの問題を回避することが重要な課題の1つとされます。そこでNTT研究所では、①学習の効率を改善して高速化を図るアルゴリズムと②RNNの学習を安定化させるアルゴリ

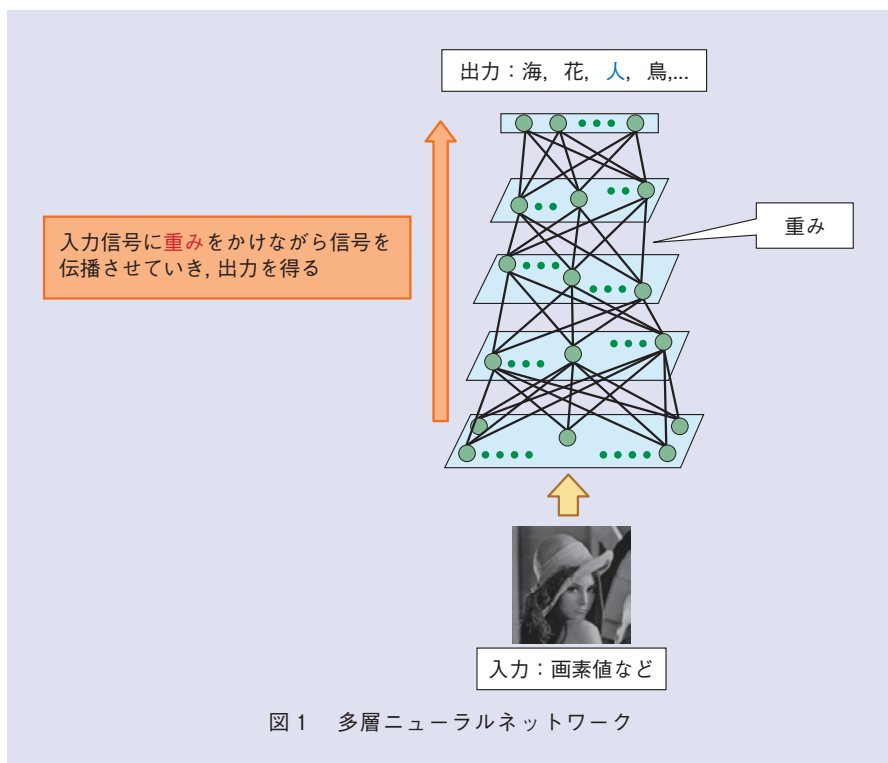


図1 多層ニューラルネットワーク

ズムを開発しました^{(1),(2)}。以降ではこれらの技術を紹介します。

学習の効率改善による高速化技術

多層ニューラルネットワークのモデルの学習では誤差が小さくなるように重みを調整します。重みの調整は図2にあるようなループを何回も適用することで徐々に行います。まずモデルにデータを入力し、予測結果を得ます(順伝播)。次にこの予測結果と正解ラベルを比較し、誤差を計算します(誤差計算)。この誤差をモデルに対して伝播させ(逆伝播)、重みの更新方向を計算します。そして重みの更新量を計算し(重み更新量計算)、最後にこの更新量を使って重みを更新します(重み更新)。重みの更新量の計算には複数のアプローチがあり、アプローチによって学習の効率は変化します。つまり、1ループ当りにどれだけ誤差を減

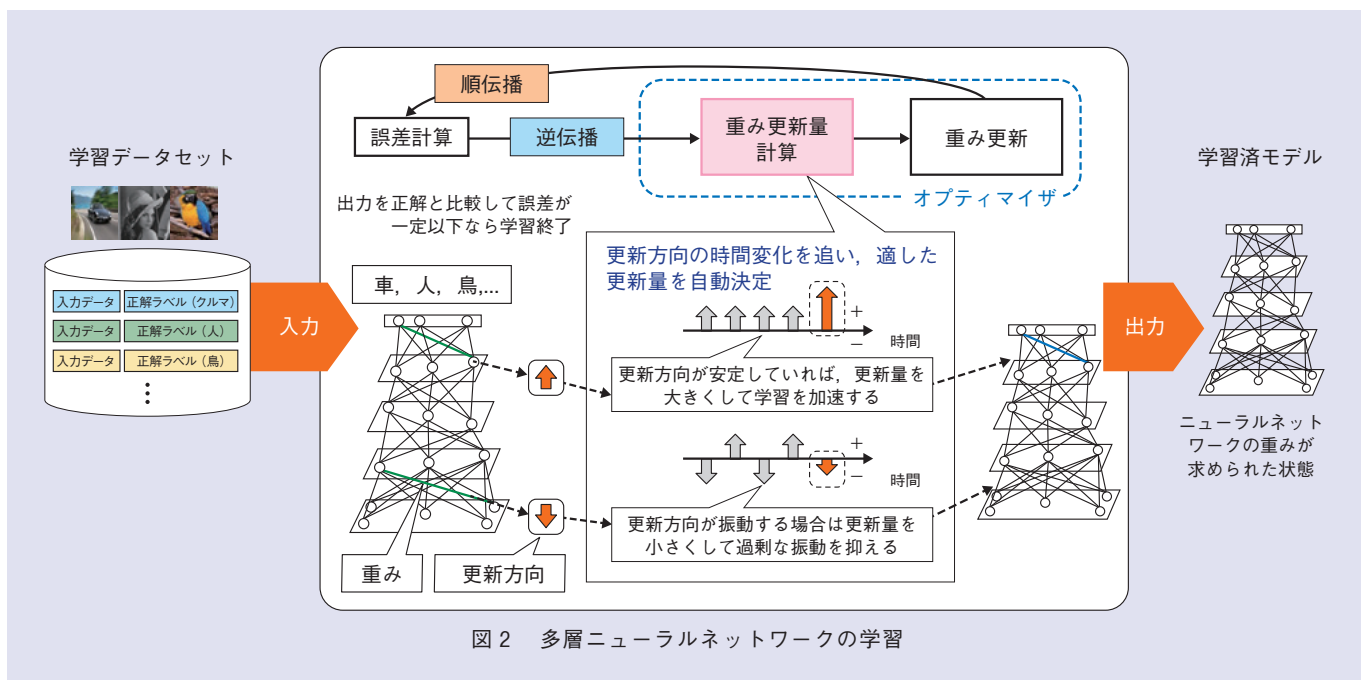
らすことができるかが変わってきます。

多くのアプローチでは過去の更新方向の情報に基に更新量を調整します。例えば現在広く使われているRMSpropやAdamといった手法は更新方向の絶対値を使った統計量で更新量を調整します。しかしこの手法では更新方向の絶対値を使うため、一般にはそのばらつきが考慮されません。これは更新方向がばらついていてもそうでなくても、絶対値が同じであれば同じ更新量を設定されてしまうことを意味します。

そこでNTT研究所では更新方向のばらつきを基に更新量を調整するアルゴリズムを開発しました(図2)。更新方向のばらつきが大きい場合、学習は不安定になりやすいため、更新量を抑えて安定化させます。ばらつきが小さい場合は更新量を大きくします。こ

れによりばらつきは大きくなりますが、局所解から抜け出せるチャンスも多くなり、誤差をより小さい値まで減らすことができます。

このアルゴリズムはシンプルであり、多くの深層学習のフレームワークで容易に実装することができます。一方で理論的には興味深い側面もあります。このアルゴリズムは前処理付きの最適化の一種とみることができます。このとき最適化の条件を整える行列(最適化の分野で前処理行列と呼ばれる行列)はFisher情報行列と呼ばれる勾配(重みの変化に対する誤差の変化量)から計算した行列の対角成分の平方根の近似となります。これは情報処理の問題を視覚的に理解するための枠組みである情報幾何学の観点から、繰り返し更新を行うアルゴリズムの場合は既存手法よりも収束が速くなる可能性があることを示唆しています(図3)。



GRU学習安定化技術

音声認識や機械翻訳などの時系列データを扱うにはRNNという深層学習のモデルが用いられます。RNNは過去のデータの情報を状態として記憶し、この状態と現在の入力を使って出力を計算します。例えば与えられた文章に対して次に出てくる単語を予測するようなタスクを考えます。次に出てくる単語はこれまでの文章の文脈に依存しており、この過去の文脈の情報をRNNの状態が記憶することで次の単語を精度良く予測することができます。

時系列データをより精度良く処理するためにはより古い情報を記憶することが必要です。そのためより古い情報を記憶する長期記憶がRNNの性能を評価する重要な指標となっています。この長期記憶を達成するため1990年代後半に Long Short-Term Memory (LSTM) というモデル構造が提案されました⁽³⁾。LSTMは過去の記憶を取り込むメモリセルという構造と、不要な情報を忘却するゲート構造を持っていて、過去の重要な情報は保存し不要となった情報は忘却することができます。LSTMはすでに多数の機械翻訳や音声認識技術に応用されています。しかしながらLSTMは複雑な構造をしているため、より簡易化して計算量を抑えたGated Recurrent Unit (GRU) が2014年に提案されました⁽⁴⁾。GRUはLSTMの持つ入力ゲートと忘却ゲートを1つにすることでより単純で計算量の少ない構造を実現していますが、ほぼLSTMと同程度の精度が

達成できることが実験的に示されています。

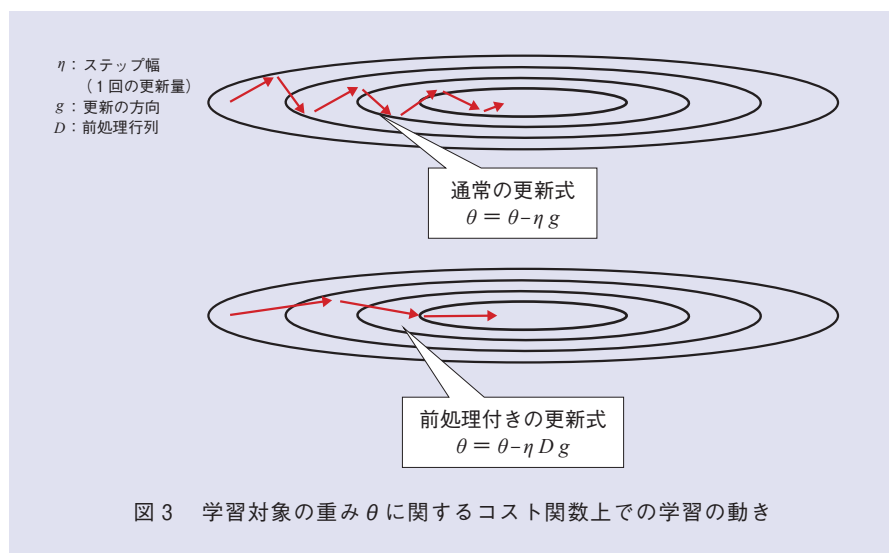
精度良く時系列データを処理可能なRNNですが、その学習は勾配爆発と呼ばれる現象により不安定なものとなっています⁽⁵⁾。勾配爆発とは勾配が非常に大きな値となり学習が失敗する現象です。これは長期間にわたるデータを扱う際に、学習に使う逆伝播時の微分計算が何度も積算されて指数関数的に増加してしまうことが原因とされています。これに対する既存技術として Gradient Clipping という勾配をしきい値で打ち切る手法がありますが、しきい値を試行錯誤的に調整しなければなりません。

NTTでは、この勾配爆発という問題に対してGRUの挙動を解析することで解決する技術を研究しています(図4)。勾配爆発は力学系の観点からみると、GRUが重みの変化により分岐(bifurcation)を生じることが原因といえます。分岐とはGRUの挙動が小さな重みの変化によって劇的に変化

する現象です。NTTのGRUの学習安定化技術ではGRUの状態の挙動を解析し、分岐を生じる点を明らかにしました。さらにこの分岐点を効率良く回避して学習させる手法を提案し、Gradient Clippingより試行錯誤が少なく高精度な学習手法を研究しています⁽²⁾。

NTTの開発した先進技術をユーザーとともに洗練化していく取り組み

私たちは研究にはじっくりと取り組みつつも、その一方で幅広いユーザーとともに技術の有効性を検証する活動も行っています。例えば深層学習の高速化技術はChainer, TensorFlow, Caffeなどの複数の代表的な深層学習のフレームワークで学習の最適化ライブラリの1つとして動くかたちを実装し、NTTグループ会社と連携して評価を行っています。これまでにさまざまな画像データを用いた画像認識サービスのための学習モデル構築を高速・高精度に行えるかを評価してきました。こ



RNNの性質の切り替わる点で勾配が爆発
GRUの性質を解析

爆発する点に重みを進ませない制約付きの学習

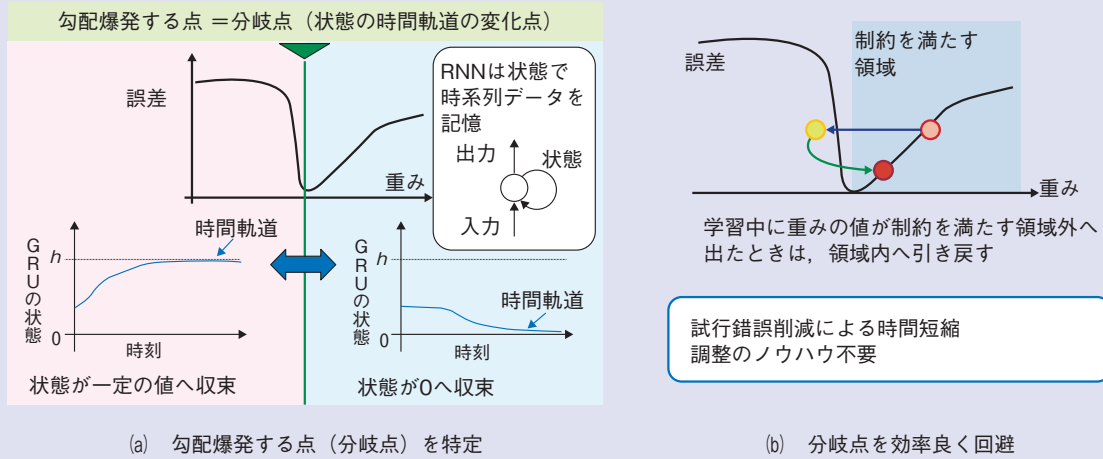


図4 GRU学習安定化技術

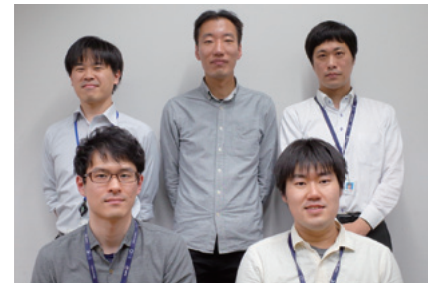
の評価を通して私たちはさまざまなデータに対してどのようなサイズ(バッチサイズ)で学習データをインプットすれば良いか、学習率や重みの初期値の与え方などのノウハウの蓄積を進めています。また学習安定化技術も音声、言語、センサデータのようなさまざまな時系列データを対象に検証し、オプティマイザとの組合せと安定性、精度との関係に関するノウハウの蓄積を進めています。

また、深層学習の高速化や安定化の技術はオープンソース化など、より広いユーザとのコラボレーションを推進する手段を幅広く検討しています。さらにNTT研究所がCCI (corevo Computing Infrastructure) として進めているAI処理基盤に深層学習の高速化や安定化の技術を組み込むことで、深層学習のフレームワーク、チューニング、分散処理基盤をユーザにワンストップで提供することも検討してい

ます⁽⁶⁾。深層学習の技術をCCIに組み込むことにより、専門的な知識を持たない人でも高度なデータ分析を行えるようになることをめざしています。

■参考文献

- (1) Y. Ida, Y. Fujiwara, and S. Iwamura: "Adaptive Learning Rate via Covariance Matrix Based Preconditioning for Deep Neural Networks," Proc. of IJCAI 2017, pp.1923-1929, Melbourne, Australia, August 2017.
- (2) S. Kanai, Y. Fujiwara, and S. Iwamura: "Preventing Gradient Explosions in Gated Recurrent Units," Proc. of NIPS 2017, pp.435-444, California, U.S.A., Dec. 2017.
- (3) S. Hochreiter and J. Schmidhuber: "Long short-term memory," Neural Computation, Vol.9, No.8, pp.1735-1780, 1997.
- (4) K. Cho, B. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio: "Learning phrase representations using RNN encoder-decoder for statistical machine translation," Proc. of EMNLP 2014, pp.1724-1734, Doha, Qatar, Oct. 2014.
- (5) R. Pascanu, T. Mikolov, and Y. Bengio: "On the difficulty of training recurrent neural networks," Proc. of ICML 2013, pp.1310-1318, Atlanta, U.S.A., June 2013.
- (6) 川島: "IoT/AIサービスの発展を支える革新的基盤技術創出に向けた4つの取り組み," ビジネスコミュニケーション, Vol.54, No.12, pp.8-13, 2017.



(後列左から) 八木 哲志/ 藤原 靖宏/
飯田 恭弘
(前列左から) 井田 安俊/ 金井 関利

深層学習はAIの基盤技術の1つです。NTT研究所では深層学習の独自技術を開発しつつ、CCIを通じてそれらを使いやすくすることで事業の優位性の確保に貢献します。

◆問い合わせ先

NTTソフトウェアイノベーションセンタ
分散基盤技術プロジェクト
TEL 0422-59-2795
FAX 0422-59-3739
E-mail iida.yasuhiro@lab.ntt.co.jp