

ONTKITのご紹介

-深層学習補完ケース-

近年、深層学習に代表されるニューラルネットワークの研究と実用化が一気に花開いておりますが、ニューラルネットワークはグラフ理論で言うところの有向かつ重み付きの複雑ネットワーク構造です。

今回、ONTROXは数々の大手企業様に採用されているグラフ理論を実践した製品の知見を基礎として、ニューラルネットワークにも適用可能な大規模大容量の複雑ネットワークを扱うことができる製品、ONTKITを開発しましたのでここにご紹介させていただきます。





近年の人口知能研究の潮流と将来

近年、人口知能研究の世界でニューラルネットワークに対する学習手法において、深層学習(Deep Learning)と呼ばれるここ20年来で最大のブレークスルーが起きています。

また、従来よりグラフィックボードに搭載されていたGPU(Graphics Processing Unit)の計算資源が近年、通常のコンピューティング処理に転用できるようになり、これがスーパーコンピュータ等に代表されるHPC(High Performance Computing)のデファクトスタンダードになりつつありますが、このGPUを利用したHPC処理がニューラルネットワーク上での演算が必要とする計算処理にベストマッチしており、一気にニューラルネットワーク分野の研究と実用化が花開いています。

現状ではそれでもまだ生体とHPCの処理速度には差がありますが、現在でも分野を特化しさえすれば生体の処理速度と精度を超える場面も多く、確実にその差は縮まってきており、ムーアの法則とこの分野の研究のスピードを考えればあと20年程度で人間が持つ汎用知能を超えた汎用超知能が誕生することが現実味を帯びて真剣に議論され、これによって世界は一変するだろうと言われています。

ニューラルネットワークとグラフ理論の相関性

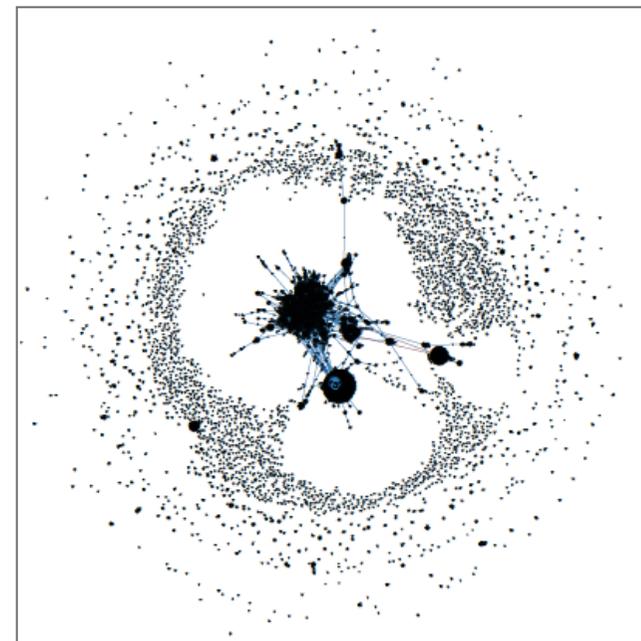
生体の脳内で表現されているニューラルネットワーク上ではニューロンとニューロンから伸びる軸索が他のニューロンへ接続する構造が複雑に絡み合い、その活動状況や接続の強弱が神経伝達物質を介して情報として伝播することで知性を獲得していることが知られています。

この構造は計算科学の分野では複雑ネットワーク構造と呼ばれ、グラフ理論という分野で研究されてきておりました。

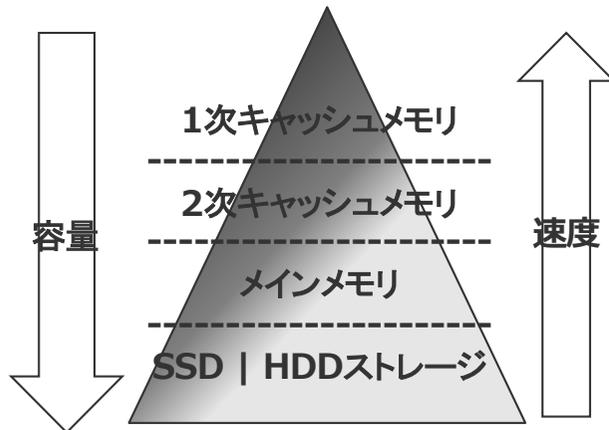
ニューラルネットワークは接続に方向があり、かつ頂点ノードと接続エッジに強弱の重みがある複雑ネットワーク構造と言えます。

弊社はこのグラフ理論を基にした製品であるIDG™やMFA™を開発しており、従来より大手企業各社様に採用されご好評をいただいております。

今回、この知見を基にして、ニューラルネットワークの分野にも適用可能な、より大規模大容量かつ高速処理可能な複雑ネットワークを扱えるライブラリ群であるONTKIT™を開発しました。



[圧縮表現による高速処理の実現]

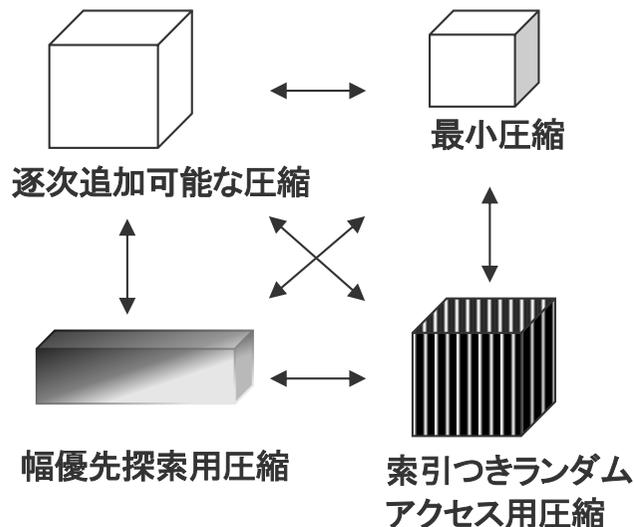


CPU/GPUいずれの処理においても、現在のコンピューティングで主要なボトルネックはデータアクセスに対する問題です。(CPU/GPUの計算処理はデータアクセス中は待ち状態であることが多いのが実態)また、特にGPU機器においてはメモリ容量を拡張すること自体が難しいのが現状です。

そこでONTROXIは、メモリ資源の効率的利用に着目し、最小限の計算資源で圧縮してその圧縮状態のまま複雑グラフを扱えるONTKITを開発しました。差分圧縮やビット演算処理によるデータの重ね合わせ表現等、さまざまなテクニックを駆使してこれを実現しハードウェア側の制限を克服しています。

これにより、大規模な複雑ネットワーク演算における2つの主要課題、
 ①例えば100億を超えるノードを扱うといった大規模性への対応、
 ②例えば1億ノードのIoTネットワークを数十分ごとに現実的なコストで全数計算といった高速性への対応、この双方の課題を同時に達成しています。

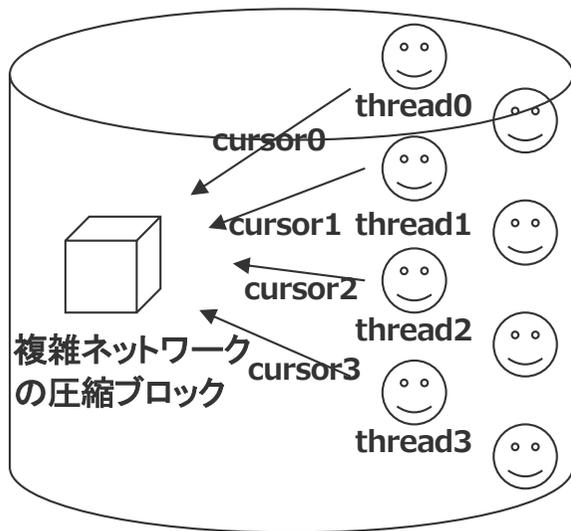
[主要探索手法に合わせた圧縮による高速処理の実現]



複雑ネットワーク系の処理では、ネットワーク探索が演算処理時間の大部分を占める為、探索処理のチューニングが高速化の鍵を握っています。

そこでONTROXIは、データ構造に着目し、主要な探索手法に合わせた最適な圧縮構造を適用可能とすることで、探索の高速化を実現しました。

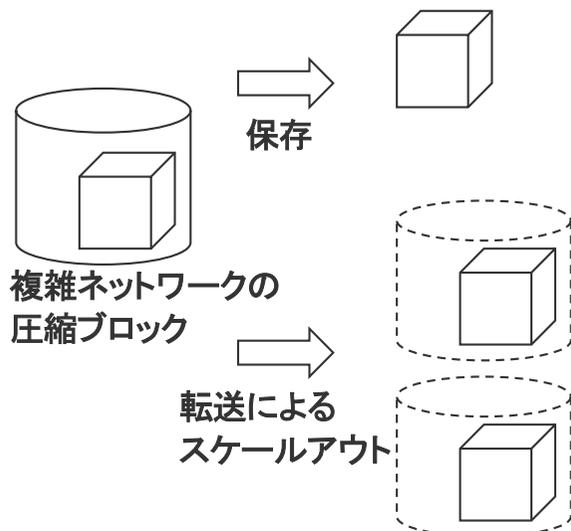
たとえば、エッジを全探索する処理で有利となる最小圧縮、ノードやエッジへのランダムアクセスが可能な索引付き圧縮、条件を満たす最短の頂点を求める処理で有利となる幅優先探索などを選択できます。



[スレッド並列時のボトルネックのない高速処理の実現]

複雑ネットワークを表現している圧縮ブロックの探索はスレッド並列の処理も想定された設計になっています。
1つのスレッドが圧縮ブロックを探索している間、別のスレッドの探索処理が待たされるようなことはありません。

通常、圧縮されたものを逐次展開するためには、現時点がどのような状態であるかを知っている必要がありますが、この情報は個々のthread側が別々に持つことができ、互いに干渉することなく探索処理が可能です。



[保存とスケールアウト]

複雑ネットワークを表現している圧縮ブロックは、圧縮されたままの単体のメモリブロックとなっており、この圧縮ブロック単体をファイルとして保存や、クラウド上の他のマシンへ転送してそのまま動作させることが可能です。

これにより、複雑グラフDBサーバの実現、計算処理のスケールアウトの実現も可能となっています。

スペック & 機能

スペック	ノードID範囲	最小値 0 から 最大値 符号なしの64Bit Intの最大値まで (UINT64_MAX)
	エッジ数範囲	0 以上から上限なし (稼働マシンのリソースに依存)
機能	ノード、エッジの追加操作	ソースノードおよびターゲットノードを指定してエッジ追加関数の呼び出し
	ノードパラメータへの対応	ソースノード、ターゲットノード別にパラメータが占有するバイト数を指定可能
	エッジパラメータへの対応	エッジパラメータが占有するバイト数を指定可能
	ノード探索操作	デザインパターン的一种であるイテレータパターンによる探索
	エッジ探索操作	デザインパターン的一种であるイテレータパターンによる探索
	探索操作のスレッド分散への対応	デザインパターン的一种であるステートパターンによる分散対応
	ノードID指定のランダムアクセス	ソースノード用またはターゲットノード用の索引付き圧縮表現でのランダムアクセスを推奨
	圧縮表現の選択適用	圧縮表現オブジェクトの指定による選択
	逐次追加可能な圧縮表現	圧縮表現メモリブロックの自動伸張のサポート
	最小圧縮表現	エッジ全探索での高速探索
	索引付き圧縮表現	ソースノードもしくはターゲットノードへの索引構築と高速アクセス
	幅優先探索用圧縮表現	幅優先探索での高速探索
	圧縮表現の変換	すべての種類の圧縮表現から別の種類の圧縮表現への相互変換
	圧縮表現のコピー	圧縮メモリブロックのコピー操作
	圧縮表現の保存	圧縮メモリブロックの保存操作
圧縮表現の読み込み	保存された圧縮メモリブロックの読み込み操作	

性能評価

[性能評価を行ったマシンのスペック: Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz 4コア、8スレッド、メインメモリ 32G]

グラフの種別	ノード数	エッジ数	圧縮ブロックサイズ (単位: bytes)	圧縮ブロック 生成速度	全エッジを舐める探索の8回繰り返し速度 並列動作スレッド数別(単位: sec)			
				逐次追加 (単位: sec)	並列数1	並列数2	並列数4	並列数8
循環グラフ	10億	10億	2,000,000,004(約2G)	9	60	30	15	15
循環グラフ	100億	100億	20,000,000,005(約20G)	96	595	296	148	150
ランダムグラフ(12%)	10万	1,249,976,784(約12億)	1,271,394,839(約1.27G)	(*注1)10	61	31	15	14
ランダムグラフ(25%)	10万	2,499,975,425(約25億)	2,500,642,464(約2.5G)	(*注1)20	99	50	25	24
ランダムグラフ(50%)	10万	4,999,963,064(約50億)	5,000,258,936(約5G)	(*注1)40	194	97	49	49
完全グラフ	10万	9,999,900,000 (約100億)	10,000,195,872(約10G)	89	388	194	97	99

(*注1) 本質的でないランダム関数処理が半分以上の時間を使ってしまうため、完全グラフの速度から逆算した参考値

[*単位: sec は1sec未満を四捨五入]

(*注2) ノードパラメータ、エッジパラメータのサイズはどちらも0バイトで評価を行った。

[循環グラフ]

ノードID1からノードID2へエッジが引かれ、ノードID2からノードID3へエッジを引くようなことを繰り返して最後のノードからノードID1へエッジが引かれたグラフ。
circleグラフとも呼ばれる。

[ランダムグラフ]

ランダムな確率を使ってノードからノードへ接続するエッジを引くか引かないかを決定したグラフ。
今回は有向で双方向にエッジを引くグラフを確率12%、25%、50%で行った。

[完全グラフ]

存在するすべてのノードからノードへエッジが引かれたグラフ。今回は完全グラフの中でも有向完全グラフで行った。これは上記のランダムグラフの確率100%とも同じグラフである。

株式会社 ONTROX

会社概要

本社：東京都世田谷区若林4丁目14-21

HP: www.ontrox.com

専門領域: 自然言語解析、複雑系解析、それ関連するミドルウェアの開発・販売、及び、テクノロジーの評価・実験検証及び、導入コンサルティング

設立: 2010年7月

サービス

プロダクト

Insight Digger : 解析及び示唆のビジュアライゼーション・エンジン提供

Mind Factor Analyzer : 思考解析エンジン提供

Zukool Feature Similarity Mapping : 特徴/関連性解析エンジン提供

ZuKool Agent : パーソナライゼーション・エンジン提供

サービス

データサイエンス コンサルティング、GPGPU化サービス

解析エンジン開発、技術審査及び技術研究開発

主要取引先

株式会社電通、株式会社電通国際情報サービス、株式会社ビデオリサーチ、GMO グループ各社
NTTコム オンラインマーケティングソリューション株式会社、GMO グループ各社 など