



株式会社 Preferred Networks

ディープラーニング研究開発を支えるGPU計算機基盤を構築 圧倒的なパフォーマンスでビジネスの創出を加速

導入サービス：深層学習用マルチノード型GPUプラットフォーム
Enterprise Cloud / Nexcenter



株式会社Preferred Networks
リサーチャー
博士(情報理工学)
土井 裕介氏
「NTTコミュニケーションズ、NTTPCコミュニケーションズは、私たちの要求を一手に引き受けて調整していただいております、高く評価しています」

- **課題**
 - ・ オンプレミスでは十分な計算資源の確保が困難だった
 - ・ AI技術の高度化により、より多くの計算資源が求められていた
- **対策**
 - ・ NVIDIA製GPUを1,024基搭載したスーパーコンピュータを構築
 - ・ ノード間接続にInfiniBandを採用し、高い処理能力を実現
- **効果**
 - ・ LINPACKのベンチマークにおいて、国内の産業領域で第1位の性能を達成
 - ・ パラレルで処理が可能となり、研究開発を効率化

課題

膨大な計算資源が要求されるAI技術の研究開発のために
高い計算能力を持った研究開発基盤の整備を決断

「IoT時代に向けた新しいコンピュータを創造する／あらゆるモノに知能を持たせ、分散知能を実現する」というミッションを掲げ、2014年3月に設立された株式会社Preferred Networks(以下、PFN)。オープンソースの深層学習フレームワークである「Chainer」の開発や提供を行うなど、AIの領域において先進的な研究開発を行う同社への注目度は高い。すでに多くの企業と協業・提携し、交通システムや製造業、バイオヘルスケアといった領域を中心に、機械学習や深層学習といった技術を用いたAI活用を推し進めている。

この機械学習や深層学習においては、計算資源の確保が重要なポイントとなる。特に大量のデータを使って学習するようなケースでは計算量も膨大になるため、高い処理能力が求められる。このニーズに応えたのがGPUと呼ばれるプロセッサであり、汎用的なCPUに比べて機械学習および深層学習で行われる計算処理を高速に実施できる利点がある。

もともとPFNでは、GPUを搭載した社内PCや、レンタルサーバーを使って機械学習や深層学習を行っていた。しかしPFNの土井裕介氏は「一般に提供されているGPUを搭載したサーバーをそのまま借りるだけでは、我々が求める規模を実現できないのではないかと考えていました」と語る。さらにAI技術の変遷も計算量に影響を及ぼすと続けた。

「これまでの深層学習は、事前に用意した正解のデータを学習させて新しい問題を解かせる『教師あり学習』が主流でした。しかし今後は、正解がない膨大なデータを学習させて法則や傾向などを導き出す、『教師なし学習』が非常に重要になってくると考えられています。この教師なし学習で、動画を使った予測学習などの研究を行っていくことを考えた場合、計算量は膨大

企業情報

社名 株式会社 Preferred Networks
事業概要 IoT時代に向けた次世代ビッグデータ技術基盤の確立を目指して2014年3月設立。独自の技術により、IoT時代に相応しいコンピュータ&ネットワークアーキテクチャーを確立すべく事業を展開している。
URL www.preferred-networks.jp/ja

なものになると考えています」

こうした技術の変化、さらにはPFNのビジネスの成長に伴い、必要となる計算資源は拡大し続けている。「それに対してキャッチアップする、あるいは追い越していくためにまとまった計算資源を投入したい」(土井氏)という考えのもと、「MN-1」と呼ばれる新たなスーパーコンピュータの整備に乗り出すことになった。

対策

NTT Com/NTTPCのGPUプラットフォームに1,024基のGPUを使ったスパコンを構築

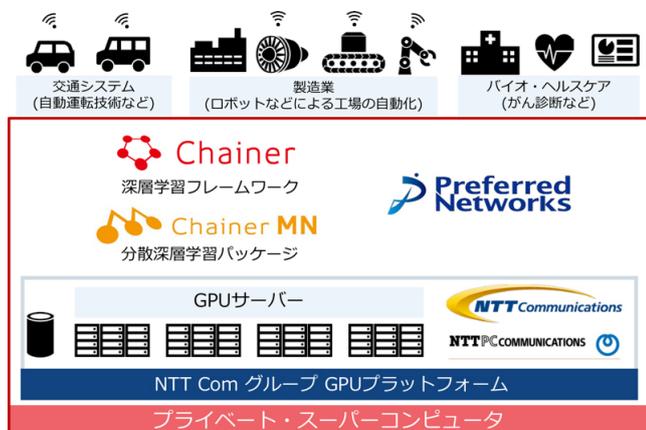
膨大な計算資源を備えたMN-1を構築するために、PFNがパートナーとして選んだのはNTTコミュニケーションズ(以下、NTT Com)とNTTPCコミュニケーションズ(以下、NTTPC)である。選定理由として、土井氏は「通信事業者としての運用の信頼性を考え、今回の構築をお願いすることに決めました」と話す。

MN-1は、NVIDIA製「Tesla P100 GPU」を1,024基搭載するという大規模なものである。これをNTT ComおよびNTTPCのマルチノード型GPUプラットフォームに構築している。

プロジェクトの途中で幾度かの方針転換もあった。そもそもAIは日進月歩で進化している技術であり、新たな研究成果が生まれたことで基盤に対する要求が変わるのは当然だ。土井氏は「我々が必要なものが何なのか、私たち自身、さらに言えば世界中の誰もわかっていないわけです。それも研究の一環であり、基盤に求められるものを明らかにしながらプロジェクトを進める必要がありました」と述べる。

こうした要件の変化の中で、特に大きかったのは個々のサーバー(ノード)のインターコネクト(相互接続)にInfiniBandを採用したことだった。極めて高性能なシステムを構築するHPC(High Performance Computing)分野において、システムを構成するコンピュータ間を接続するための技術として広く使われているのがInfiniBandであり、高速化が図れる上にレイテンシーも低く、信頼性

図 PFNの深層学習用研究開発基盤イメージ



や可用性も高いといったメリットを持つ。

PFNでは複数のノードで分散して深層学習を行う、分散深層学習の研究を進めており、分散深層学習のためのパッケージ「ChainerMN」も開発している。この研究を進める中でノード間接続のためのインターコネクトの重要性が明らかとなり、高性能なInfiniBandを使ってすべてのノードを接続することになった。

土井氏は「InfiniBandの全ノードへの導入は、非常に難易度が高かったと思います。これだけの数のGPUを使ってシステムを構築するだけでも難しいのに、さらにInfiniBandを導入することになった。なおかつ全ノードでの試験も実行することになり、NTT ComおよびNTTPCの皆さんには多大な努力をしていただきました」と振り返る。

効果

国内産業領域で第1位の性能を達成し新規ビジネスの創出や研究開発に活用

MN-1のカットオーバーに先立ち、PFNではスーパーコンピュータの世界で標準的なベンチマークとなっている「LINPACK」と、「ImageNet」と呼ばれる一般に広く使われている画像分類データセットを自社のChainerMNで学習し、それに要する時間を計測するといったテストを実施している。

LINPACKは大規模行列計算を行うベンチマークであり、パラメーターのチューニングによって結果は大きく変化する。PFNではそのチューニングに機械学習の手法を用いた。その結果、MN-1は1.39ペタフロップスという性能を記録し、2017年11月のスーパーコンピュータ性能ランキングを示すTOP500リストにおいて、日本国内の産業領域で第1位、研究用などすべてのスーパーコンピュータを含むランキングにおいても世界91位を獲得した。

ImageNetの学習についても、MN-1とChainerMNの組み合わせによってわずか15分で学習を完了させることができた。これは同様の研究報告として、それまで最速とされていた学習時間を大幅に短縮するものだった。

実際の研究開発においても、MN-1がもたらす価値は大きかったようだ。土井氏は「研究ではさまざまなパターンを試す必要があります。従来はそれをシーケンシャルに行っていましたが、MN-1であればパラレルに処理することができるため、研究開発の効率化が図れています」と話す。さらにPFNはNTT Com、NTTPCと連携し、MN-1を拡張した「MN-1b」の構築を進めている。MN-1bでは「NVIDIA Tesla V100 32GB」を512基稼働させる予定であり、拡張部分だけで、半精度浮動小数点数における性能では、従来の約3倍のピーク性能が見込まれている。

PFNのAI関連技術の研究開発やビジネスの推進において、NTT ComおよびNTTPCのマルチノード型GPUプラットフォームで運用されるMN-1が果たす役割は極めて大きい。

お問い合わせ先

NTTコミュニケーションズ株式会社

ホームページ www.ntt.com/business

●記載内容は2018年7月現在のものです。
●表記のサービス内容は予告なく変更することがありますので、お申し込み時にご確認ください。
●記載されている会社名や製品名は、各社の商標または登録商標です。