



# 革新的なAIプロダクトの創出を NVIDIA DGX-1™ がバックアップ HEROZ株式会社様

将棋のAI(人工知能)開発で知られるHEROZは、AIを活用して新たなサービスの創造するプロフェッショナルのエンジニア集団だ。GPUコンピューティングで革新的な高速処理を実現する「NVIDIA DGX-1™」が、エンジニア達の開発を最大限に加速させる。



「驚きを心に」をコンセプトに掲げ、世界を驚かすサービスの創出を目指して2009年に創業したHEROZ株式会社(以下HEROZ)。将棋の現役名人に勝利した将棋AI「Ponanza」の開発者も所属するなど、AIアルゴリズム開発のトップランナーが集うIT企業である。

BtoCでは「将棋ウォーズ」や「囲碁ウォーズ」、「ポケモンコマスター」など、AI関連技術を導入した多彩な頭脳ゲームを提供。BtoBでは、将棋AIの研究・開発で培ったAIテクノロジーから独自AI「HEROZ Kishin」を開発し、AIを組み合わせたインターネットサービスの企画・開発・運営を手掛けることで、さまざまな業界の課題解決にひと役買っている。

HEROZ 執行役員 開発部長 井口圭一氏によれば、将棋以外の分野にAIを適用し始めたのは2015年頃から。現在は、開発メンバーの半数以上がBtoBに関連した業務を手掛ける体制にシフトしている。

「業界などを絞るような考えはとくになく、幅広い分野でいろいろなことにチャレンジしていきたいと思っています。もともと得意だったエンターテインメント系はもちろんのこと、建築や金融、品質管理、物流、人材関連などで取り組みを始め、すでいくつかの領域で成果が出始めています」(井口氏)

## サービスの機能向上を見据え 128GPU環境を実現

現在、HEROZが注力しているサービスのひとつに自動監視・異常検知ツール「HEROZ Kishin Monitor」がある。このツールではまずプラットフォームを監視してさまざまなデータを取得し、そのデータをAIに学習さ

せて将来の動向を予測。そして、実際の数値がこの予測の範囲から大きく外れた場合は「何か異常が起きているのではないかと判断してアラートを出す仕組みを備えている。

例えば、Webサーバーの監視であればアクセス数やCPU使用率、ネットワークの状態などのデータをAIが取得・学習して将来の傾向予測を出力し、その予想と比較して不自然なアクセスなどがあった場合には管理者へ通知する。Webサーバー以外にも、生産ラインや金融取引など、幅広い分野での監視や異常検知に利用できるのがポイントだ。

サービスとしてはすでにCPUベースで展開しており、Webサーバーの場合は監視タイミングが5分に1回、将来の傾向予測が過去3ヵ月分のデータを使って1時間おきに再学習をかけている。ただし、Webサーバーに限らず、あらゆるトラブルやインシデントは可能な限り早く発見することが求められることから、「本来であれば監視頻度は少なくとも1分に1回が望ましいですし、再学習のタイミングも同じレベルで実現したい」というのが井口氏の本音だった。

「今回のAI学習ではニューラルネットワークが非常に効果的であることから、リカレントニューラルネットワークの一種であるLSTM(Long Short-Term Memory)が採用されています。しかし、CPUでは『LSTMを効率的に処理できない』というウィークポイントがあるため、CPUでは1分間隔での再学習はほぼ不可能です。そのため、より高速にLSTMを処理できる優れたGPUコンピューティングの環境が求められました」(井口氏)

また、HEROZではもともと「GeForce GTX 1080」や「GeForce GTX 960」など、



HEROZ株式会社 執行役員 開発部長 井口圭一氏

市販のGPUを複数台用意して開発を進めていたが、リソース不足が否めないことから2018年5月に高性能GPU「NVIDIA® Tesla® V100-PCI-Express」を8基搭載したジュー・アドバンス製の「MAS-XE5-SV4U/8X」を導入し、開発環境を大幅に強化していた。しかし、それだけの環境を用意してもGPUの利用状況はつねにフル回転。社内エンジニアの飽くなきニーズを満たすまでには至らなかった。自身もエンジニアである井口氏は、「普通であれば、NVIDIA® Tesla® V100×8基×4台でも十分な環境だと思います。でも、弊社のエンジニア達にはこれでも不十分だったようです。それだけ、チャレンジしたいことがあるということでしょう」と苦笑する。

そういった理由が重なったことから、HEROZでは約半年という短いスパンで、GPUコンピューティング環境のさらなる増強を決行。2018年11月にディープラーニング専用スーパーアプライアンスコンピュータ「NVIDIA DGX-1™」をデータセンターに12台追加導入し、先の導入と合わせて計算機16台でGPUの合計が128基という飛び抜けた計算機環境を構築した。

## 1日かかるシミュレーションを 数分程度でこなせる高性能

前出の「自動監視・異常検知ツール」は、CPUベースでWebサーバーの監視が5分

に1回、傾向予測の再学習が1時間おきとなるが、井口氏によれば「NVIDIA DGX-1™によるGPUコンピューティングによって、実験ベースながら1分での再学習に成功しています」とのこと。今後はSaaS(Software as a Service) 展開中の「HEROZ Kishin Monitor」への適用を見込んでおり、開発だけでなくサービス基盤としてのNVIDIA DGX-1™活用も見据えている。

さらに、取得するデータは項目が増えるため、CPUベースでの運用時は取得項目を絞らなければならないケースもあるそうだ。しかし、NVIDIA DGX-1™によって「さまざまな項目を多様に組み合わせた監視も可能になるので、より精度の高い監視や、ユーザーの細かいニーズにも対応した幅広い運用が可能になるでしょう」と(井口氏)。

また、高性能な処理能力を手にしたことで、建設業界でのAI活用にも新たな光がさしているようだ。昨今、業界ではパソコン上でさまざまな建物の3次元モデルを設計したあと、それを構造計算ソフトにかけて建物の安全性を検証するようになっている。

HEROZでは、その構造設計の分野で強化学習を利用し、より最適な設計を効率的に生み出そうというサービスを研究、開発している。ただし、そのAIの強化学習を進めるにあたっては、「CPUによる処理だと、大規模な建物の構造計算は数時間かかり、1日1回程度しかできません」と(井口氏)というボトルネックがあった。そのため、強化学習を適用するには、例えば1000通りのパターンを事前に作成し、それを1000台のサーバーで並行処理。処理が終わったら、その結果を集計して改善を施し、また新しいパターンを作成して次のサイクルに移るという方法をとっていた。このやり方では1000通りの構造パターンが非同期で処理されるため、改善のサイクルを1000回回すよりも「効率は良くありません」と井口氏は分析する。

しかし、NVIDIA DGX-1™を利用すれば「1回のシミュレーションが数分単位で処理

できると見込まれます」と(井口氏)。これにより、1つのパターンで連続的にサイクルを回す手法をとって改善効率を高められるとともに、AIの強化学習も飛躍的にスピードアップできると予想されている。

### 導入しやすさも魅力のひとつ 魅力的な環境は採用にも影響

GPUコンピューティングによる処理性能の向上以外にも、井口氏はいくつかメリットを感じている。例えば、今回のケースでは12台のNVIDIA DGX-1™を外部のデータセンターに導入したため、性能や規模感の面からやや特殊な準備も求められることとなった。しかし、そういった部分はNVIDIAエリートパートナーのジーデップ・アドバンスが柔軟かつ迅速に対応。「手厚いサポートがあったおかげで、特に困ったこともなくスムーズに進められました」と井口氏は振り返る。

また、意外なメリットとなったのが「人事採用への影響」。自分の実力を活かしたいと考えるAIエンジニアにとって、128基のGPUを備えた計算機環境はとても魅力的に映るようだ。

「大企業であれば同クラス的环境を持っているところもあるでしょうが、人数比で考えると必ずしも充実した環境とはいえません。また、研究所レベルだと1台のマシンを何十人単位で共有する場合もあります。そういったケースと比較すると、弊社は非常に充実した環境といえるでしょう」と(井口氏)

HEROZは、今後も「HEROZ Kishin」で新しいプロダクトを創出し、BtoBのサービスでさらに幅広い展開を進めていく方針だ。例えば、建設業界で対応できているのは構

### HEROZ株式会社様 導入モデル

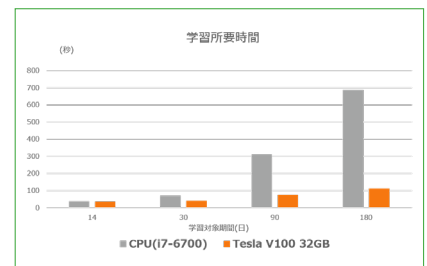
#### NVIDIA DGX-1™



NVIDIA Volta™と共に提供されるNVIDIA DGX-1™は、AI研究で不可欠となる高度な計算処理において、データセンターでのディープラーニングワークフローを合理化するように設計されています。業界をリードする大規模AI性能を余すところなく引き出し、通常のCPUサーバー25ラック分の性能を1台で実現します。



NVIDIA® Tesla® V100を8基搭載するNVIDIA DGX-1™は、データセンターに設置されている



LSTMにおいて、CPUとGPUの学習所要時間を比較したグラフ。学習対象期間が180日の場合、GPUでは84%の時間削減になる

造設計の一部だけにとどまらず、施工管理や現場の安全確保、作業の最適化なども、AI解析によって効率化がはかれると考えている。

「まだ試していませんが、今回は12台をEDRのインフィニバンドで連携させたマルチノード構成も実現しているだけに、どこまで大規模な処理に対応できるのかも楽しみです。今後もNVIDIA DGX-1™でさまざまな取り組みにチャレンジし、AIの研究・開発を加速させてあらゆる業界の課題解決をサポートしていきます」と(井口氏)