



ディープラーニングをあまねく世に

LeapMindが OKIとの共創で AIエッジ時代を加速する



エッジ端末に適したディープラーニング技術を得意とする2012年創業のLeapMind(リープマインド)。高い推論精度と応答性を維持しながら、コンパクトかつローパワーなハードウェア実装を実現する、独自の「極小量子化」技術が特長だ。この技術を産業分野などに幅広く応用していくため、同社は沖電気工業(以下、OKI)とAIエッジパートナーを結び、エッジ用ハードウェアの一つとして、OKIのAIエッジコンピューター「AE2100」を採用した。

センサーを使って装置や環境からデータを集め、それらのデータを分析し、価値を創出するIoT(Internet of Things)が、実際の応用フェーズへと移ってきた。その過程で明らかになってきたのがエッジ処理の重要性だ。装置やセンサーに近いところ(エッジ側)で分析処理を行うことで、サーバーやクラウドへのデータ・トラフィックを削減しながら遅延のないフィードバックを実現できるからだ。そのためいわゆるエッジコンピューティングが、現在のIoTアーキテクチャーの潮流の一つになっている。

さらに最近になって、エッジ部分にディープラーニングを応用しようという動きが加速してきた。従来のパターンマッチングや機械学習といった手法に対して、画像や音声などの認識精度を高めるのが狙いである。

2012年に創業したLeapMindは、エッジ部分におけるディープラーニング技術の先駆けともいえるAIベンチャー企業の一社だ。サーバーやクラウドに比べてコンピューティング・リソースや使える電力が限られるエッジにおいても、高速かつ高精度な推論を実現する独自の軽量化技術を強みとしている。また、上流ネットワークとの接続がない環境にも対応しており、オフラインでもエッジ側だけで処理を完結できるのも同社のソリューションの特長である。

「当社は、『機械学習を使った今までにないデバイスをあまねく世に広める』をミッションに、コンパクトかつローパワーな組み込み向けディープラーニング技術の提供や共同研究を行っています」と、LeapMindの安村修一氏は説明する。業種や業界を問わず、画像解析のほか、音声解析や時系列解析など、幅広い応用を手掛けているという。



LeapMind株式会社
General Manager,
Co-development Division

安村 修一氏

推論モデルサイズを最大99%削減

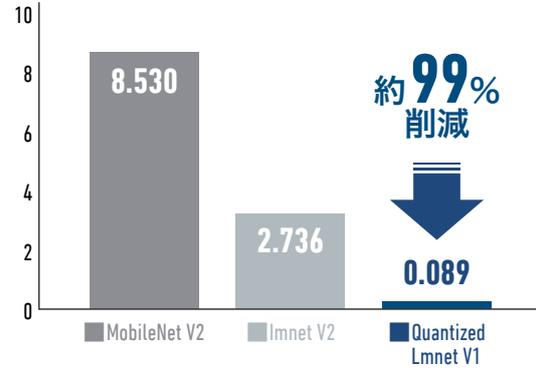
LeapMindは、DNN(Deep Neural Network)の「極小量子化」に強みを持っている。

DNNは、入力層、複数の中間層(隠れ層)、および出力層で構成されている。ある層のノードと次の層のノードは重み付きシナプスで結ばれており、一般的にはこの重みは32ビット浮動小数点で表現されるが、計算量が多くなってしまい、リソースが限られる組み込み用途には向いていない。

LeapMindでは、活性化関数を2ビット+重みを1ビットで表現するなど、DNNの大幅な軽量化を実現する方式を開発した(図1)。「推論モデルの軽量化に関しては、ビット数削減のほかに、シナプスを省略する枝切り(プルーニング)などのさまざまな手法が提案されていますが、当社の極小量子化は性能や精度の面でも世界トップレベルと考えています」と安村氏は説明する。

モデルサイズを大幅に圧縮

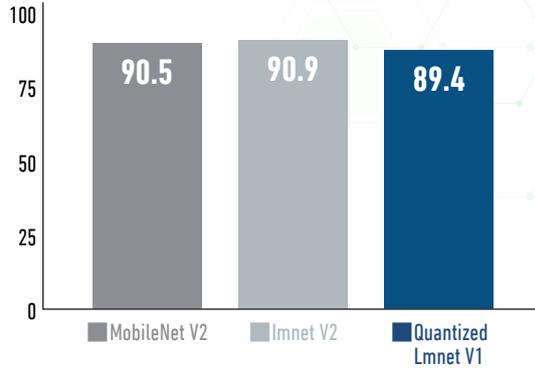
モデルサイズの比較
(MB)



※各モデルの実行には同一FPGAボードを使用

精度は同等

精度の比較
(%)



※CIFAR-10を使ってデータ検証

図1 最高で99%もの推論モデルの小型化を実現するLeapMindの極小量子化技術の例

この技術を応用した推論器のソフトウェアを「Blueoil (ブルーオイル)」として提供しており、従来の軽量化環境に対してモデルサイズの99%削減を実現した例もあるという。その場合でも認識精度はわずか1%程度の低下にとどまっており、実用的に問題ないことが確認されている。

推論モデルの軽量化はエッジのハードウェア実装に多くのメリットをもたらす。ロジックサイズの小小型化、消費電力の低減、メモリ容量の削減などが図れるため、リソースの限られるエッジに最適といえる。マイクロコンピュータのコアを内蔵した最もローエンドのFPGA (Field Programmable Gate Array) デバイスでも、十分な性能が得られる場合が多いという。

同社はこの極小量子化技術を軸に、すべてのモノにディープラーニングを組み込む「DoT (Deep Learning of Things)」の実現を掲げながら、ディープラーニングのコンサルティングに始まり、データ作成、推論モデル構築、モデル圧縮、ハードウェア向け最適化、IP 開発、実装、チューニングまで、一連の開発をトータルで提供している。

OKIとAIエッジパートナーシップを提携

LeapMindは、同社のソリューションを載せるエッジハードウェアとして、これまでのFPGA、ASIC(Application Specific Integrated Circuit)、ハードIPなどの実装方法に加え、OKIが開発したAIエッジコンピューター「AE2100」(図2)を選択肢の一つに追加した。

「産業分野で多くの実績を持つOKIが推進するAIエッジコンピューティングの考え方に賛同し、今回パートナーシップを結ぶ運びになりました。OKIの『AE2100』は、産業用として開発されただけあって信頼性が高く、入出力インターフェースも豊富で、エッジコンピューティングに最適なハードウェアの一つと捉えています」と安村氏は理由を説明する。



図2 OKIが開発したAIエッジコンピューター「AE2100」の外観

ここで「AE2100」の概要について説明しておこう。「AE2100」は、AIエッジに最適化された小型の産業用コンピューターで、Intel Atom E3950 プロセッサをベースに、AIアクセラレーターである「インテル® Movidius™ Myriad™ X VPU (Vision Processing Unit)」をオプションで搭載できる。センサーやアクチュエーターなどの接続用に複数のI/Oポートを備えているほか、OKIが開発した920MHz帯マルチホップ無線「SmartHop」を利用できるのも特長だ。

ソフトウェアとしては、IoTの実行環境としてMicrosoftの「Azure IoT Edge」と、ディープラーニングのツールキット(オプティマイザー)としてインテルの「OpenVINO™ ツールキット」が用意される。なお、LeapMindでは、アプリケーションや顧客要件に応じて、OpenVINO™ ツールキットとBlueoilとを使い分けていく考えだ。

OKIでは、AIを必要とする顧客に対してより幅広いソリューションを提供するために、「AE2100」を中心にエコシステムの構築を進めている。57社がOKIのAIエッジパートナーとして名を連ねており(2020年2月末時点)、LeapMindもその一社である。

ディープラーニングをあまねく世に広める

安村氏は今後の取り組みを次のように展望する。「ディープラーニングやAIに対する注目が高まっている中で、さまざまなアイデアやニーズを具体的な価値へとつなげていくことがこれからの課題です。当社の強みである軽量化技術にとどまらず、実運用までカバーするワンストップサービスを通じて、ディープラーニングをあまねく世に広めていきたいと考えています」。

また、とくに産業分野への応用に関しては、AIエッジコンピューティングを推進するOKIとの「共創」に期待していると述べている。

AIエッジの時代に向けて、独自の技術力に裏打ちされたLeapMindのこれからの取り組みが注目される。



パートナー企業 概要



LeapMind株式会社

所在地：東京都渋谷区円山町28-1 渋谷道玄坂スカイビル 3F

設立：2012年

従業員数：70名(2020年3月12日時点)

事業内容：ソフトウェアとハードウェアの両領域での、ディープラーニングのコンサルティング、モデル構築、実装、運用と、半導体IPの開発

<https://leapmind.io/>

お問い合わせ

OKI *Open up your dreams*

沖電気工業株式会社

<https://www.oki.com/jp/Aledge/>