

## 奈良先端科学技術大学院大学が、 機械学習とクラウドソーシングの融合により 多元ビッグデータ解析の応用を推進

AMD EPYC™ プロセッサ搭載HPE ProLiant DL385 Gen10を採用し、  
教師データの事前作成を最小限に非構造化データのカテゴリ分類を可能に

“『機械学習とクラウドソーシングを融合させた分類モデル構築』では大容量メモリが大きな威力を発揮しました”

—国立大学法人  
奈良先端科学技術大学院大学  
先端科学技術研究科 情報科学領域  
特任准教授  
博士(工学) 鈴木 優 氏

知能コミュニケーション研究室  
ビッグデータプロジェクト  
データ駆動型サイエンス創造センター



### 目的

「多元ビッグデータ解析」による新たな知見の創出。  
これを支えるデータマイニング・機械学習および  
解析のためのシステム環境を構築し、高速かつ高精度  
の処理を追求する。

### アプローチ

データマイニング・機械学習および解析のパフォー  
マンス向上に結びつく、大容量メモリを搭載可能な  
サーバーを選定。CPUとメモリを結ぶインターコネ  
クトの帯域幅を重視。

### ITの効果

- AMD EPYC™ プロセッサ搭載HPE ProLiant DL385 Gen10を採用し、最大4TBメモリを搭載可能な「多元ビッグデータ解析」環境を構築
- 高密度かつコストパフォーマンスに優れた仮想サーバー環境を構築し複数のシステムを統合
- HPE DirectPlusで仕様を相談しながら見積依頼・仕様決定・納入までを短期間化

### ビジネスの効果

- 機械学習とクラウドソーシングを融合させた分類モデルの確立に寄与
- 「訪日外国人向け観光情報のリアルタイム配信サービス」において京都府下での実証実験を推進
- 投資対効果に優れた「多元ビッグデータ解析」環境を実現、新たな知の創出への貢献が期待

国立大学法人  
奈良先端科学技術大学院  
大学

教育



国立大学法人  
奈良先端科学技術大学院大学  
先端科学技術研究科 情報科学領域  
特任准教授  
博士(工学) 鈴木 優氏

知能コミュニケーション研究室  
ビッグデータプロジェクト  
データ駆動型サイエンス創造センター



日本AMD株式会社  
エンタープライズ・ソリューション営業本部  
セールスエンジニアリング担当  
マネージャー 関根 正人氏

奈良先端科学技術大学院大学 (NAIST) 先端科学技術研究科 情報科学領域では、鈴木優 特任准教授を中心とする研究グループが「多元ビッグデータ解析」を応用した新たな知見の創出に取り組んでいる。研究グループでは、AMD EPYC™ プロセッサ搭載HPE ProLiant DL385 Gen10を採用し、データマイニング・機械学習および解析のためのシステム環境を構築。CPUとメモリを高速に結ぶアーキテクチャーの優位性を活かした高速かつ高精度の処理を実現している。また、HPEのオンラインストア「HPE DirectPlus」を利用して短期間でスムーズな調達を行った。

## チャレンジ

### 「多元ビッグデータ解析」による 新たな価値創出と実社会への貢献

奈良先端科学技術大学院大学 (NAIST) は、情報科学、バイオサイエンス、物質創成科学の3分野に特化した国立大学院大学である。1991年の創設以来、世界レベルの研究活動と人材育成の成果が各界より高く評価されている。2018年、NAISTは3分野の融合領域教育を強化するため「先端科学技術研究科」による1研究科体制へ移行した。情報科学領域で特任准教授を務める工学博士の鈴木優氏は次のように紹介する。

「より大きな研究成果と実社会への貢献を追求するために、3分野の研究者が有機的に連携する体制が整えられました。領域横断的にひとつの研究テーマに取り組む機会も増えています。情報科学領域は、バイオや物質の研究者とともに実験データをより効果的に解析する手法を検討し、得られた知識をもとにそれぞれの研究を新しいステージに導く役割も担っています」

鈴木氏が所属する「知能コミュニケーション研究室」では、人と人、人とITのコミュニケーションを支援する多様な技術を扱っている。実現性を重視したプロトタイプシステムの構築から、サービスシステムへの実装・評価までを網羅するとともに、産業界との連携も担う。その中で、鈴木氏を中心とする研究グループが取り組んでいるのは、「多元ビッグデータ解析」の応用による新たな知見の創出である。

「私たちは、多様な生活社会情報、バイオ情報、物質情報などをデータアナリティクスやデータマイニングの対象として扱っています。中でも、センサー、ソーシャルネットワーク、ライフログなど複数のデータを組み合わせた『多元ビッグデータ解析』には、大きな可能性があると考えています」(鈴木氏)

Wikipediaの情報信頼性検証、マルチメディア情報検索システムの開発、行動履歴の検索システム、メタ検索エンジンの開発、クラウドソーシングにおける品質管理など、鈴木氏が取り組む「多元ビッグデータ解析」の応用範囲は幅広い。現在進めているプロジェクトのひとつを、鈴木氏は次のように紹介する。

「多元ビッグデータ解析を応用し、『訪日外国人向け観光情報のリアルタイム配信サービス』の実用化を目指しています。観光地に設置されたセンサーの情報とTwitterに書き込まれた情報を結びつけ、たとえばある観光名所の見どころや天候、混雑具合などをリアルタイムで案内するサービスを想定しています。固定的な情報でなく、時々刻々と変わっていく状況を常に把握しながら、観光客が求めている情報を適切に配信することが大きな狙いです」

2018年2月、データマイニング・機械学習とクラウドソーシングをより積極的に活用するための環境が構築された。採用されたのはAMD EPYC™ プロセッサ搭載「HPE ProLiant DL385 Gen10」である。

## ソリューション

### AMD EPYC™ プロセッサ搭載の HPE ProLiant DL385 Gen10サーバーを採用

HPE ProLiant DL385 Gen10サーバーは、ハードウェアレベルでセキュリティ機能を強化し、シリコンレベルの信頼性 (Silicon Root of Trust) を備えたHPE Gen10 サーバープラットフォームである。AMD EPYC™ プロセッサを採用し、2ソケット/2Uサーバーながら最大64コアと最大4TBメモリを搭載可能だ。

「観光のプロジェクトでは、センサー情報とツイートという異なるデータを組み合わせ、機械学習とクラウドソーシングにより新しい知識を得ることが出発点となります。食事、交通、天気、混雑、景観といった観光に役立つツイートの分類には、クラウドソーシングにより延べ5,000人以上が参加しました。そして、HPE ProLiant DL385 Gen10上に構築した環境で、意味表現学習 (word2vec) などの手法を用いて膨大なツイートのカテゴリ分類を行っています」(鈴木氏)

鈴木氏が取り組む「多元ビッグデータ解析」の実行環境では、サーバーのメモリ容量とCPUコアあたりの性能が重要になる。期待通りのパフォーマンスを発揮させるには、メモリやI/Oにボトルネックが発生をさせないこともポイントだ。

「ランダムな読み書きが頻繁に発生するワークロードなので、可能な限りディスクへのアクセスを減らしてメモリ上で処理したいと考えました。

AMD EPYC™ プロセッサ搭載のHPE ProLiant DL385 Gen10サーバーは、同等クラスのサーバーと比較してより大容量のメモリを搭載できることにまず注目しました。そして、CPUとメモリを広帯域で結ぶアーキテクチャーに優位性があると考えました」(鈴木氏)

日本AMDの関根正人氏は次のように話す。

「AMD EPYC™ プロセッサは、CPUコア、メモリチャネル、PCI Express、セキュリティチップ等を統合したSoC (System-on-a-Chip) であることが大きな特長です。8チャネルのDDR4メモリ、128レーンのPCI Expressを備え、チップセット不要でCPUとメモリ間を170GB/sで直結します。まさに、本要件に最適な資質を備えていると言えるでしょう」

HPE ProLiant DL385 Gen10サーバーは、HPEとAMDが開発段階から緊密に協力してハイパフォーマンス&ローコストを徹底的に追求し、2ソケットサーバーの「コストパフォーマンスの常識を変える製品」として完成させた。AMD EPYC™ プロセッサの高いパフォーマンスは、SPECrate2017 FP baseで257、SPECrate2017 FP Peakで270という2つの世界記録を達成するなど、これまでの2ソケットサーバーの常識を打ち破るものだ。

#### 機械学習とクラウドソーシングを 融合させた分類モデルの構築

京都府とのプロジェクトでは、10万以上のツイート中から京都観光に関係のあるツイートだけを分類するテストを実施した。

「教師あり機械学習により分類したツイートを、クラウドソーシングによって人に評価してもらい、関係性の有無を学習データとしてリアルタイムで取り込みます。このプロセスを、繰り返し分類器に反映させることで分類精度を高めていくのです。この「機械学習とクラウドソーシングを融合させた分類モデル構築」では大容量メモリが大きな威力を発揮しました。優れたメモリアーキテクチャーが、AMD EPYC™ プロセッサの

性能を最大限引き出していると思います」と鈴木氏は話す。

大容量メモリを活かすには、メモリそのものの性能を最大化することが重要だ。このテストでは、「機械学習とクラウドソーシングを融合させた分類モデル構築」での大容量メモリの有効性が改めて確認された形だが、サービス化に向けた要求も明らかになった。

「学習を重ねるごとに分類器が消費するメモリは大きくなっていきます。リアルタイムでのサービス提供に向けて、より大容量のメモリが必要になるはず。最大4TBまでメモリを拡張可能なHPE ProLiant DL385 Gen10サーバーの優位性が、さらに発揮されることになるでしょう」(関根氏)

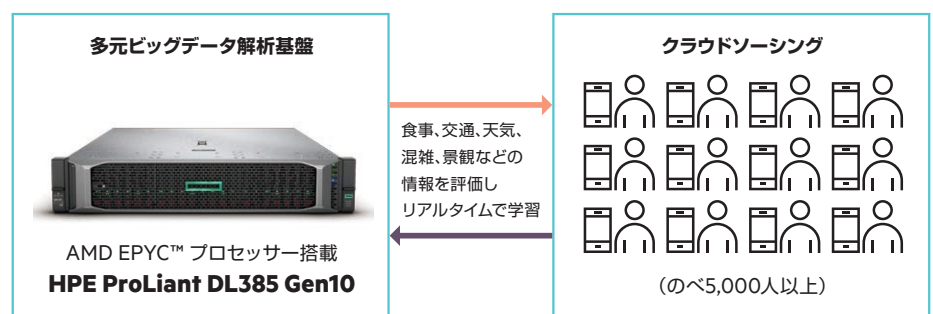
また、サーバーの選定に際しては、HPE ProLiant DL385 Gen10サーバーが備えるセキュリティ機能は「大きな安心材料になった」(鈴木氏)という。

「自治体や企業との共同プロジェクトの中で機密情報を扱う機会が増えていますので、サーバー自体が備えるセキュリティ機能は今後ますます重要になるはず」

「世界標準の安心サーバー」を掲げるHPE Gen10 サーバー プラットフォームは、マネジメントプロセッサIntegrated Lights-Out 5 (iLO 5) を信頼の起点 (Silicon Root of Trust) として、ハードウェアレベルでシステムの完全性を保証している。HPE ProLiant DL385 Gen10サーバーでは、iLO 5によるファームウェア/ハードウェアの保護に、AMD EPYC™ プロセッサならではのメモリ保護という安心が加わった。

「AMD EPYC™ プロセッサ上には『AMD Secure Processor』が統合されています。OSと独立して動作し、暗号キーの管理を行うことができます。セキュア・メモリ・エンクリプション (SME) は、暗号キーとAESエンジンによってオーバーヘッドを1%程度に抑えながらページ

### 京都府との共同プロジェクト「訪日外国人向け観光情報のリアルタイム配信サービス」



#### 機械学習とクラウドソーシングを 融合させたツイート分類

- ・教師あり機械学習によりツイートを分類
- ・クラウドソーシングによってのべ5,000人以上が評価
- ・関係性の有無をリアルタイムで学習
- ・繰り返し分類木に反映させることで分類精度を向上

## Case study

国立大学法人  
奈良先端科学技術大学院  
大学

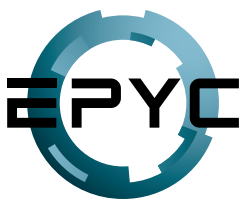
## 業界

教育

## ソリューション概略

### 導入サービス

- HPE ProLiant DL385 Gen10



### AMD EPYC™ プロセッサ搭載 HPE ProLiant DL385 Gen10

### Integrated Lights-Out 5搭載 「世界標準の安心サーバー」

- AMD EPYC™ プロセッサー  
(最大2CPU/64コア)
- 8チャンネルのDDR4メモリ(最大4TB)
- 128レーンのPCI Express
- チップセット不要でCPUとメモリ間を  
170GB/sで直結

### お問い合わせはこちら

カスタマー・インフォメーションセンター  
**0120-268-186** (または03-5749-8279)  
月曜日～金曜日 9:00～19:00  
(土曜日、日曜日、祝日、年末年始、および5月1日お休み)

日本ヒューレット・パッカード株式会社  
〒136-8711 東京都江東区大島 2-2-1



ぜひ登録ください

**Hewlett Packard  
Enterprise**

“企業や組織で管理されている構造化データに、まったく違う視点から非構造化データを組み合わせることで、新しい知識を創出できる可能性は極めて大きく幅広いと考えています。また、機械学習とクラウドソーシングを融合するアプローチによって、教師データを逐次的に作成しながら非構造化データのカテゴリ分類を可能にする道がひらかれました”

国立大学法人 奈良先端科学技術大学院大学 先端科学技術研究科 情報科学領域  
特任准教授 博士(工学) 鈴木 優 氏  
知能コミュニケーション研究室 ビッグデータプロジェクト データ駆動型サイエンス創造センター

テーブル単位でメモリ上のデータを暗号化し物理メモリへの攻撃を保護することができます。

また、セキュア・エンクリプテッド・バーチャライゼーション(SEV)と呼ばれる、仮想OS単位で暗号化キーを変え、仮想OS毎の物理メモリを暗号化することも可能になります。この機能を使っても、パフォーマンスへの影響は1%程度に抑えることができます(AMD 関根氏)

## ベネフィット

### HPE DirectPlusで仕様を相談しながら スムーズかつ短期間での導入を実現

HPE ProLiant DL385 Gen10サーバーの調達に際して、鈴木氏はオンラインストア「HPE DirectPlus」を利用した。HPE DirectPlusでは、ユーザー自身が構成を決めてオンラインで発注できるが、コールセンターの専門スタッフと電話やメールで相談しながら構成を検討することも可能だ。鈴木氏の最初のアクションは電話だったという。

「まずメモリ容量を決めてサーバーの構成を詰めていきました。以前からHPE DirectPlusを利用してきましたが、電話で仕様を相談しながら翌日には見積を提示してもらえるなど、今回もスムーズかつ短期間で導入することができました」と鈴木氏は評価する。

HPE DirectPlusでは、見積は翌営業日までに提供され、正式受注から5営業日(一部地域を除く)で、CPU・メモリ・ディスクなど構成済みのサー

バーが納品される。また、HPE DirectPlusで作成した見積・構成IDを引き継いで、販売代理店経由で調達することも可能だ。

「学内での検討を含めて、最初の間合せから2週間かからずに仕様通りのHPE ProLiant DL385 Gen10サーバーが届きました。HPE DirectPlusでは手間なく発注でき納品も早いので、忙しい年度末の調達時にはとても助かります」と鈴木氏は笑顔を見せる。

鈴木氏とその研究グループは、「多元ビッグデータ解析」のさらに幅広い応用に取り組んでいく考えだ。2018年秋には、京都府のプロジェクトの大規模な実証実験も計画されている。鈴木氏は次のように語って締めくくった。

「企業や組織で管理されている構造化データに、まったく違う視点から非構造化データを組み合わせることで、新しい知識を創出できる可能性は極めて大きく幅広いと考えています。また、機械学習とクラウドソーシングを融合するアプローチによって、教師データを逐次的に作成しながら非構造化データのカテゴリ分類を可能にする道がひらかれました。HPEサーバー製品には、より強力なコンピュータパワーを、より優れたコストパフォーマンスで実現してもらえることを期待しています」

### 詳しい情報

**HPE ProLiant DL385 Gen10**についてはこちら  
[hpe.com/jp/gen10](http://hpe.com/jp/gen10)

お客様から販売パートナー様まで強力にサポート

### HPE DirectPlusコールセンター

オンラインストア「HPE DirectPlus」はWEBでもコールセンターでもメールでも、さらにはお近くの販売代理店からでも、ご購入いただけます。コールセンターでは経験・知識豊富な専門スタッフが、製品選定、構成、お見積をサポート。お客様から販売パートナー様までのお悩みにお応えします。まずは、お気軽にご相談ください。



**0120-215-542**

携帯電話・PHS **03-5749-8462**

受付時間: 月曜日～金曜日 9:00～19:00  
(土曜日、日曜日、祝日、年末年始、および5月1日お休み)

© Copyright 2018 Hewlett Packard Enterprise Development LP

AMD、AMD Arrowのロゴ、AMD EPYC、その組み合わせは、米国および/または他の地域での Advanced Micro Devices, Inc.の商標です。本書の内容は、将来予告なく変更されることがあります。日本ヒューレット・パッカード製品およびサービスに対する保証については、当該製品およびサービスの保証規定書に記載されています。本書のいかなる内容も、新たな保証を追加するものではありません。日本ヒューレット・パッカードは、本書中の技術的あるいは校正上の誤り、脱字に対して、責任を負いかねますのでご了承ください。記載されている会社名および商品名は、各社の商標または登録商標です。

CPX13324-01 記載事項は個別に明記された場合を除き2018年5月現在のものです。