

数十万の選択肢の中から適切な解を 世界最速かつ高精度に予測するAI技術をOSSとして公開

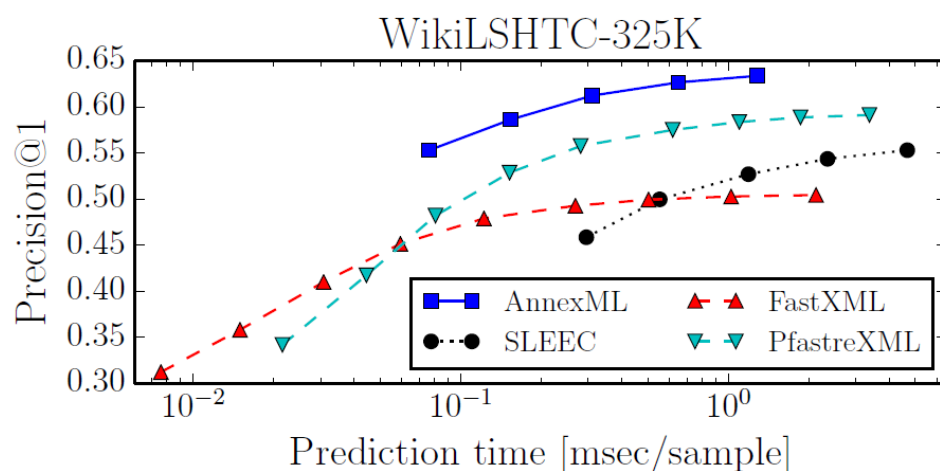
～従来技術の約58倍の速度を記録、コマー্সや広告のパーソナライズ精度向上へ～

ヤフー株式会社（以下、Yahoo! JAPAN）は、数十万の選択肢の中から適切な解を、世界最速かつ高精度に予測するAI（機械学習）技術「AnnexML（アネックスエムエル）」を開発し、11月15日（水）よりオープンソースソフトウェア（OSS）として公開しました。

「AnnexML」（以下、本技術）は、“ユーザーのサービス利用情報”などのビッグデータと、“そのユーザーがクリックした広告”などの数十万の選択肢との組み合わせを効率的に学習でき、“サービス利用情報のあるユーザーに対して、どの広告がクリックされやすいか”といった有力な候補の組み合わせを世界最速かつ高精度に予測するAI（機械学習）技術です。

Yahoo! JAPANは、一人ひとりのユーザーに合わせて数十万の選択肢の中から適切な解を掲出する、コマー্সや広告のパーソナライズ機能において、本技術が精度向上に有効と考え、今後サービスへの応用を検討しています。

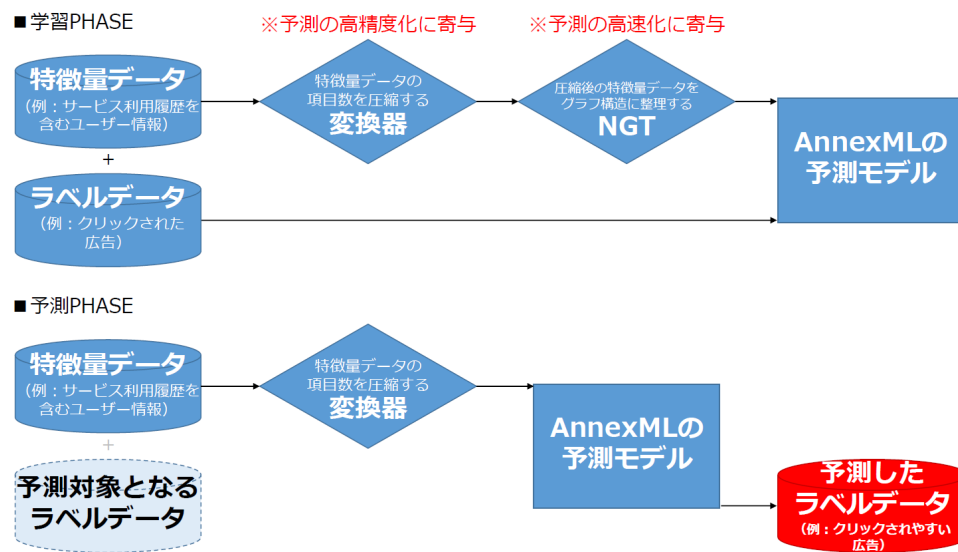
本技術を同種のビッグデータ分類技術と比較検証したところ、既存技術の中でさまざまな研究に引用されている「SLEEC」（Microsoft Research, India、Indian Institute of Technologyの研究者らによる技術）よりも、最大で約58倍の予測速度を記録し、論文投稿時点（2017年2月）における世界最速の技術となりました。ゼロコンマ数秒でユーザーに合わせた情報を選び出し配信するウェブサービスにおいては、予測速度の高速化が、より高精度な予測アルゴリズムの導入につながります。



【図1】他の既存技術との高速性比較のグラフ

本技術では、当社が提供するAI技術関連のOSSでもある高速検索技術「NGT」を用いて、項目数を数十に圧縮したビッグデータを「グラフ構造」に整理・学習することにより、高い予測精度の条件下においても高速化を実現しています。

詳細については、2017年8月にカナダで開催されたデータマイニング領域で最も権威のある国際会議（トップカンファレンス）「KDD2017」にて、論文として発表しています。



【図2】 AnnexMLの処理フロー

なお本技術は、導入後の特許侵害の発生リスクを抑え、安心して利用いただくため、特許権を取得しています。研究用途だけでなく商業用途も含め、その特許権のライセンスを無償提供する形で、11月15日（水）よりGitHub上に公開しました。OSS公開を通じて、本技術のさらなる利便性向上を図り、データサイエンス領域の研究者・エンジニアコミュニティへ貢献したい考えです。

Yahoo! JAPANは、メディア・コマー্স・決済などにおいて国内トップクラスのユーザーを抱えるサービスを多数提供しており、その裏側でさまざまな種類のビッグデータを蓄積しています。このような「マルチビッグデータ」を保持している企業は、世界的にも稀有な存在であり、近年、ますますデータの重要性が高まる中、国内外のデータサイエンティストや企業より、注目を集めています。

さらに「マルチビッグデータ」を保持するだけでなく、その利活用のために「技術で世界TOP10」を掲げ、データサイエンス領域における先端研究・開発も推し進めています。特に、大学や研究機関との共同研究やAI技術関連のOSS公開など、アカデミア領域への情報発信を強化しており、本技術のOSS公開もその取り組みの一環です。

Yahoo! JAPANは今後も、強みである「マルチビッグデータ」をいかした先端研究・開発の推進を通じて、一人ひとりのユーザーが望んでいる、さらには望んでいることにまだ気づいていない情報を適切な内容で提案し、最上級のおもてなしを実現するサービス提供を目指してまいります。

■論文の参照先

AnnexML: Approximate Nearest Neighbor Search for Extreme Multi-label

Classification

https://research-lab.yahoo.co.jp/ir/20170813_tagami.html

■GitHub 内の「AnnexML」公開ページのアドレス

<https://github.com/yahoojapan/AnnexML>

■高速検索技術「NGT」のプレスリリース

<https://about.yahoo.co.jp/pr/release/2016/11/24a/>