

## Yahoo!JAPAN におけるビッグデータ利活用のご紹介

ヤフー株式会社 Yahoo! JAPAN 研究所  
主任研究員 坪内 孝太 氏



### ■ コンテキストウェアネスの研究について

私は Yahoo! JAPAN 研究所で現在、「コンテキストウェアネス」領域を牽引する取組として、コンピュータ、Web サービスが理解する人間像を私たちがリアルの世界で理解する人間像に近づけようとする研究を行っている。思考や嗜好が表れやすいウェブ世界のログ（閲覧履歴、購買履歴等）と、見た目、環境・行動が表れやすいリアル世界のログ（位置情報、活動情報等）を併せて統合的に解析させることで、コンピュータが個人をより詳細に認識できるようになると考える。

例えば、人間関係の緊密さの分析では、これまで主に SNS や位置情報、無線通信機器の共起情報が用いられてきたが、Yahoo! JAPAN 研究所では、行動を共にする者同士の間関係は緊密だろうとの仮説を立て歩数の波形の類似性など万歩計で収集した活動量情報からの分析を試みて、精度の高い「人間関係グラフ」を作成することができた。

### ■ Yahoo!のサービスおよび Yahoo! JAPAN 研究所におけるデータ利活用

ヤフー株式会社は、社会の課題を解決し人々の役に立つ「課題解決エンジン」という役割を担うことを掲げ、日本最大級のポータルサイト Yahoo! JAPAN の運営をはじめ、様々な Web やアプリのサービスを提供している。Yahoo! JAPAN 研究所では、それらのサービスを先端技術により支えるため、約 15 名の研究員が各自テーマを持って研究を行っている。

サービスの利用者は非常に多く、年齢・職業も幅広いため、ヤフー株式会社は Web における行動情報を日本で最も多く保有していると考えられる。そのような多様なデータを利活用し、広告のクリック率向上サービスや、世の中のリアルな動きを分析・予測するビッグデータレポートの公表などを行っている。

### ■ 情報利活用事例 1：広告の CTR（クリックスルーレイト）向上

Yahoo! JAPAN では、検索クエリに入れた用語に準じて広告を表示するスポンサードサーチをはじめ様々な広告を提供しているが、データを利活用しオンライン広告を最適化しようとする事例としては「興味連動型広告」のサービスを提供している。このサービスでは、機械学習等の技術を使い、ページおよびユーザーの検索・ページ閲覧行動（検索した用語、閲覧記事といった情報）に表れる興味・嗜好に適した広告を検索エンジンから取得し、予測クリック率（CTR）に応じてランキングロジックで順番を決定し配信している。

このようにユーザーに適切な広告を適切な文脈で効果的に表示することにより、ユーザーにとっては興味のある広告が表示される、広告を掲載する Web ページの運営者にとっては広告

の掲載収入が多く得られる、広告主にとっては広告を興味のある人に効率的に届けられるという「Win-Win-Win」の姿を目指しており、実際に広告のクリック率向上や売上高増加という成果につながっている。

## ■情報利活用事例2：ビッグデータレポートの公開

Yahoo! JAPAN 研究所では、多くの人が関心を持つ、リアルな世の中の様々な事象について、Yahoo! JAPAN のサービスを通じて匿名化した上で蓄積した Web 上の行動情報（検索・広告・ショッピング・地域情報・ソーシャル上のトレンド情報などあらゆるカテゴリーの膨大なデータ）を用いて分析・予測した「ビッグデータレポート」を定期的に作成・公開している。これまでに以下のような4種類のレポートを公表してきた。

### （1）インフルエンザの流行

ヤフー検索エンジンでのある治療薬の検索数と、厚生労働省が発表したインフルエンザの患者数や「インフルエンザ」という用語の検索数との相関性が分かったため、治療薬名の検索数と検索が行われた地域の位置情報から、その時期にインフルエンザの流行している地域や終息時期、日本全体に蔓延しているかどうか、などを分析した。

### （2）東日本大震災の振り返り

Yahoo! JAPAN サービスのログを使い、2011年の東日本大震災の際の日本全域や被災地の状況について様々な観点から分析している。例えば都道府県別のヤフー検索エンジンでの検索とPV（ページビュー）のログの分析から、3月11日直後の数日間は、被災地からのパソコンからの検索数は大きく減少したが、スマートフォンからの検索数はそれほど減少しておらず、情報入手の手段としてスマートフォンがよく使われたことが分かった。

### （3）選挙結果の予測

選挙前の政党名などの用語のヤフー検索エンジンでの検索数という Web 上の行動と実際の獲得議席数というリアルな行動との相関性が2012年の衆議院選挙で分かったため、国政選挙や首長選挙の結果予測を行った。2013年の参議院選挙では42都道府県で当選政党が予測と一致するなど、専門家でも困難な選挙の投票行動の予測に成功した。

### （4）景気の予測

月末に明らかになるヤフーでの検索語ごとの月間検索数データを用いて、調査時点の2か月程度後に内閣府が発表する景気動向指数よりも早く景気を予測できないか研究した。2012年以降に毎日一定数以上検索されている約60万語の月間検索数を対象に重回帰分析を使って景気動向一致指数モデルを構築し、予測を行った。

ビッグデータレポートでは、本セミナーで紹介した内容以外にも様々な内容の分析をまとめているので、是非ご活用いただきたい。

このように、ヤフー株式会社では多様な情報を利活用しているが、その際に、どのように利活用と利用者のプライバシーへの配慮を両立させるかも重要な課題である。セキュリティ、プライバシーの面で適切な対策を講じた上で、適切なデータ利活用で快適な世の中にしていこうと取り組んでいる。

参考 URL : Yahoo! JAPAN ビッグデータレポート <http://docs.yahoo.co.jp/info/bigdata/>