

平成 29 年（2017 年）4 月 6 日

## web コンテンツのアクセス数を予測する技術を開発

行動リズムや外部ソーシャルメディアの効果など取り入れ、精度を向上

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所（NII、所長:喜連川 優、東京都千代田区）情報学プリンシプル研究系 助教・小林 亮太（こばやし・りょうた）は、web コンテンツへのアクセス数を高い精度で予測する手法を国際共同研究で開発しました。この研究成果は 4 月 6 日、オーストラリア・パースで開催されている web に関する国際的なトップ会議「第 26 回 International World Wide Web Conference (WWW2017)」<sup>(\*1)</sup> で発表されました。

web 上では膨大な量のコンテンツが生まれています。1 分ごとに、YouTube には合計で 300 時間以上にのぼる動画が投稿され、Twitter では 30 万以上のツイートが行われています。しかし、この中で人々に注目される web コンテンツはほんの一部です。web コンテンツ（web サイト、Tweet など）が公開後にどれだけ人々の興味を引くかを予測できるようになれば、マーケティングや魅力的なコンテンツ開発を効率的に行うことが可能になります。このため、web コンテンツが公開された後にどれだけ注目を集めるかを予測することは、データマイニング分野における重要な課題の一つです。

本共同研究で開発したのは、特定の期間（今後 12 時間、48 時間、96 時間など）に web コンテンツがどれだけアクセスされるかを予測する技術です。この技術によって、人間の 1 日の行動リズムや外部要因を取り入れたアクセス数の将来予測が初めて可能になり web コンテンツへのアクセスを定量的（数量的）、かつ、高精度に予測できる可能性を示しました。

### 【今回開発した技術の特徴】

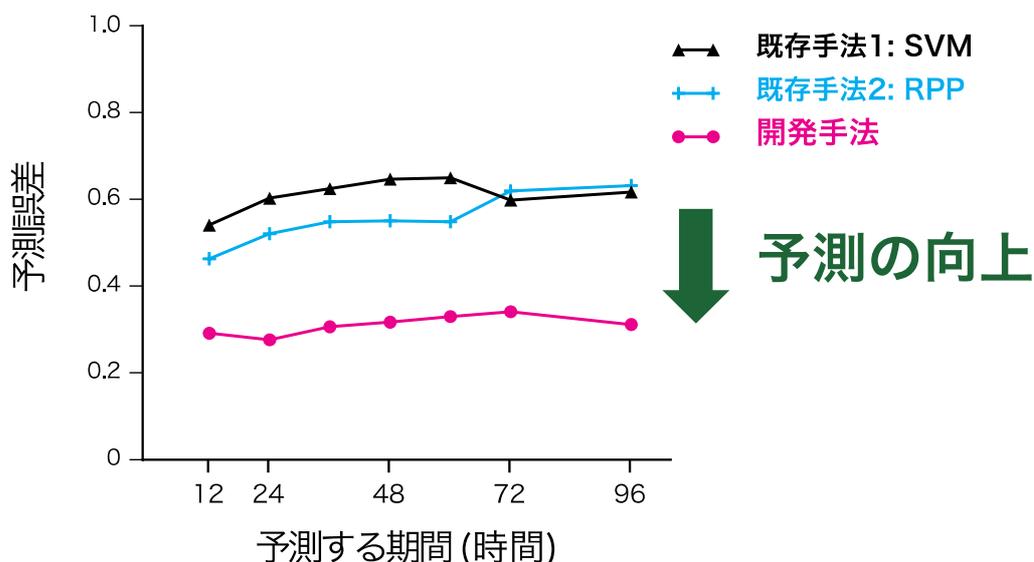
web コンテンツへの将来アクセスを予測する標準的な方法として、サポートベクターマシン（SVM）などの機械学習を適用することが考えられます。この方法は、人気が出そうかどうかは予測できますが、アクセス数そのものを予測することはできませんでした。また、アクセス数を予測するための強化ポアソン過程モデル（RPP）も提案されていますが、人間の行動リズムや外部要因が考慮されていないために高い予測精度を実現することは困難でした。

本共同研究で開発したのは、web コンテンツのアクセス数そのものを、特定の期間（今後 12 時間、48 時間、96 時間など）について予測する技術です。小林助教がナミュール大学（ベルギー）の Renaud Lambiotte 教授と開発した数理モデル<sup>(\*3)</sup>を拡張することで、睡眠や仕事などの 1 日の行動リズム、外

部のソーシャルメディアによる宣伝効果といった外部要因を取り入れることが可能な時系列モデルを新たに提案しました。そして、適切なモデルパラメータを学習して予測を行う技術を開発しました。

この手法の有効性を調べるため、本共同研究では、インターネットの代表的な署名サイトである「Petition.com」(<http://www.thepetitionsite.com>)の署名の履歴データとトピックについての関連ツイートを収集しました。そして、特定のトピックについて、これまでの履歴データと関連ツイートデータから今後の署名数を予測するという評価実験を行いました。その結果、既存手法では全体の変動の50%程度しか予測できない一方、本提案手法では70%程度が予測可能になることを確認しました(図参照)。

今回の共同研究によって、人間の行動リズム、ソーシャルメディアによる宣伝効果といった外的要因を取り込んで分析する手法が初めて実現され、これによって web コンテンツへのアクセスをより高い精度で予測することの可能性が示されました。



〈図〉「Petition.com」の署名数を予測した結果。  
 既存手法 (SVM、RPP) に比べて予測精度が大幅に向上した

本研究は「JST ACT-I 情報と未来」<sup>(\*3)</sup>の研究の一環として JST、ACT-I の支援を受け、ローザンヌ工科大学 (スイス) の Julia Proskurnia 氏と Karl Aberer 教授、マックスプランク研究所 (ドイツ) の Przemyslaw A. Grabowicz 博士、Eurecat (スペイン) の Carlos Castillo 博士、フリブール大学 (スイス) の Philippe Cudre-Mauroux 教授と共同で取り組んだものです。

以上

メディアの皆様からのお問い合わせ先

情報・システム研究機構 国立情報学研究所

総務部企画課 広報チーム

TEL:03-4212-2164 E-mail : [media@nii.ac.jp](mailto:media@nii.ac.jp)

- 
- (\*1) 「WWW2017」： 研究者や産業界、標準化団体などが最新の動向や web の未来について議論するトップ会議。会期は4月3～7日。<http://www2017.com.au/>。
- (\*2) 「Renaud Lambiotte 教授と開発した数理モデル」： 「TiDeH: Time-dependent Hawkes process for predicting retweet dynamics」(Kobayashi and Lambiotte, "Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016" P.191—200、2016年、<http://www.scopus.com/inward/record.url?scp=84979609049&partnerID=8YFLogxK>)
- (\*3) 「JST ACT-I 情報と未来」： ACT-I (Advanced Information and Communication Technology for Innovation) は国立研究開発法人 科学技術振興機構 (JST) の戦略的創造研究推進事業の一つ。小林助教は平成 28 年度 (2016 年度) に研究領域「情報と未来」の研究者 (課題名 : 時系列データの自動解析技術の実現、 Grant 番号 : JPMJPR16UC) として採択された。