

このつぶやきを書いたのはどんなひと？

～マイクロブログの特徴を活用したユーザ属性推定技術～

どんな研究

マイクロブログには個人の意見が多数存在し、マーケティング等での活用が望まれています。しかし、既存のアンケート調査と異なり、それらを書いたユーザの属性(年齢、性別等)が明らかでないという問題があります。本研究では、それらを自動的に推定する技術に取り組んでいます。

どこが凄い

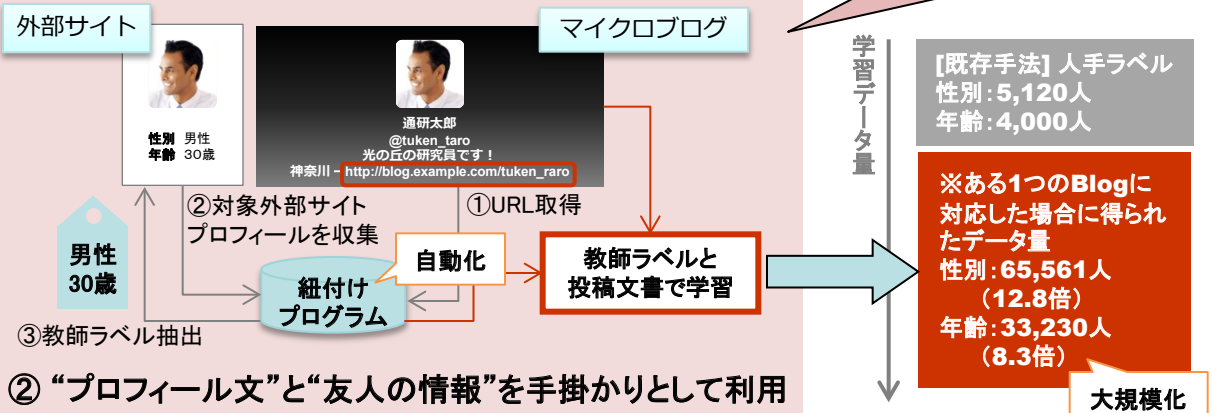
従来は人手で作成した学習データを利用し、投稿記事中の単語を元にユーザの属性を推定する手法が提案されてきました。本研究では、単語表記の背後に潜む話題や、マイクロブログの構造に注目することで、性別を約90%の精度で推定する等、主要なユーザ属性を高い精度で推定可能としました。

めざす未来

現在の推定対象はマーケティングでよく利用される年齢、性別、職業、興味等ですが、今後、より幅広い属性を、より深掘りした形で提供することを目指しています。それにより、マーケティングでの活用はもとより、個人の行動支援やコミュニケーション支援での活用も期待できます。

① 外部サイトURLを利用した学習データ収集の低コスト化

マイクロブログの構造的な特徴に注目



② “プロフィール文”と“友人の情報”を手掛かりとして利用

マイクロブログ (プロフィール文)

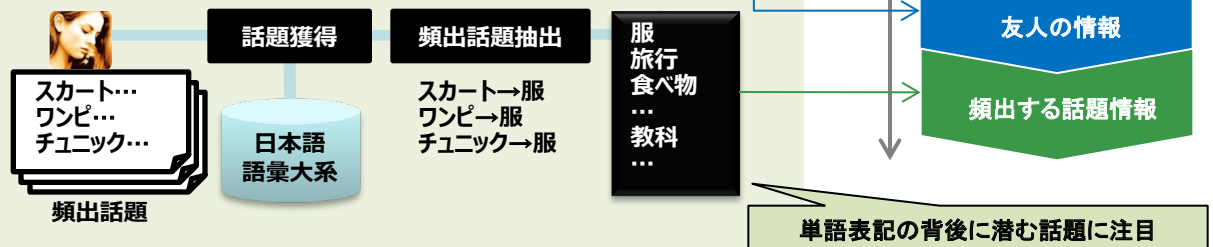


ソーシャルグラフ (友人の情報)



利用する特徴量

③ “頻出する話題”を手がかりとして利用



関連文献

[1] 伊藤淳, 西田京介, 星出高秀, 戸田浩之, 内山匡, “TwitterとBlogの共通ユーザおよび会話ユーザの同類性に着目したTwitterユーザ属性推定,” 第5回データ工学と情報マネジメントに関するフォーラム, C5-3, 2013.

連絡先

伊藤 淳 (Jun Ito)
NTT サービスエボリューション研究所
サービスハーモナイズプロジェクト
E-mail: ito.jun[at]lab.ntt.co.jp

平野 徹 (Tohru Hirano)
NTT メディアインテリジェンス研究所
音声言語メディアプロジェクト
E-mail: hirano.tohru[at]lab.ntt.co.jp

((at)の部分をも@に置き換えてください)