5N-4

マイクロブログを対象とした ユーザの習慣的な行動の解析に関する研究

田中　成典[†]　　中村　健二[‡]　　寺口　敏生[‡†]　　中本　聖也[‡†]　　加藤　諒[‡†]

関西大学総合情報学部[†]　　大阪経済大学情報社会学部[‡]　　関西大学大学院総合情報学研究科[‡†]

１．はじめに

近年，携帯端末の普及に伴い，ユーザの生活や行動に応じて様々な情報を提供するサービスに注目[1]が集まっている．そのため，携帯端末に搭載された各種センサやマイクロブログなどの CGM(Consumer Generated Media)を用いて，ユーザの生活や行動を分析，推定する研究が行われている．既存研究では，ユーザの行動推定に GPS(Global Positioning System)から取得した位置情報を用いる手法[1]や CGM の投稿内容を用いる手法[2-5]が提案されている．しかし，前者の手法では，GPS をオフにしている場合や位置情報を取得できない場合に行動を推定できない．また，後者の手法では，投稿内容に行動に関する情報が含まれていない場合や，CGM 上にユーザの投稿が存在しない場合に行動を推定できない．このように，既存研究では，推定に必要となる情報が取得できない場合に対応できない問題がある．

そこで，本研究ではユーザの日々の行動の多くが習慣的な行動(以下，習慣行動)であることに着目し，ユーザの行動推定に習慣行動を用いる方策を検討する．現代社会では，時間を基準に行動することが多く，習慣行動を抽出できれば，情報を取得できない場合でもユーザの行動を推測できると考えられる．そこで，本研究ではマイクロブログを対象に，ユーザの投稿履歴から習慣行動を解析し，その結果に基づき各時間の行動を推定する手法を提案する．これにより，投稿内容に行動に関する情報が含まれていない場合や投稿が存在しない場合においても，習慣行動に基づき行動を推測することが可能である．

Research on Analysis of Users' Habitual Behavior in Microblog
†Shigenori Tanaka
Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka 569-1095, Japan
‡Kenji Nakamura
Faculty of Information Technology and Social Science, Osaka University of Economics, 2-2-8 Osumi, Higashiyodogawa-ku, Osaka 533-8533, Japan
‡†Toshio Teraguchi, Seiya Nakamoto, Ryo Kato
Graduate School of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

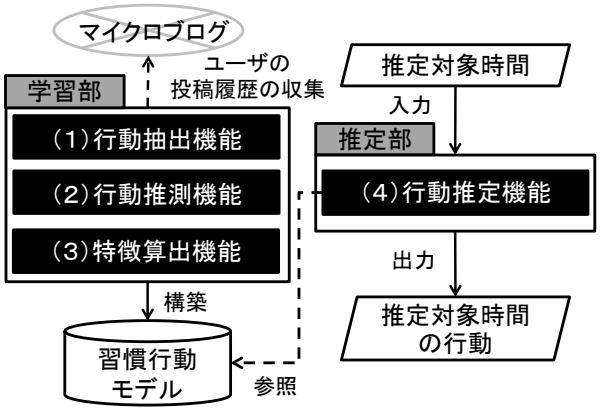

図１　本手法の流れ

２．研究の概要

本研究では，マイクロブログを対象にユーザの習慣行動に基づく行動推定手法を提案する．本手法の流れを図 1に示す．本手法は，学習部にて，マイクロブログから収集したユーザの投稿履歴から，各曜日(7 曜日)，各時間(24 時間)における行動の発生確率(以下，行動確率)を算出し，習慣行動モデルを構築する．そして，推定部にて，習慣行動モデルを用いて推定対象時間のユーザの行動を出力する．ユーザの行動は，既存研究[1]を参考に睡眠中，出勤中，勤務中，食事中，帰宅中とその他の 6 種類に分類した．

２．１　行動抽出機能

本機能では，収集したユーザの投稿履歴から投稿内容に基づき行動を抽出する．行動の抽出には，各行動に関する用語を登録した辞書を用いる．この辞書中に登録された用語と投稿に含まれる用語の一致により，ユーザの行動を抽出し，これらの投稿時間と投稿数を取得する．

２．２　行動推測機能

投稿内容に基づく行動抽出では，睡眠中や勤務中など投稿が行われない行動を抽出できない．そこで，本機能では，ユーザの投稿傾向に基づきこれらの行動を推測する．睡眠中や勤務中の時間帯は，投稿数が少なくなる傾向がある．そこで，投稿数の少ない時間帯に対して，行動抽出機能の結果を暫定的に付与することで，行動を推測する．付与する情報には，投稿数の少ない時間帯の直前の時間の投稿数を付与する．

２．３　特徴算出機能

本機能では，行動抽出機能と行動推測機能の結果から，各時間に特徴的な行動を補正する．補正には，投稿時間と投稿数から tf-idf を用いて時間ごとの行動のスコアを算出する．そして，tf-idf のスコアを各曜日，各時間で確率化することで，行動確率を算出する．算出した行動確率は，習慣行動モデルに格納する．

２．４　行動推定機能

本機能では，習慣行動モデルに基づき，推定対象時間の同曜日と同時間の行動確率を算出し，最も確率の高い行動を推定結果として出力する．

３．実証実験と考察

本手法の有用性を実証するために，マイクロブログの Twitter を解析対象として，行動推定の精度を評価する．

３．１　実証実験

実証実験では，投稿内容に基づき算出した行動確率(行動抽出機能)，投稿傾向に基づき行動を推測して算出した行動確率(行動推測機能)と各時間に特徴的な行動を補正した行動確率(特徴算出機能)の 3 つの行動確率を比較することで，本手法を評価する．実験データには，行動に関する内容が記述されている投稿の投稿時間を用いる．実験データ数としては，Twitter から 10 ユーザを選定し，各ユーザの各行動の実験データを約 50 件，合計 2,935 件を用意する．これらの実験データを推定対象時間として入力する．評価指標には，F 値を用いる．

３．２　結果と考察

実証実験における行動推定の精度を表 1 に示し，算出したユーザの行動確率を図 2 に示す．表 1 から，本手法の特徴算出機能は平均 0.711 の精度であり，習慣行動を考慮することで概ねの行動を推定できることがわかった．この結果から，本手法の有用性を確認した．また，睡眠中と勤務中については，行動抽出機能では低かった精度が行動推測機能を考慮することで大きく向上した．この結果から，本手法は投稿内容に現れない行動についても，投稿傾向に基づき適切に行動を推測できることを確認した．一方で，食事中，帰宅中とその他については，精度が 6 割程度と低い結果になった．特に，食事中については，食事中と判定される時間帯が多数抽出され，tf-idf による特徴の補正が十分に効果を発揮せず，精度があまり向上しなかった．この問題については，食事を朝食，昼食，夕食と一定時間間隔で区別し，それぞれを別々の行動として学習することで解決できると考えられる．

表 1　行動推定の精度

	行動抽出機能	行動推測機能	特徴算出機能
睡眠中(500 件)	0.568	0.791	0.824
出勤中(463 件)	0.680	0.609	0.793
勤務中(504 件)	0.306	0.611	0.734
食事中(494 件)	0.547	0.604	0.653
帰宅中(489 件)	0.494	0.517	0.630
その他(485 件)	0.505	0.538	0.630
平均(2,935 件)	0.517	0.612	0.711

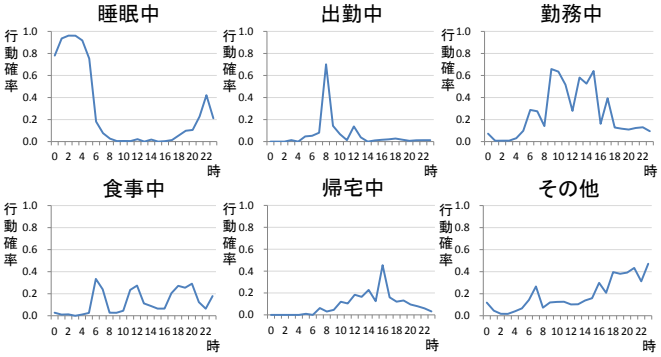

図 2　算出したユーザの行動確率

４．おわりに

本研究では，マイクロブログを対象にユーザの習慣行動に基づく行動推定手法を提案した．そして，実験により本手法の有用性を確認した．本手法を用いることで，CGM やセンサから情報が取得できない時間帯の行動を補間し，任意の時間における行動の推定技術が実現できると考えられる．今後は，実験により明らかとなった問題点に対処する手法について研究する．

参考文献

[1] 宮崎雄一郎，山田直治，住谷哲夫，磯谷佳徳：ユーザの行動に合わせたサービス実現のための行動推定技術の開発，NTT DoCoMo テクニカル・ジャーナル，NTT ドコモ，Vol.17，No.3，pp.55-61，2009.

[2] グェンミンティ，川村隆浩，中川博之，中山健，田原康之，大須賀昭彦：条件付確率場と自己教師あり学習を用いた行動属性の自動抽出と評価，人工知能学会論文誌，人工知能学会，Vol.26，No.1，pp.166-178，2011.

[3] 倉島健，藤村考，奥田英範：大規模テキストからの経験マイニング，電子情報通信学会論文誌，電子情報通信学会，Vol.J92-D，No.3，pp.301-310，2009.

[4] Cheng, Z., Caverlee, J. and Lee, K.: You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users, Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM, pp.759-768, 2010.

[5] Eisenstein, J., O'Connor, B., Smith, N. and Xing, E.: A Latent Variable Model for Geographic Lexical Variation, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, SIGDAT, pp.1277-1287, 2010.