

マイクロブログの職業属性を考慮した ユーザの行動推定に関する研究

加藤諒[†] 田中成典[‡] 中村健二^{‡†} 坂本一磨[‡]

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

1. はじめに

マイクロブログ上に投稿されたデータを用いてユーザの状況を推定し、関連した情報をユーザに対してリアルタイムに推奨するサービス[1]が注目されている。その実現には、ユーザのライフログを解析し、習慣行動を把握することが重要である。ユーザの習慣行動を推定する研究[2]-[4]では、マイクロブログの投稿内容に加えて位置情報を解析する手法[2][3]と、投稿数の変化を解析する手法[4]がある。前者は、投稿内容と移動経路を行動履歴として蓄積し、移動先を推定するが、この手法では移動履歴が必要である点で汎用性が低い。一方、後者は、投稿内容と前後の投稿数の変化とを関連付けて習慣行動を抽出し、指定した時間帯の行動を推定するが、その精度は投稿数に依存する点に課題がある。そこで、本研究では、投稿数が少ないユーザを対象に習慣行動を推定する手法を提案する。

2. 研究の概要

本研究では、投稿数が少ない時の対応として、職業ごとに異なる習慣行動[5]に着目し、既存研究[4]の推定精度を高める手法を提案する。本システムの概要を図1に示す。学習部は行動確率モデル構築機能と投稿パターンモデル構築機能、推定部は行動推定機能で構成される。

2.1 行動確率モデル構築機能

本機能では、ユーザの投稿の内容と時間（以下、投稿履歴）から行動推定に用いる行動確率モデルを構築する。職業習慣ベクトル作成処理では、職業ごとのユーザ群の投稿履歴から職業習慣ベクトル V_{job} を作成する。 V_{job} は、6次元（習慣行動） \times 24次元（時間帯）の144次元で構成される。習慣ベクトル作成処理では、習慣行動の分析のために投稿履歴を解析して推定対象ユーザの習慣ベクトル V_{user} を作成する。職業

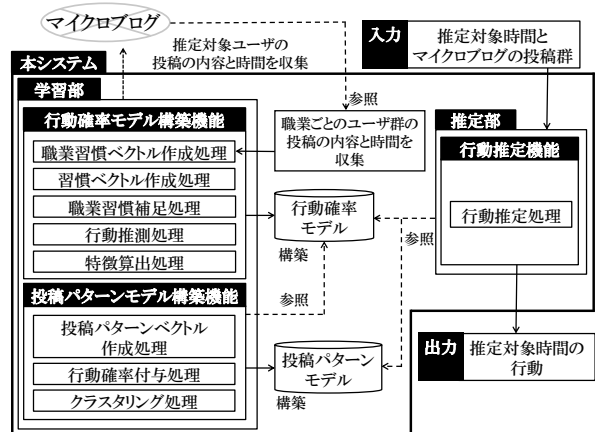


図1 本システムの概要

習慣補足処理では、 V_{job} を用いて投稿数が少ないユーザの行動情報を式(1)にて補足する。

$$V'_{user} = V_{user} \times e + V_{job} \times (1 - e) \quad (1)$$

式(1)の e は習慣ベクトルの内容を考慮する割合を示す。行動推測処理では、職業習慣補足処理で抽出した行動情報に睡眠中と勤務中を追加し、ユーザの行動情報を補足する。特徴算出処理では、投稿内容から抽出した行動と推定した行動情報を用いて各時間帯におけるユーザの行動確率を算出する。

2.2 投稿パターンモデル構築機能

本機能では、ユーザの投稿履歴から抽出した投稿数の変化と、行動確率モデルに格納した曜日ごとの各時間帯の行動確率とを関連付ける投稿パターンモデルを構築する。投稿パターンベクトル作成処理では、投稿履歴を用いて指定した時間帯のパターンベクトルを作成する。行動確率付与処理では、投稿パターンベクトルを曜日ごとの各時間帯に分類し、行動確率モデルを参照して各行動の行動確率を付与する。クラスタリング処理では、Ward法を用いて、類似する投稿パターンベクトルから特徴的なベクトルを抽出し、投稿パターンモデルを構築する。

2.3 行動推定機能

本機能では、推定対象時間とマイクロブログの投稿群を入力とし、推定対象時間の行動を出力する。行動推定処理では、行動確率モデルと投稿パターンモデルに基づき、指定した時間帯におけるユーザの行動を推定する。

Research for Reasoning Users' Activities in Consideration of Occupation Attributes in Microblogs
† Ryo Kato

Graduate School of Informatics, Kansai University, 2-1-1 Ryozenji-cho, Takatsuki City, Osaka 569-1095, Japan

‡ Shigenori Tanaka, Kazuma Sakamoto
Faculty of Informatics, Kansai University

‡† Kenji Nakamura

Faculty of Information Technology and Social Science, Osaka University of Economics

3. 実験

本システムの有用性を確認するため、マイクロブログの1つである Twitter を解析対象とし、行動推定精度を評価する。

3.1 実験内容

既存手法[4]と提案手法の行動確率の推定精度を F 値にて比較する。本実験では、習慣行動に基づく行動推定手法の有効性の評価を目的とするため、実験データとして規則的な生活リズムがあると思われる社会人ユーザを採用した。習慣行動は、既存研究[4]を参考に出勤中、勤務中、帰宅中、睡眠中、食事中とその他の6種類とした。投稿数の少ないユーザでの有効性を評価するため、ユーザの投稿の中から無作為に 1,000, 5,000, 10,000, 30,000 件の投稿を抽出して行動推定を行い、精度を評価した。ユーザの行動の正解データは、全投稿履歴を手手で解析して設定した。

提案手法では、職業習慣補足処理において、職業習慣ベクトルを用いて、解析対象ユーザの行動情報を補足する。本実験では、式(1)のパラメータ e を決定するため予備実験を実施した。予備実験の結果を図 2 に示す。結果より、1,000 件は $e=0.6$ 、5,000 件は $e=0.8$ 、10,000 件は $e=0.9$ 、30,000 件は $e=0.7$ の時に最良の結果が得られた。そのため、本実験では、それぞれの最良の値を最適値として処理を行う。

3.2 結果と考察

本実験の結果を表 1 に示す。表 1 に示す通り、ユーザの投稿件数が 1,000, 10,000, 30,000 件の場合に提案手法の方が高精度であることが分かった。特に 1,000 件時の精度は大きく向上しており、提案手法の方が、出勤中で 0.58 ポイント、勤務中で 0.23 ポイント、帰宅中で 0.11 ポイント向上した。これは職業習慣を考慮することで、社会人の規則的な生活リズムの特徴が表れたからであると考えられる。10,000 件と 30,000 件時にも平均で 0.05 ポイント精度が向上していることから、提案手法は投稿数が多いユーザに対しても有用であることが分かった。

提案手法において精度が低下した行動は、職業習慣での補足によってユーザごとの特徴が失われたためであると考えられる。この点に関しては、行動に応じて職業習慣による補足の有無の選択や投稿時間の傾向を考慮した補足手法の検討などを行う必要があると考えられる。

4. おわりに

本研究では、マイクロブログの投稿履歴に加え、職業ごとの習慣行動を考慮することで、投稿数の少ないユーザでも行動の推定が可能となる手法を提案した。評価実験の結果より、投稿

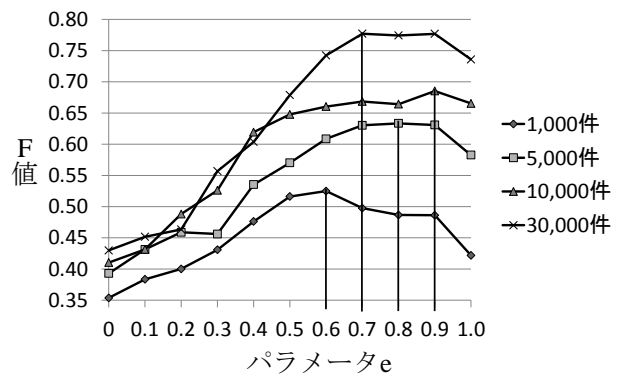


図 2 投稿データ数ごとのパラメータ e の最適値

表 1 行動推定の結果

		投稿数ごとの F 値			
		1,000 件	5,000 件	10,000 件	30,000 件
既存手法	睡眠中	0.54	0.74	0.82	0.85
	出勤中	0.07	0.69	0.71	0.85
	勤務中	0.18	0.61	0.54	0.81
	食事中	0.67	0.73	0.72	0.78
	帰宅中	0.48	0.59	0.58	0.59
	その他	0.16	0.44	0.46	0.53
	平均	0.35	0.63	0.64	0.73
提案手法	睡眠中	0.61	0.82	0.89	0.87
	出勤中	0.65	0.73	0.76	0.84
	勤務中	0.41	0.56	0.58	0.83
	食事中	0.73	0.74	0.80	0.83
	帰宅中	0.59	0.51	0.58	0.67
	その他	0.15	0.44	0.50	0.62
	平均	0.52	0.63	0.69	0.78

数が少ない場合に提案手法が有用であることを確認した。今後は、実験対象のユーザ数を増やして評価実験を行い、提案手法の汎用性を確認する予定である。

参考文献

- [1] 鳥居, 赤塚, 落合, 角野: 生活情報を提供するリアルタイム検索サービスの開発, NTT DOCOMO テクニカル・ジャーナル, Vol.20, No.4, pp.12-17, 2013.
- [2] 斉藤, 高山, 山上, 戸部, 鉄谷: マイクロブログのジオタグと発言コンテキスト解析による行動予測手法, 情報処理学会論文誌, Vol.55, No.2, pp.773-781, 2014.
- [3] 高山, 藤, 山上, 戸部, 鉄谷: マイクロブログのコンテキストを用いた行動予測への確立過程モデルの適用と評価, マルチメディア通信と分散処理研究会報告, 情報処理学会, Vol.154, No.24, pp1-6, 2013.
- [4] 田中, 中村, 寺口, 中本, 加藤: マイクロブログから抽出したユーザの習慣に基づく行動推定に関する研究, 情報処理学会論文誌: データベース, Vol.6, No.3, pp.73-89, 2013.
- [5] 田中, 中村, 加藤, 寺口: マイクロブログの投稿時間に着目したユーザの職業推定に関する研究, 情報処理学会論文誌: データベース, Vol.6, No.5, pp.71-84, 2013.