

# マイクロブログサービスの位置情報タグと 発言テキスト解析を用いた 行動推定システムの設計

山田 和貴<sup>†1</sup> 斉藤 裕樹<sup>†1</sup>

GPS機能を備えた携帯端末等による位置取得技術の普及により、人々が訪れる場所同士を結ぶ行動履歴を解析する研究が活発に行われている。経路情報を蓄積することによって得られる行動履歴は、人々の興味と移動との関係を解析する手段として注目が集まっている。一方、Twitterをはじめとするマイクロブログサービスは、テキストとともにGPSセンサによる地理位置情報を付与し、公開することが可能なサービスとして注目されている。本論文では、蓄積された人々の行動履歴を元に利用者の行動とコンテキストを解析し、将来の利用者の行動を予測する手法の提案を行い、その情報提示を行うリコメンダシステムの設計を行う。また、基礎的な評価実験による提案手法の有効性の確認を行う。

## Inferring User Behavior System by Mining Geotags and Text Messages on Micro-blog Services

KAZUKI YAMADA<sup>†1</sup> and HIROKI SAITO<sup>†1</sup>

The advance of GPS-enabled portable devices such as PDAs and smart phones facilitates people to record their location histories. Location trajectories imply human behaviors and preferences related for their interests. On the other hand, micro-blog services such as Twitter and Google Buzz enable us to publish text messages (e.g. Tweets) and location-tags (e.g. Geo-tags) to subscribers. This paper proposes a schema for inference user behavior by analyzing its location trajectories and its contexts. And we design a location recommendation system by mining multiple users' location trajectories. Also, we confirm effectiveness of the schema through basic experiments.

## 1. はじめに

GPS機能を備えた携帯端末等による位置取得技術の普及により、人々が訪れる場所同士を結ぶ行動履歴を解析する研究が活発に行われている<sup>1)-4)</sup>。経路情報を蓄積することによって得られる行動履歴解析は、人々の興味と移動との関係を抽出する手段として注目が集まっている。一方、TwitterやGoogle Buzzをはじめとするミニブログサービスは、テキストとともにGPSセンサによる地理位置情報を付与し公開することが可能なサービスとして注目されている。

本論文では、マイクロブログサービスで発信されるメッセージと位置タグから、メッセージと行動の関係を確率モデルで表現し、過去の人々の行動履歴とコンテキストから利用者の未来の行動をベイズ推定によって予測する手法を提案する。また、提案手法を用いて、位置情報サービスと連携し情報提示を行うシステムの設計について述べる。また、観光地の観光経路を対象とし、過去の行動履歴とコンテキストを用いて行動推定実験を行った結果、単純な経路の履歴から求められたものよりも有効性が高いことを示す。

本論文の構成は次のとおりである。まず第1章に引き続き、第2章でマイクロブログサービスを用いた行動推定の概要について示す。次に第3章で提案の中核となる行動推定の指標の定義、確率モデル、推定方式と、推奨経路抽出アルゴリズムについて述べる。第4章では、提案手法の有効性を確認するために行った基礎的な評価実験とその結果について述べる。さらに第5章では、関連研究と本研究の位置づけを示し、第6章では、本論文の内容をまとめ結論づける。

## 2. マイクロブログサービスを用いた行動予測

GPS機能を備えた携帯端末の普及により、多くの位置情報サービス (Location Based Service: LBS) がiPhoneやAndroidといった様々なプラットフォーム上で実現されている。初期の位置情報サービスでは、地図上でのナビゲーションを行うことや、位置を指定した最近傍の施設やサービスの検索などが提供されたが、位置情報は利用者の携帯端末内で用いられるのみで、他の利用者とは共有は行われていなかった。しかし、最近の位置情報サービスでは、人やものの位置情報を蓄積し利用者間で情報共有することで、位置情報に新たな価値

<sup>†1</sup> 東京電機大学未来科学部情報メディア学科  
Department of Information Systems and Multimedia Design, Tokyo Denki University

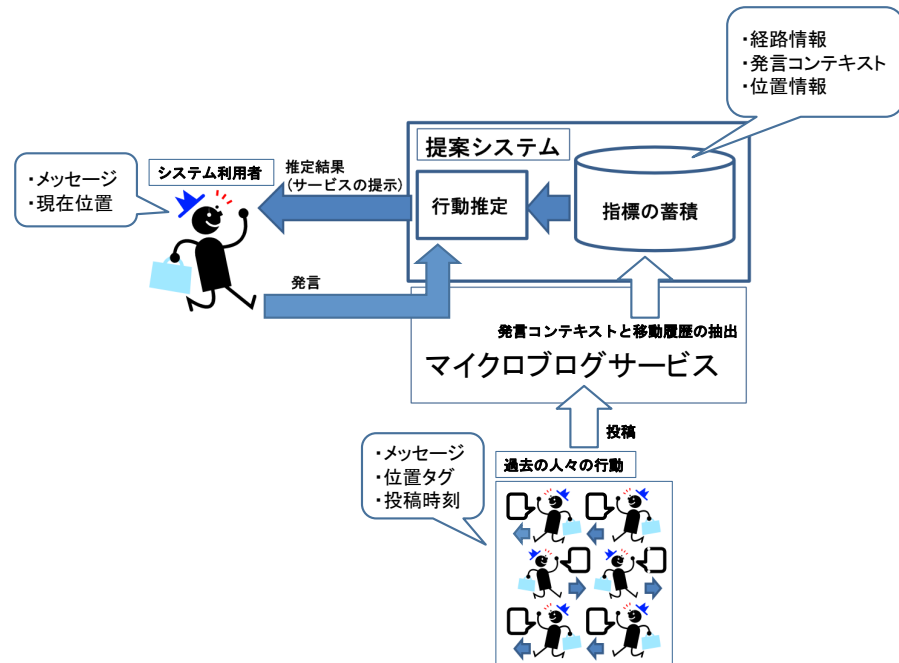


図1 マイクログのメッセージとジオタグを用いた行動予測システム

Fig.1 User Behavior Inferring System among Text Messages and Geo-tags in Micro-blog Services.

値を生み出す動きが活発に行われている。また、GPSによって取得された人々が訪れる場所同士を結んだ行動履歴を蓄積することで、街中の人々の動線を解析することやナビゲーションなどへの応用が期待されている。

一方、Twitterをはじめとするマイクロブログサービスは、短いテキストを発信するサービスであり、その場の状況に関するメッセージを即座に発信できる特徴がある。特に、スマートフォン等の携帯端末上で用いられることにより、即時性と臨場感の高い情報発信が可能である。また、メッセージを発信する際にジオタグと呼ばれる位置情報を付加し、実世界と直接リンクした情報を扱うことも可能である。このように、人がセンサとなって即座に情報発信することにより、人々の日常活動や目標指向の行動を知ることが可能になることから、新たな価値のある情報を作り出す研究が活発である<sup>5)-7)</sup>。

本研究では、まず、人々がマイクロブログサービスに発信したメッセージとジオタグ情報から、行動とそのコンテキスト情報を抽出し統計情報として蓄積することで、利用者のコンテキストから未来の行動を推定する手法を検討する。具体的には、人々がマイクロブログ上で発信するメッセージと行動には一定の関係があるとの仮定に基づき、行動とメッセージの組み合わせを条件付確率による定式化を行い、蓄積された情報に基づいたベイズ推定により利用者の未来の行動予測を行うものである。

次に、推定の結果から利用者に次の行動に対する適切な情報提示を行うリコメンダシステムの提案を行う。図1に提案システムのシステムモデルを示す。システムでは、過去の人々のメッセージ、ジオタグおよび行動の一連のデータを蓄積し、利用者の行動とこれまで発信したメッセージから上述の予測手法に基づき未来の行動の推定を行う。システムは、位置情報サービスを用いて、推定された移動先に関する利便性の高い情報を取得し利用者に提示する。

本手法は、街や都市内での一定の目的を持った一連の行動経路を対象とする。例えば、一つの観光地での一日の観光ルートなどが想定される。例えば、史跡を巡る観光、というコンテキストを有する移動であれば、歴史的建造物や遺跡などを通ることが考えられる。また、絵画を観るというコンテキストであれば、美術館や博物館を通る行動が考えられる。このようなコンテキストによって異なる行動を過去のマイクログ上の発言と行動履歴から分析することで、同様のコンテキストを有しこれから移動を開始する利用者に対して、どのような経路でどこに訪問するのが適切かを推定し、目的地へのナビゲーションや目的地情報を案内する等の適切な情報提示を行うことが可能である。

### 3. 発言コンテキストと移動履歴による行動推定手法

本章では、まず行動推定の基本的方針を述べ、行動推定に必要な指標の定義を行った後、これらの指標を用いベイズ推定による行動推定手法について述べる。

#### 3.1 行動推定の基本方針

マイクロブログは、その場その瞬間に見たことや感じたことを気軽に公開し共有するなど、まさに利用者の関心を即時的に反映するメディアであると言える。マイクロブログでの発言とジオタグによって、利用者その場所でどんなことを感じたかを知ることができる。利用者のジオタグを時間軸上に並べ、その発言を集約することで、利用者の行動と背景にある目的を知ることができる。例えば、京都を観光で訪れた人々の場合、南禅寺や三十三間堂や清水寺などを通る位置履歴とともに、南禅寺で湯豆腐を食べたことや、三十三間堂の仏像を見たことや、清水寺の舞台からの景色に関する発言が送信される可能性が高い、といったことがあげられる。この情報を多くの人々から集約することにより、移動経路とその経路の意味を知ることが可能であると考えられる。

このような状況を踏まえ、本論文では、蓄積された移動と発言を条件付確率モデルで表現し、ベイズ推定によって、ある利用者の発言に対して未来の利用者の行動を推定する手法を提案する。ある移動経路を通った人々のうち特定の発言をした人の割合を、条件付確率で計算すると、移動経路に対して関係の強いキーワードが高い割合を占めることとなる。このことを元に、発言から未来の移動経路の確率を求める。例えば、過去に五重塔から二寧坂を経て清水寺に行くルートを通った人は、高い割合で歴史的な建物や風情や景色に関する発言をしたと考えられる。このような割合を元に、現在五重塔で風情や景色を楽しんでいる発言をする人に対して、推定される目的地である清水寺の観光情報を提示することや目的地へのナビゲーションを行う。

以下、行動推定の手法を説明する。

#### 3.2 前提となる指標の定義

本節では、行動推定に用いる指標の定義を行う。提案手法では、図2のようにマイクロブログ上のメッセージとジオタグおよび、2地点間の移動を指標として用いる。

##### (1) 位置 ( $L$ )

利用者はGPS機能を備えた携帯端末を持ち、マイクロブログサービスへのメッセージ送信とともに「現在地の座標-タイムスタンプ」のペア ( $location_k, time_k$ ) を記録する。位置  $L$  はシステム上で以下のように集約される。

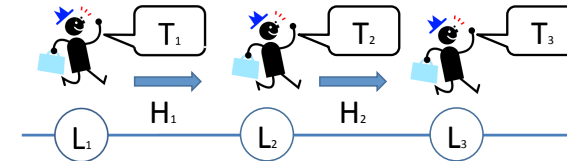


図2 行動推定に用いる各指標  
Fig.2 Indices for Inferring User Behavior.

$$L = \{(location_k, time_k) \mid k \in U\} \quad (1)$$

##### (2) 発言コンテキスト ( $T$ )

利用者の発信したメッセージは、単語ごとに分解し重み付けし、ベクトル空間上に表す。この情報とタイムスタンプのペア ( $message_k, time_k$ ) を記録する。発言コンテキスト  $T$  は以下のように集約される。

$$T = \{(message_k, time_k) \mid k \in U\} \quad (2)$$

##### (3) 移動 ( $H$ )

利用者のある位置から別の位置への移動は、移動元の位置  $L_k$  と移動先の位置  $L_{k+1}$  のペアによって表す。

以上の情報の蓄積により、移動の起こる割合  $P(H_k)$  と発言コンテキストが現れる割合  $P(T_k)$  を得ることができる。次に、行動と発言の関係について、移動が発言の根拠であると仮定し条件付確率で表すと、移動を行ったうち、ある発言コンテキストを有する割合は  $P(T_k|H_k)$  となる。

#### 3.3 ベイズ推定による行動推定

提案する行動推定手法は、前節の条件付確率モデルで表現された事前確率から、ベイズ推定により未来の行動予想の指標である事後確率を求めるものである。ここで、過去蓄積された情報として、移動の起こった割合  $P(H_k)$ 、発言コンテキストが現れた割合  $P(T_k)$ 、移動を前提とした発言の割合  $P(T_k|H_k)$  が与えられたとき、利用者の発言  $T_l$  がなされた後に移動  $H_l$  起こる確率  $P(H_l|T_l)$  を求める式を以下に示す。

$$P(H_l|T_l) = \frac{P(H_k)P(T_k|H_k)}{P(T_k)} \quad (3)$$

上式の確率を空間上のすべての2地点の移動  $H_{i \rightarrow j} (i, j \in \mathbf{W})$  について求め、2地点間の移動確率を表す状態遷移行列  $\mathbf{M}_{i,j}$  を得る。

$$\mathbf{M} = \begin{pmatrix} 0 & \cdots & P(H_{L1 \rightarrow L_i}|T_k) & \cdots & P(H_{L1 \rightarrow L_n}|T_k) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ P(H_{L_i \rightarrow L_1}|T_k) & & 0 & & P(H_{L_i \rightarrow L_n}|T_k) \\ \vdots & & \vdots & \ddots & \vdots \\ P(H_{L_n \rightarrow L_1}|T_k) & \cdots & P(H_{L_n \rightarrow L_i}|T_k) & \cdots & 0 \end{pmatrix} \quad (4)$$

そして、状態遷移行列  $\mathbf{M}$  より、位置情報の列で表現される利用者の任意の行動の確率は次式で求めることができる。

$$P(L_0, L_1, L_2, \dots, L_n|T_k) = \mathbf{M}_{L_0, L_1} \cdot \mathbf{M}_{L_1, L_2} \cdots \mathbf{M}_{L_{n-1}, L_n} \quad (5)$$

### 3.4 推奨経路抽出アルゴリズム

前節の予測手法で得られた状態遷移行列から、利用者に対して推奨される経路を求めるアルゴリズムを図3に示す。このアルゴリズムでは、状態遷移行列と利用者の現在地から次に進む可能性の高い経路を求めるものである。アルゴリズムでは、まず、利用者の現在地  $L$  を始点として最も確率の高い移動先を、状態遷移行列  $\mathbf{M}$  から選択する。次に、推定された移動先からさらに次の移動先を同行列から選択する。このことを指定された回数繰り返す。なお、同じ経路が繰り返し選択され閉路となることを防ぐため、推定経路の重複を防ぐようにしている。以上のアルゴリズムにより、利用者がこれから移動する経路を求め、移動先に関する適切な情報提示を行うことが可能である。

## 4. 行動推定手法の評価

提案手法の有効性を確認するため基礎的な評価実験を行った。

### 4.1 実験データセット

代表的なマイクロブログサービスである Twitter における、現時点でのジオタグを付加した投稿は全体の約 0.6% と非常に少ないため、今回は過去の人々の発言コンテキストに旅行記の文章を利用した。具体的には、比較的短いテキストで簡潔に旅行時の状況が記載されている旅行記サイト「フォートラベル<sup>8)</sup>」より記事を収集した。旅行記の文章から名詞、動詞のみを抽出し発言コンテキストとして利用する。また位置情報は最低 2 か所、最高 8 か

---

### Algorithm: Inference Route Recommendation

---

**Input:**  $M$ : state transition matrix,  $L$ : location,  $n$ : number of hops

**Output:**  $H = \{H_1, H_2, \dots, H_n\}$ : route sequence

```

1  current ← L
2  for i=0 to n do
3    for each loc ∈ Mcurrent do
4      if <current, loc> exists in H then continue
5      next = max index of Mcurrent.loc
6    done
7    if next is empty then return H
8    add <current, next> to H
9    current ← next
10 done
11 return
```

---

図3 経路抽出アルゴリズム

Fig. 3 Inferring Route Recommendation.

所の観光地を移動した順に記録することとした。

#### 4.2 評価方法

提案手法である発言コンテキストを用いてベイズ推定した結果と、発言を考慮せずに統計的に割合の多い経路を求めた結果を比較することで、発言コンテキスト解析の有無による違いを評価する。そこで、ある位置を基準とした次の目的地へ至る確率を、提案手法と単純統計に基づく手法とで比較を行った。

#### 4.3 結果

実験では、高知駅を起点とし利用者が「竜馬」を含む発言をした状況で、この利用者の移動先を推定することを行った。

まず、「高知駅」を起点とし、坂本龍馬に関する観光地が点在する「桂浜」に移動する確率を考える。高知市内周辺の旅行記 50 件中、高知駅を通過したケースは 30 件、高知駅から桂浜までの移動を行っていたケースは 13 件、さらに、そのうち旅行記に「竜馬」を含んでいたのは 9 件であった。

よって、 $H = \langle \text{高知駅} \rightarrow \text{桂浜} \rangle, T = \text{竜馬}$  としたとき、 $P(H) = \frac{13}{50}, P(T|H) = \frac{9}{13}$  となる。また、高知駅を起点とした移動先は桂浜を含めて 5 通りあり、これらの移動  $H_k (1 \leq k \leq 5)$  についての割合はそれぞれ、 $P(H_k) = \{\frac{13}{50}, \frac{10}{50}, \frac{4}{50}, \frac{2}{50}, \frac{1}{50}\}$ 、 $P(T|H_k) = \{\frac{9}{13}, \frac{3}{10}, \frac{1}{4}, \frac{1}{2}, \frac{0}{1}\}$  であった。次に、次式によって桂浜まで移動を行う確率を求める。

$$P(H|T) = \frac{P(H)P(T|H)}{P(T)} = \frac{P(H)P(T|H)}{\sum_k P(T|H_k)P(H_k)} \quad (6)$$

この結果、提案手法では  $P(H|T) = \frac{9}{14} \approx 0.6428$  となり、単純統計では  $\frac{13}{30} \approx 0.4333$  となる。よって、「竜馬」という発言コンテキストを用いて推定した場合の方が、桂浜を移動先とする確率が高いという結果となった。

次に、旅行記から得られた元データの高知駅を起点とした移動先「桂浜」、「高知城」、「はりまや橋」、「ひろめ市場」、「足摺岬」について、提案手法と単純統計の比較を行った。結果を図 4 に示す。単純統計と比較し、提案手法では坂本龍馬に関する観光地が点在する桂浜への移動確率が他の観光地と比べて高いことがわかる。

#### 4.4 考察

実験では、発言コンテキストによる推定を行うことで推定精度が高まることを確認した。今回の実験では、1 つの単語のみを用いたこととデータ数が少なかったため、足摺岬の移動確率が 0% となるなど極端な結果となった。実際には、複数の単語を考慮する必要があると考えられる。しかし単純に単語数を多くすると、移動経路に特有の単語以外が含まれること

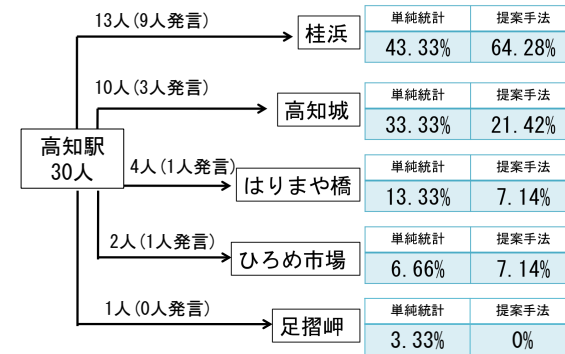


図 4 高知駅からの移動確率

Fig. 4 Probability of movement from Kochi station.

となり推定精度が悪化する。この実験では、「竜馬」という移動目的と関連性の高いと考えられる単語を選択したが、実際の環境では、有用な単語を選別する必要があると考えられる。

## 5. 関連研究

GPS などの位置計測技術の発展により、位置とタイムスタンプからなる位置履歴を記録することが可能である。この位置履歴を、ユーザの興味や趣向に関する手がかりと見なして分析する研究が行われている。Counts ら<sup>1)</sup> は、センサによってキャプチャされた情報に位置タグをつけ複数ユーザで共有するシステムを提案している。平井ら<sup>2)</sup> は、GPS ナビゲーションツールを利用して目的地へたどり着く過程を分析し、実世界との知覚的相互関係を分析している。また、Geolife<sup>3)</sup> では、位置履歴からユーザの行動を分析し、複数のユーザ間でコミュニティを形成できるようなソーシャルメディアを提案している。本研究では、位置の履歴だけを分析するのではなく、マイクロブログ上で人々が送信した発言とジオタグを蓄積し、位置履歴だけでなく同時に発言から得られるコンテキストを用いる。これにより、行動を起こす背景や目的を含めた分析を行うことが可能である。

また、ブログや SNS, Twitter などは、その瞬間に見たことや感じたことを気軽に Web 上に公開し共有するなど、世の中の関心を即時的に反映する新たなメディアであることから、これらから有用な実世界情報の抽出を目的とした研究が行われている。Kim ら<sup>5)</sup> は、ある出来事に対するソーシャルメディアでの複数の発言状況から注目する出来事の発生を推定するための、引用やテキストの類似度に基づく情報伝搬経路解析について提案している。また、田中ら<sup>6)</sup> は、Twitter の複数発言間の構造を HITS アルゴリズムにより解析し、authority と hub の 2 つの尺度を用いて情報取得に有用なユーザを抽出する手法を提案している。風間ら<sup>7)</sup> は、Twitter のツイートを収集し返信、リツイート、ハイパーリンクを手がかりに情報伝搬ネットワークを抽出する手法を提案している。Twitter は発言の文字数が短く、目的のトピックを示す単語が省略されていることや、リツイート元の内容を前提とした発言がされていることが多いため、発言間の情報伝搬を展開してコンテキストを解析することは今後の課題と言える。

## 6. おわりに

本論文では、ミニブログサービスから利用者のコンテキストを取得し、地理位置情報と組み合わせることで利用者の行動を予測する情報提示システムを提案した。提案手法は、人々がマイクロブログサービスに発信したメッセージとジオタグから、行動とそのコンテキスト情報を抽出し統計情報として蓄積することで、利用者のコンテキストから未来の行動の推定を行うものである。また、提案手法の基礎的評価を行った。

今後は、システムの実装を行い、実際のマイクロブログの発言と位置タグを大規模に収集し、行動推定の評価実験を行う予定である。

## 参 考 文 献

- 1) Counts, S. and Smith, M.: Where were we: communities for sharing space-time trails, *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems (GIS 2007)*, pp.1-8 (2007).
- 2) 平井浩将, 森 傑: 経路探索における GPS ナビゲーションツールの利用とアクション生起との関係-都市空間におけるアクトファインディングに関する研究-, 日本都市計画学会都市計画論文集, No.42-3, pp.541-546 (2007).
- 3) Zheng, Y., Xie, X. and Ma, W.: GeoLife: A Collaborative Social Networking Service among User, location and trajectory, *IEEE Data Engineering Bulletin*, Vol.33(2), pp.32-39 (2010).
- 4) Hung, C.-C., Chang, C.-W. and Peng, W.-C.: Mining trajectory profiles for dis-

covering user communities, *Proceedings of the 2009 ACM International Workshop on Location Based Social Networks (LBSN 2009)*, pp.1-8 (2009).

- 5) Kim, J.W., Candan, K.S. and Tatemura, J.: Efficient overlap and content reuse detection in blogs and online news articles, *Proceedings of the 18th ACM International Conference on World Wide Web (WWW 2009)*, pp.81-90 (2009).
- 6) 田中淳史, 田島敬史: twitter のツイートに関する分類手法の提案, 日本データベース学会他 第 2 回データ工学と情報マネジメントに関するフォーラム (DEIM2010) A5-4 (2007).
- 7) 風間一洋, 今田美幸, 柏木啓一郎: Twitter の情報伝搬ネットワークの分析, 第 24 回人工知能学会全国大会 1F2-OS8-4 (2010).
- 8) フォートラベル: <http://4travel.jp/>.