

行動要因情報源抽出のための 潜在的属性および位置情報に基づく行動予測精度の評価

河村 光則¹ 齊藤 裕樹¹

概要: GPS や無線強度などを用いることで屋内外において測位が可能となり、位置情報を利用した様々なサービスの実用化が進んでいる。また、ユーザのデモグラフィック属性や過去の行動の分析を元に、システムがユーザに有用な情報をレコメンドするサービスも普及し、ユーザが未来の行動を決める一つの要因となっている。しかし、システムが限定された属性に従って、提供する情報を選別しているため、本来なら、受け取ったユーザの行動に影響を与え得る情報の一部が失われるという問題がある。これに対して本研究では、同一の興味・関心を持ち、類似した行動をとるユーザ間で情報を流通させることで行動要因情報を得る機会を高めることを検討する。具体的には、情報を要求する際に、システムが情報源となるユーザを行動予測に基づいて選択を行ったうえで、これらの情報源に対して問い合わせを行う手法を提案する。本論文では、行動予測の結果が同一の情報源を検出するために、複数ユーザの興味・関心などの性質から潜在的属性を定義し、潜在的属性および実際の行動履歴に基づいた行動予測の精度評価実験の結果を示す。精度評価の結果、正解データに対して、潜在的属性を考慮したクラスタモデルの正答率がすべてユーザを含む混同モデルの正答率よりも高いことが示され、クラスタモデルの有用性が確認された。

1. はじめに

GPS や無線強度などを用いることで屋内外において高い精度での測位が可能となり、このような機能を備えた携帯端末が普及している。それに伴い、ユーザの位置情報を用いたナビゲーションや情報提示などの様々なサービスが実用化されている。また、Twitter におけるジオタグ付きツイートや Foursquare, Gowalla などの LBSN (Location Based Social Network) サービスの普及に伴い、その場その時の状況を発信したり、位置情報の共有や周辺情報の取得をしたりすることが可能である。

一方で、過去の行動の分析およびユーザのデモグラフィック属性を元にした行動予測の結果に応じて、システムがユーザに有用な情報や販売促進の向上を目的とした広告などを配信するレコメンドサービスも普及し、ユーザが未来の行動を決める一つの要因となっている。しかし、現状のシステムは、限定された属性に従って、ユーザに提供する情報を選別しているため、本来なら、受け取ったユーザの行動に影響を与え得る情報の一部が失われるという問題がある。これに対して本研究では、測位が可能で小型無線通信端末を所持するユーザが、実世界情報のセンシングを行い、相互に情報を交換し活用する情報基盤を想定したう

えで、同一の興味・関心を持ち、類似した行動をとるユーザ間で情報を流通させることで、行動の要因となる情報を得る機会を高めることを検討する。

社会心理学における社会的証明の原理によると、人間が他者の行動を参考にして自分の行動を決める条件の一つとして他者との類似性が挙げられる。人間がどのように振る舞うかを考えるとき、最も参考しているのが自分と類似した他者の行動であるということがいくつかの研究によって明らかになっている [1]。したがって、人間は自分とまったく異なる性質の他者よりも、同一の行動や嗜好をもつ類似の他者が示す行動に従いやすいといえる。

こうした背景から、同一の興味・関心を持ち、類似した行動をとるユーザ間における情報は、それを受け取ったユーザのその場その時の意思決定に影響を与えられられる。このような意思決定において、ユーザにとって今後の目的に関わる情報を事前に知っておくことは、非常に有用である。例えば、外出前に今後の気象情報が得られることによって、ユーザの行動の質が向上するというように、未来のイベントに関する情報を前もって取得できることが、意思決定において影響があることは明らかである。

そこで本研究では、まずあらかじめ蓄積された各ユーザの興味・関心などに基づく属性および位置情報履歴から、確率過程モデルを用いて得られる行動予測の結果に応じて、

¹ 明治大学大学院先端数理科学研究科先端メディアサイエンス専攻

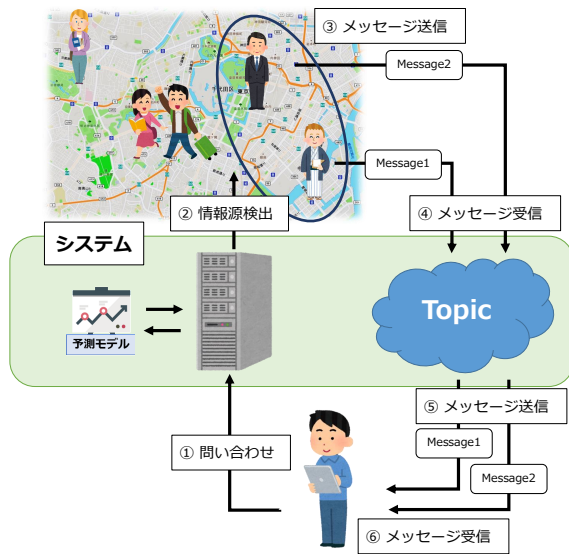


図 1 行動予測とメッセージ制御システム

情報源となるユーザを選択し、次に、これらの情報源に対して問い合わせメッセージの制御を行うことによって、情報を流通させるシステムの実現を目的とする。例えば、あるユーザと同一属性であり、かつ過去の行動履歴から類似した行動をとると予測された複数のユーザを情報源としてシステムが検出し、ユーザ間で情報を流通させるようにする。これにより、情報を受け取ったユーザの未来の行動に影響を与え得る有用な情報の取得を実現する。図 1 に提案システムのモデルを示す。本システムは、予測モデルを元に情報源の検出を行う機能とメッセージの送受信を行う機能を備える。

我々はこれまでに、過去の人々の経路データから性別による行動予測モデルの構築を行い、予測モデルが属性ごとに異なる傾向を持つことを確認した。しかし、人間は、性別や年齢、職業などといったデモグラフィック属性のほかに、興味・関心、趣味、性格などといった隠れた属性、つまり潜在的属性を有している。そのため、行動要因となる情報源を抽出するうえで、潜在的属性に基づく行動予測の検証を行う必要がある。さらに、正解データと比較したときの予測モデルの精度についての議論が不十分であった。そこで本論文では、提案システムの行動予測モデルにおける潜在的属性に焦点を当て、検証を行う。複数ユーザの興味・関心などの性質によって潜在的属性を定義し、潜在的属性および実際の行動履歴に基づいた行動予測モデルの精度評価実験を行った。

本論文の構成は以下のとおりである。本章に引き続き、2章では位置情報に基づく情報推薦技術や行動予測および社会的影響の数学モデルに関する先行研究を取り上げる。次に3章では、潜在的属性の定義および確率過程に基づく行動予測の原理について説明する。4章では潜在的属性に

基づく行動予測実験により提案手法の実現可能性を議論する。最後に5章で本論文の内容をまとめ、今後の課題を考察する。

2. 関連研究

これまでに、目的地予測や動線解析などといった研究が数多く行われている。その一つとして、LBSN へのチェックインデータを元に、ユーザに対して POI (Point of Interest) のレコメンドを行う研究がある。Cheng らの研究 [2] では、ユーザのチェックイン確率からモデルを構築し、さらに友人関係などによる社会的影響を組み込んだことによって、従来の POI レコメンド手法に比べて優れた結果が得られている。Yang らの研究 [3] では、チェックインから抽出された場所に対する嗜好と、投稿されたコメントのテキストを解析し、それらを組み合わせモデルによって、パーソナライズされた POI のレコメンドを行う手法を提案している。また、位置情報に基づく行動予測に関する様々な研究が行われており、Do らの研究 [4] では、将来ユーザがどの場所を訪れて、携帯端末でどのアプリケーションを使用するのかといったことを複数のセンサを用いた統計的手法によって予測を行っており、パーソナライズモデルに用いる学習データが限られている場合、パーソナライズモデルとジェネリックモデルを組み合わせることで予測精度が向上することを示している。和田らの研究 [5] では、位置予測の結果に基づくターゲティング広告の配信を目的とし、ユーザの行動履歴から隠れマルコフモデルを用いて予測を行なっている。これによって、ユーザが将来訪れる地点の情報を配信する手法を提案している。これらに対して本研究では、ユーザの行動の要因となる情報を得る機会を高めるために、情報源となるユーザを潜在的属性および行動予測の結果から検出することを目的とする。

一方で、社会システムの中で人間の行動が他者からどのような影響を受けているかという問題について様々な研究が行われてきた [6]。Pan らの研究 [7] では、影響モデルの構築を行なっている。実世界における人々の集団 C を構成するある個人 c において、時点 t での行動を促す観測不可能なアイデアを $h_t^{(c)}$ 、観測可能な行動を $O_t^{(c)}$ としたとき、ある時点での観測可能な行動がどのように変化するかについては、各ユーザの観測不可能なアイデアによって左右されるため、 $h_t^{(c)}$ のときに $O_t^{(c)}$ が表れる確率は $P(O_t^{(c)} | h_t^{(c)})$ と表すことができる。そして社会的影響について、ある人物の状態が別の人物の状態に影響を与えることと考え、各ユーザのアイデア $h_t^{(c)}$ とその直前の時点での集団全員のアイデア $h_{t-1}^{(1)}, \dots, h_{t-1}^{(C)}$ との間にある条件付き確率を、 $P(h_t^{(c)} | h_{t-1}^{(1)}, \dots, h_{t-1}^{(C)})$ と表し、モデル化を行なっている。

以上を踏まえ、本研究が目指すユーザの未来の行動に影響を与え得る有用な情報の流通において、行動や興味・関心などに基づく情報源の選択は効果的であるといえる。

3. 潜在的属性および位置情報による行動予測

本章では、潜在的属性および位置情報に基づく行動予測手法について述べる。

3.1 手順

本研究では、1地点を1つの状態とみなしたとき、過去の蓄積された人々の移動経路による状態遷移のモデル化を行う。モデルを元に情報源となり得るユーザを予測および検出して、ユーザ間でメッセージの送受信が可能となる手法を提案する。本論文では、潜在的属性に着目し、事前にその属性を有するユーザがどのような経路をとって行動したのかという経路データを収集し、データを元にマルコフ連鎖により予測モデルを構築する。そして、情報を得たいユーザはシステムに対して問い合わせを行い、問い合わせを行った時点において、同一の潜在的属性であり、かつ類似した行動をとると予測された複数ユーザをシステムがリアルタイムに検出する。検出された複数のユーザを集約し、グループ化することで情報の流通が可能となる。以降、3.2章では潜在的属性の定義について、3.3章では行動予測のための確率過程を用いたモデルの構築手法についてそれぞれ述べる。

3.2 ユーザに対する潜在的属性の定義

蓄積された情報から階層的クラスタリングを用いて各ユーザを潜在的属性に分類する。ここで、潜在的属性とは、興味・関心などの性質によって定まる未知の属性のことをいう。人間の行動は、各個人の潜在的属性に起因する。そのため、同一の潜在的属性を有するユーザは似通った行動をとることが考えられる。このような潜在的属性の近いユーザを集約することによって、抽象化された階層構造を持つユーザ群を形成することが可能である。

そこで、ある指標に従って個体間の類似度から、クラスタを形成する分析手法である階層的クラスタリングを用いて、分類を行う。本研究では、その指標として任意の地点における訪問頻度に着目する。あるユーザの任意の訪問地における訪問頻度が少ないとき、その訪問地に対するユーザの興味・関心度は高いとはいえないが、訪問頻度が多いとき、ユーザの興味・関心度は高いと考えられることから、各ユーザの興味・関心度を表すものと仮定する。各ユーザの訪問地に対する訪問頻度に基づいて、ユーザ間の類似度からクラスタリングを適用し、ユーザに対して属性を定義する。

3.3 属性に基づくマルコフ連鎖による行動予測手法

蓄積された過去の人々の移動経路データから、マルコフ連鎖によって情報源となる可能性の高いユーザの予測を行う。マルコフ連鎖とは、次の状態が過去の状態に依存せず

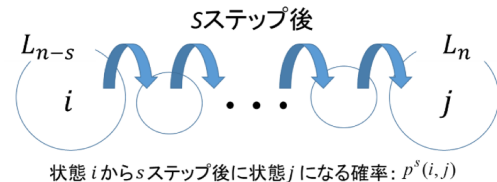


図 2 マルコフ連鎖による確率過程

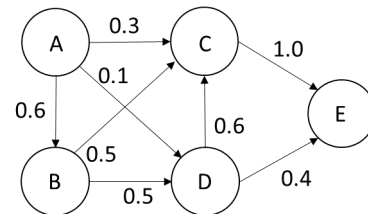


図 3 状態遷移の例

現在の状態のみによって決まるマルコフ性を持つ離散的な確率過程である。例として、気象や株、為替など予測が難しいものに対して、マルコフ連鎖で近似することによってモデル化を行っている。提案手法では、このようなマルコフ性を適用して、人間の属性、行動といった事象から確率モデルを構築し、予測を行う。

ある属性 $Attr$ を有する m 人のユーザ u それぞれの移動に伴う位置 \mathbf{L} の時系列 $\mathbf{L}_u = \{L_{u,1}, L_{u,2}, \dots, L_{u,n}\} (u = 1, 2, \dots, m)$ を保持する。 m 人のユーザがある地点 L_k を訪問した回数と地点 L_{k+1} を訪問した回数から、連続する 2 地点間の状態遷移確率を算出する。 L_k から L_{k+1} への状態遷移確率は、式 (1) のように表せる。

$$P(L_{k+1} | L_k, Attr) = \frac{\sum^m \text{地点 } L_{k+1}(u, Attr) \text{ 訪問回数}}{\sum^m \text{地点 } L_k(u, Attr) \text{ 訪問回数}} \quad (1)$$

ある地点 L_{n-s} から任意の地点 L_n に至るまでの状態遷移確率が既知であるとき、 L_{n-s} から L_n に至る確率が式 (2) により得られる。なお、変数 s は L_{n-s} から L_n までの遷移回数を表す。図 2 にマルコフ連鎖による遷移確率導出過程を示す。

$$P(L_n = j | L_{n-s} = i, Attr) = p^s(i, j) \quad (2)$$

ステップ数 s を $m+n$ と表すとチャップマン-コロモゴロフ方程式より、式 (3) が得られる。この式は、状態 i から状態 j に $m+n$ ステップで至る確率を表している。指定された状態 j およびステップ数に対して、尤度が最大となる状態 i は、状態 j に至る可能性の高いユーザが存在する地点となる。

$$p^{m+n}(i, j) = \sum_k p^m(i, k) p^n(k, j) \quad (3)$$

ある空間におけるユーザの行動履歴について、図 3 のような状態遷移を例にすると、2 地点間の遷移事前確率行列

M_{pri} は以下のように表される。

$$M_{pri} = \begin{pmatrix} p(A, A) & \dots & p(A, E) \\ \vdots & \ddots & \vdots \\ p(E, A) & \dots & p(E, E) \end{pmatrix} = \begin{pmatrix} 0 & 0.6 & 0.3 & 0.1 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 1.0 & 0 \end{pmatrix} \quad (4)$$

ここで、地点 E に 3 ステップで到達するユーザの予測を考える。式 (2) より、3 ステップ後に地点 E に到達する現在地点ごとの確率は、式 (5) で計算することができる。

$$P(L_n = E | L_{n-3} = i, Attr) = p^3(i, E) \quad (5)$$

よって、到達する現在地点ごとの事後確率行列 M_{post} を以下のように得ることができ、地点 E に至る最も尤度が高い (0.48) 地点 A が予測結果となる。

$$M_{post} = \begin{pmatrix} 0 & 0 & 0.18 & 0.34 & \boxed{0.48} \\ 0 & 0 & 0 & 0.7 & 0.3 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0.24 & 0.6 & 0.16 \\ 0 & 0 & 0 & 0.4 & 0.6 \end{pmatrix} \quad (6)$$

以上の議論に基づき、潜在的属性に基づく行動予測手法の評価を行う。

4. 潜在的属性に基づく行動予測実験

潜在的属性に基づく行動予測結果の精度を確認するための実験と結果について述べる。

4.1 実験

マルコフ連鎖を用いた潜在的属性に基づく行動予測モデルによって、行動予測結果の精度を確認する実験のため、データセットとして、2017 年 4 月 1 日～8 月 31 日の北海道におけるジオタグ付きツイートのうち、Foursquare を介して Twitter に投稿されたツイートを Twitter Streaming API^{*1}を用いて収集を行なった。潜在的属性および行動予測に関する実験方法、制約条件は、以下のとおりである。

潜在的属性について Foursquare を介して投稿されたツイート本文には訪問地名称が含まれており、Foursquare API^{*2}を利用することで訪問地に付与されているカテゴリを取得することができる。表 1 に、訪問地に対するカテゴリの例を示す。それを元に、各ユーザの訪問地における訪問回数をカテゴリ別に算出し、訪問回数の多かった上位 15 個のカテゴリ (Arcade,

*1 <https://dev.twitter.com/docs/streaming-apis>

*2 <https://developer.foursquare.com>

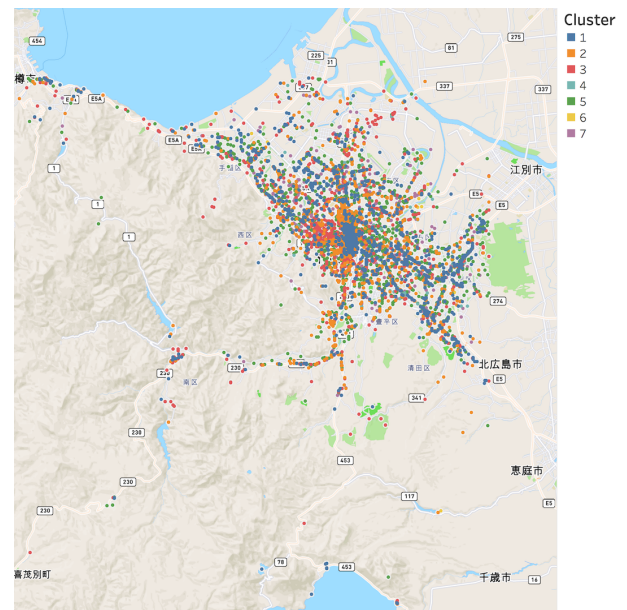


図 4 2017 年 4 月 1 日～8 月 31 日までに北海道札幌市周辺で収集されたジオタグ付きツイート

表 1 訪問地に対するカテゴリ例

訪問地	カテゴリ
Starbucks Coffee 札幌西町店	Coffee Shop
ドトールコーヒーショップ JR 札幌北口店	Coffee Shop
札幌駅	Train Station

表 2 各クラスタのユーザ数, 時系列数

クラスタ	1	2	3	4	5	6	7
ユーザ数	141	189	290	57	377	29	135
学習用 時系列数	2247	3112	2635	1039	8345	360	3744
正解用 時系列数	250	346	293	116	927	40	416

Scenic Lookout, Bridge, Shopping Mall, Hot Spring, Electronics Store, Cafe, Japanese Restaurant, Sushi Restaurant, Coffee Shop, Sake Bar, Park, Fast Food Restaurant, Ramen Restaurant, Stadium) をクラスタリングに用いる指標とした。それぞれのカテゴリの訪問回数を総訪問回数で正規化し、相対頻度に従ってクラスタリングを行うことでユーザに対して潜在的属性を定義した。ただし、駅やバス停、ホテルなどユーザの興味・関心には無関係であると考えられる訪問地については、クラスタリングを行う際には使用しないものとした。個体間距離にユークリッド距離、クラスタ間距離に Ward 法を使用して階層的クラスタリングを適用した結果を図 5 に示す。本論文では、最大距離の 5 割を基準にクラスタの分割を行ない、7 種類の潜在的属性にユーザを分類した。

行動予測について 北海道札幌市を中心とする地域を分析対象領域とし、10 × 10 の格子状の 100 エリアに区

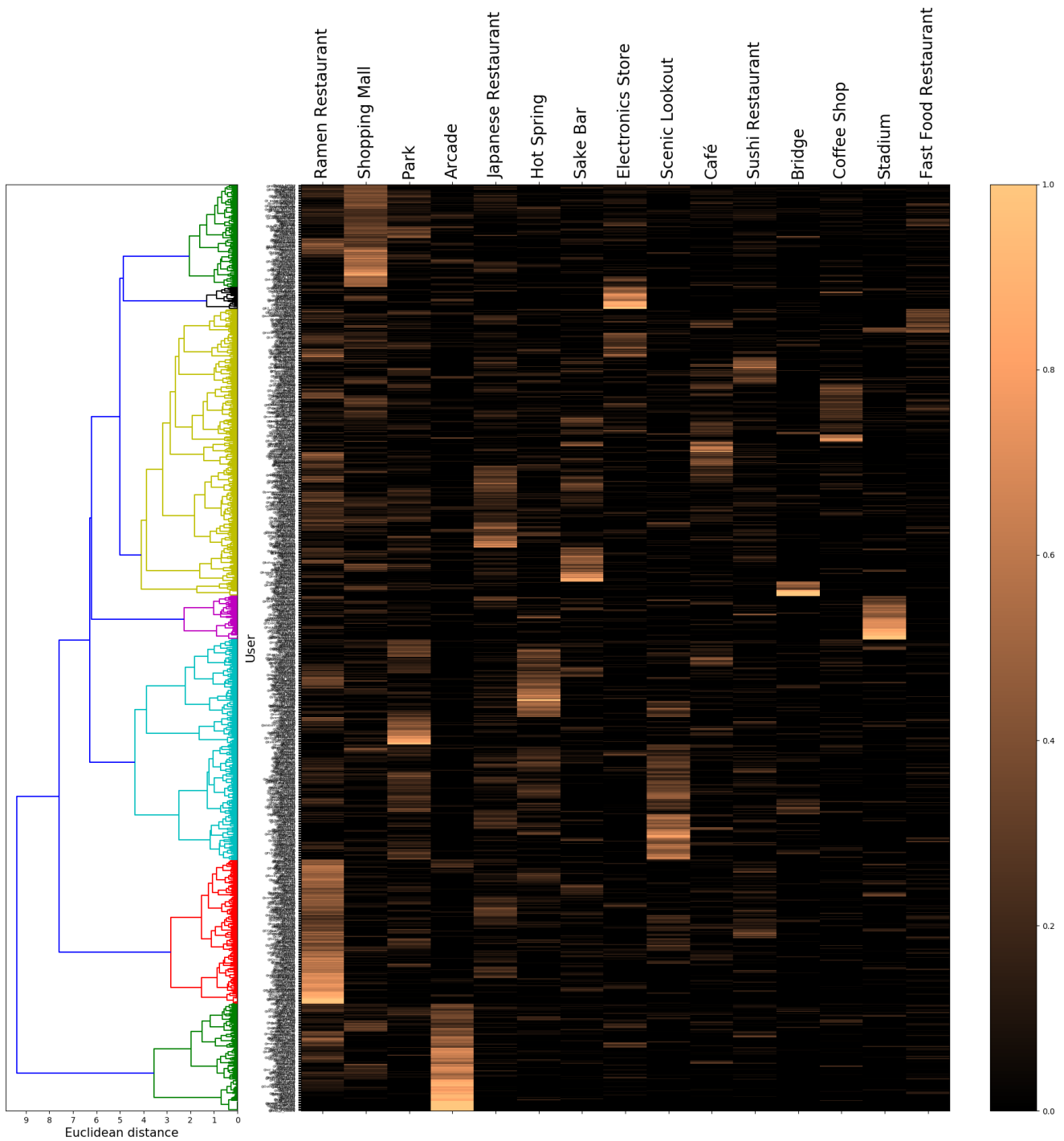


図 5 階層的クラスタリングを用いたユーザの分類

表 3 クラスタモデルおよび混同モデルの予測精度

クラス	1	2	3	4	5	6	7
クラスタモデル	38.5 %	32.3 %	30.2 %	21.4 %	34.0 %	82.6 %	42.4 %
混同モデル	29.5 %	30.1 %	31.3 %	17.9 %	34.5 %	43.5 %	31.4 %

表 4 統計量

	平均	標準偏差	分散
クラスタモデル	40.2 %	19.8	337.1
混合モデル	31.2 %	7.57	49.1

切ったエリア間の移動を状態遷移とする。また、同一ユーザであっても異なる日のツイートは別ユーザと見なしたときのユーザ u の移動に伴う位置時系列 $L_u(u = 1, 2, \dots, m)$ を保持する。分析対象領域にお

る4月1日～8月31日までの期間に収集された68,284件のツイートを時系列 L_u に変換し、各クラスタの時系列数の9割をモデル構築用、残りの1割を正解データとし、交差検証法によって評価を行なった。対象領域で収集されたツイートを図4*3に示す。各クラスタにおけるユーザ数、時系列数を表4に示す。ただし、時系列の長さ1であるものについては、移動と見なすことができないため、除外するものとした。行動予測モデル構築にあたって、学習データの各クラスタから50件の時系列をランダムに選択する操作を10回繰り返すことによって得られた学習データに従って、モデルの構築を行なった。

4.2 実験結果と考察

各潜在的属性に基づくクラスタモデルおよびすべてのユーザが混同したデータから同様の方法で構築した混同モデルそれぞれを正解データに適用し、予測精度の評価を行った。正解データの時系列 $L_u = \{L_{u,1}, L_{u,2}, \dots, L_{u,n}\}$ の初期地点 $L_{u,1}$ における終点 $L_{u,n}$ を予測したときの正答率を表3に示す。全クラスタにおいて混同モデルを適用した正答率よりもクラスタモデルを適用した場合のほうが高い正答率となっている。また、クラスタ6においては、非常に高い正答率を示しており、これは、表4からわかるように、クラスタ6のユーザ数が他と比べて少なく、クラスタモデルがユーザの行動に近いモデルであると考えられる。つまり、より細分化したクラスタを形成することで、精度の高い予測モデルの構築が可能であると考えられる。

5. おわりに

本論文では、行動予測の結果が同一の情報源を検出するために、複数ユーザの興味・関心などの性質から潜在的属性を定義し、潜在的属性および実際の行動履歴に基づいた行動予測の精度評価実験の結果を示した。精度評価の結果、正解データに対して、潜在的属性を考慮したクラスタモデルの正答率がすべてのユーザを含む混同モデルの正答率よりも高いことが示され、クラスタモデルの有用性が確認された。混同モデルの平均正答率は31.2%に対して、クラスタモデルの平均正答率が40.2%となり、平均9.0%の精度向上が確認された。今後は、行動要因情報源抽出のためのクラスタ分割基準値の最適解の導出および情報流通のためのシステム設計を行う。

参考文献

- [1] R. B. Cialdini, Influence: Science and practice. Scott, Foresman and Company (1988).
- [2] C. Cheng, H. Yang, I. King, M. R. Lyu. Fused matrix factorization with geographical and social influence

- in location-based social networks, *Proceedings of AAAI* (2012).
- [3] D. Yang, D. Zhang, Z. Yu, Z. Wang. A sentiment-enhanced personalized location recommendation system. *Proceedings of the 24th ACM Conference on Hypertext and Social Media(HT 13)*, pp. 119-128 (2013).
- [4] T. M. T. Do, D. G. Perez, Where and what: using smartphones to predict next locations and applications in daily life, *Pervasive and Mobile Computing*, Vol. 12, pp. 79-91 (2014).
- [5] 和田 計也, 増田 早紀, 梅林 泰孝, 行動クラスタリングと隠れマルコフモデルを用いたユーザの位置予測モデル, 第29回人工知能学会全国大会論文集, 3C4-1 (2015).
- [6] A Pentland, ソーシャル物理学-「良いアイデアはいかに広がるか」の新しい科学- (2015).
- [7] W. Pan, W. Dong, M. Cebrían, T. Kim, J. Fowler, A. Pentland, Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems, *Signal Processing Magazine, IEEE*, Vol. 29, No. 2, pp. 77-86 (2012).

*3 Map data ©OpenStreetMap contributors.