

個人の移動履歴と群衆の移動傾向に基づく 未訪問地も扱える目的地予測

西村 拓哉[†] 吉田 康仁^{††} 赤木 康紀[†] 倉島 健[†] 戸田 浩之[†]

[†] 日本電信電話株式会社 NTT サービスエボリューション研究所 〒239-0847 神奈川県横須賀市光の丘 1-1

^{††} 電気通信大学大学院 情報理工学研究所 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: †{takuya.nishimura.fk,yasunori.akagi.cu,takeshi.kurashima.uf,hiroyuki.toda.xb}@hco.ntt.co.jp,

††yosida@kookai.pc.uec.ac.jp

あらまし スマートデバイスなどを用いて得られる位置情報を用いて移動中のユーザの目的地を予測することは、情報配信サービスの高度化などにおいて重要である。従来の、ユーザの過去の移動履歴を事前に学習することで予測モデルを構築する手法では、移動履歴の少ないユーザの移動や未知の目的地への移動を扱うことが難しいという問題があった。本研究では群衆全体の移動傾向データを用いた大域予測モデルを導入し、それと個人の移動履歴を用いた個人予測モデルとのアンサンブルによってこの問題を解決する。本研究の貢献は、(1) 出発地と目的地の組み合わせし分からない移動傾向データを用いて高精度に個人の目的地を予測する手法の提案、(2) 個人の履歴に基づく目的地予測手法について確信度を出力するように拡張し、それを考慮して二つのモデルをアンサンブルさせることによって、未知の目的地・既知の目的地のどちらへの移動も高精度に予測できる手法の提案、の二つである。首都圏に住むユーザ 30 人の 3 週間分の移動軌跡データと、首都圏の人口分布データから算出した群衆の移動傾向データを用いて実験を行い、提案手法は従来手法よりも高い精度で目的地を予測できることを確かめた。

キーワード 目的地予測, GIS, 人流, LSTM, モバイル空間統計

1. はじめに

スマートデバイスなどを用いて得られる位置情報を用いて移動中のユーザの目的地を予測することは、情報配信サービスの高度化などにおいて重要である。情報配信サービスの例としては、NTT ドコモ社の My Daiz に代表される個人向けエージェントサービスなどが挙げられる。目的地を予測して行う情報配信の例としては、移動中のユーザに対して店舗の広告情報を配信する際に現在地周辺の情報ではなく目的地周辺の情報を配信するなどが考えられる。

移動中のユーザの目的地を予測する先行研究には様々なものが存在する [1] [2] [3] [4]。これらは、現在移動中のユーザの出発地点から現在地点までの移動軌跡をクエリとして、それに対して最終的な到達地点がどこになるかを予測して出力するものである。これらの研究の多くが予測対象のユーザや類似したユーザの過去の移動軌跡を学習し、それに基づいて予測を行うものである。そのような手法には以下の二つの問題がある。

- 移動履歴データがない、あるいは少ないユーザの目的地を予測することができない
- 過去訪問したことがない場所への移動を予測することが出来ない

これらの問題は、目的地の予測結果に基づいてユーザに対して情報配信を行う場合などには高い頻度で発生する。ユーザ向けの情報配信サービスにおいて目的地予測を用いる場合、新規ユーザはその移動履歴を取得できない。また過去の移動履歴を多く取得しているユーザについても、観光地などでは過去に訪

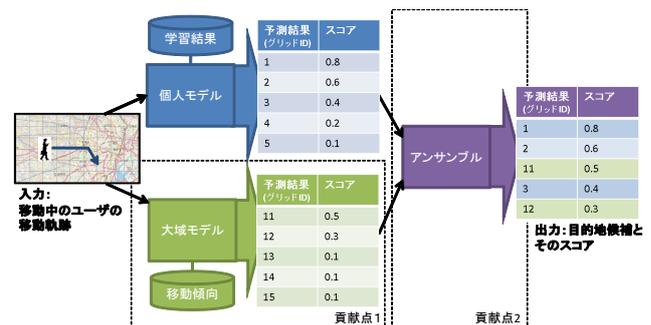


図1 本研究では、移動中のユーザの移動軌跡を入力クエリとして、目的地候補とそれらのスコアを出力することに取り組む。本研究の貢献点は、(1) 群衆の移動傾向データに基づく大域予測モデルの提案、(2) 個人の履歴に基づく個人モデルと大域予測モデルのアンサンブルの提案、の2点である。

問したことがない場所を目的地として移動することがある。

本研究では上記の問題を解決し、移動中ユーザの移動軌跡を入力クエリとして目的地を予測するタスクにおいて、移動履歴がないユーザの移動や過去に訪問したことがない場所への移動においても目的地予測を可能にする技術の実現に取り組む。

本研究では2つのアプローチの組み合わせによって上記課題に取り組む。本研究における入出力と手法の全体像を図1に示す。一つ目のアプローチとして、移動履歴がないユーザの目的地を予測するために、群衆の移動傾向データを用いる。移動傾向データとは、その地域にいる人々全体がどのように移動しているのかを記録したデータであり、特に本研究ではある時間帯

にあるグリッドからグリッドに何人が移動したのかを記録したデータを用いる。このデータを用いると、履歴がないユーザについても、どこに向かう確率が高いのかを推定することが可能になる。二つ目のアプローチとして、個人履歴に基づく予測モデルと移動傾向データに基づく予測モデルのアンサンブルによる予測を行う。移動傾向データに基づく予測は、個人履歴に基づく予測に比べると精度が低くなるため、個人履歴に基づく予測が可能であればそちらを用いたい。しかし履歴が少ない場合や未訪問地への移動を予測する場合などは、移動傾向データに基づく予測を用いたい。これを実現するために、二つのモデルに目的地候補と候補それぞれの予測の確かさを示すスコアを出力させ、それらの重み付き線形和によって最終的なランキングを取得する。

本研究の貢献は以下の3点である。

- 出発地と目的地の組み合わせしか分からない移動傾向データを用いて高精度に個人の目的地を予測する手法を提案した。
- 個人の履歴に基づく目的地予測手法とのアンサンブルによって、未知の目的地・既知の目的地のどちらへの移動も高精度に予測できる手法を提案した。
- 首都圏に住むユーザ30人の3週間分の移動軌跡データと、首都圏の人口分布データから算出した群衆の移動傾向データを用いた実験によって、提案手法の有効性を検証した。

2. 関連研究

2.1 ユーザの移動履歴に基づく目的地予測

移動中のユーザの最終的な目的地を予測する研究は多数存在するが、基本的な考え方は同一であり、ユーザ自身の過去の移動履歴と現在の移動軌跡のマッチングを何らかの方法で取り、類似したケースでの目的地を予測結果として出力するものである。Marmasseらは現在地までの軌跡と部分一致する軌跡をユーザの過去の移動軌跡から抽出して予測を行っている[2]。Xueらは移動軌跡を部分軌跡に分割してから合成することにより、プライバシーに配慮しつつ問い合わせや目的地の制限を緩和することに取り組んだ[3]。瀧本らはその手法を拡張し、ユーザ個人の履歴が少ないために本来は移動していない部分について補完するような軌跡を導入してデータスパースネスの問題に取り組みつつ、さらに時間帯を考慮することで精度を向上させた[4]。Endoらは過去の移動軌跡をLSTMでモデリングすることによって、現在の移動軌跡と完全に合致した移動軌跡が過去にない場合でも類似した軌跡の検索を行えるようにすることで高精度での予測を実現している[1]。

本研究ではこれらの先行研究の手法を利用しつつ、これらの先行研究では扱うことのできないユーザの履歴が存在しない場合や訪問したことがない場所への移動も扱うために移動傾向データを用いた予測とのアンサンブルを提案する。

2.2 群衆の移動傾向データ

本研究で移動傾向データと呼ぶものは、ある時間帯にあるグリッドからあるグリッドに移動した人数を示すものである。例えば、7時台から8時台にかけて、東京駅周辺の500m四方の

グリッドから横浜駅周辺の500m四方のグリッドには150人が移動した、というようなデータである。日本国内での移動傾向データについては、政府や自治体が都市計画を目的としてアンケートによって1日の移動内容などを取得しているパーソントリップ調査[5]や、NTTドコモ社の提供しているモバイル空間統計の動態統計[6]などが挙げられる。しかしこれらのデータは得られる日時やエリアの組み合わせが非常に限定的であるという問題がある。これはデータ取得のコストやユーザプライバシーなどが原因であり、解決することが難しい。

近年では上記の問題を解決するために、人口統計データから移動傾向データを推定することに取り組む研究が行われている。人口統計データとは、各時間帯における各エリアの人口を記録したデータであり、日本国内であればNTTドコモ社の提供しているモバイル空間統計の分布統計[7]などが例として挙げられる。人口統計データは移動傾向データと比較してプライバシーの問題などが少ないため、より広いエリアや時間帯について取得することが可能である。IwataらはCollective Graphical Modelと呼ばれる確率モデルを用いて人口分布データから移動傾向データを推定することに取り組んだ[8]。Akagiらはエリア間の遷移確率を直接求めるのではなく、エリア間の距離とエリア固有のパラメータによってモデリングすることでパラメータ数を減らし高精度に移動傾向データを推定することに取り組んだ[9][10]。

本研究では移動傾向データとしてAkagiらの手法で推定したものをを用いて、目的地予測を行う。

2.3 モデルアンサンブル

一般的にアンサンブル学習と呼ばれるものは教師データから多数のモデルを同時に学習するものであり[11]、本研究のアンサンブルは性質の異なる二つのモデルのアンサンブルを行うものであり、これらとは異なる。

本研究に近い研究としては、ランク学習(Learning to rank)が挙げられる[12]。これは情報検索分野において、複数のモデルがあるときに、入力となる検索クエリにたいしてそれぞれのモデルが出力する文書のランキングから最終的な文書のランキングを得るためのモデルへの重みづけを、教師あり学習として最適化するものである。本研究における移動中ユーザの現在の移動軌跡を検索クエリ、グリッドを文書とみなせば、ランク学習と類似した考え方でモデルの重みを最適化することが可能である。ただし、情報検索分野ではクエリに対する明確な正解は存在せず、候補文書ごとにクエリへの適合度がスコアとしてそれぞれ与えられているのに対し、本研究では正解のグリッドが一つしかないという違いがある。

3. 提案手法

本章ではモデルアンサンブルに基づく目的地予測手法について述べる。本章以降で用いられる記号の一覧を表1に示す。

最初に、移動傾向に基づいた目的地予測を行うモデルである大域予測モデルに関して、基本的な推定の考え方と、推定対象のユーザの移動内容に基づいて目的地予測精度を向上させる方法について述べる。次に、ユーザの過去の移動履歴に基づく予

表 1 本稿で用いられる記号一覧

記号	説明
A	全てのグリッドの集合
\vec{u}	ユーザ u の現在の移動方向
$\vec{i}j$	グリッド i から j への方向
$P_{dist}(D)$	残りの移動距離が D である確率
$dist(i, j)$	グリッド i と j の直線距離
M_{ij}	移動傾向データにおける、 その時間帯におけるグリッド i から j への移動人数
t	入力移動軌跡中において出発時刻を 1 としたときの現在時刻
$P_g(i)$	大域予測モデル g が出力した、グリッド i が目的地である確率
$P_p(i)$	個人予測モデル p が出力した、グリッド i が目的地である確率
C_p	個人予測モデル p の出力全体に対する確信度
$S(i)$	アンサンブルモデルが出力した、グリッド i のスコア

測を行う個人予測モデルの拡張について述べる．本研究では個人予測モデルとして Endo らの手法 [1] を用いる．また本研究では、アンサンプルの前提としてそれぞれのモデルが予測結果として目的地候補のグリッドとそれらに対するスコアを出力しているものとする．このスコアとは、目的地候補がどれほど正しいと考えられるかを示す指標として機能するものである．しかし、Endo らの手法ではスコアを取得できないため、スコアを取得できるように拡張を行う．その後、二つのモデルのアンサンブル方法について述べる．

3.1 群衆の移動傾向に基づく大域予測モデル

群衆の移動傾向データとは、2 章で述べたように、ある時間帯のあるグリッドからあるグリッドへの移動人数を、集計するか何らかの他の情報から推定したデータである．これを用いれば、履歴のないユーザに関してもどのような移動を行うか予測することが可能になる．

移動傾向データを用いたシンプルな目的地予測手法としては、移動中のユーザの現在地点からその時刻に最も移動している人数が多いグリッドを出力することが考えられる．しかし、この手法はユーザの現在の移動内容を考慮していないため、以下の二つの問題がある．

移動方向を考慮することが出来ない：東から西へとまっすぐ移動中のユーザの目的地が、現在のグリッドからずっと東にあるといったようなことは考えにくい．こういったことを目的地の予測に反映させる必要がある．

残りの移動距離を考慮することが出来ない：電車で移動中のユーザの目的地は、徒歩で移動中のユーザと比較すると遠距離にある可能性が高いと考えられる．入力クエリの移動軌跡から残りの移動距離がどの程度であるかを推定して予測に反映させる必要がある．

上記の問題を解決するために、ユーザの移動方向と残りの移動距離を考慮する手法を提案する．まず、移動方向の考慮方法について述べる．移動中のユーザの今後の移動方向について、本研究では現在の移動方向と類似すると考える．現在の移動方向とは、クエリによる問い合わせが行われた瞬間の移動方向のことである．これ以外の考え方として、出発地点からの現在地点までの移動方向と類似するといった考え方などもありうるが、予備実験の結果、現在の移動方向への類似度の方が高いことが

確かめられた．この考えに基づき、ユーザの現在の移動方向をクエリの移動軌跡から取得し、その方向にあるグリッドほど目的地になる確率が高くなるようにする．次に、残りの移動距離の考慮方法について述べる．残りの移動距離に影響する要素として、本研究では、出発から現在までの経過時間と、現在の移動速度の 2 点を考慮し、それらから残りの移動距離に関する確率分布が得られるものとする．経過時間を考慮する理由は、日常における移動において総移動時間が 1 時間以上になることは少ないなどの知識を考慮すると、出発からの経過時間によって残りの移動距離が変化すると考えられるためである．移動速度を考慮する理由は、現在の移動手段によって残りの移動距離が変化するためである．例えば車や電車で移動中のユーザの残りの移動距離は、徒歩で移動中のユーザに比べると長いと考えられる．予備実験の結果、残りの移動距離に関する確率分布は指数分布の形状をしていることが分かったので、経過時間と移動速度の二つから指数分布のパラメータが決定されるモデルとする．パラメータの学習は、経過時間と移動速度をそれぞれビンで区切り、それらの組み合わせ毎にデータを分割してそれぞれ学習を行う．経過時間は 20 分幅のビンで区切り、移動速度についてはユーザの移動状態が徒歩であるか乗物であるかを考慮するために徒歩の速度である時速 2km よりも速いか遅いかで区切るものとした．

上記を踏まえ、現在グリッド i にいるユーザの目的地が j である確率 θ_{ij} を、以下のように定式化する．

$$\theta_{ij} = \frac{(\vec{u} \cdot \vec{i}j + 1) \times P_{dist}(dist(i, j)) \times M_{ij}}{\sum_{k \in A} (\vec{u} \cdot \vec{i}k + 1) \times P_{dist}(dist(i, k)) \times M_{ik}} \quad (1)$$

ここで、 \vec{u} はユーザの現在の進行方向を示し、 $\vec{i}j$ はグリッド i から j への方向を示す． $P_{dist}(D)$ は残りの移動距離が D である確率を示す関数であり、現在の移動速度と経過時間によって決定される．この関数は指数分布の形状であり、指数分布のパラメータは教師あり学習で最適化する． $dist(i, j)$ はグリッド i と j の直線距離を示す． A は全てのグリッドの集合を示す． M_{ij} は移動傾向データにおける、その時間帯におけるグリッド i から j への移動人数を示す．

3.2 個人モデルの拡張とモデルアンサンブル

3.2.1 個人モデルの拡張

まず本研究で個人モデルとして用いる既存手法 [1] について説明する．これは移動軌跡をグリッドの系列としてみなし、LSTM [13] を用いて目的地を予測するものである．実際の予測は以下のように行われる．

(1) 対象ユーザの過去の移動を学習した LSTM に、出発時点の時刻 1 から現在時点の時刻 t までのグリッドの系列を入力し、次の時刻 $t+1$ にいるグリッドがどこになるかの確率分布を予測する．

(2) 得られた分布に従って次のグリッドをサンプリングし、得られたグリッドを LSTM に入力しさらに次のグリッドを予測する．

(3) これを、過去に訪問したグリッド (目的地候補) に到達するまで繰り返す．

(4) 上記のプロセスを十分な回数繰り返し、到達した回数を試行回数で割ったものを各目的地候補が目的地になる確率とみなす。

この手法は各グリッドが目的地になる確率は出力できるが、これは過去に訪問したグリッドのいずれかが目的地であると仮定した場合の確率である。しかし、実際には未訪問地への移動も存在する。そのため、本手法の出力する確率値をスコアとしてそのまま用いることはできない。そこで、予測時の状況を入力としてその予測全体が正しいかどうかを確信度として推定し、その確信度と確率値の積を各目的地候補グリッドのスコアとする。

本研究ではランダムフォレストで予測の確信度を推定する。用いる特徴量は以下とした。

(1) 学習した移動軌跡の数：この値が小さい場合、目的地予測に失敗している可能性が高いと考えられる。

(2) 過去に訪れたことのあるグリッドまでの距離の平均：予測時に、LSTM にクエリ移動軌跡をグリッドの系列として入力する際に、学習データ中にないグリッドは学習データに出現したグリッドの中でもっとも近くにあるグリッドに置き換えられる。その際に本当に通過していたグリッドと置き換えられたグリッドの距離の平均を特徴量として用いる。この値が大きい場合、クエリ移動軌跡の近くを通った移動軌跡が学習データ中に存在しないことになるため、目的地予測に失敗している可能性が高いと考えられる。

(3) クエリ移動軌跡の最後のグリッドの生成確率：LSTM にクエリ移動軌跡中のグリッドを順に入力していくと、入力毎に次のグリッドがどこになるかの確率分布が得られる。クエリ移動軌跡を入力していき、クエリ移動軌跡中の最後のグリッドの時刻 t にどこにいるかに関する確率分布において、実際にいるグリッドの確率の大きさを特徴量として用いる。この値が大きい場合、クエリ移動軌跡と類似した移動軌跡が学習データ中に含まれていた可能性が高いため、目的地予測にも成功している可能性が高いと考えられる。

(4) $t+1, t+2, t+3$ のグリッド予測の確率分布のエントロピー：時刻 1 から t までのクエリ移動軌跡が与えられたときの、時刻 $t+1, t+2, t+3$ の各時刻のグリッドに対する予測の確率分布のエントロピーを用いる。これは、正しい予測が行われている際にはグリッドの予測の精度が高くエントロピーが下がるが、そうではない場合にはエントロピーが上がることを用いるためである。各サンプリング時にそれぞれの時刻に対するエントロピーを計算し、最終的に全サンプリング時のエントロピーの平均を特徴量として用いる。この値が大きい場合、各時刻に対する予測のばらつきが大きいということであり、目的地予測に失敗している可能性が高いと考えられる。

(5) 出力された予測移動軌跡の生成尤度：各サンプリング時に目的地候補までの系列が併せて得られるが、これの生成確率である。この値が大きい場合、目的地予測にも成功している可能性が高いと考えられる。

予測結果の Top-5 に正解の目的地グリッドが含まれているものを真、含まれていないものを偽とした二値分類をランダムフォ

レストで学習し、推定時に真に投票した決定木の割合を推定の確信度として用いる。

モデル中でのグリッドの扱いについて述べる。一般的に移動傾向データはプライバシーの問題などから、グリッドサイズが最低でも 500m 四方程度と大きい。一方で、個人モデルで移動軌跡をグリッドの系列に変換する際に大きなグリッドに変換してしまうと、移動に関する情報が損なわれてしまい、精度が下がってしまう。しかしアンサンブルのためには大域モデルと個人モデルで出力するグリッドのサイズは同一にする必要がある。そこで本研究では、個人モデルの予測途中のグリッドサイズは先行研究のままとし、目的地候補グリッドのみを大域モデルに揃えることとする。

3.2.2 アンサンブルモデル

アンサンブルの方法について述べる。大域モデルと個人モデルの予測結果を組み合わせ、最終的な目的地候補のランキングを得ることが目的である。それぞれのモデルは各モデルにおける各グリッドが目的地になる確率と、出力全体の確信度スコアを出力する。本研究ではこれらの線形結合によって最終的な予測結果を得る。つまり、各モデルに対する重みを設定し、それぞれのモデルの出力した目的地になる確率に重みをかけてから足し合わせたものを最終的にそのグリッドが目的地になる確率とみなす。重みパラメータの最適化については、本研究では最適化すべきパラメータが二つと少ないため、グリッドサーチで行う。

アンサンブル方式の定式化について述べる。まず、それぞれのモデルの出力について述べる。大域モデルと個人モデルはそれぞれ、図 1 の左側に示すように目的地候補のグリッドとそれらのスコアを出力する大域モデル g と個人モデル p がそれぞれ出力した、グリッド i が目的地になる確率を $P_g(i), P_p(i)$ と示す。また個人モデルの確信度を C_p と示す。この時、個人モデルでのグリッド i に対するスコアは $C_p P_p(i)$ となる。このとき、あるグリッド i に対するスコア $S(i)$ を以下のように出力する。

$$S(i) = w_g P_g(i) + w_p C_p P_p(i) \quad (2)$$

ここで、 w_g と w_p はそれぞれのモデルの重みである。このスコアが高い順にグリッドを並べたものを、最終的な予測結果として出力する。

4. 実 験

提案手法の有効性を検証するため、各種モデルの予測精度について検証する。

4.1 用いたデータと実験設定

実験で用いた個人移動軌跡データと大衆の移動傾向データについて述べる。

個人の移動軌跡データについて述べる。30 人のユーザの実際の移動軌跡をタブレットの GPS および Wi-Fi で取得した。記録時には 3 秒間隔で測位を行ったが、実験時には 5 分間隔にリサンプリングしたものを用いた。これは、実際の位置情報はバッテリーの消費量などの都合で 5 分間隔程度での取得が多いからである。詳細を表 2 に示す。

ユーザ数	30人
期間	3週間
測位デバイス	Nexus7 ver.2012のGPSおよびWi-Fi
測位頻度	3秒間隔で取得して5分間隔にリサンプリング
居住地域	関東地域

表2 実験に用いた個人移動軌跡データの詳細

範囲	東京都と神奈川県
グリッドサイズ	500m 四方
日付	2015/04/01(平日), 2015/04/05(休日)

表3 移動傾向データの算出に用いた人口分布データ

次に移動傾向データについて述べる。移動傾向データは限られた地域や時間帯については取得可能であるが、2章に示したように日本全域について取得することは現状では難しい。そこで本稿では、人口分布データを基に、先行研究の手法[9]を用いて移動傾向データを取得した。人口分布データとは時間帯毎の各エリアの人数データであり、先行研究の手法は人口分布データを入力として各時間帯の各エリア間の移動人数を教師なし学習で推定する手法である。本研究では人口分布データとして、NTTドコモ社の提供している、携帯端末の通信情報を基に取得されたデータであるモバイル空間統計データを用いた。表3に用いた人口分布データの詳細と、移動傾向データ推定時のハイパーパラメータを示す。

次に、アンサンブルモデルの実験の手順について述べる。アンサンブルモデルに関わる求めるべきパラメータは、個人モデルと大域モデルのパラメータと、アンサンブル時のモデルごとの重みパラメータがある。個人モデルのパラメータには既存の予測手法[1]のパラメータと、予測確信度の算出を行うランダムフォレストのパラメータがある。これらの学習と手法の評価を、以下の手順を5回繰り返すことでの五交差検定で行う。

(1) ユーザを均等に5グループに分割し、そのうちの1グループを試験用データ、1グループをモデル重み学習用データ、残りを各モデル学習用データとする。今回のデータではユーザ数は30人であるため、一つの試験用データには6ユーザの移動軌跡データが含まれる。

(2) 各モデル学習用データに含まれる各ユーザの移動軌跡について、個人モデルのうちのLSTM部分だけを使って、古いものから一つ学習して一つ推定することを繰り返す。推定する際に、3.2.1に示した特徴量と推定結果を記録する。次に、推定結果のTop-5に正解の目的地が含まれている場合を推定成功、含まれていない場合を推定失敗として、得られた特徴量から推定の成否を予測するようなモデルを学習し、これを信頼度推定モデルとする。また、各モデル学習用データを用いて大域モデルの学習も行う。

(3) モデル重み学習用データに含まれる各ユーザについて、まず各ユーザのデータのうちの前半部分で個人モデルの学習を行う。この際に信頼度推定を用いた予測スコアの出力も行う。次に、各ユーザデータの後半部分を用いて個人モデルと大域モデルでの予測を行い、その結果を記録する。その際に、入力する移動軌跡の長さを出発から到着までのうちの25%、50%、

モデル種別	P@1	P@2	P@5
移動傾向データ	0.3220	0.3416	0.3881
距離考慮	0.3220	0.3529	0.4197
方向考慮	0.3220	0.3500	0.3976
方向・距離考慮	0.3220	0.3599	0.4226

表4 大域モデルの精度評価。五交差検定における、top-kでのPrecision評価を示している。P@1(Precision@1)はベースライン手法でも提案手法でも変わらない。これは、移動傾向データに基づく予測を行う場合、ユーザの現在地点のグリッドが目的地であると予測するのが最も高い精度になるためである。P@2とP@5においては提案手法がベースライン手法よりも高い精度を示した。

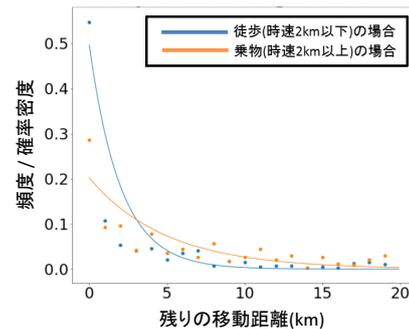


図2 速度と経過時間に基づく残りの移動距離のフィッティング結果の一部。経過時間が15分の場合のグラフを示している。

75%として3回予測を行う。これを、学習に使うデータ量についてユーザのデータ全体のうち10%から50%まで10%ずつ変えながら繰り返す。これで得られた各モデルの推定結果を用いてモデルの重みを学習する。モデルの重みは、0:1から1:0までの0.01刻みでのグリッドサーチによって学習する。

(4) 試験用データに含まれる各ユーザについて、そのユーザのデータのうちの前半部分で学習を行い、後半部分でアンサンブルでの推定を行い、精度評価する。その際に、入力する移動軌跡の長さを出発から到着までのうちの25%、50%、75%として3回予測を行い、それらの精度の平均を最終的な精度とする。学習時に使うデータ量について、ユーザのデータ全体のうち10%から50%まで10%ずつ変えながら実験を繰り返す。

4.2 結果と考察

4.2.1 大域予測モデルの評価

3.1節で述べた、群衆の移動傾向に基づく大域予測モデルの精度評価について述べる。表4に、各手法の精度を示す。移動傾向データは、移動傾向データにおける移動人数の割合をそのまま移動確率とみなしたものであり、ベースライン手法に相当する。方向考慮と距離考慮は3.1節で述べた移動方向と残りの移動距離についての考慮をそれぞれ推定に加えたものであり、方向・距離考慮はその双方を考慮したものであり提案手法に相当する。予測精度はTop-1, Top-2, Top-5でのPrecision評価を用いた。Precision@1の評価では、ベースライン手法と提案手法の精度は完全に同一であった。これは、移動傾向データに基づく予測を行う場合、ユーザの現在地点のグリッドが

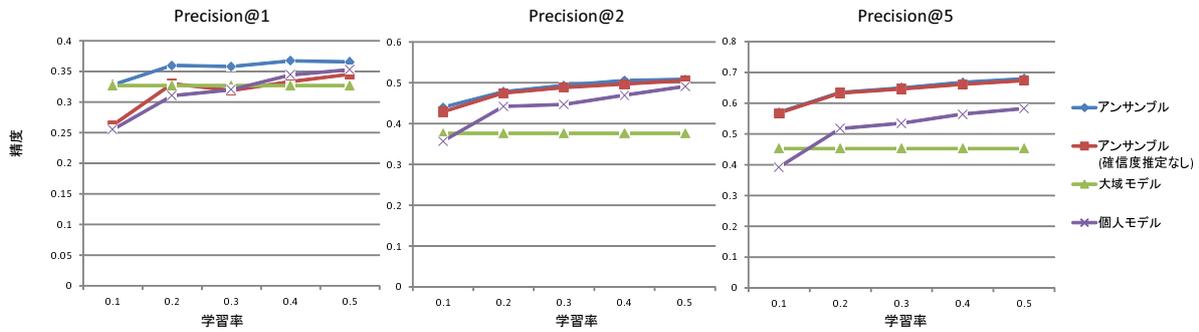


図 3 各手法の予測精度．左側のグラフは Precision@1 を，右側のグラフは Precision@5 を示している．横軸は学習率を示し，これは各ユーザの全データのうち学習データとしたデータ量の割合を示している．例えば学習率が 0.1 の場合は各ユーザの最初の 10%の移動軌跡で個人モデルの学習を行い，後半 50%のデータで精度検証している．いずれの評価結果においても個人モデルよりも大域モデルやアンサンブルモデルの方が精度が高いことを示している．また Precision@1 では個人モデルにおける確信度推定が精度向上に寄与していることも示されている．

目的地であると予測するのが最も高い精度になるためである． Precision@2 と Precision@5 においては，提案手法である方向・距離考慮がもっとも高い精度を示した．方向考慮と距離考慮はそれぞれベースライン手法よりも高い精度を示した．これは，移動方向と移動速度の考慮がそれぞれ予測精度を向上させており，同時に考慮することで最も精度が高くなることを示している．

次に，残りの移動距離の予測がどのように学習されているかについて述べる．図 2 に，経過時間が 15 分の場合の残りの移動距離を時速 2km 以下と以上の二つに分割して指数分布でフィッティングした結果を示す．点は残りの移動距離を 1km 毎に区切った場合の各残り移動時間の頻度を表し，線はそれをフィッティングした指数分布を示す．現在の移動速度に応じて残りの移動距離の分布が異なっていることと，頻度の点が指数分布に近い形で分布しており適切なフィッティングが出来ていることが示されている．

4.2.2 アンサンブルモデルの評価

アンサンブルモデルと，個々のモデルの精度評価の結果について，図 3 に示す．確信度推定なしとなっているアンサンブルモデルは，3.2 で示した確信度推定を省いたものである．横軸は学習率を示しており，これは各ユーザの移動軌跡データのうち何割で学習させたかを表している．各ユーザのデータは 3 週間分のデータであるため，例えば 0.1 は平均して約 2 日分の移動軌跡を学習させて，後半約 10 日間の移動軌跡で予測精度を検証した際の精度を示している．縦軸は予測精度で，左側の図は Precision@1 を，真ん中の図は Precision@2 を，右側の図は Precision@5 を示している．いずれの評価においても，提案手法であるアンサンブルモデルが他の手法特に，学習率が低い際に個人モデルとの差が大きい．このことは，提案手法が学習データが少ないユーザの目的地を予測する際に有効であることを示している．また，確信度推定は特に Precision@1 において精度向上に効果があることが分かる．

次に，未訪問地への予測精度を示す．精度検証の際にテスト

モデル種別	P@1	P@2	P@5
アンサンブルモデル	0.2622	0.2744	0.3164
大域モデル	0.2937	0.3409	0.4108

表 5 未訪問地への移動に限定して評価した予測精度．個人モデルによる予測の精度はいずれも 0 になるため示していない．提案手法では未訪問地への移動であっても一定の精度で予測が可能であることを示している．

データとした移動軌跡量は 30 ユーザあわせて 809 個あるが，そのうち未訪問地への移動は 272 個であった．それら未訪問地への移動のみをテストデータとしたときの予測精度を表 5 に示す．学習率は 0.5 としている．個人モデルの結果はいずれも 0 になるため示していない．いずれも個人モデルが全く予測できないことに比較すれば一定の精度での予測を実現しているが，全データでの検証と比較すると精度が下がっており，未訪問地への移動を予測することの難しさを示している．また図 3 が示す通り全体に対する予測結果はアンサンブルモデルの方が高いが，未訪問地に限るとアンサンブルモデルは大域モデルよりも精度が低い．これは，未訪問地に対する予測では個人モデルが正解することがないが，アンサンブル時に個人モデルを考慮してしまうために起きる問題である．

5. おわりに

本研究では過去の履歴がないユーザの移動や，過去に訪問したことがない場所への移動においても目的地予測が可能な手法の実現に取り組んだ．まず群衆の移動傾向データを利用した目的地予測手法を提案し，ユーザの移動速度と経過時間を考慮することでシンプルなベースライン手法よりも高い精度で予測できる手法の提案を行った．次に，先行研究の個人履歴に基づく予測についてその確信度を出力するように拡張し，移動傾向データに基づく予測モデルとのアンサンブルの提案を行った．首都圏に住むユーザ 30 人の 3 週間分の移動軌跡データと，首都圏の人口分布データから算出した群衆の移動傾向データを用

いて実験を行い、提案手法は従来手法よりも高い精度で目的地を予測できることを確かめた。特に従来手法で対応できなかった学習データが少ない場合や、未訪問地への移動の場合でも一定の精度で予測することが可能であることが確かめられた。また、二つのモデルのアンサンブルによって単独のモデルと比較して高い精度で目的地予測が可能になることも確かめられた。本研究で提案された予測手法は、モバイル系端末におけるユーザへの情報配信の高度化において有効であり、特にパーソナライズにおいて問題になることの多い履歴の少ないユーザに対するコールドスタート問題を解決することが出来ると考えられる。

今後の課題について述べる。大域予測モデルに関して、残りの移動距離を推定する際に、本研究では移動速度を条件として用いたが、移動手段を考慮することが高精度化につながる可能性がある。関連研究としてより高度な移動手段推定を行う研究[14]もあるため、これを用いて手法を拡張することが考えられる。モデルアンサンブルに関しては、現在は二つのモデルのアンサンブルだが利用可能な目的地予測モデルは本研究で扱ったもの以外にも存在するため、それらもさらに組み合わせるアンサンブルすることが考えられる。本研究ではモデルの重みはグリッドサーチを行ったが、仮にモデル数の増加により最適化すべきパラメータが多くなった場合には、場合には、ランク学習[12]における Pair-wise な最適化手法[15]などが適用可能である。

文 献

- [1] Yuki Endo, Kyosuke Nishida, Hiroyuki Toda, and Hiroshi Sawada. Predicting destinations from partial trajectories using recurrent neural network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 160–172. Springer, 2017.
- [2] Natalia Marmasse and Chris Schmandt. A user-centered location model. *Personal and ubiquitous computing*, Vol. 6, No. 5-6, pp. 318–321, 2002.
- [3] Andy Yuan Xue, Jianzhong Qi, Xing Xie, Rui Zhang, Jin Huang, and Yuan Li. Solving the data sparsity problem in destination prediction. *The VLDB Journal The International Journal on Very Large Data Bases*, Vol. 24, No. 2, pp. 219–243, 2015.
- [4] 瀧本祥章, 西田京介, 遠藤結城, 戸田浩之, 澤田宏, 石川佳治. 時間帯を考慮したパーソナライズ目的地予測. *電子情報通信学会論文誌 D*, Vol. 100, No. 4, pp. 472–484, 2017.
- [5] パーソントリップ調査. <https://www.tokyo-pt.jp/person/> (2019/01/09 確認).
- [6] モバイル空間統計 動態統計. https://www.mobaku.jp/service/jpn_flow/ (2019/01/09 確認).
- [7] モバイル空間統計 分布統計. https://www.mobaku.jp/service/jpn_distribution/ (2019/01/09 確認).
- [8] Tomoharu Iwata, Hitoshi Shimizu, Futoshi Naya, and Naonori Ueda. Estimating people flow from spatiotemporal population data via collective graphical mixture models. *ACM Trans. Spatial Algorithms Syst.*, Vol. 3, No. 1, pp. 2:1–2:18, 2017.
- [9] Yasunori Akagi, Takuya Nishimura, Takeshi Kurashima, and Hiroyuki Toda. A fast and accurate method for estimating people flow from spatiotemporal population data. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pp. 3293–3300. International Joint Conferences on Artificial Intelli-

- gence Organization, 7 2018.
- [10] 赤木康紀, 西村拓哉, 倉島健, 戸田浩之. 人の移動の特徴を考慮した人口統計情報からの移動人数推定. 第 10 回データ工学と情報マネジメントに関するフォーラム, 2018.
- [11] David Opitz and Richard Maclin. Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, Vol. 11, pp. 169–198, 1999.
- [12] Tie-Yan Liu, et al. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*, Vol. 3, No. 3, pp. 225–331, 2009.
- [13] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [14] 金村卓郎, 西村拓哉, 倉島健, 戸田浩之. 疎な移動軌跡を対象とした移動手段推定. 第 10 回データ工学と情報マネジメントに関するフォーラム, 2018.
- [15] Thorsten Joachims. Optimizing search engines using click-through data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 133–142. ACM, 2002.