

2017 年度 修士論文

MAP 推定を用いた歩行者の位置情報予測

－POI 推薦を目的として－

提出日： 2018 年 1 月 30 日

指導： 山名 早人 教授

早稲田大学 基幹理工学研究科 情報理工・情報通信専攻

学籍番号：5116F059

田中 博己

概 要

スマートフォンに代表される GPS 搭載機器が広く普及したことに伴い、位置情報サービス (LBS) の重要性はますます向上している。位置情報を予測することによって、観光地の推薦や目的地をベースとした広告の推薦、自動的に目的地を設定するナビゲーションシステム提供が可能になる。従来の研究では、位置情報の予測手法として主に隠れマルコフモデルが用いられている。隠れマルコフモデルの構築には道路ネットワークの構築が必要であるため、道路上を走行することがほぼ保証されている自動車に対する位置情報の予測に限られている。そのため歩行者に対しては、必ずしも道路上を通らない場合があるため、道路ネットワークを利用した隠れマルコフモデルの適用ができない。これに対して本研究では、道路ネットワークを用いず、100m 四方のメッシュに分割し、MAP 推定を用いて予測をし、10 分後の精度が 3.23% となった。

目次

第1章	はじめに.....	1
第2章	用語定義.....	3
2.1	Trajectory	3
2.2	Stay Point.....	3
2.3	過去の Trajectory, 未来の Trajectory	4
2.4	他人の Trajectory	4
第3章	推定に用いるアルゴリズム	5
3.1	最尤推定.....	5
3.2	MAP 推定	5
3.3	EM アルゴリズム	6
3.4	マルコフ連鎖.....	7
第4章	関連研究.....	8
4.1	概要	8
4.2	頻出パターンマイニングを利用した手法.....	8
4.3	Trajectory モデルを構築した手法.....	8
4.3.1	他人の Trajectory を利用した手法	9
4.3.2	個人の移動履歴を利用した手法.....	9
4.3.3	Trajectory モデルを構築した手法のまとめ	10
4.4	深層学習を利用した手法	10
4.5	関連研究のまとめ	10
第5章	メッシュを用いた歩行者の位置情報予測手法の提案	12
5.1	概要	12
5.2	メッシュを用いた Trajectory の表現.....	12
5.3	移動時間分布.....	13
5.4	メッシュ間遷移確率.....	14
5.5	Trajectory の尤度	14
5.6	EM アルゴリズムによるパラメータ推定.....	15
5.6.1	最尤推定	15
5.6.2	MAP 推定	17
5.7	移動予測.....	20
5.7.1	移動モード推定	21
5.7.2	移動先メッシュの予測	21
第6章	評価実験.....	22
6.1	データセット.....	22

6.2	前処理.....	23
6.2.1	位置情報の整形.....	23
6.2.2	Trajectory の分割.....	23
6.2.3	移動モードの推定.....	23
6.3	評価手法.....	24
6.4	実験結果.....	24
6.4.1	EM アルゴリズムを用いた MAP 推定によるパラメータ推定.....	24
6.4.2	メッシュの移動予測.....	25
6.5	評価実験のまとめ.....	26
第 7 章	おわりに.....	28

第1章 はじめに

スマートフォンを始めとした Global Positioning System(GPS)を搭載した機器の普及に伴い、大量の位置情報の取得が容易になった。位置情報サービス (LBS) とは、取得した位置情報を利用して提供するサービスのことをいう。LBS を用いたアプリケーションには、カーナビゲーションシステムを代表として、様々なアプリケーションが存在し、日々の生活に浸透している。

また、位置情報を用いた研究も盛んに行われている。Zheng[1]によると、位置情報を用いたデータマイニングの枠組みは図 1.1 のように示される。図 1.1 によると、Trajectory data mining の構成は、Preprocessing, Indexing and Retrieval, Uncertainty, Pattern Mining, Classification, Outlier Detection に分類できる。

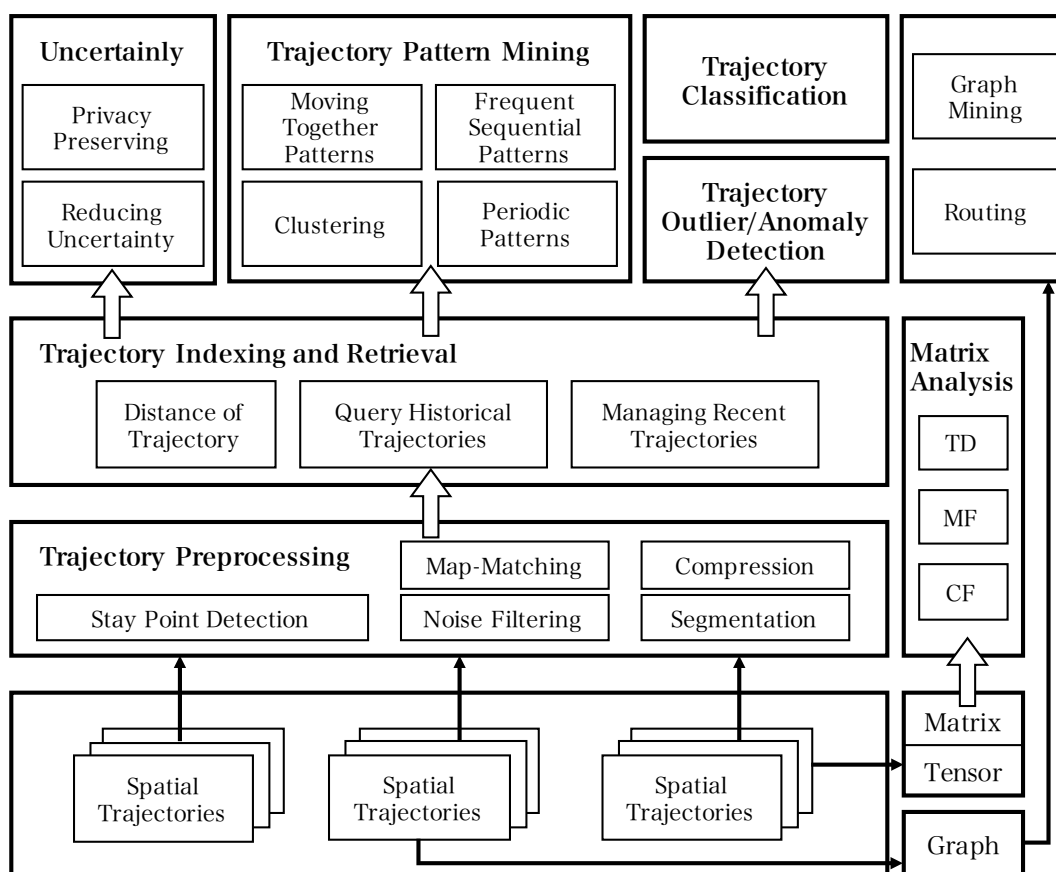


図 1.1 Trajectory data mining の枠組み ([1] Fig. 1. よりトレース)

とりわけ近年では、多くの位置情報を用いたアプリケーションの要素技術として、位置情報予測の研究が必要とされている。位置情報予測の研究は図 1.1 の Trajectory Pattern Mining の立ち位置であるといえる。例えば、観光地の推薦や目的地をベースとした広告の配信、自動的に目的地を設定するナビゲーションシステムがある[8]。位置情報の予測によ

り，LBS の提供者は現在位置から目的地や未来の位置に基づく広告の配信が可能になり，位置情報ベースの広告配信の効果がより向上すると考えられる。

従来，位置情報の予測に関する研究は，隠れマルコフモデルを利用したものが複数行われている[14][15]。隠れマルコフモデルを用いた予測は道路ネットワークを構築して行う。そのため，位置情報の予測対象が常に道路上に存在することを前提とする。したがって，位置情報の予測対象は自動車に限定され，歩行者に対する位置情報の予測に適用することができない。これは，歩行者は道路上の他に公園や広場，大学などを移動する可能性があるためである。

本研究では，従来手法が対象としなかった歩行者に対する現在地から 10 分以内の位置情報の予測を行うことを目的とする。具体的には，既存研究で多く用いられている道路ネットワークを用いずに予測を行う。はじめに，近似した位置を同一として扱うために予測対象地域を 100m 四方のメッシュに分割する。次に，MAP 推定を行い位置情報の予測を行う。

本稿では以下の構成を取る。第 2 章で本研究において用いる用語定義を行い，第 3 章で本研究に用いる推定手法の数学的説明を行う。第 4 章で関連研究を説明し，位置情報の予測の研究に対する本研究の立ち位置を明確にする。第 5 章で提案手法の説明を行い，第 6 章で評価実験を説明する。第 7 章で本研究をまとめる。

第2章 用語定義

2.1 Trajectory

Zheng ら [2]によると, Trajectory の定義は地図上のスペースを移動することによって生成される軌跡と定義される. 時系列順に並べられて表記される. $P = p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$ という一連のデータは, $p_i = (x, y, t) (i = 1, 2, \dots, n)$ と表され, (x, y) は, 緯度と経度の組を示し, t は時刻を示す. また, p_i の緯度 $p_i.x$, p_i の経度 $p_i.y$, p_i の時刻を $p_i.t$ と表す. このとき, $\forall i, 0 \leq i \leq k$ に対して, $p_{i+1}.t > p_i.t$ である.

本研究における Trajectory の定義とは, 2.2 節で説明する, ある Stay Point から, 次の Stay Point までの一連の軌跡と定義する.

2.2 Stay Point

Zheng ら [2]によると, Stay point s とは, 一定の距離 D_r 以内に一定時間 T_r 以上滞在した場所を指す. s の定義は p_i と同様に, $s = (x, y, t) (i = 1, 2, \dots, n)$ と表され, (x, y) は, 緯度と経度の組を示し, t は時刻を示す. また, s の緯度 $s.x$, s の経度 $s.y$, s の時刻を $s.t$ と表す. Stay Point を図示したものを図 2.1 に示す. Stay point s は以下のように定義される. はじめに, $Dist(p_m, p_n)$ は p_m と p_n 間の距離, $Int(p_m, p_n)$ は p_m と p_n 間の経過時間を示す.

一連の Trajectory $P = \langle p_j, p_{j+1}, \dots, p_j \rangle$ が, $\forall i, j \leq i \leq k, Dist(p_j, p_k) \leq D_r, Int(p_j, p_k) \leq T_r$ を満たしたとき, s は次のように求められる.

$$s.x = \sum_{i=j}^k p_i.x / |P| \quad (1)$$

$$s.y = \sum_{i=j}^k p_i.y / |P| \quad (2)$$

$$s.t_a = p_j.t \quad (3)$$

$$s.t_l = p_k.t \quad (4)$$

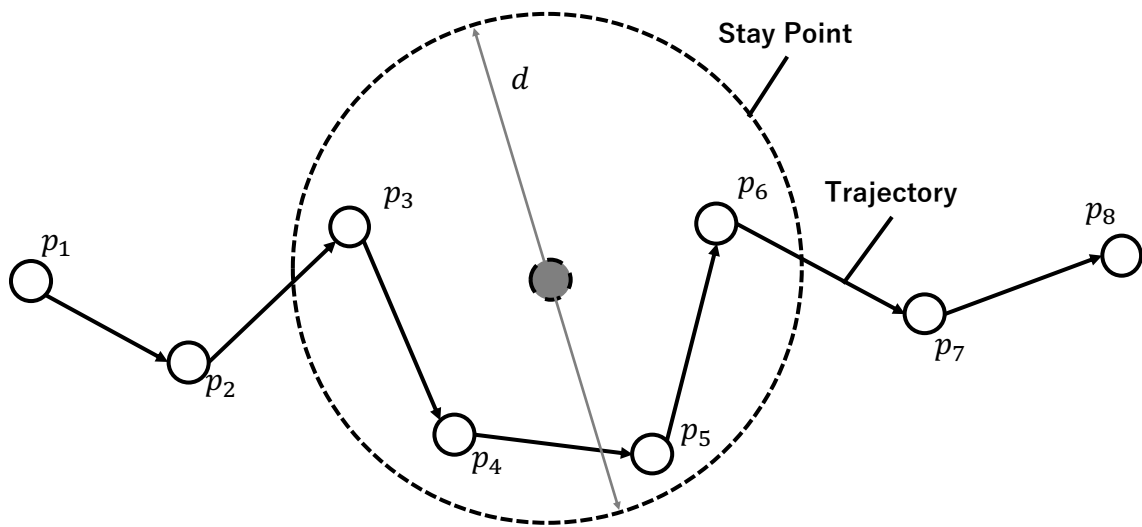


図 2.1 Trajectory と Stay Point ([2]Fig. 3.よりトレース)

2.3 過去の Trajectory, 未来の Trajectory

自分の過去の Trajectory とは, Trajectory $p = \langle p_1, \dots, p_{k-1}, p_k, p_{k+1}, \dots, p_n \rangle$ とある観測点 $p_k \in p$ が与えられたとき, 過去の Trajectory $p_{\text{past}} = \langle p_1, \dots, p_{k-1} \rangle$, 未来の Trajectory $p_{\text{future}} = \langle p_{k+1}, \dots, p_n \rangle$ と定義する.

2.4 他人の Trajectory

Trajectory の組 $PD = \{p_1, \dots, p_n\}$ と, ある Trajectory p_k が与えられたとき, 他人の Trajectory を $PD_{\text{others}} = \{p_1, \dots, p_{k-1}, p_{k+1}, \dots, p_n\}$ と定義する.

第3章 推定に用いるアルゴリズム

本章では、位置情報予測に用いる各種アルゴリズムの定義及び数学的説明を行う。

3.1 最尤推定

観測データ \mathbf{x} の確率分布がパラメータ $\boldsymbol{\theta}$ と決められているとき、 \mathbf{x} の確率分布は $p(\mathbf{x}; \boldsymbol{\theta})$ と表せる。ここで、 $p(\mathbf{x}; \boldsymbol{\theta})$ に従う独立同一分布な観測データ集合 $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ が観測できたと仮定したときこの観測データ集合、 \mathcal{D} において

$$L(\boldsymbol{\theta}, \mathcal{D}) = \prod_{n=1}^N p(\mathbf{x}_n; \boldsymbol{\theta}) \quad (5)$$

を尤度と呼ぶ。対象とする観測データの定義が明白である場合は $\boldsymbol{\theta}$ の関数とみなして $L(\boldsymbol{\theta})$ と書く。尤度を最大化するパラメータ $\boldsymbol{\theta}$ の値 $\hat{\boldsymbol{\theta}}_{ML}$ を求めることは最尤推定と呼ばれ、以下の式(6)によって定義される。

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} L(\boldsymbol{\theta}) \quad (6)$$

3.2 MAP 推定

パラメータ $\boldsymbol{\theta}$ の事前分布 $p(\boldsymbol{\theta})$ が与えられているとき、独立同一分布な観測データ集合 $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ を観測したという条件下でパラメータ $\boldsymbol{\theta}$ の事後確率はとなる。この事後確率 $p(\boldsymbol{\theta}|\mathcal{D})$ を最大化する $\boldsymbol{\theta}$ 、すなわち $\hat{\boldsymbol{\theta}}_{MAP}$ を求めるパラメータ推定法を MAP 推定と呼ぶ。

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{ML} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} p(\boldsymbol{\theta}|\mathcal{D}) \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \frac{p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \frac{p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathcal{D})} \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \end{aligned} \quad (7)$$

3.3 EM アルゴリズム

EM アルゴリズム[20]とは、以下の式で表される混合ガウス分布を代表とした、潜在変数があるモデルのパラメータ推定方法として広く知られている。

$$p(x) = \sum_{i=1}^M \pi_i p_i(x) \quad (8)$$

$$p_i(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right) \quad (9)$$

$$\sum_{i=1}^M \pi_i = 1, \quad \pi_i \geq 0 \quad (i = 1, \dots, M) \quad (10)$$

以下のように文字を定義した時、EM アルゴリズムは以下のように定義できる。

- 確率変数 x に対する観測データ集合 $D = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ と定義する。 \mathbf{x}_i は K 次元ベクトルとする。
- 潜在変数は z とし、 \mathbf{x}_n に対応する潜在変数の値は z_n とする。
- 確率分布を決めるパラメータの集合をベクトル θ とする。

θ を D から推測することがEM アルゴリズムの目的である。

Algorithm: EM アルゴリズム

初期化	$\theta^{(1)}$ の初期値を適当に決める
更新の繰り返し	以下の E step, M step を θ が収束するまで繰り返す
E step	$p(z \mathbf{x}, \theta^{(t)})$ を計算する
M step	$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} Q(\theta^{(t)} \theta^{(t+1)})$, $\theta^{(t)}$ を $\theta^{(t+1)}$ に置き換える $t \leftarrow t + 1$

EM アルゴリズム内の Q 関数とは、以下の式(11)で表される事が多い。

$$Q(\theta|\theta^{(t)}) = \sum_{z \in Z} p(z|\mathbf{x}, \theta^{(t)}) \log p(\mathbf{x}, z|\theta) \quad (11)$$

3.4 マルコフ連鎖

マルコフ過程とは、次に発生する事象の確率が、過去の状態と無関係に、現在の状態のみで決定される性質をもつ確率過程のことである。マルコフ連鎖とは、マルコフ過程のうち、取りうる状態が離散的なものを指す。

本節では、一連の確率変数を $\mathbf{X} = \langle X_0, X_1, \dots, X_n \rangle$ とし、時点 i の状態を $x_i, \forall i, 0 \leq i \leq n$ とする。現在の状態が決定されていれば過去及び未来の状態は独立である。形式的には式(12)という形で示される。

$$\begin{aligned} P(X_{n+1} = x | X_n = x_n, \dots, X_1 = x_1, X_0 = x_0) \\ = P(X_{n+1} = x | X_n = x_n) \end{aligned} \tag{12}$$

式(12)を、次で説明する N 重マルコフ連鎖と区別する際は単純マルコフ連鎖と呼ぶ。

次の状態が現在を含めた過去 N 個の状態履歴に依存して決定される確率過程を N 重マルコフ連鎖と呼ぶ。式(13)で形式的に示される。

$$\begin{aligned} P(X_{n+1} = x | X_n = x_n, \dots, X_1 = x_1, X_0 = x_0) \\ = P(X_{n+1} = x | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_{n-N+1} = x_{n-N+1}) \end{aligned} \tag{13}$$

第4章 関連研究

4.1 概要

位置情報予測に関する研究は、2012年以前に行われていた trajectory の頻出パターンマイニング、2010年以降に行われている trajectory モデルの構築をした手法、2016年以降に研究されている深層学習を用いたものの3種類に分類できる。

本章では、4.2節で trajectory の頻出パターンマイニングに関して説明を行い、4.3節で trajectory モデルの構築を行ったもの、4.4節で深層学習を用いたものにおける説明を行い、4.5節でまとめる。

4.2 頻出パターンマイニングを利用した手法

Trajectory の頻出パターンマイニングを利用した手法は、主に決定木を構築し、位置情報の予測している。ピサ大学の Monreale ら[9]は、2009年に目的地をベースに決定木の構築を行い、その後国立成功大学の Ying ら[10]が2011年に目的地だけではなく、個人の移動履歴を利用した決定木の構築を行っている。

Trajectory の頻出パターンマイニングを用いた手法では決定木構築の計算コストが高くなるという問題点がある。また、決定木を構築する手法であるため、予測の対象が目的地に限られる問題がある。そのため、2012年以降はほぼ行われていない手法である。

4.3 Trajectory モデルを構築した手法

Trajectory モデルの構築を利用した手法は、2010年以降に研究が開始された。Trajectory モデルの構築には、主に隠れマルコフモデルが用いられている。Trajectory モデルの構築した手法に対する予測対象は、ルート予測と目的地予測の2種類である。ルート予測と目的地予測それぞれの予測方法の概略を以下に示す。

ルート予測

1. Trajectory を地図上の道路に結びつける。
2. 道路を Road Segment に分割し、それぞれ番号をつける。
3. Road Segment を元に道路ネットワークを構築する。
4. 道路ネットワークから隠れマルコフモデルを構築する。

目的地予測

1. 目的地の集合を元に目的地ネットワークを構築する.
2. 目的地ネットワークから隠れマルコフモデルを構築する.

Trajectory モデルの構築方法には, 他人の Trajectory を利用した手法と, 個人の移動履歴を利用した手法の 2 種類ある.

4.3.1 他人の Trajectory を利用した手法

近傍の他人のデータを利用した手法について表 4.1 にまとめた. 他人の Trajectory を利用した手法に関しては, 多数の自動車が含まれているデータセットを用いた研究が多い. 自動車の位置情報データセットの場合, T-Drive[6][7]のように, 14,000 台を超えるタクシーの位置情報が取得可能であり, 個人のプライバシーとは無関係な大量の位置情報が収集可能であるためと考えられる. また, 予測対象が目的地のものが多い.

表 4.1 他人の Trajectory を利用した方法

著者	提案年	手法	データセット	予測対象
Gambs ら[11]	2012	MC	Pheonetic, GeoLife, Synthetic	Destination
Asahara ら[12]	2011	Mixed MC	Simulation, Original	Destination
Xue ら[8]	2013	MC	T-Drive	Route
Huang ら[13]	2015	MC	GeoLife	Destination
Ye ら[14]	2016	HMM	Taxis' Data	Route

MC: Markov Chain

HMM: Hidden Markov Model

4.3.2 個人の移動履歴を利用した手法

個人の移動履歴を利用した手法を表 4.2 にまとめた. 個人の移動履歴を利用した手法では, 予測対象の人数が少人数であるものについて, 長期間の位置情報を対象にしている. これは, 個人の移動履歴を用いた手法であるため, 位置情報のデータセットには, 個人のプライバシーと大きく関係がある. そのため, 実験協力者を募って独自にデータセットを作成したものが多く, 実験に利用したデータセットも公開されていないことがある[15]. 本手法を用いた研究の数はそれほど多くないが, 予測対象が目的地予測, ルート予測ともに存在する.

表 4.2 個人の移動履歴を利用した方法

著者	提案年	手法	データセット	予測対象
Alvarez-Garcia ら[15]	2010	HMM	Original (6 people)	Destination
Yuan ら[16]	2013	MC	CD dataset	Route

4.3.3 Trajectory モデルを構築した手法のまとめ

Trajectory モデルを構築した手法は、2010 年から現在に渡って広く行われている手法である。しかし、ルート予測をする場合、道路ネットワークを元に予測をしているため、位置情報のデータが道路上を通ることを前提としている。そのため、どの道路を通るか道路上を通過することがほぼ保証される自動車を対象とした予測に限られてしまう。目的地を予測する場合は道路ネットワークの構築が不必要だが、目的地を予め指定しておく必要がある。したがって、歩行者を対象とした予測には適していないと言える。

4.4 深層学習を利用した手法

深層学習を利用した手法は 2016 年頃から研究され始めている。研究が行われ始めた段階であり、論文の数は少ない。画像を利用し、群衆における数秒後の位置情報の予測を行うもの [17]、全体の混雑予測を行うもの [18] がある。しかしながら、プライバシーの問題やデータ収集の難易度から、データセットが十分に収集できて行きな分野であるため、個別のデータの予測を行うほどのデータセットが揃っていない。そのため、個々の対象者における位置情報の予測が行われていないと考えられる。それでも、2017 年にかけて論文数が増加している手法であるため、今後研究数が増加すると考えられる手法である。

4.5 関連研究のまとめ

位置情報の予測に用いられている手法において代表的なものを時系列順にまとめ、表 4.3 に示す。現時点で位置情報の予測に用いられている手法は、4.3 節で述べたような、隠れマルコフモデルを利用した手法が多く、位置情報の予測に道路情報が必要になる。そのため、歩行者のように必ずしも道路上を移動しない対象に対する位置情報の予測には不適切である。

また、深層学習を利用した予測手法に関しては、目的地やルート予測に必要なほどの位置情報の収集が困難である。そのため、位置情報の予測に関する研究が多く行われていないと考えられる。

表 4.3 既存研究のまとめ

著者	出版年	手法			予測対象
		FPM	ETH	DL	
Monreale ら[9]	2009	✓			People's destination
Ying ら[10]	2011	✓			People's destination
Gambs ら [11]	2012		✓		People's / vehicles' destination
Huang ら[13]	2015		✓		People's / vehicle's destination
Ye ら[14]	2016			✓	Vehicles' route
Zhang ら[18]	2017			✓	Crowd flows

FPM: Frequent Pattern Mining

ETM: Establish Trajectory Methods

DL: Deep Learning

NPD: Nearby Past Data

PTH: Personal Travel History

第5章 メッシュを用いた歩行者の位置情報予測

手法の提案

5.1 概要

本研究の提案手法の概要を述べる。必ずしも道路上を通過するとは限らない歩行者に対する位置情報予測するための手法を提案する。具体的には以下の手順で行う。

1. メッシュを用いた Trajectory の表現
2. 移動時間の分布
3. メッシュ間遷移確率
4. EM アルゴリズムによるパラメータ推定
5. マルコフ連鎖を用いたメッシュ間移動予測

5.2 メッシュを用いた Trajectory の表現

本研究においては、図 5.1 で示すように地図上の平面を格子状に分割する。各格子をメッシュと呼ぶ。位置情報は、数 cm 単位で取得が可能であり、純粹に緯度、経度の値を利用した場合、数 cm 離れた点同士でも異なる位置と認識する。また GPS の取得誤差のため、必ずしも正確に位置情報を取得できるものではない。そのため、100m 四方のメッシュに分割することで、100m 四方の範囲に関して同じ位置として認識させる。100m 四方というメッシュサイズは、POI 推薦という目的において十分に小さいサイズだといえる。また、GPS で取得した位置情報自体も誤差を含んでいるため、極端にメッシュの大きさを小さくした場合、Ground Truth が正しいメッシュに含まれない可能性が増大するためである。

メッシュの集合を G と表す。また、Trajectory に含まれる個々の位置情報を表す点を含むメッシュ $g \in G$ と定義する。

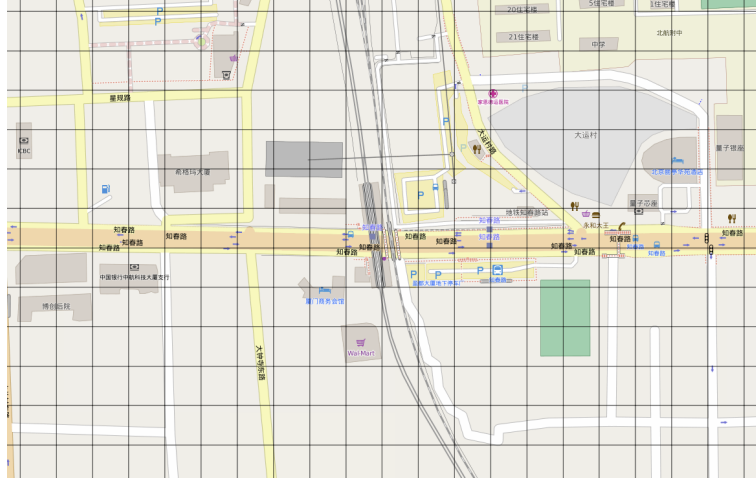


図 5.1 地図をメッシュ分割したイメージ

(map tiles ©OpenStreetMap Fig. 1. contributors, CC BY-SA 2.0).

メッシュを用いて Trajectory をメッシュ位置の列およびメッシュの移動に必要とした時間の列の組 (\mathbf{g}, \mathbf{t}) と表す. ここで, $\mathbf{g} = \langle g_1, g_2, \dots, g_n \rangle$ は Trajectory に含まれる個々の位置情報を表す点に対応するメッシュ位置を表し, $\mathbf{t} = \langle t_1, t_2, \dots, t_n \rangle$ における各時間 t_i はある Trajectory がメッシュ g_i に入ってから g_i を出るまでの時間を示す. 以下では, メッシュ内の移動時間と表現する.

5.3 移動時間分布

メッシュ内の移動時間 t_i は, 移動手段や移動目的に応じて異なる. 本研究においては, 歩行者のみを対象としているため, 目的地に向かって歩いている場合と, 観光をしているような場合とでは移動速度が異なる. そのため, 本研究では歩行者の中でも移動モード m が複数あると仮定する. そこで本研究では, 各メッシュ g を時間 t で移動する確率を式(12)の混合ガウス分布で表す.

$$p(t|g) = \sum_{m \in M} \theta_{gm} \mathcal{N}(t; \mu_{gm}, v_{gm}^2) \quad (14)$$

ここで, M は混合ガウス分布の成分集合を, θ_{gm} は, メッシュ g の成分のうち, 先程述べた移動モード $m \in M$ の混合比を表す. μ_{gm} および v_{gm}^2 は式(13)に示すように, メッシュ g のモード m のガウス分布の平均と分散を表す.

$$\mathcal{N}(t; \mu_{gm}, v_{gm}^2) = \frac{1}{\sqrt{2\pi v_{gm}^2}} \exp\left(-\frac{(t - \mu_{gm})^2}{2v_{gm}^2}\right) \quad (15)$$

5.4 メッシュ間遷移確率

Trajectory があるメッシュ g からモード m で隣接するメッシュへの遷移確率を ϕ_{gmd} と表す。ここで d は隣接するメッシュ 4 方向を表し、 $d = \{\text{north}, \text{east}, \text{south}, \text{west}\}$ のいずれかを表す。

メッシュ間の遷移確率を用いることによって、Trajectory を

$$\mathbf{c}_i = (g_i, t_i, d_i) \quad (16)$$

という 3 項の組で表すこととする。ここで d_i は g_i からの移動方向を表しており、 $\mathbf{d} = \langle d_1, d_2, \dots, d_n \rangle$ は、メッシュ間の遷移方向の列を表している。移動方向 d_i は、 g_i と g_{i+1} より求めることができる。

5.5 Trajectory の尤度

本節では、本研究で用いるパラメータの集合を

$$\Theta = \{\theta_g\}_{g \in G} \cup \{\phi_{gm}\}_{g \in G, m \in M} \cup \{\mu_{gm}\}_{g \in G, m \in M} \cup \{\nu_{gm}\}_{g \in G, m \in M} \quad (17)$$

と表す。Trajectory $\mathbf{x} = \langle c_1, c_2, \dots, c_n \rangle$ について、それぞれのメッシュ g_i の移動モード m_i とし、また、Trajectory \mathbf{x} の移動モード列を $\mathbf{z} = \langle m_1, m_2, \dots, m_n \rangle$ として表す。 \mathbf{z} は直接観測することができないことに注意する必要がある、この Trajectory の移動モード列 \mathbf{z} に対する尤度は以下のように表される。

$$p(\mathbf{x}, \mathbf{z} | \Theta) = \prod_{i=1}^l \mathcal{N}(t_i; \mu_{g_i m_i}, \nu_{g_i m_i}) \theta_{g_i m_i} \phi_{g_i m_i d_i} \quad (18)$$

また、移動モードに関する周辺尤度は以下の式で表される。

$$p(\mathbf{x}, \mathbf{z} | \Theta) = \prod_{i=1}^l \sum_{m \in M} \mathcal{N}(t_i; \mu_{g_i m}, \nu_{g_i m}) \theta_{g_i m} \phi_{g_i m d_i} \quad (19)$$

5.6 EM アルゴリズムによるパラメータ推定

5.6.1 最尤推定

学習データとして Trajectory の集合 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ が与えられたとき, パラメータ $\boldsymbol{\theta}$ の最尤推定

$$\underset{\boldsymbol{\theta}}{\operatorname{argmax}}[\ln p(\mathbf{X}|\boldsymbol{\theta})] \quad (20)$$

を求める.

本モデルの最尤推定値は EM アルゴリズムを用いて求めることが出来る. 学習データに含まれるそれぞれの Trajectory \mathbf{x}_i に対する移動モード列 \mathbf{z}_i と表す. また, 学習データに対する移動モード列を $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ と表すことにする. \mathbf{z}_i は学習データ中の Trajectory に対するモード列を表す. ここで, 訓練データ \mathbf{X} と移動モード列 \mathbf{Z} に対する変分事後分布 $q(\mathbf{Z}|\mathbf{X})$ を導入する.

Jensen の不等式を用いると, 尤度 $p(\mathbf{X}|\boldsymbol{\theta})$ の下界 $L(q(\mathbf{Z}|\mathbf{X}), \boldsymbol{\theta})$ は以下のようなになる.

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) \geq L(q(\mathbf{Z}|\mathbf{X}), \boldsymbol{\theta}) \quad (21)$$

以下では変分事後分布 $q(\mathbf{Z}|\mathbf{X})$ を q と略記することにする.

EM アルゴリズムは, 対数尤度 $\ln p(\mathbf{X}|\boldsymbol{\theta})$ を最大化する代わりに上記の下界 $L(q(\mathbf{Z}|\mathbf{X}), \boldsymbol{\theta})$ を最大化する変分事後分布 q とパラメータ $\boldsymbol{\theta}$ を求める. アルゴリズムは, 尤度 $p(\mathbf{X}|\boldsymbol{\theta})$ が収束するまで q と $\boldsymbol{\theta}$ を交互に更新する. t 回目に求められた変分事後分布とパラメータをそれぞれ q^t および $\boldsymbol{\theta}^t$ とする.

E ステップ

E ステップではパラメータを $\boldsymbol{\theta}^{t-1}$ に固定し, 対数尤度の下界 $L(q, \boldsymbol{\theta}^{t-1})$ を最大化する. これを満たす変分事後分布は以下の式(22)で与えられる.

$$q^t = \underset{q}{\operatorname{argmax}}[L(q, \boldsymbol{\theta}^{t-1})] = p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{t-1}) \quad (22)$$

EM アルゴリズムにおいて $t-1$ 回目に求められたパラメータを $\boldsymbol{\theta}^{t-1}$ で表す. また, t 回目の最適変分事後分布を $q^t(\mathbf{Z}|\mathbf{X})$ と表すことにする. 式(22)より, E ステップでは, Trajectory に含まれる個々の位置情報における点 $\mathbf{c} = (g, t, d)$ に対する移動モードの事後確率は以下の式

(23)で与えられる.

$$\begin{aligned}
q^t(m|\mathbf{c}) &= p(m|\mathbf{c}, \boldsymbol{\theta}^{t-1}) \\
&\propto p(\mathbf{c}, m|\boldsymbol{\theta}^{t-1}) \\
&= \theta_{gm}^{t-1} \phi_{gmd}^{t-1} \mathcal{N}(t; \mu_{gm}^{t-1}, \nu_{gm}^{t-1})
\end{aligned} \tag{23}$$

M ステップ

M ステップでは, 変分事後分布を q^t に固定して, 対数尤度の下界 $L(q, \boldsymbol{\theta}^t)$ を最大化するパラメータ $\boldsymbol{\theta}$ を求める.

$$L(q, \boldsymbol{\theta}^t) = \sum_{\mathbf{Z}} q^t \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) - q^t \ln q^t \tag{24}$$

ここで, 変分事後分布 q^t は固定されているため, 式(24)の第2項は定数となる, したがって, 第1項を最大化するパラメータ $\boldsymbol{\theta}$ を求めれば十分である. 以下では, 式(24)の第1項を $Q(\mathbf{X}, \boldsymbol{\theta})$ と表す.

$$\begin{aligned}
Q(\mathbf{X}, \boldsymbol{\theta}) &= \sum_{x \in \mathbf{X}} \sum_{i=1}^{|\mathbf{x}|} \sum_{m \in M} q^t(m|\mathbf{c}_i) \cdot \\
&\ln[\theta_{gim} \phi_{gimd} \mathcal{N}(t_i; \mu_{gim}, \nu_{gim})]
\end{aligned} \tag{25}$$

加えて, パラメータ $\boldsymbol{\theta}$ を求めるためには, 移動時間とメッシュ間遷移確率の制約を満たす必要がある.

1) 移動時間

各メッシュ $g \in G$ において移動モードの分布 θ_{gm} は以下の式(26)の制約を満たす必要がある.

$$\sum_{m \in M} \theta_{gm} = 1 \tag{26}$$

Lagrange の未定乗数法を用いて式(26)の制約のもと, 式(25)の $Q(\mathbf{X}, \boldsymbol{\theta})$ を最大にする移動モード分布 θ_{gm} を求めると以下の更新式が得られる

$$\theta_{gm}^t \propto \sum_{x \in \mathbf{X}} \sum_{i=1}^{|\mathbf{x}|} q^t(m|\mathbf{c}_i) \cdot \delta(g, g_i) \tag{27}$$

となる。同様に,

$$\mu_{gm}^t = \frac{\sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) \cdot t_i \cdot \delta(g, g_i)}{\sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) \cdot \delta(g, g_i)} \quad (28)$$

となり,

$$v_{gm}^t = \frac{\sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) (t_i - \mu_{gm}^t)^2 \delta(g, g_i)}{\sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) \cdot \delta(g, g_i)} \quad (29)$$

となる。

2) メッシュ間遷移確率

各メッシュ $g \in G$ の各移動モード $m \in M$ において, 方向 $d \in D$ に移動する確率は, 以下の制約を満たす必要がある。

$$\sum_{d \in D} \phi_{gmd} = 1 \quad (30)$$

式(30)の制約のもと式(25)の $Q(X, \theta)$ を最大にする遷移確率は以下のとおり求まる。

$$\phi_{gmd}^t \propto \sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) \cdot \delta(g, g_i) \cdot \delta(d, d_i) \quad (31)$$

5.6.2 MAP 推定

最尤推定を用いると, データ量が少ない場合に推定確率が低下するという問題が発生する。そこでメッシュ間の遷移確率を向上させるために, MAP 推定を利用する。

まず, 各パラメータについて以下の共役事前分布を仮定する。

- 各メッシュ $g \in G$ において $\theta_g \sim \text{Dir}(\alpha)$
- 各メッシュ $g \in G$ と各移動モード $m \in M$ において $\phi_{gm} \sim \text{Dir}(\beta)$
- 各メッシュ $g \in G$ と各移動モード $m \in M$ において $\mu_{gm} \sim \mathcal{N}(\xi, \eta)$

ここで, $\text{Dir}(\alpha), \text{Dir}(\beta)$ は以下の対象ディリクレ分布を表している。

$$\begin{aligned}
Dir(\boldsymbol{\theta}_g; \alpha) &= \frac{\Gamma(\alpha \cdot C_m)}{\Gamma(\alpha)^{C_m}} \prod_{m \in M} \theta_{gm}^{\alpha-1} \\
Dir(\phi_{gm}; \beta) &= \frac{\Gamma(\beta \cdot C_d)}{\Gamma(\beta)^{C_d}} \prod_{d \in D} \phi_{gmd}^{\beta-1}
\end{aligned} \tag{32}$$

式(32)において、 C_m は移動モードの総数を表し、 C_d は移動方向の総数を表す。本研究では4である。

以下では、上記の事前分布のハイパーパラメーターをまとめて、 Λ と表すことにする。

$$\Lambda = \{\alpha, \beta, \xi, \eta\} \tag{33}$$

ここで、事前分布に基づいてパラメータ Θ が求められる確率を以下の式(34)に示す。

$$\begin{aligned}
p(\Theta|\Lambda) &= \prod_{g \in G} [Dir(\boldsymbol{\theta}_g; \alpha)] \cdot \prod_{g \in G, m \in M} [\mathcal{N}(\mu_{gm}; \xi, \eta)] \\
&\quad \cdot \prod_{g \in G, m \in M} [Dir(\phi_{gm}; \beta)]
\end{aligned} \tag{34}$$

次に、学習データとして Trajectory の集合 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ が与えられたとき、パラメータの MAP 推定値 Θ^*

$$\begin{aligned}
\Theta^* &= \underset{\Theta}{\operatorname{argmax}} [\ln p(\mathbf{X}|\Theta) \cdot p(\Theta|\Lambda)] \\
&= \underset{\Theta}{\operatorname{argmax}} [\ln p(\mathbf{X}|\Theta) + \ln p(\Theta|\Lambda)]
\end{aligned} \tag{35}$$

を求める。式(35)の第1項は、5.6.1項で説明した最尤推定の目的関数と同じであり、その下界は、式(20)の $L(q, \Theta)$ となる。そこで、MAP推定の目的関数である式(35)の下界

$$LM(q, \Theta, \Lambda) = L(q, \Theta) + \ln p(\Theta|\Lambda) \tag{36}$$

をEMアルゴリズムを用いて最大化するパラメータを求める。5.6.1項の最尤推定の場合と同じように、アルゴリズムは、尤度 $p(\mathbf{X}|\Theta)$ が収束するまで q と Θ を交互に更新する。 t 回目に求められた変分事後分布とパラメータをそれぞれ q^t および Θ^t と表す。

E ステップ

E ステップではパラメータを Θ^{t-1} に固定し、 $LM(q, \Theta, \Lambda)$ を最大化する。ここで、下界 $LM(q, \Theta, \Lambda)$ を最大とする変分事後分布は、最尤推定と同じく、以下の式(37)で与えられる。

$$q^t = \underset{q}{\operatorname{argmax}}[LM(q, \Theta^{t-1}, \Lambda)] = p(\mathbf{Z}|\mathbf{X}, \Theta^{t-1}) \quad (37)$$

また、Trajectory に含まれる個々の位置情報における点 $\mathbf{c} = (g, t, d)$ に対する移動モードの事後確率も最尤推定と同じく式(38)として与えられる。

$$\begin{aligned} q^t(m|\mathbf{c}) &= p(m|\mathbf{c}, \Theta^{t-1}) \\ &\propto p(\mathbf{c}, m|\Theta^{t-1}) \\ &= \theta_{gm}^{t-1} \phi_{gmd}^{t-1} \mathcal{N}(t; \mu_{gm}^{t-1}, \nu_{gm}^{t-1}) \end{aligned} \quad (38)$$

M ステップ

M ステップでは、変分事後分布を q^t に固定して、対数尤度の下界 $LM(q, \Theta, \Lambda)$ を最大化するパラメータ Θ を求める。

$$LM(q, \Theta, \Lambda) = \sum_{\mathbf{Z}} q^t \ln p(\mathbf{X}, \mathbf{Z}|\Theta) - q^t \ln q^t + \ln p(\Theta|\Lambda) \quad (39)$$

ここで、変分事後分布 q^t は固定されているため、式(39)の第2項は定数となる、したがって、第1項と第3項を最大化するパラメータ Θ を求めれば十分である。また、第1項は最尤推定と同じであるため、式(25)の $Q(\mathbf{X}, \Theta)$ を用いることができる。式(40)では、MAP 推定の M ステップで最大化する関数を表す

$$QM(\mathbf{X}, \Theta, \Lambda) = Q(\mathbf{X}, \Theta) + \ln p(\Theta|\Lambda) \quad (40)$$

加えて、パラメータ Θ を求めるためには、移動時間とメッシュ間遷移確率の制約を満たす必要がある。

1) 移動時間

各メッシュ $g \in G$ において移動モードの分布 θ_{gm} は以下の式(41)の制約を満たす必要がある。

$$\sum_{m \in M} \theta_{gm} = 1 \quad (41)$$

この時、移動モード分布 θ_{gm} の更新式は

$$\theta_{gm}^t \propto \sum_{x \in X} \sum_{i=1}^{|x|} [q^t(m|\mathbf{c}_i) \cdot \delta(g, g_i)] + \alpha - 1 \quad (42)$$

となる。同様に、

$$\mu_{gm}^t \propto \frac{\eta \sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) \cdot t_i \cdot \delta(g, g_i) + \xi v_{gm}^{t-1}}{\eta \sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) \cdot \delta(g, g_i) + v_{gm}^{t-1}} \quad (43)$$

となる。

分散については、事前分布を導入していないため、最尤推定と同じ更新式になる。

$$v_{gm}^t \propto \frac{\sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) (t_i - \mu_{gm}^t)^2 \delta(g, g_i)}{\sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) \cdot \delta(g, g_i)} \quad (44)$$

となる。

2) メッシュ間遷移確率

各メッシュ $g \in G$ の各移動モード $m \in M$ において、方向 $d \in D$ に移動する確率は、以下の制約を満たす必要がある。

$$\sum_{d \in D} \phi_{gmd} = 1 \quad (45)$$

したがって、以下の更新式が求められる。

$$\phi_{gmd}^t \propto \sum_{x \in X} \sum_{i=1}^{|x|} q^t(m|\mathbf{c}_i) \cdot \delta(g, g_i) \cdot \delta(d, d_i) + \beta - 1 \quad (46)$$

5.7 移動予測

5.5節で求めたパラメータを用いて、移動先のメッシュの推定を行う。マルコフ連鎖を用い

て移動先のメッシュの位置を予測する.

5.7.1 移動モード推定

5.5 節で求めたパラメータを用いて移動モード m の事後確率を求める. その際, テストデータとして, 自分の過去の Trajectory を用いる. 確率が最も高いモードが自分の Trajectory における移動モードとする. すなわち, Trajectory の最中に移動モードが変わることはないと仮定する.

モード m が決定された場合, そのモードにおけるメッシュ内平均滞在時間が定義可能である. 本研究では, 予測するすべての Trajectory に対するメッシュ内滞在時間は, 各モード m におけるメッシュ内平均滞在時間であると仮定し, メッシュ内滞在時間まではメッシュの移動は行われぬものとする.

5.7.2 移動先メッシュの予測

5.5 節では, 各モード m に対してメッシュ間の遷移確率の推定を行った. 各モード m に対する遷移確率を元に, 3.6 節で説明した単純マルコフ連鎖を用いて推定を行う.

第6章 評価実験

本章では，第5章で説明した提案手法における評価実験を行う。

6.1 データセット

本研究で使用するデータセットは Microsoft Research の GeoLife GPS Trajectories[3][4][5] (以下 GeoLife)である。GeoLife とは最大で2007年4月から2012年8月までの期間182人に対して位置情報を収集したものである。GeoLifeの特徴として，移動モードが付与されていることにある。移動モードとは，どのような手段によって移動したかを表すもので，全11種類存在する。そのモード名は，「walk」, 「subway」, 「car」, 「bike」, 「motorcycle」, 「airplane」, 「taxi」, 「train」, 「boat」, 「run」, 「bus」であり，モードが一部の Trajectory に対して付与されている。182人のうち64人の Trajectory には，少なくとも1点以上の位置情報に対して移動モードが付与されたデータが存在する。なお，GeoLifeの移動モードは，GPS Point に直接ラベル付けされているのではなく，ある Trajectory のある時間からある時間まで，どの移動モードであったかを示すファイルが別途存在する。

1つの.pltファイルには，ユーザ1人の1日分のデータが含まれている。表6.1にGeoLifeにおける各々のGPS Pointのデータ形式を表A.1に示す。

表 6.1 データ形式

列番号	項目
1	緯度 (10進数)
2	経度 (10進数)
3	未使用 (常に 0)
4	標高 (整数/-777 はエラー値)
5	1899/12/30 からの日数 (10進数/小数点以下最大 10桁)
6	日付 (YY/MM/DD 形式)
7	時刻 (GMT/hh:mm:ss 形式)

また，移動モードは全11種類存在し，そのモード名は，「walk」, 「subway」, 「car」, 「bike」, 「motorcycle」, 「airplane」, 「taxi」, 「train」, 「boat」, 「run」, 「bus」である。なお，GeoLifeの移動モードは，GPS Point に直接ラベル付けされているのではなく，ある Trajectory のある時間からある時間まで，どの移動モードであったかを示すファイルが別途存在する。

6.2 前処理

6.2.1 位置情報の整形

付録 A で示したように、移動モードは位置情報に直接付与されているのではなく、ある Trajectory のある時間からある時間まで、どの移動モードであったかを示すファイルが別途存在している。そのため、GPS Point に移動モードを付与した。また、一部の位置情報にのみ移動モードが存在する。そのため、移動モードが不明な場合は移動モードを「undefined」とした。

また、表 6.1 の列番号 3,4,5 の部分は本研究では不必要なため削除し、表 A.1 の列番号 6,7 をまとめて UNIX time¹として時刻を取り扱うことにした。整形後の GPS Point の形式を表 6.2 に示す。

本研究では、歩行者を対象とした研究であるため、歩行者の GPS データが多く存在する、北京市内の Trajectory のみを対象とした。具体的には、北緯 39.80 度～北緯 40.05 度、東経 116.20 度～東経 116.55 度までの南北約 27.8km、東西約 29.9km の範囲に限定した。

表 6.2 整形後のデータ形式

列番号	項目
1	緯度 (10 進数)
2	経度 (10 進数)
3	時刻 (GMT/10 進数/UNIX time)
4	移動モード (文字列/不明の場合は「undefined」)

6.2.2 Trajectory の分割

GeoLife の .plt ファイルには、ユーザ 1 人の 1 日分のデータが含まれている。そのため、一つのファイルの中には 2.1 節で定義した Trajectory が複数存在していると考えられる。したがって、 $T_r = p_{k+1}.t - p_k.t \geq 300(s)$ である場合、 p_k が属している Trajectory と p_{k+1} が属している Trajectory は別のものとみなし、 p_k と p_{k+1} の間で Trajectory を分割する。

6.2.3 移動モードの推定

移動モードは一部の GPS Point にのみ付与されているため、GeoLife の「walk」「run」ラ

¹ 1970/1/1 00:00:00 からの形式的な経過秒数を表す

ベルが付与されている Trajectory のみを使用するとデータ量が不足する。そのため、研究に利用可能なデータ量を増加させる目的で移動モードを推定する。移動モードを推定する研究はこれまでも行われてきた[19]が、本研究においては、簡単のため、Trajectory の移動速度をもとに移動モードを擬似的に付与する。図 6.1 に示すように、各移動モードに移動速度の差がある。そのため、本研究では平均 5km/h 以下で移動している Trajectory を歩行者とみなした。結果、本研究で利用可能な Trajectory の数は 2,597 となった。

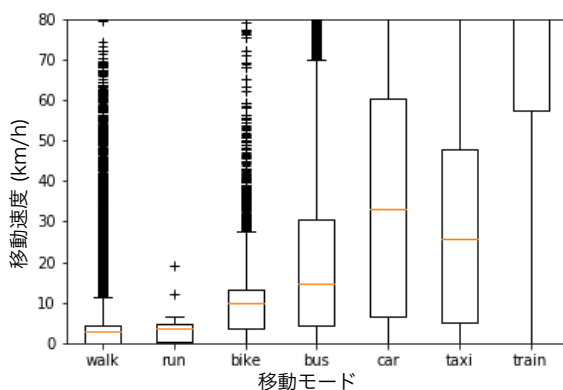


図 6.1 主な移動モードと移動速度の関係

6.3 評価手法

MAP 推定や最尤推定を用いた位置情報予測を行うため、評価手法はメッシュに対する確率を用いる。具体的には、ある観測点 p_k の n 分後($0 < n \leq 10$)に Ground Truth が属しているメッシュと同じメッシュを予測した確率を用いる。各々の Trajectory に対して予測を行い、その平均を求める。6.1 節から 6.3 節で利用した 2,597 の Trajectory のうち、90%を学習データ、10%をテストデータとし、10-cross-validation を行う。評価する際は、過去の Trajectory を使用する時間の長さを m 分間($0 \leq m \leq 10$)と可変させて評価を行う。

6.4 実験結果

6.4.1 EM アルゴリズムを用いた MAP 推定によるパラメータ推定

5.5 節で説明した EM アルゴリズムを用いた MAP 推定を行う。本研究ではメッシュ間の遷移確率を求めるために、それぞれのメッシュに対して 15 個以上の Trajectory を通過したのに対してパラメータ推定を行った。0 個を含む 15 個未満の Trajectory に関しては、 d

の遷移確率をすべて 0.25 であるとみなした。

またパラメータ推定で、歩行者の中での移動モード m の数を予め決定した。そのために、すべてのメッシュに対して、メッシュ内滞在時間 t の分布を表すグラフを示した。その結果を図 6.2 に示す。

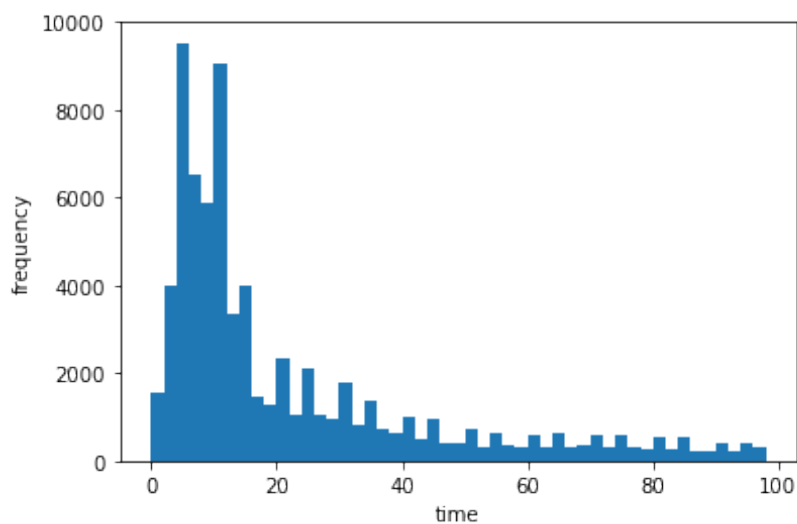


図 6.2 メッシュの滞在時間の度数分布

図 6.2 より、移動モードの総数 C_m は 2 と決定した。

6.4.2 メッシュの移動予測

EM アルゴリズムと単純マルコフ連鎖を組合せて実験を行った。実験結果を図 6.3 に示す。本研究の結果は予測開始から 2 分後の精度が 20.2%，10 分後の精度が 3.23% となった。

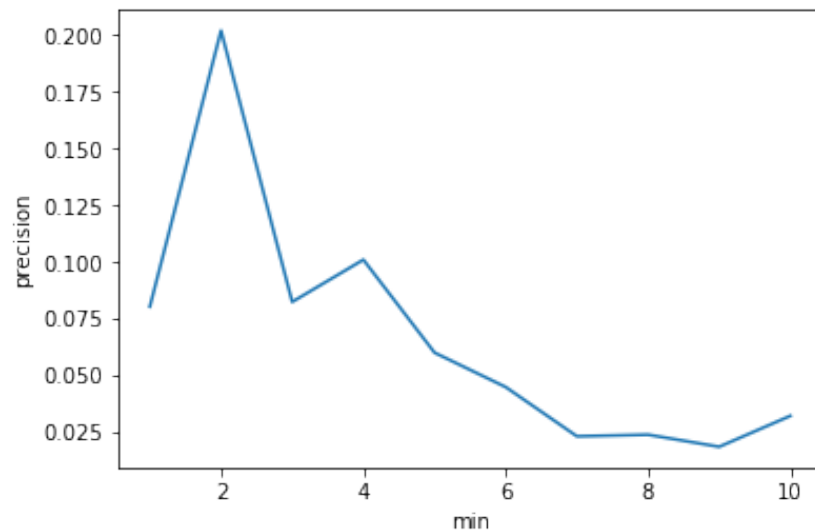


図 6.3 時間経過と精度の関係

6.5 評価実験のまとめ

本研究において、10-cross-validation の結果、2 分後の精度が 20.2%、10 分後の精度が 3.23%となった。精度があまり向上しなかった原因として、データ不足と、単純マルコフ連鎖を利用したことが考えられる。単純マルコフ連鎖を利用したということは、予測開始時点に滞在するメッシュの 1 つ前のメッシュのみ考慮しているためである。3.4 節で説明した N 重マルコフ連鎖を用いることで、複数前の状態を把握することができるため、精度の向上を図ることが出来るとか考えられる。

データ不足の問題に関しては、多くのメッシュに Trajectory が通過していないという問題があった。それぞれのメッシュに対する Trajectory の通過した個数の度数分布を図 6.4(a) に示す。また、度数分布を見やすくするために、図 6.4(b)に Trajectory が 2 つ以上通過したメッシュのみを取り出して度数分布を示した。

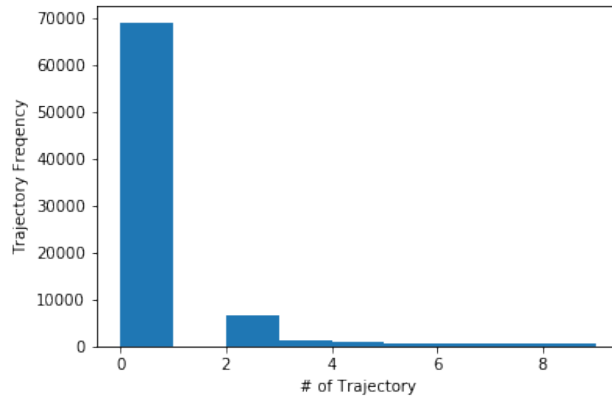


図 6.4(b) Trajectory の度数分布

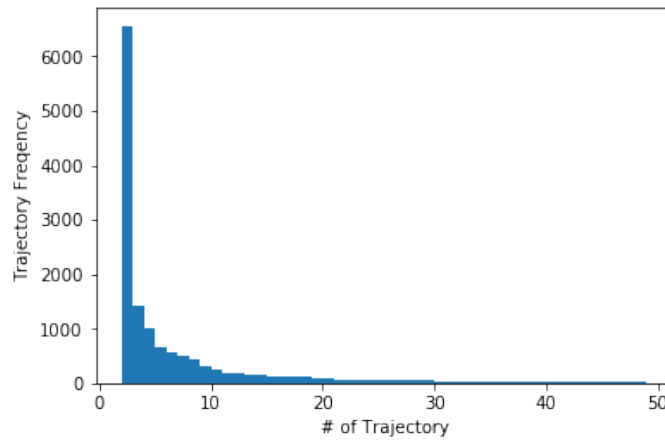


図 6.4(b) Trajectory の度数分布 (2 以上)

データ不足を解決するためには、本研究では、歩行者の位置情報予測を行った研究であるが、Trajectory モデルの構築に歩行者以外の Trajectory を用いてデータ量を増加させ、より広範な範囲に対して予測が可能になると考えられる。

また、考えられる問題点に、すべてのメッシュに対して、メッシュ内滞在時間が同じであると仮定した。そのため、1 分間にメッシュの移動が行われていない状態の仮定が行われておらず、特に 1 分後の結果に関して精度が低くなった。

第7章 おわりに

本研究では、POI 推薦に利用するため、歩行者に対する位置情報の予測を行う手法を提案した。既存研究では、主に自動車を対象としたルートの予測、また歩行者を対象とした研究であっても、目的地の予測のみであった。これら既存研究の問題点を、100m メッシュを用いて遷移確率を MAP 推定で予測し、単純マルコフ連鎖を用いることで、歩行者に対する位置情報予測が目的地に依存せず可能になった。本研究では、メッシュを用いた位置情報予測を行い 10-cross-validation の結果、3.23%の精度を達成した。

本研究においては、単純マルコフ連鎖を用いたため、また歩行者のみの Trajectory を対象としたため、精度があまり高く求められなかった。3.4 節で説明した N 重マルコフ連鎖を用いることで、複数前の状態を把握することができるため、精度の向上を図ることが出来る。また、本研究では、歩行者の位置情報予測を行った研究であるが、Trajectory モデルの構築に歩行者以外の Trajectory を用いてデータ量を増加させ、より広範な範囲に対して予測が可能になると考えられる。

その他の今後の課題は、メッシュのサイズを可変にすることで高密度に位置情報がある箇所については精密な予測が、歩行者以外における位置情報の予測への拡張、POI 推薦の具体的なアプリケーション作成が考えられる。

謝辞

本研究を行うにあたり、大変多くの方のご協力を得ることができました。山名早人教授には、全ゼミやディスカッション、論文の添削など、数多くのご指導をいただきました。心より御礼申し上げます。研究の相談や助言をいただいた高須淳宏教授には心より感謝いたします。また、研究室の同期の方々には普段から仲良くしていただき、研究が行き詰った時など、心の支えになりました。深く感謝いたします。

参考文献

- [1] Y. Zheng: “Trajectory data mining: an overview”, *ACM Trans. on Intelligent Systems and Technology*, vol. 6., No. 3, No. 29, 2015.
- [2] Y. Zheng and X. Xing: “Learning travel recommendations from user-generated GPS traces”, *ACM Trans. on Intelligent Systems and Technology*, vol. 2, No.1, article. 2, 2011.
- [3] Y. Zheng, L. Zhang, X. Xie and W. Y. Ma, “Mining interesting locations and travel sequences from GPS trajectories”, *Proc. of International conference on World Wild Web*, pp.791-800, 2009.
- [4] Y. Zheng, Q. Li, Y. Chen, X. Xie and W. Y. Ma: “Understanding Mobility Based on GPS Data”, *Proc. of ACM conference on Ubiquitous Computing*, pp.312-321, ACM, 2008.
- [5] Y. Zheng, X. Xie, W. Y. Ma: “GeoLife: A Collaborative Social Networking Service among User, location and trajectory”, in *IEEE Data Engineering Bulletin*. Vol.33, No.2, pp. 32-40. 2010.
- [6] J. Yuan, Y. Zheng, X. Xie and G. Sun: “Driving with knowledge from the physical world”, *Proc. of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 316-324, 2011.
- [7] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun and Y. Huang: “T-drive: driving directions based on taxi trajectories”, *Proc. of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 99-108, 2010.
- [8] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Huang and Z. Xu: “Destination Prediction by Sub-trajectory Synthesis and Privacy Protection against such Prediction”, *Proc of International Conference on Data Engineering*, pp. 254-265, 2013.
- [9] A. Monreale, F. Pinelli, R. Trasarti and F. Giannotti, Wherenext: “A Location Predictor on Trajectory Pattern Mining”, *Proc. of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 637-646, 2009.
- [10] J. J. C. Ying, W. C. Lee, T. C. Weng and V. S. Tseng, “Semantic Trajectory Mining for Location Prediction,” *Proc. of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 34-43, 2011.
- [11] S. Gambs, K. Marc-Olivier and N. Miguel: “Next Place Prediction using Mobility Markov Chains”, *Proc of the First Workshop on Measurement, Privacy, and Mobility*, pp.1-6, ACM, 2012.
- [12] A. Asahara, K. Maruyama, A. Sato and K. Seto: “Pedestrian-movement Prediction

- based on Mixed Markov-chain Model,” *Proc. of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 25-33, 2011.
- [13] W. Huang, S. Li, X Liu and Y. Ban: “Predicting human mobility with activity change”, *International Journal of Geographical Information Science*, vol.29, no.9, pp. 1569-1587, 2015.
- [14] N. Ye, Y. Zhang, R. Wang and R. Malekian: “Vehicle trajectory prediction based on Hidden Markov Model”, *KSII Trans. on Internet and Information Systems*, vol. 10, no. 7, pp. 3150-3170, 2016.
- [15] J. A. Alvarez-Garcia, J. A. Ortega, L. Gonzalez-Abril and F. Velasco: “Trip destination prediction based on past GPS log using a Hidden Markov Model,” *Expert Systems with Applications* vol.37, no.12, pp.8166-8171, 2010.
- [16] H. Yuan, Y. Qian, R. Yang and M. Ren: “Human Mobility Discovering and Movement Intention Detection with GPS Trajectories,” *Decision Support Systems*, no.63, pp.39-45, 2013.
- [17] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei and S. Savarese: “Social lstm: Human Trajectory Prediction in Crowded Spaces,” *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 961-971, 2016.
- [18] J. Zhang, Y. Zheng and D. Qi: “Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction,” *Proc. of the 31st AAAI Conference on Artificial Intelligence*, pp.1302-1308, 2017.
- [19] A. Kinoshita, A. Takasu, K. Aihara, J. Ishii, H. Kurawasa, H. Sato, M. Nakamura and J. Adachi: “GPS Trajectory Data Enrichment based on a Latent Statistical Model,” *Proc. of The International Conference on Pattern Recognition Applications and Methods*, pp.255-262, 2016.
- [20] A. P. Dempster, N. M. Laird and D. B. Rubin. “Maximum likelihood from incomplete data via the EM algorithm.” *Journal of the royal statistical society. Series B (methodological)*, pp.1-38, 1977.

付録: 業績

修士の業績を以下に示す.

1. 塩浦 尚久, 田中 博己, 山名 早人, “構図を用いた写真と絵画間の類似判定”, DEIM2017, B4-5, 2017.
2. M. Takahashi, H. Tanaka, H. Yamana and T. Nakajima, “Virtual Co-Eating: Making Solitary Eating Experience More Enjoyable”, *Proc. of International Conference on Entertainment Computing*, pp. 460-464, 2017.
3. 田中 博己, 高須 淳宏, 山名 早人, “MAP 推定を用いた歩行者の位置情報予測 –POI 推薦を目的として–”, DEIM2018 (発表予定)