

移動先エントロピーを用いた車両の走行履歴による到着地の推定

吉岡元貴[†] 小澤 順[†]

本稿では、走行履歴から到着地を推定する場合の困難性を示す指標として、移動先エントロピーを提案し、走行履歴から到着地を推定する精度との関係を明らかにする。走行履歴の移動先エントロピーは、出発からの所定時間における移動先の不確定さを表現しており、走行時間とともに減少する。従来の車両の走行履歴から到着地を推定する手法においては、運転者の走行履歴における到着地の頻度の情報を基に到着地の推定を行っている。そのため、到着地の推定性能は、過去の走行履歴の内容に大きく依存し、各運転者によっても推定精度のばらつきが大きい。そこで、走行履歴の移動先エントロピーを算出することで、走行履歴から到着地を推定する困難性の定量化を行い、到着地の推定への可能性を探る。車両走行履歴の移動先エントロピーと到着地の推定性能の関係を明らかにするため、10名の運転者の走行履歴を取得し、実走行履歴に対する移動先エントロピーを算出する実験を行った。その結果、移動先エントロピーと到着地の推定精度に関しては大きな相関があり、走行開始から5~10分で移動先エントロピーの減少および移動先の推定精度の向上が見られた。また、走行数の増加とともに移動先エントロピーは増加するが、ある所定値の走行数で収束が見られた。収束する走行数は被験者によって異なり、到着地の推定に必要な走行数が推定できる可能性のあることが分かった。

Destination Entropy for Arrival Place Presumption from Car Driving Route History

MOTOTAKA YOSHIOKA[†] and JUN OZAWA[†]

In this paper, the relation between the destination entropy — the index showing the difficulty to estimate the arrival place — and the accuracy of the arrival place presumption from driving history is investigated. In the previous methods for destination estimation from driving history, the destination was presumed based on the arrival place frequency in each driver's history. Therefore, it was greatly dependent on the content of the history runs as well as on each driver. In our method, the difficulty of presuming the arrival place was quantified by calculating the destination entropy from the driving history. In order to clarify the relation between the destination entropy and the performance of the proposed method, we calculated the entropy using real driving history of 10 drivers. A strong correlation between the destination entropy and the accuracy of the presumed destination was shown. In the first 5 to 10 minutes after the driving start we have obtained a significant entropy reduction and accuracy improvement. Although the destination entropy increased with the increase of number of runs, it converged after a certain predetermined value. The number of runs at which the convergence appears changes with each driver and despite the increase of the accumulated driving history data, the presumption accuracy did not improve.

1. はじめに

近年、携帯電話やカーナビゲーション(以下、カーナビ)等の屋外で利用するモバイル端末において、インターネット等による情報取得が可能となり、ユーザは時間や場所を越えて情報と接する機会を得るようになっている。これらモバイル端末には、GPS(Global Positioning System)機能が搭載され、検出されたユーザの現在位置に応じて情報を提供するサービスも行わ

れている^{1),2)}。

一方、GPS等の位置情報取得デバイスの普及を背景に、ユーザの行動特徴を抽出するために行動履歴を解析し、ユーザの行動をモデル化しようとする研究が数多く行われている³⁾⁻⁵⁾。これら行動分析では、ユーザの行動履歴から「行動A、行動Bが起これば、行動Cが起こる」等、所定の単位行動の組合せや遷移関係を抽出する。さらに、あらかじめユーザによって登録されたスケジュール情報を用いたユーザの行動単位の遷移関係の抽出や⁶⁾、ユーザが歩行していた道路の勾配や移動手段の情報を用いて移動特徴を分析する研究も行われている^{7),8)}。このようにユーザの行動を分析

[†] 松下電器産業株式会社先端技術研究所
Advanced Technology Research Laboratories, Matsushita Electric Industrial Co., Ltd.

することで、モバイル分野でのマーケティングへの利用や、あるいは将来のユーザの移動先を予測し、移動先に関する情報を事前に提供することで、モバイル端末での情報サービスの高機能化を目指している⁹⁾。

また、車両向けの情報サービスの高度化を目指し、車両の移動履歴を解析する研究も多く行われている。たとえば、車両運転者が通過した交差点等の系列と、さらに交差点を通過したときの天気、交通量、日付等の付帯情報を蓄積し、ユーザの将来の移動先を推定する研究がなされている^{10),11)}。また、運転者に有用な情報を提供するため、蓄積された移動履歴から運転者のプロフィールに応じて移動先を推定し、推定した移動先に基づく情報提供手法の有効性を示している^{12),13)}。また、VICS (Vehicle Information and Communication System) 情報の履歴を解析することで、将来の交通渋滞を推定する研究もある¹⁴⁾。

一方、車両の走行履歴からユーザの到着地を推定する研究において、その推定の精度はユーザによって様々であることが示されている^{12),15)}。たとえば、普段同じ経路を走行するユーザと比較して、渋滞等の交通状況に応じて迂回経路を選択する車両の到着地を推定することは困難である¹²⁾。また、車両を通勤に利用するか否かによって、車両の移動先を推定する精度に違いが生じる¹⁵⁾。

従来までの到着地の推定手法は、蓄積された走行履歴をもとに類似経路の検索と、その結果の経路における到着地を用いて行うのが一般的である。具体的には、ユーザの車両の走行履歴を蓄積し、現在地点までの経路が与えられたとき、その経路と一致する走行履歴を検索し、検索された走行履歴の到着地のうち、到達頻度の多いものを推定結果として出力する。そのため、推定の精度はあらかじめ蓄積される走行履歴に大きく依存し、ユーザによって推定精度の差が大きい。また到着地の推定の評価は、現在地点までの走行で推定された到着地と、その後の実際の走行での到着地の一致度で評価したものであり、履歴を取得した時期や、評価に用いた走行によっても大きく影響を受ける。したがって実験的に算出された推定精度のみからでは、到着地の推定手法の性能が保証できない。

本稿では、走行履歴から到着地の推定を行う困難性を示す指標として、あらたに移動先エントロピーを提案し、移動先エントロピーと到着地推定精度との関係について議論する。移動先エントロピーとは、移動開始からの所定時間経過後における、移動先の不確定さを情報量を用いて表したものである。移動先エントロピーは、まず、ある出発地から所定の走行時間までの

推移系列の集合を取り出し、類似した推移系列を1つのクラスに分類し、各クラスの利用確率と、各クラスにおける移動先の分散度合いの積によって算出される。

走行時間ごとに変化する移動先エントロピーの収束性は、移動先のばらつき度合いの収束性を示しているため、車両走行履歴における移動先エントロピーの値から、運転者の走行を多面的に評価することが可能となる。また、移動先エントロピーの推移を観測することで、所定の推定精度を実現するために必要な走行時間を算出することが可能となる。また、蓄積された走行履歴の走行数に応じたエントロピーの変化を観測することで、所定の推定精度を実現するために必要な走行数を算出することが可能となる。さらに、移動先エントロピーは各運転者ごとに異なり、この値を比較することで運転者ごとに到着地の推定の可能性を比較できる。

以下、2章で従来までの到着地の推定手法について説明し、その問題点と、移動先エントロピーの必要性について述べる。3章では移動先エントロピーの算出方法を定式化し、具体例を用いて算出方法を説明する。4章では車両走行データに対して本稿で提案する移動先エントロピーを算出し、その結果について述べる。最後に5章でまとめと今後の課題について述べる。

2. 車両走行履歴を用いた到着地の推定とその課題

2.1 車両走行履歴を用いた到着地の推定

従来、ノードの推移からなる時系列データから、頻出パターンを抽出し、所定のノード系列が与えられたときに、与えられた系列以降のノードの推移を推定する手法が提案されている^{16),17)}。車両走行履歴から到着地を推定する場合も、これらの手法を適用することができる。次に到着地を推定する一般的な手法について説明する。

図1に示すように、ユーザは車両を用いて出発地から到着地へ、交差点を経由しながら走行する。このとき、出発地、交差点、到着地を1つのノードとして表現し、ユーザの走行をノード間の推移系列としてとらえ、その推移系列を走行履歴として蓄積する。たとえば、「ノードAを出発してノードB、そしてノードD、ノードG、ノードH、と経由し、ノードXへ到着した」という走行履歴が、「A → B → D → G → H → X」というノードの推移系列で蓄積される。

ユーザの所定時間までの走行経路として、「A → B → D」というノードの推移系列が検出されたとする。このとき、走行経路と一致する推移系列を走行履歴よ

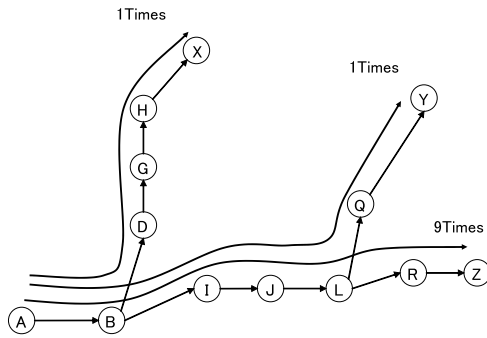


図 1 車両の走行履歴とノードの推移系列
Fig. 1 Driving history and node sequence.

り検索する．たとえば，前述の推移系列「A → B → D → G → H → X」が走行履歴として蓄積されている場合，この推移系列をもとに到着地は X である，と推定することとなる．

なお，上記例は現在地点までの推移系列と，過去の走行履歴における推移系列の 1 つとが一致し，その推移系列の到着地が 1 つであるため，到着地を一意に推定できた．しかしながら，現在地点までの推移系列によっては，過去の走行履歴で到着地が一意に決まらない場合もある．たとえば，現在ユーザの車両が「A → B」と走行してきたとする．ここで，到着地として考えられるのは，ノード X，ノード Y，ノード Z と複数にわたり，一意に推定することができない．一方，その後，車両がさらに移動し，「A → B → D」と走行したとする．この場合，蓄積された走行履歴によると，到着地はノード X と限られており，到着地は X と推定することが可能となる．また，その後ノード G，ノード H，と走行し続けても，到着地はノード X であり，その後の走行によって到着地は変わらない．

次に，推移系列の頻度を利用して，到着地を推定する例について述べる．現在ユーザの車両が「A → B → I」と移動してきたとする．この場合，到着地として考えられるのは，ノード Y，ノード Z である．さらにノード J，ノード L と走行を続けても到着地を一意に推定することができず，ノード Q，またはノード R まで走行しなければ，一意には推定することができない．しかし，ここで走行履歴をもとに到着地への到達した頻度を考慮すると，一意に推定することができる．たとえば走行履歴より，ノード Z への到達した頻度が 9 回，ノード Y への到達した頻度が 1 回であったとする．前述の「A → B → I」と走行してきた場合，過去 1 度しか到達したことがないノード Y ではなく，9 回と頻度高く到達しているノード Z であろうと一意に推定することが可能となる．

2.2 到着地の推定精度とその課題

前節で述べたように，従来までの到着地推定手法は，現在地点までの推移系列が与えられたとき，蓄積された走行履歴をもとに，その推移系列と一致する推移系列を走行履歴より検索する．検索された推移系列の到着地のうち，到達した頻度の多いものを推定結果として出力し，この推定結果が実走行での到着地と一致しているか否かで評価を行っている．そのため，到着地の推定の性能は，車両の走行履歴に依存する．たとえば，ユーザが普段通勤に車両を利用し，所定の経路に対して到着地が一意に決まる場合には，推定精度が向上するが，1 つの経路に対して多数の到着地が存在する場合には，推定精度が低くなってしま¹⁵⁾．さらに，同じユーザの車両の走行履歴を利用しても，到着した頻度をもとに到着地を推定する場合，過去の走行履歴と同一の走行をした場合でも，推定に成功しない場合もある．このように，走行履歴によって到着地推定が容易な走行履歴と，困難な走行履歴が存在する．

また，一般には，推定を行うための走行履歴を多く蓄積することにより，推定精度の向上が期待できる．しかしながら，各走行履歴に対して，どれだけの走行を蓄積すれば十分な推定精度が実現できるかも分からない．また，車両が出発してから走行時間が長ければ，推定精度が向上すると予想される．一方で，走行時間が長いと，目的地へ到着してしまう可能性があり，推定の意味がなくなる．そのため，到着地を推定するための適切な走行時間を設定する必要があり，この値は走行履歴に応じて異なるものと予想される．

以上のように，走行履歴から到着地の推定を行うためには，その走行履歴の特性と推定性能との関係を定量的に解析する必要がある．そこで，車両走行における移動先の分散度合いを表現する指標として移動先エントロピーを提案する．走行履歴の移動先エントロピーを算出することで，車両の走行履歴ごとに到着地を推定する困難性や，推定を行う変数を決定するための指針の導出を行う．

3. 移動先エントロピー

3.1 移動先エントロピーの提案

車両の走行履歴に基づいて出発から所定の時間経過後における到着地の不確定さを表した移動先エントロピーを提案する．車両の走行を出発地，交差点，到着地のノードの系列で表す．1 つのノードを s とすると，出発地 s_1 ，交差点 s_2, \dots, s_{V-1} ，到着地 s_V までの V 個のノードからなる 1 回の推移系列は，

$$S = \langle s_1, s_2, \dots, s_V \rangle$$

で表現される．また，これら推移系列 S が N 個の蓄積された推移系列の集合は，

$$S = \{S_1, S_2, \dots, S_N\}$$

で表現される．

ここで，走行を開始してから到着地の推定を行うまでの時間を T とする．このとき到着地推定は，現在走行における時間 T までのノードの推移系列と，蓄積された走行履歴のうち，時間 T まで推移系列と一致する推移系列を検索して行う．さらに，その推移系列の集合のうち最も頻度の高い到着地を，今後，到着する到着地と推定する．次に，出発してから時間 T までの推移系列を有する推移系列の集合の生成について説明する．出発してから時間 T までの推移系列を，

$$S(T) = \langle s_1, s_2, \dots, s_U \rangle \quad (U \leq V)$$

とすると，走行履歴 S における，時間 T までの推移系列の集合は，

$$S(T) = \{S_1(T), S_2(T), \dots, S_N(T)\}$$

で表現される．ただし，走行履歴 S においては異なる推移系列であったが，走行を開始してから時間 T までの走行になることにより，同一となる $S(T)$ が存在するために， $S(T)$ をあらためて，代表推移系列 $F(T)$ の集合，

$$F(T) = \{F_1(T), F_2(T), \dots, F_M(T)\} \quad (M \leq N)$$

として定義する．次に各 $F(T)$ を接頭語 (*prefix*) に持つ推移系列 S の集合を， $C_{F(T)}$ とする．そして到着地推定は，時間 T までの現在走行が検出された場合，検出された現在までの推移系列と一致する $F(T)$ のクラス $C_{F(T)}$ に含まれる推移系列の到着地を用いて到着地の推定を行う．したがって，出発から所定時間 T までの走行における到着地推定の評価分析は，各クラス $C_{F(T)}$ に属する履歴を分析することと等しくなる．そこで，各クラス $C_{F(T)}$ の到着地のエントロピーを算出し，そのクラスの利用率と，エントロピーの積を移動先エントロピーと定義する．

まず，各クラス $C_{F(T)}$ に含まれる推移系列 S の走行数を $\text{Freq}(C_{F(T)})$ で表現する．また，走行履歴 S における K 個の到着地のノードを s_k ($k = 1, 2, \dots, K$) とすると，各クラス $C_{F(T)}$ においてノード s_k へ到着する確率 P_k は，

$$P_k = \frac{\text{Freq}(s_k | C_{F(T)})}{\text{Freq}(C_{F(T)})} \quad (1)$$

となる．ただし， $\text{Freq}(s_k | C_{F(T)})$ は， $C_{F(T)}$ に含まれる走行数のうち，ノード s_k へ到着した走行数を表している．したがって，各クラス $C_{F(T)}$ のエントロピー $E_{C_{F(T)}}$ は，上記 P_k を用いて以下の式で表すことができる．

表 1 走行履歴の例

Table 1 Example of driving route history.

01	(A,0) → (B,3) → (D,5) → (G,7) → (H,8) → (X,10)
02	(A,0) → (B,3) → (I,5) → (J,7) → (L,9) → (R,12) → (Z,13)
03	(A,0) → (B,3) → (I,5) → (J,7) → (L,9) → (Q,12) → (Y,14)
04	(A,0) → (B,3) → (I,5) → (J,7) → (L,9) → (R,12) → (Z,13)
05	(A,0) → (B,3) → (I,5) → (J,7) → (L,9) → (R,12) → (Z,13)

$$E_{C_{F(T)}} = \sum_{k=1}^K \frac{-P_k \log P_k}{\log K} \quad (2)$$

さらに，上記エントロピー $E_{C_{F(T)}}$ に，到着地推定の際，クラス $C_{F(T)}$ が選択される割合を考慮して平均をとることで，所定時刻 T までの走行における到着地推定の評価値である，移動先エントロピー $E(T)$ が得られる．各ユーザの走行履歴 S より，クラスが M 個構築されているので，移動先エントロピー $E(T)$ は，

$$E(T) = \sum_{m=1}^M \frac{|C_{F_m(T)}|}{N} E_{C_{F_m(T)}} \quad (3)$$

となる．この $E(T)$ は，出発してから走行時間 T における，到着地のあいまいさの平均を示す値となっており，最もあいまいな場合， $E(T) = 1$ ，到着地が 1 つに絞りきれいている場合は $E(T) = 0$ となるようになっている．また，時刻 T を変更することで，時間の経過にともなう，到着地のあいまいさとの関係を表すことも可能である．

3.2 移動先エントロピーの計算例

走行履歴が与えられたときに，移動先エントロピーを算出する方法について，具体例を用いて説明する．表 1 に示すように 5 つの推移系列が蓄積されているとする．たとえば 1 行目の推移系列は，ノード A を出発し，時刻 3 (単位は分とする) 後にノード B，5 分後にノード D へ移動し，最終的に 10 分後にノード X へ到着したことを示している．これらの推移系列は，

$$S_1 = \langle A, B, D, G, H, X \rangle$$

$$S_2 = \langle A, B, I, J, L, R, Z \rangle$$

$$S_3 = \langle A, B, I, J, L, Q, Y \rangle$$

$$S_4 = \langle A, B, I, J, L, R, Z \rangle$$

$$S_5 = \langle A, B, I, J, L, R, Z \rangle$$

となる．そして走行開始から 5 分までの推移系列は，

$$S_1(5) = \langle A, B, D \rangle$$

$$S_2(5) = S_3(5) = S_4(5) = S_5(5) = \langle A, B, I \rangle$$

となる．したがって，代表推移系列は，

$$F_1(5) = \langle A, B, D \rangle$$

$$F_2(5) = \langle A, B, I \rangle$$

表 2 各運転者のプロフィール
Table 2 User's driving profile.

	走行数(回)	平均走行時間(分/回)	平均走行数(回/月)	到着地数(箇所)	休日利用率(%)	運転者タイプ
User1	185	24.7	49.0	47	0.25	通勤者
User2	107	22.8	34.6	46	0.71	サンデードライバ
User3	617	17.5	78.5	104	0.37	通勤者
User4	286	41.8	57.0	75	0.29	通勤者
User5	427	8.4	111.6	84	0.30	主婦
User6	236	25.3	79.0	62	0.31	主婦
User7	156	12.9	150.0	16	0.30	主婦
User8	106	40.2	78.0	44	0.63	サンデードライバ
User9	189	11.1	74.0	47	0.36	主婦
User10	285	39.9	47.1	59	0.18	通勤者

となる．次に上記 2 種類の推移系列と一致する推移系列を 1 つのクラスとして集めると，

$$C_{F_1(5)} = \{S_1\}$$

$$C_{F_2(5)} = \{S_2, S_3, S_4, S_5\}$$

となる．ここで，実際の走行により 5 分までの走行によって $\langle A, B, I \rangle$ が検出されたとする．この場合，クラス $C_{F_2(5)}$ に属する推移系列を用いて到着地を予測することとなる．たとえばクラス $C_{F_2(5)}$ に属する推移系列の各到着地 Y, Z の頻度は， Y の 1 回に対し， Z は 3 回と多く存在するため，到着地を Z と推定することとなる．

一方，移動先エントロピーは，以下のようにして算出される．まず，各クラスにおける移動先のエントロピーを算出する．ユーザが到達しうるゴール数は X, Y, Z の 3 種類であり， $C_{F_1(5)}, C_{F_2(5)}$ は，式 (2) よりそれぞれ，

$$E_{C_{F_1(5)}} = \frac{1 \log 1}{\log 3} = 0$$

$$E_{C_{F_2(5)}} = \frac{0 - \frac{1}{4} \log \frac{1}{4} - \frac{3}{4} \log \frac{3}{4}}{\log 3} = 0.51$$

となる．また，各クラスが利用される確率は， $C_{F_1(5)}$ が $1/5$ ， $C_{F_2(5)}$ が $4/5$ ，であるため，移動先エントロピーは，式 (3) より，

$$E(5) = \frac{1}{5} E_{C_{F_1(5)}} + \frac{4}{5} E_{C_{F_2(5)}} = 0.41$$

となる．

4. 車両走行履歴データへの適応実験

4.1 実験概要

車両の走行履歴データに対して 3 章で定義した移動先エントロピーを算出し，到着地の推定精度との関連性を明らかにすることで，移動先エントロピーの有効性を示す．本実験では松下電器産業(株)製カーナビ「CN-HD9000WD」を改良することで，車両の走行履歴データを取得した．カーナビの GPS ユニット

から車両の緯度経度の位置情報，およびその位置情報が検出された時刻の情報の組を，1 秒間隔で取得した．取得した位置情報から，到着地推定に用いるノードの推移系列を以下の定義により得た．

- (i) エンジンをスタートした地点(出発地)，およびエンジンをストップした地点(到着地)を中心とする所定の半径以内の位置情報を 1 つのノードとする．
- (ii) 2 方向以上に進路選択をしたことのある交差点を中心とする所定の半径以内の位置情報を 1 つのノードとする．

次に取得した緯度経度，および検出時刻情報を，上記で定義したノードとその通過時刻へと変換し，走行開始(エンジン始動)から走行終了(エンジン停止)までの移動を 1 つの系列データとした．

本実験では，被験者 10 名の自家用車に走行履歴データ取得用カーナビを装着し，ユーザの普段の走行データを取得し，走行履歴として用いた．

ユーザの車両の利用環境や取得されるデータに応じて，移動先エントロピーや推定精度は大きく異なる．今回の実験での各ユーザの車両の利用環境と取得したデータの概要を表 2 に示す．表 2 において走行数は，ユーザが走行開始から終了までの走行を 1 走行とした場合の総数である．また，1 走行の平均走行時間である平均走行時間，1 カ月あたりの平均走行数，取得した走行履歴の到着地の種類数を示している．さらに休日利用率は，土曜日，日曜日，祝日等，休日の利用率であり，各ユーザが普段，どのような目的で車両を利用するかという質問を行い，その結果をタイプとして記載している．User1, User3, User4, User10 は，車両を主に通勤に利用するユーザ(通勤者と記す)である．また，User5, User6, User7, User9 は主婦であり，通勤者とは異なり，買い物等，1 日の平均走行時間が短い．また，User2, User8 は，主に日曜のみにしか利用しないユーザ(サンデードライバ)であり，

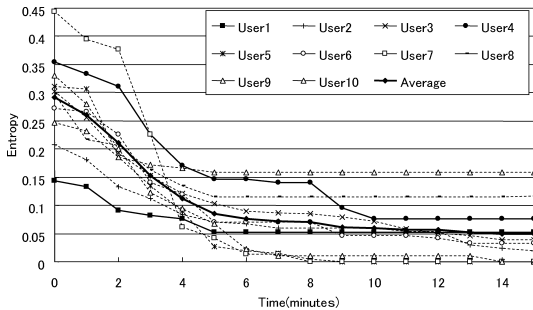


図 2 移動先エントロピーと走行時間

Fig. 2 Driving time and Destination Entropy.

休日利用率が、他の User と比較して高い値となっている。以上のようにバリエーションに富むデータが得られた。

運転開始から所定時間の走行における到着地の推定は、次のようにして成否の判定を行う。はじめに、運転開始から現在地点までの走行による位置情報が、随時検出されているため、その推移系列をノードの推移系列に変換する。次に、変換されたノードの推移系列に対して、過去に走行した推移系列の検索を行う。検索の結果の推移系列において、その到着地の最も多かった地点を到着地として推定する。最後に、実走行において到着した地点と、推定された到着地とが同じであれば、到着地の推定が成功したことになる。一方、異なっていた場合には、到着地の推定に失敗したことになる。

4.2 移動先エントロピーの走行時間による変化

各被験者 (User) の走行開始からの時間経過とともに移動先エントロピーの変化と、その平均値 (Average) を図 2 に示す。図 2 のグラフは、横軸に走行開始からの経過時間 (単位は分で表す)、縦軸に移動先エントロピーの値を示している。移動先エントロピーは、所定時間における移動先の不確定さを表しているため、後に到達する地点があいまいな場合その値は高く、一方、到着地が定まっている場合は低い値となる。

被験者 10 名の移動先エントロピーの平均値は、走行時間 0 分において 0.29、走行時間 5 分において 0.08 と、エントロピーの値が走行時間とともに減少している。また、走行時間 0 分の移動先エントロピーの値は被験者ごとに異なるが、すべての被験者において走行開始から移動先エントロピーが走行にともない減少している。これは、被験者ごとに到着地数が異なるため、後に到達する地点のあいまいさに個人差はあるが、走行時間の経過にともない到着地の候補が絞れ、結果、移動先エントロピーが減少していることを表している。

次に、被験者 10 名の移動先エントロピーの平均値

表 3 移動先エントロピーの収束条件と推定精度との相関
Table 3 Convergent condition of Destination Entropy and correlation between Destination Entropy and arrival place accuracy.

	収束に必要な走行時間 (分)	移動先エントロピーと推定精度の相関係数	収束に必要な走行回数 (回)
User1	5	-0.997	160
User2	4	-0.987	70
User3	4	-0.970	150
User4	10	-0.931	70
User5	5	-0.992	90
User6	5	-0.986	170
User7	4	-0.994	90
User8	3	-0.980	80
User9	6	-0.992	70
User10	3	-0.991	160
Average	5	-0.996	170

に着目すると、出発から 5 分を経過するまで減少を続け、5 分以降は、ある程度一定の値を保ち続けている。これは、所定時間走行することで到着地の候補が絞れ、移動先エントロピーが定常的になり、収束していることを示している。一方、各被験者の移動先エントロピーが収束する時間については個人差があることが分かる。そこで、移動先エントロピーが収束する時間を定義し、各被験者の移動先エントロピーが収束する時間を算出する。まず、出発してからの走行時間 t (分) における移動先エントロピーを $E(t)$ とする。出発時 $E(0)$ に対する、時間 t (分) の移動先エントロピー $E(t)$ の減少率は $(E(0) - E(t))/E(0)$ (%) と表すことができる。また、時間 $t-1$ (分) における減少率は $(E(0) - E(t-1))/E(0)$ (%) となる。よって時間 $t-1$ (分) と、時間 t (分) の減少率の差分 ρ は、

$$\rho = \frac{E(t-1) - E(t)}{E(0)} \quad (4)$$

となる。

ここで、ある時間以降、減少率の差分 ρ の値が所定の値より小さくなった場合に移動先エントロピーが収束し、定常的となったとする。

本実験においては ρ の値が 10% 未満となる状態を移動先エントロピーが収束し、定常的となったと解釈する。各被験者の移動先エントロピー、およびその平均値が定常的となる時間 (分) を表 3 に示す。表 3 より、User4 を除く、すべての被験者が、走行開始から 3 分 ~ 6 分と比較的早い時間で移動先エントロピーが収束し、定常的となることが分かる。また、各被験者の移動先エントロピーの平均値においても、定常的となる時間は 5 分となっている。一方、User4 の移動先エントロピーが定常的となる時間は、10 分と他の被験者と比較して遅く、到着地の候補を絞るために必要

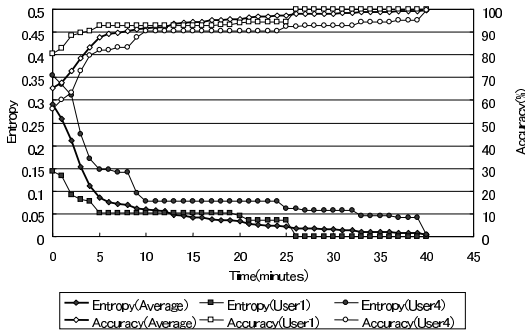


図 3 移動先エントロピーと推定精度

Fig. 3 Arrival place accuracy and Destination Entropy.

な時間はおよそ 5 分から 10 分と、ユーザによって異なることが分かる。

走行開始からの時間経過にともなう移動先エントロピーの変化と、到着地推定の精度との関係の結果を図 3 に示す。図 3 のグラフは、横軸に走行開始からの経過時間（単位は分で表す）、縦軸に移動先エントロピーの値を示している。また、到着地の推定精度を算出するため、被験者ごとに 40 走行の系列を用いて推定精度（単位はパーセントで表す）を算出した。ここでは、移動先エントロピーが定常的となる時間が 5 分と比較的早いユーザの代表として User1 と、定常的となる時間が 10 分と比較的遅い User4 との結果を示す。また、被験者 10 名の移動先エントロピーの平均値と、推定精度の平均値との走行時間に対する関係を示す。

図 3 より、移動先エントロピーの平均値は走行時間 0 分において 0.29、走行時間 5 分において 0.08 と、走行時間とともに減少しているのに対し、推定精度は走行時間 0 分において 65%、走行時間 5 分において 88% と増加している。同様に、User1 の移動先エントロピーは、走行時間 0 分において 0.15、走行時間 5 分において 0.05 と減少しているのに対し、推定精度は走行時間 0 分において 80%、走行時間 5 分において 91% と、移動先エントロピーが減少するにつれて増加しており、移動先エントロピーと推定精度に相関があると考えられる。

そこで次に、各被験者の移動先エントロピーと推定精度の相関関係について計算した結果を表 3 に示す。表 3 より、各被験者の相関係数はすべて -0.9 以下と、高い負の相関があることが分かる。なお、これらの相関係数が -1 とはならないのは、以下の理由によるものである。

推定精度は到達頻度が最大である到着地を推定結果として出力し、その到着地が実走行での到着地と一致

しているか否かで決定される。一方、移動先エントロピーは、到着頻度が最大である到着地と、他の到着地の頻度の相対関係により決定される。その結果、推定された到着地以外の到達地の数や推定された到着地の頻度によって、推定精度は変化しないが、移動先エントロピーの値が変化するために起きている。

次に、User1 および User4 の移動先エントロピーの推移に着目すると、User1 の場合、出発から 5 分を経過するまで減少を続け、5 分以降は、ある程度一定の値を保ち続けている。一方、推定精度は増加し、5 分以降はある程度一定の値を保ち続けている。また、User4 の移動先エントロピーは、10 分を経過するまでは大きく減少を続け、10 分以降の減少は少ない。これらは、所定時間走行することで、ある程度到着地の候補が絞れ、移動先エントロピーが定常的になることを示している。逆に、10 分以上走行しても、到着地の推定精度の向上は少ない。表 3 において、最も相関係数の絶対値が小さい User4 においても、図 3 より移動先エントロピーと推定精度において大きな相関が見られる。

User1 と User4 との移動先エントロピーの推移、および推定精度を比較すると、User1 に対し、User4 の移動先エントロピーの値は、全体的に高くなっていることが分かる。逆に推定精度においては、User1 と比較し、User4 の値は低くなっていることが分かる。これは、User1 に対し、User4 の移動先の不確定性が高く、到着地の推定が困難であることを示している。

以上の結果より、次のことが考察される。走行履歴の移動先エントロピーと到着地の推定精度には大きな相関関係がある。また、走行開始からの時間が 5 分から 10 分で、移動先エントロピーの減少と、推定精度の向上が見られる。しかし、所定時間以降は移動先エントロピーの減少、推定精度ともに向上は見られない。

4.3 移動先エントロピーの走行数による変化

次に移動先エントロピーと、走行数による変化について示す。図 4 は、User10 の所定の走行数における移動先エントロピーを示したものである。横軸は走行開始からの経過時間、縦軸は移動先エントロピーの値である。また、評価した走行数が 30 の場合、90 の場合、120 の場合と、各走行数におけるエントロピーの値を示している。図 4 より、各走行数における移動先エントロピーの値が、走行開始からの経過時間とともに減少しながら推移していることが分かる。

一方、各走行数における移動先エントロピーを比較すると、走行数が増加すると移動先エントロピーも増加していることが分かる。たとえば、走行数が 30 の

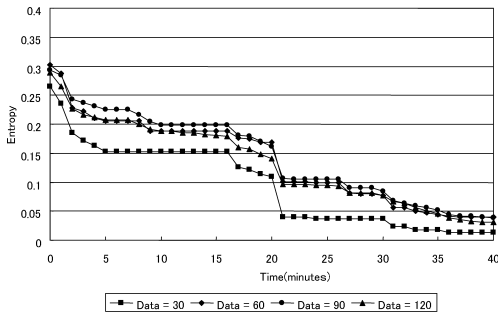


図4 走行数の変化に対する移動先エントロピーと走行時間
Fig. 4 An increase of driving data and Destination Entropy.

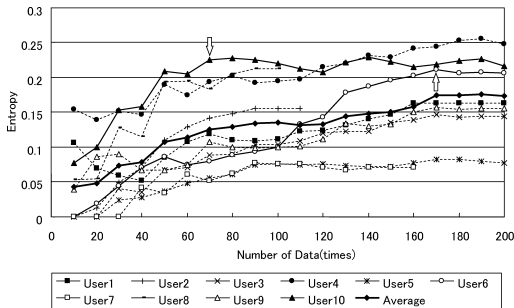


図5 移動先エントロピーと走行数

Fig. 5 Destination Entropy and number of driving data.

場合における走行時間0分の値は0.27, 走行時間5分の値は0.15となっているのに対し, 走行数が60の場合, 走行時間0分において0.29, 走行時間5分において0.21と, 走行数が30の場合と比較して, 移動先エントロピーの値が高くなっている。これは, 走行数が増加するにつれ, 移動のパターンや到着地数が増え, その結果, 移動先エントロピーが増加することを示している。一方, 走行数が60, 90, 120の場合を比較すると, 移動先エントロピーの値や, 時間の経過における推移に大きな差はないことが分かる。これは, 評価する走行数の増加にともない, 移動先エントロピーの値は増加するが, ある程度の走行履歴が蓄積されると, ユーザの移動パターンのおよそ半が蓄積され, ばらつきが定常化することを示している。また, 走行開始から5分で移動先エントロピーが減少し, その後, 15分から20分で再び移動先エントロピーが減少するというパターンについては, 走行数が増加しても保持されている。走行時間の経過にともなう移動先エントロピーの推移の類似性は, User10以外の被験者においても同様の傾向が見られた。

次に各被験者の移動先エントロピーと走行数の関係について示す。図5は, 被験者10名の出発から5分経過後における移動先エントロピーおよびその平均値

と, 走行数との関係を示したものである。横軸は走行数(回), 縦軸はエントロピーの値である。

図5における被験者10名の移動先エントロピーの平均値より, 走行数の増加にともない, 移動先エントロピーが増加している。同様に, 各被験者においても, 走行数の増加にともない, 移動先エントロピーが増加している。これは, 蓄積される走行数が増加することにより, 各ユーザの様々な走行パターンが蓄積され, その後到達する地点のあいまいさが増加していることを示す。

しかしながら, 所定の走行数以降, 各被験者およびその平均値において, 移動先エントロピーの値がほぼ定常的に推移している。これは, 走行データが蓄積されことにより, 各ユーザの移動パターンがほぼ蓄積され, その後は走行数を増加しても, 移動先のあいまいさは増加しないことを示している。

移動先エントロピーが収束し, 定常化する走行数は, 被験者ごとに異なる。移動先エントロピーが収束する走行数を定義し, 各被験者に対して計算する。

まず, 走行数 n における移動先エントロピーを $E(n)$ とする。さらに10回の走行を行った後の移動先エントロピーは $E(n+10)$ となる。そこで, 各走行数 $n(n=10, 20, \dots, 200)$ における移動先エントロピーの増加率を,

$$\xi = \frac{E(n+10) - E(n)}{E(n)} \quad (5)$$

と定義する。

ここで, ある走行数以降, 増加率 ξ が所定の値より小さくなった場合, 移動先エントロピーが収束し, 定常的となったとする。

本実験においては, 10走行の増加による移動先エントロピーの増加率 ξ の値が, 10%未満となる走行数を収束する走行数とする。各被験者の移動先エントロピーが定常的となる走行数(回)を表3に示す。表3より, 被験者ごとに移動先エントロピーが収束し, 定常的となる走行数が異なることが分かる。また, 最も早い段階で移動先エントロピーが収束するのは, User2, User4, User9であり, 走行数は70となっている。一方, 移動先エントロピーの収束に最も走行データを必要とする被験者はUser6であり, 走行数170となっている。また, 各被験者の移動先エントロピーの平均値も, 定常的となる走行数は170であり, その後は定常的となっている。

以上の結果より, 次のことが考察される。走行数が増加するとともに移動先エントロピーは増加し, 所定の値で収束する。また, 走行数が増加しても変化のパ

ターンは一致している。たとえば、User10 の場合、5 分までの走行で移動先エントロピーが減少し、10 分から 15 分では変化がない。また、移動先エントロピーが収束するのに必要な走行数はユーザによって異なり、約 70 走行から 170 走行で推定するための走行が収束し、これ以上履歴を取得しても推定精度の向上は期待できないことが分かる。

5. おわりに

走行履歴から到着地を推定する場合の困難性を示す指標として、移動先エントロピーを提案した。走行履歴の移動先エントロピーは、所定時間における移動先の不確定さを表現し、走行時間とともに減少する。また、10 名の運転者の走行履歴を取得し、実走行履歴に対する移動先エントロピーを算出する実験を行った。その結果、移動先エントロピーと到着地の推定精度に関しては大きな相関があり、走行開始から 5~10 分で移動先エントロピーの減少および移動先の推定精度の向上が見られた。その後は、移動先エントロピーの減少が顕著には見られないことから、移動先の推定は走行開始から数分で行うことが望ましいことが分かった。また、走行回数の増加とともに移動先エントロピーは増加するが、ある所定値の走行数で収束が見られる。収束する走行数は被験者によって異なるが、それ以上、走行履歴を蓄積しても、到着地の推定精度は向上しないことが分かった。

一方、本稿における上記結果は被験者 10 名によるものであり、必ずしもすべての運転者に該当する傾向とは結論づけられない。さらに今後、被験者数を増加させ、移動先エントロピーの走行時間に対する変化のパターンを分析することにより、走行履歴の分類を行うことが必要である。またこのパターンと運転者のプロフィールとの関連を分析することで、運転者に応じて到着地を推定するための変数を設定することが可能になる。さらに本稿では、車両の走行履歴に付随する情報を考慮せずに移動先エントロピーを定義した。到着地を推定するためには、走行にともなう天候や、出発時刻、走行時刻、乗車人数等の位置情報以外の付帯情報も有用となる。そこで、それらの付帯情報をも考慮した移動先エントロピーを定義することで、さらに実用的な移動先エントロピーを定義することが可能になる。

参 考 文 献

- 1) http://g-book.com/pc/service_menu/top/
- 2) <http://itp.ne.jp/contents/mobile/>

- 3) 佐藤雅春：個客行動を予測する「データマイニング」、日刊工業新聞社 (2001)。
- 4) 柳沢 豊, 赤埴淳一, 小暮 潔：位置センサ情報を用いたユーザの行動傾向の取得, 第 2 回情報処理学会知的都市基盤研究グループ研究発表会講演予稿集, pp.19-26 (2002)。
- 5) 大田智数, 波多野賢治, 吉川正俊, 植村俊亮：ウェアラブル環境における行動履歴を用いた情報提示, 信学技報, DE2001-1, pp.1-7 (2001)。
- 6) 鈴木和宏, 小笠原直人, 佐藤 究, 布川博士：位置情報とスケジューリング情報を用いた人の行動推測に関する研究, 情報処理学会研究報告, 2004-GN-50(6), pp.37-42 (2004)。
- 7) 白川 洋, 歌川由香, 福井良太郎, 重野 寛, 岡田謙一：歩行者ナビゲーションのための歩行履歴情報の分析手法, 情報処理学会研究報告, 2003-MBL-25(10), pp.69-76 (2003)。
- 8) 峰野博史, 肥田一生, 水谷美穂, 宮内直人, 楠和 浩, 水野忠則：移動機器連携ロケーショントラッキング方式の提案, 情報処理学会研究報告, 2005-DBS-135(17), pp.119-126 (2005)。
- 9) Peng, W.-C. and Chen, M.-S.: Developing Data Allocation Schemes by Incremental Mining of User Moving Patterns in a Mobile Computing System, *IEEE Trans. Knowledge and Data Eng.*, Vol.15, No.1, pp.70-85 (2003)。
- 10) Peeta, S. and Yu, J.W.: Adaptability of a Hybrid Route Choice Model to Incorporating Driver Behavior Dynamics Under Information Provision, *IEEE Trans. Systems, Man and Cybernetics-Part A: Systems and Humans*, Vol.34, No.2, pp.243-256 (2004)。
- 11) 工藤貴弘, 小澤 順：目的地予測におけるエントロピーを利用した日時カテゴリ決定方式, 日本知能情報ファジィ学会誌, Vol.16, No.6, pp.551-560 (2004)。
- 12) 小林 誠, 嶋 恵子, 小針里美, 曾布川靖, 伊東幸宏, 酒井三四郎：車載端末向け情報選別配信システムにおける到着地推定機構, 情報処理学会論文誌, Vol.45, No.12, pp.2688-2689 (2004)。
- 13) 長尾光悦, 川村秀憲, 山本雅人, 大内 東：観光動態情報の獲得を意図した GPS ログデータマイニング, 情報処理学会研究報告, 2004-ICS-135(2), pp.7-12 (2004)。
- 14) 金澤明浩, 杵渕哲也, 毛利仁士, 小澤智章, 市河研一, 荒川賢一：決定木を利用した交通渋滞予測手法に関する考察, 情報処理学会研究報告, 2004-ITS-16(21), pp.141-148 (2004)。
- 15) 吉岡元貴, 工藤貴弘, 小澤 順：行動の Frequent Pattern を用いたクラスタリングによる予測手法, 情報処理学会第 66 回全国大会講演論文集, pp.221-224 (2003)。
- 16) Mannila, H., Toivonen, H. and Verkamo, A.: Discovery of Frequent Episodes in Event Se-

quences, *Data Mining and Knowledge Discovery*, Vol.1, pp.259–289 (1997).

- 17) Han, J., Pei, J. and Yin, Y.: Mining Frequent Patterns without Candidate Generation, *ACM-SIGMOD Int. Conf. on Management of Data (SIGMOD'00)*, Dallas, TX (May 2000).

(平成 17 年 4 月 8 日受付)

(平成 17 年 10 月 11 日採録)



吉岡 元貴

昭和 52 年生。平成 15 年東京工業大学大学院総合理工学研究科物理情報システム専攻修了。同年松下電器産業(株)入社。松下電器産業(株)先端技術研究所に勤務。カーインテ

リジェンス, ユーザモデリングの研究に従事。



小澤 順(正会員)

昭和 40 年生。平成 2 年東京工業大学大学院総合理工学研究科システム科学専攻修了。同年松下電器産業(株)入社。平成 5 年より 2 年間国際ファジー工学研究所へ出向。現在,

松下電器産業(株)先端技術研究所に勤務。工学博士。情報検索インタフェース, 個人適応モデル, ネットワークエージェントに関する研究に従事。