

GPS 信号履歴を用いた建物出口特定法の提案

上嶋 祐紀^{1,a)} 内山 彰^{1,2,b)} 山口 弘純^{1,2,c)} 東野 輝夫^{1,2,d)}

概要: 本研究では、GPS 信号の特性を用いて通過した建物の出口を特定する手法を提案する。迅速かつ高精度な出口の特定が可能になれば、自律航法の蓄積誤差修正や GPS 測位開始前に現在位置を提示するなど、様々なサービスへの活用が期待される。提案手法では、ユーザが屋内から屋外に移動する際に信号を捕捉した GPS 衛星を時系列順に記録する。この GPS 履歴間に対して、時空間的な特性を考慮した乖離度を定義することで GPS 履歴群のクラスタリングを行い、出口の特定を実現する。多数のビルが存在する都市環境で実証実験を行った結果、14 秒間の GPS 履歴に対して分類可能率は 45%であり、そのときの調整ランド指数は約 0.9 と非常に高い値を示した。

A Method for Exit Identification Using GPS Signal History

Abstract: In this paper, we propose a method for identifying exits by using GPS signal history. Rapid and accurate exit identification is valuable to many services such as position fix in dead reckoning and GPS. In the proposed method, a time series of acquired GPS signals is recorded when a user in a building moves to the outside through an exit. We define a rate of divergence between a pair of GPS time series based on spatiotemporal characteristics that are dependent on exits. The proposed method performs exit identification by clustering in terms of the divergence rates. We have conducted an experiment at subway exits in an urban area. The result indicates that exit identification is performed at the probability of 0.45 with high Adjusted Rand Index of 0.9 for 14 second GPS time series.

1. はじめに

スマートフォンの普及とともに、GPS を利用した様々な位置情報サービスが提供されるようになった。GPS が利用できない屋内でも測位を可能とするため、無線 LAN [1], [2], [3] や自律航法 [4], [5], [6] を用いた位置推定手法も多数提案されている。しかし、これらの測位手法はいずれも状況に応じて誤差が大きくなる。例えば、GPS の測位誤差は測位に利用する衛星の幾何的な配置に依存するため、都市部などでは特定の方向がビルに遮られ、誤差が大きくなりやすい。無線 LAN を用いた方式も、基地局が適切に配置されていないと測位誤差は大きくなる。したがって、状況を

問わず位置情報サービスを提供するためには、状況に応じて様々な測位手法や補助情報を併用し、測位精度を向上させることが重要である。

そのような補助情報の一つとして、ユーザが屋内から屋外に移動する際にどの出口から屋外に移動したのか、という情報（**出口情報**）がある。出口情報が分かれば、自律航法の蓄積誤差を修正したり、GPS 測位開始前に迅速に現在位置を提示するなどの活用が可能である。現状では出口情報を得る方法としていくつかの方法が考えられるが、いずれもコストや精度の面で問題がある。例えば RFID タグ [7] や Place Sticker [8] のように数 m 程度の至近距離に端末が存在する場合のみ、電波が受信可能となる機器を出口に設置すれば、高精度に出口情報を得られる。しかし、多数の機器を設置する必要があり導入・維持コストがかかる。また、GPS や WiFi、自律航法などの測位結果から出口情報を得るには測位時間がかかったり、測位できても状況によっては誤差が大きいという問題がある。

そこで本研究では、迅速かつ高精度な出口情報の取得を目標に、屋内から屋外へ移動する際の GPS 信号の時空間的な特性に着目した方式を提案する。提案手法は地点ごと

¹ 大阪大学大学院情報科学研究科
Graduate School of Information Science & Technology,
Osaka University,
1-5 Yamadaoka, Suita, Osaka 565-0871, Japan

² 独立行政法人科学技術振興機構, CREST
Japan Science Technology and Agency, CREST

a) y-uejima@ist.osaka-u.ac.jp

b) utiyama@ist.osaka-u.ac.jp

c) h-yamagu@ist.osaka-u.ac.jp

d) higashino@ist.osaka-u.ac.jp

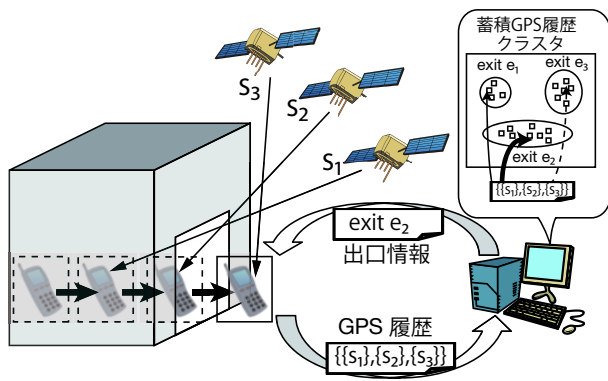


図 1 提案手法の概要

の特徴量としてこれまでに無い GPS 信号の時空間的な類似性に着目し、乖離度を定義している点に新規性がある。本研究では、提案手法による出口判定の精度を確認するため様々な出口において GPS 信号特性を記録し、定義した乖離度に基づきクラスタ解析を行った。

GPS 信号は壁などの障害物により遮蔽されるため、屋内では受信不可能な場合がほとんどである。屋内から屋外へ移動する場合、出口に近づくにつれて壁などの遮蔽物が無くなり、GPS 信号を徐々に受信できるようになる。このとき、出口ごとに空が開けている方向が異なるため、受信可能になる GPS 衛星の順序は出口ごとの特徴量として定量化できると考えられる。提案手法ではこの特徴量を **GPS 履歴** と呼び、GPS 履歴間の類似性を **履歴乖離度** として定義する。履歴乖離度は 2 つの GPS 履歴を等しくするために必要な衛星の入れ替え回数によって算出される。

図 1 に提案手法の概要を示す。まずユーザは GPS 信号の取得が開始されてから一定期間の GPS 履歴を取得し、サーバに送信する。サーバには過去にユーザから送信された GPS 履歴が蓄積されており、送信された GPS 履歴との乖離度が最も小さい出口をユーザに返信する。各出口の GPS 履歴の構築は、クラウドソーシングによるデータ収集と機械学習を組み合わせる。判定方法と同様に、まずユーザは GPS 履歴をサーバに送信する。サーバでは蓄積されている全 GPS 履歴に対して履歴乖離度に基づくクラスタリングを行う。その結果得られた各クラスタにはラベル付けがされていないため、そのままでは出口情報を得られない。そこでユーザによるラベル付けを促し、ごく一部のユーザがラベル付けを行えば、以降、それらのラベル付けされた GPS 履歴を基に出口情報をユーザにフィードバックできるようにする。

性能評価のため、大阪大学大学院情報科学研究科の研究棟に存在する 3 つの出口において GPS 履歴を収集し、パラメータの影響を調べた。さらに、大阪市内の地下鉄駅の 9 つの出口で実証実験を行った。その結果、15 秒間の GPS 履歴に対してクラスタリング精度を示す調整ランド指数は約 0.4 であった。また、14 秒間の GPS 履歴に対して分類

可能率は 45% であり、そのときの調整ランド指数は約 0.9 と非常に高い値を示した。以上の結果より、提案する履歴乖離度によって出口ごとに異なる特徴量をとらえられることが分かった。

2. 関連研究

2.1 GPS とその関連技術

GPS 測位では衛星からの距離に基づき測位を行うため、衛星の軌道情報が必要である。軌道情報は GPS からの信号に含まれているが、全ての軌道情報を得るためには約 12 分の時間が必要となる。したがって、通常の GPS では最初の測位に時間がかかるという問題がある。この問題を解決するために A-GPS (Assisted GPS)[9] が考案された。A-GPS は多くの携帯電話でも採用されている方式であり [10]、全衛星の軌道情報をサーバに蓄積しておき、受信機は携帯通信網などを介してサーバから軌道情報を取得するため、迅速に測位を開始することができる。また、精度向上を目的とした D-GPS (Differential GPS)[11] も利用されている。D-GPS では電離層などの大気の影響によって生じる GPS 信号の伝搬遅延は、近隣の受信機で同じように現れるという特徴を利用する。地理的に分散配置された基地局など正確な位置が分かる地点で GPS 信号の伝搬遅延を推定し、近隣の GPS 受信機が測位を行う際に推定された伝搬遅延情報を用いることで精度向上を図る。上記の GPS に関連する技術は測位精度向上を目的とするものである。これらの技術を用いて本研究の目的である迅速な出口判定を実現するためには、出口付近あるいは出口から屋外に出て数秒以内に高精度な測位が行えなければならない。しかし、建物の出口付近では建物が必ず一方に存在するため、直接衛星が見える方向が偏り、測位精度が低下しやすい。

日本のみちびき [12] や欧州の Galileo、中国の北斗など、利用可能な衛星数を増やすことで都市部における測位精度を向上する研究開発も行われている。また、屋内 GPS (IMES) と呼ばれるシステムの開発も進んでいる [13]。IMES では屋内や地下街にあらかじめモジュールを設置し、それらのモジュールが発信する部屋などの位置情報から位置を特定する。発信する信号は GPS と同様のデータ構造のため、携帯端末の改良が最小ですむことから屋内位置推定への利用が期待されている。これらのインフラ設備を増強する取り組みは大きな効果が期待されるが、既存の GPS 受信機では利用できず、新たな GPS 受信機の導入やインフラ整備のコストがかかるため、普及には時間がかかる。

2.2 屋内向け位置推定法

屋内向け位置推定法を利用して出口判定を実現することも可能である。特に無線 LAN を用いた手法は、電波強度を用いた方式 [1], [2] や、電波強度の地点ごとの特性をフィン

ガープリントとして利用する方式 [3] が多数提案されている。無線 LAN は多くの家庭やオフィスなどで普及が進んでいるため、Place Engine[14] や Skyhook technology[15] など、実際に利用されているサービスも多い。これらの方式では、平均誤差数 m の位置推定精度を達成できることが分かっているが、基地局の配置や壁などの影響により、場所によっては精度が低下してしまう。RFID タグ [7] や Place Sticker[8] のように、近距離無線機器を各出口に設置し、出口情報を発信する方式も存在する。これらは高い精度を実現できるが、単一機器のサービス範囲が狭いため多数の機器を設置する必要があり導入・維持コストがかかる。電子コンパスや加速度センサを利用して移動方向と歩数を取得し、移動軌跡を推定する自律航法 [4], [5], [6] により高精度に現在位置を推定し、出口を判定することも可能である。しかし、自律航法では移動方向や移動距離の推定誤差が時間経過とともに蓄積するという問題がある。また、電子コンパスは磁場の影響を受けるため、電子機器の多い屋内では特に移動方向の推定精度が悪化する可能性が高い。

2.3 場所依存の特徴量に基づく方式

場所依存の特徴量に基づき、特定地点の存在を検知する方式も存在する。前述の無線 LAN フィンガープリントに基づく位置推定法もこのカテゴリに属する。文献 [16] では、扉の通過により無線 LAN の信号強度が大きく変化することを利用した方式を考案している。UnLoc[17] では、エレベーターやエスカレーター、扉などにおいて現れる加速度、磁場、無線 LAN の電波強度の特徴量を検知し、自律航法における推定位置の補正に利用する。提案手法はこれらのアプローチと同様に場所依存の特徴量に基づく方式であり、GPS 信号の時空間的な特徴量を利用するという点に新規性がある。

3. 提案手法

3.1 提案手法の概要

提案手法ではユーザ i が保持するスマートフォンを用いて、GPS 信号が一切受信できない（屋内に存在する）状態において定期的に GPS 信号の受信状態を確認する。GPS 信号の受信を開始してから T 秒経過後、記録した GPS 履歴 $hist_i$ をサーバに送信する。サーバには蓄積された GPS 履歴集合 \mathcal{D} が存在し、新たに送信された $hist_i$ を加えた $\mathcal{D} \leftarrow \mathcal{D} \cup hist_i$ に更新される。次に全ての GPS 履歴 $hist \in \mathcal{D}$ に対してクラスタリングを行い、新たに加えられた GPS 履歴 $hist_i$ が属するクラス $C(i)$ を決定する。クラスタリングに必要な GPS 履歴間の非類似度には、後述する履歴乖離度を用いる。

ここで、GPS 履歴に出口情報が付与されているデータをラベル付き GPS 履歴と呼ぶ。 $C(i)$ に属するラベル付き GPS 履歴 $hist_1, hist_2, \dots, hist_k$ のラベルを

$e(hist_1), e(hist_2), \dots, e(hist_k)$ とする。このとき、 $C(i)$ のラベル $e(C(i))$ は $\cup_{j=1}^k e(hist_j)$ で定義され、ユーザ i には出口情報 $e(C(i))$ がサーバから返信される。 $e(C(i)) = \phi$ の場合、ユーザ i に出口情報の入力を促し、その後の出口判定に利用可能なラベル付き GPS 履歴とする。全ユーザが出口情報を入力することは考えにくい、一部のユーザが出口情報を入力すれば、 $e(C(i)) = \phi$ となる確率は低くなる。このようにして、十分な GPS 履歴ならびにラベル付き GPS 履歴がサーバに蓄積されれば、以降はユーザによる入力無しでの出口判定が可能となる。

3.2 GPS 履歴

1 つ以上の衛星から信号を受信開始した時刻を 0 とし、時刻 T 秒までの時間を 1 秒単位の T 個のタイムスロットに分割する。先頭から t 番目のタイムスロットには、時刻 $[t-1, t)$ の期間に受信を開始した衛星集合 $TS(t)$ が記録される。 $TS(t)$ をタイムスロット順に列挙したものが GPS 履歴

$$hist = \{TS(1), TS(2), \dots, TS(T)\}$$

である。

例えば、端末 i が屋外に移動する際、時刻 5, 7, 8, 8, 10, 15 に衛星 s_1, s_2, \dots, s_6 からの信号をそれぞれ受信開始したとする。このときの $T=4$ の GPS 履歴 $hist_i$ は、時刻 5 を開始スロット $TS_i(1)$ として、

$$hist = \{\{s_1\}, \phi, \{s_2\}, \{s_3, s_4\}, \phi, \}$$

となる。 s_5 および s_6 は T 秒以降に現れるため、 $hist$ には含まれない。

3.3 履歴乖離度

$hist_i$ が記録された期間で仰角が 0 度より大きい衛星の集合を全衛星集合と呼び $SAT(i)$ で表す。今、GPS 履歴 $hist_i$ および $hist_j$ が全衛星の配置が変わらない程度の同一期間に取得したものであると仮定する。このとき $hist_i, hist_j$ の記録位置が大きく離れていなければ、 $SAT(i) = SAT(j)$ である。したがって本節では全衛星集合 $SAT(i), SAT(j)$ を単純に SAT と表記する。ここで、GPS 履歴 $hist_i$ に現れる全ての衛星集合を $HIST_i = \cup_{t=1}^T TS_i(t)$ とする。まず足りない GPS 衛星を補完し、GPS 履歴 $hist_i$ の大きさ $|HIST_i|$ を $|SAT|$ にそろえる。GPS 履歴 $hist_i$ の補完では、衛星 $s \in SAT - HIST_i$ は全て $T+1$ 秒に現れるものとして $hist_i$ の最後に付け加える。すなわち、補完後の履歴を \hat{hist}_i とすると、

$$\hat{hist}_i = \{TS_i(1), TS_i(2), \dots, TS_i(T), \{SAT - HIST_i\}\}$$

である。同様に $hist_j$ も補完し、 \hat{hist}_j を得る。

履歴乖離度は 2 つの GPS 履歴を同一の GPS 履歴にする

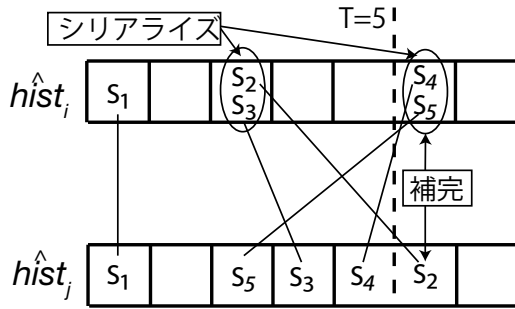


図 2 同一衛星配置での履歴乖離度の計算例

ために必要な衛星の入れ替え回数で定義される。しかし、同一のタイムスロットに複数の衛星が存在する場合、そのままでは入れ替え回数が定義できない。そこで、提案手法では \hat{hist}_i に対してシリアライズを行い、先頭から順に空白の無い GPS 衛星の PRN 列にする。このとき空白を無視する理由は、同一の出口でも移動速度や信号の揺らぎによって衛星の出現タイミングが前後するためである。以降、補完後の GPS 履歴 \hat{hist}_i に対してシリアライズを行った後の GPS 衛星の PRN 列を **GPS 系列** と呼び、 seq_i と表記する。

同一タイムスロットに存在する衛星は同時に到着したと見なせるので、シリアライズの際に任意の順序づけを行って良いものと考えられる。そこで、入れ替え回数の増加が最小になるようにシリアライズを行う。すなわち、 \hat{hist}_i のシリアライズでは、2つ以上の衛星を含む全てのタイムスロット $T_i(t)$ において、乖離度の計算対象である \hat{hist}_j に出現する順番に $T_i(t)$ に含まれる衛星を順序づけすることで seq_i を構成する。同様に \hat{hist}_j もシリアライズする。

以降では衛星 s が GPS 系列 seq_i の先頭から k 番目に現れることを $rank(s, i) = k$ で表し、**順位** と呼ぶ。GPS 履歴 $hist_i$ と $hist_j$ の非類似度 (距離) を **履歴乖離度** と呼び、 $d_{hist}(i, j)$ と表記する。 $d_{hist}(i, j)$ は $hist_i, hist_j$ から算出される GPS 系列 seq_i, seq_j の各衛星の順位を用いて以下の式で定義される。

$$d_{hist}(i, j) = \frac{1}{2} \sum_{s \in SAT} |rank(s, i) - rank(s, j)| \quad (1)$$

履歴乖離度の計算例を図 2 に示す。まず、GPS 履歴 $hist_i$ に対して $\{s_4, s_5\}$ を $TS_i(T+1)$ に補完する。 $hist_j$ には $\{s_2\}$ を $TS_j(T+1)$ に補完する。次にシリアライズを行い、GPS 系列

$$seq_i = \{s_1, s_3, s_2, s_5, s_4\}$$

$$seq_j = \{s_1, s_5, s_3, s_4, s_2\}$$

を得る。したがって履歴乖離度は 3 となる。

3.4 異なる衛星配置を考慮した履歴乖離度

式 (1) では、衛星の配置が $hist_i, hist_j$ 間で同一であると

いう仮定の下、履歴乖離度を定義した。しかし実際には衛星配置が時刻によって変化するため、衛星の PRN に基づき GPS 履歴間での入れ替え回数を求める前述の方式では、異なる時刻の GPS 履歴間の乖離度は求められない。そこで提案手法では、衛星配置が異なっても空間的に近い衛星があれば、それらの衛星が履歴中に現れる順序も近いと仮定し、異なる衛星配置でも履歴乖離度を算出できるようにする。

まず、3.3 節で述べた GPS 履歴の補完と同様に、 $hist_i$ に $SAT(i) - HIST_i$ を補完し、 \hat{hist}_i を得る。同様に $hist_j$ も補完を行う。次に、 $\hat{hist}_i, \hat{hist}_j$ 間での衛星の対応付けを行う。対応付けには天球座標で衛星 s と r がなす角度 $\theta(s, r)$ を用いる。 $SAT(i)$ の衛星 s に対応する $SAT(j)$ の衛星集合 $CSAT_i(s)$ を以下で定義する。

$$CSAT_i(s) = \{r | r \in SAT(j) \wedge \theta(s, r) < TH_{deg}\}$$

$CSAT_i(s) = \phi$ の場合は対応づける衛星が存在しないため、衛星 s は \hat{hist}_i から除外される。一方、 $CSAT_i(s)$ に属する衛星が 2 つ以上存在する場合は、衛星 $s \in hist_i$ に対応する疑似衛星 \hat{s} を用意し、 $s \in hist_i$ を $\hat{s} \in hist_j$ に対応づけることで、単一の衛星への対応付けに変換する。 \hat{s} を追加すべきスロット $TS_j(\hat{s})$ は $CSAT_i(s)$ に属する衛星の平均スロットとして定義する。すなわち、 $TS_j(\hat{s})$ は衛星 r が含まれる $hist_j$ のスロット番号を表す記号 $TS_j^{-1}(r)$ を用いて、

$$TS_j(\hat{s}) = \lfloor \sum_{r \in CSAT_i(s)} \frac{TS_j^{-1}(r)}{|CSAT_i(s)|} \rfloor$$

と表せる。

以上の操作を全衛星 $s \in SAT(i)$ について行い、 \hat{hist}_i から \hat{hist}_j に対して、単一の対応付けがなされた GPS 履歴 $hist_{i \rightarrow j}$ を構成する。対応付けがなされた GPS 衛星 $s \in HIST_i$ は、 $HIST_{i \rightarrow j}$ において $\hat{hist}_i, \hat{hist}_{i \rightarrow j}$ の双方で同一かつ固有の識別子 s で表されるものとする。 \hat{hist}_i に含まれる衛星 s は $\hat{hist}_{i \rightarrow j}$ に含まれる単一の衛星 s に必ず対応付けされているため、3.3 節で述べた方法と同様にシリアライズを行い、履歴乖離度を算出可能である。なお、履歴に現れる衛星数 $|HIST_{i \rightarrow j}|$ が計算可能衛星数 TH_{div} 未満の場合は、GPS 履歴の比較に十分な特徴量が得られないものとし、履歴乖離度は未定義とする。

3.3 節で述べた同一衛星配置での履歴乖離度と異なり、対応付けの方向によって異なる結果が得られる可能性がある。そこで一般的な距離の定義に基づき対応付けを双方で行い、それぞれの履歴乖離度を算出した結果、最小のもの $hist_i, hist_j$ 間の履歴乖離度として以下のように再定義する。

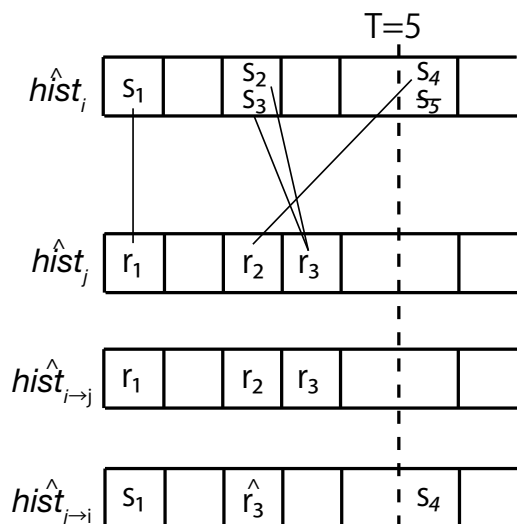


図 3 衛星配置の異なる履歴間での乖離度の計算例

$$d_{hist}(i, j) = \min \left\{ \frac{1}{2} \sum_{s \in HIST_i} |rank(s, i) - rank(s, j)|, \frac{1}{2} \sum_{r \in HIST_j} |rank(r, i) - rank(r, j)| \right\} \quad (2)$$

図 3 に衛星配置が異なる場合の履歴乖離度の計算例を示す。

$$SAT(i) = \{s_1, s_2, \dots, s_5\}, SAT(j) = \{r_1, r_2, r_3\}$$

であり、衛星間の角度が閾値 TH_{deg} より小さい衛星ペアは図 3 中の線分で結ばれている。 $CSAT_i(s_5) = \phi$ であるため、 s_5 は $hist_i$ から除外される。まず、 $hist_i$ から $hist_j$ の履歴乖離度を計算する。 $HIST_i$ に含まれる全ての衛星は単一の衛星 $r \in HIST_{i \rightarrow j}$ に対応づけられる。 $hist_i, hist_{i \rightarrow j}$ をそれぞれシリアライズした結果、

$$seq_i = \{s_1, s_2, s_3, s_4\}$$

$$seq_{i \rightarrow j} = \{r_1, r_2, r_3, r_3\} = \{s_1, s_4, s_2, s_3\}$$

となり、 $hist_i$ から $hist_j$ の履歴乖離度は 2 となる。次に、 $hist_j$ から $hist_i$ の履歴乖離度を計算する。同様にして $HIST_j$ に含まれる全ての衛星を単一の衛星 $s \in HIST_{j \rightarrow i}$ に対応づける。その際、 $CSAT_j(r_3) = \{s_2, s_3\}$ であるため、疑似衛星 \hat{r}_3 に対する平均スロットを計算し、

$$TS_i(\hat{r}_3) = \left\lfloor \frac{TS_i^{-1}(s_2) + TS_i^{-1}(s_3)}{2} \right\rfloor = 3$$

を得る。この結果、

$$seq_{j \rightarrow i} = \{s_1, \hat{r}_3, s_4\} = \{r_1, r_3, r_2\}$$

$$seq_j = \{r_1, r_2, r_3\}$$

となる。したがって、 $hist_j$ から $hist_i$ の履歴乖離度は 1 となる。以上より、 $hist_i, hist_j$ 間の履歴乖離度は

$$d_{hist}(i, j) = \min\{2, 1\} = 1$$

である。

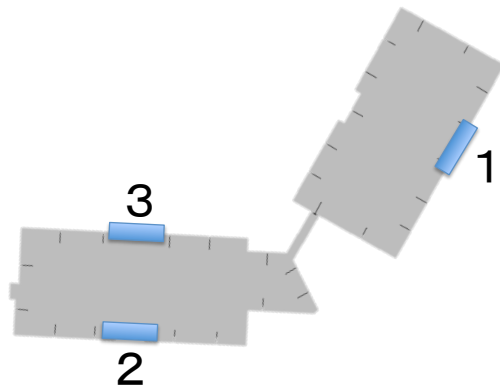


図 4 実験場所

3.5 GPS 履歴のクラスタリング

サーバに蓄積された全 GPS 履歴 \mathcal{D} に対してクラスタリングを行い、各 GPS 履歴が属するクラスタを決定する。クラスタリング手法には様々な方式が存在するが、本稿では GPS 履歴を特徴量として出口判定が可能なことを確認するため、階層的な手法 [18] により行うこととした。また、それぞれのクラスタリングにおいて指定したクラスタ数 C_k が実現されるよう、クラスタ間距離閾値を決定した。クラスタ間距離は代表的な最長距離法、群平均法の 2 種類を用いる。

クラスタ C_1, C_2 間のクラスタ間距離を計算する際には、全クラスタメンバ組 ($x_1 \in C_1, x_2 \in C_2$) 間の距離計算が必要となる。しかし、GPS 履歴間の距離は履歴に現れる衛星数が計算可能衛星数より少ない場合に未定義となるため、そのままではクラスタ間距離が定義できない場合がある。この理由は、衛星の対応付け角度閾値 TH_{deg} に対して計算可能衛星数 TH_{div} が短い場合に、GPS 履歴から得られる特徴量が不十分になってしまうためである。

この問題を解決するため、本稿では距離の定義が可能な全クラスタメンバ組間に対して算出したクラスタ間距離を計算に用いる。同一クラスタメンバ内で距離が定義可能なメンバ間に辺を定義すると、同一クラスタメンバ x_1, x'_1 間には必ず経路が存在する。したがって、クラスタ C_1, C_2 にそれぞれ属する x_1, x_2 について、直接距離 $d(x_1, x_2)$ が定義できない場合でも、他のクラスタメンバ $x'_1 \in C_1$ との距離 $d(x'_1, x_2)$ が定義可能であれば、 $d(x'_1, x_2)$ と $d(x_1, x_2)$ には正の相関があると見なしクラスタ間距離を定義する。もし直接距離が定義可能なクラスタメンバ組が存在しない場合は、クラスタ間距離 $d(C_1, C_2) = \infty$ とする。なお、GPS 履歴 $hist$ と全ての GPS 履歴との距離が定義できない場合、 $hist$ は分類不能となる。

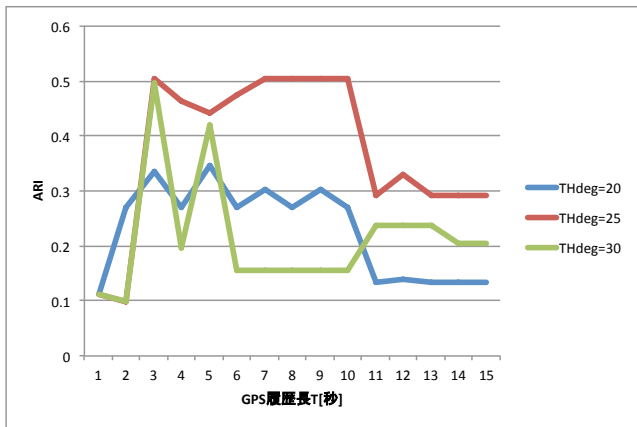


図 5 GPS 履歴長 T に対するクラスタリング精度

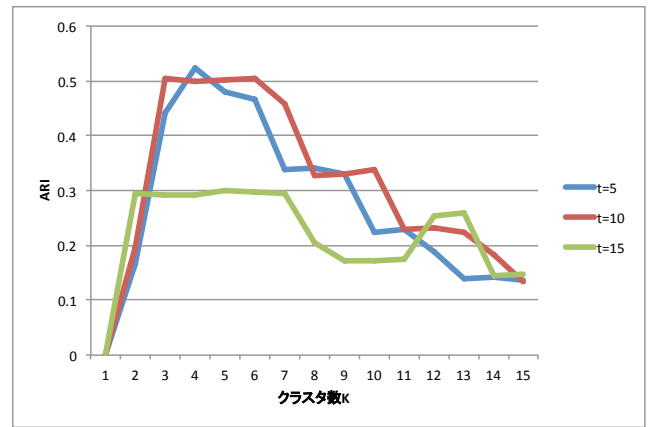


図 6 クラスタ分割数 K に対するクラスタリング精度

4. 性能評価

4.1 評価環境

性能評価のため、Android 4.1.1 を搭載した Nexus S を用いて実験を行った。まず、パラメータの影響を調べるため大阪大学情報系総合研究棟でデータを収集した。実験は図 4 に示す出口 1, 2, 3 でそれぞれ 10 回, 5 回, 5 回ずつ屋内から屋外へ移動し、GPS 履歴を記録した。同じ実験を異なる日にもう一度行い、クラスタリング精度を調べた。クラスタリング精度の指標としてここでは調整ランド指数 (ARI: Adjusted Rand Index)[19] を用いた。ARI はクラスタリング精度を測るための 1 つの指標で、クラスタリング結果の各ペアについて、ラベルが同一のペアが同一のクラスターに属し、ラベルが異なるペアが異なるクラスターに属している割合を示している。ARI は $[-1, 1]$ の値をとり、ランダムにクラスター分類を行った際の分類精度を 0 として、相対的にクラスタリングの精度を表す指標である。パラメータは明示されない限り時間 $t = 5, 10, 15$ 秒、クラスタ数 $K = 3$ 、計算可能衛星数 $TH_{div} = 3$ 、衛星の対応付け角度閾値 $TH_{deg} = 25^\circ$ を用いている。

4.2 パラメータの影響

図 5 は GPS 履歴長 T に対するクラスタリング精度を表している。衛星の対応付け角度閾値 TH_{deg} が 30° の場合は、角度閾値が大きすぎて信号特性が似ていない衛星との対応付けを行ってしまう場合があるため、精度が大きく変動する場合がある。一方、 TH_{deg} が $20^\circ, 25^\circ$ のときは同様の傾向を示しており、 $T = 1$ から $T = 3$ 秒にかけて精度が上昇し、最大になった後、 $T = 11$ 以降、急激に精度が低下している。この理由は、 T が短すぎると観測される衛星の系列数が少なく、出口ごとの特徴量を十分に得られないためである。一方、 T が長すぎても、出口から離れていくため出口ごとの特徴量が薄くなることから、精度が低下していると考えられる。

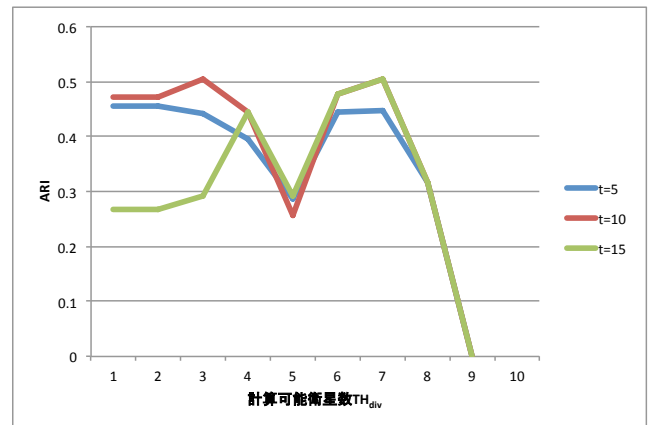


図 7 計算可能衛星数 TH_{div} に対する精度

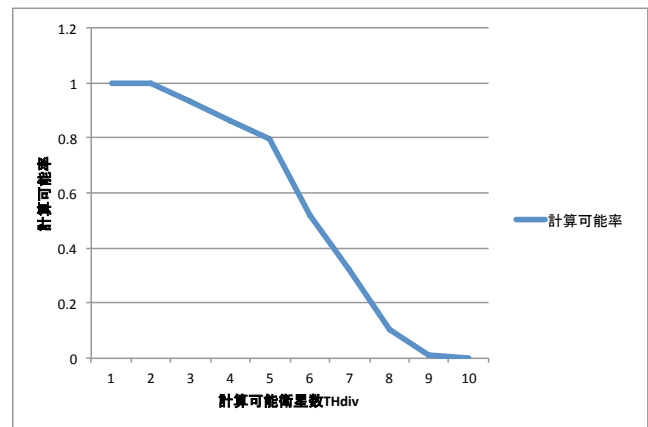


図 8 計算可能衛星数 TH_{div} に対する距離定義可能率

図 6 はクラスタリングにおけるクラスタ分割数 K を変化させたときの精度を示している。いずれの場合もほぼ $K = 3$ の時に最も良い精度を示しており、この結果は実際の出口数である 3 とも一致している。 $K > 3$ の場合でも、比較的高い精度が得られているが、 K が大きすぎると同一の出口であっても強制的に異なるクラスターに分割されてしまうため、精度が低下している。

図 7 は GPS 履歴同士の距離計算において、その距離を定義するために必要な計算可能衛星数 TH_{div} を変化させ

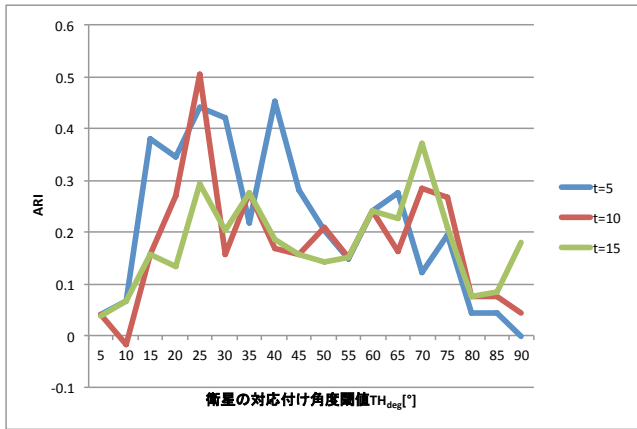


図 9 衛星の対応付け角度閾値 TH_{deg} とクラスタリング精度

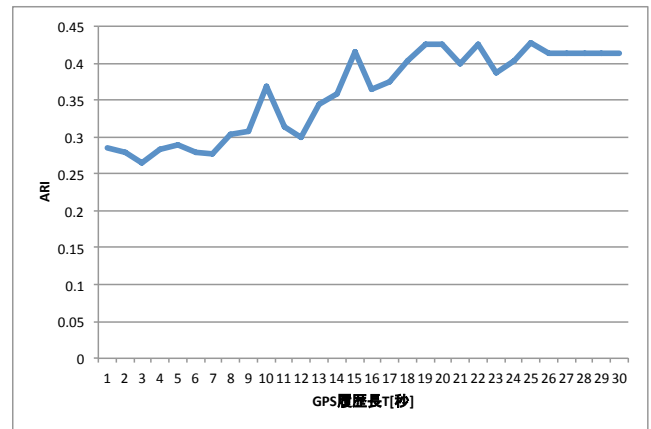


図 11 肥後橋駅におけるクラスタリング精度

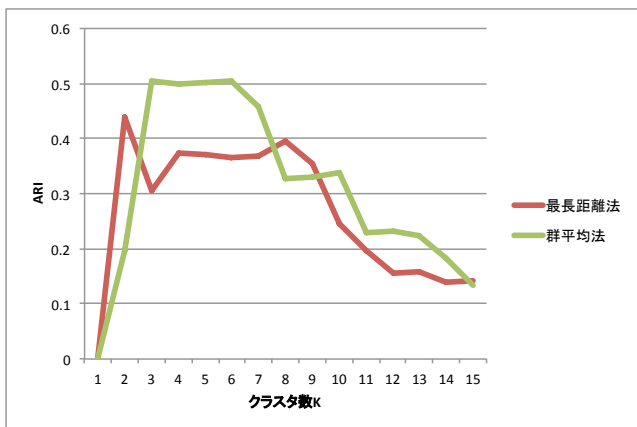


図 10 最長距離法と群平均法のクラスタリング精度比較

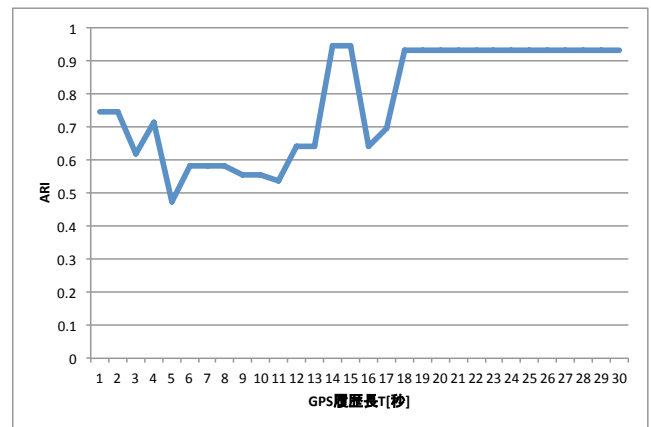


図 12 肥後橋駅における最大クラスタリング精度

たときのクラスタリング精度を示している。 $TH_{div} = 5$ の場合のみ大幅に精度が低下している理由は、偶然特定の出口での全衛星集合が 5 個であり、 $TH_{div} = 5$ の時点でこの出口で取得したデータセットに関してクラスタ分類ができなくなるためである。図 8 は計算可能衛星数 TH_{div} に対する距離定義が可能であった GPS 履歴組の割合を示している。図 8 の $TH_{div} = 5$ の場合を見ると距離定義可能率が 0.8 に低下しており、これに連動してクラスタリング精度が低下したものと考えられる。

図 9 は衛星の対応付け角度閾値 TH_{deg} を変化させたときのクラスタリング精度を示している。データセットには同一の出口でも異なる時間のものが存在している。したがって、角度閾値 TH_{deg} が小さいときは時間が異なるデータセット間で衛星の対応付けができないため精度が低くなっている。 TH_{deg} が大きくなるにつれて異なる時間のデータセット間での衛星対応付けが可能となり、徐々に精度は上昇する。一方、 TH_{deg} が大きすぎる場合には GPS 信号の特性が異なる衛星間での対応付けがなされてしまうため、精度はランダムに近い不安定な値を示していると考えられる。

図 10 はクラスタ間距離の定義に最長距離法と群平均法を用いた場合のクラスタリング精度を示している。クラス

タ分割数 K の変化に合わせてどちらも似た傾向を示しているが、多くの場合で群平均法がわずかに良い精度を達成している。最長距離法では 2 つのクラスタ C_1, C_2 にそれぞれ属するメンバ x_1, x_2 のうち、 $d(x_1, x_2)$ が最大のものをクラスタ間距離として定義する。したがって、最長距離法がわずかに精度で劣る理由は、同一の出口でも衛星配置が異なる GPS 履歴同士の場合など、距離が遠くなりやすい場合に同一のクラスタにできない場合が存在するためと考えられる。

4.3 都市部における実証実験

次に大阪市営地下鉄の肥後橋駅の出口にて実験を行った。実験は肥後橋駅の出口の内 9 箇所においてそれぞれ 10 回屋内から屋外に出ることで GPS 履歴を記録した。このデータに対してクラスタ分割数 $K = 9$ 、計算可能衛星数 $TH_{div} = 2$ 、衛星の対応付け角度閾値 $TH_{deg} = 30^\circ$ で評価すると 15 秒で ARI が 0.4 を超えることがわかった (図 11)。さらに、制約を厳しくし計算可能衛星数 $TH_{div} = 9$ 、衛星の対応付け角度閾値 $TH_{deg} = 5^\circ$ にするとクラスタリングを行える数が 90 個中 41 個となるがクラスタリング判定精度は 14 秒で 0.94 と非常に高い値を示した (図 12)。以上の結果より、提案する履歴乖離度によって出口ごとに異な

る特徴量をとらえられることが分かった。

5. まとめ

本稿では、ユーザが屋内から屋外へ移動する際に迅速かつ高精度な出口判定を実現するため、GPS信号の特性を利用する手法を提案した。出口付近においては出口から見通せる衛星の状況が変化する。したがって、出口ごとの特徴量は信号を捕捉したGPS衛星の時系列として現れる。提案手法ではGPS系列間に対して乖離度を定義し、クラスタリングを行うことにより出口判定を実現する。都市部の地下鉄出口9個を対象に実験を行った結果、15秒間のGPS履歴に対してクラスタリング精度を示す調整ラウンド指数は約0.4であった。

今後の課題として、GPS信号の不安定性を考慮することが挙げられる。このため、衛星によって入れ替えに重み付けを行うなどの方法を検討している。クラスタリングの方法やラベル付けの方法についても詳細な検討を進める予定である。また、より多くのデータを収集し、多様なデータや蓄積データ量に対する精度を評価する必要があると考えている。

参考文献

- [1] Patwari, N. and Hero, A. O.: Using proximity and quantized RSS for sensor localization in wireless networks, *Proceedings of ACM International Conference on Wireless Sensor Networks and Applications (WSNA)*, pp. 20–29 (2003).
- [2] Lim, H., Kung, L.-C., Hou, J. C. and Luo, H.: Zero-configuration indoor localization over IEEE 802.11 wireless infrastructure, *Wireless Networks*, Vol. 16, No. 2, pp. 405–420 (2010).
- [3] Bahl, P. and Padmanabhan, V. N.: RADAR: An in-building RF-based user location and tracking system, *Proceedings of Annual Joint Conference of the IEEE Computer and Communications Societies (Infocom)*, pp. 775–784 (2000).
- [4] Agarwal, A. and Das, S. R.: Dead Reckoning in Mobile Ad Hoc Networks, *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1838–1843 (2003).
- [5] Jin, Y., Motani, M., Soh, W.-S. and Zhang, J.: Sparse-Track: Enhancing Indoor Pedestrian Tracking with Sparse Infrastructure Support, *Proceedings of Annual Joint Conference of the IEEE Computer and Communications Societies (Infocom)*, pp. 1–9 (2010).
- [6] Steinhoff, U. and Schiele, B.: Dead reckoning from the pocket - An experimental study, *Proceedings of IEEE International Conference on Pervasive Computing and Communications (Percom)*, pp. 162–170 (2010).
- [7] Kouroggi, M., Sakata, N., Okuma, T. and Kurata, T.: Indoor/Outdoor Pedestrian Navigation with an Embedded GPS/RFID/Self-contained Sensor System, *Proceedings of International Conference on Artificial Reality and Telexistence*, pp. 1310–1321 (2006).
- [8] Nishio, N., Oka, R., Duc, T. X. and Arai, I.: Place Sticker: Energy Harvesting Approach for Low Cost Wi-Fi AP Positioning Infrastructure, *Proceedings of International Conference on Pervasive Computing (Pervasive)* (2011).
- [9] Djuknic, G. M. and Richton, R. E.: Geolocation and Assisted GPS, *IEEE Computer*, Vol. 2, No. 2, pp. 123–125 (2001).
- [10] KDDI: Technology Point — 研究開発 (R&D) — KDDI 株式会社, KDDI (オンライン), 入手先 (http://www.kddi.com/corporate/r_and_d/tech/question/2.html) (参照 2012-2-2).
- [11] Kojiroi, R.: わかりやすいGPS測量, Ohmsha (2010).
- [12] Kawano, I., Mokuno, M., Kogure, S. and Kishimoto, M.: Japanese Experimental GPS Augmentation using Quasi-Zenith Satellite System (QZSS), *Proceedings of International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS)*, pp. 175–181 (2004).
- [13] Kohtake, N., Morimoto, S., Kogure, S. and Manandhar, D.: Indoor and Outdoor Seamless Positioning using Indoor Messaging System and GPS, *Proceedings of International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 21–23 (2011).
- [14] Rekimoto, J., Miyaki, T. and Ishizawa, T.: LifeTag: WiFi-based Continuous Location Logging for Life Pattern Analysis, *Proceedings of International Symposium on Location- and Context-Awareness*, pp. 35–49 (2007).
- [15] Skyhook, I.: Skyhook Wireless, Skyhook, Inc. (online), available from (<http://www.skyhookwireless.com/>) (accessed 2012-10-18).
- [16] 梶克彦, 河口信夫: 無線LAN環境特異点に基づくゲート通過検出手法, マルチメディア, 分散, 協調とモバイル (DICOMO2012) シンポジウム, pp. 340–348 (2012).
- [17] Wang, H., Sen, S., Elgohary, A., Farid, M., Youssef, M. and Choudhury, R. R.: No need to war-drive: unsupervised indoor localization, *Proceedings of International Conference on Mobile Systems, Applications, and Services (Mobisys)*, pp. 197–210 (2012).
- [18] Duda, R. O., Hart, P. E. and Stork, D. G.: *Pattern Classification*, Wiley-Interscience, 2nd edition (2000).
- [19] Yeung, K. Y. and Ruzzo, W. L.: Details of the Adjusted Rand index and Clustering algorithms Supplement to the paper “An empirical study on Principal Component Analysis for clustering gene expression data”, *Bioinformatics*, Vol. 17, No. 9, pp. 763–774 (2001).