

GPS ログにおける欠損とノイズの分析手法

北沢 匠¹ 坂本 大輔² 安積 卓也³ 望月 祐洋⁴ 西尾 信彦²

概要: 近年、携帯端末の高機能化に伴い、ユーザやその周りの環境に関するライフログの常時取得が可能になりつつあり、中でも特に GPS レシーバから得られる位置情報は様々なサービスに広く利用されている。この位置情報は常に安定して取得できるわけではなく、環境や、端末の不具合などによって欠損や精度の低下が起き、サービスにおける処理精度の低下要因になってしまう可能性がある。本研究ではこの問題に対処するため、取得されたログ中に存在する不完全な GPS ログを分析し、それぞれの特徴に合わせて類別する手法を提案する。類別は GPS ログと歩数情報を併用して段階的に行う。提案手法を用いて類別を行うことで、GPS ログを用いたサービスやシステムにおいて、ログの特性を正確に把握したり、不完全なログの占める割合によって利用するログの取捨選択をしたりできるため、ユーザの行動認識をより正確に行うことが可能となる。提案手法を用いて、三週間分の GPS ログに対して不完全な GPS ログの類別を行い、その結果からユーザが建物内に滞在していた区間の推定を行った結果、ユーザの実際の行動を約 97%再現できていたことを確認した。

Analysis method of noise and loss in GPS logs

KITAZAWA TAKUMI¹ SAKAMOTO DAISUKE² AZUMI TAKUYA³ MOCHIDUKI MASAHIRO⁴
NISHIO NOBUHIKO²

1. はじめに

近年の携帯端末の高機能化は目覚ましく、様々な種類のセンサデバイスが搭載されるようになってきている。携帯端末のユーザは、所有する携帯端末を日常生活の中で持ち歩くことが多いため、このようなデバイスから得られる位置情報やセンサ情報、無線基地局情報などは、ユーザやその周りの環境に関するライフログである。

このようなライフログを活用したアシストサービスの実現を目的として、ライフログを用いたユーザの行動認識をする研究が行われている。Satonaka ら [1] は、携帯端末の GPS レシーバから得られる位置情報や、加速度計から得られる加速度情報を解析することで、ユーザの日常的に立ち寄る場所や移動手段を推定している。さらに、推定された情報を統計的に利用することで、アドバイスをシステム側から提示するような、能動的な移動支援を可能としてい

る。山田ら [2] は、ユーザの行動を支援するサービスにおいて、携帯端末の GPS レシーバから得られた位置情報を基にしてユーザが停滞していた場所やその間の移動経路を推定している。

ライフログの中でも特に、GPS レシーバから得られる位置情報 (以下、GPS ログ) は既に様々なサービスやシステムに広く利用されている。しかし、この GPS ログを取得・蓄積する際、携帯端末の位置やそれを持つユーザの周囲の環境、端末の不具合などによって一時的にログが取得できなくなり、測位結果に欠けた箇所が生じてしまったり、不正確な位置情報が取得されてしまったりといった、不完全な GPS ログが観測されてしまう場合がある。蓄積データ中にこういった不完全な GPS ログが含まれていた場合、先に挙げた研究などにおいて、ユーザの状況認識が困難になってしまう問題がある。

そこで、本研究ではこれらの不完全な GPS ログを分析し、その特徴によって類別することを目的とする。この類別によって、GPS ログを利用したサービスやシステムは、利用対象のログのうちどのログが不完全なログか、それは

¹ 立命館大学 大学院 情報理工学研究科

² 立命館大学 情報理工学部

³ 大阪大学 大学院 基礎工学研究科

⁴ 立命館大学 総合科学研究機構

どの程度の割合で含まれているかなどといった情報を把握できる。この把握によって、不正確な位置情報を使わないなど利用するデータの取捨選択が可能になり、ユーザーの状況認識の精度も向上すると考えられる。

2. GPS ログの不完全性

本研究では不完全な GPS ログとして、消失、欠損、ノイズの3つを定義している。それぞれの定義を以下に示す。

消失 GPS の正常動作が期待できない場所においてデータが取得されていなかった区間

欠損 かつて正常にデータが取得できた場所においてデータが取得できなかった区間

ノイズ 通常 GPS が取得されないはずの場所において取得されたデータ

ここで GPS の正常動作が期待できない場所とは屋内や地下といった、GPS 衛星からの信号を安定して受信できない、あるいは受信しづらい場所を指す。定義した3つの不完全な GPS ログの中でも消失に関しては、不完全とはいえ、ユーザーがよく立ち寄る場所の推定など状況認識に利用できる有用なログであると考えられる。このような状況認識への利用を鑑み、本研究では消失区間中にユーザーが移動していたか、あるいは停滞していたかによって移動中の消失と停滞中の消失という異なる性質の区間として扱う。ここまで述べてきた各条件による分類分けのまとめを表1に示す。ここで移動中の消失とは、GPS ログが一定頻度で取得出来ていなかった区間において、その端点の位置情報と時刻差からヒュベニの法則を用いて算出した速度(以下、ヒュベニ速度)がユーザーの歩行速度以上であった区間を指す。

表1 不完全な GPS ログの分類

	衛星波が受信できない	衛星波が受信不良	
		屋内	屋外
移動時	移動中の消失	欠損	
停滞時	停滞中の消失	ノイズ	欠損

一方で欠損やノイズに関してはマイニング処理等に弊害を及ぼす可能性があるため、何らかの対処が必要となる。図1に欠損の例を示す。この例ではユーザーは電車を利用して移動をしているが、端末のGPS レシーバは途中の駅で数サンプルのGPS ログを取得したのみであった。次に図2にノイズの例を示す。この例ではユーザーは円で囲んだ場所にある建物内に停滞をしているため、本来であればログが取得されない区間が存在するはずである。しかし、図2の例では携帯端末のGPS レシーバが窓から入ってきた衛星波を受信してしまったため、ユーザーが屋内にいるにも関わらずGPS ログが取得され続けてしまっている。

本研究では、携帯端末によって取得されたGPS ログを、前節で説明した4種類の不完全なGPS ログにそのまま利



図1 欠損の例



図2 ノイズの例

用可能な観測データを加えた5種類に類別することを目的とする。類別はGPS ログそのものに加え、歩数計から得られる歩数情報を用いて段階的に行う。本研究では類別の中でも状況認識の弊害になるノイズと欠損の類別に焦点を当て、この二つの正確な判定を目指す。

3. 関連研究

西田ら[3]は、端末に搭載されたGPS レシーバや携帯基地局、Wi-Fi 基地局などから得られる位置情報を基に、ユーザーが訪問した場所(POI)の推定を行っている。この研究では、取得された位置情報からユーザーの滞留点を抽出し、その位置から訪問したPOIを推定しており、このユーザーの滞留点抽出の過程で、2章で挙げた停滞中の消失とノイズの判別を行っている。そして、評価においてはKangら[4]の提案する位置情報と時刻情報のみを利用した手法よりも優位であったとしている。しかし、西田らの研究ではノイズが依然として判別・除去できておらず、一度の訪問に対していくつも滞留点が生成されてしまっている。

勝田ら[5]は、GPS レシーバから得られた情報を利用して、ユーザーが屋内にいるか屋外にいるか、つまり、2章で挙げた停滞中の消失とノイズの判別を行っている。この研究では、GPS 衛星から受信した信号に対するノイズ量を表すSNR 値や、端末と衛星のなす仰角値、測位の可否判定結果を利用してGPS の受信状態を数値化し、その数値から判定を行っている。さらに、SNR の変化量を利用して、屋内停滞中に発生したノイズへの対処も行っている。しかし、評価結果において屋内屋外判定を行えていなかったデータも存在していたことから、この手法を単純に用いるだけではノイズの判定を十分に行えないと考えられる。

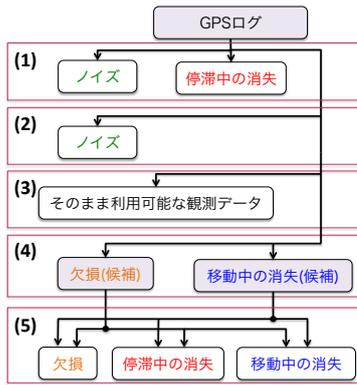


図 3 各状態の分類手法

4. 要件とアプローチ

4.1 要件

本研究が満たすべき要件として以下の2つを挙げる。
 携帯端末から得られた GPS ログに対して、ノイズを正確に判定できること

3章で挙げた GPS ログの分析を行う研究では、ノイズの判定が十分にできていない。この結果を利用すると、ユーザが屋内で停滞していた場所や時間が正確に算出できなくなってしまうため、ノイズを正確に判定する必要がある。携帯端末から得られた GPS ログに対して、欠損を正確に判定できること

2章で述べたように、欠損もマイニング処理に悪影響を及ぼす可能性がある。そのため GPS ログ中に存在する欠損を正確に分類することによって、よりマイニングの精度を高めることができると考えられる。したがって、欠損を正確に判定する必要がある。

4.2 不完全な GPS ログの分類に対するアプローチ

本研究で提案する GPS ログの分析の流れを図 3 に示す。図 3 の各段階の番号は、以下の説明の番号と対応している。

- (1) ログの取得間隔が一定以上の区間を手がかりとして停滞中の消失、ノイズを判別する。GPS ログ中には、停滞中の消失とそれ以外の欠損・消失が混ざったログも存在するが、それに関しては五段階目で対処を行う。
- (2) (1) の判別では、停滞中の消失を起点を手がかりとしているため、判別できないノイズが存在する。この二段階目では、残ったノイズを GPS から得られる精度・速度情報に加え、歩数計の情報を用いて判別する。
- (3) 停滞中の消失とノイズの判別が終わると、残るのは移動区間、あるいは屋外での停滞区間である。これらの区間に関して、データの取得間隔が最頻に近くそのまま利用可能な観測データを判別する。
- (4) (3) で判別したデータと、その他のデータを比較し欠損が起きているかどうか確認する。ここで欠損と判別されたデータは欠損(候補)とし、それ以外のデータを



図 4 屋内停滞中に観測される Lost&Found

移動中の消失(候補)とする。

- (5) (4) で一度判別したデータに関して、気圧計を用いた地下鉄の判定や W-Fi 基地局情報を用いた移動判定をして、判別の確度の向上、誤認識の発見などを行う。本研究で焦点を当てているのは(1)-(4)に含まれているノイズと欠損の判定であり、(5)の判定は今後の課題とする。各段階の分類手法については次章で詳しく説明する。

5. GPS ログの分析アルゴリズム

5.1 ノイズの判別アルゴリズム

本研究では、ノイズを二段階に分けて類別する。各段階に該当するアルゴリズムを本節の各項で説明する。

5.1.1 GPS ログのみを用いたノイズ判定

図 3 の(1)に示した一段階目では、停滞中の消失とそれに付随するノイズを判定する。まず、GPS ログ中の連続する2サンプルの取得間隔が閾値以上であった区間(以下、Lost&Found)を探し、その区間の平均移動速度が歩行速度以下であった場合に、それを停滞中の消失の起点とする。本研究では Lost&Found が生じた際、その取得間隔が開く直前に取得されたサンプルを Lost、GPS が再取得され始めた直後に観測されたサンプルを Found と呼ぶ。ユーザが屋内で停滞していた場合、基本的にこの Lost&Found は一度しか起きないはずであるが、ノイズが発生していた場合には複数の Lost&Found が発生してしまう。図 2 の例に対して閾値を2分として Lost&Found を判別した結果を図 4 に示す。4 から、一度の屋内での停滞中に、Lost&Found とノイズが複数混在して観測されていることが見てとれる。このようなノイズ(図 4 中 A)と、屋内での停滞の前後に観測されるログ(図 4 中 B)とを判別する必要がある。

屋内停滞の際最初に起きた Lost 地点と、図 4 中の B のような屋内停滞後の移動に際する GPS ログ中の各サンプルとの距離の関係を示したグラフを図 5 に示す。ここで、Lost 地点は停滞していた場所の代表地点として用いている。図 5 より、屋内停滞後には建物から離れていくようなログが断続的に取得されていることがわかる。次に、Lost 地点と図 4 の A 内に存在するあるノイズ中のサンプルとの距離の関係をグラフしたものを図 6 に示す。このグラフ

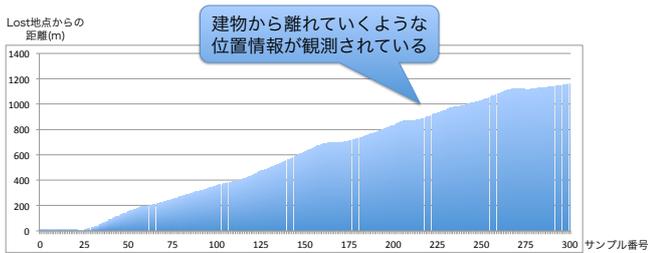


図 5 Lost 地点と屋内停滞後の GPS ログに含まれる各サンプルとの距離関係

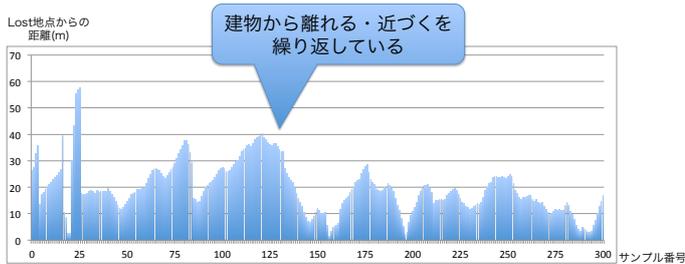


図 6 Lost 地点とノイズに含まれる各サンプルとの距離関係



図 7 GPS ログのみを用いた判定では発見できないノイズの例

は図 5 に示したグラフと異なり、建物から離れたり近づいたりを繰り返すログが取得されていることがわかる。この違いから、Found 後に建物から離れていく位置情報が取得されなかった場合、それらのログをノイズと判定する。

5.1.2 GPS ログとセンサ情報の併用によるノイズ判別

図 3 の (2) に示した類別の二段階目で判定すべきノイズは、前節と異なり Lost&Found を伴わない。この例を図 7 に示す。この例でユーザは、緑の円で囲んだ箇所に存在する建物内に滞在しているが、GPS ログは断続的に取得され続けている、つまり、ノイズが発生している。このノイズの各サンプルを GPS の精度情報によって色分けしたものを図 8 に、歩数の頻度によって色分けしたものを図 9 に示す。図 8 から、ユーザが屋内にいる間は精度の値が大きくなる、つまり、精度が悪くなっていることが見てとれる。一方で図 9 から、ユーザが屋内にいる間は歩数が頻繁に観測されなくなっていることがわかる。この特徴から、図 7 のように GPS ログが停滞していた区間に対し、精度の平均値が閾値以上かつ、歩数の頻度が一定値以下であった区間をノイズの候補と判定する。

前述した手法でノイズの候補を判定すると、その中には



図 8 精度で色分けしたノイズ例



図 9 歩数の頻度で色分けしたノイズ例



図 10 経由駅におけるノイズの誤判定

非常に多くの誤判定が混ざっている。その一例を図 10 に示す。図 10 では、電車乗車中に乗降を行わない経由駅に電車が停車していた際、誤判定が起きている。そこで、図 10 のような経由駅の問題に対しては、ノイズ候補の前後の区間に対し、速度情報や歩数情報などを用いてその区間を電車で移動していたか否か推定する。この推定において電車移動していた区間内にノイズ候補が存在していた場合、そのノイズ候補は誤判定であったと判定する。このような例以外の誤判定に関しても、それぞれの代表的な特徴から誤判定を見つけ、対処を行う。

以上のノイズ探索と、誤判定を行った結果判定されたノイズの中には、一段階目で既に発見された停滞区間の一部が含まれている可能性がある。これらの同一停滞区間は統合を行う必要がある。具体的には、図 11 に示すように停滞区間とノイズの間の屋外と判定された区間が非常に短い場合や、その区間中歩数計によって歩数が全く観測されなかった場合に、その前後のノイズと停滞区間の統合を行う。また、図 11 の推定結果中一番左に存在するような

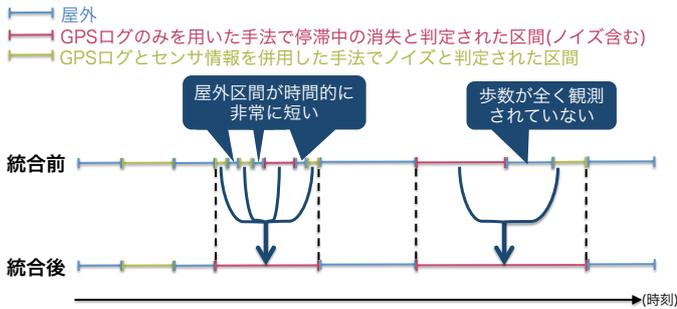


図 11 停滞中の消失とノイズ判定結果の統合

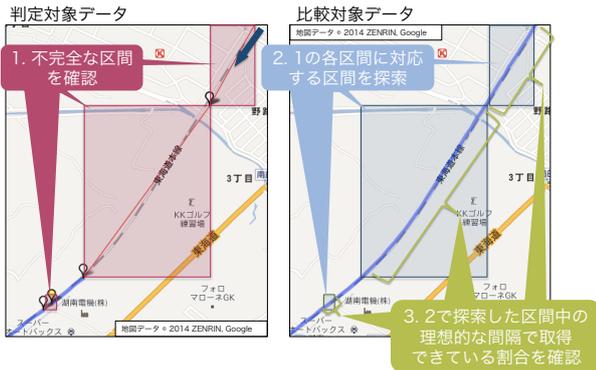


図 12 欠損・消失の判別手法

独立したノイズに関しても、屋内で停滞をしていた区間であることから、最終的には停滞中の消失として扱う。

5.2 そのまま利用可能な観測データの判定と、欠損・移動中の消失の判別アルゴリズム

図 3 の (3) に示した段階にくるログは主にユーザの移動の際に取得されたログである。この内、サンプルが理想的な間隔で取得されていた区間をそのまま利用可能な観測データとする。そして図 3 の (4) では、残った区間に対し、その区間が欠損か移動中の消失かを判別する。

5.2.1 欠損・移動中の消失の判別

欠損の判定手法の例を図 12 に示す。図 12 の左に示したデータが欠損や移動中の消失の判定を行う対象の GPS ログ (以下、判定対象データ)、図 12 の右に示したデータが判定対象データと比較するための、別の時刻に同じ箇所を取得された GPS ログ (以下比較対象データ) である。図中の青色の線は理想的な間隔で取得されている区間を、赤色の線はそれ以外を示している。この例では、判定対象データにおいて欠損が 3 回起きており、欠損どうしの間では数サンプルではあるが GPS ログが正常に取得できている区間が存在する。以下に欠損判定の手順を示す。ここで各手順の番号は図 12 中に示した番号と対応している。

- (1) 判定対象データ中の不完全な区間を確認する。
- (2) 1 で確認した区間に対し、その端点の位置情報と時刻差から比較対象データ中で対応する区間を探索する。
- (3) 2 で探索した対応区間に関して、理想的な間隔で取得

されている割合を確認する。

- (4) 3 で確認した割合が一定値以上であった場合、判定対象データ中の当該区間は欠損であると判定し、それ以外の場合には、当該区間は消失であったと判定する。

5.2.2 欠損・移動中の消失かの最終的な判定

前述した欠損・移動中の消失判定は存在する比較対象データの数だけできる。そのため、それぞれの判定回数から最終的な判定を行う。以下に判定回数による区分を示す。

- 一度でも欠損と判定された不完全区間
消失・欠損区間が完全に独立して取得されているという前提であれば、消失区間において欠損判定がされることは有り得ない。そのため、一度でも欠損と判定された不完全区間は最終的に欠損として扱う。
 - 一度も欠損と判定されず且つ、消失と判定された回数が比較回数の半数を超えていた不完全区間
前述した手法では、比較対象データの該当区間において理想的な間隔でデータの割合が多かった場合、当該区間を消失と判定した。しかし、この手法ではある区間においてたまたま欠損が発生していた場合にも消失であったという判定がなされてしまう。そのため、最終的に消失と判定された回数が全比較回数の半数を超えていた場合にも、当該区間を消失として扱う。
- 上記の二区分に含まれない区間は類別不能区間とする。

6. 評価

本節では、4.2 節、5.2 節で述べた各アプローチを実装した結果、要件として挙げた、ノイズと欠損の正確な判定が実現できていたかを評価する。アプローチの実装には Java を利用した。この評価のため、1 人のユーザに 3 週間に渡って端末を携帯してもらい、GPS ログの取得・蓄積を行った。また、歩数推定のため、加速度計から得られる情報の取得・蓄積も行った。以下に利用した端末の詳細を示す。; さらに、判定が正しいものであるかを確認するため、ユーザには端末の携帯に加え、自身が屋内で停滞をした際にその開始時刻と終了時刻の記録 (以下、正解行動記録) を手動で行ってもらった。記録は三週間行ってもらい、その記録中に含まれる屋内停滞の回数は 129 回であった。

表 2 GPS ログ・加速度情報の取得に利用した端末の詳細

利用端末	Galaxy S4
端末 OS	Android 4.2.2
GPS チップセット	Broadcom BCM47521 GPS device

6.1 ノイズ判定の評価

4.2 節でも述べたように、判定したノイズは最終的に停滞中の消失として扱う。そのためノイズ判定の評価として、一段階目でのノイズ除去率、二段階目のノイズ発見率の他に、正解行動記録と停滞中の消失の推定結果から適合率・

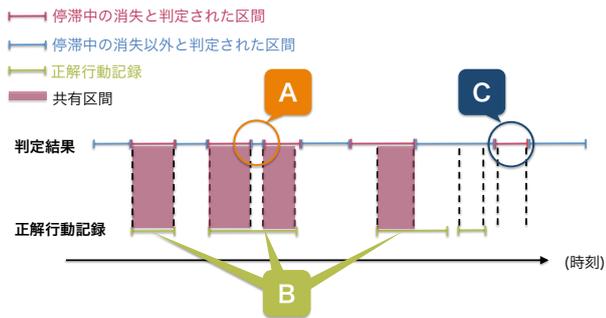
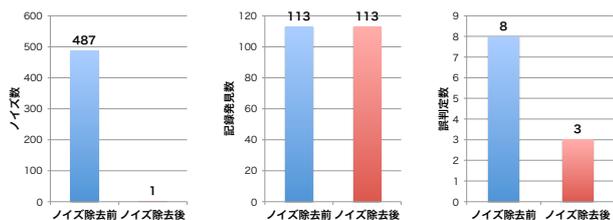


図 13 GPS ログのみを用いたノイズ判定評価の比較項目



(a) ノイズ数の変化 (b) 記録発見数の変化 (c) 誤判定数の変化

図 14 GPS ログのみを用いたノイズ判定の評価結果

再現率を算出してその評価を行う。

6.1.1 GPS ログのみを用いたノイズ判定に関する評価

5.1.1 項で述べたように、GPS ログのみを用いたノイズ判定で行っているのはノイズの除去である。そのため、Lost&Found をそのまま停滞中の消失としたノイズ除去前と、このノイズ除去後で、ノイズの数、記録発見数、誤判定数の比較を行う。比較項目の説明を図 13 に示す。ここで、ノイズ数は図 13 の A に示すような本来一つの停滞中の消失になるべきだった区間の分断要因となっていたノイズの数、記録発見数は図 13 の B に示すように推定結果と少しでも時間的な共有区間を持っていた行動記録の数を指す。さらに、誤判定数は図 13 の C に示すように、推定結果と時間的な共有区間を持たない行動記録の数を意味する。

比較結果を図 14 に示す。図 14 の (a) からノイズ除去前と比較して、5.1.1 項で述べたノイズ除去を行った後にノイズ数が 1 つにまで減っていることがわかる。3 節で紹介した西田らの研究では 1 日分の GPS ログを用いた際、その中に存在した計 4 つのノイズを判定・除去することができていなかったことから、本研究で提案する手法はノイズ除去の面で提案手法が優れているといえる。加えて、この結果から算出したノイズの除去率は約 99.8%であった。この結果から、本研究の提案したノイズ除去手法によって GPS ログ中のノイズをほぼ全て除去することができるといえる。

一方で、図 14 の (b) からは記録発見数は変動していないことが見て取れる。これは、ノイズ除去の有無によってその数が変動しないからである。加えて、この結果からは Lost&Found を用いた判定を行うだけで行動記録中に 129 箇所あった屋内停滞区間の内、約 87.60%を発見できていたことがわかる。ここで発見されていた屋内停滞区間とし

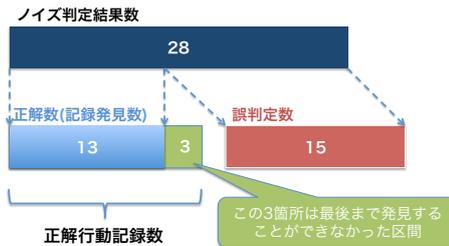


図 15 GPS ログとセンサ情報を併用したノイズ判定の評価結果



図 16 提案手法で除去し切れなかったノイズの誤判定の例

ては、端末を携帯していたユーザのよく滞在する自宅や、大学の研究棟などが存在した。この結果より、GPS ログのみを用いた分析だけでもサービスによってはそのマイニング処理になど有用である可能性があると考えられる。

図 14 の (c) からは、誤判定数が半分以下に減少していることがわかる。しかし、この減少は GPS ログのみを用いたノイズと停滞中の消失における誤判定が、その前後に存在した停滞中の消失と統合されてしまったことによるものであった。こういった誤判定は、非常に少数ではあるものの、その発生や、停滞中の消失との統合が起きてしまうと、ユーザの状況認識精度の低下要因になってしまう恐れがあるため、今後さらなる対策が必要になると考えられる。

6.1.2 GPS ログとセンサ情報の併用によるノイズ判定に関する評価

GPS ログと歩数情報を併用した手法を用いたノイズ判定の結果を図 15 に示す。図 15 において、ノイズ判定結果数は提案手法を用いてノイズと判定された区間の総数を意味する。また、正解数(記録発見数)、誤判定数は前節の図 13 に示したものと同様である。この結果から Lost&Found を用いた手法で判定することができなかった 16 箇所の屋内停滞の内、81.3%にあたる 13 箇所を発見することができていることがわかる。一方で、その記録発見数に対して誤判定数は 15 箇所、誤判定率は約 53.6%という結果であった。除去し切れなかった誤判定の一例を図 16 に示す。図 16 の例では、ユーザがトンネルを通過している区間がノイズと判定されてしまっている。この誤判定が起きた原因としては、トンネルが地下で蛇行しており、それによって GPS の停滞が発生したことが原因であると考えられる。GPS ログとセンサ情報を併用したノイズ判定手法による結果のみに注目した場合、GPS ログのみを用いた手法で発見できな

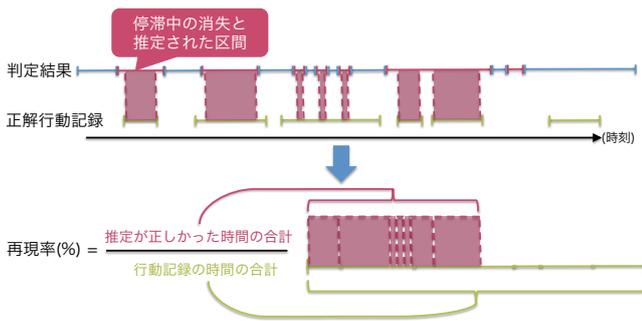


図 17 再現率の算出方法

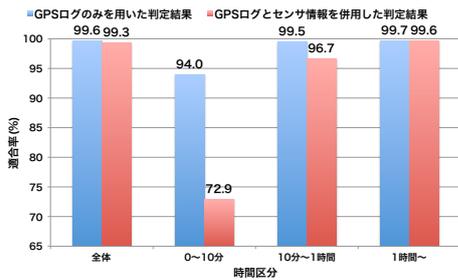


図 18 判定結果の時間に関する適合率

かった正解行動記録をほとんど発見できていることがわかる。しかし、発見できた行動記録の数に対して非常に多くの誤判定が生じてしまっている。この誤判定は前節のものと同様に、ユーザの状況認識精度の低下要因になる可能性がある。そのため、このような誤判定の除去にも継続して取り組む必要があると考える。

6.1.3 判定結果の時間に関する適合率・再現率の評価

本節の冒頭で述べたように、推定結果と正解行動記録の時間に関する適合率と再現率を算出する。それぞれの算出式を式 1 と式 2 に、再現率の算出法の説明を図 17 に示す。

$$\text{適合率 (\%)} = \frac{\text{推定が正しかった時間の合計}}{\text{判定結果の時間の合計}} \quad (1)$$

$$\text{再現率 (\%)} = \frac{\text{推定が正しかった時間の合計}}{\text{正解行動記録の時間の合計}} \quad (2)$$

図 17 において、推定が正しかった時間の合計とは、判定結果と正解行動記録の共通区間(図 17 中の赤く塗りつぶされている区間)の時間の合計である。一方で、行動記録の合計は図中緑の線で示される時間の合計である。適合率を算出する際には判定結果(図中赤線)の時間の合計を用いる。長時間の停滞だけでなく、比較的短時間の停滞に関しても詳細に評価を行うため、全ての判定結果・行動記録を用いた評価の他、判定結果・行動記録の時間が 10 分未満、10 分以上 1 時間未満、1 時間以上の 3 区分に分けて評価を行った。図 18 に適合率、図 19 に再現率の算出結果をそれぞれ示す。

図 18 から、Lost&Found を起点としたノイズの判定結果では全ての区分で適合率が 90% 以上であり、特に、判定結果の時間が 1 時間以上の場合には 99.7% と非常に高い値

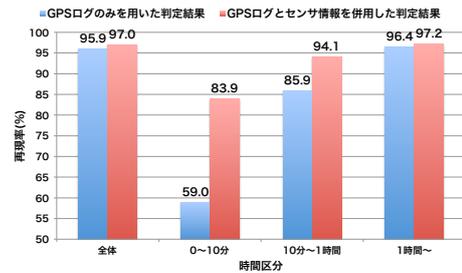


図 19 判定結果の時間に関する再現率

になっていることがわかる。適合率が高い理由としては、Lost&Found が生じ、その間の平均移動速度が歩行速度以下であったところではほぼ確実に屋内での停滞が生じていたことや、提案手法によるノイズの除去ができていたことが要因であると考えられる。一方で、二段階目のノイズ判定を行った結果においては、全ての区分で一段階目の結果よりも適合率が低下している。特に 10 分未満の区分に関しては適合率が 20% 以上低下しているが、これは前節で述べた誤判定の発生によるものであると考察できる。さらに、10 分以上の 2 区分において適合率が下がっている。この要因としては、屋内での停滞の前後でノイズの誤判定が起これ、それが停滞中の消失と統合されてしまったことが考えられる。適合率の低下は、全体的に見ると微少ではあるが、短時間の停滞も正確に認識するため、誤判定削減の継続が必要であることを示している。

図 19 から、GPS ログのみを用いたノイズ判定終了時における再現率は、全体的に見ると 95.9% と高い数値を示しているが、10 分未満の区間に限定すると 59% まで落ち込んでいることがわかる。この原因としては、この区分に 5.1.2 項で示したようなノイズが多く含まれていることが挙げられる。これに対し、二段階目のノイズ判定後の再現率を見ると、83.9% と 20% 以上向上していることがみてとれる。再現率向上の要因としては、一段階目で判定できていなかったノイズを発見できたことが挙げられる。一方で、10 分以上の二区分に関しても再現率の向上が見られた。これは、図 11 で説明した停滞中の消失とノイズとの統合により、一段階目の判定結果をより正解行動記録に近づけるような修正が行えたことによるものと考えられる。

6.2 欠損判定の評価

移動中の消失の生じる回数が既知の特定区間において 5.2 において説明した手法を用い、欠損・移動中の消失を正確に判定できるか評価を行う。この評価に関しては異なるユーザの GPS ログ・加速度情報を利用した。このユーザの利用していた端末の情報を以下に示す。

利用端末	Galaxy Note 2
GPS チップセット	Broadcom 4752 GNSS

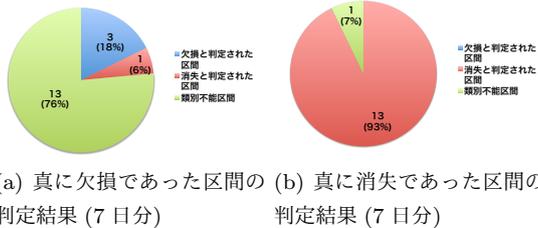


図 20 欠損・移動中の消失の判定結果 (7 日分)

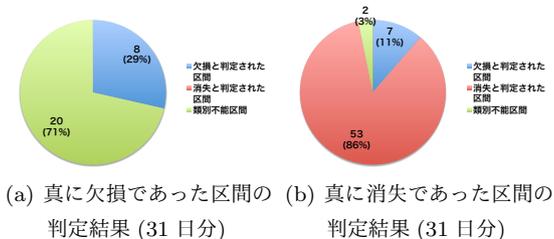


図 21 欠損・移動中の消失の判定結果 (31 日分)



図 22 消失と判定されてしまった欠損の例

この評価において、5.2 節で述べた理想的な間隔の閾値は 30 秒に設定した。また、位置情報の端点どうしを比較する距離閾値には、GPS ログから算出したユーザの移動速度 (時速) を 4 倍した値と、端点のサンプルが持つ精度の値を足した数値を用いた。評価に利用したのは京都駅から南草津駅に移動する電車移動の際に取得された GPS ログであり、この間に必ず生じる移動中の消失は 2 箇所のトンネル区間である。また、データ量は 7 日分と、その 7 日間を含む 31 日分の 2 種類を利用した。7 日分の GPS ログを用いた際の評価結果を図 20 に、31 日分の GPS ログを用いた際の評価結果を図 21 にそれぞれ示す。

7 日分の GPS ログを用いた結果においては、消失がある程度正確に判別できていることがわかる。しかし、消失と判定された区間の中に一つだけ欠損が振り分けられてしまっている。この誤判定を図 22 に示す。図 22 に示された GPS ログは、ユーザは電車に乗っており、その電車が駅で停車中の際に取得されたものである。これは、ユーザが電車に乗っていて、かつ駅の屋根などに遮られる、非常に GPS の信号を受信し辛い環境下にいたことが原因であると考えられる。また、類別不能区間は欠損と消失合わせて 14 区間あり、抽出された不完全な区間の半数近くを占めている。これは比較データが少ない、あるいは精度や

速度の問題により、判定データに対応するログを比較対象データ内から発見できなかったことが原因として挙げられる。図 21 に示す 31 日分の GPS ログを用いた結果においては消失が全て正確に判別されていることがわかる。これは比較数が増えたことによって誤判定されていた欠損が正確に判定されたことによると考えられる。しかし、欠損である区間に関して、類別が不能な場合も非常に多かった。これは、欠損・消失どちらも図 22 と同様に比較対象のログを発見できないことによるものであった。これに関しては今後距離閾値の算出方法や比較方法を改善するなどして、誤差が大きい場合や非常に短い区間であっても比較が行えるようにすべきである。

7. 結論

本稿では、ライフログを活用したサービス・システムにおいて不完全な GPS ログが問題となることを説明し、その不完全な GPS ログとして四種類を定義した。また、この四種類の不完全な GPS ログを、GPS ログと歩数情報を利用して類別するアプローチを説明した。さらに、これらのアプローチを実装し、実データを用いた評価を行った結果、大部分のノイズと欠損の類別が可能になったことを示した。

今後としては、他センサを用いて類別の精度向上や誤判定の削減を行っていく必要があると考える。加えて、このような類別結果を利用して、欠損したログの補完や、さらにはその補完を利用した GPS ログの省電力化などが期待できると考える。

参考文献

- [1] Yusuke Satonaka, Takumi Kitazawa, Kazuki Suzuki, Yuki Fukuzaki, Takuya Azumi, and Nobuhiko Nishio : LifeLog-Based Active Movement Assistant System, 1st IEEE International Conference on CyberPhysical Systems, Networks, and Applications(CPSNA2013), Taipei, Taiwan Aug, 2013
- [2] 山田直治, 磯田佳徳, 南正輝, 森川博之 : 屋外行動支援のための GPS 携帯電話を用いた移動経路の逐次の精練手法, 情報処理学会論文誌 Vol. 52 No. 6, pp1951-1967, June 2011
- [3] 西田京介, 戸田浩之, 倉島健, 内山匡 : 確率的訪問 POI 分析:時空間行動軌跡からのユーザモデリング, マルチメディア, 分散, 協調とモバイル (DICOMO2013) シンポジウム, 平成 25 年 7 月
- [4] Kang, Jong Hee and Welbourne, William and Stewart, Benjamin and Borriello, Gaetano : Extracting Places from Traces of Locations, WMASH '04, pp110-118, 2004
- [5] 勝田悦子, 内山彰, 山口弘純, 東野輝夫 : GPS 受信状態を用いた屋内外判定法, 情報処理学会研究報告 2011-MBL-60(18) pp1-8, 2011-11-11