

# 位置推定のための健全性維持手法の 大規模無線 LAN データベースへの適用

何 韜<sup>†</sup> 梶 克彦<sup>†</sup> 河口信夫<sup>†</sup>

名古屋大学大学院工学研究科<sup>†</sup> 〒464-8601 名古屋市千種区不老町

E-mail: <sup>†</sup> tacchan@ucl.nuee.nagoya-u.ac.jp, kaji@nuee.nagoya-u.ac.jp, kawaguti@nagoya-u.jp

あらまし 我々は無線 LAN を用いた位置情報・測位に関するポータル Locky.jp を運営している。一般ユーザが観測し投稿された無線 LAN 情報をデータベース化し、そのデータベースをもとに現在位置を推定することが可能である。しかし、無線 LAN 観測における不具合や、引越し、新幹線内基地局、モバイル WiFi など移動する基地局が存在し、これらのデータを適切に処理しなければ位置推定精度に悪影響を及ぼしてしまう。本研究では、Locky.jp で収集された大規模基地局データベースを対象として、無線 LAN 観測における不具合と移動基地局を検出し、データを修正することで、無線 LAN データベースの健全性を維持する。評価実験の結果、これらの修正を施すことにより、位置推定精度が向上することを確認できた。

キーワード 無線 LAN, 位置推定, データベース

## Adaptation of Integrity Maintenance Method for Location Estimation to Large Scale Wireless LAN Observation Database

Tao He<sup>†</sup> Katsuhiko Kaji<sup>†</sup> and Nobuo Kawaguchi<sup>†</sup>

<sup>†</sup> Graduate School of Engineering, Nagoya University Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan

E-mail: <sup>†</sup> tacchan@ucl.nuee.nagoya-u.ac.jp, kaji@nuee.nagoya-u.ac.jp, kawaguti@nagoya-u.jp

**Abstract** We have operated a portal Locky.jp on location information and estimation using wireless LAN. It is possible to estimate the current location based on the database of wireless LAN information observed and submitted by general users. However, the failure in wireless LAN observation, together with the base station which has been moved, base station in Shinkansen, and mobile WiFi, would cause adverse effect in accuracy of location estimation if they did not have the proper handle. In this paper, intended with the large base station database collected by Locky.jp, we detect the failure in wireless LAN observation and base station which has been moved, and fix them to maintain the integrity of wireless LAN database. The results of evaluation confirmed the improvement of the location estimation accuracy with applying these corrections.

**Keyword** Wireless LAN, Location Estimation, Database

### 1. はじめに

近年、無線 LAN の普及が急速に進み、あらゆる場所において、無線 LAN の電波を受信することが可能になりつつある。同時に、無線 LAN を用いた位置推定システムや、それらを用いた位置情報サービスが多く提案されている[1-5]。これらの無線 LAN 位置情報システムは、事前に収集された基地局の位置情報や電波強度等の情報をもとに位置推定が行われる[3]。ここで、観測された無線 LAN 情報の中には、観測時の不具合で不正な値が記録された情報や、引越し・新幹線内基地局・モバイル WiFi のように移動する基地局の情報が含まれる。これらの情報を適切に修正し、データベースの健全性を維持しなければ、位置推定の精度低下の原因となる。

我々はこれまでに、位置推定のための無線 LAN 観測データベースの健全性維持手法を提案した[6]。本手法では、廃棄された基地局、引越した基地局、新幹線車内基地局の検出により、基地局観測データベースを修正するというものである。本手法を評価するため、Locky.jp のデータベースを用いて評価実験を行い、位置推定精度の向上を確認したが、使用したデータは一部の地域の小規模なデータであった[6]。

しかし本手法が大規模なデータベースに適用できなければ、データベース全体の健全性を維持することはできない。また、従来研究では無線 LAN 観測時の不具合に関する修正は行っておらず、移動基地局の検出手法にも改善の余地があると考えられる。

本稿では、大規模無線 LAN データベースの健全性維

持を目的とする。観測時の不具合の検出、移動基地局の検出を行い、データベースを適切に修正する。また、評価実験として、Locky.jp で収集された無線 LAN 観測データベース全体の健全性維持を行い、修正前のデータベースとの位置推定精度の比較を行う。

## 2. 大規模データベースの健全性維持

### 2.1. Locky.jp

本研究は、大規模無線 LAN データベースの例として、我々が運営している無線 LAN を用いた位置情報・測位に関するポータル Locky.jp のデータベースを対象とする。Locky.jp のデータベースは、ボランティアによって観測・投稿された無線 LAN 情報を統合して構築されている。

Locky.jp の無線 LAN 情報の収集は、Locky Stumbler という PC 用ツール、または iSpotter という iPhone アプリを通して行う。無線 LAN 観測は一定間隔で行われ、その時点で受信した無線基地局の BSSID や電波強度等の情報を GPS から取得される緯度・経度情報や時刻情報と共にログファイルとして保存される。ユーザがこのログファイルを Locky.jp にアップロードすることにより、Locky.jp のデータベースの更新が行われる[7]。Locky.jp データベースの観測情報は、基地局情報 (BSSID, ESSID) とその電波強度、観測時間、GPS による観測位置 (緯度, 経度) が含まれる。

本研究で用いたデータベースの概要を表 1 に示す。トレースとは、ユーザが観測を開始して終了するまでの一連の無線 LAN 観測情報をまとめたものである。

表 1. 本研究で利用した Locky.jp の無線 LAN 観測情報

観測情報の数	33,056,474 件
観測点の数 (GPS)	4,542,464 個
トレースの数	8,838 個
BSSID の数	885,287 個
観測期間	2005-07-06 から 2010-04-09 まで

### 2.2. 大規模データベース処理

大規模なデータベースを対象とし、様々な統計分析、比較、実験を行う場合には、様々な処理に利用可能なデータベースから中間ファイルを柔軟に生成・利用できる必要がある。また、中間ファイルから別の中間ファイルを生成することも必要である。トレースごとに観測情報をまとめたものや、特定の時間や区域の観測情報を抽出したものなど、汎用的な中間ファイルを容易に生成可能である必要がある。

従来我々は、MySQL に格納された無線 LAN 観測デ

ータベースを Java で処理することにより、移動基地局や新幹線内基地局を検出してきた。しかし、RDB において中間ファイルに該当するテーブルを新規に生成する場合には、スキーマの定義やチューニングなどをその都度行わなければならない、煩雑であった。

そこで、本研究では、Linux におけるシェルスクリプトでテキストファイル进行处理する手法を採用した。無線 LAN 観測情報はテキストファイルとしてファイルシステムに保管される。マシンに十分なメモリが確保され、テキストファイルがメモリ上に展開されていれば、RDB と遜色ない速度での検索や比較等の処理が可能である。生成された中間ファイルは、閲覧、修正、コピー、削除、バックアップなどを容易に行うことが可能であり、特別なチューニングも必要ない。本方式を採用することで、実時間で Locky.jp の全国規模の無線 LAN 観測情報の分析や実験が可能となった。



図 1 無線 LAN 観測時の不具合の例

### 2.3. 位置推定に影響を及ぼすデータ

Locky.jp の無線 LAN 観測情報を分析した結果、無線 LAN 観測時の不具合と思われるデータを多数発見した。図 1 は、ある基地局が観測された地点と電波強度のトレース (約 2km) を示しているが、これら全ての観測点において同じ電波強度が観測されている。一般に、無線 LAN の電波強度は揺らぎが生じ、電波強度は距離が遠くなるほど弱くなるため、同じ電波強度が観測されることはない。よって、この観測データはデバイス

もしくはデバイスドライバの不具合によって正常な観測ができなかったと考えられる。このような誤った電波情報がデータベース中に存在すれば、位置推定精度が低下してしまうおそれがある。このような不具合は、無線 LAN 情報を収集する際に、突然無線 LAN 電波が途絶えたり、デバイススリープ等の操作により、最後に収集した情報がデバイスのキャッシュに残ったりする場合に発生すると考えられる。

また、引っ越しを行って基地局が別の地点に移動する場合は考えられる(図2)。さらに、近年では新幹線(N700系)内基地局(図3)やモバイルWiFiなど、特定の地点に駐在し続けるとは限らない基地局が数多く普及してきている。位置推定の際にこれらの基地局の情報を適切に処理しなければ、位置推定精度の低下につながってしまうだろう。



図2 移動した基地局の観測点分布 (左: 移動元、右: 移動先)



図3 新幹線に搭載された基地局の観測点分布

### 3. 無線 LAN 観測の不具合の検出

本章では、無線 LAN 観測の不具合を検出するための手法を述べる。

まず、無線 LAN 観測の不具合が含まれるデータの傾向を把握するために、データベースからランダムに349件のトレースデータを抽出し、不具合の有無を調査した。その結果、100件のトレースデータで不具合が確認された。これらのデータより、不具合が生じているトレースでは、7回以上連続して同一の電波強度が観測されていることが観察された。そこで本研究では、トレース内で、同一基地局の電波強度が7回以上

連続して等しかった場合に、不具合が生じたと判定することにした。

無線 LAN 観測の不具合検出手法を Locky.jp データベース全体に適用した結果を表2に示す。表より、全体のトレースの約44%が不具合を含むトレースと判定された。データベースの健全性維持のためには、不具合のデータを修正する必要がある。不具合を含むトレースであっても、不具合が発生するまでの観測情報は正常であり、位置推定に利用可能である。そこで、トレース内の不具合が発生している部分の観測情報のみを削除することでデータベースの健全性が維持される。

表2 無線 LAN 観測の不具合

	全体	不具合
トレースの数	8,838	3,893
観測情報の数	33,056,474	14,411,449

### 4. 移動する基地局の検出

Locky.jpの無線 LAN 観測情報の観察により、移動する基地局の存在が認められた。位置推定において、常駐している基地局の情報は信頼性が高いと考えられるが、逆に移動する基地局は信頼性を低く設定する必要がある。そこで本研究では、全ての無線 LAN 観測情報から、常駐している基地局と移動する基地局の2種類に分類する。さらに、移動する基地局には、転居等により引っ越しした基地局、モバイル WiFi のような移動と停止を繰り返す基地局、新幹線車内基地局のような常時移動している基地局の3種類が存在することが分かっている。引っ越し、新幹線、モバイル WiFi の3種類の移動基地局には、移動の特徴に違いが存在すると考えられる。移動の特徴量として以下の4種類の情報を分析し、移動基地局の種類を分類可能であるかを検討する。

1. 移動回数
2. 滞在期間
3. 移動距離
4. ベンダーコード、機種コード

#### 4.1. 移動に関する情報の抽出手法と検出結果

まず、ある基地局が観測された地点間の距離が1km以上離れていた場合に、その基地局は移動する基地局であると分類する。ここで、一般的に無線 LAN の電波が届く範囲は、基地局を中心とした直径1kmの範囲内であるため、この値を閾値とした。

以降、前述した4つの移動の特徴量の抽出手法について述べる。

##### (1) 移動回数

基地局が1km以上離れた2地点で観測された場合に

移動したと判別するアルゴリズムを用い、各基地局について以下の手順で移動回数を求める。

1. ある基地局の観測点集合を対象とする
2. 集合内で最も古い観測点を基準点とする
3. 基準点との距離が 1km 以内の観測点を観測点集合から削除する
4. もし集合内に観測点が残っているなら、2 から 4 を繰り返す。一回繰り返したら、移動回数を 1 増やす。

この手法を全 885,287 件の基地局に適用したところ、57,252 件の移動基地局が検出された。移動基地局の割合は約 6.47%である。この結果の中で、最大移動回数は 484 回であった。移動回数による基地局の数の分布を図 4 に示す。ここで、本手法で図 3 に示すような新幹線内基地局をカウントした場合、1km の移動ごとに 1 回の移動とカウントすることになるため、移動距離に比例して移動回数が増加する。よって、高い移動回数となっている基地局は新幹線のように長距離移動を行う基地局と考えられる。

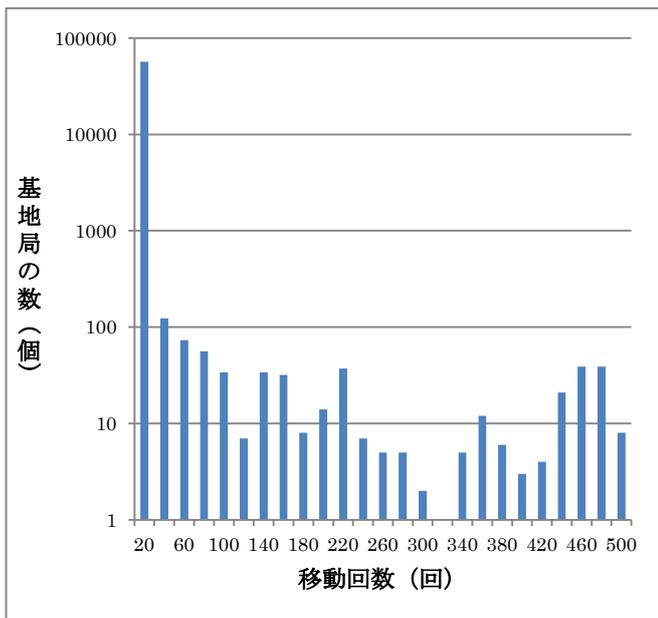


図 4 移動回数の分布

### (2) 滞在時間

基地局の引っ越しは、複数回行われる可能性があるが、常駐期間はモバイル WiFi や新幹線内基地局よりも長期であると予想される。そこで、57,252 件の移動した基地局について、最終的に移動した地点からどの程度の期間その地点に常駐しているかを分析した。検出結果を図 5 に示す。図より、1 カ月以内に移動を行うような滞在期間の短い基地局が多く、また半年以上のような長い滞在期間の基地局も多く存在することが

わかる。

### (3) 移動距離

新幹線のようにルートが決まっている基地局であれば、移動距離の長さによって基地局の種類を判別可能ではないかと考え、移動した基地局の最大移動距離を分析した。結果を図 6 に示す。最大の距離は 1,590.65km であった。これはほぼ沖縄-東京間の距離に等しい。新幹線内基地局は、走行距離から 1,100km 以上になることはない(東京-博多間は約 1,060km)と考えられるため、飛行機に乗ったユーザが飛行機内にあったモバイル WiFi を観測したものと考えられる。

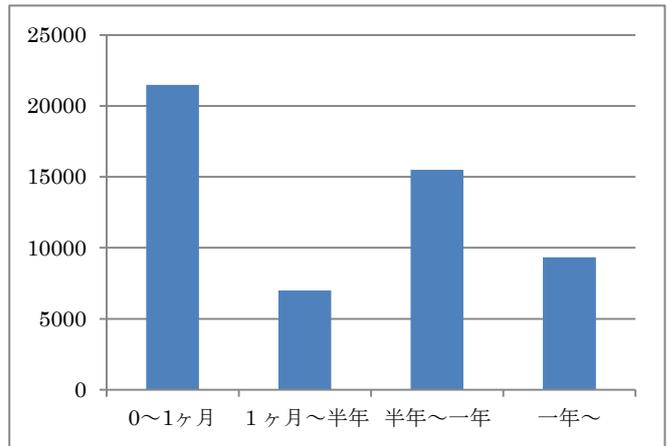


図 5 最終移動地点での滞在時間

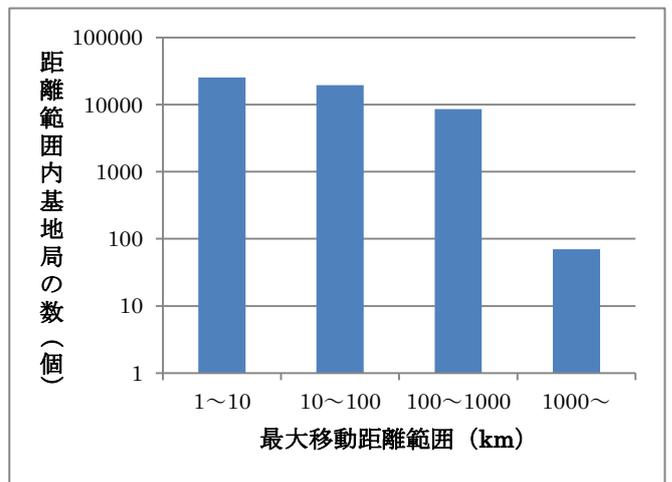


図 6 最大移動距離

### (4) ベンダーコード、機種コード

BSSID には法則性があり、そこからベンダーや機種を推定することが可能である。無線 LAN の MAC アドレス 48 ビットのうち、最初の 24 ビットがベンダーコード、次の 8 ビットが機種コードである。よって、BSSID の先頭から 32 ビットが同一の場合、同じタイプの基地局であると推測できる。たとえば基地局 MAC

アドレス AB:CD:EF:BA:DC:FE を例とすると、その前半部分の AB:CD:EF:BA が、その基地局のベンダーコードと機種コードである。

これらの移動に関する特徴量から、本研究では新幹線に搭載された基地局を分類することができた。まず、移動回数の段落で述べたように、本手法で算出した移動回数が大きい基地局は、長距離の移動を行っているといえる。そこで、移動回数が 100 回以上となる基地局 289 件のベンダーコードと機種コードを調査したところ、その種類は二つのみであった。これらの基地局の中で、ランダムに 10 件の基地局のトレースを GoogleMaps 上にプロットしたところ、新幹線の経路上に観測点が配置された。よってこれらの基地局は新幹線内基地局であるとみなし、データベースを修正することが可能である。

## 5. 評価実験

本研究では無線 LAN 観測の不具合の検出、移動基地局の検出手法を提案した。これらを用いたデータベースの修正が位置推定精度に及ぼす影響を検証するため、評価実験を行った。

### 5.1. 実験設定

Locky.jp データベースの全観測点約 450 万件の中の最新約 1%(約 1 ヶ月間)の観測情報を評価データとし、残る観測情報を用いて無線 LAN データベースを構築する。各検出手法の影響を調べるため、以下の 4 種類のデータベースを構築した。

1. 修正を施さない(DB1)
2. 観測の不具合を削除(DB2)
3. 移動基地局を削除(DB3)
4. 観測の不具合、移動基地局を削除(DB4)

これらのデータベースと、評価データの無線 LAN 情報に基づいて位置推定を行う。位置推定手法は Proximity 手法[8]を用いた。Proximity 手法は、端末が

観測している基地局からの受信電波強度をリストにし、最も強い受信電波強度を観測した基地局の位置を端末の位置として推定する手法である。評価データの GPS 座標との差が位置推定精度との誤差となる。

### 5.2. 評価実験結果

実験結果を表 3 に示す。カバー率とは、位置推定できる地点の数が評価データの地点数に占める割合である。データベースの修正のために、無線 LAN 観測情報を削除すると、カバー率は低下することになる。よって、修正を施さない DB1 が最もカバー率が高い。しかし、100m 以内の誤差となる割合は、修正前の無線 LAN 観測の不具合を削除し、移動基地局を削除した DB4 が最も高く、DB1 は最も低いという結果となった。

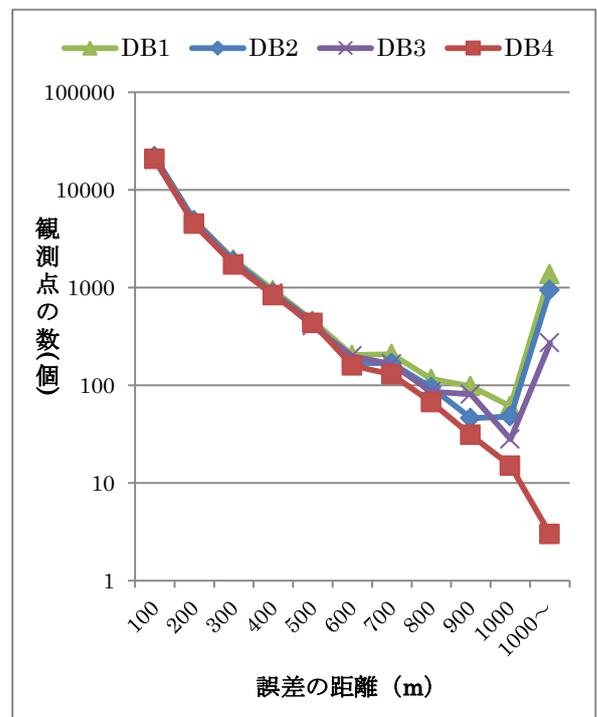


図 7 位置推定誤差分布の比較

表 3 位置推定精度の比較

	DB1	DB2	DB3	DB4
観測情報の数 (件)	33,056,474	18,645,025	24,294,139	13,117,633
基地局の数 (個)	861,839	861,839	802,257	804,674
評価用位置点の数 (個)	45,040	44,042	45,040	44,042
位置推定できる位置点の数 (個)	32,207	31,437	28,725	28,479
カバー率	71.5%	71.38%	63.78%	64.66%
100m 以内の誤差の割合	68.1%	69.7%	71%	72.4%

図 7 は位置推定誤差分布の比較を示している。0～1km まで 100m 毎に、また 1km 以上は一か所にまとめている。図より、以内の誤差範囲については大きな変化は見られなかった。しかし、不具合と移動基地局を削除した DB4 は 1km 以上の誤差になる回数が他の DB と比べて非常に少なくなっていることが分かる。DB1 では 1km 以上の誤差となる場合が 941 件あったが、DB4 では 3 件であった。以上より、無線 LAN 観測の不具合と移動基地局の修正を施したデータベースを用いることで位置推定精度を向上させることができたといえる。

## 6. おわりに

本稿では、無線 LAN を用いた位置推定のための無線 LAN データベースの健全性維持手法を提案し、無線 LAN 観測の不具合と移動基地局について、その検出手法を提案した。また大規模データベースである Locky.jp のデータベースに提案手法を適用し、位置推定の精度を向上させられることを確認した。

今回の評価実験で使用した位置推定手法は非常に単純な手法であるため、今後は複数の基地局情報や時系列的な文脈も含めた高度な位置推定手法を用いて位置推定精度を検証する必要があると考える。

## 文 献

- [1] 河口信夫, “Locky.jp: 無線 LAN を用いた位置情報ポータルとその応用”, ヒューマンインタフェース学会誌, Vol.10, No.1, pp.15-20, 2008
- [2] Anthony LaMarca, Jeff Hightower, Ian Smith and Sunny Consolvo, “Self-Mapping in 802.11 Location Systems”, Ubicomp2005, pp.87-104, 2005.
- [3] 暦本純一, 塩野崎敦, 末吉隆彦, 味八木崇, “PlaceEngine: 実世界集合知に基づく WiFi 位置情報基盤”, インターネットカンファレンス 2006, pp.95-104, 2006.
- [4] 藤田迪, 梶克彦, 河口信夫, “Gaussian Mixture Model を用いた無線 LAN 位置推定手法”, 情報処理学会論文誌, Vol.52, No.3, pp.1069-1081, 2011.
- [5] 梶克彦, 河口信夫, “indoor.Locky: UGC を利用した無線 LAN 屋内位置情報基盤”, 情報処理学会論文誌, Vol.52, No.12, 2011.
- [6] 何韜, 梶克彦, 河口信夫, “位置推定のための無線 LAN 観測データベースの健全性維持手法”, 情報処理学会第 73 回全国大会講演論文集, v.3, pp.281-282, 2011.
- [7] 吉田廣志, 伊藤誠悟, 河口信夫, “無線 LAN を用いた位置推定ポータル Locky.jp と基地局データ収集手法”, マルチメディア, 分散, 協調とモバイル (DICOMO2006) シンポジウム, 3B3, Jul.2006.
- [8] John Krumm, Ken Hinckley, “The Nearest Wireless Proximity Server”, In Proceeding of the Sixth International Conference on Ubiquitous Computing (Ubicomp 2004), pp.283-300, 2004.