

# レイリー分布を用いた Wi-Fi 電波強度分布に基づく 少数標本時位置推定

藪 圭輔<sup>1,a)</sup> 榎堀 優<sup>2</sup> 間瀬 健二<sup>2</sup>

**概要:** Wi-Fi 電波強度を用いた位置推定において、電波強度の揺らぎを考慮して、電波強度分布の一致度から位置を推定する場合、少数回の観測では十分な電波強度分布が得られず、推定精度が低下する問題がある。そこで本研究では、電波強度分布に、通信理論において、多重経路で受信機に到達する散乱信号のモデルに使われるレイリー分布を用いることで、少数回の観測でも精度の高い位置推定を実現することを目指す。電波強度分布にレイリー分布を想定した場合とガウス分布を想定した場合とを比較した結果、観測回数 2 回の場合の位置推定精度を表す F 値は、レイリー分布利用で  $0.770 \pm 0.235$ 、ガウス分布利用で  $0.055 \pm 0.054$ 、5 回で同じく  $0.871 \pm 0.174$  と  $0.647 \pm 0.195$  となった。従って、レイリー分布を利用することで、少数回の観測でも高い精度で識別できることが確かめられた。また、レイリー分布は 1 サンプルでも分布の推定が可能であり、F 値で  $0.522 \pm 0.235$  の精度を得た。

**キーワード:** レイリー分布, 少数標本, Wi-Fi 位置推定

## Wi-Fi Based Small-sample-size Positioning using Rayleigh Distribution

**Abstract:** In Wi-Fi based positioning, RSSI distribution matching is required to realize robustness for RSSI noise that is occurred by humidity change, human movement, and so on. However, RSSI distribution matching is not effective in small-sample-size positioning if it use common distribution such as Gaussian. Thus, in this paper, we propose use of Rayleigh distribution in order to solve the issue. The Rayleigh distribution is a model of scattering signals arrival with multiple paths in communication theory, and is able to estimate its distribution with small-size samples. As an evaluation result, in two samples, the F-measure is  $0.770 \pm 0.235$  with our proposed method, and it is  $0.055 \pm 0.054$  with Gaussian; and in five samples, the F-measure is  $0.871 \pm 0.174$  with our proposed method, and it is  $0.647 \pm 0.195$  with Gaussian. In addition, our proposed method is able to estimate position in only one sample with  $0.522 \pm 0.235$  in F-measure.

**Keywords:** rayleigh distribution, small-sample-size, Wi-Fi positioning

### 1. はじめに

屋外だけでなく屋内においてもユーザの位置を推定するために、Wi-Fi 電波強度を用いた位置推定の研究が多く行われ、PlaceEngine[1] や locky.jp[2] などのサービスがすでに公開されている。

電波強度を用いた位置推定に関する従来の研究では、ある地点で観測される受信電波強度 (以下, RSSI) の分布か

ら、その地点における RSSI の確率密度関数を事前に算出し、実際に観測した RSSI から、該当地点における存在確率を計算して位置を推定する方法 [3] や、ある地点における RSSI の観測ベクトルを事前に算出し、ベクトル空間内距離を用いて位置推定する方法 [4] などがある。

しかし、Wi-Fi 電波強度は水蒸気などの気象条件による減衰、近くに立つ人による減衰、移動物によるマルチパス等の影響により同じ位置でも観測ごとに変化する。そのため、ある地点で観測される RSSI は同地点でも観測ごとに変化するため、一回の観測結果から位置を推定する方法では、推定精度が低下してしまう。そのため、既存研究のように、識別器側のみに RSSI の確率分布を持たせ、識別器

<sup>1</sup> 名古屋大学工学部

School of Engineering, Nagoya University

<sup>2</sup> 名古屋大学大学院情報科学研究科

Graduate School of Information Science, Nagoya University

<sup>a)</sup> yabu@cmc.ss.is.nagoya-u.ac.jp

の入力は単一回の Wi-Fi 観測結果から位置を推定するのでは無く、識別器の入力でも RSSI の確率密度分布を構築し、位置を推定する必要がある。しかしながら、谷内らのように RSSI の分布にガウス分布を用いると少数のサンプルで分布を十分に特定することができず、短時間で位置を推定することができないという問題があった。

そこで本研究では少数のサンプルでも分布の外形を推定しやすいレイリー分布を用いて、分布による一致度による RSSI の揺らぎを考慮した位置推定を実現し、短時間で高精度な位置推定をすることを旨とする。レイリー分布とは、通信理論において、多重経路で受信機に到達する散乱信号のモデルに使われる分布であり [5]、Wi-Fi の RSSI 分布はガウス分布よりも対数正規分布やレイリー分布に近いことが知られている [6]。しかし、対数正規分布は少数のサンプル数で分布を十分に特定できないという問題がある。そこで、本研究では少数のサンプルでも分布の外形を推定しやすいレイリー分布を用いて位置推定をした。

なお、本稿で扱う位置推定とは、ユーザが一定期間とどまっていた地点を意味のある場所(エリア)として学習し、その場所に居るか否かを識別することである。

## 2. 関連研究

Wi-Fi 電波を用いた位置推定に関する研究として、藤田らは受信した Wi-Fi 電波強度の分布を混合ガウスモデルでモデル化することによって位置推定を行った [7]。名古屋大学構内で得られた Wi-Fi 電波強度ログを混合ガウスモデルでモデル化し、パーティクルフィルタで位置を推定したところ、6m~10m の程度でのリアルタイムな位置推定を可能にした。谷内らは周辺の環境の突発的な変化に対応できるように、さまざまな弱位置推定器からなる Boosted 位置推定器を用いて推定精度を上げる手法を提案し、特定のアクセスポイントが取り除かれた環境でも精度の悪化を防ぐという結果を得ている [3]。伊藤はノンパラメトリックな GaussianProcess で Wi-Fi 電波強度の分布を表現し Wi-Fi 電波強度の分布を問わない位置推定をした [8]。

しかし、上記の既存研究は、識別器と識別器の入力の両方で分布を推定して位置識別する手法をとっておらず、デバイス側で単一回の Wi-Fi 観測結果から位置を推定するため、RSSI の揺らぎに対応していない。その対策として、デバイス側でも RSSI の分布を推定して、分布の一致度で位置識別する方法が有効である。

RSSI の分布として、既存研究ではさまざまな分布が使用され、谷内らはガウス分布を使用している [3]。一方で Sklar らは RSSI に関する研究を行い、RSSI の分布はレイリー分布や対数正規分布に類似しているとした [6]。しかし、ガウス分布や対数正規分布は少数のサンプル数で分布を十分に特定できないという問題があった。

そこで、本研究では、RSSI の揺らぎを考慮して分布同

士を比較し、かつ、少数サンプルでも十分に分布を特定できるレイリー分布を用いて短時間での位置を推定する手法に着目した。

## 3. 推定に使用した WI-FI 電波からの情報

本研究ではアクセスポイントから発信される Wi-Fi 電波を用いてエリア識別をする。Wi-Fi 電波を受信することでさまざまな情報が得られるが本研究ではその中の Basic Service Set Identifier(BSSID) と Received Signal Strength Indication(RSSI) を用いた。

BSSID はネットワークの識別子であり、これによりどのアクセスポイントから発信された電波なのかが判明する。通常、BSSID はアクセスポイントの、ネットワークの各ノードを識別するために割り当てられた物理アドレスである MAC アドレスと同一のものに設定されている。この場合、先頭の 24 ビットはアクセスポイントの製造メーカーを表しており、後半の 24 ビットは製造メーカー内でのユニーク ID を表しているという特徴がある。

RSSI は dBm という単位が用いられる。受信電力を Power(単位:W) とすると RSSI は式 1 で求められる。

$$RSSI = 10 \log_{10} (\text{Power}/0.001) \quad (1)$$

アクセスポイントから発信された電波から観測される受信電力は、通常 0.001W より小さいので RSSI の値は負の値をとる。アクセスポイントから発信された電波から観測される受信電力 Power は Wi-Fi 電波の発信点と受信点の間の距離  $d$  と定数  $\alpha$ ,  $C$  を用いて式 2 で求められる。

$$\text{Power} = Cd^{-\alpha} \quad (2)$$

$\alpha$  の値は減衰係数と呼ばれ、自由空間では 2、実空間では 2~4 程度であることが知られている [9]。受信電力は距離に依存する値となるため、RSSI も距離に依存する値となる。

## 4. エリア識別器

### 4.1 エリア識別器の設計

本研究で用いたエリア識別器の概要を図 1 に示す。本研究ではユーザが一定期間にわたり静止していたならば、その場所は意味がある場所(エリア)であり、位置識別を実施する価値がある場所であるとし、エリアとして学習することとした。エリア識別器は Wi-Fi 電波ログを得るごとに逐次的に学習する。これは、取得する Wi-Fi 電波ログに、緯度経度や建物内での位置といった教師信号が付加されていないためである。しかし、教師信号が付加されていないため、ある静止期間が複数存在すると、同じエリアで静止していたのが異なるエリアであったのが区別できず、同じエリアに対する識別機が重複して構築されるなどの問題があった。そこで、エリア識別器はエリアを基準に学習する。エリア識別器の学習方法は 4.4 節で述べる。

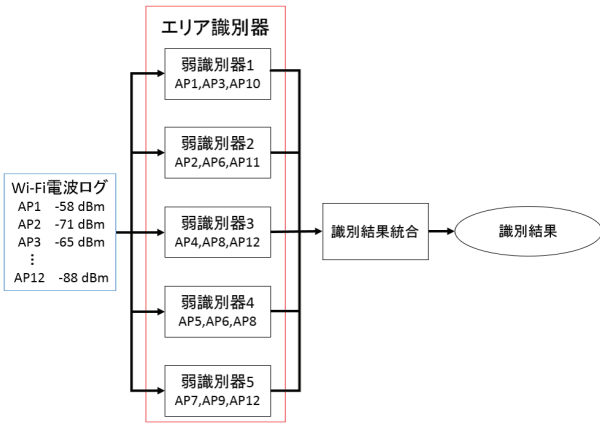


図 1 エリア識別器の概要  
 Fig. 1 Overview of the area identifier

本研究で用いたエリア識別器は5つの弱識別器を持たせ、結果を統合させて識別する手法をとった。エリア識別器は入力されたWi-Fi電波ログが同一エリアで観測されたものかそうでないかの2値判別を行う。弱識別器の数は予備実験により十分な識別精度を得られたため5とした。各弱識別器は事前に割り当てられたアクセスポイントのみを使用してエリア識別をする。あるアクセスポイントが周辺環境の変化によってRSSIが突発的に変化してもそのアクセスポイントを使用していない弱識別器は識別に影響を受けないため、環境の変化に強くなるからである[3]。

弱識別器に推定させるアクセスポイントの割り当てにはアクセスポイントが発信するWi-Fi電波から得られるBSSIDを使用する。割り当てるBSSIDの数は、異なる3つの位置からの相対位置が求まれば観測点が一意に決定されるという三点測量の考えに基づき最低3つとし、3つの異なる位置にあるアクセスポイントを用いて識別した。ただし、観測されたBSSIDの数が3つ以下の場合には割り当てるBSSIDが3つに満たない場合もある。BSSIDの割り当て方法は4.3節で述べる。

BSSIDを割り当てられた各弱識別器は入力されたWi-Fi電波ログの内、割り当てられたBSSIDをもつWi-Fi電波ログのみを使用し、各BSSIDごとにRSSIの分布を推定する。1章で述べたとおり、本研究ではRSSIの分布の推定にレイリー分布を用いて推定した。図2にレイリー分布の例を、式3にこの分布をあらわす確率密度関数の式を示す。スケール $\sigma$ は観測値から求められる値である。図のようにレイリー分布は確率変数に0以上の値をとり、0を始点とする分布である。

$$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (3)$$

エリア識別器の生成に使用する学習用Wi-Fi電波ログで求めたレイリー分布の確率密度関数を $f_{\text{area}}(x)$ とする。 $f_{\text{area}}(x)$ の算出方法は4.3節で述べる。

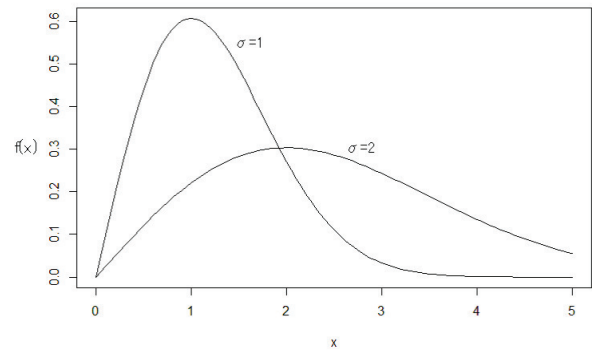


図 2 レイリー分布の例  
 Fig. 2 Examples of Rayleigh distribution

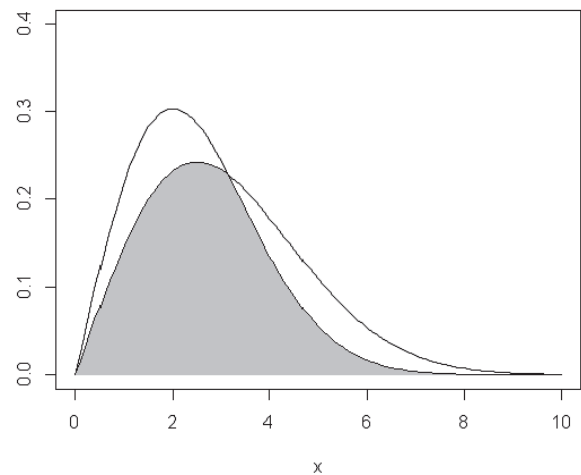


図 3 二つの分布の一致部分  
 Fig. 3 Matching portion of the two distribution

#### 4.2 エリア識別方法

エリア識別器の入力はある位置で収集したWi-Fi電波ログとする。エリア識別器は、5個の弱識別器の結果を統合し、エリア識別器を生成したエリアで観測されたWi-Fi電波ログか否かを判別する。各弱識別器はエリア識別器に入力されたWi-Fi電波ログの内、弱識別器に割り当てられているBSSIDを持つ電波から得たRSSIのみを使用し、各BSSIDごとにRSSIの分布を算出する。これによって算出されたRSSIの確率密度関数を、エリア識別器に入力した未識別Wi-Fi電波ログを観測した地点で得られるRSSIの確率密度関数とし、 $f_{\text{loc}}(x)$ で表す。

弱識別器は $f_{\text{loc}}(x)$ と $f_{\text{area}}(x)$ を用いてエリア識別する。エリア識別の指標として二つの確率密度関数の共通部分を使用した。この共通部分の大きさのことを一致度とする。図3に灰色で二つのグラフの共通部分を示す。

式4で二つの分布において重なっていない部分の面積 $S_{\text{diff}}$ を算出する。

$$S_{\text{diff}} = \int_{-\infty}^{\infty} |f_{\text{area}}(x) - f_{\text{loc}}(x)| dx \quad (4)$$



確率密度関数  $f_{loc}(x)$  の面積  $S_{loc}$  と、確率密度関数  $f_{area}(x)$  の面積  $S_{area}$  はそれぞれ 1 となる。そのため、二つの確率密度関数の一致度  $S_{coin}$  は式 5 となる。

$$S_{coin} = \frac{S_{loc} + S_{area} - S_{diff}}{S_{loc} + S_{area}} = 1 - \frac{S_{diff}}{2} \quad (5)$$

式 5 で求められる一致度  $S_{coin}$  を弱識別器に割り当てられている BSSID 別の Wi-Fi 電波ログから算出し、算出された一致度の平均を求める。一方で、入力された未識別 Wi-Fi 電波ログ中にはある弱識別器に割り当てられたある BSSID の Wi-Fi 電波ログが存在しないこともある。この場合は一致度  $S_{coin}$  の値は 0 とする。求めた一致度の平均が閾値以上になった場合は入力された Wi-Fi 電波ログは同一位置で観測されたものとして弱識別器の出力を 1 とし、閾値より小さい値だった場合は入力された Wi-Fi 電波ログは同一位置で観測されたものではないとして弱識別器の出力を -1 とする。

最後に各弱識別器  $w_i (i = 1 \sim 5)$  の出力  $a_i$  を統合したものを、最終的なエリア識別器の出力  $A$  とする。各弱識別器には 4.4 節で述べる方法で求めた重み  $w_i$  が付加されている。式 6 で求めた値が 0 以上となった場合、入力された Wi-Fi 電波ログは同位置で観測されたとして、エリア識別器の出力を TRUE とし、0 より小さかった場合は同位置で観測された Wi-Fi 電波ログではないとして、FALSE を出力する。

$$A = \sum_{n=1}^5 a_n w_n \quad (6)$$

### 4.3 Wi-Fi 電波ログによるエリア識別器構築

4.1 節で述べたように、弱識別器に割り当てる BSSID の数は、異なる 3 つの位置からの相対位置が求まれば観測点が一意に決定されるという三点測量の考えに基づき最低 3 つとし、3 つの異なる位置にあるアクセスポイントを用いて識別した。

しかし、アクセスポイントによっては異なる BSSID を持つ Wi-Fi 電波を複数発信しているものがある。このような BSSID の組を同じ弱識別器に割り当ててしまうと同一地点から発信される電波を使用してしまうため、三点測量の異なる 3 つの位置との相対位置を求めるという制約を満たさなくなる。一方で、3 章で述べたとおり、BSSID の先頭 24 ビットはアクセスポイントを製造したメーカーを表しており、BSSID を変更しない限りは同じアクセスポイントから発信される Wi-Fi 電波が持つ BSSID の先頭 24 ビットは同一のものとなる。従って、BSSID の非類似度の大きい 3 つの BSSID 同士を弱識別器に割り当てることによって同一地点から発信された Wi-Fi 電波同士を同じ弱識別器に割り当てられる可能性を軽減した。割り当てる 3 つの BSSID をそれぞれ BSSID<sub>1</sub>, BSSID<sub>2</sub>, BSSID<sub>3</sub> とする。非類似度を測る指標としてユークリッド距離を使用した。

ユークリッド距離を用いたのは、異なる 3 つの BSSID を決定する指標として、3 つの BSSID のベクトルがなす三角形の面積を使用するためである。

BSSID<sub>1</sub> は割り当てられている BSSID の種類の偏りを軽減するために割り当てに使用する BSSID の内、弱識別器に割り当てられている数が最も少ないものにした。BSSID<sub>2</sub> は BSSID<sub>1</sub> とのユークリッド距離が最大となるものとした。BSSID<sub>2</sub> を求めた後、BSSID<sub>1</sub> と BSSID<sub>2</sub> となす三角形の面積が最大となる BSSID<sub>3</sub> を式 7 で表されるヘロンの公式で求めた。a, b, c は三角形の各辺の長さを表し、S は求める三角形の面積である。

$$s = \frac{1}{2}(a + b + c)$$

$$S = \sqrt{s(s-a)(s-b)(s-c)} \quad (7)$$

3 つの BSSID の非類似度が大きいほど三角形の面積  $S$  は大きくなる。したがって、面積を最大にする 3 つ BSSID は同一地点から発信された BSSID である可能性は低いといえる。

この方法では、観測した BSSID の数が多いとどの弱識別器にも割り当てられずに識別に使用できない Wi-Fi 電波ログが発生する可能性がある。そのため、どの弱識別器にも割り当てられていない BSSID を、弱識別器に割り当てられる BSSID の数が均等になるように割り当てた。

上記のアルゴリズムでは入力される Wi-Fi 電波ログの中には BSSID の数が 3 種類の場合、全ての弱識別器に同一の BSSID が割り当てられてしまい、BSSID の数が 2 種類以下の場合、弱識別器に 3 つの BSSID を割り当てることできない。BSSID の数が 3 種類以下の場合には上記のアルゴリズムを使用せずに、弱識別器に割り当てる BSSID の数を 1 ~ 3 に変更し、割り当てられた BSSID の種類が同一となる弱識別器の組ができるだけ少なくなるように BSSID を割り当てた。

BSSID の割り当て後は BSSID ごとに RSSI の分布を推定する。しかし、3 章で述べたように RSSI は通常負の値をとる。そのため RSSI の分布をレイリー分布に対応できるように、各 RSSI に -1 をかける。この値を rssi とする。また、0 を始点とした分布という特性に対応するため、式 8 のように、rssi をあるオフセットで引くことによりレイリー分布に対応できるようにした。

$$X = rssi - \text{offset} \quad (8)$$

この  $X$  が観測値を表す確率変数となる。オフセットは弱識別器に割り当てられている各 BSSID ごとに求める。オフセットの決定方法は次のようになる。

全ての rssi の値を用いて観測値の確率密度関数  $f(x)$  を算出する。式 8 で求められる、実際に観測した全ての RSSI で算出される観測値  $x_i$  を用いて確率密度関数  $f(x)$  のスケール  $\sigma$  の最尤推定値を式 9 で算出する。

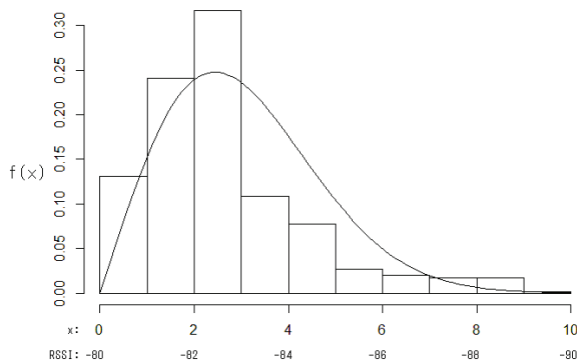


図 4 推定したレイリー分布の確率密度関数の例

Fig. 4 Examples of the probability density function of the Rayleigh distribution estimated

$$\hat{\sigma} = \sqrt{\frac{1}{2n} \sum_{i=1}^n x_i^2} \quad (9)$$

ただし、 $x_i$  が 0 以下になった場合はスケールの計算には使用しなかった。式 9 の  $n$  の値はスケール計算に使われた観測値の数である。これによって求められたスケール  $\sigma$  を式 3 に適用させることでレイリー分布の確率密度関数  $f(x)$  が求められる。この  $f(x)$  を用いて各観測値  $X_i$  の出現回数の推定値  $E_{X_i}$  を式 10 で算出する。

$$E_{X_i} = n f(X_i) \quad (10)$$

この  $E_{X_i}$  と各観測値の実際の出現回数  $O_{X_i}$  からカイ 2 乗検定に使われるカイ 2 乗値を算出する。カイ 2 乗値はカイ 2 乗検定に使用される値で、ある二つの分布形状が類似しているほど小さな値となる。カイ 2 乗値は式 11 で求められる。

$$\chi^2 = \sum \frac{(O_{X_i} - E_{X_i})^2}{E_{X_i}} \quad (11)$$

このカイ 2 乗値をオフセットを変化させながら算出し、最小のカイ 2 乗値となるオフセットを求め、これを採用する。オフセットの探索は rssi の最小値から最大値の範囲で変化させる。RSSI は整数の値で観測されるため、刻み幅は 1 とした。カイ 2 乗値を用いた手法をとるのは、ノンパラメトリックな手法であり、計算コストが少ないことからスマートフォンのようなデバイスに適しているためである。

採用されたオフセットから求められる確率密度関数を  $f_{area}(x)$  とする。図 4 に RSSI のヒストグラムの例とヒストグラムにより推定したレイリー分布の確率密度関数の例をプロットした例を示す。横軸の 1 段目はレイリー分布の確率変数の各値を表しており、2 段目は確率変数に対応する RSSI を表している。

エリア識別器に入力された Wi-Fi 電波ログの各 BSSID の RSSI の分布  $f_{loc}(x)$  も同様の方法で推定するが、オフセットの値は  $f_{area}(x)$  を推定する際に使用した値を使用

した。

上記のようにエリア識別器を構築したあと、学習データとして新しい観測が行われるごとにエリア識別器を更新する。もし、エリア識別器の出力が TRUE となった場合は入力された Wi-Fi 電波ログを併せて各 BSSID ごとの  $f_{area}(x)$  を再計算する。また、入力された Wi-Fi 電波ログにどの弱識別器にも割り当てられていない BSSID が含まれていたら、上記で述べたアルゴリズムで再度 BSSID を割り当てなおす。

その後は、各弱識別器に付加されている重み  $\omega$  を更新する。重みの更新は、学習データとして使う Wi-Fi 電波ログの内、すでにエリア識別器に入力され、どのエリアで観測された Wi-Fi 電波ログなのか識別済である Wi-Fi 電波ログのみを使用する。

本研究での重みの付け方は adaboost で用いられている計算式を用いた。使用する全ての Wi-Fi 電波ログを弱識別器に入力し、弱識別器が正しく識別した回数  $S$  と、弱識別器で識別された回数  $T$  を求め、式 12 を用いて弱識別器のエラー率  $e$  を算出した。

$$e = 1 - \frac{S}{T} \quad (12)$$

このエラー率  $e$  を式 13 に適用して重み  $\omega$  を算出し、新たな弱識別器の重みとした。

$$\omega = \frac{1}{2} \log \frac{1-e}{e} \quad (13)$$

#### 4.4 すべての Wi-Fi 電波ログによるエリア識別器群の構築

4.1 節で述べたように、取得する Wi-Fi 電波ログに、緯度経度や建物内での位置といった教師信号が付加されていないため、本研究におけるエリア識別器は Wi-Fi 電波ログを得るごとに逐次的に学習する。

本研究では、まず一つ目の静止期間の Wi-Fi 電波ログから一つ目のエリア識別機を構築する。以後の静止期間の電波ログは、構築済みの全てのエリア識別機で識別を行い、エリア識別器の出力が TRUE だった場合、該当エリア識別機を更新する。もし、すべてのエリア識別機の出力が FALSE だった場合には、新しいエリアとして、新たなエリア識別機を構築する。

### 5. エリア識別器の実験

#### 5.1 使用したデータ

実験に使用する Wi-Fi 電波の収集にはサムスン電子製の Nexus S を使用して収集した。Nexus S での Wi-Fi 電波のサンプリング周期は 3 秒に 1 回とした。データの収集は 2013 年 12 月 21 日に行い、収集場所は名古屋大学 IB 電子情報館南棟 3~5F の図 5, 図 6, 図 7 の白丸で示す観測点 a から x の計 24 箇所で行った。各階における各観測点間の距離は 5m である。観測点付近はアクセスポイントの数が



図 5 データ観測点 (3F)  
 Fig. 5 Data observation point (3rd floor)

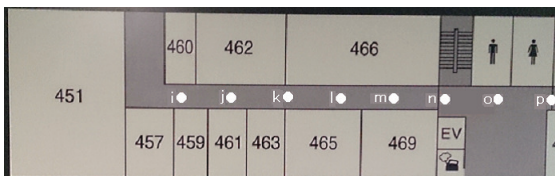


図 6 データ観測点 (4F)  
 Fig. 6 Data observation point (4th floor)

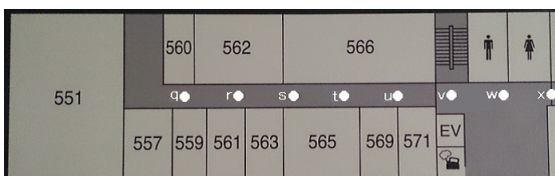


図 7 データ観測点 (5F)  
 Fig. 7 Data observation point (5th floor)



図 8 観測点の様子  
 Fig. 8 View of the observation point

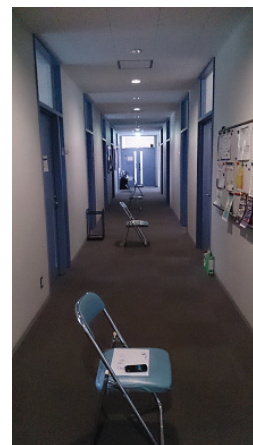


図 9 観測地帯の様子  
 Fig. 9 View of the observation zone

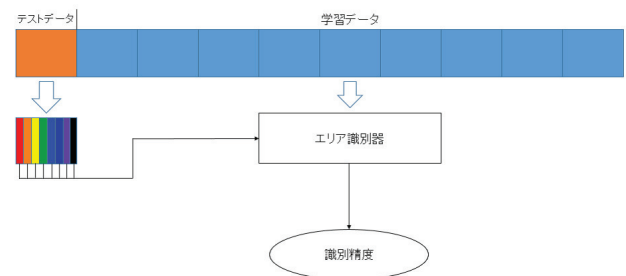


図 10 10 分割交差検定の概要  
 Fig. 10 Overview of 10-fold cross-validation

多く、常に 10 種類以上の Wi-Fi 電波を観測することができた。

ユーザが実際にデバイスを用いてデータを観測する場合は地面からある程度の高さの地点で観測する場合が多い。そこで観測点では図 8 のようにパイプ椅子の上に Nexus S を置いて 30 分間静止させた状態で Wi-Fi 電波を収集した。観測地帯の様子を図 9 に示す。

## 5.2 実験内容

本研究のエリア識別器はレイリー分布を用いて RSSI の分布を推定し位置識別する。本実験では RSSI の分布をレイリー分布で推定する場合とガウス分布で推定する場合での少数サンプリング時の識別精度を比較する。4.1 節でも述べたように本研究におけるエリア識別器は逐次学習して

利用することを想定している。しかし、逐次学習をしても同じデータを使用してもレイリー分布で推定した場合とガウス分布で推定した場合との間で、エリアの数や同一エリアでの弱識別器の重みが変わるためレイリー分布で推定した場合とガウス分布で推定した場合の識別精度の単純比較ができない可能性がある。そこで、本実験では逐次学習でエリアを生成するのではなく、エリア識別器を任意の地点で生成した。また、弱識別器の重みを初期値に設定して実験した。

精度の評価では実利用において短時間でエリア識別するという状況を考え、テストに用いるデータ数をサンプリング 1~5 回分にそれぞれ設定して実験をした。評価の方法として 10 分割交差検定を採用した。この実験における 10 分割交差検定の概要を図 10 に示す。実験は 3 つに分けて行った。

実験 1 では、ある一つの観測点で得られた Wi-Fi 電波ログを学習データとテストデータに分け、学習データを用いてエリア識別器を生成し、エリア識別することで True Positive(以下, TP) と False Negative(以下, FN) を求めた。テストデータをサンプルを観測した時間順に並べ、設定したサンプル数になるように複数に分割した。この分割したデータをそれぞれエリア識別器に入力し正しく識別できるかテストし、正しく同一エリアと識別できたテスト



データの数を求めた。これを、学習データとテストデータを変えて 10 回分言い、正しく同一エリアと識別できたテストデータの数を TP とし、同一エリアと識別できなかった数を FN とした。

実験 2 では、ある観測点で得られた Wi-Fi 電波ログ全てを用いてエリア識別器を生成し、その地点の両隣の観測点で得られた Wi-Fi 電波ログをテストデータとして実験 1 と同じく 1~5 回のサンプルに分割しエリア識別器識別することで True Negative(以下, TN) と False Positive(以下, FP) を求めた。ただし、観測点 a のように隣の観測点が一つかない観測点ではその隣の観測点のみをテストデータとした。分割したデータをエリア識別器に入力して正しく別エリアと識別されたテストデータの数を TN, 別エリアと識別できなかった数を FP とした。

実験 1, 2 で算出された TP, FN, FP を用いて式 14 で求められる適合率 Precision, 式 15 で求められる再現率 Recall, 式 16 で求められる F 値を用いてエリア識別器の評価をした。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (14)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (15)$$

$$F = \frac{2\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (16)$$

観測点 a,i,q の組や観測点 b,j,r の組などは位置的に真上または真下にある観測点の組である。そこで、実験 3 では各観測点の真上や真下の観測点とエリア識別を行い、他のエリアと判別されるか実験した。

全実験における、識別に用いる面積一致度の閾値は 0.6 とし、エリア識別はレイリー分布を用いた場合とガウス分布を用いた場合の両方で行った。

### 5.3 実験結果

レイリー分布を用いた場合とガウス分布を用いた場合の適合率, 再現率, F 値の平均(棒グラフ)と標準偏差(エラーバー)をそれぞれ図 11, 図 12, 図 13 に示す。

各観測点の真上や真下の観測点とエリア識別を行ったところ、レイリー分布を用いた場合とガウス分布を用いた場合の両方で、全ての観測点において、真上と真下の地点とは別のエリアと判別された。

### 5.4 考察

レイリー分布の F 値の平均とガウス分布の F 値の平均を比べるとレイリー分布の方が高い精度で識別できたことが確認できる。従って、分布形状にレイリー分布を用いた場合のほうが高い精度で識別できることが示唆された。さらに今回の実験では、少ないサンプル数でも高い精度で識

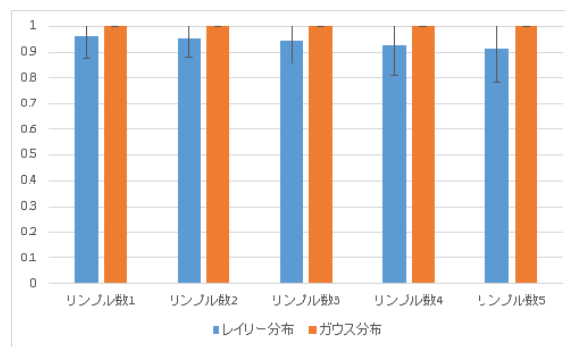


図 11 適合率の平均と標準偏差

Fig. 11 The mean and standard deviation of the precision

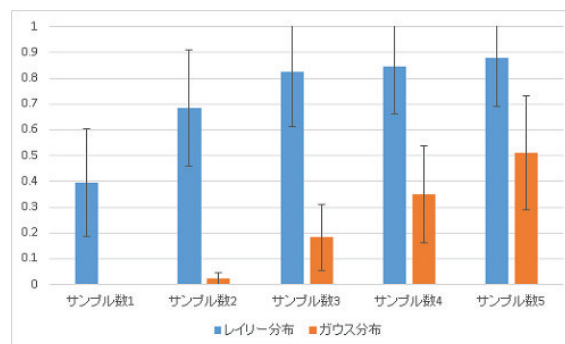


図 12 再現率の平均と標準偏差

Fig. 12 The mean and standard deviation of the recall

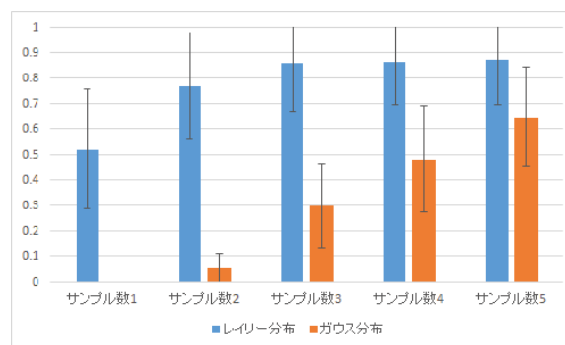


図 13 F 値の平均と標準偏差

Fig. 13 The mean and standard deviation of the F-measure

別できることが示された。従って、レイリー分布を用いることで短時間で高精度なエリア識別が可能であることが示唆された。

詳しく分析すると、レイリー分布によるエリア識別でも同エリア識別の実験において観測点 a での F 値が非常に低い結果となった。表 1 に各サンプル数ごとの観測点 a での適合率, 再現率, F 値を示す。この原因は、観測点 a において RSSI の分布が正しく推定できていないものが多かったことが挙げられる。正しく RSSI の分布を推定できていなかった例を図 14 に示す。ヒストグラムが実際に観測された RSSI の分布であり、曲線が推定された RSSI の分布を表している。図 14 において、レイリー分布の変数  $x$  と RSSI の対応が  $x = 0$  のとき  $\text{RSSI} = -45$  付近となる場合に正しく

表 1 サンプル数別の観測点 a での適合率, 再現率, F 値

Table 1 Precision, recall and F-measure of each sample for each number at the observation point “a”

|     | 1     | 2     | 3     | 4     | 5     |
|-----|-------|-------|-------|-------|-------|
| 適合率 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 再現率 | 0.003 | 0.017 | 0.042 | 0.093 | 0.075 |
| F 値 | 0.006 | 0.033 | 0.081 | 0.170 | 0.140 |

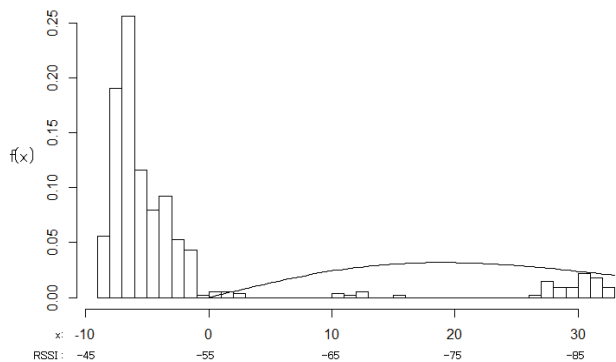


図 14 正しく RSSI の分布を推定できていない例

Fig. 14 Examples of failure to estimate the distribution of RSSI

推定できたといえる。しかし、この例では実際は  $x = 0$  のとき  $RSSI = -55$  となっており、実際に観測された RSSI のうち、8 割以上が受信できないものとして扱われていた。このような推定結果となった理由は RSSI の値が非常に小さい電波を複数受信してしまったためにオフセットの決定に大きな誤りが生じたと考えられる。このような誤りを防ぐためには推定に使うデータを制限したり、オフセットの決定方法を現在のカイ 2 乗値を用いたものから実際の分布と推定した分布の一致度から求める方法に変更するといった改善方法が有効だと考えられる。

RSSI の分布にレイリー分布を用いた場合とガウス分布を用いた場合の両方において真上の地点と真下の地点とは別の位置と識別された。これは学習に使用した観測点とテストに使用した観測点で観測された BSSID の種類が大きく異なっていたためであると考えられる。観測された BSSID が大きく異なるのは天井や床によって電波が遮断されているためであると考えられる。

## 6. おわりに

Wi-Fi 電波強度は周辺の環境の変化により同じ場所でも観測するごとに変化するため、一回の観測結果から位置を推定する方法では、推定精度が低下する問題があった。

そこで、本研究では、Wi-Fi 電波の揺らぎを考慮して分布同士の比較で位置推定をする場合に、小サンプル数でも分布を推定できるレイリー分布を用いることで、短時間でも高精度な位置識別が可能であることが示唆された。

今後の課題として、オフセットの決定方法の改善が挙げられる。今回の実験ではある 1 観測点において、オフセットの決定方法がよくなかったために識別精度が大きく低下してしまうエリアが存在した。そのため、今後はカイ 2 乗値によるオフセットの決定方法ではなく、分布の一致度を用いたオフセット決定方法をとって精度の向上を図る。また、今回はレイリー分布の比較対象にガウス分布を用いて実験したが、より実際の Wi-Fi 電波分布に近いとされている対数正規分布と比較検証を行う。また、実験では逐次学習をせず任意にエリア識別器を構築したので、逐次学習の有用性の評価を行なう。

謝辞 本研究は JSPS 科研費 24700062 の助成を受けたものである。

## 参考文献

- [1] 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. インターネットコンファレンス, Vol. 2006, pp. 95–104, 2006.
- [2] 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= Journal of Human Interface Society: human interface, Vol. 10, No. 1, pp. 15–20, 2008.
- [3] 谷内大祐, 前川卓也, 鈴木潤, 岸野泰恵. Boosting を用いた環境変化に頑健な wi-fi 屋内位置推定手法の提案. 情報処理学会研究報告. HCI, ヒューマンコンピュータインタラクション研究会報告, Vol. 2013, No. 5, pp. 1–7, 2013.
- [4] 岡本篤樹, 朝倉康夫, 内田敬, 近藤勝直, 田名部淳. Phs システムを用いた狭域での消費者回遊行動追跡のための位置特定手法. 土木計画学研究・講演集 Vol. 26, .
- [5] レイリー分布 - matlab & simulink - mathworks 日本. <http://www.mathworks.co.jp/jp/help/stats/rayleigh-distribution.html>.
- [6] Bernard Sklar. Rayleigh fading channels in mobile digital communication systems. i. characterization. *Communications Magazine, IEEE*, Vol. 35, No. 7, pp. 90–100, 1997.
- [7] 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.
- [8] 伊藤誠悟. Gaussian process particle filter を用いた無線 lan 位置推定手法. 情報学マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO 2011) 論文集, pp. 439–448, 2011.
- [9] 高島雅弘, 趙大鵬, 柳原健太郎, 福井潔, 福永茂, 原晋介, 北山研一. センサネットワークにおける受信電力と最尤法を用いた位置推定 (ネットワーク). 電子情報通信学会論文誌. B, 通信, No. 5, pp. 742–750, 2006.