

論文

# Wi-Fi パケットセンサと クラスター分析を用いた屋内での混雑度推定手法

荻原 崇 諏訪 敬祐

2011年に起きた東日本大震災や2020年に東京オリンピックが開催されることになったなど、屋内・地下での混雑度を推定するシステムの需要が高まっている。本研究では、Wi-Fi パケットセンサとクラスター分析を用いた混雑度推定手法の検討と提案を目的とする。具体的にはモバイル端末が放出するプローブ要求をキャプチャし、RSSI（受信信号強度）の揺れをクラスター分析する。そうして「範囲内にどの程度のグループ（集団）が存在するか」を分析し、範囲内に存在するテーブル数などに関連付けすることで混雑度推定手法の実質的検証を行った。大教室・食堂での評価実験を行った結果、RSSIの分散値を用いてクラスター分析することで有用なテーブル占有率を推定でき、空き状態はテーブル占有率12.1%、大混雑時はテーブル占有率93.1%のように有効な混雑率を定量的に明らかにした。

キーワード：Wi-Fi, 混雑度, クラスター分析, リアルタイム

## 1 はじめに

### 1.1 研究の背景

#### (1) 混雑度推定システムの需要向上

2012年10月1日に国土交通相が公表した「平成23年度の三大都市圏における鉄道混雑率について」では、三大都市圏全てにおいて混雑率が過去30年間の中で最も低い数値になっているなど、混雑の緩和が進んでいると思われる。しかし実際は、東京圏では2003年以降混雑率の低下が見られず、以前として180%以上の混雑率となっている区間が多く見られる。

また2014年3月にはJR東日本から山手線の電車内の混雑度を知ることができる「JR東日本アプリ」が、各観光地の混雑度をWebから確認できる「混んでる？.com」がそれぞれ公開され、多くのユーザーから利用されている。今後もこのようなシステム混雑度推定システムが公開、利用されていくと考えられる。

#### (2) 公衆無線LANサービス利用者の増加

ICT総研の「公衆無線LANサービス利用者動向調査」によると、2012年度末には公衆無線LANサービス利用者数1,269万人であり、2013年度末には約34%増の1,702万人に拡大する見通しである。この傾向は今後も続くと考えられており、2016年度には3,000万人

に近づくとも予測されている。

また本大学の学生を対象にWi-Fi機能利用状況に関するアンケート調査を行った。(図1)結果はスマートフォンのWi-Fi機能を常時利用している人が32%、頻繁に利用している人が45%と、75%以上の学生がWi-Fi機能を頻繁に利用していることがわかった。Wi-Fi機能を利用していない学生は1%のみであった。以上のことから、非常に多くの人々はスマートフォン端末のWi-Fi機能を利用していることがわかる。

#### (3) 無線LAN対応モバイル端末出荷台数の増加

同じくICT総研の「公衆無線LANサービス利用者動向調査」によると、2011年度は出荷台数3,749万台であったが、翌年の2012年度は4,000万台を超えた。今後もWi-Fi通信機能モバイル端末は増え続けると考えられ、2016年度の年間出荷台数は5,423万台に達すると予測されている。より多くの人々がWi-Fi機能搭載モバイル端末を所持し、利用すると考えられる。

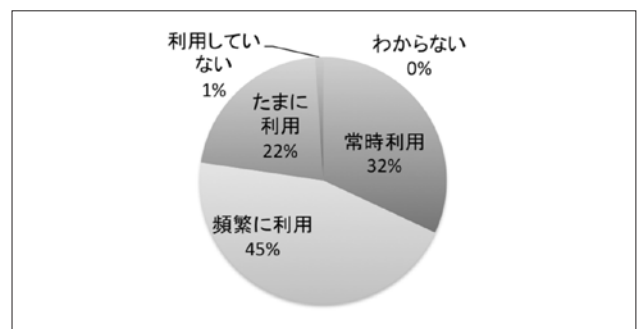


図1 Wi-Fi利用状況に関する調査結果 (n=97)

OGIWARA Takashi

東京都市大学 環境情報学部 情報メディア学科 2014年度卒業生  
SUWA Keisuke

東京都市大学 メディア情報学部 情報システム学科 教授

## 1. 2 研究の目的

Wi-Fi パケットセンサとクラスター分析を用いた混雑度推定手法の検討と提案を目的とする。具体的に、モバイル端末が放出するプローブ要求をキャプチャし、RSSI (受信信号強度) の揺れをクラスター分析する。そうして「範囲内にどの程度のグループ (集団) が存在するか」を分析し、テーブル数などに関連づけすることで混雑度推定手法の実質的検証を行う。

## 2 既存サービス, 研究について

### 2. 1 既存サービスについて

#### (1) 音波を利用した混雑度推定サービス

2014年3月にJR東日本がリリースした「JR東日本アプリ」に、音波を利用した電車内の混雑度を推定する機能がある。電車内に音波送信機を設置し、その音をスマートフォンのマイクが拾い、減衰の仕方などから混雑度を推定している。しかし、マイク機能を解放する必要があるため、専用のアプリケーションをインストール・利用する必要がある。

#### (2) Twitter を利用した混雑度推定サービス

Twitter 上のツイートから全国の各目的地の混雑状況を推定する「混んでる?.com」というサービスがある。ツイートをリアルタイムで分析しているため、常に最新の混雑状況を確認することができる。しかし、常に「混んでいる」の割合が高い傾向にあるなど、精度に問題がある。

### 2. 2 既存研究について

#### (1) 二酸化炭素センサによる混雑度推定

中村ら (2008) が行った、二酸化炭素の濃度によって電車内の混雑度を推定する研究がある。スマートフォンなどの特別な端末を必要とせず、ユーザーそのものを発信源としているため、センサを置くだけで混雑度推定を行うことができる。しかし、一度二酸化炭素濃度が高くなると、人が減っても濃度は高いため、精度に問題がある。

#### (2) 加速度センサを用いた混雑度推定手法

米村ら (2013) が行った、スマートフォンの加速度センサを用いた混雑度推定手法に関する研究がある。加速度センサから人々の歩幅を検出し、その違いから混雑度を推定した。結果、混雑していない状態を 83.0%、混んでいる状態を 77.5% の精度で推定することができる。しかし、スマートフォンに専用のアプリケーションをインストールし起動する必要があるため、実用性が低いという問題がある。

#### (3) 無線 LAN アクセスポイントへの探索要求を用いた屋内混雑度推定手法

中野ら (2013) が行った、スマートフォンなどのモバイル端末から放出される探索要求 (プローブ要求) をキャプチャすることで、混雑度を推定する研究がある。探索要求によって送信元の Mac アドレスや RSSI (受信信号強度) を取得し、混雑度を推定している。電車内と大学の大教室で有用性が確認されている。しかし、人々が自由に出入りできる環境での有用性が確認されていない、ユーザーへの混雑度に関する情報量が少ないといった問題がある。

### 2. 2 本研究の既存サービス・研究との違い

既存のものとの違い・特徴を大きく三つある。一つ目はユーザーの特別な操作を必要とせずに、リアルタイムで推定が可能であること。二つ目はクラスター分析を用いることでテーブル占有率という新たな付加価値を創造すること。三つ目は人々が自由に出入りできる環境での有用性の示唆を示すことである。

## 3 提案手法

観測場所内のモバイル端末から放出されるプローブ要求を三つのセンサを用いてキャプチャし、各端末の RSSI の揺らぎからクラスター分析を行う。そうして、端末数とクラスター数の推定を行う。クラスター数とは、人同士が近くに集まり、固まっている集団のことである。

RSSI とは、受信した電波の強度を示す指標であり、人の有無や周辺環境に大きな影響を受ける性質を持っている。しかし、阿瀬川らの研究により人混みが発生すると RSSI の分散が大きくなること、進藤らの研究により RSSI の大きさと人数の間には負の相関関係があることが示されている。そのため、距離の近い端末は似た減衰・揺らぎで RSSI が推移するという仮説を立て、似たもの同士を分類するクラスター分析を行うことで有用性の検証を行う。

### 3. 1 混雑度の定義

本研究の混雑度の定義として、テーブル占有率の値から表 1 のように設定する。

また、テーブル占有率とは観測場所内にあるテーブル

表 1 テーブル占有率と混雑度の対応

| テーブル占有率 | ~20% | 21%~50% | 51%~80% | 81%~ |
|---------|------|---------|---------|------|
| 混雑度     | 空き   | 小混雑     | 中混雑     | 大混雑  |

が人々によりどの程度埋まっているかを示す割合であり、以下の計算式で求める。  
 テーブル数：

$$\text{テーブル占有率} = \frac{\text{クラスター数の推定値}}{\text{テーブル数} \times \text{テーブルの列数}}$$

観測場所範囲内にあるテーブルの数。テーブル間隔が空いている部分を分岐点とする。

テーブルの列数：

一つのテーブルにおいて、人の座る列数のことである。

クラスターの推定値：

クラスター分析の結果値である。

本研究では、テーブルが規則的に並んでおり、人々がそのテーブル前にある椅子に座る環境を観測場所とする。このような環境にある場所の例として、「大学などの大教室」「食堂」「フードコート」などが挙げられる。

### 3. 2 クラスター分析

クラスター分析とは、異なる性質を持った要素が集まった集団の中から、互いに似たものを集めてクラスター(集団)を作り、その要素を分類する分析手法のことである。本研究では、最適なクラスター数を自動で推定し、クラスター分析を行うことができる「X-Means 法」を用いている。図2にクラスター分析のイメージ図を示す。

### 3. 3 プローブ要求

モバイル端末は定期的にアクセスポイントを検索するために、プローブ要求と呼ばれるパケットを放出している。このプローブ要求には様々な情報が取得可能であり、本研究では「放出された時間」「放出元の端末の Mac アドレス」「RSSI」の三つの情報を用いて分析を行う。

### 3. 4 システム構成

システムの流れを図3、システム構成を表2に示す。各モバイル端末から放出されるプローブ要求を三つの

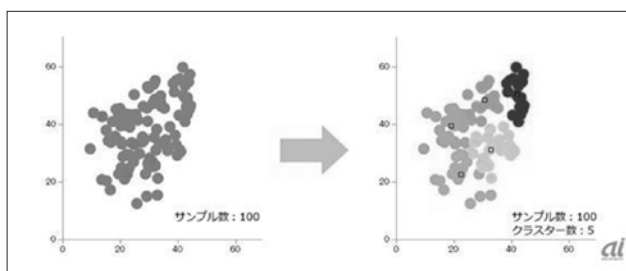


図2 クラスター分析のイメージ図

Wi-Fi パケットセンサを用いてキャプチャし、データを10秒間隔でサーバー上のデータベースへ登録する。その後、データ分析・可視化を行う。

### 3. 5 データ分析

データ分析は全てスクリプト言語 Python を用いて行った。Python には「Numpy/Scipy (数値演算ライブラリ)」「pandas (データ分析環境)」「matplotlib (グラフ/可視化ライブラリ)」といったデータ分析に関するライブラリがあり、ノイズ除去から分析、可視化まで行うことが可能である。いかに分析の流れを説明する。

#### (1) ノイズ除去処理

サーバーからパケットデータを取得後、ノイズ除去処理を行う。ノイズ除去処理ではまず観測場所範囲外のモバイル端末から放出されたパケットデータを除去する。アルゴリズムとして、三つのパケットセンサ全てからキャプチャされなかったデータを除去している。パケットセンサは約30～40m程の間隔で三角形に置かれており、観測範囲外のモバイル端末から放出されるパケットデータを三つのパケットセンサ全てがキャプチャすることは少ない。

次にモバイル端末以外の無線端末から放出されたパケットデータを除去する。キャプチャデータには Mac アドレス情報が含まれており、その情報からベンダーを知ることができる。もしベンダーからモバイル端末ではないと判断した時、そのデータを除去する。本研究では、以下のベンダーの Mac アドレスのみを残し、その他をノイズとして除去している。表3にモバイル端末として登録したベンダーを示す。

次に同じ端末から放出されたデータ数が2未満の時、その端末のデータをノイズとして除去する。理由は、本研究では各モバイル端末から放出されたパケットデータの RSSI の「標準偏差」「分散」「平均」を用いるためである。このような値はデータ数が2未満では求める

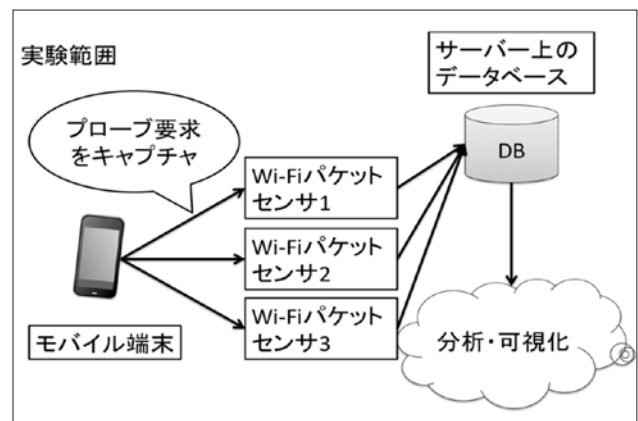


図3 システムの流れ



表 3 ベンダーリスト

|         |         |           |
|---------|---------|-----------|
| Sony    | Lenovo  | Panasonic |
| Casio   | Nec     | Fujitsu   |
| Pantech | Apple   | Kyocera   |
| LG      | Samsung | Motorola  |
| Toshiba | Htc     |           |

ことができない。

(2) データの整形・分析処理

ノイズ除去後、クラスター分析を行うためのデータを準備する。具体的に各モバイル端末の packets データの RSSI から「標準偏差」「分散」「平均」を計算する。このような値を利用する理由として、「標準偏差」「分散」は変量の散らばりを示すものであり、既存研究において散らばりに関して何かしらの法則が示唆されていたためである。また「平均」については、集団の代表値を示す値であるためである。検証実験において、「標準偏差」「分散」「平均」のどの値を用いることが妥当であるかも検証していく。計算後、X-Means 法を用いてクラスター分析を行う。分析結果のクラスター数からテーブル占有率を計算し、可視化アプリにプロットする。

(3) 可視化アプリ

可視化アプリについて、アプリ上には主に 3 つの情報が表示される。1 つ目はクラスター分析結果を表す散布図である。この散布図を表示することでユーザーはより直感的に混雑度を知ることができる。2 つ目は予想テーブル占有率である。この値により、ユーザーはどの程度のテーブルが埋まっているかを知ることができる。3 つ目は混雑度度数である。この度数もテーブル占有率と同様、簡単に混雑度を知ることができる。

図 4, 図 5 に可視化アプリの混雑していないとき、混雑しているときの画面を示す。

4 評価実験

本大学内の「大教室」「食堂」において、評価実験を行った。以下に実験環境や結果について述べる。

4.1 本大学内の大教室

大教室では諏訪教授の協力のもと、実際に講義が行われている中、評価実験を行った。大教室での実験は計 2 回行った。

(1) 実験環境

テーブルが 45 つ、椅子は 133 席あり、その他にホワ

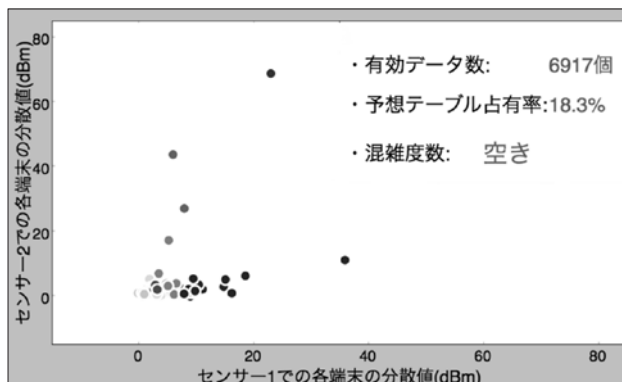


図 4 混雑していないとき

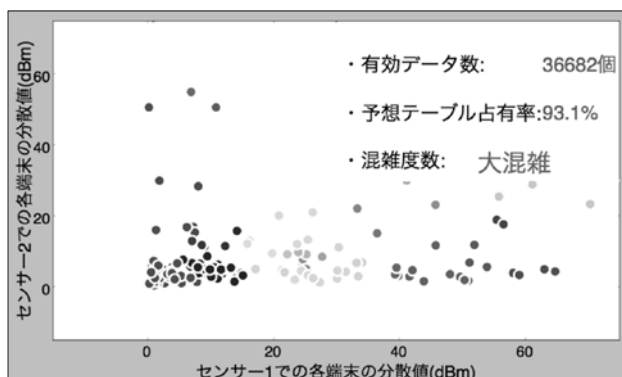


図 5 混雑しているとき

イトボード、プロジェクターなどがある。テーブル列数は 1 列である。学生が多くいるが講義中であり、移動がない環境である。図 6 に大教室のレイアウトとパケットセンサの設置場所を示す。また、下記に各実験の詳細情報を示す。

実験 1

- ・ 検証日 : 10/9
- ・ 収集時間 : 約 20 分
- ・ 在室人数 : 114 人
- ・ 理想値 : 36
- ・ 混雑度 : 大混雑

実験 2

- ・ 検証日 : 11/26
- ・ 収集時間 : 約 20 分
- ・ 在室人数 : 35 人
- ・ 理想値 : 27
- ・ 混雑度 : 小混雑

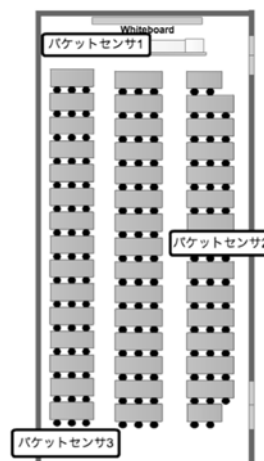


図 6 大教室のレイアウトとパケットセンサの配置場所

(2) クラスター数の理想値

クラスター分析結果の評価のため、クラスター数の理想値を設定する。本環境での理想値は「大教室内のテーブル占有率」とした。このような値を設定した理由は、

テーブル占有数に近いクラスター数であると、精度の良いテーブル占有率を推定することが可能であるためである。学生に利用されているテーブル数に近いクラスター数が分析される程、有用な結果と言える。

(3) 実験結果

分析結果と理想値との誤差と割合を図7に示す。2つの実験における差の割合の平均は標準偏差値が-71.8%、分散値が-19.0%、平均が-52.8%と分散を用いたクラスター数が良い結果となった。混雑率・混雑度数について、実験日11/26については分散値で、混雑率48.9%、小混雑と良い推定結果が出ている。しかし、実験日10/9については、大混雑時にも関わらず分散値での推定混雑率64.4%、推定混雑度数中混雑と、実際より低い推定結果となった。

次に時間毎のモバイル端末検出数の割合を図8に示す。約5分(300秒)で70%以上の観測場所範囲内のモバイル端末を検出することができた。割合については、一人一台のモバイル端末を保持していると仮定し、検出端末数/全体人数で求めている。

4.2 本大学内の食堂

食堂ではランチ前で混雑していない時間帯、ランチ途

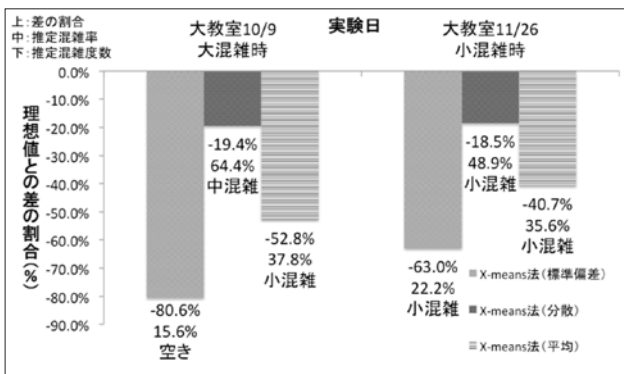


図6 大教室でのクラスター数と理想値の差と割合

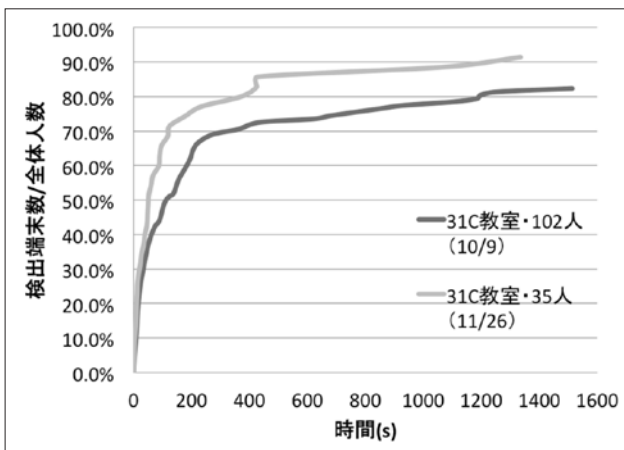


図7 時間毎のモバイル端末検出の割合

中で混雑している時間帯において計6回実施した。

(1) 実験環境

テーブル数は29つ、席数は264席あり、テーブル列数は2列ある。食堂であることから学生は移動制限がなく、自由に席を移動することができる環境である。また学生からの特別な操作をしてもらうといった協力はしてもらわず、普段の生活状態で検証実験を行った。図8に食堂のレイアウトとパケットセンサの配置場所を示す。また、表4に各実験の詳細情報を示す。

(2) クラスター数の理想値

大教室と同様、クラスター分析結果の評価のため、クラスター数の理想値を設定した。食堂においては「テーブル一列において、1席以上空いた点」を分岐点とし、その基準でのクラスター数を理想値と設定した。この理

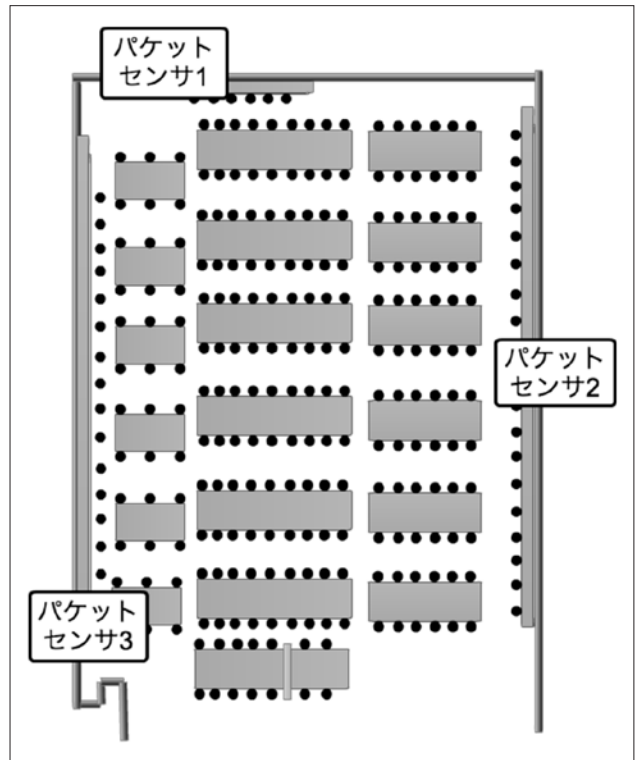


図8 食堂のレイアウトとパケットセンサの配置場所

表4 各実験環境の詳細情報

| 検証日    | 在室人数 | 理想値 | 状態         |
|--------|------|-----|------------|
| 12月12日 | 15   | 12  | 混雑前, 歩行者なし |
| 12月12日 | 69   | 46  | 混雑前, 歩行者多数 |
| 12月12日 | 68   | 51  | 混雑時, 歩行者少数 |
| 12月16日 | 13   | 9   | 混雑前, 歩行者少数 |
| 12月16日 | 108  | 58  | 混雑時, 歩行者少数 |
| 12月16日 | 144  | 68  | 混雑時, 歩行者少数 |

想値に近いクラスター結果が得られるほど、精度の高いテーブル占有率を推定することができる。

### (3) 実験結果

分析結果のクラスター数と理想値の差の割合とテーブル占有率を図9に示す。①は1回目、②は2回目を示す。全ての実験における差の割合の平均は標準偏差値が58.8%、分散値が23.5%、平均値が69.1%と分散を用いたクラスター数が良い結果となった。また、差が大きく出た実験日12/12の中混雑時1回目に行った実験の値を除くと、12.1%と85%以上の精度で分析を行うことができている。このことから、各端末のRSSIの分散値を用いてクラスター分析することで、理想値に近い分析結果を得ることができ、有効なテーブル占有率を推定することが可能であることがわかった。混雑度についても、実験日12/16を例に、空いている状態はテーブル占有率15.5%、大混雑時はテーブル占有率93.1%と推定することができている。

## 5 考察

「大教室」での実験結果について、験日11/26を例に混雑率48.9%、混雑と各モバイル端末RSSIの分散値を用いたクラスター分析を用いることで、有効なテーブル占有率を推定することが可能であると考えられる。実験日10/9において、推定値が実際よりも少なくなってしまうことについて、Wi-Fi機能を無効にした端末が多くあったことが原因ではないかと考える。Wi-Fi機能が無効であるとプローブ要求は放出されないため、本システムでは検知することができない。しかし背景に多くあるように、これからWi-Fi機能を利用するユーザーが増えていくことが予想されているため、この問題は解決されていくと考える。

時間毎のモバイル端末の検出率について、約5分程度で70%以上の範囲内でのモバイル端末を検出するこ

とができていることから、リアルタイム性の高い混雑度推定システムの構築が可能であると考えられる。

「食堂」での実験結果について、「大教室」での結果と同様、実験日12/16を例に、分散値を用いることで空いている状態はテーブル占有率15.5%、大混雑時はテーブル占有率93.1%と推定することができている。実験日12/12の中混雑時1回目において、大きな誤差が生じてしまった。原因として、歩行者が多いたことが挙げられると考える。歩行者が持つモバイル端末から放出されたパケットを用いてクラスター分析した場合、RSSI値はその都度変動しているため、違うクラスターとして分析される。その結果、理想値よりも多くのクラスター数に分割されたと考える。

分散値の平方根である標準偏差を用いたクラスター分析結果が有用ではない理由を述べる。標準偏差については、数値の範囲が分散と比べ小さいためだと考える。X-Means法では各数値の距離が近い物をグループとしてまとめる事から、その距離の範囲が広い分多くのグループに分けられる。実際に人数が少ない混雑前の食堂での実験では、分散値だけではなく、標準偏差値を用いたクラスター分析結果も良い精度であった。しかし、混雑時の食堂での実験結果では、非常に大きな誤差が生じた。

## 6 おわりに

### 6.1 まとめ

本研究ではモバイル端末から放出されるプローブ要求をキャプチャし、各端末のRSSIからクラスター分析を行った。その結果、RSSIの分散値を用いてクラスター分析を行うことで、有用なテーブル占有率を推定することができ、混雑度推定システムとしての付加価値を創出することができた。また空いている状態はテーブル占有率12.1%、大混雑時はテーブル占有率93.1%などと有効な混雑率を推定できた。リアルタイムで視覚的な分析結果を表示することで、混雑度推定システムの有用性を高めることができた。

### 6.2 今後の展望

今後の展望としては歩行者をノイズとし、除去することによる精度向上と、範囲外へ移動した端末を除去するアルゴリズムの検討を行う。

## 謝辞

本研究にて、様々な協力をして頂いた、本研究室修士2年生の猪俣史也氏、学部4年生の田中健氏、諏訪研究室の各位に心より御礼の言葉を申し上げます。

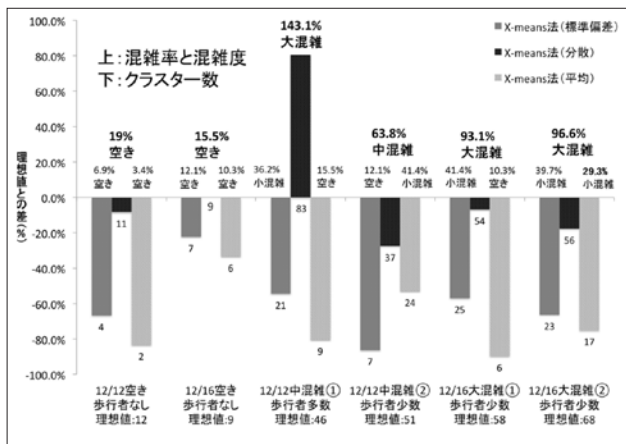


図9 クラスター数と理想値の差の割合とテーブル占有率