

# Wi-Fi パケットセンサを用いた 代表的観光パターン抽出の検討

明光 就平<sup>1</sup>・倉内 文孝<sup>2</sup>・伊藤 伸<sup>3</sup>

<sup>1</sup>学生会員 岐阜大学大学院 自然科学技術研究科環境社会基盤工学専攻 (〒501-1193 岐阜市柳戸1-1)

E-mail: x4523037@edu.gifu-u.ac.jp

<sup>2</sup>正会員 岐阜大学教授 工学部社会基盤工学科 (〒501-1193 岐阜市柳戸1-1)

E-mail: kurauchi@gifu-u.ac.jp

<sup>3</sup>非会員 西日本高速道路エンジニアリング関西株式会社 土木事業本部 交通技術部

E-mail: s\_ito@w-e-kansai.co.jp

近年我が国における観光流動はインバウンド観光を中心に増加の一途をたどっており、それに対応する交通施策が求められている。しかし、季節変動や経年変化が大きい観光流動をアンケート調査等の従来の調査方法で把握することには限界がある。一方、近年では情報通信技術の発達により、交通の分野においても位置情報を常時観測可能なビッグデータの蓄積が始まっている。本研究では観光地において多くの観光客が利用する経路（ゴールデンルート）を特定することを目標として、従来手法では調査が困難であった移動のつながり（トリップチェーン）の抽出方法の開発を試みる。本稿では、系列パターンマイニングを採用し最初観測地と途中訪問地との関係を分析することで出発地ごとの流動特性に着目し分析を行う。

**Key Words:** *Wi-Fi packet counter, Golden route, Pattern mining*

## 1. はじめに

OD 交通量などの人がどこからどこへ移動し、どの程度滞在するのかが示すデータである移動データは、交通計画の立案や評価、災害発生時の避難計画の策定などに利用される重要な基礎情報である。移動データの収集は従来、道路交通センサス、パーソントリップ調査といったアンケート形式の調査により取得することが一般的である。しかし、これらの調査はデータの収集を人手に頼っており、調査人件費の増大が問題となっている。さらに、近年の調査では詳細な移動データを取得するために調査項目が細分化する傾向にあり、調査対象者への負担の増大、そして個人情報保護に対する意識の向上などの複合的な要因により取得率が低下するという状況も発生している。一方で、近年では情報通信技術の発展により様々なビッグデータの収集が実施されており、交通の分野でも、携帯電話基地局情報を用いた人口統計情報であるモバイル空間統計、ETC2.0 対応車載機を用いて自動車の経路情報や挙動情報を含むデータを収集する ETC2.0 プローブデータ、そして Wi-Fi 通信機器の位置情報を収集する Wi-Fi パケットセンサデータなどが存在する。これら交通ビッグデータは従来の交通流動調査では収集が困難であった詳細かつ長期間の移動データを収集可能であり、注目が高まっている。本研究では、これらの交通ビッグデータのうち、Wi-Fi パケットセンサデータに着目する。Wi-Fi パケットセンサはスマートフォンなど

の Wi-Fi 通信機器がアクセスポイント探索の際に発信される識別子を収集することで Wi-Fi 通信機器の移動の推移から移動や滞留のデータ収集を可能とした装置である。交通調査手法としては設置が容易かつ低コストで調査対象者、実施者ともに大きな負担をかけることなく長期間の観測が可能であり、さらに収集データは交通手段によらないことから幅広い用途に活用可能である。そのため、現在では Wi-Fi パケットセンサがその名称が一般名詞として活用されるほど国内外で幅広い調査に活用されている<sup>1)</sup>。しかし、Wi-Fi パケットセンサの特性上、収集されるデータは Wi-Fi 通信機器の位置情報であることから実際の交通量との関係が不明であるなどデータの利用には課題も多い。

本研究では Wi-Fi パケットセンサデータを用いて観光客がどの程度の人数、どのように移動しているのか等の情報を把握すること（観光流動把握）を目的とし、系列パターンマイニングを採用し、Wi-Fi パケットデータの中から頻出するトリップパターンの抽出を試みる。

本論文の構成は以下のとおりである。1.では、本研究の背景と目的を述べた。2.では Wi-Fi パケットセンサの概要と既往の研究についてまとめる。3.では使用したデータの詳細について述べる。4.では Community Detection 法を用いたセンサ位置のグルーピングについて解説する。5.ではトリップチェーン抽出手法である系列パターンマイニングについて述べる。最後に 6.では本研究で得られた知見をまとめ、

今後の課題について整理する。

## 2. Wi-Fi パケットセンサ

### (1) Wi-Fi パケットセンサの概要

携帯電話やスマートフォンなどの Wi-Fi 通信機器は、Wi-Fi のアクセスポイント探索のために、30 秒から 2 分程度の間隔で Probe Request とよばれる信号を発信している。そして、Probe Request には Mac アドレスとよばれる機器固有の識別子が含まれる。

Wi-Fi パケットセンサはこの Probe Request を受信し、Mac アドレスとタイムスタンプ、センサの位置情報、受信時の電波強度などの情報を関連付けたデータを作成することにより Wi-Fi 通信機器の位置情報データの収集を可能とした装置である。Mac アドレスとその位置情報、取得時刻の 3 点が収集可能であることから Wi-Fi パケットセンサを複数地点に設置することによって Wi-Fi 通信機器の移動の推移が取得可能となり、移動データとして扱うことができる。なお、前述のように Mac アドレスは各機器固有のアドレスであり、その移動軌跡の収集は個人の特定に繋がる恐れがある。そのため、個人情報保護の観点より Mac アドレスは収集段階で A-Mac アドレスとよばれる匿名のアドレスに変換される仕組みとなっている (図 2-1)。このことから Wi-Fi パケットセンサは AMP センサ (Anonymous Mac-address Probe センサ) ともよばれる<sup>2)</sup>。なお、匿名であっても長期間収集する場合は移動パターンより個人が特定可能となる可能性がある。そのため、今回使用したセンサではさらに毎週月曜日の 0 時に A-Mac アドレスの更新を実施することで同一アドレスについて、1 週間以上追跡ができない仕様となっている<sup>3)</sup>。

Wi-Fi パケットセンサは、匿名化された Mac アドレスを自動的に収集するため、膨大な数のデータを得ることができる。しかし、匿名であるがゆえに課題も存在する。Wi-Fi 通信機器には、FAX や据え置き PC など (固定パケット) も含まれるため、人の移動でないデータも収集してしまう。さらに、移動手段や目的についても匿名データであることから、特定することはできない。また、Wi-Fi 通信機器を所有していればデータが取得できるため、スマートフォンとゲーム機といったように複数所有している人を区別することも困難である。

このように Wi-Fi パケットセンサには、多くの利点がある一方で課題も存在する。それらを従来手法とあわせて表 2-1 に示す。従来手法において欠点とされてきた情報収集の際にかかる手間やコストという課題に対し、Wi-Fi パケットセンサは、無人により 24 時間観測が可能であり、調査対象の移動手段に影響を受けない点において有用であるといえる。ただし、データ利用に関して上記のような課題が存在することが現状といえる。

### (2) 関連する既往研究の整理

Wi-Fi パケットセンサを活用した研究はここ数年、

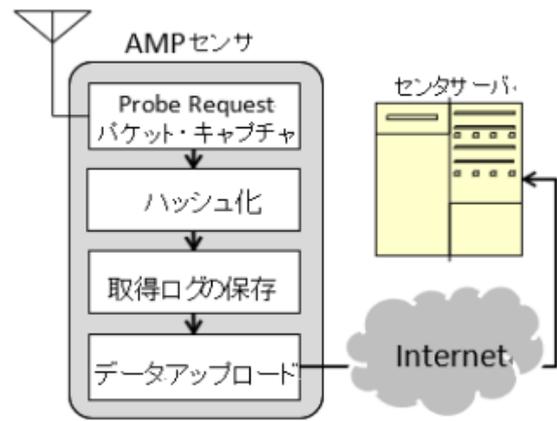


図 2-1 パケット取得のメカニズム

表 2-1 主な人の移動データ収集手法の利点と欠点

	利点	欠点
アンケート調査	多くの情報が取得可能	多数の調査員が必要
人手による方法	高価な測定機器は不要	調査員の確保、事前教育が必要
車両感知器による方法	設置地点の連続観測が可能	車両以外の観測は困難
Wi-Fi パケットセンサ	無人により 24 時間観測が可能	移動目的や手段は不明

活発に実施されている。本節ではそれらの取り組みについて紹介する。Wi-Fi パケットセンサ活用の手法として交通流動の把握が挙げられる。これは、複数の Wi-Fi パケットセンサを設置することでセンサ間の移動を長期間とらえることで、従来の紙ベースの手法では把握が困難であった移動の実態を把握しようという試みである。Wi-Fi パケットセンサの開発者である森本らは、Wi-Fi パケットセンサ 20 台をグランフロント大阪に設置した上で 2 ヶ月間の連続観測を実施した。その結果として、収集データを用いて人流解析とその可視化が可能であることを示している。より大規模かつ長期間観測の事例としては浅尾ら<sup>4)</sup>の研究があげられる。この研究では、海の京都観光圏とよばれる京都府北部に Wi-Fi パケットセンサを最大 60 台設置し、年単位の連続観測を実施した。その結果としてデータ蓄積により、さまざまな分析の可能性が拓けることを示唆している。また、本研究と同様に観光流動把握を目的としてトリップパターン抽出を試みた例として有村ら<sup>5)</sup>、伊藤ら<sup>6)</sup>の論文がある。これらの論文では系列パターンマイニングという手法を用いて各スポット間の訪問羅列を考慮した分析を実施している。その結果として系列パターンマイニングの適用により周遊パターンの全体像を把握できることを示している。しかし、Wi-Fi パケットセンサによる取得データには先述のようにデータが匿名であるがゆえの課題点も存在する。伊藤ら<sup>7)</sup>は、海の京都観光圏で取得された Wi-Fi パケットセンサデータを用いて、匿名データからの属性分類を目的としたクラスタ分析によるデータ

分類を実施した。その結果としてクラスごとに異なる移動の傾向があらわれることから匿名データである Wi-Fi パケットセンサデータの分類にクラス分析が有効であることを示している。

(3) 本研究の位置づけ

Wi-Fi パケットセンサは全国で実証実験が実施され、交通ビッグデータとしての実用化に向けた模索が実施されており、上記のように、人の行動把握が可能といえるようになりつつある。本研究では、Wi-Fi パケットセンサによるデータのさらなる活用方法として、トリップパターンの抽出を試みることで、京都駅周辺の広域な観光流動パターンを明らかにすることを旨とする。同様の検討を行っている文献 5)、6) の分析では途中訪問地と最終観測地の関係を容易に把握できることを示しており終着地点に着目したパターン抽出が検討されている。これに対し本研究では、最初観測地と途中訪問地との関係を分析することで出発地ごとの流動特性に着目し分析を行う。また、抽出パターンの評価において、単一の評価指標が上位のパターン、もしくは各評価指標が平均以上のパターンという相対評価に留まっているため、本研究では主成分分析を行い、複数の指標からより総合的に代表的といえるパターン抽出を行う。

3. データ概要

(1) 使用するデータについて

本研究では京都市が観光流動の把握を目的として 2018 年 11 月より計測した観光流動調査データを用いる。本調査では、京都駅を中心に主要観光地や鉄道駅等に 39 個の Wi-Fi パケットセンサを設置し観測が行われた。詳細な設置位置については 図 3-1 および表 3-1 に示すとおりである。

(2) データクレンジングおよび集計

先述のように Wi-Fi パケットセンサデータは匿名であることから、例えば目的別の交通流動の把握を実施するためには得られた情報を用いた分類などの工夫が求められる。そこで、移動傾向より調査対象者の属性を推定する手法として Wi-Fi パケットセンサデータの特性を踏まえた上で非階層クラスターリングによる分類を行う。

クラスターリングを実施するにあたり、事前処理として、移動データとして取り扱うことができない「1 回のみしか観測されていないアドレス」および調査結果に影響を与える可能性がある「ランダム化されたアドレス」を除去した。そして、データは 1 パケットごとに観測されているためそのままではクラスターリングを実施することができない。そこでアドレスごとに観測された時間、地点（捕捉されたセンサ）などの情報を元に表 3-2 に示した項目で集計した。



図 3-1 センサ設置地点

表 3-1 設置センサ名とセンサ番号

1	清水寺	21	バス停:祇園(南行き)
2	銀閣寺参道	22	バス停:五条坂
3	錦市場	23	円山駐車場
4	二条城	24	高台寺駐車場
5	嵐山渡月橋	25	清水坂駐車場
6	先斗町	26	JR京都 地下中央口
7	八坂神社	27	JR京都 西洞院口
8	高台寺	28	JR京都 北口広場
9	平安神宮	29	JR山科
10	伏見稲荷大社裏参道	30	JR東福寺 乗換口
11	市営-東山三条	31	JR東福寺 改札口
12	市営-三条京阪	32	JR稲荷
13	市営-京都(中央)	33	JR二条
14	市営-京都(南)	34	JR円町
15	京阪-三条	35	JR嵯峨嵐山
16	京阪-四条	36	JR京都 西口
17	京阪-五条	37	JR京都 中央口
18	阪急-河原町	38	JR京都 八条東口
19	知恩院駐車場	39	JR京都 地下東口
20	バス停:祇園(北行き)		

表 3-2 データの集計項目

分類要因	概要
観測時間	同一アドレスについて最後に観測された時刻と最初に観測された時刻の差
総観測数	同一アドレスが捕捉された回数
捕捉日数	同一アドレスが捕捉された日数
捕捉センサ数	捕捉されたセンサの数
第一観測時	最初に観測された時刻
最終観測時	最後に観測された時刻
時間帯別観測割合	ある時間帯の観測回数を総観測数で除したものの 1時間ごとに 24 要因
曜日別観測割合	ある曜日の観測回数を総観測数で除したものの 各曜日に 7 要因
センサ別観測割合	あるセンサにおける観測回数を総観測数で除したものの センサ設置地点とし 20 要因

### (3) クラスタリング手法

データの分類手法として K-Means 法による非階層クラスタリングを採用する。K-Means 法を用いる場合、クラスタ数をあらかじめ設定する必要がある、その設定が分類に大きな影響を与える。今回はクラスタ数設定の指標としてエルボー法とよばれる指標を採用する。これは、横軸にクラスタ数、縦軸にその二乗誤差和 (SSE 値) をプロットした図を作成し、その誤差がそれ以上小さくならない点をクラスタ数と設定する手法である。SSE 値は次の式で求めることができる。

$$\begin{aligned} D_k &= \sum_{X_i \in C_k} \sum_{X_j \in C_k} \|X_i - X_j\|^2 \\ &= 2n_k \sum_{X_i \in C_k} \|X_i - \mu_k\|^2 \end{aligned} \quad (3.1)$$

K-Means 法によって分類されたクラスタより、平均観測時間や平均補足日数が少なくアドレス数が多いクラスタを抽出することで観光客の所持する Wi-Fi 機器だと考えられる移動データに限定してトリップパターン抽出を行う。

分類結果の詳細については、発表時に報告する。

## 4. センサのグルーピング

本研究で取り扱う Wi-Fi パケットセンサの設置位置は京都駅周辺や東山地区周辺のように近接しているものから、嵐山方面や JR 山科駅のように非常に離れて設置してあるものまで様々であり設置密度は一様ではない。近接し合うセンサ間程、双方で捕捉される確率は高くなるため広域な観光流動よりも狭域での移動が相対的に多く見えてしまう。そのため、このまま頻出するトリップパターンの抽出を試みると近接するセンサ間の移動が多く取り上げられてしまい、広域観光行動に着目することができなくなる恐れがある。この課題に対し本研究では、近接しているセンサ間をグルーピングしたうえで、エリア間のトリップチェーンを抽出することを試みる。グルーピングには、Community Detection を採用する<sup>9)</sup>。この手法は社会ネットワーク分析等の分野で提案されている手法で、ネットワークをコミュニティ (集団内での繋がりが多く、周りとの繋がりが少ないところ) に分割する手法である。グルーピング結果には物理的距離は考慮されておらず、内内での移動が多くなるようにノード群が形成される。本研究ではセンサをノード、全センサ間をリンク、各センサ間の移動量を重みとしてネットワークを定義して計算を行っている。対象とする 39 センサすべてが起動し始めた最初の 1 週間である 2018/12/03~2018/12/9 のデータを対象に計算を行った。計算結果を図 4-1 に示す。同色のノードで同じ赤枠内にあるノード群が同じグループに所属することを示している。この結果は 1 週間の平均的なグルーピングであり、多少の物理的距離を加味した結果である。京都駅に設置してあるセンサは非常に近接しているにも関わらず

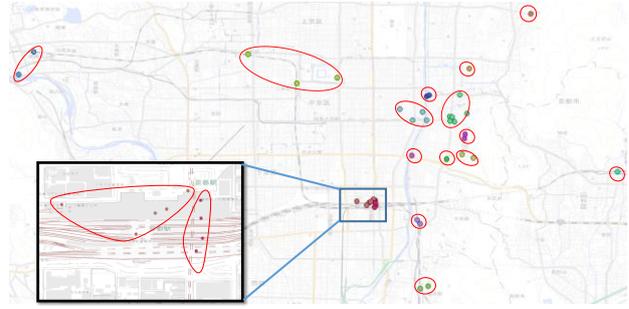


図 4-1 グルーピング結果

グルーピングにより二分されているのは、駅を横断する人が少なく、東西でその後の移動先が異なっているためと考えられる。

## 5. 移動パターン抽出手法の検討

### (1) 系列パターンマイニング

本研究では Wi-Fi パケットセンサデータから移動パターンを抽出する手法として系列パターンマイニングを採用する。

系列パターンマイニングとは、時間順に並べられたイベントの順列の中から頻出するアイテムセットを抽出するデータマイニングの手法である。なお、あくまでも順序を抽出する手法であり、時間そのものは一般的には問われない。頻出するアイテムの組み合わせ (頻出パターン) を抽出する手法としては頻出パターンマイニング (バスケット分析) という手法も存在する<sup>9)</sup>。系列パターンマイニングはこのバスケット分析の拡張であり順序を含めた頻出パターンを抽出する点が特徴である。いずれも主にマーケティングの分野で用いられる概念であり、例えば商品 A と商品 B を「同時に」購入する可能性が高いというパターンを抽出するのがバスケット分析であり、商品 A を購入した「後に」商品 B を購入する可能性が高いというパターンを抽出するのが系列パターンマイニングである。そして、頻出パターンを表現するためのルールを相関ルールといい、抽出されたパターンは相関ルールで評価される。

データベースにおける一体不可分の処理単位のデータのことをトランザクションといい、系列パターンマイニングにおいて取り扱うデータはこのトランザクションが対象である。今、あるアドレス  $r$  のデータ列を  $T_r = \{t_{1r}, t_{2r}, \dots, t_{n_r r}\}$  という集合とする。また、データベース全体をトランザクションの集合として  $D = \{T_1, T_2, \dots, T_m\}$  と定義する。そしてトランザクションには最初に捕捉されたセンサ群を示す前提部  $X$ 、ある訪問列の順列を示す帰結部  $Y$  という部分集合パターンが含まれているとする。このように定義した場合、相関ルールは次のように表現される<sup>10)</sup>。

#### (a) 支持度 (Support)

全体パターンのうち、前提部  $X$ 、かつ帰結部  $Y$  を含むパターンの割合を示す。1 に近いほど頻出するパターンを示している。

$$\text{Support}(X, Y) = \frac{X \text{かつ} Y \text{を満たす訪問パターン数}}{\text{全体訪問パターン数}} \quad (5.1)$$

### (b) 確信度 (Confidence)

前提部  $X$  が観測された際に帰結部  $Y$  が発生する条件付き確率を示す。1 に近いほど  $X$  と  $Y$  が同時に発生、つまり  $X$  を満たすパターンは、 $Y$  に向かう可能性が高いことを示している。

$$\text{Confidence}(X, Y) = \frac{X \text{かつ} Y \text{を満たす訪問パターン数}}{X \text{を満たす訪問パターン数}} \quad (5.2)$$

### (c) リフト値 (Lift)

前提部  $X$  かつ帰結部  $Y$  の確信度を  $Y$  の発生確率で割った値である。前提部の影響により帰結部の発生率がどの程度増加するのかを示しており、1 以上であれば  $X$  の影響により  $Y$  が発生しやすくなったといえる。

$$\text{Lift}(X, Y) = \frac{\text{Confidence}(X, Y)}{Y \text{の発生確率}} \quad (5.3)$$

これらの指標を試算したうえで、三指標に対し主成分分析をおこない、三つの指標から総合的に見て代表的と考えられるパターンを抽出する。

## (2) 使用アルゴリズム

本研究では系列パターンマイニングを実装するために cSPADE というアルゴリズムを用いて分析を実施する。これは、Zaki<sup>11)</sup>が 2000 年に提案したアルゴリズムであり、同氏の提案した SPADE アルゴリズム<sup>12)</sup>を元に制約条件を付加する形で拡張したものである。

cSPADE アルゴリズムは時間間隔やパターンの長さ、パターン内に出現するアイテムの数、パターン全体の時間制約を考慮できるため、より実態に近いパターンを列挙することが可能であるアルゴリズムである。

Wi-Fi パケットデータに適応する場合、ある A-Mac アドレスについて、最初の観測から最後の観測をトリップチェーンとする。トリップチェーンは移動についての一体不可分の処理単位のデータと考えることができるためトランザクションとして取り扱うことができる。また、トリップチェーンの中で経由する観測地点は同様にトランザクション内に出現するアイテムとして取り扱うことができる。また、タイムスタンプは観測地点の出現時刻そのものである。

## 6. まとめ

本研究では Wi-Fi パケットセンサデータを用いた代表的観光パターン抽出について検討した。はじめに、CommunityDetection 法を使うことで内内の移動

量が多いノード群を抽出することができ、移動量に基づいたセンサのグルーピングができることを示した。また、トリップパターンの抽出手法として系列パターンマイニングを採用し観光の出発地と回遊場所の関係について分析を行っている。こちらの結果については発表時に報告する。今後の課題として、1 時点だけでなく、複数時点で試算を行うことで、交通施策や観光施策の有無、時期の違いについて比較検討していきたいと考えている。また、訪日外国人の移動の特性を把握するために、外国籍の方に限った分析等を進める必要があると考えている。

**謝辞：**本研究は、新道路技術会議「観光流動把握を目的とした交通流動推定システムの研究開発」（研究代表者 京都大学 宇野伸宏）の研究成果の一部である。記してここに謝意を表す。

### 参考文献

- 1) 西田純二, 宇野伸宏, 倉内文孝, 中川義也, 望月祐洋, Wi-Fi パケット観測の精度と個人情報保護, 第 57 回土木計画学研究・講演集, 2018
- 2) 森本哲郎, 辻本悠佑, 白浜勝太, 上善恒雄, Wi-Fi パケットセンサを用いた人流解析と可視化, DEIM Forum 2015 F8-3, 2015
- 3) 西田純二, 足立智之, 牧村和彦, 森本哲郎, 上善恒雄, Wi-Fi パケットセンサーによる交通流動解析, 第 49 回土木計画学研究発表会・講演集, 2014
- 4) 浅尾啓明, 西田純二, 安東直紀, 前田繁, Wi-Fi パケットセンサーによる交通常時観測システムの実用化, 第 55 回土木計画学研究発表会・講演集, 2017
- 5) 遠藤幹大・高橋央亘・浅田拓海・有村幹治: Wi-Fi パケットセンシングによる広域観光圏における時空間周遊パターン分析, 土木計画学研究・講演集, 57, 2018
- 6) 伊藤伸・倉内文孝・西田純二: Wi-Fi パケットセンサデータを用いた観光流動把握のための系列パターンマイニング, 土木学会中部支部研究発表会講演概要集, 2019
- 7) 伊藤伸, 倉内文孝, 安東直紀, 西田純二, Wi-Fi パケットセンサデータによる観光行動把握の可能性に関する研究, 第 56 回土木計画学研究発表会・講演集, 2017
- 8) Joerg Reichardt・Stefan Bornholdt: Statistical Mechanics of Community Detection, Physical Review E Vol. 74, 016110, 2006
- 9) 鷲尾隆, 用語解説 (バスケット分析), 日本ファジィ学会誌 13 巻, p54, 日本ファジィ学会, 2001
- 10) 篠原透・沼尾雅之: シーケンシャルパターンマイニング拡張による特徴的なコード進行の抽出手法, DEIM Forum, 2015
- 11) Mohammed J. Zaki, "Sequence mining in categorical domains: incorporating constraints", Proceedings of the ninth international conference on Information and knowledge management, pp422-429, 2000
- 12) Mohammed J. Zaki, "Efficient enumeration of frequent sequences", Proceedings of the seventh international conference on Information and knowledge management, pp68-75, 1998

MINING TYPICAL TOURISTS' TRAVEL PATTERNS  
USING WI-FI PACKET SENSOR

Shuhei MYOKO, Fumitaka KURAUCHI and Shin ITO