

1. はじめに

近年、大量のデータから有益な情報を取り出すデータマイニングが注目されている。本研究では、奈良飛鳥地域における既存の GPS 調査データに対してデータマイニングを行い、有益な情報を取り出すための分析手法を提案する。分析は移動滞在判別、及び滞在スポット抽出を行う。移動滞在判別では観光客の滞在の概要を把握し、観光スポット抽出では、集客効果の高い場所の発見を試みる。また、分析結果から、観光地における周遊行動の分析を行う。

2. 分析手法の説明

2. 1. 移動滞在判別

30 秒毎に取得した各サンプルの位置データに対して移動滞在判別を行うことにより、各点が移動点であるか滞在点であるかを判別する。特異点除去、1 次判別、2 次判別の 3 段階に分けて行う。

1) 特異点除去

GPS が誤って取得した特異点を取り除く。連続する 2 点 $i(x_i, y_i)$, $i+1(x_{i+1}, y_{i+1})$ 間の移動速度 $V_{i,i+1}$ を求める。これを、速度閾値 V (m/s) と比較し、 $V_{i,i+1} > V$ ならば、点 $i+1$ を特異点とみなし、これを除去する。

2) 1 次判別

点 i までの判別は終わっているものとし、点 $i+1$ の判別について考える。

(i) 点 i が滞在点の場合

<STEP1-1>

連続する 2 点 $i(x_i, y_i)$, $i+1(x_{i+1}, y_{i+1})$ 間の距離 $D_{i,i+1}$ を求める。これを、1 次判別用の距離閾値 D_{1st} (m) と比較し、 $D_{i,i+1} > D_{1st}$ ならば <STEP1-2>へ、そうでないなら <STEP1-3>へ進む。

<STEP1-2>

点 $i+1$ を暫定的移動点とし、 i を更新して <STEP2-1>へ進む。

<STEP1-3>

点 $i+1$ を滞在点として確定させ、 i を更新して <STEP1-1>を繰り返す。

(ii) 点 i が暫定的移動点の場合

<STEP2-1>

連続する 2 点 $i(x_i, y_i)$, $i+1(x_{i+1}, y_{i+1})$ 間の距離

$D_{i,i+1}$ を式(2)によって求め、 $D_{i,i+1} > D_{1st}$ ならば <STEP2-2>へ、そうでないなら <STEP2-3>へ進む。

<STEP2-2>

点 i を移動点として確定させ、点 $i+1$ を暫定的移動点とし、 i を更新して <STEP2-1>を繰り返す。

<STEP2-3>

点 i , 点 $i+1$ を共に滞在点として確定させ、<STEP1-1>へ戻る。

全点について判別を行った後、連続する滞在点のまとまりを滞在スポットと定義し、その滞在時間、平均座標を計算する。滞在スポット m の平均座標 (x_m, y_m) は、各滞在点の滞在時間 t_m によって重み付けを行い、以下の式により求める。

$$x_m = \frac{\sum_{k=1}^i x_k t_k}{\sum_{k=1}^i t_k}, \quad y_m = \frac{\sum_{k=1}^i y_k t_k}{\sum_{k=1}^i t_k} \quad \dots \text{式(1)}$$

3) 2 次判別

滞在スポット m における滞在時間 t_m を、時間閾値 T_{stay} (s) と比較し、 $t_m > T_{stay}$ ならば、滞在スポット m を真の滞在スポットとする。そうでなければ、一時的な滞在とみなし、 m を滞在スポットとしない。

2. 2. 観光スポット抽出

移動滞在判別によって取り出された複数のサンプルの滞在スポットを、最短距離法によるクラスター分析を行うことで、集客効果の高い場所を発見する。

<STEP1>

全サンプルの滞在スポットの組み合わせのうち、非類似度 D_{mn} が最小となる滞在スポットの組 $m(x_m, y_m)$, $n(x_n, y_n)$ を抽出する。なお、非類似度は以下の式により計算する。

$$D_{mn} = t_m t_n \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2} \quad \dots \text{式(2)}$$

t_m, t_n : 滞在スポット m , n での総滞在時間

<STEP2>

滞在スポット m , n を束ねる。新たな座標 (x'_m, y'_m) は、各滞在スポットに含まれる滞在点数によって重み付けを行い、以下の式により求める。

$$x'_m = \frac{x_m c_m + x_n c_n}{c_m + c_n}, \quad y'_m = \frac{y_m c_m + y_n c_n}{c_m + c_n}$$

・・・式(3)

c_m, c_n : 滞在スポット m, n に含まれる
滞在点数

<STEP3>

非類似度の最小値 $MinD_{mn}$ を終了条件 D_{spot} と比較し, $MinD_{mn} > D_{spot}$ ならば観光スポット抽出を終了する. そうでなければ, <STEP1>を繰り返す.

3. 分析結果

3. 1. 移動滞在判別

図-1 に任意の1サンプルについての移動滞在判別結果を示す. なお, 距離閾値は 10m, 時間閾値は 5 分とした.

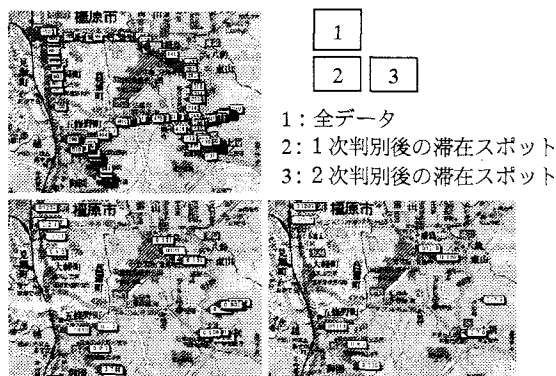


図-1 移動滞在判別結果の一例

図-2 に, 距離閾値を変えた場合に, 1 サンプル当たりの平均滞在回数がどのように変化するかを示す.

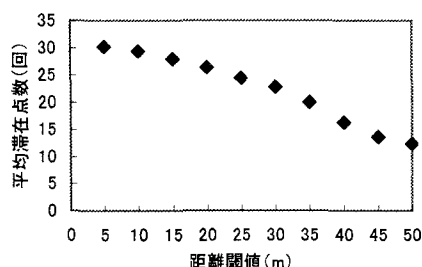


図-2 距離閾値と1サンプル当たりの平均滞在回数の変化

図-2 より, 平均滞在点数は距離閾値に比例して減少することがわかる. このことから, 距離閾値を大きく設定すれば, 滞在スポットは広い面積を持つ面的滞在スポットとして抽出されることが考えられる. 逆に, 距離閾値を小さく設定すれば, 滞在スポット

は面積を持たない点的な滞在スポットとして抽出されることが考えられる. なお, 以下の理由から, 距離閾値は 10m~30mの範囲が適当であると考えられる.

- ・ GPS の誤差が約 10mであるため, 距離閾値として有効な値の最小値は約 10mと考えられる.
- ・ 一般的な人の歩行速度は時速 3~4km/s (30 秒当たり約 30m) であると考えた場合, 距離閾値として有効な値の最大値は約 30mと考えられる.

各滞在スポットにおける滞在時間のうち, 最も頻度が高かったのは1分以上2分未満の滞在であった.

3. 2. 観光スポット抽出

図-3 に観光スポット抽出結果の一例を示す. なお, 終了条件は $D_{spot} = 1000000 (ms^2)$ とした.

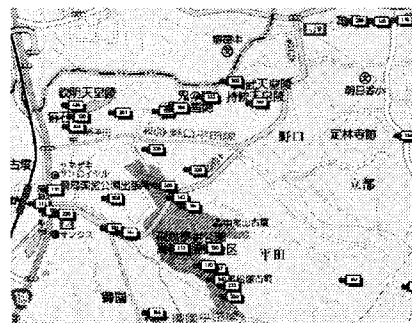


図-3 観光スポット抽出結果の一例

自販機や看板, ベンチなどを設置する際に, 図-3 で示される観光スポットを参考にするなどでその有効性を高めることができるなどといった活用方法が考えられる.

4. おわりに

本研究では, 奈良飛鳥地域における観光客の移動データから, 観光地の周遊行動の分析を行った. その結果, 一日の観光行動では1サンプル当たり約 20~30 回の滞在を繰り返すこと, 1~2 分間の滞在が最も多いことがわかった. また, 観光スポット抽出を行うことで, 従来は観光資源と認識されていなくても, 何らかの理由により集客効果の高い場所を観光スポットとして発見する手法を提案した.

【参考文献】朝倉康夫・羽藤英二・大藤武彦・田名部淳: PHS による位置情報を用いた交通行動調査手法, 土木学会論文集, No.653/IV-48, pp.95-104, 2000.7