

観光地推薦における R 木を用いた目的別推薦の効率化

平山篤[†] 鈴木優[†] 川越恭二[†]

[†]立命館大学 情報理工学部

1 はじめに

近年、旅行が身近になり、観光をする機会が増加している。また、インターネットの普及によって多くの人がインターネットを利用している。そのため、観光地を調べる際、利用者はインターネットを用いて、寺、遊園地などの観光地を検索することが考えられる。このとき、観光地を検索する方法として、目的の観光地名や観光地周辺の地理情報を入力として観光地を検索する方法や、利用者の移動経路を入力として観光地を検索する方法がある。移動経路を用いる方法では、移動経路と観光地の距離を算出し、移動経路に近い観光地を出力する。ただし、利用者の入力した移動経路と、すべての観光地との距離を計算しなければ、移動経路と観光地の近さ、遠さは判断できない。このため、計算量が多くなり、検索結果が表示されるまで時間がかかるという問題がある。

そこで、利用者の目的に応じた観光地の種類から R 木 [1] を作成する観光地推薦手法を提案する。R 木にはデータを空間的に分割することができ、分割された領域を探索していくため、効率的に検索対象を絞り込むことができるという利点がある。そのため、R 木を使用することにより、検索結果が表示されるまでの時間を短縮できると考える。さらに、利用者の目的が漠然としていて観光地の種類を決めかねている場合は、ある一定のルールを適用することにより観光地を特定する。

2 目的別 R 木を用いた観光地推薦手法

本提案手法は利用者の移動経路を入力として用いる観光地推薦において、検索結果表示までの時間短縮を行う。

本提案手法の概要を図 1 に示す。STEP1 において、

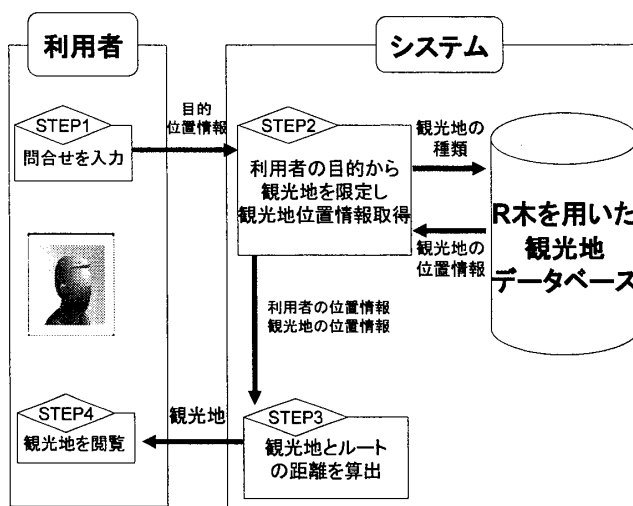


図 1: 観光地推薦システムの概要

利用者は目的と、位置情報を問合せとして入力する。STEP2 では、システムは目的別 R 木を用いて利用者が入力した目的から観光地の特定を行う。目的別 R 木については、2.1 節において詳細を述べる。また、利用者の目的が曖昧であった場合、システムは特定のルールを適用し観光地の特定を行う。特定のルールについては、2.2 節において詳細を述べる。システムは特定した観光地をもとにしてデータベースに問い合わせ、特定の観光地の位置情報を取得する。STEP3 において、システムは取得した利用者の位置情報と観光地の位置情報から 2 点間の距離を算出する。STEP4 において、システムは算出された距離をもとに、利用者の位置に近い観光地を利用者に出力する。なお、出力される観光地の件数は 10 件とする。

2.1 目的別 R 木

利用者が観光をする際、利用者の目的がはっきりしており、観光地の種類が決定している場合がある。例えば、「寺を回りたい」、「遊園地に行きたい」といった観光地の種類を特定できる場合である。その場合、システムは利用者の目的によって観光地の種類を特定して観光地を絞り込むことが可能である。本提案手法では、

Efficiency improvement of recommendation according to purpose to use R-tree in sightseeing spot recommendation

Atsushi HIRAYAMA[†], Yu SUZUKI[†] and Kyoji KAWAGOE[†]
[†]College of Information Science and Engineering, Ritsumeikan University.

hirayama@coms.ics.ritsumei.ac.jp

suzuki@ics.ritsumei.ac.jp

kawagoe@is.ritsumei.ac.jp

利用者の目的である観光地を N 種類とする。また、 R 木を使用することによって、観光地を領域に分割してその領域を探索することが可能となる。そのため、分割した領域から検索対象を絞り込むことが可能である。これらのことにより、目的別に R 木を作成することにより、より検索対象を絞り込むことができ、検索結果表示までの時間が短縮可能であると考えられる。

2.2 ルールの適用

利用者の目的が曖昧である場合、利用者は観光地の種類を特定することが困難であることが考えられる。例えば、「観光」、「暇つぶし」といった目的のように観光地の種類を特定できない場合である。この場合、利用者の目的から観光地の種類を限定することができず、すべての観光地とルートとを計算しなければならないため、計算量は削減できない。

そこで本研究は、ルールを適用し観光地の種類を限定する。ルールとは、利用者の曖昧な目的に対して、観光地の種類を適用させ、観光地の種類を限定することである。具体的には、表1によって利用者の目的と観光地の種類を対応付けを示す。このように、ルール $R_h (1 \leq h \leq M)$ に対して $I (1 \leq I \leq N)$ 種類の観光地を適用する。ルールを適用することによって、観光地の種類を限定することが可能となり、検索結果表示までの時間が短縮可能であると考えられる。

表 1: ルールの対応表

ルール	利用者の目的	観光地の種類
R_1	観光	「寺」、「遊園地」、「動物園」
R_2	暇つぶし	「飲食店」、「ショップ」

2.3 距離算出方法

本稿で用いる距離算出方法について述べる。利用者の位置情報を (X_1, Y_1) とする。また観光地の位置情報を (X_2, Y_2) とする。利用者の位置情報と観光地の位置との距離 D の算出方法を式 (1) とする。

$$D = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \quad (1)$$

(1) 式を用いて利用者の位置情報と観光地の位置情報との距離を算出し、距離が近い観光地を出力する。

3 評価実験

3.1 観光地データの生成

緯度、経度をランダムに指定した 60,000,000 件の座標を生成し、これらの座標を観光地の座標とみなす。

観光地データを 2 章で述べた手法にもとづいたデータベース、従来手法 (2.2 節において述べたルールを考慮しない R 木による位置座標管理手法) にもとづいたデータベース、および木を構築しないデータベースの三つにそれぞれ格納する。

3.2 評価方法

本提案手法と、従来手法、および線形探索手法の三つの手法を比較する。比較基準としては、ある利用者の位置座標および目的が入力されてから、周辺の観光地位置情報を取得するまでの時間とする。実験では使用する観光地の種類として、「寺」、「ホテル」、「飲食店」、「遊園地」、「動物園」、「ショップ」を用いた。また、特定のルールとして「暇つぶし」という目的に対して、「飲食店」、「ショップ」を用いた。同様に「観光」という目的に対して、「寺」、「遊園地」、「動物園」を用いた。取得する位置情報の上限は 10 件とする。

なお、評価実験は上記の評価方法で実施中である。

4 おわりに

本稿ではルールを考慮した R 木を用いた効率的な観光地位置座標管理手法の提案を行った。今後の検討課題として、評価実験を行い本提案手法の有用性を検証する。また、利用者からの入力情報として座標だけでなく、利用者の移動経路 [2] にも対応できるようにすることを考えている。さらに本提案手法のルールは、システム側であらかじめ作成されており利用者の嗜好を考慮していない。利用者の目的に対する観光地の種類は、利用者個人によって変化すると考えられる。そのため、利用者の嗜好を考慮してルールを作成することにより、利用者の嗜好に一致する観光地の推薦をすることを考えている。

参考文献

- [1] A. Guttman: "R-Trees: A Dynamic Index Structure for Spatial Searching", ACM SIGMOD Record, 14, 2, pp. 47-57 (1984).
- [2] 塚本祐一, 石川佳治, 北川博之: "移動オブジェクトへの周辺情報提供のための空間データベース検索システムの実装と評価", 電子情報通信学会論文誌 D, 87, (2004).