

## 京都観光支援のための移動軌跡データ類似検索手法

石塚 淳<sup>†</sup> 鈴木 優<sup>††</sup> 川越 恭二<sup>††</sup>

本研究では、京都における観光を支援するために、観光者の移動軌跡データ類似検索を行うための手法を提案する。現在、位置情報を基にした移動軌跡データ類似検索手法が数多く提案されている。ところが、利用者の観光支援のためには必ずしも位置情報が有効であるとはいえず、その位置に関する歴史的背景や地名など、位置情報に関するテキスト情報が有効である場合があると考える。そこで、位置情報だけではなくその位置情報に関するテキスト情報を利用して移動軌跡データの類似検索を行うための手法を提案する。提案手法を利用することによって、移動軌跡データに含まれる位置に関する利用者の移動意図を的確に特徴量に反映することができるため、より利用者の意図に適合した移動軌跡データの類似検索を行うことが可能となる。

### A Similarity Search of Trajectory Data for Supporting Kyoto Sightseeing

JUN ISHIZUKA,<sup>†</sup> YU SUZUKI<sup>††</sup> and KYOJI KAWAGOE<sup>††</sup>

In this paper, we propose a method of similarity search for trajectory data sets for supporting the users' sightseeing in Kyoto. Recently, many similarity search engines for trajectory data have been proposed by many researchers. However, the algorithms of these similarity search engines only deal with the position of the trajectory data. We think that the other data, such as historical background and the name of the place, are also essential to calculate similarities between two sets of trajectory data. In this paper, we propose a novel method for calculating similarities of trajectory data using textual metadata. Using our proposed method, the system can calculate accurate similarity values for measuring trajectory data.

#### 1. はじめに

現在、位置情報サービスの発展により、利用者は移動中に場所に関する情報を取得することを携帯端末によって容易に行うことができる。また、GIS (Global Information Service) の発展により、位置情報に関連した情報を得ることが容易になってゆくと考えられる。位置情報を利用したサービスとして、利用者にとって興味があると考えられる情報を推薦するシステムが多くの研究者によって考案されている。例えば、多くの利用者が同一の道順をたどることをあらかじめ利用者の行動履歴から収集しておき、利用者へその道順を提示する手法である。これらのシステムを利用することによって、利用者は多くの利用者の道順を参考にしながら、利用者の今後の道順を選択、決定することがで

きる。

従来の情報推薦システムでは、移動軌跡データを構成する位置情報として緯度・経度が用いられてきた。これは、位置が近い観光スポットへ行った観光客同士が互いに類似した興味を持っているという仮定により設定された手法である。また、位置情報を緯度、経度の多次元時系列データとして扱うことによって、類似度を定義することが容易であるという側面もある。ところが、利用者の位置を緯度、経度だけによって判断した場合には、利用者の興味から特徴を抽出する際に十分な特徴を得ることができない場合があると考えられる。例えば、京都のように美術館、博物館と寺院が混在している状況を考える。このような場合、多くの利用者が寺院へ移動している場合であっても、その寺院に隣接している美術館の情報は推薦されてしまうが、その寺院とは離れているが寺院に関係した博物館は推薦されないといった問題点がある。つまり、位置関係による距離と利用者の興味による類似度は必ずしも一致しないという点が従来の手法の問題であるといえる。

本研究では、位置情報からその位置に関するメタデー

<sup>†</sup> 立命館大学大学院 理工学研究所

Graduate School of Science and Engineering, Ritsumeikan University

<sup>††</sup> 立命館大学 情報理工学部

College of Information Science and Engineering, Ritsumeikan University

タを特徴として抽出し、移動軌跡データ類似検索に利用する手法の提案を行う。本手法では緯度経度情報を問合せとして Web 検索エンジンによって検索結果を得る。次に、その検索結果からその緯度経度情報に関連するキーワード抽出を行い、その緯度経度情報に対してメタデータを付加する。これらの手順を全ての緯度経度情報に対して行い、移動軌跡データを構成する位置情報に対してメタデータを付与する。最後に、メタデータであるキーワード同士をベクトル検索モデルによって類似度を算出し、二つの移動軌跡データの類似度を得ることができる。

提案手法では、移動軌跡の緯度経度情報ではなく、キーワードを移動軌跡データ類似検索に利用した点が既存の手法と大きく異なる点である。我々は、移動軌跡データから得られる特徴の一つではなく多角的に抽出することによって、利用者の考える類似度に近い類似度をシステムによって自動的に算出することができると考えている。本稿では、移動軌跡データから得られる特徴として、位置情報以外の情報として、その位置情報に関するテキスト情報を利用することを考えた。そのため、例えば京都において観光を行い、その移動軌跡を入力した上で京都以外の観光地における行動支援が必要であると考えたとき、その時代背景や利用目的等が類似した観光スポットを利用者に提示することができる。つまり、より利用者にとって直感的に類似していると考えられる観光スポットを利用者へ提示することができると思われる。

また、提案手法の利点として、入力となる移動軌跡データが必ずしも十分な大きさを持っている必要が無い点が挙げられる。既存の移動軌跡データに比べて十分な長さを持った移動軌跡データをシステムに入力しなければ、十分な検索結果を得ることができなかった。提案手法では、移動軌跡が一つであってもその移動軌跡に関するテキスト情報から十分な特徴を得ることができるため、類似検索を行った場合に十分な精度を得ることができると思われる。

提案手法を利用することによって、ある利用者が観光地において移動した場合に、その移動軌跡をシステムに入力することによって、次に利用者が目指す目的地をシステムから提案することが可能となると考えられる。

以下、2章では、提案手法の詳細について述べる。次に、3章において提案手法の類似手法として、時系列データの類似検索手法、場所に関する施設名を取得するための手法を紹介する。最後に、4章では本論文のまとめと今後の課題について述べる。

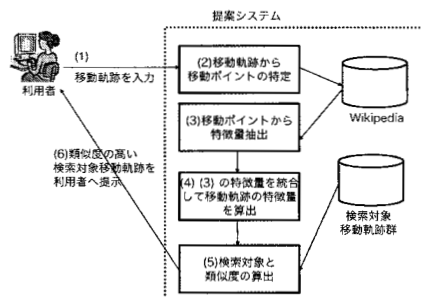


図 1 提案手法の概要

Fig. 1 Overview of our proposed method

## 2. 移動軌跡データの類似検索手法

本章では、二つの移動軌跡データの類似度を測定する手法について述べる。利用者は入力として一つの移動軌跡データ  $T_q$  を入力する。検索システムは、 $T_q$  に類似している移動軌跡データを検索対象データ群  $T_i (i = 1, 2, \dots, N)$  から選択し、類似している順に利用者へ提示する。

ここで、移動軌跡間の類似の定義について述べる。一般的に移動軌跡の類似度とは、物理的にどの程度移動軌跡間の距離が近いことによって定義されていることが多い。提案手法では、移動軌跡となる観光スポットの性質や特徴が類似しているかどうかによって類似度を定義する。また、ここで観光スポットのような、移動軌跡中で特徴量を抽出することができる場所を移動ポイントと呼ぶ。

提案手法は、図 1 で示すように、次のような流れで行われる。

- (1) **利用者が一つの移動軌跡を入力する。**
  - 利用者は、観光の出発地点から途中となる地点までの移動軌跡データを、提案システムへ入力する。ここで、利用者が入力する移動軌跡データは、緯度・経度を一つの要素とした二次元時系列データである。
- (2) **移動軌跡から移動ポイントを特定する。**
  - 利用者の移動軌跡は、多くの部分が利用者の移動であると考えられる。そこで、これら移動部分を削除し、移動ポイントを抽出する。本処理により、利用者が訪れた地域名、建物名である移動ポイントを一つの要素とした時系列データへ変換する。
- (3) **各移動ポイントから特徴ベクトルを抽出する。**
  - 各々の移動ポイントから、その移動ポイントに関する解説を Web 検索エンジンによ

り抽出し、テキスト情報を得る。ここで、検索対象として Wikipedia<sup>\*</sup> を利用することを考える。解説記事に記述されている単語から、その単語の出現頻度を TF-IDF 法により算出し、出現頻度の高い単語をキーワードとして抽出する。最後に、それらキーワードの出現頻度をベクトルとした表現したものである特徴ベクトルを、各移動ポイントごとに求める。

(4) **移動ポイントの特徴ベクトルから移動軌跡データの特徴ベクトルを算出する。**

- 移動軌跡データには複数の移動ポイントが存在するため、一つの移動軌跡データには複数の特徴ベクトルが算出されることとなる。そこでこれら複数の特徴ベクトルを統合し、一つの特徴ベクトルを算出する。特徴ベクトルを統合する際には、各要素の平均をとる手法を選択し、本稿で詳細について述べる。また、複数の特徴ベクトルを一つにまとめたものを、移動軌跡データ行列と呼ぶ。

(5) **検索対象データ群の移動軌跡特徴ベクトルと問合せ移動軌跡特徴ベクトルを比較する。**

- ベクトル検索モデルでは一般的に利用されているコサイン尺度を利用して、検索対象データの特徴ベクトルと問合せ特徴ベクトルとの類似度を求める。

(6) **類似度の高い移動軌跡データから順に利用者へ提示する。**

- 上のステップで求めた類似度のうち、最も類似していると考えられる移動軌跡データは、利用者の必要となる移動軌跡データであると考えられるため、類似度の高い移動軌跡データから順に利用者へ提示を行う。

ここで、移動軌跡から移動ポイントを特定するためには、参考文献 10) などの既存手法を活用することができる。そこで、本稿では既存手法によって移動ポイントを特定できると仮定し、3. 以降の手法について以下に述べる。

本章では、まず 2.1 節において本研究で用いる移動軌跡データについて述べる。ここではまた、位置情報が連続した移動軌跡データから文書へと変換する手法を用いて利用者の移動軌跡データを Wikipedia の文書の履歴へと変換する事についても述べる。次に、2.2 節では、観光スポット文書を用いてベクトル空間を生

成する手法を述べる。さらに、2.3 節では、2.2 節で生成されたベクトル空間に基づきベクトルを生成する手法について述べる。最後に利用者の移動軌跡と、検索対象データの類似度を求めるための手法について述べる。

### 2.1 前提条件

入力された移動軌跡データ  $T_q$  には、一つ以上の移動ポイント  $T_q(j)(j = 1, 2, \dots, L_q)$  が存在し、それぞれ  $T_q = \{t_q(1), t_q(2), \dots, t_q(L_q)\}$  とする。同様に、検索対象移動軌跡データ  $T_i$  にそれぞれ移動ポイント  $T_i = \{t_i(1), t_i(2), \dots, t_i(L_i)\}$  が存在する。ここで、移動軌跡データに含まれる移動ポイント群には時系列である全順序関係が存在するため、 $T_i$  および  $T_q$  は全順序集合であるといえる。

また、各々の移動ポイントには、メタデータが付与されている。ここで移動ポイントのメタデータとは、時代背景や所在地、建立年など移動ポイントに関する様々な情報が自然文で記述されたものである。移動ポイント  $t_i(j)$  に付与されているメタデータを  $M(t_i(j))$  とする。

以上の定義より、移動軌跡データ間の類似度を算出する。移動軌跡データ  $T_q$  と  $T_i$  との類似度を  $S(T_q, T_i)$  とする。

次に、移動軌跡データ  $T_i$  に含まれる移動ポイント  $t_i(j)$  から特徴量を算出する手法について述べる。

#### 2.1.1 索引語の抽出

本研究では、移動軌跡データより得られた Wikipedia の文書の内容を文書中に含まれる単語の集合で近似する。このように文書の内容を特徴づける単語を索引語と呼ぶ。

本研究では、文書から単語を抽出するために形態素解析を行う。また、観光スポットの内容が記述された Wikipedia の文書を構成する重要な単語として、名詞のみを索引語として抽出する事とする。

#### 2.2 ベクトル空間の生成

ベクトル空間モデル<sup>3),11),12)</sup> は、情報検索における代表的な検索モデルである。本研究では、観光地の情報を Wikipedia の文書として置き換えることにより、その文書を用いてベクトル空間を生成する。つまり、文書を多次元のベクトルに置き換える事により  $x$  と  $y$  の連続した移動軌跡データをベクトル空間上の多次元のベクトルに変換する。これにより、ベクトル間の類似度を計算することで観光の内容を考慮した移動軌跡データ間の類似検索を実現することが可能となる。ベクトル空間を生成する際の処理の流れを表したものが図 2 である。以下にベクトルを生成するまでの流れに

<sup>\*</sup> <http://ja.wikipedia.org/>

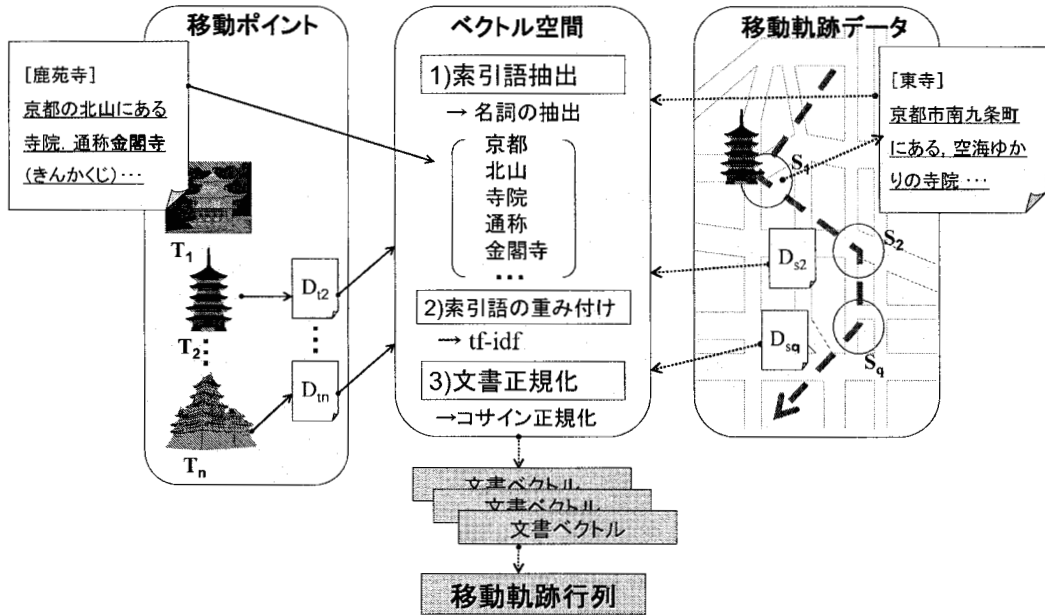


図2 移動軌跡行列生成処理  
Fig. 2 Process of generating feature array of moving point

ついて述べる.

### 2.2.1 移動ポイントからの特徴量抽出

自然文であるメタデータ  $M(i, (j))$  から特徴量を抽出する. ここでは, メタデータから抽出される特徴量として単語の出現頻度情報を用いる. 本研究では, 移動ポイントに関する Wikipedia の文書の内容をメタデータとして扱う. つまり, 移動ポイントの特徴量として文書に含まれる単語の集合とする. このように, 移動ポイント文書の内容を特徴づける単語を索引語と呼ぶ.

本研究では, 文書から単語を抽出するために形態素解析を行う. また, 観光スポットの内容が記述された Wikipedia の文書を構成するにあたり重要な単語として, 名詞のみを索引語として抽出する事とする.

### 2.2.2 索引語の重み付け

2.1.1 節で求めた索引語の中には, 文書の内容と密接に関連したものもあれば, 文書の内容とは関係の薄いものも存在する. そこで, 索引語の出現頻度 (Term-Frequency) と, 文書頻度の逆数 (Inverse Document Frequency) の TF-IDF を用いて重み付けを行う. 重み付け手法を 図3 に示す.

本研究では,  $n$  個の観光スポット数, つまり  $n$  個の Wikipedia の文書を  $D_j (j = 1, 2, \dots, n)$  とし, これらの文書集合から全部で  $m$  個の索引語  $w_i (i = 1, 2, \dots, m)$

が抽出されたとする. また, 索引語  $w_i$  の文書  $D_j$  における出現頻度を  $f_{ij}$  とし, 索引語  $w_i$  を含む文書数である文書頻度を  $n_i$  とする. このとき, 索引語  $w_i$  の文書  $D_j$  における重み  $d_{ij}$  は, 以下の式で表す.

$$d_{ij} = \log(1 + f_{ij}) \cdot \log \frac{n}{n_i} \quad (1)$$

また, 観光スポットの内容が記述された Wikipedia の文書集合中には, 単語数の多い文書と少ない文書が存在する. しかし一文中の単語数が多い場合には, 同じ索引語が多く含まれている傾向があるために, 必然的に単語数の多い文書に含まれる索引語の方が大きな重みを持つ傾向がある. そこで, この単語数の違いによる影響を除くために, 単語数の違いに対応するため単語の重みの正規化が必要となる. 本研究では, 一般的によく用いられる文書正規化手法としてコサイン正規化<sup>13)</sup>を用いる. コサイン正規化は文書に含まれる全ての索引語の重みの二乗和を1にするというものである. 正規化後の重み  $d'_{ij}$  は以下の式で求められる.

$$d'_{ij} = \frac{d_{ij}}{\sqrt{\sum_{i=1}^m (\log(1 + f_{ij}) \cdot \log \frac{n}{n_i})^2}} \quad (2)$$

### 2.3 移動軌跡行列の生成

2.2.2 節により, Wikipedia の観光スポット文書の履歴へと変換された移動軌跡データを, ベクトル空間上のベクトルに変換する. 任意の移動軌跡データ

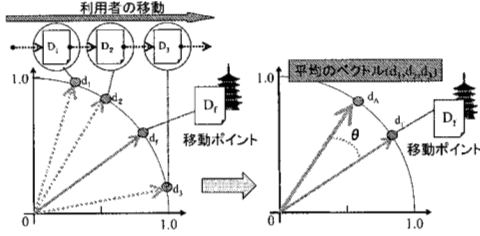


図3 移動ポイントからの類似度算出手法

Fig. 3 A method of calculating similarities between moving point and query point.

$T_k (k = 1, 2, \dots, p)$  の滞在した観光スポットの文書を  $D_{kl} (l = 1, 2, \dots, q)$  とする。つまり、 $T_k$  には  $q$  個の文書として  $D_{k1}, D_{k2}, \dots, D_{kq}$  が存在する。

2.2.2 節で述べた方法により索引語の重みを要素とするベクトルで文書を表現すると、文書  $D_l$  は以下の文書ベクトルで表す。

$$d'_l = [d'_{1l}, d'_{2l}, \dots, d'_{ml}]^T \quad (3)$$

ここで、 $d'_{ij}$  は索引語  $w_i$  の文書  $D_j$  における重みである。また、移動軌跡データ  $T_k$  は、以下のような  $m \times q$  の行列  $D$  によって表すことができる。

$$D = [d'_1, d'_2, \dots, d'_q]$$

$$= \begin{pmatrix} d'_{11} & d'_{12} & \dots & d'_{1q} \\ d'_{21} & d'_{22} & \dots & d'_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ d'_{m1} & d'_{m2} & \dots & d'_{mq} \end{pmatrix} \quad (4)$$

このようにすることで、文書集合全体を行列へと変換することができる。また、この行列の各行は索引語に関する情報を表したベクトルであり、Wikipedia の観光スポット文書の履歴に変換された移動軌跡データを  $T_k$  を  $w$  次元ベクトル空間  $R^w$  上の移動軌跡データと考えることができる。

## 2.4 類似度の定義

本章では、類似度を定義するための手法を述べる。まず提案手法では、複数の移動ポイント上の履歴である移動履歴行列から、移動軌跡データの特徴量である特徴ベクトルを抽出する。本研究では、移動軌跡行列から特徴ベクトルへ変換する際に、平均を利用する。次に、問合せ特徴ベクトルと検索対象特徴ベクトルとの類似度を算出するために、コサイン尺度を用いるための手順を述べる。

### 2.4.1 平均による移動軌跡行列のベクトル化

コサイン尺度はベクトル間の類似度を定義するものであるために、本研究で用いる移動軌跡のような幾つかの文書を持った行列同士の類似度を定義する事ができない。そこで、本研究では文書ベクトルの平均を用

いることによって、移動軌跡行列を一つのベクトルとして類似度を求める。

2.3 節で述べたような  $m \times q$  の行列  $D$  の移動軌跡データ  $T_k$  の索引語に対する各行の値を加算し、観光スポット数  $q$  の値で除算した平均値を求め、ベクトル  $d''$  とする式を以下に表す。

$$d'' = \frac{\sum_{l=1}^q d'_l}{q} = \begin{pmatrix} \frac{\sum_{l=1}^q d'_{1l}}{q} \\ \frac{\sum_{l=1}^q d'_{2l}}{q} \\ \vdots \\ \frac{\sum_{l=1}^q d'_{ml}}{q} \end{pmatrix} \quad (5)$$

### 2.4.2 コサイン尺度による類似度

二つの任意の移動軌跡データ  $T_\alpha, T_\beta$  のコサイン尺度を用いた類似度は以下の式で表す。

$$\cos(d''_\alpha, d''_\beta) = \frac{d''_\alpha \cdot d''_\beta}{\|d''_\alpha\| \|d''_\beta\|}$$

$$= \frac{\sum_{i=1}^m d''_{i\alpha} d''_{i\beta}}{\sqrt{\sum_{i=1}^m d''_{i\alpha}{}^2} \sqrt{\sum_{i=1}^m d''_{i\beta}{}^2}} \quad (6)$$

これらの手法を、問合せ移動軌跡データと検索対象移動軌跡データ全てに対して計算を行い、類似度を算出することによって、利用者に対して移動軌跡データを推薦することができる。

次に、3章において関連研究に関する説明を述べる。

## 3. 関連研究

本章では、従来の類似検索に関する技術、移動軌跡データより場所に関する情報を抽出する技術の紹介をする。

### 3.1 時系列データの類似検索

移動軌跡データの類似検索とは検索対象となるデータの中から、問合せを行う移動軌跡データと何らかの類似性のあるデータをデータベースの中から検索することである。我々が扱う移動軌跡データは、時間と共に緯度、経度の位置情報が変化するという点において時系列データの一種であるといえる。また、移動軌跡データは時間、緯度、経度を持つ3次元構造であり、多次元データの一種であると言える。そこで、本節では時系列データの類似検索手法に関する研究と、多次元データの類似検索手法に関する研究について挙げる。また、従来提案されている移動軌跡データ類似検索手法に関する研究と本提案手法との差異について述べる。

時系列データの類似検索を行うための手法が、多くの研究者によって考案されている。最も単純なものとしては二つの時系列データ間のユークリッド距離を

類似度として定義し、ユークリッド距離が近い時系列データの類似度を高くする方法である。しかし、一般に時系列データは検索対象となるデータが大量となり、検索に時間がかかる。そこで、Agrawal らによる離散フーリエ変換 (DFT: Discrete Fourier Transform) を用いた手法<sup>1)</sup> や、Faloutsos らによる離散ウェーブレット変換 (DWT: Discrete Wavelet Transform) などの特徴量抽出関数を用いた手法<sup>8)</sup> により、次元圧縮を行って索引を用いることによって検索速度を改善するための提案が行われている。一方、検索速度を短縮させるという目的は同一であるが、ユークリッド距離以外の方法で類似度を定義する方法も提案されている。さらに Agrawal らは、二つのデータがある時に、一方のデータからもう一方のデータへ変換操作を行い、その変換に要する手順の回数によって類似性を定義する手法<sup>2)</sup> の提案が行われている。

しかし、移動軌跡データは三次元の時系列データであり、上記に挙げた関連研究の手法をそのまま用いることはできない。三次元のデータは空間上の点であり、こうした多次元空間上の点の連続したものに対する問合せを高速を行うために空間インデックスを用いる手法<sup>4)</sup> が数多く提案されている。代表的なものには R-Tree<sup>9)</sup> があり、空間上に存在する点を MBR (Minimum Bounding Region) で集約し、MBR を階層状に索引付けを行う方法が提案されている。

移動軌跡データに対する類似検索手法は、上記の時系列データの類似検索手法や、多次元データの類似検索を移動軌跡データに応用することによって類似度を定義する手法が提案されている。また、移動軌跡データから得られる時間情報を用いることによって、移動体の速度や加速度を考慮した類似度定義を行う手法<sup>15)</sup> や、多数の移動オブジェクトの状況をマルコフ遷移確率の形で要約し、エントロピーを用いる事で曖昧度を用いて移動統計量を算出する手法<sup>17)</sup> が提案されている。

以上に挙げた研究では、利用者の行動を位置情報や時間情報を用いることによって類似度の定義を行っているが、移動の内容を考慮していない点が問題であると考えられる。移動にはその移動した利用者の意図があり、移動した内容を考慮することが必要だと考える。そこで本研究では、位置的に近いものの類似度が高いのではなく、内容を考慮した類似度の再定義を行うことで新たな類似検索の手法を提案する。

### 3.2 場所に関する施設名を取得する研究

GPS 等の位置計測機器等から得られる利用者の移動軌跡データから、利用者の行動や状況を示すコンテキ

ストへと変換する研究が進められている<sup>10),14),16)</sup>。Liu らの手法<sup>10)</sup> では、GPS 等の位置計測機器から得た緯度、経度の位置情報からセマンティックロケーションへと変換する手法の提案が行われている。セマンティックロケーションにはその滞在箇所の施設名、住所、施設種別等の情報を持ったものである。本研究では、セマンティックロケーションにより得た、滞在箇所の名称を入力とすることで得られる滞在箇所の説明文を類似検索に利用する。

## 4. おわりに

本稿では、京都における観光を支援するために、利用者の行動履歴と類似した行動履歴を検索するための手法を述べた。提案手法では、行動履歴における位置情報だけではなく、その位置における時代情報などのテキスト情報から類似度算出を求めた。そのため、利用者の意図に適合した移動軌跡データを検索することが可能となった。

最後に、今後の課題について述べる。本研究では、移動軌跡データの特徴量を抽出する際に、移動軌跡データを構成する移動ポイントから特徴量を抽出し、その平均を求めた。ところが本稿で述べた手法では、移動ポイントの順序を考慮して移動軌跡データの特徴量を抽出していない。そのため、順序が異なるが同一の移動ポイントが含まれる移動軌跡データからは、同一の特徴量が抽出される。ところが、我々は移動ポイントの順序によって利用者の移動意図が異なると考えており、そのため異なる特徴量が抽出されるべきであると考えている。そこで、移動ポイントの順序を利用した移動軌跡データの特徴量を抽出する手法が課題である。

また提案手法では、類似度を算出するための特徴量抽出のために、非常に多くの時間がかかる。これは、位置情報だけではなくメタデータである単語の特徴量による計算を行っているためである。計算時間に関する問題を解決するための手法として、top-k<sup>5)-7)</sup> を利用した問題解決の手法が考えられる。top-k とは、複数の順位付きリストを統合する際に、その順位付けに利用した数値であるスコアの高いものから順に高速に出力するためのアルゴリズムである。例えば、複数の検索エンジンへ同一のキーワードを入力し、検索結果集合を得た際に、それぞれの検索エンジンが最も高いスコアを与えたものから順に出力する際に利用される。提案手法では、これらの手法を利用することによって、利用者の位置情報と利用者の移動ポイントに関するメタデータ双方を考慮した検索結果をそれぞれ求め、top-k

アルゴリズムにより統合を行う手法を検討している。これらの手法を実現することが可能となれば、高速に利用者へ必要な位置情報を提供することができる。

提案手法は移動軌跡データだけではなく、実世界オブジェクトの動きを利用者の興味によって類似検索を行う手法に応用することが考えられる。

- ロボットや人間の行動や動線に関する情報を類似検索へ応用
- 行動履歴を用いたマルチメディア処理への応用
- 画像や映像認識後の行動分析の応用

謝辞 本研究の一部は、21世紀COEプログラム「京都アート・エンタテインメント創成研究」の支援によるものである。ここに記して謝意を表す。

### 参 考 文 献

- 1) Agrawal, R., Faloutsos, C. and Wami, A.: Efficient Similarity Search in Sequence Databases, *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithm (FODO '93)*, pp.69-84 (1993).
- 2) Agrawal, R., Lin, K.I., Sawhney, H.S. and Shim, K.: Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases, *Proceedings of Very Large Databases (VLDB '95)*, pp.490 - 501 (1995).
- 3) Baeza-Yates, R. and B.Ribeido-Neto: *Modern Information Retrieval*, ACM Press (1999).
- 4) Bertino, E. and et al.: *Indexing Techniques for Advanced Database Systems*, Kluwer Academic Publishers (1997).
- 5) Cai, D., Yu, S., Wen, J.R. and Ma, W.Y.: Extracting Content Structure for Web pages based on visual representation, *Proceedings of the 5th Asia Pacific Web Conference* (2003).
- 6) Dwork, C., Kumar, R., Naor, M. and Sivakumar, D.: Rank Aggregation for the web, *Proceedings of the 10-th International World Wide Web Conference (WWW10)*, pp.613 - 622 (2001).
- 7) Fagin, R., Kolatis, P.G. and Popa, L.: Efficient similarity search and classification via rank aggregation, *Proceedings of the 2003 ACM International Conference on Management of Data (SIGMOD)*, pp.pp. 301 - 313 (2003).
- 8) Faloutsos, C., Ranganathan, M. and Manolopoulos, Y.: Fast Subsequence Matching in Time-Series Databases, *Proceedings of the Conference on Management of Data (SIGMOD '94)*, pp. 419 - 429 (1994).
- 9) Guttman, A.: R-trees: a dynamic index structure for spatial searching, *SIGMOD '84: Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, New York, NY, USA, ACM Press, pp.47-57 (1984).
- 10) J.Liu, O.Wolfson and H.Yin: Extracting Semantic Location from Outdoor Positioning Systems, *MDM2006 workshop MCISME*, pp.34 - 41 (2006).
- 11) Salton, G. and Lesk, M.E.: *Introduction to Modern Information Retrieval*, McGrawHill Book (2000).
- 12) Salton, G., Wong, A. and Yang, C.S.: A Vector Space Model for Automatic Indexing, *Communications of the ACM*, Vol.18, No.11, pp.pp. 613 - 620 (1975).
- 13) 北 研二, 津田和彦, 獅子堀正幹: 情報検索アルゴリズム, 共立出版 (2002).
- 14) 長尾光悦, 川村秀憲, 山本雅人, 大内 東: GPS ログマイニングに基づく観光動態情報の獲得, 観光と情報, Vol.1, No.1, pp.38 - 46 (2005).
- 15) 河内聡恵, 増永良文: ムービングオブジェクトの速度変化パターンを識別できる類似検索機能の導入, 日本データベース学会 Letters, Vol.2, No.1, pp.15 - 18 (2003).
- 16) 岡 峰正, 國頭吾郎, 高橋竜男, 田中 聡: GPS 携帯電話による位置情報ログを用いた訪問施設抽出方法, マルチメディア, 分散, 協調とモバイル (2006) シンポジウム, pp.273 - 276 (2006).
- 17) 塚本祐一, 石川佳治, 北川博之: 索引付けされた移動軌跡データからの効率的な移動統計量抽出法, 日本データベース学会 Letters, Vol.2, No.1, pp.27 - 30 (2003).