DEWS2007 E1-7

内容を考慮した移動軌跡データの類似検索手法

石塚　淳† 鈴木　優†† 川越　恭二††

† 立命館大学大学院 理工学研究科　〒 525-8577 滋賀県草津市野路東 1-1-1
†† 立命館大学 情報理工学部　〒 525-8577 滋賀県草津市野路東 1-1-1
E-mail: †ishiduka@coms.ics.ritsumei.ac.jp, ††{yusuzuki,kawagoe}@is.ritsumei.ac.jp

あらまし　本稿では，利用者の移動の特徴や意図を考慮した新たな移動軌跡データの類似度の提案を行う．現在，利用されている移動軌跡データの類似検索手法では，類似度として位置情報を用いるものが多い．ところが，問合せと検索対象の位置が近いことが必ずしも利用者にとって有効な検索結果であるとはいえない．そこで本研究では，位置情報だけを類似度として利用するのではなく，利用者の移動に含まれる特徴や意図を移動軌跡データと移動ポイント間の類似度へ反映するために，利用者の滞在箇所に関する説明文書を用いる手法を提案する．利用者の移動の特徴や意図を抽出し，利用者にとってより直感的に類似した移動軌跡データ類似検索を行うための手法を提案する．また，利用者の移動軌跡に存在する複数の滞在箇所から得られる特徴量の変化を反映することによって，さらに検索精度を上げることを試みる．評価実験により，利用者の意図に合致した目的地を提案することが可能であることを確かめた．
キーワード　移動軌跡データ，類似検索，類似度，メタデータ

A Content Based Similarity Search for Trajectory Data

Jun ISHIZUKA†, Yu SUZUKI††, and Kyoji KAWAGOE††

† Graduate School of Science and Engineering, Ritsumeikan University.
Nojihigashi 1-1-1, Kusastsu, Shiga, 525-8577 Japan
†† College of Information Science and Engineering, Ritsumeikan University.
Nojihigashi 1-1-1, Kusastsu, Shiga, 525-8577 Japan
E-mail: †ishiduka@coms.ics.ritsumei.ac.jp, ††{yusuzuki,kawagoe}@is.ritsumei.ac.jp

Abstract　In this paper, we propose a method of a content based similarity search of trajectory data sets for assisting decision of the users' best destination. Recently, many similarity search engines of 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. The algorithms use only physical locations for similarity search are not effective, because most users have interest of moving. We believe that the data of users' interest of moving are also essential to calculate similarities between trajectory data and users' best destination. In this paper, we propose a novel method for calculating similarities of trajectory data using textual metadata. We use descriptive document of the spot the user had stayed. In our proposed method, we use average function to integrate the users' trajectory data and use slope of the similarities. Consequently, we confirmed that the system can calculate accurate similarities for trajectory data. We also confirmed the precision of our proposed method for trajectory data.
Key words　- Trajectory data, Similarity search, Similarities, Metadata

1. はじめに

近年，GPS(Global Positioning System) 等の位置計測機器の発達により，利用者の移動軌跡データの取得が容易になった．また，屋外での利用者の位置情報を利用することによって位置情報サービスが普及し，次の目的地を決定する際の情報推薦が携帯端末によって行われている．

しかし，従来の位置情報サービスにおける情報推薦は現在地を入力とした近傍検索によるものが多い．例えば，京都で金閣寺にいる利用者を例とする．従来の位置情報サービスでは，利用者は現在地である金閣寺と距離の近い観光施設，売店，駐車場等の周辺情報を取得することができる．ところが，周辺情報は距離が近いというだけで利用者の興味を考慮した情報推薦であるとはいえない．さらに，現在地の入力に加えて，利用者の

興味を利用者が入力することにより，利用者の興味を考慮した情報推薦を行うことも可能であると考えられる．しかし，利用者がその観光地ついて詳しい知識が無い場合は入力が困難であったり，利用者が利用者自身の興味を正確に把握しているとは限らない場合がある．つまり，位置情報サービスは今後利用者の興味を考慮し，利用者が次の目的地を決定する際のきっかけを与える情報推薦を行う必要があると考える．

そこで，従来の位置情報サービスに加え，あらかじめ多くの利用者の行動履歴として移動経路を収集しておき，利用者は収集された移動経路を参考にしながら，利用者の今後の移動経路を選択，決定することを考えた．例えば，観光客にとって初めて訪れた場所では，過去に利用者と同じ観光施設を訪れた観光客の行動履歴は重要である．利用者と同じ場所を訪れた観光客は利用者の興味と類似していることとなり，利用者にとって目的地を決定する際に参考になる．利用者の行動履歴には移動軌跡データを用いることが可能となるため，過去のその場所における多くの利用者の移動軌跡データと利用者の移動軌跡データの類似検索を行うことによって利用者にとって有効な情報を取得することが可能となり，利用者が次に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．

移動軌跡データの類似検索手法は多くの研究者によって研究開発されており，一般的な移動軌跡データの類似検索を行う際の類似度として緯度，経度の位置情報を用いるものが多い．しかし，位置情報サービスを受ける利用者にとって位置が近いことが必ずしも利用者にとって有効であるとはいえない．なぜなら，位置情報を類似度として用いることによって，位置が近い観光施設を訪れ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ことになる．例えば，京都のように美術館，博物館，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が推薦される．ところが，その寺院と位置が離れているがその寺院に関係した博物館は推薦されないといった問題がある．つまり，位置が近いことと，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位置情報を類似度として用いた問題であるといえる．

そこで，本研究では利用者の移動の特徴や移動意図を考慮した新たな類似度の提案を行うことによって，利用者にとってより直感的な移動軌跡データの類似検索手法を実現する．我々は，移動軌跡データから得られる特徴を緯度，経度の位置情報だけを用いるのではなく多角的に抽出することにより，利用者にとって直感的に類似していると考える類似度を自動的に算出することができると考えている．提案手法では，移動軌跡データから得られる特徴として，利用者の移動の滞在箇所に関する説明文書をメタデータとして用いることを考えた．利用者の移動を位置情報のみで考えるのではなく，説明文書を用いることによって，利用者の移動の特徴を十分に得ることができるため，類似検索を行った場合に十分な検索精度を得ることができると考えられる．さらに，利用者の移動軌跡に存在する複数の滞在箇所によって得られる類似度の変化量を反映することにより，さらに検索精度を高めることを試みる．

2. 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行う．

2.1 類似検索に関する研究

移動軌跡データの類似検索とは，検索対象となる複数のデータの中から，問合せを行う移動軌跡データと類似性のあるデータを検索することである．ここで，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ため，時系列データの一つである．また，移動軌跡データは時間，緯度，経度を持つ三次元構造であるため，多次元データの一つである．このように，移動軌跡データの類似検索には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手法を用いるため，これらの研究について述べる．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には様々な検索モデルが提案されている．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，距離が近いもの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の類似検索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 らにより提案された離散的フーリエ変換 (Discrete Fourier Transform;DFT) 用いた手法 [1] や Faloustsos らにより提案された離散的ウェーブレット変換 (Discrete Wavelet Transform;DWT) 等の特徴抽出関数を用いた手法 [2] により，次元圧縮を行って索引を作成することによって，検索速度を向上させる提案が行われている．しかし，移動軌跡データは三次元のデータであり，二次元のデータを対象とした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．

一方，移動軌跡データは三次元構造であり，こうした多次元空間上の検索にかかる時間を短縮するために，空間インデックスを用いる手法 [3] が数多く提案されている．代表的なものには R-Tree [4] があり，空間上にちらばる点を MBR(Minimum Bounding Region) でまとめ，それらの図形を階層状にして管理する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，上記に挙げた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を提案している．また，河内らは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，速度や速度の変化量を表す加速度を考慮した類似度定義を行う手法 [5] や，石川らは多数の移動オブジェクトの状況をマルコフ遷移確率を用いて要約し，エントロピーを用いることによって曖昧度を用いて移動統計量を算出する手法 [6] を提案している．

以上に挙げた研究では，位置情報や時間情報を用いて移動軌跡データ間の類似度の定義を行っているため，移動の内容を考慮していない．そこで本研究では，移動軌跡データ間の距離を類似度とするのではなく，利用者の移動の内容として利用者の移動の特徴や移動の意図を考慮した類似度の定義を行うことによって新たな移動軌跡データの類似検索の手法を提案する．

2.2 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

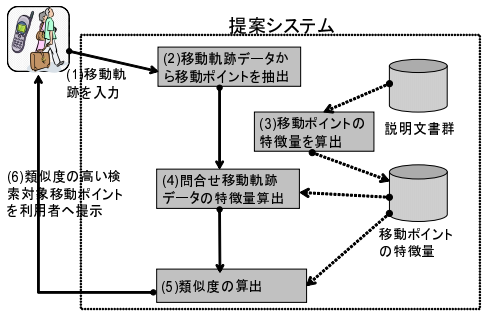

図 1　提案手法の概要
Fig. 1　Overview of our proposed method.

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．Liu らにより提案された手法 [7] では，GPS 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．利用者の移動軌跡データは移動部分と，施設等に滞在している部分に別れる．セマンティックロケーションは，移動軌跡データより滞在箇所を判別し，その滞在箇所の施設名，住所，施設種別等の情報を抽出したものである．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 Web 検索エンジンに入力とすることによって抽出した滞在箇所の説明文書を類似検索に利用する．

3. 移動軌跡データの類似検索手法

本章では，問合せ移動軌跡データと，検索対象となる利用者の次の目的地との類似度を算出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．提案システムでは，T_q に類似している目的地を検索対象データ群 $P(l)(l = 1, 2, \cdots, N)$ から選択し，類似している順に利用者へ提示する．ここで，利用者の次の目的地となる移動箇所および，移動軌跡データより得られる滞在箇所のことを**移動ポイント**と呼ぶこととする．本提案システムを用いることによって，システムに対して移動軌跡データを送信する処理のみで，利用者の目的地となる移動ポイントを提案することが可能である．

提案手法の概要を図 1 に示し，以下に提案手法の流れを説明する．

（1） **利用者が一つの移動軌跡データを入力する**

利用者は移動の出発となる地点から，現在地となる地点までの移動軌跡データを入力する．ここで，利用者が入力する移動軌跡データは，緯度，経度の位置情報を要素とした二次元時系列データである．

（2） **移動軌跡データから移動ポイントを抽出する**

参考文献 [7] 等の既存手法を活用することによって，位置情報を要素とした時系列データであった移動軌跡データを，移動ポイントを要素とした時系列データへと変換する．

（3） **移動ポイントの特徴量を算出する**

説明文書群を利用することによって，移動ポイントの特徴量を算出する．ここで，説明文書群とは利用者の移動軌跡より抽出された移動ポイントおよび，検索対象移動ポイントに関する説明の記述された文書のことである．情報検索の分野において利用されている方法を用いて，文書の特徴ベクトルを算出することによって移動ポイントの特徴量とする．

（4） **問合せ移動軌跡データの特徴量算出**

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，(3) の方法によって算出された移動ポイントの特徴ベクトルを用いて行列を生成する．ここで，複数の特徴ベクトルのことを**移動軌跡データ行列**と呼ぶこととし，この行列を問合せ移動軌跡データの特徴量とする．

（5） **類似度の算出**

入力された移動軌跡データ行列と，検索対象データの特徴ベクトルとの類似度を求める．ここでは，二つの方法により類似度を算出する．類似度算出の詳細は 3.3 節で述べる．

（6） **利用者への移動ポイントの提案**

上述の提案手法の流れで求めた類似度のうち，類似度の高い移動ポイントは，利用者にとって必要な移動ポイントであると考えられるため，類似度の高い移動ポイントから順に利用者へ提示を行う．

本章では，まず 3.1 節において本研究で用いる移動ポイントと，移動軌跡データについて述べる．次に，3.2 節では，説明文書群を用いてベクトル空間を生成し，特徴ベクトルを生成する手法について述べる．最後に，3.3 節では，利用者の入力した移動軌跡データと，検索対象移動ポイントとの類似度を算出するための手法について述べる．

3.1 前 提 条 件

検索対象となる n 個の移動ポイントを $p(l)(l = 1, 2, \cdots, n)$ とする．また，入力された移動軌跡データ T_q には，一つ以上の k 個の移動ポイント $p_q(k)(k = 1, 2, \cdots, L_q)$ が存在し，それぞれ $T_q = \{p_q(1), p_q(2), \cdots, p_q(L_q)\}$ とする．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_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 $p(l)$ 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，その場所に関する説明が自然言語で記述された説明文書である．例えば，京都を例にすると，「金閣寺」の移動ポイントのメタデータとしては，金閣寺を入力として Web 検索エンジンによって抽出された，金閣寺の説明の記述された説明文書のことである．また，移動ポイント $p(l)$ に付与されている説明文書を $D(p(l))$ とする．同様に，移動軌跡データ T_q に存在する移動ポイント $p_q(l)$ のメタデータを $D(p_q(l))$ とする．

3.2 特徴ベクトルの生成

ベクトル空間モデル [8] は，情報検索における代表的な検索モデルである．本提案手法では，移動ポイントのメタデータとしての説明文書群を用いることによってベクトル空間を生成する．また，緯度，経度の位置情報を要素とした移動軌跡データを移動軌跡データ行列へと変換する．これにより，問合せとなる移動軌跡データと検索対象移動ポイントとの特徴ベクトル間の類似度を算出することが可能となる．

検索対象となる移動ポイント群の文書を用いることによってベクトル空間を生成し，移動ポイントの特徴ベクトルと移動軌跡データ行列を生成するまでの処理の流れを表したものが図 2 である．図 2 を用いて以下の処理を説明する．

3.2.1 索引語の抽出

移動ポイントのメタデータである説明文書の内容を文書中に

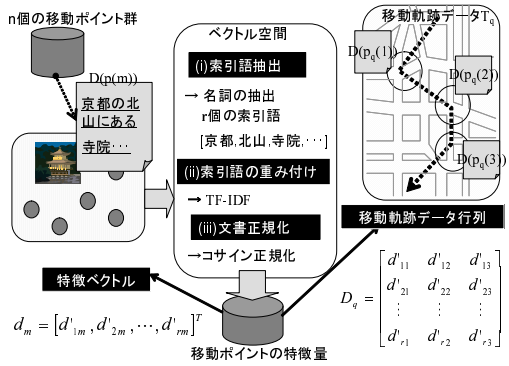

図 2　移動ポイントの特徴ベクトルと移動軌跡データ行列の生成までの流れ

Fig. 2　Process of generating feature vector of moving point and feature matrix of trajectory data.

含まれる単語の集合で近似する．ここで，文書の内容を特徴づける単語として索引語の抽出を行う．また，索引語には名詞を抽出することとする．図 2 のステップ (i) では，移動ポイント群の説明文書 $D(p(m))$ の一部分より名詞を抽出し，索引語として [京都，北山，寺院，$\cdots$] を抽出する様子を表す．さらに，移動ポイント群の全ての説明文書より，r 個の索引語を抽出する．

3.2.2　索引語の重み付け

3.2.1 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，図 2 のステップ (ii) では， 索引語の出現頻度 (Term Frequency) と，文書頻度の逆数 (Inverse Document Frequency) の積である TF-IDF を用いて重み付けを行う．ここで，3.2.1 節の方法により r 個の索引語 $w_i(i = 1, 2, \cdots, r)$ が抽出されたとする．また，索引語 w_i の文書 $D(p(m))$ における出現頻度を f_{im} 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 $D(p(m))$ における重み d_{im} は，以下の式で表す．

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

また，移動ポイントの説明文書群には，一般に文字数の多い文書と，文字数の少ない文書の双方が存在する．文字数が多くなるにつれ，同じ索引語が多く含まれている傾向があるために，必然的に文字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，図 2 のステップ (iii) ではこの文字数の違いによる影響を除くため，文書正規化を行う．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 [8] を用いる．コサイン正規化は文書中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'_{im} は以下の式で表す．

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

移動ポイント $p(m)$ に付与されているメタデータ $D(p(m))$ の特徴ベクトルは以下の式で表す．

$$\boldsymbol{d'}_m = [d'_{1m}, d'_{2m}, \cdots, d'_{rm}]^T \tag{3}$$

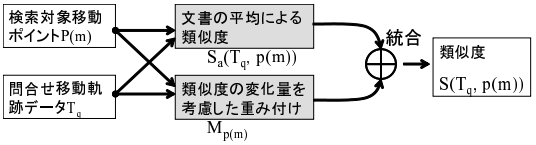

図 3　本手法での類似度算出方法

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

3.2.3　移動軌跡データ行列の生成

移動軌跡データ T_q には複数の移動ポイントが存在する．図 2 では，三つの移動ポイントの存在する $T_q = \{p_q(1), p_q(2), p_q(3)\}$ は，索引語の語数 r によって $r \times 3$ の移動軌跡データ行列 D_q を生成している．k 個の特徴ベクトルが移動軌跡データに存在する場合の，移動軌跡データ行列 D_q は以下の式によって表す．

$$D_q = [\boldsymbol{d'}_1, \boldsymbol{d'}_2, \cdots, \boldsymbol{d'}_k] = \begin{pmatrix} d'_{11}, & d'_{12}, & \cdots & d'_{1k} \\ d'_{21}, & d'_{22}, & \cdots & d'_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ d'_{r1}, & d'_{r2}, & \cdots & d'_{rk} \end{pmatrix} \tag{4}$$

3.3　類似度算出

本節では，図 3 にあるように，一つの問合せ移動軌跡データ T_q と，検索対象移動ポイント $p(m)$ 間の類似度 $S(T_q, p(m))$ の定義の方法を述べる．本手法では，文書の平均による類似度と類似度の変化量を考慮した重みの二つの値を統合することによって類似度を算出する．文書の平均による類似度では，利用者の移動全体の特徴を抽出することを目的とする．文書の平均を用いることによって，問合せ移動軌跡データ行列に存在する複数の特徴ベクトルの平均特徴ベクトルを生成し，検索対象移動ポイントの特徴ベクトルとの類似度を算出する．さらに，類似度の変化量を考慮した重み付けでは，検索対象移動ポイントと利用者の移動に含まれる複数の移動ポイント間の類似度の変化量を用いる．一般的に，利用者の移動には移動意図があると考えられ，利用者の移動の意図に合致した検索対象移動ポイントに対する類似度の変化量は少ないと考えられる．つまり，類似度の変化量を用いることによって，検索対象移動ポイントの重みを算出することが可能となる．最後に，二つの値の和を用いることによって類似度の統合を行い，本手法の類似度とする．また，これら二つの手法で用いる特徴ベクトル間の類似度には，ベクトル検索モデルで利用されるコサイン相関値を利用する．

3.3.1　文書の平均による類似度

本手法では，複数の特徴ベクトルの存在する移動軌跡データ行列を統合し，一つの特徴ベクトルに変換する際に平均を用いる．図 4 は，三つの特徴ベクトルの存在する移動軌跡データ $T_q = \{D(p_q(1), D(p_q(2), D(p_q(3)\}$ と，検索対象となる任意の移動ポイント $p(m)$ との類似度 $S_a(T_q, p(l))$ を算出する際の図である．T_q に存在する特徴ベクトルは $\{\boldsymbol{d_1}, \boldsymbol{d_2}, \boldsymbol{d_3}\}$ であり，$p(m)$ に存在する特徴ベクトルは $\boldsymbol{d_m}$ である．ここで，$\boldsymbol{d_m}$ と $\{\boldsymbol{d_1}, \boldsymbol{d_2}, \boldsymbol{d_3}\}$ とのコサイン相関値を一つずつ求めるのではなく，移動軌跡データに存在する複数の特徴ベクトルの平均となる平均特徴ベクトル $\boldsymbol{d_A}$ を算出する．つまり，移動軌跡データより得られる複数の文書の平均を用いることによって，利用者の移動全体の特徴を抽出する．次に，平均特徴ベクトル $\boldsymbol{d_A}$ と

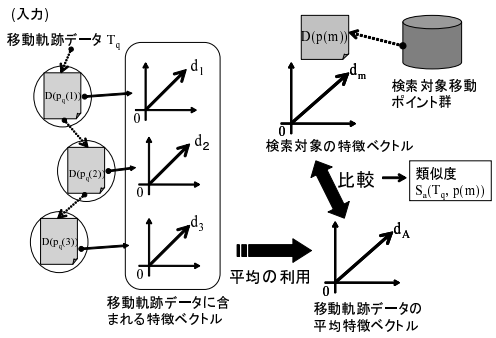

図 4　平均を用いた類似度

Fig. 4　A method of average calculating similarities.

$\boldsymbol{d_f}$ を比較し，コサイン相関値 $cos(\boldsymbol{d_A}, \boldsymbol{d_f})$ を算出することによって類似度とする．この際，$cos(\boldsymbol{d_A}, \boldsymbol{d_f})$ の値が小さいほど，$\boldsymbol{d_A}\boldsymbol{d_f}$ 間の類似度が高いと考えることができる．

平均には以下の三つの手法を採用する．一つの移動軌跡データに存在する k 個の移動ポイントの特徴ベクトルを $\boldsymbol{d_k}(k = 1, 2, \cdots, L)$ とし，統合された相加平均の特徴ベクトルを $\boldsymbol{d_{AA}}$，相乗平均の特徴ベクトルを $\boldsymbol{d_{GA}}$，調和平均の特徴ベクトルを $\boldsymbol{d_{HA}}$ として以下に表す．

- 相加平均 (arithmetic mean value;AA)

$$\boldsymbol{d_{AA}} = \frac{1}{k}\sum_{i=1}^{k} d_i \tag{5}$$

- 相乗平均 (geometric mean value;GA)

$$\boldsymbol{d_{GA}} = \sqrt[k]{\prod_{i=1}^{k} d_i} \tag{6}$$

- 調和平均 (harmonic mean value;HA)

$$\boldsymbol{d_{HA}} = \frac{k}{\sum_{i=1}^{k} \frac{1}{d_i}} \tag{7}$$

ここで，問合せ移動軌跡データ T_q と検索対象移動ポイント $p(m)$ の類似度を 以下に表す．

$$S_a(T_q, p(m)) = \frac{1}{cos(\boldsymbol{d_A}, \boldsymbol{d_f})} \tag{8}$$

3.3.2　類似度の変化量を考慮した重み付け

検索対象の一つの移動ポイントと利用者の移動軌跡データに存在する複数の移動ポイントごとに類似度を算出することによって，対象とした検索対象の移動ポイントにおける類似度の変化量を抽出する．実際の利用者の移動には滞在した移動ポイントごとにその利用者の移動の目的や意図が含まれると考えられる．つまり，利用者の移動軌跡データに存在する移動ポイントの特徴には関連性があることが多い．例えば，観光客が京都の庭園に興味があり，龍安寺，銀閣寺等の観光施設を訪れて観光している状況を考える．この場合，利用者にとって有効な次の目的地の提案はデパートやショッピングモール等の利用者の移動意図に関連性の少ない施設ではなく，庭園に関連した観光施設である．つまり，利用者が過去に訪れた複数の観光施設と利用者にとって有効な情報となる庭園に関連した観光施設には類似性があると考えられる．一方，利用者が過去に訪れた複数の観光施設と利用者にとって有効な情報とはならない施設には類似性が少ないと考えられる．本節ではこの違いを利用することによって，検索対象移動ポイントに重み付けを行う．

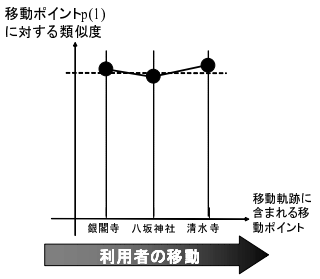

図 5　正解集合に対する類似度の変化例

Fig. 5　A modification of similarities in case of correct moving spot.

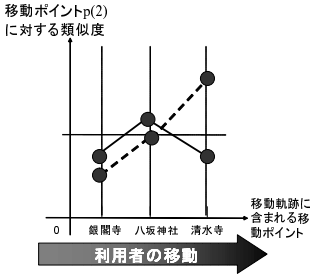

図 6　不正解集合に対する類似度の変化例

Fig. 6　A modification of similarities in case of incorrect moving spot.

ここで，利用者の意図と合致し，利用者にとって有効な提案となる移動ポイント群のことを正解集合と呼ぶこととする．一方，利用者の意図とは合致していなく，有効な提案とはならない移動ポイント群のことを不正解集合と呼ぶこととする．つまり，正解集合となる移動ポイントが，利用者が次に目指す目的地として有効な情報である． 図 5 は正解集合に対する類似度の変化例を表した図であり，図 6 は不正解集合に対する類似度の変化例を表した図である．これらの図では，銀閣寺，八坂神社，清水寺と順に移動した利用者の移動軌跡データを表している．ここで，図 5 は金閣寺 (p(1)) を対象とした類似度の変化とし，図 6 はデパート (p(2)) を対象とした類似度の変化とする．観光を行っている利用者にとって有効な情報となるのは観光スポットである金閣寺のような移動ポイントであり，この場合の利用者の移動の意図と特徴の類似した検索対象データを対象とするため，類似度の変化は少ないものが多いと考えられる．一方，利用者の移動の意図とは全く異なるデパートのような移動ポイントを対象とした場合，利用者の移動の意図とは関係が無く，類似していない移動ポイントを対象とするため，正解集合を対象とした類似度の変化よりも変化が大きいものが多くなると考えられる．つまり，正解集合である利用者の移動の意図に合致した検索対象移動ポイントに対する類似度の変化は少ない．そこで本節では，ある検索対象の移動ポイントと移動軌跡データに含まれる移動ポイントごとに算出される類似度の変化量を用いることによって，上述の重みを算出する．類似度の変化量が少ない程，検索対象移動ポイントに重みを与えることとする．図 7 は，一つの問合せ移動軌跡データ T_q と，検索対象移動ポイント $p(m)$ 間の類似度の変化量を用いることによって重み $M_{p(m)}$ の値を算出する方法を表した図である．T_q には，

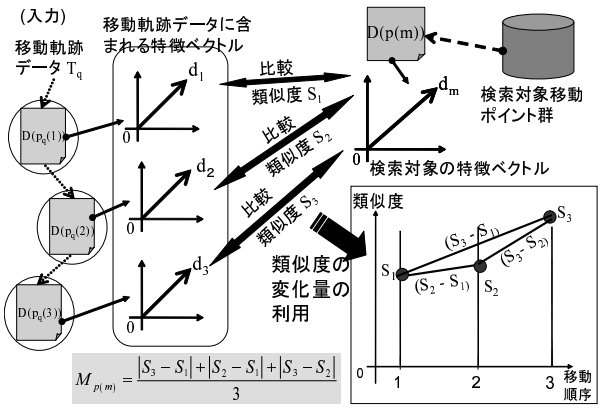

図 7　類似度の変化量を考慮した重み付け
Fig. 7　A modification of similarities based weighting.

$\{\boldsymbol{d_1}, \boldsymbol{d_2}, \boldsymbol{d_3}\}$ の三つの特徴ベクトルが存在し，$p(m)$ の特徴ベクトル $\boldsymbol{d_m}$ とそれぞれ比較することによって，コサイン相関値による類似度 $\{S_1, S_2, S_3\}$ を算出する．さらに，全ての組合せにおける類似度の変化量の絶対値を加算し，その合計を全ての組合せの個数で除したものを重み $M_{p(m)}$ とする．

k 個の移動ポイントが存在する移動軌跡データより全ての組合せとして $(k-1)!$ 個の類似度の変化量が存在する．それらを $a_j (j = 1, 2, \cdots, (k-1)!)$ とし，本節で算出される移動ポイント $p(m)$ の重みの値は以下の式で表す．

$$M_{p(m)} = \frac{\sum_{i=1}^{(k-1)!} |a_i|}{(k-1)!} \tag{9}$$

3.3.3　類似度の統合

3.3.1 節で求めた類似度と，3.3.2 節で算出された重みは異なる値であり，類似度を算出する際には二つの値を統合する必要がある．本稿では，二つの値を加算することによって値を統合する．加算する際，変数 $\alpha (0 \leqq \alpha \leqq 1)$ との積を加算することとする．さらに，α の値の最適値は 4. 章の評価実験によって導出する．また，3.3.2 節で算出された $M_{p(m)}$ の値は最大の重み $\max(M_{p(m)})$ によって正規化し，対数を取ることにより統合する．移動軌跡データ T_q と移動ポイント $p(m)$ の類似度を $S(T_q, p(m))$ とし，以下に式を表す．

$$S(T_q, p(m)) = S_a(T_q, p(m)) + \alpha \times \left(-log \left(\frac{M_{p(m)}}{\max(M_{p(m)})} \right) \right) \tag{10}$$

4. 評価実験

本論文における提案手法が有効であることを確かめるために，評価実験を行った．本手法では，3.3 節で述べたように，利用者の移動の特徴や意図を考慮するために二つの方法により問合せと検索対象との類似度を算出した．提案手法における類似度は，既存の移動軌跡データの類似検索手法との類似度とは異なるものであるため，既存の手法との比較をすることができない．そこで，以下の二つの実験を行うことによって有効性を証明する．
(実験 1)　移動軌跡データ行列から特徴ベクトルへ統合する際に採用された三つの異なる平均の検索精度の比較．
(実験 2)　検索精度の最も良い平均を用いた類似度と，類似度の変化量を考慮した重み付けを行った類似度との比較．また，その際の最適な変数 α の値の検討．
これらの評価を行うことによって，本手法で提案する新たな類似度が利用者にとって直感的に類似しているものであるかを評価することができると考えられる．

4.1　実験方法

本実験では，京都において観光を行っている利用者の移動を想定し，問合せとして利用者の移動軌跡データを入力することにより，その利用者の移動の特徴や意図を考慮した次の目的地となる京都の観光施設を提案することとした．移動ポイントは京都の主要観光施設 108 件とし，メタデータとなる説明文書は，移動ポイントにおける施設名を入力とした Wikipedia(注1) のテキスト文書を用いることとした．利用者の移動軌跡データはあらかじめ観光する際の意図を持たすようにし，京都市観光文化情報システム(注2)の「おすすめコース」にある推薦観光経路を参考にして意図とした．また，移動軌跡データに存在する移動ポイントは 4 箇所とし，研究室の学生 5 名に，あらかじめ設定された意図に合致すると考えられる移動ポイントを利用者の移動軌跡とし，このデータを問合せとした．今回のような問合せの検索結果は，情報が適合するかどうかという評価が主観的なものとなってしまい，被験者によって異なるものである．そこで正解集合は，検索対象データとなる移動ポイント群の中からそれぞれの被験者が正解と判断する移動ポイントをその被験者の正解集合とすることとした．以下に本実験の手順を示す．
（1）　利用者の問合せとなる移動軌跡を考える．本実験では以下の二つの問合せとなる意図を設定する．
(a)　庭園鑑賞コース
(b)　仏像鑑賞コース
利用者はそれらの意図に合致していると考えられる移動ポイントを用いて移動軌跡とする．
（2）　問合せに対して，あらかじめ人手で正解集合を求めておく．
（3）　本提案手法を用いてランキング表示を行う．
（4）　あらかじめ求めておいた正解集合と比較することによって以下の評価を行う．
(実験 1)　11 点再現率適合率曲線による比較
(実験 2)　11 点平均適合率による比較

4.2　実　験　1

4.2.1　実験の目的

移動軌跡データに含まれる複数の特徴ベクトルを統合する際，利用者の移動の全体の特徴を抽出するために三つの異なる平均を用いることによって特徴ベクトルを統合する．三つの異なる平均で実際に実験を行うことによって検索精度の高い平均を求め，最も利用者の特徴を抽出することのできる平均を求める．

4.2.2　実験結果と考察

11 点再現率適合率曲線の実験結果を図 8 に示す．図中にある AA は相加平均，GA は相乗平均，HA は調和平均における検索結果であり，被験者 5 名の 11 点再現率適合率の平均をもとめたグラフである．

実験結果を見ると，(a)，(b) 二つの問合せが共に，相加平均を用いた手法が再現率の変化に関わらず高い値であることが分

(注1)：http://ja.wikipedia.org/
(注2)：http://raku.city.kyoto.jp/sight.phtml

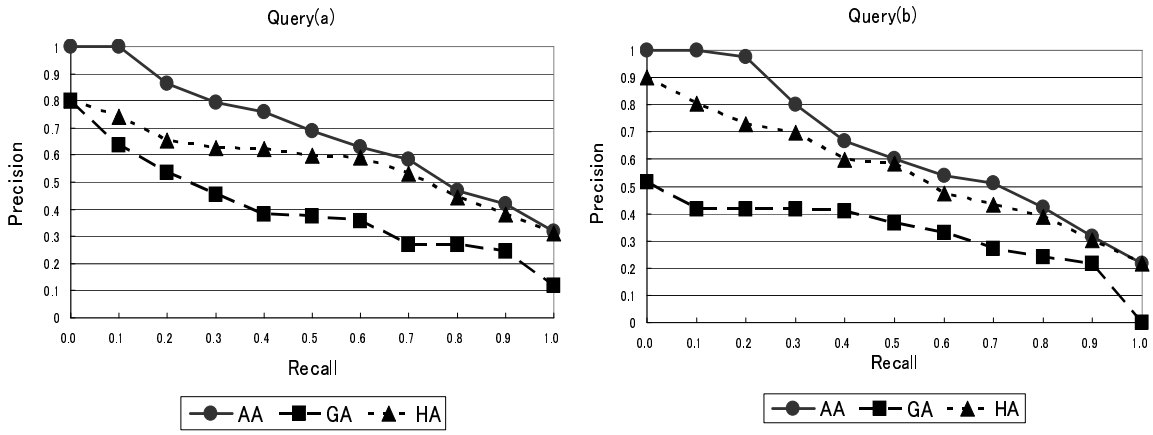

図 8　実験 1 の 11 点再現率-適合率曲線
Fig. 8　11-point recall precision graph of experiment 1.

かる．つまり，相加平均が利用者の移動全体の特徴を考慮していることが分かる．また，二つの問合せとも再現率が 0〜0.1 の部分では，適合率の値が 1 となっている．本提案手法では，利用者に移動ポイントを提示する際，類似度の高い移動ポイントから順に利用者に提示する．その際，再現率の低い部分での適合率が高いということは，順位が高い部分で利用者が正解だと考える移動ポイントを提示することが可能であり，本手法で利用者にとって有効な提案ができたと考えられる．しかし，利用者にとっての正解集合が全て出力されているわけではないために，実験 2 では，再現率の高い部分だけではなく平均適合率を向上させることを目的とする．

一方，相乗平均の値が再現率の変化に関わらず低い値となっているのは，3.2.2 節で重みの二乗和を 1 にする正規化行うことによって，ベクトルの値に差が小さくなる．そのベクトルの値の相乗平均を求める際に積を求めても重みに差が出ないため，検索精度が低くなったと考えられる．また，調和平均は代表的な使用用途として速さの平均を算出するものがあるが，使用方法が本提案手法には合っていない．そこで，今回は評価値の高い相加平均が利用者の移動全体の特徴を考慮した平均手法だと考え，相加平均を用いることによって実験 2 を進めることとする．

4.3 実　験　2

4.3.1　実験の目的

相加平均を用いることにより算出した類似度の検索精度と，被験者ごとに類似度の変化量を考慮することにより重み付けした類似度の検索精度の比較を行うことによって，どのように検索精度が変わるかを考察する．また，重み付けの際，変数 α の値の最適値を求める．

4.3.2　実験結果と考察

本実験の実験結果を図 9 に表す．横軸を α の値とし，縦軸を相加平均のみを用いた検索精度と重み付け後の検索精度の差を表す．実験結果を見ると，傾きを考慮して重み付けを行うことにより 2.9%〜0.0% 検索精度に差があることが分かる．問合せ (a) の被験者 2 のように，大きく検索精度が上がった検索結果の順位を見ると，相加平均だけを用いることによって高い順位だった正解集合は，傾きを考慮して重み付けを行った後も同様に高い順位となっていた．これは，相加平均によって求めた移動軌跡全体の特徴と正解集合の文書の特徴が似ているためだからだと考えられる．また，傾きを考慮した重み付けを行うことによって順位の上がった正解集合は，移動軌跡に存在する複数の移動ポイントによって得られる類似度の傾きの値が小さかったものであるといえる．つまり，それらの順位を上げた正解集合は，文書の特徴が利用者の移動軌跡全体の特徴と類似していなくても，類似度の変化量が少ないものである．それらの文書の特徴を見てみると，問合せ (a) の意図を表す庭園の記述もあるが，その庭園鑑賞コースという意図以外の関係の無い情報が多く記述された文書であった．つまり，移動軌跡全体の特徴に類似していないが，庭園の記述があることによって傾きの値が小さいものとなり，順位を上げて検索精度を上げたと考えられる．

本実験において，α の値が小さいということは，利用者が意図を理解して観光施設である移動ポイントを選択することによって移動軌跡とし，主観によって正解集合かどうかを判断する際に文書の特徴量と類似して正解だと判断していることだと考えられる．そのため，検索精度を向上させることができなかった被験者も存在する．一方，α の値が大きいということは，主観によって判断した正解集合の文書の特徴が意図とは類似していないことや，利用者が正解だと判断した移動ポイント以外にも文書の特徴が類似した移動ポイントが存在することを表す．つまり，傾きを考慮した提案を行うことによって，利用者が気づかないが意図と合致したきっかけとなる目的地を提案することが可能であることが分かる．しかし，利用者にきっかけとしての提案を行う場合，提案された移動ポイントがなぜ高い順位で提案されたのかを利用者に提示する必要があるのではないかと考える．また，本実験では α の最適値は利用者によって異なり，一意に決まることは無かった．3.3.3 節で述べた (10) 式の検討や，使用場所や，使用用途に合わせ多くの被験者を対象に実験を行うことによって最適な α の値を決定することが今後の課題である．

5. 終わりに

本論文では，移動軌跡データの類似検索において，利用者の

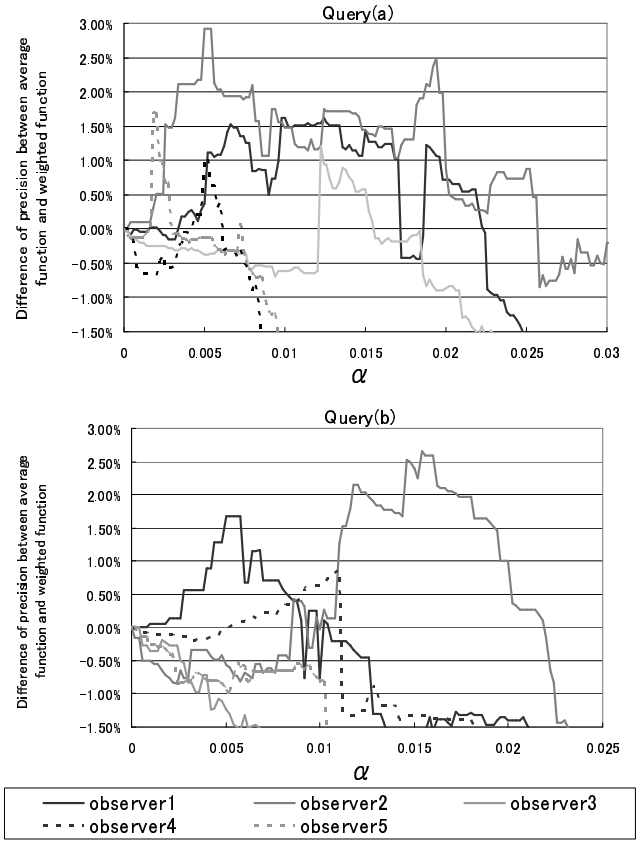

図 9　実験 2 の相加平均の検索精度と重み付け後の検索精度の差

Fig. 9　Difference of precision between average function and weighted function of experiment 2.

移動の特徴や意図を抽出することによって新たな類似度の提案を行った．提案手法では，利用者の移動軌跡データに存在する滞在箇所の説明文書をメタデータとして利用することによって類似度を算出した．そのため，利用者にとって直感的に類似した移動軌跡データの類似検索を行うことが可能となった．

評価実験を行った結果，移動軌跡データに存在する複数の特徴ベクトルを統合する際は相加平均を用いることが良いことが分かった．さらに，平均を用いることによって，再現率が低い部分で利用者にとって有効な提案が行えることが分かった．また，平均を用いる事に加え，類似度の変化量を利用することによって検索精度の向上を考え，利用者によっては検索精度を向上させることが可能となった．

今後の課題としては以下のものが挙げられる．

(1)　本研究では，類似度を算出する際に，移動軌跡データに存在する複数の特徴ベクトルの平均を算出することと，移動軌跡データに存在する複数の特徴ベクトルと検索対象の特徴ベクトルとの類似度の変化量を用いた．しかし，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，順序が逆順で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，順序が異なる移動軌跡データは異なる特徴量が抽出されるべきだと考えている．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．さらに，順序関係以外にも移動軌跡データの特徴として移動した時刻や，施設に滞在した時間等が考えられる．それらの移動軌跡データ固有の特徴を考慮することが可能となれば，より利用者の直感にあった類似度の提案が行えると考えている．

(2)　検索結果の信用度 [9] として，内容の公平さ・妥当性，社会的受容度，作者の信用度・信頼性が挙げられる．今回の評価実験では Wikipedia を用いたが Wikipedia ではそれらの信用度が欠けている点があると考えられる．さらに，信頼できるサーチエンジン [9] としてサーチエンジンの検索方法やランキング方法等の仕組みが利用者に見えていること，個々の利用者がサーチエンジンの仕組みをカスタマイズできることが挙げられる．本提案はサーチエンジンとは異なるが，利用者の行動支援を目的とした提案を行う際，検索理由等のランキング方法の仕組みを利用者に提示することによって利用者の信頼が増すと考えられる．そこで，文献 [10] に述べられている手法を用いることによって，利用者の興味を分類することによって利用者に提案理由を含めた提案が行うことが可能であると考える．また，検索方法のカスタマイズという点において，利用者によって類似度を算出する際に用いる説明文書を変更することを可能にすれば，個々の利用者が文書を選択することによって利用者にとって信頼のできる提案を行うことができると考えている．

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．

文　献

[1] R.Agrawal, C.Faloutsos, and A.Swami. Efficient similarity search in sequence databases. *In Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithm*, pp. 69–84, 1993.

[2] C.Faloutsos, M.Ranganathan, and Y.Manolopoulos. Fast subsequence matching in time-series databases. *In Proceedings of the ACM SIGMOD Conference on Management of Data*, pp. 419–429, 1994.

[3] E. Bertino and et al. *Indexing Techniques for Advanced Database Systems*. Kluwer Academic Publishers, 1997.

[4] Antonin Guttman. R-trees: a dynamic index structure for spatial searching. *In Proceedings of ACM SIGMOD Conference of Management of Data*, pp. 47–57, 1984.

[5] 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.

[6] 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.

[7] J.Liu, O.Wolfson, and H.Yin. Extracting Semantic Location from Outdoor Positioning Systems. *MDM2006 workshop MCISME*, pp. 34–41, 2006.

[8] 北研二, 津田和彦, 獅子堀正幹. 情報検索アルゴリズム. 共立出版株式会社, 2002.

[9] 田中克己. サーチエンジンにおける信用度・品質評価について. データベースと Web 情報システムに関するシンポジウム (DBWeb2006), pp. 107–108, 2006.

[10] 河合由起子, 官上大輔, 田中克己. 個人の選好に基づく複数ニュースサイトの記事収集・閲覧システム. 情報処理学会論文誌データベース, Vol. 46, No. SIG8(TOD26), pp. 14–25, 2005.