

実世界中で行動する人間の移動軌跡データからの特徴抽出

柳沢 豊[†] 山田 辰美[†] 佐藤 哲司[†]

近年、GPSなどの位置取得デバイスの発達により、人間や車の移動軌跡データを高い精度で取得できるようになり、これを分析して行動の特徴を取り出す研究が進んでいる。例えば、店舗内の来客の移動軌跡データを分析し、来客の購買行動の特徴を取り出せば、商品の陳列の最適化を行うことができる。このような分析には、一般にシンボル列化された行動情報に対し、データマイニングの手法を適用する方法が用いられる。しかし、一般にこのシンボル化の作業は手動で行う必要があり、この作業がユーザに大きな負担になっていた。そこで本研究では、ある程度自動でシンボル化を行える手法を提案し、また一連の分析作業を視覚的に行えるインタフェースシステムの実装を行った。これによりユーザの労力が軽減され、比較的容易に移動軌跡データからその特徴を抽出することが可能になった。

Extracting Motion Patterns from Trajectories of Moving Objects

Yutaka YANAGISAWA[†] Tatsumi YAMADA[†] Tetsuji SATOH[†]

In this paper, we describe the trajectory analyzing system for extracting motion patterns from trajectories of moving objects, such as walking people, driving cars, and so on. In recent years, we can get high accurate trajectory data of moving objects using GPS and other positioning sensors. The trajectory data include the motion patterns of the moving objects, and the motion patterns can be used in various application systems. For example, shop managers can arrange the products in their shops referring the motion patterns of customers visiting to the shop. The several sequence mining methods has been proposed for extracting motion patterns from trajectories. Therefore, we proposed an automatic symbolizing method for extracting motion patterns from trajectories. Moreover, we developed a visual interface system for supporting users to analyze trajectories.

1 はじめに

近年、GPSなどの位置取得デバイスにより継続的に取得された人間の移動軌跡データを分析することで、ユーザの行動の特徴を抽出しようとする研究が進んでいる。たとえば、オフィス内を移動する多くのユーザの移動軌跡を分析すれば、オフィスの構造を最適化するための情報を得られると考えられる。同様に、車の移動軌跡データから交通渋滞の予測や、道路整備計画を立てる手がかりが得られる[10]。このほかにも、店舗やイベント会場での来客の行動パターン分析、タクシーやトラックの配車管理[8]、また動物の生態調査への応用[5]など、数多くの研究が進んでいる。

従来の行動分析に関する研究の多くは、人間などが行った行動に予めいくつかのシンボルを割り当てておき、そのシンボル間の順序関係[6, 1]や遷移状態[4, 9]を取り出す方法が用いられている。いずれの方法も、分析対象

となるデータに対して、分析者がある程度手動で行動名を表すシンボルを割り当てられていることが前提である。例えば、オフィス内の行動であれば「会議」「食事」「電話」などのシンボルが割り当てられる。そしてシンボル列に対し、様々な特徴抽出アルゴリズムを適用することで、主に「どのイベントからどのイベントへ順番に遷移していったか」という特徴が取り出せる。このシンボルの与え方によって分析結果は大きく変化するため、ユーザは必要な結果が得られるまでシンボルを与えて分析しなおすという作業を繰り返す必要がある。しかし、取得された移動軌跡データは一般に膨大な量である、これら全てに対してシンボルを手動で割り当てる作業を頻繁に繰り返すことは、ユーザにとって大きな負担となる。この作業にかかるコストが高いことが、移動軌跡データを用いた特徴分析技術の実用化を妨げている最大の原因となっている。

このため、本研究では移動軌跡データに対するシンボルの割り当てを自動的にしながら、シンボル列から移動軌跡データの特徴を抽出する手法の開発を行った。またこの手法を用いて分析の作業を視覚的に行えるインタ

[†]日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
[‡]NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation.

フェースも実装した。本手法では、まず物体が移動した範囲を格子状に区切り、各格子にシンボルを与えておく。次に移動軌跡データの通過した格子の順序などを利用して、各移動軌跡データに対応するシンボル列を生成する。そして最後に、データマイニングのアルゴリズムを応用し、統合された各格子の間を移動するパターンを数え上げることで、移動軌跡データに含まれる「どこからどこへ行ったか」という特徴を取り出せる。分析作業を行うユーザは、実装したシステムを用いてシンボルの割り当て方を変化させながら試行を繰り返すことで、所用の結果を得ることができる。また本研究では、実際に GPS を用いて取得した、奈良市内を走る人力車の移動軌跡データを用い、実際にデータの分析作業を行った。

以下 2 章では、まず本研究で対象とする移動軌跡データについて簡単に説明する。次に 3 では、移動軌跡データに対して格子を利用してシンボルを割り当てながら、特徴抽出を行う手法について述べる。さらに 4 章で実装したシステムについて、実際にセンサにて取得した移動軌跡データへ適用した結果を示しながら説明する。

2 移動軌跡データ

本研究で用いる移動軌跡のためのデータモデル [11] について概略を示す。実世界中の移動する物体の移動軌跡は、図 1(左) に示すように時間 t から空間 \mathbf{R}^n への連続関数として表すことができる [3]。しかし、センサなどを用いて物体の移動軌跡を取得すると、実際には図 1(右) に示すように離散的な値しか得られない。このため、本研究では、移動軌跡データは不定間隔でサンプリングされた座標値の列としてモデル化し、連続化が必要な場合は下記の近似関数 $\tilde{\lambda}(t)$ を用いる方法をとる。

ある移動するオブジェクトの識別子を o とし、時刻列を $\mathbf{T} = \langle t_1, t_2, \dots, t_m \rangle$ と表し、 n 次元空間 \mathbf{R}^n 上のベクトルを $\mathbf{p} = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$ とする。このとき \mathbf{T} の各時刻に記録されたオブジェクト o の n 次元空間上の位置を表すベクトル列を $\mathbf{P} = \langle \mathbf{p}_1, \dots, \mathbf{p}_m \rangle$ とすると、オブジェクト o の移動軌跡データ λ は $(o, \mathbf{T}, \mathbf{P})$ の三つ組として表すことができる。なお $\lambda(t_i) = \mathbf{p}_i$, $|\dot{\lambda}| = m$ とする。

また、本研究では簡単化のため移動軌跡データは、特に断らない限り 1 秒間隔で測定されているものとする。つまり $t_{i+1} - t_i = 1\text{sec}$ である。実際に GPS を用いて取得した移動軌跡データの例を図 1 に示す。これは、奈良市内で営業する人力車に GPS 受信機を取り付けて取

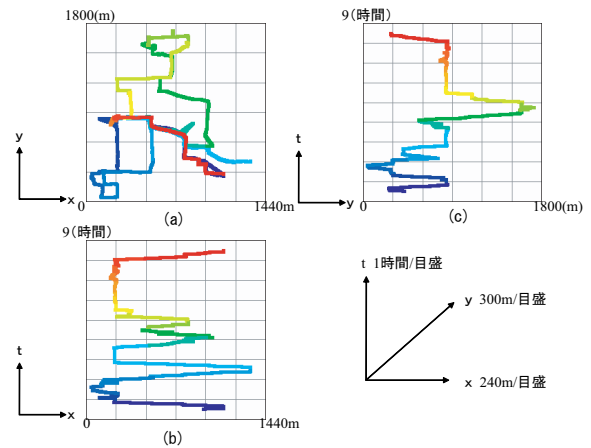


図 1: GPS を用いて取得した移動軌跡データ

得した、1 台の 8 時間分の連続した移動軌跡データである。図 1(a) は移動軌跡データを平面上にプロットしたものである。図 1(b), 図 1(c) はそれぞれ時間軸を縦にとり、 x 軸、 y 軸の動きを表したものである。

3 移動軌跡データからの特徴抽出

移動軌跡データは 2 章で述べたように、センサなどで取得された人間などの移動物体の通過したポイントを、空間上に時間順序に従ってプロットしたものである。このデータには、例えば個々の人間の移動にみられる癖、嗜好、傾向といった、いわゆる人間が移動するときに見られる一定のパターン（特徴）が含まれていると考えられる。このパターンを取り出すことができれば、それに基づいてオフィスの設備配置の最適化、店舗内の商品の配置の検討、あるいは道路の建設箇所や公共施設の配置場所の検討といった、さまざまな用途に利用できる。

従来より、こうした移動軌跡データの分析に使える手法として、時系列データからのパターン発見手法がある [7]。この手法は主に次の 2 つがある。

1. イベントなどのシンボル列から、シンボル間の遷移確率 [4] や推移関係 [1] を見つけ出す方法で、シーケンスマイニングと呼ばれる。Web ページの巡回パターン発見、道路ネットワーク内の車の遷移パターンの発見などの用途でよく用いられている。
2. 時系列データの中から、時間に沿ったデータの推移状態の近い部分を、データ全体の中から見つけ出す方法 [2] で、サブシーケスマッチなどと呼ばれる。株価の推移パターンを過去の推移と比較し、似てい

る部分を探すなどの用途に用いられる。

移動軌跡データは時系列データの一種であり、いずれの手法も適用が可能である。ただし本研究では、移動する物体が「どこからどこへ」移動したかということに着目した分析を目的としているため、1の方法に基づいた分析手法の開発を行った。

移動軌跡データを、イベント（シンボル）列に基づいて分析する場合に問題となるのは、2章で述べたような座標と時刻の組の列である移動軌跡データを、どのようなシンボル列に変換するかという点である。従来の研究の多くは、物体が移動した範囲の地図上の区画名を利用してシンボル化する手法を用いることが多かった。例えばオフィス内の移動軌跡データの分析を行う場合、オフィス全体の地図にある区画名を用いてシンボル化を行う。地図上にある区画名、例えば「会議室」「廊下」「休憩室」などといったものをシンボルとして使い、そのシンボル間をどの順番で移動していったか、という形でシンボル化される。この手法は簡単であるが、例えば駅構内や公園、広場のように比較的自由に移動できる空間の場合、予め区画割りをすることが難しい。こうした場合、移動軌跡データに対してどのようなシンボルを与えると、どのような特徴がえられるかということを試行し、分析者の求める特徴を求めていくことになる。つまり、シンボルを予め固定的に与えてから特徴を抽出するのではなく、シンボル化と特徴抽出を同時に行う必要がある。

以下本節では、まず移動軌跡データのシンボル化について簡単に説明する。次にシンボル列化された移動軌跡データから特徴を取り出す方針について述べ、さらに特徴抽出の結果をシンボル化の過程にフィードバックさせながら、シンボル化を再定義する手法について述べる。

3.1 シンボル化

移動軌跡データは2章で述べたように、座標と時刻の組の列 λ であらわされる。これを、適当なシンボルの集合 $S = \{s_1, s_2, \dots, s_n\}$ を用いてシンボル化列に変換したものを $S_\lambda = \langle s_1, s_2, \dots, s_m \rangle$ で表す。 S の各シンボルに適当な空間上の領域を割り当てておき、移動軌跡データ λ が通過した領域に割り当てられているシンボルを、通過した時間順序に書き並べたものが S_λ である。図2にシンボル化の例を示す。なお、簡単のためここでは各領域は重複した部分を持たないこととする。すなわち、物体がある時刻に同時に異なる2つの領域内に存在することはない。この制限のもとでは、 λ と S の元で生成

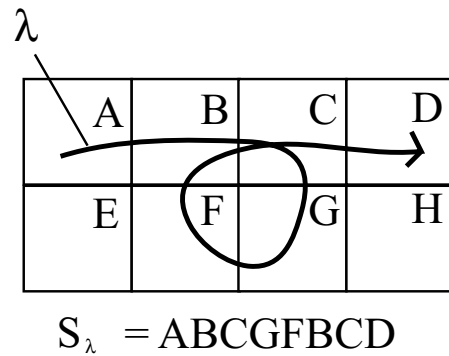


図 2: 移動軌跡データのシンボル化例

される S_λ は一意に定まる。

通常、シンボル集合 S は固定的に与えられるが、本研究では特徴抽出の結果に応じて S を変化させることで、柔軟な分析を実現する。具体的な手法については次の節で述べる。

3.2 推移を用いた特徴付け

シンボル列からの特徴抽出法としては、シンボル間の推移を見つけ出す手法が一般的に用いられる。シンボル間の推移を見つける方法は、従来のシーケンスマイニングの手法として、いくつか提案されている [1]。一般にシンボルの推移とは、シンボル列 $S_\lambda = \langle s_1, s_2, \dots, s_m \rangle$ について、 $s_i \rightarrow s_j (i < j)$ となるような組全てのことである。図2を例にとると、例えば $A \rightarrow B$, $A \rightarrow C$, $B \rightarrow C$, $B \rightarrow G$ のような組のことを言う。直観的には $A \rightarrow B$ は「 A を訪問したあとに B を訪問した」と解釈できる。このような遷移は、シンボル列の長さを m とすると、最大で $m(m+1)/2$ 種類存在することになる。

この最も単純な遷移の定義を用いて遷移を取り出すと、図3(a)のようになる。しかし、この結果には冗長性がある。例えば $ABAB$ というシンボル列からは $A \rightarrow B$ という遷移が3つ存在することになる。直観的には $A \rightarrow B$ という遷移が2回あるように見えるが、先頭の A と末尾の B の組み合わせも1回の遷移と数え上げるため、3回とかぞえられる。このような冗長性をなくすために、遷移を $s_i \rightarrow s_j (i < j)$, ただし $s_i = s_k$ または $s_j = s_k$ となるような $s_k (i < k < j)$ が存在しないものとして定義しなおす。このように定義すると、図3(b)のような結果となり、冗長性がなくなる。本研究では、再定義した推移を移動軌跡データの特徴として用いる。シンボル列から取り出された推移の集合は、「物体がどこからどこ

ABCBACB

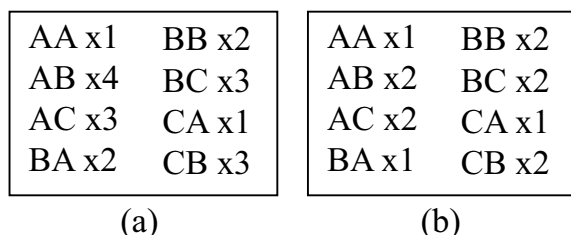


図 3: シンボル間の遷移

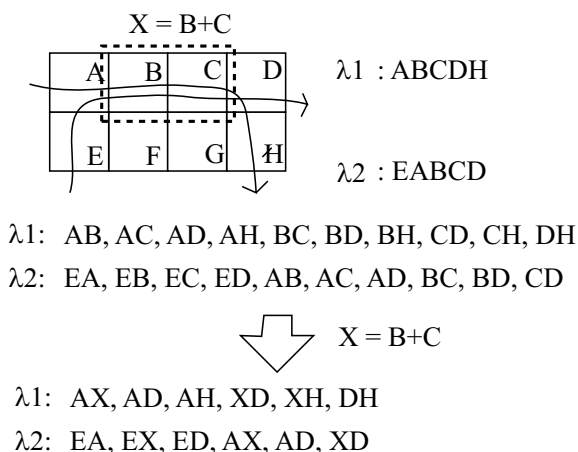


図 4: 再シンボル化

に移動したか」という特徴を表すデータとなっている。

3.3 シンボルの統合

先に述べたように、公園などの広い空間では移動軌跡データをシンボル化するにあたって、どのような区画を公園内に割り付ければよいか、分析前には未知であることが多い。このような場合、初期に適切な区画リスト(シンボル集合)を与えて一旦シンボル化しておき、特徴抽出の過程で次第にシンボルを統合化していくことで、シンボル化をある程度自動化できる。

シンボルの割り当てが冗長になる例を図 4 に示す。この図は 2 つの移動軌跡データについて同時に推移を抽出した例を示している。それぞれにデータに対して A から H までのシンボルを使ってシンボル化すると、図の中段のような多くの推移が見つかる。ここでよく見ると、任意の B, C 以外のシンボル Z について、 $Z \rightarrow B$ が存在すれば $Z \rightarrow C$ が存在し、その逆も成り立っている。同時に $B \rightarrow Z$ が存在するときは $C \rightarrow Z$ も必ず存在し、その逆も成り立つ。このとき B, C をひとつの別の

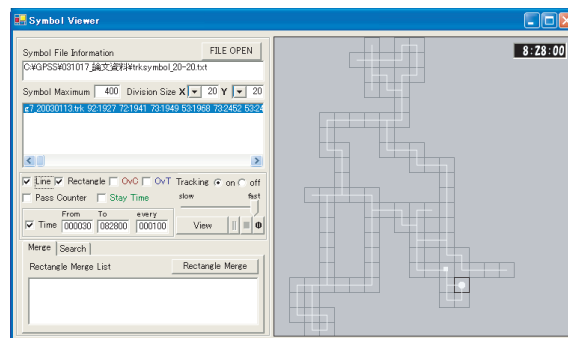


図 5: 移動軌跡の表示例

シンボル X に統合することで、冗長性を削減することができる。具体的に置き換えて推移を再抽出すると、図の下段のようになる。

この統合作業を再帰的に繰り返すことで、最終的に推移に含まれる冗長性を削減しながら、シンボルを統合していくことができる。

4 実装

本研究では、センサで取得された移動軌跡データに対して、前章までで述べた手法に基づいて分析するための、インタフェースシステムを実装した。本章では、これら実装したシステムについて説明する。システムは次の 3 つのツールから構成される。なお、図中に表示されているデータは、図 1 中のデータと同じものである。

1. 移動席データ修正ツール

センサから得られた移動軌跡データは、データの測定誤差やデータの欠落などがある。これらについて、分析の前段階として誤差の除去やデータの補完を行うツールである。

2. シンボル化・推移発見ツール

3 章で述べた手法に基づいて、データのシンボル化、推移の数え上げ、シンボルの統合を行うことが出来るツールである。このツールでは、シンボルを与える方法として、ユーザが指定した区画リストを用いる方法と、全体の領域を適当な数の区画に仕切って、初期の区画リストとして使用する方法の両方を用いることができる。ユーザは、得られた結果に応じて初期区画リストを変更することで、より望ましい結果に近づけることができる。図 6 に縦 20 区画、横 20 区画になるように区画割りした場合の、移動軌跡データのシンボル化例を示す。

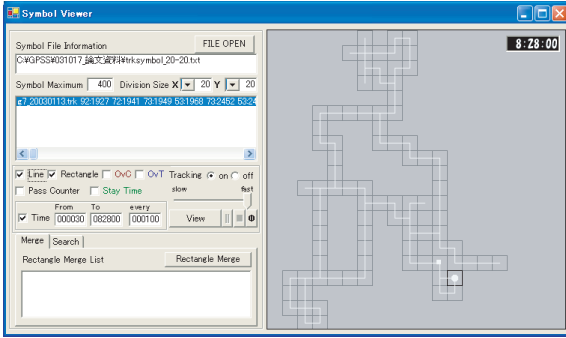


図 6: シンボル化された移動軌跡の表示例

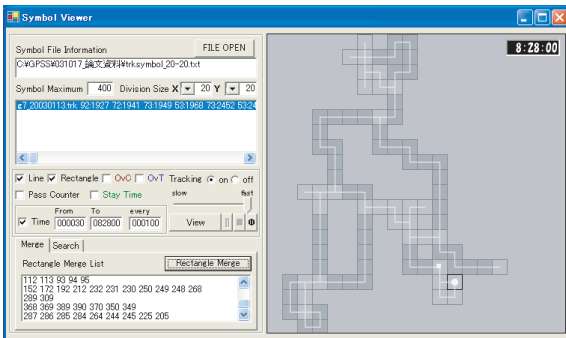


図 7: シンボル統合の例

また、このツール上では推移の発見とシンボルの統合を行うことができる。前章で述べた統合手法を再帰的に繰り返し、統合可能なシンボルを全て統合した状態を図 7 に示す。この図で、薄い色のついている区画は、統合可能であることを示している。

3. 推移結果可視化ツール

えられた特徴情報である、推移の様子を可視化するツールである。実際に得られた結果を表示した例を ?? に示す。この例では、ツールの左側のリストに、移動軌跡データから得られた推移の種類と数が一覧で表示されている。右側の表示スペースには、ユーザが選択した区画からどの区画に推移したかを、線で表示している。また、推移数がいくつあったかも、各区画内に数字で表示される。このツールを使うことで、対象の移動物体がどこからどこへよく移動しているのか、ということは視覚的に容易にわかる。

5 まとめ

本研究では、移動軌跡データの分析のためのシンボルを自動的に割り当てる手法と、訪問順序に基づく特徴抽

出の方法を新たに開発した。また、分析支援を行うためのインタフェースシステムも実装した。これらを用いることにより、移動軌跡データをさまざまな目的で分析することが可能になった。今後は、さらに実データを用いて分析実験を行い、手法の有効性の評価を進める予定である。これらに関しては現在、小売店店舗内での人間の移動軌跡データを用いた行動パターン取得実験など、いくつかの実験を実施中である。

参考文献

- [1] Rakesh Agrawal and Ramakrishnan Srikant. Mining sequential patterns. In Philip S. Yu and Arbee S. P. Chen, editors, *Eleventh International Conference on Data Engineering*, pp. 3–14, Taipei, Taiwan, 1995. IEEE Computer Society Press.
- [2] Kaushik Chakrabarti, Eamonn J. Keogh, Sharad Mehrotra, and Michael J. Pazzani. Locally adaptive dimensionality reduction for indexing large time series databases. *ACM Transactions on Database Systems*, Vol. 27, No. 2, pp. 188–228, 2002.
- [3] E. Clementini and P. Di Felice. Topological invariants for lines. *IEEE Transaction on Knowledge and Data Engineering*, Vol. 10, No. 1, pp. 38–54, 1998.
- [4] Philip Hingston. Using finite state automata for sequence mining. In *Twenty-fifth Australasian conference on Computer science*, Vol. 4, pp. 105–110. Australian Computer Society, Inc, 2002.
- [5] Patrick Laube and Stephan Imfeld. Analyzing relative motion within groups of trackable moving point objects. In *GIScience 2002 Conference Proceedings*, pp. 132–144, 2002.
- [6] Heikki Mannila, Hannu Toivonen, and A. Inkeri Verkamo. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, Vol. 1, No. 3, pp. 259–289, 1997.
- [7] Srinivasan Parthasarathy, Mohammed Javeed Zaki, Mitsunori Ogihara, and Sandhya Dwarkadas. Incremental and interactive sequence mining. In *CIKM*, pp. 251–258, 1999.

- [8] Michalis Vazirgiannis and Ouri Wolfson. A spatiotemporal model and language for moving objects on road networks. In Christian S. Jensen, Markus Schneider, Bernhard Seeger, and Vasilis J. Tsotras, editors, *SSTD 2001*, Vol. 2121 of *Lecture Notes in Computer Science*, pp. 20–35. Springer-Verlag, 2001.
- [9] 服部加奈子, 小磯貴史, 今崎直樹. 滞在時間を考慮した主要行動パターン抽出方法の検討. 第 17 回 人工知能学会大会 (講演番号 2F1-02), 2003.
- [10] 杉野勝敏, 朝倉康夫, 羽藤英二. 移動体位置情報を用いた交通行動シミュレーションモデルの開発. 土木学会, 第 26 回土木計画学研究発表会 (講演番号 188), 2002.
- [11] 柳沢豊, 赤埴淳一, 佐藤哲司. 移動軌跡データに対する類似度検索手法. FIT2002 一般講演論文集第二分冊, pp. 37–38, 2002.