

人間の行動知識を用いた移動軌跡データからの固有行動検出

Eigenbehavior Detection from Trajectory Data by Human Behavior Knowledge

田中 優子^{*1} 関 和広^{*2} 上原 邦昭^{*2}
 Yuko Tanaka Kazuhiro Seki Kuniaki Uehara

^{*1}神戸大学工学部情報知能工学科

Department of Computer Science and Systems Engineering, Kobe University

^{*2}神戸大学大学院システム情報学研究科計算科学専攻

Graduate School of System Informatics, Kobe University

The conventional analysis of behavior of people has used principal component analysis (PCA). However, PCA detects the characteristic behavior consisting of the information only regarding time and places, i.e., such as when and where people are. This paper proposes a method to detect eigen behavior using human behavior knowledge and data mining techniques. The method deals with information of transportation and relationship between places and builds an ontology of eigen behaviors. The first step is to give meanings to trajectory because the moving trajectory data is geometric. This step uses clustering and conditional random fields. Then, by mining patterns from provided semantic trajectory, eigen behaviors are detected. We perform an experiment using trajectory data in Tokyo metropolitan area and build an ontology of office workers, university students and high school students.

1. はじめに

近年, GPS の普及によって, 位置情報データが容易に集められるようになった. 位置情報データからは, 物体の移動軌跡を得ることができる. したがって, 都市部での人間の移動軌跡データを用いて行動分析をすれば, 都市計画やマーケティングに活かすことができる.

従来の行動分析では, 主成分分析などの統計的な手法が用いられてきた. 文献 [Eagle 09] では, 主成分分析を用いたグループの固有行動を検出する手法が提案されている. 人々をクラス別にグループ構成し, グループの行動を表す行列から, 主成分分析によって特徴的な行動を得ることができる. しかしこの方法では, どの時間にどの場所に居るかといった, 時間と場所の情報のみからなる行動しか得ることができない. 人の行動を把握する上では, 移動中の交通機関など, その他の情報や場所間の関係性が必要である. そこで本研究では, 人間行動に関する知識とデータマイニングによってオントロジーを構築する手法を提案する.

オントロジー構築では, まず幾何的なデータである位置情報データに対し, 階層的に概念を定義し, 移動軌跡の意味付けを行う. 具体的には, 移動軌跡を stop と move という 2 つの概念から構成する. それぞれについて, 行動知識をもとに場所と移動手段の推定をすれば, 移動軌跡に関する意味付けを行うことができる. そして, 意味付けされた移動軌跡データからパターンマイニングを行い, 概念間の関係性を発見する. 人のクラスごとに固有な関係性を発見し, オントロジーを構築すれば, そのクラスの人々に見られる特徴的な行動を検出できる. 実験では, 東京都市圏の人の流れデータ [人の流れプロジェクト 13] を用いて, オントロジー構築を行う.

2. 提案手法

2.1 軌跡の定義

移動軌跡データは, 座標と時間から成る時系列データであるため, 意味的な情報は付けられていない. 移動軌跡に意味付けを行うために, 階層的に概念を定義する. 移動軌跡に関する定義は [Renso 12] が定義したように, 観測軌跡・統語的軌跡・意味的軌跡の 3 種類の階層に分けられる.

定義 1 (観測軌跡) 観測軌跡とは, 生の移動軌跡データのことである. 具体的には, 座標と位置測定時間から成る時系列データである. 物体 O の観測軌跡 T は $T_O = \langle p_1 \dots p_n \rangle$ で表すものとする. ここで, $p_k = (x_k, y_k, t_k)$, x, y は座標, t は時刻であり, n は物体 O が移動する間に観測したサンプル点の数である.

定義 2 (stop) 観測軌跡 $T_O = \langle p_1 \dots p_n \rangle$ から, 空間的しきい値 $th_{spatial}$ と時間的しきい値 $th_{temporal}$ を用いて, 次のように stop を定義する.

$$S = \{p_m \dots p_k\} | 1 < m \leq k \leq n \wedge m \leq i \leq k \text{ Dist}(p_m, p_i) \leq th_{spatial} \wedge \text{Dist}(p_m, p_i) \geq th_{temporal}$$

ここで Dist は 2 点間のユークリッド距離を求める関数, Dur は時間差を求める関数である. この式は, 人が距離 $th_{spatial}$ 以内の範囲に, 時間 $th_{temporal}$ 以上滞在していることを表している.

定義 3 (move) move は, 連続する 2 つの stop 間の最大部分列である.

$$M = \{p_a \dots p_b\} > | S_k = \{p_m \dots p_k\} \wedge S_{k+1} = \{p_{m'} \dots p_{k'}\} \wedge a = k + 1 \wedge b = m' - 1$$

この式は, ある stop から次の stop までの移動を move とすることを表している. stop と move は移動軌跡を構成する重要な要素である [Alvares 07b].

連絡先: 田中 優子, 神戸大学工学部情報知能工学科, 神戸市灘区六甲台町 1-1, tanaka@ai.cs.kobe-u.ac.jp

定義 4 (統語的軌跡) stop と move の繰り返しで移動軌跡を表したものを統語的軌跡と呼ぶ。統語的軌跡 T_{syn} は $(T_{id}, O_{id}, B.E.S, M)$ から成る。ここで T_{id}, O_{id} はそれぞれ移動軌跡とその物体の ID である。 $B.E.S$ は移動軌跡の始点・終点・stop の集合であり、 M は move の集合を表している。

定義 5 (意味的軌跡) 意味的軌跡は、統語的軌跡に意味的情報を付加したものである。移動軌跡の文脈によって与えられる場所と移動手段のラベルの集合から、それぞれ stop と move にラベル付けされる。

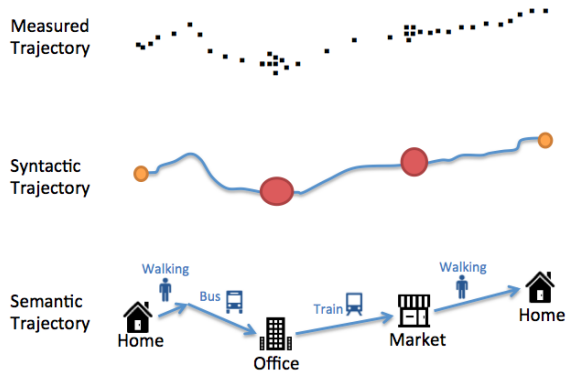


図 1: 3 段階の移動軌跡の定義

図 1 は、3 種類の軌跡を図示したものである。stop には場所 (家、会社、店など) の情報がつけられ、move には移動手段 (徒歩、バス、電車など) がラベル付けされる。移動軌跡データにこれらの意味付けを行えば、移動軌跡全体の意味を解釈することができる。たとえば、家から会社へ行く移動は通勤の移動軌跡であり、お店に行く移動は買い物の移動軌跡であるということが分かる。

2.2 クラスタリングによる場所の検出

都市部における人々の行動には、年齢や職業といったクラス別のグループごとに集まる場所がある。例えば、大学生が昼間に集まる場所は大学であるし、会社員が集まる場所は会社である。このように、クラス別に人の集まる場所が分かれば、行動知識にもとづいて意味付けを行うことができる。人の集まる場所は、密度に基づくクラスターを求めることで検出できる。このクラスタリングには DenStream [Cao 06] を用いている。DenStream は、密度に基づくクラスタリングアルゴリズム DBSCAN を、ストリームデータに適用したものである。DBSCAN はユークリッド距離からクラスターを発見するアルゴリズムである。具体的には、距離のしきい値以下の点同士を隣接していると見なし、ある密度以上に隣接する点が集まっている箇所を検出する。このようにすれば、人の集まる場所を検出できる。また、DenStream はストリームアルゴリズムであるため、大量の移動軌跡データを扱うことができる。

2.3 移動手段の推定

都市部でよく使われる移動手段として、徒歩・自転車・バス・車・電車の 5 つを考える。これらの移動手段は、速度をもとに特定できると考えられる。移動軌跡データの座標と時間から速度の系列データを求め、それぞれに対して最適な移動手段を推定することを考える。これは系列ラベリング問題として捉え

ることができる。そこで、自然言語処理の分野においてよく使われる機械学習の手法である CRF [Lafferty 01] を用いる。

自然言語処理で CRF を用いる例として、単語列に品詞のラベルをつけるという問題がある。この問題の場合、より精度よくラベル付けを行うためには、各単語のみから品詞を推定するだけでなく、前後の品詞を考慮する必要がある。そこで CRF では、素性として前後の単語や品詞を与え、品詞の順序のような構造を学習する。そして、学習した重みによって、品詞を求めたいデータに対し、各品詞が割り当てられる確率を求めることができる。

同様に、移動手段の推定でも、構造学習を行うことが有用であると考えられる。たとえば、バスと車は速度に関して大きな違いがない。しかし、人の行動知識として、電車の移動の後に車に乗る確率は低いが、バスに乗る可能性は高い、といった移動手段の前後の関係性が考えられる。この関係性を考慮して移動手段の推定を行えば、より精度よく推定できる。また、速度のみから移動手段を推定する場合のもうひとつの問題点として、同じ移動手段の移動の間に、速度が頻繁に変化するという点がある。例えば、車の移動は信号のたびに停止するため、一時的に速度が低下する。速度のみを見ると、これは徒歩や自転車の移動と見なされてしまう。そこで CRF を用いて、移動手段は頻繁に入れ替わることがないということを学習すれば、この問題を解決することができる。また、CRF では、時間や座標などのその他の情報も素性として簡単に加えることができる。

2.4 移動パターン

オントロジーを拡張するためにパターンマイニングを行い、概念間の関係性を発見する。まず [Alvares 07a] のように、支持度と移動パターンを定義する。

定義 6 (支持度) 移動軌跡の集合を $T = \{t_1, t_2, \dots, t_n\}$ とする。 $\overrightarrow{AB}$ は stop A から stop B へ移動する move を表す。ただし A, B は stop のラベルである。move $\overrightarrow{AB}$ の支持度を $sup(\overrightarrow{AB})$ で表し、次のように定義する。

$$\overrightarrow{AB} = \frac{|\{t \in T | \overrightarrow{AB} \in t\}|}{|T|}$$

ここで $|X|$ は集合 X の要素数である。

定義 7 (移動パターン) move の支持度が ς であるとき、 ς が与えられた最低支持度 $minsup$ より大きければ移動パターンとする。

意味的軌跡からの移動パターンの検出によって、場所間の関係性が検出される。

2.5 時間パターン

主成分分析では、まず、図 2 のようにグループの行動を行列で表す。この図は、会社員のグループの行動を表しており、1 行が 1 人の 1 日の行動を表すベクトルである。各滞在場所について居るか居ないかを 0 か 1 で表している。この行列を主成分分析すれば、図 3 のような固有ベクトルが得られる。このベクトルが、グループの特徴的な行動を表していることになる。たとえば、8 時までと 20 時以降は家、9 時から 19 時は会社に居るといった行動が検出される。この情報から、stop の属性として時間の概念をオントロジーに加える。

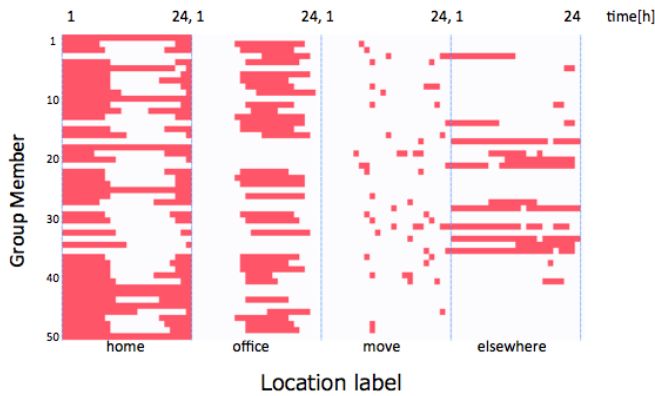


図 2: 主成分分析に用いる会社員のグループの行動行列

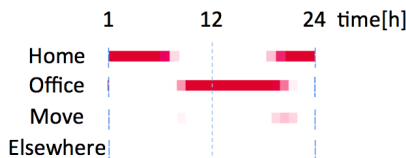


図 3: 主成分分析で得られる固有ベクトル

3. 評価

3.1 実験データ

実験データには、東京大学空間情報科学研究センター (CSIS) が提供している平成 20 年度東京都市圏の「人の流れデータ^{*1}」を用いる。このデータは、パーソントリップ調査 (PT 調査) データ^{*2}に対し、GPS データのような座標の時系列データとなるように、ジオコーディング処理を施したものである。PT 調査とは、どのような人が、どのような目的で、どのような移動手段で、どこからどこへ移動したかを把握するために、都市圏で行われているアンケート調査である。そのため、人の流れデータには、時間と座標以外に、性別や年齢、職業といった人のカテゴリと、移動手段、移動目的といった情報が含まれている。今回の実験では、会社員・高校生・大学生の 3 つの職業別クラス的数据を用いる。また、移動手段の情報をを用いて CRF による構造学習を行う。移動目的は、クラスタリングによる場所推定の精度を求めるために使用する。

3.2 実験環境

CRF の実装には、オープンソースソフトウェアである CRF++^{*3} (Ver. 0.58) を利用する。DenStream には、Massive Online Analysis (MOA)^{*4} (Ver. 2013.11) を用いる。MOA は、データストリームマイニングを行うためのオープンソースフレームワークである。さらに、主成分分析の実装には京 MapReduce (KMR)^{*5} を利用する。MapReduce とは、データをキーとバリューのセットとして扱い、キーごとにデー

タを集約して並列処理を行うフレームワークである。本実験では、スーパーコンピュータ富士通 Fx10 上で、KMR を用いて主成分分析の行列演算を並列処理で実装している。

3.3 実験結果

クラスタリングによって、高校を検出した結果を図 4 に示す。また、大学を検出した結果を図 5 に示す。図中の黒いアイコンがクラスタリングで検出したもの、緑のピンが正しい座標である。両方が重なっていることから、正しい場所が検出できたことがわかる。会社員・高校生・大学生の場合について、精度を求めた結果が表 1 の適合率である。いずれも高い精度で正しい場所が検出できている。しかし、図 4 中では緑のピンのみが表示されている場所も多くあり、検出できていない場所があるということが分かる。正しい場所のうち、どれだけ検出できたかを表すのが表 1 の再現率である。いずれも低い値であり、密度に基づくクラスタリングでは、データを取った人が少ない高校などは見つからないことが分かる。



図 4: クラスタリングで検出した高校

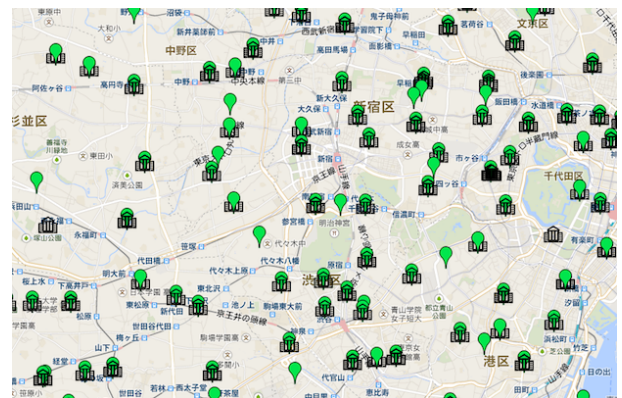


図 5: クラスタリングで検出した大学

表 1: クラスタリングによる場所検出の精度

	高校	大学	会社
適合率	96.3 %	77.9 %	96.1 %
再現率	57.6 %	32.1 %	40.6 %

*1 <http://pflow.csis.u-tokyo.ac.jp/index-j.html>

*2 <http://www.tokyo-pt.jp/person/index.html>

*3 <http://crfpp.googlecode.com/svn/trunk/doc/index.html>

*4 <http://moa.cms.waikato.ac.nz/>

*5 <http://mt.aics.riken.jp/kmr/>

次に、CRF による移動手段の推定を行い、推定の精度を求める。訓練データには 40,000 個、テストデータには 10,000 個の移動軌跡を用いている。学習に用いた素性は、速度、前後の移動手段、座標、時間である。推定の結果を図 6 に示す。図中の下の数字は各時刻における時速を表しており、速度の変化によって、正しく移動手段が特定されていることが分かる。推定の精度は、適合率が 75.4 % であった。CRF を用いて、前後の移動手段も素性として構造学習を行うことで、移動手段を求めることができた。



図 6: CRF によって推定した移動手段

次に、高校生の移動パターンの検出結果を図 7 に示す。最低支持度は $minsup = 0.1$ とする。各場所間の移動の支持度を求めることによって、家から学校に登校する人は 83 %, 学校から家ではなく、他の場所へ行く人は 10 % というように、移動の割合が分かる。また、その他から学校へ行く移動は、最低支持度より低い値であり、移動パターンとして検出されなかった。移動パターンは移動目的で表すことができる。例えば、家→学校は登校、学校→その他は寄り道、家→その他は外出となり、移動の属性としてオントロジーに加えられる。

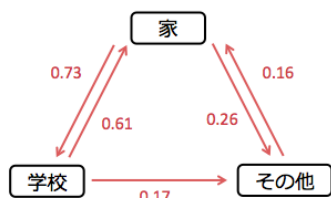


図 7: 高校生の移動パターン

以上の結果から、高校生の移動軌跡について固有行動を検出し、特徴的行動を表すオントロジーを構築した結果を図 8 に示す。生の移動軌跡データを stop と move の統語的軌跡に変換し、さらに意味付けを行って、階層的に概念が構築されている。また、時間や移動目的の概念によって固有行動を表すことができた。

4. おわりに

本稿では、人間の行動知識を利用して GPS による移動軌跡データから行動分析を行う手法を提案した。行動知識をもとにクラスタリングや機械学習によってデータに意味付けを行った。意味付けされた移動軌跡から移動パターンの検出を行い、クラス別の固有行動のオントロジーを構築した。今後、さらに概念を細分化して行動分析して、より特徴的な固有行動の発見を目指す。

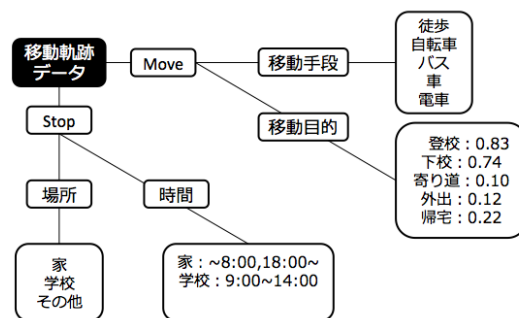


図 8: 高校生の特徴的行動を表すオントロジー

参考文献

- [Alvares 07a] Alvares, L. O., Bogorny, V., Macedo, de J. A. F., Moelans, B., and Spaccapietra, S.: Dynamic Modeling of Trajectory Patterns using Data Mining and Reverse Engineering, in *Proc. of the 26th International Conference on Conceptual Modeling*, Vol. 83, pp. 149–154 (2007)
- [Alvares 07b] Alvares, L. O., Bogorny, V., Kuijpers, B., Macedo, de J. A. F., Moelans, B., and Vaisman, A.: A Model for Enriching Trajectories with Semantic Geographical Information, in *Proc. of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems*, p. 22 (2007)
- [Cao 06] Cao, F., Ester, M., Qian, W., and Zhou, A.: Density-Based Clustering over an Evolving Data Stream with Noise., in *Proc. of SIAM International Conference on Data Mining*, pp. 328–339 (2006)
- [Eagle 09] Eagle, N. and Pentland, A. S.: Eigenbehaviors: Identifying Structure in Routine, *Behavioral Ecology and Sociobiology*, Vol. 63, No. 7, pp. 1057–1066 (2009)
- [Lafferty 01] Lafferty, J., McCallum, A., and Pereira, F. C.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in *Proc. of the 18th international conference on machine learning (ICML)*, pp. 282–289 (2001)
- [Renso 12] Renso, C., Baglioni, M., Macedo, de J. A. F., Trasarti, R., and Wachowicz, M.: How You Move Reveals Who You Are: Understanding Human Behavior by Analyzing Trajectory Data, *Knowledge and Information Systems*, Vol. 37, No. 2, pp. 331–362 (2012)
- [人の流れプロジェクト 13] 人の流れプロジェクト 東京大学 空間情報科学研究センター: <http://pflow.csis.u-tokyo.ac.jp/index-j.html> (2013)