

移動軌跡群のパターン分析による特異性検出及び予測

秋山 直輝[†] 渡邊 優太[†] 太田 昌克^{††} 宮崎 敏明[†]
 会津大学コンピュータ理工学部[†]
 日本電信電話株式会社 NTT 未来ねっと研究所^{††}

1. はじめに

本稿では、複数の移動体の移動軌跡群についてその時系列変化を分析し、都市や街の状況を把握する手法について報告する。街の状況として、道路の混雑状況を扱い、周回バスの移動軌跡を利用して街の道路状況を把握する。関連研究として、人や車の移動パターンを分析し、経路予測を行うものがある[1][2]。それらは個々の動線上での経路予測を対象にしたものである。一方、本研究では、各時間帯における街全体の道路状況を1つのパターンとして扱い、その時間変化を分析することにより、時間変化の特異性検出や予測を行う。具体的には、周回バスの動線群を、グラフを用いて表現し、それらの時間変化をN-gram 言語モデル[3]によってモデル化することにより、将来の道路状況の予測、及び既に起きた道路状況変化の特異性評価を行う。

2. 動線群の表現

街の道路状況の変化を把握するために、周回バスに搭載したGPS (Global Positioning System) 受信機から得た軌跡データを用いる。道路状況としてGPS データを基に抽出したバスの動線群を扱う。今回の実験ではある一定速度以上で移動した動線群のみを抽出している。これは、道路の混雑状況を反映する動線群を抽出するためである。ある時間内の動線群から、図1に示すような動線をノード、接続した動線のノード間をエッジで結んだ無向グラフを作成する[4]。各ノードは動線の地理的情報(動線上のスポットの系列)を属性として持っている。このグラフ表現により、動線の地理的情報だけでなく、動線間の接続関係を考慮した動線群の時系列分析が可能となる。

3. 道路状況変化のモデリング

3-1 動線群の分類

動線群をグラフ表現することにより、動線群

Extraction and Predicting Singularity in Trajectory Patterns Based on a Pattern Analysis

[†]Naoki Akiyama, [†]Yuta Watanabe, ^{††}Masakatsu Ohta, [†]Toshiaki Miyazaki

[†]School of Computer Science and Engineering, The University of Aizu

^{††}NTT Network Innovation Laboratories, NTT Corporation

間の類似性をグラフ編集距離[5]によって求め、動線群を定量的に比較・分類することができる。動線群の分類には、k-medoids 法[6]を用いて、各パターンの代表グラフを決定し、代表グラフとのグラフ編集距離が最も小さいパターンに各動線群を分類する。各動線群にはそれぞれが属するパターンのラベルを付与する。

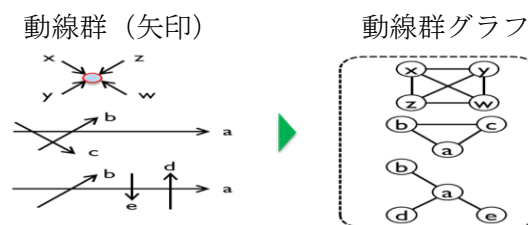


図1 動線群グラフの作成

3-2 N-gram 言語モデルの作成

ラベリングした動線群を、時系列にソートすることにより、動線群の変化を文字列として扱うことができる。本文字列に、N-gram 言語モデルを適用し、動線群の時間変化の特異性検出や予測を行う。

ラベルを W_i とすると、ラベル系列 (W_1, \dots, W_n) が出現する確率は、式(1)となる。ここでは、各ラベルの出現確率をその直前の $N-1$ 個のラベル系列のみに依存すると仮定し、式(1)を計算する。

$$P(W_1, \dots, W_n) = P(W_1)P(W_2|W_1)P(W_3|W_1, W_2)P(W_4|W_1, W_2, W_3) \dots P(W_n|W_1, W_2, \dots, W_{n-1}) \quad (1)$$

例として 3-gram 言語モデルを使用すると式(2)により動線群の時系列の出現確率を求めることができる。

$$P(W_1, \dots, W_n) = P(W_1)P(W_2|W_1)P(W_3|W_1, W_2)P(W_4|W_2, W_3) \dots P(W_n|W_{n-2}, W_{n-1}) \quad (2)$$

4. ケーススタディ

2014年4月から2014年10月上旬に収集した4台の市内周回バスのGPS データを用いて分析手法の評価実験を行った。実験条件を表1に示す。また、時速15km以上の動線群の例を図2に示す。

表1 使用したパラメータの値

動線群の作成単位	1時間毎
動線群抽出	時速15km以上
パターン数	8
ラベル数	10 (パターン数+1 日の始点終端ラベル)
N-gram	3-gram



図2 一定速度(時速15km)以上の動線群

N-gram 言語モデルを構築する前に、1時間毎に動線群を抽出してパターンに分類しラベル系列を作成した。さらに、1日の初めと終わりを示すラベルをラベル系列に挿入した。N-gram 言語モデルの作成には PALMKIT[3]を使用した。パターン数は、1日の計測時間が11時間であり、朝夕のパターンが類似していることを考慮し、8とした。テストデータに対しては、各パターンの代表グラフとのグラフ編集距離が最も小さいパターンに各動線群を分類しラベリングした。N-gram 言語モデルを構築したデータに出現しなかった動線群の時系列に対しては、出現確率が0になるのを防ぐために、バックオフスムージング[3]を行った。6月のデータでN-gram 言語モデルを構築し、9月から10月上旬のデータでテストを行った。N-gram 言語モデルの複雑さを表すパープレキシティ(平均分岐数)は約4であった。ラベル数は10であり、ランダムなモデルに比べて簡単なモデルが構築できている。

図3に動線群の1日の時系列の出現確率を示す。

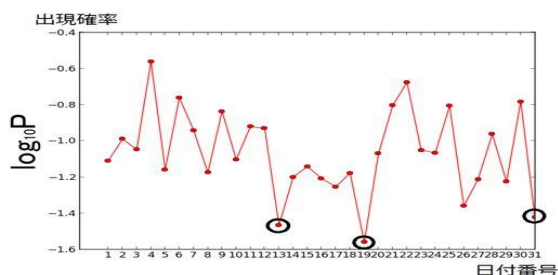


図3 動線群の時系列変化の出現確率

出現確率が低いほど街の道路状況変化がいつもと違うことを表している。

出現確率が低い9月23日(日付番号19)と10月5日(日付番号31)では、実際にお祭りやマ

ラソン大会がそれぞれ行われており、これらが周回バスの運行に影響を及ぼしていたことが分かる。

動線群の予測結果について表2に示す。ここでは、すべてのラベルについて条件付き確率を計算し、最も高いラベルを1時間後の動線群の状況として選択し、正しかった割合を正答率とした。

表2 動線群の予測結果

日時	正答率
6月1日~6月30日 (N-gram 言語モデル構築データ)	50.0%
9月1日~10月5日 (テストデータ)	32.3%

ラベル数が10であること、及び1つのラベルだけを選択していることを考慮すると、良好な結果が得られていると考える。

5. おわりに

複数の移動体の軌跡群についてその時系列変化を把握するための手法を提案した。また、実データを用いた評価実験により、手法の有効性を確認した。本手法は、日々の道路状況の予測や時間変化の特異性を分析できるため、街作りにおけるプランニングやその評価に役立つ。

本稿では、動線群は必ず1つのパターンに分類したが、どのパターンにも属さない動線群も存在する可能性がある。これらは、未知のパターンとして扱う必要があり、LOF[7]等による外れ値検出を行っていく予定である。

文献

[1] A. Monreale, F. Pinelli, R. Trasarti, F. Giannotti, "WhereNext: a Location Predictor on Trajectory Pattern Mining," Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09), pp.637-646, 2009.
 [2] F. Giannotti, M. Nanni, F. Pinelli, D. Pedreschi, "Trajectory Pattern Mining," Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07), pp.330-339, 2007.
 [3] 荒木雅弘, "フリーソフトでつくる音声認識システム," 森北出版, 2007.
 [4] 渡邊優太, 太田昌克, 宮崎敏明, "イベント空間における人の移動軌跡収集とその解析," 情報処理学会第76回全国大会, 5V-8, 2014.
 [5] K. Riesen, S. Emmenegger, H. Bunke, K. Riesen, S. Emmenegger, H. Bunke, "A Novel Software Toolkit for Graph Edit Distance Computation," Proceedings of 9th International Workshop on Graph Based Representations in Pattern Recognition, LNCS7877, pp.142-151, 2013.
 [6] Noor Kamal Kaur, Usvir Kaur, Dr. Dheerendra Singh, "K-Medoid Clustering Algorithm- A Review," International Journal of Computer Application and Technology, pp.42-45, 2014.
 [7] Markus M. Breuning, Hans-Peter Kriegel, Raymond T. Ng, Jorg Sander, "LOF: identifying density-based local outliers," Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, PP.93-104, 2000.