

高速道路実測データの機械学習による分析

只木進一

佐賀大学理工学部

概要

日本の高速道路では、およそ 2km 毎にインダクションループ計測器が設置され、流量と速度を計測している。これらのデータの特徴を抽出することは、交通流モデルが再現すべき特性を定めるうえで重要である。本研究では、これらのデータに対して機械学習によるクラスタリングを行った。その結果として、渋滞データを含む月のデータを自然な 3 つの状態への分類が行えること、及びそのモデルを渋滞の少ない月のデータに対して適用することで、同様の分類が行えることを示す。また、渋滞発生の前兆を捉える可能性についても議論する。

Machine Learning Analyses of Observed Highway Traffic Data

Shin-ichi TADAKI

Department of Information Science, Saga University

Abstract

Japanese highways feature induction loop devices, spaced approximately every 2 kilometers, to monitor traffic flow and speed. Extracting the characteristics of these data is crucial for defining the features that traffic flow models should reproduce. This report employs a clustering method using machine learning for applying to those data. We show that the data containing traffic jams are reasonably classified into three clusters, and the trained model can also classify the data with few traffic jams into the same three clusters. Furthermore, the possibility for capturing early signs of traffic jams is discussed.

1 はじめに

日本の高速道路では、およそ 2km 毎に、インダクションループ (induction loops) という計測器が埋設されている。そこでは、通過する車両の数と速度を計測している。筆者らは、東名高速道路のデータを中心に、5 分間の通過車両数と平均速度のデータを取得し、高速道路の車両の流れの全体像を把握するために、データの分析を行ってきた [1, 2, 3]。例えば、密度流量相関である基本図、速度と流量の時系列、あるいは渋滞クラスタの空間的移動の分析等を行い、その後の交通流モデル構築への手がかりを与えてきた。

しかし、これまでの実測データ分析では、データ

の可視化を通じて、その特徴を定性的に記述することしか出来てこなかった。本研究では、これらの実測データに対して機械学習を用いてデータ分析を行い、その特徴を客観的に抽出することを試みる。それらの分析を通じて、交通流の状態を捉える指針を得ることを目的とする。

今回用いるデータは、東名高速道路 172.65 キロポスト上り車線において、1996 年に取得されたものである。この地点は、日本坂 PA の西にあたり、2 車線の区間である。下流側 (東京側) には、長さ約 2.4km の日本坂トンネルがある。対象とするデータには、走行車線と追越車線、それぞれにおける 5 分毎の流量と平均速度が、時刻とともに記録されている。

2 K 平均法

K 平均法 (K-means method) は、データ空間内の標本点を自動的にグループ (クラスタ) 分けする機械学習の一手法である [4, 5]。ラベル付けなどの事前のモデル訓練を必要としないことから、広く利用されている。今回は、流量と速度が 2 車線分あることから、4 次元空間に分布するデータをクラスタに分ける。

K 平均法では、 k 個の中心とデータの距離に応じて、それぞれのクラスタへと分割する。つまり、 k 個の中心点に対するボロノイ分割を行う。分類を繰り返しながら、中心がクラスタ内データ点の重心と合うように調整するところが特徴である。クラスタ分割にはデータ点間のユークリッド距離を用いる。そのため、今回は、4 つの指標をそれぞれの平均値により規格化して分析する。

今回利用するデータの中には、データが連続して欠落している部分があり、事故や工事等によると想像できる。本研究では、一方の車線でも欠損値がある時刻のデータを、データ分析から除いている。

なお、異なる k に対して、silhouette score を計算することで、分割の質を評価し、適切な k を求めることができる。

あるデータ点 i の silhouette score $s(i)$ は、その点が属するクラスタ A に属する他のデータ点への平均距離

$$a(i) = \langle d_{ij} \rangle, j \in A \setminus \{i\} \quad (1)$$

と、点 i と最も近い他のクラスタ内の点 j が属するクラスタ B の各点への平均距離

$$b(i) = \langle d_{ij} \rangle, j \in B \quad (2)$$

を用いて

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

で定義する。この値は $(-1, 1)$ の範囲をとる。

3 分析結果

3.1 8月のデータ

今回分析するデータが観測された地点では、8月には、はっきりした渋滞が毎日 2 回程度発生している。つまり、自由流から渋滞への転移をデータ分析の観点から理解するのに適したデータである。このデータを K 平均法で分析することから始める。

図 1 は、各クラスタ内のデータ点の silhouette score の平均値 $\langle s(i) \rangle$ を、異なる k に対して表したものである。 $k = 3$ のところで最大となることから、 $k = 3$ が最もよい分類であることが分かる。

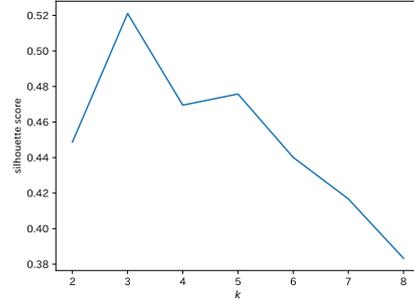


図 1: 各 k に対する silhouette score の平均

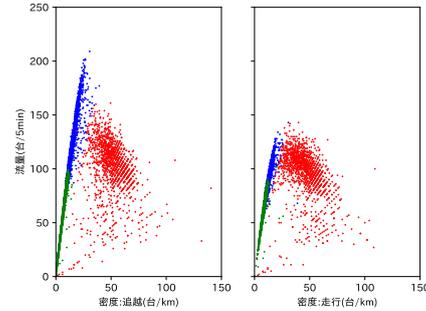


図 2: 8月データの $k = 3$ とした基本図。左は追越車線、右は走行車線。

$k = 3$ とした場合の分析結果を、基本図の形で示したものが図 2 である。緑の点は、流量が少ない自由流である。平均速度が高いことが、傾きの大きさとして現れている。流量の上限は、5 分間に概ね 100 台である。青の点は、流量が多い自由流である。平均速度は、緑の点よりもやや低い。

赤の点は、密度とともに、流量が下がる渋滞流である。広い範囲にデータ点が分散しているが、一つのクラスタに分類される。また、追越車線では、自由流を表す青の点と渋滞流を表す赤の点の間は、きれいに分かれているように見える。一方、走行車線では、二つの領域は連続的になっている。

流量時系列 (図 3) と速度時系列 (図 4) を見ると、流量の多い自由流に対応する青の点で、時間とともに平均速度が低下し、渋滞へと至る場合が多いことが分かる。図 5 は、各クラスタ間の移動頻度を表し

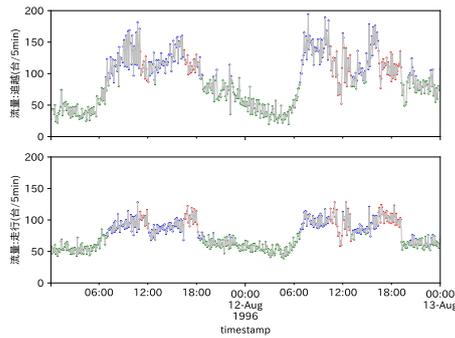


図3: 8月の流量時系列。流量だけでは、赤の渋滞発生は分からない。

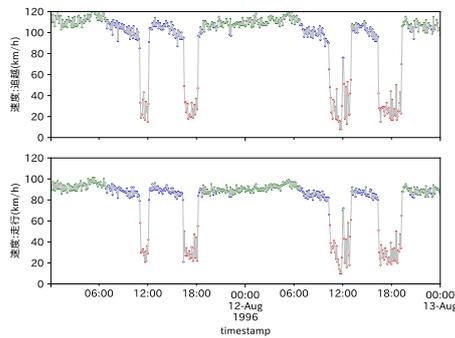


図4: 8月の速度時系列。青い点で速度が低下し、渋滞に至ることが分かる。

ている。流量の多い自由流から渋滞への転移が多いことが分かる。従って、青の点に相当する状態の特徴を分析することで、渋滞の前兆を捉えることができるであろう。

図6に、流量逆転の状況を示す。流量が多い自由流(青)と渋滞流(赤)で、追越車線の流量が走行車線の流量を超える、流量逆転が発生していることが分かる。速度低下は追越車線への移動の誘因となることから、速度と流量逆転には相関が現れる。流量逆転も渋滞発生の前兆と捉えることができる。

3.2 他の月のデータ

当該観測点では、8月には毎日のように渋滞が発生している。一方、他の月ではそれほど渋滞は発生していない。渋滞の少ない例として2月のデータの分析を行った。silhouette scoreは、 $k=2$ と $k=3$ に対して、ほぼ同じ値となっている(図7)。

$k=2$ 及び $k=3$ の場合の基本図を図8及び9に示す。渋滞流に相当するデータ点が少なく、ままと

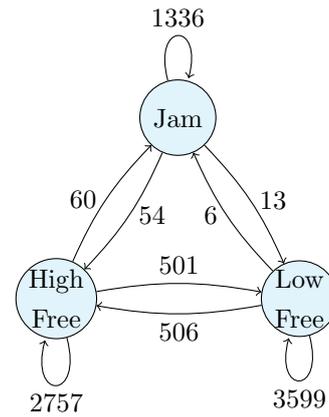


図5: クラスタ間の移動。"Jam"、"HighFree"、及び"LowFree"は、渋滞、流量の多い自由流、流量の少ない自由流をそれぞれ表す。

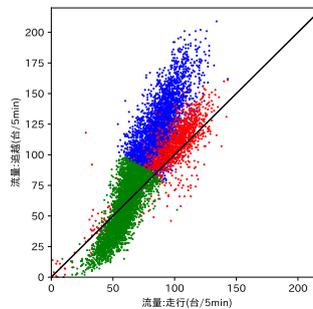


図6: 流量逆転。横軸は走行車線の流量、縦軸は追越車線の流量。

りとして認識されていない。 $k=2$ では流量によって、二つに分かれている。一方、 $k=3$ では、流量によって、三つに分類される。つまり、自由流と渋滞流を分離することが出来ていない。

4 予測

機械学習によるクラスタリングは、単にデータを分類するだけではなく、訓練されたモデルを使って、未知のデータをグループ分けすることができるということが重要である。

8月のデータに対するクラスタリングは、従来の交通流現象の理解を客観的に示すことに役立つものであった。そこで、8月のデータで訓練したクラスタリングモデルが、2月のデータを分類する様子を確認する。図10に示すように、少ないデータ点である赤の点を、渋滞として分類し、8月と同様のクラスタリングとなった。

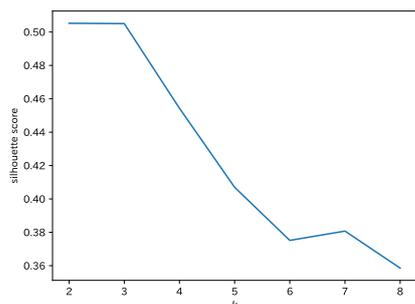


図 7: 2 月のデータの silhouette score。 $k = 2$ と $k = 3$ がほぼ同じとなっている。

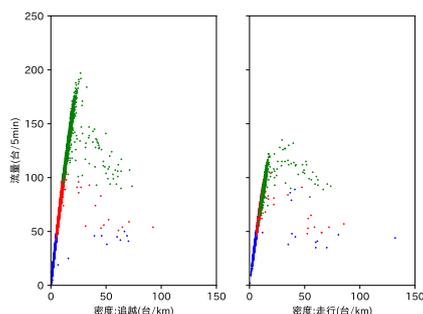


図 9: 2 月データの基本図。左は追越車線、右は走行車線。 $k = 3$ の場合。

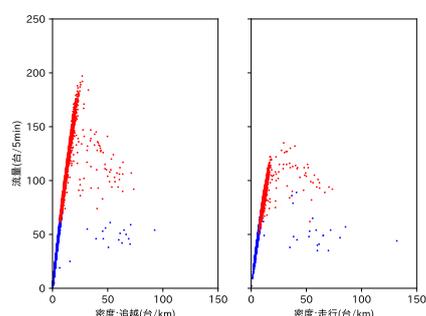


図 8: 2 月データの基本図。左は追越車線、右は走行車線。 $k = 2$ の場合。

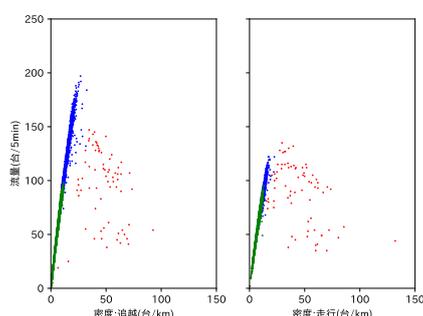


図 10: 8 月のクラスタリングモデルを 2 月データに適用した基本図。左は追越車線、右は走行車線。3 つのクラスタに分かれている。

5 まとめ

最も単純はクラスタリング手法である K 平均法を用いて、高速道路での実測データの分類を実施した。渋滞データを十分に含む 8 月のデータでは、データ点は、低流量の自由流、高流量の自由流、及び渋滞流の三つに分類された。特に、高流量の自由流では、速度低下が発生するとともに、追越車線の流量が走行車線の流量を上回る。これらは、渋滞への前駆現象の可能性がある。

しかし、総流量と平均速度に次元縮退しても同様のクラスタリング結果となる。クラスタリングにおける二車線路であることの効果の有無については、更なる分析が必要である。

一方、渋滞データの少ない 2 月のデータは、単純に流量による分類となり、渋滞データを取り出すことができなかった。しかし、渋滞が発生している 8 月のデータで訓練すると、渋滞発生が少ない 2 月のデータも、三つの状態への分類が可能となった。

また、今回の分析では時系列としての要素を含ん

でいない。時間変化の要素を取り入れたクラスタリングも、今後の課題である。

参考文献

- [1] 西成活裕, 林幹久, 「東名高速道路における交通量資料集 I」, 交通流数理研究会 (1999).
- [2] S. Tadaki, K. Nishinari, M. Kikuchi, Y. Sugiyama, and S. Yukawa, J. Phys. Soc. Japan, **71** (2002) 2326.
- [3] S. Tadaki, K. Nishinari, M. Kikuchi, Y. Sugiyama, and S. Yukawa, Physica A **315** (2002) 156.
- [4] S. Lloyd, IEEE Transactions on Information Theory, **28** (1982) 129.
- [5] <https://scikit-learn.org/stable/modules/clustering.html#clustering>