

渋滞情報データのマイニング：札幌都市圏の事例*

The data mining for the congestion data analysis: the case of the Sapporo urban area*

木村洋平**・有村幹治***

By Yohei KIMURA**・Mikiharu ARIMURA***

1. はじめに

ITSの展開により、システムの過程において記録される交通データ量は増大化している。また近年のニュー・パブリック・マネジメントの潮流により、行政サービスの指標化、可視化技術に対するニーズが高まっている。

しかし、ITSの過程において記録されるデータは、そもそも道路交通パフォーマンス計測のための統計的利用を目的としたものではない。マネジメントの価値が高い渋滞データはよく利用されるものの、その多くは、事業実施前後比較による効果の把握や、優先度明示曲線に代表される事業立案の意志決定支援等、整備量に着目した集計的利用が主となっている。

一方、インフラの賢い使い方が議論される時代においては、道路容量の増強だけではなく、ソフトを含む適切な施策の選定、その実施位置やタイミング、効果的な対象範囲の設定等も重要な評価項目となる。細密かつ大量なITSデータは、これらのニーズに応えるものであるが、その解析については、ある仮説をもって集計し、それを統計的に検証するアプローチだけではなく、統計的に検証可能なデータ群や特徴的なパターンを機械的に選定して、その背後の仮説を推測する仮説発見型アプローチも重要な試みとなる。

本研究では、札幌市内約2,200箇所に設置されたセンサーから大量に得られる渋滞情報データを用いて、時間的・空間的に特徴のある渋滞発現パターンの抽出を行った。特定箇所・時間の渋滞パターンが特定されることにより、効果的な対策の選定と対象の推定が容易になることが期待される。

2. 交通情報データ分析の動向

本研究で用いた道路交通情報通信システム (Vehicle Information and Communication System、以

*キーワード：ITS、交通マネジメント、データマイニング

**学生員、室蘭工業大学大学院工学研究科博士前期課程、
公共システム工学専攻 (室蘭市水元町27番1号、TE
L&FAX:0143-46-5289)

***正員、工博、(株)ドーコン、交通部 (札幌市厚別区厚別
中央1条5丁目4番1号、TEL:011-801-1520)

下 VICS と記す) データは財団法人日本道路交通情報センターによって提供されており、利用者はカーナビを通して渋滞や交通規制情報をリアルタイムで取得することができる。VICS データを用いた研究として、山根¹⁾は VICS データを加工し、将来時点において渋滞が悪化するか緩和するかを推定する手法を提案し、道路利用者への情報提供の質向上を図っている。安藤ら²⁾は VICS データ等から得られる旅行時間情報を蓄積し、履歴情報として旅行時間分布を用いる確率論的配車配送計画 (VRPTW-P) モデルを構築した。塚原ら³⁾はリンクの旅行時間を用い、時刻類似度を考慮した最近隣法による VICS 情報の予測手法を提案した。

以上の VICS データを利用した研究は、主に道路利用者への情報提供の質向上を目標として進められてきたように思われる。しかし、蓄積が進むデータの相互に潜むマネジメント上有用なパターンや仮説を発見する視点から検討された研究はほとんどない。そこで本研究では、蓄積されたデータを機械的に選定し、得られた結果から分析者が推測を行う仮説発見型のアプローチで、VICS データを用いて中長期の都市交通現象の傾向分析を行った。

3. データの全体傾向

(1) データの概要

本研究では、各都道府県の交通管理者及び道路管理者が収集し、(財)日本道路交通情報センターで統合される渋滞情報データを用いた。渋滞情報データは、VICS・テレビ・ラジオ等の媒体を通して外部に提供されている。渋滞情報データは、札幌市内約2,200箇所に設置されたセンサーにより、2003年4月1日から2007年3月31日までの4年間に5分間隔で観測されたものである。データは渋滞が観測された場合に限り記録され、当該期間中には5,091,077レコードの渋滞情報データが蓄積されている。なお、渋滞情報データは"日付"、"時刻"、"路線名"、"方向"、"住所"、"渋滞長"、"経度(度分秒)"、"緯度(度分秒)"の計8フィールドを持つ。以下、この渋滞情報データを生データと呼称する。

(2) データの前処理

生データのままで分析に不向きなため、年度ごとに以下に示す手順で前処理を行った。まず、セン

サー設置地点ごとに、1) 0時から23時まで1時間単位の年間渋滞観測回数、2) 4月から3月まで月単位の年間渋滞観測回数を集計した。

次に、渋滞が観測された日付ごとに、全地点の渋滞観測回数の総和を求め、日付別の渋滞観測回数を集計した。

表-1 に前処理後のデータの種類と本研究内での用途を示す。

表-1 前処理後データの種類と用途

データ名		用途
地点別	時間別年間渋滞回数	時間変動パターンの把握
	曜日別年間渋滞回数	曜日変動の把握
	月別年間渋滞回数	月変動の把握
	年間渋滞回数	分析対象とするデータの絞り込み
日付別	全地点合計渋滞回数	年間傾向の把握

(3) 渋滞の発生回数と箇所数の分布

図-1 に渋滞発生回数と発生地点数の分布を示す。図の縦軸は一日に観測された渋滞回数の合計、横軸は渋滞発生地点数である。なお、4月から11月を夏期、12月から3月を冬期とした。夏期の渋滞発生地点数は400箇所程度で、渋滞が発生する地点はほぼ決まっている。しかし、冬期の地点数は1,000を超える日もあり、降雪状況次第により渋滞発生箇所数は増加し、面的な広がりをみせるものと考察される。

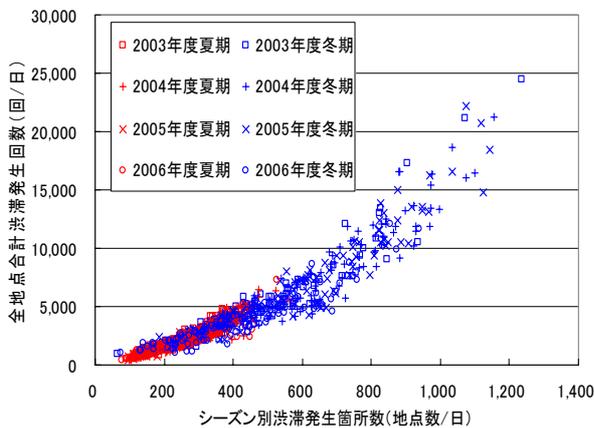


図-1 日別渋滞発生回数と日別発生箇所数の分布

次に、2003年度から2006年度までの全センサー箇所を対象として渋滞発生回数を日別に昇順に並べたグラフを図-2 に示す。4月から11月を夏期、12月から3月を冬期とした。上位200位程度の日付は冬期間に集中している。

冬期の渋滞は突発的かつ面的に拡大する事象であり、実際の移動時間がドライバーの認知上の移動時間と乖離することが多い。ある程度予測される移

動時間の変動は行動に織り込むことができるものの、このような突発性の事象は結果的に移動のサービスレベルを大きく損なわせる可能性が高い。

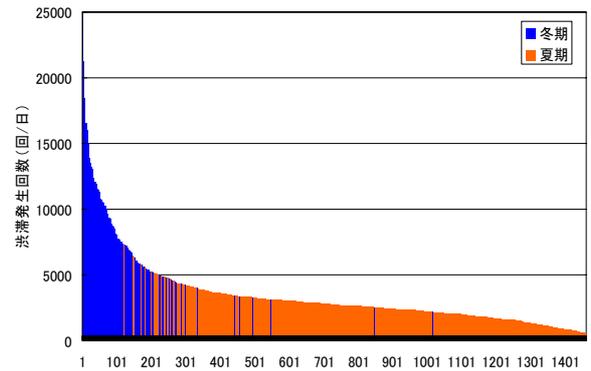


図-2 日別渋滞発生回数 (全センサー箇所：2003-2006年度)

(4) 空間的な傾向の把握

次に、渋滞の空間的な傾向の把握を試みる。ここからは2006年度に観測されたデータのみを用いる。

図-3 は2006年度の地点別年間渋滞発生回数を多い順に並べ替えたものである。全地点の年間総渋滞回数の総和に占める累積率が80%未満を赤で、80%以上を黒でプロットした。

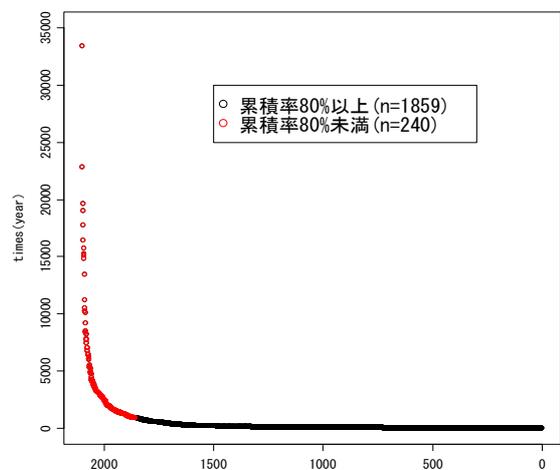


図-3 2006年度年間渋滞回数の分布

累積率が80%未満のセンサーは240箇所であり、その他のセンサーが観測した渋滞は非常に少ない。なお、この240箇所の渋滞回数について夏期と冬期で比較すると、夏期の渋滞回数が多く観測されている。本研究では、累積率80%未満のセンサー箇所を対象として渋滞の空間的な傾向の把握を試みた(図-4)。ただし、累積率80%以上の地点に関しても、同じ日に同時に渋滞が発生する等、対応策の検討が必要となる事例が含まれている可能性がある。このような、センサー相互の渋滞発生パターンの抽出等、詳細な分析については今後の課題とする。

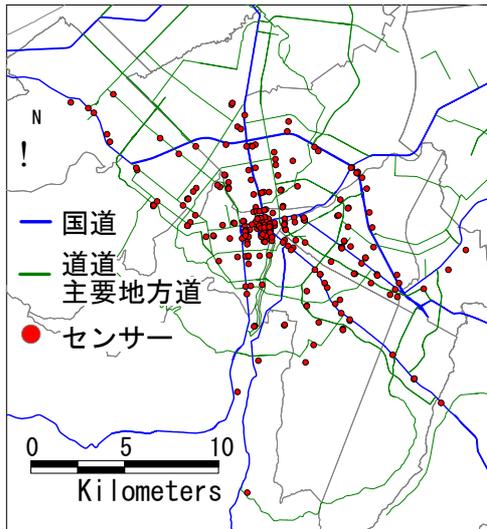


図-4 札幌市内に設置されたセンサー分布図

4. 渋滞情報データからのパターン発見

(1) 時間変動傾向の把握

渋滞発生頻度の高いセンサー箇所の時間変動を把握するため、前節で抽出した 240 地点の時間帯別渋滞回数を各地点の総渋滞回数で正規化し、グラフ化した (図-5)。図-5 で示されるように、地点別の時系列データは複数のパターンが混在しており、このままでは傾向の把握は難しい。そこで本研究では、この地点別渋滞観測時間変動を非階層型クラスタリング手法の一つである k-means 法を用いて時系列パターンを分類することを試みる。

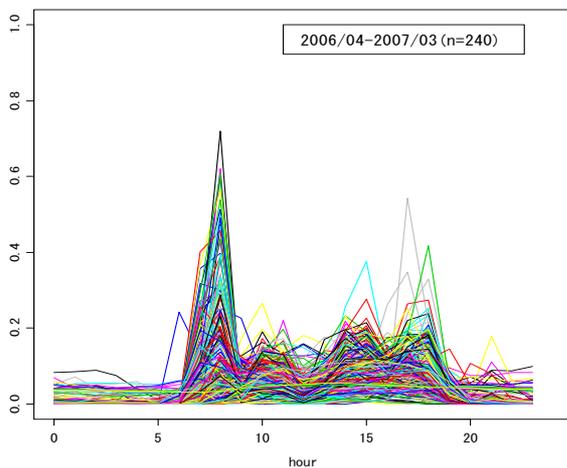


図-5 地点別渋滞観測時間変動

(2) k-means 法について

クラスタ分析の手法には、階層的な手法と非階層的な手法があり、非階層的な手法の代表例として k-means 法が挙げられる。

k-means 法では、あらかじめ指定した k 個のクラスタ中心をランダムに設定し、クラスタリングを行う。それぞれの個体を最も近い中心に割り当てる

ことで、クラスタに分類をする。分類の基準として、クラスタの中心と各個体との間のユークリッド距離の 2 乗を用いる。固体が割り当てられたら、その中からまた新たな中心を設定する。新たな中心の設定と個体の割り当てが収束するまで繰り返すことにより、適切なクラスタ中心を得て、データの分類が出来る。階層的な手法とは異なり、最終的にいくつかのクラスタを得る。

(3) k-means 法によるクラスタリング

時間帯別渋滞発生回数を入力変数として k-means 法で渋滞発生時系列パターンが似ているものをクラスタリングした。各クラスタの特徴を表-2 に、各クラスタのセンサー位置と時系列パターンを図-6 に示す。クラスタ数の設定としては、4~6 個のクラスタに分類する作業を繰り返し行い、比較検討した。その結果より、札幌市の渋滞発現パターンを解釈するために適しているクラスタ数は 5 個であった。以下にクラスタ毎の特徴を挙げる。

表-2 Cluster の特徴

Cluster	渋滞の特徴	空間的な特徴
1	・ 10 時台と 15 時台に渋滞	36 号線 石山通り付近 市内中心部 琴似
2	・ 朝と夕方へのピーク	環状線 新道 国道 12 号 豊平橋 北 7 西 10JR 高架周辺
3	・ 目立ったピークはない ・ 8 時ピークの分布と混在 ・ 全般的には緩やかに渋滞	札幌市中心部 国道 5 号手稲方面 環状 東 8 丁目
4	・ 朝に集中 ・ 渋滞回数が多い	環状線 豊平川沿い (橋)
5	・ フラット (特徴無)	—

(4) 考察

各クラスタの特徴を整理すると以下のようになる。

- ・ クラスタ 1 の特徴は、11 時、15 時、17 時をピークに渋滞が発生する時系列パターンである。業務系交通であることが推測される。札幌中心部、石山通り、新道や環状線 (伏古-雁来-札幌 JCT 間)、南郷通、12 号大谷地周辺で多く観測されている。12 時周辺で、一度渋滞回数が低下し、午後再び渋滞が増加する。市内の経済活動に依存する渋滞発生パターンである。

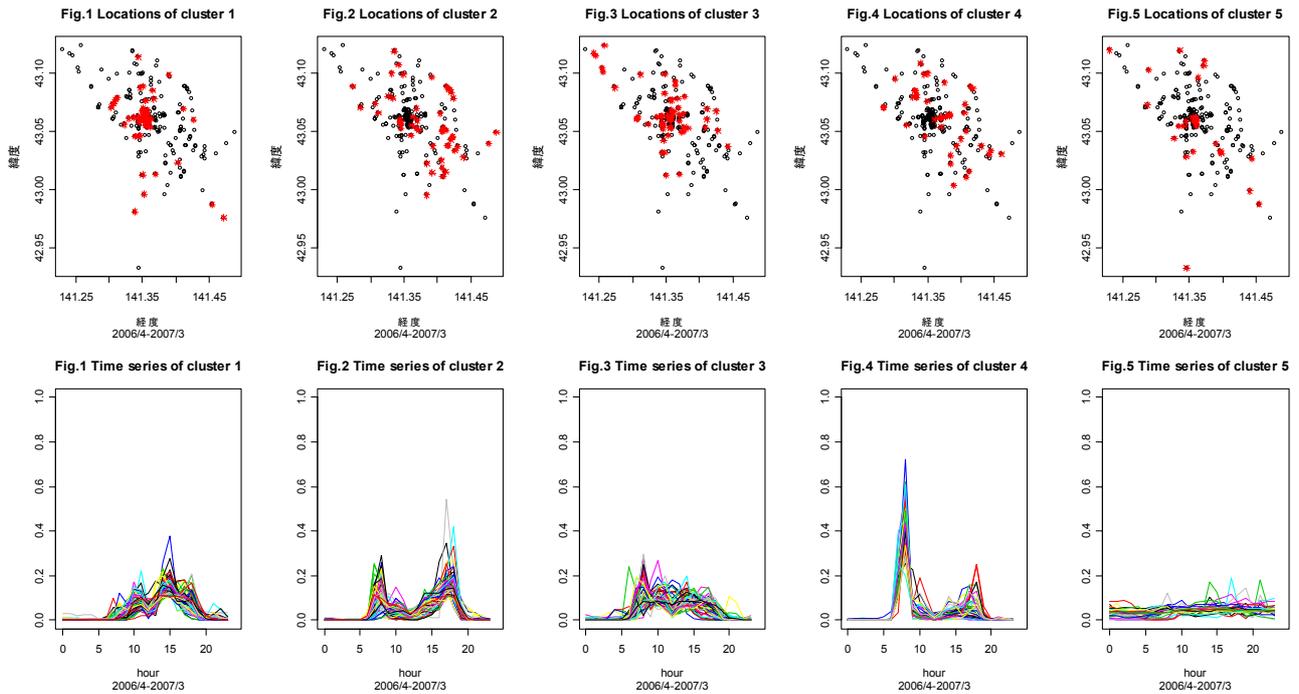


図-6 クラスター毎の空間分布と時系列パターン

- ・クラスター2の特徴は、朝一タに渋滞のピークが発生する渋滞発生パターンである。通勤交通であり、伏古-雁来-札幌 JCT 間、南郷通、12号大谷地周辺で多く観測されている。
- ・クラスター3は、8時台にピークを迎える分布と、昼間連続的な渋滞が発生する分布が混在しているものと考えられる。どちらの分布も緩やかであり、特徴的な量としては大きくはない。場所は東区役所～北IC間、内環状、石山通り、国道5号手稲方面となった。
- ・クラスター4の特徴は、特に朝に渋滞のピークが発生する時系列パターンである。やはり通勤交通であることが推測される。札幌新道や環状線(雁来-札幌 JCT 間、新川 IC 周辺)、12号札幌駅～東橋周辺、環状線(南郷-地下鉄白石駅-美園駅)周辺で多く観測される。

時系列的渋滞パターンの分類により、渋滞発生ポイントとその時間帯、時系列的なパターンの特徴が抽出された。これにより、渋滞の性格に合致した施策の立案を支援できる。

例えば、荷捌き集約化等都心系渋滞対策の実施範囲とその時間帯の設定(クラスター1)、広報等による渋滞発生回数が少ない平行ルートに関する情報の提供(クラスター2)、土日等、業務交通が通過しない時間帯を狙った道路の一般開放・イベントの実施(クラスター3)、朝ピーク時間までの除雪の完了、違法駐車を取り締まりを強化する時間帯の設定(クラスター4)等、が考えられる。

5. おわりに

本研究では、札幌市内で観測された渋滞情報データを用いてデータセットを作成し、k-means法を用いた分類を行い、空間的渋滞傾向と日変動渋滞観測傾向を得た。結果、札幌市の渋滞発生傾向と発生地点分布を数パターンに分類し、効果的な対策の選定と対象の推定が可能になった。

今後の課題としては、1)過去の渋滞発生パターンを用いた渋滞予測モデルの構築、2)気象情報等とのマッチングによる渋滞予測モデルの精緻化、3)各エリアの渋滞の特徴に合致した住民に理解しやすいアウトカム指標の構築、4)抽出データではなく、全箇所データをを用いた分析による渋滞現象の総合的な理解、等が挙げられる。

謝辞: 本研究の遂行にあたっては、北海道開発局札幌開発建設部から渋滞情報データをご提供頂いた。ここに記して感謝の意を表します。

参考文献

- 1) 山根憲一郎: VICS データを利用した渋滞傾向情報に関する検討、第 29 回土木計画学研究発表会・講演集、vol. 29、CD-ROM、2004.
- 2) 安東直紀、谷口栄一、山田忠史: ITS を活用した高度な配車配送計画について、第 33 回土木計画学研究発表会・講演集、vol. 33、CD-ROM、2006
- 3) 塚原荘一、古川 武志、原 健太、狩野 均: 時刻類似度を用いた最近隣法による広域 VICS 情報の予測手法、情報処理学会第 67 回全国大会 (2005).