

渋滞情報データからの知識発見

Knowledge Discovery from Traffic Congestion Database

室蘭工業大学 ○学生員 木村洋平 (Yohei KIMURA)
 室蘭工業大学 学生員 長谷川裕修 (Hironobu HASEGAWA)
 (株) トーコン 正 員 有村幹治 (Mikiharu ARIMURA)
 室蘭工業大学 フェロー 田村 亨 (Tohru TAMURA)

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。マネジメント的価値が高い渋滞データはよく利用されるものの、事業実施の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 箇所を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を抽出を行った。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

2. 1 既往の研究と本研究の位置づけ

現在、我が国で利用されている交通情報データとしては、道路交通情報通信システム (Vehicle Information and Communication System、以下 VICS と記す) データ、トラフィックカウンタ (以下トラカンと記す) データ、プローブデータが挙げられる。

VICS データは財団法人道路交通情報通信システム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¹⁾。

トラカンデータは道路管理者及び交通管理者が設置する車両感知器によって収集されるデータであり、交通量、

走行速度、占有率などが記録される。辻井らは空港利用者の遅延・欠航への対応行動を把握する方法として、空港に出入りする自動車の断面交通量を継続的に測定し、そこから利用者の遅れに対する余裕時間の設定行動などを推測した²⁾。

プローブデータとは、車両に搭載したプローブ車載器が走行中の車両の緯度経度情報、速度、その他の車両制御情報を情報センターへ送信するデータである。王らはプローブカーデータと VICS データを融合させ、従来の VICS 旅行時間速度よりも精度の高い旅行時間推計の推計を行った³⁾。プローブカーデータだけでなく、プローブパーソンデータの収集も進められており、今後の活用が見込まれる交通情報データである。

これまでの交通情報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、蓄積された大量データに対して仮説発見型アプローチであるデータマイニング手法の適用を試みた。

2. 2 データマイニングとは

データマイニングとは、大量のデータからルールやパターンといった知識を発見することである。データマイニングは予測だけではなく、人間に理解できるルールを提供することが大切である。

以下に示す図-1 は、Fayyad⁴⁾が示したデータマイニングの基本プロセスである。対象データの選択、データの前処理、形式変換処理、データマイニング中核処理、結果形式変換処理や表示インターフェイスなど、非常に広範囲な内容とそれらを実現する要素技術からなっている。

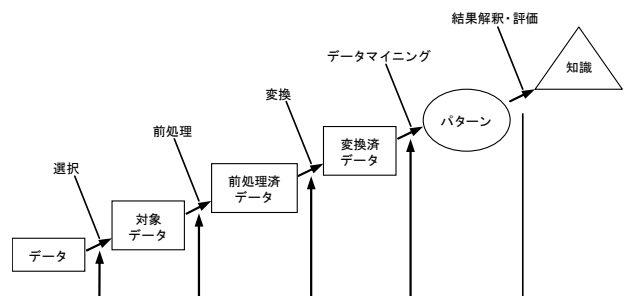


図-1 データマイニングの基本プロセス

3. 渋滞情報データからのパターン発見

本研究は渋滞情報データベースにデータマイニング手法を適用することにより、特定箇所・時間の渋滞パター

ンを発見することを試みるものである。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。

3. 1 データについて

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 箇所を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。

3. 2 データの前処理

生データのままでは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) 月曜日から日曜日まで曜日単位の年間渋滞観測回数、3) 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. 3 データの観察

交通状況は曜日、天候、突発的事象などによって影響を受け、日々変動する。データを観察し年間傾向を把握するため、4 年分の日付別全地点合計渋滞回数の推移を図-4 に示した。

図より夏期の渋滞傾向に決定的な影響を及ぼす要因は“曜日”であり、週変動を見ることで夏期渋滞傾向を掴む事が出来ると考えられる。また、夏期の日付別渋滞回数が 6,000 回を超える事例は 4 年間で 3 件のみであり、極めて少ないことがわかった(図-4、①～③)。この 3 件を突発的事象として考察する。

①2004 年 7 月 30 日(金) 6,459 回：札幌市内において花火大会が行われた。②2004 年 9 月 8 日(水) 6,387 回：札幌に台風が直撃し、観測史上最高となる最大瞬間風速(50.2m/s)を記録した。③2006 年 11 月 24 日(金) 7,326

回：勤労感謝の日を含んだ 3 連休の初日であった。

①と③の事例においては、イベントもしくは連休に絡んだ金曜日という共通点がみられた。

冬期の渋滞傾向は、年度ごとに波形が大きく変化しており、特徴的なパターンは掴めない。これは冬期の渋滞傾向に決定的な影響を及ぼす要因が“降雪”であるためだと考えられる。例えば、観測されたデータの中で最も渋滞回数の多かった⑤2004 年 1 月 15 日(木) 24,452 回と 4 番目に多く 21,170 回であった 2004 年 1 月 16 日(金)の二日間は、大雪が降った期間であった。⑥2005 年 2 月 28 日～3 月 5 日の渋滞頻度の高さも、季節外れの大雪によるものである。また、全データの中で渋滞回数が 20,000 回を越えた事例は 5 件であったが、降雪による影響が大きいと推測される。その中でも冬期の突発的事例として、④2004 年 12 月 24 日(金) 21,294 回に注目したい。同時期に大量の降雪を観測した 2005 年 12 月 28 日 13,185 回と比較しても渋滞回数が多い。②、③の事例と同様に、クリスマスと金曜日が関わって渋滞を増幅させている可能性がある。

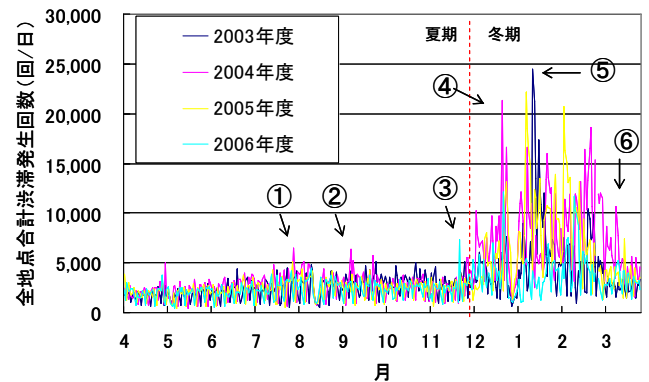


図-4 日付別全地点合計渋滞回数の推移

図-5 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。夏季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、冬季の地点数は 1,000 を超えているときもあり、降雪状況次第により渋滞発生箇所数が増加し、面的な広がりをみせることが分かる。

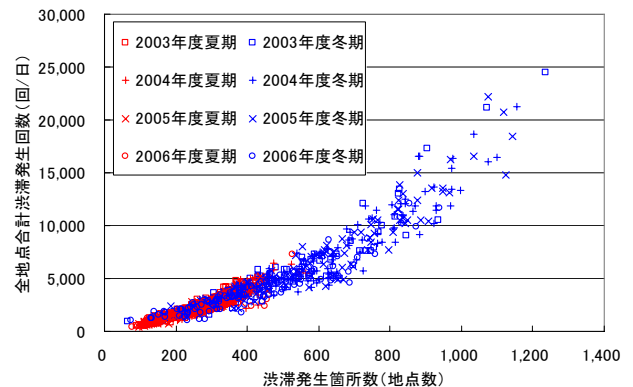


図-5 渋滞発生回数と地点数の分布

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-6 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。

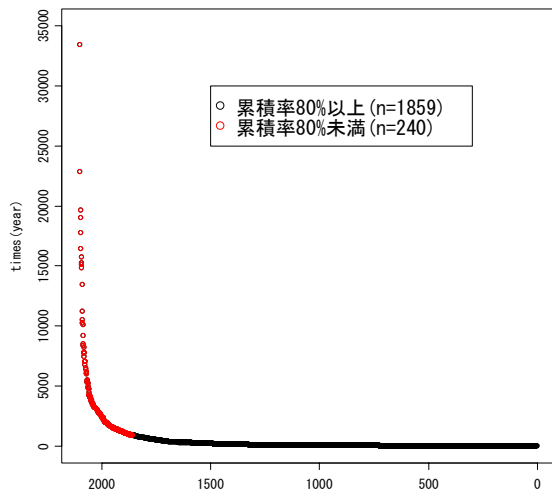


図-6 2006 年度年間渋滞回数の分布

累積率が 80%未満のセンサー数は 240 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。そこで本研究では、累積率 80%未満のセンサー箇所のみを対象として渋滞の空間的な傾向の把握を試みる (図-7)。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。この検討は今後の課題となる。

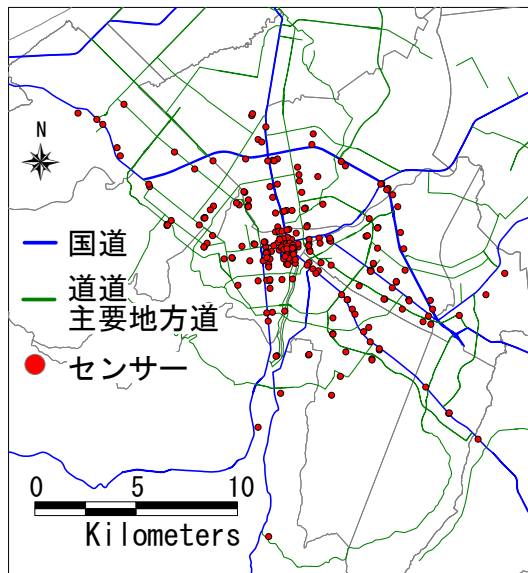


図-7 札幌市内に設置されたセンサー分布図

時間変動を把握するため、240 地点の時間帯別渋滞回数を総渋滞回数で正規化し、グラフ化した (図-8)。図-8 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クラスタリング手法の一つである k-means 法を用いてパターン化することを試みる。

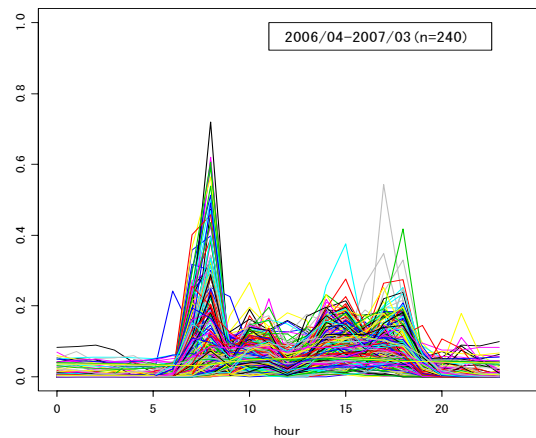


図-8 地点別渋滞観測時間変動

3. 4 k-means 法について

クラスター分析の手法には、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 個のクラスター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、クラスターに分類をする。分類の基準として、クラスター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、適切なクラスター中心を得て、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ーを得る。

3. 5 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。各クラスターの特徴を表-2 に、各クラスターのセンサー位置と時系列パターンを図-9 に示す。

表-2 Cluster の基本情報

Cluster	渋滞の特徴	渋滞の冬期比率 (%)	地点	交通形態
1	朝に集中 渋滞回数が多い	38.28	環状線 豊平川沿い (橋)	通勤 交通
2	夕方に集中	41.64	36 号線 石山通り付近	通勤 交通
3	分類なし	50.59	—	—
4	朝と夕方で同程度のピーク	42.91	札幌市中心街 新札幌方面 北広島・恵庭方面	通勤 交通
5	10 時台と 15 時台に渋滞 ランチ時に減少 個体数が最多	45.77	札幌市中心街	業務 交通

Cluster 1 の特徴は、8 時台の渋滞回数が飛び抜けて多いことである。正規化された値も 0.6~0.8 と高い数値を示している。分布地点は環状線、及び豊平川沿いに分布

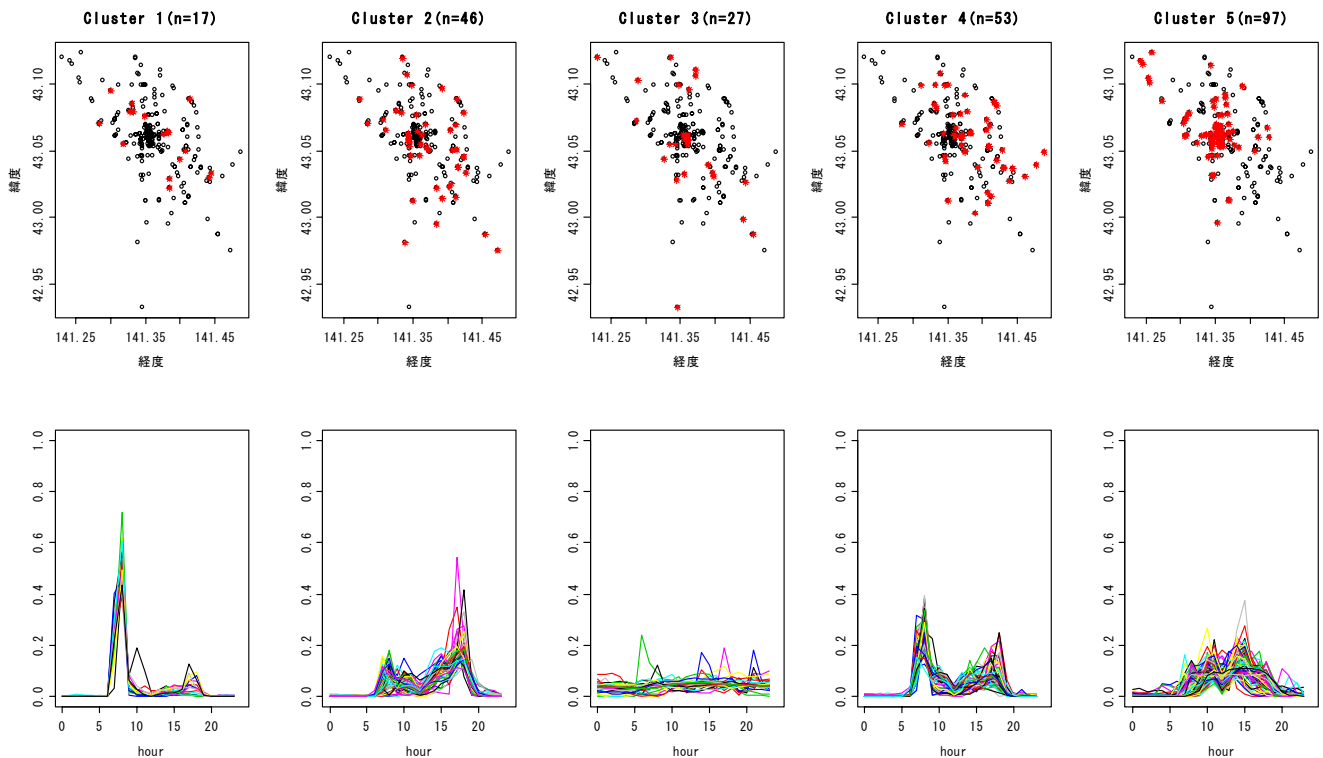


図-9 k-means 空間的な渋滞傾向と日変動傾向

しており、橋がボトルネックとなる渋滞発生パターンである。朝に渋滞が集中することから通勤交通であるものと推測される。

Cluster2 の特徴は、8 時台のピークに比べ、17 時台の渋滞のピークが大きいことである。分布地点も国道 36 号線に多く見られることから、通勤交通による渋滞であることが伺える。

Cluster 3 は特徴的な分類がなされなかったクラスターである。

Cluster 4 の特徴は、8 時台と 17 時台にピークを迎えていることである。分布地点は中心街を囲むように点在していることから、通勤交通であることが伺える。8 時台と 17 時台のピークがほぼ同程度あることから、出勤と帰宅のピーク時道路利用者が同じである可能性が高い。

Cluster 5 の特徴は、10 時台と 15 時台にピークを迎え、12 時台（ランチ時）に減少している。分布地点は札幌市中心街に集中しており、札幌市中心街の業務交通であることが伺える。また最も個体数が多いクラスターである。

3. 6 考察

k-means 法のクラスタリング結果から、効果的な施策の選定を試みる。

1) Cluster1 の 8 時台や Cluster2 の 17 時台のように特定の時間帯に集中している渋滞に対しては、羊が丘通りのようにリバーシブル・レーン規制を実施する等、道路の弾力的運用による渋滞対策が考えられる。具体例として Cluster1 では、8 時台の渋滞が通勤交通であることと札幌市の道路形状が中心街に向かって集中していることから、上りの交通渋滞が多く発生していることが指摘できる。ここで、Cluster1 に属する地点では、渋滞がピークを迎える 8 時台に上り方向の車線数を増やし、交通の

円滑化を図ることが出来る。

2) 冬期の除雪作業において、Cluster1 と Cluster4 に属する地点では、朝 8 時までに除雪作業を行い、次に、地点数の多い Cluster5 に属する地点を優先して作業を行うなど、作業優先順位の検討が出来る。

4. 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 法でクラスタリングし、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、夏期冬期で分類して、クラスタリング毎の施策を検討していくことである。

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) 辻井啓、塚井誠人、吉村充功、奥村誠：大分空港到着自動車交通の時間分布特性、第 34 回土木計画学研究発表会・講演集、vol. 34、CD-ROM、2006。
- 3) 王立暁、姜美蘭、山本俊行、森川高行：プローブカーデータと VICS データの融合による旅行時間推定に関する研究、土木計画学研究・論文集、Vol.22、pp.1011-1018、2006。
- 4) Fayyad U., Piatetsky-Shapiro G. and Smyth P.: From data mining to knowledge discovery in databases、AI magazine、Vol.17、pp.37-54、1996。