

## 科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 6 月 21 日現在

機関番号：10103  
 研究種目：若手研究(B)  
 研究期間：2011年度～2012年度  
 課題番号：23760469  
 研究課題名（和文）構造データマイニングを用いた大規模交通データセットからの知識獲得に関する研究  
 研究課題名（英文）A study on knowledge acquisition from a large-scale traffic data set by using structured data mining techniques  
 研究代表者  
 有村 幹治 (ARIMURA MIKIHARU)  
 室蘭工業大学・工学研究科・准教授  
 研究者番号：40548062

研究成果の概要（和文）：本研究は、交通計画に資する情報を得るために各種統計データや情報通信システムにより記録される交通起因の大規模データに対して構造データマイニング手法を応用し、特徴的な移動パターンを自動的に抽出する分析手法を構築するものである。本研究では、都市内の人の移動を記述した統計データであるパーソントリップデータ、及び都市の新しい移動手段であるコミュニティ・サイクル（都市型自転車シェアリングシステム）の利用履歴データに対して、各種のデータマイニング手法の適用を試みた。その結果、大規模なデータセットに内在する特徴的な移動パターンの推定、高精度の交通手段選択モデルの構築、日常・非日常時における移動パターンの推移といった交通現象全体に係る知識の抽出が可能となった。

研究成果の概要（英文）：The aims of this study are to apply the structure data mining techniques for large-scale traffic data recorded by an information and communication system to obtain information to contribute to a transportation planning and to build analysis technique to extract a characteristic movement pattern automatically. In this study, we tried to apply various data mining techniques for the person trip data which were the statistics data described the movement of the person in the city and log data of community cycle (bicycle sharing system in an urban area) which was the new transportation mode in the city. As a result, the extraction of the knowledge to modify the estimate of a characteristic movement pattern to be in large-scale data set, the construction of a highly precise transportation choice model, the whole traffic phenomenon such as the change in the traffic pattern at the daily traffic and the non-daily traffic was enabled.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
交付決定額	1,500,000	450,000	1,950,000

研究分野：土木工学

科研費の分科・細目：土木計画学・交通工学

キーワード：データマイニング、大規模交通データセット、ソフトコンピューティング

## 1. 研究開始当初の背景

本研究は、VICSやETC、またプローブカー、プローブパーソン等に代表される、情報通信技術を用いて得られる大量のデータから、特徴的な移動パターンや渋滞発生パター

ンを抽出し、交通計画に資する情報を自動抽出する方法論を構築するものである。

近年の情報通信技術の進展に伴い、交通分野においても、沿道や自動車や携帯端末に搭載される各種センサー群により得られるデ

ータ量が飛躍的に増大している。このようなデータは、主として交通や車体の制御に用いるためのものであり、交通計画へのフィードバックを目的としたものではない。旅行速度を連続的に計測できるプローブカーデータはよく利用されているものの、事業実施の前後比較による事業効果の推定や、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、ハード整備に着目した集計的分析が主に実施されている。またその一方で、現在の交通施策は、包括的な施策のパッケージングが重要視される。道路交通容量の拡大だけではなく、適切な施策の選定、施策の実施位置や時間帯、範囲の設定等も、計画立案上重要な検討項目となっている。

細密かつ大量な IT データはこれら計画上の要求に応えることができるが、その分析方法は大きく二つのアプローチに分けることができる。一つは、既存の、ある仮説をもってデータを集計・モデル化し、それを統計的に検証する集計的アプローチである。そしてもう一つは、データ群の背後に潜む安定したデータの関係性を機械的に抽出し、問題の所在と原因、対策の推定を行う、仮説発見型の分析アプローチである。この二つのアプローチは相互に補完する関係にあるが、後者は機械学習やデータマイニング技術に密接に関係しており、その学際的性格から研究蓄積は前者と比較してあまり進んでいない状況にある。

申請者は現在までに、都市部を対象とした都市交通施策の立案支援を目的として、仮説発見型分析アプローチにより、道路ネットワーク上の渋滞発生パターンをデータマイニング手法を応用することで、時間的・空間的特徴から抽出・分類する方法を構築している。この研究により、札幌都市圏において過去5年間約 2200 個所のセンサーにより蓄積された各地点毎の渋滞発生回数を示す時系列データから、夏期及び冬期の主要渋滞個所の渋滞発生箇所と渋滞発生の時系列パターンが特定化された。しかし、この研究で用いられた一般的な分類手法である k-means 法は、①多様なクラスタ形状を扱えない②実数値データ、整数値データ、カテゴリデータ等、異なる属性を同時に扱えない③最適なクラスタ数の発見やクラスタ境界が非線形な問題の扱いが困難④初期値にクラスタリングの結果が依存する、などの問題があった。またこの研究は道路沿いに固定されたトラフィックカウンターから得られた渋滞データの分類を試みたものであるが、近年の自動車や人の連続位置情報を収集するプローブカーやプローブパーソンとの普及を顧みると、移動体に関する大規模データセットの分析方法についてが課題として残っていた。

## 2. 研究の目的

渋滞状況を表す時系列データや、プローブカーやプローブパーソンといった連続位置情報(図-1参照)は、道路容量や走行車線数といった走行環境、被験者属性といった離散的な属性を持つ他のデータソースと組み合わせることで、より詳細な施策立案に適応させることができる。また、複数渋滞箇所の時系列データや移動軌跡データは、データ間の関係に、なんらかの内的、もしくは外的な構造を持つことが想定される(図-2参照)。データ間の構造を明示的に取り込めるデータマイニング手法の応用により、大規模交通データセットの交通計画的利用価値の向上が期待できる。



図-1 プローブカーデータの走行軌跡と速度分布(旭川市周辺)

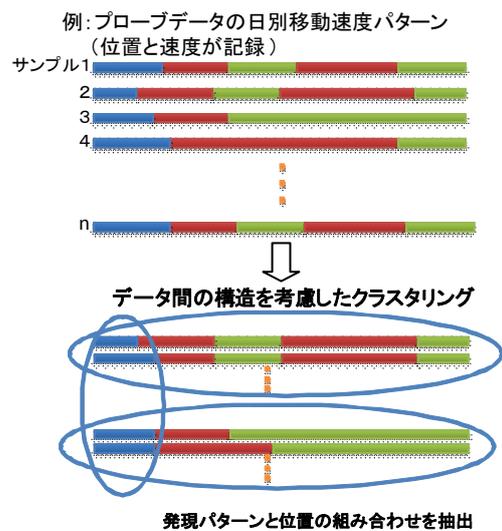


図-2 プローブデータの内部構造

本研究では、上記の問題に対応するため、データ間の構造を効率的に扱えるデータマイニング手法の適用を念頭に、地域内を流動する自動車や人間の特徴的な移動パターンや、都市内部の渋滞発生パターンを自動的に抽出・分類し、地域や都市交通施策の立案を支援する分析手法を構築する。

## 3. 研究の方法

データマイニングに係る研究分野では、カーネル k-means 法やサポートベクターマシンの応用等、多くの手法が提案されている。本研究は、交通工学・交通計画の進展を目的と

しているが、情報通信技術、及び情報科学の知識を要する学際領域の研究となるため、まず各分野における技術開発状況をレビューした上で、既存の交通データセットに適応可能な構造データマイニング手法の選定を行った。具体には、人工知能や機械学習等の知的情報処理分野において注目されるアンサンブル学習手法の適用を試みた。アンサンブル学習は複数の単純なモデルを構築し、それらを用いた分類結果を統合して最終的な分類結果を得る手法であり、高精度だが計算コストが高いニューラルネットワークやサポートベクターマシン等の機械学習手法に劣らない精度を発揮することが知られている。

使用データとしては、札幌市パーソントリップデータ、帯広市都市計画基礎調査及び人口統計といった既存統計データ、また現在札幌市において展開しているコミュニティ・サイクル(図-3)の年間利用データを用いた。入手した大規模データセットにデータマイニング手法を適用するために、データのクリーニングを施した。また 利用するデータセットの時間的・空間的な精度と、交通計画的に対応可能な時間・空間スケールが異なる可能性があるため、都心の交通計画に必要な誤差の範囲について検討した。



図-3 コミュニティサイクル (札幌)

#### 4. 研究成果

既存統計データに対するデータマイニングの試行として、平成 18 年に実施された道央都市圏パーソントリップ調査データに対して、アンサンブル学習の一手法であるランダムフォレスト (RF) を用いて交通機関選択モデルを構築し、多項ロジットモデル (MNL)、サポートベクターマシン (SVM) との比較を行った。その結果、以下の 5 点が明らかとなり、アンサンブル学習による交通機関選択モデルの有用性が示された。

- (a) 訓練データ・テストデータともに機械学習である RF モデル・SVM モデルの的中率が MNL モデルを上回った
- (b) RF モデルのみテストデータの的中率が訓練データの的中率を上回り、高い汎化能力を示した
- (c) RF モデル・SVM モデルともに全ての交

通手段の再現率で MNL モデルを上回った

- (d) 各手段に分類されたサンプル数を見ると、RF モデル・SVM モデルは実際の手段別シェアに近い値を示したが、MNL モデルは大きく乖離した値となった
- (e) RF モデルは分類結果に対する各変数の重要度を解釈しやすい形で出力できた

なお本研究では、上記ランダム・フォレストモデルによる交通機関選択モデルの有用性を確かめたうえで、ランダム・フォレストモデルのパラメータ探索に遺伝的アルゴリズムを用いるハイブリッド型モデルを構築し、その適用を試みている。

以上は既存統計データセットに対してデータマイニングの適用を試みた事例であるが、より大規模な交通系データセットへの適用を試みた事例として、札幌市のコミュニティサイクルデータの事例を挙げる。

コミュニティサイクルは、低炭素な交通モードである自転車を用いること、また都市内のモビリティを高めることから、環境的側面や交通まちづくりの視点から注目されるが、その事業実施には、料金設定、ポート配置計画とそれに連動した地権者との交渉等、事業全体と都市の状況を俯瞰した丁寧なエリアマネジメントが必要であり、そのための知見の蓄積が求められている。本研究では、ICカードにより得られる平成 21 年度のコミュニティサイクル利用記録データセットを用いて、集計分析及びデータマイニングアプローチを用いた分析を行った。その結果、特定の利用パターンを持つ複数の利用者層の抽出及びピーク時間に特定ポートへ集中する需要の特定が可能となった。これにより、運営負荷を緩和するための効率的な運営技術の開発や利用者の協力のもと負荷分散を行う運営手法の立案に有用な情報をデータマイニングにより抽出できることが確認できた。

コミュニティサイクルの計画立案支援のためにデータマイニングにより分析すべき課題としては以下が挙げられる。

- 1) 利用者層別の価格弾力性の確認
- 2) 定性的な利用者アンケートと利用記録データを結びつけたコミュニティサイクル利用実態の把握
- 3) 時間帯別ポート間 OD による自転車再配置計画の検討
- 4) 登録者の年間を通した都市内の空間認知の変化の追跡
- 5) 都心部の他の交通モードの IC データ等との統合による都市内モビリティ全体の質の評価

このように、大規模交通データセットに対するデータマイニング手法の適用により、量的な交通計画から、よりきめの細かい質的計画を立案できることが確認できた。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計3件)

(1)長谷川裕修, 内藤利幸, 有村幹治, 田村亨, アンサンブル学習による交通機関選択モデルの構築とその評価, 土木学会論文集 D3 (土木計画学), Vol. 68, No. 5 (土木計画学研究・論文集第29巻), Vol. 6, No. 5, pp773-780, 2012 (査読有)

(2)有村幹治, 猪股亮平, 田村亨, 帯広都市圏を対象とした将来居住分布の推定, 土木学会論文集 D3 (土木計画学), Vol. 68, No. 5(土木計画学研究・論文集第29巻), pp375-382, 2012 (査読有)

(3)丸山翔大, 松田真宜, 有村幹治, 交通まちづくりのためのコミュニティサイクル利用記録の活用, 土木計画学研究・講演集 (CD-ROM), No. 45, 2012 (査読無)

[学会発表] (計4件)

① Hironobu HASEGAWA, Mikiharu ARIMURA, Tohru TAMURA, Hybrid Model of Random Forests and Genetic Algorithms for Commute Mode Choice Analysis, The 10th Eastern Asia Society for Transportation Studies, Conference 2013, 2013年09月09日~2013年09月12日, Taipei (Taiwan)

② 松田真宜・平川貴志・有村幹治, 利用履歴データに基づくサイクルシェアリングの戦略的な運営手法の構築, 第47回土木計画学研究発表会(春大会), 2013年06月01日~2013年06月02日, 広島工業大学

③ 平川貴志・松田真宜・山本郁淳・有村幹治・仲田田, クラウド型プローブパーソンシステムの交通課題への適用とその応用可能性, 第47回土木計画学研究発表会(春大会), 2013年06月01日~2013年06月02日, 広島工業大学

④ 丸山翔大・松田真宜・有村幹治, 交通まちづくりのためのコミュニティサイクル利用記録の活用, 第45回土木計画学研究発表会春大会, 2012年06月02日~2012年06月03日, 京都大学

## 6. 研究組織

### (1) 研究代表者

有村 幹治 (ARIMURA MIKIHARU)

室蘭工業大学・工学研究科・准教授

研究者番号：4 0 5 4 8 0 6 2

### (2) 研究協力者

長谷川 裕修 (HASEGAWA HIRONOBU)

秋田工業高等専門学校・環境都市工学科・助教

研究者番号：0 0 5 3 3 3 7 4

松田 真宜 (MATSUDA MASANORI)

(株) ドーコン 交通事業本部交通部  
主任技師

研究者番号：7 0 6 2 9 8 5 7