

SVM AdaBoostを援用した 交通機関識別手法の開発

今泉 孝章¹・羽藤 英二²

¹学生会員 東京大学 工学系研究科 都市工学専攻 (〒113-8656 東京都文京区本郷7-3-1 工学部14号館221号室)

E-mail: imaizumi@bin.t.u-tokyo.ac.jp

²正会員 東京大学 工学系研究科 社会基盤学 専攻 (〒160-0004 東京都文京区本郷7-3-1 工学部1号館341)

E-mail: hato@bin.t.u-tokyo.ac.jp

本研究では、加速度センサーを用いた交通機関識別手法の開発を行う。判別器にはサポートベクターマシン、AdaBoostアルゴリズムを用いた。データクリーニングによる正規化を行った上で、大局的な識別と微視的な識別を同時に行うことで、プローブパーソン調査をもとにした交通機関自動識別を実現する。

Key Words : transportation mode distinction, accelerometer, SVM, AdaBoost

1. はじめに

従前の交通計画では大規模な都市間の移動に着目していた。都市をゾーンに区切った上で四段階推定法を用いて、OD間のトリップ数、経路、移動手段の推計に主眼が置かれてきた。一方、高齢化社会が進行に伴い、都市交通計画において歩行環境や街路空間の関係といった都市交通計画に関わる分析ニーズが増大している。

従前の分析に用いられるデータは、アンケート調査が主であるため、ゾーン内の微視的な移動活動パターンを分析することは難しい。また、手記入のアンケート調査では、1)時間が区切りのよい時間にまとめられ不正確である、2)ショートトリップが省略される傾向にある、3)トリップ内での挙動が把握困難である、4)調査そのものが手間である、などの問題が存在する。(Hato¹⁾2010)

そこで本研究では、個人の移動軌跡、時間、加速度など多くの多様な情報を取得できるプローブパーソン調査(以下PP調査)のデータを用いて加速度データなどを用いた移動手段の自動判別を試みる。移動手段を自動判別に関する研究は大村²⁾の既往研究が挙げられる。本研究では、1) サンプルング方法を変化させる、2)誤差、移動文脈を考慮しデータを正規化する、3)複数判別器を組み合わせる、といった判別手法の提案を行い、現実の加速度データを用いて比較検証を行う。

2. データ概要

表1 取得データ概要

データ種類	取得データ
トリップ	出発到着時刻、移動手段、目的、付帯者
ロケーション	測位日時、緯度経度、気圧、3軸加速度データ、歩数、運動強度

表2 取得データ数

データ種類	生データ数	紐づけデータ数
トリップデータ	984	965
ロケーションデータ	341,713	229,544

(1) 調査概要

本研究におけるデータは、2011年1月12日から2011年2月11日の期間に、主に高齢者を対象として、BCALs、GPS機能付き携帯電話、行動日誌により取得された。データは各ODごとのトリップデータと、トリップ内のロケーションデータがある。表1にデータ概要を示す。

(2) データクリーニング

本研究では、ロケーションデータの移動手段判別を、加速度データ(3軸方向)、運動強度、単位歩数(手動で計算)から行ったが、判別精度検証に使える真値データは取得されていない。そこで、トリップデータの移動手段をロケーションデータの真の移動手段として、トリップデータとロケーションデータを取得時刻から紐づけを行った。分析に用いたデータ数を表2に示す。

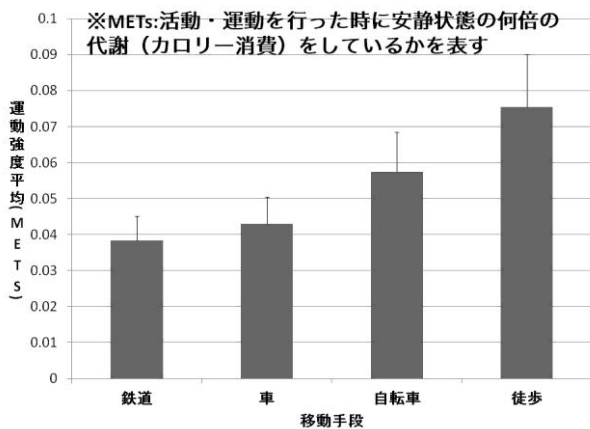


図1 移動手段別運動強度平均、標準偏差

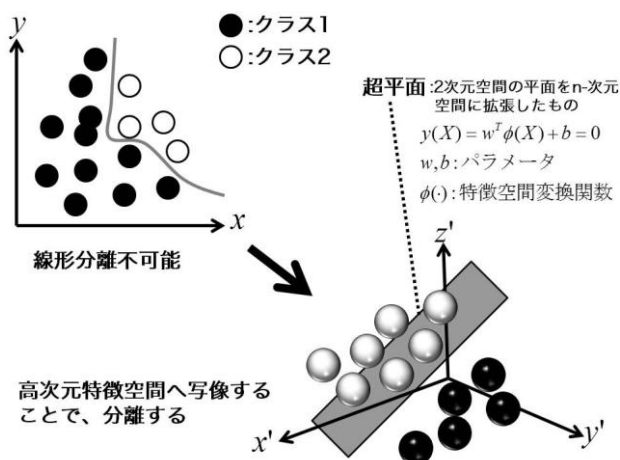


図2 SVMによる高次元特徴空間への写像

3. 移動手段別データ特徴

(1) 移動手段別のデータの平均、分散

本研究では徒歩、自転車、車、鉄道の4値判別を行った。移動手段判別を行うに先立ち、移動手段ごとに特徴量にどのような違いがあるのか確認を行う。判別に用いたデータのの一つである、”運動強度”(単位:METs)について例を示す。運動強度とは、取得された加速度データから算出されるデータである。運動強度の移動手段ごとの平均を図1に示す。徒歩、自転車、車、鉄道の順に運動強度の平均、分散ともに小さくなっていることがわかる。私的なモビリティであるほど、移動時に自らの体を使うためと考えられる。こうした移動手段ごとのデータの特徴に違いがあることに着目して移動手段判別を行う。

4. SVM、AdaBoostアルゴリズム

(1) SVM(サポートベクターマシン)

本研究での移動手段判別には、教師付き学習を用いる

識別手法で、最も認識性能が優れた手法のひとつである、SVM(Support Vector Machine,以下SVM)を用いた。SVMの特徴を表3に示す。

表3 SVMの特徴

- 1 マージン最大化によって高い汎化能力を持った識別平面を生成する
- 2 線形分離不可能なデータに対し、高次元特徴空間へ写像

図2に示すようにサンプルが線形分離不可能なとき、SVMは高次元に写像することでサンプルを分離する。但し、一般には高次元特徴空間に写像することで計算は煩雑になる傾向がある。但しSVMではカーネル関数を用いて内積計算を行うことで簡略化が可能となる。

(2) AdaBoostアルゴリズム

AdaBoostアルゴリズム(以下AdaBoost)は教師付き学習を行うためのメタアルゴリズムであり、一般の機械学習アルゴリズムを精度よく実行するためのフレームワークといえる。もともになる機械学習アルゴリズムに何を用いるかは制限されていないものの、SVMなどのすでに高性能な学習器を用いるよりも、古典的なものに対して利用する方が有利とされている。主な特徴を表4に示す。図3にAdaBoostによる判別手順のフローを示す。

表4 AdaBoostの特徴

- 1 判別結果をもとに判別器に重み付けを行う
- 2 判別結果をもとデータにも重み付けを行い、誤判別されたデータが重点的に学習されるようにする
- 3 重み付けされた複数の判別器の多数決で最終的に判別結果を出力する

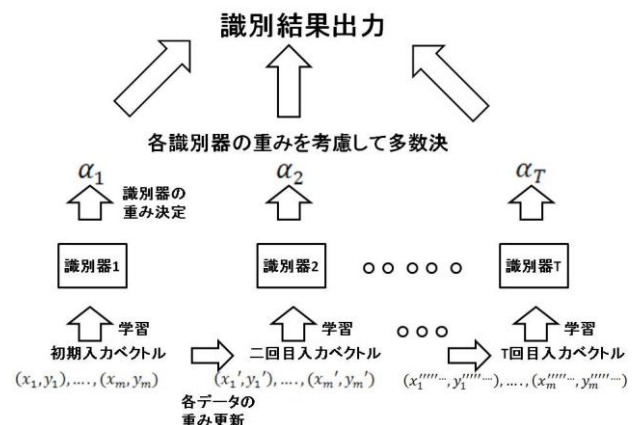


図3 AdaBoostアルゴリズムの判別フロー

表5 移動手段別データ、教師データ数

移動手段	全データ数	ランダム 教師データ 数	移動手段別 教師データ 数
徒歩	164,591	865	300
自転車	3,185	2	300
車	60,596	318	300
鉄道	354	15	300

表6 抽出方法別判別結果

移動手段	ランダム抽出	手段ごと抽出
徒歩	95.6%	85.2%
自転車	0.0%	89.9%
車	93.6%	79.4%
鉄道	0.0%	96.3%
全体	93.6%	83.8%

(表中は適合率(%))

表7 移動平均別判別結果

移動手段	生データ	n=10	n=50	n=75	n=100
徒歩	78.1	85.2	88.3	87.3	86.6
自転車	75.9	89.9	95.1	94.1	94.0
車	67.3	79.4	87.0	86.0	85.3
鉄道	92.4	96.3	100.0	100.0	100.0
全体	75.3	83.8	88.0	87.3	86.6

(※表中は適合率(%), nは平均個数を表す)

SVM、AdaBoostの二つを用いて移動手段判別を行った。計算には統計解析フリーソフト「R」の「e1071」と「ada」を用いた。また判別のための説明変数は、単位歩数(手動で計算)、運動強度、加速度(3軸最大、最小、平均、合成加速度)である。判別結果の検証は式(1)で表される適合率を用いる。

$$\text{適合率(\%)} = \frac{\text{正しく判別されたサンプル数}}{\text{全サンプル数}} \quad (1)$$

5. 実データを用いた移動手段判別

(1) 教師データサンプリング方法

移動手段判別を行う際、教師データとして全データの中から1200サンプルを抽出することを考える。その際、抽出方法として全学習データの中からランダムに1200個抽出した場合と移動手段別に300個ずつ、計1200個抽出した場合の2パターンの比較を行う。表5に全データの中で各移動手段毎のデータ数、ランダムに1200個抽出したときの移動手段別教師データ数、移動手段別に300個ずつ抽出したときの教師データ数を示す。SVMのみを用いてこれらのデータで交通手段判別を行った(表4)。ランダム抽出方法で自転車、鉄道の適合率が0%となっていることがわかる。教師データ中に十分な数の自転車と鉄道のデータが含まれていないことが理由と考えられる。

本研究では以後、移動手段ごとに300個ずつ、1200個抽出したものを教師データとして利用する。

(2) データの正規化

実空間移動を想定した場合、停止→移動→停止...といった行動パターンがどんな移動手段でも生起していると考えられる。徒歩と鉄道の加速度が完全に一致することがある。また、データ取得条件によっては無視できない誤差が含まれている可能性も考えられよう。

本研究では、連続した移動データの前後平均したものを使った。(今回平均したデータ数は10,50,75,100個である。)例えば10個平均する場合、ある時系列データの前の10個のデータと後の10個のデータを、計21個の平均データを $n=\pm 10$ として、 $n=\pm 50$ 、 $n=\pm 75$ 、 $n=\pm 100$ についてそれぞれ判別計算を行い最適な移動平均個数を検証した。結果を表7に示す。データを平均化している場合の方が、生データをそのまま利用した場合よりも適合率がよいことが明らかになった。平均する個数としては前後50個のときの適合率が最もよい。データを平均化することで、誤差、前後移動文脈を考慮し移動手段判別を行うことができたと考えられる。本研究では以後、前後50個平均データを利用することにする。

(3) SVMとAdaBoostの組み合わせによる移動手段判別

前節までSVM単体で移動手段判別を行ってきた。本節ではSVMをAdaBoostと組み合わせる移動手段判別を行った結果を示す。AdaBoostでの徒歩、自転車、車、鉄道の4値判別は2値判別の繰り返しにより行い、その後SVMによる4値判別を行うことを試みる。具体的な手順を表8に示す。

表8 SVM×AdaBoost判別手順

1	全てのテストデータで鉄道とその他の2値判別を行う(AdaBoost)
2	手順1でその他と判別されたデータで車とその他の2値判別を行う(AdaBoost)
3	手順2でその他と判別されたデータで自転車とその他の2値判別を行う(AdaBoost)
4	手順3でその他と判別されたデータで徒歩とその他の2値判別を行う(AdaBoost)
5	手順4でその他と判別されたデータで鉄道、車、自転車、徒歩の4値判別を行う(SVM)

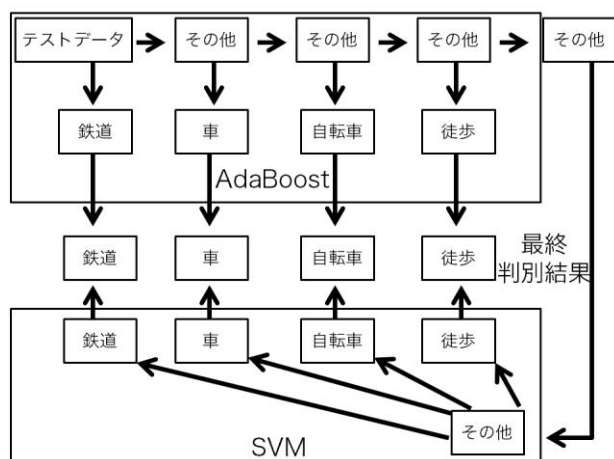


図4: SVM×AdaBoostアルゴリズムの判別フロー

表8 SVM単体、SVM×AdaBoost判別結果

移動手段	SVM	SVM×AdaBoost
徒歩	88.3	90.8
自転車	95.1	95.1
車	87.0	92.6
鉄道	100.0	100.0
全体	88.0	91.4

(表中は適合率(%))

この手順のフローを図4に示す。表8によりSVM単体の判別結果より、SVMとAdaBoostを組み合わせると判別を行ったときの方が適合率がよい。特に徒歩、車での適合率が増加しており、全ての移動手段と全体での適合率がともに90%を超えている。組み合わせ判別により、単体の判別器で誤判別されたデータが、異なる判別器によって正しく再判別されたと考えられる。

6. まとめと今後の課題

SVM、AdaBoostなどの判別器を用いて、BCALsから得られる加速度データから移動手段判別を行った。教師データサンプリングの際に、移動手段ごとに抽出を行うことで教師データ中に各判別クラスごとに十分なデータ数が得られること、データを正規化することで誤差、移動文

脈を考慮でき、生データを判別器に入力するよりも高い判別精度を得られることがわかった。また、一つの判別器だけでなく、複数の判別器を組み合わせることで単一の判別の場合よりも判別精度が向上することを示すことができた。

今後の課題としては二つ考えている。一つ目は経路データによる判別結果の修正である。例えば、鉄道軌道がないところでも、加速度データをもとにしているので、鉄道と判別されてしまう可能性がある。しかし、このような誤判別は緯度経度データが取得できているので、あらかじめ鉄道軌道が分かっているならば、鉄道軌道から離れているデータに関しては鉄道と判別する機会をなくすことで修正できる。

二つ目はLOSデータを用いた判別結果の修正である。現実的な選択確率から移動手段を修正する。例えばあるトリップで、徒歩で移動する距離ではないほど遠い場合でも、徒歩と判別される可能性がある。このような誤判別はLOSデータから移動手段選択確率を算出し、それを判別器の説明変数の一つとして入力することで修正できる。加速度データのみからの判別ではその特徴量から判別を行うので現実的な意思決定を考慮できていない。そこで経路データ、LOSデータなどの現実のデータから補正を行うことが可能であると考えている。

参考文献

- 1) Eiji Hato: Development of behavioral context addressable loggers in the shell for travel-activity analysis, Transportation Research Part C, 18, pp. 55-67, 2010
- 2) 大村朋之、位置加速度情報を用いた交通機関判別と移動負荷モデルの開発、修士課程学位論文、東京大学、2011
- 3) Lie Guo, Ping-Shu Ge, Ming-Heng Zhang, Lin-Hui Li, Yi-Bing Zhao: Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine, Transportation Research Part C, 39, pp. 4272-4286, 2012
- 3) 大村朋之、羽藤英二：高齢者行動に着目した都市内のうろつきの観測識別可能性、土木計画学研究・講演集(CD-ROM) Vol44, pp ROMBUNNO.79

DEVELOPMENT OF METHODOLOGY TO DISCRIMINATE THE MEANS OF TRANSPORTATION USING SUPPORT VECTOR MACHINE AND ADABOOST ALGORITHM

Takaaki Imaizumi and Eiji HATO

In this paper, I try to develop the methodology to discriminate the means of transportation by the accelerometer. After normalizing the data, I aim to realize that the transportation mode is recognized automatically based on the Probe Person survey by macro and micro distinction.