

携帯GPSデータを用いたスパースモデリングによるリアルタイムな人の行動予測

古澤京, 須藤明人, 関本義秀, 足立龍太郎

背景

GPSセンサや加速度センサなど、各種センサを搭載した携帯電話の普及により、携帯電話ユーザーの行動履歴を把握することが可能になった

↓
行動履歴に注目したビジネス・サービスの増加

↓
リアルタイムな目的地予測の需要が高まる

GPSデータはユーザーの行動を移動時については5分間隔で取得可能であり、目的地予測において注目されている

サービス例

Uber

Uber Technologies Inc.が運営する自動車配車アプリは利用者の目的地を自動で予測し74%の正解率



課題と目的

GPSデータから分析可能な情報は様々

- GPSログの取得間隔と移動距離を用いた交通モード推定
- 滞在点をクラスタリングすることによる家・職場・よく行く場所の推定
- 人口密度

GPSデータが持つ情報

- 日付
- 現在時刻
- 緯度・経度

これらを組み合わせることができる特徴量の数は膨大で、行動予測に最適な特徴量を抽出する必要がある。しかし、目的地予測にSVMを用いたJingjing Wang et al.(2012)や、条件付確率を用いたアンサンブル法を使用したTrinh Minh Tri Do et al.(2012)は最適解を総当たりで探しており、効率的に抽出することができていない

目的

- 所有者の行動をGPSデータから分析可能な特徴量を用いてリアルタイムに予測する
→現在地を滞在・移動のどちらの行動を取るのか
- 行動予測に必要な特徴量を効率的に抽出する

使用データ

株式会社ゼンリンデータコム提供の2012年度における日本全国約150万人の携帯電話ユーザーのGPSデータについて、Apichon et al.(2013)ユーザー毎に行動、主要滞在点の推定が行われたデータを活用した。

・トリップデータ

GPSデータからユーザーの行動をトリップ毎に記録したデータ

フォーマット

ユーザーID・日付・現在時刻・交通手段・移動距離・行動開始時刻・行動終了時刻・座標

・主要滞在点データ

ユーザーが一年間で訪れた主要な滞在点を総滞在時間が長い順にリストアップしたデータ。家・職場・よく行く場所が推定されている

フォーマット

ユーザーID・ランク(総滞在時間が長い順)・滞在点の種類・座標

手法

所有者の行動を過去のGPSデータをもとにLassoを用いた機械学習により予測する

曜日 w 、滞在点 l で n 回観測されたデータ $\{(y_i^{wl}, x_i^{wl}); i = 1, 2, \dots, n\}$ について、以下の式を満たす回帰係数ベクトル β_{wl}^{Lasso} を求める。

$$\hat{\beta}_{wl}^{Lasso} = \operatorname{argmin}_{\beta_{wl}} \left\{ \frac{1}{2n} \|y_{wl} - X_{wl}\beta_{wl}\|^2 + \lambda \|\beta_{wl}\| \right\}$$

x^{wl} : 時刻 t において曜日 w 、滞在点 l で観測された特徴量

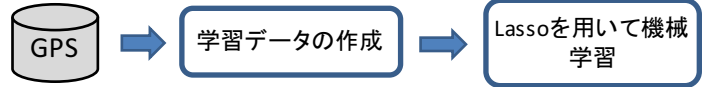
y^{wl} : 時刻 $t + \Delta t$ において曜日 w 、滞在点 l で観測された行動

$$y = \begin{cases} +1 (\text{移動}) \\ -1 (\text{滞在}) \end{cases}$$

λ : チューニングパラメータ

$$X^{wl} = (x_0^{wl}, x_1^{wl}, \dots, x_n^{wl})^T, y = (y_1^{wl}, \dots, y_n^{wl})^T$$

λ の値を調整することで $\hat{\beta}_{wl}^{Lasso}$ の成分のいくつかを0にすることができる。このとき、 $\hat{\beta}_{wl}^{Lasso}$ はスパース性を持つといい、 λ の値を大きくするほどより多くの成分が0になる。これによりLassoは回帰係数ベクトルの推定と同時に特徴量抽出が可能となる。

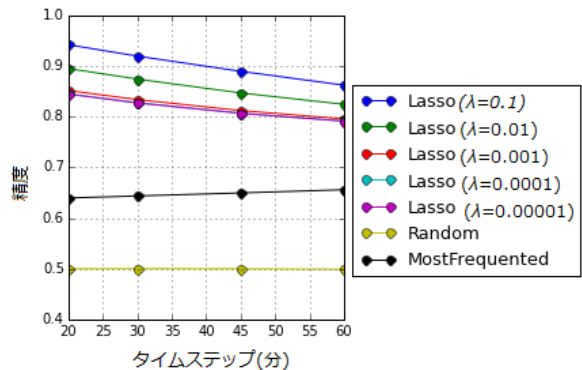


結果

過去のGPSデータ1年のうち、3/4を学習データ、1/4をテストデータに用いてクロスバリデーションによる評価を行い、比較手法よりも予測精度に優れていることを確認した。

比較手法

- ランダムな推定法...行動をランダムに推定
- MostFrequented法...所有者の各時間における最も滞在頻度の高い場所を分析し、その場所と同じなら滞在、異なるなら移動と推定



各手法における予測精度