

トピックモデルの拡張モデルを利用した スマホ型回遊データの判別分析

木崎 凜太郎¹・円山 琢也²

¹学生会員 熊本大学大学院自然科学教育部土木建築学専攻（〒860-8555 熊本県熊本市中央区黒髪 2-39-1）

E-mail:158t4837@st.kumamoto-u.ac.jp

²正会員 熊本大学准教授 くまもと水循環・減災研究教育センター

（〒860-8555 熊本県熊本市中央区黒髪 2-39-1）

E-mail:takumaru@kumamoto-u.ac.jp

近年、Wi-Fiパケットセンサによる行動調査データや携帯電話会社が取得する位置情報ビッグデータといった膨大な移動軌跡や位置情報を得られるデータが増加している。しかし、これらのデータはプライバシーの保護の観点から性別や年齢、交通手段といった属性情報を付与されていない場合が多い。そこで、移動軌跡や位置情報といったデータを基に属性情報を推測する手法の開発が望まれる。本研究では、トピックモデルを拡張したsupervised Latent Dirichlet Allocation (sLDA)を利用し、2013年熊本都心部スマホ型回遊調査の軌跡データの分析を行う。具体的には、教師あり学習の枠組みで軌跡データを学習データとテストデータに分け、学習データからモデルを推定し、テストデータから属性情報の予測を試みた。その結果、一部の属性情報は予測の可能性が示されたが、属性情報によって予測精度に違いがあることが示された。

Key Words : travel behavior analysis, GPS-based travel survey, topic modeling, sLDA

1. はじめに

(1) 研究の背景と目的

交通計画・都市計画・マーケティングにおいては人々がどのような経路で回遊しているか、ある地点でどのような人がどの程度の時間滞在しているか、といった回遊行動の実態を把握する調査が行われる。従来、回遊行動調査にはアンケート調査やプローブパーソン調査（以下、PP調査）、スマホアプリ型調査が行われてきた。しかし近年、Wi-Fiパケットセンサによる行動調査や携帯電話会社などが取得する位置情報ビッグデータ、交通系ビッグデータといったデータを用いた研究も多く見られる。前者のデータは、個人属性・トリップ属性といった属性情報を収集できるという利点があるが、調査対象者に負担がかかる点や長期間の調査を行うのが難しいといった点で課題がある。一方で、後者のデータは調査のコストが小さく、調査対象者の負担も小さい。また、長期間に渡りデータを取得することも可能である。しかし、これらのデータはプライバシー保護の観点から属性情報を付与されていない場合が多い。

このようなデータから属性情報を推定する方法として、例えばトリップ目的を推定する研究としては、GIS

を利用して立ち寄った場所を特定することでトリップ目的を推定したり、PP調査における被験者の手入力による部分的な回答と、それに基づく逐次的な機械学習による目的推定を組み合わせたトリップ目的推定システムなどがある^{1,2}。また、携帯電話の通信履歴と移動速度を特徴量とした交通手段と移動経路の推定もみられる³。

本研究では、属性情報の推定にトピックモデルを用いる。トピックモデルは、教師データなしで文書中から単語をセグメンテーションすることによって、潜在変数（トピック）の推定が可能である。文書データの解析手法として開発されながらも、数多くの分野で応用され、古屋⁴や川野⁵の研究のように観光周遊行動や回遊行動データへの適用もある。また、トピックモデルは拡張モデルも多く提案されており、それらの特性を生かした適用蓄積が望まれる。さらに先行研究で十分な検討がされていない、交通手段、来街目的の項目別の属性推定や属性予測の精度評価も求められる。

以上の背景のもと、本研究では、2013年秋に熊本市都心部で実施されたスマホ型回遊調査のデータを対象に、トピックモデルの拡張モデルである supervised Latent Dirichlet Allocation (以下、sLDA)を用いた属性情報の推定と推定結果の評価を行うことを目的とする。

(2) 本研究の構成

ここで、本研究の構成を述べる。ここまで、1章では、主に本研究の背景と目的について述べた。2章では、既往研究のレビューを行う。Wi-Fiパケットセンサや携帯電話会社などが取得する位置情報ビッグデータ等を用いた研究事例、トピックモデルを用いた推定について既往研究を整理する。3章では、トピックモデルの拡張モデルであるsLDAについて述べる。4章では、本研究で用いる2013年秋に熊本市都心部で実施されたスマホ型回遊調査の概要を述べる。5章では、sLDAを用いた回遊行動データからの属性推定と推定結果の評価を行う。最後に、6章で本研究から得られた成果と今後の課題をまとめる。

2. 既往研究のレビューと本研究の位置づけ

1章で述べた通り、本章では、Wi-Fiパケットセンサや携帯電話会社などが取得する位置情報ビッグデータ、交通系ビッグデータを用いた研究事例、トピックモデルを用いた推定についての既往研究を整理する。

(1) 位置情報ビッグデータを用いた分析

ここでは、Wi-Fiパケットセンサや携帯電話会社が取得する位置情報ビッグデータ、交通系ビッグデータの研究についてまとめる。

Wi-Fiパケットセンサで得られたデータを用いた研究として、中西ら⁶⁾は沖縄・本部半島の観光地13か所にWi-Fiパケットセンサを設置し、施設の来場者数や滞在時間が推定可能であることを示すとともに、観光地間のOD表と地点間の所要時間分布、トリップチェーンの集計を行っている。壇辻ら⁷⁾は奈良県長谷寺参道にWi-Fiパケットセンサを設置し調査を行った。この調査では観測ポイントによって鉄道利用者と車利用者を判別し、滞在時間のヒストグラムを描き比較している。携帯電話会社が取得するビッグデータに関する研究としては、渋川ら⁸⁾が携帯電話の基地局データを用い、立地適正化計画の評価指標としての活用可能性について検討している。また、北川ら⁹⁾はNTTドコモが提供しているモバイル空間統計を高度化させた人口流動統計を用いて、都道府県を超える長距離スケールのトリップに焦点を当て、新幹線を利用したか否かを判定する手法を提案している。細江ら¹⁰⁾は香川県高松市で利用可能な交通系ICカード「IruCa」の利用履歴データを用い、非負値テンソル因子分解によって、交通系ICカード利用者がどの時間帯に、どのような人が、どの駅から、どの駅へ移動していたのかという、4つの異なる属性情報を同時に考慮した行動パターンの把握が可能であることを示唆している。

以上のように、Wi-Fiパケットセンサや携帯電話ビッ

グデータ、交通系ビッグデータを用いた研究においては、様々な手法で属性情報が判別・判定されている。よって本研究も、このような研究の基礎的なものの一つに位置づけられる。

(2) トピックモデルを用いた推定分析

トピックモデルは国内外の土木計画・都市計画分野において、幅広い適用・応用事例が見られるようになってきている手法である。神谷ら¹¹⁾は、モバイル空間統計データに、文書をメッシュ、滞在者の居住区を単語として、トピックモデルを適用し、地域別人口特性の解釈を行っている。位置情報に適用した例としては、古屋ら⁴⁾がトピックモデルや、トピックモデルの拡張モデルの一つであるHierarchical Pachinko Allocation Model (hPAM) を利用し、訪日外国人旅行者の訪問地の組み合わせを分析している。また川野ら³⁾は古屋らが非常に広域な観光周遊行動を対象としたのに対し、市街地レベルの回遊行動データへトピックモデルを適用し、回遊行動パターンの抽出を行っている。そしてトピックモデルを用いた推定に関する研究では、山本¹²⁾らはTwitter上に投稿された実生活ツイートに対し適切な局面(現象や行動から考えられる場面)を付与するために、トピックモデルを用いた推定を行っている。具体的には第1段階でLDAを用いて大量のツイートからトピックを抽出し、第2段階で局面を特徴付けるトピックを抽出するため局面ラベルが付与された少量のツイートを用いてトピックと局面の対応関係を構築する。入力されたツイートから抽出した単語のトピック中の生起確率と、トピックと局面の関連度を用いてスコアを算出し、スコアが閾値を超えた局面を未知のツイートに対して推定している。また堂前¹⁴⁾らもTwitter上の投稿を対象にして半教師ありトピックモデルにより選択された地域特徴語を用いたユーザーの生活に関わる地域の推定を行っている。

しかし、1章でも述べた通り、トピックモデルの応用はこのように幅広い一方で、行動データへの適用例は少ない。また、属性情報の付与していないデータを想定し、トピックモデルを用いて推定した研究は筆者の知る限り存在しない。よって、本研究はトピックモデルの応用という点でも萌芽的な研究といえる。

3. トピックモデル

ここで、本研究で用いるsLDAについて、参考文献¹⁵⁾に基づいて解説する。なお、トピックモデルの最も基礎的なモデルであるLDAの解説は先行研究¹⁶⁾を参考にしたい。

(1) Supervised LDA

a) 概要

sLDAは、LDAの拡張モデルの一種である。LDAが文書のみを生成するのに対して、sLDAは、文書集合中の各文書に観測可能な補助情報が関連付けられる場合、文書と同時に補助情報を生成することが可能なモデルである。文書に補助情報が関連付けられているものとして、例えば、インターネット上の記事とそれにブックマーク付けた人の数やそのカテゴリを示すタグ、または映画とそれに評価として与えられる星の数等が考えられる。このように補助情報は連続値や離散値、順序付きの離散値、非負に制限された整数のいずれもとらうる。sLDAは補助情報を推定するのに線形の回帰モデルを利用しているが、labeled-LDA¹⁷⁾などの他の関数を用いた様々な拡張モデルが提案されている。本研究は、トピックモデルを利用した移動軌跡データからの個人属性推定の基礎的研究と位置づけられるものである。よって、ここでは最も基礎的なモデルであるsLDAを用いることにする。なお、上記の拡張モデルの適用は今後の課題としたい。

b) Supervised LDAによる文書の生成過程

図-1にLDAとsLDAのグラフィカルモデルを示す。図-1のとおり、単語そのものの生成過程は、LDAとsLDAで共通している。そのため、ここでは補助情報の生成過程について解説を加える。

補助情報 y_d は、パラメータ η と σ^2 の正規分布に従って生成されるため、以下のように示される。

$$y_d \sim N(\eta^T \bar{z}_d, \sigma^2) \quad (1)$$

ここで、 \bar{z}_d は経験トピック分布と呼ばれ、以下のよう示される。

$$\bar{z}_d = \frac{1}{N_d} \sum_{i=1}^{N_d} z_{d,i} \quad (2)$$

$z_{d,i}$ は、トピック分布 θ_d に従う多項分布から生成される。文書 d 中の i 番目の単語がどのトピックに属するかを示す離散型変数であった。よって、 \bar{z}_d はトピック k 毎に $z_{d,i}$ を数え上げ、文書 d の単語の総数 N_d で除したものである。言い換えると、 \bar{z}_d はトピック分布 θ_d とは対照的に、実際に文書 d 中に割り振られたトピックの頻度分布である。

この \bar{z}_d を用いて、以下の線形回帰モデルにより、補助情報 y_d を推定する。

$$y_d = \eta_k \bar{z}_{d,k} \quad (3)$$

先述のとおり、単語の生成過程はLDAのそれと共通するものの、パラメータ推定時には、単語と補助情報の分布を同時に満足するように $z_{d,i}$ を学習するため、LDAより適切なモデルが推定できるとされている。

また、パラメータの推定においては、本研究では、

変分ベイズ法を用いる。回遊行動データをトピックモデルに適用させる方法については先行研究¹⁶⁾を参考にされたい。

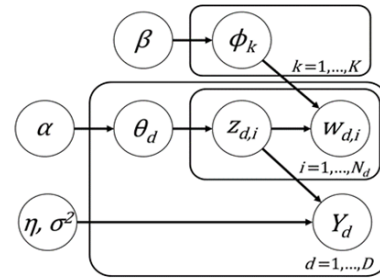


図-1 Supervised LDAのグラフィカルモデル

4. データセット

(1) くまもとまち歩き調査

調査の詳細について表-1に示す。調査日は2013/11/23(土)、24(日)、30(土)、12/1(日)、7(土)、8(日)の6日間を設定し、熊本市中心市街地における回遊行動の記録を行った。基礎情報として、性別、年齢、居住地等を入力してもらい、調査後にアンケートも実施した。位置情報の取得にはスマートフォンアプリの「スマくま」を用い、スマホを所持していない高齢者を中心とした方々にはタブレット端末(Nexus7)の貸出を行い調査に参加していただいた。調査の詳細やアンケート結果の基礎分析については、野原ら¹⁸⁾を参照されたい。

表-1 くまもとまち歩き調査概要

調査日	2013年11月～12月の土・日曜日の6日間
調査時間	午前10時～午後7時
調査エリア	熊本都心部(上通り、下通り、新市街)
調査対象	高校生(16歳)以上
調査主体	熊本県、熊本市、熊本大学
調査方法	スマホアプリ「スマくま」

(2) 分析に利用するサンプルの抽出

本調査に参加し、ポートでのアンケートに回答したのは6日間で延べ1,086サンプルであった。このうち、分析対象をAndroid端末により取得されたサンプル(705サンプル)に限定する。Android端末のGPSが10秒毎に位置情報を測位するのに対し、iOS端末のGPSは10mの移動毎に位置情報を測位する。本研究で用いるトピックモデルでは、位置情報の共起性と観測回数に基づいて行動パターンの抽出を行う。

この場合、回遊行動トピック中で滞在・滞留地点を

表現するには、端末が静止していても位置情報が測位され続けることが必要とされる。このため、本研究では、分析対象を Android 端末のみに限定した。

また、本研究では、1メッシュあたりの観測される測位点に閾値を設け、その閾値を下回るデータは移動状態であるとみなして除去することとした。詳細については先行研究¹⁰⁾を参考にされたい。以上の処理により、最終的な分析対象のサンプルは 680 サンプルとなった。

5. Supervised LDAを用いた属性推定の試み

本章では、トピックモデルより抽出されるトピックを空間的な特徴量として、教師あり学習の枠組みで、回遊行動データからの属性推定を試みる。詳細な分析手法については先行研究¹⁰⁾を参考にされたい。また本研究では予測精度の評価を行う。

予測精度の評価には、ROC 曲線と AUC を用いる。これらは情報処理や自然言語処理の分野で、分類器の性能評価や、医学分野では検査の結果を評価する手法として用いられる。予測精度は、表-2 に示す混同行列を基に算出される。詳細な用語の説明については藤田ら¹⁹⁾を参考にされたい。ここから、偽陽性率 (False positive rate) と真陽性率 (True positive rate) を以下の式から算出し、ROC 曲線を求める。

$$\text{False positive rate} = TP / (TP + FN) \quad (4)$$

$$\text{True positive rate} = FP / (FP + TN) \quad (5)$$

そして AUC (Area Under the Curve) は、図-2 に示すように ROC 曲線下の面積を表す。予測が正確に的中、つまり真陽性率が 1 かつ偽陽性率が 0 になる場合は AUC は最良値の 1 となり、予測が完全にランダムで行われている (全く予測していない) 場合、ROC が対角線となり AUC は 0.5 となる。トピック数は $k=30$ として推定を行った。

(1) 年齢・職業

まず、年齢・職業について sLDA を用いて推定を試みた。図-3 には年齢・職業に関する ROC 曲線を示す。またそれぞれの AUC の値は (学生 : 0.72, 30 歳以上 : 0.63, 40 歳以上 : 0.67, 50 歳以上 : 0.67, 65 歳以上 : 0.74) となった。これより学生と 65 歳以上の予測精度において他より良い値が得られたことがわかる。ここで、学生と 65 歳以上の予測結果について詳細を示す。

図-4 に学生を補助情報としたときのトピック別のパラメータ推定結果を示す。パラメータの大きさは各トピックの属性を表現している。ここでは、補助情報の設定として、学生を 1、その他を 0 としているため、パラメータの推定値が 1 に近いほど学生に関連するトピックとなり、

0 に近づくほど、その他に関するトピックとなる。図-4 はトピック別の棒でパラメータの推定値を表現しており、棒の中心が推定値、棒の長さは標準誤差、太さは t 値を示している。図-4 中の上部のトピックほど 1 に近く、下部のトピックほど 0 に近くなっている。つまり、トピック 19 やトピック 22 は学生に関連するトピックであることを意味する。 t 値を見ると、トピック 12 : $t=3.575$, トピック 14 : $t=4.137$, トピック 19 : $t=4.237$, トピック 22 : $t=2.939$ 。

図-5 には、図-4 で推定したモデルを用いて、職業のラベル (学生 : 1, その他 : 0) を予測し、その予測値を密度分布で表現したものである。両ラベルが重なっている区間はあるが閾値の設定によっては判別が可能と考えられる。

また図-6 に示されるトピック 19 のメッシュ分布確率は、熊本パルコ付近や上通商店街の一部といった学生が訪れやすい場所を示している。よって、学生がよく訪れる店舗の関係から学生という属性情報を予測できる可能性が考えられる。

表-2 混同行列

正解 \ 予測	A	Not A
A	TP (True Positive)	FP (False Positive)
Not A	FN (False Negative)	TN (True Negative)

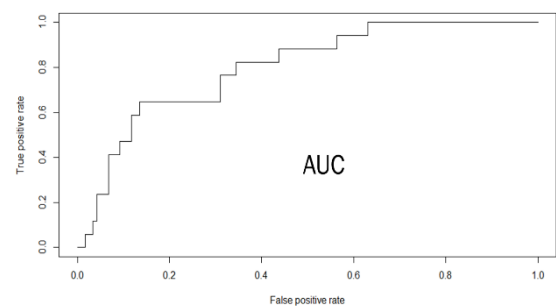


図-2 ROC 曲線と AUC の例

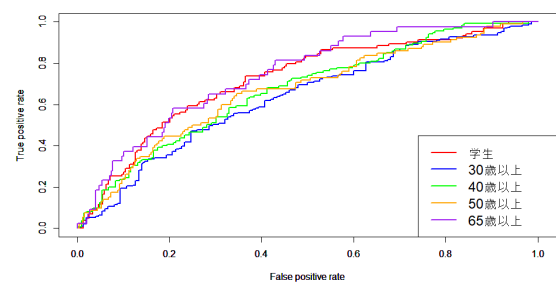


図-3 年齢・職業に関する ROC 曲線

図-7, 図-8には65歳以上を1, 65歳未満を0とした場合のトピック別パラメータ推定結果と予測値の密度分布を示す。パラメータ推定結果のt値を見ると, トピック2, 20等で1に最も近い値で1%有意となっている(トピック2: $t=5.857$, トピック20: $t=3.741$)。ここで図-9のトピック2のメッシュ分布確率を示す。確率が高いメッシュは, 交通センターや県民百貨店の周辺でみられる。これは学生の結果と同様に, 高齢者が利用しやすい施設と考えられる。

これらの結果から, 対象とするエリアに特定の年齢層に人気がある, もしくは利用されやすい, 共に利用されやすい施設がある場合には, 一部の属性データと位置情報データのみから年齢層を予測できる可能性が考えられる。

(2) 交通手段

次に交通手段についての結果を述べる。交通手段別のROC曲線を図-10に示す。各手段のAUCの値は(自動車: 0.64, 二輪車: 0.56, バス: 0.55, 熊本電鉄: 0.75, 市電: 0.55, 徒歩: 0.51)となった。

ここで図-11, 図-12に自動車とそれ以外の交通手段を1, それ以外の交通手段を0とした場合のトピック別パラメータ推定結果と予測の密度分布を示す。パラメータの推定結果のt値はトピック2, 17, 24で1%水準での有意差が確認できた(トピック2: $t=3.223$, トピック17: $t=3.382$, トピック24: $t=2.769$)。ここでトピック17とトピック24のメッシュ分布確率を図-13, 図-14に示す。トピック17で最も分布確率が高いメッシュ付近にはコインパーキングが位置していることから, コインパーキングに駐車し, 回遊を始め

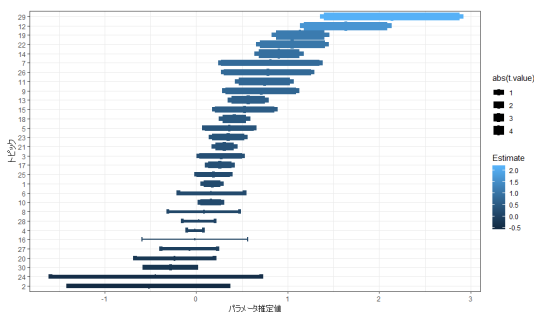


図-4 パラメータ推定結果 (学生)

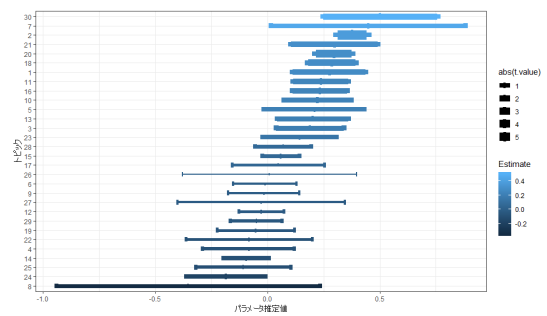


図-7 パラメータ推定結果 (年齢: 65歳以上)

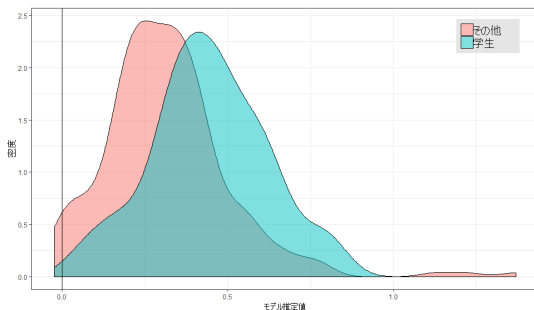


図-5 予測値の密度分布 (学生)

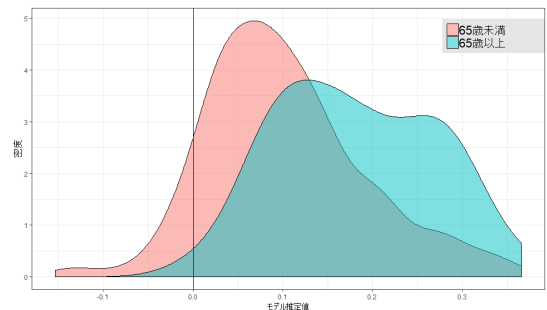


図-8 予測値の密度分布 (年齢: 65歳以上)

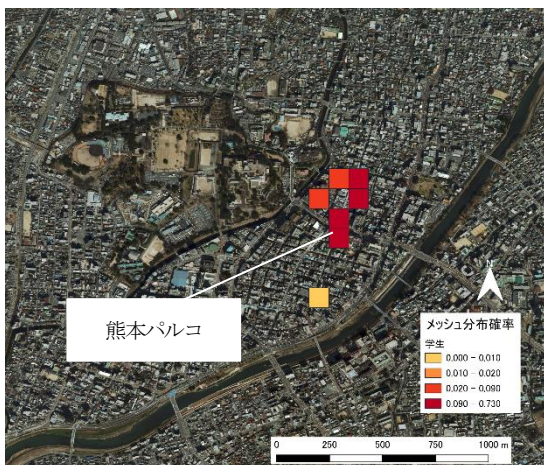


図-6 トピック(19)のメッシュ分布確率

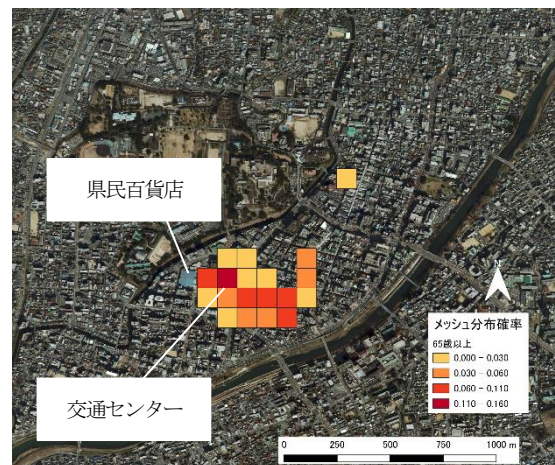


図-9 トピック(2)のメッシュ分布確率

たとえられる。またトピック24は鶴屋百貨店とその駐車場付近で分布確率が高い。これは、トピック17と同様に駐車後の回遊と考えられる。駐車場は商店街から少し離れた場所に位置し、他の交通手段と判別されやすいため、予測の精度が高くなったと考えられる。

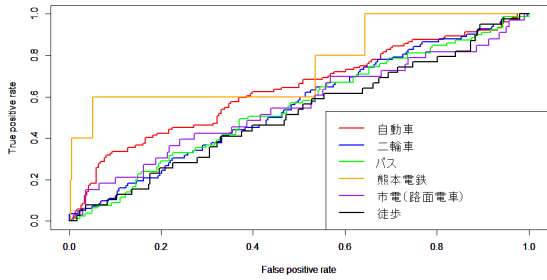


図-10 来街交通手段に関するROC曲線

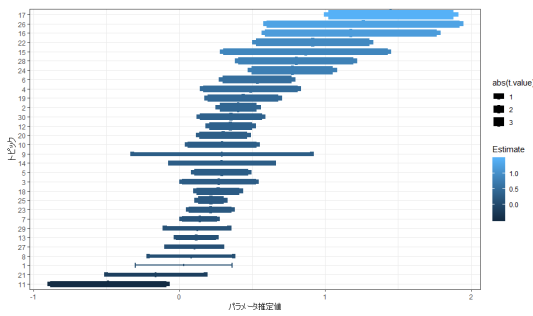


図-11 パラメータ推定結果 (交通手段: 自動車)

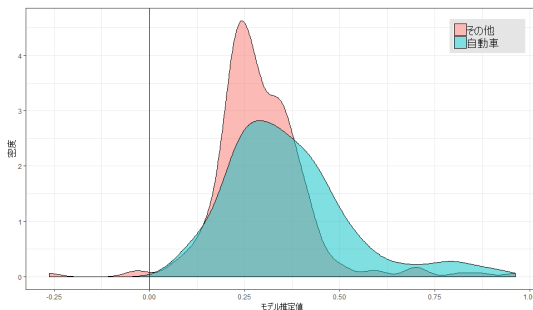


図-12 予測値の密度分布 (交通手段: 自動車)

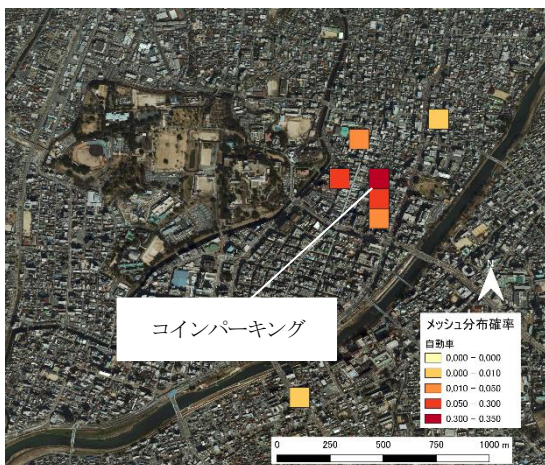


図-13 トピック(17)のメッシュ分布確率

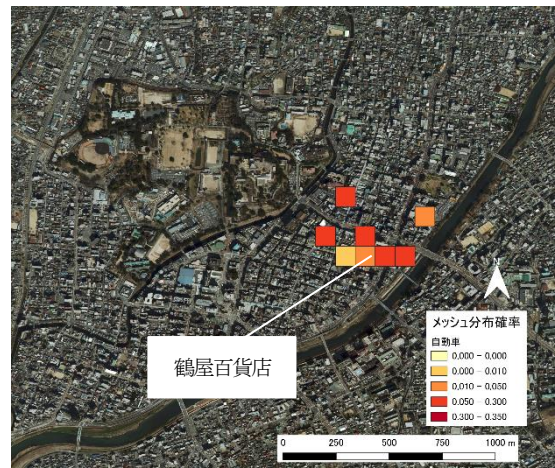


図-14 トピック(24)のメッシュ分布確率

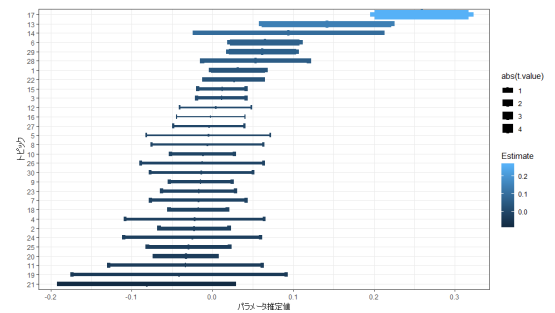


図-15 パラメータ推定結果 (交通手段: 熊本電鉄)

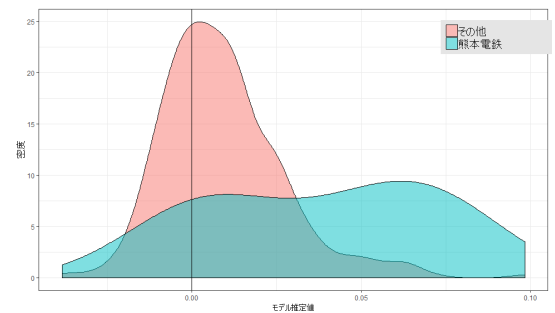


図-16 予測値の密度分布 (交通手段: 熊本電鉄)

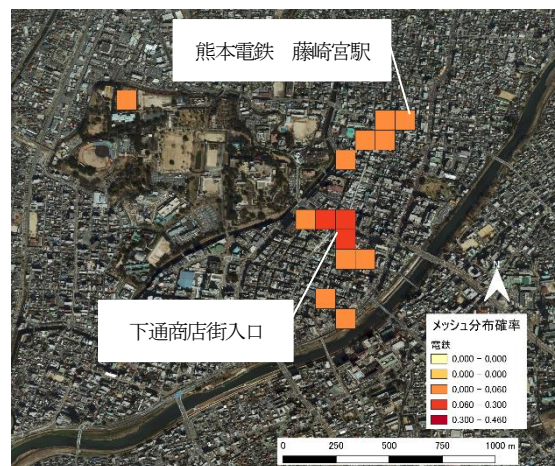


図-17 トピック(13)のメッシュ分布確率

図-15, 図-16には熊本電鉄での来街を補助情報にした場合のトピック別パラメータ推定結果と予測の密度分布を示す。t値はトピック17で1%有意が確認できた（トピック17：t=4.457）。ここでトピック13のメッシュ分布確率を図-17に示す。トピック13は藤崎宮駅から下通商店街入口付近や熊本城にも確率の高いメッシュが存在している。熊本電鉄の予測精度が高かった要因として、自動車の予測結果が他に比べ高くなった要因と同じように、回遊の中心である商店街と離れた場所に駅が位置していることが考えられる。

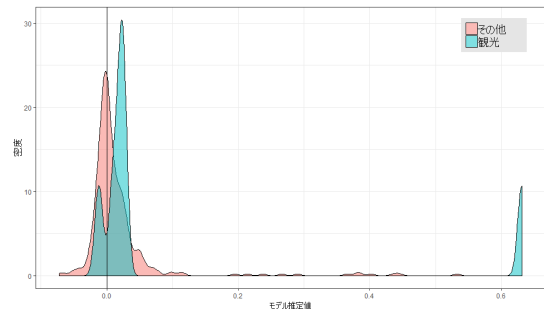


図-20 予測値の密度分布（来街目的：観光）

(3) 来街目的

次に来街目的についての結果を述べる。来街目的に関するROC曲線は図-18のようになった。また各目的のAUCの値は（買い物：0.63，食事：0.60，娯楽：0.63，散歩・散策：0.60，観光：0.68）である。これより，観光目的での来街について詳細を示す。

図-19, 図-20には観光目的での来街を補助情報にした場合のトピック別パラメータ推定結果と予測の密度分布を示す。t値はトピック6，11，29で1%有意が確認できた（トピック6：t=2.885，トピック11：t=10.154，トピック29：t=5.397）。図-21, 図-22はトピック11とトピック29のメッシュ分布確率を示したものである。観光での来街に関しては，熊本城周辺の回遊を含むトピックであることがわかる。熊本市街地のように観光地が商業地域周辺にある場合は，観光目的での来街は他の目的での来街に比べ予測精度が高いことが示された。

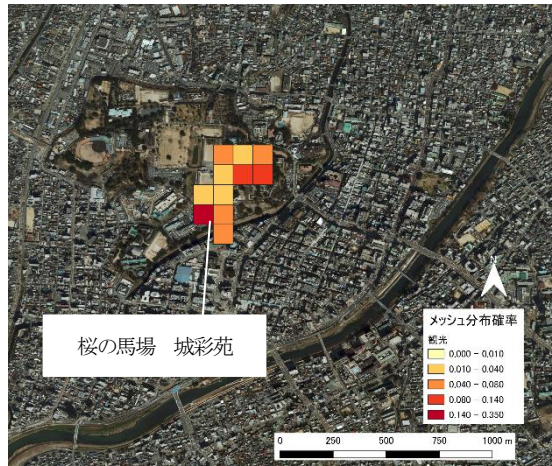


図-21 トピック(11)のメッシュ分布確率

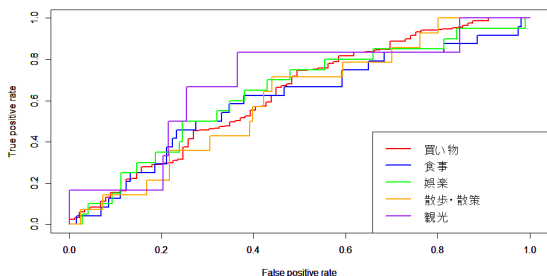


図-18 来街目的に関するROC曲線

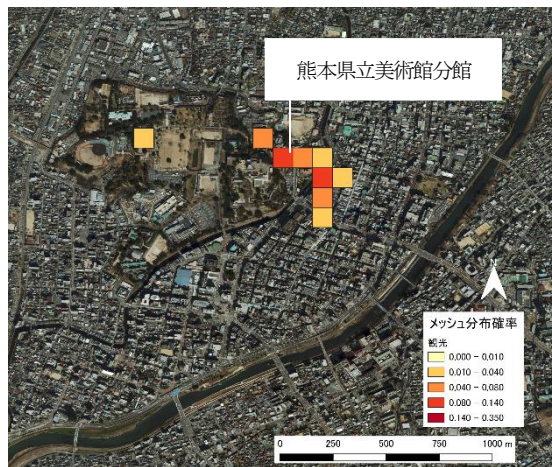


図-22 トピック(29)のメッシュ分布確率

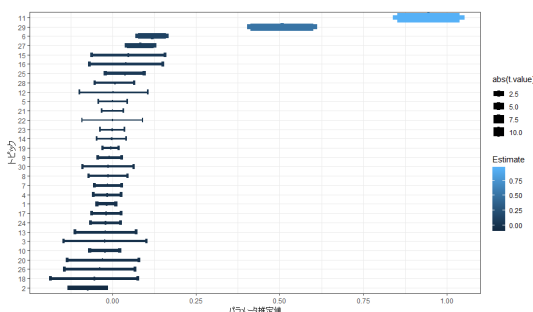


図-19 パラメータ推定結果（来街目的：観光）

(4) 同行者

図-23に同行者に関するROC曲線を示す。AUCの値は（家族：0.60，友達：0.66，1人：0.59）となった。ここで友達との来街について詳細を示す。図-24, 図-25はトピック別のパラメータ推定値と予測の密度分布である。パラメータ推定値のt値はトピック1，10，11等が1に近い推定値で，1%有意となった（トピック1：t=3.912，トピック10：t=4.681，トピック11：t=3.394，トピック24：t=2.721）。図-26にトピック24のメッシュ分布確率を示す。トピック24では熊本城周辺での回遊が抽出された。こ

のトピックの解釈については、分布確率の高いメッシュ周辺には小中学校や高校が位置しているため、友達という属性のパラメータ推定において、このようなトピックが抽出されたと考えられる。

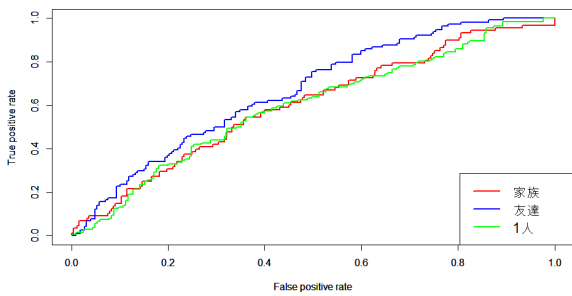


図-23 同行者に関するROC曲線

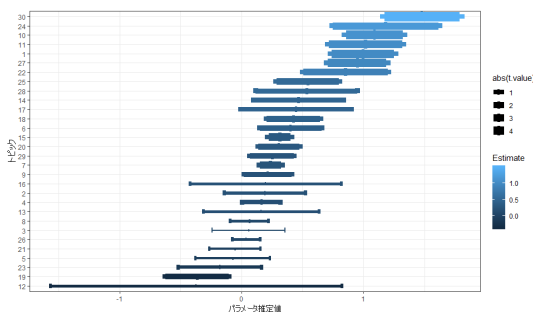


図-24 パラメータ推定結果 (同行者: 友達)

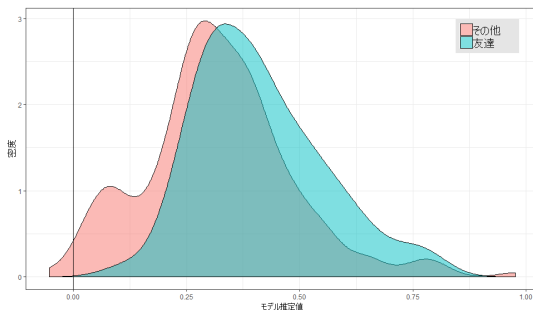


図-25 予測値の密度分布 (同行者: 友達)



図-26 トピック(24)のメッシュ分布確率

6. 結論

本研究では、2013年秋に熊本市都心部で実施されたスマホ型回遊調査のデータを対象に、トピックモデルの拡張モデルであるsLDAを用いた属性情報の推定・予測と予測精度の評価を行った。具体的には、教師あり学習の枠組みで軌跡データを学習データとテストデータに分け、学習データからモデルを推定し、テストデータから属性情報の予測を試みた。本研究から得られた成果を以下にまとめる。

- 1) sLDAを用いたトピック別のパラメータ推定で有意性が確認されたトピックは、各属性の空間的な特徴が表現されていた。
- 2) 予測精度の評価にROC曲線とAUCを用いることで属性間の予測精度の違いを明瞭的に表現することができた。
- 3) 年齢・職業に関する予測では、学生と65歳以上の予測精度が高い値となった。この結果から調査対象とするエリアに特定の年齢層に人気がある、もしくは共に利用されやすい施設がある場合には、一部の属性データと位置情報データのみから年齢層を予測できる可能性が示された。
- 4) 自動車と熊本電鉄利用者の予測精度が高くなった要因として、自動車のパーキングや今回対象にした熊本電鉄の駅は他のバスの停留所などと距離があるため空間的に特徴的で、比較的判別しやすい可能性が示唆された。
- 5) 来街目的、同行者に関する予測では、観光目的以外の属性については予測が難しいという結果が得られた。

本研究では、教師あり学習の枠組みで属性推定を行ったが、属性が付与されていない回遊行動データの活用がより一層期待される中では、教師なし学習を用いて属性を推定することが望ましい。そこで、本研究では、一部のサンプルのみで属性を取得し、モデルを推定、個人属性を取得できなかったサンプルに推定したモデルを適用して属性を推定することを想定して、教師あり学習の枠組みで分析を行った。属性の取得に関して、先に述べた通り、詳細な属性取得ニーズとサンプル数はトレードオフの関係にあると考えられる。従って、入力を求める事項は可能な限り少数で、適切に取捨選択されるべきである。ここで、本研究の貢献としては、属性予測精度の視点から、どの属性を優先的に取得すべきかを示したことにある。ただし、本研究での考察は、今回用いたデータや手法に基づいており、一般性を十分に担保できていないことや、sLDAモデルの適切なパラメータを設定できていない等の問題があるが、これは今後、別稿での課題としたい。

参考文献

- 1) Wolf, J., Guensler, R. and Bachman, W. : Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1768, pp. 125-134, 2001.
- 2) 瀬尾 亨, 日下部 貴彦, 朝倉 康夫: プローブパーソン調査のための逐次学習による交通目的推定法, 土木学会論文集 D3 (土木計画学), Vol. 73, No. 5, pp. I_517-I_526, 2017.
- 3) 山田遊馬, 内山彰, 廣森聡仁, 山口 弘純, 東野輝夫: 携帯電話の基地局通信履歴と地理情報を用いたパーソントリップ推定法の提案, 情報処理学会論文誌, Vol. 57, No. 8, pp. 1826-1834, 2016.
- 4) 古屋秀樹, 岡本直久, 野津直樹: GPS ログデータを用いた訪日外国人旅行者の訪問パターンの分析手法の開発, 運輸政策研究, Vol. 20, pp. 20-29, 2017.
- 5) 川野倫輝, 円山琢也: トピックモデルを用いたスマホ型回遊調査データの基礎分析, 土木計画学研究発表会・講演集, Vol. 58, 2018.
- 6) 中西航, 小林巴奈, 都留崇弘, 松本拓朗, 田中謙大, 菅芳樹, 神谷大介, 福田大輔: Wi-Fi パケットセンサーによる観光周遊パターンの把握可能性: 沖縄・本部半島における検討, 土木学会論文集 D3 (土木計画学), Vol. 74, No. 5, pp. I_787-I_797, 2018.
- 7) 壇辻貴生, 杉下佳辰, 福田大輔, 浅野光行: Wi-Fi パケットデータを用いた観光客の滞在時間特性把握の可能性に関する研究-奈良県長谷寺参道における試み, 都市計画論文集, Vol. 52, No. 3, pp. 247-254, 2017.
- 8) 洪川剛史, 浅野周平, 十河孝介, 森本章倫: 携帯電話基地局データを用いた立地適正化計画の評価指標に関する研究, 都市計画論文集, Vol. 53, No. 3, pp. 408-415, 2018.
- 9) 北川大喜, 関谷浩孝, 糸氏敏郎, 池田大造, 永田智大, 福手垂弥, 新階寛恭, 今井 龍一: 携帯電話基地局の運用データを用いた新幹線トリップの推計手法に関する一考察, 第 56 回土木計画学研究発表会・講演集, Vol. 56, 2017.
- 10) 細江美欧, 桑野将司, 谷本圭志: 非負値テンソル因子分解を用いた交通系 IC カードデータからの移動パターンの抽出に関する研究, 都市計画論文集, Vol. 53, No. 3, pp. 1320-1326, 2018.
- 11) 神谷啓太, 布施孝志: トピックモデルを利用した地域別人口特性の把握手法の提案, 第 55 回土木計画学研究発表会・講演集, Vol. 55, 2017.
- 12) 古屋秀樹: 類似性を考慮した訪日外国人旅行者の訪問パターン抽出に関する基礎的研究, 第 58 回土木計画学研究発表会・講演集, 2018.
- 13) 山本修平, 佐藤哲司: トピックと局面の対応関係に基づく実生活ツイートのマルチラベル分類, 情報処理学会論文誌データベース, Vol. 7, No. 2, pp. 24-36, 2014.
- 14) 堂前友貴, 関洋平: 半教師ありトピックモデルにより選択した地域特徴語を用いた Twitter ユーザの生活に関わる地域の推定, 情報処理学会論文誌データベース, Vol. 7, No. 3, pp. 1-13, 2014.
- 15) Blei, D.M. and Lafferty, J.D.: Supervised topic models, *The Annals of Applied Statistics*, Vol.1, No.1, pp. 17-35, 2007.
- 16) 川野倫輝, 木崎凜太郎, 円山琢也: トピックモデルの拡張モデルを利用したスマホ型回遊データの分析, 土木計画学研究発表会・講演集, Vol. 59, 2019.
- 17) Ramage, D., Hall, D., Nallapati, R. and Manning, C.D.: Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora, *Proceeding EMNLP '09 Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Vol.1, pp. 248-256, 2009.
- 18) 野原浩大朗, 福所誠也, 井村祥太郎, 円山琢也: スマホ・アプリを利用した熊本都心部回遊調査の分析, 第 49 回土木計画学研究発表会・講演集, Vol. 49, 2014.
- 19) 藤田広志, 志村一男, 白石順二, 西原貞光, 東田善治, 山下一也: ROC 解析の基礎と最近の進歩, 日本放射線技術学会雑誌, 第 49 巻, 第 9 号, pp. 1685-1703, 1993.

(2019.10.4 受付)

INFERRING ATTRIBUTE INFORMATION IN SMARTPHONE-BASED TRAVEL-SURVEY USING EXTENDED TOPIC MODELS

Rintaro KIZAKI, Takuya MARUYAMA