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

川野　倫輝[1]・木崎　凛太朗[2]・円山　琢也[3]

[1]学生会員　熊本大学大学院自然科学研究科社会環境工学専攻（〒860-8555 熊本県熊本市中央区黒髪 2-39-1）
E-mail:178d8811@st.kumamoto-u.ac.jp

[2]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

[3]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

人や車の膨大な移動軌跡を取得したデータは増加しているが，そのデータを適切に分析する手法の検討は重要である．文書分析におけるトピック抽出手法のトピックモデルを軌跡データの分析に利用した研究事例も見られるが，その手法の特性を生かした適用蓄積が望まれる．本研究では，トピックモデルを拡張したCorrelated Topic Model (CTM) と supervised Latent Dirichlet Allocation (sLDA) を利用した2013年熊本都心部スマホ型回遊調査の軌跡データの分析事例を示す．CTMにより，事前の地域の特定なしに，回遊で訪問しやすい滞在エリアの組み合わせを抽出できた．sLDAによる分析からは，軌跡データによる個人属性等の推定の可能性を示唆する結果が得られた．

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

1. はじめに

(1) 研究の背景と目的

地方都市が共通して抱える課題に，中心市街地の活力衰退がある．この原因として，郊外住宅地開発等による中心市街地の居住人口の減少や，市街地の拡散・郊外化の進展による中心商業地の機能低下が指摘されている[1]．例えば，熊本市においては，中心市街地の商店数，年間商品販売額の減少や市全体に占める割合の低下が報告されており[2]，今後の人口減少及び高齢化社会において，居住人口の減少に伴う都市活力の低下が懸念されている．このような課題に対して，熊本市中心市街地活性化基本計画[2]では，「にぎわいあふれる城下町」を基本方針の1つとして掲げ，中心商店街や観光拠点である熊本城，再開発地区の回遊性の向上や，ニーズにあった施設の立地を促進することが明記されている．

ここで，中心市街地の回遊性や施設立地の満足度を向上させるためには，来街者の回遊パターンを把握することが重要と考えられる．具体的には，来街者のトリップチェイン中で出現しやすい訪問地の組み合わせを把握できれば，共に訪問されやすい施設を離れた場所に配置して回遊を促すなどの政策的な知見を得られうると考える．また，逆にそれらの施設を隣接させることで，訪問時の負担を低減して来街頻度や訪問施設数を増加させることが出来ると期待される．このような分析は，バスケット分析とも呼ばれ，主にマーケティング分野において，スーパーマーケットでの商品陳列の問題として盛んに取り組まれてきた．しかし，都市の施設配置をこのバスケット分析と同様の視点で扱う研究は決して多くはなく，新手法を適用した更なる知見の蓄積が求められている．

また，このようにビッグデータを収集・活用し，個人の移動特性を把握したうえで，都市の施設配置や道路空間の配分を検討することは，スマートプランニング[3]と呼ばれて推進されている．スマートプランニングが推進されるようになった背景に，膨大な交通系ビッグデータが容易に取得可能となったことがある．Wi-Fiパケットセンサ機器を用いた調査では，スマホ等が発信する電波を受信し，その中に含まれる固有の識別情報 (MACアドレス) を匿名化して取得，それを複数の Wi-Fi パケットセンサ機器で取得することで移動の履歴を把握することができる．スマホ等機器のWi-Fi を有効化することのみ

で移動履歴が取得可能であり，対象者に特別な作業・操作を要求しないことからパッシブな調査とも呼ばれている．またWi-Fiパケットセンサ機器を設置してしまえば，実施者から特別な働きかけをせずとも半永久的にデータが取得できるため，実施者の負担も低く，サンプルあたりのコストも非常に小さい．以上のように，調査対象者と実施者双方の負担とコストを最小限に抑え，24時間365日の膨大なデータを取得できることが期待される．既に，神戸市[4]や奈良県桜井市[5]での適用事例が報告されているが，今後もこの調査方式の適用が増えていくことが予想される．

しかし，その一方で，Wi-Fiパケットセンサ機器で取得される交通行動データは，個人・トリップ属性が取得できないという欠点を持つ．Wi-Fiアクセス時にパスワードと同時に個人属性の入力を求めることや，別途アンケート調査等を実施して個人属性を収集することは可能ではある．しかし，詳細な個人属性を取得しようとすることと，調査協力者数はトレードオフの関係にあると考えられるため，個人属性の取得は慎重に行われる必要がある．具体的には，プライバシー保護の意識の高まりや，入力事項の増加に伴う対象者の負担の増大により，詳細な情報を求めるほど，有効なサンプルが減少していくことが予想される．都市・交通計画の立案に向けたデータ分析においては，サンプルの個人・トリップ属性は非常に重要な情報となる．期待される膨大なサンプル数を保持しつつ，データの有用性を向上させるためには，Wi-Fiパケットセンサから得られる位置情報や測位時刻のみを用いて個人属性を推定することが考えられる．

回遊パターンの抽出においては，膨大な回遊行動データを集約し，教師なし学習の枠組みで，類似したパターンを適切にセグメントすることが必要とされる．また，回遊行動データからの個人属性の推定においても，回遊行動を適切にセグメントすることで，個人属性を表現できると考えられる．以上のように，データを適切にセグメンテーションする手法が求められる．

そこで，本研究では，セグメンテーション手法として，自然言語解析の分野で開発された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，古屋ら[6]の研究のように広域の観光周遊行動データへの適用もある．トピックモデルは，その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．本研究では，トピックモデルの拡張モデルとして，Correlated Topic Model (CTM) と supervised Latent Dirichlet Allocation (sLDA)を取り上げる．

本研究の目的を以下に設定する．

1) 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とその拡張モデルを用いた回遊行動パターン抽出を行う．
2) 上記データの位置情報のみをWi-Fiパケットセンサより得られたデータと仮想的に見なし，トピックモデルの拡張モデルを利用して，個人属性の推定を試みる．

(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．回遊行動分析，交通行動データからの意味情報の推定，そして本研究で用いるトピックモデルの適用事例についての3点について既往研究を整理する．3章では，トピックモデルについての理論と位置情報データへの応用について述べる．4章では，本研究でデータを用いる2013年秋に熊本市都心部で実施されたスマホ型回遊調査について概要を述べるとともに，得られたデータの基礎的な分析を行う．5章では，トピックモデルを用いた移動軌跡データからの回遊行動パターンの抽出を行い，その結果と考察を記す．6章では，同じくトピックモデルを用いて，回遊行動データからの個人属性の推定に取り組み，結果から得られた知見を述べる．最後に，7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，回遊行動分析，交通行動データからの意味情報推定，トピックモデルについての3点について既往研究を整理する．

(1) 回遊行動分析

ここでは，GPSやWi-Fiパケットセンサ等の移動体通信によって取得された回遊行動データの分析についてまとめる．このようなデータの分析は，特に観光科学の分野で盛んに進められており，矢部ら[7]や相[8]によって，その分析手法が詳細にレビューされている．

回遊行動データの代表的な分析手法として，カーネル密度分布がある．カーネル密度は，密度を計算する地点を中心として，任意に指定した検索半径内の点密度を，計算地点の距離減衰効果による重みづけを伴って計算する手法である．佐藤・円山[9, 10]は，本研究でも用いるデータを対象に，カーネル密度を推定して滞在地点や回遊圏域の可視化を行っている． また，推定した回遊圏域を回遊を示す指標として，回帰分析の説明変数に組み込むなどしている．古谷[11]は，カーネル密度に基づく時空

間クラスタリングとハイブリッド階層クラスタリングを組み合わせることで，滞在場所と滞在時刻，個人属性の関係を分析している．

カーネル密度分布を用いた分析では，位置情報が移動状態であるか，または滞在状態であるかの判別を行うことなく，膨大なデータの回遊圏域，特に滞在・滞留地点を簡易に推定することを可能である．しかし，カーネル密度分布では，推定した滞在・滞留地点の共起性，組み合わせまでは把握できない．よって，より高度な分析を行うためには，上記の既往研究のように，推定された密度分布を他の統計的手法や機械学習的手法に組み込む必要がある．

一方で，滞在・滞留地点の共起性，組み合わせを把握するには，1章で触れたバスケット分析を含む空間データマイニングが有効である．この中でも，バスケット分析を用いた分析としては，出水ら[12]の研究がある．出水らは，PP調査のデータに対して，バスケット分析を適用し，市街地内の移動手段と数か所の訪問施設の組み合わせを抽出している．また，近年では，系列パターンマイニングを援用した研究事例も多く存在する．例えば，遠藤ら[13]は，Wi-Fiパケットセンサを用いて取得された北海道内の観光周遊行動データに系列パターンマイニングを適用し，主要な観光周遊パターンを抽出している．

しかし，市街地内での回遊行動においては，観光周遊行動以上に目的地が多様で，混在していると考えられる．これは，訪問地数が膨大で，ベクトル表現化した際に，スパースなデータとなることを意味する．そこで，本研究では，膨大な要素を持ち，スパースなデータの分析に適しているとされるトピックモデルを用いることとした．

(2) 交通行動データからの意味情報分析

交通行動データにおける意味情報とは，トリップ目的等のトリップ属性と，サンプルの性別や年齢といった個人属性を指している．1章で紹介したように，スマホ型調査やWi-Fiパケットセンサを用いた調査では，これらの意味情報を得られないことも多い．よって，これらの推定を試みた研究は数多くあり，関本[14]によって包括的にレビューされている．交通行動の意味情報推定に関する研究の中でも，主な関心を集めるのは，トリップ目的と交通手段の推定に関する研究である．

例えば，トリップ目的を推定する研究として，車両の軌跡データを対象に，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Wolfら[15]の研究がある．また，瀬尾ら[16]は，PP調査において，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を提案している．交通手段の推定に関する研究としては，移動速度を特徴量としたものにYanら[17]の研究があり，加速度を特徴量としたものにFengら[18]やShafiqueら[19]の研究がある．

一方で，本研究と同様に個人属性の推定を行った研究としては，田中ら[20]の研究が挙げられる．田中らは，帰宅時刻や通勤時間などの行動時間に複数のクラスター分析を統合するクラスタリングアンサンブルを適用し，ジオコーディングを行ったPT調査のデータから，就業状態の推定を行っている．

以上のように，交通行動データの意味情報推定においては，滞在地及び交通手段の推定が関心を集めており，個人属性の推定を行った研究事例は少ない．よって，本研究は，このような研究の基礎的なものと位置づけられる．また，田中ら[20]の研究では，行動時間を特徴量としているが，本研究では，回遊パターンという空間的な指標を特徴量として用いている点で新規性を有している．

(3) トピックモデル

先述の通り，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．文書データ分析に用いたものとしては，学術論文の研究トピックの分析として， Sun and Yin [21] は，交通研究分野の22の国際誌に掲載された論文のアブストラクトを対象にトピックモデルを適用し，研究トピックの増減傾向や研究トピック間の関連を把握している．レポート会議やワークショップなどの発言データの解析として，塚井ら[22]は，地域公共交通会議の討議録にトピックモデルを適用し，討議中のトピック抽出やトピック推移の把握を通して同手法の適用可能性を検証している．また，アンケート調査の自由記述分析として，川野ら[23]は聞き取り調査の自由意見にトピックモデルと離散連続モデルを適用し，トピックと個人属性の関係を統計的に分析する手法を提案している．

文書データ以外の解析への応用事例では，神谷ら[24]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．塚井ら[25]は，詳細地理情報の解析へのトピックモデル適用の妥当性を議論している．そして，位置情報に適用した例としては，古屋ら[9,26]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．

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，移動体通信で取得された行動データへの適用例は少ない．また，古屋ら[9,26]の研究は，非常に広域な観光周遊行動を対象としており，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

ここでは，本研究で用いる3種のトピックモデルについて解説する．なお，ここでの解説は参考文献[27]に基づいて示す．

(1) Latent Dirichlet Allocation

a) 概要

回遊パターンの抽出には，トピックモデルの一種であるLatent Dirichlet Allocation[28] (以下，LDA) を用いる．ここで，LDA の概要について説明する．

LDA は本来，Bag of Words (BoW) 表現された文書集合を生成するための確率モデルである．BoW表現とは，文章中に現れる単語のベクトル表現である．また，BoW表現は文章の構造は無視しており，単語の出現回数と共起性を表している．LDAは，BoWから得られる単語の共起性を用いて単語と文書をクラスタリングする手法として用いられる． 極めて汎用的な分析手法であるため，先述のとおり，多様な分野においてクラスタリングの一手法として用いられている．

b) LDAによる文書の生成過程

LDA では，文書中の各単語に，BoWからは直接得ることのできない潜在変数（トピック）を仮定する．また，LDAの特徴として，文書は複数のトピックから構成され，トピックの構成比としての確率分布をもつ．具体的には，文書 d の i 番目の単語を $w_{d,i}$ として，対応する潜在変数を $z_{d,i}$ と定義する．ここで，トピック数を K とし，$\theta_{d,k}$ ($k = 1, 2, \ldots, K$) を文章 d でトピック k が出現する確率とする．トピック分布は $\theta_d = (\theta_{d,1}, \ldots, \theta_{d,K})$ となる．また，各トピックはそれぞれに対応した単語の出現分布 ϕ_k ($k = 1, 2, \ldots, K$) を有している．文書数を D，文書 d の文章長（総単語数）を N_d とする．$\phi_{d,v}$ をトピック k における単語 v の出現確率とし，単語の出現分布を $\phi_k = (\phi_{k,1}, \ldots, \phi_{k,V})$ とする．

θ_d や ϕ_k はDirichlet分布による生成を仮定をするので，以下のように整理できる．

$$\theta_d \sim \mathrm{Dir}(\alpha), d = 1, \ldots, M \quad (1)$$
$$\phi_k \sim \mathrm{Dir}(\beta), k = 1, \ldots, K \quad (2)$$

ここで，ハイパーパラメータ α, β はそれぞれトピック数 K，単語数 V の次元をもつ．潜在トピックと各単語は，Dirichlet分布から導かれるパラメータ θ_d と ϕ_k に従う多項分布から以下のように生成される．

$$z_{d,i} \sim \mathrm{Multi}(\theta_d), i = 1, \ldots, N_d \quad (3)$$
$$w_{d,i} \sim \mathrm{Multi}(\phi_{zd,i}), i = 1, \ldots, N_d \quad (4)$$

以上から，文書 d の生成確率は以下のように示される．

$$p(w_d|\theta_d, \phi) = p(\theta_d|\alpha_d) \prod_{i=1}^{N_d} p\,(z_{di} = k|\theta_d) p(w_{di}|\phi_{zd,i}) \quad (5)$$

また，全文書データの生成確率を以下に示す．

式(6)より，LDAでは，文書のトピック分布を表すハイパーパラメータ α と θ_d，単語の分布を示すハイパーパラメータ β と ϕ_k によって規定される．パラメータの推定に

$$p(w|\theta, \phi) = \prod_{d=1}^{N} \int p(\theta_d|\alpha) \prod_{i=1}^{N_d} p\,(z_{di} = k|\theta_d) p(w_{di}|\phi_{zd,i}) d\alpha \quad (6)$$

は，変分ベイズ法等が考えられるが，本研究では，崩壊型ギブスサンプリングを用いる．

(2) Correllated Topic Model

Correlated Topic Model [29] (以下，CTM) は， LDAの拡張モデルの一種である．LDAではトピック間の独立を仮定していたのに対して，CTMではトピック間の相関を考慮することが可能なモデルである．

LDAでは，トピック分布がDirichlet分布から生成されていたが，Dirichlet分布ではトピック間の相関を考慮することは不可能である．そこで，CTMでは，トピック分布を正規分布から生成している．具体的には，まず K 次元の実数ベクトル $\eta_d = (\eta_{d1,\ldots,}\eta_{dK})$ を平均 $\mu = (\mu_1, \ldots, \mu_K)$，共分散行列 Σ の正規分布Normal (μ, Σ) から生成する．ここで Σ は大きさ $(K \times K)$ の行列で，その要素 $\sigma_{kk'}$ がトピック k と k' の間の相関を規定する．ここで，正規分布から生成された η_d の要素は負の値も取りうる．

また，η_d の全要素の和をとっても1とはならないため，η_d の指数をとり，正規化する．

$$\theta_{d,k} = p(k|\eta_d) = \frac{\exp(\eta_{dk})}{\sum_{k'=1}^{K} \exp(\eta_{dk'})} \quad (7)$$

この変換関数は，ソフトマックス関数と呼ばれる．**図-2** のグラフィカルモデルに示すように，実数ベクトル η_d をソフトマックス関数によってトピック分布 θ_d に変換した後の単語の生成過程はLDAと同様である．

よって，CTMにおける文書 d の生成確率は以下で表される．

$$p(w_d|\eta_d, \phi) = \prod_{d=1}^{N_d} \sum_{k=1}^{K} p\,(z_{di} = k|\eta_d) p(w_{di}|\phi_k) \quad (8)$$

なお，パラメータの推定には変分ベイズ法を用いる．

(3) Supervised LDA

a) 概要

Supervised LDA[30](以下，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りうる．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-LDA[31]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．

c) 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．

$$y_d \sim \mathrm{N}(\eta^{\mathrm{T}}\bar{z}_d, \sigma^2) \tag{9}$$

ここで，$\bar{z}_d$は経験トピック分布と呼ばれ，以下のように示される．

$$\bar{z}_d = \frac{1}{N_d}\sum_{i=1}^{N_d} z_{d,i} \tag{10}$$

$z_{d,i}$は，式(3)に示す通り，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．

$$y_d = \eta_k \bar{z}_{d,k} \tag{11}$$

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．

(3) 回遊行動調査データへのトピックモデルの適用

回遊行動データにトピックモデルを適用するにあたり，本研究では分析対象エリアをメッシュ化したうえで，文書dをサンプルの移動軌跡，単語wをメッシュ，文書dにおける単語wの出現回数$n_{w,d}$をメッシュ内で測位されるポイント数と考える．**図-2**にGPSより得られる回遊行動データをBoW表現化する際のイメージ図を示す．即ちここでは，wはメッシュ，dは回遊行動データ，Dは総サンプル数，Vは分析対象地域内の総メッシュ数である．なお，この考え方は，古屋ら[9]の手法に従っている．

文書分析におけるトピックモデルが単語の集合としてトピックを定義するのに対して，本研究では，メッシュの集合としてトピックを定義する．ここでは，トピックは回遊パターン，または回遊エリアと考えることが出来る．5章では，このトピック内のメッシュの組み合わせやトピックの組み合わせを分析する．6章では，このトピックから教師あり学習で個人属性を推定する．

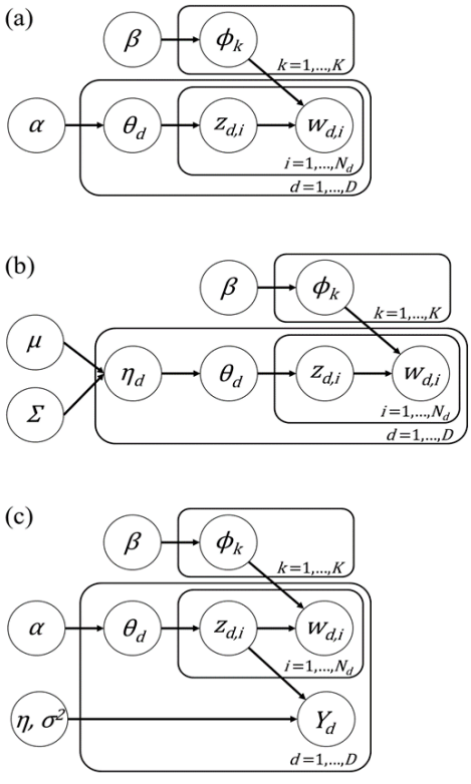
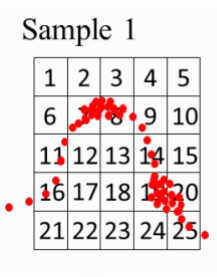


図-1 (a) LDA，(b) Correlated Topic Model，(c) Supervised LDAのグラフィカルモデル



BoW表現化

	sample 1	sample 2	sample 3	sample 4	sample 5	sample D
メッシュ1	0	0	1	0	1	$n_{1,D}$
メッシュ2	0	0	0	1	3	$n_{2,D}$
メッシュ3	0	0	0	1	4	$n_{3,D}$
メッシュ4	0	0	0	0	2	$n_{4,D}$
メッシュ5	0	0	0	0	1	$n_{5,D}$
メッシュ6	0	1	0	0	0	$n_{6,D}$
メッシュ7	11	5	2	4	0	$n_{7,D}$
メッシュ8	8	6	3	7	0	$n_{8,D}$
メッシュ9	2	3	0	3	0	$n_{9,D}$
メッシュ10	0	2	0	1	2	$n_{10,D}$
メッシュ11	2	0	0	0	0	$n_{11,D}$
メッシュ12	0	0	0	0	2	$n_{12,D}$
メッシュ13	0	0	0	0	2	$n_{13,D}$
メッシュ14	3	0	2	0	1	$n_{14,D}$
メッシュ15	0	0	0	0	2	$n_{15,D}$
メッシュ16	2	0	0	0	0	$n_{16,D}$
メッシュ17	0	0	0	0	4	$n_{17,D}$
メッシュ18	0	0	0	0	5	$n_{18,D}$
メッシュ19	13	0	12	0	0	$n_{19,D}$
メッシュ20	7	0	14	0	0	$n_{20,D}$
メッシュ21	0	0	0	0	1	$n_{21,D}$
メッシュ22	0	0	0	0	1	$n_{22,D}$
メッシュ23	0	0	0	0	2	$n_{23,D}$
メッシュ24	0	0	1	0	0	$n_{24,D}$
メッシュ25	3	0	0	0	0	$n_{25,D}$
メッシュ V	$n_{V,1}$	$n_{V,2}$	$n_{V,3}$	$n_{V,4}$	$n_{V,5}$	$n_{V,D}$

図-2　回遊行動データのBoW表現

4.　データセット

(1)　くまもとまち歩き調査の概要

a)　調査目的

2012年に熊本都市圏においてパーソントリップ調査（以下，PT調査）が実施された．PT調査は，都市圏レベルの人の動きを把握しているが，市街地レベルでの詳細なの回遊行動は十分には調査されていない．そこで熊本市中心市街地における回遊行動調査が2013年秋に実施された．この調査では，スマートフォン（以下，スマホ）を利用し，まちなかでの人の動きを把握するもので，にぎわいの向上や，歩きやすい歩行環境の整備計画，交通調査の高度化を検討するための基礎的なデータを収集することを目的としている．

b)　調査概要

調査の詳細について**表-1**に示す．調査日は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．

c)　調査モニターの募集

今回調査では調査モニターを事前登録型と当日登録型の2通りで募集した．事前登録型はポスターやチラシに記載されたURLやQRコードから登録サイトにアクセスし，事前に性別や年齢等の情報を登録した上で参加する方法である．当日登録型は調査当日にまちなかの駐車場や駐輪場，公共交通乗降場で20人/日程度の学生による調査参加依頼（まちなかキャッチ）を行い，現地で登録サイトにアクセスし参加してもらう方法である．どちらのモニターにも登録時にIDとパスワードが割り振られ，IDから個人属性の区別を行う．

d)　調査方法

調査のメインはスマホアプリの「スマくま」を用いた位置情報の取得である．参加者個人のスマホにアプリをインストールしてもらい，スマホのGPS機能を活用し，測位を行う．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．出発時にボタンを押してもらい調査中はアプリを起動した状態で回遊を行う．回遊終了後にまちなか4箇所に設けられたポートに立ち寄ってもらい，アプリの到着時にボタンを押してもらう．

また，スマホ調査への意識を調べるため回遊終了後のポートにてヒアリングによる事後アンケートも行う．ポートでは調査の内容についての事後アンケートを行い，粗品の受け渡しを実施して調査が終了となる．

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[32] を参照されたい．

(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

表-1　熊本都心部回遊調査概要

調査日	平成 25 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

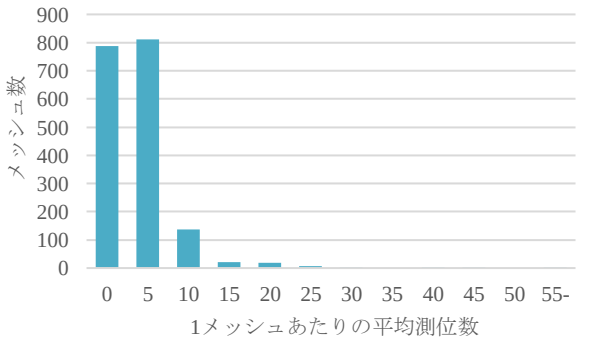


図-3　1メッシュあたりの位置情報の平均測位数の頻度分布

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，回遊パターンを抽出するに際して，極力移動状態のデータは除かれることが望ましい．移動状態のデータを除去するために，移動軌跡を移動状態と滞在状態に判別する手法は，D-Star アルゴリズム [33]をはじめとして数多くの手法が提案されている．しかし本研究では，1 メッシュあたりの位置情報の測位点数が少ないほど移動状態に近いデータであることを利用し，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．このような事前処理を要する点で，本研究に用いたデータは既存研究 [6]で対象とされてきた広域な周遊データは異なると言える．

図-3 にメッシュ別の平均測位点数の頻度分布を示した．平均測位点数が 5 以下のメッシュが多く存在することがわかる．そこで，本研究では，1 メッシュあたりの測位点数が 5 以下の場合，そのメッシュ内の測位点を移動状態とみなし，分析対象から除くこととする．これは，文書分析において，単語の出現回数に閾値を設け，低頻度語を除去することと同様である．これにより，サンプル中に測位点が存在しなくなったサンプルが 25 サンプル発生した．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．なお，分析対象エリアについては先行研究 [34]を参考にされたい．

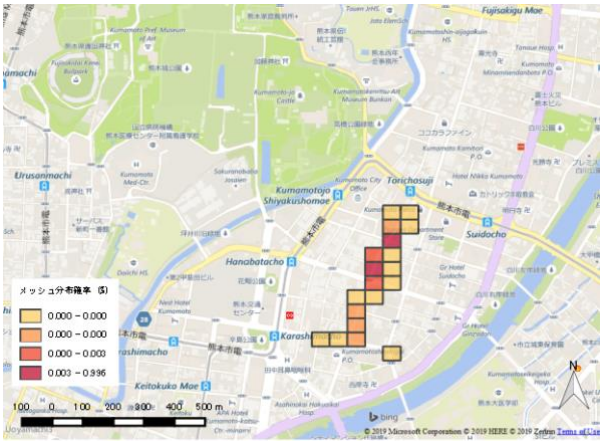


図-4　トピック (1)

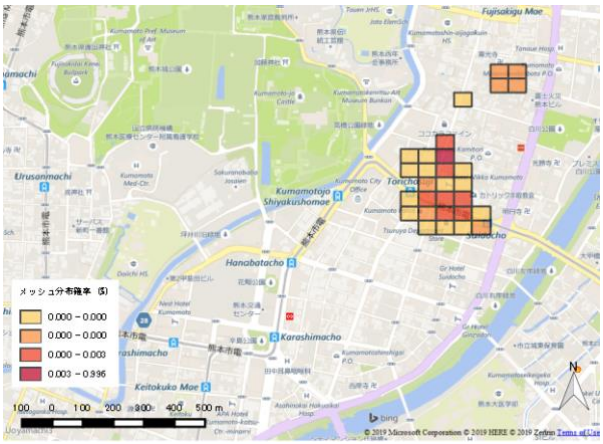


図-5　トピック (4)

5.　LDA及びCTMを用いた回遊パターンの分析

(1)　抽出されたトピック（回遊パターン）の可視化

まず，抽出されたトピックを地図上で可視化する．トピックモデルの推定アルゴリズムに関しては多くの研究例が蓄積されており，ある程度の文章量の文書が多数得られる場合，ハイパーパラメータが安定的に推定できることが示されている[35]．先述したように，本研究では，崩壊型ギブスサンプリングを用いた．ハイパーパラメータの初期値は一様に $\alpha=(0.1,\ldots,0.1)$，$\beta=(0.1,\ldots,0.1)$，サンプリング数は1000回，トピック数はk=25と設定した．

図-4から**図-10**に，抽出されたトピックを可視化したものを示す．ここでは，紙面の都合上一部のトピックのみを取り上げている．なお，各メッシュは $\phi_k=(\phi_{k,1},\ldots,\phi_{k,V})$ に基づき色付けされている．ϕ_kはトピックkにおける単語 v の出現確率を示す．回遊行動データの分析においては，トピックkに対するメッシュvの寄与率，または構成確率と解釈できる．また，図中のϕ_kの閾値は，プリティブレーク分類法に基づいて決定されている．

図-4は抽出されたトピック(1)を可視化したものである．メッシュ分布確率ϕ_kの空間的分布から，下通アーケードから，サンロード新市街の一部までの回遊を含むトピックであることがわかる．具体的には，調査実施時に出店していたダイエーから銀座通りまでに，メッシュ分布確率が高いメッシュが集中している．この区間のメッシュには，カラオケ店等の娯楽施設が多く立地している．よって，娯楽目的の来街や，比較的に年齢層が低い来街者による回遊パターンとも考えれる．

図-5にトピック (4) を示す．このトピックは，上通アーケードから通町筋一帯のメッシュを含んでいる．また，これらのメッシュとは連続していないが，上乃裏のメッシュも含んでいる． ここから，このトピックが上通付近の施設と上乃裏を共に訪問する回遊行動を示しているということが読みとれる．**図-4**のような下通での回遊と比較して，このトピックに示す回遊行動は高齢の来街者のものであることが考えられる．

図-6にはトピック (11) を示した．このトピックでは，交通センターを中心に，新市街方面へ回遊が広がっているトピックである．また，このトピックには，県民百貨店も含まれており，高齢来街者の買い物目的の回遊行動パターンと推察される．また，ここに含まれる新市街にはパチンコ店も立地しているため，男性の娯楽目的の来街も考えられうる．

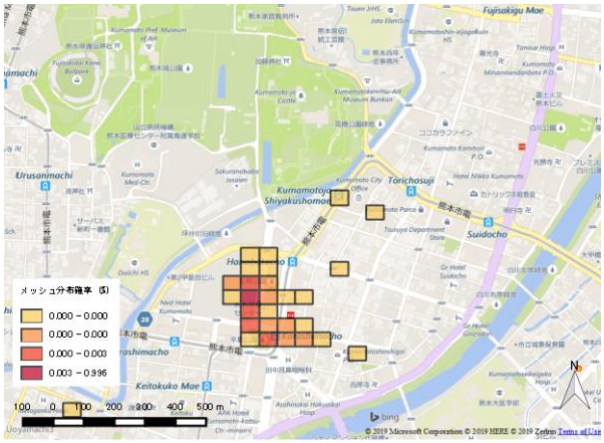


図-6　トピック (11)

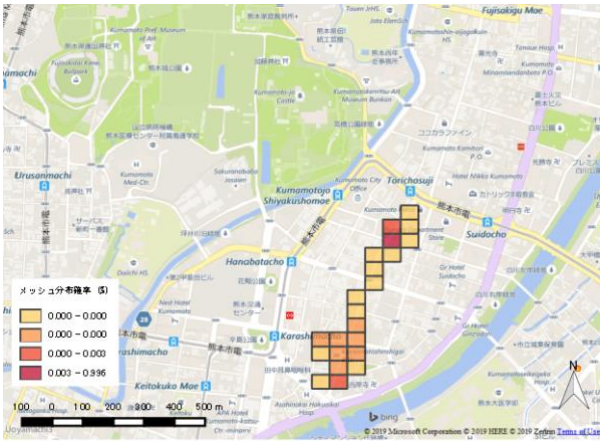


図-8　トピック (17)

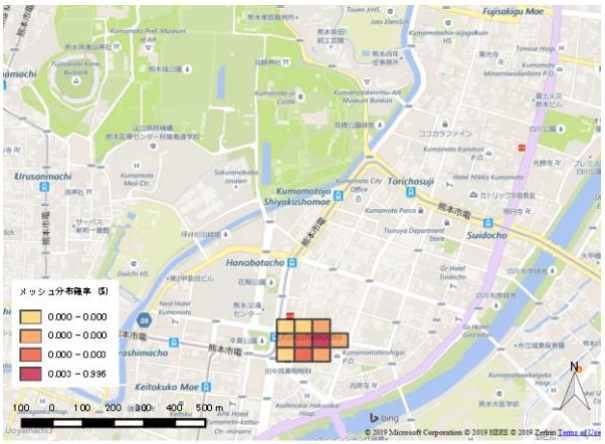


図-7　トピック (16)

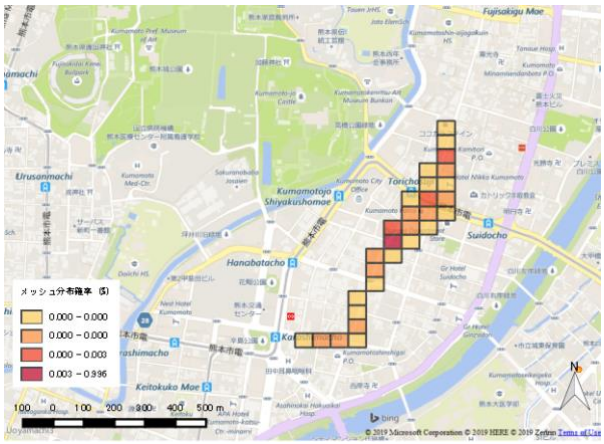


図-9　トピック (18)

図-7にトピック (16) を示す．このトピックは，トピック (11) と同じく，新市街を含むトピックであるが，交通センターは含んでいない．よって，トピック (11) と同じく，娯楽目的の男性の来街パターンと予想される．また，付近には飲食店も多数立地している．ここから，食事目的の来街も含んでいるとも考えられる．

図-8にトピック (17) を示す．このトピックは，下通アーケードから新市街 の一部，シャワー通りまでの回遊を含むトピックである．メッシュ分布確率が高いメッシュとしては，ダイエー前とシャワー通り一帯のメッシュがある．前者に関しては，ここが待ち合わせ場所としてよく利用されることも考えられるが，ここには，調査終了後にアンケートを回答してもらうためのポートも設置されていたため，その影響も存在することには留意が必要である．後者のシャワー通り一帯にメッシュに関しては，図-4に示したトピック (1) には含まれていない特徴的なメッシュである．シャワー通りには，お洒落なブティックやカフェ，雑貨店などが立ち並ぶ通りとされており，若年層，特に若い女性の回遊を含むことが予想される．また，買い物目的の来街に関するトピックであることや，友人等の同伴を伴っていることも予想される．

図-9にはトピック (18) を示した．このトピックは，上通アーケードから下通アーケード，新市街までを含むトピックである．また，上通アーケードを含み，県道28号線を横断して南に回遊が広がっているという点で，同じく下通アーケードを中心とするトピック (1) やトピック (17) とは異なるトピックである．メッシュ分布確率の高いメッシュも，それらの2トピックと異なっており，ダイエー前より北側のメッシュに集中している．ここには，鶴屋百貨店やパルコ等の大型商業施設が立地しているため，買い物目的での来街が多いことが予想される．

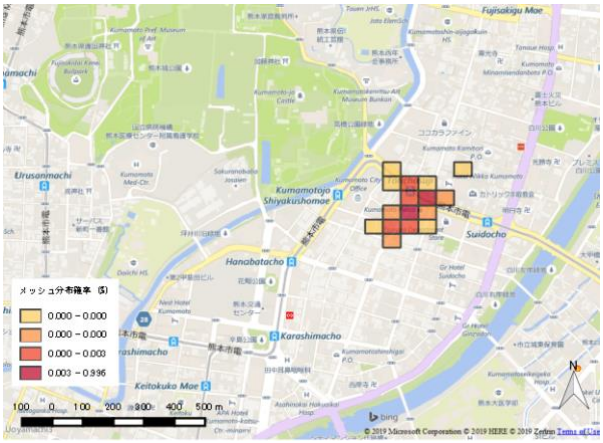


図-10　トピック (23)

図-10にはトピック (23) を示した．これまでに示してきた下通アーケードを中心としたトピックと比較して，狭域なトピックとなっている．熊本市の中心市街地で最も交通量が多い地点であり，このトピックを含むサンプルの個人・トリップ属性は多様であると考えられるため，その推測は困難である．

ここでは，図-4と図-8，図-9のような空間的に類似した回遊パターンが異なるパターンとして抽出された．これは，メッシュに対して，トピックが一意に決まらないというLDAの性質によるものである．また，この結果はカーネル密度分布のような密度ベースの手法からは得られないものであると考えられる．

(2) 抽出されたトピック（回遊パターン）と移動軌跡の比較

ここで，前節で抽出されたトピックが，実際の移動軌跡の空間的特徴をどれだけとらえているのか確認が必要である．また，LDAでは，サンプルをトピックの混合物で表現する．1サンプルが複数トピックから構成される場合，トピックの共起によって，サンプルを表現できるか確認する必要もある．

そこで，トピックと実際に取得された移動軌跡を比較する．図-11にLDAで抽出した回遊行動トピック (4) と1サンプルの移動軌跡の比較を示す．ここで取り上げた移動軌跡は，$\theta_{d,4}$=1となるサンプルdの移動軌跡である．

$\theta_{d,k}$は，文章dをトピックkが構成する確率であるが，本研究においては，サンプルdを回遊行動トピックkが構成する確率と捉えることが出来る．すなわち，$\theta_{d,k}$=1となるサンプルdとは，サンプルに対して，トピックが一意に決まるサンプルである．ここでは，取り上げたサンプルがトピック (4) のみによって構成されるということである．

図-12を見ると，トピック (4) を構成するメッシュは移動軌跡と重なっており，抽出されたトピックは概ね妥当と考えられる．また，このトピック中では，メッシュが完全に連続しておらず，3群に分かれている．移動軌跡を見ると，これらのメッシュ群上で位置情報の測位点が集中しており，このサンプルが滞在状態，または低速の移動状態であることが読みとれる．ここから，抽出されたトピックが訪問地，回遊エリアの組み合わせを表現できていると言える．ここでは，上通アーケードから通町筋一帯，上通アーケード北端，上乃裏の組み合わせの回遊であると判断できる．一方で，サンプルが比較的高速な移動状態にあると考えられる部分では，メッシュに確率が割り振られていない．ここから，4章で示した分析サンプルの抽出処理により適切に移動状態のデータを省けていることもわかる．

図-13は，このトピック (4) に関して，$\theta_{d,4}$=1となる全てのサンプルdの移動軌跡を重ねて表示したものである．ごく一部の部分で滞在が抽出出来ていないが，回遊パターンとして概ね適切にトピックが抽出出来ていることがわかる．

ここまでは，$\theta_{d,k}$=1となるサンプルのみを扱った．次に，$\theta_{d,k}$≠1のサンプルと抽出したトピックを比較する．$\theta_{d,k}$=1となるサンプルでは，サンプルに対してトピックが一意に定まっていたが，$\theta_{d,k}$≠1のサンプルでは，複数トピックの組み合わせで，サンプルdが構成されることを意味している． LDAはクラスター分析等の他のセグメンテーション手法とは異なり，サンプルとトピック (セグメント) が一意に決まらない状況を確率的に許容しているためである．つまり，図-12のように，$\theta_{d,k}$が高いときは，そのトピックkがサンプルdを代表して表現していると言えるが，$\theta_{d,k}$が低いと，トピックkのみでサンプルdを説明することが難しく，複数のトピックを組み合わせて説明することが必要となると考えられる．

図-10のトピック (23) を例に，トピックとトピック分布確率$\theta_{d,23}$別の移動軌跡の比較を図-13に示した．図中の上段左が$\theta_{d,k}$=1となるサンプルの移動軌跡になっており，下段右が0.20＜$\theta_{d,23}$＜0.39となるサンプルの移動軌跡

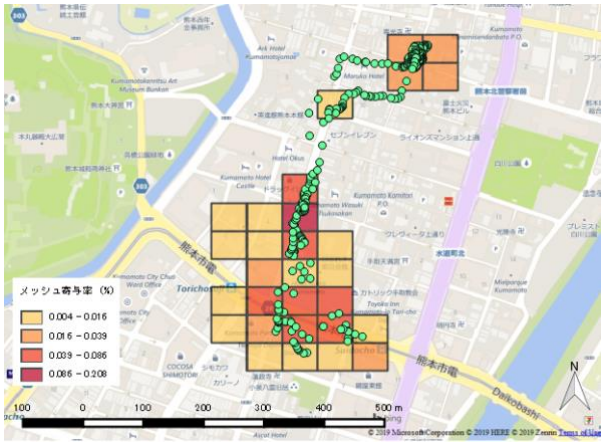


図-11　抽出されたトピック (4) と$\theta_{d,4}$=1となるサンプルの比較

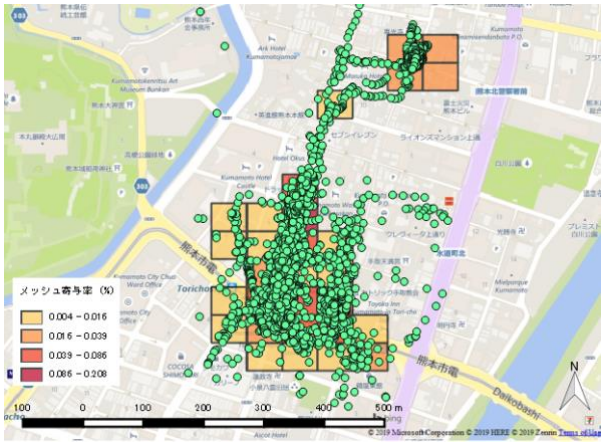


図-12　抽出されたトピック (4)と$\theta_{d,4}$=1となる全てのサンプルの比較 (N=22)

である．ここから，トピック分布$\theta_{d,23}$が大きいほどトピック (23) が示すメッシュに集中している一方で，$\theta_{d,23}$が低いほど移動軌跡が広範囲に広がっており，トピック (23) のみでは説明できないこととが確認できた．より具体的には，$\theta_{d,k}$=1，0.80＜$\theta_{d,23}$＜ 0.99となるサンプルでは，そのサンプルの滞在エリアがトピック (23) で説明できるが，0.60＜$\theta_{d,23}$＜ 0.79と0.40＜$\theta_{d,23}$＜0.59，0.20 ＜$\theta_{d,23}$＜0.39となるサンプルには，トピック (23) には含まれないメッシュへの滞在が確認できる．とりわけ，0.20 ＜$\theta_{d,23}$＜0.39となるサンプルでは，トピック (23) には含まれない熊本城や交通センター，新市街等への滞在が多く確認できる．すなわち，トピック (23) とは別のトピックがサンプルを構成しており，そのトピックに上記の施設が含まれていることが推測される．

そこで，0.20 ＜$\theta_{d,23}$＜0.39のサンプルから，θ_d=($\theta_{d,16}$=0.73, $\theta_{d,23}$=0.22) となるサンプルdを**図-14**に示した．言い換えると，このサンプルはトピック (16) によって約73%，トピック (23) によって約22%の確率で構成されているということである．サンプルの移動軌跡から，このサンプルが新市街と下通の一部に滞在したことが推測できる．この滞在エリアとトピック (16) と (23) はほぼ重なっており，サンプルdが複数のトピックで構成される場合でも，サンプルdを適切に説明できることがわかった．

(3) トピック（回遊パターン）別の個人・トリップ属性の分析

本スマホ調査の特徴として個人・トリップ属性が取得されていることがある．そこで，この個人・トリップ属性とトピック (回遊行動パターン) の関係を分析する．

この際，サンプルとトピックを一対一で結びつける必要があるが，サンプルdに対してトピックは$\theta_{d,k}$で表されるようにk次元のベクトルである．よって，θ_d = ($\theta_{d,1}$…, $\theta_{d,K}$) のうち，最も確率$\theta_{d,k}$が高い回遊行動トピックkを，サンプルdに対する優勢トピックと設定し，この優勢トピックと個人・トリップ属性の関係を分析する．

しかし前節の結果より，$\theta_{d,k}$＜ 0.8のサンプルdは，トピックk以外の複数のトピックで構成されており，トピックkを評価するのにふさわしくはないことが分かった．よってここでは，$\theta_{d,k}$＞0.8となる優勢トピックを用いて分析を行った．この結果，総サンプル数は289となっている．

また，25トピック全てを扱うと図が煩雑になるため，ここでは，本章1節で取り上げたトピックのみを取り上げることとする．

図-15～図-19に回遊行動トピック別の性別比，年齢比，来街目的，同行者の有無，就業状況を示す．性別以外で，トピック毎に個人・トリップ属性の特性を有していることがわかる．以下に，トピック毎の特性をまとめる．

図-4のトピック (1) は，下通アーケード一帯を含み，特にその中でもカラオケ店等が立地する区間のメッシュ確率が高くなっていた．ここから若年層の娯楽目的の来街が多いとの予想を立てていたが，概ねこれに一致する結果が得られた．特に，このような来街者には学生が多いことも分かった．また，このトピックでは，食事目的の来街割合も高くなっている．なお，この来街目的は，トリップごとに付与されるトリップ目的とは異なり，トリップチェインに対して1つ付与される目的である．食事目的で来街したが，娯楽活動も行った，等の来街目的

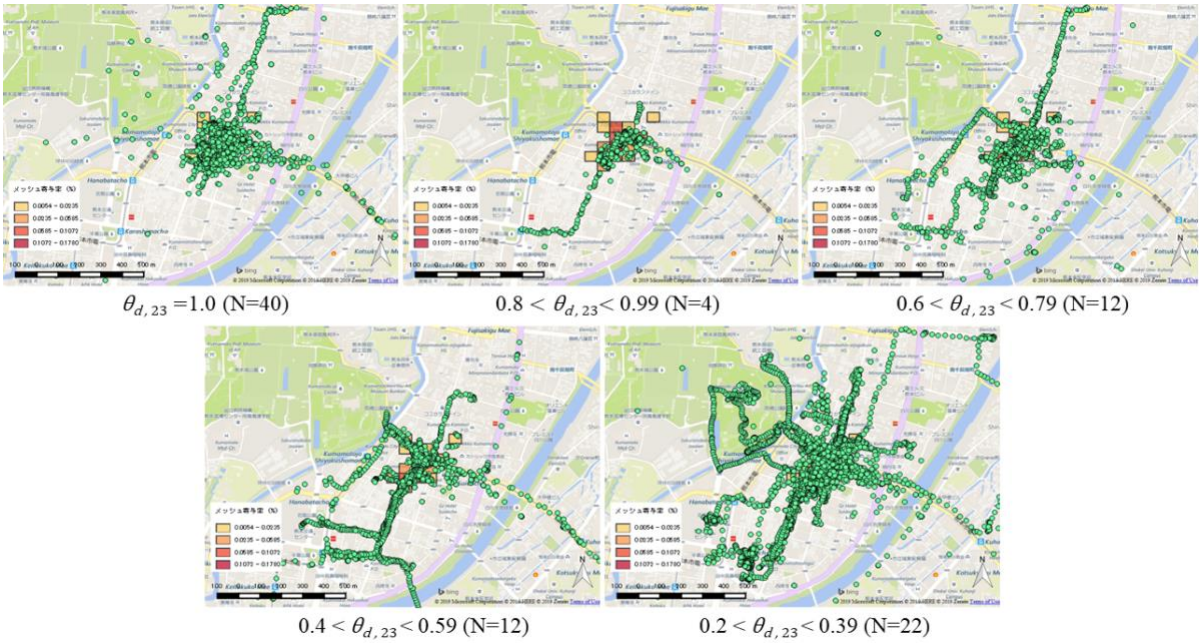


図-13　抽出されたトピック (23) と $\theta_{d,23}$別のサンプルの比較

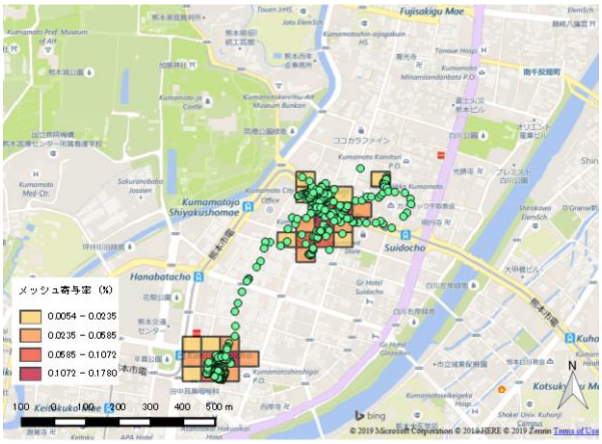


図-14　トピック(16)と(23)，$\theta_d=(\theta_{d,16}=0.73, \theta_{d,23}=0.22)$となるサンプルの比較

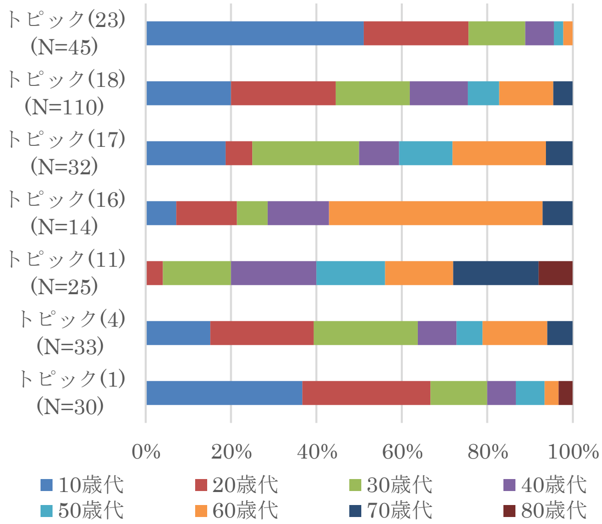


図-16　トピック別の年齢分布

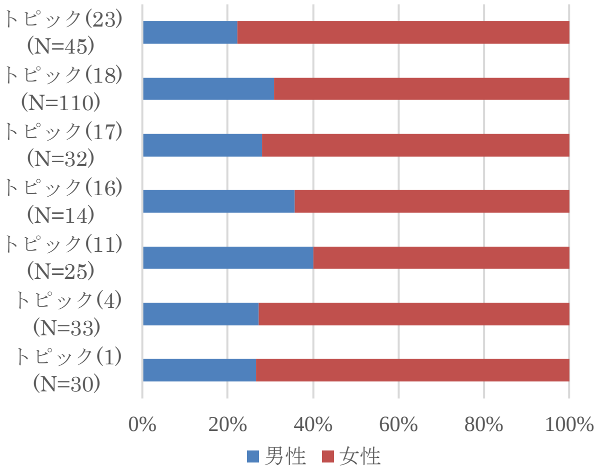


図-15　トピック別の性別分布

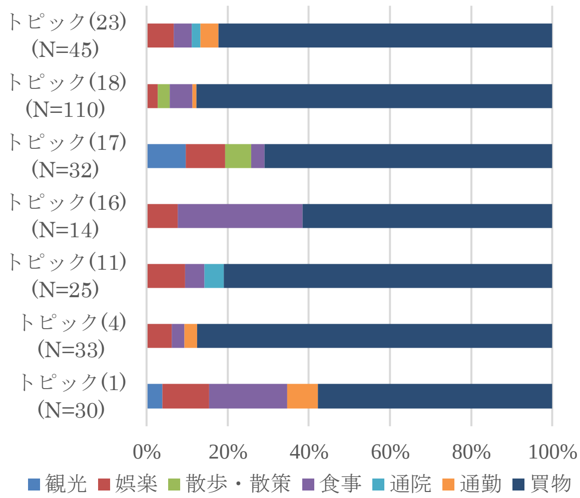


図-17　トピック別の来街目的分布

と実際の活動目的が異なる場合もありうることに留意が必要である．

図-5のトピック (4) は上通と上乃裏の回遊が見て取れるトピックである．ここでは，トピック (1) と比較して年齢層が高くなっており，仕事をしている20歳代～40歳代の若中年層がボリューム層となっている．また，来街目的としては，買物の割合が高い．上通には，ファッションビル等も立地しているため，ここに買物を行い，その後上乃裏にも足を延ばすといった回遊行動が含まれていると考えられる．

図-6のトピック (11) は，交通センターから新市街に回遊が広がるトピックであった．このトピックで特徴的な点としては，高齢者と単独での回遊が多いことである．主な来街目的は，買物と娯楽であるため，事前の 予想通り，高齢者の県民百貨店への買物や新市街へひとりでパチンコへ行くといった回遊行動が含まれていると考えられる．

図-7のトピック (16) は，新市街のみに回遊が集中するトピックである．このトピックはトピック (11) と空間的に類似していたが，特性はいくつかの点で異なっている．具体的には，より高齢者が多くなっている．また，友人を連れ立った食事目的の来街が多いと解釈できる．

図-8のトピック (17) は，下通一帯からシャワー通りまでの回遊を含むトピックであった．空間的に類似していたトピック (1) と比較すると，年齢層が高くなっている．また，その他のトピックと比較すると，ここでは観光目的の来街が特徴的であると言える．カーネル密度分布を用いた既存研究[10]でも，シャワー通りで観光目的の滞在があると示されていたことから，これは観光目的の活動はシャワー通りに集中していると考えられる．

図-9のトピック (18) は，上通アーケードを含み，県道28号線を横断して下通，新市街にまで回遊が広がるトピックである．回遊の範囲が広いためか，その個人・トリップ属性は多様である．来街目的のみ，買物が特徴的となっており，鶴屋百貨店やパルコ等での買物が主な目的

の回遊行動と推察される．

図-10のトピック (23) は，下通アーケードの狭域な部分に回遊が集中するトピックである．ここで特徴的なのは，年齢層が低く，友達連れの学生がボリューム層となっていることである．

ここまで，抽出されたトピックと個人・トリップ属性の分析を行った．LDAで抽出されたトピックと実際の個人・トリップ属性の関係は，既存手法のカーネル密度分布より得られた結果[9]に合致しており，本手法が回遊パターンの抽出手法として有用であることを示していると考えられる．一方で，カーネル密度分布のような密度ベースの分析手法では発見できない，空間的に類似しているが，異なった性質を持つ回遊パターンの抽出が可能であることが示された．

(4) Correlated Topic Model (CTM) を用いたトピック（回遊パターン）の相関の分析

ここまで，LDAを用いて回遊行動パターンの抽出を行った．ここで，滞在エリアの組み合わせを発見するためには，トピックの組み合わせ，すなわちトピック間の相関を分析することが有効であると考えられる．しかし，先述の通り，LDAではトピック間の独立を仮定しているため，LDAにより抽出されたトピックの相関を分析することは適切ではない．

そこで，トピック間のLDAの拡張モデルで相関を考慮可能なCTMを用いて，トピックを抽出し，その相関を分析する．なお，ここで抽出されるトピックはLDAで抽出されたトピックとは異なることに留意されたい．

a) トピック（回遊行動パターン）の抽出とトピック別の個人・トリップ属性の分析

まず，LDAと同様に回遊行動データからトピックを抽出する．トピック数はLDAと同じくk=25とした．**図-20**～**図-22**に，CTMによって抽出された観光行動に関すると考えられるトピックを示す．

図-20に，CTMより抽出されたトピック (1) を示した．このトピックには，シャワー通り一帯のメッシュが含まれている．一方で，**図-15**のLDAにより抽出されたトピックでは，シャワー通りは下通アーケードと同じトピックとして抽出されていた．これは，CTMではトピック間の相関が考慮できるため，1サンプルをLDAより複数のトピックで構成しており，1つ1つのトピックが狭域になったことが原因と考えられる．先に示した通り，シャワー通りは観光目的での回遊が確認されている．よって，このトピックは観光に関するものであると考えた．

図-21にトピック (11) を示す．このトピックに関しては，LDAでも同様のトピックが抽出されている．これは熊本城を含んでいる特徴的なトピックである．熊本城は，熊本市の中心部で最も有名な観光地であるため，観光に関する回遊行動であると考えた．よって，観光目的での来街や，熊本県外居住者による回遊であると推察される．

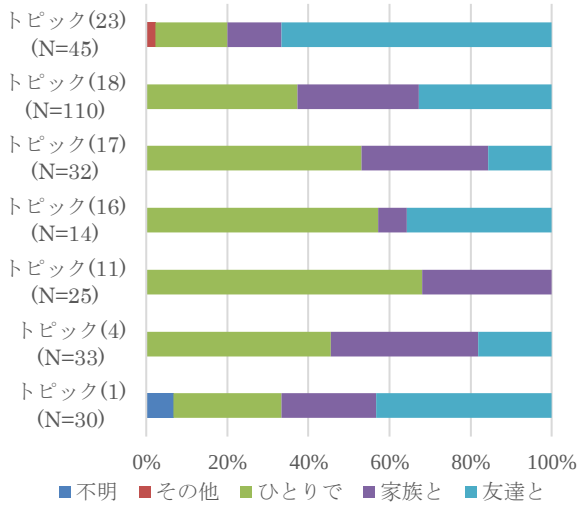


図-18　トピック別の同行者の分布

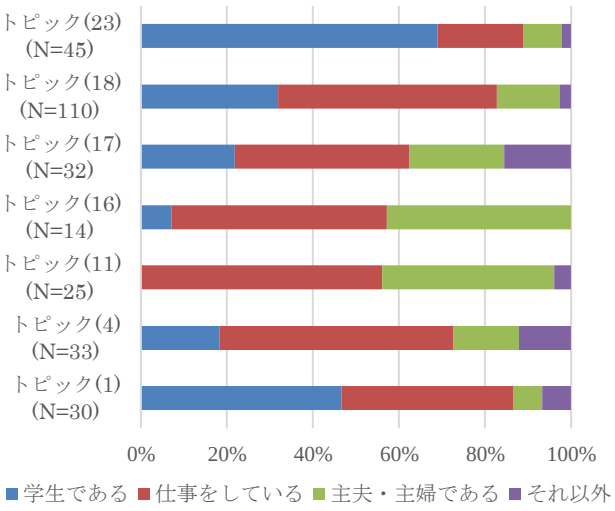


図-19　トピック別の就業状況分布

図-22はトピック (15) を示す．このトピックに関しても，LDAでも同様のトピックが抽出されている．このトピックには，比較的に高い確率で桜の馬場城彩苑を含むメッシュが出現している．桜の馬場城彩苑は，熊本城付近に立地する熊本城と連動した観光施設である．熊本城を含む，上記のトピック(11) と同じく，観光目的の来街者や熊本県外居住者の回遊であることが予想される．

ここで，LDAでの分析と同様に，トピックと個人・トリップ属性の関係を分析して，これらのトピックが観光に関するトピックか確認する．なお，LDAでは優勢トピックkに対して$\theta_{d,k}$> 0.8となるサンプルに限定して取り上げたが，ここでは，すべてのサンプルを用いることとする．これは，上記のトピックに関して，$\theta_{d,k}$>0.8となるサンプルが極めて少なかったためである．

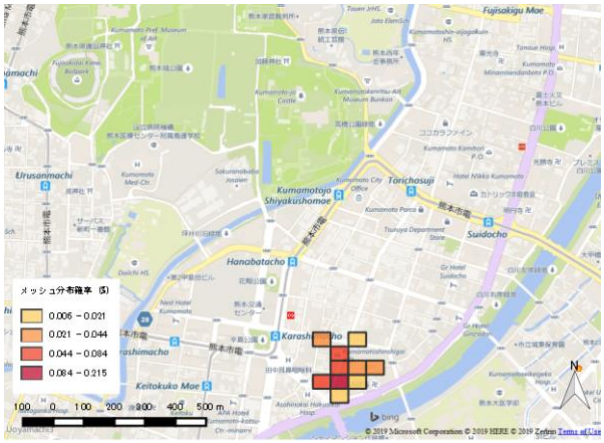


図-20　CTMで抽出されたトピック (1)

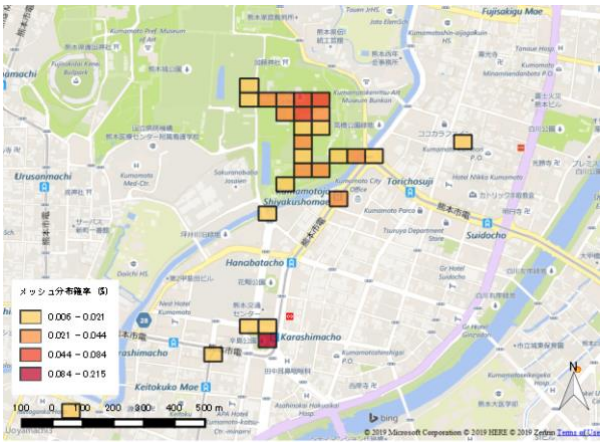


図-21　CTMで抽出されたトピック (11)

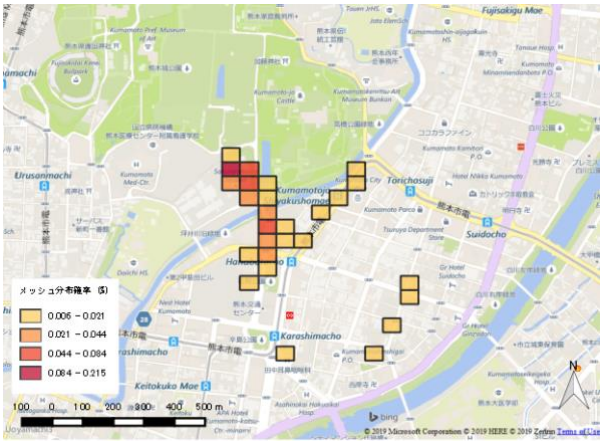


図-22　CTMで抽出されたトピック (15)

図-23にトピック別の来街目的を示した．LDAの結果や既往研究の結果から予想したように，上記の3トピックでは観光目的での来街が存在しており，これらが観光に関するトピックといえることがわかる．また，トピック (15) に関しては，食事目的の来街が多いことが特徴的であるが，これも既往研究[11]で示された結果と一致している．

b)　トピック（回遊パターン）間の相関分析

先述したように，LDA，及びCTMでは，サンプルが

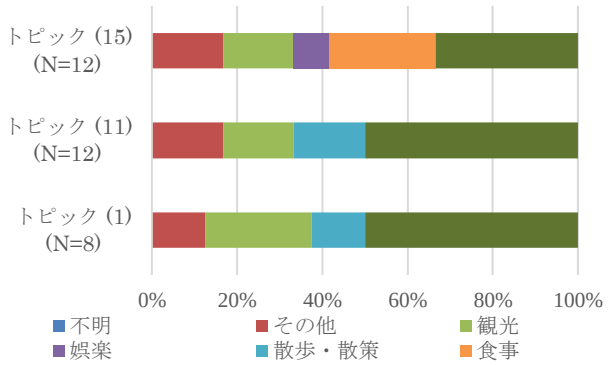


図-23　トピック別の来街目的分布 (CTM)

複数トピックで構成されることを許容している．さらにCTMではトピック間の相関を考慮可能である，すなわち，これを分析することは，来街者が訪問するエリアの組み合わせを分析するということである．

結果より，トピック (11) とトピック (15) が，1%水準で共に訪問されやすいことが統計的に有意に確認できた ($p = 0.0099$)．一方で，トピック (1) に関しては，他の2トピックとの統計的に有意な相関は見られなかった．この結果から，熊本城と桜の馬場城彩苑は1つのツアーで訪問されやすい観光地であるのに対し，シャワー通りは他の観光地とは独立した観光地であると推察できる．

6.　Supervised LDAを用いた個人属性推定の試み

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個人属性推定を試みる．ここでは，トピックと個人属性の結び付けが可能なsLDAを用いる．sLDAでは抽出したトピックを説明変数とした線形回帰で個人属性を予測する．

始めに，680サンプルを学習用データとテスト用データに分割する．680サンプル中，8割に当たる543サンプルを学習用データとし，2割に当たる137サンプルをテストデータとする．すなわち，まず，543サンプルの学習用データの位置情報と個人属性を用いてモデルを推定する．次に，サンプルの137テストデータの位置情報のみを推定されたモデルに入力し，個人属性を推定する．そして，推定された個人属性について議論することで位置情報データの個人属性推定におけるsLDAの有用性を確認する．なお，sLDAの補助情報として，性別や年齢層といった2値離散型の個人属性を用いることとする．また，トピック数はk=25として推定を行った．

(1)　性別

表-2に性別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個人属性の表現している．ここでは，補助情報の設定として，男性を1，女性を0としているため，パラメータの推定値が1に近いほど男性に関連するトピックとなり，0 に近づくほど女性に関するトピックとなる．パラメータの大小をより解釈しやすいように，**表-3**の推定表を**図-24**のように図示した．具体的には，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-24**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5やトピック16は男性に関連するトピックであり，一方で，トピック8やトピック14は女性に関するトピックである． t値を見ると，トピック5のみで，5%水準での有意差が確認できた．しかし，他のトピックでは，パラメータの推定値に大きな差が見られなかった．**図-25**には，**表-2**，**図-24**で推定したモデルを用いて，性別のラベル（男性：1，女性：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．若干男性の分布のほうが1に近いが，予測値

表-2　パラメータの推定結果（性別）

説明変数	パラメータ	標準誤差	t 値	
トピック 1	0.440	0.143	3.073	***
トピック 2	0.215	0.278	0.774	
トピック 3	0.653	0.503	1.297	
トピック 4	0.878	0.268	3.283	***
トピック 5	1.226	0.621	1.973	**
トピック 6	0.371	0.057	6.558	***
トピック 7	0.570	0.381	1.497	
トピック 8	-0.054	0.333	-0.163	
トピック 9	0.362	0.219	1.654	*
トピック 10	0.498	0.134	3.713	***
トピック 11	-0.029	0.332	-0.086	
トピック 12	0.789	0.233	3.389	***
トピック 13	0.447	0.188	2.381	**
トピック 14	-0.076	0.472	-0.160	
トピック 15	0.149	0.175	0.848	
トピック 16	0.997	1.188	0.839	
トピック 17	0.118	0.519	0.228	
トピック 18	0.386	0.230	1.676	
トピック 19	0.210	0.322	0.652	
トピック 20	0.515	0.390	1.321	
トピック 21	0.007	0.396	0.018	
トピック 22	0.275	0.754	0.365	
トピック 23	0.325	0.249	1.306	
トピック 24	-0.034	0.323	-0.106	
トピック 25	0.473	0.304	1.555	
修正済みρ^2		0.339		
サンプルサイズ		543		

注)***1%有意，**5%有意，*10%有意

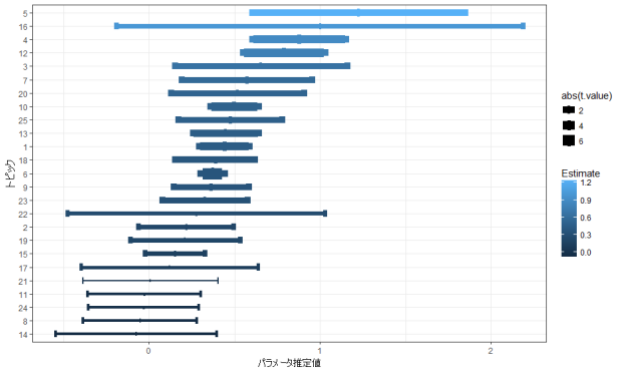


図-24　パラメータ推定結果（性別）

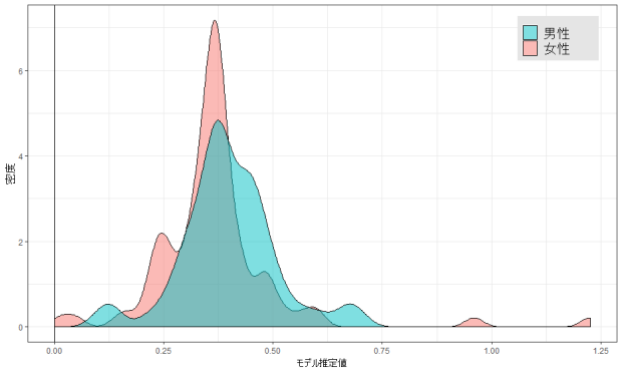


図-25　予測値の密度分布（性別）

が密となる部分は性別で大きな違いはない．ここから，**表-2**，**図-24**で推定したモデルでは，男性と女性の判別は極めて困難であるということが分かった．

(2)　年齢（65歳以上）

図-26に，65歳以上，または65歳未満のラベル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．65歳以上を1，65歳未満を0としているので，パラメータの推定値が1に近づくほど，65歳以上に関連し，0に近づくほど，65歳未満に関連するトピックである．性別を補助情報とした図と比較して，トピック間で推定値にばらつきがあるのが特徴である．一方で，標準誤差は大きくなっている．t値を見ると，トピック10，トピック11，トピック4，トピック12等の推定値が1に近いトピックで1%水準での有意差が確認された（トピック10：t=6.184，トピック11：t = 3.625，トピック4：t = 3.067，トピック12：t = 2.889）．ここでトピック10とトピック11のメッシュ分布確率を確認する．トピック10を**図-27**に，トピック11を**図-28**に示した．トピック10は交通センターを中心に桜の馬場城彩苑や調査時に県民百貨店が立地していたメッシュが含まれている．トピック11は新市街に関するトピックである．熊本市電の辛島町駅を含むメッシュも確率が高くなっているため，市電での来街が考

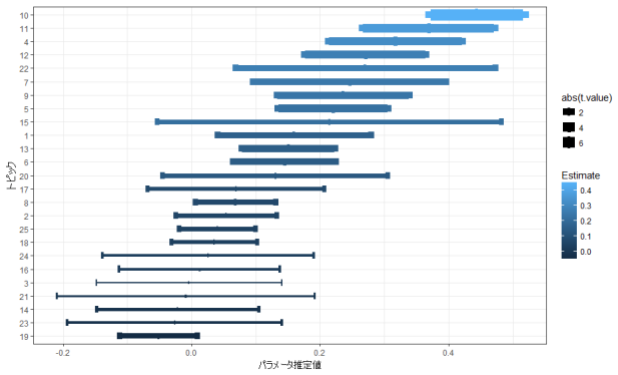


図-26　パラメータの推定結果（年齢：65歳以上）

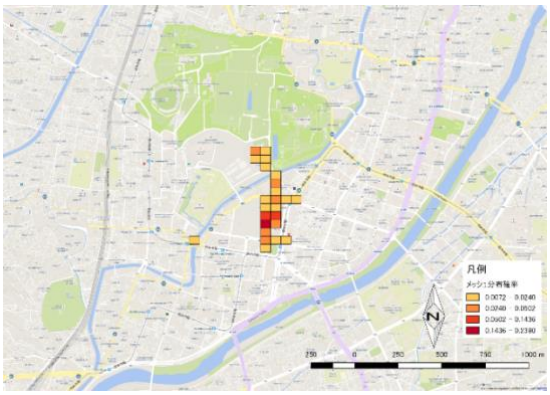


図-27　トピック(10)のメッシュ分布確率（年齢：65歳以上）

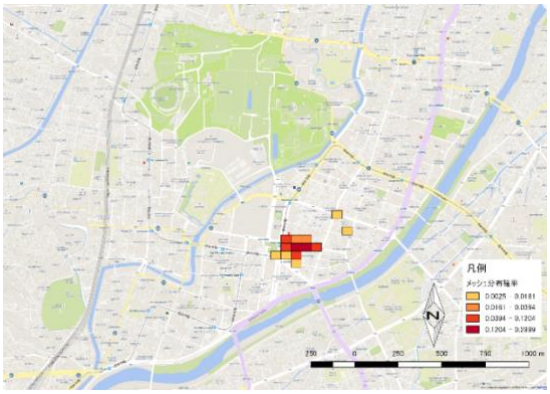


図-28　トピック(11)のメッシュ分布確率（年齢：65歳以上）

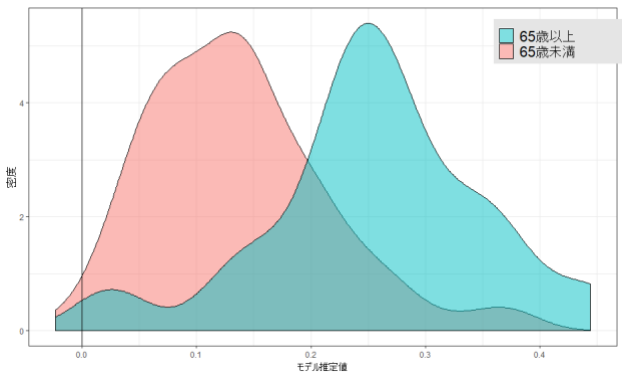


図-29　予測値の密度分布（年齢：65歳以上）

えられる．上記2つのトピックは，共に交通センター周辺に分布するトピックである．交通センター周辺に高齢者向けの施設を立地させることで，高齢者の歩行距離を低減させ，来街の満足度を拡大できる可能性がある．**図-29**には，推定したモデルを用い，65歳以上，または65歳未満のラベルの予測，その予測値の密度分布を示した．65歳以上と65歳未満の間で，予測値が密になる区間が異なることがわかる．両ラベルが重なっている0.2周辺では，適切な判別が困難ではあるが，0.1以下，0.3以上の区間では，適切な判別が可能と考えられる．

(3)　職業（学生）

図-30に学生か否かのラベルを補助情報としたときのトピック別パラメータ推定結果を示す．1が学生，0がその他となっている．比較的1に近いトピック20やトピック4，トピック22などは，1%水準以上での優位さが確認されている（トピック20：t = 5.873，トピック4：t = 3.291，トピック22：t = 4.028）．負の方向にあるトピック17は標準誤差が大きく，統計的に有意な傾向はなかった（トピック17：t = -0.186）．その他のトピックのパラメータ推定値には，トピック間で大きな違いは見られなかった．**図-31**は予測値の密度分布である．職業別に予測値の分布が密になる部分は若干の違いが見られるも，大部分が

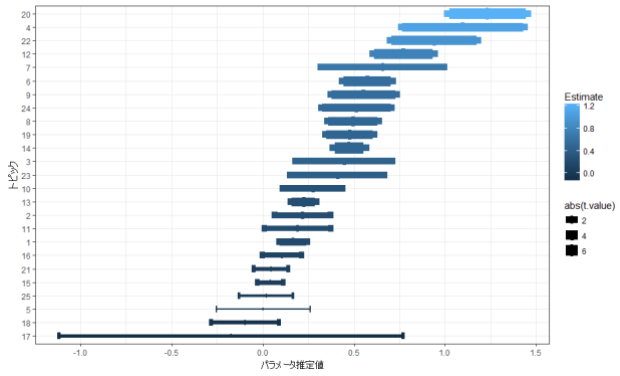


図-30　パラメータの推定結果（職業：学生）

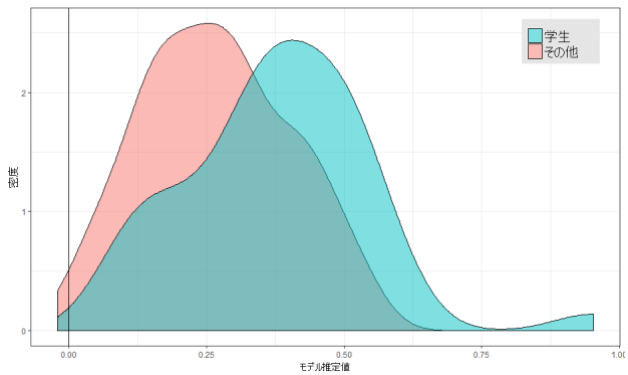


図-31　予測値の密度分布（職業：学生）

重なっているため，これらを判別するのは困難と考えられる．

(4) 来街目的（観光）

図-32に観光目的で来街したか否かを示すラベルを補助情報としたときのトピック別パラメータ推定結果を示す．1が観光目的の来街を示すラベルであるので，トピック4，トピック9，トピック6は観光に関するトピックであると考えられる．その他トピックは全て推定値0.0周辺に集中しているため，上記の3トピックは非常に特徴的である．また，この3トピックのパラメータ推定値は1%水準で統計的に有意との結果になっている．（トピック4：t=6.901，トピック9：t=7.877，トピック6：t=3.880）．ここで，トピック4のメッシュ分布確率を図-33に示す．このトピックは熊本城を含むトピックであり，観光目的での来街という解釈に沿う結果となった．図-34は予測値の密度分布である．両ラベル共に，予測値は0方向に偏ってはいるが，重なりがほぼ存在せず，非常に高い精度で判別出来ることが予想される．

(5) 居住地

図-35に居住地を補助情報としたときのトピック別パラメータ推定結果を示す．トピック10のパラメータ推定値は10%水準で統計的に有意との結果になっており（トピック10：t=1.908），その他全てのトピックでは，1%水準で有意との結果を得た．パラメータ推定値を見ると，トピック20の推定値が相対的に小さくなっている．これらトピックは，熊本県外居住者に多いトピックであることが推察される．その他トピックに関しては，標準誤差も考慮すると推定値に大きな差が見られなかった．これは，サンプル中に熊本県内居住者が多いためと考えられる．

図-36にトピック10のメッシュ分布確率を示す．熊本城含むトピックである．LDAの結果からも，熊本城を含むトピックは，熊本県外居住者に多いことが示されている．sLDAからも同様の結果が得られることとなった．

図-37にはモデル予測値を示す．熊本県内居住者の予測値はそのほとんどが1.0周辺に集中しているのに対し，熊本県外居住者の予測値分布は裾が広く，特に，パラメータが比較的小さいサンプルも存在している．これは，パラメータの予測値が小さい場合，高い確率で熊本県外居住者と予測できることを示している．しかし，熊本県外居住者の予測値は1.0周辺にも広がっているため，予測値が大きい場合は，熊本県外居住者と熊本県内居住者を判別することが困難であることもわかる．

ここまでの結果をまとめると，高齢者と観光目的の来街は比較的高い精度で判別出来ることが明らかとなった．一方で，性別等は判別が困難であることも分かった．

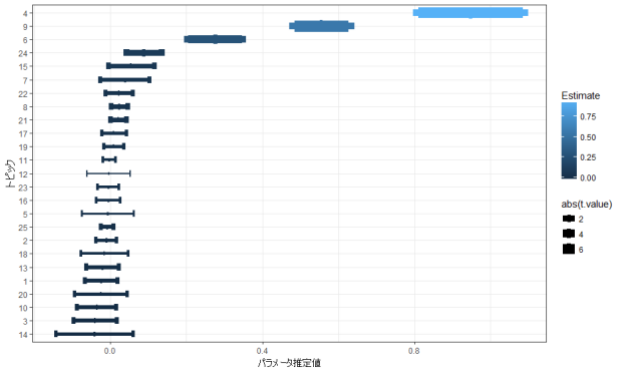


図-32　パラメータの推定結果（来街目的：観光）

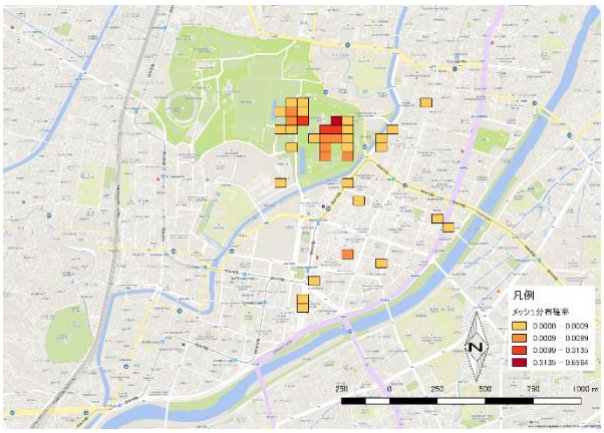


図-33　トピック(4)のメッシュ分布確率（来街目的：観光）

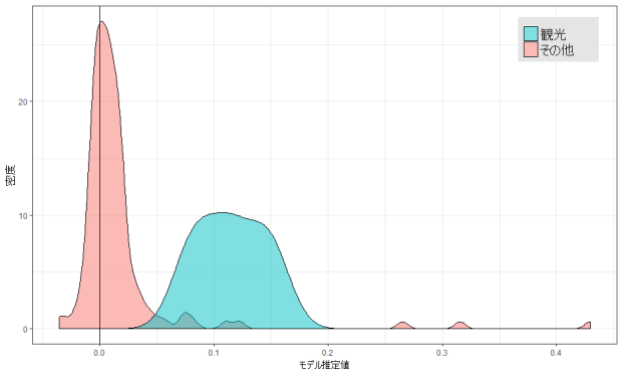


図-34　予測値の密度分布（来街目的：観光）

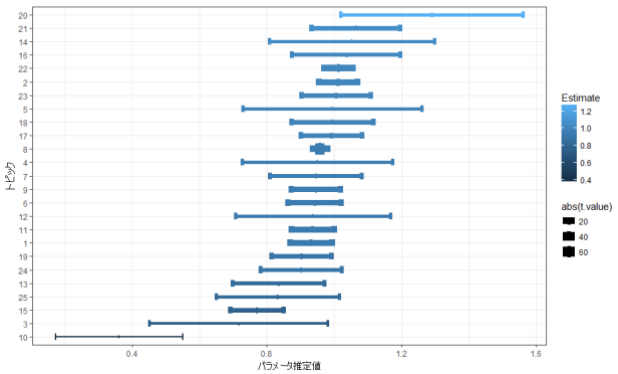


図-35　パラメータの推定結果（居住地）

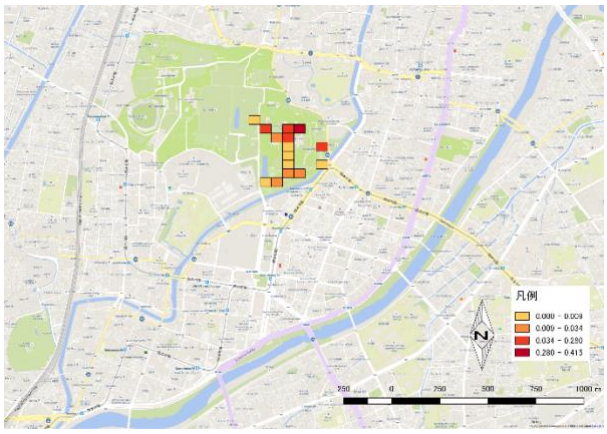


図-36　トピック(4)のメッシュ分布確率（居住地）

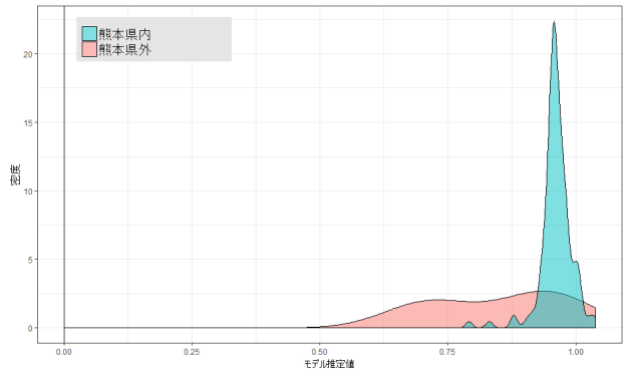


図-37　予測値の密度分布（居住地）

本研究では，教師あり学習の枠組みで個人属性推定を行った．しかし，個人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個人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個人属性を取得し，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個人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．個人属性の取得に関して，先に述べた通り，詳細な個人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極力少数で，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，個人属性予測精度の視点から，どの個人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，個人属性の正解率を直接的に算出できていないこと等の問題があるが，これは今後，別稿での課題としたい．

7.　結論

本研究では，熊本市都心部で実施されたスマホアプリ型回遊調査のデータを対象にトピックモデルを用いた分析を行った．具体的には，テキスト分析における文書と単語のそれぞれをサンプルの移動軌跡とメッシュと捉えてトピックモデルを適用し，回遊行動パターンの抽出と，個人属性の推定を試みた．本研究から得られた成果を以下にまとめる．

1) LDAを用いて，25のトピック (回遊行動パターン) を抽出した．抽出されたトピックは実際の移動軌跡の滞在，回遊を概ね適切に表現しており，LDAの回遊パターン抽出手法としての有用性と，抽出されたトピックの回遊行動パターンとしての妥当性を確認できた．
2) 移動状態と滞在状態が混在する移動軌跡データでも，適切に事前処理を行うことで，滞在状態や低速の移動状態をトピックとして抽出できることを示した．
3) 抽出されたトピックと個人・トリップ属性の関係を分析すると，抽出されたトピックがそれぞれ異なる性質を有しており，その性質は既存手法のカーネル密度分布から得られるものと概ね一致することが分かった．しかし，空間的に類似するが，異なる性質を持つパターンが抽出され，密度ベースの分析手法では得られない結果を得ることが出来た．
4) CTMを用いてトピックを抽出し，トピック間の相関を分析した．この分析より，トピックを回遊エリアと捉えた時に，回遊エリアの組み合わせを発見できることを示した．具体的には，観光回遊に関して，熊本城と桜の馬場城彩苑は同時に訪問される傾向にあるのに対し，シャワー通りはそれらと独立した観光地である可能性を示した．
5) sLDAを用いて，教師あり学習の枠組みで，回遊行動データから個人属性の抽出を試みた．個人属性間で予測値の密度分布を比較し，個人属性間で予測精度に違いがある可能性を示した．具体的には，性別の判定は困難であるのに対し，高齢者や居住地，観光目的の来街の判別は比較的容易であることが分かった．

ここで，今後の課題を以下に述べる．本研究では，トピックモデルの回遊行動データへの適用にあたって，メッシュサイズやトピック数をアドホックに設定している．しかし，これらについては合理的な設定指針を検討する必要がある．また，本研究で用いた手法は，施設や道路が整備された後の回遊行動を評価することはできない．このような事業の影響を考慮して，事業後の回遊パターンを予測するような手法の検討も望まれる．

個人属性の推定に関しては，先述したように，正解率を算出することや他の手法と正解率を比較することが求

められる．また，今回用いたsLDAは線形の回帰モデルで補助情報を予測するため，本来，二値の離散値の予測には不適である．また，sLDAは同時に複数の個人属性を扱うことは不可能である．この2点を克服するモデルも既に提案されているので，今後はそのようなモデルの適用も期待される．

参考文献

1) 国土交通省：中心市街地再生のためのまちづくりのあり方について [アドバイザリー会議報告書]，http://www.mlit.go.jp/kisha/kisha05/04/040810/02.pdf. (2019 年 1 月閲覧)
2) 熊本市：熊本市中心市街地活性化基本計画(熊本地区), http://www.city.kumamoto.jp/common/UploadFileDsp.aspx?c_id=5&id=806&sub_id=15&flid=152790. (2019 年 1 月閲覧)
3) 国土交通省都市局都市計画課都市計画調査室：スマート・プランニング実践の手引き～個人単位の行動データに基づく新たな街づくり～ (第 2 版)，http://www.mlit.go.jp/common/001255640.pdf. (2019 年 1 月閲覧)
4) 井澤佳那子，羽藤英二，菊池雅彦，石神孝裕，川名義輝，杉本保男：観測精度の異なるデータを用いた 3 次元経路選択モデルの推計法，第 55 回土木計画学研究発表会・講演集，Vol. 55, 2014.
5) 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.
6) 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.
7) 矢部直人，有馬貴之，岡村祐：GPS を用いた観光行動調査の課題と分析手法の検討，観光科学研究，Vol. 3, pp. 17-30, 2010.
8) 相尚寿：観光研究への位置情報ビッグデータ展開の可能性，観光科学研究，Vol. 7, pp. 11-19, 2014.
9) 佐藤貴大，円山琢也：スマホ・アプリ型回遊調査データによる熊本都心部回遊行動圏の分析，都市計画論文集，Vol. 50-3, pp. 345-351, 2015.
10) 佐藤貴大，円山琢也：カーネル密度推定法を応用したスマホ型回遊調査データの時空間分析，都市計画論文集，Vol. 51-2, pp. 192-199, 2016.
11) 古谷 知之：携帯型位置情報端末を用いた観光行動動態 の時空間データマイニング-箱根地域を事例として，都市計画論文集，Vol. 41, No. 3, pp. 1-6, 2006.
12) 出水浩介，羽藤英二：プローブパーソンデータを用いた移動-活動パターンのバスケット分析，第 30 回土木計画学研究発表会・講演集，Vol. 30, 2004.
13) 遠藤幹大，高橋央亘，浅田拓海，有村幹治：Wi-Fi パケットセンシングによる広域観光圏における時空間周遊行動パターン分析，第 57 回土木計画学研究発表会・講演集，Vol. 57, 2018.
14) 関本義秀：解説：人々の流動データの基礎的な処理・分析手法について，写真測量とリモートセンシング，Vol.52, No. 6, pp. 321-326, 2013.
15) 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.
16) 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.
17) Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S. and Aberer, K.: Semantic trajectories: Mobility data computation and annotation, *ACM Transactions on Intelligent Systems and Technology*, Vol. 4, No. 3, 2013.
18) Feng, T. and Timmermans, H. J.: Transportation mode recognition using GPS and accelerometer data, *Transportation Research Part C: Emerging Technologies*, Vol. 37, pp. 118-130, 2013.
19) Shafique, M. A. and Hato, E.: Use of acceleration data for transportation mode prediction, *Transportation*, Vol. 42, No. 1, pp. 163-188, 2015.
20) 田中優子，上原邦昭：人の行動把握のための教師なし学習による意味情報推定，土木学会論文集 D3 (土木計画学)，Vol. 72, No. 4, pp. 356-367, 2016.
21) Sun, L. and Yin, Y.:Discovering themes and trends in transportation research using topic modeling, *Transportation Research Part C: Emerging Technologies*, Vol. 77, pp. 49-66, 2017.
22) 塚井誠人，椎野創介：討議録に対するトピックモデルの適用，土木学会論文集 D3 (土木計画学)，Vol. 72, No. 5, 2016.
23) 川野倫輝，佐藤嘉洋，円山琢也：トピックモデルと離散連続モデルを用いた自由記述の量的分析法，土木学会論文集 D3 (土木計画学)，Vol. 72, No. 5, pp. I_277-I_284, 2018.
24) 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.
25) 塚井誠人，塚野裕太：トピックモデルによる詳細地理情報分析，土木学会論文集 D3 (土木計画学)，Vol. 74, No. 2, pp. 111-124, 2018.
26) 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.
27) 岩田具治：トピックモデル，講談社，2015.
28) Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent dirichlet allocation, *Journal of Machine Learning Research*, Vol.3, pp. 993-1022, 2003.
29) Blei, D.M. and McAuliffe, J.D.: Correlated Topic Models, *Neural Information Processing Systems*, Vol.21, pp. 121-128, 2007.
30) Blei, D.M. and Lafferty, J.D.: Supervised topic models, *The Annals of Applied Statistics*, Vol.1, No.1, pp. 17-35, 2007.
31) 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.
32) 野原浩大朗，福所誠也，井村祥太朗，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.
33) Nishida, K., Toda, H. and Koike, Y.:Extracting arbitrary-shaped stay regions from geospatial trajectories with outliers and missing points, *In proceedingsof the 8th ACM SIGSPATIAL International Workshop on Computational Transportation Science (IWCTS 2015)*, pp. 1-6. ACM, 2015.
34) 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.
35) 佐藤一誠，奥村学：トピックモデルによる統計的潜在意味解析，コロナ社(自然言語処理シリーズ)，2015.

(2019. 3. 10 受付)

ANALYZING SMARTPHONE-BASED TRAVEL-SURVEY DATA USING EXTENDED TOPIC MODELS

Tomoki KAWANO, Rintaro KIZAKI and Takuya MARUYAMA