

トピックモデルによる熊本大学構内の スマホ型回遊行動調査データの基礎分析

木崎 凜太郎¹・川野 倫輝²・円山 琢也³

¹ 学生会員 熊本大学工学部社会環境工学科 (〒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 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

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

GPS による人の回遊行動データの取得・分析は都市や観光に関する計画を検討する際に有効である。回遊行動データの分析においては、回遊行動パターン抽出が主な関心となることも多く、トピックモデルなど抽出手法の検討も進んでいる。しかし、大学のキャンパス内のような比較的狭い範囲で取得されたデータを対象としてパターン抽出を行う研究は数少なく、トピックモデルについてもこのようなデータに対する適用実績は存在しない。本研究では、2016年に熊本大学黒髪キャンパスで実施された回遊調査のデータにトピックモデルを適用した。抽出された回遊行動パターンの特徴は、トリップ属性の集計結果の解釈と概ね一致しており、その有用性が確認できた。

Key Words : travel behavior survey, GPS, smart phone, topic model, pattern mining

1. はじめに

(1) 研究の背景・目的

大学が抱えるキャンパス内の問題に駐輪場の不足や設置場所の問題があり、乱雑にとめられた自転車は通行を妨げるとともに構内の景観を悪くする原因にもなっている。また、昼食時の食堂の混雑などの問題が挙げられ、これらの解決策を検討する際には、大学内の学生・教職員の行動データの取得が有用と考えられる。近年では、回遊行動データの取得には、スマートフォン(スマホ)のGPSやWi-Fiパケットデータを利用した回遊・周遊調査が注目を集めており、実施例として2013年秋に熊本市都心部で実施されたスマホ型調査がある。佐藤・円山¹⁾²⁾は、この調査を対象に、カーネル密度推定法を用いた回遊圏域の可視化手法を提示している。これらの手法は、移動滞在判別を行うことなく回遊圏域、特に滞在・滞留地点を簡易に推定することを可能にしている。また濱澤³⁾は2016年1月に熊本大学でスマホ型調査を実施し、学生の回遊行動をヒートマップにより可視化し実態把握を行っている。

回遊行動データの分析においては、滞在地点のパター

ン分析が主な関心となることが多い。しかしながら、佐藤・円山¹⁾²⁾の研究においては、回遊行動パターン抽出は行われておらず、濱澤³⁾の研究においても十分な分析は行えていない。このような商業地域や大学キャンパス等の比較的狭い範囲で取得された回遊行動データに対してパターン抽出を適用する研究は、例えば矢部⁴⁾の研究が挙げられるが、適用事例が限られている。従って、十分な知見が得られてはおらず、適用事例の蓄積が求められている。

一方で、広域の観光周遊行動に関しては、観光振興に向けた知見の発見を狙って観光客の観光地訪問パターンを抽出する研究事例は数多く存在する。さらにデータマイニング手法の発展と共にパターン抽出の手法も多様になってきている。これらの手法のひとつにトピックモデルがある。トピックモデルとは、文書データの解析手法として開発された機械学習の一手法であり、教師データなしでデータ中からトピック(潜在変数)を推定することが可能である。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⁵⁾の研究のように回遊行動データへの適用もある。しかし、適用事例が少ないため、他の回遊行動データ、特に先述の通り比

較的狭い範囲で得られた回遊行動データへの適用が望まれる。これに対して川野・円山⁹⁾は熊本都心部で実施された回遊行動データへの適用を試みているが、さらなる実証的な事例の蓄積が求められる。

以上の背景を踏まえ、本研究では2016年1月に熊本大学で実施された回遊行動調査のデータを対象に、トピックモデルを用いた回遊行動パターン抽出とその結果に基づく分析を行う。適用事例の少なかった位置情報、中でも学内という比較的狭い範囲にトピックモデルを適用し、その有用性を検証することを目的とする。

(2) 既存研究のレビュー

本章ではトピックモデルに焦点を当て、既往研究を整理する。トピックモデルは、先述の通り、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 Yin⁷⁾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。また、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⁸⁾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。文書データ以外の解析への応用事例では、神谷ら⁹⁾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¹⁰⁾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⁾が訪日外国人のGPSデータにトピックモデルを適用し、外国人の観光地の訪問パターンを分析している。しかし、1節でも述べた通り、トピックモデルの応用はこのように幅広い一方で、回遊行動データへの適用例は少ない。また、古屋ら⁹⁾の研究は、日本全国が対象であり、比較的狭い範囲への適用は川野・円山⁹⁾の市街地レベルでの適用が挙げられる。しかし大学キャンパスで得られた回遊行動データをトピックモデルに適用した例は筆者の知る限り存在しない。本研究は、大学キャンパスにおける回遊行動データに対する回遊行動パターンの抽出に関する研究の基礎的なものと位置づけられる。

2. 分析方法

(1) トピックモデルの概要

行動パターンの抽出には、トピックモデルの一種であ

るLatent Dirichlet Allocation¹¹⁾ (LDA) を用いる。ここで、LDAの概要について説明する。なお、LDAについての解説は参考文献¹²⁾に基づいて示す。

LDAは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。

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、 $\theta_{d,k}$ ($k=1, 2, \dots, K$)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 $\theta_d = (\theta_{d,1}, \dots, \theta_{d,K})$ 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 ($k=1, 2, \dots, K$)を有している。文書数を D 、文書 d の文章長(総単語数)

$$\theta_d \sim \text{Dir}(\alpha), d=1, \dots, M \quad (1)$$

$$\phi_k \sim \text{Dir}(\beta), k=1, \dots, K \quad (2)$$

を N_d とする。 $\phi_{k,v}$ 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 $\phi_k = (\phi_{k,1}, \dots, \phi_{k,V})$ 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

$$z_{d,i} \sim \text{Multi}(\theta_d), i=1, \dots, N_d \quad (3)$$

$$w_{d,i} \sim \text{Multi}(\phi_{z_{d,i}}), i=1, \dots, N_d \quad (4)$$

のように整理できる。

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以下のように生成される。

(2) 回遊行動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 をトリップチェーン、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-1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GPSログデータ、 D は総トリップチェーン数、 V は分析対象地域内の総メッシュ数である。なお、LDAではトピック(行動パターン)間の独立を前提としている。実際の行動パターン間には空間的な相関が存在すると考えられるが、本研究においては行動パターン間には相関はないものと仮定して分析を行う。

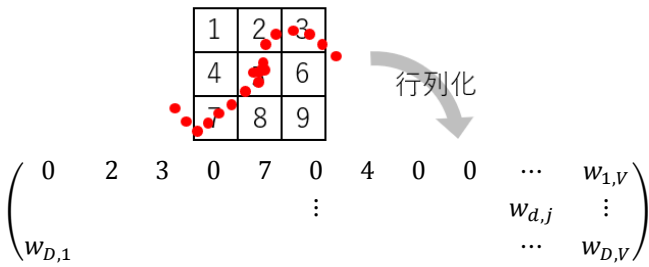


図-1 GPSログデータのBoW表現

3. 分析の対象

分析の対象とする調査概要を表-1に示す。本調査は熊本大学黒髪キャンパスで行われ、原則として19日・20日の2日間に参加するものとした。講義が無い者や、熊本大学黒髪キャンパスへのトリップが無い場合(1日を通して在宅する場合など)は、21日・22日に参加することで、計2日以上参加を依頼した。参加者は2016年1月時点で熊本大学および熊本大学大学院に在籍する学生で、かつAndroidまたはiOS端末を所持する者とした。参加者は専用アプリをインストールし、参加当日の自宅出発から帰宅までの間アプリを起動することで位置情報を取得し、回遊軌跡を記録する。この軌跡には参加者の自宅と対象地への往復経路も含まれるが、プライバシー保護のため除外している。調査で得られる情報として、位置情報に加えアプリ登録時に発行したIDと、アンケート回答時に回収した学部・学年・性別・OS等がある。対象エリアは図-2に示されるエリアで、本研究では10mメッシュに分割した。キャンパスは北側に文系(文学部・法学部・教育学部)の施設が、南側に理系(工学部・理学部)の施設が立地しており、キャンパス内の施設のうち比較的用户の多い施設を列挙している。

調査後の集計結果を表-2、学部別・学年別参加者数を表-3に示す。表-2における参加数(2日間以上)は、「調査期間のうち、2日間以上参加した人数」と定義し、1日間のみの参加や、誤作動等で測位が見られなかった学生は含まれていない。これらの場合を考慮した観測数は、参加数(1日間以上)に含まれている。ID登録数は、調査依頼時にWeb上で利用登録を行い、IDを発行した方の総数である。表-3における無回答は、調査には参加したが、IDが確認できず、属性が不明となったサンプルである。

4. 分析結果と考察

(1) トピック(回遊パターン)の抽出結果

サンプリング方法には、崩壊型ギブスサンプリングを

表-1 調査概要

調査名	熊本大学黒髪キャンパス学内回遊調査
調査日	2016年1月19日(火)～1月22日(金) (※21・22日は予備日)
対象者	熊本大学在学中の学生 かつ AndroidまたはiOS端末を所持する者 (2016年1月現在で在学中の学生)
対象地域	熊本大学黒髪キャンパス・ 熊本大学教育学部附属特別支援学校 およびその周辺
使用端末	参加者個人のスマホまたはタブレット
使用アプリ	交通調査アプリ「スマくま」



図-2 分析対象エリア

表-2 学内回遊調査の各観測数

		男性	女性	計
Android	(台)	51	10	61
iOS	(台)	61	27	88
参加数 (2日間以上)	(人)	112	37	149
参加数 (1日間以上)	(人)	174	50	224
ID登録数	(人)	387	88	475
参加率 (1日間以上)	(%)	45.0	56.8	47.2
実質参加率 (2日間以上)	(%)	28.9	42.0	31.4

表-3 学部別・学年別参加者数一覧

単位: 人	文	教	法	工	理	無回答	計
1年		1		26	3		30
2年		1		42	1		44
3年	4	2	1	35	10		52
4年				17	19		36
修士1年				8			8
修士2年				6	8		14
博士3年					2		2
無回答						38	38
総計	4	4	1	134	43	38	224

用いた。また、ハイパーパラメータは一律に $\alpha=(0.1, \dots, 0.1)$, $\beta=(0.1, \dots, 0.1)$, サンプル数は1000回とし、トピック(回遊行動パターン)数は10とした。図-3から図-12に回遊行動パターンの抽出結果を示す。なお、各メッシュは $\phi_k=(\phi_{k,1}, \dots, \phi_{k,V})$ に基づき色付けされている。 $\phi_{k,v}$ 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、回遊行動パターン k に対するメッシュの寄与率と解釈できる。この寄与率が高いと滞在・滞留、寄与率が低いと移動に近い行動であると考えられる。

図-3では、北キャンパスと南キャンパスを分断する県道337号線から工学部1号館を中心に回遊行動パターンが抽出されている。この回遊パターンは工学部1号館を利用する学生の通学または帰宅であり、工学部の学生が占める割合が高いと推察される。

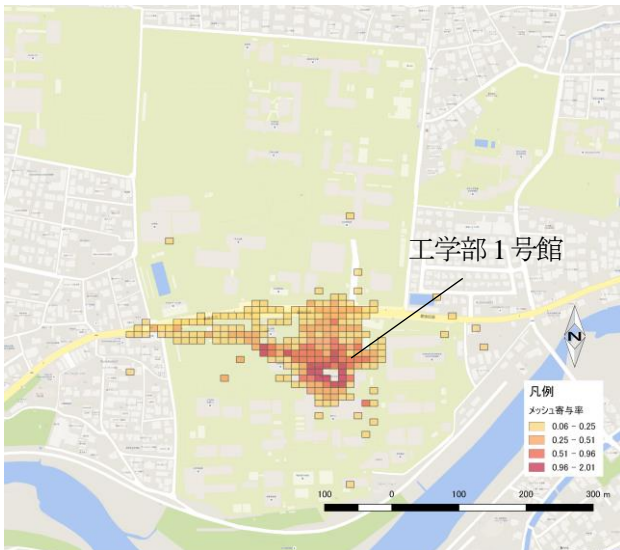


図-3 回遊行動パターン(1)

図-4は、県道337号線から理学部2号館を中心とする回遊行動パターンであり、これは理学部2号館を利用する学生の通学や帰宅であると考えられる。

図-5では、FORICO(食堂)を中心に工学部1号館、工学部2号館周辺と南キャンパス東側の道路においても回遊行動パターンが抽出された。寄与率が高いメッシュはFORICO(食堂)や工学部1号館、工学部2号館周辺のメッシュで工学部の学生が食事や買い物を行うための回遊パターンと推察される。

図-6は、寄与率が高いメッシュが理学部2号館、理学部総合研究実験棟付近と道路を挟んだ学外に見られる。これは調査で測定されたiOS特有の測位誤差であり、実際は上記の施設を利用する理学部の学生による回遊と考えられる。

図-7は、北キャンパスの全学教育棟周辺と南キャンパ

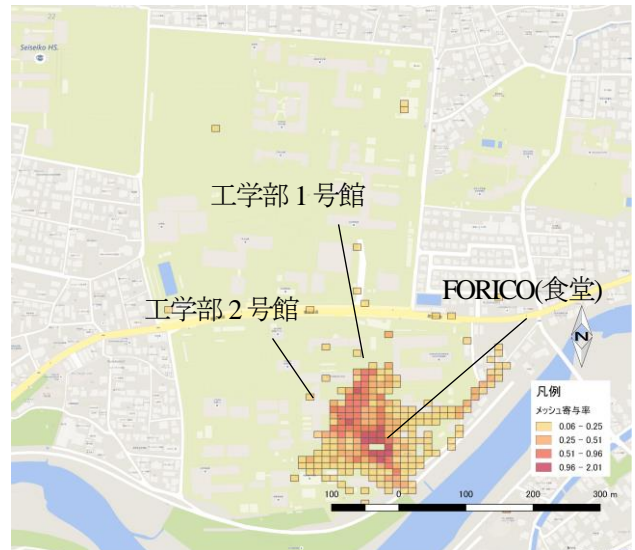


図-5 回遊行動パターン(3)

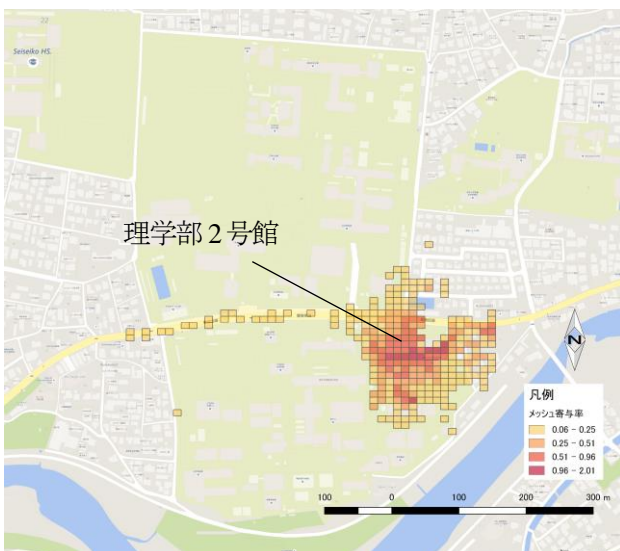


図-4 回遊行動パターン(2)



図-6 回遊行動パターン(4)

スの総合情報統括センター周辺での回遊行動パターンである。寄与率が高いメッシュである全学教育棟，総合情報統括センターは学部に限らず，低学年時の教養，情報の講義で利用されている。よって文系学生や工・理学部低学年の回遊パターンであることが推察される。

図-8は，図書館と学生会館食堂の周辺での滞在が確認できるパターンであり，図書館を利用する学生の食事や買い物による回遊パターンと推察する。また県道 337 号線や南キャンパスから上記の施設への移動と見られる回遊も確認できる。

図-9は，教育学部や文学部・法学部周辺と北キャンパス東側の道路での回遊行動パターンが抽出された。道路上の寄与率が高いメッシュは弁当屋であり，南キャンパスから弁当屋を利用する際に北キャンパス東側の道路を移動したことが考えられる。

図-10は，工学部2号館を中心にFORICO(食堂) や学外

との回遊行動パターンが抽出された。寄与率の高い工学部2号館は主に工学部1～3年生の講義に使われていることから，これらの学生の食事や買い物による回遊パターンと考えられる。

図-11は南キャンパスの南北に広がる回遊行動パターンが抽出された。最も寄与率の高いメッシュは総合研究棟付近のメッシュとなっており，工学部の学生による利用が考えられる。また南側に広がるメッシュに関しては工学部4年生や大学大学院生の実験棟の利用と推察される。

図-12では，南キャンパス西側の道路に沿った回遊行動パターンが抽出されている。これは通学や帰宅といった移動によるものと考えられる。また，学内から学外にかけて寄与率が高いメッシュが横に伸びているが，これは図-6の考察でも述べた iOS 特有の測位誤差であると考えられる。

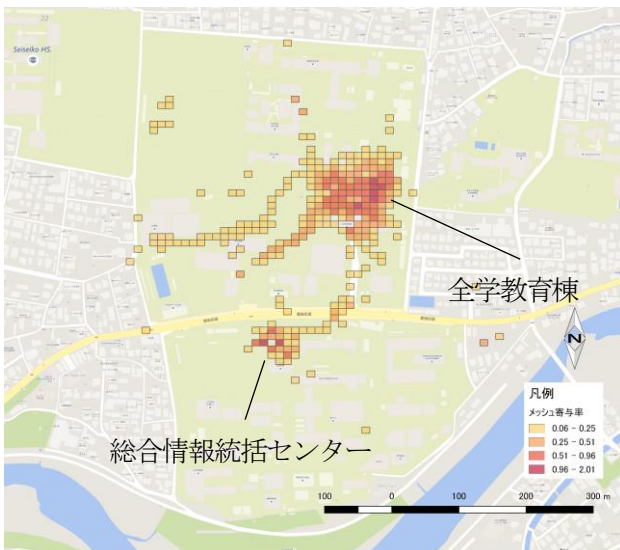


図-7 回遊行動パターン(5)

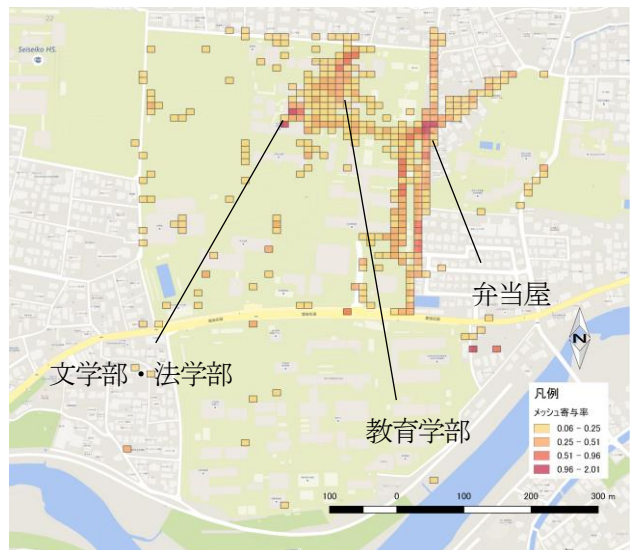


図-9 回遊行動パターン(7)

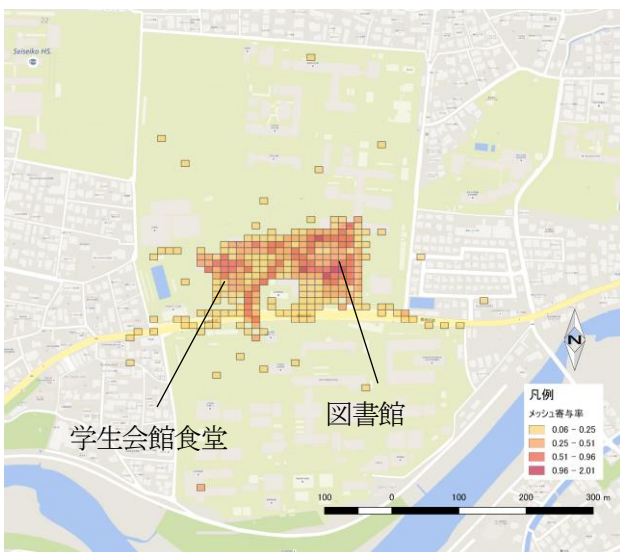


図-8 回遊行動パターン(6)

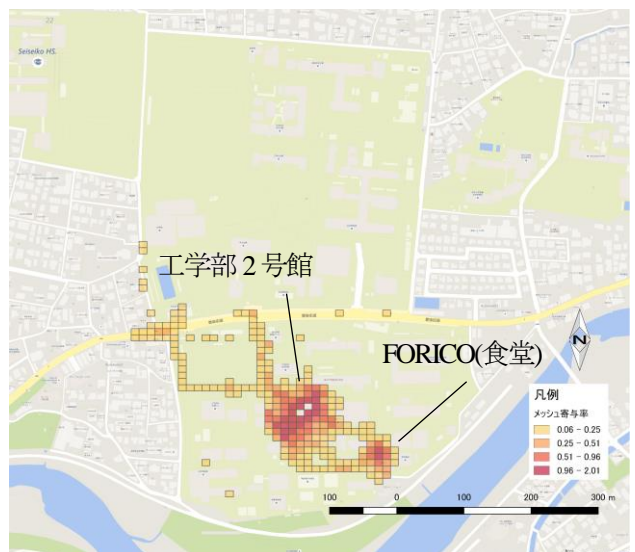


図-10 回遊行動パターン(8)

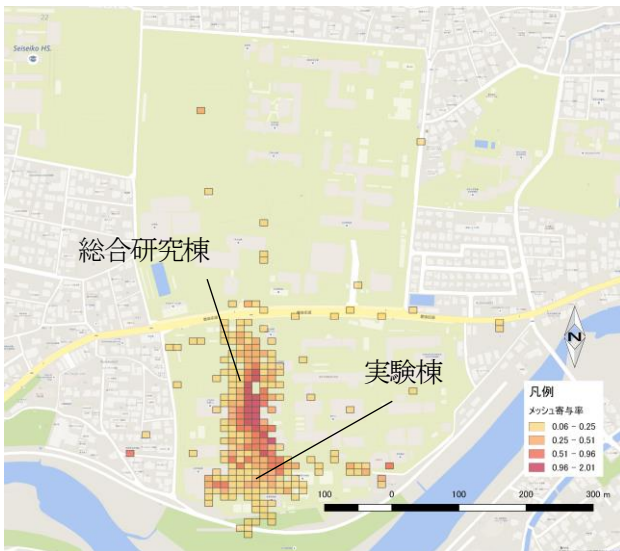


図-11 回遊行動パターン(9)

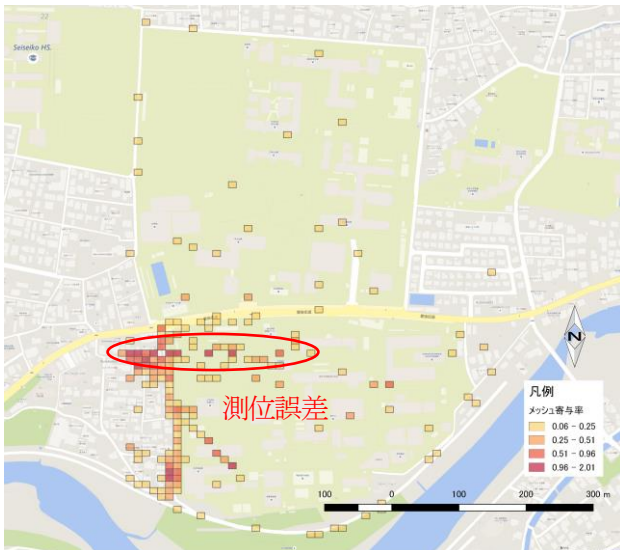


図-12 回遊行動パターン(10)

(2) トピック(回遊パターン)別の属性分析

スマホ型調査の特徴として属性の取得がある。そこで、この属性とトピック(回遊パターン)の関係を分析する。文章 d でトピック k が出現する確率 $\theta_{d,k}$ は、本研究において、サンプル d が回遊行動パターン k に分類される確率と捉えることができる。

図-13には確率 $\theta_{d,k}$ に基づき分析を行った結果を示す。本研究では比較的回遊行動が類似していると想定される学部・学年別の5つのグループに分け分析を行った。文系学生については学部が異なるもののサンプル数が少ないため1つのグループにまとめている。パターン(1)やパターン(3)、パターン(8)、パターン(9)で工学部の学生の割合が高く、パターン(2)とパターン(4)では理学部の学生の割合が高くなっている。これはそれぞれのパターンでメッシュ寄与率が高い施設を日頃利用している学生の割合が高くなっており、考察に対応している。パターン

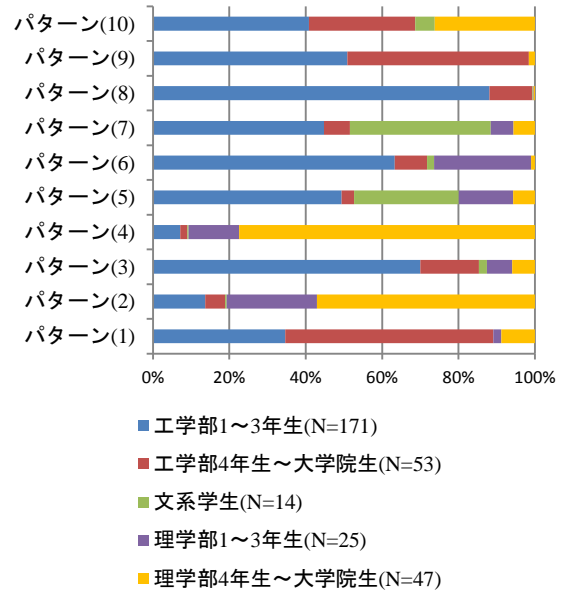


図-13 回遊行動パターン別の学部・学年の構成比率

(5)は工・理学部1~3年生と文系学生の割合が高く、図-7に関する考察と合致している。パターン(6)に関しては工・理学部1~3年生の割合が高くなっている。これは1~3年生がインターネットの利用や自主学习等で図書館に足を運ぶのに対し、4年生~大学院生は研究室に属し図書館を利用することがなくなるためだと考える。

5. おわりに

本研究では、熊本大学黒髪キャンパス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トリップチェーンとメッシュと捉えてトピックモデルを適用し、10の回遊行動パターンを抽出して解釈を行った。この回遊パターンと属性の関係を分析すると、抽出された回遊パターンは属性の集計結果と概ね整合しており、狭い範囲で取得された回遊行動データに対してもトピックモデルの有用性が確認できた。

本研究では属性が判明しているデータを対象に分析を行ったが、トピックモデルに入力した情報はGPSにより得られた位置情報のみである。本研究の結果は、属性が判明していないデータにおいて、モデルより得られたトピック分布 $\theta_{d,k}$ (サンプル d が回遊行動パターン k に分類される確率) より、パターン抽出と同時に個人属性が推察できる可能性を示しているとも考えられる。

今後の課題としては、学部・学年以外の属性情報からの分析や測位誤差の影響の考慮などが挙げられる。属性情報としては通学に利用する交通手段の把握から駐輪場

の適切な設置場所の検討が行えると考えられる。

参考文献

- 1) 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.
- 2) 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.
- 3) 濱澤憲駿: スマホ型調査に基づく熊本大学構内回遊行動分析, 熊本大学大学院自然科学研究科社会環境工学専攻修士論文, 2017.
- 4) 矢部直人: GPS データに対する配列解析の援用, 地理情報システム学会講演論文集, Vol.19, pp.4F-3, 2010.
- 5) 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.
- 6) 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の基礎分析, 第 58 回土木計画学研究発表会秋大会・講演集, 発表予定, 2018
- 7) 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.
- 8) 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.
- 9) 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.
- 10) 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.
- 11) Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent dirichlet allocation, *Journal of Machine Learning Research*, Vol.3, pp. 993-1022, 2003.
- 12) 岩田具治: トピックモデル, 講談社, 2015.

(2018.7.31 受付)

TOPIC MODELING APPROACH TO ANALYZE STUDENTS' ACTIVITIES ON KUMAMOTO UNIVERSITY CAMPUS USING SMARTPHONE-BASED SURVEY DATA

Rintaro KIZAKI, Tomoki KAWANO and Takuya MARUYAMA