

非負値テンソル因子分解 を用いた観光行動データ からの情報抽出

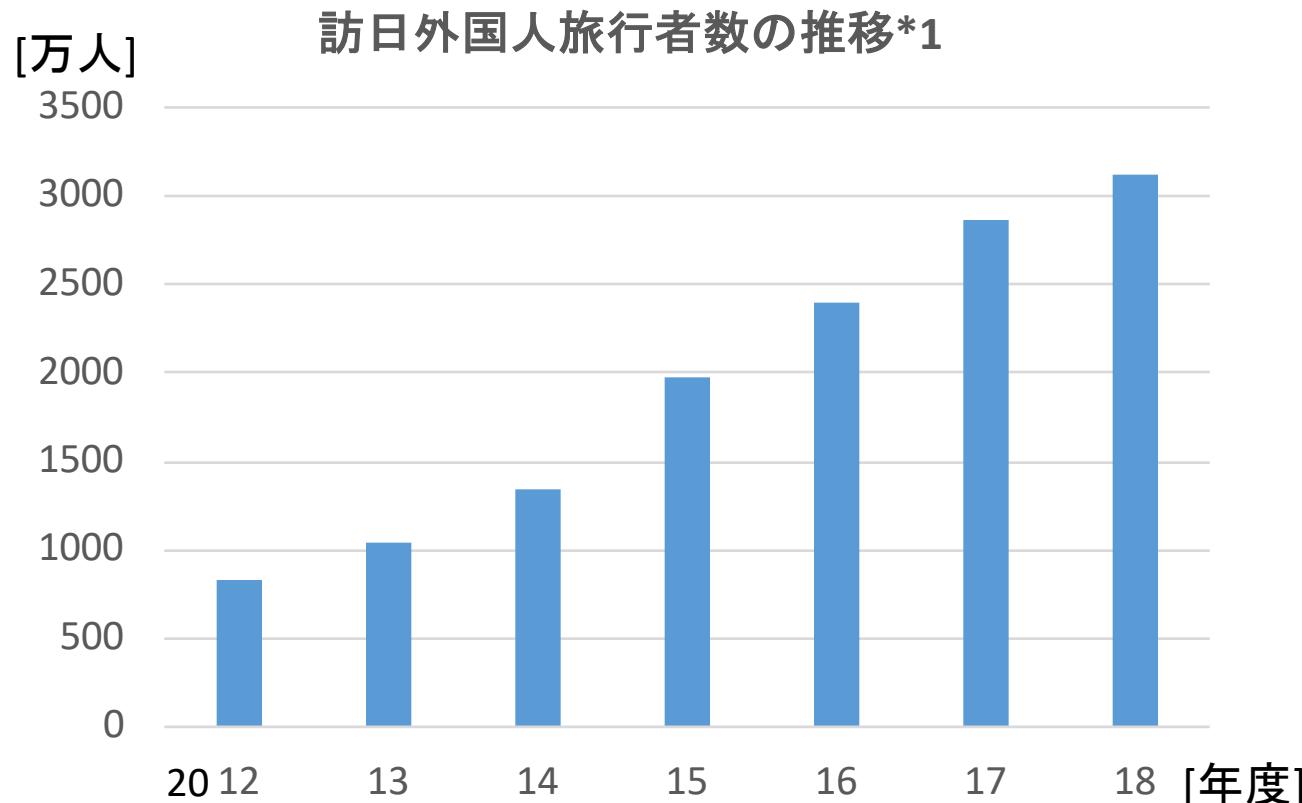
○久保 基^{1, 2}, 田中 宏季^{1, 2, 3}, 中村 哲^{1, 2, 3}

1. 奈良先端科学技術大学院大学
2. 理化学研究所 革新知能統合研究センター 観光情報解析チーム
3. NAISTデータ駆動型サイエンス創造センター

研究背景: インバウンド

●インバウンド: 訪日外国人旅行

- ・年々増加、2020年までに4000万人到達する目標
- ・重要な成長産業として注目



(*1):2018年度 観光白書 (<http://www.mlit.go.jp/common/001294467.pdf>) p.7 より抜粋

研究背景: 観光促進のために

● インバウンドへのプロモーション

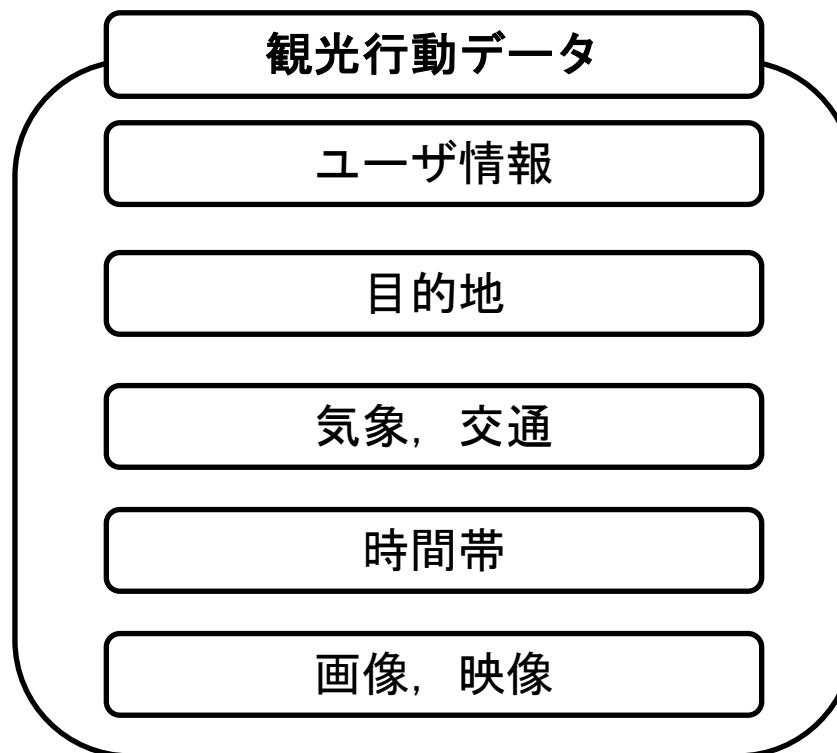
- ユーザがどのような行動を取っているのか?
 - 例: 位置情報から
- ユーザの嗜好を汲み取る必要
 - 例: 行動範囲のスポットから

データを分析し、新たな知見を得ることで観光促進につなげる

観光データの問題

● 観光行動データ間の関係性は複雑かつ高次元

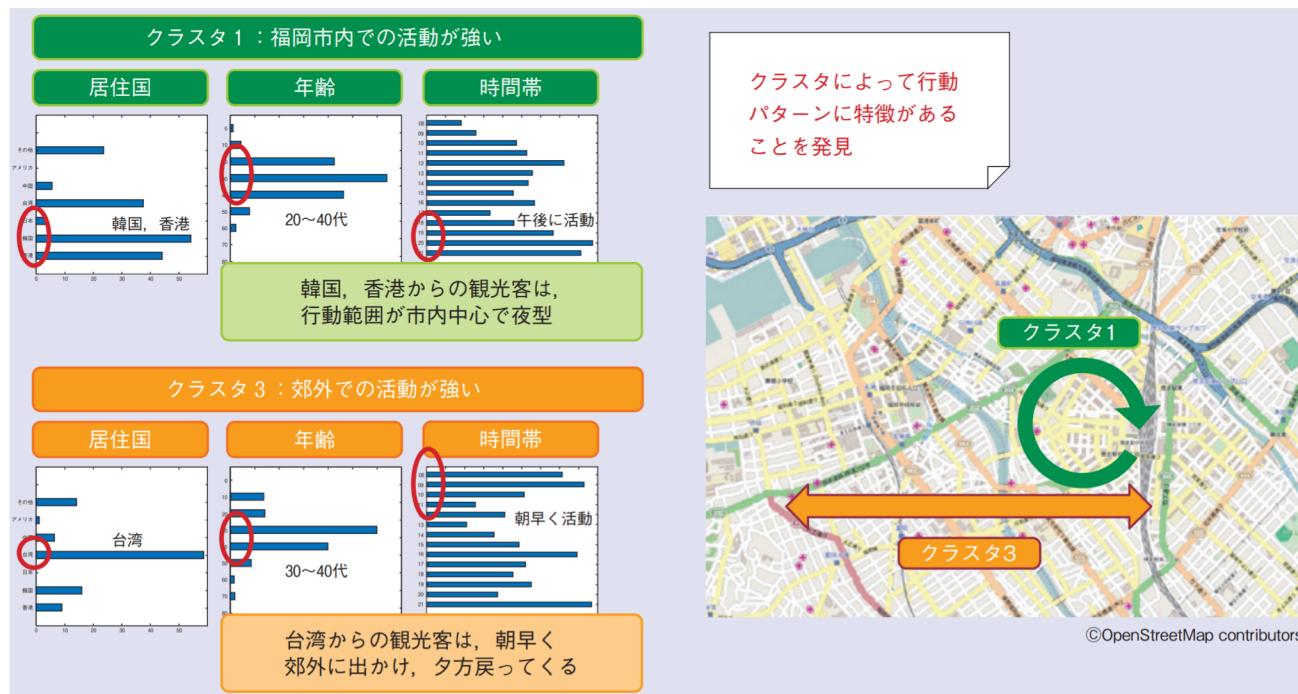
- いつ/誰が/どこで/何を
- 複数項目で交絡しており、扱いが難しい



関連研究

● 福岡の観光客における行動分析 [熊谷他, 2015.]

- 非負値複合テンソル因子分解を用いたアプローチ



ユーザ情報間の関係を抽出 [野口他, 2015]

[熊谷他, 2015] 熊谷雄介, 今井良太, 松林達史, 佐藤吉秀, 堀岡力.

非負値複合テンソル因子分解を用いた訪日外国人観光客の回遊行動分析. IEICE, IEICE 信学技報, 2015.

[野口他, 2015] 野口賢一, 佐藤吉秀, 塩原寿子. 高度高性能ビッグデータ活用技術とトライアル検証.

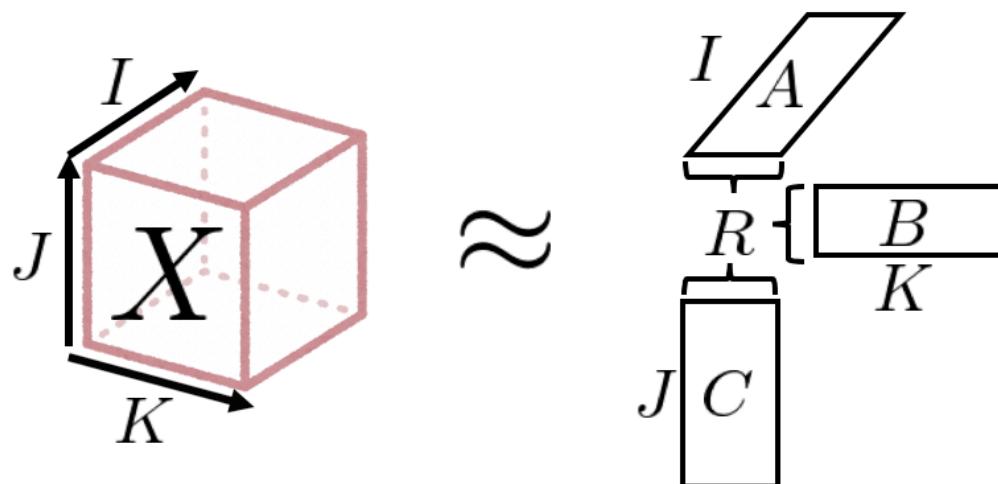
NTT技術ジャーナル, pp.34-38 2015.

非負値テンソル因子分解

[Shashua et al, 2005.]

● データの多項関係を抽出

- テンソルを3つの因子行列に分解することで近似

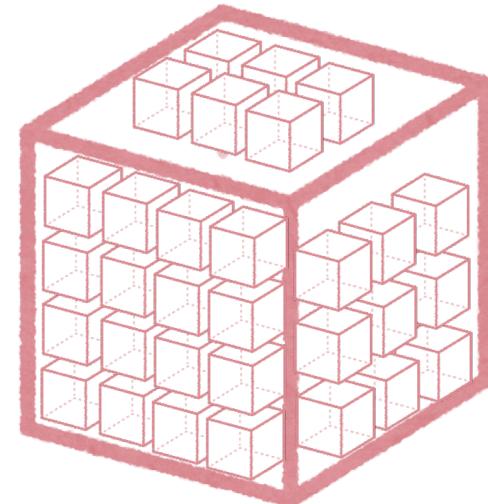
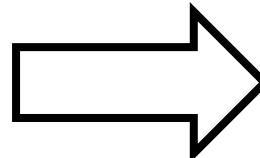
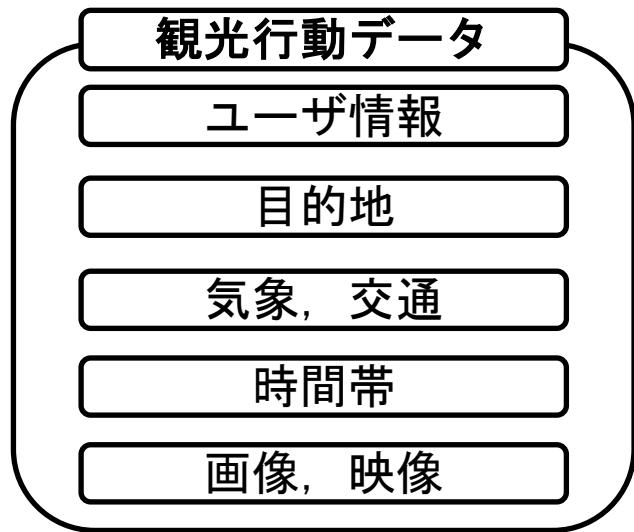


I:因子Aのラベル数
K:因子Bのラベル数
J:因子Cのラベル数
R:基底数

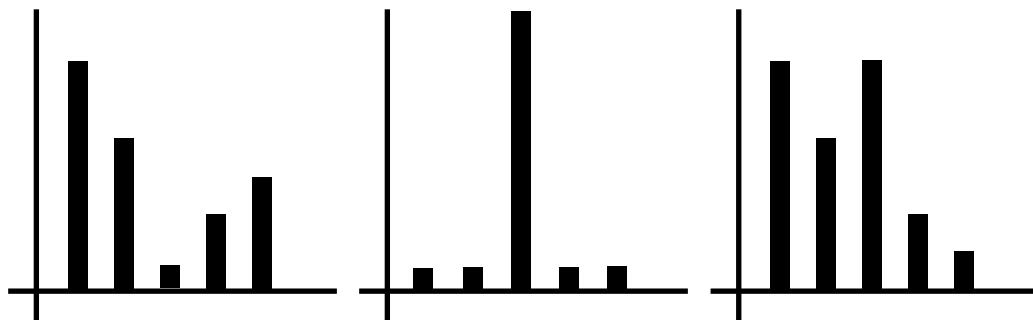
基底ごとの潜在的パターンの抽出

研究目的

- 新規の観光行動データを分析・行動傾向を調査し、新たな知見を得ることを目指す



テンソル



因子間の分析
(テンソル分解)

観光行動データの概要(1/2)

● インバウンド行動データ*

- 対象者: 日本全国を訪れた外国人観光客
- 対象期間: 2-4月, 7-8月, 11月 (2018年)
- カラム(23項目)
 - 時間情報: 年月日/時・分/曜日
 - 位置情報: 100mメッシュID, 緯度経度
 - ユーザ情報: 居住地域/勤務先(日本国内), 推定居住国

*データ提供元: (株) Agoop

観光行動データの概要(2/2)

●今回用いたインバウンド行動データ

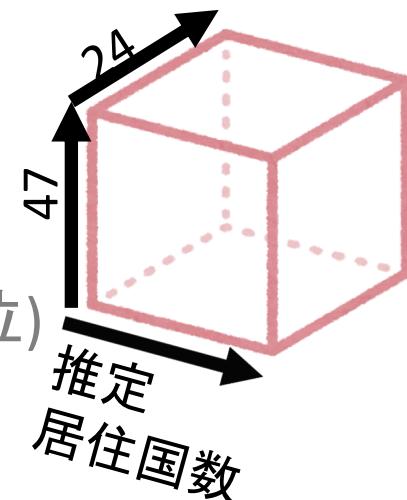
- 対象者:主に関西(京都/大阪)を訪れた外国人観光客
- 対象期間:2018年7月及び8月

7月(ユーザの居住国/目的地) TOP 5 (全ログ数 = 168657件)				8月(ユーザの居住国/目的地) TOP 5(全ログ数 = 217426件)			
居住国	割合[%]	目的地	割合[%]	居住国	割合[%]	目的地	割合[%]
アメリカ	36.4	京都	52.3	イタリア	27.2	京都	53.5
イタリア	7.8	大阪	11.8	アメリカ	18.2	大阪	9.3
スペイン	7.6	東京	6.8	フランス	15.5	東京	6.7
フランス	7.5	静岡	4.3	スペイン	12.1	奈良	4.0
オーストラリア	7.2	愛知	3.8	イギリス	3.5	広島	3.5

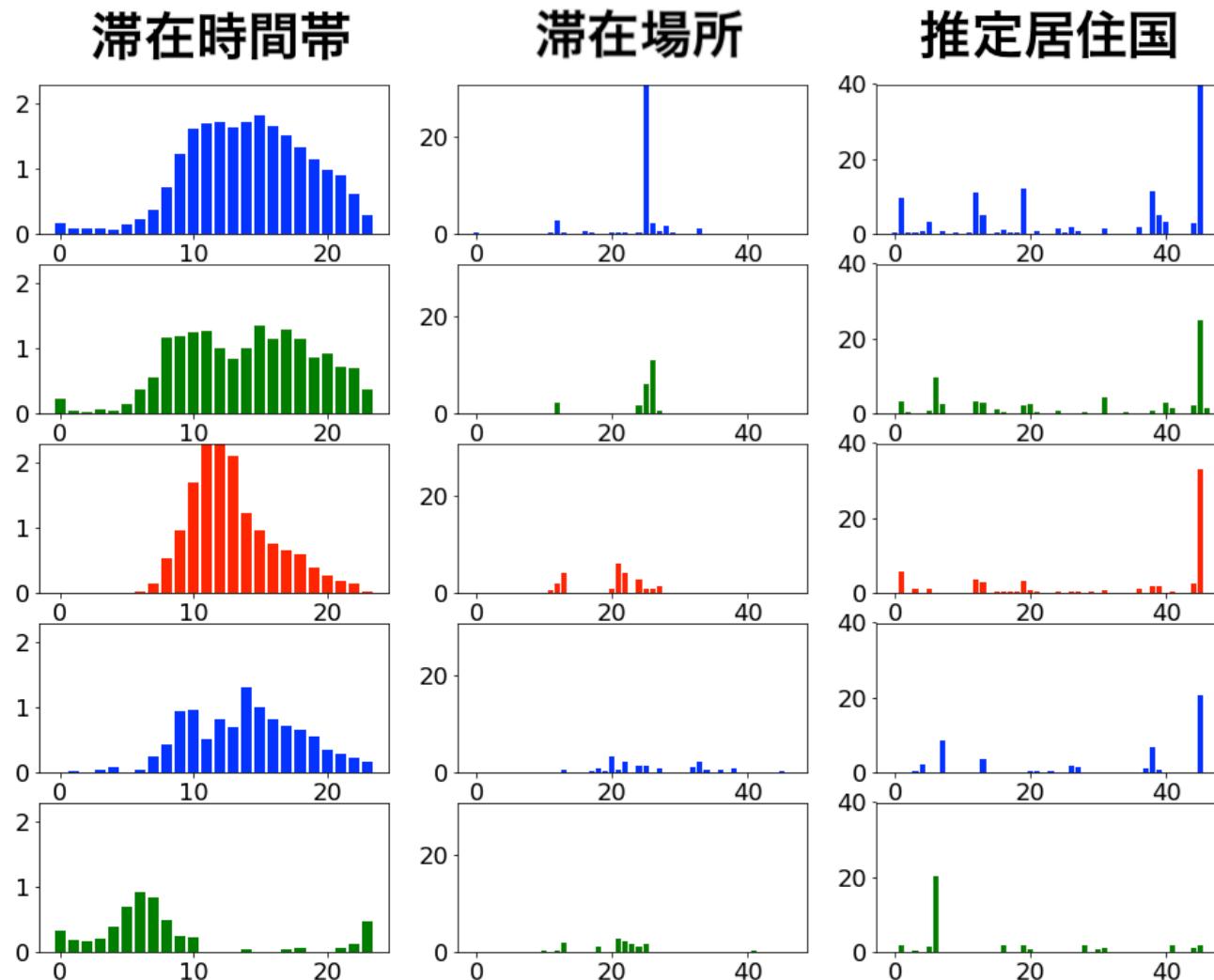
観光行動データの分析

● 概要

- 入力: 3階テンソル {滞在時間帯
滞在場所 (都道府県単位)
推定居住国}
- 滞在時間帯: 24時間幅(0-24時)
- 滞在場所: 北海道から沖縄まで(1~47のIDラベル付与)
- 推定居住国: 国名でアルファベット順にID付与(1, 2 ...)
- 基底数: 5
- 出力: 3つの潜在因子テンソルを基底数分



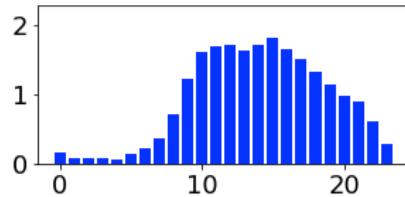
分析結果: イタリア人の行動(7月)



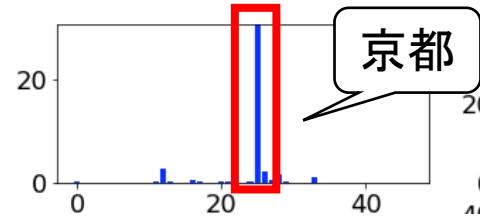
縦軸: ラベルの出現頻度 横軸: 因子の全ラベル

分析結果: イタリア人の行動(7月)

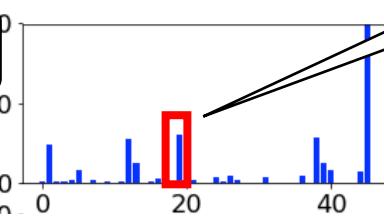
滞在時間帯



滞在場所



推定居住国



飲食店

宿泊施設周辺

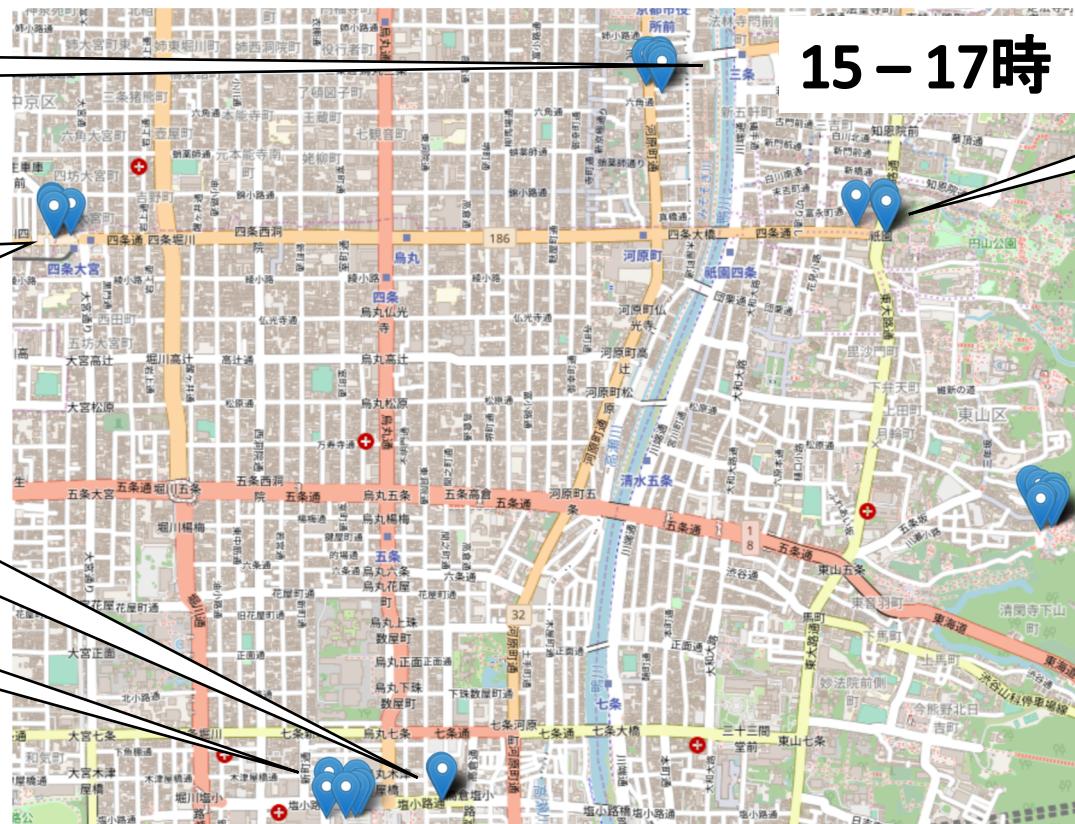
パーキング施設

京都駅前

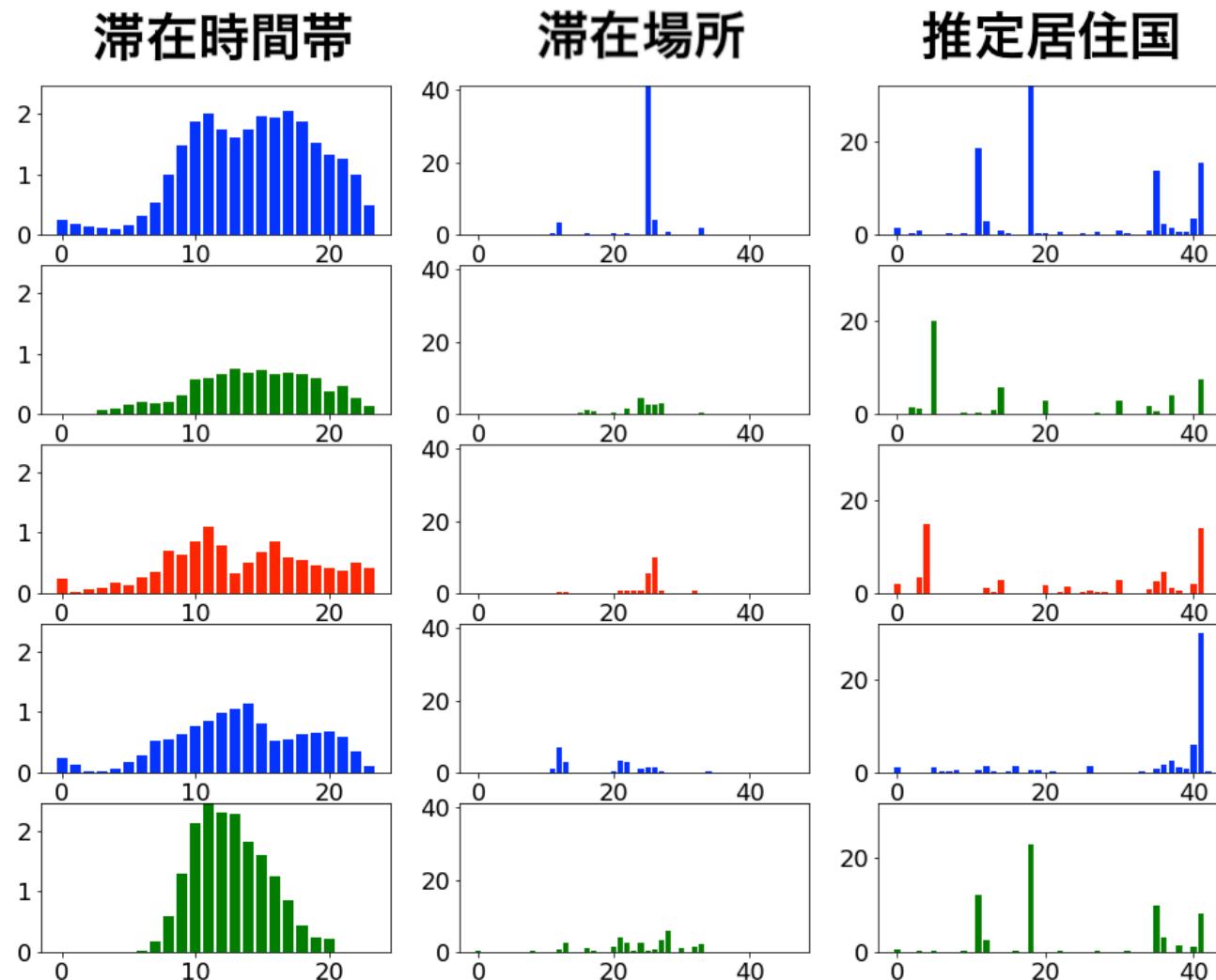
15 - 17時

八阪神社

清水寺周辺



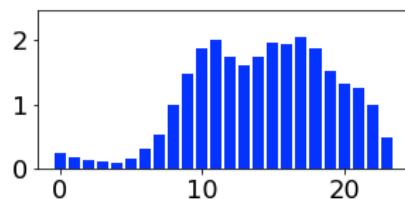
分析結果: イタリア人の行動(8月)



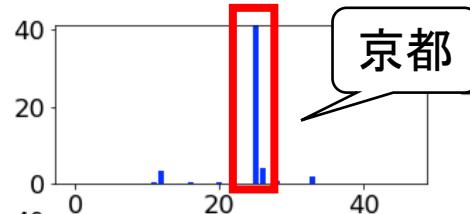
縦軸: ラベルの出現頻度 横軸: 因子の全ラベル

分析結果: イタリア人の行動(8月)

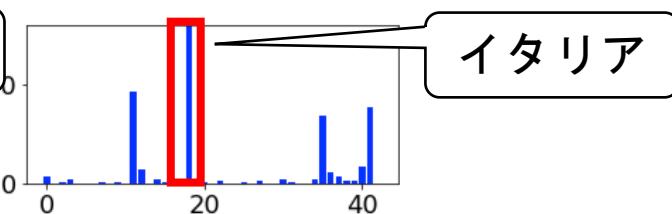
滞在時間帯



滞在場所

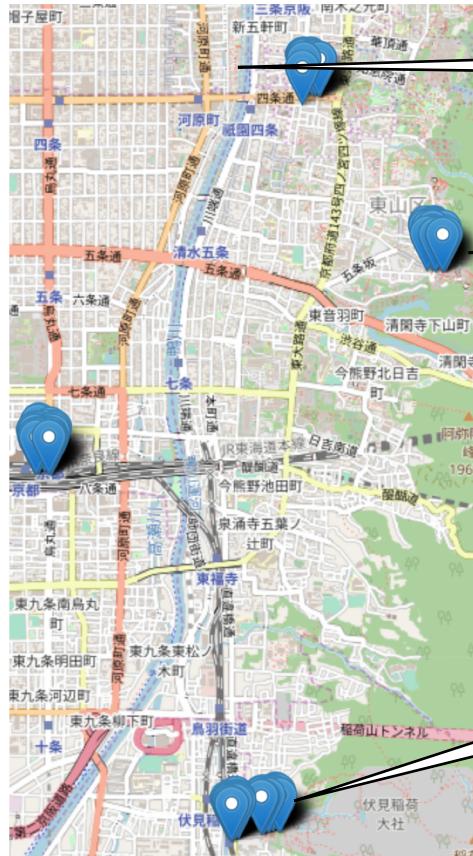


推定居住国



15 – 17時

京都駅
周辺



飲食店

清水寺
周辺

飲食店
伏見稻荷周辺

まとめ

- 非負値テンソル因子分解を用いて新規の観光行動データについてインバウンドの行動分析を行った
- 今後の予定
 - テンソルの次元数を増やし、その他の関係性を調査
 - 適切な基底数を選択する手法を適用
 - スパースなテンソルに対する手法の実装
 - ユーザ情報の追加など詳細な分析