

トピックモデルを用いた訪日外国人周遊分析

広島大学 学生会員 ○辰巳嘉大
 広島大学 正会員 塚井誠人

1. はじめに

国土交通省は、訪日外国人増加に向けた政策立案のために、訪日外国人流動データ（以下 FF データ）¹⁾ を作成した。個人属性に加えてトリップチェーンが記録されている FF データを用いれば、国籍、滞在日数などの各属性のクロス分析や、目的地間の周遊の分析が可能である。ただし、2014~2016 年の FF データのサンプル数は、約 4.3~6.3 万人、総トリップ数は約 17.9~22.8 万と膨大である。このため、周遊パターンやその他の属性の膨大な組み合わせから効率よく代表的なパターンを抽出する手法の開発が必要である。

本研究では、古屋²⁾や柴田³⁾らの既往研究では扱われていない訪日外国人の訪日目的、訪日時期などの旅行特性を扱った FF データを用いる。また、FF データにトピックモデルを適用して、訪日外国人の訪問地、訪日目的などの旅行特性を経年分析する。

2. トピックモデル

トピックモデルは情報検索分野で開発された手法であり、膨大な文書データを内容の特徴に基づいて分類する手法として知られる。トピックモデルは、特異値分解を応用した階層ベイズ型の統計手法である。特異値分解の原理を図 1 に示す。同図に示すように、データ属性の特徴抽出は $K \times V$ 行列より得られ、サンプル分類は $D \times K$ 行列から得られる。これらを同時推定することで、データの潜在的なパターンを効率よく分析できる。

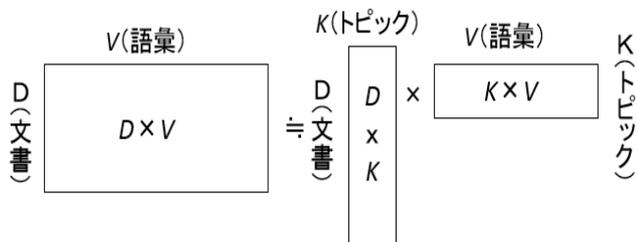


図 1 特異値分解

表 1 本研究で用いた属性

属性	内容/カテゴリ
出国港	新千歳空港, 旭川空港, 函館空港, 青森空港, 仙台空港, 秋田空港
	茨城空港, 羽田空港, 成田空港, 新潟空港, 富山空港, 小松空港, 静岡空港, 中部空港
	関西空港, 米子空港, 岡山空港, 広島空港, 高松空港, 松山空港, 福岡空港
	福岡空港, 佐賀空港, 長崎空港, 熊本空港, 大分空港, 宮崎空港, 鹿児島空港
	那覇空港, 石垣空港, 北九州空港, 博多空港, 下関空港, 厳原空港, その他空港
国籍	韓国, 台湾, 香港, 中国
	タイ, シンガポール, マレーシア, インドネシア, フィリピン, ベトナム
	インド, その他アジア
	イギリス, ドイツ, フランス, イタリア, スペイン, ロシア, その他ヨーロッパ
	アメリカ, カナダ, その他北アメリカ, 南アメリカ
オーストラリア, アフリカ, その他オセアニア, 無国籍	
旅行目的	観光・レジャー, 家族知人の訪問, 業務, 研修・学会等, 留学, 乗り継ぎ, その他, 不明
旅行手配方法	団体旅行, 個人旅行, 不明
訪日経験回数	1回目, 2回目, 3回目, 4回目, 5回目, 6~9回目, 10~19回目, 20回以上, 不明
出発地	47都道府県, 不明
	新千歳空港, 旭川空港, 函館空港, 青森空港, 仙台空港, 秋田空港
	茨城空港, 羽田空港, 成田空港, 新潟空港, 富山空港, 小松空港, 静岡空港, 中部空港
	関西空港, 米子空港, 岡山空港, 広島空港, 高松空港, 松山空港, 福岡空港
	福岡空港, 佐賀空港, 長崎空港, 熊本空港, 大分空港, 宮崎空港, 鹿児島空港
那覇空港, 石垣空港, 北九州空港, 博多空港, 下関空港, 厳原空港, その他空港	
目的地	47都道府県, 不明
	新千歳空港, 旭川空港, 函館空港, 青森空港, 仙台空港, 秋田空港
	茨城空港, 羽田空港, 成田空港, 新潟空港, 富山空港, 小松空港, 静岡空港, 中部空港
	関西空港, 米子空港, 岡山空港, 広島空港, 高松空港, 松山空港, 福岡空港
	福岡空港, 佐賀空港, 長崎空港, 熊本空港, 大分空港, 宮崎空港, 鹿児島空港
那覇空港, 石垣空港, 北九州空港, 博多空港, 下関空港, 厳原空港, その他空港	
トリップ数	サンプルIDごとのトリップ数
訪日時期	1~3月期, 4~6月期, 7~9月期, 10~12月期
滞在日数	0日, 1~2日, 3日, 4日, 5日, 6日, 7日, 8~10日
	11~14日, 15~30日, 31~90日, 91~364日
宿泊数	0~2日, 3~7日, 8日以上
利用交通機関	バス, 鉄道, 国内線飛行機, 自動車
利用交通機関分担率	0~9%, 10~19%, 20~29%, 30~39%, 40~49%, 50~59%
	60~69%, 70~79%, 80~89%, 90~99%
拡大係数	サンプルIDごとの拡大係数

3. トピックの抽出

3.1 入力データの加工手順

トピックモデルの入力データは、文書単位の語彙別カウントデータに限らず、多属性のカウントデータであれば、トピック抽出ができる⁴⁾。本研究では、FF データをトピックモデルに入力する疎構造データ形式 (Bag of Words) に変換するため、宿泊数、滞在日数、利用交通機関分担率の連続量を離散化する。離散化の閾値は、オリジナルデータを様々な閾値で離散化して、

キーワード LDA Topic Model FF-Data

連絡先 〒739-8527 広島市鏡山 1-4-1 広島大学大学院工学研究科 社会基盤環境工学専攻 事務室

TEL : 082-424-7819

表2 トピックと名前 (2014年)

順位	14.8.北海道	14.2.関東1	14.3.関東2	14.5.中部
1位	北海道 (0~2日宿泊) 0.18	東京都 (0~2日宿泊) 0.74	国内線飛行機分担率0~9% 0.07	国内線飛行機分担率0~9% 0.06
2位	鉄道分担率10~19% 0.06	成田空港出国 0.14	東京都 (0~2日宿泊) 0.06	愛知県 (0~2日宿泊) 0.06
3位	バス分担率70~79% 0.04	成田空港入国 0.13	バス分担率20~29% 0.05	鉄道分担率40~49% 0.05
4位	新千歳空港入国 0.03	東京都 (3~7日宿泊) 0.10	成田空港出国 0.05	観光・レジャー 0.04
5位	新千歳空港出国 0.03	自動車分担率10~19% 0.06	成田空港入国 0.05	中部空港出国 0.03
6位	バス分担率60~69% 0.02	神奈川県 (0~2日宿泊) 0.05	自動車分担率10~19% 0.05	中国 0.03
7位	5日滞在 0.02	バス分担率20~29% 0.05	鉄道分担率50~59% 0.05	10~12月期 0.03
8位	台湾 0.02	千葉県 (0~2日宿泊) 0.04	観光・レジャー 0.04	中部空港入国 0.03
9位	北海道 (3~7日宿泊) 0.01	東京都 (8日以上宿泊) 0.03	羽田空港入国 0.03	07~09月期 0.02
10位	5トリップ 0.01	鉄道分担率50~59% 0.02	羽田空港出国 0.03	2トリップ 0.02
11位	自動車分担率0~9% 0.01	鉄道分担率60~69% 0.02	東京都 (3~7日宿泊) 0.03	自動車分担率10~19% 0.02
12位	旭川空港出国 0.01	7トリップ 0.02	04~06月期 0.02	6日滞在 0.02
13位	旭川空港入国 0.01	8トリップ 0.02	自動車分担率0~9% 0.02	自動車分担率20~29% 0.02
14位	函館空港出国 0.01	山梨県 (0~2日宿泊) 0.02	神奈川県 (0~2日宿泊) 0.02	岐阜県 (0~2日宿泊) 0.01
15位	青森県 (0~2日宿泊) 0.01	9トリップ 0.01	01~03月期 0.02	長野県 (0~2日宿泊) 0.01
16位	観光・レジャー 0.01	15~30日滞在 0.01	業務 0.02	バス分担率30~39% 0.01
17位	函館空港入国 0.01	国内線飛行機分担率0~9% 0.01	中国 0.02	業務 0.01
18位	国内線飛行機分担率30~39% 0.00	04~06月期 0.01	07~09月期 0.02	鉄道分担率30~39% 0.01
19位	青森空港出国 0.00	カナダ 0.01	8~10日滞在 0.02	石川県 (0~2日宿泊) 0.01
20位	秋田県 (0~2日宿泊) 0.00	羽田空港入国 0.01	バス分担率10~19% 0.02	6トリップ 0.01
0.01以上の属性	なし	11~14日滞在, 01~03月期	10~12月期, バス分担率30~39%, 15~30日滞在, 5トリップ, 11~14日滞在, 4トリップ, 6トリップ, 千葉県 (0~2日宿泊), 鉄道分担率60~69%	富山県 (0~2日宿泊), 3トリップ, 台湾

順位	14.4.関西	14.6.九州	14.7.沖縄	14.1.解釈困難
1位	大阪府 (0~2日宿泊) 0.40	鉄道分担率20~29% 0.06	沖縄県 (0~2日宿泊) 0.07	国内線飛行機分担率0~9% 0.19
2位	京都府 (0~2日宿泊) 0.32	福岡県 (0~2日宿泊) 0.06	鉄道分担率0~9% 0.05	観光・レジャー 0.14
3位	関西空港出国 0.26	韓国 0.04	バス分担率0~9% 0.03	自動車分担率0~9% 0.11
4位	関西空港入国 0.24	バス分担率40~49% 0.04	不明空港入国 0.03	07~09月期 0.07
5位	鉄道分担率60~69% 0.16	福岡空港出国 0.04	訪問地不明 (0~2日宿泊) 0.03	10~12月期 0.07
6位	バス分担率20~29% 0.14	福岡空港入国 0.03	0日滞在 0.03	2トリップ 0.06
7位	自動車分担率0~9% 0.13	自動車分担率20~29% 0.03	鉄道分担率10~19% 0.02	バス分担率30~39% 0.06
8位	兵庫県 (0~2日宿泊) 0.08	バス分担率50~59% 0.03	那覇空港出国 0.02	自動車分担率10~19% 0.05
9位	大阪府 (3~7日宿泊) 0.07	大分県 (0~2日宿泊) 0.03	那覇空港入国 0.02	鉄道分担率50~59% 0.05
10位	奈良県 (0~2日宿泊) 0.07	熊本県 (0~2日宿泊) 0.02	2トリップ 0.02	5日滞在 0.05
11位	観光・レジャー 0.05	3日滞在 0.02	自動車分担率40~49% 0.01	台湾 0.04
12位	京都府 (3~7日宿泊) 0.05	長崎県 (0~2日宿泊) 0.02	自動車分担率50~59% 0.01	韓国 0.04
13位	国内線飛行機分担率0~9% 0.04	鹿児島県 (0~2日宿泊) 0.01	沖縄県 (3~7日宿泊) 0.01	中国 0.04
14位	広島県 (0~2日宿泊) 0.02	香川県 (0~2日宿泊) 0.01	家族・知人の訪問 0.01	3トリップ 0.04
15位	8~10日滞在 0.02	観光・レジャー 0.01	バス分担率30~39% 0.00	4トリップ 0.04
16位	和歌山県 (0~2日宿泊) 0.02	博多海港入国 0.01	茨城空港出国 0.00	4日滞在 0.04
17位	15~30日滞在 0.02	博多海港出国 0.01	新潟空港出国 0.00	業務 0.03
18位	11~14日滞在 0.01	福岡県 (3~7日宿泊) 0.01	07~09月期 0.00	バス分担率20~29% 0.03
19位	04~06月期 0.01	宮崎県 (0~2日宿泊) 0.01	新潟空港入国 0.00	01~03月期 0.03
20位	4トリップ 0.01	高松空港出国 0.01	香港 0.00	5トリップ 0.02
0.01以上の属性	鉄道分担率70~79%, 大阪府 (8日以上宿泊), 中国, 鉄道分担率50~59%, バス分担率10~19%, 5トリップ, 10~12月期	なし	なし	04~06月期, 成田空港入国, 成田空港出国, 6日滞在, 3日滞在, 家族・知人の訪問, 8~10日滞在, 7日滞在, 大阪府 (0~2日宿泊), 関西空港出国, 香港, 関西空港入国

表3 トピックと名前 (2015)

順位	15.8.北海道	15.2.東京	15.3.関東	15.4.中部・関西
1位	北海道 (0~2日宿泊) 0.19	個人旅行 0.11	東京都 (0~2日宿泊) 0.52	観光・レジャー 0.10
2位	団体旅行 0.08	国内線飛行機分担率0~9% 0.09	成田空港出国 0.23	自動車分担率0~9% 0.10
3位	鉄道分担率0~9% 0.06	鉄道分担率50~59% 0.08	成田空港入国 0.23	国内線飛行機分担率0~9% 0.08
4位	5日滞在 0.05	自動車分担率10~19% 0.06	東京都 (3~7日宿泊) 0.11	中国 0.06
5位	観光・レジャー 0.05	業務 0.05	神奈川県 (0~2日宿泊) 0.07	団体旅行 0.06
6位	台湾 0.04	2トリップ 0.05	千葉県 (0~2日宿泊) 0.07	訪問回数不明 0.05
7位	訪日回数不明 0.04	羽田空港入国 0.04	訪日1回目 0.06	10~12月期 0.05
8位	新千歳空港入国 0.04	羽田空港出国 0.04	鉄道分担率40~49% 0.05	6日滞在 0.04
9位	新千歳空港出国 0.03	01~03月期 0.04	15~30日滞在 0.04	07~09月期 0.04
10位	自動車分担率0~9% 0.03	バス分担率20~29% 0.03	バス分担率20~29% 0.04	鉄道分担率0~9% 0.03
11位	バス分担率70~79% 0.03	04~06月期 0.03	11~14日滞在 0.04	大阪府 (0~2日宿泊) 0.03
12位	5トリップ 0.03	3トリップ 0.02	鉄道分担率50~59% 0.03	京都府 (0~2日宿泊) 0.03
13位	鉄道分担率10~19% 0.02	バス分担率30~39% 0.02	個人旅行 0.03	静岡県 (0~2日宿泊) 0.03
14位	鹿児島県 (0~2日宿泊) 0.02	韓国 0.02	米国 0.03	5トリップ 0.02
15位	07~09月期 0.02	訪日20回以上 0.02	自動車分担率10~19% 0.02	関西空港入国 0.02
16位	香川県 (0~2日宿泊) 0.02	東京都 (0~2日宿泊) 0.02	山梨県 (0~2日宿泊) 0.02	関西空港出国 0.02
17位	10~12月期 0.01	自動車分担率0~9% 0.02	東京都 (8日以上宿泊) 0.02	4トリップ 0.02
18位	北海道 (3~7日宿泊) 0.01	訪日10~19回目 0.02	04~06月期 0.02	7日滞在 0.01
19位	宮崎県 (0~2日宿泊) 0.01	10~12月期 0.01	国内線飛行機分担率0~9% 0.02	5日滞在 0.01
20位	高松空港出国 0.01	3日滞在 0.01	バス分担率10~19% 0.02	愛知県 (0~2日宿泊) 0.01
0.01以上の属性	なし	訪日2回目, 07~09月期, 訪日6~9回目, 訪日1回目, 観光・レジャー	8~10日滞在, 01~03月期, 広島県 (0~2日宿泊), 7トリップ, 国内線飛行機分担率10~19%, オーストラリア, イタリア, 栃木県 (0~2日宿泊), 8トリップ	台湾, 岐阜県 (0~2日宿泊), 石川県 (0~2日宿泊), 関西空港出国

順位	15.1.関西	15.5.愛知・広島	15.7.九州	15.6.沖縄
1位	大阪府 (0~2日宿泊) 0.11	国内線飛行機分担率0~9% 0.19	福岡県 (0~2日宿泊) 0.06	沖縄県 (0~2日宿泊) 0.05
2位	自動車分担率0~9% 0.10	個人旅行 0.17	韓国 0.05	4日滞在 0.05
3位	関西空港出国 0.08	2トリップ 0.14	鉄道分担率20~29% 0.03	観光・レジャー 0.05
4位	バス分担率30~39% 0.08	自動車分担率0~9% 0.11	大分県 (0~2日宿泊) 0.03	韓国 0.05
5位	関西空港入国 0.08	07~09月期 0.10	福岡空港出国 0.03	国内線飛行機分担率0~9% 0.03
6位	国内線飛行機分担率0~9% 0.08	訪日回数不明 0.10	福岡空港入国 0.03	バス分担率30~39% 0.03
7位	観光・レジャー 0.07	10~12月期 0.09	3日滞在 0.02	鉄道分担率10~19% 0.03
8位	京都府 (0~2日宿泊) 0.06	バス分担率0~9% 0.09	バス分担率40~49% 0.02	個人旅行 0.02
9位	個人旅行 0.06	1~2日滞在 0.06	自動車分担率20~29% 0.02	那覇空港入国 0.02
10位	訪日1回目 0.06	愛知県 (0~2日宿泊) 0.05	熊本県 (0~2日宿泊) 0.02	那覇空港出国 0.02
11位	鉄道分担率60~69% 0.04	家族・知人の訪問 0.05	観光・レジャー 0.01	自動車分担率50~59% 0.02
12位	3トリップ 0.04	鉄道分担率30~39% 0.05	バス分担率50~59% 0.01	5日滞在 0.01
13位	4トリップ 0.03	バス分担率50~59% 0.04	長崎県 (0~2日宿泊) 0.01	台湾 0.01
14位	大阪府 (3~7日宿泊) 0.03	中部空港出国 0.04	自動車分担率10~19% 0.01	訪日2回目 0.01
15位	兵庫県 (0~2日宿泊) 0.02	鉄道分担率0~9% 0.04	4日滞在 0.01	10~12月期 0.01
16位	04~06月期 0.02	バス分担率40~49% 0.04	国内線飛行機分担率0~9% 0.01	07~09月期 0.01
17位	8~10日滞在 0.02	中部空港入国 0.04	博多海港入国 0.01	自動車分担率40~49% 0.01
18位	鉄道分担率50~59% 0.02	観光・レジャー 0.03	博多海港出国 0.01	香港 0.01
19位	訪日2回目 0.02	中国 0.03	4トリップ 0.00	沖縄県 (3~7日宿泊) 0.01
20位	01~03月期 0.02	不明空港入国 0.03	福岡県 (3~7日宿泊) 0.00	訪日3回目 0.01
0.01以上の属性	韓国, 07~09月期, 10~12月期, 5日滞在, 奈良県 (0~2日宿泊), 京都府 (3~7日宿泊)	研修・学会等, 3トリップ, 米国, その他, 広島県 (0~2日宿泊), 8~10日滞在, 7日滞在, 訪日1回目, 業務, フィリピン, 鉄道分担率40~49%, 31~90日滞在, 留学	なし	なし

表 4 トピックと名前 (2016 年)

順位	16.8. 北海道	16.1. 関東	16.4. 関東・中部・関西・広島	16.5. 中部
1位	北海道 (0~2日宿泊) 0.12	東京都 (0~2日宿泊) 0.17	訪日1回目 0.23	国内線飛行機分担率0~9% 0.11
2位	鉄道分担率10~19% 0.02	成田空港出国 0.09	自動車分担率0~9% 0.18	団体旅行 0.11
3位	新千歳空港入国 0.02	成田空港入国 0.09	観光・レジャー 0.17	中国 0.10
4位	バス分担率60~69% 0.02	鉄道分担率50~59% 0.08	京都府 (0~2日宿泊) 0.16	観光・レジャー 0.10
5位	新千歳空港出国 0.02	個人旅行 0.05	東京都 (0~2日宿泊) 0.15	自動車分担率0~9% 0.08
6位	自動車分担率10~19% 0.02	業務 0.04	個人旅行 0.15	訪日回数不明 0.07
7位	団体旅行 0.02	自動車分担率10~19% 0.04	国内線飛行機分担率0~9% 0.14	10-12月期 0.06
8位	観光・レジャー 0.01	バス分担率30~39% 0.04	バス分担率20~29% 0.12	鉄道分担率0~9% 0.06
9位	台湾 0.01	羽田空港入国 0.04	8~10日滞在 0.10	愛知県 (0~2日宿泊) 0.06
10位	5日滞在 0.01	東京都 (3~7日宿泊) 0.04	11~14日滞在 0.07	6日滞在 0.06
11位	香川県 (0~2日宿泊) 0.01	羽田空港出国 0.04	15~30日滞在 0.07	07-09月期 0.06
12位	5トリップ 0.01	国内線飛行機分担率0~9% 0.03	東京都 (3~7日宿泊) 0.06	5日滞在 0.04
13位	バス分担率70~79% 0.01	千葉県 (0~2日宿泊) 0.02	04-06月期 0.06	静岡県 (0~2日宿泊) 0.04
14位	北海道 (3~7日宿泊) 0.01	神奈川県 (0~2日宿泊) 0.02	5トリップ 0.06	台湾 0.04
15位	函館空港出国 0.01	01-03月期 0.02	広島県 (0~2日宿泊) 0.05	中部空港出国 0.03
16位	高松空港出国 0.01	米国 0.02	バス分担率10~19% 0.05	6トリップ 0.03
17位	高松空港入国 0.00	04-06月期 0.01	京都府 (3~7日宿泊) 0.05	中部空港入国 0.03
18位	愛媛県 (0~2日宿泊) 0.00	訪日20回以上 0.01	成田空港入国 0.04	岐阜県 (0~2日宿泊) 0.02
19位	旭川空港出国 0.00	鉄道分担率40~49% 0.01	4トリップ 0.04	7日滞在 0.02
20位	函館空港入国 0.00	東京都 (8日以上宿泊) 0.01	7トリップ 0.04	石川県 (0~2日宿泊) 0.02
0.01以上の属性	なし	なし	07-09月期, 山梨県 (0~2日宿泊), 鉄道分担率60~69%, 成田空港出国, 鉄道分担率70~79%, 奈良県 (0~2日宿泊), 鉄道分担率40~49%など	鹿児島県 (0~2日宿泊), 鉄道分担率30~39%, 5トリップ, 富山県 (0~2日宿泊), バス分担率40~49%, 香港, 個人旅行, 4トリップ
順位	16.2. 関西	16.7. 九州	16.6. 沖縄	16.3. 解釈困難
1位	大阪府 (0~2日宿泊) 0.14	福岡県 (0~2日宿泊) 0.08	沖縄県 (0~2日宿泊) 0.12	2トリップ 0.11
2位	関西空港出国 0.09	韓国 0.06	観光・レジャー 0.11	国内線飛行機分担率0~9% 0.09
3位	関西空港入国 0.09	福岡空港出国 0.04	4日滞在 0.10	個人旅行 0.09
4位	自動車分担率0~9% 0.09	福岡空港入国 0.04	国内線飛行機分担率0~9% 0.09	訪日回数不明 0.07
5位	鉄道分担率60~69% 0.07	鉄道分担率20~29% 0.04	韓国 0.09	10-12月期 0.06
6位	国内線飛行機分担率0~9% 0.07	大分県 (0~2日宿泊) 0.03	個人旅行 0.08	07-09月期 0.06
7位	個人旅行 0.06	3日滞在 0.03	バス分担率20~29% 0.07	自動車分担率0~9% 0.05
8位	観光・レジャー 0.06	バス分担率40~49% 0.03	鉄道分担率0~9% 0.05	バス分担率0~9% 0.04
9位	3トリップ 0.05	バス分担率50~59% 0.03	那覇空港入国 0.04	1~2日滞在 0.03
10位	京都府 (0~2日宿泊) 0.04	自動車分担率20~29% 0.02	那覇空港出国 0.04	鉄道分担率0~9% 0.02
11位	バス分担率30~39% 0.03	自動車分担率10~19% 0.02	訪日2回目 0.03	家族・知人の訪問 0.02
12位	大阪府 (3~7日宿泊) 0.03	観光・レジャー 0.02	5日滞在 0.03	観光・レジャー 0.02
13位	韓国 0.03	長崎県 (0~2日宿泊) 0.01	4トリップ 0.03	バス分担率50~59% 0.01
14位	バス分担率20~29% 0.03	国内線飛行機分担率0~9% 0.01	01-03月期 0.02	不明空港入国 0.01
15位	4トリップ 0.03	個人旅行 0.01	自動車分担率40~49% 0.02	中国 0.01
16位	訪日1回目 0.02	熊本県 (0~2日宿泊) 0.01	04-06月期 0.02	3日滞在 0.01
17位	兵庫県 (0~2日宿泊) 0.02	4日滞在 0.01	台湾 0.02	韓国 0.01
18位	04-06月期 0.02	01-03月期 0.01	5トリップ 0.02	自動車分担率50~59% 0.01
19位	訪日2回目 0.02	佐賀県 (0~2日宿泊) 0.01	訪日3回目 0.02	その他 0.01
20位	01-03月期 0.02	博多海港入国 0.01	07-09月期 0.01	5日滞在 0.01
0.01以上の属性	5日滞在, 奈良県 (0~2日宿泊)	なし	訪日4回目, 沖縄県 (3~7日宿泊), 3トリップ, 訪日1回目, 自動車分担率30~39%	なし

その分布特性を確認しながら決定した。そのほかのカテゴリカルデータは、全てダミー変数で表現した。連続量を離散化した変数ではサンプルごとに属性別階級への該当/非該当をダミー変数(1/0)で示し、カテゴリカルデータも同様の形式として行列データを作成する。

本研究では、表1に示す属性をデータとして用いた。なお、2014年のFFデータには、旅行手配方法、訪日経験回数の属性はない。

3. 2 トピック数の決定

トピックモデルでは、抽出するトピック数 K を、モデル推定に先立って設定しなくてはならない。多くの場合、 K を変更すると最終尤度が変わる。そこで、候補とする K の範囲を定めて計算を繰り返して、最適な K を選択する。ただし、トピックの最適性は統計的性質のみで定まるわけではなく、解釈の容易性に配慮しなくてはならない。本研究は、まず尤度比を用いて K の検

討範囲の上限となるトピック数 K_{max} を決定する。統計的には尤度比が極大値をとるトピック数が最も適合性が高いが、そのときのトピックの解釈性も併せて検討しなくてはならない。そこで以下の分析では、 K_{max} は検討範囲の上限として、推定したトピック群から類似したトピックを合成する。本研究は類似性の指標として、トピックベクトル間のコサイン類似度を用いる。トピック k とトピック k' の類似度 $M_{kk'}$ は、式(1)で定義される。 $M_{kk'}$ の値が大きいほど、トピック k とトピック k' の類似性が高く、冗長なトピックが得られたことになる。

以下の分析では、トピック間の類似度を定量的に判定する閾値を設定した上で、 K_{max} からそれぞれ類似トピック数を合成していき、残った非類似トピック数が最大となるトピック数を、最適値とした。なお、類似度の閾値は、 $\cos 45^\circ$ となる 0.7 を閾値とする。

$$M_{kk'} = \frac{U_k \cdot U_{k'}}{|U_k| |U_{k'}|} \quad (1)$$

3. 3 トピックの解釈・名付け

各トピックは、以下の手順で命名する。まず2014年は14、2015年は15、2016年は16のように、トピック名の末尾にそれぞれ年号の識別記号を付けた。次に構成比率が大きいトピックから順に1, 2, 3とした。そのうえで、周遊地方が明確な場合にはその地方名、または都道府県名を付した。ただし、地方、都道府県が複数の場合は、場合はそれらを付した。周遊している地方や都道府県情報が不明のトピックは、解釈困難と名付けた。その手順の結果は、表2~4に示す。

4. 訪日外国人の旅行特性と経年変化

4. 1 地方別訪日外国人旅行特性

各トピックの中で寄与率が上位の属性に着目する。北海道トピックは、台湾、観光レジャー目的、団体旅行、5日滞在、バス分担率60~69、70~79%などが上位である。関東トピックは、中国、アメリカ、業務、4~6月期、鉄道分担率50~59%などが上位である。中部トピックは、中国、台湾、観光レジャー、6日滞在、7~9月期、10~12月期などが上位である。関西トピックは、韓国、観光レジャー、個人旅行、訪日2回目、4~6月期、バス分担率20~29%、鉄道分担率60~69%などが上位である。九州トピックは、韓国、観光レジャー、3日滞在、バス分担率40~49、50~59%、自動車分担率20~29%などが上位である。沖縄トピックは、台湾、観光レジャー、個人旅行、訪日2回目、自動車分担率40~49、50~59%などが上位である。複数地方トピックは、観光レジャー、訪日1回目、8~10、11~14、15~30日滞在などが上位である。解釈困難トピックは、中国、韓国、台湾、観光レジャー、業務などが上位である。

4. 2 訪日外国人旅行特性の経年変化

推定した全トピックベクトル間のコサイン類似度を算出した。類似度の閾値は0.7とする。北海道、中部、九州トピックは、3年ともほぼ同じ属性で構成される。関東、関西トピックは2014年では東京都(0~2日宿泊)、大阪府(0~2日宿泊)、京都府(0~2日宿泊)の属性の寄与が大きい。2015、2016年では寄与が減少している。沖縄トピックは、2014年では0日滞在が上位である。同トピックは、2015年からは4日滞在、5日滞在が上位である。

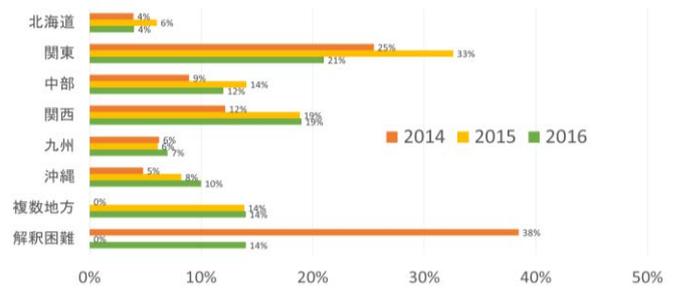


図2 トピック構成比率の経年変化

4. 3 トピック構成比率の経年変化

トピック構成比率は、DK行列の係数にFFデータの拡大係数を乗じて算出した。図2に、2014~2016年のトピックの構成比率を示す。関東、関西トピックの構成比率が大きく、観光地である北海道、沖縄トピックの構成比率が、一定程度大きいことも確認できた。これらは、分析結果の妥当性を示していると考えられる。

関西、九州、沖縄、複数地方トピックの構成比率は、経年的に増加している。一方、北海道、関東、中部トピックの構成比率は増減している。解釈困難トピックの構成比率は、年によるばらつきが大きい。

5. 結論

本研究では、トピックモデルを用いて、訪日外国人の旅行特性の経年分析を行った。関東、関西を周遊する旅行では、東京都、大阪府以外の周遊割合が高まっている。沖縄を周遊する旅行では、滞在日数が増加している。2015年では、複数の府県を周遊するトピックが現れる。2015年、2016年では、関東以外を周遊する旅行の割合が、大きく増加している。以上の結果より、訪日外国人の周遊先は、多様化していることが分かった。

今後の課題として、本研究のようにFFデータを年次別に分析するのではなく、季節ごとや出国海空港ごとに分析する方法を試行して、それらの推定トピックの安定性を検討する必要がある。

参考文献

- 1) 国土交通省：FF-Data（訪日外国人流動データ）
http://www.mlit.go.jp/sogoseisaku/soukou/sogoseisaku_soukou_fr_000022.html
- 2) 古屋秀樹・劉瑜娟：潜在クラス分析を用いた訪日外国人旅行者の訪問パターンの分析，土木学会論文集D3，Vol.72，No.5，pp571-583，2016
- 3) 菱田のぞみ・日比野直彦・森地茂：訪問地選択の多様性に着目した訪日中国人旅行者の居住地別観光行動の時系列分析，土木学会論文集D3，Vol.68，No.5，pp667-677，2012
- 4) 塚野裕太：トピックモデルを用いた土地利用メッシュ特性の抽出，平成28年度卒業論文