

2016年9月14日

報道関係各位

GMO リサーチ株式会社

訪日中国人は東京で何をする？

～SNS 投稿データから『築地市場』『上野公園』『新宿』での行動を分析～

GMO インターネットグループでインターネットリサーチ事業を展開する GMO リサーチ株式会社（代表取締役社長 細川 慎一 以下、GMO リサーチ）は、株式会社ナイトレイ（代表取締役社長 石川 豊 以下、ナイトレイ）より、訪日外国人観光客の行動解析サービス「インバウンド インサイトinbound insight」のデータ提供を受け、訪日中国人観光客（以下、訪日中国人）の行動分析を行いました。

SNS に投稿された発言内容・位置情報等から、今春、東京都内各所を訪れた訪日中国人の目的が明らかになりました。

- 分析テーマ：訪日中国人観光客の日本での行動
- 分析地域：東京都区部（23区）
- 分析対象：訪日中の中国人観光客の Weibo および Twitter 上での発言 543 件
- 分析期間：2016年3月～4月の2ヶ月間
- 分析方法：KH Coder ^(※1)を利用した共起ネットワーク分析 ^(※2)

(※1) 立命館大学・樋口耕一氏が自作し公開している、テキストデータを統計的に分析するためのフリーソフトウェア。

(※2) 共起ネットワーク分析とは、文章の中で使われている特徴的な言葉(単語)や出現数の多い言葉などを抽出し、それと共起関係にある言葉(単語)を視覚的にネットワーク化すること。

【分析背景】

訪日外国人観光客の数は、2015年に約1,974万人に達し^(※3)、2016年の上半期累計は対前年同期比28.2%増の約1,171万人と過去最高を記録する^(※4)など、引き続き増加傾向にあります。中でも訪日中国人は、上半期累計で300万人超、対前年同期比41.2%増と国別でトップの伸び率となっており、今後もさらなる増加が期待されています^(※4)。

過去にGMOリサーチが実施した調査によると、訪日中国人は、滞在中に「日本食」、「ショッピング」、「自然・景勝地観光」を楽しんでいる人が多いことが分かっています^(※5)。

そこでGMOリサーチは、訪日中国人が東京都内において、具体的にどのエリア・施設でどのような観光行動をしているのかを探るべく、ナイトレイの訪日外国人観光客の行動解析サービス「インバウンド インサイトinbound insight」で解析された訪日中国人のSNS投稿データをもとに、その行動を分析しました。

GMOリサーチとナイトレイは、訪日外国人観光客のSNSでの発言内容・投稿画像・位置情報等を参照したリサーチサービスの提供を予定しており、本分析は、今後のサービス提供に向けたテスト調査となります。

(※3) 日本政府観光局 (JNTO) : 2016年1月19日「訪日外客数 (2015年年間推計値)」

(※4) 日本政府観光局 (JNTO) : 2016年7月20日「訪日外客数 (2016年6月推計値及び上半期推計値)」

(※5) 2016年6月1日発表「訪日中国人の消費行動調査」(URL : <http://www.gmo.jp/news/article/?id=4860>)

【分析結果】

■全体構造：エリアごとにクラスタが存在（図・表 1）

分析の結果、『築地』『上野』『新宿』といった地名の出現数が多く、地名を中心としたクラスタが存在することがわかった。例えば、『築地』は「食べる」「寿司」など食関連、『上野』は「花」「桜」などのお花見関連、『新宿』は「御苑」や「西新宿」といった施設・地名関連の言葉（単語）と共起されている。一方で、他の大多数の単語には一定の傾向は見出しにくい。

■部分構造：クラスタごとの共起ネットワーク分析

以下に、上位3クラスタ『築地市場』『上野公園』『新宿』のネットワーク分析結果を示す。

出現数（全体）は、分析対象となった SNS 投稿データ（543 件）の中で、その単語が出現した回数を、共起数は各クラスタ内での単語の出現回数および割合を示している。

●クラスタ 1 築地市場（図・表 2）

『築地市場』クラスタでは、地名である「築地」（共起率：100%）および「市場」（96.1%）がネットワークの中心となり、他の単語と広く共起されていることがわかる。この2語に続いて、「寿司」「朝食」「朝」「刺身」が共起数3回（5.9%）以上となっており、訪日中国人は新鮮な海産物を食べるために、朝早くに築地市場を訪れている様子がうかがえる。

●クラスタ 2 上野公園（図・表 3）

『上野公園』クラスタでは、地名と関係する「上野恩賜公園」（95.7%）、「上野」（93.5%）の2語が9割以上の発言に含まれており、「公園」（43.5%）とともにネットワークの中心となっている。これに次いで「桜」が共起数16回（34.8%）、関連語と思われる「花」が同4回（8.7%）となっており、上野公園に桜の鑑賞目的で訪問している人が多いことが読み取れる。

また、「桜」や「公園」とともに「東京近代美術館」「東京国立博物館」「正岡子規記念球場」といった施設が共起されており、花見とともに上野公園内のさまざまな施設へも足を運んでいることがうかがえる。

●クラスタ 3 新宿（図・表 4）

『新宿』クラスタでは、地名である「新宿」（100%）を中心に、「御苑」（55.3%）も比較的大きな共起ネットワークを形成しており、新宿御苑周辺で観光する人が多数いることがわかる。この「御苑」ネットワークでは、少数ではあるが、「桜」「花」「花びら」といった単語が共起されており、新宿御苑でも花見をしている様子がうかがえる。

そのほか、「ZARA」「丸井」「蟹道楽」「瀬里奈」といった新宿に点在する店舗名が共起されており、新宿周辺で食事やショッピングを楽しんでいる様子も見受けられる。

【総論】

この度の SNS 投稿データ分析により、今春、日本を訪れた中国人が、東京都内のどこで何を目的に行動していたかが明らかになりました。3月～4月という季節柄、特徴的なキーワードは「食」と「桜」となっています。

「食」というキーワードに代表されるのは、クラスタ1として挙げた築地市場です。ネットワーク図を見ると、多くの訪日中国人が「マグロ」「トロ」「ウニ」「ホタテ」といった海産物を朝早くから楽しんでいることが分かりました。海産物を中心とした市場や食堂などの専門店が多数連なる築地市場は、昨今では訪日外国人観光客向けのツアーも組まれていることから、観光スポットの1つとして定着していると考えられます。先月8月31日には、築地市場の移転が来年2月以降に延期することが発表されたため、今秋～冬にかけて訪日中国人の駆け込み客が増えるかもしれません。

一方、日本の春を代表する「桜」という単語も多数出現しており、訪日中国人は上野公園、新宿御苑で花見を楽しんでいることが明らかとなりました。投稿の中には「#sakura」「#cherryblossom」といったハッシュタグとともに画像 URL が添付されている事例も複数見られました。このように、日本の風物詩にも関心が高いことから、これからの季節は紅葉が話題として挙がりそうです。

【分析方法】

1. 共起ネットワーク分析で全体構造を把握

訪日中国人の SNS 投稿データを以下の仕様に整理した後、KH Coder を利用して、単語同士の共起関係を視覚的に表す共起ネットワーク分析を行った（図 1・表 1）。その際、形態素解析^(※6) エンジンとして Chasen（茶筌）を利用している。

(※6) 形態素解析とは、意味を持つ最小単位（単語）に文章を分節化し、その品詞や内容等を判別すること。

■ SNS 投稿データの仕様について

2016 年の 3 月～4 月に東京都区部（23 区）に訪れた訪日中国人が、Weibo および Twitter に投稿した発言 543 件の「①発言内容のテキスト」（ナイトレイが機械翻訳により和訳）、発言時の GPS 情報から推定した「②位置情報（地名をタグとして使用）」、「③発言に含まれるハッシュタグ」の 3 点を紐付けて活用した。

「#japan」「#tokyo」といった大多数の発言に共通するハッシュタグや、同一施設でありながらチェックイン表記が異なるなどの投稿は、分析者がタグや施設名等を統一し再設定している。また、タグおよび位置情報が示す地名・施設名等計 367 語については、共起ネットワーク分析時に分節化せず、一語として抽出している（例：東京＋タワーで一語）。なお、発言内に存在する URL は分析対象に含めていない。

2. クラスタごとの分析で部分構造を把握

さらに、全体構造から想定されるクラスタ（集団）をより詳しく分析するため、共起の頻度が高い単語（ノード）をグルーピングする「クラスタ分析」を行い、『築地市場』『上野公園』『新宿』『原宿・神宮前・表参道』『その他』の 5 つのクラスタに分類した。このうち、上位 3 つのクラスタで共起ネットワーク分析を行った。（図・表 2～4）「クラスタ分析」の際には、発言同士の距離（類似性）を Jaccard 係数^(※7)を用いて算出し、Ward 法^(※8)を用いてクラスタリングしている。なお、どのクラスタにも属さない分類不能サンプルが 27 サンプル発生している。

(※7) Jaccard 係数とは、集合に含まれる各要素の単独出現数と共起回数から、集合間の類似度を測る測度。文章間の類似度を測定する際には、形態素解析後の各単語を要素として算出する。

(※8) Ward 法とは、クラスタ分析において、クラスタ間の距離（類似度）を測定する方法のひとつ。類似したクラスタをグルーピングしていくために用いる。

【GMO リサーチ株式会社について】

市場調査・分析および調査データを基にしたコンサルティングサービスを提供しております。従来通りの市場調査手法はもちろん、MROC やアイトラッキング、スキャナマインドなど、最先端の技術と手法を駆使した GMO リサーチの市場調査サービスは、企業の迅速かつ最適な意思決定のお手伝いをいたします。

また、GMO リサーチでは、アジア最大規模となる 2,000 万人超の会員を保有する「ASIA Cloud Panel」^{アジア クラウド パネル}のネットワークを活用し、訪日経験者や訪日予定者を対象とした消費実態や動向調査を多数行っています。今後はインバウンド消費者調査に特化したサービスを展開し、お客様のインバウンドビジネスを支援してまいります。

- 「インバウンド消費者調査サービス」 URL : <http://www.gmo-research.jp/inboundlp/>
- 「ASIA Cloud Panel」 URL : <http://www.gmo-research.jp/acp>

【株式会社ナイトレイについて】(URL : <http://nightley.jp/>)

2011年1月創業。ロケーションインテリジェンス分野において独自のビッグデータや解析・活用・可視化ノウハウを保有するスタートアップです。

主力事業となる訪日外国人旅行者の行動解析サービス「inbound insight」では、主に SNS 上に公開されている投稿内容を大規模にリアルタイム解析することで、訪問施設やクチコミ・国籍・性別・移動経路などをデータベース化し、「SNS 解析データ」、「統計データ」、「訪日消費データ」、「駅データ」のメニューを ASP 型分析ツールや解析結果データとして提供しています。これまで把握することが困難だった訪日外国人観光客の行動傾向や嗜好性の裏付けデータを提供することで、各種企業や自治体で進められているインバウンド対策の戦略策定・実行・検証プロセスをサポートしています。

- 「inbound insight」 URL : <http://inbound.nightley.jp/>

以上

【報道関係お問い合わせ先】

- GMO リサーチ株式会社 国内事業本部 担当 白鳥
TEL : 03-5962-0037(代表)
E-mail : pr@gmo-research.jp
- GMO インターネット株式会社
グループ広報・IR 部 石井・島田
TEL : 03-5456-2695 E-mail : pr@gmo.jp

【GMO リサーチ株式会社】 (URL : <http://www.gmo-research.jp/>)

会社名	GMO リサーチ株式会社 (東証マザーズ 証券コード : 3695)
所在地	東京都渋谷区桜丘町 26 番 1 号 セルリアンタワー
代表者	代表取締役社長 細川 慎一
事業内容	■インターネットリサーチ事業
資本金	2 億 9,903 万円

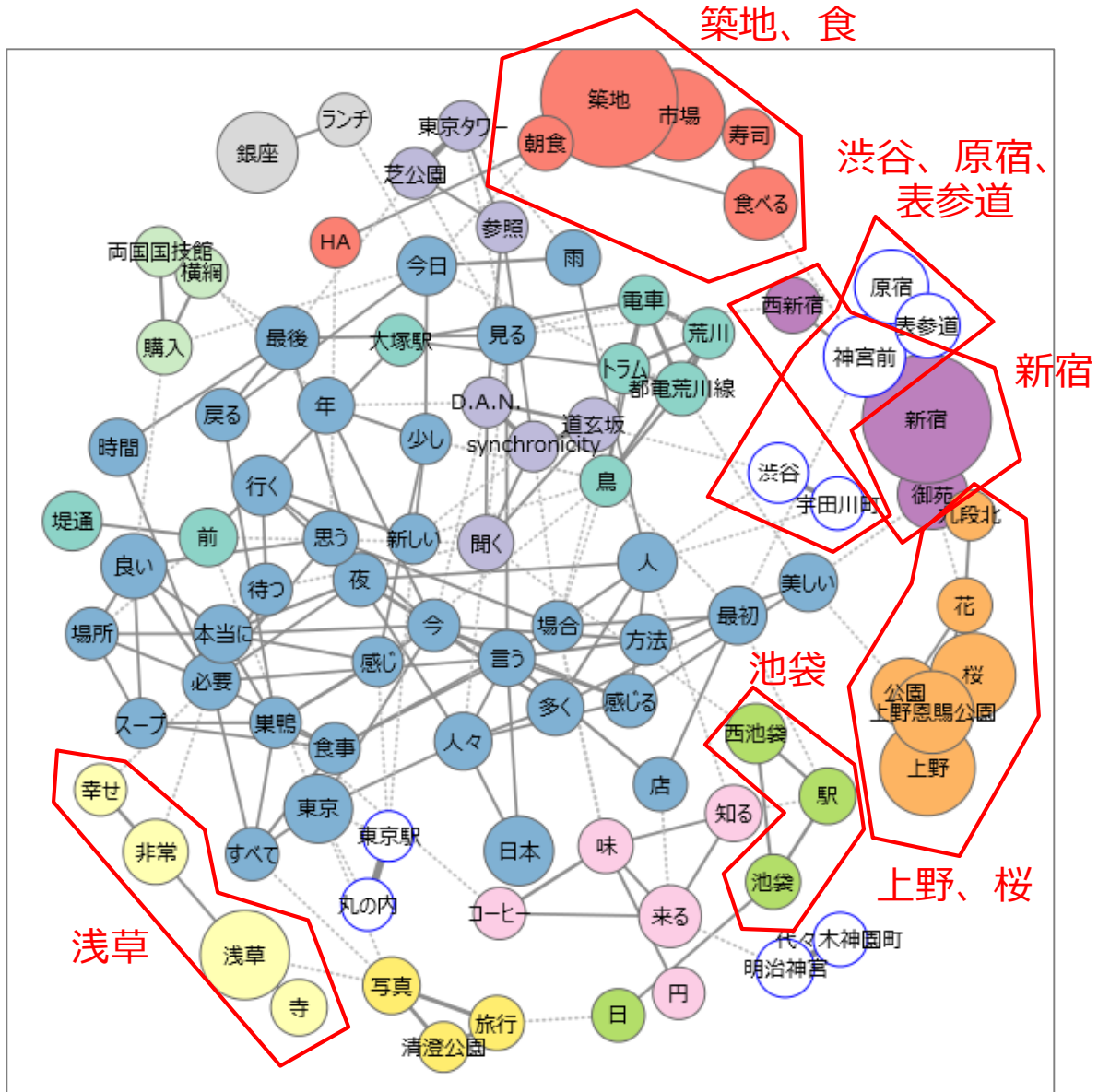
【GMO インターネット株式会社】 (URL : <http://www.gmo.jp/>)

会社名	GMO インターネット株式会社 (東証第一部 証券コード : 9449)
所在地	東京都渋谷区桜丘町 26 番 1 号 セルリアンタワー
代表者	代表取締役会長兼社長・グループ代表 熊谷 正寿
事業内容	■インターネットインフラ事業 ■インターネット広告・メディア事業 ■インターネット証券事業 ■モバイルエンターテインメント事業
資本金	50 億円

Copyright (C) GMO Research, Inc. All Rights Reserved.

【参考資料】

(図 1) 全体構造 ネットワーク図 最少出現回数= 7 / ノード数 = 95 / リンク数 = 203



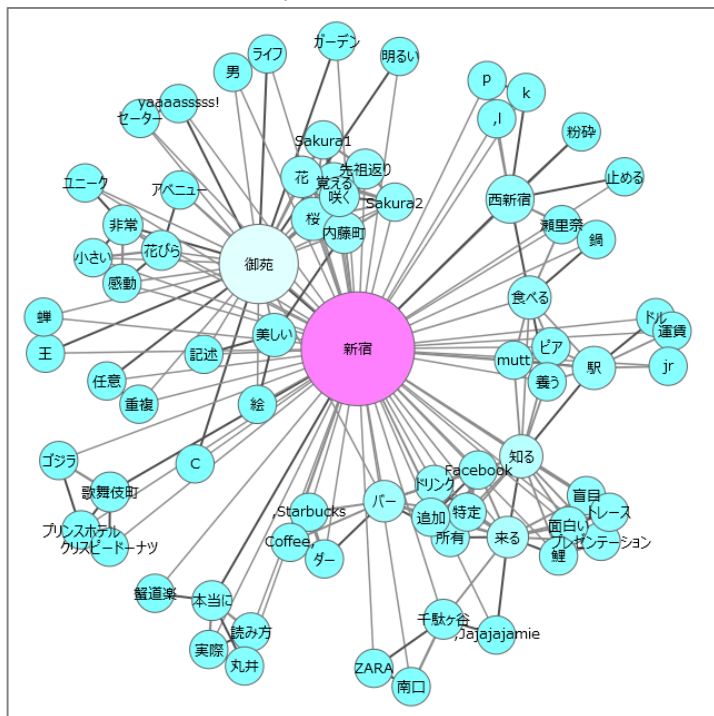
(図 1) *ノードの大きさは出現回数の多さを、リンクの太さは共起関係の強さを意味する。
 *類似度が高いと思われるノードごとに色分けをしている。
 *ノードはリンクを通じてのみ他のノードと関係性があり、ノード間の距離は特に意味を持たない。

(表 1) 全体構造 クラスタ分析結果

クラスタ番号	クラスタ名	サンプル数
1	築地市場	51
2	上野公園	46
3	新宿	47
4	原宿・神宮前・表参道	33
5	その他	339
	分類不能	27
	計	543

(図4) クラスタ3 新宿 ネットワーク図

ノード数 = 75 / リンク数 = 248



(表4) クラスタ3 新宿 抽出語 (共起数2以上)

順位	抽出語	品詞	出現数 (全体)	共起数	Jaccard 係数
1	新宿	タグ	57	47 (100.0%)	0.825
2	御苑	名詞	28	26 (55.3%)	0.531
3	西新宿	タグ	13	7 (14.9%)	0.132
4	知る	動詞	13	4 (8.5%)	0.071
5	駅	名詞C	14	4 (8.5%)	0.070
6	美しい	形容詞	15	4 (8.5%)	0.069
7	バー	名詞	4	3 (6.4%)	0.063
8	内藤町	タグ	4	3 (6.4%)	0.063
9	食べる	動詞	30	4 (8.5%)	0.055
10	花	名詞C	12	3 (6.4%)	0.054
11	桜	タグ	38	4 (8.5%)	0.049
12	来る	動詞	19	3 (6.4%)	0.048
13	花びら	名詞	2	2 (4.3%)	0.043
14	先祖返り	名詞	2	2 (4.3%)	0.043
15	咲く	動詞	3	2 (4.3%)	0.042
16	覚える	動詞	4	2 (4.3%)	0.041
17	歌舞伎町	タグ	5	2 (4.3%)	0.040
18	千駄ヶ谷	タグ	5	2 (4.3%)	0.040
19	本当に	副詞	16	2 (4.3%)	0.033
20	非常	形容動詞	23	2 (4.3%)	0.029

(図2~4)

- * ノードの大きさは出現回数の多さを意味する。
- * 他のノードとのリンク数が多い順にピンク→白→水色で表している。
- * ノードはリンクを通じてのみ他のノードと関係性があり、ノード間の距離は特に意味を持たない。
- * 太線のリンクは、ネットワークの維持に最低限必要な「全域木 (スパニングツリー)」と呼ばれるもので、当該ネットワークにおいて重要性が高いと考えられる。