

GPS データと SNS データの組み合わせによる マーケティング情報の抽出

藤澤 息吹* 長尾 光悦* 吉田 千鶴† 鈴木 裕一‡

*北海道情報大学 †ウイングアーク 1 s t 株式会社 ‡株式会社日立ソリューションズ東日本

s1512016@s.do-johodai.ac.jp

1. はじめに

本稿では、GPS データと SNS データの組み合わせによるマーケティング情報の抽出方法を提案する。ここでは、スマートフォン用フォトムービー作成アプリの 1 つである「たびろくムービーズ」に付帯する位置情報記録機能によって取得される GPS データと Twitter、及び、Instagram から獲得される SNS データを組み合わせることにより、有効なマーケティング情報の抽出を実現する。

2. たびろくムービーズ

たびろくムービーズは、株式会社日立ソリューションズ東日本が提供している旅行向けフォトムービー作成アプリである。このアプリは、動画と BGM で構成されている背景動画テンプレートとユーザが旅行先で撮影した写真を組み合わせることによって、ミニ映画のようなフォトムービーが作成できるものである[1]。

たびろくムービーズは位置情報の記録機能を有しており、アプリを起動している間、30 秒毎に位置情報が取得される。この取得された位置情報から、ユーザの移動ルートや滞在地に関する情報を抽出することは可能である。しかしながら、ユーザがどこでどのような観光を実施しているのかといった詳細な観光動態情報を位置情報のみから抽出することは困難である。

3. GPS データと SNS データの組み合わせによる マーケティング情報の抽出

3.1 概要

本提案手法を図 1 に示す。まず、たびろくムービーズの GPS データからユーザが滞在したエリアの抽出を行う。次に、Twitter からツイートが多くなされているエリアの抽出を行う。次に、Instagram から写真が撮影されているエリアの抽出を行う。更に、マッチングによ

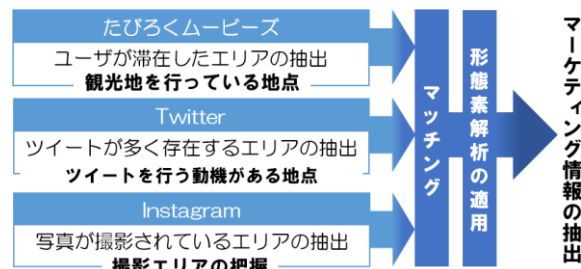


図 1 提案手法

り獲得されたエリア内のツイート集合に対して形態素解析を適用する。これら処理により、ユーザがどこでどのような観光をしているのか詳細な観光動態情報を獲得可能とする。

本研究における分析対象範囲は、札幌市、及び、その近隣都市を含む範囲とし、250m×250m の 5 次メッシュでの分析を行う。

3.2 GPS データからの滞在エリアの抽出

たびろくムービーズにおける全 GPS データは 5,551,903 件であり、このうち分析対象範囲のデータは 566,725 件であった。これらの GPS データから、各ユーザが 15 分以上滞在したエリアの抽出を行う[2]。すなわち、ユーザが観光を実施している可能性が高いエリアを抽出する。

3.3 Twitter からのツイートエリアの抽出

Twitter API により取得した全ツイートデータは 240,827,082 件であり、分析対象範囲のデータは 4,125,656 件であった。ここから、5 回以上のツイートがなされたエリアの抽出を行う。

3.4 Instagram からの撮影エリアの抽出

Instagram において「#札幌旅行」とタグ付けされた写真 3,271 枚のデータから、3 枚以上撮影されたエリアの抽出を行う。但し、この抽出は手動により実施する。また、業者による投稿などは、投稿内容に基づき予め除去する。このため Instagram のデータは、ノイズが含まれないものを利用する。

Extraction of Marketing information based on combination of
GPS data and SNS data

*Ibuki FUJISAWA • Hokkaido Information University

*Mitsuyoshi NAGAO • Hokkaido Information University

†Chizuru YOSHIDA • WingArc 1st Inc.

‡Yuichi SUZUKI • Hitachi Solutions East Japan, Ltd.

3.5 マッチング

各データのマッチングは、ウイングアーク 1st 株式会社 が提供する BI ツール MotionBoard を利用することにより実施する。また、マッチング後、各エリアにおけるツイート集合に対して、Mecab による形態素解析を適用する。

4. マーケティング情報の抽出結果

たびろくムービーズにおいて 15 分以上の滞在が行われたエリアとして、734 箇所が抽出された。抽出されたエリアとしては、札幌市を代表する観光地である、大通公園や羊ヶ丘展望台が含まれていたが、住宅地や JR 駅付近といった観光地である可能性が低い地点も多く抽出されていた。これは、アプリ利用者が当該地域への旅行者だけではなく、その地域に居住するユーザも含まれることが原因であると考えられる。

また、Twitter のツイートデータから、5 回以上ツイートがなされたエリアとして 1,006 箇所が抽出された。滞在エリアの結果と同様、範囲内の代表的な観光地である大通公園や羊ヶ丘展望台が含まれるエリアが抽出されていた。しかし、観光地の可能性が低い JR 駅付近、観光地が存在しない山中や農地においても多くのツイートが存在することが確認された。これは、Twitter が当該地域への旅行者が利用するアプリというわけではなく、日常的に利用されるものであることが原因と考えられる。

Instagram における撮影エリアの抽出結果として、滞在エリア、ツイートエリアと同様に札幌駅、大通公園といった主要観光地を含むエリアが抽出された。一方、ツイートエリアにのみ存在する藻岩山展望台やノースサファリサッポロ等が抽出された。

これら滞在エリア、ツイートエリア、撮影エリアのマッチングを行うことにより、ノイズを除去しつつ、詳細な観光動態を獲得することが可能である。マッチング結果を図 2 に示す。図 2 から、札幌市を代表する観光地である、札幌駅から中島公園にかけてのエリアや羊ヶ丘展望台といった観光地として人気の高いエリアが抽出された。また、たびろくムービーズ、及び、Twitter 単体に現れたノイズがマッチングを行うことにより除去されることが確認された。

更に、マッチングにより抽出されたエリアにおけるツイート集合に対して形態素解析を適用する。ここでは、まず、抽出されたエリア全体におけるツイート集合に対して形態素解析を適用し、上位 20 位までの名詞および形容詞を抽出する。更に、各エリアにおけるツイート集合に対して形態素解析を適用し、全体の形態素集合からの差分を取ることで、各エリアに独自の形態素集合の抽出を行う。これにより、この地域における全体としての観光概念、及び、各エリアを特徴づける観光概念を獲得する。



図 2 抽出結果(札幌地区)

順	名詞	回数	形容詞	回数	順	名詞	回数	形容詞	回数
1	https	134	美味しい	14	1	https	199	ない	8
2	丼	46	ない	10	2	羊	64	白い	6
3	市場	40	うまい	5	3	丘	43	いい	5
4	海鮮	33	良い	5	4	ヶ	41	寒い	5
5	札幌	30	楽しい	3	5	クラーク	34	かわいい	4
6	場外	25	おいしい	2	6	展望	33	っぽい	4
7	朝	19	でかい	2	7	台	33	美味しい	3
8	ごはん	11	大きい	2	8	北海道	19	よい	3
9	北海道	11	遅い	2	9	札幌	16	でっかい	2
10	いくら	9	美味しい	2	10	博士	14	楽しい	2

※ 全体の形態素集合と同じ箇所を色付き背景で示す

表 1 形態素解析結果

表 1 に、札幌場外市場付近(左)と羊ヶ丘展望台付近(右)における結果を示す。頻出名詞、及び、形容詞の上位 10 位までの結果を示している。表 1 から、このエリアでは、海鮮丼やネタが大きいといった特徴、羊ヶ丘展望台付近(右)では、クラーク博士像や羊がかわいいといった、そのエリアにおける観光地の特徴を推測可能である。

5. おわりに

本稿では、GPS データと SNS データの組み合わせによるマーケティング情報の抽出方法を提案した。本提案手法により、対象とするアプリのための有効なマーケティング情報を抽出可能であることが示された。今後の課題として、季節性分析、撮影写真の内容分析が挙げられる。

参考文献

- [1] 黒澤 彩華, 長尾 光悦, 斎藤 一, 佐藤 信吾, 辺見 吉克: 観光振興に向けた学生による地域の魅力発信～たびろくムービーズ動画テンプレート制作～, 観光情報学会第 14 回全国大会講演予稿集, pp.25-26(2017)
- [2] 青山 賢, 廣田 雅春, 石川 博, 横山 昌平: ジオタグ付き写真を用いた様々な観光地の平均的な滞在時間の推定, WebDB Forum 2015, pp.72-79(2015)