

Instagram を用いた流域毎の河川関心度評価

東北大学工学部 学生会員 ○安西 聡
 東北大学大学院工学研究科 正会員 峠 嘉哉
 東北大学大学院工学研究科 正会員 風間 聡

1. はじめに

河川に対する関心を定量的に評価するための研究は多く行われているが、その多くは単一の河川、特定の時間に限定したデータを用いている¹⁾。そこで地域的制限が無く、時間的連続性があるソーシャルメディアデータを一次データとして用いる。川守田ら²⁾はソーシャルメディアを用いて河川関心度の評価を試みたが、流域毎の議論がなされていない。よって、本研究ではソーシャルメディアの一つである Instagram を用いて流域毎の河川の興味の傾向を分析した。

2. Instagram の概要とその選択理由

ソーシャルメディアデータの一つである Instagram は、他のユーザーと相互に写真を共有できるアプリケーションである。投稿した写真にはキャプションテキストやハッシュタグが付けられる。ハッシュタグとはハッシュ記号「#」のあとにキーワードをつけたものである。ハッシュタグをつけることは投稿の意図を明確化するだけでなく、よりその投稿に対するリーチ数を高められる。

他のソーシャルメディアと比較し、Instagram は利用者が多く、非公開設定にするユーザー数が少ないため解析に必要なデータ数が得られること、画像、テキスト、位置情報等の多様な情報が含まれていること、データがサービス開始時から無料利用できること、自動投稿システムがないため同じ投稿の重複によるノイズが少ないことなどの理由から解析に有効である。

本研究では関東の代表的な河川である多摩川と関西の代表的な河川である淀川を対象とし、その河川名を持つ投稿から関心の評価を行った。

3. 形態素解析を用いた河川比較

3.1 手法

形態素解析によって河川の特徴や河川行動の抽出と分類を試みた。形態素解析は文章を品詞に分解し各単語の出現数を把握することができる。得られた各単語の出現数を投稿数によって除した値を出現率とし、5%以上の出現率の単語の中において活動を連想させるものを抽出した。興味の傾向を把握するために抽出した単語をイベントとアクティビティに、分類した。イベン

表-1 各河川の抽出された単語と投稿数/流域人口

河川名	投稿数/流域人口	イベント	出現率	アクティビティ	出現率
多摩川	0.070	花火	12%	散歩	7%
		大会	8%	BBQ	6%
淀川	0.058	花火, fireworks	26%	BBQ	11%
		大会	16%		

トは短い期間に高い関心があり、アクティビティは長い期間に小さな関心がある行動とした。また、川を比較する指標として投稿数を流域人口によって除した値を用いた。流域人口には Instagram ユーザーの年齢層を考慮し、10歳~39歳までの流域人口を用いた。流域人口データは年齢別人口のメッシュデータ³⁾を元に作成した。

3.2 結果・考察

表-1 に多摩川と淀川の抽出した単語をイベントとアクティビティに分類し、「投稿数/流域人口」の値と共に示す。どちらの河川においても「花火」、「大会」、「BBQ」の単語が確認できた。「花火」と「大会」の出現率が高い理由は多摩川と淀川において行われる花火大会の影響が大きい。また、都市河川において活発に行われている「BBQ」の活動をどちらの投稿中からも確認できた。しかしながら、「散歩」は多摩川にのみ確認された。淀川より多摩川の方が日常的に多く利用されている傾向が見える。また、そのために「投稿数/流域人口」の値は多摩川の方が高い。

4. 時系列データによる河川比較

4.1 手法

時系列データを用いて河川の特徴や定量的な河川行動の分類を試みた。分析に多摩川と淀川の2015年のデータを用いた。

一日の投稿数の標準偏差を平均によって除した値を変動係数とした。全投稿と抽出した各単語を含む投稿に対して変動係数を求め、河川行動の分類を試みた。

FFT(高速フーリエ変換)を用いたスペクトル解析による河川行動の分類も行った。FFTを行う際に、データ数を2ⁿ個にする必要があるため、投稿数を約3倍し1024個に変形させ解析した。また、全投稿と形態素解析により抽出された各単語を含む投稿についても同様にFFTを行った。

4.2 結果・考察

表-2 に河川の一日当たりの投稿数に関する平均値と標準偏差, 変動係数を記す. 形態素解析によって抽出し, イベントに分類した「花火」や「大会」は全ての投稿の変動係数よりも大きい. アクティビティに分類した「散歩」や「BBQ」の変動係数は全ての投稿の値以下か近い値をとる. よって変動係数による行動分類は可能であると考えられる.

図-1 と図-2 に多摩川と淀川 の FFT の結果をのせた. 縦軸に規格化した振幅を取り, 横軸に周期をとる. 図-1 より多摩川の「BBQ」は週周期とともに「花火」と同様に年周期が強い. BBQ は夏季に好まれるためである. また, 「散歩」は週周期とともに半年周期が強く影響していた. 散歩は春や秋等の気候条件が適度な時期に好まれるからである. 図-2 において淀川の「BBQ」の投稿は多摩川の「BBQ」と比較して半年周期がより影響していた. よって淀川は多摩川よりも BBQ が盛んに行われていることがわかる.

図-3 において多摩川と淀川 の全投稿のグラフを比較した. 多摩川において週周期や月周期の強い影響を確認し, 淀川において 100 日周期より短い様々な周期の影響を確認した. これは大阪における近年の盛んな水辺の取り組みによって, 活発な水辺利用の影響が考えられる.

5. まとめ

- 1) 多摩川と淀川の同じ都市河川においても関心の傾向に差が見られた. したがって都市河川でさえも画一的な開発は不適である.
- 2) BBQ と花火の活動は FFT によって確認すると, どちらも年周期が強い. 変動係数を用いると BBQ と花火において違いを確認できた. 今後活動の分類を定量的に行う場合は二つの分析を用いる必要がある.

謝辞

本研究（の一部）は, 科学研究費補助（15K14036, 代表: 風間聡）, および東北地域づくり協会の援助を受け実施されました. 平野勝也先生と福本潤也先生の助言に謝意を表します.

参考文献

- 1) 和田安彦, 陸奥康治, 和田有朗: 住民の暮らしからみた水辺環境の評価, 土木学会論文集, No.776/VII-33 pp.83-95, 2004.
- 2) 川守田智, 風間聡, 小森大輔: 河川関心度定量化に向けたソーシャルメディア分析の試み, 土木学会東北支部技術研究発表会, 平成 27 年度
- 3) 株式会社 日本統計センター: 推計年齢階級別人口付き人口・世帯数データ, 2015 年

表-2 河川的全投稿及び抽出単語の変動係数

河川名	単語名	平均値(投稿)	標準偏差(投稿)	変動係数
多摩川	全ての投稿	52.95	67.55	1.28
	花火	6.69	58.37	8.73
	大会	4.82	43.91	9.11
	散歩	3.53	3.45	0.98
	BBQ	3.87	6.56	1.69
淀川	全ての投稿	25.74	96.34	3.74
	花火, fireworks	8.26	89.38	10.82
	大会	5.16	57.13	11.07
	BBQ	3.18	6.01	1.89

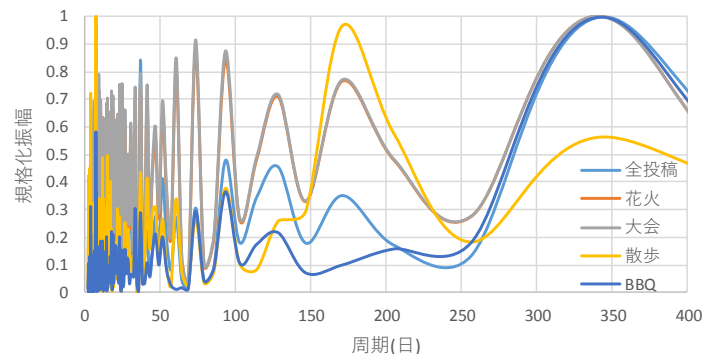


図-1 多摩川的全投稿と抽出した単語の FFT 解析

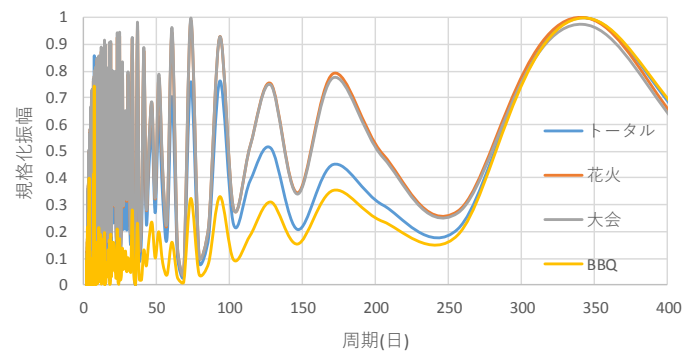


図-2 淀川的全投稿と抽出した単語の FFT 解析

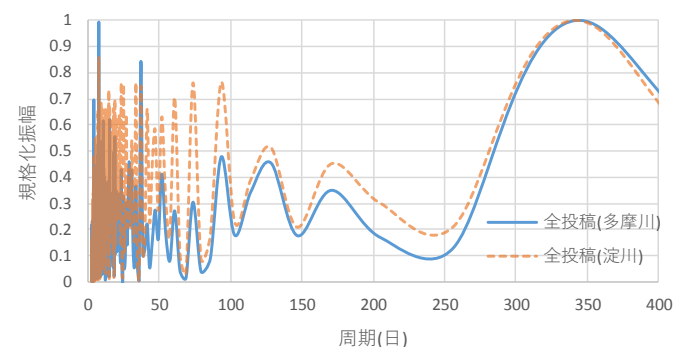


図-3 FFT 解析による多摩川と淀川 の全投稿比較