テキスト解析手法を用いた河川概念に対する関連単語分析

○東北大学大学院工学研究科　学生会員　安西　聡
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数多く行われている.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]．また，川島らは有識者の講演に基づいて，河川概念の構造化を行っていた [2]．しかしながら，これらの研究は限定的なデータを用いて行っており，地域的制限や時間的連続性の制約を受ける．そこで，ソーシャルメディアデータを一次データとして評価を試みる．同様の理由から，川守田ら [3]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，単語の使用回数のみを考慮していた．そこで，本研究では単語間の関連性に着目し，河川概念を把握することを試みた．

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，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#川がついた投稿のテキストを取得し，解析を行った．

表—1：カテゴリー覧

カテゴリ	代表的な言葉		
拠点型活動	キャンプ	花火	
写真一般	写真	カメラ	
線形移動型活動	散歩	ドライブ	
河川内活動	川遊び		
非日常型	旅行		
人間系	子供	友達	
河川一般	川	河	
景観	風景	景色	
動物	鳥	鮎	
植物	木	森	
山一般	山	渓谷	
海一般	海		
天候	空	天気	
土木構造物	橋		
都市部地名	京都	東京	大阪
地方部地名	長野	四国	
河川名	多摩川	鴨川	
河川周辺	河川敷		
観光地	温泉		
春	春	桜	
夏	夏	お盆	
秋	秋	紅葉	
冬	冬		
朝	morning	朝	
夕方	夕焼け		
夜	夜景		

3. 手法

#川が付いた投稿のテキストを MeCab によって各単語に分けた．MeCab は日本語の形態素解析エンジンである．例えば「すもももももものうち」を「すもも/も/もも/も/もも/の/うち」に分解できる．また，その際に各単語の使用率(使われている回数÷全投稿数)とその品詞を抽出した．各単語について使用率が 0.01 以上でありかつ，品詞が固有名詞，一般名詞，サ変接続名詞である単語を抜き出した．MeCab によって形態素解析する際に用いられる IPA 辞書の IPA 品詞体系では「一般名詞」や「固有名詞」，「サ変接続名詞」は他の語の一部でない名詞とされている．これは単一の語から，意味の読み取りが可能である．よって，「一般名詞」や「固有名詞」，「サ変接続名詞」のみを抜き出した．抜き出した各単語を意味の近いものでまとめ，カテゴリを作成した．表-1 に作成したカテゴリの一覧

キーワード　Instagram，ソーシャルメディア，共起関係，ビッグデータ
連絡先　〒980-8579　宮城県仙台市青葉区荒巻字青葉 6-6-06　TEL:022-795-7459

を示す．カテゴリの作成の際には川島ら[2]や建設省土木研究所河川部都市河川研究室[4]，鶴田ら[5]を参考にした．単語間の関連性を定量的に評価するため，以下の手法を用いた．作成したカテゴリ内の単語が投稿中に使われているかを数えた．その結果を用いて，投稿中に複数のカテゴリが使われている場合，そのカテゴリ間に関連性があるとし，関連回数に数えた．その結果を集計した．

4. 結果・考察

全ての関連性を描画することは図からの意味の読み取りが難解になる．そこで，今回は関連回数が 4500 以上のものだけ**図−１**を示した．各カテゴリのノードの大きさは出現回数の数を示している．出現回数が大きくなるほどノードの大きさも大きくなる．また，各カテゴリ間をつなぐエッジの太さは関連回数の数を示している．関連回数が大きくなるほど，エッジの太さは太くなる．結果として，「河川一般」と「天候一般」のカテゴリがもっとも強い関連性をもっていた．よって，河川概念に気象条件が強く，影響していることが分かった．

図−1 において，「河川一般」が他のカテゴリに強く影響している．また，「天候一般」も「河川一般」と関連性が強いため，その他のカテゴリと強い関連性を持つと予想できる．これら二つを描画することは全体の見通しを悪くする．そこで，「河川一般」と「天候一般」を除き，関連回数が 2500 以上の結果を**図−2** に示した．結果として風景に着目している群と活動に着目している群に分けてられた．拠点型の活動は主に夏季との関わりがあった．地方部で川の風景を楽しむ際には山との関わりが強い．また，「写真一般」と「ヒト・カラダ」のカテゴリが活動に着目している群と風景に着目している群をつないでいることが理解された．

5. まとめ

本研究から以下の結論を得た．

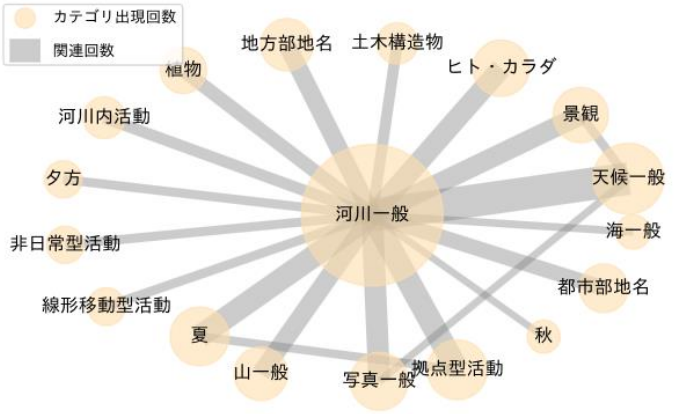
1) 気象条件が河川概念に強く影響している．
2) 河川概念は大きく二つに分けて活動に着目したものと風景に着目したものに分けられる．
3) 活動は夏との関連性強いが，場所に関連したものは少ない．
4) 地方部で好まれる風景は山との関連性が強かい．

謝辞

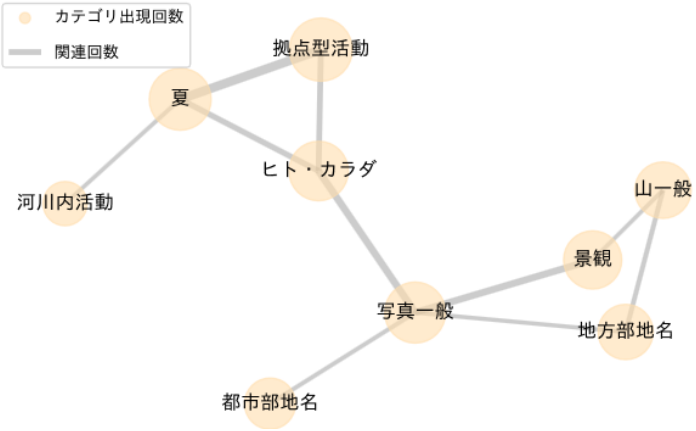
本研究の一部は，科学研究費補助金（15K14036，代表：風間聡）の助成を受けたものである．また，東北地域づくり協会の援助を受けた．ここに謝意を表する．

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) 川島隆徳，高田知紀，桑子敏雄，村井源，往往彰文：テキスト解析手法を用いた河川文化概念の構造化，情報知識学会誌，Vol.24 No.1, 2014
3) 川守田智，安西聡，風間聡：ソーシャルメディアを用いた河川関心度評価，水文・水資源学会誌，Vol.30, No.4, pp.209-220, 2017.
4) 建設省土木研究所河川部都市河川研究室：通常時の河川における人間活動（親水活動）と河川構造調査報告書，土木研究所資料第 2206 号，1985
5) 鶴田舞，中村晋一郎，萱場祐一：子どもが描いた川の将来像は川づくり計画に有効か?, 応用生態工学，20(1) , 107-115, 2017



図−1：関連回数 4500 以上の全カテゴリ間の関係



図−2：「河川一般」と「天候一般」を除いた関連回数 2500 以上のカテゴリ間の関係