

Twitter のリツイート情報に基づく観光情報アカウントの比較分析

The Analysis of Retweet Information of Tourism Account by Twitter

三田村 保^{*1}土田 邦彦^{*1}渡辺 功^{*1}大堀 隆文^{*1}鈴木 康広^{*1}

MITAMURA, Tamotsu

TSUCHIDA, Kunihiro

WATANABE, Isao

OOHORI, Takahumi

SUZUKI, Yasuhiro

近年、インターネット上での情報発信が活発となっており、北海道の基幹産業である観光産業においても、Twitter やブログ等のソーシャルメディアからの情報発信が増えており、個人観光客においてもその利用を積極的に行うことが多くなっている。本研究では、Twitter により観光に関連するデータ収集を行い、北海道に関する Twitter アカウントのリツイート情報に着目し、対象アカウントと類似アカウントの差異を抽出し、その比較検討を行った。

キーワード：北海道観光、Twitter、テキストマイニング

1 はじめに

近年、インターネットの急速な普及によりホームページやブログなどによる情報発信や Twitter や Facebook などのソーシャルメディアなどを利用した、ユーザーが主体となって情報を発信する機会が増加している。

総務省平成 24 年度情報通信白書では、情報通信機器の普及状況は、「携帯電話・PHS」及び「パソコン」の世帯普及率は、それぞれ 94.5%、77.4%と高く、「スマートフォン」は、29.3%（前年比 19.6 ポイント増）と急速に普及が進んでいる[1]。これにより、利用者もパーソナルコンピュータからスマートフォンなどの情報端末へと所有形態が移り、双方向の通信が活発になってきている。また、行政機関も東日本大震災時にはソーシャルメディアを通じてユーザーが直接情報を発信するなど災害や避難場所の情報をソーシャルメディアを通じて発信してきている。このように、ソーシャルメディアがこれまで関心を持たなかったユーザー層にも広く認知されるようになり、重要な情報受発信ツールとしてなりつつある。

観光に関しては、日本の長期的な成長戦略として観光が位置付けられるようになった[2]。観光産業は地域の活性化と一体であり、観光が復興にもたらす効果は大きく、人材や観光資源を有効的に活用することは不可欠であり、観光産業の重要性はますます高まっている。そのため地方自治体や観光産業においても、観光情報の発信、ホテルの案内、予約といったインターネットの活用によるサービスが提供されている。また、近年「サービス」を工学や科学の対象として捉え、サービスを研究する「サービス工学／科学」が注目を浴びており、観光分野においても様々な研究がおこなわれている[3]。

*1 北海道科学大学

著者らはインターネット上での北海道に関する観光情報に着目し、データの収集および分析を行っている[4~6]。これまでに Web 上に発信された北海道に関する観光情報の収集および分析を行い、Web 検索を利用した北海道観光情報を集積し、時系列グラフとして視覚化を行い、結果として市町村ごとの特徴や観光カテゴリに関する特徴を把握することができた。そして、ソーシャルメディアの中でも情報の受発信が主目的とされる Twitter を対象とした分析手法について研究を行っている[7]。この研究では、Twitter のアカウント間の類似性を定義し、類似アカウントのクラスタリングによって対象アカウントの特徴を抽出し、フォロワー/フレンド(被フォローアカウント/フォローアカウント) の関係から類似アカウントを抽出した。これはアカウントの情報発信や、そのライバル関係にあるアカウントを把握する際に有効な情報である。しかし、より詳細な分析には、ツイートの分析や時系列を考慮した分析が必要と考えられる。

本研究では、Twitter の対象アカウントと類似アカウントのリツイート情報に着目する。リツイートされた情報から特徴語を抽出し、その重みを算出する。算出された各特徴語に順位付けを行い、その特徴語のアカウント間での共通性の比較検討を行う。

2 研究背景

平成 18 年に「観光立国推進基本法」が成立、平成 19 年に「観光立国推進基本計画」が閣議決定されるなど、日本の長期的な成長戦略として観光が位置付けられるようになった[2]。観光産業は地域の活性化と一体であり、観光が復興にもたらす効果は大きい。人口減少にある我が国において人材や観光資源を有効的に活用することは不可欠であり、観光産業の重要性はますます高まっている。

また国際化や個人化が進む現代において、観光産業の活性化にはインターネットや情報技術の活用が不可欠である。インターネット上での情報発信の手法として、ホームページやブログなどの Web サイトを開設し、観光地に興味を持っている人(以下、興味者とする) の方から Web サイトに訪問してもらうサービスが提供されている。このサービスは、興味者に対する継続的なアプローチや、新たに興味者となる可能性のある人へのアプローチ機会が少ないという問題がある。

近年、Twitter や Facebook などに代表されるソーシャルネットワークサービス(SNS) を観光情報に活用する動きが多く見られる。例えば、ニセコ町では Twitter に町公式の Twitter アカウント(@nisekocho)を開設し、町の職員がニセコ町に関する観光情報や行政情報などを発信している。Twitter では観光産業者と興味者がフォローという関係で直接結ばれることより、継続的な情報提供が可能となり、情報を一方的に提供するのではなく興味者からの情報も得られる双方向コミュニケーションが可能となる。多くの観光アカウントでは、Twitter で興味者への質問に答えるなど、能動的な情報発信を行っている。また、情報が興味者にとって有益であればリツイートによって興味者のフォロワーにも伝搬され、ニセコ町と直接結ばれていない人に対して情報を伝えることも期待できる。

さらに、実際に観光地に訪れた人(以下、観光者とする) が観光情報の発信者となることも期待できる。これらの情報は観光者による「生の声」であり、興味者にとって有益な情報となり得る。近年では、ツイートに写真や位置情報を付加する機能も追加されており、観光者が発信する情報

の影響力がますます強くなっており、これらの情報を分析する研究の重要性も高まっている。しかし、観光と情報を結ぶ研究はいまだ発展途上にあり、今後さらなる研究が求められる。

近年、「サービス」を工学や科学の対象として捉え、サービスを研究する「サービス工学/科学」が注目を浴びており、観光分野においても様々な研究がおこなわれている[3]。

SNS を対象とした研究もおこなわれており、Twitter の興味分析の研究として、遠藤らはツイート情報を用いたユーザー間の興味の重なりを研究しており、返信(リプライ)によるソーシャルグラフの構築と解析、及びタグ情報を用いた興味の重なり解析を行っている[8]。リプライを用いたソーシャルグラフではグラフの定量化と性質を把握することを目標としていたが、抽出したグラフからスモールワールド性やスケールフリー性を確認することができなかった。一方、興味の重なり解析ではツイート情報からタグ情報を抽出し、ユーザー間の距離と興味の重なりについて調査し、ユーザー間の距離が離れるほど興味の重なりが低くなることが確認された。

また、後迫らはフォロー関係による分類とその集団の可視化の研究を行っている[9]。後迫らは、Twitter に対しグラフ理論を適用することでフォロー関係の情報からユーザーを集団に分類する方式を提案している。具体的には、関連性の強いフォロワーを辿ることで、関連性の強いユーザーで構築されるネットワークを抽出し、Kamada-Kawai 法を用いてクラスタに分類し、プロフィール情報をもとに特徴を推定している。この提案方式では発言内容やプロフィールといった具体的な情報を用いることなくユーザーを意味のある集団に分類することに成功している。

著者らはインターネット上での北海道に関する観光情報に着目し、データの収集および分析を行っている[4~7]。これまでに Web 上に発信された北海道に関する観光情報の収集および分析を行い、Web 検索を利用した北海道観光情報を集積し、時系列グラフとして視覚化を行い、結果として市町村ごとの特徴や観光カテゴリに関する特徴を把握することができた[4]。

キーワード入力によって得られた URL や検索結果数を収集し、北海道観光情報の発信状況について、Web リンク構造の観点からの調査を行った[5]。さらに、ブログマイニングを利用して、収集したブログに観光キーワードがどの程度の頻度で出現しているかを調査し、観光とブログとの関連に注目して検討を行ってきた[6]。そして、現在はソーシャルメディアを観光産業に効果的に活用することが出来れば、地域の活性化に繋がると考えており、ソーシャルメディアの中でも情報の受発信が主目的とされる Twitter を対象とした分析手法について研究を行っている[7]。この研究では、Twitter のアカウント間のフォローされているアカウントをフォロワー、フォローしているアカウントをフレンドと定義し、分析対象とするアカウントのフォロワー/フレンドの関係から類似アカウントを抽出し、その特徴を抽出し、分析した。これにより、フォロワー/フレンド(被フォローアカウント/フォローアカウント)の関係から類似アカウントを抽出し、アカウントの特徴抽出が可能であることを示した。これはアカウントの情報発信や、そのライバル関係にあるアカウントを把握する際に有効な情報である。しかし、この研究ではツイートの内容分析や時系列を考慮した分析は行っておらず、より詳細な分析にはこれらの分析が必要と考えられる。

一方、徳久らはブログ記事からの評判分析としてテキストマイニングを行っている[10]。また辻井らはテキストマイニングを用いて宿泊レビューの分析を行い、注目情報の抽出を試みている[11]。荒木らは Web のブログ解析による定量的な解析を広告換算金額の補完に利用する提案を行

っている[12]. また、山本らは特定ジャンルのブログに対する共起とユーザー別特徴語抽出を用いた話題抽出の研究を行っている[13]. この研究では、特徴語抽出の手法として一般的に用いられている TF-IDF を特定ジャンルの各ユーザーに対して用い、ブログの共起関係を調べることによって対象期間の話題を抽出する手法を提案している. その結果、提案手法がより多くの特定ジャンルに関する話題を提示する点で他の手法を上回る結果を示しており、ブログ解析における共起語ペアのスコア算出の有効性を示している.

3 Twitter 観光分析システム

本章では、本分析で構築した観光分析システムについて述べる.

3.1 システム概要

本論文では分析対象のアカウントの「強み」と「弱み」となる情報の抽出を目指す. ここでいう「強み」と「弱み」とは、分析アカウントのタイムラインと他のタイムラインを比較し、分析アカウントの方が興味を持たれている情報を「強み」、他のタイムラインの方が興味を持たれている情報を「弱み」とする. 比較するアカウントには分析アカウントとフォロワー集合が類似するアカウントを使用し、その上位 N 件を類似アカウントとして定義する. 類似アカウントについては著者らの論文[7] で検討を行っており、類似度におけるクラスタリングでプロフィール情報毎のクラスタに分割できることを確認した. 類似アカウントの抽出方法については第 3.4 節で述べる. そして、分析アカウントと類似アカウントのタイムラインから特徴語となる単語を抽出し、それぞれを比較することによって「強み」と「弱み」を抽出する.

3.2 システム構成

本システムは以下の三つの機能によって構成される.

- データ収集機能
分析で必要となるデータを Twitter から収集する. このとき、データは Twitter API を使用して取得している. 取得したデータは必要に応じてデータベースやテキストファイルとしてローカルストレージに保存する.
- アカウント分析機能
フォロワー/フレンドの情報をもとに類似アカウントを取得する. 類似アカウントの抽出結果は HTML 形式で出力され、ブラウザで確認することが出来る.
- 特徴ツイート分析機能
アカウント分析機能の結果を基に、分析アカウントと類似アカウントのツイートを分析する. 分析はツイートの取得、対象ツイートの抽出、特徴語の抽出、重み付け、比較の順で行われ、分析結果は CSV 形式で出力される.

3.3 データ収集

各機能が要求する内容に応じ、Twitter API を使用してデータを取得する. アカウント分析機能で

は、アカウントに関する情報として以下の情報を必要とする。

- 分析アカウントのプロフィール情報
- 分析アカウントのフォロワー集合
- フォロワー集合のフレンド集合
- フレンド集合のプロフィール情報

アカウント分析機能は分析者から分析アカウントの名称を受け取り、その名称をデータ収集機能に渡す。データ収集機能は Twitter から渡された名称のプロフィール情報を取得する。このレスポンスにエラーがある場合、分析アカウントが無効なアカウントと判断してシステムはエラーを返す。次に、分析アカウントのフォロワー集合と、その各アカウントのフレンド集合を取得しデータベースへと保存する。そして、最終的な出力結果を確認するために、各フレンド集合のプロフィール情報を取得し、アカウント分析機能へと処理を返す。

ツイート分析では、分析アカウントと類似アカウントの取得可能な全てのツイートを取得する。TwitterAPI では、取得可能なツイート数に上限が定められており、最大 3、500 件まで過去のツイートを取得することが可能である。なお、ツイートを取得する API のレスポンスには、ツイート本文の他にユーザー情報やリツイートされた回数などの情報も含まれている。

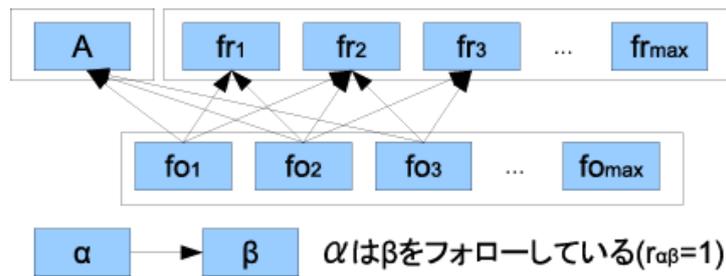


図 1: 類似アカウントの抽出

3.4 類似アカウント抽出

分析アカウントのフォロワー関係を図 1 に示す。類似アカウントはフォロワー関係において、共有するフォロワーの傾向によって以下の手法で抽出する。

1. 分析対象のアカウント A のフォロワー集合 FO と、集合 FO の各アカウントのフレンドを集合 FR を以下に定義する。ここで、アカウント A は集合 FR にも属するとする。

$$FO = \{fo_i | i = 1, 2, \dots, fo_{max}\} \quad (1)$$

$$FR = \{fr_j | j = 1, 2, \dots, fr_{max}\} \quad (2)$$

2. フォロー関係 R を以下に定義する。

$$R = \{R_{ij} | i \in FO, j \in FR\} \quad (3)$$

$$r_{ij} = \begin{cases} 1: fr_j \text{ は } fo_i \text{ のフレンドである} \\ 0: \text{それ以外} \end{cases} \quad (4)$$

3. アカウト間の類似性 S_{ij} を以下に定義する. 類似性 S_{ij} はアカウント i とアカウント j のフォロワーとフレンドの共通アカウントを用いて算出する. 1 から 0 の実数値を取り、値が大きいほどアカウント間の類似性が高いことを示す.

$$S_{ij} = \frac{\sum_{k=1}^{f_{max}} |\min(r_{ik}, r_{jk})|}{\sum_{k=1}^{f_{max}} |\max(r_{ik}, r_{jk})|} \quad (5)$$

4. 類似アカウント抽出件数を件数 N とし、アカウント A と集合 FR の類似性 S_{Aj} の上位 N 件をアカウント集合 S_N とする.

$$S_N = \{s_k | k = 1, 2, \dots, N\} \quad (6)$$

3.5 特徴ツイートの分析

本論文では以下の手順でツイートの分析を行う.

- 対象ツイートの抽出
- 特徴語の抽出
- 重み付け
- 比較

はじめに、分析対象となるツイートを抽出する. Twitter には大量の情報が投稿されているが、その中でユーザーが注目する情報は一部だけである. そのため、本論文のようにユーザーの興味情報を分析する場合、全てのツイートを対象とする分析では効率が悪く、また分析結果を絞り込むのも困難となる. そこで、本論文ではユーザーから一定の注目を集めているツイートを対象に分析を行い、その閾値として公式リツイートが行われた回数を使用する. 公式リツイートとは、ユーザーが自身のフォロワーに情報を伝搬する機能であり、一般的にユーザーがツイートに対して興味を持った時に行う動作である. なお、ツイートに RT を付けて再投稿する非公式リツイートと呼ばれるツイートも存在するが、Twitter API で取得することができないため本論文では考慮しないものとする.

次に、対象ツイートから特徴語を抽出する. 特徴語抽出では、対象ツイートから特徴語となる形態素を抽出するために形態素解析を行う. 形態素解析を行うテキストは対象ツイート内から単独では意味を成さない単語、URL、ハッシュタグ(#で始まる語)、リプライ(@で始まる語) を除いたテキストとする. 形態素解析エンジンには MeCab[14] を使用し、MeCab の辞書には IPA 辞書と Wikipedia の全記事のタイトルリストを固有名詞として登録した辞書を使用した. そして、解析結果から品詞情報が名詞(サ変接続、一般、形容動詞語幹、固有名詞) の形態素を抽出する. この時、連続で出現する名詞は複合語が分割されたものと考え、連続して出現した名詞は結合して 1 つの名詞として扱う. また、本論文では同一ツイート内で隣接している特徴語を共起しているものと考え、共起している単語の組を特徴語として抽出する. 例えば、「自然エネルギーの利活用による地域内経済効果」というツイートからは「自然エネルギー」、「利活用」、「地域」、「経済効果」の単語が抽出され、特徴語として(自然エネルギー、利活用)、(利活用、地域)、(地域、経済効果) が抽出される(図 2). なお、特徴語内での単語の出現順序は考慮しないものとし、出力結果

は特徴語を昇順に並べ替える.

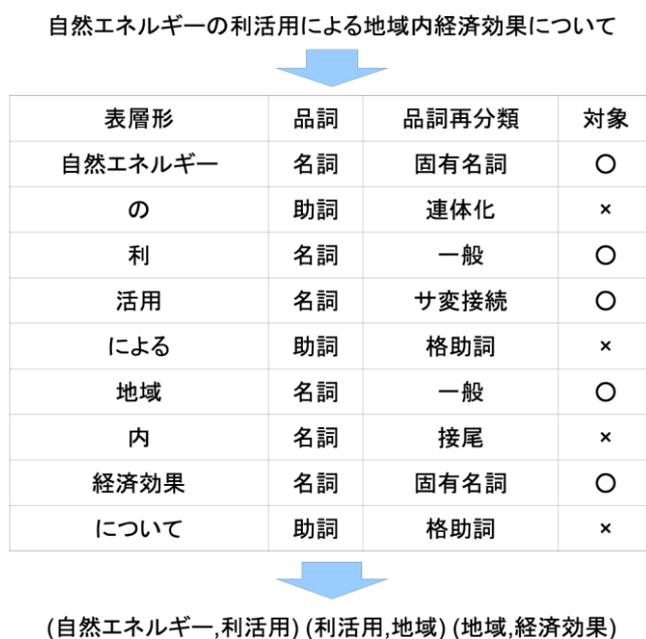


図 2: 特徴語の抽出

重み付けでは、タイムラインにおける特徴語の重みを算出し、各特徴語に順位を付ける。重みには特徴語の出現頻度を用いるが、類似アカウント集合では集合内で一般的となる特徴語の重みを下げるため、TF-IDF 法を用いた重み付けを行う。以下に TF-IDF 法の式を示す。

$$TF - IDF(w, d) = tf_{wd} \times \log_2 \left(\frac{N}{df_w} \right) \quad (7)$$

全アカウントの特徴語集合 d における特徴語 w の出現頻度を tf_{wd} とし、類似アカウント総数 N における特徴語 w を含む類似アカウント数を df_w とする。つまり、出現頻度の高い特徴語ほど重みが上がるが、多くのアカウントが共有する特徴語ほど重みは下がる。

最後に、分析アカウントと類似アカウントの特徴語を比較する。分析アカウントと類似アカウントにおける特徴語の上位 200 件から、特徴語の一方でも共通するものを取り除く。例えば、図 3 の例では、共通する単語は T3、T4 である。従って、分析アカウントの「強み」として(T1, T2)が、「弱み」として(T5, T6)が抽出される。



図 3: 特徴語の比較

4 結果・考察

4.1 分析対象

本分析では、アカウント分析の対象を北海道に関する情報を発信しており、 $f_{o_{max}}$ が 2,000～6,000 件程度アカウントとした。これらの条件を満たすアカウントとして、4 件の分析アカウントを表 1 のように定めた。なお、各アカウントのデータは 2012 年 11 月 1 日から 2012 年 11 月 3 日までの期間で収集し、分析に要したデータ容量は 7,378.1MB となった。

表 1: 分析アカウント

アカウント	平均ツイート数	$f_{o_{max}}$	$f_{r_{max}}$
nisekocho	2.04	3,774	3,489
OtaruCity	5.01	2,864	0
tv103	6.30	4,211	2
bemall sapporo	15.22	6,317	6,888

「nisekocho」はニセコ町役場の公式アカウントであり、ニセコの情報発信の他、まちづくりのため積極的に行政情報を発信している。「OtaruCity」は小樽市総務部広報広聴課が運営するアカウントであり、小樽市ホームページや小樽市 Facebook ページの新着情報をボット配信している。「tv103」はさっぽろテレビ塔をモチーフにした非公式キャラクターであり、ツイートの内容はフォロワーとのコミュニケーションが主体となっている。「bemall sapporo」は無料メルマガ「あなた情報マガジンびもーる」のアカウントであり、札幌近郊の地域情報を発信している[15]。

表 1 の平均ツイート数は 1 日あたりのツイート数の平均であり、総ツイート数をアカウント作成からの経過日数で除した値である。分析アカウントの中では bemall sapporo の平均ツイート数が最も高く、1 日の情報発信数が最も多いことになる。 $f_{o_{max}}$ は分析アカウントのフォロワー集合 FO の総数、 $f_{r_{max}}$ は集合 FO のフレンド集合 FR の総数である。本分析では類似アカウント抽出件数 N を 10 件とし、分析アカウントと類似アカウント S_{10} の特徴語を比較する。

4.2 類似アカウント抽出結果

類似アカウントの抽出結果を表2に示す。

表2: 類似アカウントの一覧

順位	nisekocho	OtaruCity	tv103	Bemall_sapporo
1	ニセコ町役場	小樽市	テレビ父さん	び〜助
2	北海道真狩村	北海道真狩村	えべチユン飼育係	北海道ファンマガジン
3	ニセコリゾート観光協会	旅なび! 網走(網走市観光課)	さっぽろテレビ塔	390 ダイニング&カフェバー BISCO
4	美唄市観光交流課	北海道上士幌町	コアックマ	北海道ディレクトリ
5	北海道上士幌町	淡路市役所	たら丸	吉田敏彦
6	小樽市	群馬県 嬭恋村	わたる	ラブちゃん
7	釧路市観光振興室	東京都八丈町	ツブキビタイツマン	SOC 総合研究所
8	旅なび! 網走(網走市観光課)	ニセコ町役場	まりもっこり	つい言ったあ〜北海道ネット枠(THN)
9	片山健也	加西市	キララちゃん	Hiroyoshi
10	クロマツツイッター【黒松内町北海道】	那須烏山市	ゴーヤ先生/福知山市みどりの親善大使	massivesapporo
11	北海道陸別町	三重県桑名市	ひこらぼ	街おこし札幌大好き!
12	社団法人小樽観光協会	長野県須坂市	シロモチくん	ひかりフォーム
13	札幌市観光コンベンション部	島田市	さけ太郎	Sapporo Campaign.com
14	十和田市役所	鱒ヶ沢町産業振興課	ゆるキャラコレクター	山本住宅ガイド札幌の不動産
15	北海道遺産協議会	西原町	アックマ	ポンパレ北海道
16	美瑛町観光協会@美瑛子	美瑛市観光交流課	つゆヤキソバン	まちなかアート
17	川上村	鳥羽市議会	801(やおい)ちゃん	キッカケマン@箕輪晃尚
18	淡路市役所	三島市	するめ〜@北海道福島町	北海道じゃらん
19	鳥取県伯耆町観光担当のつぶやき	北本市	おいちごちゃん	高橋料理店
20	福井県庁ふるさと営業課	小樽ジャーナル	サイパんだ!	平館典仁

nisekocho の類似アカウントとしてはニセコリゾート観光協会やニセコ町の町長である片山健也氏などのニセコ町に関するアカウント、真狩村や上士幌町などの北海道内の地方自治体や観光

協会アカウントなどが抽出された。また、類似度が下がるに連れて北海道外の地方自治体や観光協会アカウントが抽出されるようになる。

OtaruCity は nisekocho 同様に北海道の地方自治体や観光協会アカウントが抽出され、類似度が下がるに連れて北海道外アカウントが抽出される。これは、nisekocho と OtaruCity は互いに北海道の地方自治体アカウントであるため、抽出される類似アカウントの傾向も近く、類似度上位 100 件で共通アカウントが 55 件抽出された。

tv103 はさっぽろテレビ塔の公式アカウントとの類似性が高く、他の類似アカウントはゆるキャラが多数を占めていた。えべちゅん(江別市)、たら丸(岩内町)などの北海道ゆるキャラと類似性が高い傾向が見られたが、北海道ゆるキャラで Twitter カウントを取得しているキャラクター数は多くはない。従って、類似アカウント全体としては北海道外ゆるキャラが占める割合が高くなる。

bemall sapporo は北海道ファンマガジン、北海道ディレクトリなどの北海道に関する情報発信アカウントが抽出された。また、札幌を中心としたグルメ、ファッション、IT など様々な分野のアカウントが類似アカウントとして抽出された。

4.3 特徴語抽出結果

考察(アカウント nisekocho)

アカウント nisekocho の特徴語の抽出結果を表 3 に示す。対象ツイートとなった割合は nisekocho が 48.1%、S10 が 38.7%であった。nisekocho の「強み」として抽出された特徴語は緑の分権改革・地域資源の活用・環境エネルギーといった環境やエネルギーに関する話題、ニセコ駅周辺の中央倉庫群の活用といったニセコ町で検討されている行政情報、積雪による通行止めといった交通情報、そして町長の片山氏や副町長である北澤氏に関する話題も抽出された。一方、「弱み」として抽出された特徴語は天気や気象に関する情報、ブログや記事の更新情報、そして Facebook のアルバムや写真に関する話題が多かった。従って、nisekocho のタイムラインでニセコ町に関する写真を積極的に投稿したり、ニセコ町役場が運営する Facebook アカウントと情報を積極的に共有すると良いと考えられる。

考察(アカウント OtaruCity)

アカウント OtaruCity の特徴語の抽出結果を表 4 に示す。対象ツイートとなった割合は OtaruCity が 27.87%、S10 が 37.45%であった。OtaruCity の「強み」として抽出された特徴語は水道局・財政部・建設部などの一般競争入札、感染性胃腸炎の集団感染、企画展「養蜂とハチの世界」や「小樽の物産と観光フェア」といった小樽市のイベントに関する話題が抽出された。一方、「弱み」として抽出された特徴語は大雨洪水警報や土砂崩れといった災害情報、節電の協力要請、東北地方太平洋沖地震の被災者支援などであった。

考察(アカウント tv103)

アカウント tv103 の特徴語の抽出結果を表 5 に示す。対象ツイートとなった割合は tv103 が 24.20%、S10 が 17.17%であった。tv103 の「強み」として抽出された特徴語はゆるキャラグランプリなどのゆるキャラに関する話題、大通り公園の冬囲いやプロ野球日本シリーズに関する特徴語も抽出された。一方、「弱み」として抽出されたのは類似アカウントとして抽出されたゆるキャラに関する話題が中心であり、有益な情報が少なかった。各ゆるキャラが発する情報は自身に関する話題が中心であり、類似アカウント全体としてまとまった話題が少ない。従って、tv103 の様な独自性の強いキャラクタには本手法による分析は効果的ではなかったと思われる。

考察(アカウント bemall sapporo)

アカウント bemall sapporo の特徴語の抽出結果を表 6 に示す。対象ツイートとなった割合は bemall sapporo が 24.28%、S10 が 13.96%であった。bemall sapporo の特徴語の「強み」としては抽出された話題はライブコンサートの告知やボランティアの募集、餃子・カレーショップ「みよしの」とゲームソフト「龍が如く 5」のタイアップ、大通公園ビアガーデン・さっぽろオータムフェスト・狸まつりといった北海道のイベント、北海道大学で行われた初音ミクとどらえものイベントなども注目されていた。一方、「弱み」として抽出された特徴語は天気に関する話題や札幌以外の北海道内の地域の話が抽出された。これは、bemall sapporo は札幌を中心として情報を発信しているが、類似アカウントとして抽出されたアカウントは北海道を中心とし、札幌以外の地域に対する情報も発信しているためだと思われる。

5 おわりに

本研究では、分析アカウントと類似性の高いアカウントを抽出し、それぞれのリツイート情報に着目し、ツイートされた情報からアカウント間の比較検討を行った。分析アカウントと類似アカウントの特徴語を抽出し、比較を行うことによって分析アカウントの「強み」や「弱み」といった知見を得ることが出来た。しかし、ゆるキャラのようなアカウントでは、「弱み」となる知見の抽出が上手くいかなかった。これは、類似アカウントそれぞれの個性が強く、タイムラインに共通する話題が少なかったためであると考えられる。

今後の課題としては分析対象のアカウントの管理者に本研究の分析結果についてのヒヤリングを行い、更なる内容分析することが挙げられる。また、取得可能なツイート数に制限があるため、各アカウントのタイムラインに季節による差が現れた。従って、データを継続して収集することで、季節ごとの分析や、特定シーズンにおける分析も可能となる。

表 3: nisekocho の特徴語抽出結果

強み		弱み	
緑	分権改革	天気	気温
中央倉庫群	活用	写真	アップロード
通行止め	解除	facebook	アルバム
区間	通行止め	記事	掲載
ニセコ高校	雪山	プラス	気温
町長	片山	写真	撮影
積雪	通行止め	網走	流水
北澤	町長	ブログ	更新
中央倉庫群	活用基本設計検討委員	網走	観光協会
会議	傍聴	写真	紹介
公民館講堂	中央倉庫群	天気	回復
ニセコアワグラス	トム	ブログ	掲載
長官	講演	札幌	小樽
地域資源	活用	スズメバチ	巣
環境	エネルギー	道	駅
区間	積雪	hp	掲載
ニセコパノラマライン	共和	予報	日中
柏木先生	講演	釧路湿原	知床
ニセコ駅	綺羅乃湯	無料	定員
外国	職員	写真	掲載

表 4: OtaruCity の特徴語抽出結果

強み		弱み	
一般競争入札	水道局発注	通行止め	解除
一般競争入札	財政	営業	再開
一般競争入札	建設	台風	影響
小樽	水道局	節電	協力
感染	発生	影響	中止
財政	入札	画像	アップ
建設	入札	作品	展示
ハチ	世界	大雨	洪水警報
水道局	月別入札	車	運転
小樽	物産	番組	放送
物産	観光フェア	お誘い	来場
クリック	公告内容	道	駅
養蜂	ハチ	避難	開設
病院	月別入札	ユーストリーム	中継
総合博物館	アイアンホース	東北地方太平洋沖地震	被災
水道局	入札	通行	注意
空間放射線量	測定数値	被災	支援
感染性胃腸炎	集団発生	道路	通行止め
感染	種類	オープン	予定

表 5: tv103 の特徴語抽出結果

強み		弱み	
投票	ヒマ	沢山	お客様
天気	心配	素敵	プレゼント
外	気温	感謝	気持ち
びっくり	札幌	ゆるキャラ	集合
風邪	母さん	コアックマ	アックマ
ゆるキャラ同士	結婚	ステージ	邪魔
初雪	予報	記念	道内各地
予報	fuyu	道内各地	ゆるきゃら
プリーズ	心	やちにゃん	ゴーヤ先生
プリン大量	発注	入場	無料
発注	ツイート拡散	ゴーヤ先生	キララちゃん
いちごチョコ	抹茶チョコ	アックマ	まんべくん
冬囲い	イルミネーション	まんべくん	チュン
イルミネーション	ツリー	キララちゃん	ぎっちよ
ビアブリッツ	好き	案内	映像
中田翔	ホームラン	ゆるキャラ	続き
試合	風呂	友達	ツブキビタイツマン
野球	試合	ツブキビタイツマン	北海道放送 hbc
試合	札幌ドーム	北海道放送 hbc	もんすけ
札幌ドーム	序の口	もんすけ	写真

表 6: bemall sapporo の特徴語抽出結果

強み		弱み	
ライブ	コンサート	気圧	影響
コンサート	pv	最高気温	記録
ボランティアスタッフ	エキストラ	マリオ出現	マリオカート
みよしの	龍が如く 5	出会い	場
龍が如く 5	登場決定	クマ	目撃
動物	観察	白	滝
観察	円山動物園	夏	節電
夏恒例	大通公園ビアガーデン	ランチ	一緒
発想	ドラえもん	うどん	そば
ドラえもん	実装	旭川	マルシェ
実装	人工知能	地域	大雨
グルメ	ラーメン	熱中	注意
ラーメン	オータムフェスト	フラミンゴ	旭山動物園
ナイトバーゲン	縁日	マリオカート	いしかり
縁日	パフォーマンス	いしかり	見聞
パフォーマンス	狸まつり狸小路	テレビ	取材
北海道最大	花火大会	氷	トンネル
花火大会	真駒内花火大会	東川	記録
初音ミク	発想	美瑛	白
便利特典	観光マップ	小樽	岡川薬局

参考文献

- [1] 総務省: 情報通信白書平成 24 年度版(2012)
- [2] 国土交通省 観光庁: 平成 23 年版観光白書(2011)
- [3] 山本雅人: 「観光とサービス工学/科学」特集にあたって, 観光情報学会誌, Vol. 7, No. 1, pp. 9(2011)
- [4] 三田村保, 大町清隆, 大堀隆文: Web 検索による北海道観光情報の分析, 観光情報学会誌, Vol. 1, No. 1, pp. 47-53(2005)
- [5] 三田村保, 斎藤翔太, 大堀隆文: Web マイニングによる北海道観光情報の調査, 観光情報学会誌, Vol. 3, No. 1, pp. 62-71 (2007)
- [6] 三田村保, 岩佐渉, 湯川恵子, 大堀隆文: ブログを利用した観光情報の調査分析, 観光情報学会誌, Vol. 4, No. 1, pp. 57-65 (2008)
- [7] 桑野孝光, 三田村保, 渡辺功, 鈴木康広, 大堀隆文: Twitter を利用した観光情報の調査分析, 観光情報学会誌, Vol. 8, No. 1, pp. 27-38 (2012)
- [8] 遠藤福富美, 武田利浩, 平中幸雄: Twitter ユーザ間の興味の重なるの解析, 第 10 回情報科学技術フォーラム, pp. 481-482 (2011)
- [9] 後迫康宏, 大久保諒, 森井昌克: Twitter のフォロー関係による分類とその集団の可視化, 第 10 回情報科学技術フォーラム, pp.585-588 (2011)
- [10] 徳久雅人, 奥村秀人, 村田真樹: 観光開発支援のためのブログ記事からの評判分析, 観光情報学会誌, Vol. 7, No. 1, pp. 85-98 (2011)
- [11] 辻井康一, 津田和彦: テキストマイニングを用いた宿泊レビューからの注目情報抽出方法, 情報処理学会デジタルプラクティス, pp. 289-296 (2012)
- [12] 荒木長照, 田口順等: パブリシティ効果測定における広告換算金額とブログ分析についての考察, 観光情報学会誌, Vol. 9, No. 1, pp. 9-20(2010)
- [13] 山本尚央, 佐藤進也, 菅原俊治: 特定ジャンルのブログに対する共起とユーザ別特徴語抽出を用いた話題抽出, 第 9 回情報科学技術フォーラム, pp. 503-506 (2010)
- [14] MeCab: Yet Another Part-of-Speech and Morphological Analyzer:
<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>
- [15] あなた情報マガジンびも〜る:
<http://bemall.jp/>

(査読論文 2013年10月16日受理)