

## クックパッドデータからの道産食材を使用したレシピの調査分析

### Survey and Analysis of Recipes using Hokkaido Foods from CookPad Data

三田村 保<sup>†</sup>

MITAMURA, Tamotsu

<sup>†</sup>北海道科学大学

北海道の将来にとって、「観光」「農業」の発展は重要である。しかし、新型コロナウイルス感染症（COVID-19）の感染拡大により地域経済、特に観光業への影響が大きく、対策が急務となっている。本稿ではクックパッドデータよりオープンデータとして公開されているレシピを用いて北海道を代表する農畜産物に対する分析を行い、特徴を分析し、観光・食育に関する検討材料を提供する。

キーワード：クックパッドデータ、北海道観光、農畜産物、食、IT

#### 1. はじめに

北海道の将来にとって、「観光」「農業」の発展は重要である。しかし、新型コロナウイルス感染症（COVID-19）の感染拡大により地域経済、特に観光業への影響が大きく、対策が急務となっている。近年、日本は産業の構造改革を進めている中、北海道では「新・北海道ビジョン推進方針」の中の産業創造プログラムとして、「農業」「観光」の発展を基本政策の一つに挙げて政策を推進している[1]。

北海道の基幹産業である農業分野として、平成30年に「道産食品輸出1,000億円」の達成をめざし、ASEAN諸国等における購買力の増加や北海道への関心の高まりを確実に捉え、官民連携により、道内各地の地域資源を活かし、道産農畜産物の新商品の開発や販路拡大を推進し海外需要の積極的な取り込みを図っている。観光振興は北海道の中核産業として重要な役割に位置付けられており、日本政府の国際観光政策に呼応し、「北海道外客来訪促進計画[2]」、「北海道観光のくまのつくり条例[3]」を制定し、東南アジアを重点対象地域として、積極的な誘致活動に取り組んでいる。主な事業としては、地域資源を活かした多彩なツーリズムの展開や外国人観光客の受入環境の整備を進め、地域の特産品を使用するなどの「ご当地グルメ」を利用して各地がまちおこしに利用している事例がある[4]。

このように観光振興は、北海道のみならず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[5]。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利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[6]がおこなわれており、今後、北海道の農業および観光の発展には、農畜産物を消費者にどのように受け入れられているかなどの調査が必要となり、消費者の行動等の情報から嗜好等の様々な特性を抽出し、得られた知見を活用した新しい施策の計画が望まれる。このように北海道の「観光」「農業」の分析には情報技術のさらなる利活用が望まれ、かつ情報技術分野においてもディープラーニングに代表される人工知能やビッグデータ処理技術の利活用に

---

<sup>†</sup>北海道科学大学

よる産業への応用についての研究が活発化している。

ビッグデータ処理技術は近年発展してきており、かつ民間企業のデータを国立情報学研究所 (NII) が「情報学研究データリポジトリ (IDR)」を通じて研究者に提供するというオープンサイエンスの推進に関する取り組みを始めている [7, 8]。クックパッド株式会社は日本最大の料理レシピサービス「クックパッド」を運営している [9]。クックパッドデータとはクックパッド株式会社が国立情報学研究所と協力して研究者に提供しているデータセットで、172 万品のレシピやそれらからなる献立に関するデータが含まれている。

本稿ではクックパッドデータより北海道を代表する農畜産物がレシピに使用されている特徴を調査する。第 2 章では、北海道の将来プランとしての「観光」「食」について述べ、本稿の提案内容についてその位置づけを説明する。第 3 章では、本稿がクックパッドデータを利用した調査分析について述べる。第 4 章では、第 3 章で明らかにした分析結果に対する提案について詳述する。第 5 章では、結論と今後の展望をまとめる。

## 2. 北海道の「観光」「食分野」「食育」「IT」

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 [5]。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。

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 [10~14]。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 [10]。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 [11]。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 [12]。また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た [13]。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および特徴を抽出し、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。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っている [14]。

北海道の観光については地域の特産物を活かした地域の魅力づくりが必要である。北海道は、高い食料供給力を持ち、農業の産出額や水産業の漁獲量・漁獲高が全国 1 位であり、カロリーベースの食料自給率は 190%前後を保ち続けている [15]。この農水産物やこれらの加工品、流通や販売などを観光と結び付けることが必要であると思われる。

北海道の第 3 次北海道食育推進計画 (平成 26 年 3 月発表) に掲載されている平成 24 年度と平

成 25 年度に生産量全国一の主な農畜産物を示す (表 1) [16]。

表 1 北海道の代表的な農畜産物

No.	名前
1	てんさい
2	いんげん
3	小豆
4	馬鈴薯
5	小麦
6	たまねぎ
7	かぼちゃ
8	スイートコーン
9	そば
10	ながいも
11	大豆
12	にんじん
13	アスパラガス
14	ブロッコリー
15	だいこん
16	生乳
17	牛肉

本稿では、北海道の観光と食分野の発展のキーワードを発見するためにクックパッドデータの調査する。次章では、クックパッドデータを用いた分析について述べる。

### 3. クックパッドデータの調査

クックパッドデータの概要について述べる。クックパッドデータとはクックパッド株式会社 [1] が国立情報学研究所と協力して研究者に提供しているデータセットで、172 万品のレシピやそれらからなる献立に関するデータが含まれている。データセットの利用対象者は大学や独立行政法人など公的機関の研究者であり、国立情報学研究所 (NII) が運営する「情報学研究データリポジトリ」で利用申請することで取得することが可能となる [8]。

データベース (2015 年 2 月 24 日公開時点) は 12 個のテーブルから構成されており、内容はレシピに関する 6 テーブル、献立に関する 6 テーブルである。テーブル名と概要、テーブルのレコード件数を示す (表 2)。

表 2 テーブル一覧

テーブル名	情報	件数
recipes	レシピのタイトルや概要	1,715,595
ingredients	レシピの材料	12,725,006
steps	レシピの手順	8,849,850
tsukurepos	レシピのつくれば	9,449,479
search_categories	レシピのカテゴリ	1,099
search_category_recipes	レシピのカテゴリ	164,912
base_kondates	献立のタイトルや調理時間など	35,928
base_kondate_recipes	献立のレシピ	144,223
base_kondate_likes	献立の「参考になった！」	536,649
user_kondate_infos	献立の作者やポイント	35,928
kondate_categories	献立のカテゴリ	16
kondate_category_user_kondates	献立のカテゴリ	79,876

レシピデータと献立データから構成されており、レシピデータはクックパッドで 2014 年 9 月 30 日までに公開されたレシピに関するデータである。なお、レシピ、献立、つくれば（「作りましたフォトレポート」の略）の投稿者に関するユーザ ID はマスキングがされており、個人情報は公開されていない。レシピのタイトルや概要、手順、つくれば、カテゴリなどのデータが含まれている。献立データはクックパッドで 2014 年 9 月 30 日までに公開された献立に関するデータである。献立のタイトルや献立に含まれるレシピ、各レシピが主菜か副菜かといったデータが含まれている。データファイルは 7z 形式で圧縮した MySQL のバックアップファイルで、サイズは約 1.8GB（展開後は約 5.5GB）である。データベースの構築および分析には MySQL 6.0 CE を使用した [17]。データベースの分析には統計処理ソフトウェア R3.1.2 を使用した [18]。

レシピデータと献立データの主なファイルと概要と件数は表に示す。

本稿では、レシピテーブル (recipes) と材料テーブル (ingredients) を使用し、分析を行う。各テーブルの内容を示す (表 3、4)。

表 3 recipe テーブル

テーブル名	内容
id	レシピの ID
user_id	レシピ作者の ID
title	レシピのタイトル
description	レシピの分量
advice	レシピのコツ・ポイント
history	レシピの生い立ち
published_at	レシピの公開日

表 4 ingredients テーブル

テーブル名	内容
recipe_id	レシピの ID
name	材料の名前
quantity	材料の分量

次に、収集したデータから北海道の食材について分析した結果について述べる。材料テーブルから農畜産物の掲載されている件数を調査するために食材の特定に使用した検索語を設定した。各農産物がどの程度材料テーブルに掲載されているかと調査する上で、材料の名前に対する類義語を定義して、その類義も検索語として利用した。主な例は以下の通りである。

いんげん（インゲン）、小豆（あずき、アズキ）、馬鈴薯（じゃがいも、ジャガイモ、メークイン、男爵、だんしゃく他）、小麦（こむぎこ、コムギコ）、たまねぎ（玉葱、玉ねぎ、タマネギ他）、かぼちゃ（南瓜、カボチャ）、スイートコーン（トウモロコシ、とうきび、トウキビ、玉蜀黍）、蕎麦（そば、ソバ）、ながいも（長芋、長イモ、ナガイモ、長いも）、大豆（だいず）、にんじん（ニンジン、人参）、だいこん（大根、ダイコン）、生乳（牛乳、ミルク）、牛肉（牛にく）

また、「アスパラガス」「ブロッコリー」については省略した用語を使用しているため「アスパラ」「ブロッコリ」を検索語として使用した。

各食材の材料テーブルに掲載されている件数を示す（表 5、6）。表 5 は農畜産物の名前、レシピテーブルに材料として掲載されているレコード件数と全件数 1,715,595 に対する割合を示す。

表 5 よりたまねぎ（15.86%）、生乳（13.18%）、にんじん（10.94%）がレシピに使用されている割合が多い。各レシピに含まれている食材数別の件数を表 6 に示す。本稿で取り上げた北海道の農畜産物が材料に含まれているレシピ数は 787,490 件であり、全レシピ数 1,715,595 に対する割合は 45.9% であり、北海道の代表的な農畜産物がレシピに含まれる割合は高い。しかし、多くは 1、2 品のみレシピの中に使われているおり、その割合は 9 割である。

表 5 レシピテーブルの農畜産物のレコード件数

No.	名前	件数	全件 (1,715,595) に対する割合
6	たまねぎ	272,089	15.86%
16	生乳	226,075	13.18%
12	にんじん	187,731	10.94%
5	小麦	108,038	6.30%
15	だいこん	83,899	4.89%
4	馬鈴薯	83,077	4.84%
7	かぼちゃ	42,452	2.47%
14	ブロッコリー	29,943	1.75%
17	牛肉	18,021	1.05%
13	アスパラガス	17,579	1.02%
2	いんげん	17,215	1.00%
10	ながいも	12,232	0.71%
11	大豆	10,541	0.61%
3	小豆	6,374	0.37%
1	てんさい	3,890	0.23%
9	そば	2,876	0.17%
8	スイートコーン	2,079	0.12%

表 6 食材数別のレコード件数

食材数	件数	割合
9	1	0.00%
8	4	0.00%
7	60	0.01%
6	562	0.07%
5	3,240	0.41%
4	14,228	1.81%
3	53,117	6.75%
2	171,534	21.78%
1	544,744	69.17%
計	787,490	100.00%

食材の類似性抽出について述べる。レシピの中にもう一品北海道の農畜産物が使用することができればより北海道食材の使用率が高まるものと考えられる。また、親和性の高い食材の分析を行った。

表7 食材との相性の良い食材（上位5位）

v1	てんさい	生乳(19.7%)、小麦(8.2%)、にんじん(8.2%)、たまねぎ(6.2%)、かぼちゃ(5.3%)
v2	いんげん	にんじん(36.4%)、たまねぎ(25.3%)、馬鈴薯(12.5%)、だいこん(5.1%)、小麦(4.9%)
v3	小豆	生乳(24.9%)、かぼちゃ(7.1%)、小麦(4.8%)、たまねぎ(0.9%)、てんさい(0.9%)
v4	馬鈴薯	たまねぎ(46.4%)、にんじん(32.1%)、生乳(17.2%)、小麦(10.8%)、ブロッコリー(5.0%)
v5	小麦	生乳(26.1%)、たまねぎ(24.3%)、にんじん(10.1%)、馬鈴薯(8.3%)、かぼちゃ(3.3%)
v6	たまねぎ	にんじん(28.7%)、生乳(15.3%)、馬鈴薯(14.2%)、小麦(9.6%)、かぼちゃ(3.7%)
v7	かぼちゃ	生乳(27.4%)、たまねぎ(23.6%)、にんじん(12.1%)、小麦(8.4%)、馬鈴薯(4.9%)
v8	スイートコーン	たまねぎ(27.3%)、生乳(20.3%)、にんじん(15.8%)、馬鈴薯(11.3%)、小麦(7.2%)
v9	そば	だいこん(16.6%)、たまねぎ(9.1%)、にんじん(8.0%)、ながいも(5.0%)、生乳(2.1%)
v10	ながいも	小麦(6.9%)、たまねぎ(6.6%)、にんじん(6.3%)、だいこん(4.6%)、生乳(3.2%)
v11	大豆	にんじん(34.3%)、たまねぎ(26.6%)、馬鈴薯(6.1%)、だいこん(5.9%)、生乳(4.9%)
v12	にんじん	たまねぎ(41.6%)、馬鈴薯(14.2%)、だいこん(11.3%)、生乳(8.1%)、小麦(5.8%)
v13	アスパラガス	たまねぎ(23.2%)、にんじん(14.9%)、生乳(9.9%)、馬鈴薯(8.5%)、小麦(7.3%)
v14	ブロッコリー	たまねぎ(28.7%)、にんじん(24.2%)、生乳(15.9%)、馬鈴薯(13.7%)、小麦(7.0%)
v15	だいこん	にんじん(25.4%)、たまねぎ(10.0%)、馬鈴薯(2.9%)、小麦(2.8%)、生乳(2.0%)
v16	生乳	たまねぎ(18.4%)、小麦(12.5%)、にんじん(6.7%)、馬鈴薯(6.3%)、かぼちゃ(5.2%)
v17	牛肉	たまねぎ(43.7%)、にんじん(28.2%)、馬鈴薯(13.8%)、小麦(6.6%)、だいこん(5.5%)

食材のクラスター分析を行った。食材間の距離行列を算出し、樹形図（図1）を示す。

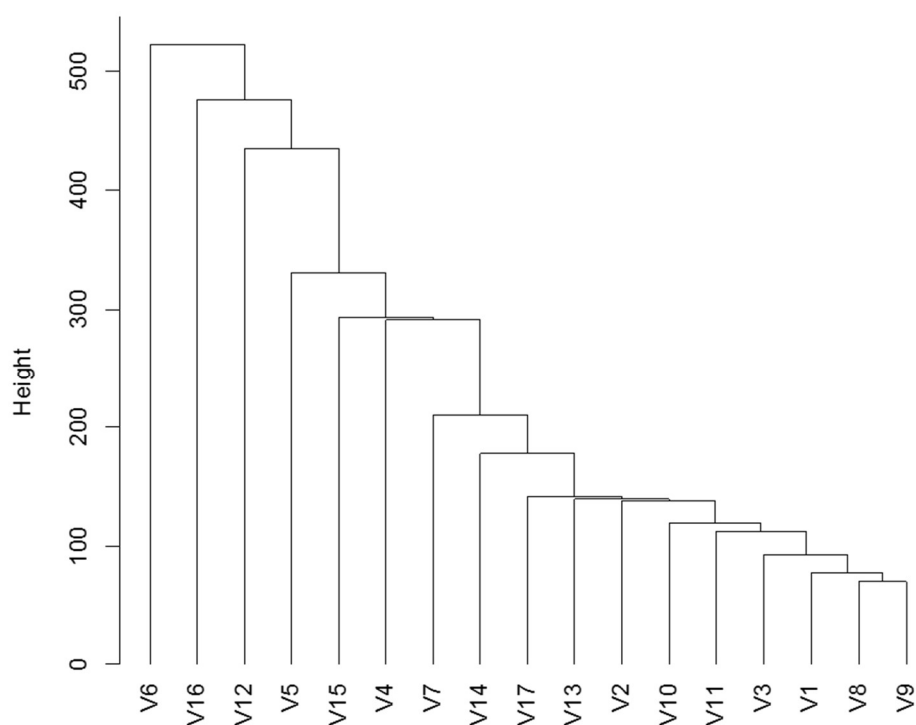


図1 食材との相性の良い食材（樹形図）

クラスター分析には最近隣法を用いた。結果、たまねぎと生乳、にんじん、小麦、だいこん、馬鈴薯が食材としての親和性が高かった。そば、スイートコーンが他の食材との親和性が低い結果となった。

#### 4. 考察と課題

前章で得られた結果より、クックパッドデータから北海道の食材に関する相関関係などが分析された。今後はそれらの食材をどのように利用したら良いか以下の事項を検討することが課題となる。

- 地域の食材に関するレシピ開発
- 食育と観光の融合
- 地域の名物レシピの開発

地域の食材に関するレシピ開発については、地域の食材を組み合わせた新たなレシピの開発をする必要があるのではないかと感じる。表1の「北海道の代表的な農産物」と表5の「レコード件数」を対照させると、例えば、「てんさい」「いんげん」「小豆」など、北海道の代表的な農産物でありながら、レシピへの取り込みが少ないことが見出せる。よって、これらの農産物に対して今後優先的にレシピ開発を展開するような政策を立案すべきであると思われる。本稿の今回の調査では特定の地域の農産物のブランド名は利用していない。そのブランド品の品質を活かしたレシピの分析は行っておらず、今後の課題である。

食育と観光の融合については、バランスの取れた食事の提供は今後、観光に関しては大事な要素となる。地産地消の観点からバランスと地域性を活かしたレシピの開発は重要である。

食育と観光の融合を図る必要があるのではないかと思う。修学旅行に食育を取り入れる取組が



行われており、非日常的な体験を通して自らの食のあり方を学ぶ機会を与えることができ、食生活改善のための意識を高めることにつながったとの結果が報告されている[19]。

地域の名物レシピの開発については、地域の農畜産物を利用して地域の食材を集めたレシピの開発をすることができる。地域の特定の食材に関するレシピの掲載状況を分析し、相性の良い食材が存在するのか、など新しいレシピの開発が可能となる。クックパッドでは「博多ブランド農産物をつかったうまかレシピ」募集など地域の農産物を使ったレシピの募集を行っており、農産物を使ったレシピ開発が行われている[20]。

## 5. おわりに

新型コロナウイルス感染症（COVID-19）の感染拡大により地域経済、特に観光業への影響が大きく、対策が急務となっている現在、本校は、北海道の農畜産物を活用したレシピ開発の一助となることを期待している。今後も北海道の将来性は「観光」「食分野」と「IT」の密接な関連が寄与されると思われる。

## 謝辞

本稿は、国立情報学研究所とクックパッド株式会社のデータを参考にした成果によるものである。

## 参考文献

- [1] 新・北海道ビジョン推進方針  
<http://www.pref.hokkaido.lg.jp/ss/ssa/tyougi/270831siryou1.pdf> (2020/5/19 accessed)
- [2] 北海道外国人観光客来訪促進計画  
<http://www.pref.hokkaido.lg.jp/kz/kkd/250-gaikyaku/gaikyaku.htm> (2020/5/19 accessed)
- [3] 北海道観光のくにつくり行動計画 [http://www.pref.hokkaido.lg.jp/kz/kkd/252-kodo\\_keikaku/koudoukeikaku.htm](http://www.pref.hokkaido.lg.jp/kz/kkd/252-kodo_keikaku/koudoukeikaku.htm) (2020/5/19 accessed)
- [4] 北海道のご当地グルメ  
<https://www.jalan.net/gourmet/010000/> (2020/7/27 accessed)
- [5] 国土交通省 観光庁：平成 23 年版観光白書(2011)
- [6] 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)
- [7] 大山敬三，大須賀智子：情報学研究資源としてのデータセットの共同利用，人工知能，Vol. 31，No. 2，pp. 254-261(2016)
- [8] <http://www.nii.ac.jp/dsc/idr/> (2020/5/19 accessed)
- [9] <http://cookpad.com/> (2020/5/19 accessed)
- [10] 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)
- [11] 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)
- [12] 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)
- [13] 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)
- [14] 三田村保，土田邦彦，渡辺功，大堀隆文，鈴木康広：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、北海道地域観光学会、Vol. 1，No. 1，pp. 18-33(2014)
- [15] 農林水産省：平成 24 年度都道府県別食料自給率の推移(2012)
- [16] 北海道食育推進プラン

[http://www.pref.hokkaido.lg.jp/ns/shs/data/advance/file/plan\\_hontai.pdf](http://www.pref.hokkaido.lg.jp/ns/shs/data/advance/file/plan_hontai.pdf) (2020/5/19 accessed)

[17] <https://www.mysql.com/jp/> (2020/5/19 accessed)

[18] <https://www.r-project.org/> (2020/5/19 accessed)

[19] 三田村理恵子：女子高校生を対象とした健康を考える観光～食育を取り入れた修学旅行～、北海道地域観光学会、Vol. 2, No. 2, pp. 10-19 (2015)

[20] 「博多ブランド農産物をつかったうまかレシピ」募集  
<http://cookpad.com/pr/contest/index/207> (2020/5/19 accessed)

(査読論文 2020 年 8 月 11 日受理)