

# 日本と中国における抹茶レシピの嗜好の分析

荒木孝治  
刘正阳

## 要旨

近年、中国で抹茶の人气が高まり、その消費量は年々増加している。中国市場で消費されている抹茶のほとんどは中国製であり、その品質は日本製の抹茶と比べて劣っていると認知されている。消費者は通常、より高品質かつ安全な食品を求めるが、抹茶に関しては、その品質の優劣を消費者が判別するのは難しい。現在、日本から中国への抹茶の輸出量は、台湾や香港と比べてかなり低い。こうした背景から、品質が良いことが認められている日本産の抹茶を中国へ輸出したり、日本の抹茶企業が中国の抹茶市場へ進出したりする余地がある。中国市場における日本産抹茶のさらなるブランド化や抹茶料理の魅力の普及なども進めていくには、中国人消費者の抹茶に対するニーズや嗜好を把握することが重要である。本研究では、日本と中国のグルメサイトに掲載された抹茶レシピをデータとして収集・分析し、日本と中国における抹茶に対する嗜好の異同を解明する。さらに、その情報に基づいて抹茶レシピの開発法を提案するとともに、中国市場への進出時における考慮すべき事項を抽出する。

## 1. はじめに

近年、中華人民共和国（以下、中国と記す）では抹茶が人気になり、抹茶市場が順調に成長し、良い経済効果をもたらされている（杜ら、2018）。しかしその市場において、日本産の抹茶の取り扱いほとんどない（ジャパンフードサイエンス、2017）。日本から海外への抹茶の輸出統計を見ても、上位20ヶ国中、中国は14位である（金額ベース）。上位には、台湾、シンガポール、香港があり、トップのアメリカ合衆国と比べて2%に満たない。<sup>1)</sup> また中国には、日本における宇治抹茶のような有名ブランドはない。刘ら（2014）は、中国産と日本産の抹茶

---

1) 日本茶輸出促進協議会の2019年度データによる (<http://www.nihon-cha.or.jp/export/news/2020/news200131.html>)。なお、従来、抹茶は緑茶の中に含まれて輸出通関統計で集計されていたため、その量はわからなかったが、2019年1月期より粉末状（抹茶+粉末茶）として分離されることになった。

の品質について調査した結果、日本産と比べて中国産の抹茶の品質は劣ると報告している。つまり、抹茶は中国で流行っているが、市場に出回っている抹茶は低品質であり、より高品質の日本産の抹茶はほとんど流通していないというのが現状である。この状況に対し、中国人は低品質の抹茶に満足せず、より高品質の日本製抹茶を求めていると判断できる。なぜなら、例えば、中国人が日本で様々な商品を爆買しているが、その中に抹茶製品も含まれるからである(村山, 2018)。

このような状況に対して、例えば静岡県焼津市にある製茶会社の大井川茶園は、衛生管理の徹底により付加価値を高めた抹茶を年10トン生産し、需要が伸びる中国や北米向けに輸出することを予定している(2018年10月現在)。同社は、品質管理の国際規格であるISO9001を取得済みであるが、より厳しい安全性が求められるFSSC 22000<sup>2)</sup>を目指すことにより、自社製品が付加価値の高い抹茶であることを国内外にPRしている。中国の卸売業者と提携を進め、現地スーパー約30社へ抹茶を供給することが決まっている(以上、日本経済新聞(電子版), 2018aより)。こうした動きから、日本の抹茶企業が、拡大している中国抹茶市場に進出することにより、日本産抹茶のシェアを拡大できる可能性が十分であると推測できる。

一方、抹茶は今や世界から注目される日本のスーパーフードであるが、将来中国国内で生産される抹茶に関して、「生産量だけでなく、技術も全体的に向上しており、抹茶の品質が上がればわざわざ日本から買わなくても良い日がくる」(ジャパンフードサイエンス, 2017)という考察もある。しかし、中国産抹茶の品質が向上しても、拡大している中国抹茶市場に、日本の抹茶企業が進出する余地はある。その理由の第1は、既に記したように現時点ではまだ中国産抹茶の品質が十分に信頼されていないことであり、第2は、日本の抹茶の歴史は長く、中国でより高品質な抹茶そのものを生産するのが簡単だとしても、それを包摂する文化を育むまでには時間がかかるからである。第3は、日本産抹茶には有名ブランドがあり、中国抹茶のブランド化には時間を要することにある。よって、日本産の抹茶が中国に進出し、市場でのシェアを拡大するためには価格による競争から離れ、高品質・日本ブランドであることを訴求しながらも、顧客となる中国人の好みを把握し、飲用に限らず、むしろ彼らが受け入れやすい食べ方をプロモートしていく必要がある。本研究では、その基礎となる情報を取得するために、中国と日本における抹茶(レシピ)に対する好みの異同を分析する。

日本と中国の抹茶に対する嗜好を知るためにこれら2国のグルメサイトにアップされているレシピを調べ、それらを比較する。データとしては、レシピで利用される食材、およびレシピに付与された説明文(テキスト)を取り上げる。食材の分析に加え、テキストデータの分析(テキストマイニング)を行うことにより中国と日本の抹茶に対する嗜好を抽出し、それらを比較することにより、日本の抹茶企業が中国に進出する際の有益な情報を獲得するとともに、中国

---

2) 一般財団法人 日本品質保証機構 [https://www.jqa.jp/service\\_list/management/service/fssc22000/](https://www.jqa.jp/service_list/management/service/fssc22000/)

の抹茶企業にとっても有益な情報となることが期待される。

本稿の構成は次のようになる。第2節で、抹茶の歴史・魅力などを紹介し、第3節で、収集したレシピデータの整理および分析を行う。食材データの分析では、アソシエーションルール分析、クラスター分析といった手法を利用する。テキストマイニングでは、共起ネットワークやクラスター分析、トピックモデリングを実行する。第4節では、総合的な考察、および具体的な提案を行う。おわりにでは、残った課題を提示する。なお、分析にはデータ解析環境R (R Core Team, 2019) を主に用い、テキストデータの分析ではKH Coder (樋口, 2014) も用いた。

## 2. 抹茶の魅力

### 2.1 日本と中国の抹茶の発展の歴史と現在の状況

#### (1) 抹茶の歴史

中国は歴史や古代文明に誇りを持っている。喫茶の歴史は千年以上あり、その習慣は、後漢(25~220年)の時代に現在の四川省近辺の茶の産地から始まったとされる(松井, 2008)。このように喫茶文化は中国を起源とし、それが朝鮮半島や日本、さらに世界へと伝播した。これに対し日本の喫茶の習慣もおよそ千年の歴史を持つ。布目(2001)によると、平安時代の初め、嵯峨天皇の在位時(809~823年)には飲茶は、中国から日本に渡来していた。このように日本と中国の喫茶の歴史はいずれも長いが、飲む形態は少し異なる。守屋(1992)によると、唐代(618~907年)のお茶は団茶(固形茶)が主流であったが、宋時代(960~1279年)になると粉末茶が台頭し、1100年頃には末茶<sup>3)</sup>がお茶の中心になっていた。しかし、抹茶法は継承されず、明代(1368年~1644年)の中期に至ると消滅した(布目, 2001)。明菴栄西は、宋時代に茶の製法や効能を日本に伝えた人物であるが、その喫茶法で用いられるのは抹茶であった<sup>4)</sup>。鎌倉時代初期にもたらされた抹茶は日本で普及し、今日の茶の湯にまで繋がる。つまり、抹茶法は中国で発祥したものの、明代に中国で消滅した一方、日本では栄西が持ち帰った抹茶法が今日まで継承されているのである。

#### (2) 抹茶の定義

上記のように抹茶は中国で発祥し、日本に伝わって継承されてきたが、中国では明代に一旦消滅した。そのため、現在、日本と中国における抹茶の定義がどのようになっているかを調べておく。公益社団法人日本茶業中央会(2019)は、抹茶について「碾茶(覆下栽培した茶葉を碾茶炉等で揉まずに乾燥したものを茶臼等で微粉末状に製造したもので、「碾茶」とは、「摘

3) 布目(2004)によると、末茶と抹茶は同じものであるが、「抹茶」は日本語で、中国語では「末茶」に相当する。そのため以下、抹茶と記す。

4) ウィキペディア「明菴栄西」<https://ja.wikipedia.org/wiki/明菴栄西>

採期前に柵施設等を利用して茶園をよしず、コモ、寒冷紗などの被覆資材で2～3週間程度覆った『覆下茶園』から摘採した茶葉を蒸熟し、揉まないで碾茶炉等で乾燥させて製造したものと定義している。一方、中国における抹茶の定義は、「采用覆盖栽培的茶树鲜叶经蒸汽(或热风)杀青后, 干燥制成的叶片为原料, 经研磨工艺加工而成的微粉状茶产品」(中华人民共和国国家标准GB/T, 2017)である。日本語では、「覆下栽培した茶葉を水蒸気(あるいは熱風)で蒸熟し、乾燥させた葉を原料とし、研磨技法によって加工された微粉末状の茶製品」となり、日本と中国における抹茶の定義は一致していると判断できる。

### (3) 抹茶の生産と消費の現状

日本において、抹茶の生産量と消費量についての明確な統計データはない。原口(2015)は、2015年の日本国内抹茶の生産量は約4000トンとし、そのうち茶道に用いられるのは全体の5%程度で、残りの95%は食品加工用であると記している。これに対し碾茶の生産量の2018年の実績は、3387トンとされており<sup>5)</sup>、少し違いがみられる。従来は、高級抹茶に対して主に茶道からの安定した需要があったが、近年は減少傾向にある一方、1980年代後半頃から食品の添加での利用が増大し、下級抹茶の生産が増加している(工藤, 2008)。この傾向は現在も継続していると思われる。

中国における抹茶の生産量についても明確な統計データはなく、張(2018)によると、2018年、中国の抹茶生産量が世界の生産量の70%を占め、年消費量は1000トンとなり、年々増加する傾向がある。また、尹・刘(2008)は、中国の抹茶市場において、抹茶の70%は食品加工用、20%は飲用、残りの10%は医薬品や化粧品に使われるとしている。

## 2.2 抹茶に含まれる成分、保健効果および魅力

### (1) 抹茶に含まれる成分と保健効果

医歯薬出版編(2019)によると、100gの抹茶には、タンパク質が29.6g(内、アミノ酸組成によるタンパク質:22.6g)、脂質が5.3g、炭水化物が39.5g(内、高分子水溶性植物繊維:6.6g、不溶性植物繊維:31.9g)、無機質(ナトリウム:6mg、カリウム:2700mg、カルシウム:420mg、マグネシウム:230mg、リン:350mg、鉄:17mg、亜鉛:6.3mg、銅:0.6mg)が含まれる。ビタミン類に関しては、A( $\beta$ -カロテン当量:29000 $\mu$ g、レチノール活性当量:2400 $\mu$ g)、E( $\alpha$ -トコフェロール:28.1mg)、Kが2900 $\mu$ g、B1が0.6mg、B2が1.35mg、ナイアシンが4.0mg、ナイアシン当量が12.0mg、B6が0.96mg、葉酸が1200 $\mu$ g、パントテン酸が3.70mg、Cが60mg含まれる。さらに、ポリフェノールの一種であるカテキンや、旨味に寄与するアミノ酸であるテアニンが含まれる。池ヶ谷ら(1984)は、価格別抹茶の成分を調べた結

5) <https://www.zennoh.or.jp/bu/nousan/tea/dekiru03.htm>

果、カテキンの含有量は8.85%～9.75%であり、テアニンの含有量は100g当たり1172.5mg～2263.4mgだとしている。つまり、抹茶の品質によって成分の含有量にばらつきがあるものの、カテキンやテアニンといった健康に良い成分を含んでいる。

上記の成分、および碾茶から作られることから抹茶には緑茶の保健効果が期待できる。それには、抗がん作用、コレステロールの吸収阻止、高血圧予防、アルツハイマー予防、心疾患予防、認知症の予防、脳卒中の予防、口臭予防、美肌効果、ダイエット効果、リラックス効果などがある(岩間, 2019)。しかも、緑茶以上に健康効果が期待できる。なぜなら、茶がもつ様々な微量栄養素や有効成分 ( $\alpha$ -トコフェロールや $\beta$ -カロテン, 食物繊維等) は、湯や水にわずかししか溶出しなため、多くは茶葉に残る。これに対して抹茶では茶葉自体を食べるため、茶葉の成分を多く吸収することができる(福司山, 2008)。

## (2) 抹茶の魅力

上記のように昔から抹茶は健康に良い薬や飲料と認識され、愛飲されてきた。しかし、抹茶は粉末状であるため、飲用以外に、調味料や添加物としてうどんやパン、ケーキなどに練り込むという利用法も注目されている。例えば、抹茶そば、抹茶チョコ、抹茶ケーキ、抹茶クッキーなど抹茶を用いた様々な加工食品がある。美味しい料理やスイーツを楽しめると同時に茶葉の栄養も摂ることができ、自然と健康も手に入れることができるというのが従来の喫茶法と異なる魅力である。だが、従来の菓子やスイーツの場合、肥満や糖尿病を心配する人はそれらを控える傾向がある。しかし、それらに抹茶を加えると、血糖値の急激な上昇を防ぐことができることがわかっている。例えば中村ら(2012)は、パンには小麦粉や砂糖由来の糖質が多く含まれているが、少量の抹茶を添加することで血糖値の上昇を有意におさえることができると記している。この事例から、パン以外の一般的な食べ物でも抹茶は血糖値の上昇を有意に抑制できると考えられる。

しかし、健康効果があっても、匂いが悪かったり、味が受け入れられなかったりすると、抹茶はそこまでブームになることはできなかった可能性が高い。原口(2015)は、抹茶の香りについて、被覆茶である抹茶は、煎茶などの露天栽培茶と比べて海苔の香り成分であるジメチルスルフィドが強く、これが良い香りをもたらし、味については、アミノ酸類がうま味や甘味を、カテキン類が渋みを、カフェインが苦みを生むとしている。このように、抹茶では香りと様々な味(甘味, 渋味, 苦味)を同時に楽しむことができる。

## 3. 中国と日本における抹茶の嗜好の異同

本研究では、中国と日本の抹茶に対する好みに違いがあるかどうかを分析するために、日本と中国のグルメサイトにアップされている抹茶レシピデータを分析対象とする。グルメサイト

としては、日本の「クックパッド」<sup>6)</sup>、中国の「美食杰」<sup>7)</sup>、「美食天下」<sup>8)</sup>、「下厨房」<sup>9)</sup>の3つ、計4つを取り上げる。

近年中国ではグルメサイトの人気が高まっている。図1に示すように2012年から2017年まで、中国のグルメサイトのアプリケーションソフトウェア（APP）のユーザー数は年々増加し、2017年には約5200万人に達すると推定されている（比达网，2017）。

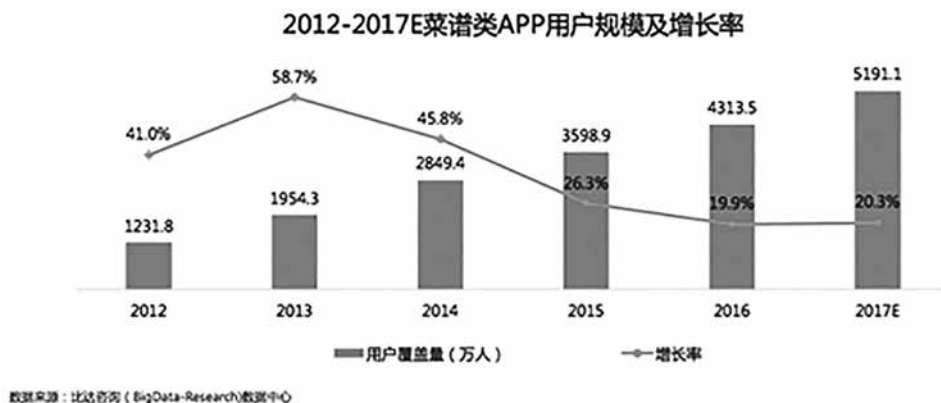


図1 中国におけるグルメAPPユーザーの規模および増加率（比达网，2017）

中国においてグルメ関連業界の競争は激しく、たくさんのサイトが存在する。サイトの人気に関しては、人気度指標が比达网で公開されている。2017年1月現在の各サイトの人気度を図2に示す。月当たりのユーザー数と一日当たりのユーザー数は下厨房の場合、各約900万人、約70万人、美食杰の場合、各約275万人、約23万人、美食天下の場合、約41万人、約3.7万人である（比达网，2017）。図3に一日平均利用回数と一回あたりの利用時間を示す。下厨房ではそれぞれ、2.5回、6.7分であり、美食天下では2.5回、5.9分、美食杰では2.2回、4.7分である（比达网，2017）。

グルメサイトの人気度の指標から、下厨房には人気がある。しかし、美食杰と美食天下に関しては、これらより人気のあるサイトも多くある。本研究で美食杰と美食天下を取り上げた理由は、人気があるグルメサイトが必ずしも抹茶に関するレシピの掲載量が多いとは限らないからである。実際、中国の全てのグルメサイトで抹茶に関するレシピ数を検索したところ、その数は美食杰（4000件）、美食天下（1028件）、下厨房（200件）の順となった（2018年9月27日現在）。レシピデータを分析するためにはデータ量がある程度必要となるため、サイトの人気

6) <https://cookpad.com/>

7) <https://www.meishij.net/>

8) <https://www.meishichina.com/>

9) <https://www.xiachufang.com/>

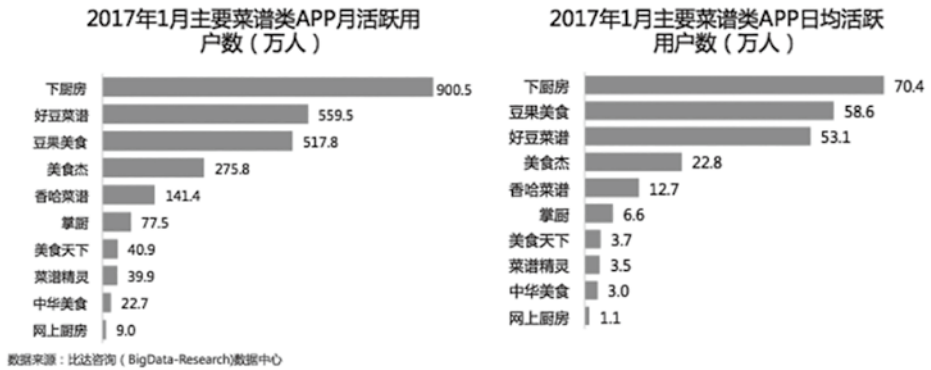


図2 2017年1月各グルメサイトのAPP利用者数(左)と日平均の利用者数(右)  
比达网(2017)より



図3 2017年1月各グルメAPPのユーザー一日平均利用回数(左)と一回あたりの利用時間(右)  
比达网(2017)より

のみではなく、アップされたレシピ数を含めて、総合的に考慮する必要がある。その結果としてこれら3つのサイトを取り上げた。

### 3.1 日本と中国の抹茶レシピの違い

#### (1) レシピから収集するデータの選択

レシピから取得するデータとして、「レシピ名」、「レシピに対する作者の説明」、「使用する材料」とその「使用量」、「公開日」の5項目を選択する。クックパッドの抹茶レシピで例示すると、図4のようになる(取得する情報は、枠で囲った部分)。中国のサイトについても、これらに対応するものである。

抹茶と卵白のガトーショコラ

バター、生クリーム不使用、しつとりふわふわで、甘さ控えめなガトーショコラです。

材料 (9人分)	
ホワイトチョコレート	85g
生油	50g
無調整豆乳	50g
卵黄 (M寸)	3個分
まじり粉	45g
卵白 (M寸)	3個分
まじり粉	45g
薄力粉	70g
抹茶	大さじ2

- (下準備)  
・卵白は細かく別んでおく  
・抹茶と薄力粉を合わせふる  
・型にペーパーを敷く
- ボウルに、米油、無調整豆乳を耐熱のボウルに入れて混ぜる。ヘラで混ぜて滑らかしたら冷ましておく。
- ボウルに卵黄ときき砂糖を入れ、泡立て器ですり混ぜる。白っぽくもつたりとしてきたら、②のボウルを加え良く混ぜる。  
な
- ヘラに代え、粉類を加え粉っぽさがなくなるまで混ぜる。
- 別のボウルに、卵白を入れハンドミキサーで軽くツノが立まで泡立て、まじり粉を4回に分けて泡立てる。その都度良く混ぜる。
- メレンゲ1/3量を④のボウル生地に入れて、ざっくり混ぜたりも混ぜたりしたら、型に流し込む。
- 170℃で熱したオーブンで約40分焼く。

このレシピの生い立ち  
あるCafeに行った時に、抹茶と卵白のガトーショコラがあって自分でも作れるかもと購入し、試してみました。

レシピID: 5201137 公開日: 18/08/09 更新日: 18/08/09

図4 収集する情報—クックパッド

2018年9月26日の時点でクックパッドにおける抹茶に関するレシピは17024品、同年同月27日の時点で中国の3つのグルメサイト美食杰、美食天下、下厨房（以下、この3サイトを総称して「中国」という）で各4000、1028、200品（計5228品）であった。

## (2) データの収集および前処理

クックパッドと中国のレシピを合わせると取得するレシピ数は2万以上となり、これらから5項目のデータを取得すると、単純に計10万項目以上となる。このような大量のデータをウェブから収集する場合、正確かつ効率的に行う必要があるため、データ解析環境Rを利用してスクレイピングした（石田ら、2017）。スクレイピングにより、各レシピの料理名、公開日、説明文、使用する材料とその使用量を収集し、食材については英語表記で統一してエクセルファイルにまとめた。次に、レシピによってはめったに使用されていない食材を含むことがあるので、比較的使用頻度が高く、一般的に認知されている食材のみにスクリーニングした。この結果、取り上げた食材は36となった。代表的なものは、matcha（抹茶）、milk、yogurt、egg、soymilk（豆乳）、butter、chocolate、cheese、redbeans（小豆）、cream、flour（小麦粉）、almondなどである。なお、レシピの作者が自由に食材を記入できるため、同じものに対して異なる表現を使っている場合がある。例えば「無調整豆乳」と「豆乳」、「卵黄（M寸）」と「卵黄」などである。このような表記揺れをなくするため、食材名を統一した。ただし、小麦粉に関しては、薄力粉、強力粉と区別せず、単に小麦粉と記載されたものがあつた。薄力粉、強力粉を分けて分析するため、重複が生じるがここではこれら3つの食材を区別して扱った。なお、整理したデータファイルの中に抹茶を使っていないレシピが存在したため、そのレシピを削除した。これが生じた理由は、抹茶でレシピを検索したとき、説明文の中に抹茶という語が単に



使われていたことによる。さらに、各食材の使用量について不明なものがあつたため、本研究では使用量にこだわらず、食材が使用されたかどうかという2値（0または1）データに変換することでこの問題を処理した。

こうした前処理の結果、クックパッドのレシピ数は15323に、中国では、1760（美食杰：598，美食天下：980，下厨房：182）となった。以降、このデータに基づいて分析していく。なお、前処理したデータにおいて、抹茶は全てのレシピに含まれるため、アソシエーションルール分析から食材としての抹茶を外す。

### 3.2 日本と中国の抹茶レシピデータの分析

#### (1) アソシエーションルール分析

アソシエーションルール (association rule) 分析は、マーケットバスケット分析とも呼ばれ、もともとは店舗のレジで自動的に収集されるPOS (Point Of Sales) データの分析のために開発された手法である。膨大なログデータから購入されたアイテム (項目) 間の意味のある関連性を抽出することができる。レシピに含まれる食材データでこの分析を行う目的は、レシピの中で各材料が使用される関連性を抽出することにある。

Lantz (2015) および山本ら (2015) を参考に、アソシエーションルール分析の考え方について簡単に説明しておく。スーパーの販売トランザクション (アイテム集合) を例とする。購入対象となる3つのアイテムの集合  $\{A, B, C\}$  があるとき、アソシエーションルールとは、購買におけるアイテム間の関連性のルール、例えば「 $\{A, B\} \rightarrow \{C\}$ 」のようなものである。ルールの左辺 (矢印 “ $\rightarrow$ ” の左で、条件部という) のアイテム集合  $\{A, B\}$  が、ルールの右辺 (結論部という) のアイテム集合  $\{C\}$  と関連性を持っているといった形のルールを求める。このルールは、AとBが同時に購入されると、Cも一緒に購入される可能性 (確率) が高いということを意味する。同時に購入された大量のアイテム集合の組み合わせをチェックすることにより有益なルールを抽出するには計算量が膨大となる。そのため、検討する組み合わせ数を絞るためのアルゴリズムが色々提案されている。その代表的なものとしてアプリアリ (Agrawal and Srikant, 1994) がある。アプリアリでは、あるアイテム集合  $\{A, B\}$  が頻出すれば、その中の各アイテム  $\{A\}$ ,  $\{B\}$  も頻出しなければならないので、もし  $\{A\}$  が頻出しないことが明確であれば、Aを含んでいる全てのアイテム集合を条件部の候補から削除するといった形で計算量を抑える。

アソシエーションルール分析では通常、多くのルールが抽出される。そのためルールを絞るための評価指標が必要となる。ここでは上記のルールの書き方を拡張し「アイテム集合Aを購入した客はアイテム集合Bも購入している確率が高い」ということを「 $A \Rightarrow B$ 」と記述する (A, Bは一つのアイテムではなく、アイテムの集合)。ルールの信頼度を評価する指標として、支持度 (support), 確信度 (confidence), リフト (lift) の3つが通常用いられる。以下、これら

3つの指標について説明する。なお、アイテム集合 $\mathbf{A}$ に含まれるアイテム数を $n(\mathbf{A})$ 、全アイテム数を $n(\mathbf{N})$ で表す。

支持度 (support) : アイテム $\mathbf{A}$ の支持度 $\text{Supp}(\mathbf{A})$ は、全アイテムの内、 $\mathbf{A}$ が含まれる割合であり、式(1)で定義される。また、ルール「 $\mathbf{A} \Rightarrow \mathbf{B}$ 」の支持度 $\text{Supp}(\mathbf{A} \Rightarrow \mathbf{B})$ は、式(2)で定義される。これは、全アイテムの内、 $\mathbf{A}$ と $\mathbf{B}$ を同時に購入したアイテムの割合である。

$$\text{Supp}(\mathbf{A}) = \frac{n(\mathbf{A})}{n(\mathbf{N})} = P(\mathbf{A}) \quad (1)$$

$$\text{Supp}(\mathbf{A} \Rightarrow \mathbf{B}) = \frac{n(\mathbf{A} \cap \mathbf{B})}{n(\mathbf{N})} = P(\mathbf{A} \cap \mathbf{B}) \quad (2)$$

確信度 (confidence) : ルール「 $\mathbf{A} \Rightarrow \mathbf{B}$ 」の確信度 $\text{Conf}(\mathbf{A} \Rightarrow \mathbf{B})$ は、式(3)で定義される。これは、 $\mathbf{A}$ を購入したという条件のもとで $\mathbf{A}$ も購入するという条件付き確率 $P(\mathbf{B} | \mathbf{A})$ である。

$$\text{Conf}(\mathbf{A} \Rightarrow \mathbf{B}) = \frac{\text{Supp}(\mathbf{A} \Rightarrow \mathbf{B})}{\text{Supp}(\mathbf{A})} = P(\mathbf{B} | \mathbf{A}) \quad (3)$$

リフト (lift) : ルール「 $\mathbf{A} \Rightarrow \mathbf{B}$ 」のリフトは式(4)で定義される。分子は式(3)の確信度であり、分母は単純に $\mathbf{B}$ を購入する確率である。よって $\text{Lift}(\mathbf{A} \Rightarrow \mathbf{B})$ は、 $\mathbf{A}$ を購入した前提での $\mathbf{B}$ を購入する割合と単純に $\mathbf{B}$ を購入する割合の比であるため、1を超えるリフト値を持つルールを意味のあるルールと判断する。

$$\text{Lift}(\mathbf{A} \Rightarrow \mathbf{B}) = \frac{\text{Conf}(\mathbf{A} \Rightarrow \mathbf{B})}{\text{Supp}(\mathbf{B})} = \frac{P(\mathbf{B} | \mathbf{A})}{P(\mathbf{B})} \quad (4)$$

### 食材データの単純集計

本格的な分析に入る前に、レシピでの利用食材の単純集計結果を見る。食材の使用頻度の棒グラフを図5(クックパッド)、図6(中国)に示す。棒グラフより、クックパッドの食材数が35個に対して、中国のそれは33であった。中国の食材の中に、soybean\_flour(きな粉)とblack\_beans(黒豆)がない。また使用率に関して、クックパッドの上位4つはsugar, milk, egg, butterで、これは中国と同じであり、以下もほぼ同じ傾向を示している。しかし、いくつかの違いがあることがわかる。第1に、クックパッドのsugar, milk, butter, cream, oilの使用率は、それぞれ中国と比べて0.1程度少ない。第2に、クックパッドにおけるchocolateの使用率は0.2弱であるが、中国では0.05程度である。第3に、クックパッドでは卵黄と卵白の使用率がそれほど変わらない一方、中国ではegg\_yolk(卵黄)がegg\_white(卵白)よりおよそ0.05大きい。第4に、クックパッドでのred\_beans(小豆)の使用率が0.11に対して、中国では0.2と少し高い。第5に、クックパッドでのpancake\_mix(パンケーキミックス)の使用率は0.1で、中国ではほぼ0である。第6に、クックパッドでのsoymilk(豆乳), honey, cocoaの使用率はそれぞれ中国と比べて少し高い。第7に、クックパッドでのcake\_flour(薄力粉)

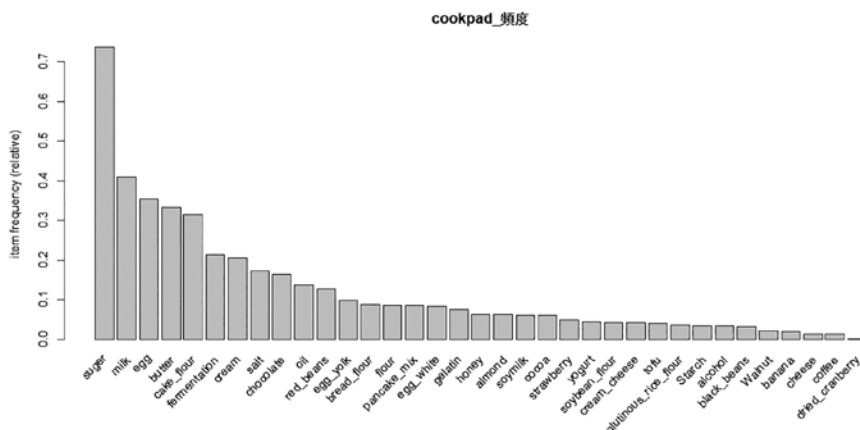


図5 クックパッドの食材使用率の棒グラフ

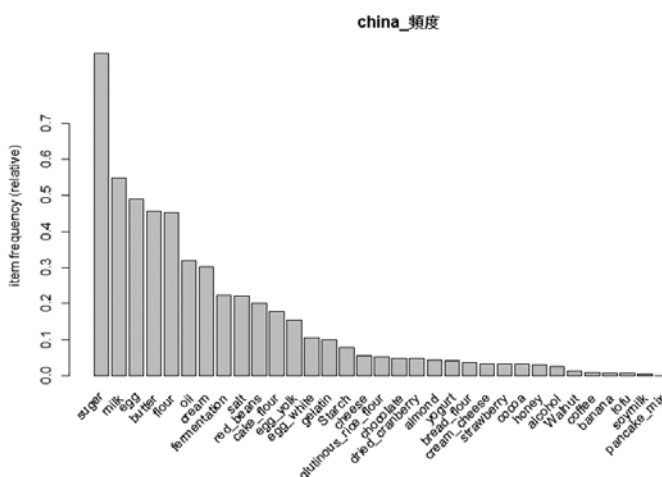


図6 中国の食材使用率の棒グラフ

の使用率は0.3で、中国では0.2である。第8に、クックパッドでのflour（小麦粉）の使用率は0.1未満で、中国では0.45である。

以上の食材の使用率の比較結果は、統計的に有意な差があるかどうかを検討したわけではないが、全体の傾向として次の4つを推察することができる。第1に、日本人が抹茶料理やスイーツを作る時に、中国人と比べると、糖分や油を控え、代わりに豆乳や蜂蜜を使用する頻度が高い。つまり、日本人の健康意識の方が高いと推察できる。第2に、日本人は中国人より多くチョコレートを利用している。第3に、日本の抹茶料理では薄力粉とパンケーキミックスの使用頻度が高く、中国では小麦粉の使用頻度が高い。日本人は薄力粉、強力粉、パンケーキミックスを使い分けているが、中国では小麦粉の種類にはそれほどこだわっていない。第4に、クックパッドと中国のいずれにおいても小豆は抹茶料理と相性が良いと認められているため人気

がある。

### アソシエーションルール分析

アソシエーションルール分析では一般に抽出されるルールが多くなり、重要なルールを発見しにくくなるため、通常、支持度と確信度の最小値を設定しておく。ここでは標準的な値（最小支持度を0.01, 最小確信度を0.5）に設定する。さらにリフトを1.5以上とすると、クックパッドでは887のルールが、中国では1382のルールが抽出された。

より直観的にルールの良し悪しを把握するために、支持度と確信度の散布図を作成する。クックパッドと中国のそれらを図7に示す。いずれも支持度が0.2以内、確信度が0.5～1の範囲にルールが多いことがわかる。さらに確信度の高いルールに注目するため、その最小値を0.8に設定して分析した結果、クックパッドでは137、中国では700に絞られた。さらに各ルールがどのような結論を導いているかを直観的に把握するため、散布図ではなく有向グラフ（矢印）でルールを表現する（図8）。図中の矢印は1つのルールを示し、矢印の先は結論部の項目である。なお、支持度が高いほど線が太く、リフトが高いほど線の濃度は濃くなっている。

クックパッドでは（図8左）、結論部（rhs: right-hand side）の食材は、egg\_yolk, egg\_white, salt, fermentation（発酵用品）、bread\_flour（強力粉）など、ほぼ料理に基本的な食材であるが、中国では（図8右）、結論部の食材はegg\_white, egg\_yolk, Starch（片栗粉）といった基本的な食材以外に、gelatin（ゼラチン）、fermentation, glutinous\_rice\_flour（白玉粉）がある。結論部が一般的な食材の場合、たとえ解析できてもほぼ当たり前の情報しか得られないので、結論部がgelatin, fermentation, glutinous\_rice\_flourの中国のルールに注目する。

中国のルールで、結論部がgelatin, fermentation, glutinous\_rice\_flourのものから、支持度、

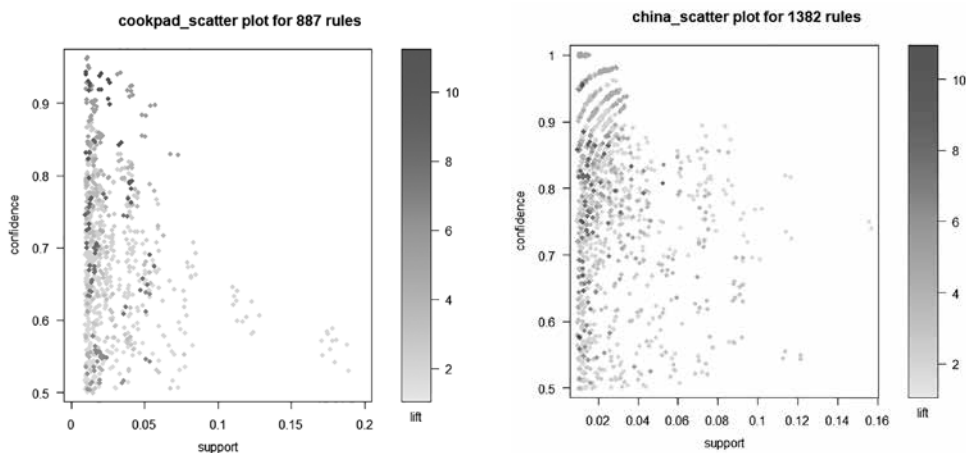


図7 横軸をsupport, 縦軸をconfidenceとした散布図（左：クックパッド, 右：中国）

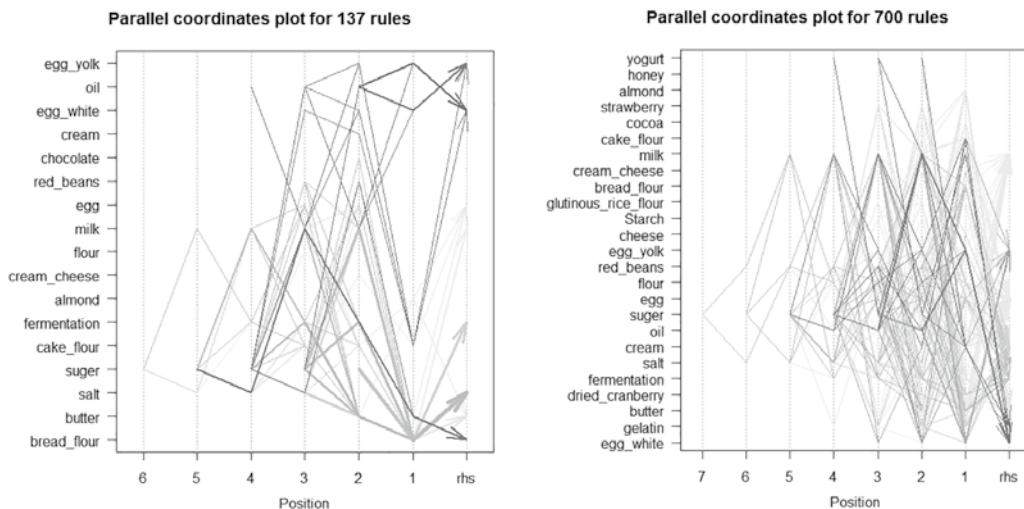


図8 有向グラフの形のルール (左: クックパッド, 右: 中国)

	A	B	C	D
1	rules	support	confidence	lift
2	{red_beans, Starch} => {glutinous_rice_fLOUR}	0.010795	0.5757576	10.896057
3	{cream, milk, yogurt} => {gelatin}	0.011932	0.9545455	9.6
4	{cream, yogurt} => {gelatin}	0.015341	0.8181818	8.2285714
5	{bread_fLOUR, red_beans} => {fermentation}	0.014773		14.4670051

図9 結論部がゼラチン, 発酵用品, 白玉粉のアソシエーションルール—中国

確信度, リフトの高い代表的な4つのルールを図9に示す。これより, 小豆と片栗粉を一緒に使用するレシピでは白玉粉も使われる確率は約0.57である。白玉粉のみを使用する割合と比べて, 小豆と片栗粉を一緒に使用する方が10倍高くなる。クリーム, 牛乳, ヨーグルトと一緒に使用されると, ゼラチンが使われる確率は約0.95である。クリーム, 牛乳, ヨーグルトと一緒に使用する時には, ゼラチンを使用する割合は約9.6倍になる。強力粉と小豆を一緒に使用する時, 発酵用品が使われる確率は1であり, 発酵用品のみを使用する割合と比べて, 強力粉と小豆を一緒に使用する時には発酵用品を使用する割合は約4.5倍になる。

上記の考察より2つのことがわかる。第1に, ヨーグルトが使用される場合, ゼラチンを使用する確率も高くなることから, レシピ作成者はヨーグルトババロアやゼリーのようなデザートを作ろうとしていると推察できる。従って, 抹茶とヨーグルトを検索する人はヨーグルトババロアやゼリーがほしい可能性が高いと考えられる。そうすると, たとえば, スーパーマーケットのヨーグルトの販売エリアに抹茶ヨーグルトババロアやゼリーなどのレシピや写真を提示すれば, 商品の購買意欲を刺激し, 売上が増える可能性がある。第2に, 小豆と片栗粉が使用される場合, 白玉粉の使用率が高くなるため, レシピ作成者は白玉ぜんざいや団子を作りたいと思っていると推察できる。一方, 小豆と強力粉を使用する場合には, 発酵用品の使用率が高

くなることから、ケーキやパンを作ろうと考えていると推察できる。従って、既に抹茶と小豆で検索しており、次に片栗粉を検索する人は、白玉ぜんざいや団子がほしい可能性が高く、強力粉を検索する人はケーキやパンがほしい可能性が高いと判断できる。

## (2) クラスタ分析

クックパッドと中国の抹茶レシピで一緒に使用される食材の異同を調べるために、食材を日本と中国別にグルーピングし、クックパッドのグループと中国のグループとを比較する。グルーピングにはクラスタ分析を用いる。以下、クラスタ分析について簡単に説明する<sup>10)</sup>。

クラスタ分析の目的は、収集したデータを類似度という指標を用いて、似ているもの同士のグループ（クラスタ）に分けることである。類似度の尺度としては、ユークリッド距離、ユークリッド距離の2乗、重み付きユークリッド距離、マンハッタン距離などが用いられる。これらは距離なので、値が小さいほどデータ同士は似ていると解釈する。ここでは、良く利用されるユークリッド距離を用いた。

クラスタ分析には階層型と非階層型の2つがあるが、ここでは階層型クラスタ分析をまず実行し、後に非階層型も実行する。階層型クラスタ分析は、凝集型（ボトムアップ型）と分割型（トップダウン型）の2つに分類されるが、ここでは凝集型を扱う。なお、データとクラスタとの距離やクラスタ同士の距離を計算する際に、一般的に使用されるウォード法を採用した。階層型クラスタ分析では、デンドログラムを作成した後、いくつのクラスタに分けるべきかを判断する。構成されたグループの良し悪しの判断指標としてエルボー曲線とCalinski-Harabaszインデックスを利用したが、一意には決定できなかったため、さらに、クラスタ安定度を評価した<sup>11)</sup>。

クラスタ安定度評価の結果より、クラスタ数を4とした。同様の分析の結果、中国のデータに対してもクラスタ数を4としたデンドログラムを図10に示す。図10では、クックパッドと中国の4つのクラスタ（枠で囲った部分）に左から、1, 2, 3, 4と番号をつけている。これらと比較すると4つのことがわかる。第1に、クックパッドおよび中国の一番大きなクラスタ1を構成する食材はほぼ同じである。異なるものは、クックパッドの小麦粉、中国の強力粉とチョコレートである。第2に、クックパッドのクラスタ2と中国のクラスタ3の内容もかなり似ており、同じものは、小豆、発酵品、卵黄、卵白、塩、油であり、異なるものは、

10) 山本ら (2015)、照井・佐藤 (2013) 参照。

11) Rのパッケージfpc (<https://cran.r-project.org/web/packages/fpc/index.html>) を利用。この統計量は、ブートストラップサンプリングによりJaccard係数を複数回計算し、その平均を取った指標であり、

・0.6以下のときクラスタは信頼できない

・0.6以上0.75未満の場合、ある程度のパターンは構成されている

・0.75以上が望ましく、0.85より大きな場合、非常に安定していると判断する (Hennig, 2020)。

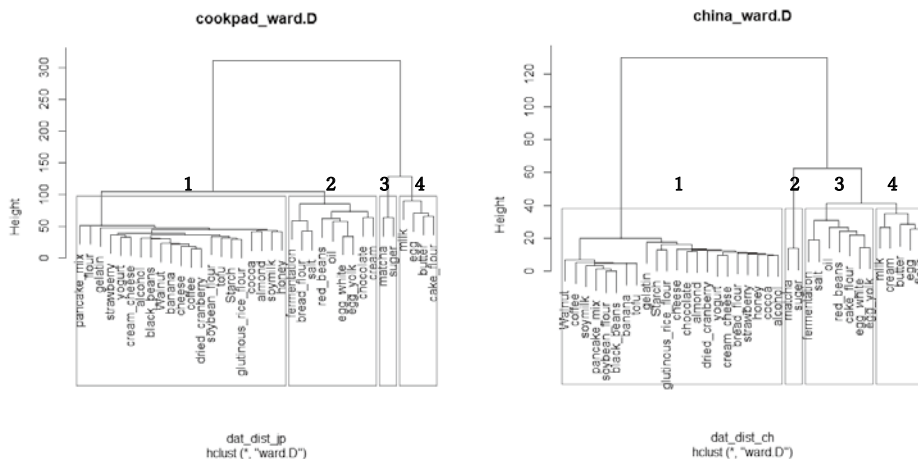


図 10 クラスタ数4のクラスタ（左:クックパッド, 右:中国）

クックパッドの強力粉、クリーム、チョコレートが、中国の薄力粉である。第3に、クックパッドのクラスタ3と中国のクラスタ2は同じで、抹茶と砂糖の2つのみで構成されている。第4に、クックパッドと中国のクラスタ4の内容はほぼ同じである。クックパッドにはミルク、卵、バター、薄力粉が含まれ、中国にはミルク、卵、バター、小麦粉、クリームが含まれている。

以上、クックパッドと中国の階層型クラスタリングを行い、各4つのクラスタに分けた。ここでは、各4クラスタの内容からクラスタを解釈する。抹茶および砂糖はレシピの重要ポイントなので、抹茶と砂糖のクラスタを「主題」クラスタと解釈する。牛乳、卵、バター、小麦粉は、料理やデザートにおいて一般的な料理にも使用される基本的な食材であるため、「基本食材」クラスタと解釈する。発酵食品、強力粉、小豆、油などは、抹茶小豆パンやケーキに使用される食材であり、抹茶と小豆の相性が良く、そのケーキやパンも一番人気あるいは定番レシピと考えているため、「定番料理の食材」クラスタと解釈する。最後のクラスタは、ココア、豆乳、チーズなどの料理の味や見た目に変化を与える食材と考え、「彩り食材」クラスタと解釈する。このようにクラスタリングの結果、食材が「主題」、「基本食材」、「定番料理の食材」、「彩り食材」と4つのテーマ（クラスタ）に分類できていることがわかった。

次に、非階層型クラスタ分析を実行し、階層型クラスタリングの結果と比較しておく。この代表的な手法としてk-means（k平均）法がある（照井・佐藤ら、2013）。階層型クラスタリングと比較するため、クラスタ数を4に設定し、クックパッドと中国の食材データをk-means法で出力した結果は次のようになった。

クックパッド

クラスタ1 : `milk`

クラスタ2 : `yogurt, egg_white, egg_yolk, cream_cheese, chocolate, cheese, banana, red`

beans, walnut, starch, soymilk, tofu, soybean\_flour, bread\_flour, honey, cocoa, almond, alcohol, oil, pancake\_mix, gelatin, strawberry, coffee, black\_beans, glutinous\_rice\_flour, dried\_cranberry, cream, flour, salt

クラスター 3 : egg, butter, fermentation, cake\_flour

クラスター 4 : matcha, suger

中国

クラスター 1 : milk, cream, oil

クラスター 2 : yogurt, egg\_white, egg\_yolk, cream\_cheese, chocolate, cheese, banana, red\_beans, walnut, starch, soymilk, tofu, soybean\_flour, bread\_flour, honey, cocoa, almond, alcohol, pancake\_mix, gelatin, strawberry, coffee, black\_beans, glutinous\_rice\_flour, dried\_cranberry, cake\_flour,

クラスター 3 : egg, butter, fermentation, flour, salt

クラスター 4 : matcha, suger

k-means法によるクックパッドと中国とのクラスタリングの結果はほとんど同じである<sup>12)</sup>。階層型クラスタリングで得られた4つのクラスターと比較すると、k-means法により得た4つのクラスターは少し異なる。しかし、k-means法で分けられた4つのクラスターも階層型によって解釈した「主題」、「基本食材」、「定番料理の食材」、「彩り食材」という4つのテーマに当てはまると判断できる。従って、クラスター数を4に設定する場合、階層型クラスター分析の結果と非階層クラスター分析の結果は大きな違いはない。

### (3) テキストマイニング

日本人と中国人の抹茶に対する好みの異同を分析するために、レシピの説明文をデータとしてテキストマイニングを行った。テキストマイニングでは、Rに加えて、KH Coder<sup>13)</sup>も利用する。日本語テキストの分析では、文を形態素と呼ばれる意味のある語単位に分解した後、その形態素の関係を分析する。KH Coderでは、形態素への分解とともに、文や文章の中にどのような形態素が出現しているか、その出現回数を集計できるのみならず、それを用いて対応分析やクラスター分析といった多変量解析を実行することもできる。また、テキストの中に出現する形態素同士または異なるテキスト間の類似度を抽出することも可能である。コーディングルールを利用した多変量解析も可能である(樋口, 2014)。

12) k-means法によるクラスタリングの結果のクラスター番号は任意のため、本来一致しないが、ここではわかりやすくするため、番号をつけ直すとともに一致する食材に網掛けしている。

13) <http://kncoder.net/index.html>



KH Coderを用いて中国語のテキストを分析することは可能であるが、中国と日本におけるレシピの異同を比較するために、分析のどこかの段階で中国語を日本語に翻訳する必要がある。そのため本研究では、グーグル翻訳<sup>14)</sup>を援用してあらかじめ中国語を日本語に翻訳した。全ての文章が文法的に正しく日本語に翻訳されることは期待できないが、テキストマイニングでは、分解された形態素が、文や段落等と一緒に含まれるかどうかという内容を利用するため、単語が正しく翻訳されている限り分析に支障は生じない。

形態素解析の後、さらなる前処理が必要となる。その1つは、食材の場合にもそうであったように、語の揺らぎを無くす必要がある。もう1つは、分析に不要と判断した絵文字や記号、意味不明な語などを削除するとともに、分解されたくない複合語を強制的に抽出することである。例えば、ホワイトチョコレートという語がテキスト中にあるとき、形態素解析により、これが「ホワイト」と「チョコレート」に分割されてしまう可能性がある。「ホワイトチョコレート」を分析対象としたい場合、KH Coderがもつ強制抽出機能を用いるとそれが可能となる。

説明文の中でよく使われている語が何かを知るために、使用単語の頻度グラフを作成した(図11)。図より、クックパッドでは主に便利さ、美味しさ、健康志向に関する語の使用頻度が高いこと、中国では健康志向に関する語があまり見られないかわり、カラーや緑といったビジュアルイメージに関する語の使用頻度が比較的高いことがわかる。

次に、1つの文の集まり（文章、段落）の中に一緒に出現する語（共起語）の関係を把握す

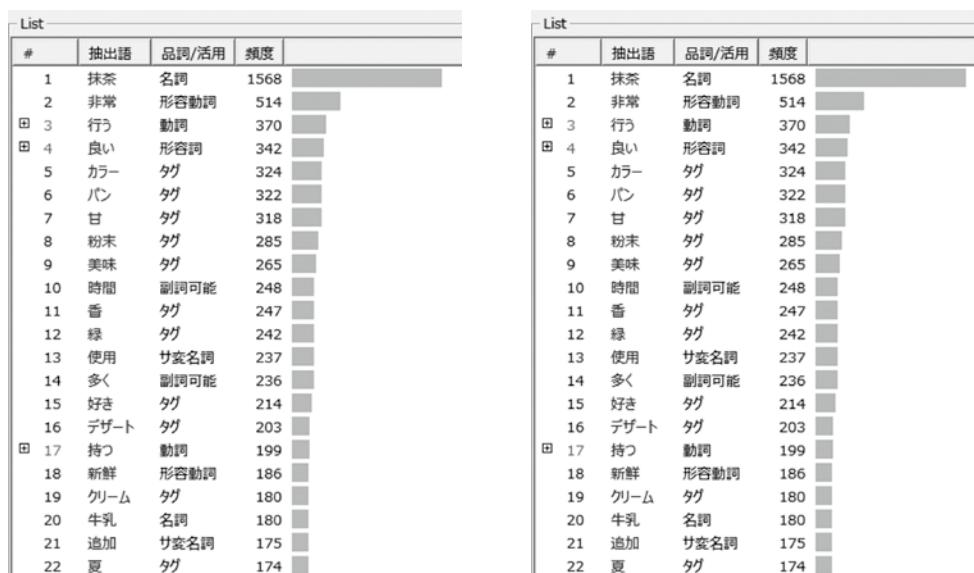


図11 抽出語頻度の棒グラフ（左：クックパッド、右：中国）

14) [https://translate.google.com/?hl=ja&ie=Shift\\_JIS&oe=Shift\\_JIS&langpair=es%7Cen](https://translate.google.com/?hl=ja&ie=Shift_JIS&oe=Shift_JIS&langpair=es%7Cen)

るため、共起ネットワークを作成した(図12, 図13)。さらに、クラスター分析を行った(図14)。図12と図14(左:クックパッド)の出力結果から、次に示す4つがわかる。

- ・抹茶料理を楽しむ際に便利さと美味しさが求められる
- ・甘さ控えめ、大人が楽しめる苦みをもつ抹茶料理へのニーズがある
- ・抹茶には和風が共起しやすく、和菓子やスイーツ等に使われる。小豆との相性もよく、パンを作る時には抹茶と小豆を一緒に生地に練り込む
- ・抹茶をケーキに使う時、味覚だけではなく、しっとり、ふわふわ、香りといった食感と香りも重視される

一方、図13と図14(右:中国)からは、次の3つがわかる。

- ・美味しい抹茶料理を楽しむとともに、見た目と香りを重視
- ・家でオープンを使って簡単にデザートを作り、友達や子供と一緒に楽しむ
- ・抹茶の緑色が視覚的な刺激となり、春のような新鮮なイメージを与える

以上より、日本と中国の両者とも基本的に美味しさと便利さを求めているが、日本では比較的健康意識が高く、和風料理が相対的に多いという特徴が、中国では視覚的な刺激を求めるといった特徴がある。

次に、特定の語が潜在する何らかの欲求や意思の表れであると考え、これを分析で利用する。例えば、肥満予防、シナモン、納豆、カテキンなどの語は潜在する健康志向への欲求の表れであり、テキパキ、簡単、便利、手軽などの語は潜在する簡便さへの欲求の表れであると考えerことは自然である。そうした潜在欲求を考慮し、分析者が主観的に語にラベルを付け(コーディングという)、それを分析の際に利用する(KH Coderで実行可能)。コーディングを単純集計した結果を図15に示す。これより、次に記す8つがわかる

- ・クックパッドおよび中国とも、潜在的には健康意識が高い。また、両者とも美味しさを求めると同時に、便利さも重視
- ・クックパッドでの和風は19%、洋風は13%。中国での和風は2%、洋風は9.5%。中国では、和風料理はまだ浸透していない
- ・クックパッドでは見た目と香りに関する集計がそれぞれ14.5%と6.3%。中国では28.3%と16.3%。中国の方が見た目と香りを重視
- ・チョコとホワイトチョコのコーディング率はクックパッドがそれぞれ4.6%と3.1%で、中国では2.2%と0.8%。中国では、チョコレートの、特にホワイトチョコレートの利用率が低い
- ・クックパッドでは春と夏の頻度率が各1.8%と1.5%に対して、秋と冬が各0.41%と0.2%。中国では春と夏の頻度率が各5.1%と9.6%に対して、秋と冬では各1.2%と1.5%。よって両者において季節により抹茶料理への人気度に違いがある。これは特に中国の方が顕著。夏に人気があるため、アイスクリームの頻度も高く、5.18%。クックパッドの2.62%と比較して、およそ2倍

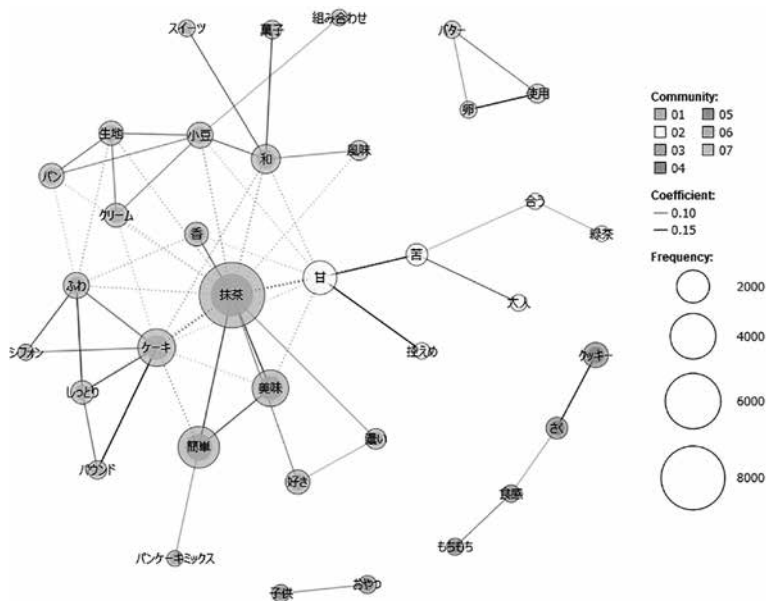


図12 共起ネットワーク—クックパッド

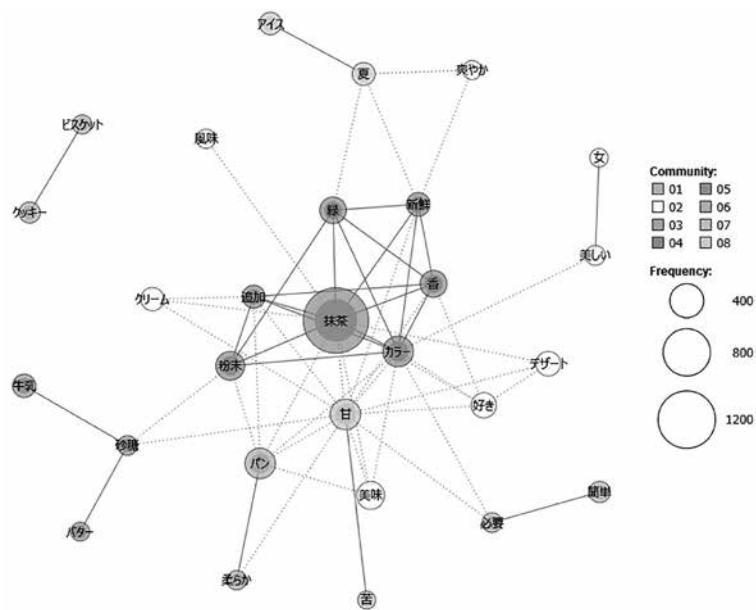


図13 共起ネットワーク—中国

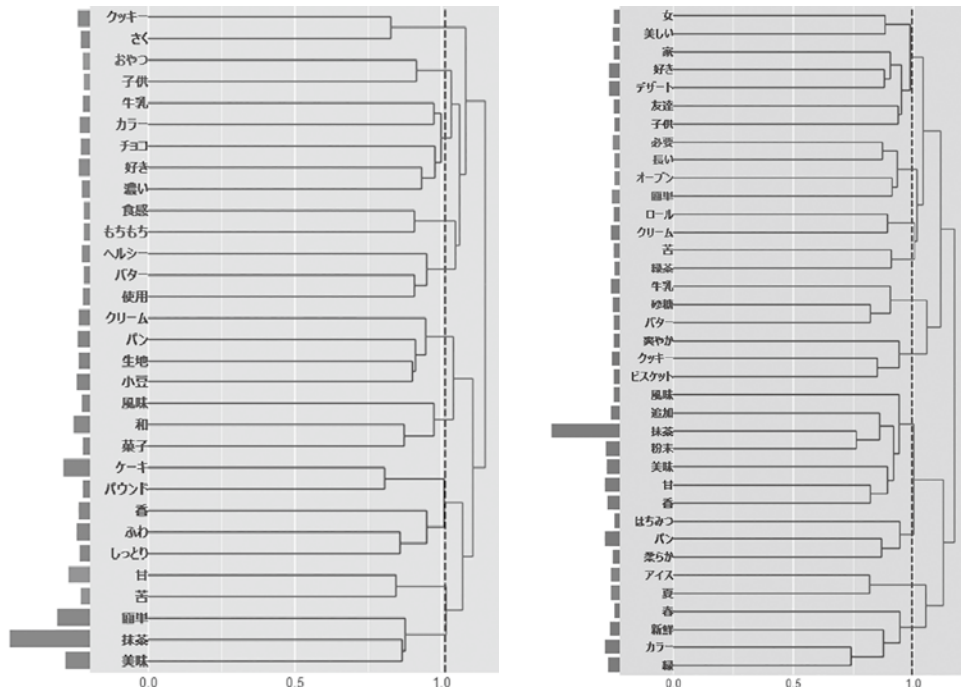


図14 抽出語のクラスター分析 (左：クックパッド, 右：中国)

コード名	頻度	パーセント	コード名	頻度	パーセント
*健康志向	3584	23.42%	*ホワイトチョコ	474	3.10%
*食感の豊かさ	582	3.80%	*ふわ	1172	7.66%
*和風	2990	19.54%	*さく	715	4.67%
*洋風	2084	13.62%	*カリ	130	0.85%
*美容、ダイエット	506	3.31%	*美味しい	2495	16.31%
*オシャレ、上品	383	2.50%	*香り	969	6.33%
*便利さ	4175	27.29%	*チーズ	301	1.97%
*大人	1814	11.86%	*酸味	132	0.86%
*子供、可愛さ	1432	9.36%	*組み合わせ	1092	7.14%
*祝日、お祝い、一緒に	949	6.20%	*しっとり	949	6.20%
*プレゼント	701	4.58%	*緑茶	363	2.37%
*オリジナリティ	52	0.34%	*春	276	1.80%
*飾り付け、外見	2218	14.50%	*夏	230	1.50%
*飲み物	541	3.54%	*冬	30	0.20%
*アイスクリーム	401	2.62%	*秋	63	0.41%
*コーヒー	143	0.93%	*宇治	98	0.64%
*好き	1010	6.60%	#コード無し	1514	9.89%
*チョコ	704	4.60%	(文書数)	15301	

コード名	頻度	パーセント	コード名	頻度	パーセント
*健康志向	316	24.07%	*ホワイトチョコ	11	0.84%
*食感の豊かさ	71	5.41%	*ふわ	4	0.30%
*和風	27	2.06%	*さく	16	1.22%
*洋風	125	9.52%	*カリ	35	2.67%
*美容、ダイエット	23	1.75%	*美味しい	223	16.98%
*オシャレ、上品	17	1.29%	*香り	215	16.37%
*便利さ	274	20.87%	*チーズ	42	3.20%
*大人	82	6.25%	*酸味	42	3.20%
*子供、可愛さ	104	7.92%	*組み合わせ	7	0.53%
*祝日、お祝い、一緒に	120	9.14%	*しっとり	10	0.76%
*プレゼント	34	2.59%	*緑茶	80	6.09%
*オリジナリティ	20	1.52%	*春	67	5.10%
*飾り付け、外見	372	28.33%	*夏	126	9.60%
*飲み物	64	4.87%	*冬	19	1.45%
*アイスクリーム	68	5.18%	*秋	16	1.22%
*コーヒー	32	2.44%	*宇治	13	0.99%
*好き	175	13.33%	#コード無し	232	17.67%
*チョコ	29	2.21%	(文書数)	1313	

図15 コーディング単純集計 (左：クックパッド, 右：中国)

- ・クックパッドでは大人が楽しめる渋みや苦みがある抹茶レシピの頻度率が11.86%。これは中国の6.25%と比べて高い
  - ・クックパッドでは、友達や家族と一緒に楽しむ場面が6.2%、プレゼントとして他人に渡す場面が4.58%。中国ではそれぞれ9.14%、2.59%
  - ・クックパッドでは抹茶料理の中に酸味がありません(0.86%)、中国では3.2%
- 続いて、コーディングに基づくクラスタリングの結果を図16に示す。これより、次に記す4

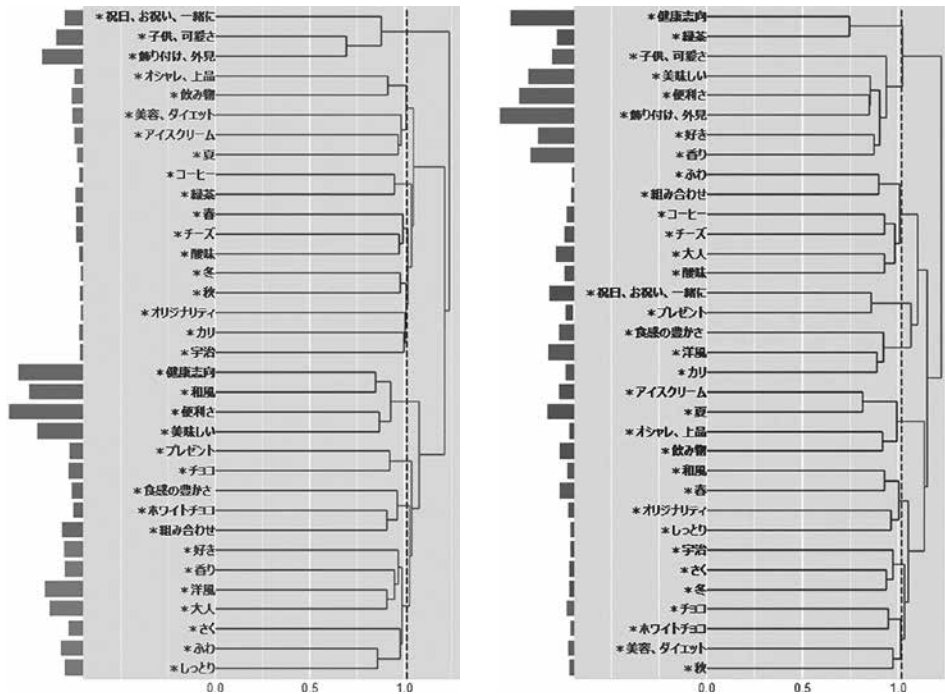


図16 コーディングのクラスター分析結果（右：クックパッド，左：中国）

つがわかる。

- ・クックパッドでは健康志向と和風，便利さ，おいしさ一つのグループを構成し，中国では健康志向と緑茶，便利さ，外見，香りが一つのグループを構成
- ・クックパッドでは大人のグループに洋風，香りあり，中国では酸味，チーズ，コーヒーがグループを構成
- ・クックパッドではプレゼントとして抹茶料理を作る時にチョコレートを利用することが多いが，中国ではそれがなく，祝日・お祝い時に利用
- ・クックパッドと中国のいずれにおいても，香りと好きが一つのクラスターに属しており，味とともに香りのインパクトを求める

コーディングではコーディング名と対象とする語を主観的に決定するため，その任意性が問題点として残る。そのためより客観的に分析するため，トピックモデリングを行う。トピックモデリングとは，テキストデータを教師なしで分類する方法であり，潜在的ディリクレ配分法(LDA:Latent Dirichlet Allocation)が広く使われる。LDAは，文章をトピックの組み合わせ，個々のトピックを単語の組み合わせとして数理的にモデル化する手法である (Silge and Robinson, 2018)。これの実行には，Rにあるトピックモデリングのパッケージtopicmodels<sup>15)</sup>

15) <https://cran.r-project.org/web/packages/topicmodels/index.html>. Grün and Hornik (2019).

を利用した。トピックモデルを作成するには、あらかじめトピック数を指定しておく必要がある。トピック数を少なく設定すると意味が曖昧になったり、トピック数を多く設定すると細分化しすぎたりすることになり、解釈が難しくなる。最適なトピック数を決定するために通常、トピック数を複数設定しておき、それらの中から最適なものを事後的に選択する（本研究では2から15に設定）。最適なトピック数を決定するために利用できる複数の指標が存在するが、ここでは、4つの指標を利用した（Nikita and Chaney, 2020）。しかしこれらを用いたとき、得られたトピックの解釈が難しかった。トピックの解釈可能性は重要であるため、さらに他の数にて試行し、最終的にクックパッドと中国のトピック数を3とした。

得られた3つのトピックの上位の頻出語を図17に示す（棒グラフの上の数字がトピック番号）。クックパッドのトピック1の頻出語を見ると（図17左）、上位にケーキ、パン、クッキー、およびそれらの材料、プレゼントといった語があるため、これは抹茶料理の種類、特にプレゼント用だと解釈する。トピック2では、上位にケーキ、和風、子供、ヘルシーといった語があるため、子供向けの和風スイーツだと解釈する。トピック3では、上位に和風、小豆、香り、カラー、可愛い、控えめといった語があるため、和風で味以外の重視、および健康志向だと解釈する。一方、中国を見ると、トピック1の頻出語としてケーキがずば抜けて多いため、トピック1はケーキだと解釈する。トピック2では、クッキー、デザート、ケーキがあるため、抹茶料理の種類だと解釈する。トピック3では、上位にカラー、小豆、緑があるため、見た目重視で爽やかなイメージだと解釈する。これらより、日本と中国との同じ点はいずれもケーキと抹茶料理の種類といったトピックがあり、異なる点はクックパッドでは健康重視や和風といったトピックがある一方、中国では見た目を重視していることがわかる。

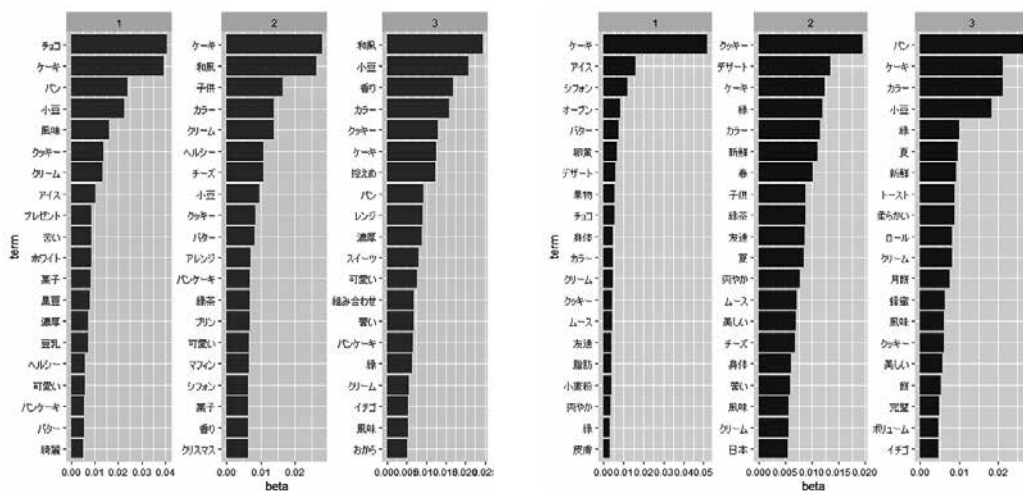


図17 各トピックの上位頻出語（左：クックパッド，右：中国）

## 4. 分析結果からの考察

第3節において、使用された食材のデータ分析（アソシエーションルール分析、階層型・非階層型クラスター分析）とレシピの説明文の分析（テキストマイニング）を行った。ここではこれらから得られた結果を総合的に考察するとともに具体的な提案を行う。

### (1) 食材データの分析

アソシエーションルール分析から、次に示す5つの情報を得た。(A1) クックパッドの利用者は砂糖や油などの食材を控え、代わりに豆乳や蜂蜜などの食材を積極的に使用し、健康意識が比較的高い。(A2) 中国と比べてクックパッドの方がチョコレートの使用率が高い。(A3) 小豆はクックパッドと中国のいずれにおいても人気が高いが、特に中国の方が顕著である。(A4) クックパッドでは薄力粉、強力粉、パンケーキミックスを細かく使い分けているが、中国ではまだその習慣はまだ確立されていない。(A5) 中国では食材にヨーグルトが含まれると、パバロアやゼリーを作ろうとしている可能性が高く、小豆が原料に入っているとき、片栗粉があるとぜんざいを作ろうとしている可能性が高い。

クラスター分析からは、クックパッドと中国の両者とも4つのクラスターが構成され、各クラスターの内容はほぼ同じである。これらのクラスターの潜在的な意味を解釈すると、主題、基本食材、定番料理の食材、彩り食材と解釈できる。

### (2) レシピの説明文の分析

共起ネットワークおよびクラスター分析から、次の6つの情報を得た。クックパッドでは：(C1) 抹茶料理を楽しむときに美味しさはもちろん、便利さと健康も求めている。(C2) 抹茶ケーキを作るときに他の料理と比べて、食感と香りにこだわる。(C3) 和風のスイーツや菓子が人気で、その中で小豆がよく使用される。(C4) 大人が楽しめる渋みや苦みをもつ抹茶料理へのニーズがある。一方中国では：(C5) 料理の美味しさと便利さを求め、家族や友達と一緒に食べるニーズが多い。(C6) 特に見た目と香りを重視し、抹茶の緑が与える春のような爽やかなイメージを好む。

コーディングルールを用いたテキスト分析から、次の7つの情報を得た。(D1) クックパッドのみならず中国でも健康志向がある。(D2) クックパッドでは洋風より和風料理が多い一方、中国では和風料理がほとんどなく、洋風が圧倒的に多い。(D3) 中国では明らかに抹茶料理の見た目と香りを重視する。(D4) クックパッドと中国のいずれにおいても秋・冬よりも春・夏の人気が高い。特に中国の方が季節による人気の差が明確である。(D5) クックパッドと比べて中国では友達や家族と一緒に楽しむことができる抹茶料理が多く、プレゼントとする場面は

少ない。(D6) クックパッドの抹茶料理にはあまり酸味を加えない一方、中国では大人が楽しめる抹茶料理として酸味を加える傾向がある。(D7) クックパッドおよび中国のテキストは3つのトピックで解釈でき、それぞれ「抹茶料理の種類、特にプレゼント用」、「子ども向け和風スイーツ」、「和風で視覚と嗅覚の重視、および健康」の3つと、「ケーキ」、「抹茶料理の種類」、「見た目」の3つである。

### (3) 総合的な考察

上記の内容は、当然のことながらかなり重複している。そのため、それらの結果を再度吟味し、ビジネスにおいて抹茶情報やレシピを中国に発信する場合にキーとなる6つの具体的な内容を提示する。(T1) クックパッドでの健康意識は高く、中国では、明示的にはあまり健康志向が見られない。しかし、コーディングルール分析の結果からは健康意識が表れており、矛盾しているように思われる。その原因として、中国では既に確実に抹茶が健康的な食材と意識されており、抹茶自体に対して健康効果を強く期待しているが、それが潜在化しているからであると判断できる。これに対してクックパッドでは、抹茶だけではなく、蜂蜜や黒豆などのような健康にいい食材もなるべく選ぶことにより、和風から洋風への食の嗜好の変化の中で、もう一段健康志向を強くする傾向が表れていると考えることができる。よって、クックパッドで提案されているより健康的なレシピを中国に発信すれば、今後さらに受け入れられやすくなる可能性がある。(T2) 中国では和風料理があまり意識されていない。これは、「抹茶=和風」という意識が高いため、和風が強く表れていない可能性がある。小豆は既にクックパッドと中国いずれにおいても抹茶と相性がいいと認められている。そのため、小豆を中心とする様々な和風抹茶料理を中国に発信すれば、あるいは逆に中国特有の料理、例えば、月餅や湯園などを抹茶と組み合わせて日本に発信すれば、ビジネスチャンスが生まれると考える。(T3) 中国では見た目が重視されるため、新たな抹茶料理を中国に発信する際、小豆の赤と抹茶の緑を組み合わせるといった色合わせを工夫すると良い。(T4) クックパッドではプレゼントとして渡す時に抹茶レシピにチョコレートを利用する習慣があるが、中国ではチョコレートの使用率が低い。中国におけるチョコレート使用率の低さの理由としてはいくつかの可能性がある。例えば、チョコレートと抹茶の色合わせを避けている、抹茶のレシピにチョコレートをうまく活用できない、そもそもチョコレート自体が嫌い、などである。色合わせに関しては、ホワイトチョコレートとの組み合わせのレシピを中国に発信していくといった方策を考えることができる。(T5) 中国では、抹茶料理の人気に季節変動があり、秋・冬と比べて春・夏の利用度が高い。中国に新しいレシピや抹茶商品を発信する時に、タイミングを考慮する必要がある。(T6) 大人が楽しめる抹茶レシピを中国に発信するとき、単に渋みや苦みのバランスのみを考慮するのではなく、酸味も加えるとより良い。中国では見た目を重視するという点も併せて考えると、酸味がある食材（レモン、すだち、梅など）との色合わせを工夫すればより受け入れられやすくなる



可能性がある。

## 5. おわりに

抹茶は中国で発祥したが、明代に消滅した。一方、日本がそれを継承し、現在まで人々に愛飲されるとともに食されてきた。昔から抹茶の保健効果について文献での記述があったが、近代になると抹茶は抗がん、脳卒中予防、高血圧予防などの保健効果をもつことが科学的に証明された。そのような健康に良い抹茶を単に飲用するだけではなく、食材として麺類やパン、スイーツなどに利用し、健康と美味しさを同時に手に入れるという新たな摂取方法が注目されている。最近では、日本・中国だけではなく、抹茶の魅力が世界で受容されている。中国でも抹茶が復活するとともにその人気が高まっている。しかし、中国市場に出回っている抹茶はほとんど中国製であり、その品質は日本産と比べて劣っている。そのため中国人は、既に高品質であると認知されている日本産の抹茶を求めている。従って、中国への日本の抹茶企業の進出、日本産抹茶の輸出、中国における日本抹茶のブランド化、抹茶料理自体の魅力の普及などもさらに拡大する余地がある。それらを実現するためにまずは、中国人が持つ抹茶料理に対する好みや拘りに関する日本人との異同、さらに、抹茶料理を食べる場面やタイミングといった事柄を知る必要があり、この情報に基づいて潜在するニーズを把握すれば、日本と中国の抹茶関連企業にとって有益な情報となる。

本研究では、このような問題意識の元、グルメサイトで公開されている日本と中国の抹茶レシピを比較することにより、日本と中国の抹茶に対して潜在する意識の異同を浮かび上がらせることを試みた。日本のクックパッド、中国の美食杰、美食天下、下厨房というグルメサイトに自発的にアップされた抹茶レシピデータをスクレイピングし、分析した。分析に用いたデータは食材と説明文である。分析結果から、日本と中国の抹茶レシピにいくつかの異同を発見した。一致するのは、抹茶料理を楽しむ時に、美味しさと便利さを重視する傾向や抹茶料理に小豆の使用率が高いことなどである。異なるのは、クックパッドでチョコレートの使用率が高く、中国では抹茶料理の見た目と香りかなり重視する点である。これらの情報に基づいて、具体的に6つの提案を行った。しかし、残された課題もいくつかあり、ここでは主に4つを記す。第1に、レシピの作者の情報を収集していないことである。レシピの作者の情報には、出身地、年代層、抹茶レシピ以外の好きなレシピなどがある。それらを分析に利用すれば、地域別、年代層別の好みを分析することが可能となるだろう。また、抹茶料理が好きな人たちは、他にどのような料理に興味があるかということもわかる可能性がある。第2に、レシピの公開日情報を利用しなかった。これを利用すると、クックパッドと中国の抹茶レシピ間に関係があると思えば、それを抽出することが可能となる。例えば、抹茶レモンプリンがまずクックパッドでアップされて、一定期間後に中国のグルメサイトでアップされるといった関連性である。さらに

は、日本と中国での抹茶レシピの伝播の関係を分析することができる可能性がある。第3に、クラスター分析とトピックモデリングの解釈に主観性がある。これらではクラスター数とトピック数を設定する際に指標を参照した。しかし、最終的には解釈のしやすさを考慮して決定したため、ある程度客観性が劣るとも考えられる。しかしこれは、探索的なデータ解析のアーティファクトな部分があつた点でもある。第4に、クックパッドでアップされたレシピ数は中国と比べて圧倒的に多い(約10倍)にも関わらず、アソシエーションルール分析で抽出されたルールは、中国の方が多くなった(約1.5倍から5倍)。これには、中国における人口の多さ、地域特性の多様性、急激なグルメサイトの成長といった要因が関係していると思われる。

なお課題ではないが、データにおける制約に関して2点記しておく。1つは、既に記したように、中国グルメサイトの抹茶レシピの量が少ない点である(クックパッドの有効データ数15323に対して、中国のそれは1760)。今後、中国における掲載レシピが増えていくことを期待するとともに、この増加率は中国における抹茶の人気の代理変数として取り扱うことができるようになるであろう。もう1つは、中国のサイトでは人気度を測る指標を公開しているところがあるが、クックパッドではそのようなデータが提供されていない点である。クックパッドには「つくれば」というサービスがあり、この数を人気度の代理変数とすることも考えられるが、現状ほとんどの抹茶レシピでは0である。もし、「いいね」の数やレシピが閲覧・ダウンロードされた数といったデータが取得できれば、さらに有益な分析が実行できると考える。

#### 参考文献

##### 日本語・中国語

- 池ヶ谷賢次郎・高柳博次・阿南豊正 (1984)「抹茶の化学成分」『短報』1984(60), 79-81, 日本茶業学会
- 石田基広・市川太佑・瓜生真也・湯谷啓明 (2017)『Rによるスクレイピング入門』シーアンドアール研究所
- 尹春英・刘乾刚 (2008)「抹茶溯源及其利用」『茶叶科学技术』2008(02), 3-15, 福建省农业科学院茶叶研究所
- 岩間真知子 (2019)「お茶と養生」『日本健康学会誌』85(1), 30-42, 日本健康学会
- 工藤康将 (2008)「碾茶」『茶大百科 I 歴史・文化/品質・機能性/品種/製茶』農山漁村文化協会
- 公益社団法人日本茶業中央会 (2019)「緑茶の表示基準」公益社団法人日本茶業中央会, <https://www.nihoncha.or.jp/pdf/hyoujikkijyun.pdf>
- ジャパンフードサイエンス (2017)「世界から注目される日本のスーパーフード『抹茶』(特集 清涼飲料をめぐる技術動向)」『ジャパンフードサイエンス: 食品加工と包装技術』56(3), 24-29, 日本食品出版
- 中华人民共和国国家标准GB/T (2017)「抹茶」中华人民共和国国家质量监督检验检疫总局・中国国家标准化管理委员会<http://www.tea.agri.cn/jspd/kpdw/201712/P020171228585766024222.pdf>
- 张士康 (2018)「抹茶の产业机会」茶和天下, [http://tea.zjol.com.cn/mjzl/201810/t20181024\\_8559829.shtml](http://tea.zjol.com.cn/mjzl/201810/t20181024_8559829.shtml)
- 照井伸彦・佐藤忠彦 (2013)『現代マーケティングリサーチ—市場を読み解くデータ分析』有斐閣
- 杜旋・郭倩文・倪穗 (2018)「我国抹茶的研究现状及前景展望」『中国野生植物资源』37(4), 29-35, 中华全国供销合作总社南京野生植物综合利用研究设计院
- 中村衣里・吉川友佳子・戸根瑛美・木戸和貴子・松浦寿喜 (2012)「抹茶含有菓子パンの摂取が健常人の食後血糖値に与える影響」『日本食品化学学会誌』19(2), 124-128, 日本食品化学学会
- 中村好志 (2012)「茶の成分と保健効果(特集 抹茶)」『月刊食生活』106(1), 40-47, カザン
- 日本経済新聞(電子版) (2018a)「大井川茶園がクリーンルームで抹茶, 高品質で海外開拓」日本経済新聞,

- <https://www.nikkei.com/article/DGXMZO36042060S8A001C1L61000/>  
日本経済新聞 (電子版) (2018b) 「佐々木製茶, 抹茶生産能力3倍 一貫加工へ新工場」日本経済新聞,  
<https://www.nikkei.com/article/DGXMZO34155630U8A810C1L61000/>  
日本茶葉輸出促進協議会 (2019) 「抹茶 (粉末状の緑茶) の輸出状況について」  
<http://www.nihon-cha.or.jp/export/news/2019/news190320.html>  
布目潮凧 (2001) 『中国茶の文化史 固形茶から茶葉へ』研文出版  
布目潮凧 (2004) 『布目潮凧中国史論集 下巻』汲古書院  
原口健司 (2015) 「抹茶の特徴」『におい・かおり環境学会誌』46(2), 121-126, におい・かおり環境協会  
樋口耕一 (2014) 『社会調査のための計量テキスト分析』ナカニシヤ  
比达咨询 (2017) 「2017年1月菜谱类APP产品市场监测报告」比达网,  
<http://www.bigdata-research.cn/content/201703/395.html>  
久永忠範・渕田孝康 (2018) 「統計処理を用いたオープンデータの述語の推薦手法の提案」『情報知識学会誌』  
28(2), 127-133, 情報知識学会  
福司山エツ子 (2008) 「茶料理」『茶大百科 I 歴史・文化/品質・機能性/品種/製茶』農山漁村文化協会  
松崎敏 (2008) 「飲料以外の利用」『茶大百科 I 歴史・文化/品質・機能性/品種/製茶』農山漁村文化協会  
松井陽吉 (2008) 「中国茶の変遷とウーロン茶」『茶大百科 I 歴史・文化/品質・機能性/品種/製茶』農山漁村  
文化協会  
村山健二 編集 (2018) 「中国を席巻する日本の『抹茶』, 中国産の『緑茶の粉末』と値段が全然違う理由 = 中国」  
Searchina  
守屋毅 (1992) 『喫茶の文明史』淡交社  
医歯薬出版編 (2019) 『日本食品標準成分表2019七訂』医歯薬出版  
山本義郎・藤野友和・久保田貴文 (2015) 『Rによるデータマイニング入門』オーム社  
刘冬娜・聂坤伦・杜晓・常菊・李双伶 (2014) 「抹茶品質的感官审评与成分分析」『食品科学』35(02), 168-172,  
北京食品科学研究院

## 英語

- Agrawal, R. and R. Srikant (1994). Fast algorithms for mining association rules, *Proceedings of the 20th International Conference on Very Large Data Bases*, 487-499  
Grün, S. B. and K. Hornik (2019). topicmodels: An R package for fitting topic models, <https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf>  
Hennig, C. (2020). fpc: Flexible procedures for clustering, <https://CRAN.R-project.org/package=fpc>  
Lantz, B. (2015). *Machine Learning with R Second Edition*, Packt Publishing. 長尾高弘訳 (2017) 『Rによる機械学習』翔泳社  
Nikita, M. and N. Chaney (2020). ldatuning: Tuning of the latent dirichlet allocation models parameters, <https://cran.r-project.org/web/packages/ldatuning/ldatuning.pdf>  
R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>  
Silge, J. and D. Robinson (2017). *Text Mining with R*, O'Reilly Media, Inc. 大橋真也監訳, 長尾高弘訳 (2018) 『Rによるテキストマイニング』オライリー・ジャパン