

レビュー情報を用いた料理レシピカテゴリの構築 およびレシピの自動分類

Construction of a Cooking Recipe Category and Automatic Classification of Recipes using Reviews

金内 萌† 難波 英嗣‡ 角谷 和俊†
Moe Kanauchi Hidetsugu Nanba Kazutoshi Sumiya

1. はじめに

近年、Cookpad や楽天レシピなどのいわゆる投稿型レシピサイトを用いて、日常的に料理をする人が増えてきている。これらの投稿型レシピサイトでは、膨大なレシピデータの中からユーザが目的のものを探すために、キーワード検索を行うか、階層的なカテゴリをたどる必要がある。このカテゴリは、材料(肉・野菜・魚介など)や料理の種類(和洋中)に基づいたものが中心であるため、例えば「鉄分補給」や「子供の運動会」などの目的に応じたレシピを探すのに十分ではない。そこで、本研究では、目的別の料理レシピカテゴリを構築し、そのカテゴリへのレシピの自動分類を目指す。

目的別料理レシピカテゴリを構築する上で、本研究ではレビュー情報に着目する。投稿型レシピサイトでは、タイトル、概要、材料、作り方、コツやポイントを投稿者が自由に記載する。さらに、あるユーザが投稿したレシピに対し、別のユーザがレビューを投稿することができる。レビュー情報には「おいしかった」や「簡単でした」などのレシピの感想だけでなく、例えば「運動会にぴったりの○○」というレシピに対してレビュー投稿者が「BBQにもっていったら喜んでくれました」や「家族でのお花見にもピッタリでした」という記述もある。こうしたレビューは、このレシピには運動会だけではなく BBQやお花見などのシーンに相応しい、または家族で食べるのに相応しいものでもあると考えられる。そこで、本研究では目的別料理レシピカテゴリの構築にレビュー情報を用いる。

ここで、すべてのレシピにレビューが存在するわけではないため、レビューのないレシピは目的別料理レシピカテゴリの分類対象外になってしまう。そこで、本研究では、レビュー情報に基づいて目的別料理レシピカテゴリに分類されているレシピを分析し、各カテゴリの特徴を明らかにした上で、レビューの存在しないレシピの分類を試みる。

本論文の構成は以下のとおりである。2節では情報推薦や特徴抽出に関する関連研究について述べる。3節では目的別料理レシピカテゴリの構築手法について述べる。4節ではレビューのないレシピを分類する手法を実現するためのいくつかの分析について述べる。5節では本研究のまとめと今後の課題を述べる。

2. 関連研究

2.1 レシピ推薦

これまでもレシピ関連の研究は数多くされてきている。レシピ推薦の研究において、門脇ら[1]は、ブログ型レシピを用いたレシピ推薦の研究を行った。ブログ型レシピには日記部分とレシピ部分が記載されている。門脇らの研究では日記部分にあるレシピを作った事由表現を含むキーワードを抽出し、ほかのユーザがレシピを推薦する手法を提案した。門脇らの研究では事由表現を抽出するために日記中の「から」「ので」「ために」「おかげ」「せい」というキーワードに着目して事由文を抽出している。本研究でも特徴や目的を抽出するため、目的を抽出するキーワードとしていくつか定義を行う。

また、池尻ら[2]の研究では意外性のあるレシピに着目したレシピ検索システムを提案している。このシステムはレシピ中の材料から、レシピに意外性があるかを判定する。このような意外性はレシピの特徴であるといえる。本研究ではレビュー情報からレシピの特徴である意外性をみつける。

2.2 レビュー情報を用いた文書のカテゴリ分類

石野ら[3]は商品レビューから語彙構文パターンを用いて購買意図抽出を行い、ユーザの目的に合わせた商品検索ができるシステムを提案している。本研究ではこのシステムのようにレシピを自動分類し、ユーザの意図によってレシピを検索できることを目的とする。

また、金子ら[4]は石野らの研究をもとに、潜在的意図から商品を絞り込むために、階層的知識の自動構築を行った。購買意図の分析からカテゴリの包含関係を構築した。本研究でも、レシピをカテゴリ分類するにあたり、カテゴリの階層化が必要となる。

2.3 レシピの特徴の抽出

Cookpad や楽天レシピなどの投稿型レシピサイトでは、ユーザが各料理タイトルを自由に決めることができる。そして、タイトルには「簡単」や「子供が喜ぶ」という表現がよく見られる。しかし、このような修飾表現は非常に曖昧なものであり、例えば「簡単」であれば、なぜ簡単なのかが一見してわからない。

橋ら[5]はこれらの修飾表現がどのような特徴に基づいてつけられたのかを明らかにする手法を提案した。提案手法として、調理器具と書くレシピの材料リストに着目し、ある一般的な手順と比べて修飾表現の意味を明らかにしている。調理器具と材料との比較はレシピ作成者の主観に基づくものである。したがって本研究ではレビュー情報を用い

†関西学院大学, Kwansai Gakuin University

‡広島市立大学, Hiroshima City University

て、レシピを利用した第三者の情報からレシピの利用目的や特徴を抽出する。

レシピの特徴は手順や見た目、材料がある。このうち、材料では、もともと入っていて特徴のあるものだけではなく、代替することによって特徴のあるレシピといえるものもある。代替の例として、アレルギー食材である牛乳を投入に替えるレシピが挙げられる。花井ら[6]は、この代替食材抽出し、代替食材を提案するレシピ検索のシステムを構築している。花井らは健康向上の効果のある食材を抽出し、レシピ内の材料と代替を行ったが、本研究では健康向上に関わらず、カテゴリにあう代替レシピを抽出することを目的とする。

3. 目的別料理レシピカテゴリの構築

レシピ分類システムの構築は、以下の4つの手順から構成される。

- (1) レビューを解析し、目的等を抽出しカテゴリを構築
- (2) カテゴリごとのレシピの特徴の分析
- (3) 楽天商品レビューの分析
- (4) 2と3を用いたレシピ分類システムの構築

これらのうち、本節では手順(1)について、次節では手順(2)と(3)について述べる。

著者ら[7]は、Cookpadのレビューに、表1に示すタグを自動付与する手法を提案している。

表1: Cookpadのレビュー解析のためのタグ

TARGET	対象者 (例: 子供, 友人, 娘)
REASON	作った意図 (例: 風邪気味だから)
PURPOSE	欲求 (例: ダイエット)
LIKE	好み (例: 好き, 嫌い, 苦手)
TIME	イベント性 (例: クリスマス)

表1の定義に基づいてレビューに人手でタグを付与した例を以下に示す。

- <PURPOSE>美肌</PURPOSE>のためにも毎日飲みます!
- 9/29の<TIME>夕食</TIME>と9/30の<TIME>運動会</TIME>のために作りました♪

上記の例では、「美肌」という文字列にPURPOSEタグが、「夕食」と「運動会」という文字列にTIMEタグが付与されている。著者らは、条件付確率場(CRF)に基づく手法で、このようなタグを自動的に付与するシステムを構築している[7]。

このシステムを用いてCookpadデータベース中のすべてのレビューを解析した。その結果、全レビューに付与されたタグは表2のとおりであった。

表2: 全レビューに付与したタグの数

タグの種類	タグ数
TARGET	21,973
REASON	1,578
PURPOSE	761
LIKE	159,639
TIME	10,685

各タグ内の文字列について頻度の高いものを表3に示す。表3の各文字列の後ろに記載されているカッコ内の数字は、レビュー中の出現頻度を示している。表3の結果に基づき、表4に示すカテゴリを構築した。このカテゴリの構築は、例えば「母」と「お母さん」のような同義語は統合、低頻度語は削除し、また、カテゴリによって極端な偏りが生じないように配慮した。

表3: タグごとの頻出文字列

上位カテゴリ	下位カテゴリ
欲求 (PURPOSE)	健康 (88)、ダイエット (69)、美容 (62)、あげる (18)、大根消費 (11)、かさ増し (10)、カロリーダウン (10)、お肌 (9)、生クリーム消費 (8)、プレゼントする (8)
作成意図 (REASON)	風邪気味 (411)、卵アレルギー (207)、ダイエット中 (117)、風邪っぴき (72)、風邪をひいた (45)、授乳中 (24)、風邪を引いた (21)、糖尿病 (19)、卵アレ (14)
イベント性 (TIME)	健康 (88)、ダイエット (69)、美容 (62)、あげる (18)、大根消費 (11)、かさ増し (10)、カロリーダウン (10)、お肌 (9)、生クリーム消費 (8)、プレゼントする (8)
対象者 (TARGET)	娘 (3999)、子供 (3415)、息子 (2370)、彼氏 (440)、旦那 (387)、夫 (299)、彼 (246)、自分 (184)、家族 (153)、お友達 (141)、友達 (122)、妹 (105)、母 (93)、パパ (93)
好み (LIKE)	野菜嫌い (2524)、私も大好き (1745)、チーズ好き (1653)、お好き (1384)、チョコ好き (1322)、納豆好き (1105)、一番好き (1048)、私の大好き (926)、私の好き (810)

表4: 目的別料理レシピカテゴリ

上位カテゴリ	下位カテゴリ
健康	風邪予防、便秘解消、鉄分補給、疲労回復、健康
美容	美容全般、肌
目的	プレゼント
	量増し
	お弁当
病気	風邪、熱、お腹、体調不良、糖尿病、骨折、貧血、疲れ、入院、病後
アレルギー	卵、乳製品、小麦、エビ、トマト、たまねぎ、筍、かに、貝、その他
イベント	誕生日 (誕生日、はじめての、子供へ、娘へ、息子へ、きょうだいへ、親戚へ、友達へ、自分へ、彼氏へ、お世話になっている人へ)
	その他 (母の日、父の日、バレンタイン、クリスマス、こどもの日、運動会、勝負の日、敬老の日、冬至、雛祭り、女子会、お月見)
食べ手	こども、娘、息子、父、母、家族、彼氏、彼女、友達

次節では、この目的別料理レシピカテゴリにレシピを分類するための基礎的調査の結果について報告する。

4. レシピ分類システムの構築に向けて

4.1 目的別料理レシピカテゴリを用いたレシピ検索プロトタイプシステム

本研究で開発中の目的別料理レシピカテゴリを用いたレシピ検索システムの動作例を示す。図 1 は、システムのトップ画面である。図では、階層的なカテゴリが表示されており、その中のひとつ「ダイエット」をクリックすると、「ダイエット」カテゴリに分類されているレシピ一覧が表示される。なお、現在のプロトタイプシステムでは、レビュー中に明示的に「ダイエット」と記載されたものだけが分類されている。

Cookpad ディレクトリ

- 健康 [風邪予防](#) [便秘解消](#) [鉄分補給](#) [疲労回復](#) [健康](#)
- 美容 [美容全般](#) [肌](#)
- [プレゼント](#)
- [量増し](#)
- [ダイエット](#)
- 病気 [風邪](#) [熱](#) [お腹痛](#) [体調不良](#) [糖尿病](#) [骨折](#) [貧血](#) [疲れ](#) [入院](#) [病後](#)
- [お弁当](#)
- アレルギー [卵](#) [乳製品](#) [小麦](#) [エビ](#) [トマト](#) [玉葱](#) [菊](#) [かに](#) [貝](#) [その他](#)

図 1: 目的別料理レシピカテゴリを用いたレシピ検索システム(トップ画面)

TITLE:お手軽朝食♪梅ぞうすい
DESCRIPTION:雑炊というか、お粥というか。簡単だしあっさりしていて朝食にはぴったりですよ☆
COMMENT:ありがとうございます〜[ダイエット](#)のために久々に私も作る☆

TITLE:ラッシー
DESCRIPTION:さっぱり冷たいヨーグルトドリンク。カレーのお供やお風呂あがりどうぞ。
COMMENT:[ダイエット](#)のためにスキムミルクを入れて食前にすっきり!!

TITLE:生クリーム不要☆桃レアチーズババロア!
DESCRIPTION:生クリーム無しババロアシリーズの新しいバージョン!缶詰を使うので安上がり。このままでも美味しいし、少し固めなので牛乳をかけて食べるのもすごく美味しかったですよ。コーゲン、カルシウムたっぷり美容と健康に!!
COMMENT:夫の[ダイエット](#)のために作りましたが、結局私の方が食べてます★

TITLE:寒天のわらびもち風
DESCRIPTION:寒天をわらびもちのように、きな粉と黒蜜で食べます。娘の[ダイエット](#)おやつです。何十回作ったことか・・・
COMMENT:[ダイエット](#)のために・・・美味しくて病みつきになること間違いなし♪

図 2: 目的別料理レシピカテゴリを用いたレシピ検索システム(「ダイエット」カテゴリのレシピの一覧表示)

4.2 カテゴリごとのレシピの特徴の分析

図 2 において、レビューに「ダイエット」という文字列が記載されていないダイエットに関するレシピもこのカテゴリに分類できるよう、カテゴリごとのレシピの特徴を分析した。その結果、各カテゴリで以下のような特徴がみられた。

- **風邪予防のカテゴリ**：生姜入りの納豆、生姜の砂糖漬け、金柑マーマレード、ニラ入り冷奴、はちみつホットミルクなどのレシピが分類されていた。レシピの内

容としては、生姜、金柑（柑橘類）、ニラ、はちみつを用いたレシピがあった。

- **便秘解消のカテゴリ**：ツナひじきご飯、おからのパウンドケーキ、寒天と黒蜜きなこ、コーヒー寒天、牛蒡と蒟蒻と人参の煮物、おからバナナ、などのレシピが分類されていた。使用されている材料には寒天やおからが目立った。
- **鉄分補給のカテゴリ**：アサイーボウル、鶏レバー料理、ひじきを使用した料理が分類されている。特にタイトルに「ひじき」を含む料理が多く見られた。
- **美容全般のカテゴリ**：人参ドレッシングやにんじんサラダ、ブルーベリーを使用したレシピ、豆乳スープや豆乳ドリンク、バナナブレッドやバナナヨーグルト、トマトを乗せた冷奴やトマトの甘酢漬けなどが分類されている。
- **アレルギー関連の各カテゴリ**：アレルギー食材を除き、別の食材で代用しているレシピが分類された。また、レビュアーが独自に代用した食材も含まれている。その一例が小麦粉を米粉に代用、牛乳を豆乳に代用、エビを豆腐に代用している。
- **運動会のカテゴリ**：手羽元の照り焼き、春巻き、いなり寿司、スイートポテト、卵焼き、サンドイッチ、ミートボール、チャーハン、クッキー、大学芋などの幅広い料理のレシピが分類されていた。

4.3 楽天商品レビューの分析

4.2 節では、各カテゴリに分類されているレシピを分析し、各カテゴリの特徴を明らかにしようとした。例えば、「鉄分補給」カテゴリでは、上述のとおり、レバーやひじきなどが有効な食材であることがわかったが、これら以外にも鉄分補給に有効な食材が存在する可能性がある。そこで、他の情報源として、楽天商品レビューデータの利用を検討した。楽天商品データは、各商品のデータがカテゴリごとに分類されており、さらに、一部の商品にはレビューが存在する。そこで本研究では、楽天商品レビューの中で、食品に関するものを取り上げ、そのレビューを解析することで、食材や料理の特徴を明らかにしようと試みた。

楽天商品カテゴリの中で、まず、「食品」カテゴリ(カテゴリ ID: 100227)から階層を下にたどり、食品に関する1,552 カテゴリを得た。次に、これらのカテゴリに属する商品のレビュー1,267,389 文を収集し、3 節で述べたレビュー解析システムをこれらの文に適用することで、楽天商品カテゴリの特徴を明らかにした。楽天商品カテゴリの中には、例えば、「トップ >食品 >乾物・粉類 >乾物 >ひじき >芽ひじき」のように、食材でまとめられているものも多いため、芽ひじきカテゴリのレビューを解析すれば、芽ひじきという食材の特徴が明らかになる。表 5 に、レビュー1,267,389 文にタグを付与した結果を示す。

表 5: 楽天商品レビューに付与したタグの数

タグの種類	タグ数
TARGET	5,557
REASON	297
PURPOSE	1,714
LIKE	40,940
TIME	4,156

次に、3.1 で作成したカテゴリ名を用いて楽天商品レビューを分析し、それぞれのカテゴリ名では、どのような楽天商品レビューが出現するのかを調査した。調査には、図3に示すような分析システムを利用した。図3は、「ダイエット」でレビューを検索した時の結果を示しており、「黒豆」、「粉寒天」、「サツマイモ」がダイエット目的で頻繁に使われることがわかる。図の「黒豆」などのカテゴリ名をクリックすると、黒豆カテゴリの商品一覧が、また「(レビュー)」をクリックすると、レビュー一覧が表示され、実際にレビュー中でダイエットの記述があることが確認できる。

タグ内キーワードマッチ

カテゴリ名	タグの種類	頻度	レビュー
黒豆	PURPOSE	25	(レビュー)
粉寒天	PURPOSE	13	(レビュー)
サツマイモ	PURPOSE	6	(レビュー)
その他	PURPOSE	5	(レビュー)
その他	PURPOSE	5	(レビュー)
糸寒天	PURPOSE	4	(レビュー)
その他	PURPOSE	3	(レビュー)
お粥	PURPOSE	3	(レビュー)
その他	PURPOSE	3	(レビュー)

図3: 楽天商品レビュー分析システム

分析結果の一部について述べる。上記のシステムを用い、「風邪予防」で検索をした結果、産地を問わず、はちみつの商品がみられた。また、温州みかんやハウスみかん、はっさくなどの柑橘類がみられ、生姜やジンジャーといった材料もみられた。「便秘解消」で検索した結果、シリアル、イチジク、寒天、麦、黒豆、サツマイモ、ブルーベリーなどがみられた。「鉄分補給」ではブルーベリー、コンポート、芽ひじき、レバー、海藻などの商品がみられた。「美容」で検索した結果、油類、雑穀、米加工品、ドイツパン、イチジク、黒豆、麦などの商品がみられた。「アレルギー」で検索した結果、小麦粉、マーガリン、ササニシキ、なたね油などがみられた。

2.3 手順2と3の分析結果に基づいたレシピ分類システムの構築

手順2と3の比較をし、3節で作成したカテゴリにそれぞれどのような特徴があるのかを分析し、その特徴からレビューのついていないレシピを分類するためのシステムを構築する。まず、手順2と3からみられるカテゴリの特徴の例を以下に示す。

- 風邪予防カテゴリに分類した Cookpad のレシピと風邪予防で楽天商品レビューを検索した楽天商品カテゴリ名を比較したところ、Cookpad レシピで使用している特徴的な材料の生姜やはちみつ、柑橘類が楽天商品カテゴリ名でもみられた。したがって生姜や柑橘類、はちみつを特徴的に使用しレシピは風邪予防のカテゴリに分類できるのではないかと考えられる。

- 便秘解消カテゴリに分類した Cookpad のレシピの特徴と楽天商品レビューから便秘解消について検索したときの楽天商品カテゴリ名を比較したところ、材料である寒天がどちらにもみられた。したがってレシピの材料に寒天を用いているレシピが便秘解消カテゴリに分類できると考えられる。
- 鉄分補給のカテゴリに分類した Cookpad のレシピの特徴と楽天商品レビューから鉄分補給について検索したときの楽天商品カテゴリ名を比較した。その結果、ひじき、レバーを主に使用しているレシピは鉄分補給のカテゴリに分類できると考える。

以上の例ではレビューのついていないレシピの分類が可能であると考えられたが、その他に美容全般やアレルギーのカテゴリではレビューのないレシピを分類するための特徴などが見られなかった。例えば、アレルギーのカテゴリについて、Cookpad のレシピは代用したレシピが分類されていたが、楽天商品カテゴリでは小麦粉など代用した根拠となる食材がみられなかった。

5. まとめ

本研究では著者ら[7]のレビュー情報の特徴抽出を用いてカテゴリの分類を行った。タグの自動付与に基づいてカテゴリ分類を行った。そして、類似するカテゴリ名に関してグルーピングを行い、手作業でディレクトリの整理を行った。また、楽天商品レビューでも同様にタグの付与とカテゴリ分類を行った。そして、Cookpad で作成したディレクトリの各カテゴリで使用されている食材と楽天商品レビューから検索できる食品の比較を行った。その結果、相互カテゴリで一致する食材がある場合と一致する食材がない場合があった。

今後の課題として、メイン食材をどのようにして定義するのか検討する。そのひとつの解決方法として、材料表記の上下関係を用いてレシピのメイン食材の定義を行うことが考えられる。また、アレルギー食材の代替を行っているレシピをどのようにして判別するのも検討する必要がある。さらに、提案手法(4)に記したように自動カテゴリ分類システムの構築とその評価方法も検討する必要がある。

6. 謝辞

本研究を遂行するにあたり、Cookpad データと楽天市場データを利用させていただいた。ここに記して謹んで感謝の意を表する。

7. 参考文献

[1]門脇拓也, 山肩洋子, 森信介, 田中克己. (2014). 誕生・使用事由によるレシピ検索～生い立ちレシピサーチ～. DEIM Forum 2014.

[2]池尻恭介, 清雄一, 中川博之, 田原康之, 大須賀昭彦. (2013). 食材に基づいたレシピの以外度算出手法の提案. 電子情報通信学会技術研究報告.

[3]石野亜耶, 村上浩司, 関根聡. (2014). 商品レビューからの購買意図の抽出とそれを用いた商品検索システムの構築. 言語処理学会 第20回次大会 発表論文集 pp.622-625.

[4]金子貴美, 村上浩司, 石野亜耶. (2016). 商品レビューからの購買意図の抽出と自動階層化. 第30回人工知能学会全国大会.

- [5] 橘明穂, 若宮翔子, 難波英嗣, 角谷和俊. (2013). 料理名のネーミングコンセプト抽出によるレシピ推薦システム. ARG Web インテリジェンスとインタラクシオン研究会, No.2.
- [6] 花井俊介, 難波英嗣, 灘本明代. (2015). 健康を意識した代替食材の発見手法. DEIM Forum 2015.
- [7] 金内萌, 難波英嗣, 角谷和俊. (2016). 投稿型レシピサイトにおけるレビュー情報に基づく料理タイトル自動生成. DEIM Forum 2016.