

レビュー情報を用いた料理レシピの特徴分析によるカテゴリ生成 およびレシピタイトルの自動生成

金内 萌† 難波 英嗣‡ 角谷 和俊†

† 関西学院大学総合政策学部 〒669-1337 兵庫県三田市学園2丁目1番地

‡ 広島市立大学大学院情報科学研究科 〒731-3194 広島市安佐南区大塚東3-4-1

E-mail: †{fgx42122,sumiya}@kwansei.ac.jp, ‡nanba@hiroshima-cu.ac.jp

あらまし 近年 Cookpad や楽天レシピなどの投稿型レシピサイトが普及している。レシピの投稿者はレシピを投稿する際に独自のタイトルをレシピにつける。しかしこれらのタイトルは必ずしもレシピの特徴を捉えることができているとは限らない。そこで本研究ではレシピの特徴を捉えたタイトル生成システムの構築を目指す。その手段として投稿型レシピサイトにあるレビュー情報を用いる。また、タイトルの自動生成の第一歩として特徴別のカテゴリ生成を行う。

キーワード レビュー分析, コンセプト生成, カテゴリ構築

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。また, それに伴い, 自作のレシピを投稿し, 多くの人に共有するユーザも増えてきている。これらの投稿型レシピサイトでは, ユーザは独自でレシピにタイトルを付与する。多くのレシピタイトルがレシピの特徴を示している修飾語とレシピのメニュー名を組み合わせたものであるが, そのタイトルが本当にレシピの特徴を示しているものとは限らない。さらに初めてレシピ投稿する人にとって, レシピの特徴を捉え, タイトルに修飾語を付与することは困難な場合もある。また, このような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るために, 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。そして分類したレシピの特徴を分析し, レシピを投稿した際に自動でタイトルを生成できるシステムの構築を目指す。

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はなく, 例えば「運動会にぴったりの〇〇」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もらえました」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いう新たな特徴を示している。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。

橘ら[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。

2.3 食材名の抽出

森ら[7]は手続き文書のフローグラフ化を試みるために、料理レシピに着目した。森らの作成するフローグラフでは、食材や道具を頂点とし、食材や道具の関係性を偏として作成を行った。頂点となる食材や道具に関しては野菜や調理器具といった通常用いられる固有表現だけではなく、食材や調理者の動作を表す用言も含まれている。関係性に該当する辺では頂点の述語項関係や参照関係などその他の関係をラベルとして示している。そして、この頂点や辺をそれぞれ基準化し、フローグラフに変換した。

本研究ではこのフローグラフを用いて文章中から食材名を抽出する。抽出した食材名からカテゴリ名の特徴の抽出を試みる。

3. 目的別料理レシピカテゴリの構築

3.1 レビューの自動解析

著者ら[8]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[8].

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.

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.

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), 風邪っぽい (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

3.2 レビューの解析結果に基づいた目的別料理レシピカテゴリの構築

我々は、これまでに、レシピのレビューを解析し、料理を作成する目的を分類・整理し、カテゴリとしてまとめている[8]。現在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 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。コラーゲン、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. レシピ分類システムの構築

料理レシピの目的別料理レシピカテゴリへの自動分類

3節で述べた手法で、目的別料理カテゴリにレシピが登録されているが、これらはCookpadの全レシピを登録することはできていない。そこで、本研究では、この目的別料理カテゴリに、料理レシピを自動分類する2種類の手法を提案する。これらの手法について、4.1節、4.2節でそれぞれ述べる。

4.1 各カテゴリの特徴に基づいた分類手法

4.1.1 カテゴリごとのレシピの特徴の分析

3節で述べた目的別料理カテゴリの各カテゴリの特徴を明らかにするため、すでに分類済のレシピを分析した。分析の結果、4つのカテゴリ「風邪予防」「便秘解消」「鉄分補給」「疲労回復」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。
- **疲労回復のカテゴリ**：酢や豚を用いたレシピが分類されていた。

これらのカテゴリでは、使われる食材に特徴があることが分かった。そこで、Yahoo!知恵袋を用いて、各カテゴリを特徴付けるより多くの食材の自動検出を行った。

4.1.2 Yahoo!知恵袋からのカテゴリを特徴付ける食材の自動検出

本研究では、知恵袋のデータの中でも、「暮らしと生活ガイド」カテゴリ中の「料理、レシピ」カテゴリを使用する。この知恵袋データには、多くの投稿者が様々な目的に対して効果のある食材やレシピを質問し、回答が行われている。


 [redacted]さん

2015/3/12 20:48:43

風邪の時にいい食べ物は何ですか？


 [redacted]さん

2015/3/12 22:51:01

ベストアンサーに選ばれた回答

- ・寒気のする風邪の時は身体を温める食材です。
⇒生姜、ねぎ、葛、鶏肉、黒砂糖、シナモンなど。白砂糖は身体を冷やすのでご注意ください。市販のしょうが湯は葛も入っている事が多いのでお勧めです。
- ・寒気は無く熱や喉の乾燥がある時は身体の熱を冷ます食材です。
⇒柑橘系の果物、トマト、大根、白菜、きゅうりなど。

寒い風邪と熱い風邪を区別する事が大切です。

図 3：知恵袋の質問と回答の例

この例の回答にある生姜やねぎなどの食材が自動抽出できれば、風邪に効果的な食材が自動的に収集できることになる。

本研究では、森ら[7]のレシピを対象にした固有表現抽出器を知恵袋データに適用することで、各カテゴリに特徴的な食材を自動収集した。その結果の一部を表[6]に示す。なお、森らのシステムは料理レシピを入力の前提としているため、知恵袋にそのまま適用すると食材以外のものも誤って抽出されてしまう。そこで森らのシステムで抽出された食材候補を、Cookpad データベースに含まれる全レシピの材料リストと比較し、材料リストに含まれないものは候補から除外した。また、例えば「鉄分」などのカテゴリ名の一

部になっているような単語も排除した。

上記の手法で知恵袋から抽出した食材名について、抽出した単語がカテゴリ名に関係のある単語であるのかを評価する。そのために、管理栄養士の有資格者にこれらの単語を評価してもらった。

評価方法として、食材名それぞれにカテゴリ名と関係がある 3 点、少し関係がある 2 点、関係がない 1 点の点数を振ってもらった。

その結果、風邪予防の項目では、44.93%の確率で、便秘解消の項目では 54.55%の確率で、鉄分補給の項目では、60%の確率で、疲労回復の項目では 50.16%の確率で項目に関係する食材が抽出された。

表 6：知恵袋から抽出したカテゴリ名ごとの食材名

カテゴリ名	自動で抽出された食材名
風邪予防	からし, 生姜, ネギ, ねぎ, マスタード, みかん, ジャガイモ, ご飯, 体, メロン, カロリー, キャベツ, ニラ, 湯, 人参, ブロccoli, 鶏肉, 梨, イチゴ, 大根, レモン, ニンニク, 皮
便秘解消	ヨーグルト, 皮, バナナ, 水分, 寒天, 水, ペクチン, オールブラン, オリブオイル, レタス, 納豆, 腸, さつまいも, 牛乳, 蕎麦, メロン, レーズン, 酢, 果物, ひじき, 小麦粉, 煮物, サツマイモ, 玄米, カスピ海ヨーグルト, プルーン, 大豆, 豚, コーンフレーク, 体, コーヒー, リンゴ, おから,
鉄分補給	レバー, 茶, 豚, プルーン, 鶏, 牛肉, 鶏レバー, 湯, 牛乳, 食材,
疲労回復	酢, 豚肉, にんにく, ニンニク, 梅干, カルシウム, 砂糖, 体, クエン酸, らっきょう, 黒酢, ネギ, バナナ, ねぎ, タマネギ, うなぎ, 水分, 牛肉, 山芋, 糖分, 梅干し, たまねぎ, 豆腐, タコ, 果糖, トマト, リンゴ, 梨, 玉ねぎ, 納豆, カロリー, レモン, 豚, イカ, 食材, B, ニラ, 乳酸, キムチ, C, 果物, ベーコン, エネルギー, ヨーグルト, 鶏肉, ナス, 味噌汁, 部分, 酢の物, C, 塩, 蜂蜜, ウナギ, 牛乳, コーヒー, 生姜, B, 炭水化物, チョコレート, パセリ, 麺, 種, 鳥, ひよこ豆, チョコ, オイル漬け, 酢漬け, 例えば, フルーツ, 玄米, ラーメン, ペクチン, しょうが, モロヘイヤ, 糠, パイナップル, 昆布, 焼き肉, ご飯, 酢豚, ラッキョウ, 梅酒, 味噌, 卵, カレー, 梅, オクラ, ほうれん草, ステーキ, グレープフルーツ, 実, 栄養ドリンク, 豚キムチ, エビ, メロン, メープルシロップ, 柿, ごま, 材料, うどん, オイル, ビール, グリーンピース, 糖, 油,

5. まとめ

本研究では著者ら[8]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、知恵袋によってカテゴリごとの食材名を抽出した結果、食材名以外の単語も抽出されたが、食材名に着目するとカテゴリ名と関係のある食材が抽出された。その結果をもとに、管理栄養士に抽出した食材を評価しても

らった結果，約半分の確率で効果のある食材が抽出されることが分かった。

今後の課題として，抽出された特徴語の評価をどのように行うかを考える必要がある。またタイトル生成を行うためにタイトルの構造について検討することが必要になる。そして，現在キュレーションメディアの情報の不正確さが問題視されている。本研究でもレビュー情報から特徴を抽出しているため，それらのレビューの信憑性・信頼性の評価が必要になるだろう。例えば鉄分補給のカテゴリについて，鉄分補給は適正な量の摂取は身体によい効果をもたらすが，過度な摂取は身体に胃腸障害などの悪影響を及ぼす場合もある。そのため専門的な補足情報も必要になる。

6. 謝辞

本研究を遂行するにあたり，Cookpad データと Yahoo! 知恵袋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] 橋明穂，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ョン研究会, No.2.
- [6] Hanai, S., Nanba, H., and Nadamoto, A. (2015) “Clustering for Closely Similar Recipes to Extract Spam Recipes in User-generated Recipe Sites”. In Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS2015) .
- [7] 前田宏邦，山形洋子，森信介。(2016). レシピテキストからのフローグラフコーパス作成. データ工学と食メディア.
- [8] 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.