

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

金内 萌[†] 難波 英嗣[‡] 角谷 和俊[†]

[†] 関西学院大学 〒669-1337 兵庫県三田市学園 2-1

[‡] 広島市立大学 〒731-3194 広島県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 現在広く普及しているクックパッド¹や楽天²レシピなどの投稿型レシピサイトでは、ユーザーが独自にレシピのタイトルをつけることが可能であるが、そのタイトルが、必ずしもレシピの特徴を適切に表しているとは限らないため、その特徴を的確に表したタイトルを自動生成することを目指す。その第一歩として、本研究では、レシピのレビューから、そのレシピの特徴を示す表現を自動的に抽出する手法を提案する。

キーワード 情報推薦, 特徴抽出, 文生成

1 はじめに

投稿型レシピサイトでは主に、タイトル、概要、材料、作り方、コツやポイント、レビューから成り立っている。本研究では、投稿型レシピサイト特有のレビュー情報を用いてタイトルを自動生成し、投稿者がタイトルを付けやすいように、さらに検索者がより検索を行いやすくすることを目的とする。

投稿型レシピサイトではユーザーが自由にレシピのタイトルをつけることができる。ユーザーはメニュー名だけをタイトルとしてつけることもできる。しかし、多くのレシピ投稿者はそのレシピを作った目的、レシピの特徴、イベント性などを考慮して「クリスマスにぴったりのケーキ」や「運動会におすすめのお弁当」というようなタイトルをつける、このようなレシピタイトル付けは、料理が得意である人や、投稿型レシピサイトに慣れている人にとっては容易である。しかし、料理が苦手な人や投稿型レシピサイトを使い慣れていない人にとって、このようなタイトルの考案は難しい。タイトル付けに不慣れな人がタイトルを自ら付けた場合、情報の不足したタイトルやレシピの特徴を捉えきれないタイトルをつけることもある。したがって、本研究ではレシピ中の情報からレシピの特徴を抽出し、抽出した特徴を用いてタイトルの自動生成を行う。

投稿型レシピサイトではレビュー機能を備えている場合が多い。レビュー情報はレシピの感想だけではなく、実際にレシピを利用した人がどのような目的で利用したかという情報やレシピがどのような特徴をもっていったかという情報も含まれている。したがって、レシピの作者でも気づかなかった特徴や目的が含まれている可能性もある。同様に商品サイトにもレビューの機能がついており、安藤ら[2]はレビューに書かれ

ている購買背景や売り手への欲望などを有効活用するためにレビューに記述される内容を分析した。本研究ではレビュー情報を用いることによって、ユーザーの欲求や目的などを反映したタイトルの自動生成を行う。

投稿型レシピサイト中のレビュー情報を用いるため、本研究ではクックパッドのレシピデータからレビュー情報に該当する「つくれば」のデータを扱う。このデータからレシピの特徴や目的を抽出し、それぞれのレシピに適切な自動タイトル生成を行う。なお、タイトルは今後投稿されるレシピや、今までに投稿されてきたレシピに反映させる。

本研究によって、投稿型レシピサイトの使用が初心者である人、レシピの特徴をつかむことが困難な人に対してタイトル生成の手助けとなる。また、投稿型レシピサイトの利用者がレシピを検索するとき、目的や特徴に合わせたレシピを検索することも可能になる。

2 節ではレシピ推薦と特徴抽出についての関連研究の紹介を行う。3 節では提案手法としてシステムの概要やプロセスを述べる。4 節では実験の内容を述べる。5 節で本研究のまとめを述べ、第 6 章で今後の研究課題を示す。

2 関連研究

本研究中ではレシピのレビュー情報から特徴や目的を抽出している。これまでの研究でもすでにレシピ情報中の材料の差異などから特徴を抽出している研究がある。また、ユーザーの意図に沿った検索を可能にする研究も行われた。本節でその関連研究の内容を述べる。

¹ <http://cookpad.com/>

² <http://recipe.rakuten.co.jp/>

2.1 タイトルの曖昧性

クックパッドや楽天レシピなどの投稿型レシピサイトでは、ユーザー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。Wakamiyaら[1]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。

2.2 レシピ推薦

これまで、レシピ関連の研究は数多くされている[3,4,5]。レシピ推薦の研究において、門脇ら[6]は、ブログ型レシピを用いたレシピ検索の研究を行った。ブログ型レシピには日記部分とレシピ部分が存在する。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、ほかのユーザーがレシピを決める事由と対応付けを行い、ユーザーに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する。

2.3 特徴抽出

本研究では投稿レシピの特徴や目的を抽出し、タイトルに反映させることを目的としている。池尻ら[7]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。システムの概要として、レシピ中の材料からそのレシピが意外性のあるものかを判定する。このような意外性のあるレシピは特徴を含むレシピであるということが出来る。本研究では材料中からの意外性ではなく、レビュー情報から意外性も含む、特徴を抽出する。

また、石野ら[6]は商品レビューからご意向文パターンを用いて購買意図抽出を行い、ユーザーの目的に合わせた商品検索ができるシステムを提案した。このような商品検索と同様に本研究では最終的にユーザーの意図によって検索をできるようにすることを目的とする。

3 提案手法

本研究ではクックパッド株式会社が提供するレシピデータからレビューに該当する「つくれば」のデータを用いて実験を行う。本研究のシステム概要を図1に示す。

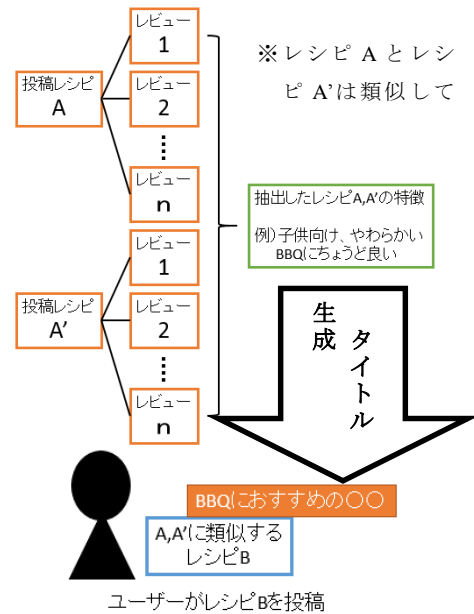


図 1：本研究のシステム概要

3.1 類似レシピの定義づけ

投稿されたレシピはメニューごとにカテゴリ分けが可能である。本研究ではメニューのカテゴリ分けについて、材料や手順を用いて分類を行う。材料によるカテゴリ分類の際に、投稿型レシピサイトでは表記方法が自由であるため、材料名の表記の違いが問題となってくる。例えば「ホットケーキミックス」は「HM」「ホケミ」など、複数の表記の違いがある。この違いを改善するために、料理オントロジー[8]を用いる。

3.2 レビュー情報からの特徴抽出

レビュー情報から特徴や目的の分析を行い、類似するレシピ集合にはどのような特徴や目的があるのかを分析する。分析の手法として、目的を導くキーワードとして「ために」「用に」「ので」「目的で」を現段階の研究では用いることとする。これらのキーワードの直前に出現する単語を目的や特徴とする、例えば「～子供用に作りました」という文章であれば、目的として「子供」を抽出する。この出現する単語は多くあると考えられるため、頻出度の高い単語を対象とする。

3.3 自動タイトル生成

自動タイトル生成のために、基本的にどのようなタイトル付けがなされているのかを検討する必要がある。これまで投稿型レシピサイトに投稿されてきたレシピはほとんどの場合「レシピの特徴または目的+レシピカテゴリ名」で成り立っている。このことを考慮してレシピカテゴリ名にレシピの特徴または目的をつなげることができるように文を変形させる必要がある。このようなレシピタイトル以外にも「レシピの特徴またはタイトル名★レシピカテゴリ名」というように、レシピの特徴または目的を変形させる必要のないタイトルもある。しかし、本研究では前述した文を変形させてレシピカテゴリ名の前に特徴や目的を修飾表現として用いるタイトル生成を最終的な目標とする。

4 実験と考察

本研究では「ために」などの文の前に記された文を特徴や目的を示している手がかり語とする。例えば「子供のために作りました」や「バレンタインデーのために作りました」という文章では、「子供」や「バレンタインデー」が特徴や目的を示す。そして、これらの特徴や目的を示す単語は人物や欲求、イベントなど、いくつかの種類に分類することができるため、それぞれの特徴や目的がどのような種類であるのかの判定を行う。本研究ではこれらの特徴や目的を判定するため、クックパッドデータのつくれぽを用いて抽出を行う。4.1節では抽出するレビュー情報の特徴や目的を表す単語の分析を考察し、4.2節ではレビュー情報から特徴や目的を自動判定する実験を行った。4.3節では自動判定による実験結果を述べる。

4.1 特徴や目的を示す単語の考察

本節ではつくれぽから特徴や目的を判定する手法について述べる。まず、すべてのレビュー情報から特徴や目的を抽出して、クラスタリングを行う。クックパッドデータのつくれぽ全体（レビューとそれに対する返信）から形態素解析エンジン MeCab を用いて「ために」という手がかり語の直前にある単語を抽出した。出てきた結果として、「父」「娘」「子ども」「クリスマス」「バレンタイン」などの目的を表す単語が導きだされた。しかし、「日」や「この」などのさらに前文に目的や特徴を示す単語も抽出された。「この」という単語はどこに目的語が存在するか不明であるため除外した。「日」という単語は直前に「誕生日」や「母」、「運動会の」などといった、比較的重要な単語を含む場合があるため、さらに形態素解析をし、分析する必要がある。

4.2 レビュー内容の分析

抽出した特徴や目的は人や欲求、イベントなどのいくつかの種類に分類することができる。したがって、まずは人手によって判定を行った。つくれぽすべてのデータの中から「ために」という手がかり語を含む1000件を抜き出し、レビュー内の目的や特徴にタグ付けを行った。表1にタグの種類とその概要を示す。表1の基準でタグ付けを行い、CRFを用いて機械学習を行った。

表1：タグの種類とその概要

| | |
|---------|------------------|
| 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） |

また、特徴や目的の判定を行う際に機械が<TARGET>と<TIME>が混同してしまうことが考えられた。例えば「子供のために」や「バレンタインデーのために」だと、機械がバレンタインデーについて<TARGET>のタグを付与する場合がある。この混同を避けるために、<TARGET>と<TIME>について、それぞれ WordNet を用いて人手で対象者やイベント性の辞書を作成した。作成した辞書を表2に示す。自動でタグを付与する際、ここで作成した辞書を用いて<TARGET>と<TIME>を分類する。

表2：<TARGET>と<TIME>の辞書

| | |
|----------|--|
| <TARGET> | 息子 長男 家族 ファミリー 親戚 ぼっちゃん 娘 ガール 子供 こども 子ども ちびっ子 ちびっこ ちび 子 ベビー 赤ちゃん 実家 夫 旦那 亭主 主人 彼氏 彼 ボーイフレンド 自分 自身 私 僕 友達 友 友だち 友人 父 お父さん 父親 父さん とうさん パパ 母 お母さん ママ 大人 |
| <TIME> | 休日 年越し 正月 お正月 七草がゆ 七草粥 節分 ひな祭り ひなまつり 雛祭り 雛祭 花見 子どもの日 記念日 体育祭 七五三 七夕 バレンタインデー バレンタインデー バレンタイン・デー ホワイトデー ホワイトデー ホワイト・デー 朝 会 デー 誕生日 バースデー バースディ バースデイ 父の日 母の日 初節句 ギフト プレゼント 手土産 お土産 祝儀 おくり物 おくりもの 贈り物 贈りもの パーベキュー BBQ パーティー パーティ 宴会 宴 運動会 |

4.3 自動判定による実験結果

人手で対象者やイベント性の辞書作成をする前の実験結果を表3に示す。表1より、LIKEタグは高い精度かつ高い再現率共でタグの検出を行うことができた。また、再現率に関してTARGET, PURPOSE, TIMEが低い結果となった。

表3：レビュー情報抽出の実験結果

| タグ | 精度 | 再現率 |
|---------|------|------|
| TARGET | 0.75 | 0.21 |
| REASON | 0.70 | 0.47 |
| PURPOSE | 0.67 | 0.25 |
| LIKE | 0.81 | 0.50 |
| TIME | 0.60 | 0.20 |

次に、自動判定の際に辞書を用いた場合についての結果を表4に示す。<TARGET>タグについて、精度は落ちたが再現率が上がった。これは、辞書によって<TIME>タグやその他のタグと選別することができたため、上がったと考えられる。<TIME>タグについては下がっており、辞書がなかった場合、対象者が<TIME>タグとしてタグ付けされていたと考えられる。

表5：辞書を用いた場合の自動判定

| タグ | 精度 | 再現率 |
|---------|------|------|
| TARGET | 0.74 | 0.84 |
| REASON | 0.61 | 0.36 |
| PURPOSE | 0.75 | 0.32 |
| LIKE | 0.88 | 0.87 |
| TIME | 0.75 | 0.10 |

5 まとめ

本研究の実験では手がかり語として「ために」だけを実験的に用いた。4.1節に示したように、直前に特徴や目的が示されることもあるが、直前ではなくそれ以前に特徴や目的が示されることも分かった。

さらにレビュー情報の文章から手作業でタグ付けを行い、CRF手法を用いた実験を行った。その結果、精度と再現率はタグによって良いものと悪いものがあった。この結果から、さらに自動判定する際に混同する可能性のある<TARGET>タグと<TIME>タグについて辞書を作成し、改めて自動判定を行ったところ、精度と再現率に変化があった。

6 今後の課題

実験について、4.1で示したように、特徴や目的を導く表現の直前に必ずしも特徴や目的がないため、それ以前にある表現の抽出について再度検討する必要がある。

また、4.2の実験について、精度や表現率について、タグの付与の仕方から検討しなおし、今回の<TARGET>タグ<TIME>とタグのように辞書を作成し、改めて実験を行う必要性も考えられた。

今回の実験では全レビュー情報から特徴や目的を抽出している。本研究の最終的な目標としては、類似レシピのグループにあるレビュー情報を用いて特徴や目的を抽出しなくてはならない。したがって、今後の研究では類似レシピグループから手がかり語を用いて特徴や目的の抽出を行う。

最終目標である自動タイトル生成には、タイトルの形がどのようなものであるか決めることと、すべてのユーザーが理解できるタイトルを生成するために、単語をどのように並べ、どのように繋いでタイトルを生成するのかを検討する必要がある。

謝辞

本研究を遂行するにあたり、クックパッド株式会社と国立情報学研究所が提供する「クックパッドデータ」を利用した。ここに記して謹んで感謝の意を表す。

参考文献

- [1] WakamiyaShoko, KawaiYukiko, NanbaHidetsugu, SumiyaKazutoshi. (2014). "Extracting Naming Concepts by Analyzing Recipes and the Modifiers in Their Titles". Transactions on Engineering Technologies, pp. 321-335, December 2014.
- [2]安藤まや, 関根聡. (2014). レビューには何が書かれていて、読み手は何を読んでいるのか? 言語処理学会 第20回年次大会 発表論文集.
- [3]荻米志帆乃, 藤井敦. (2008). 料理どうしの類似と組合せに基づく関連レシピ検索システム. 言語処理学会第14回年次大会発表論文集.
- [4]三田村祐介, 塩澤秀和. (2007). 食材の優先度を考慮した料理レシピの検索. 情報処理学会研究報告ヒューマンコンピュータインタラクション (HCI).
- [5]上田真由美, 石原和幸, 平野靖, 梶田将司, 間瀬健二. (2008). 食材利用履歴に基づき個人の嗜好を反映するレシピ推薦手法. DBSJ Letters Vol.6, No.4.
- [6]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回年次大会 (NLP2014).

- [7]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意外度算出手法の提案. **IEICE Technical Report**.
- [8]土居洋子, 辻田美穂, 難波英嗣, 竹澤寿幸, 角谷和俊. (2014). 料理レシピと特許データベースからの料理オントロジーの構築. 信学技報 **IEICE Technical Report**.
- [9]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**.