

レシピサイトのユーザコメントデータを対象とした レシピにおける代替食材抽出方式

瀧 日悠雅[†] 田丸 翔大[†] 臼木 瑠音[†] 中西 崇文[†]
武蔵野大学 データサイエンス学部 データサイエンス学科[†]

1. はじめに

近年、一般的に料理経験が浅い人や初めて作る料理の場合、インターネット上のレシピサイトを参照して料理をすることが多い。レシピサイトには、実際にその料理を作った人によるコメントが付いており、その料理の難易度や、アレンジした点、注意点などが記載されている。その中には、味付けや、値段や手に入らなかったなどの事情から別の食材を使用したことが記載されている。特に、別の食材を提案しているコメントについては、料理経験が浅い人では到底思いつかないような食材を提案している。

一方、代わりとなる食材の情報を得るためには、レシピサイトにある膨大なコメントを読み、その中から探す必要がある。そのため、レシピサイトから代わりとなる食材を知ろうとするとそのレシピサイトのコメントを読むための時間と手間がかかる。このことから、レシピサイトのコメント欄を対象として、そのコメントデータから代替食材を自動的に抽出し、ユーザに提示するシステムが実現できれば、ユーザにとって、代替となる食材を使った料理体験の向上が期待されると考えられる。

本稿では、レシピサイトのユーザコメントのテキストデータを対象とした、レシピにおける代替食材の抽出方式について示す。本方式では、レシピサイトのユーザコメントのテキストデータから、そのレシピで利用可能な代替食材を示す語を抽出する。さらに、本方式では、レシピと抽出された代替食材のデータから、新たなレシピの代替食材を推定することを可能とする。本方式は、料理の知識の浅いユーザでも、レシピを応用し、代替食材を使った新たなレシピを創造することができる。

2. 関連研究

レシピサイトのテキストから代替食材案を提示する研究としては、素材と調理動作の関連付けを行い、素材間の類似性から新たな代替素材を得る

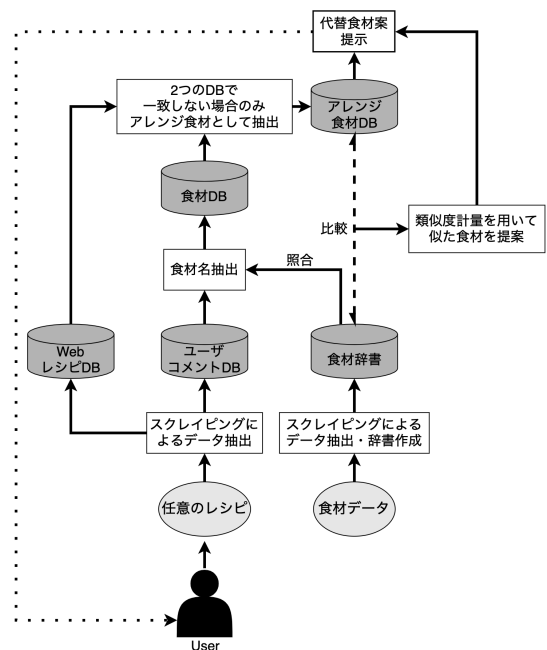


図1：レシピサイトのユーザコメントデータを対象としたレシピにおける代替食材抽出方式

というものがある[1]。この研究は調理動作に大きく依存するため、調理動作が具体的に記述されていることが前提となっている。そのため、調理動作が曖昧な場合には代替食材をうまく提示できない。それに対し、本研究は調理動作に依存せず、実際に調理を試みたユーザの経験が記述されたコメントに基づき、代替食材を提示することができる点が先行研究と異なる。

3. レシピサイトのユーザコメントデータを対象としたレシピにおける代替食材抽出方式

3.1 代替食材抽出方式の全体像

図1に代替食材抽出方式の概要を示す。本方式は、ユーザが指定した任意のレシピからのスクレイピングによるデータ抽出・辞書作成、食材辞書とユーザコメントDBから食材名抽出、WebレシピDBと食材DBからアレンジ食材抽出、類似度計量を用いたアレンジ食材抽出によって構成される。

本方式は、レシピサイトの一つである楽天レシピ[2]のレシピそれぞれにユーザが投稿した「つくったよレポート」[3]というユーザコメントを対象として、そのユーザコメントに見られる食材とレ

Alternative Ingredient Extraction Methods in Recipes for User Comment Text Data

Hyuga Taki[†], Shota Tamaru[†], Rune Usuki[†], Takafumi Nakanishi[†]

[†] Musashino University, Department of Data Science

レシピの材料欄に記載されている食材を比較し、一致しない場合、その食材をアレンジ食材として抽出をする。また、ユーザコメント内に存在しない食材をアレンジ食材として提案するために、食材辞書内の食材とアレンジ食材 DB 内の食材の類似度計量を行い、類似度計量の数値が大きいものをアレンジ食材として提案する。これにより単純なパターンマッチングだけではなく類似性に基づく食材の提案が可能となる。

3.2 ロコミデータからの代替食材抽出

レシピサイト内のユーザが選択した任意のレシピのユーザコメント欄、レシピページの材料欄をスクレイピングし、ユーザコメント DB と Web レシピ DB に格納する。また、ユーザコメント DB から食材を抽出するため、食材名とその特徴を示すデータを食材辞書とする。食材辞書の作成手順は 3.3 節で述べる。この食材辞書の食材名はパターンマッチングにより抽出することができる。抽出した食材名を食材 DB に格納する。この食材 DB と Web レシピ DB 内の食材名を文字列として捉え、パターンマッチングにより一致しないものをアレンジ食材としてアレンジ食材 DB に追加し、そのアレンジ食材を代替食材案として提示する。

3.3 類似度計量による新たな代替食材抽出

レシピのユーザコメントに存在しないアレンジ食材を類似度計量により抽出する。食材の特徴を用いた類似度計量を実現することにより、代替食材の抽出が可能になる。

食材辞書の作成方法については次の通りである。まず、食材辞典サイトである e 食材辞典[4]から、それぞれの食材の特徴をテキストで記載したデータをスクレイピングにより抽出する。食材の特徴を示すテキストデータには、食材名が記載されていない場合がほとんどであるため、食材名を付与する。その後、文章の分かち書きを行い、助詞や助動詞を除去した処理を得て作成されたものを食材辞書とする。この食材辞書と 3.1 節のアレンジ食材 DB に格納されているデータを word2vec に適用する。これを用いて食材の類似度を計量することにより、3.2 節の手法で見つけることのできない代替食材案を発見することが可能となる。

4. 実装

3 節で示した本方式のプロトタイプシステムを実装した。表 1 に 3.3 節で述べた類似度計量による新たな代替食材抽出の実行結果を示している。表 1 により、油揚げと厚揚げの類似度は 0.98、アナゴとウナギの類似度は 0.93 と高くなっており、これらは一般的に代替食材としても使うことが可

表 1: 類似度計量による新たな代替食材抽出

	食材 1	食材 2	類似度計算結果
0	油揚げ	厚揚げ	0.989401
1	アナゴ	ウナギ	0.932655
2	アオサ	アズキ	0.672734
3	ミカン	リンゴ	0.922539
4	しそ	アズキ	0.840493

能であることから、類似度計量による新たな代替食材抽出が実現できることを示唆している。

5. おわりに

本稿ではレシピサイトのユーザコメントデータから食材名を食材辞書を用いて抽出し、Web レシピ DB と比較して、アレンジ食材を抽出する方法と、アレンジ食材 DB を用いて類似度計量し、アレンジ食材を抽出する手法を示した。本方式により特定レシピのユーザコメントデータ内のアレンジ食材が 2 つの DB を比較し、一致しないものをアレンジ食材として提示できる。また食材間の距離感からユーザコメントデータに存在しない、新たな代替食材を提示することも可能となる。これにより、料理経験が浅く代替食材などが思い浮かばないユーザも気軽に代替食材を知ることができ、料理の幅が広がることが期待される。

今後の課題としては、ユーザの料理経験を考慮したレシピの提示、調理工程での注意点などをユーザコメントデータから抽出し提示、本方式を使用したサービスを実際に利用したユーザのアンケートによる本方式の有効性などの検証が挙げられる。

参考文献

- [1] 志土地由香, 高橋友和, 井手一郎, 村瀬洋, 調理レシピテキストからの代替素材の発見. In 人工知能学会全国大会論文集 第 22 回全国大会, 1B1-2, 2008
- [2] 楽天レシピ: レシピを見つける, 料理する, 投稿する, みんなで作る
<http://recipe.rakuten.co.jp/>
- [3] つくったよレポート | 人気メニューのレシピ | 楽天レシピ
<https://recipe.rakuten.co.jp/category/30/report/>
- [4] e 食材辞典 - e ヘルシーレシピ - 第一三共株式会社
<https://www.ehealthyrecipe.com/meal/syokuzai/index.php>