

材料とその調理法に基づく料理間類似度計量による レシピデータ検索方式

有馬 直孝[†] 岡田 龍太郎[†] 峰松 彩子[†] 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、インターネット上に、膨大で多様なレシピデータが生成、蓄積、配信されている。これらのレシピデータによって、ユーザ自身が調理したことがない未知のレシピに触れる機会が増大している。これらのレシピデータの活用によって、ユーザが多様な料理を楽しむ機会が増大した一方で、これらのレシピデータを対象として、ユーザの嗜好に合致したレシピ検索や、料理法を効率的に見つける手法が重要となっている。

一般的に、料理の味覚や食感は、材料とその調理法によって決定されると考えられる。インターネット上に生成、蓄積、配信されている膨大で多様なレシピデータから、料理の味の特徴に関与する材料と調理法の単語を自動的に抽出する機構が実現できれば、ユーザの味覚の嗜好に合致したレシピデータ検索、推薦システムが実現可能になると考えられる。

本稿では、材料とその調理法に基づく料理間類似度計量によるレシピデータ検索方式について示す。本方式では、料理の味覚や食感を材料とその調理法によって決定されるものとし、レシピテキストから材料とその調理法をメタデータのペアとして抽出し、そのメタデータのペア同士の類似性を計量することで、料理間の味覚や食感の類似するレシピの検索を実現する。本方式を用いることで、ユーザにとって未知の料理のレシピから類似する料理のレシピを検索す

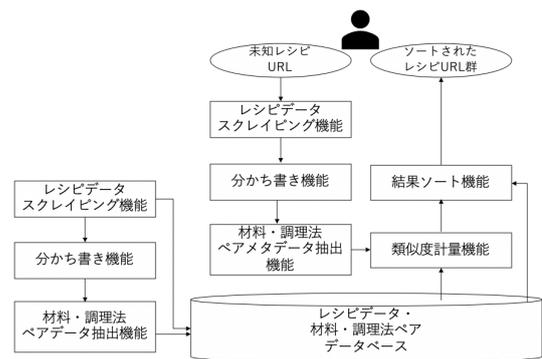


図 1：システム構成図

ることができ、出力された既知の料理から仕上がりを類推することで、安心して料理に取り組むことができるのに加え、この類推を元にしたユーザの好みに合わせた自作料理作成の手助けになると考えられる。

2. 関連研究

料理間類似度を計量する研究として、福本[1]らはレシピにおける材料の量を重要度として抽出し、それを元に料理間の類似度を計量している。またレシピテキストの構造を解析するものとして、浜田[2]らの研究がある。彼らは調理の手順を考慮したレシピフローグラフの作成を行っている。

本方式では料理間類似度を、材料とあらかじめ定義した調理法 (3.3 節) を元に計量することで、より料理本来の特性に沿った類似度計量および検索をすることが可能となっている。

3. 材料とその調理法に基づく料理間類似度計量によるレシピデータ検索方式

図 1 に、材料とその調理法に基づく料理間類似度計量によるレシピデータ検索方式の全体像を示す。

本方式は、レシピデータスクレイピング機能、分かち書き機能、材料・調理法ペアデータ抽出機能、類似度計量機能、結果ソート機能からなる。事前にレシピデータをレシписクレイピング機能によりテキストデータとして取得し、分

Recipe data search method by measuring similarity between dishes based on ingredients and their cooking methods

Arima Naotaka[†] Okada Ryoutaro[†] Minematsu Ayako[†] Nakanishi Takafumi[†]

[†]Musashino University, Department of Data Science

から書き機能，材料・調理法ペアデータ抽出機能により，取得したレシピデータごとに材料・調理法からなる単語のペア（以下，材料・調理法ペア）を抽出し，これらをレシピデータ・材料・調理法ペアデータベースに格納する．一方，ユーザはレシピが掲載されている URL を入力として与え，前記のデータベースに格納されていない場合は，スクレイピングによりレシピテキストデータを取得し，分かち書き機能，材料・調理法ペアデータ抽出機能により，そのレシピの材料・調理法ペアを抽出した上で，データベースに格納されているレシピの材料・調理法ペアデータと類似度計量を行うことで，類似度の大きい料理を推定し，結果ソート機能により検索結果をユーザに返す．

3.1. レシピデータスクレイピング機能

レシピサイトの情報を，スクレイピングを用いて抽出する．一般的なレシピサイトでは，材料と調理手順が示されたレシピテキストの二つが存在するため，それぞれを分割して取得する．

3.2. 分かち書き機能

3.1 節で取得したテキスト情報を，形態素解析器 MeCab を用いて単語に分割する．本方式では，材料と調理法に注目しているため，抽出する品詞を名詞と動詞に絞っている．

3.3. 材料・調理法ペアデータ抽出機能

レシピの材料欄から抽出した材料名と，あらかじめ定義した調理法を，3.2 節で抽出した単語データの順序に則ってペアデータとして抽出する．調理法として定義する言葉が細分化されすぎると類似度計量が困難となるため，本方式では，調理法を「焼く，茹でる，炒める，揚げる，煮る，和える，蒸す，切る，その他」の9種類として定義した．その上で，「みじん切り」や「薄切り」は「切る」に含まれるといったように，細分化された調理法名に対して一般的な調理法名との対応関係を定義し，調理法名の抽出を行う．

3.4. 類似度計量機能

3.3 節において抽出したペアデータから料理毎の特徴ベクトルを作成し，コサイン類似度を用いて類似度の計量を行う．

3.5. 結果ソート機能

ユーザが入力したレシピと，データベースに格納されているレシピとの類似度を3.4節で示した手法により算出し，類似度の高い順にソートして出力する．

4. 実験

図1で示した方式を実現するシステムを構築し，インターネット上にあるクラシルのレシピサイ

表1：類似するレシピの材料・調理法ペア

	玉ねぎ 切る	じゃがいも 切る	にんじん 切る	ブロッコリ 切る	塩 和える
レシピ1	2	2	1	1	1
レシピ2	2	2	2	0	0
	赤ワイン 煮る	水 煮る	じゃがいも 和える	じゃがいも 煮る	玉ねぎ 煮る
レシピ1	1	1	1	2	0
レシピ2	0	1	0	1	1
コサイン類似度				0.79115480528524	

ト[3]から，ビーフシチューのレシピである55件のレシピテキストデータを収集して，それぞれのレシピで類似度計量を行った．

その結果，類似度の高かったレシピの組みの一例における特徴ベクトルの成分（レシピ内での材料と調理法の組みの出現頻度）についての検証結果を表1に示す．

表1より，「玉ねぎ，切る」，「じゃがいも，切る」，「にんじん，切る」，「水，煮る」，「じゃがいも，煮る」といった材料とその調理法のペアが両方に出現している．一方，この2つのレシピにおいて材料と調理法のペアの差分は「じゃがいも，和える」，「玉ねぎ，煮る」であることが分かる．本方式では，料理の味覚や食感を，材料とその調理法によって決定されるものとしているため，これらのペアが味覚や食感の差になる．「煮る」や「和える」は，食材を調味液によって加熱することにより食感を変化させたり，食材を調味料と共に混ぜ合わせるにより味覚を変化させたりするといった工程であるため，本方式を用いてその差を明確にすることが可能になったと考えられる．

5. おわりに

本稿では材料とその調理法に基づく料理間類似度計量によるレシピデータ検索方式について示した．また，本方式を実現する実験システムを構築した．

今後の課題として，レシピを書く際によく行われる調理工程の省略によって生じるペアデータ作成の不備を修正することで，より正確なペアデータの抽出の実現，大規模データによる検索性能の検証が挙げられる．

参考文献

[1] 福本 亜紀，井上 悦子，中川 優，食材の重要度と食品群を考慮したレシピ間類似度の算出方法，DEIM Form 2012 D9-2，2012．
 [2] 浜田 玲子，井手 一郎，坂井 修一，田中 英彦，料理テキスト教材における調理手順の構造化，電子情報通信学会論文誌，D, 85(1)，pp. 79-89，2002．
 [3] クラシル | 料理レシピ動画サービス <https://www.kurashiru.com>