

外国料理を対象とした原材料推定によるアレルギー検出

大坪 愛佳*

アルハビブ ファティマ*

延澤 志保*

*東京都市大学知識工学部

*東京都市大学情報工学部

1 研究背景

海外旅行の楽しみの1つとして食があげられる。本研究では、外国料理を安心して食べられるよう、外国料理の食材に含まれるアレルギーを検出する手法を提案する。

2 食材とアレルギー

2.1 レシピ材料欄の記述

レシピの材料欄に記述される食材は3種類に分けられる(表1)。複合原材料とは、2種類以上の原材料からなる

表1: 材料欄の記述粒度

原材料	(卵, 牛肉, 桃)
複合原材料	加工品など (マヨネーズ)
原材料総称	複数の原材料を含む (肉, フルーツ)

ものである¹。例えば、マヨネーズは複合原材料であり、その原材料は油や卵である。複合原材料は、アレルギーを発見するために原材料を調べる必要がある。また、レシピでは「肉」や「フルーツ」のように原材料の総称で記載されている場合もあり、この場合もアレルギーの有無を推定する必要がある。

2.2 対象アレルギー

レシピの材料欄では、原材料として大まかな分類の「ゴマ」が記載されている場合と細かい種類の「黒ゴマ」が記載されている場合が混在する。同じ「ゴマ」と記載される食材でも、黒ゴマはアレルギーだが、トウゴマはアレルギー対象外である。食材の抽出については、レシピの希少性と一般性を測った意外食材の抽出手法[1]や、健康効果を考慮した代替食材提案のための食材抽出手法[2]などが提案されているが、ゴマの例のようにアレルギー検出ではより正確な食材抽出が必要である。

表2にアレルギー28品目²(2列目)の詳細な範囲³をまとめる。対象アレルギー名、学名で判断するのは、表2に示すとおりアレルギーの対象範囲があるためである。

3 料理に含まれるアレルギーの自動検出

3.1 提案手法

本研究は対象料理のアレルギー(表2)の含有の有無を自動推定することを目的とする。提案手法では、外国語

Allergen Detection for Foreign Cuisine Based on Ingredient Estimation.

Manaka Ootsubo*, Fatima Abdulhakim A Al Habib*, and Shihō Hoshi Nobesawa*.

* Faculty of Knowledge Engineering, Tokyo City University

* Faculty of Information Technology, Tokyo City University

¹ 消費者庁, 加工食品における複合原材料の表示方法について, https://www.cao.go.jp/consumer/history/03/kabusoshiki/syokuhinhyouji/doc/k140605_shiryou3.pdf.

² 食品衛生の窓, 東京都福祉保健局, https://www.fukushihoken.metro.tokyo.lg.jp/shokuhin/hyouji/shokuhyouhou_kakou_allergy.html.

³ 食物アレルギー表示, 消費者庁, https://www.caa.go.jp/policies/policy/food_labeling/food_sanitation/allergy/assets/food_labeling_cms204_210514_01.pdf.

表2: アレルギーとその記述粒度

総称	28品目	対象に含まれる食材	対象外食材
卵	卵	鶏卵, アヒル卵, うずら卵	魚卵, 爬虫類卵, 昆虫卵
肉	牛肉	牛脂	内臓, 骨, 皮
	豚肉	ラード	内臓, 骨, 皮
	鶏肉		内臓, 骨, 皮
ゼラチン	ゼラチン		
乳	乳	牛乳	水牛乳
魚介類	さけ	サケ科サケ属, サケ科サルモ属 (ベニマス)	サケ科イワナ属, サケ科タイヘイヨウサケ属 (ニジマス, イワナ, ヤマメ)
	いくら	いくら, すじこ	
	さば		
	いか		
	エビ	十脚目 (イセエビ, ロブスター)	口脚目, アミ目, オキアミ目 (シャコ, アミ)
	カニ		
	あわび	ミミガイアワビ属 (クロアワビ)	チリアワビ, トコブシ
小麦	小麦	小麦, デュラム小麦	大麦, ライ麦
大豆	大豆	枝豆, もやし	緑豆, 小豆
そば	そば		
ごま	ごま	ゴマ科ゴマ属 (黒ゴマ, 白ゴマ)	トウゴマ属 (エゴマ, トウゴマ)
木の実	落花生	ピーナッツ, 南京豆	
	アーモンド		
	カシューナッツ		
	クルミ		
きのこ	まつたけ		
芋	やまいも		
果物	りんご		
	桃		
	バナナ		
	オレンジ		夏みかん, はっさく, グレープフルーツ, レモン
	キウイフルーツ		

で入力した料理名に対してアレルギー検出結果を含有レベルと共に出力する。日本語のレシピでは食材を日本人向けに変更している場合があるため、提案手法では外国語レシピを日本語に翻訳して食材を抽出し、各食材について原材料推定を行ってアレルギーを検出する。

料理は店舗ごとにレシピが異なり、外食時に実際のレシピを確認することは困難なため、提案手法では複数のレシピを基にアレルギー含有推定を行う[3, 4]。一般的なレシピには含まれないアレルギーを明示するため、含有するレシピの割合に応じてアレルギーの有無を4段階の含有レベル[3, 4]に分類する(表3)。表3は同じ料理の

表3: アレルギー含有レベル

含有確実	当該料理が必ず含有する食材 (出現レシピ率 $\theta = 100\%$)
含有濃厚	当該料理が含有する可能性が高い食材 ($60\% \leq \theta < 100\%$)
含有可能	当該料理が含有する可能性を持つ食材 ($0\% < \theta < 60\%$)
含有僅少	当該料理が含有する可能性が十分に低い食材 ($\theta = 0\%$)

複数のレシピについて各食材が出現するレシピの割合を示すものであり、本稿では表3に示す出現レシピ率 θ で含有レベルを判定する。

3.2 食材のアレルギー検出

提案手法では、入力された料理名に対して複数の外国語レシピの材料欄を翻訳して食材の初期リストを作成し、

これに含まれる食材全てについてアレルゲン検出を行う。アレルゲン検出に当たっては、それぞれの食材について Wikipedia などの食材記事から原材料の記載のある文を取得する [5, 6]。図 1 に原材料記載文からのアレルゲン検出の処理の流れを示す。対象の食材記事に原材料記載

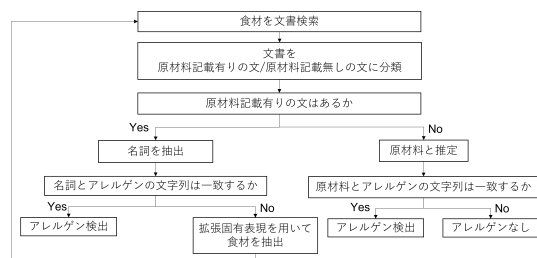


図 1: アレルゲン検出の処理の流れ

文がある場合には、アレルゲンリストの文字列と一致する名詞があればアレルゲン検出とする。一致するものがない場合は、さらに複合原材料が含まれている可能性があるため、拡張固有表現 [7] を用いて原材料記載文中の名詞から食材を抽出し、それぞれについて食材記事の検索と原材料記載文の取得を繰り返す。対象の食材記事に原材料記載文がない場合、この食材は原材料であると推定し、アレルゲンリストの文字列と一致すればアレルゲン検出、一致しなければアレルゲンなしとする。

原材料総称の場合は、複数レシピの中での鶏肉や豚肉の出現割合を表 3 の含有レベルとみなす。

4 アレルゲン検出実験

アラビア料理 (justfood¹) とアメリカ料理 (allrecipes²)、比較のため日本料理 (楽天レシピ [8]) を対象とし、提案手法でアレルゲン検出を行った。食材記事としては、多言語インターネット事典でさまざまな国の食材の情報が記載されている Wikipedia³ を使用した。

表 4 に正解のアレルゲンの数と検出されたアレルゲンの数、食材のアレルゲン検出の精度を示す。アラビア語

表 4: 食材のアレルゲン検出結果

料理言語	料理名	レシピ数	アレルゲン数 正解	アレルゲン数 検出	アレルゲン 検出精度
アラビア語	コシャリ	18	4	4	100%(5/5)
アラビア語	オマール	7	6	6	100%(14/14)
アラビア語	カブサ	59	9	9	89%(17/19)
英語	ジャンバラヤ	30	9	9	94%(15/16)
英語	ガンボ	30	9	8	83%(10/12)
英語	コブラー	30	9	9	100%(17/17)
日本語	親子丼	46	6	6	100%(11/11)
日本語	たこ焼き	46	8	7	81%(17/21)
日本語	フルーツゼリー	46	8	8	100%(16/16)

レシピを対象とした実験では 3 種類の料理の全てでアレルゲンの検出に成功した。英語レシピを対象とした実験では、ガンボに含まれる「ブイヨン」のアレルゲン検出に失敗した。Wikipedia のブイヨンの記事の原材料記載文は「牛などの動物質」であり、ここから拡張固有表現で「牛」が抽出されたが、これだけでは牛肉と判断できな

かった。日本語レシピを対象とした実験では、たこ焼きに含まれる豚肉の検出に失敗した。これは、ウインナーとソーセージの原材料記載文の記述が「刻み肉」で、複数レシピ中にも具体的な肉種の記述がなく、豚肉の含有レベル推定に失敗したためである。同じ食材総称でも、「フルーツ」については桃などのアレルゲンの含有可能性を検出することに成功した。

原材料は出現した 16 種類全て検出することができた。肉、ひき肉の 2 種類は食材の総称として出現したが、ソーセージ同様複数レシピ中に具体的な肉種の記述がなく検出に失敗した。複合原材料 56 種類のうちアレルゲン検出ができなかったものは、ウインナー、ソーセージ、たこ焼き粉、お好み焼き粉、ブイヨンの 5 種類であり、そのうちたこ焼き粉とお好み焼き粉は Wikipedia 記事がなかった。その反面、カルダモンなど外国で一般的な食材は Wikipedia 記事がある例が多かった。

5 まとめ

アレルギーを持つ人は、アレルゲンを摂取してしまう危険性があるため、海外では料理を気軽に楽しむことができない。そこで本研究では、食材に含まれるアレルゲンを検出する手法を提案した。レシピの材料欄に現れる食材には加工食品を含むため、材料欄の食材全てに対して原材料の推定を行う。加工食品については再帰的に文書検索を行い、原材料と推定された食材についてアレルゲン検出を行った。3 言語のレシピで実験し、料理のアレルゲン検出の成功を確認した。

謝辞

本研究では、国立情報学研究所 IDR データセット提供サービスを通して楽天グループ株式会社様よりご提供いただいた「楽天データセット (https://rit.rakuten.com/data_release/)」を利用しました。心より感謝いたします。

参考文献

- [1] 池尻 恭介, 清 雄一, 中川 博之, 田原 康之, 大須賀 昭彦, “希少性と一般性に基づいた意外性のある食材の抽出,” コンピュータソフトウェア, Vol.31, No.3, pp.70–78, 2014.
- [2] 花井 俊介, 難波 英嗣, 瀧本 明代, “健康を意識した代替食材の発見手法,” 第 7 回データ工学と情報マネジメントに関するフォーラム, pp.1–8, 2015.
- [3] 大坪 愛佳, アルハビブ ファティマ, 大和矢 悠仁, 延澤 志保, “アラビア語レシピ中のアレルゲン認識に関する一考察,” 第 20 回情報科学技術フォーラム, no.E-005, pp.217–218, 2021.
- [4] 大坪 愛佳, アルハビブ ファティマ, 大和矢 悠仁, 延澤 志保, “外国語レシピからのアレルゲン自動検出,” NLP 若手の会第 16 回シンポジウム, no.P4-12, 2021.
- [5] アルハビブ ファティマ, 大坪 愛佳, 延澤 志保, “Wikipedia の食材記事からの原材料自動抽出,” NII-IDR ユーザフォーラム, no.P12, 2021.
- [6] アルハビブ ファティマ, 大坪 愛佳, 延澤 志保, “外国料理のハラル推定のための食材記事からの原材料記載文抽出,” 情報処理学会第 84 回全国大会, 2022.
- [7] 関根 聡, 小林 暁雄, 安藤 まや, 乾 健太郎, “拡張固有表現に基づく Wikipedia 項目の分類と構造化,” SIG-SWO-043-05, P1-6, 2018.
- [8] 楽天グループ株式会社, 楽天レシピデータ, 国立情報学研究所情報学研究データリポジトリ, データセット, <https://doi.org/10.32130/idr.2.4>, 2021.

¹justfood, <https://www.justfood.tv/>.

²allrecipes, <https://www.allrecipes.com/>.

³Wikipedia, <https://ja.wikipedia.org/wiki/>.