

健康を意識した代替食材の発見手法

花井 俊介[†] 難波 英嗣^{††} 灘本 明代^{†††}

[†] 甲南大学大学院 自然科学研究科 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

^{††} 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

^{†††} 甲南大学 知能情報学部 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†]m1424007@center.konan-u.ac.jp, ^{††}nanba@hiroshima-cu.ac.jp, ^{†††}nadamoto@konan-u.ac.jp

あまし 近年、生活習慣病の増加や食育の推進により、食に対する健康意識が高くなっている。また、レシピサイトやブログ型レシピサイトなどの普及により、健康を目的としたレシピの取得が容易となってきている。しかし、これらのレシピは日常的に作る料理ではないことが多く、日々の食事で習慣的に健康を意識した食事をとることは困難である。本研究では、日常的に作る料理の食材の一部を代替食材を用いることで、健康を意識した料理になるようアレンジを加えることを目的とする。典型的な食材、食材に含まれる栄養成分、栄養成分による効能、過剰摂取により悪影響を及ぼす成分、食材の食感の抽出を行い、特定の食材に対して他の食材との類似度を算出することで、健康目的別に代替可能な食材を発見する手法を提案する。

キーワード 代替食品, 健康, 情報抽出

1. はじめに

メタボ検診（特定健康診査・特定保健指導）や食育への取り組みにより、食に対する健康志向が高くなっている [1]。そこで今後、健康を意識した食の情報発信がより必要になると考えられる。また、Cookpad^(注1)や楽天レシピ^(注2)に代表されるユーザ投稿型のレシピサイトやレシピブログ^(注3)に代表されるブログ型レシピポータルサイトなどの普及に伴い、「風邪予防」や「便秘改善」のような健康を意識したレシピの取得が容易となってきている。しかし、このような健康効果を示す単語をクエリとして検索を行った場合、スムージーやポタージュなど日常的に作る料理ではないことが多い。日々の食事で健康効果を得ることが出来れば手間が少なく、ユーザの負担も少ないと考えられる。しかしながら、食事毎に健康効果のあるレシピを検索し、レシピを見ながら調理することは非常に手間のかかる作業である。例えば、対象となる作る料理が「カレー」や「ハンバーグ」のように普段から作っている料理の場合、作り方を知っているためレシピを見ながら調理することは面倒である。いつも作るレシピの食材を代替しただけで健康効果が向上すると便利であると考えられる。

そこで本研究では、料理名と「風邪予防」や「便秘改善」のような健康効果をクエリとしたとき、目的的健康効果を得ることができる代替食材を提示する検索システムの開発を目指す。例えば、クエリをカレーと風邪予防としたときにカレーに一般的に使われるじゃがいもを用いるよりかぼちゃを用いた方が風邪予防に良い。そのため、かぼちゃを代替食材として提示することで風邪予防の効果の向上を図る。ここで、代替食材を用いた

結果、料理に違和感が出ないよう元の食材に類似した食材である必要がある。そこで、元の食材と代替食材との食感の類似の考慮も行う。本論文ではこのシステムを構築するために、ユーザ投稿型レシピサイト、ブログ型レシピポータルサイト、特許データから典型的な食材、食材に含まれる栄養成分、栄養成分による効能、過剰摂取により悪影響を及ぼす成分、食材の食感の抽出を行い、他の食材との類似度計算を行うことで、「風邪予防」や「便秘改善」のような健康目的別に効果を見込める代替可能な食材を発見する手法を提案する。

以降、2章では関連研究について述べる。3章では、提案手法と抽出結果について述べる。4章では、まとめと今後の課題について述べる。

2. 関連研究

貝沼ら [2] の調査によると、料理に携わる人の約半数が「献立を考えることが面倒である」と感じている。これは、日々の食事の栄養バランスや食材の有無、同居者の好みなどを考慮して献立を考えることは非常に労力を要する作業であると考えられるためである。そのため、レシピ検索やレシピ推薦に関する研究は盛んに行われている。

類似したレシピの推薦に関する研究として以下のような研究がある。福本ら [3] は、食材の分量を食材の重要度、食品群ごとの平均分量によって重み付けし、それぞれのコサイン類似度を求め、コサイン類似度を合算した値をレシピ間類似度とすることで類似レシピの抽出を行っている。苅米ら [4] は、材料、手順に着目し、コサイン距離、DP マッチングを用いることでレシピ間の類似度を算出している。また、村瀬ら [5] は、食材の出現頻度と手順に着目し、食材ごとの共起数から料理を構成要素に分割し、構成要素内でさらに分割を繰り返すことで、料理における材料のカテゴリ構造を算出し類似レシピの抽出を行っている。さらに、Wang ら [6] は、料理の工程をグラフ化する

(注1): <http://cookpad.com/>

(注2): <http://recipe.rakuten.co.jp/>

(注3): <http://www.recipe-blog.jp/>

ことにより、酷似した部分グラフがあることを示している。Liら [7] はユーザの好みにあったレシピを抽出するため、料理レシピをグラフ化し、類似度を計算することでユーザごとの好みの食材、調理法が用いられているレシピを抽出する手法を示している。しかし、これらの研究は代替食材を用いてレシピの変更を行うのではなく、元のレシピと類似した他のレシピを推薦するという点で本研究と異なる。

代替食材に関する先行研究も数多く存在している。志土地ら [8] は、同一料理カテゴリ中の料理レシピにおける特徴的な調理方法と食材との共起関係に基づいて、食材の類似度を算出することによって、代替可能な素材を発見する手法を提案している。橘ら [9] は材料と調理器具に着目し、典型的な要素との差異を抽出し、確信度を用いて代替関係を判定することでレシピの差異要素を分類し、特徴パターンを用いてレシピを分類することで、ネーミングコンセプトの抽出を行っている。野沢ら [10] は、調理手順中に出現する食材、調理法を特徴とし word2vec を用いることで、特定の食材に対して他の食材との類似度を算出することで、代替可能な食材の発見を行っている。しかし、いずれの研究においても類似した調理方法を用いて代替食材の発見を行っており、食感の類似などは考慮していない為、本研究とは異なる。

食の健康に着目したレシピ推薦として以下のような研究がある。野田ら [11] は Web から健康効果と食材に関する知識を抽出することで検索語拡張を行い、ブログに対して検索を行うことで健康効果のあるレシピが記載されているブログ記事の提示を行っている。しかし、これは健康効果のあるレシピを抽出するものであり、既存のレシピに健康効果を付与するものではない。また、苅米ら [12] は健康的な食生活の支援を目的として、人々の日常の食事記録の可視化を行い、栄養をバランス良く摂取できるレシピ推薦を行うシステムを提案している。しかしこれは、栄養をバランス良く摂取するための支援システムであり、「便秘改善」のような特定の健康目的のためのシステムではなく、本研究とは異なる。Youri ら [13] は、レシピから重要となる特徴を抽出し、特徴ベクトルを生成することによって類似度を算出し、健康的なレシピに変更できるようにしている。しかし、代替レシピを推薦して健康効果を得ている点で本研究とは異なる。また、健康効果については言及されておらず本研究と異なる。

3. 提案手法と抽出結果

本研究では健康を意識した代替食材を用いたレシピ検索システム構築の為に、以下のデータの抽出を行い、類似度計算を行うことで健康を意識した代替食材の発見を行う。

- 典型的な食材
- 食材に含まれる栄養成分
- 栄養成分による効能
- 過剰摂取により悪影響を及ぼす成分
- 類似した食感を持つ食材

3.1 提案システム

提案システムの流れを図 1 に示す。クエリとして料理名と健康

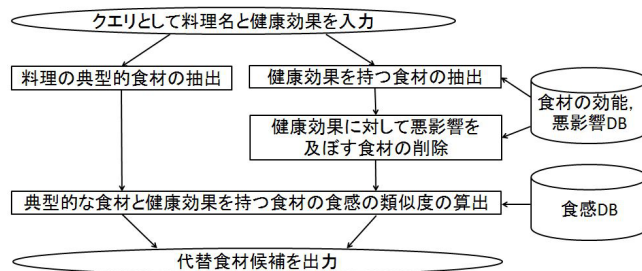


図 1 システムフロー

康効果を入力する。次にクエリで入力された料理に用いられる典型的な食材の抽出を行う。典型的ではない食材から代替を行った場合、その食材を用いない料理が多数存在する。その場合、代替前の食材がなく、代替できないためである。例えば、カレーの典型的な食材として、牛肉、玉ねぎ、人参、じゃがいもなどが考えられる。提示する代替食材として牛肉を鶏肉に変更すると提示された場合、カレーに牛肉を用いていることが多いため、問題なく代替できる。しかし、ちくわのようにその料理において典型的でない食材から代替すると提示された場合、カレーにちくわを用いていないことが多いため、代替前の食材が見つからず、代替を行うことができない。このように代替前の食材が見つからないことを防ぐため、代替前の食材の候補として典型的な食材を用いる。次にクエリとして入力された健康効果を向上させるためにその健康効果を持つ食材の抽出を行う。また、代替する食材がクエリの健康効果に対して悪影響を及ぼす食材であった場合、健康効果が低下するため代替食材候補から除外する。次に代替前の典型的な食材と代替食材候補である健康効果を持つ食材との食感の類似度計算を行い、類似順に提示する。代替する食材は代替前の食材に類似しているほど代替後の料理の違和感が少ないためである。

3.2 典型的な食材の抽出手法

以下に料理ごとの典型的な食材の抽出手法を示す。

- (1) ユーザ投稿型レシピサイトに投稿されているレシピを、料理の種類ごとに分類し、食材（調味料を除く）を抽出する。
- (2) 分類した料理の種類ごとの食材の使用頻度と 1 レシピ当たりの平均食材使用数を算出する。
- (3) 使用頻度の上位から平均使用数までの食材を取得し、典型的な食材とする。

調味料を代替した場合、その料理の味に大きく変化を及ぼす可能性があるため、調味料を除いた食材を用いる。以下に抽出した典型的な食材について述べる。

データセット

典型的な食材の抽出のためのデータには楽天レシピに投稿された料理レシピ約 44 万件を用いた。また、料理ごとの分類のための料理名には基本のレシピ 100 [14] に掲載されている料理名（デザートと飲み物を除く）を用いた。

抽出結果

抽出した料理の種類ごとの典型的な食材を表 1 に示す。表 1 より、どの料理においてもその料理において典型的な食材として

表 1 料理ごとの典型的な食材

料理名	典型的な食材	料理名	典型的な食材
カレー	玉ねぎ	青椒肉絲	ピーマン
	人参		牛肉
	じゃがいも		豚肉
	にんにく		たけのこ
	牛肉		にんにく
料理名	典型的な食材	料理名	典型的な食材
牛丼	玉ねぎ	ポテト	じゃがいも
	牛肉		きゅうり
	ご飯	サラダ	ハム
	卵		人参
	-		玉ねぎ

適切と考えられる典型的な食材が抽出されている。料理の種類ごとにレシピを分類し、食材の使用頻度を求めたため、典型的な食材が適切に抽出されたと考えられる。青椒肉絲は通常、豚肉が用いられるが「牛肉」と「豚肉」の2種類の肉が抽出されている。牛肉を用いた場合は青椒牛肉絲と呼ばれるが、日本では牛肉を用いた場合においてもこれらを区別せず、青椒肉絲と呼ぶことが多いため、「牛肉」と「豚肉」が共に抽出されたと考えられる。

3.3 食材に含まれる栄養成分、栄養成分による効能の抽出手法

近年、ブログの書き方の特徴として、段落を分ける際に改行後に1文字下げるのではなく、複数の改行を用いることで段落を分けることが多い。また、少ない文章量で細かく段落を分けることが多い。ブログ記事例を図2に示す。本研究ではこの段落を用い、同一段落中での食材名と栄養成分名との共起、栄養成分名と効能名との共起に着目する。以下に食材に含まれる栄養成分、栄養成分による効能の取得手法を示す。

(1) 対象ブログ記事を段落ごとに分割する。

(2) 段落ごとに食材名と共起している栄養成分名、栄養成分名と共起している効能名を正規表現によるマッチングによって取得する。

(3) ある閾値以上の共起頻度を持つものをそれぞれ、食材に含まれる栄養成分、栄養成分による効能とする。今回、実験により食材に含まれる栄養成分の閾値 $\alpha=5$ 、栄養成分による効能の閾値 $\beta=50$ とする。

(4) (3) で取得した食材に含まれる栄養成分、栄養成分による効能を用いて食材ごとの効能を決定する。

以下に抽出した食材に含まれる栄養成分、栄養成分による効能について述べる。

データセット

食材に含まれる栄養成分、栄養成分による効能の抽出のためのデータにはレシピブログを用いた。2013年10月1日から2014年9月31日までの1年間に投稿されたアメーバブログ^(注4)の記事65,192件を用いた。また、食材名には3.2節で述べたように調味料を代替するとその料理の味に大きく変化を及ぼす可

さつまいもと言えば、便秘の解消！
そんなイメージが強いですね^^
このさつまいもと同じくらいか、それ以上にお腹
スッキリの効果があるものがゴボウ！
今日はこの両方を使って、お腹の中からキレイ
にしちゃう一品を作ります♪

さつまいもにはセルロースやペクチンなどの食物
繊維が豊富に含まれています。この食物繊維が
整腸作用を促し、便秘予防に効果があります。

また食物繊維には、コレステロールを下げたり、
血糖値をコントロールする働きもあり、大腸がんや
動脈硬化、高血圧、糖尿病などの予防に効果が
あります。

そして便秘の予防には、食物繊維の働きと、もう
ひとつヤラビンの働きが関係しています。ヤラビ
ンはさつまいもを切った時にでる白い汁で、これは
さつまいもにしか含まれていないと言われています。

図 2 ブログ記事例

表 2 食材に含まれる栄養成分

食材名	含有成分	食材名	含有成分
レバー	鉄	豆腐	タンパク質
	カロテン		カルシウム
	タンパク質		食物繊維
	カルシウム		鉄
	ビタミン A		コレステロール
食材名	含有成分	食材名	含有成分
人参	カロテン	トマト	ビタミン C
	食物繊維		鉄
	鉄		食物繊維
	カリウム		カロテン
	ビタミン B1		カルシウム

能性があるためである。また、調味料は液体や粉末状であることが多く、食感を感じないためである。栄養成分名には食品成分データベース^(注5)に記載されている成分名を用い、効能名はレシピブログから人手により収集した。

抽出結果

食材に含まれる栄養成分の抽出結果の一部を表2に示す。また、栄養成分による効能の抽出結果を表3に示す。表2より、レバーでは「鉄」や「タンパク質」、豆腐では「タンパク質」や「カルシウム」などその食材に多く含まれている成分が抽出されていることがわかる。また、表3より、食物繊維では「便秘改善」や「美肌」、鉄では「貧血」や「骨粗鬆症」などその栄養成分による主な効能が抽出されていることがわかる。栄養成分について書かれたブログ記事は、食材の効能について書かれたブログ記事であることが多いため、含有量の少ない栄養成分やあまり効果の見込めない効果について記載されることは少ない。そのため、その食材に豊富に含まれる栄養成分や栄養成分の主な効能が取得できたと考えられる。また、ブログ記事のよ

(注4): <http://ameblo.jp/>

(注5): <http://fooddb.mext.go.jp/>

表 3 栄養成分による効能

栄養成分名	効能	栄養成分名	効能
食物繊維	便秘改善	鉄	貧血
	美肌		骨粗鬆症
	美容		疲労
	ダイエット		抗酸化
	ガン		便秘
栄養成分名	効能	栄養成分名	効能
ビタミン B1	疲労	ビタミン C	美肌
	夏バテ		風邪予防
	風邪予防		疲労
	免疫力		免疫力
	動脈硬化		ガン

表 4 過剰摂取ワード

食感系シズルワード
過剰摂取, 過剰な摂取, 過剰に摂取, 大量摂取, 大量に摂取, 多量摂取, 多量に摂取, 摂り過ぎ

うに改行を用いて段落を分ける場合、段落を少ない文章量で細かく分けることが多いため、同一段落中に食材名と栄養成分名が同時に出現した場合、その食材に含まれる栄養成分である可能性が高いと考えられる。そのため、誤った栄養成分が抽出されることが少なかったと考えられる。栄養成分による効能についても同様であったと考えられる。

3.4 過剰摂取することで悪影響を及ぼす成分の抽出

食材に含まれる栄養成分の中には過剰摂取することによって、悪影響を及ぼす成分が存在する。クエリとして入力された健康効果に対して悪影響を及ぼす成分を持つ食材を優先して代替することが出来ればより健康効果を見込めると考え、過剰摂取することで悪影響を及ぼす成分の抽出を行う。悪影響を及ぼす成分はレシピサイトやブログサイトに記載されていることが少なく、記載されていた場合においてもどの成分が悪影響を及ぼすか、過剰摂取によってどのような影響（症状）があるか記載されていないことが多い。そこで本研究では食に関する特許データに着目し、特許データから悪影響を及ぼす成分の抽出を行った。悪影響を及ぼす成分の抽出は過剰摂取に関する語（以下、過剰摂取ワードと呼ぶ）を用いて正規表現のマッチングを行い、過剰摂取ワードを含む文を取得し、その後人手にて原因となる成分、その影響（症状）を取得した。過剰摂取ワードを表 4 に示す。以下に抽出した過剰摂取することで悪影響を及ぼす成分について述べる。

データセット

過剰摂取することで悪影響を及ぼす成分の抽出には特許データを用いた。特許データは、国際特許分類 (IPC) のサブクラスレベルで A23L(食品, 食料品)、A47J(台所用具)、H05B(電気加熱) が付与された料理分野の特許明細書 (1994~2009 年公開特許公報) の 91,736 件を用いた。

抽出結果

抽出結果の一部を表 5 に示す。3.3 節の栄養成分による効能と比較すると抽出された過剰摂取による悪影響は「高脂血症」や

表 5 過剰摂取することで悪影響を及ぼす成分

栄養成分名	悪影響	栄養成分名	悪影響
脂質	肥満	ナトリウム	高血圧
	高血圧		心疾患
	高脂血症		脳梗塞
	糖尿病		腎臓疾患
	動脈硬化		動脈硬化

「動脈硬化」「腎臓疾患」のように具体的な病名であることが多い。これは、使用したデータによるものであると考えられる。栄養成分による効能を抽出したブログでは多くの人が悩みを抱えている症状であり、一般的に通院して治療するような症状でないことに対する効能が多く見受けられる。これに対して悪影響を及ぼす栄養成分を抽出した特許データは、技術的に解決すべき課題として過剰摂取すると悪影響がある成分について記述されているため具体的な病名が多く抽出されたと考えられる。

3.5 食感の抽出手法

食材の代替を行う際、料理に違和感が出ないように元の食材に類似した食材である必要がある。そこで本研究では食感の類似に着目し、食材ごとの食感の抽出を行う。食感の抽出にはユーザ投稿型レシピサイト、Google による「Web 日本語 N グラム 第 1 版」[15] を用い「とろとろ」や「ぷりぷり」といった“おいしさ”を連想させる食に関する言葉であるシズルワード [16] を用いて食材ごとの食感を抽出し、類似度計算を行う。大橋ら [16] はシズルワードを味覚系、食感系、情報系という 3 つの領域に分類している。本研究では、食感の抽出を行うにあたって食感系に分類されるシズルワード（以下、食感系シズルワードと呼ぶ）を用いる。

3.5.1 ユーザ投稿型レシピサイトを用いた食感の抽出手法

食感の抽出としてユーザ投稿型レシピサイトである Cookpad を用いる。これはユーザ投稿型レシピサイトは料理レシピに特化したサービスであり、食品以外の話題、適切でない表現が少ないため食品とシズルワードとの共起が良好である [17] ためである。また、食感系に分類されるシズルワードは 240 語存在するが、「いがいが」や「がががん」のようにあまり馴染みのない語も含まれる。そこで本論文では、Cookpad 上で食感系シズルワードをクエリとし検索を行い、投稿数が 1,000 件を超える食感系シズルワードを用いる。その食感系シズルワードは 71 語となった。その食感系シズルワードを表 6 に示す。食材名と食感系シズルワードを用いて食材の食感抽出を行う。その手順を以下に示す。

(1) 食材名をクエリとし Cookpad 上で検索を行い、最大で新着 1,000 件のレシピを取得する。クエリとして用いた食材名の一部を表 7 に示す。

(2) 各レシピタイトルに食材名、食感系シズルワードが含まれているか正規表現によるマッチングを行い、そのレシピ数を取得する。

(3) 各食材名が含まれるタイトル件数、各食感系シズルワードが含まれるタイトル件数を用い、Dice 係数を算出する。

表 6 投稿数が 1,000 件を超える食感系シズルワード

食感系シズルワード
あったかい, 熱い, あつあつ, 溢れる, かすかす, かりかり, 軽い, くずれる, ごく, こしこし, こりこり, さくさく, さっさく, さくふわ, さっくり, さらさら, しっとり, しゃきしゃき, ジューシー, しゅわしゅわ, 新食感, すーすー, つぶつぶ, つるつる, つるん, とろーり, とろける, とろっと, とろとろ, とろり, なめらかな, ねっとり, ねばねば, 歯応えのある, ぱさぱさ, はじける, はむはむ, ぱらぱら, ぱりぱり, ひえひえ, ぴりぴり, ひんやり, ふかふか, ぷちぷち, ふっくら, ふにふに, ふりふり, ぶりぶり, ぷりんぷりん, ふるふる, ぶるぶる, ふわさく, ふわっと, ふわとろ, ふわふわ, ふんわり, ほかほか, ぽかぽか, ほくほく, ほっくり, ほっこり, ほろっと, ほろほろ, みるきー, もちっと, もちもち, もっちもち, もちふわ, もっちり, もりもり, やわらかい,

表 7 クエリとして用いた食材例

食材名
紫蘇, 青りんご, アサリ, アスパラガス, アジ, 小豆, アナゴ, アヒル, アボカド, アロエ, アワビ, アンコウ, 鮫鱈, スルメイカ, イナゴ, イノシシ, イワシ, ウズラ, 牛タン, 牛すじ, 牛肉, ウナギ, ウニ, ウメ, エシャロット, エゾシカ, エビ, ボタンエビ, オイカワ, 大麦, オレンジ, かたつむり, カツオ, カニ, カボチャ, カニカマ, カモ, カリフラワー, きくらげ, キャベツ, きゅうり, くじら, クランベリー, クリ, 小女子, ゴボウ, 小松菜, 玄米, 米粉, こんにゃく, しらたき, 昆布, さくらんぼ, ザクロ, サザエ, ささみ, サツマイモ, サトイモ, フカヒレ, 秋刀魚, しいたけ, ししとう, しじみ, ししゃも, しめじ, ジャガイモ, 春菊, 生姜, すじこ, すだち, スッポン, 大根,

ここで Dice 係数を用いたのは, 加藤ら [18] の研究より, Dice 係数が食品とシズルワードの抽出を行うのに最も適していたためである.

(4) (3) で求めた Dice 係数を用い, 代替したい食材とその他の食材との類似度を算出する.

Dice 係数は以下の式によって与えられる.

$$dice(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

このとき, X は食材名が含まれるタイトル件数, Y は各食感系シズルワードが含まれるタイトル件数である.

また今回, 類似度計算にはコサイン類似度, ユークリッド距離, マンハッタン距離を用い比較を行った. それぞれの式を以下に示す.

$$cosine(x, y) = \frac{\sum x_i, y_i}{\sqrt{\sum x_i^2, y_i^2}} \quad (2)$$

$$euclid(x, y) = \sqrt{\sum (x_i - y_i)^2} \quad (3)$$

$$manhattan(x, y) = \sum |x_i - y_i| \quad (4)$$

このとき, x は代替したい食材と食感系シズルワードとの Dice 係数, y は他の食材と食感系シズルワードとの Dice 係数, i は食感系シズルワードの番号である. 以下にユーザ投稿型レシピサイトを用いて抽出を行った食材の食感について述べる.

データセット

食感の抽出のためのレシピデータには Cookpad を用い, 食材名をクエリとし, 取得した新着最大 1,000 件のレシピタイトルを用いた. クエリとして用いた食材名は 3.3 節と同様に調味料名を除いた食材名を用いた.

抽出結果

レシピデータを用いて, 各食材名が含まれるタイトル件数, 各食感系シズルワードが含まれるタイトル件数から Dice 係数を求めた結果の上位 5 件を表 8 に示す. 表 8 より, えびでは「ぶりぶり」, レンコンでは「しゃきしゃき」や「もちもち」のよ

うにその食材の食感として適切と考えられる食感を取得できていることがわかる. しかし, レンコンの「ジューシー」のようなその食材として不適切と考えられる食感も上位に出現していることがわかる.

次にレシピタイトルに含まれる食材名と食感系シズルワードとの Dice 係数を用いて類似度計算を行った結果を表 10 に示す. 表 10 より, いずれの食材においても食感が類似している食材が少なくなっていることがわかる. これは, えびのように食感のイメージがしやすい食材においても「ふわふわ」や「さくさく」のようにその食材名ではなく料理名や他の食材名と共起した食感系シズルワードをその食材の食感として取得したため, 類似した食感を持つ食材を取得できなかったと考えられる. また, レンコンは具材として使用している場合には「しゃきしゃき」や「さくさく」という食感は適切であると考えられるがすりおろして使用した場合においては「もちもち」や「ふわふわ」という食感についてもレンコンの食感として適切であると考えられる. このように食材の調理方法によって食感が大きく変わる食材であったため, 類似した食感を持つ食材を取得できなかったと考えられる.

3.5.2 Web 日本語 N グラムを用いた食感の抽出手法

次に「Web 日本語 N グラム第 1 版」[15] を用いた食感の抽出について示す. N グラムは一般に公開されている日本語の Web ページの約 200 億文が抽出対象となっており, 出現頻度 20 回以上の 1 ~ 7 グラムを収録している. 本研究では食材名と食感系シズルワードを含む 7 グラムの出現頻度を用い, 類似度計算を行う. 類似度計算には 3.5.1 節と同様にコサイン類似度, ユークリッド距離, マンハッタン距離を用い比較を行った. 以下に食材名と食感系シズルワードを含む 7 グラムの出現頻度を用いた食材の食感の抽出について述べる.

データセット

N グラムを用いた食感の抽出には Google による「Web 日本語 N グラム第 1 版」[15] の出現頻度 20 回以上の食材名と食感系シズルワードを含む 7 グラムデータを用いた. クエリとして用い

表 8 食材ごとの Dice 係数の上位 5 件

食材名	食感系シズルワード	Dice 係数	食材名	食感系シズルワード	Dice 係数	食材名	食感系シズルワード	Dice 係数
えび	ぶりぶり	0.035	レンコン	しゃきしゃき	0.146	ピーマン	ジューシー	0.01
	ふわふわ	0.014		もちもち	0.025		とろり	0.006
	ふんわり	0.012		さくさく	0.022		しゃきしゃき	0.004
	さくさく	0.006		ジューシー	0.019		ふわふわ	0.004
	かりかり	0.006		ふわふわ	0.012		もりもり	0.002

表 9 7 グラムデータによる食材ごとの出現頻度の上位 5 件

食材名	食感系シズルワード	出現頻度	食材名	食感系シズルワード	出現頻度	食材名	食感系シズルワード	出現頻度
えび	ぶりぶり	9178	レンコン	ばりばり	82	ピーマン	しゃきっと	871
	ジューシー	2446		さくさく	76		しんなり	355
	ほくほく	1331		もちもち	52		ふっくら	150
	やわらかい	833		ねっとり	23		とろける	112
	ふんわり	870		糸を引く	20		とろっと	66

た食材名は 3.3 節と同様に調味料名を除いた食材名を用いた。抽出結果

食材名と食感系シズルワードを含む 7 グラムデータを用いた食材ごとの食感系シズルワードの出現頻度の上位 5 件を表 9 に示す。表 9 より、食材名と食感系シズルワードを含む 7 グラムを用いた場合においても、食材ごとに適切と考えられる食感が抽出できていることがわかる。しかし、ピーマンの「とろける」のような不適切と考えられる食感も抽出されていることがわかる。これは、Dice 係数を用いた食感の類似度と同様に食材名ではなく料理名や他の食材名に共起したシズルワードをその食材の食感として取得しており、また調理方法によって食感が大きく変化する食材が存在するため食感が類似していない食材も抽出されたと考えられる。

次に食材名と食感系シズルワードを含む 7 グラムによる出現頻度を用いて類似度計算を行った結果を表 11 に示す。表 11 より、コサイン類似度では、えびとアンコウ、しめじの「ぶりぶり」やピーマンとキャベツ、わかめの「しゃきっと」などのように元の食材と同様のシズルワードで表現されることが多い食材が抽出されていることがわかる。これは、食感系シズルワードごとの出現頻度の値に大きな差があったため、類似した食感を持つ食材が取得できたと考えられる。

3.6 健康を意識した代替食材の抽出

以下に抽出したデータを用いて抽出した健康を意識した代替食材について述べる。

データセット

料理名をカレー、健康効果をダイエット、また料理名を青椒肉絲、健康効果を便秘改善の 2 種類のペアをクエリとして健康を意識した代替食材の抽出を行った。また、類似した食感の抽出には食材名と食感系シズルワードを含む 7 グラムによる出現頻度を用い、類似度計算にはコサイン類似度を用いる。

抽出結果

健康を意識した代替食材の抽出結果を表 12 に示す。表 12 より、カレーではクエリの健康効果であるダイエットに対して過剰摂取すると悪影響を及ぼす食材として牛肉が抽出されている。また、青椒肉絲ではクエリの健康効果である便秘改善に対して、

表 10 Dice 係数を用いた食材ごとの類似した食感の食材上位 5 件

食材名	コサイン類似度	ユークリッド距離	マンハッタン距離
えび	ニラ	紫蘇	鱒
	きくらげ	鱈	紫蘇
	高野豆腐	小松菜	あさり
	紫蘇	高野豆腐	ちくわ
	オレンジ	いかなご	ししゃも
食材名	コサイン類似度	ユークリッド距離	マンハッタン距離
レンコン	ひじき	ひじき	ひじき
	もやし	ツナ	ツナ
	栗	小松菜	ブロッコリー
	もち米	紫蘇	紫蘇
	米粉	たけのこ	ピーマン
食材名	コサイン類似度	ユークリッド距離	マンハッタン距離
ピーマン	なす	しいたけ	ちくわ
	もも肉	なす	アンチョビ
	しいたけ	もやし	しいたけ
	スペアリブ	鯉	軍鶏
	豚ロース	鱒	ヤシガニ

代替食材候補として、食物繊維が多く含まれるゴボウやキャベツ、わかめなどが抽出されている。このことから、クエリの健康効果を持つと考えられる食材が適切に代替食材候補として抽出されていると考えられる。また、どちらの料理においても野菜類の代替食材候補は適切に抽出されているのに対して、肉類は不適切な代替食材候補が多くなっている。肉類は「ジューシー」や「あつあつ」といった曖昧なシズルワードが食感として取得されていたため、適切でない代替食材の候補が多かったと考えられる。また、どちらの料理においてもにんにくの代替食材候補として挙げられている食材は適切ではない。これは、にんにくはすりおろしや細かく切って使用することが多いため、食感の抽出が適切に行われなかったためであると考えられる。カレーでは健康効果をダイエットとしたため代替食材候補として他の肉類が抽出されず、牛肉の代替食材として不適切であると考えられるオレンジや、昆布などが代替食材の候補として抽出されたと考えられる。また、カレーの玉ねぎの代替食材候補

表 11 7 グラムデータを用いた食材ごとの類似した食感の食材上位 5 件

食材名	コサイン類似度	ユークリッド距離	マンハッタン距離
えび	カツオ	カツオ	カニ
	アンコウ	ホタテ	カツオ
	カニ	カニ	ささみ
	ボタンエビ	イカ	みかん
	しめじ	キムチ	牛タン
食材名	コサイン類似度	ユークリッド距離	マンハッタン距離
レンコン	きゅうり	タコ	タコ
	りんご	さんま	さんま
	桃	水菜	水菜
	ゴマ	平茸	鰯
	数の子	たるいも	レーズン
食材名	コサイン類似度	ユークリッド距離	マンハッタン距離
ピーマン	キャベツ	にんにく	にんにく
	わかめ	バジル	バジル
	にんにく	椎茸	わかめ
	バジル	わかめ	椎茸
	ネギ	かぶ	瓜

表 12 健康を意識した代替食材の抽出結果

料理名: カレー 健康効果: ダイエット

典型食材	代替食材候補
玉ねぎ	白菜, ゴボウ, なす, 人参, ニラ
人参	玉ねぎ, 白菜, アスパラガス, ゴボウ, なす
じゃがいも	さつまいも, 椎茸, 人参, なす, わかめ
にんにく	わかめ, バジル, 椎茸, キムチ, さつまいも
牛肉	マンゴー, オレンジ, 昆布, 梅, 白菜

料理名: 青椒肉絲 健康効果: 便秘改善

典型食材	代替食材候補
ピーマン	キャベツ, ゴボウ, わかめ, にんにく, バジル
牛肉	ベーコン, マンゴー, アボカド, パナナ, チーズ
豚肉	鶏肉, オレンジ, ベーコン, 鰯, キャベツ
たけのこ	ししとう, 栗, 豆腐, パン, ブロッコリー
にんにく	わかめ, バジル, キャベツ, レモン, しいたけ

: 健康効果に対して悪影響を及ぼす食材

として人参, 人参の代替食材候補として玉ねぎが挙げられているがこれらの食材は互いに代替前の典型的な食材であるため代替食材候補としては不適切である。これらの典型的な食材は代替食材候補から除外しておく必要がある。

4. まとめと今後の課題

本研究では, 日常的に作る料理を健康を意識した料理にするために典型的な食材, 食材に含まれる栄養成分, 栄養成分による効能, 過剰摂取により悪影響を及ぼす成分, 食材の食感の抽出を行い, 特定の食材に対して他の食材との類似度を算出することで, 健康目的別に代替可能な食材を発見する手法を提案した。

今後の課題として, 料理名や他の食材名と共に食感系シズルワードが各食材の食感として取得されることを防ぐため, 対策を行う必要がある。また, 料理によって食材の適切さは異

なる。しゃきしゃきという食感を持つ食材に白菜やレタス, りんごなどが挙げられるが, 鍋料理に白菜の代替としてレタスを用いることは可能であると考えられるが, りんごを代替として用いることはできない。今後, 料理ごとに食材の適切さを考慮する必要がある。

謝 辞

本研究の一部は JSPS 科研費 24500134, 2633034 及び, 私学助成金 (大学間連携研究補助金) の助成によるものである。また, 本研究を遂行するにあたり, 解析対象となるレシピデータを楽天株式会社よりご提供いただいた。ここに記して謹んで感謝の意を表する。

文 献

- [1] 健康と食に関する意識調査 - 日清オイリオグループ <http://www.nisshin-oillio.com/report/report/images/120723/120723.pdf>
- [2] 貝沼やす子, 江間翔子, “ 日常の献立作りの実態に関する調査研究 (第一報) ”, 日本料理科学会誌, pp.364-371, 1997 .
- [3] 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手法 ”, DEIM Forum 2012 D9-2, 2012 .
- [4] 苅米志帆乃, 藤井敦, “ 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, pp. 959-962, 2008 .
- [5] 村瀬秀, 牛尾剛聡, “ 同一料理に対する多様なレシピ集合からの効率的な選別を目的とした可視化手法 ”, DEIM Forum 2014 F8-2, 2014 .
- [6] Liping Wang, Qing Li, Na Li, Guozhu Dong, Yu Yang, “ Substructure similarity measurement in chinese recipes ”, WWW '08 Proceedings of the 17th international conference on World Wide Web, Pages 979-988, 2008 .
- [7] Qing Li, Wei Chen, Lijuan Yu, “ Community-Based recipe recommendation and adaptation in peer-to-peer networks ”, ICUI MC '10 Proceedings of the 4th International Conference on Ubiquitous Information Management and Communication, Article No. 18, 201 .
- [8] 志土地由香, 井手一郎, 高橋友也, 村瀬洋, “ 料理レシピマイニングにおける代替可能食材の発見 ”, 電子情報通信学会論文誌料理を取り巻く情報メディア技術論文特集 Vol. J94-A No.7 pp.532-535, 2011 .
- [9] 橘明穂, 若宮翔子, 難波英嗣, 角谷和俊, “ 料理名の修飾表現の関係性に基づくレシピのネーミングコンセプト抽出 ”, 信学技報, vol.113, no.214, DE2013-36, pp. 19-24, 2013 .
- [10] 野沢健人, 中岡義貴, 山本修平, 佐藤哲司, “ word2vec を用いた代替食材の発見手法の提案 ”, 信学技報, vol.114, no. 204, DE2014-31, pp. 41-46, 2014 .
- [11] 野田雄也, 高橋哲朗, 橋本力, 鳥澤健太郎, “ WWW から獲得した知識による検索語拡張とレシピ検索タスクにおける評価 ”, 言語処理学会第 16 回年次大会, 2010 .
- [12] 苅米志帆乃, 藤井敦, “ 栄養素等摂取バランスの分析に基づく食生活支援システム ”, 日本データベース学会論文誌, Vol.8, No.4, pp.1-6, 2010 .
- [13] Youri van Pinxteren, Gijs Geleijnse, Paul Kamsteeg, “ Deriving a recipe similarity measure for recommending healthful meals ”, IUI '11 Proceedings of the 16th international conference on Intelligent user interfaces Pages 105-114, 2011 .
- [14] 基本のレシピ 100 - プロから学ぶ簡単家庭料理 シェフごはん <http://chefgohan.gnavi.co.jp/base100/>
- [15] 工藤拓, 賀沢秀人, Web 日本語 N グラム第 1 版, 言語資源協会発行, 2007.

- [16] 大橋正房, 武藤彩加, 山本真人, 爲国正子, 汲田亜紀子, 洪澤文明, 小川裕子, “「おいしい」感覚と言葉 食感の世代”, BMFT 出版部, 2010.
- [17] 加藤大介, 宮部真衣, 荒牧英治, 灘本明代, “ Twitter, Web, レシピサイトにおけるおいしさ表現の比較分析”, 信学技報, vol.114, no.204, DE2014-31, pp. 19-24, 2014.
- [18] 加藤大介, 宮部真衣, 荒牧英治, 灘本明代, “ シズルワードに着目した Twitter 上のおいしさの表現の分析”, DEIM Forum 2014 B6-6, 2014.