

メニューごとの共通食材に着目した料理レシピ検索手法の提案

崔 赫仁[†] 塩井 隆円^{††} 楠 和馬^{††} 波多野賢治^{†††}

[†] 同志社大学文化情報学部 〒610-0394 京都府京田辺市多々羅都谷 1-3

^{††} 同志社大学大学院文化情報学研究科 〒610-0394 京都府京田辺市多々羅都谷 1-3

^{†††} 同志社大学文化情報学部 〒610-0394 京都府京田辺市多々羅都谷 1-3

E-mail: [†]bim0324@mail4.doshisha.ac.jp, ^{††}{shioi,kusu}@ilab.doshisha.ac.jp, ^{†††}khatano@mail.doshisha.ac.jp

あらまし 料理をする際に情報源として使用される料理レシピサイトの中で最も使用されるのはユーザ投稿型料理レシピサイトである。誰でも料理レシピの投稿が可能であり、年々料理レシピの数が増加している。そのため、膨大な料理レシピの中でユーザの目的に合う料理レシピを見つけることが困難になりつつある。特に、現在の料理レシピサイトの検索補助では料理レシピ間の差がわかりづらいため、基準となる料理レシピと基準から離れている料理レシピの差がわかりやすく提示することにより料理レシピの選びに悩むユーザの負担の削減に貢献でき、レパートリー増加にもつながる。そこで本研究では、料理メニューごとに共通して使用される食材に着目して基準となる料理レシピを抽出し、その料理レシピとの差分を用いた料理レシピ検索手法を提案する。その後、基準となる料理レシピとの差分が大きい順に並べたものを新着順と比較するアンケート調査を行い、どの並替え順が新しいレパートリーが見つかりやすいかを確認する。

キーワード アソシエーション分析, 相関ルール, 料理レシピ検索

1. はじめに

クックパッド社が実施した「料理に関するアンケート」^(注1)によると、料理する際の悩みは「つい同じものばかり作ってしまう」「レパートリーが少ない」「得意といえる料理がない」「料理レシピを見ないと作れないことが多い」「メニューを決めるのに時間がかかってしまう」のような、調理可能な料理のレパートリーが少ない状態で、料理するメニューを決めることであると報告されている。

料理をする際に情報源として使用される料理レシピサイトは無料で使用でき、料理レシピ数も豊富であるが、その中で最も使用されるのは楽天レシピ^(注2) やクックパッド^(注3) のようなユーザ投稿型料理レシピサイトである。ユーザ投稿型料理レシピサイトは誰でも料理レシピの投稿が可能だという特徴があり、年々料理レシピの数が増加している。そのため、膨大な料理レシピの中でユーザは自分の目的に合う料理レシピを見つけることが困難になりつつある。

現在、ユーザ投稿型料理レシピサイトが提供している検索補助は「人気順」「新着順」の並替え、カテゴリの分類や時間・カロリーなどで絞込み、献立作成などを提供している。しかし、このような検索支援は料理レシピの特徴を捉えておらず、料理レシピ間の差異をユーザが認識しづらいため、レパートリー不足とメニュー決めが困難というユーザの悩みを解決することができない。

ユーザの悩みを解決する方法の一つに、基準となる料理レシピ（以下、定番レシピ）を設定し、定番レシピから離れている料理レシピを上位に表示する方法が挙げられる。ユーザによって投稿される料理レシピの内容は多様ではあるが、食材や調理手順がアレンジされていても同一の料理名が付けられている以上、共通の食材が存在するため、その食材の組合せを見つけることによって定番レシピを定義することは可能である。

そこで本研究では、料理レシピ間の差をわかりやすくするために、メニュー別に頻出する食材の組合せを用いて定番レシピを設定し、定番レシピとの差分を考慮した料理レシピ検索手法を提案する。また、提案手法の有用性を確かめるために既存の料理レシピサイトで用いられている並替え方法との比較評価を行う。

2. 関連研究

ユーザのレパートリー拡大のためには本研究で設定する定番レシピの基準となる典型度と同じメニューの料理レシピでもレパートリーがわかりやすいように類似レシピをグルーピングすることなどが考えられる。したがって本節では、典型度に関する研究と類似レシピの分類に関する研究について説明する。

2.1 料理レシピの典型度に関する研究

横井らは食材の組合せの典型度を分析することによって食材の組合せを考慮した食材推薦と、ユーザが意図した料理レシピ検索を目的として食材の組合せの典型度分析手法を提案している [1]。料理群の分類は「料理名の語尾が同一であることが、ある程度共通の手順により作られる料理である」という仮定 [2] をもとに、形態素解析で料理レシピのタイトルの語尾を抽出し、その結果に従って料理レシピを各料理群に分類している。

料理レシピを各料理群に分類した後、各料理群の中の全料理

(注1)：料理に関するアンケート，<https://cf.cpcdn.com/info/assets/wp-content/uploads/2014030600000/pr130723-survey.pdf> (2017/03/27 閲覧)

(注2)：楽天レシピ，<https://recipe.rakuten.co.jp/> (2017/03/27 閲覧)

(注3)：クックパッド，<https://cookpad.com/> (2017/03/27 閲覧)

レシピで使用されている全ての食材を要素とする食材ベクトルを作成し、各料理群において食材ベクトル群を主成分分析を利用して典型度を求めている。主成分分析で構築される固有空間は、その料理群に起きる典型的な食材の組合せを表す空間であり、第1固有ベクトルの要素には各料理群における食材の出現頻度の平均が表れ、第2固有ベクトル以降には固有ベクトルごとに出現しやすい食材の組合せの要素の値が大きく表れるという性質を利用している。しかし、料理レシピの典型度を考慮しているが、主成分分析で得られた主成分の解釈が主観的であるため、客観的に料理レシピの典型度を計算する必要がある。

また、中岡らは調理者の調理経験によってレパートリーを拡大する際の食材や調理法が異なると考え、料理レシピに出現する食材と調理手順に含まれる調理法の出現頻度、およびそれらの関係性を分析し、よく使用される食材や調理法を優先して推薦する手法を提案している [3]。中岡らはより多くの料理レシピで使用される食材は使い勝手がよく、多くの料理レシピで必要とされる調理法は経験しておくことで調理者の負担を軽くすることができると考え、食材の使い勝手や調理法の馴染み深さの指標として定番度を提案している。また、調理時には食材に対して調理法を実行することから、一つの動作として食材と調理法の組合せに関しても定番度を算出している。料理レシピに使用される食材と料理レシピをノードとして食材を使用している料理レシピとの間にリンクを張ることで2部グラフを作成し、Webページのランキング手法であるHITSアルゴリズムを適用することで、定番度を算出している。食材の組合せの定番度を計算しているが、メニューの中の定番度でなく、全てのメニューでよく使われる食材の組合せの定番度を計算されている。

2.2 類似レシピに関する研究

花井らは、細かい部分では違いが見受けられるがほとんど同一の料理と言える酷似レシピが数多く存在することにより、ユーザの求めるレシピを検索することが困難になっていることに着目し、酷似レシピ抽出のためのクラスタリング手法を提案している [4, 5]。花井らはユーザが検索を行い、提示された各料理レシピのタイトルとスニペットを見た状態で、違いを感じないような料理レシピを酷似レシピと定義している。具体的にはユーザが入力したキーワードから料理レシピ検索を行い、その検索結果の各料理レシピのタイトル・スニペット・材料・作り方・料理レシピの生立ちを取得している。次に、それらを変数として二つのクラスタリング手法の分類結果を比較することで適した料理レシピのクラスタリング手法の考察を行っている。その考察を基に、料理レシピから主食材と主調味料を抽出し、抽出した主食材・主調味料および食材の希少度を用いて類似レシピをクラスタリングする方法を提案している。類似レシピを抽出することで、料理レシピ同士の共通点はわかるが、その料理レシピ群が定番レシピであることまでは判断できない。

3. 提案手法

ユーザ投稿型料理レシピサイトの料理レシピは同じメニューの料理レシピであっても使用食材や手順など、料理レシピの構成要素が異なる。しかし、ユーザがあるメニュー名を付けてい

るのは料理レシピの構成要素のうち、そのメニューの特徴となる何らかの共通点が存在することである。その特徴として考えられるのが使用食材であり、共通して使用されている食材はそのメニューの特徴になると考えられる。

したがって、本研究の目的は多様なユーザが投稿した料理レシピから各メニューの共通の食材を含む定番レシピを見つけ、定番レシピから離れている料理レシピをユーザに提示することである。これは前述の、検索結果の料理レシピ間の差異がわかりづらいことに加え、アンケートに挙げられている料理する際の悩みであるメニューの選択やレパートリー拡大にも貢献できると考えられる。

そこで、本研究では各メニューごとに共通する食材の組合せを見つけるために、各メニューの食材の相関ルールを検出する。検出した相関ルールを用いて定番レシピの設定と定番スコアの計算を行い、定番レシピとの差分を考慮したレシピ検索手法を提案する。

3.1 料理レシピの分類

本研究では、楽天技術研究所と国立情報学研究所が提供している楽天レシピの料理レシピデータ 796,274 件^(注4)を用いる。なぜなら、楽天レシピは全ての料理レシピにカテゴリ情報が付与されているからである。本研究ではメニューごとに頻出する食材の組合せを抽出するため、料理レシピデータのうち、カテゴリ情報と材料情報のみを用いる。カテゴリ情報は各料理レシピごとに大・中・小の三つが与えられており、それらのカテゴリはユーザが料理レシピを投稿する際に投稿フォームの選択肢から選んだものである。それらはユーザが想定した料理群として料理レシピに付与しているため、メニューの分類にカテゴリ情報を利用する。大カテゴリは43個に分けられ、小カテゴリを基準として1,416個のカテゴリに料理レシピを分類して分析に用いる。

3.2 食材名の表記ゆれの統一

ユーザ投稿型料理レシピサイトの料理レシピは全てユーザが自由に記述しているため、投稿者によって食材名の表記や手順の説明の仕方が異なる。本研究で用いる料理レシピデータの材料情報も同じ食材を使用している料理レシピでも投稿者によって材料名に表記ゆれが生じる。

表1 素材の表記例

材料
じゃがいも
じゃがいも 中
メークイーン
馬鈴薯
冷凍ポテト
●新じゃが芋
野菜 (玉葱・じゃが芋・人参)
ジャガイモの皮・ナスのヘタ
じゃがいも、人参、かぶ、かぼちゃ など

(注4) : 楽天データセット, <https://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html> (2017/03/27 閲覧)

表1のように食材名をひらがな・カタカタ・漢字で表記する場合や食材の品名で表記する場合や食材名以外にサイズや用途など説明を加えている場合がある。そのため、材料情報を生データのまま分析に用いると、同じ食材にも関わらず別の食材として扱われ正確な結果が得られないため、材料名から食材名のみを取得し同じ食材は表記を統一させる必要がある。

そこで本研究では食材の表記の統一に料理オントロジーを用いて対処する。この料理オントロジーは同じ食材を使用しているにもかかわらずユーザによって漢字・英語・ひらがな・カタカナ・品名など、書き方が異なる問題点を解決するために構築されている [6]。

料理オントロジーの概念階層を表す「カテゴリーエントリー-同義語」の3階層の概念辞書を構築し、カテゴリは楽天レシピのカテゴリを参考に「材料-魚介」「材料-肉」「材料-野菜」「材料-その他」「調味料」「調理器具」「動作」の7種類に設定してから、エントリ辞書・同義語辞書・属性辞書・部分辞書を構築しているが、本研究では材料と調味料の同義語辞書のみを使用している。表2に料理オントロジー構築プロジェクト^(注5)に公開されている料理オントロジーのうち、本研究で用いるカテゴリの情報を示す。

表2 料理オントロジーの同義語辞書

カテゴリ	エントリー	代表表現の数
材料-魚介	魚類, 魚介類, 海産物, 水産物	62
材料-肉	肉類, 食肉, 食肉類, 原料肉	6
材料-野菜	野菜, 果菜類, 野菜類, 果菜物, 農産物	112
材料-その他	いずれにも当てはまらない用語	55
調味料	調味料, 香辛料, 薬味, スパイス類	51

料理オントロジーの同義語辞書は、ある食材の代わりに別の食材を使うことが料理レシピの特徴となるかどうかを基準として人手で同義語を選定して構築している。料理レシピデータの素材に出現する食材名を料理オントロジーの上位概念を代表表現とし、下位概念の食材名を代表表現に置換して分析に用いる。

料理レシピデータの材料情報は閲覧者の理解度を深めるために記号を表記する場合や素材の状態の説明を加える場合、複数の素材名を一つの素材欄に表記する場合など、食材名以外の文字が含まれているという問題点を解決するために形態素解析を行う。

食材名以外の文字を除外して必要な食材名のみを抽出するためにまず、特殊記号と数字、アルファベットは削除した素材名を形態素解析に用いる。特殊記号と数字、アルファベットは食材名に含まれておらず、商品名やサイズ・個数などの付加説明のために用いられるためである。

次に、辞書追加が容易である形態素解析器 MeCab [7] を用いて形態素解析を行う。形態素解析を通して食材名と食材名でない形態素に分けた後、必要な食材名のみを抽出する。料理オントロジーの同義語の代表表現を MeCab の辞書に登録し、形態素解析の結果から食材名の代表表現を抽出し、分析に用いる食

材名して扱う。形態素解析後の代表表現の取得は以下の規則に従って行う。

【case 1】 形態素解析の結果、分割された形態素が一つ

- 代表表現が存在する場合
代表表現を抽出して食材名とする。
- 代表表現が存在しない場合
そのまま食材名とする。

【case 2】 形態素解析の結果、分割された形態素が二つ

- いずれも代表表現が存在する場合
各々の代表表現を抽出し、複数の食材に分けて食材名とする。
- どちらかにしか代表表現が存在しない場合
代表表現のみを抽出して食材名とする。
- いずれも代表表現が存在しない場合
形態素解析前の素材名を食材名とする。

【case 3】 形態素解析の結果、分割された形態素が三つ

- 「A の B」の形式の場合
 - A, B いずれも代表表現が存在する場合
「代表表現の代表表現」のように食材名とする。
 - A, B どちらかにしか代表表現が存在しない場合
代表表現のみを抽出して食材名とする。
 - A, B いずれも代表表現が存在しない場合
形態素解析前の素材名を食材名とする。
- 代表表現が一つ以上存在する場合
各々の代表表現を抽出し、複数の食材に分けて食材名とする。
- 代表表現が存在しない場合
形態素解析前の素材名を食材名とする。

【case 4】 形態素解析の結果、分割された形態素が四つ以上

- 代表表現が一つ以上存在する場合
各々の代表表現を抽出し、複数の食材に分けて食材名とする。
- 代表表現が存在しない場合
形態素解析前の素材名を食材名とする。

3.3 関連ルール検出

各料理レシピの共通の食材を含む定番レシピを設定するために定番レシピを構成する定番食材を抽出する必要がある。食材の組合せの抽出にはアイテム間の関係性を抽出するアソシエーション分析を用いる。

アソシエーション分析は店舗の POS (Point Of Sales : 販売時点情報管理) データの分析のために開発された手法で、膨大なデータの中から意味のある関係性を抽出することを目的としている。アソシエーション分析は2種類のオブジェクトの多対多の関係の共通の形式を記述するのに使われ、「アイテム (item : 商品)」とトランザクション (transaction : 取引) とも呼ばれる「バスケット (basket : 買い物かご一つに相当)」を持っている [8]。本稿では、料理レシピの食材リストをトランザクションとして扱う。

(注5) : 料理オントロジー構築プロジェクト, <http://www.ls.info.hiroshima-cu.ac.jp/cgi-bin/cooking/wiki.cgi> (2017/03/27 閲覧)

相関ルールは「商品 X を買うと商品 Y も買う」のようなルールのことで、 $X \Rightarrow Y$ と記述する。ここで、ルールの左辺 X を条件部、右辺 Y を帰結部と呼び、 X と Y はアイテム集合を表す。相関ルールは二つのアイテム集合間の関係性を表し、評価指標としては支持度・確信度・リフト値がある。有効な相関ルールとは支持度と確信度は高いほど、リフト値が 1 以上のルールを指す [9]。

支持度 (support)

$$\text{supp}(X \Rightarrow Y) = \frac{\text{「}X \text{ かつ } Y\text{」を含むトランザクション数}}{\text{全トランザクション数}}$$

確信度 (confidence)

$$\text{conf}(X \Rightarrow Y) = \frac{\text{「}X \text{ かつ } Y\text{」を含むトランザクション数}}{\text{「}X\text{」を含むトランザクション数}}$$

Agrawal らが提案したユーザが最小確信度と最小支持度を与えて、最小確信度以上の確信度と最小支持度以上の支持度を持つ相関ルールを重要な相関ルールとして、全て発見するアプリアルゴリズムを用いて相関ルールを検出する [10]。

本研究では定番レシピの設定に最小支持度 0.3、最小確信度を 0.7 に指定してその以上の値を持つ相関ルールを検出した。こちらは、次節で説明する定番レシピが抽出できるカテゴリの個数が最も多かった値であったためである。定番レシピ設定に用いた相関ルールに関しては、全 1,416 個のカテゴリのうち、約 8 割である 1,145 個のカテゴリから相関ルールを検出することができた。

3.4 定番食材の定義と定番スコア算出

検出された相関ルールを用いて各カテゴリにおける定番レシピを設定する。定番スコアが低い順 (以下、非定番順) に料理レシピを表示させるために、設定した定番レシピが抽出されたカテゴリのみを用いて定番スコアを計算する。

支持度 0.3、確信度を 0.7 以上の相関ルールに出現する食材は各カテゴリにおいて、出現頻度が高い順と一致するカテゴリが多かったため、そのカテゴリの定番食材とする。この結果から、出現頻度が高い食材同士の組合せが料理レシピ中に頻出することが分かった。相関ルールが抽出できた 1,145 個のカテゴリから各カテゴリの定番食材を抽出し、抽出された定番食材のみを含む料理レシピを定番レシピとして抽出した。その結果、全カテゴリ 1,416 個のカテゴリのうち、346 個のカテゴリから定番レシピを抽出することができた。

料理レシピの並替えに用いる定番スコアは式 (1) により算出する。ここで、 n_C はある料理レシピ X が属しているカテゴリ C の定番食材の個数、 $n_{\bar{X}}$ は料理レシピ X が含んでない定番食材の個数、 n_{X_C} は料理レシピ X が含んでいるカテゴリ C の定番食材以外の食材の個数である。 n_{X_C} を考慮した理由は料理レシピ X がカテゴリ C の、定番食材以外を含んでいた場合も料理をアレンジしたものと考えられるためである。

$$\text{定番スコア} = \frac{n_C - n_{\bar{X}}}{n_C + n_{X_C}} \quad (1)$$

4. 評価実験

本節では、提案した検索結果の並替え順が実際に新しいレパートリーを見つけることに有用であるかを確かめるため、現在ユーザ投稿型料理レシピサイトで提供している新着順との比較実験を行う。比較実験は性別・職業などの属性が多様な人から回答を得るために、日本最初のクラウドソーシングサービスであるランサーズ^(注6) を使用してアンケート調査を行った。

4.1 アンケート調査

比較実験ではユーザに提案手法の非定番順と新着順のうち、「どちらが新しいレパートリーを見つけやすいか」を回答するアンケート調査を実施した。アンケート回答者の負担にならないように、10 個の質問から構成されるアンケート A~E の五つに回答してもらうことにした。各質問は定番食材のみを使用している料理レシピが存在する 346 個のカテゴリを対象とした。そのうち、小カテゴリ名が料理名である 148 個のカテゴリを料理レシピが多い順で並べたグラフが表 1 のようにロングテール型になり、148 個のカテゴリの総料理レシピ 8 割を含んでいるのは上位 43 個のカテゴリであったが、アンケートの作成を用意するために上位 50 個のカテゴリを使用した。アンケートに使用したカテゴリを表 3 に示す。

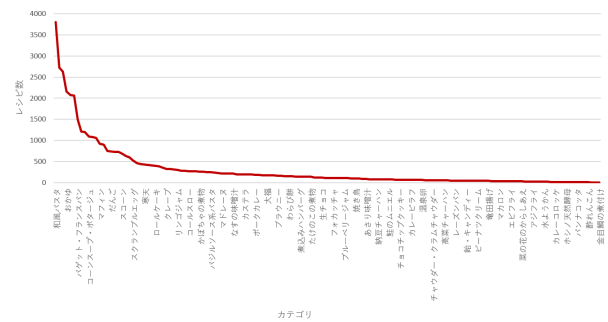


図 1 小カテゴリが料理名であるカテゴリの料理レシピ数

アンケートは各カテゴリ内の料理レシピを投稿日が早い順で並替えした新着順と提案手法の非定番順に各 10 件並べたものをユーザに提示した。並べた料理レシピにはユーザ投稿型料理レシピサイトのように料理レシピのタイトル、ユーザ ID、料理の簡単な説明、材料一覧、料理の画像を含めるようにし、料理レシピのタイトルをクリックすれば実際の楽天レシピの料理レシピが見えるようにした。これは、ユーザが実際の料理レシピサイトを使用しているような場面を設定したためである。

また、ユーザの属性に関する付帯質問は以下の通りである。

- 性別
- 年齢
- 職業
- 料理頻度
- 婚姻状況 (子ども)

具体的なアンケートの例を図 2 に示す。あるカテゴリの新着順と非定番順の並替えをアンケート回答者にはわからないよう

(注6) : ランサーズ, <http://www.lancers.jp/> (2017/03/27 閲覧)

表 3 実験に用いたカテゴリ

アンケート	大カテゴリ	中カテゴリ	小カテゴリ
A	ご飯もの	オムライス	オムライス
	お菓子	その他のお菓子	クレープ
	汁物・スープ	コンソメスープ	コンソメスープ
	お菓子	シュウクリーム	シュウクリーム
	卵料理	スクランブルエッグ	スクランブルエッグ
	パン	ハードブレッド	ポケット・フランスパン
	サラダ	マカロニサラダ	マカロニサラダ
	お菓子	チーズケーキ	レアチーズケーキ
	人気メニュー	唐揚げ	鶏のから揚げ
	お菓子	和菓子	羊羹
B	ご飯もの	チャーハン	アレンジチャーハン
	お菓子	クリーム・ジャム	イチゴジャム
	人気メニュー	煮物	かぼちゃの煮物
	汁物・スープ	コンソープ・ポタージュ	コンソープ・ポタージュ
	お菓子	スコーン・マフィン	スコーン
	サラダ	素材で選ぶサラダ	トマトサラダ
	パン	フレンチトースト	フレンチトースト
	お菓子	ホットケーキ・パンケーキ	ホットケーキ・パンケーキ
	お菓子	焼き菓子	マドレーヌ
	卵料理	目玉焼き	目玉焼き
C	ご飯もの	おかゆ・雑炊類	おかゆ
	お菓子	和菓子	おはぎ
	サラダ	かぼちゃサラダ	かぼちゃサラダ
	定番の肉料理	とんかつ	とんかつ
	汁物・スープ	味噌汁	なすの味噌汁
	パスタ	バジルソース系パスタ	バジルソース系パスタ
	パン	プレーンなパン	ペーグル
	お菓子	スコーン・マフィン	マフィン
	お菓子	クリーム・ジャム	クリーム
	人気メニュー	肉じゃが	肉じゃが
D	汁物・スープ	その他のスープ	オニオンスープ
	サラダ	コールスロー	コールスロー
	お菓子	ケーキ	スポンジケーキ
	お菓子	ゼリー・寒天・ムース	ゼリー
	お菓子	和菓子	ぜんざい
	汁物・スープ	味噌汁	なめこの味噌汁
	人気メニュー	野菜炒め	もやし炒め
	パン	プレーンなパン	丸パン
	ご飯もの	おかゆ	雑炊類
	パスタ	和風パスタ	和風パスタ
E	汁物・スープ	クリームスープ	クリームスープ
	パン	サンドイッチ	サンドイッチ
	卵料理	だし巻き卵・卵焼き	だし巻き卵・卵焼き
	お菓子	和菓子	だんご
	サラダ	ポテトサラダ	ポテトサラダ
	中華料理	マーボー豆腐(麻婆豆腐)	マーボー豆腐(麻婆豆腐)
	お菓子	ケーキ	ロールケーキ
	麺・粉物料理	焼きそば	塩焼きそば
	お菓子	ゼリー・寒天・ムース	寒天
	ご飯もの	おにぎり	焼きおにぎり

(問題 1) 以下の二つの並び替えはカテゴリ「ご飯もの>オムライス>オムライス」のレシピをある基準によって並び替えたものです。

以下URLをクリックすると、並び替えの画像(pdf)が表示されます。
料理レシピのタイトルをクリックすると、実際の料理レシピが閲覧できます。

- 並び替え 1 : <https://drive.google.com/open?id=0B3W0vUlm20jyVDNc0g1dUQ2Q1U>
- 並び替え 2 : <https://drive.google.com/open?id=0B3W0vUlm20jyZ1BnYmQ1ZkZ0cmRk>

- 1-1. 新しいレポートリを見つけやすい並び替えをお選びください
- 並び替え 1
- 並び替え 2
- 1-2. 選択した理由を教えてください

作業の進捗チェック



図 2 アンケートの例

表 4 付帯質問

属性	回答	人数	属性	回答	人数
性別	女性	33	料理頻度	ほぼ毎日	29
	男性	22		週 4~5 回程度	6
年齢	20 代	5	週 2~3 回程度	6	
	30 代	16	週 1~2 回程度	7	
	40 代	25	2 週間に 1 回程度	3	
	50 代	7	月 1 回程度	3	
	60 代以上	2	料理しない	1	
職業	学生	0	婚姻状況	未婚	30
	会社員	16		既婚	25
	専業主婦	16	子ども	いる	20
該当なし	23	いない		35	

表 5 新しいレポートリを見つけやすいと回答した割合

	A	B	C	D	E
提案手法	0.59	0.49	0.59	0.43	0.49
新着順	0.41	0.51	0.41	0.57	0.51

に表示し、どちらの並替えが新しいレポートリを見つけやすいかを選択する質問と選択した理由を記述式で回答するアンケートとなっている。

4.2 結果

一つのアンケートにつき 11 名、合計 55 名(回答者の属性は表 4 を参照)にレポートリを見つけやすいかを判断させた結果、表 5 のようになった。

表 5 に示すとおり、実験結果は新着順と提案手法の非定番順のうち、新しいレポートリの見つけやすさには差がないという結果であり、各回答者の回答の割合を対応のある t 検定を有意水準 5%で行った結果、p 値が 0.05 を上回ったため、帰無仮説が採択され、提案手法と新着順の結果には有意差があると

は言えず、新着順より提案手法の方が新しいレポートリを見つけやすいとは言えない結果となった。対応のある t 検定は平均値の差が得られる確率を推定する場合に式 (2) によって求める [11]。D は対応する値の差を示している。

$$t = \frac{\bar{D}}{\sqrt{\frac{\sum (D_i - \bar{D})^2}{n(n-1)}}} = \frac{\bar{D}}{\sqrt{\frac{\sum D_i^2 - (\sum D_i)^2}{n^2(n-1)}}} \quad (2)$$

4.3 考察

回答理由を考察すると、「楽しそうなレシピが並んでいるから」や「好きなかぼちゃメニュー中心だから」など、質問の意

図を理解せずに回答した場合があった。また、「見た目が変わっているから」や「見た目も良く、美味しそうに感じたため」のように、料理レシピのレポートリでなく、表示された画像のイメージで判断しているケースもあったため、回答者にアンケートの意図が理解できるように質問内容の伝達法を工夫する必要がある。

さらに、「料理が得意でなくてもまねできそうなレシピが多そうだから」や「作りやすそうなものが多かったから」のように新しいレポートリの判断基準が使用食材でなく、作りやすさという基準で判断される場合もあった。この場合は料理頻度が低い男性の回答者も多かったため、料理レシピの種類が豊富であることに着目されず、作ることができるかできないかで新しいレポートリの見つけやすさが判断されたと考えられる。したがって、レポートリを判断する着目点がわかるような質問項目の明確化や注意事項に着目すべきところを明記するなど、アンケートの構成を工夫する必要がある。

最後に、回答の割合がほぼ 0.5 に近く、新着順と非定番順に新しいレポートリの差がないという結果であった。このような結果が得られたのはアンケートの構成にも問題があったが、比較対象として用いた新着順の並替えは投稿された日が基準となり、料理レシピの要素が並替えの基準になっていなかったため、新着順ではなく人気順などの並替えとの比較を見直す必要もある。

5. おわりに

本稿では、定番レシピの設定と定番スコアの算出のためにカテゴリごとに食材の相関ルールを検出し、相関ルールに出現する食材のみを含む料理レシピを定番レシピとその差分を用いた料理レシピ検索手法の提案を行った。食材名の表記ゆれを統一させるために料理オントロジーの同義語辞書を用いて形態素解析を行って食材名を統一し、カテゴリごとに相関ルールを検出してカテゴリごとに頻出する食材の組合せを取得することができた。また、取得した食材の頻出組合せから定番食材と定番レシピを設定し、定番スコアを計算して非定番順に料理レシピを並替えすることができ、既存の新着順より新しいレポートリを見つけやすいかを確かめるための評価実験を行った。

今後の課題としては、より多いカテゴリで定番レシピが抽出できるように料理群の分類および食材名抽出法の見直しや定番食材以外の食材を用いた定番レシピ設定方法の見直しなどが考えられる。また、評価実験の結果からアンケートの内容を工夫して修正する必要がある。具体的には、アンケート回答者の理解を考慮した質問項目の明確化や注意事項の明記などアンケート構成の再考、評価実験の比較対象の再検討などが挙げられる。

謝 辞

本研究を遂行するにあたり貴重なご意見いただきました同志社大学文化情報学部の矢野環教授に感謝します。本研究の一部は JSPS 科研費 15H02701 の助成を受けたものである。また、楽天技術研究所が国立情報学研究所に提供した「楽天レシピデータ」を利用した。

- [1] 横井聡, 道満恵介, 平山高嗣, 井手一郎, 出口大輔, 村瀬洋. 料理レシピにおける食材の組み合わせの典型度分析. 電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎, 第 114 巻, pp. 49–54, 2015.
- [2] 志土地由香, 高橋友和, 井手一郎, 中村裕一, 村瀬洋. 調理レシピテキストからの代替素材の発見. 人工知能学会全国大会 (第 22 回) 論文集 6-6, 2008.
- [3] 中岡義貴, 佐藤哲司. 調理レポートリ拡大のためのレシピ推薦手法の提案と評価. 電子情報通信学会他共催, 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 DEIM2015 論文集 D3-1, 2015.
- [4] 花井俊介, 灘本明代. 酷似レシピ抽出のためのクラスタリング手法の提案. 電子情報通信学会他共催, 第 6 回データ工学と情報マネジメントに関するフォーラム DEIM2014 論文集 F8-6, 2014.
- [5] 花井俊介, 難波英嗣, 灘本明代. 主食材と主調味料を考慮した類似レシピクラスタリング. 電子情報通信学会他共催, 第 8 回データ工学と情報マネジメントに関するフォーラム DEIM2016 論文集 E2-7, 2016.
- [6] 土居洋子, 辻田美穂, 難波英嗣, 竹澤寿幸, 角谷和俊. 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. 電子情報通信学会 MVE/CEA 研究会, 第 113 巻, pp. 37–42, 2014.
- [7] 工藤拓, 山本薫, 松本裕治. Conditional random fields を用いた日本語形態素解析. 情処学 NL 研報, pp. 161–13, 2004.
- [8] 岩野和生, 浦本直彦. 大規模データのマイニング. 共立出版株式会社, 2014.
- [9] 福田剛志, 森本康彦, 徳山豪. データマイニング (データサイエンスシリーズ). 共立出版, 2001.
- [10] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, 1993.
- [11] 森敏明, 吉田寿夫. 心理学のためのデータ解析テクニカルブック. 北大路書房, 1990.