

## 意外性のあるレシピを推薦するエージェントの提案

池尻 恭介<sup>†a)</sup> 清 雄一<sup>†</sup> 中川 博之<sup>††</sup> 田原 康之<sup>†</sup>  
大須賀昭彦<sup>†</sup>

## A Proposal of Recommendation Agents for Surprising Recipes

Kyosuke IKEJIRI<sup>†a)</sup>, Yuichi SEI<sup>†</sup>, Hiroyuki NAKAGAWA<sup>††</sup>, Yasuyuki TAHARA<sup>†</sup>,  
and Akihiko OHSUGA<sup>†</sup>

あらまし 毎日の食事のレシピを考えることは非常に大変なことである。近年では独自のレシピを自由に書き込むことができる投稿型レシピサイトが多数存在しており、サイト内のレシピ数、ユーザ数は年々増加している。投稿型レシピサイトには、普通とは多少異なる食材を用いた意外性のあるレシピが存在するが、通常の検索機能を使って発見するのは困難である。そこで、本研究では投稿型レシピサイトから意外性のあるレシピを抽出するための推薦エージェントを提案する。このレシピ推薦エージェントは TF-IDF の考えを応用した RF-IIF (Recipe Frequency-Inverse Ingredient Frequency) を利用し、ユーザから指定された料理カテゴリーにおける食材の希少度と一般度から意外度を算出する。次にレシピに出現する食材の意外度からレシピの意外度を算出するが、各レシピの料理カテゴリーを誤判定すると、普通のレシピが意外レシピと誤判断されてしまうため、別カテゴリーのレシピをいかに除外するかが重要である。最後に、レシピ間の類似度を計り類似したレシピを除去することで多様性に富んだ意外性のあるレシピを抽出する。アンケートによる評価を実施し、提案するレシピ推薦エージェントの有用性を示した。

キーワード データマイニング, 情報推薦, 食メディア, レシピ

## 1. ま え が き

近年、投稿型レシピサイトの利用者数が増加している。ユーザは独自のレシピを自由に投稿することが可能であり、キーワードや食材名からレシピを検索する機能を有し、該当するレシピを取得することが可能である。このような点から毎日の献立を考える主婦などを中心に利用され、サイト内のレシピ数、ユーザ数は年々増加している。

投稿型レシピサイトの中には一般的なレシピの他に普通では使用しない食材を用いたレシピや普通とは異なる調理工程のレシピのような、意外性のあるレシピが多数存在する。いつも食べているレシピは飽きたと

感じた際に意外性のあるレシピを閲覧することができれば、料理のバリエーションが増え食事の幅が広がる。しかし、意外性のあるレシピは「意外」という単語を含むレシピばかりではないため、検索機能を利用して発見できないレシピが多数存在している。そのような意外性のあるレシピを発見するのは困難であり、多くのレシピが埋もれているのが現状である。そこで本研究では、投稿型レシピサイトから普通では使用しない食材を用いた意外性のあるレシピ（以下、意外レシピとする）を抽出する、レシピ推薦エージェントを提案する。

本研究では、直観的には「この食材を、この料理にも利用することができるのか」という新たな発見があるようなレシピを意外レシピの一つとして考える。より詳細には、「『通常、料理において利用されることが多いが、ユーザが指定した特定の料理カテゴリーにおいてはほとんど使われることのない食材』を含んでいる、当該料理カテゴリーのレシピ」を意外性のあるレシピの一つと考える。

以下、2. では想定シナリオと先行研究について、3.

<sup>†</sup> 電気通信大学大学院情報システム学研究科, 調布市  
Graduate School of Information Systems, The University of  
Electro-Communications, Chofu-shi, 182-8585 Japan

<sup>††</sup> 大阪大学大学院情報科学研究科, 吹田市  
Graduate School of Information Science and Technology,  
Osaka University, Suita-shi, 565-0871 Japan

a) E-mail: k-ikejiri@ohsuga.is.uec.ac.jp

DOI:10.14923/transinfj.2014SWP0007

では課題について記述する。4., 5., 6. では提案手法についての説明, 評価実験, 考察について記述する。7. では関連研究について記述し, 最後に 8. ではまとめと今後の課題について記述する。

## 2. 背景

### 2.1 想定シナリオ

自分のレパートリーの範囲で毎日料理をしている主婦が, 普段とは少し異なる料理を作りたいと考えている状況を想定する。

例えば, ハンバーグが食べたいがいつものレシピには飽きたという状況を想定する。本研究の提案システムを利用すれば, 料理カテゴリー名<sup>(注1)</sup>として「ハンバーグ」を指定することで, きゅうりをソースに利用することでさっぱりさを味わうことができるハンバーグのレシピなどが提示される。また, パスタ料理であれば, 米粉を利用した簡単なクリームパスタのレシピなど, 意外性のあるレシピが提示される。これは, いつもとは異なる料理が食べたいが, 全く調理したことがない料理に挑戦するのは敷居が高いと感じる際に, 有用であると考ええる。

なお, 意外かどうかの基準は, 本来はユーザのレパートリーに依存して決定されるものであるが, 本論文では, 多くのユーザが共通に意外であると考えられるレシピを, 意外レシピの正解データとして利用する。このようなレシピを抽出した後, 各ユーザのレパートリーに合致するものを除外する(各ユーザが過去に料理したことのあるレシピ情報を入力として受け付けることができれば, 機械的に除外することがある程度可能である)ことで, 各ユーザにとっても意外なレシピを推薦できる。

また, 意外性のあるレシピを抽出することで, 食材の意外な利用方法を学ぶことは, 食育の観点からも有用であると考えられる。

### 2.2 レシピサイトにおけるレシピの構成

レシピはタイトル, 完成画像, 投稿者 ID, 食材一覧, 調理工程, 投稿者によるコメント(概要, アドバイス, 生い立ち, 等)の六つの要素で構成されていることが多い。

また, 多くの投稿型レシピサイトではレシピを投稿する際にレシピがどのような料理カテゴリーに分類されるのか登録する項目がない。レシピのキーワードをタグとして登録するものなどは存在するが, レシピの明確なカテゴリー分類はなされていない。

### 2.3 先行研究

筆者らの先行研究[1]で提案された TF-IDF (Term Frequency-Inverse Document Frequency) を応用した RF-IIF (Recipe Frequency-Inverse Ingredient Frequency) について説明する。

まず, TF-IDF について簡単に述べる。TF-IDF は, 文書内の特徴語を得る際に用いられる。文書内での単語の出現頻度である TF と, 文書集合内での単語の逆文書出現頻度である IDF を掛け合わせることで単語の重みを算出する。つまり, ある文書内では頻出する単語だが, 他の文書においてはあまり出現しない単語の重みが大きくなる。これによって, その文書のみの特徴語を得ることが可能である。

RF-IIF は, TF-IDF の考え方を応用したものである。ある料理カテゴリー  $m$  の食材として, あまり使われない食材は料理カテゴリー  $m$  において意外な食材である可能性がある。しかし, あまり使われない食材は, 意外な食材ばかりではなく食材自体の知名度が低く, ユーザの知らない食材である可能性が高く, ユーザが調理方法を知らないことが危惧される。また, スーパーマーケットなどで取り扱っておらず, 手軽に手に入れることができない場合もある。そこで, それらの食材の中で, 更にレシピ全体での出現頻度の高い食材を料理カテゴリー  $m$  における意外食材とする。

$RF-IIF_{i,m;p}$  (式 (1)) は食材  $i$  について, 料理カテゴリー  $m$  のみにおいての珍しさを示すものである。 $RF-IIF_{i,m;p}$  は, 食材  $i$  の一般度を示す  $RF_i$  (式 (2)) と, 料理カテゴリー  $m$  における食材の希少度を示す  $IIF_{i,m}$  (式 (3)) から算出する。ここで  $R_{all}$  は全レシピ数,  $R_i$  は全レシピの中で食材  $i$  が出現するレシピ数であり,  $R_m$  は料理カテゴリー  $m$  のレシピ数,  $R_{i,m}$  は料理カテゴリー  $m$  のレシピの中で食材  $i$  が出現するレシピ数である。

$$RF-IIF_{i,m;p} = RF_i \times IIF_{i,m}^p \quad (1)$$

$$RF_i = R_i / R_{all} \quad (2)$$

$$IIF_{i,m} = \log(R_m / R_{i,m}) \quad (3)$$

(注1): 「カレー」, 「パスタ」等。提案手法では, 指定された料理カテゴリー名を含むレシピタイトルのものをまず抽出する。料理カテゴリー名は, レシピタイトルに含まれ得る名称であれば, 粒度は問わない。また, 料理に関する辞書を利用することで, 「パスタ」を指定したときに, 自動的に「スパゲティ」というサブの料理カテゴリーも対象とすることができると思われるが, 将来の課題とする。

しかし、先行研究では意外性のある食材は提示できても、意外性のあるレシピ自体は提示できていない。食材を提示するだけでは直接意外性のあるレシピには結びつかないため、意外性のあるレシピの抽出ができることが理想であり、先行研究の問題点である。

### 3. 予備実験と課題

#### 3.1 意外というキーワードを用いたレシピ検索

意外レシピを見つけるためには、1.でも述べたように、「意外」というキーワードで検索を行うことがまず考えられる。本節では、投稿型レシピサイトの一つであるクックパッドを利用して、「意外」というキーワードでレシピ検索を行っても、意外レシピを見つけることが容易ではないことを示す。

まず、ハンバーグという料理カテゴリーにおいて、「意外」というキーワード検索を行ったところ、103件のレシピが検索された。しかし、「意外と簡単」という趣旨で使われているものもあり、レシピ自体が意外ではないものもある。筆者らが確認したところ、103件中21件は意外レシピに相当し得ると判断されたが、それ以外のレシピについては、レシピ自体が意外であるとの判断はできなかった。また、ユーザは検索結果の上位から順にレシピを見ていくものと想定されるが、103件のうち上位20件に限定すると、意外レシピに相当し得ると判断されるレシピは3件のみであった。

同様にパスタという料理カテゴリーにおいても確認を行った。「意外」というキーワード検索結果では860件のレシピが得られたが、上位100件のうち、意外レシピに相当し得ると判断されたものは7件、上位20件に限定すると3件のみであった。

#### 3.2 RF-IIFの単純な応用と課題

筆者らの先行研究[1]で提案された意外食材は、「指定された料理カテゴリーにおいて」意外であると判断されるものである。したがって、各レシピがどの料理カテゴリーに属しているかをまず判定する必要がある。

ただし、この料理カテゴリー判定を誤ると、全く意外でないレシピが意外レシピと誤判断されてしまう。例えば、ユーザがハンバーグのレシピを欲している状況を考える。一方で、レシピのデータベース上には、食材に「強力粉」を含む、パンのレシピがあるとする。このレシピは、本来は「パン」という料理カテゴリーであるはずなのに、「ハンバーグ」という料理カテゴリーに誤判定されてしまった場合、強力粉がハンバーグにとっては意外食材であるため、このレシピが意外

レシピであると判断されてしまう。しかしながら、パンという料理カテゴリーの中では全く意外ではなく、意外レシピの推薦とはならず、そもそもユーザが欲する料理カテゴリーのレシピでもない。

特定の料理カテゴリーのレシピを得ようとした際に料理カテゴリー名を含むレシピを抽出する方法が考えられる。しかし、レシピタイトルに複数の料理カテゴリー名が存在する場合もあり、本来別の料理カテゴリーに属すべきレシピも同時に抽出されてしまう可能性がある。ここで、この問題を確認するために、ユーザが指定した料理カテゴリー名を含むレシピタイトルを抽出し、各レシピ内の「食材一覧」に記されている各食材の意外度を足し合わせて意外レシピを導出する、Naive Approachを実装し、予備実験を行った。対象の料理カテゴリーを「ハンバーグ」及び「パスタ」に設定した結果を表1に示す。予備実験で利用したデータは、5.にて利用したものと同一のものである。また、4.1に示す事前処理も行っている。

ハンバーグについては、パンのレシピが多く抽出されてしまっている。これらパンのレシピについて、RF-IIFの値が大きい食材（強力粉等）は、パンのレシピであることを考えれば、意外であるとは言えなかった。パスタにおけるピザのレシピも同様である。

このように、RF-IIFを単純に応用しただけでは、別カテゴリー料理の、特に意外とは考えられないレシピが多数抽出されてしまう、という課題がある。

また、表1には現れていないが、単純にランキングを行うと類似したレシピが複数抽出される可能性が高い。ユーザの閲覧できるレシピ数には限りがあるため、提示するレシピは多様性に富んだものである方が望ま

表1 Naive approachによる意外レシピ抽出  
Table 1 Surprising recipes extracted by Naive approach.

順位	ハンバーグ	パスタ
1	ハンバーグパン&焼きカレーパン	カレーパスタ♪
2	ハムチーズロールとハンバーグパン	パスタソースでピザ♪
3	ウインナーロール(&ハンバーグパン)	ぷりぷり エビチリパスタ
4	全部に塩麹美味ハンバーグランチ	パスタソースde簡単おもてなしケーキサレ
5	かぶとごぼうの和風豆腐ハンバーグ	15分りんごケーキ♡レンジパスタ容器
6	ちょっと嬉しい！チーズハンバーグと玉子♪	パスタソースで簡単HKMピザ
7	お花のハンバーグパン	海賊冷製パスタ
8	なす入りハンバーグ(やよい会)	マ○マパスタの!?チョコバナナタルト

しく、推薦するレシピ群からの類似したレシピの除去も課題であると言える。

#### 4. 提案手法

本研究では、ある料理カテゴリーにおける食材の意外度を利用することで意外レシピを抽出するレシピ推薦エージェントを提案する。意外レシピとは、意外食材を使用しているレシピと定義する。ある料理カテゴリーにおける意外食材とは、先行研究と同様に多数の料理カテゴリーレシピにおいて出現しているが、その料理カテゴリーにおいてはあまり出現しない食材と定義する。

レシピ推薦エージェントはまず、投稿型レシピサイトの全レシピからユーザーが指定した料理カテゴリーに関するレシピを抽出する。抽出したレシピから、4.1に示すように事前処理を行って表記ゆれなどを取り除く。次に、4.2に示すように紛れている別料理カテゴリーのレシピの大部分を除去する。その後、4.3で各食材の意外度を算出した後、4.4で述べるように食材の意外度を利用しレシピの意外度を算出するとともに、4.2で取り除けなかった別料理カテゴリーのレシピの除去を試みる。最後に、4.5で述べるようにレシピ間の類似度を測り、多様性に富んだ意外レシピを抽出する。提案手法の概要を図1に示す。

なお本章の各節のうち、「4.1 事前処理」及び「4.3 食材の意外度算出」の内容は、先行研究[1]の結果を利用している。残りの「4.2 別カテゴリー除去」「4.4 更なる別カテゴリー除去を考慮したレシピの意外度算

出」「4.5 抽出レシピの多様性向上」及び、これらの処理を組合せて意外レシピ推薦を行う Web システムを開発したことが本研究の新規部分である。

##### 4.1 事前処理

全レシピから、指定された料理カテゴリー、例えばハンバーグのレシピを抽出するためにレシピタイトルに「ハンバーグ」を含むレシピを抽出する。

また、投稿型レシピサイトでは一般的に、食材や調理工程はユーザーが自由に記述することができる。したがって、同じ食材、調理工程でも様々な表記の仕方が存在する。まず、抽出した全食材に対して○や\*などの記号等を除去することで食材名のみを取得する。その後、「合い挽き肉」、「豚豚合い挽き肉」などの表記ゆれを除去する。最後に、明らかに食材ではないものを除去する<sup>(注2)</sup>。

##### 4.2 別カテゴリー除去

レシピタイトルに複数の料理カテゴリー名が存在する場合は、レシピタイトルの“最後に”出現する料理カテゴリー名を優先する、という別カテゴリー除去アルゴリズムを考案した。このアルゴリズムを適用すると、表1のハンバーグについては、「パン」が「ハンバーグ」よりもレシピタイトルの後半に出現している順位1, 2, 3, 7のレシピのように、明らかに別カテゴリーであるレシピが除外される。

一方、パスタにおける「カレーパスタ」のように、別カテゴリー名としてカレーが入っているレシピであっても、「カレー」よりも「パスタ」がレシピタイトルの後半に出現しているため、このレシピは除外されない。しかし、パスタの順位4のレシピはケーキのレシピであるため除外されるべきであるが、別カテゴリー除去アルゴリズムでは除外できない。これは、「ケーキサレ」という珍しい名前には対応できないためである。また、順位5のレシピも本来は除去されるべきである。

更にこの別カテゴリー除去アルゴリズムの精度を詳細に計測するため、クックパッドのレシピサイトから、レシピタイトルに複数の料理カテゴリー名がある100レシピをランダムに取得し、予備実験を行った。複数の料理カテゴリー名のうちどのカテゴリーに属すべきであるか判定が難しかったもの4レシピを除き、89レシピは正しく料理カテゴリーを判定できた。また、4

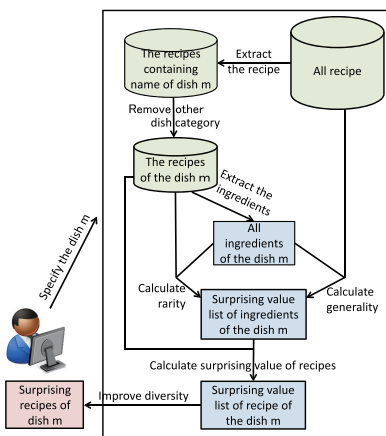


図1 提案手法概要

Fig. 1 Overview of proposed method.

(注2)：本論文の実験では手作業で実施したが、辞書等を用いることで機械的に実施することができると考えられる。



レシピは誤って除去してしまい、残り3レシピは除去すべきレシピであるが除去できなかった。誤って除去してしまうレシピは、意外レシピの推薦対象とならず、本研究の将来課題と言える。

一方、取り除けなかった別料理カテゴリーのレシピについては、全く意外でなかったとしても意外レシピとして推薦されてしまう可能性が高いため、可能な限り除去してしまう必要がある。これについては、更に4.4で提案する手法を用いることで、それらのレシピの意外度は低くなるのが期待される。

#### 4.3 食材の意外度算出

2.3において述べたRF-IIFを用いて、事前処理で得られた各食材の意外度を算出する。

#### 4.4 更なる別カテゴリー除去を考慮したレシピの意外度算出

4.2で述べたように、別カテゴリー除去アルゴリズムでは取り除けない別料理カテゴリーのレシピが存在する。食材一覧に出現する食材の意外度を合計することでレシピの意外度を算出する場合、3.で述べたとおり、別カテゴリーの、特に意外であるとは考えられないレシピが意外レシピとして推薦されてしまう。

そこで本研究では、この問題を解消するために、レシピに含まれている「投稿者によるコメント」に着目する。レシピ投稿者は、ある食材を意外食材として利用する場合その食材をアピールしようとし、投稿者によるコメントに意外食材名を記載する可能性が高いと推測される。一方、食材一覧に意外食材が含まれていたとしても別カテゴリー料理にとって意外な食材でないならば、投稿者によるコメントに出現する可能性は低いと言える。

例えば、「ハンバーグ」料理において「強力粉」は意外食材であるが、「パン」料理においては意外ではない。このため、「ハンバーグカルツォーネ」の食材一覧には強力粉が出現するが、投稿者によるコメントに出現することは少ないと推測される。つまり、概要や投稿者によるコメントに出現する食材からレシピの意外度を算出すれば、取り除き損ねた別カテゴリー料理レシピの意外度が高くなるのを防ぐことができ、結果的に別カテゴリー料理レシピの除去につながる。

そこで、各レシピの概要・投稿者によるコメントを形態素解析し、食材名が出現している場合、その食材の意外度をそのレシピの意外度に加算していくことでレシピの意外度を算出する。具体的には次のとおりである。あるレシピ $r$ において、そのレシピ中の投稿者

によるコメントに含まれる食材名の集合を $I$ と定義する。ただし、重複を許して集合を作成する。料理カテゴリー $m$ のレシピ $r$ についての意外度を $S_{r,m;p}$ と表し、式(4)と定義する。

$$S_{r,m;p} = \sum_{i \in I} RF-IIF_{i,m;p} \quad (4)$$

ユーザが料理カテゴリー $m$ として「ハンバーグ」を指定した場合における、具体例を以下に述べる。

例えば、「りんごとレンコンのハンバーグ」というタイトルのレシピ $r$ があり、投稿者コメントとして、「レンコンの食感が倍おいしい！りんごの隠し味も意外とおいしい！レンコン嫌いのお子様にもオススメです！」と記載されているものとする。このとき、式(4)において、 $I = \{\text{“りんご”}, \text{“レンコン”}, \text{“レンコン”}\}$ となる。その結果、式(4)の値は

$$S_{r,m;p} = RF-IIF_{\text{“りんご”},m;p} + 2 \times RF-IIF_{\text{“レンコン”},m;p} \quad (5)$$

と計算される。ハンバーグ料理において「りんご」や「レンコン」はRF-IIF値が高い意外食材であり、ユーザも投稿者コメントでこれらの食材をアピールしているため、式(4)の値が高くなり、意外レシピとして推薦されやすくなる。

次に、「手作りパン de ハンバーグ」というタイトルのレシピがあり、投稿者コメントとして、「ハンバーグを手作りパンで挟んだハンバーガーです。サクふわの食感がやみつきになります。」と記載されているという状況を考える。また、当該レシピの食材一覧には、「パン」の材料として「強力粉」や「ドライイースト」等が含まれていると想定する。この場合、強力粉やドライイーストは「ハンバーグ」の料理カテゴリーにおいては意外食材である(RF-IIF値が高い)。しかしながら、「パン」の材料であることを考えると意外ではない。投稿者コメントに含まれないため、式(4)による計算には「強力粉」や「ドライイースト」は出てこない。したがって、式(4)の値は高くなり、意外レシピとして推薦されにくくなる。

#### 4.5 抽出レシピの多様性向上

意外度の高いレシピの中には同じ食材を用いた類似したレシピが存在する。しかし、ユーザが閲覧できるレシピ数には限界があるため、意外レシピとして提示するレシピは多様性に富んでいる方が有用であると考える。本研究では、意外食材のみに注目して二つの

レシピ間の類似度を計算する。何故なら、意外食材以外の食材が全て一致していたとしても、意外食材が異なっていれば、意外レシピを推薦する上では有用であると考えられるからである。

料理カテゴリー  $m$  におけるレシピ  $r_a$  とレシピ  $r_b$  間の類似度をコサイン係数により計算する。具体的には  $\cos(r_a, r_b, m)$  と表し、式 (6) と定義する。ここで  $V_m$  は指定された料理カテゴリー  $m$  内で、意外度が意外度のしきい値より大きい食材の集合である。また、 $R_a$  及び  $R_b$  は、レシピ  $r_a$  及び  $r_b$  の食材一覧に含まれている食材の集合を表す。算出した類似度が「類似度のしきい値」を上回る場合、二つのレシピは類似していると判定する。この類似度は 0 から 1 までの値を取り、1 に近いほど類似していると言える。

$$\cos(r_a, r_b, m) = \frac{|R_a \cap R_b \cap V_m|}{\sqrt{|R_a \cap V_m|} \cdot \sqrt{|R_b \cap V_m|}} \quad (6)$$

この類似度算出を利用して、レシピの提示リストを作成する。まず、意外度ランク 1 位のレシピを提示リストに加える。その後、意外度ランク 2 位以降のレシピは提示リストのレシピそれぞれとの類似度を計算し、いずれかと類似している場合は除去し、全てと類似していない場合は提示リストに加える。この処理を提示リストが一定件数に達するまで続ける。提示リスト作成のアルゴリズムを Algorithm 1 に示す。

#### Algorithm 1 Make display list

```

1:  $list_{cand}$  : List of recipes sorted in descending order of
   surprising values.
2:  $list_{display}$  : List of recipes to be displayed to the user.
3:  $N$  : Number of recipes to be displayed to the user.
4:  $m$  : Target recipe category
Input:  $list_{cand}$ ,  $N$ ,  $threshold_{sim}$ ,  $m$ 
5: Add first element of  $list_{cand}$  to  $list_{display}$ 
6: for  $r_a$  in  $list_{cand}$  do
7:   for  $r_b$  in  $list_{display}$  do
8:     if  $\cos(r_a, r_b, m) > threshold_{sim}$  then
9:       Break
10:    else if  $r_b ==$  last element of  $list_{display}$  then
11:      Add  $r_a$  to  $list_{display}$ 
12:    end if
13:  end for
14: if length of  $list_{display} == N$  then
15:   Break
16: end if
17: end for
Output:  $list_{display}$ 

```

## 5. 評価実験

提案するレシピ推薦エージェントを Web サービスとして実装し、実験を行った。Web サービスのユーザインタフェースを図 2 に示す。なお著作権の関係上完成画像は実際のものとは異なる。

### 5.1 データセット

投稿型レシピサイトの中でユーザ数、レシピ数が国内トップであるクックパッドからレシピデータ 1,492,366 件、総食材数 12,826,094 件を取得した。

全レシピデータから、レシピタイトルに「ハンバーグ」を含むレシピデータ 12,327 レシピ、161,307 食材と、「パスタ」を含む 37,426 レシピ、371,797 食材を抽出した。

### 5.2 実験概要

対象とする料理カテゴリーは、比較的定番の食材が存在するハンバーグ料理と、バリエーション豊かな味



(a) Search screen.



(b) Search results screen.

図 2 Web サービスのユーザインタフェース  
Fig. 2 User interface of the Web service.

付けや食材が存在するパスタ料理を取り扱った。

提示リストのトップ8件を意外レシピとして抽出し、アンケートを実施した。アンケートでは意外レシピ全文を提示し各レシピに対して、意外性、適正性について5段階での評価を依頼した。意外性とはそのレシピが被験者にとって意外であるかどうかを示すものである。適正性は、このレシピの料理を食べたいかという質問をすることで評価した。例えばハンバーグのレシピの場合、意外性を評価する項目として「このレシピはハンバーグ料理として意外ですか?」という質問項目を、適正性を評価する項目として「このレシピの料理を食べてみたいですか?」という質問項目を設定した。更に、性別、年代、提案システムの有用性についての項目を設けた。

被験者は非専門家と専門家の二つのグループに分けた。非専門家とは日常的に料理をする人からあまりしない人まで様々な人を、専門家とは調理師、あるいは栄養士の資格を有する人を対象とした。主に筆者らの知人を通じ、ハンバーグについては非専門家20人、専門家5人のデータが、パスタについては非専門家23人、専門家5人のデータが得られた。中には未回答のレシピがある被験者もいたが、全ての回答項目において回答率は60%以上であった。

なおクックパッド社によると、2012年7月時点でクックパッド利用者のうち20代から30代が72.6%を占めているため[14]、被験者の多くをこの層に設定している(20~30代が9割)。また、クックパッド利用者の男女比は13:87であるが、被験者の男女比は6:4となった。

食材の希少度の重みパラメータ  $p$  の値は先行研究において良い結果を得た  $p = 5$  とし、類似度のしきい値は予備実験で良い結果が得られた0.4とした。

また、多様性向上における類似度算出手法の評価を行うため非専門家7人に提示リストのトップ20件のレシピを提示し、類似レシピをグループ化してもらった。二つのレシピが同一グループに含まれていればその被験者データ上では二つのレシピは類似しているとし、20件のレシピでできるペア全パターンに対して類似しているか判別した。その後、半数以上の被験者が類似していたとしたレシピペアを類似しているとみなし正解データとした。また、提案手法によってもレシピペア全パターンの類似度を算出し、類似度がしきい値を上回るレシピペアは類似していたとした(以下、算出データとする)。正解データと算出データを比較

し、適合率、再現率、F値を求めることで評価した。

なお、別カテゴリー除去アルゴリズム単独での評価は、4.2に示したとおりの結果が得られている。

### 5.3 実験結果

抽出した意外レシピトップ8のタイトルを表2に示す。Algorithm 1を適用することで除去できたレシピの例として「夏にピッタリ!おろしきゅうりハンバーグ」がある。これは「夏にぴったり!キュウリと長芋のハンバーグ」と類似しているため除去された。

各料理カテゴリーについての意外性平均、適正性平均を表3に示す。意外性平均とは意外性の項目において5段階のうち上位二つの評価をした人の割合をそのレシピの意外性とし、全意外レシピ8件の意外性の平均値を取ったものである。適正性平均についても同様である。意外性平均に関して、非専門家による評価においては5割以上であり半数以上のレシピが意外と評価されたのに対して、専門家による評価においては約2割と低い評価となった。専門家による評価が低くなったのは、専門家は非専門家に比べ食材や料理に対する知識が豊富であり、様々なレシピを知っているからであると言える。また、適正性に関しては、どれも7割程度の評価が得られていることから、抽出したレシピは有用なものであると言える。

なお、意外性については、年齢が若くなるほど高い評価を行う傾向にあった。これは食材に関する知識の

表2 意外レシピタイトル一覧  
Table 2 Titles of extracted surprising recipes.

順位	ハンバーグ	パスタ
1	夏にぴったり!キュウリと長芋のハンバーグ	簡単!トロピカルカレースープパスタ
2	♪ひとくち豆腐ハンバーグ◆旨みぎっしり◆	イソフラボン・ライスパスタ
3	私のお気に入り☆納豆ハンバーグ	米粉で簡単クリームパスタ
4	魔法の粉!?誰でもふくらハンバーグ	クリームトマトパスタ
5	簡単時短ランチ*チーズハンバーグライス	ほたて・じゃこ・ガーリックパスタ
6	ハンバーグ・和風りんごソース	☆アボカド・岩のり・納豆・たこのパスタ☆
7	おからハンバーグ	えのきでダイエット♪麻婆茄子パスタ
8	ソース要らず! ハムバーグ~ハンバーグ	簡単☆じゃこねぎ和風パスタ♪

表3 意外性平均と適正性平均  
Table 3 Results of average surprising measure values and proper measure values.

	ハンバーグ		パスタ	
	意外性平均	適正性平均	意外性平均	適正性平均
非専門家	0.60	0.66	0.56	0.77
専門家	0.25	0.69	0.15	0.65

表 4 提案システムの有用性  
Table 4 Usefulness of the proposed method.

	ハンバーグ	パスタ
非専門家	0.85	0.91
専門家	0.40	0.60

表 5 類似度算出の評価結果  
Table 5 Evaluation results of calculation similarity.

	ハンバーグ	パスタ
適合率	0.75	0.25
再現率	0.30	1.00
F値	0.43	0.40

差が影響している可能性があるが、有意差があるとまでは言えなかった。また、適正性については、男性よりも女性のほうが高い評価を行う傾向にあったが、これも有意差があるとまでは言えなかった。意外性に関する男女差や、適正性に関する年齢差はほとんど見られなかった。

提案システムの有用性項目において5段階のうち上位二つの評価をした人の割合を表4に示す。これより、特に本研究が対象とする非専門家層の値は85%以上となっており、提案システムには需要があると言える。なお、年齢差及び男女差による有用性の評価にはほとんど差が見られなかった。

多様性向上に関する算出データの適合率、再現率、F値を表5に示す。適合率、再現率はハンバーグとパスタで異なる傾向が出ているがどちらもF値は4割程度であり、食材名に基づく類似度判定の精度はあまり高くないことが分かった。

## 6. 考 察

表3について、意外性平均に関して専門家は非専門家よりも様々な食材や料理に関する知識があるため、意外だと評価しないレシピが多いと思われる。また、これに加え高い適正性平均が得られていることより、専門家は意外ではないが食べてみたいレシピ、つまり食材の組合せ等としては妥当なレシピである可能性が高いと考える。更に、非専門家が意外であり食べてみたいレシピであると評価していることを考慮すると、妥当なレシピであり、一般的には知られていないレシピが抽出できていると考える。

また、ハンバーグの意外性平均に比べパスタの意外性平均の方がやや低い評価となっている。ハンバーグは定番の食材があり、ある程度決まったレシピが存在するのに対し、パスタは様々なソースが存在し、更に

はパスタ自体の使用方法も多岐に渡る。このためパスタのレシピの方が、実際に食べたことや、調理したことがないレシピであっても意外ではないと感じる人が多い可能性がある。このことから、提案手法はある程度定番のレシピが存在する料理に対して用いる方が有効であると考えられる。

また、表4から分かるようにユーザーにおける提案システムの有用性の評価は、非専門家層では85%以上となっており多くの人是有用であると回答している。前述したようにパスタのレシピの方が意外ではないと感じる人が多い。このため、パスタの意外なレシピが存在するのであれば知りたいと思う人が多いと考える。このような期待から、パスタの方が有用性の評価が高くなったと推測される。

レシピの類似度計算に関する評価結果を表5を見ると、ハンバーグに関しては適合率が高く、再現率が低い。この要因として次のことが挙げられる。例えば、タイトルが「お豆腐ハンバーグ in ヘルシー油揚げ」のようなレシピは意外なポイントは油揚げであり、提案手法の計算上では油揚げを使用しているレシピかどうかで類似度が大きく変化する。しかし、被験者からは油揚げを使用していなくても豆腐を使用していれば類似していると判断するという意見が得られた。他にも、おからと豆腐のようにものは異なるが類似している食材が使用されていると、提案手法では類似していないレシピペアと判断するが、被験者は類似しているレシピペアと判断すると推測される。今回抽出したレシピ群においてはハンバーグの方が上記のようなレシピが含まれていたため、表5のような評価結果となったと考える。

もしレシピデータベースに十分な数の意外レシピが含まれているのであれば、提示するレシピの多様性を向上させるために、意外度の高いレシピを多少除去してしまったとしても、できるだけ多くの類似したレシピを除去できる方が良いと考える。つまり、適合率より再現率が高い方が好ましいと言える。逆に、意外レシピ数が少ない場合は、適合率が高い方が好ましいと考えられる。本論文では、指定されたカテゴリーに含まれるレシピ数を考慮しなかったが、それを考慮することで、類似度判定のしきい値を変動させ、適合率と再現率の調整ができるよう拡張することが考えられる。

## 7. 関連研究

料理のレシピに関する研究は数多く行われている。



上田ら [2] はユーザの調理履歴から TF-IDF を応用し、各食材の得点を算出している。その後、各食材の得点からレシピの得点を算出し、得点の高いものをユーザの嗜好に適したレシピとして推薦する。また、Ueda ら [3] はユーザの閲覧履歴と調理履歴から食材に対する好みの推定を行い、レシスコアを算出している。更に、類似レシピを調理した日が何日前なのかで重みをつけることで、類似レシピが連日推薦されないように考慮している。また、Maruyama ら [4] はユーザがモバイルデバイスの写真を撮って画像処理を行い、認識した食材を使用したレシピを推薦するシステムを提案している。

嗜好ではなく状況を考慮したレシピ推薦システムも存在する。Yajima ら [5] は、個人の所持する調味料や食材、嗜好を入力することで各レシピの食材や調理方法から調理難易度を算出する。そして、スケジュールを入力することで現在の調理に充てられる時間に適したレシピを推薦する。また、赤澤ら [6] は冷蔵庫内の食材の分量、賞味期限を入力することで、できる限り冷蔵庫内の食材を使いきれようようなレシピを推薦している。

多川ら [7] は食事療法支援を行う推薦システムを提案している。レシピ情報と日本食品標準成分表を Linked Data 化することでレシピの栄養計算を自動で行う。飲食店のメニュー名を入力することでレシピタイトルが類似しているレシピの栄養計算結果を取得し、飲食店メニューの栄養成分を推定する。このシステムにユーザに年齢や体重、運動量などを登録してもらうことで、栄養成分として適切なメニューであるかどうかを判断し、食事療法支援を行っている。また、van Pinxteren ら [8] は健康的なレシピの推薦システムを提案している。レシピテキストから重要な特徴を識別し、特徴ベクトルを作成し、レシピ間の類似度を算出する。類似レシピが健康指標的に高いものであれば、そちらに変更することで健康的な食事を行うことができるようにしている。

レシピにおける代替食材についての研究も存在する。唐沢ら [9] はレシピを分類し TF-IDF を適用することで各レシピ群の食材、調理動作の特徴度を算出する。特徴度の低い食材を一般食材としそのレシピ群の頻出一般食材同士を代替可能な食材であるとして抽出している。また、Shidochi ら [10] は各レシピテキストから食材とその調理動作を抽出し、TF-IDF を用いて各食材の特徴的な調理動作を取得する。更に同一料理カテゴリ内における食材の特徴的な調理動作と合わせて

食材の調理動作ベクトルを作成する。そして、コサイン類似度の高い調理動作ベクトルをもつ食材を代替可能食材として抽出している。

Fang ら [11] は晩御飯などのメニュー（複数のレシピの集合）を生成するシステムを提案している。

また、近年の情報推薦の分野では、個人の嗜好に適した情報を推薦する手法は必ずしもユーザの満足度を向上させるとは限らないと指摘されており、ユーザにおける新規性や意外性などが注目されている [12], [13]。しかし、これらの手法はユーザ情報を利用し、ユーザが閲覧したことのない情報や、興味から外れているジャンルの情報を推薦するため、本研究とは目的が異なり利用できない。

このように多様で多くの料理レシピに関する研究、情報推薦に関する研究が存在するが、レシピの意外性や食材の意外性に注目したものはなく、本研究とは目的、手法ともに異なっている。

## 8. む す び

本研究では投稿型レシピサイトに対して、意外食材の出現頻度からレシピの意外度を算出し、更にレシピ間の類似度を測ることで、多様性に富んだ意外レシピを抽出するレシピ推薦エージェントを提案した。

レシピ推薦エージェントはまず、指定された料理カテゴリ名をタイトルに含むレシピを抽出し、別料理カテゴリレシピの除去を行う。次に RF-IIF を用いて、指定料理カテゴリにおける各食材の意外度を算出した後、レシピの投稿者によるコメントに出現する食材の意外度を合計することで、各レシピの意外度を算出する。これにより、取り除き損ねていた別料理カテゴリレシピの除去を行い、意外食材をアピールしている意外レシピの抽出が可能となった。また、レシピ内の意外食材を用いてレシピ間の類似度を測り、類似レシピの除去を行うことで、多様性に富んだレシピをユーザに提示する。

以上のようにして得られたレシピの意外度の上位 8 件のレシピに対して評価実験を行った。その結果、提案手法によって抽出した意外レシピは、約 7 割が妥当なレシピであり、非専門家にとって半数以上が意外性のあるレシピであることを示した。

今後は、手元にある食材が限られている状況の中でも活用することができるように、料理カテゴリを指定するのではなく食材名を指定することで、意外レシピが抽出できるエージェントへの改良について検討し

ていく。

また、一つのレシピが複数のサブレシピから構成されている場合（例えばハンバーグパン）、複数のサブレシピに分解し、各サブレシピにおける意外性を計算するアルゴリズムを提案することを考えている。これにより、意外レシピ抽出の精度が向上する可能性がある。

**謝辞** 本研究は JSPS 科研費 24300005, 26330081, 26870201 の助成を受けたものです。本研究を遂行するにあたり、レシピ投稿・検索サイトであるクックパッドのレシピデータを提供して頂いたクックパッド株式会社様に感謝致します。

## 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,” 日本ソフトウェア科学会「コンピュータソフトウェア」, vol.31, no.3, pp.70-78, 2013.
- [2] 上田真由美, 石原和幸, 平野 靖, 梶田将司, 間瀬健二, “食材利用履歴に基づき個人の嗜好を反映するレシピ推薦手法,” 日本データベース学会 Letters, vol.6, no.4, pp.29-32, 2008.
- [3] M. Ueda, M. Takahata, and S. Nakajima, “User’s food preference extraction for personalized cooking recipe recommendation,” Workshop of ISWC, pp.98-105, 2011.
- [4] T. Maruyama, Y. Kawano, and K. Yanai, “Real-time mobile recipe recommendation system using food ingredient recognition,” Proc. IMMPD, pp.27-34, 2012.
- [5] A. Yajima and I. Kobayashi, ““Easy” cooking recipe recommendation considering user’s conditions,” Web Intelligence/IAT Workshops, pp.13-16, IEEE, 2009.
- [6] 赤澤康之, 宮森 恒, “冷蔵庫食材を考慮した料理レシピ検索システムの提案,” 第 3 回データ工学と情報マネジメントに関するフォーラム, E3-2, 2011.
- [7] 多川勇介, 田中 改, 南 裕也, 並河大地, 下村道夫, 山口高平, “レシピと食品標準成分表を連携させた食事療法支援サービス,” 第 27 回人工知能学会全国大会, 1N3-OS-10a-5, 2013.
- [8] Y. van Pinxteren, G. Geleijnse, and P. Kamsteeg, “Deriving a recipe similarity measure for recommending healthful meals,” Proc. IUI, pp.105-114, 2011.
- [9] 唐沢 隆, 浜田玲子, 井手一郎, 坂井修一, 田中英彦, “料理教材テキストからの素材と調理法に関する知識の抽出,” 第 66 回情報処理学会全国大会講演論文集, vol.2, pp.119-120, 2004.
- [10] Y. Shidochi, T. Takahashi, I. Ide, and H. Murase, “Finding replaceable materials in cooking recipe texts considering characteristic cooking actions,” Proc. MM, pp.9-14, 2009.
- [11] F.-F. Kuo, C.-T. Li, M.-K. Shan, and S.-Y. Lee, “Intelligent menu planning: Recommending set of recipes by ingredients,” Proc. MM, pp.1-6, 2012.
- [12] T. Murakami, K. Mori, and R. Orihara, “Metrics for evaluating the serendipity of recommendation lists,” Proc. JSAI, pp.40-46, 2007.
- [13] Y.C. Zhang, D.Ó. Séaghdha, D. Quercia, and T. Jambor, “Auralist: Introducing serendipity into music recommendation,” Proc. WSDM, pp.13-22, 2012.
- [14] クックパッド株式会社, 2013 年 4 月期第 1 四半期決算補足説明資料, 2012.  
(平成 26 年 8 月 27 日受付, 11 月 18 日再受付, 27 年 3 月 5 日早期公開)



池尻 恭介

1988 年生。2014 年電気通信大学大学院情報システム学研究科博士前期課程修了。同年 NTT ソフトウェア (株) 入社, 現在に至る。データマイニング技術の研究に従事。



清 雄一 (正員)

1981 年生。2009 年東京大学大学院情報理工学系研究科博士後期課程修了。同年 (株) 三菱総合研究所入社。同社情報技術研究センター, 金融ソリューション本部等に所属。2013 年より電気通信大学助教, 現在に至る。分散コンピューティング, セキュリティ, プライバシー保護技術等の研究に従事。情報処理学会, IEEE Computer Society 各会員。



中川 博之 (正員)

1974 年生。1997 年大阪大学基礎工学部情報工学科卒業。同年鹿島建設 (株) に入社。2007 年東京大学大学院情報理工学系研究科修士課程修了, 2008 年同大学院博士課程中退。同年電気通信大学助教, 2014 年大阪大学大学院情報科学研究科准教授, 現在に至る。博士 (工学) (早稲田大学)。要求分析, 形式手法, エージェント及び自己適応システム開発手法の研究に従事。情報処理学会, IEEE CS 各会員。



田原 康之

1966 年生。1991 年東京大学大学院理学系研究科数学専攻修士課程修了。同年 (株) 東芝入社。2003 年国立情報学研究所着任。2008 年より電気通信大学大学院情報システム学研究科准教授。博士 (情報科学) (早稲田大学)。エージェント技術, 及びソフトウェア工学などの研究に従事。情報処理学会, 日本ソフトウェア科学会各会員。



大須賀昭彦 (正員)

1958年生. 1981年上智大学理工学部数学科卒. 同年(株)東芝入社. 同社研究開発センター, ソフトウェア技術センター等に所属. 1985~1989年(財)新世代コンピュータ技術開発機構(ICOT)出向. 2007年より, 電気通信大学大学院情報システム学研究科教授. 2012年より, 国立情報学研究所客員教授兼任. 工学博士(早稲田大学). 主としてソフトウェアのためのフォーマルメソッド, エージェント技術の研究に従事. 1986年度情報処理学会論文賞受賞. IEEE Computer Society Japan Chapter Chair, 人工知能学会理事, 日本ソフトウェア科学会理事を歴任. 情報処理学会, 人工知能学会, 日本ソフトウェア科学会, IEEE Computer Society 各会員.