

希少性と一般性に基づいた意外性のある食材の抽出

池尻 恭介 清 雄一 中川 博之 田原 康之 大須賀 昭彦

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。しかし、それらは「意外」というキーワード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普通とは多少異なる意外性のある食材を抽出する。具体的には、TF-IDF の考えを応用した RF-IIF (Recipe Frequency-Inverse Ingredient Frequency) を利用し、食材の希少度と一般度から意外度を算出する。評価実験では、意外度の高い食材 20 件を意外性のある食材として抽出し、被験者複数名に抽出した食材に関するアンケートを実施した結果から、提案手法の有用性を示した。

Many surprising recipes exist in the user-generated recipe site. The simplest way to find surprising recipe is to use a search function. However, the title of its recipe does not always contain the keyword “surprise”. Thus we cannot find surprising recipes in an easy way. In this paper, we propose a system to extract surprising ingredients as the first step of extracting surprising recipes from the recipe site. We propose RF-IIF (Recipe Frequency-Inverse Ingredient Frequency) based on TF-IDF in our system. RF-IIF calculates a surprising value of the ingredient about the meal based on the generality and the rarity of the ingredient. We extract ingredients whose RF-IIF ranks are in top 20 as surprising ingredients about the meal. Through questionnaires to evaluate the extracted ingredients, we verified the effectiveness of the proposed method which extracts surprising ingredients.

1 はじめに

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は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り、多くのユーザが料理をする際に利

用している。

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少し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数多く存在している。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」というワードを含むレシピばかりではないため、検索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意外性のあるレシピを抽出するための先駆けとして、意外食材を抽出することを目的とする。

本研究の提案手法では、TF-IDF の考えを応用した RF-IIF(Recipe Frequency-Inverse Ingredient

Surprising Ingredient Extraction based on Rarity and Generality.

Kyosuke Ikejiri, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga, 電気通信大学大学院情報システム学研究科, Graduate School of Information Systems, The University of Electro-Communications.

Hiroyuki Nakagawa, 大阪大学大学院情報科学研究科, Graduate School of Information Science and Technology, Osaka University.

コンピュータソフトウェア, Vol.31, No.3(2014), pp.70-78.

[研究論文] 2013 年 7 月 23 日受付.

Frequency) を導入し、食材の意外度を算出することで、ある料理における意外食材の抽出を行なう。評価実験では、被験者複数名に抽出した意外食材に関するアンケートを実施する。

以下、2章では想定シナリオについて記述する。3章、4章、5章では提案手法についての説明、提案手法の評価実験、考察について記述する。6章では関連研究について記述し、最後に7章ではまとめと今後の課題について記述する。

2 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。このような状況において、ある料理における普段のレシピにはあまり出現しないような意外な食材を閲覧することが出来れば、満足度は向上すると考える。

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。そこで、本研究の提案システムを利用すれば、くるみ、きゅうり、ヨーグルトなど食感の変化や清涼感が得られそうな食材が提示される。また、パスタ料理であれば、いちご、チョコレート、こんにゃくなどデザート感覚なものや、低カロリーなものが出来そうな食材など、一般的な食材だが普段その料理にはあまり利用しないような意外性のある食材が提示される。

また、本研究の提案手法ではユーザーが料理名を指定することを想定している。したがって、普段よく調理している料理における、いつもとは少し異なる食材を閲覧することが可能である。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、有用であると考える。

3 提案手法

本研究では、意外性のあるレシピを抽出する先駆けとして、意外食材を抽出する。意外性のあるレシピとは、意外食材を使用しているレシピや、一般的な調理工程とは異なるレシピのことであるが、本研究では意外食材を使用しているレシピに注目する。意外食材と

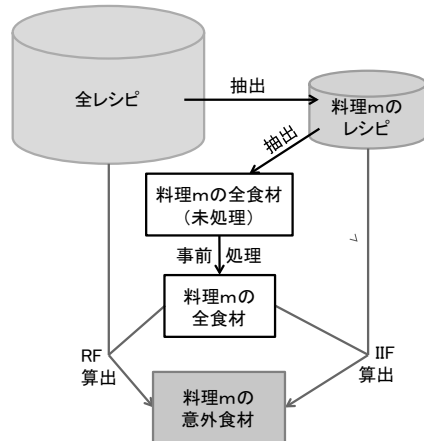


図1 提案手法概要

は、多数の料理レシピにおいて出現しているが、ある料理においてはあまり出現しない食材と定義する。

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に関するレシピを抽出し、その料理に関するレシピの全食材を抽出する。次に、抽出した全食材に対して3.1に示す事前処理を行う。その後、3.2で述べるように各食材の意外度を算出することで意外食材を抽出する。提案手法の概要を図1に示す。

3.1 事前処理

投稿型レシピサイトでは一般的に、形式は固定されているが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、(みじん切り)や(大きめ)などの括弧書き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」、「牛豚合挽き肉」などの表記ゆれを除去する。しかし、同じ食材を表していても表記の仕方によって同一食材とすべき場合と異種食材とすべき場合がある食材が存在する。このため、単純に表記を統一することができない。本研究では手作業で表記ゆれと見なすべきか否か判断し、処理する。最後に、明らかに食材ではないものを手作業で判別し、除去する。

3.2 意外食材の抽出

ある料理 m (例:ハンバーグ) のレシピについて, 出現頻度の低い食材は料理 m において意外な食材である可能性があるが, 出現頻度の著しく低い食材はスーパーなどでは手に入らないような希少な食材である可能性やユーザが全く調理方法を知らない食材である可能性もある. 以上のことから, 比較的一般的な食材を抽出するために, 出現頻度の低い食材の中で, さらにレシピ全体での出現頻度の高い食材を料理 m における意外食材とする. これによって, 料理 m 以外では普通に使用される一般的な食材であるが, 料理 m においてはあまり利用されていない食材, つまり料理 m における意外食材の抽出が可能となる.

3.2.1 食材の意外度

文書内の特徴語を得る際に用いられる TF-IDF(Term Frequency-Inverse Document Frequency) という技術が存在する. 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. 本研究における, 意外食材の定義は「多数の料理において出現しているが, ある料理においてはあまり出現しない食材」であり, 前者は DF と, 後者は TF の逆数と似ている. しかし, TF-IDF とは逆数をとる対象が異なる. そこで, 本研究では意外食材を抽出するために TF-IDF の考えを応用した, RF-IIF(Recipe Frequency-Inverse Ingredient Frequency) を提案し, 各食材の意外度を算出する. $RF-IIF_{i,m;p}$ (式 1) は食材 i について, 料理 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 と, ある料理 m における食材の希少度を示す $IIF_{i,m}$ から算出する. 変数の一覧を表 1 に示す. また, 抽出する食材が最も意外な食材となるような RF と IIF の最適なバランスが存在すると考える. そこで, 最適なバランスを取得するために希少度に重みパラメータ p を設置した. 希少度を p 乗する設計になっており, p の値を小さくすればより一般的な食材を, p の

表 1 RF-IIF の変数一覧

| 変数 | 意味 | 説明 |
|--------|-------|--------------------------------|
| RF-IIF | 意外度 | 食材 i の一般的かつ料理 m における希少な度合い |
| RF | 一般度 | 食材 i の全レシピにおける一般性 |
| IIF | 希少度 | 料理 m における食材 i の珍しさ |
| i | 食材 | 例: 挽き肉 |
| m | 料理 | 例: ハンバーグ |
| p | パラメータ | IIF の重みを調節する |

値を大きくすればより希少な食材を抽出することが可能となる.

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

3.2.2 食材の一般度

食材 i が食材として一般的であるかどうかを RF_i として式 2 によって算出する.

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

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る. つまり, RF_i の値が大きい程, 食材 i が一般的な食材であると言える.

3.2.3 食材の希少度

食材 i が料理 m において珍しい食材であるかどうかを $IIF_{i,m}$ として式 3 によって算出する.

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

ここで R_m は料理 m のレシピ数, $R_{i,m}$ は料理 m のレシピの中で食材 i が出現するレシピ数である. つまり $IIF_{i,m}$ の値が大きい程, 食材 i は料理 m において希少な食材であると言える.

4 評価実験

4.1 実験概要

提案手法によって得られる意外食材の妥当性を評価し, 提案手法の有効性を示すために評価実験を行った. 投稿型レシピサイトであるクックパッド [1] から提案手法を用いてハンバーグとパスタに関する意外食材を抽出した. このとき料理 m における食材 i の希少度である $IIF_{i,m}^p$ の重みパラメータ p の値を 3 段階変化させた. p の値は, あらかじめ実施した予備実験において, 良い結果を得た $p=1, 3, 5$ とする.

抽出した意外食材の妥当性を評価するために, 被験者に対してアンケートを実施した. 料理名, 抽出した

表 2 アンケート項目一覧

| アンケート項目 | 選択肢 |
|------------|----------------------------------|
| 食材の希少性 | 希少ではない、 どちらとも言えない、 希少である |
| 食材の一般性 | 一般的ではない、 どちらとも言えない、 一般的である |
| 性別 | 男、女 |
| 年代 | 10代、20代、30代、40代以上 |
| 料理頻度 | ほとんどしない、あまりしない、 する、とてもする |
| 提案システムの有用性 | 感じない、 どちらとも言えない、 感じる |

食材、手作業で抽出した食材の使用方法を提示し、各食材について希少性と一般性の2つの評価指標について3段階での評価を依頼した。実際のレシピを提示しなかったのは、本実験の目的が意外レシピではなく、意外食材の妥当性の検証であるからである。希少性とは料理 m における食材 i の珍しさを示し、一般性とは食材 i 自体の知名度を示すものとした。また、アンケート項目には食材についての希少性と一般性の他に、被験者の性別、年代、料理頻度、抽出された食材を踏まえた上での提案システムの有用性、自由記述欄を設けた。アンケート項目の一覧を表2に示す。

被験者は非専門家と専門家の2つのカテゴリに分けた。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対象とし、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。ハンバーグについては非専門家のうち料理頻度項目に「する」「とてもする」と回答した18人、専門家20人のデータを、パスタについては非専門家のうち料理頻度項目に「する」「とてもする」と回答した19人、専門家18人のデータを解析対象とした。

男女比は3:7であり、20代と30代が8割を占めた。クックパッド株式会社の調査によると、クックパッド利用者は女性が9割で、20代と30代が7割を占めているため、被験者の男女比、年代は妥当であると言える。また、サンプル数に関して、情報推薦や要約文生成などのデータマイニング分野においてアンケート評価の被験者数は多くの論文で10人程度~20人程度に設定されている[2]-[4]。本研究の被験者数は20人程度であり、妥当な人数であると言える。

4.2 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。その中から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付けや食材が存在するパスタ料理を取り扱った。

全レシピデータから、レシピ名に「ハンバーグ」を含むレシピデータ12,327件、食材数161,307件と、「パスタ」を含むレシピデータ37,426件、食材数371,797件を抽出した。その後、提案手法によって食材の表記ゆれを除去し、各料理の各食材の意外度を算出した。得られた意外度の上位20件の食材をその料理の意外食材とした。

4.3 結果

提案手法によって抽出した意外食材を重みパラメータ p の値別に表3に示す。 p の値によって抽出される食材が変化していることが分かる。ハンバーグにおいてはきゅうりやさつまいもなど、パスタにおいてはバナナやこんにやくなどが得られ、各料理の一般的には使用しないと思われる食材を抽出することができたと言える。

アンケートの結果を解析した。各 p についての希少性平均、一般性平均、意外性平均を表4に示す。希少性平均とは、各意外食材について希少性の項目に「希少である」と回答した人の割合をその食材の希少性とし、全意外食材の希少性の平均値を取ったものである。一般性平均とは各意外食材について一般性の項目に「一般的である」と回答した人の割合をその食材の一般性とし、全意外食材の一般性の平均値を取ったものである。意外性平均とは、各意外食材について希少性の項目に「希少である」と回答し、なおかつ一般性の項目に「一般的である」と回答した人の割合をその食材の意外性とし、全意外食材の意外性の平均値を取ったものである。

希少性に関して、非専門家の方が高い評価を得ており、 $p=1$ の場合においては3割程度と低い評価であるが、 $p=5$ の場合においては7割と高い評価が得られた。一般性に関して、専門家の方が高い評価を得て

表 3 抽出した意外食材一覧

| (a) ハンバーグ | | | | (b) パスタ | | | |
|-----------|--------------|----------------------------|----------------------------|---------|--------------|----------------------------|----------------------------|
| 順位 | RF*(IIF) 食材名 | RF*(IIF ³) 食材名 | RF*(IIF ⁵) 食材名 | 順位 | RF*(IIF) 食材名 | RF*(IIF ³) 食材名 | RF*(IIF ⁵) 食材名 |
| 1 | 薄力粉 | ベーキングパウダー | ベーキングパウダー | 1 | ご飯 | ベーキングパウダー | パニラエッセンス |
| 2 | ごま油 | 強力粉 | ドライイースト | 2 | 片栗粉 | パニラエッセンス | ベーキングパウダー |
| 3 | ご飯 | きゅうり | 強力粉 | 3 | ベーキングパウダー | ホットケーキミックス | ホットケーキミックス |
| 4 | オリーブオイル | ドライイースト | 鶏もも肉 | 4 | 薄力粉 | バナナ | バナナ |
| 5 | きゅうり | 薄力粉 | きゅうり | 5 | しょうが | ココア | ココア |
| 6 | レモン汁 | 鶏もも肉 | 油揚げ | 6 | 豆腐 | ご飯 | いちご |
| 7 | 強力粉 | 無塩バター | 無塩バター | 7 | ホットケーキミックス | ドライイースト | チョコレート |
| 8 | ベーキングパウダー | レモン汁 | スキムミルク | 8 | 酢 | チョコレート | 春巻きの皮 |
| 9 | 生クリーム | ご飯 | ちくわ | 9 | じゃがいも | 卵白 | ドライイースト |
| 10 | 酢 | 油揚げ | ハム | 10 | ドライイースト | 食パン | 白玉粉 |
| 11 | 小麦粉 | ハム | 揚げ油 | 11 | パニラエッセンス | いちご | 黒糖 |
| 12 | じゃがいも | スキムミルク | さつまいも | 12 | 強力粉 | シナモン | 卵白 |
| 13 | ドライイースト | 揚げ油 | りんご | 13 | はちみつ | 春巻きの皮 | 天ぷら粉 |
| 14 | ベーコン | パスタ | 薄力粉 | 14 | バナナ | こんにやく | ご飯 |
| 15 | だし汁 | 豚肉 | パスタ | 15 | だし汁 | りんご | シナモン |
| 16 | はちみつ | ちくわ | 豆板醤 | 16 | きゅうり | 薄力粉 | ショートニング |
| 17 | トマト缶 | ベーコン | かつお節 | 17 | ココア | 白玉粉 | こんにやく |
| 18 | ネギ | クリームチーズ | 明太子 | 18 | 食パン | 揚げ油 | 抹茶 |
| 19 | パスタ | さつまいも | クリームチーズ | 19 | パン粉 | 豆腐 | きな粉 |
| 20 | 豚肉 | 生クリーム | プレーンヨーグルト | 20 | 味噌 | 黒糖 | あんこ |

表 4 希少性平均と一般性平均および意外性平均

| (a) ハンバーグ | | | | (b) パスタ | | | | | | | | | | | |
|-----------|------|------|------|---------|------|------|------|------|------|------|------|-----|------|------|------|
| 非専門家 | | | | 専門家 | | | | 非専門家 | | | | 専門家 | | | |
| p | 希少性 | 一般性 | 意外性 | p | 希少性 | 一般性 | 意外性 | p | 希少性 | 一般性 | 意外性 | p | 希少性 | 一般性 | 意外性 |
| 1 | 0.32 | 0.85 | 0.24 | 1 | 0.25 | 0.90 | 0.22 | 1 | 0.41 | 0.84 | 0.31 | 1 | 0.39 | 0.91 | 0.34 |
| 3 | 0.49 | 0.78 | 0.36 | 3 | 0.45 | 0.84 | 0.37 | 2 | 0.61 | 0.74 | 0.40 | 3 | 0.60 | 0.85 | 0.50 |
| 5 | 0.58 | 0.74 | 0.43 | 5 | 0.52 | 0.82 | 0.37 | 3 | 0.70 | 0.68 | 0.44 | 5 | 0.70 | 0.77 | 0.55 |

おり、 $p=1$ の場合においては 8 割以上、 $p=5$ の場合においても 6.8 割以上であり、抽出した食材は一般的な食材であると言える。また、全てのパターンに関して、 p が大きくなると希少性が向上し、一般性が低下しているのが分かる。これらより、重みパラメータを用いた希少度の重み調節は有効であると言える。

$p=5$ における、上位 x 位までの意外食材による希少性平均、一般性平均、意外性平均をグラフ化したものを図 2 に示す。例えば $x=5$ のときは、意外度ランクが上位 5 位までの意外食材に対する平均値である。希少性平均は右肩下がりの傾向が見られ、一般性平均は右肩上がりの傾向が見られる。また、専門家に関するグラフについて、意外性平均は右肩下がりの傾向が見られ、算出した意外度は実際の食材の意外性に相関があることが分かる。

提案システムの有用性項目に「感じる」と答えた人の割合を表 5 に示す。これより 5 割以上の方が有用であると回答していることが分かる。

また、自由記述欄からは「一般的な食材が代用品になるならありがたい」や、「新商品開発のアイデア探

表 5 提案システムの有用性

| | 有用性があると答えた人の割合 |
|-------|----------------|
| ハンバーグ | 0.75 |
| パスタ | 0.50 |

索に役立ちそう」、「意外なレシピであってもカロリーや栄養の偏りがあれば作らない」や、「食材よりも、調理方法に意外性を受けた」などのコメントを得た。

5 考察

表 4 に関して、希少性の評価は非専門家の方が高く、一般性の評価は専門家の方が高くなる傾向となったのは、専門家は非専門家に比べ食材に対する知識が豊富であり、様々な食材の様々な調理方法を知っているからであると言える。また、意外性の評価について、パスタに比べハンバーグの方が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、さらにはパスタ自体の使用法も多岐に渡る。このため、投稿型レシピサイトにおいてパ

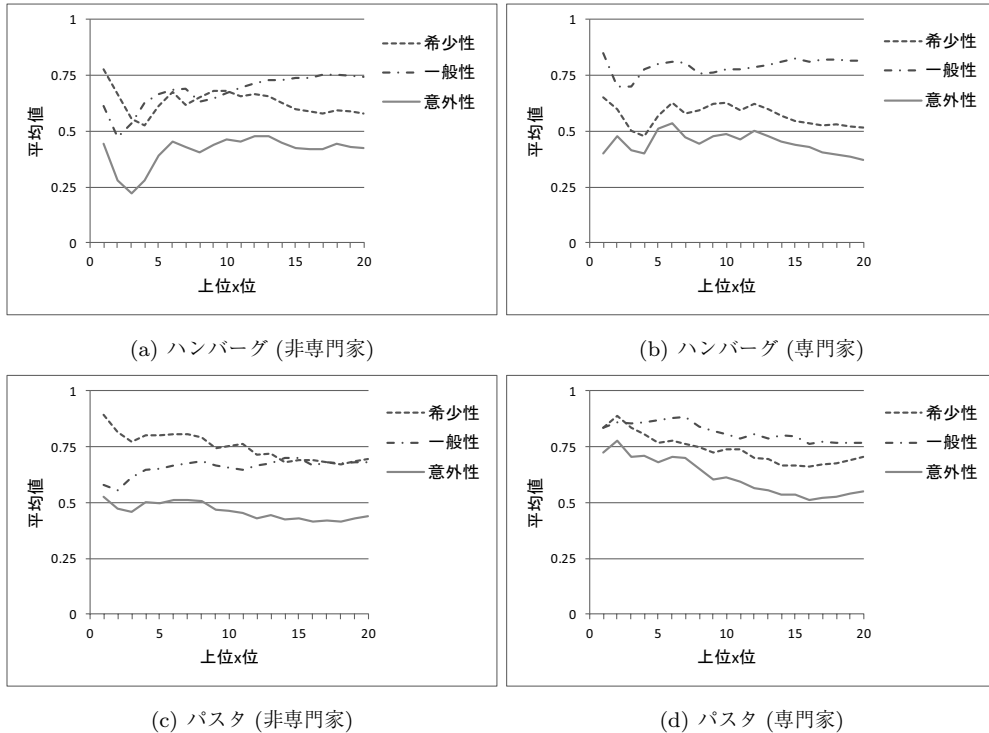


図 2 上位 x 位までの希少性平均, 一般性平均, 意外性平均

スタの方が様々なレシピが存在し、意外性のあるレシピが多く存在していると考えられる。したがって、意外な食材の数が多く、提案手法によって抽出できる可能性が高くなり、評価値が高くなったと思われる。このことから、提案手法はある程度バリエーションの豊富な料理に対して用いる方が有効であると推測される。

さらに、表 5 から分かるように有用性の評価は 5 割以上であり提案システムは有用であることが分かる。この有用性は意外食材の抽出精度を向上することでさらに向上すると推測される。

また、提案システムは料理名をインプットすることで意外な食材が提示されるシステムである。つまり、冷蔵庫にある食材でできる料理が知りたいという状況には対応できない。このため、献立に困っているという状況においては有用性が低下する可能性がある。したがって、インプットを料理名ではなく食材名とし、意外なレシピが検索できるようなシステムに改良していくことで、さらに有用性が向上する可能性があると考えられる。

図 2 に関して、非専門家は専門家に比べ、食材に対する知識が個人個人によって大きく異なると思われる。したがって、非専門家による評価において、ひとつの食材に対して複数の非専門家から同一の評価を得るのは難しく、右肩下がり傾向があまり得られなかったと考える。

意外性のあるレシピの投稿者は意外食材がレシピのポイントと考え、意外食材名をレシピのタイトルや説明文にも記述する可能性が高いと推測される。このため、意外食材は意外性のあるレシピにおいて、レシピのタイトルや説明文などにもその食材名が出現する傾向が見られる。したがって、食材一覧のみにおける出現数をカウントするのではなく、ひとつのレシピ内全体での出現回数をカウントすることで、ハンバーグなどの定番レシピが存在する料理においても意外食材の抽出精度を向上することができると考える。

6 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。

ユーザは個々に様々な食材の嗜好を持っており、同一のものを推薦するのでは満足な結果は得られない。そこで、個人の嗜好を考慮したレシピ推薦システムが存在する。上田ら [5] 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、高畑ら [6] はユーザの食材に対する好き嫌いを考慮した料理レシピの推薦システムを提案している。ユーザの閲覧履歴と調理履歴から食材に対する好きの推定、嫌いの推定を行い、レシビスコアを算出している。さらに、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。

状況を考慮したレシピ推薦システムも存在する。矢嶋ら [7] 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 [8] は冷蔵庫内の食材の分量、賞味期限を入力することで、出来る限り冷蔵庫内の食材を使いきれようなレシピを推薦している。買い足しがある場合には、買い足した食材がどの程度余るかをレシピの分量から算出し、次回推薦時には入力動作を必要としないようにしている。

多川ら [9] は食事療法支援を行う推薦システムを提案している。レシピ情報と日本標準食品成分表を 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名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別し、食事療法支援を行なっている。また、Youri ら [10] 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する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 [11] 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、そのレシピ群の類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、志土地ら [12][13] 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。さら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持つ食材を代替可能食材として抽出している。

Fang ら [14] は晩御飯などのメニュー（複数のメニューの集合）を生成するシステムを提案している。各レシピの使用されている食材からレシピ間の類似度を算出する。また、おすすめメニューの投稿が可能なサイトを利用し、レシピの共起関係を取得する。これらからユーザの所持する食材を入力することで、最適なレシピ群を持つメニューを生成している。

また、近年の情報推薦の分野では、個人の嗜好に適合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 [15][16]。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、手法共に異なっている。

7 まとめと今後の課題

本研究ではレシピにおける食材の出現頻度からある料理における意外食材を抽出するシステムを提案した。食材の、ある料理における出現頻度である希少度と全レシピにおける出現頻度である一般度から、RF-IIF を用いることによって各食材の意外度を算出し、上位 20 食材を意外食材とした。このとき、IIF の重みパラメータ p を変化させることによって抽出

精度の向上を図った。評価実験は非専門家と専門家を対象とし、抽出した各食材の希少性と一般性の評価を依頼した。その結果、 p を大きくすると、抽出した意外食材において、希少性があり一般性もあると評価される食材数が増加することを示した。また、算出した意外度が実際の食材の意外性と相関があることを示し、提案手法の RF-IIF が妥当であることを示した。

今後は、意外食材の組合せを考慮した意外レシピの抽出をし、さらにそのレシピのおいしさを自動評価するアルゴリズムを構築する。また、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、料理名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ようなシステムへの改良や、表記ゆれ除去の自動化について検討していく。

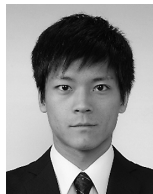
謝辞 本研究は JSPS 研究費 24300005, 23500039, 25730038 の助成を受けたものです。

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。

本研究を遂行するにあたり、研究の機会と議論・研鑽の場を提供して頂き、ご指導頂いた国立情報学研究所/東京大学本位田真一教授をはじめ、活発な議論と貴重なご意見を頂いた研究グループの皆様には感謝致します。

参考文献

- [1] クックパッド株式会社：クックパッド, <http://cookpad.com>.
- [2] Nichols, J., Mahmud, J. and Drews, C.: Summarizing Sporting Events Using Twitter, in *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, 2012, pp. 189–198.
- [3] 西川仁, 長谷川隆明, 松尾義博, 菊井玄一郎: 文の選択と順序付けを同時に行う評価文書要約モデル, *人工知能学会論文誌*, Vol. 28 (2013), pp. 88–99.
- [4] 村上知子, 吉岡信和, 折原良平, 古川康一: CAM 法を用いた個人嗜好モデルに基づく商品推薦システム, *人工知能学会論文誌*, Vol. 20 (2005), pp. 346–355.
- [5] 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 (2008), pp. 29–32.
- [6] 高畑麻理, 上田真由美, 中島伸介: 食材に対する好き嫌いを考慮した料理レシピ推薦手法の提案, 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, 2011, E3-5.
- [7] 矢嶋亜紗美, 小林一郎: 個人の状況を考慮した“かんたん”なレシピの推薦, *ファジィシステムシンポジウム講演論文集*, 25th, 2009, 1C1-01.
- [8] 赤澤康幸, 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, *データ工学と情報マネジメントに関するフォーラム*, 2011, E3-2.
- [9] 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, 2013, 1N3-OS-10a-5.
- [10] van Pinxteren, Y., Geleijnse, G. and Kamsteeg, P.: Deriving a Recipe Similarity Measure for Recommending Healthful Meals, in *Proceedings of the 16th international conference on Intelligent user interfaces*, 2011, pp. 105–114.
- [11] 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, *情報処理学会全国大会講演論文集*, 66th, 2004, pp. 119–120.
- [12] Shidochi, Y., Takahashi, T., Ide, I. and Murase, H.: Finding Replaceable Materials in Cooking Recipe Texts Considering Characteristic Cooking Actions, in *Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities*, 2009, pp. 9–14.
- [13] 志土地由香, 井手一郎, 高橋友和, 村瀬洋: 料理レシピマイニングによる代替可能食材の発見, *電子情報通信学会論文誌 A*, Vol. J94-A, No. 7 (2011), pp. 532–535.
- [14] Kuo, F.-F., Li, C.-T., Shan, M.-K. and Lee, S.-Y.: Intelligent menu planning: recommending set of recipes by ingredients, in *Proceedings of the ACM multimedia 2012 workshop on Multimedia for cooking and eating activities*, 2012, pp. 1–6.
- [15] Murakami, T., Mori, K. and Orihara, R.: Metrics for Evaluating the Serendipity of Recommendation Lists, *New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science*, Vol. 4914, 2008, pp. 40–46.
- [16] 福島良典, 大澤幸生: ソーシャルメディアを利用したセレンディピティな情報推薦, *人工知能学会全国大会*, 2012, E1-R-6-6.



池尻 恭介

1988 年生。2012 電気通信大学電気通信学部卒業。同年より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在学中。データマイニング技術の研究に従事。



清 雄一

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大学院情報システム学研究科助教、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、電子情報通信学会、IEEE Computer Society 各会員。



中川 博之

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、エージェントおよ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、電子情報通信学会、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、およ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

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、電子情報通信学会、人工知能学会、日本ソフトウェア科学会、IEEE Computer Society 各会員。