

調理履歴に基づくユーザのレシピ選択モデルの提案

山本 修平^{1,3} 神門 典子² 佐藤 哲司¹

概要：調理で使用する食材や調理手順が書かれたレシピを共有し、検索するレシピサイトが普及・発展してきている。特に、生活者が直接投稿する投稿型のレシピ共有検索サイトの発展は著しく、膨大なレシピが日々投稿されるとともに、検索したレシピを使って調理するユーザも増えてきている。著者らはこれまでに、ユーザの調理レパートリーを拡大することを目的に、ユーザに未経験食材を含む挑戦的なレシピ推薦手法を提案している。本論文では、ユーザの連続的なレシピ選択をモデル化するため、ユーザにとって挑戦的なレシピ選択モデルと、嗜好的なレシピ選択モデルを組み合わせた手法を提案する。ユーザ毎の調理履歴から、挑戦的選択が発生する確率を指数分布のパラメータとして学習し、嗜好的選択と挑戦的選択を調理回数に応じて調整したスコアを算出する。ユーザが実際に調理したレシピ系列を用いて提案手法の有効性を評価した結果、嗜好的選択のみを用いた手法に比べて、挑戦的選択を組み合わせた手法が高い推定精度を示したので報告する。

Recipe Selection Model based on Cooking History

SHUHEI YAMAMOTO^{1,3} NORIKO KANDO² TETSUJI SATOH¹

1. はじめに

料理レシピを共有するレシピサイトの普及により、膨大な数のレシピがインターネット上に公開されている。生活者が自らのレシピを投稿するレシピ共有・検索サイト COOKPAD^{*1} には、200 万件を超える膨大なレシピが蓄積され、レシピを検索して利用するだけでなく、レシピの調理経験を介したユーザ間の交流も盛んに行われている。

このようなレシピサイトの普及・発展に伴って、生活者は日常の生活の中で、調理の機会毎にレシピサイトを利用できるようになってきた。著者らはこれまでの研究で、連続するレシピの検索過程において、レパートリー拡大を狙いとしたレシピ推薦法を提案している [15]。そこでは、ユーザのレパートリー拡大に効果的なレシピを推薦するため、ユーザのこれまでの調理経験と比べて、使用する食材がわずかに増えていることに加え、多くのレシピ中に頻出する食材が含まれている汎用性の高いレシピを推定している。

一方で、ユーザは連続するレシピの調理過程において、

常にレパートリー拡大を目的としたレシピ選択をするとは限らない。満足できる食事をするために、自身の持つ料理レパートリーの中からレシピを選択したり、前回レパートリー拡大を狙ったレシピ選択で上手に調理ができなかったために、再度そのレシピを調理し熟練度を上げる選択などが考えられる。著者らの先行研究 [17] では、同じレシピを何日も繰り返し調理するユーザや確認していることから、この仮説も十分に支持できる。

本論文では、このようなユーザの連続する調理過程を考慮したレシピ推薦を目的に、ユーザのレシピ選択行動をモデル化し、これまでの調理経験から次回に選択するレシピの推定手法を提案する。ユーザは嗜好的選択、挑戦的選択の2つのレシピ選択モデルを持ち、またその2つの選択モデルを遷移しながら、連続的なレシピ選択をしていると仮定する。嗜好的選択はユーザが自身が得意とするレシピを選択するモデル、挑戦的選択はユーザが未経験レシピ群からレシピを選択するモデルである。ユーザはレシピを選択する際にいずれかの選択モデルの観点に立ち、各モデル毎に異なる選択行動をとる。

本論文の構成を以下に示す。第2章では関連研究について整理する。第3章ではユーザの調理履歴に基づくレシピ

¹ 筑波大学 図書館情報メディア系

² 国立情報学研究所 情報社会相関研究系

³ 日本学術振興会 特別研究員

*1 <http://cookpad.com>

選択行動について説明する。第4章では実際の調理履歴を用いた評価実験をし、第5章では実験結果について考察をする。最後に、第6章で結論と今後の課題を述べる。

2. 関連研究

本論文で提案するレシピ選択モデルは、ユーザ毎の過去の調理経験から、現在のレシピ選択の観点が「嗜好的選択」と「挑戦的選択」のどちらかを推定した後、それぞれの選択モデルに合わせたレシピを推定する。このことから、レシピ推薦やレシピ構造に関する研究を2.1節に、食事の嗜好に関する研究を2.2節にまとめる。

2.1 レシピ推薦やレシピ構造に関する研究

ユーザの調理経験や調理状況に合わせたレシピ推薦に関する研究は盛んに行われている。矢島ら[10]は、レシピの構成要素と調理者の状況を考慮することで、その調理者が簡単に調理できるレシピを推薦している。大山ら[20]は、ユーザの調理に関する知識や技術に基づいたレシピ推薦システムを開発している。調理に関する知識と技術を「調理スキル」、レシピを構成する各作業を「調理作業」と定義し、ユーザに対する調理作業の難易度を推定している。あらかじめ、いくつかの調理作業を評価することで、ユーザの調理スキルと調理作業の難易度を推定し、ユーザに適したレシピの推薦を目指している。苅米ら[3]は、レシピに出現する食材から栄養バランスに優れたレシピの推薦を試みている。食事制限やアレルギーで苦しむ人々に適したレシピや献立の推薦を目指している。平川ら[18]は、レシピ中のテキスト情報だけでなく、レシピ画像の色情報とレシピに含まれる修飾語を用いて、調理目的に応じたレシピ選択の支援を目指している。

このように、従来のレシピ推薦手法は、ユーザの嗜好や状況に合ったレシピ単体の推薦を目的としている。一方、本論文で提案するユーザのレシピ選択モデルでは、ユーザのレシピ選択は「嗜好的選択」と「挑戦的選択」からなると仮定し、それぞれの選択モデルを逐次推定することで、連続的なレシピ推薦を目指す。

食材や調理法に着目して調理手順の構造を解析する研究も盛んに取り組まれている。Suら[7]は、食材と調理法の関係を解析してレシピ推薦に活かす手法を提案している。Wangら[9]やYuら[12]は、調理の手順をワークフローとしてレシピグラフで表すことを試みている。レシピグラフの部分グラフの類似性を用いることで、推薦の精度を向上できるとしている。FreyneとBerkovsky[1]は、レシピと食材の間で二部グラフを作成し、ユーザの食品に対する嗜好を推定するとともに、ダイエットに適したレシピを生成する手法を提案している。彼らの実験によれば、幅広い範囲で妥当な精度を達成できたと報告している。Kuoら[6]は、いくつかのレシピを組み合わせる知的なレシピ構成法

を提案している。ユーザが指定した食材を含むレシピとなるように、レシピと食材の共起関係から生成するグラフベースのアルゴリズムを実装している。Hamadaら[2]は、料理番組のテキスト教材の構造化を行っている。食材や調理法、調理器具等に関する固有の辞書を構築し、構築した辞書を用いて調理手順を構造解析し、フローダイアグラムを作成している。杉山ら[19]は、調理手順の似通ったレシピの中から、典型的な調理手順や特徴的な調理手順の発見を目指している。検索結果で得られた各レシピの調理手順を構造化し、多くのレシピに共通する手順と、各レシピ間で差異のある手順とを検出し、料理の典型的な調理手順と、各レシピにおける特徴的な調理手順をユーザに提示する。その際、出現頻度が高い、あるいは重要性の高い食材もレシピの特徴として考慮している。Kuoら[6]は、レシピに出現する食材に着目し、レシピサイトに掲載されている複数のレシピを組み合わせ、新たな献立を作成するシステムを構築している。複数のレシピを組み合わせられた献立を掲載しているレシピサイトも存在するが、異なる食材を幅広く摂取する、あるいは、決められた食材だけからなる献立などを考えると、決してバリエーションが豊富とは言えない。このため、レシピ間で出現する食材のJaccard係数を算出し、出現する食材の種類や数を献立内で制御する手法を提案している。

以上述べたように、食材や調理法に着目して調理手順を解析する研究も多く知られているが、レパートリー拡大を狙いとした挑戦的選択に関する研究は知られていない。本論文では、嗜好的観点と挑戦的観定の2つを組み合わせ、ユーザの連続的なレシピ選択行動の適切なモデル構築を目指す。

2.2 食事の嗜好に関する研究

食事は生活に必要な栄養素を摂取するだけが目的ではなく、食事をすることでストレス解消などの効果を持つことが知られている。食事には娯楽的側面があり、ユーザが好きな食事レシピを選択できるよう、食事の嗜好に関する研究も広く行われている。Yangら[11]は、ユーザ毎の食材の嗜好を学習することを目的に、料理が写った画像をユーザに提示し、ユーザが好きな食材をクリックすることで、ユーザの嗜好を抽出するシステムを開発している。Uedaら[8][16]は、食材の使用履歴と調理者のレシピサイト閲覧履歴から、ユーザの好きな食材と嫌いな食材を考慮したレシピ推薦手法を提案している。一度推薦したレシピをユーザに連続して提示しないよう、推薦済みのレシピのスコアを低くするためのパラメータを導入している。千葉ら[14]は、食事レシピを生理学的に定められている基本となる味である、「甘味」、「酸味」、「塩味」、「苦味」、「うま味」の5つに分類している。食事レシピから抜き出した材料や調理手順から、それぞれの分類における特徴的な単語の頻度

を求め、材料の味の決定影響度を算出し、それぞれのスコアを付与している。玉田ら [13] は、健康的な生活習慣を実現する料理レシピを推薦することを目的に、各レシピの持つ栄養素をあらかじめ算出しておき、1週間の理想的な栄養素を摂取するようレシピ推薦するシステムを開発している。各レシピを「和風」「洋風」「中華」「韓国風」のいずれかに分類し、ユーザーが好きな項目のレシピの推薦を実現している。

本論文では、ユーザーの調理履歴からユーザーの嗜好を抽出する。このため、Ueda ら [8] の研究を参考に、TF-IDF 法に基づき各食材の嗜好を定量的に決定する。一方で、本研究は嗜好を考慮するだけでなく、ユーザーの挑戦的な観点も考慮したレシピ選択モデルを提案しているところに特徴がある。

3. 連続的レシピ選択モデル

3.1 概要

本節では、本論文で提案するユーザーの連続的レシピ選択モデルの概要を説明する。ユーザーは選択したレシピを調理し食べるという行為を伴う。多くのユーザーは食事に好き嫌いという嗜好を持つだけでなく、同じレシピを連続的に食べることに飽きたり、健康を考慮するため、可能な限り多様なレシピを調理すると考えられる。このことから、ユーザーの食事嗜好を食材単位でモデル化するだけでは、ユーザーの好きな（調理経験が比較的豊富な）食材を使うレシピを繰り返し推定し、ユーザーの現実的なレシピ選択とは異なるものになると考えられる。

一方で、全てのユーザーが全てのレシピを失敗なく調理できるとは限らない。ユーザーは、自らの調理経験に基づいて、調理できそうなレシピを選択していると考えられる。例えば、過去に経験したことのあるレシピと類似したレシピや、自分が得意な食材が含まれているレシピを選択することなどが予想される。特に、未経験の食材が含まれているレシピを調理することにより、ユーザーの調理可能なレシピを拡張していく「レパートリー拡大」の狙いも考えられる。

以上の考察より、本論文ではユーザーのレシピ選択行動は、「嗜好的選択」と「挑戦的選択」という2つの潜在的な状態からなり、一定の確率でそれぞれを遷移しながら連続的に行われていると仮定する。嗜好的選択では、ユーザーが自身のレパートリーとするレシピを選択する。挑戦的選択では、ユーザーが未経験レシピ群からレシピを選択する。以下、3.2節では、ユーザー毎のそれぞれの状態推定方法について説明する。3.3節では、それぞれ嗜好的選択モデルと挑戦的選択モデルにおけるレシピ推定手法について説明する。

3.2 レシピ選択行動系列の推定方法

ユーザーは嗜好的選択と挑戦的選択のそれぞれの観点からレシピ選択をし、ユーザー毎に異なる確率でそれぞ

れの潜在的な選択行動を遷移していると仮定する。本説の目的は、ユーザーの調理したことのあるレシピ系列を $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_T\}$ としたとき、各時刻でレシピが選択された状態系列 $\mathbf{q} = \{q_1, q_2, \dots, q_T\}$ を推定することである。ここで、 T はユーザーのそれまでの調理回数であり、 q_t は $\{0, 1\}$ いずれかの2状態を持ち、 $q_t = 0$ のとき嗜好的選択、 $q_t = 1$ のとき挑戦的選択であるとする。それぞれのレシピ \mathbf{r}_t は、 I 次元の食材集合からなり、 $\mathbf{r}_t = \{r_{t,1}, r_{t,2}, \dots, r_{t,I}\}$ と定義し、レシピで使われた食材を1、使われなかった食材を0とする。

食材の集合からなるレシピ系列 \mathbf{R} が与えられたとき、各レシピの選択行動の系列 \mathbf{q} は、次のコスト関数 $c(\mathbf{q}|\mathbf{R})$ を最大化することによって得られる。

$$c(\mathbf{q}|\mathbf{R}) = \sum_{t=2}^T \left(q_t \log p_{q_t}(\mathbf{r}_t) + (1 - q_t) \log p_{q_t}(\mathbf{r}_t) \right), \quad (1)$$

ここで、 $p_{q_t}(\mathbf{r}_t)$ は状態 q_t でレシピ \mathbf{r}_t を調理する利得で、次のように定義する。

$$\log p_{q_t}(\mathbf{r}_t) = \sum_{i=1}^I r_{t,i} \log p_{q_t}(t, i), \quad (2)$$

$$p_0(t, i) = \frac{n_{t,i} + 1}{t + 2}, \quad p_1(t, i) = \frac{m_t + 1}{t + 2}, \quad (3)$$

p_0 は嗜好的選択 ($q_t = 0$) をする利得で、これまでの調理回数 t で食材 i の調理回数 $n_{t,i} (= \sum_{k=1}^{t-1} r_{k,i})$ を除した確率として計算される。 p_1 は挑戦的選択 ($q_t = 1$) をする利得で、ユーザーが時刻 t までの全ての食材に対する平均調理回数 $m_t = \frac{1}{I} \sum_{k=1}^{t-1} n_{k,i}$ を調理回数 t で除した確率として計算される。すなわち、全ての食材が一様の確率値を持つ。 I_t は時刻 t までに経験したことのある食材の数である。

式 (1) は、ユーザーが選択したレシピ \mathbf{r}_t に、多く経験したことがある食材が豊富に含まれている場合に嗜好的選択が推定され、未経験の食材や経験の少ない食材が多く含まれている場合に、平均調理回数に基づき計算される p_1 が p_0 に比べ高い利得を持つため、挑戦的選択が推定される。調理経験を重ねる度に利得 p_0 と p_1 はそれぞれ更新され、ユーザー毎に時刻 t までの調理経験を反映したレシピ選択行動の推定が可能となる。

3.3 レシピスコアの算出方法

本節では、ユーザーが次に調理する可能性の高いレシピ \mathbf{r}_{t+1} を推定するため、各レシピにスコアを算出する手法を説明する。嗜好的観点と挑戦的観点から得られたスコアをそれぞれ $pref$ と $challenge$ としたとき、それらを統合したスコアを以下の式で定義する。

$$score(\mathbf{r}_{t+1}) = challenge(\mathbf{r}_{t+1})^\beta \cdot pref(\mathbf{r}_{t+1})^{(1-\beta)}, \quad (4)$$

ここで、 β は挑戦的選択が生起する確率である。最後に挑戦的選択がされてから、ユーザが次に挑戦的選択をする確率は、指数分布の累積分布関数 $\beta = 1 - e^{-\lambda x}$ に従うとする。ここで、 x は最後に挑戦的選択がされてからの調理回数である。挑戦的選択がされた直後の調理では、調理回数が 0 であるため、挑戦的選択が生起する確率は $\beta = 1 - 1 = 0$ となる。 λ は、ユーザごとに決定されるパラメータで、最後に挑戦的選択がされてから次に挑戦的選択がされるまでの調理回数の平均の逆数によって推定される。 λ が大きいほど、挑戦的選択をし易いユーザであるとみなせる。

3.3.1 嗜好的選択におけるレシピスコアの算出方法

本節では、嗜好的選択時のユーザのレシピ選択行動における、各レシピのスコアを算出する。Ueda[8] らはユーザのレシピに対するスコアを、ユーザのレシピに対する嗜好度と、過去にまだ推薦していないレシピに対する重みをパラメータを調節することで算出している。嗜好度は、ユーザの各食材に対する嗜好度を、文書の重み付け手法として知られる TF-IDF 法に基いて予め算出しておき、その値に基づいて各レシピにスコアを算出している。本論文でも、ユーザの次回調理するレシピ \mathbf{r}_{t+1} に対する嗜好度 $tfidf(\mathbf{r}_{t+1})$ を、TF-IDF 法に基いて次のように算出する。

$$tfidf(\mathbf{r}_{t+1}) = \sum_{i=1}^I tf_i \cdot idf_i \cdot r_{t+1,i} \quad (5)$$

$$= \sum_{i=1}^I p_0(t, i) \cdot \log\left(\frac{R}{n_i}\right) \cdot r_{t+1,i}, \quad (6)$$

tf_i は食材 i に対するユーザの嗜好度を表し、これは調理回数で除し確率値とした p_0 とする。 idf_i は、使用されやすい食材（調味料など）の重みを低くする役割を持ち、全てのレシピ R において、食材 i が登場した回数 n_i で除した逆頻度として計算される。

一方で、ユーザは好みの食材が含まれているレシピを繰り返し調理するだけでなく、前回調理したレシピとは似ていないレシピを調理することによって、多様な食事をしていと考えられ、ユーザの直近の調理経験から離れているレシピに対してスコアを与えることも必要である。本論文では、次のような調理回数に応じた減衰関数によって、直近で調理したレシピと類似するレシピの重みを低くすることによって実現する。

$$div(\mathbf{r}_{t+1}) = \sum_{k=1}^t \frac{1 - sim(\mathbf{r}_{t+1}, \mathbf{r}_k)}{\log_2(t - k + 2)}, \quad (7)$$

ここで、 $sim(\mathbf{r}_{t+1}, \mathbf{r}_k)$ は両レシピ間の類似度で、両レシピが持つ食材の IDF 値を特徴量としたコサイン類似度で与える。調味料などの頻繁に使用される食材に比べ、あまり使用されない特徴的な食材が片方のレシピに存在しない場合に、レシピ間の類似度は低くなる。両レシピが含む食材

集合が完全に一致している場合に 1 を、全く一致していない場合に 0 を与える。直近に調理したレシピほど類似度を直接与え、調理履歴を遡るほど類似度の影響は低くなる。

以上の両指標をパラメータ α によって組み合わせた以下の式でスコアを算出することにより、ユーザが嗜好的選択で選ぶレシピを推定する。

$$pref(\mathbf{r}_{t+1}) = tfidf(\mathbf{r}_{t+1})^\alpha \cdot div(\mathbf{r}_{t+1})^{(1-\alpha)}, \quad (8)$$

α は、 $0.0 \leq \alpha \leq 1.0$ の範囲でユーザに合わせて調節する。 α が 1.0 に近づくほど推定結果を良く近似できるユーザは、直近に調理したレシピとは関係なく嗜好するレシピを選択している。 α が 0.0 に近いユーザは、直近に調理したレシピと似ていないレシピを選択しているとみなせる。

3.3.2 挑戦的選択におけるレシピスコアの算出方法

著者らの先行研究 [15] では、ユーザにとって挑戦的なレシピを推定するために、個々の食材の馴染み深さを定番度として算出し、定番度が低いレシピを挑戦的なレシピとして推薦している。定番度は、使い勝手の良い食材は多くのレシピに登場するという仮説に基づき、Web ページのスコアリング手法として知られる HITS アルゴリズム [4] を用いてそれぞれの食材に対して算出される。

本論文でも、このような仮説に基づきレシピの難易度を推定する。すなわち、ユーザにとって挑戦的なレシピとは、これまで調理したことがない食材が含まれているだけでなく、多くのユーザにとって馴染みのない食材が含まれているレシピであると仮定する。ここで、前節でも算出した各食材の IDF 値に着目すると、 idf_i は食材 i がレシピに登場した回数が多いほど低くなり、登場した回数が少ないほど高い値を持つ特徴がある。本論文では、あまり登場しない食材を含むレシピは挑戦的と想定し、 idf_i に基づいて挑戦的なレシピを推定する。ただし、レシピに含まれている全ての食材が挑戦的である場合、ユーザは調理を躊躇することも予想される。そこで、各レシピが含む食材から、 idf_i の最大値から最小値を引いた次のスコアをレシピの挑戦度とする。

$$challenge(\mathbf{r}_{t+1}) = \max_{i \in I_t^+} idf_i - \min_{i \in I_t^-} idf_i, \quad (9)$$

ここで、 I_t^+ はレシピ \mathbf{r}_{t+1} に登場し、ユーザが時刻 t で調理経験がない ($n_{t,i} = 0$) 食材の集合である。調理経験がない食材のうち、レシピ全体であり登場しない食材と頻繁に登場する食材が同時にレシピに含まれているとき、レシピの挑戦度は高い値を持つ。

4. 評価実験

4.1 データセット

提案手法の有効性を評価するため、国立情報学研究所が

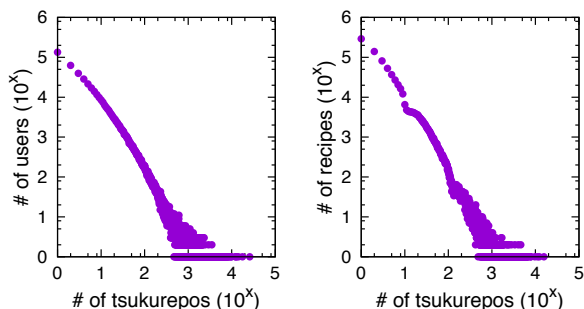


図 1 つくレポユーザ数分布 図 2 つくレポレシピ数分布

提供する COOKPAD データセットを使用した*2。COOKPAD では、ユーザの日常的な調理履歴が「つくレポ」として投稿されている。各つくレポには、つくレポを投稿したユーザ、そのつくレポの対象となったレシピ、つくレポを投稿した日時が紐付いている。本論文では、ユーザ毎のつくレポを調理経験の系列とみなし評価実験をする。つくレポを投稿した回数毎のユーザの頻度分布を図 1 に示す。つくレポ投稿が 1 度だけのユーザは 10^5 程度存在し、ユーザによっては最大で 10,000 件以上つくレポを投稿している。継続的に調理経験を投稿しているユーザを対象とするため、つくレポを 500 件以上かつ 1,000 件以下投稿した 1,234 ユーザからランダムに 100 ユーザを抽出した。

評価対象とするレシピは、他のユーザによって再現できた（つくレポが投稿された）レシピとし、その種類数は 805,018 件であった。それぞれのレシピのつくレポ被投稿回数の頻度分布を図 2 に示す。つくレポの被投稿回数が 1 件のレシピは 10^5 を超過して存在し、10,000 件以上のつくレポが投稿されたレシピもいくつか存在する。このレシピデータセットからレシピ推定に使用する IDF 値を算出した。

4.2 食材の表記統一処理

個々のレシピに登場する食材は、レシピ投稿者によって自然言語で記述されている。このことから、レシピ中の食材名は様々な表記で記述されている。例えば、食材の「たまねぎ」については、「玉ねぎ」や「タマネギ」のように、様々な表記で用いられることがある。このため、単純に食材名を素性として用いるだけでは、同じ要素を指す食材でも別の食材として扱われる可能性がある。

本論文では、日本語形態素解析器 MeCab[5] を用いて、食材名を形態素に分割する。形態素解析器の辞書は Wikipedia のタイトルリストで予め拡張している。ここで、動詞と名詞で場合分けして処理をする。動詞は単語を基本形に変換し、動詞を活用させることによって表記の揺れを吸収する。名詞は、単語をカタカナに変換し、同じ食材でも表記が揺

れる問題に対処する。加えて、直前の単語が一般名詞やサ変接続名詞、固有名詞である場合は、前の単語に文字列連結させることで、「オリーブオイル」のように、「オリーブ」と「オイル」という 2 つの単語に分割される食材についても、1 つの単語として扱うようにする。

4.3 実験方法

提案手法の有効性を評価するため、評価対象とするユーザのそれぞれのつくレポ投稿系列の前半 9 割を訓練データとし、後半 1 割を評価データとした。訓練データに対してレシピ選択の状態系列 \mathbf{q} を推定し、挑戦的選択が発生する確率を調整するパラメータ λ を学習する。また、提案手法は直近の調理履歴との類似度もスコアの一部としていることから、後半 1 割のデータも評価後は逐次訓練データとしてモデルを更新する。嗜好的選択におけるパラメータ α は、前半 9 割の訓練データで得られたレシピのスコアが、ランダムに抽出した 100 件のレシピの中で高くなるようユーザ毎に調整する。

評価では、次の時刻で実際にユーザがつくレポを投稿したレシピ 1 件に加えて、つくレポが投稿されたことがあるレシピ 805,018 件からランダムに 100 件抽出し、提案手法で算出したスコアの降順にランキングを作成する。実際にユーザがつくレポを投稿したレシピが、より上位にランク付けされているほど、提案手法は有効に機能しているとみなせる。これを定量的に評価するため、システムが出力した上位 K 件の平均精度の全ユーザの平均 (MAP@K) を計算する。

嗜好的選択と挑戦的選択を組み合わせることの有効性を評価するため、嗜好的選択のみで得られたスコア $pref$ と、挑戦的選択のみで得られたスコア $challenge$ でもランキングを作成し、それぞれの平均精度を比較する。

4.4 実験結果

評価対象としたユーザの調理履歴に対し、嗜好的選択か挑戦的選択の状態系列 \mathbf{q} を推定し、挑戦的選択を発生させる確率 β を調整する λ を算出した。この結果、 λ が最低値と最高値を示したユーザの状態系列を図 3 に示す。図では、縦軸に q_t の値を、横軸に調理回数を示している。調理を始めてからすぐは、モデル p_0 が十分にユーザの嗜好を反映できていないため、いずれのユーザも頻繁に挑戦的選択が推定されている。調理履歴が蓄積されてからは、挑戦的なレシピを頻繁に選択するユーザと選択しないユーザで差が現れ、その結果が λ の値として反映されている。

訓練データから推定したパラメータ α と λ の頻度分布を、図 4 と図 5 に示す。 α は 0.1 刻みで調節し、 λ は小数点第 1 位で四捨五入し頻度を求めた。 α はその値が大きくなるほど、直近に調理したレシピと関係なく嗜好するレシピを選択するユーザであるとみなせる。100 ユーザの α 分布

*2 <http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html>

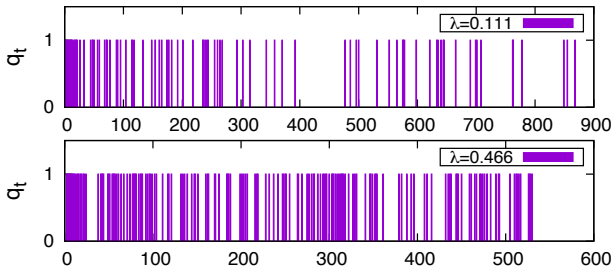


図 3 q の系列と推定された λ の例

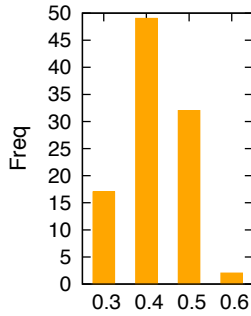


図 4 α の頻度分布

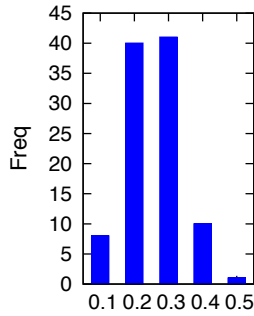


図 5 λ の頻度分布

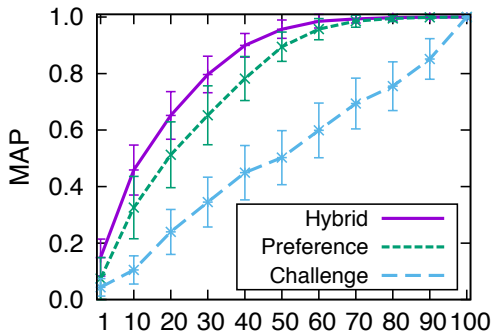


図 6 各手法の MAP 値の推移

は、0.3 から 0.6 の範囲に取まっており、ユーザが最も多くなつた値は $\alpha = 0.4$ であった。 λ はその値が大きくなるほど、挑戦的選択をしやすいユーザであるとみなせる。 λ は 0.1 から 0.5 の範囲に取まっており、評価対象としたユーザのうち、8 割ほどのユーザは $0.2 \leq \lambda \leq 0.3$ であった。

これらのユーザ毎に推定したパラメータを用いて、実際につくレポで調理したレシピを推定できるか評価した結果を図 6 に示す。 図では横軸にランキングの上位 K 、縦軸に上位 K を対象としたときの MAP 値をその標準偏差と共に示している。 K の刻みは 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 とした。 嗜好的観点と挑戦的観点を組み合わせた Hybrid が、いずれの順位においても嗜好的観点のみで推定した Preference と挑戦的観点のみで推定した Challenge よりも高い値を示した。 Preference と Challenge では、1 位のときは MAP 値に大きな差がないものの、10 位以降は Preference が Challenge に比べ高い値を示している。

5. 考察

図 6 より、挑戦的選択のみでスコアを算出する方法に比べ、嗜好的選択のみでスコアを算出する方法が高いレシピ推定性能を示した。 図 5 から、 $\lambda \leq 0.3$ に 9 割近くのユーザが含まれている。 図 7 に、評価対象としたユーザから確認できた λ の値ごとに、挑戦的選択が発生する確率 β がどのように上昇していくかを示す。 図から $\lambda = 0.3$ であるときは、挑戦的選択が発生する確率 β が 0.5 を越えるのは 3 回目の調理、 $\lambda = 0.2$ であるときは 4 回目の調理である。 すなわち、ほとんどのユーザは主として嗜好的観点でレシピ選択をしていると考えられる。 ただし、両観点を組み合わせた提案手法がいずれの手法に比べても高い推定性能を示していることから、挑戦的観点もレシピ選択において稀に発生していることが示唆される。 図 8 に、あるユーザが実際に調理したレシピを、それぞれの手法が第何位に推定したかを示す。 図は縦軸が順位、横軸が調理回数である。 2 つの選択モデルを組み合わせた Hybrid は 3 つの手法の中で最も性能が良く、実際に調理したレシピを高い順位で推定している。 挑戦的選択モデルは実際に調理したレシピを低い順位に推定することも多いが、何度か第 1 位で推定することに成功していることから、本論文で仮定したユーザの挑戦的選択も何度か発生していることが確認できる。

図 4 では $\alpha \leq 0.5$ にほとんどのユーザが含まれていることから、ユーザはレシピの嗜好的選択において、直近に調理したレシピも考慮してレシピ選択していることが示唆される。 ユーザに推定される α はいずれの時刻においても一定値であるとしたが、このパラメータも β と同様に時刻毎に動的に決定する方が有効であることも考えられる。 図 9 に、あるユーザの評価データに対するレシピ推定結果の順位に加え、もしも時刻毎に最適な α を選択した場合の順位と、そのときの α を同一の横軸上に示している。 このユーザに推定された α は 0.5 で、レシピの嗜好と直近の調理履歴を同等の比率で考慮しているとみなせる。 一方、時刻ごとに最適な α を見ると、 $\alpha \geq 0.9$ と $\alpha \leq 0.2$ の両端に偏って頻出している。 この結果から、このユーザは常に $\alpha = 0.5$ として直近の調理履歴を常に考慮しているわけではなく、素朴に好きなレシピを選択するどときと、直近に調理したレシピとは似ていないレシピを選択するときを切り替えながら、連続的なレシピ選択行動をしている。 以上の結果から、嗜好的選択におけるパラメータ α を、調理履歴から直近のレシピとの類似度の系列を生成し、調理回数と共に調整することによって、提案手法はより高い推定性能を得られると考えられる。

6. 結論

本論文では、ユーザの連続する調理過程を考慮したレシピ推薦を目的に、ユーザのレシピ選択行動をモデル化し、

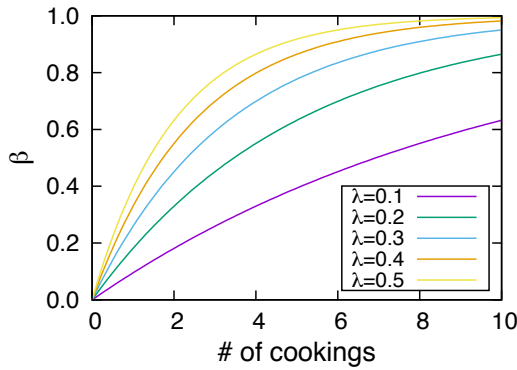


図 7 λ の値別に見た調理回数毎の β

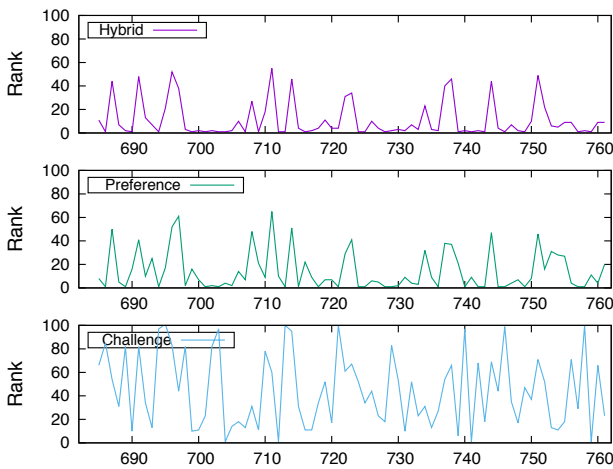


図 8 あるユーザーに対してそれぞれの手法が推定したレシピ順位系列

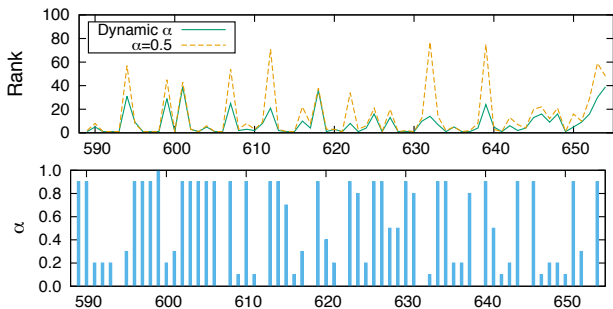


図 9 動的な α としたときのあるユーザーのレシピ推定順位系列

これまでの調理経験から次回に選択するレシピの推定手法を提案した。ユーザーは嗜好的観点と挑戦的観点という2つのレシピ選択モデルを持ち、またその2つの選択モデルを遷移しながら、連続的なレシピ選択をしていると仮定する。これまでの調理履歴から挑戦的選択が発生する確率を指数分布のパラメータとして学習する。次回に調理するレシピを嗜好的観点と挑戦的観点からそれぞれスコアを算出した後、事前に学習したパラメータによって組み合わせたスコアでレシピを推定する。

COOKPAD に投稿された「つくレポ」を調理履歴とみな

し評価実験を行った結果、嗜好的選択と挑戦的選択を組み合わせた提案手法が、それぞれで独立した手法に比べ高い推定性能を示した。嗜好的観点においては、直近の調理履歴を考慮して似ていないレシピを選択する状態と、素朴に好きなレシピを選択する状態のいずれかを遷移しながら、レシピ選択が行われている示唆が得られた。

今後の課題は、嗜好的選択における直近の調理履歴を考慮するパラメータ α を動的に変化させることと、調理履歴が少ないユーザーに提案手法を適用し、ユーザーのレシピ選択行動を支援できるか評価することが挙げられる。

謝辞

本研究は、クックパッド株式会社と国立情報学研究所が提供する「クックパッドデータ」を利用した。本研究は、NII 戦略研究公募型共同研究、および JSPS 科研費 25280110、15J05599 の助成を受けた。ここに記して謝意を示す。

参考文献

- [1] J. Freyne and S. Berkovsky. Recommending food: Reasoning on recipes and ingredients. In *Proceedings of the UMAP 2010*, pp. 381–386. Springer, 2010.
- [2] R. Hamada, I. Ide, S. Sakai, and H. Tanaka. Structural analysis of cooking preparation steps in Japanese. In *Proceedings of the IRAL 2000*, pp. 157–164, New York, NY, USA, 2000.
- [3] S. Karikome and A. Fujii. A system for supporting dietary habits: Planning menus and visualizing nutritional intake balance. In *Proceedings of the ICUIMC 2010*, pp. 56:1–56:6, 2010.
- [4] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, Vol. 46, No. 5, pp. 604–632, 1999.
- [5] Taku Kudo. Yet another part-of-speech and morphological analyzer. <http://mecab.sourceforge.net/>.
- [6] Fang-Fei Kuo, Cheng-Te Li, Man-Kwan Shan, and Suh-Yin Lee. Intelligent menu planning: Recommending set of recipes by ingredients. In *Proceedings of the ACM Multimedia 2012 Workshop on Multimedia for Cooking and Eating Activities, CEA '12*, pp. 1–6, New York, NY, USA, 2012. ACM.
- [7] H. Su, M. Shan, T. Lin, J. Chang, and C. Li. Automatic recipe cuisine classification by ingredients. In *Proceedings of the UbiComp 2014*, pp. 565–570, 2014.
- [8] M. Ueda, M. Takahata, and S. Nakajima. User's food preference extraction for personalized cooking recipe recommendation. In *Proceedings of the SPIM 2011*, pp. 98–105, 2011.
- [9] L. Wang, Q. Li, N. Li, G. Dong, and Y. Yang. Substructure similarity measurement in Chinese recipes. In *Proceedings of the WWW 2008*, pp. 979–988, 2008.
- [10] A. Yajima and I. Kobayashi. “easy” cooking recipe recommendation considering user's conditions. In *Proceedings of the WI-IAT 2009*, pp. 13–16, 2009.
- [11] Longqi Yang, Yin Cui, Fan Zhang, John P. Pollak, Serge Belongie, and Deborah Estrin. Plateclick: Bootstrapping food preferences through an adaptive visual interface. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*,

- CIKM '15, pp. 183–192, 2015.
- [12] L. Yu, Q. Li, H. Xie, and Y. Cai. Exploring folksonomy and cooking procedures to boost cooking recipe recommendation. In *Proceedings of the APWeb 2013*, pp. 119–130. Springer, 2011.
 - [13] 玉田雄基, 佐藤哲司. 栄養素に基づく嗜好ラベル付きレシピ推薦システムの提案. 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, 2016.
 - [14] 千葉祐輔, 本田真望, 大島邦夫. 味の嗜好に応じたレシピの検索. 全国大会講演論文集, Vol. 70, pp. 417–418, mar 2008.
 - [15] 中岡義貴, 佐藤哲司. 食材の偏りと調理法に基づくレパートリー拡大のためのレシピ推薦システムの提案. In *DICOMO2014*, pp. 1653–1660, 2014.
 - [16] 中川明莉沙, 上田真由美 and 高畑麻理, 中島伸介. 好き嫌いラベル付き食材分量を考慮したレシビスコア算出方式. 第 4 回データ工学と情報マネジメントに関するフォーラム, 2012.
 - [17] 平松淳, 山本修平, 伏見卓恭, 佐藤哲司. レシピコミュニティにおける投稿傾向の分析. In *DE2015*, 2015.
 - [18] 平川芽依, 牛尼剛聡, 角谷和俊. 料理画像の色情報に基づく目的に応じたレシピ選択支援. 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, 2016.
 - [19] 杉山祐一, 山肩洋子, 田中克己. 手順情報としてのレシピデータに対する類似レシピの要約と微小で重要な差異の発見. 第 5 回データ工学と情報マネジメントに関するフォーラム DEIM2013 論文集, D3-5, 2013.
 - [20] 大山裕也, 塚原みな, 中内靖. ユーザの調理スキルを考慮した調理支援システムに関する研究. ロボティクス・メカトロニクス講演会講演概要集, Vol. 2009, pp. 2A2–H08(1)–2A2–H08(4), may 2009.