

# 使用食材・器具・調理法に着目した 料理レシピの難易度推定アルゴリズム

藤崎 優理 三好 康夫

高知大学理学部応用理学科情報科学コース

あらまし：本研究の目的は、料理スキルを習得したいユーザに対し、適切な難易度の料理レシピを推薦することである。我々はレシピの難易度を手軽さ、技術的な難しさ、レシピ文章の分かりやすさの3つに分類し、それぞれの難易度の推定を目指す。本稿では、まずレシピの手軽さの推定を行うため、使用している食材や調理器具、調理法を表す単語に着目した。そして、これらの単語の出現頻度からレシピの難易度を推定するアルゴリズムを開発した。

## 1. はじめに

初心者が料理を覚えたいとき、まずは簡単なものからと考える。そして、知らない料理を作るときには、たいいていレシピを見ながら作りかたを覚える。しかし、初心者にとってはレシピを探す際に、どれが簡単で、どれが難しいのかを判断することは難しい。そのため、レシピの難易度を求めることができれば、初心者が簡単なレシピを探すことが容易になり、料理を学びやすくなると考えた。そこで本研究では、難易度を推定する手法について提案する。レシピのデータはクックパッド株式会社が国立情報学研究所と協力して研究者に提供しているデータセット (<http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html>) を使用した。

## 2. 難易度推定手法

### 2.1. 先行研究

我々は先行研究において、学習コンテンツの難易度を推定するアルゴリズムを開発している[1]。このアルゴリズムは、どのユーザがどのアイテム（学習コンテンツ）を読んだかという2部ネットワークから、ユーザの習熟度とアイテムの難易度を推定するものである。基本的な考え方は、多くのアイテムを読んだユーザは詳しいはず、難しいアイテムを読んだユーザは詳しいはず、詳しくないユーザに読まれたアイテムは易しいはず、等といった仮説に基づき、収束するまで反復計算するというものである。そこで我々はまず、このアルゴリズムをレシピの難易度推定に適用してみた[2]。

このアルゴリズムの評価のため、アルゴリズムによりレシピを難易度順に並び替えたものと、無作為に抽出した30のレシピを人が難易度順に並び替えたもの（正解データ1）とを比較した。なお、正解データ1は料理の技術的な難しさ順に並び替えたものである。比較にスパマンの順位相関を用いた結果、相関係数は  $-0.54$  で、逆順ではあるが、やや強い相関関係が認められた。

### 2.2. 結果が逆順となった原因の考察

難易度順が逆となってしまった原因は、アルゴリズムの対象ユーザの特徴が異なることにあると考えられる。このアルゴリズムは、習熟度が高くなるにつれ、ユーザはどんどん難しいアイテムに挑戦していくという想定で開発されたものである。しかし、クックパッドの主な利用者は20～40代の女性で、そのうち子供がいる利用者は約5割とされている[3]。このような利用者は、多忙のため、手軽に作れる料理のレパートリーを増やしたいと考

えていると思われる。調理技術を高めたいという動機ではなく、手軽に作れる料理を見つけたいという動機で利用するユーザにとっての習熟度は、手軽なレシピを見つければスキルレベルということになる。つまり、クックパッドを多く利用して習熟度が高くなるにつれ、より簡単なものを求める傾向があると考えられる。そうであれば、結果が逆順になってしまったことの説明がつく。

### 2.3. 難易度の種類と求め方

レシピの難易度推定に先行研究のアルゴリズムを適用するにあたり、我々は難易度を技術的な難しさであると想定していた。しかし先に述べたように、クックパッドユーザにとって有益なレシピの難易度は手軽さであるかもしれない。そこでレシピの難易度の種類について考察した。人がレシピの内容を基に難易度を判断する際に基準となる要素には、ステップ数が多いかどうか、扱いにくい食材や道具が使われているかどうか、難しい技が使われているかどうか、レシピ本文の表現がわかりやすい（難しい表現や手順の省略が見られないか）かどうか、などがあると考えられる。これらの要素をふまえ、我々はレシピの難易度は、

- ① 手軽さ
- ② 技術的な難しさ
- ③ レシピの文章の分かりやすさ

の3種類があると定義する。本稿ではこれらの難易度のうち、手軽さを推定する手法について説明する。

先行研究はレシピの内容を用いずに難易度を推定する手法であったが、本稿で提案するレシピの手軽さ推定アルゴリズムは内容に基づく手法を用いる。

## 3. 提案アルゴリズム

### 3.1. アルゴリズムの概要

本稿で提案するアルゴリズムでは、レシピ中に使われている単語それぞれに得点をつけて手軽さを計算する。単語に得点をつける手法については、類似研究の石原ら[4]や矢嶋ら[5]の手法を参考にした。

アルゴリズムの流れを次に示す。

- (1) レシピ中の動詞、名詞だけを抜き出し、その単語が使われているレシピのIDとともにデータベースに格納する。
- (2) 抜き出した単語を4つのカテゴリ（食材、調味料、調理器具、調理技法）に分類する。
- (3) 単語の出現頻度などに基づき単語の得点を求める。

(4) カテゴリ毎に係数を求めた後、単語それぞれにカテゴリの係数を掛け合わせ、合計することでレシピの手軽さを表す得点を求める。

技術的な難しさやレシピの文章の分かりやすさによる難易度についても、今後この手法を応用することで求めたいと考えている。なお、本稿ではカテゴリの係数は全て1として計算している。

### 3.2. 単語の得点の求め方

クックパッドデータセットの全1,715,595件のレシピに含まれる単語をカテゴリに分類したところ、表1のように分類された。食材と調味料カテゴリの種類が多いのは、データセットの材料テーブルから自動抽出する際に表記のゆれ（しょうゆ、醤油、正油、しょう油など）を考慮していないためである。調理器具と調理技法カテゴリの種類が少ないのは手動で抽出したからである。

紙面の都合上、本稿では単語の得点の求め方についてのみ解説する。食材、調味料、調理器具のカテゴリに属する単語については、使われている頻度の多いものほど手軽なものであるとして、低い得点を与えるようにする。調理技法のカテゴリに属する単語については、時間のかからない技法ほど手軽であるとして、低い得点を与える。

頻度に基づき得点をつける単語については、カテゴリ中の単語を頻度の高いもの順に並びかえて10等分し、頻度の高い順に1~10の得点をつけた。調理技法の場合は、時間のかかる技法かどうかを手動で判定し、1~10の範囲で時間がかかる技法ほど高い得点をつけている。

## 4. 提案アルゴリズムの推定精度の評価

### 4.1. 正解データとの比較結果

提案アルゴリズムの精度を評価するため、レシピを手軽さ順に並び替えた正解データを作成した。正解データの作成方法として、比較対象のレシピを複数人が並び替え、その平均をとる方法と、一対比較法[6]を用いる方法の2種類を考えた。先行研究の評価で使用した正解データ1は前者の並び替えの方法で作成していた。今回作成する正解データ2と3は後者の一対比較法で作成する。

一対比較法において、20のレシピの比較の組み合わせは  $20 \times 19 \div 2 = 190$  であるが、1つの比較に対し3人に評価してもらうと計570回の比較を行う必要がある。今回は協力者が20名であったため、1人あたりの比較回数は28.5回である。我々は2つのレシピを見比べるWebインタフェースを用意し、このインタフェース上で協力者に手軽であると思ったレシピを選択してもらった。正解データ2は無作為に抽出した20のレシピを元に、正解データ3は豚肉が使われたレシピの中から無作為に20を抽出したものを元に作成した。2つの正解データを作成したため、協力者には1人当たり計57回の比較を行ってもらった。

### 4.2. 正解データとの比較結果

提案アルゴリズムで推定したレシピの難易度順と正解データの難易度順とをスパイアマンの順位相関で比較した。正解データ2との比較による相関係数は0.59、正解データ3との比較では0.37であった。一方、正解データ1との比較では0.74と強い相関があることを示した。

表1. カテゴリに分類された単語の種類

カテゴリ	食材	調味料	調理器具	調理技法
単語の種類	3,072	2,824	48	47

### 4.3. 考察

手軽さを求める提案アルゴリズムが、技術的な難しさ順として作成した正解データ1よりも手軽さ順の正解データ2や3と比較した方が悪い結果が得られてしまったことは想定外であった。特に制限（豚肉を使っていること）を設けた正解データ3を用いた結果が悪くなった原因は特定しなければならない。

アルゴリズム側で考えられる原因としては、現時点では3.1節の(4)の係数を1で固定してしまっていることや、食材や調味料の表記のゆれを考慮していないことなどが挙げられる。また、正解データ側で考えられる原因としては、人によって手軽さの基準が異なる可能性が挙げられる。3人による比較の評価は、ほとんどが3人も一致すると予想していたが、実際には正解データ2では190の比較のうち105の比較で評価が2:1に割れていた。正解データ3においても110の比較で評価が割れていた。

ただしこれらの原因は、正解データ3だけが特に良くない原因とはならない。今回作成した正解データがたまたま提案アルゴリズムの結果と一致しなかっただけかもしれない。あるいは、制限を設けた正解データは精度が低くなる傾向があるのかもしれない。そのため、さらに多くの正解データを作成し、検証する必要がある。

## 5. おわりに

本稿では、クックパッドのレシピ本文中の単語を用いてレシピの難易度を推定する手法について述べた。今後はさらに多くの正解データを用意して検証を繰り返し、アルゴリズムの精度を向上させたい。

### 謝辞

本研究では、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。また、JSPS 科研費 25330364 の助成を受けた。

### 参考文献

- [1] Miyoshi, Y., Suzuki, K., Shiota, K. and Okamoto, R.: Evaluation of Difficulty Estimation for Learning Materials Recommendation, Liu, C.-C. et al. (Eds.): Proc. of the 22nd International Conference on Computers in Education (ICCE2014), pp.71-76, 2014.
- [2] Miyoshi, Y., Fujisaki, Y., Suzuki, K., Shiota, K. and Okamoto, R.: Estimating the Difficulty of Cooking Recipes by Analyzing User-Recipe Relationship in the Social Network, Proc. of World Conference on E-Learning (E-Learn2015), pp.420-425, 2015.
- [3] クックパッド株式会社: “ニュースリリース”, [https://info.cookpad.com/pr/news/press\\_2014\\_1015](https://info.cookpad.com/pr/news/press_2014_1015), 2014. (2016/2/8 現在)
- [4] 石原和幸, 上田真由美, 平野靖, 梶田将司, 間瀬健二: “FF-IRFを用いた個人嗜好レシピ推薦手法の有効性検証”, 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, pp.51-56, 2008.
- [5] 矢嶋亜紗美, 小林一郎: “個人の状況を考慮した“かんとたん”なレシピの推薦”, 日本知能情報ファジィ学会ファジィシステムシンポジウム講演論文集, pp.31-31, 2009.
- [6] 岩原信九郎: “教育と心理のための推計学”, 日本文化科学社, 1965.