

## レシピ検索のためのクラスタリングによる調理難易度推定 Estimating Cooking Difficulty Level by Clustering for Recipe Search

高橋 里紗<sup>†</sup> 高畑 眞子<sup>†</sup> 来住 伸子<sup>†</sup>  
Risa Takahashi Mako Takahata Nobuko Kishi

### 1. はじめに

様々なレシピ検索サイトがある。しかし、これらのサイトでは、レシピの難易度をはっきり分けていないため、ユーザに合った難易度のレシピを探しにくいという問題がある。本研究では、ユーザの調理技能に合うレシピ検索など、様々な検索ができるレシピ検索ツールの作成を目指した[1]。その一環として、調理難易度推定を行なった。推定に使用するレシピは、【味の素パーク】レシピ大百科[2]の人気レシピ上位 100 件のレシピである。これらのレシピから、工程数、調理時間などを抽出し、それらの値を利用してクラスタリングを行ない、難易度を推定した。評価実験として、アンケートによる難易度と本研究で推定した難易度の比較などを行なった。

### 2. 関連研究

岩本ら[4]は、調理動作の難易度に着目し、難易度ランキング順にレシピを表示するレシピ検索システムを提案した。レシピの難易度は、調理動作数と料理技能検定に基づいて定めた調理動作難易度別に重みから算出した。牧野ら[5]は、調理動作を用いた難易度算出方法を提案した。文部科学省の検定を経た家庭科の教科書に基づいて調理動作難易度を割り当て、その難易度の重みや頻出度などを考慮した 9 つの計算手法を提案した。矢嶋ら[6]は、個人の状況に特化したレシピ推薦システムを提案した。研究では、レシピの“かんたん”さを表すために、調味料・食材・動詞から、ポイントを算出して付与した。永井ら[7]はレシピ動画の時間特性に基づき、レシピの難易度を判定する手法を提案した。レシピ動画から作業種類の多さ・調理動作の並列性・編集意図の 3 つの要素を取り出し、設定した難易度定義に基づいてレシピ難易度を算出した。

### 3. 調理難易度の推定方法

レシピ大百科から取得した 100 件のレシピデータから、調理時間・工程数・食材数・調理動作数を抽出した。この 4 つの指標を説明変数としてクラスタリングを行ない、レシピを分類し、各レシピが分類されたクラスタから難易度を推定した。

#### 3.1 使用データ

レシピ大百科のレシピデータは、クックパッド[3]のようなユーザ投稿型のレシピサイトと比べて、レシピの言葉の表記が統一されている。そのため、調理動作数は、調理手順テキストの句読点の数を数えることで、簡易的に算出できる。調理動作数とレシピの句読点数が対応していることを図 1 に示す。調理時間・工程数・食材数は、レシピが掲載されているページからスクレイピングで抽出した。

<sup>†</sup> 津田塾大学 Tsuda University

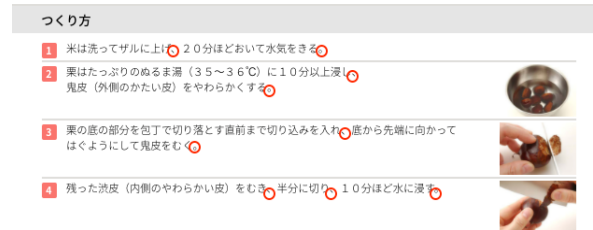


図 1 レシピ大百科のレシピの「作り方」

#### 3.2 クラスタリング方法

Python のオープンソース機械学習ライブラリである scikit-learn を利用してクラスタリングを行なった。

##### 3.2.1 階層的クラスタリング

代表的な階層的的手法として、ward 法、最長距離法、群平均法、最短距離法がある。それぞれの手法で分析したところ、ward 法、最長距離法での分析はクラスタを分けることができた。本報告では、クラスタ分けが成功した 2 つの手法のうち、代表して ward 法を取り上げる。ward 法によるクラスタリング結果を図 2 に示す。

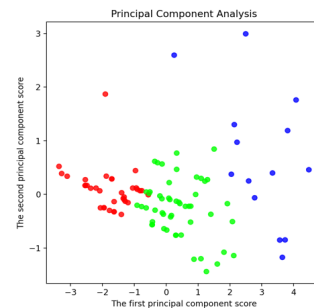


図 2 ward 法によるクラスタリング結果

##### 3.2.2 非階層的クラスタリング

代表的な非階層的的手法の k-means 法を改良したアルゴリズムである k-means++法でもクラスタリングを行なった。

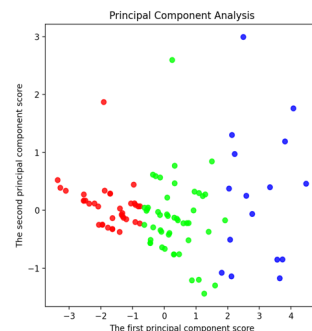


図 3 k-means++法によるクラスタリング結果

エルボー法によりクラスタを 3 と推定し、クラスタリングした結果を図 3 に示す。

### 3.3 難易度の推定

3.2.1, 3.2.2 でクラスタが 3 つに分けられたことから、難易度を「かんたん」「ふつう」「むずかしい」の 3 段階と分けることにした。この 3 段階の難易度をクラスタ番号に対応させた。この対応づけは、よく知られているレシピの難易度と図 2, 3 を参考にしている。

## 4. 評価実験

本研究の対象としたレシピの一部について、アンケート調査で難易度について尋ね、クラスタリングで推定した難易度と比較した。

### 4.1.1 実験内容

料理を作る頻度が異なる 20 代から 70 代の男女 18 名を対象にアンケートを実施した。

アンケートは、目的ごとに 3 つのセクションに分け、最後にレシピの難易度を判断する際に重視したポイントについて回答してもらった。対象としたレシピは、セクションごとに 6 件の計 18 件(重複 2 件)である。

セクション A: 難易度を 3 段階で判断してもらうことで、クラスタリングによる難易度の妥当性を評価

セクション B: k-means++法と ward 法で算出した難易度が異なったレシピについて、どちらの手法で付与した難易度がより妥当かを 2 択で評価

セクション C: レシピ大百科のサイトで「#簡単・時短」タグがついているレシピと比較

### 4.1.2 評価結果

アンケート結果で最も票が多かった難易度を正解とし、クラスタリング結果の正答数を算出した結果を図 4 に示す。なお、セクション A は 6 問であるが、A-4 のレシピは正解を定められないため除いた。よって、セクション A の問題数は 5 問、全体の問題数を 17 問として算出した。ward 法による難易度のほうが、k-means++法よりも人が考える難易度に近い結果となった。

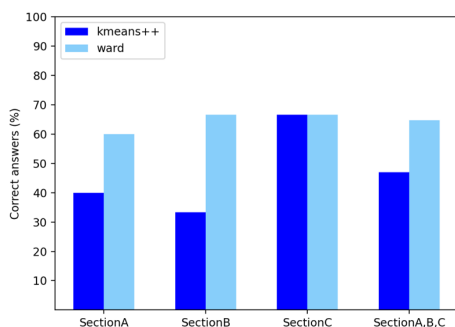


図 4 アンケート結果を正解として算出した正答率

## 5. 考察

人の実感に合う難易度を完全に推定できたわけではなかったが、アンケート調査の結果からは、ward 法を用いたクラスタリングによって、60 パーセント程度の確率で難易度が算出できると考えられる。

人の考える難易度とクラスタリング結果が異なった理由として、推定に他の指標や要素が必要である可能性が考えられる。例えば「大根のべっこう煮」は、クラスタリングでは「ふつう」または「むずかしい」であるが、アンケート結果は「かんたん」が多かった。クラスタリングで「ふつう」以上の難易度と判定した理由として、工程数・食材数・調理動作数は少ないが、調理時間が長いという理由がある。しかし、アンケート結果を個別にみると、調理時間を難易度判断のポイントに挙げた 6 人中 4 人が、A-5 でこのレシピの難易度を「かんたん」としていた。A-6, B-6 の「ホットケーキ」に関して、調理時間が長く動作数が多いレシピのため、クラスタリングでは「ふつう」または「むずかしい」だが、アンケート結果では「かんたん」または「ふつう」が妥当な難易度である。これらの結果は、難易度を推定するには、今回採用した 4 つの指標以外の要素も必要であることを示唆している。

## 6. まとめ

【味の素パーク】レシピ大百科の人気レシピ上位 100 件のデータを対象とし、工程数などによるクラスタリングによって、レシピを分類した。分類をもとにレシピの難易度を推定し、評価実験としてアンケート調査を行なったところ、ward 法を用いたクラスタリングによって、60 パーセント程度の確率で難易度が算出できると考えられる。レシピの難易度推定には、レシピテキストから推定できる難易度に加えて、料理経験等の多くの要素が影響すると考えられ、人の実感に合うレシピ難易度を定めることは困難であった。しかし、ごく簡単な指標を用いた難易度推定方法として、クラスタリングによる推定は有用な可能性があると考えられる。今後の課題として、推定に用いる指標を増やすことなどが挙げられる。

### 参考文献

- [1] 高畑真子, 高橋里紗, 来住伸子, “レシピ検索のため栄養バランスの可視化”, 第 20 回情報科学技術フォーラム(FIT2021), D-023 (2021).
- [2] 【味の素パーク】レシピ大百科, <https://park.ajinomoto.co.jp/recipe/>, (参照 2020-10-6)
- [3] クックパッド『レシピ検索 No.1/料理レシピ載せるならクックパッド』, <https://cookpad.com/>, (参照 2020-10-6)
- [4] 岩本純也, 宮森恒, “調理の難易度を考慮したレシピ検索システムの提案”, DEIM Forum 2012 E1-3 (2012).
- [5] 牧野望, 塩井隆円, 楠和馬, 波多野堅治, “調理動作に基づく料理レシピ検索のための難易度算出法の提案” DEIM Forum 2017 C4-5 (2017).
- [6] 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”, DEIM Forum 2009 E5-6 (2009).
- [7] 永井廉人, 大杉隆文, 松下光範, “料理レシピ動画の時間特性抽出による難易度判定”, JAAI2018 203-OS-24b-01 (2018).