

# 概念階層を考慮した拡張 TF-IDF 法に基づく料理タイプの分析

水谷 真子<sup>†</sup> 尾崎 知伸<sup>†</sup>

<sup>†</sup> 日本大学 文理学部

## 1. はじめに

レシピ検索投稿型サイトであるクックパッド<sup>☆1</sup>には、日々多くのレシピが投稿されている。クックパッドに投稿される各レシピには、タイトルや食材、調理工程が記されているが、レシピを区別する一つの基準として、和風や洋風といった「〇〇風」、すなわち料理タイプが考えられる。本研究では、タイトルに料理タイプを含むレシピに注目し、各料理タイプの特徴を求めることで、タイトル中に「〇〇風」を含まないが料理タイプに合致するレシピの検索を実現することを目指す。より具体的には、料理タイプ間の概念階層を考慮した上で、料理タイプごとの食材や調理法-食材対の TF-IDF 値を求めることで、料理タイプを特徴づけることを試みる。

## 2. 関連研究

これまでに、料理レシピ間の類似度やレシピの特徴抽出に関する研究が数多く行われている。池尻ら<sup>1)</sup>は、TF-IDF の考えを応用した RF-IIF (Recipe Frequency-Inverted Ingredient Frequency) を導入し、食材の意外度を算出する手法を提案した。また福本ら<sup>2)</sup>は、食品ごとの概念階層や食品群を考慮することで、食品の重要度を考慮したレシピ間類似度を提案した。一方、横井ら<sup>3)</sup>は、各料理カテゴリに属するレシピで使用される食材集合に主成分分析を適用し、食材の典型度を求める手法を提案した。Pinxteren ら<sup>4)</sup>は、料理カテゴリに基づき重みを考慮したユークリッド距離を用いてレシピ間距離を求める手法を提案した。さらに Wang<sup>5)</sup>らは、中華料理のレシピをモデル化し、その特徴を明らかにした。

このように、先行研究では食材や分量、料理手順などのレシピ情報に基づいて料理カテゴリ間の類似性を求めている。それに対し本研究では、料理タイプに注目し、料理タイプ毎の食材や手順に基づいてその特徴を抽出することに焦点を当てる。

## 3. データセット

クックパッド株式会社と国立情報学研究所より提供されているクックパッドデータセット<sup>☆2</sup>において、レシピタイトルに料理タイプが含まれるレシピ約 9.8 万件から、図 1 に示す料理タイプに属するレシピを対象とした。なお、料理タイプ数は約 130 である。

レシピを特徴づける食材としては、対象データセット中で 10 以上のレシピに出現する約 1800 種を対象とした。なお、

事前にカタカナに変換した上で表記ゆれへの対応を行っている。一方、調理手順情報から、調理法と対象となる食材の対を抽出し、料理タイプを特徴づける属性として利用する。調理法-食材対の抽出には、係り受け解析器 Cabocha<sup>☆3</sup>を利用した。具体的には、レシピ用語辞書<sup>☆4</sup>に登録されている動詞が、事前に抽出した約 1800 の食材に係る場合、両者を対として抽出する。また、「手順 (1) で混ぜたものを焼く」など、手順情報においては中間物等が記号化されることも多いが、これらはすべて中間物として扱うこととした。得られた調理法-食材対の数は約 4000 である。

料理タイプには、日本風(和風)と関西風など、内容的にはある程度類似しているタイプが存在する。しかし単純に分析してしまうと、表層的な表記が異なるため、両者は全く異なる料理タイプとして認識されてしまう。この問題に対処するため、料理タイプ間に概念階層を導入することを行った。実際の階層を図 1 に示す。図中において、タイプに続く数字は該当するレシピ数である。また今回の研究では、アジア風とヨーロッパ風をそれぞれ根とし、国・地域・料理の順に概念階層があると仮定している。

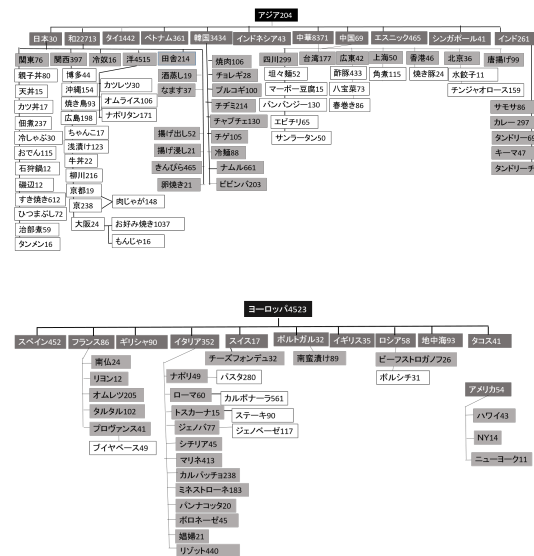


図 1 料理タイプに関する概念階層

## 4. TF-IDF 法とその拡張

本研究では、一般的な TF-IDF 法の考え方を、料理タイプの分析へと応用する。具体的には、ある料理タイプに属するレシピの集合を文書、食材または調理法-食材対を単語と考える。これにより、各料理タイプを食材または調理法-食

Analysis of cuisine types by an extended tf-idf for hierarchical data by Mako Mizutani and Tomonobu Ozaki (College of Humanities and Sciences, Nihon University)

<sup>☆1</sup> <https://cookpad.com/>

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

<sup>☆3</sup> <https://taku910.github.io/cabocha/>

<sup>☆4</sup> <http://plata.ar.media.kyoto-u.ac.jp/data/recipe/cookpaddata.html>

食材を属性とするベクトルとして表現でき、各料理タイプにのみ頻出する食材（または調理法-食材対）を特徴的な食材（調理法-食材対）として特定することができると同時に、料理タイプ間の類似性を算出することが可能となる。また、基本的な TF-IDF 法を概念階層を考慮するように拡張することで、概念階層において下位に位置する料理タイプの影響を考慮した分析を行う。以下、形式的な定義を行う。

各レシピ  $r$  を食材または調理法-食材対の集合  $I_r$  と料理タイプ  $t_r$  の対を用いて  $r = \langle I_r, t_r \rangle$  と表記する。また全料理タイプの集合を  $T$ 、全レシピの集合を  $R$ 、料理タイプ  $t$  を持つレシピの集合を  $R_t = \{ \langle I_r, t_r \rangle \in R \mid t_r = t \}$  と表記する。概念階層において、料理タイプ  $t$  の  $n$  段階下位までの料理タイプの集合を  $T_t^n$  とし、 $T_t^n$  中のいずれかの料理タイプを持つレシピの集合を  $R_t^n = \bigcup_{t' \in T_t^n} R_{t'}$  と表記する。

以上の準備の下、概念階層を考慮した拡張 TF-IDF 法として、 $n$  段階までの下位概念を考慮した料理タイプ  $t$  における食材または調理法-食材対  $i$  の評価値を以下のように計算する。すなわち、料理タイプ  $t$  を考える際に、TF の計算においては  $n$  段階下位の料理タイプまでを一塊として考え、IDF の計算においては  $n$  段階下位の料理タイプを無視することである。

$$TFIDF(t, i, n1, n2) = TF(t, i, n1) \times IDF(t, i, n2)$$

$$TF(t, i, n) = \frac{|\{ \langle I_r, t_r \rangle \in R_t \cup R_t^n \mid i \in I_r \}|}{|R_t \cup R_t^n|}$$

$$IDF(t, i, n) = \log \frac{|T| - |T_t^n|}{\sum_{t' \in T \setminus T_t^n} TF(t', i, 0)}$$

### 5. 実験

提案した拡張 TF-IDF 法の有効性を評価するため、与えられたレシピに対する料理タイプの予測問題を考える。レシピ  $r = \langle I_r, t_r \rangle$  に対し、 $r$  の料理タイプの予測値を

$$\hat{t}_r = \arg \max_{t \in T} \sum_{i \in I_r} TFIDF(t, i, n1, n2)$$

とし、正答数 ( $t_r = \hat{t}_r$  となるレシピ数) を集計した。結果を表 1 に示す。なお、食材を属性とした場合の全レシピ数は 57,535、調理法-食材対を属性とした場合のレシピ数は 56,450 である。

表 1 料理タイプの予測結果 (正答数)

手法	食材	調理法-食材対
TFIDF(t,i,0,0)	5,995	3,529
TFIDF(t,i,1,0)	5,438	3,281
TFIDF(t,i,2,0)	5,467	3,277
TFIDF(t,i,0,1)	6,325	3,622
TFIDF(t,i,0,2)	6,317	3,624
TFIDF(t,i,1,1)	5,438	3,281
TFIDF(t,i,2,2)	5,438	3,277

結果より、属性として食材を利用する方が、調理法-食材対を利用する、より良い結果となっていることが分かる。ま

た、食材、調理法-食材対の双方において、IDF のみを変更した場合に最も正解数が大きくなっていることが分かる。

次に、料理タイプ間の関連性を確認するため、多次元尺度法を用いて、料理タイプを 2 次元に配置した。図 2 に、調理法-食材対の  $TFIDF(t, i, 0, 2)$  を対象としたコサイン距離を用いた場合の結果を示す。

今回、オムライスやカツカツなどの洋食は、日本風の下位概念として配置したが、分析の結果、ヨーロッパ風により近いという結果となった。また、全体的にアジア系の料理タイプが近くに配置される一方で、中・韓風と日本風の距離が大きくなった。

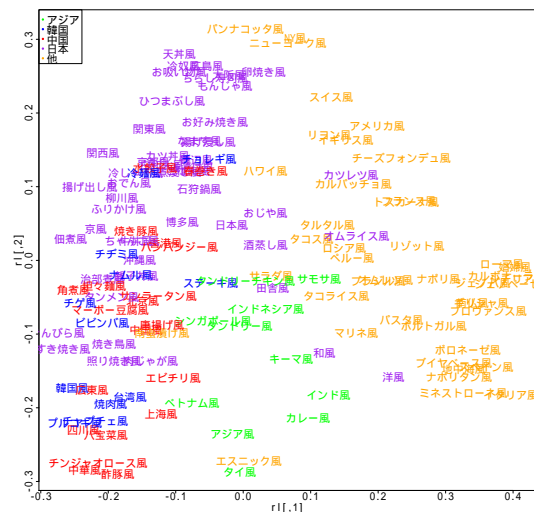


図 2 多次元尺度法による料理タイプ間の関連性の可視化

### 6. おわりに

本研究では、食材や調理法-食材対を用い、料理タイプの特徴づけを行った。今後は、料理タイプ予測の精度向上等を目指し、更なる分析を進める予定である。

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。

### 参考文献

- 1) 池尻恭介, 清雄一, 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, コンピュータソフトウェア Vol.31, No.3, p.3.70-3.78, 2014.
- 2) 福本亜紀, 井上悦子, 中川優: 食材の重要度と食品群を考慮したレシピ間類似度の算出手法, DEIM Form 2012 D9-2, 2012.
- 3) 横井聡, 道満恵介, 平山高嗣, 井出一郎, 出口大輔, 村瀬洋: 料理レシピにおける食材の組み合わせの典型度分析, 信学技報 Vol. 114, No, 487, MVE2014-84, pp.49-54, 2015.
- 4) Y. van Pinxteren, G. Geleijnse and P. Kamsteeg: Deriving a recipe similarity measure for recommending healthful meals, Proc. of the 16th International Conference on Intelligent User Interfaces, pp.105-114, 2011.
- 5) L. Wang, Q. Li, N. Li, G. Dong and Y. Yang: Substructure similarity measurement in chinese recipes, Proc. of the 17th International Conference on World Wide Web, pp.979-988, 2008.