

## 料理タイプに特徴的な食材集合の抽出について

高 霞<sup>†</sup> 尾崎 知伸<sup>†</sup>

<sup>†</sup> 日本大学 文理学部

### 1. はじめに

クックパッドや楽天レシピに代表されるユーザ投稿型レシピ検索サイトには、日々数多くのレシピが投稿されている。本研究では、ユーザにより投稿されたレシピの内、レシピタイトルに和風・中華風のように「〇〇風」といった料理タイプを含むデータを分析対象とし、各料理タイプに特徴的な食材の組み合わせを抽出することを目的とする。

### 2. データセット

本研究では、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>\*</sup>を利用した。データセットには、レシピタイトルに「風」が含まれるレシピが約12万件存在するが、そこから「風邪」や「風味」など、料理タイプを表さないレシピを取り除き、約9.8万件を特定した。加えて、レシピタイトルに「風」が含まれていないレシピデータを約10万件を比較用データとして抽出した。準備したデータセットから、表1に示す、比較的出现数が多い16の料理タイプを分析の対象とする。

一方、食材データに対しては、特徴がない調味料は削除するとともに、食材名に対する表記ゆれを統一した。その上で、出現数50回以上の食材3,277種を分析の対象とする。

表1 分析対象の料理タイプとレシピ数

料理タイプ	レシピ数	料理タイプ	レシピ数
和風	22713	中華風	8371
関西風	397	四川風	299
すき焼き風	612	韓国風	3434
洋風	4515	ナムル風	661
イタリアン風	415	タイ風	1442
グラタン風	785	スペイン風	452
カルボナーラ風	561	インド風	261
マリネ風	413	ベトナム風	361

### 3. 料理タイプ間の関係性の分析

16種の各料理タイプを、各食材の出現頻度を属性とする3,277次元のベクトルで表現し、統計的分析（主成分分析と対応分析）を用いて料理タイプ間の関係性を分析した。

#### 3.1 主成分分析

主成分分析は、多次元データの情報をできるだけ損わずに低次元空間に縮約する手法である。今回の分析では、各料理タイプと食材を2次元に縮約して平面上に配置することで、データの全体を俯瞰しその関係性の解釈を行った。主成分

分析の結果を図1に示す。図1においては、料理タイプ間、または料理タイプと食材が近い距離にある場合、両者の間に強い関係があることを示している。例えば、洋風には牛乳、じゃがいもが多く使われており、スペイン風とイタリアン風ではコショウ、オリーブオイルが多く使われており、カルボナーラ風とグラタン風ではチーズが特徴的な食材であり、ベトナム風ではナンプレーが特徴的な食材であることが分かる。

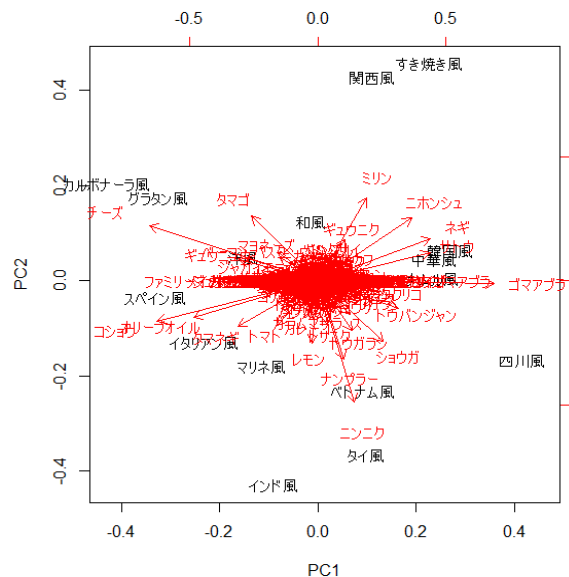


図1 食材と料理タイプの主成分分析の結果

#### 3.2 対応分析

対応分析は、単純な2次元表や多重表の行と列間の対応する測定値を分析する探索的データ解析の手法であり、複雑なデータを単純化して、データ行列の行と列の幾何的図形による解釈ができ、関連性の探索に役立つ利点を持つ。料理タイプに対する対応分析の結果を図2に示す。図2から各料理タイプには差異があることが分かる。それぞれの位置・距離から判断すると、「洋風」、「スペイン風」、「イタリアン風」の間には強い関係があり、また同様に「韓国風」、「ナムル風」、「中華風」の間と「ベトナム風」、「タイ風」の間に高い関係性があることが分かる。一方で、逆の位置に配置されている料理タイプ「すき焼き風」と「インド風」、「カルボナーラ風」と「四川風」には、関係が非常に薄い（逆の関係がある）ことが示唆される。

### 4. 特徴的な食材集合の抽出

主成分分析等によって得られた各料理タイプにおける特徴的な食材の妥当性を評価するとともに、特徴的な食材の組み合わせ（食材集合）を抽出するため、頻出パターン・相関

Extraction of characteristic sets of ingredients on recipe types by Xia Gao and Tomonobu Ozaki (College of Humanities and Sciences, Nihon University)

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

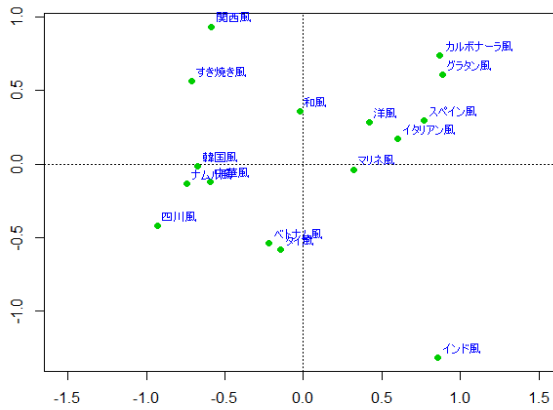


図2 料理タイプの対応分析の結果

ルール分析<sup>4)</sup>を適用した。

各レシピ  $r$  を食材集合  $I_r$  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。なお、料理タイプを持たない場合、疑似的なラベルを用い  $t_r = \phi$  とする。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  $R_t = \{ \langle I_r, t_r \rangle \in R \mid t_r = t \}$ 、 $\phi$  以外の料理タイプを持つレシピの集合を  $R_{-\phi} = \{ \langle I_r, t_r \rangle \in R \mid t_r \neq \phi \}$ 、食材集合  $X$  を含むレシピの集合を  $R^X = \{ \langle I_r, t_r \rangle \in R \mid X \subseteq I_r \}$  と表記する。以上の準備の下、各料理タイプ  $t$  ごとに、以下に示す評価関数に従って高い評価値を持つ食材の組み合わせ  $X$  や、ルール  $X \rightarrow t$  の抽出を行った。なお、Sup は相関ルールにおける支持度、Conf は確信度、AllConf は全確信度 (All Confidence) に相当する。また、料理タイプを持たないレシピを一般的なレシピとし、それとの違いを算出することで、より特徴的な組み合わせの抽出を試みている。

- (1)  $Sup(X, t) = |R^X \cap R_t| / |R_t|$
- (2)  $SupD(X, t) = Sup(X, t) - Sup(X, \phi)$
- (3)  $Conf(X \rightarrow t) = |R^X \cap R_t| / |R^X \cap R_{-\phi}|$
- (4)  $AllConf(X, t) = \min_{i \in X} |R^X \cap R_t| / |R^{(i)} \cap R_t|$
- (5)  $AConfD(X, t) = AllConf(X, t) - AllConf(X, \phi)$

ところで、ある料理タイプにおいて特徴的な食材  $A$  と、そうではない食材  $B$  との組み合わせを考える場合、上記の評価関数の値は  $A$  に大きく影響されることが懸念される。このことを避け、組み合わせたときにはじめて特徴的となる状況を捉えるため、下記に示す差分の差分に着目した評価関数も準備した。

- (1)  $SupD^2(X, t) = SupD(X, t) - \max_{P \subset X} SupD(P, t)$
- (2)  $AConfD^2(X, t) = AConfD(X, t) - \max_{P \subset X} AConfD(P, t)$

16種類の料理タイプの内、洋風に関する各評価値上位の

結果を表2に示す。結果より、洋風においては「ギョウニユウ」、「チーズ」、「コショウ」は特徴的食材(集合)として抽出されていることが分かる。特に「ギョウニユウ」に関しては、前述の主成分分析の結果と一致しており、ある程度の整合性が確認された。

表2 洋風における評価値上位の食材集合

順位	Conf	Sup	SupD
1	[チーズ, ライス]	[コショウ, タマネギ]	[コショウ, コンソメ]
2	[トリニク, ライス]	[コショウ, コンソメ]	[オリーブオイル, コショウ]
3	[ギョウニユウ, チーズ]	[オリーブオイル, コショウ]	[コショウ, タマネギ]
4	[ギョウニユウ, ニク]	[コショウ, ニンジン]	[コンソメ, タマネギ]
5	[ギョウニユウ, ナマクリーム]	[タマネギ, ニンジン]	[コショウ, ニンジン]
6	[ギョウニユウ, ライス]	[コショウ, チーズ]	[コショウ, チーズ]
7	[ギョウニユウ, コンソメ]	[コンソメ, タマネギ]	[コンソメ, ニンジン]
8	[ギョウニユウ, トリニク]	[コショウ, ニンニク]	[タマネギ, ニンジン]
9	[ギョウニユウ, トマト]	[コショウ, タマゴ]	[コショウ, ベーコン]
10	[ギョウニユウ, パター]	[コショウ, パター]	[コンソメ, ベーコン]

順位	SupD <sup>2</sup>	AConfD	AConfD <sup>2</sup>
1	[ゴマアブラ, サトウ]	[ノリ, ライス]	[キリボシダイコン, ベーコン]
2	[サトウ, ショウガ]	[ギョウニユウ, コショウ]	[オカラ, コショウ]
3	[コムギコ, ニホンシユ]	[ギョウニユウ, タマゴ]	[コンソメ, ミズ]
4	[ネギ, ミリン]	[コショウ, ライス]	[ウドン, コショウ]
5	[サトウ, ナマクリーム]	[キリボシダイコン, ベーコン]	[コンソメ, ハクサイ]
6	[ゴマアブラ, ニホンシユ]	[オリーブオイル, シオコウジ]	[コショウ, ミズ]
7	[サトウ, ハクリキコ]	[コショウ, シオコウジ]	[コショウ, ゴーキ]
8	[ス, タマゴ]	[オカラ, コショウ]	[アブラアゲ, コショウ]
9	[タマゴ, ミリン]	[コンソメ, ミズ]	[コンソメ, ダイコン]
10	[ス, レモン]	[アゴアブラ, コショウ]	[オカラ, コンソメ]

16種類の料理タイプに関して各上位30件の食材集合を、3人の被験者に提示し、その妥当性を定性的に評価した。その結果、概ね特徴的な食材集合が抽出できていることが確認できた。

## 5. おわりに

本研究では、食材や食材集合の出現頻度に着目し、料理タイプとの関係性を分析した。一定の知見は得られたが、より大規模な評価や定量的評価、評価関数間の違いや優劣等に関する評価などが、今後の課題として残されている。加えて、調理法や調理手順なども料理タイプを特徴づける大きな要因と考えられる。今後はこれら食材とは異なる情報を用いて料理タイプを特徴づけることを計画している。

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, 中川博之, 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, 31(3), 2014
- 2) 横井聡, 道満恵介, 平山高嗣, 井手一郎, 出口大輔, 村瀬洋: 料理レシピにおける食材の組合せの典型度分析, 電子情報通信学会技術研究報告 IMQ, イメージ・メディア・クオリティ 114(485), p.49-54, 2015.
- 3) 苅米志帆乃, 藤井教: 料理どうしの類似と組合せに基づく関連レシピ検索システム, 言語処理学会第14回年次大会発表論文集, p.959-962, 2008
- 4) R. Agrawal, T. Imielinski, A. N. Swami: Mining Association Rules between Sets of Items in Large Databases. Proc. SIGMOD Conference 1993, pp.207-216 (1993)