

食材と時季に基づいた料理レシピの人気分析

Popularity Analysis of Cooking Recipes Based on Ingredients and Seasons

池田真大 熊野雅仁 小堀聡 木村昌弘*

Masahiro Ikeda, Masahito Kumano, Satoshi Kobori, Masahiro Kimura

龍谷大学理工学部電子情報学科

Department of Electronics and Informatics, Ryukoku University

Abstract: 近年、料理レシピ共有サイトが世界的に普及しており、かつては観測が困難であった一般家庭の料理を反映して大量に投稿されたレシピ群がビッグデータとして成長するに従い、人々が料理に対して行う行動の背後に潜む原理や未知の要因を理解することを目指す上で、各地域の料理レシピに含まれる特に利用食材などの観点から人々の日常生活における料理レシピを科学的に分析する研究が注目されている。我々は、日本における一般家庭の料理に対して人々が行う行動と料理レシピとの関係に注目しており、特に、人気を得る料理レシピの要因について食材だけでなくその利用時季および順序性に注目している。そこで本研究では、日本の料理レシピ共有サイトに投稿された大量のレシピデータに着目し、人気を得る料理レシピの要因として、食材そのもの、ある時季に頻出する食材、食材の利用順序性、ある時季に頻出する食材の利用順序性という4つの観点から料理レシピの人気分析を行う。そして、日本の料理レシピ共有サイト CookPad の実データを用いた実験で、人気分析の結果を報告する。

1 はじめに

ソーシャルメディアの発展は、人々の日常生活に大きな影響を与えているだけでなく、実世界の人々の日常の動向を捉えるための様々な分析を行う研究にも影響を与えつつある。近年、料理レシピ共有サイトが世界的に普及しており、これまで観測が困難であった一般家庭の料理を反映して大量に投稿されたレシピ群がビッグデータとして成長するに従い、一般的な人々が料理に対して行う行動の背後に潜む原理や未知の要因を理解することを目指す上で、利用食材などの観点から人々の日常生活における料理レシピを科学的に分析する研究 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] が活発化している。

Ahn ら [2] は、食材どうして共有される香料化合物に着目したネットワークを分析し、西洋料理では香料化合物を共有した食材ペアを利用するという食材ペアリング仮説の検証を世界規模で行った。Jain ら [5] は、Ahn らの研究の応用としてインド料理の香料化合物ネットワークを分析した。Teng ら [3] は、レシピ内における共起食材のネットワークとレシピ作成における代替食材のネットワークを用い、一部の食材を代替食材に置き換えたレシピが人気を得るかを調べ、レシピ推薦へ

応用した。Sajadmanesh ら [8] は、レシピ内の食材群に着目することで、世界各地の郷土料理を判別することができる可能性を示した。これらの研究は、人々が料理に対して行う行動の背後に潜む原理や未知の要因を理解する上でレシピに含まれる食材群に着目するアプローチの可能性を押し広げていると考えられる。

ところで、Ahn ら [2] は、西洋料理が多くの香料化合物を共有した食材ペアを利用する傾向を持つことを明らかにしたが、東アジア料理が香料化合物を共有した食材ペアの利用を避ける傾向を持つことも示した。東アジアに含まれる日本の和食の特徴は、食材そのものの持ち味や新鮮さを生かすことや、季節や伝統的な行事などの時季を重んじる傾向がある（詳しくは [11] を参照）。しかし、世界の様々な料理が取り込まれている日本において、その傾向が料理レシピ共有サイトに投稿される一般家庭料理の動向に反映されているかは不明である。また、我々は、人気を得る料理レシピの要因に注目しているが、人気の要因としては、一般的にレシピ投稿者に関する人気に左右される場合もあれば、見た目の良さ、食材そのものの人気、ある時季によく使われる食材を用いることでの人気など、これら以外にも様々な要因が存在し得ると考えられる。特に我々は、料理の味や風味、色合いなどに関連すると思われる調理における順序性にも注目している。本研究では、日本の料理レシピ共有サイトに投稿された大量

*連絡先：龍谷大学理工学部電子情報学科
〒520-2194 大津市瀬田大江町横谷 1-5
E-mail: kimura@rins.ryukoku.ac.jp

表 1: 2×2 分割表

	D_1	\overline{D}_1
a	$n_{a,1}$	$n_{a,\bar{1}}$
\bar{a}	$n_{\bar{a},1}$	$n_{\bar{a},\bar{1}}$

のレシピデータに着目し、人気を得る料理レシピの要因を調べる第一歩として、食材そのものに着目した食材ユニグラム、ある時季での利用状況が活発化する食材に着目したアクティブ食材ユニグラム、また、順序性として、ある食材とその次に使われる食材に着目した食材バイグラム、ある時季での出現状況が活発化する食材バイグラムに着目したアクティブ食材バイグラム、という4つの観点から料理レシピの人気分析を行う。そして、日本の料理レシピ共有サイト CookPad の実データを用いた実験で、人気分析の結果を報告する。

2 分析法

対象とする日本の料理レシピ共有サイトにおいて、ここ数年間に投稿されたレシピ全体の集合 D を考える。レシピ総数を $|D| = N$ とする。 D 内の各レシピを調査し、どのようなレシピが人気を得たかについて分析する。 D に属するレシピのうち、人気を得たレシピ全体の集合を D_1 とし、それ以外を \overline{D}_1 とする。すなわち、 $D = D_1 \cup \overline{D}_1$, $D_1 \cap \overline{D}_1 = \emptyset$ である。本研究では、レシピに関する属性 a を導入し、レシピ $d \in D$ に対して、 d が属性 a をもつことと d が人気レシピであること (i.e., $d \in D_1$) に関係性があるかどうかを、表1の様な 2×2 分割表を作成しカイ二乗検定に基づいて調べる。ここに、 $n_{a,1}$ は属性 a をもつ D_1 の元の数、 $n_{a,\bar{1}}$ は属性 a をもつ \overline{D}_1 の元の数、 $n_{\bar{a},1}$ 属性 a をもたない D_1 の元の数、 $n_{\bar{a},\bar{1}}$ は属性 a をもたない \overline{D}_1 の元の数それぞれ表す。このとき、 $N = n_{a,1} + n_{a,\bar{1}} + n_{\bar{a},1} + n_{\bar{a},\bar{1}}$ である。本研究では、特に、利用食材や時季に関わるようなレシピ属性 a について考える。食材全体の集合を V とする。

2.1 食材ユニグラムの分析

まず、その食材が利用されたことがレシピの人気獲得に大きく貢献したというような食材 (すなわち、食材ユニグラム) の抽出を目指し、レシピに関する属性 a として、特定の食材がレシピで利用されている (すなわち、食材ユニグラムがレシピに出現する) という考えを考える。各食材 $v \in V$ に対して、カイ二乗値 $\chi^2(v)$ と対応する p 値 $p(v)$ を求め、レシピの人気獲得と関係

性があったと考えられるような食材の集合 V_1 を抽出し分析する。

2.2 アクティブ食材ユニグラムの分析

2.2.1 概要

日本には四季があり旬の食材を利用したレシピは人気を獲得する傾向があると推測されるので、次に、レシピの人気獲得と関係するような旬食材の抽出について考える。我々は、季節変化の周期性を考慮して、各食材 $v \in V$ のレシピ出現頻度の時系列データを数年にわたって調べ、 v がその平年における平均頻度と比べて顕著に出現しているような時季を検出して、それを v のアクティブ時季 \mathcal{I}_v と呼び、 v の旬の時季とみなすことにする。そして、食材 $v \in V$ とそのアクティブ時季 \mathcal{I}_v の組 (v, \mathcal{I}_v) 全体の集合 \mathcal{B} を考える。任意の $(v, \mathcal{I}_v) \in \mathcal{B}$ に対して、食材 v がアクティブ時季 \mathcal{I}_v 内に利用されていることをレシピに関する属性 a とし、カイ二乗値 $\chi^2(v, \mathcal{I}_v)$ と対応する p 値 $p(v, \mathcal{I}_v)$ を求め、レシピの人気獲得と関係性があったと考えられるような食材とそのアクティブ時季の組の集合 \mathcal{B}_1 を抽出し分析する。

2.2.2 アクティブ時季の検出法

各食材 $v \in V$ に対して、そのアクティブ時季 \mathcal{I}_v の検出を考える。本研究では、食材ネットワーク分析において我々 [9, 10] が与えた食材ペアのアクティブ期間抽出法を応用して、以下の様に \mathcal{I}_v を求める。

まず、時間単位を指定すると、1年は区間 $[0, T)$ 上の整数点の集合で表すことができることに注意する。ここで正整数 T は、その指定した時間単位における1年の長さである。本研究では、平年を円 $\mathbb{R}/T\mathbb{Z}$ 上の格子点の集合、

$$\mathcal{S} = \{[t] \in \mathbb{R}/T\mathbb{Z} \mid t = 0, 1, \dots, T-1\}$$

と同一視する。ここに、 \mathbb{R} は実数全体の集合で \mathbb{Z} は整数全体の集合である。平年 \mathcal{S} における時季という概念を次のように定義する。任意の整数 t, t' , $(0 \leq t, t' \leq T-1, t \neq t')$ に対して、時季 $[[t], [t']]$ は、 $t < t'$ のとき $[t, t')$ ($\subset [0, T)$) であり、 $t > t'$ のとき $[t, T) \cup [0, t')$ ($\subset [0, T)$) である。我々は、食材 v のアクティビティに関する観測時系列データ、

$$\mathcal{X}(v) = \{x_{[t]}(v) \mid [t] \in \mathcal{S}\}$$

に基づいて \mathcal{I}_v を検出する。ここに、 $x_{[t]}(v)$ は時間ステップ $[t]$ に投稿されたレシピのうち v を利用していたものの総数である。また、時間ステップ $[t]$ に投稿されたレシピの総数を $N_{[t]}$ とし、時間ステップ $[t]$ に投稿さ

れるレシピに食材 v が利用されている確率を $P(v|t)$ とする。

各 $t = 0, 1, \dots, T-1$ に対して, $x_{[t]}(v)$ はパラメータが $N_{[t]}$ と $P(v|t)$ である二項分布から生成されると仮定する。さらに, v のアクティビティは J_v 個の時季に分割できると仮定する。ここに, J_v は正の整数である。 $J_v = 1$ の場合, v はアクティビティ変化点をもたないと言い, $\mathcal{T}^1(v) = \emptyset$ とし,

$$P(v|t) = p^1(v), \quad \forall t \in \mathcal{S}$$

とする。 $J_v \geq 2$ の場合, v はアクティビティ変化点の集合,

$$\mathcal{T}^{J_v}(v) = \{t_1(v), \dots, t_{J_v}(v)\}$$

($0 \leq t_1(v) < \dots < t_{J_v}(v) \leq T-1$) をもち, それに対応する J_v 個の確率の組,

$$\mathbf{p}^{J_v}(v) = (p_1(v), \dots, p_{J_v}(v))$$

が存在すると仮定する。ただし, $j = 1, \dots, J_v$ に対して,

$$P(v|t) = p_j(v), \quad \forall t \in \mathcal{S}_j(v) = [t_j(v), t_{j+1}(v))$$

であり, また, $p_j(v) \neq p_{j+1}(v)$, $t_{J_v+1} = t_1(v)$, $p_{J_v+1}(v) = p_1(v)$ である。このとき,

$$\mathcal{S} = \bigcup_{j=1}^{J_v} \mathcal{S}_j(v) \quad (\text{disjoint union})$$

が成り立つことに注意する。 $(\mathcal{T}^{J_v}(v), \mathbf{p}^{J_v}(v))$ を v のアクティビティ変化パターンと呼び, 各 $\mathcal{S}_j(v)$ を v の第 j 時季と呼ぶ。

観測時系列データ $\mathcal{X}(v)$ からアクティビティ変化パターン $(\mathcal{T}^{J_v}(v), \mathbf{p}^{J_v}(v))$ を検出することを考える。 $J_v = 1$ の場合, $\mathcal{X}(v)$ の対数尤度は,

$$\mathcal{L}^1(\mathcal{X}(v); p^1(v)) = \sum_{[t] \in \mathcal{S}} \{x_{[t]}(v) \log p^1(v) + (N_{[t]} - x_{[t]}(v)) \log(1 - p^1(v))\}$$

であるので, $p^1(v)$ の最尤推定値 $\hat{p}^1(v)$ は,

$$\hat{p}^1(v) = \frac{\sum_{[t] \in \mathcal{S}} x_{[t]}(v)}{\sum_{[t] \in \mathcal{S}} N_{[t]}}$$

で与えられる。また, $J_v \geq 2$ の場合, $\mathcal{X}(v)$ の対数尤度は,

$$\begin{aligned} \mathcal{L}^{J_v}(\mathcal{X}(v); \mathcal{T}^{J_v}(v), \mathbf{p}^{J_v}(v)) \\ = \sum_{j=1}^{J_v} \sum_{[t] \in \mathcal{S}_j(v)} \{x_{[t]}(v) \log p_j(v) \\ + (N_{[t]} - x_{[t]}(v)) \log(1 - p_j(v))\} \end{aligned}$$

であるので, $\mathcal{T}^{J_v}(v)$ が指定された時, $\mathbf{p}^{J_v}(v)$ の最尤推定値 $\hat{\mathbf{p}}^{J_v}(v) = (\hat{p}_1(v), \dots, \hat{p}_{J_v}(v))$ は,

$$\hat{p}_j(v) = \frac{\sum_{[t] \in \mathcal{S}_j(v)} x_{[t]}(v)}{\sum_{[t] \in \mathcal{S}_j(v)} N_{[t]}} \quad (j = 1, \dots, J_v)$$

で与えられる。我々は, 目的関数として,

$$\begin{aligned} f(\mathcal{T}^{J_v}(v)) \\ = \mathcal{L}^{J_v}(\mathcal{X}(v); \mathcal{T}^{J_v}(v), \hat{\mathbf{p}}^{J_v}(v)) - \mathcal{L}^1(\mathcal{X}(v); \hat{p}^1(v)) \end{aligned}$$

を定義し, f を最大化することにより最適なアクティビティ変化点の集合 $\mathcal{T}^{J_v}(v)$ を検出する (詳細は [9, 10] を参照)。

さて, 観測時系列データ $\mathcal{X}(v)$ から, アクティビティ変化パターン $(\mathcal{T}^{J_v}(v), \hat{\mathbf{p}}^{J_v}(v))$ が検出されたとする。まず, $\hat{p}^1(v)$ は平年において v がレシピで利用される平均頻度を表し, 各 $\hat{p}_j(v)$ は第 j 時季 $\mathcal{S}_j(v)$ において v がレシピで利用される平均頻度を表していることに注意する。よって, 各 $j = 1, \dots, J_v$ に対して, $\hat{p}_j(v) > \hat{p}^1(v)$ ならば $\mathcal{S}_j(v)$ はアクティブであると呼び, $\hat{p}_j(v) \leq \hat{p}^1(v)$ ならば $\mathcal{S}_j(v)$ は非アクティブであると呼ぶ。我々は, アクティブな $\mathcal{S}_j(v)$ の全体を v のアクティブ時季 \mathcal{I}_v として検出する。

2.3 食材バイグラムの分析

料理レシピは食材リストを含んでいるだけではなく, 一般に, 調理における食材の利用の仕方を説明しており, その食材の利用の仕方には順序性があると考えられる。その食材の順序性がレシピの人気獲得に大きく貢献したというような食材の利用順序 (すなわち, 食材バイグラム) の抽出を目指し, レシピに関する属性 a として, 特定の食材 $u \in V$ の次に $v \in V$ が利用されている (すなわち, 食材バイグラムがレシピに出現する) ということを考える。ここに, 特定の食材 $u \in V$ の次に $v \in V$ が利用されることを $u \rightarrow v$, ($u \neq v$) と表す。

図 1 は, 食材リストと調理手順が記載された一般的な料理レシピから食材バイグラムを抽出する処理過程を示している。まず, 事前準備の処理について述べる。料理レシピの食材リストには, 野菜などの主食材やそれ以外の食材 (調味料・風味調味料) だけでなく「キッチンペーパー」などの道具として使う材料が含まれる場合がある。本研究では, 道具として使われる材料を食材リストから事前に除外した。次に, 調理手順が記載された文章は順序立てて (番号順に) 説明が行われている。その順序に沿って説明された各文章を順序性を保ちながら 1 つの文章として連結し, 形態素解析を行って順序性を維持しながら名詞群を抽出する。事前準備

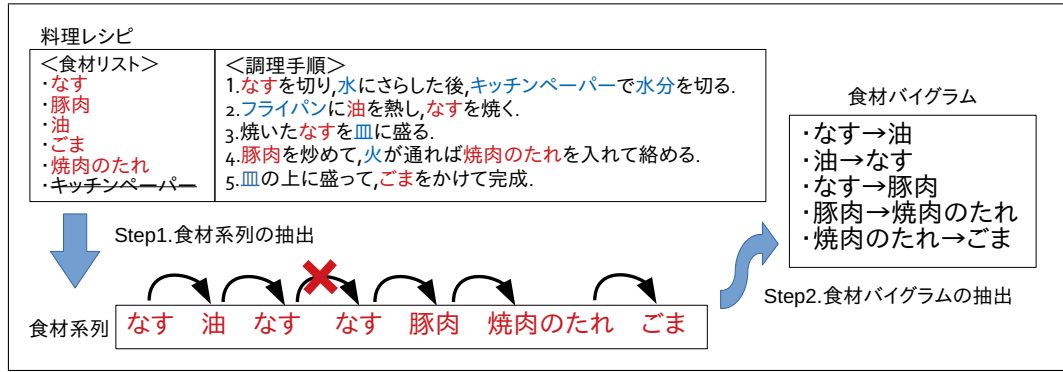


図 1: 食材バイグラムの抽出手順

備を終えると、Step 1 として料理レシピそれぞれの説明文書から抽出された名詞群とその食材リストに現れる各食材とを照合し、食材リストに存在する食材（図 1 の赤字で表された食材）のみを順序性を保ちながら食材系列として生成する。なお、図 1 の調理手順を説明する文章では、主食材や調味料・風味調味料以外に、青い字で表した「フライパン」、「皿」、「火」などの名詞が存在しているが、これらの名詞は例示した料理レシピの食材リストには存在しないため、食材系列には加わらないことに注意しておく。ただし、文書中の「水」は、食材リストに含まれていれば食材系列に加わることに注意しておく。次に、Step 2 では、食材系列に沿って先に現れた食材から隣り合う次の食材への有向リンクを結び、これを食材バイグラム $u \rightarrow v$ とする。ただし、「ナス」から「豚肉」や「ナス」から「ごま」など、食材系列上で 2 つ以上離れた食材どうしの順序関係については本研究では対象外とした。また、「ナス」から「ナス」など、同じ食材が続く場合は食材バイグラムから除外している。以上の条件のもとに食材バイグラムの集合を抽出する。各食材バイグラムに対して、カイ二乗値 $\chi^2(u \rightarrow v)$ と対応する p 値 $p(u \rightarrow v)$ を求め、レシピの人気獲得と関係性があったと考えられるような食材バイグラムの集合を抽出し分析する。

2.4 アクティブ食材バイグラムの分析

本研究において、調理における食材の利用の仕方に順序性があると考え、食材バイグラムを抽出しているが、アクティブ食材ユニグラムにおいて時季性を考えたように、食材バイグラムについても時季性が存在し得ることが予想される。そこで、レシピの人気獲得と関係するような旬の食材バイグラムとしてアクティブ食材バイグラムの抽出を行う。食材バイグラムのアクティブ時季の検出については、アクティブ食材ユニグラムにおけるアクティブ時季の検出法を適用し、食材バイグラム $u \rightarrow v$ とアクティブ時季 $\mathcal{I}_{u \rightarrow v}$ の全体の集合

\mathcal{C} を考える。任意の $(u \rightarrow v, \mathcal{I}_{u \rightarrow v}) \in \mathcal{C}$ に対して、食材バイグラム $u \rightarrow v$ がアクティブ時季 $\mathcal{I}_{u \rightarrow v}$ 内に利用されていることをレシピに関する属性 a とし、カイ二乗値 $\chi^2(u \rightarrow v, \mathcal{I}_{u \rightarrow v})$ と対応する p 値 $p(u \rightarrow v, \mathcal{I}_{u \rightarrow v})$ を求め、レシピの人気獲得と関係性があったと考えられるような食材とそのアクティブ時季の組の集合を抽出し分析する。

3 実験

3.1 実験設定

日本の料理レシピ共有サイト「Cookpad」の 2011 年 1 月から 2013 年 12 月までの 3 年間に投稿された料理レシピに関する実データを用いた。Cookpad では、レシピが投稿された時間および利用された食材、調理手順などの情報が得られる。Cookpad の料理レシピに含まれる食材の名前には「たまご」「卵」「玉子」のように漢字やひらがな、カタカナなどによる表記揺れが存在するため、食材用語に関する辞書を作成することで表記揺れを手動で解消し、調味料・風味調味料を含む 489 種類の食材を抽出した。また、料理レシピに用いられる食材や手順の傾向は「肉料理」や「野菜料理」、「お菓子」など、料理のジャンルによって異なり得ると考えられることから、実データの期間において料理ジャンルに関する情報を持つ料理レシピのみを抽出し、料理ジャンルごとに分析を行う。本研究では、Cookpad に存在する料理カテゴリーを料理ジャンルと見なし、レシピ数が多い 11 種類の料理カテゴリーに焦点を当てるが、本稿では、頁数の制限から、特に「野菜のおかず」カテゴリーについて分析を行った結果を報告する。「野菜のおかず」カテゴリーに含まれるレシピ数 $N=14,590$ であった。また、「野菜のおかず」カテゴリーに含まれる料理レシピ群について、料理レシピごとにその投稿日から 1 年間に付与されたつくレポ数の多さに基づいて人気レシピとそうでないレシピを仕分ける。本研究で

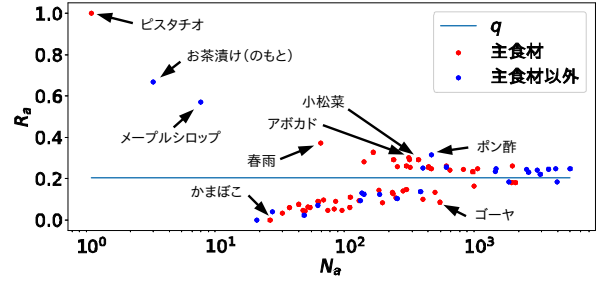
表 2: 食材ユニグラムの χ^2 値ランキング

(a) 主食材 Top 10.

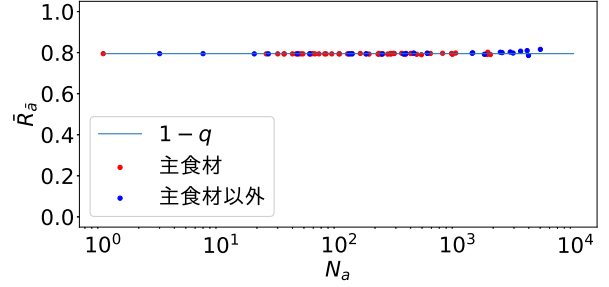
主食材	χ^2	R_a	\bar{R}_a	N_a
ゴーヤ	45.448	0.085	0.790	494
ごま	38.322	0.261	0.802	1,799
昆布	25.770	0.099	0.792	363
菜の花	16.217	0.084	0.793	178
アボカド	16.211	0.301	0.796	282
豚ばら肉	15.647	0.292	0.797	336
しいたけ	14.961	0.133	0.792	451
パプリカ	13.861	0.106	0.793	226
小松菜	13.635	0.293	0.796	287
かにかま	13.633	0.327	0.796	150

(b) 主食材以外 Top 10.

主食材以外	χ^2	R_a	\bar{R}_a	N_a
醤油	80.299	0.247	0.816	5,014
砂糖	59.133	0.248	0.810	3,875
油	45.059	0.246	0.807	3,413
ボン酢	32.332	0.315	0.798	425
酒	27.888	0.242	0.803	2,787
みりん	25.072	0.244	0.802	2,297
塩	17.381	0.183	0.786	3,979
片栗粉	16.027	0.248	0.799	1,341
ワイン	14.640	0.104	0.793	230
塩こしょう	12.186	0.232	0.800	2,380



(a) R_a と N_a の関係.



(b) \bar{R}_a と N_a の関係.

図 2: 食材ユニグラムの分析結果

は, 各料理カテゴリーの料理レシピ総数に対するつくレポ数の平均値を上回るものを人気レシピ集合 D_1 と定めて分析を行った. 料理レシピ総数 $|D| = N$ に対する人気レシピ数 $|D_1| = n_{a,1} + n_{\bar{a},1}$ の比を,

$$q = |D_1|/|D|$$

と表す. $q = 0.205$ であった.

分析を行うにあたり, 属性 a をもつレシピの数を,

$$N_a = n_{a,1} + n_{a,\bar{1}}$$

と表し, そのようなレシピが人気レシピになる確率を,

$$R_a = n_{a,1}/N_a$$

と表す. 属性 a をもつレシピについては, $R_a > q$ ならば R_a の値が大きいほど人気レシピになる傾向が強く, $R_a < q$ ならば R_a の値が小さいほど人気レシピにならない傾向が強いと考えられる. また, 属性 a をもたないレシピが人気レシピにならない確率を,

$$\bar{R}_a = n_{\bar{a},1}/(N - N_a)$$

と表す. このとき, 属性 a のカイ二乗値は,

$$\chi^2(a) = q(1-q)N_a(R_a - q)^2 + q(1-q)(N - N_a)\{\bar{R}_a - (1-q)\}^2 \quad (1)$$

と表されることに注意する. 本分析では, カイ二乗検定においてレシピの人気獲得との関係が有意であると判定された属性のみを対象とする.

3.2 食材ユニグラムに基づいた分析

表 2 は, 各食材ユニグラムに対応する χ^2 値でランキングを行った Top 10 を示しており, 表 2a は主食材のみ, 表 2b は主食材以外を抽出しており, それぞれの R_a , \bar{R}_a , N_a を併記している. また, N_a に対して R_a と \bar{R}_a に着目して分析した結果を図 2 に示す. 図 2a および図 2b ともに赤い点は食材ユニグラムのうち主食材を, 青い点は食材ユニグラムのうち調味料や風味調味料などの主食材以外を表している. ここで, 図 2b の傾向から, $\bar{R}_a \simeq 1 - q$ と見なせるため, 式 (1) の第 2 項は無視できることから, 料理レシピの人気獲得に貢献するものとして, $|R_a - q|$ や N_a の大きさが χ^2 値に貢献することがわかる. ただし, 図 2a の傾向から, N_a が大きくなるほど $|R_a - q|$ が小さくなる傾向があるため, N_a が大きすぎても人気獲得には貢献しない傾向があることが見てとれる. 図 2a において, R_a が最も高い食材ユニグラムは主食材の「ピ스타チオ」であった. しかし, そのレシピ数 N_a は非常に少ないことから, 人気レシピの主要な要因とは言いがたい. また, 主食材で R_a が低い「ゴーヤ」や主食材以外の「ワイン」, さらに図 2a から, 「かまぼこ」などについては利用するとむしろ人気に対して逆効果となる傾向がある. 一方, 主食材の「アボカド」や「春雨」(図 2a) などは N_a がある程度数に達しているだけでなく R_a が大きめであり, これらの食材ユニグラムは料理レシピが人気を得ることについてある程度貢献する可能性が期待できると考えられる.

表 3: アクティブ食材ユニグラムのカイニ乗ランキング

(a) 主食材.				
主食材	χ^2	R_a	\bar{R}_a	N_a
ゴーヤ	33.198	0.087	0.791	378
ごま	17.087	0.257	0.798	998
菜の花	14.876	0.085	0.793	165
パプリカ	13.731	0.082	0.793	146
昆布	13.365	0.089	0.793	158
しいたけ	11.911	0.119	0.793	253
小松菜	11.337	0.321	0.796	137
かにかま	11.324	0.336	0.795	107
ほうれんそう	10.323	0.278	0.796	317
玉ねぎ	9.967	0.170	0.791	1,167

(b) 主食材以外.				
主食材以外	χ^2	R_a	\bar{R}_a	N_a
醤油	40.149	0.242	0.807	3,664
砂糖	24.895	0.249	0.801	1,876
酒	16.761	0.251	0.799	1,220
水	14.562	0.243	0.799	1,546
片栗粉	12.979	0.261	0.797	655
粒マスタード	11.839	0.425	0.795	40
みりん	10.761	0.245	0.798	1,036
オリーブオイル	9.407	0.168	0.792	1,039
油	8.573	0.233	0.798	1,667
からし	8.380	0.080	0.794	87

3.3 アクティブ食材ユニグラムに基づいた分析

表 3 に, アクティブ食材ユニグラムにおける主食材 (表 3a) と主食材以外 (表 3b) のカイニ乗ランキングを示す. また, 図 3 は, アクティブ食材ユニグラムに基づいた分析結果である. アクティブ食材ユニグラムについても $|R_a - q|$ や N_a の大きさが χ^2 値に貢献することがわかる. 主食材では「小松菜」, 「かにかま」, 「ほうれんそう」, 「ごま」などが人気獲得にある程度貢献することがわかるが, 図 3a から, 「はんぺん」なども人気獲得への貢献度が高い可能性がある. しかし, 「しいたけ」, 図 3a から「ザーサイ」や「黒豆」, 特に「ゴーヤ」や「玉ねぎ」などは N_a が大きい, すなわち多くの料理レシピで利用されているにもかかわらず, 利用すると人気に対して逆効果であることがわかる. 図 4 は表 3 の各食材 v のアクティブ時季 \mathcal{I}_v を黒い区間として表わしている. 料理レシピの人気獲得への貢献が期待される「小松菜」, 「かにかま」, 「ほうれんそう」, 「ごま」はそれぞれ異なる時季にアクティブになることがわかる. 主食材以外では「粒マスタード」が人気獲得に貢献する度合いが高い傾向を持ち, そのアクティブ時季が2月から4月に限られることがわかる. また, 「酒」, 「みりん」も人気獲得への貢献が期待されるが, 「ごま」

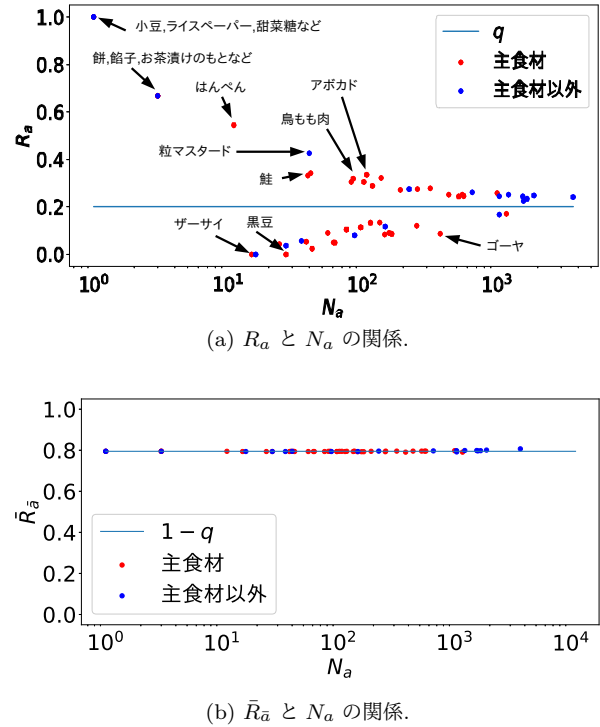


図 3: アクティブ食材ユニグラムの分析結果

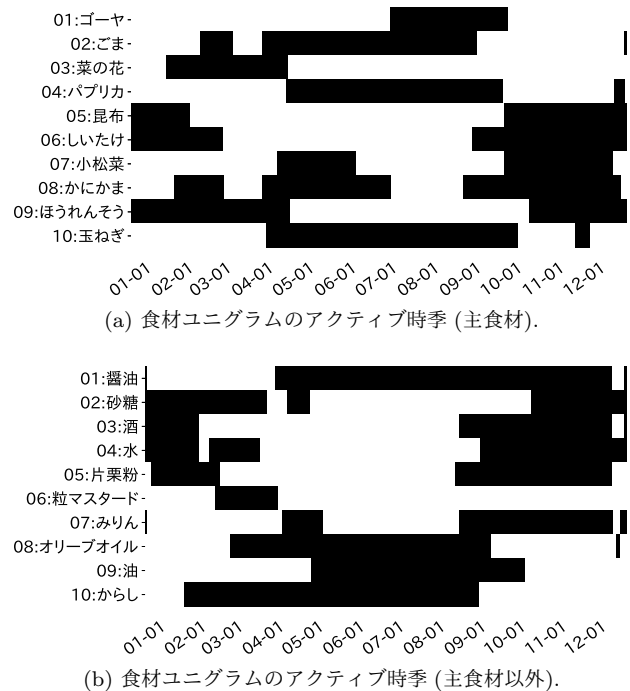
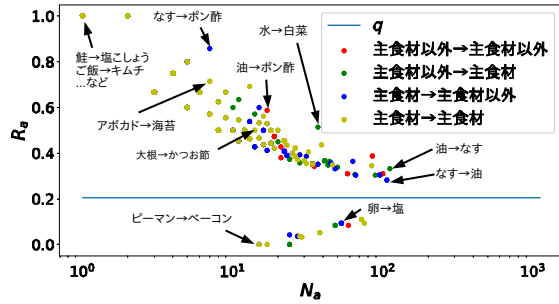
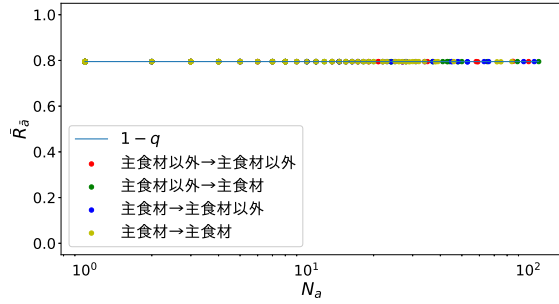


図 4: アクティブ食材ユニグラムの時季

も含め, 年末から年始にかけてアクティブになっており, 「酒」, 「みりん」, 「ごま」は, 日本の伝統文化である正月用の料理に関連した人気レシピの要因となっている可能性が示唆される.



(a) R_a と N_a の関係.



(b) \bar{R}_a と N_a の関係.

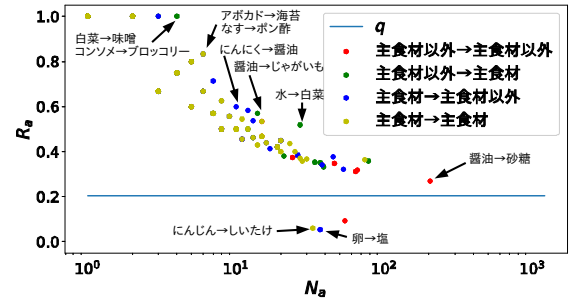
図 5: 食材バイグラムの分析結果

表 4: 食材バイグラムのカイ二乗ランキング

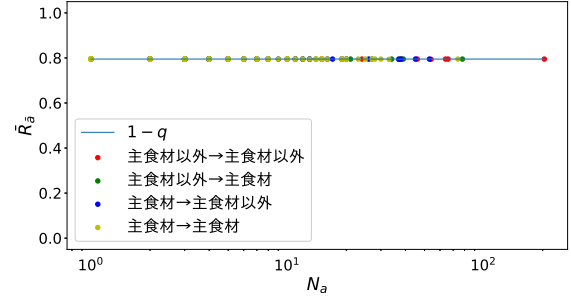
食材バイグラム	χ^2	R_a	\bar{R}_a	N_a	種類
水 → 白菜	21.558	0.514	0.795	37	主食材以外 → 主食材
小松菜 → 卵	18.888	0.692	0.795	13	主食材 → 主食材
なす → ポン酢	18.217	0.857	0.795	7	主食材 → 主食材以外
水 → コンソメ	17.490	0.388	0.796	85	主食材以外 → 主食材以外
油 → ポン酢	15.273	0.588	0.795	17	主食材以外 → 主食材以外
えりんぎ → 油	14.315	0.600	0.795	15	主食材 → 主食材以外
バター → ベーコン	12.716	0.436	0.795	39	主食材 → 主食材
片栗粉 → じゃがいも	12.519	0.636	0.795	11	主食材以外 → 主食材
バター → コーン	12.505	0.562	0.795	16	主食材 → 主食材
海老 → 塩こしょう	11.732	0.667	0.795	9	主食材 → 主食材以外

3.4 食材バイグラムに基づいた分析

図 5 は、食材バイグラムに基づいた分析結果である。また、図 2a の食材ユニグラムと図 5a を比較すると、 $|R_a - q|$ が大きくなっており、人気の獲得への貢献やその逆効果となる食材バイグラムがより顕著に現れる傾向を持つことがわかる。表 4 に食材バイグラムのカイ二乗ランキングを示す。人気への貢献度が高いことが期待されるものとして、「主食材 → 主食材以外」パターンの「なす → ポン酢」や「海老 → 塩こしょう」などがあり、図 5a から、「主食材 → 主食材」パターンの「アボカド → 海苔」、「大根 → かつお節」、「主食材以外 → 主食材以外」パターンの「油 → ポン酢」などがあることもわかる。また、「油 → なす」はその逆の「なす → 油」より人気獲得への貢献度が高いことわかる。一方、「主食材 → 主食材」パターンの「ピーマン → ベーコン」は、利用すると人気に対して逆効果となる傾向をもつことがわかる。



(a) R_a と N_a の関係.



(b) \bar{R}_a と N_a の関係.

図 6: アクティブ食材バイグラムの分析結果

表 5: アクティブ食材バイグラムのカイ二乗ランキング

アクティブ食材バイグラム	χ^2	R_a	\bar{R}_a	N_a	種類
水→白菜	16.236	0.519	0.795	27	主食材以外 → 主食材
サラダ油 → キャベツ	15.471	1.000	0.795	4	主食材以外 → 主食材
アボカド → 海苔	14.493	0.833	0.795	6	主食材 → 主食材
なす → ポン酢	14.493	0.833	0.795	6	主食材 → 主食材以外
コンソメ → ブロッコリー	11.602	1.000	0.795	3	主食材以外 → 主食材
白菜 → 味噌	11.602	1.000	0.795	3	主食材 → 主食材以外
なす → ピーマン	11.573	0.365	0.795	74	主食材 → 主食材
醤油 → じゃがいも	11.495	0.571	0.795	14	主食材以外 → 主食材
油 → なす	11.316	0.359	0.795	78	主食材以外 → 主食材
海老 → 塩こしょう	11.105	0.714	0.795	7	主食材 → 主食材以外

3.5 アクティブ食材バイグラムに基づいた分析

図 6 は、アクティブ食材バイグラムに基づいた分析結果である。また、表 5 に、アクティブ食材バイグラムのカイ二乗ランキングを示す。人気への貢献度が高いことが期待されるものとして、「主食材以外 → 主食材」パターンの「サラダ油 → キャベツ」、「コンソメ → ブロッコリー」、「主食材 → 主食材以外」パターンの「白菜 → 味噌」が非常に高い人気獲得率を持つことがわかる。それらは N_a については相対的に少ないが、利用されると人気レシピとなっていることから、人気レシピの要因として高い貢献度があると考えられる。しかし、「主食材 → 主食材」パターンの「にんじん → しいたけ」や「主食材 → 主食材以外」パターンの「卵 → 塩」などは人気に対して逆効果となる傾向があることがわかる。図 7 は、表 5 の各食材バイグラムのアクティブ時季を示したものである。「サラダ油 → キャベツ」は 4

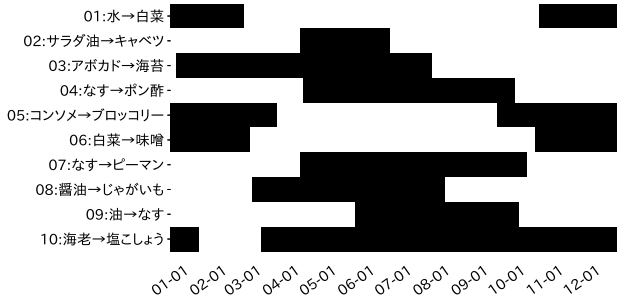


図 7: アクティブ食材バイグラムのアクティブ時季

月前から7月前の期間だけにアクティブ時季があることがわかる。「コンソメ → ブロッコリー」,「白菜 → 味噌」はいずれも秋から春までの期間にアクティブ時季があることがわかる。すなわち, アクティブ食材バイグラムの人気への貢献度が高い時季はか限られる傾向があると考えられるため, 時季とともに順序性に注目することは, 人気分析を行ううえで意義深いことが示唆された。

これらの結果から, 食材バイグラムは食材ユニグラムよりも人気の要因を捉える点でより深い知見が得られる可能性があり, 時季性を捉えることも人気の要因を探求する上で意義深いことを示唆していると考えられる。すなわち, 食材単体の分析を行うだけでなく, 利用食材の調理上の順序性や時季を捉えるような本研究が着目した料理レシピの人気分析における新たな観点は, 人々が料理に対して行う行動の背後に潜む原理や未知の要因を理解する上でレシピに含まれる食材群に着目するアプローチの可能性をより押し広げている点で, 本研究の有効性を示唆していると思われる。

4 まとめ

本研究では, 日本の料理レシピ共有サイトに投稿された大量のレシピデータに着目し, 食材ユニグラム, アクティブ食材ユニグラム, 食材バイグラム, アクティブ食材バイグラムという4つの観点から人気を得る料理レシピとの関連性について分析を行った。そして, 料理レシピの人気分析において, 食材だけでなく, 食材利用の順序性を捉えることの有効性を示唆する結果を示した。今後は, 食材利用の順序性について, 新たな観点を導入し, 料理レシピの人気の要因についてさらなる探求を行う予定である。

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] Kinouchi, O., Diez-Garcia, R. W., Holanda, A. J., Zambianchi, P., and Roque, A. C.: “The nonequilibrium nature of culinary evolution,” *New Journal of Physics*, vol. 10, No. 7, pp. 073020:1–073020:8, 2008.
- [2] Ahn, Y. Y., Ahnert, S. E., Bagrow J., and Barabasi, A. L.: “Flavor Network and the Principles of Food Pairing,” *Scientific Reports*, vol. 1, pp. 196:1–196:7, 2011.
- [3] Teng, C.-Y., Lin, Y.-R., and Adamic, L.: “Recipe recommendation using ingredient networks,” in *Proceedings of WebSci’12*, pp. 298–307, 2012.
- [4] Nedovic, V.: “Learning recipe ingredient space using generative probabilistic models,” in *Proceedings of CwC’13*, pp. 13–18, 2013.
- [5] Jain, A., N K, R., and Bagler, G.: “Analysis of food pairing in regional cuisines of India,” *PLoS ONE*, vol. 10, no. 10, pp. 1–17, 2015.
- [6] Jermurawong, J. and Habash, N.: “Predicting the structure of cooking recipes,” in *Proceedings of EMNLP’15*, pp. 781–786, 2015.
- [7] Jiang, Y., Skufca, J. D., and Sun, J.: BiFold visualization of bipartite datasets, *EPJ Data Science*, vol. 6, pp. 2:1–2:19, 2017.
- [8] Sajadmanesh, S., Jafarzadeh, S., Osia, S. A., Rabbie, H. R., Haddadi, H., Mejova, Y., Musolesi, M., Cristofaro, E. D., and Stringhini, G.: “Kiss-ing cuisines: Exploring worldwide culinary habits on the Web,” in *Proceedings of WWW’17 Companion*, pp. 1013–1021, 2017.
- [9] Kikuchi, Y., Kumano, M., and Kimura, M.: “Analyzing dynamical activities of co-occurrence patterns for cooking ingredients,” in *Proceedings of IEEE ICDMW’17*, pp. 17–24, 2017.
- [10] 菊地悠樹, 熊野雅仁, 木村昌弘, “料理レシピ共有サイトにおける食材のアクティブ共起パターンの抽出,” 情報処理学会論文誌「数理モデル化と応用」, vol. 11, no. 2, pp. 30–40, 2018.
- [11] 熊倉功夫, “和食の神髄: ユネスコ無形文化遺産に登録された本当の理由,” *ヘルシスト*, 226, pp. 2–7, 2014.