

# 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

菊地 悠樹<sup>1</sup> 熊野 雅仁<sup>2,a)</sup> 木村 昌弘<sup>2,b)</sup>

受付日 2017年11月10日, 再受付日 2017年12月27日,  
採録日 2018年2月7日

**概要:** 料理レシピ共有サイトが人気を集めていることや複雑ネットワーク科学の成功もあり, 近年, レシピで使われる食材の組合せの特徴を分析するためのネットワークベース手法の研究が注目されている. 静的な性質を分析する従来アプローチと異なり, レシピ内でいっしょに使われる食材ペアの時季による動的变化を分析することを目指して, 食材ペアの共起アクティビティ変化パターンを検出する有効な手法を提案する. そして, 検出したアクティビティ変化パターンに基づき, 各時間ステップにおいて食材集合のアクティブネットワークを構築し, 食材の主要アクティブ共起パターンを抽出する. さらに, 抽出した食材の主要アクティブ共起パターン群に対して, レシピの観点から解釈する手法を与える. 日本の料理レシピ共有サイトの実データを用いて, 食材ペアのアクティビティ変化パターン検出における提案法の有効性を実証するとともに, 提案法が抽出した食材の主要アクティブ共起パターンの妥当性を示す. さらに提案法を用いて, 日本の料理レシピで使われる食材の組合せに関する季節変化の特徴を明らかにする.

**キーワード:** レシピ分析, 変化パターン検出, 動的ネットワーク分析, 可視化, ソーシャルメディアマイニング

## Extracting Active Co-occurrence Patterns of Cooking Ingredient Usage in Recipe Sharing Sites

YUUKI KIKUCHI<sup>1</sup> MASAHITO KUMANO<sup>2,a)</sup> MASAHIRO KIMURA<sup>2,b)</sup>

Received: November 10, 2017, Revised: December 27, 2017,  
Accepted: February 7, 2018

**Abstract:** Due to the increasing popularity of cooking-recipe sharing sites and the success of complex network science, attention has recently been devoted to developing an effective network-based method of analyzing the characteristics of ingredient combinations used in recipes. Unlike previous approaches dealing with static properties, we aim at analyzing the dynamical changes in ingredient pairs jointly used in recipes, and propose an efficient method of detecting the change patterns for co-occurrence activities of ingredients. Based on the activity change patterns detected, we construct an active network among ingredients at every time-step, and extract active co-occurrence patterns. Moreover, we provide a method of interpreting active co-occurrence patterns in terms of recipes. Using real data from a Japanese recipe sharing site, we demonstrate the effectiveness of the proposed method for extracting the activity change patterns for ingredient pairs, and confirm the validity of active co-occurrence patterns extracted. Moreover, we uncover the characteristics of the seasonal changes in ingredients jointly used in Japanese recipes by applying the proposed method.

**Keywords:** recipe analysis, change pattern detection, dynamical network analysis, visualization, social media mining

<sup>1</sup> 龍谷大学大学院理工学研究科電子情報学専攻  
Division of Electronics and Informatics, Ryukoku University,  
Otsu, Shiga 520–2194, Japan

<sup>2</sup> 龍谷大学理工学部電子情報学科  
Department of Electronics and Informatics, Ryukoku Uni-  
versity, Otsu, Shiga 520–2194, Japan

a) kumano@rins.ryukoku.ac.jp

b) kimura@rins.ryukoku.ac.jp

## 1. はじめに

料理レシピ共有サイトの登場によって, 「人々が普段の家庭料理でどのようなレシピを創作し, そして, それが他の人々にどのように受け入れられているのか」というようなデータが大規模に得られるようになり, レシピ推薦システム

に関する研究が活発化している [1], [2], [3], [4], [5]. また近年, 生物学, 生態学, 社会学などの様々な分野に新たな知見をもたらしている複雑ネットワーク科学 [6], [7] は, 食品科学分野への応用に関しても注目されつつあり [8], [9], [10], [11], Ahn ら [12] や Teng ら [13] は, 料理レシピ共有サイトの大量データを用いてレシピで使われる食材の組合せの特徴を分析するネットワークベース手法を研究している.

Ahn ら [12] は, レシピ内で頻繁に使われる食材の組合せを香料化合物の観点から調べるためのネットワークベース手法を与え, 世界規模で食文化の多様性を分析した. そして, 西洋料理は多くの香料化合物を共有した食材ペアを利用する傾向があり, 一方, 東アジア料理は香料化合物を共有した食材ペアの利用を避ける傾向があることを明らかにした. このように, レシピ内で共起した食材ペアの特徴を分析するネットワークベース手法の開発は, 重要な研究課題である. 本研究では, 静的な性質を分析する従来アプローチと異なり, 食材が季節の影響を強く受けると推察される日本の料理レシピ共有サイトを対象として, そこに投稿されるレシピ内で共起した食材ペアの動的変化を分析することを目指す. 図 1 に, 食材ペア (ごま, ゴーヤ) の共起アクティビティの変動を示す<sup>\*1</sup>. 食材ペアのレシピ内での共起については, アクティブな期間とそうでない期間が存在すること, また, そのような期間の構造 (いくつかの期間があり, それらの始まりと終わりはいつであるか) は食材ペアに依存していることが示唆される. したがって我々は, 各食材ペア  $(u, v)$  に対して, 平年における共起アクティビティの変化パターンを調べることを考える. そのようなパターンは, アクティビティ変化点の集合とそれに対応する共起確率の組を検出することでとらえられる.

本論文では, 各食材ペア  $(u, v)$  に対して,  $u$  と  $v$  がいっしょに使われたレシピの数についての観測時系列データから, アクティビティ変化パターンを検出する手法を提案する. そして, 検出したアクティビティ変化パターンに基づいて, 各時間ステップにおける食材集合のアクティブネッ

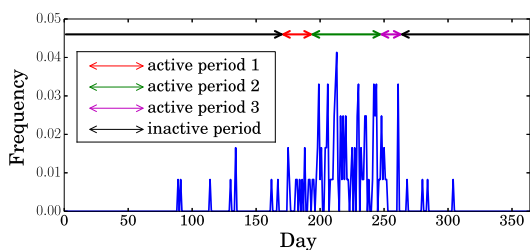


図 1 平年における食材ペア (ごま, ゴーヤ) の共起アクティビティの変動

Fig. 1 Co-occurrence activity of ingredient pair “sesame and bitter melon” in an average year.

<sup>\*1</sup> ここでは, 日本の料理レシピ共有サイトである「Cookpad」(<http://cookpad.com/>) に 2011 年と 2012 年に投稿されたレシピを調べた. 図 1 では, レシピ内でのごまとゴーヤの共起頻度を日の関数として表示している.

トワークを構築する. ここに, ノードは食材に対応し, リンク  $(u, v)$  の重みは, その時間ステップにおける食材  $u$  と  $v$  の共起バースト度であり,  $u$  と  $v$  の共起頻度が年間平均と比べてその時間ステップではどのくらい異常に大きいかを表している. 次に, 構築した食材集合のアクティブネットワークを主要なモジュールに分解することにより, 各時間ステップにおける食材の主要アクティブ共起パターンを抽出する. ここで, 与えられた時間ステップにおける食材集合のアクティブネットワークとは, その時間ステップに同じレシピで用いられる頻度が年間平均と比べ高い食材ペアから構成される食材ネットワークである. そして, その時間ステップにおける食材の主要アクティブ共起パターンとは, そのアクティブネットワーク内のコミュニティを形成する部分ネットワークを意味しており, その時間ステップにおける食材の特徴的な利用パターンを表し, ある種の料理レシピジャンルを表すと考えられる. 我々は, 季節変化の周期性を考慮して, 各食材ペアのレシピでの共起頻度時系列データを数年にわたって調べることにより, 平年に対して食材の主要アクティブ共起パターン群を抽出する. したがって, それらが季節に依存した食材の旬の利用パターン群をとらえる可能性もあるということに注意する. 一般にデータマイニングでは, その結果を解釈可能な形で提示することが求められ, 可視化分析やヒューマン・コンピュータ・インタラクションのようないくつかのフレームワークが利用されている [14], [15], [16], [17], [18], [19], [20], [21]. 本論文では, 抽出した食材の主要アクティブ共起パターン群に対して, 各パターンをレシピの観点から解釈する手法を与える. 日本の料理レシピ共有サイトの実データを用いて, 提案法を評価する. まず, 食材ペアのアクティビティ変化パターン検出において, 提案法はベースライン法よりも性能が優れていることを実証する. 次に, 提案法が抽出した食材の主要アクティブ共起パターンの妥当性を評価するとともに, 日本の料理レシピで使われる食材の組合せに関する季節変化の特徴を明らかにする.

本論文の構成は次のとおりである. 2 章では, 問題設定および記号法について述べる. 3 章ではレシピデータから食材の主要アクティブ共起パターン群を抽出してそれを分析する手法を提案し, 4 章では実データを用いた評価実験の結果を報告する. 5 章では, 関連研究について論じる. 6 章はまとめである.

## 2. 準備

この章では, 問題設定について述べ, 本論文を通して用いる記号法を導入する.

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. たとえば, 日を時間単位に採用する場合, (区間  $[0, 1)$  に相当する) 時間ス

ステップ0は1月1日であり、(区間  $[T-1, T]$  に相当する) 時間ステップ  $T-1$  は12月31日である。日本には春夏秋冬の4つの季節があり、これらが毎年繰り返されている。料理で使われる食材の多くは、このような季節による影響を受け、独自の旬期間を持っている。本研究では、食材の使われ方を平年において季節性の観点から分析する。そこで、平年を円  $\mathbb{R}/T\mathbb{Z}$  上の格子点の集合

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

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り、商集合  $\mathbb{R}/T\mathbb{Z}$  では、任意の  $t \in [0, T)$  に対して、 $[t] = [t+nT] \in \mathbb{R}/T\mathbb{Z} (\forall n \in \mathbb{Z})$  である。平年  $S$  における期間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、期間  $[[t], [t']]$  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。

与えられた料理レシピ共有サイトについて、ここ数年間に投稿されたレシピで使われた食材を調査し、どのような食材の組合せが平年  $S$  において利用されているかを分析する。食材の集合を  $V$  で表す。各  $t = 0, \dots, T-1$  に対して、時間ステップ  $[t]$  に投稿された料理レシピの集合を  $\mathcal{R}_{[t]}$  とし、 $\mathcal{R}_{[t]}$  の要素数を  $N_{[t]}$  とする。すなわち、 $N_{[t]} = |\mathcal{R}_{[t]}|$  である。任意の異なる食材  $u, v \in V$  に対して、時間ステップ  $[t]$  に投稿されたレシピで  $u$  と  $v$  の両方を含んでいたものの数を  $x_{[t]}(u, v)$  とする。また、時間ステップ  $[t]$  に投稿されるレシピでの  $u$  と  $v$  の共起確率を  $P(u, v | [t])$  で表す。 $\mathcal{X}(u, v)$  を  $u$  と  $v$  の両方を含むレシピの数に関する観測時系列データとする。すなわち、

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

である。

### 3. 提案法

任意の  $u, v \in V (u \neq v)$  に対して、まず、食材ペア  $(u, v)$  のアクティビティ変化パターンを検出する。異なる食材の各ペア  $(u, v) (u, v \in V)$  に対して、まず、そのアクティビティ変化パターンを検出する。次に、検出した変化パターンに基づき、各時間ステップ  $[t] \in S$  において、食材集合  $V$  のアクティブネットワーク  $G_{[t]} = (V, \omega_{[t]})$  を構築し、食材の主要アクティブ共起パターンを抽出する。ここに、 $\omega_{[t]}(u, v)$  はノード  $u$  とノード  $v$  の間の無向リンク  $(u, v)$  の重みであり、食材  $u$  と食材  $v$  の時間ステップ  $[t]$  での共起バースト度を表している。

#### 3.1 食材ペアのアクティビティ

異なる食材のペア  $(u, v)$  を任意に固定し、そのアクティビティについて調べる。

まず、各  $t = 0, 1, \dots, T-1$  に対して、 $x_{[t]}(u, v)$  はパラ

メータが  $N_{[t]}$  と  $P(u, v | [t])$  である二項分布から生成されていると仮定する。さらに、ペア  $(u, v)$  のアクティビティは  $J(u, v)$  個の期間に分割できると仮定する。ここに、 $J(u, v)$  は正の整数である。 $J(u, v) = 1$  の場合、 $(u, v)$  はアクティビティ変化点を持たないといい、 $\mathcal{T}^1(u, v) = \emptyset$  とし、

$$P(u, v | [t]) = p^1(u, v), \quad \forall [t] \in S \quad (1)$$

とする。 $J(u, v) \geq 2$  の場合、 $(u, v)$  はアクティビティ変化点の集合

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

を持ち、それに対応する  $J(u, v)$  個の共起確率の組

$$\mathbf{p}^{J(u, v)}(u, v) = (p_1(u, v), \dots, p_{J(u, v)}(u, v)) \quad (2)$$

が存在すると仮定する(図2参照)。ただし、 $[t] \in S_j(u, v) \subset S (j = 1, \dots, J(u, v))$  に対して、

$$P(u, v | [t]) = p_j(u, v)$$

であり、また、 $j = 1, \dots, J(u, v)$  に対して、

$$0 \leq t_1(u, v) < \dots < t_{J(u, v)}(u, v) \leq T-1, \quad (3)$$

$$S_j(u, v) = [[t_j(u, v), [t_{j+1}(u, v)]], \quad (4)$$

$$t_{J(u, v)+1} = t_1(u, v),$$

$$0 \leq p_j(u, v) \leq 1,$$

$$p_j(u, v) \neq p_{j+1}(u, v),$$

$$p_{J(u, v)+1}(u, v) = p_1(u, v)$$

である。式(4)より、

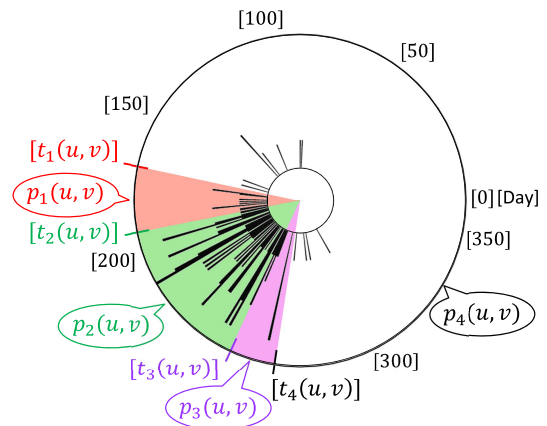


図2  $J(u, v) = 4$  の場合のアクティビティ変化パターン  $(\mathcal{T}^4(u, v), \mathbf{p}^4(u, v))$  の例。時間単位に日を用い、平年を円で表している。動径方向の棒は各時間ステップ  $[t]$  で食材ペア  $(u, v)$  が含まれるレシピ数  $x_{[t]}(u, v)$  を表している

Fig. 2 Example of activity change patterns  $(\mathcal{T}^4(u, v), \mathbf{p}^4(u, v))$  in the case of  $J(u, v) = 4$ , where, day is used as the unit of time and the circle shows an average year. Each bar in radial directions indicates the number  $x_{[t]}(u, v)$  of recipes including ingredient pair  $(u, v)$  at time-step  $[t]$ .

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

が成り立つ。\$(\mathcal{T}^{J(u,v)}(u,v), \mathbf{p}^{J(u,v)}(u,v))\$ を \$(u,v)\$ のアクティビティ変化パターンと呼び、各 \$j\$ に対して \$\mathcal{S}\_j(u,v)\$ を \$(u,v)\$ の第 \$j\$ 期間と呼ぶ。ここで、式 (1) の \$p^1(u,v)\$ と式 (2) の \$p\_1(u,v)\$ の相違点について強調しておく。\$p^1(u,v)\$ に関しては、\$J(u,v) = 1\$、すなわちアクティビティ変化点がないこと、そして平年において、任意の時間ステップでの食材ペア \$(u,v)\$ の平均共起確率が \$p^1(u,v)\$ であることを表している。一方 \$p\_1(u,v)\$ に関しては、アクティビティ変化パターンが \$J(u,v) (\ge 2)\$ 個の期間に分割され、その第 1 期間の共起確率が \$p\_1(u,v)\$ であることを表している。

### 3.2 アクティビティ変化パターンの検出

異なる食材の各ペア \$(u,v)\$ に対して、観測時系列データ \$\mathcal{X}(u,v)\$ からアクティビティ変化パターン \$(\mathcal{T}^{J(u,v)}(u,v), \mathbf{p}^{J(u,v)}(u,v))\$ を検出する手法を与える。以下で検出法を述べる際には、表記を単純にするために \$(u,v)\$ を省略する。

\$J = 1\$ の場合、\$\mathcal{X}\$ の対数尤度は、

$$\begin{aligned} \mathcal{L}^1(\mathcal{X}; p^1) &= \sum_{[t] \in \mathcal{S}} \{x_{[t]} \log p^1 + (N_{[t]} - x_{[t]}) \log(1 - p^1)\} \quad (5) \end{aligned}$$

であるので、\$p^1\$ の最尤推定値 \$\hat{p}^1\$ は、

$$\hat{p}^1 = \frac{\sum_{[t] \in \mathcal{S}} x_{[t]}}{\sum_{[t] \in \mathcal{S}} N_{[t]}} \quad (6)$$

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

$$\begin{aligned} \mathcal{L}^J(\mathcal{X}; \mathcal{T}^J, \mathbf{p}^J) &= \sum_{j=1}^J \sum_{[t] \in \mathcal{S}_j} \{x_{[t]} \log p_j + (N_{[t]} - x_{[t]}) \log(1 - p_j)\} \quad (7) \end{aligned}$$

であるので、\$\mathcal{T}^J\$ が指定されたとき、\$\mathbf{p}^J\$ の最尤推定値 \$\hat{\mathbf{p}}^J = (\hat{p}\_1, \dots, \hat{p}\_J)\$ は、

$$\hat{p}_j = \frac{\sum_{[t] \in \mathcal{S}_j} x_{[t]}}{\sum_{[t] \in \mathcal{S}_j} N_{[t]}} \quad (j = 1, \dots, J) \quad (8)$$

で与えられる。我々は、式 (6) を式 (5) に式 (8) を式 (7) にそれぞれ代入することによって、目的関数

$$f(\mathcal{T}^J) = \mathcal{L}^J(\mathcal{X}; \mathcal{T}^J, \hat{\mathbf{p}}^J) - \mathcal{L}^1(\mathcal{X}; \hat{p}^1) \quad (9)$$

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\$ を検出することを考える。

まず、網羅的に探索することで 2 つのアクティビティ変化点の最適集合 \$\mathcal{T}^2\$ を見つける。ところで、このような網羅的な探索法は、その計算量が \$O(T^J)\$ であることから、

\$J \ge 3\$ ならば実用的でないことに注意する。したがって、固定された \$J \ge 3\$ に対して最適解 \$\mathcal{T}^J\$ を得るために、貪欲法を適用することを考え、それに局所探索を組み込んだ手法を採用する。また、\$J\$ の値を推定するために、\$\chi^2\$ 検定を利用することを考える。

まず、\$J\$ の値を推定する手法について述べる。\$J \ge 2\$ に対して、最適集合 \$\mathcal{T}^J, \mathcal{T}^{J+1}\$ が得られているとする。\$J\$ 個のアクティビティ変化点モデルに対する \$J+1\$ 個のアクティビティ変化点モデルの対数尤度比

$$\begin{aligned} g(J+1) &= \mathcal{L}(\mathcal{X}; \mathcal{T}^{J+1}, \hat{\mathbf{p}}^{J+1}) - \mathcal{L}(\mathcal{X}; \mathcal{T}^J, \hat{\mathbf{p}}^J) \\ &= f(\mathcal{T}^{J+1}) - f(\mathcal{T}^J) \quad (10) \end{aligned}$$

を考える (式 (9) 参照)。このとき、\$2g(J+1)\$ は、サンプルサイズが増大すると、自由度 2 の \$\chi^2\$ 分布に漸近することが知られている。有意水準 \$\alpha\$ (\$0 < \alpha < 1\$) を設定し、\$2g(J+1)\$ と \$\chi\_{2,\alpha}\$ を比較することにより、\$J+1\$ 個のアクティビティ変化点モデル \$\mathcal{T}^{J+1}\$ が、\$J\$ 個のアクティビティ変化点モデル \$\mathcal{T}^J\$ よりも有意に適合するかを検定する。ここに、\$\chi\_{h,\alpha}\$ は自由度 \$h\$ の \$\chi^2\$ 分布の上側 \$\alpha\$ 点である。ところで \$2g(2)\$ については、サンプルサイズが大きいとき、それは自由度 3 の \$\chi^2\$ 分布に従うので、\$\mathcal{T}^1\$ と \$\mathcal{T}^2\$ のどちらを選択するかを決めるとき、それを \$\chi\_{3,\alpha}\$ と比較することに注意する。

次に、アクティビティ変化点の最適集合 \$\mathcal{T}^J\$ を見つけるための貪欲法について述べる。そのアルゴリズムは、下記のとおりである。

**G0.** If \$2g(2) \le \chi\_{3,\alpha}\$ (see Eq. (10)), output \$\mathcal{T}^1\$ and stop.

**G1.** Initialize \$J \leftarrow 2\$.

**G2.** Search for

$$[t_{J+1}] = \arg \max_{[t] \in \mathcal{S} \setminus \mathcal{T}^J} f(\mathcal{T}^J \cup \{[t]\}).$$

**G3.** Update \$\mathcal{T}^{J+1} \leftarrow \mathcal{T}^J \cup \{[t\_{J+1}]\}\$.

**G4.** If \$2g(J) \le \chi\_{2,\alpha}\$ (see Eq. (10)), output \$\mathcal{T}^J\$ and stop.

**G5.** Set \$J \leftarrow J + 1\$, and return to Step G2.

ここにステップ G3 において、\$\mathcal{T}^{J+1}\$ の要素のインデックスは式 (3) の条件を満たすように再構築される。

我々は、より高品質な解を得るために貪欲法に局所探索を組み込む。集合 \$\mathcal{T}^J\$ が貪欲法により得られたとき、\$\mathcal{T}^J\$ の各要素 \$[t\_j]\$ をそれ以外の要素を固定して修正することによって、解品質を改善することを目指す。すなわち、上述の貪欲法のステップ G3 とステップ G4 の間に以下の 5 ステップを追加することを提案する。

**L1.** Initialize \$j \leftarrow 1, h \leftarrow 0\$.

**L2.** Search for

$$[t'_j] = \arg \max_{[t] \in \mathcal{S} \setminus \{\mathcal{T}^{J+1} \setminus \{[t_j]\}\}} f(\mathcal{T}^{J+1} \setminus \{[t_j]\} \cup \{[t]\}).$$

**L3.** If \$[t'\_j] = [t\_j]\$, set \$h \leftarrow h + 1\$, otherwise set \$h \leftarrow 0\$,



and update  $\mathcal{T}^{J+1} \leftarrow \mathcal{T}^{J+1} \setminus \{[t_j]\} \cup \{[t'_j]\}$ .

**L4.** If  $h = J + 1$ , go to Step G4.

**L5.** If  $j = J + 1$ , set  $j \leftarrow 1$ , otherwise set  $j \leftarrow j + 1$ , and return to Step L2.

ここにステップL3において、 $\mathcal{T}^{J+1}$ の要素のインデックスは式(3)の条件を満たすように再構築される。

このようにしてアクティビティ変化点の集合 $\mathcal{T}^J$ を探索し、アクティビティ変化パターン $(\mathcal{T}^J, \hat{p}^J)$ を検出する。

### 3.3 食材集合のアクティブネットワーク

異なる食材の各ペア $(u, v)$ に対して、アクティビティ変化パターン $(\mathcal{T}^{J(u,v)}(u, v), \hat{p}^{J(u,v)}(u, v))$ が観測時系列データ $\mathcal{X}(u, v)$ から検出されたとする。

さて、 $\hat{p}^1(u, v)$ は、任意の時間ステップ $[t] \in \mathcal{S}$ において $u$ と $v$ が同じレシピに含まれる確率を表している(式(6)参照)、すなわち、 $u$ と $v$ の共起の平均頻度を表していることに注意する。各 $j = 1, \dots, J(u, v)$ に対して、 $(u, v)$ の第 $j$ 期間 $\mathcal{S}_j(u, v)$ (式(4)参照)を考える。我々は、バースト現象に注目し、 $\hat{p}_j(u, v) > \hat{p}^1(u, v)$ ならば $\mathcal{S}_j(u, v)$ は $(u, v)$ のアクティブ期間であるといい、 $\hat{p}_j(u, v) \leq \hat{p}^1(u, v)$ ならば $\mathcal{S}_j(u, v)$ は $(u, v)$ の非アクティブ期間であるという。

我々は、任意の $[t] \in \mathcal{S}$ に対して、時間ステップ $[t]$ における食材集合のアクティブネットワーク $G_{[t]} = (V, \omega_{[t]})$ を以下のように構築する。まず、このネットワークは自己ループがない、すなわち、任意の $v \in V$ に対して $\omega_{[t]}(v, v) = 0$ とする。次に、任意の異なる $u, v \in V$ に対して、ノード $u$ とノード $v$ の間のリンク重み $\omega_{[t]}(u, v)$ は、時間ステップ $[t]$ における $u$ と $v$ の共起バースト度とする。すなわち、 $[t] \in \mathcal{S}_j(u, v)$ であるとき $\omega_{[t]}(u, v)$ は、 $\mathcal{S}_j(u, v)$ が $(u, v)$ のアクティブ期間ならば

$$\omega_{[t]}(u, v) = \frac{\hat{p}_j(u, v)}{\hat{p}^1(u, v)}$$

と定義され、 $\mathcal{S}_j(u, v)$ が $(u, v)$ の非アクティブ期間ならば

$$\omega_{[t]}(u, v) = 0$$

と定義される。ただし、 $\omega_{[t]}(u, v) = 0$ ならば時間ステップ $[t]$ において $u$ と $v$ の間にリンクは存在しないことにする。

### 3.4 食材の主要アクティブ共起パターンの抽出

食材集合 $V$ のアクティブネットワーク群 $\{G_{[t]} = (V, \omega_{[t]}) \mid [t] \in \mathcal{S}\}$ を分析する。まず、時間ステップ $[t] \in \mathcal{S}$ における食材の主要アクティブ共起パターンを、ネットワーク $G_{[t]} = (V, \omega_{[t]})$ の主要モジュールとしてコミュニティを同定することにより抽出することを考える。ただし、ネットワークのコミュニティとは部分ネットワークであり、コミュニティ内の結合が密でコミュニティ間の結合が疎であるという性質を持つものである。我々は、多くの領域で広く使われているNewman法[6], [22], [23]を適用

し、任意の $[t] \in \mathcal{S}$ に対してネットワーク $G_{[t]} = (V, \omega_{[t]})$ のコミュニティ、すなわち、食材の主要アクティブ共起パターンを抽出する。時間ステップ $[t] \in \mathcal{S}$ における食材の主要アクティブ共起パターン群を、

$$\mathcal{AP}_{[t]} = \{G_{[t],k} = (V_{[t],k}, \omega_{[t],k}) \mid k = 1, \dots, K_{[t]}\}$$

とする。ここに、 $K_{[t]}$ は正整数であり、各 $G_{[t],k} = (V_{[t],k}, \omega_{[t],k})$ は $G_{[t]} = (V, \omega_{[t]})$ の部分ネットワークであり、

$$V = \bigcup_{k=1}^{K_{[t]}} V_{[t],k} \quad (\text{disjoint union})$$

$$\omega_{[t],k}(u, v) = \begin{cases} \omega_{[t]}(u, v) & u, v \in V_{[t],k} \\ 0 & \text{otherwise} \end{cases}$$

である。我々は、各ネットワーク $G_{[t],k}$ を可視化することにより、時間ステップ $[t]$ における食材の主要アクティブ共起パターン群の構造を分析する。

最後に、抽出した食材の主要アクティブ共起パターン群の各々 $G_{[t],k} = (V_{[t],k}, \omega_{[t],k})$ について、それをレシピの言葉で説明するために、 $\mathcal{R}_{[t]}$ 内のレシピを以下のような手法でランキングする。2つの異なる食材の共起に注目し、レシピ $r \in \mathcal{R}_{[t]}$ がどの程度 $G_{[t],k} = (V_{[t],k}, \omega_{[t],k})$ に関連するかを

$$d(r) = \sum_{u,v \in V_{[t],k} \cap V(r)} \omega_{[t],k}(u, v)$$

によって測定し、 $d(r)$ に従って $\mathcal{R}_{[t]}$ 内のレシピをランキングする。ここに、 $V(r)$ はレシピ $r$ に含まれている食材の集合を表す。

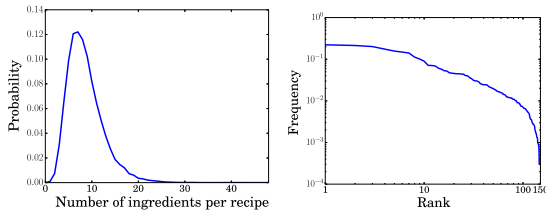
## 4. 実験

料理レシピ共有サイトの実データを用いて、提案法を評価する。この章では、まず、実験で利用するデータセットについて説明する。そして、アクティビティ変化パターンの検出および食材の主要アクティブ共起パターンの抽出という観点において、提案法を定量的に評価する。さらに、提案法により抽出された食材の主要アクティブ共起パターン群について、それらの特性を分析する。

### 4.1 実験データ

実験では、料理レシピ共有サイト「Cookpad」\*2の実データを用いた。Cookpadでは、レシピがいつ投稿され、それにはどのような食材が使われているかという情報が得られる。食材名には「卵」「王子」「たまご」「タマゴ」や「鮭」「シヤケ」「サーモン」などのような漢字、ひらがな、カタカナ、外来語などによる表記揺れが存在するので、食材用語に関する辞書を作成することによりそれを解消している。

\*2 <http://cookpad.com/>



(a) Distribution of recipe size. (b) Frequency-rank plot of ingredients.

図 3 Cookpad データセットの基本統計

Fig. 3 Fundamental statistics for the Cookpad dataset.

我々は、2011年1月から2012年12月までに投稿されたすべての料理レシピを調べた。本論文では、これを Cookpad データセットと呼ぶ。Cookpad データセットには、22,301 個のレシピと 155 種類の食材が含まれていた。

図 3 にレシピに含まれていた食材数の分布と、食材がそれが含まれたレシピ数に関してランキングしたときの「ランク・頻度プロット」を示す。図 3(a) より、レシピで使われる食材数の平均値はおよそ 7 であったこと、また、非常に多くの食材を使うレシピやほとんど食材を使わないレシピはまれであったことが分かる。また図 3(b) より、レシピに使われる頻度で食材を比較すると、高いものと低いものでは 3 桁以上もの大きな違いがあることが分かる。たとえば、最も使用された回数が少なかった食材は「塩昆布」で 12 レシピであったが、「ごま」は全レシピの約 3 分の 1 に相当する 8,784 レシピで使用されていた。

#### 4.2 アクティビティ変化パターン検出法の評価

食材ペアのアクティビティ変化パターン検出に関して、提案法の性能を評価した。この実験では、時間の単位に日を採用した。すなわち、 $T = 365$  である。3 章で述べたように、提案法は  $\chi^2$  検定を利用する貪欲法に局所探索を組み込んだものである。したがって、そのような貪欲法と提案法を比較した。提案法により得られたアクティビティ変化点の集合を  $T_*^{J*}$  とし、その貪欲法により得られたものを  $T_0^{J_0}$  とする。図 4 に、対数尤度比を表す目的関数  $F(T_*^{J*}; T_0^{J_0}) = f(T_*^{J*}) - f(T_0^{J_0})$  (式 (5), (7), (9) 参照) の分布を示す。具体的には、 $\forall z \geq 0.01$  に対して、 $F(T_*^{J*}; T_0^{J_0}) \geq z$  を満たす食材ペアの割合をプロットしている。具体的には、 $z$  の値に対して、 $\{F(T_*^{J*}; T_0^{J_0}) \geq z \text{ を満たす食材ペア数} / \{\text{食材ペアの総数}\}$  の値をプロットしている。図 4 より、まず提案法は、目的関数においてつねに比較対象の貪欲法を改善していたことが分かる。特に、約 80% の食材ペアに対しては、提案法による解  $T_*^{J*}$  はその貪欲法による解  $T_0^{J_0}$  よりも、尤度  $\exp\{\mathcal{L}^J\}$  (式 (5), (7) 参照) に関して 2 倍以上優れていた。さらに、約 20% の食材ペアに対しては、 $T_*^{J*}$  の尤度は  $T_0^{J_0}$  の尤度より 2 桁以上も高かった。これらの結果は、提案法の有効性を実証している。

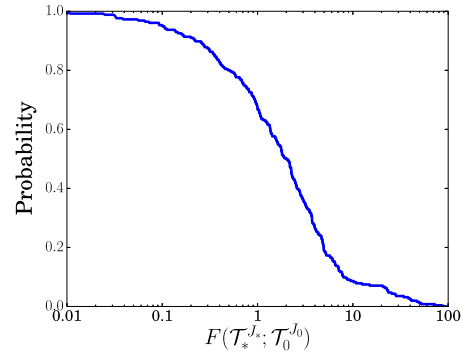


図 4 食材ペアのアクティビティ変化パターン検出に関する提案法の評価結果

Fig. 4 Evaluation results of the proposed method for detecting activity change patterns of ingredient pairs.

#### 4.3 食材の主要アクティブ共起パターン抽出法の評価

本節と次節では、Cookpad データセットに対し、提案法が抽出した食材の主要アクティブ共起パターン群  $\{AP_{[t]} | [t] \in \mathcal{S}\}$  について調べる。実験では、日常生活スタイルを考慮した分析を行うために、時間の単位を 1 週間に設定した。すなわち、 $T = 52$  である。まず本節では、抽出した各  $G_{[t],k} = (V_{[t],k}, \omega_{[t],k}) \in AP_{[t]}$  について、時間ステップ  $[t]$  における食材の主要アクティブ共起パターンとしての妥当性を定量的に評価する。

ところで、3.4 節では  $G_{[t],k}$  をレシピの言葉で説明するために、レシピ  $r \in \mathcal{R}_{[t]}$  が  $G_{[t],k}$  とどのくらい関連しているかをそのリンク構造に基づいて定量化することを考えた。ここではそれを一般化して、 $G_{[t],k}$  に関連のあるレシピを次のように定義する。ネットワーク  $G_{[t],k}$  において、リンクが存在している食材ペア  $(u, v)$  (すなわち、 $\omega_{[t],k}(u, v) > 0$  であるもの) に注目する。そして、このような食材ペアを少なくとも 1 組合むレシピ  $r$  は、 $G_{[t],k}$  に関連していると定義する。時間ステップ  $[t'] \in \mathcal{S}$  に投稿されたレシピ集合  $\mathcal{R}_{[t']}$  に対して、 $G_{[t],k}$  に関連しているレシピの集合を  $\mathcal{R}_{[t']}(G_{[t],k})$  とする。このとき、 $\beta(G_{[t],k}; [t']) = |\mathcal{R}_{[t']}(G_{[t],k})| / |\mathcal{R}_{[t']}|$  は、 $G_{[t],k}$  に関連するレシピが時間ステップ  $[t']$  に出現する確率と考えられる。そこで本論文では、 $G_{[t],k}$  に関連するレシピが時間ステップ  $[t]$  に出現する確率  $\beta(G_{[t],k}; [t])$  と平均値  $\bar{\beta}(G_{[t],k}) = \sum_{t'=0}^{T-1} \beta(G_{[t],k}; [t']) / T$  の比  $Q(G_{[t],k}) = \beta(G_{[t],k}; [t]) / \bar{\beta}(G_{[t],k})$  によって、 $G_{[t],k}$  が時間ステップ  $[t]$  における食材の主要アクティブ共起パターンを確かに表しているかを定量的に評価した。

提案法により抽出されたすべてのパターン  $G_{[t],k}$  に対して、 $Q(G_{[t],k})$  の分布を図 5 に示す。つねに  $Q(G_{[t],k}) > 1$  であったので、 $\forall z \geq 1$  に対して、 $Q(G_{[t],k}) \geq z$  を満たすパターン  $G_{[t],k}$  の割合をプロットしている。具体的には、 $z$  の値に対して、 $\{Q(G_{[t],k}) \geq z \text{ を満たすパターン } G_{[t],k} \text{ の数} / \{\text{パターン } G_{[t],k} \text{ の総数}\}$  の値をプロットしている。約 70% の  $G_{[t],k}$  に対しては  $\beta(G_{[t],k}; [t])$  は平均値の約 3 倍以

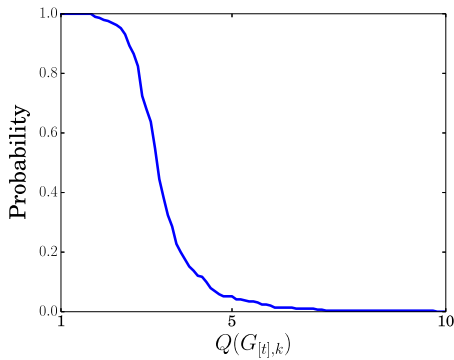


図 5 食材の主要アクティブ共起パターン抽出に関する提案法の評価結果

Fig. 5 Evaluation results of the proposed method for extracting major active co-occurrence patterns for ingredients.

表 1 共起バースト度の高い食材ペア (u, v)

Table 1 Most active ingredient pairs (u, v).

	(u, v)	アクティブ期間	共起バースト度
1	(パプリカ, 野菜*3)	12月28日 - 1月4日	17.1
2	(りんご, レモン)	1月5日 - 1月11日	16.2
3	(パプリカ, 豚肉)	4月27日 - 5月3日	16.1
4	(れんこん, 海苔)	3月2日 - 3月8日	13.9
5	(カニかま, ご飯)	2月2日 - 2月8日	12.1

上であり、約 20%の  $G_{[t],k}$  に対しては  $\beta(G_{[t],k}; [t])$  は平均値の約 4 倍以上であった。これらの結果は、各時間ステップにおける食材の主要アクティブ共起パターンの獲得に関して、提案法の有効性を実証している。

#### 4.4 食材の主要アクティブ共起パターンの分析

本節では、提案手法を Cookpad データセットに適用することにより、日本のレシピで利用された食材の組合せにおける季節変化の特徴について分析する。ここでも、時間の単位は 1 週間とした。

まず、検出した食材集合のアクティブネットワーク群  $\{G_{[t]} = (V, \omega_{[t]}) \mid [t] \in \mathcal{S}\}$  に対し、食材  $u$  と食材  $v$  の共起バースト度  $\omega_{[t]}(u, v)$  について調べた。表 1 に、上位 5 位までの食材ペア  $(u, v)$  とそれに対応するアクティブ期間  $S_j(u, v)$  と共起バースト度  $\omega_{[t]}(u, v)$  を示す。非常に共起バースト度の高い食材ペアが存在することが分かる。

次に、抽出した食材の主要アクティブ共起パターン群  $\{\mathcal{AP}_{[t]} \mid [t] \in \mathcal{S}\}$  の構造変化について分析した。それらの構造変化の可視化においては、平年  $\mathcal{S}$  を  $T$  個の動径ベクトルからなる円として表現した。ただし、各動径ベクトルは時間ステップ  $[t] \in \mathcal{S}$  ( $t = 0, 1, \dots, T-1$ ) に対応している。各ノード集合  $V_{[t],k}$  を対応する動径ベクトルの半直線上に配置することによって、 $\{\mathcal{AP}_{[t]} \mid [t] \in \mathcal{S}\}$  を可視化して

\*3 ここでは、レシピに登場した「お好みの野菜」「付け合わせの野菜」「好きなサラダ野菜」という 3 つの食材名を、「野菜」という食材名で統一した。したがって、「野菜」という食材名は料理人の嗜好に沿った適当な野菜を意味している。

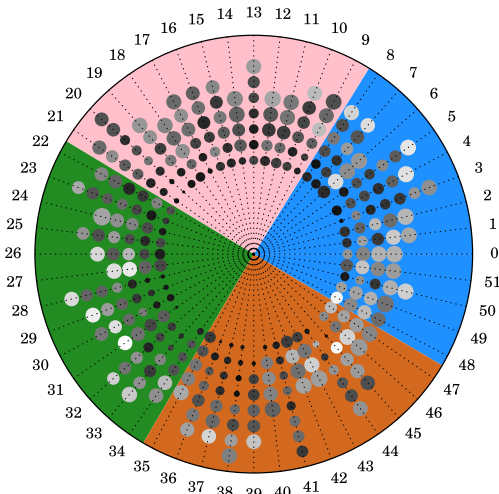


図 6  $\{\mathcal{AP}_{[t]} \mid [t] \in \mathcal{S}\}$  の構造変化の可視化

Fig. 6 Visualization of  $\{\mathcal{AP}_{[t]} \mid [t] \in \mathcal{S}\}$ .

いる。ここに、各ノード集合  $V_{[t],k}$  を円盤で描画し、その半径をノード数  $|V_{[t],k}|$  に比例するように、また、そのグレースケールをネットワーク  $G_{[t],k} = (V_{[t],k}, \omega_{[t],k})$  のリンク密度  $\sum_{u,v \in V_{[t],k}} \omega_{[t],k}(u, v) / |V_{[t],k}|$  に比例するように描いた。すなわち、各動径ベクトル上の円盤の個数は時間ステップ  $[t] \in \mathcal{S}$  において抽出された主要アクティブ共起パターンの総数に対応し、各円盤のサイズは大きいほどノード数が多く、円の描画色は白いほどリンク密度が高いことを示す。図 6 に、その構造変化の可視化結果を示す。ここでは四季との関係を把握しやすくするために、春、夏、秋、冬をそれぞれ、桃色、緑色、橙色、青色で着色している。食材の主要アクティブ共起パターン群  $\mathcal{AP}_{[t]} = \{G_{[t],k} = (V_{[t],k}, \omega_{[t],k})\}$  について、パターンの総数  $K_{[t]}$  および各パターン  $G_{[t],k}$  のノード数やリンク密度が時間ステップ  $[t]$  とともにどのように変動しているのかが概観できる。我々は、さらに各  $[t]$  に対して、 $\mathcal{AP}_{[t]} = \{G_{[t],k}\}$  の構造をより詳しく調べた。以下では例として、 $[t] = [1]$  (すなわち、1月5日から1月11日の週) の場合の結果について報告する。

図 7 に、 $\mathcal{AP}_{[1]} = \{G_{[1],1}, \dots, G_{[1],5}\}$  の構造を示す。ここでは、食材集合のアクティブネットワーク  $G_{[1]} = (V, \omega_{[1]})$  が可視化されており、その中で主要アクティブ共起パターンの各々  $G_{[1],k}$  が異なる色 (すなわち、 $k = 1$  は緑色、 $k = 2$  は橙色、 $k = 3$  は黄色、 $k = 4$  は青色、 $k = 5$  は紫色) で区別されている\*4。ただし、各リンク  $(u, v)$  の太さは  $\omega_{[1]}(u, v)$  の大きさに比例している。表 2 に、各  $G_{[1],k}$  の代表食材として、その中で次数が上位 5 位までの食材を示す。図 7 では、各  $G_{[1],k}$  で次数が最大の食材について、そのフォントサイズを大きくしている。また、1月の旬食材\*5を外枠が赤色のノードで表している。各  $G_{[1],k}$  は旬食材をある程

\*4  $K_{[1]} = 5$  であり、ネットワーク  $G_{[1]}$  の可視化において、異なるパターン間のリンクは薄灰色で表示されている。

\*5 「旬の食材百科」 (<http://foodslink.jp/syokuzaihyakka>) に従って 1 月の旬食材を決定した。



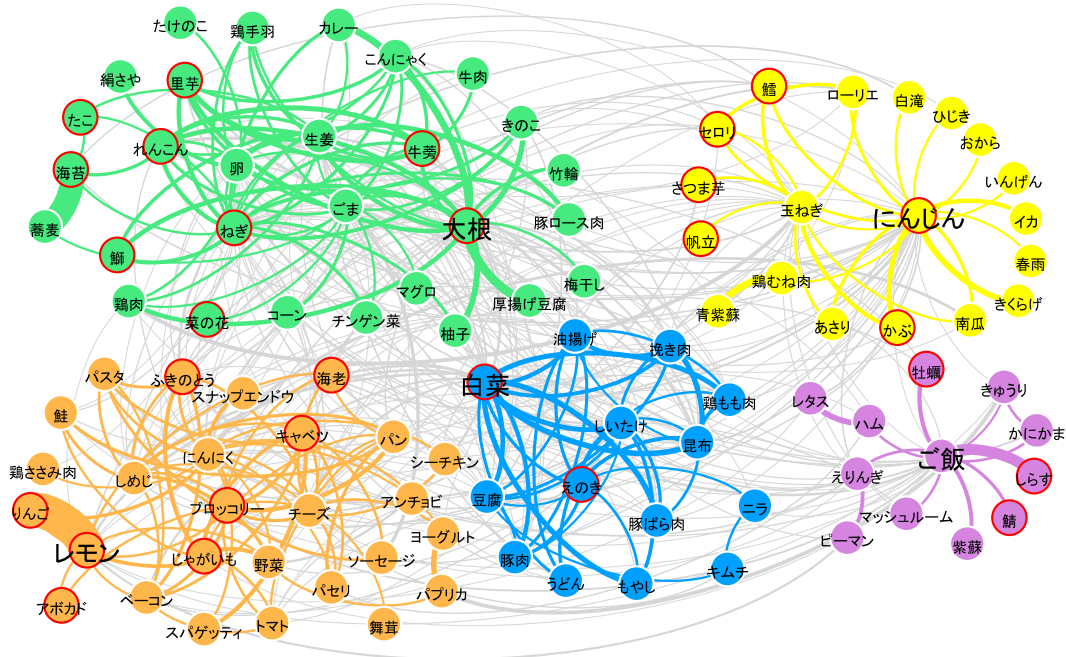


図 7 時間ステップ  $[t] = [1]$  における食材の主要アクティブ共起パターンの構造

Fig. 7 Structure of major active co-occurrence patterns for ingredients at time-step  $[t] = [1]$ .

表 3  $\mathcal{AP}_{[1]} = \{G_{[1],k}\}$  における代表レシピ

Table 3 Most representative recipes for  $\mathcal{AP}_{[1]} = \{G_{[1],k}\}$ .

(a) $k = 2$		(b) $k = 4$		(c) $k = 5$	
	Title		Title		Title
1	ローストビーフ・水菜とリンゴのサラダ添え	1	スープが自慢の美味しい★もつ鍋★	1	ちょっと贅沢?納豆しらすイクラ丼
2	レンジで!白菜と鶏挽肉と林檎の蒸しサラダ	2	ヘルシー簡単1人ししゃぶししゃぶ気分♪	2	柚子と梅のさっぱり雑炊
3	簡単♪家族大好き豆・まめサラダ♪	3	水菜使い切り♪豚肉と大根のはりはり鍋	3	風邪の時に食べやすい 白いおじや

表 2  $\mathcal{AP}_{[1]} = \{G_{[1],k} | k = 1, \dots, 5\}$  における代表食材

Table 2 Most representative ingredients for  $\mathcal{AP}_{[1]} = \{G_{[1],k} | k = 1, \dots, 5\}$ .

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
1	大根	レモン	にんじん	白菜	ご飯
2	ねぎ	りんご	玉ねぎ	しいたけ	しらす
3	生姜	にんにく	かぶ	昆布	ハム
4	ごま	ブロッコリー	ローリエ	油揚げ	えりんぎ
5	牛蒡	チーズ	鶏むね肉	えのき	レタス

度含んでいたことが観察される。特に、 $G_{[1],2}$  では旬食材の「りんごとレモン」の共起が特徴的であった。すなわち、食材の主要アクティブ共起パターンは、食材の旬の利用パターンと関係しうるものが示唆される。表 3 に、 $G_{[1],2}$ ,  $G_{[1],4}$ ,  $G_{[1],5}$  に対して、提案法によって抽出されたそれらの上位 3 位までの代表レシピを示す。食材の利用パターン  $G_{[1],2}$  は「サラダ料理」に関連していた。また、食材の利用パターン  $G_{[1],4}$  と  $G_{[1],5}$  は、それぞれ「鍋料理」と「手軽に作れる丼や雑炊」に関連していた。正月休み明けの週である  $[t] = [1]$  においては、正月料理に飽きたことや正月休み明けの気怠さなどから、このような料理レシピが

日本で多くなるのは自然だと考えられる。このようにして我々は、提案手法を適用することにより、日本の平年における食材の主要アクティブ共起パターンについていくつかの興味深い構造を見つけることができた。実際、それらは日本の季節性を強く反映しており、また日本の食習慣の傾向を如実に示していた。

レシピにおける食材の組合せの分析に関して、提案手法と従来の静的なネットワークベース手法を比較した。従来法では、全期間を通して食材ペアのレシピでの共起頻度を調べ、それに基づいて食材集合  $V$  の共起ネットワーク  $G^1 = (V, \omega^1)$  を構築し分析する。ここに、各リンク  $(u, v)$  の重み  $\omega^1(u, v)$  は  $\hat{p}^1(u, v)$  (式 (6) 参照) で定義される。ネットワーク  $G^1$  に Newman 法 [6], [22], [23] を適用してその主要モジュール群を同定することにより、高頻度に起こった食材の利用パターン群を抽出した。図 8 に、 $G^1$  の可視化と抽出された食材の利用パターンの構造を示す。ここで、各リンクの太さは重みの値に比例して描かれており、抽出された 4 個の利用パターン群は異なる色で区別されている。図 7 と同様に、各利用パターンにおいて最も次数が高い代表的な食材のフォントサイズを大きくしている。さ



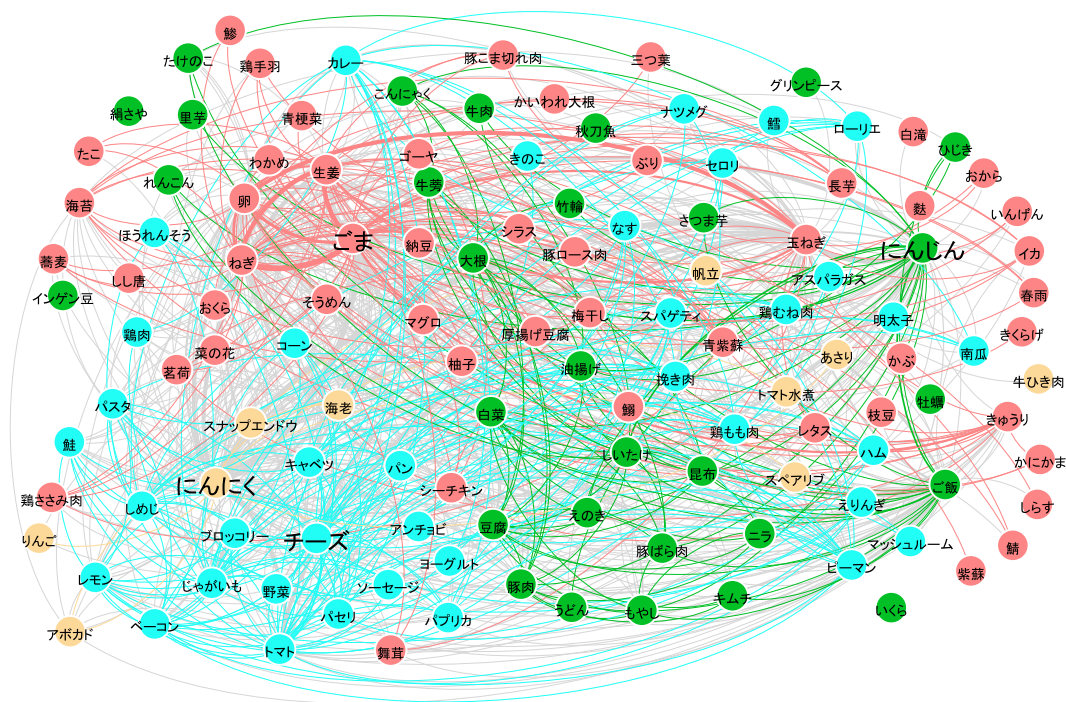


図 8 2年間を通して高頻度で起こった食材の利用パターンの構造  
 Fig. 8 Structure of frequent co-occurrence patterns for ingredients through two years.

らに図7と図8を比較しやすくするために、図7に含まれる食材は図8においても同じ座標に配置している。図7での食材の主要アクティブ共起パターンは、図8での高頻度で起こった食材の利用パターンと大きく異なっていることが観察される。たとえば、図7の提案法による  $G_{[1],2}^1$  では、「りんごとレモン」が強く結合しており、「にんにくとチーズ」もそれに属していたが、図8の従来法による  $G^1$  では、「りんごとレモン」は異なる利用パターンに属しており、「にんにくとチーズ」は異なる利用パターンの代表的な食材であった。また、 $G_{[1],5}^1$  において「ご飯としらす」は強く結合していたが、 $G^1$  ではそれらは異なる利用パターンに属していた。これらの結果は、従来法でとらえられない新たな重要構造を提案法が同定できることを示している。

### 5. 関連研究

まず、ネットワークベース手法を用いた大規模レシピデータにおける食材利用パターンの分析については、いくつかの研究がある。1章において述べたように、Ahnら[12]は、食材どうしで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[10]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、インド料理では食材ペアリングにおいて香辛料の選択が重要であり、特に香辛料の良くない組合せがほとんどないことを示した。Jiangら[24]は、レシピ群と使用される食材群との二部グラフに着目し、食材空間においてレシピ群を可視化

する手法を与え、世界の食文化を比較した。そして食材利用分布に関して、北アメリカと西ヨーロッパに類似性が見られること、ラテンアメリカ、北ヨーロッパおよび東アジアはそれらと異なる独自性を有することを示した。上記のように先行研究では、レシピに付随する時間情報を考慮せず食材利用パターンの静的な性質をもっぱら議論していたのに対し、本論文では、それらの動的変化をとらえる手法を提案していることに注意する。ところで、ネットワークベース手法を用いてブラジルにおける食材変化を約半世紀の長期にわたって分析したKinouchiら[25]のような研究もあるが、平年における食材利用パターンの季節変化の特徴を分析する研究は、本研究が最初であると考えている。

一方、利用されている食材の組合せに着目したレシピ推薦の問題が研究されている。レシピや食材に関するユーザの好みや評価を調査してレシピ推薦への戦略を比較した研究[1]、滋養のためのレシピプランニング支援の研究[2]、ダイエットを支援するためのレシピ推薦の研究[3]、レシピ推薦に向けてユーザの料理履歴から食材の好みを推定する研究[4]、[5]などが行われてきた。Tengら[13]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て、栄養やカロリーなどにおける品質向上の観点も加えてレシピを推薦する手法を与えた。本論文では、指定された期間（時間ステップ）に対して平年における食材の主要アクティブ共起パターンを抽出する手法を提案しているが、我々の提案手法をレシピ推薦の問題に適用することは、今後、取り組むべき重要な課題である。近年、人工知能研究の発展とともに、コンピュー

タにより料理レシピを創作するという研究が注目されつつある [9]. そのようなレシピ創作問題に本研究を拡張していくことも、今後の挑戦的な課題の 1 つと考えている。

本研究は、各食材ペアのレシピ出現頻度の時系列に対して、その変化パターンにおけるバースト検出が基本となっている。アイテム出現頻度の時系列データのバースト検出については、いくつかの研究がある。Zhu ら [26] は、閾値をあらかじめ設定することによって対象イベントのバースト期間を調べた。Swan ら [27] は、 $\chi^2$  検定に基づいてキーフレーズのバースト期間を検出することによりニュース記事ストリームデータからホットピックを抽出した。Kleinberg [28] は、バースト現象を隠れマルコフモデルにおける状態遷移としてとらえ、対象話題の電子メール送受信に関するバースト現象の階層構造を同定した。Sun ら [29] は、関連性のある複数のアイテムのバースト検出に Kleinberg の手法を拡張した。Saito ら [30] は、ソーシャルネットワークにおける情報拡散バーストを情報拡散モデルにおける拡散確率のバーストとしてとらえ、Twitter でのツイートへのリツイート拡散現象に関するバースト期間を同定した。与えられた期間内でアイテム出現頻度の時系列を調べその期間内でのバーストを検出する上記の従来研究と異なり、本研究では平年における性質を分析するために、季節変化の周期性を考慮して平年を円で表し、平年における期間の概念を導入して、数年にわたる食材ペアのレシピ出現頻度の時系列データから、平年における食材ペアのバースト期間を検出する手法を提案していることに注意する。

## 6. まとめ

料理レシピでいっしょに使われる食材ペアの季節による動的な変化を分析することを目指して、食材ペアの共起アクティビティに関する観測時系列データからその変化パターンを効率的に検出する問題を論じ、 $\chi^2$  検定を利用する貪欲法に局所探索を組み込んだ手法を提案した。そして、検出したアクティビティ変化パターンに基づき、各時間ステップにおいて食材集合のアクティブネットワークを構築し、それを主要なモジュールに分解することにより食材の主要アクティブ共起パターンを抽出した。さらに、抽出した食材の主要アクティブ共起パターン群に対して、各パターンをレシピの言葉で説明する手法を与えた。

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により、提案法の有効性を評価した。まず、食材ペアのアクティビティ変化パターン検出において、提案法はベースラインである貪欲法よりも高性能であることを実証し、次に、提案法が抽出した食材の主要アクティブ共起パターンの妥当性を示した。さらに提案法を適用して、Cookpad データセットにおける食材の主要アクティブ共起パターン群の平年での動的特性を分析し、食材の組合せに

関して季節性の観点からいくつかの興味深い構造を見つけた。また、提案法が抽出した構造は特異なものであり、静的な性質を扱う従来のネットワークベース手法では同定するのが困難であることを示した。

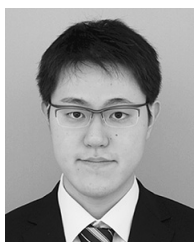
本研究では、日本の食習慣を研究対象とし、日本の料理レシピ共有サイトを調べた。もちろん提案法は、レシピの投稿時間とそこで使われた食材に関する情報が提供されているレシピ共有サイトであれば、どこの国のものにも適用することが可能である。提案法を様々な国のレシピ共有サイトに適用し、食材の主要アクティブ共起パターンの観点から世界規模で食文化の多様性を明らかにすることが、我々の今後の重要な研究課題である。

**謝辞** 本研究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、JSPS 科研費 JP17K00433 の助成を受けたものである。

## 参考文献

- [1] Freyne, J. and Berkovsky, S.: Intelligent food planning: Personalized recipe recommendation, *Proc. IUI '10*, pp.321-324 (2010).
- [2] Karikome, S. and Fujii, A.: A system for supporting dietary habits: Planning menus and visualizing nutritional intake balance, *Proc. ICUIIMC '10*, pp.56:1-56:6 (2010).
- [3] Mino, Y. and Kobayashi, I.: Recipe recommendation for a diet considering user's various conditions, *SCIS & ISIS*, Vol.2010, pp.471-476 (2010).
- [4] Ueda, M., Takahata, M. and Nakajima, S.: User's food preference extraction for personalized cooking recipe recommendation, *Proc. SPIM '11*, pp.98-105 (2011).
- [5] Ueda, M. and Nakajima, S.: Cooking recipe recommendation method focusing on the relationship between user preference and ingredient quantity, *Trans. Engineering Technologies*, pp.385-395, Springer (2015).
- [6] Newman, M.E.J.: *Networks: An Introduction*, Oxford University Press (2010).
- [7] Barabási, A.-L.: *Network Science*, Cambridge University Press (2016).
- [8] Nedović, V.: Learning recipe ingredient space using generative probabilistic models, *Proc. CwC '13*, pp.13-18 (2013).
- [9] Shao, N., Murali, P. and Sheopuri, A.: New developments in culinary computational creativity, *Proc. ICCCC '14*, pp.324-327 (2014).
- [10] 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).
- [11] Jermurawong, J. and Habash, N.: Predicting the structure of cooking recipes, *Proc. EMNLP '15*, pp.781-786 (2015).
- [12] Ahn, Y.-Y., Ahnert, S.E., Bagrow, J.P. and Barabási, A.-L.: Flavor network and the principles of food pairing, *Scientific Reports*, Vol.1, pp.196:1-196:7 (2011).
- [13] Teng, C.-Y., Lin, Y.-R. and Adamic, L.: Recipe recommendation using ingredient networks, *Proc. WebSci '12*, pp.298-307 (2012).
- [14] Janetzko, H., Jackle, D., Deussen, O. and Keim, D.A.: Visual abstraction of complex motion patterns, *Proc. SPIE 9017, Visualization and Data Analysis 2014* (2014).

- [15] Keim, D.A.: Information visualization and visual data mining, *IEEE Trans. TVCG*, Vol.8, No.1, pp.1-8 (2002).
- [16] Andrienko, G., Andrienko, N. and Wrobel, S.: Visual analytics tools for analysis of movement data, *SIGKDD Explor. Newsl.*, Vol.9, No.2, pp.38-46 (2007).
- [17] Nagel, H.R., Granum, E., Bovbjerg, S. and Vittrup, M.: Immersive visual data mining: The SDVDM approach, *Visual Data Mining*, pp.281-311, Springer (2008).
- [18] Dey, S., Simon, G.J., Westra, B.L., Steinbach, M. and Kumar, V.: Mining interpretable and predictive diagnosis codes from multi-source electronic health records, *Proc. SDM '14*, pp.1055-1063 (2014).
- [19] Huang, D., Tory, M., Aseniero, B.A., Bartram, L., Bateman, S. and Carpendale, S.: Personal visualization and personal visual analytics, *IEEE Trans. TVCG*, Vol.21, No.12, pp.420-433 (2015).
- [20] Bau, D., Zhou, B., Khosla, A., Oliva, A. and Torralba, A.: Network dissection: Quantifying interpretability of deep visual representations, *Proc. CVPR '17* (2017).
- [21] Tong, X., Li, C. and Shen, H.-W.: GlyphLens: View-dependent occlusion management in the interactive glyph visualization, *IEEE Trans. TVCG*, Vol.23, No.1, pp.891-900 (2017).
- [22] Newman, M.: Fast algorithm for detecting community structure in networks, *Physical Review E*, Vol.69, pp.066133:1-066133:5 (2004).
- [23] Newman, M.: Modularity and community structure in networks, *Proc. Natl. Acad. Sci. USA*, Vol.103, No.23, pp.8577-8582 (2006).
- [24] Jiang, Y., Skufca, J.D. and Sun, J.: BiFold visualization of bipartite datasets, *EPJ Data Science*, Vol.6, pp.2:1-2:19 (2017).
- [25] Kinouchi, O., Diez-Garcia, R.W., Holanda, A.J., Zambianchi, P. and Roque, A.C.: The non-equilibrium nature of culinary evolution, *New Journal of Physics*, Vol.10, No.7, pp.073020:1-073020:8 (2008).
- [26] Zhu, Y. and Shasha, D.: Efficient elastic burst detection in data streams, *Proc. KDD 2003*, pp.336-345 (2003).
- [27] Swan, R. and Allan, J.: Automatic generation of overview timelines, *Proc. SIGIR 2000*, pp.49-56 (2000).
- [28] Kleinberg, J.: Bursty and hierarchical structure in streams, *Proc. KDD 2002*, pp.91-101 (2002).
- [29] Sun, A., Zeng, D. and Chen, H.: Burst detection from multiple data streams: A network-based approach, *IEEE Trans. Systems, Man, and Cybernetics, Part C*, Vol.40, pp.258-267 (2010).
- [30] Saito, K., Ohara, K., Kimura, M. and Motoda, H.: Change point detection for burst analysis from an observed information diffusion sequence of tweets, *Journal of Intelligent Information Systems*, Vol.44, pp.243-269 (2015).



菊地 悠樹

2016年龍谷大学工学部電子情報学科卒業。2018年同大学大学院理工学研究科電子情報学専攻修士課程修了。同年株式会社NTT データアイ入社、第一事業部基盤システム担当、現在に至る。



熊野 雅仁 (正会員)

1991年立命館大学理工学部基礎工学科卒業。1991年龍谷大学理工学部実験助手。2008年龍谷大学理工学部実験講師、現在に至る。メディア論、社会ネットワーク可視化分析の研究と教育に従事。博士(工学)[神戸大学大学院工学研究科情報知能学専攻]。電子情報通信学会、認知科学会、人工知能学会、ACM、IEEE、他各会員。



木村 昌弘

1989年大阪大学大学院理学研究科数学専攻修士課程修了。同年日本電信電話株式会社入社。NTTコミュニケーション科学基礎研究所を経て、2005年退社。現在、龍谷大学理工学部電子情報学教授。複雑ネットワーク科学、データマイニング、機械学習の研究と教育に従事。博士(理学)。日本数学会、日本応用数理学会、人工知能学会、電子情報通信学会各会員。