

料理レシピ共有サイトにおける 食材の旬の利用パターンの抽出

菊地 悠樹^{1,a)} 熊野 雅仁^{2,b)} 木村 昌弘^{2,c)}

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み合わ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季節による動的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、食材の旬の利用パターンを抽出する。さらに、抽出した食材の旬の利用パターン群に対して、動的変化を可視化分析するフレームワークおよび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抽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旬の利用パターンの妥当性を示す。

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 Seasonally Active Patterns of Cooking Ingredient Usage in Recipe Sharing Sites

YUUKI KIKUCHI^{1,a)} MASAHITO KUMANO^{2,b)} MASAHIRO KIMURA^{2,c)}

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, and present a framework for visually analyzing their dynamical changes. 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.

Keywords: Recipe Analysis, Change Pattern Detection, Dynamical Network Analysis, Visualization, Social Media Mining

¹ 龍谷大学大学院理工学研究科電子情報学専攻
〒520-2194 大津市瀬田大江町横谷 1-5
Division of Electronics and Informatics, Ryukoku University,
Seta, Otsu 520-2194, Japan

² 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
Department of Electronics and Informatics, Ryukoku Uni-

versity, Seta, Otsu 520-2194, Japan

a) t16M013@mail.ryukoku.ac.jp
b) kumano@rins.ryukoku.ac.jp
c) 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み合わ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み合わ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^{*1}. 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 $\mathcal{T}^{J(u,v)}(u, v)$ とそれに対応する共起確率の組 $\mathbf{p}^{J(u,v)}(u, v)$ を検出することで捉えられる. ここに、 $J(u, v)$ は食材ペア (u, v) のアクティビティ変化点の数を表している.

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 $(\mathcal{T}^{J(u,v)}(u, v), \mathbf{p}^{J(u,v)}(u, v))$ 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トワークを構築する. ここに、

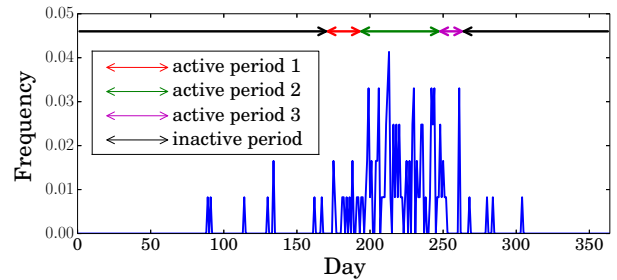


図 1: 平年における食材ペア（ごま，ゴーヤ）の共起アクティビティの変動

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旬の利用パターンを抽出する. 本論文では、抽出した食材の旬の利用パターン群に対して、利用パターン群の構造が時間ステップとともに動的に変化する様子を可視化するフレームワークおよび、各利用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.

本の構成は次の通りである. 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旬の利用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. 例えば、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、独自の旬期間をもっている. 本研究では、食材の使い方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 $\mathbb{R}/T\mathbb{Z}$ 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

*1 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.

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 $[t, t'] (C [0, T))$ であり、 $t > t'$ のとき $[t, T) \cup [0, t'] (C [0, T))$ 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み合わせが平年 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

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、食材の旬の利用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 アクティビティ変化パターンの検出

異なる食材ペア (u, v) 、 $(u, v \in V)$ 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 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もたないと言い、 $\mathcal{T}^1(u, v) = \emptyset$ とし、

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

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

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

をもち、それに対応する $J(u, v)$ 個の共起確率の組

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

が存在すると仮定する。 $(\mathcal{T}^{J(u, v)}(u, v), \mathbf{p}^{J(u, v)}(u, v))$ を (u, v) のアクティビティ変化パターンと呼び、 $j = 1, \dots, J(u, v)$ に対して $S_j(u, v) = [[t_j(u, v), [t_{j+1}(u, v)]]$ を (u, v) の第 j 期間と呼ぶ。ただし、 $t_{J(u, v)+1}(u, v) = t_1(u, v)$ である。

$\mathcal{T}^{J(u, v)}$ が指定されたならば、 $\mathcal{X}(u, v)$ の対数尤度 $\mathcal{L}^{J(u, v)}$ に対して、 $\mathbf{p}^J(u, v)$ の最尤推定値 $\hat{\mathbf{p}}^J$ は解析的に求められることに注意する。そこで我々は、次のようにして、目的関数 $f(\mathcal{T}^{J(u, v)}) = \mathcal{L}^{J(u, v)}(\mathcal{T}^{J(u, v)}, \hat{\mathbf{p}}^J)$ を最大化することにより最適なアクティビティ変化パターン $(\mathcal{T}^J(u, v), \hat{\mathbf{p}}^{J(u, v)})$ を検出する。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 $\mathcal{T}^2(u, v)$ を見つける。ところで、このような網羅的な探索法は、 J 個の変化点を見つかるときの計算量が $O(T^J)$ であることから、 $J \geq 3$ ならば実用的でない。したがって、与えられた $J \geq 3$ に対して最適解 $\mathcal{T}^J(u, v)$ を得るために、貪欲法を適用することを考え、さらに解品質を改善するために局所探索を組み込む。また、 $J(u, v)$ の値を推定するために、 χ^2 検定を利用する。

3.2 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 $(\mathcal{T}^{J(u, v)}(u, v), \hat{\mathbf{p}}^{J(u, v)}(u, v))$ 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 $[t] \in S$ 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、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 $S_j(u, v)$ 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 $S_j(u, v)$ は (u, v) のアクティブ期間であると言い、 $\hat{p}_j(u, v) \leq \hat{p}^1(u, v)$ ならば $S_j(u, v)$ は (u, v) の非アクティブ期間であると言う。

我々は、任意の $[t] \in S$ 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、 $[t] \in S_j(u, v)$ であるとき $\omega_{[t]}(u, v)$ は、 $S_j(u, v)$ 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、 $S_j(u, v)$ 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。

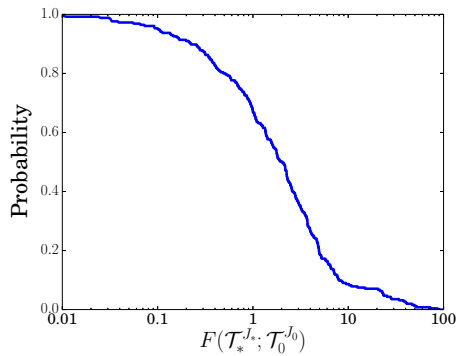


図 2: 食材ペアのアクティビティ変化パターン検出に関する提案法の評価結果

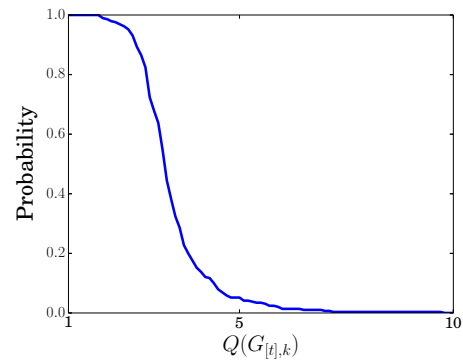


図 3: 食材の旬の利用パターン抽出に関する提案法の評価結果

3.3 食材の旬の利用パターンの抽出と可視化

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 $[t] \in \mathcal{S}$ における食材の旬の利用パターンを, ネットワーク $G_{[t]} = (V, \omega_{[t]})$ の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, 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もつものである. 我々は, 多くの領域で広く使われている Newman 法 [6], [14], [15] 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, 食材の旬の利用パターンを抽出する. 時間ステップ $[t] \in \mathcal{S}$ における食材の旬の利用パターン群を,

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

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 $[t]$ における食材の旬の利用パターン群の構造を分析する.

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旬の利用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旬の利用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. 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.

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 χ^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 $\mathcal{T}_*^{J_*}$ とし, その貪欲法により得られたものを $\mathcal{T}_0^{J_0}$ とする. 図 2 に, 対数尤度比を表す目的関数 $F(\mathcal{T}_*^{J_*}; \mathcal{T}_0^{J_0}) = f(\mathcal{T}_*^{J_*}) - f(\mathcal{T}_0^{J_0})$ の分布を示す. 具体的には, $\forall z \geq 0.01$ に対して, $F(\mathcal{T}_*^{J_*}; \mathcal{T}_0^{J_0}) \geq z$ を満たす食材ペアの割合をプロットしている. 図 2 より, まず提案法は, 目的関数において常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 $\mathcal{T}_*^{J_*}$ はその貪欲法による解 $\mathcal{T}_0^{J_0}$ よりも, 尤度 $\exp\{\mathcal{L}^J\}$ 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, $\mathcal{T}_*^{J_*}$ の尤度は $\mathcal{T}_0^{J_0}$ 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.

4.3 食材の旬の利用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旬の利用パターン群 $\{\mathcal{AP}_{[t]} \mid [t] \in \mathcal{S}\}$ 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

^{*2} <http://cookpad.com/>

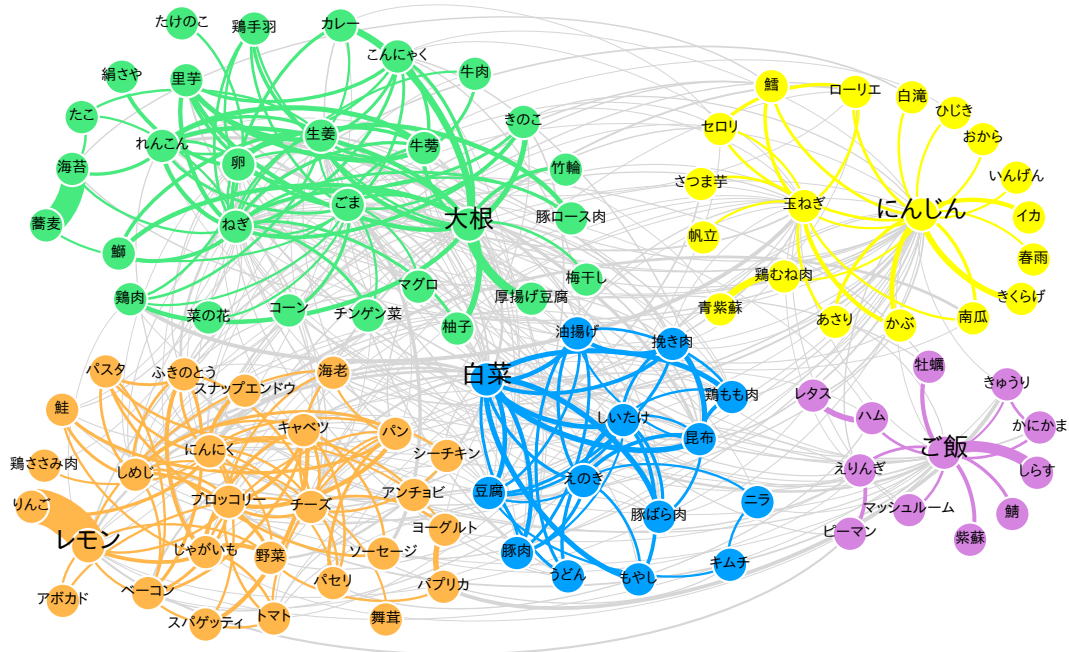


図 4: 時間ステップ $[t] = [1]$ における食材の旬の利用パターンの構造

表 2: $\mathcal{AP}_{[1]}$ における代表レシピのタイトル

(a) $k = 2$: 橙色の部分ネットワーク

(b) $k = 4$: 青色の部分ネットワーク

(c) $k = 5$: 紫色の部分ネットワーク

	Title
1	ローストビーフ・水菜とリンゴのサラダ添え
2	レンジで！白菜と鶏挽肉と林檎の蒸しサラダ
3	簡単♪家族大好き豆・まめサラダ♪

	Title
1	スープが自慢の美味しい★もつ鍋★
2	ヘルシー簡単 1 人ししゃぶししゃぶ気分♪
3	水菜使い切り♪豚肉と大根のはりはり鍋

	Title
1	ちょっと贅沢?納豆しらすイクラ丼
2	柚子と梅のさっぱり雑炊
3	風邪の時に食べやすい 白いおじや

$G_{[t],k} = (V_{[t],k}, \omega_{[t],k}) \in \mathcal{AP}_{[t]}$ について, 時間ステップ $[t]$ における食材の旬の利用パターンとしての妥当性を定量的に評価する.

ところで, 3.3 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一組含むレシピ 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 $[t]$ における食材の旬の利用パターンを確かに表しているかを定量的に評価した.

表 1: $\mathcal{AP}_{[1]}$ における代表食材

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
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

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 3 に示す. 具体的には, $\forall z \geq 1$ 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. まず, 常に $Q(G_{[t],k}) > 1$ であった. また, 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旬の利用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 食材の旬の利用パターンの分析

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み合わ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.

図 4 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, その中で旬の利用パターンの各々 $G_{[1],k}$ 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*3. 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 1 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 4 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. また表 2 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」に関連しており, そこでは「りんごとレモン」の共起が特徴的であった. 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, これら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旬の利用パターンについて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.

5. 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, χ^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旬の利用パターンを抽出した. さらに, 抽出した食材の旬の利用パターン群に対して, 利用パターン群の構造の動的変化を可視化分析するフレームワークおよび, 各利用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, 食材の旬の利用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 JSPS 科研費 JP17K00433 の助成を受けたものである. 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] Freyne, J. and Berkovsky, S.: Intelligent food planning: Personalized recipe recommendation, *Proceedings of 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, *Proceedings of ICUIMC'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, *Proceedings of 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, *Transactions on Engineering Technologies*, Springer, pp. 385–395 (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, *Proceedings of CwC'13*, pp. 13–18 (2013).
- [9] Shao, N., Murali, P. and Sheopuri, A.: New developments in culinary computational creativity, *Proceedings of ICCV'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, *Proceedings of 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, *Proceedings of WebSci'12*, pp. 298–307 (2012).
- [14] Newman, M.: Fast algorithm for detecting community structure in networks, *Physical Review E*, Vol. 69, pp. 066133:1–066133:5 (2004).
- [15] Newman, M.: Modularity and community structure in networks, *Proc Natl Acad Sci USA*, Vol. 103, No. 23, pp. 8577–8582 (2006).

*3 $K_{[1]} = 5$ であり, ネットワーク $G_{[1]}$ の可視化において, 異なる利用パターン間のリンクは薄灰色で表示されている.