

# パターンマイニング技術を用いた 特徴的食材構造の抽出に関する基礎検討

## A study of extracting characteristic structures among ingredients from recipe databases by using pattern mining techniques

尾崎知伸<sup>1\*</sup> 金城敬太<sup>2</sup>  
Tomonobu Ozaki<sup>1</sup> Keita Kinjo<sup>2</sup>

<sup>1</sup> 日本大学 文理学部

College of Humanities and Sciences, Nihon University

<sup>2</sup> 沖縄国際大学 経済学部

College of Economics and Environmental Policy, Okinawa International University

**Abstract:** Because of the recent increase in health consciousness and a rapid growth of user generated recipe data in social networking services, recipe mining, *i.e.* data mining in recipe databases, has been paid a wide attention as one of attractive research fields. In this paper, as a first step towards comprehensive and interpretable knowledge discovery from recipe databases, we investigate the applicability of the techniques in association rule mining to the recipe domain. We attempt to find interesting patterns which represent important structures among ingredients and characteristic words based on advanced association rules.

### 1 はじめに

爆発的な SNS の発展と健康志向の高まりに伴い、近年、クックパッド<sup>1</sup> や楽天レシピ<sup>2</sup> に代表されるようなレシピ投稿サイトが日常的に利用されるようになってきている。また、これらの Web サービスから得られるレシピ情報を対象とした情報処理技術の開発も精力的に行われており、レシピマイニングとして数多くの研究が行われている。

例えば文献 [1, 2, 3] では、食材の組み合わせにおける典型性や意外性について議論している。また文献 [4, 5] では、食材の組み合わせから代替可能な食材を特定する手法が提案されている。一方文献 [6] では、食材間に存在する構造を捉えるために食材間ネットワークを構築し、レシピ推薦へと応用している。また同様に文献 [7] では、料理における特徴的な単語に着目した食材間の依存関係抽出を行っている。食材と特徴語との関係を抽出する試みとして、特にオノマトペに着目した研究も盛んに行われている [8, 9, 10]。

レシピに含まれる食材や単語ではなく、手順そのものに着目した研究も数多く行われている。文献 [11] で

は、料理手順を表すグラフ表現が考案されている。また文献 [12] では、レシピテキストから構造化された手順を自動的に生成する手法が提案されている。加えて文献 [13] では、手順に着目することで類似レシピを要約するとともに、重要な差異を特定する手法が提案されている。

本研究では、文献 [7, 8, 10] などと同様、利用される食材とレシピに含まれる特徴的な単語を分析の対象とし、これらの間に存在する多様な関係性を捉え、構造化するための第一歩として、各種の相関ルールの適用可能性について考察を行う。相関ルール分析 [14] はデータマイニングの基本技術ではあるが、ルールの評価値や組み合わせを工夫することで、多様かつ明示的な関係性の抽出が可能であり、レシピ集合に対する理解を助けることが期待できる。

### 2 レシピデータベースへの相関ルール発見の適用

本章では、対象となるレシピデータベースの形式的な表現を導入する。その後、相関ルール [14] に関する基本事項及び本研究で採用した各種の拡張相関ルールについて概観し、それらのレシピデータへの適用可能性について述べる。

\*連絡先：日本大学 文理学部 情報科学科  
〒156-8550 東京都世田谷区桜上水 3-25-40  
tozaki@chs.nihon-u.ac.jp

<sup>1</sup><http://cookpad.com/>

<sup>2</sup><http://recipe.rakuten.co.jp/>

## 2.1 レシピデータベース

本研究では、レシピ  $r = \langle I_r, W_r \rangle$  を、(1) 利用される食材の集合  $I_r$  と (2) レシピ中に現れるオノマトペ  $O_r$  と形容詞  $A_r$  の各集合から構成される特徴的単語の集合  $W_r = O_r \cup A_r$  の対で表す。

レシピの集合  $R = \{r_1, r_2, \dots, r_{|R|}\}$  をレシピデータベースと呼ぶ。  $R$  に現れる全食材の集合を  $I_R = \cup_{\langle I, W \rangle \in R} I$ 、全単語の集合を  $W_R = \cup_{\langle I, W \rangle \in R} W$  とそれぞれ表記する。また、各レシピ  $r \in R$  から食材集合  $I_r$  のみを抽出したデータベースを  $R_I = \{I \mid \langle I, W \rangle \in R\}$ 、食材集合  $I_r$  と単語集合  $W_r$  の両方を抽出したデータベースを  $R_{IW} = \{I \cup W \mid \langle I, W \rangle \in R\}$  とそれぞれ表記する。

## 2.2 相関ルール

データベース中の全アイテムの集合を  $I$  と表記する。トランザクション  $t_i \subseteq I$  の集合をデータベース  $D = \{t_1, t_2, \dots, t_{|D|}\}$  と呼ぶ。データベース  $D$  に対し、アイテム集合であるパターン  $P$  の支持度を

$$\text{sup}_D(P) = |\{t_i \in D \mid P \subseteq t_i\}| / |D|$$

と定義する。

条件  $X, Y \subseteq I$  かつ  $X \cap Y = \emptyset$  を満たす 2 つのパターン  $X, Y$  間の関係を表すルールを相関ルールと呼び、 $X \rightarrow Y$  と表記する。  $X \rightarrow Y$  の支持度 (support) と確信度 (confidence) は、それぞれ

$$\begin{aligned} \text{sup}_D(X \rightarrow Y) &= \text{sup}_D(X \cup Y) \\ \text{conf}_D(X \rightarrow Y) &= \text{sup}_D(X \rightarrow Y) / \text{sup}_D(X) \end{aligned}$$

と定義される。一方、ルール  $X \rightarrow Y$  に対するリフト値 (lift) と説得性 (conviction) [15] は

$$\text{lift}_D(X \rightarrow Y) = \frac{\text{sup}_D(X \cup Y)}{\text{sup}_D(X) \cdot \text{sup}_D(Y)}$$

$$\text{conv}_D(X \rightarrow Y) = \frac{1 - \text{sup}_D(Y)}{1 - \text{conf}_D(X \rightarrow Y)}$$

と定義される。

一般に相関ルール発見問題では、支持度及び確信度やリフト値、説得性に関する閾値が与えられ、それらを満たすルールを発見することが目的となる。これまでに、アイテム間の支持度のばらつきを考慮しながら閾値を設定することの困難性を軽減するために、アイテム毎に閾値を設定する多重最低支持度 (multiple minimum support) [16] の考え方が提案されている。具体的には、データベース  $D$  中の各アイテム  $i \in I$  に対して最低支持度  $\sigma(i)$  を設定し、条件

$$\text{sup}_D(X \rightarrow Y) \geq \text{msup}_D(X \cup Y) = \min_{i \in X \cup Y} (\sigma_D(i))$$

を満たす相関ルールを頻出なものとして扱う。

本研究では、この考え方を採用するとともに、各アイテムに対する閾値設定を簡略化するため、ユーザが与える支持度に対する割合  $r$  ( $0 < r \leq 1$ ) と最低出現数  $c > 0$  を用い、各閾値  $\sigma(i)$  を

$$\sigma_D(i) = \max \left( \text{sup}_D(\{i\}) \times r, \frac{c}{|D|} \right)$$

と定義する。

## 2.3 ハイパークリークパターン

パターンを構成する各要素が高い相互依存性を有するパターンを、ハイパークリークパターン (hyper clique pattern) [17, 18] と呼ぶ。より形式的には、下記の基準  $hconf$  を用いてパターン  $P$  の相互依存性を評価し、ユーザによる閾値  $\tau$  以上の値を持つ場合に、ハイパークリークパターンと判断する。

$$\begin{aligned} hconf_D(P) &= \min_{i \in P} (\text{conf}_D(\{i\} \rightarrow P \setminus \{i\})) \\ &= \text{sup}_D(P) / \max_{i \in P} (\text{sup}_D(\{i\})) \end{aligned}$$

評価関数  $hconf$  は、「パターン  $P$  中のアイテム  $i$  がトランザクションに現れたときに、残りのアイテムもすべて同一のトランザクションに現れる確率が最低  $hconf_D(P)$  であること」を表しており、共起の観点からパターンを評価するものである。

ハイパークリークパターン  $P = \{i_1, \dots, i_{|P|}\}$  は、その評価関数から (パターンではなく) 相関ルールの集合

$$\left\{ \begin{array}{l} \{i_1\} \rightarrow P \setminus \{i_1\} \\ \{i_2\} \rightarrow P \setminus \{i_2\} \\ \dots \\ \{i_{|P|}\} \rightarrow P \setminus \{i_{|P|}\} \end{array} \right.$$

と捉えることも可能であり、相関ルールの一つの拡張とみなすことが出来る。

本研究では、レシピデータベース  $R$  に含まれる食材間の強い相互依存関係を抽出するため、頻出なハイパークリークパターンを発見することを提案する。すなわち、支持度に関するパラメータ  $r$  と  $c$  及び相互依存性に関する閾値  $\tau$  を与え、データベース  $R$  から、条件

$$\text{sup}_{R_I}(P) \geq \text{msup}_{R_I}(P) \wedge hconf_{R_I}(P) \geq \tau \wedge |P| \geq 2$$

を満たすパターン  $P$  を抽出する。

単純なハイパークリークパターンの発見では、料理全般に見られる普遍的な食材の組み合わせが発見され、各料理種に特化した意味のある情報が埋もれてしまう可能性が懸念される。この問題に対処するため、本研

究では料理種  $C$  ごとにデータベース  $R_C$  を準備し、その差を考えることを提案する。具体的には、パターンを抽出するデータベースとして料理種  $C_1$  に関するデータベース  $R^{C_1}$  及び制約として利用するデータベースとして料理種  $C_2 (\neq C_1)$  に関するデータベース  $R^{C_2}$  をそれぞれ準備し、パターン  $P$  に対する評価値の比  $hconf_{R^{C_1}}(P) / hconf_{R^{C_2}}(P)$  を考慮することで、料理種に依存しない普遍的な組み合わせをフィルタリングし、 $C_1$  に特化したパターンの抽出を試みる。

## 2.4 間接相関ルール

間接相関ルール (indirect association rule) [19] とは、ある条件の元での対比関係を表す相関ルールの対である。形式的には、最小支持度  $\sigma_{min}$  と最大支持度  $\sigma_{max}$ 、最小依存度  $\tau$  に対し、パターン  $M \subseteq I$  と二つのアイテム  $a, b \in I$  が、条件

$$\begin{aligned} sup_D(M \rightarrow \{a\}) &\geq \sigma_{min} \wedge d_D(M, \{a\}) \geq \tau \wedge \\ sup_D(M \rightarrow \{b\}) &\geq \sigma_{min} \wedge d_D(M, \{b\}) \geq \tau \wedge \\ sup_D(\{a, b\}) &< \sigma_{max} \end{aligned}$$

を満たすとき、2つの相関ルールの対

$$\langle M \rightarrow \{a\}, M \rightarrow \{b\} \rangle$$

を間接相関ルールと呼ぶ。ここで  $d_D(X, Y)$  はデータベース  $D$  中でのパターン  $X, Y$  間の依存度を表す関数であり、一般に、確信度やリフト値が用いられる。

間接相関ルールは、「条件  $M$  の元で  $a$  が起こりやすく、また同じ条件  $M$  の元で  $b$  も起こりやすいが、 $a$  と  $b$  が同時には起こることは少ない」ことを表しており、ある意味で  $a$  と  $b$  が同じような役割を果たす、またはライバルの関係にあることを表すルールと解釈することが出来る。

本研究では、間接相関ルールの考え方をレシピデータベースにおける代替食材の発見に適用することを提案する。具体的には、間接相関ルールでの閾値  $\sigma_{min}$  を関数  $msup$  に置き換えるとともに、依存度を表す関数  $d$  としてリフト値を採用し、レシピデータベース  $R$  から、条件

$$\begin{aligned} sup_{R_I}(M \rightarrow \{a\}) &\geq msup_{R_I}(M \cup \{a\}) \wedge \\ &\quad lift_{R_I}(M \rightarrow \{a\}) \geq \tau \wedge \\ sup_{R_I}(M \rightarrow \{b\}) &\geq msup_{R_I}(M \cup \{b\}) \wedge \\ &\quad lift_{R_I}(M \rightarrow \{b\}) \geq \tau \wedge \\ sup_{R_I}(\{a, b\}) &< \sigma_{max} \end{aligned}$$

を満たす間接相関ルールを抽出する。

## 2.5 制約付き相関ルール

料理において、味覚や食感は非常に重要視される要素である。本研究では、味覚や食感を表す単語と食材との間の依存関係を表すもっとも単純なルールとして、制約付き相関ルールを考え、これを食材-単語ルールと呼ぶ。

レシピデータベース  $R$  における食材-単語ルールとは、以下の条件を満たす食材集合  $X$  と特徴語  $w$  に関する相関ルール  $X \rightarrow \{w\}$  である。

$$\begin{aligned} X \subseteq I_R \wedge w \in W_R \wedge d_{R_{IW}}(X, \{w\}) &\geq \tau \wedge \\ sup_{R_{IW}}(X \rightarrow \{w\}) &\geq msup_{R_{IW}}(X \cup \{w\}) \end{aligned}$$

ここで  $d_{R_{IW}}(X, \{w\})$  は  $X$  と  $\{w\}$  の依存度を表す関数であり、本研究では確信度、リフト値、説得性の3種をそれぞれ用いることとする。また  $\tau$  は最小依存度を表す。

## 2.6 相関アクションルール

相関アクションルール (association action rule) [20] とは、「ある条件のもとでの前提の変化が、帰結の変化をもたらす」ことを表す相関ルールの対であり、帰結を望ましい状態に変化させるために必要となるアクション (前提の変化) を明示的に示すルールである。本来、属性=値表を対象としたルール表現であるが、本研究では形式を制限した上で、2値のトランザクションデータベースであるレシピデータベースへと適用する。

データベース中の各アイテム  $i \in I$  に対し、その否定を  $\bar{i}$  と表記する。レシピデータベース  $R$  において、以下の条件を満たす相関ルールの対

$$X \rightarrow \{\bar{w}\} \Rightarrow X \cup \{i\} \rightarrow \{w\}$$

を相関アクションルールと定義する。

$$\begin{aligned} X \cup \{i\} \subseteq I_R \wedge i \notin X \wedge w \in W_R \wedge \\ sup_{R_{IW}}(X \cup \{i\} \rightarrow \{w\}) &\geq msup_{R_{IW}}(X \cup \{w\}) \wedge \\ conf_{R_{IW}}(X \cup \{i\} \rightarrow \{w\}) &\geq \tau \wedge \\ conf_{R_{IW}}(X \rightarrow \{\bar{w}\}) &\geq \tau \end{aligned}$$

ここで、 $\tau$  はユーザによる閾値である。また  $conf_D(X \rightarrow \{\bar{w}\})$  の値は、否定を伴うルールの支持度

$$sup_D(X \rightarrow \{\bar{w}\}) = |\{t_i \in D \mid X \subseteq t_i, w \notin t_i\}| / |D|$$

を用いて計算される。

相関アクションルール  $X \rightarrow \{\bar{w}\} \Rightarrow X \cup \{i\} \rightarrow \{w\}$  は、「食材集合  $X$  に新たな食材  $i$  を加えることで、感覚 (特徴語)  $w$  が得られる」ことを表しており、状況 ( $X$ ) に応じて  $w$  をもたらすために必要となる食材  $i$  を特定できることが期待される。

### 3 実験

#### 3.1 実験データ

本研究で提案したレシピデータベースにおける相関ルール発見の妥当性を検証するため、各アルゴリズムを Java 言語によって実装し、計算機実験を行った。実験データには、情報学研究データリポジトリ<sup>3</sup>にてクックパッド株式会社と国立情報学研究所により提供されているクックパッドデータ<sup>4</sup>を用いた。

実験では、タイトルがうどん、そば、そうめんで行われるレシピを収集し、それぞれデータベースを構築した。また、オノマトペと形容詞は、各レシピに含まれるタイトルと概要、コツ・ポイント、手順から抽出した。抽出対象のオノマトペは、論文 [8] に従い、味覚や食感を表現するためによく利用される 126 語をひらがなとカタカナで表記した 252 語を採用した。一方、形容詞の特定には、形態素解析器 Mecab<sup>5</sup> [21] を利用した。表 1 に、実験対象となるデータセットの基本的な統計量を示す。

表 1: データセットの概要

	種類数	トランザクション毎の		
		平均値	中央値	最大値
うどん: データ数 9843				
食材	33	10.52	10	36
オノマトペ	8	0.46	0	6
形容詞	17	2.64	2	15
そば: データ数 7683				
食材	34	11.49	11	33
オノマトペ	7	0.59	0	5
形容詞	17	2.68	2	15
そうめん: データ数 3654				
食材	32	10.09	10	30
オノマトペ	6	0.51	0	4
形容詞	16	2.67	2	14

#### 3.2 実験結果

実験では、支持度や依存度に関するパラメタを変化させながら、各種の相関ルールの抽出を行った。表 2~7 に、各パラメタ設定において獲得された相関ルールの数を示す。

実験結果より、一般的に高頻度な相関ルールが少なく、多くのルールが支持度 1%未満であることが分かる。これはデータセットが比較的疎であることに加え、多重最低支持度基準を採用したことで、高頻度アイテム同士の組み合わせが排除されたことによるものと考えられる。しかしその一方で、多くの食材-単語ルールが導出されている。今回の実験では、ルール前提部に

食材間の頻出パターンを採用しているが、頻出パターンの任意の部分集合は頻出パターンとなるので、その観点から冗長なルールが多数導出されていると考えられる。ルールの意味を変えずに冗長なパターンを排除する意味でも、飽和パターン (closed pattern) や極小パターン (minimal generator) といった頻出パターンの代表元に限定する必要があると考えられる。

他の相関ルールと比較し、ハイパークリークパターンの数が少ない結果となった。この結果は、レシピデータベースを対象とした場合、相互に依存するという制限はある意味で強すぎる可能性があることを示唆するものであると考えられる。

食材-単語ルールに関しては、リフト値を基準とする場合と比べ、説得性を基準とする場合の方が少ないルールが抽出される結果となった。これは、リフト値はルールの方向に関係なく評価値が決まるのに対し、説得性ではルールの方向が重視されることに起因するものと考えられる。一方、確信度基準を用いた場合のルール数と、相関アクションルール数を比較すると、多くの場合で、相関アクションルール数が多いことが分かる。このことは、主に相関アクションルールが食材-単語ルールの対で構成されることに起因する。また、多くの場合で、導出された食材-単語ルールに対応するアクションルールが生成されたことを表していると考えられる。

次に、抽出された特徴的なルールについて概観する。ハイパークリークパターンとして、うどんを対象とした { キュウリ, マヨネーズ, レタス } や、そうめんを対象とした { オイスターソース, ゴマアブラ, ショウユ } が導出された。なお、{ キュウリ, マヨネーズ, レタス } は、他の料理 (そば, そうめん) での相互依存度が低く、うどんに特化したルールと判断できる。この他にも、そばにおける { エビ, ハクサイ } やそうめんにおける { ズッキーニ, パプリカ } が特有のルールとして抽出された。

一方、食材-単語ルールとしては、{ モヤシ, ブタニク, ニンジン, タマネギ, ミズ } → { パリパリ } (うどん, リフト値) や、{ ウメボシ, カツオブシ, ブタニク } → { さっぱり } (うどん, 説得性), { ウズラ, ハクサイ, タマゴ, モト, タマネギ, ブタニク, ニンジン, ゴマアブラ } → { パリッ } (そば, リフト値), { ウメボシ, カイワレ } → { さっぱり } (そうめん, 説得性) など、概ね妥当な結果が抽出された。

### 4 まとめと今後の課題

本研究では、食材間及び食材-単語間での特徴的な構造を抽出する第一歩として、各種の相関ルールを導入し、レシピデータベースへと適用した。またこれによ

<sup>3</sup><http://www.nii.ac.jp/dsc/idr/>

<sup>4</sup><http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html>

<sup>5</sup><http://taku910.github.io/mecab/>

表 2: 抽出されたハイパークリークパターン数

$s \setminus c$	うどん					そば					そうめん				
	0.5	0.4	0.3	0.2	0.15	0.5	0.4	0.3	0.2	0.15	0.5	0.4	0.3	0.2	0.15
0.05	0	2	9	27	49	1	9	38	117	151	1	3	6	30	39
0.03	0	2	12	50	89	1	9	38	145	266	1	3	7	40	77
0.01	1	4	18	66	156	1	10	42	165	329	1	5	10	61	122
0.00	3	7	21	77	196	3	14	50	178	364	1	5	10	64	142

各セルの値は、パラメタ  $r = 0.01, c = 10, \tau = 0.15$  において抽出されたパターン  $P$  のうち  $sup(P) \geq s, hconf(P) \geq c$  を満たすパターンの数

表 3: 抽出された間接相関ルール数

$s \setminus c$	うどん				そば				そうめん			
	10	5	2	1.5	10	5	2	1.5	10	5	2	1.5
0.05	0	0	0	0	0	0	0	0	0	0	0	0
0.03	0	0	0	1	0	0	0	4	0	0	0	0
0.01	0	0	113	561	0	1	201	881	0	2	150	523
0.00	1170	12220	237735	517969	1033	12273	225838	477823	23	548	18348	45231

各セルの値は、パラメタ  $r = 0.01, c = 10, \tau = 1.5, \sigma_{max} = 0.0025$  において抽出された間接相関ルール  $\langle M \rightarrow \{a\}, M \rightarrow \{b\} \rangle$  のうち、 $\min(sup(M \rightarrow \{a\}), sup(M \rightarrow \{b\})) \geq s$  及び  $\min(lift(M \rightarrow \{a\}), lift(M \rightarrow \{b\})) \geq c$  を満たすルール数

表 4: 抽出された食材-単語ルール数 (確信度)

$s \setminus c$	うどん				そば				そうめん			
	0.5	0.4	0.3	0.15	0.5	0.4	0.3	0.15	0.5	0.4	0.3	0.15
0.05	0	0	2	18	0	0	13	53	0	0	10	23
0.03	0	0	21	87	0	1	91	189	0	0	39	65
0.01	0	15	264	718	0	180	1404	2201	0	32	344	601
0.00	4670	12917	33273	58490	27104	61558	113927	147328	545	1656	4503	8853

各セルの値は、パラメタ  $r = 0.01, c = 10, \tau = 0.15$  において抽出されたルール  $X \rightarrow \{w\}$  のうち  $sup(X \rightarrow \{w\}) \geq s, conf(X \rightarrow \{w\}) \geq c$  を満たすルール数

表 5: 抽出された食材-単語ルール数 (リフト値)

$s \setminus c$	うどん				そば				そうめん			
	5	3	2	1	5	3	2	1	5	3	2	1
0.05	0	0	0	39	0	0	0	55	0	0	0	69
0.03	0	0	0	135	0	0	1	227	0	0	0	180
0.01	2	14	96	1204	0	11	89	2724	8	13	24	1559
0.00	22862	69230	145046	309212	46499	155940	400161	764737	2992	9757	25942	76362

各セルの値は、パラメタ  $r = 0.01, c = 10, \tau = 1.001$  において抽出されたルール  $X \rightarrow \{w\}$  のうち  $sup(X \rightarrow \{w\}) \geq s, lift(X \rightarrow \{w\}) \geq c$  を満たすルール数

表 6: 抽出された食材-単語ルール数 (説得性)

$s \setminus c$	うどん				そば				そうめん			
	5	3	2	1	5	3	2	1	5	3	2	1
0.05	0	0	0	38	0	0	0	55	0	0	0	64
0.03	0	0	0	133	0	0	0	226	0	0	0	171
0.01	0	0	0	1178	0	0	0	2710	0	0	0	1521
0.00	2578	6844	19560	306593	16496	42320	107460	762060	402	1130	3097	75761

各セルの値は、パラメタ  $r = 0.01, c = 10, \tau = 1.001$  において抽出されたルール  $X \rightarrow \{w\}$  のうち  $sup(X \rightarrow \{w\}) \geq s, conv(X \rightarrow \{w\}) \geq c$  を満たすルール数

表 7: 抽出された相関アクションルール数

$s \setminus c$	うどん				そば				そうめん			
	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2	0.5	0.4	0.3	0.2
0.05	0	0	4	18	0	0	0	0	0	0	13	23
0.03	0	0	29	100	0	2	134	194	0	0	60	94
0.01	0	44	627	1472	0	551	3556	5104	0	82	780	1288
0.00	15390	51803	125592	208496	91742	280915	500891	624370	1599	5453	13939	26699

各セルの値は、パラメタ  $r = 0.01, c = 10, \tau = 2.0$  において抽出された相関アクションルール  $\langle X \rightarrow \{x\}, X \cup \{i\} \rightarrow \{x\} \rangle$  のうち、 $\min(sup(X \rightarrow \{x\}), sup(X \cup \{i\} \rightarrow \{x\})) \geq s$  及び  $\min(conf(X \rightarrow \{x\}), conf(X \cup \{i\} \rightarrow \{x\})) \geq c$  を満たすルール数

り、断片的ではあるが、一定の構造が取り出せることを確認した。

今後は、食材の重要度や希少性、分量に着目した構造化や料理手順を考慮した構造化も検討する。これらには、価値や量を対象とした相関ルール発見技術 [22, 23] や、系列やグラフといった構造データを対象とした手法 [24, 25] が有効であると考えている。また、LeGo アプローチ [26] を用い、断片的に抽出された多様な構造を組み合わせ、大域的な構造を特定することを検討している。さらにその上で、アナロジーに基づくルールアブダクション [27] を展開することを予定している。

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] 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み合わせの典型度分析, 電子情報通信学会, 信学技報, Vol.114, No.487, MVE2014-84, pp.49-54, 2015.
- [2] 池尻恭介, 清雄一, 中川博之, 田原康之, 大須賀昭彦: 食材に基づいたレシピの意外度算出手法の提案, 電子情報通信学会, 信学技報, Vol.113, No.214, DE2013-33, pp.1-6, 2013.
- [3] 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, Vol.31, No.3, pp.70-78, 2014.
- [4] 志土地由香, 井手一郎, 高橋友和, 村瀬洋: 料理レシピマイニングによる代替可能食材の発見, 電子情報通信学会論文誌, Vol.J94-A, No.7, pp.532-535, 2011.
- [5] 野沢健人, 中岡義貴, 山本修平, 佐藤哲司: word2vec を用いた代替食材の発見手法の提案, 電子情報通信学会, 信学技報, Vol.114, No.204, DE2014-30, pp.41-46, 2014.
- [6] C.-Y. Teng, Y.-R. Lin and L. A. Adamic: Recipe recommendation using ingredient networks, Proc. of the 4th Annual ACM Web Science Conference, pp.298-307, 2012.
- [7] 鈴木貴文, 山田雅之, 宮崎慎也, 中貴俊, 遠藤守: レシピの特徴を考慮した食材間の依存関係抽出に関する研究, 電子情報通信学会, 信学技報, Vol.114, No.485, MVE2014-110, pp.167-171, 2015.
- [8] 渡辺知恵美, 中村聡史: オノマトペロリ: 味覚や食感を表すオノマトペによる料理レシピのランキング, 人工知能学会論文誌, Vol.31, No.1, pp.340-352, 2015.
- [9] ラームサムルアイパン カンウィパー, 渡辺知恵美, 中村聡史: オノマトペロリ: オノマトペを利用した料理推薦システム, 情報処理学会研究報告, Vol.2009-DD-73, No.6, 2009.
- [10] ラームサムルアイパン カンウィパー, 渡辺知恵美, 中村聡史: レシピ検索システムにおけるオノマトペとレシピ用語集合の関連付け, 情報処理学会研究報告, Vol.2010-DBS-150, No.15, 2010.
- [11] S. Mori, H. Maeta, Y. Yamakata and T. Sasada: Flow Graph Corpus from Recipe Texts, Proc. of the International Conference on Language Resources and Evaluation 2014, pp. 2370-2377, 2014.
- [12] 山肩洋子, 今堀慎治, 前田浩邦, 森信介: 調理手順文書の自然言語解析結果からの食材・加工からなる作業ツリーの構築, 電子情報通信学会, 信学技報, Vol.114, No.204, DE2014-27, pp.25-30, 2014.
- [13] 杉山祐一, 山肩洋子, 田中克己: 手順情報としてのレシピデータに対する類似レシピの要約と微小で重要な差異の発見, 第5回データ工学と情報マネジメントに関するフォーラム, D3-5, 2013.
- [14] R. Agrawal and R. Srikant: Fast Algorithms for Mining Association Rules in Large Databases, Proc. of 20th International Conference on Very Large Data Bases, pp. 487-499, 1994.
- [15] C. Tew, C. Giraud-Carrier, K. Tanner and S. Burton: Behavior-based clustering and analysis of interestingness measures for association rule mining, *Journal of Data Mining and Knowledge Discovery*, Vol.28, pp.1004-1045, 2014.
- [16] B. Liu, W. Hsu and Y. Ma: Mining association rules with multiple minimum supports, Proc. of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.337-341, 1999.
- [17] H. Xiong, P.-N. Tan, and V. Kumar: Mining strong affinity association patterns in data sets with skewed support distribution, Proc. of the 3rd IEEE International Conference on Data Mining, pp.387-394, 2003.
- [18] H. Xiong, P.-N. Tan, and V. Kumar: Hyperclique pattern discovery, *Data Mining and Knowledge Discovery*, Vol.13, No.2, pp.219-242, 2006.
- [19] P.-N. Tan, V. Kumar and J. Srivastava: Indirect Association: Mining Higher Order Dependencies in Data, Proc. of the 4th European Conference on Principles of Data Mining and Knowledge Discovery, pp.632-637, 2000.
- [20] Z. W. Ras, A. Dardzinska, L.-S. Tsay, and H. Wasyluk: Association Action Rules, *IEEE International Conference on Data Mining Workshop*, pp.283-290, 2008.
- [21] T. Kudo, K. Yamamoto and Y. Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing, pp.230-237, 2004.
- [22] C. W. Wu, P. Fournier-Viger, P. S. Yu and V. S. Tseng: Efficient Mining of a Concise and Lossless Representation of High Utility Itemsets, Proc. of the 11th IEEE International Conference on Data Mining, pp.824-833, 2011.
- [23] 光永悠紀, 鷲尾隆, 元田浩: 適応的密度基準に基づく部分空間クラスタリングを用いた定量的多頻度アイテム集合のマイニング, 人工知能学会論文誌, Vol.21, No.5, pp.439-449, 2006.
- [24] G. Dong and J. Pei: *Sequence Data Mining*, Springer, 2007.
- [25] D. J. Cook and L. B. Holder: *Mining Graph Data*, WILEY, 2007.
- [26] J. Fürnkranz and A. J. Knobbe (Eds.): Special Issue: Global Modeling Using Local Patterns, *Data Mining and Knowledge Discovery*, Vol.21, No.1, 2010.
- [27] 金城敬太, 尾崎知伸, 古川康, 原口誠: アナロジーを組み込んだルール発想推論によるスキル獲得支援, 人工知能学会論文誌, Vol.29, No.1, pp.188-193, 2014.