

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

カンウィパー ラートサムルアイパン[†]
中村聡史^{††} 渡辺知恵美[†]

日本では料理や味覚を擬音語、擬態語を表すオノマトペを用いて「ふわふわケーキ」のように表現することが多い。そこで我々は、オノマトペを利用した料理レシピ推薦システムの開発をしている。これまで、オノマトペとレシピ用語の一対一の関連を求めてオノマトペをキーワードにしてレシピを検索するシステムを開発した。今後は、複数の用語やオノマトペと用語の組合せでも検索できるようにすべきである。そこで、本研究はオノマトペと用語の集合の関連を求め、これらの結果に基づいた検索システムのインターフェースを提案する。

Association between Onomatopoeic words and Other types of words Set In Recipe Recommendation System

Kanwipa Lertsumruaypun[†] Satoshi Nakamura^{††}
and Chiemi Watanabe[†]

Japanese language is filled with onomatopoeic words, which describe sounds or actions like “click” or “bow-wow”. Especially, when talking about foods, Japanese people frequently use onomatopoeic words to express vague taste or sense of the foods. There, we develop a system for searching recipes by using onomatopoeic words as the search keyword. We collected recipes from a posted recipe website, named “COOKPAD”, and performed Japanese language morphological analysis. Then, we calculated relation between onomatopoeic words and other types of words. In this paper, we calculate association rules of them and used these rules to find matched recipes to recommend to the user.

1. はじめに

近年 Web ページの爆発的増加と検索エンジンの発展により、分からないことがあればすぐに検索エンジンを利用することが常識になりつつある。しかしながら検索目的があいまいで、頭にイメージがあるがキーワードをうまく言語化できないといった経験がないだろうか。我々はその解決法としてオノマトペ（擬音語、擬態語）に着目している。日本語はオノマトペ（特に擬態語）の種類は豊富であり、物事の漠然としたイメージを表現する際にオノマトペを用いることが多い。特に料理に関して「とろとろとした料理」「じっくり焼く」などのように、味や感触などの表現にオノマトペが非常に利用される。

我々は投稿型料理レシピサイトを利用し、オノマトペから料理レシピを検索できるようなシステム「オノマトペロリ」を開発している。先行研究では、キーワードのオノマトペに関連度が高いレシピを推薦するため、レシピページを収集し、料理に関する用語（食材や料理名、料理器具、調理法、味や食感など）とオノマトペの共起度をレシピページから算出した。例えば、食材「唐辛子」とオノマトペ「ピリッ」の共起度は強く、「ふわふわ」との共起度は弱いなどの値により、オノマトペと用語との関連を TF-IDF 法に基づいて求め、これらの値を用いてオノマトペによる料理レシピ検索システム「オノマトペロリ」を開発した。しかし、これまでのオノマトペロリでは TF-IDF 法によるオノマトペとレシピ用語の一対一の関連度に基づいてレシピを推薦する。しかし、一つの用語は別の用語との組合せでオノマトペとの関連が異なる。例えば、「大根」を「煮込み」にすれば、「とろとろ」になるが、「大根」を生のまま「サラダ」にすれば、「シャキシャキ」になる。

そこで、本研究はオノマトペとレシピ用語の集合の関連度を求める。例えば、「大根」と「シャキシャキ」をキーワードにして検索を行うと、「煮込み」類のレシピのレシピではなく、大根のシャキシャキの食感を与える「サラダ類」などのレシピを推薦する。

本稿では、オノマトペとレシピ用語の集合の関連度算出結果を報告し、またその結果に基づき、新しいシステムインターフェースを提案する。

[†] お茶の水女子大学大学院人間文化創成科学研究科
Graduate School of Humanities and Sciences, Ochanomizu University

^{††} 京都大学大学院情報学研究科
Department of Social Informatics, Graduate School of Informatics, Kyoto University

2. 先行研究

2.1 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

「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」[1]では、実験データとして Cookpad[2]のレシピデータを利用した。理由は、Cookpad はユーザによって自分のオリジナル料理レシピを投稿・検索できるウェブサイトであり、一般の料理レシピと比較してオノマトペによる表現が多いからである。初めに、Cookpad のレシピを収集してレシピから料理に関する 5 種類の用語（食材、料理名、調理法、調理器具そして味・食感）及びオノマトペを抽出した。例えば、「唐辛子」や「豚」は食材、「煮込み」は料理名、「柔らかい」は味・食感などに分類される。

次に、TF-IDF（Term Frequency-Inverse Document Frequency）法を用いて、用語の種類別にオノマトペと用語の関連度を計算した。ここでは、レシピに含まれる単語 w のオノマトペ o に対する関連度 $Weight(o, w)$ を以下の式によって算出する。

$$Weight(o, w) = TF(o, w) * IDF(o, w)$$

- TF

$$TF(o, w) = \frac{\text{オノマトペ}o\text{を含む全レシピの用語}w\text{の出現数}}{\text{オノマトペ}o\text{を含む全レシピの用語の出現数}}$$

- IDF

$$IDF(o, w) = \log \frac{\text{全オノマトペの種類数}}{\text{用語}w\text{と同じレシピに出現するオノマトペの種類数}}$$

例として、「とろとろ」というオノマトペに対して、関連度上位 3 種類の食材、料理名、調理法、調理器具そして味・食感を図 1 に示す。

TF-IDF 法に基づいた計算のように、オノマトペが含まれているレシピのみが計算対象となっているため、求められたオノマトペと用語の関連の因果関係が「オノマトペ⇒用語」となる。この関連に基づき、オノマトペをキーワードにした料理レシピ推薦システム「オノマトペロリ」を開発した。

図 2 で示しているトップページのように、ユーザはオノマトペの入力欄またはタグクラウドからキーワードのオノマトペを入力し、検索を行う。しかし、オノマトペからのみではなく、複数のキーワードの組合せからの検索も必要であると考えられる。そこで、本研究ではオノマトペとレシピ用語の集合の関連度も計算する。

オノマトペ	食材	関連度	オノマトペ	調理法	関連度
とろとろ	大根	0.06771	とろとろ	かける	0.01900
	酒	0.02119		煮る	0.01646
	水	0.01952		煮込む	0.01396

オノマトペ	料理名	関連度
とろとろ	スープ	0.10139
	うどん	0.08474
	カレー	0.07050

オノマトペ	調理器具	関連度	オノマトペ	味・食感	関連度
とろとろ	鍋	0.13082	とろとろ	柔らかい	0.11524
	圧力	0.04512		美味しい	0.10276
	器	0.04389		甘い	0.09341

図 1TF-IDF 法による関連度の計算結果例

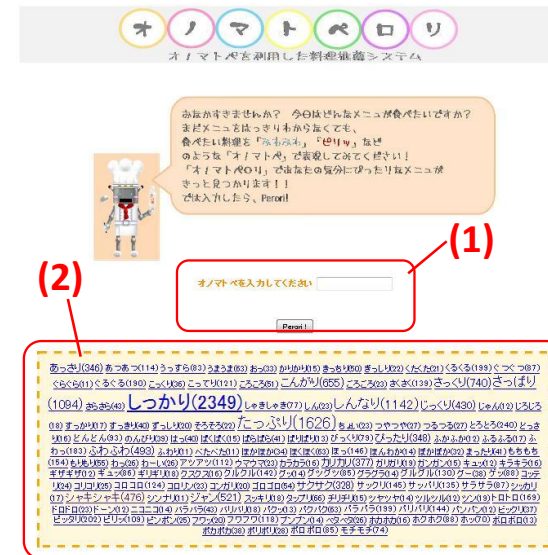


図 2 オノマトペロリのトップページ

2.2 TF-IDF によるオノマトペと用語集合の計算の問題点

オノマトペとレシピ用語の集合の関連度を計算する方法として、TF-IDF を利用する方法も考えられるが、前節の計算式からわかるように、条件部と結論部を何にすることによって計算が異なる。そのため、条件部と計算部を逆にすることはまた計算し直さなければならない。例えば、「シャキシャキ⇒大根」の関連を求めたい場合は条件部である「シャキシャキ」が含まれるレシピのみを絞ってから計算を行う。また、逆の関連の「大根⇒シャキシャキ」を求めたい場合は計算し直す必要がある。さらに、本研究でオノマトペと用語集合の関連を求める際、条件部と結論部を決定する際に組合せ爆発が起こる可能性がある。例えば、{シャキシャキ, 大根, サラダ} のオノマトペと用語の集合があるとすると、「シャキシャキ」と「{大根, サラダ}」の共起回数を 2.1 節で述べた TF-IDF の式に適用して、「シャキシャキ⇒{大根, サラダ}」の関連が求められても、「{シャキシャキ, 大根}⇒サラダ」の関連を求めるには、初めから計算する必要がある。

これに対して、相関ルールは、まず頻出アイテム集合を求める。そして求められた頻出集合に対して「 $X \Rightarrow Y$ 」(X, Y はアイテム集合とする) となる関連をすべてのパターンを求めた上で、確信度の高いパターンのみ選択して相関ルールとする。そのため、TF-IDF のように条件部と結論部の異なる関連を求める際に計算し直さなければならない手間が省くことができる。よって、本研究は相関ルールを用いてオノマトペと用語集合の関連を求めることにする。また、相関ルールの検出には Apriori アルゴリズムを利用する。

3. オノマトペとレシピ用語の集合の関連

オノマトペロリにおける相関ルール検出について説明する前に、相関ルール及び Apriori アルゴリズムについて簡潔に述べる。

3.1 相関ルールと Apriori アルゴリズムの概要

相関ルール[3]とは、トランザクションデータベースに頻出するアイテム間の何らかの組合せの規則である。例えば、「商品 A を買うと商品 B も買う」のようなルールを簡潔に、 $\{A\} \Rightarrow \{B\}$ と表す。相関ルールは通常「 $X \Rightarrow Y$ 」で示す。ただし、ルールの「 $\Rightarrow$ 」の左辺を条件部、また右辺を結論部と呼ぶ。

相関ルールを検出するために何らかの評価指標が必要である。多く使われる評価指標としては、支持度(Support)、確信度(Confidence)、リフト(Lift)がある。

ただし、ここで扱うデータベースは、トランザクションの集合 $D = \{t_1, t_2, \dots, t_m\}$ であり、また各トランザクションはアイテム集合 $I_{all} = \{i_1, i_2, \dots, i_k\}$ の子集合より構成されると

する。つまり、任意のトランザクション t_j は、アイテムの集合 I_j と持つ($I_j \subseteq I_{all}$)であり、かつ、その子集合は空集合ではない。

これから、データベースの中のトランザクションやアイテム集合における支持度、確信度、そしてリフトについて簡単に説明する。

- 支持度

例として、アイテム集合 X の支持度とは、データベースの中のアイテム集合 X を含むトランザクションの数である。よって、相関ルール $X \Rightarrow Y$ の支持度とは、アイテム集合 X と Y を含むトランザクションが全トランザクションに占める割合となる。

$$\text{supp}(X \Rightarrow Y) = \frac{\text{条件}(X) \text{と結論}(Y) \text{を共に含むトランザクション数}}{\text{全トランザクション数}}$$

- 確信度

確信度とは、相関ルール $X \Rightarrow Y$ のとき、ルールの条件部 X が発生した際に結論部 Y が起こる割合を示す。つまり、相関ルールの条件部と結論部と結び付きが強いことを意味する。

$$\text{conf}(X \Rightarrow Y) = \frac{\text{条件}(X) \text{と結論}(Y) \text{をともに含むトランザクション数}}{\text{前提}(X) \text{を含むトランザクション数}}$$

- リフト

リフトは、相関ルールの確信度を結論部の支持度で割った値である。

$$\text{lift}(X \Rightarrow Y) = \frac{\text{conf}(X \Rightarrow Y)}{\text{結論}(Y)}$$

リフトは、結論部または条件部の支持度が高くても、そのアイテム集合はいつも起こる、いわゆる一般性が高いものであればリフト値が低くなり、リフト値を利用して一般性の高すぎるアイテム集合を含む相関ルールを省くことができる。

例えば、オノマトペロリでは、レシピに良く登場する「塩」を含むアイテム集合の相関ルールは支持度が高いが、一般性の高い「塩」にはあまり意味がないルールであるため、リフト値を考慮することによってこのようなルールを省くことができる。

本研究ではこれらの評価指標を総合的に考慮し、Apriori アルゴリズムに基づいて相

関ルールを検出するが、次に Apriori アルゴリズムの概要について説明する。

Apriori アルゴリズムは高速で関連ルールを検出するアルゴリズムとして、R.Agrawal 氏らによって開発された[4]。Apriori アルゴリズムには頻出アイテム検出そして関連ルール検出の以下の 2 ステップがある。

- Step2: 頻出アイテム集合を求める。

頻出アイテム集合を効率的に求めるには Apriori アルゴリズムではアイテム集合の組合せを作成しながら、アイテム集合の支持度を計算する。そして、アイテム集合の支持度は単純減であることを利用して、あらかじめ決めておく最小支持度以上の支持度をもつアイテム集合のみに絞り込む。

例えば、アイテム集合 $I_m \subset I_n$ ならば $\text{supp}(I_m) > \text{supp}(I_n)$ である。つまり、 $\{i_1, i_2\}$ が頻出アイテム集合でなければ、 $\{i_1, i_2, i_5\}$ も頻出アイテム集合ではない。

- Step2: 頻出アイテム集合から関連ルールを導出する。

Step1 で求めた頻出アイテム集合の中から、あらかじめ決めておいた最小確信度以上の確信度をもつルールのみが関連ルールとする。

3.2 オノマトペロリにおける Apriori アルゴリズムの適用及び拡張

次にオノマトペロリで Apriori アルゴリズムをどのように適用するかを説明する。はじめに、ここでいったトランザクションをレシピとする。また、レシピの文章に含まれるオノマトペ及び他のレシピ用語（食材、料理名、調理法、調理器具、味覚・食感）はアイテムとなる。

図 3 は計算対象となるトランザクション及びアイテム例を示す。

TID(レシピ番号)	アイテム(オノマトペ, レシピ用語)
500006	{ふわふわ, 強力粉, もちもち, 塩, 焼く, 水, ...}
500009	{さっぱり, 豚肉, 酢, 水, 焼く, 切る, ...}
500032	{もちもち, 水, 白玉粉, チョコレート, 溶かす, 強力粉}
500034	{シャキシャキ, 水, 酢, 蓮根, 切る, 炒める, ...}
500042	{さっぱり, ぱくぱく, 塩, 水, 豆腐, 焼き, ひっくり返す, ...}
...	...

図 3 トランザクション及びデータアイテム例

ただし、計算対象となるオノマトペはレシピのイメージを表現する「レシピのタイトル」また、「レシピの説明」に含まれるもののみとする。理由は、レシピ内の調理手

順などに含まれる、「さっくり」混ぜるや「しっかり」炒める等のようなオノマトペは直接に料理の印象や食感に関係がないため、レシピのタイトルやレシピの説明文に含まれるもののみを計算対象とする。

次に、オノマトペロリにおける頻出アイテム集合の検出について説明する。本研究は SQL によって頻出アイテム集合を求める[5]ことにした。つまり、以下の図 4 及び図 5 で示しているオノマトペのテーブル、そしてレシピ用語のテーブルの結合を繰り返してアイテム集合の組合せを作成する。

Ono_id	Word_ono	Recipe_nbr
4	もちもち	500006
5	ふわふわ	500006
10	サクサク	500010
11	たっぷり	500010
4	もちもち	500032
5	ふわふわ	500038
...	...	...

「レシピタイトル」または
「レシピ説明文」にオノマトペ
が含まれるレシピID

図 4 オノマトペテーブル

Vocab_id	Vocab	term	Recipe_nbr	Sentence_source
66	シナモン	food	500006	step_3
65	ロール	dish	500006	step_3
1680	焼く	cook	500006	description
60	バター	food	500006	item
...	...	...	...	...

用語が含まれる
「レシピパーツ」

図 5 レシピ用語テーブル

ただし、レシピの内容に適切な関連ルールを検出するために従来の Apriori アルゴリズムの拡張として、ここでは、オノマトペと用語を結合する際の結合条件として、「オノマトペと用語が同じレシピに出現する」ことにした。しかし、用語と用語の組合せを作成する際には、「用語 1 と用語 2 が同じレシピ、かつ同じレシピのパーツに出現す

る」ことを条件とする．理由は，異なるレシピパーツに出現する用語同士には関連が低いと考えられる．例えば，調理手順 1 には「魚を揚げる」また調理手順 2 には「玉ねぎを炒める」と記述されたとき，「魚」は同じパーツに出現した「揚げる」には関連が高いが，「炒める」とは関連が低い．

また今回はオノマトペとレシピ用語の集合の関連を求めたいため，頻出アイテム集合を求めるときに作成する組合せは「オノマトペ 1 語と用語 2 語」が含まれることを条件とする．

さらに，頻出アイテム集合を検出する際，あらかじめ決めておくしきい値の最小支持度以下の支持度をもつ組合せを除く必要がある．本研究ではその基準となる支持度を経験的に「0.035%」と決めた．また，一般性の高い用語が含む意味のないルールが生じないため，ここでは最小支持度以外にリフト値を考慮する．今回は経験的に「0.008」のリフト値をしきい値とする．つまり，頻出アイテム集合候補となる組合せは「支持度が 0.035%以上」かつ「リフト値が 0.008」でなければならない．さらに，関連ルールを検出する際に評価指標とする最小確信度を経験的に「0.05%」とする．

3.3 オノマトペと用語集合の関連ルールの検出結果及び考察

次に，オノマトペと用語集合の関連ルールを検出した結果を示す．ここでは，20,000 件のレシピデータを用い，「オノマトペ 1 語と用語 2 語」の関連ルールを求めた．また，結果を以下の 4 つのパターンの関連ルールに分けて示す．

- A : 用語の集合 $\Rightarrow$ オノマトペ 1 語
- B : 用語とオノマトペの集合 $\Rightarrow$ 用語 1 語
- C : オノマトペ 1 語 $\Rightarrow$ 用語の集合
- D : 用語 1 語 $\Rightarrow$ オノマトペと用語の集合

図 6 は第 3.2 節で述べた評価指標（最小支持度：0.035%，最少リフト値：0.008，最小確信度：0.05%）によって，検出できた各パターンの関連ルールの数及びオノマトペの数を示す．ただし，ここでは「o」をオノマトペ，また「w」を用語と表記する．

パターン	関連ルールの数	オノマトペの数
A: $\{w,w\} \Rightarrow o$	110	20
B: $\{o,w\} \Rightarrow w$	1904	37
C: $o \Rightarrow \{w,w\}$	533	34
D: $w \Rightarrow \{o,w\}$	84	11

図 6 検出できた関連ルールの数

図 6 で示している結果により，関連ルールの数は開発中のレシピ検索システムに十分であると考えられるが，ルールに含まれるオノマトペの種類は全体の約 10%であり，豊富ではない．改善方法として，「関連ルールを検出する評価指標の再検討」また，計算対象となるレシピの件数を増やすことを考えている．

また，図 7 - 図 14 は結果例として，「シャキシャキ」及び「サクサク」というオノマトペを含む各パターンの関連ルールを示す．

用語1	用語2	オノマトペ	確信度
水菜	盛る	シャキシャキ	0.2174
水菜	サラダ	シャキシャキ	0.2055
酢	れんこん	シャキシャキ	0.1791
酢	さらず	シャキシャキ	0.1250
大根	サラダ	シャキシャキ	0.1008
千切り	さらず	シャキシャキ	0.0661

図 7 パターン A の関連ルール例(シャキシャキ)

用語1	用語2	オノマトペ	確信度
包む	まとまる	サクサク	0.3103
抜く	伸ばす	サクサク	0.2133
ラップ	まとまる	サクサク	0.1852
冷蔵庫	伸ばす	サクサク	0.1644
生地	寝かせる	サクサク	0.1455
バター	クッキー	サクサク	0.1449
生地	クッキー	サクサク	0.1444
ラップ	伸ばす	サクサク	0.1410
生地	パイ	サクサク	0.1379
バター	すり混ぜる	サクサク	0.1204

図 8 パターン A の関連ルール例(サクサク)

オノマトペ	用語1	用語2	確信度
シャキシャキ	ナッツ	焙る	0.9231
シャキシャキ	ナッツ	ごま	0.9231
シャキシャキ	れんこん	酢	0.6316
シャキシャキ	大根	サラダ	0.4800
シャキシャキ	水菜	サラダ	0.4545
シャキシャキ	盛る	水菜	0.3704
シャキシャキ	ごま	煎る	0.3143
シャキシャキ	水菜	盛る	0.3030
シャキシャキ	さらす	千切り	0.2857
シャキシャキ	ごま	サラダ	0.2286

図 9 パターン B の関連ルール(シャキシャキ)

オノマトペ	用語1	用語2	確信度
サクサク	オープン	焼く	0.8870
サクサク	イースト	強力粉	0.8667
サクサク	あがる	焼く	0.8667
サクサク	包む	ラップ	0.8235
サクサク	すり混ぜる	バター	0.8125
サクサク	付く	焼く	0.7647
サクサク	卵黄	バター	0.7586
サクサク	まとめる	生地	0.7500
サクサク	トースター	焼く	0.7391
サクサク	バニラ	バター	0.7391

図 10 パターン B の関連ルール(サクサク)

オノマトペ	用語1	用語2	確信度
シャキシャキ	ごま	だれ	0.0829
シャキシャキ	水菜	サラダ	0.0829
シャキシャキ	ナッツ	だれ	0.0718
シャキシャキ	大根	サラダ	0.0663
シャキシャキ	酢	蓮根	0.0663
シャキシャキ	ごま	煎る	0.0608
シャキシャキ	酢	さらす	0.0608
シャキシャキ	水菜	盛る	0.0552

図 11 パターン C の関連ルール(シャキシャキ)

オノマトペ	用語1	用語2	確信度
サクサク	オープン	焼く	0.4340
サクサク	薄力粉	バター	0.2681
サクサク	粉	バター	0.1787
サクサク	オープン	生地	0.1319
サクサク	ラップ	冷蔵庫	0.1277
サクサク	生地	伸ばす	0.1149
サクサク	薄力粉	ふるう	0.1106
サクサク	生地	冷蔵庫	0.1106
サクサク	薄力粉	ベーキングパウダー	0.0979
サクサク	バター	卵黄	0.0936

図 12 パターン C の関連ルール(サクサク)

用語1	オノマトペ	用語2	確信度
れんこん	シャキシャキ	酢	0.7010

図 13 パターン D の関連ルール(シャキシャキ)

用語1	オノマトペ	用語2	確信度
すり混ぜる	サクサク	バター	0.0594
下げる	サクサク	焼く	0.0592
まとめる	サクサク	生地	0.0579
麺棒	サクサク	伸ばす	0.0530
パイ	サクサク	生地	0.0526
アーモンド	サクサク	バター	0.0517

図 14 パターン D の関連ルール(サクサク)

パターン A - C の結果より、オノマトペと用語の集合の関連を求めることによって、より高精度のレシピ検索ができると考えられる。例えば、「シャキシャキ」が含まれる

関連ルールについては、「大根、サラダ」や「水菜、サラダ」などのシャキシャキの食感を与える用語集合が含まれている。また、図 9 で示している「{シャキシャキ、大根} ⇒ サラダ」のルールについては、レシピを検索するときに、「シャキシャキ 大根」をキーワードにすると、「煮物」類ではなく、大根のシャキシャキとする食感がある「サラダ」類のレシピを推薦することできる。

一方、パターン D の関連ルールを用いれば、用語をキーワードにしたときに検索結果にはオノマトペのグループごとでレシピを推薦することができると考えられる。例えば、「大根」で検索したとき、結果は「シャキシャキ」のグループに「サラダ類」の大根料理、また「とろとろ」のグループには「煮込み」類の大根料理が推薦される。しかし、現段階では、パターン D の関連ルールの数が不十分である。ゆえに、今後は対象データを増やすこと、またパターン D のみに適切な評価指標を検討する必要がある。

他に、「サクサク」に関する関連ルールの結果をみると、焼菓子の作り方に関する用語が非常に多いことがわかった。原因は、「サクサク」が含むレシピのほとんどが焼菓子のレシピであることが考えられる。また、ここで結果例を挙げていないが、「ふわふわ」のオノマトペに関する関連関係も同じような現象である。ふわふわの場合は、焼菓子の他に卵料理や蒸し料理などもふわふわの食感を与えるが、それらのレシピの数が少ないため、関連ルールが検出できないと考えられる。解決方法として、今後はより大規模のレシピデータを対象とし、焼菓子類のレシピを省いてから関連ルールを検出することが考えられる。

4. 新しいインターフェースの提案

これまでのオノマトペロリはオノマトペのみを検索キーワードにしたが、本研究で求めたパターン A 及びパターン C のオノマトペと用語集合の関連ルールを利用することによって、複数のキーワードからレシピを検索することができる。

また、パターン A と D の結論部にはオノマトペが含まれているので、これは、結果レシピをオノマトペのグループに分けて推薦することができる。例えば、「餃子 皮」で検索を行うと「パリパリ」また「もちもち」のグループに分けてレシピを推薦する。

他に、パターン B の結果を用いて、料理に使いたい食材とオノマトペの組合せで検索を行うと、調理法または、料理の種類別レシピを推薦することができる。例えば、「水菜 シャキシャキ」で検索すると、「炒める」また「サラダ」でレシピを推薦していく。

5. まとめ及び今後の課題

これまでオノマトペとレシピ用語の一对一の関係を計算して、その結果を料理推薦システムに利用したが、複数のキーワードからレシピ検索を行えるために、オノマトペとレシピ用語の集合の関連を求めることにした。関連を計算する方法として、先行研究で利用した TF-IDF 法と比較して計算コストが低い Apriori アルゴリズムによってオノマトペの用語集合の関連ルールを検出することにした。本稿ではその計算結果を報告し、また結果を用いてレシピ検索システムの新しいインターフェースを提案した。

今後の課題として、今回ではまだ不十分である関連ルールをより有効な関連ルールを検出するために計算対象のレシピデータ数を増やすことが考えられる。また、関連ルールを検出する段階で使われる評価指標の決め方についても再検討する必要がある。

また、今回は「オノマトペ 1 語と用語 2 語」の関連ルールを検出したが、より高精度のレシピ検索のため、今後「オノマトペ 1 語以上」と「用語 1 語以上」の集合の関連ルールも検出すると考える。

参考文献

- 1) カンウィパー ラートサムルアイバン, 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, 情報処理学会第 73 回デジタルドキュメント研究会(2009)
- 2) 日本最大の料理サイト「COOKPAD」<http://cookpad.com>
- 3) 金明哲:アソシエーション分析(1), [連携]フリーソフトによるデータ解析・マイニング第 40 回, ESTRELA, 2006 年 11 月, No.152, pp.52-57.
- 4) R. Agrawal and R. Srikan (t 1994): Fast algorithms for mining association rules: In J. B. Bocca, M. Jarke, and C. Zaniolo, editors: Proc. 20th Int. Conf. Very Large Data Bases, VLDB, Morgan Kaufmann, 12-15 September, pp.487-499.
- 5) Maurice Houtsma and Arun Swami: Set-Oriented Mining for Association Rules in Relational Databases: Eleventh International Conference on Data Engineering, Taipei 1995.