

# クックパッドデータの分析

---

尾崎研究室

石井悠加里

川口美香

# 先行研究: レシピマイニング

代替可能食材  
の特定

[志土地11, 野沢14]

食材間依存  
関係の  
抽出

[鈴木15]

食材間  
ネットワーク  
の  
構築と推薦

[Teng12]

食材の  
組み合わせに  
おける  
典型性, 意外性

[横井15, 池尻13, 池尻14]

オノマトペ  
との関係抽出

[渡辺15, カンウィパー  
09, カンウィパー10]

フローグラフ  
(調理手順を  
表すグラフ)

[Mori14, 山肩14]

類似レシピの  
要約

[杉山13]

# クックパッドデータ

クックパッド株式会社が国立情報学研究所と協力して研究者に提供しているデータセット  
172万品のレシピやそれらからなる献立に関するデータ

The screenshot shows the website interface for the National Institute of Informatics (NII) IDR Data Repository. The header includes the NII logo and the IDR logo. The main navigation menu contains links for HOME, データ一覧 (Data List), 組織 (Organization), 関連リンク (Related Links), and お問い合わせ (Contact Us). The breadcrumb trail indicates the current page is HOME > データ一覧 > クックパッドデータセット (Cookpad Dataset). On the left sidebar, there is a list of datasets including Yahoo!データセット, 楽天データセット, ニココデータセット, リクルートデータセット, and クックパッドデータセット. The main content area features a title 'クックパッドデータセット' and a description: 'クックパッド株式会社が国立情報学研究所と協力して研究者に提供しているデータセットです。' (This is a dataset provided by Cookpad Inc. in collaboration with the National Institute of Informatics for researchers). The update date is '2015/05/29 更新'. Below this is a 'データ概要' (Data Overview) section with the text: 'クックパッドに掲載されたデータで、172万品のレシピやそれらからなる献立に関するデータが含まれています。' (The data includes 1.72 million recipes and related menu data from Cookpad). A list of updates follows, with the first item being '1. レシピデータ 2015-02-24 新規' (1. Recipe Data 2015-02-24 New), which describes the data as recipes published on Cookpad before September 30, 2014, including titles, summaries, ingredients, and categories.

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

# 本研究の目的

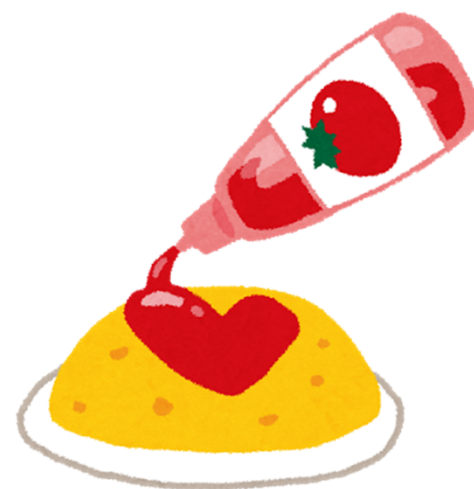
レシピ投稿サイトからの知識発見や  
レシピデータの効果的な利用を目指す

- 代替食材発見
- 調理手順の構造化



# 間接相関ルールを用いた 代替食材の抽出

---



# 先行研究: レシピマイニング

代替可能食材  
の特定

[志土地11, 野沢14]

食材間依存  
関係の  
抽出

[鈴木15]

食材間  
ネットワーク  
の  
構築と推薦

[Teng12]

食材の  
組み合わせに  
おける  
典型性, 意外性

[横井15, 池尻13, 池尻14]

オノマトペ  
との関係抽出

[渡辺15, カンウィパー  
09, カンウィパー10]

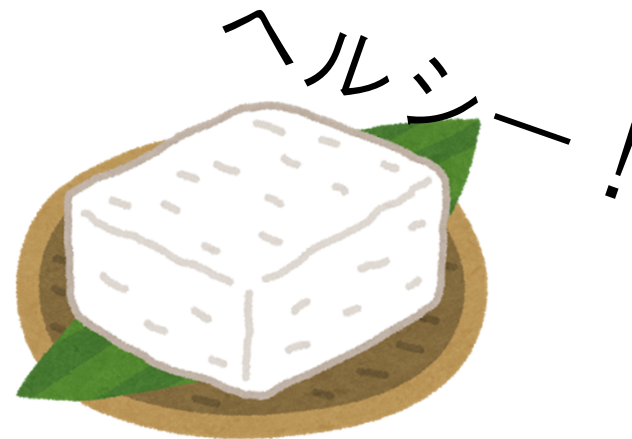
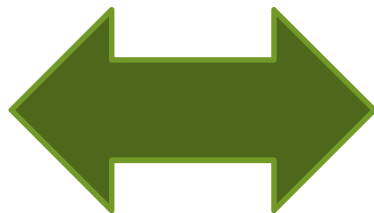
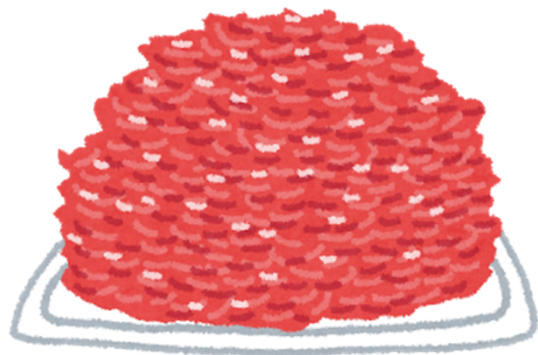
フローグラフ  
(調理手順を  
表すグラフ)

[Mori14, 山肩14]

類似レシピの  
要約

[杉山13]

# 代替食材について



簡単に!



# 代替食材を探す目的

多様なニーズに応えたい！

間接相関ルール

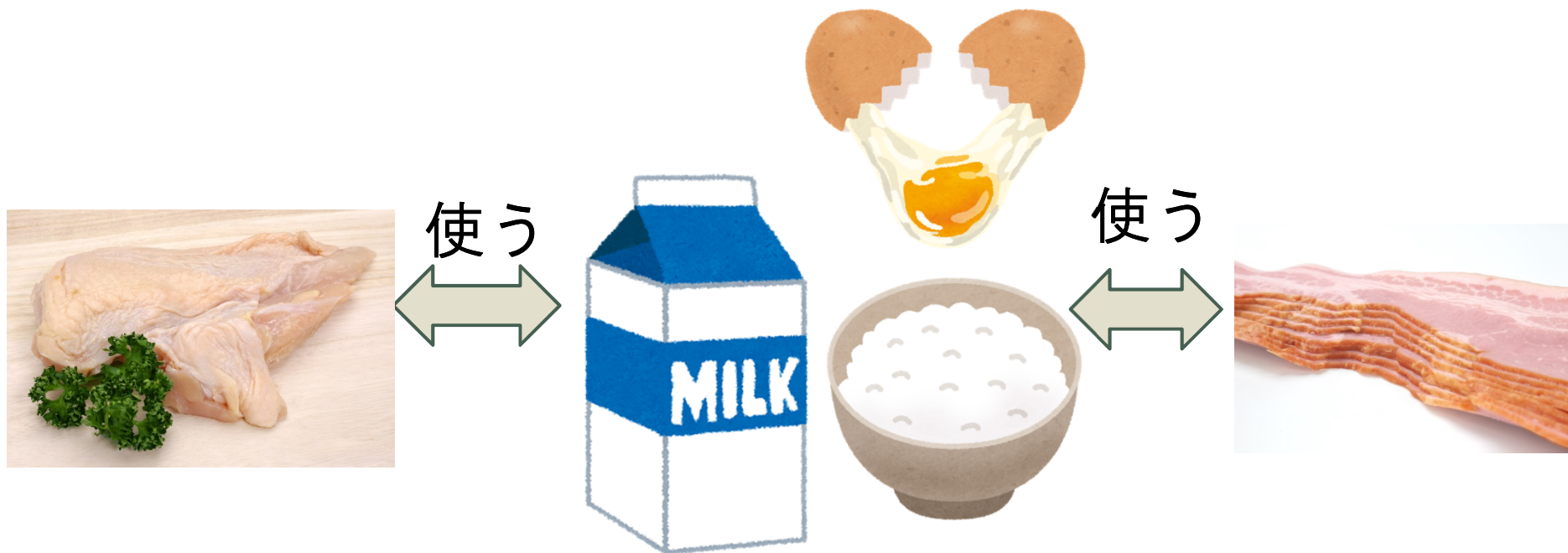
目的に応じた  
評価関数によるランキング

評価：順位相関



# 間接相関ルール

( Indirect Association Rules )



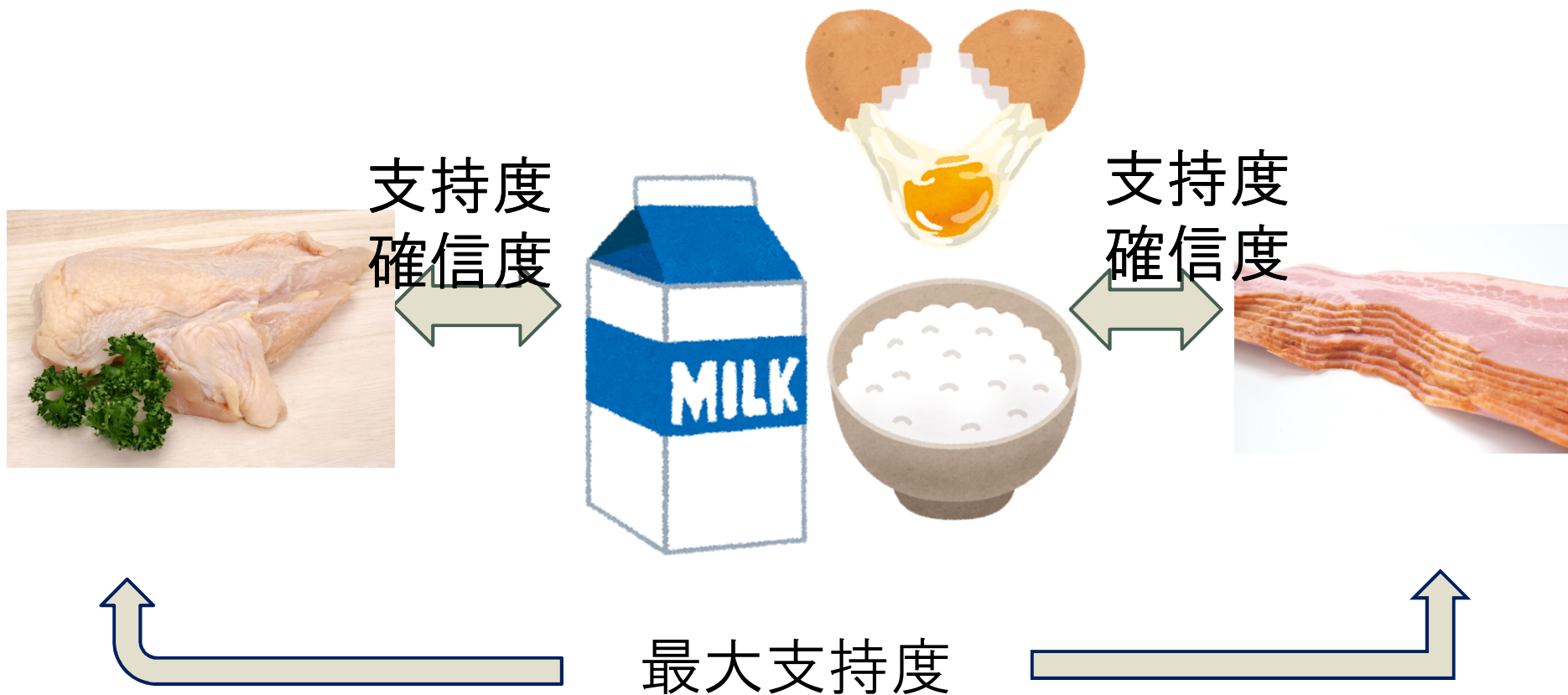
使わない!!!

同じ役割を果たす対ライバルの関係にある対 この条件を満たす対 **代替可能**

# 間接相関ルール

( Indirect Association Rules )

支持度：ルールが登場する頻度を表す  
確信度：ルールの条件と結論の結付き  
最大支持度：共起の関係がない



同じ役割を果たす対ライバルの関係にある対 この条件を満たす対 **代替可能**

## 使用するデータ:クックパッドデータ

データ名	レシピ数	食材種数	食材数 平均	手順数 平均
オムライス	4535	2275	11.8	6.5
ハンバーグ	14277	4530	12.5	6.2

「オムライス」  
「ハンバーグ」  
レシピ名が終わる

# データ成形の流れ

5つ以上のレシピに  
使用されている食材

表記ゆれ

表記の統一

たまねぎ 玉ねぎ 玉ネギ 玉葱 → タマネギ

ライス ご飯 冷や飯 白ごはん → ゴハン

サケ さけ 鮭 シャケ サーモン → ス

醤油 しょう油 しょうゆ お醤油 → ショウユ

プチトマト ミニトマト → ミニトマト

# 抽出された間接相関ルール

閾値を変化させていく

オムライス				ハンバーグ			
支持度	確信度	最大支持度	LAR	支持度	確信度	最大支持度	LAR
	0.1	0.05	61688		0.1	0.05	155265
0.005	0.1	0.005	<u>3300</u>		0.1	0.01	172
	0.3	0.05	61688		0.3	0.05	155265
	0.3	0.005	3300		0.3	0.01	172
	0.5	0.05	32534	0.005	0.5	0.05	<u>2049</u>
	0.5	0.005	3215		0.5	0.01	5



# 抽出されたルール例: オムライス



砂糖, 塩,  
ケチャップ,  
ご飯, 卵

ミックスベジタブル    ニンニク

細かい置換

牛乳,  
バター, ご飯,  
卵

バジル    パセリ

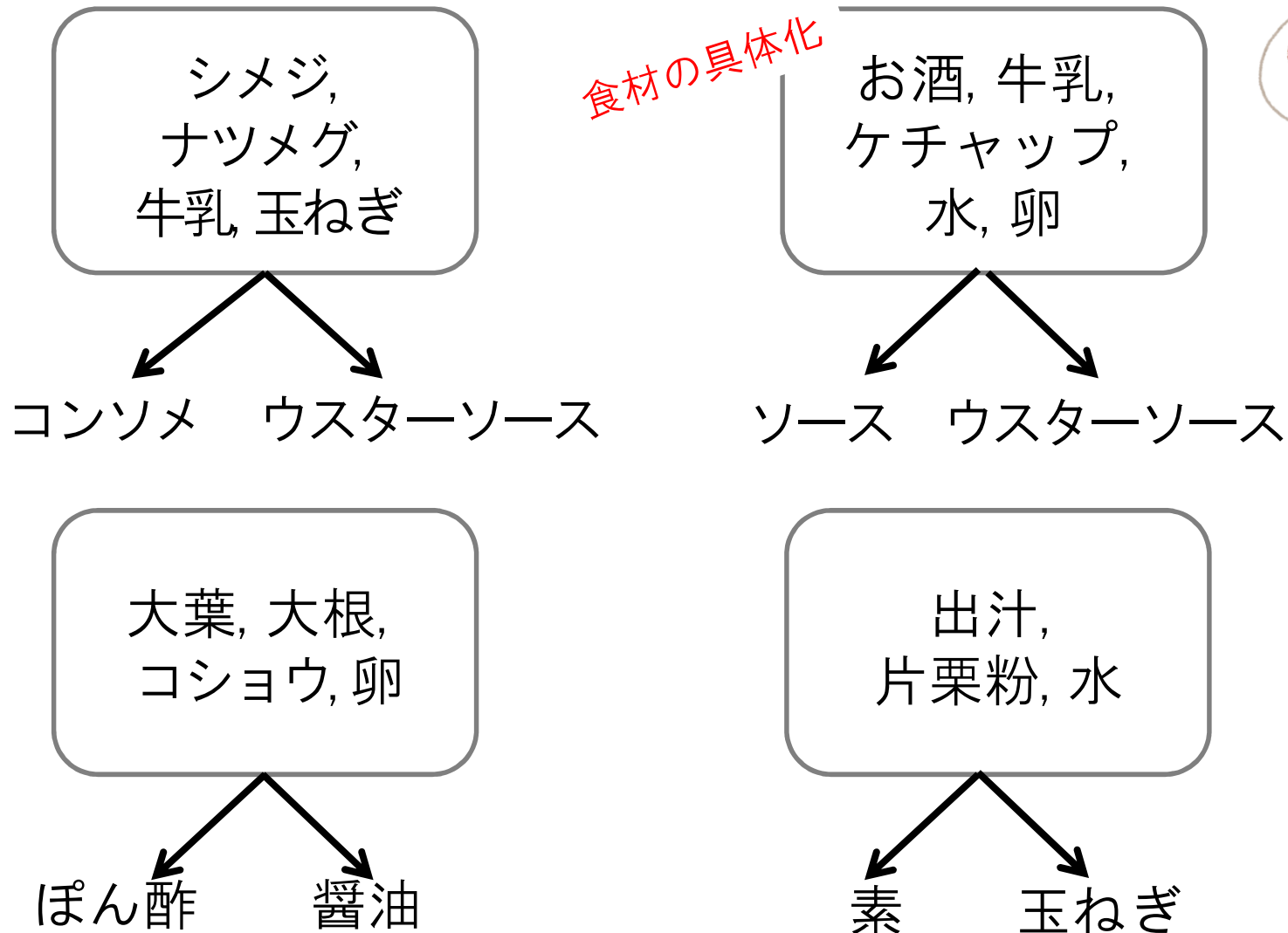
牛乳,  
バター, ご飯,  
卵

トマト    ネギ

サラダ油,  
コショウ,  
ご飯, 卵

お酒    ベーコン

# 抽出されたルール の 例 : ハンバーグ



間接相関ルール



目的に応じた  
評価関数によるランキング



評価：順位相関



# 評価関数

調理手順数  
の  
削減量

食材の多様性

食材  
オントロジー

共起

食材の  
一般性,希少性

# 評価関数: Steps

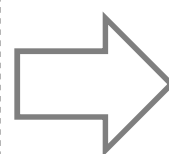
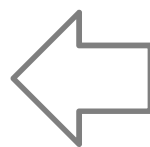
調理手順数の  
削減量

食材  
オントロジー

共起

食材の  
一般性希少性

レシピ集合



代替することによって  
手順が減っているか



| ステップ数平均 - ステップ数平均 |

# 評価関数: OntologyOne

調理手順数の  
削減量

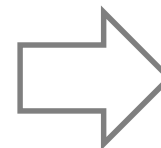
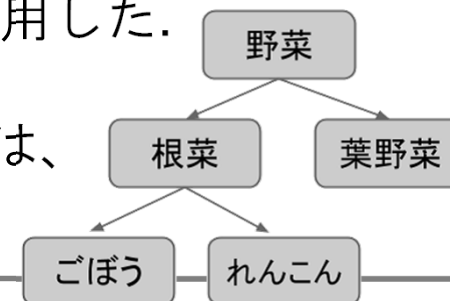
食材  
オントロジー

## オントロジーの構築

料理オントロジー構築プロジェクト

(<http://www.ls.info.hiroshima-cu.ac.jp/cgi-bin/cooking/wiki.cgi>)を利用した.

載っていない食材は、  
人手で足した.



食材オントロジー距離

# 評価関数: OntologyAll

代替する前の食材と  
代替した食材の  
レシピ内の他食材との距離

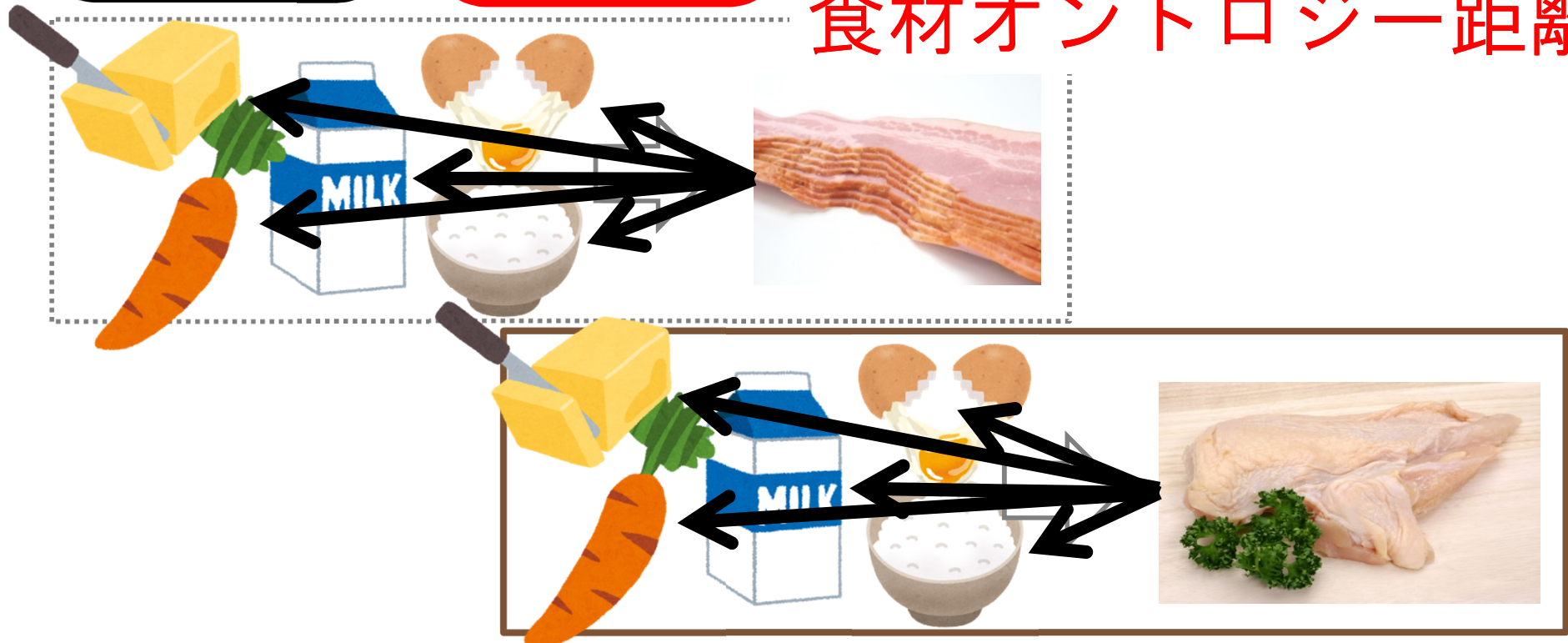
調理手順数の  
削減量

食材  
オントロジー

共起

食材の  
一般性,希少性

食材オントロジー距離



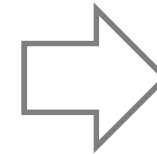
# 評価関数: Pop

調理手順数の  
削減量

食材  
オントロジー

共起

食材の  
一般性,希少性



頻度の差

間接相関ルール



目的に応じた  
評価関数によるランキング



評価：順位相関

# 関数によってどれくらい 順位が異なるのか？

相関が  
弱いものが多い

ハンバーグ



	Pop	Steps	Cosine One	Cosine All	Ontology One	Ontology All
Pop		0.00	0.76	-0.02	-0.42	0.44
Steps	0.03		0.02	0.09	0.10	-0.11
CosineOne	0.08	0.17		0.13	-0.63	0.45
CosineAll	0.03	0.02	0.42		0.19	-0.21
OntologyOne	0.18	-0.10	-0.12	-0.14		-0.62
OntologyAll	0.00	0.06	0.12	-0.04	-0.30	

オムライス



スピアマンの順位相関

$$\rho = 1 - \frac{6 \sum D^2}{N^3 - N}$$

-1 ~ 1 (0は無相関)

負 正

# 全体だけでなく 上位のランキングの類異性は？

種々の目的に応じた  
ランキングを実現

ハンバーグ



	Pop	Steps	Cosine One	Cosine All	Ontology One	Ontology All
Pop		0.00	0.74	0.10	0.00	0.01
Steps	0.00		0.00	0.05	0.01	0.02
CosineOne	0.09	0.01		0.00	0.00	0.00
CosineAll	0.02	0.00	0.19		0.00	0.04
OntologyOne	0.00	0.00	0.00	0.00		0.00
OntologyAll	0.00	0.02	0.00	0.00	0.00	

オムライス



Jaccard類似度 (50位タイ)

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

0 ~ 1



# 利用者アンケート

得られたランキングの  
妥当性を評価する

データ：オムライス

6つの基準の各ランキング上位10位タイまで  
目視で良いと選定したルール，計90件.

## 実験内容

ランダムに15個のルールを解答者が  
独自の感覚で良いと思った順に並び替える.  
10人の各解答者毎に12セット.



## 利用者アンケート：順位相関

	90ルール	上位10位タイ
Pop	0.00	-0.83
Steps	-0.20	-0.38
CosineOne	-0.50	-0.47
CosineAll	-0.39	0.47
OntologyOne	0.50	0.08
OntologyAll	-0.52	-0.12



利用者アンケートで得られたランキングと、  
各評価関数でのランキングが  
どれくらい似ているのか？



利用者アンケート結果から

# 平均順位高得点順:オムライス

3.5位

牛乳,バター,  
ケチャップ,  
ご飯,卵

鶏胸肉

ベーコン

3.95位

牛乳,バター,  
ケチャップ,  
ご飯,卵

ハム

鶏肉

4.15位

牛乳,バター,  
ケチャップ,  
ご飯,卵

トマト

コーン缶

4.2位

牛乳,バター,  
ご飯,卵

鶏もも肉

ウィンナー

4.3位

牛乳,塩,  
バター,  
ケチャップ,卵

ケチャップライス

チキンライス

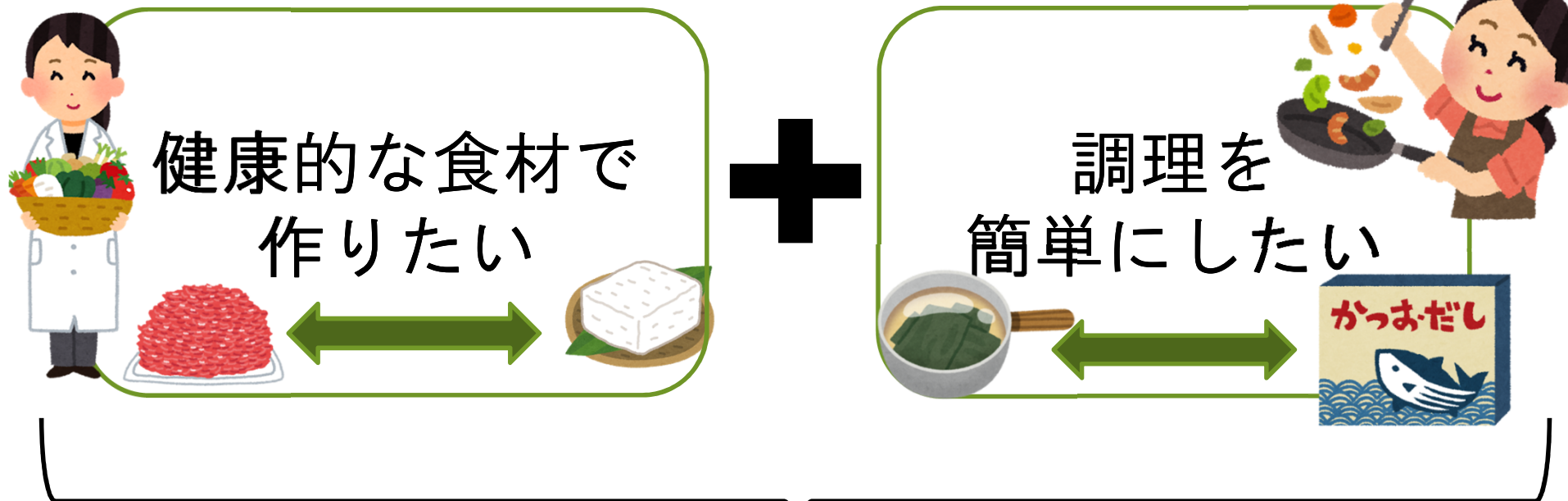
4.4位

牛乳,バター,  
ご飯,卵

鶏もも肉

ベーコン

# 代替可能食材を選ぶ際. . .



複数の基準を同時に考慮する

## 複合的な基準を使う

各評価関数を重回帰分析を用いて複合的に評価する

# 回帰分析結果

利用者が何を基準を  
代替食材を  
選んでいるのか？

目的変数：アンケート結果によるランキング  
説明変数：各基準によるランキング



	係数	P-値
切片	0.076964	0.01
Pop	0.001033	0.08
Steps	-0.00015	0.40
CosineOne	-0.00107	0.05
CosineAll	-0.00021	0.24
OntologyOne	9.13E-05	0.60
OntologyAll	-0.00033	0.07

# まとめ

代替可能食材を表す  
間接相関ルールの評価関数を複数提案し、  
種々の目的に応じたランキングを実現した

# 今後の課題

- ・ 食材の重要度,  
調味料との相性,  
オントロジーの上位下位  
を考慮した代替食材の提案
- ・ 様々なレシピに対応し、精度の高い代替食材発見





# 調理手順本に基づく プリンレシピクラスタリング

---

# 研究目的

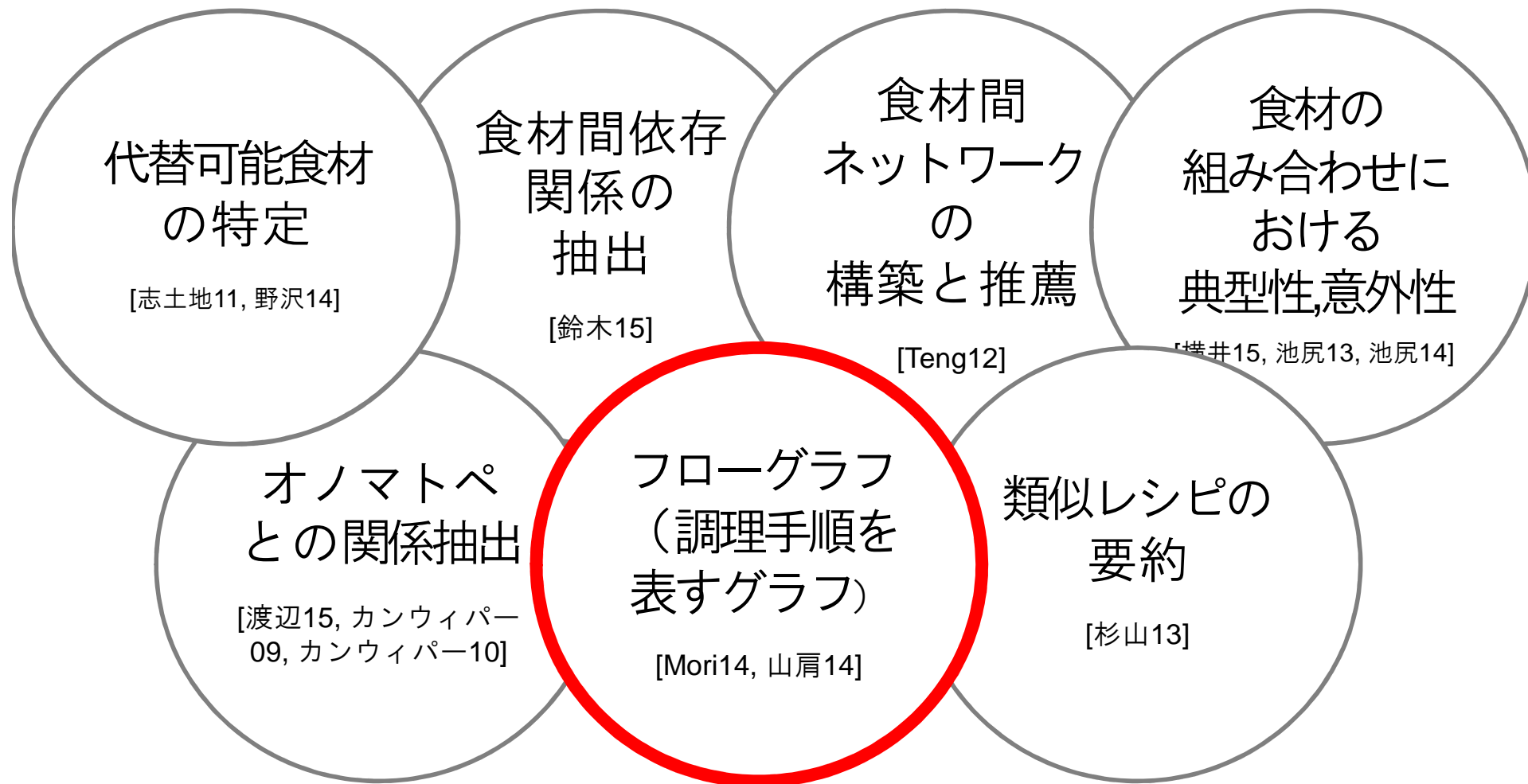
レシピデータを  
より効果的に  
利用したい！



料理手順に着目した  
分類・構造化の有用性



# 先行研究: レシピマイニング



# 使用するプリンデータ

## クックパッド

- レシピタイトル

“～プリン”



約6000件



148件選択

## 書籍

自由が丘の  
プリン専門店 ケミユ著  
『プリン屋さんの  
プリンレシピ  
(マイライフシリーズ)』



36件選択

# 研究の流れ



調理手順木  
コーパス作成

調理手順における

食材の  
Jaccard距離

類似性

木の  
編集距離

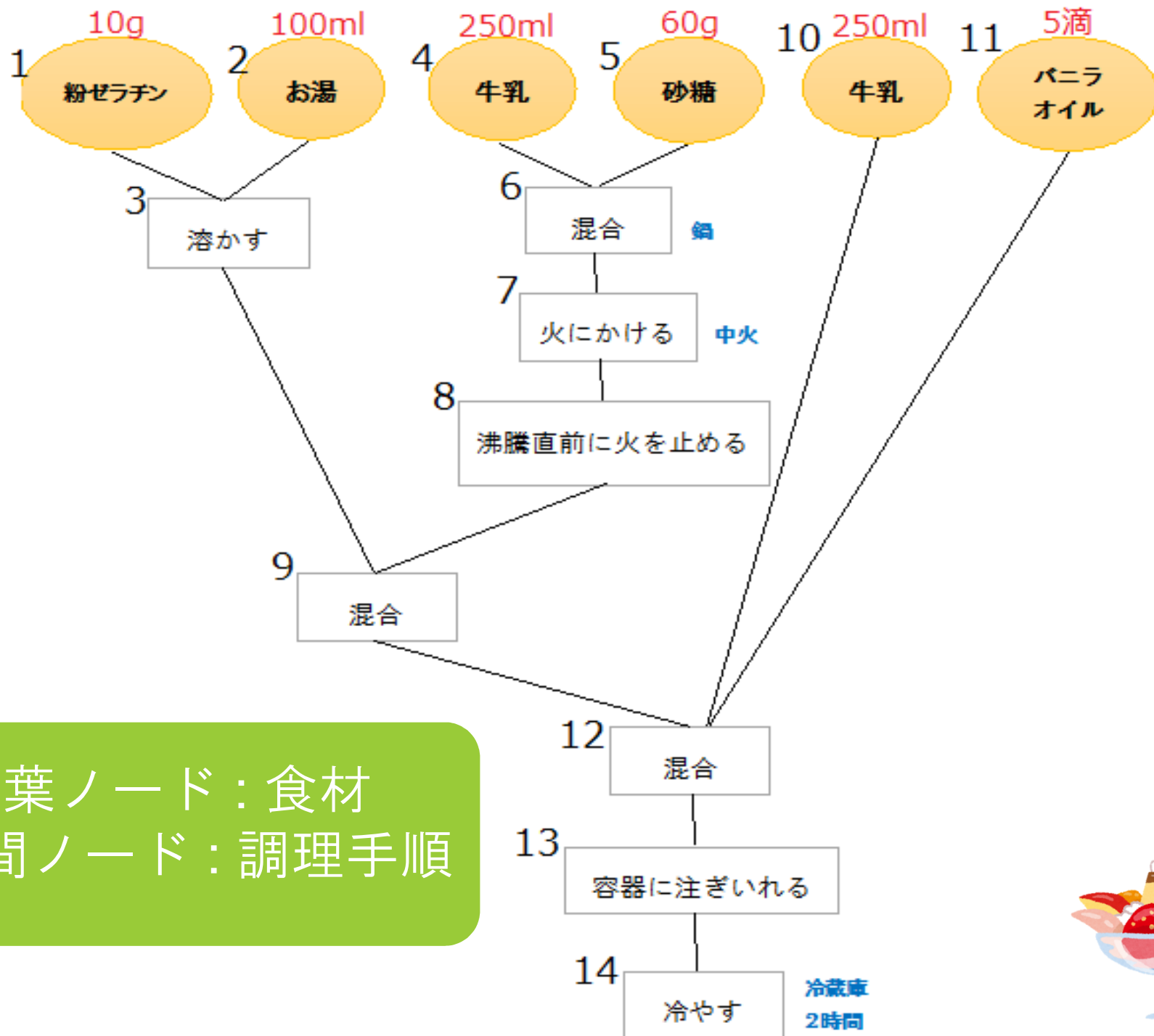
文書間距離 ×



中心的なレシピ  
の抽出

例外的レシピ  
の検出





葉ノード：食材  
 中間ノード：調理手順



# 調理手順木コーパスの作成にあたって

手作業での  
構造化

自動化の精度の問題を回避

表記揺れ  
への対処

例) お湯, 湯 → お湯  
牛乳, ミルク → 牛乳

有用性

手順のコーパスはありません！



# 調理手順木コーパスの概要



平均手順数  
= 全体-食材数

## クックパッド

データ数 : 148

10.1

書籍が  
手順数  
多い

	平均	標準偏差	最大	最小
ノード数 (全体)	16.8	5.6	32	5
リーフ数 (食材)	6.7	2.2	12	2
高さ	8.7	2.5	21	3

## 書籍

データ数 : 36

平均手順数  
= 16.4

書籍	平均	標準偏差	最大	最小
ノード数 (全体)	24.3	6.4	34	12
リーフ数 (食材)	7.9	2.4	14	3
高さ	12.5	2.8	21	6

# 研究の流れ



食材の  
Jaccard距離

調理手順木  
コーパス作成



木の  
編集距離

類似性

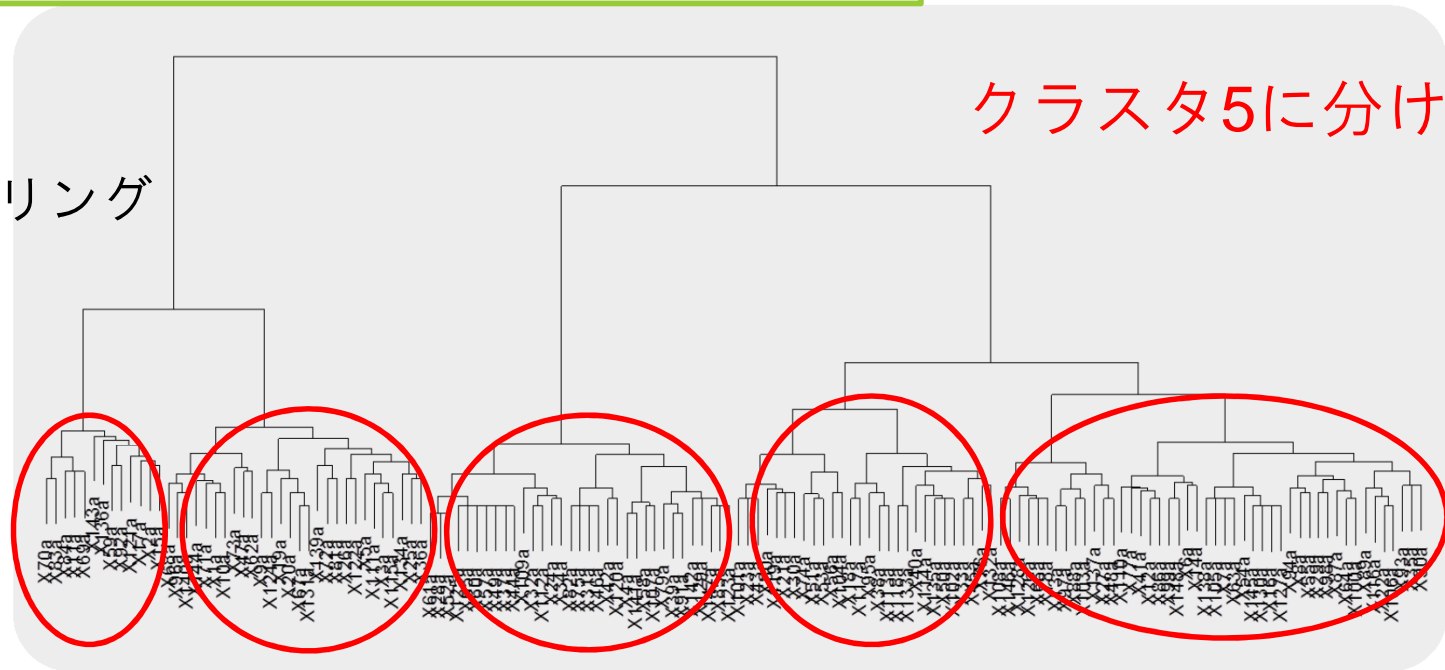


中心的なレシピ  
の抽出

例外的レシピ  
の検出

# ①階層的クラスタリング-ward法-

階層的クラスタリング  
における  
デンドログラム



## ② $k = 5$ で medoids 法実行

5つにクラスタリングし、  
各クラスタの中心的なレシピ（セントロイド）  
を出す

手順距離

VS

食材距離

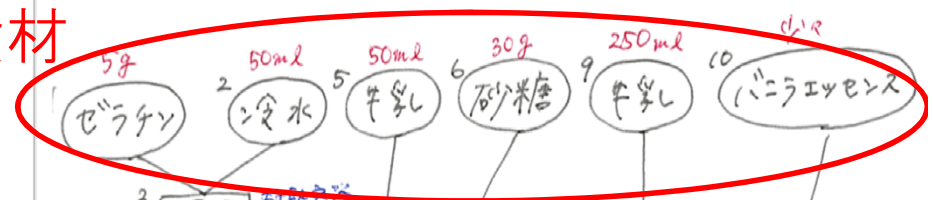


# 手順距離のセントロイド①

## 構造化

<21> id: 04c895896047ba3352f30ea80349e2f4

食材



3.混合 耐熱容器

4.冷やす 冷蔵庫 15分

7.混合

8.温める 電子レンジ 500W (1分30秒)

11.混合

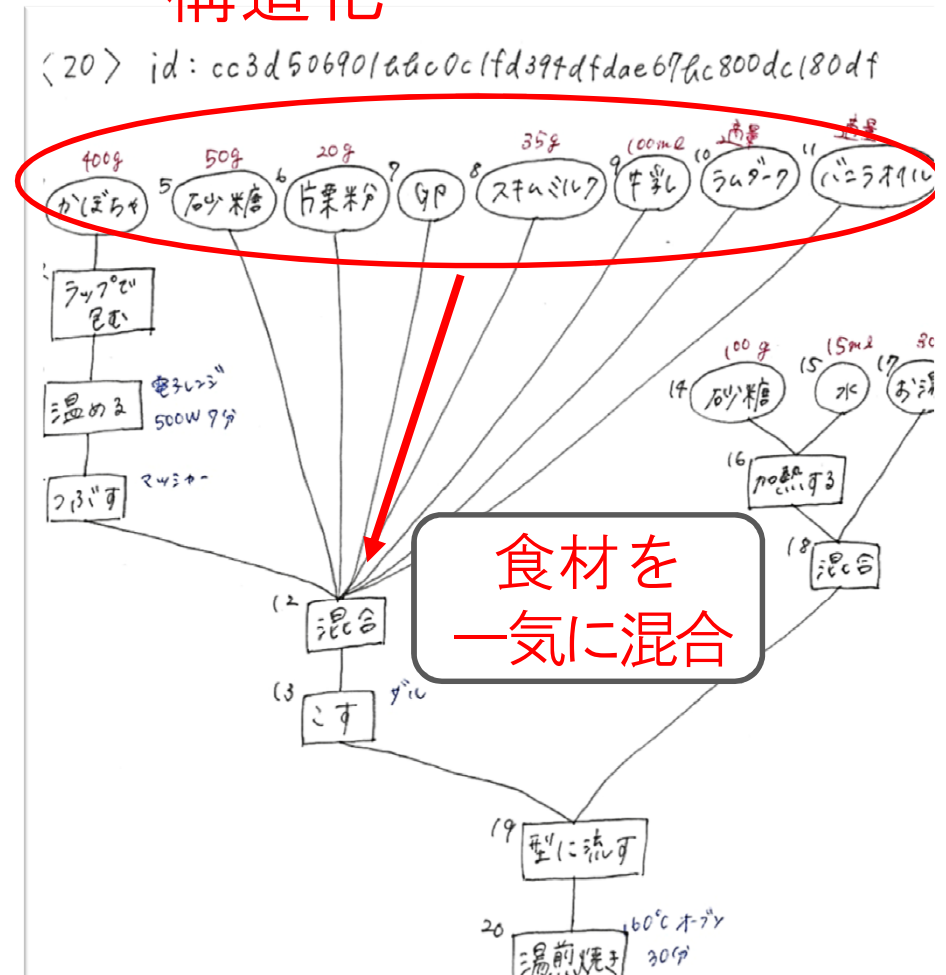
プリン  
の  
固め方

12.冷やす 冷蔵庫

一般的な材料で  
簡単な調理手順

# 手順距離のセントロイド②

## 構造化



手間がかかる料理だが  
その中でも手順がシンプル  
なレシピ

# セントロイド比較

手順距離

食材距離

手順はシンプル  
で簡単！

牛乳や卵，ゼラチンなどの  
一般的な食材  
食材・手順数が小さい

セントロイドの重複がない



手順を考慮することで，  
食材集合では得られない  
特徴的なレシピが得られた

# クラスタリング結果の類似性

値が大きい程  
類似性が高い

同一クラスタになる判  
定の正解率を求める

**Rand**尺度

		Rand尺度		手順距離		食材距離	
				Ward法	K-medoids	Ward法	K-medoids
		クラスタ	5	10	5	10	
手順距離	Ward法	5	0.77	0.64	0.71		
	K-medoids	10	0.38	0.67	0.76		
食材距離	Ward法	5	0.06	0.12	0.76		
	K-medoids	10	0.09	0.19			

相互情報量における類似性を  
求める

正規化相互情報量 (**NMI**)

他の値に比べて  
手順距離と食材  
距離による値が  
小さい

# 研究の流れ



食材の  
Jaccard距離

調理手順木  
コーパス作成



木の  
編集距離

類似性



中心的なレシピ  
の抽出

例外的レシピ  
の検出

# 例外性レシピの検出

密度に基づく  
例外性尺度

## 例外性ランキング上位5レシピ

順位	手順距離		食材距離	
	例外値	他基準での順位	例外値	他基準での順位
1	2.8	21	37	142
2	2.6	140	1.8	14
3	2.3	63	1.7	115
4	2.3	48	1.6	128
5	2.3	144	1.6	

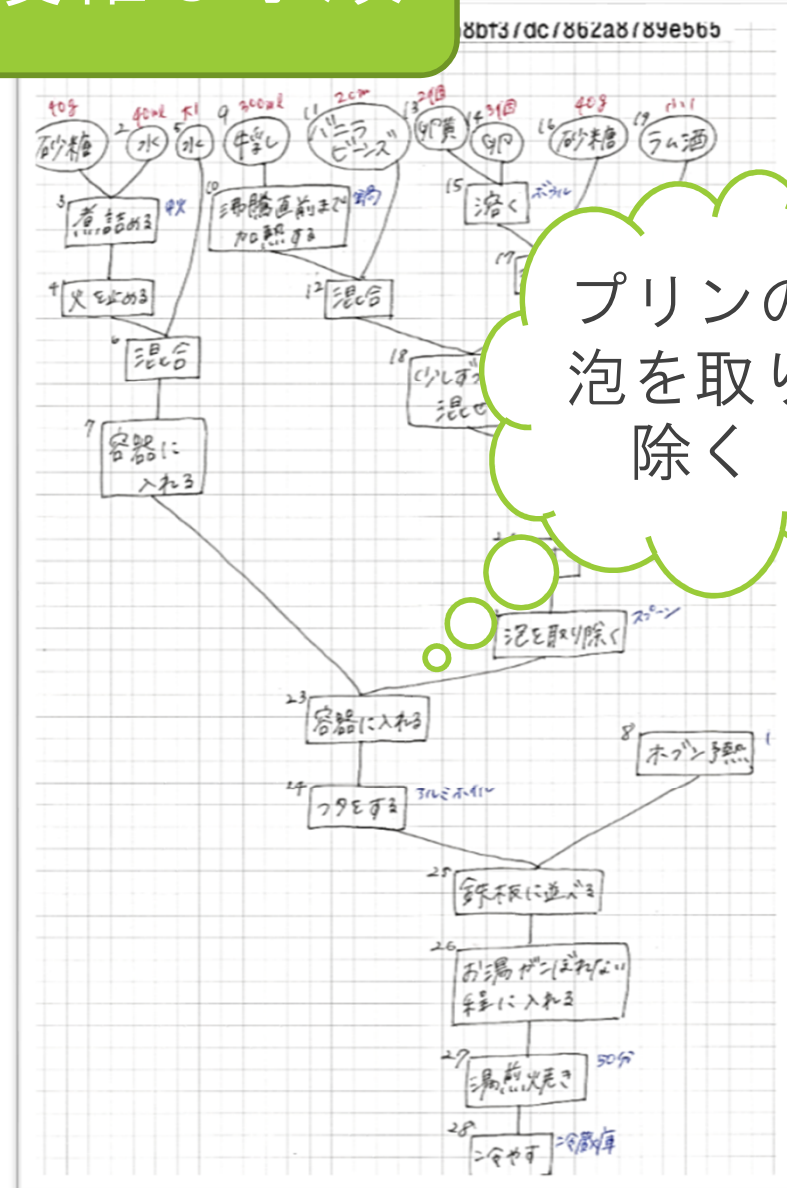
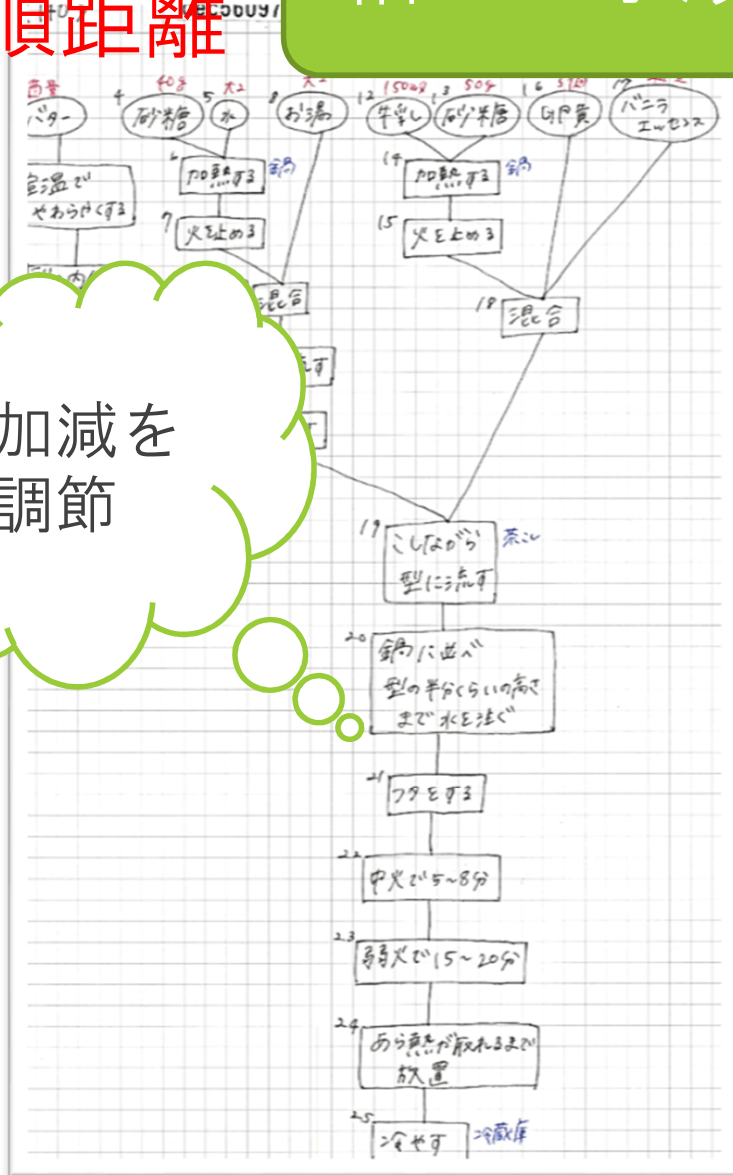
他基準とは違う

# 例外性 手順距離

## 細かい手順や複雑な手順

火加減を  
調節

プリンの  
泡を取り  
除く



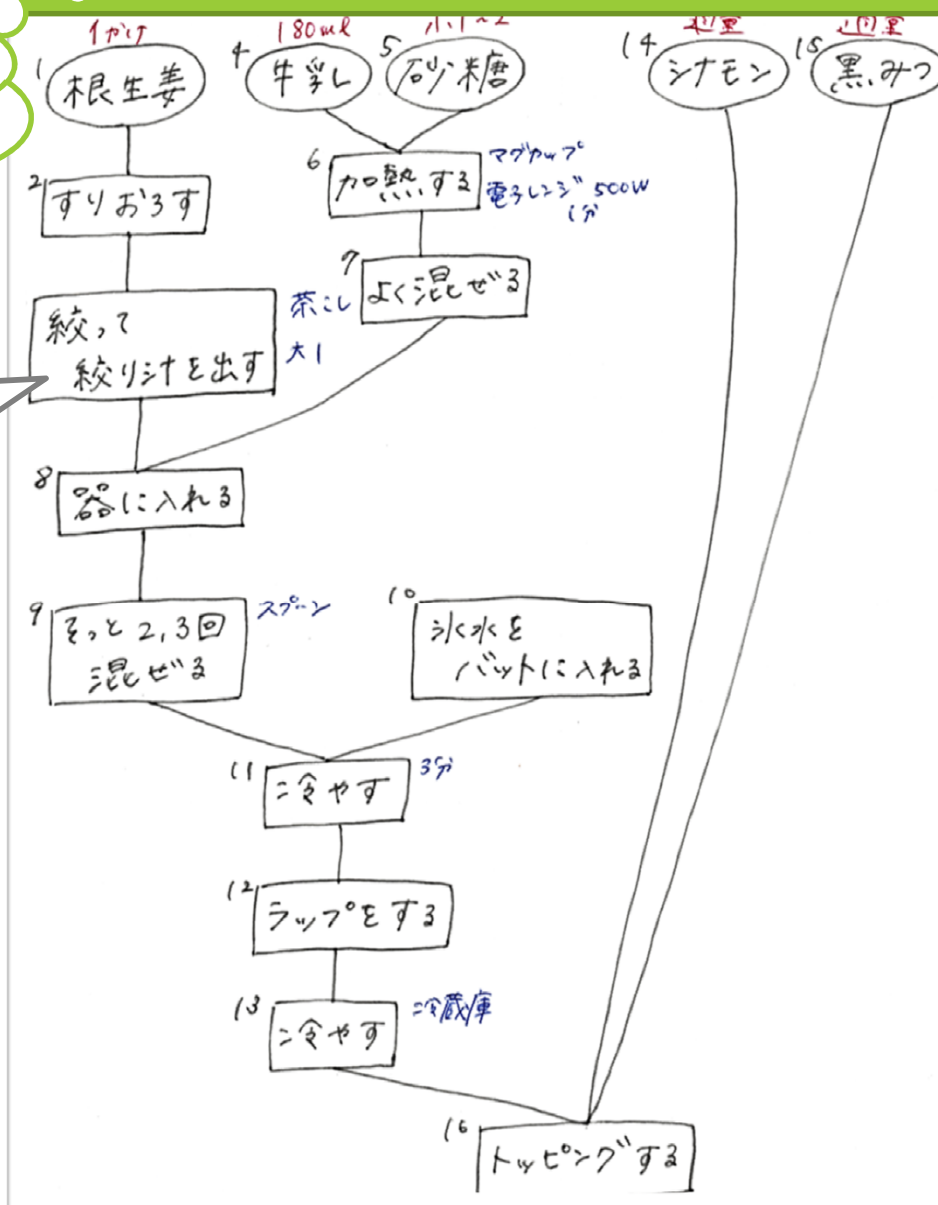
# 例外性 食材距離

## ダイエット食品などの 特殊な食材を使用

葛粉, 甘酒, 根生  
姜, 黒蜜など

例外値4位のレシピは  
手順距離のセントロイ  
ド

以上より  
調理手順の考慮は1つの  
レシピに対し, 様々な評  
価が与えられることに繋  
がる





# まとめ

- プリン184件構造化できた.
- 食材集合と調理手順では特徴を抽出.

# 今後の課題

- 他の料理を対象とした評価
- 文書間距離の比較
- 食材・手順の重要性の考慮





## まとめ

代替可能食材を表す間接  
相関ルールの評価関数を  
複数提案し，種々の目的  
に応じたランキングを実  
現した

- ・プリン184件構造化できた.
- ・食材集合と調理手順では特徴を抽出.

## 今後の課題

- ・食材の重要度，調味料との相性，オントロジーの上位下位を考慮した代替食材の提案
- ・様々なレシピに対応し，精度の高い代替食材発見

- ・他の料理を対象とした評価
- ・編集距離の正規化を試したが，上手くいかないため更なる改善をする.