

## 調理手順木に基づくレシピクラスタリング

石井 悠加里<sup>†</sup> 尾崎知伸<sup>†</sup>

<sup>†</sup> 日本大学 文理学部

### 1. はじめに

近年、レシピサイトの利用者の増大に伴い、大規模なレシピデータが容易に収集・蓄積されるようになってきている。蓄積されたレシピデータをより効果的に利用するためには、レシピデータを分類・構造化した上で、共通点や差異など、各レシピが持つ特徴を抽出することが重要となる。本研究では、レシピデータを構造化する一手法として、調理手順に基づくレシピクラスタリングを取り上げる。代表的なスイーツであるプリンを題材に、調理手順を木構造で表現し、木の編集距離に基づき種々のクラスタリングを実行する。また、利用する食材に基づいたクラスタリングとの比較を行い、手順や木構造を考慮することの有用性などについて考察を行う。

### 2. 調理手順木コーパスの作成

調理手順の精密な構造化には、調理の流れをグラフで表すフローグラフ<sup>1)</sup>などが用いられるが、本研究ではより簡易な構造化として、木構造を用いた構造化を採用する。

クックパッド株式会社が国立情報学研究所と協力して提供しているクックパッドデータセット<sup>☆</sup>に含まれる、レシピ名が“プリン”で終わる約6000件のレシピから無作為に約150件を選択し、それぞれの調理手順を文献<sup>2)</sup>等を参考に木構造に変換することで調理手順木コーパスを作成した。なお、レシピテキストからフローグラフを自動的に構築する手法<sup>3)</sup>も提案されているが、本研究では自動化の精度の問題を回避するため、変換は手作業で行っている。その際、食材や手順、調理器具を木のノードとして採用し、また表記ゆれへの対処を行っている。

調理手順木の例を図1に示す。図に示す通り、今回構築した調理手順木では、食材を葉ノードに配置し、それらに対する各調理手順が中間ノードとなる。また本来、根ノードには出来上がった料理が対応するが、今回はすべてが“プリン”となるため、省略することとした。得られたコーパスの基本的な情報を表1に示す。

表1 調理手順木コーパスの概要

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

データ数：148

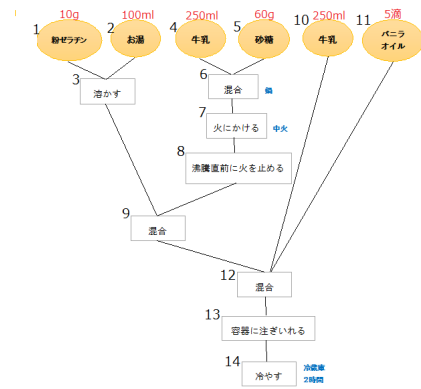


図1 調理手順木の例

### 3. レシピクラスタリング

調理手順に着目することの有効性を確認するため、調理手順木を対象としたレシピクラスタリングと、食材集合（調理手順木における葉集合）を対象としたクラスタリングをそれぞれ行い、その結果を比較、考察する。本研究では、調理手順木のクラスタリングにおける非類似度尺度として木の編集距離<sup>4)</sup>を、また食材集合のクラスタリングにおける非類似度尺度として Jaccard 距離を採用した。また、クラスタリング手法として、ward 法による階層的クラスタリングと、k-medoids 法による非階層的クラスタリングを用いた。

#### 3.1 獲得されたクラスタの概要

図2に、階層的クラスタリングにより得られたデンドログラムを示す。また表2に、階層的クラスタリング（クラスタ数5）と非階層的クラスタリング（クラスタ数5）により得られる各クラスタの概要を示す。

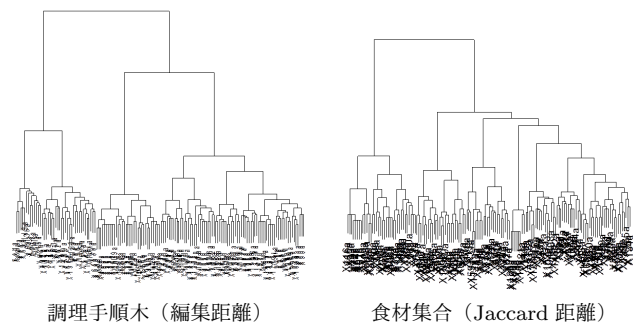


図2 階層的クラスタリングの結果

図2に示すデンドログラムより、調理手順木及び食材集合の双方で、大きく4~5のクラスタが形成されていることが分かる。その一方で、階層的クラスタリング結果からk=5としてクラスタを抽出した場合（表2上）、食材集合と比較

Recipe clustering based on flow trees  
Yukari Ishii and Tomonobu Ozaki (College of Humanities and Sciences, Nihon University)  
<sup>☆</sup> <http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html>

表2 得られたクラスタの概要

	調理手順木 (編集距離)				食材集合 (Jaccard 距離)			
	#R	N	L	H	#R	N	L	H
階層的クラスタリング								
0	28	15.7	6.2	8.0	8	11.3	4.4	6.5
1	13	19.4	5.9	12.0	23	21.5	8.3	10.4
2	33	17.8	7.5	8.9	36	13.9	5.9	7.6
3	46	16.2	6.6	8.2	55	16.5	6.6	8.6
4	28	16.4	6.6	8.2	26	18.8	7.2	9.3
非階層的クラスタリング								
0	49	16.7	6.2	9.0	49	14.9	5.5	8.1
1	52	16.9	6.9	8.7	25	21.3	6.4	13.0
2	16	15.7	6.6	8.7	16	16.1	6.6	7.9
3	22	16.3	6.7	8.5	34	17.8	7.5	8.9
4	9	19.2	7.9	8.4	24	14.9	6.3	7.7

#R: レシピ数, N: ノード数平均, L: 葉数平均, H: 高さ平均

し、調理手順木の方がクラスタを構成するレシピ数の分散が小さく、均等なクラスタが得られていることが分かる。

K-medoids 法で得られる調理手順木のセントロイドに着目すると、一般的な材料を利用し簡単に調理するレシピが3件、最後にトッピングするタイプのレシピが1件、かぼちゃプリンのような手順数が多いレシピ1件となった。これに対し、食材集合のセントロイドはすべて、牛乳や卵、砂糖、ゼラチン、生クリームなど、一般的な食材のみを利用したレシピであり、また食材数・手順数ともに小さなものであった。なお、調理手順木と食材集合間でセントロイドに重複はなく、手順を考慮することで、食材集合だけでは得られない特徴的なレシピが得られる可能性が示唆された。

### 3.2 クラスタリング結果の類似性

各基準で得られたクラスタリング結果に対し、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) 及び同一クラスタになる判定の正解率を求める Rand 尺度を用い、その類似性を確認する。表3に結果を示す。

表3 クラスタリング結果の類似性: NMI (左下), Rand 尺度 (右上)

	$T_5^W$	$T_{10}^M$	$J_5^W$	$J_{10}^M$
$T_5^W$		0.77	0.64	0.71
$T_{10}^M$	0.38		0.67	0.76
$J_5^W$	0.06	0.12		0.76
$J_{10}^M$	0.09	0.19	0.42	

$T, J$  は調理手順木と食材集合を、 $W, M$  は階層的クラスタリングと非階層的クラスタリングを、添字 5, 10 はクラスタ数を表す。

表3より、全体的に Rand 尺度の値は大きいものの、NMI の値は小さく、手法間で異なるクラスタが得られていることが分かる。特に、クラスタ数や手法 (ward 法, K-medoids 法) の違いによる結果の違いと比較し、(同じ設定での) 調理手順木と食材集合の違いによる結果の違いの方が Rand 尺度・NMI の値が小さく、大きく異なるクラスタが得られていることが分かる。これらのことは、手順を考慮することで、食材集合を用いた場合とは異なるレシピ集合の構造化が実現できる可能性を示していると考えられる。

## 4. 例外的レシピの検出

クラスタリングにおけるセントロイドとは逆に、例外的レシピ検出の観点から、調理手順木を利用する効果を検証する。表4に、密度に基づく例外性尺度である局所異常因子<sup>5)</sup>を用いて検出した、各非類似度尺度における例外的レシピ上位5件に対し、当該レシピの LOF 値と他方の非類似度尺度における例外性順位を示す。

表4 例外性ランキング上位5レシピ

順位	調理手順木 (編集距離)		食材集合 (Jaccard 距離)	
	LOF 値	他基準での順位	LOF 値	他基準での順位
1	2.8	21	2.1	37
2	2.6	140	1.8	142
3	2.3	63	1.7	14
4	2.3	48	1.6	115
5	2.3	144	1.6	128

表4より、手順と食材のどちらに着目するかで、同一レシピの例外性が大きく異なることが分かる。調理手順木において例外性が高いと判断されたレシピは、全体的にノード数が多く、火の調整を細かくするなど複雑な手順が含まれているものであった。また、食材集合において例外値4位のレシピは、調理手順木におけるセントロイドであり、食材集合が特殊でも手順としては一般的なレシピであった。これらの結果から、調理手順の考慮は、同一レシピに対し多様な評価を与えることにつながると考えられる。

## 5. おわりに

本研究では、調理手順木を用いたプリンレシピ集合の構造化の可能性について、クラスタリング及び例外発見を用いて実験的に考察を行った。今後の課題としては、他の料理を対象とした評価や、(カラメルやプリン本体など) 部分手順木の抽出とそれらの組み合わせ利用、調理手順木と食材集合の併用の可能性の追求などがあげられる。

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) S. Mori, H. Maeta, Y. Yamakata and T. Sasada : Flow Graph Corpus from Recipe Texts, International Conference on Language Resources and Evaluation, 2014.
- 2) 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 (DEIM), No.D3-5, 2013.
- 3) 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, 信学技報, vol.114, no.204, DE2014-27, pp.25-30, 2014.
- 4) K. Zhang and D. Shasha : Simple Fast Algorithms for the Editing Distance between Trees and Related Problems, SIAM Journal on Computing, Vol.18, No.6, pp.1245-1262, 1989.
- 5) M. M. Breunig, H.-P. Kriegel, R. T. Ng and J. Sander : LOF: Identifying Density-based Local Outliers, Proc. of the 2000 ACM SIGMOD International Conference on Management of Data, pp.93-104, 2000.