

# cookpad の献立データからのパターン抽出

奈須 日向太<sup>1,a)</sup> 風間 一洋<sup>2,b)</sup>

**概要:** 近年の cookpad のようなユーザ投稿型レシピサイトの普及により、ユーザは多彩なレシピを活用できるようになったにも関わらず、日常的に献立を考える際のユーザの負担はまだまだ解消されているとは言い難い。そこで、本稿ではユーザの嗜好に応じた柔軟な献立推薦を目指して、献立データから献立パターンを抽出する手法を提案する。献立パターンは、経験則に基づいた料理の抽象的な組み合わせであり、レシピに付与されたカテゴリで表現する。このカテゴリの木構造を、献立データから献立パターンへの抽象化や、献立パターンから実際の献立への具体化に利用する。そこで、カテゴリ木の部分木として表現した献立データに宇野らの LCM アルゴリズムを用いて頻出飽和部分木を抽出した後に、そのカテゴリ階層のばらつきを調整して、献立パターンを抽出する。実際に、cookpad から提供された献立データに提案手法を適用し、その有効性を示す。

## 1. はじめに

料理のレシピは、以前は料理本のような紙媒体で流通していたが、近年の cookpad<sup>\*1</sup> や楽天レシピ<sup>\*2</sup> などのユーザ投稿型レシピサイトの普及により、投稿された多彩なレシピを活用できるようになった。

しかし、逆にそのレシピの膨大さゆえに目的にあったレシピを探すことはより困難になり、日常的に食事の献立を考える際のユーザの負担は解消されているとは言い難い。例えば、献立データは複数の Web サイトで公開され、cookpad ではユーザが投稿できるが、実際に献立データを利用しようとしても、レシピと比べて利用できる献立データ数ははるかに少なく、一部のレシピがユーザの嗜好に合わない、価格が安い・冷蔵庫に残っている食材の活用が難しいなどの問題があり、そのまま利用するのは難しい。

本稿では、ユーザの嗜好に応じた柔軟な献立推薦を目指して、複数のレシピを組み合わせた献立データから献立パターンを抽出する手法を提案する。献立パターンとは、経験則に基づいた料理の抽象的な組み合わせであり、レシピに付与されたカテゴリで表現する。このカテゴリの木構造を、献立データのレシピの組み合わせから抽象的な献立パターンの抽出や、ユーザの嗜好を考慮した献立パターンからの具体的な献立の生成に利用する。

## 2. 関連研究

レシピの推薦に関して、様々な研究が存在する。木原らは、既存のレシピ検索システムから収集したデータを用いて食材を独自に分量換算し、余剰食材を消費する手法を提案した [1]。三野らは、ダイエットという観点から、カロリーや栄養素、野菜摂取量を制約条件とする推薦システムを提案した [2]。酒井らは、ユーザに入力したクエリの関連語群を提示し、選択された関連語に基づいて検索結果をランキングすることで嗜好や意思を反映させるシステムを提案した [3]。上田らは、レシピの閲覧履歴と摂食履歴から、閲覧しても調理しなかった料理や頻りに調理した料理の食材情報を用いてスコアを算出することで、ユーザの嗜好を推定する手法を提案した [4]。

さらに、複数のレシピを組み合わせた献立の推薦に関しても、様々な研究が存在する。西脇らは、献立のレシピ入れ替えの妥当性を使用食材の類似性を用いて分析することで、ユーザの好みに合わせて献立を変更する手法を提案した [5]。南谷らは、献立を決められないユーザを支援するために、レシピの題名と手順の類似性から入れ替え候補を求めて、対話的に献立を作成する方法を提案した [6]。本稿では、レシピデータの規模と比較すると献立データの規模は小さく限られていることから、利用可能なレシピデータを抽象化したパターンを用いて、ユーザと対話しながら献立を具体化するアプローチを想定している。

## 3. 提案手法

レシピデータと違って、献立データにはユーザ投稿型レ

<sup>1</sup> 和歌山大学大学院システム工学研究科

<sup>2</sup> 和歌山大学システム工学部

a) s206185@wakayama-u.ac.jp

b) kazama@ingrid.org

\*1 <http://cookpad.com/>

\*2 <http://recipe.rakuten.co.jp/>

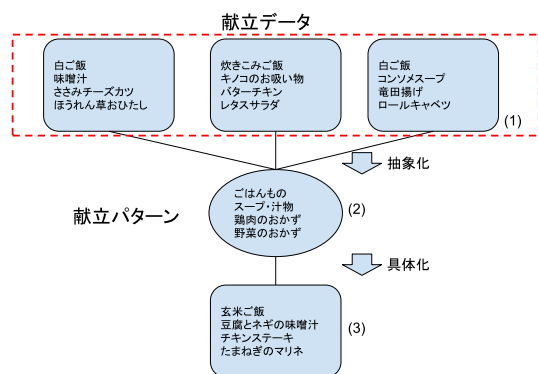


図 1 献立パターンによる抽象化と具体化の例

レシピの利点であるレシピの多様性をうまく活用できないという問題が存在する。まず、ユーザには食材の好き嫌い、賞味期限が近い食材の活用、調理器具の有無などのさまざまな要求条件が存在するが、既存の献立がその要求条件をすべて満たす可能性は低い。もちろん、レシピデータと同様に膨大な献立データが存在するならば、その中から要求条件に合わせて選択すればよいが、実際には献立データは妥当なレシピの組み合わせのごく一部でしかない。

そこで、献立データから、抽象的な献立パターンを抽出して再利用することを考える。献立の作成には、例えば「肉料理には野菜のサラダを付け合わせるとよい」などの、必ずしも明示的ではない多くの経験則が存在するが、献立パターンはその経験則を表現すると考えられる。抽象化・具体化には、レシピに付与されたカテゴリ情報を利用する。このカテゴリは単一の根を持つ木構造として関係づけられ、上位ほど抽象度が高く、下位ほど具体的である。

献立パターンの抽出・利用に関する概念図を図 1 に示す。まず、献立における経験則に基づいた料理の組み合わせを、既存の献立データから抽出した一般化したパターンとして表現する。カテゴリはレシピの特徴に基づいて命名・分類されているので、献立を各レシピに付与されたカテゴリ集合として表現することで、類似レシピをまとめる。さらに、このカテゴリ集合をカテゴリ木の部分木とみなして、頻出する類似部分木の具体的なカテゴリ（例、食材）をまとめて、図 1 の (2) のように上位の抽象的なカテゴリで表現しなおす。ユーザが献立を考える時に、献立パターンを選択した後に、図 1 の (3) のようにユーザの「玉ねぎを消費したい」などの要求を考慮して具体化すれば、多彩な献立を推薦することが可能になる。

以降で、献立パターン抽出手順について述べる。

### 3.1 献立のカテゴリ部分木への変換

献立をカテゴリ部分木として表現するために、献立の各レシピに付与されたカテゴリと、カテゴリ木でそのカテゴリより上位のカテゴリの集合を作成する。ここでカテゴリ部分木を単なるカテゴリ集合として扱う理由は、どのカテ

ゴリ木にも含まれる根カテゴリが一番出現回数が多く、葉に近づくにつれて出現回数が少なくなるために、アイテム集合として扱っても部分木として抽出されることが期待でき、特に木構造を前提としない頻出飽和アイテム集合抽出アルゴリズムを用いることができるからである。もちろん、欠落したカテゴリ間の関係情報を補うことで、抽出したアイテム集合をカテゴリ部分木としても扱うことができる。

### 3.2 頻出飽和部分木の抽出

レシピに付与されるカテゴリは、素材に由来するものが多いが、実際には「オクラ」や「トマト」のような多種類の個別の素材ではなく、「旬野菜（夏）」のようなより抽象度の高いカテゴリにまとめて表すことが望ましい。

そこで、複数の献立のカテゴリ部分木の共通部分を抽出すれば、より抽象度が高い表現が得られる点に着目して、頻出アイテム集合抽出アルゴリズムを適用する。頻出アイテム集合マイニングは、データベース  $T$  で出現頻度が閾値  $\sigma$  を超える頻出アイテム集合を列挙する手法である [7]。ただし、 $\sigma$  を越える全てのアイテム集合を列挙すると、冗長なアイテム集合まで抽出されるので、字野らの頻出飽和アイテム集合を高速に列挙する LCM アルゴリズムを用いる [8][9]。あるアイテム集合  $X$  に対して、 $X$  を真に含み、かつ  $X$  と頻度が等しいアイテム集合が  $T$  中に存在しない場合に、 $X$  を飽和集合とよぶ。LCM は、深さ優先探索に基づくバックトラック型アルゴリズムであり、飽和集合発見の多くの先行アルゴリズムと同様に閉包計算に基づく。通常、閉包計算による飽和アイテム集合の列挙では重複した解が生成されてしまうが、LCM はプレフィックス保存飽和拡張手法を用いることで、頻出飽和集合数に対して線形時間の計算量で列挙できる。

### 3.3 頻出部分木の階層調整

LCM アルゴリズムで抽出した頻出飽和部分木には、2つの問題がある。まず、葉のカテゴリの出現頻度が高い場合には、想定していたような上位カテゴリへの統合が起こらないことがある。例えば、葉カテゴリには食材のカテゴリなどが多く、そのままではユーザの嗜好や時期、入手価格などを考慮して別の代替食材に換えて具体化することはできない。さらに、献立のカテゴリ部分木の同一カテゴリが、頻出飽和部分木によっては別の抽象度のカテゴリとして抽出されることがある。例えば、ある食材カテゴリが上下関係にあるカテゴリに分散してしまうと、献立パターンの利用が難しくなる。

そこで、LCM を用いて抽出された頻出飽和部分木に対して、2段階の階層調整を行う。

まず、下位カテゴリを抽象度の高い上位カテゴリに統合する。カテゴリはレシピに付与するカテゴリと、カテゴリ分類のためのカテゴリに分類できる。前者は葉カテゴリで

あり、直上カテゴリに統合する。

次に、異なる抽象度のカテゴリが存在した場合に、抽象度の高い上位カテゴリを下位カテゴリに分割し、階層を統一する。これは、次の手順で検出・解消する。

- (1) 全部分木に対して、カテゴリの出現頻度と、葉カテゴリだけの出現頻度を調べる。
- (2) カテゴリ木上で深さ優先探索を行い、葉までの経路上で出現頻度が0より大きいカテゴリ集合を求める。
- (3) カテゴリ集合のサイズが1より大きい場合は矛盾が生じているので、上位カテゴリを、出現頻度が閾値 $\rho$ を越える下位カテゴリに分解する。

なお、頻出飽和アイテム集合抽出時に出現頻度が閾値を満たさなかったために上位カテゴリに統合されてしまったので、この処理では抽出時より条件を緩和することになる。ただし、一つのカテゴリが複数のカテゴリに分解されることから、どの献立の部分木にも出現しないようなカテゴリまで加えられて、献立パターンの処理コストが上がらないように、出現頻度が閾値 $\rho$ を越えるカテゴリに限定する。

### 3.4 部分木の献立パターンへの変換

最後に、階層調整を終えた献立の部分木を表すカテゴリ集合から、部分木における最下位カテゴリだけを抽出した集合を作成する。このカテゴリ集合が、献立パターンである。献立パターンは複数の献立に該当するために、抽出に用いた献立数よりも数は少ない。また、各献立のレシピに付与されたカテゴリよりも、カテゴリ木で上位、つまりより抽象的なカテゴリで表現される。

## 4. 評価

### 4.1 データセット

評価には、クックパッド株式会社と国立情報学研究所が提供しているクックパッドデータセットを用いた。このデータセットは、MySQLのバックアップファイル形式で提供され、2014年9月30日までに公開された1,715,595件のレシピと35,928件の献立が含まれる。レシピに付与されるカテゴリは、カテゴリID、カテゴリ名、親カテゴリIDの情報を含むテーブルとして表現され、「カテゴリー」という名前のカテゴリを根とする木構造を持つ。なお、カテゴリは1つのレシピに複数付与される場合もある。

ただし、カテゴリが付与されていないレシピや単品または一部しか含まれていない献立が存在するために、カテゴリが付与されているレシピを2つ以上含む献立10,764件を評価に使用した。

### 4.2 提案手法の定量評価

LCMアルゴリズムで抽出した頻出飽和部分木から直接献立パターンを作成した場合（以降では既存手法と呼ぶ）と、提案手法で献立パターンを抽出した場合を比較して、

表 1 抽出結果の統計値の比較

	既存手法	提案手法
総カテゴリ数	301	54
献立パターン数	938	1603
献立の平均レシピ数	2.15	2.64
根からの平均パス数	3.42	2.92
葉からの平均パス数	1.33	2.00
葉ノード数	234	0
葉ノードの親ノード数	50	50
葉ノードの親の親ノード数	17	4
それ以降の先祖ノード数	0	0

階層調整の有効性を評価した。LCMアルゴリズムの閾値 $\sigma$ は、抽出される献立パターンが1000件程度になる30とした。また、階層調整の閾値 $\rho$ は、LCMの閾値 $\sigma$ 付近の献立パターンに含まれるカテゴリ頻度よりも小さくなるように、300とした。

各手法の抽出結果の統計値を表1に示す。提案手法の方が総カテゴリ数が少ない。これは葉カテゴリの統合処理が有効に働いているからだと思われる。それに対して、パターン数は多い。これは下位カテゴリへの分割時には上位カテゴリを含む1つのパターンが複数のパターンに分割されるからである。また平均カテゴリ数が多いのは、パターン数の増加と共に、カテゴリ数が多いパターンの割合が増えたからだと考えられる。平均パス数から、階層調整のある提案手法の方が、献立パターンがより上位のカテゴリで表現されていることがわかる。その内訳を見ると、提案手法は葉ノード数が0になり、葉ノードの親の親ノードの件数が減少している。すなわち、カテゴリ木上のカテゴリの分散が解消され、同一階層に統一されたことがわかる。なお、提案手法では処理の簡略化のために、分割時のノード選択の基準として頻出飽和パターンにおけるカテゴリ出現頻度を用いたが、階層調整後のカテゴリ数が増加しなかったことから、実際よりも多くの献立パターンが抽出されるような問題は発生していないことがわかる。

### 4.3 提案手法の定性評価

次に既存手法と提案手法を比較して、抽出された献立パターンの妥当性を定性的に評価する。まず、各手法で抽出される献立パターンの一部を図2に示す。図2(a)から、既存手法では「たまご・大豆加工品」カテゴリに加えて、その下の「たまご」カテゴリと、さらに下の葉カテゴリである「卵焼き」カテゴリが別々に抽出される。しかし、図2(b)から、提案手法では「たまご」カテゴリに統一されていることがわかる。同様に、「野菜のおかず」カテゴリも、「旬野菜(春)」と「旬野菜(夏)」などの下位カテゴリに置換され、野菜の季節性が利用できるようになったことがわかる。

さらに、献立パターン群のカテゴリ間の関係を献立パ

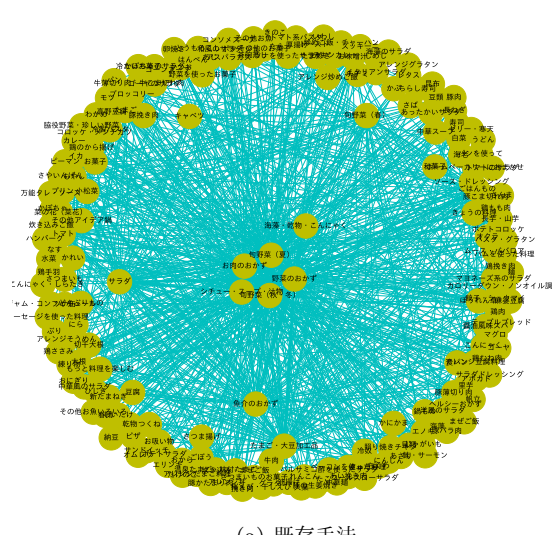
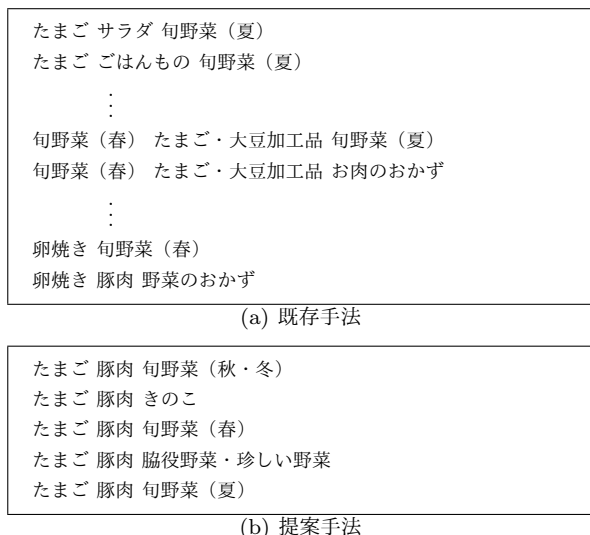


図 2 既存手法と提案手法の抽出結果の例

ターン中のカテゴリ共起数に応じてエッジが太くなるように NetworkX の Spring Layout で可視化して、図 3 に示す。既存手法では、図 3(a) のように「キャベツ」や「鶏もも肉」のような食材のカテゴリが多いが、提案手法では図 3(b) のように上位の抽象的なカテゴリに統合されていることがわかる。さらに、パターンが統合されると共に、閾値以下でも出現可能性があるパターンも加わったことでカテゴリ間の関係に強弱が生じ、カテゴリ間の関係がより明確になっていることがわかる。このような性質は、献立推薦におけるレシピカテゴリの選択に利用できると考えられる。

### 5. おわりに

本稿では、カテゴリ木の部分木として表現した献立データに LCM アルゴリズムを用いて極大頻出部分木を抽出した後に、抽出された部分木の階層のばらつきを調節することで、献立作成時の経験則を抽象的な献立パターンとして抽出できることを示した。

今後は、献立パターンに主食・主菜・副菜の属性を組み込み、実際にユーザの嗜好に合わせて献立パターンを具体化して献立を推薦する手法を実現する予定である。

**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セット」を利用した。また、JSPS 科研費 17H01826 の助成を受けた。

#### 参考文献

- [1] 木原ひかり, 上田真由美, 中島伸介: 余剰食材の使い切りを考慮したレシピ推薦手法の提案, 第 3 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 Forum 2011) E3-3 (2011).
- [2] 三野陽子, 小林一郎: 様々な状況を考慮したユーザに適するダイエットのためのレシピ推薦, 人工知能学会第 26 回全国大会 3C3-2 (2010).
- [3] 犬塚真太郎, 酒井哲也: 拡張クエリを用いたレシピ検索

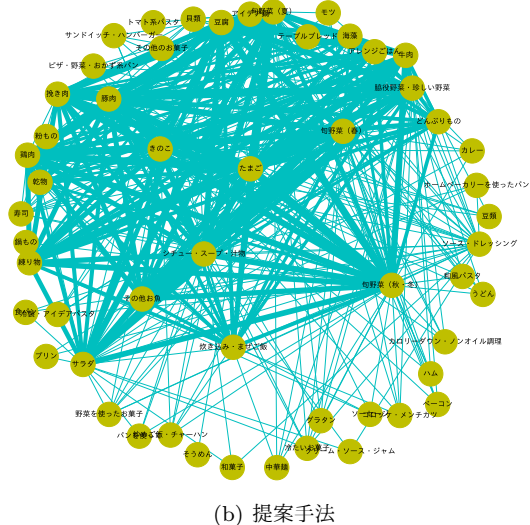


図 3 カテゴリ間の関係の可視化

のパーソナライゼーション, 第 9 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 Forum 2017) C4-2 (2017).

- [4] 上田真由美, 高畑麻理, 中島伸介: レシピ閲覧・摂食履歴を用いた嗜好の抽出, 第 4 回 Web とデータベースに関するフォーラム (WebDB Forum 2011) 3G-1-2 (2011).
- [5] 西脇崇文, 清 雄一, 田原康之, 大須賀昭彦: ユーザの好みと献立バランスを考慮したレシピ入替えを行うグループ向け献立推薦, 第 8 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 Forum 2016) E2-1 (2016).
- [6] 南谷悠大, 前田 亮: 組み合わせを考慮した献立検索手法の提案, 第 10 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 Forum 2018) P8-7 (2018).
- [7] 松澤裕史: 大規模データベースからの頻出構造化パターンの抽出, 情報処理学会論文誌: データベース, Vol. 42, No. SIG 08(TOD 10), pp. 21-35 (2001).
- [8] Uno, T., Kiyomi, M. and Arimura, H.: LCM ver. 2: Efficient Mining Algorithms for Frequent/Closed/Maximal Itemsets, *Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI '04)* (2004).
- [9] 宇野毅明, 有村博紀: 飽和集合列挙アルゴリズムを用いた大規模データベースからのルール発見手法, 統計数理, Vol. 53, No. 2, pp. 317-329 (2005).