

料理教材テキストからの素材と調理法に関する知識の抽出*

唐沢 隆†, 浜田 玲子††, 井手 一郎†††, 坂井 修一††, 田中 英彦††

{mak,reiko,sakai,tanaka}@mtl.t.u-tokyo.ac.jp ide@nii.ac.jp

† 東京大学工学部 †† 東京大学大学院情報理工学系研究科 ††† 国立情報学研究所

1 はじめに

近年、情報家電に関する研究開発が盛んであるが、インフラに関するものが多い。しかし、インフラのみでは有用な情報家電は実現できず、提供する情報こそが重要である。このような立場から、筆者らは料理支援のために提供する情報に関する研究を行ってきた[1, 2]。本稿ではその要素技術として、料理番組の教材テキストから料理に関する特徴的な素材や調理動作に関する知識を抽出する手法を提案する。実应用到に耐え得る知識を得るためには、料理に特化した抽出手法の適用が不可欠である。そこで本稿では料理の特徴となる「素材」に注目する。まず、料理名・素材の一覧・それに対する調理方法が書かれている多量のレシピテキストを、料理群に分類する。次に、分類された各料理群において特徴的な素材・調理動作を抽出する。更に、その料理群において特徴的でない素材に対してその代替可能性を考察する。

2 概要

本稿で知識抽出の対象とする料理教材テキストのレシピの例を図1に示す。

料理名: 肉じゃが	
素材一覧	調理手順
素材 分量	1 赤こんにゃくは食べやすい長さ(15cm位)に切ります
赤こんにゃく 200g	2 ジャガイモは洗って皮をむき水につけます(大きいものは食べやすい大きさに切る)
じゃがいも 400g	3 玉ねぎは皮をむき縦半分に切り、繊維に沿って約1cm幅に切ります
たまねぎ 1個	
...	
にんじん 1本	4 みりん・醤油を加え、弱火の中火で約15分煮て出汁上がりです

図1: レシピの例

このように、レシピは料理名、素材一覧、調理手順から構成されている。

本稿で提案する知識抽出手法の流れを図2に示す。まずレシピのサンプルを集め、それらを料理名に基づいて料理群に分類する。分類された各料理群において、レシピの素材の一覧に注目し、使われている素材から

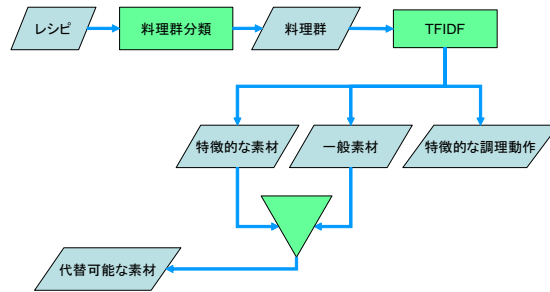


図2: 提案手法

料理群を特徴づける情報を得る。さらに調理手順部分に注目し、特徴的な素材に対する調理動作を抽出する。

次に料理群を特徴付けない一般の素材に注目し、その料理群に含まれる別の一般の素材との比較を行なう。

これらにより、データベースにない新たなレシピの生成といった、調理支援に有用な料理に関する知識を得ることを目指す。

3 料理名による料理群への分類

本稿では、料理名の末尾の名詞列が料理の種類を表すと仮定して、料理群に分類することにした。無作為に選んだ118件の料理名を日本語形態素解析システムJUMAN[4]を用いて品詞分解し、人手で分析した。その結果、91%の料理名においてこの仮定が適切であることを確認した。

また、上記の分類を行なった後、一定数以上の料理が属し、かつ一定長以上の料理名をもつ料理群を有力であると仮定し、それらの料理群名を含む全料理名を再分類した。この際、料理名の末尾のみでなく、先頭の名詞列も照合の対象にした。これにより、分類に失敗したカタカナ語が連続したもの¹や、料理群名が先頭にきたもの²をより大きな料理群に吸収した。

4 料理群を特徴付ける素材と動作の抽出

各料理群を特徴付ける素材を得るために、料理群中の全レシピの素材一覧中に現れる素材に対してTFIDFを適用した。その際、TFはある素材が対象の料理群に属するレシピ中で使用される頻度とした。また特殊な調理法がその料理群を代表してしまうことを防いだ

¹ 「トマトスパゲティ」など

² 「ニョッキのセージバターソース」など

* "Extraction of knowledge on ingredients and cooking steps from cookbooks"

Yutaka Karasawa, Reiko Hamada, Ichiro Ide, Shuichi Sakai, Hidehiko Tanaka
Department of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

め、IDF は料理群を問わず全てのレシピに関して計算した。

$$TFIDF_{d,t} = TF_{d,t} \cdot IDF_t$$

$$TF_{d,t} = \frac{FREQ_{d,t}}{N_d}$$

$$IDF_t = \left(1 + \log \frac{N}{DFREQ_t}\right)$$

N : 全レシピの数
 N_d : 料理群 d に属する料理の数
 $FREQ_{d,t}$: 素材 t の料理群 d における出現頻度
 $DFREQ_t$: 素材 t の全レシピにおける出現頻度

さらに料理群を特徴付ける調理動作を発見するために調理動作に関する情報についても同様に TFIDF を適用した。

例えば、2,057 件のレシピに適用したところ、素材「小麦粉」・「パン粉」・「卵」を他の材料と共に「揚げる」という調理動作を行なうことが料理群「コロッケ」の特徴であり、また、素材「スパゲティ」・「オリーブ油」・「にんにく」を「茹でる」という調理動作が料理群「スパゲティ」を特徴付けていることなどが得られた。

5 代替可能な素材の発見

TFIDF により、特徴的でない³と判定されたものの、その料理群において頻繁に使われている素材が存在した。レシピ中の素材には、料理を特徴付けるものとそれ以外の一般のものがある。そこで、料理の多様性は一般の素材の組み合わせ方によってもたらされる、との仮定のもとに、代替可能な素材に関する知識を抽出する。そこで、レシピ群中の頻出一般素材に注目し、それと排他的に使われている素材を抽出した。

4 節と同様に 2,057 件のレシピに適用した結果を表 1 に示す。括弧内は食品成分表 [5] による分類である。同じ分類に属する素材同士は代替可能な食材として良好な結果が得られた。またコロッケにおいて「かぼちゃ」と「じゃが芋」は異なる分類に属しているにもかかわらず妥当な代替可能性が示された。

6 おわりに

本稿では料理テキストからそれぞれの料理に対する素材・調味料・調理動作・素材の代替可能性に関する知識を得る手法を提案し実験した。

本稿で知識抽出に利用したレシピ数が少なかったため、今後より多くのレシピを対象にすることによりさらに多くの代替可能な食材の発見が可能になると期待する。さらに、分類が異なる代替可能な素材候補の妥

表 1: 代替可能素材の候補例

料理群	注目した一般の素材	排他的に使われる素材
蒸し	にわとり (肉類)	ぶた (肉類)
カレー	クミンシード (-)	上白糖 (砂糖類)
コロッケ	じゃが芋 (いも類)	かぼちゃ (野菜類)
スープ	根深ねぎ (野菜類)	たまねぎ (野菜類)
スープ	スープの素 (-)	スープ (-)
サラダ	たまねぎ (野菜類)	純米酒 (し好飲料類)
サラダ	調合油 (油脂類)	オリーブ油 (油脂類)
ソース	レモン汁 (-)	ぶた (肉類)
ごはん	上白糖 (砂糖類)	にわとり (肉類)

当性を判断するために apriori[3] アルゴリズムを用いた共起情報を用いる予定である。また、複数の料理群において同一の代替可能性が示されたならば、普遍的な代替可能性の発見につながる。

このようにして得られる知識をもとに、既存のレシピデータベースと素材から、調理可能なレシピを生成するといった料理支援を考えている。

参考文献

- [1] 浜田 玲子, 井手 一郎, 坂井 修一, 田中 英彦 “料理テキスト教材における調理手順の構造化”, 電子情報通信学会論文誌, Vol. J85-D-II, No. 1, pp.79-89, 2002.
- [2] 浜田 玲子, “索引付けのための映像とテキスト教材の対応付けシステム”, 博士論文, 東京大学大学院 工学系研究科, 2003.
- [3] Rakesh Agrawal, R. Srikant: ”Fast Algorithms for Mining Association Rules”, Proc. of the 20th Int'l Conference on Very Large Databases, Santiago, Chile, Sept. 1994. Expanded version available as IBM Research Report RJ9839, June 1994.
- [4] “日本語形態素解析システム JUMAN 3.6”, <http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html>
- [5] 香川 芳子, “五訂食品成分表 2003”, 女子栄養大学出版社, 2003.

³本稿では TFIDF の値が 1 より小さいとした