

料理名のネーミングコンセプト抽出によるレシピ推薦システム

橋 明穂^{†,a}若宮 翔子^{‡,a}難波 英嗣^{‡,b}角谷 和俊^{†,c}

[†] 兵庫県立大学環境人間学部 [‡] 兵庫県立大学大学院環境人間学研究科 ^{‡‡} 広島市立大学情報科学研究科

a) {nc10x087, ne11n002}@stshse.u-hyogo.ac.jp b) nanba@hiroshima-cu.ac.jp c) sumiya@shse.u-hyogo.ac.jp

概要 ユーザ投稿型レシピサイトの料理名には、料理の特徴を表すために多様な修飾表現が用いられている。しかし、料理名だけではレシピの特徴が分かりにくいレシピも存在する。本研究では、料理の典型的な手順や材料などの要素とあるレシピの要素との差異に着目し、料理名のネーミングコンセプトを抽出する。このとき、典型的な要素との差異には料理の特徴であるものとそうでないものが混在しているため、料理名に同一の修飾表現が付与されているレシピ集合において、典型的な要素との差異が類似するレシピから共通的な要素をネーミングコンセプトとして抽出する。また、ネーミングコンセプトを活用したレシピ推薦システムを提案する。

キーワード 情報抽出, 修飾表現, 料理レシピ, ネーミングコンセプト

1 はじめに

現在、楽天レシピ¹のようなレシピサイトには大量のレシピが投稿されている。料理名には各レシピの特徴を表すために様々な修飾表現が用いられており、同じ修飾表現でもレシピによってその根拠は多様である。そのため、レシピを検索して提示される料理名だけではレシピの特徴が分かりにくいという問題がある。先行研究 [1] において、我々は修飾表現の付与傾向をネーミングコンセプトと定義し、材料の差異によって抽出可能であることを示した。本稿では、典型的な要素と比較することで材料と手順の差異を抽出し、同じ修飾表現のレシピ間で差異を比較することで、料理名における修飾表現を特徴付ける要素を抽出し、それをネーミングコンセプトとしてレシピとともに提示するシステムを提案する。

提案システムの概要を図1に示す。ユーザが任意の料理名を入力してレシピを検索すると、結果一覧にネーミングコンセプトを付与して出力する。本研究では、ある料理において対象レシピと典型的な要素との差異となる要素がそのレシピの特徴を表すと考える。例えば、「簡単」ケーキ」というレシピの材料において、ホットケーキミックスが典型的な材料にはない差異として抽出された場合、それが「簡単」という修飾表現を表す根拠となると考えられる。さらに、同じ修飾表現を使用する類似レシピ間の、詳細化、簡略化、追加、削除、代替、見だ目の変化など、差異の役割を見つけることでネーミングコンセプトを抽出する。

修飾表現に着目した研究として、高橋ら [2] はオブジェクトの内容について書かれた部分から、修飾表現と適合する語と相反する語を抽出することで、Webテキストと修飾表現との適合度を判定する手法を提案した。本稿

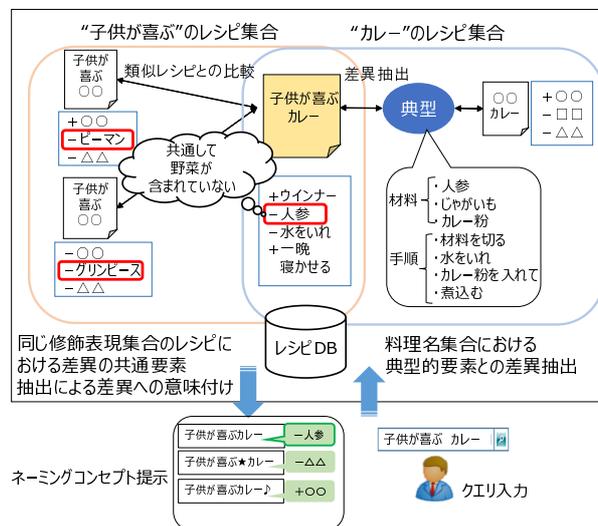


図1 ネーミングコンセプト抽出によるレシピ推薦

では、典型的な要素との差異を抽出し、同じ修飾表現の共通の特徴を分析することで、なぜその修飾表現が使用されているのかを提示することを目的としている点で異なる。佃ら [3] は、認知心理学における典型性の観点に着目し、分析対象と典型性のタイプという二つの軸から典型性の観点の整理を行った。本研究はレシピの差異を特徴として抽出するために典型的な要素を求めている点で特徴的である。

2 ネーミングコンセプト抽出手法

2.1 料理の典型的要素の定義

本研究で対象とするレシピ r_i は、属する料理カテゴリ c_i 、料理名に含まれる修飾表現の集合 E_i 、材料を表す語集合 M_i 、手順を表すリスト P_i の四つ組で表され

ものとし、以下のように定義する.

$$r_i = (c_i, E_i, M_i, P_i) \quad (1 \leq i \leq n), r_i \in R, c_i \in C$$

$$E_i \subset E, M_i \subset M, P_i \subset P$$

$$E = \{e_1, e_2, \dots\}, M = \{m_1, m_2, \dots\}, P = \{p_1, p_2, \dots\}$$

$$P_i = \langle p_{i1}, p_{i2}, \dots, p_{in} \rangle, p_k = (V_k, N_k, A_k) \quad (1 \leq k \leq n)$$

ここで、 R はレシピの全体集合、 C はカテゴリの全体集合、 E は修飾表現の全体集合で e_k は各修飾表現、 M は材料の全体集合で m_k は各材料とし、 P は全ての手順リストで一つ以上のステップ p_k から構成されているものとする。 p_k は、動詞集合 V 、名詞集合 N 、修飾語集合 A の三つ組からなる。

次に、ある料理の複数レシピを要約することにより典型的要素を抽出する。典型的要素 t_j は、属するカテゴリ c_j 、典型的な材料を表す語集合 M'_j 、典型的な手順を表すリスト P'_j の三つ組で表されるものとする。

$$t_j = (c_j, M'_j, P'_j) \quad (1 \leq j \leq m), t_j \in T, c_j \in C$$

$$M'_j \subset M, P'_j \subset P, M' = \{M_1, M_2, \dots\},$$

$$P'_j = \langle S'_1, S'_2, \dots \rangle, S' = \{p_1, p_2, \dots\}$$

ここで、 M' は典型的な材料の全体集合で、 P'_j は典型的な手順リストで S'_i は各ステップ。 S' は全ての典型的なステップ集合とする。

2.2 料理の典型的要素との比較による差異抽出

典型的要素との差異には、修飾表現のネーミングコンセプトを表す差異が含まれていると考えられる。そのため、あるレシピ r_i が属する料理カテゴリ c_i の典型的要素 t_j ($r_i.c_i = t_j.c_j$) を取得し、差異を抽出する。今回は、レシピ r_i で追加された材料 ($M_i - M'_j$)、削減された材料 ($M'_j - M_i$)、追加されたステップ ($P_i - P'_j$)、削減されたステップ ($P'_j - P_i$) を差異として抽出する。例えば、カレーの典型的な材料集合 M'_{curry} が {人参, じゃがいも, カレー粉}, 典型的な手順集合リスト P'_{curry} が <“材料を切る”, “水をいれ”, “カレー粉を入れて”, “煮込む”> であるとき, “子供が喜ぶ”カレーのレシピと比較すると、追加された材料として {もち}, 含まれていない材料として {人参}, 追加されたステップとして {一晩寝かせる}, 含まれていないステップとして {水をいれ} が抽出される。

2.3 同一修飾表現のレシピの差異による特徴抽出

同一の修飾表現が付与されたレシピであれば、料理が異なる場合でも、それぞれの典型的要素との差異にはいくつかの共通する特徴が存在すると思われる。しかし、料理が異なれば、材料や手順なども異なり、共通する特徴の判定が難しいため、レシピごとに差異を定量的に表現したベクトルを生成し、差異に基づくレシピ間の類似

度を算出する。今回は、材料の非網羅度 a_1 と不正確度 a_2 、手順の非網羅度 a_3 と不正確度 a_4 、順序不一致度 a_5 の五つの属性を持つ差異ベクトルを定義する。

$$Comp_a_1(r_i, t_j) = 1 - \frac{|M_i \cap M'_j|}{|M'_j|}, Comp_a_2(r_i, t_j) = 1 - \frac{|M_i \cap M'_j|}{|M_i|}$$

$$Comp_a_3(r_i, t_j) = 1 - \frac{|P_i \cap P'_j|}{|P'_j|}, Comp_a_4(r_i, t_j) = 1 - \frac{|P_i \cap P'_j|}{|P_i|}$$

$$Comp_a_5(r_i, t_j) = 1 - \left(1 - \frac{\sum D^2}{N^3 - N} \times \frac{N}{max(P_i, P'_j)} \times \frac{1}{2}\right)$$

ここで、 $Comp_a_5$ は、スピアマンの順位相関係数に基づき順序の不一致度を算出する関数であり、 D は対応する P_i と P'_j の手順数の差、 N は対応する手順ペア数を表す変数とする。また、他のベクトル特徴量と同様に値を 0 から 1 までの範囲に調整するための重みを付与している。

次に、差異ベクトル間の類似度が高いレシピ集合において、共通的な差異の特徴を抽出する。例えば、“子供が喜ぶ”レシピ集合において、差異の中で典型的な材料に含まれる野菜カテゴリに属する材料を含んでいないレシピが多いと判定した場合、それを共通の特徴として判定し、子供が苦手な材料を含まないため“子供が喜ぶ”という観点で一つのネーミングコンセプトを抽出する。

3 おわりに

本研究では、レシピの料理名の修飾表現を付与した根拠となるネーミングコンセプトを、料理の典型的要素との差異に基づき抽出し、レシピを推薦するためのシステムを提案した。特に、あるレシピ要素と典型的要素には多様な差異があるが、材料と手順に着目し、料理に関わらずある修飾表現に共通する差異に基づきネーミングコンセプトを抽出する手法を提案した。今後の課題として、材料が記述されている順や典型的な手順と対応するステップの分布を考慮した差異分析や特徴抽出について検討する。

謝辞

本研究の一部は、平成 24 年度特別研究員奨励費 (24.9154) によるものである。また、本研究を遂行するにあたり、解析対象となるレシピデータを楽天技術研究所よりご提供いただいた。ここに記して謹んで感謝の意を表する。

参考文献

- [1] 橘 明穂, 若宮 翔子, 角谷 和俊: レシピサイトにおける料理名の修飾表現に着目したネーミングコンセプト, DEIM Forum 2013, P3-5, 2013.
- [2] 高橋良平, 小山聡, 大島裕明, 田中克己: Web テキストと修飾表現との適合度判定手法, DEIM Forum 2010, C3-3, 2010.
- [3] 佃 洗撰, 中村 聡史, 山本 岳洋, 田中 克己: オブジェクトの典型度分析とその検索への応用, WebDB Forum 2011, 2G-1-2, 2011.