

オノマトペロリ：オノマトペを利用した料理の印象検索

カンウィパー ラートサムルアイパン (指導教員：渡辺知恵美)

1 はじめに

Web 検索を行う際に検索対象のイメージはあっても、うまく言語化できないと期待する結果にたどり着けない。ところで、日本語ではオノマトペ(擬音語、擬態語)を用いて、「ふかふかの枕」「シャキシャキのサラダ」のようにイメージをよく表現する。そこでオノマトペを利用し、イメージに基づいた検索を実現することを目指す。まずは、ユーザのレシピ検索を支援するシステム「オノマトペロリ」を提案する。従来のレシピ検索でもオノマトペでレシピを検索できるが、オノマトペを含まないレシピは検索対象にならない。そこで、オノマトペロリではあらかじめオノマトペとレシピに含まれる食材や料理名などの用語の関係を数値化しておく。そして、その関係を用いてオノマトペ自体を含まないレシピでも検索対象とし、さらに求めた関係を用いて検索結果のレシピのランキングを行う。

2 オノマトペロリの基本方針

2.1 オノマトペとレシピ用語の関連度

初めに、オノマトペと料理に関する用語の関連度を求める。具体的には、投稿型レシピサイト COOKPAD[1]に掲載されているレシピの文章を形態素解析・構文解析し、オノマトペとレシピに関する用語を食材、料理名、調理器具、調理法、味・食感の5つのジャンルに分類する。分類例を表1に示す。

ジャンル	例
オノマトペ	ピリッ, こんがり
食材	七味唐辛子, 油, なす
料理名	炒め
器具	フライパン
調理法	切る, 煮詰める
味, 食感	辛い, 柔らかい

表 1: 料理レシピの重要な要素

次にレシピから解析できたオノマトペとレシピ用語の関連度を数値化する。オノマトペと関連が高い用語の特徴は以下の2点が挙げられる。

- そのオノマトペと共に同じレシピに出現することが多い。
- どのようなレシピにも出現している一般性の高い用語ではない。

上記を踏まえて、以下の計算式で表されるリフト値を用いて、関連度とすることとした。

$$lift(X, Y) = \frac{P(X, Y)}{P(X)P(Y)} \quad (1)$$

リフト値は条件 X (オノマトペ) に対する結論 Y (レシピ用語) の関連度を示す。リフト値は、オノマトペ

とレシピ用語の共起度が高いほど高く、一般性が高いオノマトペ及びレシピ用語を用いた場合は低くなり、実情に即した結果を得られやすい。さらに、本研究ではリフト値の基準値を経験的に 0.008 とし、基準値以上の組合せのみレシピ検索に利用する。

以上の手法に基づいて求められたオノマトペ「とろとろ」に関する相関関係とその関連度の結果例を図1に示す。

関連		関連度 (リフト値)
オノマトペ	レシピ用語	
とろとろ	{卵}	0.0083
とろとろ	{牛乳}	0.0094
とろとろ	{片栗粉}	0.0110
とろとろ	{大根, 煮物}	0.0966
とろとろ	{白菜, シチュー}	0.0673
とろとろ	{片栗粉, 溶く}	0.0250
とろとろ	{スープ, 鶏, 豚}	0.0560

図 1: 「とろとろ」に関する相関関係とその関連度

2.2 関連とその関連度を用いたレシピランキング手法

「オノマトペ 食材/料理名」のキーワードに対して、オノマトペと関連の高い順でレシピを推薦するために、まず、キーワードの食材または料理名を含んだレシピを絞り込む。そして、絞り込んだ各レシピの文章について、キーワードのオノマトペに対する関連が何個含まれているかを算出し、含まれていた全関連の関連度を加算し、結果をそのレシピのスコアとする。このスコアを用いてレシピのランキングを行うことができる。

3 オノマトペロリのプロトタイプシステム

提案した手法に基づき、オノマトペロリのプロトタイプシステムを実装した。図2ではオノマトペロリのトップページを示す。なお、プロトタイプシステムは、PHPとMySQLを用いて実装した。また、データセットについては COOKPAD から取得した2万件のレシピデータをベースとしている。

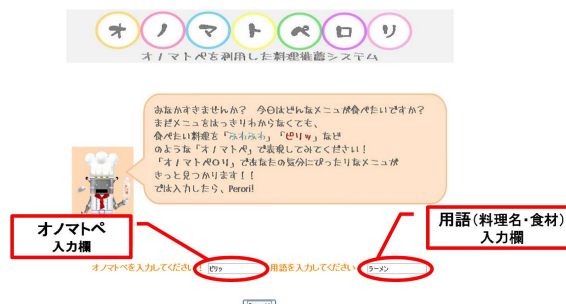


図 2: オノマトペロリのトップページ

オノマトペロリでは、トップページから探している

レシピの料理名や食材に加え，オノマトペをキーワード入力して検索を行える．レシピ検索結果の表示ページは図3の通りである．結果ページでは，キーワードのオノマトペに対して，関連の高い順で表示される．ユーザは検索結果の中から目的とするレシピを選択することにより，COOKPAD 上の詳細なレシピを閲覧することが可能である．



図 3: オノマトペロリの結果ページ

4 従来のレシピ検索と比較実験

従来のシステムとオノマトペロリの比較分析を行う．従来のシステムとは，料理名のみを検索キーワードとした場合であり，料理を「ラーメン」として COOKPAD で検索する「オノマトペロリ」の検索システムは，料理を「ラーメン」，オノマトペを「さっぱり」「ピリッ」「とろとろ」「シャキシャキ」の4つとする．両システムにより求めた検索結果の上位 10 件のレシピについて，ユーザ評価により，各オノマトペと適合するかを調べた．適合の判断は，4人中3人以上が適合すると評価したものとす．その結果を図4に示す．

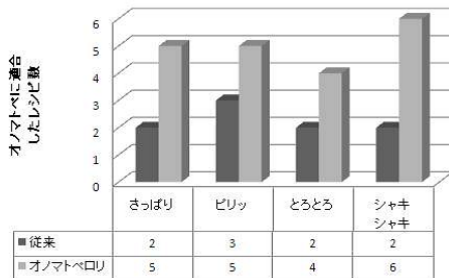


図 4: 従来のレシピ検索とオノマトペロリの結果の適合率の比較

明らかに，「オノマトペロリ」のほうが各オノマトペに適合するレシピ数が多いことがわかるので，よりユーザが食べたいレシピを上位に推薦できていることがわかる．

さらに，適合するレシピについて，オノマトペが含まれているかを調べ，その結果を図5に示す．ユーザによりオノマトペに適合すると評価されたレシピの大多数が，オノマトペを含んでいないことがわかる．つまり「オノマトペロリ」では，オノマトペをキーワードとして含んでいないレシピも，オノマトペで検索可能であると言える．

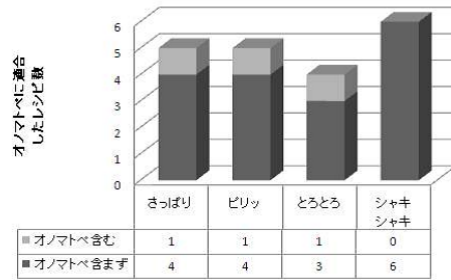


図 5: オノマトペロリの結果におけるオノマトペが含まれるレシピの割合

5 考察

プロトタイプシステムは，オノマトペに基づくインデックスを保持しておらず，毎回計算を行っているため，多くの時間がかかってしまう．これは，あらかじめ関連を計算しておき，データとして保持しておくことにより高速化が可能であると考えている．

現在は「パリパリ」と「パリッ」のように類似しているオノマトペも別のものとして扱っている．そのため，類似しているオノマトペでは多くのレシピがあるものの，ユーザが指定したオノマトペでは十分なレシピを得ることができず，十分な関連を作れないため，ランキングに問題のあることがある．小松ら [2] は，オノマトペの音からオノマトペ間の類似性を計算する手法を提案している．我々の研究でもこうした手法を応用し，類似するオノマトペをまとめることによって十分な件数を取得できるようになると考えられる．また，現在では「しそ」と「大葉」のように類似している用語や「れんこん」と「蓮根」のような表記ゆれの用語も全く別のものとして扱っている．これに関しても用語の類似及び表記ゆれを辞書ベースのマッチングにより吸収することで，ランキングの精度を向上させることが可能であると考えられる．

6 まとめと今後の課題

本論文では「オノマトペロリ」がオノマトペを含んでいないレシピでもオノマトペをキーワードとして検索することが可能であることを明らかにした．また，結果レシピに対するランキング手法を提案して実験を行った．一般的ではないオノマトペと用語の組合せで検索することによって，ユーザが新鮮なレシピを発見することができるという使い方を見つけることができた．しかし，現段階では精度に関して，まだ改良の必要性がある．そのため今後の課題として，オノマトペの類似や用語の類似，表記ゆれを考慮し，対象レシピ及びランキングに利用する関連を増やすことによって，ランキングの精度を改善する．また，本システムのプロトタイプの動作を高速化し，システムの精度を向上させていきたい

参考文献

- [1] 日本最大の料理サイト「COOKPAD」:
<http://cookpad.com>
- [2] 小松孝徳, 秋山広美, ユーザの直感的表現を支援するオノマトペ表現システム, 電子情報通信学会論文誌. A, 基礎・境界 J92-A(11), 752-763, 2009-11-01