

オノマトペに基づく料理レシピ検索における再現率向上の一検討

オノマトペ

松本 敦志[†] 宮森 恒[†]

[†] 京都産業大学コンピュータ理工学部宮森研究室

〒 603-8555 京都府京都市北区上賀茂本山

E-mail: †{g0847173,miya}@cse.kyoto-su.ac.jp

あらまし 「ふわふわ」や「さくさく」といった食感などを表す表現はオノマトペと呼ばれ、近年、漠然とした料理のイメージに基づき効率的にレシピを検索する一手段として研究が行われている。例えば、オノマトペロリシステムでは、クエリとして入力されたオノマトペを関連度の高い食材のキーワードに置き換えて検索する手法や、レシピに含まれる各用語のオノマトペとの関連度合計を計算することで、クエリであるオノマトペと各レシピとのランキングを実現する手法が提案されている。しかし、オノマトペと強く共起する特定の食材の有無がランキングに大きな影響を与えるため、その食材を含まないレシピが検索結果に含まれにくいという問題点があった。本稿では、これらの問題点を解消するため、機械学習に基づくアプローチを検討する。これにより、従来と同等な精度を維持しつつ、再現率を向上させることが期待される。従来システムとの精度比較、および、被験者実験により有効性を検証する。

キーワード 料理レシピ, 検索システム, オノマトペ, 再現率, 機械学習

1 examination of the improvement

Onomatopoeia

Atsushi MATSUMOTO[†] and Hisashi MIYAMORI[†]

[†] Faculty of Engineering, Kyotosangyo University

Kamigamo-Motoyama, Kita, Kyoto, 603-8555 Japan

E-mail: †{g0847173,miya}@cse.kyoto-su.ac.jp

Abstract "lightly" – "snappily" . to which research is done as a way stage which the expression showing the texture etc. which were said is called onomatopoeia and searches a recipe efficiently in recent years based on the image of a vague dish – for example, By calculating the degree-of-association sum total with the onomatopoeia of the technique of transposing the onomatopoeia inputted as a query to the keyword of foods with a high degree of association, and searching it with an onomatopoeia stem, and each term contained in a recipe, The technique of realizing ranking of the onomatopoeia and each recipe which are queries is proposed. However, since the existence of the specific foods which coincide with onomatopoeia strongly had big influence on ranking, there was a problem that the recipe which does not contain the foods was hard to be contained in search results.

Key words cooking recipe, search system, onomatopoeia, recall, machine learning

1. はじめに

各家庭で料理を行う際にレシピサイトを利用してどの料理を作るか考える材料として利用されるようになってきた。日々の生活の中で、献立を考えるとという大きな手間を省くことができる。例えば、代表的なレシピ検索サイトである「COOKPAD」の場合、2011年12月現在で112万件以上のレシピが登録されており、「キャベツ」で検索すると約6万件の結果が返される。この中からユーザが

このようなシステムを利用する際にイメージとしてはあるがキーワードをうまく言語化できずに検索できなかった経験はないだろうか。

そのような場合オノマトペを利用した検索を行うことが多い。オノマトペとは、擬音語、擬態語の総称であり味に関する表現を相手に伝えるには効果的である。たとえば「さくさくクッキー」や「ふわふわオムレツ」「ぱりぱりキャベツ」などである。料理検索サービスの検索窓でそのときの気分で食べたいものをオノマトペで検索することで料理レシピを格段に検索しやすくしてくれる。

しかし、従来の料理レシピ検索はオノマトペ検索を行う場合問題点

が存在した。ユーザが検索窓に「さくさく」というキーワードをクエリとして検索した場合、「タイトルにキーワードが含まれている」または「料理レシピの記述にキーワードが含まれる」の2パターン「さくさく」クエリの結果が返ってくる。

ここではタイトルや料理レシピ内の記述にオノマトペが含まれない場合、検索結果に現れないといった問題点がある。料理レシピの記述にオノマトペが含まれなくても使われる食材や調理手順からオノマトペを決定することが可能であると考えられる。

これらの問題点を解消するため、機械学習に基づくアプローチを検討する。これにより、検索結果に表れることのない料理レシピが機械学習によってより多くをカバーすることができると期待される。

2. 関連研究

近年の研究でオノマトペを利用した検索システムが増えつつある。その一例としてオノマトペロリシステムがある。オノマトペロリシステムでは、クエリとして入力されたオノマトペを関連度の高い食材のキーワードに置き換えて検索する手法や、レシピに含まれる各用語のオノマトペとの関連度合計を計算することで、クエリであるオノマトペと各レシピとのランキングを再現する手法が提案されている。

例えば図1のように「さくさく」で検索した場合、料理レシピの記述に「さくさく」を含まず、さらに「さくさく」の食感である料理レシピの検索を可能としている。

レンジでクッキー☆スタバの味☆

レシピID: 516073



電子レンジで数分！スタバのスコーンの味が再現できます♪話題入り感謝です！ID:91063を参考にさせていただきました。

mari-boo

材料 (30枚)

ホットケーキミックス	100g
薄力粉	60g
マーガリン	80g
砂糖	30g
卵	1ヶ
お好みのチョコレート	適宜

図1 レシピ画像

このオノマトペロリシステムは、オノマトペと強く共起する特定の食材の有無がランキングに大きな影響を与えるため、その食材を含まないレシピが検索結果に表れにくいといった問題点が存在する。例えば、「さくさく」に強く共起する食材は「薄力粉」または「小麦粉」などが多い。これらの食材が含まれているまたは、強く共起する食材を多く含んでいた場合、食感がさくさくと判定される。

ここでは本来なら「さくさく」であるレシピがオノマトペと強く共起する特定の食材を含まないために「さくさく」と判定されない場合が存在する。そういった今まで検索結果に現れて来なかったレシピを機械学習による手法で解決していく。

本稿では、以下の構成で述べる。2節で関連研究について述べる。3節で機械学習による提案手法を説明する。4節で提案手法を用いて評価実験を行い結果を述べる。5節で問題点・課題点について述べる。

3. 提案手法

本節では提案手法である機械学習における手順を説明していく。その前処理として Cookpad の HTML データ約 6 万件を取得し、料

理に関する用語を素性として識別器を作成する。作成された識別器を基に実験を行なっていく。

3.1 レシピの構造

まず Cookpad に掲載されているレシピの構造に着目した。Cookpad に掲載されているレシピ構造は大きくわけて4つに分割される。図2に示されるようにタイトル、説明、材料そして調理法のパーツに分けられる。

それぞれの料理レシピデータの HTML ファイルを解析し、料理に

図2 レシピの構造

関する用語を利用する。素性として利用するデータは4つのレシピ構造の内、食材を使用する。

3.2 機械学習

ある特定の食感ごとに振り分ける必要がある。そこでこの問題を解決するために、種々の素性を用いた教師あり機械学習法を用いた。機械学習法として SVM 法を用いた。この SVM は、入力された情報の特徴を用いて、正しく分類することができる。

今回の実験では SVM-Light を使用する。

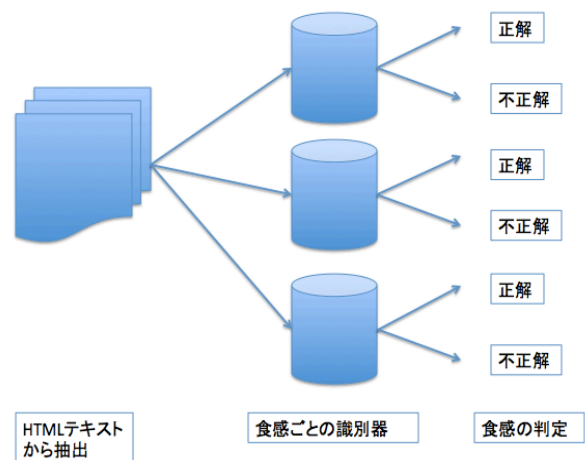


図3 提案手法

3.3 食感ごとの振り分け

特定の食感ごとに識別器を作成するためある基準をもって振り分けた。

まず、どのように人は特定の食感を決めているかを考えた。Cookpadには図4のように画像や材料、調理動作などが記載されている。これらの情報を総合して各々のレシピの食感を決定していると思われる。この記載されている情報から以下の3つの基準から特定の食感を決定していく。

- 見た目
- 使われる材料
- 調理動作

見た目、使われる材料、調理動作で、これら3つの要素から一番妥当である食感を判断して振り分ける。

ここで筆者の基準だけ食感を決めるのではなく、アンケートをとり最終的な食感を振り分けていく。アンケート事項には3つの基準を満たしたものをその食感として振り分けてもらった。

混ぜるだけ、スコーン。 レシピID:1432495

スコーンが好きで、いろいろ試しましたが、いちばん簡単な作り方を。 🍪 風音220

材料 (約4-5cm×20個)

■ A	
生クリーム	300cc
牛乳	250cc
卵	1個
■ B	
小麦粉	700g
砂糖、塩	各小さじ1杯
ベーキングパウダー	大さじ3杯

1 Aをすべてボウルに入れ、よく混ぜ合わせる。

2 BにAを注ぎ、手でさっくりと混ぜ合わせる。

3 生地を伸ばし、型で抜いていく。

4 200°Cのオーブンで、約20分ほど焼く。

見たい目

使われる材料

調理動作

図4 食感の基準

3.4 食感ごとの識別器作成

特定の食感ごとに機械学習を行う。例えば、さくさく識別器、ふわふわ識別器のように特定の食感ごとに機械学習によって分類器を作成していく。特定の食感とはさくさくやふわふわ、シャキシャキのように各々のオノマトペを表す食感のことをここではいうとする。

素性として与えるデータは食材を用いるが、Cookpadのような投稿型料理レシピサイトでは投稿者によって表記方法が異なったものが存在する。例えば、図2のように食材名の前に記号をつけ食材名を目立たせる記述がされているものも存在する。これらの表記ゆれは学習結果に影響を及ぼす可能性があるため食材ごとに食材名を統一する必要がある。

4. 評価実験

機械学習によって作成された識別器を基に実験を行なっていく。特定の食感ごとに分類器を作成していくが今回の実験ではさくさくに關する2値識別器を作成した。

4.1 実験データ

学習における正例データと負例データをそれぞれ300件ずつ用意して学習を行う。テストデータは100件用意し実験を行う。ただし、負例データは明らかに不正解であるものだけでなく、正解になりそうな間違えやすいデータを負例データとして学習データに与える。

4.2 評価方法

評価方法は、適合率 (precision) と再現率 (recall)、F 値で評価を行う。適合率とは、全検索結果に対しての、検索要求を満たす検索結果の割合である。再現率とは、検索要求を満たす全ドキュメントに対しての、検索要求を満たす検索結果の割合である。

特定の食感ごとに実験を行うため、4つの食感の識別器を作成する。今回実験で作成した識別器は”ピリッ”と”さくさく”、”ふわふわ”、”シャキシャキ”を用意した。識別器を食感ごとの結果の違いを見るため異なる食感の識別器を使用した。

4.3 結果

実験結果を表に示す。本実験では、”ピリッ”、”さくさく”、”ふわふわ”、”シャキシャキ”の4つの識別器から評価方法に則り適合率、再現率、F 値を算出する。

機械的にテストを行い結果からその判定は正解であったか、不正解であったかを確認していく。そのときの正解、不正解かはアンケートを基にする。

表1 ピリッ

	再現率	適合率	F 値
提案手法	0.93	0.85	0.887

表2 さくさく

	再現率	適合率	F 値
提案手法	0.56	0.47	0.516

表3 ふわふわ

	再現率	適合率	F 値
提案手法	0.51	0.41	0.463

表4 シャキシャキ

	再現率	適合率	F 値
提案手法	0.68	0.59	0.634

それぞれの表の結果から高い再現率、適合率を出したのは食感が”ピリッ”の識別器である。”さくさく”や”ふわふわ”は精度が悪く、同じような結果となった。これらの結果から考察を行う。

4.4 考 察

4つの食感識別器を用いて実験を行い提案手法の検証を行なった。今回の検証で用いた素性は食材名で識別器を作成した。実験の結果、食感ごとに再現率の高さに大きく異なる結果となった。

今回の実験で良い結果となったのは”ピリッ”とした食感レシピである。結果から高いスコアを出すことに成功した。食材名に香辛料や辛いものに関連した食材を含むものは料理手順や調理方法で他のオノマトペに変化はほとんどないため高い結果となった。しかし、料理レシピに食材名が非常に多く辛いものに関連した食材が1品であった場合検索結果に表れないものも存在した。これは学習数が少なく判断不可能なものだったため不正解としたものとする。

特に目立った結果は”さくさく”と”ふわふわ”の例である。これらの食感は使われる食材名が非常に似ているため各々の識別器を作成した場合、それぞれの食感が含まれてしまう結果となった。例えば、さくさく、ふわふわで主に利用される食材名は薄力粉、卵、砂糖、牛乳の記載が多い。これらから食感に関する決定要因は食材名だけでなく調理の方法によって大きく左右されることがわかる。

これらのことから食感ごとに識別器の問題点が存在した。今回の実験では識別器1つで食材名を素性として結果を得た。人間の食感に関する判断は食材名で判断されることはなく、使われる食材名、調理手順によって大きく左右される。今回の結果から調理動作を考慮した機械学習方法を試すことでより精度の高い識別を可能にすると予想される。

5. 問題点・課題点

料理レシピの解析結果で得られた食材名の表記ゆれを解消することができていない部分が存在する。投稿型レシピサイト Cookpad はユーザが自由に料理レシピを投稿することができ、食材表記方法も様々存在する。そのため表記ゆれを解消しなければ学習結果に大きな影響がある。辞書作成を行い表記ゆれをなくし、多く網羅したものを実装段階である

今回の実験では機械学習における素性は食材を使用した、料理レシピの食感を決めるオノマトペは食材のみで判断できるとは限らない。ふわふわ分類器を作成し実験を行なった際、食材に粉物が存在していたためにレシピの食感判定に不正解となるものが存在した。これは食材のみの素性で学習を行なったため判断材料が不足していたまた、使われている食材が似通っているために起こる結果である。そのため調理動作を考慮した機械学習方法の検討も必要である。他にも識別器を複数利用して分類するその実験を現在行っており、さらなる精度向上を目指す。

6. おわりに

本稿では、オノマトペに基づく料理レシピ検索における再現率向上の一検討を行なった。各々の食感から識別器を作成して、再現率、適合率の結果から考察を行なった。現段階では食材を素性として識別器を作成したが、各々の食感に関する決定は見た目や調理法など総合的に判断されるため識別精度の低さが目立った。そのため調理動作を考慮した学習を行うことでより精度の高い識別器の作成ができると期待できる。今後の研究につなげていく予定である。

文 献

- [1] 日本最大の料理サイト「COOKPAD」<http://cookpad.com>
- [2] SVMlight Support Vector Machine <http://svmlight.joachims.org/>
- [3] ラートサムルアイパン, カンウイパー, 渡辺知恵美, 中村聡史 オノマトペロリ:オノマトペを利用した料理推薦システム 情報処理学会研究報告 IPSJ SIG Technical Report Vol.2009-DD-73 No.6
- [4] 矢島亜沙美, 小林一郎 個人の状況を考慮した”かんたん”なレシピの推薦, DEIM Forum 2008.
- [5] Mecab: Yet Another Part-of-Speech and Morphological Analyzer 京都大学情報学研究所-日本電信電話株式会社コミュニケーション科学基礎研究所 共同研究ユニットプロジェクト <http://mecab.sourceforge.net/>