

レシピ検索のための文書頻度を用いたレシピの味特徴の抽出

塩谷 祐樹[†] 北山 大輔[†]

[†] 工学院大学情報学部システム数理解学科 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: tj317138@ns.kogakuin.ac.jp, tkitayama@cc.kogakuin.ac.jp

あらまし 近年、料理レシピの検索において、さまざまな料理レシピ投稿サイトが利用されている。それに伴い、投稿サイトには数多くの料理レシピが存在している。しかし、その中からユーザ個人の味の好みに合うようなレシピを検索することは困難である。そこで我々は、すべてのレシピに味特徴ベクトルを付与することで、ユーザ個人の味の好みからレシピを検索することを目指す。本稿では、料理と食材に対して味の感想が述べられている文書頻度を求め、それをその料理と食材に対しての味の値とする。その後レシピに含まれているすべての食材から、料理と食材に対しての値の平均値を、そのレシピの特徴として付与する。また、レシピの味特徴についての正解データを求める。レシピの特徴と正解データの相関関係を求めた結果、平均で 0.631 の正の相関があることを確認した。

キーワード レシピ検索, 味

1 はじめに

近年、料理レシピの検索において、クックパッド¹や楽天レシピ²をはじめとするさまざまな料理レシピ投稿サイトが利用されている。それに伴い、投稿サイトには数多くの料理レシピが存在している。これらの料理レシピサイトでは、料理名や材料名で検索することが可能であり、例えば、ピーマンが好きな人は、「チャーハン ピーマン」のように検索することでピーマンを含むチャーハンを検索することができる。また、逆にピーマンが苦手な人は、「チャーハン (not ピーマン)」のような検索要求により、ピーマンを含まないチャーハンを検索可能である。

材料に対する要求と同様に、料理の味に対する要求も考えられる。例えば、辛いのが苦手な人でも食べられる麻婆豆腐を作りたい場合、「麻婆豆腐 辛いくない」や「麻婆豆腐 甘い」のような検索をすることが考えられる。しかし、クックパッドのような一般的なレシピ検索システムにおいて、これらの検索キーワードでは、「麻婆豆腐」と検索するより、検索結果数が極端に少なくなり、うまく検索されない。また、「麻婆豆腐 辛い 濃い」のように、複数の味を検索キーワードに含めた場合、検索結果数が更に少なくなる。この原因は、検索キーワードで入力した語がレシピの文書内に含まれていないと、そのレシピが検索されないことにある。クックパッドで味を検索キーワードに含めた麻婆豆腐の検索数を表 1 に示す。このことから、一般的なレシピ検索システムではユーザ個人の味の好みに合うようなレシピを検索することは困難である。

そこで我々は、ユーザ個人の味の好みからレシピを検索することを目指す。本研究では、「料理名」と「味特徴の重み」を入力することで、それに適したいくつかのレシピを検索するシステムを構築する。このシステムの概要を図 1 に示す。ユーザが入力した味特徴の重みに類似する味のベクトルを持つ料理を検

表 1 味を検索キーワードに含めた麻婆豆腐の検索結果数

検索キーワード	検索結果数
麻婆豆腐	10,509
麻婆豆腐 辛いくない	447
麻婆豆腐 甘い	99
麻婆豆腐 辛い	1869
麻婆豆腐 辛い 濃い	23

索結果として提示する。ここでの味特徴ベクトルとは、それぞれの味を要素とし、味の強さを値として持つベクトルのことである。これを実現するため、まずすべての食材の味特徴ベクトルを作成する。次に、食材の味特徴ベクトルを用いて、レシピに味特徴ベクトルを付与する。

入力として用いる味特徴の重みは 0 から 10 の範囲で選択させる。味特徴の種類は、生理学的に味覚が認識する感覚である基本味の 5 種類（甘味、酸味、塩味、苦味、旨味）、味の感覚として認識しやすい辛味、味の濃淡（濃い、薄い）の 8 種類を用いる。

本研究で構築するシステムでは、文書内に味が明記されていないレシピも検索対象とすることができる。これにより、検索対象のレシピが増加することが期待できる。また、レシピごとに味のベクトルを与えていることから、複数の味からの検索や、味の引き算のような検索も可能となることが期待できる。

本研究の貢献を以下に示す。

- メニューに対するレビューから、食材の味特徴ベクトルを生成するという新しいアプローチに取り組み、その精度を示す
- レシピから想像される料理の味をクラウドソーシングを用いて作成する

本稿の構成を述べる。2 節では関連研究の紹介を行う。3 節で文書頻度を用いたレシピの味特徴の抽出について述べる。4 節で実験について述べる。5 節で本稿をまとめる。

1 : <https://cookpad.com/>

2 : <https://recipe.rakuten.co.jp/>

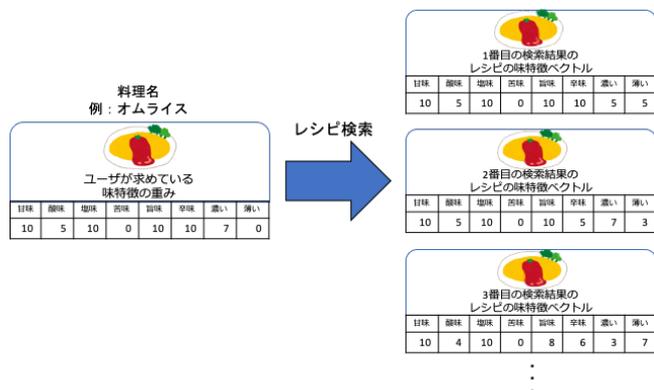


図 1 提案システム概要図

2 関連研究

食材やレシピから味の特徴を取得する研究は数多く行われており、以下に紹介することで本研究の位置づけを示す。

渡辺ら [1] やラートサムルアイパンら [2] は、レシピから味覚と食感を表すオノマトペを抽出する研究を行っている。レシピに含まれる材料、調理法、レビューに含まれる語とオノマトペを共起度を用いて、個々のレシピとオノマトペとの適合度を求めている。長谷川ら [3] は、シズル語を含むレシピサイトのレビュー文から、シズル語に関する因子情報の抽出を行っている。シズル語間の利用頻度の差異を緩和するために BM25 による単語の重み付けを施した。学習結果に対して、ワードクラウドおよびそれらの関係性をネットワークとして表示することで、シズル語の性質に関する情報を可視化した。本研究では、味の特徴をレシピとは直接関係の無い、飲食店レビューから抽出し、食材の味からレシピの味へというボトムアップなアプローチに取り組んでいる点で異なる。

加藤ら [4] は、五感に関連したオノマトペの抽出方法および当該オノマトペを特徴量とした飲食店の推薦方法を提案している。LDA を用いて、飲食店カテゴリとオノマトペの依存関係を探し出している。石橋ら [5] は、米菓を対象としてシズル語の印象評価に関する研究などを行っている。シズル感を伝える際に、しばしばオノマトペが利用されることから、オノマトペに焦点を当てて解析を行っている。刈谷 [6] らは、味覚を印象付ける語と、料理のジャンルの意味的差異に応じて、飲食物の味覚印象検索を行っている。料理のジャンルごとに作成した標準化したものと、味覚の印象を数値化したもので、相関量を計量している。本研究とは、味の特徴の取得方法が文書頻度でない点で、印象ではなく味の特徴をベクトルとして取得している点で異なる。上記の研究における知見をもとに、オノマトペやシズル語と味特徴の関係を明らかにすることで、これらの研究の成果を利用することも可能であると考えている。

3 文書頻度を用いたレシピの味特徴の抽出

本手法の処理の流れを以下に説明する。まず初めに、クックパッドデータから全レシピと全食材を取得する。次に、ぐるなび

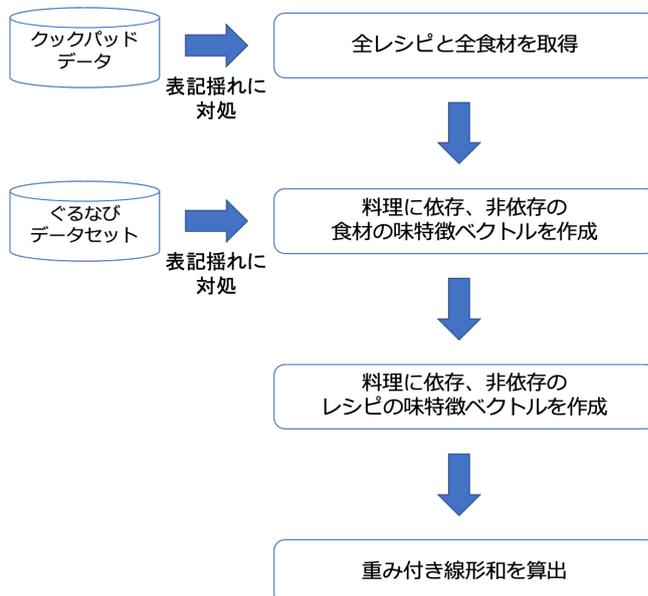


図 2 レシピの味特徴ベクトル作成

び³のデータセットにある「応援フォト」データを用いて、料理に依存、非依存の食材の味特徴ベクトルをレビュー数から作成する。これらの処理では、MeCab やオントロジーを用いた処理を施すことで表記揺れに対処する。次に、料理に依存、非依存のレシピの味特徴ベクトルを、そのレシピに含まれている食材の味特徴ベクトルから作成する。次に、料理に依存、非依存のレシピの味特徴ベクトルから重み付き線形和を算出する。本手法の処理の流れを図 2 に示す。

味の属性ベクトルを A とする (式 1)。式 1 の (甘味, 酸味, 塩味, 苦味, 旨味, 辛味, 濃い, 薄い) は A の要素である。

$$A = \begin{pmatrix} \text{甘味} \\ \text{酸味} \\ \text{塩味} \\ \text{苦味} \\ \text{旨味} \\ \text{辛味} \\ \text{濃い} \\ \text{薄い} \end{pmatrix} \quad (1)$$

3.1 使用データ

本研究では、食材とレシピのデータとして、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」⁴の 2014 年 9 月 30 日以前に公開されたデータを使用している。また、ぐるなび Web Service⁵が提供する「応援口コミ API」を用いて 2016 年 7 月 9 日までに収集したデータを使用している。

3.2 全レシピ名・食材を取得

全レシピ名と食材は、クックパッドデータから取得する。な

3 : <https://www.gnavi.co.jp/>

4 : <https://www.nii.ac.jp/dsc/idr/cookpad/>

5 : <https://api.gnavi.co.jp/api/>

お、「チキンカツカレー」のように、レシピ名に複数の料理名が使われている場合、一番後ろにあるものをそのレシピの料理名として取得する。これらの情報には、レシピ投稿サイトの性質上、レシピの投稿者によって表記揺れが存在する。そのため前処理として、大仁田ら [7] の研究を参考に、品詞絞り込み、オントロジーを用いた処理を施すことで表記揺れに対処する。各手法について、それぞれ以下のような処理を施している。

- 品詞絞り込み：MeCab により、単語を名詞に絞り込む処理
- オントロジー：品詞絞り込みにより取得した名詞に対してオントロジーを用い、料理名や食材名のみを絞り込む処理

オントロジーでは、大仁田らの研究 [8] と同じく、クックパッドデータの材料リストと、料理オントロジー構築プロジェクトで公開されている土居ら [9] の研究を基にした料理オントロジー、京都大学の黒橋・河原研究室より提供されている清丸ら [10] の研究を基にした基本料理知識ベースを用いて、対象単語の絞り込みを行った。また、食材の表記揺れをなくすために、料理オントロジーにおいて上位語が存在する単語に関しては、上位語に置換している。

3.3 食材の味特徴ベクトルの作成

食材の味特徴ベクトルは、ぐるなびデータセットの「応援フォト」から得たレビューを用いて作成する。応援フォトの例を図 3 に示す。応援フォトのレビューはある料理に対して付けられることから、応援フォトのレビューには料理の味が含まれていることがある。また、レビューに食材が含まれていた場合、その料理のその食材の味を述べていると仮定できる。よってそのレビューに食材と味の属性が含まれていれば、その食材はその味である可能性が高いと仮定する。

食材は調理法によって、大きく味が変化するものもある。たとえば玉ねぎは、生のまま食べると辛いですが、熱を通すと甘くなる。このため、料理を考慮せずに食材の味特徴ベクトルを作成すると、意図しない味特徴ベクトルが作成されてしまうと考えた。また、全てにおいて料理を考慮して食材の味特徴ベクトルを作成すると、たとえば「チャーハン+梅」のように、珍しい料理と食材の組み合わせのレシピには対応できないと考えた。以上のことから、料理に依存する食材の味特徴ベクトルと、料理に非依存な食材の味特徴ベクトルの作成を行う。

料理に依存する食材の味特徴ベクトルに対して、重みを与える応援フォトの条件を以下に示す。

- メニューに料理名が含まれている
- レビューに食材が含まれている
- レビューに味の属性が含まれている

たとえば、「当店自慢のオムライス」というメニューのレビューに、「ケチャップの酸味がきいてよく合う」と書かれていた場



図 3 応援フォトの例⁶

合、「料理名：オムライス」「食材名：ケチャップ」「味の属性：酸味」と、重みを与える条件を満たす応援フォトであったため、(オムライス, ケチャップ, 酸味) の特徴に 1 の重みを付与する。

料理 d と味 A に対応した食材 f の味特徴ベクトルを $S_{d,f,A}$ と表す。なお、集計後に極端な数値の偏りを緩和するため対数をとって、 $\log(S_{d,f,A})$ とする。また、料理に非依存な食材の味特徴ベクトルの対数をとったものを $\log(S_{*,f,A})$ とする。ここで、* は料理に依存しないことを示す。

3.4 レシピの味特徴ベクトルの作成

レシピに含まれるそれぞれの食材の、食材の味特徴ベクトル上位 m 個の平均を求めることで、レシピの味特徴ベクトルを算出し、これを味特徴 8 種類で作成する。平均化することで、食材の数が多いレシピの味特徴ベクトルが大きくなってしまいう問題を解決している。

料理に依存するレシピの味特徴ベクトル $V_{d,A}$ を算出する関数を以下のように定義する (式 2)。 $(S_{d,f,A})_j$ は上位 j 番目の食材の味特徴ベクトルであり、レシピの味特徴ベクトル算出には上位 m 個の j の平均をとる。なお、レシピに含まれる食材 f の数が m より小さい場合、 $m = f$ とする。

$$V_{d,A} = \frac{1}{m} \sum_{j=1}^m \log(S_{d,f,A})_j \quad (2)$$

同様に、料理に非依存のレシピの味特徴ベクトル $V_{*,A}$ を算出する関数を以下のように定義する (式 3)。

$$V_{*,A} = \frac{1}{m} \sum_{j=1}^m \log(S_{*,f,A})_j \quad (3)$$

3.5 重み付き線形和の算出

あまりその料理で組み合わせられない食材は、レビュー中に出現しないことも多い。これに対応するため、料理 d に依存するレシピの味 A 特徴ベクトル $V_{d,A}$ と、料理に非依存のレシピの味特徴ベクトル $V_{*,A}$ を得て、それぞれの重み付き線形和 $G_{d,A}$ (式 4) を最終的な食材の味特徴ベクトルとする。重みは α , $(1 - \alpha)$ としている。

6 : <http://mr.gnavi.co.jp/a960500/review/1/2193078/?p=1>

$$G_{d,A} = \alpha V_{d,A} + (1 - \alpha) V_{*,A} \quad (4)$$

4 実 験

4.1 食材の上位数と料理名の依存度合いの決定

正解データを作成し、それと重み付き線形和の相関係数を求め、食材の上位数である式2と式3の m 、料理名の依存度合いである式4の α を決定する。

4.1.1 正解データの作成

正解データを作成する。クラウドソーシングで募集したアンケート結果により作成した。実験協力者には、あるレシピに対して味特徴8種類の重みを0から10の範囲で選択させた。レシピ数は100で、1レシピには30人、合計3000データを募集した。また、実験協力者の料理頻度も同時に調査することで、正解データとして信頼性を確認した結果、各レシピ平均で27データを採用した。回答されたレシピごとの味特徴の数値を平均して、それをそのレシピの正解データ V_{answer} とする(式5)。 $r_{a,k}$ はそのレシピ r 内の k 番目の採用データであり、採用データで入力された味特徴の種類を A としている。

$$V_{answer,A} = \frac{1}{l} \sum_{k=1}^l r_{A,k} \quad (5)$$

4.1.2 相関係数を算出

正解データ V_{answer} と重み付き線形和 G_d の相関係数 R_d を算出する(式6)。味特徴の総数は8で、 $G_{d,A}$ 、 $V_{answer,A}$ はそれぞれの味特徴の A 番目の数値、 $\overline{G_d}$ 、 $\overline{V_{answer}}$ はそれぞれの平均値である。

$$R_d = \frac{\frac{1}{8} \sum_{A=1}^8 (G_{d,A} - \overline{G_d})(V_{answer,A} - \overline{V_{answer}})}{\sqrt{\frac{1}{8} \sum_{A=1}^8 (G_{d,A} - \overline{G_d})^2} \sqrt{\frac{1}{8} \sum_{A=1}^8 (V_{answer,A} - \overline{V_{answer}})^2}} \quad (6)$$

4.1.3 食材の上位数と料理名の依存度合いの決定

食材の上位数である m 、料理名の依存度合いである α を、相関係数の平均が1番高くなるような組み合わせにする。表2はそれぞれの m で相関係数の平均が最大のときの α である。これにより、 $m = 3$ 、 $\alpha = 0.65$ に決定した。また、相関係数の平均が0.631であることから正の相関があると言え、本手法は適切なものであると考える。

表2 それぞれの m で相関係数の平均が最大のときの α

m	相関係数	α
2	0.629	0.62
3	0.631	0.65
4	0.621	0.68
5	0.618	0.71

4.2 考 察

「あったか〜いカレーうどん」というレシピ(図4)と「塩レモン☆梅ゴーヤチャーハン」というレシピを用いて提案手法を



図4 「あったか〜いカレーうどん」のレシピ画像⁷

表3 「あったか〜いカレーうどん」の料理に依存する食材の味特徴ベクトル一覧

食材名	甘	酸	塩	苦	旨	辛	濃	薄
ネギ	1.8	0.7	1.4	0.0	2.4	1.4	2.3	1.1
うどん	3.4	2.4	3.4	2.6	3.7	3.5	4.1	2.6
出し	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0
醤油	0.7	0.0	0.7	0.0	0.0	0.0	0.0	0.0
素	1.1	0.0	0.0	0.0	1.1	0.0	0.0	0.7
鳥肉	1.1	0.0	1.1	0.0	2.0	0.0	1.8	0.0
水	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0
カレー粉	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
片栗粉	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
酒	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
みりん	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
水溶き	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

考察する。表3は料理に依存する食材の味特徴ベクトル一覧である。縦軸が食材、横軸が味特徴である。

表4は、「あったか〜いカレーうどん」の味特徴ベクトル作成の実行例であり、料理に依存するレシピの味特徴ベクトル、料理に非依存なレシピの味特徴ベクトル、重み付き線形和、正解データが記されている。重み付き線形和から、このレシピは甘味、塩味、旨味が高く味が濃い、酸味、苦味が低いと判断できる。正解データからは、このレシピは旨味、辛味が高く味が濃い、酸味、苦味が低く味が薄いと判断できる。また、このレシピの相関係数は0.92となり、強い正の相関があると言える。

表5は、「塩レモン☆梅ゴーヤチャーハン」の味特徴ベクトル作成の実行例であり、料理に依存するレシピの味特徴ベクトル、料理に非依存なレシピの味特徴ベクトル、重み付き線形和、正解データが記されている。重み付き線形和からは、このレシピは塩味、旨味が高く、酸味、苦味が低く味が薄いと判断できる。正解データからは、このレシピは酸味、塩味が高く、甘味、辛味が低いと判断できる。このレシピの相関係数は-0.04となり、ほとんど相関がないと言える。正解データでは、レシピに含まれている「塩」、「レモン」、「梅」、「ゴーヤ」の影響が強く出ており、酸味や塩味が高い結果となったが、重み付き線形和

⁷ : <https://cookpad.com/recipe/6550398>

表4 「あったか〜いカレーうどん」の味特徴ベクトル作成の実行例

	甘	酸	塩	苦	旨	辛	濃	薄
料理に依存する味特徴ベクトル	1.8	1.0	1.9	0.2	2.4	1.5	1.5	0.9
料理に非依存な味特徴ベクトル	4.6	2.9	4.9	3.0	4.9	4.9	4.8	3.0
重み付き線形和	3.0	1.7	3.0	1.6	3.5	2.8	3.5	2.0
正解データ	4.0	2.1	5.1	1.6	7.0	6.5	6.4	2.8

表5 「塩レモン☆梅ゴーヤチャーハン」の味特徴ベクトル作成の実行例

	甘	酸	塩	苦	旨	辛	濃	薄
料理に依存する味特徴ベクトル	0.0	0.0	1.0	0.0	1.1	0.6	0.2	0.2
料理に非依存な味特徴ベクトル	4.5	2.7	4.6	3.4	4.8	4.7	4.9	2.8
重み付き線形和	1.5	1.0	2.3	1.2	2.4	2.0	1.9	1.1
正解データ	2.6	7.9	6.7	5.8	6.0	2.5	5.3	3.5

では、苦味が低い結果になっているなど、この傾向が強く現れていない。このように、そのレシピの味に大きな影響を与える食材が、応援フォトでは出現しにくい食材の場合、相関が現れにくいのが原因であると考えた。

5 おわりに

本研究では、ユーザ個人の味の好みからレシピを検索するため、料理名と味特徴ベクトルを入力することで、それに適したレシピを検索するシステム構築を提案した。具体的には、最初にクックパッドデータから全料理リストと全食材リストを取得した。次に、ぐるなびのデータセットを用いて、食材の味特徴ベクトルをレビュー数から作成した。次に、レシピの味特徴を、そのレシピに含まれているすべての食材の味特徴ベクトルから作成した。そして、料理に依存、非依存を考慮したレシピの味特徴を重み付き線形和として抽出した。また、レシピの味特徴についての正解データを求めた。抽出した重み付き線形和と正解データの相関関係を求めた結果、平均で0.631の正の相関があることを確認した。今後の課題として、作成したレシピ検索システムから、本手法が適切なレシピ検索を行っているかのユーザ評価実験の実施、一般的なレシピ検索システムとの比較実験の実施を目指す。また、文書頻度によって料理名ごとの依存度合いを変更したり、食材の分量を考慮することで、より精度の高いシステムの開発を目指す。

6 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」を利用しました。また、本研究の一部は、2020年度科研費基盤研究(C)(課題番号:18K11551)によるものです。ここに記して謝意を表すものとします。

文 献

[1] 渡辺知恵美, 中村聡史. オノマトペロリ: 味覚や食感を表すオノマトペによる料理レシピのランキング. 人工知能学会論文誌, Vol. 30, No. 1, pp. 340-352, 2015.

[2] ラートサムルアイパンカンウィパー, 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. 研究報告デジタルドキュメント (DD), Vol. 73, No. 6, pp. 1-7, 2009.

[3] 長谷川永奈, 小宮香乃, 齊藤史哲, 石津昌平. 食における言語資源に基づいたシズル感に関する因子情報の抽出. 日本感性工学会論文誌, Vol. 17, No. 2, pp. 299-308, 2017.

[4] 加藤亜由美, 深澤佑介, 森武俊. 五感と関連するオノマトペを用いた意外性の高い飲食店推薦. 人工知能学会論文誌, Vol. 30, No. 1, pp. 216-228, 2015.

[5] 石橋賢, 深瀧創, 宮田一乗. 米菓を対象としたシズル語の印象評価. 人工知能学会論文誌, Vol. 30, No. 1, pp. 229-236, 2015.

[6] 苺谷花子, 倉林修一, 清木康. 味覚印象を対象としたメタデータ生成方式と印象検索方式の実現. 情報処理学会研究報告データベースシステム, Vol. 134, No. 71, pp. 145-152, 2004.

[7] 大仁田龍也, 北山大輔. アレンジ抽出のための手順と材料を考慮したレシピの典型度算出手法. 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 2020) 論文集, 2020.

[8] 大仁田龍也, 北山大輔. レシピのアレンジ抽出のための調理手順の典型度算出手法. 電子情報通信学会技術研究報告, Vol. 119, No. 201, pp. 7-11, 2019.

[9] 土居洋子, 辻田美穂, 難波英嗣, 竹澤寿幸, 角谷和俊, 関洋平. 料理レシピと特許データベースからの料理オントロジーの構築. 電子情報通信学会技術研究報告, Vol. 113, No. 468, pp. 37-42, 2014.

[10] 清丸寛一, 黒橋禎夫, 遠藤充, 山上勝義. 料理レシピとクラウドソーシングに基づく基本料理知識ベースの構築. 言語処理学会第24回年次大会, 2018.