

料理のアレンジに着目したレシピへのメタデータ付与手法

山本 啓太[†] 那須日向太^{††} 風間 一洋[†]

[†] 和歌山大学システム工学部システム工学科 〒640-8510 和歌山県和歌山市栄谷 930

^{††} 和歌山大学大学院システム工学研究科 〒640-8510 和歌山県和歌山市栄谷 930

E-mail: †{s216299,s206185,kazama}@wakayama-u.ac.jp

あらまし 近年クックパッドのようなユーザ投稿型レシピサイトが広く普及し、収録レシピ数も膨大になったが、同時にレシピの特徴や違いの判別がつきにくくなったことで自分の好みのレシピを探し出すことが困難になるユーザ投稿型レシピサイトにおける迷子問題が深刻になった。本論文では、この問題を解決するために、投稿者がレシピに施した独自のアレンジをアピールするためにタイトルや説明中で用いるアレンジ表現を半自動的に抽出し、それをユーザの好みを表す嗜好ラベルとして教師あり機械学習を用いて自動的にレシピに付与する手法を提案する。具体的には、「～風」や「～風味」のようなアレンジされた料理の特徴を表す文字列と、「簡単」や「ヘルシー」のようにアレンジの目的を表す文字列の2種類をアレンジ表現として、複数の料理に対して、アレンジ表現をタイトルに含むレシピを正例、どれも含まずつくれば数が多いレシピを負例として、レシピから抽出した特徴で分類器を学習させ、レシピに嗜好ラベルを付与する。実際に、cookpad データセットに提案手法を適用して、アレンジ表現や特徴抽出方法による分類性能の違いを評価する。

キーワード レシピ, アレンジ, メタデータ, クックパッドデータセット

1 はじめに

インターネットや SNS の普及により、クックパッドや楽天レシピのようなユーザ投稿型レシピサイトが広く普及し、利用されるようになった。例えば、現時点でクックパッドの月間利用者数は約 5400 万人、レシピ数は約 320 万品を超えており、さらに日々多くのレシピ数が投稿され続けている。しかし、例えばハンバーグが 38581 レシピもあることからわかるように、ユーザがある料理を作ろうとしてカテゴリや検索結果から見つけようとしても、表示された多くのレシピの特徴や違いがよくわからず、目的のレシピを探し出すことが困難になるユーザ投稿型レシピサイトにおける迷子問題が深刻になった [1]。このために、逆に収録数が少なくても選別されているようなレシピ集から探すために、Google 検索で“-cookpad”を指定して検索結果から除外することさえおこなわれている¹。

ただし、根本的な問題はユーザに提示されるレシピ数とその中のノイズの多さにあり、決して収録レシピ数の多さではない。これは Google 以前のサーチエンジンが「使いものにならない」と酷評されていた状況と同じであり、もし典型的またはユーザの嗜好に合ったレシピだけに絞り込んだり、それらを優先的に提示できれば、逆に収録数の多さは利点となるはずである。

この問題は、投稿者が既存の料理を独自にアレンジしていることが主な原因であり、さらに投稿者も多くのレシピから自分のレシピを選んでもらうために、実際に調理するユーザにアレンジの特徴を簡潔にわかりやすく説明する努力を強いられている状況に着目する。

本論文では、投稿者がレシピに施した独自のアレンジをア

ピールするためにタイトルや説明中で用いるアレンジ表現を半自動的に抽出し、それをユーザの好みを表す嗜好ラベルとして教師あり機械学習を用いて自動的にレシピに付与する手法を提案する。具体的には、「～風」や「～風味」のようなアレンジされた料理の特徴を表す文字列と、「簡単」や「ヘルシー」のようにアレンジの目的を表す文字列の2種類をアレンジ表現として、複数の料理に対して、アレンジ表現をタイトルに含むレシピを正例、どれも含まずつくれば数が多いレシピを負例として、レシピから抽出した特徴で分類器を学習させ、レシピに嗜好ラベルを付与する。このように、レシピを調理するユーザ側の視点に立った、頻繁に使われるような嗜好ラベルを抽出し、レシピに付与することができれば、ユーザの嗜好に沿ったレシピの検索・提示を支援できると考えられる。

2 関連研究

レシピの推薦・検索、レシピの特性の抽出、レシピの分類の3つの観点から、それぞれに関連する研究を紹介する。

2.1 レシピの推薦・検索

崔らはメニューごとの共通食材に着目し、多数の類似レシピから基準となるような典型的レシピを抽出し、そのレシピとの差分を用いたレシピ検索法を提案した [2]。崔らの手法では、基準レシピとの使用食材の差分が大きいレシピを提示することで、料理のレパートリーを拡大することを目的としている。

高畑らは料理レシピを推薦する際にはユーザの好き嫌いといった嗜好を考慮することが重要だと考え、食材に対する好き嫌いを考慮した料理レシピ推薦システムを提案した [3]。レシピの閲覧履歴を利用し、食材に対する嗜好として、好きの推定に

1 : <https://ceron.jp/url/www.moneypost.jp/586434>

は TF-IDF の考えを応用した食材スコアを用い、嫌いの推定には閲覧したが実際には調理しなかったレシピに含まれる食材を利用している。食材利用頻度 (FFk:Foodstuff Frequency) と食材の特異度 (IRFk:Inverted Recipe Frequency) から算出した食材スコア (FFk-IRFk) を使用する。

横井らは食材の組合せを考慮した食材の追加・削除の推薦とユーザが意図した料理レシピを検索するために、食材の組合せの典型度を分析する手法を提案した [4]。各料理のレシピで使用される食材集合に主成分分析を適用し食材の典型度を求め、その結果を用いて、より典型的な食材の組合せに更新するための食材、またはより非典型的な食材の組合せに更新するための食材の追加・削除の推薦と、ユーザが意図的に含めた食材である非典型的な食材を利用して、ユーザの意図を含む料理レシピの検索を目指した。

2.2 レシピの特性の抽出

花井らは類似レシピを判断する上で料理の主となる食材と味の決め手となる調味料の一致が重要であるとし、レシピデータからその料理の主食材と主調味料を抽出し、それらとそのレシピの料理名に一般的に用いられない食材を用いた類似レシピをクラスタリングする手法を提案した [5]。例えばカレーという料理名に対して一般的に用いられる食材はじゃがいもやたまねぎ、一般的に用いられない食材はちくわであり、カレーでちくわを用いているレシピを類似レシピとしてクラスタリングする。

重田らはある料理に関して、複数のレシピを手順を比較し、その共通部分を同定することで、その料理の典型的な手順を出力するシステムを構築した [6]。

高らは「～風」のような料理タイプを対象とし、各料理タイプに特徴的な食材の組み合わせを抽出する手法を提案した [7]。食材や食材集合の出現頻度に着目し、料理タイプとの関係性を分析する。本論文では料理のアレンジの判定に、「～風」に加え「～風味」にも着目し、さらに、「簡単」や「ヘルシー」などのレシピタイトル中の料理を説明する頻出表現を加えた 2 種類を組み合わせる。

矢嶋らは料理レシピが「かんたん」であるとは「調理手順が少ないこと」、「使用する調味料や食材が少ないこと」、「買い足す調味料が少ないこと」、「調理時間が短いこと」であると定義した [8]。

橘らは対象レシピの材料と対象料理の典型的な材料との差異に修飾表現を用いた意図があると考え、「子供が喜ぶ」のようなレシピの特徴的な表現を含む料理名のレシピと含まないレシピを抽出し、それぞれのレシピに含まれているすべての材料の RF (Recipe Frequency) 値を算出した [9]。RF 値とは、重要語抽出に用いられる DF (Document Frequency) の考えに基づく値であり、RF 値をそれぞれの全レシピ数で割った値をその材料の使用率とし、その使用率が閾値以上となる材料を対象とする料理名の典型的な材料として抽出し用いた。最後にその表現を含むレシピの材料と含まないレシピの材料を比較することでその表現に特徴的な材料を導いた。

表 1 アレンジ表現

レシピタイトル	アレンジ表現
きのこたっぷりヘルシー和風チャーハン	ヘルシー, 和風
簡単ヘルシーえのき入り韓国風牛そばろ	簡単, ヘルシー, 韓国風
カレー風味のヘルシーナン	カレー風味, ヘルシー

2.3 レシピの分類

高田らは既存のメタデータ付きレシピデータのレシピタイトル、食材、調理手順を考慮した特徴ベクトル間のコサイン類似度を用いて、未知のレシピデータにメタデータを自動で付与する方式を提案した [10]。

水谷らはタイトルに「～風」のような料理タイプを含むレシピに注目し、各料理タイプごとの TF-IDF 値を求めることで「～風」を含まないレシピに対しても特徴づけを試みた [11]。信本らは人手で収集した任意の料理をある国の類似した料理名と差分となる要素によって「お好み焼き = チヂミ - ニラ + キャベツ」のような数式で表す手法を提案した [12]。対象となる料理名の食材とその料理名とのコサイン類似度が高い料理名の食材の特徴ベクトルの差を求めることで、どちらか一方の料理でのみ出現する食材を抽出した。

3 アレンジに着目したメタデータ付与手法

3.1 アレンジ表現

アレンジ表現とは、レシピ投稿者がレシピに加えた変更を、そのレシピを調理するユーザに簡潔・的確に伝えるために用いる文字列表現である。実際にレシピのタイトルで用いられているアレンジ表現の例を、表 1 に示す。この例からわかるように、アレンジ表現は「和風」、「韓国風」、「カレー風味」のようにアレンジされた料理の特徴を表す文字列（これをタイプ 1 と呼ぶ）と、「ヘルシー」、「簡単」のようにアレンジの目的を表す文字列（これをタイプ 2 と呼ぶ）に大きく分けられる。また、一つの料理に対して複数のアレンジが施されることがある。

このようなアレンジ表現は、ユーザ投稿型レシピサイトに存在する多くの類似レシピの中から自分の投稿したレシピを調理してもらえるように独自のアレンジをアピールするために、タイトルや説明文の中で用いられる。新しい料理の創作を積極的に試みる先進的な職業料理人と異なり、ユーザ投稿型レシピサイトのほとんどのユーザの場合は既存の料理に独自のアレンジを加えることが多いことから、投稿されたレシピの独自性は、そのようなアレンジの組み合わせによって決定されると仮定できる。そこで、アレンジ表現からレシピに施されているアレンジを決定できれば、ユーザの嗜好や目的に合ったレシピに絞り込むことができ、ユーザ投稿型レシピサイトの迷子問題を解決に役立つと考えられる。

ただし、アレンジ表現が使われていないからと言って、そのアレンジが施されているとも限らない。また、「ヘルシー」や「健康的」のように同一のアレンジを指すアレンジ表現が複数存在することから、アレンジ表現をそのままレシピのアレンジ判定に用いることはできない。

3.2 提案手法の概要

クックパッドでは、図1に示すように、レシピ投稿者が既存の料理の標準レシピに独自のアレンジを加えて調理した結果特に自信があるレシピを投稿し、ユーザはそこから自分や家族の嗜好にあったレシピを探し出して調理する。さらに、特にそのレシピが気に入った場合には、その感想を料理の写真付きで投稿できる「つくれば」と呼ぶ仕組みが提供されている。中村らは、つくればではレシピ投稿者に対する感謝が示されていることが多く、「美味しくなかった」や「難しかった」のようなネガティブな投稿は少ないと分析している[13]ことから、つくれば数はレシピに対するポジティブな評価指標として使用できる。

そこで、投稿されたレシピとつくればのデータを利用して、以下の手順でユーザの嗜好をメタデータとしてレシピに自動的に付与する。

- (1) アレンジ表現を用いたアレンジレシピ群の抽出 (図1の [1])
- (2) つくれば数を用いた標準レシピ群の抽出 (図1の [2])
- (3) 教師あり学習によるレシピへのメタデータ付与 (図1の [3])

3.3 アレンジ表現を用いたアレンジレシピ群の抽出

まず、レシピのタイトルから2種類のアレンジ表現を抽出して、レシピとアレンジ表現集合の組を抽出する。何らかのアレンジ表現を含むレシピを、本論文ではアレンジレシピと呼ぶ。すでに述べたようにアレンジ表現はレシピの説明文にも出現するが、タイトルにおける出現率の方が高いこと、タイトルの方が解析が容易で高い適合率で抽出できることから、タイトルだけを用いる。なお、アレンジ表現の再現率は低くなるが、抽出されなかったレシピに対しても抽出結果から教師あり学習を用いて付与するために問題にはならない。

「和風」、「カレー風味」のようなタイプ1のアレンジ表現は、アレンジされた料理の特徴を表す文字列に「～風」、「～風味」のような表現が付加される特徴を持つ。そこで、これらを手掛かり表現として、キューとスタックを用いて以下の手順で抽出する。

- (1) レシピのテキストの英字・数字は半角、カタカナは全角に統一し、記号類は除去する。
- (2) タイトルを日本語形態素解析した結果をキューに追加する。
- (3) キューから形態素を一つ取り出す。キューが空なら終了する。
- (4) 形態素が「風」または「風味」の場合は、スタックから直前の形態素をポップし、結合した文字列をアレンジ表現とする。
- (5) 形態素の文字列の末尾が「風」または「風味」の場合には、それをアレンジ表現とする。
- (6) 形態素をスタックにプッシュして、(2)に戻る。

「ヘルシー」、「簡単」のようなタイプ2のアレンジ表現は、タイプ1のように手掛かり表現を使うことはできない。そこで、レシピのタイトルを日本語形態素解析してから、すべてのレシ

ピにおける形態素の頻度を計算し、次の条件を満たす文字列を手手で選択してアレンジ表現とした。

- 形態素の品詞が名詞である。
- 形態素の文字列がアレンジの目的を表している。
- 形態素の頻度が閾値 T 以上である。

なお、「簡単」や「手軽」のように同じ意味で使われる単語がある場合は、出現頻度が多い方だけを用いる。この理由は、メタデータの付与を教師あり学習でおこなうために、同じ意味である場合は必ずしも必要とは限らないからである。

3.4 つくれば数を用いた標準レシピ群の抽出

次に、各料理に対する一般的なレシピ群を抽出する。本論文では、これを標準レシピと呼ぶ。このようなレシピを抽出する既存研究が多く存在するが、本論文ではすでに抽出されたアレンジ表現を利用して、以下の2条件を満たすレシピを標準レシピとみなす。

- アレンジされていないレシピ。これはアレンジ表現の有無で判定する。
- ある料理のレシピの中で代表的と言えるレシピ。これはつくれば数の大小で判定する。

実際には、人手で選定した料理名集合に対して、以下の手順で標準レシピを抽出する。

- (1) 指定された料理名をタイトルに含むレシピ群を抽出する。
- (2) すでにアレンジ表現が抽出されているレシピを除外する。
- (3) つくれば数で降順にソートする。
- (4) さらに不適切と思われるレシピ (例、「中華サラダ」) を人手で除く。

この結果、料理名とその標準レシピの順位付きリストが得られる。

3.5 教師あり学習によるレシピへのメタデータ付与

まず、用意した料理名集合から、以下の手順で指定されたアレンジ表現の教師データを作成する。

- (1) 料理名集合から、料理名を取り出す。
- (2) 指定されたアレンジ表現のアレンジレシピ群から、料理名を含むアレンジレシピをランダムに最大 N 件抽出する。これを正例とする。
- (3) 指定された料理名の標準レシピ群から、標準レシピを上位から最大 N 件抽出する。これを負例とする。

なお、料理名集合は、以下の条件を満たすように人手で調整する。

- 正例と負例の件数がほぼ同数であり、上限の N に近い。
- アレンジ表現が示すアレンジと料理が競合しない。例えば、中華料理である「麻婆豆腐」に「中華風」のアレンジを施すことは矛盾していて、適切な教師データにはならないからである。

次に、レシピの特徴ベクトルは以下の手順で作成する。

- (1) レシピのテキストの英字・数字は半角、カタカナは全

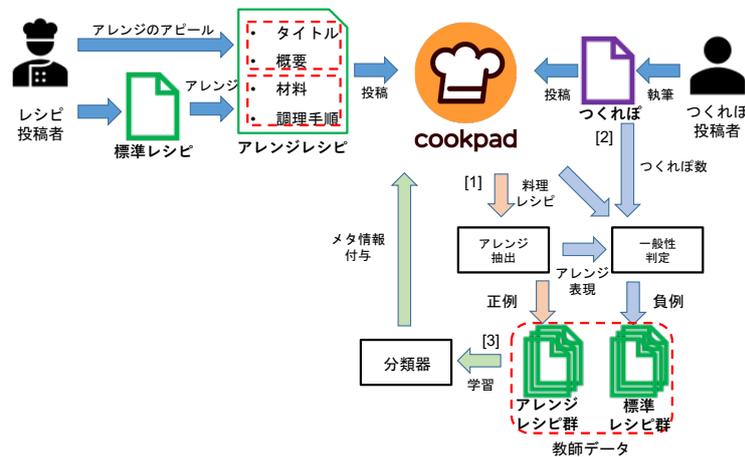


図 1 提案手法の概要

角に統一し，記号類は除去する。

(2) テキストを日本語形態素解析して，品詞が名詞と形容詞である形態素を抽出する。

(3) 形態素の大文字を小文字に，カタカナをひらがなに変換する。日本語形態素解析の後で行う理由は，この変換は解析ミスを引き起こすからである。

(4) ストップワードリストに含まれる形態素を除去する。例えば，食器・調理器具名（例，スプーン，包丁），人称名詞（例，みんな）などを含む。

(5) 形態素のレシピ中の出現頻度をカウントし，BoW (Bag of Words) ベクトルを作成する。

レシピのテキストとしては，本論文ではタイトル，概要の 1 段落目，材料のデータだけを使用する。この理由は，アレンジは材料，つまり使用する食材や調味料の違いによることが多く，題名や概要の最初の部分にはレシピ作成者がユーザの目に止まるようにアレンジの特徴をアピールすることが多いからである。なお，今回は手順は使用しない。例えば「ヘルシー」というアレンジでは調理法も重要な素性となると考えられるが，表記の揺れも多く，現時点では我々がその問題に対応できていなかったからである。また，本手法では表記揺れに対応するために単語の分散表現を利用できない。この理由は，例えば「イタリア風」ではサラダ油をオリーブオイルに，「ヘルシー」では牛乳を豆乳に変更することが考えられるが，word2vec の Skip-gram モデルでは，単語の意味はその周辺の単語（文脈）によって決定されるという分布仮説に基づくために，このような代替食材の類似度が高くなり，分類性能が低くなるからである。

最後に SVM (Support Vector Machine) と決定木 (Decision Tree) を用いて，教師データから上記の手法で特徴ベクトルを作成して学習し，学習した分類器を用いて，各レシピに嗜好ラベルを付与する。2 種類の学習アルゴリズムを用いる理由は，以降で述べる特徴ベクトルの作成方法を作成された決定木を調べて，特徴ベクトルの作成方法を検討するためである。

4 評価

クックパッド株式会社と国立情報学研究所により提供された

表 2 アレンジ表現と出現頻度

タイプ 1	頻度	タイプ 2	頻度
和風	22706	簡単	164118
中華風	8187	ヘルシー	11973
洋風	4431	甘辛	7126
韓国風	3412	手軽	6477
カレー風味	2619	離乳食	5904
タイ風	1410	濃厚	5795
ピザ風	1106	本格	4956
お好み焼き風	1022	定番	4776
レモン風味	831	シンプル	4728
胡椒風味	828	自家製	4259

クックパッドデータセットのうち，1998 年 4 月 23 日から 2014 年 9 月 30 日までに公開されたレシピ・つくればデータを使用した [14]。タイプ 2 では頻度が閾値 $T = 1000$ 以上のアレンジ表現を使用した。

4.1 アレンジ表現抽出法の妥当性評価

アレンジ表現を含むレシピがどの程度存在し，またそれぞれのアレンジ表現について教師データを作るために十分なレシピ数を確保できるかを評価するために 2 種類のアレンジ表現とそのレシピ数を求めた。タイプ 1 とタイプ 2 のアレンジ表現の上位 10 件を表 2 に示す。タイプ 1 はタイトルから自動抽出できた表現とそのレシピ数，タイプ 2 はタイトルを形態素解析して得られた頻出表現とそのレシピ数を示す。結果を見るとタイプ 1 では「和風」，タイプ 2 では「簡単」が一番多かった。これは日本人向けの味付けにアレンジされたレシピや簡単に作れるレシピが好まれやすいからと考えられる。タイプ 2 では，「簡単」，「手軽」，「シンプル」のようなおおよそ同じ意味で使われる単語が複数存在することが確認できた。また，タイプ 1 では「和風」，「中華風」，「洋風」のような主な表現は確保できたが，アレンジ表現の出現頻度が違うため，頻度が少ない場合にはさらなる検討が必要である。

4.2 レシピの分類精度の評価

レシピの特徴ベクトルをどのように作成すると分類性能が高

表3 レシピの分類精度

	SVM			決定木		
	手法1	手法2	手法3	手法1	手法2	手法3
和風	0.785	<u>0.988</u>	0.819	0.751	<u>0.988</u>	0.752
ヘルシー	0.672	<u>0.896</u>	0.740	0.594	0.879	0.685
中華風	0.859	<u>0.988</u>	0.840	0.800	<u>0.988</u>	0.803
簡単	0.573	<u>0.958</u>	0.621	0.528	0.956	0.551

くなるのかを調べるために以下の3つの手法で比較した。

手法1 材料のみに含まれる単語で特徴ベクトルを作成する

手法2 タイトル, 概要の1段落目, 材料に含まれる単語で特徴ベクトルを作成する

手法3 手法2より「アレンジ表現」を除いた単語で特徴ベクトルを作成する

本論文では, Pythonのオープンソース機械学習ライブラリであるscikit-learn²のSVM(RBFカーネル)と決定木を使用し, アレンジ表現にはタイプ1より「和風」と「中華風」をタイプ2より「ヘルシー」と「簡単」の4種類を使用した。これらは, 表2において上位2つの単語である。

10分割交差検証で求めた平均精度の結果を表3に示す。アレンジ表現を用いる手法2がもっとも平均精度が高かった。また, 平均精度が手法1よりも手法3の方が高いことからアレンジ表現以外にもタイトルや概要の1段落目にある単語が判別に影響することがわかった。

4.3 判別理由の分析

判定根拠を検討するために「和風」と「ヘルシー」の手法1と手法3の場合の可視化結果を分析した。手法2の結果は過学習していると思われたため今回は省略した。和風では, 図2(a)と図2(b)より, 醤油やつゆ, だしがあり, gini係数も高いことから, 材料, 特に和風調味料が識別に影響すると考えられる。ケチャップが含まれている理由は, 日本における洋食に用いられることが多いからであり, 和風アレンジにも使われると考える。ヘルシーでは, 図3(a)と図3(b)より, 豆腐, 豆乳, ヨーグルトのような健康的な食材が含まれていた。また, 食材以外にも「簡単」や「離乳食」, 「弁当」のような調理目的を表す単語も含まれていた。

4.4 未知レシピに対する分類の評価

実際のレシピに対する提案手法の有効性を検証するため, 手法1, 手法2, 手法3に対して教師データ以外のレシピからサンプリングしたレシピを分類した結果を分析した。本論文では, テストデータを以下の2点に設定し, 「和風」と「ヘルシー」に対してそれぞれ5種類の料理名を10件ずつ合計50件のレシピを使用した。

条件1 もともとそのアレンジ属性を持つ料理名または似たようなアレンジ表現を含むレシピの場合

条件2 そのアレンジ属性となりえない料理名またはアレンジ表現を含むレシピの場合

表4 未知レシピの分類精度

	和風			ヘルシー		
	手法1	手法2	手法3	手法1	手法2	手法3
条件1	0.88	0.04	0.92	0.38	0.14	0.38
条件2	0.96	1.00	0.78	0.84	0.94	0.72

和風では条件1に親子丼, すき焼き, 味噌汁, ぶり大根, おでん, 条件2にトムヤムクン, ビビンバ, タコス, ボルシチ, パエリアのレシピを使用し, ヘルシーでは条件1にノンオイル, 低カロリー, そうめん, 白和え, ナムル, 条件2にこってり, 濃厚, ラーメン, 焼肉, アヒージョのレシピを使用した。テストデータでは, 対象となるアレンジ表現を含むレシピを除いているが, レシピの詳細は検討していない。評価については, まず条件1では再現率(Recall), 条件2では特異率(Specificity)を求め, 次に誤って判別したレシピについて詳細を確認し, 考察を行う。

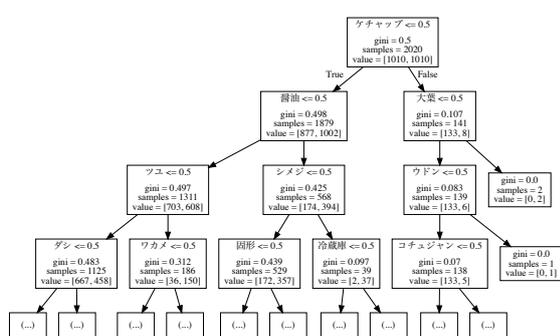
結果を表4に示す。和風の場合, 再現率と特異率はともに高かったが, 条件1では手法2が明らかに他と異なる結果となった。醤油やつゆ(めんつゆ)のような和風の判別に用いられる和風調味料が使われることが多い料理を対象としたにも関わらず, 手法2では和風であると判別されたレシピはほとんどなかった。これは単に和風というアレンジ表現があるかないかで判別されている可能性が考えられる。手法1, 手法3では醤油やつゆ(めんつゆ)のような和風調味料が使われている料理はある程度問題なく分類できた。和風でない判別されたレシピは和風調味料を使っていなかったり, 本だしやすき焼きのたれを使用したりしていた。条件2において, 手法1や手法3が和風であると分類しているレシピの詳細を見ると材料にしめじのようなきのこ類のような食材や醤油のような和風調味料が含まれていたため, これらの分類は妥当であると考えられる。

ヘルシーの場合, 特異率は高かったが再現率は低かった。これはヘルシーアレンジでは, 高カロリーな材料を減らしたり, ヘルシーな別の食材で代替したり, 量を減らしたり, 肉を減らして豆腐を加えるといったような配分率を変えたりするケースが多く, これらは材料から判断することができないためだと考えられる。実際, 今回上手く判別できたレシピは材料より判別できるヘルシーな別の食材で代替したレシピが多かった。条件2において, ヘルシーであると判別されたレシピは手法1では豆腐やトマトのような健康的な材料が含まれていた。手法3ではどちらかといえば材料よりもタイトルや概要の1段落目の表現が分類に依存していたようであり, ヘルシーなレシピだと判断されたレシピは例えば, たまごの白身を抜いた「全卵」という単語や「ダイエット」, 「メレンゲなし」のような単語がタイトル中に含まれていた。

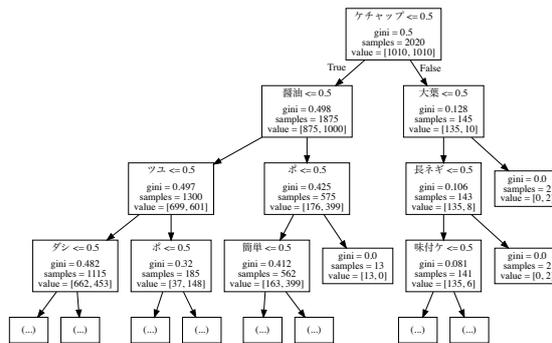
5 おわりに

本論文ではアレンジ表現とつくれば数に着目してアレンジレシピと標準レシピを抽出し, それらを教師データとして学習したSVMと決定木を用いて, レシピに料理のアレンジに関する

²: <http://scikit-learn.org/stable/>

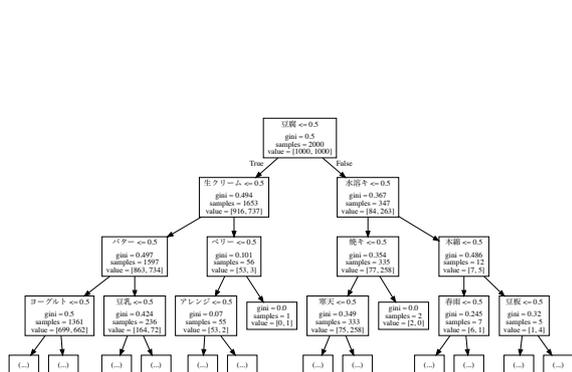


(a) 手法 1

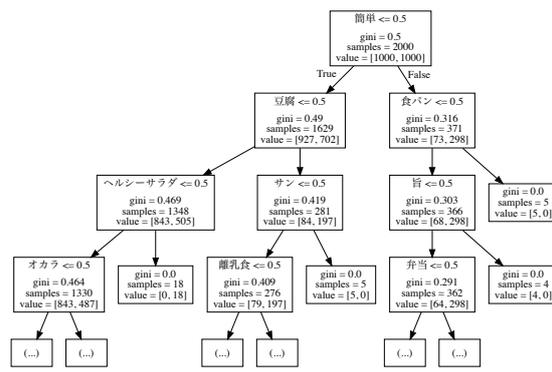


(b) 手法 3

図 2 「和風」の場合の決定木



(a) 手法 1



(b) 手法 3

図 3 「ヘルシー」の場合の決定木

メタデータを付与する手法を提案した。さらに複数の特徴ベクトル作成手法を比較検討した結果、食材に加えてタイトルや概要を用いることで性能を向上できることがわかった。ただし、決定木を可視化して分析した結果、アレンジ表現のタイプにより適している特徴抽出法は異なり、タイプ1は食材、特に調味料が重要な手がかりになることにに対し、タイプ2は食材だけでは性能が低く、タイトルや概要中の調理の目的を表す単語が重要な手がかりになることがわかった。

今後の課題として、まず表記揺れ解消のために単語の分散表現を使用できない提案手法では、料理オントロジーなどのデータを用いた正規化をおこなう。さらに、今回使用しなかった調理手順や材料の分量なども用いて、異なる性質を持つアレンジに対しても妥当な性能を確保できる特徴ベクトル作成手法を検討する予定である。

謝 辞

本研究では、国立情報学研究所の IDR データセット提供サービスによりクックパッド株式会社から提供を受けた「クックパッドデータセット」を利用した。また、本研究は JSPS 科研費 17H01826 の助成を受けた。

文 献

- [1] 「クックパッド離れ」に歯止めはかかるか かつての愛好者たちの本音。マネーポスト WEB, 2019. <https://www.moneypost.jp/586434>.
- [2] 崔赫仁, 塩井隆円, 楠和馬, 波多野賢治. メニューごとの共通食材に着目した料理レシピ検索手法の提案. 第 9 回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 2017) C1-1, 2017.
- [3] 高畑麻理, 上田真由美, 中島伸介. 食材に対する好き嫌いを考慮した料理レシピ推薦手法の提案. 第 3 回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 2011) E3-5, 2011.
- [4] 横井聡, 道満恵介, 平山高嗣, 井手一郎, 出口大輔, 村瀬洋. 料理レシピにおける食材の組合せの典型度分析. 電子情報通信学会技術研究報告 = IEICE technical report : 信学技報, Vol. 114, No. 487, pp. 49-54, mar 2015.
- [5] 花井俊介, 難波英嗣, 灘本明代. 主食材と主調味料を考慮した類似レシピクラスタリング. 第 8 回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 2016) E2-7, 2016.
- [6] 重田謙博, 難波英嗣, 竹澤寿幸. 複数料理レシピからの典型手順の自動生成. 第 9 回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 2017) C2-1, 2017.
- [7] 高霞, 尾崎知伸. 料理タイプに特徴的な食材集合の抽出について. 第 79 回全国大会講演論文集, Vol. 2017, No. 1, pp. 405-406, mar 2017.
- [8] 矢嶋亜紗美, 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. 日本知能情報ファジィ学会 ファジィ システム シンポジウム 講演論文集, Vol. 25, pp. 31-31, 2009.
- [9] 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抽出. 第 5 回データ工学

と情報マネジメントに関するフォーラム (DEIM Forum 2013) P3-5, 2013.

- [10] 高田夏彦, 上田真由美, 森下幸俊, 中島伸介. 各種メタデータの特徴を考慮した未知レシピへのメタデータ自動付与方式の提案. 第7回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 2015) C1-3, 2015.
- [11] 水谷真子, 尾崎知伸. 概念階層を考慮した拡張 tf-idf 法に基づく料理タイプの分析. 第79回全国大会講演論文集, Vol. 2017, No. 1, pp. 407-408, mar 2017.
- [12] 信本健輔, 廣田雅春, 加藤大受, 石川博. アナロジーによる料理の国際的な表記方式の提案. 知能と情報, Vol. 31, No. 1, pp. 526-533, 2019.
- [13] 中村裕太, 佐藤弘喜. 料理レシピサイト内のクチコミのテキスト解析:クチコミの信頼性に関する研究. 日本デザイン学会研究発表大会概要集, Vol. 65, pp. 16-17, 2018.
- [14] クックパッド株式会社. クックパッドデータ. 国立情報学研究所情報学研究データリポジトリ, (データセット), 2015. <https://doi.org/10.32130/idr.5.1>.