

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^[1] 奈須 日向太^[2] 風間 一洋^[1]

[1] 和歌山大学システム工学部 [2] 和歌山大学大学院システム工学研究科

はじめに

背景

- 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た
- 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てきている
 - 同じ名前の料理でも、細部が異なるレシピが膨大に存在する
 - ある料理のレシピについて、他のレシピと何が違うのか、どのような特徴があるのかわかりにくい

目的

- ユーザの嗜好にあったレシピを見つけやすくすることで、ユーザ投稿型レシピサイトにおける迷子問題の解決を図る
 - レシピのタイトルとつくればのデータを使って教師データを自動作成
 - その教師データを用いて、各レシピにメタ情報を付与

類似レシピの関連研究

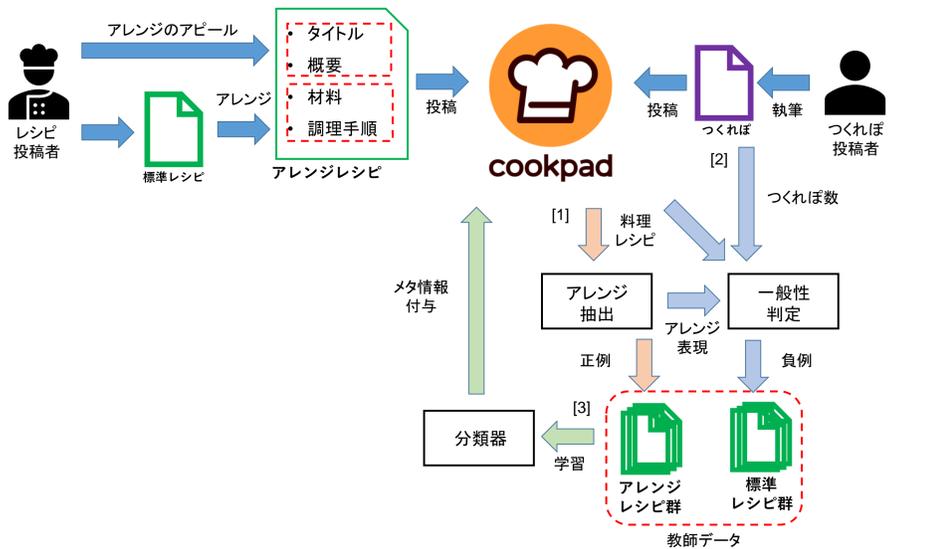
- 複数料理レシピからの典型手順の自動生成[重田ら 2017]
 - 複数レシピから典型的なレシピを作成する
- 主食材と主調味料を考慮した類似レシビクラスタリング[花山ら 2016]
 - 類似レシビをグループ化する
- 各種メタデータの特徴を考慮した未知レシピへのメタデータ自動付与方式の提案[高田ら 2016]
 - 人手で作成した教師データを用いてメタデータを付与する

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 韓国風牛そぼろ	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

- レシピ投稿者がレシピに加えた変更を、簡潔・的確に伝えるタイトル中の表現
- すべてに付けられているとは限らない
- 高らは、レシピの料理タイプの判定に「～風」のような表現に注目した[高ら 2017]
- 本研究では料理のアレンジの判定に、次の2種類を料理名と組み合わせて用いる
 - 「～風」, 「～風味」: タイトルから自動抽出
 - 「ヘルシー」(「健康的」なども含む), 「簡単」など: 事前選択したタイトル中の頻出表現と照合

提案手法の概要



1. アレンジレシピの抽出 (正例)

- 対象とする料理名集合を、タイトル中の出現レシピ数が多いものから選択する
- タイトルで対象とする料理名をアレンジ表現が修飾しているレシピを、アレンジレシピとして抽出する

2. 基本レシピの抽出 (負例)

- タイトルに料理名が含まれても、アレンジ表現がなく、つくればの投稿数が上位のレシピを、基本レシピとして抽出する

3. 学習モデルによるレシピへのタグ付け

- 料理ごとにアレンジレシピと基本レシピがほぼ同数となるように選択する
- レシピのタイトル, 概要 (第一段落のみ), 材料のテキストを正規化後に形態素解析し, 特徴ベクトルを作成する
- アレンジ表現ごとに正例・負例で学習したSVMを用いて, 各レシピにメタデータを付与する

データセット

- 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クックパッドデータ
- 2014年9月30日までに公開されたレシピ・つくればデータ

アレンジ表現の妥当性の評価

アレンジ表現	頻度	頻出表現	頻度
和風	22706	簡単	164118
中華風	8187	ヘルシー	11973
洋風	4431	手軽	6477
韓国風	3412	濃厚	5795
カレー風味	2619	本格	4956
タイ風	1410	定番	4776
ピザ風	1106	シンプル	4728
お好み焼き風	1022	手抜き	2686
レモン風味	831	節約	2638

- 表1はタイトルから自動抽出できたアレンジ表現とそのレシピ数を示す
- 表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
- レシピ数が500以上のアレンジ表現数は表1で20種類, 表2で23種類であった
- 表2では「簡単」, 「手軽」, 「シンプル」, 「手抜き」と同じ意味の単語が複数存在する
- 表1, 表2共に上位の単語に関しては教師データを作成できる量が確保できる

表1: 抽出されたアレンジ表現 表2: タイトル中の頻出表現

レシピの分類性能の評価

- レシピのタイトルと概要がどの程度判定に影響があるのかを調べるために以下の2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

- 調理手順はまだ使用できていない

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

- レシピ投稿者の意図が反映されている

- 学習にはSVM(RBFカーネル)と決定木を使用した
- アレンジ表現には「和風」と「ヘルシー」を使用した
- 性能評価には10分割交差検証で求めた平均精度を用いた

	SVM	決定木
手法1	0.999	0.969
手法2	0.993	0.994

「和風」の場合

	SVM	決定木
手法1	0.497	0.528
手法2	0.721	0.868

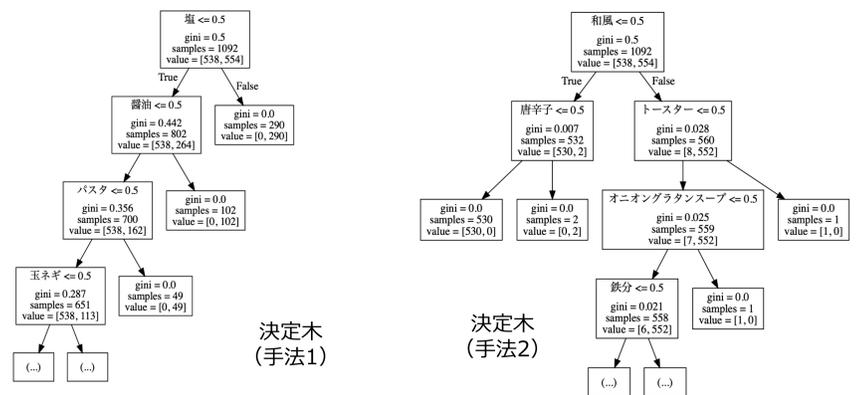
「ヘルシー」の場合

- 「和風」は現時点でも高い精度で判定できる
- 「ヘルシー」は材料だと高くないがタイトルと概要をいれると精度が向上した

判別理由の分析

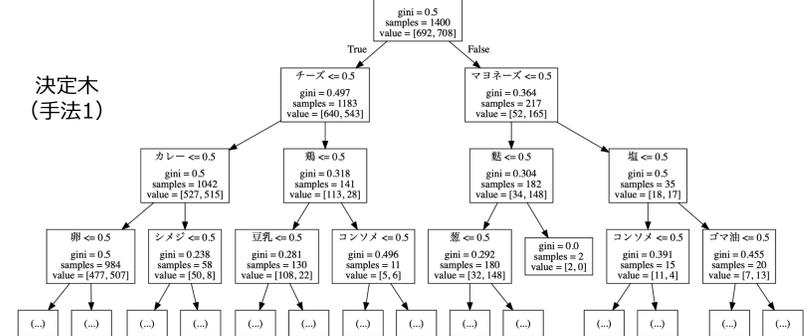
- 判定根拠を知るために手法1と手法2の決定木を可視化し, 結果を比較した

和風

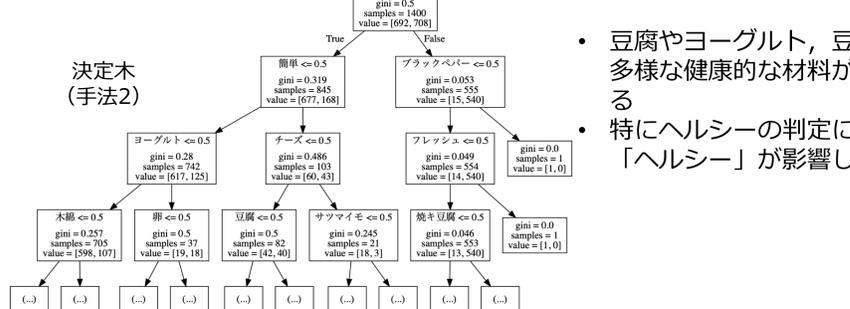


- 和風では特に調味料 (塩, 醤油) が判定に影響する

ヘルシー



決定木 (手法2)



- 豆腐やヨーグルト, 豆乳のような多様な健康的な材料が影響している
- 特にヘルシーの判定には「ヘルシー」が影響している

今後の課題

- 調理手順に対応する
 - 「簡単」などの判定に必要
- 表記ゆれを料理オントロジーなどで正規化する
 - word2vecの分散表現は使えない
 - 例: 牛乳と豆乳, 味噌と豆板醤などが同一視される
- 献立推薦システムに組み込む

参考文献

- 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, 2017.
- 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”, DEIM Forum 2015 C1-3, 2015.
- 花井俊介, 難波英嗣, 難本明代, “主食材と主調味料を考慮した類似レシビクラスタリング”, DEIM Forum 2016 E2-7, 2016.
- 重田識博, 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”, DEIM Forum 2017 C2-1, 2017.

連絡先

- 山本 啓太(和歌山大学システム工学部4年)
- Email: s216299@wakayama-u.ac.jp