

類似度分析に基づく未知レシピへのメタデータ自動付与方式の提案

高田 夏彦[†] 佐々江 駿[†] 上田真由美^{††} 森下 幸俊^{†††} 中島 伸介[†]

[†] 京都産業大学コンピュータ理工学部 〒603-8555 京都府京都市北区上賀茂本山

^{††} 流通科学大学総合政策学部 〒651-2103 兵庫県神戸市西区学園西町 3-1

^{†††} 大日本印刷株式会社 C & I 事業部 〒141-8001 東京都品川区西五反田 3-5-20

E-mail: [†]{g1144696,g1044516,nakajima}@cse.kyoto-su.ac.jp, ^{††}Mayumi __ Ueda@red.umds.ac.jp,
^{†††}morishita-y4@mail.dnp.co.jp

あらまし 近年、料理レシピの検索・推薦に関する取り組みが盛んに行われている。レシピ推薦サービスにおいて、各レシピに対してメタデータを付与することで推薦精度を向上させることが考えられる。しかしながら、全てのレシピに手動でメタデータを付与することは容易ではない。そこで本稿では、既にメタデータが付与されているマスターレシピとの類似度分析に基づいた、未知レシピへのメタデータ自動付与方式について提案する。

キーワード レシピ推薦, レシピ類似度判定, 特徴ベクトル抽出, メタデータ付与

Natsuhiko TAKATA[†], Shun SASAE[†], Mayumi UEDA^{††}, Yukitoshi MORISHITA^{†††}, and Shinsuke

NAKAJIMA[†]

[†]

^{††}

^{†††}

E-mail: [†]{g1144696,g1044516,nakajima}@cse.kyoto-su.ac.jp, ^{††}Mayumi __ Ueda@red.umds.ac.jp,
^{†††}morishita-y4@mail.dnp.co.jp

1. はじめに

近年、料理レシピの検索・推薦に関する取り組みが盛んに行われている。クックパッドでは、登録レシピ約 120 万品、月間利用者数 2000 万人以上 (2012 年 7 月)、また、味の素レシピ大百科では、約 9800 品のレシピが登録されており (2013 年 2 月現在)、料理レシピの検索・推薦サービスの需要が伺える。

レシピ推薦サービスにおいて、高精度に推薦アイテムの選定を行うためには、各レシピデータを分析し、適切なメタデータを付与するというアプローチが考えられる。例えば、“パーティメニュー”や“夜食に最適”、“子供が喜ぶ弁当メニュー”等のメタデータを付与することで、利用者の目的に合致するアイテムの推薦が可能となる。他にも、森下らはその日の気分あった献立を検索するシステムを開発している [1]。これは利用者が「元気 or お疲れ」「あっさり or こってり」「お手軽 or 本格」等の軸に基づいて、その日の自分の気分を入力すると、これに合致した献立を提示するシステムである。

ここで問題となるのは、各レシピデータに対して適切なメタデータを自動で付与することは容易ではなく、多くの場合、人

手によってメタデータが付与されている。したがって、大量のデータに適用することが難しいという問題がある。そこで本稿では、既にメタデータが付与されているマスターレシピとの類似度分析に基づいた、未知レシピへのメタデータ自動付与方式の提案を目的としている。これは、未知レシピへのメタデータ付与を考える際、この未知レシピと類似しているマスターレシピのメタデータを参考にしようとするものである。

2. 関連研究

レシピの類似度分析に関する研究として、苅米らは料理レシピの種類を豊富にすることを目的に、材料や調理手順の類似度を用いた関連検索を提案している [2]。関連の種類として、材料の類似、調理手順の類似、組合せの良さをあげ、材料の類似は各料理に含まれる食材をベクトル表現したもののコサイン類似度、調理手順の類似は調理手順のテキスト部分から抽出した動詞の出現順序の類似度から算出している。白川らは料理レシピからアニメーションを作成し、ユーザに提示することを提案している [3]。この取り組みでは、料理の基本的な動作に関して、レシピ中の表層表現とアニメーションを対応付ける情報を網羅

的に格納した料理動作辞書を作成している。志土地ら [4] は、同一料理カテゴリ中の料理群における特徴的な調理手順の類似度に基づく代替可能食材の発見手法を提案している。岩上ら [5] は、使用食材については tanimoto 係数を用い、料理手順については編集距離を用いて、その類似度を計算している。我々の手法は、食材、調理手順、レシピ名の 3 つの特徴ベクトルを抽出することで、より精密なレシピ間類似度を算出しようとするものであり、その新規性は高いと考えている。

3. レシピデータの特徴ベクトルの抽出とレシピ間類似度の算出

本研究では、メタデータ付きレシピデータを用いて、未知のレシピデータへのメタデータ自動付与を可能とする手法の提案を目指している。既にメタデータが付与されたレシピデータ（マスターレシピデータと呼ぶ）に基づいて、これらマスターレシピとの類似性を考慮して、未知レシピデータへのメタデータ付与を試みる。そこで、本節では、レシピデータの特徴ベクトルの抽出とレシピ間類似度の算出方法について説明する。

3.1 レシピデータの特徴ベクトル抽出

本節では、まずレシピデータの特徴ベクトル抽出手法について述べる。レシピデータの特徴量としては、レシピ名、食材、調理手順、栄養素、写真データなどが挙げられるが、全ての特徴量を考慮することは容易でない。したがって、本研究では、レシピ名、食材、調理手順を考慮した特徴ベクトル抽出に取り組む。なお、特徴ベクトルは各々に対して抽出する。すなわち、レシピ名特徴ベクトル、食材特徴ベクトル、調理手順特徴ベクトルである。また、マスターレシピとの類似度算出を目的としているので、特徴ベクトル次元はマスターレシピのベクトル次元を採用する。各特徴ベクトルの抽出方法順を以下に示す。

3.1.1 レシピ名特徴ベクトル

マスターレシピ群のレシピ名に対して形態素解析 (MeCab [6] を使用) を行い、レシピ推薦に不必要な単語 (以下、ストップワードと呼ぶ) を取り除いた後の全ての語をレシピ名特徴ベクトルの次元として採用する。新規レシピデータに関しては、同等にレシピ名に対して形態素解析を行い、出現回数に基づいて各次元の値を決定する。

3.1.2 食材特徴ベクトル

マスターレシピデータに対して形態素解析を行い、ストップワードを取り除いた後、少なくとも複数回出現するものを食材特徴ベクトルの次元として採用する。まだベクトルが作成されていない未知レシピのデータに関しては形態素解析を行い、同等の次元に関してベクトルを作成する。

3.1.3 調理手順特徴ベクトル

マスターレシピデータに対して形態素解析を行い、ストップワードを取り除いた後の動詞及び名詞を調理手順特徴ベクトルの次元として採用する。新規レシピに関しては形態素解析を行い、同等の次元に関してベクトルを作成する。この際、「水をきる」などのようなレシピ特有の表現については形態素辞書をカスタマイズすることで効率的に抽出できるように工夫する。未知レシピに関しては、同様に手順に対して形態素解析を行い、

対象単語の出現回数に基づいて各次元の値を決定する。以下に実際の例に基づいて、説明する。

対象となるレシピデータに対して MeCab [6] を用いて形態素解析し、名詞、動詞、形容詞を抽出する。なお、事前にレシピの特徴を表現するのに不要な語句をストップワードとしてリストアップしておき、これらを取り除くことで、できる限りレシピの特徴を表現するのに適切な語句に限定したベクトル次元を決定する。以下に、レシピ『若竹汁』を例に説明する。

=====

レシピ名：若竹汁

調理手順：

筍は薄切りにし、わかめは戻しておく。鍋に出し汁と筍を入れ、一煮立ちさせる。わかめを加え (A) で調味する。汁椀に注ぎ、吸い口に木の芽を浮かべる。

=====

これに対し形態素解析を行い、ストップワードを取り除くと表 1 の結果を得ることが出来る。

表 1 若竹汁のレシピから形態素を抽出したもの

口	薄切りにする	入れる	汁
浮かべる	木の芽	筍	椀
加える	鍋	調味する	出し汁
吸う	注ぐ	おく	させる
わかめ	煮立つ		

この処理をマスターレシピデータ全てに実施した上でとりまとめて、重複を整理することで、調理手順特徴ベクトルの次元を決定することができる (以下、マスターベクトル次元と呼ぶ)。各マスターレシピの特徴ベクトルは、マスターベクトル次元のうち、当該レシピに存在する語句に関しては数値 1、存在しない語句に関しては数値 0 とすることで得られる。

次に未知レシピの特徴ベクトル抽出方法について述べる。最終的にはマスターレシピのメタデータを考慮した、未知レシピへのメタデータ付与を目指している。未知レシピの特徴ベクトル次元は、マスターレシピのベクトル次元をそのまま継承する。手順としては、未知レシピについても形態素解析を行い、マスターベクトル次元のうち、当該レシピに存在する語句に関しては数値 1、存在しない語句に関しては数値 0 とすることで得られる。すなわち、マスターレシピに存在しない語句に関しては、未知レシピに存在したとしても特徴ベクトルには採用されない。以下に、『豚汁』を例に説明する。

=====

レシピ名：豚汁

調理手順：

大根、人参はいちょう切りにし、ごぼうはさがぎにする。冷凍里芋は流水で半解凍し、半分に切る。こんにゃくは下茹でし、短冊切りにする。ねぎは小口切りにする。鍋に水とねぎ以外の材料をすべて入れ、やわらかくなるまで煮る。具材がやわらかくなったら、味噌を加えてさっと煮る。器に盛り、ねぎを散らす。

=====

『豚汁』の未知レシピに対し形態素解析を行い、ストップワードを取り除いたものを表2に示す。

半分	解凍	小口切りにする
以外	里芋	ごぼう
やわらかい	水	ねぎ
短冊切り	冷凍	流水
散らす	入れる	煮る
すべて	材	味噌
加える	さがぎにする	豚汁
こんにゃく	大根	する
いちよう切りにする	具	茹でる
人参	器	切る
鍋	材料	盛る

これが未知レシピ『豚汁』の特徴を表現する語句集合であるが、特徴ベクトルの次元としては前述のマスターベクトル次元を採用するため、表2で示した語句のうち、マスターベクトル次元に存在する語句のみ、未知レシピ『豚汁』の特徴ベクトルとして考慮することになる。

ここで、レシピ特徴ベクトル次元としては採用しない語句リストであるストップワードリストを以下に示す。現在のストップワード数は、187語である。

~~~~~<ストップワードリスト>~~~~~  
か所, 供する, 人, 針, れる, きする, 前, れん, がく, く, のく, せる, ねぐ, は種, じす, まあ, スー, c m, よい, よう, ほう, 部, うち, さとい, なる, こ, の, め, とも, とわ, もの, す, げ, さ,, ぼる, いぬ, そう, つぎ, つま, きん, えむ, つる, ちる, ら, い, ん, てる, つけ, とさつ,, \*, 普通, か所, やすい, バッド, 二, 人, 縦, 度, からす, スー, もともの, 休む, にくい, その他, 保存, つる, 翌日, せる, 刷毛, 空気, ない, できる, ところ, いる, 向う, 止める, おきる, プロセッサー, 左, 時, ひく, ふう, 他, れる, 画びよう, 使用, とめる, 位, 1つ, 同様, 安定, く, 半量, 中央, 的, 直径, cm, よい, よう, ほう, 七, うち, 以上, これ, さ, 々所, 下, 上, 三, ビニール, くる, 一, 適当, 台, 取りはずす, 座る, 市販, 立つ, 立て, てる, 前方, ねる, たんぼ, 終わる, いかめしい, 残り, 残る, 操作, ころがす, 見る, 分, 各種, 浮き上がる, の, 順番, 時間, 管理, 各々, 全て, スナップ, ごと, 手早い, 大, 入手, 角, えむ, 戻す, 7, 外, えらい, ほう, かぶる, 次, 倍, 化粧, 日, それぞれ, 秒, 幅, ある, 長い, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, サイズ, 放す, 四隅, とも, 耐熱, 一緒, 用  
~~~~~

このように、レシピや食材に関係のない語句や記号などが多く含まれている。なのでストップワードを充実させることにより、データの類似度をより正確に出来ると考えている。

表3に、マスターレシピ次元および未知レシピ『豚汁』に含まれる語句から作成される『豚汁』の特徴ベクトル次元の一部を示す。レシピ間類似度を計算するため、ベクトル次元はマスターレシピ次元を使うことで統一されている。したがって、表

3に示すとおり、「小口切りにする」や「こんにゃく」のように、マスターレシピ次元に存在し、レシピ内にも記述されている語句は、数値1を与える。表3内で数値が0である語句は、マスターレシピ次元には含まれているが、『豚汁』レシピには含まれない語句である。

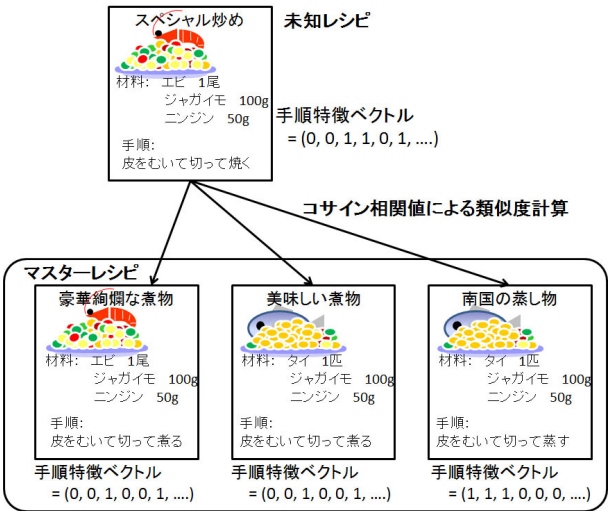
表3 マスターレシピ次元を用いた『豚汁』レシピ特徴ベクトルの一部

小口切りにする	1	材料	1
ほぐす	0	こんにゃく	1
すべて	1	以外	1
大根	1	器	1
お湯	0	切る	1
味噌	1	いちよう切りにする	1
茹でる	1	おく	0
散らす	1	人参	1
やわらかい	1	加える	1
石づき	0	ボウル	0
水	1	一口	0
短冊切り	1	酒粕	0
入れる	1	ねぎ	1
鮭	0	する	1
小房に分ける	0	盛る	1
注ぐ	0	鍋	1
溶く	0	しめじ	0
取る	0	/	/

以上の通り、マスターベクトル次元の抽出を行い、未知レシピの特徴ベクトル抽出を行う。なお、我々が用意したマスターレシピ493件に対して、マスターベクトル次元の抽出を行ったところ、次元数は1185次元となった。

3.2 レシピ間類似度算出方式

本節では、抽出された特徴ベクトルを用いたレシピデータ間の類似度計算方法について述べる。図1に、レシピ間類似度計算のイメージを示す。



各レシピは、レシピ名特徴ベクトル、食材特徴ベクトル、調理手順特徴ベクトルを有する。すなわち、レシピ間の類似度判

定としては、レシピ名類似度、食材類似度、調理手順類似度を個別に計算し、以下の式により統合することになっている。

$$Rsim = \alpha \cdot Tsim + \beta \cdot Isim + \gamma \cdot Psim \tag{1}$$

ただし、 $Rsim$ はレシピ間類似度、 $Tsim$ はレシピ名類似度、 $Isim$ は食材類似度、 $Psim$ は調理手順類似度である。それぞれ、3.1.1 節のレシピ名特徴ベクトル、3.1.2 節の食材特徴ベクトル、3.1.3 節の調理手順特徴ベクトルを用いて算出されるものである。また、 α, β, γ は重みであり、 $\alpha + \beta + \gamma = 1$ を満たす。なお、特徴ベクトル間の類似度計算には、コサイン類似度を用いる。

このように、1つのレシピを1つの特徴ベクトルで表すのではなく、レシピ名、食材、調理手順のそれぞれに対して特徴ベクトルを抽出し、個々の類似度を算出可能にすることで、種々の目的に応じた類似度計算が可能になる。例えば、「食材は類似しているが、調理手順は似ていないレシピの推薦」などが実現可能となる。

ここで、実際に未知のレシピと 493 件のマスターレシピとの類似度計算を行い、最も類似度の高いマスターレシピ上位 10 件を示す。表 4 には、未知レシピ「肉じゃが」との類似レシピ、表 5 には、未知レシピ「カレーライス」との類似レシピ、表 6 には、未知レシピ「粥」との類似レシピ、を示す。

表 4 「肉じゃが」と似ているレシピ上位 10 件

レシピ名	類似度
おからの煮物	0.420
かぼちゃの含め煮	0.389
ビーフカレー	0.369
里芋の煮ころがし	0.363
クリームシチュー	0.356
チンジャオロース	0.348
さやいんげんと豚バラ肉の煮物	0.348
味噌煮込みうどん	0.342
がんもどきと小松菜の含め煮	0.338
牛肉の柳川風	0.335

表 5 「カレーライス」と似ているレシピ上位 10 件

レシピ名	類似度
肉じゃが	0.445
ビーフカレー	0.400
おからの煮物	0.388
ポトフ	0.380
クラムチャウダー	0.375
金時豆煮	0.350
鶏肉と蕪のクリーム煮	0.349
味噌煮込みうどん	0.342
白菜と豚バラ肉の重ね蒸し	0.342
オニオンスープ	0.327

類似度算出結果より、肉じゃがと似ているレシピについては調理方法が類似しているものが多く、カレーライスと似ているものは食材が類似しており、粥はレシピ名や調理方法が類似しているものが上位となった。それぞれの結果について、ある程度レシピの類似性が考慮できているとも思えるが、ランキング最上位のレシピが最も類似しているとは言い難い状況である。

表 6 「粥」と似ているレシピ上位 10 件

レシピ名	類似度
中華粥	0.764
牛肉の柳川風	0.420
うなぎの柳川風	0.420
いか飯	0.414
赤飯	0.395
黒まめ	0.393
キムチ鍋	0.378
ひじきの煮物	0.364
肉ご飯	0.364
卵雑炊	0.364

したがって、まだリストアップできていないストップワードがノイズになっていることを含めて、特徴ベクトルの作成方法に改良の余地があると考えている。今後はストップワードリストの充実を含め、類似度算出方式の改良に取り組む。

4. 未知レシピへのメタデータ自動付与方式

本節では、未知レシピへのメタデータ自動付与方式について説明する。本研究では、メタデータ付与済みの 493 件のマスターレシピを用いて、未知レシピへメタデータ自動付与を目指す。自動付与の方法としては幾つか考えられる。

- 1) 類似度が最も高いマスターレシピのメタデータを、該当する未知レシピへコピーする。
- 2) 類似度が上位数件（例えば 5 件）のレシピ集合のメタデータの平均を、該当する未知レシピへコピーする。
- 3) 類似度が上位数件（例えば 5 件）のレシピ集合のメタデータに対して、類似度順位の重みを考慮したメタデータ平均値を、該当する未知レシピへコピーする。

ただし、2) および 3) については、メタデータが数値表現できる場合に限る。

今後は、メタデータの自動付与に関する実験を行い、その妥当性について検証すると共に、類似度算出方法も含めて、未知レシピへのメタデータ自動付与を精度良く行う手法の開発に取り組む。

5. おわりに

本稿では、既にメタデータが付与されているマスターレシピとの類似度分析に基づいた、未知レシピへのメタデータ自動付与方式の提案を行った。提案手法は、未知レシピへのメタデータ付与を考える際、この未知レシピと類似しているマスターレシピのメタデータを参考することで実現するものである。

なお、正確なメタデータの自動付与を行うためには、類似度算出の精度を上げる必要がある。したがって、今後の課題として、不必要なベクトルの次元の削除のためのストップワードの充実と、各ベクトルへの重みの検討を行う予定である。また、ユーザ評価による提案手法の妥当性の検証についても取り組む。

文 献

- [1] 森下幸俊, 中村富予, 気分により献立検索システムの検索軸の評価とレシピを活用した食品販売機能の市場ニーズの評価, 電子情報通信学会技術研究報告 Vol.112, No.75, DE2012-14, pp.79-84, 2012.
- [2] 荻米志帆乃, 藤井敦, 料理どうしの類似と組合せに基づく関連レシピ検索システム, 言語処理学会第 14 回年次大会発表論文集, pp.959-962, 2008.
- [3] 白井清昭, 大川寛志, アニメーション生成のための料理動作辞書の構築, 情報研報 自然言語処理研究会報告, Vol.2004 No.108, pp.123-128, 2004.
- [4] 志土地由香, 井手一郎, 高橋友也, 村瀬洋, 料理レシピマイニングにいる代替可能食材の発見, 電子情報通信学会論文誌 (料理を取り巻く情報メディア技術論文特集 Vol. J94-A No.7 pp.532-535, 2011)
- [5] 岩上将史, 伊藤孝行, ユーザの嗜好順序を利用した料理推薦システムの試作, 第 24 回人工知能学会全国大会, 1D2-4, 2010
- [6] <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>