

組み合わせを考慮した献立検索手法の提案

南谷 悠大[†] 前田 亮[‡]

[†]立命館大学情報理工学部 〒525-0058 滋賀県草津市野路東 1-1-1

E-mail: [†] is0272kv@ed.ritsumei.ac.jp [‡] amaeda@is.ritsumei.ac.jp

あらまし 本研究では、作りたい料理は決まっているが、献立全体を決められないユーザの献立作成支援をする手法を提案する。ユーザからの料理の入力をもとに、その料理に似た料理の類似度計算を行う。そして、その料理またはその料理に似た料理が含まれている献立の中から他の料理を出力する。ユーザは出力された料理の中から、自分が食べたいと思う料理を選択する。この行為を繰り返すことで、献立を作成する手法を提案する。

キーワード レシピ, 情報推薦, cookpad, doc2vec

1. はじめに

近年、Web を利用したレシピ検索サイトが普及している。cookpad に代表されるこれらのサイトの中には、単品の料理だけでなく、料理を献立単位でまとめて紹介しているものもある。

ユーザが一食分の献立を考える為にこれらのサイトを利用する際、食べたい料理の検索を行うことはもちろん、食べたい料理に合う付け合わせなどを自分で考え組み合わせることがある。単品料理が紹介されているサイトの場合、献立内の料理の組み合わせをユーザが考える必要がある。しかし、献立において食べたいものを思い浮かべ、組み合わせただけでは適切な組み合わせを作ることは難しく、誰もができることではない。また、献立単位で紹介されているサイトであっても、ユーザの嗜好にあった献立が必ずしも見つかるとは限らず、献立に好みではない料理が含まれている場合、その料理の代案を考える必要がある。しかし、料理の代案を考えるという行為も困難なことであり、料理の知識の少ないものには容易にできることではない。

そこで、本研究では、作りたい料理は決まっているが、献立全体を決められないユーザの献立作成支援をする手法を提案する。ユーザからの料理の入力をもとに、その料理に似た料理の類似度計算を行う。その類似度計算結果が上位のものを似た料理とする。そして、その料理またはその料理に似た料理が含まれている献立の中から他の料理を出力する。ユーザは出力された料理の中から、自分が食べたいと思う料理を選択する。この行為を繰り返すことで、献立を作成する手法を提案する。

2. 関連研究

大野らの研究[1]では、作りたい料理や献立の雰囲気は決まっているが、献立全体を決められないユーザの献立作成を支援することを目的とし、献立全体雰囲気や作りたい料理を起点とした料理推薦システムを開発している。献立作成支援システムの実現に向けて、ど

のような特徴を持つ料理同士が一つの献立を形成するのか、分析を行う。K-means 法を用いてレシピ情報をもとに料理のクラスタリングを行い、献立内に出現する料理レシピ共起頻度とクラスタリング結果を用いて料理クラスタ間の相性と雰囲気を分析した結果を報告している。

苅米らの研究[2]では、料理検索において、柔軟な検索をするために関連検索の手法を提案している。まず始めに、料理名や材料などの検索条件を入力すると条件に合致するレシピが出力される「初期検索」と呼ばれる検索を行う。次に、条件に合致したレシピがない場合には条件として入力した材料名の関連語を用いる「拡張検索」と呼ばれる検索を行う。さらに、出力された各レシピと何らかの関係がある別のレシピを検索する。この出力された各レシピと何らかの関係がある別のレシピを検索することを「関連検索」と呼ぶ。関連の種類として、「材料が類似」、「調理手順が類似」、「組み合わせの良さ」の3つを対象としている。「材料が類似」は、材料が類似している別のレシピを検索することで、ユーザが気が付かなかった別の料理を発見できる可能性がある。「調理手順が類似」は、調理手順テキストを形態素解析し、動詞を抽出する。出現する動詞の出現順序が似ているほど関連度が高くなる。「組み合わせの良さ」は、栄養バランス等を考慮して、複数の料理を組み合わせで検索する。レシピ中の材料名を食品群に分類する。

これら3つの条件を用いて料理レシピを対象とした関連検索システムを構築している。

3. 提案手法

図1に本研究で提案する献立推薦システムの概要図を示す。本システムはユーザからの料理の入力をもとに、その料理に似た料理を選出する。そして、その料理またはその料理に似た料理が含まれている献立の中から他の料理を出力する。ユーザは出力された料理の中から、自分が食べたいと思う料理を選択する。この行為を繰り返すことで献立を作成していく。料理の入

力は既存のレシピから選択する。

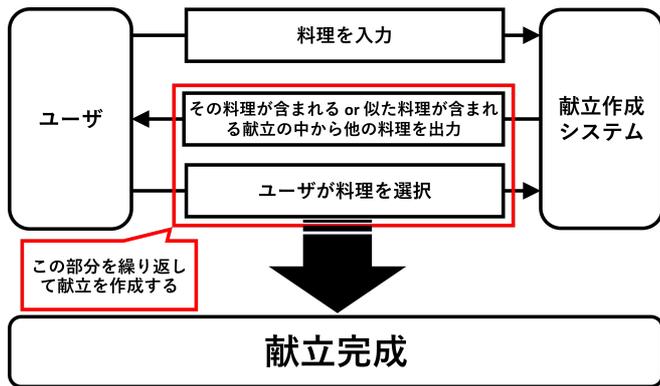


図 1 献立推薦システムの概要図

似た料理の算出には、Jey Han Lau ら[3]が提案した doc2vec とコサイン類似度を使用する。doc2vec は、任意の長さの文書をベクトル化する技術で、文やテキストに対して分散表現を獲得することができる。doc2vec をレシピのタイトルとレシピの手順に行い、それらの類似度を計算する。それらの類似度の合計が高いものを似た料理と定義する。doc2vec を使用する理由としては、doc2vec は他の類似度計算ツールとは違い、単語レベルではなく、文や文章といった任意の長さを扱えるところにある。これによって、レシピの手順をベクトル化の際に、単語の出現回数ではなく、単語の前後関係に注目して計算が行われるため、食材や調味料の入れるタイミングや、煮たり焼いたりなどの動作のタイミングを考慮することができる。doc2vec では、MeCab を用いて形態素解析を行い、名詞と動詞を用いて計算する。その理由としては、cookpad におけるレシピの手順は、投稿型のサイトであるため、書き手によって文章の形態が多様である。よってその問題を解決するために、名詞と動詞のみを用いることで、書き手の癖を減らすことができると考えられるからである。

また、コサイン類似度とは、ベクトル空間モデルにおいて、文書同士を比較する際に用いられる類似度計算手法の 1 つである。2 つの文書を A, B のコサイン類似度はとした場合、cos を式(1)に表す。この cos の値が 1 に近いほど類似度が高く、0 に近いほど類似度が低いことを表す。コサイン類似度も doc2vec と同じく MeCab を用いて形態素解析を行い、名詞と動詞を用いて計算する。

$$\cos\theta(A,B) = \frac{A \cdot B}{|A||B|} \dots (1)$$

ユーザーが入力した料理に対して類似度計算を行い、その料理とその料理と近いベクトルの料理を抽出する。それらの料理が含まれる献立を献立のレシピの情報が含まれているリストから選択し、それらの料理以外料理の部分を入力する。図 2 は例として、ユーザーがハン

バーグを食べたいと入力した場合、そのハンバーグが含まれる献立とハンバーグに似た料理が含まれている献立から、それらの料理以外の料理を出力している。そして、その出力の中から副菜を選択した場合までの流れを表している。

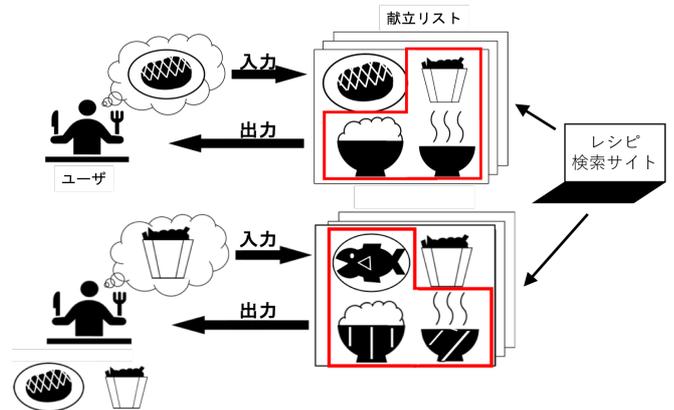


図 2 入出力の流れの例

4. 使用するデータセット

レシピのテキストデータなどのデジタルデータは、インターネット上で公開され始めている。cookpad は料理に関する研究を促進させるため、レシピに関するデータを研究者に提供している¹。提供しているデータは以下の通りである。レシピデータは、クックパッドで 2014 年 9 月 30 日以前に公開されたレシピに関するデータである。レシピのタイトルや概要、手順、つくれば(「作りましたフォトレポート」の略)、カテゴリなどのデータが含まれている。献立データは、クックパッドで 2014 年 9 月 30 日以前に公開された献立に関するデータである。献立のタイトルや献立に含まれるレシピ、各レシピが主菜か副菜かといったデータが含まれている。cookpad_data データベースには 12 個のテーブルが含まれており、6 個はレシピに関するテーブル、残りの 6 個は献立に関するテーブルになっている。そのうち使用するテーブルは以下の(a)~(c)の 3 つである。

- (a) recipes: レシピのタイトルや概要などの情報が含まれている。一つのレコードが一つのレシピに対応する(表 1)。

¹ 情報学研究データリポジトリ クックパッドデータセット

<http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html> (参照 2018/1/28)

表 1 recipes

カラム名	内容
id	レシピの ID
user_id	レシピ作者の ID
title	レシピのタイトル
description	レシピの概要
serving_for	レシピの分量
advice	レシピのコツ・ポイント
history	レシピの生い立ち
published_at	レシピの公開日

- (b) steps: レシピの手順の情報が含まれている. 一つのレコードがレシピ中の一つの手順に対応する. レシピが複数の手順から構成される場合, 各手順に対して一つのレコードが存在する (表 2).

表 2 steps

カラム名	内容
recipe_id	レシピの ID
Position	手順の位置
Memo	手順の内容

- (c) base_kondate_recipes: 献立のレシピの情報が含まれている. 一つのレコードが献立に属する一つのレシピと対応する. 献立が複数のレシピから構成される場合, 各レシピに対して一つのレコードが存在する (表 3).

表 3 base_kondate_recipes

カラム名	内容
base_kondate_id	献立の ID
recipe_id	レシピの ID
dish_type_id	レシピのタイプ

本実験では「recipes」のデータの中から, 無作為に抽出した 80 件の献立データに含まれている 200 件のレシピデータを使用する.

5. 評価実験

3 章に示した手法により, まず, 似た料理の選出を行う. doc2ve とコサイン類似度を料理のタイトルと手順のテキストに使用し, 類似度の合計値が閾値 0.7 以上の料理を似た料理とする. 本実験では, 最初の料理の入力を「むね肉やわらか! 揚げないチキン南蛮」とする. doc2vec を使用した場合の実験結果である, 料理のタイトルに対する結果の上位 5 品を表 4, 料理の手順のテキストに対する結果の上位 5 品を表 5 に示す.

表 4 doc2vec を用いた料理のタイトル

title	結果
あっさり美味☆黒酢で煮込む『豚の角煮』	0.28148
コレで決まり! 黄金比☆玉子丼のつゆ♡	0.23807
クリスマスツリーサラダ	0.22113
(^^)パプリカの肉詰め(^^)	0.21628
レンジ de チャーハン	0.18115

表 5 doc2vec を用いた料理の手順

title	結果
あっさり美味☆黒酢で煮込む『豚の角煮』	0.74861
レンジ de チャーハン	0.70550
胡桃とレーズンの全粒粉パン (H・B で)	0.70240
我が家の太巻き寿司★7種の具★恵方巻き★	0.70080
ロモ・サルタード (ペルーの牛肉野菜炒め)	0.68990

この結果により, doc2vec を用いた場合「むね肉やわらか! 揚げないチキン南蛮」は, 料理のタイトル, 手順, ともに, 「あっさり美味☆黒酢で煮込む『豚の角煮』」と近いという結果になる. しかし, 料理のタイトルは全体的に全ての結果の数値が低いため, 類似度がうまく算出できていないことがわかる. また, 料理の手順では, 最も類似度が高い料理は「0.74861」と大きな数値を出力している. そこで, それぞれの料理の手順を比較する. その内容を表 6, 7 に示す.

表 6 むね肉やわらか! 揚げないチキン南蛮の手順

むね肉やわらか! 揚げないチキン南蛮
鶏肉は脂を取り除き, 均一の厚さになるよう食べやすい大きさに削ぎ切りする. 筋や厚い部分を包丁で叩く. 下味のんにくは摺りおろし [1] の鶏肉と○の材料をビニール袋に入れ, よく揉み込み 10 分程置く. ★を合わせておく. 茹で卵, 玉葱, ピクルス, パセリは微塵切り. ☆を全て合わせてタルタルソースを作る. [2] の鶏肉に片栗粉を全体に薄くまぶし, フライパンに多めの油を熱して両面をカリッと焼く. 余分な油は拭き取っておく. 中まで火が通ったら, 合わせておいた南蛮酢を回しかけ全体に絡める. お皿に盛り付け, [3] のタルタルソースをたっぷりのせて出来上がり♪お弁当には冷めるまで待ってね!

表 7 あっさり美味☆黒酢で煮込む『豚の角煮』の結果

あっさり美味☆黒酢で煮込む『豚の角煮』
豚バラ肉は、脂身の面からフライパンでじっくり 10～20 分程度焼きます（弱火）。大量に出る脂は除きます。●の調味料と香味野菜を合わせた鍋に豚肉と香味野菜を入れ、かぶる程度の水を加えて煮立てます。煮たら灰汁をひき、すうっと竹串が通るまでとろ火で煮ます。柔らかくなった肉は一旦取り出し粗熱がとれたら、冷蔵庫に入れて冷やしておきます。その間、煮汁を好みの濃度になるようとろっと煮詰めます。尚、煮汁を一旦冷やして脂を除くとヘルシーです。【付け合わせ】蓮根とアスパラガスはごま油でソテーし、軽く塩で味つけます。【仕上げ】冷えた肉をスライスし、煮汁に戻し入れて軽く温め、付け合わせと一緒に盛り付けたら完成。好みで白髪ねぎを添えて。【以下圧力鍋用のレシピ】ブロック肉は鍋に入る大きさにカットし、●の材料、香味野菜、かぶるくらいの水を加えます。ふたをセットし強火にかけ重りがふれたら、適当な火力に切り替えて加圧時間 20 分間、火を消し圧がなくなるまで置きます。ふたをオープン！箸がすうんと入る程柔らかくなっているはず。この肉を取り出して一旦冷やします。残りの煮汁の脂をすくい、強火で半分以下の濃さになるまで煮詰めます。適当な大きさに切った肉を鍋に戻し入れ、軽く温めて完成です。

これら二つの料理を見比べた結果、この二つの料理の共通点を見つけることができなかった。つまり、この 2 つが似た料理であるとするのは難しい。よって doc2vec では似た料理を算出することはこの方法では難しい。

次に、コサイン類似度を使用した場合の実験結果である、料理のタイトルに対する結果の上位 5 品を表 8、料理の手順のテキストに対する結果の上位 5 品を表 9 に示す。

表 8 コサイン類似度を用いた料理のタイトル

title	結果
大人気☆揚げないチキン南蛮	0.71034
さわらの塩焼き	0.27441
鶏むねの甘酢煮	0.23687
蒸し揚げパンケーキ	0.20239
*簡単*カレー風味のチキンカツレツ*	0.18950

表 9 コサイン類似度を用いた料理の手順

title	結果
*失敗なし*ホワイトソース*	0.21290
ヘルシー*鶏肉と三つ葉の梅肉和え	0.20778
鶏肉とたけのこの煮物	0.18320
大人気☆揚げないチキン南蛮	0.15595
醤油麺でこっくりブリの照り焼き	0.14853

この結果により、コサイン類似度を用いた場合は、

doc2vec に反して、料理のタイトルに対して「大人気☆揚げないチキン南蛮」が「0.71034」と高い数値を出している。タイトル名から、どちらもチキン南蛮であるので、この料理は似た料理であると言える。また調理手順に対しては、全体として、類似度の数値は低い。タイトルで高い類似度を出していた「揚げない！チキン南蛮」も 4 番目に高いが、「0.15595」と数値自体は非常に低い。

これらの結果からコサイン類似度を用いた方が有用であると判断した。この実験において、料理のタイトルと手順の数値を足し合わせた結果を用いて似た料理の選出を行うため、足し合わせた結果を表 10 に示す。

表 10 類似度計算結果の合計

title	結果
大人気☆揚げないチキン南蛮	0.86373
*簡単*カレー風味のチキンカツレツ*	0.38537
鶏むねの甘酢煮	0.31569
さわらの塩焼き	0.29909
蒸し揚げパンケーキ	0.28771

この結果により、閾値 0.7 以上である似た料理は「大人気☆揚げないチキン南蛮」の 1 つだけである。よってこの料理が含まれている献立から他の料理と入力に使用した「むね肉やわらか！揚げないチキン南蛮」が含まれる献立の他の料理を出力する。その結果を表 11 に示す。

表 11 他の料理の出力

title
干し梅と蓮根の炊き込みごはん
かぼちゃの胡麻和え
★我が家の定番★おから&豆腐のサラダ
我が家の味噌汁

次に、この結果の中から自分の食べたいものを選択する。今回は、「我が家の味噌汁」を選択した。これらの作業を繰り返すことにより献立を組み立てていく。一例ではあるが、この手法を用いて作成した献立を表 12 に示す。

表 12 完成した献立の例

title
むね肉やわらか！揚げないチキン南蛮
我が家の味噌汁
ゴーヤの詰めチーズ入り♪
いんげんと人参のマヨごまサラダ

6. 考察

5 章の実験結果を元に考察する。1 品のみの選出ではあったが、似た料理を選出することに成功した。その後の手順では、ユーザの好みによって料理を選択していくことになるが、主食、副菜、サラダ、汁物とバラ

ンスの良い献立を作成することができた。

似た料理の選出において、doc2vec を使用した場合、レシピのタイトル、手順ともに似た料理を選出することができなかった。レシピの手順に対して高い類似度を算出していたが、レシピの内容を確認した結果、似ているとは言えなかった。これは、doc2vec は長めの文章には強いが短い文章には弱いと言われている。今回使用した、レシピのタイトル、手順はともに長い文章とは言えない。これが doc2vec を使用した時に、正確に似た料理を高い類似度で算出できなかった原因の一つではないかと考えられる。

コサイン類似度を使用した場合では、料理のタイトルの類似度計算結果では、同じチキン南蛮である、「大人気☆揚げないチキン南蛮」が上位にきているが、料理のテキストの類似度計算結果では、1位が「*失敗なし*ホワイトソース*」となっており、似た料理がうまく選出できているとはいえない。これは、「*失敗なし*ホワイトソース*」と「むね肉やわらか！揚げないチキン南蛮」の手順のテキストの中でタルタルソースを作成する部分によって類似度が高くなっていると考えられる。同じチキン南蛮であるので、むね肉やわらか！揚げないチキン南蛮」と「大人気☆揚げないチキン南蛮」では手順も類似していると期待したが、投稿型サイトであるということもあり、同じ料理でも、書き手によって表現に違いがあることが問題に挙げられる。

本実験で、最初に入力した料理である「むね肉やわらか！揚げないチキン南蛮」が含まれる本来の献立の他の料理は、「◆サクサク揚げない かぼちゃコロッケ」、「じゃがいもゴロゴロコンソメスープ♪」、「お箸が進む♪ごま油でほうれん草のおひたし」の4品である。作成した献立と比較した結果、どちらの献立も、副菜2品と汁物1品であり、本来の「むね肉やわらか！揚げないチキン南蛮」が含まれる献立と遜色ない適切な献立ではないかと考えられる。

7. おわりに

本稿では献立作成システムの実現に向けて、料理のタイトルと手順を元に類似度計算を行い、似たような料理を含む献立に含まれる他の料理を出力する方法を提案した。使用者の選択もあるが、作成した献立は適切な組み合わせの献立を作成することができた。使用しているデータセットが投稿型のサイトであるため、書き手の癖をより減らす手順が必要になる。

また、この他にも料理が似ていると判断できる部分を探し、より似た料理の選出の精度を高めていきたいと考えている。その例としては、レシピに使用されている材料である。今回、手順の中で材料の名前が出現すると考えていたので、使用しなかったのだが、レシ

ピの中には「●塩」のように記号をつけているものがある。このような材料を含んでいるレシピの手順では「次に●のつく材料を入れてください」のように書かれている場合が多い。このように表記されてしまうと現段階の研究では材料に何が含まれているか判別することができていなかった。よって、材料も似た料理の選出に使用できるのではないかと考えられる。

また、献立を作成するにあたり、主菜や副菜などの判別も行うことができれば、よりユーザが見やすい出力を行うことができのではないかと考える。

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」を利用した。ここに記して感謝の意を表する。

参 考 文 献

- [1] 大野礼儀, 福原知宏, 山田剛一, 増田英考: 献立の雰囲気考慮した料理推薦システムの提案, 2015年度 人工知能学会全国大会(第29回), pp1-4(2015)
- [2] 苅米志保乃, 藤井敦: 料理どうしの類似と組み合わ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, pp.959-962 (2008)
- [3] Jey Han Lau, Timothy Baldwin, An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation, 1st Workshop on Representation Learning for NLP(2016)
- [4] 情報学研究データリポジトリ クックパッドデータセット
<http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html> (参照 2018-1-28)