

レシピ頻度特徴量と LexRank に基づくアレンジ抽出手法の評価

大仁田龍也[†] 北山 大輔[†]

[†] 工学院大学工学研究科情報学専攻 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: [†]fem19002@ns.kogakuin.ac.jp, ^{††}kitayama@cc.kogakuin.ac.jp

あらまし 我々は、材料や手順を用いたレシピのアレンジ抽出手法に取り組んでいる。これまでに、ある料理の標準的な調理法が記載された基本レシピに対する差異を料理のアレンジとして抽出する手法について研究を行ってきた。この手法は、ある料理に関するレシピを1手順ごとに分解し、LexRank を用いて手順の重要度を判断し、その料理の基本レシピと抽出対象レシピの手順の対応付けを行うことで、対応関係と重要度を用いて、アレンジを含む手順を特定する。本稿では、システムの評価のため、正解データを作成後、重要度計算について比較手法を用意し、比較実験を行なった。また、対応付についても成果データを作成し、システムの出力と正解データの一致率を算出することでシステムの精度を評価した。評価の結果、重要度計算については平均適合率を用いて比較を行なった結果、提案手法の RF_{pn} よりも比較手法である TF-IDF を用いた場合の方が良い数値を得られる結果となってしまった。また、対応付の評価については、正解データとの一致率を適合率で計算した結果、約半分が一致する結果となった。こちらについては、手順に対する処理などに問題が残っているにもかかわらず、高い数値を得られたため手法として有効であると判断した。

キーワード 料理レシピ, アレンジ抽出, LexRank

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をはじめとする様々な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。投稿されるレシピは同一の料理に対するものも多く、それらのレシピは基本的に忠実なものからアレンジを多分に含むものまでさまざまである。また、投稿者によって材料の表記や手順の順序関係が異なっている場合もある。

我々はそれらのレシピから料理のアレンジを抽出することを目指す。ここで、料理のアレンジは、その料理の基本的なレシピを1つ決定すれば、そのレシピと他のレシピとの差異を求めることで自動的に抽出できると考えられる。これを実現するには基本レシピを自動的に抽出できるシステムが必要である。我々はこれまでの研究でこのシステムの構築を行なった[1][2][3]。このシステムおよび本稿で扱う基本レシピとは、対象料理に関するレシピ集合の、共通した一定の手順のみで構成されたレシピを指している。

簡単に、基本レシピ判定のアルゴリズムを説明する。まず、対象料理のレシピ中に書かれた手順を1文書として扱う。この文書を、Doc2Vec による分散表現、もしくは食材と調理動作を抜き出した文字列で表現する。次に、コサイン類似度もしくは、編集距離によって類似する文書間にエッジを構築する。その結果レシピをノード、類似関係をエッジとしたグラフを構築し、LexRank を用いることで、対象料理を代表するレシピを抽出し、基本レシピとする。このアルゴリズムは、食材にも適

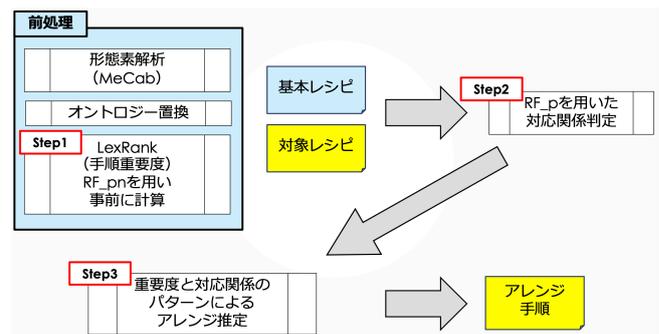


図1 アレンジ抽出手法

応することが可能であり、対象料理のレシピを1文書とし、食材集合をもつものとする。次にjaccard係数によって類似度を判断し、エッジを構築する。これにより、LexRankによって、食材に基づく基本レシピを抽出することも可能である。手順によるLexRank値と食材によるLexRank値を合算することで、両方を加味した基本レシピとすることもできる。このアルゴリズムによる基本レシピの抽出精度については、先行研究[1]において他手法と比較し、有効性を確認している。

本稿では、基本レシピ抽出アルゴリズムなどによって、基本レシピが抽出されたことを前提に、基本レシピとアレンジ抽出対象レシピ間の差異を用いて、アレンジを含む手順を特定する手法を提案する。具体的な手順を説明していく。まず、手法の概要を図1に示す。

- (1) 対象料理に関するレシピを1手順ごとに分解し、LexRankを用いて手順の重要度を判断する。
- (2) その料理の基本レシピと抽出対象レシピの手順の対応付けを行う。

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

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

(3) 対応関係と重要度のパターンを用いて、アレンジを含む手順を特定する。

(1) に関して、基本レシピ抽出の際には、レシピに書かれた複数の手順をまとめて1文書として扱っていたのに対し、レシピに書かれたそれぞれの手順ひとつひとつを1文書として扱う。すなわち、5手順からなるレシピから5文書生成される。

(1) および (2) において、手順はレシピ頻度特徴量によって重み付けした文書ベクトルで表現する。詳細は2節で述べるが、レシピ頻度は単語が対象料理のレシピでどれだけ出現しているかを表す指標であり、高いほど対象料理にとって一般的な単語であることを表現する。

本稿では、提案する手法の重要度の計算について正解データと比較手法を準備し有効性を評価する。また、手順同士の対応付けについて、正解データとの一致率を算出することで手法の評価を行う。

本稿の構成を述べる。2章でレシピ頻度特徴量の説明、3章でアレンジ抽出のアルゴリズムを述べた後、4章で実験について述べる。5章で関連研究の紹介を行う。最後に6章で本稿をまとめる。

2 レシピ頻度特徴量

2.1 正負レシピ頻度特徴量

アレンジ抽出の議論に入る前に、本稿で用いるレシピ頻度特徴量について説明する。まず、レシピ頻度は単語が対象料理のレシピでどれだけ出現しているかを表す指標であり、高いほど対象料理にとって一般的な単語であることを表現する。文書の単語特徴量で用いられる TF-IDF の DF (文書頻度) に相当する特徴量である。

本稿では2種類のレシピ頻度特徴量を定義する。まず1つ目は、手順の重要度を評価するためのレシピ頻度特徴量であり、正負レシピ頻度特徴量 (RF_{pn}) と呼ぶ。この RF_{pn} は、-1 から 1 までの値をとり、以下の式で定義される。

$$RF_{pn}(t) = 1 - 2 \times \frac{IRF(t)}{\max IRF} \quad (1)$$

$$IRF(t) = \log\left(\frac{R}{rf(t)}\right) \quad (2)$$

IRF(t) はレシピの総数を単語 t の出現数で割り、log をとったものである。max IRF は IRF(t) の最大値を取り、RF_{pn}(t) は各単語の IRF を 0 から 1 に収まるように正規化した後、2倍して 1 から引くことで -1 から 1 までの値をとるようにしている。

RF_{pn} では、一般的な単語と特徴的な単語が、正方向ならびに負方向に大きな値をとる。そのため、コサイン類似度を計算すると、共通する一般的な単語を含むが、共通しない特徴的な単語が含まれると、類似度が低く評価されることになる。この性質により、その料理で一般的な単語のみを含む手順は、多数の手順と類似度が高くなり、特徴的な単語、例えばアレンジを表す単語を含む手順は類似度が低くなるのが期待できる。

2.2 正レシピ頻度特徴量

2つ目の特徴量は、手順の対応関係を評価するための特徴量であり、正レシピ頻度特徴量 (RF_p) と呼ぶ。この RF_p は、0 から 1 までの値をとり、以下の式で定義される。

$$RF_p(t) = 1 - \frac{IRF(t)}{\max IRF} \quad (3)$$

RF_p(t) は RF_{pn}(t) と同様の IRF(t) を用いて各単語の IRF を 0 から 1 に収まるように正規化した後、1 から引いたものである。

RF_p では、一般的な単語が、正方向に大きな値をとり、特徴的な単語は 0 に近い値をとる。そのため、コサイン類似度を計算すると、共通する一般的な単語を含んでいれば、類似度が高く評価されることになる。この性質により、その料理での基本的な食材、調理動作が一致している手順を対応付けすることが期待できる。

3 レシピ頻度特徴量を用いたアレンジ抽出

3.1 概要

料理のアレンジは、その料理の基本的なレシピを1つ決定し、そのレシピと他のレシピとの差異を求めることで自動的に抽出できると考えられる。基本レシピは、先行研究の基本レシピ抽出手法などを使うことで決定可能であり、本稿では何らかの手段で決定されていることを前提とする。差異がアレンジであるかどうかを判断するために、2つの観点が必要となる。

- (1) その手順がどれだけその料理にとって一般的であるか
- (2) 基本レシピに含まれる手順であるか

まず、その料理にとって一般的ではない手順であるならば、アレンジを含む手順である可能性は高いと考えられる。また、基本レシピに含まれない手順はアレンジを含む可能性が高いと考えられる。更に、抽出対象レシピでは削除された基本レシピの手順も、それを行わないというアレンジであると考えられる。

(1) を判断するために、一般的な手順としての重要度を LexRank を用いて算出する。次に (2) を判断するために、基本レシピの各手順と抽出対象レシピの各手順の対応関係を判定する。最後に、対応関係と重要度を用いることで、アレンジを含む手順を抽出する。

各レシピに対して行う前処理について説明する。まず、レシピの各手順に対し形態素解析を行なった後、単語をオントロジーで置き換えている。本研究でのオントロジーは、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」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 [4] 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 [5] の研究を基にした基本料理知識ベースを用いている。

3.2 LexRank に基づく手順重要度算出

一般的な手順としての重要度を求めるために LexRank を用いて算出する。LexRank とは、Erkan ら [6] が提案した PageRank を応用したテキスト要約手法である。

PageRank とは Web ページ間のハイパーリンク関係を利用し、ページ毎の典型度を測るアルゴリズムである。Web のハイパーリンク構造は、Web ページをノード、ハイパーリンクをエッジとした有向グラフとして表現される。この有向グラフに基づき PageRank を計算し、そのスコアによって各 Web ページの順位付けを行う。ここで、PageRank の概念では、「他の多くの Web ページからリンクを張られている Web ページは重要であり、そのような重要な Web ページからリンクを張られている Web ページも重要である」と考えられている。そのため、ある Web ページの PageRank は、その Web ページを指す他の全てのページがもつ PageRank の総和と考えられ、PageRank 値を未知数とする連立一次方程式と捉えられる。各 Web ページへの嗜好性を一様に捉えるとき、初期値として全ての Web ページに対して PageRank を与え、上記の考えのもと反復的に計算していくことで、最終的に収束し、全ての Web ページの PageRank を計算できる。

このような PageRank に対し、LexRank は対象テキスト内の文をノードとし、文間の類似度を重みとしたエッジでつなぎ、グラフ表現で表す。このグラフ表現における固有ベクトル中心性の概念に基づいて文の典型度を計算する。LexRank は、コサイン類似度に基づく連結性行列を隣接行列で表す。この行列の各値を定めた閾値と比較し 1 または 0 に置き換え、行列にべき乗法を行うことで典型度を計算するアルゴリズムとなっている。

本稿では、文の類似度を RF_{pn} を用いたベクトルによるコサイン類似度で算出する。RF_{pn} は 2.1 節で説明した通り、対象料理において、一般的な単語と特徴的な単語に正および負方向に大きな特徴量を与える。このため、LexRank においては、一般的な単語のみで構成された手順同士にはエッジが貼られ、特徴的な単語を含む手順においてはエッジが貼られにくくなることになる。

そのため、このように構築したグラフによる LexRank では、一般的な単語のみで構成された手順が高い LexRank 値となり、特徴的な単語を含む手順は低い LexRank 値となると考えられる。この事により、一般的な手順としての重要度を与える。

3.3 RF_p を用いた手順の対応付け

基本レシピと抽出対象レシピの各手順の対応付けを行う。手順の対応付けとしては、アレンジの有無に関わらず対応付けできることが望ましい。例えば、チキンライスという料理において「玉ねぎを炒めた後、鶏肉を加えて炒める」という手順と「玉ねぎを炒めた後、豚肉を加えて炒める」であれば、「豚肉」はそのレシピにおいて特徴的な単語であるが、その有無に関わらず対応する手順と判断できるのが望ましい。

ここでの手順は、LexRank による重要度計算とは異なり、RF_p を用いたベクトルで表現する。RF_p は 2.2 節で説明したとおり、その料理にとって一般的な単語であれば高い正の特徴量を持ち、特定のレシピでしか使われない単語であれば、0 に近づく。このことにより、アレンジを含むかどうかのコサイン類似度を与える影響を小さくできる。

対応関係判定の概要を図 2 に示す。まず、基本レシピと抽出

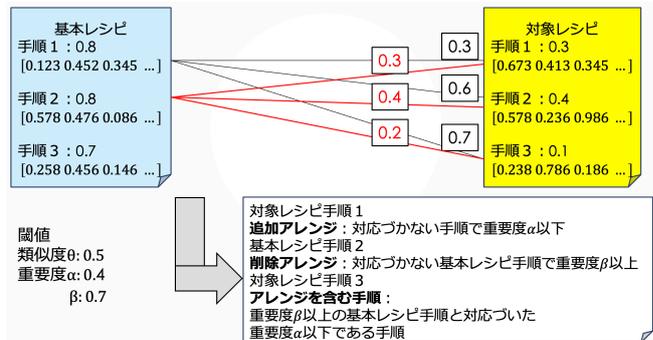


図 2 対応関係判定

対象レシピの各手順間の類似度を求める。基本レシピの手順から見て、類似度が最大かつ閾値 θ 以上となる抽出対象レシピの手順を対応する手順とする。

3.4 重要度と対応関係を用いたアレンジの判定

LexRank に基づく手順重要度算出と RF_p を用いた手順の対応関係を用いて、アレンジの判定を行う。

以下のパターンに当てはまる手順を、アレンジを含む手順として抽出する。

追加手順 どの基本レシピとも対応つかない手順でかつ重要度が閾値 α 以下である手順

削除手順 どの抽出対象レシピの手順とも対応つかない基本レシピの手順でかつ重要度が閾値 β 以上である手順

アレンジを含む手順 重要度が閾値 β 以上の基本レシピの手順と対応ついた重要度が閾値 α 以下である手順

追加手順は、その手順を追加すること自体がアレンジである可能性が高い手順であり、削除手順は省略や時短などの理由から省かれた手順であり、削除することでアレンジとなる可能性がある手順である。アレンジを含む手順は、基本レシピ中にある手順であるがその工程にアレンジが含まれる可能性のある手順である。

4 評価実験:重要度計算

提案手法の有効性を確認するため、システムの要となる部分である「重要度計算」部分について、先行研究である重田らの研究を再現し、比較手法とした。手法再現後、クラウドソーシングサービスの CrowdWorks を利用して作成した正解データを用いて、それぞれの手法で出力された結果と正解データとの一致率を平均適合率により算出し、手法の比較を行った。

4.1 使用データ

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」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。ここから、4 種類の料理を実験データとして選出した。各料理のレシピの件数を、表 1 にまとめる。後述する対応付の評価実験においても同様のデータを用いている。

表 1 実験データ

料理名	件数
親子丼	2172
豚肉の生姜焼き	297
酢豚	2267
麻婆豆腐	3223

表 2 平均適合率による比較結果

	提案手法 (RF_pn)	比較手法 (TF-IDF)
親子丼	0.50	0.62
豚肉の生姜焼き	0.37	0.40
酢豚	0.39	0.44
麻婆豆腐	0.55	0.61
平均	0.46	0.52

4.2 正解データ作成

各料理の手順を 200 件ずつランダムに抽出し、これを 20 件ずつの 10 グループに分けた。1 グループを 1 タスクとし、それぞれのタスクについて約 40 人に手順を見てもらい、その手順が対象の料理にとって必要であるかどうかを回答してもらった。回答を集計し、被験者の 6 割にあたる 24 名以上が必要だと判断した手順を正解データとした。

4.3 結果

正解データとのそれぞれの一致率を表 2 に示す。結果から、TF-IDF を用いた場合の方が数値が良いことがわかる。よって、本システムにおいては TF-IDF を用いた手法の方が適しているといえることとなった。

4.4 考察

まず、豚肉の生姜焼きにおいて、正解データに用いた手順のうちそれぞれの手法で上位に出力された手順を表 3、4 に示す。下線は正解データと一致していることを示す。傾向として、提案手法での出力結果に比べ、比較手法での出力結果は長い文章も多く上位に出力されている。手順の中身に注目すると、「フライパン」が含まれている手順が多く存在しており、これらの手順はどれも正解データと一致している。ここから、「フライパン」のような調理器具に関係する値が RF_pn と TF-IDF の差異であり、結果の差につながっていると考えられる。また、TF-IDF 値の高い単語はその性質からアレンジに近いものを抽出するのに対し、提案手法である RF_pn の値が高い単語は典型的な、基本に近いものを抽出する。そのため、このような性質が関係していることも考えられる。よって、一度 RF_pn と TF-IDF によって単語に与えられる数値にどのような差異があるのかを検証する必要があると考えられる。

5 評価実験:手順同士の対応付に関する評価実験

提案手法の有効性を確認するため、システムの要となる部分である「手順同士の対応付」部分について、評価実験を行なった。まず、重要度計算の実験と同様に CrowdWorks を利用して作成した正解データを作成した。正解データ作成後、提案手

表 3 提案手法 (RF_pn) による出力上位

順位	手順
1	豚肉は、筋切りをする
2	豚肉を焼きモヤシも焼く
3	そこに肉と葱をつける
4	人参はピーラーでスライスする
5	キャベツは千切りにする
6	<u>豚肉は、両面色付くまで</u>
7	<u>キャベツは、千切りにしておく</u>
8	ボールや容器などに、漬け汁の調味料を入れる
9	長ねぎはみじん切り (適当な大きさでも w) にする
10	<u>玉ねぎは適当にくし切りにする</u>

表 4 比較手法 (TF-IDF) による出力上位

順位	手順
1	豚肉を焼きモヤシも焼く
2	豚肉は、筋切りをする
3	フライパンにオリーブオイルを敷き、生姜、豚肉を入れ、炒める
4	キャベツは千切りにする
5	<u>豚肉に片栗粉をかるーくまぶして、フライパンに油を入れ、豚肉を焼く</u>
6	フライパンに油をしき、豚肉を焼く
7	豚肉をフライパンに入れ両面軽く焼いたら中華野菜ミックスを入れ炒める
8	<u>フライパンを熱してサラダ油を入れ、一度弱火にしておく</u>
9	玉葱を入れたままのフライパンに豚肉を入れ、中火で焼く
10	<u>キャベツは、千切りにしておく</u>

法で出力された結果と正解データとの一致率により手法の比較を行った。

5.1 正解データ作成

基本レシピと比較レシピを 1 組とし、比較レシピだけを入れ替えたものを 5 組用意した。1 組を 1 タスクとし、1 タスクについて約 10 人ほどに基本レシピ側の手順から見たときに、対応付くと思う比較レシピの手順を回答してもらった。回答後、対応付くと判断された票数の集計を行い、それぞれの組において 6 割にあたる 6 名が対応付くとして判断した手順同士の組を正解データとした。

5.2 結果

正解データとの一致率を適合率で計算したものを表 5 に示す。結果から、どの料理においても約半分の一致率となっており、システムが有効であることは確認できたが、手法として最適であると断言はできない結果となった。

5.3 考察

今回実験を行うにあたり、手順内の省略や記号を考慮せず、前処理段階のオントロジー置換時に削除している。しかし、そのような状態であっても、約半分の出力が正解データと一致する結果となった。今後、そのような省略などの問題に対し対処を行なっていくことで精度の向上を行えると考えている。

表 5 適合率による正解データとの一致率

	適合率
親子丼	0.50
豚肉の生姜焼き	0.46
酢豚	0.54
麻婆豆腐	0.53
平均	0.50

また、類似しているかの判断に用いている閾値について、料理ごとに最適な値を与えるべきであるが、これを満たすアルゴリズムの考案に至っていない。そのため、今回の実験では全料理について閾値 0.5 で統一して実験を行なっている。この点も、今後精度の向上を目指す上で考慮すべき点だと考えてる。

6 関連研究

6.1 レシピの基本手順生成に関する研究

レシピの基本手順生成に関する研究に関する研究は様々に行われている。[7][8][9] 重田ら [7] は、ある料理に関するレシピを手順レベルに分解し、その中で典型的な手順を取り出す手法を提案している。本研究で行う、LexRank に基づく手順重要度算出は、重田らの手法と基本的には同じであるが、手順の特徴ベクトルを生成する方法が異なる。重田らは一般的な文書特徴量である TF-IDF を用いているが、本研究では、典型的な手順とアレンジを含む手順を分離するために、一般的な語に高い特徴量、特徴的な語に低い特徴量を与える正負レシピ頻度特徴量を用いている点で異なる。

瀧本ら [8] は、レシピ間の手順から 2 つの操作に対して、出現位置を考慮したペアアライメントを行うことで操作を対応付け、施設配置問題の考えに基づいて基本操作の選択を行い、アライメントの結果を利用して操作の順序付けを行う、という手順で基本手順の抽出を行っている。アライメントの考え方は、本研究で行う手順の対応付けに対応し、基本操作の選択は手順の重要度の計算と対応する手法となっている。本研究ではアレンジの抽出を目的とするため、手順の重要度の計算時に、アレンジを含むほど類似度が低くなるようなベクトルを設計している点でアプローチが異なる。

これらの、類似手法との比較実験は今後の課題である。

6.2 レシピのアレンジ抽出に関する研究

崔ら [10] は、特定の料理の共通の食材の組合せを見つけ定番レシピを定義し、そのレシピとの差分を考慮した料理レシピ検索手法を提案している。料理レシピの分類、食材名の表記ゆれの統一を料理オントロジーを用いて行い、アソシエーション分析を利用して、各メニューの食材の相関ルールを検出している。相関ルールをもとに抽出された食材のみを含むレシピを定番レシピとしている。この研究では、定番レシピを求めることで食材の差分から非定番レシピを定め、アンケートにより比較を行っていたが結果が芳しくなかった。この研究での非定番レシピは、本研究でのアレンジを含むレシピの概念に近いと考えられる。本研究では、食材ではなく、手順レベルでのアレンジを

抽出する点においてアプローチが異なる。

6.3 レシピ情報処理に関する研究

レシピの前処理や、応用に関する研究を紹介する。

前処理としては、久保ら [11] が、複数のレシピ投稿サイトのある特定の料理のレシピと、そのレシピとほとんど変わらない重複したレシピの発見を行う研究を行っている。重複レシピを調理内容の一部が変更されているもの、料理の目的が異なるにも関わらず目的に応じた調理内容に変化していないものとし、文字 n-gram と Jaccard 係数を用いて類似度を計算後、高い順にランキングして上位 200 件のペアが重複レシピかどうかを人手で判定することで検証を行っていた。また、同時に調理内容の一部の代替を考慮する手法も同時に行い検証している。このような重複レシピを予め取り除くことや、記号で書かれた手順等の置き換えを実現することで、我々の提案手法の精度はより良くなることが考えられる。

応用に関する研究として、以下の研究が挙げられる。福本ら [12] は、食材の典型度と食品群を考慮したレシピ間類似度の算出手法を提案している。食品群の分類とグラム変換表を事前に準備し、食材の典型度を料理における食材単体での出現割合と、その食材が属する食品群についての出現割合を求めて合算することで算出している。レシピの各食材を要素とするベクトルと各食材の属する食品群を食材の典型度で重み付けたものを要素とするベクトルに変換し、2 種類のベクトル間のコサイン類似度を合算することでレシピ間類似度としている。

花井ら [13] は、食材と調味料、食材の希少度をもとに類似レシピのクラスタリングをする研究を行っている。レシピごとに主食材と主調味料の抽出を行うことで類似レシピを抽出する。1 段階目のクラスタリングで、料理名、調味料名ごとにレシピを分類し、2 段階目のクラスタリングにて、食材名、調味料名、主食材、主調味料、材料の希少度、単語の出現場所の重みを考慮し再びクラスタリングを行いレシピのクラスタの作成を行っている。花井らは他にも、料理名、調理法、食材名、調味料を用いたクラスタリング手法 [14] なども提案している。

このような類似度による検索結果や、クラスタリング結果に対し、我々の手法により抽出したアレンジ手順を検索結果中に追加で表示することで、ユーザが結果中からレシピを選択しやすくなるということが考えられる。このように、アレンジ抽出は、他の手法と独立に実施することが可能であり、レシピアプリケーションに追加の機能をもたらすものである。

7 まとめ

本研究では、料理のアレンジは、その料理の基本的なレシピを 1 つ決定すれば、そのレシピと他のレシピとの差異を求めることで自動的に抽出できる、という考えのもとアレンジ抽出を試みた。これまでの研究で構築した基本レシピ検索システムによって、基本レシピが抽出されたことを前提に、正負レシピ頻度特徴量 RF_{-pn} を定義し、レシピの手順単位の重要度を LexRank を用いて決定し、正レシピ頻度特徴量 RF_p を定義

し、手順ごとにベクトルを作成することで手順同士のコサイン類似度を求め、基本レシピとアレンジ抽出対象レシピを手順単位位で対応付けることで、レシピ間の差異を明確にし、アレンジを含む手順を特定する手法を提案した。

システムの評価のため、正解データを作成後、重要度計算について比較手法を用意し、比較実験を行なった。また、対応付についても成果データを作成し、システムの出力と正解データの一一致率を算出することでシステムの精度を評価した。

評価の結果、重要度計算については平均適合率を用いて比較を行なった結果、提案手法の RF_pn よりも比較手法である TF-IDF を用いた場合の方が良い数値を得られる結果となってしまった。また、対応付の評価については、正解データとの一致率を適合率で計算した結果、約半分が一致する結果となった。こちらについては、手順に対する処理などに問題が残っているにも関わらず、高い数値を得られたため手法として有効であると判断した。

今後の課題として、RF_pn と TF-IDF が単語に与える数値にどのような差異や傾向があるのかを検証することを考えている。また、手順の省略などの問題に対し対処を行なっていくことを考えている。また、対応付の判断に用いる閾値について、料理ごとに最適な値を与えるアルゴリズムを考案しなければならぬこと、アレンジの抽出単位の細分化などについても検討することを考えている。

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] 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. 第 12 回データ工学と情報マネジメントに関するフォーラム, pp. B7-2, 2020.
- [2] 大仁田龍也, 北山大輔. レシピ間類似度を用いた lexrank による基本レシピの検索. 第 11 回データ工学と情報マネジメントに関するフォーラム, pp. C2-5, 2019.
- [3] 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.
- [4] 土居洋子, 辻田美穂, 難波英嗣, 竹澤寿幸, 角谷和俊. 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. 470, pp. 37-42, 2014.
- [5] 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.
- [6] Günes Erkan and Dragomir R Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, Vol. 22, pp. 457-479, 2004.
- [7] 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. 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, pp. B2-1, 2017.
- [8] 瀧本洋喜, 笹野遼平, 高村大也, 奥村学. 施設配置問題に基づく同一料理のレシピ集合からの基本手順の抽出. 言語処理学会第 21 回年次大会発表論文集, pp. 1092-1095, 2015.
- [9] 難波英嗣, 土居洋子, 辻田美穂, 竹澤寿幸, 角谷和俊. 複数料理レ

- シビの自動要約. 電子情報通信学会技術研究報告, Vol. 113, No. 338, pp. 39-44, 2013.
- [10] 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, pp. C1-1, 2017.
- [11] 久保遥, 関洋平. 投稿型レシピサイトを横断した重複レシピの判別. 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, pp. C8-3, 2016.
- [12] 福本亜紀, 井上悦子, 中川優. 食材の重要度と食品群を考慮したレシピ間類似度の算出手法. 第 4 回データ工学と情報マネジメントに関するフォーラム, pp. D9-2, 2012.
- [13] 花井俊介, 難波英嗣, 瀧本明代. 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, pp. E2-7, 2016.
- [14] 花井俊介, 瀧本明代, 難波英嗣. スパムレシピ抽出のための酷似レシピクラスタリング手法. 情報処理学会研究報告, Vol. 2014, No. 26, pp. 1-7, 2014.