

複数料理レシピの自動要約

難波 英嗣[†] 土居 洋子[†] 辻田 美穂[†] 竹澤 寿幸[†] 角谷 和俊[‡]

[†] 広島市立大学情報科学部 〒731-3194 広島市安佐南区大塚東 3-4-1

[‡] 兵庫県立大学環境人間学部 〒670-0092 兵庫県姫路市新在家本町 1-1-12

E-mail: [†] {nanba, doi, tsujita, takezawa}@ls.info.hiroshima-cu.ac.jp, [‡] sumiya@shse.u-hyogo.ac.jp

あらまし 本稿では、複数テキスト要約の技術を用いて、レシピサイトに投稿された特定の料理に関する複数のレシピから、その料理で使う典型的な材料と調理手順を出力する手法を提案する。レシピ間の類似個所の検出には、同義語辞書、格フレーム辞書、述語項構造解ツールなど、様々な言語資源が必要とされる。そこで、我々は統計的言語処理技術を用いて、レシピデータからこれらのリソースを構築し、複数レシピ要約に用いた。

キーワード レシピ, 複数テキスト要約, オントロジー, 述語項構造解析

Summarization of Multiple Cooking Recipes

Hidetsugu NANBA[†] Yoko DOI[†] Miho TSUJITA[†]

Toshiyuki TAKEZAWA[†], and Kazutoshi SUMIYA[‡]

[†] Faculty of Information Sciences, Hiroshima City University
3-4-1 Ozukahigashi, Asaminamiku, Hiroshima 731-3194 Japan

[‡] School of Human Science and Environment, University of Hyogo
1-1-12 Shinzaikē-honcho, Himeji, Hyogo 670-0092 Japan

E-mail: [†] {nanba, doi, tsujita, takezawa}@ls.info.hiroshima-cu.ac.jp, [‡] sumiya@shse.u-hyogo.ac.jp

Abstract In this paper, we propose a method for generating a typical procedure and an ingredient list for a certain dish from a set of recipes for the dish. For the identification of similarities among recipes, various linguistic resources, such as a synonym dictionary, a case frame dictionary, and a predicate argument structure analyzer, are required. We therefore constructed these resources from recipe data using statistical natural language processing techniques, and used them for multi-recipe summarization.

Keyword Recipe, Multi-Document Summarization, Ontology, Predicate Argument Structure Analysis

1. はじめに

食は人間の日々の生活と密接に結びつく営みである。そして料理レシピとは、より美味しいものを食べたい、より変化に富んだ食卓を演出したい、という欲求を満たすべく、経験や工夫の蓄積を経て生まれたものであり、それによって知識は共有され、継承される。今日、家庭料理のレシピは、おふくろの味として親から子や孫へ教え伝えられるものから、レシピサイトを通して多くの人々が手軽に検索・閲覧できるものとなった。例えば、代表的なユーザー投稿型レシピサイトのひとつである楽天レシピにおいて、「親子丼」というクエリでレシピを検索すると、2013年11月の時点で、計1,002件のレシピが見つかる。

レシピサイトを利用する人々の目指すものはその

各人の料理の技術や経験等によって異なる。料理の経験が少ない者ならば、より簡単に美味しくできる基本的なレシピを求めたろうし、逆に経験豊富な者ならそれまでとは異なる何らかの応用、アレンジが加わったレシピに魅力を感じるであろう。

本稿では、ある特定の料理に関する複数のレシピから、その料理で使う典型的な材料と調理手順を出力するシステムについて述べる。これはその料理の基本レシピを知りたいという層ばかりでなく、応用レシピを探したいというユーザにも有用なシステムである。入力された複数のテキストからひとつの要約を作成する、いわゆる「複数テキスト要約」では、入力テキスト間の類似点と相違点を検出することが必須の処理のひとつであると言われている[1]。今、ある料理に関する複

数のレシピを複数テキスト要約システムの入力と考えるならば、その料理の典型的な作り方と個々のレシピのアレンジ箇所を認識することは、複数テキスト要約における類似点と相違点の検出に該当する。そこで、本研究では、複数テキスト要約という観点から、ある料理の典型的な作り方を出力するシステムの開発を目指す。

本論文の構成は以下のとおりである。2 節では、関連研究について述べる。3 節では、複数レシピの自動要約システムを構築するために作成した評価用データと、このデータを用いた予備実験について述べる。4 節では、複数レシピを要約するために必要となる料理オントロジーについて述べる。5 節では、複数レシピ要約開発のための基礎的な実験について報告し、6 節で本稿をまとめる。

2. 関連研究

橋ら[2]は、レシピの料理名にある「簡単」や「子どもが喜ぶ」や「ヘルシー」といった修飾表現に着目し、それらの修飾表現の根拠(ネーミングコンセプト)をレシピから抽出する手法を提案している。橋らの手法では、ネーミングコンセプトを抽出する際、レシピの材料と調理器具に着目して典型的な要素との差異を抽出しているが、本研究で目指している複数レシピ間で類似点や相違点を検出する技術は、ネーミングコンセプトの抽出にも利用できる可能性がある。

レシピを要約する試みとして、西原らの研究[3]がある。この研究では、料理の大まかな流れを把握するため、手順の段階ごとに主要な動作表現を抽出し、それらを連結することで要約を生成する。これに対し、本研究では、ある料理に関する複数のレシピを比較することで、その料理の典型的な作り方を出力することを試みる。

山肩ら[4]は、「肉じゃが」や「カルボナーラ」などのクエリを用いて検索したレシピ集合に対し、各レシピをその調理手順を表したフローグラフに変換・統合することで、典型的な調理手順(レシピツリー)を導出する手法を提案している。さらに、典型的なレシピツリーと個々のレシピを比較することで、個々のレシピの特徴を抽出している。これらは、1 節で述べた複数テキスト要約における類似点と相違点の検出の一種と捉えることができる。

3. 複数レシピの分析

「複数レシピから、ある料理の典型的な材料と調理手順を出力する」というタスクを複数テキスト要約で扱うために、新たに何を検討しなければならないのかを明らかにするため、まず、評価用データを作成し、

次に、一般的な複数テキスト要約手法で実験を行い、最後に、複数レシピ要約では新たに何を検討しなければならないのか調査した。3.1 節では評価用データの作成について、3.2 節では一般的な要約手法を用いた予備実験について、3.3 節では実験の結果得られた知見について、それぞれ述べる。

3.1. 複数レシピ要約のための評価用データの作成

1 節でも述べたとおり、複数レシピ要約では、入力された複数のレシピ間の類似点と相違点を明らかにする必要がある。最終的に出力される要約そのものの質を評価するだけでなく、その要素技術である類似点の検出についても評価を行うため、テキスト自動要約の評価型プロジェクト TSC-3[5]の枠組みに沿って、評価用データの作成を行った。TSC-3 は、NII が主催する評価ワークショップ NTCIR-4 のタスクのひとつとして、2003~2004 年に実施された。TSC-3 では、新聞記事を対象に複数テキスト要約を作成する課題が設定された。本研究では、対象テキストをレシピに変える以外は、基本的には TSC-3 の枠組みに沿って評価用データを構築し、評価を行う。以下に、評価用データの構築方法について説明する。

トピックの設定

TSC-3 では、1 トピックあたり平均約 10 記事、計 30 トピックについて新聞記事データを用いて、評価用データを構築している。本研究では、1 トピックあたり平均約 10 レシピ、計 35 トピックのデータを作成し、このうち 5 トピックをシステム作成(訓練)用、残り 30 トピックを評価用に用いた。図 1 にトピックの一覧を示す。トピックの選定は、和食/洋食/中華、あるいはメインディッシュ/サイドディッシュ/デザートなど、35 トピックが特定のカテゴリの料理に偏らないよう配慮した。

001 酢豚	019 梅酒
002 角煮	020 バンバンジー
003 ミートソースパスタ	021 高野豆腐
004 豚肉のしょうが焼き	022 エビフライ
005 パンプディング	023 大学いも
006 イチゴジャム	024 麻婆豆腐
007 ハンパロア	025 だし巻き卵
008 みたらし団子	026 おぼろ
009 おからハンバーグ	027 ゴーヤチャンプルー
010 パンプキンスープ	028 水餃子
011 カレーピラフ	029 スパニッシュオムレツ
012 カステラ	030 カスタードプリン
013 あんぱん	031 唐揚げ
014 かぼちゃコロッケ	032 カルボナーラ
015 サバの味噌煮	033 親子丼
016 カキフライ	034 茶碗蒸し
017 切干大根	035 エビマヨ
018 いなりずし	

図 1 複数レシピ要約に用いたトピック一覧

Fig. 1 A list of topics used for multi-recipe summarization

評価用データの作成

図 1 に示すトピックごとに用意した約 10 レシピについて、レシピ間で共通する個所の同定を行った。同定を行う際、TSC-3 では句点で区切られる文をひとつの単位として扱っていたが、レシピの場合、1 文中に複数の手順が記載されることが少なくないため、日本語構文解析器 CaboCha を用いてレシピ中の各文を単文に分割した上で、同定を行った。図 2 に実際に同定された文の例を示す。図 2 では、文 1020000538-1-1-1 と 1020000310-1-1-1 , 1020000538-2-1-1 と 1020000310-1-1-2 がそれぞれ同定されている。同定された文集合は半角空白で区切られ、全体を p タグで挟んで表現される。なお、あるレシピ中の 2 文 A と B が別のレシピの 1 文 C と同定される場合もある。この場合は「<p>A,B C</p>」のように、A と B は半角カンマで区切って表現する。

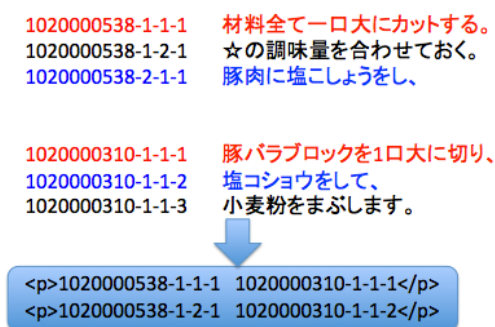


図 2 複数レシピ間で同定された文の例

Fig. 2 An example of manually identified sentences among multiple recipes

各レシピには、手順リストの他に、材料リストも存在している。材料リストについても、手順リストと同様に同定を行う。

3.2. 複数レシピ要約の予備実験

一般的な複数テキスト要約手法を用い、どの程度の精度が得られるか実験を行った。

複数テキスト要約システム

1 節でも述べたとおり、一般的な複数テキスト要約手法では、入力されたテキスト間の類似点を検出する。その理由のひとつは、類似した内容の文が出力される要約に何度も出現すると冗長であるからである。もうひとつの理由は、複数の入力テキストの中で共通して言及されている内容は重要な話題に違いない、という仮定に基づき、重要個所を検出するための手段として利用できるからである。この考えに基づき、複数レシピ要約を行う。

まず、対象レシピの各文から内容語(名詞句と動詞)を抽出し、次に、tf*idf で単語の重み付けを行った後、クラスタリングを行った。文間の類似度にはコサイン

距離を、クラスタリング手法には Repeated Bisection 法を用いた。クラスタリングの結果、入力レシピのうち過半数のレシピに出現する手順を典型的な手順として出力した。なお、クラスタリングを実行する際の閾値の設定は、後述する訓練用データを用いて決定した。

実験データ

図 1 に示す 35 トピックのうち、トピック 001~005 を訓練用(パラメータチューニング用)に、残りの 30 トピックを評価用に用いた。

評価尺度

TSC-3 で定義されている coverage と precision を用いて評価を行った[5]。

実験結果

実験の結果、評価用データにおいて、手順の同定では coverage=0.663, precision=0.607 が、材料の同定では coverage=0.880, precision=0.838 が得られた。

3.3. 複数レシピ要約を作成する上で考慮すべき点

3.2 節で述べた実験結果について考察を行った結果、レシピの手順の同定を行う際、次に述べる 3 点が、coverage と precision を下げる要因になっていることが分かった。

(要因 1) 表記の揺れ・同義語

今回実験に用いた投稿型レシピサイトのひとつである楽天レシピには、ユーザによって同一の材料や動作であっても表記方法が様々であり、それが同定を困難にする要因のひとつになっていた。図 3 は、酢豚に関する 274 レシピで「豚肉」がどのように表現されているかを調査し、まとめたものである。図では、274 レシピ中での表現ごとの出現頻度もあわせて示している。レシピ間で正しく同定を行うためには、このような表記の揺れや同義語を正しく認識する必要がある。

64 豚肉	3 かたまり肉
18 豚	3 豚肉用
12 豚こま切れ肉	3 豚角切り
8 豚ひき肉	3 豚ブロック肉
8 豚もも肉	3 豚肉かたまり
7 豚肩ロース	3 豚角切り肉
6 豚ヒレ肉	2 豚小間肉
5 豚バラブロック	2 トンカツ用豚肉
5 豚バラ肉	2 豚コマ
4 豚こま	2 豚肉薄切り
4 豚肩ロース肉	2 薄切り豚肉
4 豚ロース肉	2 豚肉角切り
4 豚かたまり肉	2 豚モモ薄切り肉
4 豚小間	2 豚小間切れ
3 豚ロースかたまり	2 豚こま
3 お肉	

図 3 「豚肉」に関する様々な表記

Fig. 3 Various expressions for "pork"

(要因 2) 省略

本研究では、同定を行う際に、原文を単文に分割している。図 4 は、実際に分割された例である。図において、2 文目と 5 文目にある「(鯛を)」という表現は、原文では省略されている。この例の場合、2 文目と 5 文目は、それぞれ「捌く」と「焼く」という内容語 1 語だけで同定を行わなければならない、これも正しく同定することを困難にするもうひとつの理由になっている。

鯛は内臓とうろこを取り除き、
(鯛を)捌いて
骨を丁寧にとった
切身にし、
(鯛を)焼きます。

図 4 レシピ中での省略の例

Fig. 4 An example of ellipses in a recipe

(要因 3) 多対多で対応する手順の同定

あるレシピでは「グラニュー糖と水でカラメルを作り」と 1 文で表現されている個所が、別のレシピでは「砂糖と水 大 2 を鍋に入れて/キツネ色になるまで/火をかける./火からおろして/お湯 大 4 を加える。」と 5 文で表現される場合がある。3.2 節で述べた一般的な複数テキスト要約手法では、このような 1 対多あるいは多対多で対応する手順の同定ができない。

4. 料理オントロジーの構築とテキスト要約への応用

3.3 節で述べた問題に対処するため、料理オントロジーの構築を試みた。ここで、料理オントロジーとは、レシピを言語解析するために利用できる知識体系のことであり、用語の同義語や上位・下位概念、典型的な調理手順を定型化したものなどを指す。本節では、料理オントロジーの構築方法と、そのテキスト要約への応用について述べる。

4.1. 概念辞書および同義語辞書の構築

本研究では、言語処理技術を用いて料理用語の概念階層を作り、さらに概念ごとに同義語リストを作成する。本研究で構築する概念辞書は「大分類-小分類-小分類の同義語」の 3 階層から構成される。このうち、大分類と小分類については「(i) 概念階層の構築」で、小分類の同義語については「(ii) 小分類の同義語」で、それぞれ述べる。

(i) 概念辞書の構築

大分類については、楽天レシピのカテゴリを参考に、一部拡張して「材料-魚介」「材料-肉」「材料-野菜」「材料-その他」「調味料」「調理器具」「動作」の 7 種類に設定した。大分類ごとに、次に述べる方法で、小分類の用語の候補を収集した。一般に、特許は権利の範囲

をなるべく広く確保するため、一般性の高い用語を用いて記述される。一方、その範囲をより明確にするため、各用語が具体的に何を指し示しているのか、明細中に記載されることが多い。その際、「A や B 等の C」といった提携表現が使われる。この表現に着目することで、「A と B の上位概念は C である」という知識が得られる[6]。特許の中には料理に関するものも数多く含まれているため、特許を利用して、小分類の用語を選定する。具体的には、「材料-魚介」に属する小分類の用語を収集するのに「A や B 等の魚類」や「C 等の魚介類」というパターンを用い、ここから A, B, C といった用語を小分類用語の候補として抽出する。こうして抽出された用語を頻度順にまとめ、頻度の高いものから順に小分類用語を選定する。上述の「魚類」や「魚介類」に相当する用語として、大分類ごとに、以下のものを利用した。

- 材料-魚介：魚類，魚介類，海産物，水産物
- 材料-肉：肉類，食肉，食肉類，原料肉
- 材料-野菜：野菜，果菜類，野菜類，果菜物，農産物
- 材料-調味料：調味料，香辛料，薬味，スパイス類
- 調理器具：調理器具，調理容器，調理器，調理具，調理道具

「材料-その他」については、上記の用語を使って収集した結果、いずれにも当てはまらないものをこのカテゴリの小分類用語とした。また、動詞については、楽天レシピ中に出現する動詞を頻度順に並べ、その上位を選定対象とした。なお、ここで構築される概念辞書は、前節で述べた要因 2 に対応するためのもので、次節で述べるレシピの述語項構造解析に用いる。

(ii) 同義語辞書の構築

これまでに、テキストデータから同義語や関連語を自動的に抽出する数多くの手法が提案されているが、その代表的な手法のひとつである分布類似度[7,8]を利用し、同義語辞書の構築を行う。分布類似度とは、2 つの用語 A と B が意味的に類似していれば、A と B の文脈に出現する語の傾向が似ている、という仮定に基づいた関連語収集方法である。文脈語の選定には、全単語を使う方法、内容語のみ使う方法など、色々な方法が考えられるが、本研究では、相澤の手法[8]に従い、対象となる語と係り受け関係にある動詞を文脈語として利用する。この方法は、ある動詞に着目し、その動詞と係り受け関係にある名詞を文脈語と考えれば、動詞の関連語を収集することも可能になる。

同義語や関連語を自動収集する方法として、分布類似度の他に、Chung が提案する手法も用いる[9]。多くのレシピでは、ある料理で使う材料のうち、主要なも

のから順に材料リストに記載する傾向にある。この特徴を利用し、例えば楽天レシピのように、各レシピが材料ごとに階層的に分類されている場合には、「エビ」というカテゴリに分類されているレシピの材料リストの先頭に記載されている材料を集めてくれば、「エビ」の同義語が効率的に集められる。Chung は、この手法を用いることで、F 値で 0.93 の精度で同義語が収集できることを実験により示している。ただ、この手法では、調味料など料理の主要材料になりえないものや、調理器具、動詞などに関しては対応できない。そこで、本研究では、Chung の手法と分布類似度を併用して幅広く同義語・関連語の候補を収集し、それらを人手で選定して、同義語リストを作成した。

概念辞書は、本論文を執筆している 2013 年 11 月現在も整備中だが、これまでに小分類レベルで 5402 語を含む辞書が作成されている。

この同義語辞書は、Zhou ら[10]や平原ら[11]が提案している言い換えを考慮した 2 つのテキスト間での類似性の判定技術を用いることで、前節の要因 1 を解決することが出来る。

4.2. レシピの述語項構造解析

3.3 節の図 4 で示した例では、2 文目と 5 文目で「鯛」という表現が省略されているが、この省略を補完する処理は、述語項構造解析と呼ばれている。本節では、4.1 節で述べた概念辞書と自動構築した格フレームを組み合わせた述語項構造解析手法を提案し、前節の要因 2 の解決を試みる。

河原ら[12]は、Web 上から収集した大量の文を構文解析し、クラスタリングにより表層格の要素をまとめることで、格フレームを自動的に構築する手法を提案している。河原らの手法をレシピに適用し、さらに得られた格要素に 4.1 節で述べた概念辞書に対応付けた例を図 5 に示す。

```
@捌く
4 イカ_を <cat1>材料-魚介</cat1><cat2>イカ</cat2>
2 いか_を <cat1>材料-魚介</cat1><cat2>イカ</cat2>
2 蟹_を <cat1>材料-魚介</cat1><cat2>かに</cat2>
2 いわし_を <cat1>材料-魚介</cat1><cat2>イワシ</cat2>
1 スルメイカ_を <cat1>材料-魚介</cat1><cat2>イカ</cat2>
1 さんま_を <cat1>材料-魚介</cat1><cat2>秋刀魚</cat2>
1 鱈_を <cat1>材料-魚介</cat1><cat2>鱈</cat2>
```

図 5 自動的に構築された格フレーム辞書の例

Fig. 5 An example of an automatically constructed case frame dictionary

図は、「捌く」という動詞のヲ格を示している。図からわかるとおり、「捌く」のヲ格は「材料-魚介」に関する用語になる場合が多い。そこで、3.3 節の図 4 の場合、2 文目と 5 文目にある「捌く」の直近の「材料-魚介」に関する用語を補完すれば、述語項構造解析

が実現できると考えられる。以下に解析手順を示す。

1. 入力されたレシピを、4.1 節で述べた概念辞書中の用語と長いものから順に照合し、用語がレシピに含まれていれば、その個所を「<大分類-小分類>タグ」で置換する。
2. 手順 1 でマッチしたタグは、スタックに積む。
3. 動詞については、同一文内にヲ格とデ格が存在しない場合、格フレーム辞書を読み込み、ヲ格とデ格になりうるタグリストを、それぞれ獲得する。
4. 手順 3 で得られたタグリストと手順 2 のスタックを、スタックの上から順に照合し、タグがマッチすれば、それをヲ格またはデ格として出力する。

4.3. 多対多で対応する手順の同定

3.3 節の要因 3 に対応するには、「天ぷら粉の作り方」や「カラメルの作り方」など、ある料理を作る上での、まとまりのある一連の手順を自動抽出し、知識として蓄積する必要がある。その第一歩として、「X を[動詞]で Y を作る。」という定型表現に着目し、「(入力 X)-動詞→(出力 Y)」の三組の抽出を試みた。この表現にマッチする文として、例えば以下のようなものがある。ここで、波下線部が動作、下線部が動作の入力、二重下線部が動作の出力に対応する。

グラニュー糖と水を合わせて、カラメルを作ります。

このような知識を蓄積すれば、「カラメル」を作るための動作の開始を検出するために、入力となる「グラニュー糖」や「水」が含まれる手順を探す、といった利用方法が考えられる。

この定型表現を楽天レシピデータに適用したところ、「X を混ぜて Y を作る」でマッチしたものが 2,630 件、「X を合わせて Y を作る」でマッチしたものが 2,146 件得られた。

5. 実験

4 節で述べた 3 種類のオントロジー構築手法のうち、4.2 節の述語項構造解析について実験を行った。

楽天レシピ約 44 万件を、CaboCha を用いて構文解析し、動詞ごとに格要素をまとめ、格フレーム辞書を構築した。動詞の異なり数は 3,450 であり、河原らが行っているようなクラスタリングにより、動詞の語義ごとに格フレームをまとめる処理は行っていない。また、今回は、述語項構造解析モジュールを評価するためのデータの作成が間に合わなかったため、定量的な評価は行っていない。

「刺身用いかの胴の部分を使用します。捌いて皮を剥ぎ 5mm の厚さのいか刺しにします。」という文を入力した時の解析例を図 6 に、また、この 2 文を解析し

た時のスタックを図7に、それぞれ示す。図6において、2文目の先頭に出現している「捌く」という動詞については、「いか」が正しく補完されていることが分かる。

```
<sentence id="1">
  刺身用<助詞>の</助詞><材料-魚介-イカ>いか</材料-魚介-イカ><
  助詞>の</助詞> 胸<助詞>の</助詞> 部分<助詞>を</助詞> 使用<
  動作-その他 org="する" wo="" de="">し</動作-その他> ます。
</sentence>
<sentence id="2">
  <動作-捌く org="捌く" wo="" de=""><材料-魚介-イカ>いか</材料-魚介-イカ>
  de="">捌い</動作-捌く><助詞>て</助詞>
  皮<助詞>を</助詞><動作-剥ぐ org="剥ぐ" wo="" de="">剥ぎ</動作-剥ぐ>
  5 mm<助詞>の</助詞> 厚さ<助詞>の</助詞><材料-魚介-イカ>いか</材料-魚介-イカ>
  wo="" de="">し</動作-その他> ます。
</sentence>
```

図6 述語項構造解析の解析例

Fig. 6 An example of predicate argument structure analysis

```
<stack>
<動作-その他 org="する" wo="" de="">し</動作-その他>
<材料-魚介-イカ>いか</材料-魚介-イカ>
<動作-剥ぐ org="剥ぐ" wo="" de="">剥ぎ</動作-剥ぐ>
<動作-捌く org="捌く" wo="" de=""><材料-魚介-イカ>いか</材料-魚介-イカ>
<動作-その他 org="する" wo="" de="">し</動作-その他>
<材料-魚介-イカ>いか</材料-魚介-イカ>
</stack>
```

図7 述語項構造解析に用いたスタックの例

Fig. 7 An example of a stack used for predicate argument structure analysis

いくつかのレシピを用いて解析を行ったところ、解析誤りには動詞の多義語の問題が影響として大きいことがわかった。特に、動詞がひらがなで記述される場合に、この問題が発生する可能性が高くなる。例えば「とじる」という表現は「卵で綴じる」という意味と「蓋を閉じる」という意味がある。このような問題に対処するためには、河原らが行っているようにクラスタリングにより語義ごとに格フレームを構築する必要があるが、レシピの場合、しばしば格要素が省略される傾向にあるため、クラスタリング手法だけで多義の問題がすべて解決できる保証はない。

6. おわりに

本研究では、ある料理に関する複数のレシピから、その料理の典型的な材料と手順を出力する手法を提案した。また、複数レシピ要約の生成に必要な料理オントロジーを構築した。

今後の課題として、今回は述語項構造解析を行う際、助詞の情報、述語からの距離など、述語項構造解析では一般的に使われる情報を使っていない。この点につ

いては、レシピのタグ付きコーパス[13]から機械学習ベースで解析器を構築する手法なども検討する必要があると思われる。

謝辞

本研究を遂行するにあたり、解析対象となるレシピデータを楽天技術研究所よりご提供いただいた。ここに記して謹んで感謝の意を表する。

文 献

- [1] 奥村学, 難波英嗣, “テキスト自動要約,” コロナ社, 2005.
- [2] 橘明徳, 若宮翔子, 難波英嗣, 角谷和俊, “料理名の修飾表現の関係性にに基づくレシピのネーミングコンセプト抽出,” 電子情報通信学会技術研究報告, DE 研第1種研究会 データ工学と食メディア, Vol. 113, No. 214, DE2013-36, pp.19-24, 2013.
- [3] 西原弘真, 苅米志帆乃, 藤井敦, “料理レシピを対象としたアウトライン型自動要約,” 情報処理学会研究報告, Vol.2013-IFAT-110, No.8, 2013.
- [4] 山肩洋子, 今堀慎治, 杉山祐一, 田中克己, “レシピフローグラフを介したレシピ集合の要約と特徴抽出,” 電子情報通信学会技術研究報告, DE 研第1種研究会 データ工学と食メディア, Vol. 113, No. 214, DE2013-36, pp.43-48, 2013.
- [5] T. Hirao, M. Okumura, T. Fukusima, and H. Nanba, “Text Summarization Challenge 3 - Text Summarization Evaluation at NTCIR Workshop 4 -,” Proceedings of the 4th NTCIR Workshop, 2004.
- [6] H. Nanba, S. Mayumi, and T. Takezawa, “Automatic Construction of a Bilingual Thesaurus using Citation Analysis”. Proceedings of the 4th International CIKM Workshop on Patent Information Retrieval (PaIR'11), pp.25-30, 2011.
- [7] D. Lin, “Automatic Retrieval and Clustering of Similar Words,” Proceedings of COLING/ACL 1998, pp.768-774, 1998.
- [8] 相澤彰子, “大規模テキストコーパスを用いた語の類似度計算に関する考察,” 情報処理学会論文誌, Vol.49, No.3, pp.1426-1436, 2008.
- [9] Y. Chung, “Finding Food Entity Relationships Using User-generated Data in Recipe Service,” Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012), pp.2611-2614, 2012.
- [10] L. Zhou, C.-Y. Lin, D.S. Munteanu, and E. Hovy, “ParaEval: Using Paraphrases to Evaluate Summaries Automatically,” Proceedings of HLT-NAACL 2006, pp.447-454, 2006.
- [11] 平原一帆, 難波英嗣, 竹澤寿幸, 奥村学, “言い換えを用いたテキスト要約の自動評価,” 情報処理学会論文誌データベース, Vol.3, No.2, pp.91-101, 2010.
- [12] 河原大輔, 黒橋禎夫, “格フレーム辞書の漸次的自動構築,” 自然言語処理, Vol.12, No.2, pp.109-131, 2005.
- [13] 森信介, 山肩洋子, 笹田鉄郎, 前田浩邦, “レシピテキストのためのフローグラフの定義,” 情報処理学会研究報告, 2013-NL-214(13), 2013.