

令和元年6月20日現在

機関番号：14301

研究種目：基盤研究(B) (一般)

研究期間：2014～2018

課題番号：26280039

研究課題名(和文) 消費者生産型レシピコンテンツの手順・記述から見た多様性の解析手法の提案

研究課題名(英文) An Analyzing Method for Diversity of Consumer Generated Recipes from the Viewpoint of Procedural and Representation

研究代表者

山肩 洋子 (Yamakata, Yoko)

京都大学・情報学研究科・特定准教授

研究者番号：60423018

交付決定額(研究期間全体)：(直接経費) 13,200,000円

研究成果の概要(和文)：本研究の目的は、Webにある膨大な数のレシピの集合が本質的にどの程度の多様性を持っているのかを明らかにすることである。そこで本研究では、自然言語処理技術によりレシピ記述から手順構造を抽出し、レシピ集合が持つ本質的な多様性を解析する手法の開発を行った。和文レシピ、英文レシピ各400件に対しアノテーションを行い、日英レシピコーパスを構築。レシピ用語の認識精度は和文F値92.6%、英文F値87.2%に到達した。またレシピ用語間の依存関係の推定精度は和文F値79.7%、英文F値76.2%を達成した。これらのツールはオンライン上で公開し、多くの研究機関にご利用いただいている。

研究成果の学術的意義や社会的意義

料理は「ものづくり」であり、レシピは料理を作る「手順書」である。本課題で解析した手法が適応できる範囲は料理レシピに限らず、裁縫やDIY、マニュアルなどあらゆる「手順書」へと拡張できるという点で学術的な意義がある。また、本課題を遂行する中で構築したレシピ言語処理ツールは、国内外でレシピを研究対象とする多くの研究グループに利用されているという点で、社会的意義は大きい。NIIクックパッドデータセットに含まれる和文レシピ172万件に対し、レシピ用語認識及びフローグラフ化の作業を完了した。和文レシピの解析データは、NIIクックパッドデータセットのサードパーティーとしての提供を目指し、協議を進めている。

研究成果の概要(英文)：The purpose of this research is to analyze diversity of millions of cooking recipes that are reachable on the Web. In this study, we developed a method to extract the procedural structure from the recipe procedural description by natural language processing technology and analyze the diversity of recipe collection. We annotated Japanese and English recipes for 400 each and constructed Japanese and English recipe corpora. The recognition accuracies of recipe named entity reached 92.6% of F0 in Japanese and 87.2% of F0 in English. Moreover, the estimation accuracy of dependencies between recipe named entities achieved 79.7% of F0 in Japanese and 76.2% of F0 in English. These tools are available online and are used by many research institutes.

研究分野：メディア情報処理

キーワード：メディア情報処理 情報検索 自然言語処理 グラフ理論

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

レシピは一般の主婦を中心として最も活発に生産・利用されている消費者生産型メディアの一つである。すでに数百万件のレシピが Web 上に存在し、なお増加の一途をたどっている。本研究の目的は、それらのレシピ集合が本質的にどの程度の多様性を持っているのか、足りないのか十分なのか、何が足りないのかを明らかにすることである。そのためには、似た二つのレシピに注目したとき、手順(作り方)は同じだが記述が異なるのか、それとも手順自体が異なるのかを分離して扱う必要がある。そこで本研究では、(i)自然言語処理技術によりレシピ記述から手順構造を抽出し、(ii)手順と記述の観点からレシピ間の関係を解析するとともに、(iii)全体の知識を使って補完可能な欠損を補完することで、レシピ集合が持つ本質的な多様性を解析する機構を構築することを目的に設定した。

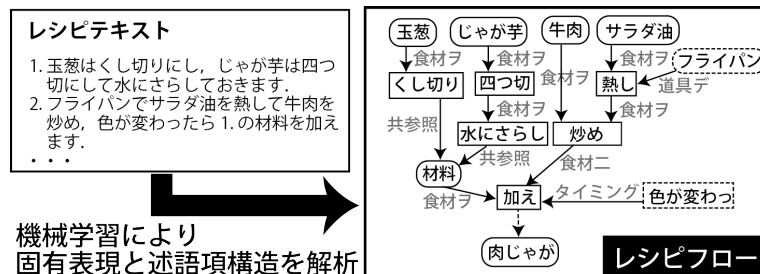


図 1) レシピの手順情報を構造化

使用する材料の組合せによりそのレシピの特徴を推定する研究が多く行われている[佃ら 2011]。しかしながら同じ組み合わせの食材でも炒め物や煮物など、その後の手順が異なると全く違う料理になる。また、材料も手順も全く同じだが、その記述の仕方が異なれば違うレシピとなる。すなわちレシピは『材料・手順・記述』の 3 つの指標によりそれぞれ多様性があると考えられる。『材料』については、材料リストからその組み合わせを獲得することが容易なため従来研究が数多くあるが、『調理手順』は文章で記述されるため、その構造を抽出することが困難であった。これに対し我々は、言語モデル学習に基づき消費者生産型レシピテキストを構造化する手法を開発してきた(図 1)。この手法で構造化されたレシピは DAG(無閉路有向グラフ)となるため、レシピ間の関係は、複数の DAG 間でのグラフマッチング問題に帰着できる。そこで、レシピテキストを構造化し、手順情報と記述の様態をそれぞれ抽出して、材料・手順・記述の 3 軸でレシピの関係を解析する手法を提案することを考えた。

2. 研究の目的

我々の調査によれば、ユーザがレシピを選ぶ際に最も重視する特徴は『調理手順』の違いである。特に「肉じゃが」などの料理名をクエリとして検索した場合、使用する材料はどれも似通っているため、同じ「肉じゃが」でも調味料を煮る段階ではなく炒める段階で加えたり、肉を先に炒めて一旦取り出しておくなどの手順の違いがレシピの選択に大きく影響するが、従来研究ではレシピから手順情報を抽出する手段がなかったために、その検索が実現できなかった。記述の詳細さについても同様である。そこで本研究では、検索結果として得られたレシピ集合から典型的な材料・手順・記述を抽出し、それと個々のレシピの特徴を抽出して提示することで、材料・手順・記述の 3 つの視点でレシピを検索することのできるシステムを構築する。

3. 研究の方法

通常、レシピは材料リスト添付されていることから、材料データの抽出は比較的容易であり、数多くの研究が存在している。タンパク質やエネルギー等を計算する上で、その解析が重要であることから、食品栄養学会等でも詳細な研究がなされていることから、本研究では手順書の解析に注力して研究を行った。本研究は次のような 4 つの流れで行った。(1) データの収集およびコーパスの作成、(2) 手順構造解析手法の開発、(3) 大規模レシピデータの解析、(4) 解析済みデータの分析とその応用手法の開発

4. 研究成果

本節では、「3. 研究の方法」で説明した 4 つの流れに沿って、その成果を説明する。

(1) データの収集およびコーパスの作成

和文レシピに関しては、国立情報学研究所よりクックパッドデータセットに 172 万件、楽天データセットに約 80 万件のレシピが収録されており、十分な数のレシピデータがあるうえ、これが国内の多くのレシピ研究のスタンダードなデータセットとなっていることから、他にデータの収集は行わなかった。クックパッドデータセットよりランダムに 412 件のレシピを抽出し、我々が定めた基準によって、レシピの固有表現タグ(表 1)と、固有表現間の依存関係構造(表 2)のタグ付けを行った。これにより図 1 のような構造を表現することが可能となる。

表 1) レシピの固有表現タグ

クラス	意味	備考
F	食材	中間・最終生成物を含む
T	道具	調理道具や器など
D	継続時間	概数表現を含む
Q	分量	概数表現を含む
Ac	調理者の動作	語幹のみ
Ac2	調理者の動作	連続しない動詞句の後ろの要素
Af	食材の動作	語幹のみ
At	道具の動作	語幹のみ
Sf	食材の状態	
St	道具の状態	

Ac2 と At は英語のレシピのフローグラフにのみ出現

→

表 2) 依存関係構造タグ

ラベル	意味
Agent	主語
Targ	対象
Dest	方向
T-comp	道具を用いて
F-comp	食材を用いて
F-eq	同一の食材
F-part-of	食材の一部
F-set	食材の集合
T-eq	同一の道具
T-part-of	道具の一部
A-eq	同一の動作
V-tm	動作を行う
other-mod	その他の修飾語句

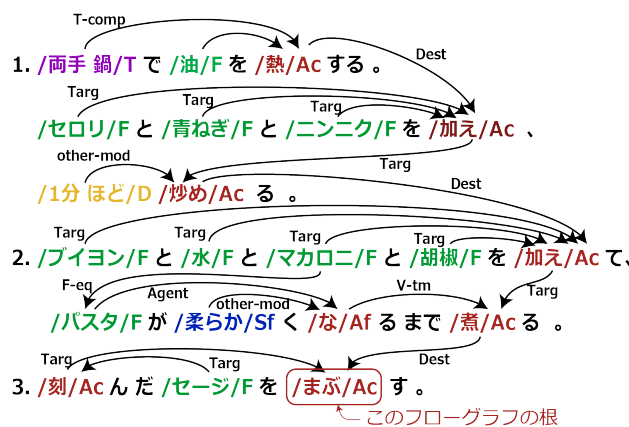


図 1) レシピのフローグラフ構造

英文レシピに関して、Allrecipes.com と food.com (現在は Genius Kitchen に改名) から英文レシピ約 50 万件を収集し、うち 400 件にタグ付けを行って英文レシピコーパスを構築した。このコーパスのガイドラインは、和文のものを翻訳したのち、英文に適合するよう微調整を加えたものであるから、日英のコーパスのガイドラインは概ね一致しており、日英レシピデータの共通規格という。

(2) 手順構造解析手法の開発

(1) で構築したコーパスを学習データとして、機械学習を適用することにより、レシピの手順文書から図 1 のような構造を抽出する手法を開発した。これは 3 つのステップからなる。1 ステップ目はレシピ用語の認識であり、(1) で構築したコーパスを学習データとして、Support Vector Machine (SVM) を学習する手法を採用した。こうして抽出されたレシピ用語の前組み合わせに対し、2 ステップ目で依存関係の有無を推定する。ロジスティック回帰モデルを学習する手法を採用した。最後に 3 ステップ目で、レシピの最後に現れるレシピ用語をルートとして、依存関係の有無の推定値を流量と見立て、Maximum Spanning Tree を抽出することにより、ツリーを抽出した。これにより、レシピ用語の認識精度は和文 F 値 92.6%、英文 F 値 87.2%、依存関係の有無の推定精度は和文 F 値 79.7%、英文 F 値 76.2% を達成した。この解析ツールと学習済みの統計モデルはホームページを通じて公開している。

(3) 大規模レシピデータの解析

(2) の手法を (1) で収集したレシピに適用し、レシピの性質を解析した。我々が開発したツールは、多くの研究者に利用されるようになったが、まだ情報分野が多数を占める。一方で、レシピの解析はむしろ人文社会系や医療・栄養学など、情報分野以外での研究が盛んであり、特にレシピから料理に含まれるエネルギーや脂質、タンパク質等を推定する栄養学分野から注目を寄せられている。我々のツールは外部の機械学習ライブラリを複数採用しており、またマシンパワーを要する処理もあることから、情報分野以外の研究者が稼働させるのは難しい場合が多い。そこで、学習済みのモデルを公開したり、共同研究として解析済みのレシピを共有する試みを始めた。NII データセットは二次配布はできないため、同データセットの利用申



図2) FoodLog アプリのスクリーンショット。画像認識で各料理の名前を推定してカロリー情報を付与する。

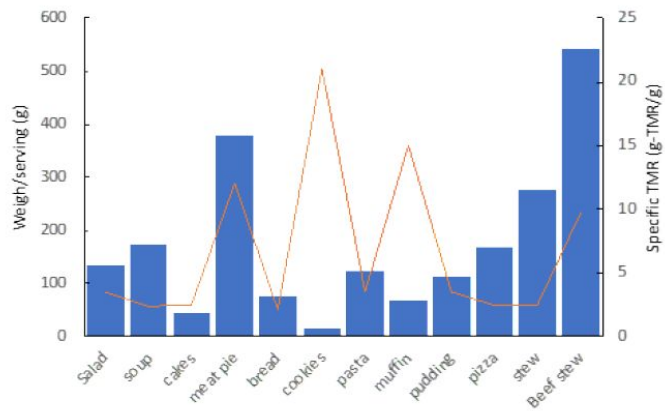


図3) 様々な料理における1食当たり、グラムあたりの関与物質総量

請を行った研究グループに対し、技術提供を行っている。

(4) 解析済みデータの分析とその応用手法の開発

当初の目的はレシピ検索・推薦であったが、研究機関が進むなかで、スマートフォンによる食事管理アプリが普及してきたことから、レシピを活用した料理の栄養素解析の研究を行っている。料理のカロリーを自動計算する際は、その材料の持つカロリーの積算で代替する方法が用いられるが、例えば同じ豚肉でも焼いたのか茹でたのかによってカロリーが大きく変わる等、調理法を考慮する必要があり、栄養士がカロリーを推定するときも同様である。また、レシピの分析においても、申請者が所属している相澤研究室が開発している食事記録アプリ FoodLog のデータを用いて、ユーザの食傾向を分析する研究を行った。FoodLog アプリには料理画像認識の機能も用意されているが、多くのユーザは料理名をテキスト入力している。料理名の名づけは、「ほうれん草カレー」のように、食材の名前や料理名、「吉野家の牛丼」などレストラン名など、様々な要素が組み合わされていることから、料理名からその料理の栄養素を抽出することは挑戦的な課題であり、本課題で構築したレシピの解析データが有用である。

2018年からは、立命館大学 山末英嗣教授と協力して、食による関与物質総量 (Total Material Requirement, TMR) のボトムアップ解析とビッグレシピデータ分析の研究を開始した。同教授は循環型材料生産プロセスに関する文理融合型研究を行っており、食 (特に農作物) が誘発する資源採掘活動 (採掘活動) を定量化し、その多様性の変化が環境影響や資源リスクに与える影響やその変容メカニズムを解明することを目的としている。本課題で開発したツールを用いて解析したレシピデータを用いて料理ごとの TMR を算出し、料理によって TMR に大きな開きがあることを示した (図3)。この成果は、2018年7月にストックホルムで開催された食のワークショップ CEA2018 において発表された。

5. 主な発表論文等

[雑誌論文] (計8件)

Yoko Yamakata, Hirokuni Maeta, Takuya Kadowaki, Tetsuro Sasada, Shinji Imahori, Shinsuke Mori, "Cooking Recipe Search by Pairs of Ingredient and Action - Word Sequence v.s. Flow-graph Representation", Transactions of the Japanese Society for Artificial Intelligence, Vol.32, No. 1, WII-F, pp.1-9, 2017.

前田 浩邦, 山肩 洋子, 森 信介, "手順文書からの意味構造抽出", 人工知能学会論文誌, Vol.32, No. 1 E, pp.1-8, 2017.

山肩 洋子, 今堀 慎治, 森 信介, 田中 克己, "ワークフロー表現を用いたレシピの典型性評価と典型的なレシピの生成", 電子情報通信学会論文誌 D, Vol.J99-D, No.4, pp. 378-391, 2016.

Mori, Shinsuke, Neubig, Graham, "A comparative study of dictionaries and corpora as methods for language resource addition", Language Resources and Evaluation. Vol. 50, No. 10, 2016

山崎 健史, 吉野 幸一郎, 前田 浩邦, 笹田 鉄郎, 橋本 敦史, 船富 卓哉, 山肩 洋子, 森 信介, "フローグラフからの手順書の生成", 情報処理学会論文誌, Vol. 57, No.3, pp. 849-862, 2016.

山崎 健史, 吉野 幸一郎, 前田 浩邦, 笹田 鉄郎, 橋本 敦史, 船富 卓哉, 山肩 洋子, 森 信介, "フローグラフからの手順書の生成", 情報処理学会論文誌, Vol. 57, No.3, pp.

849-862, 2016.

門脇拓也, 山肩洋子, 森信介, 田中克己, “誕生・使用事由によるレシピ検索～生い立ちレシピサーチ～,” 日本データベース学会和文論文誌, Vol. 13-J, No. 1, pp.78-85, 2014.

井上仁, 橋本敦史, 中村和晃, 船富卓哉, 山肩洋子, 上田真由美, 美濃導彦, “食材認識のための画像と食材切断時の振動音及び荷重の利用,” 電子情報通信学会論文誌, Vol. J97-D, No.9, pp.1490-1502, 2014.

[学会発表](計13件)

Yuji Goda, Sosuke Amano, Yoko Yamakata, Kiyoharu Aizawa, "Bag-of-foods: analysis of personal foodlogging data", IJCAI Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management, pp.63-66, July 15 - 15, 2018, Stockholm, Sweden

Yoko Yamakata, Eiji Yamasue, Benjamin McLellan, Kiyoharu Aizawa, "Resource intensity for menu items: how much land is required to provide for each dish?", IJCAI Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management, pp.15-20, July 15 - 15, 2018, Stockholm, Sweden

Yoko Yamakata, Shinsuke Mori, John Carroll, "A Comparison of Cooking Recipe Named Entities between Japanese and English," 9th Workshop on Multimedia for Cooking and Eating Activities (CEA2017), 2017 (CEA2017 Best Paper Award). 査読有り

Yoko Yamakata, Shinji Imahori, Hirokuni Maeta, Shinsuke Mori, "A Method for Extracting Major Workflow composed of Ingredients, Tools, and Actions from Cooking Procedural Text," 8th Workshop on Multimedia for Cooking and Eating Activities (CEA2016), 2016. 査読有り

Atsushi Hashimoto, Jun Harashima, Yoko Yamakata, Shinsuke Mori, "Design in Everyday Cooking: Challenges for Assisting with Menu Planning and Food Preparation, HCI, the 18th International Conference on Human-Computer Interaction, 2016.

Shinji Imahori, Shizuka Kawade, Yoko Yamakata, "Escher-Like Tilings with Weights," 18th Japan Conference on Discrete and Computational Geometry and Graphs, 2015. 査読有り

Takuya Kadowaki, Yoko Yamakata, Katsumi Tanaka, "Situation-Based Food Recommendation for Yielding Good Results," 7th Workshop on Multimedia for Cooking and Eating Activities (CEA2015), Full presentation, proc. of ICME2015, 2015. (CEA2015 Recognition of merit) 査読有り

Tetsuro Sasada, Shinsuke Mori, Tatsuya Kawahara and Yoko Yamakata, "Named Entity Recognizer Trainable from Partially Annotated Data," The 2015 Conference of the Pacific Association for Computational Linguistics (PACLING), 2015. 査読あり

Takuya Kadowaki, Yoko Yamakata, Shinsuke Mori, Katsumi Tanaka, "Recipe Search for Blog-type Recipe Articles Based on a User's Situation," Workshop on Smart Technology for Cooking and Eating Activities (CEA2014), Full paper, UbiComp '14 Adjunct Proceedings, pp. 497-506, 2014. 査読有り

Yoko Yamakata, Asuka Miyazawa, Atsushi Hashimoto, Takuya Funatomi, Michihiko Minoh, "A Method for Detecting Gaze-required action while Cooking for Assisting Video Communication," Workshop on Smart Technology for Cooking and Eating Activities (CEA2014), Short paper, UbiComp '14 Adjunct Proceedings, pp. 577-582, 2014. 査読有り

Atsushi Hashimoto, Tetsuro Sasada, Yoko Yamakata, Shinsuke Mori, Michihiko Minoh, "KUSK Dataset: Toward a Direct Understanding of Recipe Text and Human Cooking Activity," Workshop on Smart Technology for Cooking and Eating Activities (CEA2014), Short paper, UbiComp '14 Adjunct Proceedings, pp. 583-588, 2014. 査読有り.

Shinsuke Mori, Hirokuni Maeta, Tetsuro Sasada, Koichiro Yoshino, Atsushi Hashimoto, Takuya Funatomi and Yoko Yamakata, "FlowGraph2Text: Automatic Sentence Skeleton Compilation for Procedural Text Generation," INLG14, pp. 118-122, 2014. 査読有り

Shinsuke Mori, Hirokuni Maeta, Yoko Yamakata, Tetsuro Sasada, "Flow Graph Corpus from Recipe Texts," The International Conference on Language Resources and Evaluation Conference (LREC2014), pp.2370-2377 2014. 査読有り

[図書](計0件)

[産業財産権]

出願状況(計0件)

取得状況（計0件）

〔その他〕

ホームページ等

・ レシピ言語処理マニュアル（京都大学 学術情報メディアセンター 森研究室）
<http://www.ar.media.kyoto-u.ac.jp/how-to/recipe-NLP/>

・ 料理レシピ処理の研究紹介（東京大学 情報理工学系研究科 山肩研究室）
<https://sites.google.com/view/yy-lab/>

6 . 研究組織

(1)研究分担者

研究分担者氏名：今堀慎治

ローマ字氏名：IMAHORI, Shinji

所属研究機関名：中央大学

部局名：理工学部

職名：教授

研究者番号（8桁）：90396789

研究分担者氏名：森 信介

ローマ字氏名：MORI, Shinsuke

所属研究機関名：京都大学

部局名：学術情報メディアセンター

職名：教授

研究者番号（8桁）：90456773

(2)研究協力者

研究協力者氏名：John Carroll

ローマ字氏名：John Carroll

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。