

分類・検索・生成技術を用いた調理レシピの器具補完

井上 結由果¹ 飯島 千晴¹ 尾崎 知伸^{1,a)}

概要: 料理初心者にとって、様々な調理シーンに応じて必要な調理器具を選択することは容易ではない。本研究では、料理初心者に対する調理器具選択の支援を目的に、レシピテキストに記載されていない調理器具を補完することを考える。具体的には、多ラベル分類・類似レシピ検索・系列変換の各技術を用いて、レシピテキストから省略された調理器具を推定する手法を提案するとともに、実データを用いてその有用性を検証する。

Estimation of missing cooking tools in recipe texts using techniques on classification, information retrieval and sequence-to-sequence learning

INOUE YUYUKA¹ IJIMA CHIHARU¹ OZAKI TOMONOBU^{1,a)}

Abstract: It is difficult for cooking beginner having insufficient knowledge to judge the appropriate cooking tools necessary for several cooking situation. In order to assist such beginners, this paper presents a study to estimate and complement the cooking tools not mentioned in recipe texts. More concretely, we conduct comparative experiments on the missing tools estimation using three methods based on classification, information retrieval and sequence-to-sequence learning, respectively.

1. はじめに

近年のレシピサイトの発展は目覚ましく、利用者は多種多様な調理レシピを日常的に閲覧することが可能となっている。これらのレシピの中には、専門家によって監修された動画像を含む非常に理解しやすいものも含まれるが、多くのレシピは利用者によって投稿されたものであり、その質や完成度は均一ではない。また多くの場合、詳細な調理工程がすべて事細かに記載されるわけではなく、しばしば様々な省略が行われる。

例として、図 1 に示すハンバーグのレシピを考える。このレシピの手順 1 では、どんな調理器具を利用して玉葱をすりおろすのか、また炒めるために利用する器具は何なのかが明記されていない。手順 2 では、玉葱が炒められたことや冷めたことの判断基準が不明であることに加え、玉葱とひき肉とを何を使ってどの程度混ぜるべきなのかも明記されていない。手順 3 では、実際にはハンバーグのたねを

適当な分量に分けた上で両面を焼く必要があるが、焼くという行為のみが示されている。料理に慣れた人であれば、これまでの経験や知識を用いてこれらの省略事項を補完しながら問題なく料理を行うことができると考えられるが、料理初心者にとってはそれが難しく、省略があるレシピに沿って料理を行うことは容易ではない。すなわちこの例が示す通り、調理レシピにおける様々な省略は、料理初心者のレシピ利用の大きな妨げになっていると考えられる。

本研究では、料理初心者の支援を目的とし、調理レシピにおける省略事項の中でも特に調理器具に焦点を当て、その補完を試みる。これまでにも、調理器具の補完に関してはいくつかの研究が行われている。例えば [1] では、調理画像を利用した器具の推定手法が提案されている。また [2]

タイトル：シンプルなハンバーグ
手順 1：すりおろした玉葱を炒める。
手順 2：玉葱が炒められたら、冷ましてからひき肉と混ぜる。
手順 3：後は焼くだけ。

図 1 ハンバーグの調理レシピの例

¹ 日本大学文理学部情報科学科
〒156-8550 東京都世田谷区桜上水 3-25-40
^{a)} ozaki.tomonobu@nihon-u.ac.jp

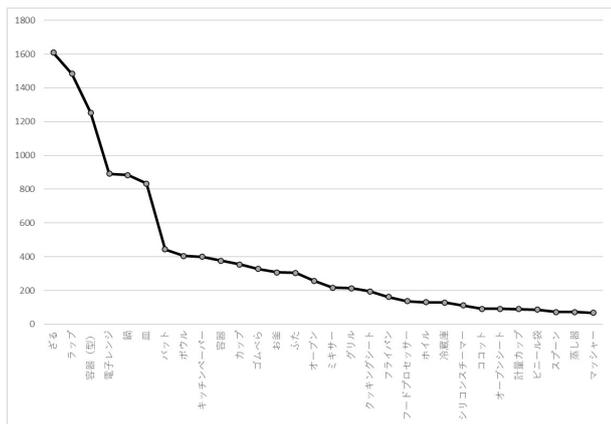


図 2 調理器具の出現回数

表 1 調理器具出現数の分布

最小値	Q1	中央値	Q3	最大値	平均値	分散
1	1	2	17	1,608	57.7	200.2

では、食材や調理動作と使用する調理器具との統計的な関連性が調査されている。これに対し本研究では、テキストのみから構成されるレシピを対象とした調理器具補完を考える。これにより、(動画画像を含まない) 多くのレシピを対象とすることができるとともに、調理工程に含まれる多様な情報を考慮した器具の補完が期待できる。

本論文の構成は以下のとおりである。2章では、本研究で用いるデータセットについて説明する。3章では、必要となる記法を導入した後、分類・類似検索・文書変換の各技術を用いた器具補完手法を提案する。4章で評価実験を行い、最後に5章でまとめと今後の課題を述べる。

2. データセットの構築

本研究では、国立情報学研究所の IDR データセット提供サービスによりクックパッド株式会社から提供を受けた「クックパッドデータセット」[3] に含まれるレシピのうち、条件 (1) 食材数：5 - 10, (2) 手順数：4 - 6, (3) 調理器具数：5 - 7 を満たすレシピを対象とした。

対象データセットは、4,200 件で構成され、約 200 種の調理器具が含まれている。もっとも出現回数が多い調理器具は“ざる”の 1,608 回であり、器具の出現回数の平均は 57.7 回である。出現数上位 30 件の調理器具を図 2 に示す。また各器具の出現回数の分布を表 1 に示す。

各レシピに対し、大学生 6 名によるアノテーションを行った。その際、アノテーションツール doccano [4] の系列ラベリング (SequenceLabeling) 機能を用い、調理器具にはラベル “tool”，隣接する修飾語や助詞等にはラベル “delete” をそれぞれ付与している (図 3 参照)。なお実際には、調理器具に加え、調理時間や火加減、状態に関するアノテーションも行っている。

アノテーション後、“tool” ラベルが付与された単語 (調



図 3 アノテーションの例

理器具) に対しては、適切な表記ゆれ対応を行っている。また、“tool” ラベルまたは “delete” ラベルが付与された部分を元のレシピから削除することで、調理器具が省略された自然なレシピが構築される。例えば図 3 の例からは、「卵を割り入れたら、混ぜます。砂糖・食用油を加えて、再び混ぜます。」というテキストが抽出される。このように、“tool” と “delete” の両ラベルを用いることで、調理器具を含むレシピからの器具特定と、器具を省略したレシピの生成を達成している。

3. 調理器具の推定

3.1 準備

手順数 r_L のレシピ $r = r_1, r_2, \dots, r_{r_L}$ を、各手順 (文字列) r_i を連結した文字列と捉える。各手順 r_i に含まれる調理器具の集合を $t(r_i)$ 、 r_i から調理器具とそれに付随する修飾語や助詞等を削除した文字列 (器具なし手順) を r_i^- とそれぞれ表記する。また、レシピ r に含まれる調理器具の集合を $t(r) = \bigcup_i t(r_i)$ 、 r に対する器具なしレシピを $r^- = r_1^-, r_2^-, \dots, r_{r_L}^-$ と定義する。一方、レシピの集合 R に対し、そこに含まれる調理器具の全体集合を $T = \bigcup_{r \in R} t(r)$ と表記する。また BERT [5] などの自然言語処理技術を用いて得られる文字列 txt の N 次元分散ベクトル表現を \vec{txt} と表記する。

以上の準備のもと、本研究では、器具なしレシピ r^- からその利用器具 $t(r)$ を推定する関数を構築することを目的とする。またそのための技術として、分類技術・類似検索技術・系列変換技術を利用する。

ところで実際のレシピにおいては、利用する調理器具がすべて省略されるとは限らない。また、前の手順で利用した調理器具をその後の手順で利用することも頻繁に行われる。このことを反映し、より現実的な調理器具補完を行うための第一歩として、本研究では手順を単位とした器具補完も考える。すなわち、レシピ r の第 i 手順のみを器具なし手順に置き換えたレシピを $r^{-i} = r_1, \dots, r_{i-1}, r_i^-, r_{i+1}, \dots, r_{r_L}$ と定義し、 r^{-i} から $t(r_i)$ を推定する関数の構築を行う。

3.2 分類技術に基づく調理器具の推定

本研究では、分類技術に基づく調理器具の推定方法として、器具なしレシピのベクトル \vec{r}^- から利用器具 $t(r) \subseteq T$ を予測する関数を構築することを考える。実際には、与えられたレシピの集合 R から訓練例集合 $\left\{ \left\langle \vec{r}^-, t(r) \right\rangle \mid r \in R \right\}$ を構築し、既存のマルチラベル分類器を用いて関数 $f: \mathbb{R}^N \rightarrow 2^T$ を構築する。

3.3 類似検索技術に基づく調理器具の推定

本研究では、類似するレシピ間では利用する調理器具も類似する傾向があると考え、第二の調理器具推定手法として、類似検索技術の利用を考える。具体的には、既存レシピの集合を R 、調理器具の補完を行いたいレシピを $q \notin R$ としたとき、分散表現ベクトルを通じて得られる R 中の q との類似レシピ上位 K 件

$$\text{sim}(q, R, K) = \{r \in R \mid |\{s \in R \mid \cos(\vec{q}^-, \vec{s}^-) > \cos(\vec{q}^-, \vec{r}^-)\}| < K\}$$

を用いて、 q で必要とされる調理器具 $t(q)$ を

$$\bigcup_{r \in \text{sim}(q, R, K)} t(r)$$

と推定する。ここで K は利用者によって与えられるパラメータである。

一方、手順を単位とする調理器具補完では、レシピの類似性に加え、手順の特徴を利用する。具体的には、補完対象手順 q_i に対し、(1) q_i と同じ調理動作（動詞）を含む、または (2) 分散表現ベクトルが \vec{q}_i と最類似する、という条件を満たす手順 r_j *s.t.* $r \in \text{sim}(q, R, K)$ をそれぞれを抽出し、そこに含まれる調理器具 $t(r_j)$ を推定結果とする。レシピ全体の類似性と手順の整合性の両方を考慮することで、より正確な器具推定が期待できる。

3.4 系列変換技術に基づく調理器具の推定

本研究における第三の調理器具推定手法として、系列変換技術 [6] を利用し、器具が省略されたレシピ r^- から、直接器具が付与されたレシピ r を生成することを考える。すなわち、レシピ集合 R から訓練例集合 $\{(r^-, r) \mid r \in R\}$ を構築し、これを用いて変換モデルを構築する。

また同様に、 R 中の各レシピの各手順を一つずつ器具なし手順に置き換えることで訓練例集合 $\{(r^-, r) \mid r = r_1, \dots, r_{r_L} \in R, 1 \leq i \leq r_L\}$ を構築し、手順に着目した変換モデルを構築する。

4. 評価実験

4.1 実験設定

3章で導入した3種の調理器具推定手法を評価するため、2章で構築した4,200件のレシピを訓練例集合(80%)とテスト例集合(20%)に分け、評価実験を行った。なお、分散表現ベクトルの獲得には、日本語 Wikipedia データを用いて訓練された東北大学乾研究室による日本語 BERT モデル*1 を用いた。またマルチラベル分類器は scikit-learn の MultiOutputClassifier (推定器は RandomForest) を、系列変換 (seq2seq) では GRU (Gated recurrent unit) [7] を基本ユニットとし、注意機構を導入している。

*1 <https://github.com/cl-tohoku/bert-japanese>

各推定手法の性能を、器具・レシピ・手順に関する再現率、適合率および調和平均である F 値を用いて評価した。ここで、テストレシピの集合を R 、調理器具の集合を T 、入力 x に対するモデル f の推定結果を $f(x)$ としたとき、器具に関する再現率、適合率は、レシピを基準に以下のように定義される。

$$R_{\text{tool}} = \frac{1}{|T|} \sum_{t \in T} \frac{|\{r \in R \mid t \in c(r) \wedge t \in f(r^-)\}|}{|\{r \in R \mid t \in c(r)\}|}$$

$$P_{\text{tool}} = \frac{1}{|T|} \sum_{t \in T} \frac{|\{r \in R \mid t \in c(r) \wedge t \in f(r^-)\}|}{|\{r \in R \mid t \in f(r^-)\}|}$$

また、手順を基準とする場合は

$$R_{\text{tool}}^s = \frac{1}{|T|} \sum_{t \in T} \frac{|\{r_i \mid r \in R \wedge t \in c(r_i) \wedge t \in f(r^-)\}|}{|\{r_i \mid r \in R \wedge t \in c(r_i)\}|}$$

$$P_{\text{tool}}^s = \frac{1}{|T|} \sum_{t \in T} \frac{|\{r_i \mid r \in R \wedge t \in c(r_i) \wedge t \in f(r^-)\}|}{|\{r_i \mid r \in R \wedge t \in f(r^-)\}|}$$

と定義される。一方、レシピに関する再現率・適合率を

$$R_{\text{recipe}} = \frac{1}{|R|} \sum_{r \in R} \frac{|\{t \in t(r) \mid t \in f(r^-)\}|}{|t(r)|}$$

$$P_{\text{recipe}} = \frac{1}{|R|} \sum_{r \in R} \frac{|\{t \in t(r) \mid t \in f(r^-)\}|}{|f(r^-)|}$$

と定義し、同様に、手順に関する再現率・適合率を以下のように定義する。

$$R_{\text{step}} = \frac{1}{|\{r_i \mid r \in R\}|} \sum_{r_i \text{ s.t. } r \in R} \frac{|\{t \in t(r_i) \mid t \in f(r^-)\}|}{|t(r_i)|}$$

$$P_{\text{step}} = \frac{1}{|\{r_i \mid r \in R\}|} \sum_{r_i \text{ s.t. } r \in R} \frac{|\{t \in t(r_i) \mid t \in f(r^-)\}|}{|f(r^-)|}$$

4.2 結果と考察

器具に関する実験結果を表 2 に、レシピに関する実験結果を表 3 にそれぞれ示す。また、 $K = 10$ とした場合の手順に関する実験結果を表 4 に示す。

識別技術では、全体的に F 値が低く、出力された器具の

表 2 器具に関する再現率・適合率・F 値

	識別	類似検索					系列変換
		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	
レシピを単位とした場合							
R	0.009	0.02	0.01	0.01	0.01	0.01	0.17
P	0.002	0.02	0.03	0.04	0.05	0.06	0.21
F	0.004	0.02	0.02	0.02	0.02	0.02	0.18
ステップを単位とした場合							
R	-	0.15	0.26	0.34	0.40	0.45	0.21
P	-	0.15	0.12	0.11	0.10	0.09	0.27
F	-	0.15	0.17	0.16	0.15	0.14	0.23

表 3 レシピに関する再現率・適合率・F 値

識別	類似検索					系列変換	
	k = 1	k = 2	k = 3	k = 4	k = 5		
R	0.03	0.15	0.26	0.34	0.40	0.45	0.23
P	0.09	0.15	0.12	0.11	0.10	0.09	0.28
F	0.05	0.15	0.17	0.16	0.15	0.14	0.25

表 4 手順に関する再現率・適合率・F 値

	類似検索		系列変換
	手順	動作	
R	0.07	0.16	0.27
P	0.09	0.19	0.33
F	0.08	0.18	0.29

表 5 補完精度の高い器具の例

分類	類似検索 (手順)		類似検索 (動作)	
フライパン	0.27	オープン	0.08	電子レンジ 0.68
鍋	0.22	ゴムベラ	0.08	蒸し器 0.40
オープン	0.14	ふた	0.08	フライパン 0.32
ボウル	0.10	電子レンジ	0.07	オープン 0.32
冷蔵庫	0.07	クッキングシート	0.03	冷蔵庫 0.22

類似検索 (レシピ)	系列変換 (レシピ)	系列変換 (手順)
ボウル 0.27	鍋 0.29	フライパン 0.38
フライパン 0.26	電子レンジ 0.29	オープン 0.37
鍋 0.23	オープン 0.27	ゴムベラ 0.36
皿 0.21	ラップ 0.26	ざる 0.35
電子レンジ 0.20	ボウル 0.25	ボウル 0.35

種類も 5 種類と少なかった。この結果から、出現頻度の高くない器具を使用するレシピの学習が足りていなかったことが考えられる。検索技術では、手順に着目して検索するよりも、動作に着目して検索する方が高い F 値が得られた。また、包丁やまな板などの使用頻度の高い器具が上位に挙がらなかったことから、使用頻度が高いことで器具を示さずに動作のみで書かれていることが多い器具の検索が難しいことが考えられる。生成技術では、識別や検索と比べ出力された器具の種類は約 50 種と多く、値も一定の推定が行えたと言える。しかし、文章の再現を目標にしていたが、文章としての完全再現は難しく調理器具の補完のみ行われた。これらの結果として、出力された器具の種類と F 値より本研究の提案手法のうち最も有効的な調理器具補完方法は生成技術であると考えられる。

次に、比較的良好な精度で補完を行うことのできた器具について考察する。各設定における精度上位 5 件の器具を表 5 に示す。表より、フライパンやオープン、ボウル、ゴムベラが、複数の設定を通じて精度よく補完されていることが分かる。これらは比較的頻度の高い調理器具であり、頻度が低い器具の推定が困難であることが伺える。

最後に、系列変換によって得られたレシピの一例を図 4 に示す。この例では、入力レシピに温度表現 (°C) や “し

入力 170° c に温め材料全部入れて混ぜる。天板にしき、生地をおとす。生地はちょっとゆるめですが、手で丸めても ok です。15分焼く。

↓

出力 オープンを 170° c に温めるボウルに材料全部入れて混ぜる。天板にオープンシートをしき、オープンシートをしき生地をスプーンでおとす。生地はちょっとゆるめですが、手で丸めても ok です。オープンで 15分焼く。

図 4 系列変換技術によって生成された手順の例

く” という動作が含まれており、それを鍵に調理器具である “オープン” や “オープンシート” が生成されていると考えられる。また生成された手順は、一部冗長な表現を含んではいるが、推定された延べ 6 個の調理器具間には一定の整合性が確認できる。

5. まとめ

本研究では、料理初心者の支援を目的に、レシピテキスト中で省略されている調理器具の推定について議論を行った。また実レシピを対象に、分類・類似検索・系列変換の各種技術に基づく推定手法を適用し、その精度を検証した。その結果、特に系列変換技術を用いることで、いくつかの調理器具に対しては一定の精度での推定が行えることが確認された。しかしその一方で、全体としては必ずしも十分な推定精度は得られなかった。

今後の課題としては、推定精度を向上させるために、特に低頻度な器具に対する学習手法の工夫が必要であると考えている。加えて、複数の推定技術の併用や、マルチモーダル情報の利用などを検討している。

謝辞

本研究では、国立情報学研究所の IDR データセット提供サービスによりクックパッド株式会社から提供を受けた「クックパッドデータセット」を利用した。

参考文献

- [1] Y. Zhang, Y. Yamakata, and K. Tajima : Complementation of Food and Tool Information in Multi-Modal Recipe Procedural Descriptions, 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, B8-2, 2020.
- [2] 永井 廉人, 大杉 隆文, 松下 光範 : 料理レシピ中に記載されていない調理器具の推定に関する基礎検討, 第 32 回人工知能学会全国大会, 203-OS-24b-01, 2018.
- [3] クックパッド株式会社 : クックパッドデータ, 国立情報学研究所情報学研究データリポジトリ (データセット), <https://doi.org/10.32130/idr.5.1>, 2015.
- [4] H. Nakayama, T. Kubo, J. Kamura, Y. Taniguchi, and X. Liang : doccano: Text Annotation Tool for Human, <https://github.com/doccano/doccano>, 2018.
- [5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova : BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv:1810.04805 [cs.CL], 2018.
- [6] I. Sutskever, O. Vinyals, and Q. V. Le : Sequence to

Sequence Learning with Neural Networks, *Proc. of the 27th International Conference on Neural Information Processing Systems*, pp.3104–3112, 2014.

- [7] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio : Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, arXiv:1406.1078 [cs.CL], 2014.