

複数料理レシピの自動要約

Automatic Summarization of Multiple Cooking Recipes

重田 識博

Norihiro Shigeta

広島市立大学大学院 情報科学研究科

Graduate School of Information Sciences, Hiroshima City University

概要 本研究では、ある料理に関する複数のレシピを比較して、その料理の典型的な手順を出力するシステムを構築する。ある料理の典型的な手順は、複数のレシピの手順を比較し、その共通部分を同定することで実現できる。ここで、料理レシピでは、しばしば項が省略される上に、同じ材料や調理器具であっても、レシピによって表記が異なるという問題がある。これらの問題を解決するため、前者の問題については述語項構造解析技術を、後者については自動同義語構築技術を用いる。本研究では、典型的な手順の出力を複数テキスト要約ととらえ、テキスト要約の評価手法に基づいて、提案手法を評価する。

1 はじめに

近年、Cookpadなどのユーザ投稿型レシピサイトの利用が盛んになり、数多くのレシピを検索するようになってきている。このようなサイトでは、ある料理に対してユーザ毎に材料の表記や手順の順序が異なるレシピが大量に存在する。このため、例えば、料理スキルの低いユーザが基本的なレシピを探すのが必ずしも容易ではない、という問題がある。そこで、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手順を出力するシステムを構築する。このようなシステムが実現できれば、本システムで出力する典型手順と個々のレシピを比較し、その差異を抽出することで、個々のレシピの特徴を検出することが可能になる。

2 システムの概要

本研究では、ある料理に関する複数のレシピを入力とし、その料理の典型的な手順と食材リストを出力する。その際、述語項構造解析と同義語抽出技術を用いる。

述語項構造解析とは、ある文書中に省略された項がある場合、述語に対応する項を補完する処理のことを指す。料理レシピでは、しばしば項の省略があり、このため複数のレシピを比較する上で、この問題を考慮する必要がある。同じ料理でも、表現が異なる場合があるため、複数のレシピを比較する際、あらかじめ同義語辞書を構築しておく必要がある。3節では、述語項構造解析について、4節では、同義語辞書構築方法について、それぞれ述べる。

3 述語項構造解析

本研究では、Moriら[1]-[4]が開発した料理レシピ用述語項構造解析ツールを利用することで、項を省略する問題に対処する。Moriらのツールでは、まず、レシピ中の食材、

動作、調理器具などの固有表現を自動抽出し、次に、述語項構造解析を行う。

図1は、親子丼のレシピの手順リストを示している。図において、手順5の「材料が煮えたら、卵を半分流し入れ」では、どこに流し入れるかの記載はない。ここで、Moriらの述語項構造解析手法を用いて図1の手順テキストを解析すると、図2のような結果が得られる。この図の点線部を見ると、「材料が煮えたものに」や「材料が煮えたものと一緒に」という補完が行われていることがわかる。

1. 鶏肉、三つ葉は適当な大きさに切り、玉ねぎは薄切りに切ります。
2. 醤油・みりん・だし汁をお鍋でひと煮立ちさせる。
3. 煮汁に鶏肉と玉ねぎを加えます。
4. 卵は混ぜずに、お箸でサッと切るようにくずした状態にする。
5. "材料が煮えたら、卵を半分流し入れ、少し煮たら残りの卵も流し入れ、
6. 三つ葉をちらし、卵が半熟な状態で火をとめます。"
7. ご飯に盛ってできあがり〜♪

図1: 親子丼の手順テキスト 1

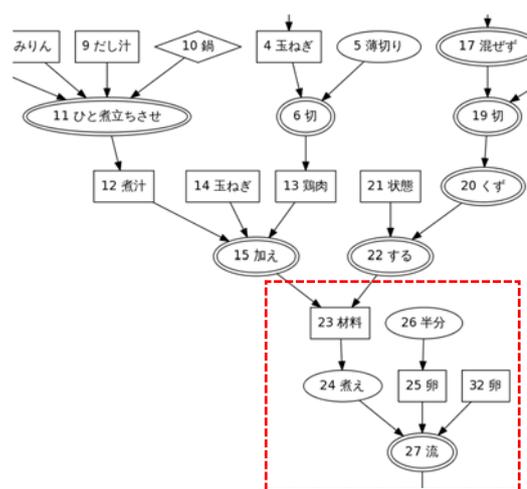


図2: 手順テキスト1の述語構造解析結果の一部

4 同義語抽出

図3は、図1とは別の親子丼の手順テキストである。図3の手順1は、図1の手順1に対応するが、これらに対応関係にあることを自動的に認識するためには、図3手順1の「鶏モモ肉」と図1手順1の「鶏肉」がほぼ同義語の関係にあるという知識が必要になる。

1. 玉ねぎは薄切り、鶏モモ肉は一口サイズに切る
2. 玉ねぎ、鶏モモ肉、水、麺つゆをフライパンに投入♪味の濃さは調整して下さい。水と麺つゆを1:3で薄めています
3. 具材に火が通ったら、水溶き片栗粉を加えてトロミをつける。 餡掛けみたいな感じで
4. 卵をまわし入れて、少しかき混ぜる。 フワトロになります(*^^*) 出来上がり*(^o^)/*

図 3: 親子丼の手順テキスト 2

本研究では、Chung[5]の提案した同義語抽出を用いる。Chung は、各料理レシピの食材リストの一番上に記載される食材は、その料理のメイン食材である、という仮定をもとに、例えば「エビ」料理カテゴリに分類されているレシピの食材リストから「エビ」の同義語を抽出している。Chung は楽天レシピデータベースで提案手法の有効性を確認しているが、この手法をクックパッドに適用し、同義語辞書を構築した。実際に構築した辞書の一部を図 4 に示す。

```
<CATEGORY>たけのこ</CATEGORY>
ゆでたけのこ 0.962 たけのこの水煮 0.909 茹で筍 0.789
タケノコ 0.784 たけのこ 0.772 筍 0.755 茹でたけのこ 0.737
竹の子 0.720
<CATEGORY>鶏もも肉</CATEGORY>
鶏もも 0.949 鶏もも肉 0.929 鳥モモ肉 0.923 鳥もも肉 0.923
鶏肉 0.921 とりもも肉 0.920 鶏モモ肉 0.892 鶏もも肉 0.636
<CATEGORY>あい挽き肉</CATEGORY>
合い挽きミンチ 0.750 合びき肉 0.727 合挽きミンチ 0.636
合いびき肉 0.609 合挽き肉 0.596 ひき肉 0.583 あいびき肉 0.571
合い挽き肉 0.568
<CATEGORY>ぶり</CATEGORY>
ブリ切り身 1.000 ぶり切り身 1.000 ぶりの切り身 0.973 ぶり 0.957
ブリ 0.950 ブリの切り身 0.938 鱈 0.909 ぶりのあら 0.882
```

図 4: Chung の手法を用いてクックパッドデータベースから構築した同義語辞書の一部

図の赤字はカテゴリ名、青字は同義語を示している。図から、カテゴリ名が「たけのこ」などの食材の場合には、良質な同義語が収集できていることがわかる。

5 評価指標

国立情報学研究所が主催する評価ワークショップ NTCIR で実施されたテキスト要約に関するワークショップ Text Summarization Challenge (TSC) では、複数テキスト要約[6]に関するタスクがある。このタスクでは、複数のテキストから要約システムが抽出した重要文を、以下に示す 2 つの尺度：精度(Precision)と被覆率(Coverage)で評価している。

$$\text{精度} = \frac{\text{最小カバーと抽出文書に含まれる文数}}{\text{抜粋した文書の長さ}} \quad (1)$$

$$\text{被覆率} = \frac{\text{抽出文書集合が要約文章をカバーしている割合}}{\text{抜粋による要約文書数}} \quad (2)$$

本研究でも、この評価尺度を用いて評価するため、TSC の枠組みに沿ってデータセットを構築する。

6 まとめ

本研究では、ある料理に関する複数のレシピを比較して、その料理の典型的な手順と食材リストを出力するシステムを構築する。複数のレシピを比較する際、述語項構造解析と同義語抽出の 2 つの要素技術を用いる。

参考文献

- [1] Hirokuni Maeta, Tetsuro Sasada, and Shinsuke Mori, “A Framework for Procedural Text Understanding” IWPT, 2015.
- [2] Shinsuke Mori, Tetsuro Sasada, Yoko Yamakata, and Koichiro Yoshino, “A Machine Learning Approach to Recipe Text Processing” Cooking with Computers Workshop, 2012.
- [3] Graham Neubig, Yosuke Nakata, and Shinsuke Mori, “Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis” ACL-HLT, 2011.
- [4] Shinsuke Mori and Graham Neubig, “A Pointwise Approach to Pronunciation Estimation for a TTS Front-end” Interspeech, 2011.
- [5] Young-joo Chung, “Finding Food Entity Relationships Using User-generated Data in Recipe Service” CIKM, 2012.
- [6] 平尾 努, 奥村 学, 福島 孝博, 難波 英嗣, 野畑 周, 磯崎 秀樹, “抜粋による複数文書要約を評価するためのコーパスと評価指標” 情報処理学会論文誌データベース, Vol.48, No.SIG 14, pp.60-68, 2007.