

機械学習を用いた実用的レシピの判別手法の提案

小野 滉太

近年、クックパッドや楽天レシピなどに代表される、ユーザ投稿型の料理レシピサイトが、急速に普及している。それに伴い、投稿されるレシピの数も膨大になり、ユーザにとっては、レシピの選択が難しくなっている。さらに、せっかくレシピを選択しても、ユーザにとって分かりづらいレシピを選択してしまい、調理がスムーズに進まないという状況がしばしば起こっている。

その状況を解消するため、本研究では、“実用的レシピ”の判別を行う。“わかりやすいレシピの書き方”という雑誌記事とレシピの書き方についての関連研究を参考にして、実用的レシピとは、“分かりやすい”、“簡潔である”、“内容がしっかり書かれている”、“コツやポイントが詳細に記述されている”という4つを満たすレシピであると定義した。実用的レシピを判別することでユーザは、膨大な数のレシピの中から、調理しやすいレシピのみを選択することが可能になる。これにより、レシピ選択の支援とスムーズな調理の支援を同時に行うことが本研究の目的である。

まず、使用する食材や調理法のバランスを考慮して選択した、“肉じゃが”、“親子丼”、“カルボナーラ”、“クリームシチュー”、“エビチリ”、“フレンチトースト”、“かぼちゃの煮物”、“麻婆豆腐”、“ポテトサラダ”、“スイートポテト”という10種類の料理について、クックパッドの人気順ランキング上位のレシピの分析を行い、人気レシピが持つ特徴を明らかにした。人気レシピの特徴には、レシピの分かりやすさや簡潔さなど、本研究で判別しようとしている実用的レシピの特徴と共通する部分が多い。そのため、人気レシピの分析で明らかになった特徴を、実用的レシピの判別に用いることを考えた。その結果得られた、“火加減の記載”や“調理時間の記載”など、レシピの内容面における7つの特徴を、実用的レシピの判別の素性として用いることを決定した。そして、それらを素性としたSVMによる判別器の実装を行った。

実験として、前述の10種類の料理それぞれについて、クックパッドの人気順ランキング上位・下位各500件ずつ、合計1,000件のデータを用いて、10分割交差検定を行った。その結果、精度・再現率・F値ともに最も高かった料理は、エビチリで、精度が0.633、再現率が0.626、F値が0.621であった。また、10種類の料理のF値の平均は、0.586という結果になった。

今回の実験では、提案手法に基づき実装を行った実用的レシピの判別器が、本研究で定義された実用的レシピを判別できていることを確認した。特に、現在のクックパッドの人気順ランキングにおいては下位に位置しているレシピの中にも、実用的レシピが存在しており、提案手法ではそれを判別できているという点は大きな成果である。また、実験を通して、10種類の料理すべてが、7つの素性全部を必要としているわけではないということが分かった。今後は、料理によって素性を工夫し、それぞれの料理に合わせた実用的レシピの判別を行っていこうと考えている。

(指導教員 関 洋平)