

調理手順と料理画像を用いた重複レシピ判別手法の提案

小邦 将輝

ユーザが投稿したレシピを掲載する「ユーザ投稿型のレシピサイト」には、同一の料理であっても、異なった手順や材料を用いて作成されたレシピが掲載されており、主婦や独り暮らしの学生など、多様なユーザが利用している。しかし、異なるレシピであるにも関わらず、調理手順が完全に一致しているレシピや、調理手順の一部が変更されている場合でも、同一のレシピとして判断できるレシピ（重複レシピ）が存在する。レシピサイトは多様なユーザが利用することから、検索結果の多様性（ダイバーシティ）が要求されるため、このようなレシピの存在は有用でない。そこで、本研究では調理手順と料理画像を用いた重複レシピの判別手法を提案する。

提案手法では、任意の1レシピを重複レシピ候補として、重複レシピ候補より投稿時間が前のレシピとの間で、調理手順の文字 3-gram 集合の Jaccard 係数（調理手順類似度）を計算する。重複レシピ候補に対して、最も調理手順類似度が高くなったレシピをオリジナルレシピ候補とし、重複レシピ候補とオリジナルレシピ候補のペアを重複レシピペア候補とする。そして、重複レシピペア候補の調理手順類似度、もしくは料理画像から抽出した特徴ベクトルのコサイン類似度（料理画像類似度）が閾値を超えた場合に重複レシピペアと判別する。

提案手法の有効性を検証するために、3つの比較手法を用いて評価実験を行った。実験では、1時間に10件以上20件未満のレシピの投稿を行ったユーザを「ユーザ群A」、20件以上30件未満のレシピの投稿を行ったユーザを「ユーザ群B」、30件以上のレシピの投稿を行ったユーザを「ユーザ群C」として、提案手法と3つの比較手法を適用した結果を比較した。ここで、提案手法は、調理手順と料理画像を用いて重複レシピを判別する手法、比較手法1は調理手順のみを用いて重複レシピを判別する手法、比較手法2は料理画像のみを用いて重複レシピを判別する手法、比較手法3はすべての重複レシピ候補を重複レシピと判別する手法である。評価実験の結果、提案手法はいずれの場合でも、比較手法1、比較手法2を上回る結果となり、提案手法と比較手法1の間ではt検定（有意水準5%、片側検定）で有意差が見られた。この結果より、重複レシピを判別する際には、調理手順と料理画像の双方を考慮することが有効であるといえる。また、提案手法はユーザ群B、Cに対して精度よく判別できるなどの差が見られたことから、投稿時間間隔を考慮することで判別精度を向上させる可能性があることを明らかにした。

投稿時間間隔を考慮することによる効果を調査するために、機械学習手法を用いて重複レシピの分類実験を行った。投稿時間間隔を考慮した場合、SVMを用いた際にF値0.728、Random Forestを用いた際にF値0.723となり、投稿時間間隔を考慮しない場合、SVMを用いた際にF値0.721、Random Forestを用いた際にF値0.712となった。この結果は、投稿時間間隔を考慮することの有効性を示唆している。

今後の課題として、調理手順内で言い換えや書き換えが行われた場合に頑健に重複レシピを判別できるよう、ニューラルネットワークを用いて重複レシピを判別する手法を検討している。

（指導教員 関 洋平）