

頻出部分グラフに基づく レシピフローグラフの分散表現の獲得

二宮 あかり

要旨

料理レシピ投稿型サイトであるクックパッドは、レシピの投稿が容易な点、自身のホームページを開設可能な点等の特徴から多数のユーザによって利用されている。類似レシピの検索や代替食材の提案など、ますます高度になる利用者の要求に精度良く応えるためには、蓄積された各レシピの特徴を抽出するとともに、大規模かつ多様なレシピデータ全体を分類・構造化することが必要となる。これまでに、調理手順をグラフ構造化することが提案されている。その一つであるレシピフローグラフは、食材や動作を表す単語列を頂点とし、頂点間の関係性を辺とすることで調理手順の意味内容を構造化しており、これにより物体-動作間や、動作-動作間の関係を明示的に捉えることが可能である。

レシピフローグラフに対する分散表現を獲得し、各レシピを実数ベクトル空間に写像することは、より詳細な手順に着目したレシピデータの分類・構造化に繋がると考えられる。一般的なグラフデータの分散表現獲得に関しては種々の手法が提案されているが、本研究では特に、料理においては手順の断片が重要であると仮定し、主要な手順の選択とそれらを通じたベクトル化を提案する。

提案する枠組みを評価するために、クックパッドデータセットを対象に、レシピフローグラフおよび調理手順文書それぞれの分散表現を獲得し、比較実験を行った。具体的には、オムライス 4013 件、コロッケ 7734 件、ハンバーグ 12934 件のレシピを対象とし、多次元尺度法による視覚化、レシピ間距離の相関係数の算出、料理種の分類、代替食材発見の各タスクを通じ、分散表現の妥当性を検証した。