

言語処理学会（2022 年 3 月）

一般発表 G1-1

『レシピ分野における動作対象の状態変化を考慮したデータセットの構築と検索モデルの提案』

白井 圭佑、橋本 敦史、牛久 祥孝、栗田 修平、亀甲 博貴、森 信介

○概要

言語的な指示を理解し行動するためには、その行動によって環境が現在の状態からどのように変化するかを予測できなければならない。そこで、この研究では、料理ドメインにおいてこの問題を調査するために、goal state identification by retrieval (GSIR) タスクを提案している。このタスクでは、視覚的な現在の状態と言語的な対象物と動作の情報を入力とし、動作後の視覚的な状態の検索を行っている。また、日本語の 200 レシピを基にデータセットを作成し、そのデータセットを用いて共有埋め込み空間の学習によって GSIR タスクを解いている。さらに、ベースラインとレシピフローを利用した発展的なモデルも提案し、これらの定量的な評価を行っている。

○主張

実験の結果から、動作前の画像のみを用いて検索した場合でもある程度の検索が可能であるが、3 項組もしくはレシピフローと合わせて用いたほうがより高い精度を実現する。さらに、動作前の画像と 3 項組、レシピフローの 3 つを合わせて用いた場合に最も精度が良くなったため、BERT 等を用いて文書上のコンテキストを考慮した表現なども利用すればより精度の向上が見込まれる。

○感想

自然言語から動作前の状態と動作後の状態を結びつけることができれば機械が言葉の意味を理解することにつながると思うため、ほぼ確実に成功するようにできればさらなる技術の進歩となると思います。完全に自立した機械の実現につながるこの研究は興味深いです。また、レシピという多くのデータが存在する料理と結びつけて研究するという発想が面白いと思いました。

○次に読むべきもの

- ・ GSIR タスクに関する書籍
- ・ 自然言語処理の構造を理解できる論文や書籍
- ・ 画像処理に関する論文や書籍