4W-02

動画データからの暗黙知抽出のための質問文自動生成手法の提案

多田　一真†　　村上　幸一‡

香川高等専門学校　創造工学専攻†　電気情報工学科‡

１．はじめに

近年，日本の農業においては，農業就業人口の減少と高齢化により，新規就農者が熟練営農者から直接指導を受ける機会が減少している．それに伴い，農作業マニュアルの整備がより一層重要となっている．そこで本研究グループでは，新規営農者への技術継承を目的として，レトロスペクティブレポート法による熟練営農者の暗黙知の表出に関する研究を行ってきた[1-3]．ここではインタビューアーのヒアリングスキルが重要であるが，ヒアリングスキルのばらつきやヒアリング自体の負担があった[1]．そこで，ヒアリングの負担軽減を目的に自動で暗黙知を抽出する手法を検討した．具体的にはインスタンスセグメンテーションアルゴリズムの一つである Mask R-CNN(Mask Region-with Convolution Neural Network) を用いて映像内の物体を明確にし，アイカメラの視線データと重ね合わせることにより，熟練営農者がその時何を見ていたかを自動で文章化させる手法について検討・解析[2,3]した．しかし，熟練農業者が眼鏡をかけている場合，眼鏡越しのアイカメラによる視線データの取得は，欠損値が多かった．

２．目的

本研究では，熟練営農者の作業動画から質問文を自動生成することで，レトロスペクティブレポート法を行う際のインタビューワーのヒアリング補助を行う手法について提案する．最終的には，熟練営農者の作業動画から熟練営農者の暗黙知を自動抽出することを目指す．

３．要素技術

本研究で使用した要素技術について説明する．

3.1 画像キャプション生成

画像キャプション生成とは，入力した画像を注釈する文章を生成するタスクである．モデルの入力として画像，出力として文章の word id 等が得られる．基本的に Encoder-Decoder モデルが使用されている．画像キャプション生成のためのデータセットとしては通称 MSCOCO と呼ばれる Microsoft の Common Objects in Context や Flicker 30k がある．

3.2　text2text-generation

text2text-generation は文章から別の文章を生成するタスクである．このタスクには機械翻訳や文章要約等のタスクが含まれている．特にコンテキストと回答から質問文を生成するタスクについて説明する．例えば“Manuel has created RuPERTa-base with the support of HF Transformers and Google”というコンテキストを考える．そして，解答として Manuel を与えると，出力する質問文として”Who created the RuPERTa-base?” という質問文を生成するタスクである．

４．　提案手法

提案手法について説明する．

4.1 概要

熟練営農者の動画データから任意の間隔で静止画をサンプリングする．サンプリングした静止画を画像キャプション生成モデルに入力し，得られた注釈文章を解析することで熟練営農者が視線を置いている物体を推定する．そして，得られた注釈文章を text2text モデルに入力することでレトロスペクティブ法のヒアリングに使用する質問文を生成することができる

4.2 画像キャプション生成モデル

本研究で画像キャプション生成を行ったモデルについて説明する．使用したモデルは Transformer ベースの画像キャプション生成モデルの Vision Transformer である．具体的な処理について説明する．まず，入力された画像

Proposed of automatic question generation method for extracting tacit knowledge from video data

†TADA, Kazuma・National Institute of technology, Kagawa College Faculty of Advanced Engineering

‡MURAKAMI, Yukikazu・National Institute of technology, Kagawa College Department of Electrical and Computer Engineering

はパッチに分け，1 次元のベクトル化する．こうして得られた 1 次元のベクトルを通常の Transformer のアーキテクチャに入力することで，Transformer の Decoder の出力として注釈文章の word id のリストを得ることができる．事前学習として MSCOCO の 2017 年データセットで学習済みの Vision Transformer を使用した．

4.3 質問文生成

本研究でコンテキストと回答を入力に質問文を出力するモデルとして通称 T5 と呼称される Text-To-Text Transfer Transformer を用いた．このモデルは，transformer をベースとしたものである．転移学習に使用したデータセットは Stanford Question Answering Dataset (SQuAD) である．このデータセットはウィキペディアの記事に対してクラウドワーカーが投げかけた質問からなる読解データセットである．

5 結果

熟練営農者の作業動画からサンプリングした 1 枚の静止画に対して提案手法を施行した結果について示す．下記にサンプリングした画像を図１として示す．作業場でパック詰め途中のイチゴと籠に入れられたイチゴが机に置いてあるのが分かる．

図１　入力画像

この画像を画像キャプションモデルに入力すると「a table topped with lots of fruit and vegetables」という文章が生成された．この文章は直訳すると「たくさんの果物や野菜がのったテーブル」という意味になる．この文章をコンテキストとして，コンテキスト内の単語「of」を解答としてさらに質問生成モデルに入力すると「What kind of table is topped with lots of fruit and vegetables?」という質問文が生成された．この質問文を直訳すると「果物や野菜がたくさん盛られたテーブルとはどんなものだろう？」 という意味になる．このようにして実際の作業画像に対して提案手法を使用することにより自動で熟練営農者の認識物の候補とその状況における質問文を生成することができた．

6 結言

本研究では画像キャプション生成と質問生成によるレトロスペクティブ法と暗黙知抽出の補助を試みる手法の提案とデモンストレーションを行った．結果として熟練営農者の認識物の候補の列挙と質問文生成を行えた．しかし，認識物の候補の列挙，質問文の生成の両方において暗黙知抽出に至るほどの精度を得られなかった．

今後は，認識物の候補の列挙に関しては転移学習を行うことにより具体性のある文章生成を行い，認識物を断定できるようになると考えている．質問文の生成においては，画像から直接的に質問文を生成するモデルを使用することにより提案手法で情報落ちしていた部分を補えるようにしようと考えている．

参考文献

[1] 渡邊修平，藤井宏次朗，村上幸一：アイカメラを用いた農作業技術継承マニュアルの提案，電子情報通信学会技術研究報告，pp.297-300(2014)

[2] 笠松 雅史，村上 幸一：Mask R-CNN を用いたアイカメラ映像解析手法の提案，人工知能学会全国大会論文集，第 34 回(2020)，セッション ID: 1M4-GS-13-03

[3] 平田 結愛，笠松 雅史，村上 幸一，脇坂颯：視線データと動画注釈システムを用いた農作業技術継承マニュアル作成手法の提案，工知能学会全国大会論文集，第 35 回(2021),セッション ID: 4I1-GS-7b-03 (2021)