【P2-10】動画キャプションモデルを用いた字幕翻訳の検討

成浦拓音[1]，品川政太朗[1]，須藤 克仁[1]，中村 哲[1]

1.奈良先端科学技術大学院大学

概要

動画キャプションモデルを用いた映像付き翻訳

従来の映像付き翻訳の課題

- 映像付きの対訳コーパスサイズが小さい

本研究のアイデア

- 学習済自然言語処理モデルをそのまま活用したい

実験結果

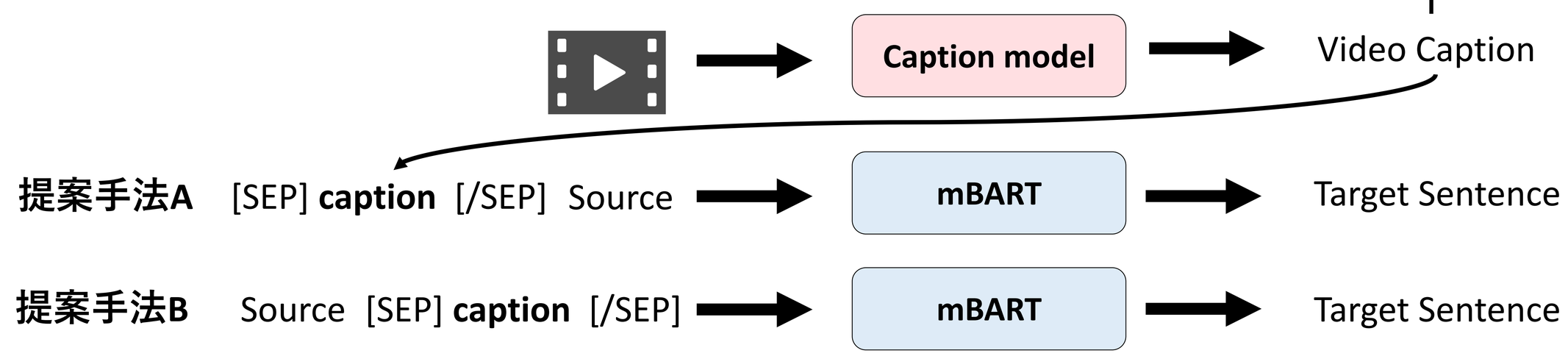
- 多義語のデータセットにおいて翻訳精度が改善

説明文例

動画ファイル名	動画説明文
omission_359950_17.mp4	a man is talking to the camera and then a man is talking.
polysemy_1204975_4.mp4	a group of people are sitting around a table and talking about poker.

提案手法

動画情報から動画説明文を取得し原言語文に結合



実験設定

- VISA データセットを使用
- CC-25 データセットで学習したmBARTをfine-turning
- 日英翻訳を実施
- SwinBERT[2] にて動画説明文を生成

データセット

- Omission
  - 主語の省略が含まれる文章を含む
- Polysemy
  - 多義語が含まれる文章を含む
- Combined
  - 上記２つのデータセットを組み合わせたもの

実験結果 (VMT†は文献[1]より引用)

	Omission			Polysemy			Combined		
評価指標	BLEU⇧	METEOR⇧	RIBES⇧	BLEU⇧	METEOR⇧	RIBES⇧	BLEU⇧	METEOR⇧	RIBES⇧
VMT†	5.63	20.10	10.98	7.40	22.33	12.64	12.89	28.45	19.45
mBART	**9.97**	**32.14**	**20.77**	11.22	33.97	21.49	12.48	36.64	24.59
提案手法A	9.67	31.73	20.48	10.34	34.28	21.53	13.05	**37.11**	25.17
提案手法B	9.79	32.11	20.71	**11.46**	**35.16**	**22.21**	**13.21**	37.07	**25.46**

結果

Omission データセットでは動画説明文を用いない場合が最も良い精度が良い

Polysemy データセット・Combined データセットでは動画説明文を付与した場合が良い

事例分析

原言語文	シャワーを浴びようとしてた	気をつけろ	金を待ってるだけだ
リファレンス	I was about to take a shower.	So just be careful, OK?	They’re expecting cash.
VMT	He was trying to get the shower.	Be careful.	Just got us there, but I’m just waiting for the money.
mBART	He was trying to get a bath.	Take care of yourself.	I’m just waiting for the money.
提案手法A	I was trying to get a bath.	Be careful.	I’m just waiting for the money.
提案手法B	I was trying to get a bath.	Be careful.	I’m just waiting for the money.

考察

多義語の場合は状況を説明するような動画説明文が有効に働いていると考えられる

映像中に発話の主語となる人物が存在することは少なく、主語の推論に動画説明文を用いるのは難しいと考えらえれる

今後の展望

- 物体キャプショニング技術の利用
- 話者情報や表情の利用
- 長い文章の翻訳時にも効果を検証

参考文献

[1] Li et al., VISA: An Ambiguous Subtitles Dataset for Visual Scene-aware Machine

[2] Lin et al., SwinBERT: End-to-End Transformers with Sparse Attention for Video Captioning