BERT を用いたアニメーション字幕における談話境界推定

Discourse Segmentation of Animation Closed Caption Using BERT

大河原 龍太朗[*1]，望月 源[*2]
Ryutaro OKAWARA[*1], Hajime MOCHIZUKI[*2]
[*1] 東京外国語大学大学院総合国際学研究科
[*1] Graduate School of Global Studies, Tokyo University of Foreign Studies
[*2] 東京外国語大学大学院総合国際学研究院
[*2] Institute of Global Studies, Tokyo University of Foreign Studies
Email: {okawara.ryutaro.s0, motizuki}@tufs.ac.jp

あらまし：本研究では，日本語 Can-do 学習に対応した会話例文をアニメーション字幕から自動的に抽出するシステムの構築を目指し，自然言語処理モデルである BERT を用いたアニメーション字幕における談話境界推定の手法を検討した．BERT をアニメーション字幕データセットでファインチューニングし，アニメーション内の談話境界を推定するモデルを複数作成，その性能を比較したところ，二連続の発話の間における談話境界の有無を推定するモデルが，最も高い性能を示した．
キーワード：談話境界推定，Can-do，教材開発，日本語学習，アニメーション

1. はじめに

近年，Can-do 型日本語学習が広まる中で，その学習スタイルに対応した教材の拡充が課題となっている．Can-do 型学習では，学習する項目に対応した会話例・例文などの教材が欠かせないが，教科書の例だけでは十分ではない．そこで，Can-do 型学習に有効な会話例（以下，Can-do 会話）の拡充をめざし，大河原・望月[(1)] は大規模日本語アニメーション字幕コーパスからの Can-do 会話の抽出を試み，人手による会話境界の推定，会話部分の抽出および Can-do 会話の選定を行った．この研究を通じ，アニメーション字幕が優れた Can-do 会話資源であることを示すとともに，字幕内の各会話に対し Can-do 会話としての適否や Can-do レベルをタグ付けした「Can-do タグ付与済みコーパス[(1)]」を作成した．

一方，手作業による拡充には限界があり，自動的に Can-do 会話を抽出するシステム（以下，抽出システム）の構築が望まれる．テキスト全体から会話部分を抽出する問題は，自然言語処理分野における談話境界推定や段落境界推定[(2)(3)(4)]と考えられる．本研究では，抽出システムの構築をめざし，アニメーション字幕における談話境界推定の手法を検討する．推定には，自然言語処理モデルである BERT[(5)]を利用する．BERT は，教師データを用いた学習（ファインチューニング）を行うことで文章分類などの様々なタスクに応用可能な事前学習モデルである．本研究では，BERT を用いた談話境界推定の手法を検討するとともに，「Can-do タグ付与済みコーパス」を教師データとして BERT をファインチューニングし，アニメーション字幕における談話境界推定を行うモデルを作成，その性能を評価する．

2. 談話境界推定モデルの作成

2.1 利用するデータセットと前処理

BERT のファインチューニングの際に教師データとして利用する「Can-do タグ付与済みコーパス」は，東京外国語大学計算言語学研究室で収集整備している「日本語テレビ字幕コーパス」[(6)]の中の日本語アニメーション字幕 35,261 文から，同大学留学生日本語教育センターの AJ Can-do リスト[(7)]を用いて一連の Can-do 会話とみなすことができる 1,600 セグメント（20,097 文）を人手で抽出したデータセットである．また，国内外で人気の高い日本語アニメーション作品群から 12 作品が無作為に選出されている[(1)]．

このデータから談話境界を学習するために，各文に付された会話の通し番号の切り替わりを手掛かりに，談話開始文となる発話に 1 のラベルを，それ以外の発話に 0 のラベルを新たに付与する．以下の図 1 に例を示す．

図 1：データセットの一部

文番号	通し番号	談話開始ラベル	発話
157	3	0	百目鬼氏 お願いします.
158	3	0	えっ まぁ… 探してみます.
159※	**4**	**1**	**お疲れさまでーす.**
160	4	0	お〜 スゴイスゴーイ!

※談話始点
出典：映像研には手を出すな！第 7 話より

2.2 談話境界推定の手法とモデルの作成

本研究では，談話境界推定を以下の 2 種類のタスクと考え，抽出システムを構築する．

タスクⅠ．ある発話が談話の開始文であるか否かを判定するタスク

タスクⅡ．連続する 2 つの発話の間に談話境界が含まれるか否かを判定するタスク

タスクⅠでは，各文について，談話の開始文，非開始文の二値分類を行う．図 1 の例では，文 159 が開始文，その他の文は非開始文である．本研究では，BertForSequenceClassification（SC1 モデル）をタスク

Ⅰに対応するモデルとして用いる．

タスクⅡでは，隣接する 2 文が連続するか否かを直接推定する BertForNextSentencePrediction（NSP モデル）と，2 文が連続関係にあるか非連続関係にあるかを分類する BertForSequenceClassification（SC2 モデル）を用いる．図 1 の例では，文 158 と文 159 のペアが入力として与えられると，その間に談話境界が存在すると判定する．

2.3 不均衡データの処理

本データセットに含まれる談話境界の総数（1,671 件）は，データセット内の発話の全数（18,426 件）に比べ大幅に少ない．本研究では，このような不均衡データへの対応として，多数派のラベルを少数派のラベルと同数になるようサンプリングし，モデルを作成する. また, サンプリングの効果を示すため, サンプリングされていないデータセットを用いて作成したモデルの性能も評価する．

3. モデルの評価と考察

計 6 個の談話境界推定モデルを作成した．各モデルの性能を以下の表 1 に示す．

表 1 ：モデルの各評価値

注：*のモデルはデータセットのサンプリングなし

		Precision	Recall	F1-Score
タスクⅠ				
SC1	train	0.7368	0.7120	0.7341
	test	0.6522	0.6184	0.6349
SC1*	train	0.0010	1.0000	0.0020
	test	0.0303	0.3632	0.0559
タスクⅡ				
NSP	train	0.7735	0.7845	0.7767
	test	0.6432	0.6363	0.6397
NSP*	train	0.0985	0.2388	0.0404
	test	0.0643	0.3321	0.1078
SC2	**train**	**0.8044**	**0.7941**	**0.8031**
	test	**0.7010**	**0.6389**	**0.6685**
SC2*	train	0.0472	0.8596	0.0901
	test	0.0122	0.2134	0.0231

評価の結果，SC2 モデルが最も高い性能を示した．まず，NSP モデルと SC2 モデルの比較から，隣接する 2 文の連続から談話境界の有無を推定するとき，BertForSequenceClassification を用いたモデルがより効果的であることが示唆された．また SC1 モデルと SC2 モデルの比較から，談話境界を推定する際，各文の談話開始文としての適否ではなく，隣接する 2 文の関係に着目することで，より効果的な推定を行えることがわかった．

また，不均衡データをサンプリングせずに学習したモデルは，すべての評価値で低値を示しており，モデルの学習における不均衡データの影響が顕在化した．一方，サンプリングを行うことで，データセットのサイズは大きく減少した（20,097→3,342）．より高精度なモデル作成のためには，少数派のラベルを含む教師データのさらなる拡張が望まれる．

4. おわりに

本研究では, アニメーション字幕内の Can-do 会話の自動抽出を目指し，BERT を利用したアニメーション字幕における談話境界推定モデルを作成した．モデルの性能評価の結果，BertForSequence-Classification を用いた，連続する二発話間における談話境界の有無を推定するモデル（SC2 モデル）の有効性が示された. ただし, 性能は precision 約 70.1%, recall 約 63.9%, F1-Score 約 66.9%であり，さらなる向上が望まれる．

また, アニメーション内の談話は時に複雑であり, 字幕の情報のみから談話境界を正確に決定することが困難な場合がある．アニメーション字幕における談話境界推定の精度を向上させるには，アニメーションのシーン変化の映像情報や，セリフの時間的間隔などを，談話境界を示す情報として利用することも検討する必要がある．

今後の研究の展望として，本研究のモデルを利用して発見された談話セグメントを会話として抽出するとともに, 抽出された会話の Can-do 会話としての適否およびその Can-do レベルを推定するシステムの構築が考えられる．

謝辞

本研究は JSPS 科研費 JP19H04224, JP20H00096 の助成を受けたものです.

参考文献

(1) 大河原龍太朗，望月源: “Can-do 型日本語学習用資源としてのアニメーション字幕の分析”，言語処理学会, 第 28 回年次大会発表論文集, pp1690 - 1694 (2022)
(2) Glavaš, G. and Somasundaran, S.: “Two-Level Transformer and Auxiliary Coherence Modeling for Improved Text Segmentation”, arXiv:2001.00891 [cs.CL] (2020)
(3) 飯倉陸，岡田真，森直樹: “Focal Loss を利用した BERT による小説の段落境界推定”，人工知能学会全国大会論文集, JSAI2020 (2020)
(4) Zhang, L. and Zhou, Q.: "Topic Segmentation for Dialogue Stream," 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China, 2019, pp. 1036-1043, doi: 10.1109/APSIPAASC47483.2019.9023126.
(5) Devlin, J., Chang, M., Lee, K. and Toutanova, K.: “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, CoRR, Vol. abs/1810.04805 (2018)
(6) Mochizuki, H. and Shibano, K.: Building Very Large Corpus Containing Useful Rich Materials for Language Learning from Closed Caption TV, World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Volume 2014, No. 1, pp. 1381-1389, Association for the Advancement of Computing in Education (AACE), 10. 2014
(7) 東京外国語大学留学生日本語教育センター, JLPTUFS アカデミック日本語 Can-do リスト（AJ Can-do リスト）