共参照情報を活用した映画のあらすじ要約生成

川原田 将之 † 長谷川 駿 † 上垣外 英剛 † 高村 大也 †§ 奥村 学 †
† 東京工業大学 § 産業技術総合研究所
{kawarada@lr., hasegawa.s@lr., kamigaito@lr., takamura@, oku@}pi.titech.ac.jp

1 はじめに

インターネット上の映画情報サイトでは，様々な映画のあらすじを読むことができる．その中には映画内容を詳細に記した長文のあらすじも多く存在する．しかし，読者には必要最低限の情報が含まれた簡潔なあらすじを読みたいという需要も大きい．そこで，本研究では，映画内容が詳細に記載されたあらすじを自動要約し，簡潔なあらすじを作成する単一文書要約手法の研究を行う．一般的な文書要約のアプローチには大きく分けて抽出型と生成型が存在する．しかし，映画のあらすじでは，１文であっても人物に関して詳細な記述がされている場合があり，文を抽出単位とした抽出型要約では冗長になることが予想される．そのため，本研究では映画のあらすじ要約を生成型文書要約として定式化する．

我々が調査したところ，映画をドメインとする要約用データセットは存在していなかった．そこでまず，映画情報サイトである **IMDb** (**I**nternet **M**ovie **D**ata**b**ase)[1)]からデータの収集を行い，映画全体のストーリーが詳細に書かれたあらすじと，それに対応した要約文書であるあらすじ要約がペアとなった映画要約データセットを構築した．

映画のあらすじには，文書要約の研究で広く用いられているニュース記事と比べて，人称代名詞が多用されるという特徴がある．しかし，現在の生成型要約モデルではそれらを十分に考慮できているとは言えない．そこで，我々は入力文書のエンコード時に共参照情報を明示的に活用した要約手法を提案する．また，本研究で作成した映画要約データセットは，CNN/Daily Mail(以下 CNN/DM) データセット [1] などのニュース記事要約用データセットと比べると小規模である．そのため，本研究では，ドメインの異なる大規模な要約データセットでモデル全体の事前学習を行った後，映画要約データセットを用いて fine-tuning を行う．

1） https://www.imdb.com/

評価実験において，共参照情報を明示的に活用する提案手法，そしてドメインの異なる大規模なデータセットを用いた事前学習による性能向上が，自動評価指標に関し確認された．

2 関連研究

単一文書要約に関する研究は幅広く行われている [2, 3] が，小説や映画などの物語を対象とした研究は少ない [4, 5]．

ニューラルネットワークによる生成型要約は，文書を入力，要約を出力とした Encoder-Decoder モデルをベースとして行われる [6]．近年では，**BERT** (**B**idirectional **E**ncoder **R**epresentations from **T**ransformers) に代表される事前学習済みエンコーダ [7, 8] を文書要約タスクに適応学習する研究が行われている．Liu ら [9] は，BERT を用いて文書全体をエンコードした後，Transformer[10] でデコードすることにより，事前学習済みのエンコーダを使った生成型文書要約を行った．

Xu ら [11] は，予め文書の談話構造と共参照関係を表すグラフを作成しておき，BERT から得られた隠れ状態を Graph Convolution Networks (GCN)[12] で更にエンコードすることにより，EDU 単位で抽出型要約を行っている．Xu らの研究は，文書の共参照関係を取り入れたモデルであり本研究と近いが，本研究は生成型要約であることから，同じ手法でモデルに共参照情報を組み込むことが出来ない．また，用いるデータセットのドメインも異なっている．

3 映画要約データセット

映画のあらすじ要約を行うにあたり，映画ドメインの要約データセットの構築をした後，作成したデータセットの分析を行う．

3.1 データセットの作成

映画データの収集先として映画情報サイトである IMDb を選択した．IMDb は，映画情報をユーザー

表 1　映画要約データセットと CNN/DM データセットの比較

データセット	文書数（括弧内は入力文書数）			入力文書長		出力文書長		人称代名詞が
	train	valid	test	単語	文	単語	文	文書に占める割合 (%)
CNN	90,266	1,220	1,093	760.54	34.00	45.69	3.56	4.32
DM	196,961	12,148	10,397	787.45	41.76	54.64	3.83	5.07
映画	43,281 (21,931)	5,461 (2,743)	5,529 (2,740)	1174.80	62.12	81.92	3.42	7.17

が自由に投稿できるようになっており，ネタバレを含まず簡潔に書かれた *Plot Summaries* と映画全体の内容が詳細に書かれた *Plot Synopses* の 2 種類のあらすじを投稿することができる．

まず，IMDb が公開している映画の固有 ID リスト[2]に記載されている作品を元に映画データを収集した．次に，収集した映画データの中で *Plot Synopses* にテキストが存在しない作品や *Plot Summaries* の文書長が *Plot Synopses* の文書長を上回っている作品を削除し[3]，最終的に 27,414 作品のデータを取得した．そして，作品単位でデータの分割を行った後，*Plot Synopsis* をあらすじ，*Plot Summaries* をあらすじ要約としてデータセットを構築した．

3.2　データセットの分析

文書数，入力文書長，出力文書長，人称代名詞が文書に占める割合のそれぞれについて，映画要約データセットと CNN/DM データセットで比較を行ったもの[4]を表 1 に示す．入力文書長，出力文書長，人称代名詞が文書に占める割合の値は，全データの平均値である．

投稿の規定上，*Plot Summaries* は複数投稿できるのに対して，*Plot Synopsis* は 1 作品につき 1 つしか投稿出来ないため，1 つのあらすじに対して複数のあらすじ要約が存在するデータセットとなっている．train/valid/test セットのそれぞれについて，あらすじは 21,931/2,743/2,740 データであるのに対して，あらすじ要約は 43,281/5,461/5,529 データとなっており，1 作品あたり平均して約 2 種類のあらすじ要約が存在していることを意味する．

また，各あらすじ文書中に存在する単語の中で人称代名詞の割合は 7.17%であった．これは，CNN データセットに含まれる割合よりも 40%程度高く，DM データセットよりも 30%程度高い値である．

2）https://www.imdb.com/interfaces/
3）テキストが存在していても記号のみの場合などは削除した．
4）単語分割には Stanford CoreNLP 3.8.0 を用いた．

4　提案手法

先行研究 [9] をベースに，共参照情報を明示的に活用できるよう改良を加えた生成型要約モデルを図 1 に示す．本手法ではまず，あらすじに対する共参照解析 [13] を行い，共参照グラフ（4.1 節で後述）を作成する．その上で，あらすじを BERT でエンコードした後，グラフエンコーダを用いて共参照グラフに基づき隠れ状態を更新する．最後に Transformer でデコードを行うことであらすじ要約を生成する．

4.1　共参照グラフ

共参照解析とは，文書内で言及されている名詞句等の表現対が同一の対象を指しているか否かを判別する処理のことである．

共参照関係にある表現は，1 単語から成る表現だけではなく，人物名など複数単語から成る表現も存在する．また，BERT の入力はサブワードに分割されている必要があり，1 単語の表現であっても複数のサブワードから構成されている場合がある．

これらを扱うため，トークン分割後のある表現 $s = \{s_1, s_2, \ldots, s_M\}$ の中で，最初と最後のサブワードである s_1, s_M をその表現を代表するサブワード s_{start}，s_{end} とする[5]．そして，同一対象を示す各表現に対して，s_{start}, s_{end} 同士をそれぞれ独立して連結する．最後に，各サブワードに self-loop エッジを追加することで，サブワードをノードとする共参照グラフを作成した．

4.2　BERT を使ったエンコード

先行研究に従い，前処理として入力テキストをサブワードに分割し，入力の最初に [CLS] ラベル，各文の最後に [SEP] ラベルを追加する．その上で，入力 $\mathbf{x} = \{x_1, \ldots, x_n\}$ は次式によってエンコードされ，隠れ状態 $\mathbf{h} = \{\mathbf{h}_1, \ldots, \mathbf{h}_n\} \in \mathbb{R}^{d \times n}$ を得る：

$$\{\mathbf{h}_1, \ldots, \mathbf{h}_n\} = \mathrm{BERT}(\{x_1, \ldots, x_n\}). \quad (1)$$

ここで，d は隠れ状態の次元を表している．

5）表現が 1 つのサブワードから成るとき ($M = 1$ のとき) は，$(s_{start}, s_{end}) = (s_1, s_1)$ とする．

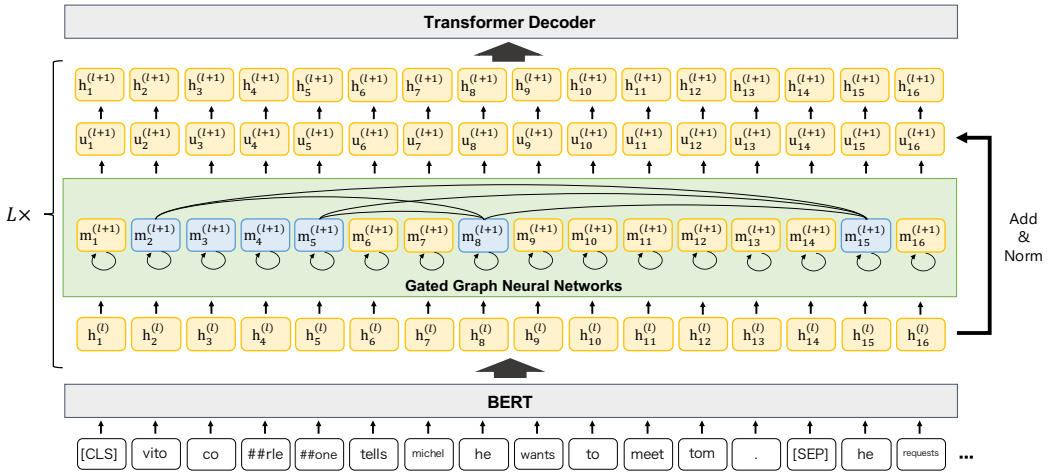

図 1　共参照情報を活用した生成型要約モデル

4.3　グラフエンコーダ

L 層のグラフエンコーダを用いて，BERT から出力された隠れ状態 $\mathbf{h}$ の更新を行う．l 層目のグラフエンコーダの出力を $\mathbf{h}^{(l+1)} = \{\mathbf{h}_1^{(l+1)}, \ldots, \mathbf{h}_n^{(l+1)}\} \in \mathbb{R}^{d \times n}$ と表すと，$\mathbf{h}_i^{(l)}$ から $\mathbf{h}_i^{(l+1)}$ への更新は次のように行われる．

まず，共参照グラフで繋がっているサブワード間で情報の交換を行うため，Gated Graph Neural Networks (以下 GGNN)[14] により $\mathbf{h}_i^{(l)}$ を $\mathbf{u}_i^{(l+1)}$ に変換する．$\mathbf{u}_i^{(l+1)}$ は次の式で計算される：

$$\mathbf{m}_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \mathbf{W}^{(l)} \mathbf{h}_j^{(l)}, \tag{2}$$

$$\mathbf{u}_i^{(l+1)} = \mathrm{GRU}^{(l)}(\mathbf{m}_i^{(l+1)}, \mathbf{h}_i^{(l)}). \tag{3}$$

ここで，$\mathcal{N}(i)$ は共参照グラフにおいて i 番目のサブワードとエッジで繋がったサブワードの集合である．$\mathbf{W}^{(l)} \in \mathbb{R}^{d \times d}$ は l 層目の学習パラメータである．

$\mathbf{h}_i^{(l+1)}$ は $\mathbf{u}_i^{(l+1)}$ と $\mathbf{h}_i^{(l)}$ を用い次式で更新される：

$$\mathbf{h}_i^{(l+1)} = \mathrm{LayerNorm}(\mathbf{h}_i^{(l)} + \mathrm{ReLU}(\mathbf{u}_i^{(l+1)})). \tag{4}$$

$\mathbf{h}^{(1)}$ には BERT の出力 $\mathbf{h}$ を用い，最終的に L 層目から $\mathbf{h}^{(L+1)}$ が出力される．

4.4　デコーダ

先行研究 [9] と同様に，デコーダにはランダムに初期化された Transformer[10] を用いる．グラフエンコーダにより出力された隠れ状態 $\mathbf{h}^{(L+1)}$ を使ってあらすじ要約の生成を行う．

4.5　CNN/DM データセットによる事前学習

3.1 節で作成した映画要約データセットは，CNN/DM データセットに比べて規模が小さい．そのため，未学習であるグラフエンコーダやデコーダを含むモデルを映画要約データセットのみで学習することは困難である．

そこで，CNN/DM データセットを用いてグラフエンコーダやデコーダを含むモデル全体を事前学習した後，映画要約データセットによる fine-tuning を行う．fine-tuning の際には，モデル全体の学習率を低くすることで事前学習の情報を保ちつつ学習する．

5　実験

提案手法の要約性能を評価するため，先行研究モデルである BertSumAbs[9] をベースラインに自動評価による比較実験を行う．

5.1　実験設定

先行研究の実装[6] を基に実験を行った．GGNN の実装には，PyTorch geometric[7] を用いた．実験に必要なハイパーパラメータは，基本的にデフォルト値を用いた．ただし，グラフエンコーダは $L = 2$，ドロップアウト率 = 0.2 とした．

CNN/DM 事前学習+映画データ：事前学習では，BERT の学習率を 0.002，warming-up ステップを 20,000 とし，グラフエンコーダとデコーダの学習率を 0.2，warming-up ステップを 10,000 とした．学習は 200,000 ステップ行い，CNN/DM データセットの開発データで Perprexity が最も小さいステップの重みを選択した．そして，選択した重みから更に映画要約データセットで fine-tuning を行う．fine-tuning では，BERT，グラフエンコーダ，デコーダの全てで学習率を 0.002，warming-up ステップを 20,000 とし

6）https://github.com/nlpyang/PreSumm
7）https://github.com/rusty1s/pytorch_geometric

表 2　自動評価結果

モデル	R-1	R-2	R-L
ORACLE	37.23	12.35	30.97
LEAD-3	25.52	6.22	21.75
映画データのみ			
ベースライン	26.33	6.05	22.95
提案手法	**26.60**	**6.26**	**23.22**
CNN/DM 事前学習+映画データ			
ベースライン	28.13	7.02	24.23
提案手法 w/o 共参照情報	27.87	6.86	23.97
提案手法	**28.33**	7.07	**24.43**

た上で 50,000 ステップの学習を行った．

映画データのみ：CNN/DM 事前学習の効果を確認するため，映画データのみでも学習を行う．先行研究に従い BERT の学習率を 0.002，warming-up ステップを 20,000 とした．また，グラフエンコーダとデコーダの学習率を 0.2，warming-up ステップは 10,000 とした．学習は 50,000 ステップを行った．

5.2　実験結果

先行研究同様，ROUGE-1，-2，-L（R-1，R-2，R-L）[15] による自動評価結果を表 2 に示す．表には異なる初期値から学習した 3 モデルの平均値を記し，太字はベースラインに対する差が有意 ($p = 0.05$) であることを表す．

映画データのみで学習した場合をみると，ベースラインと比べて，提案手法は ROUGE スコアがそれぞれ 0.27, 0.21, 0.27 改善しており，共参照情報によって要約性能が向上したことがわかる．次に，CNN/DM で事前学習を行った場合と映画データのみで学習した場合を比較する．ベースライン同士の比較で ROUGE スコアがそれぞれ 1.80, 0.97, 1.28 改善しており，事前学習の効果は大きいことがわかる．CNN/DM で事前学習を行った場合においてベースラインと提案手法を比較すると，ROUGE スコアで 0.20, 0.05, 0.20 改善しており，ROUGE-1 と ROUGE-L では有意な差があった．事前学習を行った場合でも共参照情報は要約性能の向上に寄与していることがわかる．また，提案手法モデルで共参照グラフを用いず，self-loop エッジのみをグラフエンコーダに入力するとベースラインよりも悪化した．これは，ベースラインよりもグラフエンコーダ部分における未学習のパラメータが増加することで，学習が困難になるためであると考えられる．

更に，ベースラインと提案手法の出力例を図 2 に示す．各モデルの出力の中で，あらすじと対応している箇所を青字で示してある．あらすじでは“decide to conceal”の主語は“Harry’s relatives”であるが，ベースラインでは“Harry’s parents”となっている．これは，あらすじの中で直前に“his parents, lily and James Potter”に関する記述があり，一般的に“parents”と“relative”は近い表現であるため，モデル側が上手く区別できていないからであると考えられる．一方，提案手法では“his relatives”と正しく出力出来ている．これは，近い表現であったしても，共参照情報によって人物の区別が容易になったと考えることができる．

あらすじ

（省略）Harry is left an orphan with a lightning-bolt scar on his forehead, Voldemort having killed his parents, Lily and James Potter. Professors Dumbledore and McGonagall and Gamekeeper Hagrid leave him on the doorstep of his ultra-conventional, insensitive, negligent Muggle relatives, the Dursley family, who take him in. **Harry’s relatives decide to conceal** his magical heritage from him and make him live in a cupboard under the stairs for ten years. （省略）

ベースライン

Harry Potter is an orphan with a lightning-bolt scar on his forehead. His parents, Lily and James Potter , leave him on the doorstep of his insensitive relatives, the Dursley family, who take him in. **Harry’s parents decide to hide** his magical heritage from him and make him live in a cupboard under the stairs for ten years.

提案法

Harry Potter is an orphan with a lightning-bolt scar on his forehead. He is left on the doorstep of his insensitive relatives, the Dursley family, who take him in. **His relatives hide** his magical heritage from him and make him live in a cupboard under the stairs for ten years.

図 2　ベースラインと提案手法の出力例

6　まとめ

本研究では，映画のあらすじ要約を生成型文書要約として行った．映画ドメインの要約用データセットを作成し，共参照情報を考慮した生成型要約手法を提案した．ROUGE スコアに基づく評価実験の結果，提案手法が従来手法を上回り，ROUGE-1 と ROUGE-L では有意に上回っていた．また，出力文書を比較すると提案手法では人物に関する記述で真実性の改善が見られた．今後は，人手による評価も交えて提案手法の有効性を確かめたい．

参考文献

[1] Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In *NIPS*, 2015.

[2] Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1073–1083, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[3] Jiwei Tan, Xiaojun Wan, and Jianguo Xiao. Abstractive document summarization with a graph-based attentional neural model. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1171–1181, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[4] Rada Mihalcea and Hakan Ceylan. Explorations in automatic book summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 380–389, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

[5] Anna Kazantseva and Stan Szpakowicz. Summarizing short stories. *Computational Linguistics*, Vol. 36, No. 1, pp. 71–109, 2010.

[6] Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Çağlar Guçehre, and Bing Xiang. Abstractive text summarization using sequence-to-sequence RNNs and beyond. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pp. 280–290, Berlin, Germany, August 2016. Association for Computational Linguistics.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[8] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, Vol. abs/1907.11692, , 2019.

[9] Yang Liu and Mirella Lapata. Text summarization with pretrained encoders. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 3730–3740, Hong Kong, China, November 2019. Association for Computational Linguistics.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, Vol. 30, pp. 5998–6008. Curran Associates, Inc., 2017.

[11] Jiacheng Xu, Zhe Gan, Yu Cheng, and Jingjing Liu. Discourse-aware neural extractive text summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5021–5031, Online, July 2020. Association for Computational Linguistics.

[12] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *CoRR*, Vol. abs/1609.02907, , 2016.

[13] Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 188–197, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[14] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks, 2015. cite arxiv:1511.05493Comment: Published as a conference paper in ICLR 2016.

[15] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.