

抽象型レビュー要約におけるドメイン適応

安藤 智大†

†東京農工大学大学院 工学府

藤田 桂英‡

‡東京農工大学大学院 工学研究院

1 はじめに

近年、Web 上のテキストは膨大であり、文書要約技術は自然言語処理における重要なタスクとなっている。特に、要約文をすべて生成する抽象型要約は優れた成果を残してきた。一方で、正解データとなる人手で要約されたコーパスが大量に必要であり、ニュースサイトなどの限られたドメインでしか利用されないという課題を持つ。

本論文では、Web レビューサイトを対象ドメインとして、既存の大規模データからレビュー向け要約を実現するためのドメイン適応に関する手法を提案し、効果を検証する。提案手法は、事前学習と再学習の二段階からなり、ニュースコーパスと教師なしレビューデータ、新たに開発した小規模レビュー要約コーパスを組み合わせた要約モデルの学習を行う。また、評価実験により事前学習が有効であることと、語彙の置換が生成文に影響を及ぼすことを示す。

2 Encoder-Decoder モデルの事前学習

本論文では、Encoder-Decoder モデルに対して既存のデータを活用して事前学習を行う。既存の大規模要約コーパス CNN/DailyMail[1] に加え、対象ドメインの教師なしレビューテキストからの擬似要約を学習データとして組み込む。

(1) 擬似要約文の作成: レビュー要約において重要とされる評価表現を要約文とするために、レビュー文書中の各文章について次式のスコア付けを実施する。

$$Score(s) = \frac{|\sum_{i=0}^n polarity(w_i)|}{len(s)}$$

polarity は感情極性辞書におけるスコアを指し、本論文では MPQA Subjectivity Lexicon[2] と SentiWordNet[3] の合計値を用いる。スコアが降順になるように文章を並び替え、パラメータで設定した文字数に達するまで抽出したものを擬似要約とする (図 1)。

(2) 擬似要約と CNN/DailyMail データによる学習: 以下の 3 つの方法で、Encoder-Decoder モデルとして Pointer-Generator Networks[4] を利用し、Coverage Mechanism は使用せずに学習を行う。

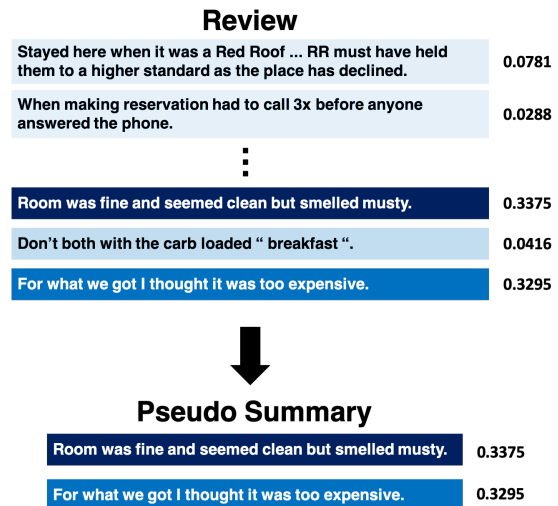


図 1: 感情極性辞書ベースによる擬似要約

表 1: アノテーション結果 (2 者によるレビュー中の語句の ROUGE 値)

	Hotel	Restaurant
平均原文抽出率	0.7336	0.7643
平均圧縮率	0.1857	0.1923
Rouge-1	0.3523	0.3587
Rouge-2	0.1235	0.1331
Rouge-l	0.3170	0.3182

S-UT: CNN/DailyMail → 擬似要約の順で学習. UT-S: 擬似要約 → CNN/DailyMail の順で学習. 相互学習 (mutual): 両データをミニバッチごとに交互に学習.

3 再学習と語句の置換

各事前学習モデルの中間層の値を引き継ぎ、小規模レビュー要約コーパスによる再学習を行う。通常の再学習に加え、データ拡張を目的とした語句の置換を 1 epoch ごとに行った類義語 (synonym) 拡張を実施する。概念辞書である WordNet から各単語の類義語を取得し、元文書・要約文のそれぞれに対してランダム (50% の確率) で置き換えを行い新たな学習データとする。また抽象化における言い換え表現と語句の上位概念との関連を検証するため、要約文に対してのみ類義語に加えて上位語 (hypernym) による拡張も行う。

Domain Adaptation for Abstractive Review Summarization
 †Faculty of Engineering, Tokyo University of Agriculture and Technology
 ‡Institute of Engineering, Tokyo University of Agriculture and Technology

表 2: 学習方法ごとの ROUGE 値

		Hotel			Restaurant		
model		Rouge-1	Rouge-2	Rouge-L	Rouge-1	Rouge-2	Rouge-L
教師なしドメイン適応 (USP)	CNN/DailyMail のみ (S)	0.2293	0.0687	0.2048	0.2259	0.0668	0.2019
	UT-S	0.2430	0.0833	0.2188	0.2148	0.0671	0.1924
	S-UT	0.3184	0.1328	0.2865	0.3095	0.1322	0.2820
	相互学習 (mutual)	0.3111	0.1235	0.2782	0.3041	0.1307	0.2777
教師ありドメイン適応 (SP)	通常教師あり学習	0.2700	0.0822	0.2449	0.2821	0.1009	0.2590
	類義語拡張	0.2079	0.0389	0.1919	0.1909	0.0366	0.1746
	類義語拡張 + 上位語拡張	0.1553	0.0170	0.1429	0.1744	0.0228	0.1607

4 レビュー文書要約コーパスの作成

再学習及び評価実験に使用するためのレビュー文書要約コーパスをアノテーションにより作成した。対象ドメインを TripAdvisor の Hotel, Restaurant の 2 種類に定め、各ドメインから 90 単語以上使用されているレビューを 200 件ずつ取得し要約タスクの対象とした。要約文の作成は 8 人のアノテータによるフリーテキスト形式によって行った。1 件のレビューにつき 2 人がそれぞれ要約文を記述し、合計 800 件の要約コーパスを作成した。使用する単語に制約は設けず、語数制限として圧縮率が 7% から 20% の内に収まるように指定している。表 1 はアノテーション結果を示し、2 者によるレビュー中の語句の一致率を ROUGE 値で示す。

5 評価実験

提案した事前学習と再学習の手法を用いて要約文生成の実験を行う。実験は事前学習モデルのみを用いた教師なしドメイン適応 (USP) とその後再学習を行った教師ありドメイン適応 (SP) の 2 種類を行う。疑似要約を作成する教師なしレビューテキストとして Hotel を 28888 件, Restaurant を 47549 件を TripAdvisor からクロールして使用する。USP ではレビュー文書要約コーパスの各ドメインの全データを評価に用い、SP では 2 分割交差検証を実施する。

表 2 は、各手法に対して 1 レビューごとに存在する 2 件の正解要約のそれぞれに対する ROUGE の平均値を示している。実験結果より S と比較して他の疑似要約による学習の値が向上している。追加実験として疑似要約を各レビューの一文目と変更し再実験を行ったところ、すべてのモデルで値が大幅に低下したため、スコア付けによる抽出の効果が確認された。一方、生成される要約はほとんどが重要文抽出となっており、抽象化としての機能は失われる結果となった (表 3 の S-UT)。SP は各事前学習モデルで再学習し ROUGE-L の値が最も大きかったモデルの結果を示している。表 2 から、いずれも再学習により ROUGE 値は低下したが、本文中

表 3: 出力結果の一例

S-UT	Front desk staff were very friendly and accommodating throughout our stay . Front desk staff were very friendly and accommodating throughout
通常教師あり学習	The hotel was also good as friendly . The rooms were big and accommodating points .
類義語拡張	The AAARP/Senior lodge surpass Park and good . The staff survive very friendly and accommodating .

に存在しない語を使用する言い換え表現が顕著に現れ、繰り返し語の抑制が確認できる。特に、表 3 の類義語拡張から、本文中に存在しない good や survive などの語が出現したが、lodge など言及していない観点も出現していることがわかる。

6 まとめ

本論文では、事前学習及び再学習により抽象型レビュー要約のドメイン適応を提案した。実験により疑似データの有用性が確認できた他、再学習における語句の置換が生成文に影響を持つことが明らかになった。

参考文献

- [1] Karl Moritz Hermann, Tomáš Kočický, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, pages 1693–1701, Cambridge, MA, USA, 2015. MIT Press.
- [2] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 347–354, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics.
- [3] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May 2010. European Language Resources Association (ELRA).
- [4] 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)*, pages 1073–1083, Vancouver, Canada, July 2017. Association for Computational Linguistics.