

テレビドラマに関する関心・感想の有無の 文・ウェブページ単位判定モデル*

川口 輝太[†] 久保 遼馬[†] 藤田 拓也[†] 前田 竜治[†]宇津呂 武仁[†] 小林 彰夫[‡] 西崎 博光[§] 河田 容英[¶][†]筑波大学大学院 システム情報工学研究科 [‡]筑波技術大学産業技術学部[§]山梨大学大学院総合研究部 [¶](株) ログワークス

1 はじめに

本論文では、テレビドラマ視聴者がドラマ視聴後にウェブ上で行うドラマ関連の関心・感想・レビュー類の情報探索過程を支援することを目的として、ブログ・ドラマ関連サイト等のウェブページからの情報収集・集約を行うウェブマイニング技術を提案する。まず、各ドラマに関連する多種多様なウェブページを収集し、その後、収集されたウェブページ中のコンテンツを、当該ドラマについての感想、および、当該ドラマについての感想以外の、キャスト、俳優、当該ドラマ特有のキーワード、視聴率・ロケ地等のドラマジャンル全体で共通の関心事項等へ分類することを目的とする。そのための要素技術として、本論文では、文もしくは、ウェブページが、ドラマに関連するか否かの自動判定、及び、感想表現を含むか否かの自動判定手法を提案する。

2 評価用データセット

評価用データセットの作成にあたっては、文献 [5] で述べた手順のとおり、表 1 中の「一ドラマあたりのウェブページ数が 10 以上のサイト対象」欄に該当するウェブページ集合から、各ドラマに対して、表 2(a) に示す数のウェブページを無作為に選定し評価実験において用いる。また、表 1 中の「一ドラマあたりのウェブページ数が 10 以上のサイト対象」欄に該当するウェブページ集合から、各ドラマに対して、表 2(b) に示す数の文を無作為に選定し評価実験において用いる。評価用文の選定においては、まず、各ドラマごとに 16

表 2: 評価用データセット (評価対象ドラマ: 「リーガルV」, 「下町ロケット」, 「僕らは奇跡でできている」, 「あなたには渡さない」)

(a) ウェブページ数

	クエリ・フォーカスのドラマに関連する	クエリ・フォーカスのドラマに関連しない	合計
感想を含む	19 / 64 / 59 / 15 (合計) 157	39 / 61	674
感想を含まない	97 / 57 / 77 / 120 (合計) 351	28 / 38 (合計) 166	
合計	508	166	674

(b) 文数

	クエリ・フォーカスのドラマに関連する	クエリ・フォーカスのドラマに関連しない	合計
感想を含む	120 / 143 / 65 / 162 (合計) 490	45 / 22 / 84 / 77 (合計) 228	718
感想を含まない	389 / 722 / 384 / 374 (合計) 1,869	787 / 685 / 761 / 639 (合計) 2,872	4,741
合計	2,359	3,100	5,459

表 3: タスクごとのモデル一覧 (相対的に高性能なモデルに 下線)

	ドラマ関連判定	感想判定
文単位	<u>BERT</u> , doc2vec	BERT+BiLSTM, <u>HAN</u>
ウェブページ単位	BERT+SVM, doc2vec	BERT+SVM, <u>HAN</u>

サイトを選定した後、16 サイトの各々から文選定用のウェブページを各 1 ページ選定し、文選定の情報源とした。

3 ドラマ関連判定

本論文の枠組みにおいて、指定されたドラマ名に対して、収集された文、および、ウェブページの話題が当

*Sentence / Web Page Level Identification Model of Concerns and Reviews on TV Drama

[†]Kota Kawaguchi, Ryoma Kubo, Takuya Fujita, Tatsuya Maeda, Takehito Utsuro, Graduate School of Systems and Information Engineering, University of Tsukuba

[‡]Akio Kobayashi, Tsukuba University of Technology, Tsukuba

[§]Hiromitsu Nishizaki, Graduate School of Interdisciplinary Research, University of Yamanashi

[¶]Yasuhide Kawada, Logworks Co., Ltd.

表 1: 各ドラマに対して収集されたウェブサイト・ウェブページの数

ドラマタイトル	サジェスト数	収集された全サイト対象		一ドラマあたりウェブページ数が 10 以上のサイト対象						
		ウェブページ数	ウェブサイト数	ウェブページ数	ウェブサイト数					クエリ・フォーカスのドラマに関連しない
					合計	クエリ・フォーカスのドラマに関連する			合計	
						感想を含むブログ	感想を含まないブログ	ブログ以外		
「リーガルV」	672	2,621	791	2,065	338	47	41	218	306	32
「下町ロケット」	782	4,688	1,393	3,287	341	52	35	229	316	25
「僕らは奇跡でできている」	630	1,912	639	1,457	279	46	34	175	255	24
「あなたには渡さない」	358	1,019	351	830	197	26	24	136	186	11

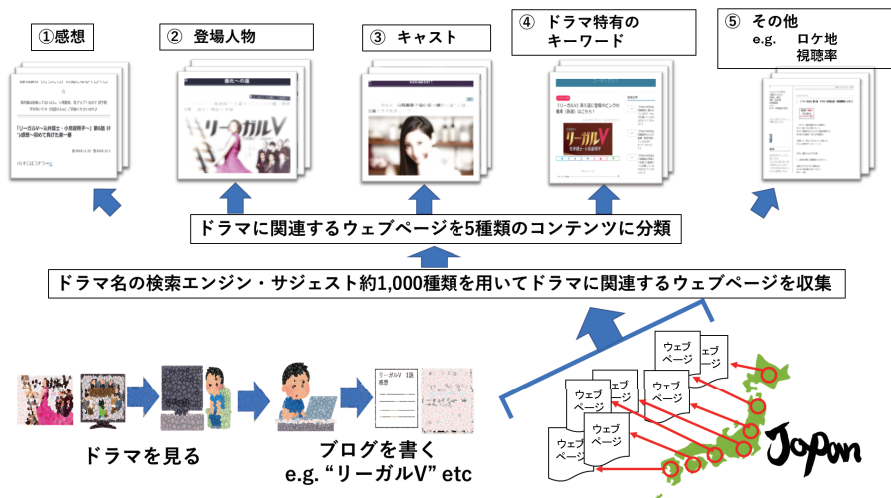


図 1: 本論文の枠組み: ドラマ関連ウェブページの収集およびドラマに関する関心・感想の分類

該ドラマに関連するかどうかを判定するドラマ関連判定モデルを図 2 に示す。

この枠組みにおいては、文の話題が当該ドラマに関連するかどうかを判定するタスクにおいて、BERT [2]、及び、doc2vec [6] による文書ベクトルの cosine 類似度を用いる手法を比較した。一方、ウェブページ単位の判定においては、BERT の文単位判定結果を特徴量とした SVM¹ モデル、及び、doc2vec の文書ベクトルによる cosine 類似度を用いた手法を比較した。BERT、及び、BERT の文単位判定結果を特徴量とした SVM モデルにおいては、文献 [5] の手法を用いた。doc2vec による文書ベクトルの cosine 類似度を用いる手法² においては、話題判定対象の文書ベクトルと、当該ドラ

マの Wikipedia エントリ中のあらすじと登場人物部分のテキストの文書ベクトルの cosine 類似度を求める。

本節のドラマ関連判定モデルの訓練および評価実験において用いるウェブページ、及び、文の数を、それぞれ、表 2(a)、及び、表 2(b)、に示す。評価実験においては、評価対象の 4 ドラマを用いて、3 ドラマを対象として収集した事例をモデルの訓練事例とし、残りの 1 ドラマを対象として収集した事例を評価事例とする交差検定により、文、及び、ウェブページ単位のドラマ関連判定モデルの評価を行った。

文、及び、ウェブページ単位のドラマ関連判定モデルの評価結果をそれぞれ図 3(a)、図 3(b) に示す。評価の際には、BERT の fine-tuning モデルにおいては softmax 関数の出力確率、doc2vec の手法においては cosine 類似度、SVM においては SVM の信頼度の下限値をそれぞれ変化させて再現率・適合率曲線を描いた。この結果から分かるように、文、及び、ウェブページ単位判

¹scikit-learn [7] を用いた。

²doc2vec モデルを訓練する際には、各ドラマにおいて、表 1 中の「一ドラマあたりのウェブページ数が 10 以上のサイト対象」欄に該当するウェブページ集合、および、当該ドラマの Wikipedia エントリ中のあらすじと登場人物部分のテキストを訓練事例として使用した。

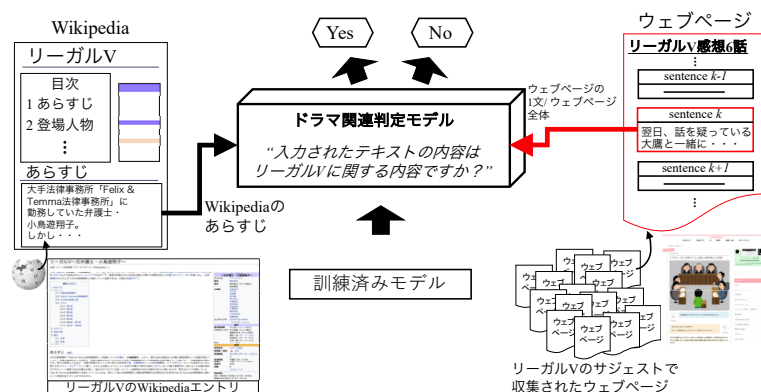
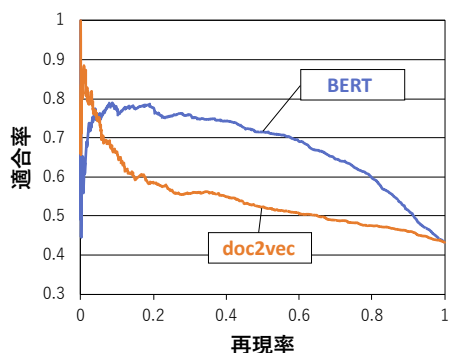
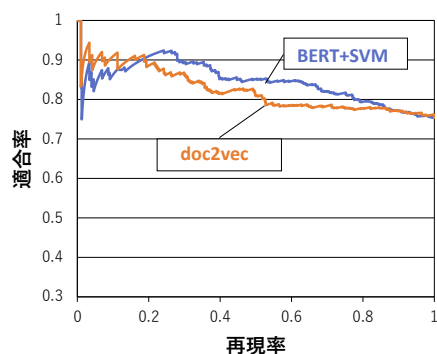


図 2: ドラマ関連判定モデル



(a) 文単位判定



(b) ウェブページ単位判定

図 3: ドラマ関連判定の評価結果

定の双方において、doc2vec を用いた文書ベクトルの cosine 類似度を用いた手法と比較して、BERT を用いた手法の方が高い適合率を達成した。

4 感想判定

次に、本論文の枠組みにおいて、収集された文、及び、ウェブページが感想を含むかを判定する感想判定モデル

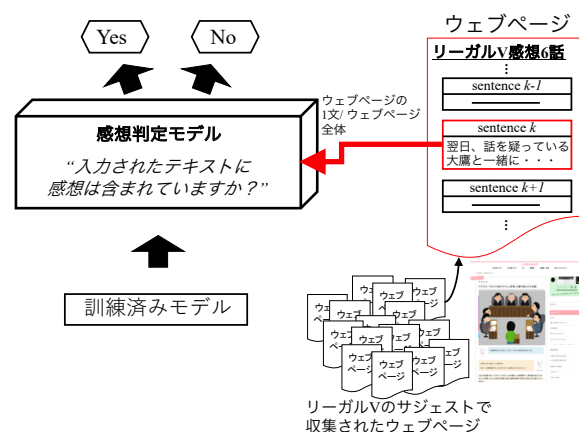


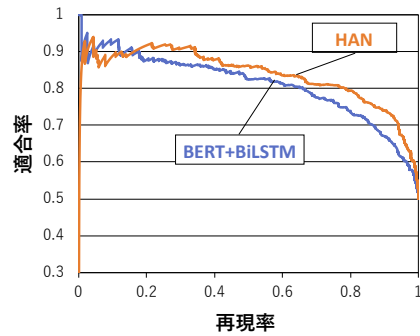
図 4: 感想判定モデル

ルを図 4 に示す。

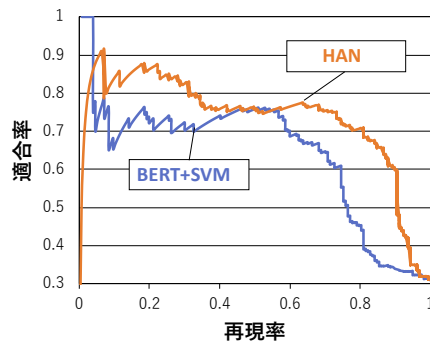
この枠組みにおいては、文が感想を含むか否かを判定するタスクにおいて、BERT の分散表現を用いた双方向 LSTM³、及び、HAN [10] を比較する。一方、ウェブページ単位の判定においては、BERT の文単位判定結果を特徴量とした SVM モデル、及び、HAN を比較する。BERT の分散表現を用いた双方向 LSTM、及び、BERT の文単位判定結果を特徴量とした SVM モデルにおいては、文献 [5] の手法を用いる。HAN は、文書単位の意味表現を構築する際、セグメント⁴単位と単語単位の 2 段階の階層型の注意機構付き双方向 GRU を持つ。HAN においては、公開されている keras 上の

³keras (<https://keras.io/>) を用いて実装した。

⁴文単位の感想判定の場合には、3 形態素を 1 セグメントとし、ウェブページ単位の感想判定の場合には、1 文を 1 セグメントとする。



(a) 文単位判定



(b) ウェブページ単位判定

図 5: 感想判定の評価結果

実装⁵を用いた。

本節の感想判定モデルの評価実験において用いるウェブページ、及び、文の数を、それぞれ、表 2(a)、及び、表 2(b) に示す。評価実験においては、評価対象の 4 ドラマを用いて、3 ドラマをクエリ・フォーカスとして収集した事例を訓練事例とし、残りの 1 ドラマをクエリ・フォーカスとして収集した事例を評価事例とする交差検定により、文、及び、ウェブページ単位の感想判定モデルの評価を行った⁶。

文、及び、ウェブページ単位の感想判定モデルの評価結果をそれぞれ、図 5(a)、及び、図 5(b) に示す。評価の際には、双方向 LSTM モデルにおいては全結合層の sigmoid 関数の出力確率、HAN においては全結合層の softmax 関数の出力確率、SVM においては SVM の信頼度の下限値をそれぞれ変化させて再現率・適合率曲線を描いた。この結果から分かるように、感想を含む文の抽出においては、モデルの信頼度が特に高い範囲の適合率において、BERT の分散表現を用いた双方向 LSTM が HAN を上回っているが、それ以外の信

⁵<https://github.com/FlorisHoogenboom/keras-han-for-docla>

⁶ただし、文単位の訓練および評価においては、無作為抽出により、感想を含まない文の数を感想を含む文の数に揃えた。

頼度の範囲の適合率においては、HAN の方が BERT の分散表現を用いた双方向 LSTM を上回った。また、ウェブページ単位感想判定タスクにおいても、HAN モデルの適合率が BERT+SVM の適合率を上回った。

5 関連研究

BERT [2] を利用した関連研究の一例として、レビュー理解・主観分析において利用するもの [1, 3]、イベント検出において利用するもの [9]、SNS マイニングにおいて利用するもの [4]、意見・主張分析において利用するもの [8] 等が挙げられる。一方、本論文においては、テレビドラマ視聴者による情報探索過程支援を目的としてウェブページ単位での当該ドラマ関連判定および感想判定を行う手法を提案しており、先行研究とは問題設定が異なる。

6 おわりに

本論文では、テレビドラマ視聴者がドラマ視聴後にウェブ上で行うドラマ関連の関心・感想・レビュー類の情報探索過程を支援することを目的として、ブログ・ドラマ関連サイト等のウェブページからの情報収集・集約を行うウェブマイニング技術を提案した。今後の課題として、ドラマ関連の関心の詳細分類として、キャスト、俳優、当該ドラマ特有のキーワード、視聴率・ロケ地等のドラマジャンル全体で共通の関心事項等への識別を行うことが挙げられる。

参考文献

- [1] S. Chi, H. Luyao, and Q. Xipeng. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proc. NAACL-HLT*, pp. 380–385, 2019.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proc. NAACL-HLT*, pp. 4171–4186, 2019.
- [3] X. Hu, L. Bing, S. Lei, and Y. Philip. BERT post-training for review reading comprehension and aspect-based sentiment analysis. In *Proc. NAACL-HLT*, pp. 2324–2335, 2019.
- [4] H. Jack and L. Lillian. Something’s brewing! early prediction of controversy-causing posts from discussion features. In *Proc. NAACL-HLT*, pp. 1648–1659, 2019.
- [5] K. Kawaguchi, R. Kubo, T. Fujita, T. Maeda, T. Utsuro, A. Kobayashi, H. Nishizaki, and Y. Kawada. BERT based Web mining of concerns and reviews for TV drama audience. In *Proc. WI Workshops*, pp. 64–71, 2019.
- [6] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In *Proc. 31st ICML*, pp. 1188–1196, 2014.
- [7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.
- [8] C. Sihao, K. Daniel, Y. Wenpeng, C. Chris, and R. Dan. Seeing things from a different angle: discovering diverse perspectives about claims. In *Proc. NAACL-HLT*, pp. 542–557, 2019.
- [9] W. Xiaozhi, H. Xu, L. Zhiyuan, S. Maosong, and L. Peng. Adversarial training for weakly supervised event detection. In *Proc. NAACL-HLT*, pp. 998–1008, 2019.
- [10] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy. Hierarchical attention networks for document classification. In *Proc. NAACL-HLT*, pp. 1480–1489, 2016.