

ツイートにおける芸能人に対する感想のマイニング

野崎 雄太[†] 菅原 琴絵[†] 宇津呂武仁^{††}

[†] 筑波大学大学院 システム情報工学研究群 知能機能システム学位プログラム

〒 305-8577 茨城県つくば市天王台 1-1-1

^{††} 筑波大学 システム情報系 知能機能工学域 〒 305-8573 茨城県つくば市天王台 1-1-1

理化学研究所 革新知能統合研究センター 〒 103-0027 東京都中央区日本橋 1-4-1

あらまし 本論文では深層学習を用いてツイートの芸能人に対する感想関係の有無を判定する手法を提案する。本論文では、芸能人ファンが芸能人に対する意見や関心動向の情報探索をウェブ上で行う過程を支援するために、芸能人に対する感想ツイートの収集・集約を目的とする。具体的にはツイート中に含まれる芸能人名と形容詞表現に着目し、BERT の読解モデルを用いて2つの単語の間の感想関係の有無を判定する。読解モデルではツイートをコンテキスト、芸能人名を質問として、感想関係がある形容詞を回答とする。既存手法の分類モデルやトークン分類モデルと比較し、評価実験の結果、読解モデルが高い性能を達成した。

キーワード ツイッター, 感想関係, 芸能人, BERT, 読解モデル, 感想マイニング

1 はじめに

本論文では、テレビドラマ視聴者や芸能人ファンが、芸能人を対象として、批評や関心動向の情報探索をウェブ上で行う過程を支援することを目的とする。そして、具体的な媒体としてツイートを情報源として、話題となっている芸能人に対する感想を表現したツイートを収集・集約するための手法を提案する。具体的には芸能人名が共起しているツイート群を収集し、ツイート内に共起している形容詞と芸能人名の間に感想関係の有無を手で判定した事例を訓練事例として、芸能人に対する形容詞表現による感想を表現したツイートを収集・集約するためのモデルをBERT [2] を用いて訓練し、感想関係の有無を判定する。モデルの枠組みとしては、図1に示す機械読解の枠組みを使用する。機械読解の枠組みを利用した読解モデルは、感想収集対象となる芸能人名のキーワード Q を質問として、キーワード Q と感想を表す形容詞 A が共起するツイートをコンテキスト C として、当該ツイート内において芸能人名 Q に対する感想が形容詞によって表現されている場合は該当する形容詞 A を出力し、感想が表現されていない場合には、「感想関係なし」を出力することにより、当該ツイート中における芸能人名 Q に対する感想の有無の判定を行う。

評価実験として、はじめにデータセットを作成する。当該データセットにおいては、特定の芸能人に言及しているツイートに対して、そのツイート内において言及されている芸能人に対して、形容詞によって感想が述べられているか否かの判定を手で行い、その判定結果を付与したものを蓄積する。データセットの構築においては、まず、一定期間を設定して、特定の芸能人名が含まれているツイートを収集する。そして、収集結果のツイート群における形容詞の頻度分布を求め、一定以上の頻度となる形容詞を含むツイートを無作為に選定し、デー

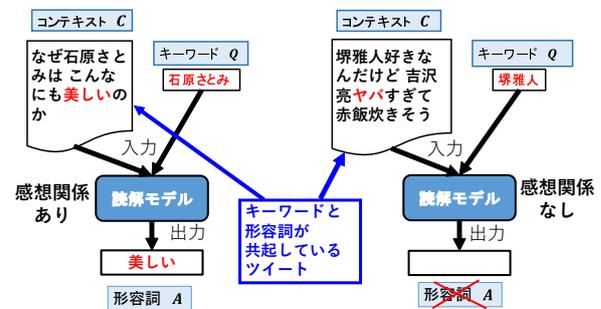


図1 芸能人名・形容詞間の感想関係の有無の読解モデルの枠組み

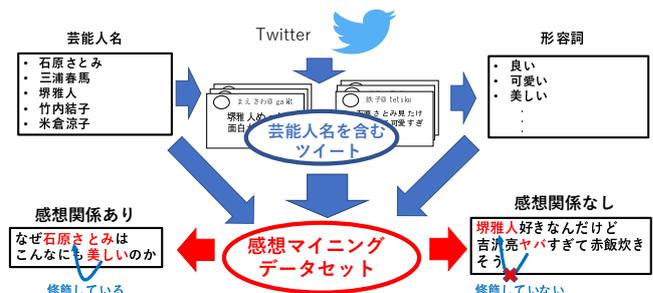


図2 芸能人名・形容詞間の感想関係の有無の手付与

タセットに含めるツイート候補とする。そして、選定された候補ツイートに対して、芸能人名・形容詞間の感想関係の有無の手付与を行い、1,500事例の感想マイニング・データセット(表1)を構築する。

当該データセットを用いて、読解モデルの訓練・評価を行った。また、同データセットを用いて既存手法の分類モデルやトークン分類モデルを訓練・評価し、読解モデルとの比較を行った。実験の結果、読解モデルが他のモデルと比較して高い性能を達成した。

表 1 収集対象、および、データセット作成対象のツイート数 (形容詞を対象とした場合)

芸能人名	ツイート数	リツイート 以外の ツイート数	収集期間	評価用ツイート数			複数の形容詞と 「感想関係あり」 (除外対象)
				一つの形容詞のみと 「感想関係あり」	「感想関係なし」	合計	
石原さとみ	1,677,692	555,701	2020/8/2 ～ 2020/10/24	112	188	300	20
三浦春馬	557,796	130,734	2020/7/22 ～ 2020/8/15	147	153	300	64
堺雅人	161,404	61,108	2020/8/2 ～ 2020/10/24	131	169	300	21
竹内結子	868,979	337,672	2020/9/27 ～ 2020/10/24	126	174	300	44
米倉涼子	37,445	14,869	2020/8/2 ～ 2020/10/24	120	180	300	20
合計	-	-	-	636	864	1,500	169

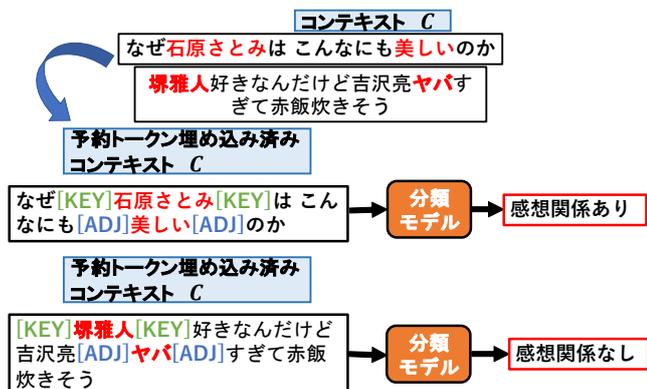


図 3 芸能人名・形容詞間の感想関係の有無の分類モデルの枠組み

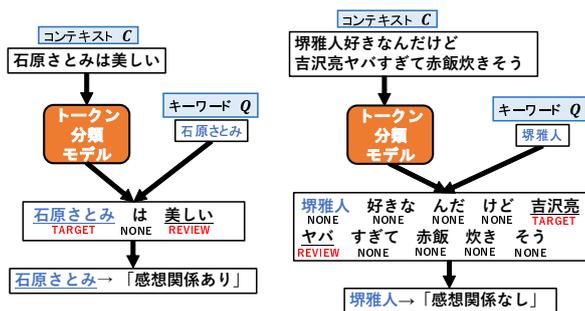


図 4 芸能人名・形容詞間の感想関係の有無のトークン分類モデルの枠組み

2 芸能人を話題とするツイートの収集、および、データセット作成候補ツイートの選定

本節では、Twitter 上で話題に上がることが多い芸能人の一例として、「石原さとみ」、「三浦春馬」、「堺雅人」、「竹内結子」、

「米倉涼子」を対象として、各芸能人に関するツイートの収集を行い、芸能人名と共に起している形容詞の頻度を集計するとともに、データセットを作成する際の候補となるツイートを選定した。

まず、Twitter Search API¹ を用いて、各芸能人名をキーワードとして、キーワードを含むツイートの収集を行う。なお芸能人名に関しては姓名共に含まれているツイートを対象とし、「石原」や「春馬」など姓のみや名のみ、「たけうちゆうこ」や「米倉りょうこ」等、「石原さとみ」の名以外がひらがなやカタカナで表されているツイートは収集の対象外とする。収集したツイートの数、リツイート以外のツイートの数、およびツイートの収集期間を表 1 に示す。

次に、本論文では、テレビドラマ出演俳優・登場人物に対するツイートの感想マイニングを行った先行研究 [5] に従い、ツイート中において感想を表現する語としては、形容詞を対象とする²。そして、収集したツイートに対して、JUMAN++³ を用いて形態素解析を行った結果を用いて、各形容詞を代表表記 (終止形) に変換した後⁴、各芸能人ごとに共に起している形容詞の頻度統計を求める。その後、各芸能人名ごとに、出現頻度の上位 30 個の形容詞をツイート収集対象の形容詞とする。

次に、各芸能人名ごとに出現頻度の上位 30 種類の形容詞を用いて、データセット作成候補のツイートを収集する。具体的

1 : <https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets>

2 : 形容詞以外の品詞は感想表現にならない場合が多く、感想の有無を同定するタスクにおいて精度を下げる要因となった [5]。

3 : <http://nlp.ist.i.kyoto-u.ac.jp/?JUMAN++>

4 : 形容詞と接尾辞の「ない」が連続した場合のみ代表表記に変換せず、形容詞と接尾辞の「ない」を一つの代表表記としてまとめる処理を行う。例えば、文字列「悪くない」の形態素解析結果を代表表記に変換すると、「悪い」、および、「ない」の列となるが、代表表記としては「悪くない」として記録する。

には、「石原さとみ」、「三浦春馬」、「堺雅人」、「竹内結子」、「米倉涼子」の各芸能人ごとに、30種類の形容詞の各々に対して10事例ずつを無作為抽出して合計1,500事例を選定し、データセットを作成する際の候補ツイートとした。

3 芸能人に対する感想マイニング・データセットの作成

本節では、前節で選定した候補ツイートに対して、芸能人名と形容詞との間での感想関係の有無の判定を人手で行った結果を蓄積し、「感想マイニング・データセット」を作成する。さらに、各芸能人名ごとに、形容詞との間の感想関係の有無の分布を集計した結果を示す。

3.1 芸能人名・形容詞間の感想関係の有無の人手付与の基準

1,500事例の候補ツイートに対し、キーワードとして指定した芸能人名と形容詞との間で感想関係が成り立つか否かを人手で判定する(図2)。なお、「形容詞によって、芸能人本人が修飾されているツイート」を芸能人名と形容詞の間で感想関係が成り立つ条件とし、この条件にあてはまらないツイートにおいては、キーワードとして指定した芸能人名と形容詞との間で感想関係が成り立たず、「感想関係なし」と判定する⁵。

3.2 感想マイニング・データセット

1,500事例の候補ツイートに対して、前節の基準に基づき、キーワードとして指定した芸能人名と形容詞との間で感想関係が成り立つか否かを人手で判定した結果は、以下の三通りに分けられる。

(a) キーワードとして指定した芸能人名と、ツイート中の一つの形容詞のみとの間で感想関係が成り立つ。ただし、ここでの形容詞は、各ツイート中で、2節においてツイート収集時点において指定された形容詞である必要はない⁶。

(b) キーワードとして指定した芸能人名との間では、ツイート中のいずれの形容詞も感想関係にない。

(c) キーワードとして指定した芸能人名と、ツイート中の二つ以上の形容詞との間で感想関係が成り立つ。

1,500事例のツイートのうち、上述の(c)に該当するツイートを除外し、その分のツイートを、2節における母集団のツイート集合から補充する⁷。以上の手順の結果、(c)に該当すると除外されたツイート数を表1「複数の形容詞と「感想関係あり」(除外対象)」欄に示す。また、表1においては、最終的に、

5: 例えば、キーワードとして指定している芸能人以外の芸能人を形容詞が修飾しているツイート、芸能人が出演しているテレビドラマ等を修飾しているツイート等が挙げられる。

6: 2節において、芸能人名 c と形容詞 a の組が共起する10事例のツイートのうちの一つとして収集されたツイート $t_i(c, a)$ ($1 \leq i \leq 10$)において、実際に芸能人名 c と感想関係にある形容詞が $a'(\neq a)$ となっても構わない。

7: 除外されたツイートが、芸能人名 c と形容詞 a の組が共起する10事例のツイートのうちの一つとして収集されたツイート $t_j(c, a)$ ($1 \leq j \leq 10$)である場合には、2節における母集団のツイート集合から、芸能人名 c と形容詞 a の組が共起するツイートを再度補充する。補充したツイートが再度(c)に該当する場合には、(a)または(b)に該当するツイートが補充されるまで、補充手続きを繰り返す。

1,500事例の範囲において(a)に該当するツイート数、および、(b)に該当するツイート数を、「一つの形容詞のみと「感想関係あり」」欄、および、「感想関係なし」欄にそれぞれ示す⁸。

4 芸能人に対する感想マイニングのモデル化

4.1 読解モデルによるモデル化

読解モデルによるモデル化では、図1の質問応答の枠組み(本論文ではBERT[2]を用いる)を利用する。この枠組みでは、芸能人名が共起しているツイートをコンテキスト C 、芸能人名をキーワード Q として、当該ツイート内において芸能人名 Q に対する感想が形容詞によって表現されている場合は該当する形容詞 A を出力し、感想が表現されていない場合には、「感想関係なし」を出力することにより、当該ツイート中における芸能人名 Q に対する感想の有無の判定を行う。

本論文では、BERTにおける予約トークンとしてコンテキスト C 内の全ての形容詞の前後に[ADJ]を追加したコンテキスト C 、芸能人名であるキーワード Q 、および、 Q との間で感想関係にある形容詞 A のコンテキスト C 内でのトークン位置 L_A 、もしくは、「感想関係なし」の組 A' をデータセットとして構築する。具体的には3.2節の(a)に該当するツイートの場合は、読解モデル[2]訓練用のコンテキスト C 、質問回答事例における質問 Q 、および、質問回答事例における回答 A のトークン位置 L_A の組として、以下の形式の事例を保存する。

コンテキスト C : 芸能人名としてのキーワード Q 、および、 Q と感想関係にある形容詞 A が共起しているツイート
質問 Q : 芸能人名としてのキーワード Q
回答 A のトークン位置 L_A : Q と感想関係にある形容詞 A のトークン位置

一方、3.2節の(b)に該当するツイートの場合は、以下の形式で、読解モデル[2]訓練用のコンテキスト C 、質問回答事例における質問 Q 、および、質問回答事例における回答 $A' = 「$

(空白)の組として、以下の形式の事例を保存する。
コンテキスト C : 芸能人名としてのキーワード Q を含み、キーワード Q とツイート内のすべての形容詞が感想関係にないツイート

質問 Q : 芸能人名としてのキーワード Q
回答 A' : 「」(空白)

一つの形容詞のみと「感想関係あり」の場合は、3.2節の(a)に該当する事例における形容詞 A のトークン位置 L_A 、「感想関係なし」の場合は3.2節の(b)に該当する事例における回答 A' を「参照用ラベル」とする。

4.2 分類モデルによるモデル化

分類モデルによるモデル化では、芸能人名が共起している

8: どの芸能人においても、(a)「一つの形容詞のみと「感想関係あり」」の事例数が、(b)「感想関係なし」の事例数を上回ることにはなかった。また、「三浦春馬」、および、「竹内結子」においては、複数の形容詞と「感想関係あり」となる事例数が、それぞれ、64例(約21%)、および、44例(約15%)と多かった。これらにおいては、例えば、「切なくて悲しい」や「寂しい、苦しい」等、ツイート投稿者が複数の感情表現をツイート中に含めている事例が多く観測された。

ツイートをコンテキスト C , 芸能人名をキーワード Q として, C 内に存在する全ての形容詞の列を A_1, \dots, A_n , コンテキスト C 内のキーワード Q の前後に予約トークン [KEY], 形容詞 $A_i (1 \leq i \leq n)$ の前後に予約トークン [ADJ] を付与し, 形容詞 $A_i (1 \leq i \leq n)$ がキーワード Q に対して, 「感想関係あり」もしくは「感想関係なし」の2クラス分類を BERT を用いて行う. 本論文で使用した分類モデルの枠組みを図3に示す. 本論文では, 最初に, コンテキスト C を n 個複製し, コンテキスト C_1, \dots, C_n とする. コンテキスト $C_i (1 \leq i \leq n)$ に対し, 形容詞 $A_i (1 \leq i \leq n)$ の前後に [ADJ], キーワード Q の前後に [KEY] を付与し, 形容詞 A_1, \dots, A_n のトークン位置の集合 AA を次式で定義する.

$$AA = \{L_i \mid i = 1, \dots, n\}$$

ただし, 本論文で対象とする事例としては, 3.2節の(a)に該当するキーワード Q が AA 中の一つのトークン位置 L_i の形容詞 $A_i (1 \leq i \leq n)$ のみと「感想関係あり」と人手で判定された場合, または, 3.2節の(b)に該当する AA 中のすべての形容詞と「感想関係なし」と人手で判定された場合の二通りのみを扱う.⁹ また, コンテキスト $C_i (1 \leq i \leq n)$ を (トークン位置 $L_i (1 \leq i \leq n)$ のみへの) 予約トークン埋め込み済みコンテキスト $C_i (1 \leq i \leq n)$ として定義する. そして, 予約トークン埋め込み済みコンテキスト $C_i (1 \leq i \leq n)$, キーワード Q , 形容詞トークン位置 $L_i (1 \leq i \leq n)$ の組 $\langle C_i, Q, L_i \rangle (1 \leq i \leq n)$ を集めた集合を

$$X(C, Q) = \{ \langle C_i, Q, L_i \rangle \mid i = 1, \dots, n \}$$

と定義する. この集合 $X(C, Q)$ を利用し, 訓練, 評価用の事例を生成し, 分類モデルによる訓練, 評価を行う. 以下に手順を示す.

「訓練時」

分類モデルの訓練データとして, 集合 $X(C, Q)$ の n 個の全要素から事例を生成する.

「集合 $X(C, Q)$ の一つの元のみがキーワード Q との間で人手による判定 h で, 「感想関係あり」となる場合¹⁰」

該当する組 $\langle C_i, Q, L_i \rangle (1 \leq i \leq n)$ を「感想関係あり」の訓練事例とし, その他の組 $\langle C_j, Q, L_j \rangle (j=1, \dots, n, j \neq i)$ を「感想関係なし」の訓練事例として, 分類モデルを訓練する.

「集合 $X(C, Q)$ の全ての元がキーワード Q との間で人手による判定 h で, 「感想関係なし」となる場合¹¹」

全ての組 $\langle C_i, Q, L_i \rangle (1 \leq i \leq n)$ を「感想関係なし」の訓練事例として, 分類モデルを訓練する.

「評価時」

集合 $X(C, Q)$ の各要素¹² に対して, 分類モデルが, 確率値 $p(\langle C_i, Q, L_i \rangle)$ で「感想関係あり」, または「感想関係なし」の判定を行う.

「集合 $X(C, Q)$ の一つ以上の元に対する分類モデル f の判定結果が「感想関係あり」となる場合¹³」

該当する一つ以上の組 $\langle C_i, Q, L_i \rangle (1 \leq k \leq n)$ のうち, 分類モデルによる確率値 $p(\langle C_i, Q, L_i \rangle)$ が最大となる組 $\langle C_{\hat{k}}, Q, L_{\hat{k}} \rangle (1 \leq \hat{k} \leq n)$ に対応する形容詞 $A_{\hat{k}}$ をコンテキスト C 内のキーワード Q に対する「分類モデルの予測結果」とする.

「集合 $X(C, Q)$ の全ての元に対する分類モデル f の判定結果が「感想関係なし」となる場合¹⁴」

コンテキスト C 内のキーワード Q に対しては, 「感想関係なし」を「分類モデルの予測結果」とする.

4.3 トークン分類モデルによるモデル化

トークン分類モデルによるモデル化では, 芸能人名をキーワード Q として, 芸能人名が共起しているツイート内に存在する全ての形態素の列を L_1, \dots, L_m として, 芸能人名が共起しているツイートを集合 C として次式で定義する.

$$C = \{L_i \mid i = 1, \dots, m\}$$

形態素 $L_i (1 \leq i \leq m)$ の「REVIEW」クラス, 「TARGET」クラス, 「NONE」クラスの3クラス分類¹⁵ を BERT を用いて行う. 本論文で使用したトークン分類モデルの枠組みを図4に示す.

「訓練時」

形態素 $L_i (1 \leq i \leq m)$ を「REVIEW」, 「TARGET」, 「NONE」の訓練事例として, トークン分類モデルを訓練する.

「評価時」

形態素 $L_i (1 \leq i \leq m)$ に対して, トークン分類モデルが, 確率値 $p(L_i (1 \leq i \leq m))$ で「REVIEW」, 「TARGET」, 「NONE」の判定を行う.

「芸能人名を表すキーワード Q に対する形態素 $L_i (1 \leq i \leq m)$ に対して, トークン分類モデル f の判定結果が「TARGET」となり, かつ, トークン分類モデル f の判定結果が「REVIEW」となる形態素 $L_j (1 \leq j \leq m)$ が存在する場合」

芸能人名を表すキーワード Q に対して, ツイート C は「感想関係あり」となるという結果を, 「トークン分類モデルの予測結果」とする.

9: 3.2節の(c)に該当する当該ツイート中でキーワード Q が AA 内の二つ以上の形容詞と「感想関係あり」, と人手で判定された場合の事例を除外した.

10: $\exists \langle C_i, Q, L_i \rangle \in X(C, Q), h(\langle C_i, Q, L_i \rangle) = \text{「感想関係あり」}$

11: $\forall \langle C_i, Q, L_i \rangle \in X(C, Q), h(\langle C_i, Q, L_i \rangle) = \text{「感想関係なし」}$

12: コンテキスト, キーワード, 形容詞のトークン位置の組 $\langle C_i, Q, L_i \rangle (i = 1, \dots, n)$

13: $\exists \langle C_i, Q, L_i \rangle \in X(C, Q), f(\langle C_i, Q, L_i \rangle) = \text{「感想関係あり」}$

14: $\forall \langle C_i, Q, L_i \rangle \in X(C, Q), f(\langle C_i, Q, L_i \rangle) = \text{「感想関係なし」}$

15: 「REVIEW」クラスは芸能人に対して感想関係を示す形態素のクラス, 「TARGET」クラスは芸能人名を示す形態素のクラス, 「NONE」クラスはそれ以外の形態素のクラスをそれぞれ示す.

芸能人名を表すキーワード Q に対する形態素 $L_i (1 \leq i \leq m)$ に対して、トークン分類モデル f の判定結果が「TARGET」とならないか、または、トークン分類モデル f の判定結果が「REVIEW」となる形態素 $L_j (1 \leq j \leq m)$ が存在しない場合

芸能人名を表すキーワード Q に対して、ツイート C は「感想関係なし」となるという結果を、「トークン分類モデルの予測結果」とする。

5 評価

5.1 評価手順

本論文では、読解モデル、分類モデル、トークン分類モデルともに BERT を PyTorch で実装した。BERT はタイトルを除く日本語 Wikipedia 全てで事前学習した、NICT BERT 日本語 Pre-trained モデル¹⁶を用いた。4節で作成した読解モデル、分類モデル、トークン分類モデルのデータセットのコンテキスト C に対し、Juman++を用いて形態素単位で分割を行った。続いて利用した BERT の仕様にに基づき、WordPiece モジュール¹⁷を適用し、語彙数 32,000 のサブワード単位へ分割を行った。fine-tuning においては、読解モデルでは Huggingface¹⁸ の question-answering モデルの fine-tuning モジュール¹⁹、分類モデルでは Huggingface の text-classification モデルの fine-tuning モジュール²⁰、トークン分類モデルでは Huggingface の token-classification モデルの fine-tuning モジュール²¹を利用した。評価においては、2つの評価方法で実験を行った。最初に評価対象の5人の芸能人を用いて、4人の芸能人のデータで訓練を行い、1人の芸能人のデータで評価を行う、5芸能人交差検証を行った。次にデータセットを芸能人名に関わらず無作為にツイート単位で5分割し、4つのデータ群で訓練し、1つのデータ群で評価を行う、無作為にツイート単位5分割交差検証を行った。

5.2 評価結果

本論文で評価対象とする読解モデル、読解モデル(予約トークンなし)²²、分類モデル、およびトークン分類モデルの予測結果に対してツイート単位で評価を行い、トークン分類モデル以外では BERT の softmax 関数の出力確率値の下限値 p_0 を降順に変化させた場合の適合率、再現率²³の推移をプロットした

結果、トークン分類モデルではモデル出力結果の適合率、再現率²⁴をプロットした結果²⁵を5芸能人交差検証で、モデル出力が「感想関係あり」の場合は、図5(a)、モデル出力が「感想関係なし」の場合は、図5(b)にそれぞれ示す。また、同様に無作為に5分割交差検証で、モデル出力が「感想関係あり」の場合は、図6(a)、モデル出力が「感想関係なし」の場合は、図6(b)にそれぞれ示す²⁶。5芸能人交差検証で、モデル出力が「感想関係あり」では、読解モデル(予約トークンなし)が高い精度を示した。一方、モデル出力が「感想関係なし」でも読解モデルが高い精度を示した。また、無作為に5分割交差検証では、モデル出力が「感想関係あり」では、分類モデルが高い精度を示した。一方、モデル出力が「感想関係なし」では読解モデルが高い精度を示した。また、5芸能人交差検証と無作為に5分割交差検証の評価の比較を行い、結果を図7に示す。この結果から、訓練事例において既知の芸能人に対する性能を評価する無作為に5分割交差検証に対しては、訓練事例において未知の芸能人に対する性能を評価する5芸能人交差検証の方がやや性能が落ちるものの、大幅な低下ではないことから、読解モデルでは、訓練事例において既知か否かに関わらず、感想の有無判定のモデル化ができていくことが分かる。

また、先行研究として、ドラマに関する研究[5]で構築されたデータセットで読解モデルによって訓練を行った。このデータセットは合計で4,327件のため、1,200件分のデータを無作為に抽出し、同様に各300件のデータを5芸能人分評価を行った。評価の結果を図8に示す。評価の結果、芸能人データで訓練したモデルの精度が高いことがわかった。

また、先行研究として、ツイートからのテーマパークなどの施設に対する感想マイニングを行った研究[7]で用いられた SVM の手法を利用し、芸能人データセットで訓練を行い、5芸能人交差検証を行った。評価の結果を図9に示す。評価の結果、BERT モデルの精度が高いことがわかった。

また、ツイート中において感想を表現する語として、形容動詞を対象に追加し、形容動詞が共起しているツイートデータセット(表2)²⁷を作成した。評価実験においては、2つの実験を行った。最初に5人の芸能人の形容詞データセット、評価

16: <https://alaginrc.nict.go.jp/nict-bert/index.html>

17: tokenization.py

18: transformers-2.2.1

19: run_squad.py を用い、訓練時のパラメータは、エポック数 2、バッチ数 8、学習率 0.00003 とした。

20: run_glue.py を用い、訓練時のパラメータは、エポック数 3、バッチ数 12、学習率 0.00002 とした。

21: run_ner.py を用い、訓練時のパラメータは、エポック数 2、バッチ数 8、学習率 0.00003 とした。

22: 4.1 節において、コンテキスト C 内の全ての形容詞の前後に [ADJ] を付与する手順を省略したモデル。

23: 読解モデルでは、モデルが出力した回答文字列と人手で感想関係の有無の判定結果を付与した答えと比較し、「完全一致」、モデルの回答に過不足があるなど

の「部分一致」、「不一致」とし、「完全一致」と「部分一致」を正答、「不一致」を誤答とした。分類モデルではモデル出力が「感想関係あり」「感想関係なし」の2クラスなので、人手で感想関係の有無の判定結果を付与した答えと同じ形容詞に対して「感想関係あり」を出力している場合、もしくは、共起する全ての形容詞に対して人手で感想関係の有無の判定結果が「感想関係なし」の事例で「感想関係なし」を出力している場合を正答、それ以外は誤答とする。

24: 人手で感想関係の有無の判定結果を付与した答えと同じ形容詞に対して「REVIEW」ラベルを出力している場合、もしくは、共起する全ての形容詞に対して人手で感想関係の有無の判定結果が「NONE」ラベルの事例で全ての形容詞に対して「NONE」ラベルを出力している場合を正答、それ以外は誤答とする。また、本論文では「TARGET」ラベルの予測結果は考慮しない。

25: トークン分類モデルでは BERT の softmax 関数の出力は形態素ごとのため、ツイート単位での評価時に PR 曲線は描けない。

26: これらの結果において、トークン分類モデルにおいては、訓練事例において未知の芸能人に対する感想有無判定が全くできず、5芸能人交差検証(図5)における再現率・適合率がゼロとなったため、図中ではプロットしていない。

27: 対象芸能人を「石原さとみ」、「堺雅人」、「米倉涼子」の3人に限定し、作成手法は形容詞が共起しているデータセットと同様である。

対象の3人の形容動詞データセットを用いて、4人の芸能人の形容動詞データで訓練を行い、1人の芸能人の形容動詞データで評価を行う、3芸能人交差検証²⁸を行った。次に評価対象の5人の芸能人の形容動詞データセット、評価対象の3人の形容動詞データセットを用いて、各芸能人の形容動詞データセットの半分、および全形容動詞データで訓練を行い、1人の芸能人の形容動詞データで評価を行う3芸能人交差検証²⁹を行った。評価の結果を図10に示す。評価の結果からモデル出力が「感想関係あり」では、形容詞が共起しているデータセットのみで訓練、形容動詞が共起しているデータセットで評価したモデルより、形容詞が共起しているデータセット・形容動詞が共起しているデータセットを混ぜて訓練、形容動詞が共起しているデータセットで評価したモデルが高い精度を示した。形容詞が共起しているデータセットのみで訓練したモデルには、訓練データに形容動詞を回答として与えていない。一方で、形容詞が共起しているデータセット・形容動詞が共起しているデータセットの両方のデータセットを混ぜて訓練したモデルには形容動詞を回答として与えていることが、後者のモデルが前者のモデルより精度が高くなった原因と考えられる。このため、読解モデルは訓練データとして与えられていない未知の形容詞以外の語を正しく回答できていないことがわかった。

6 関連研究

ツイートからのレビュー抽出の関連研究として、テーマパークに対する意見抽出に関する研究[7]、および、主観に基づく主張や意見を抽出する研究[1]等がある。一方、本論文におけるタスクは、施設や主張ではなく、芸能人に対する感想のマイニングであり、問題設定が異なる。また、レビューマイニングの関連研究として、ツイート内で言及されているテレビドラマ出演俳優・登場人物に対して感想関係の有無を判定する研究[5]、テレビドラマに関するウェブページにおいてレビューの有無を判定する研究[4]、テレビドラマに関するウェブページの記事タイトルにおけるインパクトを判定する研究[3]が挙げられる。一方、本論文では、テレビドラマ出演俳優に限らず、芸能人全般を対象として、ツイート中の感想マイニングをタスクとしており、これらの関連研究とは対象が異なる。また、本論文の類似タスクとして、ツイートから評価対象や評価表現を同定する研究[6]が挙げられるが、この関連研究がトークン分類モデルでモデル化しているのに対し、本研究では読解モデルでのモデル化を提案しているため、手法が異なる。

7 おわりに

本論文では、媒体としてツイートを情報源として、話題となっている芸能人に対する感想を表現したツイートを収集・集約する手法を提案した。具体的にはツイート、芸能人名をそれ

ぞれコンテキスト、キーワードとして、BERTによって、この二つからコンテキスト内に共起している形容詞に感想関係がある場合は該当する形容詞を出力、感想関係がない場合は空白を出力することで機械読解の枠組みによってモデル化を行い、一定以上の性能で感想関係の有無を判定できることが示された。

文 献

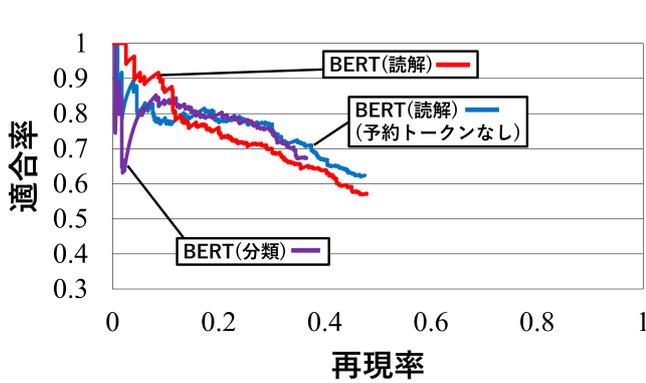
- [1] S. Chen, D. Khashabi, W. Yin, C. Callison-Burch, and D. Roth. Seeing things from a different angle: Discovering diverse perspectives about claims. In *Proc. NAACL-HLT*, pp. 542–557, 2019.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional Transformers for language understanding. In *Proc. NAACL-HLT*, pp. 4171–4186, 2019.
- [3] 藤田拓也, 久保遼馬, 川口輝太, 前田竜治, 宇津呂武仁, 小林彰夫, 西崎博光, 河田容英. テレビドラマに関する関心動向・感想のウェブマイニングにおけるインパクトの分析. 第12回 DEIM フォーラム論文集, 2020.
- [4] 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.
- [5] 久保遼馬, 藤田拓也, 宇津呂武仁, 小林彰夫, 西崎博光, 河田容英. テレビドラマ出演俳優・登場人物を話題とするツイートにおけるBERTを用いた感想マイニング. 第34回人工知能学会全国大会論文集, 2020.
- [6] 栗原理聡, 水本智也, 乾健太郎. Twitterによる評判分析を目的とした評価対象-評価表現データセット作成. 言語処理学会第24回年次大会発表論文集, 2018.
- [7] 野崎雄太, 櫻井義尚. Twitterからの意見抽出モデル構築のための教師データ作成手法. 情報処理学会論文誌: 数理モデル化と応用, Vol. 13, No. 2, pp. 69–83, 2020.

28: 訓練: 4 芸能人形容動詞データセット (1,200 件), 評価: 1 芸能人形容動詞データセット (300 件)

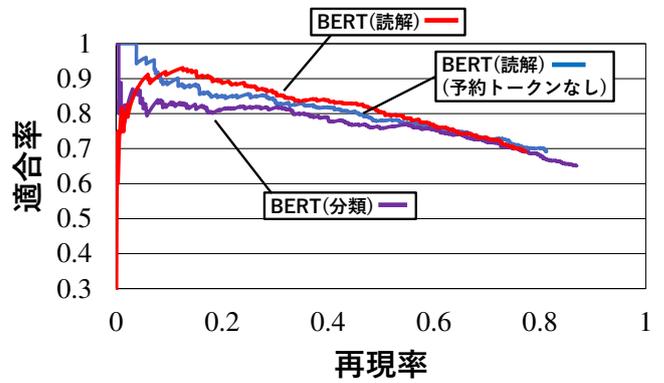
29: 訓練: 4 芸能人の形容動詞データセットの半分 (600 件), 2 芸能人の形容動詞データセット (600 件), 評価: 1 芸能人の形容動詞データセット (300 件)

表 2 収集対象, および, データセット作成対象のツイート数 (形容詞・形容動詞を対象とした場合)

芸能人名	ツイート数	リツイート以外のツイート数	収集期間	訓練・評価用ツイート数			複数の形容詞 (または, 形容動詞) と「感想関係あり」 (除外対象)
				一つの形容詞 (または, 形容動詞) のみと「感想関係あり」	「感想関係なし」	合計	
石原さとみ	1,677,692	555,701	2020/8/2 ~ 2020/10/24	55	245	300	8
堺雅人	161,404	61,108	2020/8/2 ~ 2020/10/24	111	189	300	8
米倉涼子	37,445	14,869	2020/8/2 ~ 2020/10/24	97	203	300	3
合計	-	-	-	263	637	900	19

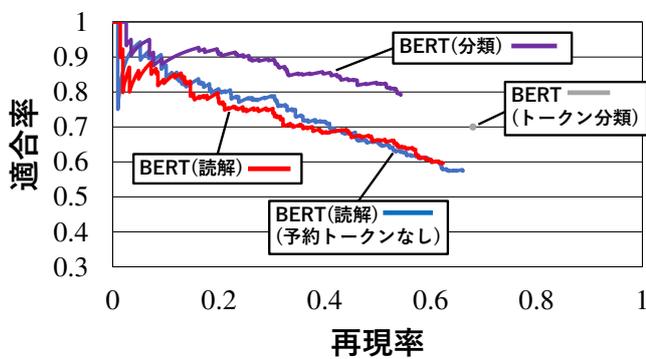


(a) モデル出力が「感想関係あり」の場合

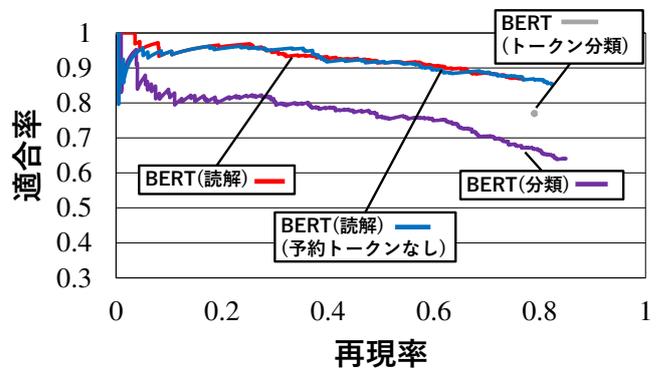


(b) モデル出力が「感想関係なし」の場合

図 5 ツイート単位の評価結果 (5 芸能人交差検証)



(a) モデル出力が「感想関係あり」の場合



(b) モデル出力が「感想関係なし」の場合

図 6 ツイート単位の評価結果 (ツイート単位無作為 5 分割交差検証)

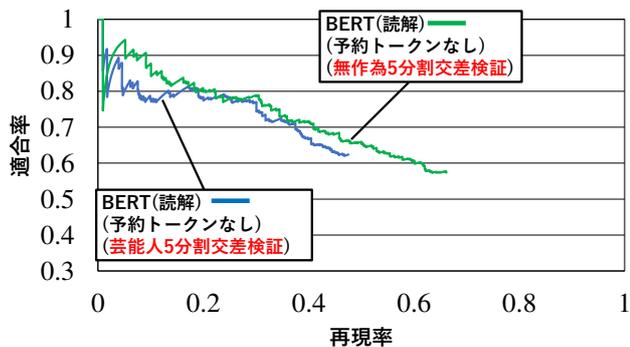


図7 無作為交差検証との比較評価結果 (モデル出力が「感想関係あり」の場合)

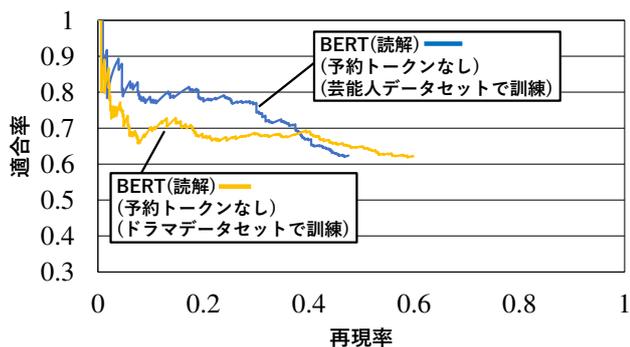


図8 ドラマデータセットとの比較評価結果 (モデル出力が「感想関係あり」の場合)

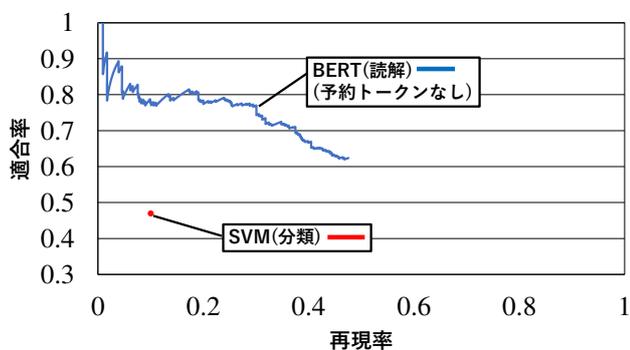


図9 SVM との比較評価結果 (モデル出力が「感想関係あり」の場合)

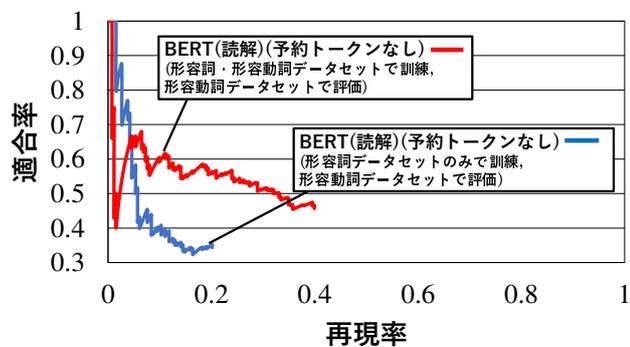


図10 形容動詞を対象とした評価結果 (モデル出力が「感想関係あり」の場合)