

ツイートにおける芸能人の評価対象に対する感想のマイニング

Mining Reviews focusing on Celebrities' Aspects in Tweets

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 本論文は、芸能人のファンが、芸能人および芸能人に関する事柄に対する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。そのため本論文では、Twitter において芸能人に関する事柄について述べているツイートを収集し、かつ、感想を述べているか否かを判定してデータとして蓄積する機械学習モデルを構築する。具体的には、収集したツイート中に含まれる芸能人の評価対象名と形容詞・形容動詞間で感想を述べているか否かに対して、BERT を用いたトークン分類モデルで判定する。評価実験の結果、高い精度を達成し、このタスクにおいて BERT を用いたトークン分類モデルを用いることが有効であることが示された。

キーワード ツイート, 芸能人, 評価対象, 評判, 感想

1 序 論

本論文は、芸能人のファンが、芸能人やそれに関する事柄について批評等の情報探索を行いやすくすることを大きな目的とする。そのために、Twitter のツイートをデータとし、芸能人に関する事柄に対してレビューマイニングを行う手法を提案する。具体的には、芸能人に関する評価対象を含むツイートを収集し、かつ、感想を述べているか否かを判定してデータとして蓄積するモデルを構築する。ここで、芸能人に関する評価対象とは、「X(芸能人名)のY(名詞)」や、「X(芸能人名)のZな(名詞の連体形, 形容詞等, 直後の「Y(名詞)」を修飾する単語)Y(名詞)」等、様々考えられるが、本論文では「X(芸能人名)のY(名詞)」を対象とし、具体的には「石原さとみのY(名詞)」, 「堺雅人のY(名詞)」, 「山田涼介のY(名詞)」とする。芸能人に関する評価対象に対して感想を述べているか否かを判定するために、本論文では、感想を表現する語として形容詞、および形容動詞を対象とする。

芸能人の評価対象を含むツイートを収集し、かつ、感想を述べているか否かを判定してデータとして蓄積するモデルとして、BERT [2] を用いる。また、モデルの枠組みとして、図 1 に示すトークン分類モデルを用いる。

具体的には、芸能人に関する評価対象名を含むツイートをコンテキスト C 、芸能人に関する評価対象名をキーワード Q 、キーワード Q に対して感想を述べている形容詞、もしくは形容動詞を A とする。まず、芸能人に関する評価対象名であるキーワード Q であるか否かに関わらず、コンテキスト C 中に出現する全ての姓名を $[LAST_NAME][FIRST_NAME]$ とトークンに変換して判定を行う。この時、キーワード Q である「X(芸能人名)のY(名詞)」は「 $[LAST_NAME][FIRST_NAME]$ の

Y(名詞)」と変換される。次に、コンテキスト C 内の全ての形態素の列 L_1, \dots, L_n に対してキーワード Q であることを表す TARGET ラベル、キーワード Q に対して感想を述べている形容詞、もしくは形容動詞 A であることを表す REVIEW ラベル、それ以外であることを表す NONE ラベルを付与する。「感想関係あり」の場合は、該当するコンテキスト C 内において芸能人に関する評価対象名 Q に対して感想を述べている形容詞、もしくは形容動詞 A に対して REVIEW ラベルを付与し、構築したモデルで出力する。「感想関係なし」の場合は、該当するコンテキスト C に対して、REVIEW ラベルを付与せずに構築したモデルで出力することによって、ツイート中の芸能人に関する評価対象名 Q に対して感想を述べているか否かの判定を行う。

データセットの作成においては、まず「石原さとみ」, 「堺雅人」, 「山田涼介」をキーワードに収集したツイートをを用いる。収集したツイート群の中から、それぞれのツイート内で形容詞と共起している各芸能人の評価対象の頻度分布を求め、一定以上の頻度で出現した評価対象を含むツイートに絞り込む。次に、出現頻度 1 位となった評価対象を含むツイートに対しては、共起している形容詞の頻度分布を求め、一定以上の頻度で出現した形容詞を含むツイートを無作為に選定し、データセットに蓄積する。それ以外の一定以上の頻度で出現した評価対象を含むツイートに対しては、それらをキーワードとして用いて無作為に選定し、データセットに蓄積する。そして、蓄積したツイートに対して、形容詞・形容動詞と、芸能人の評価対象との間で感想関係があるか否か人手で判定することによって、データセットを構築する。

当該データセットを用いてトークン分類モデルの訓練・評価を行い、芸能人名・名詞がいずれも未知であったとしても、TARGET ラベル、REVIEW ラベルそれぞれで一定の精度を

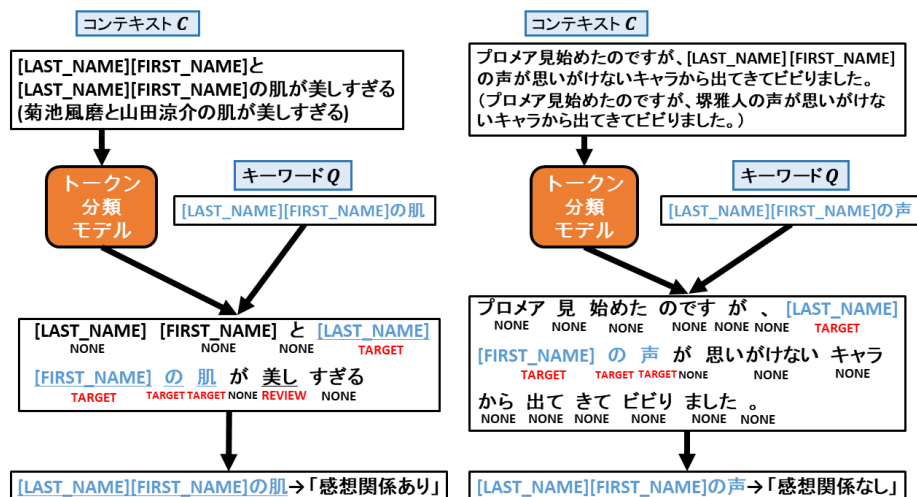


図 1: 「芸能人を話題とするツイートにおける評価対象」と形容詞・形容動詞間の感想関係の有無のトークン分類モデルの枠組み

表 1: 収集したツイート数とその期間

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
石原さとみ	1,677,692	555,701	2020/8/2 ～ 2020/10/24
堺雅人	161,404	61,108	2020/8/2 ～ 2020/10/24
山田涼介	378,656	73,611	2022/8/22 ～ 2022/10/23

示すことが確認された。

2 芸能人に関する評価対象を含む感想ツイート候補の選出

本節では、芸能人の評価対象を含む感想ツイートの選出方法について述べる。

選出方法について、はじめに、Twitter 上で話題になることが多い芸能人の一例である「山田涼介」をキーワードにツイートの収集を行った¹。収集には、Twitter Search API²を用いた。さらに、同条件で先行研究[10]で収集された「石原さとみ」と「堺雅人」をキーワードにツイートの収集を行ったデータを本研究では使用した。データセット中における各芸能人をキーワードとする「ツイート数」、「リツイート以外のツイート数」、「収集期間」は表 1 の通りである。次に、JUMAN++³を用

いて、収集されたツイートに対して形態素解析を行う。そして、その結果を用いて各ツイート内に「X(芸能人名)のY(名詞)」が含まれ、かつ、形容詞と共起している「Y(名詞)」の出現頻度を求め、収集対象として絞り込んだ。具体的には、各ツイート内で形容詞と共起している「石原さとみのY(名詞)」、「堺雅人のY(名詞)」、「山田涼介のY(名詞)」のそれぞれの「Y(名詞)」の出現頻度を求めた。結果の一部である出現頻度上位 31 位までの「Y(名詞)」の一覧とその頻度は図 2 に記載された通りとなっている。

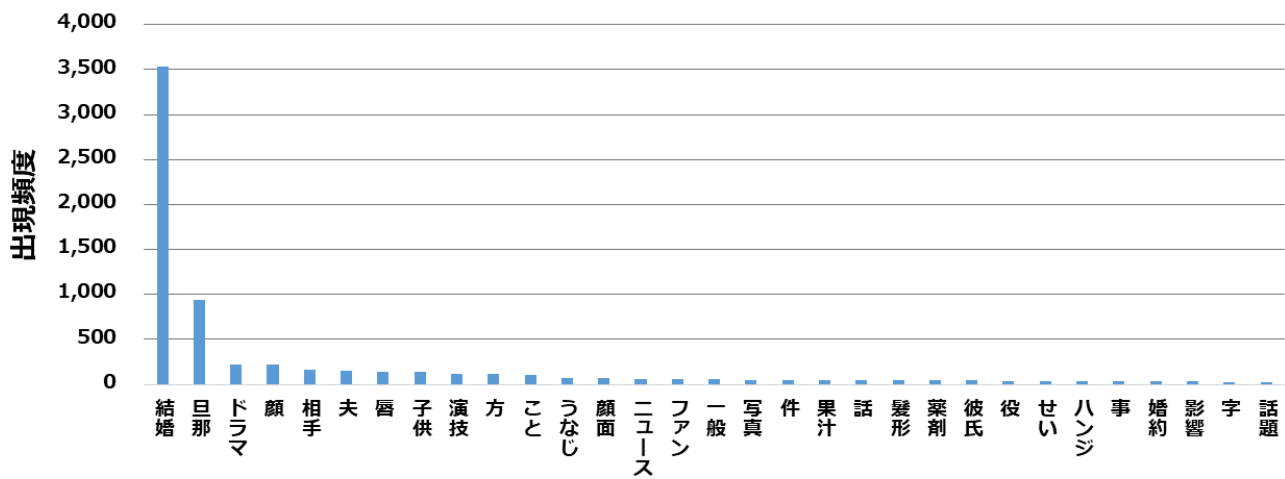
この後の処理は、各芸能人で出現頻度 1 位となった名詞と、それらを除いた上位 30 個の名詞 (出現頻度上位 2 位～31 位の名詞) で分かれる。

まず、各芸能人で出現頻度 1 位となった「石原さとみの結婚」、「堺雅人の演技」、「山田涼介のドラマ」はそれぞれを上記で収集したツイート群に対してキーワードとして用いる。そして、JUMAN++で形態素解析を行った結果から、各形容詞を代表表記 (終止形) に変換して共起している形容詞の頻度統計を求めた。結果の一部である出現頻度上位 30 位までの形容詞の一覧とその頻度は図 3 に記載された通りである。そして、それぞれで出現頻度上位 30 位となった形容詞をキーワードとし

1: 「山田」や「涼介」などの姓名どちらか一方のみ、「山田リョースケ」や「やまだりょうすけ」などの平仮名表記やカタカナ表記で芸能人名を記したツイートは収集の対象外としている。

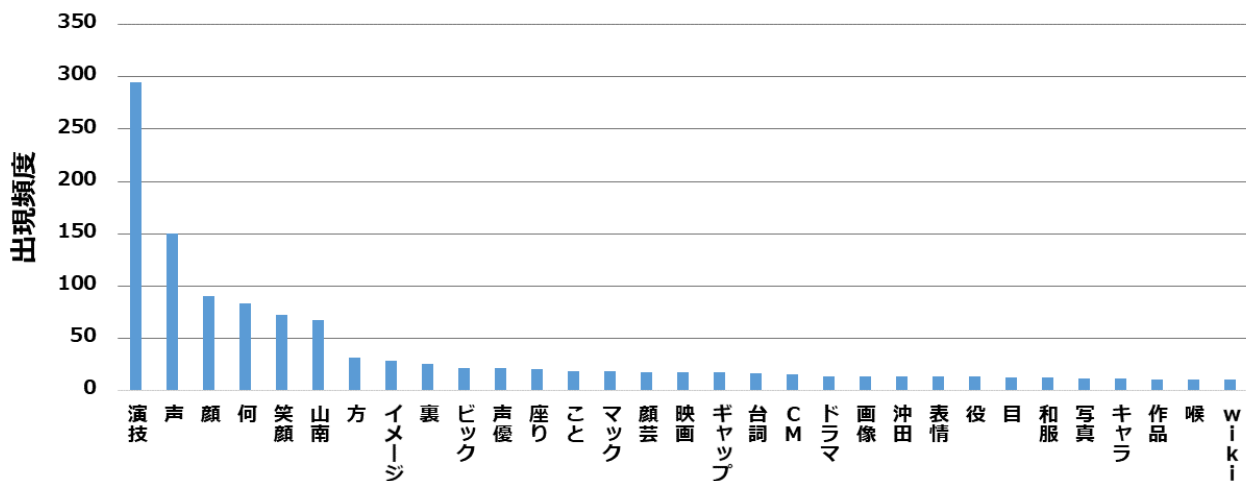
2: <https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets>

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



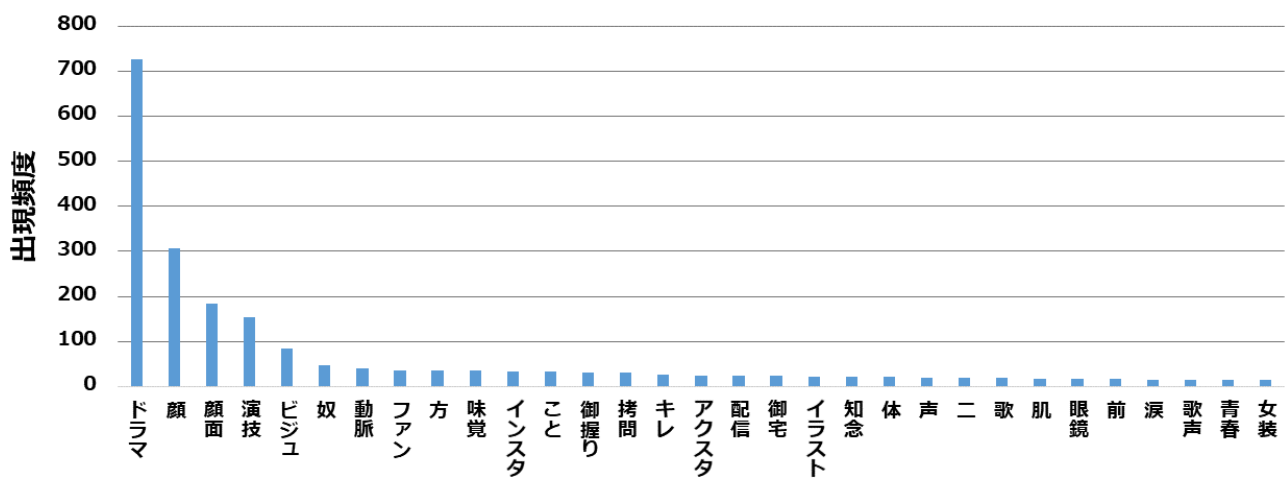
「石原さとみのY (名詞)」

(a) 「石原さとみの Y(名詞)」



「堺雅人のY (名詞)」

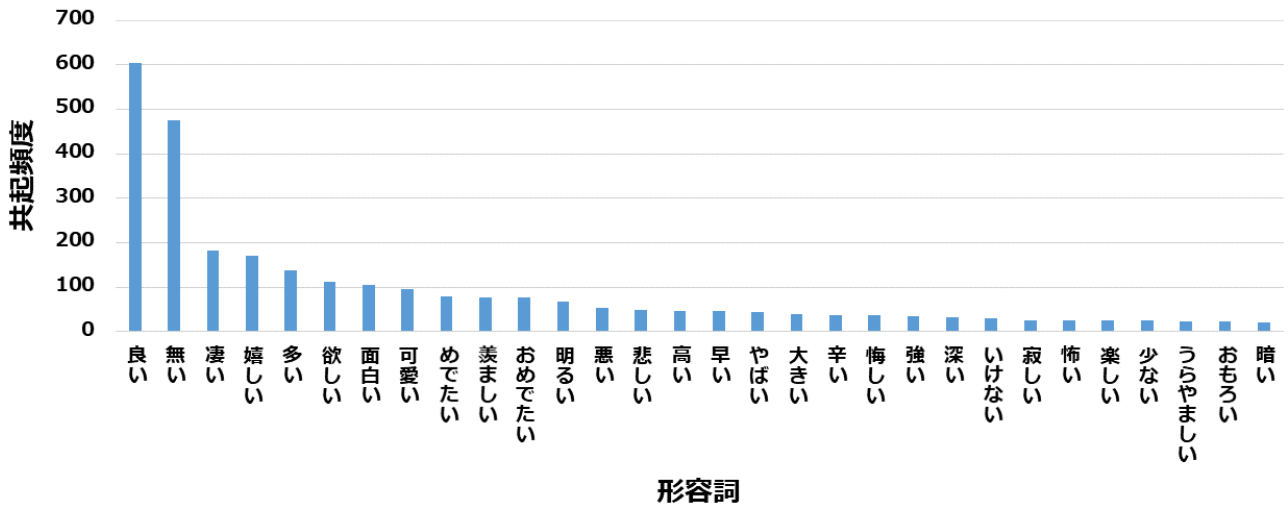
(b) 「堺雅人の Y(名詞)」



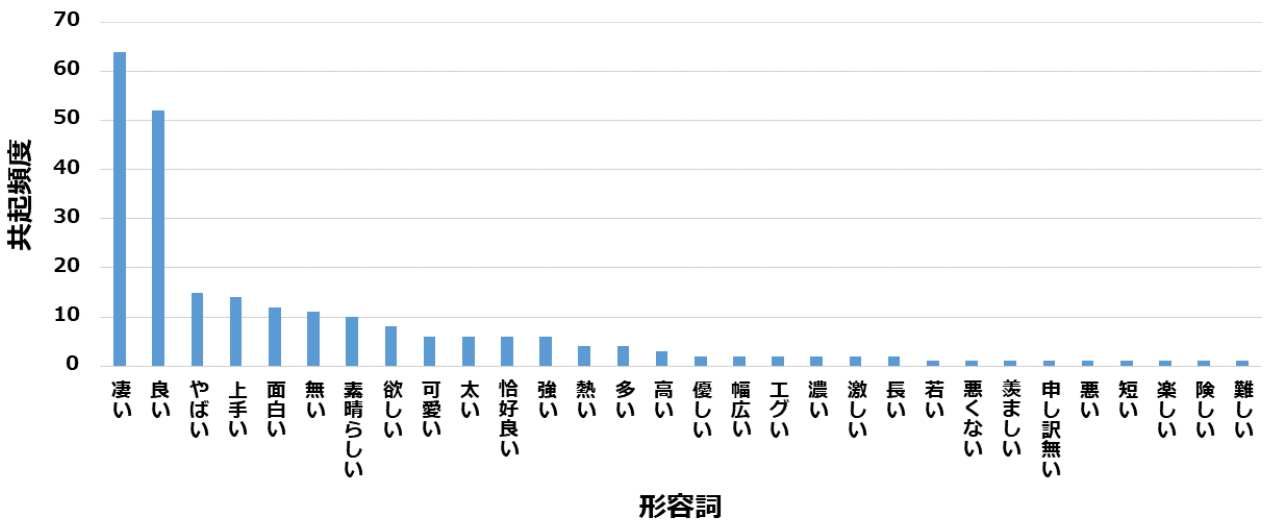
「山田涼介のY (名詞)」

(c) 「山田涼介の Y(名詞)」

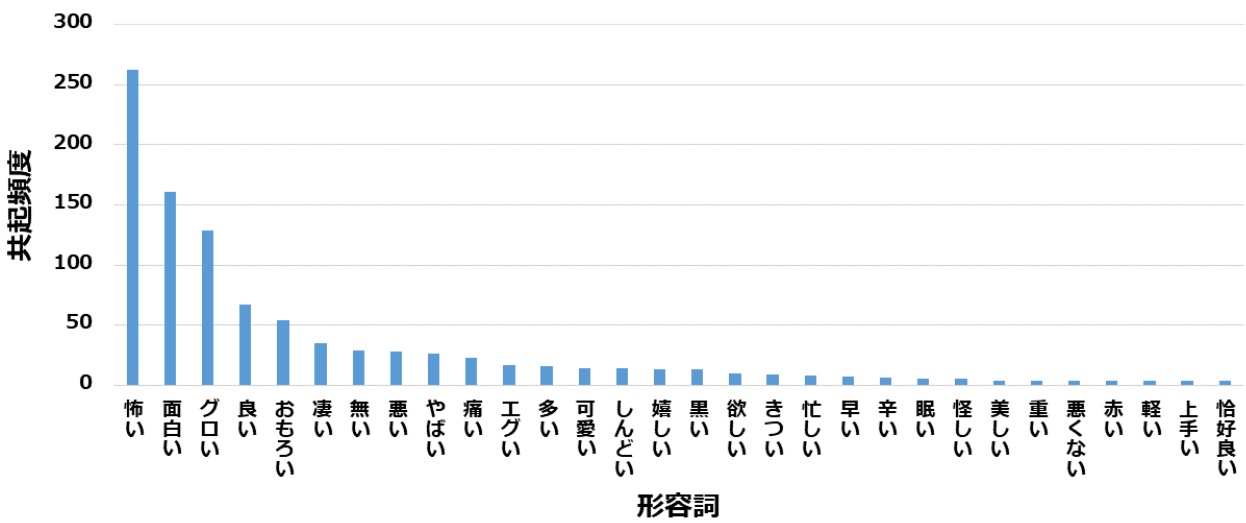
図 2: 各芸能人ごとの「X(芸能人名)のY(名詞)」における「Y(名詞)」の出現頻度統計 (頻度上位 30 種類)



(a) 「石原さとみの結婚」



(b) 「堺雅人の演技」



(c) 「山田涼介のドラマ」

図 3: 出現頻度 1 位となった各芸能人の評価対象名と共起して出現した形容詞の頻度統計 (頻度上位 30 種類)

て用い、「石原さとみの結婚」,「堺雅人の演技」,「山田涼介のドラマ」を含むツイートから最大 10 事例ずつ無作為抽出し、結果として 684 事例を抽出した。

「石原さとみの結婚」と「堺雅人の演技」,「山田涼介のドラマ」を除いた上位 30 個の名詞 (出現頻度上位 2 位~31 位の名詞) は、「石原さとみの Y(名詞)」,「堺雅人の Y(名詞)」,「山田涼介の Y(名詞)」のようにキーワードとして用いて 10 事例ずつ無作為抽出し、結果として 900 事例を抽出した。

最終的に 1,584 事例を抽出し、3 節でデータセットの構築に使用した。

3 芸能人を話題とするツイートにおける評価対象・感想の有無のデータセットの作成

本節では、2 節で抽出した 1,584 事例のツイートに対して、芸能人の評価対象と、形容詞・形容動詞との間で感想関係があるか否かを人手で判定し、データセットを作成する。

3.1 データセットの形式

本論文でデータセットを構築するにあたり、芸能人の評価対象名をキーワード Q 、キーワード Q に対して感想を述べている形容詞、もしくは形容動詞を A として、芸能人の評価対象名 Q を含むコンテキスト C 内に存在する全ての形態素の列を L_1, \dots, L_n とする。その上で、芸能人の評価対象名を含むコンテキストを集合 C として次式で定義する。

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

まず、図 1 のモデルに入力するために、ツイート中に出現する全ての姓名を [LAST_NAME][FIRST_NAME] とトークンに変換する処理を施す。具体的には、コンテキスト C 内に存在する全ての形態素の列 L_1, \dots, L_n に対して、JUMAN++中の形態素解析結果にある「意味情報」欄で下記のいずれかを含むものを姓、もしくは名として判定した。

(i) 「人名: 日本: 姓: …」, もしくは「人名: 日本: 名: …」という記載が含まれているもの^{4 5}。

(ii) 「自動獲得: Wikipedia…Wikipedia 人名 Wikipedia 姓」, もしくは「自動獲得: Wikipedia…Wikipedia 人名 Wikipedia 名」という記載が含まれているもの^{6 7}。

上記の条件で姓と判定され、かつ直後の形態素が上記の条件で名と判定された全ての形態素列に対して、「姓名 (例えば、筑波太郎など)」を「[LAST_NAME][FIRST_NAME]」とトークンに変換した。この時、「X(芸能人名) の Y(名詞)」は「[LAST_NAME][FIRST_NAME] の Y(名詞)」と変換された。

次に、「感想関係あり」に分類されたツイートに対しては、

4: 続く記載には、当該姓名の頻度順位と相対頻度を表す数字が載せられている。

5: JUMAN++中で用いられている辞書のうち、基本語彙辞書を参照した姓名に対してこの結果が出力される。

6: 途中の記載には、当該姓名の読み方が載せられている。

7: JUMAN++中で用いられている辞書のうち、Wikipedia 辞書を参照した姓名に対してこの結果が出力される。これは、基本語彙辞書に含まれていない姓名を Wikipedia のエントリ中から自動的に選択、および引用をして結果を付与したものである。

TARGET ラベル, REVIEW ラベル, NONE ラベルを用いて、以下の形式でデータセットに蓄積する。

付与するラベルの種類

TARGET, REVIEW, NONE

形態素列 L_1, \dots, L_n 内での各ラベルの付与位置

TARGET ラベル:

芸能人の評価対象名であるキーワード Q に該当する形態素列
REVIEW ラベル:

Q と感想関係が成立する形容詞、もしくは形容動詞 A に該当する形態素列

NONE ラベル:

上記 TARGET ラベル, REVIEW ラベルが付与されていない全ての形態素列

具体的な例として、図 1 の「[LAST_NAME][FIRST_NAME] と [LAST_NAME][FIRST_NAME] の肌が美しすぎる」(姓名をトークンに変換する前の元の文章は「菊池風磨と山田涼介の肌が美しすぎる」である) という文章を挙げる。この文章の場合、キーワード Q に該当する「[LAST_NAME][FIRST_NAME] の肌 (山田涼介の肌)」に TARGET ラベルが付与される。さらに、それに出現して共起している形容詞「美し」が「[LAST_NAME][FIRST_NAME] の肌 (山田涼介の肌)」に対して感想を述べていることから、REVIEW ラベルが付与される。それ以外の形態素には NONE ラベルが付与される。

一方、「感想関係なし」に分類されたツイートは、芸能人の評価対象名であるキーワード Q の間で感想関係が成立する形容詞・形容動詞 A が存在しないことになる。そのため、REVIEW ラベルは用いず、TARGET ラベル, NONE ラベルのみを用いて、以下の形式でデータセットに蓄積する。

付与するラベルの種類

TARGET, NONE

形態素列 L_1, \dots, L_n 内での各ラベルの付与位置

TARGET ラベル:

芸能人の評価対象名であるキーワード Q に該当する形態素列
NONE ラベル:

上記 TARGET ラベルが付与されていない全ての形態素列

具体的な例として、図 1 の「プロメア見始めたのですが、[LAST_NAME][FIRST_NAME] の声が思いがけないキャラから出てきてビビりました。」(姓名をトークンに変換する前の元の文章は「プロメア見始めたのですが、堺雅人の声が思いがけないキャラから出てきてビビりました。」である) という文章を挙げる。この文章の場合、キーワード Q に該当する「[LAST_NAME][FIRST_NAME] の声 (堺雅人の声)」に TARGET ラベルが付与される。しかし、それに共起して出現している形容詞「思いがけない」は「[LAST_NAME][FIRST_NAME] の声 (堺雅人の声)」に対して感想を述べていないことから、REVIEW ラベルではなく NONE ラベルが付与される。それ以外の形態素にも同様に NONE ラベルが付与される。

3.2 データセット作成・分類基準

2 節で抽出した 1,584 事例のツイートに対して、感想を表現

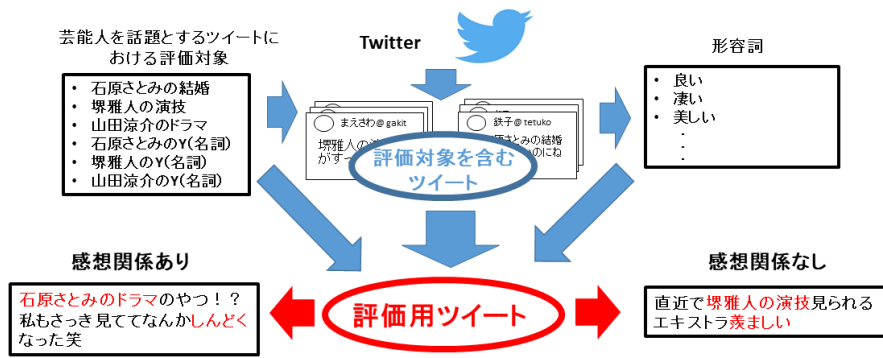


図 4: 芸能人を話題とするツイートにおける評価対象と形容詞・形容動詞間の感想関係の有無の人手付与

表 2: データセットを作成する際のキーワードとそのツイート数

収集キーワード	評価用ツイート数			複数の形容(動)詞と「感想関係あり」(除外対象)
	1つの形容(動)詞のみと「感想関係あり」	感想関係なし	合計	
「石原さとみの Y(名詞)」	109	191	300	20
「堺雅人の Y(名詞)」	148	152	300	18
「山田涼介の Y(名詞)」	139	161	300	20
「石原さとみの結婚」	75	225	300	15
「堺雅人の演技」	106	35	141	12
「山田涼介のドラマ」	131	85	216	71
合計	708	849	1,557	156

する語として形容詞、および形容動詞を対象とし、芸能人の評価対象との間で感想関係があるか否かを図 4 のように人手で判定する。

感想関係があるか否かの判定基準は下記の基準 (1)(2) を用いる。基準 (1)(2) に該当するツイートの場合、感想関係が成り立つと判定し、「感想関係あり」としてデータセットに蓄積する。一方、基準 (1)(2) に該当しないツイートの場合、感想関係は成り立たないと判定し、「感想関係なし」としてデータセットに蓄積する。

(1) 形容詞・形容動詞が芸能人の評価対象そのものに対して修飾しているツイート。

(2) 芸能人の評価対象そのものを修飾している投稿者の感情表現中に形容詞・形容動詞が含まれるツイート。

さらに、判定したツイートは以下の (a)(b)(c) の 3 通りに分類する。

(a) キーワードとして指定した芸能人の評価対象と、1 つの形容詞、もしくは 1 つの形容動詞との間で感想関係が成立する。1 つの形容詞との間で感想関係が成立する場合、必ずしも 2 節でのツイート抽出時に各ツイートで共起しているとされた形容詞間で感想関係が成立していても良いものとする⁸。

8: 2 節において、芸能人の評価対象 c と形容詞 a の組が共起しているとして、例えば、10 事例あるうちの 1 つとして抽出されたツイート $t_i(c, a)$ ($1 \leq i \leq 10$) において、実際には芸能人の評価対象 c と感想関係にある形容詞が $a'(\neq a)$ となっても良いものとする。

(b) キーワードとして指定した芸能人の評価対象と、2 つ以上の形容詞・形容動詞との間で感想関係が成立する。

(c) キーワードとして指定した芸能人の評価対象と、いずれの形容詞・形容動詞においても感想関係が成立しない。

1,584 事例のうち、(b) に分類されたツイートを除外し、その分のツイートを 2 節で述べた母集団のツイート集合から補充する⁹。この時、「堺雅人の演技」と「山田涼介のドラマ」は、出現頻度上位 30 位の形容詞の中で出現頻度が 10 事例にも満たない形容詞が出現したため、補充はせず、結果的に 1,557 事例をデータセットに蓄積した。

この集計結果は、表 2 では (a) に分類されたツイート数は『1 つの形容(動)詞のみと「感想関係あり」』の項目、(b) に分類されたツイート数は『複数の形容(動)詞と「感想関係あり」(除外対象)』の項目、(c) に分類されたツイート数は「感想関係なし」の項目に記載された通りとなっている。

9: 例えば、除外されたツイートが出現頻度 1 位の芸能人の評価対象 c と形容詞 a の組が共起する 10 事例の中の 1 つとして抽出されたツイート $t_j(c, a)$ ($1 \leq j \leq 10$) とする。その場合、2 節で述べた母集団のツイート集合から再度芸能人の評価対象 c と形容詞 a の組が共起するツイートを補充する。そして、感想関係有無の判定を行い、補充したツイートが (b) に該当する場合には、(a) または (c) に該当するツイートが現れるまで補充作業を繰り返す。除外されたツイートが出現頻度 2 位~31 位の芸能人の評価対象 c の場合も同様に、母集団のツイート集合から再度該当する芸能人の評価対象 c と形容詞が共起するツイートを補充し、(a) または (c) に該当するツイートが現れるまで補充作業を繰り返す。

4 評価

4.1 評価手順

本論文では、訓練・評価をするにあたり、トークン分類モデルで用いる BERT を Pytorch で実装した。BERT は日本語の Wikipedia のタイトルを除く全文章を事前学習した、NICT BERT 日本語 Pre-trained モデル¹⁰を使用した。3 節で作成したデータセット内のコンテキスト C に対しては、JUMAN++ で形態素解析を行った。その後、BERT の仕様に基づき、Word-Piece モジュール¹¹ で語彙数 32,000 のサブワード分割を行った。fine-tuning においては、Huggingface¹² の token-classification モデルの fine-tuning モジュールを用いた¹³。

訓練および評価においては、無作為にツイート単位 5 分割検証を行った。具体的には、出現頻度 1 位となった評価対象である「石原さとみの結婚」、「堺雅人の演技」、「山田涼介のドラマ」のデータ全 657 件を訓練データとし、出現頻度 2 位～31 位となった評価対象である「石原さとみの Y(名詞)」、「堺雅人の Y(名詞)」、「山田涼介の Y(名詞)」のデータ全 900 件に対して「Y(名詞)」を基に訓練データ、評価データのいずれかに分類した。まず、全データ 1,557 件に対して芸能人名に関わらず出現する全ての姓名を [LAST_NAME][FIRST_NAME] とトークンに変換した後、出現頻度 2 位～31 位となった評価対象である「[LAST_NAME][FIRST_NAME] の Y(名詞)」のデータ全 900 件に対して名詞の種類数を求めた。それを基に名詞を無作為に 5 グループに分類し、4 グループを訓練データ用、1 グループを評価データ用のグループとした。そして、ツイート中でキーワード Q として指定した「[LAST_NAME][FIRST_NAME] の Y(名詞)」の「Y(名詞)」がいずれかのグループ中に含まれる名詞に該当した場合、当該ツイートを訓練データ、評価データの該当するいずれかに分類し、訓練および評価を行った。

評価指標においては、TARGET ラベルは感想関係の有無に関わらず、全てのツイートに対し必ず付与されるため、予測ラベルで参照ラベルと同一形態素に TARGET ラベルを付与したツイート数の割合と定義され、Accuracy を求める。この時、「X(芸能人名)の Y(名詞)」および「Y(名詞)」についてそれぞれ完全一致で評価を行い、Accuracy を求めた。「X(芸能人名)の Y(名詞)」のみならず、「Y(名詞)」を完全一致で評価することによって、本手法で有効性があるか検証を行った。

一方、REVIEW ラベルは「感想関係あり」と人手で判定され、評価データ中に含まれるツイートに対して部分一致で Precision, Recall を求める。具体的な TP, TN, FP, FN の定義は下記の通りである。TP が参照ラベルで「感想関係あり」であり、予測ラベルで参照ラベルと同一箇所に 1 形態素以上 REVIEW ラベルを付与したツイート、TN が参照ラベルで「感

想関係なし」であり、予測ラベルで REVIEW ラベルを付与しなかったツイート、FP が参照ラベルで「感想関係なし」であり、予測ラベルで REVIEW ラベルを付与したツイート、FN が参照ラベルで「感想関係あり」であり、予測ラベルで参照ラベルと同一箇所に REVIEW ラベルを付与しなかったツイートと定義される。これを踏まえ、REVIEW ラベルの Precision, Recall は下記の通り定義される。

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

これらの実験条件を基に、ツイート単位で評価を行った。

4.2 評価結果

TARGET ラベルにおいて「X(芸能人名)の Y(名詞)」に対して完全一致で評価を行った結果は Accuracy=0.78、「Y(名詞)」に対して完全一致で評価を行った結果は Accuracy=0.78 となった。REVIEW ラベルにおいて、部分一致で評価を行った結果は Precision=0.80, Recall=0.67 となった。

TARGET ラベルにおいて誤って予測した結果を分析すると以下の通りになる。「X(芸能人名)の Y(名詞)」の完全一致の評価と「Y(名詞)」は正例および負例が一致し、「X(芸能人名)の Y(名詞)」中の「Y(名詞)」のみに予測ラベルで正しく TARGET ラベルを付与することはないことが示された。このことから、今回用いたトークン分類モデルは、芸能人の評価対象名「X(芸能人名)の Y(名詞)」を判別し、TARGET ラベルを付与していることが示唆された。また、誤って予測した事例の傾向として、ツイート中の参照ラベルで TARGET ラベルを付与した「[LAST_NAME][FIRST_NAME] の Y(名詞)」以外に「[LAST_NAME][FIRST_NAME] の A(名詞)」¹⁴ が含まれており、それに対して TARGET ラベルを誤って付与したり、「Y(名詞)」が 2 形態素以上に渡って表記されるもので、その内の一部しか誤って TARGET ラベルを付与しないという傾向が見られた。

一方、REVIEW ラベルにおいて誤って予測した結果を分析すると以下の通りになる。FP では、①「感想関係あり」として人手で判定された頻度が多い形容詞・形容動詞(「良い」、「凄い」、「かわいい」など)を含むツイート、②「X(芸能人名)の Y(名詞)」が出現した直後に形容詞・形容動詞が出現するツイート、③動詞や名詞で感想を述べているツイートに対して誤って REVIEW ラベルを付与する傾向が見られた。一方、FN では、①「感想関係あり」として人手で判定された頻度が少ない形容詞・形容動詞(「几帳面だ」、「楽しい」など)を含むツイート、②「X(芸能人名)の Y(名詞)」とは離れた場所に感想を述べている形容詞・形容動詞が出現するツイートに対して誤って

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

11: tokenization.py

12: transformers-2.2.1

13: 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。

14: 誤って予測ラベルで TARGET ラベルを付与した [LAST_NAME][FIRST_NAME] は、必ずしも参照ラベルで TARGET ラベルを付与した「[LAST_NAME][FIRST_NAME] の Y(名詞)」の姓名と同一ではない。また、A=Y の場合もあれば、A≠Y の場合もある。

REVIEW ラベルを付与しない傾向が見られた。FP, FN ともに②に関しては、「X(芸能人名)のY(名詞)」が出現した直後に感想関係を表す形容詞・形容動詞が出現するツイートが、「感想関係あり」として人手で判定されたツイート中で多い傾向を反映したものだと考えられる。

これらの結果から、芸能人の評価対象名と形容詞・形容動詞間で感想関係の有無を同定するタスクにおいて、本提案手法を用いて BERT を用いたトークン分類モデルで評価を行うことは有効であることが示された。

5 関連研究

本研究に関連して、ある観点に基づき、テキスト群に対して評価対象の付与をし、データセットの構築を行った関連研究としては、以下のものが挙げられる。

まず、日本語のデータを対象とした研究では、公共交通機関に関するツイート群に対して評価情報(事物とそれに対する体験に基づく意見や感想等)を含むか否かを判定した後に、該当するツイートに対して評価対象(事物)と評価表現(事物に対する意見や感想等)を付与してデータセットを構築した研究[6]が挙げられる。また、宿泊施設のレビュー群に対して評価対象のカテゴリの付与(例えば、「大浴場のお湯」という評価対象を含むレビューの場合は「風呂」というカテゴリを付与する)や評価対象を含む文章に対して感情の極性分類を行い、データセットを構築した研究[11]が挙げられる。

一方、英語では、Twitter をデータセットの対象として実験を行った研究は以下のものが挙げられる。まず、評価対象を含む文章に対して感情の極性を判定するために Adaptive Recursive Neural Network というニューラルネットワークモデルを提案し、その精度を検証する過程で、有名人、製品、企業等のキーワードを含むツイート群に対して感情の極性分類を行った後に、評価対象を付与してデータセットの構築を行った研究[7]が挙げられる。また、糖尿病に関するツイート群に対して、評価対象のカテゴリ(糖尿病の症状、治療薬、合併症など)を付与した後、評価対象のカテゴリの同定および評価対象前後の単語を対象とした n-gram 法で感情の極性分類を行った研究[1]が挙げられる。

Twitter 以外のデータセットを対象として実験を行った研究は以下のものが挙げられる。SemEval 2016 task5 [8] のノート PC とレストランのレビュー群から、ユーザーから与えられた質問に対してレビュー中の単語を抽出して応答するデータセットを作成し、その評価の一環で BERT を用いて評価対象の抽出および評価対象に対する感情の極性分類を行った研究[4]が挙げられる。また、Citysearch New York データセット[3]のレストランのレビュー群に対して、1文中に感情極性がそれぞれ異なる複数の評価対象が含まれるデータセットを構築し、評価を行った研究[9]が挙げられる。

また、ドラマや出演俳優・登場人物におけるレビューマイニングの関連研究では、ツイート内で述べられているテレビドラマ出演俳優や登場人物を対象に感想関係の有無を判定した研

究[5]や、特定の芸能人と形容詞・形容動詞との間で BERT を用いた読解モデル、トークン分類モデル、分類モデルで感想関係の有無の判定を行った研究[10]が挙げられる。

一方、本論文では、「X(芸能人名)のY(名詞)」のみに評価対象を絞り、形容詞・形容動詞間で BERT を用いたトークン分類モデルで芸能人名、および名詞が未知の場合に有効性があるか訓練・評価を行った点が上記の研究と異なる点である。

6 結論

本論文では、芸能人の評価対象である「X(芸能人名)のY(名詞)」を含むツイートを収集し、感想を述べているか否かを判定したデータセットを作成し、BERT を用いたトークン分類モデルで訓練・評価を行った。その結果、芸能人名・名詞が未知であったとしても、一定の有効性があることが示された。今後は、「X(芸能人名)のY(名詞)」に対して広範的にツイートを収集・分析することや、他のモデルの枠組みを用いて感想マイニングを行うタスクに対して評価を行うことが課題である。

文献

- [1] S. Z. María del Pilar, M. M. José, L. O. Katty, L. A. Harry, R. G. Miguel Ángel, and V. G. Rafael. Sentiment analysis on tweets about diabetes: An aspect-level approach. *Computational and Mathematical Methods in Medicine*, p. 5140631, 2017.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional Transformers for language understanding. In *In Proc. NAACL-HLT*, pp. 4171–4186, 2019.
- [3] G. Ganyu, N. Elhadad, and A. Marian. Beyond the stars: Improving rating predictions using review text content. In *WebDB*, 2009.
- [4] X. Hu, L. Bing, S. Lei, and Philip S. Yu. Bert post-training for review reading comprehension and aspect-based sentiment analysis. In *Proc. ACL*, 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, pp. 344–347, 2018.
- [7] D. Li, W. Furu, T. Chuanqi, T. Duyu, Z. Ming, and X. Ke. Adaptive recursive neural network for target-dependent Twitter sentiment classification. In *Proc. ACL*, pp. 49–54, 2014.
- [8] P. Maria, G. Dimitris, P. Haris, A. Ion, M. Suresh, A. S. Mohammad, A. A. Mahmoud, Z. Yanyan, Q. Bing, D. C. Orphée, H. Véronique, A. Marianna, T. Xavier, L. Natalia, K. Evgeniy, B. Nuria, J. Z. Salud María, and E. Gülşen. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proc. 10th SemEval*, pp. 19–30, 2016.
- [9] J. Qingnan, C. Lei, X. Ruifeng, A. Xiang, and Y. Min. A challenge dataset and effective models for aspect-based sentiment analysis. In *Proc. EMNLP-IJCNLP*, pp. 6280–6285, 2019.
- [10] 野崎雄太, 菅原琴絵, 宇津呂武仁. ツイートにおける芸能人に対する感想のマイニング. 第 14 回 DEIM フォーラム論文集, 2022.
- [11] 中山祐輝, 村上浩司, Ikuko Hardaway. アスペクトベース意見分析における日本語評価コーパスの構築. 言語処理学会第 27 回年次大会, pp. 375–380, 2021.