

読み手の印象推定に基づく ツイートのフィルタリングに関する研究

黒木 クルミ[‡] 牛尼 剛聡[‡]

[†]九州大学芸術工学部 〒815-8540 福岡県福岡市南区塩原 4-9-1

[‡]九州大学大学院芸術工学研究院 〒815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: [†]1DS11187Y@s.kyushu-u.ac.jp, [‡]ushiana@design.kyushu-u.ac.jp

あらまし 現在, Twitter は広く普及し, 様々な目的や用途で使われている. こうした中で, ユーザの目的に合致したツイートを検索・推薦するためには, ユーザの目的に合致したツイートをフィルタリングすることが必要となる. 本研究では, 「気晴らし」や「息抜き」という目的に着目し, ツイートに対する読み手の印象を推定し, ユーザが希望する印象が得られるツイートをフィルタリングする手法を提案する. 本手法では, ユーザの希望する印象の種類を推定するために, ユーザがお気に入りとしたツイートに対して, 他ユーザの反応を利用するクラスタリングを行い, それぞれの結果がユーザの希望する印象の種類を表していると考え. そして, それぞれのクラスタの特徴を利用して, 未知のツイートのフィルタリングを行い, ユーザが希望する印象を与えるツイートを推薦する.

キーワード Twitter, ユーザ反応, 印象推定, フィルタリング, クラスタリング

1. はじめに

近年, ソーシャル・ネットワーキングサービス (SNS) は広く一般的に普及した. SNS には様々な種類が存在する. 代表的な SNS の一つに Twitter がある. Twitter は 140 文字以内のメッセージを投稿するサービスである. ユーザは, 友人や興味がある情報を発信する人をフォローすると, フォロー先のユーザの投稿がリアルタイムに配信される. Twitter では, 多くのメッセージを自由に閲覧でき, 自由にフォローできるユーザも多い. こうした中で, Twitter は, 「情報発信」, 「情報収集」, 「コミュニケーション」, 「気晴らし」, 「息抜き」など, 様々な目的に使われている. そのため, ユーザの目的に合致したツイートを検索・推薦することは重要である. この目的を実現するためには, ユーザの目的に合致したツイートをフィルタリングすることが必要となる.

ユーザが特定のトピックに関する情報を取得したい場合には, そのトピックに関する単語をキーワードとした検索により実現できることが多い. 一方, 「気晴らし」や「息抜き」を目的とする場合には, トピックに関するキーワードを利用した検索では適切な結果が得られない場合がある. これは, 「気晴らし」や「息抜き」は, ツイート自体を対象としているのではなく, ツイートが読者に与える印象に注目しているからである. ツイッターと印象の関係を図 1 に示す.

ユーザが「気晴らし」や「息抜き」の目的のために Twitter を使う場面においては, ユーザが希望する印象が得られるツイートを得られることができれば, ユーザにとっての満足度が高くなると期待できる. そこで,

本研究では, ユーザが「息抜き」を目的のために Twitter を使う場面において, ユーザが希望する印象が得られるツイートを推薦することを目的とする. ここで言う印象とは, ツイートを読んだ時に読み手の内面に生じる情動のことを指す.

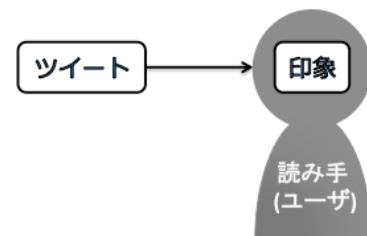


図 1 ツイートとその印象

ユーザに与える印象に基づいてフィルタリング行うためには, 未知のツイートに対する印象推定が必要となる. 本研究では, 未知のツイートの印象を推定するために, ツイートの本文ではなく, ツイートに対する他のユーザの反応に注目する.

ツイートと, それに対する反応ツイートの例を図 2 に示す. 図 2 に示した例では, 本文における「今日」や「猫」といった単語の特徴では読み手との反応に関連性がなく, 本文だけでは印象を推定することは難しい. しかし, 対象ツイートに対する反応ツイートには, 「可愛い」, 「癒される」などの読み手の印象を表す単語が含まれている場合がある. このように, 反応ツイートはユーザの反応として様々なものが考えられるが, ツイートに対する印象に関する記述が含まれていることが多いと考えられる. そこで, 本研究では反応ツイートを利用してツイートの印象を推定する.

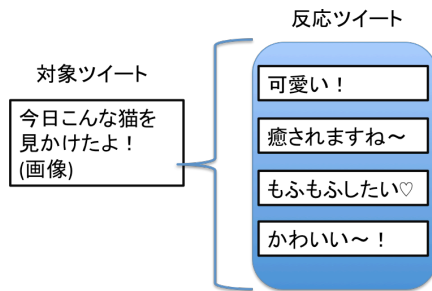


図2 反応ツイートの例

2. 関連研究

林田ら [1]は、ユーザの目的や価値に基づいてツイートを自動的に分類する手法を開発することを目的とする研究であり、ユーザにとっての価値や目的に応じたツイートのカテゴリの代表例として、「ネタツイート」を取り上げている。ネタツイートに対する反応が持つ特徴を利用して、複数のツイートの中からネタツイートを分類している研究である。この研究では、反応ツイートに対し 2-gram で分解し、その出現頻度を足し合わせたものを反応の特徴ベクトルとしている。ネタツイートの判定においては、機械学習であるサポートベクターマシン (SVM) を用いている。

堀宮ら[3]の研究では、人間とコンピュータのよりインタラクティブなコミュニケーションを実現することを目的とした研究であり、人間の他者に対する推測能力に着目している。ユーザの発言に対する反応であるリプライを利用して感情推定を行う手法を提案している。2-gram で素性に分解したのち、tf-idf で重みを計算し、SVM で学習を行なっている。この研究では学習データのラベルを基本 6 感情（「幸福感」、「驚き」、

「恐れ」、「悲しみ」、「怒り」、「嫌悪」）に人手で分類し、それを正解データとしている。

[5]の研究での目的は、プレゼンテーション動画に対して聴衆が抱く印象を予測することであり、動画において話者が話す文書と音声を利用し、プレゼンテーション動画に対し聴衆が抱く印象の予測を行う手法を提案するものである。文書分類の特徴量として Bag-of-Words (BoW), Latent Semantic Indexing (LSI), Latent Dirichlet Allocation(LDA), word2vec, word2vec に Spherical K-means (spk-means)を適用した手法、そこに更に Approximated locality-constrained linear coding (LLC)を組み合わせた手法を提案し、計六つの特徴量を用いた比較を行なっている。word2vec と LLC を組み合わせた手法が従来の文書分類に用いられてきた手法より有効であることを示した。

本研究では、これらをふまえ、ユーザの目的や価値に基づいたツイートをユーザに提供することを目指し

た。

3. 提案手法

3.1. システムの概要

本手法では、ユーザが要求する印象の種類を推定するために、ユーザが過去に「お気に入り」としたツイートを利用する。ユーザが「お気に入り」としたツイートに対して、他ユーザの反応に基づいてクラスタリングを行うと、それぞれのクラスがユーザの希望する印象の種類（カテゴリ）を表していると考えられる。

ユーザはクラスを指定することで、ユーザが要求するツイートの種類を指定し、それぞれのクラスの特徴を利用して、未知のツイートのフィルタリングを行うことで、ユーザが希望する印象を与えるツイートを推薦する。図3にシステムの全体像を示す。

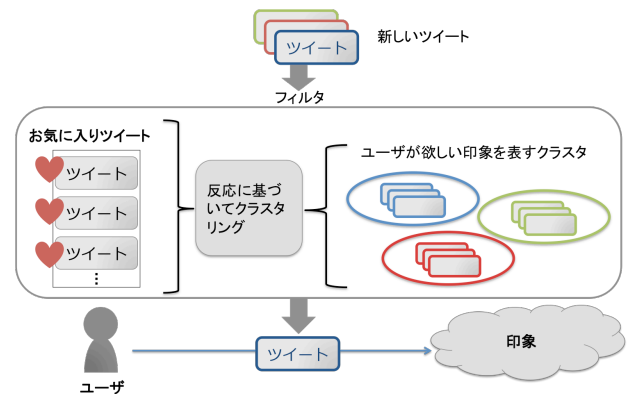


図3 システムの全体像

3.2. 反応ツイートの取得

本研究では、対象とするツイートの内容をふまえて投稿されるツイートを「反応ツイート」と呼ぶ。Twitterでは、反応ツイートを投稿するためにいくつかの方法がある。具体的には、リプライ、引用ツイート、リツイート直後のツイートなどである。本研究では、対象とするツイートが明確であり、ユーザの反応が必ず記載されているリプライを対象とする。

本研究では、ツイートの取得は Twitter API を用いる。Twitter APIには特定のツイートに対するリプライを直接取得するメソッドが提供されていない。そこで、以下の手順で反応ツイートを取得する。図4に、本研究でのリプライ取得の概要を示す。まず、Twitter APIでツイートの検索を行う際に、to 演算子を指定することで、特定のユーザに発信されたツイートのみを検索できる。これを用いて、対象ツイートを発信したユーザに対する反応ツイートを取得する。次に、ツイートがリプライである場合には、ツイートのステータス内の in_reply_to_status_id に反応先のツイートの ID が登録

されていることを利用して、対象ツイートのツイート ID と一致するかどうかで判断する。

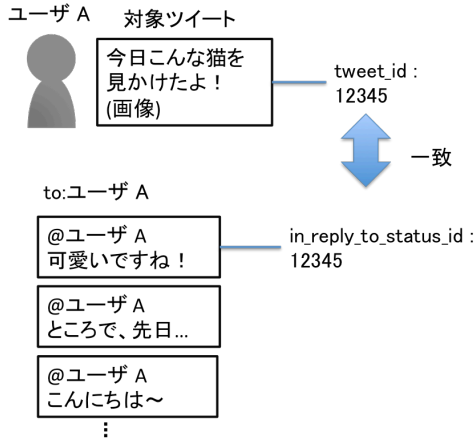


図 4 リプライの取得

Twitter API には 15 分あたりの制限回数が定められている API レートリミットがある。そのため、制限により取得に失敗した際にもう一度同じことを繰り返してくれる python の `retry` というパッケージ [12] を用いる。パラメータは `tries=4`, `delay=900` とする。これは失敗した時に 900 秒待ち、4 回目の失敗まで繰り返す。

Twitter API の仕様により、一回に取得できる件数が最大 200 件程度である。リプライとなるツイートを探すために、複数回 API にアクセスする必要がある場合がある。この処理において、API にアクセスする回数をできるだけ少なくするために、以下のように行った。ツイートを検索する際のパラメータとして、指定したツイート ID より過去のツイートのみを取得する `max_id` と、指定したツイート ID より未来のツイートのみを取得する `since_id` がある。また、ツイート ID は未来に投稿されたものほど数が大きくなるように設定されている。また、検索結果は未来のものから取得される。そこで、`since_id` に対象ツイートのツイート ID を指定し、2 回目以降の検索ではそれに加え `max_id` に前回取得した反応ツイートのツイート ID から 1 を引いたものを指定する。これを繰り返すことによってより効率的に反応ツイートを取得する。

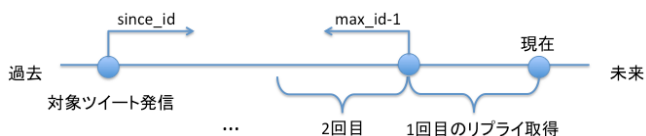


図 5 効率的なリプライ取得

さらに、ノイズを除去するために、以下の条件に当

てはまるものを分析の対象外として除去した。

- お気に入り数が 10 以下
- 取得できたリプライ数が 10 以下

お気に入り数が 10 以下のものは、炎上などのようなツイートと、個人的なやり取りによるリプライが続くツイートを除去するためである。リプライ数が 10 以下のものは、反応による特徴量の分析においてデータ量が不十分であると考えられるため、除外した。

3.3. 反応ツイートからの特徴量抽出

反応ツイートから、ツイートのフィルタリングに利用する特徴量を取得することを考える。

ツイート t に対するリプライ集合 $R(t)$ を以下のように定義する。

$$R(t) = \{r_{t,1}, r_{t,2}, \dots, r_{t,|R(t)|}\}$$

ここで、 $r_{t,i}$ は、ツイート t に対する i 番目のリプライを表す。 $R(t)$ に含まれるリプライを全て連結した文章をツイート t に対するリプライ文書と呼ぶ。 t のリプライ文書を rd_t と表記する。

本研究では、ツイート t のリプライ文書の特徴ベクトルを利用して、ツイートのクラスタリングとフィルタリングを行う。

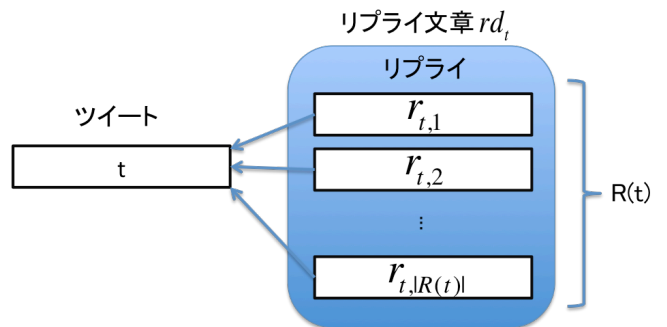


図 6 変数の定義

反応ツイートの特徴量を取るために、形態素を利用する。形態素とは、図 7 の例のように品詞に分解した単語のことである。今回は形容詞と形容動詞の原型を取得したため、図 7 の例文においては「いい」が特徴量として抽出される。

原文	いいお天気ですね			
	形容詞	名詞	助動詞	助詞
形態素	いい	お天気	です	ね

図 7 形態素の例

このように単語を分割してできた特徴量をパターン文字列 p とする。tf-idf を算出し、正規化したものを特

微量とする．tf-idf は単語の出現頻度を考慮した重み付けの手法であり，以下のように定義される．

$$w_{p,t} = tf_{p,rd_t} \times idf_p \quad (3.2)$$

$$idf_p = \log \frac{N}{df_p} + 1 \quad (3.3)$$

$$idf_p = \log \frac{N}{df_p} + 1$$

ここで， $tf_{p,rd}$ は，文書 rd_t 内に出現したパターン文字列 p の出現頻度を表す．一方， idf_p はパターン文字列 p が出現した文書の数， N はすべての文書数である．

上記の重みを利用して，文書 t の特徴ベクトル f_t を以下の式によって定義する．

$$f_t = \begin{pmatrix} w_{p1,t} \\ \vdots \\ w_{pn,t} \end{pmatrix} \quad (3.4)$$

特徴ベクトルは，コサイン正規化を用いて正規化する．コサイン正規化とは，文書 t の特徴ベクトル f_t のノルムを計算し，その文書の各タームの頻度をノルムで割ったものである．ノルムとはベクトルの大きさを表す数値で，次の式で計算される

$$\|f_t\| = \sqrt{\sum_{p \in P} (w_{p,rd_t})^2} \quad (3.5)$$

特徴ベクトルに対して，ユークリッド距離を用いて文章同士の距離行列を生成する．距離行列はベクトル同士の非類似度表す行列である．ユークリッド距離とは一般的な 2 点 x, y 間の距離であり，以下の式で定義される．

$$d_{x,y} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.6)$$

3.4. お気に入りツイートのクラスタリング

上記で得られた反応ツイートに基づいた特徴ベクトルを用いて，ユーザが「お気に入り」に登録したツイートを，その反応に基づいてクラスタリングを行う．これは，ユーザが気晴らしのためにツイートを選別する際には，ユーザがお気に入りに入れたツイートと類似したツイートを期待することが多いと考え，ユーザは自分のお気に入りに入れたツイートのクラスタを指定することで，自分の要求するツイートを簡単に指定できると考えたからである．

クラスタリングには様々な手法が提案されている

が，本研究では，DBSCAN(Density-Based Spatial Clustering)を利用する．DBSCAN は，密度に基づいたクラスタリングの代表的な手法である．DBSCAN は，ノイズに強く，予めクラスタの個数を指定せずにクラスタリングが可能という特徴がある．DBSCAN では，対象とするオブジェクト x から半径 ϵ 内に， $MinPts$ 個以上のオブジェクト y を含むものを直接密度到達可能とし，その範囲内にあるオブジェクトが同じクラスタに所属するとみなす． $N_\epsilon(x)$ を，ある点 x から距離 ϵ 内にある点集合とするとき，直接到達可能な条件を表す式と概念図を図 8 に示す．

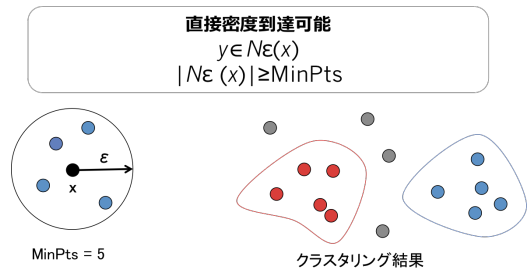


図 8 DBSCAN の説明

3.5. 未知ツイートのフィルタリング

本研究では，お気に入りツイートのクラスタを正解データとして機械学習を行い，未知のツイートに対するクラスタリングを行う．つまり，対象とするクラスタに含まれるものを正解データ，それ以外のクラスタのものを全て不正解データとし学習器を構築する．正解データは，ユーザがお気に入りにしたツイートの中でリプライ 10 以上のリプライが取れたツイートをを用いた．クラスタ毎に構築される学習器を利用して，未知データの分類を行う．

すでに抽出されたクラスタを正解データとするために，本研究では教師あり学習を利用する．教師あり学習は，正誤が既に分かっている訓練データから学習器を作成し，その学習器を用いて未知のデータを判別する．概要を図 9 に示す．本研究では，代表的な教師あり学習アルゴリズムであるニューラルネットワーク (NN) を用いて実験を行い，それぞれの性能の比較を行う．

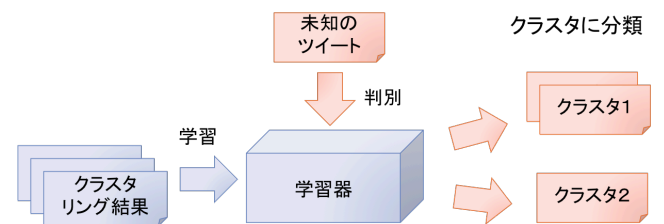


図 9 機械学習

4. 評価

4.1. 実験手法

提案手法のクラスタリングとフィルタリングの有効性を検証するために、取得したツイートに対し、3節で説明した手順に沿って実験を行った。

本実験では、著者自身がお気に入りに登録したツイートを対象にした。ここで、対象となるツイートは117件であり、未知の15件のツイートを分類することを考える。

本実験ではRを利用し、形態素解析はRMeCabパッケージ[13]、DBSCANはfpcパッケージ[14]を用いた。

本実験では、まず、DBSCANにおいて密度を決定するためのパラメータである最小個数 $MinPts$ と、半径 ϵ を決定するために実験を行なった。次に、お気に入りのクラスタリングの結果に基づいて未知のツイートを分類する実験を行った。

4.2. 実験結果

4.2.1 クラスタリングのためのパラメータ

DBSCANによるクラスタリングのためのパラメータとして $MinPts$ を3から6まで変化させ、 ϵ を1.4から2.0まで変化させながら、DBSCANの結果であるクラスタ数の変化を観察した結果を図10に示す。

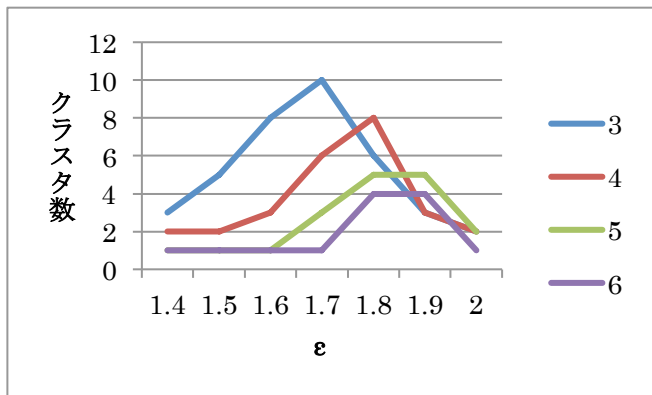


図10 パラメータによるクラスタ数の変化

上記の結果のうち、クラスタ数が8以上である以下の3つの組み合わせについて具体的に考察を行う。

- ① $MinPts = 3, \epsilon = 1.6$
- ② $MinPts = 3, \epsilon = 1.7$
- ③ $MinPts = 4, \epsilon = 1.8$

① $MinPts = 3, \epsilon = 1.6$ の場合のDBSCANの結果を表1に示す。0は外れ値であり、クラスタの数とそれに含まれる件数を示した。

表1 ①の場合におけるクラスタリング結果

クラスタ	0	1	2	3	4	5	6	7	8
件数	77	4	15	4	3	3	4	4	3

以下に、それぞれのクラスタについて幾つかの例を、それに含まれるツイートの特徴について著者の主観的な印象を述べる。クラスタ1は、今日のご飯を紹介したツイートや、柑橘フェアを紹介しているツイートなどがあつた。このクラスタを名付けるとすれば「美味しそう系」と言える。クラスタ2は、芸能人がぬいぐるみを抱いているツイートや、駅に迷い込んだ猫が車掌に捕まえられているツイートなど、「かわいい」、もしくは「かわいくて笑える」ツイートが多かった。クラスタ3は「懐かしい系」、クラスタ5は、「感嘆系」、クラスタ6は、「笑える系」と思えるツイートが多かった。

② $MinPts = 3, \epsilon = 1.7$ の場合の結果を表2に示す。

表2 ②の場合におけるクラスタリング結果

クラスタ	0	1	2	3	4	5	6	7	8	9	10
件数	57	4	3	19	4	11	4	3	5	3	4

クラスタ9は、歌うと猫が囃んでくるという旨の漫画のツイートや、流行しているダンスを面白可笑しくアレンジした動画など、「笑える」ツイートが多かった。クラスタ1は、「美味しそう系」、クラスタ3は、「かわいい」もしくは「かわいくて笑える」、クラスタ5は「すごい系」、クラスタ6は、「懐かしい系」と思えるツイートが多かった。

③ $MinPts = 4, \epsilon = 1.8$ の場合において同様に評価した。

表3 ③の場合におけるクラスタリング結果

クラスタ	0	1	2	3	4	5	6	7	8
件数	40	8	5	21	22	4	7	4	6

クラスタ1は、鬱発症までの様子を图示したツイートや、共謀罪が成立する事案について解説したツイートなどがあり、「問題提起系」と言える。クラスタ6は、リーゼントの正しい位置について解説したツイートや、自転車の飛び出し問題など、「議論したくなる系」と言えるものが多かった。クラスタ2は「美味しそう系」、クラスタ3は「かわいい系」もしくは「可愛くて笑える系」、クラスタ4は、「すごい系」、クラスタ5は、「懐かしい系」と思えるツイートが多かった。

評価の方法は、特定のクラスタに含まれているあるツイートが、そのクラスタに含まれている他の複数のツイートと比較してそのクラスタに含まれることが妥当であるかどうかで判断した。また、再現率においては、クラスタの印象と思われるものが、そのツイートに対して適切な判定であると思われるものを正解とした。

それぞれの適合率、再現率、F 値を表 4 に示す。

表 4 クラスタリングに対する評価値

① MinPts=3, $\epsilon=1.6$

クラスタ	1	2	3	4	5	6	7	8	合計
適合率	1	1	1	0	1	0.75	0	0	0.725
再現率	1	0.455	1	0	0.188	0.176	0	0	0.392

② MinPts=3, $\epsilon=1.7$

クラスタ	1	2	3	4	5	6	7	8	9	10	合計
適合率	1	0	0.947	0	0.455	1	0	0.4	1	0	0.6
再現率	1	0	0.633	0	0.313	1	0	0.5	0.176	0	0.493

③ MinPts=4, $\epsilon=1.8$

クラスタ	1	2	3	4	5	6	7	8	計
適合率	0.625	0.8	1	0.409	1	0.571	1	0	0.662
再現率	0.5	1	0.692	0.5	1	1	0.25	0	0.574

また、それぞれの場合における F 値は以下のようになった。

表 5 F 値

	①	②	③
F 値	0.509	0.541	0.615

これを見ると、③の場合において一番クラスタリングが適切に行えていると考えられる。よって、③のクラスタリングを用いてフィルタリングを行う。

ニューラルネットワーク (NN) による機械学習の結果は以下のようになった。

表 7 機械学習によるクラスタ判定

ツイート	クラスタ							
	1	2	3	4	5	6	7	8
1	0	0	1	0	0	0	0	0
2	0	0	0	1	0	0	0	0
3	0	0	0	1	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0
8	0	0	0	0	0	1	0	0
9	0	0	1	0	0	0	0	0
10	0	0	0	0	0	0	0	0
11	0	0	1	0	0	0	0	0
12	0	0	0	1	0	0	0	0
13	0	0	0	0	0	0	0	0
14	0	0	0	1	0	0	0	0
15	0	0	0	0	0	0	0	0

フィルタリングの結果、ツイートはそれぞれ一つの

クラスタに分類されている。そのため、機械学習の正解率を求めた。

表 9 正誤判定

ツイート	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
正解	3	7	7	1	7	3	4	6	3	0	3	1	0	7	3
NN結果	3	4	4	0	0	3	4	6	3	0	3	4	0	4	0

表 9 は、ツイートが含まれると思われる正解クラスタと、機械学習により判定したクラスタを示している。赤で色付けした部分は、正解クラスタと、機械学習によるクラスタ判定のクラスタが一致している正しい判定である。正解率は、0.53 であった。

4.3. 考察

実験結果から、提案手法である印象によるフィルタリングの考察を述べる。

実験で比較した中で一番良かった DBSCAN のパラメータである $eps = 1.8$, $MinPts = 4$ の F 値が 0.615 であり、ニューラルネットワークによるフィルタリングの正解率は 0.53 と、正誤が半々程度で、評価はあまり良いとは言えないものであるが、印象を推定する方法としては可能性がある。

問題点として、今回は単一のラベルを用いたが、必ずしも単一のラベルではなく複数ラベルを考えたほうが適切な場合がある。例えば、図 12 に示した例では、「かわいい」ツイートだと判定されたが、同時に「笑える」という印象をあたえる場合もある。また、それ以外にも「驚き」と「笑える」という印象を与える場合がある。



図 12 複数の印象を内包するツイート例[15]

また、特徴量に反応ツイート文書の形態素における形容詞の tf-idf を正規化したものを用いたが、word2vec など文章の特徴量を生成する他の手法や、単語以外の特徴を用いて結果を比較したいと考える。印象は反応ツイートの形容詞のみに現れるものではなく、例えば絵文字や顔文字、「!」「w」などの記号にも表現されていると考えられるためである。

今回、機械学習のパラメータにおいてはパッケージのデフォルトのものを採用したが、SVM における C (コスト) の値や、ナイーブベイズにおけるゼロ頻度問題を解消するためのスムージング処理を行うなどの調整を行う必要がある。

今回の実験では、著者のお気に入りしたものを対象にしたが、本研究の結果が一般性を持つのかを複数の被験者のお気に入りを利用して検証する必要がある。また、ユーザ自身が気に入りしたものの以外にも、リプライが一定数以上の一般的なツイートのものと比較し、どのような印象のクラスが生成されるのかを比較することが必要である。本実験で利用したデータセットのツイート件数は 100 件強程度であり、より精度を高めるために件数を増やすことも課題である。また、今回の実験は主観的な評価によるところが大きいので、複数の被験者実験により、客観性のある評価に基づいて有効性を検証する必要がある。

5. おわりに

本論文では、Twitter において「息抜き」を目的とするように、ユーザが特定の印象を与えるツイートを取得することを目的に、「お気に入り」に登録されたツイートに対してクラスタリングを行い、それをユーザが欲しい印象とみなすことによって、機械学習を用いてフィルタリングする方法を提案した。

ユーザが特定のトピックに関する情報を取得したい場合には、そのトピックに関する単語をキーワードとして検索して実現できることが多い。しかし、ツイートが読者に与える印象を求める際にトピックに関するキーワードを利用した検索では適切な結果が得られない場合がある。一方、本文に対する反応ツイートは、ツイートに対する印象に関する記述が含まれていることが多いと考えられる。そこで、本研究では反応ツイートを利用してツイートの印象を推定する。本手法では、反応ツイートを形態素解析し、形容詞の出現頻度に基づいて対象ツイートの特徴量を生成し、ノイズに強い DBSCAN を用いてクラスタリングを行なう。分類は、代表的な機械学習手法であるサポートベクターマシン、ニューラルネットワーク、ナイーブベイズを用いて行った。実験の結果、印象を推定する方法としては可能性があることがわかった。

今後、特徴量を絵文字や顔文字などの形態素以外のものを利用する予定であり、より高い精度のクラスタリングやフィルタリングが実現できる用に改善を行っていく。

参 考 文 献

- [1] 林田宗一郎, 牛尼剛聡, “ユーザの反応を利用したネタツイート自動分類手法”, DEIM Forum 2014, <http://db-event.jpn.org/deim2014/final/proceedings/B6-5.pdf>
- [2] 田中淳史, 田島敬史, “twitter のツイートに関する分類手法の提案”, DEIM Forum 2010, <http://db-event.jpn.org/deim2010/proceedings/files/A5-4.pdf>
- [3] 堀宮ありさ, 坂野 遼平, 佐藤 晴彦, 小山 聡, 栗原 正仁, 沼澤 政信, “Twitter における発話者へのリプライを用いたユーザ感情推定手法”, DEIM Forum 2012, <http://db-event.jpn.org/deim2012/proceedings/final-pdf/f2-1.pdf>
- [4] 福元伸也, 湊田孝康, “単語の共起関係に基づく機械学習による文書分類,” DEIM Forum 2015 <http://db-event.jpn.org/deim2015/paper/335.pdf>
- [5] 福島悠介, 山崎俊彦, 相澤清晴, “文書と音声解析に基づくプレゼンテーション動画の印象予測”, 2016, 電子情報通信学会論文, Vol.J99-D, No.8, pp.699-708
- [6] “後藤充裕, 木全英明, 中村浩司, “日本語機能表現を用いたユーザのツイート発信意図の推定”, 大会講演論文集 2015, page60
- [7] 塚田文哉, 鈴木徹也, “Twitter におけるツイート間の反応を考慮した話題分類法”, 第 76 回全国大会講演論文集 2014, page109-110
- [8] 山下澄枝, 川喜田佑介, “Twitter のための単語の出現頻度を用いたツイート有益度推定(「コミュニティ上での知識の共有・再利用」及び一般)”, 電子情報通信学会技術研究報告. AI, 人工知能と知識処理 110(301), 7-11
- [9] Tadahiko Kumamoto, Hitomi Wada, Tomoya Suzuki, “Visualizing Temporal Changes in Impressions from Tweets”, iiWAS2014, page116-125
- [10] M.Ester, H.-P.Kriegel, J.Sander, and X.Xu "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", KDD1996
- [11] Twitter Developer Documentation, <https://dev.twitter.com/docs>
- [12] retry 0.9.2 : Python Package Index, <https://pypi.python.org/pypi/retry/>
- [13] RMeCab の使い方, 石田基広 <http://rmecab.jp/wiki/index.php?plugin=attach&refer=RMeCab&openfile=manual.pdf>
- [14] Package ‘fpc’, <https://cran.r-project.org/web/packages/fpc/fpc.pdf>
- [15] https://twitter.com/kanon_yr/status/819911880214839296