

バーチャル YouTuber をもっと活気づけ
たい！
～「バズる」VTuber とは～

数理科学研究会 3年 中本幸佑
スペシャルサンクス なぽりたん

平成 30 年 11 月 4 日

目次

1	前提知識	1
2	研究内容	1
3	方針	1
4	Bag-of-words とは	2
4.1	形態素解析とは	2
4.2	順位付けの方法	2
4.3	TF-IDF	2
5	「人気となったきっかけ」の指標について	3
6	成果	3
6.1	システムを作る	3
6.2	VTuber のハッシュタグのツイートを取得する	4
6.3	結果	4
7	方針 2	4
7.1	YouTube からコメントの取得	4
8	結果	5
8.1	各 VTuber がバズった時期を求めるその 1	5
8.2	各 VTuber がバズった時期を求めるその 2	7
8.3	それぞれがバズったきっかけとなる言葉	10
9	今後の課題	13

研究背景

作者はバーチャル YouTuber が少なくとも大好きである。しかしこのコンテンツが長命であるかはわからない。そこでバーチャル YouTuber 界がもっと活気づくと良いと考えた。活気づくにはどうすれば良いか考えた結果多くの人の目に留まればバーチャル YouTuber は長生きするのではないかと考え、「人気のバーチャル YouTuber が増えれば様々な趣向のバーチャル YouTuber が増えるため、より多くの人の目に着くのではないか」という仮説に行き着いた。そのため、人気のバーチャル YouTuber が流行った理由を見つければ良いと考えた。

1 前提知識

この資料を手にとってくれた方はおそらく「バーチャル YouTuber」という単語に反応した方が多数だと思う。議論を進める関係で YouTube 自体の機能を知っている必要があるが、バーチャル YouTuber 自体の知識はなくても読める様に作ったつもりである。また、今後「バーチャル YouTuber」を「VTuber」と表記する。以下わからない人の為に簡単に説明する。

- バーチャル YouTuber：バーチャル YouTuber とは「外見がコンピューターグラフィックス (CG) やイラストのキャラクター」であるユーチューバーを指す用語である。例として挙げると「キズナアイ」や「電脳少女シロ」、「輝夜月」などが該当する。
- (YouTube) チャンネル：動画を挙げている人が動画をまとめておく場所である。このチャンネルを気に入った人がチャンネル登録（お気に入りのようなもの）をしていくとチャンネル登録者数が増える。登録者数が多いほどその YouTuber は人気であるといえる。^{*1}

2 研究内容

どのようにして VTuber が流行るかについて調べる。調査方法だが、人気の VTuber が何を発言してチャンネル登録者数が急激に増えたかを調べる。(登録者数が多いことを今回は「バズる」と呼ぶ。) その後どんな発言をすると人気になるかを集計する。ここで人気の VTuber とは@User Local 様の「バーチャル YouTuber ランキング」に書かれたチャンネル登録者数 1 万人以上の VTuber とする。^{*2}

3 方針

まず、先の節でも述べたが、「人気の VTuber」を「登録者数が 1 万人以上である VTuber」と定義する。この人気の VTuber 達が何が原因で 1 万人を超えたかを調査する。調査の仕方だが、今回は、何人かの VTuber を選び、その VTuber の関連したハッシュタグのツイート数が急激に伸び始めた日付を探す。次にその日付から 1 週間を目安にその VTuber の関連するハッシュタグで Twitter で検索をかけてそのあたり一帯のツイートと、その日付から前のツイートを抽出する。その中から大

^{*1} 明確にどこからが人気という定義はない

^{*2} 10/6 時点で 255 人の VTuber がチャンネル登録者数 1 万人を超えている。

量に発言された単語を集計し各 VTuber の流行ったきっかけにあたる単語がどういうものなのかを Bag-of-words モデルを用いて推定する。

4 Bag-of-words とは

Bag-of-words モデルとは文書に単語が含まれているかどうかのみを考え、単語の並び方などは考慮しないモデルのことである。今回はその中でも Term Frequency という文書中にある単語が含まれている回数をそのまま値として用いる方法で行う。得られた文書に対して形態素解析を行い、文書を形態素に分ける。その後、その単語から有用な情報だけを得たいので単語に対して順位付けを行う。その後、文書行列化してこの行列から特徴後を抽出するものである。

4.1 形態素解析とは

そもそも形態素とは、言語学の用語で、意味をもつ表現要素の最小単位である。ある言語においてそれ以上分解したら意味をなさなくなるまで分割して抽出された、音素のまとまりの 1 つ 1 つを指す。形態素解析は文書などを形態素に分類し、その品詞を特定する自然言語処理における基礎技術のことである。

4.2 順位付けの方法

単語に分類できたのでその単語から有用な情報だけに絞っていく。今回は「バズったときにおける特徴的な言葉」に関心があるため、名詞や動詞については重要度が高いが、助詞や接続詞などは重要度があまり高くない。従って、特徴語を抽出するために「名詞」と「動詞」などを残し、助詞や接続詞は除いて結果を出す。また、VTuber ごとにそもそもの特徴^{*3}は「バズる」じたいにはあまり関係ないと思われるためそれも省きたい。従って TF-IDF という手法を用いることで情報の順位付けを行う。

4.3 TF-IDF

TF-IDF をするにあたって単語に分割したリストを文書行列化する必要があるためそれを行う。

4.3.1 文書行列化

文書を単語に分割できるのでこれで単語のリストに変換できるようになったため、特徴後を抽出する。

4.3.2 TF-IDF について

TF-IDF とは「TF」と「IDF」という 2 つの値を掛け合わせた指標のことである。先ほど変換した文書行列を用いて以下のように計算したものである。

1. 「TF」は文書内における単語の出現頻度を表す。これは「ある文書中である単語が何回出現したか」で定義される。1 つの文書に多く出現する単語ほど重要度が高くなる。

^{*3} にじさんじ公式ライバーの一期生である「月ノ美兎」であれば、公式設定である「委員長」など

- 「IDF」は多数の文書に出現する単語ほど重要度が低くなるようなスコアである。「ある単語が含まれている文書数を全ての文書数で割ったものの逆数」で定義される。
- 「TF-IDF」は上記の「TF」と「IDF」を掛けあわせた指標である。

以上から、TF-IDF とは以下のものである。

$$\begin{aligned} \text{TF-IDF}(w, d) &= \text{TF}_{w,d} \cdot \text{IDF}_w \\ \text{TF}_{w,d} &= \frac{n_{w,d}}{\sum_k n_{k,d}} \\ \text{IDF}_w &= \log \frac{|D|}{|\{S : t_w \in S\}|} \end{aligned}$$

$n_{w,d}$ は文書 S_d における単語 t_w の出現回数, $\sum_k n_{k,d}$ は文書 S_d におけるすべての単語の出現回数の和, $|D|$ は総文書数, $|\{S : t_w \in S\}|$ は単語 t_w を含む文書数である。

5 「人気となったきっかけ」の指標について

人気となったきっかけの判断方法だが抽出した特徴語のうち人気になった時の特徴語と人気でない時の特徴語の差集合をとる。このとき、人気でないときにのみ現れた特徴語は消去する。こうすることによって、人気になった時の特徴語がわかるため流行った原因であると推定できる。今回この特徴語を「バズる」と定義する。

6 成果

成果についてだがシステムを作りその後 VTuber に関するデータを作ってどのようなになるか調べることにした。

6.1 システムを作る

まず、特徴語を出すシステムについてだが、まずシステムを作る実験に用いたのは友人の Twitter の過去のツイートである。一通りシステムを完成させるのが最初の目的であったため、まずは特徴語を出す対象として名詞だけで行った。そして形態素解析して名詞だけにしたのが以下の図 1 である。次にこの名詞に対して文書行列化し、TF-IDF を計算して 1 日毎の特徴語を抽出してみた。今回は

```
[ '富山', '東京' ],
[ 'iPad', 'Pro', '人気' ],
[ '中古', '6', '万', '本体', 'ペンシル', '日常', '生活', 'みんな' ],
[ 'goodnotes', '他', '人', 'goodnotes', 'ノート', '自分', 'ノート' ],
[ 'iPad', '奴ら', '同士', 'ルール', 'ノート', '更新', 'ノート', 'ふり', 'Twitter', 'わけ' ],
[ '誰', 'ファボ', 'ん', '言葉', 'せい', 'ファボ', 'こと' ],
[ 'amazon', 'プライム', 'ビデオ' ],
[ '過去', '記憶', '時', '数', '年', '後', '方', '現象', 'ん' ],
[ 'ディーン', '版' ],
```

図 1 形態素解析して名詞だけ取り出したものの一部 (5000 ツイート近くあったため省略した)

python のモジュールである scikit-learn を用いることで特徴語の抽出まで一度に行った。その結果以下の図 2 のようになった。

人種
レパートリー
バッテリー
姿勢
得意
自信
敗北
トーク
iPhoneX
むぎ

図 2 図 1 と tf-idf から求めた特徴語の一部

6.2 VTuber のハッシュタグのツイートを取得する

VTuber のハッシュタグのツイートを取得しようとしたが、Twitter の API によると過去 1 週間分しかとってこれないことがわかった。今欲しいのはバズった周辺とその前のツイートなのでこれは当てはまらない。色々試してみたが取得方法が見当たらないため、ここで詰まってしまった。

6.3 結果

該当の文書を集めること出来ず、実験できなかった。

7 方針 2

Twitter から該当文書を集めることはできなかったため、次に YouTube に投稿された動画についてのコメントを取得してその動画が投稿された 1 週間以内のコメントを集めてそのコメントから特徴語を出して「バズる」きっかけを調べることにした。

7.1 YouTube からコメントの取得

今回対象にした VTuber は以下の 8 人である。

- 物述有栖
- 月ノ美兎
- 樋口楓
- 家長むぎ
- 名取さな
- 笹木咲
- 静凜

それぞれのチャンネルのコメント欄から YouTube の API を用いてコメントの投稿日時とコメントを取得した。

8 結果

各 VTuber がバズった時期と「バズる」きっかけは以下のようになった。

8.1 各 VTuber がバズった時期を求めるその 1

それぞれのコメント欄のコメント数をグラフにプロットすることでコメント数が急激に増えたところが「バズる」瞬間だととらえることにする。以下は各 VTuber の動画投稿後 1 週間以内にコメント数を実際にプロットしたものである。

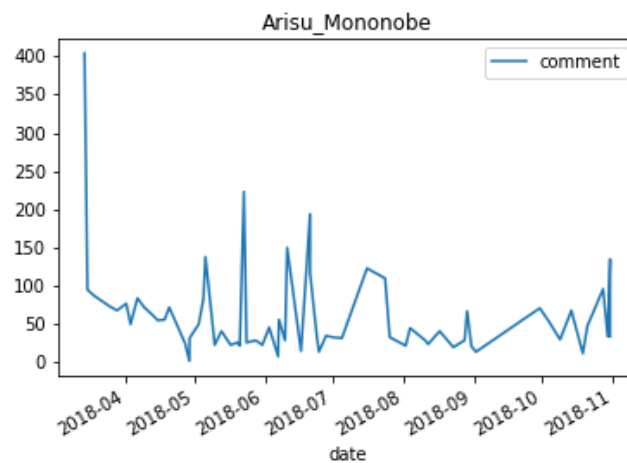


図3 「物述有栖」のコメント数

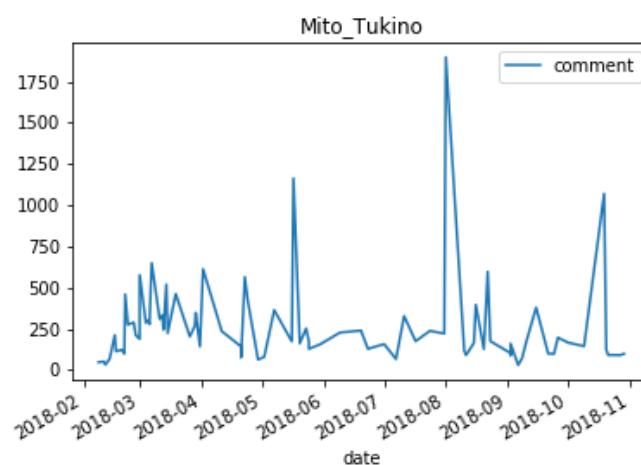


図4 「月ノ美兎」のコメント数

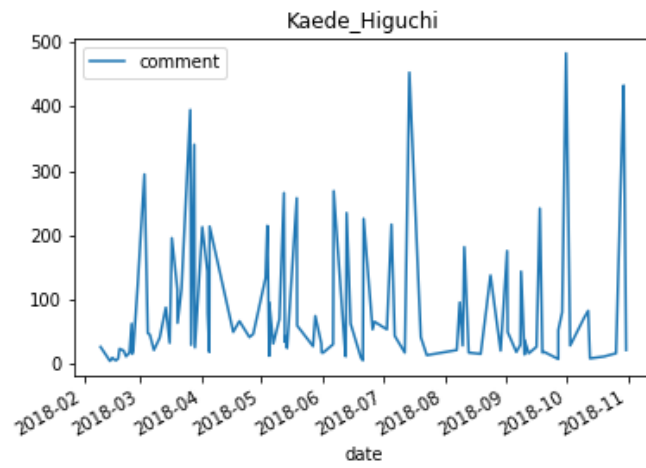


図5 「樋口楓」のコメント数

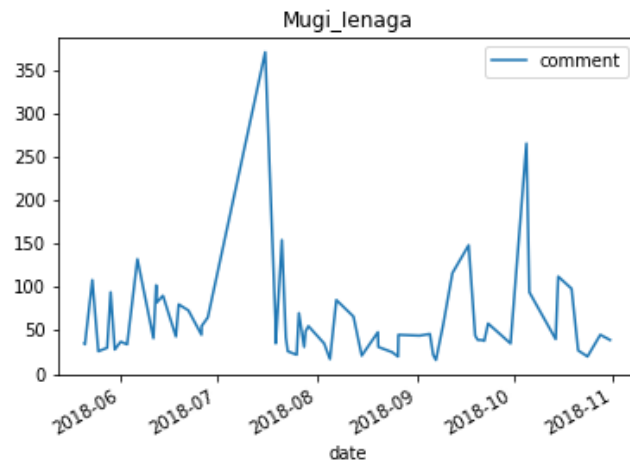


図6 「家長むぎ」のコメント数

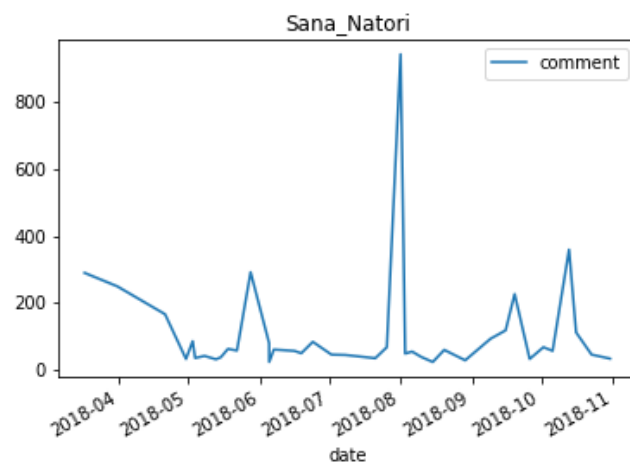


図7 「名取さな」のコメント数

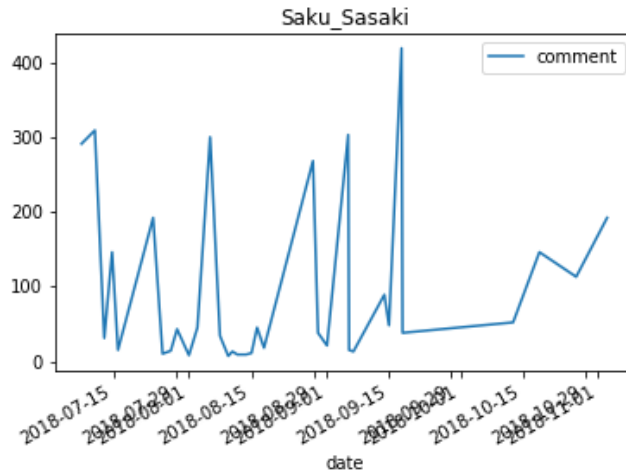


図8 「笹木咲」のコメント数

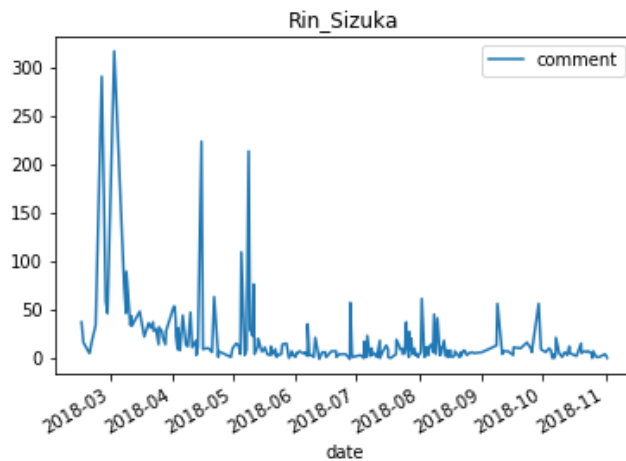


図9 「静凜のコメント数

8.1.1 考察

コメント数をグラフにして見てみたが、いまいちわからなかった。「家長むぎ」や「名取さな」はある程度この時期に「バズる」何かがあったのではないかと考察できるが、他のVTuberは怪しい。従って、この「バズる」瞬間について少し変化を加えた。

8.2 各VTuberがバズった時期を求めるその2

ただ単にコメント数を数えるのではなく、各動画の特徴語をまとめて、その特徴語が何語含まれているかをカウントすることにした。その結果以下ようになった。

8.2.1 考察

先ほどと同様にいまいちわからないのは「静凜」である。これは「動画」を挙げているか「生放送」を動画化してしているかでコメントの質が異なる(というのも生放送中にもコメントが出来るが、そ

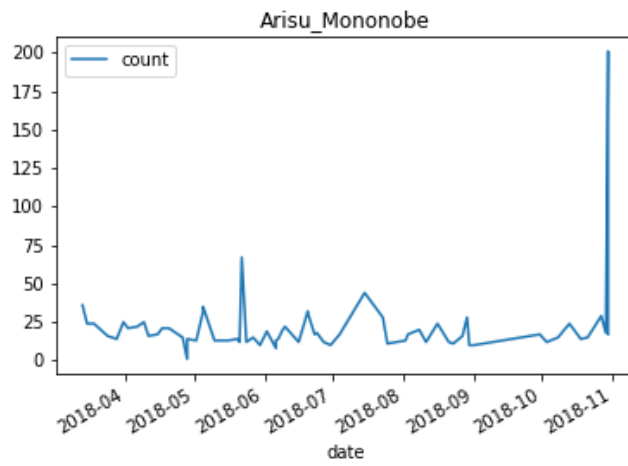


図 10 「物述有栖」の特徴語のカウント数

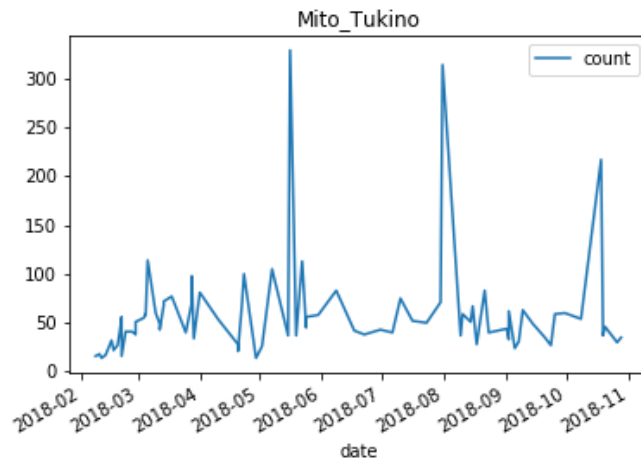


図 11 「月ノ美兎」の特徴語のカウント数

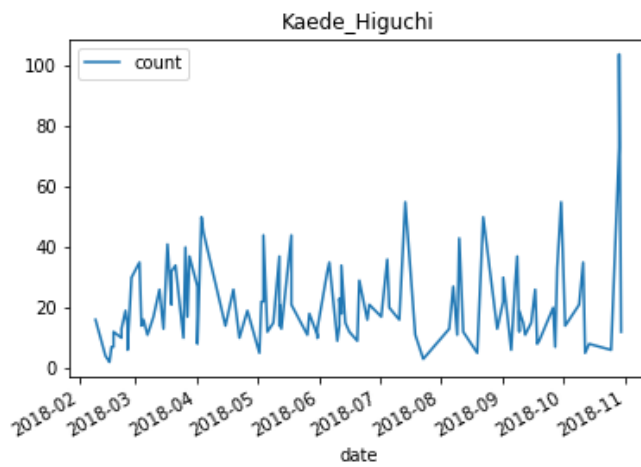


図 12 「樋口楓」の特徴語のカウント数

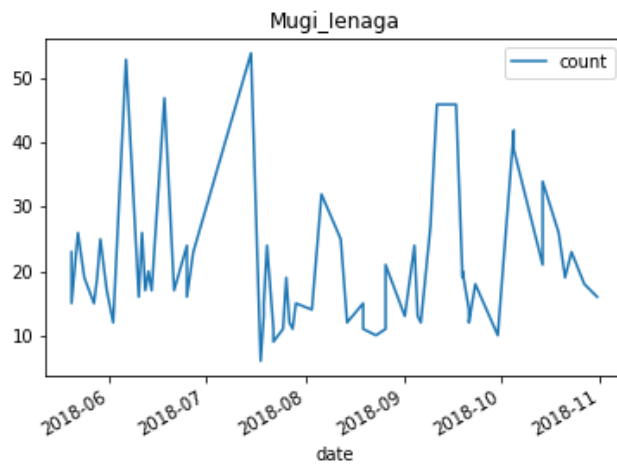


図 13 「家長むぎ」の特徴語のカウント数

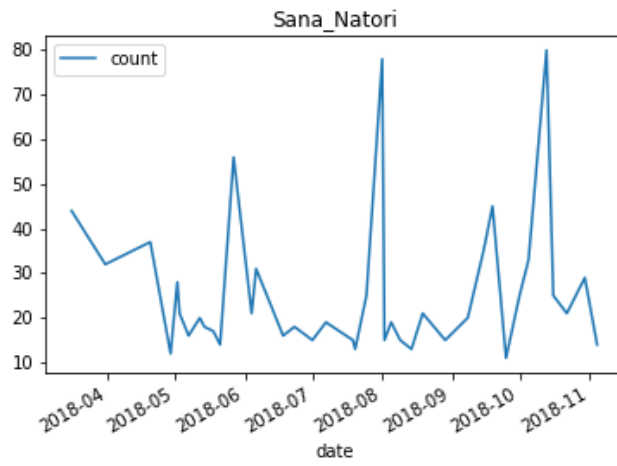


図 14 「名取さな」の特徴語のカウント数

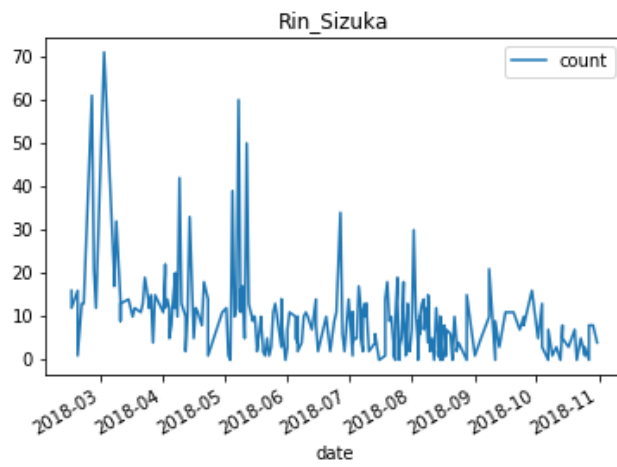


図 15 「静凛」の特徴語のカウント数

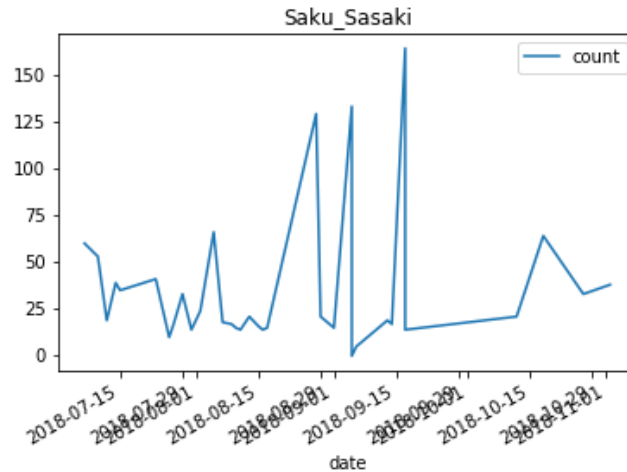


図16 「笹木咲」の特徴語のカウント数

のコメントは YouTube のコメント欄に残らないため) この方法ではあまり効果がない。だが、他の VTuber は「バズる」ポイントを

8.3 それぞれがバズったきっかけとなる言葉

上の結果より「静凜」とは「バズる」ポイントがわからないため今回は結果が出せなかった。従って残りの VTuber について差集合を求める。「バズる」ポイントだが一番最初に急にカウント数が増えたところとする。

8.3.1 結果

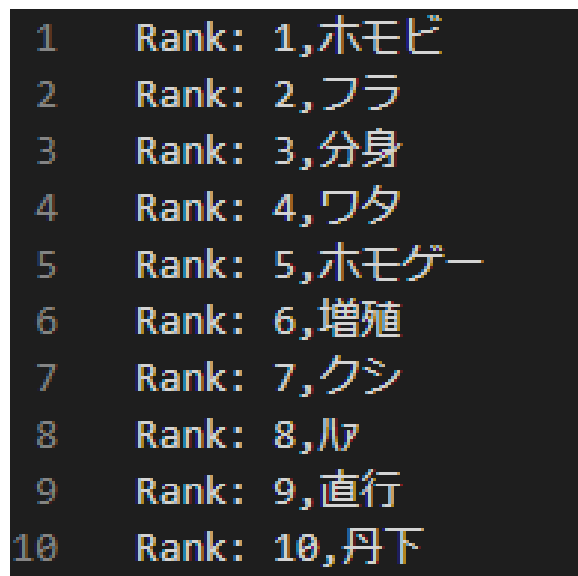
差集合を求めた結果、全て元の特徴語と完全一致だったため実は差集合を求める必要がないことがわかった。

物述有栖 「物述有栖」はグラフをみると5月上旬地点で「バズる」ポイントが来ているように見える。そしてこのときの特徴語は以下ようになった。

- 1 Rank: 1,ホース
- 2 Rank: 2,AI
- 3 Rank: 3,ゲキ
- 4 Rank: 4,ハッ
- 5 Rank: 5,ヌンッ
- 6 Rank: 6,絵心
- 7 Rank: 7,ダイジェスト
- 8 Rank: 8,フォント
- 9 Rank: 9,よほほほほろりりおひひい
- 10 Rank: 10,Vtube

図17 「物述有栖」の「バズる」ポイントでの特徴語

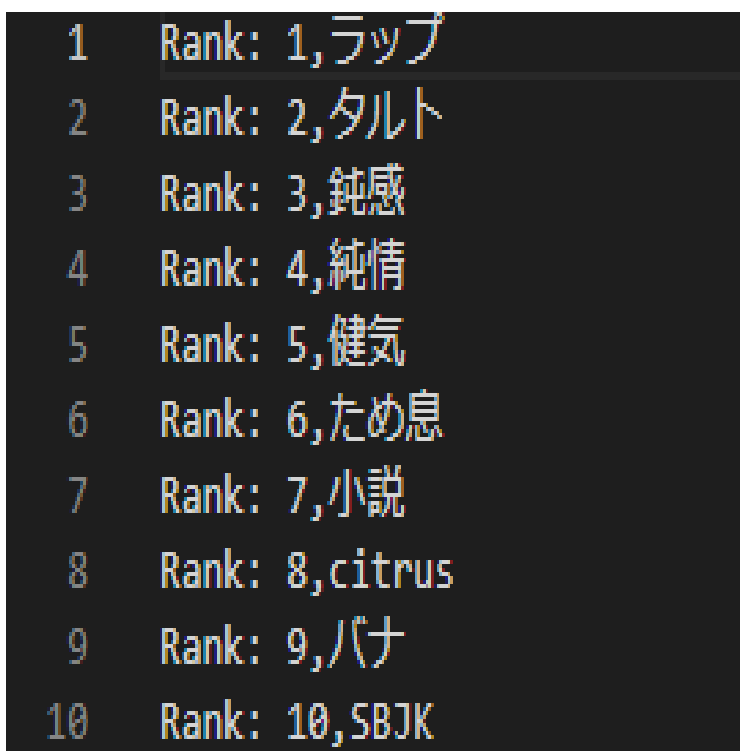
月ノ美兔 「月ノ美兔」はグラフをみると2月下旬地点で「バズる」ポイントが来ているように見える。そしてこのときの特徴語は以下のようになった..



1	Rank: 1,ホモビ
2	Rank: 2,フラ
3	Rank: 3,分身
4	Rank: 4,ワタ
5	Rank: 5,ホモゲー
6	Rank: 6,増殖
7	Rank: 7,クシ
8	Rank: 8,ルア
9	Rank: 9,直行
10	Rank: 10,丹下

図 18 「月ノ美兔」の「バズる」ポイントでの特徴語

樋口楓 「樋口楓」はグラフをみると2月末地点で「バズる」ポイントが来ているように見える。そしてこのときの特徴語は以下のようになった.



1	Rank: 1,ラップ
2	Rank: 2,タルト
3	Rank: 3,鈍感
4	Rank: 4,純情
5	Rank: 5,健気
6	Rank: 6,ため息
7	Rank: 7,小説
8	Rank: 8,citrus
9	Rank: 9,バナ
10	Rank: 10,SBJK

図 19 「樋口楓」の「バズる」ポイントでの特徴語

家長むぎ 「家長むぎ」はグラフをみると6月上旬地点で「バズる」ポイントが来ているように見える。そしてこのときの特徴語は以下のようになった.

1	Rank: 1,5902
2	Rank: 2,マジック
3	Rank: 3,ばば
4	Rank: 4,殺戮
5	Rank: 5,MM
6	Rank: 6,呆気
7	Rank: 7,jc
8	Rank: 8,ラー
9	Rank: 9,美感
10	Rank: 10,ぎむぎすこだあ

図 20 「家長むぎ」の「バズる」ポイントでの特徴語

名取さな 「名取さな」はグラフをみると5月下旬地点で「バズる」ポイントが来ているように見える。そしてこのときの特徴語は以下のようになった..

1	Rank: 1,パピルス
2	Rank: 2,グレータードッグ
3	Rank: 3,上半身
4	Rank: 4,アンダ
5	Rank: 5,貧弱
6	Rank: 6,下半身
7	Rank: 7,やんすこだあ
8	Rank: 8,バンダナ
9	Rank: 9,252
10	Rank: 10,saying

図 21 「名取さな」の「バズる」ポイントでの特徴語

笹木咲 「笹木咲」はグラフをみると8月上旬地点で「バズる」ポイントが来ているように見える。そしてこのときの特徴語は以下のようになった..

1	Rank: 1, 特技
2	Rank: 2, 商品
3	Rank: 3, ドラムロール
4	Rank: 4, セブン
5	Rank: 5, バター
6	Rank: 6, 20000
7	Rank: 7, レポ
8	Rank: 8, イレブン
9	Rank: 9, ビスケット
10	Rank: 10, ホリッ

図 22 「笹木咲」の「バズる」ポイントでの特徴語

9 今後の課題

特徴を取り出し今回の「バズる」という指標を調べるシステムは完成した。しかし、VTuber が流行る指標として「バズる」を定義したが、この方法でとった特徴語でほんとうに「バズる」という指標が取れたか疑問である。今後この指標を変えたり、調査対象を変更することでより良い「バズる」の指標を作りたい。また、動画と生放送の動画化でコメントの質が変わってしまうことが考察の時に判明したためその差別化も図りたい。さらに、最後の特徴語の出し方はより工夫することでもっと納得のいく「バズる」が作れるのではないかと思う。

宣伝

ここから先は宣伝だが、にじさんじバーチャルライバー 2 期生の家長むぎちゃんがかわいいので見てください。また、にじさんじの公式番組「にじさんじのくじじゅうじ」が AbemaTV で毎週水曜 21:00～ウルトラゲームスにて好評配信中! はちゃめちゃに面白いので見てください!

謝辞

本研究をするにあたって、プログラムを書く上で友人のなぼりたん氏がいなければおそらく形態素解析の実装で終わっていたと思う。今回最後まで面倒を見てくれたなぼりたん氏に感謝。

紹介

名前はなぼりたん、VTuber が好きで月ノ美兔を追いかけて破産しかけている大学三年生。

参考文献

- [1] @User Local バーチャル YouTuber ランキング, <https://virtual-youtuber.userlocal.jp/document/ranking?page=6> 2018/10/06 アクセス.
- [2] SOCIALBLADE, <https://socialblade.com/>
- [3] 松浦 健一郎, 司 ゆき, わかる Python, SBCreative, 2018
- [4] Wes McKinney, Python によるデータ分析入門 Numpy、Pandas を使ったデータ処理, 株式会社オライリー・ジャパン, 2013

付録

以下に各 VTuber の出た特徴語等の csv ファイルを配信する. 興味がある方は見てください.
https://www.dropbox.com/sh/721d24o4ip897bg/AAAw_WHHknDCTCtGXMxW9Ph-a?dl=0