

# SNSに支配されず SNSを支配しよう

思索駆動コース 古田 陸太

20代の9割が利用しているSNS。

みんなやっているから、置いていかれるのが怖いからやめられない。

見たっていいことないのに。疲れるだけなのに。

そんなSNSは、機械に任せてしまいましょう。

今回は、作者がよく使うTwitterを対象にします。

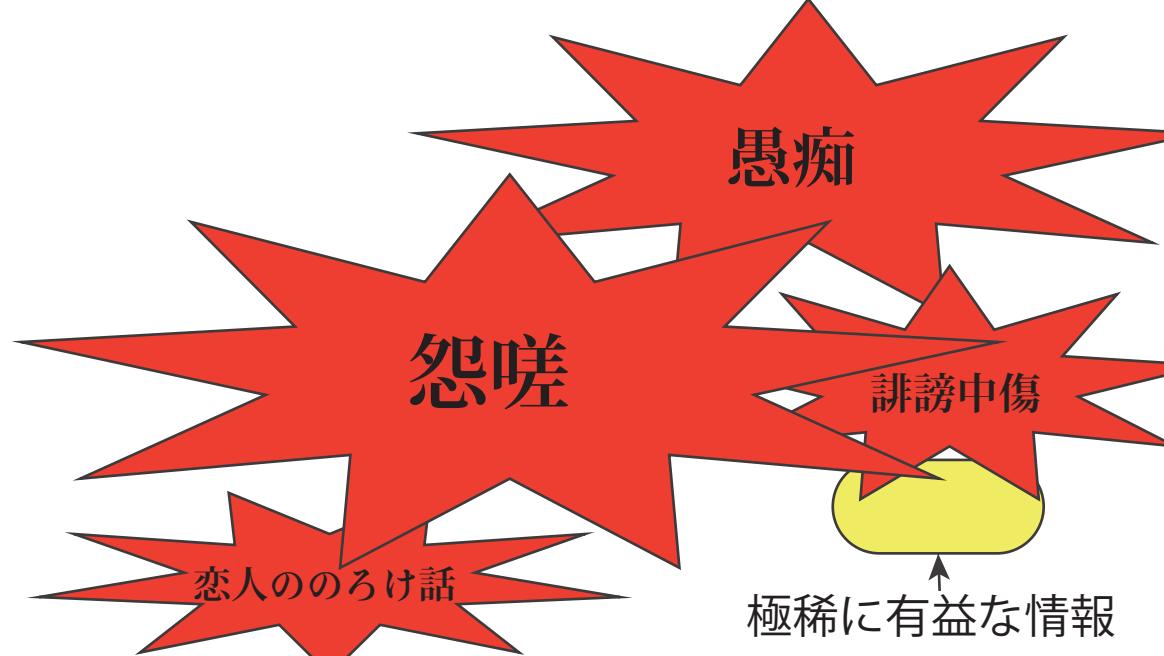
## SNSとの付き合い方を考える

SNSの長期利用は個人の幸福度に悪影響を及ぼす一方、「常に確認しないと大事な情報を見落とす気がする」などの理由でやめられない人は多い。確かに重要な速報をSNSから入手することはあるが、大多数はどうでもいい情報である。また、常にSNSを見ているわけではないのだから、肝心の情報を見落とすことも考えられる。この、SNSから得たいものと実際に得られるものが完全に逆転している状態をどうにかできないかと考え、今回の開発に至った。

### 理想のSNS

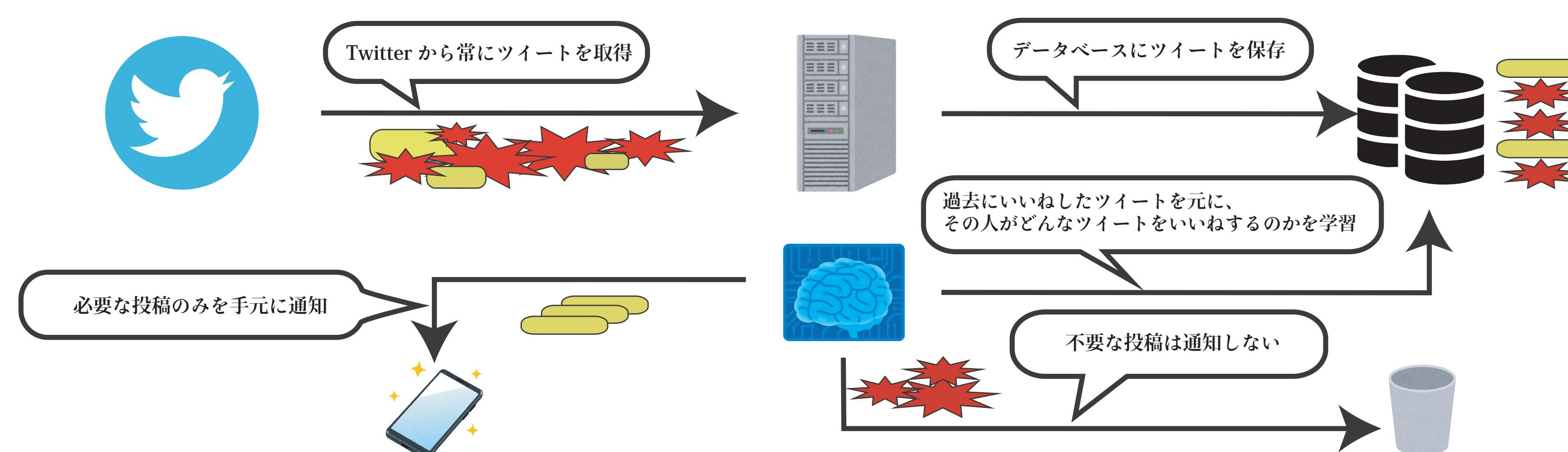


### 現実のSNS



## 設計

サーバー上のbotが定期的にTwitterのタイムラインを監視し、誰かが新しい投稿をしていたらそれをサーバー上に保存しておく。また、自分が新しく「いいね」をつけた投稿が存在したときも、後に使うためその情報を保存しておく。サーバー上には自分のタイムラインに存在したフォロワー達の投稿がすべて保存されるので、保存された投稿の中から自分にとって「必要な投稿」のみを抽出する（後述）。毎日タイムライン上に投稿される中で、「必要な投稿」と判定された投稿のみを、1日の終りに通知することで、余計な投稿を眺めること無く、自分に「必要な投稿」のみを得ることができる。通知の手段として、今回はLINEを実装したが、メールや他のアプリで通知する実装も可能である。



## 自分にとって「必要な投稿」の判定法

### ①自分が「いいね」をつけるくなるツイートを判定する

自分が過去にTwitter上でいいねした投稿をすべて取得し、BERTを用いて過去のツイートを分析する。その後に更新された投稿について自分が「いいね」をつけるかどうかを、過去のデータを元に教師あり学習を用いて自動で判定する。機械学習アルゴリズムとして、今回は勾配Boostingを使用した。

### ②リアルタイムな速報を取得する

速報性はSNSの大きな強み。ただし、SNSが一次情報であることは殆どないため、投稿には基本的にURLの形でソースが記載されている。リンク先のページに書かれている文章のカテゴリを分析することで、自分が興味あるテーマに関する投稿のみを取得することができる。文章カテゴリの分析法として、現在は「脆弱性」などの特定の単語が文章に出てくるかどうかを判定している。

## 結果

フォロー、フォロワーが共に1000人強の自分のアカウントでは、平均して1日3000件以上のツイートが流れてくるが、その中で自分に必要な投稿はせいぜい1ツイート程度であることがわかった。これまで1日1時間以上ツイッターを眺めていたが、これは丸ごと不要な時間である。また、数少ない自分に必要な投稿のうち半数近くを見落としていたことが判明した。すべてのツイートを自動で監視しておくことで、これらの取りこぼしを拾うことも可能であった。

結果として、SNSに使われるのではなくSNSを使う生活が実現できることが期待できる。

### BERTとは

SNSの投稿のような、人間が扱う自然言語をコンピューターに処理させる技術はNLP(Natural Language Processing)と呼ばれる。コンピューターは数字しか扱うことができないため、文章を何らかの形で数値データに変換する必要がある。様々な手法が提案される中で、2018年にGoogleにより公開されたBERTというモデルがそれまでの手法を圧倒的に上回る性能を叩き出した。NLPのモデルの多くは、大量の文章を学習させる必要があるが、今回利用したモデルは東北大学が作成した3万2千語の単語を含む文章を訓練させた大規模なモデルである。一方BERTには一度作ったモデルを使い回しやすいという特性があり、今回作成したTwitter用に調整されたモデルを他のSNS用に作り直すことなどは比較的容易である。

なお、2021年現在では、BERTよりさらに高性能とされるGPT-3がOpenAIにより作成されているが、こちらはモデル作成に途方も無いマシンパワーが必要であり、作成時の電気代が5億円を超えたとのことで今回は利用を断念した。いつか使ってみたい。



その日SNSに投稿された中で重要な投稿のみがLINEに届く様子

## NEXT!

今回はTwitterを対象としたが、今後は複数のSNSを快適にすることで生活をより向上させる。また、URL先の記事の内容判定を現在はパターンマッチで実装しているが、将来はウィキペディアの記事と記事のジャンルなどを教師データに、「いいね」をつけるかどうかの判定と同様の技術を用いて記事ジャンルを詳細に推定する予定である。開発は左のQRコード先にて継続中。



開発レポジトリのリンク