文章からワードクラウドを作成して頻出単語を視覚的に確認する

【重要】以下の欄をすべて埋めてから応募してください

氏名：[redacted]

所属学部学科(専攻)：[redacted]

email：[redacted]

【考察記述欄】作業をすべて終えてから考察を自由に記述してください。実行結果の考察だけでなく、データを集めたりプログラムを実行するにあたって苦労した点や、こんなデータを収集、解析してみたい、プログラムを改造してこんな解析や予測もしてみたいというアイデアなどでも構いません。また、実際にプログラムを改造するなど特別に工夫したことがあれば記載してください。

- - - 記述欄はじめ

◯背景

「ビットコイン」、「ブロックチェーン」、「仮想通貨」というワードがあります。微妙な差異はありますが、実はこの３つのワードはかなり近い意味を持ちます。「ビットコイン」とは「仮想通貨」（正しくは「暗号通貨」と呼びますが、一般的にはこのように呼ばれています）の一種であり、「ブロックチェーン」とは「仮想通貨」を実現する技術の一つです。

私はブロックチェーン技術に興味を持っており、この技術に関する卒業論文を執筆するために、これまでに多くの資料を読んできました。その中で、この３つのワードは近い意味を持つ一方で、その使われ方や印象は大きく異なっているのではないかと考えました。そこで、このレポートでは Twitter でこの３つのワードを含むツイートを検索し、ワードクラウドとすることで、３つの言葉のイメージがどのように異なるのか、その違いを可視化し、考察を加えてみます。

◯手法

このレポートでは、ツイートの収集には Twitter API を使いました。プログラミング言語の python を用いて API を叩き、2021/10/08 の 14:00~15:00 にかけてのツイートを収集しました。また、ワードクラウドの作成に当たって、分析の邪魔になりそうなツイート特有の文言（「#」、ユーザー名）や URL、当該ワード（一番大きく表示されることが自明であるため））を取り除きました。ワードクラウドはページ下部に表示されています。

◯考察

まず、「仮想通貨」のワードクラウドを見ると、他のものと比較して「俗っぽい」言葉が多いことが分かります。「資産」、「増税」、「為替」、「所得」といった言葉が多いことから、「仮想通貨」という言葉は「投資（投機）」の文脈で使われていることが推測されます。実際、2017 年前後に起きた「仮想通貨バブル」では「億り人」と呼ばれる、高額所得者が誕生しました。それ以降、バブルの崩壊や取引所のハッキングによる仮想通貨流出、いわゆる「コインチェック事件」を経て、仮想通貨投資は下火になっていました。しかし、最近ではビットコインの価格が 600 万円を越え[1]、再びハイリスク・ハイリターンな投資先として注目され始めているものと考えられます。また「今後」という言葉が比較的大きいことも印象的です。仮想通貨はボラティリティ（価格の上下）が激しいことでも知られており、現在の高値もいつ、どういう要因で崩れるのか予想

できません。そういった不安がこの言葉に反映されているのではないでしょうか。全体を通してみると、「仮想通貨」というワードの印象はやや「俗っぽい」または「雑多」なイメージがあると言えるでしょう。

次に、「ブロックチェーン」のワードクラウドを見てみると、今度はテクニカルな言葉が多いのが特徴的です。「ETF」は「上場投資信託」という意味ですが、これはビットコインを相当量保有する会社が組み込まれた ETF が「SEC」（米証券取引委員会）で承認されたニュース[2]を反映してのものと思われます。その他に、他のワードクラウドではあまり見られない言葉として「トレーサビリティ」や「技術」、「デジタル」、「暗号」といったものもあります。「ブロックチェーン」とは確かに「仮想通貨」（特定の主体に依存せずデジタルに価値の移転を可能にする技術）を実現する技術の一つですが、その高い可用性（システムがダウンする確率が低いこと）、情報の完全性（取引記録が改ざんされる確率が低いこと）は、フィンテック、即ち金融分野だけでなく、医療情報の管理、トレーサビリティの管理、IoT など、幅広い分野での応用が期待されています[3]。スケーラビリティ等の問題を抱えつつも、確実に世の中に浸透しつつあるこの新技術は、一般市民だけでなく、ブロックチェーンを応用しようとしているエンジニアや研究者などの注目も集めていることが、このワードクラウドから推測されます。総じて、「ブロックチェーン」というワードは暗号通貨の「テクニカルな」イメージを強く持っているように感じ取れます。

最後に「ビットコイン」のワードクラウドを見てみます。「ビットコイン」のワードクラウドからは、投資関連の言葉が目立つものの、社会との関連も示唆されるようなものも多いことが分かります。最も目立つ「米」、「上院」、「報告」、「財務」等は前述の ETF のニュースを受けてのものと思われます。一方、他のワードクラウドから見られなかった特徴的な言葉としては、「登録」、「ビッコレ」、「プレゼント」などが挙げられます。「ビッコレ」は広告を見ることで発生する報酬をビットコインで受け取るサービスのようです[4]。「登録」、「プレゼント」はこれに関連するワードと考えられます。この辺りは「仮想通貨」に近い「俗っぽさ」を感じます。一方で、「IOST」や「リップル」などの言葉もあります。これはビットコイン以外の主要な仮想通貨の一種です。このようなビットコイン以外の仮想通貨は「アルトコイン」などと呼ばれています。諸説ありますが、アルトコインとビットコインの値動きには、ある程度相関関係があると言われており、このことがワードクラウドにも反映されたのではないでしょうか。「仮想通貨」と「ブロックチェーン」と比較すると、「ビットコイン」はやや「仮想通貨」に近い言葉ですが、より相場や社会情勢を反映したワードが多いことが特徴です。このワードのイメージを言語化するのは難しいですが、新たな投資先としての「先進的」というイメージがあると言えるのではないでしょうか。

◯まとめ

このレポートでは、「ビットコイン」、「ブロックチェーン」、「仮想通貨」という３つのワードについて、ツイートを収集し、ワードクラウドを用いて各言葉のイメージを可視化しました。その結果、「仮想通貨」という言葉は、特に仮想通貨投資に関連する俗っぽい言葉が多く、「ブロックチェーン」は仮想通貨に関連する社会情勢や、ブロックチェーン技術のテクニカルな側面を反映した言葉が多く、「ビットコイン」は仮想通貨の「相場」に関連する言葉が多いことが分かりました。

以上から、「仮想通貨」は「お金儲け」やハイリスク・ハイリターンな「投資」というやや「俗っぽい」イメージ、「ブロックチェーン」は仮想通貨を実現する技術としての「テクニカルな」イメージ、「ビットコイン」は新たな投資先としての「先進的な」イメージがあると言えるでしょう。

◯感想

私が興味を持っているブロックチェーン技術に関連して、「仮想通貨」、「ブロックチェーン」、「ビットコイン」の３つのワードに対する、大衆のイメージをワードクラウドから推測してみました。ワードクラウドの作成自体は用意されていたプログラムを走らせるだけなので簡単でしたが、ツイートを集めるところでやや

手間がかかりました。TwitterAPI は Python 等から動かすことが出来ますが、そのためにアクセスコードなどを取得する必要があります。また、大量のツイートは集められないことや、最大でも直近の 1 週間以内のツイートしか検索できないことなど、様々な制約もありました。

また、集めたツイートから余計な語句を取り除く処理にも Python を使いましたが、正規表現の扱いに慣れていなかったので、これもなかなか骨が折れる作業でした。しかし、結果的にはワードに対する大衆のイメージや、暗号通貨に纏わる直近の変化など、様々な要素を取り込んだきれいなワードクラウドを作ることが出来たので、この点はまず満足です。

これは TwitterAPI を使う上の制約なので、どうしようもないのですが、ワードに対するイメージを抽出するという目的なので、過去のツイートも全て反映したワードクラウドを作ることができれば、一貫したイメージが何なのか推測することができただろうと思います。現状では時期的な変化が大きく反映されすぎており、本質的なイメージを捉えることが出来ていないようにも感じます。また、今回は技術的に出来ませんでしたが、時系列の変化などを動的に反映させた gif なんかを作れることができれば、イメージの変化を可視化でき、面白い分析ができるだろうなと思いました。

◯参考文献

[1]CoinMarketCap, Bitcoin, https://coinmarketcap.com/ja/currencies/bitcoin/（確認　2021/10/08）

[2]Yahoo!Finance（2021）「米 SEC、ビットコインを保有する起業で構成される ETF を承認」https://finance.yahoo.co.jp/news/detail/20211008-01077002-fisf-market（確認　2021/10/08）

[3]Casino, Fran & Dasaklis, Thomas & Patsakis, Constantinos. (2018). A systematic literature review of blockchain-based applications: Current status, classification and open issues. Telematics and Informatics. 36. 10.1016/j.tele.2018.11.006.

[4]ビッコレ, https://bikkore.jp/（確認　2021/10/08）

記述欄おわり---

(ここからプログラムが始まります)

最近良く見かけるワードクラウドは、文章の中から単語を出現頻度に応じた大きさで可視化する手法です。MATLAB のテキストマイニング機能を利用して文章データから簡単にワードクラウドを作成できます。

身の回りにあるテキストデータを使ってワードクラウドを作成してみてください。よく使われている単語を視覚で確認することで、思わずおおっ！と声が出そうになる驚きや気づきがあるかもしれません。

事前準備:

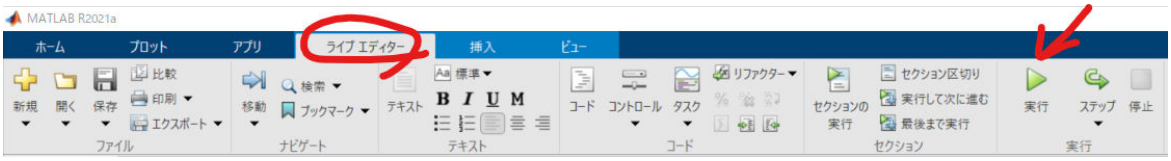
- テキストファイルを用意してください。英語でも日本語でも大丈夫です。自分が毎日つけている日記やメモ書き、ブログを開設していたらブログ記事でも良いでしょう。インターネット上で公開されているパブリックドメインの文章、演説や小説などでは話し手や書き手の癖が垣間見えたりするかもしれませ

ん。ファイル形式がテキスト形式ではない場合は、Windows メモ帳を開いて、該当の文章をすべてコピーしてメモ帳に張り付けてから、テキスト形式(.txt)のファイルとして保存してください。

プログラムの実行:

下のステップ 1 を読んでプログラムを書き換えてから、このプログラムを実行してください。

Tip: メニューに実行ボタンが見当たらない時は、[ライブエディター]タブを選択してください。



プログラム

ステップ 1: テキストファイルの読み込み

以下の作業を済ませてからプログラムを実行してください。

- 事前準備で用意したテキストファイルをこのプログラムと同じ場所に保存してください。
- MATLAB の「現在のフォルダー」をこのプログラムとテキストファイルが保存されているフォルダーに変更してください。
- 下のプログラムの filename = の後のファイル名を実際に読み込むファイル名に書き換えてください。

同梱のサンプルファイル"hakusho.txt"は令和 2 年度版情報通信白書「はじめに」の本文です。

出典：「令和 2 年度版情報通信白書」(総務省) 「はじめに」 https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/pdf/n1000000.pdf

令和 2 年度版情報通信白書の二次利用について(二次利用可能)：https://www.soumu.go.jp/main_content/000700124.pdf

```
% 初期化します
clear
%  「ビットコイン」、「仮想通貨」、「ブロックチェーン」に関するツイートを含むファイルを開きます。
%  ファイルは 2021/10/08 14:00~15:00 に行われたツイートを Twitter api で取得し、URL や「RT」、
%  「#」、ユーザー名、「仮想通貨・ビットコイン・ブロックチェーン」といった文言を抜いた純粋なテキストフ
vc = ['tweets_virtualcurrency_strip.txt'];
bc = ['tweets_blockchain_strip.txt']
```

```
bc =
'tweets_blockchain_strip.txt'
```

```
bt = ['tweets_bitcoin_strip.txt']
```

```
bt =
'tweets_bitcoin_strip.txt'
```

```
% 上で指定したファイルを文字列として変数 xx_text に格納
vc_text = extractFileText(vc);
bc_text = extractFileText(bc);
bt_text = extractFileText(bt);
```

ステップ **2:** ワードクラウドの作成

MATLAB では文章データから自動的に単語を抽出、カウントしてワードクラウドを作成してくれます。

```
% 「仮想通貨」を含むツイート（「仮想通貨」というワードは中心に来ることが自明なため除外）のワードクラウド
wordcloud(vc_text);
```


```
% 「ブロックチェーン」を含むツイート（「ブロックチェーン」というワードは中心に来ることが自明なため除外）
wordcloud(bc_text);
```


```
% 「ビットコイン」を含むツイート（「ビットコイン」というワードは中心に来ることが自明なため除外）のワード
wordcloud(bt_text);
```

考察:

実行結果に対して、冒頭の考察記述欄に考察を自由に記述してみましょう。

保存(レポート化):

ここまでの作業お疲れさまでした。最後に作業の結果を残しておきましょう。

このプログラムは実行結果とともにレポート出力できます。

[ライブエディター]タブの[エクスポート]ボタンを押して、PDF を選んでエクスポートしてください。

保存された PDF 形式のファイルをコンテスト窓口に送付してください。