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

はじめに

2021 年 10 月 7 日に千葉県北西部で発生した地震に関するツイートを収集し，ワードクラウドで可視化した．本レポートでは，今回の地震に関するツイートにどのような単語が含まれているのかを明らかにする．

使用したデータについて

データ収集には TwitterAPI を用い，「地震」を検索キーワードとした．検索投稿時間の設定は，地震発生直後の 22:45 とその後は 7 日の 23:00 から 8 日の 22:00 まで 1 時間ごとをとし，これを検索基準時間とした．なお，TwitterAPI の取得数制限により，取得できたのは各検索基準時間から，投稿時間順に遡って 150000 件までのツイートであった（例：8 日の 8:00 を検索基準時間とした場合は 7:59:59~7:47:03 までに投稿されたツイートデータ）．リツイートは除いた．分析には 7 日 22:45(22:44:59~22:44:44{150000 件})[検索基準時間 (取得ツイートの投稿時間{取得件数})], 8 日 8:00(7:59:59~7:47:03{150000 件}), 12:00(11:59:59~11:14:35{150000 件}), 16:00(15:59:59~15:00:01{9693 件})に投稿されたツイートのテキストデータを用いた．ただし，そのまま可視化すると検索単語である「地震」が最も出現回数が多い単語として表示されてしまうので，前処理の段階で「地震」を空白に置き換えた．

考察

どの時間でも「大丈夫」が，出現回数が多い単語として表示された．災害発生時に投稿される「大丈夫」には「こっちは大丈夫」といった自身の安否を知らせるという意味と「みなさん地震大丈夫ですか？」といったフォロワーを心配する２つの意味があると考えられる．さらに後者に関しては，「地震」が Twitter のトレンドに表示されるから，流行りに乗るために「地震」を入れてツイートをする人も多い．もとのデータを見たところ，ほとんどの「大丈夫」が後者の意味で使われており，「地震」とともに投稿される「大丈夫」の多くが緊急性を要しないツイートであると示唆された．

次に「大丈夫」の単語以外の結果をみていく．7 日の発生直後は「怖い」「揺れる」「速報」が出現回数上位の単語にあがった．関東で最大震度 5 強を観測した地震は東北地方太平洋沖地震以来であったことから，ひさしぶりの地震や緊急速報に驚いた人々が多かったと考えられる．

翌日 8 日になると，どの時間帯も「昨日」が「大丈夫」とともに上位にあがった．翌日でも昨夜の地震の心配をするツイートが多く投稿されていると考えられる．また，「怖い」や「影響」も上位の単語としてどの時間帯にもあがったことから，人々の地震や今後の余震に対する恐怖が読み取れる．「電車」については，時間がたつにつれて表示サイズが小さくなっていったことから，遅延が解消していったことが伺えた．

おわりに

今回の分析を通して，検索にかける単語を組み合わせることで，緊急性の低いツイートを除外し，有益な情報を得ることができるようになると考えた．今後は単語のサイズだけではなく，細かな出現回数や単語同士の関連性についても分析していきたい．

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

```
fileattrib
```

```
             Name: 'C:\Users\miyuk\Desktop\NiigataDSContest2021\1_TextWordCloud'
          archive: 0
           system: 0
           hidden: 0
        directory: 1
         UserRead: 1
        UserWrite: 1
      UserExecute: 1
        GroupRead: NaN
       GroupWrite: NaN
     GroupExecute: NaN
        OtherRead: NaN
       OtherWrite: NaN
     OtherExecute: NaN
```

次利用について(二次利用可能)：https://www.soumu.go.jp/main_content/000700124.pdf

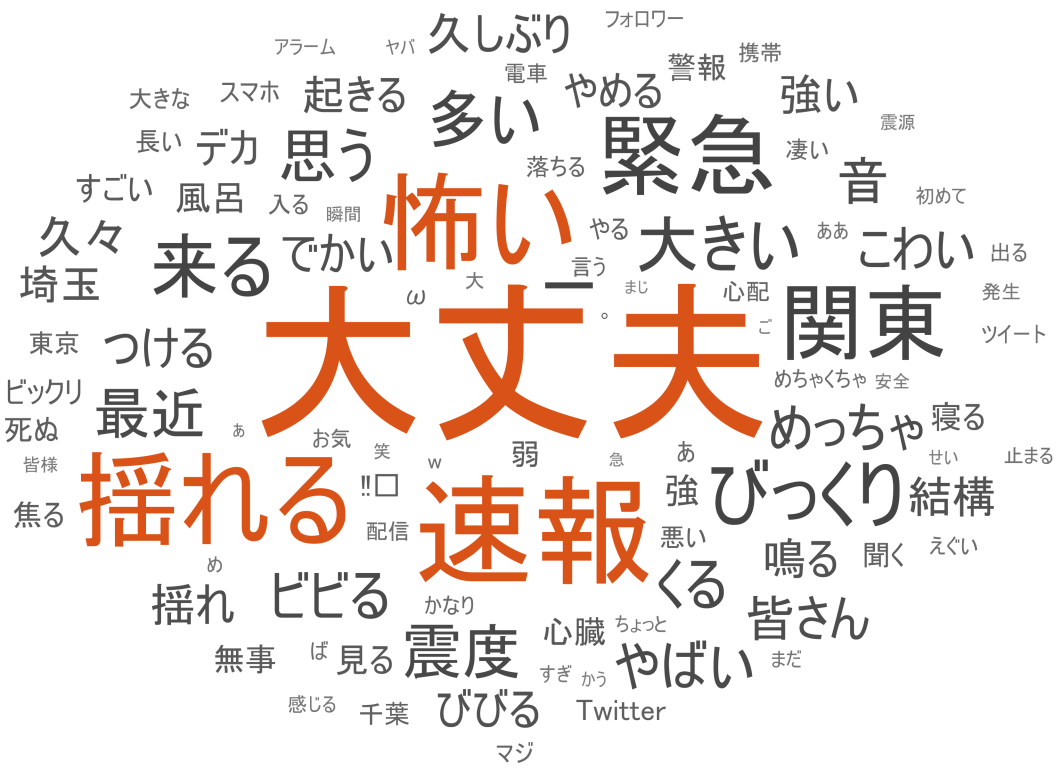
```
% 初期化します
clear

% 実際に読み込むファイル名に書き換えてください
filename = ['関東 1007_h2245.txt'];

% 上で指定したファイルを文字列として変数 text に格納します
text = extractFileText(filename);

droptext = ["地震"];
text = replace(text,droptext," ");
```

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。

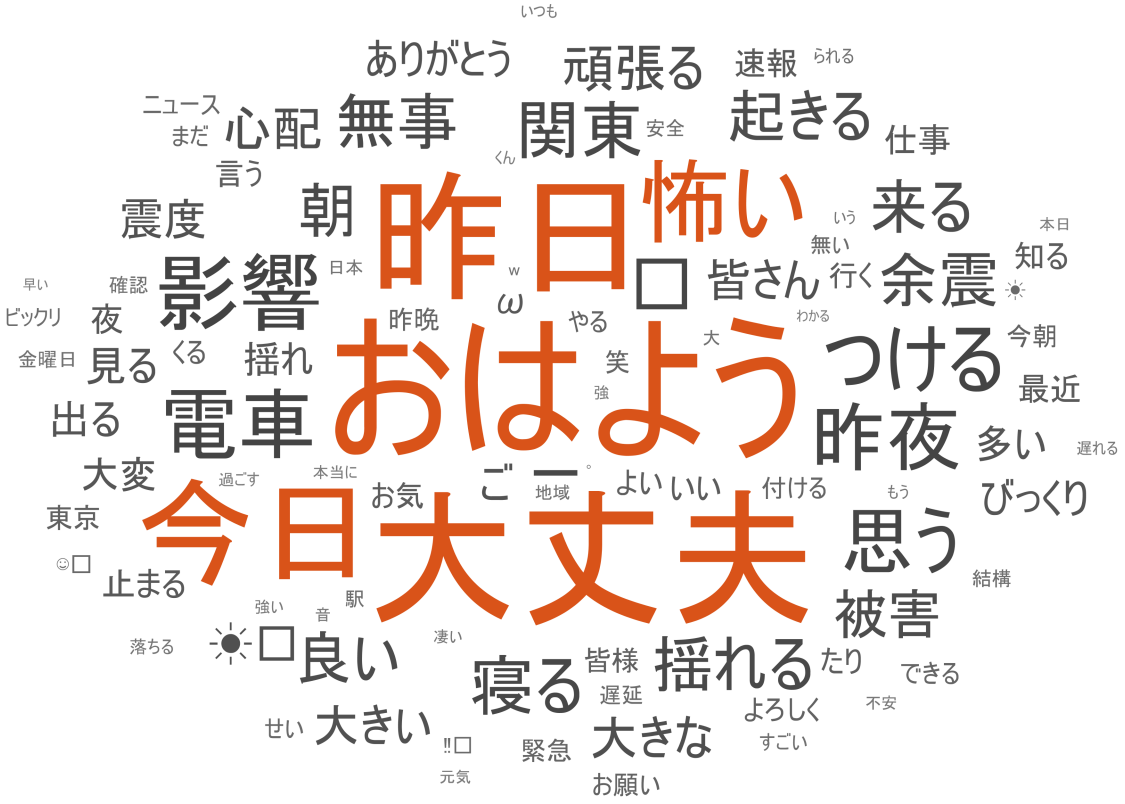
```
wordcloud(text);
title("関東 10 月 7 日地震 7 日発生直後 22:45");
```

関東10月7日地震 7日発生直後22:45



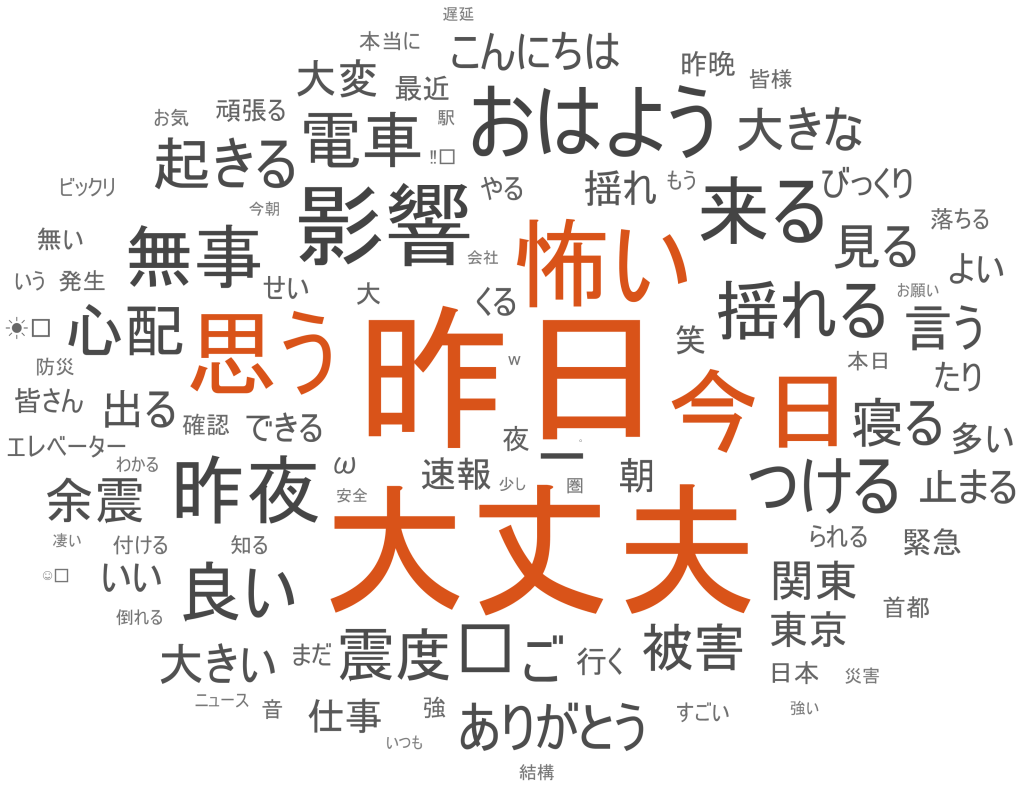

```
%朝
% 初期化します
clear

% 実際に読み込むファイル名に書き換えてください
filename = ['関東 1007_h0800.txt'];

% 上で指定したファイルを文字列として変数 text に格納します
text = extractFileText(filename);

droptext = ["地震"];
text = replace(text,droptext," ");

wordcloud(text);
title("関東 10 月 7 日地震 8 日朝 08:00");
```

関東10月7日地震 8日朝08:00



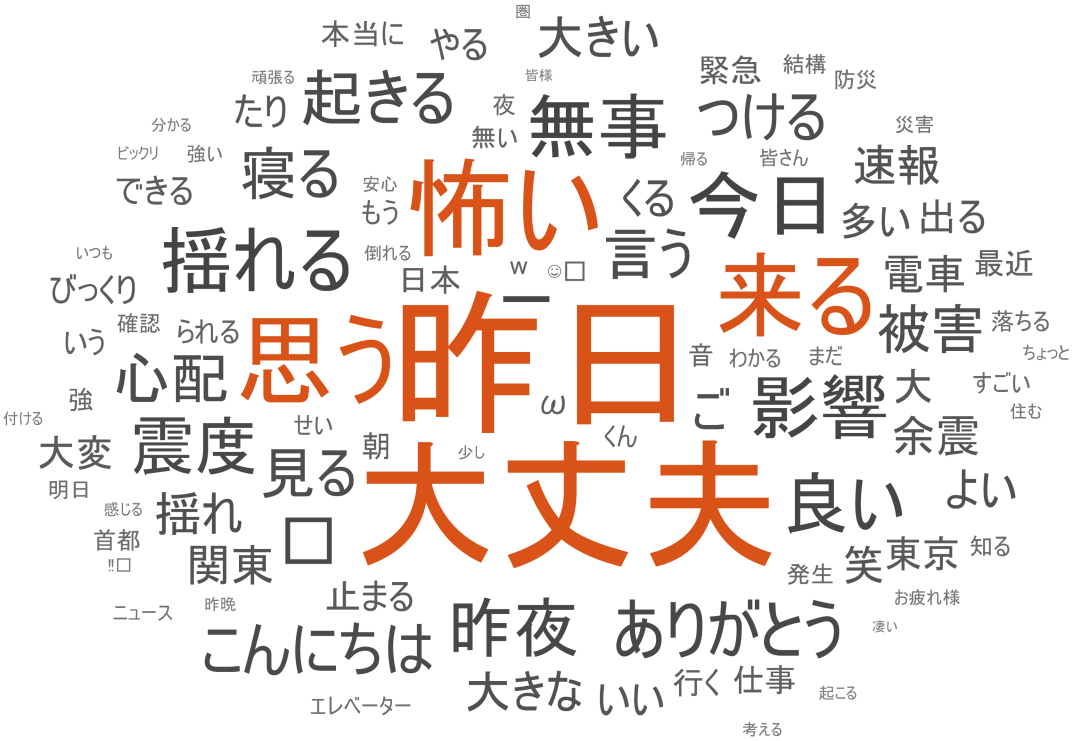

```
%昼
% 初期化します
clear

% 実際に読み込むファイル名に書き換えてください
filename = ['関東 1007_h1200.txt'];

% 上で指定したファイルを文字列として変数 text に格納します
text = extractFileText(filename);

droptext = ["地震"];
text = replace(text,droptext," ");
wordcloud(text);
title("関東 10 月 7 日地震 8 日昼 12:00");
```

関東10月7日地震 8日昼12:00




```
%夕方
% 初期化します
clear

% 実際に読み込むファイル名に書き換えてください
filename = ['関東 1007_h1600.txt'];

% 上で指定したファイルを文字列として変数 text に格納します
text = extractFileText(filename);

droptext = ["地震"];
text = replace(text,droptext," ");

wordcloud(text);
title("関東 10 月 7 日地震 8 日夕方 16:00");
```

関東10月7日地震 8日夕方16:00



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。

Copyright 2021 The MathWorks, Inc.