Performing Text Analysis on Tweets

Introduction

We want to create a flexible framework for analysing a collection of tweets. For now, we want to do the following:

First, we will make a simple webpage which contains a form asking for a Twitter handle. We will then perform some text analysis of, say, the last 1000 of tweets from this handle. For now, the analysis will consist of the most common hashtags used by the handle, and the most common words used in the text of the tweet. For the latter, we will exclude common English words like articles and also collapse different forms of the same word (like “going” and “gone”).

The rest of the document lays down the steps to achieve our objective.

Basic framework

We use Python for coding. The website is created using the Django framework, and the text analysis is carried out with using the nltk library. We use the Twitter API for accessing the Twitter feed and extract metadata from it, and some nltk modules for tweet analysis.

We deploy the code on a server using Google Cloud, Heroku etc. (TODO).

Using Django

Django is a Python framework for web development. We will cover all the required steps in this document; Django also has a step-by-step [tutorial](https://docs.djangoproject.com/en/3.1/intro/tutorial01/) on its website. We assume that Django is already installed. If not, here are the [directions](https://docs.djangoproject.com/en/3.1/intro/install/) to install it.

All the code in this document is stored in github at the following link (TODO).

Creating a Django project

Type the following code:

django-admin startproject tweet\_analysis

This step will create a directory named tweet\_analysis in your working directory.

Creating Django app enter\_handle

Now we create an app. We first show the user a form which takes a Twitter handle as input. We then access the handle’s Twitter feed, extract the last 1000 tweets, and then perform text analysis on the tweets. Since the Twitter API part is not completed yet, for demonstration purposes, we use a corpus of Tweets in the nltk library instead.

First step is to create the app:

python manage.py startapp enter\_handle

This step creates a directory called enter\_handle under the project folder tweet\_analysis. (A project is a collection of apps, and an app can be reused in multiple projects.)

Handling the URL for the input page

We then create a simple webpage for user input. First we need to handle the URL for the webpage. To do this, we need the following two steps.

First, we edit the file enter\_handle/urls.py

**from** **django.urls** **import** path

**from** **.** **import** views

urlpatterns = [

path('', views.index, name='index'),

]