Final Project

Huajun Chai

998584845

Traffic data mining with social media-using twitter as an example

Introduction

Nowadays, more and more people are using smartphone and having their social media account, like twitter and facebook. People using these social media post a lot amount of posts online concerning their life and other things. As these data is so huge, it has great value in term of data mining in many fields. And there are many applications concerning of using these data sets.

As a transportation student, I really want to find applications in my area with those huge social media data. The most useful information will be those about accidents, congestion, low speed, workzone ahead and so on. Since there are so many users, and the posts are relatively real-time, it will be a great supplyment to our existing traffic monitoring system, or even better than the existing systems.

In my application, I will first use twitter API to crab current twitters from the website by some keywords, like "congestion". After that, a sentiment analysis is performed to decide if a certain post is "positive", "negative" or "nuetual". At the same time, the geo-location information is recorded as well if it is available(some twitter uses don't allow twitter application to record their location information, in this case, this information is unavailable). After that, there is an option for the users to save the result to a database, for later research usage.
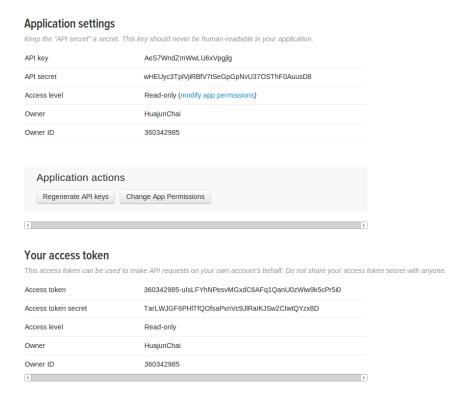

Instructions on platform set up

Twitter developer tools

In here, https://dev.twitter.com/ login with your twitter account, and create an application within that website. After that, you will get your API key and API secret, as well as consumer key and consumer secret under the account information. We need these information to login twitter through API.

Application settings

Keep the "API secret" a secret. This key should never be human-readable in your application.

API key	AeS7WndZmWwLU6xVpgjlg
API secret	wHEUyc3TplVjiRBfV7tSeGpGpNvU37OSThF0AuusD8
Access level	Read-only (modify app permissions)
Owner	HuajunChai
Owner ID	360342985

Application actions

Regenerate API keys   Change App Permissions

Your access token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access token	360342985-uIsLFYhNPesvMGxdC6AFq1QanU0zWiw9k5cPr5i0
Access token secret	TarLWJGF6PHlTfQOfsaPxnVc9JlRaIKJSw2ClwtQYzx8D
Access level	Read-only
Owner	HuajunChai
Owner ID	360342985

DATUMBOX

Besides twitter API, we also use an other API called DATUMBOX. It is mainly used as sentimant analysis tools. You can find information about this tool in here: http://datumbox.com.

Within that website, you need to register an account. And in your account information, you can find your API key:

Click "See your API key":

You will see:

API Key

API Credentials

User Id

1361

API Key

7908aa9d442bacb5da824446f8779298

Subscription Id

1384

Write down this API key. You will need it shortly.

In the file "config.php", there are:

<?php

// Here in the second argument, enter your API key from datumbox.com

define('DATUMBOX_API_KEY', 'your datumbox key');

// Configure your consumer key and consumer secret here. They can be found in your twitter account.

define('TWITTER_CONSUMER_KEY', 'your twitter API key');

define('TWITTER_CONSUMER_SECRET', 'your twitter API secret');

// Configure your authentication credentials. They can be created in your twitter account too.

// For detailed information, please refer to my documentation.

define('TWITTER_ACCESS_KEY', 'your twitter access key');

define('TWITTER_ACCESS_SECRET', 'your twitter access key');

// Please note that, all these keys and secrets are all from my accounts.

// If you want to make the applicaiton to work in your machine, you need to create all these

// information by yourself and fill them in.

After filling those information, the account set up is done. Then we need to set up the softwares.

Apache & PHP

We use Apache as the web server, you can download and install it from here: http://httpd.apache.org/.

For the website inqury, we use PHP. If you are using Linux machine, you can simply install this by:

sudo apt-get install php5 libapache2-mod-php5

sudo /etc/init.d/apache2

Actually, if you are using ubuntu, you can use the following command to install all the required softwares and packages at once:

```
sudo tasksel install lamp-server
```
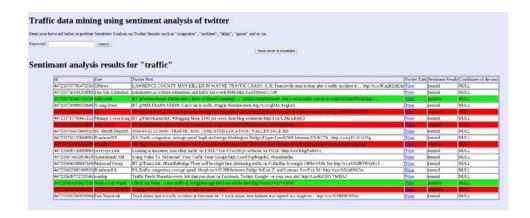
After getting these packages and softwares ready, you are good to run the webpage. But before that, it is good to set the root direcotry for Apache. You can get this instructions here: https://netbeans.org/kb/docs/php/configure-php-environment-ubuntu.html.

Creating a New Virtual Host

1. To launch the Terminal, choose Applications>Accessories>Terminal. The Terminal window opens.
2. To copy the configuration file of the default virtual host to a new file (mysite), type the following command at the command prompt:

```
sudo cp /etc/apache2/sites-available/default /etc/apache2/sites-available/mysite
```

3. Run the gedit application and edit the new configuration file (mysite) in it:

```
gksudo gedit /etc/apache2/sites-available/mysite
```

If asked, enter the password that you specified for the root user during the installation of your operating system.

4. Change the Document Root to point to the new location:

```
/home/<user>/public_html/
```

5. Change the Directory directive, replace

```
<Directory /var/www/>
```

with

```
<Directory /home/user/public_html/>
```

You should put all the files in a folder, for example named twitter, under the root folder. After all these, open your web browser, and enter the following url:

127.0.0.1/twitter/index.php

You will see a webpage like this:

Features

The application can search and get the twitters by some key words through twitter API. We can get the post ID, the user name, the content of that post, the actual link of the post, the sentiment result and the geo-location information when the user posts that post(this function is not available for now, but will definitely be added shortly as it is a key information when we do transportation research).

1. Will provide option to user to save the search result to some database. Or this can be automatically done as the web crawler will do. This functionality is also not available for now. But it is relatively easily to implement. Just add a connection to the local database, and store all the information in that database using SQL queries.

Future things

Now the sentiment analysis only can predict three status: "positive", "negative" and "neutral". But in our case, more analysis will be needed. For example, if one user reports low speed, we might be looking forward to a speed value. If one user reports accident, there is a high possibility that will be the information about how many lanes are closed. We need these information in transportation research.

And also the searching should be more intelligent. We might provide the application some keywords, but as we know we can't provide all the relevant words. We need some machine learning method to get the application smarter to "learn" what information is related to transportation research.

The saving of information should be made automatically.