

YouTube留言

AI情緒判讀

自主學習成果

Mood Bot 情緒機器人

搭載功能：

- 爬取YouTube留言，並做正負面情緒的判斷與統計。
- 亦能對單篇文章或者一串文字進行正負面的情緒判斷。

使用技術：

- 爬蟲與YouTubeAPI
- 機器學習(單純貝氏)

小組成員：

- 李庭宇
- 林敬哲
- 林秉弦

介紹

MoodBot 使用新浪微博之 weibo_senti_100k 資料庫, 廣泛蒐集了正負向評論約各 5 萬條, 並且搭載了 Multinomial Naive Bayes 機率模型做作為機器學習演算法, 是一種監督式的分類演算法, 對文字具有不錯的準確度(雖然說這隻得準確率 accuracy 只來到了 70%)。

動機

滑著YT永無止盡的留言, 看著主旨大多相同的內容, 於是我就有了統計所有留言情緒的想法。

過程

1. 首先要學會Python的基礎語法。

來題簡單的數字排序:

求使用者輸入10個數字並存放在串列中。接著由大到小的順序顯示最大的 3個數字。

```
a = []
for i in range(0,10):
    a.append(int(input()))
a.sort(reverse=True)
print(*a[:3])
```

2. 再來是需要使用Flask套件來當 Line 的 WebHook server。

基礎的 Flask Server 範例:

```
app = Flask(__name__)

@app.route('/')
def hello_world():
    return 'Hello, World!'

app.run()
```

作為 Line 的 WebHook Server:

```
# 載入需要的模組
from __future__ import unicode_literals
import os
```

```
from flask import Flask, request, abort
from linebot import LineBotApi, WebhookHandler
from linebot.exceptions import InvalidSignatureError

app = Flask(__name__)

# LINE 聊天機器人的基本資料
line_bot_api = LineBotApi('聊天機器人的 Channel access token')
handler = WebhookHandler('聊天機器人的 Channel secret')

# 接收 LINE 的資訊
@app.route("/callback", methods=['POST'])
def callback():
    signature = request.headers['X-Line-Signature']

    body = request.get_data(as_text=True)
    app.logger.info("Request body: " + body)

    try:
        handler.handle(body, signature)
    except InvalidSignatureError:
        abort(400)

    return 'OK'

if __name__ == "__main__":
    app.run()
```

3. 接著就是爬蟲的部分

由於 YouTube 爬起來相當複雜，於是後來改為採用 API 的方式。

4. 最後就是上我們機器人的大腦拉～

首先把文字經過頻率轉換成數字，但我們這裡使用的 TfidfVectorizer 與一般的 CountVectorizer 不一樣的地方是它會在字的出現次數上多乘上常用程度的反比。因為是整數特徵，所以採用的是 Multinomial Naive Bayes。

貝葉斯定理：

下方式子是對的：

$$P(A) \times P(B|A) = P(B) \times P(A|B)$$

接著將其移項：

$$P(B|A) = \frac{P(B) \times P(A|B)}{P(A)}$$

單純貝氏就是假設樣本的每個特徵都是獨立的。

結果

流程總覽

將從 Line 收到的指令訊息爬取對應的 YouTube 影片，再丟進 AI 裡後，將結果透過 Line 傳回給使用者。

使用者 → Line → YouTube → AI → Line → 使用者

