**NLP 文本分類 Text Classification**

Input : text / paragraph

Model : CNN

Output : category

**Text Classification with CNN and RNN**

資料集：中文文本分類資料集 THUCNews <http://thuctc.thunlp.org/#%E4%B8%AD%E6%96%87%E6%96%87%E6%9C%AC%E5%88%86%E7%B1%BB%E6%95%B0%E6%8D%AE%E9%9B%86THUCNews>

輸出：類別 (體育、財經、房產、家居、教育、科技、時尚、時政、遊戲、娛樂)

URL: <https://github.com/gaussic/text-classification-cnn-rnn>

**TextClassify**

資料集：MR電影評論數據 <http://www.cs.cornell.edu/people/pabo/movie-review-data/>

教學：<https://github.com/Kevinwenya/TextClassify/blob/master/text_classify_details.pdf>

**NLP-使用CNN進行文本分類**

1. News data: crawled historical news headlines from Reddit WorldNews Channel (/r/worldnews). They are ranked by reddit users' votes, and only the top 25 headlines are considered for a single date. (Range: 2008-06-08 to 2016-07-01)
2. binary classification task

Dow Jones Industrial Average (DJIA) 道瓊工業平均指數

"1" when DJIA Adj Close value rose or stayed as the same. (指數上升或維持一樣)

"0" when DJIA Adj Close value decreased. (指數下降)

1. Training Set : from 2008-08-08 to 2014-12-31 / Testing Set : from 2015-01-02 to 2016-07-01.

This is roughly a 80%/20% split.

1. use AUC as the evaluation metric

**A Comprehensive Guide to Understand and Implement Text Classification in Python**

**手把手教你在Python中實現文本分類（附代碼、資料集）**

資料集：亞馬遜的評論資料集

<https://gist.github.com/kunalj101/ad1d9c58d338e20d09ff26bcc06c4235>

<https://drive.google.com/drive/folders/0Bz8a_Dbh9Qhbfll6bVpmNUtUcFdjYmF2SEpmZUZUcVNiMUw1TWN6RDV3a0JHT3kxLVhVR2M>

**Text Classification With Python and Keras**

資料集：sentiment labelled sentences/ yelp\_labelled.txt

sentiment labelled sentences/ amazon\_cells\_labelled.txt

sentiment labelled sentences/ imdb\_labelled.txt

模型：CNN

輸入：從3個網站收集句子 (imdb.com / amazon.com / yelp.com)

輸出：1 (for positive) or 0 (for negative)

URL: <https://www.kaggle.com/kernels/scriptcontent/8006802/data>

**Classify Kaggle San Francisco Crime Description**

Goal: Predict the category of crimes that occurred in the city by the bay

Data: Kaggle San Francisco Crime <https://www.kaggle.com/c/sf-crime/data>

Input: Descript

Output: Category

Model: CNN

Example:

|  |  |
| --- | --- |
| GRAND THEFT FROM LOCKED AUTO | LARCENY/THEFT |
| POSSESSION OF NARCOTICS PARAPHERNALIA | DRUG/NARCOTIC |
| AIDED CASE, MENTAL DISTURBED | NON-CRIMINAL |
| AGGRAVATED ASSAULT WITH BODILY FORCE | ASSAULT |
| ATTEMPTED ROBBERY ON THE STREET WITH A GUN | ROBBERY |

URL: <https://github.com/jiegzhan/multi-class-text-classification-cnn-rnn>

參考資料：

* A Comprehensive Guide to Understand and Implement Text Classification in Python

<https://www.analyticsvidhya.com/blog/2018/04/a-comprehensive-guide-to-understand-and-implement-text-classification-in-python/>

* 手把手教你在Python中實現文本分類（附代碼、資料集）

<https://zhuanlan.zhihu.com/p/37157010>

* Classify Kaggle San Francisco Crime Description

<https://github.com/jiegzhan/multi-class-text-classification-cnn-rnn>

* Text Classification with CNN and RNN

<https://github.com/gaussic/text-classification-cnn-rnn>

* TextClassify

<https://github.com/Kevinwenya/TextClassify>

* NLP-使用CNN進行文本分類

<https://blog.csdn.net/spring_willow/article/details/80011161>

資料集Combined\_News\_DJIA.csv

<https://www.kaggle.com/aaron7sun/stocknews>

* 文本處理——基於 word2vec 和 CNN 的文本分類：綜述 & 實踐（一）

<https://blog.csdn.net/u010417185/article/details/80649356>

* Python-3-Data-Analysis-Basics

<https://github.com/yenlung/Python-3-Data-Analysis-Basics>

* Text Classification through CNN, RNN & HAN using Keras

<https://github.com/jatana-research/Text-Classification>