今晚我想來點異國料理 - 食譜分類模型

ZC-106802513張書榕

ZC-106502504倪滙渝

一、動機

每個國家都有獨特的飲食文化，此次我們藉由料理的食材與佐料判斷各國料理的特性，期許電腦依原料來判斷此料理的出處為哪一個國家。

二、方法

1. 資料集：

2016年由Kaggle所提供的Recipe Ingredients Dataset，其中有39774組共20個國家的料理，我們將其切割成0.8的訓練集與0.2的測試集，以下為兩組範例。

|  |  |  |
| --- | --- | --- |
| id | cuisine | ingredients |
| 10259 | greek | ['romaine lettuce',......] |
| 25693 | southern\_us | ['plain flour',......] |

1. 前處理與特徵選取：

為了消除某些特殊材料，我們先將全部材料裁剪成「單一國家中最常使用的前幾種材料」後再進行文字向量化。但由於特徵維度仍舊太大，因此使用PCA降維到低維度。

1. 分類與預測：

使用SVM、KNN進行模型訓練與預測。

三、實驗與分析

1. 以accuracy評斷各項分類器的結果。
2. 討論不同維度對其分類結果之影響。