FLIP01 FINAL PRESENTATION

Zhaoyang Wang

Xi'an Shiyou University, China

Introduction

Some of our strongest geographic and cultural associations are tied to a region's local foods. This playground competitions asks you to predict the category of a dish's cuisine given a list of its ingredients. This is a natural language processing problem, so we need to use related methods to deal with it.

data set presentation

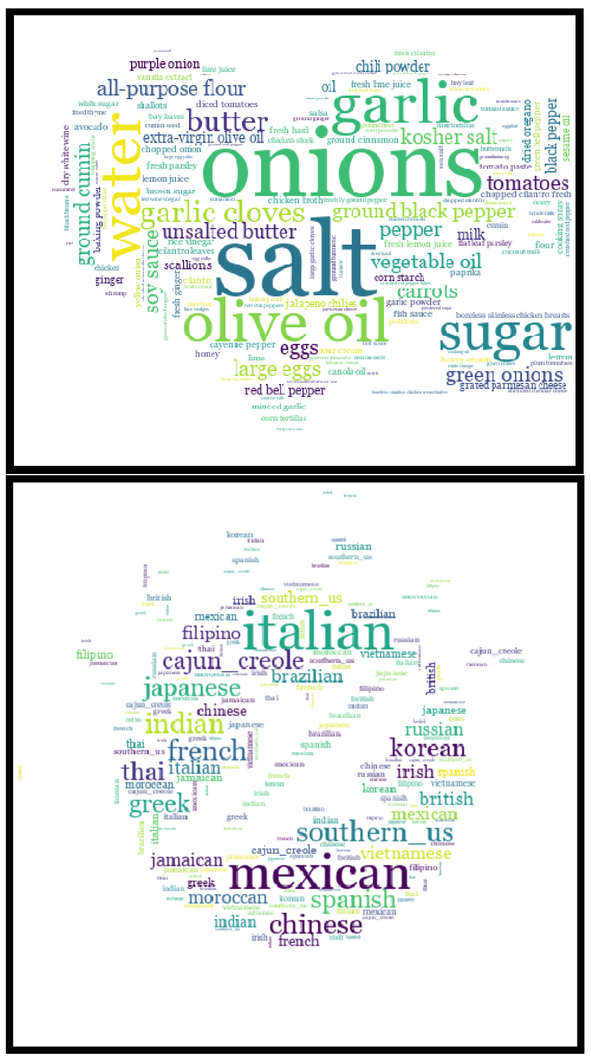
cuisine - Represents the country of each recipe

ID - item ID

ingredients - Recipes for each country.

Data Visualization

By using wordcloud to describe the data.And Observe the frequency of the text data.



Modeling

There are many machine learning methods for text classification. We have selected the following five methods:

- Logistic Regression
- KNN
- Random forest
- SVM
- CNN

The discribe of the model

Among them, in KNN, random forest, support vector machine, a grid search method is used to adjust the parameters. Among them, the kernel function of the support vector machine is LinearSVC.
In the convolutional neural network model, it is constructed as an embading layer, two convolutional base layers, and an output layer. It uses dropout technology. and batch normallization technology.

model score

Figure of left is the first prediction based on the model. Figure of center shows ten features with high feature importance. Figure of the right is a new prediction based on the model to get the results needed for this problem.

Logistic Regression 0.729

KNN 0.740

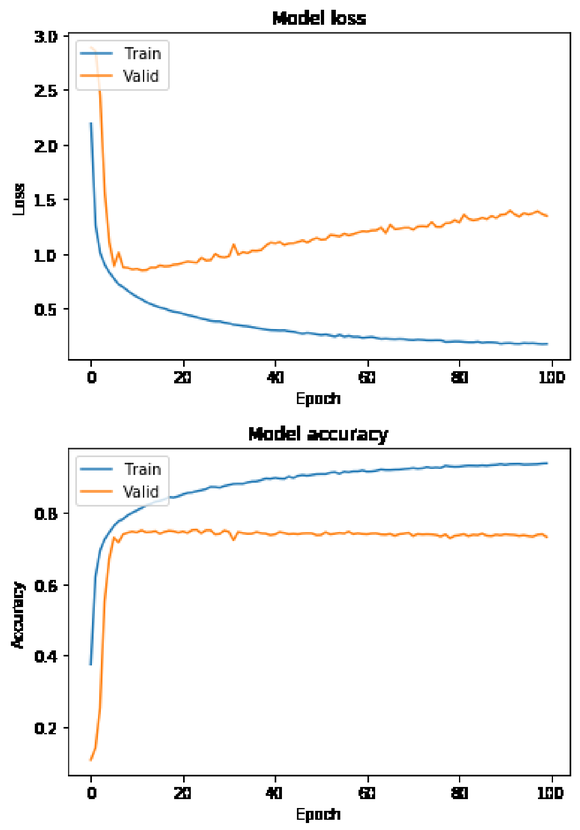
Random forest 0.739

SVM 0.736

CNN 0.753

CNN

It can be seen that the model gradually started to stabilize when iterating about 10 times.



Conclusion

1 Using the Word2vec to help us process the textdata.If the text data is Chinese, we can use jieba for word segmentation.

2 There are many ways to deal with text classification in machine learni... can select suitable ways on combination with the problem.

3 In this problem, i use the mean of each words v... vector. Maybe this is the question wh...

Acknowledgement

- Thank you