第一次实习项目书

一、实习目标

通过本次项目快速掌握分类任务的数据处理、模型训练和评估。

二、项目介绍

本次项目提供英文训练集与测试集，用于分类任务，数据格式为：

label3 rick denzien rick denzien is a songwriter singer and studio technician from buffalo new york .
label14 talking to strange men talking to strange men is a 1987 novel by british writer ruth rendell .
label6 porsche 550 the porsche 550 was a sports car produced by porsche from 1953-1956 .
label11 mendoncella mendoncella is a genus of flowering plants from the orchid family orchidaceae .
label5 gino matrundola gino matrundola (born july 21 1940) is a former politician in ontario canada .
......

上述例子中**label**3，表示文本rick denzien rick denzien is a songwriter singer and studio technician from buffalo new york所属的类别标签。这里无需知晓**label**3的具体含义，对最终结果无任何影响。

三、项目要求

算法设计

基于TensorFlow或Pytorch开发出fastText、CNN和LSTM算法模型。

效果评估

所有算法获得的准去率不低于85%，F1值不低于0.8。

代码提交

所有代码均提交至github（提前注册github），并写清楚README。

四、参考项目

- https://github.com/facebookresearch/fastText
- https://github.com/gaussic/text-classification-cnn-rnn