实习报告

1. 实习名称

构建聊天机器人

1. 实习目的

经历和实践一个项目的完整开发过程;

初步具备

1）应用软件或系统软件的设计与开发能力；

2）计算机算法的设计与优化能力；

3）运用主流计算机科学技术（如ML、DL、NLP 等）分析解决实际问题的能力。

培养团队合作能力和协作精神 (github);

1. 实习内容

基于特定语境的聊天机器人开发

通过机器学习和统计学方法，结合大数据分类，对社交群体等平台上的文本数

据进行获取和分析。在此基础上，针对特定交互文本场景，对用户语言分类并对

关键词进行抽取，探索如何有效地生成会话。

I 数据获取

II 数据预处理

III 数据标注

IV 文本分类与意图识别

V 实体关键词识别抽取

VI 生成对答

1. 实习过程、

**rasa的框架意图分类embedding算法**

**算法思想**

把训练样本和意图编码到同一个向量空间，设计损失函数，使得样本与真实意图更相近，样本与其他意图更相反，意图之间编码更相反，达到意图分类的目的。

举个例子说明，假设有两条训练样本“我要充话费”和“我要订机票”，有四个意图“订机票”、“查天气”，“充话费”，“查运势”，意图分类算法的目的就是把“我要充话费”分到“充话费”上，“我要订机票”分到“订机票”上。

样本与真实意图更相近就对应“我要充话费”和“充话费”向量相似度更高，同理，“我要订机票”和“订机票”向量相似度更高；

样本与其他意图更相反就对应“我要充话费”和“订机票”、“查天气”、“查运势”向量相似度更小；

意图之间编码相反就对应“订机票”、“查天气”，“充话费”，“查运势”向量相似度更小。

**网络结构**

1、样本输入x

输入层采用bag-of-word方式，只考虑词频，n\_words是词汇表词语总数

MLP有两个隐含层，采用relu激活函数，并配置 dropout和L2正则

2、意图输入y

输入层是意图数\*意图数的方阵，对角线元素为1，其余为0，方阵的第一行是样本的真实意图，剩下行数是其他意图（负样本）

3、embedding层

样本输入和意图输入的embedding层维度均为20

本次实习我负责在docker中部署整个聊天机器人的所有应用，以下是我的整个工作流程：

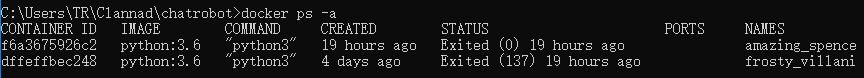
首先我们进行docker进行镜像拉取，我们所需要的镜像为：python:3.6

通过 docker pull python:3.6 命令拉取镜像

再在命令行输入命令 docker run -ti python:3.6

来运行当前拉取的镜像。

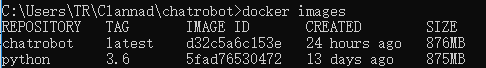
这时候我们得到一个新的容器，可以通过如下命令查看当前所有容器：docker ps -a



执行docker exec -it f6a3675926c2 /bin/bash进入容器

然后执行docker commit f6a3675926c2 chatrobot 将这个新容器制作成一个新的镜像

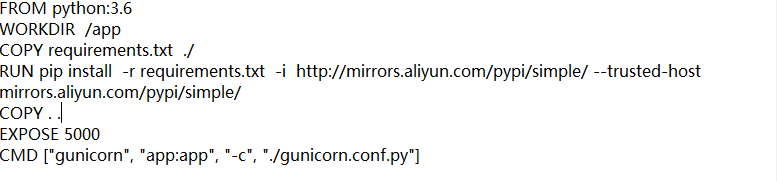
执行dockers images查看已经存在的镜像



可以看到新容器已经被我们制作成一个镜像

证明docker运行环境正常，制作的镜像能够正常运行。

后续我们编写dockerfile来制作新的镜像并部署我们的flask应用



在dockerfile文件目录下执行docker build -t chatrobot .制作新镜像

执行docker run -p 5000:5000 chatrobot来执行flask项目

然后通过网页来访问127.0.0.1:5000，来看能否访问成功。

1. 心得体会

在这次实习中，我们组主要是做一个基于rasa的关于星座的聊天机器人。我的主要工作是对于docker应用来部署相关的flask应用以及对于镜像的拉取使用。通过这么多天的实习，在实习的过程中不断摸索，攻克一个个困难，学习到了许多有用的知识，对于整个工作的流程有了一个大概的认识，还收获很多课外知识。我们相信，只要一直努力奋斗，不断学习，就能克服困难，收获成功。