**计算机学院生产实习报告**

|  |  |
| --- | --- |
| 专业名称 | 计算机科学与技术 |
| 实习题目 | 聊天机器人开发 |
| 组 名 | Clannad |
| 组 员 | 田丁辰 |
| 指导教师 | 邱晨 |

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

任务分配：

袁粼：负责使用flask在网页上布置rasa聊天机器人

林万超：负责收集语料，整理数据，训练模型

王凯：负责在Docker中部署整个聊天机器人的所有应用

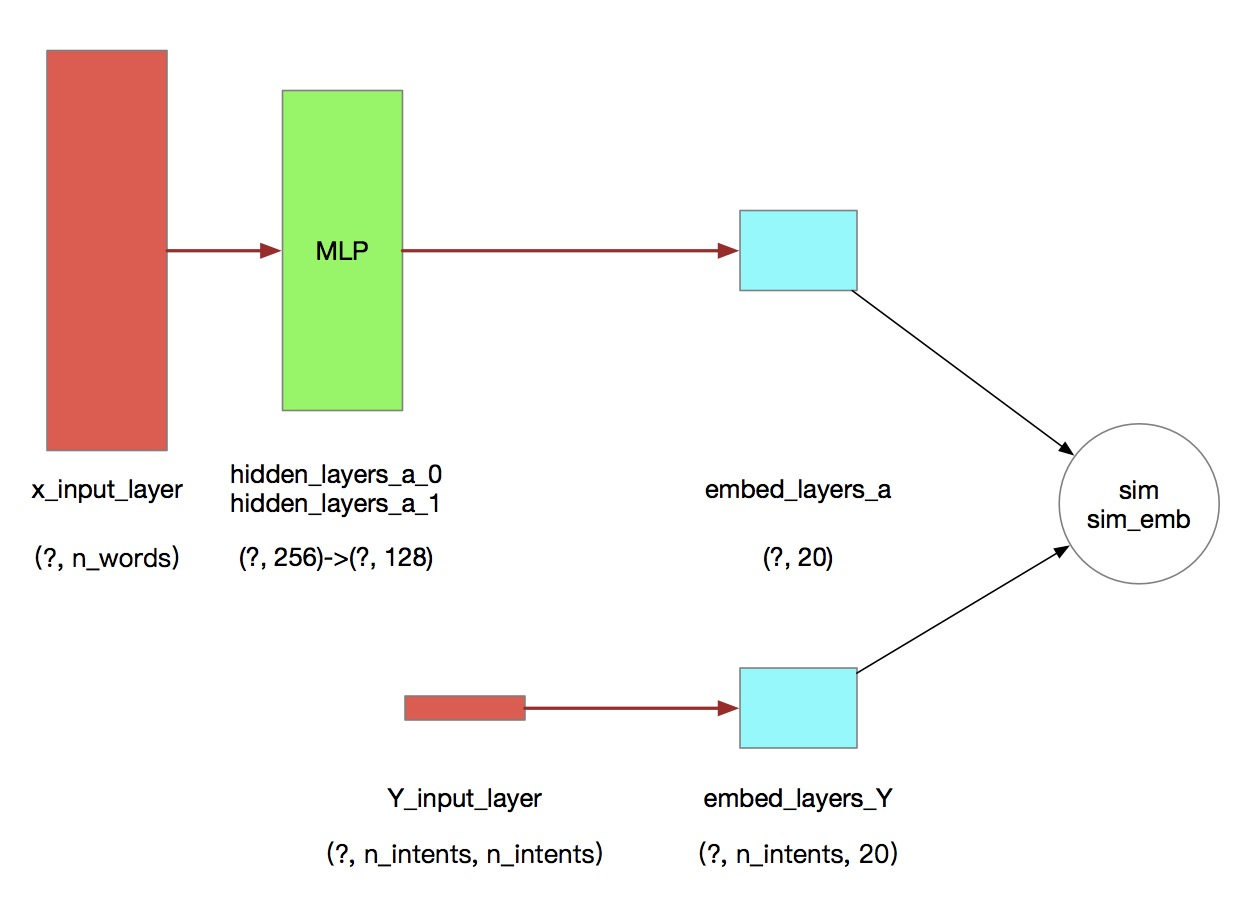
田丁辰：负责收集语料

李涵：负责收集语料

1. 实习过程、

**rasa的框架意图分类embedding算法**

**算法框架：**



**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

**相似度计算**  
向量相似度计算采用常见的余弦相似度。但是本算法计算两个相似度，一个是sim，即样本与意图之间的相似度；另一个是sim\_emb，即意图之间的相似度。

**损失函数**

损失函数定性理解，损失=样本与真实意图的相似度取反+样本与最相近的其他意图的相似度取正+真实意图与最相近的其他意图的相似度取正  
最小化损失相当于，训练过程中逐渐让样本的编码向量与真实意图的编码向量更相近，样本的编码向量与其他意图的编码向量更相反，不同意图之间的编码向量更相反。

**网页部分**

**使用方法**

Host Machine

Flask

Browser

$.get("/get", ...)

:5000

Rasa

localhost:5005

predict\_url → response → action/messages

messages\_url → send message to rasa

execute\_url → execute action

index.html

style.css

pictures

web-

api

web app结构

使用方法：

1. 打开rasa项目所在文件夹，在命令行运行：

python -m rasa run --enable-api

开启rasa服务。默认url = 127.0.0.1:5005。

1. 等待rasa开启后，打开app.py所在的文件夹，运行：

python app.py

开启flask服务。默认端口5000。

1. 现在可以通过浏览器访问聊天机器人了。

设计理念：星座

以个人按现代心理占星术的理解，十二星座是指人的十二种情感行为模式，人人都是小宇宙，每个人身上都会普遍存在十二星座的情感行为模式（本身包含诸多子人格再加上大运流年之类的影响），只是多少强弱隐露的问题，只是我们通常说的星座是太阳星座，太阳星座对我们的人格影响很大，所以也会显示的挺准的样子，实际上要复杂的多。你最后呈现出来的真实行为习惯其实是各种力量综合博弈的结果。

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。在实习中，我主要是负责相关语料的收集与对语料集的创建及编写。通过这次长达两周的实习，我对聊天机器人相关知识有了更加深刻的认识，对以后我的工作及学习有深刻影响。