**开课吧人工智能课学院**

**自然语言片与人工智能第5期 项目三报告书**

项目名：面向服务的对话机器人的构建

队名：冯・诺依曼队

在本项目中，我们实现了一个面向服务的对话机器人的构建及优化，并以WEB应用的形式编写了可以与用户进行交互的展示程序。

在实现项目目标的过程中，我们小组运用了多种从NLP课程中学习到的方法和技术，并根据本项目自身的特点，灵活地做了一些调整和优化，成功地达到了预期的目标。

但是，由于项目时间和规模有限，以及我们对自然语言处理技术还处在逐步的学习探索过程中，因此，实现的项目成果还有一些有待改进的地方。

以下，我们简单总结了一下我们小组的项目三成果的优点和不足之处。

**【优点】**

**1、选取的模型和算法简单可靠**

对于语言模型和算法，我们尝试过GPT2、BM25等多种，并在反复实验并观察评分结果和训练时间后，大致确定了最终使用的模型结构和算法。

选定模型和主要算法的过程如下：

1.1.关于QS分类 尝试过KNN模型，但效果较差。手动打标-监督学习。

1.2.关于分类效果 尝试了lightgbm和朴素贝叶斯，发现效果差不多。选用朴素贝叶斯，并尝试取同样数量的正负例问答样本提升分类效果。

1.3.关于业务（基于搜索的问答系统）尝试过如下：

boolen search 选10个candidates、BM25选top10 candidates、

query与候选的20个candidate通过取bert预训练句向量计算相似度取top1

但是结果显示速度较慢。

尝试过单独使用boolen search。结果显示，效果一般，但速度很快。

也尝试过只通过BM25计算出来的top1。这种情况效果更好，速度适中。

1.4. 关于闲聊类问题，尝试过使用GPT2来生成。

1.5. 关于添加默认QA ，尝试了正则匹配。

经过如上的反复尝试和测试，最终选定了当前使用的模型和算法。选定当前模型和算法的主要考虑是，他们的结构相对简单可靠，易于实现和修改，并且，训练时间和训练后实测的精确度都在可接收的范围内。对比其它模型结构，我们认为最终选定的模型结构是能在复杂度、精确度和训练时间三者之间取得平衡并达到系统整体最优的解决方案。因此，我们在这种模型结构的基础上，进行后续更进一步的调参和测试。

**2、训练块，迭代快，调整快**

对于本项目的任务场景来说，模型训练所需的时间相对较长是一个很大的挑战。对于需要反复多次进行的【分析现象 -> 提出假设 -> 改变模型做实验 -> 观察是否符合期望】这样的调参过程来说，长时间的模型训练使得调参的结果很难在短时间内得到反馈，从而使得整个模型的迭代效率大大降低。

为了从根本上提高模型的优化迭代速度，我们从关键性的“模型训练时间”着手，采取了多种方法来提高模型训练的效率，有效果地减少了在此环节上的时间浪费，并提高了训练结果的精确度。经过以上努力，我们将训练的时间控制在了合理的范围内，并且使得训练结果的有效性在很大程度上得到了保证。这就为我们后续反复调整参数，优化模型精确度的工作打下了基础。

**3、使用了爬虫功能实现问题的检索**

对于在QA库中匹配不到相似的问题的情况，我们使用爬虫技术，借助百度搜索，爬取相关问题的检索结果。这使得我们的对话机器人的使用范围进一步提高，对于QA库中没有现成答案的问题也可在一定程度上满足用户的查询需求。

**4、有效运用框架**

在将模型构建成WEB应用程序的过程中，我们使用了「Django」框架，使得前端和后端能在很短时间内搭建起来。这大大节省了搭建WEB应用所需时间精力，使得我们可以集中精力进行模型的持续改进工作中。并且，框架本身的功能结合互联网上的一些资源，使得前端的展示页面能够拥有漂亮简洁的UI界面。

**5、良好的交互效果**

为了提高用户的交互体验，我们设计了清晰友好的输入界面和简单明了操作方式。用户与对话机器人的互动是基于一个类似于常见聊天界面的UI下实现的，操作简单，输入输出的结果清晰直观。

**【有待改进的地方】**

1. **对于多种技术的综合使用还不够成熟**

本次问题比较综合，牵涉到的知识点比较多，涉及到NLP领域目前为止所学的大部分知识。对于将这些多种技术的综合来说，我们还有很多运用不熟练地地方。各个模块的输入输出如何更好更高效地衔接起来，今后可以作为我们的课题进行持续优化改善。

1. **性能问题。**

对于模型和算法的选取，目前还是在精度与性能之间做了折衷。运算的精度实际上还有提升的空间。今后考虑在保证运算精度的基础上，进一步提高精度。

以上关于本次项目三成果的分析报告。

谢谢各位老师的指导和辛勤付出。

队名：冯・诺依曼队

成员：李鹏飞、吕汶颖、刘迪

时间：2020年3月5日