软件需求工程LAB1

开源社区SmartIDE的需求分析分类

学校：南京大学

院系：计算机科学与技术系

姓名：许家铭、黄逸维、陈永健

学号：171860581、171860542、171860568

邮箱：[381026323@qq.com](mailto:381026323@qq.com)

1. 概述

本次实验是针对SmartIDE的需求分析分类工作。通过对开源社区的Smart IDE抓取他们开发和维护过程中的commit与issue，从中分析获取他们所对应的需求，并进行分类。由于我们三年级学生还没有学习和使用过自然语言处理（Normal language process）和机器学习（machine learning），我们选择了使用LUIS来进行自然语言处理和分类，尽我们所能的完成了本次实验。

1. 实验背景

在IT行业日益发达的现今，一个智能的IDE会成为很多程序员的得力助手，帮助他们迅速而简洁的完成任务。因此smartIDE在现今日益重要。而今天我们就将针对SmartIDE中的一个：开源社区中的Visual Studio Code进行需求分析与分类。

实验任务：

1. 数据获取

从GitHub的开源代码中，我们爬取了其中的issue和commit的数据，根据提交的描述作为我们分类所使用的原始数据。

1. 需求分析

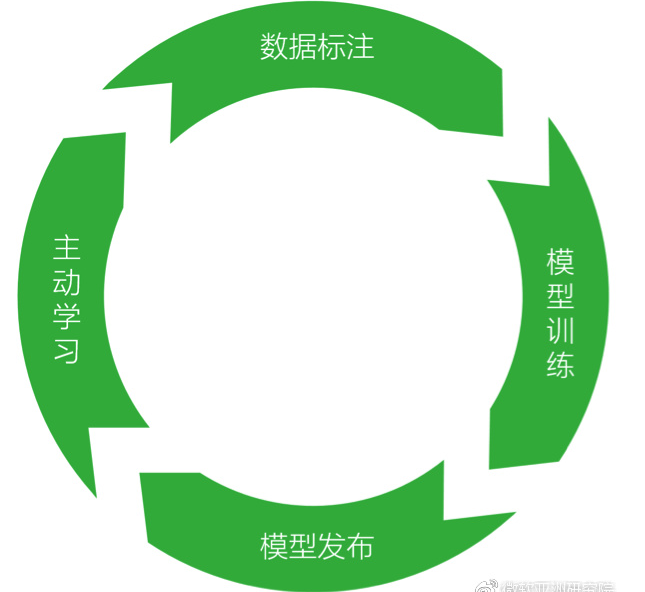
对抓取的数据进行分析，分析他们主要包括的种类，主要有哪些需求，从而可以找出对应的应该有的分类标准，建立起NLP所需要的语料库

1. 需求分类

根据之前建立起来的分类标准，对于所有的原始数据进行分类，放入他们所属的类别。

1. 名词解释
2. LUIS

LUIS (Language Understanding Intelligent Service，https://www.luis.ai) 是微软发布的面向开发者的自然语言语义理解模块开发服务。



LUIS开发流程首先是标注一定的起始数据，后训练得到语义理解模型，再对模型进行必要的测试，发布模型并应用到真实用户场景。根据应用日志继续标注更新模型，使其更加精确。

1. 分类标准

由于本次实验没有给出一个统一的分类标准或语料库，我们只能建立自己的分类标准，并且遵照其进行语料库的建立。

分类标准：通过研究原始数据中的需求，我们小组通过讨论制定了一套需求分类标准。

①.编辑功能：与编辑代码有关的所有功能，包括锁紧功能，代码补全等。

②.文件管理：有关文件与文件夹的功能，包括文件的输入输出，保存，二进制文件支持等

③.编译功能：对于各种语言环境的支持，支持对于程序的编译等功能

④.扩展功能：支持各类插件和扩展按钮功能，包括git支持，API支持等功能

⑤.界面设计：对于UI可视化等的支持，包括主题、字体、颜色等

⑥.软件稳定性：对于软件的稳定性的支持，如能够正确启动，不会出现失败或错误

语料库的建立：

本次实验中，为了使用NLP对数据进行处理，我们使用了自主建立的语料库。建立的方式主要是通过对于原始数据进行人工的贴标签分类。在小组合作后，将我们选出的语料库放入了在线文档中，这部分语料库我们也复制了一份，放在了文件夹中。

1. 实验方法

首先，我们使用Python编写了两份爬虫，分别爬取了vscode项目中的issue和commit。

其次，我们使用了腾讯在线文档，进行分工贴标签，并进行汇总，得到我们的语料库。

腾讯在线文档地址：

<https://docs.qq.com/sheet/DSE1EZEJxcG5NdUJk?tab=BB08J2&c=N6K0A0>

最后，我们将语料库输入LUIS，建立起NLP模型，并且训练，测试后发布，形成API供我们进行调用。由于微软限制每个用户每个月免费测试条数为1000条，我们将测试数据分割成五份进行了分类。

1. 实验过程

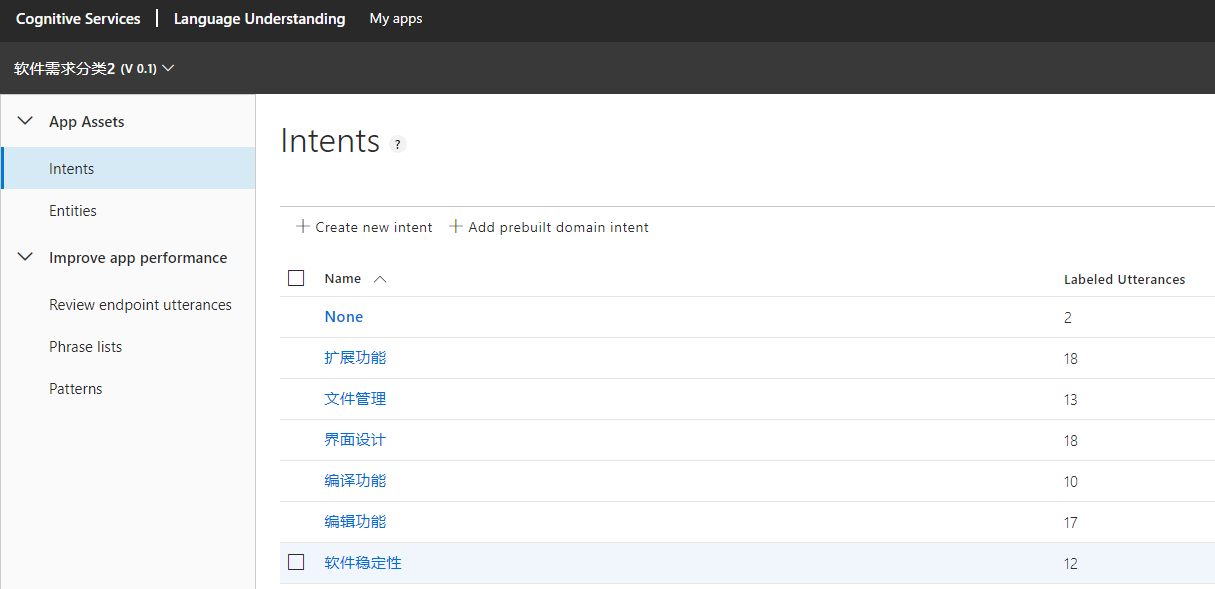
爬虫程序（~/ lab1\_vscode需求分类/数据抓取）：



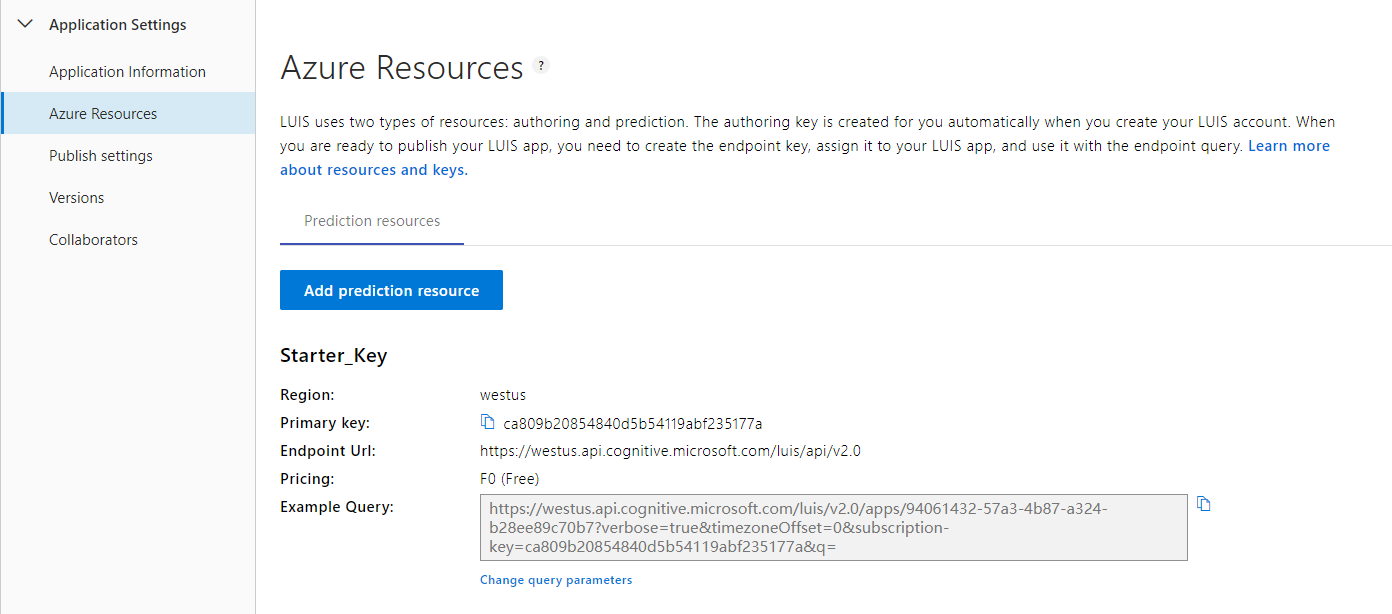
语料库建立：



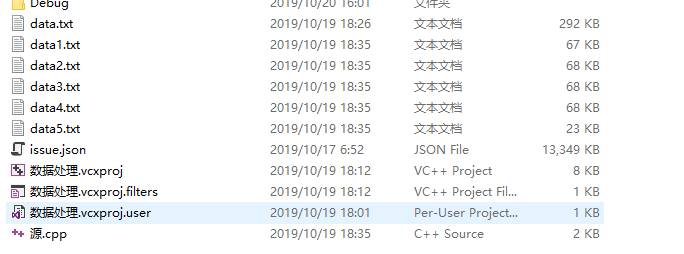
LUIS模型建立：



API的建立：



数据分割（~/ lab1\_vscode需求分类/数据处理）：



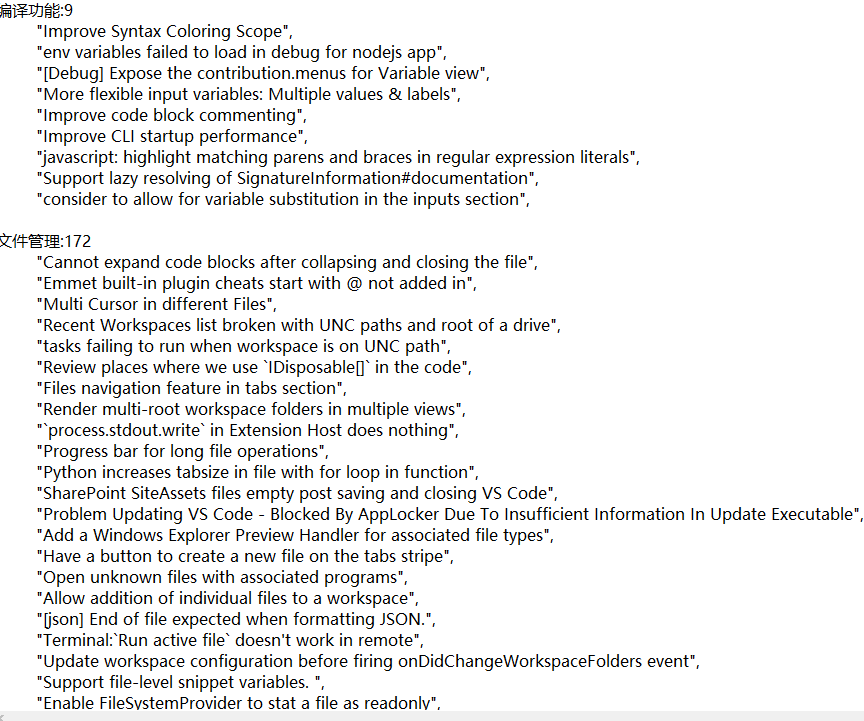
调用API的程序：



以上使用的所有程序均附在GitHub中

1. 实验结果

根据初步测试的结果来看，分类的精确度还是有一定的。结果示意图如下：



具体的结果我已经放入了~/ lab1\_vscode需求分类/数据存储/分类后数据中

1. 结果分析

从我们测试所得的数据来看，我们分类的数据虽然有一定的精确度，但还是和我们人工分类结果有一定区别。

这些误差主要来源分别如下：

首先最主要的是，我们的语料库不够完整和全面。由于我在网上没有能够找到完整的软件需求方面的语料库，我们只能自主建立一个语料库。

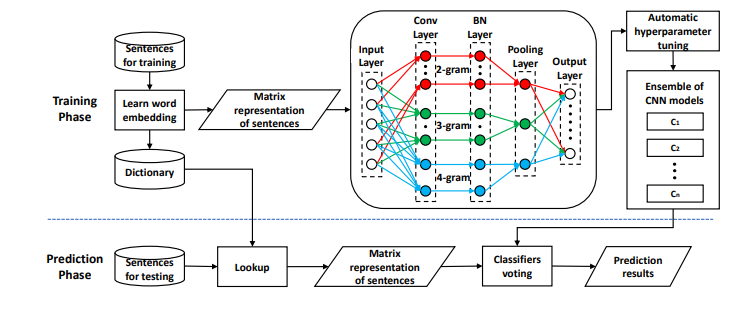
很显然，我们这种方法是不够的完整而精确的。我们的模型例句不够充足，由于我们没有学习过NLP和ML，确定我们实验的方法并且学习就用掉了很大一部分时间，我们没有能够建立起完整精确的模型。若要进行后续实验，我们可以通过增加例句的形式来使我们的模型更加精确。

并且我们使用的是每个人各自去找例句分类的方法，相当于只有一个人在进行label的工作，这显然会导致我们贴上的标签不够客观，会有很强的个人理解在其中。如果有更多时间，我们可以通过从中抽取一份样本，每个人给样本进行分类，然后对我们的分类结果进行汇总，如果出现分类结果不同的情况，我们可以通过讨论或者进行投票，请更多人进行评判，从而获得一份更加符合大众价值观的分类标准和语料库。

接着是我们的分类标准不够清晰，由于没有统一的规范，我们查询论文时，也没有找到特别适合使用的分类标准，因此我们只能通过对于我们抽选的样本进行了分类，并且我们的分类之间的界定也没能做到很清晰，这一点我们可以通过对于分类进行精确定义来实现。后续也有待改进。

1. 相关的实验

在查阅论文的时候，我注意到浙江大学的一位教授也在做这一方面的研究。他们使用的模型是这样的。



先对句子进行分割，再学习其中嵌入的词，通过矩阵计算其代表语句的权值从而算出该语句的最终属于的intension。由于其中使用的分类标准和模型方法过于复杂，我们也没能复现出来。

论文的地址如下：

<https://xin-xia.github.io/publication/tse185.pdf>

1. 结论

本次实验，我们使用了微软的NLP模块LUIS进行了自然语言处理与分类，取得的结果还是比较精确的，但是还是存在一定误差。在后续实验中，我们可以通过增加例句，界定分类标准，多人贴标签汇总等方式进行实验的细化。

在本次实验中，我们组的分数分配为每个人得分1/3。

虽然我们没有学过NLP和ML，但我们还是用我们自己的方式尽力地完成了实验。

1. 参考文献

Huang Q, Xia X, Lo D, et al. Automating intention mining[J]. IEEE Transactions on Software Engineering, 2018.

<https://docs.microsoft.com/zh-cn/azure/cognitive-services/luis/what-is-luis>