

实验一：语言模型

实验目的：

理解并实践语言模型

实验内容：

- 1: 用 python 编程实践语言模型 (uni-gram 和 bi-gram) ,加入平滑技术。
- 2: 计算 test.txt 中句子的 PPL, 对比 uni-gram 和 bi-gram 语言模型效果。

数据集：

train_LM.txt

test_LM.txt

Example: （每行数据是一段对话，句子间用__eou__分隔）

How much can I change 100 dollars for ? __eou__ What kind of currency do you want ? __eou__ How much will it be in Chinese currency ? __eou__ That's 680 Yuan .
__eou__

What kind of account do you prefer ? Checking account or savings account ?
__eou__ I would like to open a checking account . __eou__ Ok , please just fill out this form and show us your ID card . __eou__ Here you are . __eou__

实验指导：

Python 教程：

<https://www.liaoxuefeng.com/wiki/0014316089557264a6b348958f449949df42a6d3a2e542c000>

python 以及 pycharm 安装：

<http://www.runoob.com/w3cnote/pycharm-windows-install.html>

分词工具：

nlTK.tokenize.word_tokenize