Kenlm使用

Kenlm是一个训练n-gram模型的工具

工具包的下载地址：<http://kheafield.com/code/kenlm.tar.gz>

编译的命令如下：

mkdir kenlm/build  
cd kenlm/build  
cmake ..  
make -j2

成功编译后进入bin文件夹下

运行./lmplz -o 2 -S 20% --verbose\_header --discount\_fallback --text test.txt --arpa test.arpa

-o 参数为 n- gram 中的n

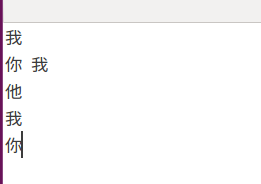
-S 为使用的最大内存

--verbose\_header 在生成的文件头位置加上统计信息

-- text 为训练的文本

--arpa 为生成的模型文件

例如测试文件如下



得到的 arpa文件结果为

