问题：

1. 2个train函数 一个定义在Sentiment类里，一个是文件里的 作用域不同
2. Marshal 二进制序列化-模型存在硬盘上 定义的一种格式

.dump –对象，.load-存入文件

1. ret, prob命名统一
2. from .sim import bm25.py \_\_init\_\_.py没东西

from . import tag \_\_init\_\_.py里有东西

1. setup.py这个库进行安装时会用到，下载源码python setup.py --install
2. test.py:

测试如何调用：

方法一

from snownlp import normal

sents=normal.get\_sentence(text)

方法二

from snownlp import snowNLP

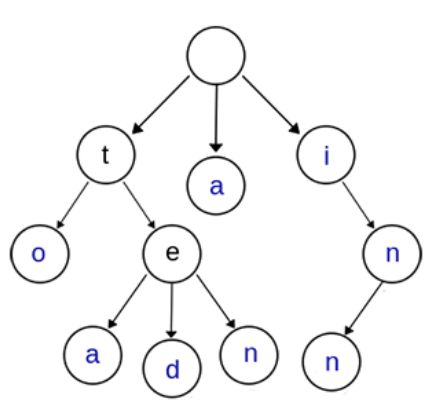
snowNLP(text).sentences

1. re\_zh=re.compile(‘([u4E00-\u9FA5]+)’)

re\_zh.split(sent) 按照前面正则表达式匹配的内容来切分，把不是汉字的扔出去

1. 文字转拼音

对trie树的遍历-逐个根节点往下找



1. 文本分类-朴素贝叶斯

log() –可以把太小的结果放大

问题：tnt算法是什么原理？？？