### 心得体会

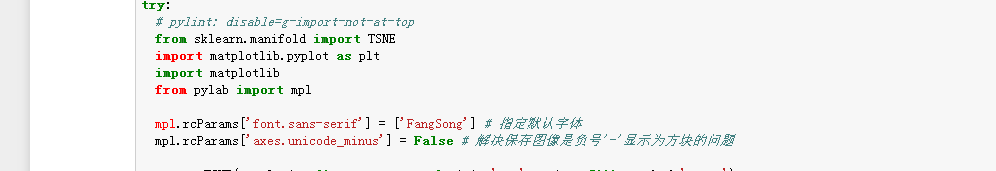
本次的任务：使用tensorflow中的rnn相关操作，以作业提供的《全宋词》为训练数据，训练一个人工智能写词机。

分俩部份，一个是使用embedding 对全宋词进行词嵌入处理得到：如下结果



以及embedding.npy文件

在使用embedding 的时候主要遇到如下困难：matplotlib 中文显示乱码问题，解决方案如下：



对其进行设置默认字体；

Embedding:是词嵌入的意思，在自然语言处理中，如果使用的one\_hot 编码的方式去处理那么如果词是10个就是10维空间词有6000个就是6k维的空间；在表示一个字的时候其中一列是1其余都是0，这样对于计算机效率来说实在太慢了；而embedding解决了这个问题它可以使用词嵌入的方式让词与词之间产生关联关系，并进行降维，比如说6000个词全部放在一个128维度上来进行向量表示，然后向量相近的词的关联性也就大，这样可以使计算效率大大提高，天然有聚类后的效果，一个向量可以编码一词多义。

通俗的翻译可以认为就是把X所属空间的单词映射为到Y空间的多维向量，那么该多维向量相当于嵌入到Y所属空间中，一个萝卜一个坑。

另一个任务就是使用rnn模型配合embedding 来进行训练：