### 心得体会

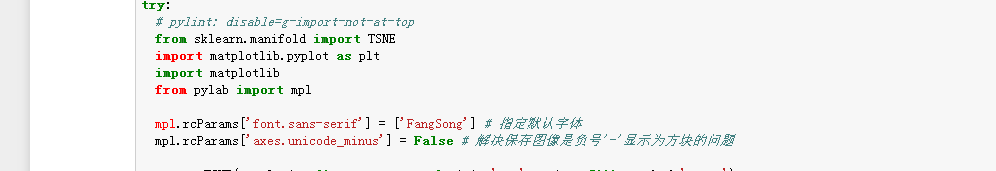
本次的任务：使用tensorflow中的rnn相关操作，以作业提供的《全宋词》为训练数据，训练一个人工智能写词机。

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：

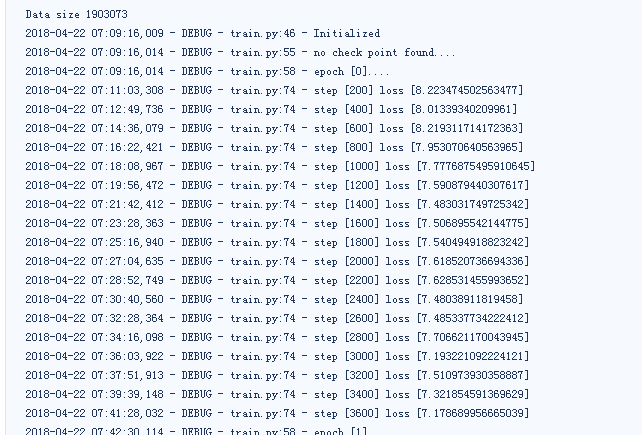


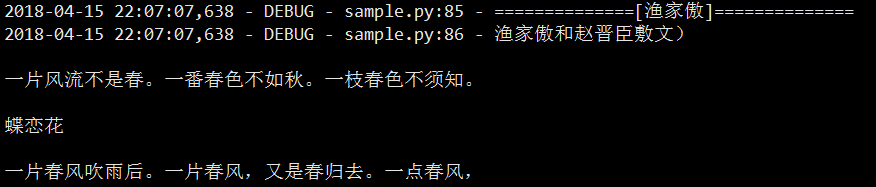
对其进行设置默认字体；

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。

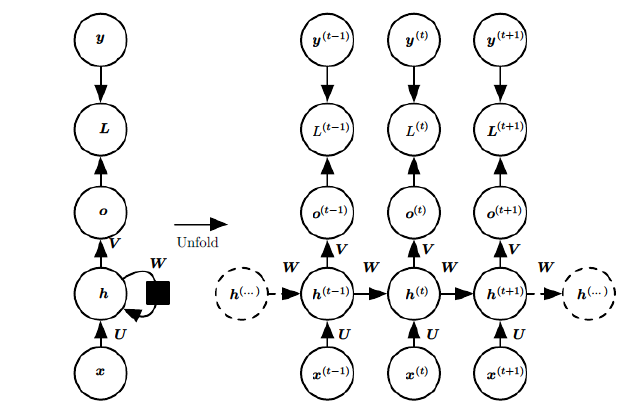
通俗的翻译可以认为就是把X所属空间的单词映射为到Y空间的多维向量，那么该多维向量相当于嵌入到Y所属空间中，一个萝卜一个坑。

另一个任务就是使用rnn模型配合embedding 来进行训练：





RNN :主要用于自然语言处理，rnn的主要流程如下



上图中左边是RNN模型没有按时间展开的图，如果按时间序列展开，则是上图中的右边部分。我们重点观察右边部分的图。

　　　　这幅图描述了在序列索引号tt附近RNN的模型。其中：

　　　　1）x(t)x(t)代表在序列索引号tt时训练样本的输入。同样的，x(t−1)x(t−1)和x(t+1)x(t+1)代表在序列索引号t−1t−1和t+1t+1时训练样本的输入。

　　　　2）h(t)h(t)代表在序列索引号tt时模型的隐藏状态。h(t)h(t)由x(t)x(t)和h(t−1)h(t−1)共同决定。

　　　　3）o(t)o(t)代表在序列索引号tt时模型的输出。o(t)o(t)只由模型当前的隐藏状态h(t)h(t)决定。

　　　　4）L(t)L(t)代表在序列索引号tt时模型的损失函数。

　　　　5）y(t)y(t)代表在序列索引号tt时训练样本序列的真实输出。

　　　　6）U,W,VU,W,V这三个矩阵是我们的模型的线性关系参数，它在整个RNN网络中是共享的，这点和DNN很不相同。 也正因为是共享了，它体现了RNN的模型的“循环反馈”的思想。

总结：

本次作业是最后的一次作业 也是耗时最长的作业，怎么说呢！感觉自己很乱；时常弄不明白 感觉自己脑袋不够用了！ 花了这么长时间让我真的有点失落；还是期待自己放宽心态吧！好好把最后的项目做好吧