自然语言处理及推荐系统测试题（每题10分）

1. 给出一句有歧义的汉语句子，并给出各自的分词方式。
2. Word2vec把某个语言里的单词映射成高维向量,例如这样：

W(‘‘cat”)=(0.2, -0.4, 0.7, …)

W(‘‘mat”)=(0.0, 0.6, -0.1, …)

…

问cat和hat，cat和pet哪一个距离更近，为什么？

1. 请用5-10句话简单介绍主题模型。
2. LSTM解决了之前RNN的什么问题？
3. 把稀疏的特征向量嵌入成为隐变量在机器学习中有什么好处？
4. 在中文中，常见的错误输入是同音不同字：例如，"苹果"输错为"平果";在英文中，常见的错误输入时拼写错误，如"latest"错输为"latst"。针对以上两种在中文和英文输入中的错误，请分别给出一种课上讲过的解决方案。
5. 计算下面两个向量的余弦相似度：

(0, 1, 1, 0, 1, 0, 1) 及 (1, 0, 1, 0, 1, 0, 1)

1. 比较基于内容和协同过滤的推荐系统各有什么优缺点。
2. 讲一讲TF-IDF的基本思想。
3. 如何解决协同过滤推荐系统中的冷启动问题？