

作业要求

- 每人完成总分值不少于100分的候选题目
- 使用的方法模型不限、编程语言不限
- 要求提交物：
 - 说明文档：任务定义、输入输出、方法描述、结果分析（性能评价）、源码运行环境
 - 代码：源码及可执行文件
- 提交方式：
 - 邮箱：yuansassignment@163.com
 - 邮件及附件命名格式：学号-姓名-作业题目（e. g. , 20190310-张三-语言模型）
- 最晚提交时间：课程结束后的一周内

作业说明

- 关于分组：
 - 原则上，每人独立完成至少60分值的作业
 - 对于一些题目（例如信息检索、机器翻译、推荐系统等），工作量较大同时可以清晰地分成几个功能模块进行分别建模的情形，可以分工完成，但需要在文档中特别说明各自的分工、各自的功能模块实现过程
- 关于加分：
 - 如果对于一个题目提供了不同的解决方案，或在一个解决方案之上提供了改进方案，则可额外加最多10分，具体根据完成情况确定
 - 作业演示：演示者通过slides向大家介绍自己的某一个或几个作业，根据具体情况，演示者则可额外加最多10分，演示需提前一周申请

Problem 9

- Sentence matching: 60 points
 - In this assignment, you will build a model to compute the similarity of two sentences. You will consider both the problem of sentence meaning representation and the problem of similarity computing with that representation.
 - The data is from a clinical record. The model you build should make medical diagnosis through comparing the similarity of a new symptom with known symptoms. You can divide the data into the training data, the development data, and the blind test data as required.

Problem 10

- Text generation: 40 points
 - In this assignment, you will build a model to generate text (news, poems, couplets and so on). Input, for example 5 words for start, the model will output the desired word sequences.
 - You can use an encoder-decoder architecture and attention mechanism for generating appropriate text. For bottom encoding, you can use an RNN or a Transformer encoder.

Problem 11

- Pretrained Transformer models: 60 points
 - In this portion of the assignment, you'll learn how to pretrain a Transformer on Wikipedia text that contains world knowledge, and find that finetuning that Transformer on the same knowledge-intensive task enables the model to access some of the knowledge learned at pretraining time.
 - You'll train a Transformer model to attempt to answer simple questions of the form "Where was person [x] born?" – without providing any input text from which to draw the answer.
 - You'll find that models are able to learn some facts about where people were born through pretraining, and access that information during fine-tuning to answer the questions.
 - Then, you'll take a harder look at the system you built, and reason about the implications and concerns about relying on such implicit pretrained knowledge.

Problem 12

- Chatbot: 60 points
 - In this assignment, you will build a chatbot that can converse with you naturally using natural language.
 - You can build a chatbot through a retrieval model by comparing the similarity of user utterance with the collected dialogues.
 - You can also build a chatbot through a generation model using sequence-to-sequence model like RNN, LSTM, etc.
 - You are also provided movie knowledge triplets, so that you can build a chatbot with background knowledge.

- Thank you!