**《智能算法及应用》课程作业**

**作业一：**模拟退火和遗传算法

在TSPLIB（http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/，多个地址有备份；其他网站还可以找到有趣的art TSP和national TSP）中选一个大于100个城市数的TSP问题，

1. 采用多种邻域操作的局部搜索local search策略求解；
2. 在局部搜索策略的基础上，加入模拟退火simulated annealing策略，并比较两者的效果；

3. 要求求得的解不要超过最优值的10％，并能够提供可视化图形界面，观察路径的变化和交叉程度。

用遗传算法求解TSP问题（问题规模等和模拟退火求解TSP实验同），要求：

1.设计较好的交叉操作，并且引入多种局部搜索操作（可替换通常遗传算法的变异操作）

2.和之前的模拟退火算法（采用相同的局部搜索操作）进行比较

3.得出设计高效遗传算法的一些经验，并比较单点搜索和多点搜索的优缺点。

**作业二：用深度学习做文本情感分类**

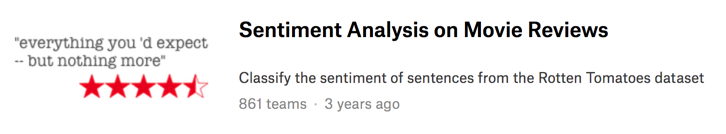
本次作业是Kaggle 上的一个影评文本情感分析的赛题，属于5分类的文本分类任务。文本情感分析是指对文本所表达的情感作情感极性的分类，本题将文本分类为消极、有点消极、中性、有点积极、积极5个类别。

请大家按要求注册账号，运用所学知识独立完成。 交作业的时候需要交一份报告(PDF 格式)、代码和最终版本的结果文件。

注意:我们鼓励互相讨论，但是作业必须独立完成，不允许抄袭，更不允许直接使用他人的运行结果交到 Kaggle 上。

**题目：Sentiment Analysis on Movie Reviews**

**网址：<https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews>**

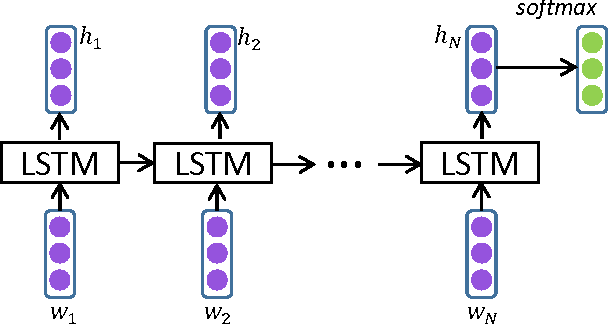


**作业要求 ：**

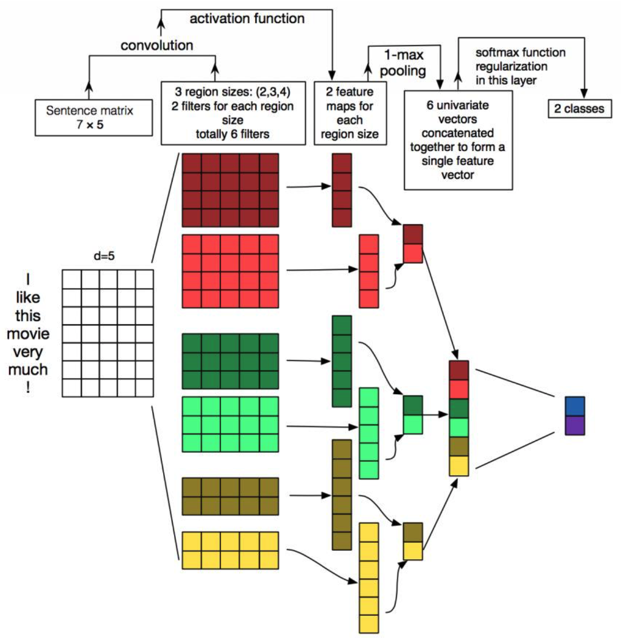
* 使用深度学习方法，编程语言不限，工具平台不限。推荐使用 PyTorch 或者 Keras 来实现。
* 禁止直接抄袭网上代码，禁止直接使用他人的结果文件交到 Kaggle 上。
* 作业包含一份报告(PDF 格式)、代码和最终版本的结果文件。
* 报告里面必须包含算法详细说明、Leaderboard 的分数以及排名。

**参考实现方案主要流程：**

1. **分词（本赛题的英文数据集已经作好了分词，跳过）**
2. **构建词汇表：所有文本语料中，每个唯一的词都对应一个索引id**
3. **序列化：将文本转化为整数序列，即词汇表索引的序列**
4. **向量化：将整数序列转为词向量序列。推荐使用预训练好的词向量,也可以使用可训练的embedding层，也可以直接使用one\_hot编码**
5. **输入RNN或CNN等深度模型，最后softmax分类**
6. 循环神经网络（RNN）



1. 卷积神经网络（CNN）



**推荐阅读的论文综述：**

* Zhang, Lei, Shuai Wang, and Bing Liu. "Deep learning for sentiment analysis: A survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (2018): e1253.
* Young, Tom, et al. "Recent trends in deep learning based natural language processing." *arXiv preprint arXiv:1708.02709*(2017).

**智能算法与应用作者报告**

――――具体实验题目

姓名与学号： 联系方式： 日期：

**摘要：**简要介绍要解决的问题，所使用的方法步骤，取得的结果或结论。

**1．导言**

要解决的问题描述，问题背景介绍；

拟使用的方法，方法的背景介绍；

**2．实验过程**

所用的具体的算法思想流程；

实现算法的程序主要流程，功能说明；

**3．结果分析**

交代实验环境，算法设计设计的参数说明；

结果（图或表格），比如在若干次运行后所得的最好解，最差解，平均值，标准差。

分析算法的性能，包括解的精度，算法的速度，或者与其他算法的对比分析。

算法的优缺点；本实验的不足之处，进一步改进的设想。

**4．结论**

简要结论或者对智能算法的体会。

**主要参考文献**(如有，三五个即可)

**实验评分说明**

为了更好地规范课程实验，达到课程实验的目标，现公布课程实验的要求如下：

1. 所有实验成果必须原创。允许提交半成品、失败品但绝不允许提交复制品。
2. 实验程序对正确的输入有正确的输出。
3. 提交实验结果，完整齐全。
4. 源代码编写符合编码规范，可读性高。
5. 源代码逻辑清晰。
6. 程序具有鲁棒性。
7. 界面友好。

第1—3是基本要求，满分70分。

第4—7属于加分项目，每达到一个要求加10分，总分最高100分。

注意事项：

1. 需要提交的实验结果包括：实验报告，源代码，编译得到的文件，说明文档，测试用例等。
2. 请避免将源代码直接复制到实验报告中，在说明算法思想和流程方面源代码并不比流程图和伪代码优胜。
3. 提交的源代码不应夹杂任何IDE的项目文件，如VS的\*.sln。如果使用JAVA则应当包含编译和运行用的批处理文件。
4. 说明文档内容为程序的运行说明，如输入输出等。
5. 需要提交至少两个测试用例，明确指出输入和期望的输出。
6. 小组完成的实验必须在源代码、实验报告中明确指出各人负责的部分，用于个人评分。
7. 实验结果通过FTP提交。要求使用压缩包，命名规则为：学号+姓名+个人实验编号，如000000001\_张三\_实验1，小组实验须注明组长姓名，如000000001\_张三（小组）\_实验1。目录组织如图1所示：

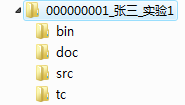


图1 目录组织

..\ 放置说明文件(readme.txt)以及批处理文件等。

..\src\ 放置源代码；

..\bin\ 放置编译得到的文件；

..\tc\ 放置测试用例；

..\doc\ 放置实验报告。

1. 源代码编写规范可参考SUN的编程规范<http://java.sun.com/docs/codeconv/>。
2. 界面友好并不等于要做图形界面，关键在于符合使用者的习惯，操作方便，表达清晰。

实验的评分分两部分：课堂演示和实验报告检查。课堂演示利用实验课进行，并完成绝大部分的评分工作，不参与课堂演示将直接影响实验成绩，请各位同学务必到场，抓紧时间。建议在完成课堂演示后再提交实验报告。

**最后重申，不允许抄袭。只要能体现个人的实验工作，即使是半成品、失败品同样会得到分数，但抄袭将会被判零分并上报学校！**