#### 中文 U97825

# 决赛

### 答题时间 2023-05-03 10:00:00 ~ 2023-05-28 18:00:00

#### 考试须知:

请各位选手务必严格按照要求进行答题,否则带来的失分后果自负。

- 1. 除题目特殊要求外, 文档应为PDF格式, 且满足以下要求:
- (1) 在文档中的任何位置不得透露个人和队伍信息, 否则按作弊处理;
- (2) 文档内容应包括:
- i. 摘要
- ii. 引言与背景介绍
- iii. 主要结果介绍
- iv. 主要方法与原理
- v. 结论和展望 vi. 参考文献 附录内容,如运行结果、其他数据等

### 2. 代码要求:

- (1)应在IDE中给定模板内作答,不得修改入口函数名、函数输入和函数输出,否则会有无法正常阅卷的风险。
- (2)版本要求:决赛作答IDE的python为3.8版本,pyqpanda版本为3.7.16版本。pychemiq 1.0.1版本,pyvqnet版本为2.0.6版本。
- (3)题目中,量子线路部分必须使用pyqpanda,禁止使用qiskit、cirq等其他量子编程模块。
- (4)除一些常见模块(如numpy、scipy、pandas、matplotlib等科学计算模块,以及xlrd、xlwt等文件处理模块)和题目特殊要求的模
- 块外,原则上禁止使用其他第三方模块。若有必要,则需要在文档中详细说明其必要性和应用范围,并自行承担无法正常阅卷的风险。
- (5)代码应具有较好的可读性,对核心函数的作用、输入、输出进行注释和说明。

### 一、算法题

#### 1. 行业背景

情感分类任务 (Sentiment Classification): 文本情感分类任务是NLP众多下游任务的一种,通过深度学习模型来提取文本情感特 征来达到对文本情感类型进行分类的目的。其可以是复杂的判别情感极性的多标签分类任务或是回归任务。 问题背景

情感分析旨在自动识别和提取文本中的倾向、立场、评价、观点等主观信息。传统的情感分析包含各式各样的任务,比如句子级 情感分类、评价对象级情感分类、评论观点抽取、情绪分类等。情感分析是人工智能的重要研究方向,具有很高的学术和应用价 值。同时,情感分析在消费决策、舆情分析、个性化推荐等领域均有重要的应用,具有很高的商业价值。

基于深度学习模型的分类方法因其强大的特征学习能力,逐渐成为解决文本情感分类问题的一种重要方法。量子机器学习是量子 计算和机器学习相结合的全新研究领域,相对于在经典机器学习算法而言,不仅可以利用量子态的可叠加性,在不增加硬件的基础 上实现并行计算, 对完成同样功能的经典机器学习算法达到二次甚至指数加速; 此外可以将经典数据编码为量子数据, 利用量子并 行性进行存储, 实现指数级节省存储硬件需求。其中, 变分量子线路是一种常用的量子机器学习算法。

变分量子解决机器学习任务包含如下步骤:

- (1)设计量子线路进行初态制备;
- (2)设计一组由可优化的参数集合heta构成的量子线路U( heta),一般可使用量子逻辑门构成;
- (3) 获取线路的某观测量H的测量值。例如初态为0的变分量子线路可由下式表示:

 $f(\theta) = \langle 0 | U^{\dagger}(\theta) H U(\theta) | 0 \rangle$ 

(4) 根据任务的目标构建损失函数进行优化。例如对于数据预测任务,若采用均方根误差(Mean Squared Error ) 作为损失, 问题 转化为找到最优的参数集合heta使得Loss( heta)最小:

 $rg \min_{ heta} rac{1}{N} \sum_{n=1}^{N} \left( \hat{y}^n - f( heta) 
ight)^2$ 

#### 赛题设计

### 1. 具体任务:

对于给定的文本d,系统需要根据文本的内容,给出其对应的情感类别s,类别s取值可以是"喜悦"、"愤怒"、"厌恶"、"低落"的离 散值0,1,2,3、表示情感类别。数据集中每个样本是一个二元组<d, s>、样例如下:

**–输入文本(d)**: 西湖雷峰塔标准旅游留念照,哈哈,人美就是硬道理据说此图片是珍藏版

**-情感类别(s)**: 喜悦

### 2. 重点关注:

(1) 比赛提供训练集和测试集,选手应该充分利用训练集的数据,合理划分数据并检验效果。

(2)选手需要保证结果复现的稳定性。选手需要在提交文档中详细阐述求解思路以及所用算法描述,需清楚分别注明量子计算 和经典计算部分,并给出该算法在测试数据上的测试结果。

(3)必须使用到量子线路,最好是纯量子线路解决问题。算法文档中需要突出算法创新性,如果是参考前人工作,需要给出相 关参考文献。同时以压缩包的形式上传相应完整可输出代码。其中量子计算部分的算法需要使用QPanda或pyqpanda设计量子线 路, 机器学习部分必须使用VQNet协助进行量子线路的参数优化。

(4)要求设计线路需要有可扩展性,对不同的情感分析数据集具有较好的鲁棒性。

(5)选手应在question函数中实现对验证数据集的预测。该函数的模板已经给好,函数接收一个string作为输入,该string为验 证集的名字,利用pandas打开该验证集,并利用你的模型对结果进行预测,输出对应情感类别(numpy ndarray格式,分别代表 对每个样本的情感类别预测结果离散值0,1,2,3)。注意,函数名称、输入格式、输出格式、输出的顺序皆不可改变,否则 带来的扣分损失自负。

(6)IDE中的代码应完整、可运行。选手可以在本地搭建对应环境进行训练,在IDE的对应文件位置上传训练好的模型。预测推 理部分的代码将会在系统服务器运行和检测,服务器规格为12核CPU,内存16G。在该配置下,总运行时间不得超过15分钟, 否则以无法运行处理。

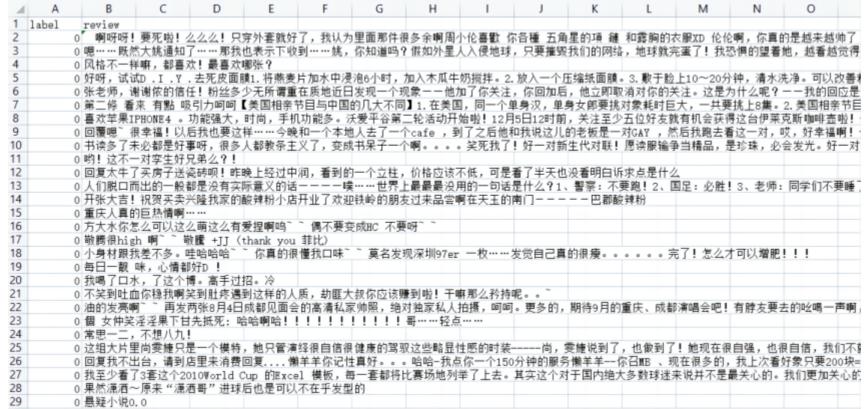
(7)最终将会根据在验证集上的预测结果的F1分数、变分量子线路参数占比、算法创新性和文档质量打分。 数据简介

## **simplifyweibo\_4\_moods数据集**: train.csv, test.csv和某验证集。

(1) 数据集简介

该数据及包括36 万多条,新浪微博情感分析数据集,包含 4 种情感,其中喜悦约 20 万条,愤怒、厌恶、低落各约 5 万条。本 比赛使用的是数据集的部分,四种情感数据数量相等,train.csv有1600条数据集,test.csv有400条数据集。

(2)数据集 数据集形式为\*.csv形式。其中label代表标签:分为0,1,2,3。对应有四种不同的情感。review代表对应的文本。



## 评分标准

该问题是探索性机器学习问题,没有固定答案。大赛将根据选手设计的量子机器学习算法原理、代码实现复杂度以及测试集上性 能指标综合评分,按以下三项进行评分,第一项占60分,第二项占20分,第三项占20分,总分满分为100分。

1.能够利用经典-量子混合算法使得4种情感数据的预测F1分数分别达到喜悦: 100%,愤怒: 100%,厌恶: 100%,低落:

100%,即得60分,4项指标每项15分为满分,按F1分数进行得分。例如:喜悦预测F1分数为0.5,则得分为15\*0.5=7.5分,以 此类推; 2.根据提交的算法代码,评定量子线路训练参数量占模型总参数量的比重进行得分,本条满分20分。例如:量子线路训练参数量

占比为30%,则该项得分为20\*30%=6分;

3.算法文档详细介绍和代码详细介绍,该项满分20分,评委对文档和代码质量进行综合评判进行打分。 注:

(1) 未使用量子算法不得分。

(2) 评价指标: F1, 查准率Precision, 查全率Recall

 $F1 = \frac{2 \times P \times R}{P+R}$ , Precision  $= \frac{TP}{TP+FP}$ , Recall  $= \frac{TP}{TP+FN}$ 

(3)本次比赛不允许选手自主使用任何外部数据集和预训练模型辅助训练比赛模型,如果抽查发现参赛队伍有造假、作弊、雷 同等行为,将取消该队伍的参赛资格及奖励。

(4) 选手禁止利用规则漏洞或技术漏洞等不良途径提高成绩排名,经发现将取消比赛成绩并严肃处理。其中划分好的数据集,测 试集只允许做预测推理,不允许以任意形式(有监督、半监督、自监督)在模型训练过程中使用。 参考文献

[1] 李胜旺, 杨艺, 许云峰, 等. 文本方面级情感分类方法综述[J]. Journal of Hebei University of Science & Technology, 2020, 41(6).

[2] 王婷, 杨文忠. 文本情感分析方法研究综述[J]. 计算机工程与应用, 2021, 57(12): 11-24.

[3] 王颖洁, 朱久祺, 汪祖民, 等. 自然语言处理在文本情感分析领域应用综述[J]. 计算机应用, 2022, 42(4): 1011.

[4] Li G, Zhao X, Wang X. Quantum Self-Attention Neural Networks for Text Classification[J]. arXiv preprint arXiv:2205.05625, 2022.

IDE作答

上传文档

编写代码

选择文件

https://contest.originqc.com.cn/contest/preliminary?contestTrackId=4&examId=10

一、算法题 1 未答 🗌 已答 临时保存 提交试卷

2023/5/12 10:48 本源量子云大赛平台

> 大赛平台 首页 大赛列表 赛事动态 学习 中文

> > 一、算法题 1 未答 🗌 已答

U97825



联系方式	支持与服务	快速链接	加入我们
0551-63836038	工单管理	本源量子官网	公司招聘
0551-63836039	帮助文档	关于本源	
oqc@originqc.com		本源量子学习机	



Copyright ©2017-2022 合肥本源量子计算科技有限责任公司 版权所有 皖ICP备 17025557号-1 皖公网安备 34011102000802号 《隐私政策》