CCKS 2022 技术评测任务书

教科书示意图问答任务

## 任务描述

**示意图是一种高度抽象的知识表达载体，常由矩形、圆形等几何形状和箭头、折线等逻辑符号组成，广泛被应用于教科书、百科、知识博客等教育场景。在上述场景中，学习者常通过视觉问答的形式来判断对知识点的掌握情况，即：给定学习者一张示意图，根据视觉信息来回答与该知识点相关的自然语言问题，我们将其定义为教科书示意图问答任务，任务样例如图1所示（实际评测中会提供标注信息）。该任务可能涉及视觉对象的检测与对齐、知识术语的理解以及知识表示的动态变化等复杂推理过程，良好的完成该任务是未来跨媒体智能、智慧教育等应用的重要基础，也有助于提升计算机对人类认知行为和学习习惯的理解。**

图1 教科书示意图问答样例

**该任务主要有两个挑战：一是示意图资源稀缺，标注代价高昂，具有天然的小样本场景，在不具备大型语料库的场景下完成该任务是一个难题；二是示意图表达抽象、特征稀疏，简单的几何形状蕴含有丰富、专业的语义信息，如何较好地提取并理解其视觉特征是一个难题。本次测评任务的示意图来源于计算机科学领域，希望选手能针对上述两个难点，深入理解示意图中的关键信息，构建模型实现准确的示意图问答表现。**

## 数据描述

* 数据来源：

本次测评任务的示意图问答数据来自Computer Science Diagrams (CSDia) ，是一个包含有丰富标注信息的计算机科学领域数据集。 CSDQA是计算机科学领域的第一个示意图问答数据集，共包含1,294张示意图和超过3,400个问答对。其中，所有的示意图均来自于教育场景中的真实数据，包含两本英文教科书，五本中文教科书，若干博客和知识百科等来源，并经过自动化收集、领域专家选取与标注，可保证数据的准确性和多样性。问答数据集中的问题，包含判断题和四选一问题，数量比例大致为1:1；所有问题依据难度还可划分为简答（一步推理）问题和复杂（两步推理）问题，数量比例大致为4:1。

* 标注内容：

数据集中，每个问题对应一张示意图，为了便于选手处理数据，我们将问题及标注信息统一整理为Q.json文件，文件中问题格式样例如下：

1

|---question: How many contents are in the array?

|---type: MC

|---difficulty: easy

|---diagram\_path: ./data/CSDia\_train\_val\_test/train/D\10.png

|---cords:

| |---0

| |---211

| |---233

| |---108

| |---44

|---answer

| |---d: 13

| |---c: 10

| |---b: 5

| |---a: 9

|---correct\_answer: a

|---split: train

其中，‘question’为问题文本；‘type’为问题类型，包含TF（判断题）和MC（四选一问题）；‘difficulty’为问题难度，包含简单和复杂；‘diagram\_path’为该问题对应的示意图相对路径；‘cords’中包含有若干对象的标注框坐标；‘answer’为该问题的候选答案；‘correct\_answer’为该问题的正确答案，需要注意，仅在训练集中包含问题的正确答案标注；‘split’为该问题所属的划分集合。

* 训练集&验证集：

在训练数据发布阶段，我们会发布约2000条标注好的数据作为训练集。同时发布约600条不含标注答案的问题作为验证集，选手可以将自己生成的验证集答案提交，比赛系统会对答案进行评测，给出得分并进行排行。

在测试数据发布阶段，我们会发布验证集的标注结果（包括答案），同时发布约600条左右不含标注结果的问题，作为测试。选手将自己生成的测试集答案提交，比赛系统会对答案进行评测，给出得分并进行排行。

## 评价指标

本任务的评价指标使用宏观准确率(Macro Precision)，最终排名以其值为基准。设$Q$为问题集合，$A\_{i}$为选手对第$i$个问题给出的答案集合，$G\_{i}$为第$i$个问题的标准答案集合，相关计算公式如下：

$Macro Precision=\frac{1}{|Q|}\sum\_{i=1}^{|Q|}P\_{i}$, $P\_{i}=\frac{|A\_{i}\bigcap\_{}^{}G\_{i}|}{|A\_{i}|}$

## 任务提交

**提交方式：**

本次任务将采取刷榜的方式，分为两个阶段进行：

第一阶段：验证集发布后，允许参赛队伍多次向平台提交结果。参赛者将所有预测结果写入一个结果文件中上传给系统（每日最多可上传3次），如果新提交结果好于之前提交结果，排行榜中的成绩将自动进行更新覆盖。

第二阶段：测试集发布后，验证集提交通道将会关闭，其余事项与第一阶段相同。

**最终提交文件要求：**

每一个参赛队需提交的材料如下。

* 1. 问答任务测试集结果文件，用result.txt命名（UTF-8格式）
	2. 相关代码及说明
	3. 方法描述文档（非评测论文，评测论文撰写要求见CCKS 2022官网）

以上三个文件需在任务提交截止日期前发送至邮箱wang97@stu.xjtu.edu.cn。邮件的标题为：“CCKS-DQA-参赛队名称”，例如“CCKS-DQA-张三队”。

代码及其文档需打包成一个文件（tar，zip，gzip，rar等均可），用code.xxx命名，要求提交所有的程序代码及相关的配置说明，**程序应当可以运行且所得结果与result.txt相符**。如果方法使用了额外资源，要求说明并提供资源文件或地址。

**评测平台：**

 本次评测将依托Biendata平台展开，请有意向的参赛队伍关注平台上的竞赛列表。

## 时间安排

* 任务征集截止：3月24日
* 任务准备时间：3月25日—4月11日
* 评测任务发布：4月11日
* 报名时间：4月11日—7月25日
* 训练及验证数据发布：4月25日
* 测试数据发布：7月25日
* 提交测试结果：7月31日
* 评测论文提交：8月12日
* CCKS会议日期(评测报告及颁奖)：8月25日—28日

## 规则

1. 报名方式：登入Biendata平台官网，完成个人信息注册，即可报名参赛。本次比赛的参赛对象仅限全日制在校大学生（本科、硕士、博士均可）和企业员工；

2. 参赛选手需确保注册时提交信息准确有效，所有的比赛资格及奖金支付均以提交信息为准；

3. 参赛选手在管理系统中组队，参赛队伍成员数量不得超过5个，报名截止日期之后不允许更改队员名单；

4. 每支队伍需指定一名队长，队伍名称不超过15个字符，队伍名的设定不得违反中国法律法规或公序良俗词汇，否则组织者有可能会解散队伍；

5. 每名选手只能参加一支队伍，一旦发现某选手以注册多个账号的方式参加多支队伍，将取消相关队伍的参赛资格；

6. 允许使用开源代码或工具，但不允许使用任何未公开发布或需要授权的代码或工具；

7. 除主办方提供的数据集外，参赛选手允许使用预训练数据（如词向量、字向量等），但是不能用额外的标注数据；

## 任务组织者及联系人

组织者：张玲玲、刘均、武亚强、魏笔凡、王绍伟

联系人：王绍伟 wang97@stu.xjtu.edu.cn

届时会开放微信群作为任务讨论平台。

## 奖项安排

一等奖（1名）：5000

二等奖（2名）：2500

三等奖（2名）：1000

创新技术奖（1名）：3000