任务：带条件的分层级多答案问答

1. **任务描述**

文本问答是语言文本分析的一项重要研究任务。随着机器阅读理解数据集的不断提出和模型水平的长足进步，问答系统得到了学术界和产业界的广泛关注。抽取式问答是一种重要的问答任务形式，其要求模型在相应的文本中抽取文本片段（span）作为答案。不同于目前常见的单答案（single-span）抽取任务，在实际应用中，一个问题的答案可能存在于文本的多个位置，被称为多答案（multi-span）抽取式问答。通过对互联网问答社区的样本进行分析，我们发现多答案样本往往伴随着以下两个特征：1）由于提问用户缺乏相关的领域知识，提出的问题不够明确，导致需要根据不同的条件分别回答；2）一个问题的多个答案可能分属于不同的粒度，并且彼此之间存在着层级关系。这些都为多答案问答提出了严峻的挑战。

为此，腾讯和中国科学院自动化研究所联合举办“带条件的分层级多答案问答”评测任务，希望通过本次评测任务以及发布的中文数据集，提升现有阅读理解模型的水平，并进一步帮助问答系统在真实场景中得到更有价值的应用。

1. **数据描述**

本次数据主要来源于腾讯外部问答社区的样本，原始数据包含一个问题和可以回答该问题的文本。其包含的问题涉及多个领域，医疗健康领域相关的问题占最大比重（77.6%）。本次评测答案均为人工标注，所有样本均需包含多个答案片段（span）。除答案片段外，我们对回答问题所需的条件片段也进行了标注，并且答案片段被标记为两种粒度：粗粒度和细粒度。即标注了三种文本片段：条件、粗粒度答案和细粒度答案。除文本片段外，我们还标注了条件与答案的对应关系，以及两种粒度答案之间的层级关系。

在训练及验证数据发布阶段，我们会发布5000条数据作为训练集，1000条数据作为验证集。在测试数据发布阶段，我们将会再发布1000条测试数据，不含标注结果。

发布的数据每行为一个json格式数据, 其中context为文本，question为问题。condition、coarse、fine分别为三种文本片段的起始/结束位置。condition\_answer、coarse\_fine为条件-答案的对应关系和不同粒度答案的对应关系。

**样例示意如下：**

**条件（**condition**）**

条件意为答案是有条件作为限制的。

例：

问题：苹果手机的电池容量是多少？

文本：苹果12的电池容量是…苹果11的电池容量是…

因为文本中不同苹果手机型号的电池容量不同，手机的型号是答案片段（具体容量）的条件。

所以标出条件片段为：苹果12、苹果11。

*注: 由于在标注结果的一致性检验中发现条件片段的边界很难统一，故在最终发布的数据中把标注的条件片段的起始和结束位置扩展到具其最近的停止符。*

此外，答案片段被标记为不同的粒度：粗粒度和细粒度。细粒度意味着答案是特定的事物，例如特定的时间、特定的人等等。粗粒度的意思是答案片段是一类事物的总称，比如含有某些营养成分的食物和具有某些特征的人。

**粗粒度答案（**coarse**）**

粗粒度答案指答案片段是一类事物的总称，比如含有某些营养成分的食物和具有某些特征的人等。

例：富含维生素的食物、老年人。

**细粒度答案（**fine**）**

细粒度答案指答案片段为特定的事物，例如某个具体的食物、某个具体药物等。

例：苹果、阿司匹林。

**条件-答案（**condition\_answer**）**

具有对应关系的条件片段和答案片段被标记为条件-答案对。

例：（苹果11，3110毫安）

**粗粒度答案-细粒度答案（**coarse\_fine**）**

具有层级关系的粗粒度答案和细粒度答案被标记为粗粒度答案-细粒度答案对。

例：（富含维生素的食物，橙子）

**训练集**

{'context': '早期肝炎肝硬化用什么药物治疗要看是什么原因引起的，如果是乙肝引起的肝硬化选择口服抗病毒药物，\n 要选用起效快、耐药率低、疗效好的药物，推荐使用恩替卡韦或替诺福韦。但如果合并的肾功能不全，替诺福韦是禁用的。\n 如果是丙型肝炎引起的肝硬化，抗病毒药物的选择应该根据丙肝的基因型来选择，但是基因分型比较复杂，\n 所以现在泛基因型用药，也就是对丙肝的1-6型基因型都有效果的药物，如丙通沙。\n 对于肝硬化来讲丙肝除了用丙通沙之外还要联合利巴韦林，这样抗病毒的效果才最佳。\n 但是不论是乙肝或者丙肝引起的肝硬化，除了口服上述的抗病毒药物之外还要选择中成药来降低肝内纤维组织，\n 可以口服扶正化瘀胶囊或者复方鳖甲软肝片，帮助肝硬化早期逆转。\n', 'question': ' 早期肝硬化吃什么药\n', 'condition': [[25, 45], [104, 117]], 'coarse': [[40, 45], [269, 272]], 'fine': [[71, 75], [76, 80], [214, 218], [205, 208], [288, 294], [296, 303]], 'condition\_answer': [[[25, 45], [40, 45]], [[104, 117], [205, 208]], [[104, 117], [214, 218]], [[25, 45], [71, 75]], [[25, 45], [76, 80]]], 'coarse\_fine': [[[269, 272], [296, 303]], [[269, 272], [288, 294]], [[40, 45], [76, 80]], [[40, 45], [71, 75]]]}

**验证集&测试集**

{'context': '要预防脑梗死，那么患者一定要注意在饮食上要清淡饮食、多吃新鲜的蔬菜和水果。对于含胆固醇和油脂比较高的食物，\n 尽量不要吃，比如说蛋黄或者是动物的内脏，像大肠、肝脏、肾脏等食物尽量不能吃。患者一定要吃富含维生素的食物。\n 患者如果存在脑梗死的危险因素，比如患者存在高血压、糖尿病、年龄也比较高。\n 这些患者如果没有禁忌症可以服用阿司匹林以及他汀类的药物进行脑梗死的一级预防治疗。此外，\n 针对高血压和糖尿病等血管病的危险因素，患者还需要应用相应的药物进行治疗\n', 'question': ' 吃什么能预防脑梗\n', }

1. **评价指标**

本次任务分别评估文本片段的抽取和关系抽取效果，采用F1值（F1-measure, F1）和精准匹配率（exactly-match, EM）作为评测指标。采用微平均计算F值，即所有样本的文本片段/关系一起计算P和R。

**精准匹配率:** 若预测包含的标准答案中所有文本片段/关系且不包含其他的文本片段/关系，视为回答正确。

$$精准匹配率（EM）=\frac{回答正确样本数量}{样本总量}$$

**F1值:** 其中元素指文本片段或关系。

$$精确率（Precission）=\frac{预测元素与标注元素重合数量}{预测元素总量}$$

$$召回率（Recall）=\frac{预测元素与标注元素重合数量}{标注元素总量}$$

$$F1值=\frac{2\*精确率\*召回率}{精确率+召回率}$$

1. **任务提交**

本次任务将采取刷榜的方式，验证集发布后，允许参赛队伍多次向平台提交结果，文件命名为“参赛队名称\_valid\_result.json”，格式与结果文件result.txt相同，排名随时更新。参赛队伍可在评测集发布之前随时上传验证集的计算结果（每日最多可上传3次），管理系统会及时更新各队伍的最新排名情况；测试集发布后，允许参赛队伍多次提交结果文件，结果文件提交格式如下。

**结果文件格式说明**：

选手将结果保存为result.json，以utf-8编码格式保存，每行为一个样本的答案，顺序与发布数据相同。

样例如下：

{'condition': [[110, 124]], 'coarse': [[28, 36], [99, 107], [169, 175]], 'fine': [[163, 167]], 'condition\_answer': [[[110, 124], [169, 175]], [[110, 124], [163, 167]]], 'coarse\_fine': []}

1. **时间安排**

•评测任务发布：4月11日

•报名时间：4月18日—7月25日

•训练及验证数据发布：4月25日

•验证集提交开放：4月25日

•测试数据发布：7月25日

•测试集提交开放： 7月31日

•评测论文提交：8月12日

•CCKS会议日期：8月25日—28日

1. **规则**

1. 所有参赛选手都必须在biendata平台管理系统中注册，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。

8. 参赛队伍可在参赛期间随时上传验证集的预测结果，一天不能超过3次 ，管理系统会实时更新各队伍的最新排名情况。

1. **奖励设置**

第一名：10000元

第二名：3000元

第三名：2000元

1. **任务组织者和联系人**

组织者：

杨韬（腾讯）

郑孙聪（腾讯）

王唯康（腾讯）

张元哲（中科院自动化所）

刘康（中科院自动化所）

赵军（中科院自动化所）

任务联系人：

王唯康（腾讯）：daxianwang@tencent.com

张元哲（中科院自动化所）：yzzhang@nlpr.ia.ac.cn

1. **任务交流平台**

请参赛选手加入QQ群交流

