



(12) 发明专利申请

(10) 申请公布号 CN 116403231 A

(43) 申请公布日 2023. 07. 07

(21) 申请号 202310398450.X

(22) 申请日 2023.04.14

(71) 申请人 福州大学

地址 350108 福建省福州市闽侯县福州大学城乌龙江北大道2号福州大学

(72) 发明人 陈羽中 陈谨雯

(74) 专利代理机构 福州元创专利商标代理有限公司 35100

专利代理师 张灯灿 蔡学俊

(51) Int. Cl.

G06V 30/412 (2022.01)

G06V 30/19 (2022.01)

G06F 17/16 (2006.01)

G06N 3/0464 (2023.01)

G06N 3/082 (2023.01)

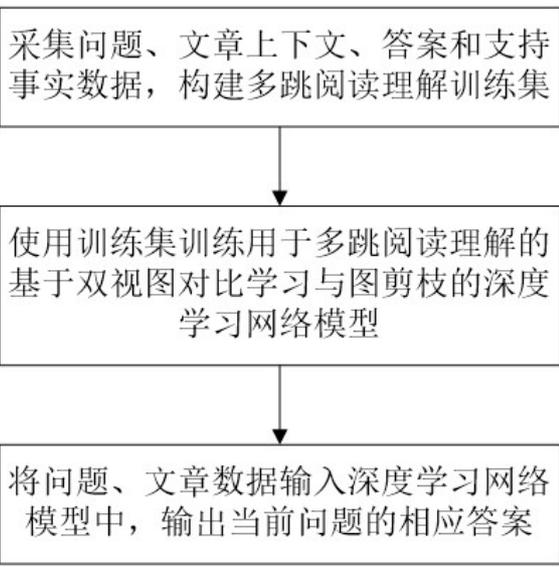
权利要求书5页 说明书14页 附图2页

(54) 发明名称

基于双视图对比学习与图剪枝的多跳阅读理解方法及系统

(57) 摘要

本发明涉及一种基于双视图对比学习与图剪枝的多跳阅读理解方法及系统,该方法包括以下步骤:步骤A:采集问题、文章上下文、答案和支持事实数据,构建多跳阅读理解训练集;步骤B:使用训练集训练用于多跳阅读理解的基于双视图对比学习与图剪枝的深度学习网络模型D;步骤C:将问题、文章数据依次输入深度学习网络模型D中,输出当前的问题的相应答案。该方法及系统有利于提高多跳阅读理解答案预测的准确性。



1. 一种基于双视图对比学习与图剪枝的多跳阅读理解方法,其特征在于,包括以下步骤:

步骤A:采集问题、文章上下文、答案和支持事实数据,构建多跳阅读理解训练集;

步骤B:使用训练集训练用于多跳阅读理解的基于双视图对比学习与图剪枝的深度学习网络模型D;

步骤C:将问题、文章数据输入深度学习网络模型D中,输出当前问题的相应答案。

2. 根据权利要求1所述的基于双视图对比学习与图剪枝的多跳阅读理解方法,其特征在于,所述步骤B具体包括以下步骤:

步骤B1:使用预训练模型Roberta对训练集中各个样本的各个段落包含黄金支持事实的概率进行计算,同时将各个样本的标题和段落的关系对与问题进行文本匹配,得到N个候选段落 $p_{candi} = \{p_1, p_2, \dots, p_N\}$;

步骤B2:将步骤B1得到的候选段落 p_{candi} 和问题、段落中句子、段落中实体词、超链接一同作为节点构造层次图G,并通过G中各节点边关系构造层次图的邻接矩阵 A^{adj} ;

步骤B3:将步骤B2得到的层次图G使用预训练模型Roberta获得所有图节点的初始表示,得到段落、句子、实体三种类型的节点表示 $(f_{p_i}, f_{s_i}, f_{e_i})$ 以及问题节点表示 f_q ,以此获得层次图的总体节点表示 $F = \{f_{p_i}, f_{s_i}, f_{e_i}, f_q\}$;

步骤B4:将步骤B2、B3得到的层次图的邻接矩阵 A^{adj} 及节点表示F用随机删除边关系和随机掩蔽节点特征的方法对层次图分别进行两次损坏,两个损坏后的层次图即为双视图,从而得到两个视图的邻接矩阵 A^1, A^2 和节点表示 F^1, F^2 ;然后将两个视图分别输入到图注意力网络中,得到更新后的节点表示 \tilde{F} ;

步骤B5:使用门控注意力机制计算初始上下文嵌入E和更新后图节点表示 \tilde{F} ,得到相关性系数 \tilde{G} 和上下文表征 E_G ;同时,被损坏的两个视图的邻接矩阵 A^1, A^2 和特征矩阵 F^1, F^2 相互作为正负样本,以余弦相似度计算每一对样本相似性,得到两个图的对比学习损失 L_{gra}^1, L_{gra}^2 ,最大化双视图正例对总体目标损失 L_{gra} ;

步骤B6:在步骤5中得到的上下文表征 E_G 中包含候选答案节点 $\tilde{F}'_e \in \mathbb{R}^{n_e \times d}$,将其与步骤B3得到的问题节点表示 f_q 采用关系过滤方法生成注意力权重稀疏矩阵 $A_{sp} \in \mathbb{R}^{1 \times n_e}$,将其用来筛选当前候选答案节点;

步骤B7:将步骤B4和B6得到的所有的不同类型节点分别输入多层感知机进行子任务预测,得到段落、支持事实、实体词、答案类型预测结果,使用交叉熵函数计算各类预测损失;随后根据答案类型预测结果返回相应类型的答案;然后根据目标损失函数,通过反向传播方法计算深度学习网络模型中各参数的梯度,并利用随机梯度下降方法更新各参数;

步骤B8:当深度学习网络模型产生的损失值小于设定的阈值或达到最大的迭代次数时,终止深度学习网络模型D的训练。

3. 根据权利要求2所述的基于双视图对比学习与图剪枝的多跳阅读理解方法,其特征在于,所述步骤B1具体包括以下步骤:

步骤B11:对训练集进行遍历,对于训练集中的每个样本,使用Roberta模型作为编码器计算样本中每个段落包含黄金支持事实的概率 $P^{ranking}$,使用 $P^{ranking}$ 对段落进行排序, $P^{ranking}$ 表示为:

$$p^{\text{ranking}} = \text{Roberta}_{\text{ranking}}(P)$$

其中, $P = (t_i, p_i)$ 为样本中的段落集, t_i 为段落的标题, p_i 为段落文本;

将问题中的短语跨度 ($q^{\text{start}}, q^{\text{end}}$) 与段落标题中的短语跨度 ($t_i^{\text{start}}, t_i^{\text{end}}$) 相匹配; 如果段落标题与问题匹配时有多个段落匹配成功, 则选择 P^{ranking} 前二的段落; 如果标题匹配没有匹配到段落, 则将问题中的实体词 q_{e_i} 匹配段落中的实体词 p_{e_i} ; 如果实体词匹配也失败, 则取 P^{ranking} 最高的段落, 选取出的段落 P_{sel} 表示为:

$$P_{\text{sel}} = \begin{cases} P_{1,2}^{\text{ranking}} & , p_{\text{sel}} > 2 \\ P_{e_{\text{sel}}} & , e_{\text{sel}} > 0 \\ P_1^{\text{ranking}} & , \text{else} \end{cases}$$

其中, p_{sel} 表示段落标题与问题匹配成功的段落数量, $P_{1,2}^{\text{ranking}}$ 表示此种情况所选择的 P^{ranking} 前二的段落; e_{sel} 表示问题中的实体词与段落匹配是否成功, $e_{\text{sel}} > 0$ 表示匹配成功; $P_{e_{\text{sel}}}$ 表示此种情况问题与段落实体词匹配到的段落; $P_{e_{\text{sel}}}$ 表示问题与段落实体词匹配结果;

步骤B12: 将步骤B11的结果 P_{sel} 相对应的数据集中的超链接来搜索第二跳段落; 第一跳选出的段落与问题构建双向边 (q, p_i), 若第二跳段落存在则构建两段落节点之间的双向边 (p_i, p_j), p_i 为第一跳中选中的段落, p_j 为第二跳中选中的段落; 得到 n 个候选段落 $p_{\text{candi}} = \{p_1, p_2, \dots, p_n\}$, 使用 P^{ranking} 选择排名前 N 的段落, 得到 N 个候选段落 $p_{\text{candi}} = \{p_1, p_2, \dots, p_N\}$ 。

4. 根据权利要求3所述的基于双视图对比学习与图剪枝的多跳阅读理解方法, 其特征在于, 所述步骤B2具体包括以下步骤:

步骤B21: 对于训练集中的每个样本, 将步骤B1中匹配到的段落、各段落中的句子、各句子中的实体词与问题一起构建层次图 G ; 定义层次图中存在四种节点和七种边关系:

$$G = \{V, E\}$$

其中, V 为问题节点 q 、匹配得到的段落节点 $p_{\text{candi}} = \{p_1, p_2, \dots, p_N\}$ 、各段落中的句子节点 $s = \{s_1, s_2, \dots, s_{n_s}\}$ 、各句子中的实体词节点 $e = \{e_1, e_2, \dots, e_{n_e}\}$ 构成的集合, N, n_s, n_e 分别表示层次图中限定的段落、句子、实体词节点的个数; E 为七种边关系组成的集合;

步骤B22: 利用层次图 G 中各节点的边关系构造邻接矩阵 A^{adj} :

$$A^{\text{adj}} = \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nn} \end{bmatrix}$$

其中, $A_{i,j}$ 非零表示两节点之间的关系属于集合 E 中的一种, 为零则表示两节点之间不存在集合 E 中的关系。

5. 根据权利要求4所述的基于双视图对比学习与图剪枝的多跳阅读理解方法, 其特征在于, 所述步骤B3具体包括以下步骤:

步骤B31: 将每个训练样本中的问题对应的选定段落合并, 并与问题连接, 然后输入 Roberta 获得初始表征; 经互注意力层获得问题表示 $Q = \{q_0, q_1, \dots, q_{m-1}\} \in \mathbb{R}^{m \times d}$ 和段落上下文表示 $P = \{p_0, p_1, \dots, p_{n-1}\} \in \mathbb{R}^{n \times d}$, 其中, m 和 n 分别为问题数量和段落数量, d 表示表征向量维度;

步骤B32: 将问题和上下文表示分开处理, 上下文表示中, 对于不同类型的结点, 均以跨度形式计算, 经 Bi-LSTM 网络提取段落、句子、实体三种类型的节点表示 ($f_{p_i}, f_{s_i}, f_{e_i}$):

$$f_{p_i} = MLP_1[LSTM(p_i^{start}); LSTM(p_i^{end})]$$

$$f_{s_i} = MLP_2[LSTM(s_i^{start}); LSTM(s_i^{end})]$$

$$f_{e_i} = MLP_3[LSTM(e_i^{start}); LSTM(e_i^{end})]$$

其中, p_i^{start}, p_i^{end} 表示段落节点的起始位置和结束位置, 句子结点和实体节点的表示类似; $[\cdot; \cdot]$ 代表两向量拼接, 左右两边分别代表Bi-LSTM的前向和后向计算;

步骤B33: 将问题节点经过最大池化层获得节点表示 f_q :

$$f_q = \text{maxpooling}(Q)。$$

6. 根据权利要求5所述的基于双视图对比学习与图剪枝的多跳阅读理解方法, 其特征在于, 所述步骤B4具体包括以下步骤:

步骤B41: 对于层次图 $G = (V, E)$, 其中, V 为图中的节点集合, E 为图中的边关系集合; 图的邻接矩阵为 $A^{adj} \in \mathbb{R}^{n_V \times n_V}$; 将邻接矩阵 A^{adj} 转换为稀疏矩阵 A^{coo} , 根据节点个数随机构造一个掩蔽矩阵 $A^{mask} \in \mathbb{R}^{n_V \times n_V}$ 作为掩蔽矩阵, 当 A^{adj} 中存在边关系的时候为掩蔽矩阵分配值, 其值根据伯努利分布计算得到:

$$A_{ij}^{mask} = \begin{cases} \mathcal{B}(1 - p_v), & A_{ij}^{adj} \neq 0 \\ 0 & , A_{ij}^{adj} = 0 \end{cases}$$

其中, $\mathcal{B}(0)$ 为伯努利分布, p_v 为自定义的掩蔽概率;

步骤B42: 将掩蔽矩阵与原邻接矩阵做哈达玛乘积得到被损坏的邻接矩阵 $\tilde{A}^{adj} \in \mathbb{R}^{n_V \times n_V}$:

$$\tilde{A}^{adj} = A^{adj} \circ A^{mask}$$

其中, (\circ) 为哈达玛积, 以此生成两个邻接矩阵: $A^1 \in \mathbb{R}^{n_V \times n_V}, A^2 \in \mathbb{R}^{n_V \times n_V}$;

步骤B43: 采样随机向量 f^{mask} 作为 d 维掩蔽向量, 向量中的值采用伯努利分布提取:

$$f_i^{mask} = \mathcal{B}(1 - p_f)$$

其中, p_f 为自定义的掩蔽概率;

步骤B44: 将节点特征 $F = \{f_{p_i}, f_{s_i}, f_{e_i}, f_q\}$ 展开为 $F = \{f_1, f_2, \dots, f_{n_V}\} \in \mathbb{R}^{n_V \times d}$, 并与掩蔽向量做哈达玛乘积生成被损坏的节点特征 $\tilde{F} \in \mathbb{R}^{n_V \times d}$:

$$\tilde{F} = [f_1 \circ f^{mask}; f_2 \circ f^{mask}; \dots; f_{n_V} \circ f^{mask}]^T$$

其中, $[\cdot]^T$ 为转置运算, 对原节点特征进行两次掩蔽损坏生成两个特征矩阵: $F^1 \in \mathbb{R}^{n_V \times d}, F^2 \in \mathbb{R}^{n_V \times d}$;

步骤B45: 设图中相邻结点集合为 \mathcal{N}_i , 计算步骤B42和步骤B44生成的两个图的节点 i 与邻居节点之间的相关性系数 e_{ij} , 再经过归一化后得到注意力系数 α_{ij} :

$$e_{ij} = ([W_e h_i; W_e h_j]), j \in \mathcal{N}_i$$

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(e_{ik}))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(e_{ik}))}$$

其中, h_i 为当前计算的节点 i 的表示, h_j 为节点 i 的邻居节点 j 表示, $W_e \in \mathbb{R}^d$ 为共享参数矩阵;

步骤B46: 将每个损坏后的节点特征矩阵的注意力系数 α_{ij} 与邻居节点 \mathcal{N}_i 进行加权计算得到更新后的节点表示 \tilde{F}' , $\tilde{F}' \in \mathbb{R}^{n_V \times d}$ 为更新后的节点特征, 计算过程如下:

$$h'_i = \text{LeakyReLU} \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W_1 h_j \right)$$

其中, $W_1 \in \mathbb{R}^{d \times d}$ 为共享参数矩阵, h'_i 为节点 i 更新后的节点表示。

7. 根据权利要求6所述的基于双视图对比学习与图剪枝的多跳阅读理解方法, 其特征在于, 所述步骤B5具体包括以下步骤:

步骤B51: 设 $E = \{Q, P\}$ 为上下文编码层经Roberta编码后的初始嵌入表示, 以注意力机制的计算方式计算出上下文嵌入和更新图节点的相关性系数 \bar{G} , 再结合门控机制获得门控上下文表征 $E_G \in \mathbb{R}^{n \times 4d}$:

$$\bar{G} = \text{Softmax} \left(\text{Relu}(EW_2) \cdot \text{Relu}(\tilde{F}W_3) \right)^T \cdot \tilde{F}$$

$$E_G = \sigma([E; \bar{G}]W_4) \cdot \text{Tanh}([E; \bar{G}]W_5)$$

其中, $W_2 \in \mathbb{R}^{2d \times 2d}$, $W_3 \in \mathbb{R}^{2d \times 2d}$, $W_4 \in \mathbb{R}^{4d \times 4d}$, $W_5 \in \mathbb{R}^{4d \times 4d}$ 是可学习参数矩阵;

步骤B52: 对于图 A^1 上的任一节点 A_{ij}^1 , 图 A^2 中对应的节点 A_{ij}^2 为正样本, 图 A^2 中的其余节点以及图 A^1 中的其余节点为负样本, A^2 对于 A^1 也是同样计算; 以余弦相似度计算两个图样本损失 L_{gra}^1, L_{gra}^2 , 正例对的训练目标为 e_{pv} , 负例对的训练目标为 e_{nv}^1 , 另一图中负例对的训练目标为 e_{nv}^2 :

$$e_{pv} = e^{\cos(A_{ij}^1, A_{ij}^2) / \tau}$$

$$e_{nv}^1 = \sum_{k=1}^{n_V} e^{\cos(A_{ij}^1, \bar{A}_k^1) / \tau}$$

$$e_{nv}^2 = \sum_{k=1}^{n_V} e^{\cos(A_{ij}^1, \bar{A}_k^2) / \tau}$$

$$L_{gra}^1 = \log \frac{e_{pv}}{e_{pv} + e_{nv}^1 + e_{nv}^2}$$

其中, \bar{A}_k^1, \bar{A}_k^2 分别为 A^1 中除 A_{ij}^1 之外的其他节点的集合以及 A^2 中除 A_{ij}^2 之外的其他节点的集合, τ 是温度系数, $\cos()$ 为余弦相似度计算;

步骤B53: 以最大化双图正例对总体目标损失 L_{gra} 为此模块任务损失:

$$L_{gra} = \sum_{i=1}^{n_V} (L_{gra}^1 + L_{gra}^2)。$$

8. 根据权利要求7所述的基于双视图对比学习与图剪枝的多跳阅读理解方法, 其特征在于, 所述步骤B6具体包括以下步骤:

步骤B61: 将上下文编码后的问题表示 f_q 与更新后的候选答案节点 $\tilde{F}'_e \in \mathbb{R}^{n_e \times d}$ 采用注意力机制生成注意力矩阵 $A_{ds} \in \mathbb{R}^{1 \times n_e}$:

$$A_{ds} = \text{Softmax} \left(\frac{f_q \cdot \tilde{F}'_e}{\sqrt{d \tilde{F}'_e}} \right)$$

步骤B62:将步骤B61生成的注意力矩阵按维度降序排序生成矩阵 $\bar{A}_{ds} \in \mathbb{R}^{1*n_e}$,再将该矩阵中联系最为紧密的 $N_{relation}$ 个关系保留,生成较为精准的注意力权重稀疏矩阵 $A_{sp} \in \mathbb{R}^{1*n_e}$:

$$\bar{A}_{ds} = sort(A_{ds})$$

$$A_{sp} = f_{Pruning}(\bar{A}_{ds})$$

其中,sort()为降序排序函数, $f_{Pruning}()$ 为剪枝函数;

步骤B63:将步骤B62生成的注意力权重稀疏矩阵与候选答案节点 $\tilde{F}'_e \in \mathbb{R}^{n_e*d}$ 相乘,得到筛选后的候选答案节点 $\tilde{F}'_{\bar{e}} \in \mathbb{R}^{n_e*d}$:

$$\tilde{F}'_{\bar{e}} = A_{sp} \cdot \tilde{F}'_e$$

9.根据权利要求8所述的基于双视图对比学习与图剪枝的多跳阅读理解方法,其特征在于,所述步骤B7具体包括以下步骤:

步骤B71:对于图中的段落结点、句子结点使用二分类多层感知机进行段落预测、句子预测;对于实体词结点使用多分类的多层感知机进行实体词预测:

$$p^{para} = MLP_4(\tilde{F}'_p)$$

$$p^{sent} = MLP_5(\tilde{F}'_s)$$

$$p^{ent} = MLP_6(\tilde{F}'_{\bar{e}})$$

其中, $p^{para} \in \mathbb{R}^{n_p}$ 表示段落中包含支持事实的概率, $p^{sent} \in \mathbb{R}^{n_s}$ 表示句子被选择为支持事实的概率, $p^{ent} \in \mathbb{R}^{n_e}$ 表示实体节点中存在正确答案的概率;

步骤B72:使用MLP基于门控注意力层的隐藏状态来进行答案类型预测:

$$p^{type} = MLP_7(E_G[0])$$

其中, p^{type} 为答案类型概率;

步骤B73:对于是非类型,直接返回答案;对于实体类型答案,返回实体词预测结果;对于跨度类型答案,使用上述门控注意力层的隐藏状态计算跨度预测:

$$p^{start} = MLP_8(E_G)$$

$$p^{end} = MLP_9(E_G)$$

其中, p^{start} 、 p^{end} 为跨度开始与结束预测概率;

步骤B74:用交叉熵作为损失函数计算损失值,通过梯度优化算法Adam对学习率进行更新,利用反向传播迭代更新模型参数,以最小化损失函数来训练模型;模型总损失为上述各项预测的损失加权相加:

$$L_{all} = L_{start} + L_{end} + \mu_1 L_{para} + \mu_2 (L_{sent} + \mu_{gra} L_{gra}) + \mu_3 L_{ent} + \mu_4 L_{type}$$

其中, $\mu_1, \mu_2, \mu_3, \mu_4, \mu_{gra}$ 为超参数, L_{start}, L_{end} 为跨度预测的损失, $L_{para}, L_{sent}, L_{ent}, L_{type}$ 分别为段落预测、句子预测、实体词预测、答案类型预测的损失。

10.一种采用如权利要求1-9任一项所述方法的多跳阅读理解系统,其特征在于,包括:
构建训练集模块,用于采集对话上下文和回答数据,构建对话训练集;
模型训练模块,用于训练基于双视图对比学习与图剪枝的深度学习网络模型D;以及
多跳阅读理解模块,将用于多跳阅读理解的过程中将问题和文章输入训练好的深度神经网络模型中输出当前的预测答案。

基于双视图对比学习与图剪枝的多跳阅读理解方法及系统

技术领域

[0001] 本发明属于自然语言处理领域,具体涉及一种基于双视图对比学习与图剪枝的多跳阅读理解方法及系统。

背景技术

[0002] 机器阅读理解是人机问答系统的重要组成部分,其目标为使计算机拥有与人类相媲美的文章理解能力。与信息检索任务不同,机器阅读理解不是简单地让机器根据问题匹配文本数据库中相似度最高的字符串,而是让机器能够理解用户所描述的自然语言问题,这些问题的答案可能存在于文本段落中,可能是“是或否”,也有可能是无法回答的,甚至需要机器根据自己的理解生成或计算出正确的答案。多跳阅读理解是指给出的问题无法在单个段落或单个文档中回答,需要经过至少两次文档跳转的推理链才能得到答案。相较于传统的阅读理解问题更需提高模型的推理能力,要求模型具有更好的解释性和拓展性。对文本语义理解的准确性会影响人机问答系统下游的各种学习任务的性能,而仅通过文本和问题词的简单匹配忽略了未能获取到具体的文章语义信息。因此获得真实的文本语义信息至关重要。

[0003] 近年来,随着GPT、BERT等预训练模型的出现,以及在阅读理解任务上的强大性能提升,诞生了一批优秀的预训练语言模型。单跳阅读理解的简单阅读理解任务如SQuAD、TriviaQA、SearchQA等已取得重大突破,研究者们逐渐将目光转移到更能检验模型的理解程度的“多跳”、“推理”情形上。多跳数据集HotpotQA提供了各种推理策略,其中包含多种问题类型,每个问题对应多段文档,模型应根据不同类型在多段文档中构造一条多跳推理链,整合推理链上的信息得出最终答案。由于单跳任务的成功,许多方法可跨任务借鉴,如多任务方法辅助主任务答案预测。然而,大多数研究依旧使用检索方式查找可能包含正确答案的段落,接着使用单文档答案预测方法。由于多跳推理需要同时应用不同粒度的信息,单一的检索方式不能有效地将收集到的推理证据进行整合。由于图神经网络能够有效整理节点之间的依赖关系,许多模型提出实体词构图、使用关系构图或使用段落和实体词混合构图,使用图神经网络方法更好的整合细粒度信息,可以在图网络层后的实体词中选择答案。但是,此类方法所构造的图难以考虑支持事实信息,仅是单一地通过整合段落和实体词信息进行答案预测。鉴于此,有研究工作使用段落、句子、实体词构造多粒度异构图,通过图注意力机制整合不同粒度信息之间的上下文依赖,利用更新后的节点标识用于不同的子任务,针对多种类型问题,引入跨度预测模块来进行最终答案预测。但是多层次细粒度图将所有线索合并成一个图,难以解释模型的决策,多粒度信息的繁多使得关键信息的获取受到挑战。图神经网络方法过于依赖邻接矩阵,缺乏扩展性。

[0004] 综上,图注意力网络在融合文本和问题之间语义表示中取得了一定的成就,但在语义理解方面依旧存在不足,在证据句子中易被相似性噪声影响。由分析人类阅读理解过程可知,在阅读回答时人们通常需要再结合问题以明确答案应该具有的关键信息,列出候选答案,再通过候选答案得出正确答案。在阅读时人类通常会通过辨别文章和问题中的关

键信息以及干扰信息来得到正确答案。

发明内容

[0005] 本发明的目的在于提供一种基于双视图对比学习与图剪枝的多跳阅读理解方法及系统,该方法及系统有利于提高多跳阅读理解答案预测的准确性。

[0006] 为实现上述目的,本发明采用的技术方案是:一种基于双视图对比学习与图剪枝的多跳阅读理解方法,包括以下步骤:

[0007] 步骤A:采集问题、文章上下文、答案和支持事实数据,构建多跳阅读理解训练集;

[0008] 步骤B:使用训练集训练用于多跳阅读理解的基于双视图对比学习与图剪枝的深度学习网络模型D;

[0009] 步骤C:将问题、文章数据输入深度学习网络模型D中,输出当前问题的相应答案。

[0010] 进一步地,所述步骤B具体包括以下步骤:

[0011] 步骤B1:使用预训练模型Roberta对训练集中各个样本的各个段落包含黄金支持事实的概率进行计算,同时将各个样本的标题和段落的关系对与问题进行文本匹配,得到N个候选段落 $p_{candi} = \{p_1, p_2, \dots, p_N\}$;

[0012] 步骤B2:将步骤B1得到的候选段落 p_{candi} 和问题、段落中句子、段落中实体词、超链接一同作为节点构造层次图G,并通过G中各节点边关系构造层次图的邻接矩阵 A^{adj} ;

[0013] 步骤B3:将步骤B2得到的层次图G使用预训练模型Roberta获得所有图节点的初始表示,得到段落、句子、实体三种类型的节点表示 $(f_{p_i}, f_{s_i}, f_{e_i})$ 以及问题节点表示 f_q ,以此获得层次图的总体节点表示 $F = \{f_{p_i}, f_{s_i}, f_{e_i}, f_q\}$;

[0014] 步骤B4:将步骤B2、B3得到的层次图的邻接矩阵 A^{adj} 及节点表示F用随机删除边关系和随机掩蔽节点特征的方法对层次图分别进行两次损坏,两个损坏后的层次图即为双视图,从而得到两个视图的邻接矩阵 A^1, A^2 和节点表示 F^1, F^2 ;然后将两个视图分别输入到图注意力网络中,得到更新后的节点表示 \tilde{F} ;

[0015] 步骤B5:使用门控注意力机制计算初始上下文嵌入E和更新后图节点表示 \tilde{F} ,得到相关性系数 \bar{G} 和上下文表征 E_G ;同时,被损坏的两个视图的邻接矩阵 A^1, A^2 和特征矩阵 F^1, F^2 相互作为正负样本,以余弦相似度计算每一对样本相似性,得到两个图的对比学习损失 L_{gra}^1, L_{gra}^2 ,最大化双视图正例对总体目标损失 L_{gra} ;

[0016] 步骤B6:在步骤5中得到的上下文表征 E_G 中包含候选答案节点 $\tilde{F}_e \in \mathbb{R}^{n_e \times d}$,将其与步骤B3得到的问题节点表示 f_q 采用关系过滤方法生成注意力权重稀疏矩阵 $A_{sp} \in \mathbb{R}^{1 \times n_e}$,将其用来筛选当前候选答案节点;

[0017] 步骤B7:将步骤B4和B6得到的所有的不同类型节点分别输入多层感知机进行子任务预测,得到段落、支持事实、实体词、答案类型预测结果,使用交叉熵函数计算各类预测损失;随后根据答案类型预测结果返回相应类型的答案;然后根据目标损失函数,通过反向传播方法计算深度学习网络模型中各参数的梯度,并利用随机梯度下降方法更新各参数;

[0018] 步骤B8:当深度学习网络模型产生的损失值小于设定的阈值或达到最大的迭代次

数时,终止深度学习网络模型D的训练。

[0019] 进一步地,所述步骤B1具体包括以下步骤:

[0020] 步骤B11:对训练集进行遍历,对于训练集中的每个样本,使用Roberta模型作为编码器计算样本中每个段落包含黄金支持事实的概率 $P^{ranking}$,使用 $P^{ranking}$ 对段落进行排序, $P^{ranking}$ 表示为:

[0021] $P^{ranking} = \text{Roberta}_{ranking}(P)$

[0022] 其中, $P = (t_i, p_i)$ 为样本中的段落集, t_i 为段落的标题, p_i 为段落文本;

[0023] 将问题中的短语跨度(q^{start}, q^{end})与段落标题中的短语跨度(t_i^{start}, t_i^{end})相匹配;如果段落标题与问题匹配时有多个段落匹配成功,则选择 $P^{ranking}$ 前二的段落;如果标题匹配没有匹配到段落,则将问题中的实体词 q_{e_i} 匹配段落中的实体词 p_{e_i} ;如果实体词匹配也失败,则取 $P^{ranking}$ 最高的段落,选取出的段落 P_{sel} 表示为:

$$[0024] \quad P_{sel} = \begin{cases} P_{1,2}^{ranking} & , p_{sel} > 2 \\ P_{e_{sel}} & , e_{sel} > 0 \\ P_1^{ranking} & , else \end{cases}$$

[0025] 其中, p_{sel} 表示段落标题与问题匹配成功的段落数量, $P_{1,2}^{ranking}$ 表示此种情况所选择的 $P^{ranking}$ 前二的段落; e_{sel} 表示问题中的实体词与段落匹配是否成功, $e_{sel} > 0$ 表示匹配成功; $P_{e_{sel}}$ 表示此种情况问题与段落实体词匹配到的段落; P_{eesl} 表示问题与段落实体词匹配结果;

[0026] 步骤B12:将步骤B1 1的结果 P_{sel} 相对应的数据集中的超链接来搜索第二跳段落;第一跳选出的段落与问题构建双向边(q, p_i),若第二跳段落存在则构建两段落节点之间的双向边(p_i, p_j), p_i 为第一跳中选中的段落, p_j 为第二跳中选中的段落;得到n个候选段落 $p_{candi} = \{p_1, p_2, \dots, p_n\}$,使用 $P^{ranking}$ 选择排名前N的段落,得到N个候选段落 $p_{candi} = \{p_1, p_2, \dots, p_N\}$ 。

[0027] 进一步地,所述步骤B2具体包括以下步骤:

[0028] 步骤B21:对于训练集中的每个样本,将步骤B1中匹配到的段落、各段落中的句子、各句子中的实体词与问题一起构建层次图G;定义层次图中存在四种节点和七种边关系:

[0029] $G = \{V, E\}$

[0030] 其中, V 为问题节点 q 、匹配得到的段落节点 $p_{candi} = \{p_1, p_2, \dots, p_N\}$ 、各段落中的句子节点 $s = \{s_1, s_2, \dots, s_{n_s}\}$ 、各句子中的实体词节点 $e = \{e_1, e_2, \dots, e_{n_e}\}$ 构成的集合, N, n_s, n_e 分别表示层次图中限定的段落、句子、实体词节点的个数; E 为七种边关系组成的集合;

[0031] 步骤B22:利用层次图G中各节点的边关系构造邻接矩阵 A^{adj} :

$$[0032] \quad A^{adj} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix}$$

[0033] 其中, A_{ij} 非零表示两节点之间的关系属于集合E中的一种,为零则表示两节点之间不存在集合E中的关系。

[0034] 进一步地,所述步骤B3具体包括以下步骤:

[0035] 步骤B31:将每个训练样本中的问题对应的选定段落合并,并与问题连接,然后输入Roberta获得初始表征;经互注意力层获得问题表示 $Q = \{q_0, q_1, \dots, q_{m-1}\} \in \mathbb{R}^{m \times d}$ 和段落上下文表示 $P = \{p_0, p_1, \dots, p_{n-1}\} \in \mathbb{R}^{n \times d}$,其中,m和n分别为问题数量和段落数量,d表示表征向量维度;

[0036] 步骤B32:将问题和上下文表示分开处理,上下文表示中,对于不同类型的结点,均以跨度形式计算,经Bi-LSTM网络提取段落、句子、实体三种类型的节点表示 $(f_{p_i}, f_{s_i}, f_{e_i})$:

$$[0037] \quad f_{p_i} = MLP_1[LSTM(p_i^{start}); LSTM(p_i^{end})]$$

$$[0038] \quad f_{s_i} = MLP_2[LSTM(s_i^{start}); LSTM(s_i^{end})]$$

$$[0039] \quad f_{e_i} = MLP_3[LSTM(e_i^{start}); LSTM(e_i^{end})]$$

[0040] 其中, p_i^{start}, p_i^{end} 表示段落节点的起始位置和结束位置,句子结点和实体节点的表示类似;[,]代表两向量拼接,左右两边分别代表Bi-LSTM的前向和后向计算;

[0041] 步骤B33:将问题节点经过最大池化层获得节点表示 f_q :

$$[0042] \quad f_q = \text{maxpooling}(Q)。$$

[0043] 进一步地,所述步骤B4具体包括以下步骤:

[0044] 步骤B41:对于层次图 $G = (V, E)$,其中,V为图中的节点集合,E为图中的边关系集合;图的邻接矩阵为 $A^{adj} \in \mathbb{R}^{n_v \times n_v}$;将邻接矩阵 A^{adj} 转换为稀疏矩阵 A^{coo} ,根据节点个数随机构造一个掩蔽矩阵 $A^{mask} \in \mathbb{R}^{n_v \times n_v}$ 作为掩蔽矩阵,当 A^{adj} 中存在边关系的时候为掩蔽矩阵分配值,其值根据伯努利分布计算得到:

$$[0045] \quad A_{ij}^{mask} = \begin{cases} \mathcal{B}(1 - p_v), & A_{ij}^{adj} \neq 0 \\ 0 & , A_{ij}^{adj} = 0 \end{cases}$$

[0046] 其中, $\mathcal{B}(0)$ 为伯努利分布, p_v 为自定义的掩蔽概率;

[0047] 步骤B42:将掩蔽矩阵与原邻接矩阵做哈达玛乘积得到被损坏的邻接矩阵 $\tilde{A}^{adj} \in \mathbb{R}^{n_v \times n_v}$:

$$[0048] \quad \tilde{A}^{adj} = A^{adj} \circ A^{mask}$$

[0049] 其中, (\circ) 为哈达玛积,以此生成两个邻接矩阵: $A^1 \in \mathbb{R}^{n_v \times n_v}, A^2 \in \mathbb{R}^{n_v \times n_v}$;

[0050] 步骤B43:采样随机向量 f^{mask} 作为d维掩蔽向量,向量中的值采用伯努利分布提取:

$$[0051] \quad f_i^{mask} = \mathcal{B}(1 - p_f)$$

[0052] 其中, p_f 为自定义的掩蔽概率;

[0053] 步骤B44:将节点特征 $F = \{f_{p_i}, f_{s_i}, f_{e_i}, f_q\}$ 展开为 $F = \{f_1, f_2, \dots, f_{n_v}\} \in \mathbb{R}^{n_v \times d}$,并与掩蔽向量做哈达玛乘积生成被损坏的节点特征 $\tilde{F} \in \mathbb{R}^{n_v \times d}$:

$$[0054] \quad \tilde{F} = [f_1 \circ f^{mask}; f_2 \circ f^{mask}; \dots; f_{n_v} \circ f^{mask}]^T$$

[0055] 其中, $[\cdot]^T$ 为转置运算,对原节点特征进行两次掩蔽损坏生成两个特征矩阵: $F^1 \in \mathbb{R}^{n_v \times d}$, $F^2 \in \mathbb{R}^{n_v \times d}$;

[0056] 步骤B45:设图中相邻结点集合为 \mathcal{N}_i ,计算步骤B42和步骤B44生成的两个图的节点i与邻居节点之间的相关性系数 e_{ij} ,再经过归一化后得到注意力系数 α_{ij} :

$$[0057] \quad e_{ij} = ([W_e h_i; W_e h_j]), j \in \mathcal{N}_i$$

$$[0058] \quad \alpha_{ij} = \frac{\exp(\text{LeakyReLU}(e_{ik}))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(e_{ik}))}$$

[0059] 其中, h_i 为当前计算的节点i的表示, h_j 为节点i的邻居节点j表示, $W_e \in \mathbb{R}^d$ 为共享参数矩阵;

[0060] 步骤B46:将每个损坏后的节点特征矩阵的注意力系数 α_{ij} 与邻居节点 \mathcal{N}_i 进行加权计算得到更新后的节点表示 \tilde{F}' , $\tilde{F}' \in \mathbb{R}^{n_v \times d}$ 为更新后的节点特征,计算过程如下:

$$[0061] \quad h'_i = \text{LeakyReLU}\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W_1 h_j\right)$$

[0062] 其中, $W_1 \in \mathbb{R}^{d \times d}$ 为共享参数矩阵, h'_i 为节点i更新后的节点表示。

[0063] 进一步地,所述步骤B5具体包括以下步骤:

[0064] 步骤B51:设 $E = \{Q, P\}$ 为上下文编码层经Roberta编码后的初始嵌入表示,以注意力机制的计算方式计算出上下文嵌入和更新图节点的相关性系数 \bar{G} ,再结合门控机制获得门控上下文表征 $E_G \in \mathbb{R}^{n \times 4d}$:

$$[0065] \quad \bar{G} = \text{Softmax}\left(\text{Relu}(EW_2) \cdot \text{Relu}(\tilde{F}W_3)\right)^T \cdot \tilde{F}$$

$$[0066] \quad E_G = \sigma([E; \bar{G}]W_4) \cdot \text{Tanh}([E; \bar{G}]W_5)$$

[0067] 其中, $W_2 \in \mathbb{R}^{2d \times 2d}$, $W_3 \in \mathbb{R}^{2d \times 2d}$, $W_4 \in \mathbb{R}^{4d \times 4d}$, $W_5 \in \mathbb{R}^{4d \times 4d}$ 是可学习参数矩阵;

[0068] 步骤B52:对于图 A^1 上的任一节点 A_{ij}^1 ,图 A^2 中对应的节点 A_{ij}^2 为正样本,图 A^2 中的其余节点以及图 A^1 中的其余节点为负样本, A^2 对于 A^1 也是同样计算;以余弦相似度计算两个图样本损失 L_{gra}^1, L_{gra}^2 ,正例对的训练目标为 e_{pv} ,负例对的训练目标为 e_{nv}^1 ,另一图中负例对的训练目标为 e_{nv}^2 :

$$[0069] \quad e_{pv} = e^{\cos(A_{ij}^1, A_{ij}^2)/\tau}$$

$$[0070] \quad e_{nv}^1 = \sum_{k=1}^{n_v} e^{\cos(A_{ij}^1, \bar{A}_k^1)/\tau}$$

$$[0071] \quad e_{nv}^2 = \sum_{k=1}^{n_v} e^{\cos(A_{ij}^1, \bar{A}_k^2)/\tau}$$

$$[0072] \quad L_{gra}^i = \log \frac{e_{pv}}{e_{pv} + e_{nv}^1 + e_{nv}^2}$$

[0073] 其中, \bar{A}_k^1, \bar{A}_k^2 分别为 A^1 中除 A_{ij}^1 之外的其他节点的集合以及 A^2 中除 A_{ij}^2 之外的其他节点的集合, τ 是温度系数, $\cos()$ 为余弦相似度计算;

[0074] 步骤B53: 以最大化双图正例对总体目标损失 L_{gra} 为此模块任务损失:

$$[0075] \quad L_{gra} = \sum_{i=1}^{n_v} (L_{gra}^1 + L_{gra}^2)。$$

[0076] 进一步地, 所述步骤B6具体包括以下步骤:

[0077] 步骤B61: 将上下文编码后的问题表示 f_q 与更新后的候选答案节点 $\tilde{F}'_e \in \mathbb{R}^{n_e * d}$ 采用注意力机制生成注意力矩阵 $A_{ds} \in \mathbb{R}^{1 * n_e}$:

$$[0078] \quad A_{ds} = \text{Softmax}\left(\frac{f_q \cdot \tilde{F}'_e}{\sqrt{d^{\tilde{F}'_e}}}\right)$$

[0079] 步骤B62: 将步骤B61生成的注意力矩阵按维度降序排序生成矩阵 $\bar{A}_{ds} \in \mathbb{R}^{1 * n_e}$, 再将该矩阵中联系最为紧密的 $N_{relation}$ 个关系保留, 生成较为精准的注意力权重稀疏矩阵 $A_{sp} \in \mathbb{R}^{1 * n_e}$:

$$[0080] \quad \bar{A}_{ds} = \text{sort}(A_{ds})$$

$$[0081] \quad A_{sp} = f_{Pruning}(\bar{A}_{ds})$$

[0082] 其中, $\text{sort}()$ 为降序排序函数, $f_{Pruning}()$ 为剪枝函数;

[0083] 步骤B63: 将步骤B62生成的注意力权重稀疏矩阵与候选答案节点 $\tilde{F}'_e \in \mathbb{R}^{n_e * d}$ 相乘, 得到筛选后的候选答案节点 $\tilde{F}'_{\bar{e}} \in \mathbb{R}^{n_e * d}$:

$$[0084] \quad \tilde{F}'_{\bar{e}} = A_{sp} \cdot \tilde{F}'_e。$$

[0085] 进一步地, 所述步骤B7具体包括以下步骤:

[0086] 步骤B71: 对于图中的段落结点、句子节点使用二分类多层感知机进行段落预测、句子预测; 对于实体词结点使用多分类的多层感知机进行实体词预测:

$$[0087] \quad p^{para} = \text{MLP}_4(\tilde{F}'_p)$$

$$[0088] \quad p^{sent} = \text{MLP}_5(\tilde{F}'_s)$$

$$[0089] \quad p^{ent} = \text{MLP}_6(\tilde{F}'_{\bar{e}})$$

[0090] 其中, $p^{para} \in \mathbb{R}^{n_p}$ 表示段落中包含支持事实的概率, $p^{sent} \in \mathbb{R}^{n_s}$ 表示句子被选择为支持事实的概率, $p^{ent} \in \mathbb{R}^{n_e}$ 表示实体节点中存在正确答案的概率;

[0091] 步骤B72:使用MLP基于门控注意力层的隐藏状态来进行答案类型预测:

$$[0092] \quad p^{\text{type}} = \text{MLP}_7(E_G[0])$$

[0093] 其中, p^{type} 为答案类型概率;

[0094] 步骤B73:对于是非类型,直接返回答案;对于实体类型答案,返回实体词预测结果;对于跨度类型答案,使用上述门控注意力层的隐藏状态计算跨度预测:

$$[0095] \quad p^{\text{start}} = \text{MLP}_8(E_G)$$

$$[0096] \quad p^{\text{end}} = \text{MLP}_9(E_G)$$

[0097] 其中, p^{start} 、 p^{end} 为跨度开始与结束预测概率;

[0098] 步骤B74:用交叉熵作为损失函数计算损失值,通过梯度优化算法Adam对学习率进行更新,利用反向传播迭代更新模型参数,以最小化损失函数来训练模型;模型总损失为上述各项预测的损失加权相加:

$$[0099] \quad L_{\text{all}} = L_{\text{start}} + L_{\text{end}} + \mu_1 L_{\text{para}} + \mu_2 (L_{\text{sent}} + \mu_{\text{gra}} L_{\text{gra}}) + \mu_3 L_{\text{ent}} + \mu_4 L_{\text{type}}$$

[0100] 其中, $\mu_1, \mu_2, \mu_3, \mu_4, \mu_{\text{gra}}$ 为超参数, $L_{\text{start}}, L_{\text{end}}$ 为跨度预测的损失, $L_{\text{para}}, L_{\text{sent}}, L_{\text{ent}}, L_{\text{type}}$ 分别为段落预测、句子预测、实体词预测、答案类型预测的损失。

[0101] 本发明还提供了采用上述方法的多跳阅读理解系统,包括:

[0102] 构建训练集模块,用于采集对话上下文和回答数据,构建对话训练集;

[0103] 模型训练模块,用于训练基于双视图对比学习与图剪枝的深度学习网络模型D;以及

[0104] 多跳阅读理解模块,将用于多跳阅读理解的过程中将问题和文章输入训练好的深度网络模型中输出当前的预测答案。

[0105] 与现有技术相比,本发明具有以下有益效果:本发明首先通过基于图的节点级正负样本对比学习任务来获取更加丰富的上下文互信息,经对比学习后的模型拥有更丰富的上下文语义信息,能够有效辨别多粒度上下文信息以及干扰信息,使得模型更能关注到与问题有关的正确答案的范围。另外,本发明通过问题指导筛选答案节点来缩小候选答案范围,使用问题表示对答案实体节点构造注意力权重矩阵,再利用关系筛选算法对该矩阵进行筛选,仅保留关联性最强的部分关系,减弱相似性表述对模型预测候选答案造成的噪声。

附图说明

[0106] 图1是本发明实施例的方法实现流程图;

[0107] 图2是本发明实施例中深度学习网络模型的架构图;

[0108] 图3是本发明实施例的系统结构示意图。

具体实施方式

[0109] 下面结合附图及实施例对本发明做进一步说明。

[0110] 应该指出,以下详细说明都是示例性的,旨在对本申请提供进一步的说明。除非另有指明,本文使用的所有技术和科学术语具有与本申请所属技术领域的普通技术人员通常理解的相同含义。

[0111] 需要注意的是,这里所使用的术语仅是为了描述具体实施方式,而非意图限制根据本申请的示例性实施方式。如在这里所使用的,除非上下文另外明确指出,否则单数形式

也意图包括复数形式,此外,还应当理解的是,当在本说明书中使用术语“包含”和/或“包括”时,其指明存在特征、步骤、操作、器件、组件和/或它们的组合。

[0112] 如图1所示,本实施例提供了一种基于双视图对比学习与图剪枝的多跳阅读理解方法,包括以下步骤:

[0113] 步骤A:采集问题、文章上下文、答案和支持事实数据,构建多跳阅读理解训练集。

[0114] 步骤B:使用训练集训练用于多跳阅读理解的基于双视图对比学习与图剪枝的深度学习网络模型D。本实施例中深度学习网络模型的结构如图2所示。

[0115] 步骤C:将问题、文章数据输入深度学习网络模型D中,输出当前问题的相应答案。

[0116] 在本实施例中,所述步骤B具体包括以下步骤:

[0117] 步骤B1:使用预训练模型Roberta对训练集中各个样本的各个段落包含黄金支持事实的概率进行计算,同时将各个样本的标题和段落的关系对与问题进行文本匹配,得到N个候选段落 $p_{candi} = \{p_1, p_2, \dots, p_N\}$ 。所述步骤B1具体包括以下步骤:

[0118] 步骤B11:对训练集进行遍历,对于训练集中的每个样本,使用Roberta模型作为编码器计算样本中每个段落包含黄金支持事实的概率 $P^{ranking}$,使用 $P^{ranking}$ 对段落进行排序, $P^{ranking}$ 表示为:

[0119] $P^{ranking} = Roberta_{ranking}(P)$

[0120] 其中, $P = (t_i, p_i)$ 为样本中的段落集, t_i 为段落的标题, p_i 为段落文本。

[0121] 将问题中的短语跨度 (q^{start}, q^{end}) 与段落标题中的短语跨度 (t_i^{start}, t_i^{end}) 相匹配;问题跨度 (q^{start}, q^{end}) 指的是问题中短语的起止位置,标题跨度 (t_i^{start}, t_i^{end}) 指的是段落标题中短语的起止位置,即每个段落都有一个标题,检索标题短语存在于问题中的段落;如果段落标题与问题匹配时有多个段落匹配成功,则选择 $P^{ranking}$ 前二的段落;如果标题匹配没有匹配到段落,则将问题中的实体词 q_{e_i} 匹配段落中的实体词 p_{e_i} ;如果实体词匹配也失败,则取 $P^{ranking}$ 最高的段落,选取出的段落 P_{sel} 表示为:

$$[0122] \quad P_{sel} = \begin{cases} P_{1,2}^{ranking} & , p_{sel} > 2 \\ P_{e_{sel}} & , e_{sel} > 0 \\ P_1^{ranking} & , else \end{cases}$$

[0123] 其中, p_{sel} 表示段落标题与问题匹配成功的段落数量, $P_{1,2}^{ranking}$ 表示此种情况所选择的 $P^{ranking}$ 前二的段落; e_{sel} 表示问题中的实体词与段落匹配是否成功, $e_{sel} > 0$ 表示匹配成功; $P_{e_{sel}}$ 表示此种情况问题与段落实体词匹配到的段落; $P_{e_{sel}}$ 表示问题与段落实体词匹配结果。

[0124] 步骤B12:将步骤B11的结果 P_{sel} 相对应的数据集中的超链接来搜索第二跳段落;第一跳选出的段落与问题构建双向边 (q, p_i) ,若第二跳段落存在则构建两段落节点之间的双向边 (p_i, p_j) , p_i 为第一跳中选中的段落, p_j 为第二跳中选中的段落;得到n个候选段落 $p_{candi} = \{p_1, p_2, \dots, p_n\}$,使用 $P^{ranking}$ 选择排名前N的段落,得到N个候选段落 $p_{candi} = \{p_1, p_2, \dots, p_N\}$ 。

[0125] 多跳阅读理解,即指的是问题的解答要结合段落中多个线索得出,且线索不是一

次就能检索出,当找到一个线索时,要接着在文中寻找下一个线索,这就是多跳意义所在。第一跳即为步骤B11中问题与段落的匹配结果 P_{sel} ,再次利用上一步结果继续检索,即为第二跳。

[0126] 步骤B2:将步骤B1得到的候选段落 p_{candi} 和问题、段落中句子、段落中实体词、超链接一同作为节点构造层次图G,并通过G中各节点边关系构造层次图的邻接矩阵 A^{adj} 。所述步骤B2具体包括以下步骤:

[0127] 步骤B21:对于训练集中的每个样本,将步骤B1中匹配到的段落、各段落中的句子、各句子中的实体词与问题一起构建层次图G;定义层次图中存在四种节点和七种边关系:

[0128] $G = \{V, E\}$

[0129] 其中,V为问题节点q、匹配得到的段落节点 $p_{candi} = \{p_1, p_2, \dots, p_N\}$ 、各段落中的句子节点 $s = \{s_1, s_2, \dots, s_{n_s}\}$ 、各句子中的实体词节点 $e = \{e_1, e_2, \dots, e_{n_e}\}$ 构成的集合,N, n_s , n_e 分别表示层次图中限定的段落、句子、实体词节点的个数;E为七种边关系组成的集合。

[0130] 步骤B22:利用层次图G中各节点的边关系构造邻接矩阵 A^{adj} :

$$[0131] \quad A^{adj} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix}$$

[0132] 其中, A_{ii} 非零表示两节点之间的关系属于集合E中的一种,为零则表示两节点之间不存在集合E中的关系。

[0133] 步骤B3:将步骤B2得到的层次图G使用预训练模型Roberta获得所有图节点的初始表示,得到段落、句子、实体三种类型的节点表示 $(f_{p_i}, f_{s_i}, f_{e_i})$ 以及问题节点表示 f_q ,以此获得层次图的总体节点表示 $F = \{f_{p_i}, f_{s_i}, f_{e_i}, f_q\}$ 。所述步骤B3具体包括以下步骤:

[0134] 步骤B31:将每个训练样本中的问题对应的选定段落合并,并与问题连接,然后输入Roberta获得初始表征;经互注意力层获得问题表示 $Q = \{q_0, q_1, \dots, q_{m-1}\} \in \mathbb{R}^{m \times d}$ 和段落上下文表示 $P = \{p_0, p_1, \dots, p_{n-1}\} \in \mathbb{R}^{n \times d}$,其中,m和n分别为问题数量和段落数量,d表示表征向量维度。

[0135] 步骤B32:将问题和上下文表示分开处理,上下文表示中,对于不同类型的结点,均以跨度形式计算,经Bi-LSTM网络提取段落、句子、实体三种类型的节点表示 $(f_{p_i}, f_{s_i}, f_{e_i})$:

$$[0136] \quad f_{p_i} = MLP_1[LSTM(p_i^{start}); LSTM(p_i^{end})]$$

$$[0137] \quad f_{s_i} = MLP_2[LSTM(s_i^{start}); LSTM(s_i^{end})]$$

$$[0138] \quad f_{e_i} = MLP_3[LSTM(e_i^{start}); LSTM(e_i^{end})]$$

[0139] 其中, p_i^{start}, p_i^{end} 表示段落节点的起始位置和结束位置,句子结点和实体节点的表示类似;[,]代表两向量拼接,左右两边分别代表Bi-LSTM的前向和后向计算。

[0140] 步骤B33:将问题节点经过最大池化层获得节点表示 f_q :

$$[0141] \quad f_q = \text{maxpooling}(Q)。$$

[0142] 步骤B4:将步骤B2、B3得到的层次图的邻接矩阵 A^{adj} 及节点表示F用随机删除边关

系和随机掩蔽节点特征的方法对层次图分别进行两次损坏,两个损坏后的层次图即为双视图,从而得到两个视图的邻接矩阵 A^1, A^2 和节点表示 F^1, F^2 ;然后将两个视图分别输入到图注意力网络中,得到更新后的节点表示 \tilde{F} 。所述步骤B4具体包括以下步骤:

[0143] 步骤B41:对于层次图 $G=(V, E)$,其中, V 为图中的节点集合, E 为图中的边关系集合;图的邻接矩阵为 $A^{adj} \in \mathbb{R}^{n_v \times n_v}$;将邻接矩阵 A^{adj} 转换为稀疏矩阵 A^{coo} ,根据节点个数随机构造一个掩蔽矩阵 $A^{mask} \in \mathbb{R}^{n_v \times n_v}$ 作为掩蔽矩阵,当 A^{adj} 中存在边关系的时候为掩蔽矩阵分配值,其值根据伯努利分布计算得到:

$$[0144] \quad A_{ij}^{mask} = \begin{cases} \mathcal{B}(1 - p_v), & A_{ij}^{adj} \neq 0 \\ 0 & , A_{ij}^{adj} = 0 \end{cases}$$

[0145] 其中, $\mathcal{B}(0)$ 为伯努利分布, p_v 为自定义的掩蔽概率。

[0146] 步骤B42:将掩蔽矩阵与原邻接矩阵做哈达玛乘积得到被损坏的邻接矩阵 $\tilde{A}^{adj} \in \mathbb{R}^{n_v \times n_v}$:

$$[0147] \quad \tilde{A}^{adj} = A^{adj} \circ A^{mask}$$

[0148] 其中, (\circ) 为哈达玛积,以此生成两个邻接矩阵: $A^1 \in \mathbb{R}^{n_v \times n_v}, A^2 \in \mathbb{R}^{n_v \times n_v}$ 。

[0149] 由于双视图就是损坏后的两个层次图,而层次图在模型层的计算中主要是以节点边关系构建的邻接矩阵和节点特征所构成,故本方法生成双视图的方式就是损坏邻接矩阵和节点特征。

[0150] 步骤B43:采样随机向量 f^{mask} 作为 d 维掩蔽向量,向量中的值采用伯努利分布提取:

$$[0151] \quad f_i^{mask} = \mathcal{B}(1 - p_f)$$

[0152] 其中, p_f 为自定义的掩蔽概率。

[0153] 步骤B44:将节点特征 $F = \{f_{p_i}, f_{s_i}, f_{e_i}, f_{q_i}\}$ 展开为 $F = \{f_1, f_2, \dots, f_{n_v}\} \in \mathbb{R}^{n_v \times d}$,并与掩蔽向量做哈达玛乘积生成被损坏的节点特征 $\tilde{F} \in \mathbb{R}^{n_v \times d}$:

$$[0154] \quad \tilde{F} = [f_1 \circ f^{mask}; f_2 \circ f^{mask}; \dots; f_{n_v} \circ f^{mask}]^T$$

[0155] 其中, $[\cdot]^T$ 为转置运算,对原节点特征进行两次掩蔽损坏生成两个特征矩阵: $F^1 \in \mathbb{R}^{n_v \times d}, F^2 \in \mathbb{R}^{n_v \times d}$ 。

[0156] 步骤B45:设图中相邻结点集合为 \mathcal{N}_i ,计算步骤B42和步骤B44生成的两个图的节点 i 与邻居节点之间的相关性系数 e_{ij} ,再经过归一化后得到注意力系数 α_{ij} :

$$[0157] \quad e_{ij} = ([W_e h_i; W_e h_j]), j \in \mathcal{N}_i$$

$$[0158] \quad \alpha_{ij} = \frac{\exp(\text{LeakyReLU}(e_{ik}))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(e_{ik}))}$$

[0159] 其中, h_i 为当前计算的节点 i 的表示, h_j 为节点 i 的邻居节点 j 表示, $W_e \in \mathbb{R}^d$ 为共享参数矩阵。

[0160] 步骤B46:将每个损坏后的节点特征矩阵的注意力系数 α_{ij} 与邻居节点 \mathcal{N}_i 进行加权计算得到更新后的节点表示 \tilde{F}' , $\tilde{F}' \in \mathbb{R}^{n_v \times d}$ 为更新后的节点特征,计算过程如下:

$$[0161] \quad h'_i = \text{LeakyReLU} \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W_1 h_j \right)$$

[0162] 其中, $W_1 \in \mathbb{R}^{d \times d}$ 为共享参数矩阵, h'_i 为节点i更新后的节点表示。

[0163] 这里的 \tilde{F}' 指的是任一被损坏后的节点特征矩阵,且后续两个特征矩阵操作都相同,故用此代称。 h'_i 与 \tilde{F}' 是包含关系, h'_i 指的是节点i的表示,而 \tilde{F}' 指的是图中所有结点的表示。

[0164] 步骤B5:使用门控注意力机制计算初始上下文嵌入E和更新后图节点表示 \tilde{F} ,得到相关性系数 \bar{G} 和上下文表征 E_G ;同时,被损坏的两个视图的邻接矩阵 A^1, A^2 和特征矩阵 F^1, F^2 相互作为正负样本,以余弦相似度计算每一对样本相似性,得到两个图的对比学习损失 L_{gra}^1, L_{gra}^2 ,最大化双视图正例对总体目标损失 L_{gra} 。所述步骤B5具体包括以下步骤:

[0165] 步骤B51:设 $E = \{Q, P\}$ 为上下文编码层经Roberta编码后的初始嵌入表示,以注意力机制的计算方式计算出上下文嵌入和更新图节点的相关性系数 \bar{G} ,再结合门控机制获得门控上下文表征 $E_G \in \mathbb{R}^{n \times 4d}$:

$$[0166] \quad \bar{G} = \text{Softmax} \left(\text{Relu}(EW_2) \cdot \text{Relu}(\tilde{F}W_3) \right)^T \cdot \tilde{F}$$

$$[0167] \quad E_G = \sigma([E; \bar{G}]W_4) \cdot \text{Tanh}([E; \bar{G}]W_5)$$

[0168] 其中, $W_2 \in \mathbb{R}^{2d \times 2d}$, $W_3 \in \mathbb{R}^{2d \times 2d}$, $W_4 \in \mathbb{R}^{4d \times 4d}$, $W_5 \in \mathbb{R}^{4d \times 4d}$ 是可学习参数矩阵。

[0169] 步骤B52:对于图 A^1 上的任一节点 A_{ij}^1 ,图 A^2 中对应的节点 A_{ij}^2 为正样本,图 A^2 中的其余节点以及图 A^1 中的其余节点为负样本, A^2 对于 A^1 也是同样计算;以余弦相似度计算两个图样本损失 L_{gra}^1, L_{gra}^2 ,正例对的训练目标为 e_{pv} ,负例对的训练目标为 e_{nv}^1 ,另一图中负例对的训练目标为 e_{nv}^2 :

$$[0170] \quad e_{pv} = e^{\cos(A_{ij}^1, A_{ij}^2)/\tau}$$

$$[0171] \quad e_{nv}^1 = \sum_{k=1}^{n_v} e^{\cos(A_{ij}^1, \bar{A}_k^1)/\tau}$$

$$[0172] \quad e_{nv}^2 = \sum_{k=1}^{n_v} e^{\cos(A_{ij}^1, \bar{A}_k^2)/\tau}$$

$$[0173] \quad L_{gra}^i = \log \frac{e_{pv}}{e_{pv} + e_{nv}^1 + e_{nv}^2}$$

[0174] 其中, \bar{A}_k^1, \bar{A}_k^2 分别为 A^1 中除 A_{ij}^1 之外的其他节点的集合以及 A^2 中除 A_{ij}^2 之外的其他节点的集合, τ 是温度系数, $\cos(\cdot)$ 为余弦相似度计算。

[0175] 由于正例对和负例对都是建立在两个视图之上的, 正例对指的是图 A^1 中的节点 i 和图 A^2 中节点 i 相对应的那个节点, 故这里的正例对指的是两个视图中所有的正例对, 训练目标相同; 负例对的类型有图内和图外两种, 故有 1、2 之分。

[0176] 步骤 B53: 以最大化双图正例对总体目标损失 L_{gra} 为此模块任务损失:

$$[0177] \quad L_{gra} = \sum_{i=1}^{n_v} (L_{gra}^1 + L_{gra}^2)。$$

[0178] 步骤 B6: 在步骤 5 中得到的上下文表征 E_G 中包含候选答案节点 $\tilde{F}'_e \in \mathbb{R}^{n_e \times d}$, 将其与步骤 B3 得到的问题节点表示 f_q 采用关系过滤方法生成注意力权重稀疏矩阵 $A_{sp} \in \mathbb{R}^{1 \times n_e}$, 将其用来筛选当前候选答案节点。所述步骤 B6 具体包括以下步骤:

[0179] 步骤 B61: 将上下文编码后的问题表示 f_q 与更新后的候选答案节点 $\tilde{F}'_e \in \mathbb{R}^{n_e \times d}$ 采用注意力机制生成注意力矩阵 $A_{ds} \in \mathbb{R}^{1 \times n_e}$:

$$[0180] \quad A_{ds} = \text{Softmax}\left(\frac{f_q \cdot \tilde{F}'_e}{\sqrt{d^{\tilde{F}'_e}}}\right)$$

[0181] 步骤 B62: 将步骤 B61 生成的注意力矩阵按维度降序排序生成矩阵 $\bar{A}_{ds} \in \mathbb{R}^{1 \times n_e}$, 再将该矩阵中联系最为紧密的 N_{relation} 个关系保留, 生成较为精准的注意力权重稀疏矩阵 $A_{sp} \in \mathbb{R}^{1 \times n_e}$:

$$[0182] \quad \bar{A}_{ds} = \text{sort}(A_{ds})$$

$$[0183] \quad A_{sp} = f_{\text{Pruning}}(\bar{A}_{ds})$$

[0184] 其中, $\text{sort}(\cdot)$ 为降序排序函数, $f_{\text{Pruning}}(\cdot)$ 为剪枝函数。

[0185] 步骤 B63: 将步骤 B62 生成的注意力权重稀疏矩阵与候选答案节点 $\tilde{F}'_e \in \mathbb{R}^{n_e \times d}$ 相乘, 得到筛选后的候选答案节点 $\tilde{F}'_{\bar{e}} \in \mathbb{R}^{n_e \times d}$:

$$[0186] \quad \tilde{F}'_{\bar{e}} = A_{sp} \cdot \tilde{F}'_e。$$

[0187] 步骤 B7: 将步骤 B4 和 B6 得到的所有的不同类型节点分别输入多层感知机进行子任务预测, 得到段落、支持事实、实体词、答案类型预测结果, 使用交叉熵函数计算各类预测损失; 随后根据答案类型预测结果返回相应类型的答案; 然后根据目标损失函数, 通过反向传播方法计算深度学习网络模型中各参数的梯度, 并利用随机梯度下降方法更新各参数。所述步骤 B7 具体包括以下步骤:

[0188] 步骤 B71: 对于图中的段落结点、句子节点使用二分类多层感知机进行段落预测、句子预测; 对于实体词结点使用多分类的多层感知机进行实体词预测:

$$[0189] \quad p^{para} = \text{MLP}_4(\tilde{F}'_p)$$

$$[0190] \quad p^{sent} = \text{MLP}_5(\tilde{F}'_s)$$

[0191] $p^{ent} = MLP_6(\tilde{F}'_e)$

[0192] 其中, $p^{para} \in \mathbb{R}^{n_p}$ 表示段落中包含支持事实的概率, $p^{sent} \in \mathbb{R}^{n_s}$ 表示句子被选择为支持事实的概率, $p^{ent} \in \mathbb{R}^{n_e}$ 表示实体节点中存在正确答案的概率。

[0193] 步骤B72:使用MLP基于门控注意力层的隐藏状态来进行答案类型预测:

[0194] $p^{type} = MLP_7(E_G[0])$

[0195] 其中, p^{type} 为答案类型概率。

[0196] 步骤B73:对于是非类型,直接返回答案;对于实体类型答案,返回实体词预测结果;对于跨度类型答案,使用上述门控注意力层的隐藏状态计算跨度预测:

[0197] $p^{start} = MLP_8(E_G)$

[0198] $p^{end} = MLP_9(E_G)$

[0199] 其中, p^{start} 、 p^{end} 为跨度开始与结束预测概率。

[0200] 步骤B74:用交叉熵作为损失函数计算损失值,通过梯度优化算法Adam对学习率进行更新,利用反向传播迭代更新模型参数,以最小化损失函数来训练模型;模型总损失为上述各项预测的损失加权相加:

[0201] $L_{all} = L_{start} + L_{end} + \mu_1 L_{para} + \mu_2 (L_{sent} + \mu_{gra} L_{gra}) + \mu_3 L_{ent} + \mu_4 L_{type}$

[0202] 其中, $\mu_1, \mu_2, \mu_3, \mu_4, \mu_{gra}$ 为超参数, L_{start}, L_{end} 为跨度预测的损失, $L_{para}, L_{sent}, L_{ent}, L_{type}$ 分别为段落预测、句子预测、实体词预测、答案类型预测的损失。

[0203] 步骤B8:当深度学习网络模型产生的损失值小于设定的阈值或达到最大的迭代次数时,终止深度学习网络模型D的训练。

[0204] 如图3所示,本实施例还提供了用于实现上述方法的多跳阅读理解系统,包括构建训练集模块、模型训练模块和多跳阅读理解模块。

[0205] 所述构建训练集模块用于采集对话上下文和回答数据,构建对话训练集。

[0206] 所述模型训练模块用于训练基于双视图对比学习与图剪枝的深度学习网络模型D。

[0207] 所述多跳阅读理解模块将用于多跳阅读理解的过程中将问题和文章输入训练好的深度网络模型中输出当前的预测答案。

[0208] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0209] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0210] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特

定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0211] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0212] 以上所述,仅是本发明的较佳实施例而已,并非是对本发明作其它形式的限制,任何熟悉本专业的技术人员可能利用上述揭示的技术内容加以变更或改型为等同变化的等效实施例。但是凡是未脱离本发明技术方案内容,依据本发明的技术实质对以上实施例所作的任何简单修改、等同变化与改型,仍属于本发明技术方案的保护范围。

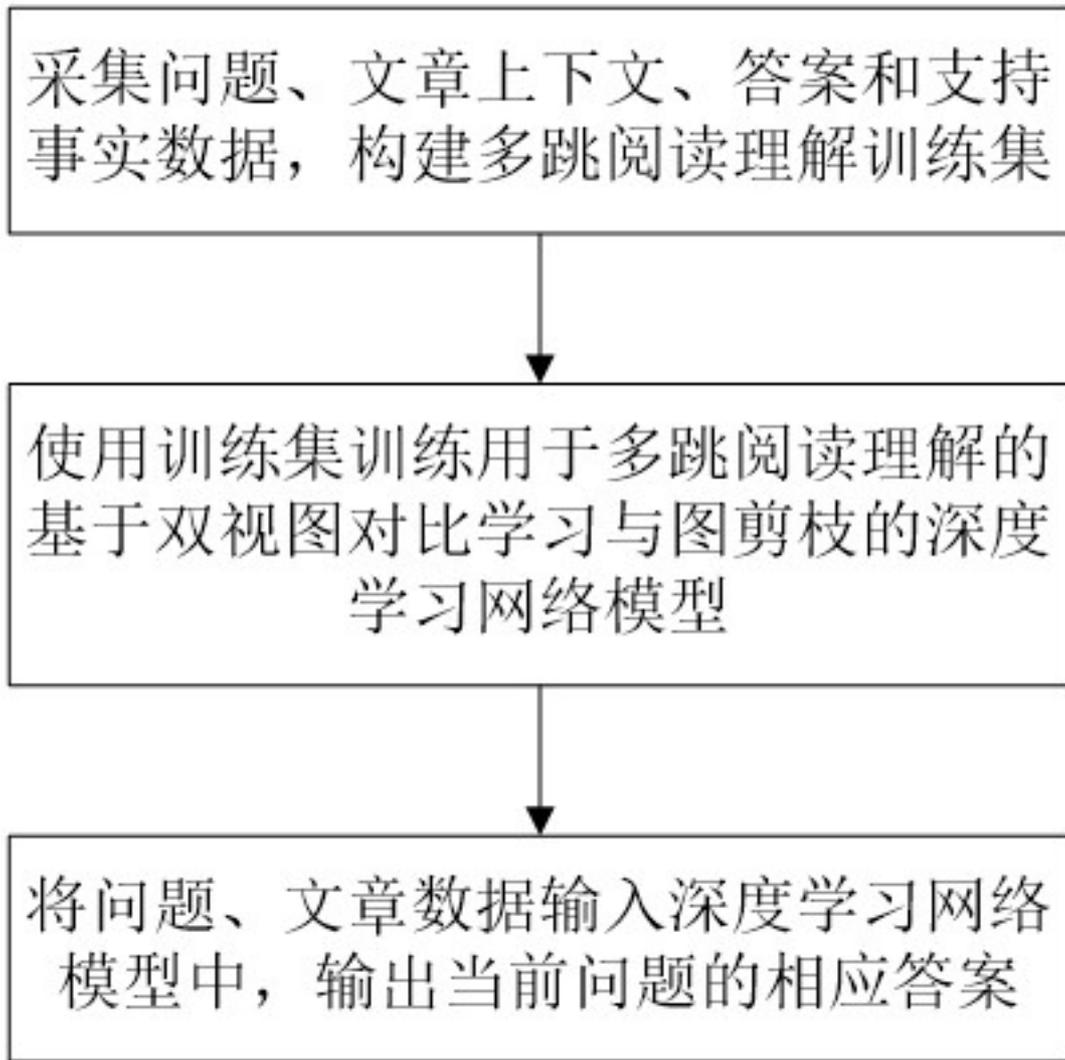


图1

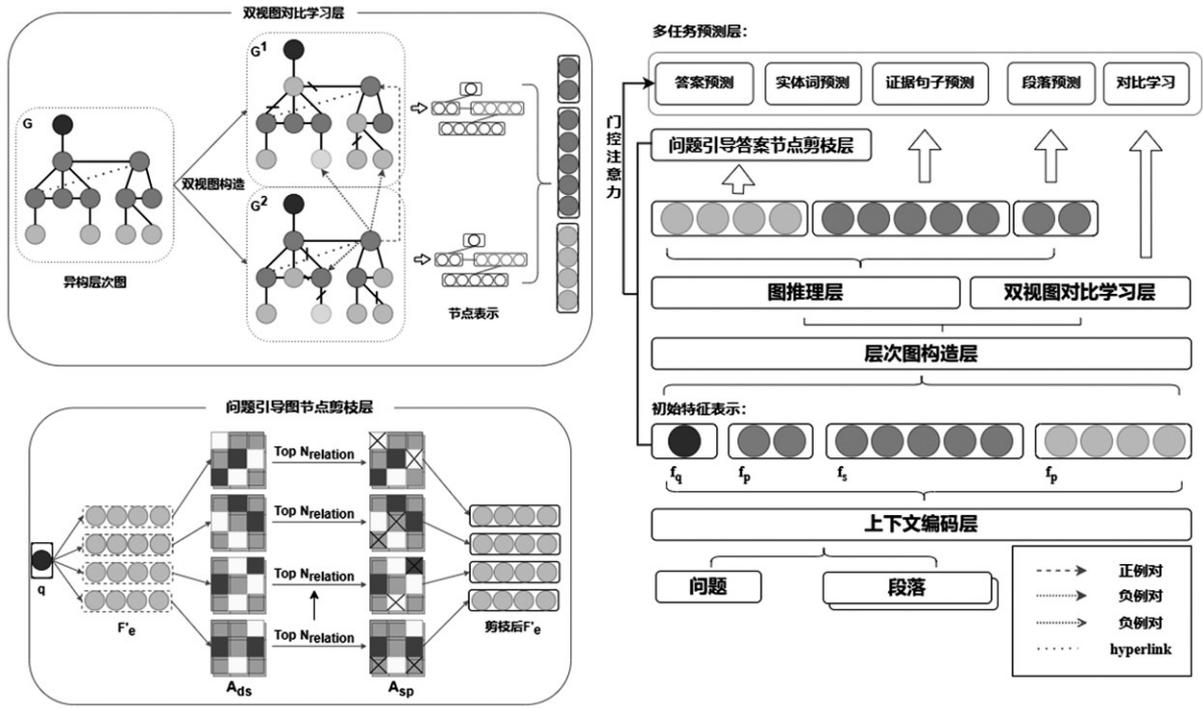


图2

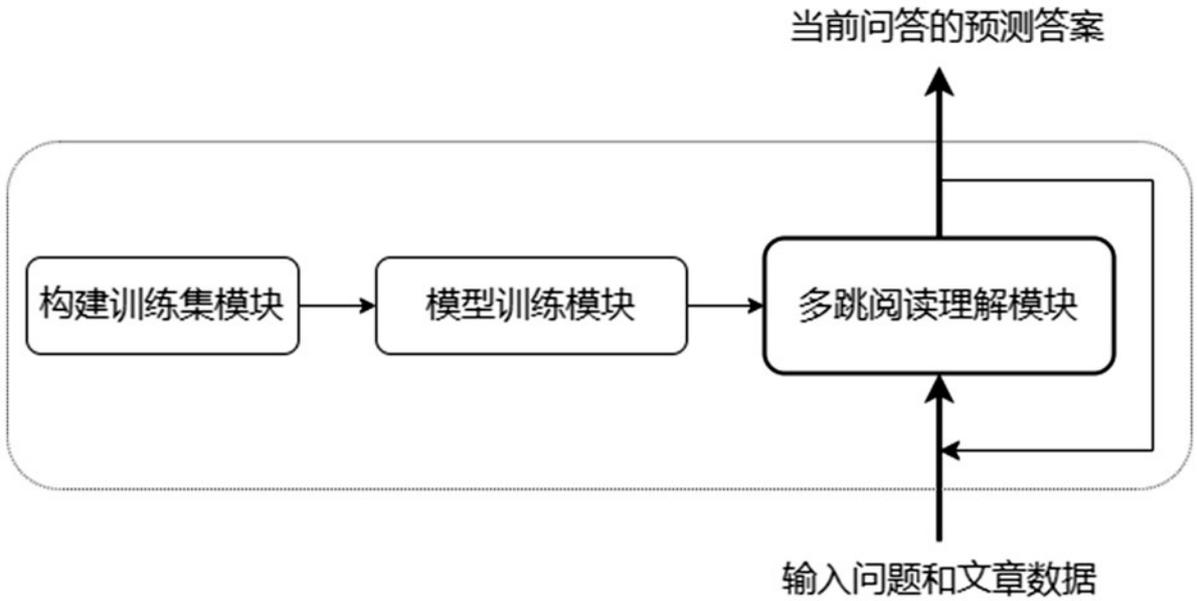


图3