

融合因果关系表征的阅读理解因果关系类选项判断

普瑞丽¹,王元龙¹,李 茹^{1,2}

(1.山西大学 计算机与信息技术学院,太原 030006; 2.山西大学 计算智能与中文信息处理教育部重点实验室,太原 030006)

摘 要: 阅读理解因果关系类选项是指存在因果线索词的选项,此类选项需要根据原文中的因果关系表征进行作答。基于高考阅读理解任务构建因果关系网络,提出融合因果关系表征的因果关系类选项判断方法。采用模式匹配方法抽取原文的因果句对,根据文章因果句对抽取出因果关系词对,并通过点互信息计算因果关系词对之间的因果关联强度,从而构建因果关系网络来表征原文的因果关系。在此基础上,将因果关系表征融入到BERT模型中,预测因果关系选项和原文是否一致。同时,根据高考阅读理解大纲结合语料库发现错误类型分为因果颠倒、强加因果、偷换原因或结果、其他类型等4类,根据每一种错误类型的特点结合预测结果确定选项的错误类型,并提供一个错误解释,以增强方法的可解释性。选用近15年全国高考试题及模拟题中的4 071个科技类阅读理解因果选项进行实验,结果显示F1值达到62.09%,验证了该方法的有效性。

关键词: 高考阅读理解;因果关系类选项;因果关系网络;因果关系表征;可解释性

开放科学(资源服务)标志码(OSID):



中文引用格式: 普瑞丽,王元龙,李茹.融合因果关系表征的阅读理解因果关系类选项判断[J].计算机工程,2022,48(7):89-96.

英文引用格式: PU R L, WANG Y L, LI R. Judgement of causality options in reading comprehension incorporating representation of causality[J]. Computer Engineering, 2022, 48(7): 89-96.

Judgement of Causality Options in Reading Comprehension Incorporating Representation of Causality

PU Ruili¹, WANG Yuanlong¹, LI Ru^{1,2}

(1.School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China; 2.Key Laboratory of Computation Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China)

[Abstract] In reading comprehension task, causality options are those with causal cues, which need to be answered according to the representation of causality in the original text. This paper constructs a causality network based on college entrance examination reading comprehension, and proposes a judgment method for the causality options based on the representation of causality. Firstly, a method based on pattern matching is used to extract the causal sentence pairs from the original text, then the causal word pairs are extracted from the causal sentence pairs of the original text, and the strength of the causality between the causal word pairs is calculated through point mutual information, so as to construct a causality network to represent the causality of the original text. On this basis, the causal representation is incorporated into the BERT model to predict whether the causal choice is consistent with the original text. According to the college entrance examination reading comprehension syllabus combined with the corpus, the error types are divided into four types: reversed cause and effect, imposed cause and effect, substitution of cause or result, and other types. According to the characteristics of each error type and the predicted results, the error type of the option is determined, and an error explanation is provided to enhance the interpretability of the method. In total, 4 071 scientific and technological reading comprehension questions and simulation questions from the last 15 years are tested. The results show that the F1 value reaches 62.09%, verifying the effectiveness of the proposed method.

[Key words] college entrance examination reading comprehension; causality options; causality network; representation of causality; interpretability

DOI: 10. 19678/j. issn. 1000-3428. 0061496

基金项目: 国家重点研发计划重点专项“基于大数据的类人智能关键技术与系统”(2018YFB1005103);国家自然科学基金“面向汉语篇章语义分析的框架推理技术研究”(61772324);国家自然科学基金“基于事件的图文数据阅读理解关键技术研究”(61806117)。

作者简介: 普瑞丽(1997—),女,硕士研究生,主研方向为自然语言处理;王元龙(通信作者),副教授、博士;李 茹,教授、博士。

收稿日期: 2021-04-28 **修回日期:** 2021-07-19 **E-mail:** ylwang@sxu.edu.cn

0 概述

随着自然语言处理研究领域的深入发展,阅读理解任务受到了越来越多研究者的关注。阅读理解任务是指让机器根据给定文本回答相关问题,从而衡量机器理解自然语言的能力。近年来,大量的阅读理解数据集被提出,如 SQuAD^[1]、RACE^[2]、Drop^[3]、CMRC^[4]等,推动了该项研究的发展。

高考语文阅读理解中的选择题来源于高考真实考试场景,由领域专家设计,以评估学生的阅读理解能力,因此在数据质量上有很大的保障。在高考阅读理解任务中,需要结合文章针对提出的问题从四个选项中选择出正确的答案。由近15年的全国高考试题和7392道模拟题中选项的统计可知,在全国高考试题中,包含因果关系选项的题目占总题目的35.43%,包含因果关系的选项占总选项的12.65%,在模拟题中,包含因果关系选项的题目占总题目的39.69%,包含因果关系的选项占总选项的13.77%。因此,本文重点研究因果关系类选项。此外,由于目前大多数阅读理解模型都是使用深度学习的方法将问题转换成一个分类或排序问题,该类方法用在因果关系类选项中存在两个缺陷:一是不能很好地捕捉到因果关系类选项特有的表征;二是只能用来判断选项是否正确,而对于模型预测的结果却无法给出一个合理的解释。

本文对高考阅读理解任务中的因果关系类选项进行分析,发现该类型选项错误的原因主要分为因果颠倒、强加因果、偷换原因或结果、其他等4种类型。针对不同的选项错误原因,提出采用因果网来表征原文,并将其融入到BERT模型^[5]中,从而提高BERT模型预测因果关系选项的性能。此外,结合每一种错误类型的特点,针对因果关系类选项判断的可解释性做进一步研究。

1 相关工作

目前,针对阅读理解提出的方法主要有两种:基于特征的方法^[6-8]和基于神经网络的方法^[9-11]。基于特征的方法核心在于选择合适的特征用以表示文本,常用特征为句法特征、词频特征、词距离特征、框架语义特征^[12-13]、词向量特征等。基于神经网络的方法大多先使用词向量表示,再使用深度学习模型学习文本的向量表示,最后将问题转换为一个分类或排序问题。现阶段,多数研究使用基于神经网络的方法来解决阅读理解任务,由于高考阅读理解中包含的类型较多,难度较大,且数据量较少,加入因果关系网络可以使模型更好地学习到选项与文章中的因果关系表征。文献[6]将高考阅读理解选择题中的选项分为4类,并针对其中的因果选项,提出基于因果网的关系支持度分析方法。本文在文献[6]的基础上,对因果关系类选项按照错误类型进行分类,并将因果关系表征融入到BERT模型中,进而提高模型预测因果关系选项的性能。

在解决因果关系类选项问题时,首要任务是抽取因果关系。目前,因果关系抽取方法主要有两种:

基于模式匹配的方法和基于机器学习的方法。基于模式匹配的方法是依据因果句子的结构特点进行模式匹配抽取因果关系,该方法仅可抽取固定模式的显式因果关系。文献[14]通过分析法语中具有因果含义的动词,实现了一个用于抽取带有标记的显示因果关系的COATIS系统。文献[15]指出,除动词之外,还有一些介词连接词(如because of、for)、状语连接词(如so、therefore)以及从句连接(如the result is)也能表达因果关系,并且使用模式匹配的方法在Wall Street Journal数据中抽取带有标记的因果关系。基于机器学习的方法主要把因果关系抽取任务分为两类:分类任务和序列标注任务,在分类任务中,依据文本中候选对之间是否存在因果关系进行分类。文献[16]利用平行维基语料库来识别已知因果短语的新标记,并使用远程监督的方法构造训练集,利用来自开放类标记的特性和提供上下文信息的语义特性训练一个因果分类器。文献[17]利用生成式对抗网络的对抗学习特性,把带有注意力机制的双向门控循环单元网络和对抗学习融合起来,提出一种融合对抗学习的因果关系分类方法。在序列标注任务中,需要标注出句子中的因果实体。文献[18]使用单词级的词向量及其语义特征,通过BiLSTM网络对句子中的原因、结果及因果连接词进行标注。文献[19]通过使用因果关系的时间特征,把因果关系抽取任务定义成一种特殊的时间进行抽取,并提出一种基于条件随机场(Conditional Random Field, CRF)的紧急情况中因果关系抽取方法。文献[20]提出了BiLSTM+CRF+S-GAT的因果关系抽取模型,将词向量输入模型中从而得到句中每个词的因果标签。与基于模式匹配的方法相比,基于机器学习的方法更加灵活,可以抽取隐式因果关系、跨句、跨段因果关系等,但前提是需要人工对语料进行大量的标注用于训练模型,而基于模式匹配的方法可以更加准确地确定文本中因果关系的因果角色,在探测到因果知识的同时还可以有效地分清何为因,何为果。因此,本文使用已创建好的模板来提取文章中的显示因果关系。

对于构建因果关系网络,文献[6]首先对句子的依存句法进行分析,然后使用3种关系类型的短语(动宾关系、定中关系和状中关系)作为因变量和果变量来构建因果关系网络。然而,由于每个句子的句法结构各不相同,因此得到的因果关系网络比较稀疏。文献[21]在英文语料上,通过保留句子中的动词、名词、形容词和副词作为网络节点来构建因果关系网络,该方法既可以保留句子中的主要信息,又可以缓解网络稀疏的问题,因此,本文借鉴该方法建立一个基于高考语文理解的因果关系网络。

2 融合因果关系表征的因果选项判断方法

在高考阅读理解选择题中,题干对解题的作用较小,多数为考察对原文的理解是否正确,因此,本文参考文献[6],忽略题干的作用,将选择题形式化描述为一个三元组<D, O, A>,其中:D表示文章;O表示因果关系选项;A表示正确答案(正解为1,否则为0)。

对应上述形式,本文针对高考阅读理解中因果关系类选项提出一种判断方法。首先使用基于模式匹配的方法分别抽取文章与因果选项中的原因句和结果句,得到文章因果句对集合与选项因果句对集合,并根据选项因果句从文章中抽取出相应的原因候选句与结果候选句;然后从得到的文章因果句对中抽取出因果关系词对,从而构建因果关系网络;最后在BERT中融入因果关系网络进行训练,并根据每一种错误类型的特点,结合BERT的预测结果来确定选项的类型,进而判断选项是否正确。

2.1 因果句子对抽取

将文章按句号划分成句子集 $D=(S_1, S_2, \dots, S_n)$, 采用因果线索词分别在文章句子集和选项中抽取因果句子对。本文共收集了105个因果线索词,根据因果线索词的特点将抽取模式分为句内因果模式和句间因果模式,具体的抽取模板如表1所示,其中,Cause和Effect分别代表原因句和结果句。

表1 因果线索模式

Table 1 Causality clue patterns

因果线索模式	因果句子对
句内因果	{Cause}是{Effect}(主要 根本 关键...原因)
	{Cause}导致 致使 造成...{Effect}
	{Effect}取决于 源于 在于...{Cause}
	{Effect}是由于 是因为 是由...{Cause}
句间因果	之所以{Effect},因为 由于 缘于...{Cause}
	因为 由于 因...{Cause},所以 故 以至于...{Effect}
	{Cause},于是 所以 故而...{Effect}
	为了 凭借 正是...{Cause},{Effect}
	{Cause},以便 为此 才...{Effect}
	因为 由于 只要...{Cause},{Effect}

对于每一篇文章及选项,得到一个文章因果句对集合 $D^{c-e}=(\dots, (S_i^c, S_i^e), \dots)$ 和选项因果句对集合 $O^{c-e}=(O^c, O^e)$, 其中: S_i^c 和 S_i^e 分别代表文章中第 i 个句子的原因句和结果句; O^c 和 O^e 分别代表选项中的原因句和结果句,具体步骤如算法1所示。

算法1 抽取文章与选项中的因果句子对

输入 文章句子集 $D=(S_1, S_2, \dots, S_n)$, 选项 O
 输出 文章因果句对 $D^{c-e}=(\dots, (S_i^c, S_i^e), \dots)$, 选项因果句对 $O^{c-e}=(O^c, O^e)$

//Pattern(S_i) 用来判断句子 S_i 是否符合因果线索模板,
 //若符合,则返回相应的因果句(S_i^c, S_i^e),否则,返回None。

1. for S_i in D do:

2. if Pattern(S_i) != None then:

3. (S_i^c, S_i^e) ← Pattern(S_i)

4. Append(S_i^c, S_i^e) to D^{c-e}

5. end if

6. end for

7. O^{c-e} ← Pattern(O)

8. return D^{c-e}, O^{c-e}

例句1 雾霾加重 {Effect} 的原因在于大气污染排放总量及集聚程度上升 {Cause}。

例句2 动物的大脑是对称的 {Cause}, 因而动物不能区分左右 {Effect}。

从例句1中可以提取到的句内因果句对为(大气污染物排放总量及集聚程度上升,雾霾加重),而从例句2中可以提取到的句间因果句对为(动物的大脑是对称的,动物不能区分左右)。

2.2 候选句抽取

在分别得到文章句子集 $D=(S_1, S_2, \dots, S_n)$ 与选项因果句 $O^{c-e}=(O^c, O^e)$ 之后,为了从文章中抽取出与选项相关的句子,本文先使用BERT对文章中的每一个句子以及选项进行编码,分别计算 O^c 和 O^e 与文章句子集 D 中每一个句子的余弦相似度,计算公式如式(1)所示:

$$S_{\text{sim}}(S_m, O^c) = \frac{\sum_{j=1}^n (x_j \times y_j)}{\sqrt{\sum_{j=1}^n (x_j)^2} \times \sqrt{\sum_{j=1}^n (y_j)^2}} \quad (1)$$

其中: x_j 和 y_j 分别表示 S_m 和 O^c 中第 j 个词的词向量; n 表示最大句子长度。在此基础上,分别选出相似度最大的句子 S_i 和 S_j 作为原因候选句与结果候选句,即选项 O 的候选句 $S_D=(S_i, S_j)$,具体步骤如算法2所示。

算法2 抽取候选句

输入 文章句子集 $D=(S_1, S_2, \dots, S_n)$, 选项因果句对 $O^{c-e}=(O^c, O^e)$

输出 选项候选句 $S_D=(S_i, S_j)$

1. $O^c \leftarrow \text{Bert.encoding}(O^c)$

2. $O^e \leftarrow \text{Bert.encoding}(O^e)$

3. max_similar_c ← 0

4. max_similar_e ← 0

5. for S_i in D do:

6. $S_i \leftarrow \text{Bert.encoding}(S_i)$

7. if Sim(S_i, O^c) > max_similar_c then:

8. max_similar_c ← Sim(S_i, O^c)

9. max_index_c ← i

10. end if

11. if Sim(S_i, O^e) > max_similar_e then:

12. max_similar_e ← Sim(S_i, O^e)

13. max_index_e ← i

14. end if

15. end for

16. return $S_{\text{max_index_c}}, S_{\text{max_index_e}}$

2.3 因果关系网络构建

本文参考文献[21]的方法构建一个基于高考语文阅读理解的因果关系网络,网络中的节点均为词,在因果句子对(Cause, Effect)中,将Cause中的一个单词 w_i 作为因(记为 i_c),Effect中的一个单词 w_j 作为果(记为 j_e),并两两相连构成因果关系对(i_c, j_e),记其出现的频率为1。统计同一因果关系对在不同因果模式下的频率,便可以得到该因果词对在语料中出现的总频率。例如,从2.1节例句1中,可以提取出的因果关系对有:(大气,雾霾),(污染物,雾霾),(排放,雾霾),(总量,雾霾),(集聚,雾霾),(程度,雾霾),(上升,雾霾),(大气,加重),(污染物,加重),(排放,加重),(总量,加重),(集聚,加重),(程度,加重),(上升,加重)。在此过程中,去除停用词,并且只保留在WordNet^[22]中作为名词、动词和形容词

出现过的词,从而在一定程度上可以减少一些无意义的因果关系对。

因果关系网络的边的权重为2个节点之间的点互信息,计算公式如式(2)~式(5)所示:

$$P_{PMI}(i_c, j_c) = \text{lb} \frac{P(i_c, j_c)}{P(i_c)P(j_c)} \quad (2)$$

$$P(i_c) = \frac{\sum_{w \in W} f(i_c, w_c)}{M} \quad (3)$$

$$P(j_c) = \frac{\sum_{w \in W} f(w_c, j_c)}{M} \quad (4)$$

$$P(i_c, j_c) = \frac{f(i_c, j_c)}{M} \quad (5)$$

其中: $f(i_c, j_c)$ 是在语料中抽取到的因果关系对 (i_c, j_c) 的数目; M 是所有因果关系的总数; W 是语料中所有单词的集合。

因果关系网络的具体构建步骤如算法3所示,该算法时间复杂度为 $O(l_{len}(D^{c-c}) \times l_{len}(S_i^c) \times l_{len}(S_j^c))$,其中 $l_{len}(D^{c-c})$ 表示 D^{c-c} 中因果句对的个数, $l_{len}(S_i^c)$ 和 $l_{len}(S_j^c)$ 分别表示原因句 S_i^c 和结果句 S_j^c 中词的个数。

算法3 构建因果关系网络

输入 文章因果句对 $D^{c-c} = (\dots, (S_i^c, S_j^c), \dots)$,需要计算因果关联强度的句对 (S_c, S_e)

输出 因果关系网络 $G=(V, R)$,句对 (S_c, S_e) 的因果关联强度 $C_{Cau}(S_c, S_e)$

```

1. wd = []
2. for  $(S_i^c, S_j^c)$  in  $D^{c-c}$  do:
3. 对  $S_i^c, S_j^c$  进行分词,并去除停用词,仅保留动词、名词和形容词
4. for  $i_c$  in  $S_i^c$  do:
5. for  $j_c$  in  $S_j^c$  do:
6. Append( $i_c, j_c$ ) to wd
7. end for
8. end for
9. end for
10. V 表示图中节点与 wd 中的词一一对应

```

11. R 表示有向边, $R = V_i \xrightarrow{\text{weight}} V_j$ 表示节点 V_i 与 V_j 之间的因果关系, $w_{\text{weight}} = P_{PMI}(V_i, V_j)$ 表示边的权重

12. 对于任意句对 (S_c, S_e) ,当 S_c 作为原因句, S_e 作为结果句时,其因果关联强度为 $C_{Cau}(S_c, S_e)$

2.3.1 词语级因果关联强度计算

在计算两个句子中的词对之间的因果关联强度时,需要同时考虑使用因果关系网络中的信息以及该词对所在的上下文信息,因此,本文首先得到句子中的每个词基于上下文的表示,并用余弦距离来计算每个词对之间的相关度,再加上每个词对在因果关系网络中的点互信息,将结果作为每个词对基于其所在句子的因果关联强度,如式(6)所示:

$$C_{\text{Causal}}((i_c, j_c) | (S_c, S_e)) = P_{PMI}(i_c, j_c) + \cos(V_{i_c}, V_{j_c}) \quad (6)$$

其中: S_c 代表词语 i_c 所在的句子; S_e 代表词语 j_c 所在的句子; V_{i_c} 代表词语 i_c 基于上下文的词向量表示; V_{j_c}

代表词语 j_c 基于上下文的词向量表示; $\cos(V_{i_c}, V_{j_c})$ 代表 V_{i_c} 与 V_{j_c} 之间的余弦距离。

2.3.2 句子级因果关联强度计算

当 S_c 作为原因句, S_e 作为结果句时,它们之间的因果关联度可根据式(7)计算得出:

$$C_{\text{Cau}}(S_c, S_e) = \frac{1}{|S_c| + |S_e|} \sum_{i_c \in S_c} \sum_{j_e \in S_e} C_{\text{Causal}}((i_c, j_e) | (S_c, S_e)) \quad (7)$$

其中: i_c 表示 S_c 中的第 i 个词; j_e 表示 S_e 中的第 j 个词; $|S_c|$ 和 $|S_e|$ 分别表示 S_c 和 S_e 的句子长度。

2.4 融合因果关系表征的BERT模型

为了使BERT能够更好地捕捉到选项内部以及候选句内部的因果关系,本文将因果关系表征融入到BERT模型中,模型结构如图1所示。

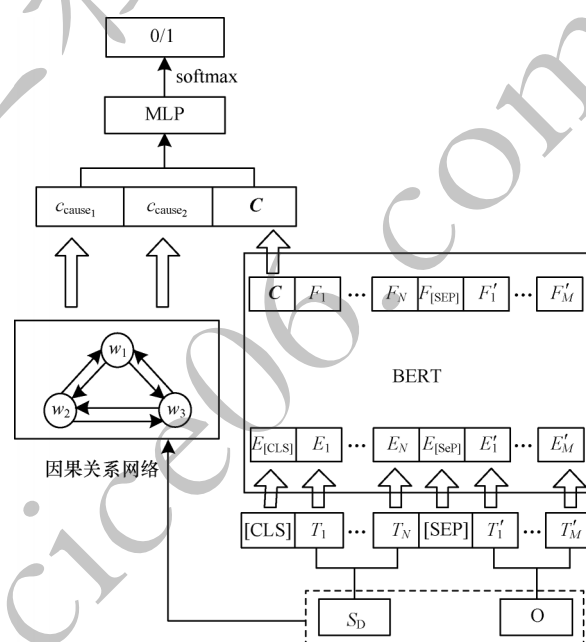


图1 融合因果关系表征的BERT模型

Fig.1 BERT model incorporating representation of causality

本文使用BERT对选项 O 与候选句 S_D 进行编码,使用BERT最后一层对应于输入中的 $[CLS]$ 的 C 向量作为句子对表示,如式(8)所示:

$$C = \text{BERT}(S_D, O) \quad (8)$$

用 c_{cause_1} 表示原因候选句与结果候选句之间的因果关联度,用 c_{cause_2} 表示选项因果句之间的因果关联度,分别如式(9)和式(10)所示:

$$c_{\text{cause}_1} = C_{\text{Cau}}(S_i, S_j) \quad (9)$$

$$c_{\text{cause}_2} = C_{\text{Cau}}(O^c, O^c) \quad (10)$$

选项与候选句间的表示由3个特征拼接而成,最后经过softmax层便可得到最终的分类结果,从而判断选项是否正确,如式(11)和式(12)所示:

$$f = (C; c_{\text{cause}_1}; c_{\text{cause}_2}) \quad (11)$$

$$\hat{y} = \text{softmax}(wf + b) \quad (12)$$

其中: w 为权重矩阵; b 为偏置值。

2.5 选项类型判断

为了增强模型的可解释性,本文在BERT的预测结果上做进一步处理,来判断每一个选项的类型,并且对于判断为错误的选项,提供一个错误解释。

因果关系类选项错误的原因主要分为因果颠倒、强加因果、偷换原因或结果和其他类型,表2为各错误类型示例。4种错误类型的特点如下:

1)因果颠倒:选项在文章中有对应的因果句,但选项中的原因和结果与文章中正好相反。

2)强加因果:选项在文章中有对应的候选句,但候选句间无因果关系,或选项中的原因或结果在原文中并无体现。

3)偷换原因或结果:选项在文章中有对应的因果句,但选项中的原因或结果与文章中不符。

4)其他:如细节错误等。

本文工作主要针对前3种错误类型。根据这3种错误类型的特点,设计图2所示流程进行错误类型判断。

表2 错误类型示例

Table 2 Examples of error types

错误类型	文章	因果选项
因果颠倒	由于各个民族、国家的经济政治发展道路不同,历史文化传统不同,哲学思想在各个民族、国家的存在和发展也必然会有自己的不同特点。	哲学思想在各个民族、国家的存在和发展有自己的不同特点,所以各个民族、国家的历史文化传统就呈现出多样性。
强加因果	绵竹年画历史悠久,它起源于北宋,到明末清初进入繁盛时期。	绵竹年画由于起源于北宋,所以它是所有年画品种中历史最悠久的。
偷换原因或结果	隐性知识是高度个人化的知识,有其自身的特殊含义,因此很难规范化,也不易传递给他人,也就是我们常说的“只能意会不可言传”。	隐性知识主要是长期积累的主观知识,因而“只能意会不可言传”。
其他	2008年起,中国文化文物系统的博物馆开始全面免费开放,此后博物馆的直接经济贡献逐渐减弱。因此,只能通过博物馆事业增加值来衡量博物馆对国民经济的直接贡献。	博物馆免费开放会使其经济贡献逐渐弱化,因此需要通过博物馆事业增加值来衡量博物馆对国民经济的全部贡献。

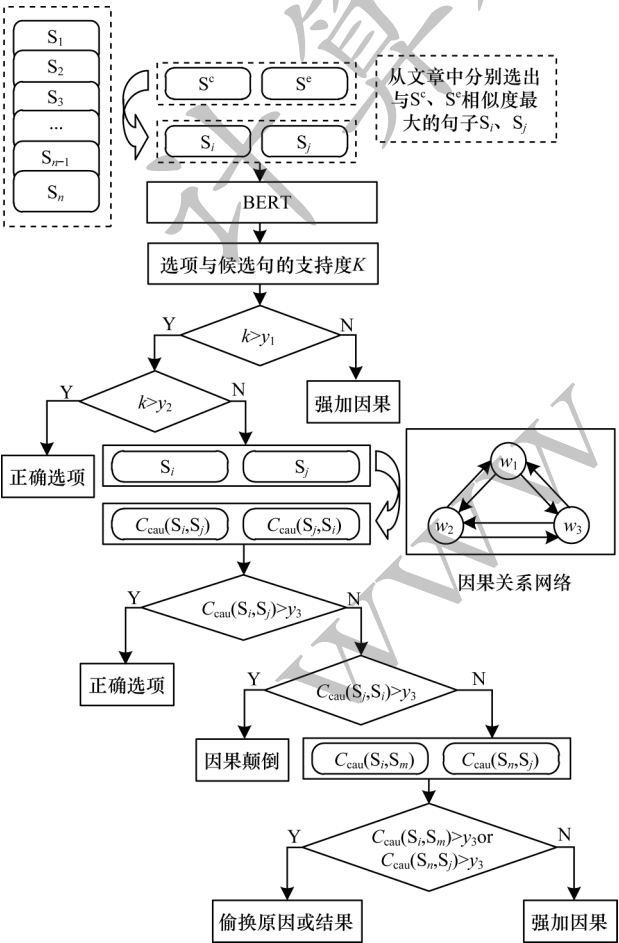


图2 选项类型判断流程

Fig.2 Procedure of option types judgement

在得到2.4节中BERT的预测结果之后,本文使用softmax预测结果为1的概率作为选项与候选句之间的支持度 k ,并设定阈值 y_1 和 y_2 来判断选项与候选句之间的支持度;使用阈值 y_3 作为句子之间是否存在因果关系的依据,若两个句子(S_i, S_j)之间的因果关联强度 $C_{\text{cau}}(S_i, S_j)$ 大于 y_3 ,则表示两个句子之间存在因果关系,否则,表示不存在因果关系。具体步骤如下:

步骤1 当选项与候选句之间的支持度 k 小于 y_1 时,则认为选项与文章内容完全不匹配,即判定为强加因果;若支持度 k 大于 y_2 ,则认为选项与文章内容完全相符,即判定为正确选项;若支持度 k 在 y_1 与 y_2 之间,则需要对选项与候选句之间是否匹配做进一步判断,即转到步骤2。

步骤2 根据因果关系网络分别计算 S_i 与 S_j 之间的因果关联强度 $C_{\text{cau}}(S_i, S_j)$ 及 S_j 与 S_i 之间的因果关联强度 $C_{\text{cau}}(S_j, S_i)$,判断 $C_{\text{cau}}(S_i, S_j)$ 是否大于特定的阈值 y_3 ,若是,说明选项中的原因和结果在文章中都能找到相应的候选句且原因与结果之间存在因果关系,即判定为正确选项,否则,转到步骤3。

步骤3 判断 $C_{\text{cau}}(S_j, S_i)$ 是否大于特定的阈值 y_3 ,若是,说明选项中的原因和结果在文章中都能找到相应的候选句但候选句间的因果关系和选项间的因果关系正好相反,即判定为因果颠倒型,否则,转到步骤4。

步骤4 用 S_i 作为原因分别与文章中的每一个句子计算因果关联强度,并选出因果关联强度最大的值 $C_{\text{cau}}(S_i, S_m)$,如式(13)所示,用 S_j 作为结果分别

与文章中的每一个句子计算因果关系强度,并选出因果关联强度最大的值 $C_{\text{Cau}}(S_i, S_j)$, 如式(14)所示。若 $C_{\text{Cau}}(S_i, S_m)$ 大于特定的阈值 y_3 或 $C_{\text{Cau}}(S_n, S_j)$ 大于特定的阈值 y_3 , 则说明选项在文章中有对应的因果句, 但选项中的原因或结果与文章中不符, 即判定为偷换原因或结果型, 否则, 说明选项在文章中有对应的候选句, 但候选句间并无因果关系, 即判定为强加因果型错误。

$$C_{\text{Cau}}(S_i, S_m) = \max(C_{\text{Cau}}(S_i, S_p)), S_p \in D \quad (13)$$

$$C_{\text{Cau}}(S_n, S_j) = \max(C_{\text{Cau}}(S_q, S_j)), S_q \in D \quad (14)$$

3 实验

3.1 实验数据及参数设置

本文实验所用环境为 CPU: Intel® Xeon® CPU E5-2620 v4@2.10 GHz, GPU: TITAN Xp (12 GB), python 版本为 3.6。本文实验所用的语料为科技类阅读理解, 包括 507 道高考题与 8 161 道模拟题, 其中因果选项有 4 071 个。实验中使用 jieba 分词工具来进行分词与词性标注。本文将选项与候选句之间的支持度阈值 y_1 的取值区间设置为 $[0, 0.5]$, 阈值 y_2 的取值区间设置为 $[0.5, 1]$, 因果关联强度阈值 y_3 的取值区间设置为 $[0, 1]$, 步长都为 0.01, 逐次进行试验, 并从中选出最好的结果作为最终的实验结果。经过多次试验, 当 $y_1=0.33$, $y_2=0.89$, $y_3=0.72$ 时, 实验效果最好。

3.2 基线模型及实验结果分析

为验证本文方法的有效性, 设置以下 5 个基线模型进行对比实验:

1) 参考文献[6], 本文设置了一个基线方法。在得到因果关系网络、选项因果句 $C^{\text{cc}}=(S^{\text{c}}, S^{\text{e}})$ 以及

选项候选句 (S_{h1}, S_{h2}) 之后, 根据因果关系网络计算出 S_{h1} 和 S_{h2} 之间的因果关联度 $C_{\text{Cau}}(S_{h1}, S_{h2})$ 。因此, 可将候选句之间的因果关系支持度 $C_{\text{Causality}}(S_{h1}, S_{h2})$ 表示为式(15), 最后, 通过判断 $C_{\text{Causality}}(S_{h1}, S_{h2})$ 是否大于阈值 y 来判断选项是否正确, 其中, 阈值 y 的取值区间为 $[0, 1]$, 步长为 0.01, 逐次进行试验, 并从中选出最好的结果作为最终的实验结果。经过多次试验, 当 $y=0.75$ 时, 实验结果最好。

$$C_{\text{Causality}}(S_{h1}, S_{h2}) = \frac{S_{\text{Sim}}(S^{\text{c}}, S_{h1}) + S_{\text{Sim}}(S^{\text{e}}, S_{h2}) + C_{\text{Cau}}(S_{h1}, S_{h2})}{3} \quad (15)$$

2) BERT: 本文使用 BERT-base 作为基线方法, 将选项与候选句作为输入, 预测每一个选项是否正确。

3) BERT+cause: 仅在 BERT 中融入因果关系表征进行实验。

4) BERT+kjs: 仅在 BERT 中融入可解释模块进行实验。

5) BERT+cause+kjs: 将因果关系表征与可解释模块都融入 BERT 中进行实验。

实验结果如表3所示。可以看出, 本文的 3 种方法相比文献[6]方法和 BERT 方法的效果均有所提升。相比文献[6]方法, 准确率最多提高 6.25 个百分点, F1 值最多提高 8.91 个百分点; 相比 BERT 方法, 准确率最多提高 1.32 个百分点, F1 值最多提高 1.3 个百分点。根据 BERT+cause 的实验结果可看出, 在 BERT 中融入因果关系表征是有效的; 根据 BERT+kjs 的实验结果可看出, 按错误类型将选项进行分类, 并根据每个错误类型的特点来进行解答是有效的。图3中展示了一个使用本文方法判断正确的示例。

表3 不同方法的性能指标

Table 3 Performance indexes of different methods

方法	准确率/%	精确率/%	召回率/%	F1 值/%	时间/s	
					训练 1 个 epoch	训练 10 个 epoch
文献[6]方法	53.05	51.86	54.58	53.18	—	—
BERT 方法	57.98	58.84	62.88	60.79	372	3 720
BERT+cause 方法	58.20	59.02	63.33	61.10	558	3 906
BERT+kjs 方法	59.08	59.92	63.77	61.78	864	4 212
BERT+cause+kjs 方法	59.30	59.61	64.78	62.09	979	4 327

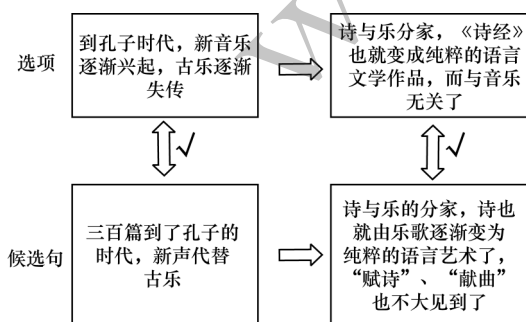


图3 正确选项示例

Fig.3 Examples of correct option

本文方法的准确率仍有进一步的提升空间, 主要原因是:

1) 本文在构建因果关系网络时, 使用的因果文本对是利用模式匹配的方法抽取出来的, 因此对于一些隐式的因果信息可能未能抽取到。

2) 一部分选项中的因果关系是根据文章中的多处信息进行归纳总结或推理之后得到的, 如在错误选项示例1中, 选项为正确选项, 但选项中的结果句“西班牙早期在美洲的主要矿产是黄金”在文章中并没有显示提到, 需要根据候选句进行推断才可得到该信息, 从而容易导致模型判断错误。

错误选项示例 1

候选句:在美洲方面,当西方走向世界寻求财富时,最早寻找的是黄金,但也是从16世纪40年代开始,西班牙在美洲转而开采白银且产量激增。

选项:西方走向世界的重要原因是寻求黄金,因此西班牙最早在美洲的主要矿产是黄金。

由于本文方法涉及到的计算较多,为了准确评估性能,对每一个关键步骤的耗时进行统计,表3中统计了1个epoch和10个epoch分别需要消耗的时间。其中抽取因果句子对、候选句抽取以及构建因果关系网络部分均可作为数据预处理阶段,由于它们在每个epoch中都是固定的,因此在整个训练过程中计算一次,不需要在每个epoch中进行计算,分别耗时126 s、672 s和498 s。与BERT相比,增加了因果关系表征的BERT+cause方法在训练1个epoch和10个epoch均多消耗186 s。增加了可解释性的BERT+kjs方法在训练1个epoch和10个epoch均多消耗492 s。而同时增加了因果表征和可解释性两个模块的方法BERT+cause+kjs方法多消耗607 s,说明本文方法所消耗的时间在整个训练过程中不会随着epoch的增加而变化。

3.3 可解释性实验

对于机器阅读理解,可解释性始终是一个非常好的性质,有助于更充分地理解问题。本文针对错误的选项提供一个错误原因的解释,从而使选项判断具有可解释性。

3.3.1 可解释性实验结果

为了验证本文方法可解释性的有效性,本文标注了一部分实验数据来对可解释性进行验证,每个类型的标注数目各20条,实验设置的评价指标为准确率,实验结果如表4所示。可以看出,本文方法具有可解释性,但是对于偷换原因或结果类型的错误准确率不高,这可能是因为偷换原因或结果类型的选项大多需要从文章中进行总结,如“一开始,蛮子大妈很疼爱这四个敌人,主要是因为四个德国士兵会让她想到自己的儿子,也会让自己的忧愁和牵挂得到一些安慰。”,而在文章中无法找到和选项对应的原因与结果,因此容易判别成其他类型的错误。

表4 可解释性实验结果

Table 4 Explanable experiment results		%
错误类型	准确率	
全部	51.25	
正确选项	51.25	
因果颠倒	45.00	
强加因果	55.00	
偷换原因或结果	40.00	

3.3.2 可解释性实验示例

从实验结果可以看出本文方法对选项错误可解释性的理解是有效的,每种错误类型的可解释性实验示例如下所示:

1)因果颠倒型

经过2.5节中的步骤1~步骤3,可判断出选项中的原因与文章中的结果相匹配,而选项中的结果与文章中的原因相匹配,如图4所示,因此,可判断出选项为错误选项,且错误原因为因果颠倒。

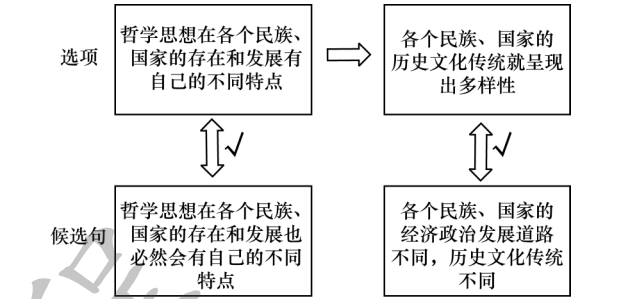


图4 因果颠倒型选项示例

Fig.4 Example of reverse cause and effect option

2)偷换原因或结果型

经过2.5节中的步骤1~步骤4,可判断出选项中的原因与文章中的原因相匹配,而结果却不匹配,如图5所示,因此,可判断出选项为错误选项,且错误原因为偷换原因或结果。

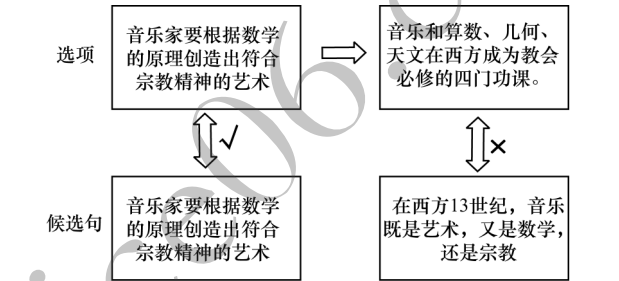


图5 偷换原因或结果型选项示例

Fig.5 Example of swapping cause or effect option

3)强加因果型

在例2中,经过2.5节中的步骤1~步骤4,可判断出选项在文章中对应的候选句间并无因果关系,因此,可判断出选项为错误选项,且错误原因为强加因果,错误选项示例如下:

错误选项示例 2

候选句:绵竹年画历史悠久,它起源于北宋,到明末清初进入繁盛时期。

选项:绵竹年画由于起源于北宋,所以它是所有年画品种中历史最悠久的。

4 结束语

本文针对高考语文阅读理解中的因果关系类选项,提出融合因果关系表征的因果关系类选项判断方法。采用因果关系网来表征原文,并将其融入到BERT模型中,同时结合每一种错误类型的特点,针对因果关系类选项的可解释性进行研究。实验结果表明,本文方法有效且具有可解释性,但其在抽取因果句子对时只能识别到包含因果线索词的显式因果关系。因此,后续将改进方法使其可识别具有隐式因果关系的因果句子对,并进一步提高判断方法的准确率。

参考文献

- [1] RAJPURKAR P, ZHANG J, LOPYREV K, et al. SQuAD: 100,000+ questions for machine comprehension of text[C]//Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2016: 2383-2392.
- [2] LAI G K, XIE Q Z, LIU H X, et al. RACE: large-scale ReAding comprehension dataset from examinationst[C]//Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2017: 785-794.
- [3] DUA D, WAN Y Z, DASIGI P, et al. Drop: a reading comprehension benchmark requiring discrete reasoning over paragraphst[C]//Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2019: 2368-2378.
- [4] CUI Y M, LIU T, CHE W X, et al. A span-extraction dataset for Chinese machine reading comprehension[C]//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2019: 5883-5889.
- [5] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL]. [2021-03-10]. <https://arxiv.org/abs/1810.04805>.
- [6] 王元龙, 李茹, 张虎, 等. 阅读理解中因果关系类选项的研究[J]. 清华大学学报(自然科学版), 2018, 58(3): 272-278.
WANG Y L, LI R, ZHANG H, et al. Causal options in Chinese reading comprehension[J]. Journal of Tsinghua University (Science and Technology), 2018, 58(3): 272-278. (in Chinese)
- [7] 郭少茹, 张虎, 钱揖丽, 等. 面向高考阅读理解的句子语义相关度[J]. 清华大学学报(自然科学版), 2017, 57(6): 575-579, 585.
GUO S R, ZHANG H, QIAN Y L, et al. Semantic relevancy between sentences for Chinese reading comprehension on college entrance examinations[J]. Journal of Tsinghua University (Science and Technology), 2017, 57(6): 575-579, 585. (in Chinese)
- [8] 关勇, 吕国英, 李茹, 等. 面向高考语文阅读理解的篇章标题选择研究[J]. 中文信息学报, 2018, 32(6): 28-35, 43.
GUAN Y, LÜ G Y, LI R, et al. Discourse title selection for Chinese reading comprehension of college entrance examination[J]. Journal of Chinese Information Processing, 2018, 32(6): 28-35, 43. (in Chinese)
- [9] 徐鹏飞, 李晓戈. 基于深度学习的机器中文阅读理解研究[J]. 计算机与数字工程, 2019, 47(12): 3126-3131.
XU P F, LI X G. Study on machine Chinese reading comprehension based on deep learning[J]. Computer & Digital Engineering, 2019, 47(12): 3126-3131. (in Chinese)
- [10] 段利国, 高建颖, 李爱萍. 机器阅读理解中观点型问题的求解策略研究[J]. 中文信息学报, 2019, 33(10): 81-89.
DUAN L G, GAO J Y, LI A P. Research on solving strategies of perspective-type problems in machine reading comprehension [J]. Journal of Chinese Information Processing, 2019, 33(10): 81-89. (in Chinese)
- [11] 张浩宇, 张鹏飞, 李真真, 等. 基于自注意力机制的阅读理解模型[J]. 中文信息学报, 2018, 32(12): 125-131.
ZHANG H Y, ZHANG P F, LI Z Z, et al. Self-attention based machine reading comprehension [J]. Journal of Chinese Information Processing, 2018, 32(12): 125-131. (in Chinese)
- [12] LITKOWSKI K. CLR: integration of FrameNet in a text representation system[C]//Proceedings of the 4th International Workshop on Semantic Evaluations. Stroudsburg, USA: Association for Computational Linguistics, 2007: 113-116.
- [13] LI R, LIU H J, LI S H. Chinese frame identification using T-CRF model[C]//Proceedings of International Conference on Computational Linguistics. Washington D. C., USA: IEEE Press, 2010: 674-682.
- [14] GARCIA D. COATIS, an NLP system to locate expressions of actions connected by causality links[C]//Proceedings of International Conference on Knowledge Engineering and Knowledge Management. Berlin, Germany: Springer, 1997: 347-352.
- [15] KHOO C S G, KORNfilt J, ODDY R N, et al. Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing [J]. Literary and Linguistic Computing, 1998, 13(4): 177-186.
- [16] HIDEY C, MCKEOWN K. Identifying causal relations using parallel wikipedia articles[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, USA: Association for Computational Linguistics, 2016: 1424-1433.
- [17] 冯冲, 康丽琪, 石戈, 等. 融合对抗学习的因果关系抽取[J]. 自动化学报, 2018, 44(5): 811-818.
FENG C, KANG L Q, SHI G, et al. Causality extraction with GAN[J]. Acta Automatica Sinica, 2018, 44(5): 811-818. (in Chinese)
- [18] DASGUPTA T, SAHA R, DEY L, et al. Automatic extraction of causal relations from text using linguistically informed deep neural networks[C]//Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue. Stroudsburg, USA: Association for Computational Linguistics, 2018: 306-316.
- [19] QIU J N, XU L W, ZHAI J, et al. Extracting causal relations from emergency cases based on conditional random fields[J]. Procedia Computer Science, 2017, 112: 1623-1632.
- [20] 许晶航, 左万利, 梁世宁, 等. 基于图注意力网络的因果关系抽取[J]. 计算机研究与发展, 2020, 57(1): 159-174.
XU J H, ZUO W L, LIANG S N, et al. Causal relation extraction based on graph attention networks[J]. Journal of Computer Research and Development, 2020, 57(1): 159-174. (in Chinese)
- [21] 沙雨辰. 常识性因果知识库构建[D]. 上海: 上海交通大学, 2018.
SHA Y C. Construction of commonsense causal knowledge base[D]. Shanghai: Shanghai Jiao Tong University, 2018. (in Chinese)
- [22] MILLER G A. WordNet[J]. Communications of the ACM, 1995, 38(11): 39-41.