

文章编号: 1003-0077(2022)05-0085-09

对话中融入丰富历史信息的回应选择

司博文, 孔芳

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘要: 对话是一个交互的过程, 回应选择旨在根据已有对话上文选择合适的回应, 是自然语言处理领域的研究热点。已有研究取得了一定的成功, 但仍存在两个突出的问题: 一是历史信息与备选回应间的关联关系未得到充分利用; 二是对话历史的潜在语义信息挖掘不够。针对问题一, 该文同时考虑对话的历史信息与备选回应信息, 借助交叉注意力机制实现两者的软对齐, 从而对它们之间的关联关系进行有效捕捉; 针对问题二, 一方面借助多头自注意力机制从多个不同视角捕获对话历史的潜在语义信息, 另一方面借助高速路神经网络实现多种信息的有效桥接, 在深度挖掘语义信息的同时保证信息的完整。在 Ubuntu Corpus V1 数据集上的对比实验表明了该方法的有效性, 模型取得了 88.66% 的 $R_{10}@1$ 值, 90.06% 的 $R_{10}@2$ 值和 95.15% 的 $R_{10}@5$ 值。

关键词: 回应选择; 交叉注意力机制; 自注意力机制; 高速路神经网络

中图分类号: TP391

文献标识码: A

Towards Better Response Selection in Dialogue via Rich Historical Information

SI Bowen, KONG Fang

(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract: Response selection for dialogue is a popular research issue in the field of NLP, which is aimed at selecting appropriate responses based on the existing dialogue. Existing researches are defected in two aspects: 1) insufficient utilization of the correlation between historical information and alternative responses and, 2) insufficient mining of potential semantic information in dialogue history. To deal with the first issue, this paper considers both the historical information and the alternative response information in the dialogue, by the cross-attention mechanism to effectively capture the relationship between them. For the second issue, this paper employs the multi-head self-attention mechanism to capture the latent semantic information of the conversation history from multiple different perspectives, and the highway network to effectively bridge a variety of information to ensure the integrity of the information. Experiments show the proposed method achieves a 88.66% $R_{10}@1$ -score, a 90.06% $R_{10}@2$ -score and a 95.15% $R_{10}@5$ -score on the Ubuntu Corpus V1 dataset.

Keywords: response selection; cross-attention mechanism; self-attention mechanism; highway network

0 引言

人机对话(Human-Computer Conversations^[1])是自然语言处理(Natural Language Process, NLP^[2])领域的关键任务, 旨在最大限度地模仿人与人之间的对话方式, 使得人类能够用更自然的方式与机器进行交流。当前, 人机对话系统主要分为任务型对话系统(Task-Oriented Dialogue Systems^[3])和聊天

机器人(Chatbots^[4]), 任务型对话系统面向垂直领域, 目的是使用尽可能少的对话轮数帮助用户完成预定任务或动作, 例如, 预订酒店、机票和电影票等, 典型的应用有 Siri^[5]、Cortana^[6]等。任务型对话针对特定任务, 这也限制了其在其他领域的使用。聊天机器人面向开放领域, 话题没有特定领域的限制, 因此对系统的要求极高, 典型的应用有微软的小冰^[7]。这两种对话系统都需要根据历史信息进行回应, 而给出的回应都需要确保回应与对话历史在语

收稿日期: 2020-05-13 定稿日期: 2020-06-17

基金项目: 国家自然科学基金(61876118, 61751206)

义上具有连贯性。

产生回应目前有两种方法：检索式和生成式。检索式又称回应选择(Response Selection),旨在针对对话上文从备选回应池中选择最相关的回应;生成式则要求模型自动生成合适的回应。本文主要关注回应选择任务。已有的回应选择方法存在对对话历史的语义挖掘不充分、对话上文和备选回应之间语义关联不够的问题。针对这两个问题,本文提出了融合丰富历史信息的回应选择方法,记为 AICRS (All Information Combination for Response Selection),具体而言,首先将对话上文拼接,形成一个长文,以保证对话历史的序列特性;其次使用交叉注意力机制(Cross-attention Mechanism^[8])将对话历史与回应进行软对齐,从而挖掘历史和回应间的关联信息;接着使用多头自注意力机制(Multi-Head Attention Mechanism^[9])挖掘对话历史及回应的语义信息并使用高速路神经网络(Highway Network^[10])进行信息桥接;最后整合各类信息进行最终的回应选择。该方法能深度挖掘对话历史及回应间的潜在语义和相互的关联关系,提升回应选择的性能。

本文的贡献包括如下几点:①提出了一个结构简单,没有使用任何外部资源的轻量级回应选择模型;②使用交叉注意力机制挖掘对话历史和备选回应之间的语义关联;③将信息桥接的方法应用到回应选择任务,在深度挖掘信息的同时保证信息的完整性;④通过 Ubuntu Corpus V1 数据集上的实验验证模型的有效性。

本文的组织结构如下:第1节介绍相关工作;第2节阐述 AICRS 模型;第3节介绍具体的实验设置,并对实验结果进行分析;第4节对本文进行总结,并给出下一步的研究计划。

1 相关研究

早期关于对话回应选择的研究主要侧重于短文本的单回合对话。2013年 Wang 等人^[11]提出了相关数据集和一个基于向量空间和语义匹配的方法。2014年 Ji 等人^[12]提出使用深度神经网络来匹配对话上文和回应之间的语义信息的方法。2016年 Wu 等人^[13]提出了一个用于短文本回应选择的主题感知卷积神经网络框架。这些方法在单回合回应选择任务取得了一定成果,但是无法解决多回合问题。

当前的研究主要集中于多轮对话的回应选择任务。该任务更具挑战性,因为模型需要整合整个对

话上文中信息。2015年 Lowe^[14]提出一个大型多轮对话数据集 UDC(Ubuntu Dialogue Corpus),此外 Lowe 提出使用 RNN 编码对话上文和备选回应的方法,此类方法又被称为平行编码方法。不久, Kadlec 等人^[15]研究了不同种类编码器在平行编码网络上的性能。2016年 Yan 等人^[16]采用了另一种做法,他们使用一个 CNN 计算对话上文和备选回应的匹配分数。但这些方法处理信息相对简单,没有充分挖掘对话上文和回应之间的深层语义信息。

之后的研究人员对平行编码方法进行了各种拓展。2016年 Zhou 等人^[17]采用了两个并行的编码器,一个处理单词级别的信息,另一个处理话语级别的信息。2016年 Wang 等人提出 Match_LSTM^[18]模型,通过基于 LSTM 的注意力加权句子表示法来提升模型的性能。Tan 等人提出的 QA-LSTM^[19],使用一个简单的注意力机制与 LSTM 编码器相结合的方法。此外,研究人员还提出一些将回应选择任务与句子评分任务相结合的方法。这些方法在挖掘文本信息方面均取得了一定的成果,但是没有将得到的信息进行充分整合,回应选择任务要求模型选出合适的回应,因此,模型必须能够充分整合出对话上文和备选回应之间的关联信息。

2017年 Wu 等人提出顺序匹配网络(Sequential Matching Network, SMN^[20]),将备选回应分别与对话上文中的每一次对话匹配,2018年 Zhang 等人^[21]提出深度话语汇聚网络(Deep Utterance Aggregation Network, DUA),该网络细化处理对话,同时使用自注意力机制来寻找每次对话中的重要信息。Zhou 等人^[22]提出深度注意力匹配网络(Deep Attention Matching network, DAM),该网络构造了具有不同的粒度的表示形式,并具有堆叠式的自我注意力和交叉注意力。2019年 Tao 等人^[23]提出交互模型(Interaction Over Interaction Model, IOI),该模型通过堆叠多个交互块来执行匹配。以上四种做法主张将每一次对话与备选回应进行匹配,使用了前文对话信息,但是忽视了对话之间的语义信息。

现有的工作取得了一定成果并推动了多回合对话回应选择任务的发展,但已有研究存在对话历史和回应间关联关系挖掘不够、潜在语义利用及多种信息整合效率不足的问题。对此,本文提出 AICRS 方法,借助交叉注意力机制挖掘对话历史和回应间的关联关系,借助多头自注意力机制和高速路神经

网络进行丰富信息的整合。

2 AICRS 模型

多回合对话回应选择任务要求模型在备选回应池中为对话上文选择合适的回应。本文将多回合对话回应选择任务转化为分类任务,要求模型在得到对话上文和备选回应的情况下判断备选回应是否为对话上文的合适回应。

2.1 任务描述

给定数据集 $D = \{(y_i, c_i, r_i)\}_{N_{i=1}}$, 其中, c_i 表示对话上文, r_i 表示备选回应, $y_i \in \{0, 1\}$ 表示标签, $y_i = 1$ 表示 r_i 是 c_i 合适的回应, $y_i = 0$ 表示 r_i 不是 c_i 合适的回应。检索式对话的目标是根据数据集 D 训练出一个匹配模型 $s(\cdot, \cdot)$ 。对于任意一个对话上文-回应对 (c, r) , 匹配模型都能给出对话上文 c 和回应 r 之间的匹配分数。

2.2 模型概述

图 1 为 AICRS 模型框架图, 该模型主要有以下结构:

(1) **编码层**: 首先将对话上文拼接成一个长文, 保证对话的序列特性, 之后使用 BiGRU^[24] 编码器, 获得长文和备选回应的内部信息。

(2) **嵌入交叉注意力机制的匹配层**: 使用交叉注意力机制获取对话上文和备选回应相关的信息, 并对获得的信息进行第一次整合。

(3) **融入多头自注意力机制的信息桥接层**: 首先, 使用多头自注意力机制深度挖掘对话上文的语义信息, 之后, 使用高速路神经网络对信息进行桥接。

(4) **输出层**: 首先, 融合得到的语义信息, 之后, 对结果进行预测。

下面我们将逐步介绍 AICRS 模型中的各个组成部分。

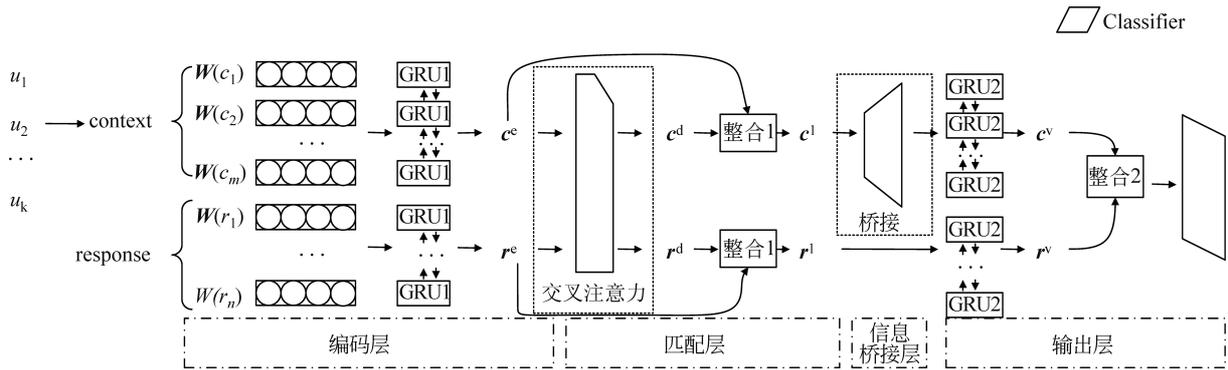


图 1 AICRS 模型框架

2.3 编码层

本文首先将对话上文拼接成一个长文并控制长文的长度, 对话之间用特殊的连接符连接, 这样做可以在编码的时候尽量将整个对话上文的语义信息整合成一个语义向量, 相比复杂的层次编码, 该做法在一定程度上保证了挖掘到的对话上文语义信息的完整。此时, 对话上文可以表示为: $c = (c_1, \dots, c_m)$, 备选的回应可以表示为: $r = (r_1, \dots, r_n)$, m 表示拼接后对话上文的长度, n 表示备选回应的长度。之后, 使用预训练词向量 $\mathbf{R}^{d_e \times |V|}$ 将每个词转为其对应的词向量。特别的, 如果某个词在表中不存在, 则赋予一个随机值。此时, 对话上文 c 和回应 r 变成词向量序列 $\mathbf{W}(c) = [\mathbf{W}(c_1), \mathbf{W}(c_2), \mathbf{W}(c_3), \dots,$

$\mathbf{W}(c_m)], \mathbf{W}(r) = [\mathbf{W}(r_1), \mathbf{W}(r_2), \mathbf{W}(r_3), \dots,$
 $\mathbf{W}(r_n)]$ 。

为了得到对话上文和备选回应的语义向量 c^e, r^e , 本文将对话上文和回应向量序列传入 BiGRU, 具体计算如式(1)、式(2)所示。

$$c_i^e = \text{BiGRU}_1(\mathbf{W}(c), i) \quad (1)$$

$$r_j^e = \text{BiGRU}_1(\mathbf{W}(r), j) \quad (2)$$

其中, i 和 j 表示对话长文中的第 i 个单词和回应中的第 j 个单词, 式(1)和式(2)中的 BiGRU_1 是同一个。

2.4 嵌入交叉注意力机制的匹配层

基于编码的方法在得到对话上文和回应的语义信息后直接将两个向量拼接, 传入输出层, 这样做只

挖掘到了对话上文和回应本身的浅层语义信息,并没有将二者的交互语义信息挖掘出来。而能否对对话上文和回应之间的局部语义关系进行正确建模是确定回应是否为正确回应的关键步骤。正确的回应一定与对话上文中一些关键信息相关,这些关键信息可以通过本地建模得到。本文使用交叉注意力机制将对话上文与备选回应进行对齐,计算对话上文与备选回应之间单词级别的语义信息,具体如图2所示。

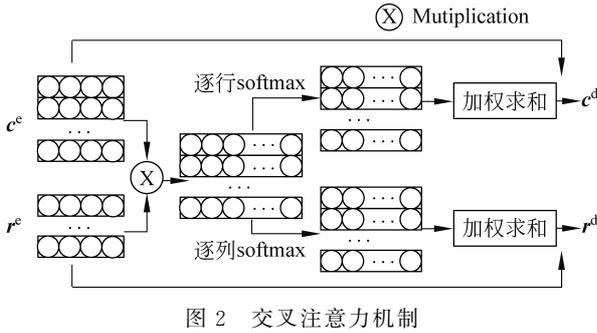


图2 交叉注意力机制

交叉注意力机制权重的计算如式(3)所示。

$$e_{ij} = (\mathbf{c}_i^e)^T \mathbf{r}_j^e \quad (3)$$

本文使用软对齐^[25]获取对话上文和回应之间的局部相关性,其通过式(8)中的注意力矩阵 $\mathbf{e} \in \mathbb{R}^{m \times n}$ 计算得到。对话上文中第 i 个单词的隐藏层向量 \mathbf{c}_i^e 与备选回应中相关的语义部分被标识为向量 \mathbf{c}_i^d ,称之为 \mathbf{c}_i^e 的对偶向量,这个向量表示所有相关信息的加权和。备选回应对偶向量的计算方法和对话上文对偶向量的计算方法相同,如式(4)、式(5)所示。

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})}, \mathbf{c}_i^d = \sum_{j=1}^n \alpha_{ij} \mathbf{r}_j^e \quad (4)$$

$$\beta_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^m \exp(e_{kj})}, \mathbf{r}_j^d = \sum_{i=1}^m \beta_{ij} \mathbf{c}_i^e \quad (5)$$

其中, $\boldsymbol{\alpha} \in \mathbb{R}^{m \times n}$, $\boldsymbol{\beta} \in \mathbb{R}^{m \times n}$ 表示标准化注意力权重矩阵。

为了融合已经得到的所有相关信息,本文使用启发式匹配方法分别处理对话上文和回应语义向量以获得局部匹配向量 \mathbf{c}_i^l 和 \mathbf{r}_j^l 。具体如式(6)、式(7)所示。

$$\mathbf{c}_i^l = F([\mathbf{c}_i^e; \mathbf{c}_i^d; \mathbf{c}_i^e - \mathbf{c}_i^d; \mathbf{c}_i^e \cdot \mathbf{c}_i^d]) \quad (6)$$

$$\mathbf{r}_j^l = F([\mathbf{r}_j^e; \mathbf{r}_j^d; \mathbf{r}_j^e - \mathbf{r}_j^d; \mathbf{r}_j^e \cdot \mathbf{r}_j^d]) \quad (7)$$

其中, $[\]$ 表示向量的拼接操作, F 是使用 ReLU 降维的单层前馈神经网络。

2.5 融入多头自注意力机制的信息桥接层

匹配层将编码后的语义向量和经过交叉注意

力机制的语义信息进行整合。得到的语义向量包含丰富的语义信息。考虑到信息过于饱和和可能会起到反作用,本文特地使用多头自注意力机制对 \mathbf{c}_i^l 进行信息挖掘、整合,之后使用高速路神经网络对经过多头自注意力机制的信息和没有经过多头自注意力机制的信息桥接,在深度挖掘文本信息的同时保证信息的完整性,具体流程如图3所示。

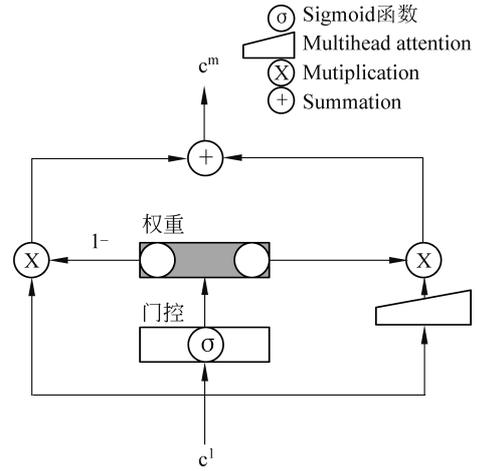


图3 信息桥接

具体计算如式(8)~式(11)所示。

$$\mathbf{c}^{m1} = \text{MultiHead}(\mathbf{c}^l, \mathbf{c}^l, \mathbf{c}^l) \quad (8)$$

$$\mathbf{r}^m = \text{MultiHead}(\mathbf{r}^l, \mathbf{r}^l, \mathbf{r}^l) \quad (9)$$

$$\text{Gate}_1 = \text{sigmoid}(\mathbf{W} \cdot \mathbf{c}^l + \mathbf{b}) \quad (10)$$

$$\mathbf{c}^m = \text{Gate}_1 \cdot \mathbf{c}^l + (1 - \text{Gate}_1) \cdot \mathbf{c}^{m1} \quad (11)$$

其中,高速路神经网络的门控机制由 sigmoid 单元组成。 \mathbf{W} 和 \mathbf{b} 分别表示权重矩阵与偏置项, Gate_1 表示控制 \mathbf{c}^l 的门控,为了使得到的信息向量能进行正则化处理,本文特地将控制 \mathbf{c}^{m1} 的门控设置为 $1 - \text{Gate}_1$ 。

2.6 输出层

输出层的主要工作是对前面得到的信息进行整合,输出最终的答案。首先,将 \mathbf{r}_j^m 和 \mathbf{c}_i^m 传入一个 BiGRU 中,如式(12)、式(13)所示。

$$\mathbf{c}_i^v = \text{BiGRU}_2(\mathbf{c}^m, i) \quad (12)$$

$$\mathbf{r}_j^v = \text{BiGRU}_2(\mathbf{r}^m, j) \quad (13)$$

其中,两式中的 BiGRU_2 是同一个。

其次,对 BiGRU_2 的输出进行最大化和平均化处理;最后,将处理后的向量合并成固定长度的向量,并传入分类器中,如式(14)~式(19)所示。

$$\mathbf{c}_{\max}^v = \max(\mathbf{c}^v, 0) \quad (14)$$

$$\mathbf{r}_{\max}^v = \max(\mathbf{r}^v, 0) \quad (15)$$

$$\mathbf{c}_{\text{mean}}^v = \text{mean}(\mathbf{c}^v, 0) \quad (16)$$

$$\mathbf{r}_{\text{mean}}^v = \text{mean}(\mathbf{r}^v, 0) \quad (17)$$

$$\mathbf{f} = [\mathbf{c}_{\text{max}}^v; \mathbf{c}_{\text{mean}}^v; \mathbf{r}_{\text{max}}^v; \mathbf{r}_{\text{mean}}^v] \quad (18)$$

$$\text{label} = \text{softmax}(\mathbf{W}_2 \cdot \tanh(\mathbf{W}_1 \cdot \mathbf{f} + \mathbf{b}_1) + \mathbf{b}_2) \quad (19)$$

其中，“0”表示在第 0 维上做 max 和 mean，“[]”表示向量的拼接操作，label 表示最终预测的结果， \mathbf{W}_1 和 \mathbf{b}_1 分别表示 tanh 层的权重和偏置， \mathbf{W}_2 和 \mathbf{b}_2 分别表示 softmax 层的权重和偏置。

2.7 优化策略

在模型训练的过程中，本文选择交叉熵误差作为损失函数，如式(20)所示。

$$\text{Loss}(\hat{y}, y) = - \sum_{s=1}^S \sum_{c=1}^C y_s^c \cdot \log \hat{y}_s^c \quad (20)$$

其中， y 表示真实答案， \hat{y} 是模型预测的概率， S 是训练样本的总数， C 是类别的数目。同时，本文使用 Adam(Adaptive Moment Estimation^[26])算法优化模型参数。

3 实验设置与结果分析

本节将分析所提出的方法在多轮对话回应选择任务上的实验结果及分析。

3.1 实验数据集

本文主要使用 Ubuntu Corpus V1^[14]数据集验证所提出的方法，该数据集中的对话主要是关于 Ubuntu 系统故障排除的多回合对话，训练集包含 50 万组多回合对话，每组对话都配对一个由人工产生的正确回应，一个随机产生的负回应，验证集和测试集都包含 5 万组多回合对话，每组对话都配对 1 个正确回应，9 个负回应。

为了弄清楚数据集中不同类型问题的分布，我们从以下角度对数据集中对话数据进行统计。首先，根据对话是否需要整合前文信息，我们将问题分为以下两种情况：单回合对话，通过最近一次对话语义信息即可得到合适的回应，如表 1 中例子所示，回应可以直接通过最近的一次对话“How it helps”直接得到；多回合对话，必须将对话上文信息整合才能得到合适的回应，如表 2 中的例子所示，想要找到合适的回应就必须找到最近一次对话“What is that like?”中“What”代指什么。

表 1 单回合对话样例

| Speaker | Sentences |
|---------|--|
| A | What different configuration it s just __number__ file |
| B | Yes i got it but if a new hardware appears how do they know about it |
| A | You don't — if new hardware appears the package maintainers update the file — you get the update file when you want to |
| B | Thank you ikonia i want to know how it works more than knowing that its just a file |
| A | What do you mean how it works. lspci is a binary application that reads device id s from a flat file and offers a human readable output |
| B | How it helps |
| Answer | How it helps ??? it shows you information about your hardware thats it yes i know that what i mean is i want to know how different hardwares are identified |

表 2 多回合对话样例

| Speaker | Senences |
|---------|---|
| A | Hi I am looking to see what packages are installed on my system, I don't see a path, is the list being held somewhere else? |
| B | Try dpkg — get-selections |
| A | What is that like? A database for packages instead of a flat file structure? |
| Answer | dpkg is the debian package manager — Get - selections simply shows you what packages are handed by it |

本文在数据集中随机抽取 100 个样例，统计每种问题的比重。统计结果如表 3 所示。从表 3 中可以看出，数据集中存在一定比例的单回合对话，但是多回合对话的比重明显更大。

根据对话长度是否均匀，我们将问题分为以下两种情况：对话长度均匀，对话之间长度相差不超过 10 个字符；对话长度不均匀，对话之间长度相差超过 10 个字符。上述 100 个随机抽取样例中两种情况的统计结果如表 4 所示。从表 4 中可以得出，数据集中有一定比例的问题对话长度相差较大。如果对每句话单独处理，在训练过程中就必须将对话填充成相同长度，这无疑增加了模型的计算量。

表3 单回合与多回合样例比重

| 问题类别 | 占比/% |
|------|------|
| 单回合 | 36 |
| 多回合 | 64 |

表4 对话长度分布是否均匀样例比重

| 问题类别 | 占比/% |
|---------|------|
| 语句长度均匀 | 77 |
| 语句长度不均匀 | 23 |

为了充分了解数据集的特点,本文随机抽取100个样例,对其中对话的回合数平均值(取整后)和对话上文长度的平均值(取整后)进行了统计,结果见表5。从表5中的数据可以得出,对话上文多为多回合、短文本。因此本文没有使用层次匹配的方法,而是首先将多回合对话拼接成一个长文,保证上文对话的序列特性。其次,使用编码器挖掘其浅层语义信息。再次,使用交叉注意力机制挖掘对话上文和回应之间相关的语义信息,并用信息桥接方法对挖掘到的方法进行整合。最后,获取结果。

表5 样例中对话的回合数和对话上文长度平均值

| 类别 | 平均值 |
|----------|-----|
| 平均回合数 | 18 |
| 对话上文平均长度 | 230 |

本文的方法是在分析已有研究的问题和Ubuntu Corpus V1数据集的特点的基础上提出的。为了验证本文方法的普适性,我们还在Douban Conversation Corpus^[20]数据集上验证了本文的方法,该数据集是从豆瓣中获取的开放域对话。

3.2 实验设置

实验采用Pytorch 0.4.0框架,并用NVIDIA的1080 GPU进行加速。具体的模型参数配置为:使用Word2Vec预训练词向量进行初始化,word_dim为300。BiGRU的输入是一个形状为[batch_size, seq_length, hidden]的张量,其中,第一个维度表示batchsize,训练中batch size设置为16;第二维表示batchsize个句子中最大句子长度,每个batch size的seq_length是不相同的,对话上文的seq_length设置为400,回应的seq_length设置为150;第三维表示隐藏层维数,实验中,hidden设置为300。Bi-GRU的输出是一个形状为[batch_size, seq_

length, hidden × 2]的张量,使用Adam(Adaptive Moment Estimation)算法优化模型参数,学习率lr设置为0.004, BiGRU的层数设置为1, dropout设置为0.5, Multi-head Attention中的多头设置为4, 损失函数为交叉熵损失函数。

3.3 实验结果

模型被要求从备选的10个回应中选出最适合的k个回应。本文通过计算k个答案中正确答案的召回率作为主要的评价指标,如式(21)所示。

$$R_{10}@k = \frac{\sum_{i=1}^k y_i}{\sum_{i=1}^{10} y_i} \quad (21)$$

考虑到Douban数据集中的回应不止一个,我们增加MAP, MRR和P@1作为Douban数据集的评价指标。

本文选取的对比模型有:

(1) 基于句子编码的方法: BiLSTM^[15]。首先,编码对话上文和各选回应;然后,计算对话上文和回应之间的语义相似度。

(2) 基于序列匹配的方法: MV-LSTM^[27]和Match-LSTM^[18]。使用注意力机制计算对话上文和回应之间单词级别的信息。

(3) 复杂的基于层次的方法: Multi-View^[17], 分别计算对话上文与回应之间单词级别和句子级别的相似语义; DAM^[22], 使用堆栈式自我注意模块和交叉注意模块获取对话上文中每句话与回应在不同级别的交互信息; IOI^[23]通过堆叠多个交互块来执行匹配。

表6给出了本文模型和各个模型的实验结果。从结果可以看到,本文给出的方法虽然结构简单,但在Ubuntu Corpus V1数据集上的性能与复杂的基于层次编码的方法相比并不逊色,在R₁₀@1和R₁₀@2两个指标上均取得了最佳性能。

但在Douban Conversation Corpus数据集上,本文方法取得的效果不如层次匹配模型。分析其中原因,我们认为,Ubuntu Corpus V1数据集是一个任务驱动型对话数据集,其话题主要集中在Ubuntu系统方面,噪声相对较少;而Douban Conversation Corpus数据集是一个开放领域对话数据集,其话题面向开放领域,噪声相对较多。我们的模型结构较为简单,处理噪声的能力有限,同时Douban Conversation Corpus中对话上文相对较长,而本文的模型采用将对话上文拼接成一个长文

之后对长文进行编码的方法。当长文达到编码器的极限时,模型性能将会下降。未来我们将针对

这一问题探讨相应的解决方案,提升本文方法的普适性。

表 6 各个模型的实验结果

| 模型 | Ubuntu Corpus V1 | | | Douban Conversation Corpus | | | | | |
|----------------------------|------------------|----------------|------------|----------------------------|-------|--------------|------------|------------|------------|
| | $R_{10}@1$ | $R_{10}@2$ | $R_{10}@5$ | MAP | MRR | P@1 | $R_{10}@1$ | $R_{10}@2$ | $R_{10}@5$ |
| BiLSTM ^[15] | 0.630 | 0.780 | 0.944 | 0.479 | 0.514 | 0.313 | 0.184 | 0.330 | 0.716 |
| MV-LSTM ^[27] | 0.653 | 0.804 | 0.946 | 0.498 | 0.538 | 0.348 | 0.202 | 0.351 | 0.710 |
| Match-LSTM ^[18] | 0.653 | 0.799 | 0.944 | 0.500 | 0.537 | 0.345 | 0.202 | 0.348 | 0.720 |
| Multi-View ^[17] | 0.662 | 0.801 | 0.951 | 0.505 | 0.543 | 0.342 | 0.202 | 0.350 | 0.729 |
| DAM ^[22] | 0.767 | 0.874 | 0.969 | 0.550 | 0.601 | 0.427 | 0.254 | 0.410 | 0.757 |
| IOI ^[23] | 0.796 | 0.894 | 0.974 | 0.573 | 0.621 | 0.444 | 0.269 | 0.451 | 0.786 |
| AICRS | 0.886 6 | 0.900 6 | 0.951 5 | 0.545 | 0.598 | 0.425 | 0.255 | 0.425 | 0.771 |

3.4 实验分析

为了验证不同模块的作用,本文在 Ubuntu Corpus 数据集上设置了四个不同的对比实验。

(1) B: 使用 BiGRU 编码,并把编码后的信息直接传入输出层。

(2) B+交叉注意力: 使用 BiGRU 编码,使用交叉注意力机制。

(3) B+交叉注意力+残差: 使用 BiGRU 编码,使用交叉注意力机制,使用多头自注意力机制和残差网络对整合后的信息进行深度挖掘和整合。

(4) AICRS: 使用 BiGRU 编码,使用交叉注意力机制,使用多头自注意力机制和高速路神经网络对整合后的信息进行桥接。

详细对比实验结果如表 7 所示。

表 7 详细实验对比结果

| 模型 | $R_{10}@1$ | $R_{10}@2$ | $R_{10}@5$ | 训练时间/h |
|--------------|----------------|----------------|----------------|-----------|
| B | 0.6891 | 0.783 2 | 0.944 0 | 14 |
| B+交叉注意力 | 0.855 1 | 0.894 4 | 0.944 6 | 15 |
| B+交叉注意力+残差 | 0.847 9 | 0.860 5 | 0.931 0 | 15 |
| AICRS | 0.886 6 | 0.900 6 | 0.951 5 | 15 |

3.4.1 交叉注意力机制效用分析

比较 B 和 B+交叉注意力的结果,可以得出使用交叉注意力机制的模型相比没有使用交叉注意力机制的模型,模型性能在 $R_{10}@1$ 和 $R_{10}@2$ 上分别提升 16.6%和 11.12%,说明交叉注意力机制在挖掘对话上文和备选回应之间的相关语义信息方面发挥

了不可忽视的作用,模型在整合到对话上文和备选回应之间相关联信息之后,性能有了可观的提升。

3.4.2 信息桥接效用分析

比较 B+交叉注意力和 AICRS 可以看出:模型在没有进行信息桥接的情况下,性能在 $R_{10}@1$, $R_{10}@2$ 和 $R_{10}@5$ 上分别下降 3.15%, 0.62%和 0.69%,说明信息桥接在整合对话上文信息时起到了一定的作用。比较 B+交叉注意力和 B+交叉注意力+残差:可以看出,使用残差连接替代信息桥接,模型的性能在 $R_{10}@1$, $R_{10}@2$ 和 $R_{10}@5$ 上反而下降了 0.72%, 3.39%和 1.36%,同为连接,残差连接虽然将多种信息进行了组合,但是此时的语义向量包含十分复杂的信息,如果不做进一步的处理:反而会对模型的性能起反作用。

3.4.3 模型处理过长对话上文能力分析

为了分析模型处理过长对话上文的能力,本文在 Ubuntu Corpus V1 数据集中随机抽取 100 个对话上文较长的样例,分析这 100 个样例在模型上的结果。实验表明,本文的模型处理过长对话上文能力不是很好。100 个样例只有 27 个可以准确预测,其余 73 个均没能预测出来,分析原因,我们发现这是编码器造成的。当文本的长度达到编码器的极限后,编码器便不能有效地挖掘文本的信息,进而导致性能的下降。未来我们将进一步细化对话历史的表征,特别是长距离依赖问题,以弥补这一缺陷。

4 结论

本文借助多头自注意力机制多视角挖掘潜在语

义,借助交叉注意力机制建立对话历史和回应间的关联,再借助高速路神经网络进行多种丰富信息的高效融合,据此构建了一个轻量级对话回应选择模型。相比基准方法,我们的系统简单有效,实验结果也证明给出的模型确实能提升对话回应选择的性能。

未来我们将进一步细化对话历史的表征,计划从挖掘历史对话中的全局和局部信息更好地融入对话的交互信息、时序信息等方面入手,尝试通过有效解决长距离依赖问题来进一步提升回应选择的性能。

参考文献

- [1] Saygin A P, Cicekli I. Pragmatics in human-computer conversations[J]. *Journal of Pragmatics*, 2002,34(3): 227-258.
- [2] Manning C D, Manning C D, Schütze H. *Foundations of statistical natural language processing*[M]. NY: MIT press, 1999.
- [3] Wen T H, Vandyke D, Mrkšić N, et al. A network-based end-to-end trainable task-oriented dialogue system[C]//*Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 2017, 1: 438-449.
- [4] Higashinaka R, Funakoshi K, Araki M, et al. Towards taxonomy of errors in chat-oriented dialogue systems[C]//*Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2015: 87-95.
- [5] Guttormsen M, Bürger A, Hansen T E, et al. The Si-Ri particle-telescope system[J]. *Nuclear Instruments and Methods in Physics Research A*, 2011, 648: 168-173.
- [6] Hoy M B. Alexa, Siri, Cortana, and more: an introduction to voice assistants[J]. *Medical Reference Services Quarterly*, 2018, 37(1): 81-88.
- [7] Shum H, He X, Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots[J]. *Frontiers of Information Technology and Electronic Engineering*, 2018, 1(19): 10-26.
- [8] Hao Y, Zhang Y, Liu K, et al. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge[C]//*Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2017: 221-231.
- [9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, 2017: 5998-6008.
- [10] Srivastava R K, Greff K, Schmidhuber J. Highway networks[J]. *arXiv preprint arXiv:1505.00387*, 2015.
- [11] Wang H, Lu Z, Li H, et al. A dataset for research on short-text conversations[C]//*Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2013: 935-945.
- [12] Ji Z, Lu Z, Li H. An information retrieval approach to short text conversation[J]. *arXiv preprint arXiv:1408.6988*, 2014.
- [13] Wu Y, Wu W, Li Z, et al. Topic augmented neural network for short text conversation[J]. *arXiv preprint arXiv:1065.00090v2*, 2016.
- [14] Lowe R, Pow N, Serban I, et al. The ubuntu dialogue corpus; a large dataset for research in unstructured multi-turn dialogue systems[C]//*Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Prague: the Association for Computer Linguistics, 2015: 285-294.
- [15] Kadlec R, Schmid M, Kleindienst J. Improved deep learning baselines for Ubuntu corpus dialogs[J]. *arXiv preprint arXiv:1510.03753*, 2015.
- [16] Yan R, Song Y, Wu H. Learning to respond with deep neural networks for retrieval-based human-computer conversation system[C]//*Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2016: 55-64.
- [17] Zhou X, Dong D, Wu H, et al. Multi-view response selection for human-computer conversation[C]//*Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Austin: the Association for Computational Linguistics, 2016: 372-381.
- [18] Wang S, Jiang J. Learning natural language inference with LSTM[C]//*Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016: 1442-1451.
- [19] Tan M, dos Santos C, Xiang B, et al. LSTM-based deep learning models for non-factoid answer selection[J]. *arXiv preprint arXiv:1511.04108*, 2015.
- [20] Wu Y, Wu W, Li Z, et al. Sequential matching network; a new architecture for multi-turn response selection in retrieval-based chatbots[C]//*Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Vancouver: Association for Computational Linguistics, 2017: 496-505.
- [21] Zhang Z, Li J, Zhu P, et al. Modeling multi-turn conversation with deep utterance aggregation[C]//*Proceedings of the 27th International Conference on Computational Linguistics*. 2018: 3740-3752.

- [22] Zhou X, Li L, Dong D, et al. Multi-turn response selection for chatbots with deep attention matching network[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 1118-1127.
- [23] Tao C, Wu W, Xu C, et al. One time of interaction may not be enough: go deep with an interaction-over-interaction network for response selection in dialogues [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 1-11.
- [24] Dey R, Salem F M. Gate-variants of gated recurrent unit (GRU) neural networks[C]//Proceedings of the IEEE 60th International Midwest Symposium on Circuits and Systems. IEEE, 2017: 1597-1600.
- [25] 韩松伯. 基于深度神经网络的英文文本蕴含识别研究[D]. 北京: 北京邮电大学硕士学位论文, 2018.
- [26] Kingma D P, Ba J L. Adam: a method for stochastic optimization[C]//Proceedings of the 3rd International Conference on Learning Representations. San Diego, 2015.
- [27] Wan S, Lan Y, Guo J, et al. A deep architecture for semantic matching with multiple positional sentence representations[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix : AAAI Press, 2016: 2835-2841.



司博文(1995—), 硕士研究生, 主要研究领域为自然语言处理、机器阅读理解、对话回应选择。
E-mail: 20185227065@stu.suda.edu.cn



孔芳(1977—), 博士, 主要研究领域为机器学习、自然语言处理、篇章分析。
E-mail: kongfang@suda.edu.cn

~~~~~  
(上接第 84 页)



林健成(1995—), 硕士, 主要研究领域为自然语言处理、对话系统等。  
E-mail: rabbit9510@outlook.com



蔺晓川(1996—), 硕士, 主要研究领域为移动群智感知、推荐系统托攻击、自然语言处理等。  
E-mail: lxc946970811@outlook.com