融入对话上文整体信息的层次匹配回应选择

司博文
苏州大学
计算机科学与技术学院
20185227065@stu.suda.edu.cn

孔芳*
苏州大学
计算机科学与技术学院
kongfang@suda.edu.cn

摘要

对话是一个顺序交互的过程，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，但仍然存在两个突出的问题。一是现有的编码器在挖掘对话文本语义信息上尚存在不足；二是只考虑每一回合对话与备选回应之间的关系，忽视了对话上文的整体语义信息。针对问题一，本文借助多头自注意力机制有效捕捉对话文本的语义信息；针对问题二，整合对话上文的整体语义信息，分别从单词、句子以及整体对话上文三个层次与备选回应进行匹配，充分保证匹配信息的完整。在Ubuntu Corpus V1和Douban Conversation Corpus数据集上的对比实验表明了本文给出方法的有效性。

关键词： 回应选择；多头自注意力机制；交叉注意力机制；对话整体信息

Learning Overall Dialogue Information for Dialogue Response Selection

Bowen Si
School of Computer
Science and Technology
Soochow University
20185227065@stu.suda.edu.cn

Fang Kong *
School of Computer
Science and Technology
Soochow University
kongfang@suda.edu.cn

Abstract

Dialogue is a sequential interactive process. Response selection aims to select the appropriate response based on the existing dialogue, which is a research hotspot in the field of natural language processing. Existing researches have achieved some success, but there are two problems. First, the existing encoders still have deficiencies in mining the semantic information of the dialogue text; second, the existing research only considers the relationship between each round of dialogue and the alternative response, ignoring the overall semantic information above the dialogue. For problem one, this article uses the multi-head self-attention mechanism to capture the semantic information of the dialogue text effectively; for problem two, integrating semantic information above the dialogue, then matching context and response from the words, sentences, and the overall dialogue levels which can fully guarantee the completeness of matching information. Experiments on the Ubuntu Corpus V1 and Douban Conversation Corpus datasets exhibit the effectiveness of the method presented in this article.

基金项目：国家自然科学基金面上项目（61876118）；国家自然科学基金人工智能应急管理项目(61751206)
*通信作者：kongfang@suda.edu.cn

第十九届中国计算语言学大会论文集，第266页-第276页，海南，中国，2020年10月30日至11月1日。

Keywords: Response selection , Self-attention , Cross-attention , Overall Dialogue Information

1 引言

人机对话（human-computer conversations(Saygin and Cicekli, 2002)），旨在促进人类与机器自然交流，是自然语言处理（natural language process,NLP(Manning et al., 1999)）领域的关键任务。该任务要求模型依据已有的对话上文，产生与对话上文相匹配的回应。当前，有两种产生回应的方法：生成式和检索式。生成式又称回应生成（response generation(Ritter et al., 2011)），旨在使用自然语言生成技术生成与对话上文相匹配的回应；检索式又称回应选择（response selection(Rowe et al., 2000)），旨在从备选回应中为对话上文选出合适的回应。这两种方法产生的回应都与对话上文密切相关，因此如何更好的整合对话上文的信息一直是这两个任务的研究重点。本文主要关注回应选择任务，因为，回应选择任务存在信息量丰富和对话内容流畅等优势。许多工业界的产品采用的都是检索式对话系统，例如微软的小冰(Zhou et al., 2020)、阿里的小蜜(Li et al., 2017)等。

已有的回应选择方法存在对话历史的语义挖掘不充分，以及只注重每一回合对话与备选回应之间的关联信息，从而忽视了完整对话上文信息的问题。针对这两个问题，本文提出了融入整体对话上文信息的回应选择方法，记为IODIRS（Integrate the Overall Dialogue Information for Response Selection），具体而言，首先，使用多头自注意力机制（Multi-head attention Mechanism(Vaswani et al., 2017)）挖掘对话文本的潜在语义信息；其次，计算每一回合对话与备选回应在单词和回合级别的关联信息，并将每个回合和备选回应的关联信息拼接到一个矩阵中；接着，使用交叉注意力机制（Cross-attention Mechanism(Hao et al., 2017)）从上到下计算对话上文的整体语义信息，挖掘整体关联信息的同时保证对话历史的序列特性；之后，将对话上文的整体语义信息与备选的回应进行关联，并将关联的信息拼接到上述矩阵中；最后整合各类信息进行最终的回应选择。该方法在深度挖掘每一回合对话文本语义信息的同时将对话上文的整体信息融入该任务中，从而提升回应选择的性能。

本文的贡献包括如下几点：（1）提出融入对话上文整体语义信息的回应选择模型。（2）使用多头自注意力机制挖掘每一回合对话的语义信息。（3）使用交叉注意力机制从上到下挖掘对话上文之间的关联信息，并将其整合，保证对话上文序列特性的前提下，整合对话上文的整体信息，并通过Ubuntu Corpus V1(Lowe et al., 2015)和Douban Conversation Corpus(Wu et al., 2017)数据集上的实验验证模型的有效性。

本文的组织结构如下：第二节介绍相关工作；第三节阐述IODIRS模型；第四节介绍具体的实验设置，并对实验结果进行分析；第五节对本文进行总结，并给出下一步的研究计划。

2 相关研究

随着深度学习的发展，建立一个数据驱动的人机对话系统越来越受到关注。已有的研究可以划分为生成式对话和检索式对话。生成式对话又称回应生成，旨在使用自然语言生成技术生成与对话上文相关的回应。检索式对话又称回应选择，旨在从备选回应中选出与对话上文最相关的回应。相较于生成式对话，检索式对话的信息更丰富，过程更流畅，所以本文重点研究检索式对话。

早期对话回应选择的研究主要集中于短文本的单回合对话。Wang (2015)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。Ji(2014)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。Wu(2016)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。目前的研究主要有以下三种方法。

基于编码的方法，Lowe (2015)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 (2015)研究了不同种类编码器在平行编码网络上的性能。Yan(2016)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。Zhou(2016)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。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。

基于匹配的方法，Wang(2016)提出MV-LSTM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 (2015)提出的QA-LSTM，使用一个简单的注意力机制与LSTM编码器相结合的方法。这些方法在挖掘文本信息方面取得一定的成果，但是将对话上文拼接成一个长文的做法使得文本变得过长，现有的编码器处理文本的能力有限，很难学习到有效的信息，同时基于匹配的方法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，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。

层次匹配的方法，Wu(2017)提出顺序匹配网络（Sequential Matching Network，SMN），将备选回应分别与对话上文中的每一次对话匹配，Zhang(2018)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。Tao(2019)提出多种表示聚合模型（multi-representation fusion network，MRFN），该模型将对话文本在多个粒度表示，之后将每个粒度上的信息进行聚合。以上做法主张将每一次对话与备选回应进行交互，之后将交互信息进行聚合。虽然使用了对话前文信息，但是忽视了对话上文的整体语义信息。实际中，对话之间存在很多指代和省略。忽视这些信息很难有效捕捉到完整的对话上文信息。

现有的工作取得了一定成果并有效推动了多回合对话回应选择任务的发展，但已有研究存在挖掘文本语义信息能力不足和忽视对话上文整体语义信息的问题。对此，本文提出IODIRS模型，借助多头自注意力机制有效挖掘对话文本的潜在语义信息，借助交叉注意力机制从上至下挖掘对话之间的关联信息，最后将关联信息进行整合进而得到对话上文的完整语义信息。

3 IODIRS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是对话上文的合适回应。

3.1 任务描述

给定数据集$D = \{(y_i, C_i, r_i)\}_{i=1}^{N}$，其中$C_i = \{u_{i1}, u_{i2}, ..., u_{iL}\}$表示对话上文，其中对话上文有$L$回合的对话，$r_i$ 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$s(.,.)$。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。

3.2 模型概述

图1为IODIRS模型框架图，该模型主要由以下三个部分组成：

1）编码层：使用多头自注意力机制对每一回合的对话进行编码，挖掘其潜在语义信息。

2）聚合层：首先，使用交叉注意力机制从上到下计算对话上文的整体语义信息，同时计算每一回合对话与备选回应在单词和回合级别的相似矩阵。之后，使用CNN整合每一回合对话的单词和回合级别相似矩阵获取其匹配信息，最后，将每一回合对话与备选回应的匹配信息与对话整体与备选回应的匹配信息进行整合，得到最终的匹配信息。

3）输出层：融合得到的语义信息，并对结果进行预测。

3.3 编码层

诸多研究证明，注意力机制在自然语言处理领域是有效的。Vaswani(2017)提出一个基于注意力机制的生成网络Transformer。相比RNN模型，Transformer不仅取得了更好的生成结果，而且其训练速度也非常快，因为其注意力计算可以并行。已有工作显示注意力机制在挖掘文本潜在语义信息上的卓越性能。因此，我们使用多头自注意力机制对文本进行编码。

任务中，对话上文可以表示为：$C = (u_1, ..., u_L)$ ，其中，$u_i = (u_{i1}, ..., u_{im})$ ，备选的回应可以表示为：$r = (r_1, ..., r_n)$，L表示对话上文的回合数，m表示每一回合对话文本的长度，n表示备选回应的长度。首先，我们使用预训练词向量$R^{d_e \times V}$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中的一回合对话u_i和回

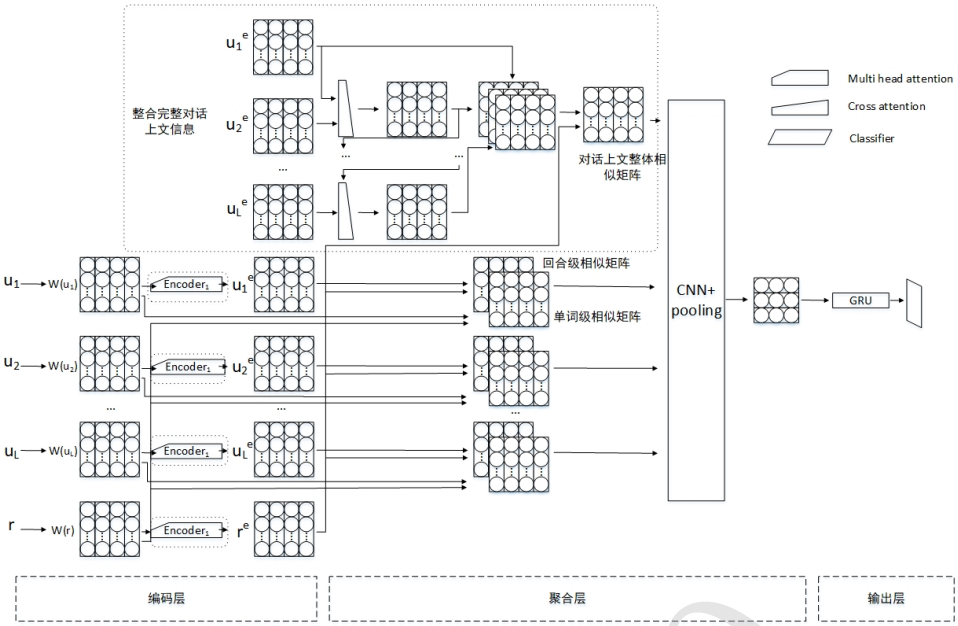

图 1 IODIRS模型

应r变成词向量序列$W(u_i) = [W(u_{i1}), W(u_{i2}), ..., W(u_{im})]$，$W(r) = [W(r_1), W(r_2), ..., W(r_n)]$。

为了得到对话上文中每一回合对话和备选回应的语义向量u_i^e，r^e，本文将对话上文和回应向量序列传入$MultiHead$，具体计算如式（1）～（2）。

$$u_i{}^e = MultiHead_1(W(u_i), W(u_i), W(u_i)) \tag{1}$$

$$r^e = MultiHead_1(W(r), W(r), W(r)) \tag{2}$$

其中，i表示对话上文中的第i个回合的对话文本，式（1）和式（2）共享一个$MultiHead_1$。

3.4 聚合层

回应选择任务的重点在于如何有效的整合对话上文的信息，并准确找到对话上文与备选回应之间的联系。正确的回应一定与对话上文中的相关信息密切相关。因此，模型能否有效整合出对话上文和备选回应之间的联系是确定回应是否是正确回应的关键步骤。已有的层次编码方法采用将每一回合对话与备选回应进行交互，并把交互的信息进行整合的方法。这样做虽然能够捕捉到每一回合对话与备选回应之间的联系，但是忽视了对话上文的整体语义信息。实际生活中，对话之间存在较多的省略和指代现象，倘若忽视这些现象，模型很难捕捉到所有关键信息。为充分挖掘对话上文与备选回应之间的语义联系，本文从对话上文整体、每一回合对话单词级别和每一回合对话整体语义三个层面将对话上文与备选回应进行交互。

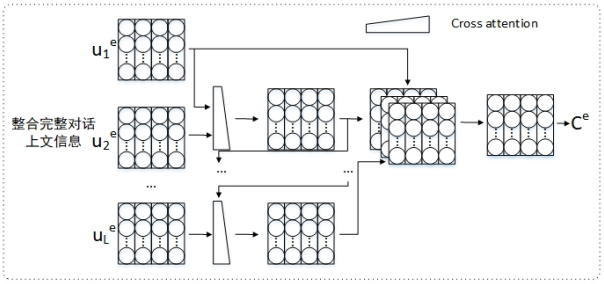

图 2 整合完整对话上文信息

3.4.1 对话上文整体语义挖掘

对于对话上文整体语义信息的挖掘，本文使用交叉注意力机制从上到下捕获对话之间的关联信息，并把关联到的信息进行整合，进而得到完整的对话上文语义信息E(C)。具体如图2所示具体计算如公式（3）～（4）所示:

$$u_{i+1}{}^{c} = CrossAttention(u_i^c, u_{i+1}^e) \tag{3}$$

$$C^e = mean_1([u_1^c; u_2^c; ...; u_L^c]) \tag{4}$$

其中$u_1^c = u_1^e$，[.] 表示向量的拼接，L表示对话上文中对话的回合数。

交叉注意力机制工作原理如下图3所示:

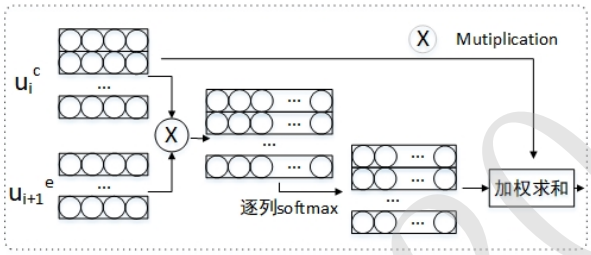

图 3 交叉注意力机制

交叉注意力权重的计算如式（5）。

$$e_{jk} = (u_{ij}^c)^T u_{(i+1)k}^e \tag{5}$$

本文使用软对齐获取对话回合之间的局部相关性，其通过式（8）中的注意力矩阵$e \in R^{m\times m}$计算得到。$u_{(i+1)}^e$中第k个单词的隐藏层向量$u_{(i+1)k}^e$与u_i^c中相关的语义部分被标识为向量$u_{(i+1)k}^l$，称之为$u_{(i+1)k}^e$的对偶向量，这个向量表示所有相关信息的加权和。具体计算如式（6）。

$$\beta_{jk} = \frac{exp(e_{jk})}{\sum_{i=1}^{m} exp(e_{ik})}, u_{(i+1)k}^l = \sum_{j=1}^{m} \beta_{jk} u_{ij}^c \tag{6}$$

其中，$\beta \in R^{m\times m}$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处理得到的语义向量。具体如式（7）～（9）。

$$u_{(i+1)k1}^l = u_{(i+1)k}^e - u_{(i+1)k}^l \tag{7}$$

$$u_{(i+1)kM}^l = u_{(i+1)k}^e \cdot u_{(i+1)k}^l \tag{8}$$

$$u_{(i+1)k}^c = F([u_{(i+1)k}^e; u_{(i+1)k}^l; u_{(i+1)k1}^l; u_{(i+1)kM}^l]) \tag{9}$$

其中，[.]表示向量的拼接操作，F是使用$ReLU$ 降低维度的单层前馈神经网络。

3.4.2 三个层次信息与备选回应的交互

为了整合有效信息，本文从对话上文整体语义，每一回合对话单词级别和每一回合对话整体语义三个层面将对话上文与备选回应进行交互。

首先，我们先计算每一回合对话文本与备选回应在单词和回合文本级别的相似信息，并将结果进行整合，得到该回合文本与备选回应的匹配向量ul_i。具体如公式（10）～（13）所示。

$$uwl_i = W(u_i) \cdot (W(r))^T \tag{10}$$

$$usl_i = u_i^e \cdot A \cdot (r^e)^T \tag{11}$$

$$ulc_i = F(Conv2d(Stack(uwl_i, usl_i))) \tag{12}$$

$$ul_i = Maxpool(ulc_i) \tag{13}$$

其中，T表示矩阵的转置，$A \in R^{m \times m}$ 表示线性变化的参数，$Stack$ 表示矩阵的堆叠，$Conv2d$(Chudanov et al., 1999)表示卷积神经网络，F 表示层前馈神经网络，$Maxpool$表示池化。

接着，我们使用同样的方法计算对话上文整体与备选回应之间的相似信息。具体如公式（14）～（16）所示。

$$Csl = C^e \cdot A \cdot (r^e)^T \tag{14}$$

$$Clc = F(Conv2d(Stack(Csl))) \tag{15}$$

$$Cl = Maxpool(Clc) \tag{16}$$

其中，$Stack$表示矩阵的堆叠，$Conv2d$表示卷积神经网络，F表示层前馈神经网络，$Maxpool$表示池化。

最后我们将每一回合的匹配信息以及对话上文整体匹配信息整合。具体如公式（17）所示。

$$match = Stack(ul_1, ul_2, ..., ul_L, Cl) \tag{17}$$

其中，$Stack$表示矩阵的堆叠，L 表示对话上文的回合数。

3.5 输出层

获取到对话上文与备选回应之间匹配信息$match$ 之后，输出层的主要工作是对上述匹配信息进行进一步整合，得到最终的匹配信息，并得到最终的匹配结果。

我们采用GRU(Dey and Salem, 2017)对$match$进行进一步编码，并使用最后一层结果作为最终的匹配向量。具体操作如公式（18）～（19）所示。

$$last = GRU(match) \tag{18}$$

$$label = Sigmoid(W_1 \cdot last + b_1) \tag{19}$$

其中，$label$ 表示预测的结果，W_1 和b_1分别表示$Sigmoid$层的权重和偏置。

3.6 优化策略

在模型训练的过程中，本文选择二分类交叉熵误差作为损失函数，具体计算如公式（20）～（21）：

$$Loss(y\prime, y) = \sum_{i=1}^{S} l_i \tag{20}$$

$$l_i = -w_i[y_i log y\prime_i - (1 - y_i) log(1 - y\prime_i)] \tag{21}$$

其中，y表示真实答案，$y\prime$是模型预测的概率，S是训练样本的总数。同时，本文使用$Adam$（Adaptive Moment Estimation(Kingma and Ba, 2015)）算法优化模型参数。

4 实验设置与结果分析

4.1 实验数据集

本文使用公开数据集Ubuntu Corpus V1和Douban Conversation Corpus数据集验证所提出的方法。Ubuntu Corpus V1数据集中的对话主要是关于Ubuntu系统故障排除的多回合英文对话。Douban Conversation Corpus数据集是从豆瓣中获取的开放域对话，其构建方式与Ubuntu Corpus V1相似。两个数据集的具体分布如表1所示。

	Ubuntu			**Douban**		
Name	Train	Val	Test	Train	Val	Test
样本数	1M	500K	500K	1M	50K	10K
备选回应数	2	10	10	2	2	10
正例数	1	1	1	1	1	1.18
回合数	10.13	10.11	10.11	6.69	6.75	6.45
回合长度	11.35	11.34	11.37	18.56	18.50	20.74

表 1　Ubuntu Corpus V1和Douban Conversation Corpus分布数据

4.2 实验设置

实验采用Pytorch 0.4.0 框架，并用$NVIDIA$的1080GPU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200。样本的回合数设置为10，$MultiHead$的输入是一个形状为$[batchsize, seqlength, hidden]$ 的张量，其中第一个维度表示$batchsize$ ，训练中不同数据集的$batchsize$设置为不同，Ubuntu设置为200。Douban设置为150；第二维表示$batchsize$ 个句子中最大句子长度，设置为50，；第三维表示隐藏层维数，实验中，$hidden$ 设置为200。$MultiHead$的输出是一个形状为$[batchsize, seqlength, hidden]$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，$dropout$ 设置为0.5，$MultiHead$ 的多头设置为4，损失函数为交叉熵损失函数，$Conv2d$ 的卷积核设置为（3，3）。本文着重验证所提方法的有效性，并没有刻意的关注模型的极限性能。因此，没有刻意对模型进行调参。

4.3 实验结果

本文主要使用数据集作者指出的评价指标作为模型性能的评价指标，其中Ubuntu Corpus V1(Lowe et al., 2015)使用$R_{10}@K$(Lowe et al., 2015)作为评价指标，Douban Conversation Corpus(Wu et al., 2017)使用$R_{10}@K$，MAP，MRR和$P@1$(Wu et al., 2017)作为评价指标。

本文选取的对比模型有：

基于句子编码的方法：BiLSTM(Lowe et al., 2015)。首先，将对话上文和回应编码；然后，计算对话上文和回应之间的语义相似度。

基于序列匹配的方法：MV-LSTM(Wang and Jiang, 2016)和Match-LSTM(Wang and Jiang, 2017)。将对话上文拼接成一个长文，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。

复杂的基于层次的方法：SMN(Wu et al., 2017)，将备选回应分别与对话上文中的每一次对话匹配，之后将匹配的信息进行聚合。DUA(Zhang et al., 2018)，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。MRFN(Tao et al., 2019)，使用多表示网络将文本特征融合。

表2给出了本文模型和各个模型的实验结果。

从表2给出的结果可以看到：

本文的方法取得了相当可观的性能。普通的层次编码方法着重强调每一轮对话与备选回应之间的关系，没有重视对话之间的联系，而实际上，人们对话之间的联系是很密切的。考虑到对话上文整体语义信息的重要性，本文使用交叉注意力机制从上到下挖掘对话上文的语义信

息，并对其进行整合。之后，从对话上文整体、每一回合对话单词级别和每一回合对话整体语义三个层面将对话上文与备选回应进行交互。相比普通的层次编码模型，本文的模型在复杂度上略高，但取得了高于普通层次编码模型的效果；同时与采用多级别表示文本特征的MFRN模型相比，本文的模型在复杂度上较低，却取得了与之相当的性能。

	Ubuntu Corpus			**Douban Conversation Corpus**					
模型	R_{10}@1	R_{10}@2	R_{10}@5	MAP	MRR	P@1	R_{10}@1	R_{10}@2	R_{10}@5
BiLSTM	0.630	0.780	0.944	0.479	0.514	0.313	0.184	0.330	0.716
MV-LSTM	0.653	0.804	0.946	0.498	0.538	0.348	0.202	0.351	0.710
Match-LSTM	0.653	0.799	0.944	0.500	0.537	0.345	0.202	0.348	0.720
SMN	0.726	0.847	0.961	0.529	0.569	0.397	0.233	0.396	0.724
DUA	0.752	0.868	0.962	0.551	0.599	0.421	0.243	0.421	0.780
MFRN	0.786	0.886	0.976	0.571	0.617	0.448	0.276	0.435	0.783
IODIRS	0.782	**0.886**	0.973	0.561	**0.617**	0.427	0.258	**0.436**	**0.791**

表 2　各个模型的结果

4.4　实验分析

为了验证不同模块的作用，本文设置了以下对比实验。

B：使用GRU编码，将每一回合对话与备选回应匹配，使用CNN聚合匹配后的信息，不考虑对话上文整体语义信息。

B+整体：使用GRU编码，使用交叉注意力机制，从上至下整合对话上文语义信息。之后从三个层面将对话上文与备选回应进行交互。

B+多头：使用MulitHead编码，将每一回合对话与备选回应匹配，使用CNN聚合匹配后的信息，不考虑对话上文整体语义信息。

IODIRS：使用MulitHead编码，使用交叉注意力机制，从上至下整合对话上文语义信息。之后从三个层面将对话上文与备选回应进行交互。

实验结果见表3。

	Ubuntu Corpus			**Douban Conversation Corpus**					
模型	R_{10}@1	R_{10}@2	R_{10}@5	MAP	MRR	P@1	R_{10}@1	R_{10}@2	R_{10}@5
B	0.755	0.865	0.961	0.558	0.597	0.425	0.255	0.424	0.753
B+多头	0.761	0.866	0.963	0.567	0.609	0.439	0.261	0.435	0.782
B+整体	0.773	0.877	0.965	**0.569**	0.608	**0.443**	**0.262**	0.434	0.780
IODIRS	**0.782**	**0.886**	**0.973**	0.561	**0.617**	0.427	0.258	**0.436**	**0.791**

表 3　详细实验对比结果

4.4.1　多头自注意力机制效用分析

比较B和B+多头的结果，可以得出使用多头自注意力机制编码的模型相比没有使用多头自注意力机制的模型，模型性能在各个评价指标上均有近1个点的提升，说明多头自注意力机制在挖掘文本潜在语义信息方面效果明显，因为多头自注意力机制可以从多角度、多层次深度挖掘文本的语义信息。

4.4.2　对话上文整体语义分析

比较B和B+整体的结果，可以得出模型在没有整合对话上文整体语义信息的情况下，性能有着近1个点的下降。说明整合到的对话上文整体语义信息在匹配回应上起到了关键作用。在实际生活中，对话之间存在较多的省略和指代现象，只有交互相关回合的对话才能有效捕捉到这些信息。而普通的层次匹配方法只重视每一回合对话与备选回应的匹配，忽视了这些信息。

4.4.3 整合后模型效用分析

上述分析了多头自注意力机制作为编码器以及整合对话上文整体语义信息对模型性能的影响。为了探究二者结合的性能，我们搭建了IODIRS模型。实验显示，相比B+多头，IODIRS模型在各个指标上均取得了提升。相比B+整体，IODIRS模型在准确率上略有下降，但是模型的整体性能却有所提升，这是因为多头自注意力机制可以挖掘到文本中不同层次，不同角度的信息，过于丰富的信息可能导致模型准确率的降低，却能提升模型整体的性能。

4.5 样例分析

为了具体说明各个模块的作用，本文验证了四个模型在一个样例上的输出，各个模型的预测结果见表4:

	样例1	B	B+多头	B+整体	IODIRS
speaker 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?				
speaker B	Try dpkg - get-selections				
speaker A	What is that like? A database for pack-ages instead of a flat file structure?				
answer	**dpkg is the debian package manager - Get - selections simply shows you what packages are handed by it**				
预测结果		**0.485**	**0.512**	**0.735**	**0.855**

表 4　对比模型在样例1上的结果

从表4中各个模型的预测结果，我们可以清晰的看出，在使用多头自注意力机制进行编码后，模型预测的准确率有着一定的提升。但因为忽视了对话之间的语义信息，未能有效的捕捉对话之间的指代信息，如“that”代指什么。在整合了对话上文整体语义信息之后，模型的准确率有着明显的提升。而本文提出的模型在使用多头自注意力机制进行编码的同时将全文信息进行整合，给出了最精确的预测结果。

5 结论

本文借助多头注意力机制多视角挖掘潜在语义，借助交叉注意力从上到下挖掘对话上文的整体语义信息，并从对话上文整体语义，每一回合对话单词级别和每一回合对话整体语义三个层面将对话上文与备选回应进行交互。据此构建了一个层次匹配对话回应选择模型。实验证明，本文的模型能够提升对话回应选择的性能。

本文将对话全文信息进行整合，考虑到全文的信息比较多，容易造成信息的冗余。未来我们将提升模型对文本信息关键信息的挖掘能力，准确寻找到相关信息，高效整合对话上文的信息。提高模型效率的同时，进一步提升回应选择的性能。

参考文献

VV Chudanov, AE Aksenova, VA Pervichko, VF Strizhov, PN Vabishchevich, and AG Churbanov. 1999. Current status and validation of conv2d and 3d code. Technical report.

Rahul Dey and Fathi M. Salem. 2017. Gate-variants of gated recurrent unit (GRU) neural networks. In *IEEE 60th International Midwest Symposium on Circuits and Systems, MWSCAS 2017, Boston, MA, USA, August 6-9, 2017*, pages 1597–1600. IEEE.

Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhanyi Liu, Hua Wu, and Jun Zhao. 2017. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In Regina Barzilay and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting*

of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers, pages 221–231. Association for Computational Linguistics.

Zongcheng Ji, Zhengdong Lu, and Hang Li. 2014. An information retrieval approach to short text conversation. *CoRR*, abs/1408.6988.

Rudolf Kadlec, Martin Schmid, and Jan Kleindienst. 2015. Improved deep learning baselines for ubuntu corpus dialogs. *CoRR*, abs/1510.03753.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Feng-Lin Li, Minghui Qiu, Haiqing Chen, Xiongwei Wang, Xing Gao, Jun Huang, Juwei Ren, Zhongzhou Zhao, Weipeng Zhao, Lei Wang, Guwei Jin, and Wei Chu. 2017. *AliMe Assist* : An intelligent assistant for creating an innovative e-commerce experience. In Ee-Peng Lim, Marianne Winslett, Mark Sanderson, Ada Wai-Chee Fu, Jimeng Sun, J. Shane Culpepper, Eric Lo, Joyce C. Ho, Debora Donato, Rakesh Agrawal, Yu Zheng, Carlos Castillo, Aixin Sun, Vincent S. Tseng, and Chenliang Li, editors, *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pages 2495–2498. ACM.

Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau. 2015. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. In *Proceedings of the SIGDIAL 2015 Conference, The 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2-4 September 2015, Prague, Czech Republic*, pages 285–294. The Association for Computer Linguistics.

Christopher D Manning, Christopher D Manning, and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. MIT press.

Alan Ritter, Colin Cherry, and William B Dolan. 2011. Data-driven response generation in social media. In *Proceedings of the conference on empirical methods in natural language processing*, pages 583–593. Association for Computational Linguistics.

James B Rowe, Ivan Toni, Oliver Josephs, Richard SJ Frackowiak, and Richard E Passingham. 2000. The prefrontal cortex: response selection or maintenance within working memory? *Science*, 288(5471):1656–1660.

Ayse Pinar Saygin and Ilyas Cicekli. 2002. Pragmatics in human-computer conversations. *Journal of Pragmatics*, 34(3):227–258.

Ming Tan, Bing Xiang, and Bowen Zhou. 2015. Lstm-based deep learning models for non-factoid answer selection. *CoRR*, abs/1511.04108.

Chongyang Tao, Wei Wu, Can Xu, Wenpeng Hu, Dongyan Zhao, and Rui Yan. 2019. Multi-representation fusion network for multi-turn response selection in retrieval-based chatbots. In J. Shane Culpepper, Alistair Moffat, Paul N. Bennett, and Kristina Lerman, editors, *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*, pages 267–275. ACM.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 5998–6008.

Shuohang Wang and Jing Jiang. 2016. Learning natural language inference with LSTM. In Kevin Knight, Ani Nenkova, and Owen Rambow, editors, *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 1442–1451. The Association for Computational Linguistics.

Shuohang Wang and Jing Jiang. 2017. Machine comprehension using match-lstm and answer pointer. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Yu Wu, Wei Wu, Zhoujun Li, and Ming Zhou. 2016. Topic augmented neural network for short text conversation. *CoRR*, abs/1605.00090.

Yu Wu, Wei Wu, Chen Xing, Ming Zhou, and Zhoujun Li. 2017. Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots. In Regina Barzilay and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 496–505. Association for Computational Linguistics.

Rui Yan, Yiping Song, and Hua Wu. 2016. Learning to respond with deep neural networks for retrieval-based human-computer conversation system. In Raffaele Perego, Fabrizio Sebastiani, Javed A. Aslam, Ian Ruthven, and Justin Zobel, editors, *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*, pages 55–64. ACM.

Zhuosheng Zhang, Jiangtong Li, Pengfei Zhu, Hai Zhao, and Gongshen Liu. 2018. Modeling multi-turn conversation with deep utterance aggregation. In Emily M. Bender, Leon Derczynski, and Pierre Isabelle, editors, *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 3740–3752. Association for Computational Linguistics.

Xiangyang Zhou, Daxiang Dong, Hua Wu, Shiqi Zhao, Dianhai Yu, Hao Tian, Xuan Liu, and Rui Yan. 2016. Multi-view response selection for human-computer conversation. In Jian Su, Xavier Carreras, and Kevin Duh, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 372–381. The Association for Computational Linguistics.

Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The design and implementation of xiaoice, an empathetic social chatbot. *Comput. Linguistics*, 46(1):53–93.