

基于回复生成的对话意图预测

王博宇 王中卿 周国栋

苏州大学计算机科学与技术学院 江苏 苏州 215006

(wby990108@163.com)



摘要 随着人机对话系统的不断发展,让计算机能够准确理解对话者的对话意图,并根据对话的历史信息对回复进行意图预测,对于人机对话系统有着十分重要的意义。已有研究重点关注根据对话文本和已有标签对回复进行意图预测,但是,在很多场景下回复可能并没有生成。因此,文中提出了一种结合回复生成的对话意图预测模型。在生成部分,使用 Seq2Seq 结构,根据对话历史信息生成文本,作为对话中未来回复的文本信息;在分类部分,利用 LSTM 模型,将生成的回复文本与已有的对话信息转变为子句级别的表示,并结合注意力机制突出同一轮次对话句与生成回复的联系。实验结果表明,所提出的模型相比简单基线模型取得了 2.54% 的 *F1-score* 提升,并且联合训练的方式有助于提升模型性能。

关键词: 对话意图;注意力机制;预测模型;文本生成

中图分类号 TP391

Dialogue Act Prediction Based on Response Generation

WANG Bo-yu, WANG Zhong-qing and ZHOU Guo-dong

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

Abstract With the continuous development of the human-machine dialogue system, it is of great significance for the computer to accurately understand the speaker's dialogue act and predict the act of response according to the history information of the dialogue. Previous research work focus on act prediction of responses based on dialogue text and existing labels. But in many scenarios, the reply has not been generated. Therefore, this paper proposes a dialogue act prediction model based on reply generation. In the generation part, the Seq2Seq structure is used to generate text based on the conversation history information as text information for future replies in the conversation; in the classification part, the LSTM model is used to express the generated reply text and the existing conversation information as clause level representations. Combined with the attention mechanism, it highlights the connection between the dialogue sentence of the same round and the generated response. The experimental results show that the proposed model achieves a 2.54% *F1-score* improvement compared to the simple baseline model, and the joint training method contributes to the improvement of model performance.

Keywords Dialogue act, Attention mechanism, Prediction model, Text generation

1 引言

近年来,随着移动互联网的快速发展,对话系统被广泛运用于各类应用和网站。常见的对话包括智能客服、聊天机器人和 Siri 之类的语音助手。准确识别对话意图是构建对话系统过程中一个很重要的子任务。多轮对话中,参与对话的双方往往有很强的目的性,多轮对话中的每一句的意图很明确。面向多轮对话进行意图分析,以及根据对话的历史信息对下一句回复的意图进行正确预测,对于人机对话系统构建、客服自动回复等应用领域具有十分重要的意义。表 1 列出了在一次对话中对话双方的意图以及对未来回复的预测。

表 1 对话意图预测示例

Table 1 Example of dialogue act prediction

子句	意图
A: Our dinner is ready.	Inform
B: What do we have for the dinner now ?	Question
A: Potatoes, tomatoes and cucumber.	Inform
B: (Do you want some juice ?)	Question

与以往的研究不同,本研究的目的是预测未知回复的意图。如表 1 所列,如果在 B 还没有回复时,就能够根据已有信息预测答复的意图为“Question”,则有助于根据对话历史信息生成个性化答复,而不是无意义的通用内容。在人机交互过程中,对用户意图进行正确分析能够有效提高人机交互的

到稿日期:2020-07-21 返修日期:2020-09-01 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61806137,61702149,61836007,61702518);江苏省高等学校自然科学研究面上项目(18KJB520043)

This work was supported by the National Natural Science Foundation of China(61806137,61702149,61836007,61702518) and Jiangsu High School Research Grant(18KJB520043).

通信作者:王中卿(wangzq. antony@gmail.com)

自然度^[1-2]。因此,对于未知回复的意图进行预测很有意义,同时也是一个全新的研究任务。本文将结合回复生成技术对未知回复的意图进行预测。

结合之前的研究工作^[3],本文提出了一种新的对话意图预测方法,具体实现过程为:将对话生成模型与对话意图分类模型相结合,根据历史对话信息生成未来回复,再将未来回复和已有对话一起投入分类部分,得到预测结果。本研究的主要贡献有:

(1)根据多轮对话中的历史信息,使用 Seq2Seq 结构生成未来回复^[4],并将其作为对话中下一句的文本信息;

(2)将未来回复与多轮对话已有文本投入模型的意图分类部分,使用长短期记忆网络(Long Short-Term Memory, LSTM)^[5]并结合注意力机制^[6]进行对话意图分类,生成的文本的分类结果就是对话下一句回复意图的预测结果。

2 相关工作

本文提出的基于回复生成的对话意图预测模型涉及到文本生成和文本意图分类技术。2013年,深度学习开始被运用于自然语言处理任务。Kalchbrenner等^[7]使用卷积神经网络(Convolutional Neural Network, CNN)学习对话的句子级别表示,并结合循环神经网络(Recurrent Neural Network, RNN)^[8]进行对话意图分类,该模型在 SwDA 数据集上取得了好于非神经网络模型的结果。Khanpour等^[9]提出了基于多层 LSTM 的模型,将对话中的每个单句作为输入,在 SwDA 语料上达到了 80.1% 的准确率,他们还讨论了不同层数对分类结果的影响。

Kumar等^[10]使用双向 LSTM 进行多级别的对话文本表示学习,学习了同一段对话中语句内部和不同语句间的文本信息,并使用条件随机场(Conditional Random Field, CRF)^[11]捕获对话意图标签的依赖关系。Raheja等^[12]在多层双向 LSTM 结合条件随机场模型的基础上,采用了自注意力机制和字符级别的词嵌入。Li等^[13]将 dual-attention 运用于多层分级学习的双向 LSTM,也使用了条件随机场进行标签的分类,值得注意的是,该模型同时对意图和对话主题极性进行分类,捕获了两类标签信息之间的联系。

目前,预测对话未来回复意图的研究工作较少,部分研究工作聚焦于如何利用对话历史信息 and 标签信息得出下一句回复的意图标签。Koji等^[14]提出的模型很好地兼顾了这两部分信息,该模型主要由 3 部分构成,分别是 Utterance Encoder, DA Encoder 和 Context Encoder, 3 个 Encoder 用到的网络均为门控循环单元(Gated Recurrent Unit, GRU)^[15]。Utterance Encoder 的输入为词向量,取最后一个 GRU 单元的输出作为整个对话单句的表示。同样地,将得到的单句向量输入到 Context Encoder 中,输出为对话历史信息的篇章级表示。考虑到不同对话者的信息,该模型设置了一个转换标记,与对话单句表示连接后投入 Context Encoder。DA Encoder 的主要作用为,对有标签的信息进行学习,其输入为标签序列的 one-hot 向量,输出设定为 128 维的向量。最终将对话的篇章级表示和标签向量连接,投入全连接层,所得的分类结果即视为对下一句回复的意图预测。

综上所述,先前的研究工作主要集中在文本意图分类模型的构建上。而本文提出的模型更关注对未知回复的意图预测,使用文本生成技术填充未知回复的空白,并在分类部分借鉴先前研究的思想。

3 基于回复生成的对话意图预测模型

3.1 模型整体框架

图 1 给出了本文模型的整体框架。输入为对话的前 $n-1$ 句历史文本信息 $D' = \langle u_1, u_2, \dots, u_{n-1} \rangle$, 输出为对话回复 u_n 的意图识别结果。首先计算 D' 的句子级别表示,将该表示的平均值投入到解码器部分,经过多次生成后得到回复 u_n 。计算 u_n 的句子级表示 h_n ,与组成完整的对话篇章 D ,然后在分类部分结合注意力机制再次进行表示学习,将最终表示投入全连接层得到分类结果。 u_n 的意图分类结果就是对话回复的预测结果。

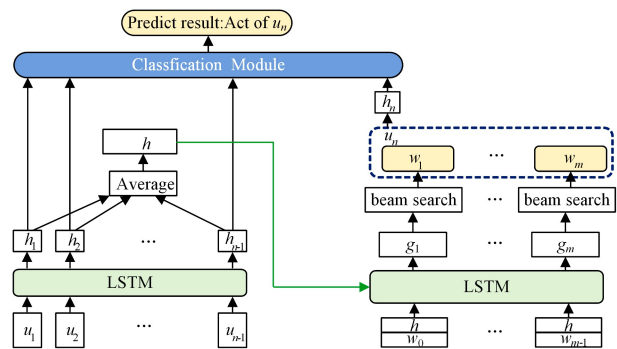


图 1 基于回复生成的对话意图预测模型

Fig. 1 Dialogue act prediction model based on response generation

3.2 对话子句的表示

我们利用 LSTM 来学习每一个对话子句的表示。每一个 LSTM 单元最主要的部分为输入门、遗忘门和输出门。相比 RNN,为了维持长期记忆, LSTM 还有一个长期状态 c 。LSTM 单元的计算公式如下:

$$i_t = \sigma(W_i x_t + U_i h_{t-1}) \quad (1)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1}) \quad (2)$$

$$g_t = \tanh(W_g x_t + U_g h_{t-1}) \quad (3)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1}) \quad (4)$$

$$c_t = \sigma(W_c x_t + U_c h_{t-1}) \quad (5)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g_t \quad (6)$$

$$h_t = o_t \circ \tanh(c_t) \quad (7)$$

其中, h_t 表示 t 时刻的隐藏状态, c_t 表示 t 时刻的单元长期状态, i_t, f_t, o_t 分别表示 t 时刻的输入门、遗忘门和输出门的输出。

经过 LSTM 神经网络,将经词嵌入层输出的原对话文本的词向量序列转变为句子级表示,且每个对话子句独立表示。

3.3 回复生成

在联合模型的回复生成部分主要使用 Seq2Seq 模型, Seq2Seq 模型由编码器和解码器组成,编码器用来学习对话历史信息的表示,将学习到的表示传入编码器进行回复生成。

编码器将已知的前 $n-1$ 句文本信息输入 LSTM 神经网络,得到每一个句子的句子级表示,取句子级表示的平均值,经过池化操作后得到表示 h 作为对话的历史文本信息,传入解码器。

解码器部分的核心思想为:根据历史文本信息和当前词 w_t , 经过 LSTM 神经网络, 生成下一个词语 w_{t+1} , 然后将 w_{t+1} 作为新的输入, 不断循环, 生成新的文本。从神经网络输出到单词 w 的转化过程中, 使用 beam search 解码^[16]。循环起始时, 输入起始标记符(start)作为第一个当前词, 当生成结束标记符(end), 或者生成序列的长度达到规定的最大值时, 循环结束。

最终, 序列 $u_n = \langle w_1, w_2, \dots, w_m \rangle$ 即为生成的对话回复, 与对话历史信息一起传入分类部分。

3.4 对话意图分类

经过文本生成部分, 已有的历史文本信息 $D' = \langle u_1, u_2, \dots, u_{n-1} \rangle$ 与生成部分的结果 u_n 组成完整的对话信息 $D = \langle u_1, u_2, \dots, u_n \rangle$, 其中 $n = 2 * k, k$ 为正整数。

将词向量级别的对话文档 $D = \langle u_1, u_2, \dots, u_n \rangle$ 输入 LSTM 层, 单句中的每个词向量即为序列中的每个元素, 在 LSTM 单元中按照 LSTM 原理进行前向计算。LSTM 网络最后一个时间步的输出包含了整个序列的信息, 可以看作整个单句的表示, 经过第一个 LSTM 层, 对话文档为 $D = \langle h_1, h_2, \dots, h_n \rangle$, 其中 h_i 是 u_i 的句子级别表示。实际上, 该层是生成部分与分类部分的共享层, 如 3.2 节所述, $\langle h_1, h_2, \dots, h_{n-1} \rangle$ 在进行回复生成时就被计算出, 我们只需计算 u_n 的句子级表示 h_n 即可。

一般而言, 未来回复的前一句与回复句有更密切的联系, 因此我们考虑在分类部分中加入句子级别的注意力机制。在注意力机制中, 最主要的是 Query 向量(Q)、Key 向量(K)和 Value 向量(V), 这 3 个向量维度相同, 分别由文本信息的表示乘以 3 个不同的权值矩阵 W_Q, W_K 和 W_V 获得。注意力机制的计算过程为: 使用 softmax 函数对 Q 和 K 向量的内积做平滑, 平滑后的结果与 V 向量相乘即为计算结果。计算公式如下:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (8)$$

多头注意力机制是 Google 提出的新概念^[17], 是对注意力机制的完善, 其基本原理为把 Q, K, V 向量通过参数矩阵映射, 然后按式(8)进行注意力机制的运算, 重复该过程 h 次, 对结果进行拼接, 具体公式如下:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h) \quad (9)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (10)$$

本文模型中, Q 对应该生成回复句的表示, K 和 V 对应回复前一句的表示。

最终, 将学习到的表示投入全连接层, 使用 softmax 函数作为激活函数进行多分类, 得到对话中每个单句的意图类别。

4 数据集及实验设置

4.1 实验数据集及预处理

本文选用 Li 等整理的 DailyDialog 数据集^[18]。

除了对话文本外, Li 等为 DailyDialog 数据集标注了丰富的信息, 包括对话主题、对话情感和对话意图。本文的主要研究对象为对话意图。DailyDialog 数据集中共有 4 类对话意图标签, 分别为 Inform, Question, Directive 和 Commissive。

本文实验设定为对话轮次固定的情况。根据数据集的基

本统计信息, 平均每段对话的单句数为 7.9, 且轮数大于或等于 5 轮的对话数只占总对话数的 32%, 使用较多轮次的对话则会使能利用的语料变少, 因此固定对话轮次为 4 轮。经过筛选, 共有 6372 段对话轮次大于或等于 4 轮, 对于轮次大于 4 轮的对话, 只取前 4 轮。筛选后的语料中, 4 类意图标签的统计情况如图 2 所示。

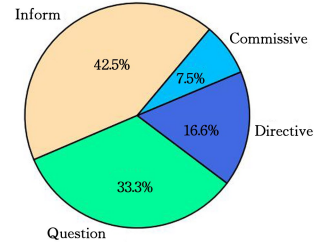
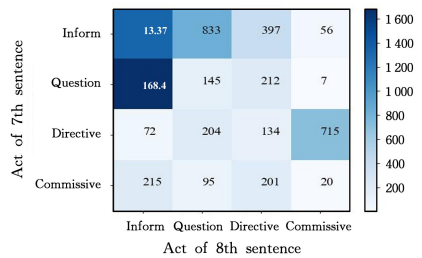


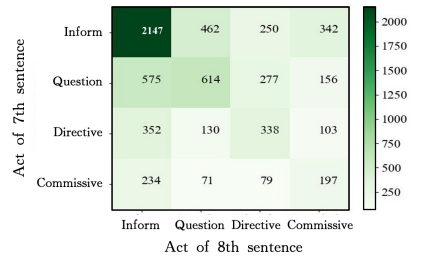
图 2 4 类意图标签占比

Fig. 2 Proportion of four types of act labels

此外, 为了探究同一个对话轮次中对话意图的相互影响和同一个对话者语句的意图联系, 我们还分别统计了第 6 句、第 7 句与第 8 句对话单句的意图交互情况, 并绘制了交互矩阵, 如图 3 所示。



(a) 第 7 句与第 8 句的意图交互矩阵



(b) 第 6 句与第 8 句的意图交互矩阵

图 3 意图交互矩阵

Fig. 3 Act interaction matrix

从意图交互矩阵可以看出, 第 7 句和第 8 句之间意图的联系更强, 即在同一个对话轮次中, 有很大可能从前一句对话的意图推断出下一句的意图。因此, 在对未来回复进行预测时, 使用同一轮次的对话语句对未来回复进行替代、模拟有一定的合理性。

4.2 基线对比方法

对于回复意图预测问题, 我们将所提模型与多个基线模型进行比较。这些基线模型没有生成第 8 句文本, 而是简单地考虑如何用历史文本信息对第 n 句文本进行替代, 没有充分利用已知信息, 或者缺少对话者、对话轮次多个角度的考虑。本文用于比较的几种基线模型如下:

(1) LSTM^{N-1} 模型。使用已知对话文本的最后一句, 即使用同一对话轮次的语句进行替代。

(2)LSTM^{N-2}。使用已知对话文本的倒数第二句,即使用同一对话者且最接近预测语句处的文本信息进行替代。

(3)LSTM^D模型。使用这个已知对话文本经过两层 LSTM 后的表示作为预测语句的替代。

(4)Koji 等^[14]提出的模型。

4.3 实验参数设置

本文实验的主要参数设置:词嵌入层输出的词向量维度设置为 128 维;每个对话单句序列长度为 35,每段对话包含 7 个对话单句,故输入的每段对话序列长度为 245;设置每批次大小为 64,优化器使用 Adam,学习率设置为 0.001,为缓解过拟合问题和梯度消失问题,设置 Dropout 比率为 0.5;对于 LSTM 隐含单元数,生成部分设置为 128,分类部分设置为 64;全连接层激活函数设置为 *softmax*,用于进行多分类。

5 实验结果分析及讨论

5.1 与基线模型的实验结果对比

本文模型与基线模型的实验结果如表 2 所列。

表 2 各模型的实验结果对比

Table 2 Comparison with experimental results of each model

模型	精确率	召回率	F1-score
LSTM ^{N-1}	0.4513	0.4312	0.4410
LSTM ^{N-2}	0.4076	0.4210	0.4142
LSTM ^D	0.4733	0.4569	0.4650
Koji ^[14]	0.4802	0.4551	0.4673
Ours	0.4858	0.4952	0.4904

根据表 2 的结果,结合意图交互矩阵可知,同一轮次的对话句和出于同一对话者的上一句,与被预测句之间的意图存在一定联系,因此使用简单的基线模型如 LSTM^{N-1} 和 LSTM^{N-2} 的预测结果作为对比。同时,第 7 句与第 8 句的意图联系强于第 6 句与第 8 句的意图联系,相应地,LSTM^{N-1} 模型的 F1 得分比 LSTM^{N-2} 模型提高了 2.68%,说明了使用这两种基线模型进行意图预测的合理性。但是,由于这两个模型只用到了原对话中少量的相关信息,F1 得分相比本文模型有 4.94% 和 7.62% 的差距。基线模型 LSTM^D 和 Koji^[14] 的方法虽然用到了全部的对话信息,但相比本文模型,没有考虑不同对话句之间的联系,因此本文模型相比 LSTM^D 和 Koji^[14] 取得了 2.54% 和 2.31% 的性能提升。本文模型使用 Seq2Seq 结构进行文本生成,在生成回复的基础上进行分类,相比基线模型结果得到了提升,说明了文本生成技术对于意图预测的有效性。

5.2 实验影响因素的讨论

对于本文实验因素的影响,主要考虑以下两点:1)生成部分对于预测效果的作用;2)训练方式的影响。对于第 1)点,为了讨论生成部分的有效性,只保留分类部分进行实验,作为完整模型的对比,这需要对第 n 句进行合理的模拟。我们将同一对话者的语句经过第一层 LSTM 后得到的语句级别表示进行加和,作为对第 n 句文本的模拟,再连接到分类模型后半部分进行训练。

对于第 2)点,本文模型为联合模型,在每一个训练回合中,生成部分和分类部分都得到了训练,并且学习句子级别表示的 LSTM 层为两部分共有。为了验证联合训练的有效性,

将生成模型和分类模型分开,按先生成再分类的次序进行训练,作为联合模型的对比实验。

5.2.1 实验影响因素的讨论

我们选取了包含 4 个轮次的对话进行实验,对第 8 个单句的意图做出预测,因此,在不使用生成回复的情况下,选择使用同一对话者的语句,即用第 2,4,6 句对话单句的表示来对第 8 句进行模拟,投入分类模块。实验结果如表 3 所列。

表 3 生成部分对实验结果的影响

Table 3 Influence of generated part on experimental results

模型	精确率	召回率	F1-score
生成回复	0.4858	0.4952	0.4904
模拟回复	0.4552	0.4467	0.4411
生成十模拟	0.4851	0.4797	0.4823

从表 3 中可以看出,只使用相关对话者的语句进行回复模拟的效果最差,生成回复比模拟回复的 F1-score 高出 4.93%,这证实了联合模型生成部分起到了作用。使用 2,4,6 句模拟是一种粗糙的处理方式,因此将两种回复进行融合后,相比只使用生成的回复,效果又有所下降。

5.2.2 训练方式的影响

本文模型将文本生成与文本分类相结合,生成部分的输出作为分类部分输入的一部分,且两部分有共享的 LSTM 层。我们将两部分拆分为两个模型,分离训练采取“先生成再分类”的顺序,对比实验结果如表 4 所列。

表 4 不同训练方式的影响

Table 4 Influence of different training methods

训练方式	精确率	召回率	F1-score
分离训练	0.4709	0.4711	0.4710
联合训练	0.4858	0.4952	0.4904

联合训练方式相比分离训练方式,在 F1-score 上有 1.94% 的提升,这主要得益于模型计算的连贯性得到保持。并且,学习对话句子级表示的 LSTM 层是生成部分和分类部分的共享层,避免了分离训练下参数的冗余,且相对于两个部分,对话信息的句子级表示是一致的,从而使模型整体对句子级表示的学习更加准确。

5.3 生成回复的评估

双语评估替换(Bilingual Evaluation Understudy, BLEU)是比较生成文本与参考文本相似程度的评价分数^[19]。BLEU 的得分计算公式如下:

$$BLEU = BP \times \exp\left(\sum_{n=1}^N w_n \log p_n\right) \quad (11)$$

$$p_n = \frac{C_m}{C} \quad (12)$$

$$BP = \begin{cases} 1, & c > r \\ e^{(1-\frac{r}{c})}, & c \leq r \end{cases} \quad (13)$$

其中, p_n 是 n -gram 的精度得分, w_n 是正向权重, C_m 是生成文本与参考文本中匹配的 n -gram 次数, C 是候选文本中匹配的 n -gram 次数。针对生成文本比参考文本短的情况,需要引入惩罚机制,即 BP。式(13)中, c 为生成文本的长度, r 为参考文本的长度。可以取 n -gram 中的 n 值为 1~4,计算相应的 BLEU 得分。表 5 列出了生成回复句、第 $N-1$ 句、 $N-2$ 句与参考文本的 BLEU 得分。

表5 不同文本的 BLEU 得分

Table 5 BLEU score for different texts

评估文本	BLEU-1	BLEU-2	BLEU-3	BLEU-4
生成文本	0.3136	0.2447	0.2195	0.1997
N-1	0.1053	0.0187	0.0056	0.0017
N-2	0.1368	0.0265	0.0076	0.0024

由表5可知,在 n 取 $1\sim 4$ 的情况下,生成回复句的 BLEU 得分都远高于 $N-1$ 句和 $N-2$ 句。这说明生成回复与真实的第8句回复更为接近,从另一个角度说明了实验结果的正确性。

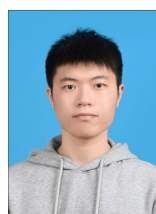
结束语 本文提出了一种包含文本生成和文本分类两部分的模型,用来解决多轮对话中下一句回复的意图预测问题。实验结果表明,使用 Seq2Seq 结构生成的对话回复语句较好地维持了对已有对话信息的记忆,并能突出与其中答复的重点。将文本生成技术与文本分类技术相结合构建意图预测模型,其预测结果的准确率和 $F1$ -score 超过了使用历史信息模拟对话回复的基线模型,相比简单基线模型取得了 2.54% 的 $F1$ -score 提升,说明了使用文本生成技术生成对话回复在本任务中的重要性。在下一步的工作中将考虑对分类部分进行结构改进,如融合条件随机场以实现意图标签更准确的判别。

参 考 文 献

- [1] TUR G, CELIKYILMAZ A, HAKKANI-TÜR D. Latent semantic modeling for slot filling in conversational understanding [C] // 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013: 8307-8311.
- [2] EYBEN F, WÖLLMER M, GRAVES A, et al. On-line emotion recognition in a 3-D activation-valence-time continuum using acoustic and linguistic cues [J]. Journal on Multimodal User Interfaces, 2010, 3(1/2): 7-19.
- [3] LIU J, LI Y L, LIN M. Review of Intent Detection Methods in Human-Machine Dialogue System [J]. Computer Engineering and Applications, 2019, 55(12): 1-7, 43.
- [4] LI W B, ZHANG L, SHU X. Application and Research on Generative Automatic Question Answering System Based on Seq2Seq [J]. Modern Computer, 2017(36): 59-62.
- [5] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [6] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [J/OL]. <https://arxiv.org/abs/1409.0473>.
- [7] KALCHBRENNER N, BLUNSOM P. Recurrent Convolutional Neural Networks for Discourse Compositionality [C] // Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality. 2013: 119-126.
- [8] MIKOLOV T, KARAFIT M, BURGET L, et al. Recurrent neural network based language model [C] // Proceedings of 11th Annual Conference of the International Speech Communication Association. 2010: 1045-1048.
- [9] KHANPOUR H, GUNTAKANDLA N, NIELSEN R. Dialogue act classification in domain-independent conversations using a deep recurrent neural network [C] // Proceedings of COLING 2016, the 26th International Conference on Computational Lin-

guistics. 2016: 2012-2021.

- [10] KUMAR H, AGARWAL A, DASGUPTA R, et al. Dialogue act sequence labeling using hierarchical encoder with crf [C] // Thirty-Second AAAI Conference on Artificial Intelligence. 2018: 3440-3447.
- [11] LAFFERTY J. Conditional random fields: Probabilistic models for segmenting and labeling sequence data [C] // Proceedings of the 18th International Conference on Machine Learning. 2001.
- [12] RAHEJA V, TETREAUULT J. Dialogue Act Classification with Context-Aware Self-Attention [C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019: 3727-3733.
- [13] LI R, LIN C, COLLINSON M, et al. A Dual-Attention Hierarchical Recurrent Neural Network for Dialogue Act Classification [C] // Proceedings of the 23rd Conference on Computational Natural Language Learning. ACL, 2019: 383-392.
- [14] TANAKA K, TAKAYAMA J, ARASE Y. Dialogue-Act Prediction of Future Responses Based on Conversation History [C] // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. 2019: 197-202.
- [15] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation [C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1724-1734.
- [16] FREITAG M, AL-ONAIZAN Y. Beam Search Strategies for Neural Machine Translation [J]. Association for Computational Linguistics, 2017, 2017: 56.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // Advances in Neural Information Processing Systems. 2017: 5998-6008.
- [18] LI Y, SU H, SHEN X, et al. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset [C] // Proceedings of the Eighth International Joint Conference on Natural Language Processing. 2017: 986-995.
- [19] PAPANENI K, ROUKOS S, WARD T, et al. BLEU: a method for automatic evaluation of machine translation [C] // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 2002: 311-318.



WANG Bo-yu, born in 1999, postgraduate. His main research interests include natural language processing and machine learning.



WANG Zhong-qing, born in 1987, Ph.D., lecturer. His main research interests include natural language processing, sentiment analysis and event extraction.