

# SMGN:用于对话状态跟踪的状态记忆图网络

张志昌,于沛霖,庞雅丽,朱 林,曾扬扬

(西北师范大学计算机科学与工程学院,甘肃兰州 730070)

**摘 要:** 对话状态跟踪是任务型对话系统的重要模块. 已有研究使用注意力机制模拟图结构来引入历史信息,但这种方法无法显式利用对话状态的结构性. 此外,如何生成复杂格式的对话状态也为研究带来了挑战. 针对以上问题,本文提出一种状态记忆图网络 SMGN(State Memory Graph Network). 该网络通过状态记忆图保存历史对话信息,并使用图结构与当前对话进行特征交互. 本文还设计了一种基于状态记忆图的复杂对话状态生成方法. 实验结果表明,本文提出的方法在 CrossWOZ 数据集上联合正确率提高 1.39%,在 MultiWOZ 数据集上提高 1.86%.

**关键词:** 任务型对话系统; 对话状态跟踪; 注意力机制; 自然语言处理; 深度学习

**中图分类号:** TP391.1

**文献标识码:** A

**文章编号:** 0372-2112(2022)08-1851-08

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20201463

## SMGN: A State Memory Graph Network for Dialogue State Tracking

ZHANG Zhi-chang, YU Pei-lin, PANG Ya-li, ZHU Lin, ZENG Yang-yang

(College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu 730070, China)

**Abstract:** Dialogue state tracking is an important module of task-oriented dialogue system. Previous studies exploited the historical dialogue information by attention-based graph structure simulation, but these methods cannot explicitly take advantage of the structure of the dialogue state. In addition, how to generate complex format dialogue states also brings challenges to research. In this paper, we propose a state memory graph network(SMGN). The network saves historical information through the state memory graph, and uses the graph to interact with the current dialogue. We also implement a complex dialogue state generation method based on state memory graph. Experimental results show that the proposed method improves the joint accuracy by 1.39% on the CrossWOZ dataset and 1.86% on the MultiWOZ dataset.

**Key words:** task-oriented dialogue system; dialogue state tracking; attention mechanism; natural language processing; deep learning

### 1 引言

随着各种智能设备的快速发展,人机对话近年来引起了学术界和产业界的广泛关注. 任务型对话系统相关技术已经在许多产品中得到运用,例如微软公司的“小娜”(Cortana),苹果公司的智能语音助手“Siri”等. 在任务型对话系统中,对话状态跟踪(Dialogue State Tracking, DST)是一个重要模块,该模块的主要目标是追踪用户意图并构建对话状态集合,集合由多个结构化的“领域-槽位-值”三元组构成. 表 1 展示了一个对话状态跟踪任务示例,对于用户输入的话语,对话状态跟踪模型需要识别出句子的对话状态(例如“酒店-房间类型-标准间”). 正确识别对话状态对任务型对话系统非常重要.

表 1 对话状态跟踪任务示例

对话文本	用户:帮我定一间香格里拉酒店的标准间. 系统:请问您哪一天入住呢? 用户:明天.再帮我预定一辆从首都国际机场到酒店的出租车. 系统:请问您几点可以从首都国际机场出发? 用户:.....
对话状态	{酒店-名称-香格里拉酒店,酒店-房间类型-标准间,酒店-入住日期-明天,出租车-出发地-首都国际机场,出租车-目的地-香格里拉酒店}

传统的静态对话状态跟踪模型仅依赖当前轮用户输入的话语<sup>[1-4]</sup>. 这类方法忽略了对话历史和当前对话的关系. 在多轮对话中,经常需要利用历史对话中的信

息来帮助理解用户当前的意图,所以历史敏感的对话状态跟踪模型成了主流研究方向<sup>[5-8]</sup>. 其中多数研究都选择通过对历史文本进行编码来引入上文信息,而文献[9~11]等研究表明,对历史时刻生成的对话状态进行编码会更有效. 与对话文本不同的是,对话状态本身具有结构性,文献[9]就利用这种结构建立模式图来帮助模型进行特征交互. 但是,过去的研究没有显式利用对话状态的结构,只是通过注意力机制中特征的连接来模拟模式图,这种方法忽略了图结构中节点自身包含的信息.

另外,随着对话系统的复杂度逐渐提升,如何生成复杂格式的对话状态也成了研究的热点. 传统方法是使用序列标注生成对话状态<sup>[12]</sup>,该方法在其他自然语言处理研究中得到了广泛的运用<sup>[13,14]</sup>. 但是这种方法无法应对复杂场景下的要求. 例如,某些槽位对应的值仅出现在历史对话中,或者槽位并不对应某个具体的值而是对应“是”或“否”(例如在预定酒店场景中,可能会出现“酒店-停车位-否”这类对话状态). 已有研究提出多种方法使模型可以生成复杂格式的对话状态. 例如文献[7,10]等通过预测4种离散的运算状态来决定对槽位执行何种操作,文献[15]提出通过编码器-解码器模型(Seq2Seq)将状态预测转化为生成式任务. 这些方法会为模型带来额外的计算负担,并且需要针对不同的对话状态调整生成规则,可扩展性差.

针对上述问题,本文提出了一种状态记忆图网络(State Memory Graph Network, SMGN). 在对话开始时,模型枚举所有可能出现的“领域-槽位”关系对,通过编码器进行编码生成槽结点,再使用槽结点与对话文本生成的值结点进行特征交互,通过两类结点连接形成状态记忆图. 图1展示了表1样例对应的状态记忆图.

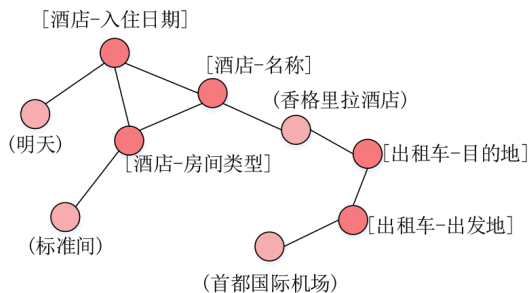


图1 状态记忆图示例

在状态记忆图中,槽结点与值结点相连的边对应一条对话状态,整个状态记忆图对应对话状态集合. 模型会使用状态记忆图与当前对话进行特征交互并更新对话状态集合,之后通过图注意力网络(Graph Attention neTwork, GAT)<sup>[16]</sup>对状态记忆图进行更新. 此外,本文通过值结点标记器和图连接层实现了一个新的对话状态生成方法,首先通过值结点标记器为当前对话

生成值结点,之后通过判断槽结点与值结点是否连接来生成对话状态. 由于状态记忆图会保留历史对话产生的值结点,语义槽可以和历史文本中的值构建对话状态. 通过预生成公共值结点,模型也可以生成类似“是”或“否”的特殊类型值,这使得本文的模型可以生成复杂格式的对话状态,并且具有很强的扩展性.

为了验证模型的有效性,本文在3个公开的任务型对话数据集(中文数据集 CrossWOZ<sup>[17]</sup>,英文数据集 MultiWOZ 2.0<sup>[18]</sup>及 MultiWOZ 2.1<sup>[19]</sup>)上进行了实验. 实验结果表明,本文提出的模型在3个数据集上的联合正确率分别为37.47%,53.03%和54.88%. 实验分析表明了本文的模型可以有效提升对话状态跟踪任务的性能. 相比其他模型,本文模型对历史对话的语义理解能力更强. 另外该模型比其他同类模型的运算效率更高.

本文的主要贡献如下:

(1)提出了一种显式利用图结构的状态记忆图网络,通过状态记忆图与当前对话进行特征交互,可以更有效地捕获对话状态的语义信息;

(2)针对复杂结构对话状态生成困难的问题,本文实现了一个基于状态记忆图的复杂对话状态生成方法,通过这种方法模型可以应对各种复杂格式的对话状态生成要求;

(3)在3个公开的多轮对话数据集上进行了对比实验,实验结果表明本文方法可以有效提高对话状态跟踪任务的联合正确率,与同类模型相比运算效率更高.

## 2 状态记忆图网络

本文提出的状态记忆图网络模型结构如图2所示. 状态记忆图网络由编码层、值结点标记器、图连接层以及图更新层4部分组成.

### 2.1 任务定义

针对任意一个多轮对话, $t$ 时刻的对话历史文本 $X$ 可以表示为

$$X = \{(R^1, U^1), (R^2, U^2), \dots, (R^{t-1}, U^{t-1})\} \quad (1)$$

其中, $R^i$ 表示系统在第 $i$ 轮生成的回复, $U^i$ 表示用户在第 $i$ 轮(或者 $i$ 时刻)输入的话语. $t-1$ 时刻的对话状态集合可以表示为 $S^{t-1} = \{(d_j, s_j, v_j), j \in J\}$ ,其中, $d_j$ 表示领域, $s_j$ 表示槽位, $v_j$ 表示槽位对应的值, $J$ 表示对话状态总数.

虽然不同对话系统间的对话状态格式有所区别,但都能通过规则转化为“领域-槽位-值”结构的三元组. 例如某些对话系统需要预测意图,此时只需要将“意图”槽位对应值设为“空”即可(形如“领域-意图-null”).

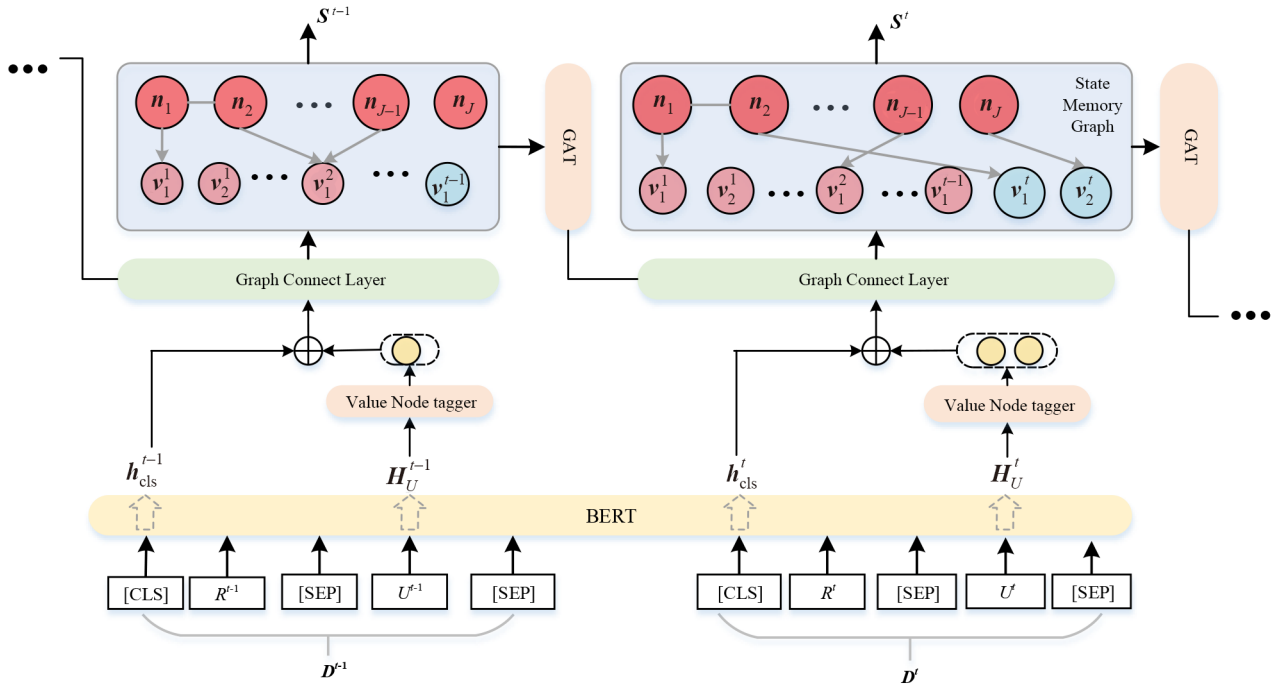


图2 状态记忆图网络模型结构图

对话状态跟踪任务是给定对话历史  $X$  和历史状态集合  $S^{t-1}$ , 根据用户  $t$  时刻的输入  $U^t$  预测对话状态集合  $S^t$ .

### 2.2 编码层

在编码层中, 模型分别进行槽节点编码以及输入编码, 两部分编码任务共享同一个编码器.

**槽节点编码:** 在对话开始时, 模型通过枚举列出所有可能出现的“领域-槽位”关系对, 并且为每个关系对编码生成一个槽节点. 本文使用 BERT (Bidirectional Encoder Representation from Transformers) 作为编码器. BERT 是一种大规模预训练语言模型<sup>[18]</sup>, 具有强大的通用上下文语意表示能力, 已在其他自然语言处理任务中得到充分运用<sup>[19]</sup>. 槽节点的具体编码计算方式如下:

$$n_i = \text{BERT}_{\text{fin}}([\text{CLS}] \oplus d_j \oplus [\text{SEP}] \oplus s_{jk}) \quad (2)$$

其中,  $[\text{CLS}]$  和  $[\text{SEP}]$  为 BERT 编码器预先设定的标记符,  $\oplus$  表示拼接操作,  $d_j$  表示第  $j$  个领域,  $s_{jk}$  表示第  $j$  个领域下对应的第  $k$  个语义槽.  $\text{BERT}_{\text{fin}}$  表示在训练期间将对 BERT 进行微调 (fine-tune). 编码后模型截取标识符  $[\text{CLS}]$  对应的上下文向量作为节点的整体表示  $n_i$ .

通过槽节点编码, 模型获得了一系列槽节点构成的集合  $N = \{n_1, n_2, \dots, n_j\}$ . 需要说明的是, 槽节点编码过程并不受对话内容影响, 槽节点的个数只与对话系统所支持的领域与槽位有关.

**输入编码:** 在  $t$  时刻, 模型将系统回复  $R^t$  及用户输入  $U^t$  进行拼接并通过 BERT 进行编码, 具体计算方式如下:

$$H^t = \text{BERT}_{\text{fin}}([\text{CLS}] \oplus R^t \oplus [\text{SEP}] \oplus U^t) \quad (3)$$

输入编码与槽节点编码共享同一个 BERT 编码器, 也会在训练期间对 BERT 进行微调. 模型截取标识符  $[\text{CLS}]$  对应的上下文向量  $h_{\text{cls}}^t$  作为当前轮对话的整体表示, 截取用户输入  $U^t$  对应的上下文向量  $H_U^t = \{h_1^t, h_2^t, \dots, h_n^t, n = |U^t|\}$  作为用户输入的代表, 其长度与  $U^t$  相等.

### 2.3 值节点标记器

在获得用户输入  $U^t$  对应的上下文向量  $H_U^t$  后, 模型通过值节点标记器提取值节点. 值节点标记器会进行一个简单的序列标注任务, 为每个上下文向量标注一个边界标签来划分出实体边界. 本文采用 BIO ((B-begin, I-inside, O-outside)) 标注方法, 标注方法的具体细节可以参考文献<sup>[12]</sup>. 在标注任务中, 模型只判断实体的边界而不进行实体分类. 在标注结束后, 模型对同一实体内的上下文向量进行平均池化, 再与当前时刻的句向量  $h_{\text{cls}}^t$  拼接获得值节点. 具体计算方式为

$$y_i^t = \text{ReLU}(W h_i^t + b) \quad (4)$$

$$o_i^t = \arg \max \left( \frac{\exp(y_i^t)}{\sum_{j=1}^3 \exp(y_{ij}^t)} \right) \quad (5)$$

$$v_i^t = \left\{ \left( \frac{1}{L} \sum_{0 \leq l \leq L} h_{i+l}^t \right) \oplus h_{\text{cls}}^t \right\} \quad (6)$$

其中,  $W, b$  是可训练参数,  $y_i^t \in R^3$  是对应 3 个实体标记

$B, I, O$  的三维实数向量,  $\mathbf{o}_i^t \in R^3$  是第  $i$  个字对应实体标记  $B, I, O$  的概率向量,  $L$  为第  $i$  个实体的字段长度. 经过值结点标记器后, 模型将新生成的值结点加入集合  $V$  中, 得到  $V = \{v_1^1, v_2^1, \dots, v_n^1, v_1^2, \dots, v_n^2, \dots, v_1^t, \dots, v_n^t\}$ . 每个时刻生成的值结点个数受到当前用户输入的影响.

## 2.4 图连接层

经过值结点标记器后, 模型将当前时刻的值结点添加到状态记忆图中, 并对每一对“槽结点-值结点”进行分类, 判断两个结点之间是否有边, 从而构建连接关系. 图连接层的具体计算方式为

$$\mathbf{e}_{ij} = (\mathbf{n}_i \oplus \mathbf{elem}_j | \mathbf{n}_i \in N, \mathbf{elem}_j \in V) \quad (7)$$

$$\mathbf{o}^{e_{ij}} = \arg \max \left( \frac{\exp(\mathbf{e}_{ij})}{\sum_{k=1}^2 \exp(\mathbf{e}_{ij_k})} \right) \quad (8)$$

$$S^t = \{\mathbf{o}^{e_{i1}}, \mathbf{o}^{e_{i2}}, \dots, \mathbf{o}^{e_{ij}}\} \quad (9)$$

其中,  $\mathbf{o}^{e_{ij}} \in R^2$  是槽结点  $\mathbf{n}_i$  与某个值结点  $\mathbf{elem}_j$  的二分类结果的概率向量,  $S^t$  是当前时刻的对话状态集合. 通过图连接层, 模型为每个槽结点修正了当前时刻对应的值结点, 并且更新了对话状态集合.

## 2.5 图更新层

在图更新层中, 模型将根据状态记忆图的连接关系对结点进行更新. 除了保留图连接层的信息外, 模型还为相似的槽结点添加连接. 具体方法是, 如果两个槽结点具有相同的领域或者槽位, 则为这一对结点添加一条边. 完成连接后, 模型得到了  $t$  时刻状态记忆图对应的邻接矩阵  $A^t$ , 之后将使用该邻接矩阵对状态记忆图进行更新.

模型采用图注意力网络 GAT 实现图更新. 对于一个拥有  $M$  个结点的图, 图注意力网络使用由槽结点和值结点拼接获得的结点状态图  $\tilde{H} = \{\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_m, m \in N \oplus V\}$  作为输入, 旨在通过邻接矩阵  $A^t$  在一阶临结点间进行特征交互, 得到结点间的抽象表示  $\tilde{H}' = \{\tilde{h}'_1, \tilde{h}'_2, \dots, \tilde{h}'_m\}$  作为输出. 图注意力网络具体可表示为

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^T [W_h \tilde{h}_i || W_h \tilde{h}_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(a^T [W_h \tilde{h}_i || W_h \tilde{h}_k]))} \quad (10)$$

$$\tilde{h}'_i = \sigma \left( \frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{N}_i} \alpha_{ij} W_h \tilde{h}_j \right) \quad (11)$$

其中,  $\mathcal{N}_i$  是结点  $i$  的所有一阶邻结点,  $W_h$  和  $a$  是可训练权重矩阵,  $\alpha$  是每个  $\tilde{h}_j$  对  $\tilde{h}_i$  的归一化注意力权重,  $\sigma$  是非线性激活函数,  $(\cdot)^T$  代表转置操作,  $||$  表示拼接操作,  $K$  表示多头注意力的个数.

经过图注意力网络后, 状态记忆图中的结点信息将被更新, 在  $t+1$  时刻被传送至图连接层, 为新一轮对话带来历史信息.

## 2.6 损失计算

整个网络的损失由值结点标记损失和状态分类损失两部分组成. 两部分损失都通过交叉熵损失函数 (cross entropy loss)<sup>[20]</sup> 计算得到, 具体计算方式如下:

$$L_v \triangleq - \sum_{i=1}^t \sum_{j=1}^{|U_i|} \hat{y}_{ij}^t \log(y_{ij}^t) \quad (12)$$

$$L_N \triangleq - \sum_{i,j=1}^{|m|} \hat{e}_{ij} \log(e_{ij}) \quad (13)$$

$$L = \alpha_1 L_v + \alpha_2 L_N \quad (14)$$

其中,  $L_v$  是值结点标记损失,  $L_N$  是状态分类损失,  $\alpha_1$  和  $\alpha_2$  是通过开发集结果微调的超参数.

## 3 实验及分析

### 3.1 数据集

本文在 3 个公开数据集 CrossWOZ, MultiWOZ 2.0 和 MultiWOZ 2.1 上进行实验. CrossWOZ<sup>[17]</sup> 是一个大规模跨领域中文任务导向对话数据集, 包含 6 000 余轮对话和 102 万句话语, 涉及 5 种领域、72 种槽位、7 000 余种槽位值, 每个对话平均涉及 3.2 种领域. MultiWOZ 2.0<sup>[18]</sup> 是英文多领域对话数据集, 包含 7 种领域、25 种槽位、4 500 余种槽位值. 为了对比的公平性, 本文效仿文献[5]丢弃两个数量过少且仅出现在训练集的领域数据. MultiWOZ 2.1<sup>[19]</sup> 是对 2.0 版本数据集的修正版本. 表 2 为 3 个数据集的具体信息统计.

表 2 数据集信息统计

统计信息	CrossWOZ	MultiWOZ 2.0	MultiWOZ 2.1
语言	中文	英文	英文
领域	5	7	5
槽位	72	25	25
槽值	7 871	4 510	4 510
对话数	5 012	8 438	8 133
轮次总数	84 692	115 424	113 556

### 3.2 参数设置

针对 CrossWOZ 数据集, 本文采用文献[23]发布的 BERT-WWM 版本作为预训练模型; 针对 MultiWOZ 2.0 及 MultiWOZ 2.1 数据集, 模型使用 Google 发布的 BERT-large-uncased<sup>[20]</sup> 版本. 模型初始学习率设为  $5e-5$ , 学习率热身系数设为 0.1, 即在前 10% 的训练中, 学习率从 0 递增至预设值, 并在之后线性衰减. 模型的优化算法使用 Adam<sup>[24]</sup>. 其他参数通过对比开发集的实验结果来确定. 模型批处理大小设置为 32, 在 Tesla P40 GPU 上训练迭代 100 个周期, 当连续 10 个周期损失没有降低时

训练过程会提前终止. 本文平均了多个随机种子下的实验结果,用以减少统计误差.

### 3.3 实验结果

表3和表4分别给出了在中文数据集 CrossWOZ 和英文数据集 MultiWOZ 上的实验结果. 针对 CrossWOZ 数据集, 本文与文献[15]提供的 TRADE 模型进行性能比较. CrossWOZ 数据集的特点是大量对话发生在跨领域场景中, 而跨领域或多领域场景会为模型带来更大的噪音, 影响模型预测. 表中 S 表示单一领域; M 表示多领域但是领域内信息并不发生交叉; M+T 表示多领域并且与交通领域相关; CM 表示跨领域, 即领域内信息会发生交叉, 多个槽位可能指向同一实体; CM+T 表示跨领域并且与交通领域相关. 本文的模型与 TRADE 模型相比在 S 场景下联合正确率提升了 0.97%, 在 M 场景下提升了 1.47%, 在 M+T 场景下提升了 1.27%, 在 CM 场景下提升了 2.15%, 在 CM+T 场景下提升了 2.09%, 总体提升了 1.39%. 可以看出 SMGN 模型在多领域和跨领域等复杂场景下性能提升更加明显.

表3 CrossWOZ 数据集实验结果

任务分类	TRADE 联合正确率/%	SMGN 联合正确率/%
Single-domain (S)	71.67	72.63
Independent multi-domain (M)	45.29	46.76
multi-domain+traffic (M+T)	37.98	39.25
Cross multi-domain (CM)	30.77	32.92
Cross multi-domain+traffic (CM+T)	25.65	27.74
Overall	36.08	37.47

表4是在 MultiWOZ 数据集上的实验结果. 为了公平进行对比, 本文将各个模型实验时使用模型的具体情况列在表4中. 其中“Ontology”列表示模型在实验中是否使用预定义本体方法. 预定义本体是指在对话开始之前提前告知模型哪些槽位类型会出现在该对话中, 使用预定义本体会大幅降低对话状态种类数目, 从而降低任务难度. “BERT”列表示该模型在实验中是否使用例如 BERT 等预训练语言模型. 使用 BERT 的模型相比其他模型在词向量编码过程中更有优势, 模型可以捕获到更丰富的语义信息.

本文主要与以下对话状态跟踪方法进行对比.

**DST-Reader**<sup>[7]</sup>: 将对话状态跟踪任务建模为阅读理解任务, 通过预测字段跨度来提取槽位值.

**HyST**<sup>[23]</sup>: 使用分层循环神经网络 (Recurrent Neural Network, RNN) 编码器, 通过混合模型结合预定义本体与开放词表.

**TRADE**<sup>[5]</sup>: 对整个对话历史文本进行编码, 使用复制增强解码器对每个槽位进行解码.

表4 MultiWOZ 数据集实验结果

模型	Ontology	BERT	MultiWOZ 2.0 联合正确率/%	MultiWOZ 2.1 联合正确率/%
DSTreader			39.41	36.40
HyST	√		42.33	38.10
TRADE			48.60	45.60
DSTQA	√		51.44	51.17
SOM-DST		√	51.38	52.57
SUMBT	√	√	48.81	52.75
SST	√		51.17	55.23
SMGN		√	53.03	54.88

**DSTQA**<sup>[6]</sup>: 将对话状态跟踪任务建模为问答任务, 并使用动态演变的知识图来学习槽值对的连接关系.

**SOM-DST**<sup>[24]</sup>: 将对话状态视为一个固定大小的临时记忆, 并提出了选择覆盖机制.

**SUMBT**<sup>[11]</sup>: 使用 BERT 作为话语, 槽位和值的编码器, 通过计算距离来对每个候选值进行评分.

**SST**<sup>[9]</sup>: 利用注意力机制模拟对话状态模式图, 通过循环网络控制状态更新.

从表4可以看出, SMGN 在 2 个数据集上的联合正确率分别为 53.03% 和 54.88%. 在 MultiWOZ 2.0 数据集中本文模型超越其他模型; 在 MultiWOZ 2.1 数据集中, 本文模型性能略弱于 SST 模型, 这是因为 SST 模型是使用预定义本体方法设计的模型. 相比起其他同种类模型, 本文模型的性能有明显提升.

### 3.4 消融实验

在本节中, 本文将进行消融实验, 通过对模型中各个模块进行替换或删除后比较性能, 从而验证 SMGN 模型中各部分改进的有效性. 表5是在 MultiWOZ 2.0 数据集上进行消融实验的结果.

表5 消融实验结果

Model	MultiWOZ 2.0 联合正确率/%
SMGN	53.03
-Graph Update Layer	45.97(-7.06)
GAT-->Transformer	50.25(-2.78)
Value Node Tagger-->State prediction	50.79(-2.24)
BERT-->Word2Vec+BiLSTM	50.74(-2.29)

首先尝试从 SMGN 模型中删除图更新层, 这将使模型变为静态模型, 仅能保留上一轮对话的历史信息, 更早的信息将被覆盖. 这种方法会大幅降低模型的正确率.

其次尝试将模型图更新层中的图注意力 GAT 网络替换为 Transformer 结构, 并使用结构图的连接关系作为注意力权重矩阵. 这使得状态记忆图退化为仅利用边结构的模拟图方法. 可以看出, 替换后性能与隐式利用图网络的 SST 模型性能相近, 这表明显式构建图网络

可以有效提升模型对上文信息的利用能力。

为了与其他不使用BERT预训练语言模型的对话状态跟踪方法进行对比,本文将BERT替换为Word2Vec+双向长短时记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)方法并将其作为模型的编码层进行对比实验。从结果可以看出,使用预训练语言模型可以提升对话状态跟踪任务性能,但即使不使用BERT,状态记忆图网络仍然比其他同类模型性能更好。

最后本文尝试使用状态预测方法代替模型的值结点标记器,实验结果表明在使用状态记忆图的模型中,“标记值结点+结点分类”的方法会比直接进行状态预测的方法效果更好。

### 3.5 效率对比

SMGN模型的另一优点是可以提高计算效率。由于使用了状态记忆图,模型无需在每一轮都对历史文本或对话状态进行重新编码,仅需要保存并传递状态记忆图中的结点信息和连接关系。为了证明以上观点,本文在相同的计算环境下,使用不同模型在MultiWOZ 2.0数据集上进行实验,并对比一轮对话状态跟踪任务所需的平均时间,实验结果如表6所示。其中SOM-DST, SUMBT和本文提出的SMGN模型使用BERT作为编码器。

表6 模型计算效率对比

Model	MultiWOZ 2.0联合正确率/%	时间/ms
TRADE	48.60	979
SOM-DST	51.38	245
SUMBT	48.81	392
SST	51.17	79
SMGN	53.03	165

可以看出,相比TRADE等对话文本编码模型,虽然BERT编码器编码所需时间比RNN要高出数倍,但是本文模型整体运算时间远低于此类模型。与其他状态信息编码模型相比,本文的模型计算效率仅略低于SST模型,这主要是由于使用BERT编码会比其他静态语言模型花费更多计算时间。但横向对比使用BERT作为编码器的SOM-DST和SUMBT模型,本文模型可以提升48%~138%的计算效率。如果对运算效率有严格要求或没有GPU运算资源,SST模型是最佳选择;如果可以使用预训练语言模型,综合考量性能和效率,本文模型优势更明显。

### 3.6 样例分析

为了进一步验证SMGN模型的有效性,本文在CrossWOZ数据集中选取了一个跨领域场景的对话作为样例进行分析,结果如表7所示。

为了便于对比,在表格中省去了两种模型均预测正确的对话状态。在样例中,用户于第二轮对话时使用

表7 CrossWOZ数据集示例结果对比

用户:请帮我在延寿寺附近找一个人均消费100元的餐馆。	
系统:为您推荐盛和轩烤鸭店,人均消费是93元,评分4分。	
用户:行。另外我再问问那里有高档型酒店吗?	
正确结果	(……,餐馆-地点-延寿寺附近,酒店-地点-延寿寺附近)
TRADE	(……,餐馆-地点-延寿寺)
SMGN	(……,餐馆-地点-延寿寺附近,酒店-地点-延寿寺附近)

“那附近”一词代替了景点名称“寿延寺”,而“寿延寺”一词仅出现在上一轮对话历史中。TRADE模型在这个样例中无法正确预测出用户希望寻找“寿延寺”景点附近的酒店,只能判断出酒店类型为高档型。另外,TRADE模型无法有效区分“延寿寺”和“延寿寺附近”这种语义非常相近的槽值,这导致餐馆领域的地点槽位没有找到完全正确的答案。本文的模型在状态记忆图中保留了第一轮对话形成的值结点“寿延寺附近”,在结合当前时刻用户的描述后模型可以正确判断出(酒店-位置-寿延寺附近)对话状态,并且“延寿寺”和“延寿寺附近”会作为两个向量相似但不同的值结点存储在状态记忆图中,使模型可以正确区分相似的槽值。这表明了SMGN模型可以更加有效地处理复杂格式的对话状态生成要求,并且对历史对话的语义理解能力更强。

## 4 总结

针对对话状态跟踪任务,本文提出了一种状态记忆图网络模型,旨在显式利用对话状态自身结构来增加历史信息的语义理解能力。此外,针对复杂格式对话状态生成困难的问题,本文实现了一个基于状态记忆图的对话状态生成方法,通过判断结点连接情况来生成对话状态。因为结点定义自由且易于扩展,该方法可以应对不同格式的生成要求。本文在3个公开的多轮对话数据集上进行了实验。实验结果表明,本文的方法可以有效提高对话状态跟踪任务的性能,同时可以提高运算效率。在今后的研究中会考虑如何将模型修改为并行的多任务学习模型,以解决现有模型会受到错误传播影响的问题。

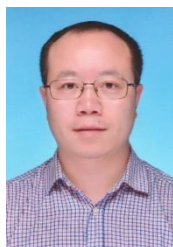
### 参考文献

- [1] HENDERSON M, THOMSON B, YOUNG S. Word-based dialog state tracking with recurrent neural networks [C]//Proceedings of the 15th Annual Meeting of the SIGDI-AL. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014: 292-299.
- [2] MRKŠIĆ N, SÉAGHDHA D Ó, WEN T H, et al. Neural belief tracker: Data-driven dialogue state tracking [C]//Proceedings of the 55th Annual Meeting of the Association

- for Computational Linguistics. Vancouver, Canada: Association for Computational Linguistics, 2017: 1777-1788.
- [3] ZHONG V, XIONG C M, SOCHER R. Global-locally self-attentive encoder for dialogue state tracking[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Volume 1: Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics, 2018: 1458-1467.
- [4] CHAO G L, LANE I. BERT-DST: Scalable end-to-end dialogue state tracking with bidirectional encoder representations from transformer[C]//Proceedings of the 20th Annual Conference of the International Speech Communication Association. Graz, Austria: ISCA, 2019: 1468-1472.
- [5] WU C S, MADOTTO A, HOSSEINI-ASL E, et al. Transferable multi-domain state generator for task-oriented dialogue systems[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 808-819.
- [6] ZHOU L, SMALL K. Multi-domain dialogue state tracking as dynamic knowledge graph enhanced question answering[EB/OL]. (2020-06-20)[2020-12]. <https://arxiv.org/abs/1911.06192>.
- [7] GAO S Y, SETHI A, AGARWAL S, et al. Dialog state tracking: A neural reading comprehension approach[C]//Proceedings of the SIGDial 2019 Conference. Stockholm, Sweden: Association for Computational Linguistics, 2019: 264-273.
- [8] ZHANG J G, HASHIMOTO K, WU C S, et al. Find or classify? Dual strategy for slot-value predictions on multi-domain dialog state tracking[C]//Proceedings of the 9th Joint Conference on Lexical and Computational Semantics. Barcelona, Spain: Association for Computational Linguistics, 2020: 154-167.
- [9] CHEN L, LV B E, WANG C, et al. Schema-guided multi-domain dialogue state tracking with graph attention neural networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(5): 7521-7528.
- [10] ZHU S, LI J Y, CHEN L, et al. Efficient context and schema fusion networks for multi-domain dialogue state tracking[C]//Findings of the Association for Computational Linguistics: EMNLP 2020. Virtual Conference: Association for Computational Linguistics, 2020: 766-781.
- [11] LEE H, LEE J, KIM T Y. SUMBT: Slot-utterance matching for universal and scalable belief tracking[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 5478-5483.
- [12] HUANG Z H, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[EB/OL]. (2015-08-09)[2020-12]. <https://arxiv.org/abs/1508.01991>
- [13] 尹存燕, 黄书剑, 戴新宇, 等. 中英命名实体识别及对齐中的中文分词优化[J]. 电子学报, 2015, 43(8): 1481-1487.
- YIN C Y, HUANG S J, DAI X Y, et al. Optimization of Chinese word segmentation in named entity recognition and word alignment[J]. Acta Electronica Sinica, 2015, 43(8): 1481-1487. (in Chinese)
- [14] 郜成胜, 张君福, 李伟平, 等. 一种基于混合神经网络的命名实体识别与共指消解联合模型[J]. 电子学报, 2020, 48(3): 442-448.
- GAO C S, ZHANG J F, LI W P, et al. A joint model of named entity recognition and coreference resolution based on hybrid neural network[J]. Acta Electronica Sinica, 2020, 48(3): 442-448. (in Chinese)
- [15] FENG Y, WANG Y, LI H. A sequence-to-sequence approach to dialogue state tracking[C]//Proceedings of the 59th ACL and the 11th IJCNLP. Virtual Conference: Association for Computational Linguistics, 2021:1714-1725
- [16] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2018-02-04) [2020-12]. <https://arxiv.org/abs/1710.10903>.
- [17] ZHU Q, HUANG K L, ZHANG Z, et al. CrossWOZ: A large-scale Chinese cross-domain task-oriented dialogue dataset[J]. Transactions of the Association for Computational Linguistics, 2020, (8): 281-295.
- [18] BUDZIANOWSKI P, WEN T H, TSENG B H, et al. MultiWOZ: A large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling[C]//Proceedings of the 2018 Conference on EMNLP. Brussels: Association for Computational Linguistics, 2018: 5016-5026.
- [19] ERIC M, GOEL R, PAUL S, et al. MultiWOZ 2.1: A consolidated multi-domain dialogue dataset with state corrections and state tracking baselines[C]//Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020). Marseille, France: European Language Resources Association, 2020: 422-428
- [20] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of NAACL-HLT 2019. Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019: 4171-4186.

- [21] 张志昌, 曾扬扬, 庞雅丽. 融合语义角色和自注意力机制的中文文本蕴含识别[J]. 电子学报, 2020, 48(11): 2162-2169.  
ZHANG Z C, ZENG Y Y, PANG Y L. A Chinese textual entailment recognition method incorporating semantic role and self-Attention. Acta Electronica Sinica, 2020, 48 (11): 2162-2169. (in Chinese)
- [22] DE BOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method[J]. Annals of Operations Research, 2005, 134(1): 19-67.
- [23] CUI Y M, CHE W X, LIU T, et al. Pre-training with whole word masking for Chinese BERT[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2021, 29: 3504-3514.
- [24] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]//Proceedings of the 3rd International Conference on Learning Representations(ICLR 2015). San Diego: arXiv.org, 2015: 1-13.
- [25] GOEL R, PAUL S, HAKKANI-TÜR D. HyST: A hybrid approach for flexible and accurate dialogue state tracking [C]//Proceedings of the 20th Annual Conference of the International Speech Communication Association. Graz, Austria: ISCA, 2019: 1458-1462.
- [26] KIM S, YANG S, KIM G, et al. Efficient dialogue state tracking by selectively overwriting memory[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Virtual Conference: Association for Computational Linguistics, 2020: 567-582.

## 作者简介



张志昌 男, 1976年4月出生, 甘肃天水人. 教授、硕士生导师. 1998年、2003年和2010年分别在西北师范大学、西北工业大学、哈尔滨工业大学获工学学士、工学硕士和工学博士学位. 研究方向为自然语言处理, 主要进行问答技术、医疗文本处理技术研究.

E-mail: zzc@nwnu.edu.cn



于沛霖 男, 1996年1月出生, 甘肃兰州人. 西北师范大学计算机科学与工程学院硕士研究生. 研究方向为自然语言处理.

E-mail: ypl670335924@gmail.com