面向任务口语对话系统中不含槽信息话语的端到端对话控制*

黄锵嘉，黄沛杰，李杨辉，杜泽峰

（华南农业大学数学与信息学院，广东 广州 510642）

摘要：端到端（end-to-end）模型因其能有效避免传统的管道式设计存在的错误传递和累积问题，成为了近年来口语对话系统（spoken dialogue system，SDS)的研究热点。在面向任务 SDS 的 end-to-end 对话控制中，携带任务领域语义信息（槽信息）的话语的处理可以结合命名实体识别、数据库查询结果等语义特征，而不含槽信息的话语由于缺乏领域语义信息以及表达多样，其有效对话控制仍然是一个挑战。本文提出一种融合“显式”话语特征和“隐式”上下文信息的 end-to-end 混合编码网络用于处理不含槽信息话语。具体上，在应用卷积神经网络（convolutional neural network, CNN）对“显式”话语序列提取得到的特征表达的基础上，通过构造和捕获对话序列中“隐式”的系统后台上下文信息，进一步丰富了系统动作分类模型的特征表达。在限定领域的面向任务中文 SDS 中的评估结果表明，与传统的管道式 SDS 和 end-to-end SDS 相比，本文的方案在不含槽信息话语的单回合处理以及对话段整体性能上都得到了显著的提升。

关键词：口语对话系统；端到端；卷积神经网络；显式话语特征；隐式上下文

中图分类号：TP391　　**文献标识码：**A

End-to-end Dialogue Control for Utterances without Slot Values in Task-oriented with Spoken Dialogue System

HUANG Qiangjia, HUANG Peijie, LI Yanghui, DU Zefeng

(College of Mathematic and Informatics, South China Agricultural University, Guangzhou 510642, China)

Abstract: Recent years, end-to-end frameworks have become research hot pot in NLP, because it can avoid errors propagation and accumulation of traditional pipeline spoken dialogue system (SDS). In end-to-end task-oriented SDS, the dialogue control for utterances with slot values can incorporate named entity recognition and database query results into end-to-end training. However, due to the lack of domain semantic information and the diversity of expression, end-to-end dialogue control of utterances without slot values remains a challenge. To address this issue, this paper proposes an end-to-end hybrid coding network that combines explicit utterance features and implicit context information to handle utterances without slot information. Specifically, beyond the basis features expression of the "explicit" dialogue sequence getting from the convolutional neural network (CNN), the feature expression of the system action classification model is further enriched by constructing and capturing the "implicit" background system context information in the dialogue sequence. The evaluation in task-oriented restricted domain Chinese SDS shows that, compared to traditional pipeline SDS and end-to-end SDS, the proposed method has got significant improvement in per-response accuracy and per-dialog accuracy.

Key words: spoken dialogue system; end-to-end; CNN; explicit utterance features; implicit context

1 引言

面向任务（task-oriented）的限定领域口语对话系统（spoken dialogue system，SDS）是自然语言理解领域内的研究热点之一，并有着广泛的应用场景，如信息咨询[1-4]、商品导购[5]、旅游虚拟助理[6]、导航系统[7-8]等自然语言智能助理。近年来，一系列端到端（end-to-end）

*收稿日期：　　定稿日期：

基金项目：国家自然科学基金(71472068)；国家级大学生创新训练计划项目(201710564154)

SDS 模型 [2-4, 9-11]取得了超过传统的管道式 SDS 模型的性能，从某种意义上归功于它们能有效避免传统的管道式设计存在的错误传递和累积问题。然而，经典的 end-to-end 训练方式比较适合面向非任务（non-task-oriented）SDS [9-11]，通过大量的数据，适合训练有效的聊天机器人类型的 SDS。对于面向任务（task-oriented）的 SDS，则需要进一步解决如何有效捕捉任务领域的信息，并融入到动作策略的选择及应答[12]。

最近的研究已经开始转向尝试以 end-to-end 方式训练面向任务的 SDS[2-4]。通过采用记忆网络的推理方法来建模对话[2]，以及将命名实体识别、数据库查询结果等语义特征结合到基于神经网络的 end-to-end 模型中[3-4]，可以在一定程度上实现含有领域语义槽信息的用户话语的系统应答动作策略的 end-to-end 学习。然而，面向任务 SDS 中表达多样的不含槽信息的话语，既不能像面向聊天 SDS 那样处理，又缺乏含有槽信息话语中的领域语义信息，其有效对话控制仍然是一个挑战。

本文在应用卷积神经网络（convolutional neural network, CNN）对“显式”话语序列提取得到的特征表达的基础上，通过构造和捕获对话序列中“隐式”的系统后台上下文信息，进一步丰富了处理不含槽信息话语的 end-to-end 系统动作分类模型的特征表达。本文的主要贡献包括：

（1）本文提出一种融合“显式”话语特征和“隐式”上下文信息的 end-to-end 混合编码网络处理不含槽信息话语的对话控制。“显式”的话语序列的特征学习有助于捕捉不含槽信息话语的对话行为，“隐式”的系统后台上下文信息支持面向任务的对话策略学习。

（2）在中文限定领域的面向任务 SDS 数据集的评估表明，在不含槽信息话语处理方面，与传统的管道式 SDS 模型和经典的 end-to-end 模型相比，本文的方案能更好地捕捉对话上下文信息，在单回合处理以及对话段整体性能上都得到了显著的提升。

本文后续部分安排如下：下一节介绍相关工作。第 3 节介绍本文提出的方法。第 4 节给出测试结果及分析。最后，第 5 节总结了本文的工作并做了简要的展望。

2 相关工作

2.1 管道式的 SDS

传统的管道式 SDS 由一系列有关联的模块组成，一般包括自动语音识别（automatic speech recognition，ASR）、口语语言理解（spoken language understanding，SLU）、对话管理（dialogue management，DM）、自然语言生成（natural language generation，NLG）、以及语音合成（text-to-speech，TTS）等 5 大模块[13-14]。管道式设计的局限性一方面在于其模块一般是独立训练的，本质上很难将系统适应到新任务领域。例如当其中一个模块采用新数据进行训练或者在设计上进行了改动，与之相关联的模块也需要重新训练或者设计，但是这样的做法需要耗费大量的精力。此外，模块间的错误传递会导致错误累积，并且前序模块的错误会向后续模块传播，不容易追踪和确定错误的来源[15]。

2.2 面向非任务的 end-to-end SDS

面向非任务（non-task-oriented）的 end-to-end SDS [9-11]主要受到序列到序列（sequence to sequence）学习[16]的启发，将对话当做是从源到目标的序列转换问题，采用编码器（encoder）网络[17]将用户话语编码为表示其语义的分布式向量，然后通过解码器（decoder）网络生成系统应答。这些模型通常需要大量的数据来训练。它们适合创建有效的聊天机器人类型的 SDS，但由于缺乏支持完成特定领域任务的能力，例如，与数据库交互[12]并将有用的信息融入到应答中。

2.3 面向任务的 end-to-end SDS

近年来，研究者们开始研究适合面向任务 SDS 的 end-to-end 可训练模型。Wen 等[4]提出

了一种模块化连接的基于神经网络的 end-to-end 可训练模型。该模型分离了用于意图估计、状态确认、策略学习和响应生成等模块。Bordes 和 Weston[2]提出了一种使用 end-to-end 记忆网络的推理方法来建模对话。他们的模型直接从响应列表中选择系统响应，而没有经过状态追踪模块。Williams 等人[3]将命名实体识别、数据库查询结果等语义特征结合到基于神经网络的 end-to-end 模型中，优化了系统动作策略学习的性能。

表 1 是一个面向手机导购的 SDS 的示例。在目前研究进展中，对于携带任务领域语义信息（槽信息）的用户话语（表 1 对话话语中的（4）、（6）、（8）、（12）、（16））， 采用已有的研究方法，可以有效地实现从用户话语到系统动作的 end-to-end 学习，无需通过管道式设计的槽信息填充、对话行为识别和对话策略等模块化的独立学习过程。具体上，如果是“陈述句”，模型输出“主动引导”的系统动作，记录下用户提供的信息并引导用户进一步提供更多必要信息；如果是疑问句，模型输出“查询并回复”的系统动作，结合话语相应的问句类型（如（8）的“特指问”、（12）的“选择问”）查询数据库并回复。

表 1　面向任务的 SDS 示例 （None 表示不含槽信息）

对话话语	用户话语的槽信息	用户对话行为 / 系统动作
(1) 系统：您好！	—	问候
(2) 用户：你好！	**None**	**问候**
(3) 系统：您想买什么品牌的手机呢？	—	主动引导
(4) 用户：我要三星的。	[品牌, 三星]	陈述句
(5) 系统：您打算买什么价位的？	—	主动引导
(6) 用户：2000 块左右的吧。	[价格, 2000 块左右]	陈述句
(7) 系统：您想要什么输入方式的？	—	主动引导
(8) 用户：触屏手机有什么好介绍？	[操作方式, 触屏]	特指问
(9) 系统：有 E120L 和 I9050。	—	查询并回复
(10) 用户：我也不知道哪个好。	**None**	**表态犹豫**
(11) 系统：您可以考虑屏幕大小。	—	选择差异参数
(12) 用户：哪个大一点？	[屏幕尺寸, 大]	选择问
(13) 系统：E120L。	—	查询并回复
(14) 用户：嗯，不过我不喜欢。	**None**	**表态否定**
(15) 系统：那要不这款 I9050。	—	选择其他手机
(16) 用户：这款摄像头像素高吗？	[摄像头像素, 高]	是非问
(17) 系统：还可以，500 万像素。	—	查询并回答
(18) 用户：那就这款吧。	**None**	**成交**
(19) 系统：很高兴能够帮到您!	—	成交结束语

然而，对于不含槽信息的用户话语（表 1 对话话语中的（2）、（10）、（14）、（18）），其表述更像面向聊天的用户话语，但又不能像面向聊天 SDS 那样处理话语，因为那样会脱离所面向的任务的流程，不利于任务的完成。另一方面，相比于含有槽信息的话语，不含槽信息话语由于缺乏领域语义信息，并且此类话语表达多样，其有效对话控制仍然是一个挑战。

本文在应用卷积神经网络（convolutional neural network, CNN）对“显式”话语序列提取得到的特征表达的基础上，通过构造和捕获对话序列中“隐式”的系统后台上下文信息，进一步丰富了 end-to-end 系统动作学习模型的特征表达，实现不含槽信息话语的 end-to-end 对话控制。

3 结合“隐式”上下文信息的 end-to-end 模型

3.1 总体技术框架

图 1 是本文提出的处理不含槽信息话语的融合“显式”话语特征和“隐式”上下文信息的 end-to-end 混合编码网络模型结构图。

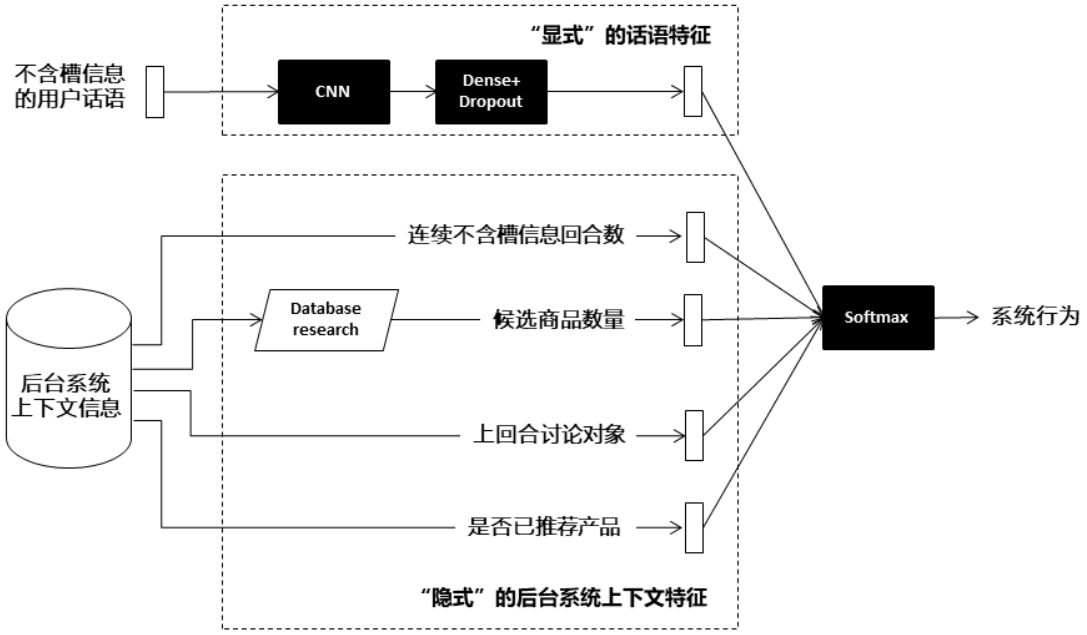

图 1　融合“显式”话语特征和“隐式”上下文信息的 end-to-end 混合编码网络模型结构（以商品导购 SDS 为例）

（1）“显式”的话语特征：口语话语是一种典型的短文本，因此本文采用了在短文本理解中广泛采用的 CNN 提取话语的特征表达。用户句子首先通过句子特征表达完成从自然语言到低维实数向量的映射，完成句子特征的表达。通过 CNN 不同的卷积窗口大小对输入特征进行卷积操作，提取不同粒度的特征组合，池化层实现对卷积特征的聚合统计，在保留重要特征的同时对特征起到降维作用。在池化层后面是 dense 层（隐藏层），dense 层能够对卷积层和池化层提取到的多粒度词语特征进行权衡，并选择出最有价值的特征，而 dropout 机制能够有效防止模型过拟合。

（2）“隐式”的后台系统上下文特征：后台系统上下文信息主要包括两个方面的作用，一方面是更为准确地“捕捉”对话上下文的概要信息，比如用户“连续不含槽信息的回合数”、“上回合讨论对象”（具体商品还是只是某一商品属性，或者是商品无关）、“是否已推荐商品”，单纯依靠对话话语序列输入的 end-to-end 模型，很难直接通过对话序列“获取”这些信息；另一方面是提供直接从对话序列无法获取的信息，比如“候选商品数量”是系统内部掌握的信息，可由系统根据当前的对话上下文已确定的商品槽信息查询数据库得到。值得注意的是，对于含有槽信息的话语的处理，当前的主流模型，不论是管道式 SDS 还是 end-to-end SDS，都带了商品槽信息获取。

最后，将“显式”的话语特征和“隐式”的后台系统上下文特征进行拼接，并连接 softmax 分类层，输出对不同系统动作预测的概率分布。

3.2 卷积神经网络模型

卷积神经网络（convolutional neural network, CNN）在短文本理解方面取得较好的应用效果[18-19]。本文采用了Kim提出的卷积神经网络结构模型[18]，该模型采用了多种核尺寸实现对句子进行不同粒度特征的提取，不同粒度的特征对实验结果有改善的作用。

在原始的全连接神经网络中，假设第 l 隐藏层有 m^l 个节点，第 $(l-1)$ 隐藏层有 n^{l-1} 个节点，这时连接第 l 层与第 $(l-1)$ 层的参数个数将达到 $n^{l-1}\times m^l$ 个。当 n^{l-1} 和 m^l 都很大的时候，参数空间会很大，也意味着计算量加大，从而导致训练过程也会特别慢。因此，采用卷积神经网络可以很好的解决这个问题，卷积神经网络第 l 层的神经元只与第 $(l-1)$ 层的局部相连接，这时连接第 l 层与第 $(l-1)$ 层的参数个数为 $(n^{l-1}-m^l+1)$：

$$\mathrm{h}^l = f(W^{(l)} \otimes h^{(l-1)} + b^{(l)}) \quad (1)$$

其中 h^l 表示第 l 层的输出，$\mathrm{W}^{(l)}$ 表示连接第 $(l-1)$ 层和 l 层的滤波器，$\otimes$ 代表卷积运算，$\mathrm{b}^{(1)}$ 为偏置，f 为激活函数。

卷积操作的应用虽然使得连接层与层之间的权重参数空间大大减少，但卷积操作后输出的神经元个数与全连接的神经网络相比并没有多大改变，如果此时在卷积结果后接一个 softmax 分类层，那么其参数空间仍然相当庞大。因此通常的做法是在卷积操作之后执行一个采样操作，也称为池化（pooling）。池化操作不仅能进一步减少参数的个数，还能降低特征的维度，从而避免了过拟合现象[20]。

$$\mathrm{h}^l = f(W^{(l)} \otimes pool(h^{(l-1)}) + b^{(l)}) \quad (2)$$

其中 $pool$ 表示池化操作，常见的有最大池化和平均池化，本文采用最大池化（max-pooling）作为采样函数。

3.3 “隐式”的后台系统上下文信息

本文提出的方案中，从对话信息流中持续采集的“隐式”的后台系统上下文信息如表 2 所示。本文期望通过增加这 4 类后台系统上下文信息，可以有助于捕捉对话流程的“隐式”上下文（相比于“理解”用户话语的文本信息得到的“显式”特征），有效地实现不含槽信息话语的 end-to-end 对话控制。

表 2 “隐式”系统上下文特征

后台系统上下文信息	特征构造说明
连续不含槽信息回合数	0：表示回合数<3；1：表示回合数≥3
是否已推荐商品	0：表示系统未推荐手机；1：表示系统已推荐了手机
上回合讨论对象	0：表示讨论与商品无关；1：表示讨论特定商品；2：表示讨论商品的属性
候选商品数量	0：表示 0 款；1：表示 1 或 2 款；2：表示>2 款

（1）连续不含槽信息回合数：该特征记录了用户连续没有主动提供商品相关信息（槽信息）的回合数。一般而言，出现这种情况或者是用户缺乏主动主导对话的习惯，或者是用户进入了连续闲聊的状态。加入这个特征有助于增加模型选择“主动引导”动作的概率，将对话“带回”导购流程。

（2）是否已推荐商品：该特征用于标识“推荐商品”这个系统动作是否出现过。系统是否已经向用户推荐过商品，潜在有助于对用户的一些对话行为，如“表态肯定”、“表态否定”等做出正确的系统动作选择。

（3）上回合讨论对象：该特征不容易直接通过模型从对话序列上下文直接“理解”得到，但系统却能做出较为准确的记录。上回合讨论的对象的不同，也会影响系统做出不同的

系统动作选择。

（4）候选商品数量：该特征结合系统当前的对话上下文已确定的商品槽信息查询数据库得到。一般而言，在候选商品依然较多时宜继续引导用户表达特定商品属性的需求，而在候选商品较少时则可能适时推荐会更合适。尤其是，候选为空时，显然需要及时告知用户。

4 实验

4.1 数据集

实验数据采用中文手机导购领域的对话语料库。训练数据从多年的系统日志中整理得到，选取了 516 段对话过程没出现系统错误，并且包含有不含槽信息话语的对话段，用于模型训练及模型超参数调节。在训练集中，用户话语 4938 句，不含槽信息的话语为 1180 句，占了 23.89%。

测试数据采用 13 名测试人员测试产生的对话过程没出现系统错误，并且包含有不含槽信息话语的对话段，共 159 段，用于本文提出方法与研究进展方法的评估和对比。训练和测试数据的总体统计结果如表 3 所示。

本文用于训练 word2vector 模型的数据是由中国中文信息学会社会媒体专委会提供的 SMP2015 微博数据集（SMP 2015 Weibo DataSet）。本文只使用了该数据集的一个子集（1.5G），该子集超过 1000 万条微博数据，词汇表的词语数为 519,734。

表 3　数据集的统计结果

数据组成	对话段数	用户话语数/不含槽信息的话语数	不含槽信息话语占比
训练数据	516	4938/1180	23.89%
测试数据	159	2144/584	27.22%

4.2 系统动作

系统动作是系统根据当前的上下文情况以及与用户交互过程中可采用的对话行为，也是本文的 end-to-end 对话控制的输出。系统确定了系统动作之后，再结合一些来自数据库的信息生成完整的系统应答。

在目前版本的中文手机导购 SDS 中，共有 12 种系统动作。其中包括社交动作的设计，使得用户对系统应答的体验更加自然，如“问候语”、“闲聊结束语”等。而任务相关动作设计的目的在于协助用户进行导购，如“主动引导”、“推荐候选手机”等动作。如表 4 所示。

表 4　系统动作（实验采用的手机导购 SDS）

系统动作	话语示例
1 问候语	"早上好"
2 主动引导	"请问您想买什么手机呢"
3 购买确认	"亲，确定要拍下这款手机了吗？"
4 成交结束语	"成交，您的肯定是我前进的动力"
5 选择其它手机	"那您看看这部***怎么样？"
6 属性值确认	"那么您对品牌有什么要求呢？"
7 闲聊结束语	"拜拜"
8 选择差异参数	"这款的屏幕比较大"（2 款及以上候选手机）
9 输出若干优势	"这台手机屏幕大"（1 款候选手机）

10 选择候选属性	"没有这样的手机，对尺寸方面有什么要求"
11 推荐候选手机	"看看华为 XX 怎么样？"
12 表情附和	"^_^ "

4.3 实验设置

本文所进行的实验包括了 CNN 的特征选择以及研究进展方法的对比，实验中模型训练的参数调节均采用 k-折（本文采用 5 折）交叉验证。

本文方法，融合“显式”话语特征和“隐式”上下文信息的 end-to-end 混合编码网络（Explicit and Implicit Context Hybrid Code Network，EIC-HCN），对比的三种研究进展方法如下：

（1）DA+DM：该方法作为对比的传统的管道式 SDS 方法，对话行为（DA）识别采用了 Wang 等人提出 CNN-RF 混合模型[21]，对话管理（DM）采用 POMDP 进行建模[14]。

（2）MemN2N：采用了 Bordes 和 Weston[2]提出的记忆网络对 SDS 进行建模，实现了 SDS 以 end-to-end 的方式进行学习。

（3）CNN(BOC(N_utterances))：该方法作为经典的 end-to-end 模型[19]的另一个代表，将原始的序列对话中的近 N 个用户话语作为输入，并最终输出系统动作的预测。在本文实验中，经过 5 折的交叉验证，该对比方法选择了 N=3。

测试采用了 Bordes 和 Weston[2]实验中的评估方法，对单回合和对话段两个层面进行评估。

4.4 评价方法

目前针对口语对话系统性能的评价没有一个统一标准[22]。有直接以人工标准（Gold-standard）作为评判[23]，较为客观，但由于特定用户话语潜在存在不止一个合适的系统应答动作，所以偏向“严格”。也有采用主观的评价，例如，人工对系统的预测结果进行满意度的评分[24-25]，但这样的做法容易受主观性影响。也有其他研究工作为每一个用户的请求定义了一个候选集[2, 26]，若系统预测结果在候选集合里，则认为是正确的，但人工定义这样的候选集需要耗费较多的人力资源。综合已有的评价标准，本文基于中文手机导购系统为背景，采用了以下的评判方法：

（1） 客观评价

对测试集中每一句不含槽信息的用户话语对应的系统应答的系统动作进行人工检查确认，然后以人工标准（Gold-standard）作为评判，模型输出和 Gold-standard 一致时为正确，不一致时为错误。

（2） 主观评价

对于不同方法预测的系统动作，人工对系统的预测结果进行满意度的评价，满意判为正确，不满意的判为错误。本文对以下三种情况判为不满意（也即是错误）。

- 答非所问：系统回答与用户询问存在逻辑或语义的错误，如系统选择“闲聊结束语”去应答用户的问候等。

- 违反上下文：系统应答动作与当前用户的历史上下文相矛盾，如在“未推荐手机”的情况下，执行“成交结束语”、“购买确认”、“选择其它手机”等动作。

- 态度消极：系统在应对导购任务进程采取消极的应对动作，具体表现如，在 “已推荐手机”的情况下没有积极地向用户做出“购买确认”动作；在满足向用户推荐手机的条件下未向用户做出“推荐候选手机”动作；或在导购任务未完成的条件下，多次持续地与用户进行任务无关的交谈。

4.5 实验结果及分析

4.5.1 客观评价

（1）原始特征选择

本文对比字袋（BOC）、词袋（BOW）、word2vector（w2v）、position encoding（PE）[12]等不同 CNN 原始特征表达的效果。结果如表 5 所示。w2v 特征表达中对句子字数的句子编码进行验证。

表 5　不同原始特征识别效果

方法	验证正确率（%）
EIC-HCN(PE)	83.37
EIC-HCN(w2v)	84.75
EIC-HCN(BOW)	90.50
EIC-HCN(BOC)	91.25

从表 5 可看出，BOC 和 BOW 的特征表达均优于 w2v 和 PE 的特征表达，字相比词在 SDS 领域中能够更好的对句子特征进行表达，这也体现了 SDS 中话语口语化的特点。

（2）研究进展方法对比（客观评价）

本文提出的方法与研究进展方法对比结果如表 6 所示。本文的方法选用了字作为原始特征，采用了字袋的向量表达方式（BOC），MemN2N 模型和 CNN(BOC(N_utterances))模型采用原始的对话序列作为上下文。

表 6　本文提出方法与研究进展方法的对比（采用客观评价）

方法	单回合正确率（%）	对话段正确率（%）
DA+DM	55.99	31.45
MemN2N	59.93	30.19
CNN(BOC(N_utterances))	66.61	40.25
EIC-HCN(BOC)	83.39	64.78

从表 6 测试结果表明，本文提出的方法比传统管道式模型（DA+DM）、记忆网络模型 MemN2N、以及经典的 end-to-end CNN 模型 CNN(BOC(N_utterances))在单回合正确率分别提高了 27.4%、23.46%和 16.78%，在对话段评估上，本文提出的方法比三者正确率分别提高了 33.33%、34.59%和 24.53%。这表明本文的方案能够较好的解决不含槽信息话语 end-to-end 对话控制的挑战。此外，虽然经典的 End-to-end 模型在非任务领域有优异的表现，但在任务领域的性能并没有显著地超过管道式 SDS 模型。本文融入模型的“隐式”上下文信息有效地弥补了经典 end-to-end SDS 在用户对话上下文刻画方面的不足。

4.5.2 主观评价

（1）研究进展方法对比（主观评价）

我们进一步以主观评价为标准，对比了本文提出的方法 EIC-HCN 与研究进展方法，结果如表 7 所示。

表 7　本文提出方法与研究进展方法的对比（采用主观评价）

方法	单回合正确率（%）	对话段正确率（%）
DA+DM	71.58	50.94
MemN2N	68.32	35.85
CNN(BOC(N_utterances))	73.80	47.17
EIC-HCN(BOC)	87.67	69.81

从表 7 可以看到，从主观评价角度，所有模型的正确率（人工评价地系统动作的满意率）都得到一定程度的提升。尤其是管道式 SDS 取得了和两个经典 end-to-end 模型相当的单回合正确率。我们的模型取得了优于对比方法的性能，正确率达到了接近 90%。

（2）细分错误类别分析

表 8 给出了不同方法的单回合动作预测错误（也即是人工评价中的不满意）分类，为了方便不同方法之间的横向比较，统计数据采用了每种错误类型的回合数占总回合数的比例。在具体的分析中，存在着同时满足 2 个或以上错误分类的例子，所以不同错误分类的占比总和会略大于总错误率。

表 8　不同方法的单回合错误分类（比例为全部测试样例的百分比）

方法	答非所问（%）	违反上下文（%）	态度消极（%）	总错误率（%）
DA+DM	18.32	11.99	7.71	28.42
MemN2N	22.77	16.61	4.11	31.68
CNN(BOC(N_utterances))	19.01	16.78	3.08	26.20
EIC-HCN(BOC)	7.02	0.86	5.48	12.33

从表 8 可以看出，在不同方法的错误分类中，“答非所问”和“违反上下文”是错误分类的主要组成部分。同时，相比研究进展方法，本文提出的方法在“答非所问”和“违反上下文”的错误率得到明显的减少，而“态度消极”需在系统任务和用户体验感之间做出权衡，相比研究进展方法，本文方法在该错误分类中并没有得到明显的改善。

具体错误分析如下：

（1）答非所问：造成“答非所问”错误的主要原因在于对用户话语的语义理解失误，MemN2N 方案采用的记忆网络并不能较好的结合用户话语中字或词之间的联系，错误最多。而 CNN 能够通过卷积操作实现对用户话语语义的有效提取，实验结果得到一定的改善。本文提出方案 EIC-HCN(BOC)显著地将该分类的错误率降低到 7.02%。

（2）违反上下文：MemN2N 和 CNN(BOC(N_utterances))方案采用经典的 End-to-end 模型通过原始对话序列去捕获对话的上下文信息，两个方案在该错误分类占比分别为 16.79%和 17.86%，这说明记忆网络和单纯依靠“显示”话语特征的 CNN 并不适应长距离的对话任务。传统管道式 SDS（DA+DM）和本文方法 EIC-HCN(BOC)都采用了一些构造性的上下文特征，能够有效的克服上下文信息难以捕获的问题。本文的方法则还避免了管道式模型的错误传递，在该错误分类中，本文方案将错误率降低到 0.86%。

（3）态度消极：本文着重对“答非所问”和“违反上下文”错误分类进行研究和分析，这是因为大部分错误是由这两类错误引起的。在“态度消极”这个分类上，本文方法与研究进展方法的实验结果并没有显著的差别，未来的工作会对该类别进行改善。

5 结束语

为了有效地处理面向任务 SDS 中不含槽信息话语，本文提出了一种融合“显式”话语特征和“隐式”上下文信息的 end-to-end 混合编码网络模型。具体上，在应用 CNN 对“显式”话语序列提取得到的特征表达的基础上，通过构造和捕获对话序列中“隐式”的系统后台上下文信息，进一步丰富了系统动作分类模型的特征表达。在中文手机导购领域 SDS 的测试表明，本文的方法取得了优于研究进展中管道式 SDS、经典的记忆网络和 CNN 训练的 end-to-end 模型的应用效果，在客观评价和主观评价两方面，都取得了单回合准确率和对话段准确率的显著性能提升。未来工作主要是完善“态度消极”错误分类的改进探索，进一步丰富后台系统上下文特征的构造，以及采用 AMT (the Amazon Mechanical Turk) 服务[27]进行

更大规模的测试。

参考文献

[1] Zue V, Seneff S, Glass J, et al. JUPITER: a telephone-based conversational interface for weather information[J]. IEEE Transactions on Speech and Audio Processing, 2000, 8(1): 85−96.

[2] Bordes A, Weston J. Learning end-to-end goal-oriented dialog[C]// Proceedings of the 5th International Conferenceon Learning Representations(ICLR 2017), 2017:1-15.

[3] Williams J D, Asadi K, Zweig G. Hybrid Code Networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), 2015: 665-677.

[4] 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 (EACL 2017), 2017: 438-449.

[5] Huang P J, Lin X M, Lian Z Q, et al. Ch2R: a Chinese chatter robot for online shopping guide[C]//Proceedings of the 3rd CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP-2014), 2014: 26-34.

[6] Pappu A, Rudnicky A. The structure and generality of spoken route instructions[C]//Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2012), 2012: 99-107.

[7] 黄寅飞, 郑方, 燕鹏举, 等. 校园导航系统EasyNav的设计与实现[J].中文信息学报, 2001, 15(4): 35-40.

[8] Reichel C S, Sohn J, Ehrlich U, et al. Out-of-domain spoken dialogs in the car: a WoZ study[C]//Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2014), 2014: 12-21.

[9] Vinyals O and Le Q V. A neural conversational model [C]// Proceedings of the ICML Deep Learning Workshop, Lille, France, 2015.

[10] Shang L F, Lu Z D and Li H. Neural responding machine for short-text conversation [C]// Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics (ACL 2015), 2015: 1577–1586.

[11] Serban I V, Sordoni A, Bengio Y, et al. Hierarchical neural network generative models for movie dialogues [C]// Proceedings of the 13th AAAI Conference on Artificial Intelligence (AAAI-16), 3776-3783.

[12] Sukhbaatar S, Szlam A, Weston J. End-to-end memory networks [C]// Proceedings of the Twenty-ninth Annual Conference on Neural Information Processing Systems (NIPS 2015), 2015:2440-2448.

[13] Raux A, Langner B, Dan B, et al. Let's go public! taking a spoken dialog system to the real world[C]// Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH 2005), 2005:885-888.

[14] Young S, Gǎsić M, Thomson, B, et al.. POMDP-based statistical spoken dialog systems: A review [J]. Proceedings of IEEE, 2013, 101(5):1160-1179.

[15] Zhao T C and Eskenazi M. Towards End-to-end Learning for dialog state tracking and management using deep reinforcement learning[C]// Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue(SIGDIAL 2016) , 2016:1-10.

[16] Sutskever I, Vinyals O, and Le Q V. Sequence to sequence learning with neural networks [C]// Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS 2014), 2014: 3104–3112.

[17] Cho K., Merrienboer B, Gulcehre C, et al. Learning phrase representations using rnn encoder–decoder for statistical machine translation [C]// Proceedings of the 19th Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), 2014: 1724–1734.

[18] Kim Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 19th Conference on

Empirical Methods in Natural Language Processing (EMNLP 2014), 2014: 1746–1751.

[19] Severyn A, Moschitti A. Learning to rank short text pairs with convolutional deep neural networks. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, (ACM 2015), 2015: 373–382.

[20] Ciresan D C, Meier U, Masci J, et al.. Flexible, high performance convolutional neural networks for image classification[C]//Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011), 2011:1237-1242.

[21] Wang J D, Huang P J, Huang Q J, et al.. Dialogue act recognition for Chinese out-of-domain utterances using hybrid CNN-RF[C]//Proceedings of the 20th International Conference on Asian Language Processing (IALP 2016), 2016:14-17.

[22] 张伟男, 张杨子, 刘挺. 对话系统评价方法综述[J]. 中国科学:信息科学, 2017,43(8): 954-966.

[23] Yang X S, Chen Y N, Hakkani-Tür D, et al. End-to-end joint learning of natural language understanding and dialogue manager[C]// Proceedings of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017), 2017: 5690-5694.

[24] Ritter A, Cherry C, and Dolan W B, et al.. Data-driven response generation in social media[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2011), 2011:583-593.

[25] Shang L F, Lu Z D, and Li H. Neural responding machine for short-text conversation[C]//Proceedings of the 53st Annual Meeting of the Association for Computational Linguistics(ACL 2015), 2015:1577-1586.

[26] 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 (SIGDIAL 2015), 2015: 285-294.

[27] Jurčíček F, Keizer S, Gašić M, et al. Real user evaluation of spoken dialogue systems using Amazon Mechanical Turk [C]// Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011), 2011: 3061–3064.

作者简介：

黄锵嘉（1994—），本科，主要研究领域为自然语言处理。
Email: qjHuang1024@163.com

黄沛杰（1980—），通讯作者，博士，副教授，主要研究领域为人工智能、自然语言处理、口语对话系统。
Email: pjhuang@scau.edu.cn

李杨辉（1995—），本科，主要研究领域为自然语言处理。
Email: lyh_liyanghui@163.com