

Mem2Seq:有效地将知识库纳入端到端
面向任务的对话系统

安德里亚·马多托* , Chien-Sheng Wu , 冯柏斯
人类语言技术中心
人工智能研究中心 (CAiRE)
电子与计算机工程系
香港科技大学,清水湾,香港
[eeandreamad,cwuak,pascale]@ust.hk

抽象的

端到端的面向任务的对话系统
通常会面临整合知识库的挑战。在本文中,我们提出了一种新颖而简单的端到端可微分模型,称为内存到序列 (Mem2Seq)来解决这个问题。Mem2Seq 是第一个将记忆的多跳注意力与

指针网络的想法。我们凭经验
展示 Mem2Seq 如何控制每个生成步骤,以及它的多跳注意力如何
机制有助于学习相关性
记忆之间。此外,我们的模型
非常通用,没有复杂的任务特定设计。结果,我们表明

Mem2Seq 可以更快地训练并达到
三个方面最先进的性能
不同的面向任务的对话数据集。

1 简介

面向任务的对话系统帮助用户实现
使用自然语言实现特定目标,例如
餐厅预订和日程安排。
传统上,它们是由多个
流水线模块:语言理解、对话日志管理、知识查询、语言

一代 (Williams 和 Young, 2007 年; Hori 等人, 2009年;李等人, 2009;莱文等人, 2000;年轻的
等人, 2013) 。此外,查询外部知识库 (KB) 的能力在面向任务的
对话系统中至关重要,因为响应是

不仅受对话历史的指导,而且受
查询结果 (如表1) 。然而,尽管
通过结合特定领域的知识和槽填充来保证此类管道系统的稳定性

* 这两位作者的贡献是相等的。

兴趣点 (poi) 距离	交通信息	然后输入	地址
威斯汀酒店 5 英里 交通状况一般		休息站拔	329 皇家路
圆桌会议 4 英里无交通		萨店	113 Anton Ct 中
Mandarin Roots 5 英里无交通		餐厅 271 Springer Street	
• Palo Alto Cafe 4 英里交通适中的咖啡或茶馆			第436座
多米诺骨牌 6 英里交通繁忙的披萨餐厅			阿拉斯特德罗路176号
斯坦福 Express Care 6 英里医院休息站	没有交通		214 皇家路
基恩酒店	2英里	交通繁忙	第578章 树博士

司机我在哪里可以喝到茶？	
Seq2Seq 我有一个很远的地方,你想要地址吗	
+Attn 我有一个 5 英里外供应茶的地方的列表	
Ptr-Unk 有一个您喜欢的地方,您想了解更多信息吗	
内存到序列	最近的茶馆是距离酒店 4 英里的 Palo Alto Cafe, 您想知道那里的路线吗?
金子	Palo Alto Cafe咖啡馆距离酒店有4英里 (6.4公里) ,供应咖啡和茶。 你想要地址吗?

司机是的。	
Seq2Seq 位于	
+收件人	地址是 329 El Camino Real,距离酒店 3 英里 并且没有交通
Ptr-Unk Palo Alto Cafe 距离酒店 4 英里,PAD 不堵车	
Mem2Seq Palo Alto Cafe 距离酒店 4 英里 (6 公里) ,地址:436 Alger Drive	
GOLD Palo Alto 位于 436 Alger Dr.	

表 1:生成的响应示例
导航领域的车载助手。

技术,对之间的依赖关系进行建模
模块很复杂,知识库解释需要人工。

最近,端到端的对话方法
建模,使用循环神经网络
(RNN)编码器-解码器模型,已显示
有希望的结果 (Serban 等人, 2016; Wen 等人, 2017年;赵等人, 2017) 。因为他们可以直接
将纯文本对话历史映射到输出响应,并且对话状态是潜在的,有

无需手工制作的国家标签。而且,
基于注意力的复制机制 (Gulcehre 等人, 2016年;埃里克和曼宁, 2017)最近
引入直接从输入复制单词
输出响应的来源。使用这种机制,即使未知令牌出现在

对话历史记录,模型仍然能够生成
正确且相关的实体。

然而,尽管上述方法取得了成功,但仍然存在以下问题:

两个主要问题:1)他们很难有效地将外部知识库信息合并到

RNN 隐藏状态 (Sukhbaatar 等人, 2015) ,

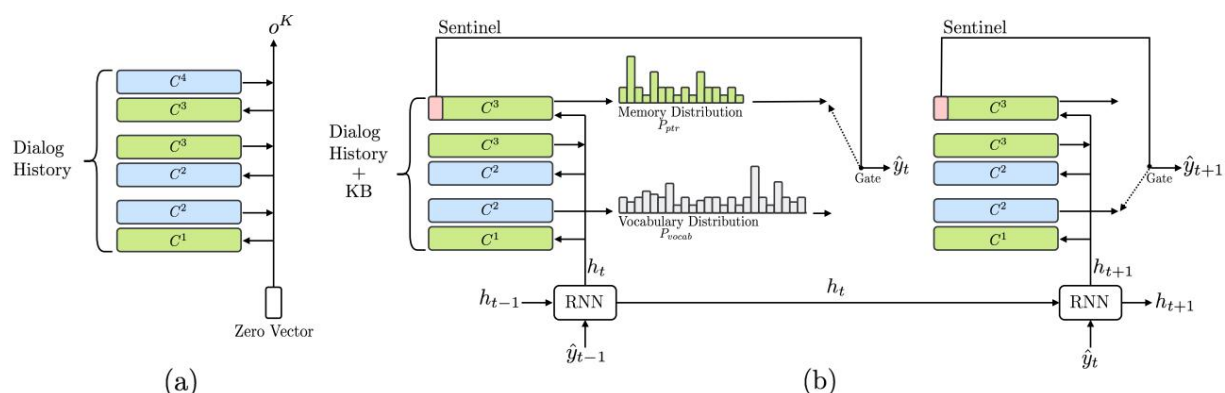


图 1: 针对面向任务的对话系统提出的 Mem2Seq 架构。(a) 内存编码器有 3 跳; (b) 两步生成的内存解码器。

因为众所周知,RNN 长期不稳定序列。2)处理长序列非常困难耗时,尤其是在使用注意力机制。

另一方面,端到端内存网络 (MemNN)是循环注意力在可能很大的外部存储器上建立模型 (Sukhbaatar 等人, 2015)。他们将外部记忆写入几个嵌入矩阵中,

并使用查询向量重复读取内存。这种方法可以记忆外部知识库信息并快速编码长对话历史。

此外,MemNN的多跳机制经验表明,对于实现推理任务的高性能至关重要 (Bordes

和韦斯顿, 2017)。然而,MemNN 只是从预定义的候选数据池中选择其响应,而不是逐字生成。在

另外,内存查询需要明确的设计而不是被学习,以及复制机制缺席。

为了解决这些问题,我们提出了一个新的我们称之为“内存到序列”的架构 (Mem2Seq) 学习面向任务的对话端到端的方式。简而言之,我们的模型增强了现有的具有顺序生成架构的 MemNN 框架,使用全局多跳注意力机制直接复制单词

来自对话历史记录或知识库。我们总结了我们的主要贡献如下: 1)Mem2Seq 是第一个将多跳注意力机制与指针网络思想相结合的模式,

使我们能够有效地整合知识库信息。2)Mem2Seq学习如何生成动态查询来控制内存访问。此外,我们还可可视化并解释了内存控制器各跳之间的模型动态

和关注度。3)Mem2Seq可以训练更快并在多个方面取得最先进的结果面向任务的对话数据集。

2 型号说明

Mem2Seq 由两个组件组成: MemNN 编码器和内存解码器为如图1所示。MemNN 编码器创建对话历史记录向量的表示。

然后内存解码器读取并复制记忆来产生响应。我们定义所有的对话历史记录中的单词以 $\text{ken } X = \{x_1, \dots, x_n, \$\}$, 其中 $\$$ 是特殊的包机用作哨兵,KB 元组用作 $B = \{b_1, \dots, b_l\}$ 。我们进一步定义 $U = [B; X]$ 作为两个集合 X 和 B 的串联, $Y = \{y_1, \dots, y_m\}$ 作为预期中的单词集系统响应,并且 $\text{PTR} = \{\text{ptr}_1, \dots, \text{ptr}_m\}$ 作为指针索引集:

$$\text{ptr}_i = \begin{cases} \max(z) & \text{如果 } \exists z \text{ st } y_i = u_z \\ n + l + 1 & \text{否则} \end{cases} \quad (1)$$

其中 $u_z \in U$ 是输入序列, $n + l + 1$ 是哨兵位置指数。

2.1 内存编码器

Mem2Seq 使用标准 MemNN 和相邻的加权捆绑 (Sukhbaatar et al., 2015) 作为编码器。编码器的输入是 U 中的字级信息。MemNN 的存储器是

由一组可训练的嵌入矩阵表示 $C = \{C^1, \dots, C^{K+1}\}$, 其中每个 C^k 地图向量的标记和查询向量 q 一个读头。该模型循环 K 跳,

1代码可在<https://github.com/HLTCHKUST/Mem2Seq>获取

它使用以下方法计算每个内存存在第 k 跳的注意力权重：

$$c_{i-k} = \text{Softmax}((q_{k-1}^{(i)} \cdot c_{i-k}^{(k)})), \quad (2)$$

其中 $c_{i-k}^{(k)} = C^{(k)}(x_i)$ 是 p_o 中的内存内容

位置 i , $\text{Softmax}(z_i) = e^{z_i} / \sum_j e^{z_j}$ 。这里, p 是一个软内存选择器, k 决定 mem

与查询向量 q 的相关性

然后,模型通过 C^{k+1} 的加权和读出内存 o_k

$$o_k = \sum_{i=1}^{k+1} c_{i-k}^{(k+1)} \cdot o_k. \quad (3)$$

然后,使用 q 更新下一跳 $k+1$ 的查询向量。编码步骤的结果是记忆向量 o_k , 它将成为解码步骤的输入克+奥。

2.2 内存解码器

解码器使用 RNN 和 MemNN。MemNN 加载了 X 和 B , 因为我们使用对话历史记录和 KB 信息来生成正确的系统响应。门控电流单元 (GRU) (Chung 等人, 2014) 用作 MemNN 的动态查询生成器。在每个解码步骤 t , GRU 获取先前生成的单词和先前的查询作为输入, 并生成新的查询向量。正式：

$$h_t = \text{GRU}(C^{(1)}(y_{t-1}), h_{t-1}); \quad (4)$$

然后将查询 h_t 传递到 MemNN, MemNN 将生成标记, 其中 h_0 是编码器向量 o_k 。在每个时间步, 都会生成两个分布: 一个分布在词汇表中的所有单词 (Pvocab) 上, 另一个分布分布在词汇表中的所有单词上。内存内容 (Ptr), 即对话历史记录和 KB 信息。第一个 Pvocab 是通过连接读出的第一跳注意力和当前查询向量生成的。

$$\text{Pvocab}(y_t) = \text{Softmax}(W1[h_t; o_k^1]) \quad (5)$$

其中 $W1$ 是可训练参数。另一方面, Ptr 是使用解码器最后一个 MemNN 跳处的注意力权重生成的: $\text{Ptr} = K_t$ 。我们的解码器通过指向内存中的输入单词来生成标记, 这与指针网络中使用的注意力机制类似 (Vinyals et al., 2015)。

我们以这种方式设计我们的架构是因为我们期望第一跳和最后一跳的注意力权重分别表现出“更宽松”和“更尖锐”的分布。详细来说, 第一跳更多地侧重于检索内存信息, 最后一跳倾向于利用指针监督来选择确切的令牌。因此, 在训练期间, 通过最小化两个标准交叉熵损失的总和来联合学习所有参数: 一个在 $\text{Pvocab}(y_t)$ 和 $y_t \in Y$ 之间 (对于词汇分布), 一个在 $\text{Ptr}(y_t)$ 和 $\text{ptr} \in \text{PTR}$ 之间 (对于词汇分布) 内存分配。

2.2.1 哨兵

如果预期的单词没有出现在内存中, 则训练 Ptr 来生成哨兵标记 $\$$, 如公式 1 所示。一旦选择了哨兵, 我们的模型就会从 Pvocab 生成标记, 否则, 它将采用使用 Ptr 分配的内存内容。基本上, 哨兵令牌用作硬门来控制每个时间步使用哪个分布。Merity 等人, 2017 年使用了类似的方法来控制语言建模任务中的软门。

使用这种方法, 模型不需要像 Gulcehre 等人那样单独学习门函数。 (2016), 并且不受 See 等人中的软门函数的约束。 (2017)。

2.3 内存内容

我们将字级内容 X 存储在内存模块中。与 Bordes 和 Weston (2017) 类似, 我们在 X 的每个标记中添加时间信息和说话人信息以捕获顺序依赖性。例如, “hello t1 \$u”表示用户在时间步骤 1 说出的 “hello”。

另一方面, 为了存储 B , 即 KB 信息, 我们遵循 Miller 等人的工作。 (2016) ; 埃里克等人。 (2017) 使用 (主题、关系、对象) 表示。例如, 我们将威斯汀酒店的信息表示在表 1 中: (威斯汀酒店, 距离, 5 英里)。因此, 我们对主语、关系和宾语的词嵌入求和以获得每个 KB 内存表示。在解码阶段, 对象部分用作 Ptr 的生成词。例如, 当指向 KB 元组 (威斯汀酒店、距离、5 英里) 时, 我们的模型会将 “5 英里” 复制为输出单词。请注意, 只有与特定对话框相关的知识库的特定部分被加载到内存中。

2 这里是 C^{k+1} , 因为我们使用相邻加权捆绑。

任务1	2	3	4	5	DSTC2 车载	
平均。用户平均转动次数	4 6.5	6.4 3.5	12.9	系统转数	6 9.5	6.7
9.9 3.5 18.4	0 24 7	23.7				9.3
平均。知识库结果	0 平均系					39.5
统字	6.3 6.2 7.2 5.7	6.5	系统字	9		10.2
指针比率	.23 .53 .46	.19 .60	9	9	8	9
						29
						.46
词汇训练对话			3747			1229
框 Val 对话框 测			1000			1618
试对话框表 2:			1000			500
3 个不同数据			1000+1000PTA			1117
						1601
						2425
						302
						304

集的数据集统计数据。

3 实验装置

3.1 数据集

我们使用三个公共多轮面向任务的对话日志数据集来评估我们的模型:bAbI dia log (Bordes and Weston, 2017)、DSTC2 (Hender son et al., 2014)和 In-Car Assistant (Eric et al., 2014)。 , 2017) 。这三个的训练/验证/测试集数据集由提供商提前分割。这数据集统计数据如表 2 所示。

bAbI 对话包括餐厅领域的五个端到端对话学习任务,其中是模拟对话数据。任务1至4是关于 API 调用、改进 API 调用、推荐选项以及提供附加信息。任务 5 是任务 1-4 的并集。那里

每个任务有两个测试集:一个遵循与训练集和其他分布相同具有词汇外 (OOV) 实体值训练集中不存在。

我们还使用了从对话状态跟踪挑战 2 (DSTC2) 中提取的对话 Bordes 和 Weston的精致版本 (2017),它忽略对话框状态注释。与 bAbI 对话框的主要区别在于数据集是从真实的人机对话中提取的,自从机器人制造以来,这变得更加嘈杂和困难由于语音识别错误或错误解释而导致的错误。

近日,车载助手数据集已发布。这是一个人与人、多领域的从 Amazon Mechanical Turk 收集的对话数据集。它具有三个不同的域:cal Endar调度、天气信息检索、和兴趣点导航。该数据集有对话轮次较短,但用户和系统行为更加多样化。此外,系统响应也各不相同,KB 信息也各不相同。

情况要复杂得多。因此,该数据集需要更强的与知识库交互的能力,而不是比对话框状态跟踪。

3.2 培训

我们使用 Adam optimizer (Kingma 和 Ba, 2015)端到端地训练我们的模型,并选择学习荷兰国际集团率之间 $[1e^{-3}, 1e^{-4}]$ 。MemNN, 编码器和解码器的跳数 K = 1, 3, 6 以显示性能差异。我们使用简单的贪婪搜索并且没有任何重新评分技术。嵌入大小,也相当于内存大小和 RNN 隐藏大小

(即,包括基线),已被选择之间 [64, 512]。辍学率设置在 [0.1, 0.4],我们还随机将一些输入词屏蔽为未知标记以模拟 OOV

相同辍学率的情况。在所有的数据集,我们通过验证集上的网格搜索来调整超参数,使用作为度量 bAbI 对话框的每次响应准确性车载助手的 DSTC2 和 BLEU 分数。

3.3 评估指标

每个响应/对话的准确性:生成响应只有与以下内容完全相同时才是正确的:黄金反应。仅当对话的每个生成的响应都正确时,对话才是正确的,可以认为是任务完成速度。请注意Bordes 和 Weston (2017)测试他们的模型通过选择系统响应预定义的响应候选者,也就是说,他们的系统解决了多类分类任务。自从 Mem2Seq 单独生成每个标记,使用此指标进行评估更具挑战性

对于我们的模型。BLEU:这是机器翻译系统常用的一种度量 (Papineni et al., 2002) ,但它也被用于评估对话系统 (Eric 和 Manning, 2017; Zhao 等人, 2017) 和聊天机器人 (Ritter 等人, 2011; Li 等人, 2016) 。此外,BLEU 分数是以下方面的相关衡量标准:面向任务的对话,因为生成的答案之间没有很大的差异,这与开放式对话不同域生成 (Liu et al., 2016) 。因此,我们在我们的评估中包括 BLEU 分数 (即使用摩西多 bleu.perl 脚本) 。实体 F1:我们对整个集合进行微平均系统响应并以简单方式比较实体文本。每个黄金系统响应中的实体是由预定义的实体列表选择。这个指标评估生成相关实体的能力从提供的知识库中获取对话的语义 (Eric 和 Manning, 2017 年; Eric 等人, 2017) 。请注意,原来的车载辅助系统

任务	QRN	MemNN	GMemNN	Seq2Seq	Seq2Seq	Seq2Seq	+Attn	Ptr-Unk	Mem2Seq H1	Mem2Seq H3	Mem2Seq H6
T1	99.4 (-)	99.9 (99.6)	100 (100)	100 (100)	100 (100)				100 (100)	100 (100)	100 (100)
T2	99.5 (-)	100 (100)	100 (100)	100 (100)	100 (100)				100 (100)	100 (100)	100 (100)
T3	74.8 (-)	74.9 (2.0)	74.9 (0)	74.8 (0)					74.8 (0)	85.1 (19.0)	87.0 (25.2)
T4	57.2 (-)	59.5 (3.0)	57.2 (0)	57.2 (0)					57.2 (0)	100 (100)	97.6 (91.7)
T5	99.6 (-)	96.1 (49.4)	96.3 (52.5)	98.8 (81.5)					98.4 (87.3)	99.4 (91.5)	96.1 (45.3)
T1-OOV	83.1 (-)		72.3 (0)	82.4 (0)			79.9 (0)		81.7 (0)	92.5 (54.7)	93.4 (60.4)
T2-OOV	78.9 (-)		78.9 (0)	78.9 (0)			78.9 (0)		78.9 (0)	83.2 (0)	81.7 (1.2)
T3-OOV	75.2 (-)		74.4 (0)	75.3 (0)			74.3 (0)		75.3 (0)	82.9 (13.4)	86.6 (26.2)
T4-OOV	56.9 (-)		57.6 (0)	57.0 (0)			57.0 (0)		57.0 (0)	100 (100)	97.3 (90.6)
T5-OOV	67.8 (-)		65.5 (0)	66.7 (0)			67.4 (0)		65.7 (0)	73.6 (0)	67.6 (0)

表 3:bAbI 对话框中每个响应和每个对话 (括号内)的准确性。Mem2Seq实现平均每次响应准确度最高,词汇外性能下降最少。

	耳鼻喉科。	F1 蓝色	每 回复。	每个 拨号。
基于规则的	-	-	33.3	-
QRN-	-	-	43.8	-
记忆神经网络-	-	-	41.1	0.0
GMemNN-	-	-	47.4	1.4
序列到序列	69.7	55.0	46.4	1.5
+收件人	67.1	56.6	46.0	1.4
+Copy	71.6	55.4	47.3	1.3
Mem2Seq H1	72.9	53.7	41.7	0.0
H3 75.3 Mem2Seq H6	72.8	55.3	45.0	0.5
4:DSTC2 评估。		53.6	42.8	0.7

	BLEU 耳鼻喉科F1 学校。	F1 威。	F1 导航F1
人类* 13.5		60.7	64.3
基于规则* 6.6		43.8	61.3
KV检索网* 13.2		48.0	62.9
Seq2Seq 8.4 +Attn		10.3	09.7
9.3		19.9	23.4
Ptr-Unk 8.3		22.7	26.9
Mem2Seq H1 11.6		32.4	39.8
Mem2Seq H3 12.6		33.4	49.3
Mem2Seq H6 9.9		23.6	34.3

表5:对车载助手的评价。报告了人类、基于规则和 KV 检索网络评估 (带 *)

报告 Seq2Seq (+attn 和 +copy)
来自埃里克和曼宁 (2017) 。

来自 (Eric et al., 2017) ,它们不具有直接可比性。
Mem2Seq 获得最高 BLEU 和实体 F1 分数
基线。

Eric 等人报告的 Tant F1 分数。 (2017)使用
规范化形式的实体,它们是
不是根据实际实体价值计算的。自从
数据集不是为插槽跟踪而设计的,我们
报告实体 F1 而不是 (Wen 等人, 2017; Zhao 等人, 2017)
中的时隙跟踪精度。

4 实验结果

我们主要比较Mem2Seq和hop 1,3,6
与几个现有模型:查询减少
网络 (QRN, Seo 等人 (2017)) ,端到端记忆网络
(MemNN, Sukhbaatar
等人。 (2015))和门控端到端内存网络 (GMemNN, Liu 和 Perez
(2017))。我们
还实现了以下基线模型:
标准序列到序列 (Seq2Seq) 模型
有或没有注意 (Luong et al., 2015) ,
和指向未知的指针 (Ptr-Unk, Gulcehre 等人。
(2016)) 。请注意,我们在表3中列出的结果
QRN 的表4和表4与原始论文不同,因为根据他们发布的代码,我
们发现每个响应的准确度为 3

和 Weston (2017)比较性能
基于每个响应和每个对话的准确性。
6 跳的 Mem2Seq 可以实现逐响应
T5 中的准确度为 97.9%,每个对话的准确度为 69.6%,以及
T5-OOV 为 84.5% 和 2.3%,远远超过了现有方法。可以发现,在
T3中,特别是根据排名推荐餐厅的任务,我们的模型可以实现

由于内存指针,结果很有希望。在
就每个响应的准确性而言,这表明
我们的模型可以很好地泛化,对于测试 OOV 数据几
乎没有性能损失,而其他模型则没有
下降约 15-20%。性能增益在
OOV数据也主要归因于使用
复制机制。此外,功效
任务 3-5 中演示了跃点,因为它们需要对知识库信息进行推理的
能力。
请注意,QRN.MemNN 和 GMemNN 已查看
bAbI 对话任务作为分类问题。尽管与我们的生成方法相比,他们
的任务更容易,但 Mem2Seq 模型仍然可以超越性能。最后,人们
可以发现,

Seq2Seq和Ptr-Unk模型也是强有力的基线,这进一步证实了生
成方法也可以在面向任务的对话系统中取得良好的性能 (Eric和
Manning, 2017) 。

未正确计算。
bAbI Dialog:在表3 中,我们遵循Bordes

3我们简单修改了评估部分并报告了
结果。 (https://github.com/uwnlp/qrn)

DSTC2:表4中,Seq2Seq模型
来自Eric 和 Manning (2017)以及基于规则的
Bordes 和 Weston (2017)的报道。
Mem2Seq 实体 F1 得分最高 75.3%
BLEU 得分高达 55.3。这进一步证实了 Mem2Seq 在检索方面可以表现良好
正确的实体,使用多跳机制而不丢失语言建模。在这里,我们

不使用匹配类型 (Bordes
和 Weston, 2017)或实体类型 (Eric 和 Man ning, 2017)特征,因为此元信息
并不常见,我们希望拥有
对普通输入输出对的评估。一
还可以发现,Mem2Seq 的每个响应精度 (即 2% 裕度)与其他类似
现有的解决方案。请注意,每个模型的每次响应准确率均低于
50%,因为
数据集非常嘈杂,很难生成
与黄金一模一样的反应。
车载助手:在表5 中,我们的模型可以
获得最高 12.6 BLEU 分数。此外,
Mem2Seq 在以下方面显示了有希望的结果
实体 F1 分数 (33.4%),一般来说,
远高于其他基线。笔记
埃里克等人报告的数字。(2017)
正如我们提到的,与我们的没有直接可比性
以下。其他基线 (例如 Seq2Seq 或 Ptr Unk)在这方面的性能尤其差
数据集,因为 RNN 方法编码较长的 KB 信息效率非常低,这就是

Mem2Seq 的优点。
此外,我们观察到一个有趣的现象,人类可以轻松达到很高的水平
BLEU 分数较低的实体 F1 分数。这意味着对实体的推理能力更强

(hops)很重要,但结果可能不相似
到黄金答案。我们相信,即使 BLEU 分数较低,人类也能给出好的答案,
因为可以有不同的方式来表达
相同的概念。因此,Mem2Seq 显示
成功选择正确实体的潜力。

注意KV Retrieval Net base的结果
表5中报告的行来自In-Car Assistant 的原始论文
(Eric 等人, 2017 年),其中
他们通过映射表达式简化了任务
使用命名实体将实体转换为规范形式
识别 (NER)和链接。因此,评估不能直接与我们的系统进行比较。

例如,他们的模型学会了生成诸如“你脚下有一场足球比赛

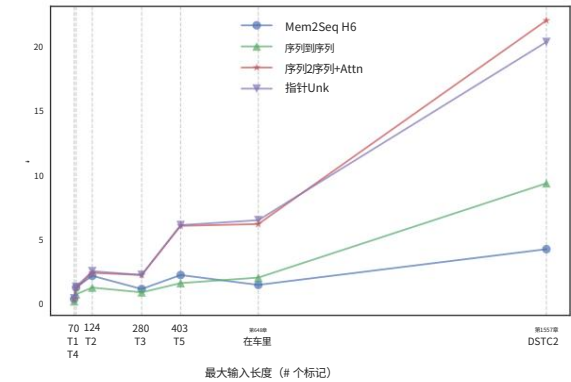


图 2:不同时期每个时期的训练时间
任务 (越低越好)。速度差异随着最大输入长度的增加而变大
折痕。

与足球派对一起玩球时间”,而不是生成诸如“你
有一场足球比赛”之类的句子
晚上 7 点和约翰一起。”因为可能有超过
一次足球派对或足球时间,他们的模式
不学习如何访问知识库,而是学习
学习规范化的语言模型。

Time Per-Epoch:我们还比较了图2中的训练时间。实验设置
4

批量大小为 16,我们报告每个模型
与超参数可以实现
最高性能。人们可以观察到
对于较短的输入长度 (bAbI 对话任务 1-4)和差距,训练时间并
没有那么不同
随着最大输入长度的增加而变大。与 Seq2Seq 相比,Mem2Seq
在车载助手和 DSTC2 中的速度大约快 5 倍

与注意。这种训练效率的差异主要归因于 Seq2Seq

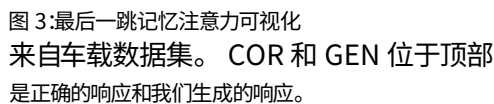
模型具有输入顺序依赖性
限制任何并行化。此外,Seq2Seq 模型不可避免地需要对
KB 进行编码,而不是
Mem2Seq 仅使用对话历史记录进行编码。

5 分析与讨论

记忆注意力:分析注意力
权重经常被用来表示
内存读出,因为它是理解模型动态的直观方式。图8显示了

每个生成的最后一跳的注意向量
令牌。每列代表Pptr向量
相应的生成步骤。我们的模型有
内存上的急剧分布,我

⁴ Intel(R) Core(TM) i7-3930K CPU@3.20GHz,使用
GeForce GTX 1080 Ti



触发了哨兵,并且“is”是由生成的词汇表中不存在该词对话历史记录。

排名从最高到最低。用户可以拒绝推荐和系统必须考虑下一个最高的餐厅。我们发现

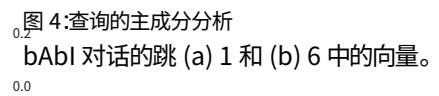


表 6:生成的响应示例
调度域上的车载助手。

查询向量:在图4 中,Mem2Seq 查询向量的主要成分分析是

我们可以发现,在 (a) 第一跳中,两种不同颜色没有明显的分离,但每种颜色都

倾向于聚集在一起。另一方面，(b) 第 6 跳中的分离变得更加清晰，因为每个颜色聚类为几组，例如位置、美食和数量。我们的模型倾向于检索第一跳中的更多信息，并指向最后一跳的回忆。

示例:表1和表6显示了来自 In-Car Assistant 数据集的两个测试集样本中不同模型生成的响应。我们报道

该数据集中的示例,因为他们答案是更加人性化,不像其他人那样结构化和重复。Seq2Seq一般无法产生相关信息,有时甚至会失败

语言建模。相反,利用注意力会有所帮助存在这个问题,但它仍然很少产生正确的实体。例如,表1中带有注意力的 Seq2Seq 生成了 5 英里,但正确的结果是

一英里是 4 英里。另外,Ptr-Unk常常不能从输入中复制正确的标记,如下所示表 1 中的“PAD” 。另一方面,Mem2Seq 能够在这两个方面产生正确的反应例子。特别是在导航领域,如表 1 所示,Mem2Seq 产生不同的但仍是正确的说法。我们在补充材料中报告了所有领域的进一步示例材料。

讨论:传统的面向任务的对话系统 (Williams 和 Young, 2007) ,它们是仍然广泛应用于商业系统中,需要系统设计中大量的人类努力和数据收集。另一方面,虽然端到端对话系统尚不完美,它们需要更少的人为干扰,特别是在数据集构建中,作为原始会话文本和KB信息可以直接使用,无需大量预处理 (例如NER,

依存分析) 。就这一点而言,Mem2Seq 是一个简单的生成模型,能够在企业知识库信息中具有良好的泛化能力。我们还发现实体 F1 分数可能是比每个响应准确度或 BLEU 更全面的评估指标

分数,因为人类通常可以选择正确的实体,但反应非常多样化。事实上,我们想强调的是,人类可能有

尽管它们是正确的,但 BLEU 分数较低,因为之间可能没有大的 n 元语法重叠给定的响应和预期的响应。然而,这并不意味着 BLEU 分数与人类评估之间不存在相关性。

事实上,与聊天机器人和开放域对话框不同其中 BLEU 分数与人类评估不相关 (Liu et al., 2016) ,在任务导向中对话的答案仅限于特定的实体和重复模式。因此,我们相信 BLEU 分数仍然可以被视为相关的措施。在未来的工作中,可以采用多种方法

被应用 (例如强化学习 (Ranzato 等人, 2016) ,波束搜索 (Wiseman 和 Rush, 2016))提高回复的相关性和实体F1分数。然而,我们更愿意保留我们的模型尽可能简单以便展示即使没有高级培训它也能很好地工作方法。

6 相关作品

端到端的面向任务的对话系统训练直接在对话文本转录上建立单一模型 (Wen et al., 2017; Serban et al., 2016; 威廉姆斯等人, 2017;赵等人, 2017;徐等人, 2017年; Serban 等人, 2017) 。在这里,RNN 发挥着由于它们能够创建潜在的表示,从而避免人工的需要,因此发挥着重要作用状态标签。端到端内存网络 (Bordes 和 Weston, 2017; Sukhbaatar 等人, 2015) , 及其变体 (Liu 和 Perez, 2017; Wu 等人, 2017年、2018年)在这方面也取得了不错的成绩任务。在每个架构中,输出是通过生成一系列标记而产生,或者通过选择一组预定义的话语。

序列到序列 (Seq2Seq) 模型具有也被用于面向任务的对话系统 (Zhao et al., 2017) 。这些架构有更好的语言建模能力,但他们没有在知识库检索方面表现良好。即使有复杂的注意力模型 (Luong et al., 2015; Bahdanau et al., 2015) , Seq2Seq 无法将正确的实体映射到生成的输入。为了缓解这个问题,复制增强的 Seq2Seq 模型Eric 和

使用曼宁 (2017) 。这些模型通过复制来执行话语选择方法

相关信息直接来自知识库。复制该机制也被用于有问题的转向任务 (Dehghani 等人, 2017; He 等人, 2017) ,神经机器翻译 (Gulcehre 等人, 2016年; Gu 等人, 2016) ,语言建模 (Merity 等人, 2017)和总结 (参见等人, 2017) 。

与对话系统关系不大,但与我们的工作,是基于内存的解码器和非循环生成模型:1)Mem2Seq 查询

用于访问我们记忆的生成阶段可以被视为Memory中使用的内存控制器增强神经网络 (MANN) (格雷夫斯等, 2014,2016) 。同样,记忆编码器已用于神经机器翻译 (Wang 等人, 2016)和元学习应用 (Kaiser 等人, 2017) 。然而,Mem2Seq

与这些模型的不同之处在于:它使用多

跳跃注意力与复制机制相结合,而其他模型使用单个矩阵

表示。2)仅依赖自注意力机制的非循环生成模型 (Vaswani et al., 2017)与多跳相关

MemNN 中使用的注意力机制。

7 结论

在这项工作中,我们提出了一个端到端的可训练的用于面向任务的对话系统的内存到序列模型。Mem2Seq 结合了多跳端到端记忆网络中的注意力机制与指针网络的思想一起工作以合并外部信息。我们凭经验表明

我们的模型能够使用外部知识库信息和预定义词汇产生相关答案,并可视化多跳注意力机制如何帮助学习记忆之间的相关性。Mem2Seq 快速、通用,并且能够在以下方面取得最先进的结果

三个不同的数据集。

致谢

这项工作部分由 ITS/319/16FP 资助
香港科技大学创新科技委员会
香港研究资助局 16214415 及 16248016,以及香港研究资助局 RDC 1718050-0 EMOS.AI。

参考

德兹米特里·巴达瑙,Kyunghyun Cho 和 Yoshua Ben gio。2015. 联合神经机器翻译学习对齐和翻译。国际学习表征会议。

安托万·博德斯和杰森·韦斯顿。2017.学习端到端的目标导向对话。学习代表大会,内部的绝对/1605.07683。

Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, 和约书亚·本吉奥。2014.实证评价序列建模上的门控循环神经网络。NIPS 深度学习和表示学习工作坊。

穆斯塔法·德加尼、萨沙·罗特、恩里克·阿方塞卡、和帕斯卡·弗勒里。2017.学习出席、抄写、并生成基于会话的查询建议。在2017 年 ACM 信息和知识管理会议论文集,CIKM 17,第 1747–1756 页,美国纽约州纽约。ACM。

米哈伊尔·埃里克、拉克希米·克里希南、弗朗索瓦·夏莱特和克里斯托弗·D·曼宁。2017.键值检索

以任务为导向的对话网络。诉讼中第 18 届 SIGdial 话语年会和对话,第 37-49 页。计算语言学协会。

米哈伊尔·埃里克和克里斯托弗·曼宁。2017.复制增强的序列到序列架构给出了在面向任务的对话中表现良好。在第十五届欧洲联盟会议论文集计算语言学协会分会:第 2 卷,短论文,第 468-473 页,西班牙巴伦西亚。计算语言学协会。

亚历克斯·格雷夫斯、格雷格·韦恩和伊沃·丹尼尔卡。2014年。神经图灵机。联合RR。
亚历克斯·格雷夫斯、格雷格·韦恩、马尔科姆·雷诺兹、蒂姆·哈雷、伊沃·达尼赫卡、阿格涅斯卡·格拉布斯卡·巴文斯卡、塞尔吉奥·戈麦斯·科尔梅纳雷霍、爱德华·格雷芬斯泰特、蒂亚戈·拉马略、约翰·阿加皮乌等。2016.使用神经网络与动态外部存储器的混合计算。自然,538 (7626) :471–476。

Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O.K. 李。2016.将复制机制纳入序列到序列的学习。在诉讼程序中计算语言学协会第 54 届年会 (第一卷:长论文),第 1631-1640 页,德国柏林。协会计算语言学。

卡格拉·古尔切雷、安圣金、拉梅什·纳拉帕蒂、Bowen Zhou, and Yoshua Bengio. 2016. Pointing 那些不知名的词。在第 54 届会议记录中计算语言学协会年会 (第一卷:长论文),页数140–149,德国柏林。计算语言学协会。

Shizhu He, Cao Liu, Kang Liu, and Jun Zhao. 2017.通过合并生成自然答案序列到序列学习中的复制和检索机制。计算协会第 55 届年会记录语言学 (第一卷:长论文),第 199–208,加拿大温哥华。计算语言学协会。

马修·亨德森、布莱斯·汤姆森和杰森·D 威廉姆斯。2014.第二次对话框状态跟踪挑战。讨论和特别利益小组第 15 届年会记录对话 (SIGDIAL),第 263-272 页。

堀千织、大竹清教、美须辉久、英树 樫冈和中村聪。2009.统计对话管理应用于基于wfst的对话系统。在IEEE国际会议上声学、语音和信号处理,2009 年。ICASSP 2009,第 4793–4796 页。IEEE。

Lukasz Kaiser、Ofir Nachum、Aurko Roy 和 Samy 本吉奥。2017.学习记住罕见事件。

国际学习表征会议。

迪德里克·P·金马 (Diederik P Kingma) 和吉米·巴 (Jimmy Ba)。 2015. 亚当:A 随机优化方法。国际的
学习代表会议。

Cheongjae Lee,Sangung Jung,Seokhwan Kim 和
加里·根培·李. 2009. 基于示例的对话框
实际多域对话系统的建模。
言语沟通,51(5):466–484。

埃丝特·莱文·罗伯托·皮耶拉奇尼和维兰德·埃克特。
2000.用于学习对话策略的人机交互随机模型。 IEEE 语音和音频处
理交易,8(1):11–23。

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao,
和比尔·多兰。 2016.[神经对话模型的多样性促进目标函数](#)。在

2016年北方会议论文集
计算语言学协会美国分会:人类语言技术,

第 110-119 页,加利福尼亚州圣地亚哥。协会
用于计算语言学。

刘家伟,瑞安·洛·尤利安·塞尔班·迈克·诺斯·沃斯·洛朗·
查林和乔尔·皮诺。 2016年。
[如何不评估你的对话系统:无监督评估指标的实证研究](#)

[对话响应生成](#)。在诉讼程序中
2016 年自然语言处理经验方法会议,第 2122-2132 页,奥斯汀,

德克萨斯州。计算语言学协会。

刘飞和朱利安·佩雷斯。 2017.[门控端到端](#)
[记忆网络](#)。协会欧洲分会第 15 届会议记录

计算语言学:第 1 卷,长文,第 1-10 页,西班牙巴伦西亚。协会

计算语言学。

Thang Luong,Hieu Pham 和 Christopher D. Manning。
2015.[基于注意力的有效方法](#)
[神经机器翻译](#)。在诉讼程序中
2015 年自然语言处理经验方法会议,第 1412-1421 页,葡萄牙里斯
本。计算语言学协会。

斯蒂芬·梅里蒂·蔡明熊、詹姆斯·布拉德伯里和
理查德·索切尔。 2017.[指针哨兵混合物](#)
楷模。国际学习代表会议。

亚历山大·米勒、亚当·费什、杰西·道奇、阿米尔·侯赛因·卡
里米、安托万·博德斯和杰森·威斯顿。 2016.[键值存](#)
[储网络直接用于](#)
[阅读文档](#)。 2016 年会议论文集
自然语言处理经验方法会议,第 1400-1409 页,德克萨斯州奥斯汀。

计算语言学协会。

Kishore Papineni,Salim Roukos,Todd Ward 和 Wei Jing
Zhu。 2002. [Bleu:一种机器翻译自动评估方法](#)。在诉讼程序
中

计算语言学协会第 40 届年会,第 311-318 页,费城,

美国宾夕法尼亚州。计算协会
语言学。

马克奥雷利奥·兰扎托、苏米特·乔普拉、迈克尔·奥利、
和沃伊切赫·扎伦巴。 2016.使用循环神经网络进行序列级训练。国
际的
学习代表会议。

艾伦·里特、科林·切里和威廉·B·多兰。 2011年。
[社交媒体中数据驱动的响应生成](#)。在
2011年实证会议论文集
自然语言处理方法,页面
583–593,英国苏格兰爱丁堡。协会
计算语言学。

阿比盖尔·西 (Abigail See)、彼得·J·刘 (Peter J. Liu) 和克里斯托弗·D·曼宁 (Christopher D. Manning)。
2017.[开门见山:用指针生成器网络进行总结](#)。计算协会第 55 届年会
记录

语言学 (第一卷:长论文),第 1073 页–
1083,加拿大温哥华。计算语言学协会。

Minjoon Seo,Sewon Min,Ali Farhadi 和 Hannaneh
哈吉希尔兹。 2017. 查询缩减网络
问题回答。国际会议
学习表示。

Iulian Vlad Serban,Alessandro Sordoni,Yoshua Bengio、
Aaron C Courville 和 Joelle Pineau。 2016年。
使用生成分层神经网络模型构建端到端对话系统。在AAAI中,

第 3776-3784 页。

尤利安·弗拉德·塞尔班、亚历山德罗·索尔多尼、瑞安·洛、
洛朗·查林、乔尔·皮诺、亚伦·C·库维尔、
和约书亚·本吉奥。 2017.分层潜伏
用于生成对话的可变编码器-解码器模型。 AAAI,第 3295-3301 页。

桑巴亚尔·苏赫巴托、杰森·韦斯顿、罗布·弗格斯等。
2015.[端到端内存网络](#)。进展中
在神经信息处理系统中,页面
2440–2448。

阿希什·瓦斯瓦尼、诺姆·沙泽尔、尼基·帕尔玛、雅各布
Uszkoreit,Llion Jones,Aidan N Gomez、 ukasz
凯撒和伊利亚·波洛苏欣。 2017.注意力就是一切
你需要。《神经信息处理系统进展》,第 6000-6010 页。

Oriol Vinyals,Meire Fortunato 和 Navdeep Jaitly。
2015.[指针网络](#)。位于北达科他州科尔特斯
劳伦斯,DD Lee,M. Sugiyama 和 R. Garnett,
编辑,神经信息处理进展
系统 28,第 2692–2700 页。柯兰联营公司,
公司

Mingxuan Wang, Zhengdong Lu, Hang Li, and Qun
刘。 2016.[神经网络的记忆增强解码器](#)
[机器翻译](#)。 2016 年会议论文集
自然语言处理经验方法会议,第 278-286 页,德克萨斯州奥斯汀。

计算语言学协会。

文宗宪、米利卡·加西奇、尼古拉·米尔克西奇、
莉娜·玛丽亚·罗哈斯-巴拉奥纳、苏佩豪、斯特凡
乌尔特斯、大卫·范戴克和史蒂夫·J·杨。 2017年。
基于网络的端到端可训练任务导向
对话系统。在EACL。

杰森·D·威廉姆斯、卡沃什·阿萨迪和杰弗里·茨威格。
2017。[混合代码网络:实用且高效](#)
[具有监督和强化学习的端到端对话控制](#)。计算协会第 55 届年会记录

语言学（第一卷:长论文）,第 665 页–
677 ,加拿大温哥华。计算语言学协会。

杰森·D·威廉姆斯和史蒂夫·杨。 2007. 部分
口语的可观察马尔可夫决策过程
对话系统。计算机语音和语言,
21 (2) :393-422。

萨姆·怀斯曼和亚历山大·M·拉什。 2016年。
[序列到序列学习作为波束搜索优化](#)。 2016 年会议论文集

自然语言处理中的经验方法,第 1296-1306 页,德克萨斯州奥斯汀。
协会
用于计算语言学。

吴建生、Andrea Madotto、Genta Winata 和
冯帕斯卡。 2017.用于实体价值独立的目标导向对话学习的端到端循
环实体网络。在对话系统技术挑战研讨会中,DSTC6。

吴建生、Andrea Madotto、Genta Winata 和
冯帕斯卡。 2018. 端到端动态查询
用于实体值独立的面向任务的对话的记忆网络。 IEEE 国际声学、语
音和信号处理会议

（国际社会保障计划）。

Steve Young、Milica Gasic、Blaise Thomson 和 Jason D Williams。
2013. 基于 Pomdp 的统计口语对话系统:综述。诉讼程序

IEEE,101(5):1160–1179。

赵天成、Allen Lu、Kyungho Lee 和 Max 在埃斯肯纳齐。 2017。[生成编
码器-解码器](#)
[面向任务的口语对话系统模型](#)
[聊天能力](#)。在第 18 届 SIGdial 话语与对话年度会议记录中，
第 27-36 页。计算语言学协会。

8桌

8.1 每个 Epoch 的时间

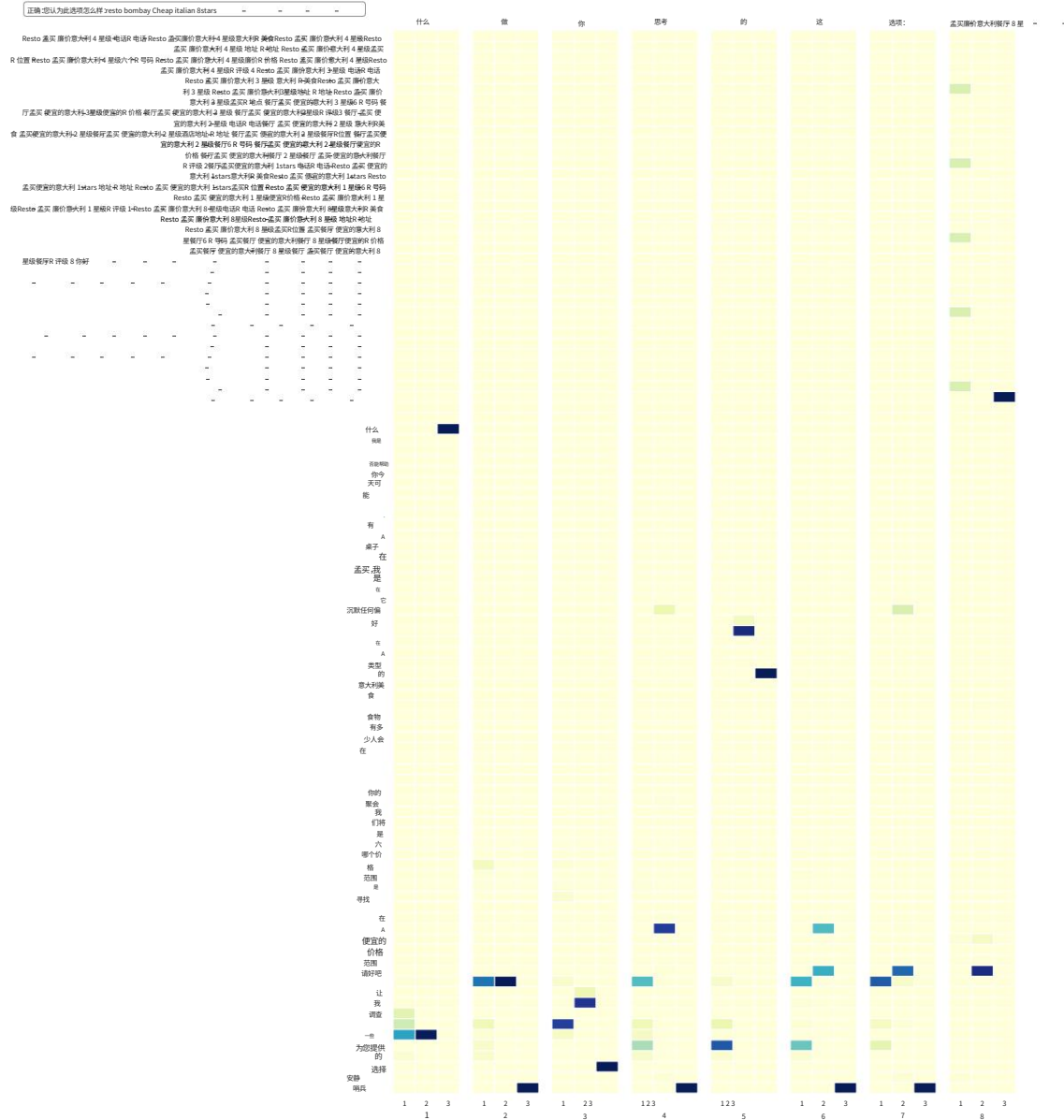
	T1	T2	T3	T4	T5	DSTC2	车内			
序列到序列	0.7	1.22	0.85	0.15	1.57				9.36	2.00
+注意	1.18	2.38	2.2	0.32	6.04	22.07				6.18
Ptr-Unk	1.28	2.5	2.21	0.38	6.10	20.39				6.48
Mem2Seq H1	0.73	1.10	0.55	0.15	1.05	Mem2Seq H3	0.97		1.49	0.68
1.48	0.77	0.21	1.33	Mem2Seq H6	1.23	2.14	1.11	0.33	2.20	1.02
									4.22	1.43

表 7:每个周期的分钟数。 Mem2Seq 比其他方法更快,尤其是对于较长的输入。

8.2 超参数

	T1	T2	T3	T4	T5	DSTC2 车载		LR
序列到序列	256 (0.1)	256 (0.1)	128(0.1)	256 (0.1)	256 (0.1)	-	512 (0.3)	0.0001
+注意	256 (0.1)	256 (0.1)	256 (0.1)	256 (0.1)	128 (0.1)	-	512 (0.0)	
Ptr-Unk	256 (0.1)	256 (0.1)	256 (0.1)	256 (0.1)	128 (0.1)	-	512 (0.0)	
Mem2Seq GRU L1	256 (0.3)	256 (0.2)	128 (0.2)	256 (0.2)	256 (0.1)	256 (0.1)	256 (0.2)	0.001
Mem2Seq GRU L3	256 (0.1)	256 (0.3)	128 (0.2)	256 (0.3)	128 (0.1)	128 (0.2)	256 (0.1)	
Mem2Seq GRU L6	256 (0.2)	128 (0.3)	256 (0.1)	128 (0.3)	256 (0.2)	256 (0.1)	256 (0.2)	

表 8:每个数据集中针对不同跃点选择的超参数。值是嵌入维度和 GRU 隐藏大小,括号之间的值是 dropout 率。对于所有的在模型中,我们使用的学习率为 0.001,衰减率为 0.5。



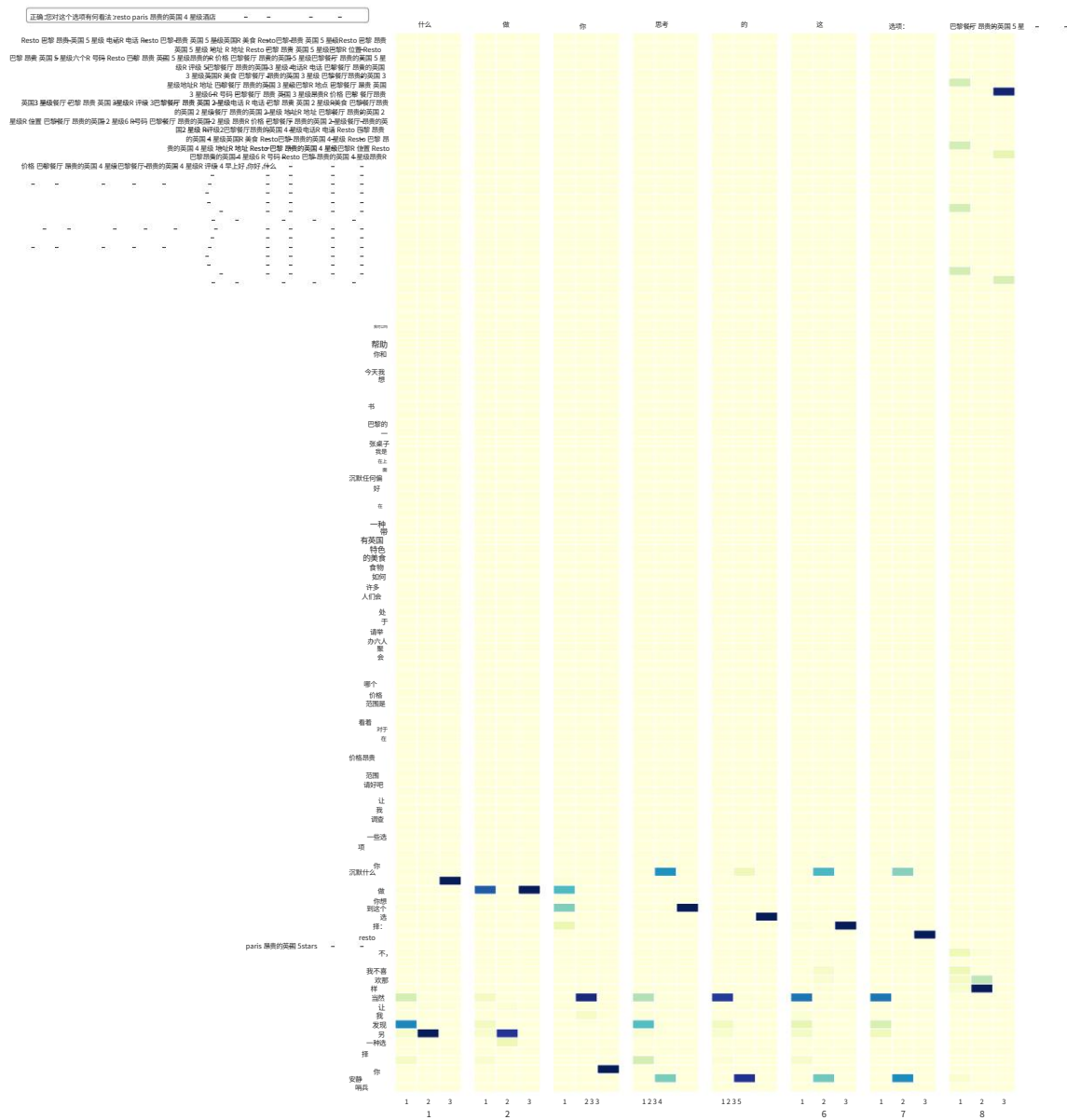


图 6:bAbI 对话任务 3 的 3 跳记忆注意力可视化。如果最后一跳的注意力不敏锐,模型往往会出错。

9.2 最后一跳注意力

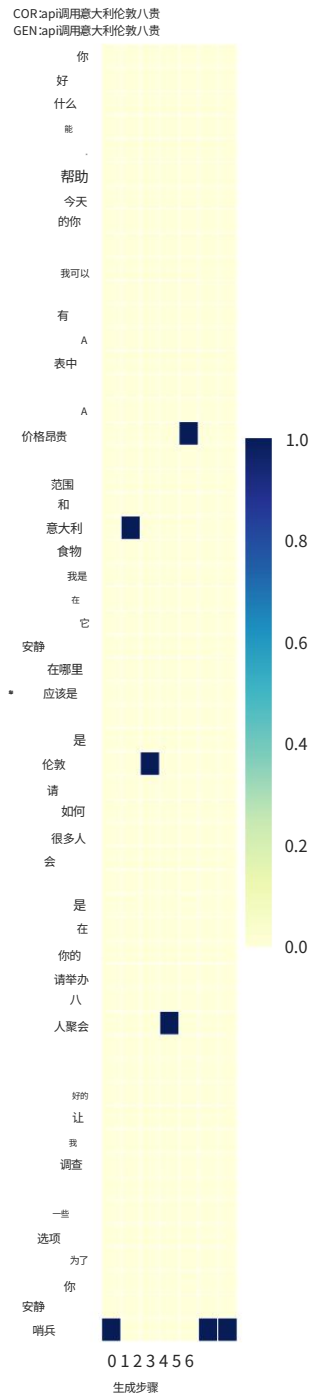


图 7:bAbI 数据集的最后一跳记忆注意力可视化。顶部的 COR 和 GEN 是正确的响应,也是我们生成的响应。

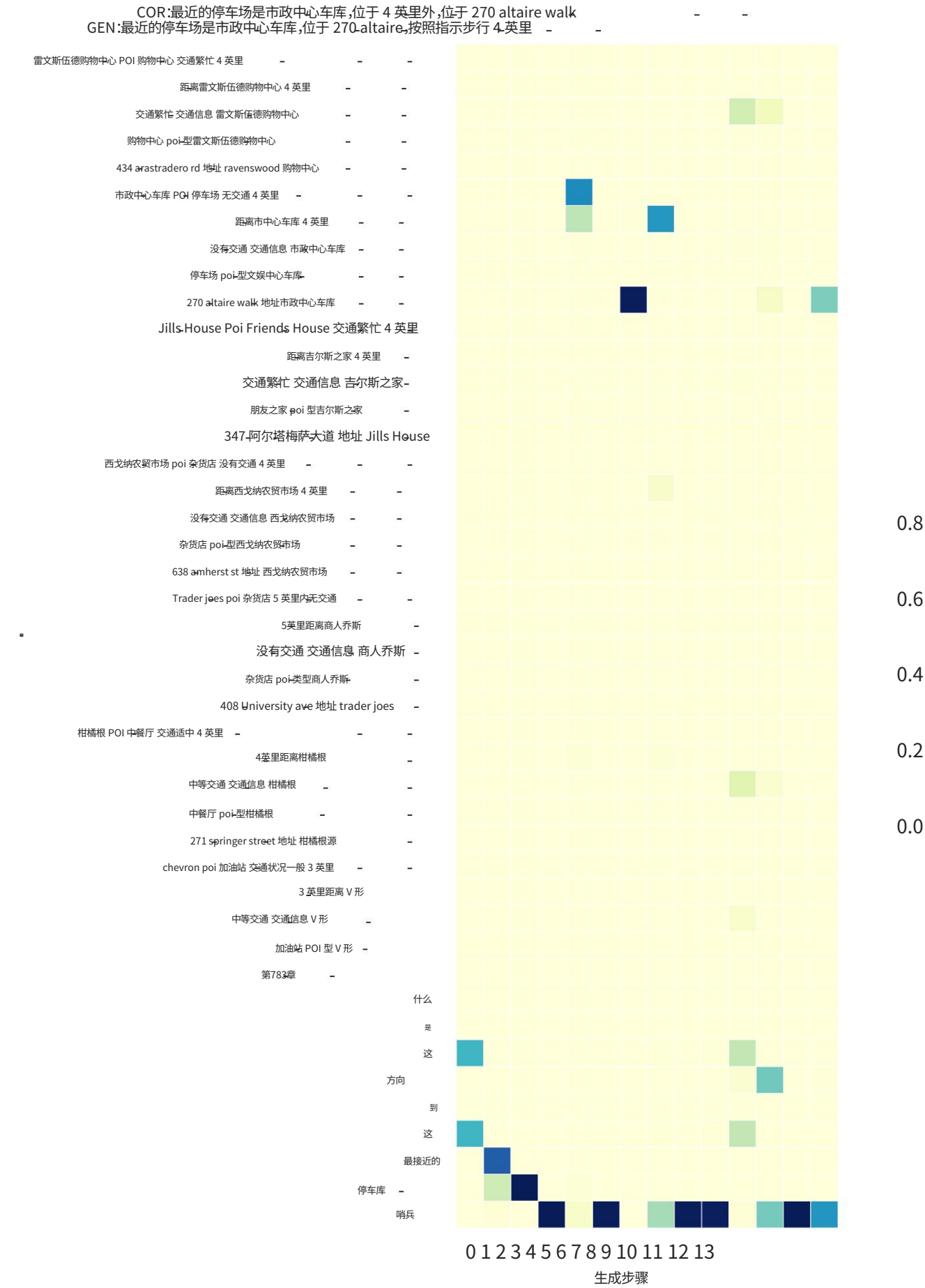


图 8:来自车载数据集的最后一跳记忆注意力可视化。顶部的 COR 和 GEN 是正确的响应,也是我们生成的响应。

10 个 InCar Assistant 数据集示例

表 9 :车载助手在天气领域生成的响应示例。

地点大急	星期一炎	周二有雨，	周三
流域 纽约 波士	热,最低 50F,最高 70F 有雾,最低	最低 60 华氏度,最高 80 华氏度雪,最低	雨,最低气温 20 华氏度,最高气温 30 华氏度
顿 达勒姆	30F,最高 50F 冰雹,最低 90F,最高	70 华氏度,最高 80 华氏度阴,最低 60	多云,最低气温 20 华氏度,最高气温 30 华氏度
	100F 炎热,最低 90F,最高 100F 有雨，	华氏度,最高 70 华氏度干燥,最低 60 华氏	雨,最低气温 50 华氏度,最高气温 60 华氏度
	最低 60F,最高 70F	度,最高 80 华氏度多云,最低 30 华氏	有雾,最低 60F,最高 80F
旧金山		度,最高 40 华氏度有雨,最低70F,最高	阴,最低气温 90 华氏度,最高气温 100 华氏度
卡森	80F 潮湿,最低 90F,最高 100F 圣何塞暴风雪,	最低 40F,最高 50F 雪,最低 90F,最高 100F	霜冻,最低 40 华氏度,最高 60 华氏度
周四周五周六晴空,最低 60F,最高 70F 温暖,低70F,最高 90F 有雾,最低 50F	最高 60F 有雨,最低 80F,最高 100F 有雨,最		阴,最低 60F,最高 80F
低 40F,最高 60F 多云,最低 30F,最高 50F 露水,最低 20F,最高30F 多云,最低	90F,最高 100F 阴,最低 50F,		星期日
最高 70F 阴,最低 80F,最高 90F			阴,最低 50F,最高 60F
			雪,最低气温 20 华氏度,最高气温 40 华氏度
有雾,最低 90 华氏度,最高 100 华氏度,冰雹,最低 30 华氏度,最高 70F,最高 90F,	最低 60 华氏度,最高 80 华氏度		
有暴风雨,最低 90 华氏度,最高 100 华氏度 阴天,最低 90 华氏度,最高 100 华氏度 晴空,最低 90 华氏度,最高 100 华氏度 有雾,最低 30 华氏度,最高 50 华氏度			
干燥,最低气温 50 华氏度,最高气温 60 华氏度	有雾,最低气温 80 华氏度,最高气温 100 华氏度 有雨，	阴,最低 50F,最高 70F 阴,最低 60F,最高 70F	
雨,最低气温 30 华氏度,最高气温 50 华氏度	最低气温 80 华氏度,最高气温 100 华氏度	干燥,最低气温 50	干燥,低温 90F,高温 100F

华氏度,最高气温 60 华氏度驾驶员这周末达勒姆会下雪吗？

Seq2seq 预计本周不会下雪
+注意达勒姆的天气预报中没有雪

Ptr-Unk 是的,达勒姆本周没有雪

Mem2Seq 没有雪,天气预报没有预测达勒姆有雪

GOLD达勒姆本周预计不会下雪

表 10:天气域车载数据集生成的答案示例。

周一周二周五有风,最低 40 华氏度,最高 50 华氏度雪,最低 40 华氏度,最高 50 华氏度霜,最低 30 华氏度,最高	周三
40 华氏度有雨,最低 30 华氏度,最高 40 华氏度有雾,最低 60 华氏度,最高 70 华氏度有暴风雨,最低 50F,最高 60F 多云,最低 20F,最高 40F	冰雹,最低 70 华氏度,最高 80 华氏度
晴空,最低 70F,最高 90F 晴空,最低 60F,最高 70F 温暖,最低 90F,最高 100F 有风,最低 50F ,最高 60F 暴风雪,最低 20F,最高 30F 雪,最低	雪,最低温度 80 华氏度,最高温度 90 华氏度
60F,最高 80F 晴空,最低 30F,最高 50F 晴空,最低 30F,最高 40F 有风,最低 50F,最高60F 有毛毛雨,最低 80F,最高 100F 晴空,最低 60F,最高	干燥,低温 40F,高温 60F
70F 有雨,最低 90F,最高 100F 周四周日下雨,最低 80F,最高 90F 有雾,最低 40F,最高 50F 有毛毛雨,最低 50F,最高 60F 晴空,最低 90F,最高	雨,最低气温 60 华氏度,最高气温 70 华氏度
100F 暴风雨,最低 40F,最高 50F 冰雹,最低 40F,最高 60F 阴天,最低 80F,最高 90F 暴风雪,最低 20F ,最高 40 华氏度有雨,最低 50 华氏度,最	干燥,低温 30F,高温 50F
高 60 华氏度有风,最低 30 华氏度,最高 50 华氏度霜,最低 40 华氏度,最高 50 华氏度毛毛	有风,最低气温 80 华氏度,最高气温 100 华氏度晴空,最低气温 50 华氏度,最高气温 70 华氏度
雨,最低 40 华氏度,最高 60 华氏度干燥,最低 20 华氏度,最高 30 华氏度多云,低20F,高40F	有雾,最低气温 20 华氏度,最高气温 40 华氏度,冰雹,最低气温 20 华氏度,最高气温 30 华氏度
	周六地点
	旧金山下雪,最低气温 40 华氏度,最高气温 50 华氏度
	曼哈顿干燥,最低气温 70 华氏度,最高气温 90 华氏度
	西雅图有风,最低气温 60 华氏度,最高气温 70 华氏度
	奥克兰
	埃克塞特
	门洛帕克
	卡森
	有风,最低气温 80 华氏度,最高气温 100 华氏度

司机今天会下雪吗？

CAR你想知道哪个城市今天会下雪吗？

司机请检查奥克兰

黄金今天奥克兰天气会很暖和

Seq2seq 今天没有下雪
+att 不,奥克兰今天会下雪

Ptr-Unk 今天将是 PAD

Mem2Seq 奥克兰今天不会下雪

表 11:在导航域上为车载助手生成的响应示例。

然后	距离	交通信息	POI型医院	地址
斯坦福快速护理中心 3 英里 回家 4 英里		没有交通 交	之家	214 皇家路
		通繁忙 交通繁		巴林杰街5671号
全食超市	2英里	忙 杂货店 6 英里 交通中等 中餐馆 669 El		苹果街819号
PF 张晶晶披萨	Camino Real			
My Heart	5英里	交通拥堵 6 英	中国餐厅	113树博士
Cafe Venetia	里 交通适中	披萨店 4 英里 交通堵塞	咖啡或茶馆	528安东康涅狄格州
				269阿尔杰博士
司机给我找一家茶馆				
Seq2seq 我们距离 2 英里 ,附近有一个路障				
+att 我们距 Cafe Venetia 4 英里 ,位于 269 Alger Drive				
PTRUNK 我们距 Whole Foods 6 英里 ,您喜欢屏幕上的地址				
Mem2Seq 我们距离 Cafe Venetia 4 英里 ,没有交通				
黄金4 英里外有一家 Cafe Venetia ,位于 269 Alger dr				

表 12:在导航域上为车载助手生成的响应示例。

距离交通信息4 英里 交通繁忙 1	poi 类型休	地址	然后
英里 交通中等 2 英里	息站 医院	第578章 树博士	基恩酒店
	杂货店 1	第899章	斯坦福儿童健康中心
没有交通	英里交通流量适	第452章	西夫韦
中的咖啡或茶馆 1 英里交通繁忙的杂货店 6 英里交通流量适中		第436章	帕洛阿尔托咖啡馆
的咖啡或茶馆 394 Van Ness Ave 6 英里交通繁忙的咖啡或		苹果街819号	全食超市
茶馆 145 Amherst St 司机 带我去一个地方我在哪里可以喝点热茶			切
			茶瓦那
Seq2seq 有一个远			
att 芝加哥有一家披萨店和 Teavana			
PTRUNK 我有一个 PAD 不在身边			
Mem2Seq 我们距离 Palo Alto Cafe 6 英里 ,距离 Coupa 6 英里 ,您更喜欢哪一个?			
GOLD Coupa 就在附近 ,供应茶			

表 13:车载助手在导航域上生成的响应示例。

距离 poi 类型 地址 交通信息	593 Arrowhead Way Chef Chu s		然后
3英里	没有交通		
1英里	没有交通	中餐厅 669 El Camino Real PF Changs	
附近 5 英里	路障 3 英里无交通	加油站 披萨	阿莱斯特大街200号 瓦莱罗
		店 休息站	528安东康涅狄格州 披萨我的心
4英里	交通繁忙 交通		753 University Ave 中餐 康福特茵酒店
5 英里 2	繁忙 交通繁忙	厅 842 Arrowhead Way Panda Express	
英里 6 英	没有交通	披萨餐厅 704 El Camino Real 朋友之家	必胜客
里			分段阅读_第 864 章 杰克之家

司机帮我找到最近的加油站。
Seq2seq 最近的停车场是从远处
+att 一个人不在了
PTRUNK PAD 最近的加油站是 是 是
Mem2Seq 最近的加油站是你的加油站
GOLD最近的加油站距离酒店 5 英里。需要更多信息吗？

表 14:车载助手在导航域上生成的响应示例。

然后	距离 交通信息 Poi 类型4 英里 无交通	休息站 2 英里 交通	地址
威斯汀酒店	中等 停车场 4 英里 交通中等 中餐厅	271 Springer	329 皇家路
帕洛阿尔托车库 R	Street		阿玛兰塔大街481号
柑橘根源			
西戈纳农贸市场	4 英里 638 阿默斯特街 交通 交	杂货店 2 英里	
全食超市	819 Alma St 通繁忙	杂货店 773	
斯坦福购物中心 2 英里 交通适中 购物中心 toms house		Alger Dr	
	1英里	没有交通	朋友的家 范尼斯大道580号

司机告诉我去最近的杂货店的路线	
车	2 英里外有 Whole Foods 超市， 西戈纳农贸市场距离酒店 4 英里。我们去哪里？
司机我需要一辆可以避开所有交通拥堵的司机	
Seq2seq 他是 是 是 不在	
+Att Chef Chu s 位于 593 Arrowhead way	
Ptr-Unk PAD 是 PAD 是	
内存到序列	西戈纳农贸市场是唯一的选择,距酒店 4 英里 没有交通,现在设置导航
金子	西戈纳农贸市场 (Sigona Farmers Market) 距离酒店 4 英里,交通畅通。 它位于阿默斯特街 638 号。

房间	议程党	时间	日期		事件
会议室 50 讨论合并 周三下午	1 点 基础设施团队 周三下午	2 点 周三 下午 2 点	会议室 50 讨论预算 周一	下午 7 点 基础	会议
-	-	设施团队		-	瑜伽活动
-	-			-	医生预约
					会议
-	-	周一	下午 5 点周四上午	-	晚餐
-	-	10 点		-	药品

<p>司机告诉我下次会议是什么时候、谁被邀请参加会议以及下次会议的议程是什么</p> <p>Seq2seq 会议是下午 1 点在 100 号会议室举行的会议</p> <p>+att 您的下一次会议将于周三下午 1 点举行</p> <p>PTRUNK 你的 PAD 会议 PAD 中的 到的</p> <p>Mem2Seq 您的下一次会议将于周一晚上 7 点举行,出席人数将超过预算</p> <p>黄金级您的下一次会议将于周一晚上 7 点与基础设施团队召开,讨论超出预算的问题</p>
--