

使用图注意力生成常识知识感知对话

Hao Zhou¹, Tom Young²、黄敏烈^{1*}、赵海舟³, Jingfang Xu³, Xiaoyan Zhu¹

¹ 清华大学计算机系人工智能实验室对话式人工智能研究组

¹ 北京信息科学技术国家研究中心, 中国

² 北京理工大学信息与电子学院

³ Sogou Inc., Beijing, China

tuxchow@gmail.com, tom@sentic.net, aihuang@tsinghua.edu.cn,

zhaohaizhou@sogou-inc.com.xujingfang@sogou-inc.com.zxy-dcs@tsinghua.edu.cn

抽象的

常识知识对于许多自然语言处理任务至关重要。在本文中,我们

提出了一种新颖的开放域对话生成模型,以展示大规模常识知识如何促进语言理解和生成。给定一个用户帖子,

模型从a中检索相关知识图

知识库,然后用

静态图注意力机制,增强了

帖子的语义信息,从而支持更好地理解帖子。然后,在单词生成过程中,模型会仔细阅读

检索到的知识图谱和知识

每个图中的三元组,以通过动态图注意机制促进更好的生成。这是首次尝试大规模使用

对话生成中的常识知识。此外,与使用的现有模型不同

知识三元组 (实体)分开且独立,我们的模型将每个知识图视为一个

整体,编码更加结构化、互联

图中的语义信息。实验

表明所提出的模型可以生成更多

比最先进的基线更适当且信息丰富的响应。

1 简介

语义理解,特别是在有利于语义理解的情况下

常识性知识或世界事实对于

许多自然语言处理任务 [Wang et al., 2017;

Lin et al., 2017],毫无疑问,这是一个关键因素

对话或会话系统的成功,因为会话交互是一种语义活动[Eggs和Slade,

2005]。在开放域会话系统中,常识知识对于建立有效的交互非常重要,因为社会共享的常识知识是

人们想要了解的一组背景信息

在对话中使用 [Minsky, 1991;马尔科瓦等人,2007年;

施佩尔和哈瓦西,2012;苏托,2015]。

*通讯作者:黄敏烈

近年来,多种神经模型被提出

用于对话生成 [Ritter et al., 2011;尚等

等,2015]。然而,这些模型往往会生成通用的

无法做出适当响应的响应,

在大多数情况下,信息丰富,因为这是一个挑战

仅从会话数据中学习语义交互

[Ghazvininejad et al., 2017] 没有深入了解

用户输入,以及背景知识和上下文

对话。模型可以更好地理解对话

如果它可以访问并充分利用,则可以做出更正确的响应

使用大量常识性知识。例如,要

理解后响应对 “不要在

餐厅,要求免费水”和 “在德国不行。水

比啤酒还贵。带上你自己的水瓶”,我们需要

常识性知识,例如 (水、地点、餐厅)、(免费、相关、费用)等。

之前已经进行了一些研究来在对话生成中引入外部知识[Han et al., 2015;

加兹维内贾德等人,2017;朱等人,2017]。知识点

这些模型中使用的要么是非结构化文本 [Ghazvininejad et al.,

2017] 要么是特定领域的知识三元组 [Zhu

等人,2017]。因此,此类模型面临两个问题

当它们应用于开放域、开放主题的对话生成时。首先,他们高度依赖

非结构化文本的质量或受到小规模的限制,

特定领域的知识。其次,他们通常会利用

分别独立的知识三元组 (实体),

而不是在图中将知识三元组视为一个整体。

因此,它们无法表示图的语义

通过链接的实体和关系。

为了解决这两个问题,我们提出一个常识

知识感知对话模型 (CCM)以促进

开放域会话系统中的语言理解和生成。我们使用大规模常识

知识 [Speer和Havasi,2012] 帮助理解

特定职位的背景信息,并利用这些知识促进回复。模型检索

每个帖子都有一些知识图,然后使用这些图

做出更丰富、更恰当的回应,如图所示

图1中。为了在对话生成中充分利用检索到的图,两种新颖的图注意机制

被设计。静态图注意力机制编码

检索到的帖子图表可以增强帖子的语义表示,这有助于理解帖子。

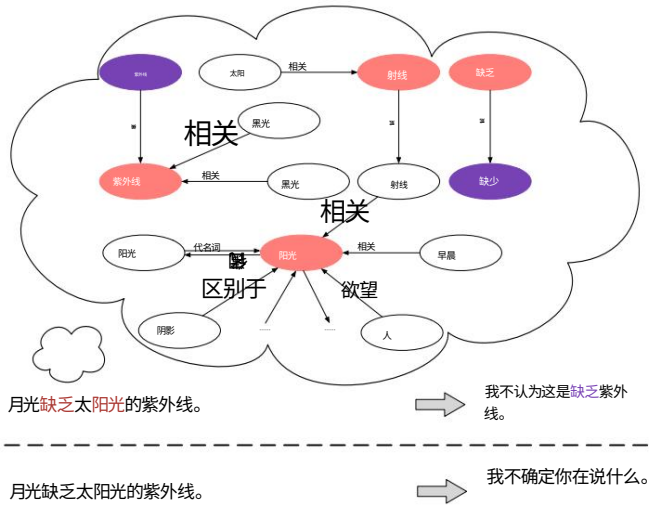


图 1: (以彩色显示效果更佳)我们的两个响应示例
分别考虑/不考虑常识知识的模型 (第一行)和 Seq2Seq (第二行)。

动态图注意力机制仔细阅读
知识图和每个图中的三元组,然后使用
来自图表和三元组的语义信息,以便更好地生成响应。

综上所述,本文做出以下贡献:

- 这项工作首次尝试在神经对话生成中使用大规模常识知识。

在这些知识的支持下,我们的模型可以更好地理解对话,从而做出更恰当、更丰富的回应。

· 我们不是单独和独立地处理知识三元组 (或实体),而是设计静态和动态图注意力机制,将知识边缘三元组视为图,从中我们可以更好地解释实体与其相邻实体的语义

实体和关系。

2 相关工作

开放域对话模型

最近,序列到序列模型 [Sutskever 等人, 2014年; Bahdanau et al., 2014]已成功应用到大规模对话生成,包括神经响应机 [Shang et al., 2015]、分层循环模型 [Serban et al., 2015] 以及许多其他模型 [Sor doni et al., 2015]。这些模型开发了各种技术来提高生成响应的内容质量,

包括多样性促进 [Li et al., 2016;邵等人, 2017],考虑到更多信息[Xing et al., 2017; Mou et al., 2016],以及处理未知单词 [Gu et al., 2016]。然而,通用或无意义的回应仍然存在由于无法良好地在这些模型中常见了解用户输入或其他上下文。

非结构化文本增强会话模型

一些研究将非结构化文本作为外部将知识转化为对话生成 [Ghazvininejad 等

等,2017;龙等人,2017]。[加兹维内贾德等人,2017]使用存储非结构化文本的记忆网络来改进对话生成。[Long et al., 2017]应用了卷积神经网络从非结构化文本中提取知识以生成多轮对话。然而,这些模型在很大程度上取决于非结构化文本的质量,如果文本不相关,这可能会在对话生成中引入噪音。

结构化知识增强对话模型

存在一些引入高质量结构化知识来生成对话的模型 [Han 等人, 2015年;朱等人,2017;徐等人,2017]。[Xu et al., 2017]将结构化的特定领域知识库整合到具有召回门机制的对话生成。[朱 et al., 2017]提出了一种端到端的知识基础使用复制网络的会话模型 [Gu et al., 2016]。然而,这些研究在某种程度上受到小规模的限制。特定领域的知识库,使得它们不适用于开放领域、开放主题的对话生成。

相比之下,我们的模型应用了大规模常识知识库以促进对帖子的理解并生成响应,具有新颖的图形注意力机制。

3 常识性对话模型

3.1 背景 :编码器-解码器框架

首先,我们介绍一个基于序列到序列 (seq2seq)的通用编码器-解码器框架

学习 [Sutskever 等人,2014]。编码器代表一个后序列 $X = x_1x_2 \cdots x_n$ 具有隐藏表示 $H = h_1h_2 \cdots h_n$, 其简要定义如下:

$$h_t = \text{GRU}(h_{t-1}, e(x_t)), \tag{1}$$

其中 $e(x_t)$ 是单词 x_t 的嵌入, GRU 是门控循环单元 [Cho et al., 2014]。

解码器将上下文向量 c_t 和先前解码的单词 $e(y_{t-1})$ 的 e_m 嵌入作为输入,并更新它的状态 s_t 使用另一个 GRU:

$$s_t = \text{GRU}(s_{t-1}, [c_{t-1}; e(y_{t-1})]), \tag{2}$$

其中 $[c_{t-1}; e(y_{t-1})]$ 是两个向量的串联,作为 GRU 网络的输入。上下文向量 c_{t-1} 是对 H 的仔细阅读,它是 H 的加权和编码器的隐藏状态为 $c_{t-1} = \sum_{k=1}^n A^{t-1}_{hk} h_k$ 和 α_k 测量状态 s_{t-1} 和隐藏状态之间的相关性香港。更多详情请参阅 [Bahdanau et al., 2014]。

解码器通过对输出进行采样来生成令牌概率分布可以计算如下:

$$y_t = \arg \max_{y_t} P(y_t | y_{<t}, c_t) = \text{softmax}(W \cdot s_t) \tag{3}$$

其中 $y_{<t} = y_1y_2 \cdots y_{t-1}$,已生成的单词。

1在整篇论文中,普通字母表示离散符号而粗体字母表示向量。

检索到的知识中的 $K(g_i) = \{k_1, k_2, \dots, k_{N_{g_i}}\}$
 $graph\ g_i$,生成图向量 g_i ,如下所示:

$$G = \sum_{n=1}^{\infty} \frac{1}{n} [n; n], \quad (4)$$

$$P_{ij} = \frac{\exp(\beta_{ij})}{\sum_{j=1}^J \exp(\beta_{ij})}, \quad (5)$$

$$Z_n^s = (W_{rrn}) \tanh(W_{hhn} + W_{ttn}), \quad (6)$$

其中 $(h_n, r_n, t_n) = k_n$, W_h, W_r, W_t 分别是头实体、关系和尾实体的权重矩阵。

注意力权重测量关系 r_n 与头实体 h_n 和尾实体 t_n 的关联。

本质上,图向量 g_i 是头向量和尾向量的加权和 $[h_n; t_n]$ 图中包含的三元组。

3.4 知识感知生成器

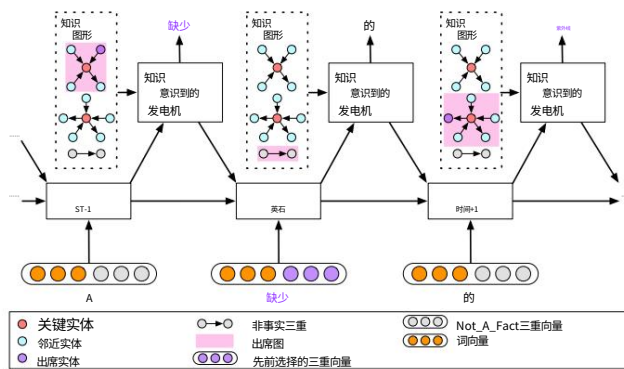


图 4:知识感知生成器动态关注图表 (粉色图表大部分被关注)。

然后,它仔细读取每个图中的三元组,以估计选择三元组的概率,其中三元组的相邻实体(紫色点/单词)用于单词生成。

知识感知生成器旨在通过充分利用检索到的知识图来生成响应,如图4所示。知识感知生成器起着两个作用:1)专心读取检索到的图以获得图感知上下文向量,并使用向量来更新解码器的状态;2)从检索到的图中自适应地选择通用单词或实体来生成单词。正式地,解码器更新其状态如下:

$$k_{st+1} = \text{GRU}(st, [ct; c_t; e(y_t)]), t_{kce}(y_t) = [w(y_t); \quad (7)$$

其中 $e(y_t)$ 是单词向量 $w(y_t)$ 和

从中选择前一个单词(yt)的先知识三元组向量kj的串联, ct是方程 1 中使用的上下文向量。k和c分别是知识图谱t向量{g1, g2, ..., gNG}和知识三元组向量{K(g1), K(g2), ..., K(gNG)}上的上下

文向量。 2、c

图表,然后仔细阅读每个图表中的所有三元组以生成最终的单词。给定解码器状态 st ,它首先关注知识图向量 $\{g_1, g_2, \dots, g_{NG}\}$ 来计算每个图 g_i 的使用概率,定义如下:

哪个

$$\frac{\text{克}}{\text{吨}} = \frac{\text{加}}{\text{蒂吉, --}} \quad (9)$$

$$\alpha_{是}^g = \frac{\exp(\beta t_i^g)}{\sum_{j=1}^{NG_j} \text{指数}(b_{tj}^g)}, \quad (10)$$

$$z_{\text{的}}^G = Vb \tanh(Wbst + Ubgi), \quad (11)$$

其中 $V_b/W_b/U_b$ 是参数, α 是在步骤t选择知识图 g_i 的概率,图上下文 g 是图向量的加权和,向量 c 权重测量解码器状态 s_t 和图向量 g_i 之间的关联。

然后,该模型关注每个图 g_i 中的知识三元组向量 $K(g_i) = \{k_1, k_2, \dots, k_{N_{g_i}}\}$ 来计算选择三元组进行单词生成的概率,形式如下:

$$c_t^k = \frac{1}{k} \sum_{j=1}^k \text{rank}_{j-1}(\mathbf{A}) \quad (12)$$

$$a_{tj}^k = \frac{\text{指数}(b_{tj}^k)}{\sum_{n=1}^N \exp(\beta t_n)}, \quad (13)$$

$$t_{j-}^k = k_j \quad \text{西斯特,} \quad (14)$$

其中 β 可以看作是第 t 个知识三元组向量 \mathbf{k}_j 与解码器状态 \mathbf{s}_t 之间的相似度， α 是在第 t 步从图 G_i 中的所有三元组中选择三元组 t_j 的概率。

最后,知识感知生成器选择通用词或实体词³具有以下分布:

$$\text{在} = [\text{st}; \text{CT}; \text{ct}; \text{c}_{\text{yt}}^{\text{G}} \quad \text{t}^{\text{k}}], \quad (15)$$

$$\text{sigmoid}(V_o \text{ at}), \quad (16)$$

$$P_c(y_t = w_c) = \text{softmax}(W_o a_t), k \quad (17)$$

$$Pe(yt = we) = \alpha t_i \alpha t_j, \quad G \quad (18)$$

$$y_t \quad o_t = P(y_t) = \frac{(1 - y_t)Pg(y_t = w_c)}{y_tPe(y_t = w_e)}, \quad (19)$$

其中 $y_t \in [0, 1]$ 是一个标量,用于平衡实体词 w_e 和通用词 w_c 之间的选择, P_c/P_e 分别是通用/实体词上的分布。最终分布 $P(y_t)$ 是两个分布的串联。

3.5 损失函数

损失函数是预测的 ken 分布 ot 与参考分布 pt 之间的交叉熵

3实体词取自知识三元组的相邻实体。

动态图注意力动态图注意力机制是一

个分层的、自上而下的过程。它首先认真阅读所有知识

训练语料库。此外,我们将监督信号应用于
感知生成层强制教师选择实体或通用单词。一个样本的损失

$\langle X, Y \rangle (X = x_1 x_2 \cdots x_n, Y = y_1 y_2 \cdots y_m)$ 的定义
作为:

$$L(\theta) = - \sum_{t=1}^m \text{ptlog}(\theta_t) - \sum_{t=1}^n (\text{qtlog}(y_t) + (1 - \text{qt}) \log(1 - y_t)), \tag{20}$$

其中 y_t 是选择实体词或
通用词, $\text{qt} \in \{0, 1\}$ 是实体的真实选择
 Y 中的单词或通用单词 第二项用于监督选择实体单词或通用单词的概率
单词。

4 实验

4.1 数据集

常识知识库
采用 ConceptNet4 作为常识知识库。它
不仅包含诸如 “巴黎是世界各国的首都” 之类的世界事实
法国” 始终是真实的, 但也是非正式的关系
作为日常知识一部分的常见概念之间
比如 “狗是宠物”。这个特征在我们的实验中是可取的, 因为识别非正式关系的
能力
在开放领域中, 共同概念之间的联系是必要的
对话设置。为了简单起见, 我们删除了包含多词实体的三元组, 并保留了 120,850
个三元组
拥有 21,471 个实体和 44 个关系。

常识对话数据集
我们采用了站点 10M reddit 单轮对话5。
由于我们的目标是使用常识知识来促进语言理解和生成, 因此我们过滤了

具有知识三元组的原始语料库。如果事后回复
对不能通过任何三元组连接 (即, 一个实体出现在帖子中, 另一个实体出现在响
应中), 该对将
被删除。统计结果见表1。

我们随机抽取了 10,000 对进行验证。去测试
常识性知识如何帮助理解常见问题
或者帖子中的罕见概念, 我们构建了四个测试集: 高频对, 其中每个帖子都具有前
25% 的频繁出现
每个帖子包含的单词、中频对
至少有一个词的频率在 25% 范围内 -
75%, 75% - 100% 范围内的低频对,
和 OOV 对, 其中每个帖子都包含词汇表之外的内容
字。每个测试集有 5,000 对随机抽样
数据集6。

4.2 实施细节

我们的模型是用 Tensorflow7 实现的。编码器
解码器具有 2 层 GRU 结构, 包含 512 个隐藏层

4<https://conceptnet.io>
5<https://www.reddit.com/r/datasets/>
评论/3bxl7/i_have_every_publicly_
可用的 reddit 评论/
6 我们的数据可在以下网址获取: <http://coai.cs.tsinghua.edu.cn/hml/dataset/#commonsense>
7 <https://github.com/tensorflow/tensorflow>

对话对常识知识库			
培训	3,384,185 实体 21,471		
验证	10,000 44	关系	
测试	20,000	三倍	120,850

表1:数据集和知识库的统计数据。

每层的单元格并且它们不共享参数。这
词嵌入大小设置为 300。词汇大小为
限量 30,000 名。我们使用 TransE [Bordes et al., 2013] 来
获取实体和关系表示。嵌入
实体和关系的大小设置为 100。

我们使用小批量大小为 100 的 Adam 优化器。
学习率为 0.0001。模型最多运行 20 个
epoch, 每个模型的训练阶段大约需要一个
在 Titan X GPU 机器上使用一周。我们的代码可在以下位置获取:
<https://github.com/tuxchow/ccm>。

4.3 基线

我们选择了几个合适的基线:

- seq2seq 模型 (Seq2Seq) [Sutskever 等人, 2014],
广泛应用于开放域会话
系统。
- 基于知识的模型 (MemNet) 改编自
[Ghazvininejad et al., 2017], 其中内存单元
存储知识三元组的 TransE 嵌入。
- 复制网络 (CopyNet) 模型 [Zhu et al., 2017],
它从知识三元组中复制一个单词或从词汇表中生成一个单词。

4.4 自动评估

指标: 我们采用困惑度 [Serban et al., 2015] 在内容层面评估模型 (内容是否

符合语法且与主题相关)。我们还计算了
每个响应的实体数量, 用于衡量模型从常识知识中选择概念的能力

一代的基础。该指标由实体分数表示。
结果: 如表 2 所示, CCM 在所有测试集上获得最低的平均复杂度, 表明 CCM 可
以更好地理解用户的帖子并生成更多语法响应。此外, CCM 从以下实体中选择最
多的实体:

生成过程中模型之间的常识知识, 证明常识知识可以真正

促进响应的生成。
更有趣的是, CCM 在低频帖子中比在高频帖子中更频繁地使用常识知识 (实
体得分: 1.196 vs. 1.156)。这是符合

根据我们的直觉, 罕见的概念需要更多背景
理解和回应的知识。对于困惑,
高频帖子的得分低于低频帖子
的 (35.36 vs. 40.67), 因为常见的概率
单词可以得到更充分的训练。

4.5 人工评估

我们求助于众包服务 Amazon Mechanical
Turk, 用于手动注释。随机抽取 400 个帖子进行手动注释。我们对 CCM 生成的
响应和由 CCM 生成的响应进行了成对比较

模型	全面的		高频。中频低频。		ent. ent. ppx。 ppx。				PTA	
	ppx。	ent.			ppx。	ent.			ppx。	ent.
序列到序列	47.02	0.717	42.41	0.713	47.25	0.740	48.61	0.721	49.96	0.669
内存网络	46.85	0.761	41.93	0.764	47.32	0.788	48.86	0.760	49.52	0.706
复制网络	40.27	0.96	CCM	39.18	36.26	0.91		40.99	0.97	
	1.180	35.36	1.156	39.64	1.191	40.67	1.196	40.87	1.162	

表 2:使用困惑度 (ppx.) 和实体得分 (ent.) 自动评估。

模型	总体信息。		高频。中频低频。信息。信息。		应用程序。应用程序。				PTA	
	应用程序。				应用程序。	信息。			应用程序。	信息。
CCM 与 Seq2Seq	0.616	0.662	0.605	0.656	0.549	0.624	0.636	0.650	0.673	0.716
CCM 与 MemNet	0.602	0.647	0.593	0.656	0.566	0.640	0.622	0.635	0.626	0.657
CCM 与 CopyNet	0.600	0.640	0.606	0.669	0.586	0.619	0.610	0.633	0.596	0.640

表 3:手动评估适当性 (app.) 和信息性 (inf.)。分数是CCM战胜竞争对手的百分比
删除 “领带”对后。CCM 明显优于所有测试集上的所有基线（符号检验,p 值 < 0.005）。

同一职位的基线。总共有 1,200 对
因为我们有三个基线。对于每个响应对,有七个
聘请法官根据以下两个指标对这两种回答做出优先选择。领带是允许的。请注意,系统标识符在表示法期间被屏蔽。

指标:我们定义了两个指标:
内容水平（回答的语法、主题和逻辑是否合适）;以及知识的信息量

级别（响应是否提供新信息和
除了帖子之外的知识）。
注释统计:我们计算协议以确保评估者间的一致性。对于适当性,至少 4 位评委给出
相同标签的配对的年龄百分比 (4/7

一致)达96.3%,且该百分比至少
5/7 一致率为 60.2%。为了提供信息,百分比
至少 4/7 的同意率为 90.4%,至少 5/7 的同意率为 90.4%
同意率为55.1%。
结果:结果如表3所示,得分为
去除 “平局”后 CCM 赢得基线的百分比
对。CCM 在以下方面显着优于所有基线
所有指标的两个指标（符号检验,p 值 < 0.005）
测试集。此外,CCM 有大约 60% 的获胜机会
最强的基线是 CopyNet,它单独且独立地使用知识边缘三元组（实体）。这

展示了我们的图注意力机制的有效性。

值得注意的是,CCM 在 Seq2Seq 上获胜的概率
OOV 数据集明显高于
高频数据集（适当性为 0.673 与 0.605,
信息量为0.716 vs. 0.656）,这进一步表明常识知识对于理解后期的罕见概念更
有用,因为 Seq2Seq 没有能力

使用这些知识。对于 MemNet 和 CopyNet,我们没有
观察这种差异,因为两个基线具有
或多或少运用知识的能力。

4.6 案例研究

表 4 显示了一个示例对话。
帖子中的 “易碎”一词是已知的实体词

邮政	你为何如此脆弱?
知识	(玻璃,相关,易碎) , (易碎,相关,易碎) 、 (规则,相关,可破坏)
Seq2Seq	我不是一个OOV,我只是一个OOV。
MemNet	我不是OOV。我只是一个非常好的人。
复制网	我不是。我只是一个潜伏者。
	因为我是个脆弱的人CCM 。

表 4 :所有模型生成的示例响应。

边缘基础以及所有词汇表之外的单词
楷模。如果无法获得常识性知识,
Seq2Seq 模型无法理解该帖子,因为
词汇表外的单词 “易碎” ,从而生成
OOV的话。MemNet 可以生成一些有意义的单词
因为它读取了内存中的三重嵌入,但仍然输出OOV。CopyNet 可以读取和复制知
识边缘三元组中的单词。然而,CopyNet 生成的实体词较少

比我们的（见表 2）,因为它只处理单独的知识边缘三元组。相反,CCM 将知识图
谱视为
整体并编码更结构化、关联的信息
通过链接的实体和关系。因此,它通过更好地利用知识产生更合理的反应。这个简单的
例子表明CCM可以生成更合适的

和比基线更丰富的反应。

5 结论和未来工作

在本文中,我们提出了常识性知识意识
会话模型（CCM）展示常识知识如何促进语言理解和

开放域对话系统中的生成。自动和手动评估表明CCM可以生成

比最先进的基线更合适、信息更丰富的响应。

作为未来的工作,我们的图注意力机制可能会在
激发其他任务的使用常识性知识。

6 致谢本工作得到了国家自然科学基金委员会 (No.61272227/61332007)和国家重点基础研究发展计划 (973计划) (No.2013CB329403)的部分支持。

参考

[Bahdanau 等人,2014] Dzmitry Bahdanau,Kyunghyun Cho 和 Yoshua Bengio.通过共同学习对齐和翻译进行神经机器翻译。 CoRR,abs/1409.0473,2014。

[Bordes 等人,2013] Antoine Bordes,Nicolas Usunier,Alberto Garcia-Duran, Jason Weston 和 Oksana Yakhnenko.翻译嵌入以多关系数据进行建模。载于 NIPS,第 2787-2795 页,2013 年。

[Cho 等人,2014] Kyunghyun Cho,Bart van Merriënboer,Changlun Gulcehre,Dzmitry Bahdanau,Fethi Bougares,Holger Schwenk 和 Joshua Bengio.使用 RNN 编码器-解码器学习短语表示以进行统计机器翻译。载于 EMNLP,第 1724-1734 页,2014 年。

[埃金斯和斯莱德,2005] 苏珊娜·埃金斯和戴安娜·斯莱德。分析随意谈话。 Equinox 出版有限公司,2005 年。

[Ghazvininejad 等人,2017] Marjan Ghazvininejad,Chris Brockett,Ming-Wei Chang,Bill Dolan,Jianfeng Gau,Wen-tau Yih 和 Michel Galley.基于知识的神经对话模型。 CoRR,abs/1702.01932,2017。

[Gu et al., 2016] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor OK Li. Incorporating copying mechanism in sequence-to-sequence learning. In ACL, pages 1631–1640, 2016.

[Han 等人,2015] Sangdo Han,Jeesoo Bang,Seonghan Ryu 和 Gary Geunbae Lee.利用知识库为自然语言对话监听代理生成响应。载于 SIGDIAL,第 129-133 页,2015 年。

[Li et al., 2016] 李继伟,米歇尔·加利,克里斯·布罗克特,高剑峰和比尔·多兰。神经对话模型的多样性促进目标函数。载于 NAACL,第 110-119 页,2016 年。

[Lin et al., 2017] Hongyu Lin, Le Sun, and Xianpei Han.使用异构知识进行推理以实现常识机器理解。 EMNLP,第 2032-2043 页,2017 年。

[Long et al., 2017] Yinong Long, Jianan Wang, Zhen Xu, Zongsheng Wang, Baoxun Wang, and Zhuoran Wang.知识增强的生成对话服务代理。 DSTC6 研讨会,2017 年。

[Markova et al., 2007] Ivana Markova,Per Linell,Michelle Grossen 和 Anne Salazar Orvig.焦点小组对话:探索社会共享知识。春分出版社,2007 年。

[明斯基,1991] 马文·明斯基。心灵社会:对四篇评论的回应。人工智能,48 (3) : 371-396,1991。

[Mou et al., 2016] 牟丽丽,宋一平,严睿,李革,张璐,金志。向后和向前序列的序列:一种生成短文本对话的内容介绍方法。《COLING》,第 3349-3358 页,2016 年。

[Ritter 等人,2011] Alan Ritter,Colin Cherry 和 William B. Dolan.社交媒体中数据驱动响应生成。载于 EMNLP,第 583-593 页,2011 年。

[Serban 等人,2015] Iulian Vlad Serban,Alessandro Soru,Yoshua Bengio, Aaron C. Courville 和 Joelle Pineau.电影对话的分层神经网络生成模型。 CoRR,abs/1507.04808,2015。

[Shang et al., 2015] Lifeng Shang, Zhengdong Lu, and Hang Li. Neural responding machine for short-text conversation. ACL,第 1577-1586 页,2015 年。

[Shao 等人,2017] Louis Shao,Stephan Gouws,Denny Britz,Anna Goldie, Brian Strope 和 Ray Kurzweil.使用神经对话模型生成时间且多样化的响应。 CoRR,abs/1701.03185,2017。

[Sordoni 等人,2015] Alessandro Sordoni,Michel Galley,Michael Auli,Chris Brockett,Yangfeng Ji,Margaret Mitchell,Jian-Yun Nie,Jianfeng Gau 和 Bill Dolan.一种用于上下文敏感生成对话响应的神经网络方法。载于 NAACL,第 196-205 页,2015 年。

[Souto,2015] Patricia Cristina Nascimento Souto.从差异中创造知识:所需的对话性和对话能力。 RAI Revista de Administraco e Inovaco [Speer 和 Havasi,2012] Robert Speer 和 Catherine Havasi.在 ConceptNet 5 中表示一般关系知识。LREC,第 12 (2) :60-89,2015。 3679-3686 页,2012 年。

[Sutskever 等人,2014] Ilya Sutskever,Oriol Vinyals 和 Quoc V Le.使用神经网络进行序列到序列学习。载于 NIPS,第 3104-3112 页,2014 年。

[Velickovic 等人,2017] Petar Velickovic,Guillem Cucurull,Arantxa Casanova, Adriana Romero,Pietro Lio 和 Yoshua Bengio.图注意力网络。 CoRR,abs/1710.10903,2017。

[Wang et al., 2017] 王冰宁,刘康和赵军。用于常识机器理解的条件生成对抗网络。 IJCAI,第 4123-4129 页,2017 年。

[Xing et al., 2017] Chen Xing, Wei Wu, Yu Wu, Jie Liu, Yalou Huang, Ming Zhou, and Wei-Ying Ma. Topic aware neural response generation. In AAAI, pages 3351–3357, 2017.

[Xu et al., 2017] Zhen Xu, Bingquan Liu, Baoxun Wang, Chengjie Sun, and Xiaolong Wang. Incorporating loose structured knowledge into conversation modeling via recall-gate lstm. In IJCNN, pages 3506–3513. IEEE, 2017.

[Zhu et al., 2017] Wenya Zhu, Kaixiang Mo, Yu Zhang, Zhangbin Zhu, Xuezheng Peng, and Qiang Yang. Flexible end-to-end dialogue system for knowledge grounded conversation. CoRR, abs/1709.04264, 2017.