基于深度神经网络的对话系统研究综述

邢春康,任勋益

(南京邮电大学 计算机学院、软件学院、网络空间安全学院,江苏 南京 210023)

摘 要: 随着深度学习技术的崛起,自然语言处理应用取得了显著进展,特别是在对话系统研究中。为此,阐述对话系统基本流程,全面梳理基于深度学习的对话系统技术,包括卷积神经网络、循环神经网络和注意力机制三大类关键技术。同时,介绍3种模型的基本原理,并从信息抽取、对话状态追踪和对话生成方面深入分析比较了各基本模型及其衍生模型在对话任务上的应用、特点和优缺点。最后,指出对话任务中依旧存在的问题,并提出可行解决方案。

关键词: 深度学习;自然语言处理;注意力机制;对话系统;神经网络

DOI: 10.11907/rjdk.231932 **开放科学(资源服务)标识码(OSID):**

中图分类号: TP18;TP391.1 **文献标识码:** A **文章编号:** 1672-7800(2024)009-0020-11

Review of Dialogue Systems Based on Deep Neural Networks

XING Chunkang, REN Xunyi

(*School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing* 210023, *China*)

Abstract: With the rise of deep learning technologies, significant advancements have been achieved in the field of natural language processing(NLP), particularly in the domain of dialogue systems. This paper begins by providing an overview of the fundamental processes involved in dialogue systems. Subsequently, it comprehensively reviews deep learning-based techniques for dialogue systems, encompassing three key categories: convolutional neural network(CNN), recurrent neural network(RNN), and attention mechanism(AM). The paper introduces the principles of these models, and then provides an in-depth analysis and comparison of the applications, characteristics, and advantages and disadvantages of various basic models and their derivative models in dialogue tasks form the perspectives of information extraction, dialogue state tracking, and dialogue generation. Finally, this paper enumerates persisting challenges within dialogue tasks, and proposes feasible solutions.

Key Words: deep learning; natural language processing; attention mechanism; dialogue system; neural network

0 引言

随着计算机软硬件技术的不断提升,互联网在全球范围内逐渐普及,互联网用户数量的高速增长也使得每天在全球范围内产生的数据量呈现爆炸性增长。这些数据为人工智能提供了充分且可靠的实验数据支持,进一步推动了人工智能的研究和发展。

自然语言处理(Natural Language Processing,NLP)是当下人工智能研究领域的一个重要研究方向,主要研究如何使用计算机处理人类书面和口头表达的自然语言信息,以理解实际含义,并生成与人类自然语言相符的输出,从而使计算机能够像人一样与人类进行交互。自然语言处理的研究范围包括但不限于文本分类、情感分析、人机对话和机器翻译。

近年来,随着互联网生成大规模数据和深度学习的兴起,自然语言处理技术取得了显著进展。这一进展主要体现在深度学习领域,神经网络模型在很大程度上优化了传统NLP技术产生的问题,包括一词多义和上下文信息缺失导致的文本分析结果误差增加。

因此,本文以人机对话为切入点,对当下深度神经网络(Deep Neutral Network, DNN)在NLP领域中人机对话的

收稿日期: 2023-10-14

扫描二维码阅读全文:

作者简介: 邢春康(1999-),男,南京邮电大学计算机学院、软件学院、网络空间安全学院硕士研究生,研究方向为自然语言处理(意图识别);任勋益(1973-),男,博士,南京邮电大学计算机学院、软件学院、网络空间安全学院副教授,研究方向为入侵检测及防御。

发展进行研究综述，并详细阐述运用于对话系统的 DNN 关键技术。

1　对话系统

对话系统是一种人工智能系统，能够与用户进行自然语言交互。其基本流程如图 1 所示，包括输入、输出、自然语言理解（Natural Language Understanding，NLU）、对话管理（Dialogue Management，DM）和自然语言生成（Natural Language Generation）等几个模块。同时，对话任务的完成需要依赖知识库的数据支持。

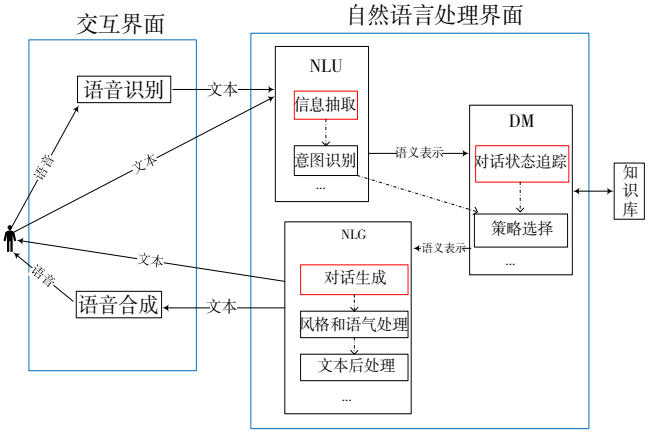

Fig. 1　Basic process of dialogue system

图 1　对话系统基本流程

用户以语音或文本形式发起对话，系统将用户输入转换成文本后，进行初步处理，如分词和去除停用词。处理后的文本传递到 NLU 模块，NLU 模块负责信息抽取和用户意图识别以获取关键信息和参数。DM 模块根据 NLU 模块的语义表示管理对话状态，包括跟踪对话上下文和历史记录。基于对话状态和用户意图，DM 模块采用一定策略确定系统如何回应用户的输入，选择合适的对话行为和生成对话响应。NLG 模块负责根据 DM 模块的语义表示规划生成内容的风格和逻辑结构，然后对话生成模块根据系统规划生成自然语言文本。生成的文本可以根据需要添加适当的语气和情感，以使其更加自然。此外，还可以进行一些后处理操作，如修正拼写错误、排版和控制文本长度等。

总之，信息抽取、对话状态追踪和对话生成模块在对话系统中扮演重要的支撑角色，因此下文将重点围绕这 3 个模块进行阐述。

2　基于 CNN 的对话系统

2.1　卷积神经网络原理

卷积神经网络（Convolutional Neural Networks，CNN）是一种深度前馈神经网络。最初，它自计算机视觉领域设计而成，后来被证明对问答任务、文本分类等 NLP 任务中也表现出色。CNN 结构如图 2 所示，主要包括输入层、输出层、卷积层、池化层和全连接层。其中，卷积层是 CNN 的核心部分，通过卷积操作计算出数据的输入特征。在卷积层提取特征后，输出的特征传递至池化层，进行特征选择和信息过滤。这有助于减少模型参数数量，提高模型鲁棒性，并在一定程度上缓解过拟合问题。接下来，全连接层将卷积层和池化层输出特征进行加权，将高维度数据转换为低维度数据，保留有用信息。该过程有助于进一步提取和总结特征，为最终任务提供有力支持。

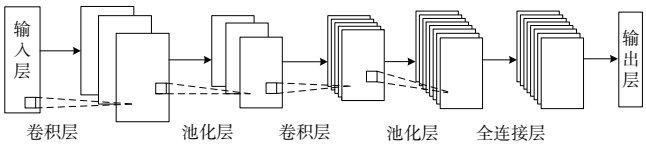

Fig. 2　CNN structure

图 2　CNN 结构

2.2　基于 CNN 的信息抽取

信息抽取旨在从非结构化文本中抽取结构化的信息，主要包括实体抽取、实体关系抽取、事件抽取和事件关系抽取等任务[1]。

实体提取，又称为命名实体识别（Named Entity Recognition，NER），在文本分类、机器翻译和自动问答等 NLP 下游任务中具有重要支撑作用。早期基于规则和字典的 NER 方法存在一些局限，包括构建成本高、规则只适用于特定领域，且扩展性差。因此，基于深度学习的 NER 模型已经成为主流。

Santos 等[2]提出一种针对 NER 任务的字符级词嵌入神经网络（Character-level Word Embedding Neural networks，CharWNN）模型。该模型通过使用单词级和字符级表示进行序列分类，CharWNN 采用了一个卷积层，能够有效提取来自任何大小单词中的字符级特征。在 HAREM I 和 SPA CoNLL-2002 语料库上，分别对葡萄牙语和西班牙语进行了实验。在西班牙语 NER 任务中，将 CharWNN 结果和 SPA CoNLL-2002 语料库中最先进的系统进行比较；在葡萄牙语 NER 任务中，将 CharWNN 和 ETLCMT 语料库（最先进的 HAREM I 语料库）进行比较。实验结果如表 1 所示，在两种 NER 任务中，CharWNN 都取得了相应语料库中最好的精准率、召回率和 F1 值。

CharWNN 相较于传统 CNN 有以下优点：①能捕获文本每个字符的信息，不需要事先构建词汇表，而是直接从字符级别构建文本表示。这使得它在处理特殊字符、日期和缩写等情况下更有鲁棒性；②通过卷积操作，能够捕获字符级别局部特征和模式，从而更细粒度地理解文本结构和语法信息；③具有一定的抗干扰性，即使存在单词拼写错误，仍能提供有用的特征。

Strubell 等[3]提出了迭代卷积神经网络（Iterated Dilated Convolutional Neural Network，IDCNN）模型。该模型通过迭代 CNN 的方式扩大卷积核的感受野，同时随着迭代层数的增加，IDCNN 扩展的卷积和有效输入宽度呈现指数增

Table 1 Comparison among the state-of-the-art for SPA CoNLL-2002 and HAREM I corpus

表 1 SPA CoLL-2002 和 HAEM I 语料库最新技术比较

语种	系统	特征	精准率/%	召回率/%	F1 值/%
西班牙语	CharWNN	词嵌入、字符嵌入	82.21	82.21	82.21
	AdaBoost	单词、正交的、POS 标签、触发单词、词袋、实体长度	81.38	81.40	81.39
葡萄牙语	CharWNN	词嵌入、字符嵌入	74.54	68.53	71.41
	ETLcmt	单词、POS 标签、NP 标签、单词长度	77.52	53.86	63.56

长。与传统的 CNN 相比，IDCNN 有以下优势：①连续扩大的卷积核的感受野有助于捕获文本长距离依赖关系，对于文本分类或序列标记等需要更多上下文信息的 NER 任务更加有益；②IDCNN 的感受野自适应扩大，使得模型更好地适应不同类型的输入；③IDCNN 通过使用不同尺寸和膨胀率的卷积核以捕获不同尺度的特征，从而有效地提取文本的多尺度信息。

实体提取之后，另一信息抽取分支任务是实体关系提取，也叫关系事实提取。其目的是辨别句子中给定实体之间的关系。Zheng 等[4]利用简单的 CNN 对句子中元素之间的多种关系进行分类，CNN 接收特定的向量矩阵并将其作为输入，经过卷积层和池化层的操作将输入转换成固定长度向量，再利用其他特征进行语义信息综合。在关系提取中，远程关系监督也是解决训练数据生成问题的一种常用方法。考虑到远程监督可能导致错误的标签等问题，Zeng 等[5]结合多实例学习提出了分段卷积神经网络（Piecewise Convolutional Neural Networks，PCNN）模型，该模型可以自动学习特征，用于远程监督关系提取，而不需要进行复杂的 NLP 预处理。该模型设计了一个分段最大池化层，用于捕获两个实体之间的结构信息。与传统 CNN 仅能返回单个最大值，无法捕获实体之间结构信息的不足相比，PCNN 的分段最大池化层能返回每个段的最大值，从而更好地捕获实体之间的结构信息。

事件提取和事件关系提取是信息领域的重要研究内容。事件提取旨在识别特定类型的事件，而事件关系提取则涉及在一段文本中提取出两个事件之间的潜在关系[6]。早期的事件识别主要依赖于规则和浅层神经网络，但存在明显不足：①基于规则的事件识别需要不断更新规则，以保持在新领域中的最佳性能，这导致了可移植性和鲁棒性欠佳；②由于基于浅层神经网络的事件识别采用梯度下降算法，其搜索方向是固定的，因此容易陷入局部最优解，而且随着神经网络层数的增加，结果误差会增大，性能也将下降。

Zhang 等[7]提出基于 DL 的中文应急事件识别（Chinese Emergency Event Recognition，CEERM）模型，该模型利用 DL 模型提取事件触发器的深层语义特征，而无需手动设定规则，同时克服了浅层神经网络容易陷入局部最优解的问题。传统 CNN 通常只能捕捉句子中最重要的信息，然而在处理包含多个事件的句子时，可能会错过其中有价值的事实。为了解决该问题，Chen 等[8]提出动态多池化卷积神经网络（Dynamic Multi-Pooling Convolutional Neural Networks，DMCNN）模型。DMCNN 使用动态多池化层捕获每个部分的最大值，这些最大值由事件触发器和事件参数分割开来，从而更有可能捕获更多有价值的线索。实验结果表明（见表 2），DMCNN 在性能上提升了 2.8% 和 4.6%，相较于 CNN 表现更优。

Table 2 Comparison of the event extraction scores obtained by traditional，CNN and DMCNN models

表 2 传统、CNN 和 DMCNN 模型获取的事件提取分数比较

阶段	方法	1/1	1/N	all
		F1	F1	F1
触发器	Embedding+T	68.1	25.5	59.8
	CNN	72.5	43.1	66.3
	DMCNN	74.3	50.9	69.1
参数	Embedding+T	37.4	15.5	32.6
	CNN	51.6	36.6	48.9
	DMCNN	54.6	48.7	53.5

这表明，DMCNN 在中文应急识别任务中表现更出色，能更有效地提取有关事件信息。

2.3 基于 CNN 的对话状态追踪

对话状态追踪（Dialogue State Tracking，DST）作为对话系统重要任务，其目的在于追踪对话中的用户意图和系统状态。在早期，DST 依赖手写规则更新对话状态，然而这种方法需要不断手动更新规则以适应新数据，非常耗时且繁琐。随着基于概率统计的 DST 模型的出现，它们通过最大化目标函数更新参数，突破了手工编写规则的限制。经典 DST 模型包括基于马尔可夫决策和贝叶斯网络的方法。目前，随着 DNN 表示学习的兴起，许多 DST 研究都基于 DNN 而开展[9]。这些方法利用 DNN 实现学习表示，从而更好地捕捉对话中的复杂关系和语境。

DSTC5 引出一种新的情景跨越状态追踪任务，要求参与者基于英语语料库构建状态追踪器，并用未标记的中文语料库进行评估。尽管数据集中提供了计算机生成的英语和中文语料库翻译，但是这些翻译中存在错误，粗心地使用这些数据集对追踪器的性能产生了很大损耗。因此，Shi 等[10]提出了多通道 CNN 模型，用于跨语言对话框状态追踪。在 DSTC5 的评估中，这种多通道 CNN 模型有效地提高了针对翻译错误的鲁棒性。

针对 DSTC6，Korpusik 等[11]提出了一种全新的端到端目标导向对话追踪方法，其目标更接近于生成下一个系统响应，而不仅仅是从餐厅预订对话的候选响应中选择最佳系统响应。该模型使用一个 CNN 对对话历史中的每个话语进行语义标记，以更新对话状态，另一个 CNN 预测最佳的行动选择，最后一个响应生成步骤用最终对话状态中的

槽值填充模板。

2.4　基于CNN的对话生成

近年来，构建能够在开放领域话题上与人类进行自然、持续对话的计算机系统成为研究的热门话题。对话机器人中的一个关键任务是响应选择，其目的是在给定对话上下文的情况下，从一组候选语句中选择最匹配的回答。早期研究主要集中在使用上下文的最后一句话以匹配回复，这种方式称为单轮对话。Li 等[12]利用树结构作为 DNN 模型的输入，取得了更好的效果；Lu 等[13]提出基于 DNN 的匹配模型以改进响应选择；Hu 等[14]通过使用 CNN 进一步优化了性能。该模型不仅很好地表示了句子的层次组合和池化结构，还在不同层次上捕获了丰富的匹配模式。这些模型具有相当的通用性，不需要先验语言知识，因而可以应用于不同性质和语言的匹配任务。

2.5　CNN 小结

总体而言，基于 CNN 的对话系统通过卷积操作和特征映射以捕获输入文本的局部特征和模式，然后通过全连接层将特征映射转化为最终语义标签。CNN 具有平移不变性，能够识别相同的模式。并且，CNN 还能共享参数，从而减少了模型参数数量，同时也降低了过拟合风险。此外，CNN 能并行处理，加快了模型训练速度。

尽管 CNN 有着诸多优点，但仍有不少无法忽略的缺点：①CNN 只能捕获局部模式，而且由于 CNN 的输入元素之间相互独立，因此无法充分考虑文本序列中的全局上下文；②CNN 不擅长建模标签之间的依赖关系，而这在实体关系提取任务中又相当重要；③CNN 的结构决定了输入大小固定，这容易导致信息损失；④CNN 模型需要大量标记数据用于训练，因此在特定领域会受到一定限制。

3　基于RNN的对话系统

3.1　循环神经网络原理

循环神经网络（Recurrent Neural Networks，RNN）是一种递归神经网络[15]，以序列数据作为输入，在序列的演进方向上递归，并且所有循环单元按链式连接在一起。与常规前馈神经网络（如 CNN）相比，RNN 具有更强的循环性。其核心思想是在每个时间步接收一个输入向量和前一个时间步的隐藏状态。然后，RNN 使用权重矩阵和激活函数计算新的隐藏状态，新的隐藏状态会传递给下一个时间步，依次递归。RNN 结构如图 3 所示，其主要包括输入层、隐藏层和输出层。

给定一个输入序列$(x1,\cdots,xT)$，对于t时刻$(1\leqslant t\leqslant T)$，RNN 通过迭代计算公式$h_t=\tanh\left(w_{hx}x_t+w_{hh}h_{t-1}+b_h\right)$和$o_t=w_{oh}h_t+b_o$以计算隐藏状态序列$(h1,\cdots,hT)$和输出序列$\left(o_1,\cdots,o_T\right)$。其中，$w_{hx}$是隐藏输入的权重矩阵；$w_{hh}$是隐藏到隐藏（递归）的权重矩阵；向量$b_h$和$b_o$是隐藏层的偏置；$h_{t-1}$为$t-1$时刻隐藏层单元的输出；$x_t$为$t$时刻隐变量输入。

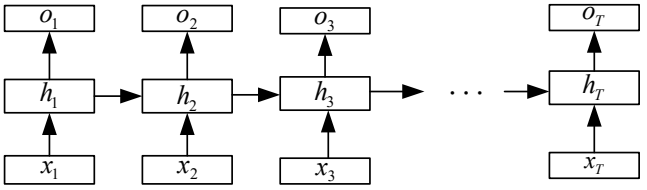

Fig. 3　RNN structure

图 3　RNN 结构

3.2　基于RNN的信息抽取

考虑到句子含义和文本单词时序的相关性，信息提取中的实体关系可以看作是一个时序任务，因此可以用 RNN 提取实体关系。然而，仅使用单一正向的 RNN 结构存在一潜在问题，即 RNN 在单向预测句子语义时无法充分利用未来时刻信息。为了解决该问题，Zhang 等[16]提出双向循环神经网络（Bidirectional Recurrent Neural Networks，BRNN）模型。该模型将单词按正向和逆向依次输入到两个方向不同的 RNN 中，然后将正向和逆向隐状态相加以生成当前时刻的隐状态输出。接着，对所有隐状态进行最大池化操作，最终得到一个向量结果表示。

然而，尽管在 RNN 结构上进行各种优化，但仍无法避免由于 RNN 在“循环”过程中涉及梯度连乘而引起的梯度消失或梯度爆炸问题。

针对梯度爆炸问题，Hochreiter 等[17]提出长短期记忆神经网络模型（Long Short-Term Memory，LSTM）模型，并且在序列模型任务中得到成功运用。LSTM 和 RNN 在基本结构上都是链式的，但 LSTM 在 RNN 神经元结构基础上引入了专门用于储存长期信息的记忆单元[18]，它包含 4 个不同的神经网络层，用于信息交互。LSTM 结构如图 4 所示，其核心部分包括遗忘门、输入门和输出门：遗忘门用来控制上一记忆单元状态c_{t-1}、上一时刻的输出h_{t-1}和当前时刻x_t一起作为当前时刻输入；输入门用来控制当前记忆单元状态c_t；输出门表示需要确定输出什么值。

邓琴等[19]提出一种基于 LSTM-SNP 的命名实体识别方法，这是首次结合非线性 SNP[20]和 LSTM，形成具有门控

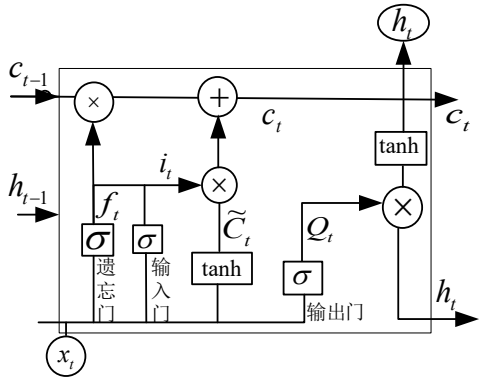

Fig. 4　Neural cell structure of LSTM

图 4　LSTM 神经元细胞结构

机制的深度学习通用模型。在文献[19]中，通过将 CRF、GloVe 和 GloVe-CNN 等不同的深度学习组件添加到 LSTM-SNP 中，对比分析了 LSTM-SNP 与传统 LSTM 及其变体 BILSTM 的性能差异。实验结果显示，LSTM-SNP 和 LSTM 模型性能相近，而与 BILSTM 相比仍有一定差距。但在加入 CRF 或 CNN 组件后，该模型性能显著提高。

在信息抽取领域的关系提取分支任务中，Xu 等[21]采用 LSTM 模型进行关系提取。在 LSTM 基础上，Graves 等[22]提出双向长短期记忆神经网络（Bidirectional Long Short-Term Memory，BiLSTM）。该模型与前面提到的 BRNN 模型类似，它包括了正向和逆向两个隐藏 LSTM 层。为了更好地提取上下文信息，Zhang 等[23]提出使用 BiLSTM 模型进行优化。该方法提取句子双向的隐状态输入，组合词汇特征和句子级别的特征，以丰富句子特征表示。

针对序列标记任务，Huang 等[24]在自身 LSTM-CRF 模型上进行改进，提出了 BiLSTM-CRF 模型，并应用于 NLP 基准序列标记数据集。该模型不仅捕获了过去输入特征和句子级语义信息，还能捕获未来的输入特征。受 Huang 等[24]工作启发，后续许多研究人员也采用了 BiLSTM-CRF 模型，并在中文、德语和葡萄牙语等多种语言 NER 任务中取得出色表现[25-27]。

然而，LSTM 模型的参数较多，因此训练 LSTM 模型通常需要更多的计算资源和数据。因此，Cho 等[28]提出门控循环单元（Gated Recurrent Unit，GRU），其结构如图 5 所示。

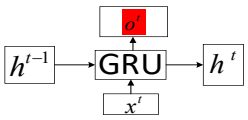

Fig. 5 The input and output structure of GRU

图 5 GRU 的输入输出结构

GRU 的一大特点是容易训练，因为它没有额外的记忆单元，只有更新门和重置门。这减小了内部结构的复杂性，减少了参数数量，提高了训练速度，并且降低了计算成本。Yang 等[29]提出 GRU-CRF 模型，用于序列标记，该模型使用 GRU 在字符级别和单词级别上编码上下文信息，而 CRF 层主要用于标签预测。常用的基于 RNN 的 NER 方法及其原理和特点如表 3 所示。

Table 3 Comparison of NER tasks based on RNN

表 3 基于 RNN 的 NER 任务比较

方法名称	原理	特点
LSTM[17]	通过门控管理细胞单元的信息保留和当前输入的信息记忆，自动学习并分配权重以控制数据依赖程度	具有长期依赖建模能力，缓解梯度消失和梯度爆炸；参数较多，计算成本高
LSTM-CRF[24]	LSTM 将输入序列映射到标签序列的概率分布，然后利用 CRF 优化整个标签序列概率	在 NER 任务中更好地处理标签依赖性和全局上下文；充分发挥 LSTM 上下文编码能力
BILSTM[22]	使用两个 LSTM 分别从相反方向对每个单词处理构建上下文相关表示，学习特征后利用 softmax 预测标签	解决长距离依赖问题，在少量标注数据上能大幅提高 NER 效果
BILSTM-CRF[24]	词嵌入作为模型输入，BILSTM 的输出为每个标签的预测分值，该分值作为 CRF 输入，在 CRF 损失函数中转移概率矩阵可学习到约束规则，使预测结构更准确	首次将该模型应用基准序列标记任务，有效融合过去和未来特征；但需要大量的特征工程，因此适用需要复杂特征工程的数据集；强调了模型对上下文的全面考虑
BILSTM-CNNS-CRF[30]	对于每个单词，字符级表示由 CNN 计算	实现了真正的端到端，无需繁琐的特征工程或数据预处理，适用于广泛的序列标记任务；无需额外手动处理步骤，简化工作流程，提高通用性
GRU-CRF[29]	使用 GRU 模型编码上下文信息，并用 CRF 预测标签。通过共享架构和参数进一步将模型扩展到多任务和跨语言联合培训	通过共享部分网络架构和参数，在多种语言 NER 任务上获得优异效果

3.3 基于 RNN 的对话状态追踪

2013 年，Henderson 等[31]首次将深度学习应用于 DST 任务，提出将 NLU 集成在 DST 中对所训练的模型进行联合建模。该模型使用 RNN，将每轮对话中的槽位—槽值对与用户文本作为输入，以提取特征。随后，该模型通过滑动窗口进行特征拼接，对每个槽值对打分，并选择最大值更新对话状态。此外，该模型还采用 Mask 机制，以提高模型泛化能力。

LSTM 作为 RNN 的一种变体，通过增加 4 个门控机制解决长距离依赖建模问题，在 DST 任务中也得到了广泛应用。Yoshino 等[32]使用了 Doc2vec 将自动语音识别（Automatic Speech Recognition，ASR）表示为句子向量，然后使用 LSTM 进行序列标注，以获得当前轮次结果。Ramadan 等[33]提出一种使用多个 BILSTM 共同追踪域和对话状态的模型，同时考虑了用户文本和本体标签之间的语义相似性，并允许跨域共享信息。

3.4 基于 RNN 的对话生成

LSTM 网络和 GRU 网络的相继应用，有效缓解了神经网络在处理具有长期和短期依赖性及梯度消失等训练问题方面的挑战，大幅度缩短了训练难度。然而，在 NLP 领域分支任务中，例如对话生成任务，会出现输入的源序列和生成的序列长短不一致的情况。由此，Sutskever 等[34]提出序列到序列（Sequence to Sequence，Seq2Seq）模型，其结构如图 6 所示。

从模型结构上看，Seq2Seq 是一种 RNN 结构，其核心部分由编码器和解码器（Encoder-Decoder）结构组成。其主要思想是通过 RNN 将源语句编码为向量，并通过另一个 RNN 将源语句解码为目标语句。

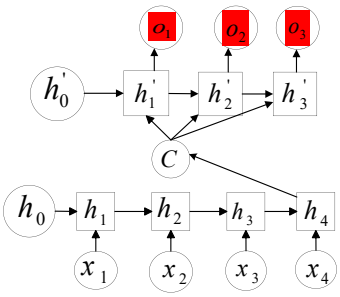

Fig. 6　Seq2Seq model

图 6　Seq2Seq 模型

Vinyals 等[35]将单轮对话看作一种 Seq2Seq 问题，并提出一种基于 Encoder-Decoder 的响应生成模型。在该模型中，首先使用 LSTM 对后响应进行编码，并将其嵌入作为另一 LSTM 的初始化状态以生成响应。然而，单轮对话形式不符合日常交流习惯，因为人类日常交流一般是包含多个回合的语句交流。由此，近些年的研究已经扩展到了多轮对话。Sutskever 等[18]提出乘法循环神经网络（Multiplicate Recurrent Neural Networks，MRNN）模型，该模型用于文本预测和生成，在该方法中，不同的输入字符使用不同的隐藏状态之间的转换函数。Bowman 等[36]研究了使用变分自编码器从连续语义空间生成句子，其中 RNN 用于编码器和解码器。这些研究表明，RNN 对各种不同结构的数据集上的文本生成工作非常有效。但这些工作都没有联系上下文信息。

Serban 等[37]提出基于 LSTM 的多轮会话响应选择模型，该模型采用 LSTM 对上下文信息和响应中的单词进行建模。类似地，Kadlec 等[38]使用时间卷积神经网络（Temporal Convolutinal Neural Networks，TCNN）模型和 BILSTM 模型以替代 LSTM。

3.5　RNN 小结

总体而言，RNN、LSTM 和 GRU 都是在对话系统中常用的神经网络架构。RNN 能够有效建模对话历史中上下文信息，有助于更好地理解用户需求和系统响应，但在训练深层 RNN 时容易出现梯度消失或梯度爆炸问题。此外，RNN 对短期的记忆影响较大，但对长期的记忆影响很小，因此无法处理很长的输入序列。

为此，LSTM 有效解决了 RNN 梯度消失和梯度爆炸问题，使得训练更加稳定。此外，LSTM 通过其内部记忆单元可以更好地捕获长期依赖关系，因此适合处理复杂的对话历史。但相对于普通 RNN，LSTM 的计算开销更高，通常需要大量的标记数据用来训练，这在人机对话系统领域会是一个挑战。

与 LSTM 类似，GRU 引入了门控机制，能够更好地处理梯度问题，并且相比 LSTM 拥有更少的参数，因此在资源受限情况下计算成本更低，更有优势。

4　基于注意力机制的对话系统

4.1　注意力机制原理

注意力机制最早被提出并在计算机视觉领域的神经网络中应用。Bahadanau 等[39]最早将注意力机制应用于 NLP 领域，在基于 Encoder-Decoder 架构的神经机器翻译模型中应用注意力机制改进 Decoder。传统 Encoder-Decoder 模型存在两方面缺陷：①Encoder 必须将所有输入信息压缩传递给 Decoder 的一个固定长度向量中，这将导致信息丢失；②该模型难以有效建模输入和输出序列间的对齐关系，而在文本翻译和文本摘要等结构化文本生成任务中，这种对齐关系非常重要。同时，Decoder 缺少机制以便有选择性地关注与生成每个输出词相关的输入。

基于注意力机制的 Encoder-Decoder 在解码过程中考虑 Encoder 对输入的每个单词的编码向量，并为每个单词赋一个权重值，形成权重分布。该权重分布使得模型有选择性地关注输入序列中的不同部分，从而更好地捕获输入序列和输出序列之间的关联信息。

但是，考虑到注意力机制因为采用 RNN 作为 Encoder 时难以编码长距离依赖关系，以及难以并行计算的限制，Vaswani 等[40]提出 Transformer 模型。Transformer 不依赖任何 RNN 或其变体，而是完全采用了自注意力机制（Self Attention Mechanism，SAM）[41]替代 RNN，从而克服了 RNN 在编码长距离依赖关系方面的困难。同时，Transformer 吸收 CNN 网络的思想，解决了 RNN 难以并行计算的缺陷。

BERT（Bidirectional Encoder Representations from Transformer，BERT）由 Google 公司提出，在神经机器翻译、问答系统、文本摘要等领域取得突出表现[42-43]。BERT 由多个 Transformer 模型的编码器（Encoder）连接而成，其具体结构如图 7 所示。

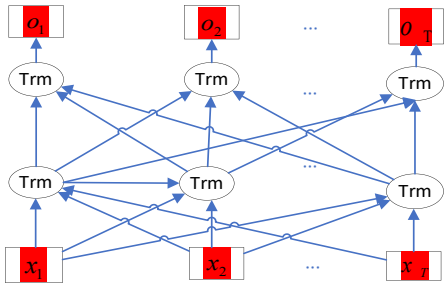

Fig. 7　BERT model

图 7　BERT 模型

其中，$\{x_1, x_2, \cdots, x_T\}$ 表示输入嵌入序列，$\{o_1, o_2, \cdots, o_T\}$ 表示通过 Transformer 编码器后对应的字向量输出序列，这里的嵌入由令牌嵌入、段嵌入和位置嵌入求和得来。

Transformer 中的 Decoder 层是用于生成目标序列的关键组件，其机制包括以下几个关键部分：①自注意力机制

(Self-Attention, SA):与 Encoder 中的自注意力机制类似,但有一些差异。在 Decoder 中,自注意力机制用于同时考虑已生成的目标序列部分和输入序列(通常是 Encoder 输出),以便模型能够理解上下文信息。这有助于确保生成的每个词语都能依赖之前生成的词语,并正确地翻译或生成目标序列。②多头注意力机制(Multi-Head Self-Attention, MHSA):Decoder 通常使用多个自注意力头,多头机制允许模型同时关注不同位置和语义信息,以更好地捕获序列的关系。③位置编码(Positional Encoding, PE):这些位置编码向量与词嵌入相加,以考虑词语在序列中的位置,因为 Transformer 模型没有内置的顺序信息。④层规范化(Layer Normalization, LN):在每个子层输出后,通常会应用层规范化以减少训练中的梯度问题和加速训练。⑤前馈神经网络(FeedForward Neural Network, FFN):用于进行特征映射和非线性变换,有助于捕捉输入和输出之间的复杂关系。⑥掩码机制(Masking):掩码机制能将未来位置的信息屏蔽掉,从而确保生成的目标序列不会包含未来信息,这有助于模型按顺序生成目标序列。这些机制的组合使得 Transformer 模型能够在各种 NLP 任务中表现出色,如文本生成和机器翻译等。

BERT 只包含了 Transformer 模型中的 Encoder 部分,是一种 Encoder-only Transformer 模型。与之相对的,Decoder-only Transformer 模型只包括 Transformer 模型中的 Decoder 部分,而没有 Encoder。Decoder-only Transformer 通常用于文本生成和自回归生成任务,其目标是生成连贯的文本序列。其中,最著名的 Decoder-only Transformer 模型之一是 GPT-3。

4.2 基于 Transformer 的信息抽取

基于自注意力机制的 Transformer 模型的提出成为近几年 NLP 领域最有影响力的研究之一。与上述 CNN、RNN 等模型相比,Transformer 具有以下潜在优势:①通过参数矩阵映射并循环进行多次权重分配操作,Transformer 能直接获取全局特征,尤其在需要考虑长距离依赖的共享情况下表现出色;②Transformer 引入了多头注意力机制,有助于更好地分辨命名实体的边界和识别实体类别;③基于 Transformer 的 NER 方法通常采用预训练的 Transformer 模型,如 BERT 以获得词嵌入;④当 Transformer 与 CRF 结合时能更好地建模标签之间的依赖关系,以确保 NER 输出标签序列连贯一致。

杨飘等[44]在中文 NER 任务中,基于 BERT 提出了 BERT-BIGRU-CRF 模型。该模型在 MSRA 中文语料上有良好效果,并且超过了 Zhang 等[45]提出的 LSTM 模型,其 F1 值达到 95.43%,比 LSTM 高出 2.25%。为解决中文 NER 任务中实体边界模糊、实体边界和非实体之间边界模糊等问题,胡叮叮等[46]提出了 BERT-BILSTM-CRF 模型。在该模型中,BERT 根据上下文信息生成每个词语的低维度稠密词向量,BILSTM 用于捕获时序特征,CRF 用于对输出标签顺序进行约束。实验结果表明,该模型下的 NER 在精准率、召回率和 F1 值 3 个方面取得了显著成效,分别达到了 89.12%、91.42% 和 90.25%。蔡翟源等[47]将 BERT 预训练模型应用于电子病历的关系语义实体识别,构建了 ALBERT-BILSTM-Attention-CRF 模型。其中,ALBERT 模型[48]在 BERT 模型基础上进行了 3 方面的改进:①嵌入的因式分解;②跨层参数共享;③句间连贯性损失。将 ALBERT-BILSTM-Attention-CRF 模型与 CNN-CRF、BILSTM-CRF 和 BERT-BILSTM-CRF 3 种模型实验结果进行比较,结果显示其 F1 值最高,达到了 96.3%。同时,实体的精准率和召回率均有所提升。

在关系抽取(Relation Extraction, RE)任务中,每个实体对关系的权重不同,因此注意力机制的优势在机器翻译、图像分类、语音识别等 NLP 应用中也得以体现。Zhou 等[49]在 Zhang 等[16]的基础上将注意力机制引入 BILSTM 模型,以捕获句子中更多信息部分,以提取实体关系。Yuan 等[50]提出一种特定关系注意力机制,通过为每种关系类型的上下文中单词分配不同的权重,以更准确地捕获不同关系的信息。Bekoulis 等[51]利用 RNN 和注意力机制,在不需要依赖解析树特征的情况下实现了实体关系抽取。在 Transformer 的基础上,Radford 等[52]提出生成式预训练模型(Generative Pre-trained Transformer, GPT),用于语言理解。但是 GPT 只能从左到右进行单词预测,仅单向捕获上下文信息,导致其在句子级别语料的文本表示效果不佳。为进一步改进这项工作,Devlin 等[53]提出 BERT 预训练模型,通过添加遮蔽任务和预测下一句的任务进行训练。Wang 等[54]通过使用 BERT 模型,只需对输入语料库进行一次编码即可提取命名实体关系。Lin 等[55]利用 BERT 模型和生物医学数据集上的监督训练,在临床时间关系提取方面取得了较好表现。此外,Kim 等[56]在 T5[57](Text-to-Text Transfer Transformer, T5)的基础上去除 Decoder 层的自注意力层,提出了 $T5_{slim\_dec}$ 模型,并选择 BERT 模型、GPT-3 模型和 T5 模型作为基准模型和 $T5_{slim\_dec}$ 模型进行对比实验,在生物医学领域的关系提取任务中取得了优异效果。这些研究都展示了注意力机制和预训练模型关系提取任务的重要性和有效性。

4.3 基于 Transformer 的对话状态追踪

一方面,注意力机制允许模型动态地关注对话历史中的不同部分,以更好地理解上下文,这对于状态追踪非常有用,因为某些对话历史可能对于当前状态的更新更为重要;另一方面,多头注意力机制使得模型能够同时关注多个不同方面的信息,有助于更全面地捕获上下文信息,并提高 DST 性能。

Jang[58]利用注意力机制和 BILSTM 的 DST 模型对对话进行状态更新,在 DSTC5 数据集上取得较高精度。GLAD 模型[59]使用基于注意力机制的 RNN 模型,用于学习全局追踪器,而全局追踪器和局部追踪器在追踪槽时共享参

数。2019 年，Wu 等[60]提出了 TRADE 模型，该模型基于 Encoder-Decoder 结构，包括一个对话文本编码器、槽门和状态生成器。TRADE 模型将系统和用户对话文本拼接编码为上下文向量，然后在状态生成器中使用复制机制为每个槽位逐词生成槽值。该模型充分考虑了上下文信息和槽位的动态性，使其在 DST 任务中表现出色。

BERT 作为一种预训练的 Transformer 模型，在对话状态追踪任务中也有着良好表现。其主要应用方式有：①对话历史表示：将多轮对话历史（包括用户的输入和系统的回复）转换为 BERT 可以理解的格式；②特殊标记：为了告诉 BERT 哪些部分是对话历史，哪些部分是特定的任务标签，通常需要添加一些特殊标记，如[CLS]（用于表示序列的开始）和[SEP]（用于分隔不同的文本段落或句子）；③任务标签：为对话状态追踪任务定义合适的任务标签，如用户意图、槽位值等，以便 BERT 可以在训练时使用这些标签，从而更好地捕获对话状态信息。

BERT 在对话状态追踪中的应用受益于其预训练的能力，它可以捕获丰富的语言表示，并可以用于不同的 NLP 任务。然而，在对话状态追踪中，BERT 在训练下游任务时往往会选择模型的最后一层进行分类，这种做法往往忽略了 BERT 其他层包含的语义信息。叶正等[61]在 BERT 微调策略方法上进行了研究，探讨了 BERT 层数对任务结果的影响，他们通过引入注意力机制对拼接后 BERT 的 12 层输出的特征权重进行微调。实验表明，该方法提高了语义信息的特征表达能力，相比仅在最后一层进行分类的 BERT 模型，取得了更好的任务效果。

4.4　基于注意力机制的对话生成

Shang 等[62]利用注意力机制优化了 Encoder-Decoder 模型。在神经机器翻译、生成式文本摘要、语音识别等生成式任务中，通常将注意力机制作为连接 Encoder 和 Decoder 的桥梁，以确保 Decoder 在生成每个词项时都可以参考源序列中最相关的部分。已有多项研究证实了在生成式任务中注意力机制的不可或缺性。Devlin 等[53]提出的 BERT 模型采用多头自注意力网络作为特征提取器，在包括序列标注在内的 11 项 NLP 任务中取得了当时的最佳成绩。Zhou 等[63]在 Trasformer 基础上，提出了深度注意匹配模型（Deep Attention Matching，DAM）模型。该模型利用注意力和交叉注意力机制提取从字级别到句子级别的上下文信息和响应之间的匹配信息；然后，DAM 将匹配信息汇总到一个 3D 匹配图中，经过卷积和池化进一步提取匹配信息，最后通过单层感知机计算匹配得分。DAM 模型打破了 RNN 和 CNN 结构的限制，在多轮对话任务中速度快，达到了目前最好的效果。Liu 等[64]提出一种基于 Transformer 模型的多模态注入对话系统（Transformer-based Multimodal Infusion Dialogue，TMID），该系统通过多模态上下文 Encoder 从对话中提取视觉或文本信息，并借助交叉注意力机制实现每个话语的图像和文本之间的信息输入。此外，TMID 还采用了自适应的 Decoder，根据状态分离器确定的用户意图生成适当的多模态响应，并将相关领域知识融合到 Decoder 中，以丰富生成的响应内容。

与普通的对话系统对比，基于角色的对话系统是一种人工智能对话系统，它模拟了不同角色之间的交流方式，这种对话系统通常用于娱乐、虚拟助手、教育等领域。Zheng 等[65]提出一种注意力路由结构，该结构有助于从人物信息中获得权重以生成响应。基于角色的对话系统旨在生成与历史背景和预定角色一致的响应。Huang 等[66]提出一种有效的角色适应性注意力框架（Persona-Adaptive Attention，PAA），该框架通过设定的注意力机制自适应地整合了人物和上下文信息的权重。与从外部预测器中计算权重的方法不同，PAA 在框架内计算权重，再对加权的交叉注意结果应用掩蔽，以减轻训练难度[67]。此外，PAA 方法在低资源环境下，使用 20%~30% 的数据训练模型，仍能够达到与全数据训练模型相媲美的出色效果。

相较于 BERT，改进的训练 BERT 模型的方法（Robustly Optimized BERT Approach，RoBERT，RoBERTa）拥有更长的预训练时间、更大的训练数据集、动态掩码和更小的批次大小等。卢幸等[68]针对对话模型缺乏情感感知和预测能力问题，提出基于 RoBERTa 的多轮对话情感识别模型。他们使用 RoBERTa 作为 Encoder 层进行文本特征提取，得到文本信息编码向量。并且，将每轮对话[CLS]字符对应的编码向量按照时间顺序输入 GRU，最后通过融合模块和分类模块得到最终情感概率标签向量。该模型在情感识别任务中有着更优秀的性能。

在文本生成领域，Li 等[69]提出了 TD-NHG（Transformer Decoder-News Headline Generation，TD-NHG）模型，用于生成新闻标题。TD-NHG 模型 12 个 Transformer Decoder 层构成，是一种自回归模型，该模型使用掩蔽式多头自注意力机制学习新闻文本中不同表示子空间的特征信息。在解码阶段，TD-NHG 模型采用 top-k、top-p 和惩罚机制的解码选择策略。通过在 LCSTS 和 CSTS 数据集上与基本模型如 RNN 和 LSTM 进行实验对比，TD-NHG 模型证明了其能够提高新闻标题的准确性和多样性。此外，TD-NHG 模型还在对话系统中表现出较好的性能，能够较好满足用户提出的文本摘要等任务需求。

4.5　注意力机制小结

总体而言，注意力机制是深度学习中的一项关键技术，相较于 CNN、RNN 等模型，注意力机制通过不同的权重有效处理长距离依赖关系并捕捉全局信息。此外，注意力机制还实现上下文感知，根据上下文自动调整关注内容，在处理复杂语境时很有帮助。Transformer 作为基于注意力机制的具体模型架构，极大改进了序列建模的效率和性能，成为 NLP 领域的一个重要里程碑。

5 问题与展望

随着深度学习人机对话的发展，越来越多优秀高效的模型被提出，但仍存在不少问题待解决。本文从多方面概述所面临的问题，并提出可行的解决方案。

（1）数据需求。深度学习模型通常需要大量的标记数据用于训练，但在人机对话领域获取大规模标记数据可能会非常昂贵且耗时。此外，NER 任务需要某些特定专业知识，例如电子病历涉及疾病、诊断和解剖等实体，识别这实体及其关系需要相关领域专家，手工生成足够的标记数据极其困难。针对该问题，可考虑利用半监督学习方法，以减少对大规模标记数据的依赖。半监督学习和主动学习同属于弱监督学习方法中的不完全监督学习，其结构如图 8 所示。

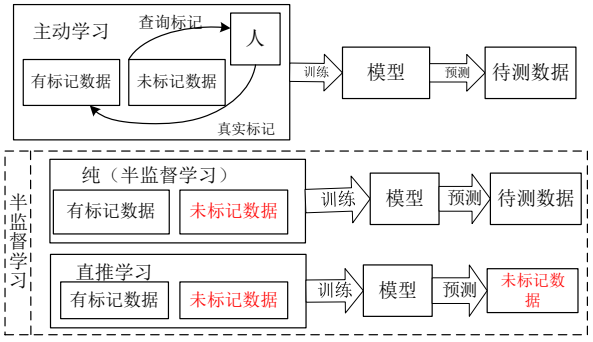

Fig. 8 Incomplete supervision structure

图 8 不完全监督结构

主动学习策略先训练有标记数据，然后根据经验对未标记数据进行聚类，自动选择最需要标记的对话样本进行标注，从而达到最小化查询数量以提高数据利用率。与主动学习相比，半监督学习没有人类干预。半监督学习在减少对标记数据依赖方面具有以下优势：①允许同时使用有标记数据和未标记数据进行训练，它将未标记数据的预测结果作为伪标签，再将伪标签数据和有标记标签数据一起训练，随后依次迭代这一过程，将高置信度的预测结果添加到有标记数据，以此扩大标记数据；②由于未标记数量大于有标记数据，半监督学习有利于防止过拟合问题，未标记数据的引入将提供更多训练样本，从而提高了模型鲁棒性。

（2）领域适应问题。构建一个通用领域的对话系统相对容易，但要在其他特定领域中有较好表现，需进行额外的调节适应。针对该问题，可考虑使用迁移学习技术，迁移学习通过将源对话领域中模型训练得到的参数和表示等信息应用到目标对话领域，从而提高目标对话领域性能。大体流程如图 9 所示。迁移学习的优势主要如下：①通过在不同领域上提取共享特征并迁移使得模型能更好地处理未见过的数据；②共享特征的迁移减少了模型对标记数据的需求，降低了训练成本；③迁移学习通过利用源对话领域的共享特征，解决了特定目标领域标记数据匮乏而引发的冷启动问题。

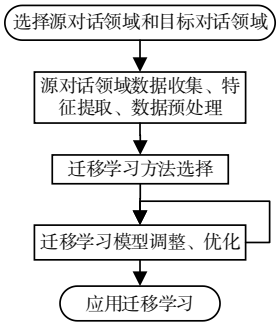

Fig. 9 Transfer learning process

图 9 迁移学习流程

（3）伦理和隐私问题。伦理和隐私问题是与人机对话系统相关的重要考虑因素。这些问题包括计算机生成歧视性回复或将敏感信息泄露给特定人群。针对该问题，需采取一系列伦理规则和技术措施。

伦理规则方面，实施伦理准则、监管和审查机制，加强科研人员伦理意识培养，确保人机对话不滥用、不偏见。隐私方面，可以引入隐私增强技术，如差分隐私，以减少敏感信息的泄露风险。人机对话差分隐私保护流程如图 10 所示，其中对话信息收集过程最小化收集和存储敏感信息，仅收集和保留必要信息，并删除不再需要的数据以降低泄露风险。噪声注入是差分隐私核心内容，在信息收集过程中引入随机性扰动或使用差分隐私算法进行数据处理。数据处理分析和文本输出过程同样使用差分隐私技术来保护用户的隐私。

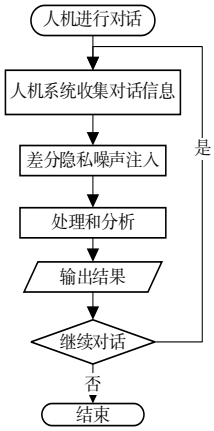

Fig. 10 Differential privacy protection process for human-machine dialogue

图 10 人机对话差分隐私保护流程

6 结语

本文对基于深度学习的对话系统技术进行了综述，并

将其归纳为三大类方法,认为人机对话具有广阔应用前景。随着深度学习技术的不断成熟和经验的不断积累,目前对话任务存在的问题都会被解决,深度学习在人机对话领域的应用将会更加完善。基于深度学习的对话系统研究的意义在于改善了人们的生活、工作和社会交流,推动了技术进步,并为各领域提供了创新的机会。随着人工智能和自然语言处理技术的不断发展,人机对话将继续发挥更大的作用。

参考文献:

[1] LIU K. A survey on neural relation extraction[J]. Science China(Technological Sciences),2020,63(10):1971-1989.

[2] SANTOS C N D,GUIMARÃES V. Boosting named entity recognition with neural character embeddings[DB/OL]. http://arxiv. org/abs/1505. 05008v2,2015.

[3] STRUBELL E, VERGA P, BELANGER D, et al. Fast and accurate entity recognition with iterated dilated convolutions[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017.

[4] ZHENG S C, HAO Y X, LU D Y, et al. Joint entity and relation extraction based on a hybrid neural network[J]. Neural Computing, 2017, 257: 59-66.

[5] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1753-1762.

[6] DAI J H, PENG R Y, XU L, et al. A survey of information extraction based on deep neural networks[J]. Journal of Southwest China Normal University(Natural Science Edition),2022,47(4): 1-11.
代建华, 彭若瑶, 许路,等. 基于深度神经网络的信息抽取研究综述[J]. 西南师范大学学报(自然科学版),2022,47(4):1-11.

[7] ZHANG Y, LIU Z, ZHOU W. Event recognition based on deep learning in Chinese texts[J]. Plos One,2016, 11(8):e0160147.

[8] CHEN Y B, XU L H, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]//Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on NLP,2015: 5916-5923.

[9] LEE S. Structured discriminative model for dialog state tracking[C]//Proceedings of the SIGDIAL 2013 Conference,2013:442-451.

[10] SHI H, USHIO T, ENDO M, et al. A multichannel convolutional neural network for cross-language dialog state tracking[C]//San Diego: 2016 IEEE Spoken Language Technology Workshop,2016:559-564.

[11] KORPUSIK M, GLASS J. Dialogue state tracking with convolutional semantic taggers[C]//Brighton: 2019 IEEE International Conference on Acoustics, Speech and Signal Processing,2019:7220-7224.

[12] LI H, WANG M, LIU Q, et al. Syntax-based deep matching of short texts[C]//Proceedings of the 24th International Conference on Artificial Intelligence,2015:1354-1361.

[13] LU Z, LI H. A deep architecture for matching short texts[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems,2013:1367-1375.

[14] HU B T, LU Z D, LI H, et al. Convolutional neural network architectures for matching natural language sentences[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems, 2015:2042-2050.

[15] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning (Vol. 1)[M]. Cambridge: MIT Press,2016.

[16] ZHANG D,WANG D. Relation classification via recurrent neural network[DB/OL]. https://arxiv. org/abs/1508. 01006, 2015.

[17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation,1997,9(8):1735-1780.

[18] SUTSKEVER I,MARTENS J,HITON G E. Generating text with recurrent neural networks[C]//Proceedings of the 28th International Conference on International Conference on Machine Learning, 2016:1017-1024.

[19] DENG Q,CHEN X L,CHEN L Q. Named entity recognition based on the LSTM-SNP[J]. Journal of Xihua University(Natural Science Edition), 2023,42(5):28-37.
邓琴,陈晓亮,陈龙齐. 基于LSTM-SNP的命名实体识别[J]. 西华大学学报(自然科学版),2023,42(5):28-37.

[20] PĂUN G. Computing with membranes[J]. Journal of Computer and System Sciences,2000,61(1):108-143.

[21] XU Y, LI M, GE L, et al. Classifying relations via long short term memory networks along shortest dependency paths[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015:1785-1794.

[22] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks,2005,18(5-6):602-610.

[23] ZHANG S, ZHENG D, HU X, et al. Bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 29th Pacific Aisa Conference on Language, Information and Computation, 2015: 73-78.

[24] HUANG Z,XU W,YU K. Bidirectional LSTM-CRF models for sequence tagging[DB/OL]. https://arxiv. org/abs/1508. 01991,2015.

[25] LAMPLE G,BALLESTEROS M,SUBRAMANIAN S, et al. Neural architectures for named entity recognition[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,2016:260-270.

[26] OUYANG E, LI Y, JIN L, et al. Exploring N-gram character presentation in bidirectional RNN-CRF for Chinese clinical named entity recognition[C]//Chengdu: CCKS 2017-China Conference on Knowledge Graph and Semantic Computing,2017.

[27] GAO X,WANG S,ZHU J W,et al. Overview of named entity recognition tasks[J]. Computer Science, 2023,50(S1): 26-33.
高翔,王石,朱俊武,等. 命名实体识别任务综述[J]. 计算机科学, 2023,50(S1): 26-33.

[28] CHO K, BART VAN M, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[DB/OL]. https://arxiv. org/abs/1406. 1078,2014.

[29] YANG Z, SALAKHUTDINOV R, COHEN W. Multi-task cross-lingual sequence tagging from scratch [DB/OL]. https://arxiv. org/abs/1603. 06270v2,2016.

[30] MA X,HOVY E. End-to-end sequence labeling via bidirectional LSTM-CNNs-CRF[DB/OL]. https://arxiv. org/abs/1603. 01354v5, 2016.

[31] HENDERSON M, THOMSON B, YOUNG S. Deep neural network approach for the dialog state tracking challenge[C]//Proceedings of the SIGDIAL 2013 Conference,2013:467-471.

[32] YOSHINO K, HIRAOKA T, NEUBIG G, et al. Dialogue state tracking using long short term memory neural networks[C]//Tokyo: 2018 IEEE 42nd Annual Computer Software and Applications Conference,2016.

[33] RAMADAN O, BUDZIANOWSKI P, GASIC M. Large-scale multi-domain belief tracking with knowledge sharing[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 432-437.

[34] SUTSKEVER I,VINYALS O, LE Q. Sequence to sequence learning with neural networks[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems,2014:3104-3112.

[35] VINYALS O, LE Q. A neural conversational model[DB/OL]. https://arxiv. org/abs/1506. 05869,2015.

[36] BOWMAN S R, VILNIS L, VINYALS O, et al. Generating sentences from a continuous space[C]//Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning,2015:10-21.

[37] SERBAN I V,LOWE R,HENDERSON P,et al. A survey of available corpora for building data-driven dialogue systems[J]. Computer Science, 2017,33(16):6078-6093.

[38] KADLEC R,SCHMID M,KLEINDIENST J. Improved deep learning baselines for ubuntu corpus dialogs[DB/OL]. //arxiv. org/abs/1510. 03753, 2015.

[39] BAHDANAU D,CHO K,BENGIO Y. Neural machine translation by jointly learning to align and translate [DB/OL]. https://arxiv. org/abs/1409. 0473,2015.

[40] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems,2017:6000-6010.

[41] HU D. An introductory survey on attention mechanisms in NLP problems [C]//Proceedings of the 2019 Intelligent Systems Conference, 2020: 432-448.

[42] KENTON J D M W C, TOUTANOVA L K. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2019: 4171-4186.

[43] KAUR K,KAUR P. Improving BERT model for requirements classification by bidirectional LSTM-CNN deep model[J]. Computers & Electrical Engineering,2023,108:108699.

[44] YANG P, DONG W Y. Chinese named entity recognition method based on BERT embedding[J]. Computer Engineering,2020,46(4):40-45.
杨飘, 董文永. 基于BERT嵌入的中文命名实体识别方法[J]. 计算机工程, 2020,46(4):: 40-45.

[45] ZHANG Y, YANG J. Chinese NER using lattice LSTM[DB/OL]. http://arxiv. org/abs/1805. 02023v3, 2018.

[46] HU D D, ZHANG C, WANG Z Y. Research on Named Entity Recognition Based on Pre-training Model[J]. Modern Information Technology, 2023,7(15): 78-82.
胡叮叮,张琛,王之原. 基于预训练模型的命名实体识别研究[J]. 现代信息科技,2023,7(15): 78-82.

[47] CAI Z Y, CHEN J, XI X F. Relational semantic entity recognition for electronic medical record[J]. Journal of Suzhou University of Science and Technology(Natural Science Edition),2023,40(3): 62-70.
蔡翟源,陈杰,奚雪峰,等. 电子病历的关系语义实体识别[J]. 苏州科技大学学报(自然科学版), 2023,40(3):62-70.

[48] LAN Z, CHEN M, GOODMAN S, et al. ALBERT: a lite BERT for self-supervised learning of language representations[DB/OL]. https://arxiv. org/abs/1909. 11942,2019.

[49] ZHOU P,SHI W,TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016:207-212.

[50] YUAN Y, ZHOU X, PAN S, et al. A relation-specific attention network for joint entity and relation extraction[C]//Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence,2020:4054-4060.

[51] BEKOULIS G, DELEU J, DEMEESTER T, et al. An attentive neural architecture for joint segmentation and parsing and its application to real estate ads[J]. Expert Systems with Applications,2018,102:100-112.

[52] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[R]. Technical Report, 2018.

[53] DEVLIN J,CHANG M W,LEE K,et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.

[54] WANG H, TAN M, YU M, et al. Extracting multiple-relations in one-pass with pre-trained transformers [DB/OL]. https://arxiv. org/abs/1902. 01030v2, 2019.

[55] LIN C, MILLER T, DLIGACH D, et al. A bert-based universal model for both within-and cross-sentence clinical temporal relation extraction [C]//Proceedings of the 2nd Clinical NLP Workshop,2019: 65-71.

[56] KIM S, YOON J, KWON O. Biomedical relation extraction using dependency graph and decoder-enhanced transformer model[J]. Bioengineering (Basel),2023,10(5):586.

[57] RAFFEL C, SHAZEER N, REBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. The Journal of Machine Learning Research,2020,21(1): 5485-5551.

[58] JANG Y, HAM J, LEE B J, et al. Neural dialog state tracker for large ontologies by attention mechanism[C]//San Diego:2016 IEEE Spoken Language Technology Workshop (SLT),2016.

[59] ZHONG V, XIONG C, SOCHER R. Global-locally self-attentive dialogue state tracker[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018:1458-1467.

[60] WU C S, MADOTTO A, HOSSEINI A 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,2019:808-819.

[61] YE Z, FU L, QIN J, et al. Dialogue state tracking based on fine-tuning strategy using BERT information at different layer[J]. Journal of South-Central Minzu University (Natural Science Edition), 2023, 42 (3): 327-333.
叶正,傅灵,覃俊,等. 基于利用BERT不同层信息的微调策略的对话状态追踪[J]. 中南民族大学学报 (自然科学版),2023,42(3): 327-333.

[62] SHANG L,LU Z, LI H. Neural responding machine for short-text conversation[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing,2015:1577-1586.

[63] ZHOU X Y,LI L, DONG D X, et al. Multi-turn response selection for chatbots with deep attention matching network[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018:1118-1127.

[64] LIU B, HE L, LIU Y,et al. Transformer-based multimodal infusion dialogue systems[J]. Electronics,2022,11(20):3409.

[65] ZHENG Y, ZHANG R, HUANG M, et al. A pre-training based personalized dialogue generation model with persona-sparse data[C]//Proceedings of the AAAI Conference on Artificial Intelligence,2020:9693-9700.

[66] HUANG Q S,ZHANG Y,KO T,et al. Personalized dialogue genereation with persona-adaptive attention [DB/OL]. https://arxiv. org/abs/2210. 15088,2022.

[67] LIU Y, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized bert pretraing approach[DB/OL]. https://arxiv. org/abs/1907. 11692,2019.

[68] LU X. Deep learning-based emotional dialogue generation research[D]. Changchun:Changchun University of Technology,2023.
卢幸. 基于深度学习的情感对话生成研究[D]. 长春:长春工业大学, 2023.

[69] LI Z,WU J, MIAO J,et al. News headline generation based on improved decoder from transformer[J]. Scientific Reports,2022,12:11648.

(责任编辑:孙 娟)