控制与决策
Control and Decision

循环神经网络研究综述

刘建伟, 宋志妍

引用本文:
刘建伟,宋志妍. 循环神经网络研究综述[J]. 控制与决策, 2022, 37(11): 2753−2768.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2021.1241

循环神经网络研究综述

刘建伟†, 宋志妍

(中国石油大学(北京) 信息科学与工程学院，北京 102249)

摘 要: 循环神经网络是神经网络序列模型的主要实现形式,近几年得到迅速发展,其是机器翻译、机器问题回答、序列视频分析的标准处理手段,也是对于手写体自动合成、语音处理和图像生成等问题的主流建模手段. 鉴于此,循环神经网络的各分支按照网络结构进行详细分类,大致分为3大类:一是衍生循环神经网络,这类网络是基于基本RNNs模型的结构衍生变体,即对RNNs的内部结构进行修改;二是组合循环神经网络,这类网络将其他一些经典的网络模型或结构与第一类衍生循环神经网络进行组合,得到更好的模型效果,是一种非常有效的手段;三是混合循环神经网络,这类网络模型既有不同网络模型的组合,又在RNNs内部结构上进行修改,是同属于前两类网络分类的结构. 为了更加深入地理解循环神经网络,进一步介绍与循环神经网络经常混为一谈的递归神经网络结构以及递归神经网络与循环神经网络的区别和联系. 在详略描述上述模型的应用背景、网络结构以及模型变种后,对各个模型的特点进行总结和比较,并对循环神经网络模型进行展望和总结.

关键词: 循环神经网络；衍生循环神经网络；组合循环神经网络；混合循环神经网络

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2021.1241

引用格式: 刘建伟,宋志妍. 循环神经网络研究综述[J]. 控制与决策, 2022, 37(11): 2753-2768.

Overview of recurrent neural networks

LIU Jian-wei†, SONG Zhi-yan

(College of Information Science and Engineering，University of Petroleum China (Beijing)，Beijing 102249，China)

Abstract: Recurrent neural networks (RNNs) are the main implementation paradigms for deep neural network sequence model, and have been developed rapidly and widespreadly in last two decades. Now, RNNs are cornerstone and foundation underpinning for machine translation, machine question answering and sequence video analysis, and RNNs are also the mainstream modeling approaches for handwriting automatic synthesis, speech processing and image generation. In this paper, the branches of recurrent neural networks are classified in detail according to the network structure, which can be roughly divided into three categories: the first one is all sorts of variants of recurrent neural networks, which are structural variants based on the basic RNNs architecture, that is, modifying the internal structure of RNNs. The second kind is combined RNNs, which combine some classical other network models or structures with the first kind of RNNs to get better modeling effect. It is a very effective means. The third one is hybrid RNNs, which not only combine different network models, but also modify the internal structure of RNNs. In order to understand the RNNs more deeply, this paper also introduces the structure of recursive neural networks which are often confused with RNNs, and the difference and connection between recursive neural networks and RNNs. After a detailed description of the application background, network structure and model variants of the above models based on RNNs, the characteristics of each model are summarized and compared. Finally, the prospect and summary of the RNNs are given.

Keywords: recurrent neural networks；derived recurrent neural networks；combined recurrent neural networks；hybrid recurrent neural networks

0 引 言

人工神经网络(artificial neural networks, ANNs)是由单个神经元或节点组成的网络,神经元分层后通过具有权值的连接边相连,表示神经元之间的相互作用关系,称为人工神经元. 浅层神经网络是指有一个输入层、一个输出层和最多一个隐层的神经网络,因此网络中没有循环连接. 随着层数的增加,网络的复杂度也随之增加,更多的层或循环连接会增加网络的深度,并使其能够提供不同层次的数据表示和特征提取,称为深度学习. 一般而言,这些网络由非线性

收稿日期: 2021-07-16；**录用日期:** 2021-11-12.

基金项目: 中国石油大学(北京)科学基金项目(2462020YXZZ023).

†通讯作者. E-mail: liujw@cup.edu.cn.

但很简单的单元组成,其中较高的层提供了更抽象的数据表示,但各层非线性的组合会带来优化和训练困难,导致对于深层网络体系结构的研究并不多. 2006年,Hinton等[1]提出基于深度信念网络可以使用非监督的逐层贪心训练算法,这为训练深度神经网络带来了希望,自此深度学习进入发展热潮,并从边缘学科变为主流科学技术. 在深度学习领域中,具有循环连接的神经网络称为循环神经网络(recurrent neural networks, RNNs),其能够为序列的识别和预测建模序列数据,使用循环迭代函数存储信息,很好地捕捉上下文信息,实现暂态依赖关系学习. 由于RNNs良好的扩展性和广泛的应用前景,近年来受到众多国内外学者的青睐,取得了较多的研究成果. 夏瑜潞[2]和杨丽等[3]发表的循环神经网络综述仅详细介绍了基本RNNs和最基础的几个变体结构,其他衍生变体结构以及与很多经典网络模型的结合并没有提及,模型的总结和分析不够全面.

鉴于此,本文对基于RNNs的各类扩展模型进行系统结构综述,通过模型的应用背景、结构以及各自的特点进行较为全面的总结. 注意到,ANNs还有一类递归神经网络(recursive neural networks, RecursiveNNs),它对结构化的输入进行操作,递归地应用变换来生成数据表示,并没有循环连接. 但是,由于两个网络首字母缩写相同,人们经常将这两种网络混为一谈. 于是,为了更深入地理解RNNs,本文后续还会介绍RecursiveNNs的网络结构以及两个网络的本质区别和联系.

1 RNNs基本原理

RNNs是传统前馈神经网络的扩展,能够处理可变长度的序列输入,它通过内部的循环隐变量学习可变长度输入序列的隐表示,隐变量每一时刻的激活函数输出都依赖于前一时刻循环隐变量激活函数的输出[4]. 给定一个输入序列 $\boldsymbol{x}=(x_1,x_2,\dots,x_T)$,RNNs隐变量的循环更新过程如下:

$$\boldsymbol{h}_t=g(W\boldsymbol{x}_t+U\boldsymbol{h}_{t-1}). \tag{1}$$

其中:g为一个激活函数(如logistics sigmoid函数或双曲正切函数),W为输入到这一时刻隐变量的权重矩阵,U为上一时刻隐变量到这一时刻隐变量的权重矩阵. 在给定当前隐状态$\boldsymbol{h}_t$的情况下,RNNs可以用来表示输入序列上的联合概率分布,即用生成式模型的观点解释RNNs的更新过程:每一个时刻的更新公式生成一个条件概率分布,由所有时刻条件概率分布的乘积得到联合概率分布. RNNs引入特殊的终止符号来探知可变长度序列的结束位置,RNNs可以很自然地表示可变长度序列上的概率分布.

2 衍生RNNs

本节介绍以RNNs为基础衍生出的结构变体,即对RNNs内部的结构进行更改,包括:双向RNNs、长短期记忆网络、微分RNNs、高速公路网、多维RNNs和嵌套堆叠RNNs. 下面对各类变体的模型背景、模型结构以及模型优点进行详细描述,并纵向和横向比较模型之间的区别.

2.1 双向RNNs

RNNs的一个缺点是只能利用先前的上下文信息,在语音识别中,一次性要转录整段话,所以必须利用未来的语境. Schuster等[5]提出的双向RNNs(bidirectional recurrent neural networks, BRNNs)用两个单独的隐层处理两个方向上的数据,充分利用了未来的语境,并且将这些隐层输出馈送到同一输出层以实现分类,提高语音识别的准确率.

2.2 长短期记忆网络

2.2.1 模 型

虽然RNNs在理论上是一个简单而强大的模型,但在实践中却很难得到较好的训练,主要原因之一是Razvan等[6]描述的梯度消失和梯度爆炸问题. 梯度爆炸问题是指训练期间由于长期分量的爆炸引起梯度范数大幅增加,这些分量的增长速度是短期分量的指数倍. 梯度消失问题指的是相反的过程,当以指数形式增长的长期分量的范数快速趋近于0时,模型无法学习大范围事件之间的依赖关系[7].

长短期记忆神经网络(long short term memory, LSTM)由Hochreiter & Schmidhuber提出,随后被Graves[8]进行了改进和推广,在很多问题上,LSTM都取得了巨大的成功.

假设LSTM第j-th单元在时刻t的记忆单元为$\boldsymbol{c}_t^j$,那么LSTM单元的隐状态输出$\boldsymbol{h}_t^j$[1]表示为

$$\boldsymbol{h}_t^j=\boldsymbol{o}_t^j\tanh(\boldsymbol{c}_t^j), \tag{2}$$

其中$\boldsymbol{o}_t^j$为输出门,用来调节记忆单元输出内容的数量. 输出门计算过程如下:

$$\boldsymbol{o}_t^j=\sigma(W_o\boldsymbol{x}_t+U_o\boldsymbol{h}_{t-1}+V_o\boldsymbol{c}_t)^j. \tag{3}$$

其中:θ为logistic sigmoid函数,V_o为一个对角矩阵. 记忆单元不仅要生成隐状态,还要进行循环更新以便计算下一时刻的隐状态,于是通过部分遗忘现存的记忆同时增加一个新的中间记忆单元$\tilde{\boldsymbol{c}}_t^j$来更新记忆单元$\boldsymbol{c}_t{}^j$,有

$$\boldsymbol{c}_t^j=\boldsymbol{f}_t^j\,\boldsymbol{c}_{t-1}^j+\boldsymbol{i}_t^j\tilde{\boldsymbol{c}}_t^j. \tag{4}$$

中间记忆单元由下式得到:

$$\tilde{\boldsymbol{c}}_t^j = \tanh(W_c \boldsymbol{x}_t + U_c \boldsymbol{h}_{t-1})^j. \tag{5}$$

遗忘门 $\boldsymbol{f}_t^j$ 的作用是调节现存记忆的遗忘程度, 输入门 $\boldsymbol{i}_t^j$ 的作用是调节增加内容到新记忆单元的程度. 遗忘门和输入门计算为

$$\boldsymbol{f}_t^j = \sigma(W_f \boldsymbol{x}_t + U_f \boldsymbol{h}_{t-1} + V_f \boldsymbol{c}_{t-1})^j, \tag{6}$$

$$\boldsymbol{i}_t^j = \sigma(W_i \boldsymbol{x}_t + U_i \boldsymbol{h}_{t-1} + V_i \boldsymbol{c}_{t-1})^j, \tag{7}$$

其中 V_f 和 V_i 均为对角矩阵. 与在每个时刻覆盖以前时刻状态的传统循环单元不同, LSTM 单元通过引入门决定是否保持现有记忆. 直观地说, 如果 LSTM 单元能够在早期阶段从输入序列中检测到重要特征, 则 LSTM 可在长时间内保留该特征信息, 从而捕获潜在的大范围序列依赖关系.

Cho 等[9] 提出了一种门限循环单元 (gated recurrent unit, GRU), 该模型使每个循环单元能够自适应地捕捉不同时间尺度的依赖关系, 广泛应用于序列建模中. 与 LSTM 单元类似, GRU 用两个门限单元调节单元内部的信息流, 但是没有单独的记忆单元. 更新门决定单元更新其内容的程度, 复位门决定遗忘其之前隐状态的程度. GRU 由于模型相对简单, 更适用于构建较大的网络. 从计算角度看, 由于只有两个门控, 该模型的效率更高, 可以节约计算成本.

GRU 功能强大, 但序列中有缺失值[10] 时只能利用均值法、正向插补或将前两个方法处理得到的输入、 掩码 (masking) 向量和时间间隔向量串联作为 GRU 的输入来填补缺失值, 但是用均值法或正向插补法对缺失值进行插补并不能区分缺失值是插补的还是真实观测的, 而简单地串联掩码向量和时间间隔向量则无法利用缺失值的时间结构, 因此 Che 等[11] 提出了 GRU-D 模型, 有效利用掩码和时间间隔这两种信息丢失的表现形式对 GRU 进行扩展以有效处理缺失值的问题. 掩码通知模型哪些输入被观察到 (或丢失), 时间间隔封装输入的观察模式. 该模型通过对 GRU 的输入和隐状态应用掩码和时间间隔 (使用衰减项) 捕获观测值及其依赖性, 引入衰变机制, 通过衰变率控制衰变机制, 并使用反向传播联合训练所有模型组件. 由此, 该模型不仅能捕捉到时间序列观测值的长期时间依赖性, 而且能够利用缺失模式改善预测结果.

Yao 等[12] 扩展了 LSTM, 使用深度门连接相邻层的记忆单元, 在下层与上层循环单元之间引入一种线性依赖关系应用于语言建模和机器翻译. 该模型具有深度门, 深度门连接上层中的记忆单元和下层中的记忆单元, 用于控制从下层记忆单元直接到上层记忆单元的流量.

由于 LSTM 和乘法 RNN[13](multiplicative RNN, mRNN) 结构存在互补性, Ben 等[14] 提出乘法 LSTM (multiplicative LSTM, mLSTM), 该模型结合了 mRNN 中隐状态的元素级更新控制与 LSTM 门限框架的特点. mLSTM 结构在 mRNN 的隐状态与 LSTM 中的每个门限单元之间添加连接, 目标是将 mRNN 上一时刻的隐状态输出用作这一时刻输入的权值以更新隐状态, 并与 LSTM 的大时滞和信息流控制相结合. 这样, LSTM 的门限单元可以更容易地控制大范围的暂态依赖关系.

RNNs 在基于骨架的人类行为识别方面取得了较好的结果[15-16]. 不同于现有的将关节信息串联起来作为整体表示的方法, Liu 等[17] 通过发现不同身体关节之间的空间依赖关系, 将循环结构扩展到空间域, 提出了一种时空 LSTM(spatio-temporal LSTM, ST-LSTM) 网络, 该模型可以同时建模不同时间帧之间的暂态依赖关系以及某一时间帧处于不同关节的空间依赖关系. 每个 ST-LSTM 单元对应一个身体关节, 它接收上一时刻自身关节的隐表示, 同时接收当前时间帧前一关节的隐表示.

针对自动语音识别问题, Li 等[18] 提出了层轨迹 LSTM (layer trajectory LSTM, ltLSTM), 该模型对目标分类任务和暂态依赖关系进行解耦建模, 有效降低了错误率. ItLSTM 使用多个时刻 LSTM 的隐状态输出构建一个层-LSTM(layer-LSTM, L-LSTM), L-LSTM 将当前层的 LSTM 隐状态输出作为输入. 每一时刻对应一个不同的 LSTM, 所以 ItLSTM 各层之间不共享权重. LSTM 引入反馈可以对暂态依赖关系进行建模, 而 L-LSTM 通过对 LSTM 隐状态多个输出进行加和作为多元音素分类问题的输入. L-LSTM 层有一个门限路径控制从输出层到底层的信息流, 可以缓解梯度消失问题.

2.2.2　小结与纵向分析

上述各模型在网络结构、模型特点以及应用领域的总结如表 1 所示. 虽然 RNNs 可以有效挖掘数据中的时序信息和语义信息, 但是由于梯度消失和梯度爆炸问题, RNNS 无法捕获大范围的序列依赖关系. 为了解决该问题, LSTM 应运而生, 其引入门控机制, 通过有选择性地保留之前的信息, 可以从输入序列中检测到重要特征, 并在长时间内保留该特征信息, 从而捕获潜在的大范围序列依赖关系. GRU 是在 LSTM 基础上进行的改进, 该模型简化了门控的数量,

在构建较大网络时能够用更少的参数获得比LSTM还好的效果,并可以节约计算成本. 然而,GRU无法解决序列中存在缺失值的问题,因此GRU-D在GRU的基础上引入衰变机制,对GRU的输入和隐状态应用掩码和时间间隔捕获观测值及其依赖性,从而提高预测结果. 深度门限LSTM引入深度门连接相邻层的记忆单元,在下层与上层循环单元之间引入一种线性依赖关系,有效地增加了模型语言建模和机器翻译的性能. 乘法LSTM将mRNN与LSTM这两种互补的结构相结合,由于引入mRNN, mLSTM的中间状态更新过程为线性,更容易实现并行计算. 由于权值函数也在更新迭代,该模型能够学习到更准确的序列联合概率分布. 然而LSTM主要用于时域,只是对上一时刻的隐状态进行循环更新,ST-LSTM通过同时引入时间和空间位置信息并增加调节输入,将循环分析扩展到时空域,相比较LSTM,该模型能够提高网络对输入序列的噪声的鲁棒性. ltLSTM将目标分类任务和暂态依赖关系进行解耦建模,能够有效降低语音识别的错误率,相比较LSTM,该模型能够考虑多个时刻的暂态依赖关系.

表1 长短期记忆网络模型总结

模型	网络结构	模型特点	应用领域
LSTM	输入门+遗忘门+输出门	相比较RNNs,能捕获潜在大范围序列依赖关系	序列建模
GRU	更新门+复位门	相比较LSTM,模型简单,在训练数据很大的情况下效果更好	序列建模
GRU-D	GRU+衰变机制	相比较GRU,解决缺失值问题	时间序列预测
深度门限LSTM	LSTM+深度门	相比较LSTM,在上层和下层循环单元之间引入线性依赖	语言建模、机器翻译
mLSTM	mRNN+LSTM	相比较LSTM,在每个门限单元和mRNN隐状态之间添加连接	离散多项式序列建模
ST-LSTM	时间信息+空间信息+调节输入	相比较LSTM,提高网络对输入序列的噪声的鲁棒性,将循环分析扩展到时空域	人体动作识别
ltLSTM	LSTM+L-LSTM	相比较LSTM,将目标分类任务和暂态依赖关系解耦,考虑多个时刻的暂态依赖关系	语音识别

2.3 微分循环RNNs

对于动作识别任务,许多时空模型(如3D-SIFT[19]和HoGHoF[20]),已用于检测并编码视频帧中与物体显著运动特点相关的时空点,并揭示动作的重要动态特性. Veeriah等[21]提出了一种新的LSTM模型,通过检测和整合显著的时空序列,自动学习动作的动态特性. 但是,LSTM中的门限单元没有明确考虑输入序列中存在的动态结构的影响,使得该模型很难学习行为状态的演变. 为了解决此问题,Veeriah等提出了一种微分RNN模型学习和整合动作的动态特性. 该模型每个记忆单元的内部状态包含关于时空结构的累积信息,即它是输入序列的长短期记忆表示. 因此,状态变量导数(derivative of states, DoS) $\mathrm{d}\boldsymbol{c}_t/\mathrm{d}t$量化时刻$t$的信息变化. 换言之,较大的DoS是一个显著的时空结构指标,其中包含由动作状态的突然变化引起的信息动态特性. 在这种情况下,门限单元应该允许更多信息进入记忆单元来更新其内部状态. 相反,当DoS的值很小时,输入信息应该被拒绝进入记忆单元,这样内部状态才不会受到当前输入的影响. 因此,该模型使用DoS作为控制信息流进出记忆单元内部状态的一个指标.

基于微分RNN的良好性能,Zhuang等[22]提出深度微分RNNs (deep differential RNN, d^2RNN). 受Graves等[23]的启发,该模型通过DoS的特定阶数调节LSTM门,在DoS阶数递增的情况下,堆叠多层LSTM单元. 具体而言,d^2RNN的第1层用零阶DoS,与传统的LSTM单元类似;第2层用带有一阶DoS的LSTM单元;第3层用带有二阶DoS的LSTM单元,以此类推. d^2RNN的每一层通过DoS的特定阶数学习信息增益的变化,随着d^2RNN的单元层变深,此模型可以学习更高阶和更复杂的动态模式.

2.4 高速公路网

Srivastava等[24]从LSTM结构中获得灵感,通过在多个隐层信息流动过程中增加门控单元缓解深层网络训练难的问题,应用于手写体识别和图像分类问题. 这些单元不仅可以在层内控制信息,而且转换门会学习输入的非线性变换对隐层的总输出的贡献有多少,进位门会学习有多少输入传给输出. 转换门和进位门类似于LSTM结构中的输入和遗忘门,而从输入到输出的权重为1的跳跃连接类似于LSTM单元里的循环自连接. 因为该模型允许信息畅通无阻地跨层流动,所以具有这种体系结构的网络称为高速公

路网.

为了利用高速公路网减缓梯度消失问题的优良性能,进一步增加网络深度,Julian 等[25] 提出了循环高速公路网络(recurrent highway network, RHN)层,在循环状态转换之间增加一个或多个(据所需循环深度确定层的个数)高速公路层,在保证LSTM 易于训练特性的同时,使RHN 利用循环变换增加网络深度,有效解决序列学习任务.

基于深度神经网络的声学模型能够大大提高自动语音识别的准确性,但是在出现混响和重叠的声学信号时,现有的语音处理技术仍然很难达到满意的效果. 因此,为了降低字误率,Zhang 等[26] 利用高速公路网络在深度网络上适应性较好的特点,提出了高速公路LSTM (HLSTM) RNNs, 该模型下层记忆单元与上层记忆单元之间具有直接门控连接,进位门控制从下层单元直接流到上层单元的信息量.

上一节描述的单向LSTM RNNs 只能利用过去的历史信息,然而,在语音识别中,未来的数据也会携带重要的信息,因此应该利用未来的信息进一步增强声学模型. 双向RNNs(如第3.1 节所示)通过在两个方向上使用两个独立的隐层处理数据,从而充分利用过去和将来的上下文信息. 文献[23, 27-28]的研究结果显示,双向LSTM RNNs 确实可以提高语音识别的效果,因此将HLSTMRNNs 从单向扩展到双向. 需要注意的是,反向层与前向层使用相同的更新公式,只是$t-1$被$t+1$代替,以便利用未来帧的信息,且模型计算是从最后一个时刻$t=T$到1,最后将反向层与前向层的输出连接起来形成下一层的输入.

2.5　多维RNNs

二维LSTM[29] 由Graves 等作为LSTM 的推广而引入,并应用于神经机器翻译. Gundram 等[30] 提出的LSTM 变体 —— 二维LSTM, 可以处理任意长度的二维序列数据. 二维LSTM 在记忆单元中保留了一些状态信息,同时使用一个额外的λ门,λ门的计算过程与其他门类似. 最后更新记忆单元时在通过遗忘门前,λ门对前两个记忆单元$\boldsymbol{c}_{j-1,i}$和$\boldsymbol{c}_{j,i-1}$进行加权,因此λ门的作用相当于另一个遗忘门,决定前两个记忆单元信息保留的程度. 以此类推,多维LSTM 便是有多个不同的遗忘门来处理多维上的信息,最终通过加权更新最后的记忆单元.

多维循环神经网络 (multidimensional RNN, MDRNN)[31] 的基本思想是利用与数据时空维数同样多的连接替换RNNs 中的单个循环连接,这些连接使网络灵活地学习周围环境的内部表示,对局部失真有很强的鲁棒性. 首先,MDRNN 隐层会扫描一维区间中的输入,并将激活函数的输出存储在缓冲区中. 一维扫描区间以如下方式排序:在每个点处,都有先前点的隐状态输出,这些先前点的隐状态输出通过循环连接与输入一起馈送到当前点. 一个这样的隐层足以使网络能够从当前点扫描的方向访问所有上下文. 然而,模型通常希望能够利用某一点所有方向的上下文信息进行建模,典型的一维解决方案是BRNNs (如第3.1 节所示):两个独立的隐层向前和向后扫描输入. 所有隐层都连接到一个输出层,因此能够接收所有周围环境的信息,然后利用反向传播随时间的n维扩展来计算MDRNN 的误差梯度. 与一维网络情况一样,数据处理的顺序与前向过程的顺序相反,每个隐层在每个时刻都接收激活函数输出的导数及其自身的n个“未来”导数.

2.6　嵌套堆叠RNNs

堆叠LSTM 目前是处理序列预测问题的成熟技术,通过增加网络的深度提高训练的效率,获得更高的准确性. 堆叠LSTM[32] 结构可以定义为由多个LSTM 层组成的LSTM 模型,每个时刻的输入对应一个输出,而不是所有输入对应一个输出.

LSTM 中的输出门编码过程实现人类的直觉观念,即在当前时刻,不相关的记忆仍然会被记住,嵌套LSTM (nested LSTM, NLSTM)[33] 便是用这种直觉创建记忆的时间层次结构. 嵌套,即一层套一层的结构,具体而言,NLSTM 单元用外层记忆单元的输出作为内层的输入和隐状态来计算内层记忆单元的值,一层一层这样循环嵌套便生成了NLSTM, 这种方法使得网络能够进行任意深度的嵌套.

2.7　小结与横向分析

前文已经对衍生RNNs 各分类下的模型进行了纵向分析比较,了解到各模型的特点以及适用领域,下面将各模型分类进行横向分析比较,以便了解各分类的研究角度和分类下各模型的共同点. 各模型分类总结如表2 所示. 双向RNNs 是在RNNs 的基础上引入反向层,实现了同时利用过去和未来信息的功能. 长短期记忆网络以RNNs 的一个经典变体LSTM 为基础,在内部结构上进行改进,其中的深度门限LSTM 与高速公路网相似,使用相同的输入线性和选通连接思想. 它与高速公路网的区别是,在非线性路径上不使用特定的门,而是保留LSTM 非线性变换的输入和输出门,同时深度门限LSTM 线性连接上下层的记忆单元,因此深度门限LSTM 中的记忆单元有从未来和顶层反向传播的误差. 微分循环网络引

入微分量化信息变化的思想,使得模型可以有效学习复杂的动态模式. 高速公路网通过在多个隐层信息流动过程中增加门控单元,实现跨层信息流的控制,缓解深层网络训练难的问题. 多维RNNs和嵌套堆叠RNNs均在维度上进行探索,多维RNNs研究多维输入数据,嵌套堆叠RNNs研究多层RNNs. 值得注意的是,衍生RNNs种类繁多,虽然大多已经进行了详细介绍,但不拘泥于以上分类. 例如,Kamil[34]提出了数组LSTM,其主要思想不是建立单个层的层次结构(如堆叠LSTM、门控反馈RNNs),而是在RNNs单元内构建更复杂的记忆结构[35]. Rocki希望通过共享内部状态创建一个函数,迫使学习过程用一个属于隐单元的多个记忆单元去池化相似或可互换的内容,因此,数组LSTM对于随机数组记忆单元对噪声输入具有弹性. Qin等[36]提出了一个单层循环神经网络解决伪凸优化问题,该模型引入的惩罚参数不需要先确定具体参数,具有更好的收敛性. Kosmatopoulos等[37]提出了应用于动态系统辨识的循环高阶神经网络,该模型通过分量乘积等形式的高阶交互模拟系统,因此,只要允许足够的高阶连接,该网络就能够近似任意动态系统.

表 2 衍生RNNs分类横向总结

模型分类	分类结构	分类角度
双向RNNs	引入反向层	双向结构
长短期记忆网络	根据LSTM增减门,添加连接	LSTM内部结构的变化
微分循环网络	引入微分量化信息变化思想	学习复杂动态模式
高速公路网	引入跨层控制思想	加深网络训练深度
多维RNNs	探索数据维度	拓宽输入数据维度
嵌套堆叠RNNs	将LSTM作为整体进行循环	建立层次结构

3 组合RNNs

本节介绍其他经典结构与RNNs或其结构变体相结合形成的组合RNNs,包括卷积RNNs、网格RNNs、图RNNs、暂态RNNs、格子RNNs、分层RNNs和记忆RNNs. 下面对各类变体的模型背景、模型结构以及模型优点进行详细描述,并纵向和横向比较模型之间的差别.

3.1 卷积RNNs

3.1.1 模 型

McLaughlin等[38]结合CNN和RNNs结构提出了循环卷积网络,以提高基于视频的人的再识别问题的准确率,并使用孪生神经网络(siamese network)[39]训练特征提取网络进行人体再识别. 因此,基于循环卷积网络的视频再识别方法准确率更高. 对于由人的全身图像组成的视频序列,每个图像通过CNN产生向量,该向量是CNN输出层激活映射的向量表示. 之后向量被传递到循环层作为输入,投影到一个低维特征空间,并与之前时刻的信息相结合. 需要注意的是,CNN在所有时刻都共享参数,即每个输入帧都由相同的特征提取网络处理,所以在CNN与循环层之间使用随机失活(Dropout)以减少过拟合. 在循环层后添加一个暂态池,这样可以在任意时刻聚合序列信息,捕获序列中存在的长期依赖信息,学习任意长度序列的特征向量表示.

针对并发活动识别问题,Li等[40]提出了一个基于多模态数据CNN-LSTM结构的系统. 该系统可识别来自不同类型多个传感器捕获的真实数据的并发活动. 识别分两步完成:第1步从多模态数据中提取空间和时间特征,将每种数据类型输入到一个提取空间特征的CNN中,然后用LSTM提取传感器数据中的暂态信息;第2步将提取的特征进行融合,利用单个分类器实现并发活动识别. 此系统是第1个使用单一模型解决多感官数据并发活动识别的系统,由4个串联的主要模块组成:感官数据预处理、空间特征提取、时间关联与融合、编码层. 多模态CNN-LSTM结构在系统中负责特征提取和时序关联提取.

在语音识别领域,CNN、RNNs和全连接深层神经网络在其建模能力方面是互补的,因此,Hsu等[41]将3个组件组合起来,提出一个新的模型——高速公路卷积循环深度神经网络(highway convolutional recurrent deep neural network,highway CLDNN). 该模型结构遵循Sainath等[42]的设计思想,由于循环层能够捕获暂态依赖关系,在每个时刻可以将没有上下文的帧输入传递给该网络. Hsu等使用滤波器组特征和基音特征表示每帧,根据Sainath等[42]的建议,为了捕捉来自不同抽象层次的输入表示,Hsu等还将原始输入特征也传递给循环层. 由于语音信号有不同时间尺度的信息[43],希望循环层也可以在不同时

间尺度上捕获暂态依赖关系,又因为DNN可以捕获这种关系[44],该模型为了充分利用深层循环结构,将Highway LSTM用于循环层. 将最后一层循环层的输出馈送到全连接的前馈层,这样可以更好地对输出目标进行分类.

基于神经网络的良好性能,Ning等[45]将深层神经网络分析扩展到时空域用于视觉目标跟踪,提出了空间监督循环卷积神经网络,首先使用YOLO[46]系统采集视觉特征并进行初步的位置推断,然后在下一阶段使用LSTM,这是因为LSTM可以捕捉大范围空间依赖关系,适合于人和物跟踪过程的序列处理. 此模型属于深度神经网络,它将原始视频帧作为输入,并返回每个帧中被跟踪对象边界框的坐标. 空间监督循环卷积神经网络用LSTM将YOLO深度卷积神经网络扩展到时空域,可以有效处理时空信息、推断区域位置.

对于音频标记问题,大多数模型使用单声道录音,或者简单地将多声道的平均值作为输入信号. 然而,这种融合策略忽略了立体声音频的空间信息,会降低识别精度. 因此,为了提高识别精度,Xu等[47-48]提出了卷积门限循环神经网络(convolutional gated recurrent neural network, CGRNN). 首先,该模型使用CNN作为特征提取器;然后,将提取的鲁棒特征馈入双向GRU(bidirectional gated recurrent unit, BGRU)中,由于GRU只能利用历史信息,该模型用BGRU学习长期音频模式进而利用未来信息;接着,选用sigmoid激活函数,通过一个前馈神经层得到目标音频事件的后验概率;最后,CGRNN用单层前馈深度神经网络(deep neural network, DNN)处理BGRU的最终序列以预测标签的后验概率.

3.1.2　小结与纵向分析

卷积RNNs结合CNN和衍生RNNs的优点,应用于多个领域,上述各模型网络结构、模型特点以及应用领域的总结如表3所示. 循环卷积网络可以自动学习提取与重新识别相关的时空特征,充分利用RNNs处理不同长度序列的优点,是成功的尝试. CNN-LSTM结构是第1个利用单一模型解决多传感器数据并发活动识别的系统,该系统具有可扩展性强、训练简单、易于部署等特点. Highway CLDNN通过将循环卷积网络中的RNNs换成HLSTM,使得信息可以从低层单元直接流到高层单元以学习更深层次的隐表示信息. 空间监督循环卷积神经网络在保持较低计算成本的同时,具有更高的精度和鲁棒性. CGRNN通过将循环卷积网络中的RNNs换成BGRU来同时利用过去和未来的信息学习音频之间的暂态依赖关系.

表 3　卷积RNNs模型总结

模型	网络结构	模型特点	应用领域
循环卷积网络	CNN+RNNs	通过CNN产生特征向量,通过RNNs利用序列内的时间信息	视频再识别
CNN-LSTM结构	CNN+LSTM	相比较循环卷积网络,用LSTM提取传感器数据中的暂态信息	并发活动识别
Highway CLDNN	CNN+HLSTM	相比较循环卷积网络,利用HLSTM学习更深层次结构的隐表示信息	语音识别
空间监督循环卷积神经网络	YOLO+LSTM	相比较循环卷积网络,利用YOLO采集视觉特征进行初步的位置推断,用LSTM捕捉大范围空间依赖关系	视觉目标跟踪
CGRNN	CNN+BGRU	相比较循环卷积网络,用BGRU同时利用过去和未来的信息学习音频暂态依赖关系	音频标记

3.2　网格RNNs

网格LSTM由Nal等[49]提出,其将LSTM块组成多维网格,因此每个网格包含每个维度的一组LSTM块. 该模型还引入了相邻单元状态之间的每一维门限线性依赖关系,缓解了在所有维度上训练时的梯度消失问题. 例如,Hsu等[50]考虑的是二维网格LSTM模型,维度分别是时间和深度两个维度.

在第3.1节中,时间维度LSTM隐单元的输出未在当前时刻进行分类,但是深度维度应知道当前时刻网格中其他维度的输出,以便在分类之前获得更多的更新信息. 深度维度LSTM的输入在同一网格的时间维度LSTM更新完成之后再进行更新,Hsu等[50]将此模型称为优先级网格LSTM.

在语音识别领域,RNNs已被扩展到同时建模时间和频率维度的信息,鉴于网格LSTM的优良性能,Kreyssing等[51]提出了频率依赖网格RNN(frequency dependent grid-RNN, FD-GRNN). 该模型使用RNNs块,将时间轴输入窗口分到相同的7个时间段中,将频率轴分成5个大小为10的块. 网格RNN由两个RNN组成,一个有非线性激活功能,另一个有线性激活功能. 非线性RNN执行特征提取,线性RNN对非线性RNN处理后的信息流进行建模,改善非线性激活函数的信息流. 整体结构以上述展开形式进行训练,类似Saon等[52]的训练过程.

3.3 图RNNs

为了从图结构和时变数据中学习时空结构,提高模型的学习速度,Youngjoo等[53]提出了两种图卷积循环网络(graph convolutional recurrent metwork, GCRN)结构. 两种结构均为图结构、卷积与LSTM结构的结合,但是融合方式不一样,GCRN I属于组合RNNs范畴,在语言建模方面有很好的表现,GCRN II属于混合RNNs,在视频预测方面表现出良好的性能.

本节主要介绍GCRN I,该模型将用于特征提取的图CNN与用于序列学习的LSTM(RNNs也可以)堆叠起来,输入矩阵$\boldsymbol{x}_t$表示动态系统在时刻t的观测值,动态系统网络结构由图$\mathcal{G}$给出,$\boldsymbol{x}_t^{\rm CNN}$为图CNN门的输出. 为了结合图CNN与LSTM, Youngjoo等[53]令$\boldsymbol{x}_t^{\rm CNN}=W^{\rm CNN}{}_{\mathcal{G}}\boldsymbol{x}_t$,其中$W^{\rm CNN}$为图卷积核的Chebyshev(切比雪夫)系数. 以上结构可以利用图CNN的局部稳定性、两个结构的组合特性以及动态属性捕获数据上的分布.

Yuan等[54]针对动作驱动的视频目标检测任务建立了一个完全可微分的时间动态图LSTM(temporal dynamic graph LSTM, TD-Graph LSTM)结构,该结构摒弃传统的使用静态图像学习检测器的方法,改用动作描述作为监督,从日常活动视频中学习检测器,同时解决了全局描述中存在的“缺失标签”的现象. 输入视频中的每一帧首先通过空间卷积网络获取候选区域的空间视觉特征;然后在两个连续的帧中连接所有语义相似的区域构造时间图结构;接着利用TD-Graph LSTM单元在整个时间图上循环传播信息,其中LSTM单元将空间视觉特征作为输入,空间视觉特征由时间上下文特征加权求和获得,而时间上下文特征由时间动态图构造获得. TD-Graph LSTM输出所有区域增强的时间感知特征,采用区域级分类模块产生分类置信度,最终将这些区域级预测汇总起来,生成帧级别的对象类别预测,并由动作标签中的对象类别进行监督.

3.4 暂态RNNs

对于诸如语音识别的任务,如果输入与标签之间的对应方式未知,则RNNs无法发挥作用. CTC(connectionist temporal classification)[55]又称为暂态连接分类[56],专门用于时序分类任务,即用于那些输入与目标标签之间对应关系未知的序列数据标签预测问题. CTC输出层使用判别损失函数训练RNNs网络连接权值矩阵和偏置,预测未知序列的类标签. 首先利用softmax层定义沿着输入序列每个时刻输出的概率分布[23];然后利用前向-后向算法对所有可能的对齐进行求和,并在给定输入序列的情况下确定目标序列的归一化概率分布. CTC的特别之处在于其完全忽略了分段,将给定标签的输出概率描述为几个可行路径的和. 利用CTC训练的RNNs通常是双向的,以确保每个输出概率分布依赖于整个输入序列,而不仅仅是由输入得到的序列段. 关于语音和手写体识别的实践表明,具有CTC输出层的双向LSTM网络能够有效预测序列标签,通常情况下效果好于标准隐马尔科夫模型.

基于视频重新识别问题的方法大多关注两点:特征学习和度量学习. 为了有选择地学习最相关的图像,应利用注意力机制探索给定图像序列的时间结构,从而进一步提高特征学习的性能. 由此,Zhou等[57]提出暂态注意力模型(temporal attention model, TAM). 假设每个图像序列表示为$\boldsymbol{x}=\{\boldsymbol{x}_t\mid\boldsymbol{x}_t\in\boldsymbol{R}^D\}_{t=1}^T$,$T$为图像序列长度,$D$为图像维数,将CNN表示为$f(\boldsymbol{x})$,$fc7$为CNN的一个全连接层. 模型由两部分组成,即注意力单元和RNNs单元,在每个时刻t,注意力单元将$\{f(\boldsymbol{x}_t)_{fc7}\}_{t=1}^T$作为输入并产生这些特征的加权平均值,即

$$\overline{\boldsymbol{x}}_t=\sum_{i=1}^{T}w_{t,i}f(\boldsymbol{x}_t)_{fc7},\tag{8}$$

其中$\{w_{t,i}\}$由子网学习. 将$\bar{\boldsymbol{x}}_t$输入到神经网络中,神经网络采用LSTM结构[58],其能够在长距离序列中捕捉大范围暂态依赖关系. 图像序列的最终表示是每次输出的暂态平均池[59].

3.5 格子RNNs

在解决命名实体识别(named entity recognition, NER)问题时,由于现有的模型可能会引起不正确分割,而不正确分割的实体边界会导致NER错误,为了提高NER效果,Zhang等[60]提出了格子LSTM,其可以看作是字符模型的扩展,集成了字符单元和附加控制信息流的门. 该模型对一系列输入字符以及与词典匹配的所有潜在单词进行编码,与基于字符的方法相比,显式地利用了单词和单词序列信息,与基于词的方法相比,格子LSTM不存在分割错误,其中的门控循环单元格可以从句子中选择最相关的字符和单词,以获得更好的结果[61-63].

格子循环单元(lattice recurrent unit, LRU)[64]模型在时间和深度维度上具有不同的信息流,学习深层模型可以看作是GRU模型的扩展,主要用于语言建模问题. 第1个LRU模型称为预状态LRU,仅解耦每个维度的投影状态;第2种模型称为复位门LRU,可以进一步解耦复位门;最后一个称为更新门LRU,可以

解耦所有投影状态、复位门和更新门3个组件.

3.6 分层RNNs

3.6.1 模　型

具有深层结构的神经网络已被引入到语音带宽扩展(speech bandwidth extension, BWE)领域,但现有方法在声音编码器参数化的过程中会降低语音质量,且由于相位翘曲问题很难对相位谱进行参数化和预测,因此,Ling 等[65] 提出了分层 RNN (hierarchical RNN, HRNN). HRNN 由 LSTM 和前馈 (feed-forward, FF)层组成,这些LSTM和FF层形成多层的层次结构,并且每层以特定的时间分辨率工作,除顶层外,每个层的输入都包含上一个层的条件向量. 该模型结构类似于样本RNN[66],主要区别在于样本RNN模型是一个无条件音频发生器,其使用输出波形的历史信息作为网络输入,并以自回归方式生成输出波形. 然而,HRNN模型直接描述了两个波形序列之间的映射关系,并没有考虑输出波形的自回归特性,这样有助于降低计算复杂度,并有利于并行计算.

从输入窄带波形中提取的一些帧级辅助特征,如瓶颈(bottleneck, BN)特征[67],在提高基于声音编码器的BWE[68]性能方面很有效. 为了将这些辅助输入与HRNN模型相结合,Gu等[68]设计了条件HRNN,与HRNN相比,前者在顶部增加了一个条件层. 条件层的输入特征是从输入波形中提取的帧级辅助特征向量,而不是波形样本,与HRNN相比,加入了附加条件的条件HRNN可以进一步提高重构宽带语音的主观质量.

对于几何场景解析问题,传统的语义场景标记方法无法充分利用局部到全局不同视角的图像信息,并且无法发现主要区域之间的交互关系,因此Peng等[69]提出了分层LSTM(hierarchical long short-term memory, H-LSTM). 该模型包含两个耦合的子网络:分别用于处理曲面标记和关系预测的像素LSTM (Pixel LSTM, P-LSTM)、多尺度超像素LSTM(Multi-scale Super-pixel LSTM, MS-LSTM),两个子网互相提供互补的信息学习层次化的场景上下文信息,同时标记几何表面并确定主要区域之间的交互关系.

目前,已经出现一些解决语境层次结构建模问题的方法,但大多研究忽视了语境中词汇和话语的重要性,可能会丢失上下文中的重要信息,因此Xing等[70]提出了分层循环注意力网络(hierarchical recurrent attention network, HRAN)模拟生成概率. 粗略地讲,在生成概率前,HRAN利用双向GRU将上下文中每个话语的信息编码为隐向量,在生成单词时,分层注意力机制会分别以词级注意力和话语级注意力来关注话语内部和话语之间的重要信息. 在这两种注意力水平下,HRAN采用自下而上的方式工作:单词级注意力处理话语的隐向量,并将其上传到话语级编码器形成上下文的隐向量;话语级注意力将上下文的隐向量进一步处理为上下文向量,并上传到GRU生成单词.

前文所述的分层模型虽然可以学习分层和暂态表示,但没有通过发现序列隐层次结构来学习暂态依赖关系,因此Chung等[71]提出了一种新的多尺度模型——分层多尺度LSTM,该模型的关键是引入了参数化段落边界检测器,该检测器在堆叠RNN的每一层中输出一个二进制值,并学习何时应以优化总体目标的方式结束某个段落. 使用段落边界状态,在每个时刻,每个层都选择如下操作之一:更新(update)、复制(copy)或刷新(flush). 该选择取决于当前时刻前一层的边界段落状态和当前层中前一时刻的段落边界状态.

3.6.2 小结与纵向分析

分层RNNs结合分层结构和衍生RNNs的优点,可以应用于多个领域. 上述模型的网络结构、模型特点以及应用领域总结如表4所示.

表 4　分层RNNs模型总结

模型	网络结构	模型特点	应用领域
HRNN	LSTM+FF+分层结构	LSTM层和FF层形成层次结构,并且每一层以特定的时间分辨率操作	语音带宽扩展
条件HRNN	HRNN+一个条件层	相比较HRNN,将帧级辅助特征和HRNN结合	语音带宽扩展
H-LSTM	P-LSTM+MS-LSTM	相比较LSTM,利用两个子网互相提供互补的上下文信息	几何场景解析
HRAN	双向GRU+注意力+GRU+分层结构	相比较GRU,利用分层注意力机制分别以词级注意力和话语级注意力来关注话语内部和话语之间的重要信息	聊天机器人多回合反应生成
分层多尺度LSTM	堆叠LSTM+段落边界检测器	相比较堆叠LSTM,利用段落检测器学习何时应以优化总体目标的方式结束某个段落	字符级语言建模和笔迹序列生成

HRNN模型利用LSTM层和FF层组成的神经网络表示每个宽带或高频波形样本在输入窄带波形样本上的分布. LSTM层形成一个层次结构,每一层以特定的时间分辨率操作有效地捕获时间序列之间的长跨度依赖. 条件HRNN利用基于DNN的状态分类器从窄带语音中提取BN特征等附加条件作为辅助输入,可以进一步提高生成的宽带语音的质量. H-LSTM利用两个耦合子网络P-LSTM和MS-LSTM分别处理曲面标记和关系预测,两个子网络相互提供互补信息以利用层次化场景上下文,并对它们进行联合优化以提高几何场景解析精度. HRAN相比于LSTM,利用分层注意机制分别以词级注意力和话语级注意力关注话语内部和话语间的重要部分,利用词级注意力将词级编码器的隐向量合成为话语向量,并反馈给话语级编码器,构造上下文的隐藏表示,语境的隐向量被话语级注意力处理并形成用于解码的语境向量.

3.7 记忆RNNs

文档阅读的问答问题之前一直利用知识库将信息组织成结构化的形式进行解决,但是知识库系统有其固有的局限性,如不可避免的不完整性和由于固定模式不能支持某些类型的问答问题答案,鉴于此,Miller等[72]提出了键值记忆网络(key-value memory networks, KVMN). 该模型将事实存储在键值结构存储器中,然后对其进行推理以预测答案,从而进行问答. 由于键值记忆网络的良好性能,Jain等[73]将该网络与CNN编码器和RNN编码器相结合应用于视频字幕任务中,网络中的键和值将视觉空间上下文信息转换为语言空间,有效地捕获视觉特征与文本描述之间的关系.

前文已经证明,RNNs可以长时间保留输入信息,但是现有的RNNs体系结构很难在每个时刻分析哪些信息需要精确地保留在其隐状态下,尤其是当数据具有复杂的结构时(这在自然语言中很常见). 基于这一难题,Tran等[74]提出了一种新的RNNs结构,称为循环记忆网络(recurrent memory network, RMN). 对于语言数据,RMN不仅能够确定哪些语言信息会随着时间的推移而保留,以及为什么会这样,而且还能发现数据中的依赖关系. RMN由两个组件组成:LSTM和记忆块(memory block, MB). MB是记忆网络[75]的一种变体,获取LSTM的隐状态并利用注意力机制将其与最新输入进行比较[76],因此,分析一个训练过的模型的注意力权值可以使LSTM中随时间推移而保留的信息更有价值.

3.8 小结与横向分析

本节的模型分类原则较为直观,不同的分类便是不同的经典模型结构与RNNs结构的结合,因此分类结构和分类角度都无需再赘述. 卷积RNNs侧重于利用CNN特征提取和RNNs学习暂态依赖关系的优点,主要应用于识别问题. 同时,CNN和RNNs结构的结合存在却不仅存在于卷积RNNs分类中. 网格RNNs的主要思想是将衍生RNNs中的模型包装成块,以块为整体进行操作,该分类与衍生RNNs中堆叠LSTM的区别便在于此. 同时,网格LSTM和HLSTM最显著的区别在于,前者深入又充分地利用了LSTM的垂直信息,能够提供LSTM在深度维度上拥有的所有功能. 图RNNs重点侧重于RNNs应用在图结构数据上的尝试. 一般而言,图结构数据比较适合用RecursiveNNs进行处理,但实验证明,图结构数据在经过CNN处理学习到特征后,完全可以用RNNs进行暂态依赖关系的建模. 暂态RNNs侧重于构造探索RNNs性能的一些辅助结构. 本文所介绍的CTC和暂态池结构能够大幅提升RNNs模型的性能,同样,这些辅助结构不仅应用于RNNs模型中,也可以与其他模型相结合以得到更好的模型效果. 格子RNNs在模型结构内部点之间(如字符之间、隐单元之间)添加格子模块连接,在有效的资源下更有效地利用信息,与以模型整体为块的网格RNNs区分开. 分层RNNs是分层结构与RNNs的结合,有了层次结构后,模型可以学习更多不同的有效信息,降低计算复杂度,有利于并行计算,提高模型性能. 记忆RNNs则是记忆结构与RNNs的结合,通过键值对或者记忆块使模型有选择地保留更有效的信息,提高预测的准确性.

4 混合RNNs

本节介绍一些同属于衍生RNNs和组合RNNs的模型,这一类网络模型既有不同网络模型的组合,又在RNNs内部结构上进行了修改. 下面对混合RNNs的模型背景、模型结构以及模型优点进行详细描述,并对这些模型进行分析.

4.1 模 型

为了更好地表示包含复杂多级关联的数据,Liang等[77]提出了结构演化LSTM. 该模型由5个门组成:输入门、遗忘门、自适应遗忘门、记忆门和输出门,另外一个节点连接边门. 结构演化LSTM单元通过对当前节点的输入状态及其隐状态进行作用,为不同相邻节点指定不同的自适应遗忘门;然后利用权重矩阵对自适应遗忘门进行加权以计算每对图节点的联合概率分布. 直观地说,自适应遗忘门用于识

别不同节点对的显著依赖关系,结构演化LSTM利用自适应遗忘门来估计每对图的合并概率. 该模型作为一个记忆系统,将信息写入记忆状态并由每个图节点按顺序记录,用于后续图节点和前一LSTM层隐状态之间的通信,因此联合概率能有效地学习并应用在下一层中生成新的高层的图结构.

针对交通预测问题, Yaguang 等[78] 提出了扩散卷积循环神经网络(diffusion convolutional recurrent neural network, DCRNN), 该模型考虑了交通流中的空间和暂态依赖性. 具体而言, DCRNN 利用扩散卷积层学习图结构数据的表示, 利用扩散卷积代替 GRU 中的矩阵乘法, 以生成目标未来时间序列的可能性作为损失函数, 使用反向传播训练整个网络. DCRNN 能够捕获时间序列之间的时空依赖性, 并且可以应用于各种时空预测问题.

本节介绍 Youngjoo 等[53] 提出的 GCRN II, 即利用图卷积 $*_{\mathcal{G}}$ 代替欧几里德二维卷积 $*$ 形成混合 RNNs 结构. 该模型由 Chebyshev 系数和定义的图卷积核支集来确定参数的个数,与节点的数量无关. 在分布式计算环境中,图卷积核支集控制通信开销,即任何给定节点为了计算局部状态而应该交换的节点数. 该模型在视频预测方面表现出较好的性能,由于 GCRN I 和 GCRN II 只是图卷积和 LSTM 嵌入方式不同,因此两个模型的优点基本相同.

4.2　小结与分析

混合 RNNs 是由于同属于衍生 RNNs 和组合 RNNs 而单独列出来的一类模型,其网络结构、模型特点以及应用领域的总结如表 5 所示. 结构演化 LSTM 可以更有效地传播远程数据依赖关系. DCRNN 利用双向图随机游走来建立空间依赖模型,使用 RNNs 捕捉时间动态. GCRN 与 DCRNN 相似,但该模型使用图卷积而非扩散卷积进行嵌入,这种操作会引起参数的维数大幅增加,影响模型效果,为了达到理想效果,可以参考 Shi 等[79] 缩小数据维度.

表 5　记忆 RNNs 模型总结

模型	网络结构	模型特点	应用领域
结构演化LSTM	图结构+LSTM+自适应遗忘门+节点连接边门	相比较图卷积循环网络I, 为不同的相邻节点指定不同的自适应遗忘门, 增加访问标志	语义对象解析
DCRNN	图结构+GRU+扩散卷积	相比较图卷积循环网络I, 用扩散卷积层学习图结构数据的表示, 用扩散卷积代替GRU中的矩阵乘法	交通预测
GCRN II	图结构+CNN+LSTM	相比较图卷积循环网络I, 用图卷积代替欧几里得二维卷积	视频预测

5　RecursiveNNs 与 RNNs

为了加深对 RNNs 的理解, 更好地区分 RecursiveNNs 与 RNNs, 本节主要介绍 RecursiveNNs 的基本结构以及两个网络的区别和联系.

5.1　RecursiveNNs 基本原理

RecursiveNNs 可以对结构化的输入进行操作, 而不仅只适用于序列结构, 本质是多个同样结构的子树递归地构造更大子树的过程, 其体系结构如下[80]: 沿着序列长度递归地应用相同的权重集. 给定一个局部有向无环图, RecursiveNNs 以拓扑顺序访问节点, 在先前计算的子表示中递归地应用变换生成进一步的表示.

给定一个固定叶子节点的二叉树结构, RecursiveNNs 计算内部节点 η 表示为

$$\boldsymbol{x}_\eta = \sigma(W_L \boldsymbol{x}_{l(\eta)} + W_R \boldsymbol{x}_{r(\eta)} + b). \tag{9}$$

其中: $l(\eta)$ 和 $r(\eta)$ 为 η 的左右子节点, W_L 和 W_R 为将左、右子节点连接到父节点的权重矩阵, b 为偏置. 假定 W_L 和 W_R 为方阵, 且不区分 $l(\eta)$ 和 $r(\eta)$ 是叶子还是内部节点, 这种情况下叶子节点的初始表示和非终止的中间表示在同一个空间里. 在解析树示例中, RecursiveNNs 组合了两个子短语, 在同一空间中生成更长的短语, 因此根 ρ 处有一个最终输出层

$$\boldsymbol{y} = \gamma(U \boldsymbol{x}_\rho + c). \tag{10}$$

其中: U 为输出权重矩阵, c 为输出层的偏置. 在监督学习任务中, 监督过程发生在该层, 因此学习过程中 $\boldsymbol{y}$ 会产生初始误差, 并从根向叶子节点反向传播.

5.2　RecursiveNNs 和 RNNs 的区别和联系

RNNs 可看作是沿着时间展开[81], 处理的是序列结构的信息, 在每个时刻, 隐层的输入包含用户的输入和前一时刻隐层的输出. RNNs 可看作是 RecursiveNNs 的一个特例, RecursiveNNs 更像一个分层网络, 在该网络中, 输入序列实际上与时间不相关, 但必须以树的方式分层处理输入. 当 RNNs 在测试过程中遇到一个与在训练过程中看到的序列长度不同

的序列时无法处理位置相关的权重,因此权重沿着序列长度共享. RecursiveNNs由于相同的原因,每个节点上的权值也是共享的.

RecursiveNNs每个父节点的子节点只是与该节点相似的一个节点,不同的神经网络适用于不同的场景. 若想一个一个生成字符,则RNNs非常适合,但若想生成一个解析树,则使用RecursiveNNs更好,因为它有助于创建更好的层次表示.

6 未来趋势及发展方向

RNNs及其结构变体与其他网络相结合,可以实现更复杂的功能,取得更好的学习效果,是当前主流的深度学习发展方向. 本文对RNNs的主要类型进行梳理,给出了模型的应用背景、构造过程以及模型特点. RNNs虽然大有潜力,但也存在很多挑战:

1) 梯度消失与爆炸. 这不仅只是RNNs存在的问题,由于链式求导法则以及存在非线性激活函数,几乎所有网络都会出现梯度消失和爆炸的问题,RNNs尤甚. 尽管出现了很多诸如高速公路网[24]、残差连接、网格LSTM[50]和层轨迹LSTM[18]等为了解决这一问题而涌现的模型变体,但在模型训练深度大幅增加时,这一问题仍然存在并且是阻碍探索模型深度的主要原因,因此彻底解决梯度消失和梯度爆炸问题的空间还很大.

2) 缺失数据处理. RNNs非常显著的特点是利用上下文之间的连续信息,但若序列是删失或截断的数据,或者当更多的噪声和碎片被用作输入时,模型效果便会大打折扣. 文献[11]通过对GRU的输入和隐状态应用掩码和时间间隔(使用衰减项)捕获观测值及其依赖性,处理缺失值的问题,但是衰减项无法完全捕获缺失模式. 因此,如何更有效地处理这种删失截断数据,提高模型的鲁棒性,还有待进一步探索.

3) RNNs无监督学习. 现有的RNNs模型大多都是有监督学习,然而现实生活中的很多序列缺乏足够的先验知识,难以进行人工标注,无监督学习是重要且必要的一个手段. 文献[54]提出一种弱监督的LSTM框架,减轻了每个单独帧的标签缺失问题,却无法实现完全无监督. 因此,如何结合主成分分析方法、等距映射方法、局部线性嵌入方法、拉普拉斯特征映射方法、Hessian局部线性嵌入方法和局部切空间排列方法等实现RNNs的无监督学习是未来一个重要的研究课题.

4) 多任务学习和增量学习. RNNs相关文献中的所有模型大多应用在特定任务环境下,但是多任务学习已经引起了人们的兴趣. 文献[40]提出了一个基于多模态数据CNN-LSTM结构的系统,该系统首次使用单一模型解决多感官数据并发活动识别的系统. 尽管人们一直对利用RNNs实现多任务学习感兴趣,但进展仍然相对较小,并且跨任务共享知识的最佳机制尚不清楚. 因此,如何实现任务增量学习[82]、域增量学习[83]和类增量学习[84-85]是值得深入研究的领域.

5) 与注意力机制融合. 注意力机制[86]作为一种特征选择机制,实现了输出对输入多个变量加权选择的过程,近几年已经成为神经网络中一个重要的概念,尤其在transformer结构[87]问世后,掀起了注意力机制的研究热潮. 注意力机制与RNNs的融合(文献[72,74]做过尝试)由于注意力专注于序列相关性的特点,可以提高RNNs学习暂态依赖关系的能力,两种模型相辅相成互相促进. 文献[88]通过在transformer结构外部加入循环,使得模型可以处理字符串或公式长度超过训练时观察到的长度的问题,增强模型通用性,解决transformer缺少归纳偏置、非图灵完备和不能处理超长输入的问题. 文献[89]通过增加段落级循环解决transformer上下文碎片化和不能建模超过段落长度依赖关系的问题. 由此可见,注意力机制与RNNs相结合是非常值得深入研究的问题.

6) 应用领域扩展. RNNs的应用范围一般包括NLP和图像问题,将RNNs的思想和成果运用到生产和生活场景中,以及加速与这些领域相融合,是未来进一步发展RNNs的关键方向. 目前,RNNs主要应用于人工智能领域,对于其他领域有一定的应用,但缺乏进一步的深入研究. 例如,文献[90]利用GRU进行微电网日前电力负荷预测,文献[91]利用RNNs对物联网中的分布式数据库访问路径进行预测分析. 此外,还有很多领域都需要进行预测分析和相似度测量,因此,RNNs应用领域的扩展是一个重要的研究方向.

7) 参数和计算量. RNNs虽然高效,但是相应的各种变体结合不同结构形成了复杂的混合结构,导致模型参数增多,计算量大幅度增加. 文献[71]引入段落边界检测器,学习何时应以优化总体目标的方式结束某个段落,避免冗余操作,节约计算量. 文献[38,69]通过在模型框架内共享参数或者共享输入来减少参数数量. 然而,参数之间是否也存在某种关系,能否在保证模型准确率的条件下进行模型混合时省去一些不必要的参数,这是一个值得深入研究的问题.

7 结 语

循环神经网络是深度学习领域中非常有分量的一部分,是机器翻译、机器问题回答、序列视频分析的标准处理手段,也是对于手写体自动合成、语音处理和图像生成等问题的主流建模手段. 本文系统地对各种类别循环神经网络进行综述,指出了各个模型的应用背景、模型特点以及模型之间的比较分析. 在各类网络发展和融合的情况下,许多复杂的序列、语音和图像问题得以解决. 从上述各类模型的总结可以看出,循环神经网络的种类相当丰富且发展迅速,尽管各类模型都存在一定的问题和限制,但不能否认的是,随着理论研究的进一步深入和应用领域的进一步扩展,循环神经网络必将会发挥越来越重要的作用.

参考文献(References)

[1] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[2] 夏瑜潞. 循环神经网络的发展综述[J]. 电脑知识与技术, 2019, 15(21): 182-184.
(Xia Y L. A review of the development of recurrent neural network[J]. Computer Knowledge and Technology, 2019, 15(21): 182-184.)

[3] 杨丽, 吴雨茜, 王俊丽, 等. 循环神经网络研究综述[J]. 计算机应用, 2018, 38(S2): 1-6.
(Yang L, Wu Y X, Wang J L, et al. Research on recurrent neural network[J]. Journal of Computer Applications, 2018, 38(S2): 1-6.)

[4] Chung J, Gülçehre Ç, Cho K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J/OL]. 2014, arXiv: 1412.3555.

[5] Schuster M, Paliwal K K. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681.

[6] Razvan Pascanu, Tomas Mikolov, Yoshua Bengio. On the difficulty of training recurrent neural networks[C]. Proceedings of the 30th International Conference on Machine Learning. Atlanta, 2013: 1310-1318.

[7] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.

[8] Graves A. Generating sequences with recurrent neural networks[J/OL]. 2013, arXiv: 1308.0850.

[9] Cho K, van Merrienboer B, Bahdanau D, et al. On the properties of neural machine translation: Encoder-decoder approaches[C]. Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Stroudsburg, 2014: 103-111.

[10] Rubin D B. Inference and missing data[J]. Biometrika, 1976, 63(3): 581-592.

[11] Che Z P, Purushotham S, Cho K, et al. Recurrent neural networks for multivariate time series with missing values[J]. Scientific Reports, 2018, 8: 6085.

[12] Yao K, Cohn T, Vylomova K, et al. Depth-gated recurrent neural networks[J/OL]. 2015, arXiv: 1508.03790.

[13] Sutskever I, Martens, Hinton G E. Generating text with recurrent neural networks[C]. Proceedings of the 28th International Conference on Machine Learning. Bellevue, 2011: 1017-1024.

[14] Ben Krause, Iain Murray, Steve Renals, et al. Multiplicative LSTM for sequence modeling[J/OL]. 2017, arXiv: 1609.07959.

[15] Du Y, Wang W, Wang L. Hierarchical recurrent neural network for skeleton based action recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition. Boston, 2015: 1110-1118.

[16] Shahroudy A, Liu J, Ng T T, et al. NTU RGB+D: A large scale dataset for 3D human activity analysis[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 1010-1019.

[17] Liu J, Shahroudy A, Xu D, et al. Skeleton-based action recognition using spatio-temporal LSTM network with trust gates[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(12): 3007-3021.

[18] Li J Y, Liu C L, Gong Y F. Layer trajectory LSTM[C]. Interspeech 2018. Hyderabad, 2018: 1768-1772.

[19] Scovanner P, Ali S, Shah M. A 3-dimensional sift descriptor and its application to action recognition[C]. Proceedings of the 15th ACM International Conference on Multimedia. New York, 2007: 357-360.

[20] Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies[C]. IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, 2008: 1-8.

[21] Veeriah V, Zhuang N F, Qi G J. Differential recurrent neural networks for action recognition[C]. IEEE International Conference on Computer Vision. Santiago, 2015: 4041-4049.

[22] Zhuang N F, Kieu D, Qi G J, et al. Deep differential recurrent neural networks[J/OL]. 2020, arXiv: 1804.04192.

[23] Graves A, Mohamed A R, Hinton G. Speech recognition with deep recurrent neural networks[C]. IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, 2013: 6645-6649.

[24] Srivastava Rupesh K, Greff Klaus, Schmidhuber Juergen. Training very deep networks[C]. Advances in Neural Information Processing Systems. Montreal, 2015:

2377-2385.

[25] Julian Georg Zilly, Rupesh Kumar Srivastava, Jan Koutník, et al. Recurrent highway networks[C]. Proceedings of the 34th International Conference on Machine Learning. Sydney, 2017: 4189-4198.

[26] Zhang Y, Chen G G, Yu D, et al. Highway long short-term memory RNNS for distant speech recognition[C]. IEEE International Conference on Acoustics, Speech and Signal Processing. Shanghai, 2016: 5755-5759.

[27] Graves A, Jaitly N, Mohamed A R. Hybrid speech recognition with Deep Bidirectional LSTM[C]. IEEE Workshop on Automatic Speech Recognition and Understanding. Olomouc, 2013: 273-278.

[28] Chen K, Yan Z J, Huo Q. Training deep bidirectional LSTM acoustic model for LVCSR by a context-sensitive-chunk BPTT approach[C]. Interspeech 2015. Dresden, 2015: 3600-3604.

[29] Bahar P, Brix C, Ney H. Towards two-dimensional sequence to sequence model in neural machine translation[C]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, 2018: 3009-3015.

[30] Gundram Leifert, Tobias StrauSS, Tobias Grüning, et al. Cells in multidimensional recurrent neural networks[J/OL]. 2020, arXiv: 1412.2620.

[31] Alex Graves, Jürgen Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks[C]. Proceedings of the 22nd Annual Conference on Neural Information Processing Systems. Vancouver, 2009: 545-552.

[32] Jihun Choi, Taeuk Kim, Sang-goo Lee. Cell-aware stacked LSTMs for modeling sentences[J/OL]. 2021, arXiv: 1809.02279.

[33] Joel Ruben Antony Moniz, David Krueger. Nested LSTMs[C]. Proceedings of the 9th Asian Conference on Machine Learning. Seoul, 2017: 530-544.

[34] Kamil Rocki. Recurrent memory array structures[J/OL]. 2021, arXiv: 1607.03085.

[35] Razvan Pascanu, Çaglar Gülçehre, Kyunghyun Cho, et al. How to construct deep recurrent neural networks[C]. The 2nd International Conference on Learning Representations. Banff, 2014: 14-16.

[36] Qin S T, Yang X D, Xue X P, et al. A one-layer recurrent neural network for pseudoconvex optimization problems with equality and inequality constraints[J]. IEEE Transactions on Cybernetics, 2017, 47(10): 3063-3074.

[37] Kosmatopoulos E B, Polycarpou M M, Christodoulou M A, et al. High-order neural network structures for identification of dynamical systems[J]. IEEE Transactions on Neural Networks, 1995, 6(2): 422-431.

[38] McLaughlin N, Martinez del Rincon J, Miller P. Recurrent convolutional network for video-based person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 1325-1334.

[39] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, 2006: 1735-1742.

[40] Li X Y, Zhang Y Y, Zhang J Y, et al. Concurrent activity recognition with multimodal CNN-LSTM structure[J/OL]. 2019, arXiv: 1702.01638.

[41] Hsu W N, Zhang Y, Lee A, et al. Exploiting depth and highway connections in convolutional recurrent deep neural networks for speech recognition[C]. Interspeech 2016. San Francisco, 2016: 395-399.

[42] Sainath T N, Vinyals O, Senior A, et al. Convolutional, long short-term memory, fully connected deep neural networks[C]. IEEE International Conference on Acoustics, Speech and Signal Processing. South Brisbane, 2015: 4580-4584.

[43] Wu S L, Kingsbury E D, Morgan N, et al. Incorporating information from syllable-length time scales into automatic speech recognition[C]. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing. Seattle, 1998: 721-724.

[44] Hermans M, Schrauwen B. Training and analysing deep recurrent neural networks[C]. Advances in Neural Information Processing Systems. Lake Tahoe, 2013: 190-198.

[45] Ning G H, Zhang Z, Huang C, et al. Spatially supervised recurrent convolutional neural networks for visual object tracking[C]. IEEE International Symposium on Circuits and Systems. Baltimore, 2017: 1-4.

[46] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 779-788.

[47] Xu Y, Kong Q Q, Huang Q, et al. Convolutional gated recurrent neural network incorporating spatial features for audio tagging[C]. International Joint Conference on Neural Networks. Anchorage, 2017: 3461-3466.

[48] Xu Y, Huang Q, Wang W W, et al. Unsupervised feature learning based on deep models for environmental audio tagging[J]. ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(6): 1230-1241.

[49] Nal Kalchbrenner, Ivo Danihelka, Alex Graves. Grid long short-term memory[J/OL]. 2019, arXiv: 1507.01526.

[50] Hsu W N, Zhang Y, Glass J R. A prioritized grid

long short-term memory RNN for speech recognition[C]. Proceedings of the 2016 IEEE Spoken Language Technology Workshop. San Diego, 2016: 467-473.

[51] Kreyssig F L, Zhang C, Woodland P C. Improved tdnns using deep kernels and frequency dependent grid-RNNS[C]. IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, 2018: 4864-4868.

[52] Saon G, Soltau H, Emami A, et al. Unfolded recurrent neural networks for speech recognition[C]. The 15th Annual Conference of the International Speech Communication Association. Singapore, 2014: 343-347.

[53] Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, et al. Structured sequence modeling with graph convolutional recurrent networks[C]. Processing 25th International Conference on Neural Information. Siem Reap, 2018: 362-373.

[54] Yuan Y, Liang X D, Wang X L, et al. Temporal dynamic graph LSTM for action-driven video object detection[C]. IEEE International Conference on Computer Vision. Venice, 2017: 1819-1828.

[55] Graves A, Fernández S, Gomez F, et al. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks[C]. Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh, 2006: 369-376.

[56] Zhai C L, Chen Z N, Li J, et al. Chinese Image text recognition with BLSTM-CTC: A segmentation-free method[C]. The 7th Chinese Conference on Pattern Recognition. Chengdu, 2016: 525-536.

[57] Zhou Z, Huang Y, Wang W, et al. See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 6776-6785.

[58] Vinyals O, Toshev A, Bengio S, et al. Show and tell: A neural image caption generator[C]. IEEE Conference on Computer Vision and Pattern Recognition. Boston, 2015: 3156-3164.

[59] McLaughlin N, Martinez del Rincon J, Miller P. Recurrent convolutional network for video-based person re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 1325-1334.

[60] Zhang Y, Yang J. Chinese NER using lattice LSTM[C]. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, 2018: 1554-1564.

[61] Buckman J, Neubig G. Neural lattice language models[J]. Transactions of the Association for Computational Linguistics, 2018, 6: 529-541.

[62] Neubig G, Mimura M, Mori S, et al. Learning a language model from continuous speech[C]. The 11th Annual Conference of the International Speech Communication Association. Makuhari, 2010: 1053-1056.

[63] Pierre Dupont, Ronald Rosenfeld. Lattice based language models[Z]. 2018.

[64] Chaitanya Ahuja, Louis-Philippe Morency. Lattice recurrent unit: Improving convergence and statistical efficiency for sequence modeling[C]. Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, 2018: 4996-5003.

[65] Ling Z H, Ai Y, Gu Y, et al. Waveform modeling and generation using hierarchical recurrent neural networks for speech bandwidth extension[J]. ACM Transactions on Audio, Speech, and Language Processing, 2018, 26(5): 883-894.

[66] Mehri S, Kumar K, Gulrajani I, et al. SampleRNN: An unconditional end-to end neural audio generation model[J/OL]. 2018, arXiv: 1612.07837.

[67] Yu D, Seltzer M L. Improved bottleneck features using pretrained deep neural networks[C]. The 12th Annual Conference of the International Speech Communication Association. Florence, 2011: 237-240.

[68] Gu Y, Ling Z H, Dai L R. Speech bandwidth extension using bottleneck features and deep recurrent neural networks[C]. The 17th Annual Conference of the International Speech Communication Association. San Francisco, 2016: 297-301.

[69] Peng Z L, Zhang R M, Liang X D, et al. Geometric scene parsing with hierarchical LSTM. Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, 2016: 3439-3445.

[70] Xing C, Wu Y, Wu W, et al. Hierarchical recurrent attention network for response generation[C]. Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, 2018: 5610-5617.

[71] Chung J Y, Ahn S J, Bengio Y S. Hierarchical multiscale recurrent neural networks[J/OL]. 2016, arXiv: 1609.01704.

[72] Miller A, Fisch A, Dodge J, et al. Key-value memory networks for directly reading documents[J/OL]. 2016, arXiv: 1606.03126.

[73] Jain A K, Agarwalla A, Agrawal K K, et al. Recurrent memory addressing for describing videos[C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, 2017: 2200-2207.

[74] Tran K, Bisazza A, Monz C. Recurrent memory networks for language modeling[C]. Proceedings of the 2016 Conference of the North American Chapter of the

Association for Computational Linguistics: Human Language Technologies. San Diego, 2016: 321-331.

[75] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, et al. End-to-end memory networks[C]. Annual Conference on Neural Information Processing Systems. Montreal, 2015: 2440-2448.

[76] Karol Gregor, Ivo Danihelka, Alex Graves, et al. A recurrent neural network for image generation[C]. Proceedings of the 32nd International Conference on Machine Learning. Lille, 2015: 1462-1471.

[77] Liang X D, Lin L, Shen X H, et al. Interpretable structure-evolving LSTM[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 2175-2184.

[78] Yaguang L, Rose Y, Cyrus S, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[J/OL]. 2018, arXiv: 1707.01926.

[79] Shi X J, Chen Z R, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[J/OL]. 2015, arXiv: 1506.04214.

[80] Irsoy O. Deep sequential and structural neural models of compositionality[D]. New York: Cornell University, 2017.

[81] Graves A. Supervised sequence labelling with recurrent neural networks[M]. Berlin, Heidelberg: Springer, 2012.

[82] Dai Z H, Peng C, Chen H J, et al. A multi-task incremental learning framework with category name embedding for aspect-category sentiment analysis[C]. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, 2020: 6955-6965.

[83] Adrian Bulat, Jean Kossaifi, Georgios Tzimiropoulos, et al. Incremental multi-domain learning with network latent tensor factorization[C]. The 34th AAAI Conference on Artificial Intelligence. New York, 2020: 10470-10477.

[84] Patra A, Noble J A. Hierarchical class incremental learning of anatomical structures in fetal echocardiography videos[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(4): 1046-1058.

[85] Hu C Y, Chen Y Q, Hu L S, et al. A novel random forests based class incremental learning method for activity recognition[J]. Pattern Recognition, 2018, 78: 277-290.

[86] Kelvin X, Jimmy B, Ryan K, et al. Show, attend and tell: Neural image caption generation with visual attention[C]. Proceedings of the 32nd International Conference on Machine Learning. Lille, 2015: 2048-2057.

[87] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. Attention is all you need[C]. Annual Conference on Neural Information Processing Systems. Long Beach, 2017: 5998-6008.

[88] Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, et al. Universal Transformers[C]. Proceedings of the 7th International Conference on Learning Representations. New Orleans, 2019: 6-9.

[89] Dai Z H, Yang Z L, Yang Y M, et al. Transformer-XL: Attentive language models beyond a fixed-length context[C]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, 2019: 2978-2988.

[90] 张诤杰. 基于门循环单元神经网络的微电网日前电力负荷预测[D]. 徐州: 中国矿业大学, 2020.
(Zhang Z J. Day-ahead load forecasting of microgrid based on GRU network[D]. Xuzhou: China University of Mining and Technology, 2020.)

[91] Yu G Z, Fu W N. Analysis of distributed database access path prediction based on recurrent neural network in internet of things[J]. Concurrency and Computation: Practice and Experience, DOI: 10.1002/cpe.5116.

作者简介

刘建伟(1966−), 男, 副教授, 博士生导师, 从事模式识别与智能系统等研究, E-mail: liujw@cup.edu.cn;

宋志妍(1997−), 女, 硕士生, 从事机器学习、神经点过程的研究, E-mail: zhiyansong@foxmail.com.

(责任编辑：郑晓蕾)