

回复式神经网络及其应用研究综述

刘丹, 叶茂

(电子科技大学 计算机科学与工程学院, 成都 611731)

E-mail: liudankaicy5@gmail.com

摘要: 回复式神经网络是深度学习领域中一类非常强大的神经网络模型,它主要用于处理和预测序列数据。得益于回复式神经网络的“记忆”功能,它和它的变体网络已经被成功应用于多种任务,比如:机器翻译、语音识别、视频行为识别等。本文首先介绍回复式神经网络的发展历程,之后阐述其网络基本原理、计算过程和网络训练方法。针对传统回复式神经网络存在的短期记忆和梯度消失问题,长短时记忆和门控回复式单元网络结构被详细地分析和对比。然后介绍回复式神经网络的热门应用领域及其相关工作。最后结合近年来工业界和学术界关于回复式神经网络的研究进展,本文总结了回复式神经网络的研究发展趋势。

关键词: 回复式神经网络; 长短时记忆; 门控回复式单元; 序列数据

中图分类号: TP399

文献标识码: A

文章编号: 1000-1220(2020)10-2024-06

Review of Research on Recurrent Neural Network and Its Applications

LIU Dan, YE Mao

(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: Recurrent Neural Network (RNN) is a kind of very powerful neural network model in the field of deep learning, which is mainly used to process and predict sequence data. Benefiting from the “Memory” function, RNN and its variant networks have been successfully applied to a variety of tasks, such as machine translation, speech recognition, video action recognition and so on. This paper first introduces the development of RNN and then elaborates on the basic principles, computational process and how to train the RNN. In terms of the shortcomings of RNN: short-term memory and gradient vanishing, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are analyzed and compared in detail. Then, the popular research areas and the related research works are summarized. Finally, based on the research progress on RNN in industry and academics, this paper discusses the development trend of RNN in the future.

Key words: recurrent neural network; long short-term memory; gated recurrent unit; sequence data

1 引言

自20世纪80年代以来,人工神经网络(Artificial Neural Network, ANN)已经成为人工智能领域的研究热点。ANN是一种旨在模拟大脑神经系统对外界复杂信息的处理机制的数学模型,其在模式识别、智能机器人、自动控制等领域表现出了良好的智能特性。传统的全连接神经网络(Full Connected Neural Network, FCNN),也称前馈神经网络(Feed-forward Neural Network, FNN),如图1(左)所示。它采用单向多层结构,同一层的神经元之间没有相互连接,是彼此独立的。换句话说,对于FNN,一个输入对应一个输出,不同的输入之间是没有联系的。感知机在1957年首次被提出,作为最早的ANN,感知机是一种最简单的网络模型。作为深度FNN的经典网络——卷积神经网络(Convolutional Neural Network, CNN)^[1]依靠其局部连接和共享权值的设计理念,在图像识别、自然语言处理等领域被广泛应用,尤其在手写数字识别任务中取得了良好的识别效果^[2,3]。但是,FNN无法处理在时间

或空间上有前后关联的信息输入问题。所以,传统神经网络的应用受到了限制。

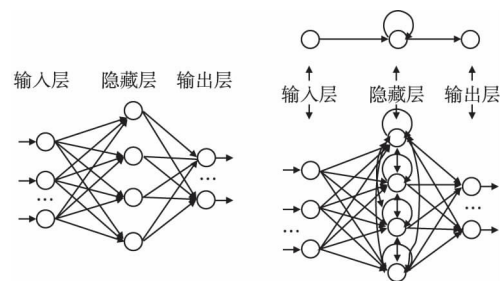


图1 (左) 前馈神经网络和(右) 回复式神经网络

Fig.1 (left) FNN and (right) RNN

针对需要处理序列信息的学习任务,如文本翻译、智能对话、股票行情预测等,它们的输入是有时间顺序的序列数据,回复式神经网络(Recurrent Neural Network, RNN)作为一类用于处理和预测序列数据的神经网络模型应运而生。RNN在

收稿日期: 2020-05-18 收修改稿日期: 2020-06-05 基金项目: 国家自然科学基金-联合基金项目(U181320052)资助; 国家自然科学基金面上项目(6177020680)资助; 国家重点研发计划项目(2018YFC0831801)资助; 四川省重点研发项目(17ZDYF3184)资助; 中德合作项目跨模态学习项目(Sonderforschungsbereich Transregio 169)资助。 作者简介: 刘丹,女,1990年生,博士研究生,CCF会员,研究方向为计算机视觉、深度学习; 叶茂,男,1973年生,博士,教授,研究方向为机器学习、计算机视觉、图像处理、智能机器人。

文本生成、机器翻译、语音识别、视频标记等任务中被广泛地应用研究,并且已经取得丰硕的研究成果^[4-6]。作为神经网络的一个重要模型,RNN 克服了传统机器学习方法对输入输出数据的许多限制,无论从理论上还是应用层面上,深入分析研究回复式神经网络(RNN)具有重要的现实意义。图2是ANN的体系结构图,它包括前馈神经网络和反馈神经网络两大类。

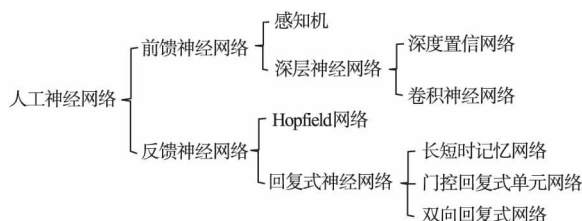


图2 人工神经网络分类

Fig.2 Classification of the ANN

本文将首先阐述 RNN 的基本原理、计算过程和训练方法,然后针对传统 RNN 在实际应用中存在的缺陷,介绍几种流行的优化结构,包括模型结构特点、组件功能;接着描述 RNN 的常用应用领域、相关工作以及对 RNN 的未来研究发展方向展开探讨;最后总结全文内容。

2 RNN

1982 年,美国加州理工学院物理学家 Hopfield 提出 Hopfield 网络模型用以解决组合优化问题,该网络拥有很强的计算能力并且具有联想记忆功能,这是最早的 RNN 的雏形。之后 Jordan 和 Elman 对其进行改进和简化,便有了最初的仅包含单个自连接节点的 RNN 模型。但是,由于梯度消失和梯度爆炸问题,训练 RNN 非常困难,致使 RNN 的应用受到限制。直到 1997 年,Hochreiter 和 Schmidhuber^[7]提出长短期记忆(Long Short-Term Memory,LSTM)网络,该网络使用门控单元以及记忆状态大大缓解了早期 RNN 训练问题,LSTM 模型的提出改进了 RNN 结构,拓宽了 RNN 的应用范围,为后续序列建模的发展奠定了基础。

与传统神经网络相比,RNN 相当于加入了人的先验知识,即将数据信息按照时间顺序输入到对应模型。RNN 最大的特点就是隐藏层神经元的状态不仅依赖于当前时刻的输入,也与当前系统前一时刻的隐藏状态相关,如图1(右)所示,RNN 在隐藏层单元间添加了横向连接,这种结构使得神经元历史信息能够传递到当前神经元中,适合处理时间序列数据,因此 RNN 可以保持数据之间的依赖关系。

RNN 的前向传播计算公式如公式(1)、公式(2)所示,其中 x 表示输入层, h 表示隐藏层, y 表示输出层, W 和 b 分别表示权重参数和偏移项, σ 为激活函数, t 表示时间索引。

$$h_t = \sigma(W_h [x_t, h_{t-1}] + b_h) \quad (1)$$

$$y_t = \sigma(W_y h_t + b_y) \quad (2)$$

由于 RNN 的循环连接方式,将公式(1)反复带入公式(2)中,可以看到 RNN 可以向前查看任意时刻的输入,因此,RNN 具有记忆功能,尤其适用于注重信息时序性的任务。

传统的 FNN 采用反向传播(Back-Propagation,BP)算法

进行有监督地训练,其训练过程包括前向计算、反向各层梯度计算以及参数更新。根据预测输出 y 和真实输出 y' 计算损失函数 $loss$,从后向前计算各参数 ω 关于目标函数 $loss$ 的梯度,即 $\frac{\partial loss}{\partial \omega}$ 。利用梯度下降法(Gradient Descent,GD)更新网络中的所有参数,设网络学习率(Learning Rate,LR)为 α ,则参数 ω 的更新过程如下:

$$\omega = \omega - \alpha \frac{\partial loss}{\partial \omega} \quad (3)$$

对于 RNN,因为要处理时序数据,所以它采用基于时间的反向传播(Back-Propagation Through Time,BPTT)算法。BPTT 算法的基本原理和 BP 算法相同^[8]。RNN 每一时刻 t 都会产生一个损失值 $loss_t$,因此网络的总损失可以表示为 $L = \sum_t loss_t$ 。在 RNN 训练过程中,梯度一方面会按照空间结构传播(输出层、隐藏层、输入层),即 $y_t \rightarrow h_t \rightarrow x_t$,这种情况与 BP 算法一致;另一方面,梯度要沿着时间通道传播,即 $h_t \rightarrow h_{t-1} \rightarrow \dots \rightarrow h_1$,此时,基于公式(1)循环迭代地计算各层梯度,这也是 BPTT 算法名称的由来。表1总结了 BP 算法和 BPTT 算法的特点和训练过程。

表1 BP 算法和 BPTT 算法比较

Table 1 Comparison of BP and BPTT algorithms

算法名称	训练过程	特点
BP 算法	1. 初始化网络参数,计算网络输出误差; 2. 按照空间结构(输出层-隐藏层-输入层)计算各层梯度; 3. 更新梯度。	用来训练多层 FNN,学习输入与输出之间的映射关系;易形成局部最优,收敛速度慢。BP 算法只考虑上层级之间的梯度传播。
BPTT 算法	1. 初始化网络参数,计算网络输出误差; 2. 不仅要沿着空间结构计算梯度,亦要沿着时间通道计算梯度; 3. 更新梯度	BPTT 算法的本质仍是 BP 算法,用来训练 RNN 网络。BPTT 算法考虑上层级和时间通道上的传播,同时在两个方向上进行参数优化。

在计算沿时间方向的梯度时,利用链式法则,会包含多个因子的连乘,即 $\frac{\partial h_t}{\partial h_1} = \frac{\partial h_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial h_{t-2}} \dots \frac{\partial h_2}{\partial h_1}$,加上 RNN 按时间展开后权重共享的设定,BPTT 算法比普通 BP 算法更容易出现梯度消失和梯度爆炸问题^[9]。对于梯度爆炸问题,通常利用梯度截断、添加正则项等技巧可以有效避免。对于梯度消失问题,最常用的手段就是使用跳跃连接(Skip Connection,SC),让梯度能够跨过一些时间步,直接向前传播。常用的 LSTM 网络和门控回复式单元(Gated Recurrent Unit,GRU)网络,都是利用这种 SC 技术缓解梯度消失问题。

3 RNN 的优化结构

RNN 受限于短期记忆问题,根据其隐藏层单元状态的计算方式,越晚的时间步输入对后续影响越大,越早的输入对后续影响越小。如果一个输入序列足够长,RNN 可能会遗漏该序列的前期重要信息,这就是 RNN 存在的短期记忆问题。另

外,在使用 BPTT 算法训练 RNN 时,梯度会在反向传播的过程中快速消失,也就是所谓的梯度消失问题。当前已经存在大量的研究工作用于解决 RNN 存在的这些问题^[7,10-12],其中最具有代表性的优化结构算法是 LSTM 网络和 GRU 网络,它们都使用跳跃连接(SC)的技巧让梯度跨过一些时间步,从而只保留重要的信息。

3.1 LSTM

RNN 隐藏层只有一个状态 h ,不能处理长期依赖问题,LSTM 网络通过在隐藏层添加一个细胞状态 c 来保存长期依赖关系,然后利用门控单元来控制外部信息向长期状态 c 的流入和流出。如果将网络看作一个黑匣子,图 3 则简单明了的说明了 RNN 与 LSTM 的根本区别,即 LSTM 添加了门控单元和存储长期记忆的细胞状态 c 。

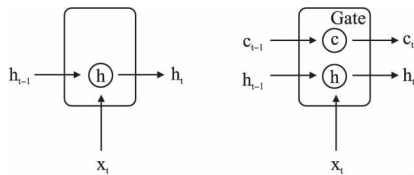


图 3 RNN 和 LSTM 的内部机制示意图

Fig. 3 Internal mechanism diagram of the RNN and LSTM

LSTM 设计三个门控开关来控制外部信息向长期状态 c 的流入和流出。图 4 展示了 LSTM 单元结构。其中,输入门 i_t

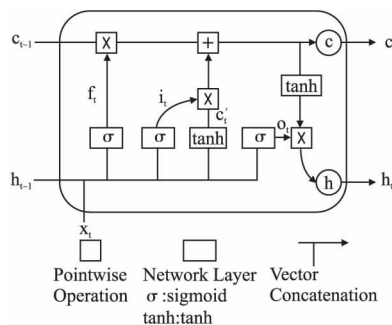


图 4 LSTM 单元结构

Fig. 4 Structure of LSTM cell

控制当前网络输入 x_t 有多少信息可以保存到记忆单元,遗忘门 f_t 决定上一时刻的单元状态 c_{t-1} 有多少信息可以保留到当前状态 c_t ,而输出门 o_t 则用来控制单元状态 c_t 有多少信息可以输出到 LSTM 的当前的输出值 h_t 。三个门的计算公式如下:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (4)$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (5)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (6)$$

其中 $[h_{t-1}, x_t]$ 表示 $t-1$ 时刻的隐藏层状态向量 h_{t-1} 和 t 时刻的输入状态向量 x_t 的拼接。LSTM 隐藏层的最终输出 h_t 的计算方式为:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (7)$$

$$h_t = o_t \odot \tanh(c_t) \quad (8)$$

其中 \odot 代表对应元素相乘。Gers 等人^[13]在 LSTM 的基础上提出窥视孔连接(Peephole Connection, PC),其主要思想是增加三个门控单元和记忆单元的连接,门控单元也可以观察

到记忆单元的内容。但是,研究者们发现^[14]在语言模型任务中,LSTM 和带 PC 的 LSTM 网络性能基本相当,但是后者需要更多的网络参数,所以现在大多数研究工作都不使用带 PC 的 LSTM 网络结构。另外,实验证明输入门、输出门和输入激活函数对 LSTM 网络结构的性能影响很大。

综上所述,LSTM 网络通过引入门控单元(输入门、输出门、遗忘门)和长期记忆细胞状态,有效克服了 RNN 中存在的梯度消失问题,尤其在长距离依赖的序列学习任务中表现优异。

3.2 门控回复式单元网络

GRU 是 RNN 的另一种流行变体,它与 LSTM 单元非常相似,它将 LSTM 中的遗忘门和输入门合成了一个门,并且只使用隐藏状态来传输信息。由于同等规模下,GRU 网络比 LSTM 网络的训练参数少,所有能够有效避免训练过程中的过拟合问题。

在 GRU 中,更新门 z_t 代替了 LSTM 中的遗忘门和输入门,它决定了要丢弃哪些信息和添加哪些信息;另一个门是重置门 r_t ,用于决定丢弃之前信息的程度。图 5 展示了 GRU 的内部结构,更新门 z_t 和重置门 r_t 的计算过程如下所示:

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z) \quad (9)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \quad (10)$$

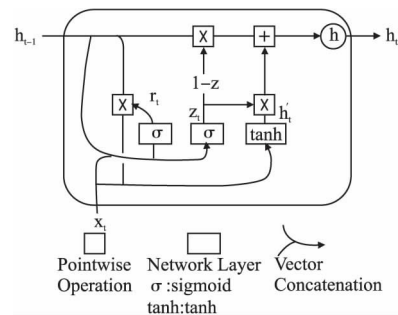


图 5 GRU 单元结构

Fig. 5 Structure of GRU cell

在更新门 z_t 、重置门 r_t 的作用下,当前时刻隐藏层的候选输入状态 h'_t ,当前时刻的隐藏层输出状态 h_t 可由公式(11)、公式(12)计算所得。

$$h'_t = \tanh(W_h[r_t \odot h_{t-1}, x_t] + b_h) \quad (11)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot h'_t \quad (12)$$

综上所述,GRU 网络可以看作是 LSTM 的变体,由于 GRU 网络中的变量参数少,因此可以用更少的时间来训练它。在手写句子识别、复调音乐识别等任务中^[14],GRU 和 LSTM 单元所表现出来的网络性能基本相同,但是在训练大数据集的情况下,LSTM 表达性更优。

3.3 双向 RNN

在经典的 RNN 中,神经元状态是从前向后单向传播的,然而,有些任务当前时刻的输出不仅与之前状态有关,也与后续状态相关。比如,要预测一个句子中缺失的词语,我们必须考虑上下文才能做出正确预测,这时就需要双向的 RNN(Bi-directional RNN, BiRNN)来综合前后文信息。

双向 RNN 的隐藏层要保存两个状态值:参与正向计算的 h_t^f 和参与反向计算的 h_t^b ,最终的输出状态 y_t 取决于隐藏层的

两个状态.

$$h_t^f = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (13)$$

$$h_t^b = \sigma(W_b[x_t, h_{t+1}] + b_b) \quad (14)$$

$$y_t = \sigma(W_y[h_t^f, h_t^b] + b_y) \quad (15)$$

图 6 表示 BiRNN 信息流向示意图. 通俗地讲, BiRNN 的隐藏层由两层神经元组成, 一层处理按时间正向排序的输入序列, 另一层处理反向排序的输入序列, 而输出则由这两者综合计算得来. 将传统的 RNN 单元替换为 LSTM 单元或者 GRU 结构, 则 BiRNN 就变成了双向 LSTM (Bi-directional LSTM, BiLSTM) 网络或者双向 GRU (Bi-directional GRU, BiGRU) 网络.

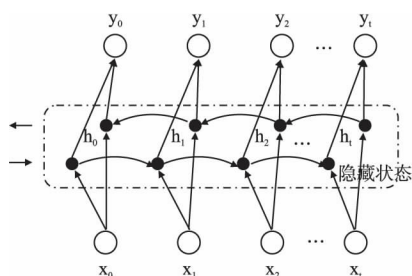


图 6 双向 RNN 示意图

Fig. 6 BiRNN schematic

综合来说, LSTM 和 GRU 都能通过门控单元将时间序列信息中的重要特征保留, 保证在长期的反向传播中梯度不会消失. 在实际应用中, 需要根据具体任务选择具体模型结构, 不能笼统地说哪一种网络结构更优.

4 应用场景

4.1 机器翻译

机器翻译 (Machine Translation, MT) 是指利用计算机实现将源语言转换成目标语言的过程, 是自然语言处理领域中的一个重要应用方向, 具有很高的科学研究价值. 传统的机器翻译方法需要依靠大量的人力和时间对众多平行语料进行分析, 最终构造翻译系统. 基于 RNN 的机器翻译模型, 往往通过一个 RNN 将源语言输入序列 (x_1, x_2, \dots, x_T) 转换成固定长度的特征向量 v , 然后通过另一个 RNN 将该固定向量转换成目标语言输出序列 (y_1, y_2, \dots, y_T) . 根据条件概率公式, 可以得到:

$$p(y_1, y_2, \dots, y_T | x_1, x_2, \dots, x_T) = \prod_{t=1}^T p(y_t | v, y_1, \dots, y_{t-1}) \quad (16)$$

编码-解码器 (Encoder-Decoder) 模型^[15, 16] 就是由两个 RNN 构造而成的机器翻译模型, 也可以叫做 Seq2Seq 模型. 该模型最重要的地方在于输入序列和输出序列的长度是可变的, 可以用于翻译、聊天机器人、句法分析、文本摘要等任务.

4.2 语音识别

语音识别 (Speech Recognition, SR) 就是将声音波形转换成计算机可读的语言, 进而通过解码技术转变为人类能够理解的文本的过程. RNN 能够对语音的长时相关性进行建模, 从而提高识别准确率^[17]. 2015 年, 百度公开发布的采用 LSTM + CTC (Connectionist Temporal Classification, CTC)^[18]

模型大幅度降低了语音识别的错误率, 采用这种技术在安静环境下的标准普通话的识别率接近 97%. Arisoy 等人^[19] 提出使用 BiRNN 来解决语音识别语言建模任务, 并证明了 BiRNN 在处理语音识别任务中优于单向的 RNN 结构. Tian 等人^[20] 建议利用层级训练和指数移动平均方法来构造一个深层 LSTM 网络, 它在语音识别任务中取得优异成绩.

4.3 行为识别

行为识别 (Action Recognition, AR) 是计算机视觉领域的一个重要研究课题, 它分为两类任务: 一类是给定一段视频, 需要定位其中每个行为的起始时间, 并为每个行为分配类别标签; 另一类是给定一段已经被预先剪切好地数据, 只需要为该数据分配一个行为类别标签即可. 行为视频中的帧与帧之间的时序关系在识别行为的过程中至关重要, 而 RNN 能够学习到长期的时间依赖关系, 已经被广泛地应用在该任务地学习中. Donahue 等人^[21] 利用 LSTM 网络学习视频中的时间动力学和长期依赖关系. Liu 等人^[22] 使用 RNN 来建模行为序列, 利用序列预测学习的方法来解决行为识别问题.

4.4 情感分析

情感分析 (Sentiment Analysis, SA) 就是判断一段文本所表达的情绪状态, 可以是正面、负面, 也可以是开心、悲伤等. 该任务属于典型的文本分类问题, 即将一段变长的文本序列映射为文本的类别. 为了学习到一个好的情感分析系统, 算法必须考虑文本中单词的顺序. Can 等人^[23] 提出利用 RNN 来训练一个情感分析模型. 另外, 一些研究者^[24] 利用多模态数据 (文本、视频、音频等) 训练多种 RNN 变体结构来解决情感分析任务.

4.5 其它应用

RNN 作为深度学习算法中极为重要的一种, 已经在众多领域大放异彩. 除了上述介绍的几种应用场景外, RNN 在对话 (问答) 系统、图像描述、推荐系统、步态识别、目标检测等领域也取得了优异成果^[25-29]. 换句话说, 一切与时序相关的计算机任务都可以尝试使用 RNN 来处理学习.

5 RNN 的发展趋势

5.1 设计新的变体结构

从 RNN 到 LSTM、GRU 单元结构, 传统的循环网络通过增加一些门控单元在一定程度上解决了 RNN 训练梯度消失问题和短期记忆问题. 但是, 这些复杂的单元内部结构以及循环网络本身的时序输入性, 严重影响了网络的训练速度. 如何有效提高循环网络训练效率、简化单元内部结构, 依然是企业和学术研究者要解决的问题.

Bradbury 等人^[30] 提出类回复式神经网络 (Quasi-Recurrent Neural Network, QRNN), 它交替使用跨时间步并行处理的卷积层来进行序列建模. 实验表明 QRNN 在保持算法性能的基础上有效提高了网络训练和测试的速度. 同理, 研究者们设计只与当前时刻输入有关的门控单元来简化 LSTM 单元内部结构. 提出简单回复式单元网络 (Simple Recurrent Unit, SRU), SRU 的门控单元丢弃时间依赖关系, 但记忆单元仍保持时间依赖关系. 这种设计能够在网络性能和训练速度之间获得权衡^[31]. Bouaziz 等人^[32] 在原始 LSTM 基础上, 设计并

行的 LSTM 来执行多个并行同步输入序列,预测相应的输出.上海交通大学的研究者^[33]提出切片回复式神经网络(Sliced Recurrent Neural Network, SRNN),它可以通过将序列分割成多个子序列来实现计算并行化,在不改变循环单元的情况下,SRNN 的训练速度是标准 RNN 的 136 倍.

5.2 引入注意力机制

人脑神经机制的一个重要特征就是选择性,它表示人类过滤掉无用信息的能力,这就是所谓的注意力机制^[34,35].RNN 通过添加跨越时间点的自连接对时序数据建模,但是对于时序信息来说,不同时刻的输入的重要性对于学习任务来说是不一样的.注意力机制已经被广泛应用在自然语言处理、图像识别等各个领域,显著提升了模型表现能力.

He 等人^[36]提出一种新型的基于注意力机制的混合神经网络模型,该模型在 BiLSTM 网络的基础上添加单词级的注意力模块来提取文本语义信息. Huang 等人^[37]将注意力机制应用在情感分析任务中,提出利用多模态注意力融合模块探索图像和文本之间的内部关联性,甚至在弱标签的数据集上也验证了算法的有效性.除此之外, Sudhakaran 等人^[38]利用长短时注意力机制(Long Short-Term Attention, LSTA)学习行为视频的时间和空间特征. Xu 等人^[39]提出基于注意力机制的 Encoder-Decoder 模型用于图像描述任务,实验证明注意力机制为模型生成过程提供了更多可解释性,如图 7 所示.

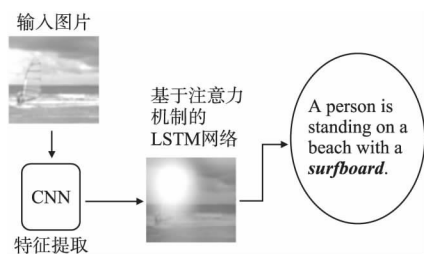


图 7 注意力的可视化

Fig. 7 Visualization of attentional map

5.3 结合其它网络

普通 RNN 主要着重对时序信息进行建模,然而,现实世界中的许多任务不仅仅与时序相关,往往集合多种网络结构优势才能更好的提升算法性能.比如, FCNN 可以将输入数据映射到更离散的空间,经常在各种网络的最上层使用. CNN 擅长学习局部信息,能够最大限度地过滤输入数据.随着深度学习热度的延续,更灵活的组合方式、更多的网络结构将会被开发出来.

Donahue 等人^[40]首次将 CNN 与 LSTM 进行结合,提出一种新型回复式卷积结构,该模型在视频识别与视频描述任务上都取得了很好的效果. Sainath 等人^[41]同时使用卷积结构、循环结构与全连接结构(即 CNN、RNN、FCNN)构建神经网络模型,该模型在论文中的语音识别任务上比 LSTM 性能提升 4-6%,充分地利用了几种网络结构的互补性. Minaee 等人^[42]结合 CNN 和 LSTM 网络来解决情感分析任务, CNN 网络学习文本的局部结构特征,而 LSTM 网络学习时序关系. Lee 等人^[43]提出条件随机场(Conditional Random Field, CRF)和 RNN 的结合使用,用以解决命名实体识别问题.

6 结束语

目前,回复式神经网络已经成为深度学习领域中一类非常重要的网络模型,本文对其进行展开讨论.首先介绍了 RNN 的发展历程,并阐述其基本原理、网络计算方式和网络训练方法.其次,面对传统 RNN 在实际应用中存在的短期记忆和梯度消失问题,详细分析了 RNN 中被广泛使用的变体结构: LSTM、GRU 和其对应的双向变体结构.接着总结了 RNN 的热门研究领域并分析相关研究工作.最后,结合近几年工业界和学术圈对 RNN 的研究进展,对 RNN 的发展趋势进行总结介绍,这也是我们未来研究的方向.

References:

- [1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems (NIPS) 2012: 1097-1105.
- [2] He K, Gkioxari G, Dollár P et al. Mask r-cnn [C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017: 2961-2969.
- [3] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences [J]. arXiv preprint arXiv: 1404. 2188 2014.
- [4] Hori T, Cho J, Watanabe S. End-to-end speech recognition with word-based RNN language models [C]//2018 IEEE Spoken Language Technology Workshop (SLT) 2018: 389-396.
- [5] Wang R, Panju M, Gohari M. Classification-based RNN machine translation using GRUs [J]. arXiv preprint arXiv: 1703. 07841, 2017.
- [6] Brito E, Lübbering M, Biesner D et al. Towards supervised extractive text summarization via RNN-based sequence classification [J]. arXiv preprint arXiv: 1911. 06121 2019.
- [7] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [8] Werbos P J. Backpropagation through time: what it does and how to do it [J]. Proceedings of the IEEE, 1990, 78(10): 1550-1560, doi: 10.1109/5. 58337.
- [9] Bengio Y, Simard P Y, Frasconi P et al. Learning long-term dependencies with gradient descent is difficult [J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.
- [10] Gers F A, Eck D, Schmidhuber J. Applying LSTM to time series predictable through time-window approaches [C]//Artificial Neural Networks (ICANN) 2001: 669-676.
- [11] Gers F A, Schmidhuber J, Cummins F et al. Learning to forget: continual prediction with LSTM [J]. Neural Computation, 2000, 12(10): 2451-2471.
- [12] Lei T, Zhang Y, Wang S I et al. Simple recurrent units for highly parallelizable recurrence [C]//Empirical Methods in Natural Language Processing 2018: 4470-4481.
- [13] Gers F A, Schmidhuber J. Recurrent nets that time and count [C]//Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN 2000), Neural Computing: New Challenges and Perspectives for the New Millennium 2000, 3: 189-194.
- [14] Chung J, Gulcehre C, Cho K H et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. arXiv preprint

- arXiv: 1412.3555 2014.
- [15] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv: 1406.1078 2014.
- [16] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Advances in Neural Information Processing Systems(NIPS) 2014: 3104-3112.
- [17] Chiu C C, Sainath T N, Wu Y, et al. State-of-the-art speech recognition with sequence-to-sequence models[C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018: 4774-4778.
- [18] 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(ICML) 2006: 369-376.
- [19] Arisoy E, Sethy A, Ramabhadran B, et al. Bidirectional recurrent neural network language models for automatic speech recognition[C]//2015 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP) IEEE 2015: 5421-5425.
- [20] Tian X, Zhang J, Ma Z, et al. Deep LSTM for large vocabulary continuous speech recognition[J]. arXiv preprint arXiv: 1703.07090 2017.
- [21] Donahue J, Anne Hendricks L, Guadarrama S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2015: 2625-2634.
- [22] Liu D, Ye M, Zhang J, et al. Improving action recognition using sequence prediction learning[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020: 2050029, <https://doi.org/10.1142/S0218001420500299>.
- [23] Can E F, Ezen-Can A, Can F. Multilingual sentiment analysis: an rnn-based framework for limited data[J]. arXiv preprint arXiv: 1806.04511 2018.
- [24] Agarwal A, Yadav A, Vishwakarma D K. Multimodal sentiment analysis via RNN variants[C]//2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD) 2019: 19-23.
- [25] Wang S, Jiang J. Machine comprehension using match-lstm and answer pointer[J]. arXiv preprint arXiv: 1608.07905 2016.
- [26] Gao L, Guo Z, Zhang H, et al. Video captioning with attention-based LSTM and semantic consistency[J]. IEEE Transactions on Multimedia 2017, 19(9): 2045-2055.
- [27] Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks[J]. arXiv preprint arXiv: 1511.06939 2015.
- [28] Liu D, Ye M, Li X, et al. Memory-based gait recognition[C]//British Machine Vision Conference(BMVC) 2016: 1-12.
- [29] Li X, Ye M, Liu D, et al. Memory-based object detection in surveillance scenes[C]//IEEE International Conference on Multimedia and Expo(ICME) 2016: 1-6.
- [30] Bradbury J, Merity S, Xiong C, et al. Quasi-recurrent neural networks[C]//International Conference on Learning Representations (ICLR) 2017.
- [31] Lei T, Zhang Y, Wang S I, et al. Training RNNs as fast as CNNs[C]//Empirical Methods in Natural Language Processing 2018.
- [32] Bouaziz M, Morchid M, Dufour R, et al. Parallel long short-term memory for multi-stream classification[C]//IEEE Spoken Language Technology Workshop(SLT) 2016: 218-223.
- [33] Yu Z, Liu G. Sliced recurrent neural networks[J]. arXiv preprint arXiv: 1807.02291 2018.
- [34] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems (NIPS) 2017: 5998-6008.
- [35] Desimone R, Duncan J S. Neural mechanisms of selective visual attention[J]. Annual Review of Neuroscience, 1995, 18(1): 193-222.
- [36] He D, Zhang H, Hao W, et al. An attention-based hybrid neural network for document modeling[J]. IEICE Transactions on Information and Systems 2017, 100(6): 1372-1375.
- [37] Huang F, Zhang X, Zhao Z, et al. Image-text sentiment analysis via deep multimodal attentive fusion[J]. Knowledge Based Systems, 2019, 167(1): 26-37.
- [38] Sudhakaran S, Escalera S, Lanz O. Lsta: long short-term attention for egocentric action recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2019: 9954-9963.
- [39] Xu K, Ba J, Kiros R, et al. Show, attend and tell: neural image caption generation with visual attention[C]//International Conference on Machine Learning(ICML) 2015: 2048-2057.
- [40] Donahue J, Anne Hendricks L, Guadarrama S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2015: 2625-2634.
- [41] Sainath T N, Vinyals O, Senior A, et al. Convolutional long short-term memory, fully connected deep neural networks[C]//2015 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP) 2015: 4580-4584.
- [42] Minaee S, Azimi E, Abdolrashidi A A. Deep-sentiment: sentiment analysis using ensemble of CNN and Bi-LSTM models[J]. arXiv preprint arXiv: 1904.04206 2019.
- [43] Lee C. LSTM-CRF models for named entity recognition[J]. IEICE Transactions on Information and Systems 2017, 100(4): 882-887.