doi:10.3772/j.issn.1002-0470.2017.01.004

大数据下的深度学习研究①

王金甲② 陈 浩 刘青玉

(燕山大学信息科学与工程学院 秦皇岛 066004)

摘 要 给出了大数据和机器学习的子领域——深度学习的概念,阐述了深度学习对获取大数据中的有价值信息的重要作用。描述了大数据下利用图像处理单元(GPU)进行并行运算的深度学习框架,对其中的大规模卷积神经网络(CNN)、大规模深度置信网络(DBN)和大规模递归神经网络(RNN)进行了重点论述。分析了大数据的容量、多样性、速率特征,介绍了大规模数据、多样性数据、高速率数据下的深度学习方法。展望了大数据背景下深度学习的发展前景,指出在不远的将来,大数据与深度学习融合的技术将会在计算机视觉、机器智能等多个领域获得突破性进展。

关键词 大数据,深度学习,卷积神经网络(CNN),深度置信网络(DBN),递归神经网络(RNN)

0 引 言

大数据(big data),是指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。在生活中的各个方面,时时刻刻都有数据在生成、传输、接收和处理。社交网络、移动设备、物联网及云计算所涉及到的数据量正在呈指数级增长。在大数据时代,国家拥有数据的规模和运用数据的能力将成为国家综合国力的重要组成部分,对数据的占有和控制将成为国家间和企业间新的争夺焦点[1]。对于人们来说,如何有效地利用大数据并从其中获取有价值的信息,是一个严峻的挑战,而机器学习,尤其是深度学习,以及日渐进步的计算能力,将成为开启大数据宝库的一把钥匙。

深度学习是机器学习中较为热门的子领域。深度学习构建了具有很多隐层的机器学习模型不是只包含一层隐层节点的浅层模型,并通过监督/非监督的方式训练海量的数据,自动地学习多层结构中的特征,用来提升分类或预测的准确性,因而可广泛应用于图像分类[2,3]、语音识别[4,5]、推荐系统[6]等领域。深度学习只需要很少的手工工程,并且很容易受益于可用计算能力和数据量的增加[7],这两点都有助于深度学习取得更多的成功。本文对大数据下的深度学习的研究进行了综述,而且也预测了研究的未来趋势。

1 深度学习介绍

深度学习架构由多层非线性运算单元组成,每个较低层的输出作为较高层的输入,可以从大量输入数据中学习有效的特征表示,学习到的高阶表示中包含输入数据的许多结构信息。深度学习是一种从数据中提取表示的好方法,能够用于分类、回归和信息检索等特定问题中[8]。本节介绍三种深度模型:卷积神经网络(convolutional neural network,CNN)、深度置信网络(deep belief network,DBN)和递归神经网络(recurrent neural network,RNN)。

① 国家自然科学基金(61273019, 61473339),中国博士后科学基金(2014M561202),河北省博士后专项(B2014010005)和首批"河北省青年拔尖人才"([1013]17)资助项目。

② 男,1978 年生,博士,教授;研究方向:信号处理,模式识别及其应用;联系人,E-mail:wjj@ysu.edu.cn

(收稿日期:2016-04-15)

1.1　卷积神经网络(CNN)

本节以卷积神经网络 LeNet-5 为例介绍 CNN。如图 1 所示,卷积神经网络的输入层为一个二维数组,大小为 32 × 32,随后是两组交替的卷积层和下采样层。每个卷积层都由多个特征层组成,特征层中的每个神经元的输入与前一层的局部感受野相连,大大减少了需要学习的参数个数并提取该局部的特征。在 C1 中,选定的卷积核大小为 5 × 5,通过使用不同的卷积核对输入的二维数组进行卷积并加上偏置后经过激活函数得到特征映射图,用公式表示为

$$y_j^{(l)} = f\left(\sum_i K_{ij} \otimes x_i^{(l-1)} + b_j\right) \tag{1}$$

其中,$y_j^{(l)}$ 是第 l 个卷积层 C_l 的第 j 个输出;K_{ij} 表示卷积核,与之前一层(即 $x_i^{(l-1)}$ 层)的特征映射图的全部或部分相连,并与 $x_i^{(l-1)}$ 层的特征映射图卷积得到当前层的特征映射图;符号⊗表示离散卷积运算;b_j 表示偏移量;$f(\cdot)$ 是一个非线性的激活函数,最常用的是双曲正切函数和 sigmoid 函数。由于相邻的接收域相互重合,所以处理后的特征映射图大小为 28 × 28。S2 对 C1 进行下采样处理,下采样并不改变特征映射图的数目,只是将特征映射图变小,使特征映射图的输出对平移、缩放、旋转等变换的敏感度下降,效果是在减少数据处理量的同时保留有用信息。若采样窗口大小为,经过一次下采样后,特征映射图的大小变为原来特征映射图大小的 $1/n \times 1/n$。下采样一般的表达式为

$$y_j^l = f(\beta_j^l down(y_j^{l-1}) + b_j^l) \tag{2}$$

其中,y_j^l 和 y_j^{l-1} 分别表示当前层和前一层的第 j 个特征映射图;$down(\cdot)$ 表示一个下采样函数;β_j^l 和 b_j^l 分别表示当前层第 j 个特征映射图的乘性偏置和加性偏置,$f(\cdot)$ 为激活函数。下采样采用 2 × 2 不重合窗与 C1 中对应特征映射图相连接,采样结束后的特征映射图大小为 14 × 14。C3、S4 与 C1、S2 的过程相似。F5 和 F6 为全连接层,最后输出由欧式径向基函数(euclidean radial basis function,ERBF)单元组成[9]。

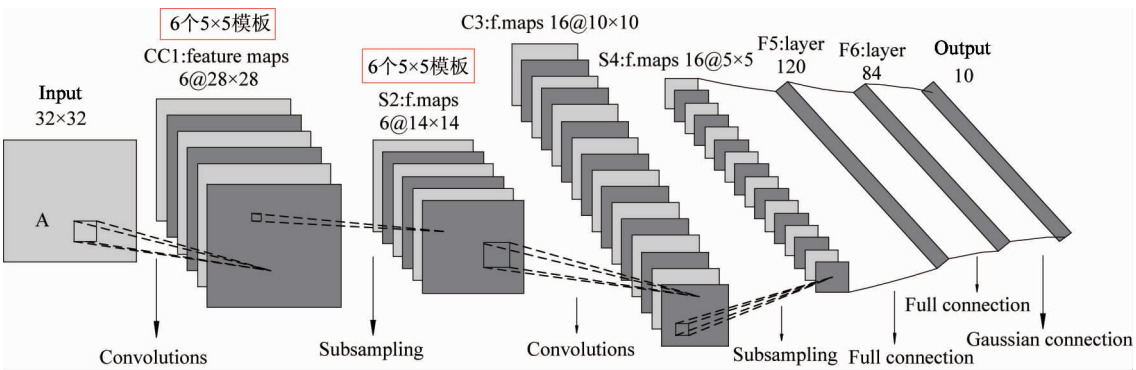

图 1　LeNet5 示意图

卷积神经网络的训练算法与传统的反向传播(back propagation,BP)算法类似,即从样本集中取一个样本输入网络,经过 CNN 后得到输出,将其与理想输出作对比,得到误差函数并进行反向传播,随后用随机梯度下降法(stochastic gradient descent,SGD)调整卷积参数和偏置量,以达到收敛状态或最大迭代次数为止。

1.2　深度置信网络(DBN)

深度置信网络模型由多层无监督的受限玻尔兹曼机(restricted boltzmann machine,RBM)和一层有监督的 BP 网络组成,结合了无监督学习和有监督学习各自的优点。DBN 的训练过程分为预训练和微调两个阶段[10]。RBM 是一个两层结构模型,底层是由可见节点 $v = \{v_1, v_2, v_3, \cdots, v_i\}$ 组成的可见层,顶层是由隐藏节点 $h = \{h_1, h_2, h_3, \cdots, h_i\}$ 组成的隐藏层,两层之间的节点全连接,每一个连接都有一个权值,且同一个 RBM 内所有的权值都相同,但处于同一层的节点之间互不连接。

深度置信网络的训练过程如图 2 所示。在进行预训练时,第一层 RBM 网络的输入是原始数据,它

由可见层 V_0 和隐藏层 H_0 组成。第一层 RBM 网络完成 V_0 与 H_0 之间权值 W_0 的学习后[11],进入第二层 RBM 网络的训练,此时 $H_0 = V_1$。随后完成 V_1 与 H_1 之间权值 W_1 的学习。依此类推,$H_1 = V_2, \cdots, H_{n-1} = V_n$,在此过程中得到各个 RBM 之间的权值 $W_i\{i = 1,2,\cdots,n\}$,即 DBN 各隐藏层之间的参数。这时训练出的参数仅仅是局部最优值,需要进行反向微调才能达到 DBN 整体最优效果。反向微调时,DBN 利用 BP 网络有监督地训练最顶层的 softmax 网络,对 RBM 网络学习到的特征进行分类,将 DBN 的实际输出与预期输出之间的误差逐层反向传递至所有 RBM 网络,微调 RBM 网络层间的参数,得到最优的 DBN[12]。

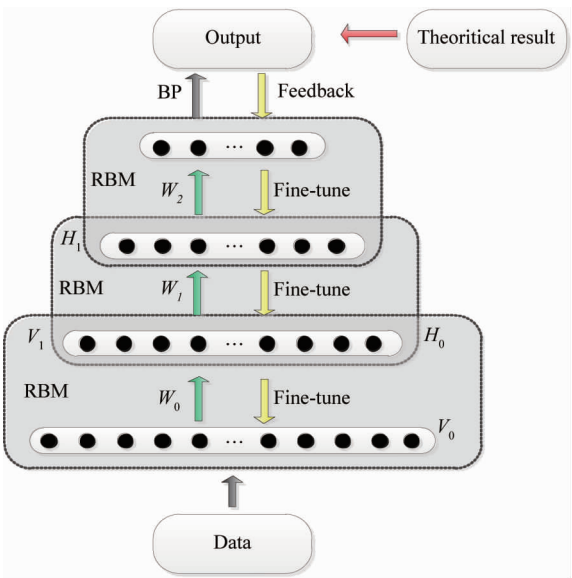

图 2　DBN 训练示意图

1.3　递归神经网络(RNN)

RNN 由输入层、隐藏层和输出层构成,结构如图 3 所示。一条单向流动的信息流从输入单元到达隐藏单元,同时另一条单向流动的信息流从隐藏单元到达输出单元。有时 RNN 会打破后者的限制,引导信息流从输出单元返回隐藏单元,这种现象被称为“反投影”,隐藏层的输入还包括上一隐藏层的状态,即隐藏层内的节点可以自连或互连。

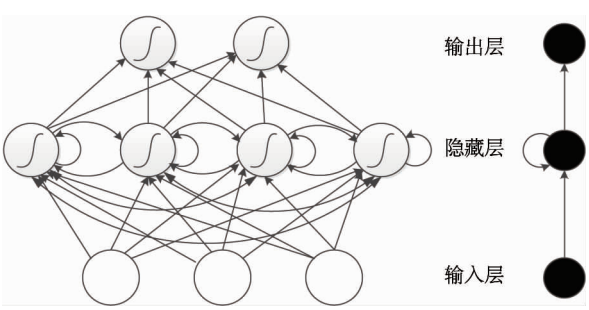

图 3　RNN 简单结构

展开的 RNN 如图 4 所示。在第 t 步中,I_t,$t = \{1,2,\cdots,n\}$ 是网络的输入,H_t 为隐藏层的状态,O_t 为输出。RNN 共享一组 $\boldsymbol{P}$、$\boldsymbol{Q}$、$\boldsymbol{R}$ 权值矩阵,大大降低了网络中参数的数量。H_t 和 O_t 的表达式为

$$H_t = f(\boldsymbol{P}I_t + \boldsymbol{R}H_{t-1}) \tag{3}$$

$$O_t = \text{softmax}(\boldsymbol{Q}H_t) \tag{4}$$

其中,f 一般是非线性激活函数,如 tanh 或 ReLU。在计算 H_0 时,没有上一步隐藏层的状态,因此将上一步的隐藏层状态设为零。输出 O_t 与 $\boldsymbol{Q}$、H_t 有关,为了降低网络复杂度,H_t 一般只包括前面若干步隐藏层状态[13,14]。

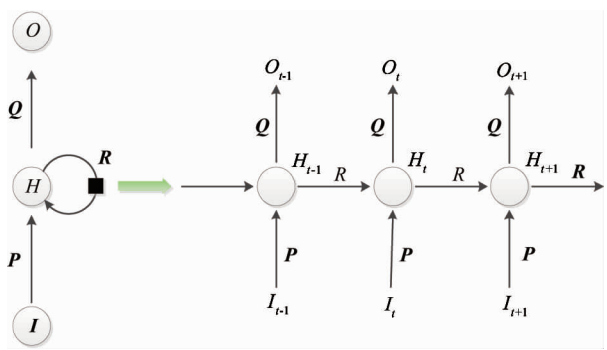

图 4　展开的 RNN

2　大数据下的深度学习框架

随着数据规模的急剧增长,人们对算法的要求也在不断提高。虽然深度学习在某些领域有着不俗的表现,但其内在的迭代运算在处理大数据时很难通过并行运算来提高处理速度。深度学习算法通常涉及到大规模的隐藏神经元以及数以百万级的参数,可以处理海量的数据并训练复杂模型。近期,研究人员结合深度学习算法提出了一些处理大规模数据的方法:例如 Sainath 等改进深度学习的内部结构,将 ReLU + dropout 结构与原结构结合,提高了大规模语音识别的准确率[15];Lv 等采用堆栈自动编码器(stacked autoencoders,SAEs)实现深度结构用

于交通流的预测[16]；雷亚国等使用去噪自动编码器训练深度神经网络(deep neural network，DNN)，对设备产生的大规模即时数据进行处理，完成了对设备的健康监测[17]；Wang 等提出了扩展的且鲁棒的深度学习框架 DistDL，该框架中的数据分层分割结构和无缝通道参数节点可以很好地达到数据/模型的并行性，并且 DistDL 利用布隆过滤器(bloom filter)在每个工作节点对数据编码和计算信息建立点阵图(Bitmap)，减少了运算开销并且实现了数据移动。此外，DistDL 还可以通过 GPU 异构资源对训练过程进行加速[18]。

下面将介绍在大规模数据背景下利用 GPU 进行并行运算的深度学习框架。

图像处理单元(graphic processors units，GPU)是以其内部计算资源为主的通用大规模并行处理器。这里以基于统一计算设备架构(compute unified device architecture，CUDA)的 GPU 为例来进行介绍。GPU 处理的基本单元是线程(Thread)，继而组成线程块(Block)、线程格(Grid)。图 5 表示的是 GPU 内存与线程的关系。每个 Thread 都有专属的寄存器和本地内存，每个 Block 内有共享内存，其内部的 Thread 都可访问。运行中的 Thread 可访问设备的全局内存(global memory，GM)。

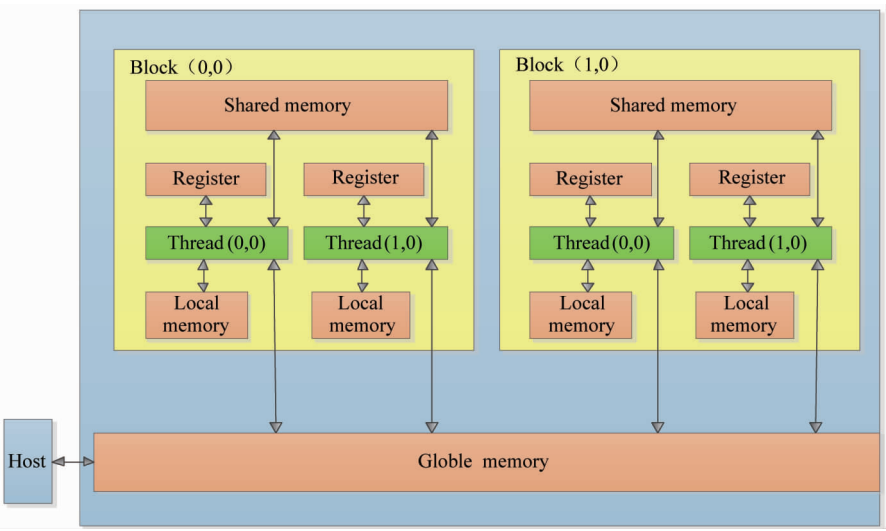

图 5　GPU 内存与线程的联系

2.1　大规模卷积神经网络

将卷积神经网络应用于大规模数据处理的过程中，使用了多核的 GPU。训练过程中所需要的线程数由所选滤波器的大小所决定。例如选择大小为 8 ×8 =64 的滤波器时，可以由 64 个线程同时对其进行处理。梯度下降算法决定了滤波器的权重。由于滤波器内核的子采样操作，卷积层反向传播并不是直接实现的。Simard 等人提出的“推动/牵拉”(pushing or pulling)方法解决了这个问题。公式为

$$\delta_{j+i}^{L} += w_i \delta_j^{L+1} (i = 1,2,3,4) \tag{5}$$

实现了反向传播中误差的计算，式中 $w_i(i = 0,1,2,3)$ 是连接 L 层 $j+i$ 单元与 $L+1$ 层第 j 单元的权重，δ_{j+i}^{L} 表示第 L 层第 $j+i$ 个神经元的误差。

反向传播误差计算如图 6 所示，$L+1$ 层的误差信号与“途中”权重因子 w_i 相结合后“推动”(pushing)至 L 层对应的误差信号，且 $L+1$ 层(较高层)“推动”至 L 层(较低层)的连接数是固定的。反之

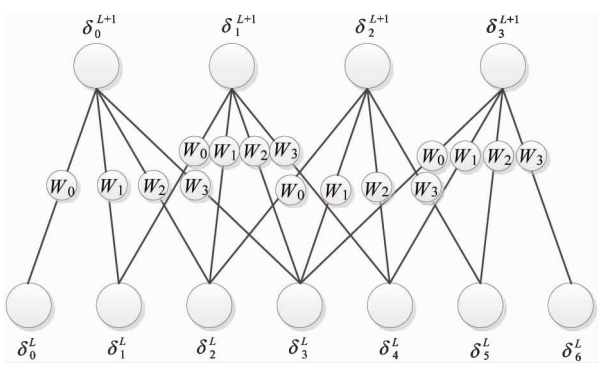

图 6　反向传播误差计算示意图

也可以理解为 L 层“牵拉”着 $L+1$ 层的误差信号。由于边界的限制,连接数是变化的(从 1 到 4 不等)[19]。

在实现阶段,程序被分为几块较小的可独立解决的部分以对使用统一计算设备架构(CUDA)框架的应用进行加速,一个 Block 被用来进行一个深度学习模型训练。因为 GPU 内分享内存的大小有着严格限制,所以在并行运算时要尽可能多地对加载数据进行再利用:在一个 Block 中将源特征映射图同时卷积运算至 8 个目标特征映射图,这 8 个目标特征映射图只需加载一次源特征映射图的活动即可完成模型的训练,这极大地减少了内存传输的数量[20];Scherer 等人同时提出将共享内存用作环形缓冲器的方法,即共享内存中只保留源特征映射图的一小部分,同样节省了内存。此外, GPU 内存有如下限制:反向传播时需要存储每张特征映射图和每个模式的活动与错误信号。对此 Scherer 等提出了一种改进的结构:将卷积与子采样一起进行实现。当与处理步骤相结合后,每张特征映射图所用的内存可以降为原来的 1/4,从而使并行运算更快进行[20]。

GPU 中应用所需的空间超过有限的内存时,需要将数据拷贝到 CPU 内存中,而 GPU-CPU 之间的传输速率相对 GPU 的处理速度来说十分缓慢,因此 Satish 等提出了将数据传输模块化为整数线性程序(integer linear program,ILP)及改进的模拟退火法/混合整数线性程序(simulated annealing/mixed integer linear program, SA/MILP)算法,明显减少了 GPU-CPU 之间的数据传输[21]。在图像处理时,他们的方法通过对数据移动的自动管理,相较未经优化的方法获得了 30 倍的数据传输量的下降,这对于 CNN 处理大规模数据的并行运算起到了有力的推动作用[21]。

2.2 大规模深度置信网络

大规模深度置信网络(DBN)的训练涉及到众多独立的受限玻尔兹曼机(RBM)和数以百万计的参数,并且大规模 DBN 中并行运算占主导地位,导致传统的 CPU 不能有效地对算法实现细粒度并行处理。而 GPU 内在的并行结构十分适合在 Block 和 Threads 两个层级结构中进行大规模的并行运算。

近期,Noel 等人提出了一种基于高度可扩展 GPU 并行运算的 CD-K 算法。CD-K 是由 Hindon 等人提出的在 RBM 抽样过程中使用的方法[22,23]。Noel 提出的算法设计了三种 CUDA 核:ComputeStatusHiddenUnits(下文简称 H)、ComputeStatusVisibleUnits(下文简称 V)和 CorrectWeights(下文简称 W)。为了充分地利用 GPU 的计算能力,他们将两个神经元之间的一个连接作为 H 核和 V 核的最小计算单位,这个决定利用了文献[24]中在 BP 层中计算隐藏层单元的算法,并与其相似,从而可以认为神经元之间的连接通过其权重进行了一个简单的函数运算,即与“压紧输入”相乘。这种情况下,每个 Block 代表一个神经元并利用快速共享内存(fast shared memory)求得每个 Thread 计算出来的值的和,并且计算出活跃样本的神经元输出值。在决定权重矩阵 $\boldsymbol{W}$ 时,其排列顺序会影响 H 和 V。在合并法则下,行优先(row-major)会加速前者,而后者不变,因为 H 核内运算时使用到的权重 W_{ji} 是按照行排列的,j 不变,i 从 1 到 I 变化,行优先可以短时间内将所需的参数全部调用,从而加快 H 核内的运算;在列优先(column-major)时,加速效果与行优先相反。

如图 7 所示,由于访问 H 核较多,作者将 $\boldsymbol{W}$ 矩阵设置为行优先。大部分的工作由 W 核处理,主要是将偏差值(ΔW_{ji},Δb_j 和 Δc_i)

$$\Delta W_{ji} = \gamma(\langle v_i h_j\rangle_0 - \langle v_i h_j\rangle_k) \tag{6}$$

$$\Delta b_j = \gamma(\langle h_j\rangle_0 - \langle h_j\rangle_k) \tag{7}$$

$$\Delta c_i = \gamma(\langle v_i\rangle_0 - \langle v_i\rangle_k) \tag{8}$$

合计来更新权值。γ 表示学习速率,v_i 和 h_j 表示一个 RBM 中显层的第 i 个节点和隐层的第 j 个节点。$\langle\cdot\rangle_0$ 表示在数据分布($p_0 = p(\boldsymbol{h}\mid\boldsymbol{v})$)下的期望,$\langle\cdot\rangle_\infty$ 表示在模型分布($p_\infty = p(\boldsymbol{h}\mid\boldsymbol{v})$)下的期望。$\boldsymbol{h}$ 和 $\boldsymbol{v}$ 分别表示一个 RBM 中的隐性单元和显性单元,其中 $\boldsymbol{v}\in\{0,1\}^I$,$\boldsymbol{h}\in\{0,1\}^J$(假设隐层有 I 个单位,显层有 J 个单位)。

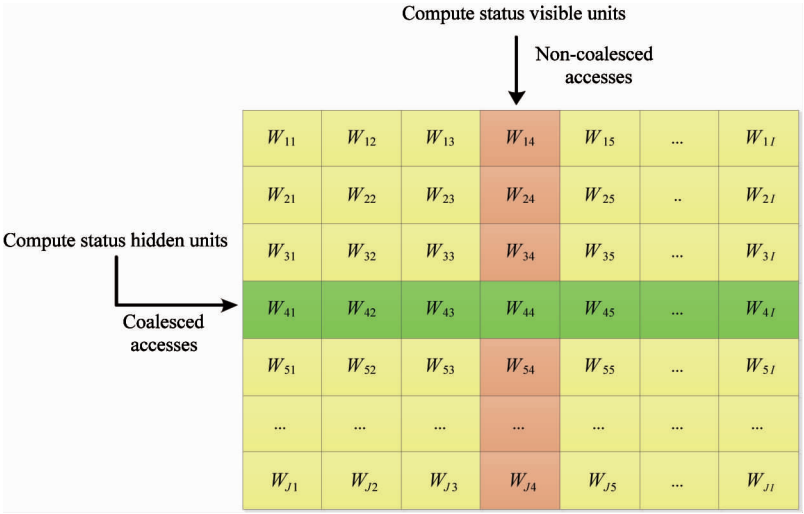

图 7　行优先(row-major)示意图

实现 W 核的方法如下:(1)为每个连接建立一个 Block,Block 内每个 Thread 对一个或多个样本(取决于实际样本数)的值进行求和运算。(2)进行删减过程(reduction process)计算 delta(),进行权重和偏差的更新。随后利用 MNIST 数据集来进行测试,数据集包括 60000 个训练样本和 10000 个测试样本,每个样本是手写数字并由 $28 \times 28 = 784$ 个像素点组成。测试取训练样本进行,在一个 RBM 学习周期内,CPU 用时超过 40min,而 GPU 用时仅约为 53s,在此次测试中 GPU 较 CPU 约加速了 45 倍[25]。

2.3　大规模的递归神经网络

递归神经网络(RNN)是一个配有额外循环连接的神经网络。这样一个独特的网络结构能够使 RNN 记住过去已经处理过的信息,同时使其成为非线性序列处理任务的表达模型。但是,大规模的计算复杂度严重限制了 RNN 的发展。随着 GPU 的出现,大规模的 RNN 随之迅速发展。考虑到 CPU 与 GPU 之间的数据传输很耗时,Li 等人在提出的在 GPU 上实现 RNN 的方法中,将参数(权重 $\boldsymbol{W}$,偏差 $\boldsymbol{B}$)和所有隐藏层及其输出层的状态存储在 GPU 的全局内存中便于调用。但他们将训练数据存储在主内存中而不是 GPU 内存中,因为在 RNN 的许多应用中,每一时步只有一个输入节点被激活,同时只有一个输出节点被检测到。但是对于大规模的训练数据,例如图像等,则应当存储在 GPU 内存中以进行高效地处理。为了充分利用 GPU 的大规模计算能力,Li 等人提出将两个权重矩阵相结合并同时将两个偏差向量相结合的方法,这样有助于 RNN 在不同管道阶段(pipeline stages)的并行运算的实现[26]。

3　处理大数据的深度学习新方法、现存挑战及发展前景

近些年,人类社会产生的数据量急剧增加,大数据技术成为科技界、企业界和各国政府关注的热点。如何分析、利用大数据并从中得出人类所需要的信息成为摆在人们面前的难题。上一节中本文提及了一些处理大数据的大规模深度学习框架,本节将介绍大数据的基本特征和对应的深度学习方法以及大数据深度学习所面对的挑战,并简要说明“深度学习 + 大数据”的发展前景。

3.1　大数据的三个基本特征

大数据的基本特征可以总结为三个 V:Volume(容量)、Variety(多样性)和 Velocity(高速化),更明确的说法就是数据的大规模、数据的多种形式以及数据流的高速传输。下面从这三个方面介绍相应的深度学习方法以及所面临的挑战。

3.2　大规模数据下的深度学习

随着数据规模的不断扩大,深度学习模型的输入、输出可能呈现指数级的增长,数据的维度也会不断增多,这些直接导致了深度学习模型运算时间的增加和模型的复杂化。由于海量的数据不可能放入

仅有单个 CPU 的深度学习模型里进行运算,因此需应用分布式框架的并行运算来对数据进行处理。例如,马焕芳等人采用了以分布式系统架构 Hadoop 为基础,结合 MapReduce 框架,通过批量更新和数据并行的方式来训练卷积神经网络(CNN)的 MR-TCNN 算法。此算法将整个训练数据集分为一个个小块,分布式存储在 Hadoop 平台的每个结点上,每个结点都存储一个完整且相同的 CNN,各个结点使用其上的数据对网络进行训练,通过 map 过程的接收输入数据,进行正向传播、反向传播计算,得出各个权值和偏置的局部改变量,接着经过 combine 和 reduce 处理过程汇总每个结点的权值和偏置的局部改变量得到全局改变量,随后将权值和全局梯度改变量以键值对的形式输出,最后 user 函数利用 reduce 的结果对网络执行批量更新。多次 MapReduce 任务后,在神经网络权值变化很小且在规定误差范围内或达到最大迭代次数时结束网络训练。实验结果表明,MR-TCNN 算法具有较好的并行性,解决了单机训练 CNN 算法耗时长、内存不足的问题[27];也可以通过改进深度学习的架构来达到更好的扩展性和稳定性,以适应大数据的背景,例如:Tang 等人将 CNN 和深度置信网络(DBN)结合用于分析大规模图表。文中采用由 5 个卷积层和 3 个全链接层组成的 CNN,使用 ISVCR-2010 数据集对其进行预训练,然后用图表对模型进行微调,随后对 CNN 的全链接层进行抽取作为深度隐藏特征,送入 DBN 内进行进一步处理;DBN 的输入层、输出层和隐藏层的维度分别为 5000、500 和 2000。深度隐藏特征送入 DBN 后,通过各个受限玻尔兹曼机(RBM)的逐一初始化及随后的微调过程,完成数据的处理。作者在自建的数据集上进行的实验表明,CNN + DBN 的结构相比于单独的 CNN 准确率提高了 2.8%[28];此外,还可以在当前的深度学习框架上扩展应用更多的 GPU,并在其单个 GPU 的计算能力和计算内存上进行改进,提高处理大数据的能力。但大规模数据下的深度学习方法还面临诸多的挑战:大数据来源的多样性有时会导致数据的不完整性,大规模数据中的标签量不足和噪声引起的标签失准等。

3.3 多样性数据下的深度学习

数据的多样性对数据处理模型提出了苛刻的要求。而深度学习可以很有效地对多种类的数据进行处理。例如:Zhao 等人提出了一种基于深度学习的针对多模式数据分类方法,提出的方法应用堆栈受限玻尔兹曼机评估多模式数据集的真实生成函数,避免了进行代价昂贵的参数估计及使用表示能力受限的参数化模型,为了使提出的模型在训练阶段或测试阶段都具有高效性,在区分不同来源的数据样本之前,一个基于聚类的预处理层被应用在堆栈 RBM 网络,在两个基准数据集上的估算证实了提出的方法与当前最成功的方法相比更具效力的结果[29]。

最近,Zhang 等人提出了一种利用深度学习来学习视频中的视觉和文本特征以用于标签定位的方法。方法如下:(1)提出一个基于使用视觉特征的深度学习的语义学标签定位方法,使用 CNN 从测试视频中获取框架 f 的视觉相关性得分 $C(f)$,过程如图 8 所示。在训练时,使用数据集中的标签在训练视频中提取相关的帧,随后相关帧被送入 CNN 中进行模型训练。在测试阶段,利用训练好的 CNN 模型预测从测试视频中提取出的帧,同时获得每个概念的帧的分类概率,一个帧的概率与其对应的标签被用来作为相关分 $C(f)$,并且当相关分超过阈值时才会被标记为正。(2)使用自动语音识别(automatic speech recognition,ASR)获取文本特征同时使用深度学习模型 Word2Vec 定位标签。其中,语义相似性被用来作为框架 f 的文本相关得分 $A(f)$,具体过程如图 9 所示。在训练过程中,作者对移除无用词后的语料库进行词干提取,接着用它来训练 Word2Vec 模型。在测试阶段,进行帧的提取并将其映射至与其关联的 ASR 文本,随后 Word2Vec 模型被用来计算标签和 ASR 文本之间的语义相似性,也就是对应帧的相关得分 $A(f)$,相关分超过阈值时才会被标记为正。(3)使用多模态融合模型结合 $C(f)$ 和 $A(f)$。融合得分 $M(f)$ 由下式决定:

$$M(f) = C(f) + \omega \times A(f) \tag{9}$$

ω 是权重因子(文节中设定的是0.2)。当融合得分 $M(f)$ 超过一定的阈值时,提取的帧才会被标记为与标签相关。当 $C(f)$ 和 $A(f)$ 同时较高的时候,$M(f)$ 才会相对较高,对应框架才会被标注为活跃。实验

使用 DUT-WEBV 数据集,采用多模态融合模型分别同视觉模型和 ASR 文本模型进行比较,前者较后者精确度分别提高了1.28% 和64.3%,取得了最好的效果[30]。

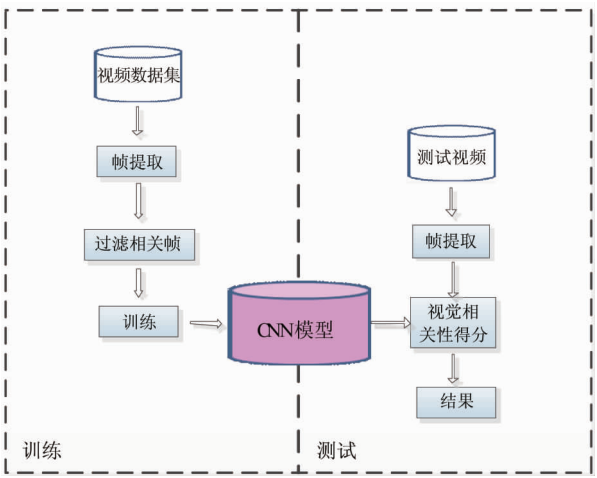

图 8　基于 CNN 的视觉深度学习

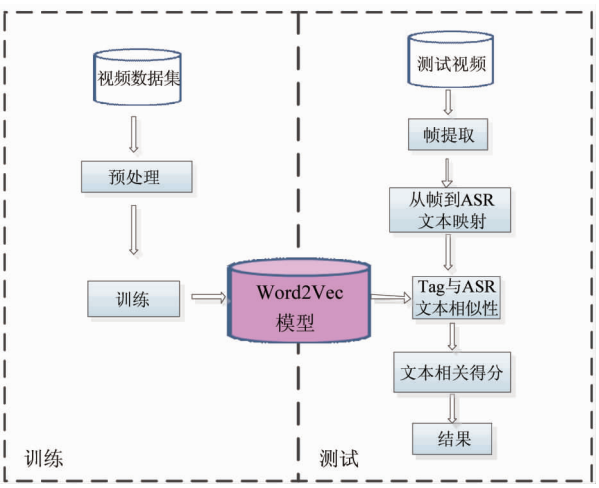

图 9　文本深度学习

虽然人们提出了一些解决方法,但是多种数据类型下的深度学习仍有许多问题亟待解决。比如当前处理多种数据的深度学习方法大部分是处理 2 种不同类型,多于 2 种数据类型的深度学习处理模型还有待研究;所要处理的多种类型数据若相互冲突时,研究者应该怎样最有效地对其进行融合,以达到整体最优等[31]。

3.4　高速率数据下的深度学习

数据的高速处理同样是大数据学习中的一个挑战:数据生成速率非常快,并且要在规定的时间内对数据进行处理,不允许较长的延迟,否则后续高速数据的加入可能会增加数据的"不稳定性"。在高数据速率下,一种从数据中进行学习的方法就是在线学习。在线学习每次学习一个实例并且能迅速得到该实例的正确标签,随后标签会被用于精炼模型。为了对在线学习进行加速,Scherer 等人提出了迷你分批(mini-batch)的方法,就是将范例分成一个个独立的批次,每个批次尽可能的相互独立,这种方法有效地解决了运行时间和运算内存之间的平衡问题[32]。近期 Zhang 等人提出了结合流处理,分批处理和深度学习的深度智能视频处理框架,这个框架能够揭示视频数据中隐藏的知识。基于 Lambda 结构和作者之前的工作[33],在设计框架时他们采用了质量驱动[34]的方法,提出 5 个关键质量属性,并采用了服务导向结构(service-oriented architecture, SOA)、Public-subscribe、MapReduce、Shared data 和 Layered architecture 等组成的软件体系结构。视频处理框架中最底部的层是数据检索层,由媒体服务器和网络摄像头应用程序界面接收来自多个摄像机的视频数据;第二层是数据处理层,它由基于Apache hadoop 的离线处理软件包和基于 Apache storm 流处理的在线处理软件包组成,离线处理软件包包括深度卷积神经网络(deep convolutional neural network, DCNN)训练和 DBN 训练的相关组件。在离线神经网络训练结果的基础上,在线处理软件包可以实现对事件、目标和行为的实时识别;最顶层的域服务层结合了离线和在线处理的结果,开发者可利用这一层创造不同领域的应用。例如智能交通和智能校园等[35];Gu 等人近期提出了在线半监督深度极限学习机(online semi-supervised deep extreme learning machine, OSDELM)方法来利用 WIFI 实现高动态的室内定位系统。

OSDELM 框架如图 10 所示,深度特征学习的每一层都是单个隐藏层前向反馈神经网络,第 k 层的网络参数为 $\beta^{(0,k)} = (H_{(0,k)}^T(C_1 + 1)l_1H_{(0,k)} + C_2l_2)^{-1}H_{(0,k)}^TX^{(0,k-1)}$ [36],用于生成对应层的输出(同时也是上一层的输入)。随后,最后一层输出的判决特征与学习参数 $\beta^{(0,k)} = [C \cdot I + H^T(J + \lambda L)H]^{-1}H^TJT^0$ 一同被送入分类器中[36]。当在新环

境中收集到第 l 层的在线数据后,就可以使用严格的数学推导获取用于第层的更新的参数 $\beta^{(l,k)}$,由下式给出:

$$\beta^{(l,k)} = \beta^{(l-1,k)} + (K_l^k)^{-1} H_{(l,k)}^T (X^{(l,k-1)} - (1 + C_1) H_{(l,k)} \beta^{(l-1,k)}) \quad (10)$$

式(10)等号右边的变量不是与现有模型有关,就是由新的在线数据产生。因此,深度特征学习部分可通过这种在线方式学习判别特征。拉普拉斯矩阵被用来学习分类器,当模型生成时,旧的数据将会被舍弃,仅使用新的在线数据和现有的模型,不能获得准确的拉普拉斯矩阵,所以本文使用近似的方法去更新分类器的参数,经过 I 次在线学习,参数 $\beta^{(l,c)}$ 变为

$$\beta^{(l,c)} = \beta^{(l-1,c)} + (K_l^c)^{-1} H_l^{c^T} (J_u \times T^l - (J_u - \lambda \boldsymbol{L}_u) \times H_l^c \beta^{(l-1,c)}) \quad (11)$$

其中,$\boldsymbol{L}_u$ 是第 l 层在线数据生成的拉普拉斯矩阵,T_l 是相关在线数据的标签。根据这个更新计划,可以利用现有模型和在线数据获得新模型的分类器[37]。

全球各地的研究人员不断地提出新的深度学习方法对高速率的数据进行处理,但高速率数据的不稳定性与对数据的实时性处理等挑战仍需要去应对。

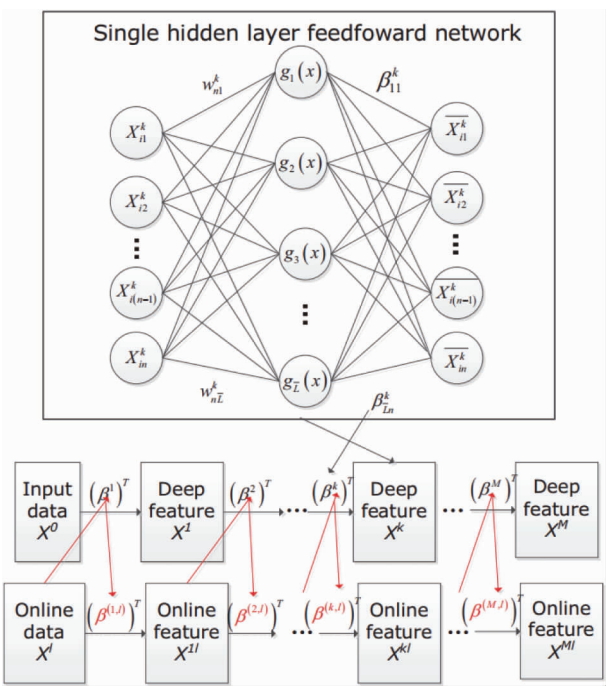

图 10 OSDELM 框架

3.5 大数据背景下深度学习的发展前景

移动互联网和多种智能设备的普及使人类社会进入大数据时代。大数据和深度学习这两者之间的微妙关系正如人工智能专家吴恩达所说的:人工智能是火箭,深度学习是火箭的发动机,而大数据则是火箭的燃料,这两部分必须同时做好,才能够顺利地发射到太空中。对于当前的深度学习神经网络来说,处理大数据的能力是优先的,当超过某个临界点时,输入更多信息并不会带来更好的表现,可能还会适得其反。因此需要对深度学习的网络系统进行开发和不断地调整以适应大数据的要求;反过来说,数据量的适当增长有时会反过来提升深度学习的表现。但还是能够在某些方面预见到深度学习理论的发展趋势,那就是:在大数据背景下,深度学习的架构将会迅速变得更大、更复杂,这些架构也会成为未来创新架构的组成部分。

4 结 论

本文介绍了深度学习的三种模型,同时介绍了每个模型在大数据环境下的应用、挑战以及大数据背景下深度学习的发展前景。在万物互连、数据暴增的时代,面对海量的数据,要考虑的首要问题就是如何对其进行有效的分析和处理并挖掘出数据的价值。深度学习方法在处理大数据的过程中扮演了关键性的角色,它能从数据中自适应地提取其内部表示,尽可能地减少人工的参与,并且用于提取特征的深度模型可以应用到多种场景下,具有更强的泛化性能。此外,深度学习模型可以根据数据量的不断增加适当地扩展其规模,从而更有效地对数据进行处理。目前,在大规模有标签数据集的支撑下,基于有监督特征的深度学习取得了很好的效果。然而世界上大多数数据都是未标记的,大数据时代给人们带来的将会是越来越多的无标签数据,因此许多研究人员认为基于无监督特征的深度学习在未来会是一个十分热门的研究领域。在大数据的环境下,随着深度学习研究的不断深入,可以预见,在不远的将来"大数据 + 深度学习"的技术融合将会在计算机视觉、自然语言理解、机器智能甚至更广泛领域获得

突破性的进展。

参考文献

[1] 程学旗,靳小龙,王元卓等. 大数据系统和分析技术综述. 软件学报,2014, 25(9):1889-1908

[2] Wu M, Chen L. Image recognition based on deep learning. In: Chinese Automation Congress, Wuhan, China, 2015. 542-546

[3] Zhang Y, Shang C. Combining Newton interpolation and deep learning for image classification. *Electronics Letters*, 2015, 51(1):40-42

[4] Sánchez-Gutiérrez M E, Albornoz E M, Martinez-Licona F, et al. Deep learning for emotional speech recognition. In: Proceedings of the 6th Mexican Conference on Pattern Recognition, Cancun, Mexico, 2014. 311-320

[5] Zhao Y, Xu Y M, Sun M J, et al. Cross-language transfer speech recognition using deep learning. In: Proceedings of the 11th IEEE International Conference of Control & Automation (ICCA), Munich, Germany, 2014. 1422-1426

[6] Wang H, Wang N, Yeung D Y. Collaborative deep learning for recommender systems. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 2015. 1235-1244

[7] Yann L C, Yoshua B, Geoffrey H. Deep learning. *Nature*, 2015, 521(7553):436-44

[8] 刘建伟,刘媛,罗雄麟. 深度学习研究进展. 计算机应用研究,2014,31(7):1921-1930

[9] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11):2278-2324

[10] 翟继友. 基于深度置信网络的语义相关度计算模型. 科学技术与工程,2014,14(32):58-62

[11] Fischer A, Igel C. An introduction to restricted boltzmann machines. *Lecture Notes in Computer Science*, 2012, 7441:14-36

[12] 陈宇. 基于深度置信网络的中文信息抽取方法:[博士学位论文]. 哈尔滨:哈尔滨工业大学计算机科学与技术学院,2014. 12-13

[13] Graves A. Supervised Sequence Labelling with Recurrent Neural Networks. *Studies in Computational Intelligence*, 2012, 385:5-13

[14] 一只鸟的天空. 循环神经网络(RNN,Recurrent Neural Networks)介绍. http://blog. csdn. net/heyongluoyao8/article/details/48636251:北京创新乐知信息技术有限公司,2015

[15] Sainath T N, Kingsbury B, Saon G, et al. Deep convolutional neural networks for large-scale speech tasks. *Neural Network*, 2015, 64:39-48

[16] Lv Y, Duan Y, Kang W, et al. Traffic flow prediction with big data: a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(2): 865-873

[17] 雷亚国,贾峰,周昕等. 基于深度学习理论的机械装备大数据健康监测方法. 机械工程学报,2015,51(21): 49-56

[18] Wang J, Cheng L. DistDL: A distributed deep learning service schema with GPU accelerating. In: Proceedings of the Web Technologies and Applications, Guangzhou, China, 2015. 793-804

[19] Simard P Y, Steinkraus D, Platt J C. Best practice for convolutional neural networks applied to visual document analysis. In: Proceedings of the International Conference on Document Analysis & Recognition, Edinburgh, UK, 2003. 958-962

[20] Scherer D, Schulz H, Behnke S. Accelerating Large-Scale Convolutional Neural Networks with Parallel Graphics Multiprocessors. *Springer*, 2010, 6354:82-91

[21] Satish N, Sundaram N, Keutzer K. Optimizing the use of GPU memory in applications with large data sets. In: Proceedings of the 16th International Conference on High Performance Computing, Kochi, India, 2009. 408-418

[22] Hinton G E. A practical guide to training restricted boltzmann machines. *Momentum*, 2010, 9(1):599-619

[23] Hinton G E. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 2002, 14 (8):1771-1800

[24] Lopes N, Ribeiro B. An evaluation of multiple feed-forward networks on GPUs. *International Journal of Neural Systems*, 2011, 21 (1):31-47

[25] Lopes N, Ribeiro B. Towards adaptive learning with improved convergence of deep belief networks on graphics processing units. *Pattern Recognition*, 2014, 47 (1): 114-127

[26] Li B, Zhou E, Huang B, et al. Large scale recurrent

neural network on GPU. In: Proceedings of the 2014 International Joint Conference on Neural Networks, Beijing, China, 2014. 4062-4069

[27] 马焕芳,赵歆波,邹晓春. 基于 MapReduce 的卷积神经网络算法研究. 中国体视学与图像分析,2015,20(4):339-346

[28] Tang B, Liu X, Lei J, et al. Deep chart: combining deep convolutional networks and deep belief networks in chart classification. *Signal Processing*, 2015, 124:156-161

[29] Zhao H, Li G, Niu W, et al. A deep learning method for multimodal data. *Journal of Computational Information Systems*, 2015, 11(12):4237-4244

[30] Zhang R, Tang S, Liu W, et al. Multimodal tag localization based on deep learning. In: Proceedings of the International Conference on Internet Multimedia Computing and Service, Zhangjiajie, China, 2015. 1-4

[31] Chen X W, Lin X. Big data deep learning: challenges and Perspectives. *Access IEEE*, 2014, 2:514-525

[32] Scherer D, Müller A, Behnke S. Evaluation of pooling operations in convolutional architectures for object recognition. In: Proceedings of the International Conference on Artificial Neural Networks, Thessaloniki, Greece, 2010. 92-101

[33] Zhang W, Hansen K M, Ingstrup M. A hybrid approach to self-management in a pervasive service middleware. *On the Horizon*, 2014, 67(3):143-161

[34] Bass L, Clements P, Kazman R, et al. Software Architecture in Practice: Addison-Wesley. 2003

[35] Zhang W, Xu L, Li Z, et al. A deep-intelligence framework for online video processing. *IEEE Software*, 2016, 33(2):44-51

[36] Gu Y, Chen Y, Liu J, et al. Semi-supervised deep extreme learning machine for Wi-Fi based localization. *Neurocomputing*, 2015, 166(C):282-293

[37] Gu Y, Chen Y, Liu J, et al. Online deep intelligence for Wi-Fi indoor localization. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and the ACM International Symposium on Wearable Computers, Osaka, Japan, 2015. 29-32

The study of deep learning under big data

Wang Jinjia, Chen Hao, Liu Qingyu

(School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

Abstract

The concepts of big data and deep learning (a subfield of machine learning) were given, and the importance of deep learning in acquiring valuable information from big data was interpreted. The deep learning framework for concurrent computation using graphics processing unit was described, and its big convolutional neural network (CNN), big deep belief network (DBN) and big recurrent neural network (RNN) were emphatically introduced. The features of big data in volume, variety and velocity were analyzed, and the methods for deep learning under large scale data, variable data and high rate data were introduced. The future development of the research on deep learning under big data was forecasted, and the possibility that the technology of fusing big data and deep learning will make an important breakthrough in the fields such as computer vision and machine intelligence was pointed out.

Key words: big data, deep learning, convolutional neural network (CNN), deep belief network(DBN), recurrent neural network (RNN)