

人工智能技术在机器设备剩余使用寿命预测中的应用

李晓黎^{1,2*}

(1. 新加坡资讯通信研究院,新加坡 138632;

2. 南洋理工大学,新加坡 639798)

摘要:机器设备剩余使用寿命预测的研究具有广泛的应用价值。文章介绍了人工智能技术在机器设备剩余使用寿命预测中的最新研究成果。在一些经典方法基础上,讨论了三个新的研究方向:一是如何从传感器时序数据中自动学习好的特征表示以及融合自动学习的特征和人工提取的特征以提升预测的准确率;二是如何通过知识蒸馏的方法从大的深度学习网络LSTM中学习小而准的CNN网络模型,以便将其应用到计算资源有限的设备端;三是如何利用其他源领域的训练数据来建立跨领域的预测模型,使得它在当前没有训练数据的目标领域取得良好的预测效果。

关键词:剩余使用寿命估计;长短期记忆网络;卷积神经网络;深度学习;生成对抗网络;领域自适应

中图分类号:TP391 **文献标志码:**A **文章编号:**0253-2395(2022)03-0626-09

AI Techniques for Applications of Equipment Remaining Useful Life Prediction

LI Xiaoli^{1,2*}

(1. Institute for Infocomm Research, A*STAR, Singapore 138632;

2. Nanyang Technological University, Singapore 639798)

Abstract: This paper introduces how latest artificial intelligence technologies can be used for prediction of equipment remaining useful life, which has widespread applications. Based on some standard approaches, we introduce three new research directions. First, in order to improve the prediction accuracy, how to build a model that is able to better learn the feature representation from time series sensor signals, as well as how to integrate features extracted by deep learning and manually curated features based on domain knowledge. Second, we elaborate how to leverage knowledge distillation to transfer the knowledge from big LSTM network to smaller yet accurate CNN network, to facilitate its usefulness at computing resource limited edge device. Finally, we study how to build a cross-domain model to leverage the training data from source domain so that it can perform well in the unlabelled target domain.

Key words: equipment remaining useful life prediction; long short-tem memory; convolutional neural network; deep learning; generative adversarial network; domain adaption

0 引言

人工智能是由计算机实现的智能,它能够执行通常需要高度人类智能才能完成的任务。

近年来,由于计算机的高速运算能力和大规模存储空间、大数据的积累、最新人工智能算法的提出,人工智能在计算机视觉、语音识别、自然语言处理、生物信息学、推荐系统、战略游

收稿日期:2021-12-21;接受日期:2022-02-25

* 通信作者:李晓黎(1969-),男,山西晋城人,博士,教授,主要研究方向为人工智能、机器学习、大数据方面的研究。E-mail: xlli@i2r.a-star.edu.sg

引文格式:李晓黎. 人工智能技术在机器设备剩余使用寿命预测的应用[J]. 山西大学学报(自然科学版),2022,45(3): 626-634. DOI:10.13451/j.sxu.ns.2022008

戏、智能感知等诸多领域取得了长足进展。在一些智能需求高度复杂的领域中取得的骄人成绩甚至大大超过人类智能,比如来自DeepMind的AlphaGo^[1]已经可以轻松战胜人类围棋的世界冠军。

此外,人工智能在提高人类的生产力和效率方面开始发挥越来越重要的作用,例如帮助医生完成费时费力的需要高度专业知识的医疗图像诊断任务,通过建立个性化推荐引擎帮助电商和新闻网站和应用在合适的时间和地点推荐合适的产品和服务,大规模的智能制造中生产设备预测维护以及根据客户的需求进行个性化的生产,智能交通和自动驾驶汽车提高交通效率,以及促进人类通过机器翻译和聊天机器人等工具进行有效的交流和互动等。

近年来,由于各种传感器技术的成熟和广泛使用,我们可以利用传感器收集机器设备的历史和实时大数据,并利用人工智能技术来建立预测模型,进行故障诊断和预测^[2]。因此,机器设备剩余使用寿命预测等具有广泛的应用领域。

机器设备剩余使用寿命(Remaining Useful Life,简称RUL)预测的研究非常重要^[3-5],它不但对于提高机器设备和工业系统的可靠性起关键作用,而且由于很多机器设备非常昂贵,这项研究还可以最大程度地提高它们的使用寿命、使用效率、降低它们的维护成本等。同时,从技术的角度上讲,如何从大量的传感器数据中学习好的特征表示?如何提高预测模型的准确率?如何建立可以应用于设备端的小而准确的模型?这些仍然是具有挑战性的课题。所以,这项研究不仅具有重要的理论价值,而且具有实际的应用价值。它的应用领域横跨生产企业、航天航空、电子、半导体、医疗卫生、工业制造、石油天然气勘探等行业。

1 一些经典的相关工作

为了预测机器设备的剩余使用寿命,近年来已有许多方法被提出。主要的方法可以分为两大类,即基于物理模型的方法和基于数据驱动的方法。基于物理模型的方法需要物理理论来理解并模拟机器设备的行为及随着机器寿命的衰老其性能下降的详细过程^[6-7],但由于工业系统种类繁多且变得越来越复杂,显式的物理建模变得极其困难。与之相比,数据驱动的方法旨在直接从机器设备产生的数据中学习,而不依赖于未知系统的物理模型^[8-9]。此外,由于传感器的大量使用并由此产生大量的数据,而人工智能技术特别是深度学习技术可以从大量数据中进行学习和预测,所以,基于数据驱动的方法有望成为对于复杂的工业系统非常有前途的剩余使用寿命预测的关键技术。

在数据驱动方法中,传统的人工智能通常通过标准的机器学习算法来进行机器设备的剩余使用寿命预测^[8,10],其主要过程分为两步:一是执行特征工程,这通常是领域知识和专家经验来手动的从传感器数据中提取代表性的特征。二是利用已有的机器学习算法对用第一个步骤来表示的训练数据进行回归分析,也就是用模型来建立传感器数据的特征表示和已知机器设备的剩余使用寿命之间的关系,从而达到预测的目的。

图1为原始传感器获得的时序数据信号随时间变化的一个示例,其中 x 轴是时间轴, y 轴表示传感器在相应时间点的数值。针对不同的应用,传感器可以采集来自于震动、温度、声音、电流等信号^[11-12]。为了执行特征工程,通常可以对传感器数据进行时域分析和频域分析^[13],其中,时域分析是衡量信号特征的重要指标,可以分析信号的各种特性和性能。比如给

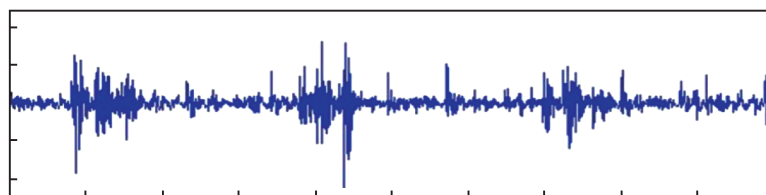


图1 传感器时序数据

Fig. 1 Time-series sensor signal changes over time

定 N 个时间点,可以提取时域的各种特征,包括平均值、峰值、均方值、均方根、峰度、和标准差等。它们的定义如下:

平均值 (mean, \bar{x}) 表示 N 个信号值的平均,可以表示为: $\bar{x} = \frac{1}{N} \sum_i x_i$, 其中 x_i 是在第 i 时间点的信号值。

峰值 (Peak value, PV) 表示信号值的最大落差 (也即最大值和最小值的不同):

$$PV = \max(x_i) - \min(x_i)。$$

均方值 (Mean square, MS) 是信号平方的平均,代表了信号的能量: $MS = \frac{1}{N} \sum_i x_i^2$ 。

均方根 (Root-Mean-Squared, RMS) 是均方值的算数平方根: $RMS = \sqrt{\frac{1}{N} \sum_i x_i^2}$ 。

峰度 (Kurtosis, K) 表示信号概率密度函数的陡峭程度: $K = \frac{1/N \sum_i (x_i - \bar{x})^4}{(RMS)^4}$ 。

标准差或方差 (Standard deviation) 是每个信号值与全体样本平均值之差平方值平均数的算数平方根,表示数据的离散程度: $SD = \sqrt{\frac{1}{N} \sum_i (x_i - \bar{x})^2}$ 。

另一方面,可以使用频域分析方法,它显示有多少信号存在于与频率范围相关的给定频带内,其主要过程如图 2 所示。对于传感器原始信号一般要先进行归一化:

$$NormalData = 2 \frac{x - \min(x)}{\max(x) - \min(x)} - 1。$$

然后将传感器信号切割成段(频带),在每一段上计算的频带的强度频谱,并将其用作特征

集。具体来说,可以使用转换函数把信号从时域转为频域,转换函数常使用傅立叶变换,它将时间转换为各种频率的正弦波的积分,每个正弦波代表一个频率分量,如利用快速傅立叶变换 (FFT) 将归一化的数据从时域变换到频域。最后,计算各个频带上的强度频谱,将输入信号在各个频带上变换成离散的带强度的特征。

一般而言,从传感器信号中可以抽取时域特征和频域特征,并可以将它们连接在一起成为联合特征表示。时域和频域的不同是:当分析使用信号的时间单位,例如秒或其倍数之一(分钟或小时)作为度量单位时,则它处于时域中;当分析涉及像赫兹这样的频率单位时,它就在频域中。

执行完第一个步骤特征工程,在第二个步骤将使用机器学习算法(例如支持向量机^[14]、决策树、随机森林、神经网络等),根据训练集对第一步提取的特征来执行回归分析,以便进行机器设备的剩余使用寿命的预测。然而针对不同的应用领域,第一个步骤需要基于领域知识和专家经验选择并提取不同的特征,也即特征工程不一定具备通用性。此外,这个第一步的特征提取和第二步的基于机器学习的预测通常是割裂的或独立的,不能进行联合优化,这也影响了传统人工智能算法的性能表现。

最近,深度学习在许多具有挑战性的领域,如图像分析、语音识别、自然语言处理,推荐系统、生物信息学、网络分析、剩余使用寿命预测^[4, 15-22]等领域取得长足进展。深度学习的最大优点是能够自动从数据中学习具有代表性的特征而无需人工干预,并可以同时执行和联合优化

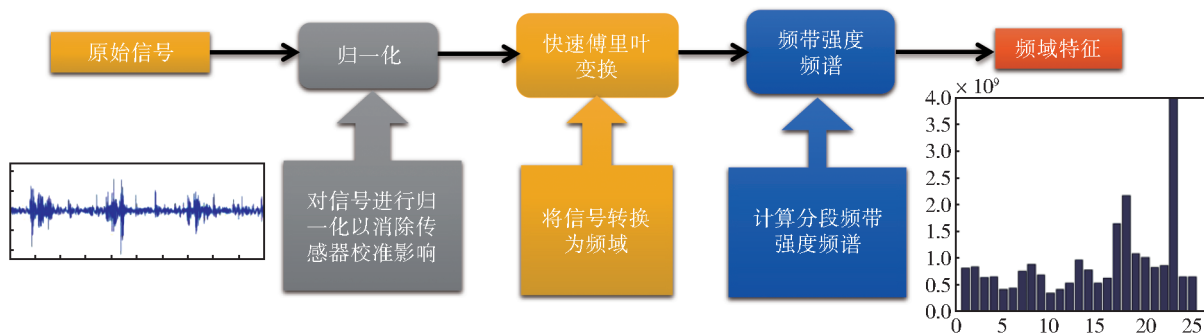


图2 频域分析方法示意图

Fig. 2 The diagram of frequency domain analysis

特征学习和机器学习两个步骤,以获得良好性能。

深度学习也可用于机器设备剩余使用寿命预测。根据机器设备的具体特点,许多如震动、温度、声音、电流传感器等,可以连接到相关机器设备上采集数据。因此可以联合优化从传感器原始数据中自动学习好的特征表示和根据训练数据中真实的机器设备剩余使用寿命执行回归分析这两个步骤。最受欢迎的深度学习算法之一是卷积神经网络(CNN),它在图像处理方面的性能得到了显著提高,在传感器数据上也取得很好的效果^[23]。由于CNN网络的独特结构,对于特征学习非常有效,并且可以并行训练,它也被用于机器设备剩余使用寿命预测并表现出比传统的机器学习算法出色的结果^[4,24]。

为了进一步提高机器设备剩余使用寿命预测的准确性,可以考虑传感器时间序列数据的特点,也即不同时间点之间的数据具有时间依赖性。同时,循环神经网络(RNN)节点沿序列连接,旨在建模时间序列中的时间依赖性。因此一个RNN网络模型自然适用于机器设备剩余使用寿命预测。然而传统的RNN经常会遇到梯度消失或在网络训练期间梯度爆炸的问题,这大大降低了其在建模时间依赖性方面的性能。为了解决这个问题,Hochreiter和Schmidhuber提出了一个新的架构,命名为长短期记忆网络(LSTM)^[25],它由单元,输入门,输出门和忘记门组成。该单元记住任意时间间隔内的值,并且三个门控制进出单元的信息流^[26]。门可以允许或阻止信息沿序列传递,可以捕获长期时间依赖。由于LSMT网络专为分析具有时间序列的数据而设计,在分析时间序列数据上取得了巨大的成功,例如占用率估计、视频分析和自然语言处理^[22,27]等。最近,它在RUL预测上也取得了不错的成绩,特别是最近的研究表明LSTM在用于机器设备剩余使用寿命预测的问题上是优于CNN的^[3]。但是,由于独特设计的结构,LSTM的计算复杂度远远高于CNN。

2 通过改进LSTM方法和特征融合框架以

实现更准确的预测

虽然LSTM可以利用深度学习从时间序列中自动学习特征表示,并已经在机器设备剩余使用寿命的预测任务上取得了良好的表现^[28-29],但是标准的LSTM网络仅使用在整个时间序列的最后时间点学习的特征进行回归分析或分类,这样做比较浪费在整个时间序列上的学习成果,这是因为如果能学习到其他时间点的特征应该会对最终的预测有一定的贡献。此外,还有一些基于人类的领域知识获取的特征可能提供对这个预测任务额外有用的信息。因此如何将这些基于人的智能构建的特征和基于深度学习LSTM网络自动学习特征进行结合是一个重要的问题。

研究人员提出了一种新的基于注意力机制的深度学习方法进行机器设备剩余使用寿命的预测^[3],主要思想是当LSTM网络从传感器信号中自动学习序列数据的特征时,用注意力机制在考虑各个时间点重要性的情况下综合学习特征。它会同时学习两个重要性:特征的重要性(哪一维更重要)和时间点的重要性,以将更大的权重分配给更重要的时间点及特征。也就是说,这个特征表示会是所有时间点所学重要特征的综合,而重要的时间点特征表示上会发挥更大的作用。此外,除了以上用深度学习到的特征,还需要一个特征融合框架来结合自动学习的特征和手工的特征,以提高系统的总体预测性能。图3显示了特征融合框架来结合自动学习的特征和手工的特征。

手工特征:从基于传感器原始信号中,可以提取一些基于专家知识的手工定制特征,例如线性回归的均值和趋势系数。其中,均值表示传感器数据的大小,趋势系数表示传感器数据的性能退化。这两个简单的手工定制的特征已被证明对机器设备剩余使用寿命的预测有用^[8]。为了充分利用所有可用信息,这里需要一个特征融合框架来结合深度学习的特征和手工特征以提高系统预测的性能。

特征融合框架:在图3中,根据原始传感器信号的输入,LSTM网络对不同的时间点(1, 2, ..., s)用于自动学习特征表示。学习到各个时间点的序列特征被视为注意力机制模型(黄

色)的输入,其输出(注意力的权重)表示了多维特征重要性和时间点的重要性。然后,将学习到的各个时间点序列特征根据注意力模型生成的权重进行加权融合。之后,通过两个全连接层(蓝色)以获得更多抽象特征,同时,从原始传感器信号中获得的手工制作的特征,输入全连接层以获得更多抽象特征。为了充分利用这两种特征,可以将它们合并起来形成一个完整的特征表示。最后,我们可以利用回归分析进行机器设备剩余使用寿命的预测。

由于这个预测问题是一个典型的回归分析问题,所提出方法的损失函数被设定为均方误差(MSE)损失。给定预测的剩余使用寿命和真实的剩余使用寿命,MSE损失可用训练数据进行计算并反向传播到为每一层(例如LSTM层、注意力层和全连接层)。然后,Adam优化方法被用来优化模型在各层的参数。注意到图3这个框架是一个端到端的可训练架构,这意味着所有模型参数都可以联合训练。同时考虑深度学习模型中的过拟合问题,可以使用正则化技术。Dropout是解决这个问题的最流行的技术,在训练阶段,Dropout会随机地部分屏蔽隐藏神经元的输出,以便这些神经元不会影响前向传播模型的训练。但在测试阶段,Dropout会被关闭,所有隐藏神经元将会输出结果。

用图3的框架训练的模型在两个真实飞机发动机性能退化的C-MAPSS数据集上进行了实验,实验结果表明,这个新的模型优于现有的技术^[31]。

3 通过知识蒸馏来建立小而准的模型

在许多实际场景中,机器设备剩余使用寿命的预测模型需要部署在边缘设备上,以支持实时决策,降低数据通信成本,并保留数据的隐私。然而由于这些机器设备具有有限的计算资源和内存,因此,工业界普遍更喜欢可以实现准确预测和快速推理的小模型。但当前的深度学习算法太复杂,导致超大的模型,并且其推理过程也需要大量计算,因此,这些模型不可以应用到计算资源有限的设备端。

为了解决这些问题,模型压缩技术已经提出,通过压缩深度神经网络来应用到边缘设备

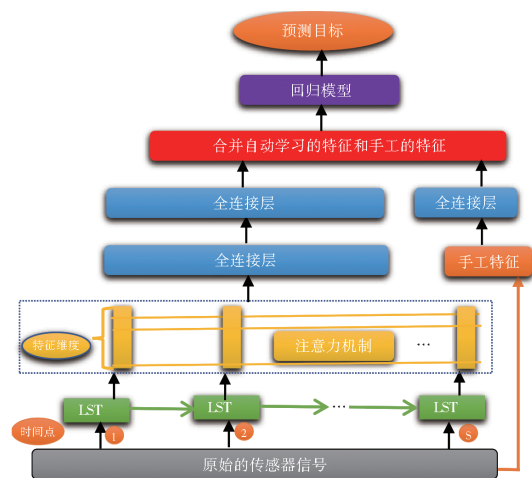


图3 特征融合框架来结合自动学习的特征和手工的特征
Fig. 3 Feature fusion framework to integrate automatically learned features and manually designed features

中。例如,参数量化方法^[30-31]使用更少的比特来压缩原始网络来表示权重,可以实现显著的加速但也会导致精度损失^[32]。另一种常见的用于模型压缩方法是权重剪枝^[33],其目的是去除神经网络中不必要的参数。虽然权重剪枝能够减少模型存储大小,但它不能提高训练效率或推理时间。其他方法,如矩阵分解^[34-35]虽然也可以减少模型大小,但它们只解决了存储复杂深度模型的问题,并且有类似权重剪枝方法的缺点。相对来说,知识蒸馏方法不仅在减少模型存储方面显示出希望,而且可以提高模型的推理效率^[36-37],所以蒸馏后的模型特别适合应用到设备端。

前面提到LSTM对传感器时序数据有较为强大的建模能力,用它来预测机器设备剩余使用寿命达到较好的效果。然而,由于基于LSTM方法的模型很大、计算复杂性高,所以它不能部署到那些具有有限计算能力和存储空间边缘的设备上。因此,如何提出一个新的方法,从大型的模型中学习小而相对准确的模型变的异常重要,它是我们是否可以将模型应用到边缘设备的关键问题。

为了解决这个问题,研究人员提出了新的知识蒸馏框架,用于把知识从复杂的基于LSTM的预测模型压缩到小规模CNN模型以便可以应用到设备端。它包括两个部分,特征蒸馏和知识蒸馏,如图4所示。

图4左边是特征蒸馏部分,借鉴了生成对

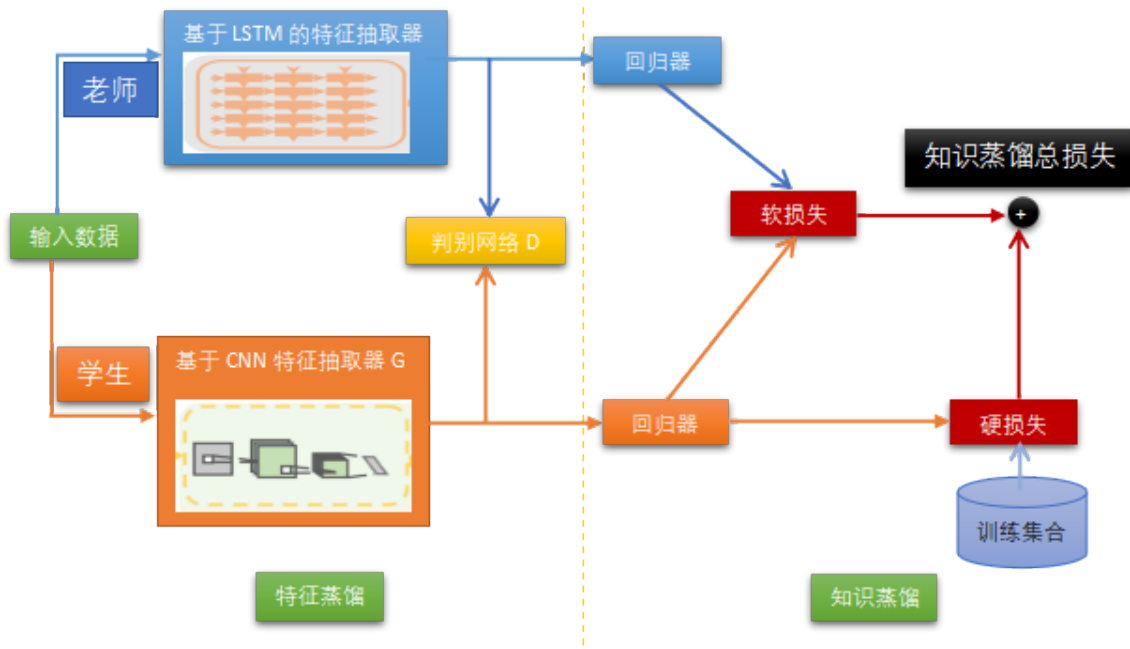


图4 知识蒸馏框架

Fig. 4 A knowledge distillation frame

抗网络 (Generative Adversarial Network, 简称 GAN) 的思想, 通过让两个神经网络相互博弈的方式进行学习。在这个设计的特征蒸馏方法中, 两个神经网络分别是生成网络 (G) 和判别网络 (D)。在图 4 中, 基于 CNN 的特征抽取器被认为是生成网络 G, 它表示学生用来从输入数据中自动提取特征的模块。与此同时, 基于 LSTM 的特征抽取器是老师用来从输入数据中自动提取特征的模块。最后, 判别网络 D 作为一个两类分类网络, 旨在识别抽取的特征是来自教师 (图 4 上面部分) 还是学生 (图 4 下面部分)。在这里, 判别网络 D 和生成网络 G 在玩一个两人极极大的博弈游戏, 其中 D 旨在最大化正确分类抽取的特征分别来自老师和学生的概率, 同时 G 旨在最小化 D 能够预测特征来自学生的概率。交替重复上面的生成网络和判别网络的训练过程, 即迭代地最小化和最大化。最终, 可以最大化基于 CNN 和基于 LSTM 的特征抽取器的相似度, 学生能够生成和老师相似的特征, 从而改进学生的基于简单 CNN 特征抽取器的性能。也就是说, 老师 (使用复杂网络) 教给学生 (使用简单网络) 如何从数据中获得好的特征表示。

图 4 右边是知识蒸馏部分, 接着处理机器设备剩余使用寿命预测的回归任务。通过利用

老师的预测来获取知识蒸馏, 因为老师的预测是从复杂网络来的, 比较准确, 所以老师的预测结果被称为软标签, 而将软损失定义为学生的预测和教师的预测之间的差异。此外, 还有硬损失, 它被定义为学生的预测和真实结果 (标签) 之间的差异。知识蒸馏的总损失函数定义为软损失和硬损失的加权组合。最后经过最小化知识蒸馏的总损失函数, 可以学到小又比较准确的学生 CNN 网络中的回归器, 并用于设备端剩余使用寿命的预测。在预测阶段, 我们用图 4 的下半部分 (学生网络) 做预测, 也即从设备端获得传感器信号, 然后通过 CNN 的特征抽取器 G 抽取特征 (它可以获得和复杂 LSTM 网络相似的特征表示), 然后用学生的回归器进行预测 (它可以获得与老师回归器相似的预测以及最小化训练集中的错误, 以取得准确的预测结果)。

使用图 4 的模型在简单和复杂数据集来验证所提出模型的有效性。结果表明, 所提出的新模型显著优于现有的方法。与复杂的大型 LSTM 老师网络相比, 最终获得的基于 CNN 的学生模型网络具有以下三个特点: (1) 网络权重减少 12.8 倍; (2) 总浮点运算数减少了 46.2 倍; (3) 它的性能达到了与大型的 LSTM 模型相当的性能^[5]。

4 建立跨领域的预测模型

现有的机器设备剩余使用寿命预测的工作通常有两个假设:1)收集的训练数据和测试数据来自相同的操作条件(相同的数据分布或领域);2)有丰富的训练数据(大量真实的剩余使用寿命的例子)可用于训练预测任务。但是在实际应用中,这两个假设通常是不成立的。主要的原因如下:一是训练数据的收集是昂贵的。对于一些复杂的关键机器,运行至失败或生命周期可能付出高昂代价和导致灾难性后果。此外,机器设备的性能下降过程可能会延续数年,这也会限制训练数据的收集。二是利用现有的训练数据集建立的模型可能只有在特定的工作条件下产生正确的预测。然而,当工作条件发生变化后,之前训练的模型往往因为不同工作条件下数据分布的不同,无法取得良好的效果。所以,在具有有限训练数据和不同工作条件(数据分布)下预测剩余使用寿命是具有挑战性的任务。

一般而言,我们可能会有其他领域(源领域)的训练数据,如何用它们来建立跨领域的预测模型,以便当前没有训练数据的当前领域(目标领域)取得良好的预测效果是这个问题的关键。领域自适应是一种将知识从训练数据丰富的源领域转移到一个不同但相关的缺乏训练数据的目标领域的方法^[38]。但是,大多数现有的领域自适应方法是为与图像分析相关的任务而设计的。

最近研究人员提出了一种无监督的对抗性领域自适应方法^[39-40],它可以高效地从拥有训练数据的源领域将知识转移到一个新的没有训练数据的目标领域。这个方法具有三个主要步骤。第一步是从源领域中进行监督学习,从训练数据中通过源领域编码器来抽取特征和建立预测模型。第二步是训练目标领域的特征编码器以产生与源领域相似的特征。通常目标领域特征编码器先采用源领域编码器的权重进行初始化。然后使用生成对抗网络来减轻源领域和目标领域的不同分布。它基于两个玩家进行跨领域的对抗性游戏(这里目标领域特征编码器是生成器),以找出一个领域不变的特征表示,使得领域鉴别器网络不能区分生成器产生的特

征是来自源领域和还是目标领域。最后一步是使用源领域预测模型(在第一步建立的模型)和第二步建立的目标编码器来对目标领域进行预测。因此,由于源领域和目标领域产生特征的高度相似性,预测器(回归分析)在源领域上训练的网络可以很好地推广到目标领域中,从而在目标领域产生准确的预测效果。在预测航空发动机12个跨领域场景的剩余使用寿命预测任务上,实验结果表明新提出的方法明显优于现有的方法。

5 展望未来和结论

近年来,由于计算机硬件和机器学习算法不断进步,人工智能的研究取得长足进展。特别是随着物联网的高速发展,大量不同类型的传感器可以从机器设备上方便地收集大量的历史和实时数据。另一方面,人工智能可以从大数据中学习,从而对机器设备的健康状况进行自动诊断,以及在对其未来健康发展趋势进行预测中发挥重要的作用。

本文主要集中探讨了人工智能在机器设备剩余使用寿命预测中的应用。这项研究不但具有重要的理论价值,而且具有广泛的实际应用。本文在介绍了一些经典的方法基础上,讨论了三个新的重要研究方向。一是如何利用一个特征融合框架来更好地从传感器时序数据中自动学习好的特征表示,并融合人的智能以进一步提升预测的准确率。二是如何通过知识蒸馏方法从大的深度学习网络中学习,并把知识通过生成对抗网络传递给小而准的网络模型,以便在计算资源有限的设备端使用,并支持实时决策,降低数据通信成本,同时保留数据的隐私。三是如何利用源领域的训练数据来建立跨领域的预测模型,以便在没有训练数据的目标领域取得良好的预测效果。其重要性在于很多情况下我们面临在当前的目标领域没有训练数据的问题,而且源领域虽然有训练数据但和目标领域有数据分布不同的问题,所以要用新的领域自适应的方法来把源领域和目标领域进行更好的匹配和对齐,以此提高预测的精度。

尽管这几个新的研究方向取得了一定的进展,但是为了提高应用效果还需要很多的努

力。很多情况下机器设备剩余使用寿命预测需要我们获得大的训练数据,例如从机器设备全新的状况下的数据一直到其生命周期结束的数据,而在一些真实世界的应用中,我们不一定有这样的数据。同时我们可能会有很多相同或相似机器设备的数据,但不一定会有标签(真实的剩余使用寿命),如何利用好这些无标签的数据是进一步研究的关键。这里只依赖少量训练数据和大量无标注数据新的半监督学习和无监督学习的方法就变得异常重要。

参考文献

- [1] SILVER D, HUANG A, MADDISON C J, *et al.* Mastering the Game of Go with Deep Neural Networks and Tree Search[J]. *Nature*, 2016, **529**(7587): 484–489. DOI: 10.1038/nature16961.
- [2] ZHANG B, LI W, LI X L, *et al.* Intelligent Fault Diagnosis under Varying Working Conditions Based on Domain Adaptive Convolutional Neural Networks[J]. *IEEE Access*, 2018, **6**: 66367–66384. DOI:10.1109/ACCESS.2018.2878491.
- [3] CHEN Z H, WU M, ZHAO R, *et al.* Machine Remaining Useful Life Prediction via an Attention-based Deep Learning Approach[J]. *IEEE Trans Ind Electron*, 2021, **68**(3): 2521–2531. DOI:10.1109/tie.2020.2972443.
- [4] SATEESH BABU G, ZHAO P L, LI X L. Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life[C]//Database Syst Adv Appl, 2016. DOI:10.1007/978-3-319-32025-0_14.
- [5] XU Q, CHEN Z H, WU K Y, *et al.* KDnet-RUL: a Knowledge Distillation Framework to Compress Deep Neural Networks for Machine Remaining Useful Life Prediction[J]. *IEEE Trans Ind Electron*, 2022, **69**(2): 2022–2032. DOI:10.1109/TIE.2021.3057030.
- [6] SINGLETON R K, STRANGAS E G, AVIYENTE S. Extended Kalman Filtering for Remaining-useful-life Estimation of Bearings[J]. *IEEE Trans Ind Electron*, 2015, **62**(3): 1781–1790. DOI:10.1109/TIE.2014.2336616.
- [7] LI N P, LEI Y G, GUO L, *et al.* Remaining Useful Life Prediction Based on a General Expression of Stochastic Process Models[J]. *IEEE Trans Ind Electron*, 2017, **64**(7): 5709–5718. DOI:10.1109/TIE.2017.2677334.
- [8] KHELIF R, CHEBEL-MORELLO B, MALINOWSKI S, *et al.* Direct Remaining Useful Life Estimation Based on Support Vector Regression[J]. *IEEE Trans Ind Electron*, 2017, **64**(3): 2276–2285. DOI:10.1109/tie.2016.2623260.
- [9] SUN C, MA M, ZHAO Z B, *et al.* Deep Transfer Learning Based on Sparse Autoencoder for Remaining Useful Life Prediction of Tool in Manufacturing[J]. *IEEE Trans Ind Inform*, 2019, **15**(4): 2416–2425. DOI:10.1109/TII.2018.2881543.
- [10] TIAN Z G. An Artificial Neural Network Method for Remaining Useful Life Prediction of Equipment Subject to Condition Monitoring[J]. *J Intell Manuf*, 2012, **23**(2): 227–237. DOI:10.1007/s10845-009-0356-9.
- [11] CAO H, LI X L, WOON D Y K, *et al.* Integrated Oversampling for Imbalanced Time Series Classification[J]. *IEEE Trans Knowl Data Eng*, 2013, **25**(12): 2809–2822. DOI:10.1109/tkde.2013.37.
- [12] CAO H, LI X L, WOON Y K, *et al.* SPO: Structure Preserving Oversampling for Imbalanced Time Series Classification[C]//2011 IEEE 11th International Conference on Data Mining. 2011: 1008–1013. DOI: 10.1109/ICDM.2011.137.
- [13] NGUYEN MN, BAO CY, TEW K L, *et al.* Ensemble Based Real-time Adaptive Classification System for Intelligent Sensing Machine Diagnostics[J]. *IEEE Trans Reliab*, 2012, **61**(2): 303–313. DOI:10.1109/TR.2012.2194352.
- [14] LI X L, LIU J M, SHI Z Z. A Chinese Web Page Classifier Based on Support Vector Machine and Unsupervised Clustering[J]. *Chin J Comput*, 2001, **24**(1): 62–68.
- [15] XIA M, LI T, SHU T X, *et al.* A Two-stage Approach for the Remaining Useful Life Prediction of Bearings Using Deep Neural Networks[J]. *IEEE Trans Ind Inform*, 2019, **15**(6): 3703–3711. DOI:10.1109/TII.2018.2868687.
- [16] YANG B Y, LIU R N, ZIO E. Remaining Useful Life Prediction Based on a Double-convolutional Neural Network Architecture[J]. *IEEE Trans Ind Electron*, 2019, **66**(12): 9521–9530. DOI:10.1109/TIE.2019.2924605.
- [17] Tran L V, Pham T A N, Yi T, *et al.* Interact and Decide: Medley of Sub-attention Networks for Effective Group Recommendation[M]. New York, NY, USA: Association for Computing Machinery, 2019, 255–264. DOI: 10.1145/3331184.3331251.
- [18] LIN W Q, XIAO X K, XIE X, *et al.* Network Motif Discovery: a GPU Approach[J]. *IEEE Trans Knowl Data Eng*, 2017, **29**(3): 513–528. DOI:10.1109/tkde.2016.2566618.
- [19] VINH TRAN L, TAY Y, ZHANG S, *et al.* HyperML: a Boosting Metric Learning Approach in Hyperbolic Space for Recommender Systems[C]//Proceedings of the 13th International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2020: DOI: 10.1145/3336191.3371850.
- [20] WANG Z Y, WEI W, CONG G, *et al.* Global Context Enhanced Graph Neural Networks for Session-based Recommendation[C]//SIGIR, 2020.

- [21] Hai Z, Zhao P, Cheng P, *et al.* Deceptive Review Spam Detection *via* Exploiting Task Relatedness and Unlabeled Data[C]//EMNLP, 2016.
- [22] HU L M, LI J Z, NIE L Q, *et al.* What Happens Next? Future Subevent Prediction Using Contextual Hierarchical LSTM[C]//AAAI, 2017.
- [23] YANG J B, NGUYEN M N, SAN P P, *et al.* Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition[C]//IJCAI, 2015.
- [24] ZHU J, CHEN N, PENG W W. Estimation of Bearing Remaining Useful Life Based on Multiscale Convolutional Neural Network[J]. *IEEE Trans Ind Electron*, 2019, **66**(4): 3208–3216. DOI:10.1109/TIE.2018.2844856.
- [25] WU J Y, WU M, CHEN Z H, *et al.* Degradation-aware Remaining Useful Life Prediction with LSTM Autoencoder[J]. *IEEE Trans Instrum Meas*, 2021, **70**: 1–10. DOI:10.1109/tim.2021.3055788.
- [26] HINTON G E, VINYALS O, DEAN J. Distilling the knowledge in a neural network[C]//NIPS Deep Learning and Representation Learning Workshop, 2015.
- [27] FU Y J, LIAO J, LI Y, *et al.* Multiple Perspective Attention Based on Double BiLSTM for Aspect and Sentiment Pair Extract[J]. *Neurocomputing*, 2021, **438**: 302–311. DOI:10.1016/j.neucom.2021.01.079.
- [28] ZHENG S, RISTOVSKI K, FARAHAT A, *et al.* Long Short-term Memory Network for Remaining Useful Life Estimation[C]//2017 IEEE International Conference on Prognostics and Health Management. IEEE, 2017: 88–95. DOI:10.1109/ICPHM.2017.7998311.
- [29] ZHANG J J, WANG P, YAN R Q, *et al.* Long Short-term Memory for Machine Remaining Life Prediction [J]. *J Manuf Syst*, 2018, **48**: 78–86. DOI:10.1016/j.jmsy.2018.05.011.
- [30] GONG Y C, LIU L, YANG M, *et al.* Compressing deep convolutional networks using vector quantization [C]//ICLR, 2014.
- [31] WU J X, LENG C, WANG Y H, *et al.* Quantized Convolutional Neural Networks for Mobile Devices[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 4820–4828. DOI:10.1109/CVPR.2016.521.
- [32] CHENG Y, WANG D, ZHOU P, *et al.* A Survey of Model Compression and Acceleration for Deep Neural Networks[EB/OL]. 2017, arXiv Preprint: 1710.09282.
- [33] HAN S, MAO H Z, DALLY W J. Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding[EB/OL]. 2015, arXiv Preprint: 1510.00149.
- [34] NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing System (Volume 2), 2013: 2148–2156. DOI:10.48550/arXiv.1306.0543.
- [35] NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing System (Volume 1), 2014: 1269–1277. DOI:10.48550/arXiv.1404.0736.
- [36] SAPUTRA M R U, GUSMAO P, ALMALIOGLU Y, *et al.* Distilling Knowledge from a Deep Pose Regressor Network[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 263–272. DOI:10.1109/ICCV.2019.00035.
- [37] CHEN W C, CHANG C C, LEE C R. Knowledge Distillation with Feature Maps for Image Classification[C]//Comput Vis - ACCV 2018, 2019. DOI:10.1007/978-3-030-20893-6_13.
- [38] PAN S J, YANG Q. A Survey on Transfer Learning[J]. *IEEE Trans Knowl Data Eng*, 2010, **22**(10): 1345–1359. DOI:10.1109/tkde.2009.191.
- [39] RAGAB M, CHEN Z H, WU M, *et al.* Contrastive Adversarial Domain Adaptation for Machine Remaining Useful Life Prediction[J]. *IEEE Trans Ind Inf*, 2021, **17**(8): 5239–5249. DOI:10.1109/tii.2020.3032690.
- [40] RAGAB M, CHEN Z H, WU M, *et al.* Adversarial Transfer Learning for Machine Remaining Useful Life Prediction[C]//2020 IEEE International Conference on Prognostics and Health Management. IEEE, 2020: 1–7. DOI:10.1109/ICPHM49022.2020.9187053.