# Application of Convolutional Neural Network to Prediction of Temperature Distribution in Data Centers

# 摘要

我们提出了一个通过使用卷积神经网络（CNN）预测数据中心温度分布的模型。温度分布的变化取决于数据中心的本地结构，例如设备位置和服务器类型。尽管数据中心中的各种物理关系在我们以前的工作中被模拟为网络，但没有自动提取数据中心结构的机制。 CNN的使用是自适应地学习局部结构的技术，它允许学习复杂的特征，例如数据中心中的各种物理关系。我们通过使用来自实验数据中心的实际数据来评估所提出的模型的性能。评估表明，所提出的模型可以预测在0.42毫秒内超过48个位置20分钟的未来温度分布，均方根误差（RMSE）为0.96度。与简单的线性预测模型相比，这种精度是一个显着的改进，并且准确度足以允许基于这些温度预测来控制空调。

关键词：- 数据中心，功耗降低，温度预测，卷积神经网络

# 1.简介

由于为满足社交网络服务和云服务等大规模在线服务需求而构建的高性能数据中心数量庞大，数据中心功耗已成为严重问题[1]。 因此，有关于节能数据中心的研究[2] - [6]。 例如，Khuller等人 提出了一个能源管理系统，将所有任务分配给最少数量的服务器，其余服务器关闭[7]。 在[8]中，Iyengar等人根据数据中心测量的温度分布提出了一种空调控制方法，以降低空调能耗。 有多种方法可以降低数据中心内各种设备的功耗，如空调，电源设备和服务器[9] - [11]。 但是，大多数现有的方法不考虑设备之间的相互影响。

为了减少数据中心的总能耗，重要的是将数据中心视为一个单元。 如果降低其他设备的功耗，则降低某些特定设备的功耗可能不一定会降低整体功耗。 服务器和空调占数据中心能源消耗的大部分，而且它们的消耗水平也是强相互依赖的（图1）。 因此，基于数据中心建模的协同控制至关重要。 但是，由于各种设备之间的相互依赖关系，数据中心并不容易建模。

机器学习技术，尤其是深度学习，近年来备受关注，因为成功应用于由于问题的复杂性而被认为几十年困难的各种“实际”问题。 文献[12]表明，深度学习在着名的“自然”图像分类任务中超越了使用手工特征提取的传统机器学习方法[13]。 有人提出，深度神经网络可以自动学习问题的重要特征，例如图像中像素的几何约束。 例如，一个在人类图像上训练过的深度神经网络产生了对人脸或身体有反应的神经元[14]。

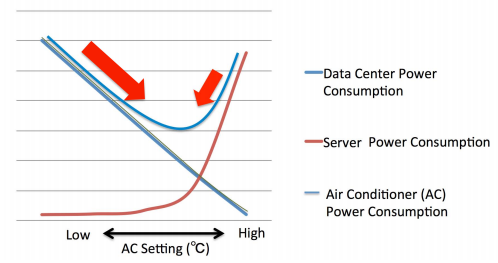


Fig1功耗与空调设置之间的关系 通过调高空调功率（较低的温度设置）可以降低数据中心的功耗。 但是，如果我们降低空调功率（较高的温度设置），服务器故障率会增加，服务器功耗增加，特别是服务器风扇功耗。

在本文中，我们提出了一种基于深度神经网络预测数据中心温度分布的方法。 预测模型对于节能数据中心管理至关重要。 由于传热取决于每件设备和气流的物理位置，因此采用能够捕捉这种几何约束的学习方法预计会是有益的。 在我们以前的工作[15]中，我们提出了一种网络模型，其中连接结构是手工设计的，以匹配数据中心的结构。 但是，没有自动提取结构的机制。 由于卷积神经网络（CNN）能够提取目标的几何结构，我们使用这个网络进行预测。

本文的其余部分安排如下。 我们解释了第二部分中数据中心的温度分布是如何预测的。 在第三部分，我们提出了评估指标和所提出的方法在预测精度方面的评估结果。 第四部分总结了本文，并展示了未来的研究工作。

# 2. 数据中心温度分布的预测

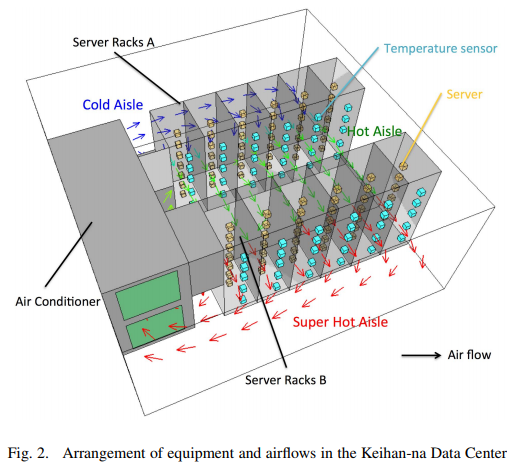
在数据中心中，ICT设备和空调使用大部分电力。 一般来说，增加空调的温度设置会降低空调的功耗，但会增加服务器的功耗，如图1所示。由于服务器的故障率随着进气温度的升高而增加，所以 从空调出来的空气通常设置为低价值，而不考虑能源效率，特别是当人力资源不足或缺乏数据中心管理资源成本意识时[16]。 自动管理数据中心的机制对于克服此问题非常重要。

为了最大限度地减少整个数据中心的总功耗，因此重要的是不仅要考虑独立降低每个设备的功耗，还要考虑对多台设备进行协调控制。 在本文中，我们提出了一种使用机器学习技术来预测数据中心温度分布的方法。 这是数据中心合作控制的必要组成部分。

我们的研究小组建立了一个串联结构数据中心（图2），用于在高能效数据中心进行演示实验; 这个中心叫做Keihanna数据中心[17]。 与传统的数据中心不同，流入冷通道的冷空气被服务器机架A和B加热两次。因此排气比常规数据中心更热，排气热量可以重复使用，例如空调周边区域。 由于热空气是该能量再利用系统中的高效热源，因此数据中心设备的协同控制对提高效率非常重要。 通过采用CNN，预计由数据中心结构的物理特性引起的传热的几何约束将能够被自动提取

## 2.1 数据中心的主动控制

我们的研究小组已经提出了对运营设置的主动控制，例如服务器和空调的任务分配，作为降低数据中心总功耗的一种方法。主动控制采用以下过程来确定数据中心的操作设置。首先，生成许多候选操作设置。其次，预测给定候选操作设置的数据中心的温度分布和功耗。最后，从提供可承受的温度分布的操作设置中，选择具有最小总功耗的设置。所选的操作设置在数据中心中执行。目标是根据总任务将总功耗降至最低。数据中心实施了各种响应时间的数台设备。因此，主动控制需要运行最佳条件以最小化功耗。当计划传入任务的进程时，或者事先从以前的案例中明显看出最佳条件时，主动控制是最有效的。



## 2.2 数据中心获得的数据

数据中心安装了一个传感器系统来测量数据中心的状态。 记录传感器值以训练预测模型，并且还用于在操作期间管理数据中心。 图2中的青色框表示放置在服务器机架A和B的出风口处的温度传感器。虽然分辨率不是特别高，但传感器值指示每个机架旁的排气的温度分布。

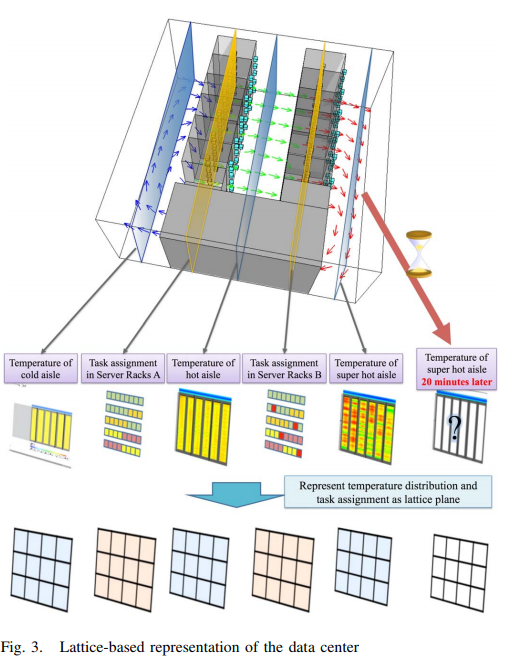
在服务器机架A和B中安装了150台和206台服务器（概念上用图2中的黄色框表示），并且还记录了每台服务器的CPU使用情况等操作条件。 这些值表示服务器上的负载分布，即热源分布。

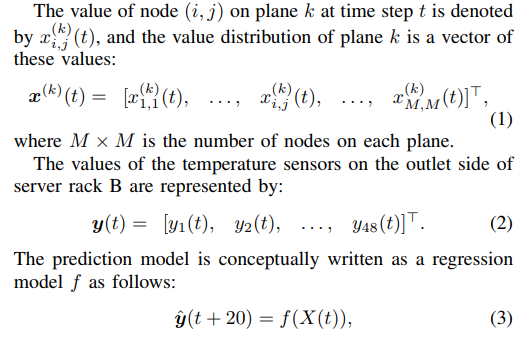
空调温度设置也以1分钟的采样间隔进行记录。 所获得的数据随时间线性内插以填充缺失的数据值并获得一致的采样率。

## 2.3 基于格的分布表示

我们假设所测量的分布可以被汇总为二维平面上的值分布。 为了简单起见，我们为每个通道和服务器机架分配一个平面，并将节点分配给每个平面的网格点。 服务器机架平面上的节点表示服务器CPU使用率，而通道平面上的节点表示温度。 因此分布近似为基于格子的表示。

索引号分配给每个平面，平面1分配给冷通道，平面2分配给服务器机架A，平面3分配给热通道，平面4分配给服务器机架B和平面5分配给超热通道（由图3中的黄色和蓝色平面表示）。



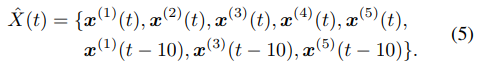


输入X（t）由传感器值和数据中心的运行设置组成。 严格地说，输入X（t）必须包含关于数据中心的所有信息，例如每个点的气流和每个组件的温度，但在本研究中，假设传感器值及其时间历史包含足够的信息 进行预测。 由于温度分布不会立即改变，我们的模型预测未来20分钟的温度。 学习的目的是获得这个函数的一个很好的近似值，但这并不容易，因为函数的输入是一个高维向量，f可能是一个复杂的非线性函数。

**因此，我们采用神经网络模型，更恰当的是卷积神经网络（CNN）[18]来逼近函数：**

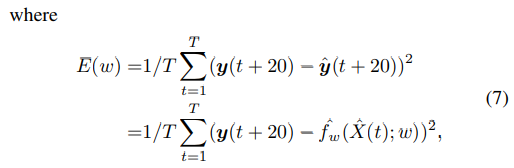


其中w是CNN的权重参数向量。 此外，网络的输入被定义为：



在图3中，y（t + 20）由右侧的白色格子状平面表示，X（t）由其他蓝色和黄色平面表示。 权重参数经过优化，以将估计值和观测值之间的误差最小化为：





T表示训练数据的样本大小，并且训练数据从运行中的数据中心收集

## 2.4 卷积神经网络

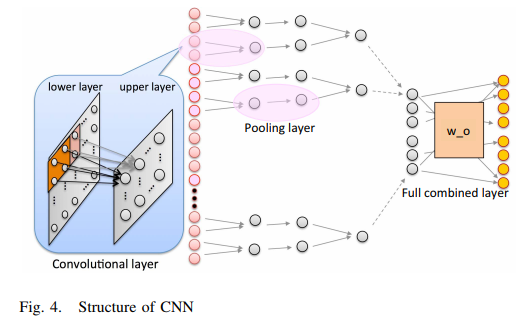
图4显示了CNN的结构。 CNN是一种多层感知器（MLP）。 传统的MLP和CNN的区别在于网络的连接结构。 传统的MLP只有少量的完全连接层，但CNN有两种类型的深层的判别层。

第一层是卷积层。 该图层具有本地连接和共享权重。 并将连续层的神经元作为过滤器。 然后网络通过训练学习诸如几何结构等特征

第二层是池化层。 该图层收集出色的特征。 由于这一层的存在，使得训练后网络的输出是平移不变的

CNN包含卷积层和汇聚层的多次重复。 CNN通常有一些完全连接的层到达网络尾部。 CNN因此能够自动学习各种局部结构。

**在我们提出的方法中，预测模型由CNN组成，每个平面都有一个通道。 CNN是可以自适应地学习局部结构并可以学习复杂特征的技术。 因此，预计CNN将能够自动学习网络节点之间复杂的相互依赖关系**



## 2.5 温度预测CNN网络结构

网络的输入由8个通道的二维阵列组成，输出是一个48维向量。 第一层是一个具有32个2×2滤波器的卷积层。 第二层是2×2池化窗口和2×2池化步幅的最大池化层。 第三层是具有16个4×4滤波器的卷积层。 第四层是2×2池化窗口和2×2池化步长的最大池层。 最后一层是具有线性激活功能的完全连接层。 总层数是5。

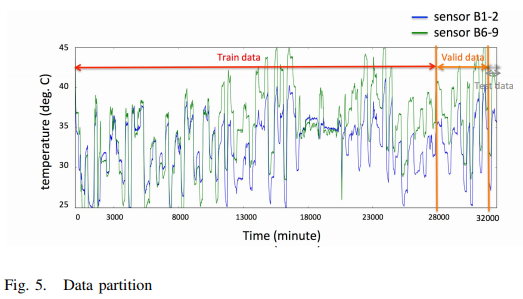
我们使用随机梯度下降（SGD）来训练网络，学习率为0.0001，批量为100.当最后10个迭代批次中的70％没有更新E(w)时，停止训练。

# 3.实验

## 3.1 数据集和评估环境

使用数据中心在运行过程中获得的数据评估了所提出方法的性能。 运行数据包括温度传感器的值，服务器的CPU使用情况以及空调设置温度数据。 所使用的数据是从2015年12月9日至2015年12月31日从数据中心获得的32,640分钟的数据，间隔为1分钟。 我们使用第一批28,000分钟的数据作为模型的训练数据，接下来的4000分钟进行验证，最后使用500分钟进行测试。

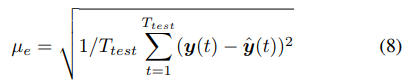
服务器机架B的出口侧有48个温度传感器，由机架顶部的机架名称和传感器位置组合来表示。 例如，机架B1顶部的第二个温度传感器是B1-2传感器。 机架B1有3个传感器，每个其他机架有9个传感器。 图5显示了传感器B1-2和B6-9处温度传感器数据的分区。 传感器B6-9的平均温度最高，传感器B1-2的平均温度最低。



我们使用pylearn2 [19]实现了所提议的模型，该模型运行在具有32GB内存的Ubuntu桌面14.04 PC，GeForce GTX TITAN X GPU和Intel Core i7-6700K 4GHz CPU上。

## 3.2 评估指标

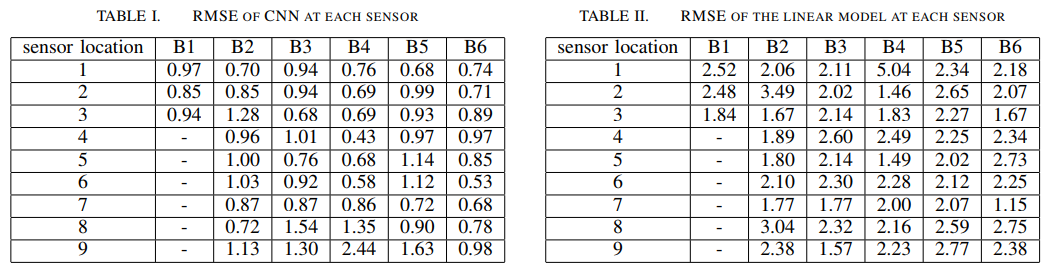
将所提出的方法与线性回归模型和MLP在学习和预测的准确性和计算时间方面进行比较。 为了评估模型精度，我们计算了方程（8）的RMSEμe，其中Ttest是预测数据的数量。



## 3.3 评价结果

### 3.3.1 预测精准度

表I和表II分别总结了CNN，Linear模型和MLP预测的温度的RMSE。 表I中平均RMSE为0.93度，表II中为2.24度，表III中为0.74度。 因此，CNN比简单的线性预测模型提供了显着的改进。 此外，根据数据中心空调器的运行数据，当设定温度增加1度时，空调器的功耗增加8％。 因此，CNN模型的RMSE值足够小以允许基于温度预测来控制空调。 而且，与MLP相比，CNN并没有提高准确性。 这是因为测试平台数据中心的相互依赖关系并不那么复杂。 然而，当我们预测数据中心的温度分布具有更多的设备和传感器以及更复杂的布局时，假定CNN的准确性改善超过MLP



在表I中，机架底部的传感器精度往往较低。 数据中心的机架下有一个通风口，导致外部空气流入机架底部。 准确性可能很低，因为该模型没有考虑到这种影响。

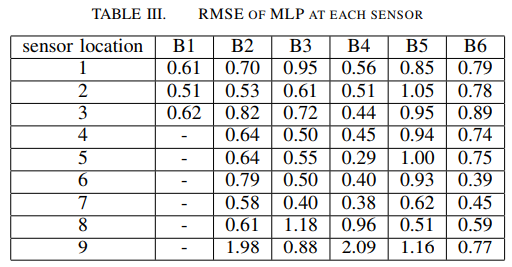
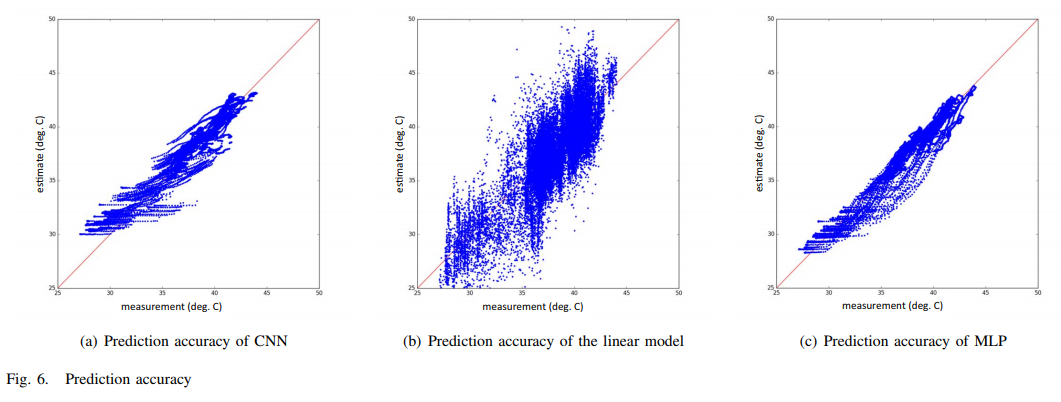


图6显示了CNN和线性模型的预测和测量温度的散点图。 如图6所示，CNN比线性预测模型有显着的改进。 但是，低预测值有水平分布的倾向，表明该模型无法获知温度变化的所有特征。 通过改变网络结构和培训更多特征，可以提高模型的准确性。

### 3.3.2 学习和预测的计算时间

使用28,000分钟的数据进行训练的CNN的计算时间以及使用500分钟的数据进行的预测（其对应于对48个温度传感器进行500次预测）分别为约4分49秒和2.1秒。 计算时间很短，因为网络结构大大简化。 由于精度和计算时间之间通常存在折衷，因此需要根据情况决定最适合模型的结构。 因为我们收集数据并使用这些数据进行训练以预先构建预测模型，所以我们不需要考虑学习时的计算时间。 根据这些结果，所提出的模型能够在几秒钟内做出预测。 因此，我们得出结论：我们提出的方法可用于数据中心的实时温度分布预测



# 4.结论和未来工作

我们提出了一个模型，用于预测基于CNN的数据中心的温度分布，实现节能数据中心管理。 实验结果表明，我们的模型提供准确的预测，所提出的模型能够以几毫秒的计算时间预测0.93度RSME内的温度分布。 对于数据中心能源管理系统来说，这种精度预计足以进行精细的控制，并且计算成本足够低，以便根据采样控制方法的需要多次重复预测[20]。

在今后的工作中，我们计划通过搜索更有效的网络结构来提高所提出模型的预测精度。 我们还打算使用此模型来降低实际数据中心的功耗。