**Title**

**标题：基于工程化方法的LSTM智能温控系统设计**

**Summary**

温度系统设计需要从系统设计抽象到数学模型。我们选择的抽象方法是基于UML统一建模语言的系统建模方法。对于温度控制系统问题域，使用用例图描述用户和温控系统的可能使用场景，然后设计MVC体系结构框架解决温控系统与用户的交互问题，使用类图描述如何控制温度这个问题域中出现的对象，建立了一个最基本的问题域模型。最终确定要研究的核心问题是在中央处理器上如何综合传感器发来的和用户发来的数据。这个核心问题就是所要研究的数学模型。下面详细阐述从智能温度系统中抽象数学问题的方案。

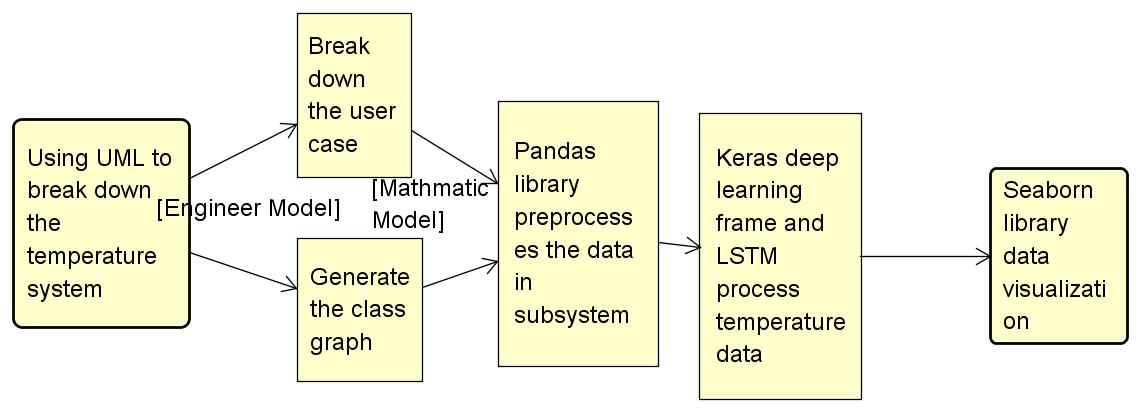


Figure 1工程化建模过程流

1、面向对象分析——用例建模

**用例建模帮助我们理解需求，明确任务。用例包括智能温度系统中的场景以及外部对象。外部对象包括用户和中央处理系统。对象的交互过程通过用例完成。用户面对的场景是编程控制温度。假设用户编制的程序是能够表征温度随时间变化的函数。然后通过程序，翻译成MVC框架中的控制器所能识别的数据流，进入学习网络，激活控制器收集传感器数据，进而控制屋内的温度。节省能量、适应用户的喜好这两个用例，在某种程度上存在矛盾，一个系统在运行的过程中必须要先保证用户的满意才能节省能量，这是系统设计的基准点。**

**用户的喜好判断由系统完成。有关用户喜好的用例包含设置温度程序的子用例。用户编程设定的温度场景通过嵌入式终端实现，如通过手机设定温度。考虑到用户手动编程设定温度非常麻烦，而且效率极低。因此，分析设计一个用户可图形编程的界面与用户交互十分必要。这也就是下文体系结构框架里所述的界面设计。**

**与系统有关的用例都能和系统作用，从而感知外部的温度变化，温度变化主要从传感器中感知。其中有一个扩展用例感知的是湿度的变化，不仅需要湿度传感器，还要感知空气污染水平的传感器。系统除了要从这几个传感器中接收数据，并计算出温度应该怎么变化，而且应该能够算出在室内有多少个人，根据人数进一步调整室内温度。人数较少时系统工作强度大，人数较多时系统工作强度小。这样能够有效节省能量。**

**综上所述，智能温度控制系统要有一个面向用户的图形界面，能够在界面上进行显式的编程。同时系统也能够根据用户的数据，学习出一个模型，根据用户的人数，根据传感器发出来的各种数据去学习固定的模式。**

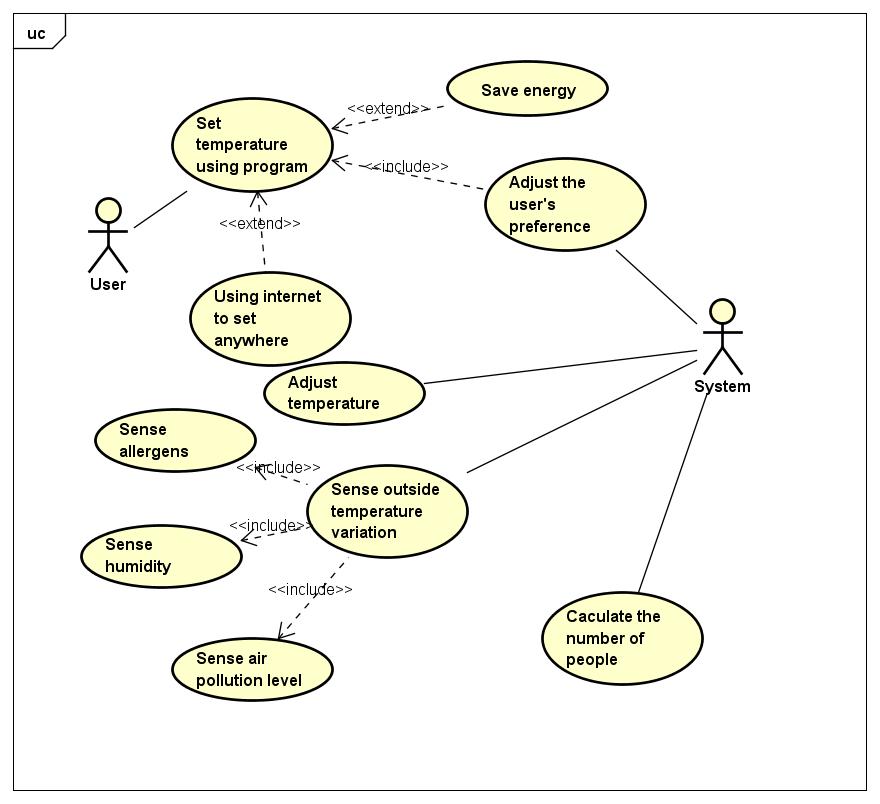


Figure 2温控系统需求分析用例（UML统一建模语言）

2、体系结构框架设计——MVC框架

我们基于MVC设计出面向用户和智能温度系统的体系结构框架。MVC框架是一个体系结构框架，由三部分组成：模型对象，视图对象和控制器对象。

其中的控制器对象实例化为温度系统中心处理器。它控制智能家庭系统中的任务，并完成核心的计算功能。

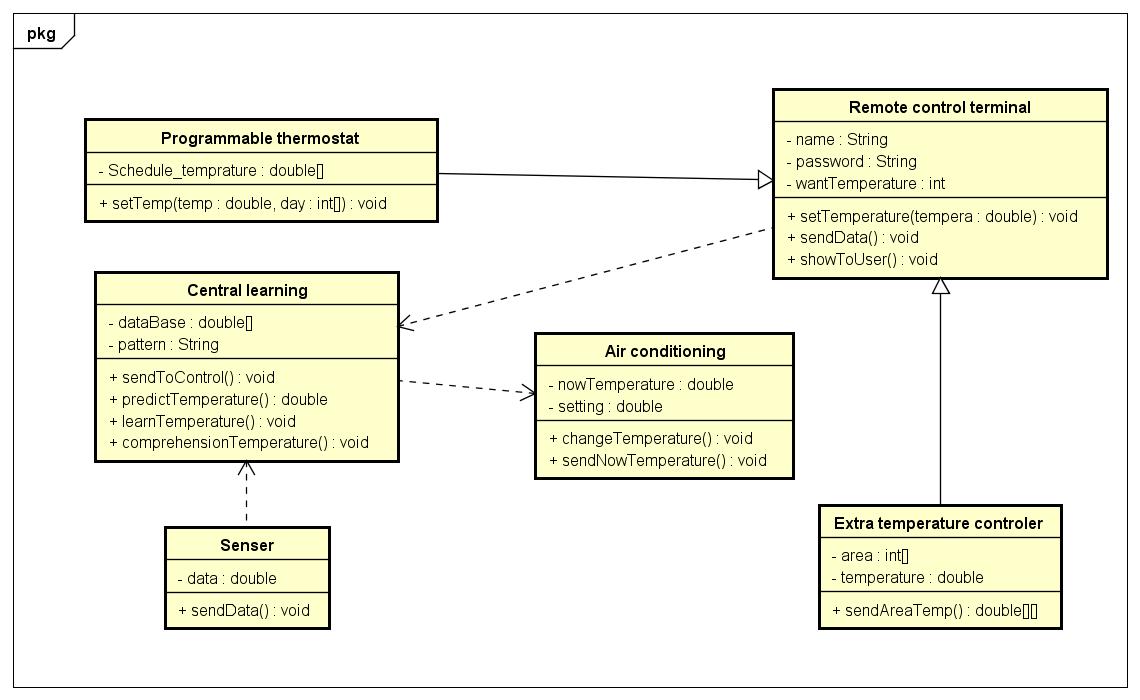
视图对象是人机交互的窗口，当用户需要访问智能控制家居系统时，首先需要通过手机应用或者电脑端应用访问系统界面，在下面解释中我们会看到，界面应该具有可编程的特性，以便让用户去图形化编程，在控制系统的温度的同时能够方便地与系统交互。

模型对象具体指LSTM神经网络模型。我们应用Keras深度学习框架实现了这个模型，具体的模型将在后文中讨论，相关的代码以及注释作为附录。

3、数学模型生成——基于类的分析

**下面根据场景建模，进一步分析需求并系统设计。类图把场景映射成抽象的对象，抽象出数学模型。程控温度设定器的实质是上述建立MVC框架中所述图形温度控制界面。程控温度器对象中的方法把用户的输入、编写的程序转换成移动数据，发送给远程终端。远程终端能够感知到消息。远程终端同时发送数据给MVC框架中的核心处理器，简而言之是一台服务器。在物联网中设置的服务器是嵌入式为主的微服务器。考虑到在建模过程中用到神经网络，在系统设计时，处理器必须预留USB神经网络加速器接口，否则使用会有较大局限。**

Figure 3基于类的建模

**另一方面，我们要考虑到题目中所给的一些其他限定，只有一个温度调节器和额外温度调节器的情况下，分别对系统有什么影响？额外的温度控制器对象继承于远程控制终端。**

**综上所述，在基于类的建模过程中，最终把智能温度系统设计这样的实际问题归结为数学问题——处理各个传感器和用户发来的数据。在接下来过程中，我们要把这个对工程的建模转化成对数学问题的建模，更详细的讨论central controler中的数学问题。**

**Contents**

1. **Introduction**
   1. **Background**

In recent years,

* 1. **Problem Statement and Analysis**

In recent years,

1. **Assumption and Symbol Explanation**
   1. **Assumption**

* In recent years,
* In recent years,
  1. **Symbol Explanation**

|  |  |
| --- | --- |
| **Symbol** | **Definition** |
| *A* |  |
| *B* |  |
| *C* |  |
| *D* |  |
| *E* |  |
| *F* |  |
| *G* |  |

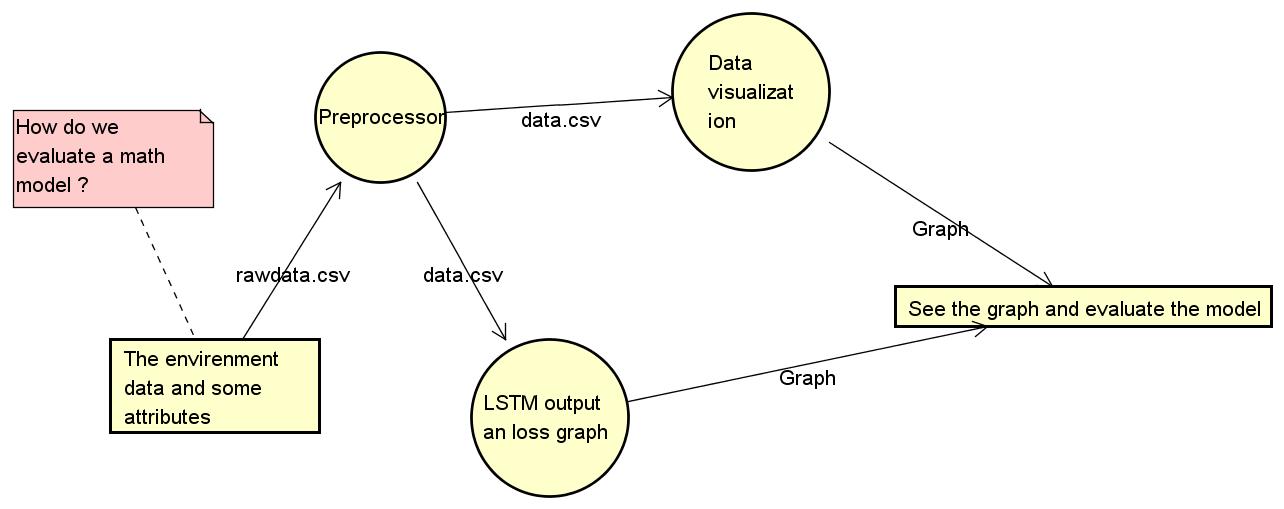
1. **Task 1**
   1. **Data Pre-processing(因为这个题正好用到了数据的预处理，就给它先列出来了)**

数据收集

本文使用深度学习的方法分析数据，因此需要收集较多的数据。收集到的数据有8760条，来自马塞诸塞大学实验室，反映室内室外的环境变化，以及人体的体感温度。数据集共有9个属性，分别为temperature,icon,humidity,apparentTemperature,pressure, windspeed, time, windBearing, dewpoint.通过数据集分析，找出数据间相互作用规律，进而对中心温度控制器建模，最后验证LSTM神经网络模型。

|  |  |
| --- | --- |
| Interpreter of Data field | |
| Temperature | The inside temperature |
| Icon | The weather condition |
| Humidity | The inside humidity |
| apparentTemoerature | Sendible temperature |
| pressure | Stmospheric pressure |
| windspeed | Wind Spees |
| time | Local Time |
| windBearing | Wind direction |
| dewpoint | Dew point temperature (Meteorological academic language) |

**数据预处理**

下面展现了通过构建数据流图显示需要进行的数据处理。通过对数据可视化图像的分析，找出数据的一般规律，进而利用LSTM神经网络回归数据。数据集存在如数据缺失、存在字符串属性等部分数据不够规范化的问题。

我们使用Python语言的Pandas库和sklearn库进行数据预处理，运行环境是Linux。代码和注释在附录中，分为两个文件解决不同类型的数据正则化问题。

缺失数据的解决方案。缺失数据集中在三列上，缺失量不大，因此我们采用缺失值填充方法而未采用缺失值剔除方法。根据学界有关学者的研究，连续型数据用该属性的平均值去填充，而离散属性使用众数填充数据。鉴于缺失数据全部为连续型的数据，我们通过平均值法填充。

用户数据的填充方案。原始数据集并未提供用户根据具体环境做出的反映，因此我们作为模拟用户尝试填充了小部分调节温度，反映我们自身对于该项环境下的体验。经过80项人工判断填充，我们大致掌握了温度的调节规律。并依规律大规模人工向下填充。经过方案可行性分析，这种人工填充方法模拟了用户的实际行为，能够较好地反映真实系统下用户的有关操作。

非数值属性的编码问题。原始数据中，表现室外天气情况的icon字段出现了{'clear-night', 'wind', 'partly-cloudy-night', 'rain', 'cloudy', 'fog', 'partly-cloudy-day', 'clear-day', 'snow'}这样的字符串。需要将其映射成为整数。考虑到最终用户的输入数据有可能与光线强弱有关，我们按照不同天气情况对应的光线强弱由弱到强依次编码为0-9.为下一步矩阵计算创造数值方面的条件。

时间序列的生成方案。因为数据集提供了时间字段，因此数据预处理时需要根据时间字段排序，以反映时间序列的特征。由于LSTM神经网络是时间序列的网络，我们需要在原始数据的基础上建立滑动窗口。这里我们设置的滑动窗口长度是2，每次窗口滑动一个单位。这里需要特别注意，建立了时间序列后，原始的数据从二维矩阵变换为三维张量。

数据的归一化问题。神经网络的基本原理是把线性不可分的数据映射到高维空间，增强数据的线性可分性。数据归一化可以让数据集中分布在较小的范围内，既可以增强激活函数的非线性映射能力，又可以增强数据集的表现效果。避免数据在经过多层网络时因为计算机浮点运算能力的问题出现梯度消失或者梯度爆炸问题。

**数据可视化分析**

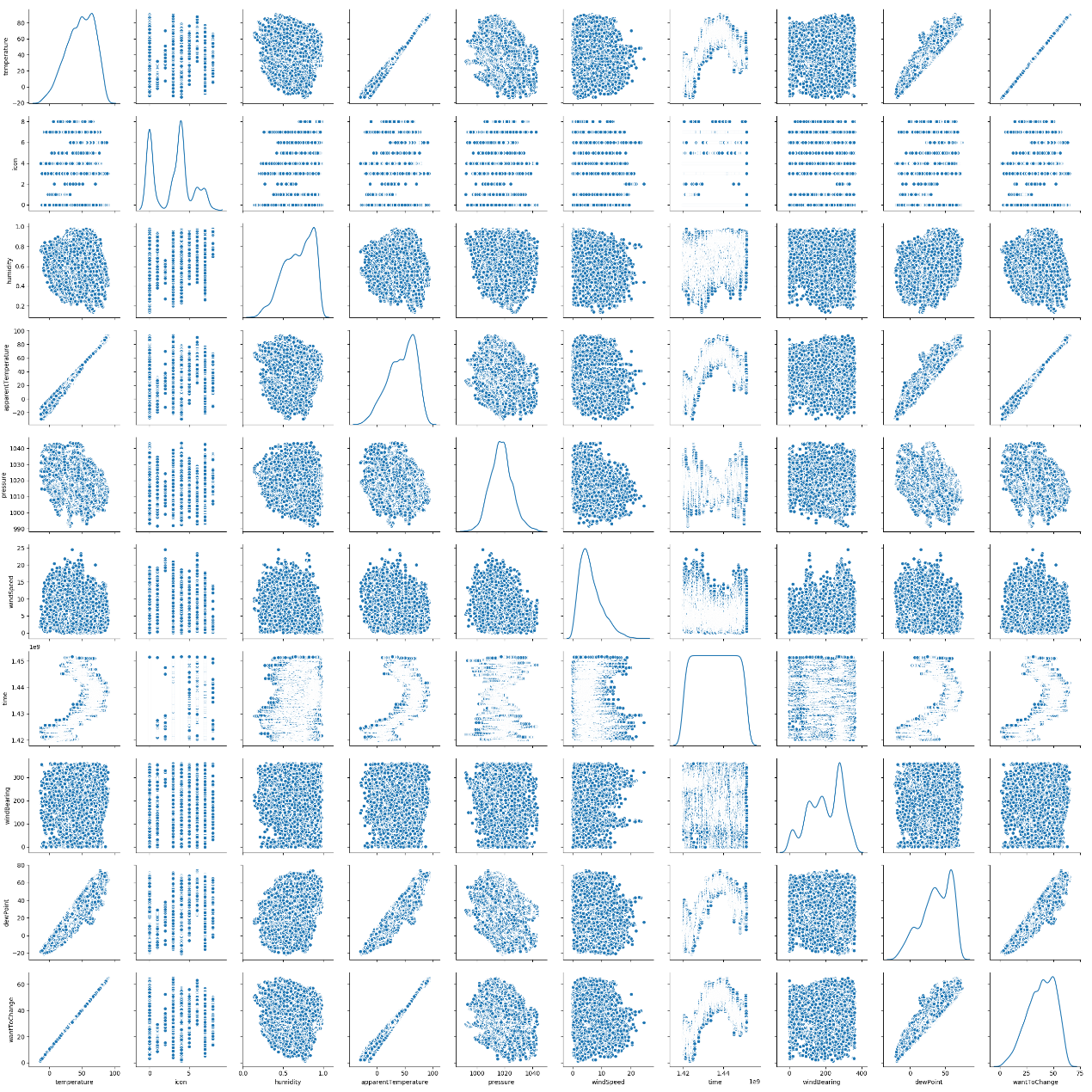
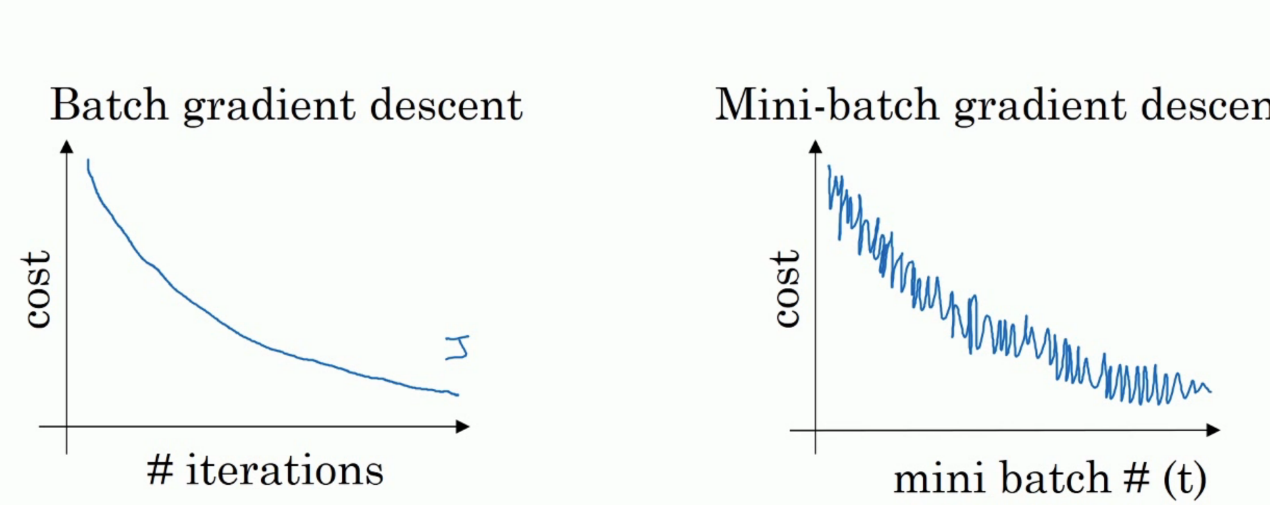


Figure The scatter and distribution Graph

上图展示的是数据集的分布情况，横轴从左到右依次为temperature, icon, humidity, apparentTemperature, pressure, windspeed, time, windBearing, dewpoint, wantToChange.纵轴从上到下与横轴相同。图中对角线上展示的是不同属性的分布情况。可以清晰的发现，数据的分布大多都向正态分布靠近，这也与大数定律的结果相吻合，实验次数越多，分布就向正态分布靠近。

散点图中的time属性具有相当明显的周期性，这说明温度和环境的有关数据和信息都在随着时间的变化呈现周期性的变迁。而温度相关的属性，apparentTemperature, temperature, wantToChange。

LSTM方法建模

数据的正则化方法。为了防止模型过拟合，需要对数据进行正则化处理，这里我们采用数量为20的minibatch方法。如图所示，在不使用minibatch的传统梯度下降算法中，损失函数的下降速率并不规则，这是由于单个梯度下降具有随机性，很难精准的向总体均方误差最小化的方向移动。而minibatch往往具有代表性，能够反映数据总体在误差上的表现。

为了精准比较误差，我们控制LSTM网络返回整个序列而不是一个序列的单个输出。

模型选择方面，全连接型神经网络和卷积神经网络不适合。卷积神经网络一般是用在图像的处理里面，解决高维数据压缩问题，但是这个数据集维度仅为9个属性。全连接神经网络无法很好的处理时间序列问题。朴素贝叶斯方法对于样本数量较少的情况下拟合效果好，但是缺乏时间序列的判断性。递归神经网络已经在很多自然言语处理任务中 表现出很强的学习能力，特别是能够对序列数据进 行很好的建模，充分挖掘出序列中隐藏的信息。如RNN循环神经网络，能够表现时间序列的变化，但是因为缺乏控制流。因此不能较好地对长时数据建模。

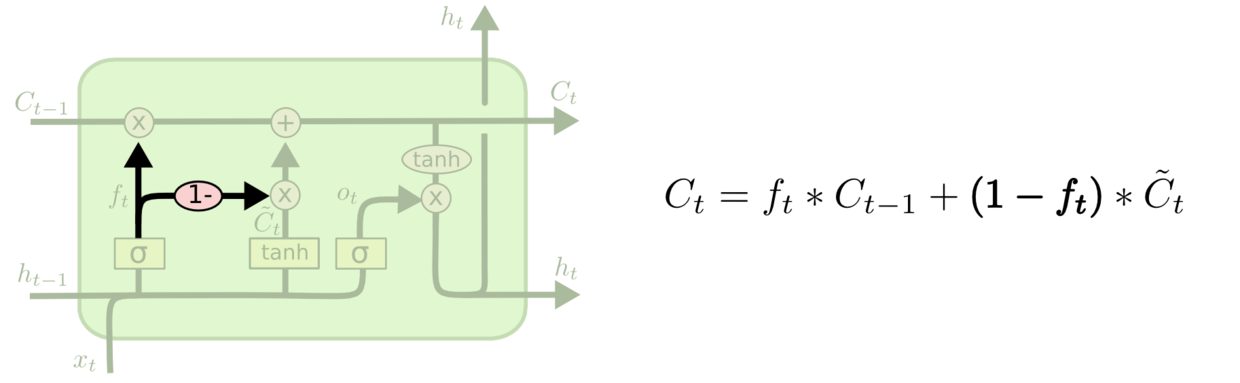
标准的LSTM神经网络模块有4个可交互的神经网络层组成，避免在迭代次数增加的情况下出现梯度消失或者爆炸的问题，保证模型得到充分的训练，从而提高模型的预测准确度。其中σ 为sigmoid，tanh为tanh层。

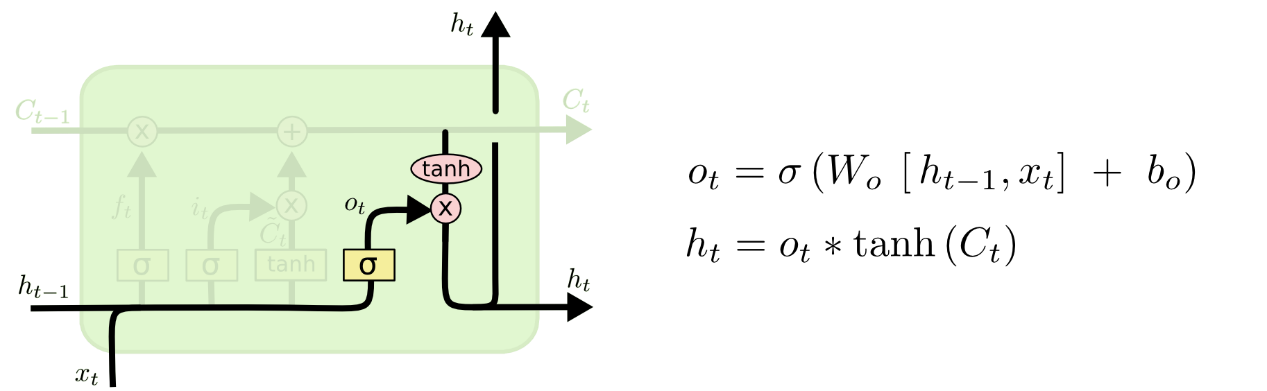
ｔａｎｈ（ ｘ ）＝(１－exp(-2x))/ (１＋ｅxp(－２ｘ))

Sigmoid（x）=1/(1+exp(-x))

每个神经网络模块可以看作一个Memory cell，其中Xt，Ct，Ht分别表示第t时刻的输入单元、细胞状态和输出单元，而Ct-1,Ht-1,分别表示第t-1时刻的细胞状态和输出单元。每个“记忆细胞”包含三个“门”结构，分别是：“Forget Gate”,”Input gate”,”Output gate”.”Forget gate “决定从细胞状态中忘记哪些信息；”输入门“决定将要更新哪些信息到细胞状态中；”输出门“决定将要输出哪些信息。“输入门”层采用与“遗忘门”层耦合的方式对向 “记忆细胞”注入的信息进行控制，共同决定细胞状 态的更新。用 Ct表示将要注入“记忆细胞”新信息 的向量，它为tanh神经网络层的输出.下面的公式中，分别用ft表示Forget gate中需要遗忘信息的比率，（1-ft）表示当前输入的信息需要传递给细胞状态的比率。与此类似，ot表示经过sigmoid函数放缩过后的预输出，其中sigmoid函数的作用对象是输入数据的一个线性映射。ht表示经过前一个细胞信息与当前输入综合后的信息流，是预输出与细胞控制流的乘积。

进一步解释模型，tanh起到信息流的非线性映射作用，sigmoid函数起到了信息流开关的控制作用。而sigmoid和tanh函数在自变量较大时都会出现梯度消失的现象，而且他们的输出是有有界的，这进一步论证了我们在数据预处理阶段对温度集所有数据按列整体归一化的原因。必须确保网络输出能够尽可能映射成为实际数据，也要确保控制流能够较好地进行放缩。如果数据做归一化，则实际应用时必须要进行反归一化才能输出正确结果。





下面是基于Keras框架设计温度集上LSTM的网络结构：

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=========================================

lstm\_1 (LSTM) (None, 2, 4) 208

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

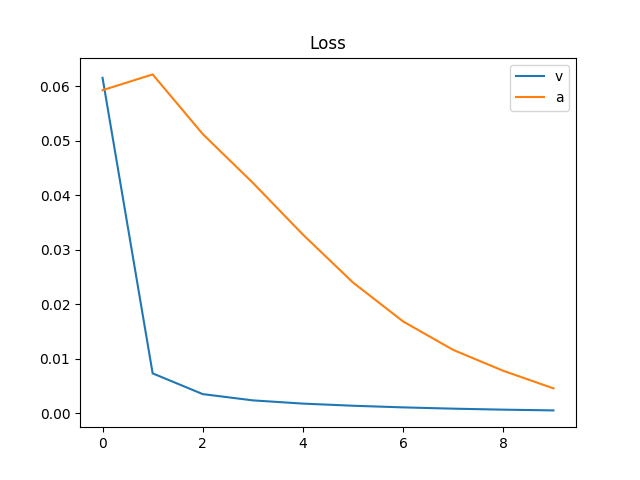
dense\_1 (Dense) (None, 2, 1) 5

=========================================

Total params: 213

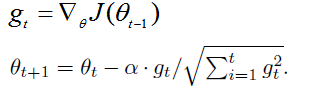
Trainable params: 213

Non-trainable params: 0



图中所示为LSTM模型拟合过程中的损失下降曲线，v stands for validation set，a stands for train set。可以看出，使用minibatch方法，在迭代少于10次时下降损失已经较低。我们认为，这是因为模拟用户温度的设置偏向于线性设置，因此模型能够在较少的迭代次数下快速完成用户数据的分析。

因为现在并不知道用户设定的实际温度是多少。所以接下来需要用自适应方法优化模型。我们仍然使用传统的均方误差作为损失函数。除了使用minibatch外（在数据预处理部分讲述），还使用Adagrad优化器加速模型的收敛速度。



随机梯度下降方法中，对于每一个参数的训练使用了相同的学习率α。Adagrad算法能够在训练中自动的对learning rate进行调整，对于出现频率较低参数采用较大的α更新；相反，对于出现频率较高的参数采用较小的α更新。因此，Adagrad非常适合处理稀疏数据，而温度数据集中存在的离散数据满足了稀疏数据的条件，适合处理此数据集。

我们设*gt*,*i*为第t轮第i个参数的梯度，即*gt*,*i*=▽Θ*J*(Θ*i)*。因此，SGD中参数更新的过程可写为：

Θ*t*+1,*i*=Θ*t*,*i*−*α*⋅*gt*,*i*

Adagrad在每轮训练中对每个参数的学习率进行更新，参数更新公式如下：

Θ*t*+1,*i*=Θ*t*,*i*−*αGt*,*ii*+*ϵ*√⋅*gt*,*i*

其中，*Gt*∈R*d*×*d*为对角矩阵，每个对角线位置（i，i）为对应参数*θi*从第1轮到第t轮梯度的平方和。ϵ是平滑项，用于避免分母为0，一般取值1e−8。Adagrad的缺点是在训练的中后期，分母上梯度平方的累加将会越来越大，从而梯度趋近于0，使得训练提前结束，但是在温度训练集上我们并没有观察到这样的情况出现。

LSTM的工程化应用

上述理论部分已经说明，LSTM用于温度数据分析在理论上是可行的，本节说明LSTM用于Central learning的子系统仍然可以提供给用户最佳体验。

假设物联网系统使用的是Intel up square board作为中心处理器。可以安装tenserflow-lite版本运行上述神经网络，但是并不时时刻刻运转，而是首先记录用户输入数据、输入时间以及传感器数据。等待用户输入的数据量达到一个minibatch然后集中送入LSTM神经网络。这样设计的优点在于可以在优化梯度下降的同时改善功耗，达到节能的目的。此时用户可以通过手机应用访问物联网中心，在手机上编程设定温度，达到温度控制的效果。物联网控制中心的控制子系统随时根据传感器发来的数据做出应答，此时控制子系统使用的是已经建好的数据库，并非正在学习中的。实现控制与计算分离，最大程度降低功耗。

Appendix 1: Python code on Data preprocessing (Based on Pandas library)

|  |
| --- |
| #--------------------------------------------------------  # Copyright 2018 Kaiyan Chang, Kaiyuan Tian, Ruilin Chen  # For the data preprocessing  # Python 3.6 for Linux. Only run in Linux system.  #--------------------------------------------------------  import numpy as np  import pandas as pd  import matplotlib.pyplot as plt  import seaborn as sns  #Read data from csv  raw\_data=pd.read\_csv("datar.csv")  temp=set(raw\_data['icon'])  print(temp)  #Choose a best index for icon  size\_mapping={label:index for index, label in enumerate(temp)}  raw\_data['icon']=raw\_data['icon'].map(size\_mapping)  #Caculate the user defined temperature  sigma=0.2  mu=0  wanttochange=0.607\*raw\_data['temperature'].values+10.092+np.random.normal(mu,sigma,raw\_data.shape[0])  #Fill NaN  raw\_data['wantToChange']=wanttochange  raw\_data[raw\_data.isnull().values==True]  raw\_data['pressure']=raw\_data['pressure'].fillna(raw\_data['pressure'].mean())  raw\_data['windBearing']=raw\_data['windBearing'].fillna(raw\_data['windBearing'].mean())  raw\_data['windSpeed']=raw\_data['windSpeed'].fillna(raw\_data['windSpeed'].mean())  #Sort by time  raw\_data.sort\_values(by="time")  raw\_data.to\_csv('datanoraw.csv',index=False)  sns.pairplot(raw\_data,diag\_kind='kde')  plt.savefig('datasetprocessingpairplot1')  raw\_data |

Appendix 2: Python code on Data regression (Based on Keras and tensorflow library)

|  |
| --- |
| #--------------------------------------------------------  # Copyright 2018 Kaiyan Chang, Kaiyuan Tian, Ruilin Chen  # For the data regression  # Python 3.6 for Linux. Only run in Linux system.  #--------------------------------------------------------  from matplotlib import pyplot as plt  from pandas import read\_csv  import math  from keras.models import Sequential  from keras.layers import Dense  from keras.layers import LSTM  from sklearn.preprocessing import MinMaxScaler  import numpy as np  look\_back=2 # A length of time window  epochs=1000 # training Epochs  batch\_size=20 #In minibatch the batch size is 20  #Create a time windows and extends matrix to tensor  def create\_dataset(dataset):  dataX, dataY = [],[]  global look\_back  for i in range(len(dataset) - look\_back - 1):  x= dataset[i:i+look\_back, :dataset.shape[1]-1]  dataX.append(x)  y = dataset[i:i+look\_back, dataset.shape[1]-1:dataset.shape[1]]  dataY.append(y)  dataX=np.reshape(np.array(dataX),((len(dataset) - look\_back - 1),look\_back,dataset.shape[1]-1))  dataY = np.reshape(np.array(dataY), ((len(dataset) - look\_back - 1),look\_back,1))  return dataX, dataY  #Define batch normalization class  scaler\_x = MinMaxScaler()  scaler\_y=MinMaxScaler()  scaler=MinMaxScaler()  #Read data from .csv  data\_set = read\_csv('datanoraw.csv', header=0, index\_col=0)  data\_set=data\_set.values.astype('float32')  #Batch normalization  Train\_x = scaler\_x.fit\_transform(data\_set[:,:data\_set.shape[1]-1])  Train\_y = scaler\_y.fit\_transform(data\_set[:,data\_set.shape[1]-1:data\_set.shape[1]])  data\_set=np.hstack((Train\_x,Train\_y))  train = data\_set[0:, :]  X\_train, y\_train = create\_dataset(train)  #Define model using keras  model = Sequential()  model.add(LSTM(units=4, input\_shape=(look\_back, data\_set.shape[1]-1),return\_sequences=True))  model.add(Dense(units=1))  model.compile(loss='mean\_squared\_error', optimizer='adam')  #Run the model  history=model.fit(X\_train, y\_train, epochs=epochs, batch\_size=batch\_size, verbose=2,validation\_split=0.33)  #Model visulization  model.summary()  plt.plot(history.history['loss'], label='train')  plt.legend('train')  plt.plot(history.history['val\_loss'], label='test')  plt.legend('validation')  plt.title('Loss')  plt.show()  plt.savefig('loss 1-10') |

**Data Collection**

In recent years,

* + 1. **Data Filling**

In recent years,

* 1. **Model Name**

**这里写模型的介绍，作用，具体的方程，图像等。。**

1. **Task 2**

**这里写模型的介绍，作用，具体的方程，图像等。。**

1. **Task 3**

**这里写模型的介绍，作用，具体的方程，图像等。。**

1. **Strengths and Weaknesses**
   1. **Strengths**

* In recent years,
* In recent years,
* In recent years,
  1. **Weaknesses**
* In recent years,
* In recent years,
* In recent years,

**References**

[1] COMAP. (2017). ICM Problems.zip. http://www.comap.com/undergraduate/contests/mcm/contests/2017/problems/2017\_ MCM-ICM\_Problems.zip

[2] EPA, “Smart Growth: A Guide to Developing and Implementing Greenhouse Gas Reductions Programs.” 2011. http://www.sustainablecitiesinstitute.org/Documents/SCI/Report\_Guide/Guide\_EPA\_SmartGrowth GHGReduction\_2011.pdf