**Title**

**标题：基于工程化方法的LSTM智能温控系统设计**

**Summary**

我们认为这是一个系统设计问题，从系统设计问题到数学模型，需要抽象。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。对于温度控制系统这样一个设计的问题域，我们使用用例图描述用户和温控系统的可能使用场景，然后设计了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。

1、面向对象分析——用例建模

**用例建模是面向对象分析的第一步。我们分析的用例包括智能温度系统中的用例以及外部对象，在这个用例图中的外部对象，有用户和中央处理系统两个对象。对象的交互过程是通过用例完成的，用例也就是所谓的场景。用户面对的场景是如何去设定温度程序，换句话说用户去编程控制温度。用户编制的程序应该是能够表征温度的函数，比如说可能是输入温度的函数，然后通过程序，把它翻译成MVC框架中的控制器所能识别的数据流，就可以控制屋内的温度和湿度等信息。对于节省能量、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。僵尸用户的，这个喜好呢，主要是由系统完成的，在使用用户喜好的用例中，它一定是包含设置温度程序的子用例。而对于用户的编程设定温度这个场景，我们的摄像师它一定是通过因特网的实现，那这个因特网应该是用户通过他自己的计算机，或者是他的强项是终端，比如说手机上的设备去设定温度的，但是用户编程设定温度是非常麻烦的一件事情，我们不能够让用户去，总是通过编写程序设计，因此，我们分析出用户操作系统，它更应当去设计一个图形界面，见面许，与用户交互，从而算出来这个温度的变化，这也就是我们在刚才的体系结构框架里面所说的，界面设计。下面将介绍与系统有关的几个用例，在这几个用例都能和就是系统要去感知外部的温度变化，这个温度变化主要是从几个传感器里面改制的，其中有一个是，感知的是湿度的变化，湿度传感器，还有一个要感知空气污染水平的这样一个传感器，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，它还应该能够去算出在这个室内有多少个人，根据人数去进一步的调整和努力打个比方，如果说在一个屋子里面，在一个教室里面吧，有，一百个人，那么这一百块钱的空气是非常非常的，污浊的，据说这个系统是要送来，我这个空气调节器，应该开的很大，但是如果说我在这有一个30个人的屋子里面，这个空气调节器是不用开这么大的，因为只有30个人的屋子里面的空气，相对来说是比较清新的，这也是我们说这个系统可以节能的一个原因，比如结合来说，我们的系统是一个要有一个面向用户的界面，这个界面是个图形界面，能够在图形界面上进行显示的编程，同时这个系统也能够根据用户的数据，去窗户上出来一个模型，比如说这是一种良好的学习，这样一个模型，根据用户的数据，根据人数，根据传感器发出来的各种数据去学习这一模式，也就是说我们去抽象出这样一个数学模型。**

我们利用了MVC框架，设计了一个面向用户和系统的体系结构模型。

具体解释，mvc框架：mvc框架是一个体系结构框架，这个框架由三部分组成，模型对象，视图对象和控制器对象，其中我们把温度控制器对象建模成表示，系统中心处理器的一个模型。我们做的工作就是说把这三部分融合在一起，形成了一个类图，这就是体系架构的相关内容。首先解释界面，界面是人机交互的一个窗口，当用户需要访问智能控制家居系统时，他首先需要通过手机的应用，或者说是电脑端的应用，去访问系统的界面，在下面解释中我们会看到，这个界面上应该具有可编程的特性，以便让用户去编程，控制这个系统的温度，下面解释模型的含义，这个模型我们选择的是lstm网络模型。这个模型是一个基于神经网络的深度学习框架的模型，在这个模型里面，我们应用了keras深度学习框架，来实现的这个模型，相关的代码在本论文的附录里面。在我们的控制器里面，我们选择的是一个核心控制器，也就是说在一个智能家庭系统中，由一个芯片来控制所有的任务，他来完成核心的计算功能，当然，这个芯片在，智能家居领域中有很多，比方说树莓派，和英特尔的凌动CPU，他们都可以完成相关的计算工作，因此这就奠定了深度学习的神经网络的可行性。

**那接下来就让我们从这个系统温度控制系统的用例图中，分析出了不同的用例，也就是场景，那么我们需要根据这个场景进一步的建模，都需要进一步的进行需求分析，这个时候需求分析类图，我们绘制出一张类图，那么这个类图中有几个对象呢？我们可以看到他的一个是程控的温度设定器，也就是刚才我们在建立MVC框架中所说的图形控制温度器，所以说我们说他是个可以图形控制的。但是它不仅是可以图形控制的，他可以在这个图形界面中去输入函数进行计算，所以它也是属于一种程控温度器。这个对象中有一定有个属性，这个属性，也有一个方法，这个方法的意思就是说，我们要把这个用户的输入法，用户的编程啊，这样的一些东西给它转换成移动数据，就发个远程终端，远程终端能够感知到，远程终端负责接收用户的数据，同时这个远程终端的产业负责去发送数据，把这个数据发送给我们说的核心处理器，核心处理器，它就是一台服务器，当然在物联网中设置服务器可能不是特别大型的，但是可能是一些更多的是嵌入式为主，微服务器，将比方说英特尔的凌动CPU，或者是爱迪生处理器，这样一些能够处理大规模数据的，但是考虑到在我们，建模的过程中，可能会用到神经网络，在这个系统设计时，我们应该可以考虑到这样的一个处理器呢，它一定是有一个USB神经网络加速器接口的，他这样子处理就会比较的，完全是，毕竟我们要处理的不是少数，而是大量的数据，和用户数据。就比方说，在这个旅途中，我们也就要考虑到题目中所给的一些其他限定，比方说只有一个温度调节器的情况下，很有多个，也就是额外温度控制器的情况下，怎么去对这个系统产生影响？我们在建模的时候要考虑到，这个额外的温度，控制器对象要有一个继承关系，继承于这个远程控制终端架子，能把他的一个模型来分析得更加透彻。**

**在刚才的过程中，我们工程建模方面考虑，在什么样的场景下使用，在这个场景中，在这个问题域中有哪些类？我们需要做些什么？最终把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，也就是说，接下来我们会更详细的讨论细节的数学问题。**

**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)**

对于任务中题中所给的这个任务，我们有了一个相对比较清晰的思路，这个清晰的思路呢，在我们所给的数据流图中所展现的是我们要进行数据处理，这个数据来源是马萨诸塞大学的，但是数据集是存在问题的，比方说数据缺失，牙数据造成影响，数据总共是有8000多条，在这8000多条中，我们需要先进行一定的数据处理，在预处理，希望能够把她送到一个比较标准的模型当中，这个模型选择呢，我们是这么选择的，而我们考虑到了很多模型，比方说，全连接型神经网络，还有一个卷积神经网络，当然神经网络肯定是不适合的，卷积神经网络一般是用在图像的处理里面，这个问题比较显著的是，在我们这个数据集中，我们特意标出了时间序列的数据，因为我们知道，在这个温度温度中，他的时间起到了非常大的阻，比方说我白天温度是什么样子呢？白天晚上温度是什么样子的，用户怎么挑，这个时候我们需要考虑温度和时间的一些特性，但是，我们在这个时间才能让我更加的去关注他，这时间有关时间有关的一个模型，我都想了一想，我是贝叶斯方法可能会有角色数这样一些，其他的一些弱分类器，或者说还有一种集成学习的方法，这些话都只适用于一种特定的分类或者回归，它与我们所想要的这种时间序列相关的方法是没有太大关系的，我们考虑到更多的是rnn循环神经网络，或者说是st m，叫一个，能够较清楚的去表现，时间序列的一个模型。为什么我们需要选择l tm这样一个网络，而不选择恩这样一个网络呢？这是因为，l tm赛相当大学者的验证下，它是一个具有长短时记忆效应的网络，它能够更好的去记住时间序列中之前读入的数据，这个特性是rnn网络不具备的，这个原理是什么呢？就是因为，感恩这个网络中不具有sim网络中所特有的1万个输入和输出吧，中，我们使用的一般都使用的是sigmoid函数，这样一个函数去作为遗忘门的一个控制，还是因为sigmoid函数的值域是在0到1之间，是这样子啊，就是相对来说比较好找一些，而真正的数据我们一般使用的tanh函数激活。单纯的针对这个数据来说，这个数据集大概是有8000条的数据，每个数据呢是有八个属性，这个属性包括了天气情况，还有温度，体感，温度，室内的具体情况，叫你一个属性，我们需要把这些属性，这成为用户所需的一个温度，或者我们需要设计，真正涉及的问题是怎么设计一个算法，让这个算法成为这个中央处理器，两种处理器能够处理，使得，这些属性，也就是说属性也对应着我刚才建模过程中所说到的传感器所来的数据，把这些数据综合起来，我能综合出一个用户所需的数据，所以说这个时候我们需要做一些合理的假设，我们需要假设出在每一条数据下，用户所做出的一个对温度的调节，这个时候，我们不得不再加一列数据，加点数据，这个数据就是用户所调节的温度，当然我们是用户，我们是用户，我们可以做用户的一个群体中的一员去调节温度，这个家的一个数据是根据我们自己的一个喜好去调节出来的，也就是说我们利用了，马萨诸塞大学的这个数据，再加上我们自己作为用户去调节的这样一个温度来看一下，这个数据到底具有什么样的特征，我们设计的系统啊，用这个iTunes的算法设计系统，合不合理？那么上面我就阐述了，对于第一个任务大致的看法，我们对于建模总体的一个思路，接下来我将阐述我们数据预处理以及利用专业模型进行建模的细节和算法的细节。

是这样的，我们考虑到建模的过程中会遇到非常多的问题，我们刚才所已经把工程化问题划到数学问题，这样一个工程化思想，而我们是用面向对象的方法进行转化的，所以说我们在这个时候用matlab可能就会力不从心，所以我选择的是python，python语言有非常好的工程能力，他对于工程的建模能力要比较强一些，运行环境是ubuntu。那么我介绍一下，我们用的是潘的数据处理库，因为考虑到我们的温度数据中有非常多的确实数据，当然这个缺失数据呢，主要也集中在于，三列数据上，可以查看我们对应的数据集，首先我们需要对这些数据做一个处理，看了一下这个缺失数据的基本上都是，连续性的数据，对于这个连续型的数据，嗯，之前有处理方法，有非常多的学者做过研究，连续型的数据一般是用在这一属性的平均值去填充，所以说我们也采用平均值填充法数据，涉及到一个编码问题，这个编码主要针对的是这个天气了，别买了，这是，几种天气的类型，在这种类型下，我没有办法去直接让网络去计算，所以我们需要把这个天气类型转换成一个整数类型，转换的过程中呢，我们也给出了一定的算法，比如说我们尽量把黑夜黑的排到一起，把白天、雪，这样的一些光线有关的，尽量的按这个顺序排的，这样子我们就形成了icon的编码方案。其次呢，我们需要填充的是他要系统预测的这样一个，用户调节的温度，所以我们作为用户的我们试图的去在数据中去，感受了一下这样的一个条件，我们可能会调到这个温度，并且把这个温度写上去了以后呢，我们为了加强数据的综合效果，使得数据能够推广起来，我们又在上都加了一个均值为零，当差是0.2的二的一个正态分布效果。这个正态分布用在用户的温度数据上。接下来还要解决的一个问题，也就是最关键的问题是，我们要做的是一个时间序列分析，做时间序列分析的话，我们在数据预处理的时候，必须把这个数据，按照时间来排序，而恰好这个数据集都有一个属性是时间，我们按照这个数据集中的时间属性，就这个数据集进行了预处理，使它变得有序，家来，就是我们对这个数据，预处理完的数据进行分析的成果，我们使用的数据可视化的分析方法，在python里面利用了seaborn，这样一个库对这个数据分析分析出来的结果应该是8×8等于64张图。

下面我介绍一下这个图，这个64张图的对角线上是这几组数据的分布情况，我们可以很明显的看到这些数据，都非常的，有规律，他们都集中在一个值的附近，我们把这种现象的很多都可以把它想象成一个正态分布现象，你在概率论上我们知道，无论是20项分布，还是其它的一些指数分布，当它的数据量增大的时候，它们的分布呢，都会逐渐的近似于正态分布打个比方，看风速这个数据封塑这个数据，它的分布非常想你的指数分布，那么可能这个指数分布增大了以后，应该是吧，天桥数据的数据量已经不是特别小了，那他会更多的像一个正态分布靠拢，可以看出来的，但是在有些方面不是你比方说在我们这个标，这显然就不是一个正态分布，因为它有三个峰值，那就是三个峰值。当然这个峰值的这个为什么不是正态分布，她是因为这个，标签是我们给他安排上，因为它本身是一个离散的字符串的属性，我不给他安排上的话，就不能在之后的数据处理上，进行相应的处理，接下来这个图中还有非常显著的是时间这个属性，时间，这个属性，我们可以看出来很多其它的属性都随着时间属性成周期性的变化。这个图中比较有代表性的，可能是几个线，看似比较是线性的，分布的线性分布，主要集中在体感温度和温度上，以及我们想要的温度和体感温度上，当然我们想要的温度是我们用户自己想的温度，它呈现性又非常大的偶然的因素，这跟我们几个的看法有关，因为我们在生成数据的时候，除了考虑到我们，在这样一个具体现象就是在这里用的还不是摄氏度的因素，还考虑到一些钱，所以生成的可能更多的是现行，这个并不是特别的重要。 而且我要声明的是，我们这个数据集里面所有的温度都是摄氏温度的，不是什么开尔文温度和华氏温度，我们考虑到在这个徒弟里面会有非常多的三点对吧，因为我们生成的是三点和分布图，他把分布和三点都包含在了一张大图里面，这个图，对角线上刚才已经说过了，是分布图，而对导线两侧的都是不同属性之间的散点图，可以很清楚的看到哪些属性之间有关联的那些属性之间，没有对应的关联。

接下来我们要介绍的是怎么样用ltm，这样的模型去处理，我们已经经过预处理的数据经过预处理后的数据，他并不是一个时间序列数据，但是它是按时间序列排序的，而这个时候因为我们喜欢神经网络和ltm，都有一个重要的特征，是它输入的必须是一个循环的数据，嗯，我是在ubuntu底下又帮徒弟下进行实验的，嗯，这个时候我们是，考虑的是建立了一个函数，这个函数create，他是，把我们所输的这8000多个数据呢，转换成了时间序列数据，比方说我在我们的例子总是比两个作为时间序列关联的，当然这个数量有点小，因为我们只是用来测试测试使用，在实际使用中，我们可以使用十个或更多，就说如果咱两个情况怎么可能8000多变成16000个这样个左右的数据，首先我们读入数据，然后对所有数据进行归一化，归一的目的是因为非线性的映射效果会比较好，在深度学习中之所以能够用深度学习的，就是因为它能把数据映射到某个高维空间，在这个高维空间里，让做数据线性可分，我们首先就要把归一化都要让它映射的更好，同时呢，我们可以通过反过一会儿让这个输出也更好一些，当然在这个问题，我要研究的是佛系的物产还是个模型，到底有多少误差啊，这个是我们研究的重点，并不是说具体的预测，因为我们现在并不知道用户是多少，接下来我们去用自适应方法优化这个模型，我们觉得还是一个很传统的均方误差，均方误差，原来你和效果还是在用的比较好，优化器我用的是，adam，就是说它的梯度下降是一种自适应的体重下降法，嗯，比方说那个我们用的更多的是我不用在这里面用的是动量题，就三角法，也就是自适应在路上基础上一个姿势，接下来我们在训练的过程中，我们训练的迭代次数的迭代了没有，从识字到一千次都有，嗯，比较明显的是在前十次的过程中很明显的看到这个题，这个损失有明显的下降，我们在这个时候也就到了一个正则化方法，对于这个模型，我们每一个minibatch是20个，当然这个正则化方法，那么它主要是加强验证数据集的效果，防止过拟合的出现，但是在我们最后看到这个结果的情况下的不存在过拟合的情况，他在数据在训练集合验证集上都呈现非常好的下降趋势，除了刚开始误差在验证集上升高了一些。详细情况可以看我们的，误差损失图像，那具体的过程呢，我们在程序的框架流程图里面也给出来，如果不清楚的话，你可以亲自去运行这个程序，这程序是在linux上跑的。哦，那我们用这个模型是要打造一个效果，就是我们所生存的这个时间序列大概是16000个左右，那在这个时间序列的情况下，我们能够准确的较为准确的预测出用户的输入是什么？进而使我们在鄂中央处理器上利用这个深度学习的模型，能够快速的计算出用户的输入。那比较细节的方面，我们是怎么把这8000个数据转换成16000个数据呢？这个中间我们还是用的非常传统的方法，滑动窗口法，每个窗口的定义，树木12，比如说我们，每次把这窗口下移一个数据单位，那这个总共能每两个一组，总共差不多就能够生成8000个数据框左右，那么这8000个数据框接着去拓展，总共能生成的是16000个数据框呢，这个时候我们就把一个二维的数据集变成了一个三维的张亮，把这张亮为入到，嗯，我们所设计的这个网络里面，它输出的肯定是也是一个张亮吧，我们把这个张亮也进行了，一定的，你和将来把这个张靓颖是哪一个误差，周五查档案均方误差吧，就比较常见了，我就不再去解释了，映射函数也都是比较常见的函数啊。猜个这个程序的最后，我们也接触在训练过程中的一些数据，去输出了这样一个图形，表示的是损失的图形，在这里面我们只给出比较明显的图形，就是一到狮子的图形，因为1到10次已经下降得相对来说比较好了，这说明这个模型相对来说在时间序列上还是比较成功的，

那么这个模型具体有什么用？他就是主要在我们在前面所说的MC框架中，这个框架中的那样一个，模型处理器，这个模型主要是用来计算的，也就是说在我们的，面向对象的类图里面，它对应的是那个中央处理器，它负责接收各传感器的数据，并且输出用户想要的数据，去调节各个温度调节器啊，这样一个数据。

**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