

# 《Web 信息处理与应用》课程实验报告

## Attention-based Convolutional Neural Networks

### 一、实验介绍

#### 1.1 实验内容

本实验利用基于注意力机制的卷积神经网络对用户的体检指标序列建模，预测用户下一年体检的收缩压，同时比较了几种不同的注意力机制所得到的预测结果。

#### 1.2 数据集描述

原数据是130多万条体检记录集合，每条体检记录有1091个属性，本实验首先基于规则将属于一个用户的体检记录根据体检时间从远到近链接起来最终形式用户的体检记录序列，然后筛去确实比例大于0.3的属性，接着针对我们要处理的问题（预测用户下一年体检的收缩压），本实验基于一些先验知识以及机器学习方法（lasso、随机森林）筛选出以下25个体检指标特征：

15#一般检查#收缩压，

15#一般检查#舒张压，

1#血常规#白细胞计数（WBC），

1#血常规#血红蛋白含量（Hb），

1#血常规#红细胞计数（RBC），

1#血常规#红细胞比积（HCT），

1#血常规#平均红细胞体积（MCV），

1#血常规#平均红细胞血红蛋白含量（MCH），

1#血常规#平均红细胞血红蛋白浓度（MCHC），

1#血常规#血小板计数（PLT），

1#血常规#平均血小板体积（MPV） ,  
1#血常规#血小板分布宽度（PDW） ,  
1#尿常规#尿葡萄糖（GLU） ,  
1#尿常规#尿胆原（URO） ,  
1#尿常规#尿酸碱度（PH） ,  
1#尿常规#尿比重（SG） ,  
1#尿常规#尿蛋白质（PRO） ,  
1#血脂检测#甘油三酯（TG） ,  
1#血脂检测#总胆固醇（TC） ,  
1#血脂检测#高密度脂蛋白胆固醇（HDL-C） ,  
1#血脂检测#低密度脂蛋白胆固醇（LDL-C） ,  
1#肾功能检测#肌酐（Cr） ,  
1#肾功能检测#尿酸（UA） ,  
1#糖尿病检测#空腹血糖（FPG） ,

体重指数

）和4个context特征：性别、省份（北京、海南）、年龄，接着筛选出体检次数大于等于3次的用户，最终本实验得到9011个用户的体检记录序列，随机划分出7208个用户体检记录序列作为训练集，1803个体检记录序列作为测试集。

### 1.3 实验环境

软硬件配置：Ubuntu 4.8.2、GTX 1080 8G、Python3

深度学习工具：CUDA Version 8.0.61、Tensorflow 1.4.1(GPU版)、Keras 2.0.6

## 二、实验知识点

### 1.1 卷积神经网络

卷积神经网络的概念最早起源于科学家提出感受野 (Receptive Field)，即每个动物的神经元只会处理一小块区域的视觉图像，相当于CNN中卷积核的处理过程。后来又提出了神经认知机的概念，神经认知机包含两类神经元，一类是用来提取特征的S-cell，对应于现在的CNN中卷积核的滤波操作；一类是用来抗变得C-cell，对应于现在的CNN中激励函数、池化等操作。一般的CNN的基本结构由输入层、卷积层、取样层、全连接层及输出层构成。每层卷积层会进行一下几个操作：通过不同的卷积核滤波，加偏置bias，提取局部特征；滤波输出结果进行激活函数处理，常用ReLU函数；激活函数输出进行最大池化或平均池化操作。

卷积神经网络不仅在语音识别、图像识别及分类领域有了广泛且成功的应用，其在自然语言处理领域也被广泛且成功地应用，受此启发，本实验采用卷积神经网络对用户的体检序列建模。

### 1.2 注意力机制

注意力机制可以使得神经网络具备专注于其输入（或特征）子集的能力：选择特定的输入。这可以是  $x \in \mathbb{R}^d$  一个输入， $z \in \mathbb{R}^k$  一个特征向量， $a \in [0, 1]^k$  一个注意力向量或  $f \phi(x)$  注意力网络， $a=f \phi(x)$ ， $za=a \odot z$ ，其中  $\odot$  是指元素对应乘法 (element-wise multiplication)。它将特征与一个值在 0 和 1 之间的掩码或当这些值被限定为 0 或 1 时的硬注意力 (hard attention) 相乘，即  $a \in \{0, 1\}^k$ 。神经网络相当于函数近似器。它的能够近似不同类别函数的能力主要依赖于它的架构。一个典型的神经网络可以被实现为一系列矩阵乘 (matrix multiplications) 和元素对应非线性乘法 (element-wise nonlinearities) 其中输入或特征向量的元素仅仅通过加法相互作用。注意力机制会对一个用于与特征相乘的掩码后进行计算，一个可以通过神经网络进行很好的近似的函数空间得到了极大的扩展，使得全新的用例成为可能。这种理论认为神经网络是一个通用函数近似器，可以对任意函数进行近似为任意精度，但只能在无限数量的隐藏单位限定条件下进行。在任何实际情况下，我们受限于可以使用的隐藏单位的数量。当我们要对N个输入的结果进行近似，前馈神经网络只能通过模拟具有许多加法（以及非线性）的乘法来实现，因此需要大量神经网络的实际空间。但如果我们引入乘法交互的理念，过程就会自然而然地变得简单而又便捷。

### 1.3 基于注意力机制的卷积神经网络

本实验在卷积神经网络中引入注意力机制，相当于对输入先进行再赋权重，根据各部分和结果的密切联系程度赋予不同的权重，再经过卷积神经网络得到最终结果，这种方法符合任务的内在结构，使其高效地得到一个比较好的结果。

### 三、实验过程

#### 3.1数据准备

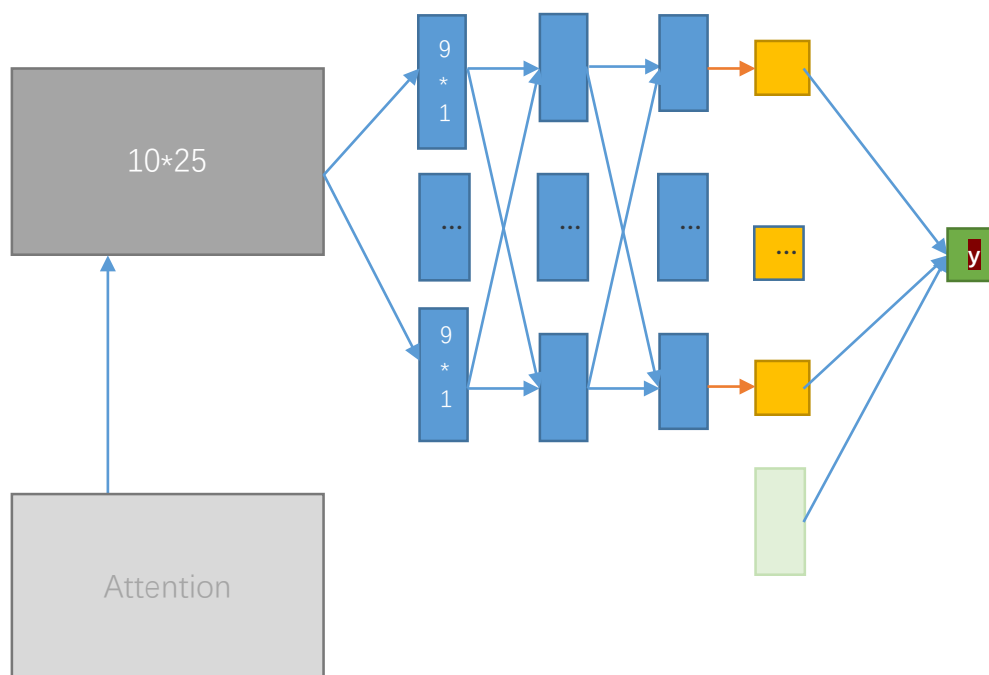
从TJ\_180403. json读入所有用户的体检序列TJ\_dict;

分别从train\_3\_bph\_180517\_2.8.txt和test\_3\_bph\_180517\_2.8.txt读入已经筛选出的训练集和测试集的用户ID;

把以上得到的数据放入produce\_single\_var\_data .py中的produce\_single\_var\_data函数，得到各项特征被归一化的训练集、测试集以及其对应的下一年的收缩压x\_train, y\_train, x\_test, y\_test, train\_context, test\_context。

#### 3.2构建模型结构

本实验利用且比较卷积神经网络以及4种基于注意力机制的卷积神经网络对用户的体检指标序列建模，预测用户下一年体检的收缩压的效果，其模型结构如下：



attention模块设计了以下四种实现：

- 以年份为整体赋予注意力 (Cnn+attention of years)

把每一年的各个体检指标组合形成特征向量放入前馈神经网络得到权重，把权重乘以相应的年份的所有指标的值。

- 以体检指标为整体赋予注意力 (Cnn+attention of indices)

把每一个体检指标在所有年份的值组合成特征向量放入前馈神经网络得到权重，把权重乘以相应的指标的在每一年的值。

- 同时考虑年份、体检指标赋予注意力

➤ Cnn+ summing attention of indices and years

对以上两种方法得到的权重加和之后得到对每个元素新的权重

➤ Cnn+ hierarchical attention of indices and years

先得到对指标的权重，使其对输入进行加权调整，再得到对年份的权重，把权重乘以相应的年份的所有指标的值。

### 3.3 训练模型

实验设置：实验使用 Keras 编写神经网络代码，使用 tensorflow 作为 backend。batch\_size = 128，每轮迭代随机 shuffle 训练数据顺序，一共有100个 epoch。损失函数为 mse，优化方法为:adam，除了attention层后面的激活函数为 softmax，其它地方的激活函数均为 ReLu。

## 四、实验结果

实验项目	mae	mse
Cnn	8.9018	130.0172
Cnn+attention of years	8.9571	129.5476
Cnn+attention of indices	8.9360	128.1652

Cnn+ hierarchical attention of indices and years	8.8846	129.2386
Cnn+ summing attention of indices and years	8.9006	128.3927

以上实验结果说明在这个问题上在传统的卷积神经网络中引入注意力机制可以有效地提高预测结果的准确性，对指标施加注意力比对年份施加注意力更有效。

## 五、参考链接

TensorFlow 英文官方网站

Keras 英文文档