# 实验记录

## 1时间：20190605，23：00-？

## 2实验目的：

随机选取网络中的几个参数进行修改和调整，根据结果查看网络调整对实验结果的影响。

## 3实验代码存放文件夹：

20190605文件夹中。

## 4实验运行时间：

1天及以上。

## 5实验使用数据信息：

药物作用位点预测，可能存在的药物作用位点标志为1，正样本。其余其他位点均为负样本。消除位于边界的负样本，包含正样本位点的负样本。

## 6.1，数据集正负样本设置，比例设置，

正负样本1：33。采用十则交叉验证模式构建10个数据集的数据，依次进行学习。目前没完成整体评估+存取模型+选取最优模型等步骤

## 6.2，网络设置：

def zyh\_CNN(trainX, trainY,valX=None, valY=None, compiletimes = 0, forkinas=False,transferlayer=1,compilemodels=None,earlystop=None,nb\_epoch=5,batch\_size = 8192,fildername = None ):

"""

:argument

:return:

"""

input\_row = trainX.shape[2]

input\_col = trainX.shape[3]

trainX\_t = trainX;

valX\_t = valX;

#print(input\_row)

#print(input\_col)

trainX\_t.shape = (trainX\_t.shape[0], input\_row, input\_col)

if (valX is not None):

valX\_t.shape = (valX\_t.shape[0], input\_row, input\_col)

#print(valX\_t)

if (earlystop is not None):

early\_stopping = EarlyStopping(monitor='val\_loss', patience=earlystop)

nb\_epoch = 1000; # set to a very big value since earlystop used

if compiletimes == 0:

input = Input(shape=(input\_row, input\_col))

filter1 = 64

filtersize1 = 2

dropout1 = 0.25

L1CNN = 0

nb\_classes = 2

batch\_size = batch\_size

actfun = "relu";

nadam = Nadam(lr=0.00001)

optimization = nadam

dense\_size1 = 128

dense\_size2 = 64

dense\_size3 = 8

dropout\_dense1 = 0.298224

dropout\_dense2 = 0

dropout\_dense3 = 0

input = Input(shape=(input\_row, input\_col))

x = conv.Conv1D(filter1, filtersize1, init='glorot\_normal', W\_regularizer=regularizers.l2(L1CNN),

border\_mode="same")(input)

x = Dropout(dropout1)(x)

x = Activation(actfun)(x)

x = core.Flatten()(x)

output = x

output = Dropout(dropout1)(output)

output = Dense(dense\_size1, init='glorot\_normal', activation='relu')(output)

output = Dropout(dropout\_dense1)(output)

output = Dense(dense\_size2, activation="relu", init='glorot\_normal')(output)

output = Dropout(dropout\_dense2)(output)

output = Dense(dense\_size3, activation="relu", init='glorot\_normal')(output)

output = Dropout(dropout\_dense3)(output)

out = Dense(nb\_classes, init='glorot\_normal', activation='softmax')(output)

cnn = Model(input, out)

cnn.compile(loss='binary\_crossentropy', optimizer=optimization, metrics=['accuracy'])

pass

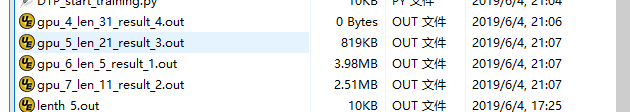
else:

print("use old net")

cnn = compilemodels

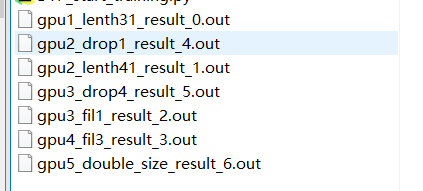
pass7实验变量：

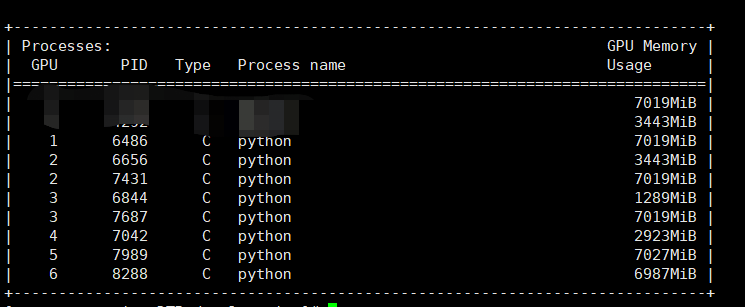
## 8实验运行记录



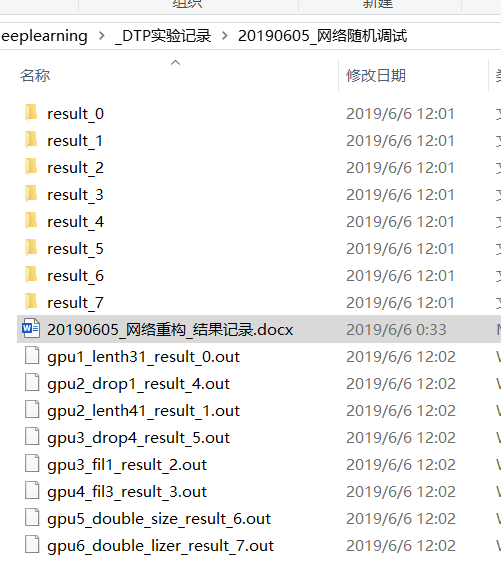
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 开始时间 | 输出文件 | 结束时间 | GPU编号（数量） | 运行id | 长度 | 文件夹 | Filtersize(kernalsize)  卷积核改变 | Densize改变 | 卷积层drop out改变 | 描述 |
| 1 | 23：00 | 如图 |  | 1 | 6486 | 31 | 0 | 2 | 128：64：8 | 0.25 |  |
| 2 | 23：00 | 如图 |  | 2 | 6656 | **41** | 1 | 2 |  |  |  |
| 3 | 23：00 | 如图 |  | 3 | 6844 | 31 | 2 | **1** |  |  |  |
| 4 | 23：00 | 如图 |  | 4 | 7042 | 31 | 3 | **3** |  |  |  |
| 5 | 23：00 | 如图 |  | 2 | 7431 | 31 | 4 | 2 |  | 0.1 |  |
| 6 | 23：00 | 如图 |  | 2 | 7687 | 31 | 5 | 2 |  | 0.4 |  |
| 7 | 23：00 | 如图 |  |  | 7989 | 31 | 6 | 2 | 256：124：16  （数量加倍） |  |  |
| 8 | 23：00 | 如图 |  |  | 8828 | 31 | 7 | 2 | 128：128：64：64：8：8（层数加倍） |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

以后根据预期主要变量由小到大设置数值，GPU编号，名字，文件夹都要由小到大设置。



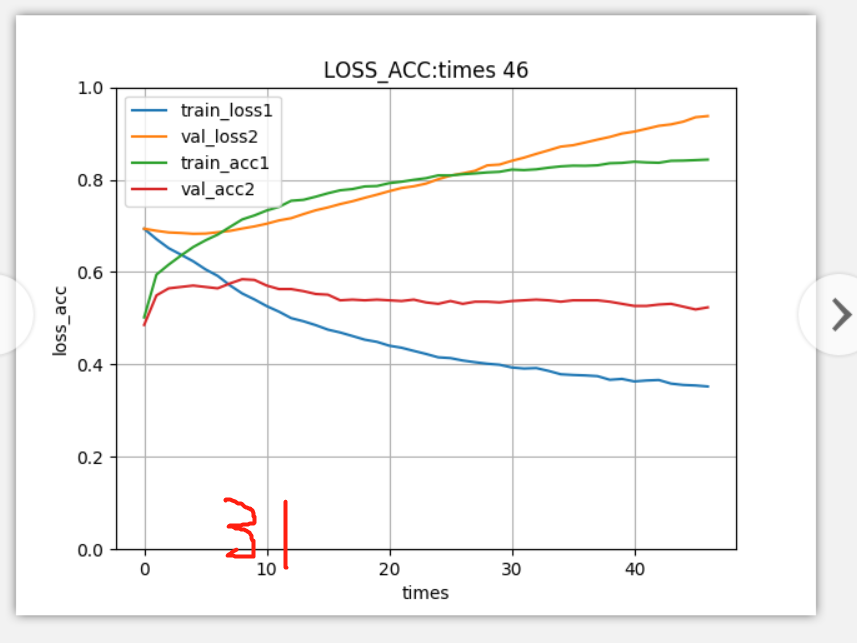


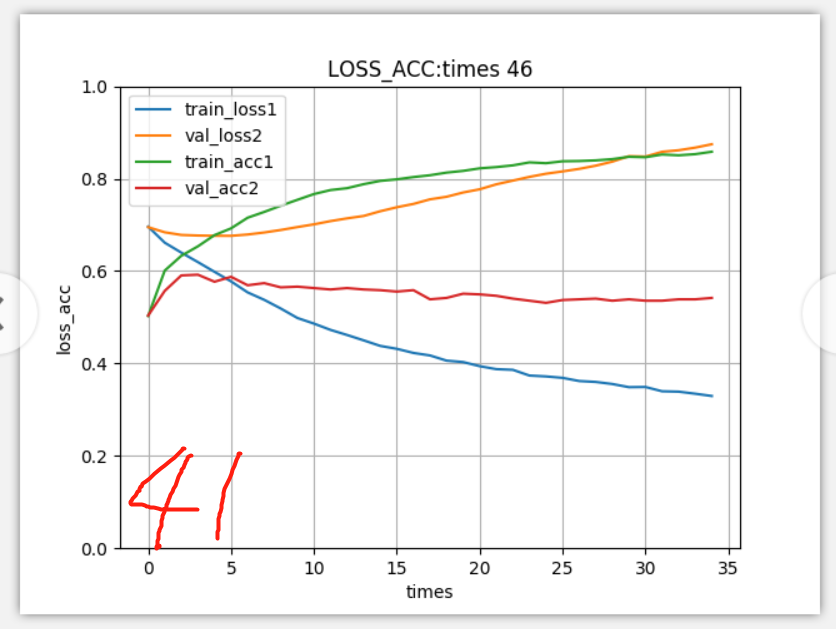
## 9实验结果存放位置



## 10结果分析

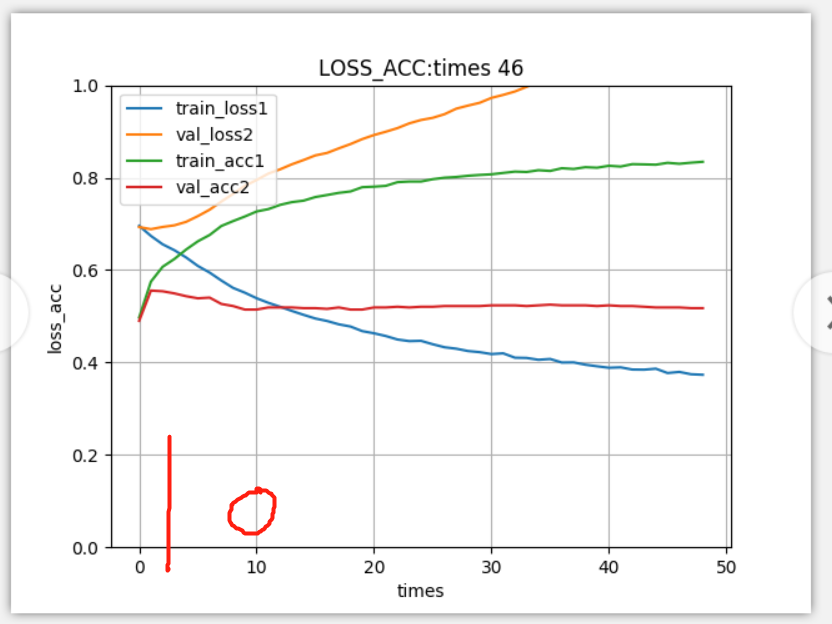
【第一组】窗口长度31、长度41

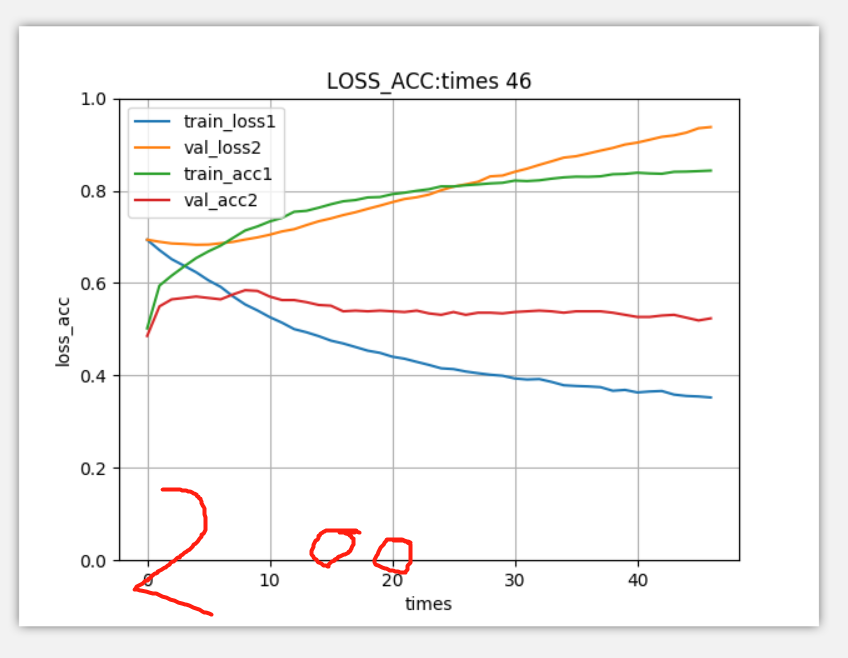


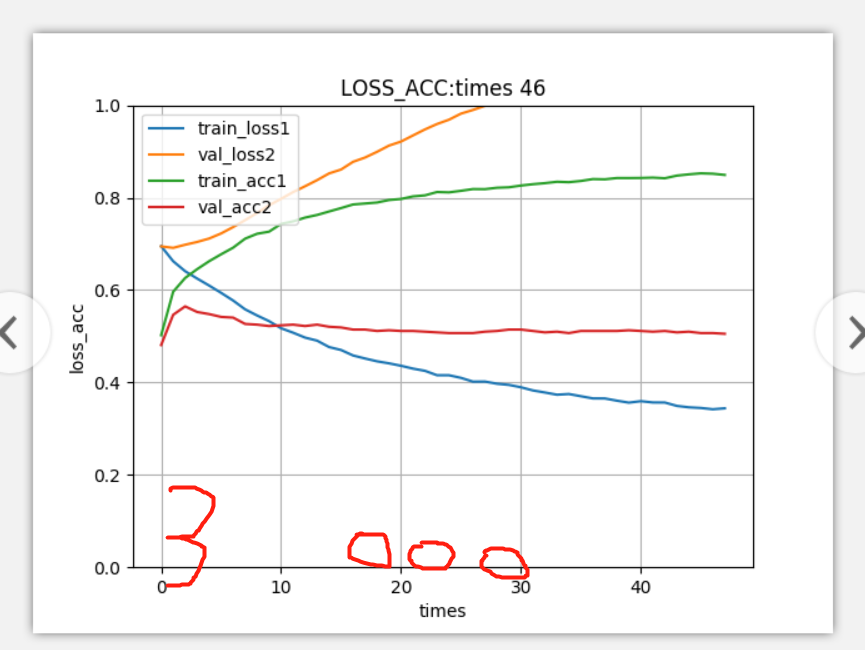


猛然一看似乎41比31loss曲线的趋势要好，其实是差不多的：长度41的因为训练慢，所以同样时间才训练了35次。所以可以看出，当前长队对训练差距不大，所以计划之后使用长度31的做训练。

【卷积核，1-2-3】

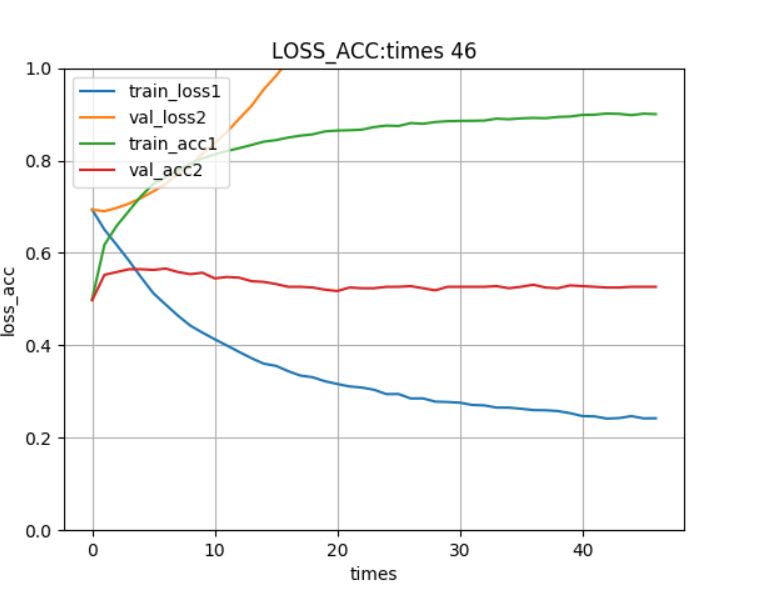


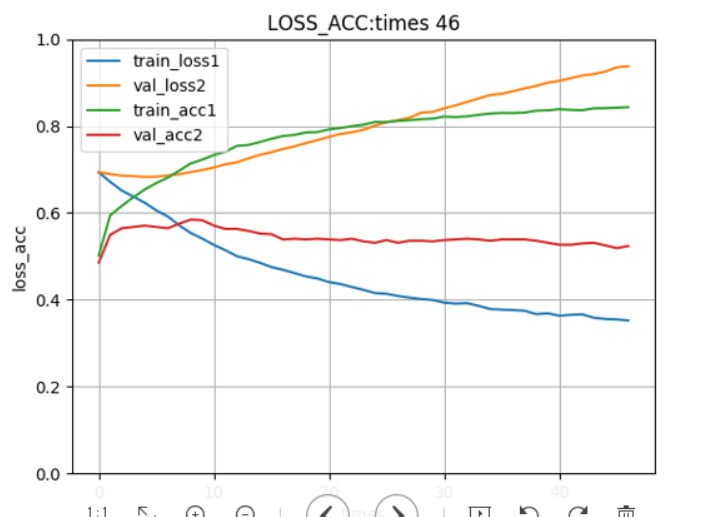


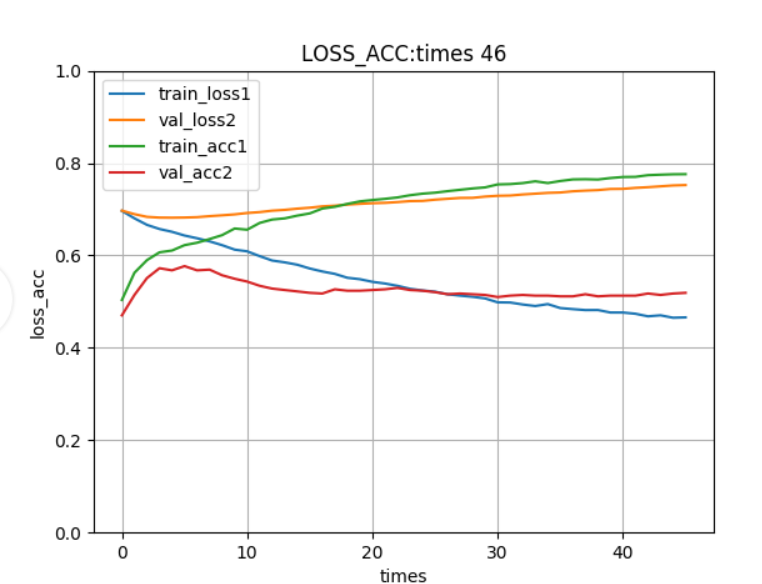


【卷积核】设置为2看起来稍微好一点，理由val\_acc高一点，loss没有另外两个差。但其实是五十步笑百步，没啥用。

【drop out】0.1-0.25-0.4

【0.1】

【0.25】

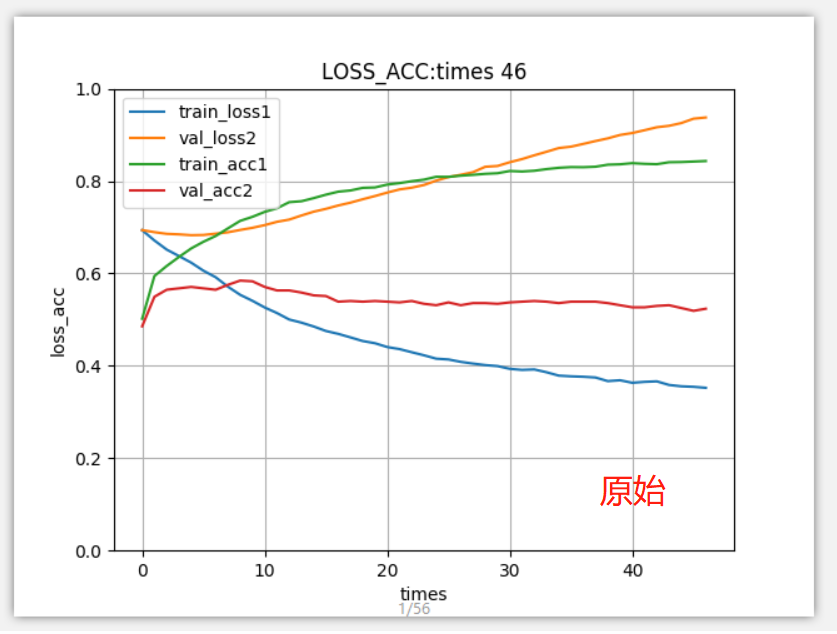
【0.4】

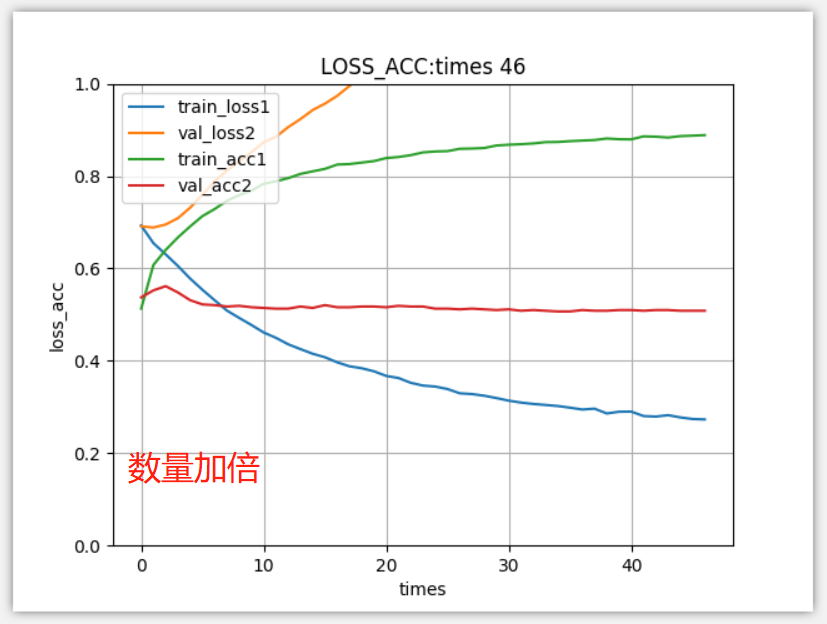
随着drop out增加，同样训练批次结果可以看到，train训练集效果变差。我这么理解，损失掉一部分神经元的连接，让模型学习效率降低。当然，调整drop out的目的是消除过拟合，

人家网上介绍的调整经验中，loss都是下降或者水平。。。咱们的是上扬

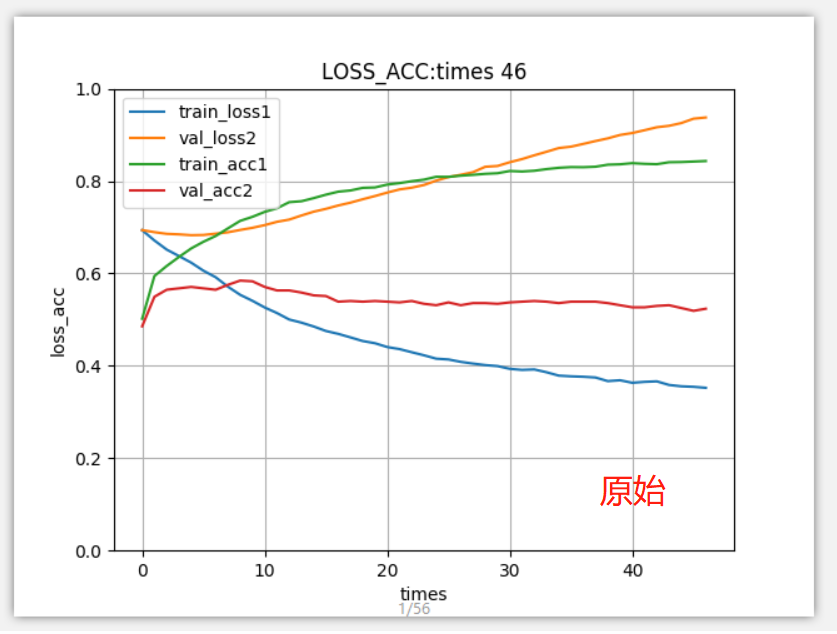


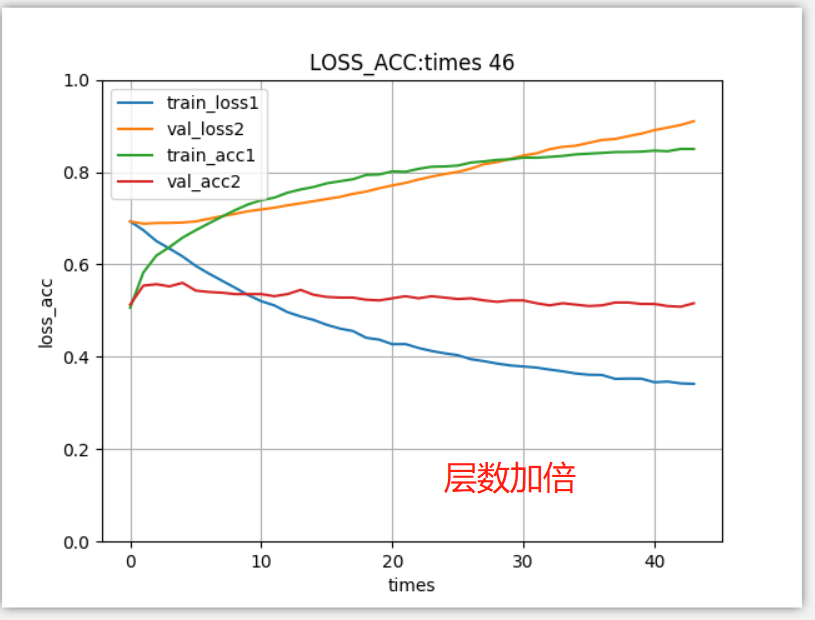
最后一组实验：网络哦结构调整。也没啥太好的效果。



结点数量加倍训练集获得了更好的效果！！！但是验证集效果更差。。。（假如减少结点，验证集和训练集的acc曲线将会结合，或相互拟合？）

【这部分需要做一下测试！！！！】（短时间内先搞数据后测试吧）但是如果拟合后，acc值下降了啊！！！





多了几个全连接层对网络没有太多影响。。。这个结论很简单。

另外，auc,roc,精准度，等数据惨不忍睹，就没有列出来。

## 11总结

1网络随意尝试效率太低。下次调整网络，计划找一个效果好的。或者根据先验知识调整。

2数据是关键，今天发现数据问题比较严重。也很心疼自己的网络：先告诉它这个是正样本，一会儿又告诉它这个这个是负样本。。。假阴性样本多？