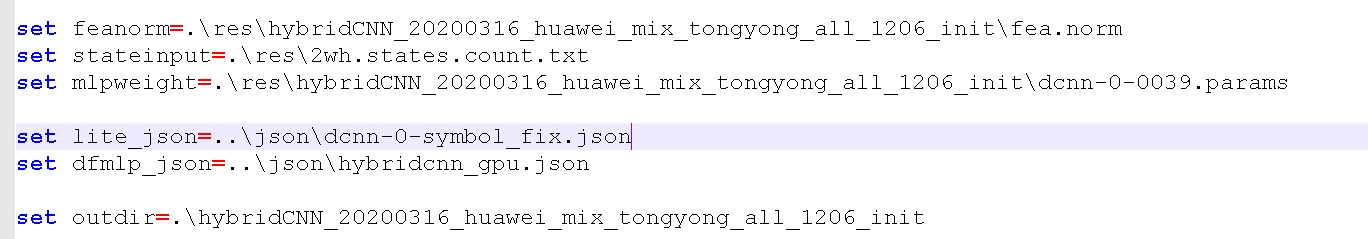
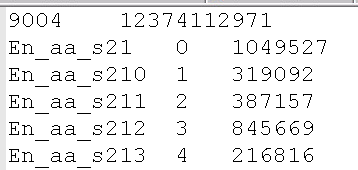
Hybridcnn打包工具说明

**1、总体运行脚本说明**

总体运行脚本的一般取名为hyper\_cnn.bat脚本，里面主要需要修改的包括以下内容：



1. feanorm 为输入特征的统计量，需要根据自己的训练数据进行统计适配；
2. stateinput 为状态文件，第一行状态的个数和状态个数总和(用于计算先验);比如英语的文件形式如下：



1. mlpweight为mlp的参数，即模型最终保存前向计算的权重，与lite\_json共同完成hybridcnn的前向计算。
2. lite\_json 为hybridcnn网络结构图的json文件，用于Maxengine的前向计算，如果网络中的卷积等OP的参数调整需要修改该文件进行适配；
3. dfmlp\_json 为dfmlp的json文件，主要包括控制整个计算流的相关参数，包括输入帧数、前后视野等参数；
4. outdir 为最终打包资源的输出文件。

**2、dfmlp json文件说明**

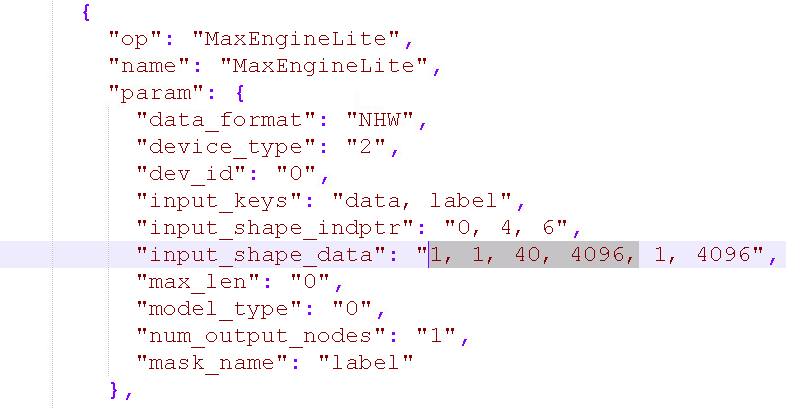
由于hybrid cnn的引擎的由dfmlp+maxengine实现，其中dfmlp完成数据流模块、maxengine完成前向计算模块，因此存在dfmlp json和maxengine json两个json文件。

Dfmlp的json主要用于数据流的控制，主要需要注意的配置如下：



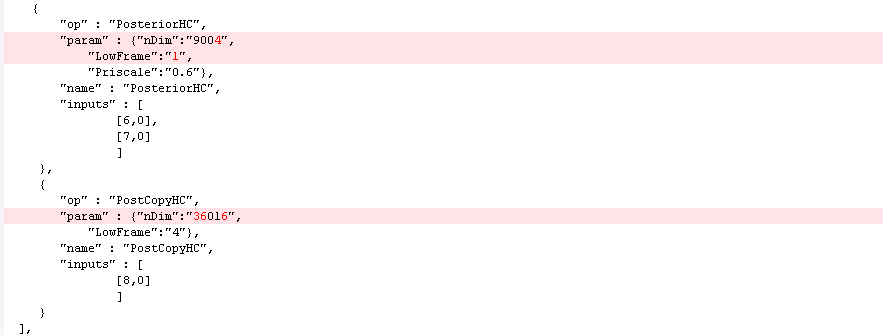
1. nDim:为输入特征维度，此处为40维特征；
2. nInputFrame：为输入的帧数，此处为4帧进模型；
3. nBufferLength:为dfmlp最多缓存的输入数据的长度；
4. nHistoryPad：为历史感受野的长度，即向前看的帧数,该值根据hybridcnn的模型具体计算得出，现有模型都为96；
5. nDataFlushWindowLength：为一次flush下发的长度，即数据满足400帧则进行一次前向计算；
6. nFuturePad：为未来感受野的长度，即向后看的帧数,该值根据hybridcnn的模型具体计算得出，现有模型都为96；
7. nMaxSingleBatchLength:因为该数据流采用单batch的方式，因此需要进行拼成单batch计算，该值设置单batch最大的数据长度；
8. nOutputDimNum:为输出数据的维度；
9. 其他配置一般采用默认值，不需要修改。

因为Dfmlp中，调用Maxengine进行前向计算，下图为Dfmlp调用Maxengine的配置：

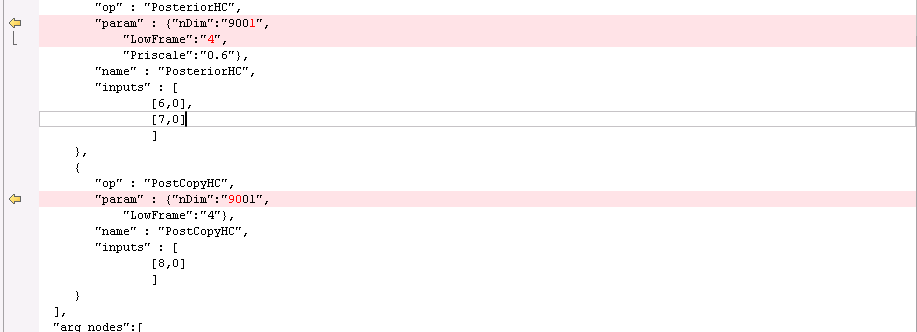


1. data\_format:为输入数据的格式，此处为NHW(batch\*H\*W)；
2. device\_type:用于设置计算平台，2为GPU平台、0为CPU平台；
3. dev\_id:为计算平台的设备卡号，此处0表示0号GPU；
4. input\_keys:为hybridcnn的输入数据的名字，与hybridcnn网络结构的json的输入相对应；
5. input\_shape\_indptr:为输入数据对应的维度信息，比如此处data为4位，label为2维(6 - 4)；
6. input\_shape\_data:为输入数据的具体维度信息，比如此处data的4维信息为（1\*1\*40\*4096），label为（1\*4096）；
7. model\_type:为输入数据的数据类型，一般取0为float32类型；
8. num\_output\_nodes:为输出数据的个数，即网络json最终的输出；
9. mask\_name:mask信息对应的参数名字，此处为label；

同时Dfmlp通过PosteriorHC、PostCopyHC两个OP用于控制最终的输出的维度，也是控制模型高低帧率的配置：



上图配置为高帧率配置，4帧进4帧出，输出36016（4\*9004），即输出4帧，每帧输出9004个状态。

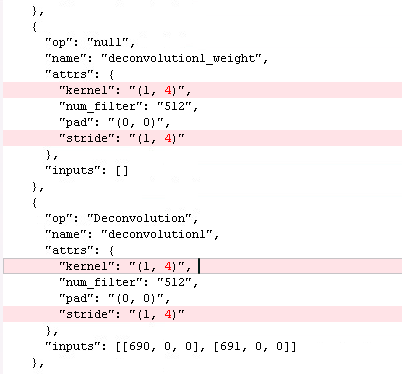


上图配置为低帧率配置，4帧进1帧出，输出9001（1\*9001），即输出1帧，每帧输出9001个状态。

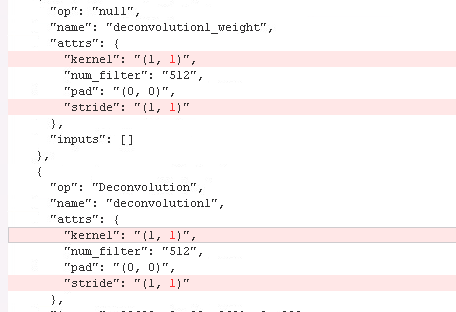
**3、hybridcnn网络结构json**

该json为hybridcnn的网络结构json，与模型的卷积个数、核大小、attention等相关，因此模型具体结构变化时需要进行适配。

目前，主要小语种中修改的json文件位置如下表，用于适配高低帧率模型：



上图配置用于高帧率模型的反卷积配置，卷积核为1\*4；



上图配置用于低帧率模型的反卷积配置，卷积核为1\*1；

同时需要根据自己的模型状态数设置下面OP中的num\_filter的个数。

