ctr和cvr分开建模，相比ctcvr直接建模的优势：电商场景的成交正样本是非常稀疏的，而负样本中又充斥着大量的easy负样本（未点击）。大量的easy负样本会影响模型的训练，具体可以参考[kaiming](https://www.zhihu.com/search?q=kaiming&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7B%22sourceType%22%3A%22answer%22%2C%22sourceId%22%3A2751181049%7D)的focal loss论文。将ctr和[cvr](https://www.zhihu.com/search?q=cvr&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7B%22sourceType%22%3A%22answer%22%2C%22sourceId%22%3A2751181049%7D" \t "_blank)分开[建模](https://www.zhihu.com/search?q=%E5%BB%BA%E6%A8%A1&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7B%22sourceType%22%3A%22answer%22%2C%22sourceId%22%3A2751181049%7D" \t "_blank)的好处就显而易见了：相对[ctcvr](https://www.zhihu.com/search?q=ctcvr&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7B%22sourceType%22%3A%22answer%22%2C%22sourceId%22%3A2751181049%7D)，ctr任务的正样本更丰富，可以让embedding等训练的更充分，cvr任务的负样本过滤掉了easy负样本（未点击）让模型训练更容易。

LSTM的神经网络输入输出究竟是怎样的

图示

描述已自动生成

可以看到中间的 cell 里面有四个黄色小框，你如果理解了那个代表的含义一切就明白了，每一个小黄框代表一个[前馈网络层](https://www.zhihu.com/search?q=%E5%89%8D%E9%A6%88%E7%BD%91%E7%BB%9C%E5%B1%82&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7B%22sourceType%22%3A%22answer%22%2C%22sourceId%22%3A%22309529532%22%7D)，对，就是经典的神经网络的结构，num\_units就是这个层的隐藏神经元个数，就这么简单。其中1、2、4的激活函数是 sigmoid，第三个的[激活函数](https://www.zhihu.com/search?q=%E6%BF%80%E6%B4%BB%E5%87%BD%E6%95%B0&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7B%22sourceType%22%3A%22answer%22%2C%22sourceId%22%3A%22309529532%22%7D)是 tanh。这个隐藏层其实就是cell中的前馈神经网络，他的参数就是可以调的，就是num\_units。他的调整会影响后面的output的输出size。Lstm可以理解为前馈神经网络外面包了一个cell外壳，多个cell组成了lstm。

另外需要注意的地方：

1. 上一次的状态 h(t-1)是怎么和下一次的输入 x(t) 结合（concat）起来的，这也是很多资料没有明白讲的地方，也很简单，concat， 直白的说就是把二者直接拼起来，比如 x是28位的向量，h(t-1)是128位的，那么拼起来就是156位的向量。
2. cell 的权重是共享的，这是什么意思呢？这是指这张图片上有三个绿色的大框，代表三个 cell 对吧，但是实际上，它只是代表了一个 cell 在不同时序时候的状态，所有的数据只会通过一个 cell，然后不断更新它的权重。
3. 那么一层的 LSTM 的参数有多少个？根据第 3 点的说明，我们知道参数的数量是由 cell 的数量决定的，这里只有一个 cell，所以参数的数量就是这个 cell 里面用到的参数个数。假设 num\_units 是128，输入是28位的，那么根据上面的第 2 点，可以得到，四个小黄框的参数一共有 （128+28）\*（128\*4），也就是156 \* 512，可以看看 TensorFlow 的最简单的 LSTM 的案例，中间层的参数就是这样，不过还要加上输出的时候的激活函数的参数，假设是10个类的话，就是128\*10的 W 参数和10个[bias](https://www.zhihu.com/search?q=bias&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7B%22sourceType%22%3A%22answer%22%2C%22sourceId%22%3A%22309529532%22%7D) 参数
4. cell 最上面的一条线的状态即 s(t) 代表了长时记忆，而下面的 h(t)则代表了工作记忆或短时记忆