3 时空循环网络

人类的行为可以描述为身体部位随着时间的移动。

LSTM网络将门机制应用到中间记忆单元以学习和表示一个更好的更复杂的输入数据中的长期依赖表示，因此它们适合对序列数据的特征学习。

3.1 LSTM的时序模型

输入门确定了“调制输入信息应该更新记忆细胞”的范围。

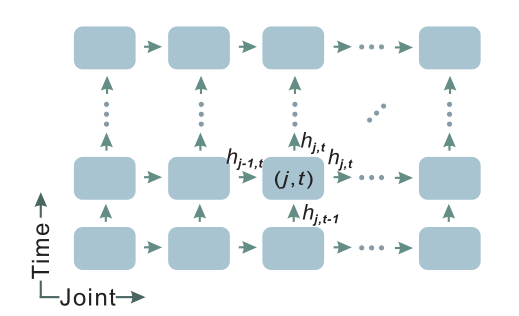
遗忘门确定了记忆单元的先前状态对其当前状态的有效性。

输出门控制了从记忆单元输出的信息量。

3.2 时空LSTM

已经有很多工作**集中在发掘复杂动态模型中的区别信息（discriminative information）**，同时证明了RNN在对人类行为在时序中的复杂动态模型的杰出性能和能力。

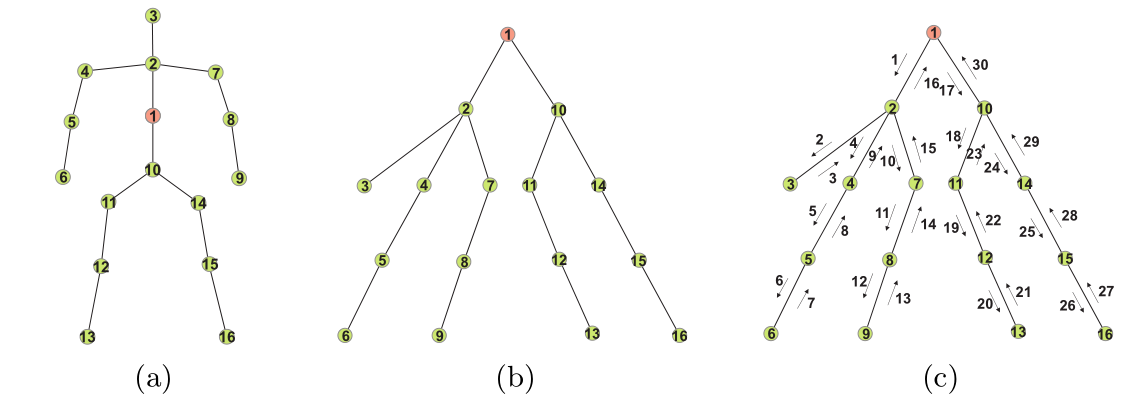
然而，单帧的静态姿态中同样包含了区别性的信息，同时骨骼数据的序列特征使得RNN类的模型能够应用到空域中。不同于其他的将节点信息合并的方式，作者通过将RNN中循环分析方法应用到空域中以发掘节点之间的空域依赖模式。



每个STLSTM神经元对应了每一个骨骼节点，同时**每个神经元获取到先前一个节点的神经元得到的隐藏表示**以及**来自上一帧中同一节点的隐藏表示。**

首先我们假定节点以一个链条序列的顺序进行排列，同时被以该顺序组织网络模型中的STLSTM神经元。之后再引入树结构的顺序序列。

**简易链的缺点**：如果仅仅是按照一个简易链的方式进行排序，相应模型会忽略掉人体节点的真实的互相连接关系，不仅如此，还添加了节点之间未曾真实存在或者连接强度不足的连接关系。



对人类结构的分析中，骨骼节点普遍被建模成基于树的图形结构。以此建模方式，会发掘2号节点即颈部应该包含更多的隐藏表示。

基于树的图形结构比简易链方式更加适合描述节点间的关系，同样更有利于模型学习到适合真实情况的节点间相关影响的信息。但是问题在于树关系不能够直接喂入STLSTM的框架中。作者提出了一种新的方式来缓解该问题，使树结构适配LSTM模型。即**对节点双向遍历得到新的节点序列，以蕴含骨骼的树结构的相邻信息**。如图（c）所示。

值得注意的是，节点的遍历遵循深度优先的遍历方式，即当遍历到节点时，遍历方向将返回。如此，**树结构中节点的连接信息将会被遍历两次，同时上下文信息将会从两个方向送入模型**，也就使得该方式能够获取到更加强烈的、长时间的空间依赖模式（stronger long-term spatial dependency）。

另外，由于每个stlstm神经元接受的数据只是单个的节点（3，），因此比其他的合并所有人体节点特征的模型，具有更加小量的参数规模，同时在训练样本规模受限时，模型具有更好的泛化能力。

与LSTM模型相似的是，模型的表示能力可以通过叠加网络层的方式得以提升，同时构建一个深度但完全可处理的网络（constructing a deep yet completely tractable network）。如图3所示。

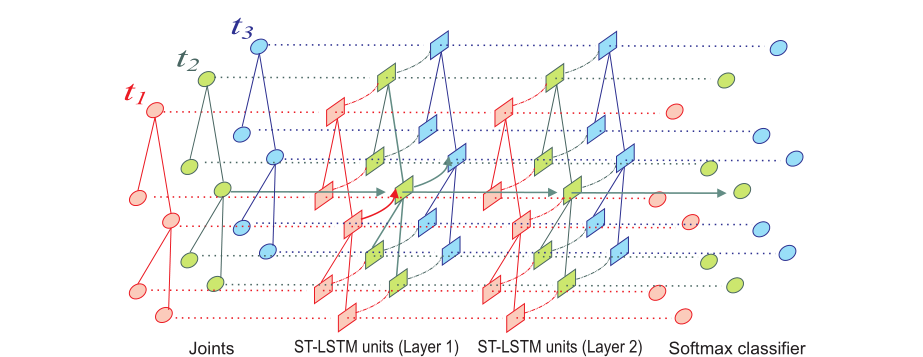


图3

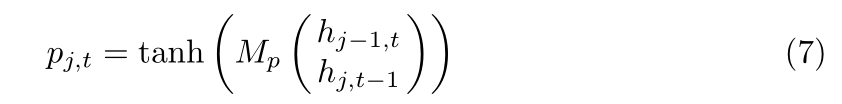
3.4 拥有信任门的STLSTM模型

目前收集到的位置坐标信息是通过某个设备实现的，但是该设备因为噪声和遮挡的原因而不总是能得到真实可信的数据。这一点就限制了模型的性能。

为了解决该问题，作者提出了信任门，改造LSTM神经元。它能够**基于对可用的上下文信息的输入的估计，**以分析每次输入的可信度。

信任门的想法是**受到了自然语言处理的一些工作的启发。**它可以告诉长期记忆机制知道什么时候以及怎样记住与遗忘**记忆单元中的内容**。

比如，如果信任门发觉当前节点的3D值有问题，它可以关闭输入门，以阻止记忆单元根据当前不可信的输入更新内容。

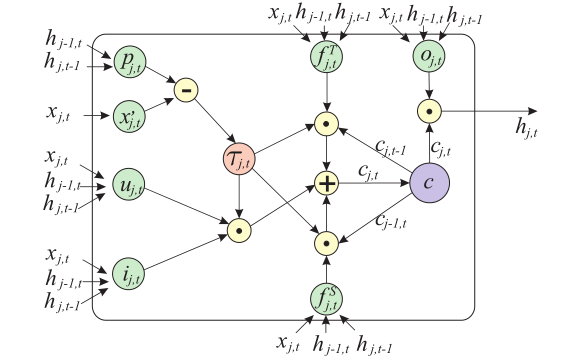


从数学的角度讲，对于每一步的输入内容，作者做了一个函数来生成预测值。值得注意的是，先前帧中相同节点的长期记忆信息以及同帧中其他已遍历过的节点包含的上下文信息**无缝的结合了（are seamlessly incorporated）**。（Therefore, we can expect this function to be able to produce good predictions.）

其实这里我觉得很扯淡，因为两类信息合并了并且由一个参数阵做了放射转换，而认为该函数能够产生好的预测值。就这~~



使用了信任门的stlstm神经元的数学公式表达为上式。



同时，上图表达了信任门对输入数据的控制过程。

3.5 分类器

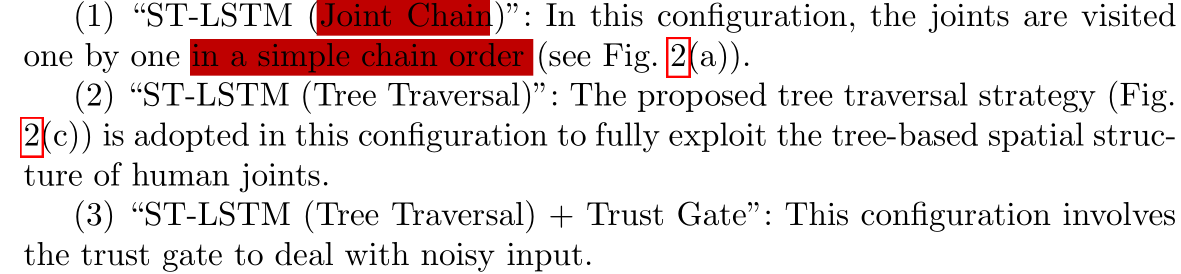
作者使用softmax层，同时经验表明，**将所有神经元的输出值平均**，相比于最后一步中的最后一个神经元的输出会取得更高的性能。



4 实验

模型在NTU RGB+D、SBU交互、UT-Kinect和MHAD数据集上**根据不同的配置**进行了**广泛**的实验。

不同的配置如下所示：



4.1 评估数据集的信息

4.2 执行细节

每个视频序列被分割成相同长度的T个子序列，每一帧随机地从每个子序列中选择而出。**如此的方法为属于生成过程增加了随机性同时提升泛化能力。实验证明，与均匀采样帧相比，该策略获得了更好的性能。**在NTU数据集上，对基于留一法协议的性能进行了交叉验证。

作者使用SGD策略，设定学习率为0.002，动量值为0.9，学习率下降率为0.95，

我们设置为神经元的size大小为128，同时高斯函数中的超参数设置了0.5，模型使用了叠加模型，dropout设置为0.5。

**尽管不同的数据集的序列长度、节点数量和数据采集设备有所不同，但我们仍然使用上述相同的参数设置。**

**这也表明了我们的方法对参数设置不敏感。因为它在具有相同配置的所有数据集上获得了promising的结果。**

4.3 实验数据

NTU RGB+D 数据集

数据集有两个评估协议，一个式交叉人物对象的评估，一半用于训练，一半测试。另一个是交叉视角的评估，两个视角用于训练，第三个用于测试。

SBU 交互数据集

使用了五折交叉验证的训练方式。

4.4 信任门的有效性

5 结论