# GCA-LSTM

## 1 摘要

LSTM具有对序列数据的**动态特征和依赖关系**进行建模的能力。

由于并不是所有节点都对行为分析具有价值，同时不相关的节点通常会带来噪音影响，因而我们需要将更多的精力集中在有价值的哪些节点上。

LSTM模型引入注意力机制，使得模型能够有选择性地集中在有价值地节点上。同时为了实现从序列数据中抽取出可信的注意力表示，模型引入了循环机制，即注意力行为可以迭代进行。

On some challenging benchmark datasets

## 1 引言

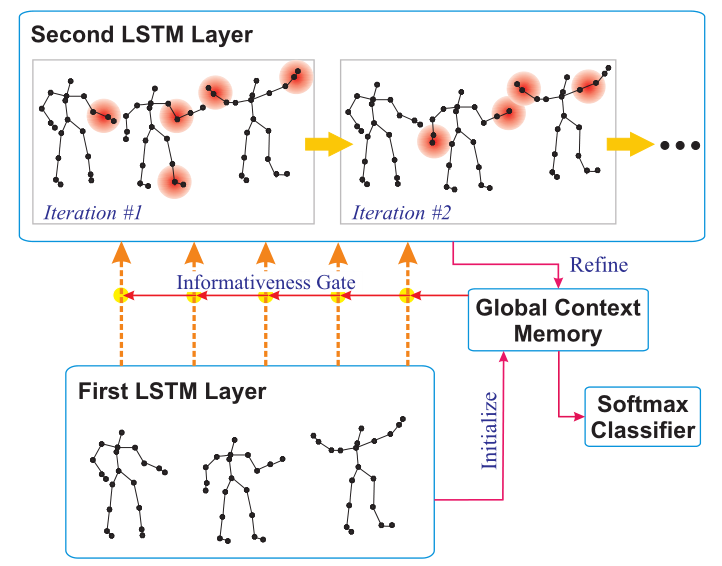
人类行为可以通过3D空间内一系列骨骼节点的移动的组合以表达。

不同的行为序列有其对应的不同的有价值的节点，同时在该序列中节点的价值程度也会随着帧的变化而改变。

LSTM模型在行为识别中缺乏足够的注意力，主要是因为LSTM在获取全局上下文信息方面存在限制。

LSTM尽管较后的神经元的隐藏表示包含了较为广泛的上文信息，但是它们内容仍然相对匮乏，因为LSTM本身对过远的信息进行记忆的能力存在问题。

在GCALSTM中，全局的上下文信息被送入每一步中，由此模型可以通过该信息来衡量每一步的新输入的价值分数，进而调整每个输入对应的注意力权重（**其实每一步对应的是每一帧数据，还是每一个节点呢？—— 准确的说是节点应该。**）。



这里的上图1所示，GCALSTM模型包含两个lstm层。其中 第一层对骨骼序列进行编码，生成一个初始的全局上下文记忆。之后全局信息会被送入第二层以协助模型有选择性的将精力集中在每一帧中更有价值的节点上，进一步提升对全局行为的注意力表示能力。接着，注意力表示信息又会对全局信息进行修正。特别地，多个注意力行为的迭代能更好的提升全局信息的记忆能力和水平，最后，修正后的全局信息会被送入分类器中以预测行为的类别。

文章的贡献点：

1 提出了带有注意力能力的模型；2 提出循环注意力机制，以大幅度提升模型的注意力性能；3 对有价值的节点的骨骼的可视化展示。

## 2 相关工作

3D行为识别的场景，第一阶段是手工提取特征，第二阶段是基于RNN\LSTM的方法

注意力机制

虽然已经有一些文章探索过了注意力机制，但是我们的方法与他们在以下方面不同：他们都**使用的是LSTM先前时间步的信息量较为局部的state，以计算下一时间步的注意力分数**，但是对于全局的分类问题，全局信息对于可信评估每个输入的重要性十分必要，因此作者提GCA-LSTM，**引入全局信息以判定每个输入的价值分数**。除此之外，作者与前者工作的不同还在于引入**循环注意力机制**，以提升行为识别任务中的注意力能力。

## 3 GCA-LSTM网络

在这一节，作者先回顾了2DST-LSTM模型，因为它是GCA-LSTM的基础模型。

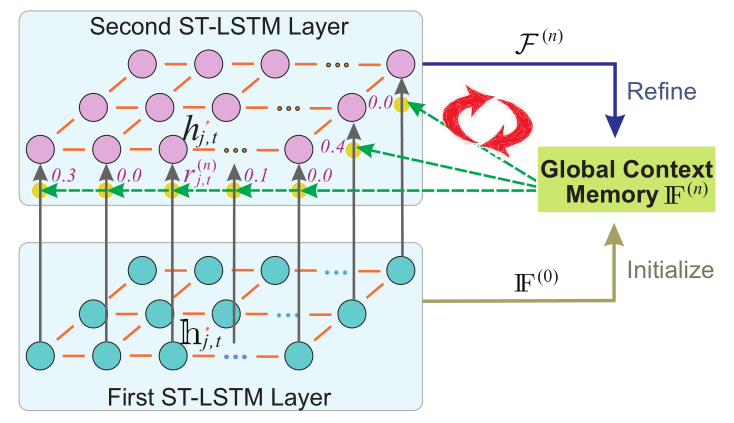
### 3.1 时空LSTM模型

不同帧中相同节点之间的时序依赖性和同一帧中不同节点之间的空间依赖性都是行为分析中的重要信息。

最近，Liu提出了一个2DST-LSTM的网络模型以同时构建空域和时序方面的依赖和上下文信息。

### 3.2 GCA-LSTM模型

图3显示全局信息对注意力计算步骤的协助的步骤图



**Overview：总览**

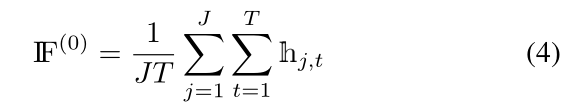
GCM包含了全局行为序列的全局表示；

第一层LSTM实现对骨骼序列的编码以及对全局信息的初始化。

第二层LSTM在时空领域中对每一步的输入执行attention机制以产生行为的注意力表示，同时该表示将会用来修正GCM。

**初始化GCM全局信息记忆内容**

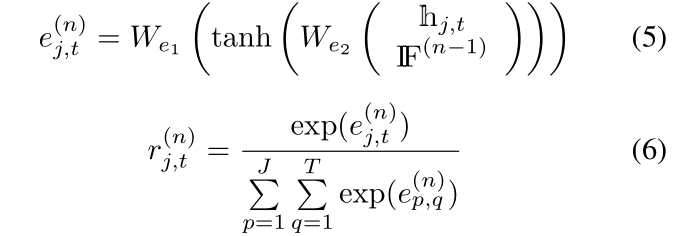
使用第一层的输出以生成一个全局信息表示，具体是：使用第一层LSTM的隐藏单元的输出的平均值以初始化一个全局信息记忆：

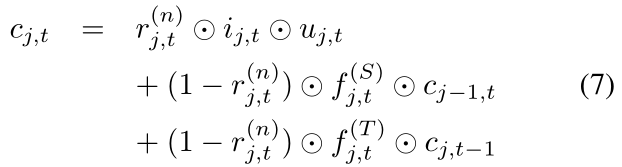


其次，第一层的输出内容将喂入到一个前向的神经网络中。

总的来说，平均化操作不涉及任何的新参数，而后者带来了大量的新参数。

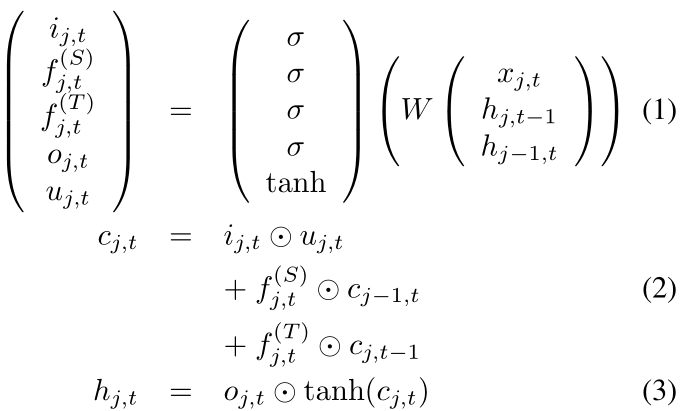
**第二个ST-LSTM层中的注意力机制**





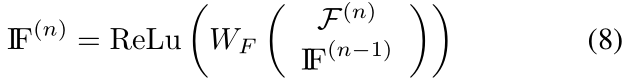
这里的cell state的更新模式可以解释为：如果在全局信息的影响下输入数据被判定为重要的或者有价值的，我们将会通过**加入更多的信息**以更新第二层的memory cell。然而，如果输入时无关的，那么我们**需要抑制它在memory中的影响力，同时充分利用更多的历史信息**，以更新memory cell的内容。

对比原始LSTM的公式：



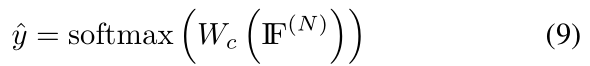
改善全局信息记忆：

第二层中的LSTM网络的最后一步的输出内容将会用来更新GCM，改善的公式如下示：



**分类器的学习**

最后一个全局信息记忆内哦让那个进入softmax中产生预测label，数学公式如下：



之后借助**负对数似然损失函数**来衡量预测值和真实值之间差异。

## 4 实验

三个benchmark datasets：NTU RGB+D、UT-Kinect、SBU-Kinect数据集。

为了调查该模型的有效性，作者使用了三个不同的模型架构在三个数据集上进行（笛卡尔积）个实验。

（1）ST-LSTM+前馈神经网络。

它类似于STLSTM，但是不同的时，第二个LSTM层中每一步的隐藏表示都被合并了，且被送入一个一层的前馈神经网络中以初始化全局信息记忆。另外，该模型中分类器时基于全局信息记忆进行的判断，而STLSTM论文中时具有与每一步的预测值的分数的平均值以预测。

（2）GCA-LSTM网络

它即是所提出的模型，其分类器基于全局信息记忆

（3）GCA-LSTM – 无attention机制

该模型具有全局信息表示，它时有每一步的隐藏表示的平均值得到，准确的说，是通过公式2以更新全局信息表示。

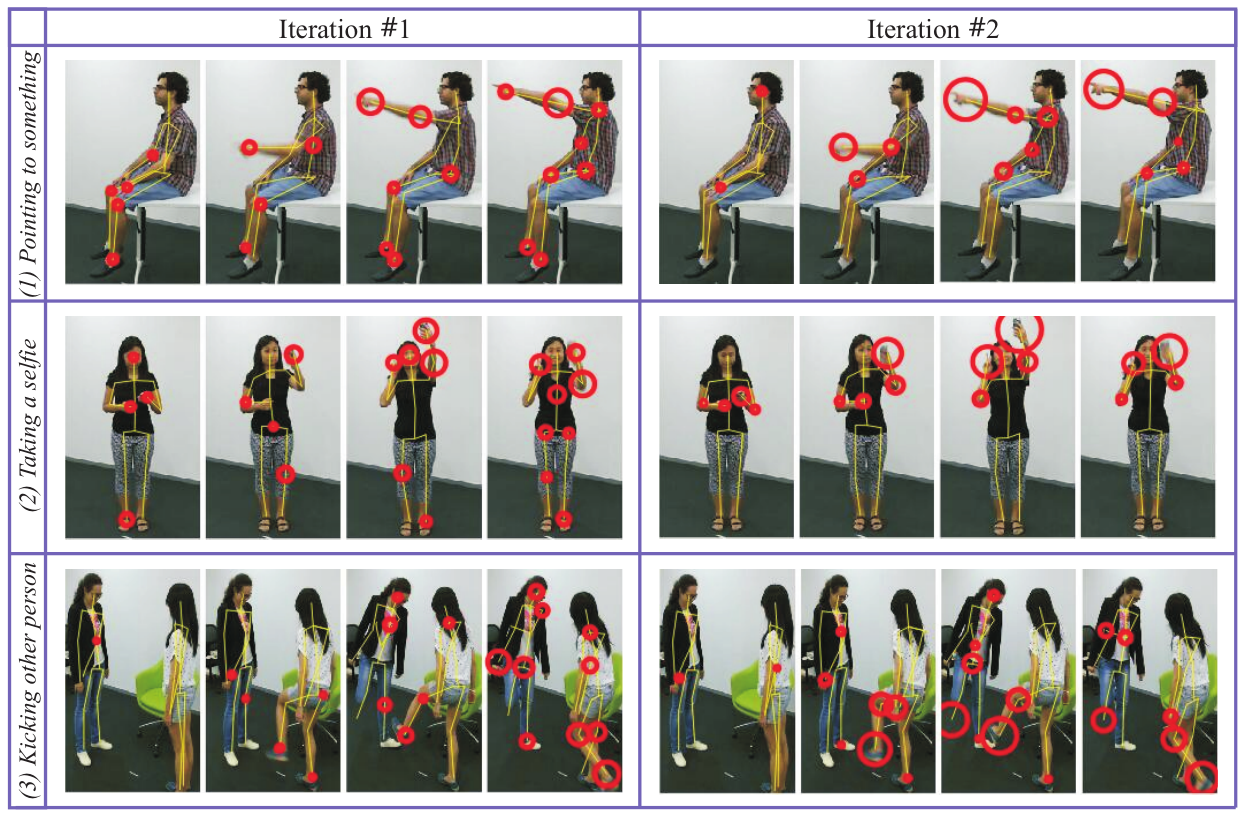
实验细节：

SGD，学习率是0.0015，学习率下降率0.95，动量为0.9。dropout率是0.5，第一层LSTM是双向的ST-LSTM。

为了公平比较，实验使用相同的帧数进行采样训练。T=20**表示什么意思？**

4.4 可视化和讨论

1 一次、两次、三次迭代后关节的价值程度的变化。



2 不区分行为的差异，综合来分析人体节点的价值程度

