摘要

脑机接口（Brain Machine Interface, BCI）是一个完成大脑神经元和外部机器互联，通信与控制的接口。脑机接口可以代替受损的神经系统，通过大脑信号采集，信号处理，解码为计算机指令，反馈这四步，为残障人士提供自动化服务。本文主要研究植入式脑电信号处理，压缩，以及xx脑电信号的解码。植入式脑电信号具有高采样率，多通道，高分辨率的性质，因此压缩对于信号存储和传输都是必不可少的环节；另一个脑机接口的核心工作是神经信号的解码，即将神经信号翻译成运动指令，只有正确地解码才能实现BCI执行符合用户意愿。

在压缩方面，过去的神经信号压缩工作主要关注于非植入式脑电信号，如EEG（脑电图），EMG（肌电图）等，这些方法都应用了相关信号的特性。但这些信号同植入式信号有很大差异，它们的有效信息都在低频部分，可以直接通过高频滤波器获得有效信号，而植入式脑机接口获得的信号中高频部分包括神经元锋电位（spike），它能有效地进行神经活动解码，因而不能丢弃。所以，EEG等信号的压缩方法不能直接套用在植入式脑机接口获得的信号中。本文通过植入式电极研究大脑运动皮层信号性质，利用信道内部神经信号特性建立了一个完整的高保真运动皮层信号压缩框架，达到在维持信噪比为36db下压缩比例为18%，并以92%的保真度保存神经元锋电位信号，保证了重建效果。

在解码方面，……本文中，我们将深度学习的方法应用于xx信号，分别在空间和时间维度建立深度神经网络，结合神经信号的时序性提高解码准确性，完善模型。

本文工作实现了植入式脑机接口神经信号的高保真压缩，xx信号的解码研究，主要创新点在于（1）对植入式脑电信号进行光谱分析，针对压缩问题提出了频谱-幅值压缩框架，有效地完成了运动皮层信号压缩，（2）通过神经网络对神经信号进行解码，利用了信号的空间结构与时序关系。

关键词：脑机接口，压缩，卷积神经网络，长短期记忆方法，循环神经网络

目录

摘要

1. 绪论
   1. 研究背景与意义
      1. 脑机接口介绍
      2. 植入式脑机接口信号分析
   2. 研究现状
      1. 脑电信号压缩现状
      2. 脑电信号解码现状
   3. 研究目标与内容
   4. 论文章节安排
2. 实验设计及信号处理
   1. 猴子抓握实验
      1. 实验设计
      2. 信号采集
   2. P300信号实验
      1. 实验设计
      2. 信号采集
3. 运动皮层神经信号压缩
   1. 运动皮层信号特性
   2. 压缩算法设计
   3. 两步编码方法
      1. LAC符号编码法
      2. HAC量化
      3. 混合无损编码方法
   4. 压缩框架总流程
4. 基于卷积神经网络的xx信号解码
   1. 深度学习介绍
   2. 神经网络的训练
      1. Feed-forward
      2. Back-propagation
   3. 卷积神经网络结构
   4. 运用卷积神经网络解码xx信号
5. 基于循环神经网络的信号解码
   1. 时序信号分类方法
   2. 循环神经网络
   3. 长短期记忆网络
   4. 运用循环神经网络解码xx信号
6. 实验结果
7. 总结与展望
8. 神经科学概述
   1. BCI 介绍

由于现代计算机技术和神经科学学科的迅速发展，人们已经可以将大脑中的运动与计算机设备相关联，通过机器捕捉大脑中各个通道的活动（Review见Van Gerven et al., 2009 和Wolpaw et al. 2002）。这种方法与应用统称为脑机接口（brain computer interface，或BCI），以探索大脑活动与特定神经状态的关系。其中特定的神经状态也叫做签名（signatures）。一个BCI需要包括：

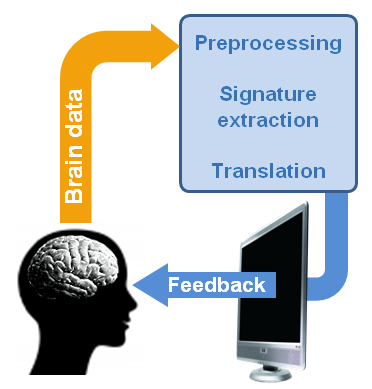
①记录大脑活动

②提取并处理签名

③将签名翻译成计算机指令

④最后返回给用户

整个过程的技术。



* 1. P300

现在有很多测量脑信号的技术，如fMRI （functional magnetic resonance imaging，功能性磁共振成像），NIRS（near-infrared spectroscopy，近红外光谱学），EEG（Electroencephalograph，脑电图），MEG（Magnetoencephalography，脑磁图）等。针对不同的采集信号，其信号预处理方法也各不相同。但是BCI最基本的任务都是正确地识别“签名”并翻译成机器指令以完成用户的意愿。在P300拼写中，可以通过“oddball task”（如图）产生P300信号。这个oddball task是一个关于字符注意的实验，展现一串字符（如SSTSSSSTSS），其中出现频率高的称为标准刺激（如该信号序列中的S），频率低的称为异常刺激（如该信号序列中的T）。当出现一个异常刺激时，300ms后就会在EEG信号中产生一个正向偏移。这里标准刺激和异常刺激的差异可以用来识别所给刺激的类别，然后基于刺激发送信号给计算机指令执行。



Figure 2. 标准刺激（S）中的异常刺激（T）

在P300实验中，字母都展现在一个matrix中，其中同一时刻只有一行或一列亮起，具体哪一行或哪一列随机闪现。如下图所示，当一行一列相继闪烁的交点为指定字母时，测试者可以集中精神在头脑中进行简单计数或者确认，相应就会有P300产生。

1. 神经信号压缩

在植入式脑机接口（invasive Brain Machine Interface）中，我们将多电极阵列（multi-electrode array, MEA）植入大脑皮层从而获取高质量的电神经信号。 这种信号的采样率为30kHz, 给数据存储和传输带来了重大负荷， 所以我们需要对数据进行压缩来降低数据量。在这一节中，我们结合大脑运动皮层神经电信号的特性，提出了一种高保真压缩算法。实验中，我们将该算法应用于哺乳动物的大脑运动皮层信号，相对原信号得到了18%的压缩率，而且没有对信号重建产生明显影响。该方法的信噪比（signal to noise ratio, SNR）达到36dB，而且spike信号也保存下来92%，大幅超过已有工作的效果。

2.1神经电信号（electroneurographic signal）

我们已经在第一章中介绍了BMI的简单构造及原理，这一节中，我们将介绍神经电信号的采集和特点。脑机接口可以分为植入式和非植入式，非植入式方法，如头皮电信号（scalp electroencephalogram, EEG）易于获取，但是信号精度很差，采样率也相应很低。相反，植入式脑机接口用外科手术的方法将电极植入大脑皮层进行信号采集，可以采集到很高精度的细胞外神经元信号。在单个神经元中这种高精度信号包含神经锋电位（spike），或者叫做动作电位（action potential）。当神经元被激发的时候就会在神经元膜上产生离子电流，导致细胞去极化（depolarize）并激发出一个spike信号。

本章中我们将专注于运动皮层神经信号。作为大脑皮层的一个重要部分，运动皮层负责计划，控制并执行人体主动行为。在运动皮层功能的相关研究中，电极阵列所采集的多通道信号通常在每个通道进行分频。通道中信号低频部分（截止频率在100Hz）对应神经信号的局部场电位（local field potential, LFP），而中频到高频部分对应于动作电位（spikes）。LFP主要源于前突触行为，反映了很多树突行为的平均电流。与之相反，Spike主要反映兴奋神经元的行为。LFP和spike信号对神经解码都很重要。对于运动皮层而言，spike通常的持续时间小于1毫秒，因此需要用高分辨率设备进行信号采集。这里我们用多电极阵列刺入细胞去采集数以百计的感兴趣神经元的信号。哺乳动物运动皮层神经元信号通常以20-30kHz的频率采集128个通道，以保证可以完好保存spike细节。这样，以16-bit的A/D分辨率计算，如果采样率为30kHz，那么128个通道的信号就会以7.68MB/s的速度进行采集。换句话说，一小时内的信号量就积累到28.8GB，这对信号的存储和传输都带来了巨大挑战。所以，我们要对信号进行压缩。

尽管BMI系统已经建立得比较完善了，脑皮层胞外信号的记录并没有深入研究过。在Electromyography（EMG）和Electroencephalography(EEG)信号的压缩上有过一些相关工作\cite{}，为了有效压缩，他们都结合了所处理信号的信号特性。但是植入式胞外信号与之相差甚远。

现有多通道压缩算法从两种思路进行实现。一种是应用通道内特性对每个通道的信号分别进行压缩，另一种是用通道间相关性同时对所有通道的信号进行压缩。从第一个思路出发，Weber等人通过基于小波的编码器对老鼠躯体感觉皮质进行压缩，然后这种方法代价是丢掉了25%的spike，对于后期的信号还原和分析并不理想。Chen等人对老鼠的S1区域进行研究，通过自适应信号量化在信噪比保持25db的时候达到的压缩率高于25%，那么信号压缩率和信号质量都得不到保证。为了改善他们的工作，Chen从第二种思路出发，利用通道间的信号相关性，在25db信噪比的情况下，将压缩率降至5%。然而，以上几种方法都丢失了太多信号细节，白费了采集来的高分辨率信号。

本章中，我们提出了一个运动皮层胞外信号的高保真压缩框架。 首先，我们讨论这种信号的3个特性：1) 信号能量集中在低频；2）离散余先变换系数中的高频部分可能对英语spike的激活模式 3）通道间相关性不稳定。根据特性（2），我们提出了一个新颖的幅值滤波器，将离散余先变换系数按幅值，而不是按频率分为两部分。低幅值成分由一个符号编码方法进行编码来降低全局失真；高幅值成分，包含主要信息和spike，由另一步骤编码。这个步骤叫混合编码，包含哈弗曼编码和一个新颖的零长编码。我们的主要工作如下：

* 设计了一种新颖的幅值滤波器，它将离散余先变换系数根据幅值分为两部分，这避免了spike信息的丢失。
* 提出了一个符号编码方法，用来对低幅值成分进行编码，而不是简单丢弃。这有效避免了全局信号失真。
* 发明了一种合并哈弗曼编码和新颖的零长编码的混合编码方法，用来对高幅值成分和低幅值成分的索引进行编码。由此，spike信息得到了精准的结构化保存。

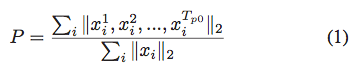
最后，我们用一系列方法测试我们提出的压缩框架，得到了平均信噪比36db，压缩比18%的效果，而且spike保真率保持在92%以上，保证了重构效果。

2. 运动皮层胞外信号特点

为了在保证信号质量的同时进行有效压缩，我们在本节中多通道胞外信号的特点。我们的数据将在实验部分进行详细描述。这里，我们从通道内特点到通道间相关性，总结出三个特点：

1. 信号能量集中在低频：

为了研究所记录信号的频域特性，我们采用离散余弦变换（discrete cosine transformation, DCT）将数据先转换到频域。作为傅里叶变换的一个变种，DCT系数得到的是一系列实数，处理起来比傅里叶变换方便。变换后的DCT系数用xi = [x1i , x2i , ..., xNi ]来表示，其中xji表示第i个通道DCT系数中的第j个元素。那么整个数据集上低频部分能量比例为：



式中，分母表示所有通道总能量，分子为所有通道的前Tp0个DCT分量的总能量，即，以Tp0为截止频率，低频部分的能量。在整个数据集上，我们将P的平均值画在图\ref{fig:power}，图中清晰表明了少量DCT分量占据了信号的主要能量。换句话说，相当大的能量集中在了低频部分。

1. 高频信号中存在显著峰值：

和其它自然信号（如图像）一样，胞外信号的主要能力也集中在低频部分。然而，这样的信号在中高频有所差异。如图\ref{fig:remarkable\_peak}所示为中频部分的一部分截断光谱，可见在7325Hz处有一个明显的峰值，对应于一个经常出现的神经元放电模式。实际上，实验表明很多通道共享这些具有峰值的频率，而一些通道没有。这可以从多电极阵列的采样原理理解，我们采集到的胞外信号的单通道信息可以由3至5个有不同spike激发模式的神经元组成。

1. 不稳定的通道间信号相关性：

运动皮层信号的第三个特点就是通道间不稳定的相关性。在这里，我们计算了通道间的信号相关性，所有样本的平均波动范围（定义为随时间平移的均值-方差比）为0.68。也就是说，相关系数

1. 信号解码
2. LSTM
3. 循环神经网络(RNN)

人类可以很快地识别节奏模式序列，即那些有时间间隔的子模式。此外，鼓手等演奏者也能根据运动命令生成有精确时间节奏的节奏序列。这促使人们研究人工系统，去分离或产生通过事件间时间间隔长度传递信息的模式。

Deep Neural Network(DNN)通常在较难的learning问题上能够达到比较好的效果。在实际问题中有很多是时序信号，如语音识别, 连续（即字符无分割的）手写识别, 蛋白质分析，股市预测等。虽然这些问题无论是否有监督，都可以交给CNN或者DNN处理并达到较好效果\cite{abdel2014convolutional, lecun1995convolutional , ciresan2011convolutional, lecun1994word , lauer2007trainable , baldi1996hybrid , zhu2014stock}，但是并不能完全利用上信号的时序性，而循环神经网络（recurrent neural network）可以利用神经元循环传递实现时序信号的分析。

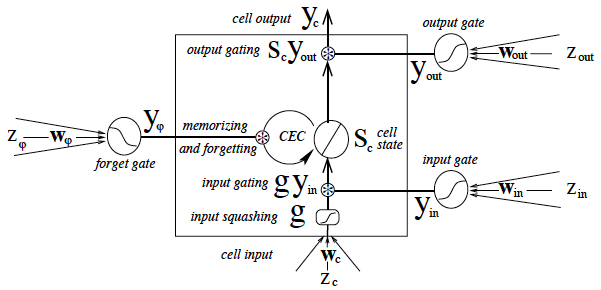
一个循环神经网络是一个结构中带反馈连接的神经网络，它可以学习并处理时序序列问题，而传统的机器学习算法由于没有考虑时间方面相关性可能并不善于解决这类问题。之前对于这种时序信号的处理方法是一些基于可学习可适应的方法，如隐马尔可夫模型\cite{ eddy1996hidden },前向网络等，但是在计算能力和物理意义上RNN都比这些方法能力强。而且事件在时域上的信息很大程度上反映了时序任务的重要信息，如运动控制和节奏检测。隐马尔可夫模型往往忽视这一信息，而循环神经网络（ Recurrent Neural Network, RNNs）可以从根本上学习应用这个信息。隐马尔可夫模型（HMM）能够成功地应用在语音识别，正是因为这种方法不喜欢特定口语单词语速产生的差异。而如节奏检测，音乐处理，和其他本文中所述的任务，都需要精确的时间测量。 虽然HMM可以通过给每个时间间隔单独创建一个内部状态来解决有限个时间间隔的问题，但是这样既麻烦又低效的，并且没有用到HMM的非线性时序伸展不变性。

事实上，尽管隐马尔克夫模型和传统的离散符号语法学习设备被限制在离散状态空间， RNNs在根本上上适用于所有的序列学习任务，因为它们具有图灵能力\cite{ siegelmann1995computational }。典型的RNN学习算法在一个潜在抗噪算法的通用空间中进行梯度下降\cite{ pearlmutter1995gradient }，这个通用算法采用分布式，用连续值表示的内部状态来将实值的输入序列映射为实值输出序列。所以循环神经网络在识别由时间距离定义的模式上更有希望。混合HMM的RNN方法\cite{ bengio1995input }可能能够结合两种方法的优点，但就我们所知，从未被用到准确的事件计时的问题。

然而，Schmidhuber的学生Hochreiter在他的博士论文\cite{ hochreiter1991untersuchungen }中指出了RNN的缺点：vanishing or exploding gradients:

在传统RNN中，反传的误差要不就收缩太快，要不就爆发超出界限。事实上误差会随着网络层数而指数型衰减或爆发，因此很难训练, 而且不能处理长时间延时事件。

之后Hochereiter提出了一种新型的RNN——长短时间记忆Long Short-Term Memory \cite{ hochreiter1997long}，它在涉及很长一段时间延时的任务上比传统RNNs更好。LSTM网络中的基本单元是包含一个或多个存储单元（memory cell）和所有cell共享的三个自适应乘法门（图2）的存储块（memory block）。每个存储单元都在其中心有一个自连接单元，我们称之为“恒误差旋转木马”（ Constant Error Carousel， CEC ），它本质上是一个无参数的线性单元。通过无限循环地进行激活和误差信号传递，CEC可以在延长的时间段内提供短期存储。Gers et al在LSTM中安置了忘记门 \cite{ gers2000learning }。输入，忘记和输出门可以通过训练分别进行学习,即学会将什么信息存入，怎样保存，什么时候读出。



LSTM的体系结构允许LSTM容忍输入的事件之间有很长的时间延时（1000步及以上）；而传统RNN用代价更高的更新算法，比如BPTT \cite{williams1990efficient}，RTRL \cite{ robinson1987utility, williams1989learning}，或它们的组合\cite{schmidhuber1992fixed, williams1995gradient}, 都无法学习超过10个时间延时的序列\cite{ hochreiter1991untersuchungen, bengio1994learning, hochreiter1997long, gers2000learning, hochreiter2001gradient}.

比如，一些我们以前任务需要用LSTM网络对发生于50个离散时间间隔之前的事件予以响应，而不受发生于之前49个时间间隔的事件的影响。就在关键时刻之前，如果有一个有用的“标记”输入通知网络，其下一步的行动将是至关重要的。因此，网络没有必要学习度量50个步骤的时间间隔；它只需要学会保存这50步的相关信息，而后一旦观察到这个标记就用这段信息即可。

但是如果没有这样的标记呢？如果网络本身需要学习测量并在内部表示这段特定任务的间隔内信息，或生成由特定时间间隔分割的序列模式，就要求在长时间连续输入流中保持网络的时间精确行和鲁棒性。显然，这样的任务通常不能由基于共同时间窗口的方法解决的，因为它们的泛化能力受时间窗口大小的限制。不幸的是，这意味着我们不能用标准基准进行测试和计时，因为就我们所知，目前没有任何其他网络学会将15个时间滞后推广到45个等。因此，我们不得不以创建一组新的比较任务。那就将用到LSTM了。

因为长短期记忆方法已被证明在长时间滞后任务上比其他RNN的方法强，而且其体系结构与人脑对信息的处理方式很类似，考虑到这种方式可能可以更好地模拟大脑皮层信号激发方式，所以我们研究该方法。LSTM通过内部细胞到多门的“窥孔连接”，在没有任何段式训练样本的情况下可以很好的区分50个和49个时间间隔的spike序列。在没有外部复位和监督学习的情况下，LSTM的变种也能学习生成时间精确的稳定spike流和其他非线性周期模式。这使得LSTM在需要的精确测量或生成有时间间隔的任务上很有效。

1. LSTM的应用

我们先来看一下LSTM的已有应用：

语音识别：

LSTM RNN是著名语音识别数据库TIMIT的benchmark (Graves et al, ICASSP 2013)，Nicole等人用LSTM网络进行RNN的retraining\cite{beringer2005classifying}，这是无法在HMM中实现的。google用LSTM RNN来改善大规模语音识别\cite{sak2014long}。在类似问题上，\cite{ sutskever2014sequence }将LSTM应用在机器翻译问题，实现了sequences to sequence学习。

连续（即字符无分割的）手写识别：

Connectionist Temporal Classification（CTC）\cite{ graves2012connectionist }训练的RNN（CTC-LSTM）\cite{ bluche2014a2ia }，在2009年跑得结果赢了很多手写识别竞赛的冠军。\cite{ graves2006connectionist }

音乐合成:

通过训练，RNN的变种RNN-RBM可以无监督地形成简单的钢琴旋律。[5] Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription.通过用音乐集训练LSTM网络，网络就可以无监督地生成有特定样式的旋律与和弦的音乐样本。而前向网络和传统RNN都无法学习出和弦音。\cite{ eck2002first }

蛋白质分析：

在RNN中还可以实现无分割的蛋白质分析\cite{ hochreiter2007fast }，其本质和无分割的手写体识别同理。

1. RNN网络结构

首先我们来看RNN的结构：

RNN包含一些用带权连接的单元，每个单元有一个在时间$t$更新的激活函数$y(t)，t=1,2,…$。对单元$i$的激活$y^{i}$通过计算网络中与$i$的相连输入$net^{l}(t)$得到：

$net^{l}(t) = \sum\_m{w\_{im}y^m(t-1)}$, 而后通过一个可导函数$f$将$net^{l}(t)$进行”压榨”，$f$比如我们之前提到的sigmoid, tanh, 或Relu函数：

$y^{i}(t) = f(net^i(t))$

其中网络的输入是随时间变化的序列，对于传统的RNN回归问题或者序列分类问题（如手写体识别），输出也是随时间变化的序列。而对于纯分类问题，输出只是一个标量。为了方便讲述传统RNN的方法，我们在这里使用输入输出都是时间序列来描述。

对有监督的RNN而言，我们将每个单元的损失计为与其输入与真实输入之差，而全局损失函数为所有样本损失的平方和，即t时刻的总误差为：

$E(t) = \frac{1}{2}\sum\_i{{e\_i(t)}^2}, e\_i(t) := Y^i(t) – y^k(t)$

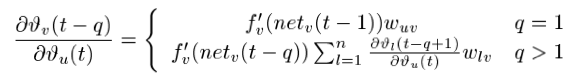
整个序列的误差为每一时刻的误差$E(t)$之和。

和CNN类似，RNN可以用梯度下降进行网络参数更新，即对于每个权重$w\_{im}$，有

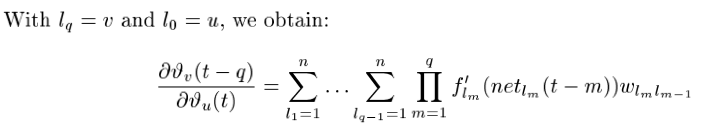
$\delta w\_{im} = -\alpha \frac{\partial E(t)}{\partial w\_{im}}$，其中$\alpha$为学习率。这种方法也是标准RNN的更新方法backpropagation through time (BPTT)方法，通过带入可得t时刻i单元的误差信息为

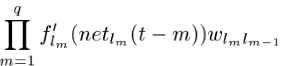
$\delta e\_i(t) = f\_i’(net\_i(t))\sum\_j{w\_{ji}e\_j(t+1)}$

然而在1990年，Schmidhuber的学生Hochreiter提出RNN在理论上很好，但是不好训练，误差会以指数增长或者消失\cite{ williams1995gradient }，原因是BPTT， real-time recurrent learning（RTRL）\cite{ williams1989learning }，这类基于梯度下降的方法都有一个问题，如果计算权重更新后的q步，则如公式（）所示，t时刻与t-q的更新误差之比与$w\_lv$有关，



进一步化简，可得



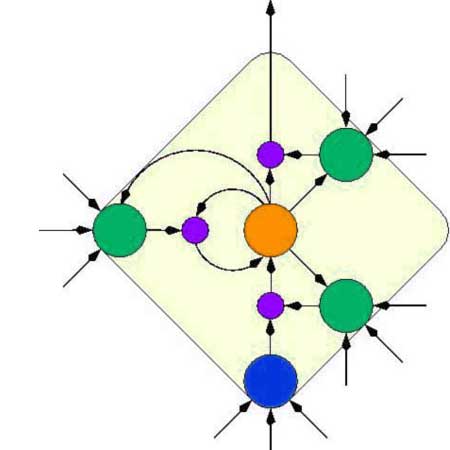
其中决定了总的传回误差。所以如果，则最大的乘积随q指数增长，就会使学习过程不稳定；相反地，如果，则最大的乘积随q指数下降以至消失，在有限时间内无法学到有效网络参数，这也被称为长时间延时问题(long time lag problem)，使得RNN的应用一度局限在短时间序列，而我们在绪论中提到的长短时间记忆方法有效克服了这一问题，我们将在下一节中予以介绍，关于RNN，更多优化方法请参考Bengio在\cite{bengio2013advances}中的综述。

1. LSTM的网络结构

LSTM通过设置恒定的传递误差来解决上面提到的问题。 在基本LSTM单元中，LSTM引入了一个自连接线性单元，称作误差旋转木马（error carousel）, 恒定的传递误差中设定权重$w$恒等于1，$f$为线性函数，目的是这各部分只起到传递误差的作用，而将学习参数作用于其他权重。

粗略地表示，我们在第一章中简单介绍了LSTM。它以memory block作为基本单元，其中包含一个至多个memory cell，典型的LSTM memory cell如下图所示：

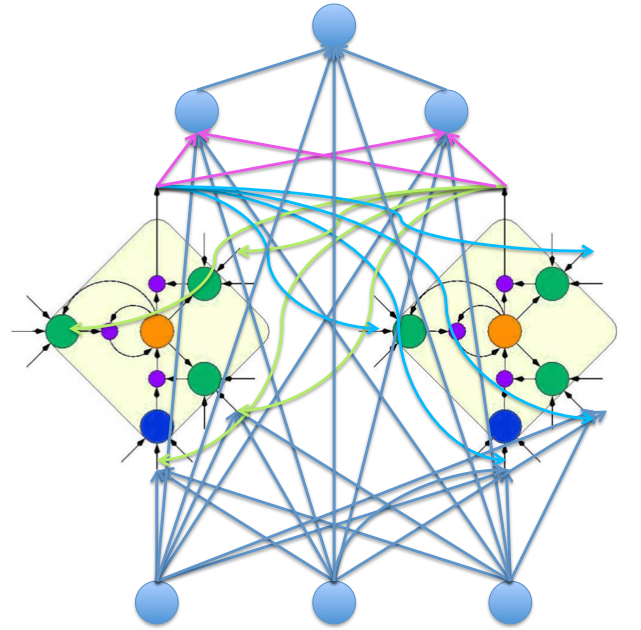
核心有一个线性单元（或者称作线性神经元，如图橘黄色所示），这是个线性单元，在任意给定时间，该单元将其所有输入通过加权连接求和。它的自循环连接权重固定为1（即与左边紫色圆点之间的半圆连接权重固定为1），这样通过确保训练信号或者误差在被传递的时候不会消失（vanish）克服了原先RNN的一个主要问题：vanish（参考3.2节）。而且这个核心神经元为线性让LSTM实现了可以识别多达1000步之前的模式。



在这个核心周围是非线性可适应神经元，用来学习非线性神经动作。图中蓝色神经元用来表示输入，三个绿色神经元分别表示三个门（左边的是忘记门，右下角为输入门，右上角为输出门），这些门通过学习来使核心线性神经元不受无关事件和误差信号的干扰，而又能通过有效信息更新网络参数，即门是否打开来决定信号的有效性。最后，紫色圆点用以表示乘积操作。LSTM的学习算法是非常高效的,每个边权的计算复杂度不超过O(1)。

LSTM的应用中，可以将多个memory cell进行组合，如图（）所示混合模型。其中有两个memory cell，分别从输入和输入门接受输入，……

。但实际应用中，通常将一至多个memory cell组合成memory block，共享输入，输出，忘记门的权重（而不共享输入的权重），这样可以减少自适应参数的数量，同时使block内部各cell起到不同作用。



本文中用

j表示memory block 的下标；

v表示block j中memory cell的下标；

表示第j个block中第v个cell；

表示unit m到unit l之间的权重；其中m是源unit，取决于网络拓扑。

一个源unit激活用表示；

Memory cell c的输出由当前cell的状态和四个输入源确定：

，；

3.2 常数误差传递：简单方法