**基于双阶段注意力机制的递归神经网络时间序列预测方法(DA-RNN)**

**Abstract**

* 非线性自回归外生(NARX)模型是根据一个时间序列以前的值以及多个驱动(外生)序列的当前值和过去值来预测时间序列当前值的模型，已经研究了几十年。尽管已经开发了各种各样的NARX模型，**但很少有模型能够恰当地捕获长期的时间依赖关系，并选择相关的驱动序列进行预测**。针对这两个问题，本文提出了一种基于双阶段注意力机制的递归神经网络(DA-RNN)。在第一个阶段，我们引入一个输入注意机制，通过在每个time step参考以前的编码器隐藏状态，来自适应地提取相关的驱动序列(输入特性)。在第二阶段，我们在每个时间步长使用时间注意机制来选择相关的编码器隐藏状态。利用这种双阶段注意方案，我们的模型不仅可以有效地进行预测，而且还可以很易于解释。基于SML 2010数据集和NASDAQ 100股票数据集的深入实证研究表明，在时间序列预测方面，DA-RNN的表现优于最先进的方法。

**1 Introduction**

* 时间序列预测算法在许多领域得到了广泛的应用，如金融市场预测[Wu et al.2013]，天气预报[Chakraborty et al.， 2012]，复杂动力系统分析[Liu and Hauskrecht，2015]。虽然著名的自回归移动平均(ARMA)模型[Whittle, 1951]及其变体[Asteriou and Hall,2011;Brockwell和Davis, 2009]已经证明了它们对各种现实世界应用程序的有效性，**它们不能对非线性关系建模，也不能在外生(驱动)输入项之间进行区分**。为了解决这个问题，各种非线性自生回归模型(NARX)被提出。通常，给定目标序列的前一个值和驱动(外生)序列的当前值和过去值，NARX旨在学习目标序列当前值的非线性映射，此处F(·)就是要学习的映射函数。
* 尽管已经通过内核方式、集成方式、高斯过程等方法进行了大量的时间序列预测。**但是基本上大多数这些方法的缺点是采用预先定义的非线性形式，可能无法适当地捕获真正的潜在非线性关系**。RNN，一个专门为了序列建模所涉及的深度神经网络，由于其在捕获非线性关系方面的灵活性而受到广泛关注。尤其，近些年来在NARX时间序列预测方面，RNN展现出了一定的成功。**然而，传统的RNN会遇到梯度消失，因为无法捕获长期的依赖关系**。近些年来，长短时记忆单元(LSTM)和门控循环单元(GRU)客服了该限制并成功在多个领域得到了应用，例如神经机器翻译、语音识别和图像处理。因此，考虑最先进的RNN网络进行时间序列预测是很自然的，例如编码解码网络[Cho et al.,2014b; Sutskever et al., 2014]和基于注意力机制的编码解码网络 [Bahdanau et al., 2014]。
* 基于LSTM和GRU，编码解码网络变得流行，因为它们在机器翻译反面所取得的成功。其关键思想是将源语句编码为固定长度的向量，然后使用解码器生成翻译。**该方法得到的机器翻译的一个问题是，随着输入序列的长度增加，编码解码网络的性能会急剧恶化**。在时间序列分析中，这个会是一个严重的问题，因为我们通常希望根据目标系列和驱动系列中相对较长的一段进行预测。为了解决这个问题，基于注意力机制的编码解码网络应用注意力机制选择所有时间步长的部分隐藏状态。最近，一个分层注意力网络被提出，它使用了两层的注意力机制去选择所有时间步长的相关编码器隐藏状态。尽管基于注意力的编码解码网络和分层注意力网络在机器翻译、图像字幕和文件分类方面取得了一定的效果，**但是他们不适合于时间序列预测。这是因为当多个驱动序列可用时，这个网络无法准确选择相关的驱动序列进行预测**。此外，他们主要应用在分类领域而不是时间序列预测。
* 为了解决上述提到的问题，受一些人类注意力理论的启发[Hubner¨等人，2010]，假定人类行为是由一个两阶段的注意机制建模(指引)，我们提出了一个新的基于双阶段注意力机制的递归神经网络(DA-RNN)进行时间序列预测。在第一个阶段，我们提出一种新的注意力机制，通过参考先前的编码器隐藏状态，在每个时间点来自适应地提取相关驱动序列。在第二个阶段，一个时间注意力机制在每个时间点被使用来，以此选择相关编码器隐藏状态。这两种注意力模型在基于lstm的递归神经网络(RNN)中得到了很好的集成，可以通过标准的反向传播进行联合训练。通过这种方式，DA-RNN可以自适应地选择最相关的输入特性，并适当地捕获时间序列的长期时间依赖性。为了证明DA-RNN的有效性，我们将其与使用了具有大量驱动序列的SML 2010数据集和NASDAQ 100股票数据集的最新方法进行了比较。大量的实验不仅证明了本文提出方法的有效性，同时证明了该网络具有较好的解释性和鲁棒性。

**2 Dual-Stage Attention-Based RNN**

* 在这一章中，我们第一次介绍在本文中将要使用的符号和我们将要研究的问题。然后我们讲解一下DA-RNN对时间序列预测的动机和细节。

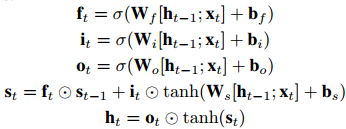
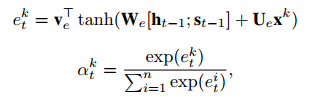
**2.1Notation and Problem Statement**

**2.2Model**

* 许多关于人类注意力的研究都认为，行为结果最好由两阶段注意力机制来建模。第一阶段选择初级刺激特征，第二阶段使用分类信息对刺激进行解码。受到这些理论的启发，我们提出一种新的基于双阶段注意力机制的RNN网络进行时间序列预测。在编码器中，我们引入一种新的输入注意力机制，该机制可以动态选择相关的驱动序列。在解码器中，时间注意力机制在所有时间点自动选择相关的编码器隐藏状态。对于目标，使用平方损失。利用这两种注意力机制，DA-RNN能够自适应地选择最相关的输入特征，并捕获时间序列的长期时间依赖性。图1显示了所提议的模型的图形说明。

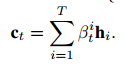
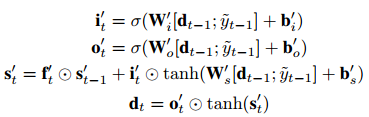
**ModelEncoder with input attention**

**编码阶段的步骤可以总结如下：**

* 编码阶段主要是进行由输入序列xt到编码器隐藏状态ht的映射，主要使用LSTM单元作为映射函数完成。使用LSTM单元的主要原因：单元状态总和随时间的活动，这可以克服渐变消失的问题，并更好地捕获时间序列的长期依赖性。  
  在这里插入图片描述
* 为了实现注意力机制的作用，使用t-1时刻的隐藏状态ht-1，LSTM单元状态st-1和driving series(即每行数据)，计算求得t时刻每行的权重，并进行归一化处理。将求得的权重赋权到inputing series(每列数据)上。  
  
* 由加权后的inputing serise(每列数据)和隐藏状态ht-1完成LSTM映射，得到t时刻的隐藏状态ht。  
  在这里插入图片描述在这里插入图片描述
* 由于输入注意力机制考虑了每个driving series对ht的作用，利用该输入注意机制，编码器可以选择性地聚焦于特定的driving series，而不是对所有的输入驱动序列都一视同仁。

**Decoder with temporal attention**

**解码阶段的步骤可以总结如下：**

* 该阶段主要求t时刻的预测值yt。使用t-1时刻解码器的隐藏状态dt-1和LSTM(注意编码器和解码器各使用了一个LSTM)的单元状态s’t-1，与编码器阶段的隐藏状态h的每列数据作用得到T(即每行数据的length)个值，并进行归一化得到权重。  
  在这里插入图片描述在这里插入图片描述
* 将编码器隐藏状态矩阵的每列数据进行加权求和得到语义向量ct。注意，每个时刻都有一个语义向量。  
  
* 使用t-1时刻y的真实值yt-1与此时的语义向量ct-1进行作用，得到y在t-1时刻的预测值。  
  在这里插入图片描述
* 使用t-1时刻的解码器隐藏状态和此时刻预测值进行映射得到t时刻的解码器隐藏状态，此时的非线性映射函数也选为LSTM。  
  在这里插入图片描述
* 最后使用如下公式进行映射得到y在整个时间段之内的预测值：

在这里插入图片描述