

华泰期货量化专题报告： 利用循环神经网络预测期货收益

循环神经网络简介

循环神经网络是人工神经网络的一种特殊类型，其最先应用是在自然语言处理当中，其地位与卷积神经网络在图像处理中类似。这种神经网络出现于20世纪80年代，在发展早期，由于计算量大应用不多，但由于近年来计算机硬件技术的发展与深度学习技术的成熟，循环神经网络变得越来越流行。这种网络特别适合时间序列数据，因为其内部神经元可以通过有效架构组合保存时间序列上的信息。

期货基本面研究员在每次思考时都不会重头开始，而是保留了之前的信息，或者经过其思考加工的结果。例如研究员会提出目前某商品的价格正处在历史低位，未来走高的可能性很大的表述。如果单从动量的角度来看处在历史低位可能是价格连续单边下跌造成的，那得出的投资决策应该是继续做空该品种，但是考虑到目前的历史低位，也就是研究员回溯了过去一段更长的历史由此得出的结论，所以在期货研究上往往也包含了对过去的认知，而单纯的动量策略是没法容纳这个信息的。而从理论上讲循环神经网络能弥补纯动量策略的这个缺陷。另外研究人员所谓的“历史低位”这个概念其实也很模糊，这个历史低位究竟是几个月还是几年的低位也很难有一个合适的客观标准。而使用循环神经网络虽然也不能完美解决这个问题，但是作为量化投资的一种手段，这类模型至少可以通过历史回测，提供一种相对客观的方式去处理这个问题。

其实循环神经网络在金融上早有应用，比较常见的状态空间模型(State-Space Model, SSM)和隐含马尔科夫模型(Hidden Markov Model, HMM)等动态贝叶斯网络经常用在金融时间序列的处理上。这些动态贝叶斯网络其实可以理解为只有一层而且包含少量神经元的循环神经网络。这类模型在自然语言处理上已逐步被新兴的循环神经网络所取代，所以循环神经网络在金融投资上的应用也并非不可能。

本报告首先介绍了循环神经网络与普通前馈神经网络的差异，然后介绍了两种常见循环神经网络的变体长短时记忆单元和门控循环单元，最后尝试应用门控循环单元构成的循环神经网络对沪金、PTA和菜粕主力期货的5日收益率进行预测，并取得一定效果。

华泰期货研究所 量化策略组

陈维嘉

量化研究员

☎ 0755-23991517

✉ chenweijia@htgwf.com

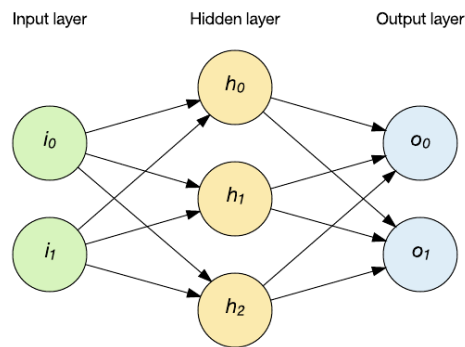
从业资格号：T236848

投资咨询号：TZ012046

循环神经网络架构

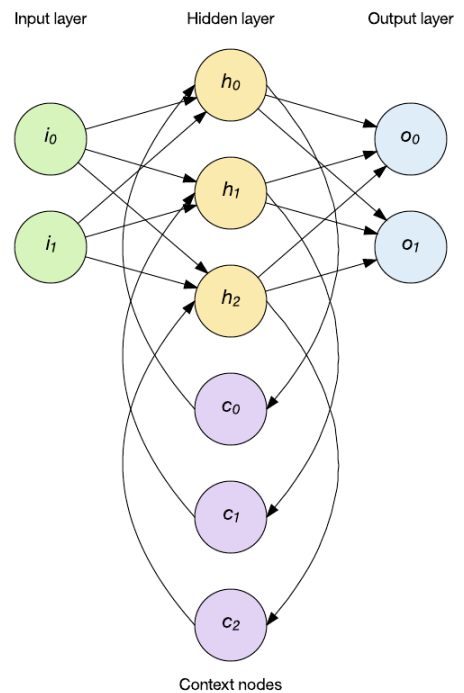
循环神经网络(Recurrent Neural Network, RNN)是人工神经网络的一种特殊组织形式,他们首先被用于自然语言相关的时间序列处理,比如机器翻译或者模仿某位作家的写作风格等。与前馈神经网络相比,循环神经网络的架构充分考虑了输入序列随时间变化的特点。下图对比了前馈神经网络与循环神经网络的架构。

图 1: 前馈神经网络架构



数据来源: 网络截图

图 2: 循环神经网络架构



数据来源: 网络截图

<https://www.ibm.com/developerworks/library/cc-cognitive-recurrent-neural-networks/index.html>

如果一个周频时间序列包含两个维度,比如期货收益率和库存,那么应用在图 1: 的前馈神经网络上就是第一周的期货收益率和库存作为输入层(Input Layer)进入 i_0 和 i_1 这两个神经元中,然后经过隐藏层(Hidden Layer)的三个神经元处理,再进入输出层(Output Layer)预测下一周的期货收益率和库存。训练这个模型需要把每一周的期货收益率和库存作为输入变量,下一周的期货收益率和库存作为输出变量。同时,采用这样一种架构就相当于这个模型认为每一周的期货收益率和库存都是独立的,不会受之前一周或几周的期货收益率和库存影响。如果按着这种思路图 1: 架构是无法充分利用期货收益率和库存存在时间维度上的变化信息,而图 2: 中的循环神经网络就是为了迎合这种需求而设计的。同样图 2: 中的输入层包含了当前一周的期货收益率和库存这两个维度。输入层的信息进入隐藏层后,一部分进入输出层,而另一部分则会进入到内容节点(Context nodes)中贮存起来,这部分可理解为网络的内部记忆,待进入下一个时间周期,即与下一周的期货收益率和库存数据一同进

入隐藏层。通过内容节点的使用，不同时间节点上的期货收益率和库存信息就能够被模型记录下来从而与输入一起影响下一周的输出结果预测。通常循环神经网络的样本都包含了多个时间步骤，这多个时间步骤其实也可以全部作为输入放入图 1: 的前馈神经网络中，这么做的话输入层神经元的数目就会变为时间步骤数量×时间序列维度，那么输入层与隐藏层之间的连接矩阵也会相应加大。这也是利用神经网络处理时间序列的一种方法，并且有研究人员用于股票动量策略的改进，取得不错的收益，详情可参考 *Applying Deep Learning to Enhance Momentum Trading Strategies in Stocks*。

但是直接使用前馈神经网络并没有充分考虑时间序列的特点，而且在机器学习的常见的应用领域自然语言处理中，前馈网络并没有优势。另一方面，在金融时间序列建模上比较常用的是隐藏马尔科夫模型(Hidden Markov Model, HMM)和状态空间模型(State Space Model, SSM)等动态贝叶斯网络(Dynamic Bayesian Network)，所以在这篇报告里不会把前馈神经网络作为处理金融时间序列的工具进行重点介绍。而金融建模中常见的动态贝叶斯网络只是循环神经网络的一种特殊形式，在其诞生领域自然语言处理中已逐渐被循环神经网络所取代，所以这里会作为重点介绍对象。

如果把循环神经网络展开如图 3: 所示，则也可把循环网络表示为一个多层的前馈网络。如果一个时间序列包含 5 周，那么这个循环网络展开后就有 5 层，每层对应一个时间步骤。下图中的 x 作为循环神经网络的输入， s 为隐藏层变量， o 为输出层变量， U 、 V 和 W 是各层间的矩阵。展开后的循环神经网络与前馈网络最大的不同是各层都是使用了相同的矩阵 U 、 V 和 W 。所以从某种意义上讲循环神经网络结合了时间序列的特点，所使用的参数更少，在一定程度上也有助于信息的压缩。如果用 x_t 向量表示在时间 t 上的输入， s_t 表示隐藏层和内容节点在时间 t 上的值，则 s_t 的计算可表示为

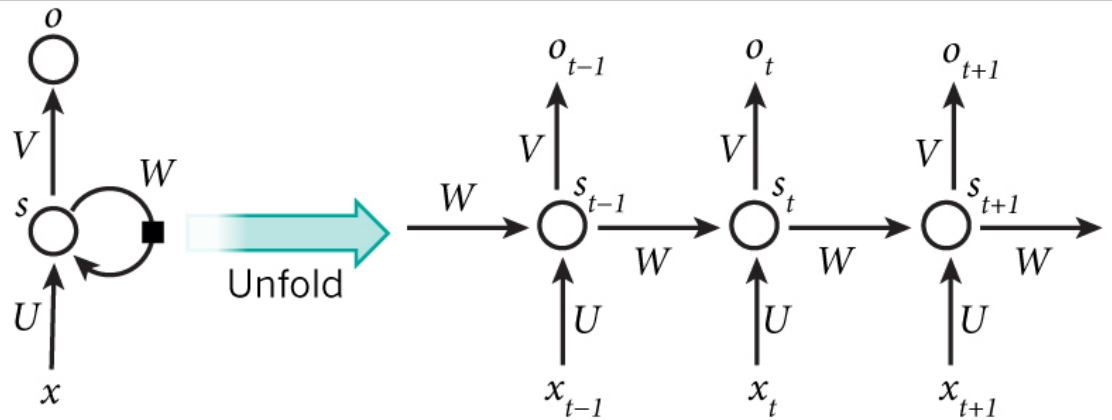
$$s_t = \sigma(Us_t + Ws_{t-1})$$

其中 σ 为激活函数

t 时刻输出层向量 o_t 则由 $o_t = Vs_t$ 计算。

虽然循环神经网络在一定程度上能够模拟时间序列，但是如果两个时间步骤之间间隔太大，则循环神经网络不能有效记录其序列特征，而这些特征很有可能是期货时间序列模拟所必须的，例如农产品库存或价格的季节性变化。造成这种现象的原因通常是由于深度神经网络训练过程中，多层神经网络的梯度不稳定形成梯度消失或者梯度激增的现象。所以直接应用这种简单循环神经网络的情况较少，而大多数情况是应用他们的变体，通过增加控制门来实现长时间信息的贮存。

图 3: 循环神经网络的展开



数据来源: Nature <http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>

长短时记忆单元

长短时记忆(Long Short Term Memory, LSTM)网络尝试使用‘门’机制来改善一般循环神经网络所产生的梯度消失问题。最早在 1997 年,由 Sepp Hochreiter 和 Jürgen Schmidhuber 提出。LSTM 其实是在图 2: 中在输入层和隐藏层之间,隐藏层和输出层之间加入了‘门’来控制隐藏层的更新和输出。LSTM 作为一个改进的隐藏层可以用以下公式来表示

$$i = \sigma(x_t U^i + h_{t-1} W^i)$$

$$f = \sigma(x_t U^f + h_{t-1} W^f)$$

$$o = \sigma(x_t U^o + h_{t-1} W^o)$$

$$g = \tanh(x_t U^g + h_{t-1} W^g)$$

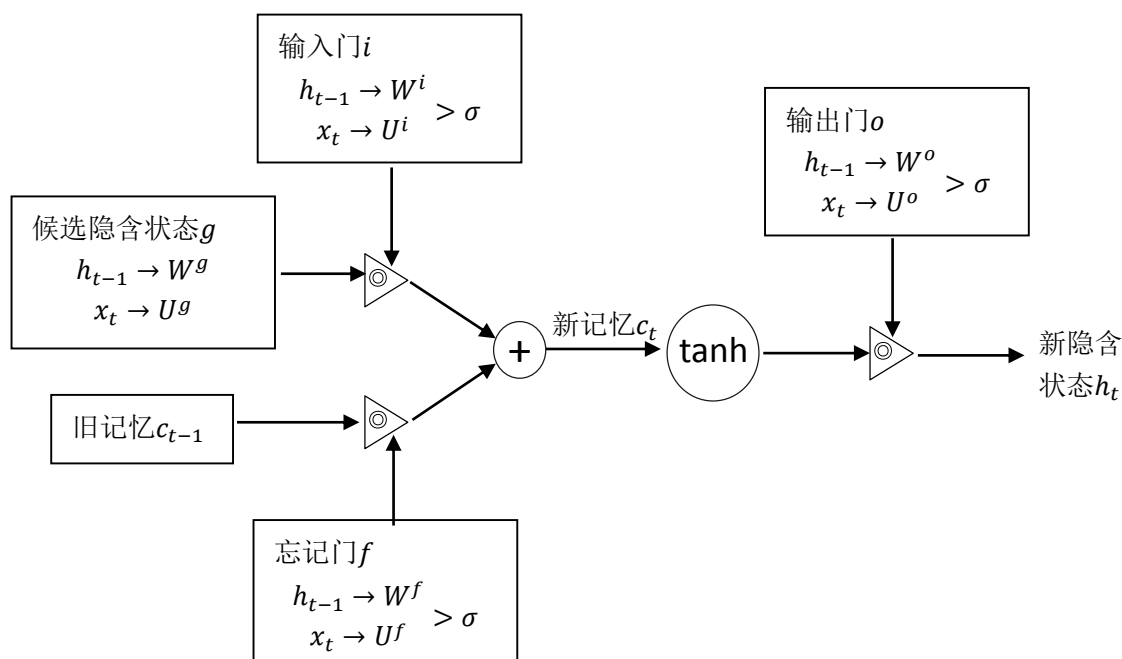
$$c_t = c_{t-1} \circ f + g \circ i$$

$$h_t = \tanh(c_t) \circ o$$

其中 x_t 代表输入层的输入, h_t 代表隐藏层中的激活神经元, c_t 为内容节点。 i, f, o 分别为输入门, 忘记门和输出门。这三个门有着相同的表达式, 但不同的矩阵系数 U 和 W , 他们的值还取决于当前的信息输入 x_t 和上一时间段的隐含状态 h_{t-1} 。他们被称为门是因为使用了 σ 即 sigmoid 函数, 取值在 0 至 1 的范围来控制门的开闭。 g 是一个候选隐含状态, 同样也取决于当前的信息输入 x_t 和上一时间段的隐含状态 h_{t-1} 。并非所有的隐含状态都会用于更新内容节点 c_t , 这由 g 与输入门 i 的各个元素相乘决定, 即 $g \circ i$ 。旧的内容节点 c_{t-1} 也会删除部分内容, 这由忘记门控制, 即 $c_{t-1} \circ f$ 。新的内容节点 c_t 就是新增内容部分与忘记后内容部分的结合。最后输出的隐含状态 h_t 由输出门 o 控制, 即 $\tanh(c_t) \circ o$ 。如果把忘记门 f 的所有值设为 0, 输

入门*i*的所有值设为 1，则全部内容节点的信息都会更新，同时输出门*o*设为 1，那么 LSTM 网络将回归到普通的循环神经网络。LSTM 的单元结构可以用以下图来表示。

图 4: LSTM 单元



数据来源：华泰期货研究院

图 4: 中的 LSTM 单元与图 2: 中循环神经网络的隐藏层相比多了三个信息门，参数大概增加为原来的 4 倍，所以也要求样本数量增加，同时训练难度也相应增加。于是在最近研究人员 Chung 和 Junyoung 等人又在 Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014 中提出了门控循环单元(Gated Recurrent Unit, GRU)的概念。

门控循环单元

门控循环单元与长短时记忆单元结构很像也是作用于循环神经网络的隐藏层，但是‘门’的数量从长短时记忆单元的输入门，忘记门和输出门三个门压缩为更新门和重置门两个门。其关系式可如下表示

$$z = \sigma(x_t U^z + h_{t-1} W^z)$$

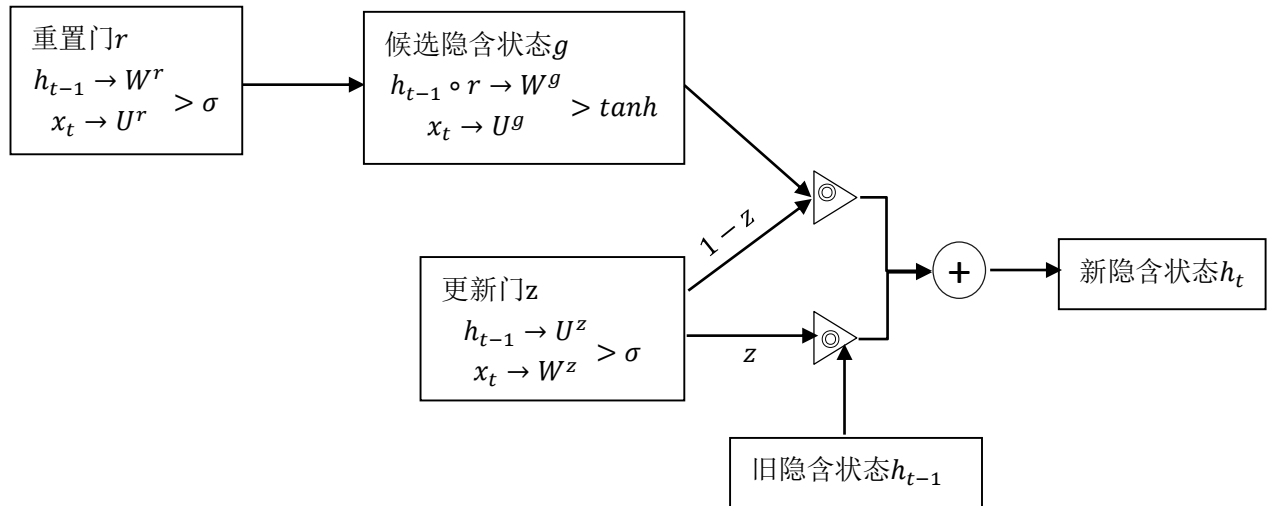
$$r = \sigma(x_t U^r + h_{t-1} W^r)$$

$$g = \tanh(x_t U^g + (h_{t-1} \circ r) W^g)$$

$$h_t = (1 - z) \circ g + z \circ h_{t-1}$$

其中 z 和 r 分别为更新门和重置门，重置门 r 决定有多少上一时间点的隐含信息 h_{t-1} 进入候选隐含状态 g ，而更新门 z 新隐含状态 h_t 中有多少信息会被更新，多少旧信息 h_{t-1} 会被保留。如果把重置门 r 的所有值设为1，更新门 z 的所有值设为0，那就回到了原始的循环神经网络状态。GRU的单元结构可以用以下图来表示。

图 5: GRU 单元



数据来源：华泰期货研究院

由图 5: 和 GRU 的公式组可以看出，除了控制门的数量减少以外，GRU 不像 LSTM 或一般循环神经网络那样具备内容节点 c_t ，GRU 的记忆内容完全贮存在隐含状态 h_t 当中。原来 LSTM 中的输入门和忘记门被合并成了更新门 z ，而重置门 r 则直接应用到之前的隐含状态 h_{t-1} 中。同时在计算新隐含状态 h_t 时并没有新的激活函数。对比 LSTM 和 GRU 的公式可以发现，由于少了一个控制门，GRU 的系数矩阵也相应减少。GRU 的系数大概是普通循环神经网络的 3 倍。

目前 GRU 这种形式的循环神经网络研究并不太多，但有初步研究认为 GRU 和 LSTM 模型在预测效果上差异并不大，而在很多机器学习任务中如何调整模型超参数其实比模型架构更为重要。显然，GRU 的模型参数更少所以训练 GRU 的时间以及样本数量也能相对少些。

GRU 模型应用实例

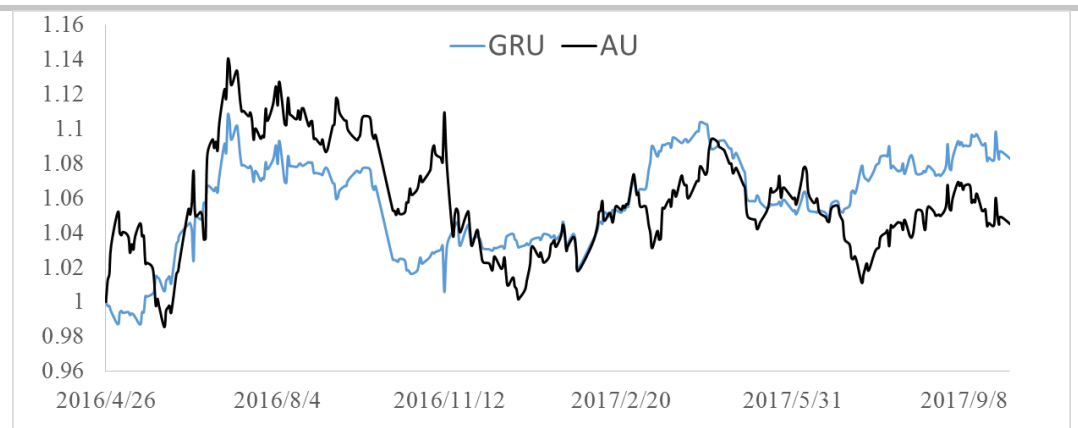
在这一节里尝试使用 GRU 单元构建循环神经网络从而预测期货主力合约的收益率。预测是基于期货品种的基本面因子包括商品库存、上下游产物价格, 相关进出口过汇率等, 以及市场因素如期货成交量、持仓量变化等。每个期货使用的市场因子数据有 20-30 个, 数据维度较大, 由于目前使用的是日频数据, 大部分期货基本面数据样本量只有 2000 个左右,

如果把高维度，少样本的数据直接放入神经网络中很容易造成过度拟合，因此这里采用了主成分分析(Principle Component Analysis, PCA)的方法进行降维减少直接输入神经网络输入层的变量数目，通常保留的主成份数量为 3-9 个。神经网络中隐藏层节点数目为 3-6 个，目前只使用了一个隐藏层，神经网络可调的参数控制在 100 个以下。神经网络的训练使用简单的交叉验证法，约 65%的样本用于训练集，15%的样本用于验证集，20%的样本用作测试集。训练数据中输入样本的时间序列长度通常为 50-200 个交易日，预测接下来 5 个交易日的收益率。训练集和验证集的时间序列样本在时间上为可获得样本的前 80%，训练时顺序为随机抽取，但是测试集的样本则按照正常交易日时间排序。神经网络训练时使用提早停止的方法，即验证集误差开始增大时停止。神经网络的超参数包括输入的时间序列长度，隐藏层神经元数量，学习速度和主成份个数等，由于神经网络训练耗时较长，这里并没有对超参数空间进行全局搜索，因此所使用的超参数并非最优。

目前该模型用于黄金、PTA 和菜粕三个期货主力合约，每个交易日收盘后对下 5 个交易日的收益率进行预测，收益为正则进场做多，反之则做空。由于持仓时间有 5 个交易日，所以存在 5 条不同路径，这里把资金分成 5 份，每条路径投入一份资金，相当于把这 5 条路径取平均收益。下面给出这三个品种在测试集上的表现，是纯样本外预测的效果。

图 6：在沪金期货上的表现。GRU 表示当前循环神经网络模型，AU 为沪金主力期货收益，数据是从 2016 年 4 月至 2017 年 10 月，虽然从图上看两条曲线的走势十分接近，但是以收益率计算的相关性却只有 0.27。尤其是在 2017 年 6 月前，两条曲线的累计收益并无太大区别，但在 2017 年 6 月后 GRU 模型却从黄金下跌的那波趋势中成功获利。

图 6： 黄金策略净值曲线



数据来源：华泰期货研究院

GRU 模型在黄金上的表现总结如下。该模型在黄金上的年化收益并不高，只有 5.29%，但是任然高于单纯做多黄金的收益，并且 GRU 模型的年化波动率要比黄金主力期货要低些。

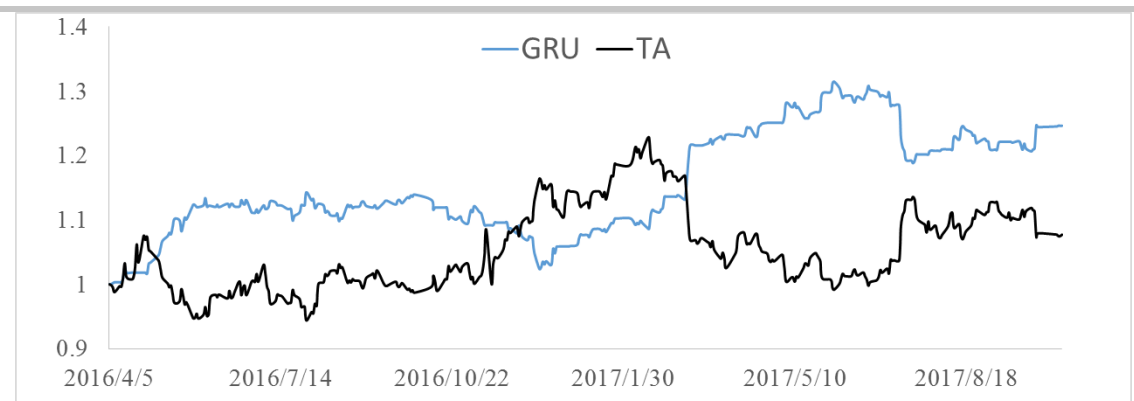
表 1: 黄金策略表现总结

	GRU	AU
年化收益	5.29%	2.92%
年化波动率	9.01%	12.43%
夏普率	0.59	0.24

数据来源: 华泰期货研究院

图 7: 是 GRU 在 PTA 上的表现, 从 2016 年 5 月底至 2016 年 10 月这段时间里虽然 PTA 主力期货有较大波动, 但是 GRU 的表现比较平稳。从 2016 年 12 月至 2017 年 7 月 PTA 主力期货在期间经历了从上涨到下跌的反转行情, 但是 GRU 模型在这段时间的表现大部分时间是稳步上涨, 而且中也从中间的暴跌中获利, 之后 GRU 的表现也回归平稳。

图 7: TA 策略净值曲线



数据来源: 华泰期货研究院

表 2 是 GRU 在 PTA 这个品种上的总结, GRU 的年化收益达到了 14.8%, 波动率为 10.3%, 明显低于主力合约的 19.25%, 而夏普率也达到了 1.35。由此看来在 PTA 这个品种上 GRU 的确体现了一定的预测效果。

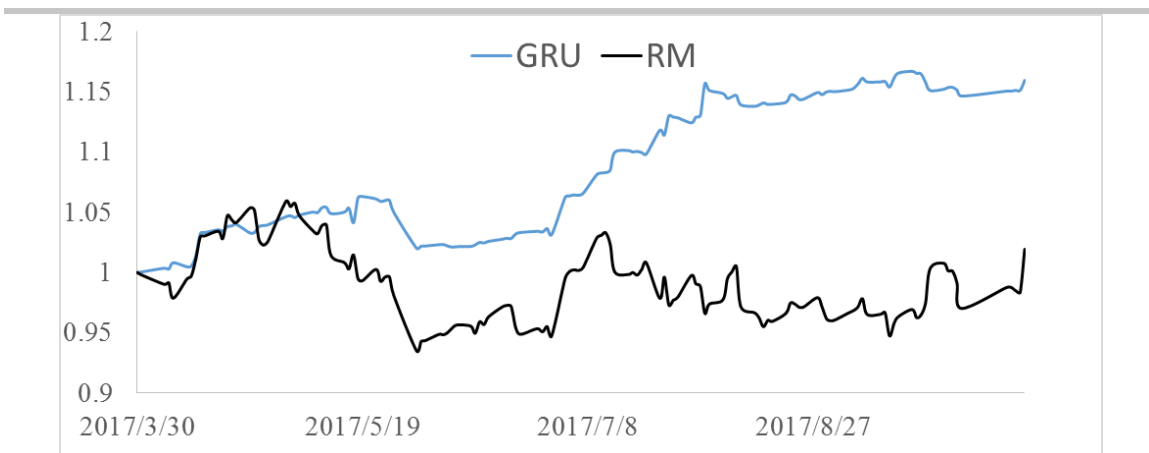
表 2: PTA 策略表现总结

	GRU	PTA
年化收益	14.80%	5.00%
年化波动率	10.93%	19.25%
夏普率	1.35	0.26

数据来源: 华泰期货研究院

图 8: 是 GRU 在菜粕上的表现, 由于菜粕是 2012 年底才上市的品种, 样本数量较少, 所以纯样本外预测数据只有半年左右。这段时间里 GRU 的表现相对前两个品种比较稳定, 大部分时间保持上升趋势, 只在 2017 年 5 月出现了较大回撤。

图 8: RM 策略净值曲线



数据来源: 华泰期货研究院

GRU 在菜粕上的表现总结如表 3 所示,在半年时间里 GRU 模型的年化收益达到 28.21% 波动率 9.52%, 与菜粕主力期货相比也是显著降低, 而夏普率则达到 2.96。

表 3: RM 策略表现总结

	GRU	RM
年化收益	28.21%	3.64%
年化波动率	9.52%	19.89%
夏普率	2.96	0.183

数据来源: 华泰期货研究院

从 GRU 模型在上面三个品种上的表现来看,该模型在不同品种上的表现并不一致,但是普遍来说,GRU 模型的波动率都要比相应的主力期货要低,而且普遍录得正的收益,说明 GRU 模型在预测上还是有一定的效果。但同时该模型的可调参数较多,要求较多的训练样本,虽然该模型在菜粕上短期内取得较好效果,但是其能否持续盈利还有待进一步考察。

结果讨论

本报告首先对循环神经网络模型进行了基本介绍,包括其原理、框架以及常见的变体。目前直接使用简单的循环神经网络已经较为少见,所以本报告更多的是介绍其变体 LSTM 和 GRU。从原理上来讲这两类模型更加接近人类的思维模式,也更加灵活,例如能有效地贮存信息。但是这些模型的训练也更加困难,模型超参数空间更大,往往需要在神经网络上有一定经验的研究人员进行手动调参或者利用庞大的计算资源对超参数空间进行搜索。在本报告研究的三个品种中 GRU 型循环网络的表现不一,部分品种例如 PTA 还是取得了较好的效果,但是这类模型的实用性仍需进一步探索。

● 免责声明

此报告并非针对或意图送发给或为任何就送发、发布、可得到或使用此报告而使华泰期货有限公司违反当地的法律或法规或可致使华泰期货有限公司受制于的法律或法规的任何地区、国家或其它管辖区域的公民或居民。除非另有显示，否则所有此报告中的材料的版权均属华泰期货有限公司。未经华泰期货有限公司事先书面授权下，不得更改或以任何方式发送、复印此报告的材料、内容或其复印本予任何其它人。所有于此报告中使用的商标、服务标记及标记均为华泰期货有限公司的商标、服务标记及标记。

此报告所载的资料、工具及材料只提供给阁下作查照之用。此报告的内容并不构成对任何人的投资建议，而华泰期货有限公司不会因接收人收到此报告而视他们为其客户。

此报告所载资料的来源及观点的出处皆被华泰期货有限公司认为可靠，但华泰期货有限公司不能担保其准确性或完整性，而华泰期货有限公司不对因使用此报告的材料而引致的损失而负任何责任。并不能依靠此报告以取代行使独立判断。华泰期货有限公司可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告反映编写分析员的不同设想、见解及分析方法。为免生疑，本报告所载的观点并不代表华泰期货有限公司，或任何其附属或联营公司的立场。

此报告中所指的投资及服务可能不适合阁下，我们建议阁下如有任何疑问应咨询独立投资顾问。此报告并不构成投资、法律、会计或税务建议或担保任何投资或策略适合或切合阁下个别情况。此报告并不构成给予阁下私人咨询建议。

华泰期货有限公司2016版权所有。保留一切权利。

● 公司总部

地址：广州市越秀区先烈中路65号东山广场东楼5、11、12层

电话：400-6280-888

网址：www.htgwf.com