

ForexRL项目规划与实施

-- Alvin 陈诺岩

1. 项目背景

1.1 强化学习简介

1.1.1 定义和关键概念

强化学习 (RL) 是机器学习的一个子领域，关注人工智能代理应如何在环境中采取行动以最大化累积奖励的概念。它涉及到连续决策的过程，每个决策都会引导到需要做出新决策。与有监督学习（模型训练时被提供正确答案）不同，RL通过奖励和惩罚在其问题空间中学习，与其环境互动。这赋予RL算法一种探索性的特征，使其适合解决事先未知解决方案的问题。

1.1.2 在人工智能中的重要性和当前应用

RL的最显著成就之一是AlphaGo，这是由Google DeepMind开发的一款电脑程序，击败了世界围棋冠军。这个成就凸显了RL解决复杂问题的能力和在众多领域（从游戏到金融，乃至更广泛领域）的潜在应用。

1.2 对Gym和Stable Baselines 3的解释

Gym和Stable Baselines 3是两个强大的库，对于开发和训练强化学习模型非常重要。

1.2.1 它们是什么以及为什么要使用

■ Gym:

Gym是OpenAI开发的一个开源库，旨在使强化学习算法的开发和对比更加简单和便捷。它提供了一个简单且通用的API，可以与各种环境进行交互。这些环境从简单的如'CartPole'到复杂的如'Atari'游戏，模拟了强化学习代理可以学习解决的不同场景。

主要使用Gym的两个原因：

- i. 它抽象化了许多强化学习的底层细节，使研究人员可以专注于创建和改进算法。
- ii. 它提供了大量的基准问题合集，使开发者可以将他们的算法性能与他人的进行比较，促进了强化学习研究社区的透明性和可复制性。

■ Stable Baselines 3:

Stable Baselines 3是一组高质量的强化学习算法在Python中的实现。它是Stable Baselines 2的继任者，提供了诸如近端策略优化（PPO），优势动作评论算法（A2C），柔性致动/评价算法（SAC）等先进的强化学习算法的实现。

使用Stable Baselines 3的原因：

- i. 提供可靠的，经过测试和优化的RL算法实现，减少了编码和调试的时间。
- ii. 为构建，训练和部署强化学习模型提供了更高级别的抽象，从而加快实验进程。

1.2.2 二者之间的关系

Gym提供了代理可以学习的环境，而Stable Baselines 3提供了让代理在这些环境中改善性能的学习算法。从本质上讲，Gym定义了代理将要玩的'游戏'和需要遵循的规则，而Stable Baselines 3则教导代理如何有效地玩这个游戏。

当一起使用时，这两个库提供了一个强大的框架用于开发，训练和测试强化学习模型。Gym的环境为代理与之交互提供了平台，而Stable Baselines 3的算法指导了代理的学习过程，使其能够随着时间的推移改善，并有可能在其设计解决的任务中达到出色的性能。

1.3 流行的RL算法

1.3.1 各种算法概述

强化学习算法可以分为三种主要类型：基于价值的方法、基于策略的方法和基于模型的方法。一些流行的算法包括深度Q网络（DQN）、演员-评论家方法（A2C, A3C）、信任区域策略优化（TRPO）和近端策略优化（PPO）。

1.3.2 近端策略优化（PPO）

我们关注的PPO，是一种策略优化方法，相对容易实现，并且已经被证明在各种任务中可以取得优越的结果。PPO在样本复杂性（即，它需要多少数据才能有效地学习）、实施的便利性和计算成本之间取得了平衡。与其他策略优化方法不同，PPO被设计为在一系列广泛的任务中具有强大的性能，只需很少的超参数调整，所以是我们项目的合适选择之一。

2. 在外汇交易中的强化学习

外汇交易领域是应用强化学习技术的沃土。这一部分将概述这些应用的动机，外汇交易固有的挑战以及强化学习如何为这些挑战提供解决方案。我们还将对外汇中的RL应用进行大致的概述，包括设置交易环境，数据预处理，以及在交易决策中平衡探索和利用的微妙行为。

2.1 在外汇交易中的应用动机

外汇交易世界的特点是快速变化，高波动性，以及巨大的盈亏潜力。在如此复杂，动态的环境中做出明智的交易决策，即使对于经验丰富的交易者也是具有挑战性的。许多人转向算法交易来帮助管理这些复杂性并增强他们的决策过程。

将强化学习（RL）应用于外汇交易的想法源于改进交易结果和效率的愿望。RL从环境中学习并随着时间改进其行动的能力与外汇交易的要求完全符合。考虑到大量的交易数据和外汇市场的非静态性质，一个能够学习并适应新数据的模型可能会显著改善交易结果。

2.2 外汇交易中的挑战以及RL如何提供帮助

外汇交易面临几个挑战，包括高波动性，实时决策的需要，以及全球事件对市场动向的影响。传统的交易算法可能无法有效地应对这些挑战。

强化学习可以帮助解决这些挑战：

- a. 适应性：RL模型旨在随着时间的推移进行学习和改进，使他们非常适合外汇市场的动态性质。他们可以根据新的数据和经验调整他们的交易策略，随着他们与环境的交互增加，他们的性能也会改善。
- b. 在不确定性下的决策制定：RL模型被训练在不确定的条件下做决策，这在外汇交易中是常见的情况。通过学习最优策略，他们即使在市场条件不确定或迅速变化时也能做出明智的交易决策。
- c. 平衡探索和利用：RL模型具有平衡探索（尝试新策略）和利用（坚持已知策略）的内在能力。这在外汇交易中至关重要，因为由于市场动态，坚持一种策略可能不会总是产生最好的结果。

2.3 在外汇交易中RL的应用概述

在外汇交易的背景下，RL代理与交易环境进行交互，该环境包括市场数据，当前的投资组合状态，以及任何其他相关信息。代理的目标是学习一个最优的策略，最大化预期的累积奖励，在这种情况下可能是交易的净利润。

2.3.1 基础交易环境设置

设置交易环境涉及定义状态空间，动作空间和奖励函数：

- 状态空间：状态空间通常由历史市场数据组成，包括价格（开盘，最高，最低，收盘），交易量和和其他技术指标。也可以包括当前的投资组合状态，如开放的仓位和现金余额。
- 动作空间：动作空间可以简单地买/卖/关闭操作，或者可以包括更复杂的交易操作，如限价和止损单。
- 奖励函数：奖励函数可能是交易的即时利润或损失，投资组合价值的变化，或其他更复杂的指标，旨在平衡利润和风险。

2.3.2 外汇交易数据预处理

外汇交易数据通常需要在用于RL之前进行预处理。这可能包括清理数据，处理缺失值，标准化，以及使用技术分析技术创建额外的特征。也可能需要对齐和重新采样不同时间框架的数据，为RL代理提供一致的输入。

2.3.3 在交易中平衡探索和利用

在RL中，代理需要平衡探索（尝试新的策略）和利用（坚持已知的成功策略）。这在外汇交易中尤其重要，因为市场条件可能会迅速变化。过多的探索可能导致过度的冒险，而过多的利用可能导致错过机会。如Epsilon-Greedy或上界信心区间（UCB）等RL算法可以帮助管理这种权衡。

通过理解应用RL在外汇交易中的这些关键方面，我们可以设计并实施一个基于RL的交易系统，该系统能够适应市场动态，随着时间的推移进行改善，并在风险和奖励之间找到平衡。

3. 行动和奖励空间

3.1 简单的买/卖/平仓

在交易强化学习代理的最简单形式的行动空间中，我们考虑三个基本操作：买入、卖出和平仓。代理可以买入单一的货币单位，卖出单一的货币单位，或者平掉任何现有的仓位。这样会导致行动空间大小为3。

- **Buy:** 这个行动涉及到代理购买一单位的货币对。这样做，代理采取"做多"的仓位，预期价格会上涨。
- **Sell:** 卖出涉及到代理抛售一单位的货币对。这里，代理采取"做空"的仓位，预测价格会下降。
- **Close:** 平仓行动表明代理关闭现有的仓位，无论是长期或者短期。通过平仓，代理锁定与该仓位相关的任何利润或损失。

3.2 扩展的买/卖/平仓

简单买/卖/平仓行动空间的扩展版本涉及到增加复杂性，扩大代理的行动范围。这可能涉及允许代理持有（即，不采取行动），甚至允许代理一次购买或卖出多个单位。

- **Hold:** 这是一个决定不交易的决定，通常是因为代理人不认为市场条件有利于盈利。
- **Multiple Units:** 代理人可能有选择一次购买或卖出多个单位的选项，这可以是固定数量的单位或者是总投资组合的一定百分比。

3.3 限价/止损买卖平仓

下一级的复杂性涉及到引入限价单和止损单的概念。这不仅允许代理决定是否购买或出售，还可以决定以什么价格进行。

- **Limit Order:** 限价单是一个以特定价格或更好的价格购买或出售的订单。买入限价单只能以限价或更低的价格执行，卖出限价单只能以限价或更高的价格执行。

- **Stop Order:** 止损单，也被称为止损订单，是一个在货币对价格达到指定价格（即止损价格）时购买或卖出的订单。当达到止损价格时，止损单变为市价单。

3.4 投资组合管理

在涉及多个货币对或其他金融工具的更复杂的场景中，每一个行动都可能代表分配给每种资产的投资组合的比例。这些比例需要加起来等于1，意味着投资组合价值的100%已经被分配出去。随着资产数量的增加，行动空间的大小可能会呈指数级增长。

3.5 多个仓位

在允许代理同时维持多个开放仓位的场景中，每一个行动都会涉及到不仅要决定执行哪种操作，还要决定在哪个仓位上进行。例如，如果代理有三个开放的仓位，它可以选择在一个仓位上购买更多，在另一个仓位上卖出一部分，并在第三个仓位上全部平仓，所有这些都在一个周期内完成。这引入了一层策略深度，但也大大扩大了行动空间。

上述所有方法都展示了在交易中使用强化学习定义行动空间的灵活性。根据在交易策略中所需要的复杂性和真实性的级别，人们可能会选择不同的行动空间定义。然而，值得注意的是，随着行动空间复杂性的增加，训练强化学习代理的挑战也随之增大。在现实性和可训练性之间进行权衡对于这种模型的成功部署至关重要。

4. Transformer模型

4.1 概念和工作原理

Transformer模型是由Vaswani等人在论文"Attention is All You Need"中提出的，主要用于自然语言处理（NLP）领域的一种模型架构。然而，其潜力超越了NLP领域，使其成为一系列任务，包括时间序列预测（对我们的外汇交易项目特别相关）的有价值的工具。

Transformer模型的关键概念在于其注意力机制，或更具体地说，其自注意力机制。自注意力机制使模型能够相对于其他输入对不同的输入进行权重分配和优先级排序，从而捕捉序列中不同元素之间的依赖关系，而不论它们彼此之间的距离如何。

在外汇交易问题的背景下，序列中的每个元素可以代表特定时间戳的历史市场数据，整个序列将代表一段时间内的这样的数据系列。通过利用自注意力机制，Transformer模型可以学习不同时间的市场事件之间的关系，从而为外汇市场的未来潜在走势提供宝贵的洞察。

4.2 优势和劣势

Transformer模型的主要优点之一是它能处理长序列的数据，这对于时间序列分析尤其有益。它的自注意力机制使其能够捕捉数据中的长期依赖性，使其成为当前输出依赖于前一次输出的任务的理想选择。此外，与RNN（递归神经网络）不同，Transformer模型不会出现梯度消失和爆炸的问题，使它们在训练过程中更稳定。

然而，Transformer模型也有其弱点。例如，它们可能在计算上很密集，可能需要大量的资源，特别是对于较长的序列。它们还需要大量的训练数据以确保良好的泛化。这意味着它们可能不适合数据稀疏的场景。

4.3 作为特征提取器的角色

在我们的强化学习项目中，我们计划将Transformer模型作为外汇交易数据的特征提取器。模型的输入将是一系列历史市场数据，模型将从这些数据中提取相关特征，然后将这些特征用作强化学习代理的输入。

通过这种方式使用Transformer模型，我们实际上是让模型自己学习外汇市场数据的有意义的表示。Transformer模型提取的特征可以被看作是对市场动态的一种"理解"，然后可以被强化学习代理用来更好地决定何时买入、卖出或持有。

5. 项目路线图

5.1 短期目标

- a. **数据收集和预处理:** 我们的目标是收集过去十年中大量的XAUUSD外汇数据。收集后，我们将对这些数据进行预处理，使其适合训练我们的强化学习模型。
- b. **开发并训练RL代理:** 下一步是使用Stable Baselines 3库建立一个RL代理。模型的主要算法是近端策略优化（PPO）。代理将在预处理的数据上进行训练，奖励函数将鼓励有利可图的交易。
- c. **评估和优化:** 模型训练完后，将在单独的验证数据集上进行评估。这个过程将帮助我们评估代理的性能并确定优化的领域。
- d. **实施:** 一旦我们优化了代理，下一个目标就是将它实施到一个真正的交易系统中。这个系统将使用RL代理的建议在外汇市场上进行实时交易。

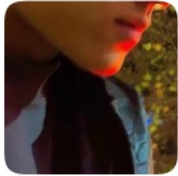
5.2 长期愿景

- a. **包含更多的金融市场:** 在成功训练和实施了XAUUSD交易的RL代理后，我们计划扩展到其他货币对，甚至可能扩展到其他类型的金融市场，包括股票和加密货币。
- b. **结合新闻数据:** 我们的目标是将新闻数据结合到模型中。我们计划从市场新闻API中提取相关信息，并将其输入到一个大型语言模型中。然后，这个模型将与初始的RL代理结合，以提高整体性能并做出更明智的交易决策。

5.3 潜在的扩展

- a. **多代理系统:** 我们可以探索开发一个多代理系统，其中多个RL代理一起进行交易，每个代理专门处理不同的市场条件或金融工具。这可能会提高系统的整体盈利能力和稳健性。
- b. **自适应奖励函数:** 在未来，我们可以研究开发一种根据市场的波动性或其他条件进行调整的自适应奖励函数。

- c. **集成其他深度学习模型:** 除了transformer模型，我们还可能尝试使用不同的深度学习架构，如循环神经网络（RNNs）或卷积神经网络（CNNs），以确定它们是否提供任何性能优势。
- d. **实时学习:** 一个重要的可能的扩展可能涉及从一个在历史数据上进行训练然后部署的模型，过渡到一个在新的市场数据到来时实时学习的模型。这将使模型能够更有效地适应市场的变化。



Alvin

Melbourne, Australia



Scan the QR code to add me as a friend.