ISSN 1009-3044
Computer Knowledge and Technology 电脑知识与技术
Vol.16, No.23, August. 2020

E-mail: xsjl@dnzs.net.cn
http://www.dnzs.net.cn
Tel:+86-551-65690963 65690964

深度强化学习股票算法交易系统应用

容梓豪

(广东工业大学,广东 广州 510000)

摘要:由于各种复杂的因素,股票价格预测与交易一直是难以有效处理的问题。现实世界不确定的因素太多,很难设计出用于自动股票交易的可靠算法。为了获得可行的交易策略,文中使用深度强化学习设计了一种方法,该方法在使用时间序列股票价格数据的基础上,加入了新闻标题进行观点挖掘,同时通过知识图来利用有关隐性关系的新闻,最后给出总结。

关键词:机器学习;强化学习;股价预测;新闻标签;知识图

中图分类号:TP389.1 文献标识码:A

文章编号:1009-3044(2020)23-0075-02

开放科学(资源服务)标识码(OSID):

DOI:10.14004/j.cnki.ckt.2020.2476

1引言

机器学习主要涉及根据数据构建预测模型,当数据是时间序列时,模型还可以预测序列和结果。预测股票市场的运作是近年来机器学习的一种应用,但是事实证明这项工作非常困难,因为参与预测的因素很多。以往机器学习在金融市场中使用主要试图通过诸如人工神经网络,支持向量机甚至决策树之类的监督学习来预测金融资产的未来回报。但是到目前为止,很多方法的效果都不太好。这其中有多种原因,例如,在有监督的机器学习中,通常使用带有平衡类分布的标记数据集。当涉及股票市场时,没有关于某人何时应该购买或出售其所持股票的此类标记数据,因此该问题适合强化学习框架,这是一种基于行为的学习,依赖于反复试验并辅以奖励机制。一旦定义了适当的奖励信号,强化学习便能够生成这种缺失的标签。但是在这种情况下,还有其他一些股票市场特有的问题。股票交易市场的变化非常频繁,不能仅从历史趋势中推断出这些变化,它们受到现实世界因素的影响,例如政治,社会甚至环境因素。在这种情况下,信噪比非常高,这会导致很难学到有意义的东西。可以将此类环境建模为部分可观察的马尔可夫决策过程,在该过程中,智能体对所有环境条件的可见性均有限。对智能体决策过程进行建模,在该过程中,假定系统是由离散时间随机控制过程确定的,但是智能体无法直接观察基础状态。该系统结合对交易公司及其相关新闻的观点挖掘,并结合强化学习算法,学习适当的策略来交易指定公司的股票。为了找到可以应用观点挖掘的相关新闻标题,使用知识图进行遍历,设计出一套强化学习交易方案,最后做出总结。

2 系统设计

该系统结合了来自不同领域的概念,因此,将对它们中的每一个步骤进行简要概述,并解释它们在系统中的使用方式。

2.1 Q-learning

Q-learning是一种基于值的强化学习算法,利用Q函数寻找最优的动作选择策略。它通过评估动作值函数应该选择哪个动作,这个函数决定了处于某一个特定状态以及在该状态下采取特定动作的奖励期望值。目标是最大化Q函数的值,即给定一个状态和动作时的未来奖励期望最大值,通过使用贝尔曼方程迭代地更新,不断优化策略。在给定环境的情况下,智能体会通过不断试错训练来学习一种策略,该策略会在episode结束时最大限度地从环境中获得总奖励。智能体试图了解处于某种状态并在该状态下采取特定行动,然后遵循到目前为止所学的行为策略直到episode结束的效用。因此,Q学习尝试学习每个状态和动作的动作值,它是通过同时探索和利用来实现的。但如果它始终遵循它认为可用的最佳选项,那么它将无法了解在该状态下采用其他可用选项的价值,这种困境被称为勘探与开发。解决此问题的一种简单但有效的方法是始终采用“贪婪”选项,除了在很小的一部分时间内随机行动,其他时候按照最优策略选择行为。

2.2 函数逼近

上述Q-learning方法存在一定缺点,它依赖于不同状态。注意,这时的值函数其实是一个表格。对于状态值函数,其索引是状态,对于行为值函数,其索引是状态—行为对。值函数迭代更新的过程实际上就是对这张表进行迭代更新。对于状态值函数,其表格的维数为状态的个数。若状态空间的维数很大,或者状态空间为连续空间,此时值函数无法用一张表格来表示,这不能很好地泛化,对于现实世界的问题也不容易处理。例如,根据当今世界的状况,股票交易智能体可能会做出一些决定并从中吸取教训,但是不太可能再次出现完全相同的状态。这时需要利用函数逼近的方法对值函数进行表示,该函数逼近当前的环境观测值,并且所选的动作会将它们映射为一个动作值。一旦观察到实际奖励,就可以类似于监督学习来更新近似器的参数。在本系统中使用人工神经网络进行函数逼近,对于大的状态空间,由于仅通过反向传播来优化人工神经网络变得不稳定,因此加入一些改良网络的技巧,修改适应于深度

收稿日期:2020-04-15

作者简介:容梓豪(1994—),男,广东江门人,硕士研究生,主要研究方向为深度学习和强化学习及其应用。

Q网络,这些修改包括经验重播,使用Q网络按比例更新独立目标网络的过程。

2.3 观点挖掘

观点挖掘用于注释预计表达正面或负面观点的文本,将文本分为正面和负面两类,文本的极性常用于分析产品或服务评论,例如网购,电影等,还用于分析其他书面文本,例如博客文章,新闻等。观点挖掘有两种主要类型,即使用具有极性的词汇词典的词汇分析方法;以及基于机器学习的,使用标记的训练数据集构建预测模型。通常,每个句子序列都具有正或负的含义,但有时是中立的。新闻标题、新闻全文本身,都会在某种程度上表达意见。自然语言处理技术被用来以自动化的方式提取这些观点。因此一旦提取出观点,它就可以用作一些重要数据点,以了解客户的意见。在该系统中,使用观点挖掘来评估新闻标题对交易股票的公司是否有利。从考虑买卖股票的公司的角度来看,每个新闻标题都被认为是正面,负面或中立的。积极的观点可以预测公司股价的总体上涨,而类似的消极观点则可以表明股价下跌。

2.4 知识图

词汇库和本体是与语义关系链接的术语数据库,用带有实体和关系的图形表示。知识库和知识图相对更复杂,其中的实体不是单纯的术语,而是知识的整合。一般来说Web搜索仅限于整个词汇库中在给定查询的字符串时匹配关键字,但现实中的实体相互有关联,并可以用不同的方式链接,因此普通的字符串匹配不能很好地进行智能搜索。这种互连以知识图为特征,该知识图表示类似图的数据结构,其中每个节点是一个实体,节点之间的边缘指示它们之间的关系。例如,仅使用简单的字符串匹配来简单地搜索“雷军”就不会出现小米。但是,使用知识图,由于“雷军”是小米的主要创始人,因此他是知识图中“小米”节点附近的一个紧密相关的节点,“小米”将作为一个相关的搜索结果出现,由此可将与公司相关但未在新闻标题中明确提及的实体识别为影响股价的潜在因素。连接的实体的标题将传递给观点挖掘,并在学习算法中利用其极性。

3 实证评估

3.1 数据

将股票的交易信息用于训练环境,即训练智能体以交易股票获取最大利益。对于新闻信息,从网站中爬取历史新闻头条,新闻标题的时间段与股票数据完全对应。对于每个新闻标题将其标记化,然后在预先确定的距离内的知识图中检查每个节点是否存在与所关注的特定公司之间的节点关系。选择距离长于此的距离会导致过多的噪音,而较短的距离意味着几乎找不到隐式关系,一旦发现标题中的所有标记都在预定距离之内就可以认为整个标题与公司相关。例如,谈论百度的新闻标题不会被视为影响股票价格的新闻,但是通过使用知识图,可以发现这种隐式关系。当找到与公司相关的头条新闻,就使用集成观点挖掘器进行观点挖掘。将每个新闻标题分为正面新闻和负面新闻,并使用上述分类中的分类,选择置信度最高的一个。如果同一天的头条新闻不止一个,就取当天所有头条新闻中的大部分观点得分。

3.2 MDP设计

智能体与股票交易环境进行交互,随着越来越多的事件用作训练,该智能体会探索不同的策略并改进其现有策略。环境使用4个变量来描述每个状态:分别是智能体拥有的当前金额,智能体现有的库存数量,当天的开盘价,收盘价以及当天对公司的平均观点,使用相关段落中表达的观点并用知识图来评估标题的相关性。行动空间:智能体每天与环境进行交互,它每次的行动可以有购买股票,卖出股票和不买不卖,即保持原状。奖励:交易期结束时投资组合的净增加应导致正回报,而净亏损将导致负奖励,最后智能体进行多个周期的训练。

4 结论

通过实验,单纯使用RNN进行初始实验很难优化网络,这可能是由于数据中的噪声以及可能没有正确的超参数所致。具有RNN的网络花费了特别长的时间来训练,并且难以分析,因为无法提取网络隐藏状态下发生的事情。另外,股票交易机器人仅限于每天买卖单只股票,这很可能限制了它可以赚到的利润。在现实世界中,交易的频率要比日内交易的频率高得多。在知识图中,关系距离阈值保持有限,以便从新闻标题的角度限制添加到数据中的噪声。如果提供了一个带有加权节点的知识图,该节点可以判断所讨论的实体与正在交易的公司股票之间是否存在正向或负向关系,以更准确的方式利用更长的距离关系。通过知识图谱从新闻头条中提取实体之间的隐式关系,并利用这些相关新闻上的正面或负面观点挖掘来训练强化学习智能体。经过训练的强化学习智能体可以在产生的利润方面取得更好的结果。这样的整个流程是一种新颖的方法,并通过实验证明了其有效性。

参考文献:

[1] 高阳,陈世福,陆鑫.强化学习研究综述[J].自动化学报,2004,30(1):86-100.

[2] 郭潇逍,李程,梅俏竹.深度学习在游戏中的应用[J].自动化学报,2016,42(5):676-684.

[3] 常亮,张伟涛,古天龙,等.知识图谱的推荐系统综述[J].智能系统学报,2019,14(2):207-216.

[4] 文丹艳,马超群,王琨.一种多源数据驱动的自动交易系统决策模型[J].自动化学报,2018,44(8):1505-1517.

[5] 杜漫,徐学可,杜慧,等.面向情绪分类的情绪词向量学习[J].山东大学学报(理学版),2017,52(7):52-58,65.

[6] 姜娜,孔浩.在线股票交易系统的分析与设计[J].计算机光盘软件与应用,2013,16(13):274-275.

[7] Kaelbling L P,Littman M L,Moore A W.Reinforcement learning:a survey[J].Journal of Artificial Intelligence Research,1996,4(1):237-285.

[8] Hessel M,Modayil J,van Hasselt H,et al.Rainbow:combining improvements in deep reinforcement learning[EB/OL].[2019-12-20].https://arxiv.org/abs/1710.02298.

[9] Mahmood A R,Korenkevych D,Vasan G,et al.Benchmarking reinforcement learning algorithms on real-world robots[EB/OL].[2019-12-20].https://arxiv.org/abs/1809.07731.

[10] Gers F A,Schmidhuber J,Cummins F.Learning to forget:continual prediction with LSTM[J].Neural Computation,2000,12(10):2451-2471.

【通联编辑:唐一东】