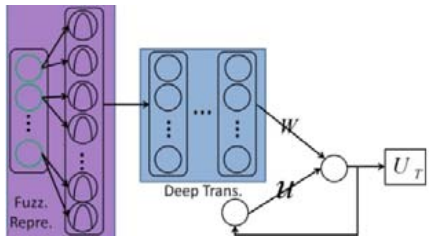

Fig.9 FRDNN framework

图 9 FRDNN 结构

在图 9 中,Fuzz Repre 代表模糊学习模块;而 Deep Trans 则代表使 DNN 的特征提取模块;W 代表 $\sum_{i=0}^{M} w_i k_{t-i}$;u 代表 $w_{M+1}F_{t-1}$;U_T 同于公式(4)的 p_T,代表时间 T 内的累计期望,即最大收益.Deng 分别在沪深 300 的期货交易数据和白银、白糖的商品期货分钟级别的高频数据上进行测试.实验结果表明,FRDNN 的收益极高,RRL 模型在某些交易上亏损非常严重.FRDNN 还与预测型 DNN 做了对比实验,分别使用 CNN,RNN,LSTM 在无交易成本时,DNN 模型的收益同 FRDNN 不相上下,一旦交易成本上升,DNN 模型的盈利能力迅速下降.可见:不能只注重模型预测能力,忽略交易成本,频繁交易的获利会被巨大的交易成本所吞没.这也进一步证明了 FRDNN 模型的合理性.同时,Deng 的实验中还对比了最高累计总利润和最高夏普比率分别作为目标函数时的收益情况.显而易见,最高夏普比率的模型收益明显要高,特别是在市场进入下行轨道时.

同样在 2017 年,Lu 等人发现,在文献[65]中使用 DNN 作为特征提取时常出现梯度消散问题,因此采用 LSTM 替换上述 DNN[66],并加入了 Dropout 技术来调试 LSTM 避免过拟合.Lu 在美元兑英镑的外汇交易数据上测试:首先,作者观察到公式(1)中的阈值 v 对交易频率和策略的影响,当 v 逐渐增大时,交易频率下降;之后,使用 LSTM 进行特征提取,并加入市场下行信号;最后,尝试使用下降偏差比率代替夏普比率作为损失函数.这些操作的结果都证明:在市场下行时,通过精确的做空,依然可以取得较高的交易利润.

文献[65,66]中可以看到:深度强化学习的算法应用在特征提取上,可以依靠确定性策略直接从采样特征中找寻下一次操作[67].无模型的策略搜索可以分为随机策略搜索方法和确定性策略搜索方法.2014 年以前,学者们都在发展随机策略搜索方法,直到 2014 年,Silver 提出了确定性策略理论[67].确定性策略意味着在应用策略函数 π_θ 时,在状态 s_t 下,下一步的动作 a 是确定的,即 $a=\pi_\theta(s_t)$.随机策略中,即使在相同的状态,每次采用的动作也很可能不一样.当然,当采用高斯策略的时候,相同的策略在同一个状态处,采样动作差别不大.确定性策略不需要像随机策略一样在空间进行大量采样.通常来说,确定性策略方法的效率比随机策略方法高 10 倍,这也是确定性策略方法最主要的优点.

2017 年,Jiang 等人将深度学和确定性策略应用在加密货币的投资组合中,通过将资金不断分配到不同的加密货币,获得更大累计收益[68].该系统包括独立评估集合(ensemble of identical independent evaluators,简称 EIIE)、投资组合内存(portfolio-vector memory,简称 PVM)、在线随机批量学习(online stochastic batch learning,简称 OSBL)和针对即时奖励的奖励函数.

Jiang 等人重新设计了 Actor-Critic 方法的状态、回报和动作,Actor 使用确定性策略梯度实现,Actor 的交易动作定义为下一个时间段 t 下各类资产分配的权重数值,用矢量 $w_t=\{x_1,\ldots,x_i\}$ 表示,x_i 的和为 1,见公式(40).

$$a_t=w_t \tag{40}$$

状态 s_t 则由当前时刻的价格张量 X_t(由最高价、最低价、收盘价组成)和前一时刻的资产分配权重 w_{t-1} 组成,见公式(41).

$$s_t=(X_t,w_{t-1}) \tag{41}$$

回报则用收益率的对数回报率表示.Jiang 采用深度神经网络作为确定性策略梯度函数 π_θ,并测试了 CNN,RNN,LSTM 这 3 个模型.例如,用 CNN 模型对输入特征(X_t,w_{t-1})进行采样,直接用 *softmax* 层的输出作为权重分

配值 w_{t-1},而在通常的分类任务中,常取 *softmax* 的最大值作为分类答案.同时,在训练过程中,依靠投资组合内存(portfolio-vector memory,简称 PVM)和小批量训练这两种机制进行训练.PVM 与强化学习的 DQN 经验回放机制非常相似:首先,通过引入外部存储机制,存储数据不断加入到训练数据中,使得训练数据尽量满足均衡分布,避免过拟合;然后,用小批量数据训练,每个批次内的数据必须是完整时间序列.对神经网络训练而言,即使它们具有显着重叠的间隔,不同时期的数据依然被认为是独特而有效的.这个系统依托在线随机批量学习方式,可以直接应用到在线上项目.在模型对比中,CNN,RNN 和 LSTM 占据了前三名,在比特币的虚拟交易中,即便在佣金率高达 0.25%的情况下,该系统仍然能够在 50 天内使收益增长为原来的 4 倍.

综上所述,深度强化学习在金融交易系统中的应用已经越来越多.随着深度强化学习在 2014 年后的强势兴起,带动了新一轮研究热潮.从模型结构上看,深度学习与强化学习的结合方式多种多样,在不同的应用领域各有优势:在单资产投资中,借助深度学习提取特征的 RRL 学习方法有效性依然很高,依托不同的目标函数应对不同的市场风格变化;而在资产组合交易中,基于策略搜索的深度强化学习方法显得更加灵活,状态和动作设计也不受模型局限.

7 结 论

本文综述了强化学习在金融交易领域的应用进展情况,包括 RRL、*Q* 学习、Actor-Critic、A3C 算法和结合深度神经网络的各类强化学习算法;以及依托强化学习构建的各类金融交易系统,在股票、指数、期货、投资组合、虚拟货币等交易领域的应用,基于强化学习的各类金融交易系统在风险控制、交易进出场时机、资金管理等方面都取得了突破.

基于强化学习将促进自动交易系统的进一步发展,可预见的趋势至少有两个方面.

(1) 经典的 RRL 模型将继续发展,但是 RRL 基于循环的自适应框架将会得到保留.在目标函数的选择上将变得更加灵活多样,在金融资产序列的特征提取上将更多地采用深度学习模型;

(2) 随着 A3C 算法的进一步发展,产业界与学术界将目光投向多智能体并行处理的方式,A3C 是在策略(on policy)算法,效果、时间和资源消耗上都优于 DQN 和 DDPG,它的应用有望部分解决强化学习策略受到的限制.

本文认为,上述研究中仍然存在着亟待解决的问题.

(1) 金融市场具有不稳定性,趋势实时变化.从历史的训练数据中学到的知识可能不会在后续测试数据中有良好的效果,这对强化学习模型的适应性提出了更高的要求,不同市场条件下如何选择合适的强化学习模型和深度学习模型仍然是一个悬而未决的问题;

(2) 构建基于强化学习的交易软件或系统.通常,一种算法不能解决全部问题,针对不同的市场情况,需要设置不同的配置模块.风险层、策略轮动层、自适应层等层次结构的设计至今没有统一解决方案,业界仍然在探索中;

(3) 大部分强化学习模型系统都是专攻某一类金融交易,单纯地做多、做空或空仓观望等,投资组合方式也仅是对各类金融资产的权重进行重新分配.但是,如股票中性、期货中性等策略需要对多种资产同时进行复杂的多空对冲操作时,仍缺少充分的研究;

(4) 强化学习领域最近提出了确定性策略和蒙特卡罗树搜索结合的算法,并应用于围棋领域[69],获得了突破.如何将蒙特卡罗树搜索策略应用在交易系统中,值得深入研究.

最后还要强调,深入研究强化学习理论、完善金融交易系统的组成结构、在提高交易的利润的同时降低交易风险,这是基于强化学习的金融交易系统研究的核心问题.

References:

[1] Fama Eugene F. Random walks in stock market prices. Financial Analysts Journal, 1965,21(5):55−59.

[2] Farmer JD. Market force, ecology and evolution. Computing in Economics & Finance, 1998,11(5):895−953(59). [doi: 10.1093/icc/11.5.895]

[3] Lo AW. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. Social Science Electronic Publishing, 2004. [doi: 10.3905/jpm.2004.442611]

[4] Lo AW. Reconciling efficient markets with behavioral finance: The adaptive markets hypothesis. Journal of Investment Consulting, 2005. http://ssrn.com/abstract=728864

[5] Sutton RS, Barto AG. Introduction to Reinforcement Learning. Vol.135. Cambridge: MIT Press, 1998. http://legacydirs.umiacs.umd.edu/~hal/courses/2016F_RL/RL9.pdf

[6] Kuleshov V, Precup D. Algorithms for the multi-armed bandit problem. Journal of Machine Learning Research, 2000,1:1−48. http://cn.arxiv.org/pdf/1402.6028

[7] Moody J, Saffell M. Reinforcement learning for trading. In: Proc. of the Conf. on Advances in Neural Information Processing Systems II. MIT Press, 1999. 917−923.

[8] Moody J, Wu L, Liao Y, Saffell M. Performance functions and reinforcement learning for trading systems and portfolios. Journal of Forecasting, 1998,17(5-6):441−470. [doi: 10.1002/(sici)1099-131x(1998090)17:5/6<441::aid-for707>3.3.co;2-r]

[9] Moody J, Saffell M. Learning to trade via direct reinforcement. IEEE Trans. on Neural Networks, 2001,12(4):875−889. [doi: 10.1109/72.935097]

[10] Gold C. FX trading via recurrent reinforcement learning. In: Proc. of the IEEE Int'l Conf. on Computational Intelligence for Financial Engineering. IEEE, 2003. 363−370. [doi: 10.1109/cifer.2003.1196283]

[11] Gorse D. Application of stochastic recurrent reinforcement learning to index trading. In: Proc. of the Esann 2011, European Symp. on Artificial Neural Networks. Bruges: DBLP, 2011. http://pdfs.semanticscholar.org/e7aa/08a404bb879cae6fcb751394a29465078e56.pdf

[12] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006,313(5786):504−507. [doi: 10.1126/science.1127647]

[13] Zhang J, Maringer D. Indicator selection for daily equity trading with recurrent reinforcement learning. In: Proc. of the Conf. Companion on Genetic and Evolutionary Computation. ACM Press, 2013. 1757−1758. [doi: 10.1145/2464576.2480773]

[14] Zhang J, Maringer D. Using a genetic algorithm to improve recurrent reinforcement learning for equity trading. Computational Economics, 2016,47(4):551−567. [doi: 10.1007/s10614-015-9490-y]

[15] Werbos PJ. Advanced forecasting methods for global crisis warning and models of intelligence. General Systems Yearbook, 1977, 22(6):25−38.

[16] Bertsekas DP, Tsitsiklis JN. Neuro-dynamic programming: An overview. In: Proc. of the IEEE Conf. on Decision and Control. IEEE, 1995. 560−564. [doi: 10.1109/cdc.1995.478953]

[17] Lewis FL, Vrabie D. Reinforcement learning and adaptive dynamic programming for feedback control. IEEE Circuits and Systems Magazine, 2009,9(3):32−50. [doi: 10.1109/MCAS.2009.933854]

[18] Liu D, Wei Q. Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. IEEE Trans. on Neural Networks and Learning Systems, 2014,25(3):621−634. [doi: 10.1109/tnnls.2013.2281663]

[19] Zhao H, Wang B, Liao J, Wang H, Tan G. Adaptive dynamic programming for control: algorithms and stability. Communications & Control Engineering, 2013,54(45):6019−6022.

[20] Atiya AF, Parlos AG, Ingber L. A reinforcement learning method based on adaptive simulated annealing. In: Proc. of the 2003 IEEE Midwest Symp. on Circuits and Systems. IEEE, 2003. 121−124. [doi: 10.1109/mwscas.2003.1562233]

[21] Jangmin O, Lee J, Lee JW, Zhang BT. Adaptive stock trading with dynamic asset allocation using reinforcement learning. Information Sciences, 2006,176(15):2121−2147. [doi: 10.1016/j.ins.2005.10.009]

[22] Dempster MAH, Leemans V. An automated FX trading system using adaptive reinforcement learning. Expert Systems with Applications, 2006,30(3):543−552. [doi: 10.1016/j.eswa.2005.10.012]

[23] Bertoluzzo F, Corazza M. Making financial trading by recurrent reinforcement learning. In: Proc. of the Int'l Conf. on Knowledge-based and Intelligent Information and Engineering Systems. Berlin, Heidelberg: Springer-Verlag, 2007. 619−626. [doi: 10.1007/978-3-540-74827-4_78].

[24] Tan Z, Quek C, Cheng PYK. Stock trading with cycles: A financial application of ANFIS and reinforcement learning. Expert Systems with Applications, 2011,38(5):4741−4755. [doi: 10.1016/j.eswa.2010.09.001]

[25] Almahdi S, Yang SY. An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. Expert Systems with Applications, 2017,87:267−279. [doi: 10.1016/j.eswa.2017.06.023]

[26] Hamilton JD. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 1989, 57(2):357−384. [doi: 10.2307/1912559]

[27] Hamilton JD, Susmel R. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics, 1994, 64(1-2):307−333. [doi: 10.1016/0304-4076(94)90067-1]

[28] Gray SF. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics, 1996,42(1):27−62. [doi: 10.1016/0304-405x(96)00875-6]

[29] Maringer D, Ramtohul T. Regime-switching recurrent reinforcement learning for investment decision making. Computational Management Science, 2012,9(1):89−107. [doi: 10.1007/s10287-011-0131-1]

[30] Maringer D, Ramtohul T. Threshold recurrent reinforcement learning model for automated trading. In: Proc. of the Applications of Evolutionary Computation, Evoapplications 2010: Evocomnet, Evoenvironment, Evofin, Evomusart, and Evotranslog. Istanbul: DBLP, 2010. 212−221. [doi: 10.1007/978-3-642-12242-2_22]

[31] Maringer D, Ramtohul T. Regime-switching recurrent reinforcement learning in automated trading. In: Proc. of the Natural Computing in Computational Finance. Berlin, Heidelberg: Springer-Verlag, 2011. 93−121. [doi: 10.1007/978-3-642-23336-4_6]

[32] Maringer D, Zhang J. Transition variable selection for regime switching recurrent reinforcement learning. In: Proc. of the Computational Intelligence for Financial Engineering & Economics. IEEE, 2014. 407−413. [doi: 10.1109/cifer.2014.6924102]

[33] Wierstra D, Förster A, Peters J, Schmidhuber J. Recurrent policy gradients. Logic Journal of Igpl, 2010,18(2010):620−634. [doi: 10.1093/jigpal/jzp049]

[34] Baird L, Moore A. Gradient descent for general reinforcement learning. In: Proc. of the Conf. on Advances in Neural Information Processing Systems II. MIT Press, 1999. 968−974.

[35] Watkins CJCH. Learning from delayed rewards. Robotics & Autonomous Systems, 1989,15(4):233−235.

[36] Jaakkola T, Jordan MI, Singh SP. On the convergence of stochastic iterative dynamic programming algorithms. Neural Computation, 1993,6(6):1185−1201. [doi: 10.21236/ada276517]

[37] Tsitsiklis JN. Asynchronous stochastic approximation and *Q*-learning. Machine Learning, 1994,16(3):185−202. [doi: 10.1007/bf00993306]

[38] Watkins CJCH, Dayan P. Technical note: *Q*-learning. Machine Learning, 1992,8(3-4):279−292. [doi: 10.1007/978-1-4615-3618-5_4]

[39] Moore AW, Atkeson CG. Prioritized sweeping: Reinforcement learning with less data and less time. Machine Learning, 1993,13(1): 103−130. [doi: 10.1007/bf00993104]

[40] Mahadevan S, Maggioni M. Proto-value functions: A laplacian framework for learning representation and control in markov decision processes. Journal of Machine Learning Research, 2007,8:2169−2231. [doi: 10.1145/1102351.1102421]

[41] Sutton RS. Policy gradient methods for reinforcement learning with function approximation. Submitted to Advances in Neural Information Processing Systems, 1999,12:1057−1063.

[42] Lee JW, Jangmin O. A multi-agent *Q*-learning framework for optimizing stock trading systems. In: Proc. of the Int'l Conf. on Database and Expert Systems Applications. Springer-Verlag, 2002. 153−162. [doi: 10.1007/3-540-46146-9_16]

[43] Lee JW, Park J, Jangmin O, Lee J, Hong E. A multiagent approach to Q-learning for daily stock trading. IEEE Trans. on Systems Man & Cybernetics—Part A: Systems & Humans, 2007,37(6):864−877. [doi: 10.1109/tsmca.2007.904825]

[44] Li J, Chan L. Reward adjustment reinforcement learning for risk-averse asset allocation. In: Proc. of the IEEE Int'l Joint Conf. on Neural Network. 2006. 534−541. [doi: 10.1109/ijcnn.2006.246728]

[45] Bertoluzzo F, Corazza M. Reinforcement learning for automatic financial trading: Introduction and some applications. Working Papers, 2012. [doi: 10.2139/ssrn.2192034]

[46] Bertoluzzo F, Corazza M. Testing different reinforcement learning configurations for financial trading: Introduction and applications. Procedia Economics & Finance, 2012,3(338):68−77. [doi: 10.1016/s2212-5671(12)00122-0]

[47] Corazza M, Bertoluzzo F. *Q*-learning-based financial trading systems with applications. Social Science Electronic Publishing, 2014. [doi: 10.2139/ssrn.2507826]

[48] Du X, Zhai JJ, Lv KP. Algorithm trading using *q*-learning and recurrent reinforcement learning. 2016. http://cs229.stanford.edu/proj2009/LvDuZhai.pdf

[49] Eilers D, Dunis CL, von Mettenheim HJ, Breitner MH. Intelligent trading of seasonal effects: A decision support algorithm based on reinforcement learning. Decision Support Systems, 2014,64:100−108. [doi: 10.1016/j.dss.2014.04.011]

[50] Konda V. Actor-critic algorithms. Siam Journal on Control & Optimization, 1999,42(4):1143−1166. http://papers.nips.cc/paper/1786-actor-critic-algorithms.pdf

[51] Li H, Dagli CH, Enke D. Short-term stock market timing prediction under reinforcement learning schemes. In: Proc. of the IEEE Int'l Symp. on Approximate Dynamic Programming and Reinforcement Learning. IEEE, 2007. 233−240. [doi: 10.1109/adprl.2007.368193]

[52] Bekiros SD. Heterogeneous trading strategies with adaptive fuzzy actor—Critic reinforcement learning: A behavioral approach. Journal of Economic Dynamics & Control, 2010,34(6):1153−1170. [doi: 10.1016/j.jedc.2010.01.015]

[53] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D. Playing atari with deep reinforcement learning. Computer Science, 2013.

[54] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J. Human-level control through deep reinforcement learning. Nature, 2015, 518(7540):529. [doi: 10.1038/nature14236]

[55] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa, Y. Continuous control with deep reinforcement learning. Computer Science, 2015,8(6):A187.

[56] Mnih V, Badia AP, Mirza M, Graves A, Lillicrap TP, Harley T. Asynchronous methods for deep reinforcement learning. 2016.

[57] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proc. of the 26th Annual Conf. on Neural Information Processing Systems. Nevada, 2012. 1097−1105. [doi: 10.1145/3065386]

[58] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S. Image net large scale visual recognition challenge. Int'l Journal of Computer Vision, 2015,115(3):211−252. [doi: 10.1007/s11263-015-0816-y]

[59] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks. In: Proc. of the IEEE Conf. on Acoustics, Speech and NAL Processing. Vancouver, 2013. 6645−6649. [doi: 10.1109/icassp.2013.6638947]

[60] Li YX, Zhang JQ, Pan D, Hu D. A study of speech recognition based on RNN-RBM language model. Journal of Computer Research a Development, 2014,51(9):1936−1944 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2014.20140211]

[61] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing. Doha, 2014. 1724−1734. [doi: 10.3115/v1/d14-1179]

[62] Yang Z, Tao DP, Zhang SY, Jin LW. Similar handwritten Chinese character recognition based on deep neural networks with big data. Journal on Communications, 2014,35(9):184−189 (in Chinese with English abstract). [doi: 10.3969/j.issn.1000-436x.2014.09.019]

[63] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Li F. Large-scale video classification with convolutional neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Columbus, 2014. 1725−1732. [doi: 10.1109/cvpr.2014.223]

[64] Sun ZJ, Xue L, Xu YM, Wang Z. Overview of deep learning. Application Research of Computers, 2012,29(8):2806−2810 (in Chinese with English abstract). [doi: 10.3969/j.issn.1001-3695.2012.08.002]

[65] Deng Y, Bao F, Kong Y, Ren Z, Dai Q. Deep direct reinforcement learning for financial signal representation and trading. IEEE Trans. on Neural Networks and Learning Systems, 2017,28(3):653−664. [doi: 10.1109/tnnls.2016.2522401]

[66] Lu DW. Agent inspired trading using recurrent reinforcement learning and LSTM neural networks. Papers, 2017. https://arxiv.org/pdf/1707.07338.pdf

[67] Silver D, Lever G, Heess N, Degris T, Wierstra D, Riedmiller M. Deterministic policy gradient algorithms. In: Proc. of the Int'l Conf. on Machine Learning. 2014. 387−395.

[68] Jiang ZY, Xu DX, Liang JJ. A deep reinforcement learning framework for the financial portfolio management problem. arXiv preprint arXiv:1706.10059, 2017. https://arxiv.org/abs/1706.10059

[69] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A. Mastering the game of Go without human knowledge. Nature, 2017,550(7676):354−359. [doi: 10.1038/nature24270]

附中文参考文献:

[60] 黎亚雄,张坚强,潘登,等.基于 RNN-RBM 语言模型的语音识别研究计算机研究与发展, 2014,51(9):1936−1944.

[62] 杨钊,陶大鹏,张树业,等.大数据下的基于深度神经网的相似汉字识别.通信学报,2014,35(9):184−189.

[64] 孙志军,薛磊,许阳明,等.深度学习研究综述.计算机应用研究,2012,29(8):2806−2810.

梁天新(1984－),男,黑龙江齐齐哈尔人,博士生,CCF 学生会员,主要研究领域为自然语言处理,深度学习,机器学习,强化学习.

王良(1963－),男,博士,副教授,CCF 高级会员,主要研究领域为智能科学,数据库管理系统,数据库系统评价和性能优化.

杨小平(1956－),男,博士,教授,博士生导师,主要研究领域为信息系统工程,电子政务,网络安全技术.

韩镇远(1993－),男,硕士生,主要研究领域为深度学习,自然语言处理.