The mean reversion strategy [20] considers asset prices always return to their mean over a past period, so it buys assets with a price under their historical mean and sells above the historical mean. The multi-factor model [7] uses factors to compute a valuation for each asset and buys/sells those assets with price under/above their valuations. Most of these financial investment strategies can only exploit a certain factor of financial markets and thus might fail in complex market environments.

Deep Learning in Finance: In recent years, deep learning approaches begin to be applied in the financial areas. In the literature, L. Zhang *et al.* proposed to exploit frequency information to predict stock prices [11]. News and social media were used in price prediction in Refs. [12, 27]. Information about events and corporation relationships were used to predict stock prices in Ref. [2, 4]. Most of these works focus on price prediction rather than end-to-end investment portfolio generation like us.

Reinforcement Learning in Finance: The RL approaches used in investment strategies fall in two categories: the value-based and the policy-based [8]. The value-based approaches learn a critic to describe the expected outcomes of markets to trading actions. Typical value-based approaches in investment strategies include Q-learning [19] and deep Q-learning [16]. A defect of value-based approaches is the market environment is too complex to be approximated by a critic. Therefore, policy-based approaches are considered as more suitable to financial markets [8]. The AlphaStock model also belongs to this category. A classic policy-based RL algorithm in investment strategy is the Recurrent Reinforcement Learning (RRL) [17]. The FDDR [3] model extends the RRL framework using deep neural networks. In the Investor-Imitator model [6], a policy-based deep RL framework was proposed to imitate the behaviors of different types of investors. Compared with RRL and its deep learning extensions, which focus on exploiting sequential dependence in financial signals, our AlphaStock model pays more attention to the interrelationships among assets. Moreover, deep RL approaches are often hard to deployed in real-life applications for unexplainable deep network structures. The interpretation tools offered by our model can solve this problem.

7 CONCLUSIONS

In this paper, we proposed a RL-based deep attention network to design a BWSL strategy called AlphaStock. We also designed a sensitivity analysis method to interpret the investment logics of our model. Compared with existing RL-based investment strategies, AlphaStock fully exploits the interrelationship among stocks, and opens a door for solving the "black box" problem of using deep learning models in financial markets. The back-testing and simulation experiments over U.S. and Chinese stock markets showed that AlphaStock performed much better than other competing strategies. Interestingly, AlphaStock suggests buying stocks with high long-term growth, low volatility, high intrinsic value, and being undervalued recently.

ACKNOWLEDGMENTS

J. Wang's work was partially supported by the National Natural Science Foundation of China (NSFC) (61572059, 61202426), the Science and Technology Project of Beijing (Z181100003518001), and the CETC Union Fund (6141B08080401). Y. Zhang's work was partially supported by the National Key Research and Development Program of China under Grant (2017YFC0820405) and the Fundamental Research Funds for the Central Universities. K. Tang's work was partially supported the National Social Sciences Foundation of China (No.14BJL028). J. Wu's work was partially supported by NSFC (71725002, 71531001, U1636210).

REFERENCES

[1] Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. 2018. Sanity checks for saliency maps. In *NIPS'18*. 9525–9536.

[2] Yingmei Chen, Zhongyu Wei, and Xuanjing Huang. 2018. Incorporating Corporation Relationship via Graph Convolutional Neural Networks for Stock Price Prediction. In *CIKM'18*. ACM, 1655–1658.

[3] Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. 2017. Deep direct reinforcement learning for financial signal representation and trading. *IEEE TNNLS* 28, 3 (2017), 653–664.

[4] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction.. In *IJCAI'15*. 2327–2333.

[5] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2016. Knowledge-driven event embedding for stock prediction. In *COLING'16*. 2133–2142.

[6] Yi Ding, Weiqing Liu, Jiang Bian, Daoqiang Zhang, and Tie-Yan Liu. 2018. Investor-Imitator: A Framework for Trading Knowledge Extraction. In *KDD'18*. ACM, 1310–1319.

[7] Eugene F Fama and Kenneth R French. 1996. Multifactor explanations of asset pricing anomalies. *J. Finance* 51, 1 (1996), 55–84.

[8] Thomas G Fischer. 2018. *Reinforcement learning in financial markets-a survey.* Technical Report. FAU Discussion Papers in Economics.

[9] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 93.

[10] Sepp Hochreiter and Jurgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[11] Hao Hu and Guo-Jun Qi. 2017. State-Frequency Memory Recurrent Neural Networks. In *ICML'17*. 1568–1577.

[12] Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. 2018. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *WSDM'18*. ACM, 261–269.

[13] Dingjiang Huang, Junlong Zhou, Bin Li, Steven CH Hoi, and Shuigeng Zhou. 2016. Robust median reversion strategy for online portfolio selection. *IEEE TKDE* 28, 9 (2016), 2480–2493.

[14] Narasimhan Jegadeesh and Sheridan Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance* 48, 1 (1993), 65–91.

[15] Narasimhan Jegadeesh and Sheridan Titman. 2002. Cross-sectional and time-series determinants of momentum returns. *RFS* 15, 1 (2002), 143–157.

[16] Olivier Jin and Hamza El-Saawy. 2016. *Portfolio Management using Reinforcement Learning.* Technical Report. Stanford University.

[17] John Moody, Lizhong Wu, Yuansong Liao, and Matthew Saffell. 1998. Performance functions and reinforcement learning for trading systems and portfolios. *Journal of Forecasting* 17, 5-6 (1998), 441–470.

[18] Tobias J Moskowitz, Yao Hua Ooi, and Lasse Heje Pedersen. 2012. Time series momentum. *J. Financial Economics* 104, 2 (2012), 228–250.

[19] Ralph Neuneier. 1995. Optimal Asset Allocation using Adaptive Dynamic Programming. In *NIPS'95*.

[20] James M Poterba and Lawrence H Summers. 1988. Mean reversion in stock prices: Evidence and implications. *J. Financial Economics* 22, 1 (1988), 27–59.

[21] William F Sharpe. 1994. The sharpe ratio. *JPM* 21, 1 (1994), 49–58.

[22] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to Sequence Learning with Neural Networks. *NIPS'14* (2014), 3104–3112.

[23] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction.* MIT press.

[24] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS'17*. 5998–6008.

[25] Jingyuan Wang, Qian Gu, Junjie Wu, Guannan Liu, and Zhang Xiong. 2016. Traffic speed prediction and congestion source exploration: A deep learning method. In *ICDM'16*. IEEE, 499–508.

[26] Jingyuan Wang, Ze Wang, Jianfeng Li, and Junjie Wu. 2018. Multilevel wavelet decomposition network for interpretable time series analysis. In *KDD'18*. ACM, 2437–2446.

[27] Yumo Xu and Shay B Cohen. 2018. Stock movement prediction from tweets and historical prices. In *ACL'18*, Vol. 1. 1970–1979.