Reinforcement Learning for Computational Finance

Yuxi Li YUXI@CS.UALBERTA.CA
Csaba Szepesvari SZEPESVA@CS.UALBERTA.CA
Dale Schuurmans DALE@CS.UALBERTA.CA
Department of Computing Science, University of Alberta, Edmonton, AB, Canada, T6G 2E8

Abstract

We give a brief overview of the application of reinforcement learning to problems in computational finance, focusing on option pricing, portfolio optimization, and the adaptive market hypothesis. We argue that computational finance is a fruitful area for further development of reinforcement learning methods.

1. Introduction

As a method for solving large-scale Markov Decision Processes (MDPs), *reinforcement learning* (RL) (Bertsekas and Tsitsiklis 1996; Sutton and Barto 1998; Powell 2007) has attracted significant attention from the academic and industrial communities in science, engineering and economics. Computational finance is a particularly attractive area for application of RL methods, which has not been a central focus of the field. We argue that computational finance can offer one of the most influential areas of application for RL.

Computational finance is a discipline that is rooted in finance, computer science and mathematics (Hull 2006; Luenberger 1997). Of the many key issues it raises, derivative pricing, portfolio optimization, and the adaptive market hypothesis are particularly amenable to an RL treatment. The values of financial derivatives, such as options, depend on the values of underlying assets. An option, in particular, gives the holder the right, but not the obligation, to buy or sell an asset at a certain price by a certain time. Portfolio optimization, on the other hand, considers how to allocate assets so as to trade off between return and risk. RL has natural applications in both these areas. The Adaptive Market Hypothesis (AMH) attempts to reconcile the Efficient Market Hypothesis and behavioral finance by asserting that the financial market adapts itself to a market equilibrium. RL can also provide a useful formalization of this process.

2. Option pricing

Options (Hull 2006; Glasserman 2004) are fundamental financial instruments with ancient origins. A central challenge is to determine the value of an option. For European options, which can only be exercised at the maturity date, prices can be calculated by the Black-Scholes formula in certain cases. However, American type options can be exercised before the maturity date, hence the key problem is to determine the conditional expected value of continuation. This is an optimal stopping problem that is a special case of an MDP and thus can be solved by RL methods. In the finance literature, least squares Monte Carlo (LSM) (Longstaff and Schwartz 2001) is a standard approach for pricing American options. However, in Li et al. (2009), we have applied a standard RL method, least squares policy iteration (LSPI) (Lagoudakis and Parr 2003), to this problem and compared it to LSM and fitted Q-iteration (FQI) (Tsitsiklis and Van Roy 2001). Our results show that LSPI and FQI both outperform LSM. We have also derived a high-probability, finite-time bound on the performance of LSPI for this problem. In this case, RL methods offer solution methods that challenge the state of the art.

3. Portfolio optimization

Mean-variance analysis by Markowitz is a classical approach to portfolio optimization in one period (Luenberger 1997). The problem of dynamic portfolio optimization in multi-periods (Campbell and Viceira 2002; Brandt et al. 2005) has received renewed attention given recent empirical evidence of return predictability (Pastor and Stambaugh 2009). This problem entails consideration of parameter and model uncertainty, transaction cost, and background risks. Brandt et al. (2005) deploy the backward dynamic programming approach in Longstaff and Schwartz (2001) to address the dynamic portfolio problem. Nevertheless, it is possible to apply recent RL methods to this problem. For example, Moody and Saffell (2001) learn to trade via direct reinforcement, without explicit forecasting. However, it may be beneficial to take advantage of return predictability in RL methods.

Another central issue is controlling risk in forming portfolios. Value-at-Risk (VaR) and conditional VaR (CVaR) are popular risk measures in the finance literature, where CVaR in particular has desirable mathematical properties (Hull 2006). Delage and Mannor (2009) show how to solve MDPs with parameter uncertainty with respect to the percentile risk measure, which is equivalent to VaR. Although operations research has also considered robust MDP planning, the generalization to continuous state and action spaces is an indispensable step for such methods to be applied to dynamic portfolio optimization, and this is an issue under current, active development in RL.

4. Adaptive Market Hypothesis

The two dominant schools of thought in finance are the Efficient Market Hypothesis (EMH) and Behavioral Finance (Lo 2004). According to the EMH, "prices fully reflect all available information" and are determined by market equilibrium. However, psychologists and economists have found a number of natural behavioral biases in human decision-making under uncertainty. For example, Amos Tversky and Daniel Kahneman demonstrate the phenomenon of loss aversion, wherein people tend to strongly prefer avoiding losses to acquiring gains. Lo (2004) has proposed the Adaptive Market Hypothesis to reconcile the EMH with behavioral finance, postulating that the market is engaged in an evolutionary process of competition, mutation, reproduction and natural selection. RL concepts can play an important role in formalizing this fundamental market paradigm. Additionally, the research led by Michael Kearns on studying market microstructure with RL methods also falls in this area.

5. Discussion

We claim that RL is an important and natural tool for computational finance, and claim further that computational finance remains a promising area of investigation for RL methods even given the current financial crisis. As elucidated by Andrew Lo in his written testimony on hedge funds for the House Oversight Committee: financial derivatives can help reduce risk; a good understanding of such instruments requires advanced training and methods; and there remains an insufficient supply of professionals with such capacity. We therefore assert that there is a need for greater attention to be paid by the RL community to this area.

References

Dimitri P. Bertsekas and John N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

Michael W. Brandt, Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud. A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *The Review of Financial Studies*, 18(3):831–873, Fall 2005.

John Y. Campbell and Luis M. Viceira. *Strategic Asset Allocation Portfolio Choice for Long-Term Investors*. Oxford University Press, 2002.

Erick Delage and Shie Mannor. Percentile optimization for markov decision processes with parameter uncertainty. *Operations Research*, to appear, 2009.

Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2004.

John C. Hull. *Options, Futures and Other Derivatives (6th edition)*. Prentice Hall, 2006.

Michail G. Lagoudakis and Ronald Parr. Least-squares policy iteration. *The Journal of Machine Learning Research*, 4:1107 – 1149, December 2003.

Yuxi Li, Csaba Szepesvari, and Dale Schuurmans. Learning exercise policies for american options. In *International Conference on Artificial Intelligence and Statistics (AISTATS09)*, April 2009.

Andrew W. Lo. The Adaptive Markets Hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management*, 30:15–29, 2004.

Francis A Longstaff and Eduardo S Schwartz. Valuing American options by simulation: a simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147, Spring 2001.

David G. Luenberger. *Investment Science*. Oxford University Press, 1997.

John Moody and Matthew Saffell. Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, 12(4):875–889, July 2001.

Lubos Pastor and Robert F. Stambaugh. Predictive systems: Living with imperfect predictors. *Journal of Finance*, (to appear), 2009.

Warren B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*. John Wiley and Sons, 2007.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

John. N. Tsitsiklis and Benjamin Van Roy. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12 (4):694–703, July 2001.