Uncertainty Decomposition for Risk Control in Automated Stock Trading

Anita Mahinpei

Abstract

Uncertainty estimation is a common method of risk control in reinforcement learning. However, most RL approaches to stock trading use other means of risk aversion such as the Sharpe Ratio or Financial Turbulence Index (FTI). This paper investigates the suitability of risk control via uncertainty estimation for stock trading problems by building on the uncertainty decomposition algorithm developed by Clements et al (2020). We examined the performance of the uncertainty based strategy on a set of volatile stocks and were able to improve upon the performance of a risk-controlled algorithm that uses FTI.

1. Introduction

Reinforcement learning (RL) is an area of machine learning concerned with learning via interaction; an agent has to find the optimal policy that maximizes cumulative rewards by interacting with the environment. One natural application of reinforcement learning is automated stock trading and portfolio management. A simple formulation of stock trading as an RL problem would aim to maximize the change in the portfolio value. This approach, however, will not be appropriate for all individuals. Depending on their financial circumstances, some individuals would prefer more conservative portfolios that do not maximize returns but have lower risk. Others might prefer high risk portfolios that have lower expected returns but have the potential, albeit with low probability, for much greater returns over time. Risk assessment is crucial for such portfolio customization. More generally, risk assessment can help improve the exploration exploitation trade-off by ensuring that exploratory actions are not too risky.

One approach to risk aversion could be to use a measure of uncertainty in the evaluated value functions. A more risk tolerant portfolio should be more likely to take actions with high uncertainty/variance in their expected return while more conservative portfolios should be less likely to select such actions even if the expected return is high. Avoiding actions with high overall uncertainty however, will limit exploration which is crucial to the success of RL algorithms; without exploration, RL algorithms can converge to suboptimal solutions. One way to balance the trade-off between risk control and exploration, is to disentangle epistemic and aleatoric uncertainties. Aleatoric uncertainty is inherent to the data and stems from stochasticity of the environment or the agent's observations of the environment, therefore it is irreducible. Epistemic uncertainty on the other hand, stems from limitations in the model or training data and can be reduced via further exploration. Once disentangled, the learning algorithm can use aleatoric uncertainty for risk control. When risk tolerance is low, states with high aleatoric uncertainty should be avoided since aleatoric uncertainty is an indirect measurement of risks associated with the stochasticity of the environment. On the other hand, the learning algorithm can more efficiently explore the state-space by using epistemic uncertainty, similar to how the variance of the posterior distribution in Thomposon Sampling allows for more efficient exploration (Clements et al., 2020).

This paper investigates the suitability of risk control via uncertainty estimation for stock trading problems by building on the uncertainty decomposition algorithm developed by Clements et al (2020). We examine the feasibility of the aforementioned strategy for risk control by testing its performance on a set of volatile stocks. Inherently, unstable stocks require better risk control in order to give good cumulative returns. Thus, if the uncertainty decomposition approach is suitable for risk control, it should be able to provide better cumulative returns when trading volatile stocks compared to an approach that does not incorporate risk aversion.

2. Related Works

2.1. Uncertainty Estimation in RL

Various approaches have been proposed for estimating uncertainty in different reinforcement learning algorithms for both model based and model free problems. Leno et al. use bootstrapped neural networks to estimate the variance in the Q-function and request human intervention when the variance/uncertainty is above a threshold (2020). Tamar et al. also use the mean-variance trade-off for risk assessment but in the context of policy gradient algorithms (2012). O'Donoghue et al. propose the Uncertainty Bellman Equation for propagating local uncertainty estimates to global estimates and suggest various techniques for approximating

local uncertainties (2018).

The aforementioned uncertainty estimation approaches do not attempt to disentangle aleatoric and epistemic uncertainties. Clements et al. however, tackle the problem of decomposing aleatoric and epistemic uncertainties in Deep Q-Learning. Rather than learning a single point estimate, they learn several quantiles of the Q-function distribution and use these quantiles to estimate aleatoric and epistemic uncertainties (2020). They show that their algorithm is capable of avoiding the cliff in a grid world and that the probability of finding the safer route could be tuned with their risk aversion parameter (2020).

2.2. RL Approaches to Stock Trading

Different RL techniques have previously been applied to stock trading and portfolio management. Neuneier was one of the first researchers to propose a deep Q-learning framework for stock trading. His original approach did not incorporate any form of risk control and was designed for trading a single asset (1996). Deep Q-learning has since been adapted to stock trading in various ways and tested against different non-RL based approaches. Sornmayura, for instance, examined the performance of deep Q-learning on stock trading compared to a buy-and-hold approach as well as experienced human traders and found better performance compared to buy-and-hold but no improvement over an experienced human trader (2019).

One of the main limitations of Q-learning based approaches to stock trading is that they only work with a discrete and finite action-space which is not practical for managing multi-asset portfolios. As a result, several actor-critic based algorithms that do not suffer from this limitation have more recently been applied to stock trading. For instance, Yang et al. developed an ensemble of three actor-critic algorithms: Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C), and Deep Deterministic Policy Gradient (DDPG); they found that the ensemble performed better than all the individual models in terms of the Sharpe ratio (2020). Although actor-critic algorithms are more suitable for stock trading, this paper uses deep Q-learning since the uncertainty decomposition method proposed by Clements et al. is for deep Q-learning. If the algorithm is found suitable for risk control in stock trading, future research can focus on extending uncertainty decomposition to actor-critic algorithms for financial applications.

3. Methodology

3.1. Problem Formulation

The stock trading problem examined in this paper is the management of a single-stock portfolio where the agent is limited to buying or selling a single share at each time step. The problem is formulated as a model-free Markov Decision Process (MDP) as follows:

State S_t: The state at time t is described by the agent's remaining balance, the number of shares owned, the most recent stock price, and a set of common market indicators from StockStats including: Moving Average Convergence Divergence, 30-day Relative Strength Index, 30-day Commodity Channel Index, Stock Bolling Bands, and Directional Movement Index.

Action A_t: At each time step t, the agent is limited to either buying a single share, selling one of their shares, or holding all their shares.

Reward R_t: The immediate reward obtained at time t, is the change in the total portfolio value (total stock value + remaining balance) at the end of a trading day.

3.2. Uncertainty Aware DQN Algorithm

In normal deep Q-learning, a deep neural network is used to approximate the action values for all possible actions given the current state. In uncertainty aware deep Q-learning, rather than getting a single point-estimate for the expected value of each action, the neural network outputs N quantile estimates of the action-value distribution for each of the plausible actions.

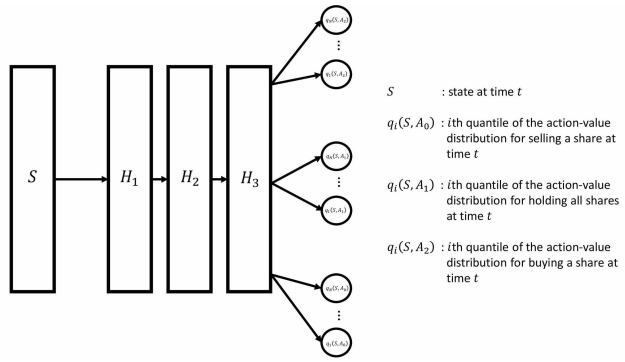


Figure 1. Uncertainty Aware Deep Q Learning Neural Network (UA-DQN) Architecture

The network is trained using the quantile regression loss function (equation 1) where $Z(s,a)$ is the distribution of returns parameterized by N quantiles $\tau_i = \frac{i}{N+1}$ with values q_i (Dabney et al., 2018). For temporal difference learning, the target z is replaced by the Bellman update: $R(s,a) + \gamma \max_{a'} Z(s',a')$.

$$\mathcal{L}(\boldsymbol{q}) = E_{z\sim Z(s,a)} \sum_{i=1}^{N} \rho_{\tau_i}(z - q_i(s,a)) \tag{1}$$

$$\text{where } \rho_{\tau_i}(x) = x(\tau_i - 1_{x<0}) \tag{2}$$

Clements et al. define epistemic uncertainty as the expected

variance in the quantile estimations and aleatoric uncertainty as the variance in the expected quantile values. They show however, that these two uncertainties can be approximated using only an ensemble of two DQNs with parameters $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ (2020):

$$\tilde{\sigma}^2_{\text{epistemic}} = \frac{1}{2} E_{i\sim\{1,N\}}[q_i(\boldsymbol{\theta_1}, s, a) - q_i(\boldsymbol{\theta_2}, s, a)]^2 \quad (3)$$

$$\tilde{\sigma}^2_{\text{aleatoric}} = \text{cov}_{i\sim\{1,N\}}(q_i(\boldsymbol{\theta_1}, s, a), q_i(\boldsymbol{\theta_2}, s, a)) \quad (4)$$

Using these uncertainty estimations, the action selection algorithm for the uncertainty aware DQN is as follows:

Algorithm 1 UA-DQN Action Selection Algorithm

for a in A **do**
  Calculate the expected action-value $\mu = E_{i\sim\{1,N\}}[q_i(s,a)]$
  Approximate $\tilde{\sigma}^2_{\text{aleatoric}}$ and $\tilde{\sigma}^2_{\text{epistemic}}$ using networks $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$
  Adjust for risk-aversion: $\mu \leftarrow \mu - \lambda\tilde{\sigma}_{\text{aleatoric}}$
  Draw a sample $\hat{Z}_a$ from $\mathcal{N}(\mu, \beta\tilde{\sigma}^2_{\text{epistemic}})$
end for
Return: $\text{argmax}[\hat{Z}_a]$

In this algorithm, risk control is achieved through penalizing the expected action-value μ by a fraction of the estimated aleatoric uncertainty for that action. Exploration is achieved by drawing action values from a normal distribution centered at the expected action value μ and with standard deviation equal to the epistemic uncertainty. Risk control can be tuned by changing the hyperparameter λ while exploration can be tuned by changing the hyperparameter β.

3.3. Baseline Algorithm

Buying and holding is a simple trading strategy that tends to perform well for long-term returns. As a result, a baseline buy-and-hold model is also built to assess the performance of risk-aware trading strategies. The baseline buy-and-hold model would buy a single new share everyday from the start of the trading period to the last day of the trading period.

3.4. DQN + Financial Turbulence Index Algorithm

Financial Turbulence Index (FTI) is one measure of extreme asset fluctuation used by the stock trading library FinRL for risk control (Liu et al., 2020). In this project, a DQN model that uses FTI for risk assessment and epsilon-greedy for action selection, is also built for comparison with the proposed uncertainty aware DQN model. All buying is halted when FTI is larger than a pre-defined threshold and shares are gradually sold until FTI drops below the tolerance threshold.

4. Experiments

When trading unstable stocks, risk control is particularly important in order to observe cumulative growth in assets. As a result, to test the risk control capabilities of the uncertainty aware model, we examine the performance of the models with historical data from Yahoo Finance for eight volatile stocks. The data from 2008 to 2015 are used for training while the data from 2016 to 2020 are used for testing. Sample plots of the total returns as a fraction of initial assets for the test data are presented in figures 3 and 4. The total returns at the end of trading for the test data are displayed in table 1 as a fraction of initial assets. Methods that incorporate some form of risk control tend to out-perform the baseline buy-and-hold model. For some stocks such as AMD and ATLC, we notice that the baseline model gives the best performance. This is likely because despite having volatility in the past, during the past 5 years which roughly align with the duration of our test data, these stocks have shown consistent growth in their value. For consistent growth scenarios, buy-and-hold strategies often perform very well. In fact, with the AMD stocks which have had the most significant growth, we observe that UA-DQN with $\lambda = 2.0$ converges to a buy-and-hold action selection strategy.

Table 1: cumulative returns at the end of a test episode as a fraction of initial assets

		Model				
		Baseline	DQN + FTI	UA-DAN 0.5	UA-DQN 1.0	UA-DQN 2.0
Stock Ticker	AMD	**1.091**	1.021	1.023	1.046	**1.091**
	BB	0.998	1.000	1.000	**1.001**	**1.001**
	MKTY	**1.005**	**1.005**	1.001	**1.005**	**1.005**
	DORM	**1.022**	1.020	1.013	1.016	1.011
	RRD	0.993	0.999	1.002	1.000	**1.004**
	ARLP	0.990	**1.009**	1.003	1.001	1.004
	ATLC	**1.027**	1.022	1.026	1.005	1.025
	DLPN	0.968	0.998	1.018	**1.024**	1.013

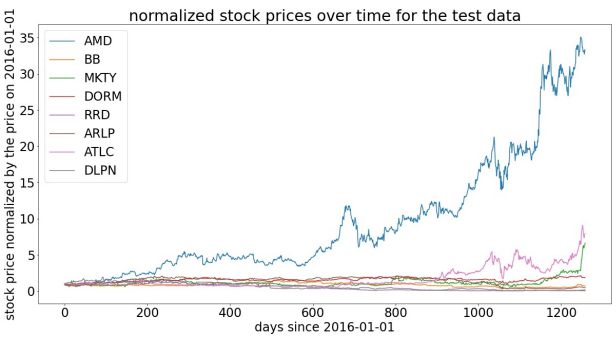
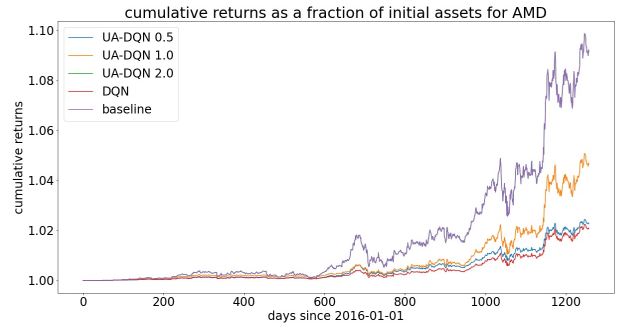
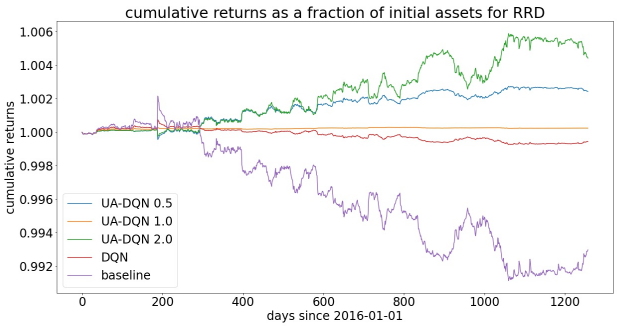
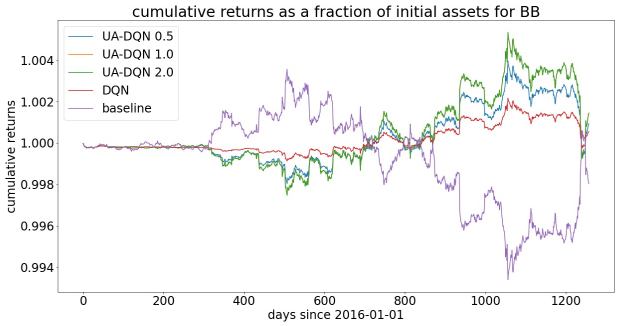
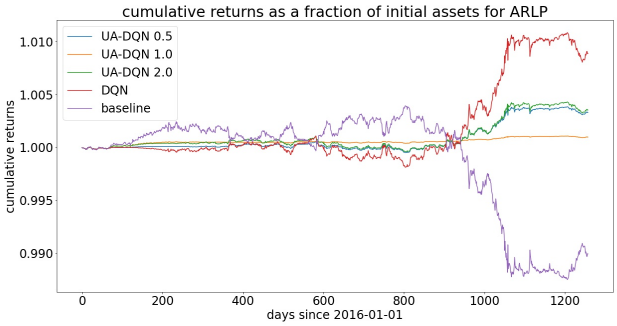


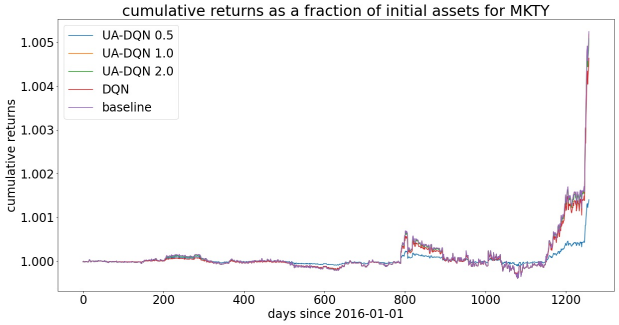
Figure 2. Stock closing prices since 2016-01-01 normalized by the stock price on 2016-01-01. Some stocks such as AMD have shown consistent growth over the past 5 years.

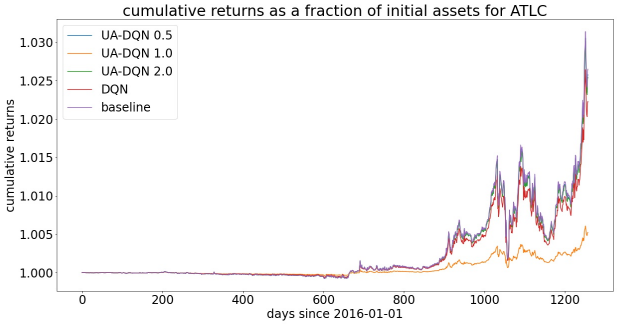


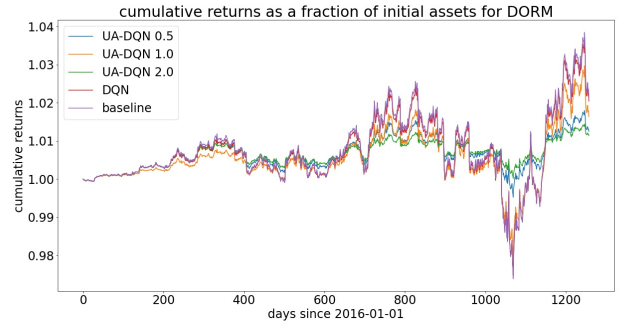












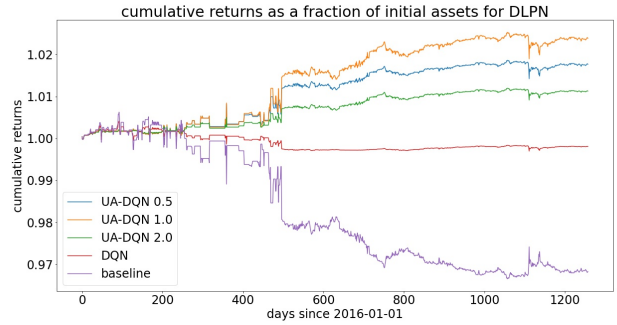


Figure 3. Plots of cumulative returns as a fraction of initial assets over time for AMD, BlackBerry, Mechanical Technology Inc, and Dorman Products Inc. stocks averaged over 4 trials for a baseline buy-and-hold, DQN with FTI for risk-control, and three UA-DQN models with different risk penalty parameters (λ).

Figure 4. Plots of cumulative returns as a fraction of initial assets over time for RR Donnelley , Alliance Resource Partners, Atlanticus Holdings, and Dolphin Entertainment Inc. stocks averaged over 4 trials for a baseline buy-and-hold, DQN with FTI for risk-control, and three UA-DQN models with different risk penalty parameters (λ).

We also observe that UA-DQN matches or out-performs risk control with financial turbulence index for all but the ARLP stocks. These experimental results suggest that uncertainty decomposition can be a suitable technique for risk control in automated stock trading, although appropriate tuning of the hyperparameter λ is necessary for achieving good results. For most volatile stocks used in this experiment, setting λ equal to 0.5 is not sufficient for risk control. $\lambda = 2.0$ performs the best for all but the DLPN stocks. In order to effectively tune λ, a framework for mapping the parameter λ to a human understandable measure of risk is necessary. A potential approach that can be investigated in future studies, is to create groups of stocks based on their volatility, for instance, as measured by the Sharpe Ratio. A suitable λ can be found for each group and the Sharpe Ratio associated with each group can be used as a rough mapping to a measure of risk that is traditionally used by human traders.

4.1. Ablation Study

In order to see whether uncertainty decomposition is necessary or using total uncertainty can also achieve the same risk control results, we perform an ablation study that uses the combined aleatoric and epistemic uncertainties for risk aversion and achieves exploration via epsilon-greedy action selection. Investigating the performance on the same set of volatile stocks as before, we do not observe any improvement over the uncertainty decomposed algorithm. In fact, for most stocks, we see slightly better performance when the uncertainty is not decomposed.

Table 2: cumulative returns at the end of a test episode as a fraction of initial assets for UA-DQNs with total and decomposed uncertainties. Risk control tuning parameter $\lambda = 2.0$ for both models.

		Model	
		UA-DQN with total uncertainty	UA-DQN with decomposed uncertainties
Stock Ticker	AMD	1.081	**1.091**
	BB	**1.001**	**1.001**
	MKTY	**1.005**	**1.005**
	DORM	**1.022**	1.011
	RRD	**1.005**	1.004
	ARLP	**1.009**	1.004
	ATLC	**1.026**	1.025
	DLPN	**1.021**	1.013

One potential explanation is that exploration based on epistemic uncertainty is somewhat masking the effects of risk penalization based on aleatoric uncertainty. In our experiments, β which controls exploration was set to 1. Further experimenting with the β and λ parameters could help us better understand if uncertainty decomposition provides any advantages over using total uncertainty for risk aversion. The paper by Clements et al. does not perform this ablation study for their grid world or contextual bandits experiments. Future studies should first verify the necessity of uncertainty decomposition in contrived, simpler experiments such as the contextual bandits in the paper by Clements et al.

4.2. Source Code

Experiment results and further exploratory plots can be found in our github repository at: https://github.com/anita76/STAT-234-project.

5. Conclusion

In this paper, we applied uncertainty decomposition to the problem of risk control in automated stock trading. Performance results on a set of volatile stocks show that uncertainty decomposition is capable of risk control and could improve upon risk control results from more traditional approaches to risk assessment in stock trading such as the financial turbulence index. Our ablation study however, did not show any improvement over methods that do not decompose uncertainty and only use a measure of total uncertainty.

Future experiments, should attempt to map the risk control parameter to a human understandable measure of risk. Furthermore, experiments should try to extend the uncertainty decomposition technique to actor-critic algorithms in order to investigate if this approach is feasible for more practical, multi-asset stock trading scenarios that are not limited to a finite action-space.

References

Clements, W. R., Delft, B. V., Robaglia, B., Slaoui, R. B., and Toth, S. Estimating risk and uncertainty in deep reinforcement learning. *ICML*, 2020.

Dabney, W., Rowland, M., Bellemare, M. G., and Munos, R. Distributional reinforcement learning with quantile regression. *AAAI Conference on Artificial Intelligence*, 2018.

Leno, F., Hernandez-Leal, P., Kartal, B., and Taylor, M. E. Uncertainty-aware action advising for deep reinforcement learning agents. *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020.

Liu, X., Yang, H., Chen, Q., Zhang, R., Yang, L., Xiao, B., and Wang, C. Finrl: A deep reinforcement learning library for automated stock trading in quantitative finance. *NeurIPS*, 2020.

Neuneier, R. Optimal asset allocation using adaptive dynamic programming. *Advances in Neural Information Processing Systems*, pp. 952–958, 1996.

O'Donoghue, B., Osband, I., Munos, R., and Mnih, V. The uncertainty bellman equation and exploration. *Proceedings of the Thirty-Fifth International Conference on Machine Learning*, 2018.

Sornmayura, S. Robust forex trading with deep q network (dqn). *ABAC Journal*, 39(1), 2019. ISSN 0858-0855.

Tamar, A., Castro, D. D., and Mannor, S. Policy gradients with variance related risk criteria. *Proceedings of the Twenty-Ninth International Conference on Machine Learning*, pp. 387—396, 2012.

Yang, H., Liu, X., Zhong, S., and Walid, A. Deep reinforcement learning for automatedstock trading: An ensemble strategy. *International Conference on AI in Finance*, 2020.