Journal of Korean Institute of Intelligent Systems
Vol. 25, No. 4, August 2015, pp. 319-326
http://dx.doi.org/10.5391/JKIIS.2015.25.4.319

ISSN(Print) 1976-9172
ISSN(Online) 2288-2324

JKIIS

# 데이터 기반 확률론적 최적제어와 근사적 추론 기반 강화학습 방법론에 관한 고찰

# Investigations on data-driven stochastic optimal control and approximate-inference-based reinforcement learning methods

박주영*†, 지승현*, 성기훈*, 허성만*, 박경욱**†
Jooyoung Park†, Seunghyun Ji, Keehoon Sung, Seongman Heo, and Kyungwook Park†

*고려대학교 과학기술대학 제어계측공학과, **고려대학교 경상대학 경영학부
*Department of Control and Instrumentation Engineering, Korea University
**School of Business Administration, Korea University

## 요 약

최근들어, 확률론적 최적제어(stochastic optimal control) 및 강화학습(reinforcement learning) 분야에서는 데이터를 활용하여 준최적 제어 전략을 찾는 문제를 위한 많은 연구 노력이 있어 왔다. 가치함수(value function) 기반 동적 계획법(dynamic programming)으로 최적제어기를 구하는 고전적인 이론은 확률론적 최적 제어 문제를 풀기위해 확고한 이론적 근거 아래 확립된바 있다. 하지만, 이러한 고전적 이론은 매우 간단한 경우에만 성공적으로 적용될 수 있다. 그러므로, 엄밀한 수학적 분석 대신에 상태 전이 및 보상 신호 값 등의 관련 데이터를 활용하여 준최적해를 구하고자 하는 데이터 기반 현대적 접근 방법들은 실용적인 응용분야에서 특히 매력적이다. 본 논문에서는 확률론적 최적제어 전략과 근사적 추론 및 기계학습 기반 데이터 처리 방법을 접목하는 방법론들을 고려한다. 그리고 이러한 고려를 통하여 얻어진 방법론들을 금융공학을 포함한 다양한 응용 분야에 적용하고 그들의 성능을 관찰해보도록 한다.

**키워드** : 데이터 기반 방법론, 확률론적 최적 제어, 근사추론, 기계학습, 금융공학.

## Abstract

Recently in the fields of stochastic optimal control (SOC) and reinforcemnet learning (RL), there have been a great deal of research efforts for the problem of finding data-based sub-optimal control policies. The conventional theory for finding optimal controllers via the value-function-based dynamic programming was established for solving the stochastic optimal control problems with solid theoretical background. However, they can be successfully applied only to extremely simple cases. Hence, the data-based modern approach, which tries to find sub-optimal solutions utilizing relevant data such as the state-transition and reward signals instead of rigorous mathematical analyses, is particularly attractive to practical applications. In this paper, we consider a couple of methods combining the modern SOC strategies and approximate inference together with machine-learning-based data treatment methods. Also, we apply the resultant methods to a variety of application domains including financial engineering, and observe their performance.

**Key Words** : Data-driven methods. Stochastic optimal control, Approximate inference, Machine learning, Financial engineering..

Received: Mar. 22, 2015
Revised : Apr. 5, 2015
Accepted: Jun. 4, 2015
†Corresponding authors
parkj@korea. ac.kr, pkw@korea.ac.kr

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(NRF-2011-0021188).


## 1. 서 론

확률론적 최적 제어 문제의 해법 중 하나로는 가치함수(value function)를 이용한 동적 계획법(dynamic programming)[1]을 들 수 있다. 이러한 동적 계획법을 실제 공학적 문제에 적용하여 최적 제어 전략을 구하고자 하는 경우, 가치함수를 위하여 어떠한 종류의 함수족(function family)을 사용할 것인가, 가치함수의 계산은 어떠한 과정을 통하여 수행할 것인가 등의 어려운 문제점들을 만나게 된다.

근사적 가치 함수와 "현재상태-입력-다음상태-보상값"으로 구성된 상태 전이 데이터를 활용하여 확률론적 최적 제어 문제에 접근하는 방식으로 강화학습(reinforcement learning) 기반 해법[2,3]이 있다. 강화학습은 불확실성을 포함하는 최적 제어 문제에 대해, 주어진

상태와 제어기에 의한 상호작용을 통한 데이터 기반의 학습 방법으로 이 상호과정 중에 얻어지는 보상신호를 최소화하는 제어 입력을 구하는 기법이다. 최근 들어 주목 받는 강화학습 기반 기법으로 근사적 추론 기법이 있다[4,5]. 추론 기법은 기계학습[6,7]의 한 분야로 확률분포를 이용해 주어진 확률모델에 대한 해법을 제시하는 방법이다. 근사적 추론기법은 이러한 추론 문제를 실제 확률분포 대신 이를 근사하는 근사 확률분포를 통해 해결하는 기법이다. 이러한 근사 추론기법을 통한 최적 제어 전략은 연속 및 이산 제어 입력 문제에 모두 적용될 수 있으며, 여러 가지 측면에서 유용한 성질을 갖는다. 본 논문에서는 불확실성과 비선형성이 존재하는 확률론적 최적제어 문제를 대상으로 확률론적 최적제어 기법, 근사 추론 및 기계학습 기반 데이터 처리 방안을 융합하여 데이터 기반의 준최적해를 구하는 기법을 고려한다.

이와 같은 융합 연구에 있어서 본 논문에서 첫 번째로 고려하는 이슈는 확률론적 최적제어 기법과 근사적 추론 기반 강화학습 개념의 융합이다. 최근들어, 확률론적 최적제어(stochastic optimal control) 및 강화학습(reinforcement learning) 분야에서는 데이터를 활용하여 준최적 제어 전략을 찾는 문제[8,9]를 위한 많은 연구 노력이 있어 왔다. 가치함수(value function) 기반 동적 계획법(dynamic programming)으로 최적제어기를 구하는 고전적인 이론은 확률론적 최적 제어 문제를 풀기위해 확고한 이론적 근거 아래 확립된바 있다. 하지만, 이러한 고전적 이론은 매우 간단한 경우에만 성공적으로 적용될 수 있다. 그러므로, 엄밀한 수학적 분석 대신에 상태 전이 및 보상 데이터를 활용하여 준최적해를 구하고자하는 데이터 기반 현대적 접근 방법은 실용적인 응용분야에서 특히 매력적이다. 본 논문의 첫 번째 이슈에서는 데이터 기반 확률론적 최적 제어 전략과 근사적 추론 기반 강화학습 접근 방법을 융합한 방법론을 고려한다. 그리고 이러한 방법론에 대한 성능 관찰을 위하여 간단한 일차원 제어 문제와 금융공학 관련 문제를 고려한다. 최근들어, 각종 제어 이론 및 기계학습 기반 인공지능 방법론은 주요 금융공학 문제[10-14]의 중요한 핵심 도구로 자리를 잡아 가고 있다. 본 논문의 첫 번째 이슈에서 고려하는 금융공학 관련 문제는 동적 옵션 헤징[13,14] 문제이다. 옵션의 헷징은 특히 금융기관에 중요한 의미를 갖는데 주로 금융기관들이 취하게 되는 옵션의 숏포지션은 이론적으로 무한손실을 가져올 수 있기 때문이다. 따라서 옵션을 매도하는 금융기관들은 자신들이 발행한 옵션의 투자수익(payoff)을 완벽하게 또는 최소한의 오차범위 내에서 복제할 수 있는 복제포트폴리오(replicating portfolio)를 구성하여 이 포트폴리오의 롱포지션을 취함으로써 옵션 숏포지션을 헷징하게 된다. 옵션 헷징은 일반적으로 정적헷징(static hedging)과 동적헷징(dynamic hedging)으로 구분되며 정적헷징은 헷징 대상이 되는 옵션과 기초자산이 동일하거나 유사한 옵션 포트폴리오를 구성하고 이 옵션 포트폴리오를 변동없이 유지하는 전략이다. 정적헷징에 대한 연구는 [15,16]등의 초기 연구이래 다양한 형태의 헷징 전략들이 제기되었는데 특히 장애물옵션(barrier option)과 같은 이색옵션(exotic options)의 경우에 다양하게 적용된 연구결과들이 있다[17,18]. 정적헷징이 옵션헷징을 위하여 다른 옵션들을 이용하여 포트포리오를 구성하고 이 포트폴리오의 포지션을 유지하는 것과 대조적으로 동적헷징의 경우는 옵션의 기초자산에 대한 포지션을 취하고 이를 주기적으로 조정(rebalancing)함으로써 헷징 목적을 달성하는 전략으로 표준적인 유로피언 옵션에 대해서는 델타헷징(delta hedging), 델타-감마헷징, 델타-감마-베가 헷징 등이 주요 전략으로 활용된다. 동적헷징의 유효성 평가는 헷징비용의 크기에 따라 평가된다[19,20]. 한편, 장애물옵션이나 아시안 옵션과 같은 이색옵션에 대한 동적헷징 전략을 제안하는 연구들에서는 기초자산의 가격변동에 대하여 Poisson Jumps, Levy Market 등의 특정한 가정을 기반으로 하여 quasi-explicit hedging 방식을 활용하는 전략들이 주류를 이루고 있다[21,22,23]. 또한, 아메리칸 옵션에 대한 동적헷징의 경우에도 [24]의 연구에서와 같이 quasi-explicit hedging 전략을 활용할 수 있다. 본 논문에서는 유럽형 콜옵션을 대상으로 하여, 데이터기반 확률론적 최적제어 전략과 근사적 추론기반 강화학습 방법론이 가미된 본 논문의 제어 방법이 어떠한 성능을 갖는지를 관찰한다.

본 논문의 융합연구에서 고려하는 두 번째 이슈는 금융공학 분야 중 트레이딩 전략을 고려한다. 공학적 기법에 의존하는 금융공학 도구 중 본 논문에서 주목하는 응용 분야 중 하나는 추세 추종형 접근 방식(trend-following approach)[10-12]의 트레이딩 전략이다. 본 논문에서는 [10]의 확률론적 최적제어 기반 추세 추종형 트레이딩 기법에 HMMUG(hidden Markov model with univariate Gaussian outcomes) 모델링[25], 확률론적 최적제어적 관점[10] 및 지수함수형 NES (exponential natural evolution strategy) 기법[26]을 함께 접목하는 방안을 고려한다. 그리고, 미국 NASDAQ 시장의 데이터를 대상으로 고려된 방법론의 적용가능성을 시험해 본다.

본 논문의 구성은 다음과 같다. 2장에서는 확률론적 최적제어와 근사적 추론 그리고 데이터 기반 데이터 처리 방안 등을 접목하는 접근 방식들을 고려한다. 그리고 이들을 간단한 일차원 확률론적 최적 제어 문제와 주요 금융 공학 문제에 응용하는 사례를 다룬다. 마지막으로 3장에서는 결론과 향후 과제를 제시한다.

## 2. 본 론

각종 공학문제에서 확률론적 최적제어 문제는 이차 다항식에 의한 가치함수의 근사와 시간에 종속하는 선형 제어기에 의해 최적 정책에 근사하는 전략을 사용하곤 한다. 이러한 근사를 사용함으로써, 지역적 최적 제어기는 실제로 효과적인 결과를 보여 준다. 본 논문에서는 참고문헌 [8]의 ITSOC(information theoretic stochastic optimal control)에서와 같이 동적 시스템을 시변 선형 시스템(time-varying linear

system) $p_t(x'|x,u)=N(x'|a_t+A_tx+B_tu,C_t)$ 으로 근사하며, 제어 정책 $\pi_t(u|x)$는 각 시간 $t$에 $\pi_t(u|x)=N(u|s_t+S_tx,\Sigma_t)$와 같은 선형적인 표현을 이용하는 방안[8]을 고려한다. 참고문헌 [8]의 ITSOC 기법 등의 방법론은 최적제어기를 찾는 과정에서 고차원의 쌍대 최적화 문제(dual optimization problem)을 풀어야 하는데 [8,9], 참고문헌 [8]에서 언급한 바와 같이 이 과정은 여러 가지 범용적 방법들이 쉽게 해결할 수 없는 어려움을 수반하게 된다. 본 논문에서는 이러한 쌍대 문제의 풀이를 대신하여 가상 데이터를 생성하여 활용하는 데이터 기반 전략과 최근에 보고된 바 있는 근사적 추론 기반 최적제어 및 강화학습 방법론 [4,5]가 사용하는 전략을 도입한다. 즉, 새로운 최적 정책을 찾기 위한 과정으로, 학습한 정책 $\pi_t(u|x)=N(u|s_t+S_tx,\Sigma_t)$과 시변 선형 시스템 모델 $p_t(x'|x,u)=N(x'|a_t+A_tx+B_tu,C_t)$을 이용해 $M$개의 가상 데이터(virtual data)를 생성하며, 새로운 정책을 구하기 위한 확률 제어기 형태로는 다음의 볼츠만 분포를 (Boltzmann like distribution) 이용한다[4].

$$\pi^{n+1}(u_t|x_t)=\exp\left\{\Psi_t^{n+1}(x_t,u_t)-\overline{\Psi}_t^{n+1}(x_t)\right\},$$

이때의 에너지는

$$\Psi_t^{n+1}(x_t,u_t)=\log\pi^n(u_t|x_t)+\log P(r_t=1|x_t,u_t)+\int_{x_{t+1}}P(x_{t+1}|x_t,u_t)\overline{\Psi}_{t+1}(x_{t+1})$$

이며, 로그 파티션 함수는 다음과 같다.

$$\overline{\Psi}_t^{n+1}(x_t)=\log\int_u\exp\left\{\Psi_t^{n+1}(x_t,u)\right\}.$$

기저 함수의 선택에서 로그 파티션 함수 $\overline{\Psi}$의 계산은 유한한 상태 입력에 대해서는 간단하게 계산될 수 있으나, 일반적인 경우에는 연속된 제어 입력 공간에 대해 고려하므로 계산이 어렵다. 이러한 어려움을 극복할 수 있는 적합한 기저 집합(basis sets)의 선택 중 하나로 $\widetilde{\Psi}(x_t,u_t,w_{t)=}-0.5u_t^TP_{uu}(x_t,w_t)u_t+u_t^Tp_u(x_t,w_t)+q(x_t,w_t)$와 같은 표현을 고려할 수 있다[4]. 여기에서 편의상 $t$의 표기를 생략하면 각 $t$ 시점에서 $P_{uu}(x,w)$은 양의 정부호 행렬이며, $p_u(x,w)$은 벡터 함수이고, $q(x,w)$는 스칼라 함수이다. 이러한 기저 집합의 선택은 로그 파티션 함수 $\overline{\Psi}$를 위한 적분이 닫힌 형태로 표현될 수 있는 장점을 가지며[4], 가상 시나리오 $\widetilde{D}_v$에 대한 데이터 처리를 수행함으로써 새로운 정책을 구하는 전략을 구축할 수 있다. 본 논문에서는, 이러한 전략을 주요 핵심 부분으로 사용하는 데이터 기반 확률론적 최적제어 방법론의 응용 가능성을 확인하기 위하여 두 개의 응용 문제를 고려한다. 첫 번째 응용 예제는 참고문헌 [27]의 6장에서 다룬 일차원 확률제어문제인데, 이 예제에서 고려하는 주요 시스템 파라메터는 다음과 같다[27]:

$$A=1,B=-0.5,Q=1,R=1,\gamma=0.95$$

본 예제에 대하여 제어기를 사용하지 않은 경우와 본 논문에서 고려한 융합적 방법론을 적용하여 구한 상태궤적이 각각 그림 1과 2에 보여졌다. 이 그림의 상태궤적은 본 논문의 방법론이 참고문헌 [27]의 간단한 예제에 대하여 정상적으로 적용될 수 있음을 보여준다.

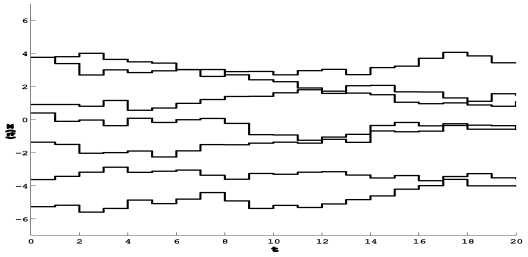

그림 1. 제어기를 사용하지 않은 경우의 상태궤적
Fig. 1. State trajectories: Uncontrolled cases

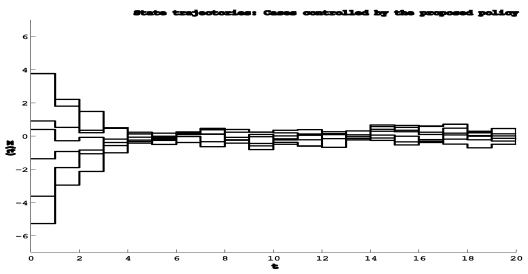

그림 2. 제어기를 적용한 경우의 상태궤적
Fig. 2. State trajectories: Controlled cases

확률론적 최적제어와 근사적 추론 기반 강화학습이 접목된 융합적 기법을 위한 두 번째 응용문제는 옵션 헤징 예제(e.g., [13,14])로서 $S_0$=10, $E$=10, $\mu$=0.0916, $\sigma$=0.3066, $r$=0.05, $T$=11[day], $\delta t$=1[day], $N$=500, $M$=1000을 사용하여 유럽형 콜옵션에 대한 동적 옵션 헤징 문제를 다루었다. 본 논문의 방법론을 각 단계별 절차에 적절한 튜닝과정을 거쳐 적용할 경우 그림 3과 같이 정해진 투자수익 다이어그램(payoff diagram)과 유사한 만기 시 포트폴리오 가치, 즉, 부의 분포를 얻을 수 있었다. 그리고, 시뮬레이션 중 고려된 특정 자산 경로에 대하여 위의 방법론이 적용될 때 제어입력(control), 현금보유량 및 포트폴리오의 가치가 시간이 경과함에 따라 어떻게 변하는 지가 그림 4에 보여졌다. 이 그림의 맨 위쪽 그래프는 자산 경로와 행사가격을 보여주고, 두 번째 그래프는 제어입력이다. 그리고, 세 번째 그래프는 헤징을 위한 포트폴리오 중 현금보유량이고, 마지막 그래프에서는 작은 원 모양의 블랙-숄즈 공식 결과와 막대 그래프 모양의 본 논문의 방법론 적용으로 얻어지는 포트폴리오의 가치를 비교하고 있다. 이 그림의 결과는 고려된 자산 경로에 대해 적절한 동적헤징이 수행되고 있음을 보여준다.

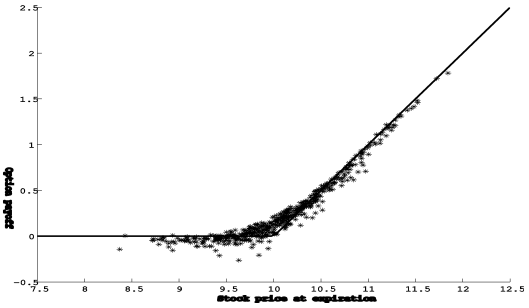


그림 3. 제안된 데이터 기반 확률론적 최적제어 방법론을 적용할 경우의 만기 시점 부의 분포
Fig. 3. Distribution of final payoff when the proposed data-based stochastic optimal control method is applied

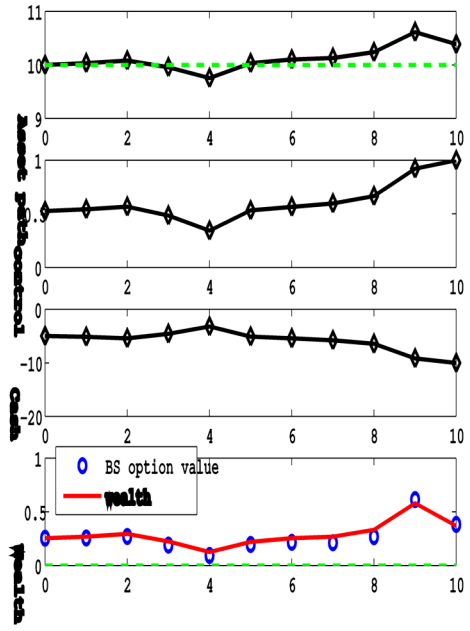


그림 4. 동적 헷징 예제를 위한 시뮬레이션 결과
Fig. 4. Simulation results for the dynamic hedging example

본 논문의 융합연구에서 고려하는 두 번째 이슈는 금융공학 분야 중 트레이딩 전략을 고려한다. 최근 제어이론의 응용 분야 중 금융공학과 관련하여 주목을 끄는 흥미로운 문제 중 하나는 추세 추종형 제어이론 기반 트레이딩 전략(control-theoretic trading strategy)[10-12]을 들 수 있다 이러한 추세 추종형 제어전략 중 본 논문에서 다룰 방법은 Dai 등에 의해 제시된 확률론적 최적제어 방법[10]이다. 이러한 방법은 가격의 변화를 모델링할때 다음과 같은 국면전환 기하 브라운 운동(regime-switching geometric Brownian motion)을 사용한다:

$$dS_r = S_r[\mu(\alpha_r)dr + \sigma dB_r],\ S_t = S,\ t \le r \le T \le \infty$$

여기서 $\alpha_r \in 1,2$는 시장의 mode를 의미한다. 즉, 두 가지의 상태를 갖는 마르코프 체인(two-state Markov chain)을 나타내며, $\alpha_r = 1$ 은 상승 국면(bull market)을 나타내고 $\alpha_r = 2$ 은 하락 국면(bear market)을 나타낸다. 또한, $\mu(i) = \mu_i,\ i = 1,2$ 는 각 장세에 대한 기대 수익률(expected rate of return)을 의미하고, $Q = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}$는 이 마르코프 체인의 생성 행렬(generator)이다. 각 $B_r$은 표준 브라운 운동(standard Brownian Motion)을 의미하며, $\{\alpha_r\}$과 $\{B_r\}$은 서로 독립이다. Dai 등에 의하면 [1]에서 S&P500, 다우존스, 나스닥 등의 인덱스 지수 $S_t$가 위와 같은 국면전환 기하 브라운 운동으로 모델링되고, 이때의 거래비용과 무위험 이자율이 각각 $100K$[%], $\rho$ 이면 목적함수가 다음과 같이 정의될 때, 최적 매수시점 $\tau_1, \tau_2, \cdots$ 과 최적 매도시점 $\tau_1, \tau_2, \cdots$ 에 대해 목적함수를 최대화 하는 해를 정지시간(stopping times)의 개념과 편미분방정식의 풀이 등을 활용하여 구할 수 있다[10].

$$J_i(S,p,t,\Lambda_i) = \begin{cases} E_t\left\{\sum_{n=1}^{\infty}\left[e^{-\rho(v_n-t)}S_{v_n}(1-K) - e^{-\rho(\tau_n-t)}S_{\tau_n}(1+K)\right]I_{\{\tau_n < T\}}\right\} \\ \quad if\ initially\ flat, \\ E_t\left\{e^{-\rho(v_1-t)}S_{v_1}(1-K)\right. \\ \quad \left. + \sum_{n=2}^{\infty}\left[e^{-\rho(v_n-t)}S_{v_n}(1-K) - e^{-\rho(\tau_n-t)}S_{\tau_n}(1+K)\right]I_{\{\tau_n < T\}}\right\} \\ \quad if\ initially\ long. \end{cases}$$

이러한 방법론은 추세추종형 트레이딩 문제에 대한 수학적인 해를 제시한 주목할 만한 결과이며, 추후에 다양한 개선이 뒷받침될 것으로 기대된다. 본 논문에서는 위와 같이 고려되는 수학적 모델의 기본적인 틀에 추가적으로 은닉 마르코프 모델과 진화전략 등의 방법론을 접목하여 관련된 확률론적 최적제어 문제를 푸는 방안을 고려해보았다. 이러한 고려 과정에서 관찰된 예비 결과 등은 본 논문의 저자들에 의해 발표된 학술대회 논문[28]에서 간략하게 소개된 바 있다. 이러한 결과를 배경 설명과 함께 보다 구체적으로 소개하면 다음과 같다: 트레이딩에서 고려되는 자산의 경로는 모델링 과정에서 HMMUG 기법에 의거하여 분석하여 $\mu_1, \mu_2, \sigma_1, \sigma_2, \lambda_1, \lambda_2$ 등의 값들을 구하게 된다. 이러한 과정에서 고려하는 HMMUG 모델링 기법을 절차에 따라 표현하면 다음과 같다[25]. 그리고, 이 절차에 등장하는 각 변수의 의미 등에 대한 상세한 사항을 위해서는 [25]를 참고하면 된다:

[HMMUG 기반 모델링 절차의 요약]
입력: $X, \theta^0$
각종 작업 변수 및 선택 사항 설정
초기 $\alpha, \nu, \beta, \log L$ 계산

루프 시작
E-step: 현재 $\beta, \gamma, \xi$ 계산
M-step: $\theta = \{\pi, A, \mu, \sigma\}$ 업데이트
로그우도관련 계산: 다음 스텝 $\alpha, \nu, b, \log L$ 계산
로그우도 history 벡터에 $\log L$ 값 추가
종료조건 확인: 로그우도값이 수렴하거나 최대 반복 횟수에 도달한 경우 루프 종료
루프 종료
출력: $\theta, \gamma, \log L$

확률론적 최적제어 형태로 기술된 추세추종을 위한 트레이딩 문제를 파라메터 탐색을 수행하는 기계학습 측면에서 보기 위한 관점은 다음과 같이 설명될 수 있다. 국면전환 기하 브라운 운동으로 표현할 수 있는 시스템에 대한 최적 정책(optimal policy)을 구하는 과정에서 에피소딕 작업(episodic task) 형식으로 문제를 다루면, 비용(cost)을 최소화하는 관점으로 해를 구할 수 있다. 즉, 자연적 기울기를 활용한 지수함수형 NES(natural evolution strategy) 학습[26]으로 탐색한 문턱을 활용한 트레이딩 전략을 구함으로써 준최적 정책(suboptimal policy)을 확보할 수 있다. 이러한 접근 방법의 장점으로는 비용 함수 혹은 탐색 알고리즘의 선택 등의 단계에서는 융통성과 접근성을 보다 폭넓게 확보할 수 있는 편리함을 들 수 있다. 본 논문에서 고려하는 HMMUG 및 NES기반 방법론의 응용 가능성을 관찰하기 위해서 [21-Aug-1990, 09-Mar-2009]기간 동안에 NASDAQ 지수를 대상으로 추세 추종형 전략을 얻는 문제를 고려하여 보았다. 참고문헌 [10]과 같이 해당 기간 동안의 거래비용 비율은 $K = 0.001$ 로 정하고, 무위험 이자율을 해당 기간의 평균값 $\rho = 5.4\%$ 로 고려하였다. 주어진 NASDAQ 지수 데이터에 대한 HMMUG[25]를 통해 모델링하는 과정에서 얻어진 파라메터 집합의 값은 다음과 같다:

$$\lambda_1 = 1.272,\ \lambda_2 = 2.647, \mu_1 = 0.240,\ \mu_2 = -0.300,$$
$$\sigma_1 = 0.143,\ \sigma_2 = 0.390,\ \sigma = 0.267$$

그리고, 이러한 결과를 얻기까지 EM(expectation maximization) 학습과정에서 관찰된 로그우도 값의 변화추이는 그림 5와 같다. 여기에서 구체적인 로그우도 값의 의미 등에 대한 설명을 위해서는 참고문헌 [25]의 도움을 받으면 된다.

본 논문에서 HMMUG[25] 기반 절차를 사용하여 구한 파라메터 추정 결과는 참고문헌 [10]에서 제시하는 결과와 약간의 차이를 보이고 있다. 이러한 차이는 파라메터 탐색과정에서 어떠한 알고리즘을 사용하였는 지와 어떠한 초기값들을 사용하였는 지 등에 의해 야기될 수 있다. 그리고, 그림 5에서 보여주는 바와 같이 본 예제에 대해 EM 알고리즘을 기초로 하는 HMMUG 모델링 과정을 적용하는 경우에 15회 반복 이전에 수렴하는 양상을 관찰하였다. 모델링 과정에서 구해진 $\mu_1, \mu_2, \sigma_1, \sigma_2, \lambda_1, \lambda_2$를 활용하여 적절한 에피소드들(episodes)을 구성하고 이를 바탕으로 자연적 기울기 방향으로 정책의 파라메터를 업데이트하는 지수함수형 NES 학습[26]을 수행하면 그림 6과 같은 학습 커브(learning curve)가 얻어졌다. 그리고, 이러한 시뮬레이션이 여러 차례 수행될 때 얻을 수 있는 평균적인 경향을 관찰하기 위하여 같은 종류의 실험을 10회 수행한 결과 관찰된 평균 학습 커브(average learning curve)를 그림 7에 보였다. 이러한 평균 학습 커브에서 파라메터 업데이트가 거듭될수록 비용의 감소가 나타나는 평균적인 경향은 본 논문에서 고려하는 방법론의 메카니즘이 바람직한 방향으로 작동하고 있음을 보여준다.

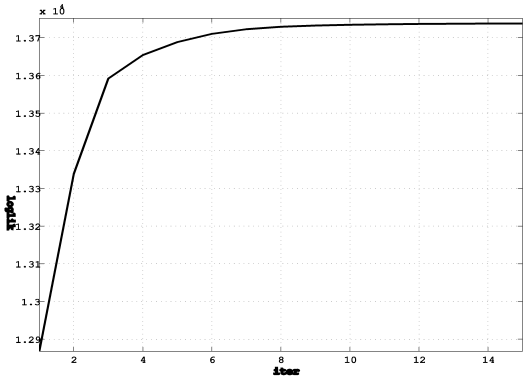


그림 5. EM 반복횟수에 따르는 로그우도 값의 변화
Fig. 5. Log-likelihood values versus the number of EM iterations

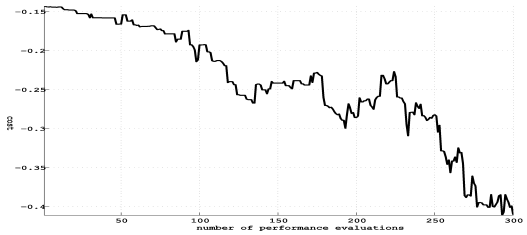


그림 6. NES 적용과정에서 관찰된 학습커브
Fig. 6. Learning curve observed in the process of NES

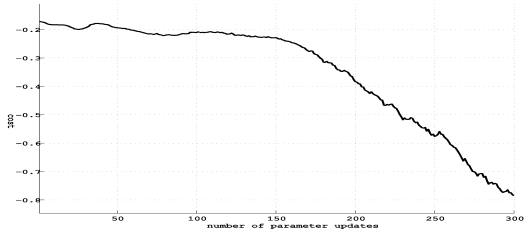


그림 7. NES 적용과정에서 관찰된 평균 학습커브
Fig. 7. Average learning curve observed in the process of NES

HMMUG 및 NES 기반 트레이딩 전략이 제공하는 long 및 flat 트레이딩 포지션 결과를 그림 8에 나타내었다. 그리고, 그림 9에서는 본 논문에서 구한 트레이딩 전략을 적용하는 경우에 얻어지는 트레이드 수익(trade returns)과 부(wealth)

의 변화를 기록하였다. 본 문제에 대해 "Buy and hold" 전략을 사용하는 경우의 투자수익이 투자액의 4.24배이고, 무위험이자율에 의한 투자수익이 투자액의 2.7배임을 고려할 때, 그림 9가 보여주는 투자수익은 확률론적 최적제어 이론을 사용하는 [10]의 성능에 어느 정도 필적하는 결과로써 상당히 고무적인 결과로 볼 수 있다. 아울러 이러한 결과가 수학적 이론을 통한 해석적 방법 대신에 모델링과 최적화 과정에서 기계학습 기반의 근사해를 통하여 얻어진 결과임을 감안하면 본 논문에서 소개한 융합적 방법론의 접근성과 범용성 측면에서 향후에 주목을 받을 수 있는 매우 의미 있는 결과라고 할 수 있다.

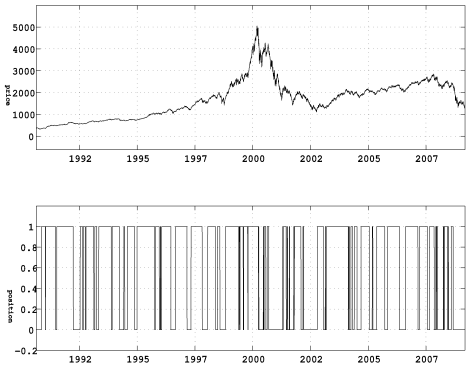


그림 8. NASDAQ 지수와 트레이딩 포지션
Fig. 8. NASDAQ indices and trading positions

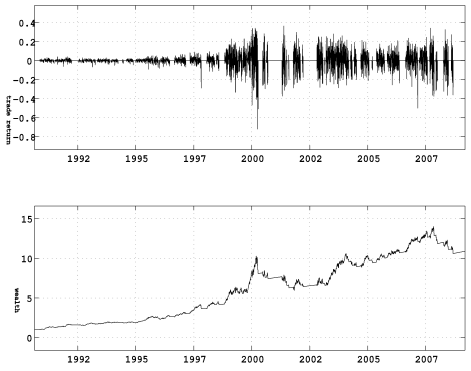


그림 9. 트레이드 수익과 부
Fig. 9. Trade returns and wealth

## 3. 결 론

본 논문에서는 불확실성과 비선형성이 존재하는 확률론적 최적제어 문제를 대상으로 확률론적 최적제어 기법, 근사 추론 및 기계학습 기반 데이터 처리 방안을 융합하여,데이터 기반의 준최적해를 구하는 기법을 고려하고, 이러한 기법의 응용 가능성을 간단한 일차원 제어문제와 금융공학 분야의 동적 옵션헤징 문제와 추세추종 트레이딩 전략 문제에 적용하여 고무적인 결과를 얻었다. 관련하여 향후에 수행할 연구로는, 보다 실용적인 제어시스템과 동적 포트폴리오 최적화 등의 관련 분야에 적용하는 것이 가능하도록 방법론을 추가적으로 개선하는 문제를 들 수 있다.

## References

[1] D.P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. II, 4th edition, Athena Scientific, 2012.

[2] R.S. Sutton and A.G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.

[3] D.P. Bertsekas and J. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996.

[4] K. Rawlik, M. Toussaint and S. Vijayakumar, "On stochastic optimal control and reinforcement learning by approximate inference", *Proceedings of International Conference on Robotics Science and Systems*, pp. 3052-3056, 2012.

[5] M.G. Azar, V. Gmez and H.J. Kappen, "Dynamic policy programming with function approximation," *Proceedings of 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

[6] C.M. Bishop, *Pattern Recognition and Learning*, Springer, 2006.

[7] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT press. 2012.

[8] R. Lioutikov, A. Paraschos, J. Peters and G. Neumann, "Sample-based information-theoretic stochastic optimal control", *Proceedings of the International Conference on Robotics and Automation*, pp. 3896-3902, 2014.

[9] J. Peters, K. Mulling and Y. Altun, "Relative entropy policy search", *Proceedings of the 24th National Conference on Artificial Intelligence (AAAI)*, pp. 1607-1612, 2010.

[10] M. Dai, Q. Zhang and Q.J. Zhu, "Trend following trading under a regime switching model," *SIAM Journal on Financial Mathematics*, vol. 1, pp. 780-810, 2010.

[11] H.T. Kong, Q. Zhang and G.G. Yin, "A trend-following strategy: Conditions for optimality," *Automatica*, vol. 47, no. 4, pp. 661-667, 2011.

[12] J. Yu and Q. Zhang, "Optimal trend-following trading rules under a three-state regime switching model," *Mathematical Control and Related Fields*, vol. 2, no. 1, pp. 81-100, 2012.

[13] J.A. Primbs, "A control systems based look at financial engineering," *Tutorial from the presentation, The*

*Control of Financial Portfolios*, 2009.
[14] D.J. Higham, *An Introduction to Financial Option Valuation: Mathematics, Stochastics and Computation*, Cambridge University Press, 2004.
[15] P. Carr, K. Ellis and V. Gupta, "Static hedging of exotic options," *The Journal of Finance*, vol. 53, pp. 1165-1190, 1998.
[16] E. Derman, D. Ergener, and I. Kani, "Static options replication," *Journal of Derivatives*, vol. 2, pp. 78-95, 1995.
[17] S. Chung, P. Shih and W. Tsai, "Static hedging and pricing american knock-out options," *Journal of Derivatives*, vol. 37, pp. 23-48, 2013.
[18] M. Nalholm and R. Poulsen, "Static hedging of barrier options under general asset dynamics: Unification and application," *Journal of Derivatives* , vol. 13, pp. 46-60, 2006.
[19] M. Kamal, "When you cannot hedge continuously: The corrections to Black-Scholes," *Goldman Sachs Equity Derivatives Research*, 1998.
[20] F. Trabelsi and A. Trad, "Discrete hedging in a continuous-time model," *Applied Mathematical Finance*, vol. 9, pp. 189-217, 2002.
[21] P. Carr, "Semi-static hedging of barrier options under Poission jumps," *International Journal of Theoretical and Applied Finance*, vol. 14, pp. 1091-1111, 2011.
[22] M. Jeannin, M. Pistorius, "Pricing and hedging barrier options in a hyper-exponential additive model," *International Journal of Theoretical and Applied Finance*, vol. 13, pp. 657-681, 2010.
[23] W. Yip, D. Stephens and S. Olhede, "Hedging strategies and minimal variance portfolios for european and exotic options in a Levy market", *Mathematical Finance*, vol. 20, pp. 617-646, 2010.
[24] J. Huang, M.G. Subrahmanyam and G. Yu, "Pricing and hedging american options: A recursive integration method," *The Review of Financial Studies*, vol. 9, pp. 277-300, 1996.
[25] R.J. Frey, "Hidden Markov models with univariate Gaussian outcomes," *Technical Report*, Stony Brook University, 2009.
[26] T. Schaul, "Benchmarking exponential natural evolution strategies on the noiseless and noisy blackbox optimization testbeds," Proceedings of GECCO' 12, 2012.
[27] Y. Wang and S. Boyd, "Approximate dynamic programming via iterated Bellman inequalities," International Journal of Robust and Nonlinear Control, vol. 25, pp. 1472-1496, 2015.
[28] J. Park, S. Ji, K. Sung, K. Park, "Trend-following based on hidden Markov model and modern evolution strategy," *Proceedings of 2015 Information and Control Symposium,* pp. 52-54, 2015.

## 저 자 소 개

**박주영(Jooyoung Park)**
1983년 : 서울대학교 전기공학과 공학사
1985년 : KAIST 핵공학과 공학석사
1992년 : University of Texas at Austin, 전기 및 컴퓨터공학과 공학박사
1993년~현재 : 고려대학교 과학기술대학 제어계측공학과 교수
관심분야 : Machine Learning, Control Theory
Phone : +82-44-860-1440
E-mail : parkj@korea.ac.kr

**지승현(Seunghyun Ji)**
2014년 : 고려대학교 제어계측공학과 학사
2014년~현재 : 고려대학교 제어계측공학과 석사과정
관심분야 : Machine Learning, Control Theory
Phone : +82-10-4753-5871
E-mail : mysky5871@korea.ac.kr

**성기훈(Keehoon Sung)**
2014년 : 고려대학교 제어계측공학과 학사
2014년~현재 : 고려대학교 제어계측공학과 석사과정
관심분야 : Machine Learning, Control Theory
Phone : +82-10-6893-8777
E-mail : skh0910@korea.ac.kr

**허성만(Seongman Hoe)**

2015년 : 고려대학교 제어계측공학과 학사
2015년~현재 : 고려대학교 제어계측공학과 석사과정

관심분야 : Machine Learning, Control Theory
Phone : +82-10-9189-8657
E-mail : hsm0099@korea.ac.kr

**박경욱(Kyungwook Park)**

1983년 : 서울대학교 경역학과 경영학 학사
1985년 : 서울대학교 국제경영학 석사
1993년 : University of Texas at Austin, 재무학 박사
1994년~현재 : 고려대학교 경상대학 경역학부 교수

관심분야 : Financial Management, Financial Engineering
Phone : +82-44-860-1520
E-mail : pkw@korea.ac.kr