Proceedings of KFIS Autumn Conference 2007 Vol. 17, No. 2.

강화학습에 의한 현가장치의 제어

Suspension Control using Reinforcement Learning

정규백, 문영준, 박주영

고려대 제어계측공학과
E-mail: {qbaek, dreamhill, parkj}@korea.ac.kr

요 약

최근에 국내외의 인공지능 분야에서는, 강화학습(reinforcement learning)에 관한 연구가 활발히 진행되고 있다. 본 논문에서는 능동형 현가장치(active-suspension)의 제어를 위하여 RLS 기반 NAC(natural actor-critic)을 활용한 강화학습 기법을 적용해보고, 그 성능을 시뮬레이션을 통해 확인해본다.

Key Words : Reinforcement learning, RLS estimation, Natural actor-critic method, Active suspension

1. 서 론

최근에 국내외의 인공지능 분야에서는, 강화학습(reinforcement learning)에 관한 연구가 활발히 진행되고 있다. 강화학습 방법론 중 하나의 부류인 액터-크리틱 학습 방법은 정책 반복을 이용하여 액터와 크리틱에 대한 학습을 진행한다. 크리틱의 학습은 정책의 평가에 관련된 부분으로 현재 상태와 다음 상태의 가치 함수의 차를 활용하여 가치 함수를 근사하며, 이 근사값들은 액터의 제어 입력을 선택하는데 이용된다. 액터의 학습은 정책 조정과 관련된 부분으로 최적의 제어 입력을 선택하는 부분이다. 본 논문에서는 액터-크리틱 방법의 하나인 NAC(natural actor-critc) 알고리즘을 개선하는 취지로 최근에 본 연구팀에 의해서 제안된 바 있는 RLS-NAC 알고리즘[1]을 Howell 등이 고려한 능동형 현가장치[2]에 적용하여 보았다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문의 주요 소재가 되는 능동형 현가장치에 대하여 간단히 설명한다. 3장에서는 제어입력이 실수 범위에서 연속적인 값을 취하는 경우를 위한 현가장치 제어 문제에 RLS-NAC[1] 알고리즘을 적용하는 단계를 간단히 소개한 후, 시뮬레이션을 통해 얻어진 성능으로 검증한다. 마지막 4장에서는 결론과 향후 연구 방향 등을 제시한다.

2. 능동형 현가장치

그림 1은 본 논문에서 고려할 1/4 현가장치이며, 운동방정식은 다음과 같다[3].

$$m_s\ddot{z}_s + b_s(\dot{z}_s - \dot{z}_u) + k_s(z_s - z_u) = 0$$
$$m_u\ddot{z}_u + k_t(z_u - z_r) - b_s(\dot{z}_s - \dot{z}_u) - k_s(z_s - z_u) = 0$$

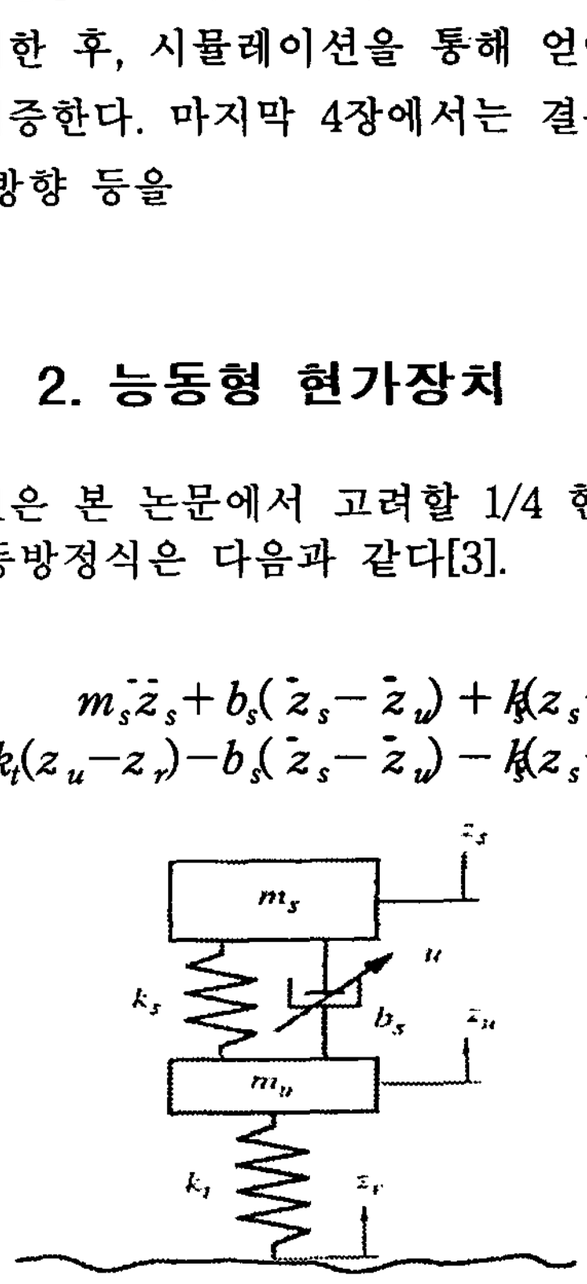


그림 1. 1/4 현가장치[3]

위의 운동 방정식을 상태 방정식의 형태로 나타내면

$$\dot{x} = Ax + Bf_{des} + \Gamma \dot{z}_r$$

의 꼴이 되고

이 때 각 상태 변수는

$$x_1 = z_s - z_u$$
$$x_2 = \dot{z}_s$$
$$x_3 = z_u - z_r$$
$$x_4 = \dot{z}_u$$

$$f_{des} = k_1 x_1 + k_2 x_2 + k_3 x_3$$

$$A = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -k_t & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ -\frac{1}{m_s} \\ 0 \\ \frac{1}{m_u} \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

여기서 z_s, z_u는 각각 차체와 휠의 변위를 의미하며, z_r은 랜덤한 노면 입력[3]을 의미한다. f_{des}는 Howell [2]에 의해 제안된 제어 규칙을 사용하였다.

본 논문의 서스펜션 관련 데이터는 [2],[4]의 경우와 같다. 차체의 질량 m_s=375 kg 이고(이하, 단위 생략), 휠의 질량 m_u=60 이다. 그리고, 타이어를 k_t=200kN/m 의 스프링 계수를 가진 스프링으로 고려한다.

3. RLS-NAC 알고리즘의 적용

본 논문에서 고려하는 RLS-NAC 알고리즘은, 본 논문에 앞서 작성된 바 있는 [1]에서 설명된 대로 다음과 같이 요약될 수 있다(각 용어에 대한 정의 및 상세한 설명을 위해서는 [1], [6] 등을 참조하기 바람):

알고리즘의 적용을 위해 미리 준비할 내용:

- 초기 상태 s_0
- 제어전략 $\pi_\theta(a|s)$과 초기 파라메터 $\theta = \theta_0$, 그리고 관련 미분 벡터 $\nabla_\theta \log \pi_\theta(a|s)$
- 상태가치함수(state value function) 근사기 $\widetilde{V}_v(s) = \phi(s)^T v$에 사용하는 기저함수 $\phi(s) = [\phi_1(s), \cdots, \phi_K(s)]^T$
- 액터 파라메터 θ 갱신 때의 학습율 $\alpha > 0$
- 망각계수(forgetting factor) $\beta \in (0,1)$
- 할인율(discount rate) $\gamma \in (0,1)$
- 트레이스 감쇠계수(trace-decay parameter) $\lambda \in [0,1]$
- 행렬 P_0를 가역으로 만들기 위한 상수 $\delta > 0$
- 각 액터 파라메터 크기를 한정시키기 위한 $M > 0$

알고리즘 적용을 통해 달성하고자 하는 목표:

- 제어전략 $\pi_\theta(a|s)$의 파라메터 벡터 θ를 위한 최적해 발견
- 상태가치함수 근사기 $\widetilde{V}_v$와 우월가치함수 근사기(advantage value function approximator) $\widetilde{A}_w(s,a) = \nabla_\theta \log(a|s)^T w$의 파라메터 벡터 v와 w를 위한 최적해 발견

알고리즘:

for $t := 0, 1, 2, \cdots$ do

- 제어전략을 위한 확률분포 $\pi_{\theta_t}(\bullet|s_t)$로부터 제어입력 a_t를 추출함
- a_t를 적용한 후 보상값(reward) r_t와 다음 상태 s_{t+1}를 관찰함
- w_t와 v_t를 갱신하기 위해 [1]의 (5)와 (6) 식에 나와 있는 RLS 규칙을 적용함.
- 제어전략을 위한 확률분포의 파라메터 벡터를 $\theta_{t+1} = \theta_t + \alpha \omega_t$를 이용하여 갱신함.
- θ_{t+1}의 원소값의 절대값이 M을 초과하는 경우에는 M으로 한정시켜 줌.

end

본 논문에서는 정규분포로 표현되는 연속제어입력 확률분포를 고려한다. 즉, 현가장치의 제어입력 선택 확률분포 π로 다음과 같은 정규분포를 고려하였다:

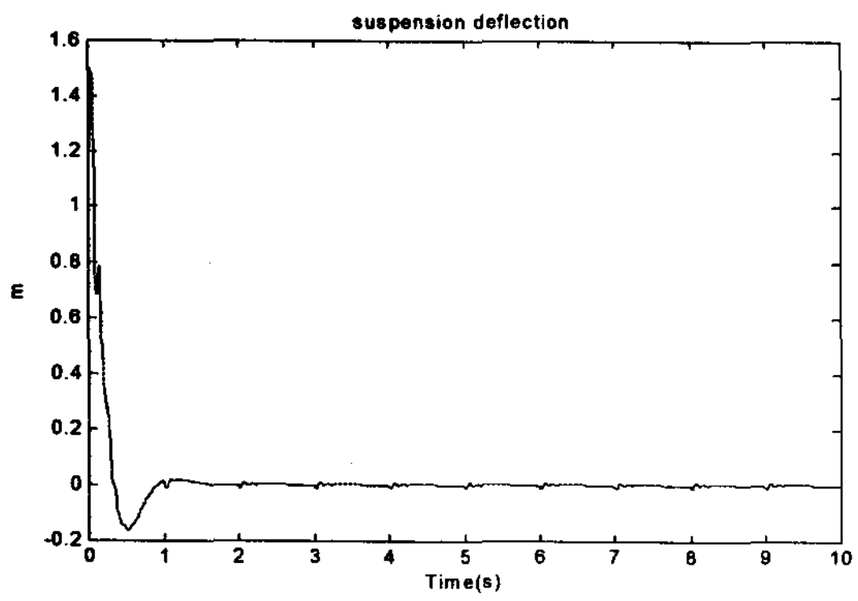
$$\pi(a,\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{-(a-\mu)^2}{2\sigma^2})$$

그리고, π의 평균 μ와 σ를 각각

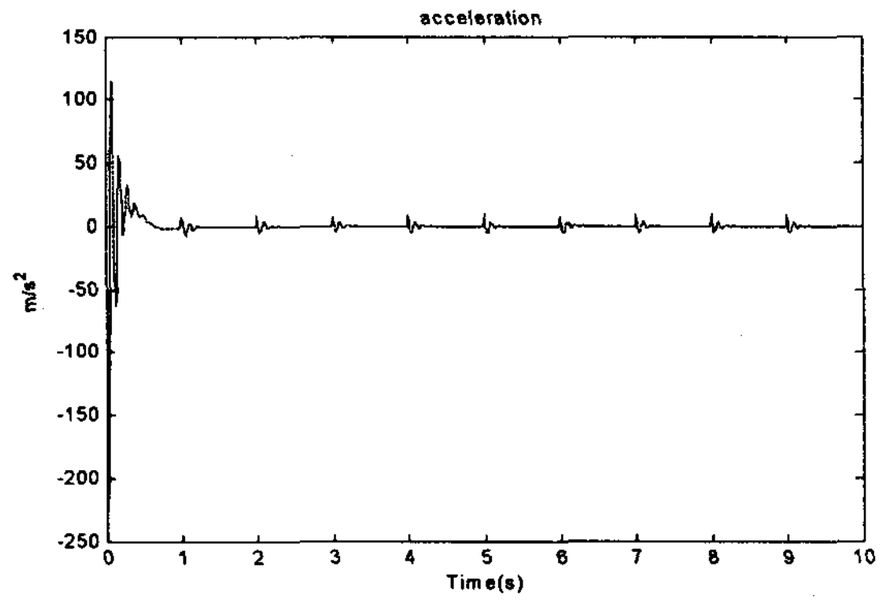
$$\mu=\theta_1 s_1+\theta_2 s_2+\theta_3 s_3+\theta_4,$$

$$\sigma=0.1+\frac{1}{1+\exp(-\theta_5)}$$

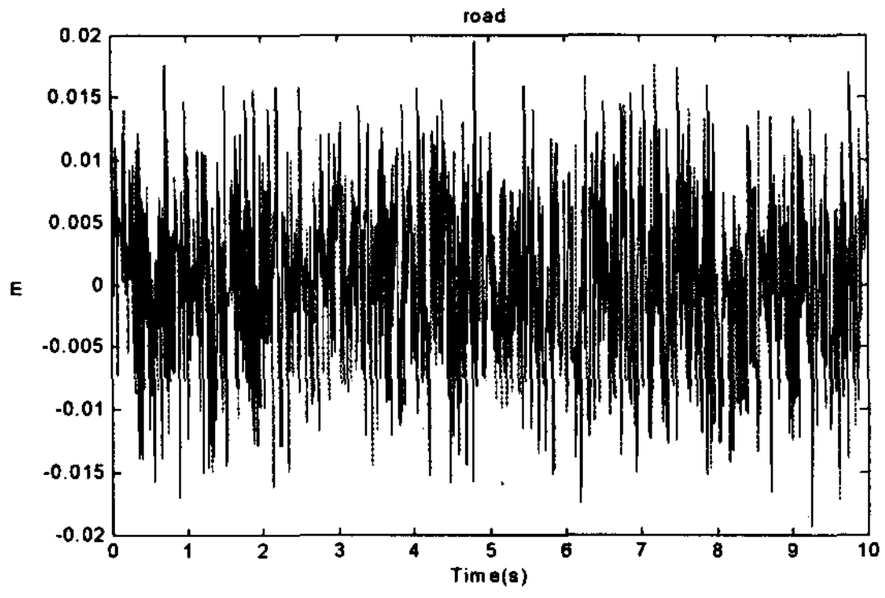
로 정하였다. 따라서, 액터를 위한 확률분포의 파라메터 개수는 총 5개가 된다. 그리고, 위의 정의에서 등장하는 s_1, s_2, s_3 는 상태변수로 현가장치 제어식의 f_{des}에 등장하는 값들과 동일한 값들이다. 제어입력으로는 확률분포 π에 의해서 선택된 값을 사용하였고, 기저함수 ϕ를 위해서는 항등함수가 사용되었다. 차량의 속도는 50m/s 로 일정하고 10초 동안의 학습 결과를 관찰하고 이 결과를 그림 2에 정리하였다. 각각 차체와 축 사이의 변위, 차체의 가속도, 노면 입력으로 가정한 랜덤 입력을 나타낸다. 시간이 지나면서 알고리즘에 의한 학습이 진행되므로써 각 변위값의 변화가 줄어드는 결과를 관찰할수 있었다.



(a) 차체와 축 사이의 변위



(b) 차체의 가속도



(c) 노면 입력

그림 2. RLS-NAC를 현가장치 제어에 적용한 결과

4. 결론 및 향후 연구 방향

RLS(Recursive Least Square)를 기반으로 한 NAC(Natural Actor Critic) 기법을 능동형 현가장치의 제어 문제에 적용시켜 보았다. 실험을 통하여 관찰해 본 결과 강화학습의 적용이 현가장치의 제어에 효과적임을 확인할 수 있었다.

향후 연구 주제로는 퍼지기법을 본 논문에서 고려한 현가장치의 제어에 접목시켜서 적용해보는 문제를 고려하고 있다.

참 고 문 헌

[1] J. Park, J. Kim, and D. Kang, "An RLS-based natural actor-critic algorithm for locomotion of a two-linked robot arm," Lecture Notes in Artificial Intelligence, vol. 3801, pp. 65-72, December, 2005.

[2] M.N. Howell, G.P. Frost , T.J. Gordon, Q.H. Wu, "Continuous action reinforcement learning applied to vehicle suspension control" Mechatronics, Volume 7, Number 3, pp. 263-276(14), April 1997.
[3] T. Busuen, The Design of Semi-Active Suspensions for Automotive Vehicles. PhD thesis, Massachusetts Institute of Technology, June, 1989
[4] G. Verros, S. Natsiavas and C. Papadimitriou, Design optimization of quarter-car models with passive and semi-active suspensions under random road excitation, J Vibr Control 11, pp. 581 - 606, 2005.
[5] T.J. Gordon, C. Marsh, and Q.H. Wu, Stochastic Optimal Control of Active Vehicle Suspension Using Learning Automata, Proc. I.Mech.E. Part I, Vol. 207, pp. 143-152, 1993.
[6] 김종호, 강대성, 박주영, "RLS 기반 Actor-Critic 학습을 이용한 로봇 이동", 한국 퍼지 및 지능시스템학회 논문지, 15권 6호, pp. 88-93, 2005년 12월.
[7] 박주영, 정규백, 문영준, "강화학습에 의해 학습된 기는 로봇의 성능 비교", 한국 퍼지 및 지능 시스템학회 논문집, 17권 1호, pp. 33-36, 2007.