

로봇 학습에서의 확산 모델과 flow matching 모델의 비교 연구

이도혁, 이정우*
서울대학교

dohyeoklee@snu.ac.kr, *junglee@snu.ac.kr

A Study on the diffusion model and flow matching model for robot learning

Lee Dohyeok, Lee Jungwoo*
Seoul National Univ.

요 약

본 논문은 로봇 조작을 위한 정책 학습에서 Flow Matching 과 확산 정책의 성능을 비교 분석한다. 확산 정책은 확산 모델을 기반으로 우수한 행동 복제 성능을 보여주었으나, 학습 효율성과 추론 속도 측면에서 개선의 여지가 있다. Flow Matching 은 확산 모델의 대안으로서, 연속적인 흐름 경로를 통해 노이즈에서 데이터로의 변환을 학습한다. 본 연구는 세 가지 실제 로봇 조작 과제에서 두 방법을 실험적으로 비교한다. 실험 결과, Flow Matching 정책은 더 빠른 학습 및 추론 속도로, 확산 정책과 유사한 성공률을 달성하였다. 실시간 로봇 제어가 요구되는 환경에서 Flow Matching 의 빠른 샘플링 속도는 실용적인 이점을 제공한다. 본 연구는 Flow Matching 이 로봇 정책 학습을 위한 효율적이고 실용적인 대안임을 실증한다.

I. 서 론

최근 확산 모델 (Diffusion Models)은 이미지, 비디오 생성뿐만 아니라 로봇 정책의 액션 생성에서도 유망한 결과를 보여주었다. 확산 정책 (Diffusion Policy)[1]은 U-Net 기반 또는 Transformer 기반 아키텍처를 사용하여 순수 노이즈에서 시작해 여러번의 역확산 과정을 거쳐 깨끗한 액션을 생성한다. 이 방법은 행동 복제(Behavior Cloning)에서 우수한 성능을 입증했으나, 학습 효율성과 추론 속도 측면에서 개선의 여지가 있다.

Flow Matching[2]은 확산 모델의 대안으로 등장한 생성 모델링 기법이다. Flow Matching 은 향상된 학습 안정성과 더 빠른 추론 속도를 제공하면서도 유사한 표현력을 유지한다. 최근 로봇 학습 분야에서 pi0[3]와 같은 모델들이 Flow Matching 을 채택하면서 그 효과가 주목받고 있다.

본 연구의 목적은 로봇 조작을 위한 정책 학습에서 Flow Matching 과 확산 정책의 성능을 비교 분석하는 것이다. 특히 학습 효율성, 추론 속도, 실제 로봇 조작 과제에서의 성능을 중심으로 두 방법을 실험적으로 비교한다.

II. 본론

Flow Matching 은 노이즈에서 데이터로의 연속적인 경로를 학습한다. 샘플 $x^*(\tau)$ 는 흐름 시간 $\tau \in [0, 1]$ 에서 노이즈($\tau = 0$)에서 데이터($\tau = 1$)로 변환된다. 가우시안 조건부 분포를 사용하는 일반적인 공식은 다음과 같다:

$$x^*(\tau) = \tau x^*(1) + (1 - \tau)x^*(0), x^*(0) \sim N(0, I)$$

여기서 $x^*(0)$ 는 표준 정규분포에서 샘플링한 노이즈이고, $x^*(1)$ 는 목표 데이터 분포에서 온 샘플이다. 조건부 벡터장 $u_\tau(x^*(\tau)|x^*(1))$ 는 흐름 시간 τ 에 대한 $x^*(\tau)$ 의 변화율을 나타낸다:

$$d/d\tau x^*(\tau) = u_\tau(x^*(\tau)|x^*(1))$$

조건부 Flow Matching 은 신경망 $v_\theta(x^*(\tau), \tau|x^*(1))$ 를 학습하여 조건부 벡터장을 근사화하는 지도 학습 프레임워크이다. 학습 목적 함수는 다음과 같다:

$$L_{CFM}(\theta) = E[||u_\tau(x^*(\tau)|x^*(1)) - v_\theta(x^*(\tau), \tau|x^*(1))||^2]$$

선형 보간의 경우, 벡터장의 목표값은 $u_\tau = x^*(1) - x^*(0)$ 로 간단하게 계산된다.

확산 정책은 DDPM 을 기반으로 한다. 순수 노이즈 $a_t(K)$ 에서 시작하여 K 번의 역확산 스텝을 통해 점진적으로 노이즈를 제거하며 액션 $a_t(0)$ 를 생성한다. 각 스텝에서 노이즈 예측 네트워크 ϵ_θ 는 현재 노이즈 레벨과 관측값 o_t 를 조건으로 받아 노이즈를 추정한다.

두 방법의 주요 차이점은 다음과 같다:

학습 효율성: Flow Matching 은 일반적으로 더 적은 예폭으로 수렴한다. 확산 모델은 다양한 노이즈 레벨에서의 노이즈 예측을 학습해야 하는 반면, Flow

Matching 은 직접적인 벡터장 회귀를 수행하여 학습이 더 안정적이다.

추론 속도: Flow Matching 은 연속적인 흐름 경로를 통해 더 적은 스텝으로 샘플링이 가능하다. DDPM 이 일반적으로 수십에서 수백 번의 반복을 요구하는 반면, Flow Matching 은 10-20 스텝으로도 고품질 샘플을 생성할 수 있다.

생성 품질: 두 방법 모두 유사한 수준의 표현력을 가지며, 최종 성능은 비슷한 수준이다. 다만 Flow Matching 이 더 적은 샘플링 스텝에서도 안정적인 품질을 유지한다.

실제 로봇 조작 과제에서 두 방법을 비교하기 위해 실험적 검증을 수행한다. 비교 대상은 다음과 같다:

- DP-C: U-Net 기반 확산 정책 (DDPM)
- DP-T: Transformer 기반 확산 정책 (DDPM)
- FMP: Flow Matching 정책 (U-Net 기반)

세 가지 실제 로봇 조작 과제에서 평가를 수행한다:

- Close Drawer (CD): 서랍을 완전히 닫는
과제로, 제약된 움직임 동역학과
비파지(non-prehensile) 조작 이해가
필요하다.
- Towel Folding (TF): 수건의 특정 모서리를
잡아 반으로 접는 변형 가능한 물체 조작
과제이다.
- Cup Arrangement (CA): 누워있는 컵을
테두리를 잡아 접시 위에 똑바로 세우는
과제로, 정밀한 조작과 적절한 방향 제어가
요구된다.

각 과제당 50-100 개의 전문가 시연 에피소드를 수집하여 학습 데이터로 사용한다. 모든 방법은 동일한 데이터셋과 학습 설정으로 훈련되며, 각 과제당 10 회 시도를 통해 성공률을 측정한다.

방법	평균 성공률	학습시 샘플 스텝	추론시 샘플 스텝
DP-C	0.567	100	16
DP-T	0.6	100	16
FMP	0.6	10	10

실험 결과, Flow Matching 정책은 평균적으로 확산 정책과 유사한 성능을 보인다. 반면 Flow Matching 정책은 학습과 추론 두 경우 모두 훨씬 적은 생성과정을 가진다. 이는 실시간 로봇 제어에서 중요한 이점이다.

III. 결론

본 연구는 로봇 정책 학습에서 Flow Matching 이 확산 정책의 효율적인 대안임을 실증한다. Flow Matching 은 다음과 같은 장점을 제공한다:

첫째, 향상된 학습 효율성으로 더 적은 학습 시간으로 유사한 성능에 도달한다. 둘째, 빠른 추론 속도로 실시간 로봇 제어에 더 적합하다. 셋째, 안정적인 학습 과정으로 하이퍼파라미터 튜닝의 부담이 적다.

특히 로봇 조작과 같이 실시간성이 중요한 응용에서 Flow Matching 의 빠른 샘플링 속도는 실용적인 이점을 제공한다. 또한 학습 안정성이 높아 다양한 과제와 환경에 쉽게 적용할 수 있다.

향후 연구 방향으로 Flow Matching 을 대규모 데이터셋과 멀티태스킹 학습에 적용하여 일반화 능력을 더욱 향상시키는 것을 고려할 수 있다. 또한 다양한 Flow Matching 변형 (Rectified Flow, Optimal Transport Flow)을 로봇 학습에 적용하여 추가적인 성능 개선을 탐구할 가치가 있다..

ACKNOWLEDGMENT

This work is in part supported by the National Research Foundation of Korea (NRF, RS-2024-00451435(20%), RS-2024-00413957(20%)), Institute of Information & communications Technology Planning & Evaluation (IITP, RS-2025-02305453(15%), RS-2025-02273157(15%), RS-2025-25442149(15%) RS-2021-II211343(10%)) grant funded by the Ministry of Science and ICT (MSIT), Institute of New Media and Communications(INMAC), and the BK21 FOUR program of the Education, Artificial Intelligence Graduate School Program (Seoul National University), and Research Program for Future ICT Pioneers, Seoul National University in 2026.

참 고 문 헌

- [1] Chi, C., Feng, S., Du, Y., Xu, Z., Cousineau, E., Burchfiel, B., and Song, S., "Diffusion policy: Visuomotor policy learning via action diffusion," in Robotics: Science and Systems (RSS), 2023.
- [2] Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M., "Flow matching for generative modeling," arXiv preprint arXiv:2210.02747, 2023.
- [3] Black, K., Brown, N., Driess, D., Esmail, A., Equi, M., Finn, C., Fusai, N., Groom, L., Hausman, K., Ichter, B., et al., " π 0: A vision-language-action flow model for general robot control," arXiv preprint arXiv:2410.24164, 2024.