View Reviews

Paper ID
5081

Paper Title
Planning with Diffusion for Flexible Behavior Synthesis

Reviewer #2

Questions

1. Summarize the contributions made in the paper with your own words
The majority of model-based methods to control learn a dynamics model separately from the control algorithm which ultimately uses the model to plan. This leads to an issue where a model which maximizes observation likelihood does not necessarily result in a good controller. Instead, this paper proposes Diffuser, which learns the two jointly by leveraging denoising diffusion probabilistic models and reframing planning as a combination of class-conditional sampling and inpainting. The diffusion model iteratively refines the entire trajectory together rather than requiring us to roll out an autoregressive dynamics model. Despite this property, they demonstrate that the learned planner is able to generalize well to trajectories outside the training data. Moreover, they demonstrate that the learned planner can generate long-horizon plans given only sparse rewards. The backbone of the diffusion model consists of 1D convolutions, which enables the planning horizon to be adjusted. Since the reward function must be available in a differentiable form, the authors propose to learn it with a separate model using the same data used to train the diffusion model. They show that the planner can successful generate plans for reward functions on which it was not explicitly trained. The authors experimentally compare the diffusion planner to model free methods on maze environments and a block stacking task with a Kuka arm. Additionally, they compare to both model-free and model-based methods on offline RL locomotion tasks.

2. Novelty, relevance, significance
Model-based reinforcement learning is an important problem which promises to be more sample-efficient than its model-free counterpart. The mismatch between model prediction accuracy and controller performance is a well known issue. While there have been a number of works which address learning a dynamics model with controller performance in mind, this paper is the first to propose jointly learning the planner and dynamics model simultaneously. Moreover, reframing goal conditioning as inpainting and performing planning as class-conditional sampling is an interesting and, to my knowledge, novel extension of the control-as-inference framework [1].

References:
[1] S. Levine, "Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review," arXiv:1805.00909v3, 2018.

3. Soundness
Overall, the paper is sound.

4. Quality of writing/presentation
The paper is well organized and clearly written. It does a good job explaining the novelty and results and provides enough information to support its claims. The appendix appears to provide most of the information necessary to reproduce the results. The main aspect which the authors should elaborate on is details about the specific benchmarks compared to in the paper and why they were chosen.

Post-rebuttal: I am satisfied with the authors' justification for the chosen benchmarks. This should be added to the final paper.

5. Literature

The paper is well situated in the literature and cites all necessary papers to the best of my knowledge.

6. Basis of review (how much of the paper did you read)?
I read the full paper, including the appendix.

7. Summary
As stated above, the major strength of the paper is that the technique is novel and provides an interesting interpretation of goal-conditioned trajectory planning and reinforcement learning as inpainting and class-conditional sampling. The results demonstrate significant performance gains by using Diffuser over existing model-free methods and model-based methods in the case of offline RL. Additionally, the authors provide a means to warm-start planning between time steps, which is crucial to running any online planner under real-time constraints. They provide an analysis of the trade-off in the number of diffusion steps used to re-plan and the resulting accuracy on one of the benchmarks.

While not a major weakness, I have a slight concern about the choice of baselines used to evaluate the proposed technique. The main hypothesis of the paper seems to be that learning the model and planner jointly should perform better than learning a model separately from the planning algorithm used. Yet, the vast majority of experimental results compare with model-free methods and the only model-based baselines are in the offline RL setting. Therefore, I think the paper would greatly benefit from more thorough comparisons with existing model-based methods that make use of learned models.

Moreover, it's not clear if Diffuser's improved performance specifically comes from jointly learning the planner and model or if it's simply that it has better long-horizon prediction capabilities. One potential way to disentangle the source of benefits would be to simply use the diffusion model as a dynamics model paired with a traditional sampling based planner, such as CEM [1] or MPPI [2], and condition on action sequences generated by the planner. If this still performs worse than planning with Diffuser directly, then I think it would be safe to say that jointly learning both performs better. This paper would greatly benefit from such a study.

References:
[1] K. Chua, R. Calandra, R. McAllister, and S. Levine, "Deep Reinforcement Learning in a Handful of Trials using Probabilistic Dynamics Models," NeurIPS, 2018.
[2] G. Williams, N. Wagener, B. Goldfain, P. Drews, J. M. Rehg, B. Boots, and E. A. Theodorou, "Information Theoretic MPC for Model-Based Reinforcement Learning," ICRA, 2017.

Post-rebuttal:
The authors provided solid justification for the chosen baselines and why the offline scenario was the focus. The additional experiments address all of my concerns, in terms of disentangling the sources of the benefits and additional model-based comparisons. These additional experiments and text should be added to the final paper.

8. Miscellaneous minor issues
The only issue is that the sentence in the related work section "applying collocation techniques to learned single-step energies" is missing a citation.

10. [R] Phase 1 recommendation. Should the paper progress to phase 2?
Yes

Reviewer #3

Questions

1. Summarize the contributions made in the paper with your own words

This paper proposes to approximate the process of trajectory optimization with a learned diffusion model such that planning is identical to sampling from this model. Several properties of this approach are identified and tested empirically.

2. Novelty, relevance, significance
As far as I know, the overall idea to apply diffusion models to trajectory optimization is novel and seems promising given all the demonstrated properties. The approach is technically simple as it seems to be a rather straightforward application of existing methods on diffusion models.

The task compositionality due to the diffusion models seems really interesting.

3. Soundness
There is no theoretical claim made in the paper. It would be useful to know for example, in the RL setting, can be planner guarantee to discover the optimal trajectory, in what conditions?

4. Quality of writing/presentation
I have a mixed feeling about the writing of this paper.

On the one hand, I find that each subsection is clearly written.

On the other hand, I find it hard to get a big picture of the paper. Since the paper started by mentioning the issues of existing MBRL methods, I had the impression that this is a standard MBRL setting where an agent is interacting with an environment while learning a model to make decisions. Then, it was very confusing for me since in the paper, there is little to none description of how the agent makes decisiosn online to collect data and how the exploration is done, and there is not much mentioned on training.

Now I guess the paper actually assumes an offline setting where all the data, in this case, trajectories, are collected in advance and given as a dataset? It could save the reader some time if you mention it explicitly early on.

Can the authors also elaborate on the limitations of this work, except the potential slow sampling/planning due to the use of diffusion models? What is the bottleneck? What can be the future improvements? Why using diffusion models in the first place?

5. Literature
When mentioning latent-space MBRL works like MuZero, the authors point out that the key difference is that methods like MuZero generate trajectories autoregressively, while this work generates all timesteps of a trajectory concurrently.

Can the authors elaborate on the implications of this difference? Will this bring advantages or disadvantages?

6. Basis of review (how much of the paper did you read)?
I read the full paper.

7. Summary
Strong points:
* the idea is novel and descriped clearly.

Weal points:
* the overall presentation can be improved.

==========

Thanks for the clarifications. I find the argument for not comparing to MuZero sensible.

10. [R] Phase 1 recommendation. Should the paper progress to phase 2?
Yes

Reviewer #4

Questions

1. Summarize the contributions made in the paper with your own words
The paper presents an approach to planning that mitigates the limitations of model-based reinforcement learning by using a diffusion probabilistic model.

2. Novelty, relevance, significance
I am not an expert in this field, so I cannot evaluate its novelty.

3. Soundness
The approach seems sound.

4. Quality of writing/presentation
The paper was hard to read and follow. Most importantly, the contribution is unclear.

5. Literature
Seems like relevant litearutre (in terms of the techniques used) is cited.
I would have liked to see a comparison to existing model-based approaches to planning.

6. Basis of review (how much of the paper did you read)?
Most of the paper.

7. Summary
The paper presents a novel adaptation of an existing ML technique to planning. The approach seems novel and relevant to many applications, but the main contributions and the key ideas of the proposed framework are not clearly explained. Specifically, it is clear that in order to achieve generality, the model should remain agnostic to reward function (a basic idea in model-based RL), but unclear how the presented idea translates to the fact that planning and sampling become identical: beyond the standard assumptions that apply to any sampling based approach.

Post rebuttal: Based on the author's response and the opinion of the other reviewers that confirm this work's novelty, I am updating my score to request the integration of the clarifications I requested regarding the way by which the proposed probabilistic model is used for planning.

10. [R] Phase 1 recommendation. Should the paper progress to phase 2?
Yes