AAAI-18
Association for the Advancement of Artificial Intelligence 2018
February 2 - 7, 2018, New Orleans, USA

Reviews For Paper

Paper ID 893

Title Learning Approximate Stochastic Transition Models

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question	
[Summary] Please summarize the main claims/contributions of the paper in your own words.	In this paper the authors describe a way to learn a stochastic transition model by using a generative adversarial network approach. In this approach, a generator network is trained as a stochastic model of next state from current state. A second network, the discriminator network, receives two states as input and is trained to output a high value for real observed states and a low score for states from the generator network.
[Relevance] Is this paper relevant to an AI audience?	Of limited interest to an AI audience
[Significance] Are the results significant?	Not significant
[Novelty] Are the problems or approaches novel?	Somewhat novel or somewhat incremental
[Soundness] Is the paper technically sound?	Technically sound
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Not convincing
[Clarity] Is the paper well-organized and clearly written?	Satisfactory
[Detailed Comments] Please elaborate on your assessments and provide constructive feedback.	Please explain the measures of performance shown in Tables 1 and 2 in the captions of those tables. Also please explain how significant the performance differences are between GP-WGAN and SGAN. The illustrations in Figures 1 and 2 are so simple they are not good demonstrations of the algorithm's abilities. The more complex demonstrations in Figure 3 are very hard to see and understand. In the fourth paragraph in the section titled "Experiments" it is not clear why the convolutional layers are 3D. At this point the data dimensional is not described. This paper includes a considerable number of mathematical definitions, derivations, and proofs. This technical detail will not serve the AAAI readership as well as additional examples and explanations.
[QUESTIONS FOR THE AUTHORS] Please provide	How do the performance measures of GP-WGAN and SGAN shown in Tables 1 and 2 vary for different runs, with different initial weights and state transitions?

questions for authors to address during the author feedback period.	Please add, in the Introduction, an explanation for why a stochastic model is needed in the reinforcement learning field. Many in the AAAI audience may not be familiar with this.
[OVERALL SCORE]	Reject
[CONFIDENCE]	Reviewer is knowledgeable in the area

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	
[Summary] Please summarize the main claims/contributions of the paper in your own words.	The paper looks at an approach to learn the transition kernel in an MDP using GAN techniques. The idea is interesting, but there are unfortunately some mistakes in the analysis and I don't see how to overcome these.
[Relevance] Is this paper relevant to an AI audience?	Relevant to researchers in subareas only
[Significance] Are the results significant?	Moderately significant
[Novelty] Are the problems or approaches novel?	Somewhat novel or somewhat incremental
[Soundness] Is the paper technically sound?	Has major errors
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Somewhat weak
[Clarity] Is the paper well-organized and clearly written?	Satisfactory
[Detailed Comments] Please elaborate on your assessments and provide constructive feedback.	In lemma 1, the authors say that x_r, x_g are random variables, but then T must also be a random variable. However, in use T in equation 14 as if it's a deterministic quantity. I don't see how to recover from this.
[QUESTIONS FOR THE AUTHORS] Please provide questions for authors to address during the author feedback period.	How do you set the threshold for convergence of \theta in algorithm 1?
[OVERALL SCORE]	Clear reject
[CONFIDENCE]	Reviewer is an expert in the area

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
[Summary] Please summarize the main claims/contributions of the paper in your own words.	The paper proposes a GAN setup to learn stochastic transition models of an environment. The authors propose an extension to WGAN / GP-WGAN by modifying the discriminator loss function. The proposed algorithm is demonstrated on 1D, 2D grid world domains as well as a physical domain.
[Relevance] Is this paper relevant to an AI audience?	Relevant to researchers in subareas only
[Significance] Are the results significant?	Not significant
[Novelty] Are the problems or approaches novel?	Not novel
[Soundness] Is the paper technically sound?	Technically sound
[Evaluation] Are claims well-supported by theoretical analysis or experimental results?	Not convincing
[Clarity] Is the paper well-organized and clearly written?	Satisfactory
[Detailed Comments] Please elaborate on your assessments and provide constructive feedback.	The motivation to learning stochastic transition models is clear. Using GANs as a way to learn them is reasonable. My concerns about this work are: 1. The primary use of learning transition models in model based RL is to use them to generate samples of states over a sequence of timesteps. The log likelihood of the sequence is one of the measures to evaluate such models as it attempts to incorporate temporal consistency. However, the GAN setup used here is optimizing the next state distribution. This is unlikely to preserve temporal consistency. See [1, 2]. The use of recurrent neural networks is essential to maintain state across timesteps so that the samples exhibit temporal consistency. 2. In general, the literature review needs to be more comprehensive. Algorithms like Dagger, PILCO are completely missing. 3. The metric used for evaluation: L1 error on the next state distribution doesn't appear to capture the essentials of good sample generation . In domains where one needs to precisely generate location of the agent/moving targets etc that are only a few pixels in size, the L1 error is unlikely to highlight the difference between correct and incorrect samples. The tolerance level used in this paper (0.1) appears to be quite high. 4. The performance of agents using these models in planning / learning would be an interesting addition. [1] Self-Correcting Models for Model-Based Reinforcement Learning: AAAI 2017 [2] Agnostic system identification for model-based reinforcement learning: ICML, 2012 .
[QUESTIONS FOR THE AUTHORS] Please provide	My questions are essentially those raised in the comments section.

questions for authors to address during the author feedback period.	
[OVERALL SCORE]	Strong reject
[CONFIDENCE]	Reviewer is knowledgeable in the area