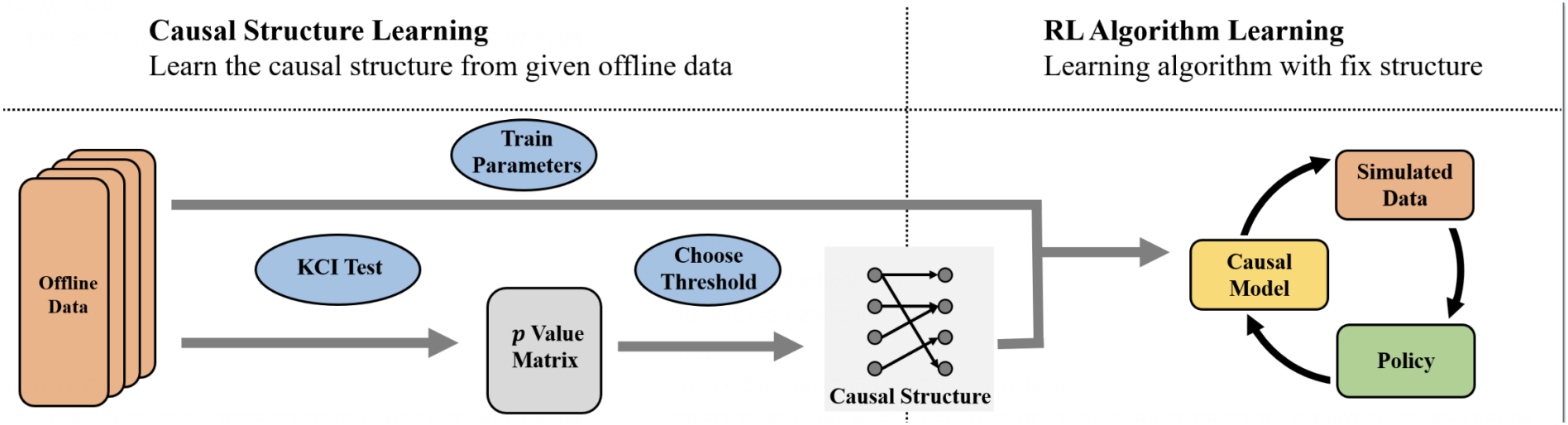
Offline Model-Based Reinforcement Learning with Causal Structured World Models

Zhengmao ZHU, Honglong TIAN, Xionghui CHEN, Kun ZHANG, Yang YU

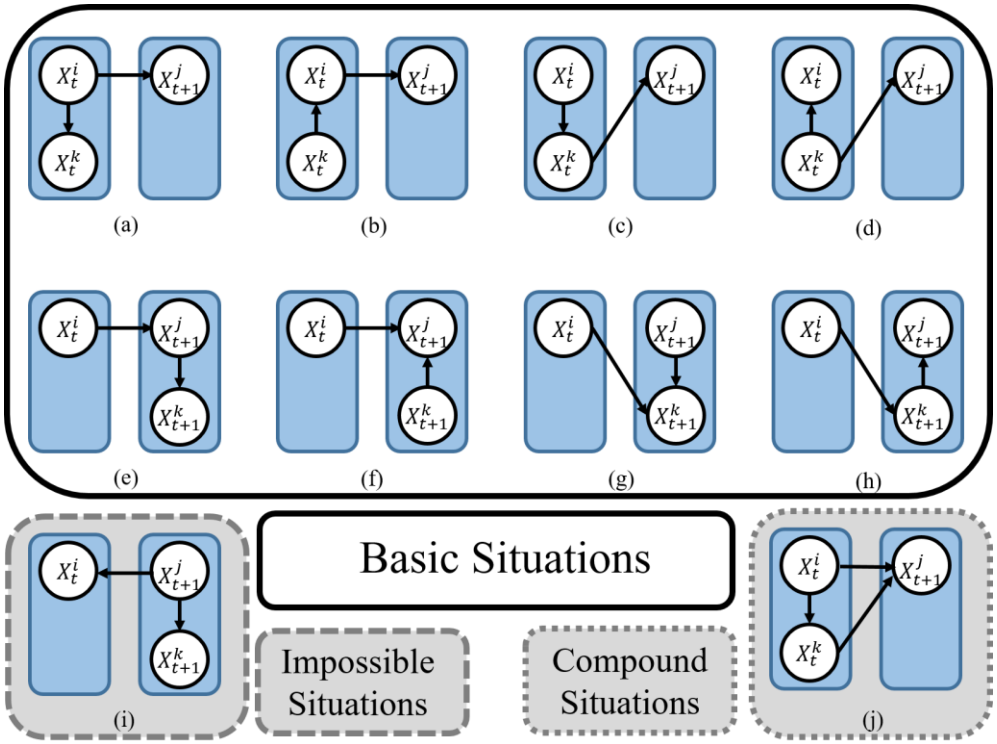
Frontiers of Computer Science, DOI: 10.1007/s11704-024-3946-y

Problems & Ideas



The article theoretically analyzes how causal world models affect policy performance in the context of offline reinforcement learning and studies how to accurately capture causal relationships and construct causal world models in offline scenarios. The analysis found that incorporating causal structures into the learning of environmental models can improve the generalization error bounds, thereby proving the positive impact of causal world models on policy performance. At the same time, the article also developed an efficient causal discovery method in the offline reinforcement learning scenario, which reduces the number of independence assumption tests by leveraging the data characteristics of reinforcement learning, thus improving the efficiency of causal discovery while maintaining the accuracy of causal findings.

Main Contributions



1.We list all possible causal relationships in triples and remove those structures that are impossible to occur in reinforcement learning scenarios and composite structures that need not be discussed.

2.Among the remaining possible structures, we find that the independence assumption test can be conducted with a very concise principle, namely: placing the variables at the current time t into the condition set, and not including the variables at the future time t+1 in the condition set.

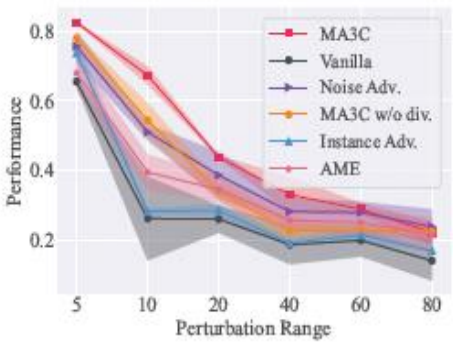
Main Results

Table 3: Results for D4RL datasets. For "ORIGIN" version, we take the results from their original papers. For "FOCUS" version, each number is the averaged score at the last iteration of training, averaged over 3 random seeds. We bold the highest score across all methods.

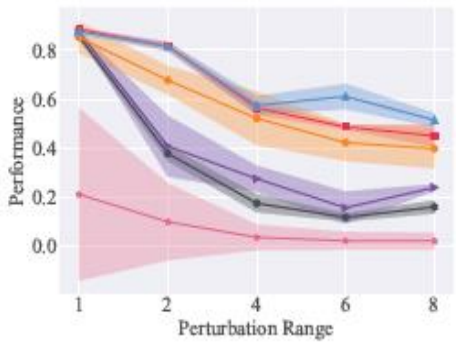
DATASET TYPE	ENVIRONMENT	MOPO		MOREL		COMBO	
		ORIGIN	FOCUS	ORIGIN	FOCUS	ORIGIN	FOCUS
RANDOM	HALFCHEETAH	35	**37**	25	**34**	38	**44**
	HOPPER	11	**30**	53	**63**	17	**23**
	WALKER2D	13	**28**	37	**54**	7	**11**
MEDIUM	HALFCHEETAH	42	**49**	42	**61**	54	**60**
	HOPPER	28	**30**	95	**102**	103	**107**
	WALKER2D	17	**27**	77	**85**	81	**94**
MEDIUM REPLAY	HALFCHEETAH	53	**58**	40	**47**	55	**60**
	HOPPER	67	**71**	**93**	90	**89**	85
	WALKER2D	39	**44**	49	**54**	56	**63**

The results indicate that the proposed offline reinforcement learning algorithm based on causal world models can achieve significant advantages over offline reinforcement learning algorithms based on non-causal models in multiple experimental environments.

Main Results

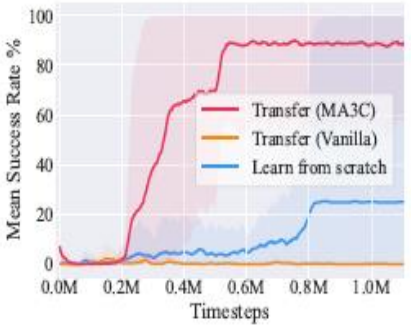


(a) SMAC-1o_2r_vs_4r

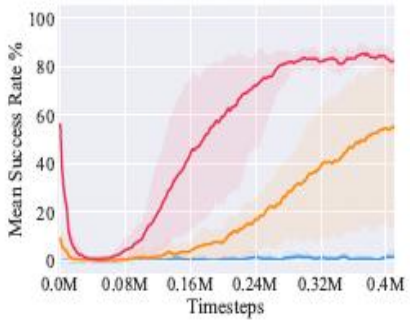


(b) GP-4r

Generalization Ability



(a) Hallway-4x5x9



(b) GP-4r

Transfer Ability

Our method MA3C enjoys high *generalization* ability to different perturbation ranges and could promote the learning phase for new tasks.