cases

doc_1		doc_2		decision	id
authors	• Ruinian Xu • Hongyi Chen • Yunzhi Lin • P. Vela	authors	• Ruinian Xu • Hongyi Chen • Yunzhi Lin • Patricio A. Vela	DUPLICATES	154
title	SGL: Symbolic Goal Learning in a Hybrid, Modular Framework for Human Instruction Following	title	SGL: Symbolic Goal Learning in a Hybrid, Modular Framework for Human Instruction Following		
publication_date	2022-02-25 00:00:00	publication_date	2022-02-25 19:04:31+00:00		
source	SupportedSources.SEMANTIC_SCHOLAR	source	SupportedSources.ARXIV		
journal	IEEE Robotics and Automation Letters	journal	None		
volume	7	volume			
doi	10.1109/lra.2022.3190076	doi			
urls	• https://www.semanticscholar.org/paper/c38b5fddfc520ebd023f97d9e819b6900b633e20	urls	• http://arxiv.org/pdf/2202.12912v1 • http://arxiv.org/abs/2202.12912v1 • http://arxiv.org/pdf/2202.12912v1		
id	id-4456920619948250140	id	id7832311250377038884		
abstract	This paper investigates human instruction following for robotic manipulation via a hybrid, modular system with symbolic and connectionist elements. Symbolic methods build modular systems with semantic parsing and task planning modules for producing sequences of actions from natural language requests. Modern connectionist methods employ deep neural networks that learn visual and linguistic features for mapping inputs to a sequence of low-level actions, in an end-to-end fashion. The hybrid, modular system blends these two approaches to create a modular framework: it formulates instruction following as symbolic goal learning via deep neural networks followed by task planning via symbolic planners. Connectionist and symbolic modules are bridged with Planning Domain Definition Language. The vision-and-language learning network predicts its goal representation, which is sent to a planner for producing a task-completing action sequence. For improving the flexibility of natural language, we further incorporate implicit human intents with explicit human instructions. To learn generic features for vision and language, we propose to separately pretrain vision and language encoders on scene graph parsing and semantic textual similarity tasks. Benchmarking evaluates the impacts of different components of, or options for, the vision-and-language learning model and shows the effectiveness of pretraining strategies. Manipulation experiments conducted in the simulator AI2THOR show the robustness of the framework to novel scenarios.	abstract	This paper investigates robot manipulation based on human instruction with ambiguous requests. The intent is to compensate for imperfect natural language via visual observations. Early symbolic methods, based on manually defined symbols, built modular framework consist of semantic parsing and task planning for producing sequences of actions from natural language requests. Modern connectionist methods employ deep neural networks to automatically learn visual and linguistic features and map to a sequence of low-level actions, in an endto-end fashion. These two approaches are blended to create a hybrid, modular framework: it formulates instruction following as symbolic goal learning via deep neural networks followed by task planning via symbolic planners. Connectionist and symbolic modules are bridged with Planning Domain Definition Language. The vision-and-language learning network predicts its goal representation, which is sent to a planner for producing a task-completing action sequence. For improving the flexibility of natural language, we further incorporate implicit human intents with explicit human instructions. To learn generic features for vision and language, we propose to separately pretrain vision and language encoders on scene graph parsing and semantic textual similarity tasks. Benchmarking evaluates the impacts of different components of, or options for, the vision-and-language learning model and shows the effectiveness of pretraining strategies. Manipulation experiments conducted in the simulator AI2THOR show the robustness of the framework to novel scenarios.		
versions		versions			