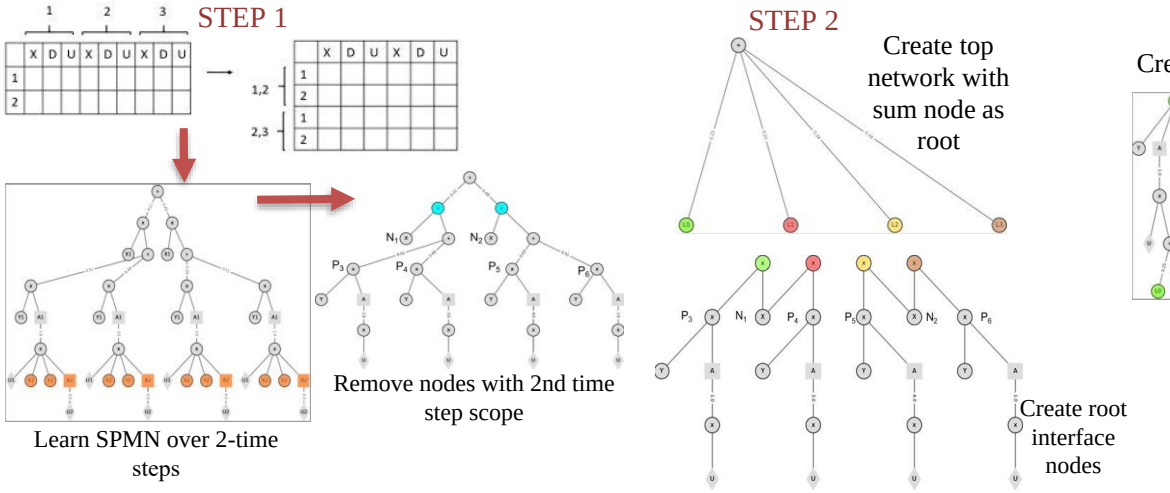
International Conference on Automated Planning & Scheduling (ICAPS), 2021

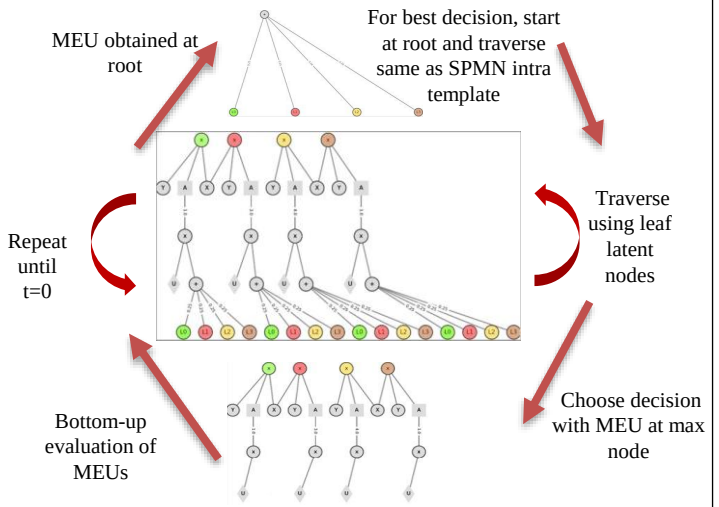
Data-Driven Decision-Theoretic Planning using Recurrent Sum-Product-Max Networks

Hari Teja Tatavarti, Prashant Doshi, Layton Hayes
Institute for AI, University of Georgia, Athens, GA

RSPMN Learning



Computing MEU and best decision



Performance Results

Comparison of learned policy and MEU

Data set	MEU			Average reward			
	Optimal	RSPMN	SPMN	RSPMN	BCQ	Δ %	LL (RSPMN)
GridUniverse	6	6	6	5.9	5.9	0	-0.87
FrozenLake	0.8	0.818	0.13	0.8	0.3	62.5	-6.17
Maze	0.966	0.966	0.052	0.96	0.96	0	-0.86
Taxi	8.9	9	-	8.9	-200	60.25	-2.45
SkillTeaching	-3.022	-3.06	-	-3.009	-7.36	83.3	-2.09
Elevators	-7.33	-7.47	-	-7.357	-9.14	80	-4.8
CrossingTraffic	-4.428	-4.425	-	-4.427	-5.91	94.7	-8.44

Comparison of network size

Data set	$\|X\|$, $\|D\|$	#Episodes	T	|Columns|	|SPMN|	|RSPMN|
GridUniverse[1]	(1, 1)	100K	8	24	138,492	(13, 210)
FrozenLake[1]	(1, 1)	100K	8	24	1,068,246	(18, 401)
Maze[1]	(2, 1)	100K	8	24	352,312	(11, 184)
Taxi[1]	(4, 1)	20K	50	150	-	(80, 1815)
SkillTeaching[2]	(12, 4)	100K	10	170	-	(137, 4878)
Elevators[2]	(13, 4)	200K	10	180	-	(143, 5390)
CrossingTraffic[2]	(18, 4)	100K	15	345	-	(82, 2349)

Average rewards from simulating the learned RSPMN's policies are close to the optimal values.
Difference in RSPMN MEU and average reward indicates whether the environment dynamics were learned accurately.
Batch-constrained Q-learning performs poorly and expects far more data.

Sizes of the SPMNs learned for the sequential data sets blow up.
For the larger RDDLSim domains, SPMNs could not be learned.
In comparison, RSPMNs are smaller because there is no disproportionate growth. Only the sizes of the top and template networks increase.

Conclusion

RSPMNs offer model-based decision-theoretic planning with the benefit that model can be learned directly from data.

MEU and policy computation are linear in the size of the network

However, network size is unbounded

This research is supported in part by NSF grant #1815598 to PD. We thank Alejandro Molina for help with the SPFlow codebase for SPNs. We also thank Pascal Poupart for discussions that helped shape the research direction.