

Curriculum Learning of Bayesian Network Structures (Supplemental Material)

Yanpeng Zhao¹

ZHAOYP1@SHANGHAITECH.EDU.CN

Yetian Chen²

YETIANC@IASTATE.EDU

Kewei Tu¹

TUKW@SHANGHAITECH.EDU.CN

Jin Tian²

JTIAN@IASTATE.EDU

¹*School of Information Science and Technology, ShanghaiTech University, Shanghai, China*

²*Department of Computer Science, Iowa State University, Ames, IA, USA*

Abstract

This document provides supplemental material for the paper “Curriculum Learning of Bayesian Network Structures”.

1. Step Size Statistics

In Table S1, we record the number of times each step size produces the best output network for each benchmark network. It can be seen that for most of the benchmark networks, step size $t = 1$ dominates other step sizes. Overall, nearly half of runs achieve best results with step size $t = 1$.

Table S1: Effect of step size

Network	Step Size (t)			
	No CL	1	2	3
alarm	1	19	9	1
andes	0	14	9	7
asia	3	22	2	3
child	3	16	8	3
hailfinder	13	9	4	4
hepar2	5	10	6	9
insurance	8	7	11	4
sachs	8	10	8	4
water	8	11	7	4
win95pts	1	14	6	9
total	50	132	70	48
ratio	16.67%	44.00%	23.33%	16.00%

ratio is the proportion of total 300 runs held by a specific step size at which the algorithm get best results. No CL corresponds to the plain greedy search.

2. Comparison of Structural Hamming Distance(SHD)

Figure S1 compares the average SHD on Alarm, Asia, Child, Insurance, Sachs and Water networks.

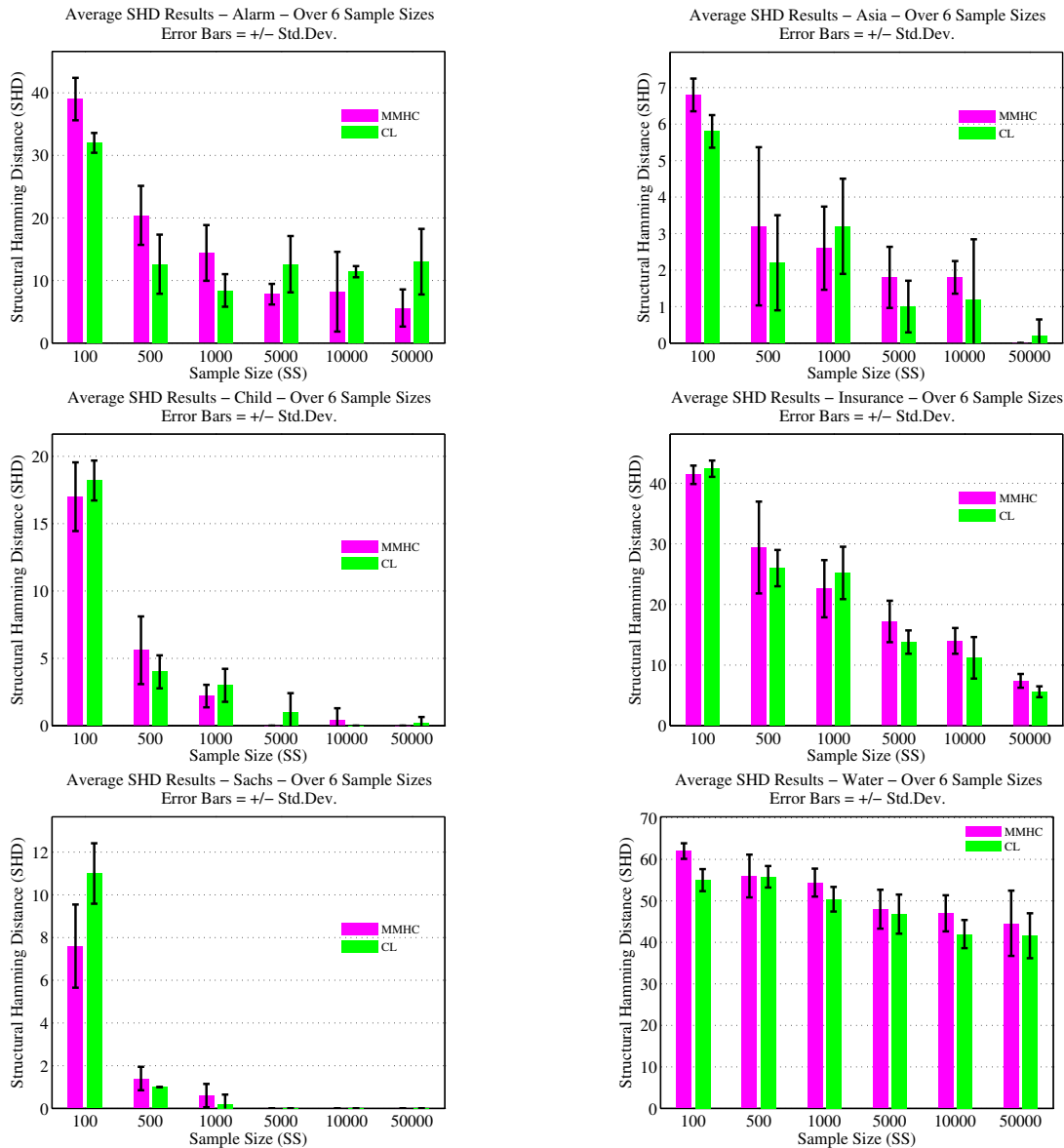


Figure S1: Comparison of the average SHD on the Alarm, Asia, Child, Insurance, Sachs and Water networks between CL and MMHC. The absence of the bar for an algorithm indicates the algorithm found the same network structure as groundtruth.