Supplement to Causal Structure Learning via Temporal Markov Networks

1. Experiments

1.1 Method Parameters

Here are the specific parameters of the methods and their rationales. We used the Center for Causal Discovery's Java implementation of the PC algorithm which is based on the Tetrad implementation from Carnegie Mellon University.

- PC
  - α: 0.01. Decided to ensure approximately one Type 1 error per 10-node graph.
  - depth: 10. Decided to correspond with the maximum number of nodes in the synthetic data and the maximum number of parents (drugs) in the OMOP data.
- BNFinder
  - score: BDe. Tuned, but the Bayesian Dirichlet equivalence score performs no differently than the minimum description length score.
  - maximum parents: 10. Decided to correspond with the maximum number of nodes in the synthetic data and the maximum number of parents (drugs) in the OMOP data.
- Temporal Markov networks using Optim.jl
  - maximum optimization iterations: 1000. Software default.
  - gradient infinity-norm bound: 1e-8. Software default.
  - L-BFGS approximation vectors: 10. Software default.

1.2 Synthetic DBN Experiments

In the synthetic data experiments, the goal was to recover the structure of random DBNs given datasets of timelines sampled from those DBNs. Each dataset received four data treatments designed to test the methods in the face of noise, missing timesteps, and confounding. The plain treatment left the data unaltered. The noisy treatment selected $X_{i,t}$ IID if Bernoulli(ε) and replaced selected values with $x_{i,t} \sim$ Bernoulli$(1/2)$. Each dataset had its own noise level $\varepsilon \sim$ Uniform$(0.1, 0.9)$. The missing treatment selected timesteps IID if Bernoulli(η) and hid their values. This was meant to imitate how patients are unobserved in real EMR data and to measure the influence of assuming unobserved values to be false. Each dataset had its own missingness level $\eta \sim$ Uniform$(0.1, 0.9)$. The confounding treatment randomly selected a subset of confounders (variables with at least two children) and removed them from the data and the ground truth graph. Specifically, the DBN graph was compressed (rolled up) (Plis et al., 2015), confounders were randomly selected so that no more than $2/5$ of the variables would be hidden and so that the class proportion remained in $[0.1, 0.9]$, the confounding variables were removed from the graph by summing them out, and the graph was uncompressed to become the new ground truth graph. For each of the four data treatments the data was represented in two ways: fully-observed and condensed, as illustrated in Figures 2b and 2c. The condensed data imitates real EMR data where negatives are typically not recorded and absolute times are not reliable, but it also simplifies the problem of modeling events that occur over widely-varying time scales.

Each DBN was generated by (1) drawing a number of variables $n \in 2:10$ from a distribution that favors numbers in proportion to their size, (2) drawing an edge probability $p_e \sim$ Uniform$(0, 1)$ and drawing each of the n^2 possible forward DBN edges IID as Bernoulli(p_e) to create a bipartite graph representing two timesteps, (3) creating conditional probability tables by sampling a probability $p \sim$ Uniform$(0, 1)$ for each setting of a node's parents, and (4) rejecting any DBN with edge density ($|E|/n^2$) not in $[0.1, 0.9]$ (which kept the class skew less than $9:1$).

The synthetic data consisted of datasets sampled from 1k random DBNs. Each dataset had 10k timelines; each timeline had 10 timesteps. Experiments were performed on the first 100, 1k, and 10k timelines of each dataset to assess statistical efficiency. The number of DBNs was determined by a power calculation for a 0.9 probability of detecting a PR area difference of 0.01 at $\alpha = 0.01$ with a two-tailed paired t-test. In total, there were 120k experiments: 1k random DBNs, 4 data treatments, 2 data representations, 3 data sizes, and 5 methods (PC, TMN-PC, BNF-DBN, TMN-DBN, TMN-Bf3).

1.3 OMOP Experiments

The number of replicates was determined by a power calculation for a 0.9 probability of detecting a PR area difference of 0.05 at $\alpha = 0.01$ with a two-tailed paired t-test.

2. Results

The following tables and figures include additional analysis of the experimental results. Table 2 contains a linear regression on the synthetic data results that includes interactions between methods and data regimes, and Table 3 contains the detailed results of the pairwise comparisons and their statistical significance. The names and descriptions of the experimental data regimes are in Table 1.

Name	Data Regime
Plain	unaltered synthetic timelines
Noisy	synthetic timelines with noise
MisTs	synthetic timelines with missing timesteps
Cnfdr	synthetic timelines with hidden confounders
OMOP	condensed OMOP timelines

Table 1: Experimental data regimes

References

S. Plis, D. Danks, C. Freeman, and V. Calhoun. Rate-agnostic (causal) structure learning. In *Advances in Neural Information Processing Systems 28*, 2015.

Rank	X	$\hat{\beta}$	se($\hat{\beta}$)	TStat	P-Value
1	BNF-DBN? * cnfdr	-0.713	0.0139	-51.2	0
2	density, e/n^2	0.663	0.0117	56.9	0
3	TMN-PC? * cnfdr	-0.553	0.0139	-39.7	0
4	PC? * cnfdr	-0.499	0.0139	-35.9	1.69e-280
5	TMN-DBN? * cnfdr	-0.475	0.0139	-34.1	7.15e-254
6	TMN-Bf3? * cnfdr	-0.454	0.0139	-32.6	3.16e-232
7	BNF-DBN? * noise	-0.414	0.00547	-75.7	0
8	BNF-DBN?	0.400	0.00188	213	0
9	TMN-PC?	0.309	0.00188	165	0
10	BNF-DBN? * mists	-0.305	0.00537	-56.8	0
11	PC? * noise	-0.287	0.00547	-52.5	0
12	TMN-PC? * noise	-0.281	0.00547	-51.4	0
13	TMN-Bf3? * noise	-0.251	0.00547	-45.8	0
14	PC? * mists	-0.248	0.00537	-46.1	0
15	PC?	0.245	0.00188	131	0
16	TMN-DBN? * noise	-0.225	0.00547	-41.1	0
17	TMN-DBN?	0.217	0.00188	115	0
18	TMN-PC? * mists	-0.216	0.00537	-40.2	0
19	TMN-Bf3?	0.209	0.00188	111	0
20	TMN-DBN? * mists	-0.169	0.00537	-31.5	2.11e-216
21	TMN-Bf3? * mists	-0.157	0.00537	-29.2	1.26e-186
22	# cnfdr $/n$	0.130	0.00996	13.1	3.99e-39
23	log # data	0.0747	0.000430	174	0
24	missingness	-0.0227	0.00380	-5.99	2.16e-09
25	noise level	-0.0213	0.00387	-5.50	3.79e-08
26	intercept	-0.0121	0.00698	-1.73	0.0833
27	avg in-deg	-0.0109	0.00820	-1.33	0.184
28	max in-deg	-0.00795	0.000607	-13.1	3.39e-39
29	# edges, e	0.00533	0.000421	12.7	1.07e-36
30	# nodes, $2n$	-0.00167	0.00117	-1.42	0.156
31	# V-structures	-0.00138	6.06e-05	-22.7	1.06e-113
32	condensed?	0.000733	0.000702	1.04	0.296
33	max edges, n^2	-0.000375	0.000184	-2.04	0.0417
34	max out-deg	-0.000243	0.000638	-0.381	0.703
35	# CPT θs	3.11e-05	1.86e-06	16.7	1.75e-62

Table 2: Linear regression of PR area on attributes of synthetic DBN experiments, including interactions between method and data regime, ranked by $\hat{\beta}$ magnitude. R^2 = 0.699. The method indicators contrast with random guessing.

Rank	Better	Worse	DiffMeans	TStatistic	P-Value
1	BNF-DBN-Plain	Random-Plain	0.480	92.2	0
2	TMN-PC-Plain	Random-Plain	0.447	89.5	0
3	TMN-Bf3-Plain	Random-Plain	0.362	83.4	0
4	TMN-DBN-Plain	Random-Plain	0.304	81.0	0
5	PC-Plain	Random-Plain	0.398	77.4	0
6	BNF-DBN-MisTs	Random-MisTs	0.367	68.5	0
7	TMN-PC-MisTs	Random-MisTs	0.313	65.5	0
8	BNF-DBN-Noisy	Random-Noisy	0.319	60.2	0
9	BNF-DBN-Cnfdr	Random-Cnfdr	0.260	59.2	0
10	TMN-PC-Noisy	Random-Noisy	0.308	59.1	0
11	TMN-PC-Cnfdr	Random-Cnfdr	0.246	56.8	0
12	TMN-DBN-MisTs	Random-MisTs	0.191	53.1	0
13	TMN-Bf3-MisTs	Random-MisTs	0.228	52.9	0
14	BNF-DBN-Plain	TMN-DBN-Plain	0.176	50.0	0
15	TMN-Bf3-Cnfdr	Random-Cnfdr	0.167	49.9	0
16	BNF-DBN-Noisy	PC-Noisy	0.107	49.8	0
17	TMN-DBN-Noisy	Random-Noisy	0.172	48.6	0
18	TMN-PC-Noisy	PC-Noisy	0.0963	47.1	0
19	BNF-DBN-MisTs	PC-MisTs	0.151	47.0	0
20	PC-Cnfdr	Random-Cnfdr	0.194	46.6	7.41e-322
21	PC-MisTs	Random-MisTs	0.216	46.6	7.91e-322
22	BNF-DBN-Plain	PC-Plain	0.0821	45.2	8.61e-308
23	PC-Noisy	Random-Noisy	0.212	44.1	1.15e-297
24	TMN-PC-Plain	TMN-DBN-Plain	0.143	43.6	5.15e-292
25	TMN-DBN-Cnfdr	Random-Cnfdr	0.140	41.9	2.36e-276
26	BNF-DBN-MisTs	TMN-DBN-MisTs	0.176	41.6	5.19e-273
27	TMN-Bf3-Noisy	Random-Noisy	0.162	40.9	6.11e-266
28	TMN-PC-MisTs	PC-MisTs	0.0973	40.0	1.39e-257
29	BNF-DBN-Noisy	TMN-DBN-Noisy	0.147	37.5	9.95e-234
30	BNF-DBN-Noisy	TMN-Bf3-Noisy	0.157	36.5	1.20e-223
31	TMN-PC-Noisy	TMN-DBN-Noisy	0.136	35.2	2.64e-211
32	BNF-DBN-Cnfdr	PC-Cnfdr	0.0657	34.3	5.39e-203
33	BNF-DBN-Cnfdr	TMN-DBN-Cnfdr	0.120	33.5	1.93e-195
34	TMN-PC-Noisy	TMN-Bf3-Noisy	0.146	33.2	6.38e-193
35	TMN-PC-Plain	PC-Plain	0.0495	33.2	1.45e-192
36	TMN-PC-MisTs	TMN-DBN-MisTs	0.122	32.3	2.95e-184
37	BNF-DBN-MisTs	TMN-Bf3-MisTs	0.139	31.8	9.87e-180
38	TMN-PC-Cnfdr	TMN-DBN-Cnfdr	0.106	30.9	3.77e-172
39	BNF-DBN-Plain	TMN-Bf3-Plain	0.118	30.4	1.65e-167
40	TMN-PC-Cnfdr	PC-Cnfdr	0.0515	29.5	3.56e-159

Table 3: Pairwise comparisons between methods within the five data regimes (Table 1) using two-tailed paired t-tests, ranked by p-value. The row with the lines indicates the significance cutoff of a paper-wise false discovery rate controlled at 0.01 with the Benjamini-Hochberg procedure.

Rank	Better	Worse	DiffMeans	TStatistic	P-Value
41	PC-Plain	TMN-DBN-Plain	0.0934	26.5	1.30e-132
42	BNF-DBN-Cnfdr	TMN-Bf3-Cnfdr	0.0933	24.4	2.49e-115
43	TMN-PC-Plain	TMN-Bf3-Plain	0.0851	22.8	2.17e-102
44	BNF-DBN-Plain	TMN-PC-Plain	0.0327	21.3	8.01e-91
45	TMN-PC-Cnfdr	TMN-Bf3-Cnfdr	0.0790	21.1	1.05e-89
46	TMN-Bf3-Plain	TMN-DBN-Plain	0.0578	20.2	4.54e-83
47	TMN-PC-MisTs	TMN-Bf3-MisTs	0.0855	20.1	5.41e-82
48	BNF-DBN-MisTs	TMN-PC-MisTs	0.0538	17.8	4.57e-66
49	PC-Cnfdr	TMN-DBN-Cnfdr	0.0546	15.8	3.29e-53
50	PC-Noisy	TMN-Bf3-Noisy	0.0493	12.6	7.45e-35
51	BNF-DBN-OMOP	Random-OMOP	0.0580	18.1	3.05e-33
52	TMN-Bf3-MisTs	TMN-DBN-MisTs	0.0368	10.9	5.90e-27
53	PC-Noisy	TMN-DBN-Noisy	0.0396	10.8	2.61e-26
54	TMN-Bf3-Cnfdr	TMN-DBN-Cnfdr	0.0270	10.4	1.19e-24
55	TMN-Bf3-OMOP	Random-OMOP	0.114	12.2	1.70e-21
56	BNF-DBN-Noisy	TMN-PC-Noisy	0.0110	9.11	2.01e-19
57	PC-Plain	TMN-Bf3-Plain	0.0356	8.79	3.29e-18
58	PC-OMOP	Random-OMOP	0.0353	10.4	1.68e-17
59	TMN-PC-OMOP	Random-OMOP	0.0655	10.3	2.73e-17
60	BNF-DBN-Cnfdr	TMN-PC-Cnfdr	0.0142	7.91	4.18e-15
61	TMN-DBN-OMOP	Random-OMOP	0.0524	8.77	5.14e-14
62	PC-Cnfdr	TMN-Bf3-Cnfdr	0.0275	7.45	1.37e-13
63	TMN-Bf3-OMOP	PC-OMOP	0.0783	8.11	1.40e-12
64	BNF-DBN-OMOP	PC-OMOP	0.0227	7.86	4.70e-12
65	PC-MisTs	TMN-DBN-MisTs	0.0249	6.80	1.39e-11
66	TMN-Bf3-OMOP	BNF-DBN-OMOP	0.0556	5.93	4.47e-08
67	TMN-Bf3-OMOP	TMN-DBN-OMOP	0.0612	5.45	3.76e-07
68	TMN-Bf3-OMOP	TMN-PC-OMOP	0.0481	4.40	2.78e-05
69	TMN-DBN-Noisy	TMN-Bf3-Noisy	0.00977	4.06	5.00e-05
70	TMN-PC-OMOP	PC-OMOP	0.0303	4.21	5.59e-05
	—	—	—	—	— BH 0.01
71	TMN-Bf3-MisTs	PC-MisTs	0.0119	2.81	0.00494
72	TMN-DBN-OMOP	PC-OMOP	0.0171	2.52	0.0132
73	TMN-PC-OMOP	TMN-DBN-OMOP	0.0131	1.97	0.0513
74	TMN-PC-OMOP	BNF-DBN-OMOP	0.00755	1.04	0.300
75	BNF-DBN-OMOP	TMN-DBN-OMOP	0.00559	0.822	0.413

Table 4: Pairwise comparisons, part 2 of Table 3