Supplementary Material for:
Causality-based Feature Selection: Methods and Evaluations

KUI YU and XIANJIE GUO, Hefei University of Technology, China
LIN LIU and JIUYONG LI, University of South Australia, Australia
HAO WANG and ZHAOLONG LING, Hefei University of Technology, China
XINDONG WU, Mininglamp Technology, China

S-1: EXPERIMENTS ON SYNTHETIC DATA

In this section, we systematically evaluate the algorithms in the CausalFS package using the four standard benchmark BNs as shown in Table 1. Using the four benchmark networks, we randomly generate two groups of data, one including five datasets with 500 data instances each, and the other also containing five datasets with 5,000 data instances each. Using the two groups of datasets, we evaluate the causality-based feature selection algorithms for learning MB and PC, respectively.

Table 1. Summary of Benchmark BNs

Network	Num. Vars	Num. Edges	Max In/out-Degree	Min/Max |PCset |	Variable Domain
Child	20	25	2/7	1/8	2-6
Alarm	37	46	4/5	1/6	2-4
Hailfinder	56	66	4/16	1/17	2-11
Pigs	441	592	2/39	1/41	3-3
Gene	801	972	4/10	0/11	3-5

All experiments were conducted on a computer with Intel(R) i5-2600, 3.4 GHz CPU, and 8 GB memory. In the following experiments, the G^2 test is used for independence tests and the significance level for the tests is set to 0.05 and 0.01, respectively. In all tables reporting experiment results, "-" denotes that an algorithm cannot deal with a dataset due to expensive computations (the running time of the algorithm exceeded the three-day time threshold and it was stopped).

The following metrics are used to compare the MB or PC of a variable learnt by a causality-based feature selection algorithms with the true MB or PC of the variable in the BN.

- Precision. The number of true positives in the output (i.e., the features in the output belonging to the true MB or PC of a variable) divided by the number of features in the MB or PC output by an algorithm. Precision reports the false positive rate in the output of an algorithm.
- Recall. The number of true positives in the output divided by the total number of true positives (the size of the true MB or PC) Recall reports the true positive rate in the output of an algorithm.
- $F1 = 2 * precision * recall/(precision + recall)$. It is the harmonic average of the precision and recall, where $F1 = 1$ is the best case (perfect precision and recall) while $F1 = 0$ is the worst case.

0360-0300/2020/09-ART111 $15.00
https://doi.org/10.1145/3409382

Table 2. Results of PC Learning Methods on Synthetic Datasets (size = 500)

Network	Algorithm	F1	Precision	Recall	Time
Child	PC-simple	0.89/0.89	0.94/0.91	0.86/0.88	0/0
	MMPC	0.90/0.87	0.93/0.87	0.89/0.89	0/0
	HITON-PC	0.91/0.88	0.94/0.89	0.89/0.91	0/0
	Semi-HITON-PC	0.91/0.89	0.94/0.90	0.89/0.91	0/0
	GetPC	0.84/0.84	0.93/0.91	0.80/0.81	0/0
	MBtoPC	0.86/0.83	0.95/0.90	0.81/0.80	0/0
	SLL-PC	0.87	0.95	0.83	0.47
	S^2TMB-PC	0.86	0.91	0.85	0.15
Alarm	PC-simple	0.82/0.84	0.89/0.88	0.79/0.84	0/0
	MMPC	0.85/0.84	0.90/0.84	0.83/0.87	0/0
	HITON-PC	0.84/0.83	0.89/0.83	0.83/0.87	0/0
	Semi-HITON-PC	0.85/0.84	0.89/0.85	0.83/0.87	0/0
	GetPC	0.77/0.82	0.88/0.89	0.73/0.80	0/0
	MBtoPC	0.86/0.85	0.95/0.92	0.81/0.82	0/0
	SLL-PC	0.91	0.93	0.91	0.38
	S^2TMB-PC	0.88	0.87	0.93	0.31
Pigs	PC-simple	0.99/0.95	0.98/0.93	1/1	0.02/0.03
	MMPC	0.91/0.77	0.86/0.66	1/1	0/0.01
	HITON-PC	0.91/0.77	0.86/0.66	1/1	0.01/0.01
	Semi-HITON-PC	0.91/0.77	0.87/0.67	1/1	0.01/0.01
	GetPC	0.99/0.95	0.98/0.92	1/1	0.04/0.05
	MBtoPC	0.93/0.89	0.97/0.90	0.92/0.92	0.02/0.03
	SLL-PC	-	-	-	-
	S^2TMB-PC	-	-	-	-
Gene	PC-simple	0.96/0.91	0.97/0.89	0.96/0.96	0.01/0.02
	MMPC	0.83/0.70	0.79/0.60	0.92/0.93	0.01/0.01
	HITON-PC	0.83/0.70	0.79/0.60	0.92/0.93	0.01/0.01
	Semi-HITON-PC	0.83/0.71	0.79/0.61	0.92/0.93	0.01/0.01
	GetPC	0.96/0.94	0.97/0.94	0.95/0.96	0.03/0.04
	MBtoPC	0.86/0.80	0.91/0.83	0.84/0.85	0.03/0.05
	SLL-PC	-	-	-	-
	S^2TMB-PC	-	-	-	-

- Efficiency. We use the time consumed in seconds to measure the efficiency of an algorithm.

For an algorithm, we first use it to learn the MB or PC of each variable in a dataset, then compute the average results of recall, precision, F1, and running time over all variables in the dataset, and finally we report these average results over five datasets. In all tables of experiment results, a value pair A/B indicates the precision, recall, or $F1$ of an algorithm when the significance level of independence tests is set to 0.01 (represented by A) and 0.05 (represented by B), respectively.

Results of PC learning. From Tables 2 and 3, for the time efficiency, we have the following observations:

Table 3. Results of PC Learning Methods on Synthetic Datasets (size = 5,000)

Network	Algorithm	F1	Precision	Recall	Time
Child	PC-simple	1/0.96	0.99/0.94	1/1	0.01/0.02
	MMPC	0.98/0.92	0.96/0.87	1/1	0.01/0.01
	HITON-PC	0.98/0.92	0.97/0.88	1/1	0.01/0.01
	Semi-HITON-PC	0.98/0.92	0.97/0.88	1/1	0.01/0.01
	GetPC	1/1	1/0.99	1/1	0.05/0.05
	MBtoPC	0.97/0.93	0.98/0.92	0.97/0.97	0.02/0.02
	SLL-PC	0.97	1	0.94	5.2
	S^2TMB-PC	0.98	0.99	0.97	1.51
Alarm	PC-simple	0.98/0.97	0.97/0.96	0.99/0.99	0.02/0.02
	MMPC	0.97/0.93	0.97/0.90	0.99/0.99	0.01/0.01
	HITON-PC	0.97/0.92	0.97/0.89	0.99/0.99	0.01/0.01
	Semi-HITON-PC	0.97/0.93	0.97/0.90	0.99/0.99	0.01/0.01
	GetPC	0.98/0.98	1/1	0.97/0.97	0.04/0.04
	MBtoPC	0.98/0.96	0.99/0.97	0.97/0.97	0.03/0.03
	SLL-PC	0.97	0.96	0.98	3.49
	S^2TMB-PC	0.95	0.94	0.98	2.83
Pigs	PC-simple	0.99/0.95	0.98/0.93	1/1	0.64/0.78
	MMPC	0.91/0.78	0.86/0.68	1/1	0.19/0.2
	HITON-PC	0.91/0.78	0.86/0.68	1/1	0.24/0.26
	Semi-HITON-PC	0.91/0.79	0.86/0.69	1/1	0.29/0.3
	GetPC	0.99/0.97	0.99/0.96	1/1	10.23/10.43
	MBtoPC	0.97/0.95	0.98/0.94	0.99/0.99	0.52/0.76
	SLL-PC	-	-	-	-
	S^2TMB-PC	-	-	-	-
Gene	PC-simple	0.98/0.92	0.97/0.90	0.99/0.98	0.40/0.59
	MMPC	0.83/0.72	0.78/0.62	0.94/0.94	0.12/0.13
	HITON-PC	0.83/0.72	0.78/0.62	0.94/0.94	0.15/0.19
	Semi-HITON-PC	0.83/0.72	0.78/0.63	0.94/0.94	0.15/0.19
	GetPC	0.98/0.96	0.98/0.95	0.99/0.98	0.64/0.87
	MBtoPC	0.91/0.90	0.91/0.89	0.93/0.93	0.69/1.22
	SLL-PC	-	-	-	-
	S^2TMB-PC	-	-	-	-

- An algorithm using a forward strategy for PC learning may be faster than an algorithm using the backward strategy. Tables 2 and 3 illustrate that among the six constraint-based PC learning methods, MMPC, HITON-PC, and semi-HITON-PC are very competitive in terms of efficiency, but they are faster than the other PC algorithms. The explanation is that MMPC, HITON-PC, and semi-HITON-PC use a forward strategy, while Recognize-PC uses a backward strategy.
- The symmetry check makes a PC learning algorithm computationally expensive. GetPC uses a forward strategy, but it is not efficient, since it employs the symmetry check using the AND rule in the PC learning phase. MBtoPC is slower than the other algorithms, because it performs the symmetry check using the OR rule. SLL-PC implements the symmetry check

for both PC and spouse learning. S^2TMB-PC is faster than SLL-PC, since it does not perform the symmetry check.

- A score-based PC learning algorithm may be slower than a constraint-based method due to the computational complexity of the BN structure learning. In Tables 2 and 3, both score-based methods are slower than any constraint-based PC learning algorithms and failed to produce results for the two large-sized networks before they were terminated after three days.

In terms of learning performance, the following conclusions are obtained:

- A PC learning algorithm using the backward strategy or the symmetry check can remove more false positives, especially on large-sized BN networks. On the two small-sized networks, Recognize-PC, GetPC, and MBtoPC are very competitive with MMPC, HITON-PC, and semi-HITON-PC in terms of precision and recall. On the two large-sized networks, Recognize-PC, GetPC, and MBtoPC significantly outperform MMPC, HITON-PC, and semi-HITON-PC in terms of precision. This indicates that on the large-sized networks, learning PC using a backward strategy or the symmetry check can remove much more false positives.
- A PC learning algorithm using the backward strategy or the symmetry check is not affected by the significance level in terms of precision (i.e., false positive rate). In Tables 2 and 3, the recall metric is not affected much by the significance level parameter, but the precision metrics of MMPC, HITON-PC, and semi-HITON-PC are influenced greatly by this parameter, while Recognize-PC, GetPC, and MBtoPC are not affected. Meanwhile, for the learning performance, SLL-PC (performing the symmetry check) and S^2TMB-PC are very competitive to PC-simple, GetPC, and MBtoPC on the two small-sized networks.

In summary, the backward strategy or the symmetry check is a double-edged sword. They will make a PC learning algorithm computationally expensive, but they are able to make the PC learning algorithms output a more accurate PC set.

Results of MB learning. From Tables 4 and 5, we have the following conclusions:

- Existing constraint-based simultaneous MB learning algorithms are inferior to the other MB learning algorithms. From Tables 4 and 5, we can see that regardless of sample size (500 or 5,000), on the four BN networks, the six simultaneous MB learning algorithms are inferior to the remaining 10 MB learning algorithms, especially on the two large-sized BN networks. Meanwhile, GSMB has the worst precisions and recalls among all rivals. STMB has the lowest precision among all the MB learning algorithms except for GSMB. The explanation is that STMB adopts the same strategy as the simultaneous MB discovery algorithms in the false positive removal phase and this leads to large data sample requirements.
- The backward strategy and the symmetry check make an MB learning algorithm more accurate using a large-sized data samples. Since GetPC, Recognize-PC, and MBtoPC outperform the other PC learning algorithms as shown in Tables 3 and 4, their corresponding MB learning algorithms, PCMB, IPCMB, and MBOR, are better than the other constraint-based 11 MB learning algorithms on the two large-sized BN networks. For example, IPCMB achieves the best performance on the two large networks, since IPCMB not only adopts the Recognize-PC algorithm to learn the PC of each variable, but also it performs the symmetry check in the spouse discovery procedure (Recognize-PC gets the best performance as shown in Tables 2 and 3).

  Duo to unreliable independence tests using small-sized data samples, an MB learning algorithm using the AND rule may remove the true MB members from its output when

Table 4. Results of MB Learning Methods on Synthetic Datasets (size = 500)

Data	Algorithm	F1	Precision	Recall	Time
Child	GSMB	0.69/0.54	0.78/0.59	0.67/0.58	**0/0**
	IAMB	0.78/0.67	0.86/0.67	0.77/0.77	**0/0**
	Inter-IAMB	0.78/0.67	0.86/0.67	0.77/0.77	**0/0**
	Fast-IAMB	0.74/0.68	0.87/0.77	0.71/0.70	**0/0**
	LRH	0.80/0.68	0.86/0.67	0.80/0.82	**0/0**
	FBED	0.79/0.70	0.87/0.71	0.77/0.77	**0/0**
	MMMB	0.87/0.82	0.94/0.84	0.83/0.84	**0/0**
	PCMB	0.79/0.79	0.93/0.88	0.74/0.76	**0/0**
	HITON-MB	0.85/0.83	0.94/0.86	0.82/0.85	**0/0**
	Semi-HITON-MB	0.85/0.83	0.94/0.87	0.82/0.85	**0/0**
	MBOR	0.84/0.81	0.92/0.83	0.81/0.84	**0/0**
	IPCMB	0.80/0.82	0.93/0.88	0.74/0.79	**0/0**
	STMB	0.81/0.71	0.85/0.66	0.82/0.86	**0/0**
	BAMB	0.86/0.81	0.93/0.82	0.83/0.84	**0/0**
	EEMB	0.84/0.80	0.92/0.81	0.82/0.85	**0/0**
	SLL	0.84	**0.95**	0.79	0.85
	S^2TMB	0.83	**0.95**	0.77	0.14
	fGES-MB	**0.88**	0.92	**0.88**	0.04
Alarm	GSMB	0.29/0.19	0.35/0.21	0.27/0.20	**0/0**
	IAMB	0.76/0.73	0.89/0.79	0.71/0.74	**0/0**
	Inter-IAMB	0.77/0.72	0.89/0.77	0.72/0.75	**0/0**
	Fast-IAMB	0.71/0.67	0.85/0.77	0.65/0.66	**0/0**
	LRH	0.75/0.70	0.85/0.71	0.73/0.76	**0/0**
	FBED	0.77/0.74	0.90/0.80	0.71/0.73	**0/0**
	MMMB	0.82/0.82	0.90/0.85	0.78/0.84	**0/0**
	PCMB	0.74/0.79	0.88/0.89	0.68/0.76	**0**/0.01
	HITON-MB	0.82/0.81	0.90/0.84	0.78/0.85	**0/0**
	Semi-HITON-MB	0.82/0.83	0.90/0.86	0.78/0.85	**0/0**
	MBOR	0.85/0.85	0.93/0.89	0.81/0.85	**0/0**
	IPCMB	0.74/0.80	0.87/0.88	0.69/0.77	**0**/0.01
	STMB	0.67/0.59	0.69/0.54	0.76/0.84	**0/0**
	BAMB	0.80/0.80	0.90/0.85	0.75/0.80	**0/0**
	EEMB	0.80/0.79	0.90/0.85	0.75/0.87	**0/0**
	SLL	**0.88**	**0.93**	0.87	0.88
	S^2TMB	0.87	0.92	0.86	0.29
	fGES-MB	0.73	0.66	**0.91**	0.22
	GSMB	0.06/0.01	0.08/0.02	0.05/0.01	**0/0**
	IAMB	0.80/0.80	0.98/0.98	0.72/0.72	0.01/0.01
	Inter-IAMB	0.80/0.80	0.98/0.98	0.72/0.72	0.01/0.01
	Fast-IAMB	0.77/0.76	0.80/0.80	0.83/0.81	0.01/0.01
	LRH	0.69/0.68	0.90/0.87	0.60/0.60	0.01/0.02
	FBED	0.71/0.69	0.89/0.87	0.64/0.62	**0/0**

(Continued)

Table 4. Continued

Data	Algorithm	F1	Precision	Recall	Time
Pigs	MMMB	0.93/0.76	0.88/0.64	**1/1**	0.04/0.05
	PCMB	**0.99**/0.95	0.98/0.92	**1/1**	0.13/0.17
	HITON-MB	0.93/0.76	0.88/0.64	**1/1**	0.07/0.08
	Semi-HITON-MB	0.93/0.77	0.88/0.65	**1/1**	0.08/0.1
	MBOR	0.95/0.85	0.92/0.77	**1/1**	0.02/0.04
	IPCMB	**0.99**/0.97	**0.99**/0.95	**1/1**	0.60/0.67
	STMB	0.59/0.25	0.47/0.15	0.97/0.98	0.05/0.08
	BAMB	0.96/0.81	0.93/0.71	**1/1**	0.05/0.06
	EEMB	0.96/0.82	0.93/0.72	**1/1**	0.04/0.08
	SLL	-	-	-	-
	S^2TMB	-	-	-	-
	fGES-MB	0.98	0.97	**1**	8.17
Gene	GSMB	0.06/0.01	0.08/0.02	0.05/0.01	**0/0**
	IAMB	0.66/0.66	0.80/0.80	0.68/0.68	0.01/0.01
	Inter-IAMB	0.66/0.66	0.80/0.80	0.68/0.68	0.01/0.01
	Fast-IAMB	0.66/0.65	0.68/0.68	0.78/0.78	0.01/0.01
	LRH	0.71/0.67	0.89/0.82	0.67/0.67	0.01/0.02
	FBED	0.66/0.66	0.80/0.80	0.68/0.67	0.01/0.01
	MMMB	0.82/0.66	0.79/0.56	0.91/0.92	0.03/0.04
	PCMB	0.94/0.93	0.97/0.93	0.93/0.95	0.08/0.12
	HITON-MB	0.82/0.66	0.79/0.56	0.91/0.92	0.03/0.05
	Semi-HITON-MB	0.82/0.67	0.79/0.57	0.91/0.92	0.03/0.05
	MBOR	0.87/0.74	0.86/0.66	0.91/0.92	0.03/0.06
	IPCMB	**0.96**/0.94	**0.98**/0.94	0.94/0.96	0.04/0.07
	STMB	0.56/0.18	0.44/0.12	0.92/0.92	0.04/0.06
	BAMB	0.82/0.68	0.79/0.59	0.90/0.91	0.03/0.05
	EEMB	0.82/0.67	0.78/0.57	0.90/0.92	0.03/0.05
	SLL	-	-	-	-
	S^2TMB	-	-	-	-
	fGES-MB	0.95	0.94	**0.97**	7.34

Table 5. Results of MB Learning Methods on Synthetic Datasets (size = 5,000)

Data	Algorithm	F1	Precision	Recall	Time
	GSMB	0.70/0.45	0.69/0.42	0.75/0.58	**0/0**
	IAMB	0.86/0.72	0.83/0.64	0.95/0.95	0.01/0.01
	Inter-IAMB	0.86/0.72	0.83/0.65	0.95/0.95	0.01/0.01
	Fast-IAMB	0.89/0.74	0.89/0.70	0.93/0.91	0.01/0.01
	LRH	0.85/0.71	0.81/0.63	0.97/0.96	0.02/0.03
	FBED	0.89/0.76	0.87/0.69	0.94/0.94	0.01/0.01
	MMMB	0.98/0.89	0.97/0.84	**1/1**	0.05/0.05

(Continued)

Table 5. Continued

Data	Algorithm	F1	Precision	Recall	Time
	PCMB	**1**/0.98	**1**/0.97	**1/1**	0.1/0.11
Child	HITON-MB	0.98/0.90	0.97/0.85	**1/1**	0.06/0.07
	Semi-HITON-MB	0.98/0.90	0.97/0.85	**1/1**	0.07/0.08
	MBOR	0.96/0.88	0.96/0.84	0.98/0.98	0.04/0.08
	IPCMB	**1**/0.98	**1**/0.97	**1/1**	0.09/0.11
	STMB	0.91/0.80	0.86/0.71	0.99/**1**	0.02/0.03
	BAMB	0.97/0.89	0.96/0.83	0.99/0.99	0.04/0.05
	EEMB	0.96/0.88	0.95/0.82	0.99/0.99	0.02/0.02
	SLL	0.96	**1**	0.93	9.1
	S^2TMB	0.98	**1**	0.96	1.3
	fGES-MB	**1**	**1**	**1**	0.14
	GSMB	0.33/0.20	0.34/0.20	0.35/0.26	0.01/**0**
	IAMB	0.92/0.80	0.94/0.75	0.91/0.92	0.02/0.02
	Inter-IAMB	0.92/0.79	0.94/0.72	0.92/0.93	0.02/0.03
	Fast-IAMB	0.91/0.79	0.94/0.75	0.91/0.92	0.01/0.01
	LRH	0.91/0.78	0.93/0.71	0.92/0.93	0.13/0.05
	FBED	0.93/0.83	0.96/0.80	0.91/0.92	0.01/0.01
	MMMB	0.97/0.93	0.98/0.90	0.97/0.98	0.04/0.04
	PCMB	0.97/**0.98**	**1/1**	0.96/0.96	0.08/0.09
Alarm	HITON-MB	0.97/0.92	0.98/0.89	0.97/**0.98**	0.04/0.05
	Semi-HITON-MB	0.97/0.93	0.98/0.9	0.97/**0.98**	0.05/0.06
	MBOR	0.97/0.96	0.98/0.95	0.97/**0.98**	0.05/0.06
	IPCMB	0.97/**0.98**	**1/1**	0.96/0.97	0.09/0.1
	STMB	0.79/0.71	0.76/0.63	0.95/0.97	0.04/0.04
	BAMB	0.96/0.92	0.98/0.91	0.94/0.96	0.03/0.04
	EEMB	0.96/0.93	0.99/0.93	0.94/0.95	0.02/0.03
	SLL	0.96	0.97	0.96	8.15
	S^2TMB	0.95	0.98	0.93	2.44
	fGES-MB	0.88	0.85	0.96	0.5
	GSMB	0.05/0.02	0.04/0.02	0.06/0.03	0.02/**0.01**
	IAMB	0.71/0.71	0.62/0.62	0.96/0.96	0.29/0.29
	Inter-IAMB	0.71/0.71	0.62/0.62	0.96/0.96	0.28/0.28
	Fast-IAMB	0.71/0.70	0.62/0.61	0.95/0.94	0.18/0.14
	LRH	0.80/0.76	0.75/0.68	0.96/0.96	0.87/1.22
	FBED	0.78/0.67	0.71/0.59	0.95/0.89	0.12/0.11
	MMMB	0.92/0.77	0.87/0.65	**1/1**	6.47/6
	PCMB	0.99/0.97	0.99/0.95	**1/1**	29.15/23.12
Pigs	HITON-MB	0.92/0.76	0.87/0.64	**1/1**	9.74/12.27
	Semi-HITON-MB	0.92/0.77	0.87/0.65	**1/1**	13.58/13.6
	MBOR	0.97/0.91	0.94/0.86	**1/1**	0.73/1.19
	IPCMB	**1**/0.98	**1**/0.97	**1/1**	11.17/11.37
	STMB	0.38/0.15	0.28/0.09	**1/1**	2.66/5.2
	BAMB	0.96/0.84	0.93/0.74	**1/1**	23.35/23.28
	EEMB	0.96/0.87	0.94/0.79	**1/1**	8.71/8.27

(Continued)

Table 5. Continued

Data	Algorithm	F1	Precision	Recall	Time
	SLL	-	-	-	-
	S^2TMB	-	-	-	-
	fGES-MB	-	-	-	-
Gene	GSMB	0.03/0.01	0.03/0.01	0.04/0.02	0.02/**0.01**
	IAMB	0.60/0.60	0.53/0.53	0.89/0.89	0.65/0.64
	Inter-IAMB	0.60/0.60	0.53/0.53	0.89/0.89	0.71/0.64
	Fast-IAMB	0.60/0.59	0.53/0.52	0.89/0.88	0.35/0.28
	LRH	0.66/0.61	0.60/0.55	0.89/0.88	0.98/3.3
	FBED	0.62/0.55	0.55/0.48	0.88/0.82	0.24/0.19
	MMMB	0.83/0.68	0.77/0.57	0.94/0.94	0.58/0.69
	PCMB	0.98/0.96	0.98/0.95	**0.99**/0.98	1.67/2.13
	HITON-MB	0.83/0.68	0.77/0.57	0.94/0.94	0.76/1.1
	Semi-HITON-MB	0.83/0.69	0.77/0.58	0.94/0.94	0.73/1.08
	MBOR	0.89/0.83	0.86/0.78	0.94/0.94	0.85/1.83
	IPCMB	**0.99**/0.97	**0.99**/0.96	**0.99/0.99**	1.92/2.96
	STMB	0.30/0.11	0.20/0.07	**0.99**/0.98	0.95/1.57
	BAMB	0.82/0.69	0.76/0.59	0.94/0.94	0.71/1.55
	EEMB	0.82/0.71	0.77/0.61	0.94/0.94	0.56/0.82
	SLL	-	-	-	-
	S^2TMB	-	-	-	-
	fGES-MB	-	-	-	-

performing the symmetry check, and thus this may degrade the performance of the algorithm on recall. For example, in Table 2, using 500 samples on the two small-sized networks, PCMB and IPCMB get much lower recall than MMMB, HITON-MB, and semi-HITON-MB. Thus, it is an interesting problem of studying under what conditions we use the AND rule or the OR rule or combining both to make MB learning more accurate.

- Score-based methods have the similar performance as constraint-based methods in terms of F1. From Tables 4 and 5, S^2TMB and SLL achieve similar performance as the best constraint-based MB learning methods in terms of both precision and recall with the two small-sized networks. These results are consistent with the results of S^2TMB-PC and SLL-PC in Tables 3 and 4. fGES-MB has a very high performance in terms of recall, and its efficiency is also the highest in score-based methods. Due to expensive computation costs, S^2TMB and SLL did not produce the MBs of all variables of the Pigs and Gene networks within three days, respectively. fGES-MB can run perfectly and its recall is the highest among all MB learning algorithms, and it can learn the MBs from the Pigs and Gene networks with 500 data samples. But when we use the fGES-MB algorithm to learn the MB of each variable in these two networks independently, we found that fGES-MB cannot produce the MBs of some variables with 5,000 data samples in reasonable time (we ran fGES-MB and fGES using their original implementations at http://github.com/cmu-phil/tetrad.), while the fGES algorithm can fast learn the Gene and Pigs network structures with 5,000 data samples.
- Simultaneous MB learning methods are the fastest, while score-based methods are the slowest. Since the six simultaneous MB learning methods do not need to perform an exhaustive subset search within the currently selected features, they are faster than the eight

divide-and-conquer MB methods. In addition, these six simultaneous MB learning algorithms have similar performance in terms of efficiency. PCMB and IPCMB are the slowest among the divide-and-conquer MB methods, since they need to perform the symmetry check. Moreover, the Recognize-PC algorithm used by IPCMB is slower than MMPC, HITON-PC, and semi-HITON-PC as shown in Tables 2 and 3. STMB and BAMB do not need to perform the symmetry check. However, they need to perform an exhaustive subset search for PC learning. These two algorithms are not significantly faster than MMMB, HITON-MB, and semi-HITON-MB. fGES-MB, S^2TMB, and SLL are slower than all constraint-based MB learning algorithms, since they need to use a BN structure learning algorithm to achieve MBs. Among three score-based methods, fGES-MB is the fastest. S^2TMB is faster than SLL, because S^2TMB removes the symmetry check.

S-2: EXPERIMENTS ON SYNTHETIC DATA WITH DENSE VARIABLES

The time complexity of existing causality-based feature selection algorithms is mainly determined by the size of the MB. Thus, in this section we evaluate the performance of existing causality-based feature selection algorithms using dense variables. A dense variable means that it has a large-sized MB or has both the large-sized MB and a large number of discrete values. If a causality-based feature selection algorithm cannot learn a large-sized MB in a reasonable time, it cannot deal with dense networks.

In this section, using the two standard benchmark BNs, Pigs and Hailfinder, we select variable 435 and variable 3, respectively, for using existing causality-based feature selection algorithms to learn their MBs. The MB of 435 contains 119 variables and the MB of variable 3 has 33 variables. Moreover, variable 3 takes 11 distinct values. We randomly generate two groups of data, one including 5 datasets with 500 data instances each, and the other also containing 5 datasets with 5,000 data instances each. Using the two groups of datasets, we evaluate the causality-based feature selection algorithms for learning MB with these two dense variables and report the average results of recall, precision, F1, and running time over five datasets. From Tables 6 and 7, we have the following conclusions:

- When a variable has a large-sized MB and a large number of discrete values, all causality-based feature selection algorithms cannot produce good results. For the Hailfinder dataset, variable 3 not only has a large-sized MB, but also takes 11 discrete values. This leads to many unreliable independence tests when the sample size is limited. Thus, all MB algorithms in Tables 6 and 7 achieve poor results for learning the MB of variable 3. In addition, either the backward strategy or the symmetry check makes the MB algorithms have worse results, such as PCMB, IPCMB, and STMB as shown in Tables 6 and 7.
- On Pigs, since the MB size of variable 435 is large, S^2TMB, and SLL are unable to produce the MB of variable 435 in the three days. fGES-MB not only has the highest recall, but also runs significantly faster than the dvide-and-conquer MB algorithms with large-sized data samples as shown in Tables 6 and 7. Due to large-sized MBs, simultaneous MB learning methods achieve the worst performance with both large-sized and small-sized samples.

Table 6. Results of MB Learning Methods on Dense Variables (size = 500)

Data	Algorithm	F1	Precision	Recall	Time
Hailfinder	GSMB	0.12/0.10	0.40/0.33	0.07/0.06	**0/0**
	IAMB	0.14/0.14	0.47/0.47	0.08/0.08	**0/0**
	Inter-IAMB	0.14/0.14	0.47/0.47	0.08/0.08	**0/0**
	Fast-IAMB	0.04/0.04	0.20/0.20	0.02/0.02	**0/0**
	LRH	0.15/0.14	0.45/0.40	0.09/0.08	**0/0**
	FBED	0.14/0.16	0.6/0.63	0.08/0.09	**0/0**
	MMMB	0.11/0.13	**1**/0.9	0.06/0.07	**0/0**
	PCMB	0/0	0/0	0/0	**0/0**
	HITON-MB	0.11/0.13	**1**/0.9	0.06/0.07	**0/0**
	Semi-HITON-MB	0.11/0.13	**1**/0.9	0.06/0.07	**0/0**
	MBOR	0.06/0.08	0.23/0.27	0.04/0.05	**0/0**
	IPCMB	0/0	0/0	0/0	**0/0**
	STMB	0/0	0/0	0/0	**0/0**
	BAMB	0.16/**0.18**	0.67/0.60	0.09/**0.11**	**0/0**
	EEMB	0.16/**0.18**	0.67/0.60	0.09/**0.11**	**0/0**
	SLL	0	0	0	0.02
	S^2TMB	0	0	0	0.01
	fGES-MB	0.07	0.60	0.04	0.04
Pigs	GSMB	0.01/0	0.10/0	0/0	**0/0**
	IAMB	0.06/0.06	**1/1**	0.03/0.03	**0/0**
	Inter-IAMB	0.06/0.06	**1/1**	0.03/0.03	**0/0**
	Fast-IAMB	0.08/0.08	**1/1**	0.04/0.04	**0/0**
	LRH	0.06/0.06	**1/1**	0.03/0.03	0.06/0.08
	FBED	0.06/0.06	**1/1**	0.03/0.03	**0/0**
	MMMB	**1**/0.98	0.99/0.96	**1/1**	0.22/0.26
	PCMB	**1**/0.99	**1**/0.99	**1/1**	0.42/0.53
	HITON-MB	0.99/0.97	0.99/0.93	**1/1**	0.41/0.43
	Semi-HITON-MB	**1**/0.97	0.99/0.94	**1/1**	0.5/0.52
	MBOR	0.98/0.96	0.98/0.94	0.99/0.97	0.29/0.38
	IPCMB	**1**/0.99	0.99/0.98	**1/1**	3.67/4.46
	STMB	0.60/0.49	0.56/0.39	0.64/0.66	6.25/10.23
	BAMB	**1**/0.99	**1**/0.99	**1**/0.99	8.06/8.32
	EEMB	0.97/0.89	0.95/0.80	**1/1**	8.84/12.35
	SLL	-	-	-	-
	S^2TMB	-	-	-	-
	fGES-MB	0.94	0.88	**1**	1.18

Table 7. Results of MB Learning Methods on Dense Variables (size = 5,000)

Data	Algorithm	F1	Precision	Recall	Time
Hailfinder	GSMB	0.20/0.20	0.44/0.44	0.13/0.13	**0/0**
	IAMB	**0.26/0.26**	0.61/0.61	0.16/0.16	0.01/0.01
	Inter-IAMB	**0.26/0.26**	0.61/0.61	0.16/0.16	0.01/0.01
	Fast-IAMB	0.10/0.10	0.32/0.32	0.06/0.06	0.01/0.01
	LRH	**0.26**/0.19	0.53/0.38	**0.18**/0.13	0.01/0.03
	FBED	0.20/0.23	0.67/0.63	0.12/0.14	0.01/0.01
	MMMB	0.11/0.13	**1**/0.73	0.06/0.07	0.01/0.01
	PCMB	0/0	0/0	0/0	0.01/0.01
	HITON-MB	0.11/0.13	**1**/0.73	0.06/0.07	0.01/0.01
	Semi-HITON-MB	0.11/0.13	**1**/0.73	0.06/0.07	0.01/0.01
	MBOR	0.10/0.10	0.33/0.32	0.06/0.06	0.01/0.01
	IPCMB	0/0	0/0	0/0	**0/0**
	STMB	0/0	0/0	0/0	**0/0**
	BAMB	0.20/0.20	0.67/0.67	0.12/0.12	0.01/0.01
	EEMB	0.20/0.20	0.67/0.67	0.12/0.12	0.01/0.01
	SLL	0	0	0	0.17
	S^2TMB	0	0	0	0.10
	FGES-MB	0.16	0.53	0.09	0.16
Pigs	GSMB	0/0	0/0	0/0	**0.01/0.01**
	IAMB	0.14/0.14	**1/1**	0.07/0.07	0.48/0.47
	Inter-IAMB	0.14/0.14	**1/1**	0.07/0.07	0.48/0.50
	Fast-IAMB	0.14/0.14	**1/1**	0.07/0.07	0.07/0.08
	LRH	0.14/0.14	**1/1**	0.07/0.07	9.30/11.12
	FBED	0.14/0.14	**1/1**	0.07/0.07	0.21/0.22
	MMMB	**1**/0.97	0.99/0.94	**1/1**	52.23/54.17
	PCMB	**1/1**	**1/1**	**1/1**	61.38/63.15
	HITON-MB	0.99/0.96	0.99/0.93	**1/1**	68.71/69.98
	Semi-HITON-MB	0.99/0.96	0.99/0.93	**1/1**	71.32/71.97
	MBOR	0.9/0.89	0.82/0.8	**1/1**	9.69/13.16
	IPCMB	**1/1**	**1**/0.99	**1/1**	115.32/126.71
	STMB	0.62/0.48	0.57/0.36	0.69/0.73	512.36/968.12
	BAMB	0.99/0.99	0.99/**1**	**1**/0.99	5126.56/5198.33
	EEMB	**1/1**	**1/1**	**1/1**	2348.98/2269.17
	SLL	-	-	-	-
	S^2TMB	-	-	-	-
	fGES-MB	0.96	0.92	**1**	13.79

S-3: EXPERIMENTS ON REAL-WORLD DATA

In this section, we validate the causality-based feature selection algorithms using eight real-world datasets from the UCI Machine Learning Repository and NIPS2003 feature selection challenge datasets as shown in Table 8. In addition, we also selected three non-causal feature selection algorithms to compare with causality-based feature selection algorithms. The threshold value of FCBF [4] is set to 0.01. And the value of k for mRMR [3] and SPEC_CMI [2] is set to 15 (for the dataset with less than 500 features) or 30 (for the dataset with more than 500 features). Among these eight datasets, three are of low dimensionality but contain a large number of samples, two are high-dimensional datasets with large number of samples, and the other three are also of high dimensionality but small size.

We use the three classifiers NB, KNN, and SVM to evaluate a selected feature subset for classification. The value of k for the KNN classifier is set to 3 and both SVM and KNN use the linear kernel. In the following, we evaluate two types of feature subsets selected by the causality-based feature selection algorithms for classification, which are the MB and PC of the class variable, using the following metrics:

- Prediction accuracy. The number of correctly predicted data samples divided by the total number of data samples in a test dataset.
- Compactness. The size of the feature subset selected by an algorithm.
- AUC. Area Under the ROC Curve for demonstrating the performance of a classifier for predicting a binary class variable with imbalanced classes.

From Tables 9 to 12, our observations are summarized as follows:

- The classification performance using the PC set of a class variable is not inferior to that of using the MB of the class variable. And learning the PC set of the class variable for feature selection is much more efficient than learning the MB of the class variable. These findings are consistent with the results in Reference [1]. Thus, in terms of feature selection, PC learning algorithms are practical in real-world applications.
- When the size of data samples is large, the simultaneous MB learning approach is significantly faster than the other MB and PC learning algorithms, and they achieve very competitive prediction accuracy with their rivals. Surprisingly, FBED is the fastest algorithm and its performance is comparable with the others as shown in Tables 9 to 11.
- Existing score-based MB or PC learning algorithms (except for fGES-MB) are still computationally expensive or prohibitive when the size of PC or MB is large, even if the number of features in a dataset is small. For example, on the *spambase* dataset, we observe that

Table 8. Summary of Real-world Datasets

Dataset	Number of features	Number of instances	Category ratio
infant	86	5,337	0.94/0.06
spambase	57	4,601	0.61/0.39
bankruptcy	147	7,063	0.89/0.11
madelon	500	2,000	0.50/0.50
gisette	5,000	6,000	0.50/0.50
arcene	10,000	100	0.56/0.44
dexter	20,000	300	0.50/0.50
dorothea	100,000	800	0.90/0.10

Table 9. Comparison of MB Methods on Real-world Datasets in Prediction Accuracy (1)

Data	Algorithm	NB	KNN	SVM	Compactness	Time
madelon	GSMB	0.58/0.55	0.51/0.52	0.56/0.55	7.00/7.00	0.01/**0.00**
	IAMB	0.58/0.58	0.62/0.62	0.63/0.63	7.00/7.00	0.21/0.22
	Inter-IAMB	0.58/0.58	0.62/0.61	0.63/0.63	7.00/7.00	0.22/0.21
	Fast-IAMB	0.57/0.57	0.57/0.56	0.61/0.60	6.00/6.00	0.06/0.06
	LRH	0.60/**0.61**	0.61/0.60	0.62/0.63	8.00/8.00	0.16/0.30
	FBED	0.60/0.59	0.61/0.57	0.62/0.61	7.10/7.90	0.06/0.03
	MMMB	0.58/0.59	0.55/0.63	0.60/**0.64**	5.70/9.40	0.14/0.23
	PCMB	0.56/0.58	0.50/0.52	0.55/0.58	1.50/3.50	0.24/0.42
	HITON-MB	0.58/0.60	0.57/0.60	0.61/0.63	5.60/8.40	0.15/0.24
	Semi-HITON-MB	0.58/0.60	0.56/0.60	0.60/0.63	5.50/8.30	0.15/0.26
	MBOR	0.60/0.60	0.59/0.60	0.61/0.61	6.40/9.40	0.38/0.78
	IPCMB	0.59/0.60	0.54/0.54	0.59/0.61	3.70/6.10	0.16/0.35
	STMB	0.60/0.60	0.56/0.52	0.62/0.60	24.00/83.40	0.13/0.31
	BAMB	0.60/0.59	0.61/**0.65**	0.63/**0.64**	7.20/9.60	0.27/0.47
	EEMB	0.60/**0.61**	0.59/0.62	0.62/0.62	6.60/8.10	0.17/0.26
	SLL	0.57	0.52	0.57	6.10	64.47
	S^2TMB	0.58	0.52	0.57	5.20	61.59
	fGES-MB	**0.61**	0.52	0.62	4.80	0.12
	FCBF	0.57	0.50	0.57	2.00	0.03
	MRMR	0.60	0.54	0.60	30.00	0.77
	SPEC_CMI	0.52	0.52	0.52	30.00	42.95
arcene	GSMB	0.71/0.70	0.68/0.71	0.66/0.67	3.60/3.70	**0.00/0.00**
	IAMB	0.69/0.69	0.64/0.64	0.66/0.66	4.00/4.00	0.10/0.11
	Inter-IAMB	0.69/0.69	0.64/0.64	0.66/0.66	4.00/4.00	0.10/0.11
	Fast-IAMB	0.70/0.67	0.66/0.64	0.70/0.68	3.00/3.00	0.05/0.06
	LRH	0.59/0.61	0.61/0.58	0.59/0.66	3.30/4.00	0.13/0.30
	FBED	0.68/0.63	0.58/0.60	0.66/0.65	3.90/4.00	0.03/0.03
	MMMB	**0.80**/0.79	0.66/**0.74**	**0.79**/0.76	3.90/6.60	1.61/2.33
	PCMB	0.61/0.62	0.60/0.60	0.61/0.63	1.70/2.00	2.88/5.01
	HITON-MB	0.74/0.70	0.67/0.69	0.69/0.73	3.70/6.90	1.71/2.48
	Semi-HITON-MB	0.69/0.71	0.62/0.62	0.66/0.73	3.50/5.80	1.56/2.20
	MBOR	0.66/0.71	0.63/0.73	0.66/0.71	3.30/6.80	0.12/0.59
	IPCMB	0.61/0.63	0.61/0.62	0.60/0.62	1.50/2.80	0.57/1.38
	STMB	0.63/0.61	0.71/0.71	0.71/0.67	56.10/293.10	0.64/0.87
	BAMB	0.73/0.70	0.63/0.65	0.66/0.70	3.90/7.30	0.53/1.28
	EEMB	0.75/0.72	0.67/0.66	0.69/0.72	3.80/5.90	0.85/1.13
	SLL	-	-	-	-	-
	S^2TMB	0.74	0.70	0.73	6.10	1,648.84
	fGES-MB	-	-	-	-	-
	FCBF	0.62	0.65	0.62	33.40	0.27
	MRMR	0.73	0.70	0.70	30.00	8.62
	SPEC_CMI	-	-	-	-	-

(Continued)

Table 9. Continued

Data	Algorithm	NB	KNN	SVM	Compactness	Time
dexter	GSMB	0.70/0.63	0.67/0.55	0.70/0.63	4.80/4.60	**0.00/0.00**
	IAMB	0.81/0.81	0.73/0.73	0.82/0.82	5.00/5.00	0.36/0.36
	Inter-IAMB	0.81/0.81	0.73/0.73	0.82/0.82	5.00/5.00	0.35/0.35
	Fast-IAMB	0.73/0.73	0.72/0.74	0.76/0.77	4.00/4.00	0.12/0.13
	LRH	0.78/0.75	0.80/0.77	0.79/0.77	4.80/5.00	0.30/1.02
	FBED	0.81/0.81	0.73/0.73	0.82/0.82	5.00/5.00	0.05/0.06
	MMMB	0.86/0.88	0.84/0.86	0.86/0.87	9.20/19.40	4.10/7.17
	PCMB	0.81/0.85	0.71/0.83	0.81/0.84	6.60/12.30	11.74/32.99
	HITON-MB	0.85/0.87	0.81/0.86	0.85/0.86	9.80/19.50	4.44/6.99
	Semi-HITON-MB	0.85/0.87	0.79/0.85	0.84/0.86	8.40/18.80	3.92/6.87
	MBOR	**0.91**/0.90	0.87/**0.89**	0.88/0.89	21.60/31.00	4.38/10.83
	IPCMB	0.82/0.86	0.71/0.85	0.82/0.85	5.90/12.40	6.19/13.08
	STMB	0.88/0.90	0.82/0.75	**0.90/0.90**	54.50/244.50	3.03/9.61
	BAMB	0.86/0.89	0.84/0.88	0.86/0.88	9.90/17.90	2.04/3.01
	EEMB	0.85/0.87	0.82/0.87	0.85/0.86	9.20/14.20	3.32/4.45
	SLL	-	-	-	-	-
	S^2TMB	-	-	-	-	-
	fGES-MB	0.86	0.86	0.88	11.30	1.11
	FCBF	0.87	0.84	0.86	40.20	0.38
	MRMR	0.90	**0.89**	**0.90**	30.00	18.51
	SPEC_CMI	0.48	0.50	0.51	30.00	1,030.95

Table 10. Comparison of MB Methods on Real-world Datasets in Prediction Accuracy (2)

Data	Algorithm	NB	KNN	SVM	Compactness	Time
dorothea	GSMB	0.93/0.90	0.92/0.90	0.93/0.90	4.30/7.00	0.05/**0.01**
	IAMB	0.93/0.93	0.93/**0.94**	0.93/**0.94**	7.00/7.00	42.48/44.64
	Inter-IAMB	0.93/0.93	0.93/0.93	0.93/0.93	7.00/7.00	40.89/41.32
	Fast-IAMB	0.93/0.93	0.92/0.90	0.93/0.92	6.00/6.00	12.64/12.89
	LRH	**0.94/0.94**	0.92/0.92	0.93/0.93	5.90/6.30	45.22/303.98
	FBED	0.93/0.93	0.93/0.93	0.93/0.93	7.00/7.00	2.72/5.32
	MMMB	0.93/0.93	0.92/0.91	0.93/0.93	8.60/16.60	288.00/555.40
	PCMB	0.45/0.92	0.45/0.92	0.45/0.92	0.60/2.50	331.04/1,259.19
	HITON-MB	0.93/0.93	0.91/0.92	0.93/0.93	10.20/19.90	333.89/638.10
	Semi-HITON-MB	**0.94**/0.93	0.92/0.92	0.93/0.93	8.90/14.60	302.03/497.61
	MBOR	0.93/0.93	0.92/0.93	0.93/0.93	10.50/31.80	171.26/1,870.99
	IPCMB	0.36/0.83	0.36/0.82	0.36/0.83	0.50/1.40	75.16/538.33
	STMB	-/-	-/-	-/-	-/-	-/-
	BAMB	0.93/0.93	0.90/0.91	0.93/0.93	10.50/20.20	329.52/2,941.22
	EEMB	-/-	-/-	-/-	-/-	-/-
	SLL	-	-	-	-	-
	S^2TMB	-	-	-	-	-
	fGES-MB	-	-	-	-	-

(Continued)

Table 10. Continued

Data	Algorithm	NB	KNN	SVM	Compactness	Time
	FCBF	0.93	0.92	0.90	92.60	5.59
	MRMR	**0.94**	0.93	**0.94**	30.00	106.39
	SPEC_CMI	-	-	-	-	-
gisette	GSMB	0.63/0.58	0.60/0.55	0.63/0.57	3.00/3.00	**0.07/0.07**
	IAMB	0.84/0.84	0.91/0.91	**0.91/0.91**	4.00/4.00	5.94/5.86
	Inter-IAMB	0.84/0.84	0.91/0.91	**0.91/0.91**	4.00/4.00	5.82/5.85
	Fast-IAMB	0.84/0.84	0.83/0.83	0.84/0.84	3.00/3.00	1.25/1.25
	LRH	0.86/0.86	0.90/0.90	**0.91/0.91**	6.00/6.00	294.35/460.95
	FBED	0.85/0.84	0.89/0.91	0.90/**0.91**	4.00/4.00	3.52/3.69
	MMMB	0.90/0.89	**0.97**/0.96	0.75/0.75	255.10/334.40	332.94/601.17
	PCMB	0.86/0.87	0.93/0.95	0.87/0.77	46.60/113.90	1,335.92/3,115.83
	HITON-MB	0.89/0.88	**0.97/0.97**	0.75/0.75	361.30/507.50	496.91/1,035.98
	Semi-HITON-MB	0.90/0.89	**0.97/0.97**	0.75/0.73	315.80/433.50	546.27/1,117.32
	MBOR	-/-	-/-	-/-	-/-	-/-
	IPCMB	0.86/0.86	0.94/0.96	0.83/0.73	80.90/192.30	2,904.15/6,057.50
	STMB	-/-	-/-	-/-	-/-	-/-
	BAMB	-/-	-/-	-/-	-/-	-/-
	EEMB	-/-	-/-	-/-	-/-	-/-
	SLL	-	-	-	-	-
	S^2TMB	-	-	-	-	-
	fGES-MB	0.85	0.95	**0.91**	73.90	4535.82
	FCBF	**0.91**	0.92	0.90	23.20	1.85
	MRMR	0.89	0.94	**0.91**	30.00	16.25
	SPEC_CMI	0.68	0.66	0.68	30.00	752.30
bankrupty	GSMB	**0.89/0.89**	0.85/0.88	0.89/0.89	7.20/6.00	0.02/**0.01**
	IAMB	**0.89/0.89**	**0.89/0.89**	**0.90/0.90**	10.00/10.00	0.30/0.30
	Inter-IAMB	**0.89/0.89**	**0.89/0.89**	**0.90/0.90**	10.00/10.00	0.30/0.31
	Fast-IAMB	**0.89/0.89**	0.86/0.86	0.89/0.89	9.00/9.00	0.08/0.09
	LRH	**0.89/0.89**	0.88/0.88	0.89/0.89	9.70/9.70	4.28/6.26
	FBED	**0.89/0.89**	**0.89/0.89**	**0.90/0.90**	10.00/10.00	0.16/0.18
	MMMB	0.84/0.82	0.88/0.88	**0.90**/0.89	60.70/76.90	32.53/79.24
	PCMB	0.86/0.85	0.88/**0.89**	0.89/0.89	38.90/54.40	99.06/225.14
	HITON-MB	0.84/0.82	0.88/0.88	**0.90**/0.89	58.30/73.70	54.81/132.83
	Semi-HITON-MB	0.84/0.83	0.88/0.88	**0.90**/0.89	55.80/72.30	58.64/161.94
	MBOR	0.87/0.85	0.88/0.88	**0.90/0.90**	32.80/39.90	67.79/80.22
	IPCMB	0.86/0.85	0.88/0.88	0.89/0.89	39.70/56.00	76.54/243.60
	STMB	0.80/0.80	0.88/0.88	0.89/0.89	89.70/113.80	125.16/488.26
	BAMB	0.85/0.84	**0.89/0.89**	**0.90/0.90**	42.60/52.80	431.61/1,523.60
	EEMB	0.85/0.85	**0.89/0.89**	**0.90/0.90**	37.00/44.90	217.67/647.90
	SLL	0.88	0.84	0.89	12.22	11,737.17
	S^2TMB	**0.89**	0.83	0.89	9.10	1,958.75
	fGES-MB	0.84	0.88	**0.90**	47.50	19.55
	FCBF	**0.89**	0.80	0.89	8.20	0.06
	MRMR	**0.89**	0.86	0.89	15.00	0.23
	SPEC_CMI	0.88	0.85	0.89	15.00	0.72

Table 11. Comparison of MB Methods on Real-world Datasets in Prediction Accuracy (3)

Data	Algorithm	NB	KNN	SVM	Compactness	Time
infant	GSMB	0.95/0.95	**0.95/0.95**	**0.96**/0.95	4.80/3.40	0.02/**0.00**
	IAMB	0.95/0.95	**0.95/0.95**	**0.96/0.96**	5.00/5.00	0.06/0.06
	Inter-IAMB	0.95/0.95	**0.95/0.95**	**0.96/0.96**	5.00/5.00	0.06/0.06
	Fast-IAMB	0.95/0.95	0.94/0.94	**0.96/0.96**	4.30/4.30	0.02/0.02
	LRH	0.95/0.95	**0.95/0.95**	**0.96/0.96**	6.00/6.60	0.10/0.15
	FBED	0.95/0.95	**0.95/0.95**	**0.96/0.96**	5.00/5.00	0.02/0.02
	MMMB	0.95/0.95	**0.95/0.95**	**0.96/0.96**	6.90/11.30	0.12/0.24
	PCMB	0.95/0.95	**0.95/0.95**	**0.96/0.96**	5.10/7.80	0.41/0.93
	HITON-MB	0.95/0.95	**0.95/0.95**	**0.96/0.96**	6.40/11.30	0.14/0.33
	Semi-HITON-MB	0.95/0.95	**0.95/0.95**	**0.96/0.96**	6.50/11.30	0.18/0.47
	MBOR	0.95/0.95	**0.95/0.95**	**0.96/0.96**	7.10/8.30	0.30/0.58
	IPCMB	0.95/0.95	0.94/**0.95**	**0.96/0.96**	3.50/7.70	0.26/0.56
	STMB	0.95/0.94	**0.95/0.95**	**0.96**/0.95	14.60/28.00	0.13/0.48
	BAMB	0.95/0.95	**0.95/0.95**	**0.96/0.96**	5.20/9.20	0.17/0.64
	EEMB	0.95/0.95	**0.95/0.95**	**0.96/0.96**	5.20/8.30	0.09/0.28
	SLL	**0.96**	0.94	**0.96**	5.17	3,087.95
	S^2TMB	0.95	0.94	0.95	6.50	94.77
	fGES-MB	0.95	**0.95**	**0.96**	12.10	2.99
	FCBF	0.95	0.94	0.95	4.80	0.01
	MRMR	0.94	**0.95**	**0.96**	15.00	0.09
	SPEC_CMI	0.94	0.94	0.94	15.00	0.56
spambase	GSMB	0.79/0.78	0.78/0.78	0.80/0.80	8.50/8.90	**0.00/0.00**
	IAMB	**0.90/0.90**	0.91/0.91	0.91/0.91	9.00/9.00	0.06/0.06
	Inter-IAMB	**0.90/0.90**	0.91/0.91	0.91/0.91	9.00/9.00	0.06/0.06
	Fast-IAMB	0.88/0.88	0.90/0.90	0.89/0.89	8.00/8.00	0.01/0.01
	LRH	**0.90/0.90**	0.91/0.90	0.91/0.91	8.20/8.90	0.88/1.12
	FBED	**0.90/0.90**	0.91/0.91	0.91/0.91	9.00/9.00	0.04/0.05
	MMMB	0.88/0.88	**0.93/0.93**	**0.93/0.93**	44.10/49.80	12.30/23.11
	PCMB	0.88/0.88	**0.93/0.93**	**0.93/0.93**	41.30/47.80	58.64/90.65
	HITON-MB	0.88/0.88	**0.93/0.93**	**0.93/0.93**	43.10/49.10	27.06/52.15
	Semi-HITON-MB	0.88/0.88	**0.93/0.93**	**0.93/0.93**	42.90/49.00	34.29/65.51
	MBOR	0.89/0.89	0.92/**0.93**	**0.93/0.93**	37.40/42.90	25.74/29.35
	IPCMB	0.88/0.88	0.92/**0.93**	**0.93/0.93**	42.00/47.80	66.02/111.47
	STMB	0.88/0.89	**0.93/0.93**	**0.93/0.93**	50.00/52.90	67.55/147.72
	BAMB	0.89/0.89	0.92/**0.93**	**0.93/0.93**	35.40/43.30	147.33/432.71
	EEMB	**0.90/0.90**	0.92/**0.93**	**0.93/0.93**	31.60/38.00	58.51/160.65
	SLL	-	-	-	-	-
	S^2TMB	-	-	-	-	-
	fGES-MB	0.89	0.92	**0.93**	45.00	3.07
	FCBF	**0.90**	0.91	0.91	10.60	0.04
	MRMR	**0.90**	0.91	0.91	15.00	0.06
	SPEC_CMI	0.83	0.82	0.85	15.00	0.08

Table 12. Comparison of the PC Methods on Real-world Datasets in Prediction Accuracy

Data	Algorithm	NB	KNN	SVM	Compactness	Time
madelon	PC-simple	**0.60**/**0.60**	0.53/0.49	0.59/**0.60**	3.60/4.70	**0.03**/0.08
	MMPC	0.58/0.59	0.52/**0.57**	**0.60**/**0.60**	4.70/6.10	**0.03**/0.04
	HITON-PC	0.59/0.59	0.52/0.54	0.59/**0.60**	4.70/5.90	**0.03**/0.05
	Semi-HITON-PC	0.59/0.59	0.51/0.54	0.59/**0.60**	4.60/5.50	**0.03**/0.05
	GetPC	0.56/0.58	0.50/0.51	0.56/0.58	1.50/2.70	0.15/0.23
	MBtoPC	**0.60**/0.59	0.52/0.52	**0.60**/**0.60**	3.40/4.50	0.24/0.43
	SLL-PC	0.57	0.52	0.57	5.10	22.94
	S^2TMB-PC	0.58	0.52	0.57	5.10	61.63
arcene	PC-simple	0.61/0.63	0.61/0.62	0.60/0.61	1.50/3.60	0.25/0.37
	MMPC	0.78/**0.80**	0.66/**0.75**	**0.79**/0.77	3.80/5.30	0.34/0.40
	HITON-PC	0.74/0.72	0.67/0.68	0.69/0.71	3.70/5.60	0.34/0.40
	Semi-HITON-PC	0.69/0.73	0.62/0.61	0.66/0.69	3.50/4.90	0.34/0.40
	GetPC	0.61/0.62	0.60/0.60	0.61/0.63	1.70/2.00	1.66/2.42
	MBtoPC	0.66/0.69	0.68/0.65	0.71/0.69	2.70/3.20	**0.12**/0.61
	SLL-PC	-	-	-	-	-
	S^2TMB-PC	0.74	0.68	0.73	5.80	1633.66
dexter	PC-simple	0.82/0.86	0.71/0.84	0.82/0.85	6.00/10.90	1.01/2.06
	MMPC	0.86/**0.87**	0.81/**0.87**	**0.86**/**0.86**	8.20/13.40	**0.49**/0.63
	HITON-PC	0.85/0.86	0.79/**0.87**	0.85/**0.86**	9.00/13.40	**0.49**/0.61
	Semi-HITON-PC	0.85/0.86	0.78/0.85	0.85/**0.86**	7.80/13.00	**0.49**/0.64
	GetPC	0.81/0.85	0.72/0.82	0.81/0.84	6.20/10.60	4.16/7.18
	MBtoPC	0.80/0.79	0.75/0.75	0.82/0.81	5.10/5.50	4.10/9.95
	SLL-PC	-	-	-	-	-
	S^2TMB-PC	-	-	-	-	-
dorothea	PC-simple	0.73/0.93	0.73/**0.92**	0.73/0.93	1.30/2.20	36.72/249.97
	MMPC	0.93/0.93	**0.92**/0.91	0.93/0.93	8.60/15.30	28.47/46.96
	HITON-PC	0.93/0.93	0.91/0.91	0.93/0.93	9.90/17.80	28.09/42.30
	Semi-HITON-PC	**0.94**/**0.94**	**0.92**/**0.92**	0.93/0.93	8.80/13.90	**28.04**/39.01
	GetPC	0.54/0.92	0.54/**0.92**	0.54/0.92	0.80/2.20	266.76/615.19
	MBtoPC	**0.94**/0.93	**0.92**/**0.92**	0.93/**0.94**	6.40/7.60	165.69/1,726.01
	SLL-PC	-	-	-	-	-
	S^2TMB-PC	-	-	-	-	-
gisette	PC-simple	0.89/0.90	0.93/0.94	**0.90**/0.87	22.30/39.60	1,154.45/3,168.57
	MMPC	**0.91**/**0.91**	0.96/0.96	0.84/0.82	53.70/73.90	**67.47**/169.92
	HITON-PC	**0.91**/**0.91**	0.96/**0.97**	0.83/0.80	48.80/67.90	121.95/329.02
	Semi-HITON-PC	**0.91**/0.90	0.96/0.96	0.84/0.82	46.50/63.60	135.25/372.99
	GetPC	0.87/0.88	0.89/0.91	**0.90**/0.88	11.50/22.70	291.91/640.68
	MBtoPC	-/-	-/-	-/-	-/-	-/-
	SLL-PC	-	-	-	-	-
	S^2TMB-PC	-	-	-	-	-
	PC-simple	0.88/0.87	0.86/0.87	0.89/0.89	23.30/31.30	29.92/92.30
	MMPC	0.87/0.87	0.87/0.87	0.89/0.89	30.00/37.40	12.56/30.69
	HITON-PC	0.87/0.87	0.87/0.88	0.89/0.89	28.50/36.00	22.43/56.45

(Continued)

Table 12. Continued

Data	Algorithm	NB	KNN	SVM	Compactness	Time
bankrupty	Semi-HITON-PC	0.88/0.87	0.87/0.87	0.89/0.89	27.00/35.00	26.36/72.68
	GetPC	**0.89**/0.88	0.86/0.87	0.89/0.89	15.40/24.30	56.52/163.50
	MBtoPC	**0.89/0.89**	0.88/**0.89**	0.89/**0.90**	8.20/9.90	**4.05**/5.17
	SLL-PC	**0.89**	0.84	0.89	9.30	2,668.09
	S^2TMB-PC	**0.89**	0.84	0.89	8.90	1,984.47
infant	PC-simple	**0.95/0.95**	**0.95/0.95**	**0.96/0.96**	3.80/6.70	0.06/0.14
	MMPC	**0.95/0.95**	**0.95/0.95**	**0.96/0.96**	5.00/7.90	**0.02**/0.06
	HITON-PC	**0.95/0.95**	**0.95/0.95**	**0.96/0.96**	5.00/8.10	0.03/0.10
	Semi-HITON-PC	**0.95/0.95**	**0.95/0.95**	**0.96/0.96**	5.00/7.90	0.04/0.15
	GetPC	**0.95/0.95**	**0.95/0.95**	**0.96/0.96**	4.00/6.40	0.11/0.25
	MBtoPC	**0.95/0.95**	**0.95/0.95**	**0.96/0.96**	5.00/5.00	0.14/0.19
	SLL-PC	**0.95**	0.94	0.95	5.40	2,653.88
	S^2TMB-PC	**0.95**	0.94	0.95	5.90	94.8
spambase	PC-simple	0.90/0.90	**0.92/0.92**	**0.93/0.93**	21.10/26.80	19.17/26.59
	MMPC	**0.91/0.91**	**0.92/0.92**	**0.93/0.93**	23.30/28.60	2.06/4.08
	HITON-PC	0.90/**0.91**	**0.92/0.92**	**0.93/0.93**	23.30/28.10	5.02/9.97
	Semi-HITON-PC	**0.91/0.91**	**0.92/0.92**	**0.93/0.93**	23.20/28.10	6.18/12.78
	GetPC	0.90/**0.91**	**0.92/0.92**	**0.93/0.93**	21.90/27.10	25.86/49.17
	MBtoPC	0.90/0.90	0.90/0.91	0.91/0.91	9.30/9.70	**0.99**/1.02
	SLL-PC	-	-	-	-	-
	S^2TMB-PC	-	-	-	-	-

when the size of MB is over 30, SLL and S^2TMB are very computationally expensive using the exact score-based BN learning algorithms. Compared to SLL and S^2TMB, fGES-MB can efficiently deal with some high-dimensional datasets. Using both synthetic and real-world datasets, existing score-based MB or PC learning algorithms do not show significant advantages over the constraint-based algorithms. The results also explain why the score-based MB learning approach is not the focus in the causality-based feature selection research.

- With the three class-imbalanced datasets, *infant*, *bankruptcy*, and *dorothea*, all MB learning algorithms achieve almost the same high prediction accuracy in Tables 9 to 11. In Table 12, on *infant* and *bankruptcy*, it is the same for all PC learning algorithms. However, in Tables 13 to 14, we can see that all MB and PC learning algorithms achieve much low values of AUC on the three class-imbalanced datasets, compared to their corresponding prediction accuracy in Tables 9 to 12. Here we do not report the value of AUC of each algorithm on the remaining five class-balanced date sets, since the MB and PC learning algorithms obtain almost the same prediction accuracy and AUC. The results indicate that the existing causality-base feature selection algorithms are not able to deal with a dataset with imbalanced classes well.
- The compactness and symmetry check. In terms of compactness, in Tables 9 to 11, STMB selects the most features than the other rivals for all datasets, while the score-based MB algorithms achieve the worst performance among all algorithms under comparison. Meanwhile, the symmetry check may not be helpful for selecting a good feature subset for classification in real-world applications. MMMB, HITON-MB, and semi-HITON-MB achieve more stable and better prediction accuracy than the other rivals, although their outputs include more false positives than the other rivals as discussed above. The possible explanation is that when we use causality-based feature selection to deal with real-world datasets for

Table 13. Comparison of MB Learning Methods on Class-imbalanced Real-world Data in AUC

Data	Algorithm	NB-AUC	KNN-AUC	SVM-AUC
dorothea	GSMB	0.73/0.50	0.68/0.51	0.73/0.50
	IAMB	0.78/0.78	0.78/0.78	**0.82/0.82**
	Inter-IAMB	0.78/0.79	0.77/**0.79**	**0.82/0.82**
	Fast-IAMB	0.76/0.76	0.74/0.67	0.78/0.76
	LRH	0.79/0.78	0.75/0.74	0.78/0.79
	FBED	0.79/0.79	0.78/0.76	**0.82/0.82**
	MMMB	0.75/0.73	0.73/0.72	0.79/0.77
	PCMB	0.28/0.68	0.28/0.68	0.28/0.69
	HITON-MB	0.74/0.74	0.73/0.73	0.77/0.71
	Semi-HITON-MB	0.78/0.74	0.75/0.72	0.79/0.72
	MBOR	0.73/0.78	0.72/0.71	0.78/0.75
	IPCMB	0.23/0.59	0.23/0.57	0.23/0.60
	STMB	-/-	-/-	-/-
	BAMB	0.72/0.73	0.68/0.71	0.77/0.74
	EEMB	-/-	-/-	-/-
	SLL	-	-	-
	S^2TMB	-	-	-
	fGES-MB	-	-	-
	FCBF	0.72	0.68	0.51
	MRMR	**0.82**	0.73	0.78
	SPEC_CMI	-	-	-
bankrupty	GSMB	0.50/0.50	0.51/0.50	0.50/0.50
	IAMB	0.55/0.55	0.70/0.70	**0.59/0.59**
	Inter-IAMB	0.55/0.55	0.70/0.70	**0.59/0.59**
	Fast-IAMB	0.50/0.50	**0.73**/0.70	0.50/0.50
	LRH	0.56/0.56	0.69/0.69	0.52/0.53
	FBED	0.54/0.55	0.69/0.70	**0.59/0.59**
	MMMB	0.78/0.78	0.66/0.64	0.56/0.54
	PCMB	0.74/0.77	0.64/0.66	0.52/0.53
	HITON-MB	0.78/0.78	0.65/0.65	0.56/0.54
	Semi-HITON-MB	0.78/0.78	0.65/0.65	0.56/0.54
	MBOR	0.74/0.76	0.65/0.65	0.56/0.57
	IPCMB	0.75/0.78	0.65/0.66	0.54/0.54
	STMB	0.78/0.78	0.63/0.63	0.53/0.51
	BAMB	0.78/**0.79**	0.68/0.68	0.58/0.57
	EEMB	0.77/0.78	0.67/0.67	0.57/0.57
	SLL	0.62	0.66	0.50
	S^2TMB	0.51	**0.73**	0.50
	FGES-MB	0.74	0.65	0.55
	fCBF	0.52	0.63	0.50
	MRMR	0.62	0.65	0.50
	SPEC_CMI	0.53	0.53	0.50

(Continued)

Table 13. Continued

Data	Algorithm	NB-AUC	KNN-AUC	SVM-AUC
infant	GSMB	0.72/0.65	0.59/0.58	0.68/0.65
	IAMB	**0.74**/**0.74**	0.66/0.66	0.69/0.69
	Inter-IAMB	**0.74**/**0.74**	0.66/0.66	0.69/0.69
	Fast-IAMB	0.72/0.72	0.53/0.53	0.68/0.68
	LRH	**0.74**/0.73	0.66/0.65	0.69/0.68
	FBED	**0.74**/**0.74**	0.66/0.66	0.69/0.69
	MMMB	**0.74**/**0.74**	**0.67**/0.65	0.69/0.69
	PCMB	0.73/0.73	0.64/0.65	0.69/0.69
	HITON-MB	0.73/**0.74**	0.66/**0.67**	0.69/0.70
	Semi-HITON-MB	0.73/**0.74**	0.66/0.66	0.69/0.69
	MBOR	**0.74**/**0.74**	**0.67**/**0.67**	0.69/0.69
	IPCMB	0.72/0.73	0.55/0.63	0.68/0.69
	STMB	0.73/**0.74**	0.66/**0.67**	0.69/0.67
	BAMB	**0.74**/**0.74**	**0.67**/**0.67**	0.69/0.69
	EEMB	0.73/**0.74**	0.65/0.66	0.69/0.69
	SLL	0.71	0.50	**0.71**
	S^2TMB	0.71	0.50	**0.71**
	fGES-MB	0.73	0.59	0.69
	FCBF	0.71	0.51	**0.71**
	MRMR	**0.74**	**0.67**	0.69
	SPEC_CMI	0.56	0.53	0.54

Table 14. Comparison of PC Learning Methods on Class-imbalanced Real-world Data in AUC

Data	Algorithm	NB-AUC	KNN-AUC	SVM-AUC
dorothea	PC-simple	0.51/0.71	0.51/0.65	0.51/0.72
	MMPC	0.75/0.73	0.72/0.72	0.79/0.76
	HITON-PC	0.74/0.73	0.72/0.73	0.77/0.73
	Semi-HITON-PC	**0.78**/0.74	**0.75**/0.72	0.79/0.74
	GetPC	0.33/0.68	0.33/0.67	0.33/0.69
	MBtoPC	0.77/**0.78**	0.72/0.73	0.81/**0.82**
	SLL-PC	-	-	-
	S^2TMB-PC	-	-	-
bankrupty	PC-simple	0.72/0.74	0.64/0.67	0.50/0.51
	MMPC	0.76/**0.77**	0.67/0.67	0.50/0.51
	HITON-PC	0.74/0.76	0.67/0.68	0.50/0.51
	Semi-HITON-PC	0.74/**0.77**	0.66/0.67	0.50/0.51
	GetPC	0.63/0.70	0.65/0.65	0.50/0.51
	MBtoPC	0.54/0.55	0.67/0.69	0.50/**0.59**
	SLL-PC	0.54	0.70	0.50
	S^2TMB-PC	0.51	**0.74**	0.50

(Continued)

Table 14. Continued

Data	Algorithm	NB-AUC	KNN-AUC	SVM-AUC
infant	PC-simple	0.73/**0.74**	0.61/0.65	0.69/0.69
	MMPC	0.73/**0.74**	**0.66/0.66**	0.69/0.69
	HITON-PC	0.73/**0.74**	**0.66/0.66**	0.69/0.68
	Semi-HITON-PC	0.73/**0.74**	**0.66**/0.65	0.69/0.68
	GetPC	0.73/0.73	0.62/0.62	0.69/0.69
	MBtoPC	**0.74/0.74**	**0.66/0.66**	0.69/0.69
	SLL-PC	0.71	0.50	**0.71**
	S^2TMB-PC	0.71	0.50	**0.71**

classification, these real-world datasets may violate the faithfulness or causal sufficiency assumption. And this is the same for MMPC, HITON-PC, and semi-HITON-PC in Table 11.

- Among the three non-causal feature selection algorithms, FCBF is the fastest. FCBF and mRMR are computationally efficient. FCBF is faster than the constraint-based MB learning algorithms and only a little slower than simultaneous MB learning algorithms. All three algorithms achieve good prediction accuracy. In most cases, the causality-based feature selection algorithms achieve better prediction accuracy than the three non-causal feature selection algorithms. In addition, the three non-causal feature selection algorithms cannot deal with a dataset with imbalanced classes.

ACKNOWLEDGMENTS

This work is partially supported by the National Key Research and Development Program of China (under grant 2016YFB1000901), Australian Research Council Discovery Projects (under grant DP170101306), and the National Science Foundation of China (under grant 61876206).

REFERENCES

[1] Constantin F. Aliferis, Alexander Statnikov, Ioannis Tsamardinos, Subramani Mani, and Xenofon D. Koutsoukos. 2010. Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation. *J. Mach. Learn. Res.* 11 (2010), 171–234.

[2] Xuan Vinh Nguyen, Jeffrey Chan, Simone Romano, and James Bailey. 2014. Effective global approaches for mutual information based feature selection. In *Proceedings of the Conference on Knowledge Discovery and Data Mining (KDD'14)*. 512–521.

[3] Hanchuan Peng, Fuhui Long, and Chris Ding. 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 8 (2005), 1226–1238.

[4] Lei Yu and Huan Liu. 2004. Efficient feature selection via analysis of relevance and redundancy. *J. Mach. Learn. Res.* 5, Oct. (2004), 1205–1224.