A Study on Comparison of Bayesian Network Structure Learning Algorithms for Selecting Appropriate Models

Jae-seong Yoo

Dept. of Statistics

January 19, 2015

Title

Goal

- In this paper, we compare the performance between the Bayesian network structure learning algorithms provided by **bnlearn** package in **R**.
- The performance is evaluated by
  - using a score

    or
  - comparing between the target network and the learnt network.

  In this paper, it was confirmed that algorithm specific performance test results using fore-mentioned methods are different.
- A data generator based on Bayesian network model using **R** is built and introduced.
- The aim of this paper is to provide objective guidance of selecting suitable algorithm in accordance to target network using synthetic data generated based on topology.

Bayesian Network

A BN defines a unique joint probability distribution over X given by

$$P_B(X_1, \cdots, X_n) = \prod_{i=1}^{n} P_B(X_i | \prod_{X_i}).$$

- A BN encodes the independence assumptions over the component random variables of X.
- An edge (j, i) in E represents a direct dependency of X_i from X_j.
- The set of all Bayesian networks with n variables is denotes by B_n.

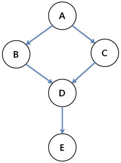


Figure: $P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|B, C)P(E|D)$

기본 개념

- **베이지안 네트워크**(이하 BN)는 확률 값이 모인 집합의 결합확률분포의 **결정모델**이다.
- 특정 분야의 영역지식을 확률적으로 표현하는 수단
- 변수들간의 확률적 의존 관계를 나타내는 **그래프**와, 각 변수별 **조건부 확률**로 구성된다.
- 하나의 BN은 각 노드마다 하나의 **조건부 확률표(CPT ; Conditional Probability Table)** 를 갖는 **비순환유향그래프(DAG ; Directed Acycle Graph)**로 정의할 수 있다.
- 노드와 노드를 연결하는 **호(arc or edge)**는 노드 사이의 관계를 나타내며, 변수의 확률적인 인과관계로 네트워크를 구성하고 조건부확률표(CPT)를 가지고 다음의 식과 같은 베이즈 정리(Bayes Theorem)을 이용하여 결과를 추론할 수 있다.

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

기본 개념

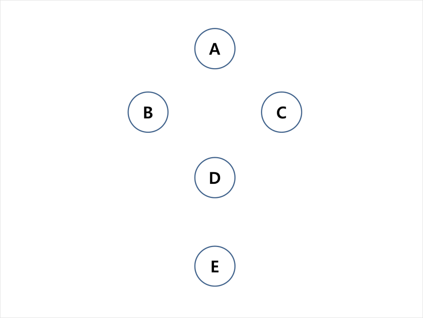


Figure: Nodes

기본 개념

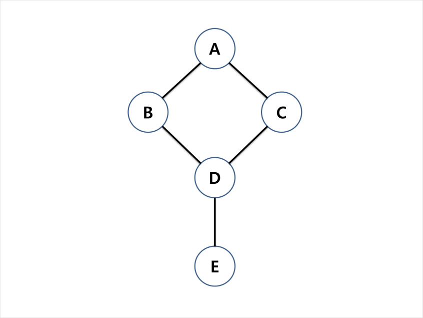


Figure: Edges

기본 개념

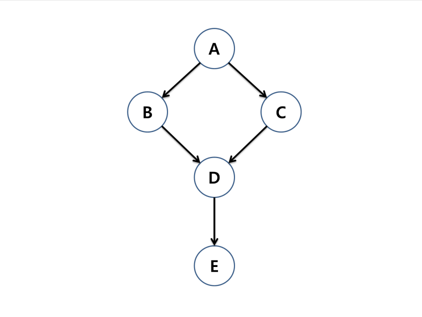


Figure: Edges = directed, (No cycles!)

기본 개념

- 일반적인 베이지안 네트워크는 베이즈 정리, 곱셈 규칙, 체인 규칙(chain rule)에 의하여 다음과 같은 식이 만들어진다.

$$P(A, B, C, D, E) = \prod_i P(x_i|\text{parent}_i)$$

여기서 $x_1, \cdots, x_n$은 특정 데이터의 속성 집합

$\text{parent}(x_i)$는 x_i의 부모 노드들의 집합

기본 개념

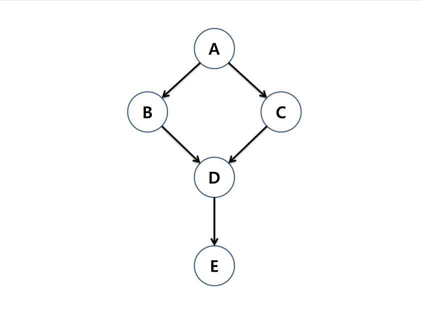


Figure: $P(A, B, C, D, E) = \prod_i P(\text{node}_i | \text{parents}_i)$

기본 개념

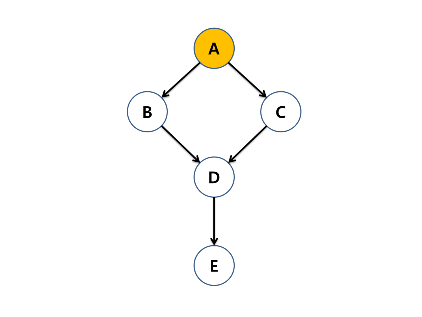


Figure: $P(A, B, C, D, E) = P(A)$

기본 개념

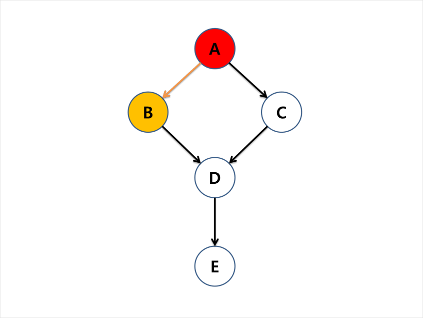


Figure: $P(A, B, C, D, E) = P(A)P(B|A)$

기본 개념

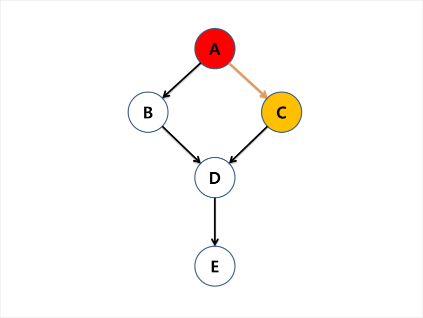


Figure: $P(A, B, C, D, E) = P(A)P(B|A)P(C|A)$

기본 개념

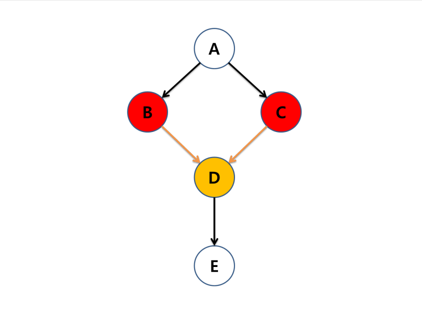


Figure: $P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|B, C)$

기본 개념

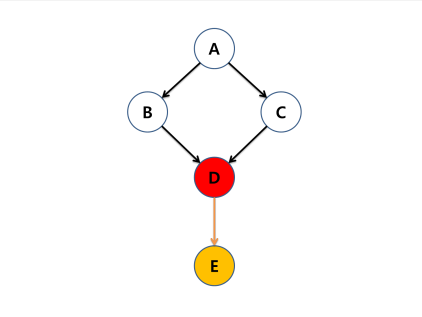


Figure: $P(A, B, C, D, E) = P(A)P(B|A)P(C|A)P(D|B, C)P(E|D)$

S. L. Lauritzen and D. J. Spiegelhalter (1988)

S. L. Lauritzen and D. J. Spiegelhalter (1988),

"Local Computations with Probabilities on Graphical Structures

and Their Application to Expert Systems",

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 50, No. 2, pp. 157-224

Abstract

Casual Network는 변수 set에 포함되어있는 영향력의 패턴을 서술하는 것을 말하며, 많은 분야에서 사용되고 있다. 이는 전문가 시스템으로부터 크지만(large) 희소(sparse) 네트워크를 이용해 국지적 계산을 하여 구해진 평균을 통해 이 영향력을 추론하는 것이 보통이다. 이는 정확한 확률적인 방법을 이용하는 것이 불가능하였다.

이 논문에서는 original network에 위상 변화를 주어 일부 범위를 결합 확률 분포로 나타낼 수 있고, 이를 이용해서 변수 사이의 국지적 영향력을 추론할 수 있다고 설명하고 있다.

S. L. Lauritzen and D. J. Spiegelhalter (1988)

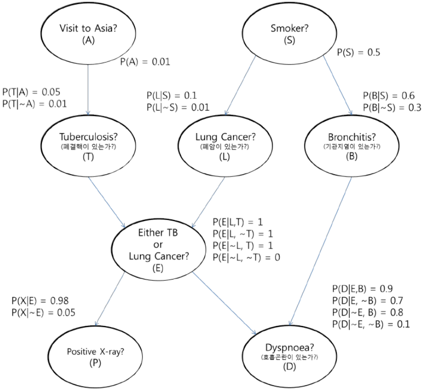


Figure: 아시아 방문 여부, 흡연 여부와 폐질환과의 관계를 도식화한 베이지안 네트워크 모형

S. L. Lauritzen and D. J. Spiegelhalter (1988)

Console ~/

```
> # same BN using data
> data(asia)
> head(asia)
   A   S   T  L   B   E   X   D
1 no yes  no no yes  no  no yes
2 no yes  no no  no  no  no  no
3 no  no yes no  no yes yes yes
4 no  no  no no yes  no  no yes
5 no  no  no no  no  no  no yes
6 no yes  no no  no  no  no yes
>
> net.data <- bn.fit(hc(asia), asia)
```

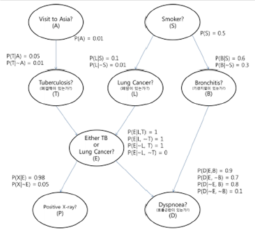
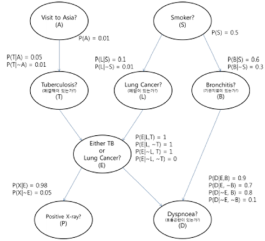


Figure: 실제 데이터의 모습과, 이의 베이지안 네트워크 그래프 모형

S. L. Lauritzen and D. J. Spiegelhalter (1988)

Unit Test : 환자 1이 있는데, 그에 대한 정보가 아무것도 없다. $P(T=1)$, $P(L=1)$, $P(B=1)$

```
Console ~/
> # Unit test
> cpquery(net.disc, (T=="yes"), TRUE)
[1] 0.0106
> cpquery(net.disc, (L=="yes"), TRUE)
[1] 0.0578
> cpquery(net.disc, (B=="yes"), TRUE)
[1] 0.4578
```

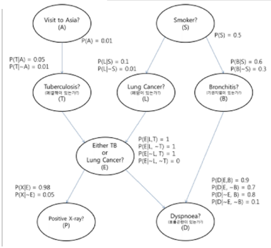


결론 : 결핵일 가능성은 약 1%, 폐암일 가능성은 약 5%, 기관지염이 있을 가능성은 약 46% 이다.

S. L. Lauritzen and D. J. Spiegelhalter (1988)

Question 1 : 환자 2는 최근 아시아를 방문했고, 비흡연자이다. $P(T=1|A=1,S=0)$

```
Console ~/
> # Question 1:
> cpquery(net.disc, (T=="yes"), (A=="yes" & S=="no"))
[1] 0.04545455
> cpquery(net.disc, (L=="yes"), (A=="yes" & S=="no"))
[1] 0.03703704
> cpquery(net.disc, (B=="yes"), (A=="yes" & S=="no"))
[1] 0.3448276
```

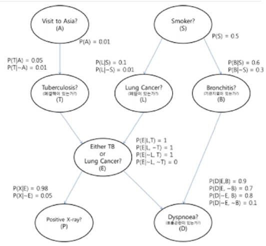


결론 : 기관지염이 있을 가능성이 약 34% 이다.

S. L. Lauritzen and D. J. Spiegelhalter (1988)

Question 2 : 환자 3은 최근 아시아를 방문했고 비흡연자이다. 호흡 곤란도 겪지 않고 있지만, X-Ray 테스트 결과 양성 반응을 보였다. $P(T = 1 | A = 1, S = 0, D = 0, X = 1)$

```
Console ~/
> cpquery(net.data, (T=="yes"), (A=="yes" & S=="no" & D=="no" & X=="yes"))
[1] 0.3333333
> cpquery(net.data, (L=="yes"), (A=="yes" & S=="no" & D=="no" & X=="yes"))
[1] 0
> cpquery(net.data, (B=="yes"), (A=="yes" & S=="no" & D=="no" & X=="yes"))
[1] 0.3333333
```



결론 : 환자에게 결핵과 기관지염이 있을 가능성이 각각 약 33%이다.

베이지안 네트워크의 특징

장점

- 특정 분야의 영역 지식을 확률적으로 표현하는 대표적인 수단
- 변수들 간의 확률적 의존 관계를 나타내는 그래프와 각 변수별 조건부 확률로 구성
- 분류 클래스 노드의 사후 확률분포를 구해줌으로써 개체들에 대한 하나의 자동분류기로 이용 가능
- 샘플이 어떤 부류로 분류되었을 때 왜 그런 결정이 내려졌는지 해석 가능

단점

- 입력값으로 수치형이 아닌 범주형을 사용
- 노드 수가 방대해지면 시간이 오래 소요될 수 있음

다른 기법과의 비교 - 로지스틱 회귀분석

장점

- 다변량 분석으로 많이 쓰임
- 다변량 변수를 독립변수로하여 종속변수에 미치는 영향을 파악 가능
- 입력값으로 수치형과 범주형 모두 취급 가능

단점

- 샘플이 어떤 부류로 분류되었을 때 왜 그런 결정이 내려졌는지 해석하기가 어려움
- 분석자료에 가장 적합한 모델을 선정하는 데 시간 투자가 필요

다른 기법과의 비교 - 신경망

장점

- 분류문제 뿐만 아니라 예측, 평가, 합성, 제어 등의 다양한 분야에 적용 가능
- 학습능력을 갖추고 일반화 능력이 뛰어나고 구현이 쉬움
- 다층 퍼셉트론은 선형분리가 불가능한 경우에도 높은 성능을 보여주는 한 단계 진보한 신경망

단점

- 샘플이 어떤 부류로 분류되었을 때 왜 그런 결정이 내려졌는지 이유를 분석하기가 어려움
- 입력값으로 수치형이 아닌 범주형을 사용

다른 기법과의 비교 - 의사 결정 트리

장점

- 의사결정규칙을 도표화하여 관심대상에 해당하는 집단을 몇 개의 소집단으로 분류하거나 예측을 수행하는 계량적 분석 방법
- 샘플이 어떤 부류로 분류되었을 때 왜 그런 결정이 내려졌는지 해석 가능
- 입력값으로 수치형, 범주형 모두 취급 가능

단점

- 반응변수가 수치형인 회귀모형에서는 그 예측력이 떨어짐
- 나무가 너무 깊은 경우에는 예측력 저하와 해석이 쉽지 않음
- 가지가 많을 경우 새로운 자료에 적용할 때 예측 오차가 큼

베이지안 네트워크의 다양한 유형

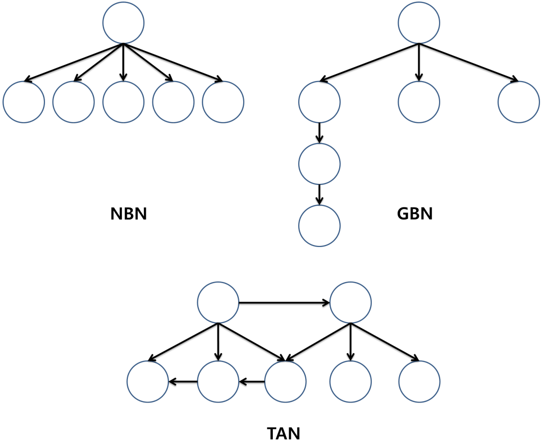
- 나이브 베이지안 네트워크 (NBN)
  가정의 단순함에도 불구하고 많은 연구를 통해 비교적 높은 분류성능을 보여준다.

- 일반 베이지안 네트워크 (GBN)
  클래스 노드조차 일반 속성 노드와의 차이를 두지 않고 모든 노드들 간의 상호의존도를 하나의 베이지안 네트워크로 표현한다.

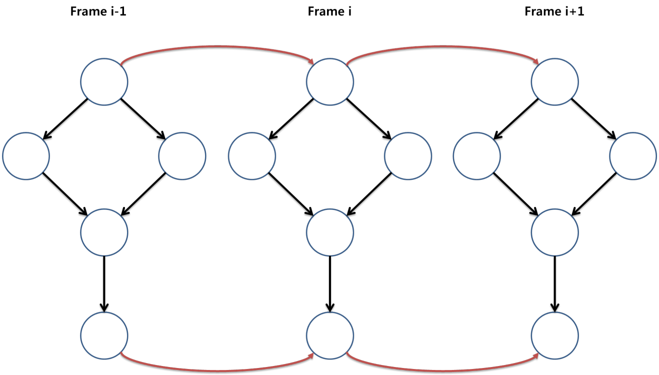
- 트리-확장 나이브 베이지안 네트워크 (TAN)
  속성 노드들 간에도 상호의존도가 존재한다고 가정하고, 이러한 속성 간 상호의존도를 하나의 일반 베이지안 네트워크 형태로 표현 가능하도록 NBN을 확장한 것이다.

- 동적 베이지안 네트워크 (DBN)
  시계열 분석을 위해 현재 변수의 확률을 계산할 때, 이전 시점의 정보를 함께 고려하는 베이지안 네트워크이다.

베이지안 네트워크의 다양한 유형



베이지안 네트워크의 다양한 유형



DBN

Bayesian Network Structure Learning

Learning a Bayesian network is as follows:

Given a data $T = \{y_1, \cdots, y_n\}$ and a scoring function ϕ, the problem of learning a Bayesian network is to find a Bayesian network $B \in B_n$ that maximizes the value $\phi(B, T)$.

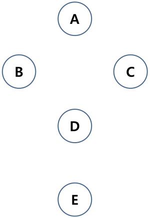


Figure: A model before learning structure

Available constraint-based learning algorithms

Grow-Shrink (GS) based on the Grow-Shrink Markov Blanket, the first (and simplest) Markov blanket detection algorithm used in a structure learning algorithm.

Incremental Association (IAMB) based on the Markov blanket detection algorithm of the same name, which is based on a two-phase selection scheme (a forward selection followed by an attempt to remove false positives).

Available Score-based Learning Algorithms

Hill-Climbing (HC) a hill climbing greedy search on the space of the directed graphs. The optimized implementation uses score caching, score decomposability and score equivalence to reduce the number of duplicated tests.

Tabu Search (TABU) a modified hill climbing able to escape local optima by selecting a network that minimally decreases the score function.

Available Hybrid Learning Algorithms

Max-Min Hill-Climbing (MHHC) a hybrid algorithm which combines the Max-Min Parents and Children algorithm (to restrict the search space) and the Hill-Climbing algorithm (to find the optimal network structure in the restricted space).

Restricted Maximization (RSMAX2) a more general implementation of the Max-Min Hill-Climbing, which can use any combination of constraint-based and score-based algorithms.

The Number of Graphical Errors in the Learnt Structure

In terms of the number of graphical errors in the learnt structure.

		Target Network	Learnt Network	Direction
C	(Correct Arcs)	exist	exist	correct
M	(Missing Arcs)	exist	not exist	
WO	(Wrongly Oriented Arcs)	exist	exist	wrong
WC	(Wrongly Corrected Arcs)	not exist	exist	

Network Scores

In all four cases, the higher the value of the metric, the better the network.

BDe $BDe(B,T) = P(B,T) = P(B) \times \prod_{i=1}^{n} \prod_{j=1}^{q_i} \left(\frac{\Gamma(N'_{ij})}{\Gamma(N_{ij}+N'_{ij})} \times \prod_{k=1}^{r_i} \frac{\Gamma(N_{ijk}+N'_{ijk})}{\Gamma(N'_{ijk})} \right)$

$$\phi(B|T) = LL(B|T) - f(N)|B|,$$

Log-Likelihood(LL) If $f(N) = 0$, we have the **LL** score.

AIC If f(N) = 1, we have the **AIC** scoring function:

BIC If $f(N) = \frac{1}{2}\log(N)$, we have the **BIC** score.

Data Generation with BN_Data_Generator in R

BN_Data_Generator {BNDataGenerator}

Description It based on a Bayesian network model to generates synthetic data.

Usage BN_Data_Generator (arcs, Probs, n, node_names)

Suppose bnlearn

Repository CRAN (Submitted at 2014-12-28)

URL https://github.com/praster1/BN_Data_Generator

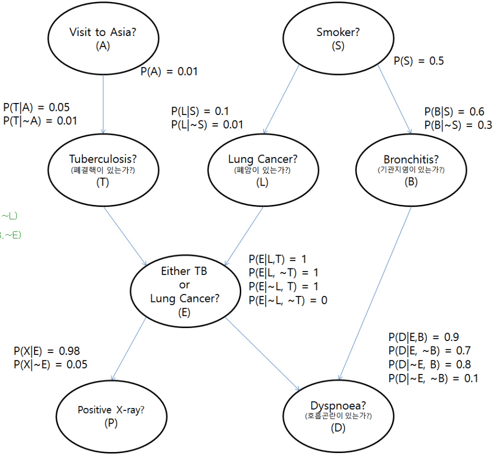
Arguments

arcs	(matrix)	A matrix that determines the arcs.
Probs	(list)	The conditional probabilities.
n	(constant)	Data Size
node_names	(vector)	Node names

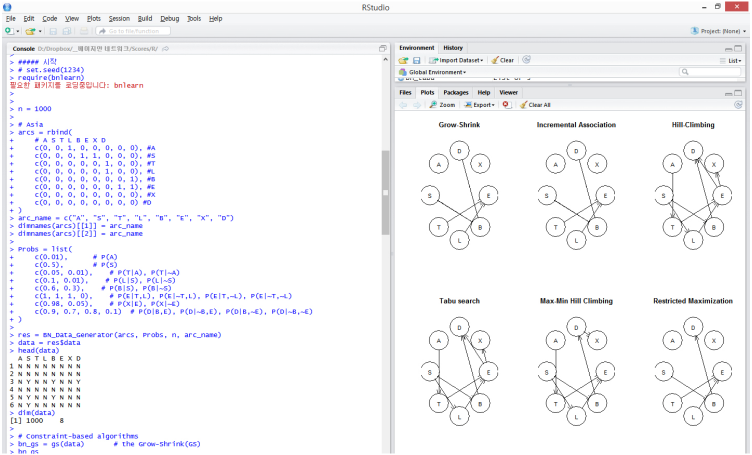
Data Generation with BN_Data_Generator in R

```
# Asia
arcs = rbind(
     #  A  S  T  L  B  E  X  D
     c(0,  0, 1, 0, 0, 0, 0, 0), #A
     c(0,  0, 0, 1, 1, 0, 0, 0), #S
     c(0,  0, 0, 0, 0, 1, 0, 0), #T
     c(0,  0, 0, 0, 0, 1, 0, 0), #L
     c(0,  0, 0, 0, 0, 0, 0, 1), #B
     c(0,  0, 0, 0, 0, 0, 1, 1), #E
     c(0,  0, 0, 0, 0, 0, 0, 0), #X
     c(0,  0, 0, 0, 0, 0, 0, 0) #D
     )
arc_name = c("A", "S", "T", "L", "B", "E", "X", "D")
dimnames(arcs)[[1]] = arc_name
dimnames(arcs)[[2]] = arc_name

Probs = list(
   c(0.01),                  # P(A)
   c(0.5),                   # P(S)
   c(0.05, 0.01),            # P(T|A), P(T|~A)
   c(0.1, 0.01),             # P(L|S), P(L|~S)
   c(0.6, 0.3),            # P(B|S), P(B|~S)
   c(1, 1, 1, 0),            # P(E|T,L), P(E|~T,L), P(E|T,~L), P(E|~T,~L)
   c(0.98, 0.05),            # P(X|E), P(X|~E)
   c(0.9, 0.7, 0.8, 0.1)     # P(D|B,E), P(D|~B,E), P(D|B,~E), P(D|~B,~E)
)
```



Data Generation with BN_Data_Generator in R



Outline

Asia DataSet

Description Small synthetic data set from Lauritzen and Spiegelhalter (1988) about lung diseases (tuberculosis, lung cancer or bronchitis) and visits to Asia.

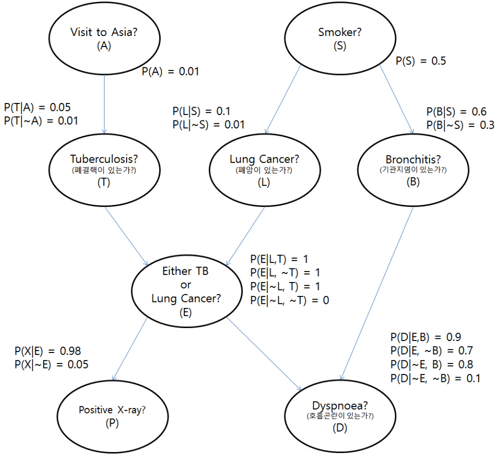
Number of nodes 8

Number of arcs 8

Number of parameters 18

Source Lauritzen S, Spiegelhalter D (1988).
"Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion)".
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 50(2), 157-224.

Asia DataSet



Insurance DataSet

Description Insurance is a network for evaluating car insurance risks.

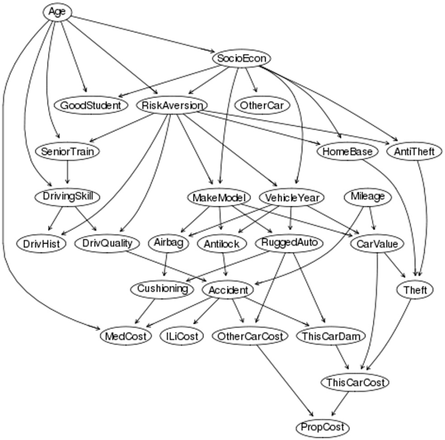
Number of nodes 27

Number of arcs 52

Number of parameters 984

Source Binder J, Koller D, Russell S, Kanazawa K (1997). "Adaptive Probabilistic Networks with Hidden Variables". Machine Learning, 29(2-3), 213-244.

Insurance DataSet



Alarm DataSet

Description The ALARM ("A Logical Alarm Reduction Mechanism") is a Bayesian network designed to provide an alarm message system for patient monitoring.

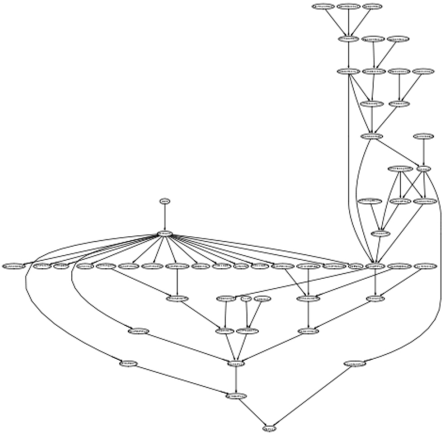
Number of nodes 37

Number of arcs 46

Number of parameters 509

Source Beinlich I, Suermondt HJ, Chavez RM, Cooper GF (1989).
"The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks."
In "Proceedings of the 2nd European Conference on Artificial Intelligence in Medicine", pp. 247-256. Springer-Verlag.

Alarm DataSet



HailFinder DataSet

Description Hailfinder is a Bayesian network designed to forecast severe summer hail in northeastern Colorado.

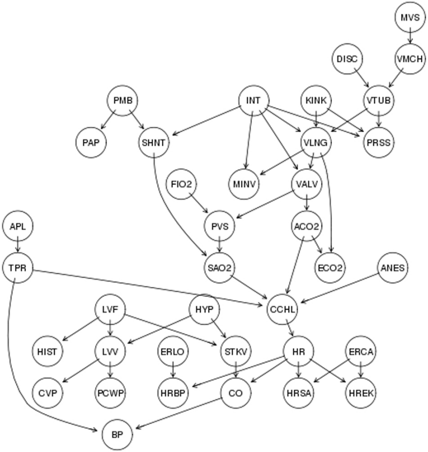
Number of nodes 56

Number of arcs 66

Number of parameters 2656

Source Abramson B, Brown J, Edwards W, Murphy A, Winkler RL (1996). "Hailfinder: A Bayesian system for forecasting severe weather". International Journal of Forecasting, 12(1), 57-71.

HailFinder DataSet



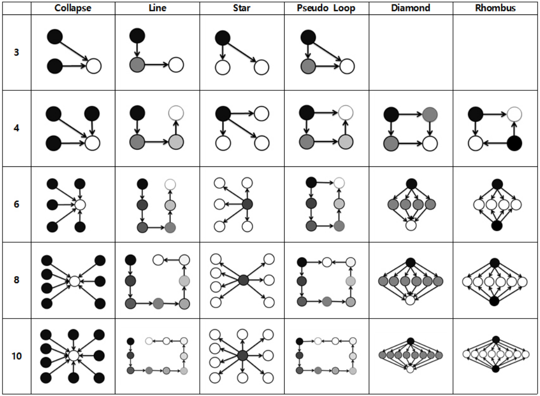
Summary

Sample Size 1000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
Asia	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	3	2	1	4
Insurance	2	1	3	4	2	1	3	4	3	4	2	1	2	1	4	3	1	2	4	3
Alarm	2	1	3	4	2	1	3	4	3	4	2	1	1	2	3	4	1	2	3	4
HaliFinder	2	1	3	4	2	1	3	4	4	4	2	1	1	2	3	4	1	2	3	4

Sample Size 5000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
Asia	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	3	2	1	4
Insurance	2	1	3	4	2	1	3	4	3	4	2	1	1	2	3	4	1	2	3	4
Alarm	1	2	3	4	2	1	3	4	4	3	2	1	1	3	2	4	2	3	1	4
HaliFinder	1	1	3	4	1	1	3	4	4	4	2	1	4	4	4	4	2	2	1	4

Sample Size 10000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
Asia	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	3	1	2	4
Insurance	2	1	3	4	2	1	3	4	3	4	2	1	1	3	2	4	1	2	3	4
Alarm	2	1	3	4	2	1	3	4	4	4	2	1	1	2	3	4	1	2	3	4
HaliFinder	2	1	3	4	1	2	3	4	4	3	2	1	2	1	4	4	3	2	1	4

Varying topologies and number of nodes

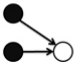
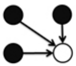


Eitel J. M. Lauría,
"An Information-Geometric Approach to Learning Bayesian Network Topologies from Data",
Innovations in Bayesian Networks Studies in Computational Intelligence Volume 156, 2008, pp 187-217

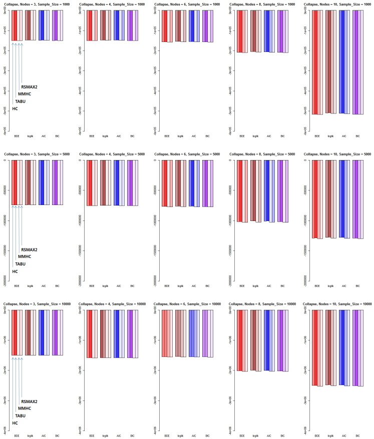
Prerequisite

- Cardinality was limited to two.
- The probability value, which is imparted optionally under U(0, 1) distribution.
- All experiments are repeated 100 times, and overall results are reported.
- Constraint-based Learning Algorithms often makes undirected arcs. So, this has been excluded from comparison.

Collapse

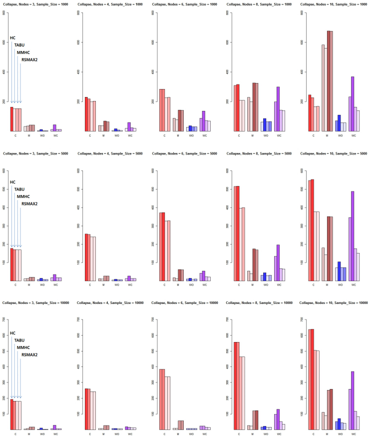
	3	4	6	8	10
Collapse					

Collapse (Score)



*The higher the value, the color is dark.

Collapse (Arcs)



*The higher the value, the color is dark.

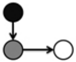
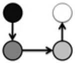
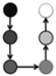
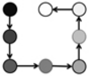
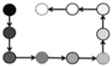
Collapse

Sample Size 1000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	2	1	4	4	1	4	2	2	4	3	1	1	4	1	4	4	4	1	4	4
4	2	1	4	3	1	2	4	3	3	4	1	2	4	1	2	4	4	1	2	4
6	2	1	4	3	1	1	4	3	3	4	1	2	4	1	2	3	2	1	3	4
8	2	1	3	4	2	1	4	4	3	4	1	2	4	1	3	2	2	1	3	4
10	2	1	3	4	1	2	4	3	3	4	1	2	2	1	4	4	2	1	3	4

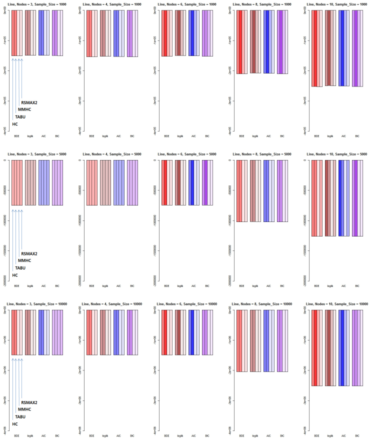
Sample Size 5000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	2	1	4	4	1	2	4	4	4	3	1	1	4	1	4	4	4	1	4	4
4	2	1	4	4	1	2	4	4	3	4	1	1	4	1	2	2	4	1	4	4
6	2	1	4	3	2	1	4	3	3	4	1	2	2	1	4	2	2	1	3	4
8	2	1	4	3	2	1	4	3	3	4	1	2	2	1	4	2	2	1	3	4
10	2	1	3	4	2	1	4	4	3	4	1	2	4	1	3	2	2	1	3	4

Sample Size 10000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	2	1	4	4	1	2	4	4	4	3	1	1	4	1	4	4	4	1	4	4
4	2	1	4	4	1	2	4	4	4	4	1	1	2	1	4	4	1	2	4	4
6	1	1	3	4	1	1	4	4	4	4	1	1	1	1	1	1	1	1	3	4
8	2	1	3	4	1	2	3	4	3	4	2	1	2	1	4	4	2	1	3	4
10	2	1	3	4	2	1	3	4	3	4	2	1	2	1	3	4	2	1	3	4

Line

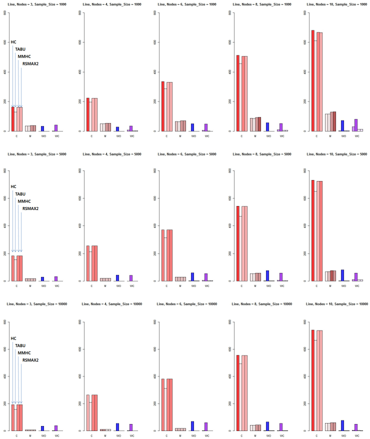
	3	4	6	8	10
Line					

Line (Score)



*The higher the value, the color is dark.

Line (Arcs)



*The higher the value, the color is dark.

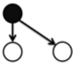
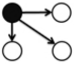
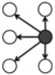
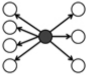
Line

Sample Size 1000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	1	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	2	1	4	4
4	1	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	2	1	4	4
6	2	1	3	4	1	4	2	2	3	4	1	1	4	1	4	4	2	1	3	4
8	2	1	3	4	1	4	2	3	3	4	2	1	4	1	4	4	2	1	4	4
10	2	1	3	4	1	4	2	3	4	4	2	1	4	1	4	4	2	1	4	4

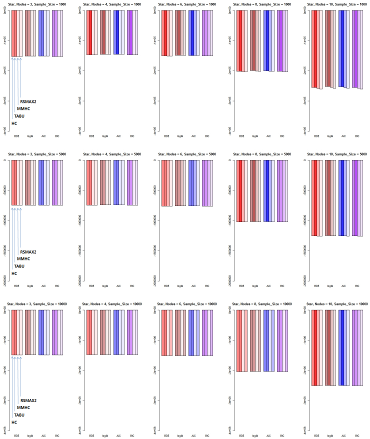
Sample Size 5000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	1	1	1	1	1	4	1	1	1	1	1	1	4	1	4	4	4	1	4	4
4	1	1	1	1	1	4	1	1	1	1	1	1	4	1	4	4	4	1	4	4
6	2	1	4	4	1	4	1	1	1	1	1	1	4	1	4	4	4	1	4	4
8	1	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	2	1	4	4
10	1	1	4	3	1	4	3	2	4	3	1	2	4	1	4	4	2	1	4	4

Sample Size 10000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	1	1	4	4	1	4	1	1	1	1	1	1	4	1	4	4	2	1	4	4
4	1	1	4	4	3	4	1	1	1	1	4	4	4	1	4	4	2	1	4	4
6	2	1	4	4	1	4	1	1	1	1	1	1	4	1	4	4	2	1	4	4
8	1	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	4	1	4	4
10	2	1	3	4	1	4	2	2	4	4	1	1	4	1	4	4	2	1	3	4

Star

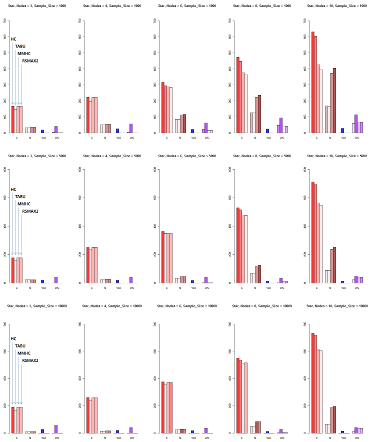
	3	4	6	8	10
Star					

Star (Score)



*The higher the value, the color is dark.

Star (Arcs)



*The higher the value, the color is dark.

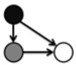
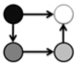
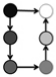
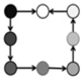
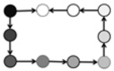
Star

Sample Size 1000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	1	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	2	1	4	4
4	2	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	2	1	4	4
6	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	2	1	4	4
8	2	1	3	4	1	2	3	4	3	4	2	1	4	1	4	4	2	1	4	3
10	2	1	3	4	1	2	3	4	3	4	2	1	4	1	4	4	4	1	3	2

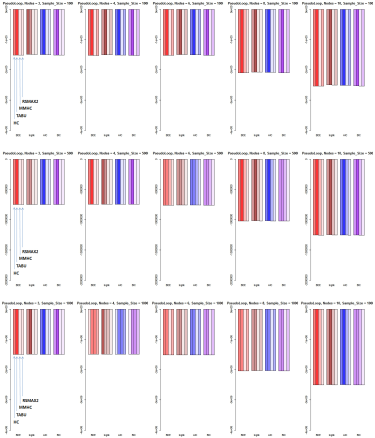
Sample Size 5000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	1	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	4	1	4	4
4	1	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	4	1	4	4
6	1	1	3	4	1	4	2	2	4	4	1	1	4	1	4	4	4	1	4	4
8	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	4	1	3	2
10	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	4	1	3	2

Sample Size 10000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	1	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	4	1	4	4
4	1	1	4	4	1	4	2	2	4	4	1	1	4	1	4	4	4	1	4	4
6	1	1	4	3	1	4	3	2	4	4	1	2	4	1	4	4	4	1	4	4
8	1	1	4	3	1	2	4	3	4	4	1	2	4	1	4	4	4	1	2	3
10	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	4	1	2	2

PseudoLoop

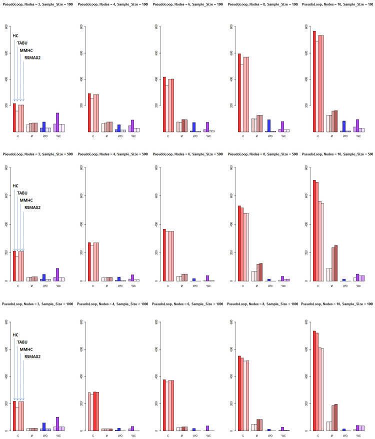
	3	4	6	8	10
Pseudo Loop					

PseudoLoop (Score)



*The higher the value, the color is dark.

PseudoLoop (Arc)



*The higher the value, the color is dark.

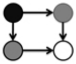
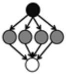
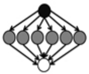
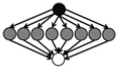
PseudoLoop

Sample Size 1000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	2	1	4	4	1	4	2	2	4	3	1	1	2	1	4	4	2	1	4	4
4	2	1	4	3	1	4	2	2	4	3	1	1	2	1	4	4	2	1	4	4
6	2	1	4	3	1	4	3	2	3	4	1	2	2	1	4	4	2	1	4	4
8	2	1	3	4	1	4	2	2	3	4	1	1	4	1	4	4	2	1	4	4
10	2	1	4	3	1	4	2	3	4	3	2	1	4	1	4	4	2	1	3	4

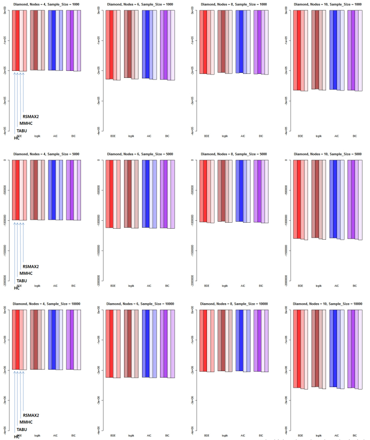
Sample Size 5000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	2	1	4	4	1	4	2	2	4	3	1	1	2	1	4	4	2	1	4	4
4	2	1	4	4	1	4	2	2	4	4	1	1	2	1	4	4	2	1	4	4
6	1	1	3	4	1	4	2	2	4	4	1	1	4	1	4	4	4	1	4	4
8	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	4	1	3	2
10	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	4	1	3	2

Sample Size 10000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
3	2	1	4	4	1	4	2	2	4	3	1	1	4	1	4	4	4	1	4	4
4	4	1	1	3	3	4	1	2	4	4	4	1	2	1	4	4	2	1	4	4
6	1	1	4	3	1	4	3	2	4	4	1	2	4	1	4	4	4	1	4	4
8	1	1	4	3	1	2	4	3	4	4	1	2	4	1	4	4	4	1	2	3
10	2	1	3	4	1	2	3	4	4	4	2	1	4	1	4	4	4	1	2	2

Diamond

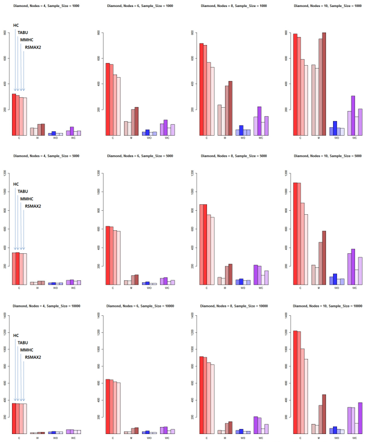
	3	4	6	8	10
Diamond					

Diamond (Score)



*The higher the value, the color is dark.

Diamond (Arc)



*The higher the value, the color is dark.

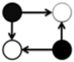
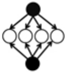
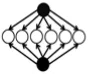
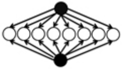
Diamond

Sample Size 1000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
4	2	1	4	3	1	2	3	4	3	4	2	1	2	1	4	4	2	1	4	3
6	2	1	3	4	1	2	3	4	3	4	2	1	2	1	4	3	2	1	4	3
8	2	1	3	4	1	2	3	4	3	4	2	1	4	1	4	4	3	1	4	2
10	2	1	3	4	1	2	3	4	3	4	2	1	2	1	3	4	3	1	4	2

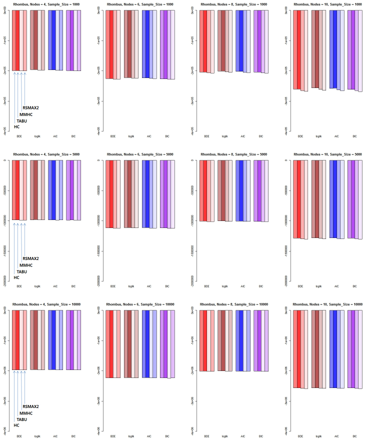
Sample Size 5000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
4	2	1	4	3	2	1	3	4	3	4	1	1	2	1	4	3	2	1	4	3
6	2	1	4	3	1	2	3	4	3	4	2	1	2	1	4	4	2	1	4	3
8	2	1	3	4	2	1	3	4	3	4	2	1	2	1	4	3	1	2	4	3
10	2	1	3	4	1	2	3	4	3	4	2	1	2	1	4	3	2	1	4	3

Sample Size 10000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
4	2	1	3	4	1	2	3	4	3	4	2	1	2	1	4	4	1	1	4	4
6	2	1	4	3	1	2	3	4	3	4	2	1	2	1	4	4	2	1	4	3
8	2	1	4	3	1	2	3	4	3	4	2	1	2	1	4	3	1	2	4	3
10	2	1	3	4	1	2	3	4	3	4	2	1	2	1	3	4	2	3	4	1

Rhombus

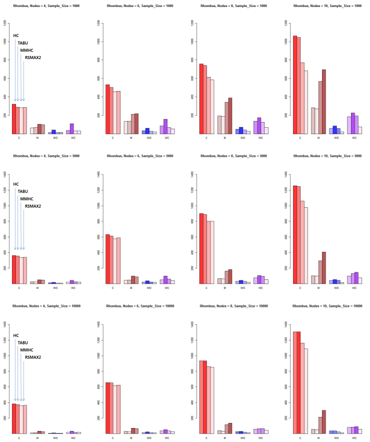
	3	4	6	8	10
Rhombus					

Rhombus (Score)



*The higher the value, the color is dark.

Rhombus (Arc)



*The higher the value, the color is dark.

Rhombus

Sample Size 1000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
4	2	1	4	3	1	2	4	3	4	3	1	2	4	1	4	4	2	1	4	4
6	2	1	3	4	1	2	4	3	4	3	2	1	2	1	3	4	2	1	3	4
8	2	1	3	4	1	2	3	4	3	4	2	1	2	1	3	4	2	1	3	4
10	2	1	3	4	1	2	3	4	3	4	2	1	3	1	2	4	3	1	2	4

Sample Size 5000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
4	2	1	4	3	1	2	4	3	3	4	1	2	2	1	4	4	4	1	2	4
6	2	1	3	4	1	2	4	3	4	3	1	2	3	1	2	4	3	1	2	4
8	2	1	3	4	1	2	3	4	3	4	2	1	2	1	2	4	3	1	2	4
10	2	1	3	4	1	2	3	4	3	4	2	1	3	1	2	4	3	2	1	4

Sample Size 10000	Score				C				M				WO				WC			
	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2	HC	TABU	MMHC	RSMAX2
4	2	1	4	3	1	2	4	3	4	3	1	2	4	1	4	4	4	1	2	4
6	2	1	4	3	1	2	4	3	3	4	1	2	2	1	2	4	2	1	2	4
8	2	1	4	3	2	1	3	4	3	4	2	1	2	1	3	4	3	1	1	4
10	2	1	3	4	2	1	3	4	3	4	2	1	1	1	3	4	3	2	1	4

Discussion

- In most cases using synthetic data according to topology,
  If comparing by score, then TABU search has good performance.
  But comparing by reference to "What C is the lot?", then HC has also good performance.
- Hybrid algorithm compared to Score-based algorithm is found to be that draw the arc more conservative.
- About Line and Star form, the performance difference due to relatively algorithm was not large compared to other topology.