Bayesian Networks

April 30, 2012

Sung-Bae Cho
Dept. of Computer Science, Yonsei University

Outline

- Bayesian Network
  - Inference of Bayesian Network
  - Modeling of Bayesian Network
- Bayesian Network Application
  - Application Example
- Summary & Review

Probabilistic Paradigm

- Advantages
  - Can accommodate inaccurate models
  - Can accommodate imperfect sensors
  - Robust in real-world applications

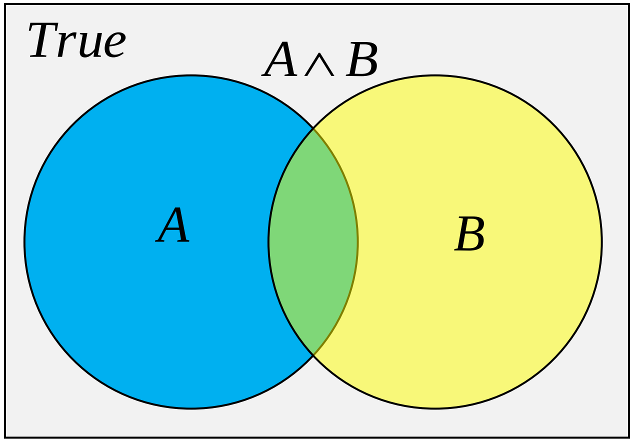
- Pitfalls
  - Computationally demanding
  - False assumptions
  - Approximate

Probabilistic State Estimation

- Axioms of Probability Theory
  - Pr(*A*) denotes probability that proposition *A* is true

$$0 \leq \Pr(A) \leq 1$$
$$\Pr(True) = 1$$
$$\Pr(False) = 0$$

$$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \wedge B)$$



- Utilization

$$\begin{aligned}
\Pr(A \vee \neg A) &= \Pr(A) + \Pr(\neg A) - \Pr(A \wedge \neg A) \\
\Pr(True) &= \Pr(A) + \Pr(\neg A) - \Pr(False) \\
1 &= \Pr(A) + \Pr(\neg A) - 0 \\
\Pr(\neg A) &= 1 - \Pr(A)
\end{aligned}$$

Joint & Conditional Probability

- Joint and Conditional Probability
  - P(X=x and Y=y) = P(x,y)
  - If X and Y are independent then P(x,y) = P(x) P(y)
  - P(x | y) is the probability of x given y P(x | y) = P(x,y) / P(y)
    P(x,y) = P(x | y) P(y)
  - If X and Y are independent then P(x | y) = P(x)

- Law of Total Probability, Marginals

Discrete case

$$\sum_x P(x) = 1$$

$$P(x) = \sum_y P(x, y)$$

$$P(x) = \sum_y P(x \mid y) P(y)$$

Continuous case

$$\int p(x)\, dx = 1$$

$$p(x) = \int p(x, y)\, dy$$

$$p(x) = \int p(x \mid y) p(y)\, dy$$

Probability Calculation

- Bayes Formula

$$P(x,y) = P(x \mid y)P(y) = P(y \mid x)P(x)$$

$$\Rightarrow$$

$$\boxed{P(x \mid y) = \frac{P(y \mid x)\; P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}}$$

- Normalization

$$P(x \mid y) = \frac{P(y \mid x)\; P(x)}{P(y)} = \eta\; P(y \mid x)\, P(x)$$

$$\eta = P(y)^{-1} = \frac{1}{\sum_x P(y \mid x) P(x)}$$

- Conditioning
  - Total probability:

$$P(x \mid y) = \int P(x \mid y, z)\, P(z \mid y)\, dz$$

  - Bayes rule and background knowledge:

$$P(x \mid y, z) = \frac{P(y \mid x, z)\; P(x \mid z)}{P(y \mid z)}$$

Outline

- **Bayesian Network**
  - Inference of Bayesian Network
  - Modeling of Bayesian Network
- Bayesian Network Application
  - Application Example
- Summary & Review

Rules of Probability

- Product Rule

$$P(X,Y) = P(X \mid Y)P(Y) = P(Y \mid X)P(X)$$

- Marginalization

$$P(Y) = \sum_{i=1}^{n} P(Y, x_i)$$

$$X \text{ binary: } P(Y) = P(Y,x) + P(Y,\bar{x})$$

- Bayes Rule

$$P(H,E) = P(H \mid E)P(E) = P(E \mid H)P(H)$$

$$\boxed{P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}}$$

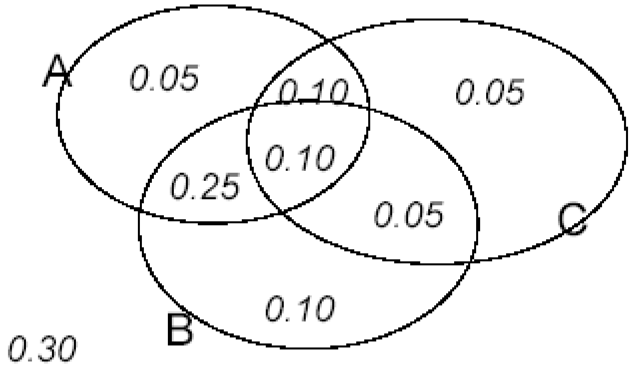
- Chain rule of probability

$$\begin{aligned} p(x) &= p(x_1,\ldots,x_n) \\ &= p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1,x_2)\ldots \\ &= \prod_{i=1}^{n} p(x_i \mid x_1,\ldots,x_{i-1}) \end{aligned}$$

The Joint Distribution

- Recipe for making a joint distribution of *M* variables
  - Make a truth table listing all combinations of values of your variables
  - For each combination of values, say how probable it is

A	B	C
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1



A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10

Using the Joint

- Once you have the joint distribution, you can ask for the probability of any logical expression involving your attribute

$$P(E) = \sum_{\text{rows matching } E} P(row)$$

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

Using the Joint

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

$$P(Poor\ Male) = 0.4654$$

$$P(Poor) = 0.7604$$

$$P(E) = \sum_{\text{rows matching } E} P(row)$$

$$P(E) = \sum_{\text{rows matching } E} P(row)$$

Inference with Joint

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

$$P(E_1 \mid E_2) = \frac{P(E_1, E_2)}{P(E_2)} = \frac{\sum_{\text{rows matching } E_1 \text{ and } E_2} P(row)}{\sum_{\text{rows matching } E_2} P(row)}$$

Inference with the Joint

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

$$P(Male \mid Poor) = 0.4654 / 0.7604 = 0.612$$

$$P(E_1 \mid E_2) = \frac{P(E_1, E_2)}{P(E_2)} = \frac{\sum_{\text{rows matching } E_1 \text{ and } E_2} P(row)}{\sum_{\text{rows matching } E_2} P(row)}$$

Joint Distributions

- Good news
  - Once you have a joint distribution, you can ask important questions about stuff that involves a lot of uncertainty
- Bad news
  - Impossible to create for more than about ten attributes because there are so many numbers needed when you build them

Bayesian Networks (1)

- In general
  - P(X_1, … X_n) needs at least 2^n – 1 numbers to specify the joint probability
  - Exponential storage and inference

- Overcome the problem of exponential size by exploiting conditional independence

REAL
Joint Probability Distribution
(2^n-1 numbers)

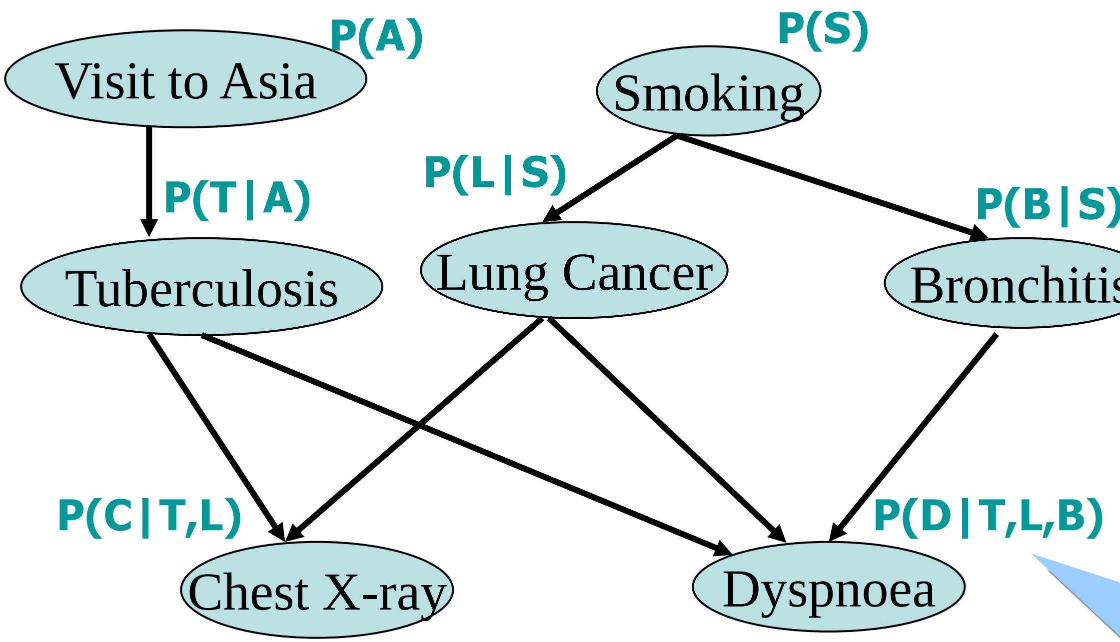
→

Conditional
Independence
(Domain knowledge or
Derived from data)

→

Graphical Representation of
Joint Probability Distribution
(Bayesian Network)

Bayesian Networks (2)



BN = (G, Θ)

CPD:

T	L	B	D=0	D=1
0	0	0	0.1	0.9
0	0	1	0.7	0.3
0	1	0	0.8	0.2
0	1	1	0.9	0.1
...				

$$P(A, S, T, L, B, C, D) = P(A)\,P(S)\,P(T|A)\,P(L|S)\,P(B|S)\,P(C|T,L)\,P(D|T,L,B)$$

Conditional Independencies ➡ Efficient Representation

[Lauritzen & Spiegelhalter, 95]

Bayesian Networks (3)

- Structured, graphical representation of probabilistic relationships between several random variables
- Explicit representation of conditional independencies
- Missing arcs encode conditional independence
- Efficient representation of joint pdf (probabilistic distribution function)
- Allows arbitrary queries to be answered
- P(lung cancer=yes | smoking=no, dyspnoea=yes)=?

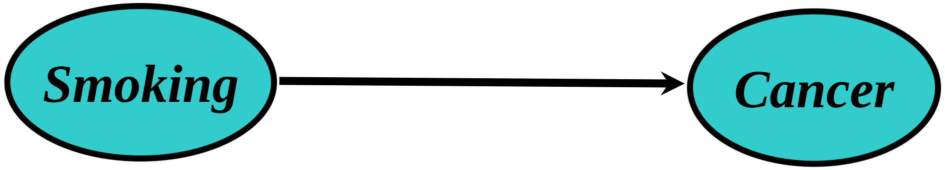
Bayesian Networks (4)

- Also called belief networks, and (directed acyclic) graphical models

- Bayesian network
  - Directed acyclic graph
    - Nodes are variables (discrete or continuous)
    - Arcs indicate dependence between variables
  - Conditional Probabilities (local distributions)

Bayesian Networks (5)

$S \in \{no, light, heavy\}$



$C \in \{none, benign, malignant\}$

P(S = no)	0.80
P(S = light)	0.15
P(S = heavy)	0.05

Smoking =	*no*	*light*	*heavy*
P(C = none)	0.96	0.88	0.60
P(C = benign)	0.03	0.08	0.25
P(C = malig)	0.01	0.04	0.15

Product Rule

- *P(C,S) = P(C|S) P(S)*

$S \Downarrow \quad C \Rightarrow$	*none*	*benign*	*malignant*
no	0.768	0.024	0.008
light	0.132	0.012	0.006
heavy	0.035	0.010	0.005

$S \Downarrow \quad C \Rightarrow$	*none*	*benign*	*malig*	total
no	0.768	0.024	0.008	.80
light	0.132	0.012	0.006	.15
heavy	0.035	0.010	0.005	.05
total	0.935	0.046	0.019	

P(Smoke)

P(Cancer)

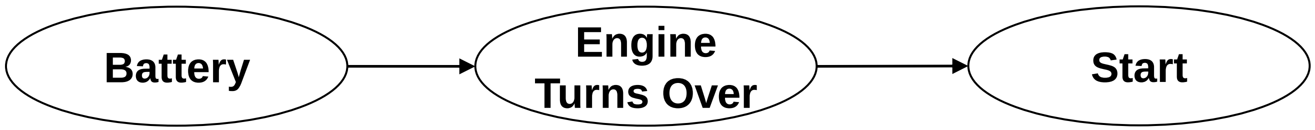
Bayes Rule

$$P(S\mid C)=\frac{P(C\mid S)P(S)}{P(C)}=\frac{P(C,S)}{P(C)}$$

$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malig*
no	0.768/.935	0.024/.046	0.008/.019
light	0.132/.935	0.012/.046	0.006/.019
heavy	0.030/.935	0.015/.046	0.005/.019

Cancer=	*none*	*benign*	*malignant*
P(S=no)	0.821	0.522	0.421
P(S=light)	0.141	0.261	0.316
P(S=heavy)	0.037	0.217	0.263

Missing Arcs Represent Conditional Independence



Start and Battery are independent, given Engine Turns Over

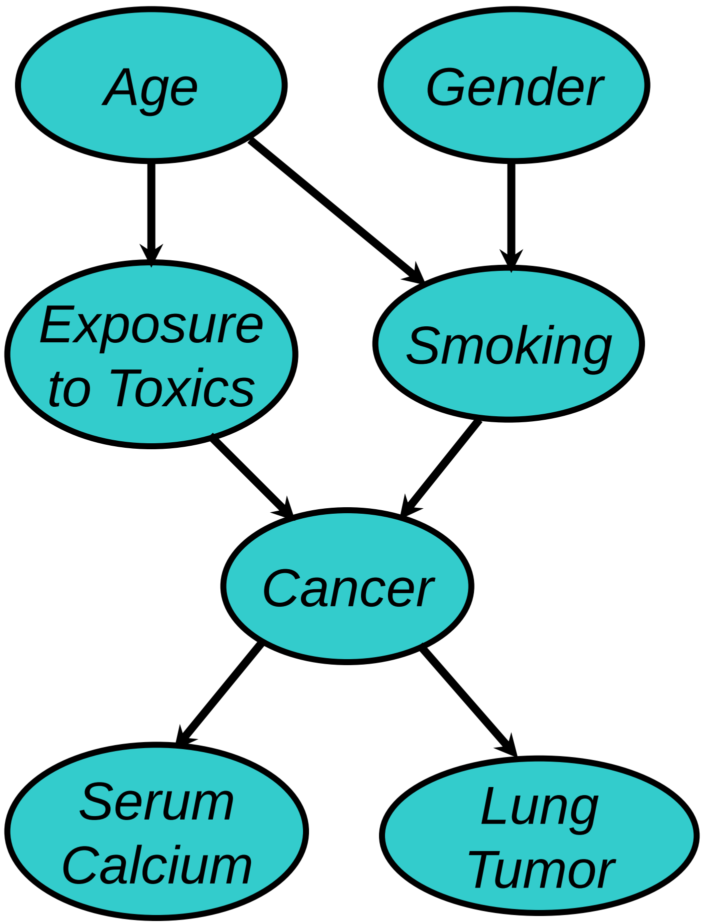
$$p(s \mid b, t) = p(s \mid t)$$

$$p(b, t, s) = p(b)\, p(t \mid b)\, p(s \mid t)$$

General product (chain) rule for Bayesian networks

$$p(x_1, x_2, \ldots, x_n) = \prod_{t=1}^{n} p(x_t \mid parents(x_t))$$

Bayesian Network



$$P(A, G, E, S, C, L, SC) =$$

$$P(A) \cdot P(G) \cdot$$

$$P(E \mid A) \cdot P(S \mid A, G) \cdot$$

$$P(C \mid E, S) \cdot$$

$$P(SC \mid C) \cdot P(L \mid C)$$

Outline

- Bayesian Network
  - **Inference of Bayesian Network**
  - Modeling of Bayesian Network
- Bayesian Network Application
  - Application Example
- Summary & Review

Inference

- We now have compact representations of probability distributions: Bayesian Networks

- Network describes a unique probability distribution *P*

- How do we answer queries about *P*?

- We use **inference** as a name for the process of computing answers to such queries

Inference: The Good & Bad News

- We can do inference
- We can compute any conditional probability
- P(Some variables | Some other variable values)

$$P(E_1 \mid E_2) = \frac{P(E_1, E_2)}{P(E_2)} = \frac{\sum_{\text{joint entries matching } E_1 \text{ and } E_2} P(row)}{\sum_{\text{joint entries matching } E_2} P(row)}$$

The sad, bad news

Conditional probabilities by enumerating all matching entries in the joint are expensive:
Exponential in the number of variables

Sadder and worse news

General querying of Bayesian networks is NP-complete

Hardness does not mean we cannot solve inference
It implies that we cannot find a general procedure that works efficiently for all networks
For particular families of networks, we can have provably efficient procedures

Queries: Likelihood

- There are many types of queries we might ask.

- Most of these involve **evidence**
  - Evidence $\boldsymbol{e}$ is an assignment of values to a set $\boldsymbol{E}$ variables in the domain
  - Without loss of generality $\boldsymbol{E} = \{ X_{k+1}, \ldots, X_n \}$

- Simplest query: compute probability of evidence

$$P(\boldsymbol{e}) = \sum_{x_1} \cdots \sum_{x_k} P(x_1, \ldots, x_k, \boldsymbol{e})$$

- This is often referred to as computing the **likelihood** of the evidence

Queries

- Often we are interested in the conditional probability of a variable given the evidence

$$P(X \mid e) = \frac{P(X, e)}{P(e)}$$

- This is the **a posteriori belief** in X, given evidence $\boldsymbol{e}$
- A related task is computing the term $P(X, \boldsymbol{e})$
  - i.e., the likelihood of $\boldsymbol{e}$ and $X = x$ for values of X
  - we can recover the a posteriori belief by

$$P(X = x \mid \boldsymbol{e}) = \frac{P(X = x, \boldsymbol{e})}{\sum_{x'} P(X = x', \boldsymbol{e})}$$

A Posteriori Belief

This query is useful in many cases:

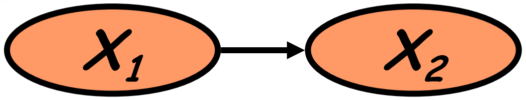
- **Prediction**: what is the probability of an outcome given the starting condition
  – Target is a descendent of the evidence
- **Diagnosis**: what is the probability of disease/fault given symptoms
  – Target is an ancestor of the evidence
- As we shall see, the direction between variables does not restrict the directions of the queries
  – Probabilistic inference can combine evidence from all parts of the network

Approaches to Inference

- Exact inference
  - Inference in Simple Chains
  - Variable elimination
  - Clustering / join tree algorithms

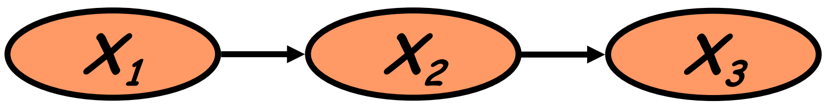
- Approximate inference
  - Stochastic simulation / sampling methods
  - Markov chain Monte Carlo methods
  - Mean field theory

Inference in Simple Chains



- How do we compute $P(X_2)$?

$$P(x_2) = \sum_{x_1} P(x_1, x_2) = \sum_{x_1} P(x_1) P(x_2 \mid x_1)$$



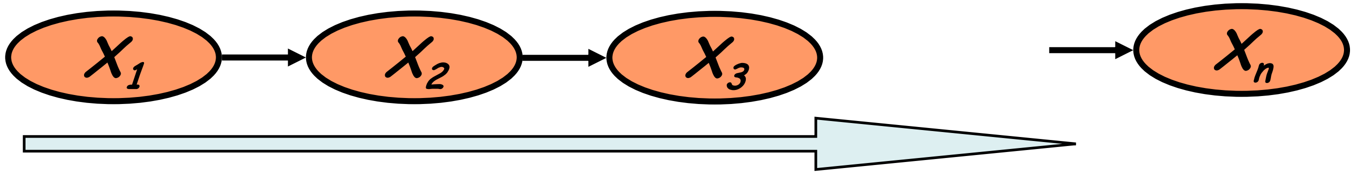
- How do we compute $P(X_3)$?

$$P(x_3) = \sum_{x_2} P(x_2, x_3) = \sum_{x_2} P(x_2) P(x_3 \mid x_2)$$

- we already know how to compute $P(X_2)$...

$$P(x_2) = \sum_{x_1} P(x_1, x_2) = \sum_{x_1} P(x_1) P(x_2 \mid x_1)$$

Inference in Simple Chains



How do we compute $P(X_n)$?

- Compute $P(X_1)$, $P(X_2)$, $P(X_3)$, …
- We compute each term by using the previous one

$$P(x_{i+1}) = \sum_{x_i} P(x_i)P(x_{i+1} \mid x_i)$$

- Complexity:
  - Each step costs $O(|Val(X_i)| \times |Val(X_i + 1)|)$ operations
- Compare to naïve evaluation, that requires summing over joint values of n-1 variables

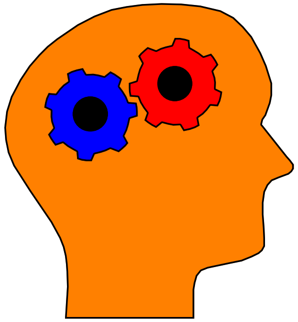
$$P(X_n) = \sum_{X_1, X_2, \ldots, X_{n-1}} P(X_1, X_2, \ldots, X_n)$$

Outline

- Bayesian Network
  - Inference of Bayesian Network
  - **Modeling of Bayesian Network**
- Bayesian Network Application
  - Application Example
- Summary & Review

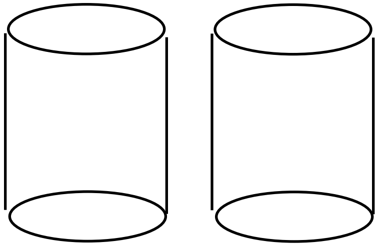
Why Learning ?

knowledge-based
(expert systems)



-Answer Wizard, Office 95, 97, & 2000
-Troubleshooters, Windows 98 & 2000

-Causal discovery
-Data visualization
-Concise model of data
-Prediction



data-based

Why Learning?

- Knowledge acquisition is bottleneck
  - Knowledge acquisition is an expensive process
  - Often we don't have an expert

- Data is cheap
  - Amount of available information growing rapidly
  - Learning allows us to construct models from raw data

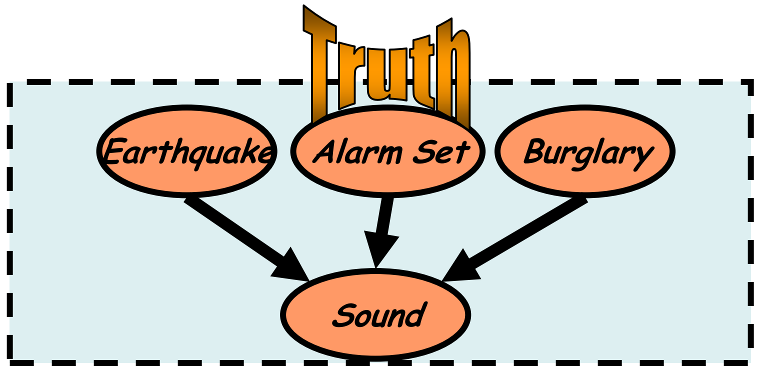
- Conditional independencies & graphical language capture structure of many real-world distributions

- Graph structure provides much insight into domain
  - Allows "knowledge discovery"

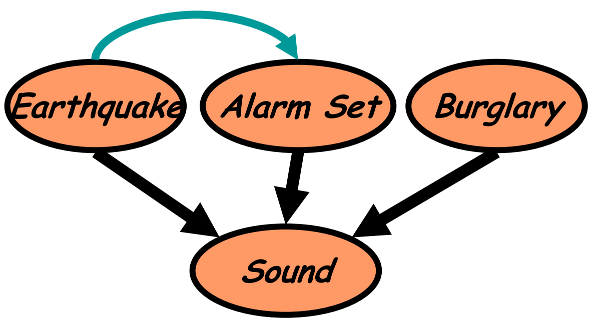
- Learned model can be used for many tasks

- Supports all the features of probabilistic learning
  - Model selection criteria
  - Dealing with missing data & hidden variables

Why Struggle for Accurate Structure?

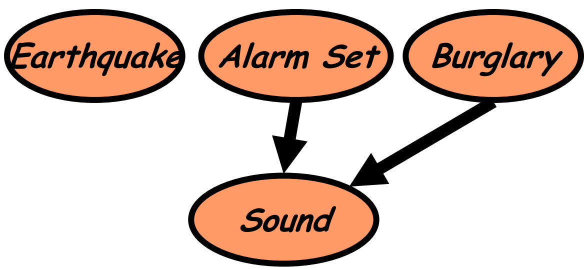


Adding an arc



- Increases the number of parameters to be fitted
- Wrong assumptions about causality and domain structure

Missing an arc



- Cannot be compensated by accurate fitting of parameters
- Also misses causality and domain structure

Learning Bayesian Networks from Data

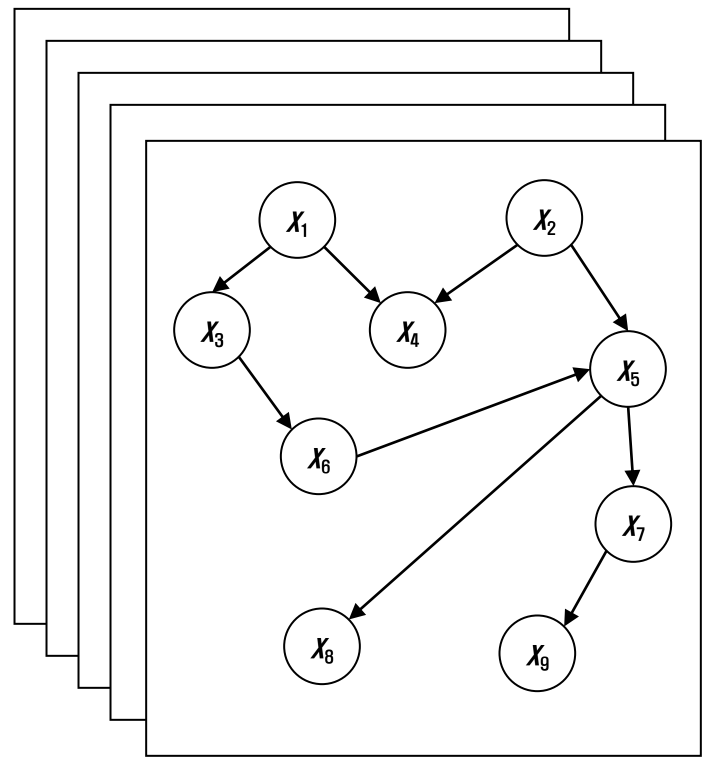
data

X_1	X_2	X_3	
True	1	0.7	
False	5	−1.6	
False	3	5.9	⋯
true	2	6.3	
	⋮		⋱

+
Prior/expert information

Bayesian network learner →

Bayesian networks



Learning Bayesian Networks

• **Known Structure, Complete Data** – Network structure is specified • Inducer needs to estimate parameters – Data do not contain missing values	• **Unknown Structure, Complete Data** – Network structure is not specified • Inducer needs to select arcs & estimate parameters – Data do not contain missing values
• **Known Structure, Incomplete Data** – Network structure is specified – Data contain missing values • Need to consider assignments to missing values	• **Unknown Structure, Incomplete Data** – Network structure is not specified – Data contain missing values • Need to consider assignments to missing values

Two Types of Methods for Learning BNs

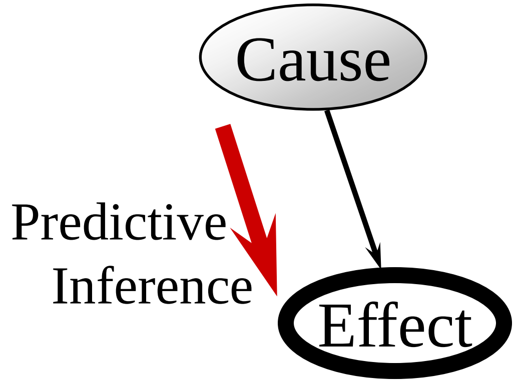
- Constraint based
  - Finds a Bayesian network structure whose implied independence constraints “match” those found in the data
- Scoring methods (Bayesian, MDL, MML)
  - Find the Bayesian network structure that can represent distributions that “match” the data (i.e., could have generated the data)
- Practical considerations
  - The number of possible BN structures is super exponential in the number of variables.
  - How do we find the best graph(s)?

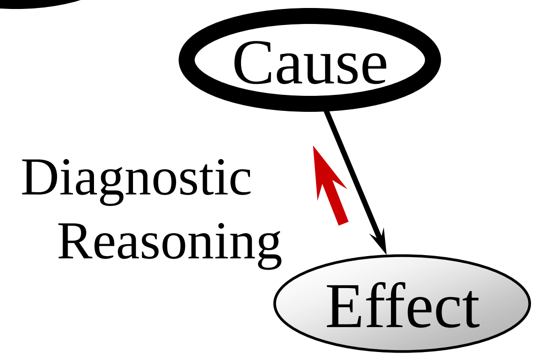
Outline

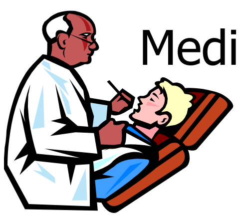
- Bayesian Network
  - Inference of Bayesian Network
  - Modeling of Bayesian Network
- **Bayesian Network Application**
  - Application Example
- Summary & Review

What are BNs useful for?

- Prediction: P(symptom|cause)=?
- Diagnosis: P(cause|symptom)=?
- Classification: $\max_{class} P(class|data)$
- Decision-making (given a cost function)
- Data mining: induce best model from data

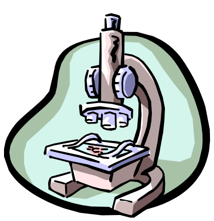






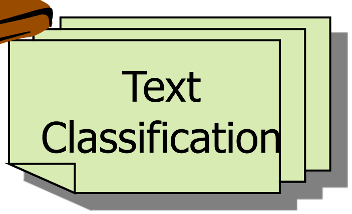
Medicine

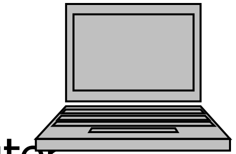
Speech recognition



Bio-informatics

Stock market





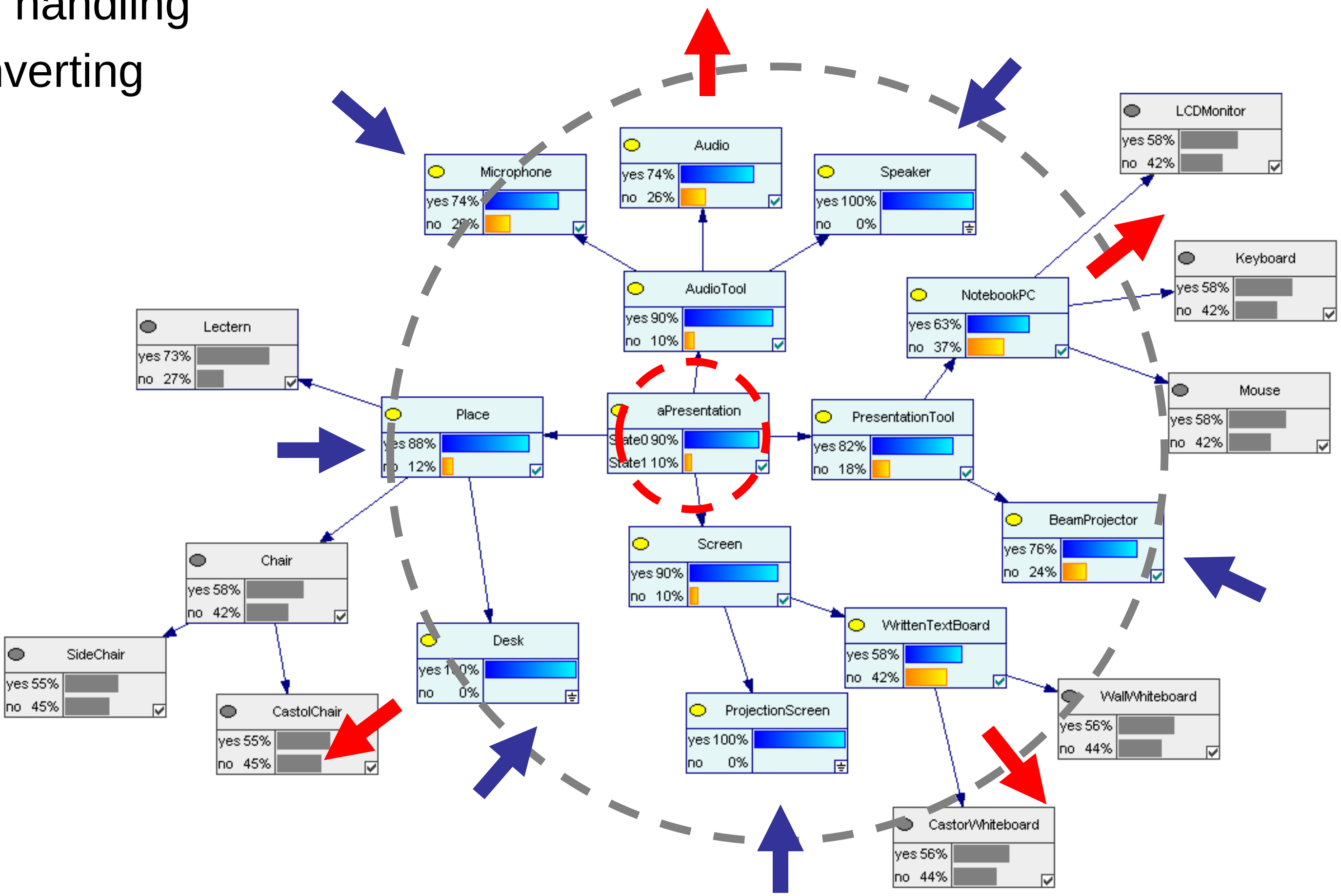
Computer troubleshooting

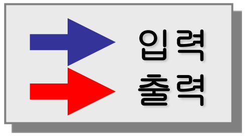
Why use BNs?

- Explicit management of uncertainty
- Predictive-Diagnostic Reasoning
- Modularity (modular specification of a joint distribution) implies maintainability
- Better, flexible and robust decision making – MEU (Maximization of Expected Utility), VOI (Value of Information)
- Can be used to answer arbitrary queries - multiple fault problems (General purpose "inference" algorithm)
- Easy to incorporate prior knowledge
- Easy to understand

Uncertainty Handling

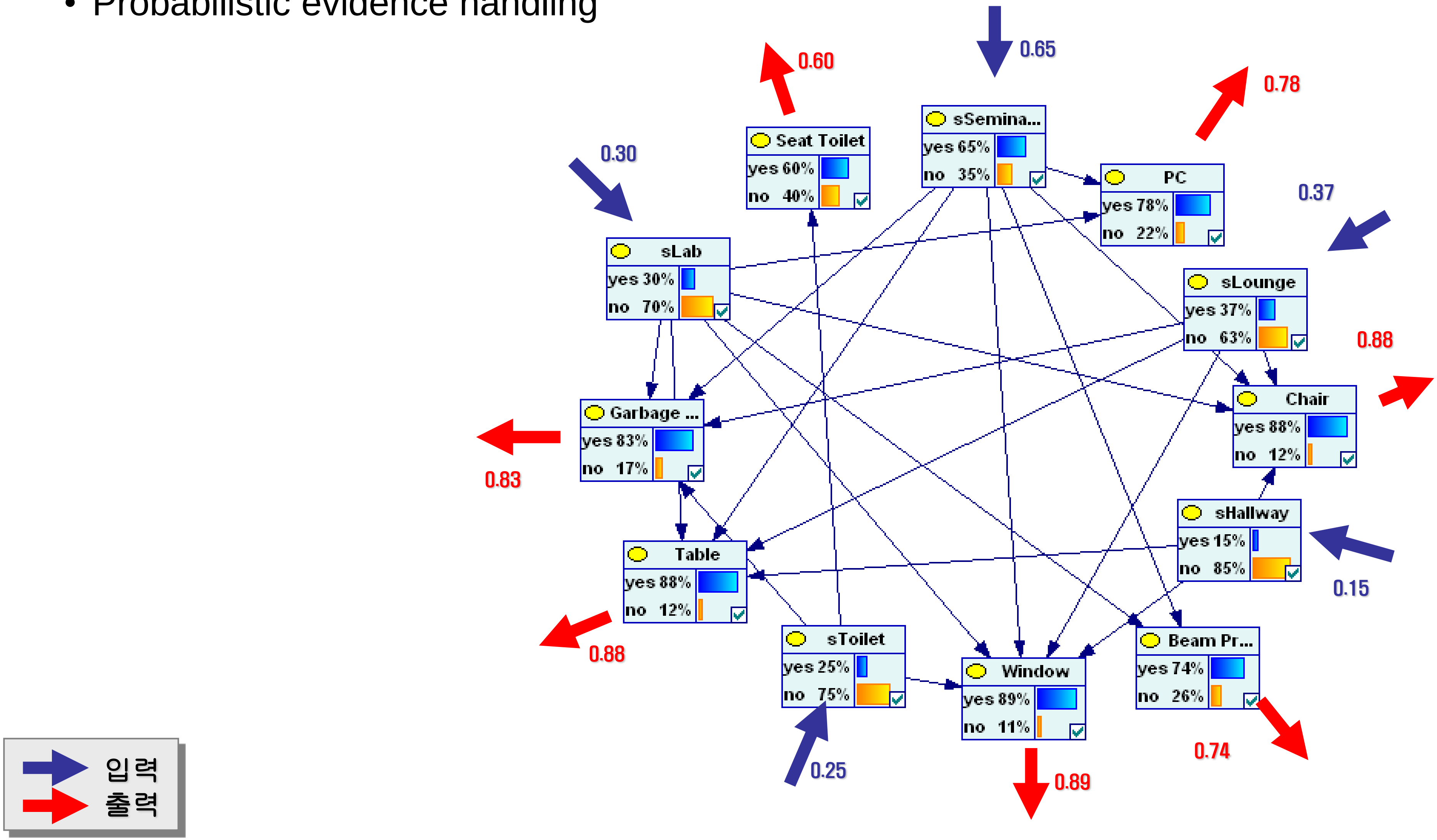
- Partial evidence handling
- Input/output converting



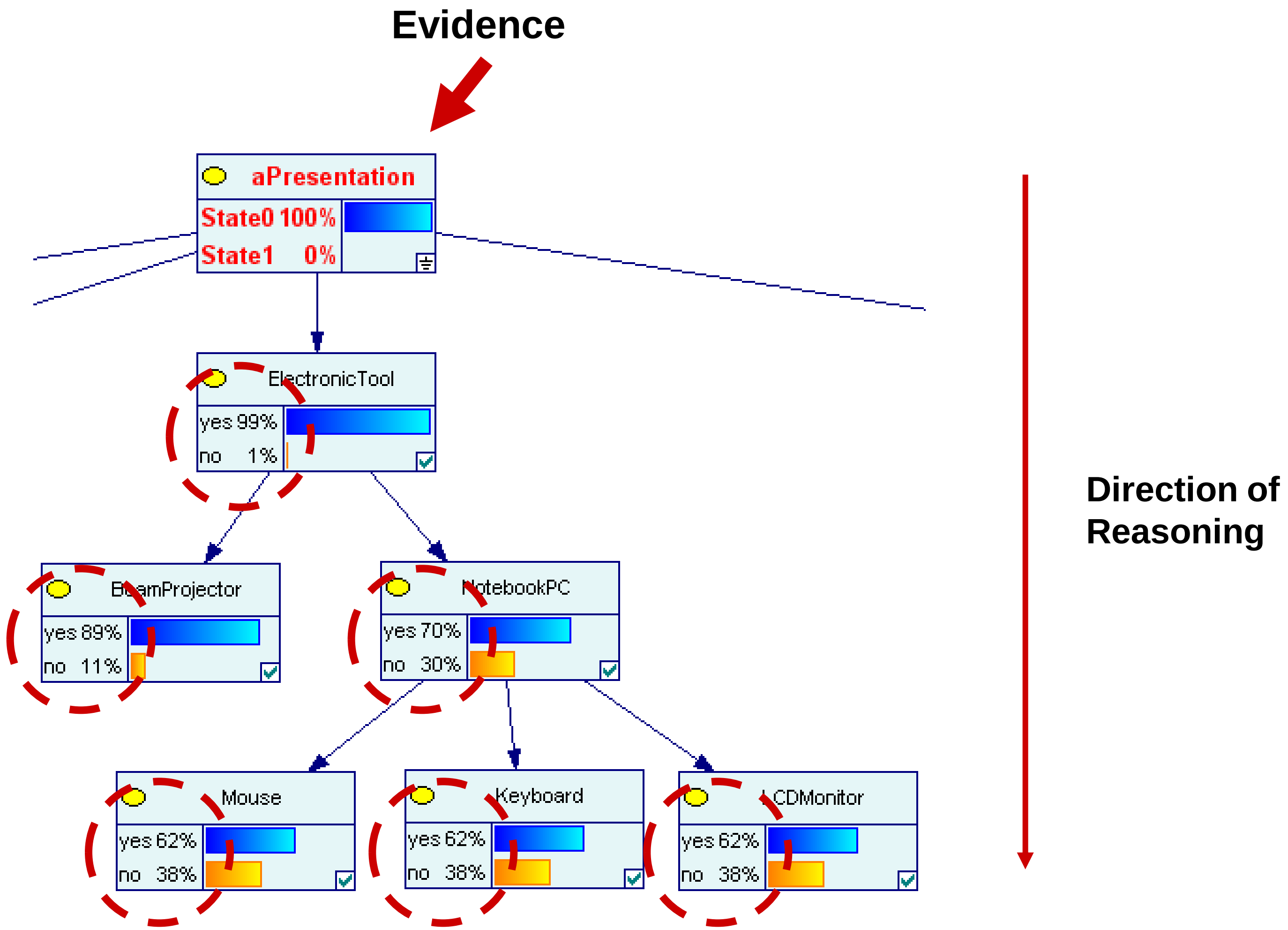


Utilization of Probability Evidence

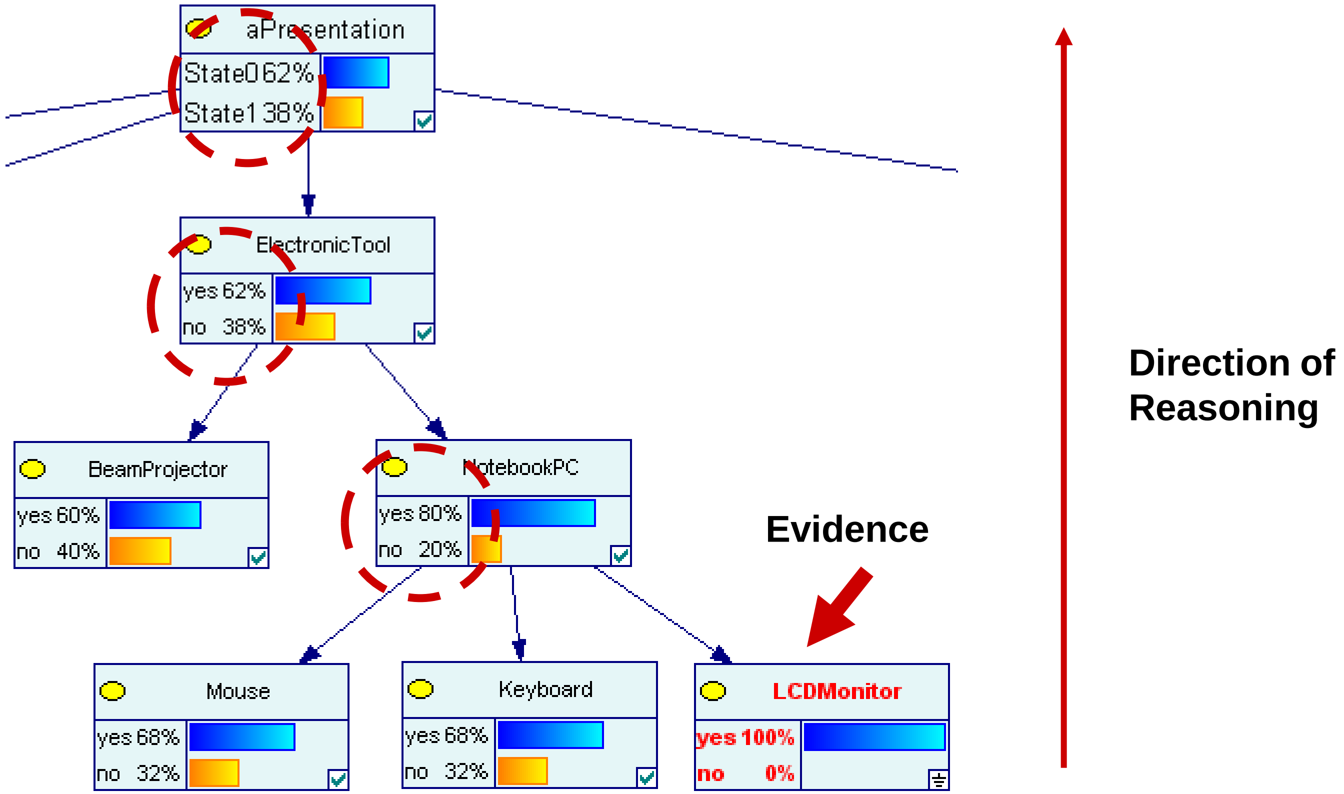
- Probabilistic evidence handling



Predictive Reasoning

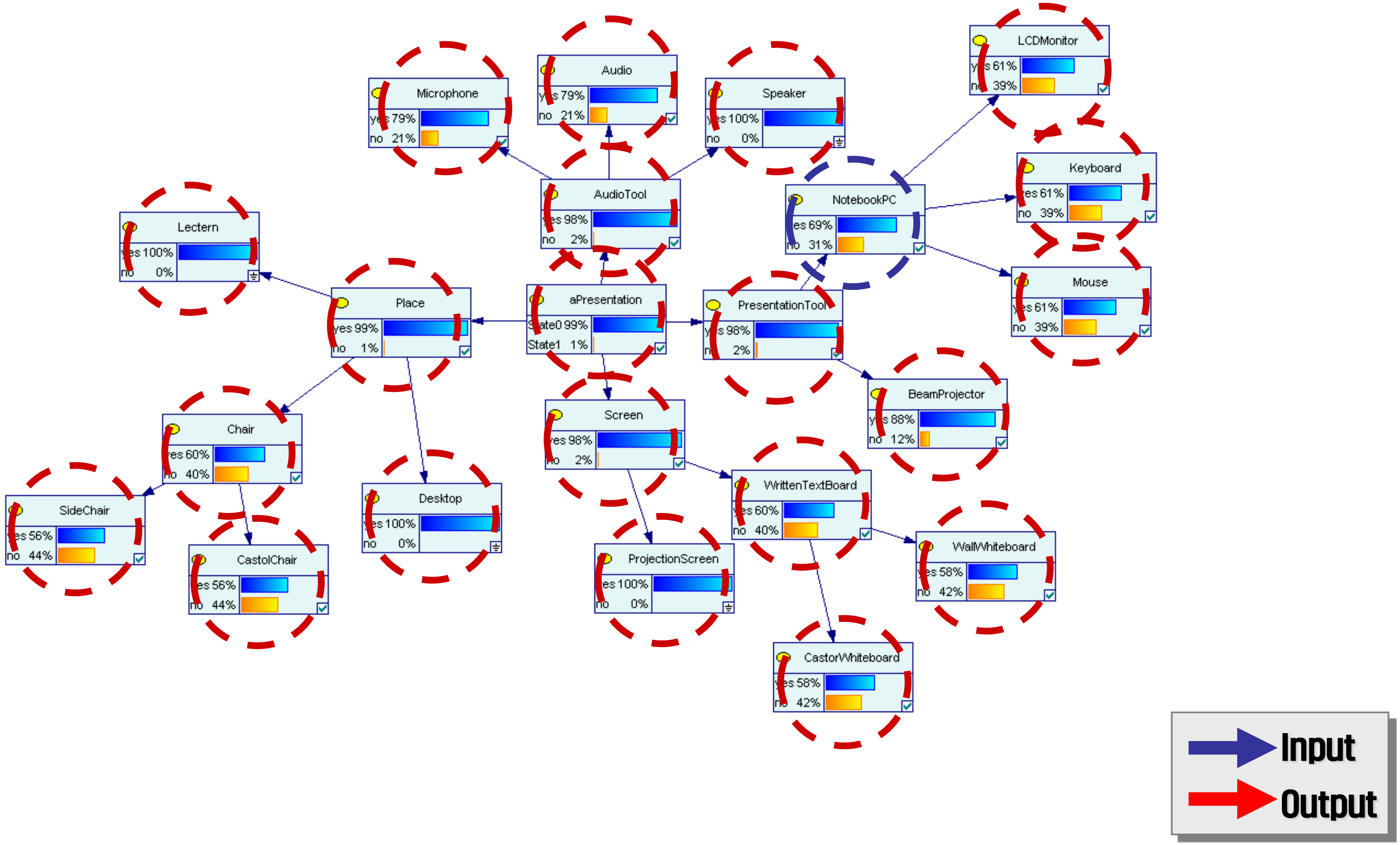


Diagnostic Reasoning



Predictive-Diagnostic Reasoning

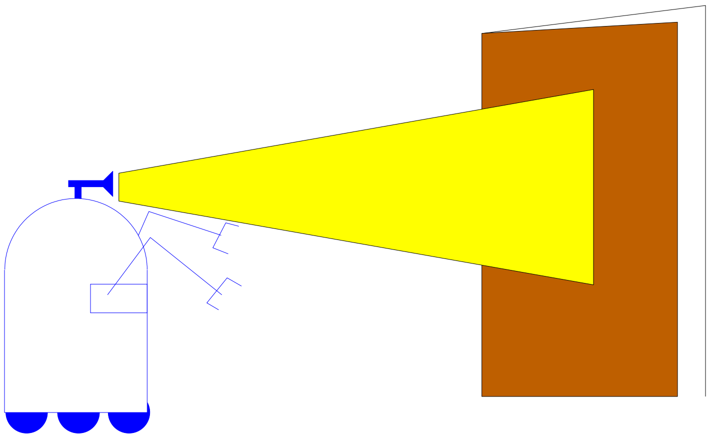
- Can get probability distribution of multiple target nodes



App. Exam: BN Based Robot Intelligence

- (플래너): 문은 어디에 있는가?
  (도메인 정보): 벽 쪽, 네모난 형태
- (플래너): 문 통과가 가능한가?
- (영상필터): 문의 종류(미는 문), 문이 열린 상태(조금 열림), 문과의 거리
- (추론): 통과 가능 판단
- (플래너): 문으로 이동, 장애물 발견
- (추론): 통과 가능성 낮아짐
- (추론): 장애물 이동 방향 입력 결과 통과 가능성 높아짐
- (플래너): 문 통과 및 이동:
  거실->거실 문->주방

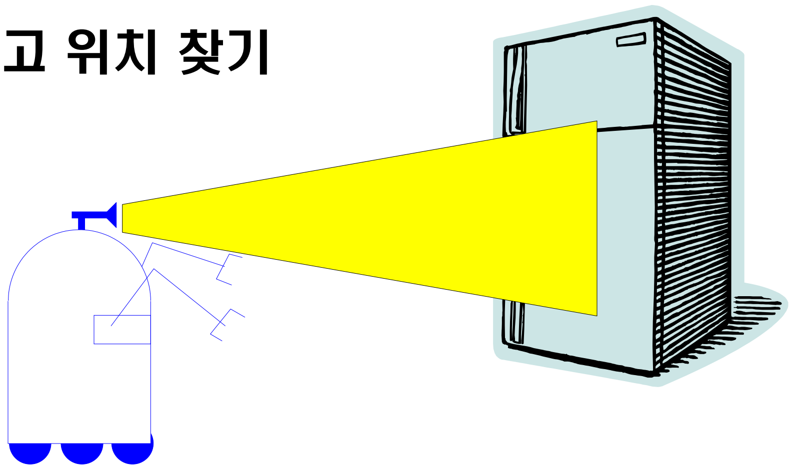
- BN을 이용한
  문 통과 가능성 추론 및
  장애물 판단



App. Exam: BN Based Robot Intelligence (2)

- (플래너): 냉장고 예상 위치는?
- (도메인 지식): 벽면 쪽
- (추론): 발견된 물체(싱크대) 근처, 벽면 쪽

- 냉장고 위치 찾기



- (영상필터): 냉장고 앞, 의자 발견
- (플래너): 냉장고 앞 이동 가능?
- (추론): 장애물 때문에 이동 불가, 치워야 함
- [영상필더]: 장애물=의자
- (추론): 의자는 로봇이 옆으로 움직일 수 있음
- (추론): 유저에게 치워도 되는지 물어보시오!

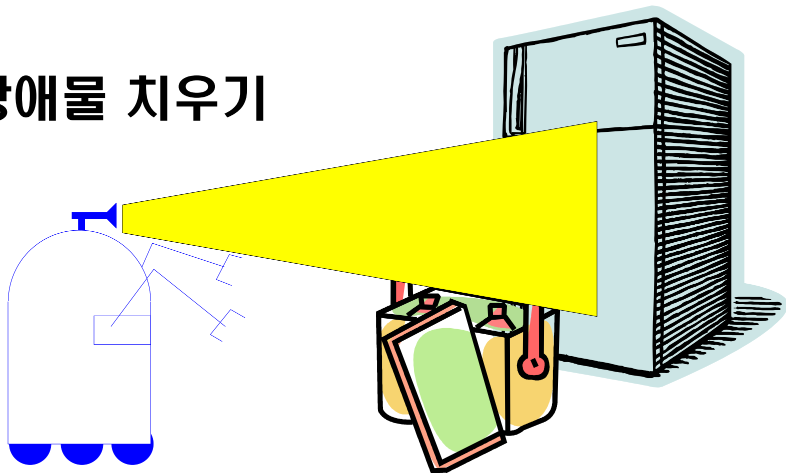
- 냉장고 앞으로 이동하기 위한 추론

App. Exam: BN Based Robot Intelligence (3)

- 로봇: "의자를 치우겠습니다. 괜찮습니까?"
- 유저: "그래 치워라."

- (플래너): 의자 옆으로 밀기
- (추론): 유저 위치=주방, 의자가 있을 위치= 식탁 옆
- (추론): 식탁이 가깝고, 식탁 옆에 공간 있으므로 그고스로 이동하시오.
- (플래너): 의자를 식탁으로 이동

- 장애물 치우기

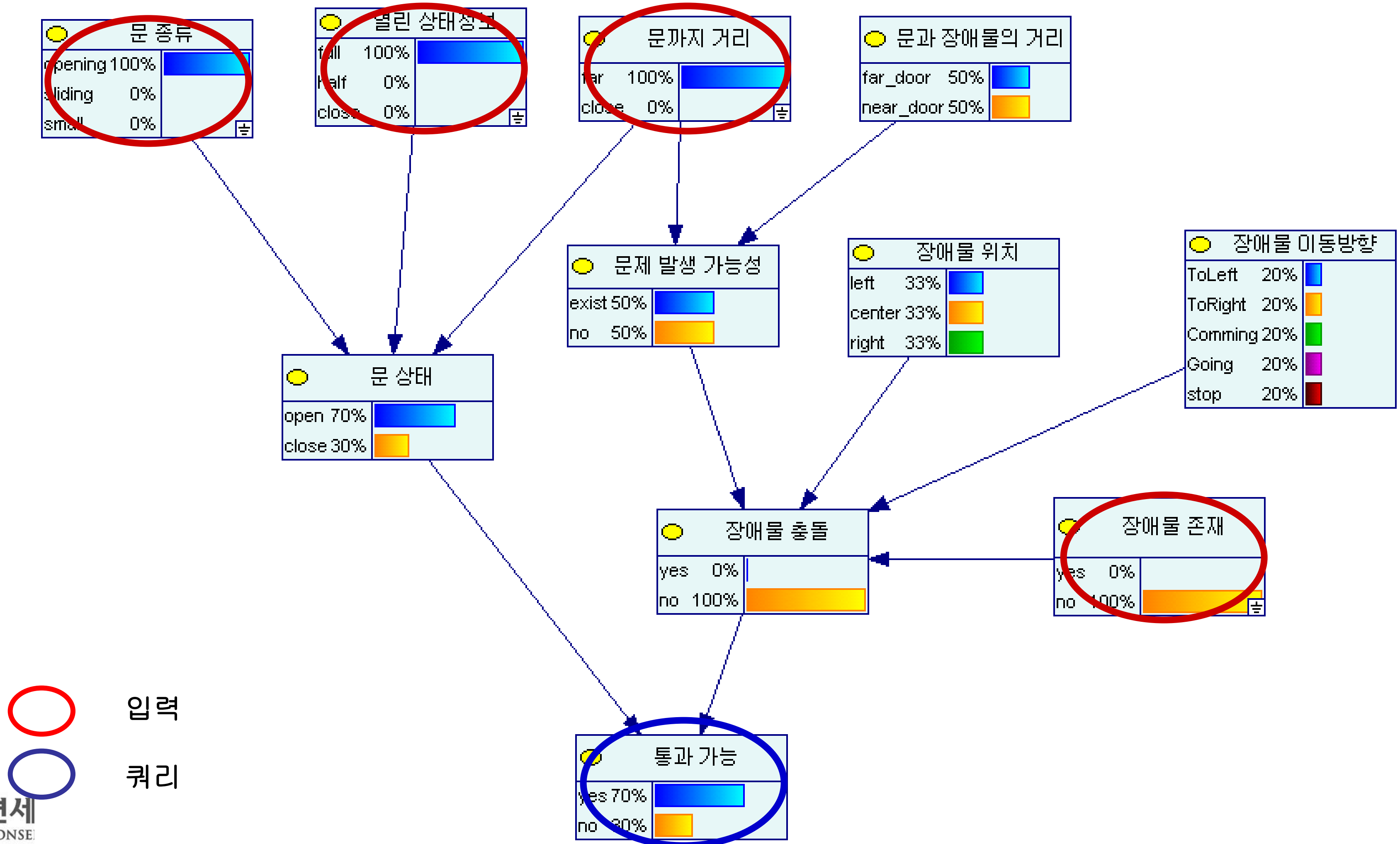


- (플래너): 냉장고 앞으로 이동, 냉장고 문 열기 수행
- [플래너]: 냉장고 문열기 중 실패
- (추론): 냉장고 옆에 다른 장애물이 있는 것으로 추정됨

- 냉장고 앞의 다른 장애물 추론

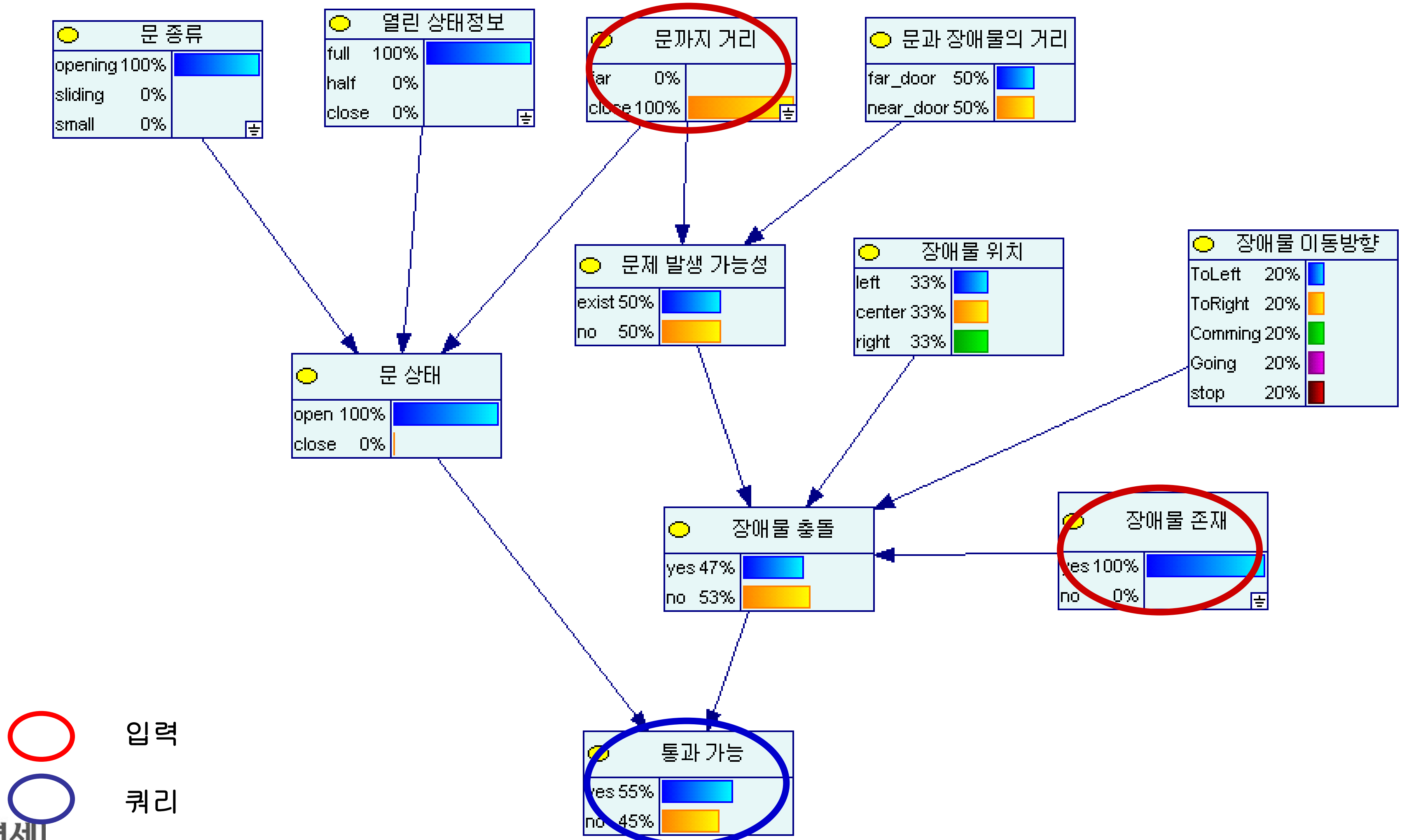
1. Door Passability

- Given: **조금 먼 거리**, **문은 여닫이 문**, **열려 있음**, **장애물은 안 보임**
- Output: **문 통과 가능성** 70%



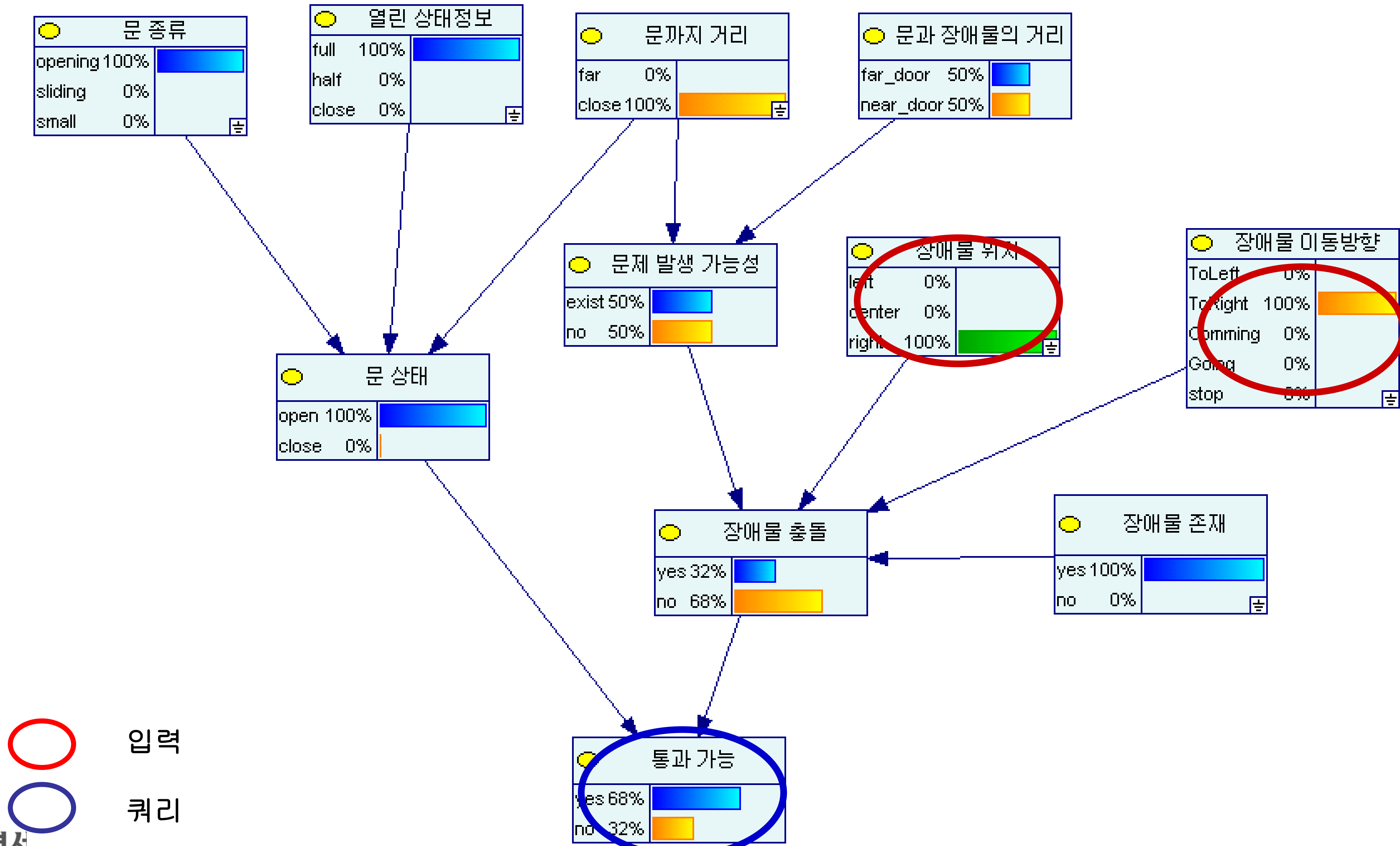
2. Approaching & Obstacle Detection

- Given: **문으로 접근**, **장애물 발견**
- Output: **문 통과 가능성** 55%



3. Inference of Obstacle's Status

- Given: **장애물 위치는 <u>오른편</u>, 장애물은 <u>오른쪽으로 이동중</u>**
- Output: **문 통과 가능성** 68%

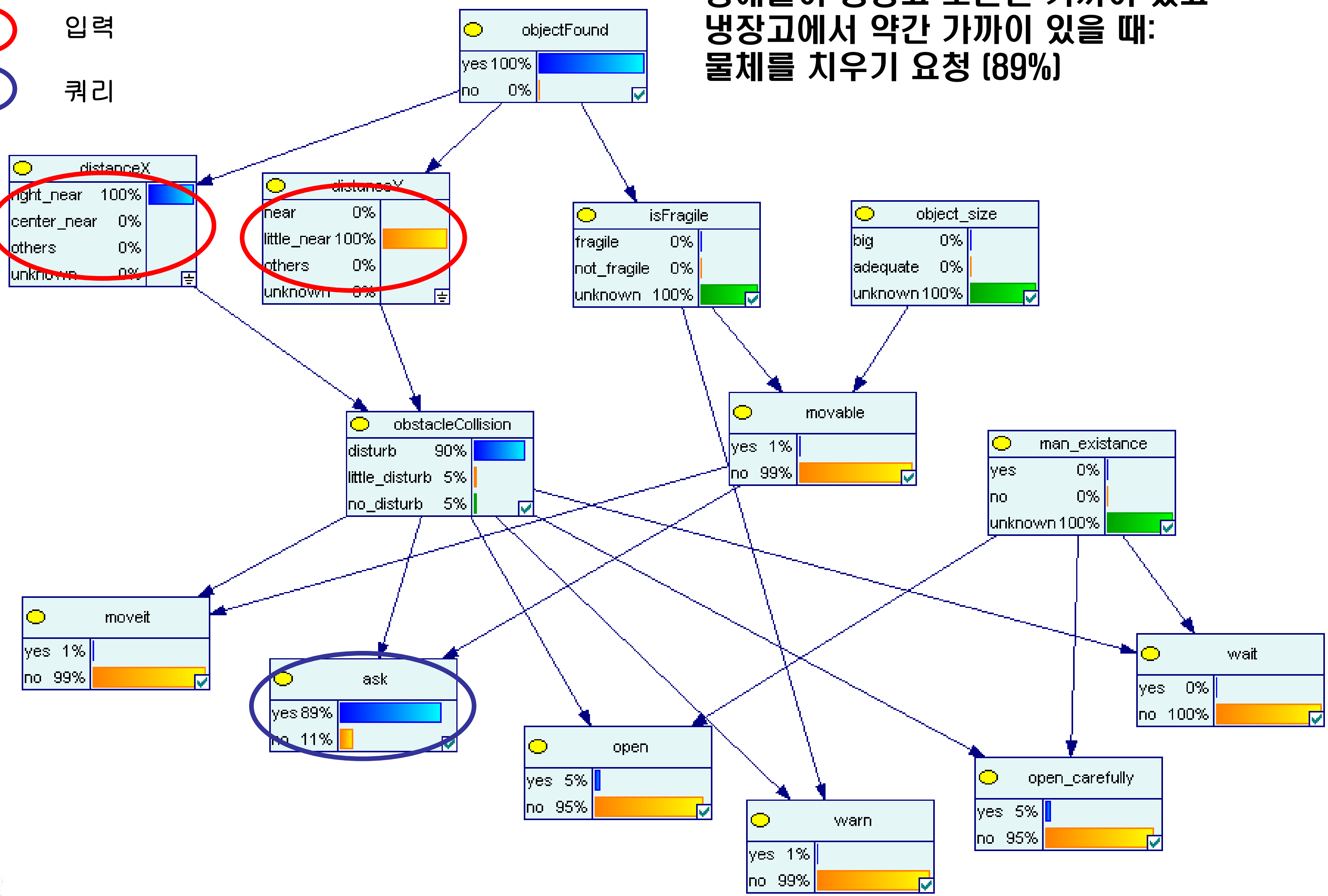


4. Approaching Refrigerator

입력

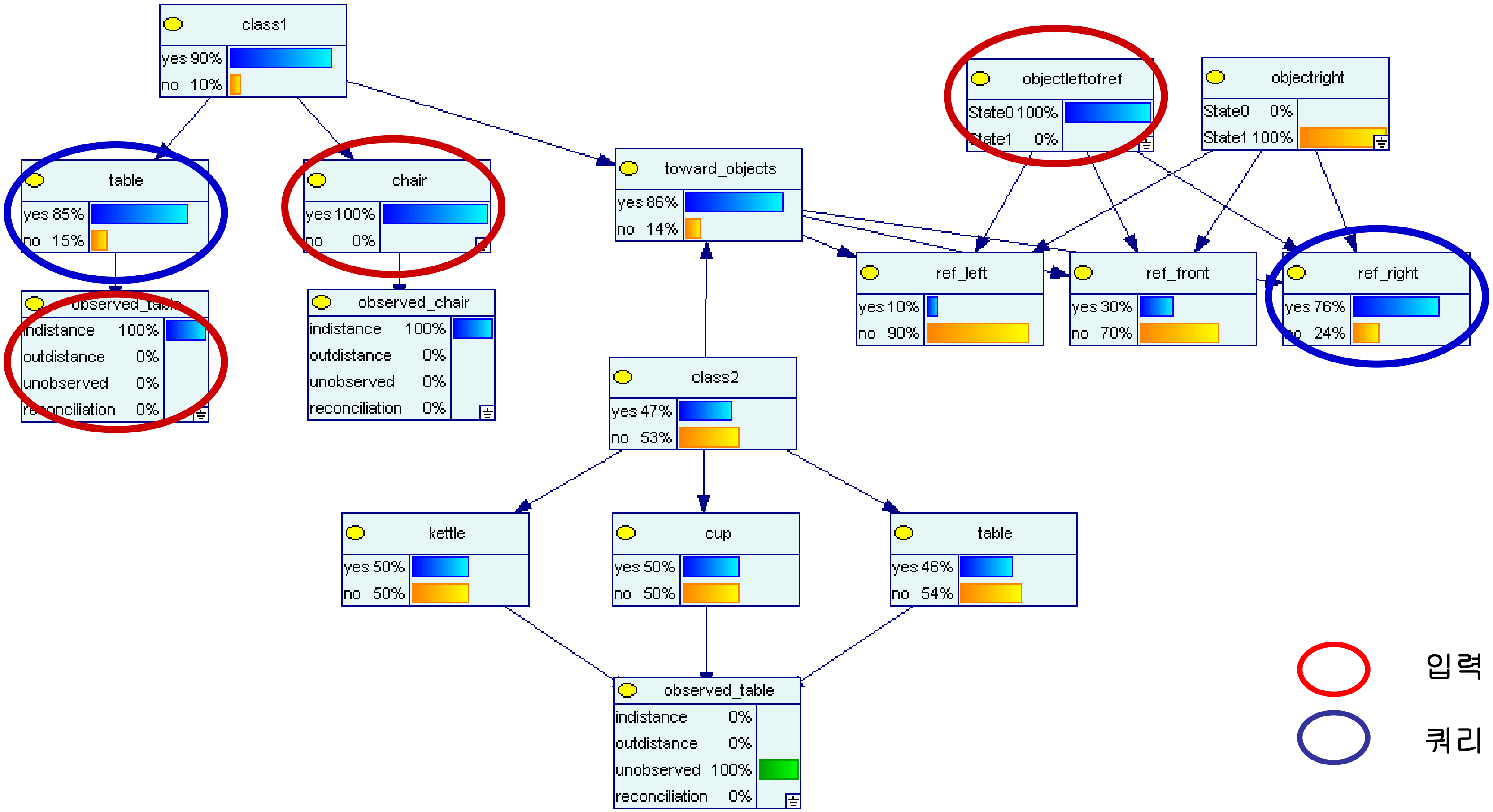
쿼리

장애물이 냉장고 오른편 가까이 있고
냉장고에서 약간 가까이 있을 때:
물체를 치우기 요청 (89%)



5. Removing Obstacle

- Given: 냉장고 왼편에 물체, 냉장고 앞에 의자. 테이블 가까움
- Output: 장애물 치우기! 추천 위치 1. 테이블 근처 85% > 2. 냉장고 오른쪽 76%



Applications

- Industrial
  - Processor Fault Diagnosis - by Intel
  - Auxiliary Turbine Diagnosis - GEMS by GE
  - Diagnosis of space shuttle propulsion systems - VISTA by NASA/Rockwell
  - Situation assessment for nuclear power plant – NRC
- Military
  - Automatic Target Recognition - MITRE
  - Autonomous control of unmanned underwater vehicle - Lockheed Martin
  - Assessment of Intent
- Medical Diagnosis
  - Internal Medicine
  - Pathology diagnosis - Intellipath by Chapman & Hall
  - Breast Cancer Manager with Intellipath
- Commercial
  - Financial Market Analysis
  - Information Retrieval
  - Software troubleshooting and advice - Windows 95 & Office 97
  - Pregnancy and Child Care - Microsoft
  - Software debugging - American Airlines' SABRE online reservation system

Outline

- Bayesian Network
  - Inference of Bayesian Network
  - Modeling of Bayesian Network

- Bayesian Network Application
  - Application Example 1
  - Application Example 2

- **Summary & Review**

Summary & Review

- Motivation
  – Explicit representation of uncertainty using the calculus of probability theory

- Bayesian Network
  – Inference & Modeling Methods
  – Applications

- Ongoing research
  – How to use Bayesian intelligence for robot control
  – How to design probabilistic models
  – How to combine the deliberative method with reactive method

향후 일정

- 5/2: IT 표준화 최신현황 및 전망 (진병문 본부장, 5시 B039) ***
  Review Paper Due

- 5/7(2): 허윤진, 김용중
- 5/9: 정원섭

- 5/14(2): 이명춘, 양견모
- 5/16: 이시혁

- 5/21(2): 박성기, 정광복
- 5/23: 박광일

- 5/30: 이석준
- 6/4: 윤성재

- 6/11: 텀프로젝트 최종 발표