هوش مصنوعی پیشرفته

فصل ۱۴

استدلال احتمالاتی

Probabilistic Reasoning

کاظم فولادی
دانشکده مهندسی برق و کامپیوتر
دانشگاه تهران

http://courses.fouladi.ir/aai

استدلال احتمالاتی

۱

بازنمایی دانایی در یک دامنه‌ی نامطمئن

شبکه‌های بیزی

Bayesian Networks

شبکه‌ی بیزی
Bayesian Network

یک نمادگذاری ساده و گرافیکی برای بیان استقلال شرطی
(و در نتیجه برای مشخص‌سازی متراکم توزیع‌های توأم کامل)

شبکه‌های بیزی

نحو

BAYESIAN NETWORKS

شبکه‌ی بیزی
Bayesian Network

یک گراف جهت‌دار بدون دور

پیوندها *Links*	**گره‌ها** *Nodes*
نشان‌دهنده‌ی **رابطه‌ی تأثیر مستقیم**	نشان‌دهنده‌ی **متغیرهای تصادفی**
(link ≈ "directly influences")	**توزیع شرطی** *Conditional Distribution*
	برای هر گره، توزیع شرطی آن گره به‌شرط والدهای آن را داریم:
	$\mathbf{P}(X_i \mid Parents(X_i))$
	در قالب جدول احتمال شرطی (CPT)

Conditional Probability Table (CPT)

توزیع شرطی به طور ساده در قالب جدول احتمال شرطی بازنمایی می‌شود.

جدول احتمال شرطی
Conditional Probability Table

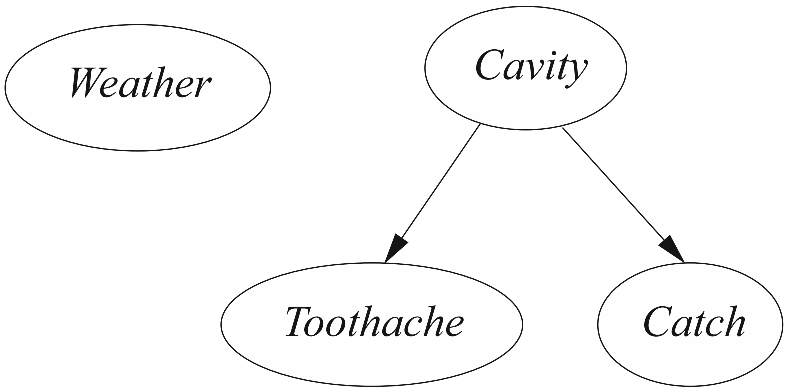
توزیع برای متغیر X_i برای هر ترکیب از مقادیر والدهای آن

CPT

شبکه‌های بیزی

مثال

BAYESIAN NETWORKS



$Weather$ is independent of the other variables

$Toothache$ and $Catch$ are conditionally independent given $Cavity$

توپولوژی شبکه‌ی بیزی بیان استقلال شرطی را کدگذاری می‌کند.

شبکه‌های بیزی

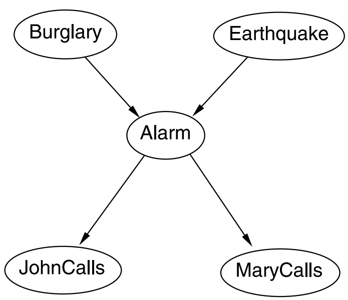
مثال: سیستم هشدار سرقت منزل

BAYESIAN NETWORKS

شما دو همسایه به نام‌های مری و جان دارید.
آن‌ها قول داده‌اند که در صورت شنیدن صدای زنگ هشدار سرقت با شما در محل کارتان تماس بگیرند.
جان همیشه وقتی تماس می‌گیرد که صدای زنگ هشدار را بشنود،
اما گاهی صدای زنگ هشدار را با صدای زنگ تلفن اشتباه می‌گیرد و با شما تماس می‌گیرد.
مری موسیقی با صدای بلند گوش می‌دهد و گاهی صدای زنگ هشدار را نمی‌شنود.
البته گاهی زمین‌لرزه‌ی خفیف هم باعث به صدا در آمدن زنگ هشدار می‌شود.
می‌خواهیم با دانستن فرد تماس‌گیرنده، احتمال **سرقت** را تخمین بزنیم.

Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$

سرقت	زمین‌لرزه	هشدار	تماس جان	تماس مری



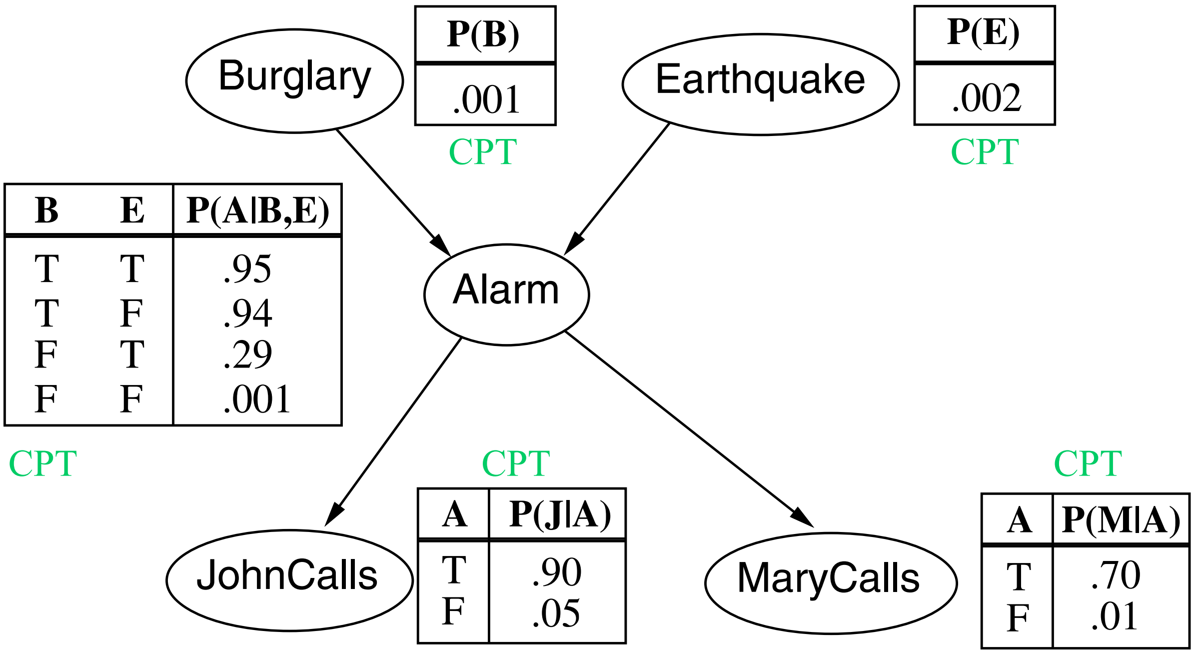
توپولوژی شبکه‌ی بیزی، دانایی «علّی» را منعکس می‌کند:

- وقوع یک سرقت می‌تواند زنگ هشدار را به صدا درآورد.
- وقوع یک زمین‌لرزه می‌تواند زنگ هشدار را به صدا درآورد.
- زنگ هشدار می‌تواند باعث شود مری تماس بگیرد.
- زنگ هشدار می‌تواند باعث شود جان تماس بگیرد.

شبکه‌های بیزی

مثال: سیستم هشدار سرقت منزل

BAYESIAN NETWORKS



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Variables: *Burglar*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

B E A J M

شبکه‌های بیزی

مفهوم «تراکم»

COMPACTNESS

یک CPT برای متغیر تصادفی بولی X با k والد بولی
دارای 2^k سطر برای ترکیب‌های مختلف مقادیر والدهاست.

هر سطر یک عدد p برای $X = true$ نیاز دارد.
(عدد $1-p$ برای $X = false$ مشخص است)

اگر هریک از n متغیر حداکثر k والد داشته باشد، کل شبکه به $O(n \cdot 2^k)$ عدد برای CPTها نیاز دارد.

$$O(2^n) \Rightarrow O(n \cdot 2^k)$$

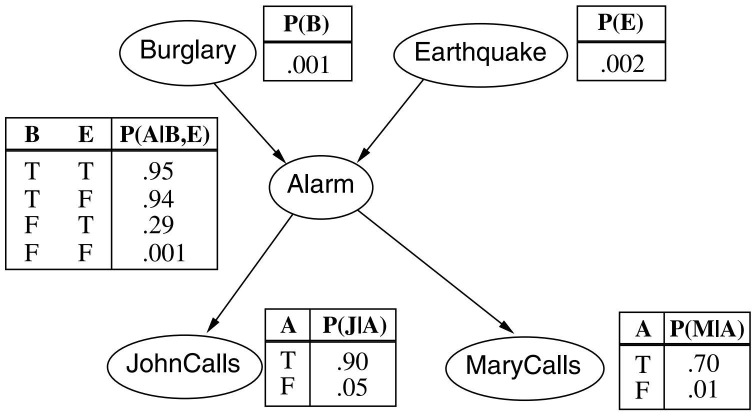
برای توزیع‌های توأم کامل (نمایی بر حسب n)	برای توزیع‌های شرطی (خطی بر حسب n)

شبکه‌های بیزی

مفهوم «تراکم»: مثال

COMPACTNESS

برای مثال سیستم هشدار سرقت



$$O(2^n)$$

برای توزیع‌های توأم کامل
(نمایی بر حسب n)

$$2^5 = 32$$

$$O(n \cdot 2^k)$$

برای توزیع‌های شرطی
(خطی بر حسب n)

$$1+1+4+2+2=10$$

هوش مصنوعی

استدلال احتمالاتی

۲

معناشناسی شبکه‌های بیزی

معناشناسی شبکه‌های بیزی

THE SEMANTICS OF BAYESIAN NETWORKS

معناشناسی شبکه‌های بیزی	
معناشناسی سراسری *Global Semantics*	معناشناسی محلی *Local Semantics*

معناشناسی شبکه‌های بیزی

معناشناسی سراسری

THE SEMANTICS OF BAYESIAN NETWORKS

معناشناسی شبکه‌های بیزی	
معناشناسی سراسری *Global Semantics*	معناشناسی محلی *Local Semantics*
توزیع توأم کامل = حاصل‌ضرب توزیع‌های شرطی محلی	

$$P(x_1,\ldots,x_n) = \Pi_{i=1}^{n} P(x_i|parents(X_i))$$

معناشناسی شبکه‌های بیزی

معناشناسی سراسری: مثال

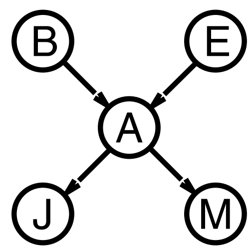
THE SEMANTICS OF BAYESIAN NETWORKS

معناشناسی شبکه‌های بیزی	
معناشناسی سراسری *Global Semantics*	معناشناسی محلی *Local Semantics*
توزیع توأم کامل = حاصل‌ضرب توزیع‌های شرطی محلی	

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

e.g., $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$$
$$\approx 0.00063$$



معناشناسی شبکه‌های بیزی

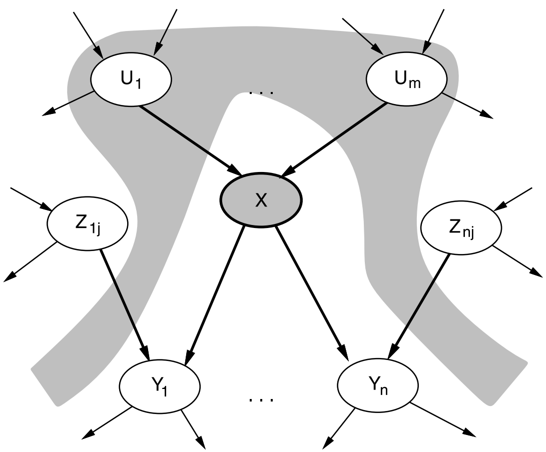
معناشناسی محلی

THE SEMANTICS OF BAYESIAN NETWORKS

معناشناسی شبکه‌های بیزی

معناشناسی محلی *Local Semantics*	معناشناسی سراسری *Global Semantics*

هر گره با داشتن والدهایش از گره‌های غیرنواده‌اش مستقل شرطی است.



معناشناسی شبکه‌های بیزی

قضیه‌ی هم‌ارزی معناشناسی سراسری با معناشناسی محلی

The Semantics of Bayesian Networks

معناشناسی شبکه‌های بیزی	
معناشناسی سراسری *Global Semantics*	معناشناسی محلی *Local Semantics*

Theorem: Local semantics $\Leftrightarrow$ global semantics

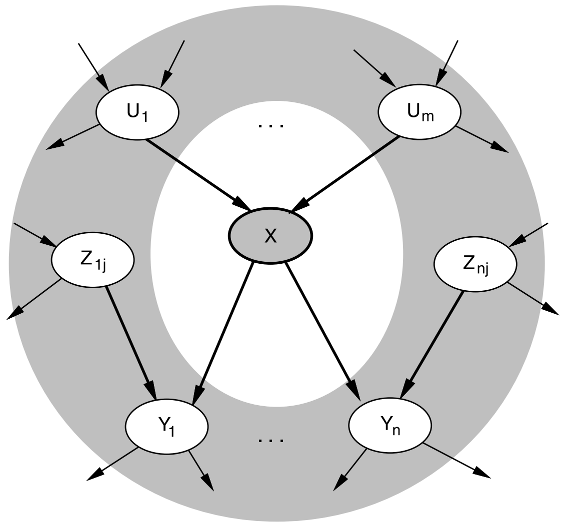
معناشناسی شبکه‌های بیزی

پتوی مارکوف

MARKOV BLANKET

هر گره از سایر گره‌ها مستقل شرطی است به‌شرط داشتن پتوی مارکوف آن گره

پتوی مارکوف *Markov Blanket*	پتوی مارکوف برای هر گره = والدهای آن + فرزندان آن + والدهای فرزندان آن



ساختن شبکه‌های بیزی

Constructing Bayesian Networks

به روشی نیاز داریم که
با بررسی یک سری از بیان‌های آزمون‌پذیر به طور محلی در مورد استقلال شرطی
بتواند معناشناسی سراسری مورد نیاز را تضمین کند.

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to n
   add X_i to the network
   select parents from $X_1, \ldots, X_{i-1}$ such that
   $\mathbf{P}(X_i | Parents(X_i)) = \mathbf{P}(X_i | X_1, \ldots, X_{i-1})$

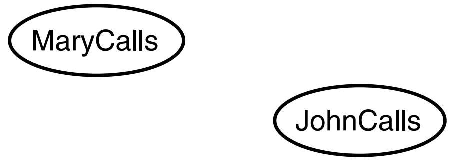
انتخاب والدها، معناشناسی سراسری را تضمین می‌کند:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i | X_1, \ldots, X_{i-1}) \quad \text{(chain rule)} \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i)) \quad \text{(by construction)}\end{aligned}$$

ساختن شبکه‌های بیزی

مثال (۱ از ۶)

Suppose we choose the ordering M, J, A, B, E

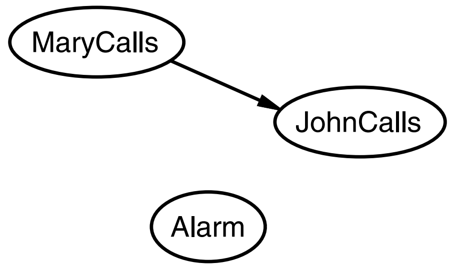


$P(J|M) = P(J)$?

ساختن شبکه‌های بیزی

مثال (۲ از ۶)

Suppose we choose the ordering M, J, A, B, E

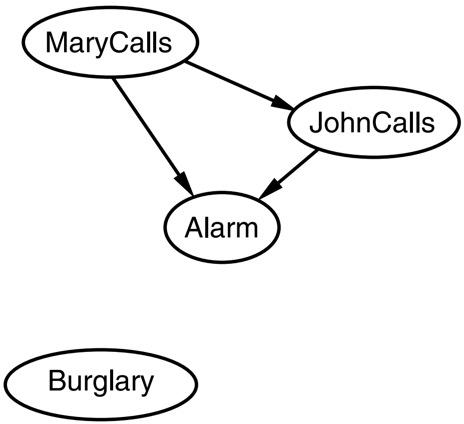


$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$?

ساختن شبکه‌های بیزی

مثال (۳ از ۶)

Suppose we choose the ordering M, J, A, B, E

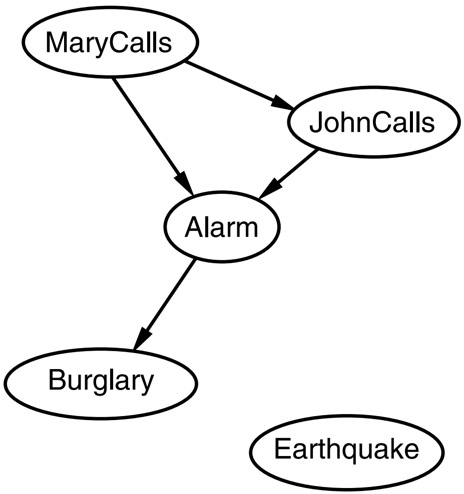


$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No
$P(B|A,J,M) = P(B|A)$?
$P(B|A,J,M) = P(B)$?

ساختن شبکه‌های بیزی

مثال (۴ از ۶)

Suppose we choose the ordering M, J, A, B, E

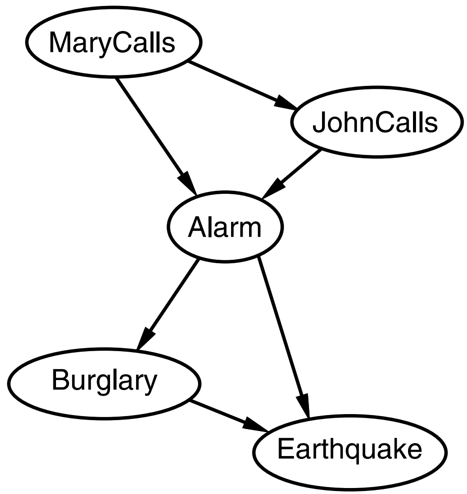


$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No
$P(B|A,J,M) = P(B|A)$? Yes
$P(B|A,J,M) = P(B)$? No
$P(E|B,A,J,M) = P(E|A)$?
$P(E|B,A,J,M) = P(E|A,B)$?

ساختن شبکه‌های بیزی

مثال (۵ از ۶)

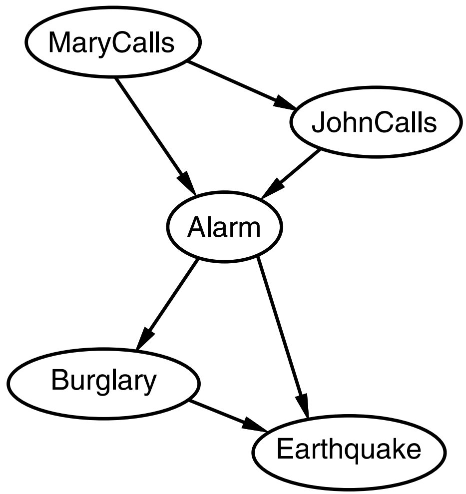
Suppose we choose the ordering M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No
$P(B|A,J,M) = P(B|A)$? Yes
$P(B|A,J,M) = P(B)$? No
$P(E|B,A,J,M) = P(E|A)$? No
$P(E|B,A,J,M) = P(E|A,B)$? Yes

ساختن شبکه‌های بیزی

مثال (۶ از ۶)



تصمیم‌گیری در مورد استقلال شرطی در جهت‌های غیرعلّی دشوار است.

(به نظر می‌رسد مدل‌های علّی و استقلال شرطی برای انسان‌ها سیم‌بندی سخت شده است)

سنجش احتمالات شرطی در جهت‌های غیرعلّی دشوار است.

❋ در این مثال تراکم شبکه کم است: فقط $1 + 2 + 4 + 2 + 4 = 13$ عدد لازم داریم.

ساخت شبکه‌های بیزی

مثال: تشخیص عیب خودرو

Example: Car Diagnosis

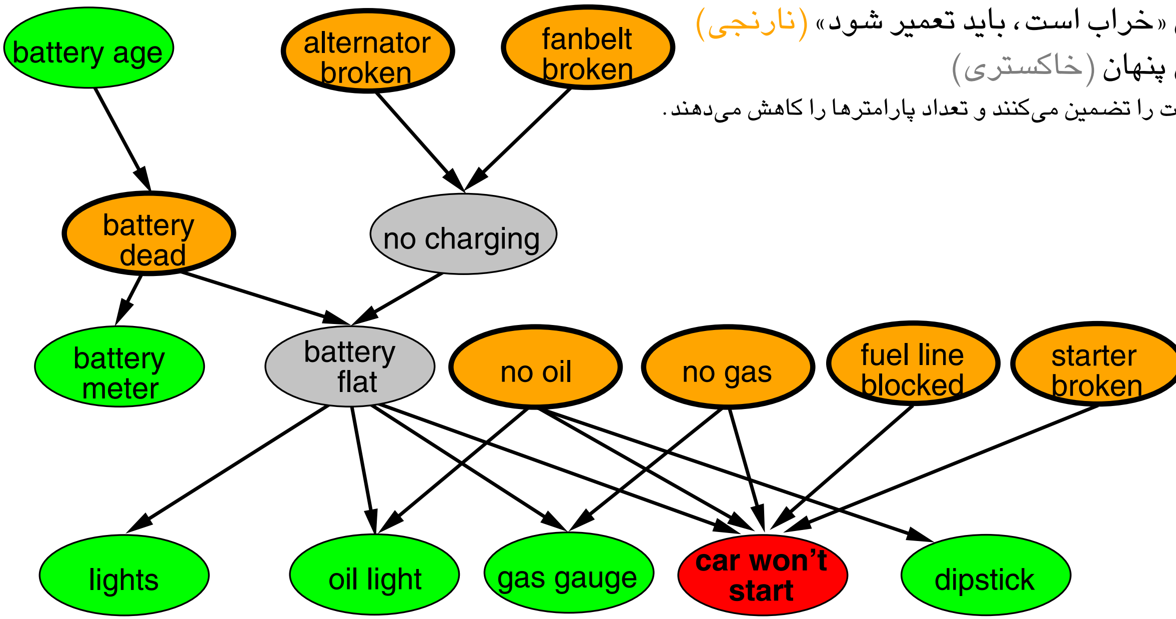
شاهد آغازین: خودرو روش نمی‌شود

متغیرهای آزمون‌پذیر (سبز)

متغیرهای «خراب است، باید تعمیر شود» (نارنجی)

متغیرهای پنهان (خاکستری)

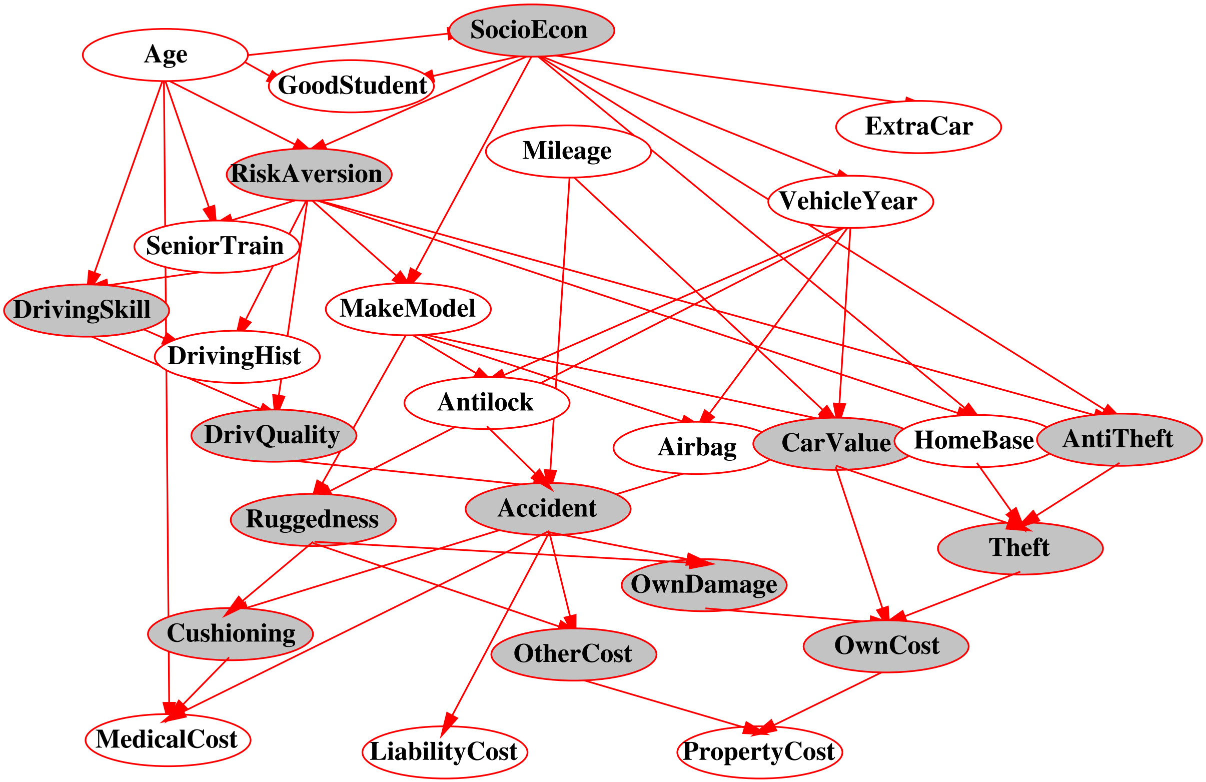
ساختار خلوت را تضمین می‌کنند و تعداد پارامترها را کاهش می‌دهند.



ساخت شبکه‌های بیزی

مثال: بیمه‌ی خودرو

Example: Car Insurance



استدلال احتمالاتی

۳

بازنمایی کارآمد توزیع‌های شرطی

توزیع‌های شرطی متراکم

Compact Conditional Distributions

مشکل

اندازه‌ی CPT با تعداد والدها به صورت نمایی رشد می‌کند.

حل مشکل

برای گره‌های قطعی **(deterministic nodes)**

استفاده از روابط قطعی بین گره‌ها

برای روابط نامطمئن **(uncertain relationships)**

استفاده از روابط منطقی «نویزی» مانند مدل **noisy-OR**

مشکل

اندازه‌ی CPT با والدها یا فرزندان پیوسته–مقدار، بی‌نهایت می‌شود.

حل مشکل

متغیرهای پیوسته، والدهای گسسته + پیوسته

استفاده از یک تابع چگالی شرطی

متغیرهای گسسته، والدهای پیوسته

استفاده از یک تابع چگالی شرطی

توزیع‌های شرطی متراکم

برای گره‌های قطعی

Compact Conditional Distributions

مشکل

اندازه‌ی CPT با تعداد والدها به صورت نمایی رشد می‌کند.

حل مشکل

برای گره‌های قطعی (deterministic nodes)

استفاده از روابط قطعی بین گره‌ها

برای روابط نامطمئن (uncertain relationships)

استفاده از روابط منطقی «نویزی» مانند مدل noisy-OR

اندازه‌ی CPT با والدها یا فرزندان پیوسته-مقدار، بی‌نهایت می‌شود

حل مشکل

متغیرهای پیوسته، والدهای گسسته + پیوسته

استفاده از یک تابع چگالی شرطی

متغیرهای گسسته، والدهای پیوسته

استفاده از یک تابع چگالی شرطی

توزیع‌های شرطی متراکم

برای گره‌های قطعی

DETERMINISTIC NODES

مقدار گره‌ی قطعی از روی مقدار والدهای آن به‌صورت قطعی مشخص می‌شود.

$$X = f(Parents(X)) \text{ for some function } f$$

برای مثال: توابع بولی:

$$NorthAmerican \Leftrightarrow Canadian \vee US \vee Mexican$$

برای مثال: روابط عددی بین متغیرهای پیوسته

$$\frac{\partial Level}{\partial t} = \text{inflow + precipitation - outflow - evaporation}$$

توزیع‌های شرطی متراکم

برای روابط نامطمئن

COMPACT CONDITIONAL DISTRIBUTIONS

مشکل

اندازه‌ی CPT با تعداد والدها به صورت نمایی رشد می‌کند.

حل مشکل

برای گره‌های قطعی (deterministic nodes)

استفاده از روابط قطعی بین گره‌ها

برای روابط نامطمئن **(uncertain relationships)**

استفاده از روابط منطقی «نویزی» مانند مدل **noisy-OR**

اندازه‌ی CPT با والدها یا فرزندان پیوسته-مقدار، بی‌نهایت می‌شود

حل مشکل

متغیرهای پیوسته، والدهای گسسته + پیوسته

استفاده از یک تابع چگالی شرطی

متغیرهای گسسته، والدهای پیوسته

استفاده از یک تابع چگالی شرطی

توزیع‌های شرطی متراکم

برای رابطه‌های نامطمئن

UNCERTAIN RELATIONSHIPS

رابطه‌های غیرقطعی را می‌توان
با استفاده از روابط منطقی «نویزی» مانند مدل **noisy-OR** مشخص کرد.

$$Child \Leftarrow Parent_1 \vee Parent_2 \vee \cdots \vee Parent_k$$

در اینکه والدها می‌توانند موجب درست شدن فرزند شوند، عدم اطمینان مجاز شمرده می‌شود.

توزیع‌های **noisy-OR** علت‌های غیرمتعامل چندگانه را مدل می‌کنند؛ با دو شرط

۱ والدها شامل همه‌ی علت‌ها باشند: $U_1, U_2, \ldots, U_k$

همیشه می‌توان سایر علت‌ها را با اضافه کردن یک **گره‌ی نشتی** **(leak node)** پوشش داد.

۲ هر علت به‌تنهایی دارای احتمال شکست مستقل q_i باشد (مستقل از سایر والدها).

q_i : **احتمال ممانعت فردی (individual inhibition probabilities)**

$$q_i = P(\neg Child \mid \neg Parent_1, \neg Parent_2, \cdots, Parent_i, \cdots, \neg Parent_k)$$

در این صورت داریم:

$$P(X|U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = 1 - \Pi_{i=1}^{j} q_i$$

توزیع‌های شرطی متراکم

برای رابطه‌های نامطمئن: مثال

UNCERTAIN RELATIONSHIPS

در منطق گزاره‌ای می‌توان گفت $Fever \Leftrightarrow Cold \vee Flu \vee Malaria$

$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$	
F	F	F	**0.0**	1.0	
F	F	T	0.9	**0.1**	داده شده
F	T	F	0.8	**0.2**	داده شده
F	T	T	0.98	$0.02 = 0.2 \times 0.1$	
T	F	F	0.4	**0.6**	داده شده
T	F	T	0.94	$0.06 = 0.6 \times 0.1$	
T	T	F	0.88	$0.12 = 0.6 \times 0.2$	
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$	

$$q_{\text{cold}} = P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6 \,,$$
$$q_{\text{flu}} = P(\neg fever \mid \neg cold, flu, \neg malaria) = 0.2 \,,$$
$$q_{\text{malaria}} = P(\neg fever \mid \neg cold, \neg flu, malaria) = 0.1 \,.$$

بر اساس احتمالات ممانعت فردی،
می‌توان کل CPT را کامل کرد.

$$P(x_i \mid parents(X_i)) = 1 - \prod_{\{j : X_j = true\}} q_j$$

تعداد پارامترهای مورد نیاز بر حسب تعداد والدها، **خطی** است.

$$O(2^k) \rightsquigarrow O(k)$$

توزیع‌های شرطی متراکم

برای متغیرهای پیوسته، والدهای گسسته + پیوسته

Compact Conditional Distributions

مشکل

اندازه‌ی CPT با تعداد والدها به صورت نمایی رشد می‌کند

حل مشکل

برای گره‌های قطعی (deterministic nodes)

استفاده از روابط قطعی بین گره‌ها

برای روابط نامطمئن (uncertain relationships)

استفاده از روابط منطقی «مزهزی» مانند مدل noisy-OR

اندازه‌ی CPT با والدها یا فرزندان پیوسته-مقدار، بی‌نهایت می‌شود.

حل مشکل

متغیرهای پیوسته، والدهای گسسته + پیوسته

استفاده از یک تابع چگالی شرطی

متغیرهای گسسته، والدهای پیوسته

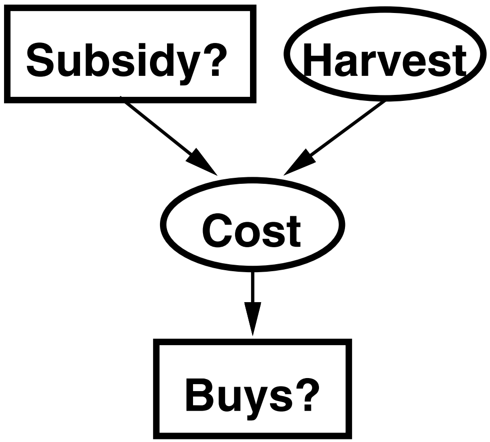
استفاده از یک تابع چگالی شرطی

توزیع‌های شرطی متراکم

برای شبکه‌های بیزی هیبرید (گسسته + پیوسته)

HYBRID (DISCRETE+CONTINUOUS) NETWORKS

Discrete (*Subsidy?* and *Buys?*); continuous (*Harvest* and *Cost*)



چگونگی برخورد با متغیرهای پیوسته	
	گزینه ۱) استفاده از گسسته‌سازی: ❋ مشکل خطای احتمالی بالا ❋ مشکل CPTهای بزرگ
	گزینه ۲) استفاده از خانواده‌های کانونیک توزیع‌های پارامتری متناهی

توزیع‌های شرطی متراکم

برای شبکه‌های بیزی هیبرید (گسسته + پیوسته): برای متغیرهای پیوسته فرزند

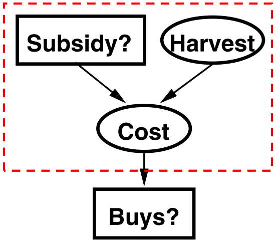
HYBRID (DISCRETE+CONTINUOUS) NETWORKS

نیاز داریم به یک تابع چگالی شرطی

برای متغیر فرزند پیوسته با داشتن والدهای پیوسته به‌ازای هر انتساب ممکن برای والدهای گسسته

معمولاً مدل **گاوسی خطی** برای تابع چگالی شرطی متداول‌ترین گزینه است، مثلاً:

$$P(Cost = c | Harvest = h, Subsidy? = true)$$
$$= N(a_t h + b_t, \sigma_t)(c)$$
$$= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{c-(a_t h + b_t)}{\sigma_t}\right)^2\right)$$



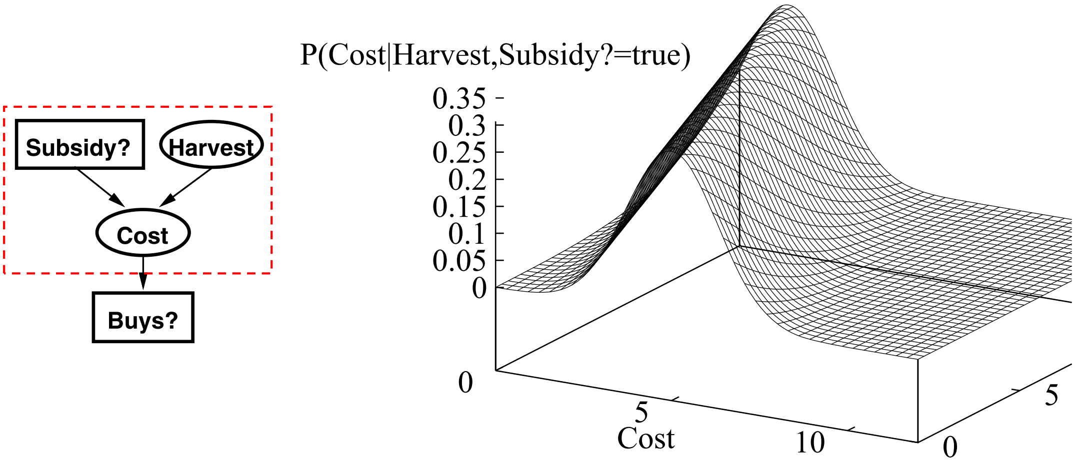
متوسط **فرزند** $(Cost)$ به صورت خطی با **والد** $(Harvest)$ تغییر می‌کند؛ واریانس ثابت است.

تغییر خطی بر روی یک بازه‌ی بزرگ غیرمنطقی است،
اما به‌خوبی کار می‌کند اگر بازه‌ی احتمالی $Harvest$ باریک باشد.

توزیع‌های شرطی متراکم

برای شبکه‌های بیزی هیبرید (گسسته + پیوسته): برای متغیرهای پیوسته فرزند: توزیع گاوسی خطی

HYBRID (DISCRETE+CONTINUOUS) NETWORKS

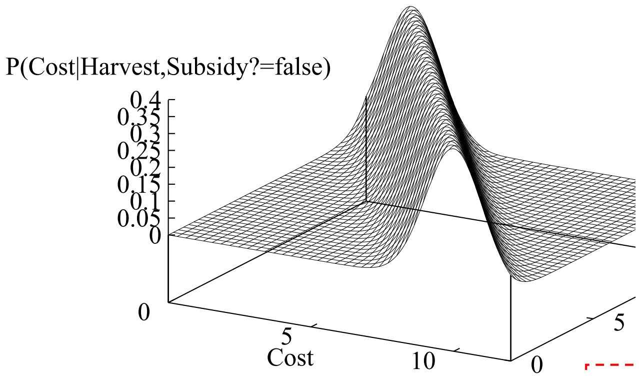


یک شبکه که تنها از متغیرهای پیوسته با توزیع گاوسی خطی تشکیل شده است
دارای یک توزیع توأم کامل با گاوسی چندمتغیره است.

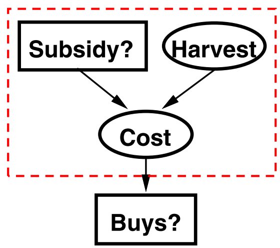
توزیع‌های شرطی متراکم

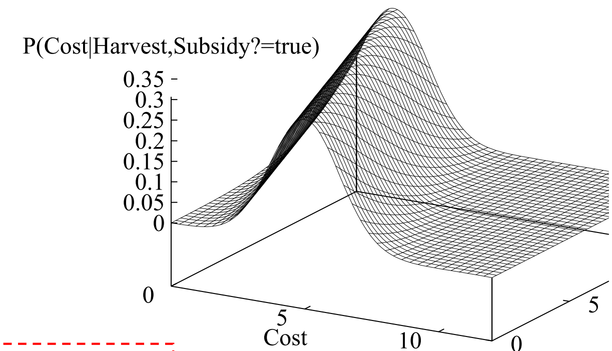
برای شبکه‌های بیزی هیبرید (گسسته + پیوسته): برای متغیرهای پیوسته فرزند: توزیع گاوسی خطی

Hybrid (discrete+continuous) Networks



a_f, b_f, σ_f





a_t, b_t, σ_t

شبکه‌ی گاوسی خطی گسسته + پیوسته، یک شبکه‌ی گاوسی شرطی است.

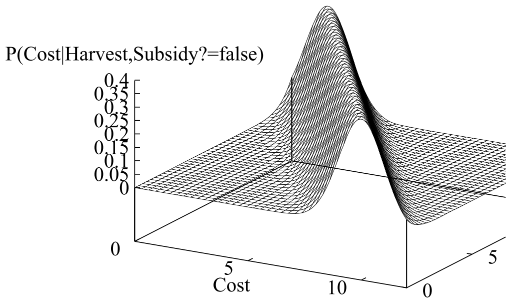
یعنی: یک گاوسی چندمتغیره بر روی همه‌ی متغیرهای پیوسته

برای هر ترکیب از مقادیر متغیرهای گسسته وجود دارد.

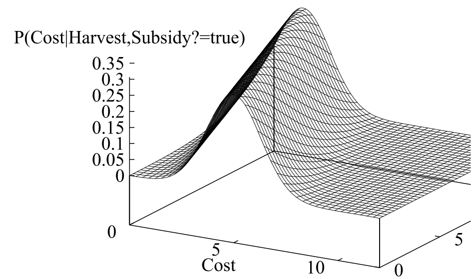
توزیع‌های شرطی متراکم

برای شبکه‌های بیزی هیبرید (گسسته + پیوسته): برای متغیرهای پیوسته فرزند: توزیع گاوسی خطی

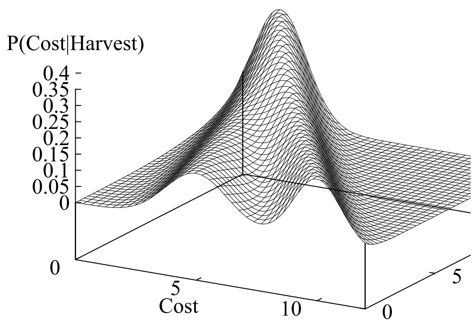
Hybrid (discrete+continuous) Networks



+



=



مثال با فرض احتمال پیشین 0.5

حاصل با متوسط گیری روی دو مقدار ممکن متغیر بولی

مجموع دو توزیع گاوسی، یک گاوسی است.

توزیع‌های شرطی متراکم

برای متغیرهای گسسته، والدهای پیوسته

COMPACT CONDITIONAL DISTRIBUTIONS

مشکل

اندازه‌ی CPT با تعداد والدها به صورت نمایی رشد می‌کند.

حل مشکل

برای گره‌های قطعی (deterministic nodes)

استفاده از روابط قطعی بین گره‌ها

برای روابط نامطمئن (uncertain relationships)

استفاده از روابط منطقی «نویزی» مانند مدل noisy-OR

اندازه‌ی CPT با والدها یا فرزندان پیوسته-مقدار، بی‌نهایت می‌شود.

حل مشکل

متغیرهای پیوسته، والدهای گسسته + پیوسته

استفاده از یک تابع چگالی شرطی

متغیرهای گسسته، والدهای پیوسته

استفاده از یک تابع چگالی شرطی

Prepared by Kazim Fouladi | Spring 2017 | 2nd Edition

توزیع‌های شرطی متراکم

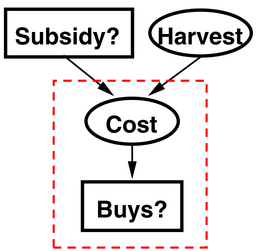
برای شبکه‌های بیزی هیبرید (گسسته + پیوسته): برای متغیرهای گسسته فرزند و والدهای پیوسته

Hybrid (discrete+continuous) Networks

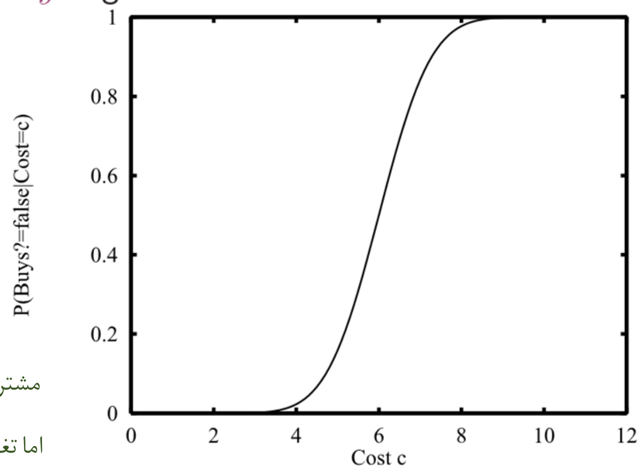
احتمال **فرزند گسسته** به شرط **والد پیوسته** باید یک مقدار آستانه‌ای نرم باشد (مثل توزیع پروبیت).

(نباید تغییر ناگهانی داشته باشد)

Probability of $Buys?$ given $Cost$ should be a "soft" threshold:



مشتری در صورتی خرید می‌کند که قیمت پایین باشد و
در صورت بالا بودن قیمت خرید نخواهد کرد؛
اما تغییرات احتمال خرید در ناحیه‌ی میانی ملایم است.



توزیع **پروبیت (Probit)** از انتگرال تابع توزیع گاوسی استفاده می‌کند:

$$\Phi(x) = \int_{-\infty}^{x} N(0,1)(x)dx$$

$$P(Buys? = true \mid Cost = c) = \Phi((-c+\mu)/\sigma)$$

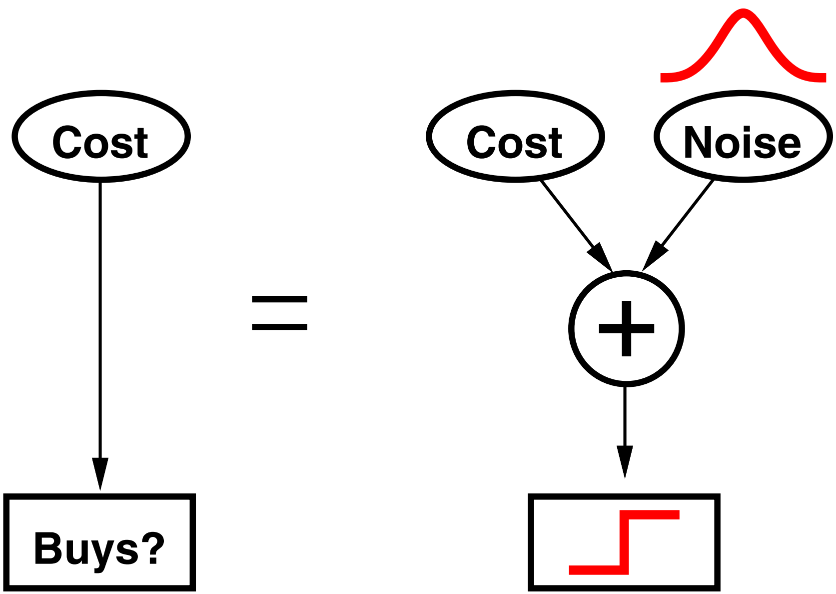
توزیع‌های شرطی متراکم

برای شبکه‌های بیزی هیبرید (گسسته + پیوسته): برای متغیرهای گسسته فرزند و والدهای پیوسته: توزیع پروبیت

HYBRID (DISCRETE+CONTINUOUS) NETWORKS

ویژگی توزیع پروبیت (Probit)

فرآیند تصمیم‌گیری دارای مقدار آستانه‌ی سخت است، اما مکان دقیق آستانه در معرض نویز گاوسی تصادفی قرار دارد.



توزیع‌های شرطی متراکم

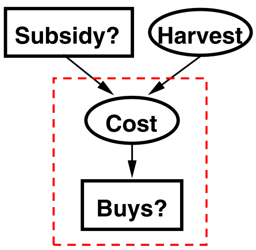
برای شبکه‌های بیزی هیبرید (گسسته + پیوسته): برای متغیرهای گسسته فرزند و والدهای پیوسته: توزیع لوگیت

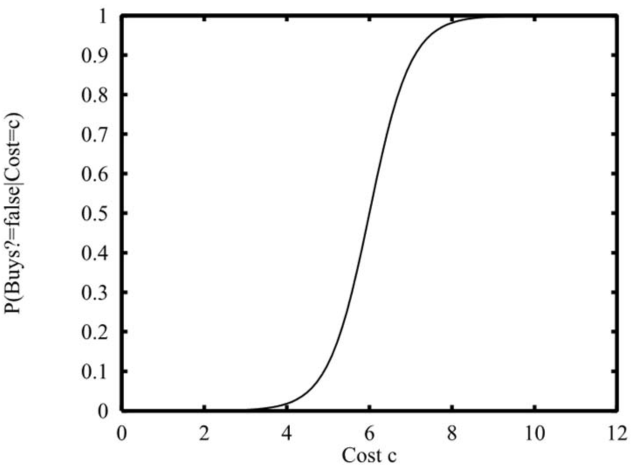
Hybrid (discrete+continuous) Networks

ویژگی توزیع لوگیت (Logit) (یا سیگموئید (Sigmoid))

مشابه پروبیت (آستانه‌ی نرم) اما دُم‌های آن طولانی‌تر است.

$$P(Buys? = true \mid Cost = c) = \frac{1}{1 + exp(-2\frac{-c+\mu}{\sigma})}$$





استدلال احتمالاتی

۴

استنتاج دقیق در شبکه‌های بیزی

وظیفه‌های استنتاج

INFERENCE TASKS

پرسش‌های ساده
Simple Queries

محاسبه‌ی توزیع‌های پسین حاشیه‌ای $\mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$

$$P(NoGas|Gauge=empty, Lights=on, Starts=false)$$

پرسش‌های عطفی
Conjunctive Queries

محاسبه‌ی توزیع‌های پسین حاشیه‌ای شامل چند متغیر

$$\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$$

تصمیم‌های بهینه
Optimal Decisions

شبکه‌های تصمیم حاوی اطلاعات سودمندی؛ استنتاج احتمالاتی لازم برای:

$$P(outcome|action, evidence)$$

ارزش اطلاعات
Value of Information

به دنبال چه شاهدی برای بعد باید باشیم؟

تحلیل حساسیت
Sensitivity Analysis

کدام مقادیر احتمال، حیاتی‌ترین آنها هستند؟

تبیین
Explaination

مثلاً: چرا من به یک استارتر موتور جدید نیاز دارم؟

استنتاج در شبکه‌های بیزی

استنتاج در شبکه‌های بیزی			
استنتاج دقیق *Exact Inference*		استنتاج تقریبی *Approximate Inference*	
برشمارش *Enumeration*	حذف متغیر *Variable Elimination*	شبیه‌سازی اتفاقی *Stochastic Simulation*	زنجیره‌ی مارکوف مونت کارلو *Markov Chain Monte Carlo*

استنتاج در شبکه‌های بیزی

استنتاج دقیق با برشماری

استنتاج در شبکه‌های بیزی			
استنتاج دقیق *Exact Inference*		استنتاج تقریبی *Approximate Inference*	
برشمارش *Enumeration*	حذف متغیر *Variable Elimination*	شبیه‌سازی اتفاقی *Stochastic Simulation*	زنجیره‌ی مارکوف مونت کارلو *Markov Chain Monte Carlo*

استنتاج با برشماری

Inference by Enumeration

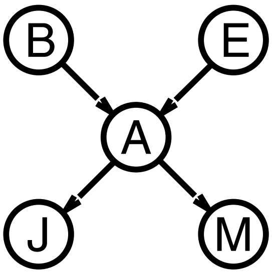
روشی با هوشمندی اندک برای مجموع‌گیری متغیرها از توزیع توأم، بدون ساخت واقعی بازنمایی صریح آن

$$\underbrace{P(X \mid e)}_{\text{پرس‌وجو}} = \alpha P(X, e) = \alpha \sum_{y} P(X, e, y)$$

استنتاج با برشماری

مثال

Inference by Enumeration



مثال

Simple query on the burglary network:

$\mathbf{P}(B|j,m)$
$= \mathbf{P}(B,j,m)/P(j,m)$
$= \alpha\mathbf{P}(B,j,m)$
$= \alpha \sum_e \sum_a \mathbf{P}(B,e,a,j,m)$

Rewrite full joint entries using product of CPT entries:

$\mathbf{P}(B|j,m)$
$= \alpha \sum_e \sum_a \mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a)$
$= \alpha\mathbf{P}(B) \sum_e P(e) \sum_a \mathbf{P}(a|B,e)P(j|a)P(m|a)$

استنتاج با برشماری

الگوریتم

Inference by Enumeration

```
function ENUMERATION-ASK(X, e, bn) returns a distribution over X
   inputs: X, the query variable
           e, observed values for variables E
           bn, a Bayesian network with variables {X} ∪ E ∪ Y

   Q(X) ← a distribution over X, initially empty
   for each value x_i of X do
       extend e with value x_i for X
       Q(x_i) ← ENUMERATE-ALL(VARS[bn], e)
   return NORMALIZE(Q(X))
```

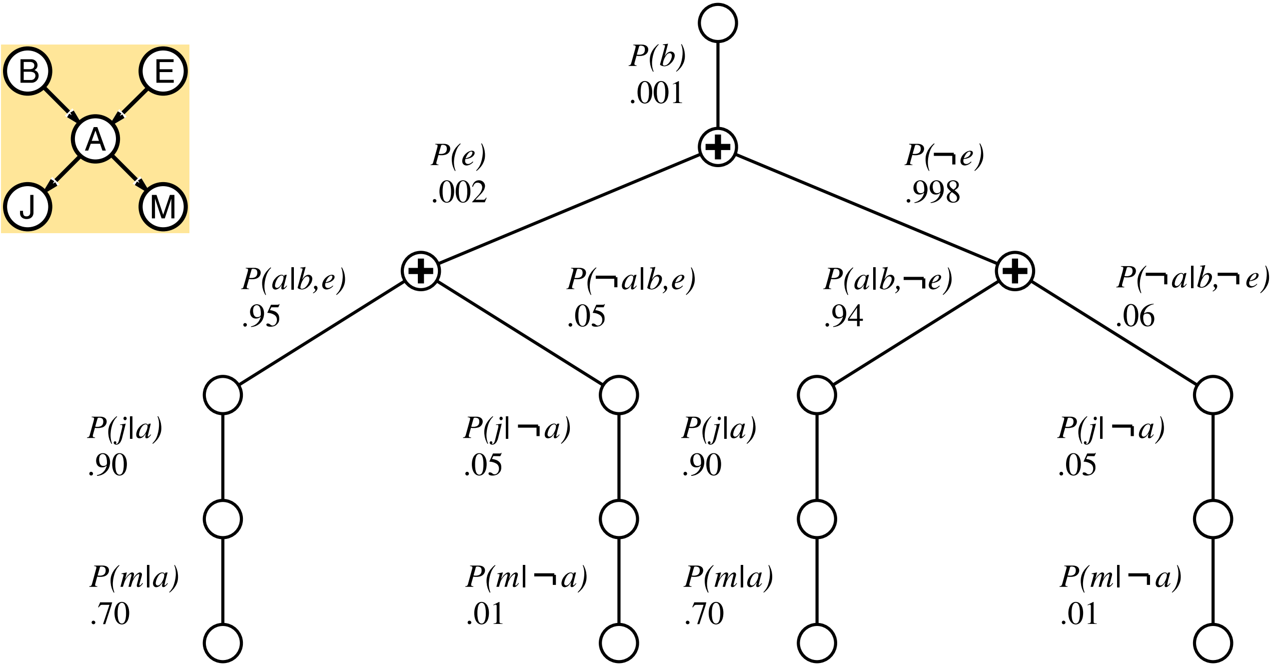
```
function ENUMERATE-ALL(vars, e) returns a real number
   if EMPTY?(vars) then return 1.0
   Y ← FIRST(vars)
   if Y has value y in e
       then return P(y | Pa(Y)) × ENUMERATE-ALL(REST(vars), e)
       else return Σ_y P(y | Pa(Y)) × ENUMERATE-ALL(REST(vars), e_y)
           where e_y is e extended with Y = y
```

Recursive depth-first enumeration: $O(n)$ space, $O(d^n)$ time

استنتاج با برشماری

درخت ارزیابی

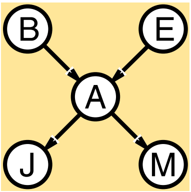
Evaluation Tree

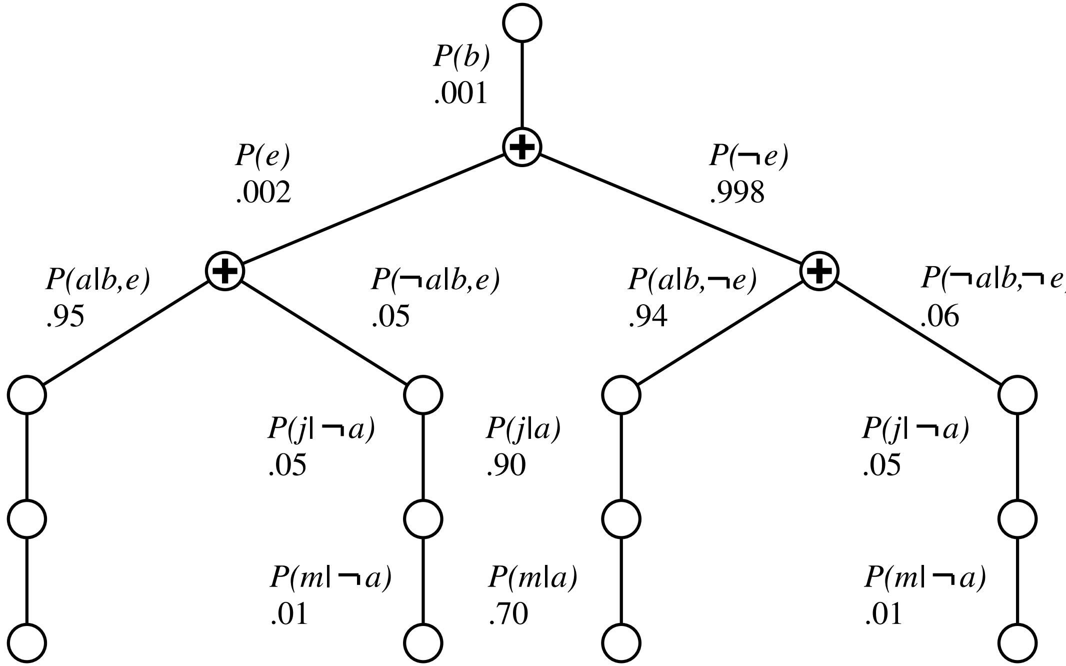


$$\mathbf{P}(B|j,m)$$
$$= \alpha \textstyle\sum_e \sum_a \mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a)$$
$$\boxed{= \alpha\mathbf{P}(B) \textstyle\sum_e P(e) \sum_a \mathbf{P}(a|B,e)P(j|a)P(m|a)}$$

استنتاج با برشماری

مشکل ناکارآمدی





برشماری **ناکارآمد** است: به دلیل محاسبات تکراری

برای مثال: $P(j|a)P(m|a)$ برای هر مقدار e محاسبه می‌شود!

استنتاج در شبکه‌های بیزی

استنتاج دقیق با حذف متغیر

استنتاج در شبکه‌های بیزی			
استنتاج دقیق *Exact Inference*		استنتاج تقریبی *Approximate Inference*	
برشمارش *Enumeration*	**حذف متغیر** *Variable Elimination*	شبیه‌سازی اتفاقی *Stochastic Simulation*	زنجیره‌ی مارکوف مونت کارلو *Markov Chain Monte Carlo*

استنتاج با حذف متغیر

Inference by Variable Elimination

حذف متغیر:

انجام مجموع‌یابی‌ها از سمت راست به چپ،

نتایج میانی (فاکتورها) ذخیره می‌شوند تا از محاسبه‌ی مجدد آنها اجتناب شود.

مثال

$$
\begin{aligned}
&\mathbf{P}(B|j,m) \\
&\quad = \alpha \underbrace{\mathbf{P}(B)}_{B} \textstyle\sum_e \underbrace{P(e)}_{E} \sum_a \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M} \\
&\quad = \alpha \mathbf{P}(B) \textstyle\sum_e P(e) \sum_a \mathbf{P}(a|B,e) P(j|a) f_M(a) \\
&\quad = \alpha \mathbf{P}(B) \textstyle\sum_e P(e) \sum_a \mathbf{P}(a|B,e) f_J(a) f_M(a) \\
&\quad = \alpha \mathbf{P}(B) \textstyle\sum_e P(e) \sum_a f_A(a,b,e) f_J(a) f_M(a) \\
&\quad = \alpha \mathbf{P}(B) \textstyle\sum_e P(e) f_{\bar{A}JM}(b,e) \text{ (sum out } A) \\
&\quad = \alpha \mathbf{P}(B) f_{\bar{E}\bar{A}JM}(b) \text{ (sum out } E) \\
&\quad = \alpha f_B(b) \times f_{\bar{E}\bar{A}JM}(b)
\end{aligned}
$$

استنتاج با حذف متغیر

فاکتور

INFERENCE BY VARIABLE ELIMINATION

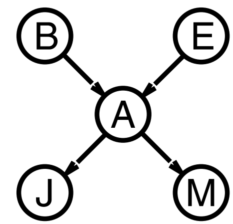
$$f_M(A) = \begin{bmatrix} P(m|a) \\ P(m|\neg a) \end{bmatrix}$$

وابستگی فاکتور به M

استنتاج با حذف متغیر

فاکتورها

INFERENCE BY VARIABLE ELIMINATION



فاکتورهای آغازین، CPTهای محلی هستند:

$$\underbrace{P(B)}_{f_B(B)} \qquad \underbrace{P(J|A)}_{f_J(A,J)} \qquad \underbrace{P(A|B,E)}_{f_A(A,B,E)}$$

در طول حذف، فاکتورهای جدیدی ساخته می‌شوند.

4 numbers, one for each value of D and E

$$f_{A\bar{B}\bar{C}D}(D,E)$$

متغیرهای وارد شده Variables introduced

متغیرهای جمع بسته شده Variables summed out

آناتومی یک فاکتور

Argument variables, always non-evidence variables

متغیرهای آرگومان: همیشه غیر از شواهد

استنتاج با حذف متغیر

عملیات پایه: الحاق فاکتورها (ضرب نقطه به نقطه)

Join Factors

ترکیب دو فاکتور

(مشابه عمل Join در پایگاه داده‌ی رابطه‌ای):

ایجاد یک فاکتور بر روی اجتماع دامنه‌ها

$$f_1(A,B) \times f_2(B,C) \Longrightarrow f_3(A,B,C)$$

$$\boxed{f_3(a,b,c) = f_1(a,b) \cdot f_2(b,c)}$$

$$\text{“}P(a,b|c) = P(a|b) \cdot P(b|c)\text{”}$$

استنتاج با حذف متغیر

عملیات پایه: الحاق فاکتورها (ضرب نقطه به نقطه): مثال

Join Factors

A	B	$\mathbf{f}_1(A,B)$	B	C	$\mathbf{f}_2(B,C)$	A	B	C	$\mathbf{f}_3(A,B,C)$
T	T	.3	T	T	.2	T	T	T	$.3 \times .2 = .06$
T	F	.7	T	F	.8	T	T	F	$.3 \times .8 = .24$
F	T	.9	F	T	.6	T	F	T	$.7 \times .6 = .42$
F	F	.1	F	F	.4	T	F	F	$.7 \times .4 = .28$
						F	T	T	$.9 \times .2 = .18$
						F	T	F	$.9 \times .8 = .72$
						F	F	T	$.1 \times .6 = .06$
						F	F	F	$.1 \times .4 = .04$

Figure 14.10 Illustrating pointwise multiplication: $\mathbf{f}_1(A,B) \times \mathbf{f}_2(B,C) = \mathbf{f}_3(A,B,C)$.

استنتاج با حذف متغیر

عملیات پایه: حاشیه‌ای‌سازی

MARGINALIZATION

گرفتن یک فاکتور و مجموع‌گیری روی یک متغیر
(مشابه عمل Projection در پایگاه داده‌ی رابطه‌ای):
تبدیل یک فاکتور به یک فاکتور کوچک‌تر

$$\boxed{f_{\bar{A}B}(b) = \sum_a f_{AB}(a,b)}$$

$$\text{``}P(b) = \sum_a P(a,b)\text{''}$$

استنتاج با حذف متغیر

عملیات پایه: حاشیه‌ای‌سازی: مثال

MARGINALIZATION

$$\begin{aligned}\mathbf{f}(B,C) &= \sum_a \mathbf{f}_3(A,B,C) = \mathbf{f}_3(a,B,C) + \mathbf{f}_3(\neg a,B,C) \\ &= \begin{pmatrix} .06 & .24 \\ .42 & .28 \end{pmatrix} + \begin{pmatrix} .18 & .72 \\ .06 & .04 \end{pmatrix} = \begin{pmatrix} .24 & .96 \\ .48 & .32 \end{pmatrix}.\end{aligned}$$

استنتاج با حذف متغیر

عملیات پایه

INFERENCE BY VARIABLE ELIMINATION

مجموع‌یابی یک متغیر از ضرب یک‌سری از فاکتورها:

- هر فاکتور ثابت را به خارج مجموع انتقال می‌دهیم.
- برای باقیمانده‌ی فاکتورها:

زیرماتریس‌های آنها را نقطه به نقطه (pointwise) ضرب می‌کنیم و حاصل آنها را با هم جمع می‌کنیم.

$$\Sigma_x f_1 \times \cdots \times f_k = f_1 \times \cdots \times f_i \, \Sigma_x \, f_{i+1} \times \cdots \times f_k = f_1 \times \cdots \times f_i \times f_{\bar{X}}$$

assuming $f_1, \ldots, f_i$ do not depend on X

Pointwise product of factors f_1 and f_2:

$$f_1(x_1, \ldots, x_j, y_1, \ldots, y_k) \times f_2(y_1, \ldots, y_k, z_1, \ldots, z_l)$$
$$= f(x_1, \ldots, x_j, y_1, \ldots, y_k, z_1, \ldots, z_l)$$

E.g., $f_1(a,b) \times f_2(b,c) = f(a,b,c)$

استنتاج با حذف متغیر

استفاده از عملیات پایه : محاسبه : مثال

INFERENCE BY VARIABLE ELIMINATION

$$
\begin{aligned}
& P(b, j, m) \\
& \quad = \underbrace{P(b)}_{B} \textstyle\sum_e \underbrace{P(e)}_{E} \textstyle\sum_a \underbrace{P(a|b,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M} \\
& \quad = f_B(b) \sum_e f_E(e) \sum_a f_A(a,b,e) f_J(a) f_M(a) \\
& \quad = f_B(b) \sum_e f_E(e) \sum_a f_{AJM}(a,b,e) \\
& \quad = f_B(b) \sum_e f_E(e) f_{\bar{A}JM}(b,e) \\
& \quad = f_B(b) \sum_e f_{\bar{A}EJM}(b,e) \\
& \quad = f_B(b) f_{\bar{A}\bar{E}JM}(b) \\
& \quad = f_{\bar{A}B\bar{E}JM}(b)
\end{aligned}
$$

استنتاج با حذف متغیر

الگوریتم عمومی

Inference by Variable Elimination

پرس‌وجو:

$$P(Q|E_1 = e_1, \dots E_k = e_k)$$

الگوریتم استنتاج با حذف متغیر

با فاکتورهای آغازین شروع می‌کنیم.
همان CPTهای محلی اما نمونه‌سازی شده با شواهد

تا زمانی که هنوز متغیرهای پنهان وجود دارند (غیر از Q یا شواهد)

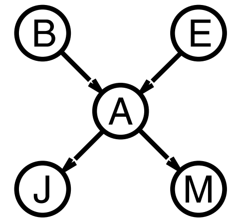
- یک متغیر پنهان H را انتخاب می‌کنیم.
- همه‌ی فاکتورهایی که به H اشاره دارند را join می‌کنیم.
- برای حذف H حاشیه‌سازی می‌کنیم (project).

فاکتورهای باقیمانده را join و نرمال‌سازی می‌کنیم

استنتاج با حذف متغیر

مثال (۱ از ۲)

INFERENCE BY VARIABLE ELIMINATION



$$\boxed{P(B|j,m)} \propto P(B,j,m)$$

$$\underbrace{P(B)}_{f_B(B)} \quad \underbrace{P(E)}_{f_E(E)} \quad \underbrace{P(A|B,E)}_{f_A(A,B,E)} \quad \underbrace{P(j|A)}_{f_J(A)} \quad \underbrace{P(m|A)}_{f_M(A)}$$

Choose A

الحاق

حاشیه‌سازی

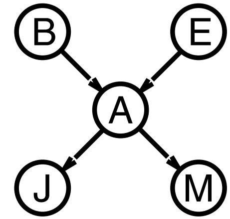
$$\begin{matrix} f_A(\underline{A},B,E) \\ f_J(\underline{A}) \\ f_M(\underline{A}) \end{matrix} \xrightarrow{\times} f_{AJM}(A,B,E) \xrightarrow{\Sigma} f_{\bar{A}JM}(B,E)$$

$$\boxed{f_B(B) \quad f_E(E) \quad f_{\bar{A}JM}(B,E)}$$

استنتاج با حذف متغیر

مثال (۲ از ۲)

INFERENCE BY VARIABLE ELIMINATION



$$f_B(B) \qquad f_E(E) \qquad f_{\bar{A}JM}(B,E)$$

Choose E

$$f_E(E), \; f_{\bar{A}JM}(B,E) \xrightarrow[\text{الحاق}]{\times} f_{\bar{A}EJM}(B,E) \xrightarrow[\text{حاشیه‌سازی}]{\Sigma} f_{\bar{A}\bar{E}JM}(B)$$

$$f_B(B) \qquad f_{\bar{A}\bar{E}JM}(B)$$

Finish

$$f_B(B), \; f_{\bar{A}\bar{E}JM}(B) \xrightarrow[\text{الحاق}]{\times} f_{\bar{A}B\bar{E}JM}(B) \xrightarrow[\text{نرمال‌سازی}]{\text{Normalize}} P(B|j,m)$$

استنتاج با حذف متغیر

شبه کد الگوریتم

INFERENCE BY VARIABLE ELIMINATION

function ELIMINATION-ASK(X, $\mathbf{e}$, bn) **returns** a distribution over X
  inputs: X, the query variable
    $\mathbf{e}$, evidence specified as an event
    bn, a belief network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $factors \leftarrow [\,]$; $vars \leftarrow$ REVERSE(VARS[bn])
  for each var **in** $vars$ **do**
    $factors \leftarrow$ [MAKE-FACTOR(var, $\mathbf{e}$)|$factors$]
    if var is a hidden variable **then** $factors \leftarrow$ SUM-OUT(var, $factors$)
  return NORMALIZE(POINTWISE-PRODUCT($factors$))

استنتاج با حذف متغیر

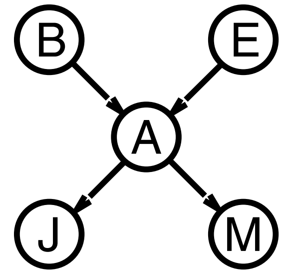
متغیرهای نامربوط

IRRELEVANT VARIABLES

Consider the query $P(JohnCalls|Burglary = true)$

$$P(J|b) = \alpha P(b) \sum_e P(e) \sum_a P(a|b,e) P(J|a) \sum_m P(m|a)$$

Sum over m is identically 1; M is **irrelevant** to the query



قضیه

$\boxed{Y \text{ is irrelevant unless } Y \in Ancestors(\{X\} \cup \mathbf{E})}$

مثال

Here, $X = JohnCalls$, $\mathbf{E} = \{Burglary\}$, and
$Ancestors(\{X\} \cup \mathbf{E}) = \{Alarm, Earthquake\}$
so $MaryCalls$ is irrelevant

اگر یک متغیر، جد یک متغیر پرس‌وجو یا مشاهده نباشد، به آن پرس‌وجو نامربوط است.

$\Leftarrow$ در الگوریتم حذف متغیر، می‌توان این متغیر نامربوط را پیش از ارزیابی پرس‌وجو حذف کرد.

استنتاج با حذف متغیر

متغیرهای نامربوط

IRRELEVANT VARIABLES

گراف مورال
Moral Graph

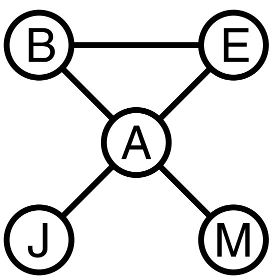
گراف مورال یک شبکه‌ی بیزی:
در شبکه‌ی بیزی، همه‌ی والدها را به هم پیوند می‌دهیم و پیکان‌ها را حذف می‌کنیم.

$\mathbf{A}$ is m-separated from $\mathbf{B}$ by $\mathbf{C}$ iff separated by $\mathbf{C}$ in the moral graph

قضیه

Y is irrelevant if m-separated from X by $\mathbf{E}$

مثال



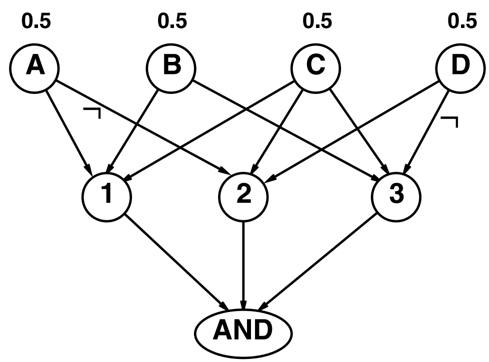
For $P(JohnCalls|Alarm = true)$, both $Burglary$ and $Earthquake$ are irrelevant

استنتاج دقیق

پیچیدگی

شبکه‌های بیزی	
شبکه‌های همبند تنها (چند درختی) *Singly Connected Networks (Polytrees)*	**شبکه‌های همبند چندتایی** *Multiply Connected Networks*
هر دو گره حداکثر با یک مسیر (بی‌جهت) به هم متصل می‌شوند. هزینه‌ی زمان و فضای حذف متغیر $O(d^k n)$	می‌توان 3SAT را به استنتاج دقیق کاهش داد NP-hard ⇐ معادل با **شمارش** مدل‌های 3SAT #P-complete ⇐

1. A v B v C
2. C v D v ¬A
3. B v C v ¬D



استنتاج با حذف متغیر

خوشه‌بندی برای افزایش کارآیی (الگوریتم درخت مشترک)

CLUSTERING (JOIN TREE ALGORITHMS)

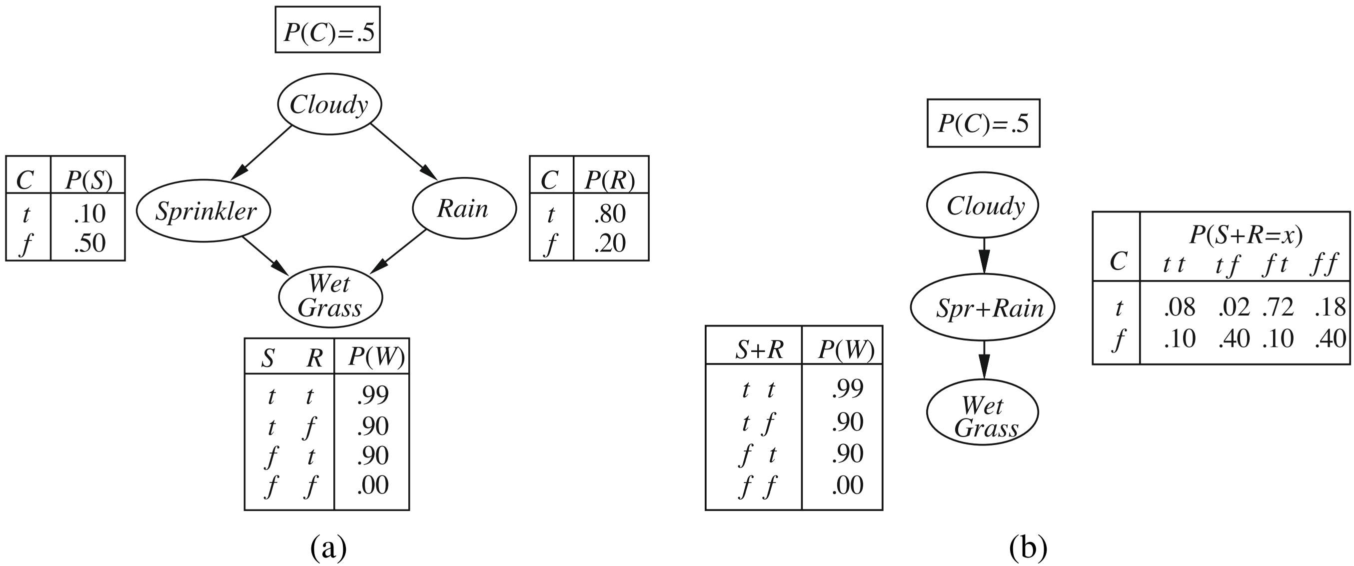


Figure 14.12 (a) A multiply connected network with conditional probability tables. (b) A clustered equivalent of the multiply connected network.

هوش مصنوعی پیشرفته

استدلال احتمالاتی

۵

استنتاج تقریبی در شبکه‌های بیزی

استنتاج در شبکه‌های بیزی

استنتاج تقریبی با شبیه‌سازی اتفاقی

استنتاج در شبکه‌های بیزی

استنتاج تقریبی
Approximate Inference

استنتاج دقیق
Exact Inference

زنجیره‌ی مارکوف مونت کارلو
Markov Chain Monte Carlo

شبیه‌سازی اتفاقی
Stochastic Simulation

حذف متغیر
Variable Elimination

برشمارش
Enumeration

استنتاج با شبیه‌سازی اتفاقی

INFERENCE BY STOCHASTIC SIMULATION

ایده‌ی پایه

N نمونه را از یک توزیع نمونه‌برداری S بیرون می‌کشیم.

↓

یک تقریب از احتمال پسین $\hat{P}$ محاسبه می‌کنیم.

↓

نشان می‌دهیم که این تقریب به احتمال واقعی P همگرا می‌شود.

$\boxed{0.5}$

Coin

طرح بحث:

- **نمونه‌برداری** از یک شبکه‌ی خالی
- **رد کردن نمونه‌برداری:** رد کردن نمونه‌های ناموافق با شواهد
- **وزن‌دهی درست‌نمایی:** استفاده از شواهد برای وزن‌دهی به نمونه‌ها
- **زنجیره‌ی مارکوف مونت کارلو (MCMC):**
نمونه‌برداری از یک فرآیند اتفاقی که توزیع ایستان آن، توزیع احتمال پسین واقعی است.

استنتاج با شبیه‌سازی اتفاقی

نمونه‌برداری از یک شبکه‌ی خالی

function PRIOR-SAMPLE(bn) **returns** an event sampled from bn
  inputs: bn, a belief network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $\mathbf{x} \leftarrow$ an event with n elements
  for $i = 1$ **to** n **do**
    $x_i \leftarrow$ a random sample from $\mathbf{P}(X_i \mid Parents(X_i))$
  return $\mathbf{x}$

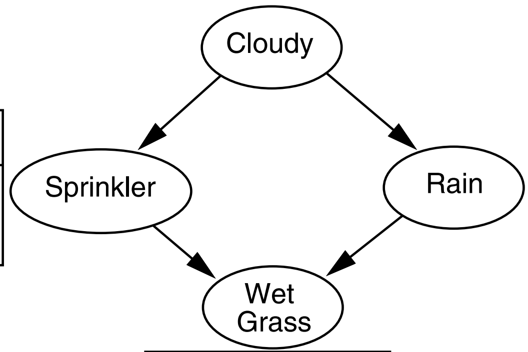
از هر متغیر به نوبت و به ترتیب توپولوژی نمونه‌برداری می‌شود.

توزیع احتمال متغیرهایی که مقادیر آنها نمونه‌برداری شده است،
به مقادیری که از قبل به متغیرهای والد نسبت داده شده است، مشروط می‌شود.

استنتاج با شبیه‌سازی اتفاقی

نمونه‌برداری از یک شبکه‌ی خالی: مثال (۱ از ۸)

P(C)
.50



C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

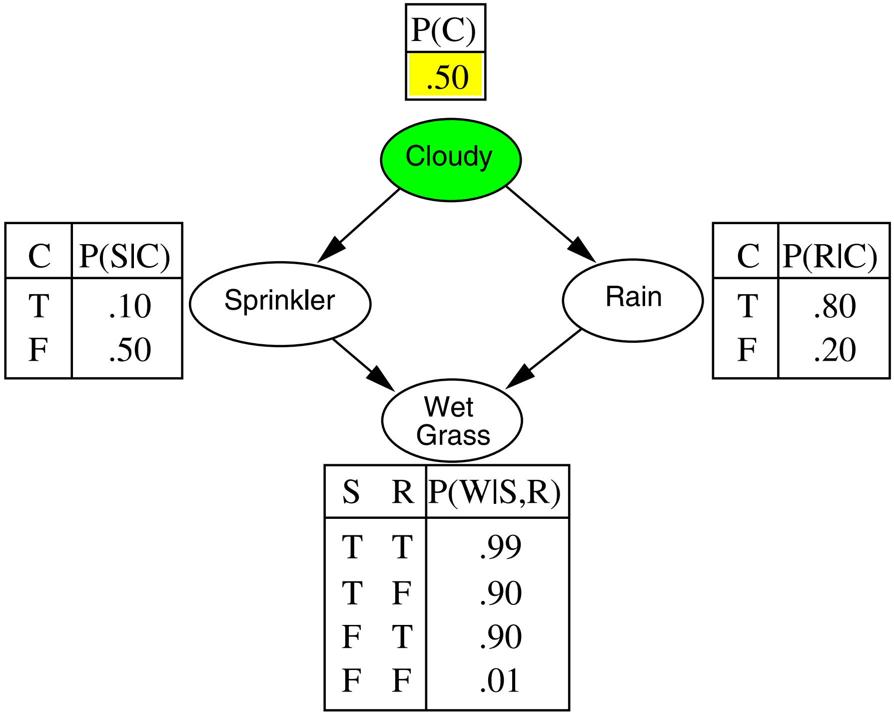
کاظم فولادی

استنتاج با شبیه‌سازی اتفاقی

نمونه‌برداری از یک شبکه‌ی خالی: مثال (۲ از ۸)

F

T



فرض می‌کنیم نمونه‌برداری از $\mathbf{P}(Cloudy) = <0.5,\ 0.5>$ مقدار $True$ برگرداند.

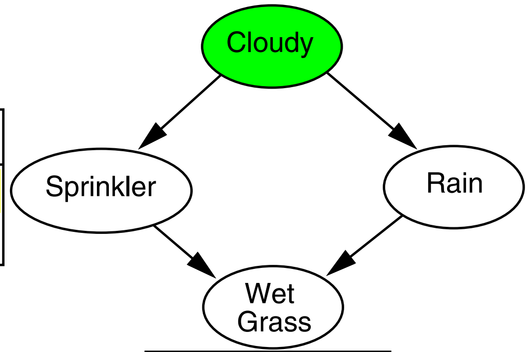
استنتاج با شبیه‌سازی اتفاقی

نمونه‌برداری از یک شبکه‌ی خالی: مثال (۳ از ۸)

F

T

P(C)
.50



C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

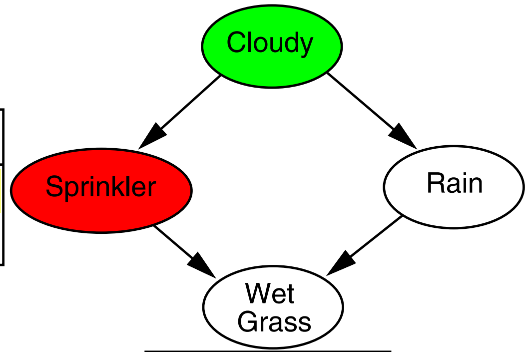
استنتاج با شبیه‌سازی اتفاقی

نمونه‌برداری از یک شبکه‌ی خالی: مثال (۴ از ۸)

F

T

P(C)
.50



C	P(S|C)
T	.10
F	.50

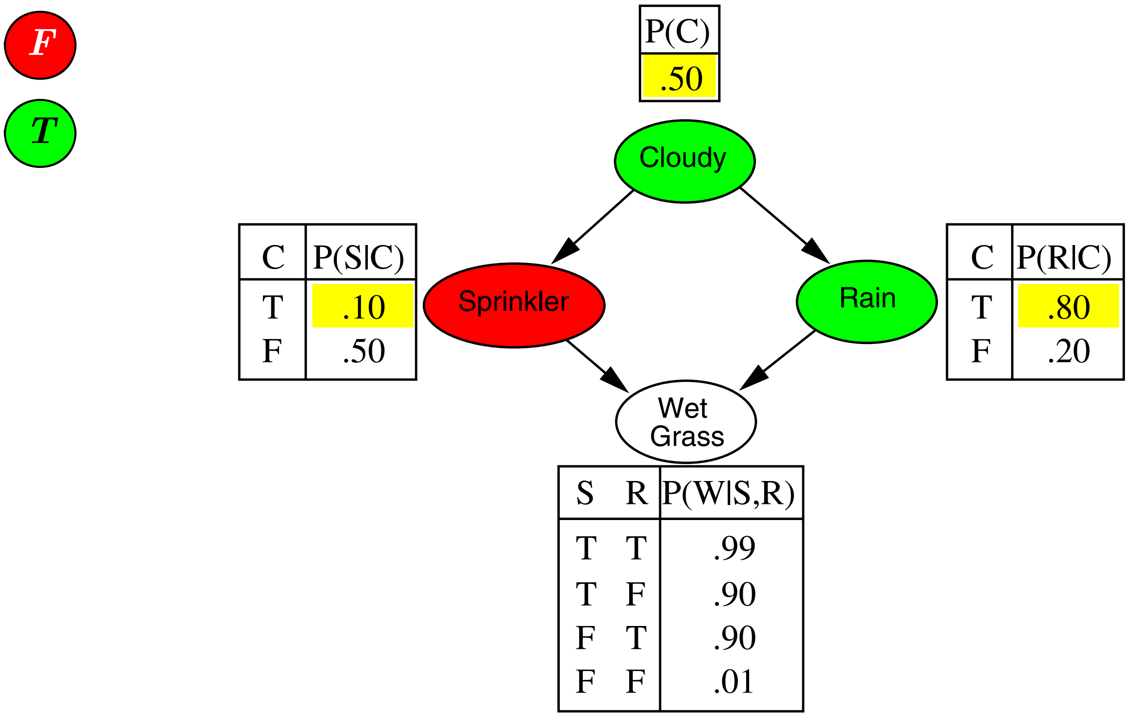
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

فرض می‌کنیم نمونه‌برداری از $\mathbf{P}(Sprinkler|\, Cloudy = True) = \langle 0.1,\ 0.9\rangle$ مقدار $False$ برگرداند.

استنتاج با شبیه‌سازی اتفاقی

نمونه‌برداری از یک شبکه‌ی خالی: مثال (۵ از ۸)



P(C)
.50

C	P(S|C)
T	.10
F	.50

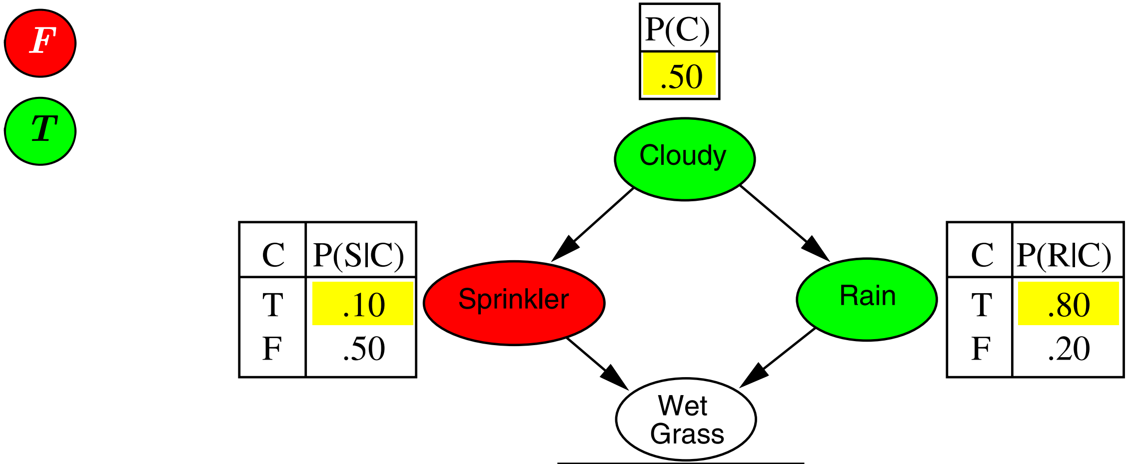
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

فرض می‌کنیم نمونه‌برداری از $\mathbf{P}(Rain|Cloudy = True) = <0.8,\ 0.2>$ مقدار $True$ برگرداند.

استنتاج با شبیه‌سازی اتفاقی

نمونه‌برداری از یک شبکه‌ی خالی: مثال (۶ از ۸)



P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

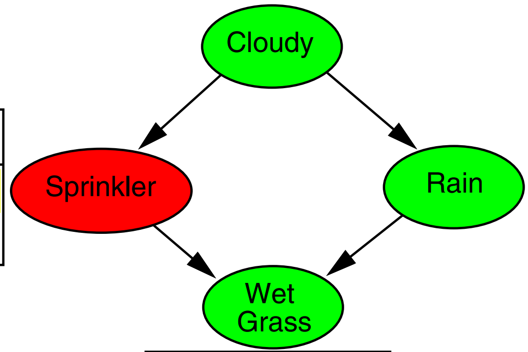
استنتاج با شبیه‌سازی اتفاقی

نمونه‌برداری از یک شبکه‌ی خالی: مثال (۷ از ۸)

F

T

P(C)
.50



C	P(S|C)
T	.10
F	.50

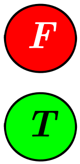
C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

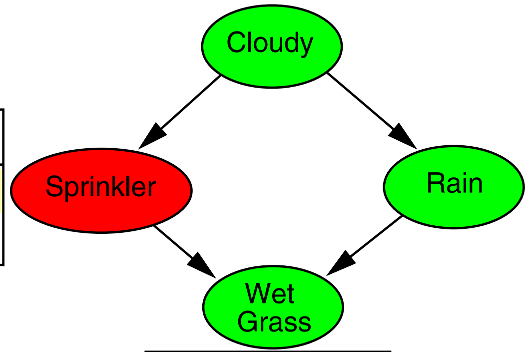
فرض می‌کنیم نمونه‌برداری از $\mathbf{P}(WetGrass|Sprinkler = False, Rain = True) = \langle 0.9,\ 0.1\rangle$ مقدار $True$ برگرداند.

استنتاج با شبیه‌سازی اتفاقی

نمونه‌برداری از یک شبکه‌ی خالی: مثال (۷ از ۸)



P(C)
.50



C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$\textsc{Prior-Sample}(bn)$ **returns** $[\mathit{True}, \mathit{False}, \mathit{True}, \mathit{True}]$

استنتاج با شبیه‌سازی اتفاقی

محاسبه

احتمال اینکه $\textsc{Prior-Sample}$ یک پیشامد خاص را تولید کند:

$$S_{PS}(x_1 \ldots x_n) = \Pi_{i=1}^{n} P(x_i|parents(X_i)) = P(x_1 \ldots x_n)$$

یعنی: احتمال پیشین واقعی

E.g., $S_{PS}(t, f, t, t) = 0.5 \times 0.9 \times 0.8 \times 0.9 = 0.324 = P(t, f, t, t)$

Let $N_{PS}(x_1 \ldots x_n)$ be the number of samples generated for event $x_1, \ldots, x_n$

Then we have

$$\begin{aligned}
\lim_{N\to\infty} \hat{P}(x_1, \ldots, x_n) &= \lim_{N\to\infty} N_{PS}(x_1, \ldots, x_n)/N \\
&= S_{PS}(x_1, \ldots, x_n) \\
&= P(x_1 \ldots x_n)
\end{aligned}$$

یعنی: تخمین‌های استخراج‌شده از $\textsc{Prior-Sample}$ سازگار **(consistent)** هستند.

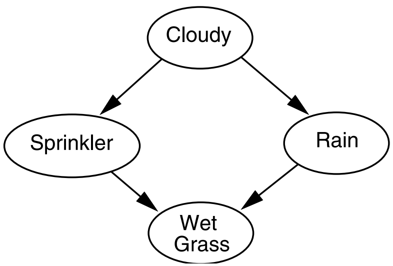
Shorthand: $\hat{P}(x_1, \ldots, x_n) \approx P(x_1 \ldots x_n)$

استنتاج با شبیه‌سازی اتفاقی

مثال

فرض کنید دسته‌ای از نمونه‌ها را از شبکه‌ی بیزی زیر گرفته باشیم:

$$[c, \neg s, r, w]$$
$$[c, s, r, w]$$
$$[\neg c, s, r, \neg w]$$
$$[c, \neg s, r, w]$$
$$[\neg c, s, \neg r, w]$$



برای محاسبه‌ی $\mathbf{P}(W)$:

با شمارش داریم: $<w\text{: } 4,\ \neg w\text{: } 1>$

با نرمال‌سازی به‌دست می‌آوریم: $\mathbf{P}(W) = <w\text{: } 0.8,\ \neg w\text{: } 0.2>$

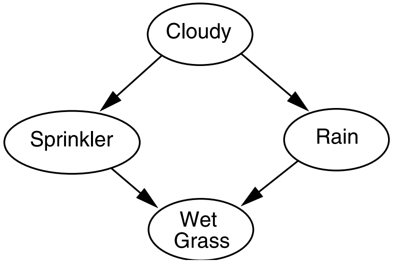
با داشتن نمونه‌های بیشتر، به توزیع واقعی نزدیک‌تر می‌شویم.

استنتاج با شبیه‌سازی اتفاقی

مثال

فرض کنید دسته‌ای از نمونه‌ها را از شبکه‌ی بیزی زیر گرفته باشیم:

$$[c, \neg s, r, w]$$
$$[c, s, r, w]$$
$$[\neg c, s, r, \neg w]$$
$$[c, \neg s, r, w]$$
$$[\neg c, s, \neg r, w]$$



برای محاسبه‌ی $\mathbf{P}(C|s)$:

برآمدهای C را شمارش می‌کنیم،

اما نمونه‌هایی که در آنها $S = s$ نیست را نادیده می‌گیریم (رد می‌کنیم: reject)

* به این روش **رد کردن نمونه‌برداری** می‌گوییم.

استنتاج با شبیه‌سازی اتفاقی

 رد کردن نمونه‌برداری

Rejection Sampling

$\hat{\mathbf{P}}(X|\mathbf{e})$ estimated from samples agreeing with $\mathbf{e}$

function Rejection-Sampling(X, $\mathbf{e}$, bn, N) **returns** an estimate of $P(X|\mathbf{e})$
  local variables: $\mathbf{N}$, a vector of counts over X, initially zero

  for $j = 1$ to N **do**
    $\mathbf{x} \leftarrow$ Prior-Sample(bn)
    if $\mathbf{x}$ is consistent with $\mathbf{e}$ **then**
      $\mathbf{N}[x] \leftarrow \mathbf{N}[x]+1$ where x is the value of X in $\mathbf{x}$
  return Normalize($\mathbf{N}[X]$)

استنتاج با شبیه‌سازی اتفاقی

رد کردن نمونه‌برداری: مثال

REJECTION SAMPLING

E.g., estimate $\mathbf{P}(Rain|Sprinkler = true)$ using 100 samples

27 samples have $Sprinkler = true$

Of these, 8 have $Rain = true$ and 19 have $Rain = false$.

$$\hat{\mathbf{P}}(Rain|Sprinkler = true) = \text{NORMALIZE}(\langle 8, 19\rangle) = \langle 0.296, 0.704\rangle$$

مشابه با یک روال تخمین تجربی در دنیای واقعی

استنتاج با شبیه‌سازی اتفاقی

رد کردن نمونه‌برداری: تحلیل

Rejection Sampling

$$\begin{aligned}\hat{\mathbf{P}}(X|\mathbf{e}) &= \alpha \mathbf{N}_{PS}(X,\mathbf{e}) && \text{(algorithm defn.)}\\ &= \mathbf{N}_{PS}(X,\mathbf{e})/N_{PS}(\mathbf{e}) && \text{(normalized by } N_{PS}(\mathbf{e})\text{)}\\ &\approx \mathbf{P}(X,\mathbf{e})/P(\mathbf{e}) && \text{(property of PriorSample)}\\ &= \mathbf{P}(X|\mathbf{e}) && \text{(defn. of conditional probability)}\end{aligned}$$

$\Leftarrow$ نمونه‌برداری با رد کردن، تخمین‌های پسین **سازگار** را برمی‌گرداند.

مشکل: این روش به طور ناامیدکننده‌ای گران و پرهزینه است اگر $P(\mathbf{e})$ کوچک باشد.
(در این صورت تعداد زیادی از نمونه‌ها دور ریخته می‌شود.)
$P(\mathbf{e})$ با تعداد متغیرهای شاهد به صورت نمایی افت می‌کند.

راه حل: **وزن‌دهی درست‌نمایی** (تنها نمونه‌های هماهنگ با شاهد $\mathbf{e}$ را تولید می‌کند.)

استنتاج با شبیه‌سازی اتفاقی

وزن‌دهی درست‌نمایی

LIKELIHOOD WEIGHTING

ایده‌ی پایه

متغیرهای شواهد را ثابت می‌گیریم.

↓

تنها متغیرهای غیر شواهد را نمونه‌برداری می‌کنیم.

↓

هر نمونه را با درست‌نمایی موافقت آن با شواهد وزن می‌دهیم.

بدون وزن‌دهی با درست‌نمایی، تخمین‌های پسین حاصل **سازگار** نخواهند بود.

استنتاج با شبیه‌سازی اتفاقی

وزن‌دهی درست‌نمایی : الگوریتم

LIKELIHOOD WEIGHTING

function LIKELIHOOD-WEIGHTING(X, $\mathbf{e}$, bn, N) **returns** an estimate of $P(X|\mathbf{e})$
  local variables: $\mathbf{W}$, a vector of weighted counts over X, initially zero

  for $j = 1$ to N **do**
    $\mathbf{x}$, $w \leftarrow$ WEIGHTED-SAMPLE(bn)
    $\mathbf{W}[x] \leftarrow \mathbf{W}[x] + w$ where x is the value of X in $\mathbf{x}$
  return NORMALIZE($\mathbf{W}[X]$)

---

function WEIGHTED-SAMPLE(bn, $\mathbf{e}$) **returns** an event and a weight

  $\mathbf{x} \leftarrow$ an event with n elements; $w \leftarrow 1$
  for $i = 1$ **to** n **do**
    if X_i has a value x_i in $\mathbf{e}$
      then $w \leftarrow w \times P(X_i = x_i \mid Parents(X_i))$
      else $x_i \leftarrow$ a random sample from $\mathbf{P}(X_i \mid Parents(X_i))$
  return $\mathbf{x}$, w

درست‌نمایی هر شاهد برای هر نمونه =
حاصل‌ضرب احتمالات شرطی آن متغیر شاهد با فرض معلوم بودن والدهای آن

استنتاج با شبیه‌سازی اتفاقی

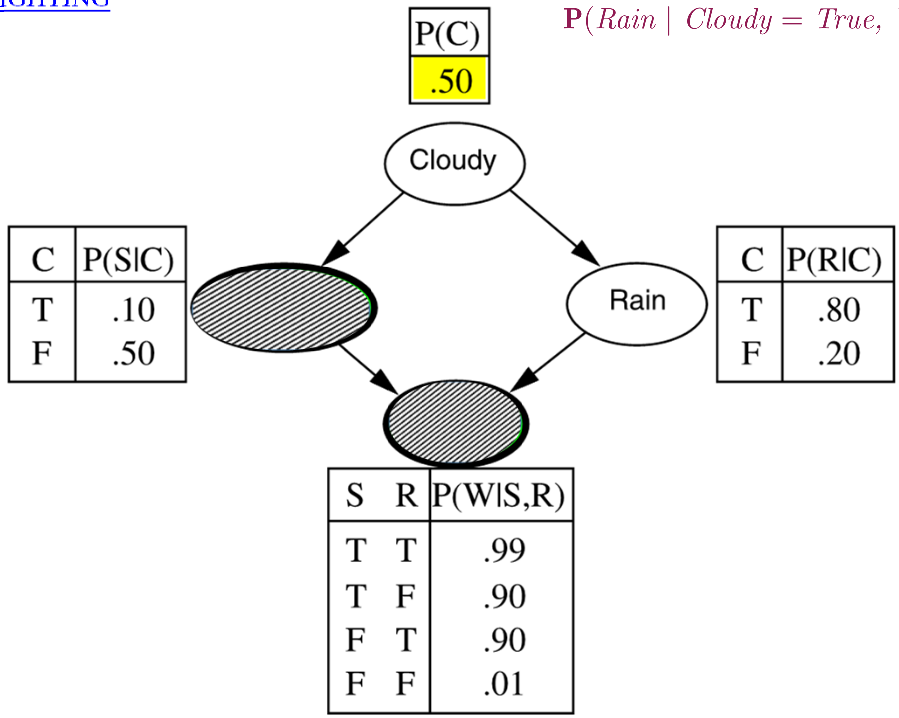
وزن‌دهی درست‌نمایی : مثال (۱ از ۸)

LIKELIHOOD WEIGHTING

$\mathbf{P}(Rain \mid Cloudy = True,\ WetGrass = True)$

F

T



$$w = 1.0$$

استنتاج با شبیه‌سازی اتفاقی

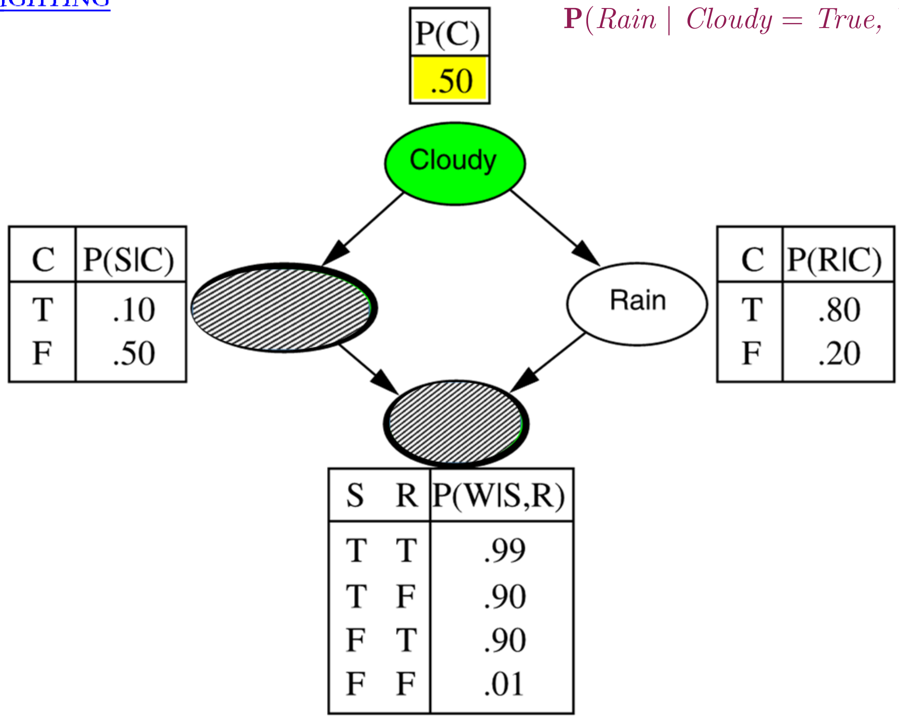
وزن‌دهی درست‌نمایی: مثال (۲ از ۸)

LIKELIHOOD WEIGHTING

$\mathbf{P}(Rain \mid Cloudy = True,\ WetGrass = True)$

F

T



$w = 1.0$

فرض می‌کنیم نمونه‌برداری از $\mathbf{P}(Cloudy) = \langle 0.5,\ 0.5 \rangle$ مقدار $True$ برگرداند.

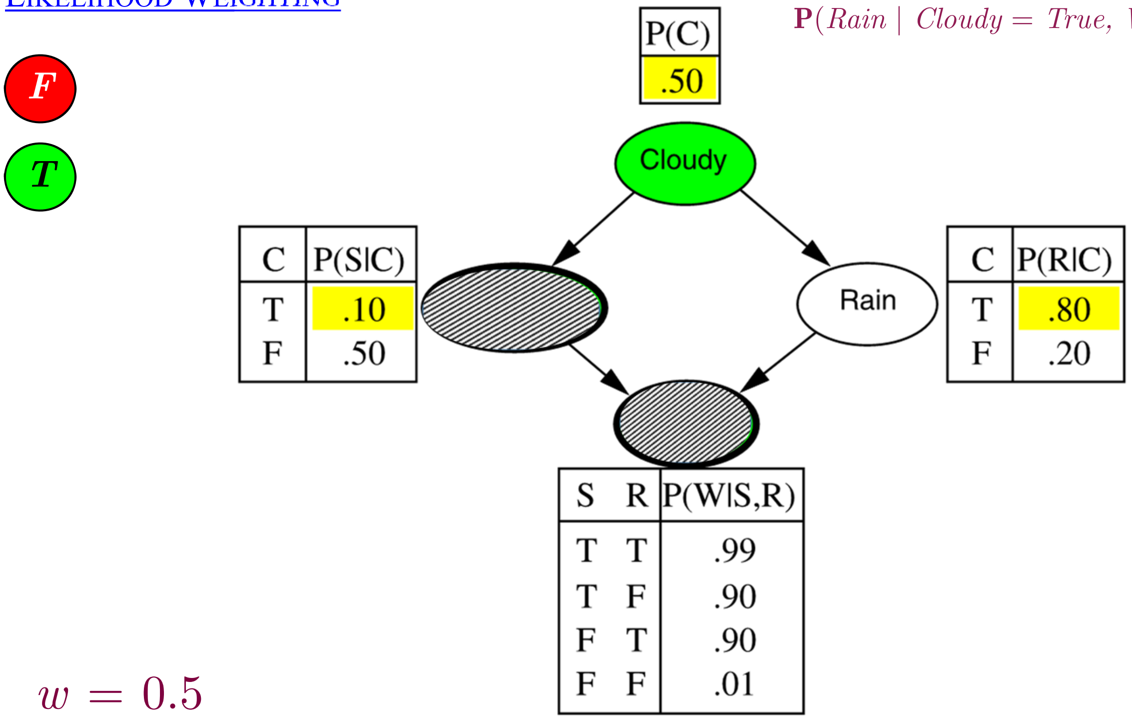
$$w \leftarrow w \times P(Cloudy = True) = 0.5$$

استنتاج با شبیه‌سازی اتفاقی

وزن‌دهی درست‌نمایی : مثال (۳ از ۸)

LIKELIHOOD WEIGHTING

$\mathbf{P}(Rain \mid Cloudy = True,\ WetGrass = True)$



$$w = 0.5$$

استنتاج با شبیه‌سازی اتفاقی

وزن‌دهی درست‌نمایی : مثال (۴ از ۸)

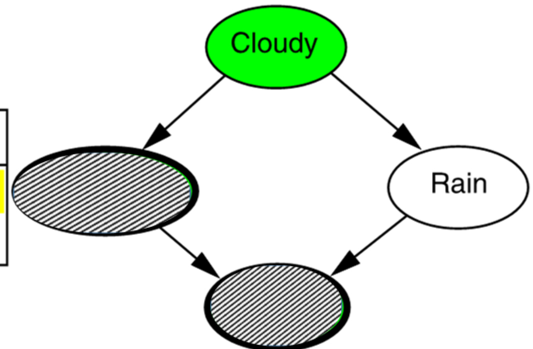
LIKELIHOOD WEIGHTING

$\mathbf{P}(Rain \mid Cloudy = True,\ WetGrass = True)$

F

T

P(C)
.50



C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$$w = 0.5$$

$Sprinkler$ یک متغیر شاهد نیست، پس از $P(Sprinkler \mid Cloudy = True) = \langle 0.8, 0.2 \rangle$ نمونه‌برداری می‌کنیم.
فرض می‌کنیم مقدار $False$ برمی‌گرداند.

استنتاج با شبیه‌سازی اتفاقی

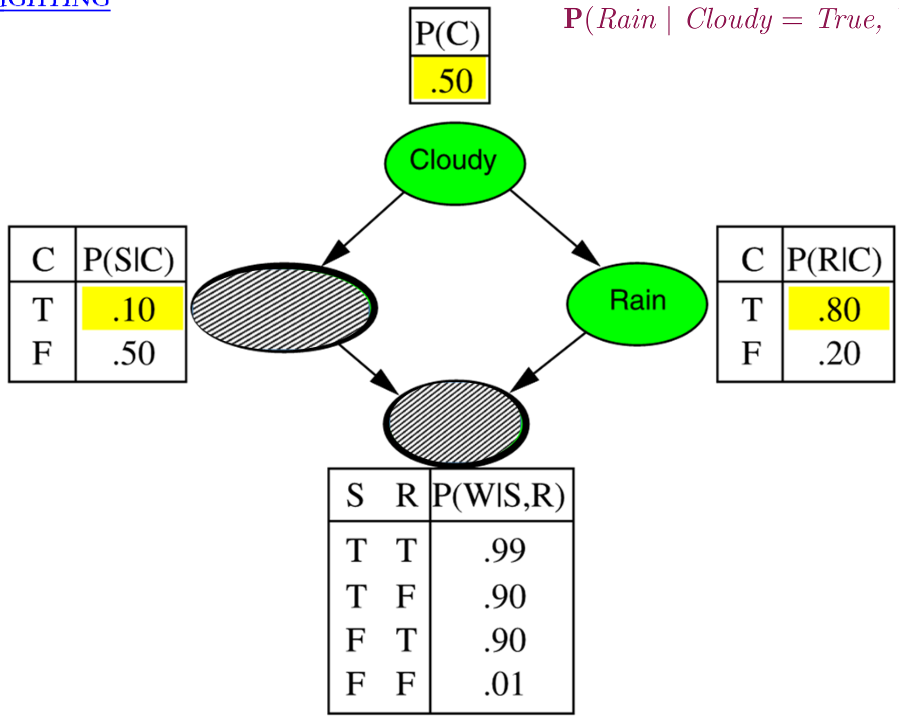
وزن‌دهی درست‌نمایی : مثال (۵ از ۸)

LIKELIHOOD WEIGHTING

$\mathbf{P}(Rain \mid Cloudy = True,\ WetGrass = True)$

F

T



$w = 0.5$

فرض می‌کنیم نمونه‌برداری از $\mathbf{P}(Rain|Cloudy = True) = <0.8,\ 0.2>$ مقدار $True$ برگرداند.

Prepared by Kazim Fouladi | Spring 2017 | 2nd Edition
کاظم فولادی

استنتاج با شبیه‌سازی اتفاقی

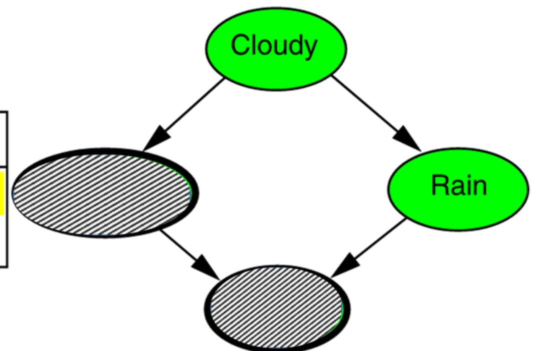
وزن‌دهی درستنمایی : مثال (۶ از ۸)

LIKELIHOOD WEIGHTING

$\mathbf{P}(Rain \mid Cloudy = True,\ WetGrass = True)$

F

T



P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

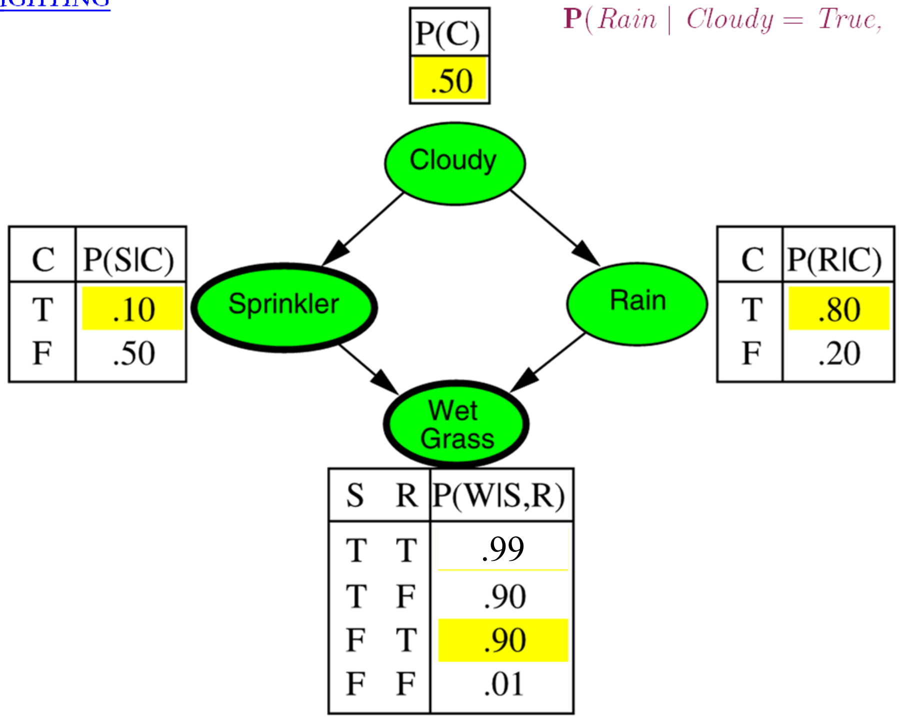
$$w = 0.5$$

استنتاج با شبیه‌سازی اتفاقی

وزن‌دهی درست‌نمایی: مثال (۷ از ۸)

LIKELIHOOD WEIGHTING

$$\mathbf{P}(Rain \mid Cloudy = Truc,\ WetGrass = True)$$



P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

$$w = 0.5$$

$WetGrass$ یک متغیر شاهد است که مقدار آن $True$ است، پس قرار می‌دهیم:

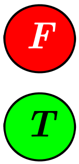
$$w \leftarrow w \times P(WetGrass = True \mid Sprinkler = False,\ Rain = True) = 0.5 \times 0.90 = 0.45$$

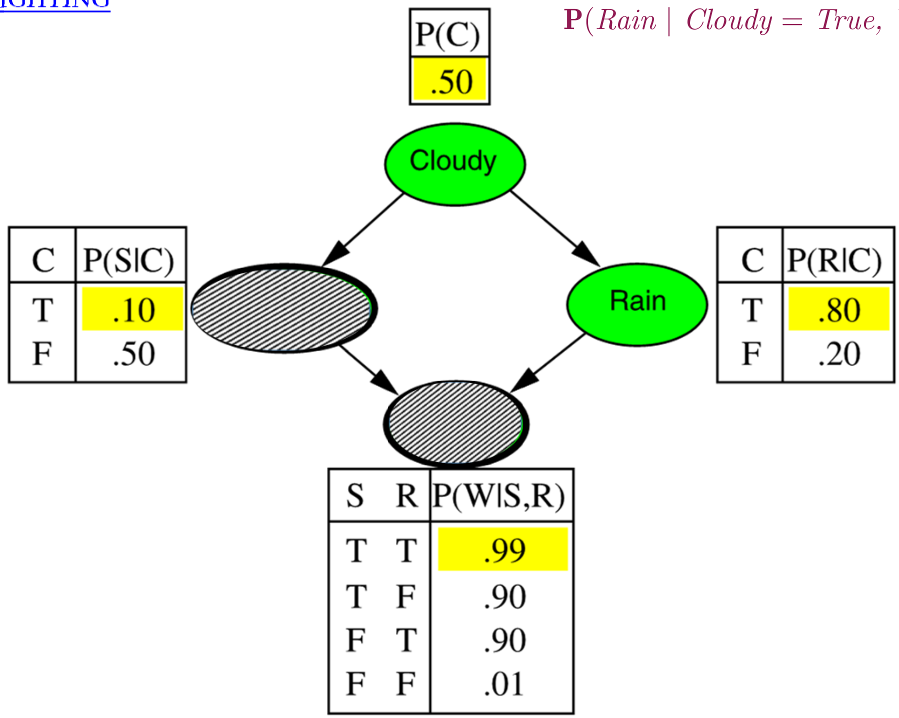
استنتاج با شبیه‌سازی اتفاقی

وزن‌دهی درست‌نمایی: مثال (۸ از ۸)

LIKELIHOOD WEIGHTING

$$\mathbf{P}(Rain \mid Cloudy = True,\ WetGrass = True)$$





$$w = 0.45$$

WEIGHTED-SAMPLE نمونه‌ی $[True, False, True, True]$ را با وزن 0.45 برمی‌گرداند:

با توصیف $Rain = True$ هماهنگ است (زیرا نمونه یک روز بارانی را توصیف می‌کند که بعید است آب‌پاشی صورت گیرد.).

استنتاج با شبیه‌سازی اتفاقی

وزن‌دهی درست‌نمایی : تحلیل

LIKELIHOOD WEIGHTING

احتمال نمونه‌برداری برای WEIGHTED-SAMPLE عبارت است از:

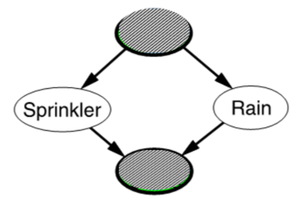
$$S_{WS}(\mathbf{z},\mathbf{e}) = \prod_{i=1}^{l} P(z_i \mid parents(Z_i))$$

سایر متغیرها

متغیرهای شواهد

تذکر: تنها به شواهد در **جد**ها توجه کنید

⇐ جایی «در میان» توزیع پیشین و پسین



وزن برای نمونه‌ی داده شده‌ی $\mathbf{z}$, $\mathbf{e}$ عبارت است از:

$$w(\mathbf{z},\mathbf{e}) = \prod_{i=1}^{m} P(e_i \mid parents(E_i))$$

احتمال نمونه‌برداری وزن‌دهی شده عبارت است از:

$$S_{WS}(\mathbf{z},\mathbf{e})w(\mathbf{z},\mathbf{e})$$
$$= \prod_{i=1}^{l} P(z_i \mid parents(Z_i)) \prod_{i=1}^{m} P(e_i \mid parents(E_i))$$
$$= P(\mathbf{z},\mathbf{e})$$ (بر اساس معناشناسی سراسری شبکه‌ی بیزی)

بنابراین، وزن‌دهی درست‌نمایی تخمین‌های سازگار را برمی‌گرداند،

اما: کارآیی هنوز با تعداد زیاد متغیرهای شواهد تنزل می‌یابد

(زیرا نمونه‌های اندکی تقریباً همه‌ی وزن مجموع را دارند)

Prepared by Kazim Fouladi | Spring 2017 | 2nd Edition
کاظم فولادی

استنتاج در شبکه‌های بیزی

استنتاج تقریبی با زنجیره‌ی مارکوف مونت کارلو

استنتاج در شبکه‌های بیزی

استنتاج دقیق *Exact Inference*		استنتاج تقریبی *Approximate Inference*	
برشماری *Enumeration*	حذف متغیر *Variable Elimination*	شبیه‌سازی اتفاقی *Stochastic Simulation*	زنجیره‌ی مارکوف مونت کارلو *Markov Chain Monte Carlo*

استنتاج با شبیه‌سازی اتفاقی

استنتاج تقریبی با استفاده از «زنجیره‌ی مارکوف مونت کارلو» MCMC

تولید هر نمونه با ایجاد یک تغییر تصادفی در نمونه‌ی تولید شده‌ی قبلی

ایده‌ی پایه

متغیرهای شواهد را ثابت می‌گیریم.

↓

حالت بعدی را با نمونه‌برداری یک متغیر با داشتن پتوی مارکوف تولید می‌کنیم.

↓

متغیرها را به نوبت نمونه‌برداری می‌کنیم.

حالت شبکه: انتساب فعلی به همه‌ی متغیرها

استنتاج با شبیه‌سازی اتفاقی

استنتاج تقریبی با استفاده از «زنجیره‌ی مارکوف مونت کارلو» MCMC: الگوریتم

MCMC-ASK

function GIBBS-ASK(X, **e**, bn, N) **returns** an estimate of $\mathbf{P}(X|\mathbf{e})$
  local variables: **N**, a vector of counts for each value of X, initially zero
    Z, the nonevidence variables in bn
    x, the current state of the network, initially copied from **e**

  initialize **x** with random values for the variables in **Z**
  for j = 1 to N **do**
    for each Z_i in **Z do**
      set the value of Z_i in **x** by sampling from $\mathbf{P}(Z_i|mb(Z_i))$
      $\mathbf{N}[x] \leftarrow \mathbf{N}[x] + 1$ where x is the value of X in **x**
  return NORMALIZE(**N**)

همچنین می‌توانیم یک متغیر را برای نمونه‌برداری تصادفی در هر زمان انتخاب کنیم.

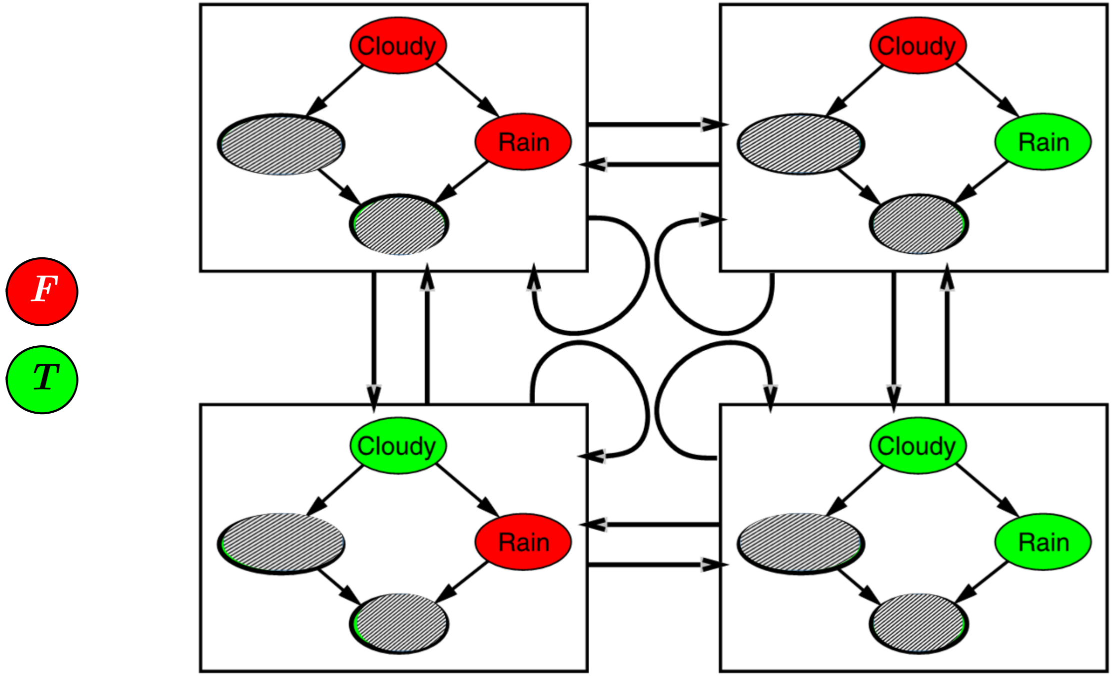
Prepared by Kazim Fouladi | Spring 2017 | 2nd Edition
کاظم فولادی

استنتاج با شبیه‌سازی اتفاقی

استنتاج تقریبی با استفاده از «زنجیره‌ی مارکوف مونت کارلو» MCMC: زنجیره‌ی مارکوف

Markov Chain

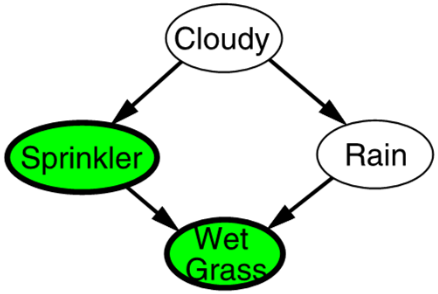
با $Sprinkler = true,\ WetGrass = true$ چهار حالت وجود دارد:



برای مدتی نگاه می‌کنیم و از آنچه دیدیم متوسط می‌گیریم.

استنتاج با شبیه‌سازی اتفاقی

استنتاج تقریبی با استفاده از «زنجیره‌ی مارکوف مونت کارلو» MCMC: مثال



Estimate $\mathbf{P}(Rain|Sprinkler = true, WetGrass = true)$

Sample $Cloudy$ or $Rain$ given its Markov blanket, repeat.
Count number of times $Rain$ is true and false in the samples.

E.g., visit 100 states
31 have $Rain = true$, 69 have $Rain = false$

$$\hat{\mathbf{P}}(Rain|Sprinkler = true, WetGrass = true)$$
$$= \text{NORMALIZE}(\langle 31, 69\rangle) = \langle 0.31, 0.69\rangle$$

Prepared by Kazim Fouladi | Spring 2017 | 2nd Edition

استنتاج با شبیه‌سازی اتفاقی

استنتاج تقریبی با استفاده از «زنجیره‌ی مارکوف مونت کارلو» MCMC: قضیه

قضیه

زنجیره‌ی مارکوف، به یک توزیع ایستان میل می‌کند:

کسر مدتی از زمان که در هر حالت صرف می‌شود،
متناسب است با احتمال پسین آن.

استنتاج با شبیه‌سازی اتفاقی

استنتاج تقریبی با استفاده از «زنجیره‌ی مارکوف مونت کارلو» MCMC: نمونه‌برداری پتوی مارکوف

Markov Blanket Sampling

Markov blanket of $Cloudy$ is
$Sprinkler$ and $Rain$
Markov blanket of $Rain$ is
$Cloudy$, $Sprinkler$, and $WetGrass$

احتمال با داشتن پتوی مارکوف به صورت زیر محاسبه می‌شود:

$$P(x_i'|mb(X_i)) = P(x_i'|parents(X_i))\Pi_{Z_j\in Children(X_i)}P(z_j|parents(Z_j))$$

به‌سادگی در «سیستم‌های موازی با گذردادن پیام» پیاده‌سازی می‌شود (مثل مغز).

مسائل محاسباتی اصلی:

۱) دشواری فهمیدن دست‌یابی به همگرایی
۲) اتلاف‌کننده است، اگر پتوی مارکوف بزرگ باشد:
$P(X_i|mb(X_i))$ چندان تغییر نخواهد کرد (قانون اعداد بزرگ).

استنتاج با شبیه‌سازی اتفاقی

استنتاج تقریبی با استفاده از «زنجیره‌ی مارکوف مونت کارلو» MCMC: تحلیل: طرح کلی

Transition probability $q(\mathbf{x} \longrightarrow \mathbf{x}\ )$

Occupancy probability $\pi_t(\mathbf{x})$ at time t

Equilibrium condition on π_t defines stationary distribution $\pi(\mathbf{x})$

Note: stationary distribution depends on choice of $q(\mathbf{x} \longrightarrow \mathbf{x}\ )$

Pairwise detailed balance on states guarantees equilibrium

Gibbs sampling transition probability:

sample each variable given current values of all others

$\Longrightarrow$ detailed balance with the true posterior

For Bayesian networks, Gibbs sampling reduces to sampling conditioned on each variable's Markov blanket

استنتاج تقریبی با استفاده از «زنجیره‌ی مارکوف مونت کارلو» MCMC: توزیع ایستان

Stationary Distribution

$\pi_t(\mathbf{x})$ = probability in state $\mathbf{x}$ at time t
$\pi_{t+1}(\mathbf{x})$ = probability in state $\mathbf{x}$ at time $t+1$

π_{t+1} in terms of π_t and $q(\mathbf{x} \longrightarrow \mathbf{x}')$

$$\pi_{t+1}(\mathbf{x}) = \Sigma_{\mathbf{x}} \pi_t(\mathbf{x}) q(\mathbf{x} \longrightarrow \mathbf{x}')$$

Stationary distribution: $\pi_t = \pi_{t+1} = \pi$

$$\pi(\mathbf{x}) = \Sigma_{\mathbf{x}} \pi(\mathbf{x}) q(\mathbf{x} \longrightarrow \mathbf{x}') \qquad \text{for all } \mathbf{x}$$

If π exists, it is unique (specific to $q(\mathbf{x} \longrightarrow \mathbf{x}')$)

In equilibrium, expected "outflow" = expected "inflow"

DETAILED BALANCE

"Outflow" = "inflow" for each pair of states:

$$\pi(\mathbf{x})q(\mathbf{x} \to \mathbf{x}') = \pi(\mathbf{x}')q(\mathbf{x}' \to \mathbf{x}) \qquad \text{for all } \mathbf{x},\ \mathbf{x}$$

Detailed balance $\Rightarrow$ stationarity:

$$\begin{aligned} \Sigma_{\mathbf{x}}\pi(\mathbf{x})q(\mathbf{x} \to \mathbf{x}') &= \Sigma_{\mathbf{x}}\pi(\mathbf{x}')q(\mathbf{x}' \to \mathbf{x}) \\ &= \pi(\mathbf{x}')\Sigma_{\mathbf{x}}q(\mathbf{x}' \to \mathbf{x}) \\ &= \pi(\mathbf{x}') \end{aligned}$$

MCMC algorithms typically constructed by designing a transition probability q that is in detailed balance with desired π

استنتاج با شبیه‌سازی اتفاقی

استنتاج تقریبی با استفاده از «زنجیره‌ی مارکوف مونت کارلو» MCMC: نمونه‌برداری گیبس

GIBBS SAMPLING

Sample each variable in turn, given **all other variables**

Sampling X_i, let $\bar{\mathbf{X}}_i$ be all other nonevidence variables
Current values are x_i and $\bar{\mathbf{x}}_i$; $\mathbf{e}$ is fixed
Transition probability is given by

$$q(\mathbf{x} \longrightarrow \mathbf{x}') = q(x_i, \bar{\mathbf{x}}_i \longrightarrow x_i', \bar{\mathbf{x}}_i) = P(x_i | \bar{\mathbf{x}}_i, \mathbf{e})$$

This gives detailed balance with true posterior $P(\mathbf{x}|\mathbf{e})$:

$$\begin{aligned}
\pi(\mathbf{x})q(\mathbf{x} \longrightarrow \mathbf{x}') &= P(\mathbf{x}|\mathbf{e})P(x_i'|\bar{\mathbf{x}}_i, \mathbf{e}) = P(x_i, \bar{\mathbf{x}}_i|\mathbf{e})P(x_i'|\bar{\mathbf{x}}_i, \mathbf{e}) \\
&= P(x_i|\bar{\mathbf{x}}_i, \mathbf{e})P(\bar{\mathbf{x}}_i|\mathbf{e})P(x_i'|\bar{\mathbf{x}}_i, \mathbf{e}) \quad \text{(chain rule)} \\
&= P(x_i|\bar{\mathbf{x}}_i, \mathbf{e})P(x_i', \bar{\mathbf{x}}_i|\mathbf{e}) \quad \text{(chain rule backwards)} \\
&= q(\mathbf{x}' \longrightarrow \mathbf{x})\pi(\mathbf{x}') = \pi(\mathbf{x}')q(\mathbf{x}' \longrightarrow \mathbf{x})
\end{aligned}$$

کاظم فولادی

PERFORMANCE OF APPROXIMATION ALGORITHMS

Absolute approximation: $|P(X|\mathbf{e}) - \hat{P}(X|\mathbf{e})| \leq \epsilon$

Relative approximation: $\frac{|P(X|\mathbf{e}) - \hat{P}(X|\mathbf{e})|}{P(X|\mathbf{e})} \leq \epsilon$

Relative $\Rightarrow$ absolute since $0 \leq P \leq 1$ (may be $O(2^{-n})$)

Randomized algorithms may fail with probability at most δ

Polytime approximation: $\text{poly}(n, \epsilon^{-1}, \log \delta^{-1})$

Theorem (Dagum and Luby, 1993): both absolute and relative approximation for either deterministic or randomized algorithms are NP-hard for any $\epsilon, \delta < 0.5$

(Absolute approximation polytime with no evidence—Chernoff bounds)

۶

مدل‌های احتمال رابطه‌ای و مرتبه اول

مدل‌های احتمال رابطه‌ای و مرتبه اول

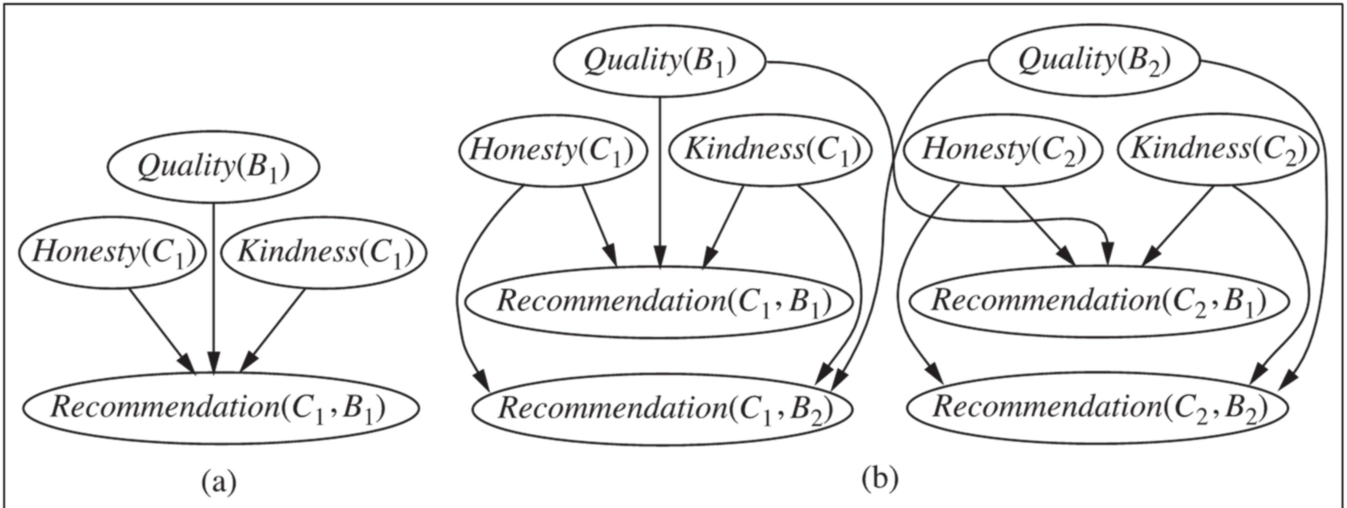


Figure 14.17 (a) Bayes net for a single customer C_1 recommending a single book B_1. $Honest(C_1)$ is Boolean, while the other variables have integer values from 1 to 5. (b) Bayes net with two customers and two books.

مدل‌های احتمال رابطه‌ای و مرتبه اول

$$P(\phi) = \sum_{\omega:\phi \text{ is true in } \omega} P(\omega)$$

مدل‌های احتمال رابطه‌ای و مرتبه اول

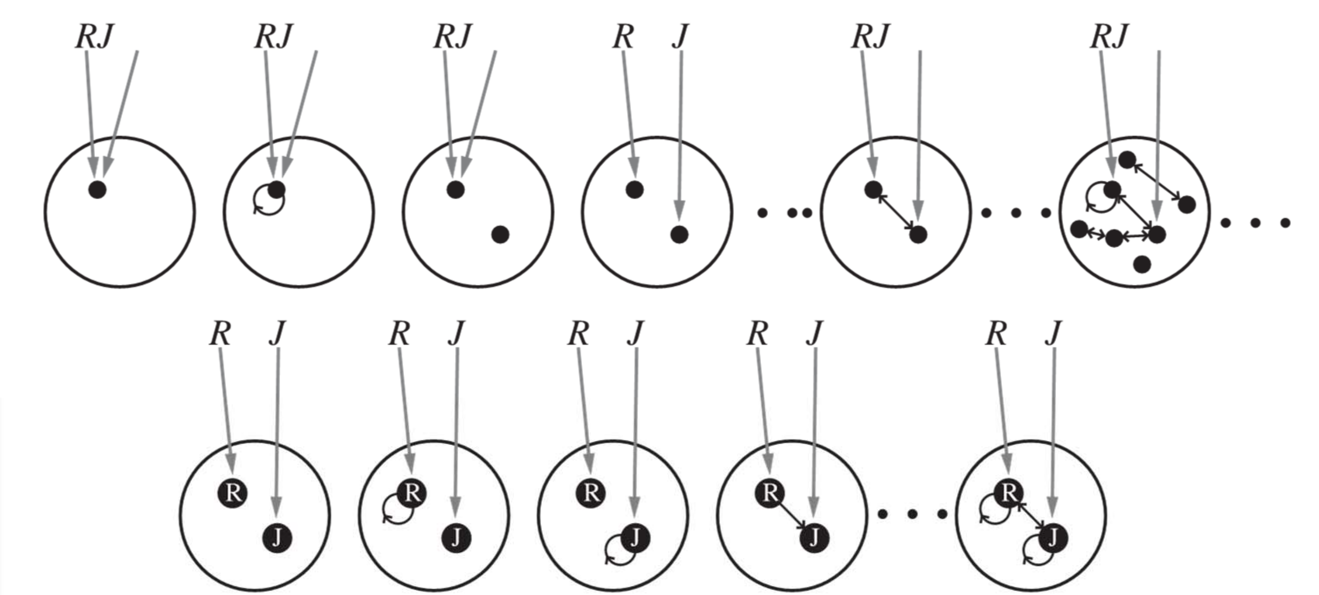


Figure 14.18 Top: Some members of the set of all possible worlds for a language with two constant symbols, R and J, and one binary relation symbol, under the standard semantics for first-order logic. Bottom: the possible worlds under database semantics. The interpretation of the constant symbols is fixed, and there is a distinct object for each constant symbol.

مدل‌های احتمال رابطه‌ای و مرتبه اول

مدل‌های احتمال رابطه‌ای

RELATIONAL PROBABILITY MODELS

$$Honest : Customer \rightarrow \{true, false\} Kindness : Customer \rightarrow \{1,2,3,4,5\}$$
$$Quality : Book \rightarrow \{1,2,3,4,5\}$$
$$Recommendation : Customer \times Book \rightarrow \{1,2,3,4,5\}$$

$$Honest(c) \sim \langle 0.99, 0.01 \rangle$$
$$Kindness(c) \sim \langle 0.1, 0.1, 0.2, 0.3, 0.3 \rangle$$
$$Quality(b) \sim \langle 0.05, 0.2, 0.4, 0.2, 0.15 \rangle$$
$$Recommendation(c,b) \sim RecCPT(Honest(c), Kindness(c), Quality(b))$$

$$\begin{aligned} Recommendation(c,b) \sim \quad & \textbf{if } Honest(c) \textbf{ then} \\ & \quad HonestRecCPT(Kindness(c), Quality(b)) \\ & \textbf{else } \langle 0.4, 0.1, 0.0, 0.1, 0.4 \rangle \ . \end{aligned}$$

$$\begin{aligned} Recommendation(c,b) \sim \quad & \textbf{if } Honest(c) \textbf{ then} \\ & \quad \textbf{if } Fan(c, Author(b)) \textbf{ then } Exactly(5) \\ & \quad \textbf{else } HonestRecCPT(Kindness(c), Quality(b)) \\ & \textbf{else } \langle 0.4, 0.1, 0.0, 0.1, 0.4 \rangle \end{aligned}$$

مدل‌های احتمال رابطه‌ای

RELATIONAL PROBABILITY MODELS

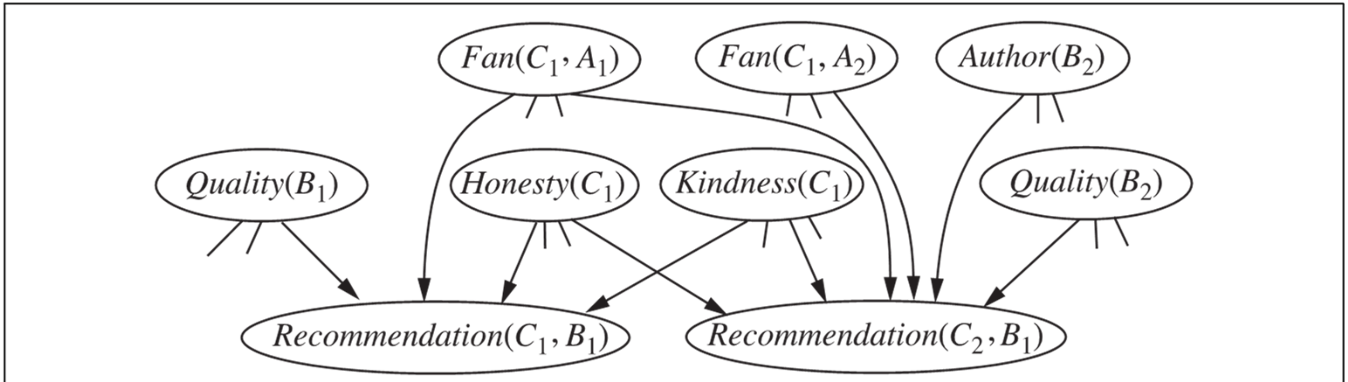


Figure 14.19 Fragment of the equivalent Bayes net when $Author(B_2)$ is unknown.

استدلال احتمالاتی

۷

سایر روش‌های استدلال نامطئن

روش‌های برخورد با عدم اطمینان

روش‌های برخورد با عدم اطمینان (استدلال نامطمئن)				
منطق غیریکنوا/پیش‌فرض *Default/Nonmonotonic Logic*	مبتنی بر قاعده *Rule-based*	نظریه دمپستر- شافر *Dempster–Shafer theory*	منطق فازی *Fyzzy Logic*	نظریه احتمال *Probability Theory*
استدلال‌های کیفی *Qualitative Reasoning*	با فاکتور فاج *Using Fudge Factor*	بازنمایی ناآگاهی *Representing Ignorance*	بازنمایی سربستگی *Representing Vagueness*	درجه‌ی باور به درستی رویداد *Belief Degree for Truth*

روش‌های برخورد با عدم اطمینان

منطق غیریکنوا / منطق پیش‌فرض

روش‌های برخورد با عدم اطمینان (استدلال نامطمئن)				
منطق غیریکنوا/پیش‌فرض *Default/Nonmonotonic Logic*	مبتنی بر قاعده *Rule-based*	نظریه دمپستر-شافر *Dempster-Shafer theory*	منطق فازی *Fuzzy Logic*	نظریه احتمال *Probability Theory*
استدلال‌های کیفی *Qualitative Reasoning*	با فاکتور فاج *Using Fudge Factor*	بازنمایی ناآگاهی *Representing Ignorance*	بازنمایی سربستگی *Representing Vagueness*	درجه‌ی باور به درستی رویداد *Belief Degree for Truth*

استفاده از استدلال‌های کیفی (مشابه منطق انسانی) به‌جای محاسبات عددی

منطق پیش‌فرض:

برخورد با نتایج به صورت باور
تا زمانی که دلیل بهتری برای باور به چیز دیگری پیدا شود.

مثال: به‌فرض، ماشین من لاستیک پنچر ندارد،
به‌فرض، A_{25} من را به‌موقع می‌رساند مگر اینکه با شاهدی تناقض پیدا کند.

مشکلات: ❋ چه فرض‌هایی مستدل هستند؟ ❋ چگونگی با تناقض‌ها برخورد کنیم؟

روش‌های برخورد با عدم اطمینان

مبتنی بر قاعده

روش‌های برخورد با عدم اطمینان (استدلال نامطمئن)				
منطق غیر یکنوا *Nonmonotonic Logic*	**مبتنی بر قاعده** *Rule-based*	نظریه دمپستر-شافر *Dempster-Shafer theory*	منطق فازی *Fuzzy Logic*	نظریه احتمال *Probability Theory*
استدلال‌های کیفی *Qualitative Reasoning*	**با فاکتور فاج** *Using Fudge Factor*	بازنمایی ناآگاهی *Representing Ignorance*	بازنمایی سربستگی *Representing Vagueness*	درجه‌ی باور به درستی رویداد *Belief Degree for Truth*

ساخت سیستم‌های مبتنی بر قاعده‌ی منطقی،
با اضافه کردن نوعی عامل فاج به هر قاعده، برای برخورد با عدم اطمینان

$$A_{25} \mapsto_{0.3} AtAirportOnTime$$
$$Sprinkler \mapsto_{0.99} WetGrass$$
$$WetGrass \mapsto_{0.7} Rain$$

مشکلات: ❋ چگونگی ترکیب نتایج: مثلاً $Sprinkler \mapsto Rain$?? با چه درجه‌ای؟

روش‌های برخورد با عدم اطمینان

نظریه دمپستر- شافر

روش‌های برخورد با عدم اطمینان (استدلال نامطمئن)				
منطق غیریکنواخت *Nonmonotonic Logic*	مبتنی بر قاعده *Rule-based*	**نظریه دمپستر- شافر** *Dempster–Shafer theory*	منطق فازی *Fuzzy Logic*	نظریه احتمال *Probability Theory*
استدلال‌های کیفی *Qualitative Reasoning*	با فاکتور فاج *Using Fudge Factor*	**بازنمایی ناآگاهی** *Representing Ignorance*	بازنمایی سربستگی *Representing Vagueness*	درجه‌ی باور به درستی رویداد *Belief Degree for Truth*

استفاده از درجه‌های باور با مقادیر بازه‌ای (کران بالا، کران پایین)
برای بازنمایی دانایی عامل در مورد احتمال یک گزاره

مشکلات: ❋ پیچیدگی بالای محاسباتی

روش‌های برخورد با عدم اطمینان

منطق فازی

روش‌های برخورد با عدم اطمینان (استدلال نامطمئن)				
منطق غیریکنوا Nonmonotonic Logic	مبتنی بر قاعده Rule-based	نظریه دمپستر-شافر Dempster-Shafer theory	**منطق فازی** *Fyzzy Logic*	نظریه احتمال Probability Theory
استدلال‌های کیفی Qualitative Reasoning	با فاکتور فاج Using Fudge Factor	بازنمایی ناآگاهی Representing Ignorance	**بازنمایی سربستگی** *Representing Vagueness*	درجه‌ی باور به درستی رویداد Belief Degree for Truth

هستی‌شناسی منطق فازی: اجازه دادن به **سربستگی**

یک گزاره می‌تواند تا مرتبه‌ای درست باشد **(درجه‌ی درستی)**

احتمالات و منطق معمولی تعهدات هستی‌شناختی یکسانی دارند:
گزاره‌های دنیا **درست یا نادرست** هستند،
حتی اگر عامل نامطمئن باشد که کدام یک است.

❋ مناسب برای کار کردن با مفاهیمی که حد و مرز دقیقی ندارند.

روش‌های برخورد با عدم اطمینان

نظریه‌ی احتمال

روش‌های برخورد با عدم اطمینان (استدلال نامطمئن)				
منطق غیریکنواخت *Nonmonotonic Logic*	مبتنی بر قاعده *Rule-based*	نظریه دمپستر-شافر *Dempster-Shafer theory*	منطق فازی *Fuzzy Logic*	**نظریه احتمال** *Probability Theory*
استدلال‌های کیفی *Qualitative Reasoning*	با فاکتور فاج *Using Fudge Factor*	بازنمایی ناآگاهی *Representing Ignorance*	بازنمایی سربستگی *Representing Vagueness*	**درجه‌ی باور به درستی رویداد** *Belief Degree for Truth*

به هر جمله یک درجه‌ی باور بین صفر و یک نسبت داده می‌شود.

احتمالات، روشی برای جمع‌بندی عدم اطمینان ناشی از تنبلی و ناآگاهی است.

مثال: A_{25} من را به‌موقع می‌رساند با احتمال 0.04

مشکلات: ❋ چه فرض‌هایی مستدل هستند؟ ❋ چگونگی با تناقض‌ها برخورد کنیم؟

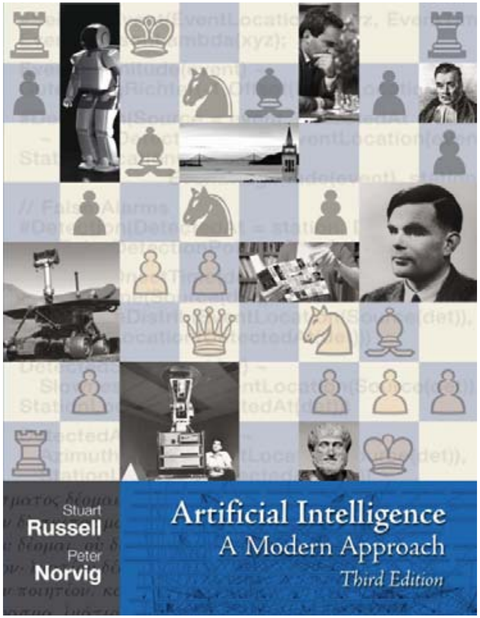
Prepared by Kazim Fouladi | Spring 2017 | 2nd Edition
کاظم فولادی

استدلال احتمالاتی

۸

منابع

منبع اصلی



Stuart Russell and Peter Norvig,
Artificial Intelligence: A Modern Approach,
3rd Edition, Prentice Hall, 2010.

Chapter 14

14 PROBABILISTIC REASONING

In which we explain how to build network models to reason under uncertainty according to the laws of probability theory.

Chapter 13 introduced the basic elements of probability theory and noted the importance of independence and conditional independence relationships in simplifying probabilistic representations of the world. This chapter introduces a systematic way to represent such relationships explicitly in the form of **Bayesian networks**. We define the syntax and semantics of these networks and show how they can be used to capture uncertain knowledge in a natural and efficient way. We then show how probabilistic inference, although computationally intractable in the worst case, can be done efficiently in many practical situations. We also describe a variety of approximate inference algorithms that are often applicable when exact inference is infeasible. We explore ways in which probability theory can be applied to worlds with objects and relations—that is, to *first-order*, as opposed to *propositional*, representations. Finally, we survey alternative approaches to uncertain reasoning.

14.1 Representing Knowledge in an Uncertain Domain

In Chapter 13, we saw that the full joint probability distribution can answer any question about the domain, but can become intractably large as the number of variables grows. Furthermore, specifying probabilities for possible worlds one by one is unnatural and tedious.

We also saw that independence and conditional independence relationships among variables can greatly reduce the number of probabilities that need to be specified in order to define the full joint distribution. This section introduces a data structure called a **Bayesian network**[1] to represent the dependencies among variables. Bayesian networks can represent essentially *any* full joint probability distribution and in many cases can do so very concisely.

BAYESIAN NETWORK

[1] This is the most common name, but there are many synonyms, including **belief network**, **probabilistic network**, **causal network**, and **knowledge map**. In statistics, the term **graphical model** refers to a somewhat broader class that includes Bayesian networks. An extension of Bayesian networks called a **decision network** or **influence diagram** is covered in Chapter 16.