چگونه با نامعلومی ، برخورد کنیم ؟

به صورت ناآشکار (ضمنی یا تلویحی)

به مواردی که نامعلوم هستند ، در زمانی که شما می توانید کار خود را انجام دهید ، بی اعتنا باشید و عملیّاتی را به وجود بیاورید که در مقابل نامعلومی ، مقاوم هستند .

به صورت آشکار

مدلی از جهان را بسازید که نامعلومی را در وضعیت ها ، حرکت ها و مشاهده هایش در نظر می گیرد و در مورد اثر عملیّاتی که مدل انجام می دهد دلیل بیاورید .

منطق و نامعلومی

ما به زودی خواهیم دید که چگونه منطق را با نامعلومی به کار ببریم ؛ مثال هایی از منطق را در زیر مشاهده می نمایید :

$$Studies(Bart) \vee WatchesTV(Bart)$$

$$Hungry(Homer) \Rightarrow Eats(Homer, HotDog) \vee Eats(Homer, Pie)$$

$$\exists x : Hungry(x)$$

متاسفانه ، نگرش منطقی دارای برخی اشکالات می باشد .

ضعف های منطق [۱]

نمی توانیم همه ی نتیجه های ممکن را تایین کنیم ؛ مثلا ، نمی دانیم که اگر الان از خانه خارج شویم ، آیا به موقع در محل مورد نظر خواهیم بود یا نه ؟ . یا اگر پدال گاز اتومبیل خود را زیاد تر فشار دهیم آیا باعث زودتر رسیدن ما به مقصد می شود یا اینکه باعث می شود که ما تصادف کنیم ؟! .

ما ممکن است همه ی نتیجه های ممکن را ندانیم . مثلا ، " در صورتی که یک بیمار دارای دندان درد باشد ، وی ممکن است پوسیدگی [۲] در دندان خود داشته باشد ، یا ممکن است دارای بیماری لثه [۳] باشد ، یا ممکن است دلیلی دیگر داشته باشد که ما چیزی در مورد آن نمی دانیم . "

ما هیچ راهی برای صحبت در مورد احتمال حوادث نداریم . مثلا ، " امروز ممکن است که من با تیر چراغ برق برخورد کنم . "

کمّیّت و کیفیّت

منطق ، یک نگرش کمّی به نامعلومی را ارایه می دهد ؛ ما می توانیم بگوییم که یک اتّفاق ، بیش تر از دیگری روی می دهد یا این که ممکن است یک اتّفاق روی دهد . این روش در مواردی که ما آماری نداریم ، مفید می باشد . احتمال به ما اجازه می دهد که به صورت کیفی بحث کنیم ؛ ما احتمال رخ دادن یک اتفاق را با یک عدد بیان می کنیم .

نامعلومی و عقلانیّت (هوشمند بودن)

---

[۱] weaknesses with logic

[۲] cavity

[۳] gum

تعریف ما از عقلانیّت را به یاد بیاورید ؛ یک عامل هوشمند (عقلانی) ، عاملی است که با بیشینه نمودن معیارکارایی عمل می نماید . ما چگونه این عامل را در یک جهان دارای نامعلومی تعریف نماییم ؟ . می گوییم ، عملکرد عامل دارای نتیجه هایی است و هر نتیجه را با یک عدد که نشان دهنده ی احتمال آن است بیان می نماییم . سپس یک عامل می تواند به هر کدام از نتایج ممکن و عدد آن ها و احتمال رخدادن نتیجه توجه نماید و عملی را که دارای بالاترین سود مورد انتظار می باشد را انتخاب نماید .

تیوری ترکیب کننده ی برتری ها با احتما ل رخداد یک نتیجه ، *تیوری تصمیم گیری*[1] نام دارد .

مثال : بیایید عملکرد A_t را ترک کردن به مقصد فرودگاه ، t دقیقه قبل از پرواز قرار دهیم . آیا در زمان A_t به موقع آن جا خواهم بود ؟

مشکلات :

۱) مشاهده پذیری جزیی[2] ؛ مثل ، حالت جاده ، برنامه ی رانندگان دیگر اتومبیل ها و .....

۲) حس گرهای پر پارازیت[3](اغتشاش) ؛ مثلا ، گزارش های ترافیک ممکن است درست نباشند .

۳) نامعلومی در نتایج عملکرد[4] ؛ مثلا ممکن است تایر اتومبیل مان پنچر شود و .....

۴) پیچیدگی بیش از حد[5] طرح ریزی[1] و پیش بینی ترافیک[2]

---

[1] decision theory
[2] partial observability
[3] noisy sensors
[4] uncertainty in action outcomes
[5] immense

بنابراین یک شیوه ی منطقی یا با ریسک های دروغ [۳] است ؛ مثلا ، [این که بگوییم] " با A_{25} به موقع خواهم رسید " یا به نتایجی منجر می شود که خیلی برای تصمیم گیری ضعیفند ؛ " A_{25} در صورتی مرا به موقع خواهد رساند که تصادفی در راه رخ ندهد و باران نبارد و لاستیک اتومبیل من بی عیب باشد و ..... " (A_{1440} معقولانه است که مرا به موقع برساند اما من باید در طول شب در فرودگاه بمانم ! .....)

روش هایی برای به کارگیری نامعلومی

منطق پیش فرض یا غیر یکنواخت [۴]: فرض کنید که اتومبیل من دارای تایر پنچر نباشد .
یا فرض کنید که A_{25} کار می کند مگر این که با تناقض [۵] ثابت شود که کار نمی کند .

احتمال

یک چارچوب مشهور و قابل فهم برای نامعلومی می باشد که دارای معنی های واضح است و برای ترکیب شواهد ، استدلال در مورد پیشگویی و تشخیص و ترکیب شواهد جدید ، جواب های اساسی را مهیّا می کند . احتمال ، برخی از سطح های انسان های ماهر (خبره های انسانی) را شناسایی می کند .

اساس احتمال

به عنوان مثال ، احتمال های P(BartStudied) = 0.01 و P(Hungry(Homer)) = 0.99 را در نظر بگیرید . توجه کنید که خود گزاره ممکن است درست یا غلط باشد و فرق دارد با اینکه بگوییم ، عبارت به صورت جزیی ، درست است .

---

[۱] modelling
[۲] predicting traffic
[۳] risks falsehood
[۴] nonmonotonic
[۵] contradiction

مقادیر تصادفی [۱]

یک متغیّر تصادفی ، متغیّر یا گزاره ای است که مقدارش ناشناخته است و دارای دامنه ای از مقادیر است که می تواند آن ها را به خود بگیرد ؛ این متغیّرها می توانند از موارد زیر باشند :

- بولین (درست و غلط) ؛ مثل : Hungry(Homer) و isRaining

- گسسته ؛ مقادیر از یک دامنه ی *قابل شمارش* گرفته می شوند ؛ مثل ، دما [۲] : < گرم [۳] ، خنک [۴] ، ملایم [۵] > و پیش بینی وضع هوا : < آفتابی [۶] ، ابری [۷] ، بارانی >

- پیوسته ؛ مقادیر ، می توانند از یک *فاصله (دامنه ی غیرقابل شمارش)* گرفته شوند ، نظیر : [۰و۱] ؛ مثل ، سرعت [۸] ، زمان ، مکان [۹]

در اینجا بیش تر تمرکز ما بر روی مورد گسسته خواهد بود .

پس ، یک متغیّر تصادفی ، یک تابع از فضای نمونه برای برای بعضی از محدوده ها می باشد ، مثال : real ها یا Boolean ها . مثل : Odd(1) = true

---

[۱] random values
[۲] temperature
[۳] hot
[۴] cool
[۵] mild
[۶] sunny
[۷] overcast
[۸] velocity
[۹] position

رویدادها (رخدادها) ی تجزیه ناپذیر یا اتمیک [1]

ما می توانیم گزاره ها را با استفاده از رابط های منطقی استاندارد و به کارگیری اجتماع و اشتراک با هم ترکیب نماییم :

$P(Hungry(Homer) \wedge \neg Study(Bart))$

$P(Brother(Lisa,Bart) \vee Sister(Lisa,Bart))$

یک عبارت که یک مقدار ممکن را برای هر متغیّر نامعین مشخص می کند یک *رویداد اتمیک* [2] نام دارد ؛ رویدادهای اتمیک دو به دو ناسازگار [3] (مستقل از هم) هستند . مجموعه ای از همه ی رویدادهای اتمیک ، شامل تمام جزییات [4] هستند . یک رویداد اتمیک ، درستی یا نادرستی هر گزاره را پیش بینی می کند . توجه کنید که رویدادهای اتمیک در تشخیص درستی موارد دارای چند متغیّر نامعین ، مفید می باشند .

نکته : منطق فازی [5] از درجه ی قطعیّت (درجه ی درستی) [6] استفاده می کند نه نامعلومی [7] .

مثال : $P(Cavity$[1]$=true) = 0.1$ و $P(Weather=sunny) = 0.72$ نمونه هایی از احتمال غیرشرطی هستند . توزیع احتمال ، مقادیر را برای همه ی انتساب های ممکن ، ارایه می دهد :

---

[1] Atomic Events

[2] atomic event

[3] mutually exclusive

[4] exhaustive

[5] Fuzzy logic : یک شکل از منطق ریاضی است که در آن مقادیر بین صفر و یک دارای ارزش درست هستند . (WordNet 2.0)

[6] degree of truth

[7] uncertainty

$$P(Weather) = <0.72,0.1,0.08,0.1>$$

توجّه نمایید که مجموع اعداد برابر یک می باشد (۰.۷۲+۰.۱+۰.۰۸+۰.۱=۱.۰۰) . توزیع پیوسته ی احتمال برای مجموعه ای از متغیّرهای تصادفی ، احتمال هر رویداد اتمیک را در متغیّرهای تصادفی ارایه می کند . به عنوان مثال ، P(Weather,Cavity) برابر است با یک ماتریس ۲*۴ از مقادیر :

Weather =	sunny	rain	cloudy	snow
Cavity = true	0.144	0.02	0.016	0.02
Cavity = false	0.576	0.08	0.064	0.08

هر پرسش درباره ی یک دامنه ، می تواند به وسیله ی توزیع پیوسته ، پاسخ داده شود ، زیرا هر رویداد ، مجموعه ای از فضای نمونه می باشد .

اگر داشته باشیم ، P(Rain)=0.1 ، P(Overcast[2])=0.4 ، P(Sunny)=0.5 ، در این صورت ، می توانیم احتمالات ترکیب های متغیّرها را لیست نماییم :

$P(Rain \wedge Humid) = 0.1$ ، $P(Rain \wedge \neg Humid) = 0.1$ ، $P(Overcast \wedge Humid) = 0.2$ ،
$P(Overcast \wedge \neg Humid) = 0.2$ ، $P(Sunny \wedge Humid) = 0.15$ ، $P(Sunny \wedge \neg Humid) = 0.25$

این ، یک *توزیع احتمال توام* [3] نام دارد . برای متغیّرهای پیوسته ، ما نمی توانیم مقادیر را بشماریم ؛ در عوض ، ما از یک تابع پارامتری به نام توزیع نرمال [1] استفاده می نماییم :

---

[1] کرم خوردگی دندان

[2] ابری

[3] joint probability distribution

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{-x^2}{2}} dz$$ (توزیع نرمال)

تصمیم گیری تحت شرایط نامعلومی – در مثال رفتن به فرودگاه که قبلا در مورد آن صحبت کردیم ، تصور نمایید موارد زیر را داریم :

P(مرا به موقع برساند A_{25})= ۰.۰۴

P(مرا به موقع برساند A_{90})= ۰.۷۴

P(مرا به موقع برساند A_{120})= ۰.۹۵

P(مرا به موقع برساند A_{1440})= ۰.۹۹۹

کدام مورد فوق را باید انتخاب نماییم ؟ انتخاب ، وابسته به صلاح دید های ما ، برای از دست ندادن پرواز ، وضعیت فرودگاه و ..... می باشد .

قضیّه ی سودمندی [۲] ، برای ارایه و استدلال در مورد صلاح دیدها استفاده می شود .

قضیّه ی تصمیم گیری [۳] برابر است با ، قضیّه ی سودمندی + قضیّه ی احتمال .

مبانی احتمال – با یک مجموعه از فضای ساده ی Ω شروع می کنیم ؛ مثال ، شش حالت ممکن یک تاس . اگر $\omega \in \Omega$ باشد .

---

[۱] normal distribution

[۲] Utility theory

[۳] Decision theory

در یک بار پرتاب یک تاس ، احتمال آمدن هر کدام از اعداد ۱ تا ۶ روی شش وجه تاس ، بین ۰ تا ۱ است ؛ پس داریم : $0 \le P(\omega) \le 1$. در ضمن مجموع احتمال های آمدن هر وجه تاس ، برابر ۱ است ؛ یعنی ، $\Sigma_{\omega} P(\omega) = 1$.

مثال : $P(1)=P(2)=P(3)=P(4)=P(5)=P(6)=1/6$ ؛ می بینید که احتمال آمدن هر وجه تاس برابر با $\frac{1}{6}$ می باشد .

یک رویداد A زیرمجموعه ای از Ω می باشد :

$$P(A) = \sum_{\{\omega \in A\}} P(\omega)$$

مثلا ، احتمال این که عددی که در یک بار پرتاب یک تاس ، ظاهر می شود کم تر از ۴ باشد ، برابر است با : احتمال ۱ آمدن + احتمال ۲ آمدن + احتمال ۳ آمدن = ۱/۶+۱/۶+۱/۶ = ۱/۲

استنتاج با توزیع های پیوسته ی احتمال

آسان ترین راه انجام استنتاج احتمالی این است که یک جدول ارایه کننده ی توزیع پیوسته ی احتمال را داشته باشیم . به هر متغیّر مستقل نگاه کنید و احتمال وابسته را پیدا نمایید .

	Humidity[1]=High	Humidity=High	Humidity=Normal	Humidity=Normal

[1] رطوبت

	Outlook[1]=Overcast	Outlook=Sunny	Outlook=Overcast	Outlook=Sunny
Rain	0.1	0.05	0.15	0.05
¬ Rain	0.2	0.15	0.1	0.2

می توانیم توزیع پیوسته ی احتمال را برای تشخیص *احتمال نهایی*[2] متغیّر وابسته ، با جمع کردن همه ی موارد وابسته به متغیّر که می توانند درست باشند به دست آوریم :

P(Rain)=0.1+0.05+0.15+0.05=0.35

اگر **P** ، یک **توزیع تصادفی** را برای هر متغیّر تصادفی نظیر **X** ، بیان نماید ؛ داریم :

$$P(X = x_i) = \sum_{\{\omega : X(\omega)=x_i\}} P(\omega)$$

مثال : در پرتاب یکبار یک تاس ، احتمال این که عدد فرد ، ظاهر شود برابر است با :

P(Odd=true)=P(1)+P(3)+P(5)=1/6+1/6+1/6=1/2

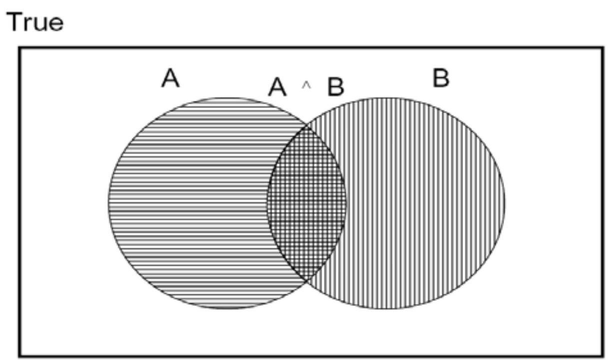
گزاره ها – یک گزاره را به صورت رویداد (مجموعه ای از فضاهای نمونه) در جایی که گزاره صحیح می باشد ، تصور نمایید . برای متغیّرهای تصادفی بولین A و B ، اگر رویداد a ، مجموعه ای از فضاهای نمونه ای به شرط آن که A(ω)=true باشد ، آن گاه رویداد ¬a ، مجموعه ای از فضاهای نمونه ای به شرط آن که A(ω) = false خواهد بود و رویداد a∧b = مواردی هستند که A(ω) و B(ω) [هر دو] ، True می باشند .

[1] پیش بینی
[2] marginal probability

اغلب در هوش مصنوعی ، *فضای نمونه ای ، به وسیله ی مقادیر یک مجموعه از متغیّرهای تصادفی تعریف می شود* . توجه نمایید که فضای نمونه ای حاصلضرب کارتزین محدوده هایی از متغیّرها می باشد . با متغیّرهای بولین ، فضای نمونه با مدل منطق گزاره ای ، برابر می باشد .

مثال :

$(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b) \Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$



چرا از احتمال استفاده می کنیم ؟ تعریف ، بر این دلالت می کند که رویدادهای منطقی وابسته ی معین ، باید روابط احتمالاتی داشته باشند .

مثال : $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

یک فرمول :

$P(A \vee \neg A) = P(A)+P(\neg A)-P(A \wedge \neg A) \Rightarrow P(True) = P(A)+P(\neg A)-P(False) \Rightarrow 1 = P(A) + P(\neg A) \Rightarrow \mathbf{P(A) = 1 - P(\neg A)}$

ظاهر گزاره ها

متغیّرهای تصادفی گزاره ای یا بولین : مثلا اگر Cavity برای آیا من کرم خوردگی دندان دارم ؟ ، باشد ؛ آن گاه ، Cavity = true یک گزاره می باشد .

متغیّرهای تصادفی گسسته (محدود یا نامحدود)

مثال : هوا (وضعیت جوی) [1] ، دارای یکی از حالت های آفتابی ، بارانی ، ابری و برفی [2] می باشد . هوا = بارانی ، یک گزاره می باشد .

متغیّرهای تصادفی پیوسته [3] ، محدوده دار [4] یا فاقد محدوده [5] می باشند . مثال : دما = ۲۱.۶ ، یا Temp<22.0 .

احتمال شرطی [6]

قبلا دیدیم که اگر هوا ابری باشد ، احتمال بارندگی بالا می رود ؛ که این یک احتمال شرطی نام دارد و به صورت P(Rain|Cloudy) نوشته می شود . برای توزیع های شرطی توجه نمایید که اگر P(Cavity|Toothache) مورد نظر باشد و ما بیش تر بدانیم ، مثلا cavity ارایه شده باشد ، در این صورت داریم : P(cavity|toothache,cavity) = 1

تعریف احتمال شرطی

---

[1] Weather

[2] snow

[3] Continuous random variables

[4] bounded

[5] unbounded

[6] conditional probability

$P(a|b)=\frac{P(a\wedge b)}{P(b)}$ ، با شرط $P(b)\neq 0$

با استفاده از فرمول بالا ، قانون ضرب [1] را به صورت زیر خواهیم داشت :

$$P(a\wedge b)=P(a|b)P(b)=P(b|a)P(a)$$

مثال :

$$P(Weather,Cavity)=P(Weather|Cavity)P(Cavity)$$

مثال : اگر $P(Rain \wedge Cloudy) = 0.15$ و $P(Cloudy) = 0.3$ باشد ؛ $P(Rain \mid Cloudy)$ چه قدر است ؟

$$P(Rain|Cloudy)=\frac{P(Rain\wedge Cloudy)}{P(Cloudy)}=\frac{0.15}{0.3}=0.5$$

قانون زنجیره ای [2] - برای کاربرد موفّقیّت آمیز قانون ضرب به وجود آمده است :

$$P(X_1,...,X_n)=P(X_1,...,X_{n-1})P(X_n \mid X_1,...,X_{n-1})$$
$$=P(X_1,...,X_{n-2})P(X_{n_1} \mid X_1,...,X_{n-2})P(X_n \mid X_1,...,X_{n-1})$$
$$=...=\prod_{i=1}^{n}P(X_i \mid X_1,...,X_{i-1})$$

توزیع پیوسته [3]

---

[1] Product rule

[2] Chain rule

[3] joint distribution

برای k متغیّر تصادفی $X_k, \ldots, X_1$ ، توزیع پیوسته ی این متغیّرها ، جدولی است که در آن هر عـدد ، احتمال یک ترکیب از مقدارهای $X_k, \ldots, X_1$ را ارایه می نماید .

مثال :

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

$P(\neg Cavity \wedge Toothache)$ $P(Cavity \wedge \neg Toothache)$

با توجه به جدول بالا داریم :

$$P(Toothache) = P((Toothache \wedge Cavity) \vee (Toothache \wedge \neg Cavity))$$

$$= P(Toothache \wedge Cavity) + P(Toothache \wedge \neg Cavity) = 0.04 + 0.01 = 0.05$$

$$P(Toothache \vee Cavity) = P((Toothache \wedge Cavity) \vee (Toothache \wedge \neg Cavity) \vee (\neg Toothache \wedge Cavity)) = 0.04 + 0.01 + 0.06 = 0.11$$

می توانیم همچنین احتمالات شرطی را محاسبه نماییم :

$$P(Cavity|Toothache) = P(Cavity \wedge Toothache) / P(Toothache)$$

$$P(Cavity \wedge Toothache) = ?$$

با توجه به جدول ، برابر است با ۰.۰۴

$$P(Toothache) = ?$$

حاصل این احتمال را هم که قبلا محاسبه کردیم ، برابر بود با : ۰.۰۵ ؛ درنتیجه داریم :

$$P(Cavity|Toothache) = 0.04/0.05 = 0.8$$

نرمال سازی

در فرمول احتمال شرطی ، مقسوم علیه [1] می تواند به صورت یک ثابت نرمال سازی α دیده شود :

P(Cavity|Toothache) = P(Cavity $\wedge$ Toothache) / P(Toothache) = α [P(Cavity $\wedge$ Toothache)] = $\alpha \times 0.04 = \frac{1}{0.05} \times 0.04 = 20 \times \frac{4}{100} = 0.8$

تعمیم

$$P(A \wedge B \wedge C) = P(A|B,C)\, P(B|C)\, P(C)$$

استدلال با شمارش

تصور نمایید X ، همه ی متغیّرها باشد . معمولاً ، ما توزیع پیوسته ی **متغیّرهای پرس و جوی**[2] **Y** را می خواهیم که توسط مقادیر e برای **متغیّرهای ثابت**[3] **E** ارایه شده اند . تصور نمایید متغیّرهـای مخفـی H=X-Y-E باشند . در این صورت ، مجموع تمام پیوند ها با جمع کردن متغیّرهای مخفی به دست می آید :

$$P(Y \mid E=e) = \alpha P(Y, E=e) = \alpha \sum_h P(Y, E=e, H=h)$$

واژگان درون مجموعه ، تماما متصل می باشند زیرا Y و E و H با هـم ، خروجـی آن هـا ، مجموعـه ای از متغیّرهای تصادفی می باشد .

مسایل آشکار :

۱) پیچیدگی زمانی در بدترین حالت در جایی که d بـزرگ تـرین مـی باشـد برابر است با $O(d^n)$

---

[1] denominator
[2] query variables
[3] evidence variables

۲) پیچیدگی فضا برای نگهداری توزیع پیوسته برابر است با $O(d^n)$

۳) چگونه می توان تعداد را برای ورودی های $O(d^n)$ پیدا نماییم ؟؟؟

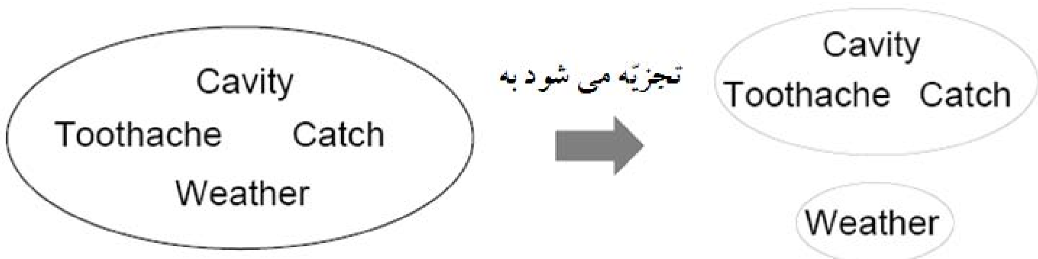
استقلال [۱]

در برخی از موارد ، ما می توانیم چیزها را با توجه به این که یک متغیّر دارای هیچ اثری بر روی دیگری نمی باشد ساده نماییم . برای مثال ، اگر ما یک متغیّر چهارم DayOfWeek را به محاسبه ی بارندگی اضافه نماییم چه تغییری به وجود می آید ؟ ؛ از آنجایی که روز هفته بر احتمال باران تاثیری نخواهد داشت ، ما داریم :

$$P(Rain|Cloudy,Monday)=P(Rain|Cloudy,Tuesday)\ldots=P(Rain|Cloudy)$$

ما گفتیم که DayOfWeek و Rain مستقل می باشند . بنابراین می توانیم توزیع پیوسته ی احتمال بزرگ تر را به زیر جدول هایی مجزا تقسیم نماییم ؛ استقلال به ما کمک خواهد کرد که دامنه را به تکه های مجزا تقسیم نماییم .

تعریف – A و B در صورتی مستقل هستند که ، $P(A|B)=P(A)$ یا $P(B|A)=P(B)$ یا $P(A,B)=P(A)P(B)$. با توجه به این تعریف برای مثال دندان درد داریم :



---

[۱] Independence

P(Toothache,Catch,Cavity,Weather) =
P(Toothache,Catch,Cavity)P(Weather)

۳۲ ورودی به ۱۲ تا کاهش داده می شود ؛ برای n سکه ی مستقل داریم : $n \rightarrow ۲^n$

استقلال مطلق ، خیلی مفید است ولی نادر یا کمیاب [1] می باشد . مثلا ، دندانپزشکی ، زمینه ی بزرگی با صد ها متغیّر می باشد که هیچکدام مستقل نیستند . در این مورد چه کاری انجام بدهیم ؟

استقلال شرطی [2]

اگر A و B مستقل شرطی باشند و C هم داده شده باشد ، داریم :

P(A,B|C) = P(A|C)P(B|C)

P(A|C,B) = P(A|C)

در بیش تر موارد ، استفاده از استقلال شرطی ، اندازه ی ارایه ی توزیع پیوسته را از حالت نمایی n به حالت خطی n کاهش می دهد . استقلال شرطی ، پایه ای ترین و نیرومندترین دانش درباره ی محیط های نامعین می باشد .

مثال اتومبیل – سه پارامتر گاز (Gas) ، باتری (Battery) و استارت (Start) را برای اتومبیل در نظر بگیرید ؛ داریم :

P(Battery|Gas) = P(Battery)

چون که گاز و باتری ، مستقل از هم هستند . ولی

P(Battery|Gas,Starts) ≠ P(Battery|Starts)

---

[1] rare
[2] conditional independence

چون گاز و باتری با وجود استارت ، مستقل از هم نمی باشند .

قانون بیز

با استفاده از قـانون ضـرب داریـم : $P(a \wedge b) = P(a \mid b)P(b)$ و $P(a \wedge b) = P(b \mid a)P(a)$ ، در نتیجه ما می توانیم این تساوی ها را برابر هم قرار دهیم :

$P(a \wedge b) = P(a \mid b)P(b) = P(b \mid a)P(a)$ ، در نتیجه قانون بیز را به صورت زیر خواهیم داشت :

$$P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)}$$

یا به شکل نرمال داریم :

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} = \alpha P(X \mid Y)P(Y)$$

مثال برای قانون بیز – این مفروضات را در نظر بگیریـد : مننژیـت [1] باعـث خشـکی گـردن [1] در ۵۰٪ از بیماران می شـود ؛ P(stiffNeck|Meningities)=0.5 ؛ احتمـال مننژیـت قـدیمی برابـر اسـت بـا ۱/۵۰۰۰۰ ؛ P(meningitis)=0.00002 ؛ احتمــال خشــکی گــردن قــدیمی برابــر اســت بــا ۱/۲۰ و P(stiffNeck)=0.05 . حال یک بیمار با خشکی گردن مراجعـه مـی نمایـد . احتمـا ل ایـن کـه وی دارای مننژیت باشد چقدر است ؟

$$P(meningitis \mid stiffNeck) = \frac{P(stiffNeck \mid meningitis)P(meningitis)}{P(stiffNeck)} = \frac{0.5 \times 0.00002}{0.05} = 0.0002$$

---

[1] meningitis

[1] stiff neck

مثال : جدول زیر را در نظر بگیرید ،

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

اگر داشته باشیم : P(Cavity)=0.1 ، P(Toothache)=0.05 و P(Cavity|Toothache)=0.8 ؛ در این صورت P(Toothache|Cavity) را به دست آورید .

$$P(Toothache|Cavity) = (0.8 \times 0.05) / 0.1 = 0.4$$

تعمیم

$$P(A \wedge B \wedge C) = P(A \wedge B|C)\ P(C) = P(A|B,C)\ P(B|C)\ P(C)$$

$$P(A \wedge B \wedge C) = P(A \wedge B|C)\ P(C) = P(B|A,C)\ P(A|C)\ P(C)$$

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

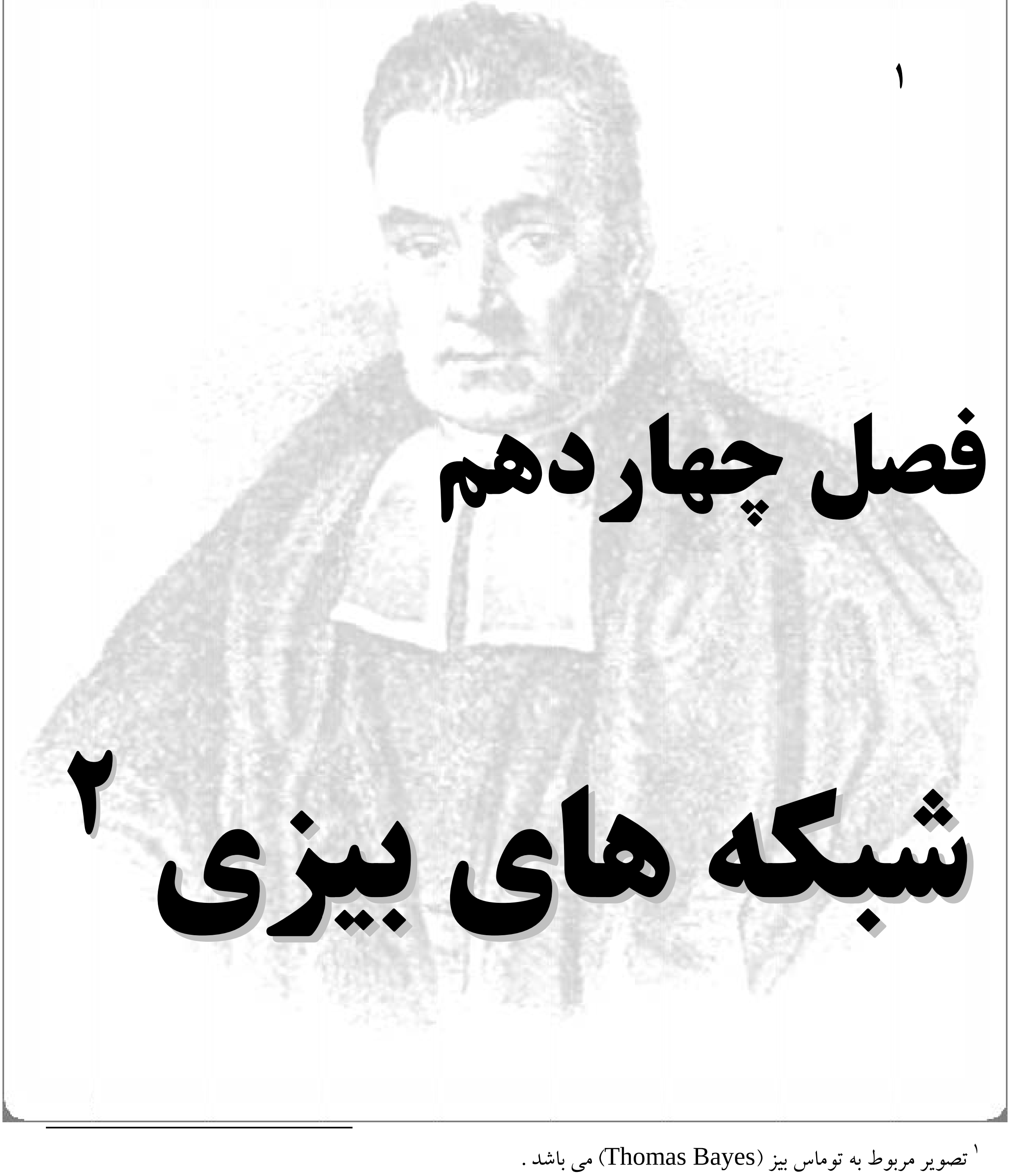
مثال : مدل ساده ی بیز [1]

[1] Naïve Bayes Model

در یک مدل قدیمی برای تشخیص بیماری ؛ علایمی که به صورت مستقل شرطی هستند برای بیماری داده شده اند ، بنابراین ، اگر $X_1,\ldots,X_n$ علایمی باشند که مریض گفته ؛ مثل ، سردرد ، تب بالا و غیره و H ، فرضیه هایی برای سلامتی بیمار باشد ، در این صورت ، داریم :

$$P(X_1,\ldots,X_n,H) = P(H)P(X_1|H)\ldots P(X_n|H)$$

این مدل ساده ی بیزی ، اجازه ی ارایه های فشرده را می دهد و فرضیه های مستقل نیرومندی دارد .



[1]

فصل چهاردهم

شبکه های بیزی [2]

[1] تصویر مربوط به توماس بیز (Thomas Bayes) می باشد .

[2] Bayesian networks

ریوس مطالب

- **ساختار**
- **معنا**
- **توزیعات پارامتربندی شده [1]**

استنتاج احتمالی

در گذشته ، ما در مورد سیستم های برمبنای قانونی که می توانند استنتاج های منطقی را انجام دهند صحبت کردیم . مثلا اگر کبوتر خانگی [2] ، گرسنه شود به بازار می رود :

---

[1] parametrized distributions

[2] Homer

$$Hungry(Homer) \Rightarrow GoesTo(Homer,Quickie\text{-}mart)$$

ما تمایل داریم این نوع از عملکرد را به جهان هایی با نامعلومی توسعه دهیم . مثلا :

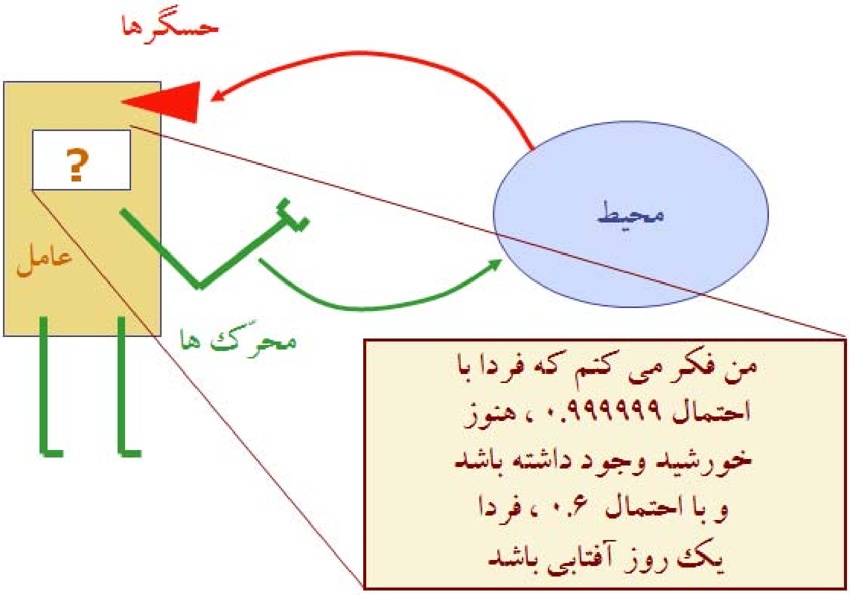
$$P(Hungry(Homer))=0.98$$

$$P(GoesTo(Homer,Quickie\text{-}mart)|Hungry(Homer))=0.5$$

$$P(GoesTo(Homer,Quickie\text{-}mart))=?$$

در کل ، ما می توانیم از قانون بیز برای این کار استفاده نماییم ، در این مورد ، مشکل در کار کردن با توزیع پیوسته ی احتمال می باشد ، چون دارای جدولی بزرگ به صورت نمایی می باشد . ما به ساختمان داده ای که در آن تعدادی متغیّر بر هم تاثیر نمی گذارند نیازمندیم . به عنوان مثال ، رنگ کلاه بارت بر این که Homer گرسنه می باشد تاثیری نمی گذارد ؛ ما این ساختار را یک *شبکه ی بیزی* می نامیم .

عامل احتمالی



مسأله

در زمان مشخص t ، پایگاه دانش (KB) یک عامل ، مجموعه ای از حدس ها است ؛ در زمان t ، حسگرهای عامل ، یک دید را به وجود می آورند که احتمال یکی از حدس ها را افزایش می دهد ؛ ***حال سوال این است که در این موقع عامل باید چگونه احتمال موارد دیگر را به روز نماید ؟***

هدف شبکه های بیزی

توصیف یک مجموعه از حدس ها را با به وجود آوردن ارتباط های صریح میان حدس ها و پیدا کردن استقلال شرطی میان حدس ها آسان می کند و یک روش مناسب تر از استفاده از جدول های توزیع پیوسته را برای به روز کردن قدرت حدس ها در زمانی که شواهد جدید مشاهده می شوند ، به وجود می آورد .

نام های دیگر شبکه های بیزی

نام های دیگر شبکه های بیزی عبارتند از : شبکه های حدسی [1] ، شبکه های احتمالی [2] و شبکه های بیان کننده ی علت [3] .

شبکه های بیزی

یک روش ساده و گرافیکی برای ادعاهای مستقل شرطی در یک نمایش فشرده برای توزیع پیوسته ی کامل است . به عبارت دیگر ، یک شبکه ی بیزی ، یک گراف مستقیم است که در آن هر گره با اطّلاعات احتمالی ، تفسیر می شود . یک شبکه دارای موارد زیر است :

---

[1] belief networks
[2] probabilistic networks
[3] causal networks

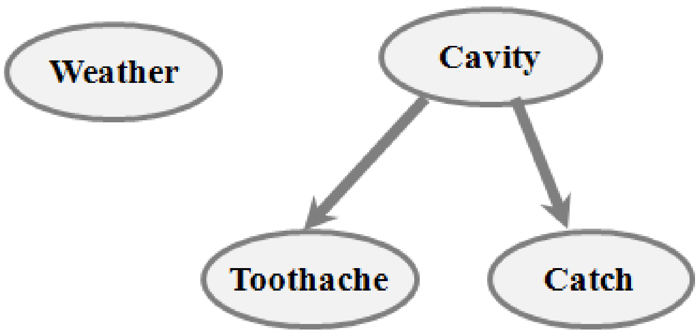
- یک مجموعه از متغیّرهای تصادفی که با گره های درون شبکه برابر است .
- یک مجموعه از یال ها (لبه ها یا پیکان ها) که گره ها را متصل می کنند . یال ها ، تاثیرات را نمایش می دهند . اگر پیکانی از X به طرف Y وجود داشته باشد ، آن گاه X ، پدر Y می باشد .

هر گره یک توزیع شرطی احتمال را که نشان دهنده ی احتمال هر مقداری که گره می تواند بگیرد و شرطی که مقادیر والدهایش می توانند بگیرند را نگهداری می کند . در این شبکه ، هیچ حلقه ای وجود ندارد ؛ به عبارت دیگر ، این شبکه به صورت یک گراف مستقیم بدون دور می باشد .

پس ، یک شبکه ی بیزی ، یک نمایش ساده و گرافیکی برای موارد مستقل شرطی می باشد .

ساختار شبکه ی بیزی : در این ساختار هر متغیّر دارای یک گره است و شکل آن به صورت یک گراف غیرحلقه ای مستقیم می باشد . در ضمن ، یک توزیع شرطی برای هر گره ی ارایه شده توسط والدهای خودش به صورت $P(X_i \mid Parents(X_i))$ می باشد . در ساده ترین حالت ، توزیع شرطی ارایه شده به صورت یک جدول احتمال شرطی [1] ، توزیع را برمبنای X_i ، برای هر ترکیب از مقادیر والد ارایه می دهد .

مثال – توپولوژی شبکه ، موارد استقلال شرطی را بیان می کند :



[1] conditional probability table (CPT)

هوا مستقل از دیگر متغیّرها می باشد . دندان درد و دچار شدن به آن به صورت شرطی مستقل از پوسیدگی دندان می باشند .

مثال دزدی [1] – من سرکار هستم و همسایه ام ، جان با من تماس می گیرد و می گوید آژیر خانه ی من فعال است ، اما همسایه ی دیگرم ، مری [2] این کار را انجام نمی دهد . بعضی از وقت ها آژیر با کم ترین لرزه ای فعال می شود . آیا در خانه ی من واقعا دزد است ؟

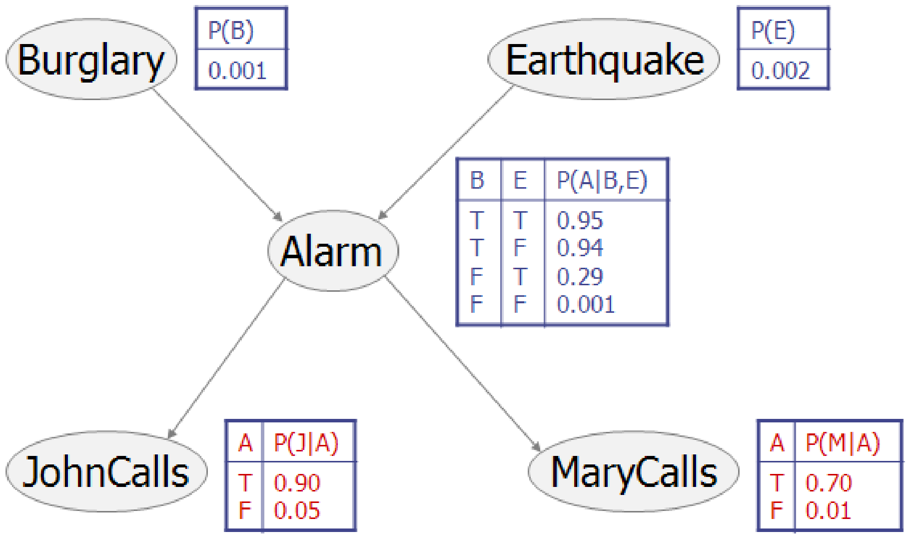
متغیّرها در این مثال عبارتند از : دزد ، لرزه [3] ، آژیر [4] ، تماس های جان [5] ، تماس های مری [6] .

توپولوژی شبکه آن چه که اتفاق افتاده است را برگشت می دهد (منعکس می کند) :

- یک دزد می تواند آژیر را در وضعیت خاموش قرار دهد .
- یک لرزه ی زمین می تواند آژیر را خاموش نماید .
- آژیر می تواند سبب تماس مری شود .
- آژیر می تواند سبب تماس جان شود .

---

[1] burglar
[2] Mary
[3] earthquake
[4] alarm
[5] JohnCalls
[6] MaryCalls



تراکم [1]

شبکه های بیزی ، ارایه ی شبکه های بزرگ را امکان پذیر می سازند ؛ در ایـن شـبکه هـا ، اطّلاعـات اضافی (تکراری) [2] برداشته می شود .

یک CPT برای X_i بولین با والدهای بولین k دارای ۲^k سطر برای ترکیب مقدارهای والد می باشد . هر سطر یک عدد p را برای $X_i = true$ نیاز دارد (عدد مورد نیاز برای $X_i = false$ فقط 1-p می باشـد) . در صورتی که هر متغیّر دارای بیش از k والد نباشد ، شبکه ی کامل به تعداد $O(n.2^k)$ سطر نیاز دارد . توجه کنید که [تراکم] به صورت خطی ، با n رشد می کند و دارای مرتبه ی $O(2^n)$ برای توزیع با اتصال کامل می باشد .

معانی عمومی

---

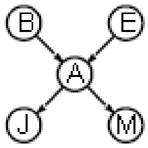
[1] compactness

[2] redundant

معانی کلی ، توزیع پیوسته ی کامل را به صورت ضرب توزیعات شرطی محلی ، تعریف می نماید :

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

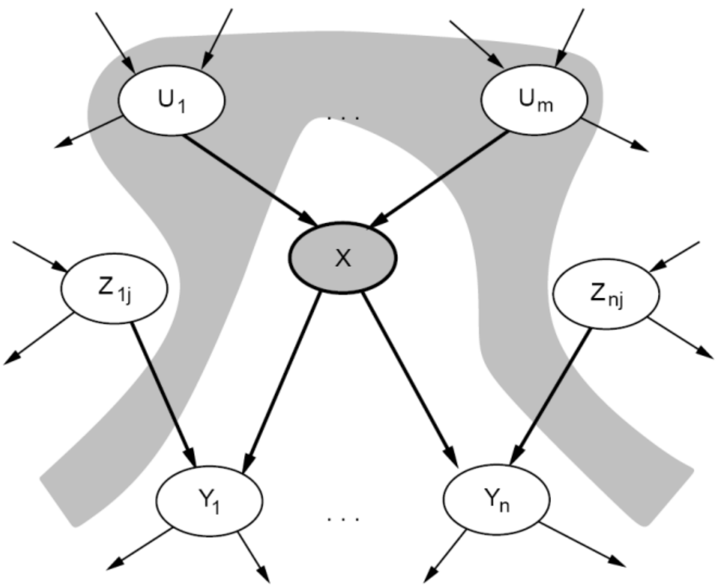
مثال :



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

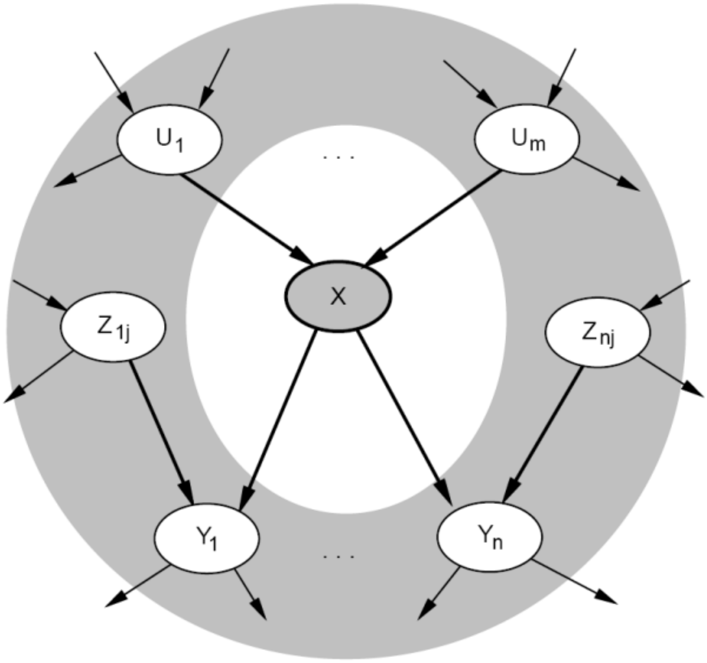
معانی محلّی

هر گره ، به صورت شرطی مستقل از فرزندانی است که توسط پدرشان به وجود آمده اند .



قضیّه : معانی محلی ⇔ معانی عمومی

پوسته ی مارکف [1] – هر گره به صورت شرطی ، مستقل از همه ی گره های دیگری است که توسط پوسته ی مارکوف ارایه می شوند ؛ پوسته ی مارکوف تشکیل شده از : والدین + فرزندان + والدین فرزندان



ساختار شبکه

هر گره دارای یک جدول احتما ل شرطی است که این جدول ، احتمال هر مقدار آن گره و مقادیر والدهای آن را ارایه می کند .

بیان نامعلومی (عدم قطعیّت) به طورخلاصه

---

[1] Markov blanket

توجه نمایید که ما نیازی به داشتن گره هایی برای همه ی مواردی که مری ممکن است تماس نگیـرد نداریم . یک روش احتمالی به ما اجازه می دهد که این اطّلاعـات را بـه صـورت $\neg M$ خلاصـه نمـاییم . ایـن روش ، به یک عامل ساده اجازه می دهد که به جهان های بزرگـی کـه دارای تعـداد زیـادی از نتـایج نـامعلوم ممکن هستند رسیدگی کند .

ارایه ی ضمنی توزیع پیوسته ی کامل

به یاد بیاورید که توزیع پیوسته ی کامل به ما اجازه می داد که احتمال هـر متغیّـر و همـه ی آن هـایی که ارایه شده اند را محاسبه نماییم . رویدادهای مستقل می توانند به جـدول هـای مسـتقل تقسـیم شـوند و ایـن رویدادها ، CPT هایی هستند که در شبکه ی بیزی دیده می شوند . بنـابراین مـا مـی تـوانیم از ایـن اطّلاعـات برای انجام محاسبات استفاده نماییم . داریم :

$$P(x_1, x_2, \ldots, x_n) = \prod P(x_i \mid parents(x_i))$$

مثلا ؛

$$P(A \wedge \neg E \wedge \neg B \wedge J \wedge M) = P(J \mid A) P(M \mid A) P(A \mid \neg B \wedge \neg E) P(\neg B) P(\neg E) =$$
$$0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.00062$$

ساخت شبکه های بیزی

اغلب ، چند روش برای ساخت یک شبکه ی بیزی وجود دارد . در این مورد ، مهندسی دانش به کشف ارتباطات *مستقل شرطی* نیازمند است . والدهای یک گره باید آن هایی باشند که به طور مستقیم بر مقدار آن برتری دارند ؛ مثلا ، JohnCalls توسط لرزه دارای برتری می شود ، اما نه به طور مستقیم و تماس جان و ماری دارای برتری بر یکدیگر نمی باشند . به صورت صریح ، ما فرض می کنیم :

P(MaryCalls|JohnCalls,Alarm,Earthquake,Burglary)=P(MaryCalls|Alarm)

در زیر ، روشی برای ساخت شبکه های بیزی آمده است :

۱) یک مجموعه ی منظم از متغیّرهای $X_1,...,X_n$ را انتخاب نمایید .

۲) برای i=1 تا n کارهای زیر را انجام بده

X_i را به شبکه اضافه کن . والدهایی از $X_1,...,X_{i-1}$ را انتخاب کن که

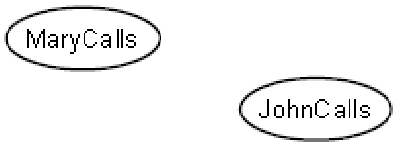
$$P(X_i \mid Parents(X_i)) = P(X_i \mid X_1,..., X_{i-1})$$

این نوع از انتخاب والد ها معانی عمومی زیر را به وجود می آورد :

$$P(X_1,..., X_n) = \prod_{i=1}^{n} P(X_i \mid X_1,..., X_{i-1})$$ (قانون زنجیری [۱])

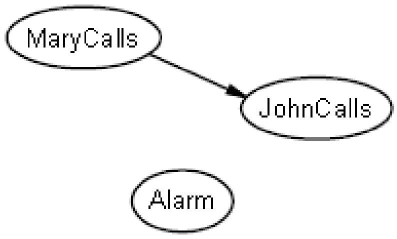
$$= \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

مثال – تصور کنید که ما مجموعه ی M,J,A,B,E را به ترتیب ، انتخاب می نماییم :

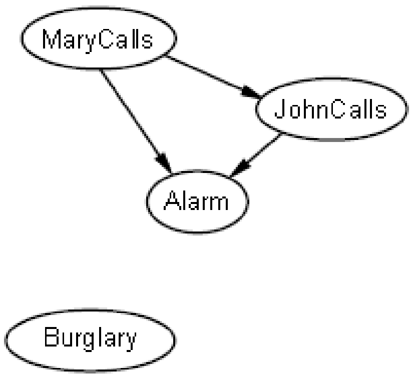


P(J|M)=P(J)? **نه**

[۱] chain rule

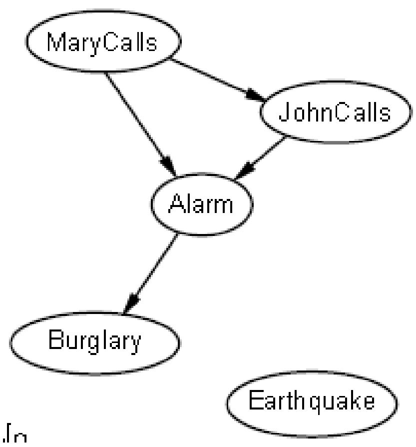


P(A|J,M)=P(A|J)? P(A|J,M)=P(A) ? نه



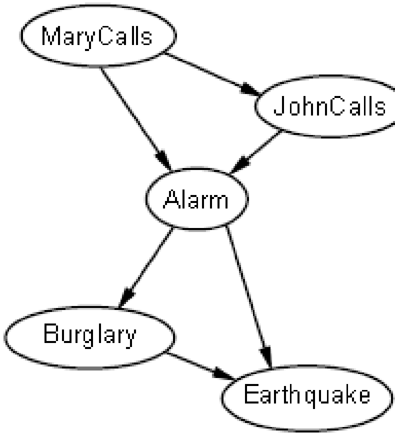
P(B|A,J,M)=P(B|A) ?بله

P(B|A,J,M)=P(B) ?نه



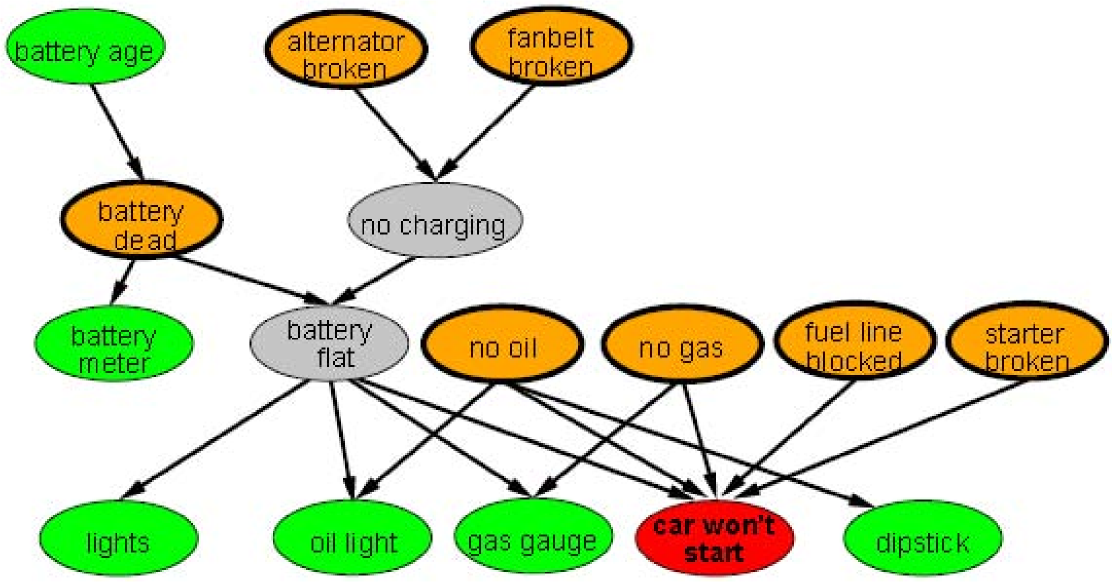
$P(E|B,A,J,M)=P(E|A)$? نه

$P(E|B,A,J,M)=P(E|A,B)$? بله

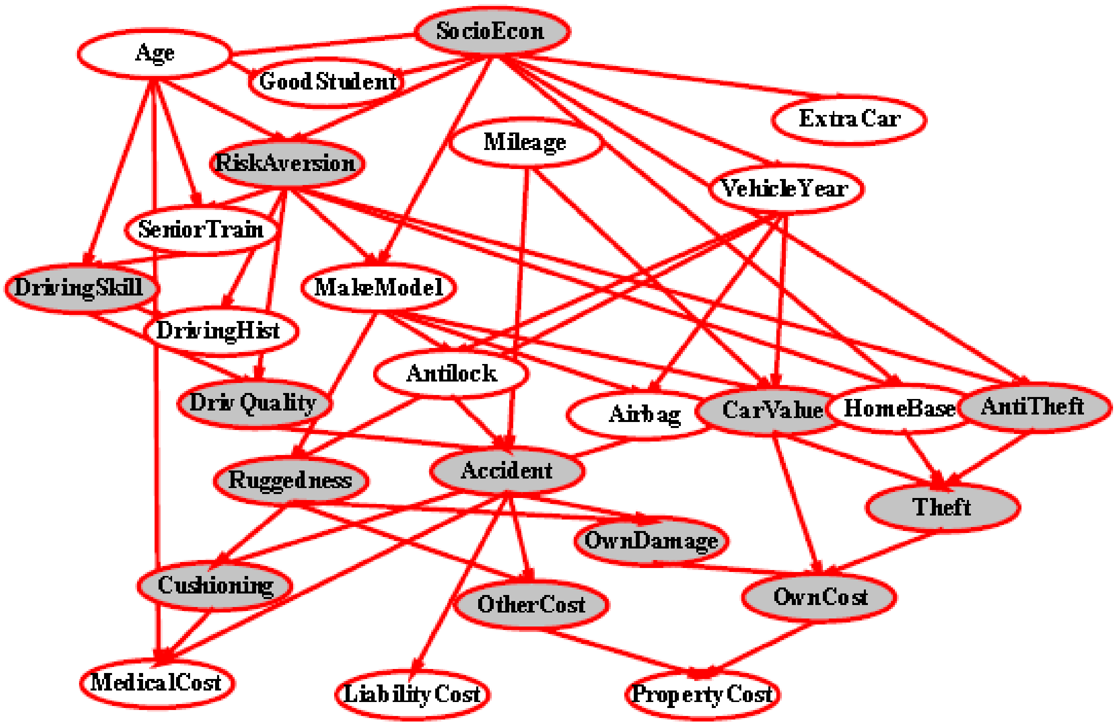


مثال : عیب اتومبیل

دلیل اوّلیه : ماشین روشن نمی شود . متغیّرهای قابل آزمایش در شکل زیر با رنگ سبز نشان داده شده اند ؛ متغیّرهایی که " شکسته اند و بنابراین آن ها را تعمیر نماییم " با رنگ نارنجی در شکل زیر نمایش داده شده اند ؛ برای متغیّر های مخفی (خاکستری) ، قطعات وابسته را چک نمایید و به این طریق پارامترها را کاهش دهید .



مثال : بیمه ی اتومبیل



ساخت یک شبکه

اوّل ، ریشه ها را به وجود بیاورید ؛ دوّم ، متغیّرهای مستقیم را اضافه نمایید ؛ سوّم ، فرزندان مستقیم آن ها را اضافه نمایید ؛ به این ترتیب ، شما یک *مدل سببی*[۱] را به وجود آورده اید . تخمین ها با این روش به مراتب آسان تر به دست می آیند . برای ساخت می توانیم از اثر به سبب هم تلاش نماییم ، امّا در این حالت ، شبکه به مراتب پیچیده تر خواهد بود .

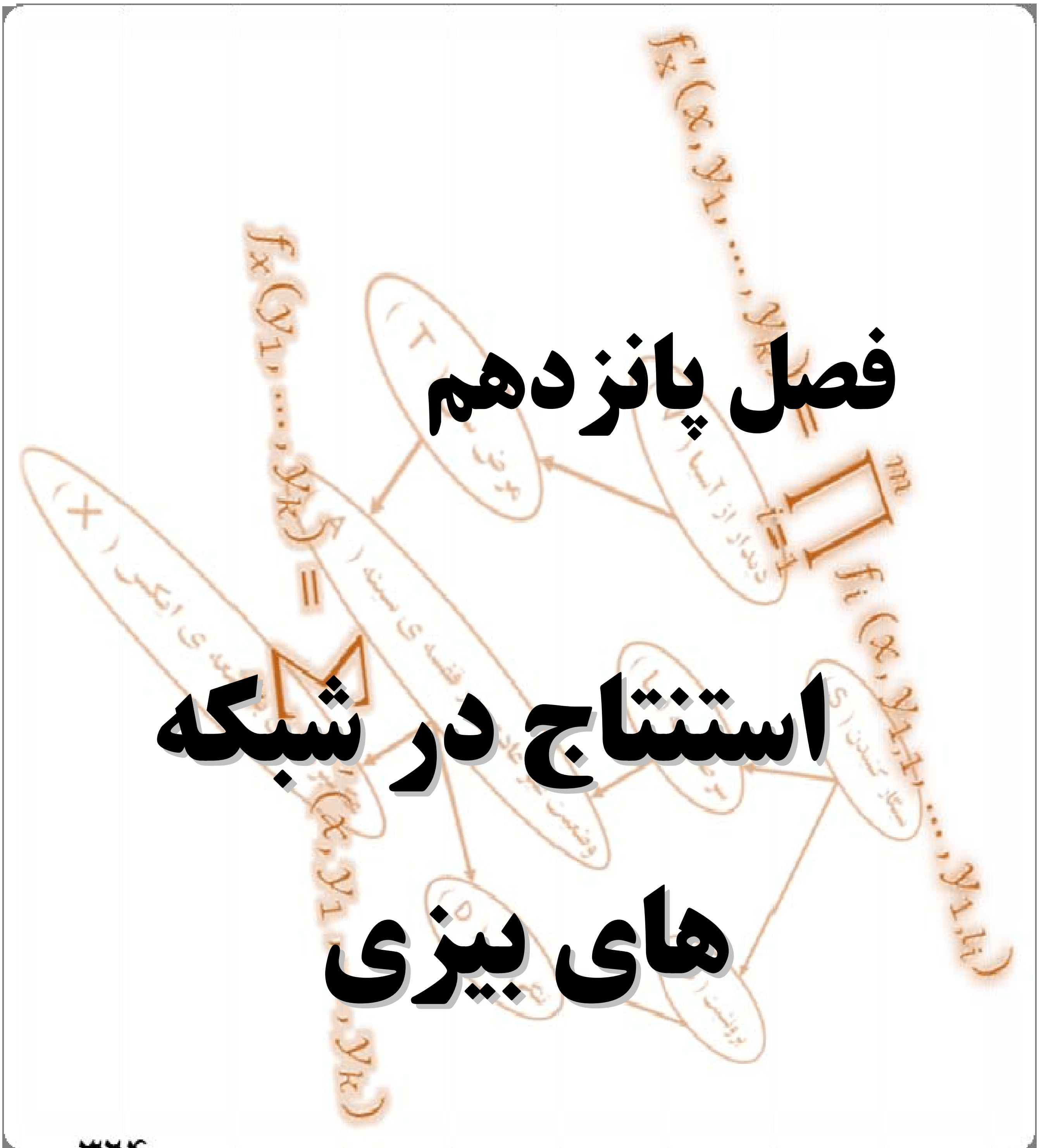
کاربردهای شبکه های بیزی

عبارتند از : تشخیص ، که به طور گسترده ای در محصولات میکروسافت استفاده می شود ؛ تشخیص طبی ؛ فیلتر نمودن نامه های الکترونیکی که ناشناس می باشند[۱] ؛ کاربرد در سیستم های خبره (کنترل وسایل ،مانیتورینگ) و کنترل روبوتیک .

---

[۱] causal model

[۱] spam filtering



فصل پانزدهم

استنتاج در شبکه های بیزی