آشنایی با شبکه‌های بیزی

بابک به‌ساز

۱- مقدمه

امروزه بسیاری از مشکلات انسان‌ها، با کمک هوش مصنوعی حل می‌شود. یکی از مهمترین خصوصیات این مشکلات وجود عدم قطعیت (uncertainty) در آنها است. روش‌های زیادی در هوش مصنوعی برای کنترل عدم قطعیت پیشنهاد شده‌اند که اکثر آنها بر پایه نظریه احتمالات و نظریه فازی بنا نهاده شده‌اند. در این مطالعه می‌خواهیم یک روش برای کنترل عدم قطعیت در مسائل بر پایه نظریه احتمالات به نام شبکه‌های بیزی (Bayesian Networks) را بررسی کنیم.

در سیستم‌های هوشمند بسیاری نیاز به جواب درخواست‌هایی (query) است که احتمال وقوع یک رویداد را براساس تعدادی از مشاهدات می‌خواهند. مثلاً در یک سیستم تصمیم‌یار دندان پزشک، احتمال خرابی دندان براساس مشاهداتی مانند دندان درد و رنگ بیرونی دندان مطلوب می‌باشد. به عنوان مثال دیگر، در سیستم‌های دسته‌بندی احتمال عضویت یک شی در هر یک از دسته‌ها براساس ویژگی‌های شی موردنظر است. به فرایند جواب دادن به درخواست‌ها، استنتاج (inference) می‌گویند و هر فرایند استنتاج نیاز به داده‌هایی در مورد قلمرویی (domain) دارد که قصد کنترل عدم قطعیت آن را داریم. در توضیحاتی که در ادامه خواهد آمد، متغیرهای تصادفی را با حرف اول بزرگ و مقادیر آنها را با حرف اول کوچک نشان می‌دهیم. همچنین، ماتریس‌ها و بردارها را با حروف پررنگ (bold) نشان می‌دهیم.

۱-۱ استنتاج با استفاده از توزیع توام کامل

ساده‌ترین روش استنتاج از توزیع توام کامل (full joint distribution) استفاده می‌کند که با یک مثال ساده آن را توضیح می‌دهیم. یک قلمرو را در نظر بگیرید که تنها شامل سه متغیر بولی Toothache، Cavity و Catch می‌باشد. Toothache نشان دهنده دندان درد، Cavity نشان دهنده سوراخ بودن دندان و Catch نشان دهنده گیر کردن ابزار دندان پزشک در دندان می‌باشد. توزیع توام کامل یک جدول ۲×۲×۲ است که در شکل ۱ دیده می‌شود.

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	**.108**	**.012**	**.072**	**.008**
$\neg cavity$	**.016**	**.064**	**.144**	**.576**

شکل ۱: جدول توزیع یک قلمرو ساده.

با استفاده از این جدول می‌توان احتمال هر رویداد دلخواه را محاسبه کرد. مثلاً احتمال رویداد $cavity \vee toothache$ برابر است با:

$$P(cavity \vee toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

همانطور که قبلاًگفته شد، در بسیاری موارد محاسبه احتمال شرطی یک رویداد در حالی که بعضی از مشاهدات داده شده است، مطلوب می‌باشد. احتمال شرطی رویدادها با استفاده از فرمول احتمال شرطی و جمع عناصری از جدول توزیع آنها قابل محاسبه است. به عنوان مثال، محاسبه احتمال یک سوراخ در دندان در صورت وجود دندان درد، بصورت زیر می‌باشد:

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

به عنوان مثال دیگر می‌توان احتمال سوراخ نبودن دندان با وجود دندان درد را نیز محاسبه کرد:

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

توجه کنید که در این محاسبات $1/P(toothache)$ بدون توجه به اینکه چه مقداری برای Cavity محاسبه کنیم، ثابت باقی می‌ماند. در واقع، می‌توان آن را یک عامل برای نرمال‌سازی فرض کرد تا توزیع $P(cavity|toothache)$ مجموعی برابر ۱ داشته باشد. در طول این بررسی از a برای نشان دادن چنین ثابت‌هایی استفاده می‌کنیم. با این نماد، دو عبارت قبلی را در یک عبارت بصورت زیر خلاصه کرد:

$$P(Cavity \mid toothache) = aP(Cavity, toothche) = a <0.12, 0.08> = <0.6, 0.4>$$

با تعمیم روش بکار رفته در این مثال به راحتی می‌توان به یک رویه استنتاج کلی دست یافت. در اینجا، حالتی که تنها یک متغیر در درخواست وجود دارد را بررسی می‌کنیم (حالت‌های دیگر بسادگی از تعمیم این حالت بدست می‌آیند). فرض کنید X متغیر درخواست (Cavity در مثال) باشد، $\mathbf{E}$ مجموعه متغیرهای مربوط به شواهد را نشان دهد (Toothache در مثال)، $\mathbf{e}$ مقادیر مشاهده شده برای $\mathbf{E}$ را نشان دهد و $\mathbf{Y}$ مجموعه بقیه متغیرهای مشاهده نشده باشد (Catch در مثال). با توجه به این فرضیات، درخواست‌ها در حالت کلی بصورت $P(X \mid \mathbf{e})$ هستند و می‌توان بصورت زیر آنها را محاسبه کرد:

$$P(X \mid \mathbf{e}) = aP(X, \mathbf{e}) = a \sum_{\mathbf{y}} P(X, \mathbf{e}, \mathbf{y}) \qquad (۱)$$

که مجموع بر روی تمامی مقادیر $\mathbf{y}$ ممکن است (یعنی تمامی حالات مقادیر متغیرهای $\mathbf{Y}$). توجه کنید که X و $\mathbf{E}$ و $\mathbf{Y}$ با هم تمامی متغیرهای قلمرو را تشکیل می‌دهند و $P(X, \mathbf{e}, \mathbf{y})$ یک زیرمجموعه از عناصر جدول توزیع می‌باشد و بنابراین، مجموع براحتی قابل محاسبه است.

۱-۲ مشکلات استنتاج با توزیع توام کامل و راه‌حل آنها

روش استنتاج با توزیع توام کامل سه مشکل اساسی دارد. اولاً، برای متغیرهای پیوسته اگر آنها یک توزیع خاص و مشخصی نداشته باشند، نیاز به یک جدول نامتناهی است. دوماً، حتی برای n متغیر بولی هم جدول توزیع 2^n خانه دارد و جواب به یک درخواست زمان $O(2^n)$ نیاز دارد. سوماً، احتمالات مربوط به توزیع را باید از داده‌های آماری تخمین زد که برای تعداد نمایی خانه در جدول نیاز به داده‌های بسیار زیادی است که در عمل موجود نمی‌باشد. برای حل این مشکلات سعی می‌شود که شکل ساده‌تری برای توزیع فرض شود تا بدست آوردن آن عملی باشد. یکی از فرض‌هایی که بسیار استفاده شده فرض مستقل بودن متغیرهای قلمرو بوده است که در عمل نتیجه‌های خوبی نیز از آن گرفته شده است.

این فرض بسیار محدود کننده است. بهمین دلیل سعی شد این محدودیت تا حدی برداشته شود و وابستگی متغیرها تا حدی در محاسبه توزیع در نظر گرفته شود. براین اساس فرض وابستگی درختی و وابستگی بصورت گراف‌های جهت‌دار بودن دور (Directed Acyclic Graph) مورد بررسی قرار گرفت. با فرض وابستگی بصورت گراف‌های جهت‌دار بودن دور به مدلی از توزیع می‌رسیم که به شبکه‌های بیزی معروف است.

مطالبی که در این بررسی خواهد آمد به ترتیب زیر می‌باشد. در قسمت ۲، شبکه‌های بیزی را تعریف کرده و مفاهیم مربوط به آنها را توضیح می‌دهیم. در قسمت ۳، روش استنتاج دقیق در شبکه‌های بیزی را توضیح می‌دهیم. در قسمت بعدی، روش استنتاج تقریبی در این شبکه‌ها را توضیح می‌دهیم. در نهایت، در قسمت آخر جمع‌بندی‌ای از مطالب گفته شده ارائه می‌دهیم.

۲- شبکه‌های بیزی

همانطور که در قسمت قبل توضیح داده شد، توزیع توام کامل می‌تواند هر درخواستی را جواب دهد، ولی می‌تواند بطور غیرقابل کنترلی بزرگ شود. همچنین می‌دانیم استقلال و یا استقلال شرطی بین متغیرها می‌تواند تعداد احتمالاتی که برای محاسبه توزیع کامل لازم است، را به طور قابل توجهی کاهش دهد. در این قسمت به معرفی یک ساختمان داده به نام شبکه بیزی می‌پردازیم که نشان دهنده وابستگی‌های فرض شده بین متغیرها است و توزیع کامل با توجه به این فرض‌های وابستگی را بطور دقیق تعیین می‌کند.

شبکه بیزی یک گراف جهت‌دار است که رئوس آن شامل اطلاعات مقادیر احتمالات شرطی هستند. بطور دقیق‌تر این شبکه شامل اجزا و خصوصیات زیر است:

۱ـ یک مجموعه از متغیرهای تصادفی، مجموعه رئوس گراف را تشکیل می‌دهند که این متغیرها می‌توانند گسسته یا پیوسته باشند.

۲ـ یک مجموعه از یالهای جهت‌دار که اگر یک یال از راس X به راس Y باشد، X را والد Y می‌نامیم.

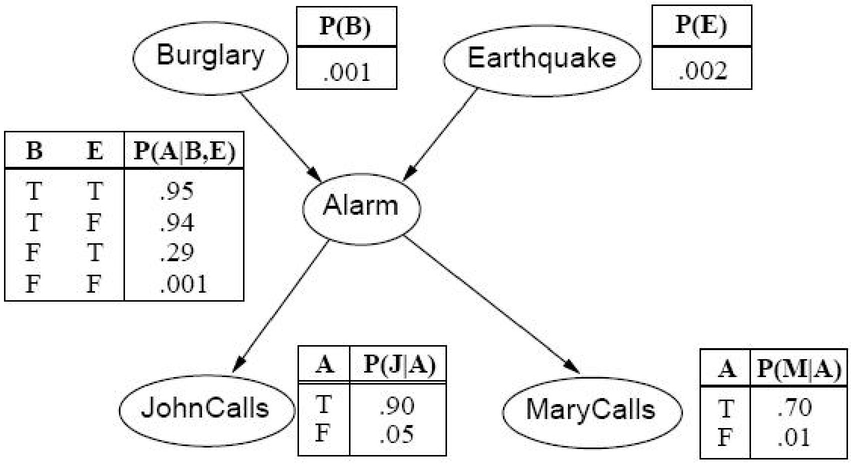
۳ـ هر گره X_i، یک توزیع احتمال شرطی $P(X_i \mid Parents(X_i))$ دارد که تاثیر گره‌های والد بر روی این گره را بصورت عددی نشان می‌دهند.

۴ـ گراف هیچ دور جهت‌داری ندارد و در واقع، یک گراف بدون دور جهت‌دار است.

ساختار شبکه نشان دهنده وابستگی‌های شرطی در قلمرو است. بصورت شهودی، معنی یک یال از X به Y وجود تاثیر مستقیم X بر Y و یا وابستگی مستقیم Y به X است. باید توجه داشت که تعیین این وابستگی‌های مستقیم برای یک فرد خبره قلمرو کار مشکلی نمی‌باشد و به همین دلیل معمولاً در صورت وجود فرد خبره تعیین ساختار شبکه آنچنان سخت نمی‌باشد. پس از تعیین ساختار، تعیین توزیع شرطی مربوط به گره‌ها، ساختمان داده شبکه بیزی را کامل می‌کند و با استفاده از آن می‌توان توزیع توام کامل را بدست آورد که بزودی توضیح داده خواهد شد.

حال به توضیح مطالب ذکر شده با یک مثال می‌پردازیم. فرض کنید شما به تازگی یک آژیر دزدی خریده‌اید که در صورت وقوع دزدی، امکان زیادی دارد که بصدا در آید. علاوه بر این، در صورت وقوع زلزله‌های ضعیف هم ممکن است که آژیر بصدا درآید. همچنین، شما دو همسایه به نام‌های ماری و جان دارید که قول داده‌اند که در صورت شنیدن صدای آژیر با شما در محل کارتان تماس بگیرند. جان همیشه در صورت به صدا درآمدن آژیر به شما تلفن می‌زند. ولی ممکن است که صدای تلفن شما را گاهی با آژیر اشتباه بگیرد و دراین صورت هم تلفن بزند. ماری موسیقی را با صدای بلند دوست دارد و

ممکن است گاهی صدای آژیر را نشنود. حالا با توجه به مشاهده اینکه چه کسی تلفن زده است یا نزده است، می‌خواهیم احتمال دزدی را محاسبه کنیم. شبکه بیزی این قلمرو در شکل ۲ آمده است.



شکل ۲: شبکه بیزی قلمرو دستگاه آژیر جدید.

در ابتدا، ساختار این شبکه را توضیح می‌دهیم و بعد در مورد جدول‌های احتمالات شرطی توضیحاتی می‌دهیم. با توجه به شبکه می‌توان دید که دزدی و زلزله بطور مستقیم بر روی بصدا درآمدن آژیر تاثیر می‌گذارند، ولی بر روی تلفن زدن جان یا ماری تاثیر ندارد، زیرا آنها تنها با شنیدن آژیر تلفن می‌زنند که این موضوع در دو یال خارج شده از گره مربوط به آژیر در شبکه معلوم است. بنابراین شبکه نشان می‌دهد که آنها از وقوع زلزله خفیف آگاه نمی‌شوند.

باید توجه داشت که گره‌ای برای نشان دادن اینکه ماری به موسیقی گوش می‌دهد و یا تلفن زنگ می‌زند در شبکه نیست. این موارد در عدم قطعیت (احتمالی بودن) مربوط به یال‌های خروجی از گره به صدار درآمدن آژیر تاثیر داده شده است. به طور کلی روشی برای اینکه به چه عواملی باید در شبکه گره اختصاص داد، وجود ندارد. در واقع، احتمالات شرطی، تمامی عواملی که در شبکه صریحاً نیامده‌اند، را به طور خلاصه در خود دارند. به این طریق، یک عامل (agent) ساده می‌تواند به طور تقریبی با یک دنیای پیچیده را مدل کند. میزان این تقریب را می‌توان با اضافه کردن اطلاعات مرتبط به شبکه افزایش داد.

توزیع‌های شرطی گره‌ها را با جدول‌های احتمالات شرطی نشان می‌دهند (البته اگر توزیع‌ها پیوسته باشد از روش‌هایی دیگری استفاده می‌شود که بعداً به آنها اشاره خواهد شد). هر سطر در این جدول‌ها نشان دهنده مقدار احتمال مقادیر متغیر راس برای یک حالت شرطی خاصی می‌باشد. هر حالت شرطی یکی از مقداردهی‌های ممکن به والدین راس را نشان می‌دهد. در جدول آخرین مقدار ممکن برای متغیر راس نمایش داده نمی‌شود، زیرا برابر یک منهای جمع احتمال بقیه مقادیر است. به طور کلی، یک جدول برای یک متغیر بولی با k والد بولی، $O(2^k)$ مقدار باید داشته باشد. یک راس که هیچ والدی ندارد تنها شامل یک سطر است که احتمالات اولیه مقادیر متغیر راس را نشان می‌دهند.

۲-۱ مفاهیم شبکه‌های بیزی

در مطالب قبلی، با شبکه بیزی آشنا شدید و در اینجا می‌خواهیم مفهوم یک شبکه بیزی را بررسی کنیم. برای یک شبکه بیزی دو مفهوم می‌توان در نظر گرفت. از دید مفهومی اول، می‌توان شبکه را تقریبی از توزیع توام کامل قلمرو دید. از

دید دوم می‌توان شبکه را بصورت ساختاری که وابستگی و استقلال متغیرها را نشان می‌دهد، دید. هر دو دید معادل هم می‌باشند و دید اول برای طراحی شبکه و دید دوم برای طراحی روال استنتاج مناسب می‌باشد.

۲-۱-۱ نمایش توزیع توام کامل

یک شبکه یک توصیف کامل از قلمرو را ارائه می‌دهد. هر عنصر توزیع احتمال توام کامل (که از این به بعد کامل را برای خلاصه‌سازی حذف می‌کنیم)، با استفاده از اطلاعات درون شبکه قابل محاسبه است. یک عنصر در توزیع را می‌توان بصورت عطف مقداردهی متغیر مانند $P(X_1 = x_1,..., X_n = x_n)$ در نظر گرفت. با توجه به اطلاعات شبکه مقدار یک عنصر بصورت:

$$P(x_1,...,x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i)) \qquad (۲)$$

محاسبه می‌شود که $parents(X_i)$ مقادیر متغیرهای درون $Parents(X_i)$ را نشان می‌دهد. بنابراین، هرعنصر توزیع توام بصورت ضرب تعدادی از عناصر جدول‌های احتمالات شرطی محاسبه می‌شود و این جدول‌ها به نوعی تجزیه توزیع توام را انجام می‌دهد. برای نشان دادن این موضوع، احتمال بصدا در آمدن آژیر در حالی که نه زلزله آمده است و نه دزدی شده است در حالیکه جان و ماری هر دو تماس گرفته‌اند را حساب می‌کنیم. از حرف اول متغیرها برای نشان دادن آنها استفاده می‌کنیم:

$$\begin{aligned} P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) &= P(j \mid a)P(m \mid a)P(a \mid \neg b \wedge \neg e)P(\neg b)P(\neg e) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.00062 \end{aligned}$$

با استفاده از فرمول (۲) می‌توان ایده‌ای برای ساختن ساختار شبکه بیزی بدست آورد. این فرمول مشخص می‌کند که یک شبکه بیزی چه معنی‌ای دارد، ولی به طور صریح روش ساختن شبکه را طوری که یک نمایش مناسب از قلمرو باشد، معلوم نمی‌سازد. باید توجه داشت در یک شبکه فرض بر نوع خاصی از وابستگی بین متغیرها است که لزوماً با واقعیت تطابق ندارد. بنابراین، ما تنها دنبال یک شبکه خوب می‌گردیم نه شبکه‌ای که دقیقاً توزیع توام را مدل کند (چون ممکن است ممکن نباشد). حال نشان می‌دهیم که فرمول (۲) منجر به فرضی در مورد استقلال‌های شرطی می‌شود که به فرایند ساخت شبکه کمک می‌کند.

با توجه به قانون زنجیره‌ای (chain rule) در احتمالات، می‌توانیم احتمال یک عنصر توزیع توام را بصورت:

$$P(x_1, x_2,..., x_n) = P(x_n \mid x_{n-1},..., x_1)P(x_{n-1} \mid x_{n-2},..., x_1)...P(x_2 \mid x_1)P(x_1)$$

بنویسیم. با مقایسه این فرمول با فرمول (۲)، پی می‌بریم که شبکه بیزی وقتی دقیقاً برابر توزیع توام است که برای هر متغیر X_i در شبکه داشته باشیم:

$$P(X_i \mid X_{i-1},..., X_1) = P(X_i \mid Parents(X_i)) \qquad (۳)$$

با این فرض که $Parents(X_i) \subseteq \{X_1,..., X_{i-1}\}$ باشد. فرمول (۳) مشخص می‌کند که شبکه بیزی تنها وقتی دقیقاً برابر توزیع توام است که هر متغیر بطور شرطی مستقل از تمامی متغیرهای رئوس بعدی در ترتیب اندیس رئوس باشد. بنابراین، در صورتی که بخواهیم یک شبکه بیزی خوب بسازیم که تا حدی خوبی به توزیع توام نزدیک باشد، نیاز داریم تا والدهای هر راس را طوری انتخاب کنیم که این خاصیت حفظ شود. یعنی باید والدین راس منتسب به متغیر X_i را، از رئوس منتسب به $X_1,..., X_{i-1}$ که بر روی X_i تاثیر مستقیم دارند انتخاب کنیم. برای مثال، فرض کنید، قسمتی از

شبکه شکل ۳ را ساخته‌ایم و تنها تعیین والدین MaryCalls باقیمانده است. والدین راس این متغیر را باید از متغیرهای قبل از این متغیر انتخاب کنیم که چون تنها Alarm بصورت مستقیم بر آن تاثیر می‌گذارد، به عنوان والد آن انتخاب می‌شود.

شبکه‌های بیزی با در نظر گرفتن استقلال متغیرها می‌توانند در همان حال که توزیع توام را بخوبی نشان میدهند، از نظر محاسباتی هم قابل قبول باشند. اما در قلمروهایی که استقلال بین متغیرها کم است، این شبکه‌ها نیز نمی‌توانند کمکی کنند. در بعضی موارد، وابستگی بین متغیرهایی که آنچنان قوی نیست در نظر گرفته نمی‌شود، تا در مقابل از دست رفتن مقداری از دقت، شبکه از نظر محاسباتی قابل قبول شود. علاوه بر این، ترتیب اضافه کردن متغیرها برای ساختار شبکه هم بسیار موثر و می‌تواند برای قلمروای که می‌تواند شبکه مناسبی داشته باشد یک شبکه با یال‌های زیاد و غیرقابل قبول از لحاظ محاسباتی نتیجه دهد. این شبکه‌ها نه تنها یال‌های بیشتری دارند، بلکه ما را با سختی حساب احتمالاتی روبرو می‌کنند که بسیار پیچیده می‌تواند باشد (مثلاً احتمال وقوع زلزله در صورت بصدا درآمدن آژیر و صورت گرفتن دزدی).

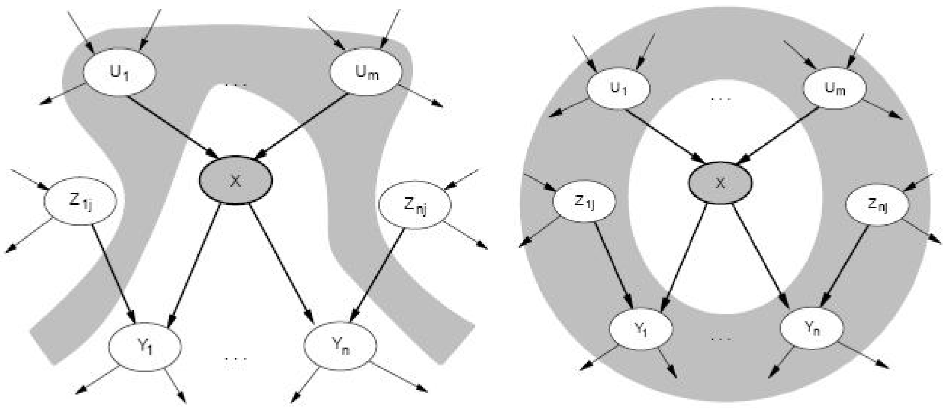
۲-۲ رابطه‌های استقلال شرطی در شبکه‌های بیزی

تا کنون یک مفهوم عددی برای شبکه‌های بیزی ارائه کردیم و نشان دادیم چگونه استفاده از این مفهوم می‌توان ساختار شبکه بیزی را بدست آورد. در حقیقت، ما برعکس این کار را هم می‌توانیم انجام دهیم. ما می‌توانیم از مفاهیم ساختاری (که وابستگی‌های شرطی کد شده در ساختار گرافی را نشان می‌دهد) برای بدست آوردن جدول‌های احتمالات شرطی و در نتیجه، بدست آوردن مفاهیم عددی استفاده کنیم. مفاهیم ساختاری با یکی از دو صورت زیر که معادل هستند می‌تواند داده شود:

۱ـ یک گره بصورت شرطی مستقل از گره‌های غیربچه خود است در صورتی که والدهای آن را داده باشند و برای مثال در شکل ۲، JohnCalls از Burglary و Earthquake در صورت داده شده بودن مقدار Alarm مستقل است.

۲ـ یک گره بصورت شرطی مستقل از تمامی گره‌های دیگر شبکه در صورت داده شده بودن والدهای آن، فرزندان و والدهای فرزندان آن (یا به عبارت دیگر در صورت داده شده بودن پوشش مارکوفی (Markov Blanket) آن) است. مثلاً Burglary مستقل از JohnCalls و MaryCalls در صورتی که مقادیر Alarm و Earthquake داده شده باشند، است.

توصیفی از این دو نوع نمایش در شکل ۳ آمده است. از این نوع بیان استقلال‌ها و جداول احتمالات شرطی می‌توان توزیع توام را بدست آورد. بنابراین مفاهیم عددی و مفاهیم ساختاری معادل هستند.



شکل ۳: ارائه مفاهیم ساختاری به دو صورت معمول.

۲-۳ نمایش کارآمد توزیع‌های شرطی

در مورد تعیین جدول احتمالات شرطی باید توجه داشت که حتی اگر تعداد والدهای یک گره یک عدد کوچک k باشد، باید $O(2^k)$ مقدار برای آن گره حساب شود که به دانش زیادی برای تعیین این تعداد مقدار نیاز است. در حقیقت، این بدترین حالات است که فرض کنیم رابطه فرزندان و والدین دلخواه است و بسیاری مواقع یک رابطه مشخص و استاندارد بین مقادیر والدها و بچه برقرار است که کار تعیین توزیع شرطی را ساده می‌کند. مثلاً یک گره ممکن است لزوماً فصل مقادیر بولی والدین خود باشد.

علاوه بر این، بسیاری از مسائل دنیای واقعی متغیرهایی پیوسته دارند. نمایش این احتمالات شرطی این متغیرها بصورت جدول ممکن نمی‌باشد. در اینگونه موارد دو روش کلی استفاده می‌شود. یک روش گسسته کردن مقادیر متغیرهای پیوسته است و ایجاد جدول‌های احتمالات شرطی برای مقادیر گسسته آنها است. مشکل این روش این است که کارایی روش و دقت آن به شدت افت می‌کند و علاوه بر آن، اندازه جداول نیز بسیار بزرگ می‌شود. روش دیگر استفاده از توزیع‌های پیوسته استاندارد که با مجموعه‌ای متناهی از پارامترها تعریف می‌شوند، است. مثلاً یک توزیع نرمال شامل دو پارامتر s ,m است و بسیار بکار می‌رود.

۲-۴ یادگیری شبکه‌های بیزی

یکی از مشکلات استفاده از شبکه‌های بیزی این است که ایجاد کامل شبکه حتی برای یک خبره هم می‌تواند مشکل باشد. بنابراین، تلاش‌های زیادی برای یادگیری شبکه‌های بیزی صورت گرفته است. در هر شبکه بیزی ساختار آن و جداول احتمالات شرطی تعیین کننده آن هستند. بنابراین، باید بتوان با فرایند یادگیری این دو مورد را تعیین کرد. برای یادگیری خودکار ساختار شبکه یکی از روش‌های اصلی بر پایه تعیین وابستگی بین متغیرها بنا نهاده شده است. برای تعیین این وابستگی‌ها معیارهای زیادی مانند معیار آنتروپی طراحی شده‌اند. با استفاده از این معیارها مانند روشی که قبلاً توضیح داده شد، وابستگی هر متغیر X_i را نسبت به متغیرهای $X_1,...,X_{i-1}$ می‌سنجیم و آنها را که معیار وابستگی‌شان از یک آستانه‌ای بیشتر بود، به عنوان والدهای X_i انتخاب می‌کنیم. پس از تعیین ساختار، اگر مقدار همه متغیرها به طور کامل قابل مشاهده باشند، از تخمین احتمال معمولی (گرفتن تعدادی نمونه و شمردن تعداد اتفاقات یک رویداد در این مجموعه نمونه) استفاده می‌کنیم. اگر بعضی از متغیرها قابل مشاهده نباشند، با استفاده از آموزش یک شبکه نورونی می‌توان مقادیر جداول احتمالات شرطی را یاد گرفت.

تاکنون توضیحات مختصری در مورد شبکه‌های بیزی، مفاهیم آنها و موارد مربوط به طراحی آنها ارائه شد. در نهایت پس از طراحی شبکه تنها نکته‌ای که باقی می‌ماند، استنتاج با این شبکه‌ها است که هدف اصلی ما در بکار بردن آنها می‌باشد.

۳- استنتاج دقیق در شبکه‌های بیزی

همانطور که قبلاً گفته شد، یک عمل پایه برای سیستم‌های احتمالی، جواب به درخواست‌هایی است که احتمال وقوع مقادیری خاص برای یک مجموعه از متغیرها را با فرض داده شده بودن تعدادی مشاهده (یک مقداردهی متغیرهای

مشاهده) می‌خواهند بدانند. در اینجا از نمادگذاری‌ای مانند آنچه در قسمت ۱ معرفی شده، استفاده می‌کنیم. X نشان دهنده متغیر درخواست، $\mathbf{E}$ نشان دهنده مجموعه متغیرهای مشاهده $E_1, E_2, \mathbf{K}, E_m$، $\mathbf{e}$ نشان دهنده مقادیر مشاهده شده برای متغیرهای $\mathbf{E}$ و $\mathbf{Y}$ نشان دهنده متغیرهایی که مشاهده نشده‌اند (که به آنها متغیرهای مخفی نیز می‌گویند) $Y_1, Y_2, \mathbf{K}, Y_l$ است.

بنابراین مجموعه کل متغیرها $\mathbf{X}$ برابر $X\,\mathbf{U}\,\mathbf{E}\,\mathbf{U}\,\mathbf{Y}$ است. یک درخواست معمول بصورت $P(X\,|\,\mathbf{e})$ است. در اینجا برای سادگی فرض کرده‌ایم که X جز متغیرهای مشاهده شده نیست، متغیرها بولی هستند و درخواست تنها در مورد یک متغیر است، زیرا تعمیم توضیحات به حالت کلی به آسانی امکان‌پذیر است.

۳-۱ استنتاج بوسیله محاسبه تک‌تک عناصر احتمالی

در قسمت ۱ توضیح دادیم که محاسبه یک درخواست می‌تواند بصورت جمع تک‌تک تعدادی از عناصر توزیع توام صورت بپذیرد. با عبارت دیگر، جواب به یک درخواست، با استفاده از فرمول (۱) امکان‌پذیر است. از آنجایی که یک شبکه بیزی دقیقاً توزیع توام را نشان می‌دهد (در صورتی که خاصیت گراف بدون دور جهت‌دار بودن گراف وابستگی متغیرها صادق باشد) و به عبارت روشن‌تر، از آنجایی که فرمول (۲) نشان می‌دهد که هر عبارت $P(X, \mathbf{e}, \mathbf{y})$ در فرمول (۱) چگونه بصورت ضرب احتمالات شرطی قابل محاسبه است، شبکه بیزی بصورت دقیق می‌تواند به درخواست‌ها جواب دهد.

باید توجه داشت که استنتاج بوسیله محاسبه تک تک عناصر از $O(n2^n)$ است. زیرا محاسبه عبارت $P(X, \mathbf{e}, \mathbf{y})$ نیاز به $n-1$ ضرب دارد و $\mathbf{y}$ می‌تواند در بدترین حالت 2^n حالت را شامل شود که به $(n-1)2^n$ ضرب منجر می‌شود. حال سعی می‌کنیم با ارائه یک مثال به توضیح ایده‌ای بپردازیم که ما را به یک الگوریتم $O(2^n)$ برای استنتاج دقیق می‌رساند.

درخواست $P(Burglary\,|\,JohnCalls = ture, MaryCalls = true)$ را در نظر بگیرید. در اینجا متغیرهای مخفی Earthquake و Alarm هستند. با توجه به فرمول (۱) با در نظر گرفتن حرف اول اسم‌ها برای نشان دادن متغیرها داریم:

$$P(B\,|\,j,m) = aP(B,j,m) = a\sum_e \sum_a P(B,e,a,j,m)$$

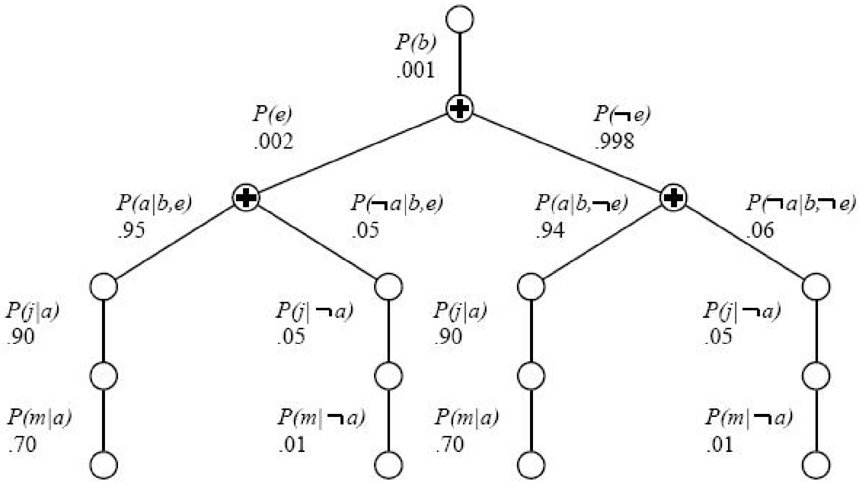
با توجه به ساختار شبکه بیزی به عبارتی بصورت زیر دست می‌یابیم:

$$P(|\,j,m) = a\sum_e \sum_a P(B)P(e)P(a\,|\,b,e)P(j\,|\,a)P(m\,|\,a)$$

که در اینجا $P(b)$ معادل احتمال یک مقداردهی دلخواه متغیر B است. برای محاسبه این عبارت نیاز به جمع ۵ عبارت که هر کدام از ۴ ضرب بدست آمده‌اند می‌باشد، ولی بهبودهایی با دقت بیشتر در عبارت می‌تواند مشاهده شود. $P(b)$ مستقل از متغیرهای مجموع گیری است و می‌تواند به بیرون مجموع‌ها انتقال یابد. همچنین $P(e)$ می‌تواند به خارج مجموع گیری بر روی a انتقال یابد. با در نظر گرفتن این نکات به فرمول زیر می‌رسیم:

$$P(b\,|\,j,m) = aP(b)\sum_e P(e)\sum_a P(a\,|\,b,e)P(j\,|\,a)P(m/a)$$

ساختار محاسبه براساس عبارت بالا که بسیار بهتر است، در شکل ۴ نشان داده شده است.



شکل ۴: جواب به درخواست با محاسبه عبارت بهینه‌تر.

۳-۲ الگوریتم حذف متغیر

همانطور که در مثال بالا دیده شد، الگوریتم محاسبه تک به تک عناصر، می‌تواند با حذف یکسری محاسبات تکراری با توجه به ساختار عبارت، بهبود یابد. در واقع ایده اصلی این است که هر عبارت را تنها یک بار محاسبه می‌کنیم و دفعات بعد که به این عبارت نیاز بود از مقدار محاسبه شده استفاده کنیم (این نوعی برنامه‌نویسی پویا می‌باشد). این ایده منجر به الگوریتم حذف متغیر می‌شود که بطور خلاصه به این صورت عمل می‌کند که از راست به چپ در عبارت حرکت می‌کند و با محاسبه حالت‌های متغیرها آنها را تک‌تک حذف می‌کند. نتایج میانی برای استفاده آینده ذخیره می‌شود و جمع کردن روی هر متغیر تنها برای آن قسمتی از عبارت که به این متغیر وابسته است، صورت می‌گیرد. باید توجه داشت که پیچیدگی زمانی و مکانی الگوریتم حذف متغیر وابسته به اندازه جداول نتایج میانی است و در بدترین حالت می‌تواند از $O(2^n)$ باشد. این الگوریتم در شکل ۵ دیده می‌شود.

```
function ELIMINATION-ASK(X, e, bn) returns a distribution over X
   inputs: X, the query variable
           e, evidence specified as an event
           bn, a belief network specifying joint distribution P(X1, ..., Xn)
   factors ← []; vars ← REVERSE(VARS[bn])
   for each var in vars do
      factors ← [MAKE-FACTOR(var, e)|factors]
      if var is a hidden variable then factors ← SUM-OUT(var, factors)
   return NORMALIZE(POINTWISE-PRODUCT(factors))
```

شکل ۵: الگوریتم حذف متغیر.

اما برای دسته‌ای خاص و پرکاربرد از شبکه‌های بیزی این الگوریتم پیچیدگی خطی براساس تعداد عناصر جداول احتمالات شرطی دارد. این دسته از شبکه‌ها، آنهایی هستند که بین هر دو راس آنها تنها یک مسیر بدون جهت وجود دارد و به آنها شبکه‌های یک همبند و یا چند درخت (polytree) می‌گویند. باید توجه داشت پیچیدگی خطی براساس مقدار عناصر جداول، اگر تعداد والدهای تمامی گره‌ها به یک حد بالا مستقل از تعداد کل گره‌ها محدود باشد، به معنی خطی

بودن کل فرایند استنتاج بر اساس تعداد گره‌های شبکه است. برای شبکه‌های بیزی که یک همبند نباشند، الگوریتم حذف متغیر می‌تواند در بدترین حالت پیچیدگی نمایی داشته باشد (حتی اگر تعداد والدهای گره‌ها به یک حد بالای ثابت محدود باشد).

۴- استنتاج تقریبی در شبکه‌های بیزی

همانطور که دیده شد، استنتاج دقیق در شبکه‌های بیزی از نظر پیچیدگی زمانی و حافظه‌ای در مسائل بسیاری عملی نمی‌باشد و در بدترین حالت نمایی است. این مشکل باعث ایجاد دسته جدیدی از الگوریتم‌ها شد که با استفاده از نمونه‌گیری از شبکه سعی در رسیدن به یک تقریب مناسب از احتمال مورد سوال درخواست دارند. در اینجا دو دسته روش معروف از الگوریتم‌ها که نمونه‌گیری مستقیم و نمونه‌گیری زنجیره مارکونی هستند، بطور مختصر توضیح داده می‌شود.

۴-۱ روش‌های نمونه‌گیری مستقیم

ایده اصلی این روش‌ها نمونه‌گیری با توجه به توزیعی است که شبکه بیزی نشان می‌دهد. مهمترین مشکل در اعمال این ایده ساده وجود مشاهدات می‌باشد، یعنی هر نمونه‌ای که از شبکه گرفته می‌شود معتبر نمی‌باشد و باید با مقدار متغیرهای مشاهده شده سازگار باشد. برای حل این مشکل دو روش ارائه شده است که دومی معمولاً کارایی بیشتری دارد.

۴-۱-۱ نمونه‌گیری با رد کردن

این روش از یک ایده قدیمی در آمار استفاده می‌کند. از شبکه بیزی تعداد نمونه می‌گیرد و تنها آنهایی که با مشاهدات سازگار هستند را در فرایند محاسبه احتمال تقریبی در نظر می‌گیرد. این الگوریتم در شکل ۶ نشان داده شده است. باید توجه داشت، اثبات می‌شود که این روش یک تقریب سازگار از توزیع شرطی را نتیجه می‌دهد. یعنی اگر تعداد نمونه‌ها به بی‌نهایت میل کند، تقریب به مقدار واقعی توزیع میل می‌کند.

```
function PRIOR-SAMPLE(bn) returns an event sampled from bn
   inputs: bn, a belief network specifying joint distribution P(X_1, ..., X_n)
   x ← an event with n elements
   for i = 1 to n do
       x_i ← a random sample from P(X_i | parents(X_i))
             given the values of Parents(X_i) in x
   return x
```

```
function REJECTION-SAMPLING(X, e, bn, N) returns an estimate of P(X|e)
   local variables: N, a vector of counts over X, initially zero
   for j = 1 to N do
       x ← PRIOR-SAMPLE(bn)
       if x is consistent with e then
          N[x] ← N[x]+1 where x is the value of X in x
   return NORMALIZE(N[X])
```

شکل ۶: الگوریتم نمونه‌گیری با رد کردن.

۴-۱-۲ نمونه‌گیری وزن‌دار

دومین دسته روش از روش‌های نمونه‌گیری مستقیم، نمونه‌گیری وزن‌دار می‌باشد که برای حل مشکل نمونه‌گیری با ردکردن طراحی شده است. یکی از مشکلات بزرگ روش نمونه‌گیری با ردکردن این است که ممکن است، احتمال وقوع مشاهدات کم باشد و در نتیجه، یک قسمت اصلی از نمونه‌ها رد می‌شود. این باعث باقیماندن نمونه‌های خیلی کمی می‌شود که برای تقریب خوب، مناسب نمی‌باشند و عملاً استفاده از این روش در قلمروهای پیچیده امکان‌پذیر نمی‌باشد.

نمونه‌گیری وزن‌دار مشکل نمونه‌گیری با ردکردن را به این شکل رفع می‌کند که تنها نمونه‌هایی ایجاد می‌کند که سازگار با مشاهدات هستند. این روش مقادیر متغیرهای E را ثابت در نظر می‌گیرد و تنها از مقادیر متغیرهای X و Y نمونه‌گیری می‌کند. باید توجه داشت که این نوع نمونه‌گیری توجه یکسانی به نمونه‌ها نمی‌کند و نمی‌توان تنها تعداد نمونه‌ها را برای محاسبه احتمالات موردنظر درخواست‌ها استفاده کرد. بهمین دلیل به هر نمونه وزنی متناسب با احتمال وقوع مشاهدات با توجه به مقادیر متغیرهای دیگر نسبت می‌گیرد و از جمع وزن‌های نمونه‌ها برای محاسبه احتمال موردنظر درخواست استفاده می‌شود. این الگوریتم در شکل ۷ آمده است. باید توجه داشت که این روش نمونه‌گیری، تقریبی سازگار را نتیجه می‌دهد.

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

  for j = 1 to N do
      x, w ← WEIGHTED-SAMPLE(bn)
      W[x] ← W[x] + w where x is the value of X in x
  return NORMALIZE(W[X])
-----------------------------------------------------------------------
function WEIGHTED-SAMPLE(bn, e) returns an event and a weight

  x ← an event with n elements; w ← 1
  for i = 1 to n do
      if X_i has a value x_i in e
          then w ← w × P(X_i = x_i | parents(X_i))
          else x_i ← a random sample from P(X_i | parents(X_i))
  return x, w
```

شکل ۷: الگوریتم نمونه‌گیری وزن‌دار.

۴-۲ نمونه‌گیری زنجیره مارکوفی

دو روش توضیح داده شده در نمونه‌گیری مستقیم، برای تولید هر نمونه تمامی مقادیر را از ابتدا نمونه‌گیری می‌کنند که ممکن است از نظر زمانی هزینه‌بر باشد. در مقابل، روش‌های مبتنی بر زنجیره مارکوفی هر نمونه را با یک تغییر تصادفی در نمونه قبلی بدست می‌آوردند. بطور دقیق‌تر، هر نمونه جدید بصورت تصادفی از تغیر یکی از متغیرهای غیرمشاهده‌ای نمونه قبلی بدست می‌آید. برای این کار از آن متغیر با فرض داده شده بودن مقدار مقادیر متغیرهای پوشش مارکوفی آن در نمونه قبلی، نمونه‌برداری می‌شود. به این ترتیب نمونه‌گیری زنجیره مارکوفی به حرکت در فضای نمونه‌ها (مقادیر مختلف متغیرهای غیرمشاهده شده) می‌پردازد. الگوریتم ساده‌ای براساس این ایده در شکل ۸ دیده می‌شود. می‌توان ثابت کرد که نمونه‌گیری زنجیره مارکوفی نیز جز نمونه‌گیرهای سازگار می‌باشد.

```
function MCMC-ASK(X, e, bn, N) returns an estimate of P(X|e)
   local variables: N[X], a vector of counts over 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 Y
   for j = 1 to N do
       for each Z_i in Z do
           sample the value of Z_i in x from P(Z_i|mb(Z_i))
               given the values of MB(Z_i) in x
           N[x] ← N[x] + 1 where x is the value of X in x
   return NORMALIZE(N[X])
```

شکل ۸: الگوریتم نمونه‌گیری وزن‌دار.

۵- جمع‌بندی و مطالعات بیشتر

شبکه‌های بیزی یک نمایش مناسب قلمرو برای کنترل عدم قطعیت می‌باشند. یک شبکه بیزی یک گراف بدون دور جهت‌دار است که رئوس آن متغیرهای تصادفی و هر راس یک توزیع شرطی براساس والدهای خود دارد. در واقع، شبکه‌های بیزی یک راه مناسب برای نشان دادن استقلال‌های شرطی که فرایند استنتاج احتمالی را سریع می‌کند، می‌باشند. شبکه‌های بیزی در واقع یک توزیع توام نیز برای قلمرو تعیین می‌کنند که هر عنصر توزیع بصورت ضرب تعدادی عناصر از جداول احتمالات شرطی رئوس می‌باشد. همچنین، بطور کلی اندازه گره‌های یک شبکه بیزی بصورت نمایی کوچکتر از تعداد عناصر درون این جداول احتمالات شرطی است. بنابراین، برای نمایش کارآمد جداول توزیع‌های شرطی راه‌های زیادی طراحی شده است که استفاده از توزیع‌های خاص با پارامترهای محدود از جمله این روش‌ها است. علاوه بر این، روش‌های زیادی برای استنتاج کارآمد در شبکه‌های بیزی طراحی شده است. روش‌هایی که سعی در استنتاج دقیق دارند، در قلمروهای پیچیده عملی نیستند و منجر به روشهای استنتاج تقریبی مانند نمونه‌گیری وزن دار و نمونه‌گیری زنجیره مارکوفی می‌شوند. برای مطالعه کاربردهای شبکه‌های بیزی به فصل شش [۱] رجوع کنید. همچنین، برای آشنایی جامع‌تر با این شبکه‌ها و مشاهده جزئیات بیشتر روش‌های استنتاج به فصول ۱۳ و ۱۴ [۲] مراجعه کنید. در نهایت، خواننده علاقمند روش‌های یادگیری شبکه بیزی، مطالعه [۳] را مفید خواهد یافت.

مراجع

[1] T. Mitchell, *Machine Learning*, McGraw Hill, 1997.

[2] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, second edition, Prentice Hall, 2003.

[3] D. Heckerman, D. Geiger, and D. M. Chickering, *Learning Bayesian Networks: The Combination of Knowledge and Statistical Data*, Machine Learning, vol. 20, pp. 197-243, 1995.