علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۷، شماره ۲، ۱۳۹۸
صفحات ۳۲- ۲۴
مقاله عادی

یادگیری ساختار شبکه‌های بیزین با استفاده از روش گردکردن قطعی

شهاب بهجتی[۱*]، حمید بیگی[۲]

*نویسنده مسئول، دریافت: ۹۸/۰۷/۱۹، بازنگری: ۹۸/۱۰/۰۱، پذیرش: ۹۸/۱۲/۱۵

[۱] دانش‌آموخته دکتری، علوم کامپیوتر، پژوهشکده علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی، تهران، ایران
[۲] دانشیار، مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران

چکیده

شبکه‌های بیزین یکی از پرکاربردترین مدل‌های گرافی احتمالاتی بوده که برای نمایش یک توزیع احتمال به‌کاررفته و دارای کاربردهای بسیار متنوع در هوش مصنوعی، داده‌کاوی و یادگیری ماشین است. یکی از مهم‌ترین مسائل در این شبکه‌ها، مسئله یادگیری ساختار از روی مجموعه داده‌ها آموزشی است. به‌طورکلی روش‌های یادگیری ساختار به سه دسته مبتنی بر محدودیت، مبتنی بر امتیاز و ترکیبی تقسیم می‌شوند. در این مقاله یک روش مبتنی بر امتیاز جهت ساخت شبکه بیزین ارائه می‌شود که مبتنی بر روش گردسازی قطعی در برنامه‌ریزی خطی است. روش پیشنهادی ابتدا مسئله یادگیری ساختار را به‌صورت یک برنامه خطی صحیح مدل‌سازی نموده و سپس آن را به یک مسئله برنامه‌ریزی خطی تعدیل می‌کند. سپس با حل برنامه خطی تعدیل شده، جواب‌های کسری به‌دست‌آمده را با استفاده از روش معرفی شده گردسازی قطعی به جواب‌های صحیح تبدیل می‌کند.

کلمات کلیدی: برنامه‌ریزی خطی، شبکه بیزین، گرد کردن قطعی، یادگیری ساختار در شبکه‌های بیزین.

۱- مقدمه

مدل‌های گرافی احتمالاتی به‌عنوان ابزاری کارآمد برای نمایش فشرده توزیع‌های احتمال توأم شناخته می‌شوند. یک مدل گرافی یا مدل گرافی احتمالاتی مدلی احتمالاتی است که در آن از یک گراف برای بیان و نمایش ساختار وابستگی‌های شرطی میان یک مجموعه از متغیرهای تصادفی استفاده می‌شود. از این مدل‌ها به‌طور گسترده‌ای در نظریه احتمالات، آمار بیزی و یادگیری ماشین استفاده می‌شود. مدل‌های گرافی احتمالاتی گراف‌هایی هستند که هر گره آن یک متغیر تصادفی و هر یال میان دو گره نشان دهنده وابستگی شرطی میان متغیرهای تصادفی متناظر با آن دو گره است. درواقع استفاده از مدل‌های گرافی احتمالاتی یک روش فشرده برای نمایش توزیع احتمال توأم میان متغیرهای تصادفی را فراهم می‌کند. برای نمونه، اگر $\boldsymbol{n}$ متغیر تصادفی دودویی داشته باشیم، آنگاه برای نمایش توزیع توأم $\boldsymbol{p(X_1,\dots,X_n)}$ نیازمند تعیین $۲^{\boldsymbol{n}}$ پارامتر هستیم که این دارای پیچیدگی فضا از مرتبه نمایی $\boldsymbol{O(۲^n)}$ است. درحالی‌که با استفاده از شبکه‌های بیزین تعداد پارامترها به نحو قابل ملاحظه‌ای کاهش می‌یابد.

سه خانواده از مدل‌های گرافی احتمالاتی وجود دارند که ارتباط بسیار نزدیکی با یکدیگر داشته اما هرکدام برای کاربرد و توزیع احتمال متفاوتی مناسب هستند. این سه خانواده عبارت‌اند از: مدل‌های گرافی احتمالاتی بدون جهت، گراف‌های عامل یا گراف فاکتور و مدل‌های گرافی احتمالاتی جهت‌دار(شبکه‌های بیزین). یک توزیع احتمال روی مجموعه متغیرهای $\{X_1,\dots,X_n\}$ مانند $p(X_1,\dots,X_n)$ را در نظر بگیرید. یک مدل گرافی احتمالاتی، یک گراف همراه با یک مجموعه از توابع احتمال روی زیرمجموعه‌ای از متغیرهای تصادفی بوده که توزیع احتمال موردنظر را تعریف می‌کند. این مدل‌ها تزویجی میان نظریه احتمال و نظریه گراف بوده و امکان نمایش فشرده و استنباط کارآمد را در هنگامی‌که توزیع احتمال موردنظر دارای ساختارهای استقلال خاصی است، فراهم می‌کند. از مدل‌های گرافی احتمالاتی بدون جهت که تحت عنوان میدان‌های تصادفی مارکوف[۱] نیز شناخته می‌شوند بیشتر در فیزیک و پردازش تصویر استفاده می‌شود. از مدل‌های گرافی احتمالاتی جهت‌دار نیز که تحت عناوینی نظیر شبکه‌های بیزین[۲]، شبکه‌های باور[۳] و شبکه‌های سببی[۴] نیز شناخته می‌شوند بیشتر در حوزه‌های هوش مصنوعی و یادگیری ماشین استفاده می‌شود. شبکه‌های بیزین دارای کاربرهایی بسیار متنوع در حوزه‌های مختلف مهندسی، آمار و هوش مصنوعی است [۲۲، ۳۴، ۳۵، ۲۰، ۳۲، ۲۹، ۱۱]. میدان‌های تصادفی مارکوف که به‌اختصار شبکه مارکوف نیز نامیده می‌شوند، مدل‌های گرافی احتمالاتی بدون جهتی هستند که گره‌های آن متناظر با متغیرهای تصادفی و یال‌های بدون جهت نشان دهنده وابستگی‌های میان متغیرهای تصادفی است. در این‌گونه از مدل‌ها، تعابیر گرافی خیلی ساده‌تر است: X و Y با توجه به Z مستقل از هم هستند هرگاه آن‌ها توسط Z از یکدیگر جدا شوند. به ازای هر مجموعه $\{X_1,\dots,X_n\}$ از

متغیرهای تصادفی تعداد $2^{\frac{n(n-1)}{2}}$ گراف بدون جهت متمایز از یکدیگر وجود دارد. توزیع توأم برای شبکه‌های مارکوف به‌صورت یک حاصل‌ضرب از توابع نامنفی خوشه‌های گراف بیان می‌شوند:

$$p\left(X_1, \dots, X_n\right) = \frac{1}{Z} \prod_{c \in \mathcal{C}(G)} \Psi_c\left(x_c\right), \quad (1)$$

که $\mathcal{C}(G)$ مجموعه خوشه‌های ماکسیمال گراف G و $\Psi_c\left(x_c\right)$ پتانسیل خوشه‌ای[5] و Z یک مقدار ثابت نرمال‌ساز است. یک توزیع p که توسط گراف بدون جهت G نمایش داده می‌شود یک توزیع گیبز[6] روی G نامیده می‌شود.

شبکه‌های بیزین روشی متفاوت و کارآمد برای نمایش توزیع احتمال توأم مبتنی بر تجزیه آن به حاصل‌ضرب توزیع‌های احتمال شرطی ارائه می‌کنند. با این شیوه نمایش، تعداد پارامترهای موردنیاز جهت ساخت مدل احتمالاتی و استنباط در مسائل بزرگ به طرز قابل ملاحظه‌ای کاهش می‌یابد. هر توزیع احتمال توأم $p(X_1, \dots, X_n)$ را می‌توان با استفاده از قاعده زنجیره‌ای زیر به حاصل‌ضرب توزیع‌های احتمال شرطی به‌صورت زیر تجزیه کرد:

$$p\left(X_1, \dots, X_n\right) = p(X_1)p(X_2 \mid X_1)p(X_3 \mid X_1, X_2) \dots p(X_n \mid X_1, \dots, X_{n-1})$$

این تجزیه توسط یک گراف جهت‌دار بدون دور[7] یا DAG قابل نمایش است. اگر متغیر X_i را توسط یک گره در گراف نشان دهیم، آنگاه متغیر تصادفی X_j به‌عنوان والد گره X_i را می‌توان به‌صورت یک یال جهت‌دار از X_j به X_i نشان داد. گراف‌های جهت‌دار و بدون دور متعددی برای نمایش یک توزیع احتمال توأم وجود دارد. برای نمایش تجزیه به‌دست‌آمده از قاعده زنجیره‌ای، می‌توان از یک گراف جهت‌دار استفاده کرد. ازآنجاکه تعداد $n!$ جایگشت متفاوت برای n متغیر وجود دارد، تعداد گراف‌های جهت‌دار بدون دور ممکن برای نمایش این توزیع نیز برابر $n!$ است. اما از طرف دیگر، اگر متغیرهای X_i و X_j مستقل(شرطی) از یکدیگر باشند آنگاه هیچ‌کدام از آن‌ها در مجموعه شرطی دیگری در قاعده زنجیره‌ای ظاهر نمی‌شوند و درنتیجه هیچ‌کدام والد یا پدر دیگری در گراف جهت‌دار بدون دور نمایش دهنده آن نیست. بنابراین هیچ یال جهت‌داری میان گره‌های X_i و X_j وجود ندارد. اگر یک گراف جهت‌دار بدون دور نشان دهنده یک تجزیه مربوط به یک توزیع احتمال توأم باشد، در این صورت این توزیع احتمال توأم را می‌توان به‌صورت زیر بازسازی کرد.

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid pa(X_i)) \quad (2)$$

که در آن منظور از $pa(X_i)$ مجموعه والدهای گره X_i در گراف جهت‌دار بدون دور مربوطه است. گراف‌های جهت‌دار بدون دور مختلف نمایش دقیق توزیع احتمال توأم یکسان هستند. به‌عنوان‌مثال تمامی $n!$گراف جهت‌دار بدون دور بحث شده در بالا همراه با پارامترهای متناسب محاسبه شده همگی برای نمایش توزیع احتمال توأم به‌طور معادل قابل قبول هستند اگرچه ساختار آن‌ها با یکدیگر متفاوت است. تعریف۱ ساختارهای هم‌ارز و معادل را مشخص می‌کند[۱۹].

تعریف۱. دو ساختار DAG مانند $\mathcal{G}_1$ و $\mathcal{G}_2$ را هم‌ارز گوییم هرگاه مجموعه توزیع‌های قابل نمایش توسط $\mathcal{G}_1$ با مجموعه توزیع‌های قابل نمایش توسط $\mathcal{G}_2$ یکسان باشد .

یک شبکه بیزین دارای دو مؤلفه اساسی است: یک گراف جهت‌دار فاقد دور یا DAG و متناظر با هر گره v از گراف یک جدول توزیع احتمال توأم $p(v, pa(v))$ که در آن منظور از $pa(v)$ مجموعه رئوسی مانند u بوده به‌طوری‌که (u, v) یک یال از گراف است. یک شبکه بیزین توزیع توأم p روی مجموعه متغیرهای $V = \{X_1, \dots, X_n\}$را کد می‌کند و آن را به‌صورت یک حاصل‌ضرب از توزیع احتمالات شرطی روی هر متغیر با توجه به والدین آن در گراف به‌صورت زیر تجزیه می‌کند:

$$p(X_1, \dots, X_n) = \prod_{i=1}^{n} p(X_i \mid pa(X_i)) \quad (3)$$

با توجه به حاصل‌ضرب بالا، اگر Z مجموعه‌ای از گره‌های غیرنسل[8] گره X_i به جزء گره‌های موجود در $pa(X_i)$ باشد، آنگاه شرط محلی مارکوف[9] (خاصیت مارکوف) به‌صورت زیر برقرار است:

$$p(X_i \mid pa(X_i), Z) = p(X_i \mid pa(X_i)) \quad (4)$$

اگرچه در دنیای واقعی داده‌ها اغلب به‌صورت ترکیبی از متغیرها با مقادیر گسسته و پیوسته است، بااین‌حال بسیاری از الگوریتم‌های یادگیری ساختار شبکه‌های بیزین فرض را بر این می‌گذارند که تمامی متغیرهای تصادفی گسسته هستند. بنابراین، متغیرهای پیوسته معمولاً هنگام یادگیری یک شبکه بیزین به متغیرهای گسسته تبدیل می‌شوند. برخی روش‌ها برای توسعه الگوریتم‌های یادگیری شبکه بیزین به متغیرهای پیوسته وجود دارد. در تحقیق قبلی[۲۲] ، ما یک الگوریتم بهینه‌سازی جدید برای حل مسائل بهینه‌سازی پیوسته معرفی کرده‌ایم که تیمی از آتوماتای یادگیر را برای ساختن یک شبکه بیزین به کار می‌گیرد. این تیم از آتوماتای یادگیر سعی می‌کند ساختار مطلوب شبکه بیزین را در طول اجرای الگوریتم فرا بگیرد. استفاده از آتوماتای یادگیر منجر به یک الگوریتم با زمان محاسبه کمتر برای ساخت شبکه بیزین می‌شود. در این مقاله فرض می‌کنیم که همه داده‌ها گسسته بوده و بدون هیچ‌گونه مقادیر گمشده هستند.

نوآوری این مقاله در ارائه یک مدل ساده، سریع و مقیاس‌پذیر برای حل مسئله یادگیری ساختار در شبکه‌های بیزین است. ما به‌جای استفاده از تکنیک‌های کُند برنامه‌ریزی خطی صحیح که قادر به حل مسائل با اندازه‌های نسبتاً بزرگ نیستند، با ترکیب روش برنامه‌ریزی خطی و گردکردن قطعی مدلی ارائه می‌کنیم که در مقایسه با روش‌های برنامه‌ریزی خطی صحیح، علاوه بر سریع بودن جواب‌های نسبتاً قابل قبولی نیز به دست می‌دهد.

این مقاله به‌صورت زیر سازماندهی شده است: ابتدا در بخش ۲ مروری بر روش‌های یادگیری ساختار در شبکه‌های بیزین و بررسی دسته‌بندی‌های آن انجام می‌شود. سپس در بخش ۳ روش برنامه‌ریزی خطی برای حل مسئله یادگیری ساختار در شبکه‌های بیزین معرفی و روش‌های مبتنی بر آن مرور می‌شود. در بخش ۴ روش پیشنهادی ما برای حل مسئله یادگیری ساختار بیان و شرح داده می‌شود. بخش ۵ نتایج تجربی حاصل از مقایسه روش پیشنهادی ما با سایر روش‌های مطرح در این زمینه نشان می‌دهد. نهایتاً، در بخش ۶ نتیجه‌گیری کار انجام شده به همراه پیشنهاداتی برای کارهای آتی را ارائه می‌گردد.

۲- یادگیری ساختار در شبکه‌های بیزین

مسئله یادگیری ساختار شبکه‌های بیزین یک مسئله اِن‌پی سخت است [۵]. برای یک مجموعه شامل n متغیر، تعداد $O(n2^{n(n-1)})$ گراف جهت‌دار بدون دور وجود دارد [۲۵]. به‌عبارت‌دیگر رشد اندازه فضای جواب برحسب تعداد متغیرها از مرتبه‌ی نمایی است. پیچیدگی روش‌های یادگیری ساختار در شبکه‌های بیزین کاملاً به فرضیاتی نظیر اینکه مجموعه داده دارای مقادیر گم‌شده بوده یا نه و اینکه متغیرهای پنهان وجود دارند یا خیر وابسته است. در حالتی که مجموعه داده دارای مقادیر گم‌شده یا متغیرهای پنهان باشد، الگوریتمی کارآمد به نام EM ساخت‌یافته[10] معرفی شده است که با ترکیب الگوریتم استاندارد بیشینه‌سازی امید[11] یعنی EM برای بهینه‌سازی پارامترها با جستجوی ساختار برای انتخاب مدل مسئله مقادیر گم‌شده را حل می‌کند [۱۳, ۱۴]. به‌طورکلی می‌توان روش‌های یادگیری ساختار در مدل‌های گرافی احتمالاتی را به سه دسته روش‌های مبتنی بر محدودیت[12]، روش‌های مبتنی بر امتیاز[13] و روش‌های ترکیبی[14] دسته‌بندی نمود که در ادامه به‌طور مختصر به شرح هر دسته خواهیم پرداخت.

۲-۱- روش‌های مبتنی بر محدودیت

در این روش به شبکه بیزین به‌صورت نمایشی از استقلال‌های شرطی نگریسته می‌شود. در این دسته از روش‌ها سعی بر این است که با انجام آزمون استقلال شرطی روی داده‌ها وابستگی و یا عدم وابستگی میان متغیرها بررسی شده و از آن برای یافتن یک شبکه (و یا به عبارت دقیق‌تر یک کلاس هم‌ارزی از شبکه‌ها) که بهترین توصیف از مجموعه وابستگی‌ها و عدم وابستگی‌های محاسبه شده را به دست می‌دهد، استفاده شود. این روش‌ها به‌طورکلی شهودی هستند: آن‌ها مسئله یافتن ساختار را از مفهوم استقلال تفکیک کرده و بیشتر به دنبال تعریف شبکه بیزین هستند. متأسفانه این روش‌ها می‌توانند حساس به شکست در یک آزمون موردی باشند. برگشت نتیجه‌ای نادرست توسط یکی از آزمون‌های تعیین وابستگی منجر به گمراهی در ایجاد شبکه بیزین خواهد شد. روش‌های مبتنی بر محدودیت شامل دو مرحله است: مرحله آزمون استقلال شرطی و مرحله جهت‌دهی یال‌ها. یکی از الگوریتم‌های پیشرو در این زمینه، الگوریتم PC[۱۵] است[۲۷]. الگوریتم PC با یک گراف بدون جهت کاملاً همبند آغاز می‌شود. در ادامه با انجام آزمون استقلال شرطی میان تمامی زوج متغیرها، اگر نتیجه آزمون استقلال شرطی میان دو متغیر X_i و X_j مثبت بود، یال متناظر با آن را از گراف حذف کرده و مجموعه شرطی که استقلال روی آن حفظ می‌شود را نگهداری می‌کند. اجرای روال آزمون استقلال شرطی با یک مجموعه متغیرهای شرطی که اندازه آن مرتباً در حال افزایش است تکرار می‌شود. در گام بعدی و برای جهت‌دهی یال‌ها، الگوریتم به دنبال v-ساختار شکل در گراف می‌گردد. این ساختار میان هر سه‌تایی از رئوس A، B و C که میان A و B و نیز میان B و C یال وجود دارد شکل می‌گیرد. یک ساختارv- شکل با جهت‌دهی یال از A به B و یک یال از C به B شکل می‌گیرد هرگاه یک مجموعه شرط S وجود داشته باشد به‌طوری‌که A مستقل از B بوده و C در S نباشد. پس‌ازاینکه تمامی v-ساختارهای شکل تعیین شدند، الگوریتم PC به‌صورت تکراری الگوهای گره/یال/یال جهت‌دار را با استفاده از دو قاعده زیر بررسی می‌کند. اگر یکی از این دو قاعده درست دربیاید آن را اجرا و ساختار شبکه را اصلاح می‌کند:

- اگر یک یال جهت‌دار از $\boldsymbol{A}$ به $\boldsymbol{B}$ وجود داشته و یک یال میان $\boldsymbol{B}$ و $\boldsymbol{C}$ وجود دارد و هیچ یالی میان $\boldsymbol{A}$ و $\boldsymbol{C}$ نیست، یال میان $\boldsymbol{B}$ و $\boldsymbol{C}$ را با یک یال جهت‌دار جایگزین کرده هرگاه این موجب ایجاد یک v-ساختار شکل جدید نشود.
- اگر یک مسیر جهت‌دار از $\boldsymbol{A}$ به $\boldsymbol{B}$ وجود داشته باشد، این به آن معنی است که یک توالی از یال‌های جهت‌دار وجود دارد که از $\boldsymbol{A}$ شروع شده و به $\boldsymbol{B}$ ختم می‌شود و یک یال میان $\boldsymbol{A}$ و $\boldsymbol{B}$ وجود دارد، آنگاه یال را با یک یال جهت‌دار جایگزین می‌کنیم.

درنهایت اگر امکان اعمال هیچ‌یک از این قواعد نباشد، الگوریتم، ساختار فعلی را به‌عنوان ساختار نهایی تولید می‌کند. الگوریتم ممکن است یک شبکه‌ی بیزین را تولید نکند اما یک گراف جهت‌دار بدون دور جزئی را تولید کند. یک گراف جهت‌دار فاقد دور جزئی گرافی است که هم شامل یال‌های جهت‌دار و هم شامل یال‌های بدون‌جهت بوده و زیرگراف‌های جهت‌دار آن فاقد دور جهت‌دار است. از سایر روش‌های این دسته می‌توان به الگوریتم‌های Grow-Shrink [۲۱] و Inter-IAMB [۳۳] اشاره کرد.

اخیراً روش WMIF[۱۶] معرفی شده است که در آن الگوریتم PC از دو جنبه یکپارچه می‌شود: یک راهبرد حذف لبه مبتنی بر اطلاعات متقابل و یک سری تنظیم شرط مبتنی بر اطلاعات متقابل. منظور از راهبرد حذف لبه مبتنی بر اطلاعات متقابل این است که همیشه یک جفت متغیر را با ضعیف‌ترین اطلاعات متقابل برای آزمایش استقلال شرطی انتخاب می‌شود. راهبرد تولید مجموعه شرط نیز ایجاد یک مجموعه شرطی است که در آن متغیرهایی که دارای درجه ضعیف‌تر اطلاعات متقابل با متغیر هدف هستند همیشه اول در نظر گرفته می‌شوند[۲۴].

۲-۲- روش‌های مبتنی بر امتیاز

در روش‌های مبتنی بر امتیاز، به شبکه بیزین به‌صورت تعیین یک مدل آماری نگریسته شده و درنتیجه مسئله یادگیری ساختار به‌صورت مسئله انتخاب مدل در نظر گرفته می‌شود. تمامی این روش‌ها مبتنی بر یک اصل مشترک هستند: ما یک فضای فرضیه[۱۷] از مدل‌های بالقوه- مجموعه‌ای از ساختارهای شبکه ممکن که می‌خواهیم در نظر بگیریم- را همراه با یک تابع امتیاز[۱۸] که میزان تطابق مدل با داده‌های مشاهده شده را نشان می‌دهد، تعریف می‌کنیم. حال تنها کار باقی‌مانده یافتن یک ساختار با بیشترین میزان امتیاز است [۱۷, ۶]. فضای شبکه‌های بیزین یک فضای ترکیبیاتی[۱۹] بوده که تعداد ساختارها در آن از مرتبه نمایی $2^{O(n^2)}$ است. درنتیجه حتی باوجود یک تابع امتیاز، یافتن یک ساختار با بیشترین امتیاز کاری ساده نیست. متأسفانه مسئله یادگیری ساختار شبکه‌های بیزین مبتنی بر امتیاز نیز یک مسئله اِن‌پی سخت بوده و درنتیجه مجبور به استفاده از روش‌های جستجوی ابتکاری[۲۰] هستیم [۵]. یکی از مشکلات این روش‌ها این است که آن‌ها یک مسئله جستجو را طرح می‌کنند که ممکن است راه‌حلی سرراست و کارآمد نداشته باشد.

یک تابع امتیاز، میزان جور بودن ساختار به‌دست‌آمده با مجموعه داده‌های آموزشی را نشان می‌دهد. برخی از توابع امتیاز متداول عبارت‌اند از: امتیاز بیز[۲۱]، طول توصیف کمینه[۲۲] یا MDL و معیار اطلاعات بیزی[۲۳] یا BIC . تمامی این توابع به‌صورت مجانبی هم‌ارز و درست هستند به این معنا که با افزایش تعداد نمونه‌های آموزشی، توزیع کد شده توسط شبکه فراگرفته شده با احتمال ۱ به توزیع درستی که نمونه‌ها از آن نمونه‌برداری شده‌اند، همگرا می‌شود. طول توصیف کمینه یا MDL یک معیار از نظریه اطلاعات بوده که تمایل آن به مدل‌هایی است که کوتاه‌ترین توصیف را برای مجموعه آموزشی فراهم می‌کنند. این معیار برای شبکه بیزین $BN = (G, \Theta)$و مجموعه داده‌های آموزشی $D = \{d_1, \dots, d_N\}$ به‌صورت زیر تعریف می‌شود:

$$MDL(G, \Theta|D) = \frac{\log N}{2}|\Theta| - \log P(D|\Theta, G) \qquad (۵)$$

که در آن $|\Theta|$ تعداد پارامترهای مدل در شبکه است. ازآنجاکه هر پارامتر توسط $\frac{\log N}{2}$ بیت قابل کد کردن است، اولین جمله از تعریف MDL توصیفی از طول شبکه بیزین یعنی تعداد بیت‌های لازم برای کد کردن تمامی پارامترهای آن است. همچنین، دومین جمله از تعریف MDL یعنی منفی لگاریتم درست‌نمایی مدل شبکه بیزین با توجه به D، تعداد بیت‌های لازم برای توصیف D وقتی از شبکه بیزین G استفاده می‌کنیم، است. کمینه‌سازی MDL معادل بیشینه‌سازی BIC بوده که دقیقاً منفی MDL است. به‌طور شهودی، هر دوی MDL و BIC متمایل به مدل‌هایی هستند که بهتر داده‌های آموزشی را به تصویر کشیده و از پیچیدگی نمایش کمتری برخوردار باشند. ازاین‌رو، یادگیری یک شبکه بیزین از روی مجموعه داده D را می‌توان به‌طور رسمی به‌صورت زیر تعریف کرد [۱۹]:

$$(G, \Theta) = \arg \min_{G,\Theta} MDL(G, \Theta) \qquad (۶)$$

یک روش مبتنی بر شاخه و تحدید توسط کمپس و همکاران [۱۰] و کمپس و جی [۴] ارائه شده است که در آن ابتدا با نادیده گرفتن محدودیت بدون دور بودن گراف ساختار، مجموعه والدین بهینه را برای تمامی متغیرها پیدا کرده و سپس تمامی دورهای جهت‌دار گراف ساختار را شناسایی و با شکست آن‌ها یک شبکه معتبر بیزین را به دست می‌آورد. الگوریتم به بهبود کیفیت ساختار شبکه بیزین ادامه می‌دهد که یا به یک جواب بهتر برسد و یا اینکه کران بالا کاهش یابد و این روند تا زمان رسیدن به جواب بهینه ادامه می‌یابد. این روش قادر به حل مسائلی تا ۱۰۰ متغیر در زمان قابل قبول است.

بااین‌حال، بهترین روش یادگیری ساختار مبتنی بر برنامه‌ریزی خطی است [۹, ۱۸] که در آن استفاده از صفحات برشی برای شکستن دورهای جهت‌دار تا یافتن جواب بهینه ادامه می‌یابد. تعیین صفحات برشی مناسب کلید موفقیت این روش است. نتایج تجربی به‌دست‌آمده حاکی از آن است که این روش قادر به حل مسائل تا ۲۰۰ متغیر در زمان قابل قبول است. یک ایده جالب توسط تیسیر و کولر [۳۰] معرفی شده که یک روش ابتکاری ساده و درعین‌حال کارآمد بر پایه جستجوی مونت کارلو روی فضای ترتیب متغیرها است. به‌عنوان یک روش تقریبی، این روش با سایر روش‌های دقیق زمانی که اندازه مسئله بسیار بزرگ باشد قابل رقابت است.

از سایر روش‌های این دسته می‌توان به الگوریتم‌های SBNR[۲۴] [۲۸] و Inter-IAMB [۳۳] اشاره کرد. اخیراً نیز یک روش مبتنی بر ترتیب متغیرها ارائه شده است که از مفهوم گراف بهترین امتیاز و مؤلفه‌های همبند قوی این گراف برای یادگیر ساختار استفاده می‌کند [۳]. ازجمله روش‌های دیگر که از برنامه‌ریزی خطی صحیح برای مدل کردن یادگیری ساختار استفاده می‌کنند می‌توان به الگوریتم‌های ارائه شده توسط جیمز گاسنس [۹] و جاکولا [۱۸] اشاره کرد.

۳-۲- روش‌های ترکیبی

دسته سوم روش‌هایی هستند که تلفیقی از دو روش قبلی بوده و از مزایای هرکدام از آن‌ها استفاده می‌کنند. روش‌های ترکیبی مبتنی بر دو گام هستند: یک گام محدود کننده که در آن با استفاده از راهبردهای مبتنی بر محدودیت فضای جستجوی گراف‌های جهت‌دار بدون دور کاهش یافته و در گام بیشینه‌سازی با استفاده از راهبردهای مبتنی بر امتیاز جستجو بر روی فضای محدود شده جهت یافتن گراف جهت‌دار بدون دور بهینه دنبال می‌شود. یکی از بهترین روش‌های این دسته الگوریتم MMHC[۲۵] است[۳۱]. این الگوریتم ابتدا گراف زمینه شبکه بیزین که در آن یال‌ها بدون جهت هستند را با استفاده از یک الگوریتم اکتشافی محلی به نام MMPC[۲۶] به دست می‌آورد. سپس با استفاده از روش جستجوی حریصانه تپه‌نوردی برای جهت دادن به یال‌ها بهره می‌گیرد. این الگوریتم ابتدا مجموعه والدین و فرزندان هر یک از متغیرها را مشخص می‌کند ، سپس یک جستجوی تپه‌نوردی حریصانه در فضای جستجوی شبکه‌های بیزی انجام می‌دهد. جستجو با یک گراف تهی آغاز می‌شود. علاوه بر این ، افزودن یال، حذف یا تغییر جهت یال که منجر به بیشترین امتیاز گراف ساختار می‌شود نیز صورت گرفته و جستجو به‌صورت بازگشتی به طریق مشابه ادامه می‌یابد. تفاوت مهم آن با روش جستجوی حریصانه استاندارد در این است که جستجو تنها محدود به یالی است که در گام اول توسط MMPC کشف شده است که در صورت بهبود امتیاز به گراف ساختار اضافه می‌شود . از سایر روش‌های این دسته می‌توان به الگوریتم‌های RSMAX2 [۲۶] و H2PC [۱۶] اشاره کرد.

۳- یادگیری ساختار در شبکه‌های بیزین

استفاده از روش برنامه‌ریزی خطی یکی از متداول‌ترین روش‌ها برای برخورد با مسائل بهینه‌سازی اِن‌پی کامل است. در این روش ابتدا مسئله موردنظر در قالب یک مسئله برنامه‌ریزی خطی صحیح مدل شده و سپس به یک مسئله خطی تعدیل می‌شود. سپس این مسئله خطی حل شده و از روی جواب‌های به‌دست‌آمده برای آن یک جواب صحیح با استفاده از روش‌هایی نظیر صفحات برشیcutting planes، شاخه و تحدید و گردسازی قطعی و تصادفی برای مسئله اصلی استخراج می‌شود[۷و۹].

مهم‌ترین چالش پیش رو برای مدل‌سازی مسئله یادگیری ساختار در شبکه‌های بیزین تعیین قید بدون دور بودن گراف ساختار شبکه است. به‌طورکلی سه دسته قید یا محدودیت خطی برای اعمال بدون دور بودن گراف ساختار وجود دارد که عبارت‌اند از: قیود خوشه، قیود ترتیب گسسته متغیرها و قیود ترتیب پیوسته که در ادامه به توضیح هرکدام می‌پردازیم.

۳-۱- قیود خوشه

یک زیرمجموعه $C \subseteq V$ از گره‌های یک گراف جهت‌دار یک خوشه[۲۷] نامیده می‌شود. این با مفهوم خوشه در گراف‌های بدون جهت فرق دارد. خوشه در گراف‌های بدون جهت یک زیرگراف کامل است درحالی‌که خوشه در گراف‌های جهت‌دار صرفاً یک زیرمجموعه از گره‌های گراف است. قیود خوشه‌ای مبتنی بر این مشاهده هستند که در هر خوشه از یک گراف جهت‌دار بدون دور، می‌بایست دست‌کم یک گره مانند uوجود داشته باشد به‌طوری‌که مجموعه والدین آن یعنی S_u جدای از خوشه باشد. یادآوری این نکته ضروری است که مجموعه والدین هر گره می‌تواند بنا بر تعریف مجموعه‌ای تهی باشد و مجموعه تهی جدا از هر مجموعه دیگری است. قیود خوشه برای اعمال بدون دور بودن[۲۸] در گراف‌های جهت‌دار اولین بار توسط جاکولا و همکارانش[۱۸] معرفی شد و سپس توسط گاسن و بارتلت[۱, ۲] توسعه یافت.

برای بیان این قید متغیر دودویی z_{S_u} را معرفی می‌کنیم که مقدار آن برابر یک است اگر و تنها اگر S مجموعه والدین u باشد. با توجه به این تعریف محدودیت خوشه تعریف شده در بالا را می‌توان به‌صورت قید خطی زیر بیان کرد:

$$\sum_{u \in C} \sum_{S: S \cap C = \emptyset} z_{S_u} \geq 1, \quad \forall C \subseteq V \tag{۷}$$

که در آن V مجموعه گره‌های گراف است. افزون بر این ما نیازمند این قید هستیم که هر گره دقیقاً دارای یک مجموعه والدین باشد که به‌راحتی توسط قید خطی زیر قابل‌بیان است:

$$\sum_{S \in \mathcal{F}_u} z_{S_u} = 1, \quad \forall u \in V \tag{۸}$$

این دو مجموعه از قیود خطی روی مجموعه متغیرهای دودویی برای اینکه بتوانند گراف ساختار یک شبکه بیزین را نمایش بدهند کفایت می‌کند. درصورتی‌که $|V| = n$ باشد، تعداد قیود خوشه برابر 2^nبوده و بنابراین ساده نیست که بتوان تمامی آن‌ها را به‌عنوان قید خطی در یک برنامه‌ی خطی اضافه کرد. درنتیجه آن‌ها به‌صورت صفحات برشی در الگوریتم شاخه و تحدید اضافه می‌شوند.

۳-۲- ترتیب گسسته متغیرها

استفاده از مدل کردن ترتیب گسسته متغیرها برای یادگیری ساختار شبکه‌های بیزین اولین بار توسط جیمز گاسنز [۸] پیشنهاد شد. در هر ترتیب توپولوژیک از گره‌های گراف، اگر (u, v) یالی از گراف باشد آنگاه بایستی گره u قبل از گره v در ترتیب ظاهر شود. ما می‌دانیم که یک ترتیب توپولوژیک از گره‌های یک گراف جهت‌دار وجود دارد اگر و تنها اگر گراف جهت‌دار بدون دور باشد. معرفی یک ترتیب توپولوژیک گره‌ها را می‌توان با استفاده از یک مجموعه متغیر کمکی بیان کرد. برای گراف جهت‌دار $G = (V, E)$ ، به ازای هر زوج مرتب از گره‌های u و v ، متغیر ترتیب گسسته o_{uv} را به‌صورت زیر معرفی می‌کنیم:

$$o_{uv} = \begin{cases} 1 & if\ u\ be\ before\ v\ in\ order\ O. \\ 0 & \text{otherwise} \end{cases} \tag{۹}$$

برای اینکه یک ترتیب از گره‌های گراف ترتیبی توپولوژیک باشد ما نیازمند ارائه یک سری قیود روی آن‌ها و مجموعه والدینشان هستیم. همان‌طور که در قید خوشه لازم بود که هر گره دقیقاً یک مجموعه والد داشته باشد، در اینجا این‌گونه است. به‌عبارت‌دیگر، به ازای هر زوج $u, v \in V$ از مجموعه رئوس گراف یا u باید قبل از

$v \prec u$) و یا v بایستی قبل از u در ترتیب توپولوژیک قرار بگیرد ($u \prec v$) که این محدودیت را می‌توان به‌صورت قید خطی زیر بیان کرد:

$$o_{uv} + o_{vu} = 1 \quad \forall u, v \in V \tag{۱۰}$$

علاوه بر این لازم است ترتیب توپولوژیک گره‌ها متعدی[۲۹] نیز باشد به این معنی که اگر $u \prec v$ و $v \prec w$ آنگاه $u \prec w$. این محدودیت را هم می‌توان به‌صورت قید خطی زیر بیان کرد:

$$o_{uv} + o_{vw} + o_{wu} \leq 2 \quad \forall u, v, w \in V \tag{۱۱}$$

درنهایت، با توجه به تعریف ترتیب توپولوژیک، اگر (u, v) یالی از گراف باشد، آنگاه بایستی گره u قبل از گره v در ترتیب توپولوژیک باشد. این محدودیت را می‌توان در قالب قید خطی زیر بیان کرد:

$$\sum_{S \in \mathcal{F}_v : u \in S} z_{S_v} - o_{uv} \leq 0 \quad \forall u, v \in V \tag{۱۲}$$

نکته قابل‌توجه آن است که اگر یال (u, v) در گراف وجود داشته باشد، آنگاه یک متغیر دودویی z_{S_u} که $v \in S$ وجود دارد که مقدار آن برابر ۱ است. قید بالا تضمین می‌کند اگر این‌گونه باشد آنگاه $o_{uv} = 1$ بوده و درنتیجه گره u در ترتیب توپولوژیک قبل از گره v است و این همان چیزی است که ما می‌خواهیم. از طرفی دیگر اگر یال (u, v) در گراف وجود نداشته باشد، آنگاه بایستی داشته باشیم:

$$\sum_{S : u \in S} z_{S_v} = 0 \tag{۱۳}$$

با توجه به مطالب عنوان شده بالا، مجموعه قیود خطی مورد نیاز برای مدل کردن ترتیب گسسته گره‌های گراف به‌صورت زیر است:

$$\begin{aligned}
&\sum_{S \in \mathcal{F}_u} Z_{S_u} = 1, \qquad \forall u \in V \\
&o_{uv} + o_{vu} = 1 \qquad \forall u, v \in V \\
&o_{uv} + o_{vw} + o_{wu} \leq 2 \qquad \forall u, v, w \in V \\
&\sum_{S \in \mathcal{F}_v : u \in S} z_{S_v} - o_{uv} \leq 0 \qquad \forall u, v \in V \\
&z_{S_u} \in \{0,1\}, \qquad \forall u \in V, S \in \mathcal{F}_u \\
&o_{uv} \in \{0,1\} \qquad \forall u, v \in V
\end{aligned} \tag{۱۴}$$

درصورتی‌که $|V| = n$ باشد، مدل کردن ترتیب گسسته گره‌ها تعداد $O(n^3)$ قید خطی از نوع قید اضافه کرده و از تعداد $n^2 - n$ متغیر کمکی استفاده می‌کند. ازآنجایی‌که تعداد قیود خطی از مرتبه چندجمله‌ای است، آن را برای اینکه تمامی قیود به ریشه درخت جستجو اضافه شوند مناسب می‌کند.

۳-۲- ترتیب پیوسته متغیرها

مشابه آنچه برای ترتیب گسسته گفته شد، ترتیب پیوسته متغیرها نیز مبتنی بر یک ترتیب توپولوژیک از گره‌های گراف جهت‌دار است. این مدل‌سازی از مسئله اولین بار توسط پهارز و فرازنز پرنکف [۲۳] معرفی و توسط فراهانی و لاگرگرین [۱۲] مورد استفاده قرار گرفت. ترتیب توپولوژیک با استفاده از یک مجموعه از متغیرهای کمکی بیان می‌شوند. به ازای هر گرهی مانند u، متغیر ترتیب پیوسته p_u که $p_u \in [0,1]$ معرفی می‌شود. حال، متغیرهای ترتیب پیوسته یک ترتیب توپولوژیک از گره‌های یک گراف جهت‌دار را این‌گونه بیان می‌کنند: اگر $p_u < p_v$ آنگاه گره u قبل از گره v در هر ترتیب توپولوژیک گره‌های گراف قرار دارد. مانند آنچه در بخش قبل بیان شد لازم است که هر گره دارای تنها یک مجموعه والدین باشد. علاوه بر آن، اگر (u, v) یالی از گراف باشد بایستی گره u قبل از گره v در ترتیب توپولوژیک ظاهر شود. بنابراین مجموعه قیود خطی زیر لازم است به مدل اضافه شوند:

$$\sum_{S \in \mathcal{F}_v : u \in S} z_{S_v} - p_v + p_u < 1 \quad \forall u, v \in V \tag{۱۵}$$

با این توضیحات مجموعه قیود خطی لازم برای فرموله کردن یک ترتیب پیوسته از متغیرها به‌صورت زیر است:

$$\begin{aligned}
&\textstyle\sum_{S \in \mathcal{F}_u} Z_{S_u} = 1, \quad \forall u \in V \\
&\textstyle\sum_{S \in \mathcal{F}_v : u \in S} z_{S_v} - p_v + p_u < 1 \quad \forall u, v \in V \\
&z_{S_u} \in \{0,1\}, \quad \forall u \in V, S \in \mathcal{F}_u \\
&p_u \in [0,1] \quad \forall u \in V
\end{aligned}$$

قابل‌توجه اینکه تعداد متغیرهای کمکی اضافه شده در این حالت برابر $n = |V|$ است. تعداد قیود خطی نیز در این فرمولاسیون برابر است با:

$$O\left(n \max_{u \in V} |\mathcal{F}_u|\right) \tag{۱۶}$$

این کران روی تعداد قیود آن را برای این هدف که تمامی قیود به ریشه درخت جستجو اضافه شوند را مناسب می‌سازد.

۴- یادگیری ساختار با استفاده از روش گرد کردن قطعی

مسائل بهینه‌سازی ترکیبیاتی اِن‌پی-سخت شامل آن دسته از مسائلی هستند که بعید به نظر می‌رسد برای آن‌ها راه‌حل کارآمدی وجود داشته باشد. بیشتر این مسائل را می‌توان به‌صورت یک مسئله برنامه‌ریزی خطی صحیح[۳۰] یا ILP مدل کرد. درواقع، در بسیاری از آن‌ها مقادیر متغیرهای صحیح تنها یکی از دو مقدار صفر یا یک است. بنابراین ما با یک مسئله برنامه‌ریزی صحیح صفر-یک مواجه هستیم. با تعدیل یک مسئله برنامه‌ریزی خطی صحیح صفر-یک، ما درواقع قیود $x_i \in \{۰,۱\}$ را به قیود خطی به شکل $۰ \leq x_i \leq ۱$ تبدیل می‌کنیم. با چنین تبدیلی ما یک مسئله برنامه‌ریزی خطی داریم که در زمان چندجمله‌ای قابل حل است. این ارتباط تنگاتنگ میان برنامه‌ریزی خطی صحیح صفر-یک (0-1 ILP) و برنامه‌ریزی خطی (LP)این امکان را برای ما فراهم می‌کند تا بتوانیم برای مسائل اِن‌پی-سخت الگوریتم‌های تقریبی کارآمد طراحی کنیم. یک جواب شدنی برای برنامه‌ریزی خطی تعدیل شده را می‌توان به‌صورت یک جواب کسری برای مسئله اصلی تصور کرد. مجموعه جواب‌های شدنی برای یک سیستم از قیود خطی را به‌عنوان یک پلی‌توپ[۳۱] شناخته می‌شوند. جستجوی یک جواب بهینه با توجه به یک تابع خطی در یک پلی‌توپ کاری دشوار به نظر نمی‌رسد زیرا ثابت شده است که نقاط بهینه روی برخی از رئوس پلی‌توپ وقع شده‌اند. بااین‌حال در مورد مسائل اِن‌پی-سخت ما نمی‌توانیم انتظار داشته باشیم که پلی‌توپ تعریف کننده مجموعه جواب‌های بهینه دارای رئوس صحیح باشند. درنتیجه تنها کاری که لازم است انجام دهیم تبدیل جواب بهینه برنامه‌ریزی خطی تعدیل شده به یک جواب بهینه صحیح نزدیک است. یک روش اساسی برای به دست آوردن الگوریتم‌های تقریبی با استفاده از برنامه‌ریزی خطی همان چیزی است که تحت عنوان گردسازی برنامه‌ریزی خطی شناخته می‌شود و آن شامل مراحل زیر است:

۱. ابتدا برنامه خطی تعدیل شده را حل می‌کنیم.
۲. سپس جواب‌های به‌دست‌آمده کسری(ناصحیح) را به جواب‌های صحیح تبدیل می‌کنیم.
۳. اطمینان حاصل می‌کنیم که جواب تابع حد خیلی افزایش نیابد.

به‌طورکلی دو روش برای گرد کردن جواب حاصل از برنامه خطی تعدیل شده وجود دارد که عبارت‌اند از: گردسازی قطعی[۳۲] و گردسازی تصادفی[۳۳]. گردسازی قطعی یکی از سرراست‌ترین روش‌ها برای تبدیل جواب‌های کسری به مقادیر صحیح است. ساده‌ترین راه برای گردسازی یک جواب کسری به یک جواب صحیح که در آن تمام مقادیر صفر یا یک هستند این است که متغیرهایی با مقادیر نسبتاً بزرگ را یک بگیریم درحالی‌که سایر موارد دیگر را به صفر گرد کنیم. به‌عنوان‌مثال مقادیر بزرگ‌تر یا مساوی $\frac{۱}{۲}$ را به یک و مقادیر کم‌تر از $\frac{۱}{۲}$ را به صفر گرد کنیم. روش دیگر برای گردسازی، گردسازی تصادفی است. یک ایده معمول برای گرد کردن جواب بهینه کسری مشاهده مقادیر کسری به‌عنوان احتمالات است. در این صورت ما می‌توانیم از پرتاب سکه با این احتمالات به‌عنوان اریب استفاده کرده و بر اساس آن گرد کنیم. به‌عنوان‌مثال اگر در جواب بهینه $x_i = ۲/۳$ به دست آمد می‌توانیم این‌گونه تصور کنیم که مقدار متغیر x_i با احتمال $۲/۳$ برابر یک و با احتمال $۱ - ۲/۳ = ۱/۳$ برابر صفر است. این ایده گردسازی تصادفی نامیده می‌شود.

ما در این بخش به شرح روش پیشنهادی یعنی یادگیری ساختار با استفاده از روش برنامه‌ریزی خطی صحیح و گردسازی قطعی می‌پردازیم. ما ابتدا مسئله یادگیری ساختار را با استفاده از محدودیت‌های خوشه ارائه شده توسط جاکولا و همکارانش[۱۸] و گاسن و بارتلت[۱، ۲] به‌صورت یک برنامه خطی مدل کرده و سپس مدل به‌دست‌آمده را توسط روش گردسازی قطعی حل می‌کنیم. همان‌طور که قبلاً اشاره شد یک خوشه C از گراف جهت‌دار و بدون دور $G = (V, E)$ زیرمجموعه‌ای از یال‌های آن است. محدودیت خوشه به ما می‌گوید که در هر خوشه از G دست‌کم گرهی مانند u وجود دارد که مجموعه والدین آن یعنی S_u جدای از خوشه هستند. برای بیان این محدودیت متغیر دودویی z_{S_u} را به‌صورت زیر معرفی می‌کنیم که مقدار آن برابر یک است اگر و تنها اگر S مجموعه والدین u باشد.

$$z_{S_u} = \begin{cases} 1 & \text{if } S \text{ be parent set of } u \\ 0 & \text{otherwise} \end{cases} \quad (۱۷)$$

با توجه به این تعریف قید خوشه تعریف شده در بالا را می‌توان به‌صورت قید خطی زیر بیان کرد:

$$\sum_{u \in C} \sum_{S: S \cap C = \emptyset} z_{S_u} \geq 1, \quad \forall C \subseteq V \quad (۱۸)$$

که در آن V مجموعه گره‌های گراف است. افزون بر این ما نیازمند این محدودیت هستیم که هر گره دقیقاً دارای یک مجموعه والدین باشد که به‌راحتی توسط قید خطی زیر قابل‌بیان است:

$$\sum_{S \in \mathcal{F}_u} z_{S_u} = 1, \quad \forall u \in V \quad (۱۹)$$

این دو مجموعه از قیود خطی روی مجموعه متغیرهای دودویی برای اینکه بتوانند گراف ساختار یک شبکه بیزین را نمایش بدهند کفایت می‌کند. فرض کنید c_{S_u} امتیاز محلی BIC گره u با مجموعه والدین S باشد. در این صورت، با توجه به مطالب عنوان شده بالا مدل کامل برنامه‌ریزی خطی صحیح برای مسئله یادگیری ساختار به‌صورت زیر است:

$$\begin{aligned} \text{maximize} \quad & \sum_{u \in V, S \in \mathcal{F}_u} z_{S_u} . c_{S_u} \\ \text{subjectto} \quad & \sum_{S \in \mathcal{F}_u} z_{S_u} = 1, && \forall u \in V \\ & \sum_{u \in C} \sum_{S: S \cap C = \emptyset} z_{S_u} \geq 1, && \forall C \subseteq V \\ & z_{S_u} \in \{0,1\}, && \forall u \in V, S \in \mathcal{F}_u \end{aligned} \quad (۲۰)$$

نسخه تعدیل شده برنامه خطی بالا نیز به‌صورت زیر است:

$$\begin{aligned} \text{maximize} \quad & \sum_{u \in V, S \in \mathcal{F}_u} z_{S_u} . c_{S_u} \\ \text{subjectto} \quad & \sum_{S \in \mathcal{F}_u} z_{S_u} = 1, && \forall u \in V \\ & \sum_{u \in C} \sum_{S: S \cap C = \emptyset} z_{S_u} \geq 1, && \forall C \subseteq V \\ & 0 \leq z_{S_u} \leq 1, && \forall u \in V, S \in \mathcal{F}_u \end{aligned} \quad (۲۱)$$

روش پیشنهادی ما برای حل مسئله یادگیری ساختار به این صورت است که ابتدا نسخه تعدیل شده برنامه خطی به‌دست‌آمده را حل می‌کنیم. فرض می‌کنیم $z^*_{S_u}$ به‌عنوان بخشی از جواب بهینه کسری به‌دست‌آمده باشد. با فرض اینکه تعداد متغیرها برابر n باشد، در قیود دسته اول که مربوط به مجموعه والدین هر گره است، به ازای هر متغیر تعداد ۲^{n-1} مجموعه والد کاندید وجود دارد و ازآنجایی‌که مجموع متغیرهای z_{S_u} بایستی برابر یک باشد، لذا ضمانت می‌شود که دست‌کم یک متغیر دارای مقدار بزرگ‌تر یا مساوی ۲^{n-1} وجود دارد.

پس روش پیشنهادی برای گرد کردن جواب کسری به‌دست‌آمده به این صورت است: اگر مقدار آن بزرگ‌تر یا مساوی $\frac{۱}{۲^{n-1}}$ باشد آن را به مقدار یک گرد کرده و S_u را به‌عنوان مجموعه والدین گره u در نظر می‌گیریم و در غیر این صورت مقدار آن را به صفر گرد می‌کنیم. درصورتی‌که بیش از یک متغیر با مقدار بزرگ‌تر یا مساوی $\frac{۱}{۲^{n-1}}$ وجود داشته باشد، تنها مقداری را به یک گرد می‌کنیم که مجموعه والدین دارای کم‌ترین تعداد نسبت به سایر متغیرها باشد. زیرا در یادگیری شبکه‌های بیزین بیشتر به دنبال ساختار با پیچیدگی کمتر هستیم تا روش‌های استنباط روی آن‌ها بتوانند در زمانی مناسب اجرا شوند. همچنین، قیود دسته دوم فاقد دور بودن گراف ساختار را ضمانت می‌کنند زیرا آن‌ها به‌نوعی قیود دسته اول را در خود دارند. درنتیجه با گرد کردن قطعی به یک گراف ساختار معتبر به دست می‌آوریم که فاقد هرگونه دور جهت‌دار است. الگوریتم ۱ شبه کد روش پیشنهادی را نشان می‌دهد.

در الگوریتم ۱، ابتدا در خط ۱ برنامه خطی ۲۱ حل شده و سپس در حلقه مربوط به خطوط ۲ الی ۳، با استفاده از قاعده گرد کردن قطعی شرح داده شده در بالا، مجموعه والدین مربوط به هر متغیر مشخص می‌شود. نهایتاً و در خطوط ۵ الی ۱۰ گراف ساختار با افزودن یال‌های موردنظر شکل می‌گیرد.

مزیت عمده روش پیشنهادی در مقایسه با سایر روش‌های مبتنی بر برنامه‌ریزی خطی صحیح مقیاس‌پذیر بودن آن است. همان‌گونه که اشاره شد در حال حاضر بهترین روش یادگیری دقیق که مبتنی بر برنامه‌ریزی خطی صحیح است می‌تواند در عمل ساختارهایی تا حداکثر ۲۰۰ متغیر را فرا بگیرد. این در حالی است که در کاربردهای واقعی و به‌خصوص در حوزه بیوانفورماتیک با شبکه‌هایی سروکار داریم که تعداد گره‌های آن بعضاً به ده‌ها و حتی صدها هزار نیز می‌رسد و اینجا هست که به‌کارگیری روش‌های دقیق عملاً غیرممکن است. علاوه بر این در یادگیری دقیق ما با پدیده بیش برازش[۳۴] مواجه هستیم که خود مشکلی عمده در تعمیم مدل به داده‌های دنیای واقعی است. به هر ترتیب آنچه مرز کار ما را با سایر روش‌های ارائه

شده مشخص می‌کند در درجه اول مقیاس‌پذیری آن است. درواقع مهم‌ترین نوآوری این مقاله ارائه روشی برای یادگیری ساختار در شبکه‌های بیزین است که قادر به حل نمونه‌های بسیار بزرگ در زمان و کیفیت قابل قبول است.

Algorithm 1: Learning the structure via deterministic rounding

Input: Variables $\mathcal{X} = \{X_1, \dots, X_n\}$ and dataset $\mathcal{D}$ on $\mathcal{X}$.

Output: A Bayesian network structure G

1. Solve LP program 21 to obtain an optimal fractional solution z^*.
2. ***for*** $u \in \mathcal{X}$ ***do***
3. $\quad pa(u) = \bigcup_{S \in \mathcal{F}_u} \{z^*_{S_u} \geq \frac{1}{2^{n-1}}\}$
4. ***end***

/* Construct the structure $G = (V, E)$ */

5. ***for*** $i \leftarrow 1$ ***to*** n ***do***
6. $\quad$ Create node X_i
7. $\quad$ ***for*** $X_j \in pa(X_i)$ ***do***
8. $\quad\quad$ add directed edge (X_i, X_j) to the E
9. $\quad$ ***end***
10. ***end***
11. ***return*** G

۵- نتایج تجربی

در این بخش ما به ارائه نتایج تجربی حاصل از مقایسه روش پیشنهادی خود با دو روش دیگر روی هشت مجموعه داده گرفته شده از مخزن شبکه‌های بیزین تعبیه شده در بسته نرم‌افزاری bnlearn که در R گسترش یافته است می‌پردازیم.

۵-۱- تنظیم آزمایش‌ها

مجموعه داده‌های مورد استفاده در آزمایش‌ها به‌صورت زیر است: ما یک شبکه کوچک (کمتر از ۲۰ گره)، سه شبکه متوسط (بین ۲۰ تا ۵۰ گره)، یک شبکه بزرگ (بین ۵۰ تا ۱۰۰ گره) و یک شبکه حجیم (بیش از هزار گره) را انتخاب کرده‌ایم. جدول ۱ جزئیات اطلاعات مربوط به شبکه‌های انتخاب شده را نشان می‌دهد. ما برای هر شبکه یک نمونه تصادفی از داده‌ها را شامل ۱۰۰۰۰ نمونه آموزشی در نظر گرفته‌ایم.

۵-۲- تحلیل نتایج

ما روش پیشنهادی خود را با دو الگوریتم تپه‌نوردی [۱۵] و الگوریتم یادگیری دقیق [۱] که در بسته نرم‌افزاری Gobnilp[۳۵] پیاده‌سازی و ارائه شده است، مقایسه کرده‌ایم. الگوریتم اخیر روشی مبتنی بر امتیاز بوده که با استفاده از روش صفحات برشی[۳۶] همان مدل برنامه‌ریزی خطی صحیح ما را برای یافتن جواب دقیق حل می‌کند. برای مقایسه این سه روش جستجو، ما امتیاز بیشینه به‌دست‌آمده توسط هر الگوریتم را به همراه زمان صرف شده برای به دست آوردن این جواب گزارش کرده‌ایم که نتایج آن در جداول ۲ و ۳ منعکس شده است. تمامی آزمایش‌ها روی یک کامپیوتر corei7 با هشت مگابایت حافظه اصلی اجرا شده‌اند.

ما از کیفیت شبکه امتیاز BIC و زمان اجرا به‌عنوان مقیاس اندازه‌گیری عملکرد الگوریتم‌ها استفاده کرده‌ایم. نتایج به‌دست‌آمده نشان می‌دهد که الگوریتم پیشنهادی ما منجر به ارائه نتایجی مناسب در زمانی مناسب می‌شود. به‌این‌ترتیب که نتایج به‌دست‌آمده ازلحاظ زمانی بر هر دو روش یادگیری دقیق و تپه‌نوردی غلبه می‌کند و به لحاظ کیفیت شبکه فراگرفته شده اگرچه با روش یادگیری دقیق فاصله معناداری دارد اما تقریباً نزدیک به روش حریصانه تپه‌نوردی است. در کاربردهای دنیای واقعی به‌خصوص بیوانفورماتیک ما با نمونه‌هایی سروکار داریم که در آن تعداد متغیرها بسیار بزرگ بوده و به‌کارگیری روش‌های دقیق برای حل مسئله در زمان قابل قبول تقریباً ناممکن است. روش پیشنهادی ما اگرچه به لحاظ کیفی با جواب‌های دقیق فاصله دارد اما قادر است مسائل با اندازه بسیار بزرگ را در زمان قابل قبول حل نماید و درعین‌حال جواب‌هایی ارائه کند که با سایر روش‌های نادقیق و ابتکاری قابل رقابت باشد.

جدول ۱- مشخصات شبکه‌های بیزین مورد استفاده برای انجام آزمایش‌های تجربی

شبکه بیزین	تعداد متغیرها	تعداد یال‌ها	بیشینه تعداد والدین هر گره	تعداد پارامترها
ASIA	۸	۸	۲	۱۸
INSURANCE	۲۷	۵۲	۳	۹۸۴
MILDEW	۳۵	۴۶	۳	۵۴۰۱۵۰
ALARM	۳۷	۴۶	۴	۵۰۹
HAILFINDER	۵۶	۶۶	۴	۲۶۵۶
DIABETES	۴۱۳	۶۰۲	۲	۴۲۹۴۰۹
PIGS	۴۴۱	۵۹۲	۲	۵۶۱۸
MUNIN	۱۰۴۱	۱۳۹۷	۳	۸۰۵۹۲

جدول ۲- مقایسه امتیاز روش پیشنهادی با سایر الگوریتم‌ها روی مجموعه داده‌ها

مجموعه داده	یادگیری دقیق	تپه‌نوردی حریصانه	روش پیشنهادی
ASIA	۲۲۶۲۹,۶۷	-۲۲۶۲۹,۶۷	-۲۲۸۳۴,۱۴
INSURANCE	-۱۳۲۵۶۳,۵۰	-۱۳۴۰۱۷,۹۰	-۱۶۲۳۳۲,۲۲
MILDEW	-۴۶۲۶۴۱,۲۲	-۵۰۳۵۶۸,۳۰	-۶۵۳۴۹۰,۵۱
ALARM	-۱۰۶۲۵۱,۴۰	-۱۰۷۱۷۶,۳	-۱۱۵۳۲۲,۲۳
HAILFINDER	-۴۷۲۳۲۲,۳۱	-۴۹۸۹۱۸,۲	-۵۲۳۳۲۸,۲۱
DIABETES	-۲۶۳۵۲۷,۲۰	-۲۸۳۳۲۴,۵	-۳۰۲۲۷۶,۲۶
PIGS	-۳۰۸۳۴۱,۴۳	-۳۴۸۰۹۴,۷۱	-۳۶۲۹۸۴,۸۷
MUNIN	-۲۳۷۵۶,۱۱	-۳۴۹۰۷,۲۳	-۳۷۱۲۲,۶۸

جدول ۳- مقایسه زمان اجرای روش پیشنهادی با سایر الگوریتم‌ها روی مجموعه داده‌ها

مجموعه داده	یادگیری دقیق	تپه‌نوردی حریصانه	روش پیشنهادی
ASIA	۲,۶۰s	۰,۰۲s	۰,۰۱s
INSURANCE	۳,۵۱s	۰,۲۷s	۰,۱۷s
MILDEW	۱۴,۳۳s	۰,۳۱s	۰,۲۷s
ALARM	۲۰,۲۲s	۰,۳۹s	۰,۲۸s
HAILFINDER	۶۷,۴۰s	۰,۶۹s	۰,۳۶s
DIABETES	۳۴۵۲۲۱s	۴۸۷۱s	۳۲۵۱s
PIGS	۳۷۴۱۲۷s	۳۹۶۰s	۳۰۶۴s
MUNIN	۱۴۵۶۳۵۷s	۲۹۷۶۰۱s	۲۴۳۰۴۸s

۶- نتیجه‌گیری و کارهای آتی

این مقاله یک روش گردسازی قطعی مبتنی بر برنامه‌ریزی خطی را برای حل مسئله یادگیری ساختار در شبکه‌های بیزین ارائه می‌دهد. روش پیشنهادی ابتدا مسئله یادگیری ساختار را به‌صورت یک برنامه خطی صحیح مدل و سپس آن را به یک مسئله برنامه‌ریزی خطی تعدیل می‌کند. سپس با حل برنامه خطی تعدیل شده، جواب‌های کسری به‌دست‌آمده را با استفاده از روش معرفی شده گردسازی قطعی به جواب‌های صحیح تبدیل می‌کند. نتایج به‌دست‌آمده حاصل از انجام آزمایش‌های تجربی روی مجموعه داده‌ها نشان می‌دهد که روش پیشنهادی در کنار برتری زمان

اجرا از نظر کیفیت شبکه فراگرفته شده برحسب معیار امتیاز BIC نیز با روش‌های برتر موجود قابل رقابت است.

یک پیشنهاد برای توسعه این روش به‌عنوان کار آتی ارائه کرانی دقیق‌تر برای گرد کردن جواب‌های کسری به‌دست‌آمده توسط مدل برنامه‌ریزی خطی است. همچنین توسعه این روش برای یافتن شبکه‌هایی با محدودیت‌های ساختاری نظیر محدودیت روی تعداد والدین هر گره و یا محدودیت اندازه پوشش رأسی[۳۷] به‌عنوان کار آتی پیشنهاد می‌شود.

۷- مراجع

[1] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques.* MIT press, 2009.

[2] B. Moradabadi and H. Beigy, "A new real-coded bayesian optimization algorithm based on a team of learning automata for continuous optimization," *Genetic programming and evolvable machines,* vol.15, no.2, pp.169–193, 2014.

[3] Z.-b. Zhou, D. D. DONG, J. L. Zhou, "Application of bayesian networks in reliability analysis," *Systems Engineering-Theory & Practice*, vol.6, pp.95–100, 2006.

[4] J. Zhu and A. Deshmukh, "Application of Bayesian decision networks to life cycle engineering in green design and manufacturing," *Engineering Applications of Artificial Intelligence*, vol.16, no.2, pp.91–103, 2003.

[5] B. Malekmohammadi, R. Kerachian, and B. Zahraie, "Developing monthly operating rules for a cascade system of reservoirs: application of bayesian networks," *Environmental Modelling & Software*, vol.24, no.12, pp.1420–1432, 2009.

[6] W. A. Wiest, M. D. Correll, B. G. Marcot, B. J. Olsen, C. S. Elphick, T. P. Hodgman, G. R. Guntenspergen, and W. G. Shriver, "Estimates of tidal-marsh bird densities using bayesian networks," *The Journal of Wildlife Management*, vol.83, no.1, pp.109–120, 2019.

[7] D. Taylor, L. Samie, and C. Champod, "Using bayesian networks to track dna movement through complex transfer scenarios," *Forensic Science International: Genetics*, vol. 42, pp. 69-80, 2019.

[8] D. Dinis, A. Barbosa-Póvoa, and Â. P. Teixeira, "Valuing data in aircraft maintenance through big data analytics: A probabilistic approach for capacity planning using bayesian networks," *Computers & Industrial Engineering*, vol.128, pp.920–936, 2019.

[9] D. M. Chickering, "Learning Bayesian networks is NP-complete," *in Learning from Data*, Springer, pp.121–130, 1996.

[10] R. W. Robinson, "Counting unlabeled acyclic digraphs," *in Combinatorial mathematics* V, pp.28–43, Springer, 1977.

[11] N. Friedman, "Learning belief networks in the presence of missing values and hidden variables," *in ICML*, vol.97, pp.125–133, 1997.

[12] N. Friedman, "The Bayesian structural em algorithm," *in Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pp.129–138, 1998.

[13] P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, prediction, and search*, vol.81. MIT press, 2000.

[14] D. Margaritis, "Learning bayesian network model structure from data," *tech. rep.*, Carnegie-Mellon University Pittsburgh Pa School of Computer Science, 2003.

[15] S. Yaramakala and D. Margaritis, "Speculative markov blanket discovery for optimal feature selection," *in Fifth IEEE International Conference on Data Mining(ICDM'05)*, pp.4 pp, IEEE, 2005.

[16] X. Qi, X. Fan, Y. Gao, and Y. Liu, "Learning Bayesian network structures using weakest mutual-information first strategy," *International Journal of Approximate Reasoning*, vol.114, pp.84–98, 2019.

[17] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data," *Machine learning*, vol.20, no.3, pp.197–243, 1995.

[18] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine learning*, vol.9, no.4, pp.309–347, 1992.

[19] C. P. De Campos, Z. Zeng, and Q. Ji, "Structure learning of Bayesian networks using constraints," *in Proceedings of the 26th Annual International Conference on Machine Learning*, pp.113–120, 2009.

[20] C. P. d. Campos and Q. Ji, "Efficient structure learning of bayesian networks using constraints," *Journal of Machine Learning Research*, vol.12, pp.663–689, 2011.

[21] J. Cussens, "Bayesian network learning with cutting planes," *arXiv preprint* arXiv:1202.3713, 2012.

[22] T. Jaakkola, D. Sontag, A. Globerson, and M. Meila, "Learning bayesian network structure using lp relaxations," *in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp.358–365, 2010.

[23] M. Teyssier and D. Koller, "Ordering-based search: A simple and effective algorithm for learning Bayesian networks," arXiv preprint arXiv:1207.1429, 2012.

[24] [24] Y. Tang and Z. Xu, "A score based approach towards improving bayesian network structure learning," *in Second International Conference on Advanced Cloud and Big Data*, pp.39–44, IEEE, 2014.

[25] [25] S. Behjati and H. Beigy, "An order-based algorithm for learning structure of bayesian networks," *in International Conference on Probabilistic Graphical Models*, pp.25–36, 2018.

[26] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing bayesian network structure learning algorithm," *Machine learning*, vol. 65, no. 1, pp.31–78, 2006.

[27] M. Scutari, P. Howell, D. J. Balding, and I. Mackay, "Multiple quantitative trait analysis using Bayesian networks," *Genetics*, vol.198, no.1, pp.129–137, 2014.

[28] M. Gasse, A. Aussem, and H. Elghazel, "A hybrid algorithm for bayesian network structure learning with application to multi-label learning," *Expert Systems with Applications*, vol.41, no.15, pp.6755–6772, 2014.

[29] J. Cussens, M. Järvisalo, J. H. Korhonen, and M. Bartlett, "Bayesian network structure learning with integer programming: Polytopes, facets and complexity," *Journal of Artificial Intelligence Research*, vol.58, pp.185–229, 2017.

[30] M. Bartlett and J. Cussens, "Advances in bayesian network learning using integer programming," *in proceeding of the 29th conference on Uncertainty in artificial intelligence (UAI2013)*, p.182–191, UAUI Press, 2013.

[31] M. Bartlett and J. Cussens, "Integer linear programming for the bayesian network structure learning problem," *Artificial Intelligence*, vol.244, pp.258–271, 2017.

[32] J. Cussens, "Maximum likelihood pedigree reconstruction using integer programming.," *in WCB@ ICLP*, pp.8–19, 2010.

[33] R. Peharz and F. Pernkopf, "Exact maximum margin structure learning of bayesian networks," *arXiv preprint* arXiv:1206.6431, 2012.

[34] H. S. Farahani and J. Lagergren, "Learning oncogenetic networks by reducing to mixed integer linear programming," *PloS one*, vol.8, no.6, p.e65773, 2013.

[35] J. A. Gámez, J. L. Mateo, and J. M. Puerta, "Learning bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood," *Data Mining and Knowledge Discovery*, vol.22, no.1, pp.106–148, 2011.

شهاب بهجتی مدارک کارشناسی، کارشناسی ارشد و دکتری خود را در رشته علوم کامپیوتر به ترتیب از دانشگاه صنعتی امیرکبیر (پلی‌تکنیک تهران)، دانشگاه صنعتی شریف و پژوهشگاه دانش‌های بنیادی (آی پی ام) اخذ نموده است. زمینه‌های تحقیقاتی مورد علاقه وی شامل: یادگیری ماشین، داده‌کاوی، علم داده و مدل‌های گرافی احتمالاتی، الگوریتم‌های تقریبی و الگوریتم‌های تصادفی است.

آدرس پست الکترونیکی ایشان عبارت است:

shbehjati@ipm.ir

حمید بیگی مدارک کارشناسی و کارشناسی ارشد خود در رشته مهندسی کامپیوتر از دانشگاه شیراز و مدرک دکتری خود را در رشته مهندسی کامپیوتر از دانشگاه صنعتی امیرکبیر (پلی‌تکنیک تهران) اخذ نموده و هم اکنون دانشیار دانشکده مهندسی کامپیوتر دانشگاه صنعتی شریف هستند. زمینه‌های تحقیقاتی مورد علاقه وی شامل: یادگیری ماشین، داده‌کاوی، داده‌های حجیم و مدل‌های گرافی احتمالاتی است.

آدرس پست الکترونیکی ایشان عبارت است:

beigy@sharif.edu

[1] Markov random fields
[2] Bayesian networks
[3] Belief networks
[4] Causal networks
[5] clique potential
[6] Gibbs distribution
[7] Directed Acyclic Graph
[8] non-descendant
[9] local Markov condition
[10] structural EM
[11] Expectation Maximization
[12] Constraint-based methods
[13] Score-based methods
[14] Hybrid methods
[15] Peter Spirtes and Clark Glymour
[16] weakestmutual-information-first
[17] hypothesis space
[18] scoring function
[19] combinatorial space
[20] heuristic search
[21] bayesian score
[22] Minimum Description Length
[23] Bayesian Information Criterion
[24] Score-based Bayesian Network Refinement
[25] Max-Min Hill Climbing
[26] Max-Min Parents and Children
[27] cluster
[28] acyclicity
[29] transitive
[30] Integer Linear Programming
[31] polytope
[32] deterministic rounding
[33] randomized rounding
[34] over fitting
[35] https://www.cs.york.ac.uk/aig/sw/gobnilp
[36] cutting planes
[37] vertex cover

The CSI Journal on Computing Science and Information Technology

The CSI Journal on
Computing Science and
Information Technology
Vol. 17, No.2, 2020

Learning the Structure of Bayesian Networks Using the Deterministic Rounding Method

Shahab Behjati[1], **Hamid Beigy**[2]

[1] School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
[2] Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

Abstract

Bayesian networks are one of the most widely used probabilistic graphic models used to represent a probability distribution and have a wide variety of applications in artificial intelligence, data mining, and machine learning. One of the most important issues in these networks is the problem of learning structure from data. In general, structured learning methods are divided into three categories: constraint-based, score-based, and hybrid. In this paper, we present a new score-based method for learning a Bayesian network from data, which is based on the deterministic rounding method in linear programming. The proposed method first models the structure learning problem as an integer linear program and then relax it to a linear programming problem. Then, by solving the relaxed linear program, it converts the obtained fractional answers into integer answers using the introduced deterministic rounding method.

Keywords: linear programming, Bayesian networks, deterministic rounding, learning structure in Bayesian networks.