یادگیری پارامترهای شبکه بیزی از داده حاوی مقادیر گم‌شده

کبری اطمینانی، محمود نقیب‌زاده، مهدی عمادی و امیررضا رضوی

چکیده: یادگیری ساختار شبکه بیزی از داده، در سال‌های اخیر توجه بسیاری از محققین را به خود جلب نموده است. از طرفی، یافتن شبکه بهینه از داده کامل، خود یک مسأله غیر چندجمله‌ای سخت می‌باشد و پیچیدگی مسأله، زمانی که داده ناقص است، بیشتر می‌شود. به طور کلی دو حالت یادگیری شبکه بیزی از داده ناقص وجود دارد: زمانی که ساختار مشخص است و زمانی که ساختار نیز نامشخص است. در این مقاله سعی بر آن است تا پارامترهای بهینه را برای یک شبکه بیزی با ساختار مشخص از داده حاوی مقادیر گم‌شده بیابیم. برای این منظور مفهوم "پارامتر مؤثر" را معرفی نمودیم، به طوری که درست‌نمایی ساختار شبکه به شرط داده کامل‌شده، بیشینه گردد. این روش می‌تواند به هر الگوریتمی همچون بیشینه‌سازی امید ساختاری که به پارامترهای بهینه برای یافتن ساختار شبکه بیزی نیاز دارند، متصل شود.

در این مقاله ثابت کردیم که روش پیشنهادی از دیدگاه تابع درست‌نمایی به پارامترهای بهینه شبکه دست می‌یابد. نتایج اعمال روش پیشنهادی به چندین شبکه بیزی استاندارد، نشان‌دهنده سرعت روش در مقایسه با روش‌های شناخته‌شده قبلی است و نیز این که به پارامترهای بهتری نسبت به آنها دست می‌یابد.

کلید واژه: **پارامترهای شبکه بیزی، راست‌نمایی، شبکه بیزی، مقدار گم‌شده.**

۱- مقدمه

شبکه بیزی[1] یک روش نمایش دانش است که اجازه استنباط در موارد بروز عدم قطعیت را می‌دهد (که یکی از ویژگی‌های مثبت این مدل در مقایسه با دیگر مدل‌ها است). شبکه بیزی در واقع یک مدل گرافیکی احتمالی است که استقلال‌های شرطی بین یک مجموعه متغیر تصادفی را نشان می‌دهد. در یک تعریف رسمی می‌توان شبکه بیزی را یک گراف بدون دور جهت‌دار[2] (DAG) دانست که گره‌های آن متغیرهای تصادفی و ساختار یال‌ها وابستگی احتمالی را نشان می‌دهند. از کاربردهای موفق شبکه‌های بیزی می‌توان به بیولوژی محاسباتی و بیوانفورماتیک [۱]، پزشکی [۲]، بازیابی اطلاعات [۳]، سیستم‌های تصمیم‌یار [۴] و ... اشاره کرد.

فرایند یادگیری شبکه بیزی از دانش افراد خبره، اغلب کار طاقت‌فرسا و پرهزینه‌ای است. لذا یادگیری ساختار شبکه بیزی از داده توجه بسیاری از محققین را در سال‌های اخیر به خود جلب نموده است. چیکرینگ و همکارانش در [۵] و [۶] نشان دادند که یافتن ساختار بهینه از داده یک مسأله غیر چندجمله‌ای سخت[3] است. تعداد کل ساختارهای ممکن از مرتبه $O(n!2^{\binom{n}{2}})$ است که n نشان‌دهنده تعداد متغیرها یا همان گره‌های شبکه است.

به طور کلی روش‌های یادگیری می‌توانند به دو دسته کلی تقسیم شوند. یک دسته با نام روش‌های مبتنی بر امتیاز است که ساختار شبکه را با توجه به داده، به کمک توابع امتیازدهی امتیاز می‌دهند. دسته دوم با نام روش‌های مبتنی بر قید است که بر اساس شباهت‌های آماری بین متغیرها (استقلال و یا وابستگی بین متغیرها) ساختار بهینه را می‌یابد. علاوه بر این، برخی روش‌هایی همچون [۷] و [۸] هستند که از هر دو دسته به صورت ترکیبی استفاده می‌کنند. در این مقاله بر روی دسته اول متمرکز می‌شویم.

تحقیقات در این حوزه اغلب بر روی یادگیری ساختار از داده کامل متمرکز شده است. از طرفی، داده ناقص به کرات در مسایل یادگیری دنیای واقعی رخ می‌دهد. همان گونه که قبلاً نیز ذکر شد، مسأله یادگیری ساختار از داده کامل، یک فرایند رام‌نشدنی است. اکنون داده گم‌شده نیز به آن اضافه شده است که منجر به ایجاد یک مسأله رام‌نشدنی با پیچیدگی زمانی و مکانی بالا خواهد شد.

روبین در [۹] مکانیسم‌های گم‌شدن داده را به سه دسته زیر تقسیم نموده است:

۱) گم‌شدن کاملاً تصادفی[4] (MCAR): احتمال هر مقدار گم‌شده مستقل از مقادیر مشاهده‌شده و نشده در مجموعه داده است.

۲) گم‌شدن تصادفی[5] (MAR): احتمال هر مقدار گم‌شده تابعی از مقادیر مشاهده‌شده در مجموعه داده است.

۳) گم‌شدن غیر تصادفی[6] (NMAR): احتمال هر مقدار گم‌شده تابعی از مقادیر مشاهده‌شده و نشده در مجموعه داده است.

لیتل و روبین در [۱۰] روش‌های برخورد با مقادیر گم‌شده را به سه دسته کلی تقسیم نموده‌اند: روش‌های تحلیل حالت کامل[7]، روش‌های مبتنی بر مقداردهی[8] و روش‌های درست‌نمایی مبتنی بر مدل[9]. دسته اول تنها رکوردهایی را بررسی می‌کند که تمامی متغیرها در آنها مشاهده شده‌اند، یعنی رکوردهای حاوی مقادیر گم‌شده را نادیده می‌گیرد. این روش تنها زمانی قابل اعمال است که میزان مقادیر گم‌شده ناچیز باشد. دسته دوم سعی می‌کند مقادیر گم‌شده را با مقادیر جدیدی تخمین زده و جایگزین کند و سپس شبکه بیزی را از روی داده کامل‌شده بسازد.

این مقاله در تاریخ ۳۰ فروردین ماه ۱۳۹۱ دریافت و در تاریخ ۲۴ بهمن ماه ۱۳۹۱ بازنگری شد. این تحقیق توسط مرکز تحقیقات مخابرات ایران (ITRC) با قرارداد شماره ۵۰۰/۸۷۴۰ پشتیبانی شده است.

کبری اطمینانی، گروه کامپیوتر، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: etminani@wali.um.ac.ir).

محمود نقیب‌زاده، گروه کامپیوتر، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: naghibzadeh@um.ac.ir).

مهدی عمادی، گروه آمار، دانشکده ریاضی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: emadi@um.ac.ir).

امیررضا رضوی، گروه انفورماتیک پزشکی، دانشگاه علوم پزشکی مشهد، مشهد، ایران، (email: razaviar@mums.ac.ir).

1. Bayesian Network
2. Directed Acyclic Graph
3. NP-Hard
4. Missing Completely at Random
5. Missing at Random
6. Not Missing at Random
7. Complete Case Analysis
8. Imputation-Based
9. Model-Based Likelihood

Archive of SID

www.SID.ir

نهایتاً دسته سوم سعی می‌کند مکانیسم گم‌شدن را مدل کند. در تمامی این سه دسته فرض می‌شود که مکانیسم گم‌شدن، از نوع گم‌شدن غیر تصادفی نیست.

تعداد روش‌های یادگیری شبکه بیزی از داده ناقص نسبت به داده کامل کمتر است. منظور از داده ناقص، داده حاوی مقادیر گم‌شده و/یا متغیرهای مخفی است. به طور کلی دو حالت در یادگیری شبکه بیزی از داده ناقص رخ می‌دهد: ساختار شبکه مشخص و ساختار نامشخص.

در این مقاله ابتدا مسأله یادگیری پارامترهای شبکه بیزی از داده حاوی مقادیر گم‌شده، با فرض ساختار شبکه بیزی مشخص به صورت یک مسأله بهینه‌سازی فرمول‌بندی شده است. سپس یک روش مستقیم برای یافتن جواب بهینه این مسأله با عنوان "پارامتر مؤثر" ارائه شده است. در واقع به ازای تمامی مقادیر ممکن قابل انتساب به هر مقدار گم‌شده، یک پارامتر مؤثر محاسبه می‌شود و سپس از بین این مقادیر، مقداری منتسب می‌شود که دارای بزرگ‌ترین پارامتر مؤثر است. در اینجا فرض می‌شود که داده ناقص، از نوع گم‌شدن غیر تصادفی نیست. روش پیشنهادی می‌تواند در هر الگوریتم یادگیری شبکه بیزی با ساختار نامشخص همچون بیشینه‌سازی امید ساختاری[1] (SEM) اعمال شود.

روش پیشنهادی و دو روش شناخته‌شده قبلی برای یادگیری پارامترهای شبکه بیزی، بیشینه‌سازی امید[2] (EM) و نمونه‌گیری گیبس[3] (GS)، بر روی دو شبکه بیزی استاندارد، ASIA و Cancer - Neapolitan، با درصدهای متفاوتی از مقادیر گم‌شده اعمال شده است. نتایج نشان‌دهنده سرعت روش پیشنهادی است.

سازمان‌بندی ادامه مقاله به صورت زیر است. در بخش دوم به بررسی تحقیقات مشابه می‌پردازیم. مسأله یادگیری شبکه بیزی از داده در بخش سوم بیان می‌شود. روش پیشنهادی، به عبارتی مسأله بهینه‌سازی و پارامتر مؤثر برای حل آن، لم‌ها و قضایای اثبات صحت و بهینگی آن و پیچیدگی زمانی در بخش چهارم مورد بحث قرار می‌گیرد. نتایج تجربی در بخش پنجم و در نهایت، نتیجه‌گیری در بخش ششم ارائه می‌شوند.

۲- تحقیقات مشابه

پروژه‌های تحقیقاتی متعددی در زمینه یادگیری شبکه بیزی از داده کامل تاکنون ارائه شده است. برخی از این موارد، رسیدن به جواب بهینه را تضمین نمی‌کنند که از این جمله می‌توان به [۱۱] و [۱۲] اشاره نمود. از طرفی روش‌هایی نیز بر پایه تکنیک‌های برنامه‌سازی پویا همچون [۱۳] تا [۱۵] و انشعاب و تحدید همچون [۱۶] تا [۱۸] ارائه شده‌اند که رسیدن به ساختار بهینه را تضمین می‌کنند.

در زمینه یادگیری شبکه بیزی از داده ناقص، چیکرینگ و هکرمن [۱۹] جزء اولین محققینی بودند که روش نمونه‌گیری گیبس را برای حل این مسأله ارائه دادند. نمونه‌گیری گیبس، یکی از روش‌های زنجیره مارکفی مونت کارلو[4] (MCMC) به حساب می‌آید. در این روش به صورت تکراری مقادیر گم‌شده را با توجه به برخی توابع توزیع شرطی، جانهی کرده و تخمین تصادفی از احتمال پسین ارائه می‌کند [۲۰]. زنجیر در این روش دیر به همگرایی می‌رسد و محاسبات زیادی نیاز دارد و تشخیص همگراشدن زنجیر خود یک موضوع مورد تحقیق است.

یکی دیگر از روش‌های مهم برخورد با داده ناقص در یادگیری شبکه بیزی، روش بیشینه‌سازی امید ساختاری است که توسط فریدمن در [۲۱] و [۲۲] ارائه شد. بیشینه‌سازی امید ساختاری، مشابه نمونه‌گیری گیبس، ابتدا سعی در جانهی مقادیر گم‌شده می‌کند. برای این منظور از الگوریتم بیشینه‌سازی امید [۲۳] استفاده می‌کند و بدین ترتیب، مسأله اولیه به حل مسأله یادگیری شبکه بیزی از داده کامل، تبدیل می‌گردد. بیشینه‌سازی امید ساختاری، کار را از یک شبکه بیزی اولیه با ساختار و پارامترهای تصادفی شروع می‌کند. سپس در هر تکرار، بهترین ساختار را از مجموعه داده کامل‌شده پیدا کرده و بهترین پارامترها را برای این شبکه محاسبه می‌کند و تا زمانی این عمل را تکرار می‌کند که دیگر تغییری در ساختار شبکه مشاهده نشود. فریدمن برای پیداکردن ساختار بهینه در هر مرحله از روش تپه‌نوردی بهره برد اما دیگر روش‌های یافتن ساختار بهینه از داده کامل نیز قابل اعمال هستند. پس از یادگیری ساختار، می‌بایست پارامترهای بهنیه شبکه را محاسبه کرد. برای این منظور در هر مرحله، مقادیر گم‌شده را طوری جانهی می‌کند که تابع امتیازدهی بیشینه گردد. این روش نیز زمان‌بر بوده و در هر مرحله نیاز به محاسبات زیادی دارد. برخی تحقیقات بعدی همچون [۲۴] سعی در بهبود این روش داشته‌اند.

الگوریتم تحدید و افتادن[5] (BC) که توسط رامونی و سباستیانی در [۲۰] ارائه شده بر تخمین پارامترهای شبکه بیزی از داده ناقص تمرکز کرده است. این الگوریتم دارای دو مرحله تحدید و افتادن است. در مرحله تحدید، حالات ممکن کامل‌شدن داده ناقص بررسی می‌شود. همچنین یک بازه (کران بالا و پایین) که تخمین پارامترهای واقعی در آن قرار می‌گیرد، محاسبه می‌شود. سپس در مرحله افتادن، یک ترکیب محدب از پارامترهایی که در مرحله قبل به دست آمدند، محاسبه می‌شود. وزن‌های ترکیب محدب از داده مشاهده‌شده محاسبه می‌شود. الگوریتم تحدید و افتادن، هزینه تخمین توزیع احتمال هر متغیر را به هزینه محاسبه بازه و ترکیب محدب به ازای هر حالت از آن متغیر کاهش می‌دهد.

در این مقاله یادگیری پارامترهای شبکه بیزی از داده حاوی مقادیر گم‌شده، به صورت یک مسأله بهینه‌سازی عدد صحیح غیر خطی، فرمول‌بندی شده است. سپس "پارامتر مؤثر" برای حل این مسأله بر مبنای بیشینه‌کردن درست‌نمایی[6] ارائه شده است، بدین صورت که ابتدا برای تمامی مقادیر قابل انتساب به هر مقدار گم‌شده، مقدار پارامتر مؤثر با شمارش رکوردهای خاصی محاسبه می‌شود. سپس به مقادیر گم‌شده مقادیری منتسب می‌شوند که دارای بیشترین مقدار پارامتر مؤثر هستند. انتساب مقادیر جدید به مقادیر گم‌شده بر اساس پارامتر مؤثر، فرایند یادگیری پارامترهای شبکه بیزی را تسریع می‌کند. برای توضیح بیشتر روش، ابتدا مسأله در بخش بعدی تشریح می‌گردد.

۳- شرح مسأله

۳-۱ شبکه بیزی

یک شبکه بیزی یک گراف بدون دور جهت‌دار است که از n متغیر تصادفی تشکیل شده و چگالی احتمالات توأم بین این متغیرها را نشان می‌دهد. یک شبکه بیزی را به صورت سه‌تایی (G,V,P) نشان می‌دهند که G نشان‌دهنده گراف بدون دور جهت‌دار است، V مجموعه متغیرهای تصادفی که گره‌های شبکه G هستند و P مجموعه چگالی‌های احتمالی

1. Structural Expectation Maximization
2. Expectation Maximization
3. Gibbs Sampling
4. Mont Carlo Markov Chain
5. Bound and Collapse
6. Maximum Likelihood

Archive of SID

www.SID.ir

جدول ۱: یک مجموعه داده نمونه حاوی مقادیر گم‌شده.

V_1	V_2
۰	?
۱	۰
۱	۱
?	۱
۰	۰
۰	۱
۱	۰
۰	۱

شرطی $p(V_i|\pi_i)$ یا همان پارامترهای شبکه بیزی می‌باشد که π_i مجموعه والدهای متغیر V_i را نشان می‌دهد. مهم‌ترین ویژگی شبکه بیزی، ویژگی مارکفی آن است. این ویژگی می‌گوید هر متغیر به شرط والدهایش مستقل شرطی از تمامی غیر نوادگانش است. لذا قانون زنجیر به صورت زیر کاهش می‌یابد

$$p(V_1,\ldots,V_n)=\prod_{i=1}^{n}p(V_i|V_{i-1},\ldots,V_1)=\prod_{i=1}^{n}p(V_i|\pi_i) \qquad (۱)$$

هر متغیر در شبکه بیزی می‌تواند گسسته، پیوسته یا ترکیبی از این دو باشد. در این مقاله متغیرهای گسسته در نظر گرفته شده‌اند. لذا اگر متغیر پیوسته‌ای وجود داشته باشد گسسته‌سازی می‌شود [۲۵]. هر متغیر گسسته دارای r_i مقدار ممکن است و لذا تعداد کل پیکربندی‌های والدانش از $q_i=\prod_{V_j\in\pi_i}r_j$ به دست می‌آید.

۳-۲ توابع امتیازدهی

یافتن ساختار بهینه شبکه بیزی، هدف یادگیری ساختار آن از داده است. توابع امتیازدهی متعددی ارائه شده‌اند که برای داده کامل قابل اعمال هستند که این توابع به ساختار بهینه، بیشترین امتیاز را اختصاص می‌دهند. از جمله این توابع می‌توان به BD[1] [۲۶]، BDe [۲۷]، BDeu [۱۲]، BIC[2] [۲۸]، MDL[3] [۲۹] و AIC[4] [۳۰] اشاره نمود. با داشتن مجموعه داده کامل D با N نمونه، یافتن ساختار شبکه بیزی بهینه به صورت مسأله بیشینه‌سازی تابع امتیازدهی بیان می‌شود

$$\max_G score(D,G) \qquad (۲)$$

توابع امتیازدهی به داده ناقص قابل اعمال نیستند. لذا بر آن شدیم تا ابتدا مقادیر گم‌شده را مورد بررسی قرار دهیم و سپس از توابع امتیازدهی استفاده کنیم.

اغلب توابع امتیازدهی مذکور بر پایه درست‌نمایی بنا نهاده شده‌اند

$$score(D,G)=ll(D:G,\theta)-w.Dim(G) \qquad (۳)$$

که θ نشان‌دهنده پارامترهای مدل است و به G بستگی دارد و w در توابع امتیازدهی مختلف مقادیر متفاوتی می‌گیرد (به طور مثال در BIC، $w=\log N/2$ و در AIC، $w=1$ است). $Dim(G)$ ابعاد مدل است و از رابطه زیر قابل محاسبه است

$$Dim(G)=\sum_{i=1}^{n}q_i(r_i-1) \qquad (۴)$$

و ll لگاریتم درست‌نمایی است که می‌تواند برای شبکه بیزی گسسته به صورت زیر محاسبه شود

$$ll(D:G,\theta)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i}N_{ijk}\log\theta_{ijk} \qquad (۵)$$

N_{ijk} تعداد رکوردهای مجموعه داده است که متغیر V_i مقدار kام خود و والدهایش در پیکربندی jام قرار دارند. پارامترهای بیشینه $\hat{\theta}_{ijk}$، فراوانی نسبی در داده هستند و لذا $\hat{\theta}_{ijk}=N_{ijk}/N_{ij}$ که $N_{ij}=\sum_{i=1}^{r_i}N_{ijk}$.

۴- روش پیشنهادی

در این بخش، ابتدا فرمول‌بندی مسأله یادگیری پارامترهای شبکه بیزی از داده حاوی مقادیر گم‌شده به صورت یک مسأله بهینه‌سازی ارائه می‌شود و سپس پارامتر مؤثر برای حل این مسأله تعریف می‌گردد.

۴-۱ مسأله بهینه‌سازی

سعی کردیم مسأله یادگیری پارامترهای شبکه بیزی از داده حاوی مقادیر گم‌شده را به صورت یک مسأله بهینه‌سازی بنویسیم. برای این منظور ابتدا مقادیر جدیدی را برای مقادیر گم‌شده طوری تخمین می‌زنیم که تابع امتیازدهی شبکه بیزی بیشینه گردد. فرض اولیه این بود که ساختار شبکه بیزی مشخص است، لذا در (۳) مقدار $w.Dim(G)$ برای تمامی حالات ممکن پارامترها یکسان است و می‌تواند حذف گردد. بنابراین تنها قسمت لگاریتم درست‌نمایی باقی می‌ماند. محاسبه درست‌نمایی برای داده حاوی مقادیر گم‌شده کار سختی به شمار می‌رود، لذا سعی کردیم تا فرمول‌بندی مناسبی ارائه دهیم تا بر این مشکل فائق آید.

برای نوشتن تابع درست‌نمایی ابتدا مقادیر گم‌شده را باید مورد بررسی قرار دهیم و برای این منظور فرمول‌بندی زیر ارائه شد. هر رکورد حاوی مقادیر گم‌شده در مجموعه داده با متامتغیر H نشان داده می‌شود. این متامتغیر با توجه به مقادیر گم‌شده و تعداد حالت‌های ممکن برای جانهی آنها به زیرمتغیرهای η تبدیل می‌شود. برای روشن‌تر شدن بحث با یک مثال شروع می‌کنیم.

فرض کنید مجموعه داده نشان داده شده در جدول ۱ که حاوی مقادیر گم‌شده است را در اختیار داریم. همان طور که گفته شد هر رکورد حاوی مقادیر گم‌شده با یک متامتغیر نمایش داده می‌شود. لذا در این مثال دو متامتغیر $H_{.x}$ و H_{x1} برای دو رکورد حاوی مقادیر گم‌شده در نظر گرفته می‌شوند. برای توضیح اندیس‌ها کافی است بدانیم که اندیس $.x$ در $H_{.x}$ نشان می‌دهد که در رکورد مورد نظر V_1 مقدار ۰ و V_2 مقدارش گم‌شده است (x نشان‌دهنده گم‌شدن است). فرض کنید در این مثال متغیرها دودویی هستند، پس $r_1=r_2=2$. در این صورت هر یک از متامتغیرها به دو زیرمتغیر تبدیل می‌شوند: $H_{.x}$ به $\eta_{.x}^{0}$ و $\eta_{.x}^{1}$ و H_{x1} به η_{x1}^{0} و η_{x1}^{1}. یعنی x دو مقدار ممکن برای دو متغیر (در اینجا ۰ و ۱) می‌تواند به خود اختصاص دهد.

با این توضیحات مشخص می‌شود که متغیرهای شبکه بیزی باید گسسته باشند. اگر متغیر پیوسته‌ای وجود داشت، ابتدا باید گسسته‌سازی شود ولی در تعداد مقادیر ممکن هر متغیر، محدودیتی وجود ندارد. گرچه برای سادگی در این مثال، متغیرها دودویی در نظر گرفته شده‌اند، ولی r_iها می‌توانند هر مقدار صحیحی داشته باشند. آنچه همواره دودویی در نظر گرفته می‌شود، زیرمتغیرها هستند و یکی از مقادیر ۰ یا ۱ را می‌گیرند. زیرمتغیرها نشان می‌دهند که کدام یک از مقادیر ممکن به مقدار گم‌شده نسبت داده شده است و هر زیرمتغیر مربوط به یکی از مقادیر ممکن برای

1. Bayesian Dirichlet
2. Bayesian Information Criterion
3. Minimum Description Length
4. Akaike Information Criterion

Archive of SID

www.SID.ir

شکل ۱: ساختار مشخص برای مجموعه داده نشان داده شده در جدول ۱.

مقدار گم‌شده است. در واقع اگر مقدار زیرمتغیری ۱ باشد، بدین معنا است که مقدار مربوط به آن زیرمتغیر در نهایت به مقدار گم‌شده نسبت داده شده است. به علاوه از بین زیرمتغیرهای یک متامتغیر تنها یکی مقدار ۱ و بقیه مقدار ۰ می‌گیرند. دلیل این امر نیز روشن است، تنها یکی از مقادیر ممکن به مقدار گم‌شده تخصیص می‌یابد.

در مثال قبل نهایتاً چهار زیرمتغیر ایجاد شد که هر یک نشان‌دهنده یکی از حالات ممکن برای مقداردهی مقادیر گم‌شده است. با این نمادگذاری، تابع درست‌نمایی می‌تواند بازنویسی شود. از (۵) برای حل مثال استفاده می‌کنیم. با فرض این که ساختار مشخص این مثال مطابق شکل ۱ باشد، برای نمونه در V_2 درست‌نمایی به صورت زیر بازنویسی می‌شود

$$\begin{aligned} N_{211} &= 1+\eta^{\cdot\cdot}_{\cdot x} \\ N_{212} &= 2+\eta^{\cdot 1}_{\cdot x}+\eta^{\cdot 1}_{x1} \\ \Rightarrow N_{21} &= 1+\eta^{\cdot\cdot}_{\cdot x}+2+\eta^{\cdot 1}_{\cdot x}+\eta^{\cdot 1}_{x1} = 4+\eta^{\cdot 1}_{x1} \end{aligned} \quad (6)$$

و

$$\begin{aligned} N_{221} &= 2 \\ N_{222} &= 1+\eta^{11}_{x1} \\ \Rightarrow N_{22} &= 3+\eta^{11}_{x1} = 3+\eta^{11}_{x1} \end{aligned} \quad (7)$$

لذا لگاریتم درست‌نمایی در V_2 به صورت زیر نوشته می‌شود

$$\begin{aligned} ll_{V_2}(D:G,\theta) &= \sum_{j=1}^{2}\sum_{k=1}^{2} N_{2jk}\log(\frac{N_{2jk}}{N_{2j}}) \\ &= (1+\eta^{\cdot\cdot}_{\cdot x})\log(\frac{1+\eta^{\cdot\cdot}_{\cdot x}}{4+\eta^{\cdot 1}_{x1}}) \\ &+ (2+\eta^{\cdot 1}_{x1}+\eta^{\cdot 1}_{\cdot x})\log(\frac{2+\eta^{\cdot 1}_{x1}+\eta^{\cdot 1}_{\cdot x}}{4+\eta^{\cdot 1}_{x1}}) \\ &+ 2\log(\frac{2}{3+\eta^{11}_{x1}}) + (1+\eta^{11}_{x1})\log(\frac{1+\eta^{11}_{x1}}{3+\eta^{11}_{x1}}) \end{aligned} \quad (8)$$

مشابه فرمول‌های فوق برای V_1 نیز نوشته می‌شود. بنابراین مسأله یافتن شبکه بیزی با بیشترین امتیاز (رابطه (۲)) برای این مثال می‌تواند به صورت زیر بازنویسی شود

$$\begin{aligned} &\max ll(D:G,\theta) = \\ &\max_{s.t.} \quad [(4+\eta^{\cdot 1}_{x1})\log(\frac{4+\eta^{\cdot 1}_{x1}}{8}) \\ &\sum_{i=0}^{1}\eta^{\cdot i}_{\cdot x}=1 \ , \ \sum_{i=0}^{1}\eta^{i\cdot}_{x1}=1 \ , \ \text{all sub-variables } \eta\in\{0,1\} \\ &+ (3+\eta^{11}_{x1})\log(\frac{3+\eta^{11}_{x1}}{8}) + (1+\eta^{\cdot\cdot}_{\cdot x})\log(\frac{1+\eta^{\cdot\cdot}_{\cdot x}}{4+\eta^{\cdot 1}_{x1}}) \\ &+ (2+\eta^{\cdot 1}_{x1}+\eta^{\cdot 1}_{\cdot x})\log(\frac{2+\eta^{\cdot 1}_{x1}+\eta^{\cdot 1}_{\cdot x}}{4+\eta^{\cdot 1}_{x1}}) \\ &+ 2\log(\frac{2}{3+\eta^{11}_{x1}}) + (1+\eta^{11}_{x1})\log(\frac{1+\eta^{11}_{x1}}{3+\eta^{11}_{x1}})] \end{aligned} \quad (9)$$

دو قید اول تضمین می‌کنند که تنها یکی از زیرمتغیرهای یک متامتغیر مقدار ۱ بگیرد و قید سوم تأکید می‌کند که زیرمتغیرها دودویی هستند. متأسفانه این مسأله بهینه‌سازی یک مسأله بیشینه‌سازی عدد صحیح

Algorithm 1. Finding the optimal parameters of BN

Input. G : BN structure, D : incomplete dataset
Output. P : optimal parameters for BN, D' : completed dataset
begin
(1) $D' = D$
(2) **for** $m = 1$ to M
//each missing value which occurs for variable V_i in D
for $j = 1$ to r_i
(3) /*each possible value that can be assigned to the missing value*/
(4) $\lambda_j = CalcEffeParam(G, m, j)$
(5) **end for**
(6) $\{\eta_j\} = 0$
(7) $\eta_k = 1$ where λ_k has the largest value among $\{\lambda_j\}$
(8) Assign v_k to the missing value in D'
(9) **end for**
(10) $P = CalcParam(G, D')$
end

شکل ۲: الگوریتم روش پیشنهادی برای یافتن پارامترهای بهینه شبکه بیزی از داده حاوی مقادیر گم‌شده.

غیر خطی است که حل آن خود دارای پیچیدگی بالایی است. در بخش بعدی راه حلی برای این مسأله، به عبارتی پارامتر مؤثر ارائه خواهیم داد.

۴-۲ پارامتر مؤثر

با معرفی پارامتر مؤثر به دنبال راه حلی برای مسأله بهینه‌سازی مطرح‌شده در بخش قبلی هستیم. ابتدا تعریف کلی پارامتر مؤثر ارائه می‌شود. سپس قضیه‌ای که پایه روش پیشنهادی است مطرح می‌شود و در نهایت تعریف دقیق و لم‌های مورد نیاز مورد بحث قرار می‌گیرند.

تعریف ۱: برای هر زیرمتغیر η، یک پارامتر مؤثر λ وجود دارد.

تعریف دقیق‌تر پارامتر مؤثر که به لایه مارکفی[1] متغیری که در آن مقادیر گم‌شده رخ‌داده بستگی دارد، بعداً ارائه می‌شود اما نکته‌ای که بین تمام لایه‌های مارکفی صدق می‌کند این است که بین تمامی زیرمتغیرهای یک متامتغیر، زیرمتغیری مقدار ۱ می‌گیرد که دارای بیشترین مقدار پارامتر مؤثر باشد. این حقیقت مبنای کار روش پیشنهادی برای یافتن بهترین مقادیر تخمینی برای مقادیر گم‌شده و در نهایت بهترین پارامترهای شبکه بیزی است. قضیه ۱ گویای همین مطلب است.

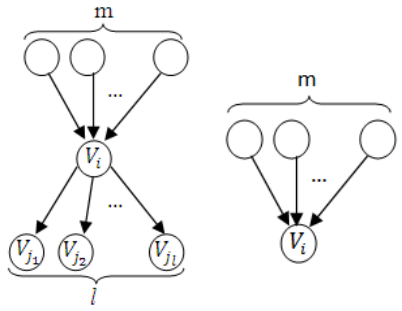
قضیه ۱: اگر η_i و η_j دو زیرمتغیر از متامتغیر H باشند به طوری که $\lambda_i > \lambda_j$، آنگاه $ll'_i > ll'_j$ که ll'_i لگاریتم درست‌نمایی است زمانی که $\eta_j = 0$ و $\eta_i = 1$ (یعنی تنها η_i مقدار ۱ گرفته است).

قبل از توضیح دقیق‌تر پارامتر مؤثر، ایده کلی روش پیشنهادی که در شکل ۲ نشان داده شده، ارائه شده است. ابتدا به ازای هر مقدار گم‌شده که در مجموعه داده D رخ داده یک پارامتر مؤثر λ به کمک تابع CalcEffeParam() برای تمامی مقادیر ممکن آن محاسبه می‌شود. سپس با توجه به قضیه ۱، مقدار گم‌شده، مقدار kام ممکن را می‌گیرد به طوری که λ_k بیشترین مقدار پارامتر مؤثر را در بین پارامتر مؤثر دیگر مقادیر ممکن، دارا بوده است. یعنی تنها η_k در بین کلیه زیرمتغیرهای H مقدار ۱ گرفته است. پس از تکمیل مجموعه داده، مقدار پارامتر بهینه به کمک تابع CalcParam() محاسبه می‌شود.

1. Markov Blanket

Archive of SID

www.SID.ir



شکل ۳: (الف) V_i دارای حداکثر m والد و بدون فرزند و (ب) V_i دارای حداکثر m والد و l فرزند.

همان طور که قبلاً گفته شد، تعریف پارامتر مؤثر به لایه مارکفی متغیری که مقدار گم‌شده در آن رخ داده، بستگی دارد. تعریف پارامتر مؤثر را برای دو حالت کلی زیر در نظر گرفته‌ایم. لازم به ذکر است که V_i متغیری است که در رکورد مزبور، مقدارش گم شده است و λ_p پارامتر مؤثر زیرمتغیر η_p است. η_p زمانی ۱ می‌شود که در آن رکورد V_i مقدار p ام ممکن خود را بگیرد.

حالت اول: V_i بدون فرزند و دارای حداکثر m والد است که $m \geq ۰$ (شکل ۳- الف).

حالت دوم: V_i دارای حداکثر m والد و l فرزند است که $m,l \geq ۱$ (شکل ۳- ب).

ابتدا لم ۱ ارائه می‌شود که در اثبات دیگر لم‌ها به کار می‌آید. سپس در زیربخش‌های بعدی دو مورد فوق مورد بحث و بررسی قرار می‌گیرند.

لم ۱: اگر $a > b \geq ۰$ و $a,b \in N$، آنگاه

$$(a+۱)\log(a+۱)+b\log b > a\log a+(b+۱)\log(b+۱) \quad (۱۰)$$

اثبات: رابطه (۱۰) را می‌توان به صورت زیر نوشت

$$(a+۱)\log(a+۱)-a\log a > (b+۱)\log(b+۱)-b\log b \quad (۱۱)$$

با تعریف $f(x)=(x+۱)\log(x+۱)-x\log(x)$، (۱۱) به شکل زیر تبدیل خواهد شد

$$a > b \Rightarrow f(a) > f(b) \quad (۱۲)$$

اگر $f(x)$ یک تابع اکیداً صعودی باشد، نامساوی فوق صحیح خواهد بود. لذا باید ثابت کنیم $f(x)$ اکیداً صعودی است

$$f'(x)=\log(x+۱)-\log(x) \quad (۱۳)$$

از آنجا که $a > b \geq ۰$ آنگاه $f'(x) > ۰$. بنابراین $f(x)$ یک تابع اکیداً صعودی است و حکم برقرار است.

٤-۲-۱ حالت اول: V_i دارای m والد و بدون فرزند

تعریف ۲: اگر V_i دارای m $(m \geq ۰)$ والد و بدون فرزند باشد (شکل ۳- الف)، آنگاه پارامتر مؤثر λ_p برای زیرمتغیر η_p به صورت زیر تعریف می‌شود

$$\lambda_p = N_{icp}/N_{ic} \quad (۱۴)$$

که N_{icp} تعداد رکوردهایی است که V_i مقدار p ام و والدهایش پیکربندی c ام را دارند و N_{ic} تعداد رکوردهایی است که والدهایش پیکربندی c ام را دارند. اگر $m=۰$ باشد، آنگاه $N_{ic}=N$ که N تعداد کل رکوردها است.

اکنون قضیه ۱ برای این حالت اثبات می‌گردد که در قضیه زیر بازنویسی شده است.

قضیه ۲: فرض کنید V_i دارای m $(m \geq ۰)$ والد و بدون فرزند باشد. اگر η_p و η_q دو زیرمتغیر از متامتغیر H باشند به طوری که $\lambda_p > \lambda_q$، آنگاه $ll'_p > ll'_q$.

اثبات: به جای نوشتن کل لگاریتم درست‌نمایی، تنها بخشی از آن که مقدار گم‌شده بر آن مؤثر است را می‌نویسیم، به عبارتی بخش‌هایی که حاوی زیرمتغیرها است. چرا که بقیه قسمت‌ها برای ll'_p و ll'_q یکسان بوده و حذف می‌گردند. با فرض این که مقدار گم‌شده برای V_i زمانی رخ داده که والدهایش در پیکربندی c ام بوده‌اند، آن بخش از لگاریتم درست‌نمایی که مد نظر است به صورت زیر خواهد بود

$$ll' = \sum_{j=۱,j\neq c}^{q_i}\sum_{d=۱}^{r_i}(N_{ijd}\log\frac{N_{ijd}}{N_{ij}}) + \sum_{d=۱}^{r_i}((N_{icd}+\eta_d)\log\frac{N_{icd}+\eta_d}{N_{ic}}) \quad (۱۵)$$

با توجه به تعریف ۲ داریم $\lambda_p = N_{icp}/N_{ic}$ و $\lambda_q = N_{icq}/N_{ic}$، بنابراین

$$\lambda_p > \lambda_p \Rightarrow \frac{N_{icp}}{N_{ic}} > \frac{N_{icq}}{N_{ic}} \Rightarrow N_{icp} > N_{icq} \quad (۱۶)$$

از طرف دیگر ll'_p به صورت زیر نوشته می‌شود، یعنی ll' که $\eta_p = ۱$

$$ll'_p = \sum_{j=۱}^{q_i}\sum_{d=۱,d\neq p}^{r_i}(N_{ijd}\log\frac{N_{ijd}}{N_{ij}})+(N_{icp}+۱)\log\frac{N_{icp}+۱}{N_{ic}} \quad (۱۷)$$

و به صورت مشابه برای ll'_q

$$ll'_q = \sum_{j=۱}^{q_i}\sum_{d=۱,d\neq q}^{r_i}(N_{ijd}\log\frac{N_{ijd}}{N_{ij}})+(N_{icq}+۱)\log\frac{N_{icq}+۱}{N_{ic}} \quad (۱۸)$$

اکنون باید ثابت کنیم که $ll'_p > ll'_q$، یعنی

$$(N_{icp}+۱)\log\frac{N_{icp}+۱}{N_{ic}}+N_{icq}\log\frac{N_{icq}}{N_{ic}} > (N_{icq}+۱)\log\frac{N_{icq}+۱}{N_{ic}}+N_{icp}\log\frac{N_{icp}}{N_{ic}} \quad (۱۹)$$

یا

$$(N_{icp}+۱)\log(N_{icp}+۱)-N_{icp}\log N_{icp} > (N_{icq}+۱)\log(N_{icq}+۱)-N_{icq}\log N_{icq} \quad (۲۰)$$

نامساوی فوق با توجه به لم ۱ صحیح می‌باشد.

در مثال قبل با توجه به (۱۴) داریم $\lambda'_{:x}=۱/۴$ و $\lambda'_{:x}=۲/۴$، بنابراین $ll'_{:x}<ll'_{:x}$. لذا از قضیه ۲ باید

$$ll'_{:x}=۴\log\frac{۴}{۸}+۳\log\frac{۳}{۸}+۲\log\frac{۲}{۴}+۲\log\frac{۲}{۴}+۲\log\frac{۲}{۳}+\log\frac{۱}{۳}=-۴/۵۱۵۳$$
$$ll'_{:x}=۴\log\frac{۴}{۸}+۳\log\frac{۳}{۸}+\log\frac{۱}{۴}+۳\log\frac{۳}{۴}+۲\log\frac{۲}{۳}+\log\frac{۱}{۳}=-۴/۲۸۸ \quad (۲۱)$$

٤-۲-۲ حالت دوم: V_i دارای m والد و l فرزند

تعریف ۳: اگر V_i دارای m $(m \geq ۰)$ والد و l $(l \geq ۱)$ فرزند باشد (شکل ۳- ب) آنگاه پارامتر مؤثر λ_p برای زیرمتغیر η_p به صورت زیر خواهد بود

Archive of SID
www.SID.ir

$$\lambda_p = (N_{isp}+1)\log(N_{isp}+1) - N_{isp}\log N_{isp} + \sum_{k=1}^{l}((N_{j_k(b_k,p)c_k}+1)\log(N_{j_k(b_k,p)c_k}+1) - N_{j_k(b_k,p)c_k}\log N_{j_k(b_k,p)c_k}) - \sum_{k=1}^{l}((N_{j_k(b_k,p)}+1)\log(N_{j_k(b_k,p)}+1) - N_{j_k(b_k,p)}\log N_{j_k(b_k,p)}) \quad (۲۲)$$

که N_{isp} تعداد رکوردهایی است که V_i مقدار p ام و والدهایش پیکربندی s ام را گرفته‌اند. $N_{j_k(b_k,p)c_k}$ تعداد رکوردهایی است که V_{j_k} مقدار c_k ام و والدهایش بدون در نظر گرفتن V_i پیکربندی حالت b_k ام و V_i مقدار p ام را گرفته‌اند و $N_{j_k(b_k,p)}$ تعداد رکوردهایی است که والدهای V_{j_k} بدون در نظر گرفتن V_i، پیکربندی حالت b_k ام و V_i مقدار p ام را گرفته‌اند.

بدین ترتیب با شمارش تعداد رکوردها در هر حالت می‌توان مقدار پارامتر مؤثر را برای هر زیرمتغیر محاسبه کرد. در نهایت زیرمتغیری مقدار ۱ می‌گیرد (یعنی مقداری برای مقدار گم‌شده در نظر گرفته می‌شود) که دارای بیشترین مقدار پارامتر مؤثر است.

اگر $m=0$ (یعنی V_i والد نداشته باشد) آنگاه (۲۲) به صورت زیر ساده خواهد شد

$$\lambda_p = \sum_{k=1}^{l}((N_{j_k pc_k}+1)\log(N_{j_k pc_k}+1) - N_{j_k pc_k}\log N_{j_k pc_k}) - (l-1)((N_{i1p}+1)\log(N_{i1p}+1) - N_{i1p}\log N_{i1p}) \quad (۲۳)$$

اکنون قضیه ۱ را برای این حالت نیز اثبات می‌کنیم:

قضیه ۳: فرض کنید V_i دارای m $(m\geq 0)$ والد و l $(l\geq 1)$ فرزند باشد. اگر η_p و η_q دو زیرمتغیر از متامتغیر H باشند به طوری که $\lambda_p > \lambda_q$، آنگاه $ll'_p > ll'_q$.

اثبات: فرض کنید $V_{j_1},\ldots,V_{j_l}$ فرزندان V_i می‌باشند و در رکورد مزبور که مقدار V_i گم شده است، فرزندان در پیکربندی c ام خود قرار داشته‌اند (یعنی هر یک از فرزندان V_{j_k} مقدار c_k ام ممکن خود را داشته‌اند) و والدهایشان بدون در نظر گرفتن V_i پیکربندی b_k ام ممکن خود و والدهای V_i پیکربندی s ام خود را داشته‌اند. مشابه اثبات قضیه ۲ تنها بخش‌هایی از لگاریتم درست‌نمایی نوشته می‌شوند که حاوی زیرمتغیرها هستند

$$ll = \sum_{j=1,j\neq s}^{q_i}\sum_{d=1}^{r_i}(N_{ijd}\log\frac{N_{ijd}}{N_{ij}}) + \sum_{d=1}^{r_i}((N_{isd}+\eta_d)\log\frac{N_{isd}+\eta_d}{N_{is}}) + \sum_{k=1}^{l}\sum_{u=1}^{t_k}\sum_{d=1}^{r_i}(\sum_{w=1}^{r_{ik}}N_{j_k(u,d)w}\log\frac{N_{j_k(u,d)w}}{N_{j_k(u,d)}+\eta_d} + (N_{j_k(u,d)c_k}+\eta_d)\log\frac{N_{j_k(u,d)c_k}+\eta_d}{N_{j_k(u,d)}+\eta_d}) \quad (۲۴)$$

که $N_{j_k(u,p)c_k}$ تعداد رکوردهایی است که V_{j_k} (یکی از فرزندان V_i) مقدار c_k ام و تمامی والدانش به جز V_i پیکربندی u ام و V_i مقدار d ام ممکن خود را گرفته‌اند و t_k تعداد کل پیکربندی‌های والدهای V_{j_k} بدون در نظر گرفتن V_i می‌باشد.

بدین ترتیب ll'_p به صورت زیر خواهد بود، یعنی ll' که $\eta_p=1$ و دیگر زیرمتغیرها مقدار ۰ گرفته‌اند

Algorithm 2. *CalcEffeParam*(G,m,p)

Input. G : BN structure, m : index of the missing value, p : index of the possible value which can be assigned to this missing value

Output. λ_p : the effective parameter for η_p

(1) Suppose in m th missing value, missingness occurs in variable V_i

(2) **if** V_i has some parents but no children **then**

(3) Calculate λ_p according to (14)

(4) **else** //if V_i has children and parents

(5) Calculate λ_p according to (22)

(6) **end if**

شکل ۴: روال محاسبه پارامتر مؤثر با توجه به ساختار شبکه بیزی (تابع CalcEffeParam()).

$$ll'_p = \sum_{j=1}^{q_i}\sum_{d=1,d\neq p}^{r_i}(N_{ijd}\log\frac{N_{ijd}}{N_{ij}}) + (N_{isp}+1)\log\frac{N_{isp}+1}{N_{is}} + \sum_{k=1}^{l}(\sum_{w=1,w\neq c_k}^{r_{jk}}N_{j_k(b_k,p)w}\log\frac{N_{j_k(b_k,p)w}}{N_{j_k(b_k,p)}+1} + (N_{j_k(b_k,p)c_k}+1)\log\frac{N_{j_k(b_k,p)c_k}+1}{N_{j_k(b_k,p)}+1}) \quad (۲۵)$$

نامساوی فوق به صورت مشابه برای ll'_q نیز نوشته می‌شود. باید ثابت کنیم که $ll'_p > ll'_q$ که پس از ساده‌سازی به صورت زیر در خواهد آمد

$$\begin{aligned}&(N_{isp}+1)\log(N_{isp}+1) - N_{isp}\log N_{isp}\\ &+\sum_{k=1}^{l}((N_{j_k(b_k,p)c_k}+1)\log(N_{j_k(b_k,p)c_k}+1) - N_{j_k(b_k,p)c_k}\log N_{j_k(b_k,p)c_k})\\ &-\sum_{k=1}^{l}((N_{j_k(b_k,p)}+1)\log(N_{j_k(b_k,p)}+1) - N_{j_k(b_k,p)}\log N_{j_k(b_k,p)})\\ &> (N_{isq}+1)\log(N_{isq}+1) - N_{isq}\log N_{isq}\\ &+\sum_{k=1}^{l}((N_{j_k(b_k,q)c_k}+1)\log(N_{j_k(b_k,q)c_k}+1) - N_{j_k(b_k,q)c_k}\log N_{j_k(b_k,q)c_k})\\ &-\sum_{k=1}^{l}((N_{j_k(b_k,q)}+1)\log(N_{j_k(b_k,q)}+1) - N_{j_k(b_k,q)}\log N_{j_k(b_k,q)})\end{aligned} \quad (۲۶)$$

چون $\lambda_p > \lambda_q$، لذا نامساوی فوق همیشه برقرار و حکم صحیح است.

۴-۲-۳ روال پیشنهادی برای محاسبه پارامتر مؤثر

اکنون که تعریف دقیق پارامتر مؤثر برای دو مدل ساختار شبکه بیزی بیان شد، بدنه تابع CalcEffeParam() مطابق شکل ۴ خواهد بود.

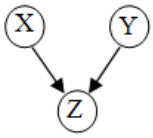
۴-۳ تحلیل پیچیدگی زمانی الگوریتم پیشنهادی

با توجه به شکل ۲ برای هر مقدار گم‌شده، بسته به ساختار شبکه بیزی در گره مربوطه، پارامتر مؤثر برای تمامی مقادیر ممکن که می‌توانند به آن مقدار گم‌شده تخصیص یابند، محاسبه می‌شود. سپس مقداری انتخاب می‌شود که بیشترین پارامتر مؤثر را دارد. با این توضیح برای هر مقدار گم‌شده که در گره V_i رخ می‌دهد، r_i پارامتر مؤثر محاسبه می‌شود. محاسبه پارامتر مؤثر از یکی از دو حالت زیر خواهد بود:

۱) V_i دارای m والد و بدون فرزند است. از تعریف ۲ برای چنین حالتی تنها یک کسر محاسبه می‌شود و لذا محاسبه پارامتر مؤثر از مرتبه $O(1)$ است.

Archive of SID

www.SID.ir



شکل ۵: تلاقی سربه‌سر، X و Y نباید مجاور باشند.

۲) V_i دارای m والد و l فرزند است. از تعریف ۳ برای چنین حالتی تعداد فرزندان نیز مهم است و لذا محاسبه پارامتر مؤثر از مرتبه $O(l)$ است.

با در نظر گرفتن بیشترین حالت (یعنی $O(l)$) برای محاسبه پارامتر مؤثر، مرتبه زمانی کل روش $O(M.r.l)$ خواهد شد که M تعداد مقادیر گم‌شده، r بیشترین تعداد مقادیر ممکنی که متغیرها می‌توانند داشته باشند $(r = \max_i r_i)$ و l بیشترین تعداد فرزند در شبکه است.

برای مقایسه بهتر، فضای جستجوی مسأله بهینه‌سازی بخش ۴، برابر $\prod_{i=1}^{M} r_j$ است که j اندیس متغیری است که iامین مقدار گم‌شده در آن رخ داده است. با این توضیح، میزان کاهش فضای جستجو به کمک روش پیشنهادی کاملاً محسوس می‌باشد.

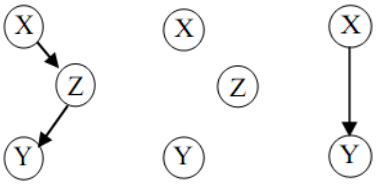
۵- نتایج تجربی

برای نشان‌دادن میزان کارایی روش پیشنهادی برای یادگیری پارامترها، در مقایسه با دو روش بیشینه‌سازی امید و نمونه‌گیری گیبس، هر دوی این روش‌ها را به دو شبکه بیزی معروف ASIA [۳۱] و Cancer - Neapolitan [۳۲] اعمال کرده‌ایم. برای این منظور داده را به کمک نرم‌افزار Netica[1]، با تعداد ۱۰۰۰ نمونه و درصدهای متفاوتی از مقادیر گم‌شده ایجاد کردیم. در هر لایه مارکفی نیز حداکثر یک مقدار گم‌شده رخ داده است، یعنی ۶۰٪ مقدار گم‌شده بدین معنا است که در ۶۰٪ لایه‌های مارکفی به طور تصادفی یک مقدار گم‌شده وجود دارد. تمامی برنامه‌ها به زبان جاوا نوشته شده‌اند و بر روی کامپیوتری با CPU ۳/۲ GHz و RAM ۱/۰ GB اجرا شده‌اند.

علاوه بر این پس از تکمیل داده به کمک روش پیشنهادی، ساختار بهینه شبکه بیزی را با فرض ساختار نامشخص از روی این داده کامل‌شده به کمک الگوریتمی که در [۱۸] ارائه داده بودیم، به دست آوردیم. سپس ساختار به دست آمده را با ساختار اولیه مقایسه کرده‌ایم تا کارایی روش پیشنهادی را از این جهت نیز بررسی کرده باشیم. برای مقایسه این دو، معیار جدیدی با نام "جریمه کل" را ارائه دادیم. این معیار، میانگین وزنی از تعداد یال‌های نادرست یعنی یال‌های کم، اضافه و جهت نادرست است که با نرمال‌سازی بین ۰ (بهترین حالت) و ۱ (بدترین حالت) خواهد بود.

۵-۱ معیار جریمه کل برای ارزیابی ساختار به دست آمده شبکه بیزی

برای به دست آوردن این معیار، تمامی حالات رخ‌دادن یک اشتباه در ساختار شبکه بیزی را بررسی کرده‌ایم: یالی کم (یالی که باید می‌بود ولی نیست)، یال اضافه (یالی که نباید می‌بود ولی هست) و یال با جهت نادرست. برای هر حالت، وزنی در نظر گرفته و سپس میانگین وزنی آنها را به دست می‌آوریم. تنها نکته‌ای که باید بیان شود این است که یال‌های با جهت نادرست همیشه نباید جزء این موارد شمارش شوند و به تابع امتیازدهی که موقع ساخت استفاده شده بستگی دارند. در اینجا ما از تابع امتیازدهی BIC استفاده نمودیم، لذا برخی از چنین یال‌هایی نباید در



شکل ۶: حالات مختلف یال‌های کم و اضافه در ساختار شبکه بیزی.

شمارش به حساب آیند. علت این امر آن است که در برخی توابع امتیازدهی که از ویژگی برابری مارکفی حمایت می‌کنند، همچون BIC، دو ساختار شبکه بیزی با لینک‌های یکسان (لینک یعنی یال‌ها بدون در نظر گرفتن جهتشان) و تلاقی‌های سربه‌سر یکسان (منظور از تلاقی سربه‌سر شکل ۵ می‌باشد) دارای یک امتیاز هستند [۳۲]. بنابراین اگر یک یال با جهت نادرست، یک یال از تلاقی سربه‌سر است، آنگاه w_{wrong} به عنوان جریمه در نظر گرفته شده و به جریمه کل اضافه می‌گردد. در غیر این صورت، این یال‌ها در جریمه کل در نظر گرفته نمی‌شوند.

حالات مختلفی برای یال اضافه یا کم ممکن است رخ دهد. مواردی در اینجا در نظر گرفته شده‌اند (شکل ۶) که بر روی ساختار شبکه بیزی تأثیر می‌گذارند. برای هر حالت وزن متفاوتی در نظر گرفته شده است. فرض کنید بخشی از ساختار شبکه بیزی مطابق شکل ۶- الف باشد. در این صورت حالات زیر ممکن است رخ دهند:

- بدترین حالت برای نبودن یال $X \to Y$ که دارای بیشترین وزن جریمه نیز هست، در شکل ۶- ب نشان داده شده است. در این حالت یال $X \to Y$ موجود نبوده و هیچ یال جبران‌کننده‌ای نیز وجود ندارد. وزن جریمه این حالت با w_{miss1} نشان داده می‌شود.
- حالت بعدی که وزن جریمه کمتری نسبت به مورد قبلی دارد، مطابق شکل ۷- ج، زمانی است که یال $X \to Y$ نبوده، اما دو یال جبران‌کننده $X \to Z$ و $Z \to Y$ به جای آن در ساختار موجود هستند. در این حالت وزن جریمه با w_{miss2} نشان داده می‌شود. توجه شود که $w_{miss2} < w_{miss1}$.
- آخرین حالت که بهترین حالت نیز هست، زمانی است که یال $X \to Y$ نبوده ولی یک یال جبران‌کننده $X \to Z$ در ساختار موجود باشد و Z یکی از والدهای Y در ساختار صحیح شبکه بیزی و نیز در ساختار جدید به دست آمده است، لذا $Z \to Y$. شکل ۶- ج مجدداً نشان‌دهنده این حالت است و وزن جریمه w_{miss3} برای این حالت در نظر گرفته شده است. توجه شود که $w_{miss3} < w_{miss2} < w_{miss1}$.

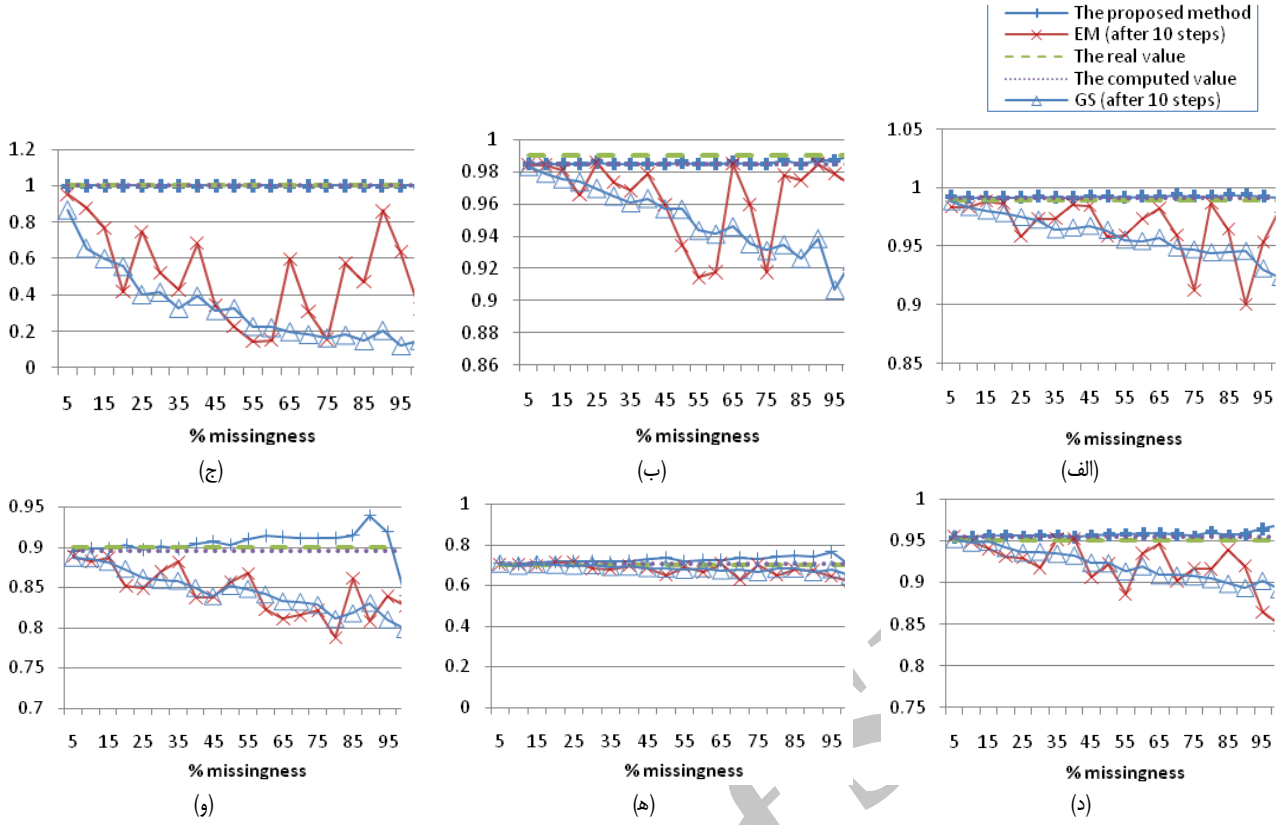
اگر یال کم یا اضافه در یکی از حالات فوق صدق نکرد، موارد زیر در نظر گرفته شده‌اند. اگر یال اضافه $X \to Y$ در ساختار به دست آمده شبکه باشد، یکی از حالات زیر ممکن است رخ دهند:

- دو یال $X \to Z$ و $Z \to Y$ (مطابق شکل ۶- ج) در ساختار صحیح شبکه بیزی وجود داشته باشند ولی در ساختار به دست آمده شبکه یال $X \to Y$ اضافه شده در حالی که یکی از یال‌های $X \to Z$ یا $Z \to Y$ حذف شده است. وزن جریمه چنین حالتی w_{extra1} در نظر گرفته شده است.
- در غیر این صورت وزن جریمه w_{extra2} در نظر گرفته می‌شود. توجه شود که $w_{extra1} < w_{extra2}$.

حالات متعدد دیگری نیز ممکن است رخ دهند که در این مقاله با w_{miss1} یا w_{extra1} در نظر گرفته شده‌اند. در نهایت جریمه کل از رابطه زیر به دست می‌آید

1. http://www.norsys.com/

Archive of SID
www.SID.ir



شکل ۷: نتیجه اجرای روش پیشنهادی، بیشینه‌سازی امید (پس از ۱۰ مرحله) و نمونه‌گیری گیبس (پس از ۱۰ مرحله) در شبکه بیزی ASIA با تعداد نمونه ۱۰۰۰ و درصدهای متفاوت از مقادیر گم‌شده برای احتمالات شرطی، (الف) $p(\text{VisitAsia} = \text{NoVisit})$، (ب) $p(\text{Tuberculosis} = \text{Absent}|\text{VisitAsia} = \text{NoVisit})$، (ج) $p(\text{TbOrCa} = \text{True}|\text{Tuberculosis} = \text{Present}, \text{Cancer} = \text{Absent})$، (د) $p(\text{XRay} = \text{Normal}|\text{TbOrCa} = \text{False})$، (ه) $p(\text{Bronchitis} = \text{Absent}|\text{Smoking} = \text{NonSmoker})$ و (و) $p(\text{Dyspnea} = \text{Absent}|\text{TbOrCa} = \text{False}, \text{Bronchitis} = \text{Absent})$.

۱. $VisitAsia[Visit, NoVisit]$
۲. $Tuberculosis[Absent, Present]$
۳. $smoking[Smoker, NonSmoker]$
۴. $Cancer[Absent, Present]$
۵. $TbOrCa[False, True]$
۶. $XRay[Normal, Abnormal]$
۷. $Bronchitis[Absent, Present]$
۸. $Dyspnea[Absent, Present]$

شکل ۷ نتایج اجرای روش پیشنهادی و دو روش بیشینه‌سازی امید و نمونه‌گیری گیبس، هر دو پس از ۱۰ مرحله بر روی این شبکه با تعداد نمونه ۱۰۰۰ و درصدهای مختلفی از مقادیر گم‌شده برای پارامترهای زیر نشان می‌دهد

$$p(VisitAsia = NoVisit)$$
$$p(Tuberculosis = Absent|VisitAsia = NoVisit)$$
$$p(TbOrCa = True \mid Tuberculosis = Present, Cancer = Absent)$$
$$p(XRay = Normal|TbOrCa = False)$$
$$p(Bronchitis = Absent|Smoking = NonSmoker)$$
$$p(Dyspnea = Absent \mid TbOrCa = False, Bronchitis = Absent)$$

دیگر پارامترهای این شبکه به علت کمبود فضا گزارش نشدند. در شکل ۷ مقدار صحیح احتمال، مقدار محاسبه‌شده از داده بدون مقدار گم‌شده و مقدار محاسبه‌شده از داده حاوی مقادیر گم‌شده گزارش شده

$$\rho = n_{miss1}W_{miss1} + n_{miss2}w_{miss2} + n_{miss3}w_{miss3} + n_{extra1}w_{extra1} + n_{extra2}w_{extra2} + n_{wrong}w_{wrong} \quad (۲۷)$$

که n_x تعداد یال‌هایی است که در حالت x صدق می‌کنند. بهترین جریمه ۰ است، زمانی که تمامی $n_x = ۰$. بدترین حالت زمانی است که تمامی یال‌های شبکه صحیح در حالتی با وزن جریمه w_{miss1} کم و تمامی یال‌های شبکه به دست آمده در حالتی با وزن w_{extra2} یا w_{wrong} قرار داشته باشند، یعنی

$$n_{\max} = \max(w_{extra2}, w_{wrong}) + |E| w_{miss1} \quad (۲۸)$$

که $|E|$ تعداد یال‌هایی است که در شبکه بیزی صحیح وجود دارد و $n_{\max}$ بیشترین تعداد یال‌های یک گراف بدون دور جهت‌دار با n گره است که پس از حذف یال‌های شبکه بیزی صحیح به دست می‌آیند، بنابراین $n_{\max} = (|V|(|V|-۱)/۲) - |E|$. لذا جریمه نهایی می‌تواند به صورت زیر نرمال‌سازی شود

$$\hat{p} = \frac{\rho}{|E| w_{miss1} + (\frac{|V|(|V|-۱)}{۲} - |E|)\max(w_{extra2}, w_{wrong})} \quad (۲۹)$$

در این مقاله وزن‌های زیر در نظر گرفته شده‌اند: $w_{miss1} = ۱$، $w_{miss2} = ۰/۷$، $w_{miss3} = ۰/۱$، $w_{extra1} = ۰/۵$، $w_{extra2} = ۱$ و $w_{wrong} = ۰/۷$.

۵-۲ شبکه بیزی ASIA

شبکه ASIA از ۸ متغیر گسسته دودویی زیر و ۸ یال تشکیل شده است

Archive of SID
www.SID.ir

جدول ۲: زمان اجرای (به ثانیه) روش پیشنهادی، بیشینه‌سازی امید (پس از ۱۰ مرحله) و نمونه‌گیری گیبس (پس از ۱۰ مرحله) برای شبکه بیزی ASIA با تعداد نمونه ۱۰۰۰ و درصدهای متفاوتی از مقادیر گم‌شده.

GS (پس از ۱۰ گام)	EM (پس از ۱۰ گام)	روش پیشنهادی	درصد گم‌شدن
۲۰۰٫۷۶۵	۷٫۵۴۷	۰٫۰۰۱	۵
۳۸۲٫۲۱۹	۱۸٫۵۷۹	۰٫۰۱۵	۱۰
۵۶۶٫۲۹۷	۳۴٫۵۱۶	۰٫۰۳۱	۱۵
۷۵۶٫۲۶۶	۵۶٫۳۴۴	۰٫۰۳۱	۲۰
۹۴۲٫۶۷۲	۸۵٫۹۶۹	۰٫۰۳۱	۲۵
۱۱۴۰٫۶۵۶	۱۲۲٫۴۳۸	۰٫۰۴۶	۳۰
۱۳۳۲٫۰۳۱	۱۶۴٫۵۰۰	۰٫۰۶۲	۳۵
۱۵۲۴٫۷۱۹	۲۰۶٫۷۸۱	۰٫۰۶۲	۴۰
۱۷۲۴٫۰۴۷	۲۷۲٫۶۷۱	۰٫۰۷۸	۴۵
۱۹۱۹٫۵۷۸	۳۳۷٫۱۲۵	۰٫۰۷۸	۵۰
۲۱۱۳٫۳۹۱	۳۹۹٫۶۰۹	۰٫۰۷۸	۵۵
۲۳۱۳٫۹۳۷	۴۷۲٫۴۸۵	۰٫۰۹۳	۶۰
۲۵۱۳٫۹۳۷	۵۳۳٫۱۲۵	۰٫۰۹۳	۶۵
۲۷۱۷٫۶۲۵	۶۱۹٫۸۷۵	۰٫۱۰۹	۷۰
۲۹۰۸٫۵۰۰	۷۳۰٫۹۶۹	۰٫۱۰۹	۷۵
۳۱۱۳٫۱۴۰	۸۲۳٫۷۹۷	۰٫۱۲۵	۸۰
۳۳۱۶٫۴۸۴	۹۲۷٫۰۱۶	۰٫۱۲۵	۸۵
۳۵۵۹٫۱۸۷	۱۰۳۸٫۰۰۰	۰٫۱۴۰	۹۰
۳۷۲۴٫۸۴۴	۱۱۷۳٫۰۰۰	۰٫۱۴۰	۹۵
۳۹۱۹٫۳۴۴	۱۲۵۶٫۲۰۳	۰٫۱۲۵	۱۰۰

جدول ۳: تعداد یال‌های کم، اضافه، جهت نادرست و جریمه کل نرمال‌شده ساختار به دست آمده شبکه بیزی در مقایسه با ساختار صحیح در شبکه بیزی ASIA با تعداد نمونه ۱۰۰۰ و درصدهای متفاوت مقادیر گم‌شده.

جریمه کل نرمال‌شده	تعداد یال‌های با جهت نادرست	تعداد یال‌های اضافی	تعداد یال‌های کم	درصد گم‌شدن
۰٫۱۶	۱	۲	۴	۰–۹۰
۰٫۱۶	۲	۳	۲	۹۵
۰٫۱۸	۳	۴	۳	۱۰۰

است. زمان اجرای این سه روش در جدول ۲ نمایش داده شده است.

علت اجرای بیشینه‌سازی امید و نمونه‌گیری گیبس برای ۱۰ مرحله، زمان اجرای بالای آنها است. همان طور که در جدول ۲ نشان داده شده است، هر دو روش بیشینه‌سازی امید و نمونه‌گیری گیبس در مقایسه با روش پیشنهادی، به ویژه با افزایش درصد مقادیر گم‌شده، زمان‌بر هستند. بیشینه‌سازی امید را برای درصدهای کم مثل ۵ و ۱۰ تا زمانی که همگرا شود، اجرا کردیم. نتیجه، همان طور که انتظار می‌رود، پارامترهای به دست آمده توسط روش پیشنهادی بود با این تفاوت که برای ۵ و ۱۰ درصد گم‌شده به ترتیب پس از ۹۶ و ۲۵۲ مرحله همگرا شد.

زمان اجرای روش پیشنهادی در برخی موارد تغییر نکرده و یا حتی کاهش می‌یابد. علت این امر این است که زمان اجرا به تعداد فرزندان متغیرهایی که در آنها مقدار گم‌شده رخ داده بستگی دارد. در برخی موارد مقدار گم‌شده برای متغیرهایی رخ داده که فرزندی ندارند و لذا زمان‌بر نیستند.

ساختار به دست آمده شبکه بیزی که به کمک الگوریتم قبلی ما در [۱۸] به دست آمده با داده کامل‌شده توسط روش پیشنهادی، با ساختار صحیح شبکه نیز مقایسه شده است و نتایج در جدول ۳ نشان داده شده‌اند. در این جدول تعداد یال‌های کم، اضافه، با جهت نادرست و جریمه کل نرمال‌شده ساختار به دست آمده نمایش داده شده‌اند.

۵–۳ شبکه بیزی Cancer-Neapolitan

شبکه بیزی Cancer-Neapolitan دارای ۵ متغیر گسسته دودویی زیر و ۵ یال می‌باشد

۱. $MetastaticCancer[Present, Absent]$
۲. $SerumCalsium[Increased, NotIncreased]$
۳. $BrainTumor[Present, Absent]$
۴. $Coma[Present, Absent]$
۵. $SeverHeadaches[Present, Absent]$

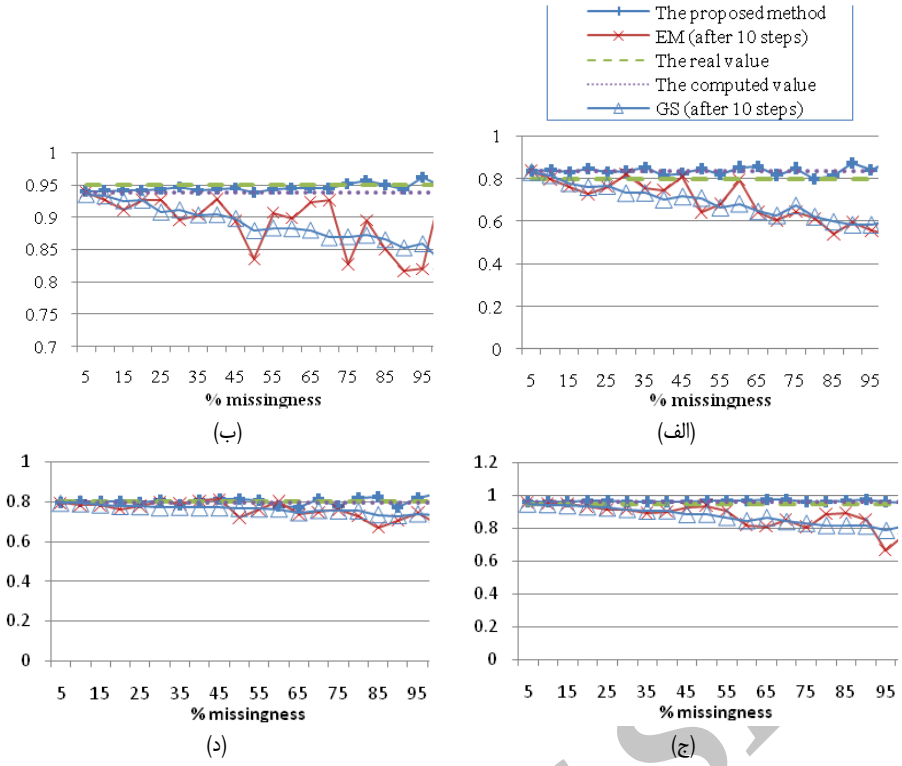
شکل ۸ نشان‌دهنده نتایج اجرای روش پیشنهادی، بیشینه‌سازی امید و نمونه‌گیری گیبس هر دو پس از ۱۰ مرحله بر روی این شبکه با تعداد نمونه ۱۰۰۰ و درصدهای متفاوتی از مقادیر گم‌شده می‌باشد. مجدداً به علت کمبود فضا برخی پارامترها گزارش شده‌اند. زمان اجرای این سه روش در جدول ۴ نشان داده شده است.

به علاوه مشابه شبکه قبلی، ساختار به دست آمده شبکه بیزی از داده کامل‌شده را نیز با ساختار صحیح آن در جدول ۵ مقایسه کرده‌ایم. از آنجا که یال‌های جهت نادرست در این مثال در تلاقی سربه‌سر شرکت ندارند، در جریمه کل مؤثر نبوده و اغلب جریمه مقدار ۰ دارد.

همان گونه که در شکل‌های ۷ و ۸ و جداول ۲ و ۴ دیدیم، روش

Archive of SID

www.SID.ir



شکل ۸: نتایج اجرای روش پیشنهادی، بیشینه‌سازی امید (پس از ۱۰ مرحله) و نمونه‌گیری گیبس (پس از ۱۰ مرحله) در شبکه بیزی Cancer - Neapolitan با تعداد نمونه ۱۰۰۰ و درصدهای مختلفی از مقادیر گم‌شده برای احتمالات شرطی، (الف) $p(\text{SerumCalsium} = -\text{NotIncreased} \mid \text{MetastaticCancer} = \text{Absent})$ ، (ب) $p(\text{BrainTumor} = \text{Absent} \mid \text{MetastaticCancer} = \text{Absent})$ ، (ج) $p(\text{Coma} = \text{Absent} \mid \text{SerumCalsium} = \text{NotIncreased}, \text{BrainTumor} = \text{Absent})$ و (د) $p(\text{MetastaticCancer} = \text{Absent})$.

جدول ۴: زمان اجرای (به ثانیه) روش پیشنهادی، بیشینه‌سازی امید (پس از ۱۰ مرحله) و نمونه‌گیری گیبس (پس از ۱۰ مرحله) برای شبکه بیزی CANCER - NEAPOLITAN با تعداد نمونه ۱۰۰۰ و درصدهای متفاوتی از مقادیر گم‌شده.

درصد گم‌شدن	روش پیشنهادی	EM (پس از ۱۰ گام)	GS (پس از ۱۰ گام)
۵	۰٫۰۰۱	۴٫۹۳۷	۱۱۹٫۳۴۴
۱۰	۰٫۰۰۱	۱۳٫۰۶۲	۲۳۵٫۳۹۰
۱۵	۰٫۰۱۵	۳۴٫۷۱۹	۳۵۱٫۰۹۴
۲۰	۰٫۰۱۵	۴۷٫۷۶۶	۴۷۲٫۵۷۸
۲۵	۰٫۰۱۵	۸۰٫۵۳۱	۵۹۱٫۰۳۲
۳۰	۰٫۰۱۵	۱۰۳٫۰۹۴	۷۰۸٫۲۰۳
۳۵	۰٫۰۳۱	۱۳۴٫۰۴۷	۷۸۳٫۴۵۳
۴۰	۰٫۰۳۱	۱۷۹٫۳۲۹	۱۰۴۷٫۱۸۸
۴۵	۰٫۰۴۷	۲۲۵٫۶۴۱	۱۱۵۲٫۹۵۳
۵۰	۰٫۰۴۷	۲۸۹٫۳۴۴	۱۲۶۰٫۰۱۶
۵۵	۰٫۰۴۷	۳۴۲٫۴۸۴	۱۳۹۹٫۵۹۴
۶۰	۰٫۰۶۲	۴۱۲٫۶۲۵	۱۵۲۹٫۵۹۳
۶۵	۰٫۰۶۲	۴۷۲٫۶۸۸	۱۶۵۱٫۰۶۳
۷۰	۰٫۰۶۲	۵۵۹٫۲۰۳	۱۷۳۱٫۵۰۰
۷۵	۰٫۰۶۲	۶۲۴٫۴۳۸	۱۷۵۱٫۳۷۵
۸۰	۰٫۰۶۲	۷۲۹٫۳۱۳	۱۸۴۱٫۵۴۷
۸۵	۰٫۰۶۲	۸۲۵٫۴۲۲	۱۹۶۶٫۲۱۹
۹۰	۰٫۰۶۲	۹۳۸٫۰۱۶	۲۰۸۳٫۶۸۷
۹۵	۰٫۰۶۲	۱۰۰۴٫۸۱۳	۲۲۰۵٫۸۷۵
۱۰۰	۰٫۰۶۲	۱۱۱۳٫۳۷۵	۲۳۲۹٫۶۵۶

پیشنهادی به پارامترهای بهینه شبکه بیزی از دیدگاه درست‌نمایی، در مدت زمان اندکی دست می‌یابد در حالی که دو روش بیشینه‌سازی امید و نمونه‌گیری گیبس، دو روش تکراری هستند که بسته به نقطه شروع به جواب‌های بهینه محلی دست می‌یابند.

Archive of SID
www.SID.ir

جدول ۵: تعداد یال‌های کم، اضافه، جهت نادرست و جریمه کل ساختار به دست آمده شبکه بیزی از داده کامل‌شده توسط روش پیشنهادی در شبکه بیزی CANCER - NEAPOLITAN با تعداد نمونه ۱۰۰۰ و درصدهای متفاوت از مقادیر گم‌شده.

درصد گم‌شدن	تعداد یال‌های کم	تعداد یال‌های اضافی	تعداد یال‌های با جهت نادرست	جریمه کل نرمال‌شده
۰–۲۰	۰	۰	۱	۰
۲۵	۰	۰	۲	۰
۳۰–۳۵	۰	۰	۱	۰
۴۰–۴۵	۰	۰	۲	۰
۵۰	۰	۰	۱	۰
۵۵	۰	۰	۲	۰
۶۰	۰	۰	۱	۰
۶۰–۸۵	۰	۰	۲	۰
۹۰	۱	۰	۱	۰/۱
۹۵	۰	۰	۲	۰
۱۰۰	۱	۱	۱	۰/۲

۶- نتیجه‌گیری

در این مقاله، یادگیری پارامترهای شبکه بیزی از داده حاوی مقادیر گم‌شده مورد مطالعه قرار گرفت. مفهوم جدیدی با نام "پارامتر مؤثر" معرفی شد که به کمک آن می‌توان بهترین مقدار برای مقادیر گم‌شده را بدون نیاز به جستجوی سراسری و تنها با داشتن ساختار شبکه پیدا کرد به طوری که درست‌نمایی بیشینه گردد. برای این منظور به ازای تمامی مقادیر ممکن برای مقادیر گم‌شده مقدار پارامتر مؤثر محاسبه گشته و مقداری تخصیص می‌یابد که دارای بیشترین پارامتر مؤثر باشد. قضیه‌هایی که این ادعا را ثابت می‌کنند نیز اثبات گردید.

برای بررسی میزان کارایی روش، از دو شبکه بیزی استاندارد ASIA و Cancer - Neapolitan با درصدهای متفاوتی از مقادیر گم‌شده استفاده کردیم. حاصل کار با نتایج دو روش شناخته‌شده بیشینه‌سازی امید و نمونه‌گیری گیبس مقایسه شد. نتایج حاکی از مدت زمان اجرای اندک روش جدید در مقایسه با روش‌های پیشین است. به علاوه پارامترهای نهایی از دیدگاه درست‌نمایی بهینه می‌باشند.

مراجع

[1] N. Friedman, M. Linial, I. Nachman, and D. Peer, "Using Bayesian networks to analyze expression data," *Computational Biology*, vol. 7, no. 3-4, pp. 601-620, 2000.

[2] J. Uebersax, *Breast Cancer Risk Modeling: An Application of Bayes Networks*, Technical Report, Ravenpack International, Spain, 2004.

[3] L. M. de Campos, J. M. Fernandez - Luna, and J. F. Huete, "Bayesian networks and information retrieval: an introduction to the special issue," *Information Processing and Management*, vol. 40, no. 5, pp. 727-733, Sep. 2004.

[4] F. J. Diez, J. Mira, E. Iturralde, S. Zubillaga, and A. Diaval, "A Bayesian expert system for echocardiography," *Artificial Intelligence in Medicine*, vol. 10, no. 1, pp. 59-73, May 1997.

[5] D. M. Chickering, "Learning bayesian network is NP - complete," *Learning from Data: Artificial Intelligence and Statistics V*, pp. 121-130, 1996.

[6] D. M. Chickering, C. Meek, and D. Heckerman, "Large-sample learning of Bayesian networks is NP-Hard," in *Proc. 19th Conf. on Uncertainty in Artificial Intelligence, UAI'03*, pp. 124-133, 2003.

[7] Z. Kebaili and A. Aussem, "A novel hybrid Bayesian network structure learning algorithm based on correlated itemset mining techniques," *Int. J. of Computational Intelligence Research*, vol. 5, no. 1, pp. 16-21, 2009.

[8] 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.

[9] D. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581-592, 1976.

[10] R. Little and D. Rubin, *Statistical Analysis with Missing Data*, Wiley - Interscience, 2002.

[11] G. Elidan, I. Nachman, and N. Friedman, "Ideal parent structure learning for continuous variable Bayesian networks," *J. of Machine Learning Research*, vol. 8, pp. 1799-1833, 2007.

[12] W. Buntine, "Theory refinement on Bayesian networks," in *Proc. 7th Conf. on Uncertainty in Artificial Intelligence, UAI'91*, pp. 52-60, 1991.

[13] T. Silander and P. Myllymaki, "A simple approach for finding the globally optimal Bayesian network structure," in *Proc. 22nd Conf. on Uncertainty in Artificial Intelligence, UAI'06*, pp. 445-452, 2006.

[14] A. P. Singh and A. W. Moore, *Finding Optimal Bayesian Networks by Dynamic Programming*, Technical Report, Carnegie Mellon University CALD-05-106, 2005.

[15] M. Koivisto and K. Sood, "Exact bayesian structure discovery in bayesian networks," *J. of Machine Learning Research*, vol. 5, pp. 549-573, 2004.

[16] J. Suzuki, "Learning Bayesian belief networks based on the minimum description length principle: an efficient algorithm using the branch and bound technique," *IEICE Trans. Inf. & Syst.*, vol. E82-D, no. 2, pp. 356-367, Feb. 1999.

[17] C. P. de Campos, Z. Zeng, and Q. Ji, "Structure learning of Bayesian networks using constraints," in *Proc. 26th Int. Conf. on Machine Learning, ICML'09*, 2009.

[18] K. Etminani, M. Naghibzadeh, and A. R. Razavi, "Globally optimal structure learning of bayesian networks from data," *Lecture Notes in Computer Science*, vol. 6352, pp. 101-106, 2010.

[19] D. M. Chickering and D. Heckerman, "Efficient approximations for the marginal likelihood of bayesian networks with hidden variables," *Machine Learning*, vol. 29, no. 2-3, pp. 181-212, 1997.

[20] M. Ramoni and P. Sebastiani, "Learning Bayesian networks from incomplete databases," in *Proc. of the Conf. on Uncertainty in AI*, pp. 401-408, 1997.

[21] N. Friedman, "Learning belief network in the presence of missing values and hidden variables," in *Proc.14th Int. Conf. on Machine Learning*, pp. 125-133, 1997.

[22] N. Friedman, "The bayesian structural EM algorithm," in *Proc. 14th Conf. on Uncertainty in Artificial Intelligence*, 1998.

[23] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of the Royal Statistical Society B*, vol. 39, no. 1, pp. 1-39, 1977.

[24] P. Leray and O. Franois, "Bayesian network structural learning and incomplete data," in *Proc. Int. and Interdisciplinary Conf. on Adaptive Knowledge Representation and Reasoning*, 2005.

[25] N. Friedman and M. Goldszmidt, "Discretizing continuous attributes while learning bayesian networks," in *Proc. 13th Int. Conf. on Machine Learning, ICML'96*, pp. 157-165, 1996.

[26] G. 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.

[27] 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.

[28] G. Schwartz, "Estimating the dimensions of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461-464, 1978.

[29] J. Rissanen, "Stochastic complexity (with discussion)," *J. of the Royal Statistical Society*, vol. 49, no. 3, pp. 223-239, 1987.

Archive of SID

www.SID.ir

[30] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. on Automatic Control*, vol. 19, no. 6, pp. 716-723, 1974.
[31] S. L. Lauritzen and D. J. Spiegelhalter, "Local computations with probabilitieson graphical structures and their application to expert systems," *J. of Royal Statistics Society*, vol. 50, no. 2, pp. 157-224, 1988.
[32] R. E. Neapolitan, *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*, John Wiley & Sons, Inc., pp. 179-180, 1990.

کبری اطمینانی تحصیلات خود را در مقاطع کارشناسی، کارشناسی ارشد و دکتری تخصصی به ترتیب در سال های ۱۳۸۴، ۱۳۸۶ و ۱۳۹۱ در رشته مهندسی کامپیوتر گرایش نرم افزار از دانشگاه فردوسی مشهد به پایان رسانیده است. در سال ۱۳۹۰ دوره فرصت مطالعاتی را در دانشگاه لینشوپینگ سوئد گذرانیده است. ایشان هم‌اکنون استادیار گروه انفورماتیک پزشکی دانشگاه علوم پزشکی مشهد می‌باشد.

زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: مدل‌های احتمالی از جمله شبکه بیزی، داده‌کاوی بویژه در داده‌های پزشکی و الگوریتم‌های یادگیری ماشین.

محمود نقیب‌زاده مدرک کارشناسی خود را در رشته آمار و علوم کامپیوتر از دانشگاه فردوسی مشهد، و کارشناسی ارشد و دکتری تخصصی خود را در رشته علوم کامپیوتر و مهندسی کامپیوتر از دانشگاه کالیفرنیای جنوبی اخذ نموده است. ایشان در حال حاضر استاد گروه مهندسی کامپیوتر دانشگاه فردوسی مشهد می‌باشد. گروه تحقیقاتی مهندسی دانش (KERG) تحت نظر ایشان در حال انجام فعالیت‌های تحقیقاتی است.

زمینه‌های تحقیقاتی مورد علاقه ایشان شامل زمان‌بندی فرایندها و مهندسی دانش است. تاکنون مقالات متعددی در کنفرانس‌ها و مجلات معتبر و نیز کتب متعددی در حوزه علوم و مهندسی کامپیوتر به چاپ رسیده است. ایشان داور بسیاری از مجلات بوده و از اعضای بسیاری از انجمن‌های کامپیوتر از جمله IEEE است.

مهدی عمادی تحصیلات خود را در مقاطع کارشناسی؛ کارشناسی ارشد و دکتری آمار به‌ترتیب در سال‌های ۱۳۷۴، ۱۳۷۶ و ۱۳۸۲ از دانشگاه فردوسی مشهد به پایان رسانده است و هم‌اکنون استادیار دانشکده علوم ریاضی گروه آمار است.

زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: استنباط آماری، استنباط بیزی و استنباط شواهدی، ناپارامتری.

امیررضا رضوی در سال ۱۳۷۷ از دانشگاه علوم پزشکی مشهد در رشته پزشکی فارغ‌التحصیل گردید. ایشان مدرک دکتری تخصصی خود را در رشته انفورماتیک پزشکی از دانشگاه لینشوپینگ سوئد در سال ۲۰۰۷ اخذ نموده است و دوره فلوشیپ را نیز در همین دانشگاه به مدت یک سال گذرانیده است. در طول مدت تحقیقات، ایشان بر روی دو پروژه DebugIT که توسط EU مورد حمایت قرار گرفته است و DKOD فعالیت کرده است. ایشان هم‌اکنون استادیار گروه انفورماتیک پزشکی دانشگاه علوم پزشکی مشهد می‌باشد. یکی از مقالات ایشان در حوزه داده کاوی در پزشکی در کتاب سال IMIA در سال ۲۰۰۹ به چاپ رسیده است.

زمینه‌های تحقیقاتی مورد علاقه ایشان، اکتشاف دانش از داده های پزشکی، داده کاوی، راهنماهای بالینی و رجیستری سرطان می‌باشند.

Archive of SID

www.SID.ir