کاهش خطای مدل‌سازی به کمک میانگین‌گیری بیزی

افشین فلاح [۱]، محسن محمدزاده [۲]

چکیده:

در مدل‌سازی غالباً نمی‌توان مدلی یافت که به صورت کامل به داده‌ها برازش داشته باشد، بلکه معمولاً دسته‌ای از مدل‌های رقیب وجود دارد که همگی به داده‌ها برازش مطلوبی دارند. با این وجود در روش‌های مرسوم مدل‌سازی با تکیه بر برخی معیارها، یک مدل به عنوان بهترین مدل انتخاب می‌شود. چنین رویکردی موجب چندگانگی در تصمیم‌گیری، کم براورد کردن عدم حتمیت و در نهایت افزایش خطا می‌شود. میانگین‌گیری از مدل‌ها روشی توانمند در مدل‌سازی است، که در آن میانگین وزنی همهٔ مدل‌ها یا دسته‌ای از بهترین مدل‌ها به عنوان مدل نهایی انتخاب می‌شود. در این مقاله به صورت نظری و همچنین با ارائه یک مثال کاربردی نشان داده می‌شود که این روش کارایی پیش‌بینی را افزایش داده، متغیرهای تأثیرگذار را شناسائی کرده و با تعیین عدم حتمیت واقعی براوردها، خطای مدل‌سازی را نیز کاهش می‌دهد.

واژه‌های کلیدی: عدم حتمیت، توزیع پیش‌بین، عامل بیز، میانگین‌گیری بیزی.

۱ مقدمه

در روش‌های معمول مدل‌سازی، کلاسی از مدل‌ها در نظر گرفته شده و براساس مجموعه‌ای از معیارها یک مدل به عنوان مدل برتر انتخاب می‌شود. در حالی که ممکن است رقبای بسیار خوبی برای مدل انتخابی در فضای مدل وجود داشته باشند. این نوع مدل‌سازی در مرحلهٔ انتخاب با تعدد معیارهای انتخاب و عدم وجود مدلی که براساس همهٔ معیارها مطلوب باشد مواجه است و در مرحلهٔ پس از انتخاب به دلیل تکیه بر یک مدل و کنار گذاشتن سایر مدل‌های رقیب، عدم حتمیت مدل انتخابی نادیده گرفته می‌شود. برای رفع این مشکل میانگین‌گیری بیزی مدل‌ها، از تمام مدل‌های موجود در فضای مدل برای دستیابی به مدلی مناسب استفاده می‌کند. در این روش بر اساس میزان حمایت داده‌ها از هر مدل، وزنی به آن اختصاص داده می‌شود. سپس مدل حاصل از میانگین وزنی همهٔ مدل‌ها برای انجام استنباط و پیش‌بینی به کار گرفته می‌شود. ریشهٔ تاریخی این روش به لیمر [۵] برمی‌گردد، که به دلیل محاسبات دشوار و پیچیده در آن زمان چندان مورد توجه قرار نگرفت. با پیشرفت رایانه‌ها و

[۱] استادیار گروه آمار، دانشگاه بین المللی امام خمینی
[۲] دانشیار گروه آمار، دانشگاه تربیت مدرس

فنون محاسبات تقریبی در اوایل دههٔ نود این روش دوباره مطرح و در کانون توجهات قرار گرفت. مادیگان و رفتری ([۸]، [۹]) دو روش پایه‌ای برای اجرای این شیوهٔ مدل‌سازی ارائه کردند. رفتری[۱۱] و رفتری و همکاران([۱۲]، [۱۳]) میانگین‌گیری بیزی را در چهارچوب مدل‌های رگرسیونی مورد بررسی قرار دادند. هوئتینگ [۲] و هوئتینگ و همکاران([۳]، [۴]) به نحوهٔ انجام محاسبات و چگونگی تعیین مؤلفه‌های لازم در اجرای این روش پرداختند. روبرت و نوبل [۱۴] و لیپکوویچ [۶] میانگین‌گیری بیزی را در حالت چند متغیره مورد توجه قرار دادند. در این مقاله روش میانگین‌گیری بیزی مدل‌ها[۳]، که در آن تمام مدل‌های مطرح در تعیین مدل نهایی نقش دارند و نحوهٔ محاسبهٔ مؤلفه‌های مختلف آن، در بخش ۲ ارائه می‌شود. سپس یکی از روش‌های اجرای آن معرفی و در بخش ۳ بصورت نظری و همچنین با ارائه یک مثال کاربردی به ارزیابی روش BMA پرداخته خواهد شد و نشان داده می‌شود که این روش در مقایسه با مدل‌های بدست آمده از سایر روش‌های رایج مدل‌سازی، از قابلیت پیش‌بینی بهتری برخوردار است، متغیرهای مؤثر را به نحو مطلوب‌تری شناسایی می‌کند و عدم حتمیت واقعی براوردها را تعیین می‌نماید و در مجموع خطای مدل‌سازی کاهش می‌دهد. بحث و نتیجه‌گیری در بخش آخر ارائه می‌شود.

۲ میانگین‌گیری بیزی مدل‌ها

فرض کنید برای متغیر وابستهٔ Y و مجموعه‌ای از متغیرهای پیش‌بین $X_1,\ldots,X_k$، هدف یافتن بهترین مدل از بین همهٔ مدل‌های خطی

$$Y=\beta_0+\sum_{j=1}^{p}\beta_{i_j}X_{i_j}+\varepsilon \qquad (1)$$

باشد، که در آن $X_{i_1},\ldots,X_{i_p}$ و $\beta_{i_1},\ldots,\beta_{i_p}$ به ترتیب زیرمجموعه‌ای از متغیرهای $X_1,\ldots,X_k$ و ضرایب نظیر آن‌ها هستند. بسته به حضور یا عدم حضور هر متغیر پیش‌بین در مدل، 2^k مدل رگرسیونی ممکن وجود دارد، که فرض می‌شود همگی در فضای مدل $\mathcal{M}$ قرار دارند. اگر $\mathcal{M}=\{M_1,\ldots M_T\}$ نشان دهندهٔ فضای مدل، یعنی مجموعه تمام مدل‌های مطرح و Δ کمیّتی مورد علاقه، مانند مشاهده‌ای در آینده باشد. در این صورت استنباط بیزی بر اساس توزیع پسین Δ صورت می‌پذیرد، که با فرض مشاهدهٔ مجموعه‌ی داده‌ی D، بنا بر قاعدهٔ احتمال کل، آمیخته‌ای از احتمال‌های پسین همهٔ مدل‌ها به‌صورت

$$Pr(\Delta|D)=\sum_{k=1}^{T}Pr(\Delta|M_k,D).Pr(M_k|D) \qquad (2)$$

است. میانگین و واریانس پسین Δ به ترتیب به‌صورت

$$\begin{aligned}E[\Delta|D] &= \sum_{k=1}^{T}E[\Delta|M_k,D].Pr(M_k|D)\\ &= \sum_{k=1}^{T}\hat{\Delta}_k.Pr(M_k|D)\end{aligned}$$

[۳] Bayesian Model Averaging (BMA)

و

$$V[\Delta|D] = E_M[V(\Delta|D,M)] + V_M[E(\Delta|D,M)] = \sum_{k=1}^{T}(V[\Delta|D,M_k] + \hat{\Delta}_k^2) \times Pr(M_k|D) - E[\Delta|D]^2$$

هستند [۱۱]، که در آن مؤلفهٔ $Var_M[E(\Delta|D,M)]$ عدم حتمیت بین مدل‌ها را نشان می‌دهد. عبارت (۲) یک میانگین وزنی احتمال پسین است، که در آن هر مدل با احتمال پسین متناظر خود، یعنی $P(M_k|D)$، وزن‌دار شده است و توزیع پیش‌بین Δ به شرط مدل M_k به‌صورت

$$Pr(\Delta|M_k,D) = \int Pr(\Delta|\theta_k,M_k,D).Pr(\theta_k)\ d\theta_k \qquad (۳)$$

است، که در آن θ_k بردار پارامترهای مدل M_k را نشان می‌دهد. احتمال پسین مدل M_k با استفاده از قاعدهٔ بیز به‌صورت

$$Pr(M_k|D) = \frac{Pr(D|M_k).Pr(M_k)}{\sum_{j=1}^{T} P(D|M_j).Pr(M_j)} \qquad (۴)$$

به‌دست می‌آید، که در آن $Pr(M_k)$ احتمال پیشین درست بودن مدل M_k و

$$Pr(D|M_k) = \int Pr(D|\theta_k,M_k).Pr(\theta_k|M_k)\ d\theta_k \qquad (۵)$$

درستنمایی جمع بستهٔ[۴] مدل M_k و $Pr(D|\theta_k,M_k)$ درستنمایی مدل M_k است. برای محاسبهٔ $Pr(\Delta|D)$ باید مؤلفه‌های تشکیل دهنده‌ی آن را مشخص و جایگزین نمود. بنابراین لازم است توزیع پیشین پارامترها، احتمال‌های پیشین و پسین هر مدل و توزیع پیش‌بین کمیّت مورد علاقه را برای همهٔ مدل‌های موجود در فضای مدل، مشخص کرد.

الف— تعیین احتمال‌های پیشین پارامترهای مدل.

از جمله مسائل دشوار در روش BMA، تخصیص پیشین به پارامترهای مدل است. استفاده از پیشین‌های ناسره[۵]، منجر به توزیع‌های پسین ناسره می‌شود، که در این‌صورت نمی‌توان این احتمال‌ها را به‌عنوان احتمال مدل و نسبت آن‌ها را به عنوان عامل بیز تعبیر نمود. به همین دلیل بسیاری از محققان گونه‌های مختلفی از پیشین‌های آگاهی بخش[۶] را پیشنهاد کرده‌اند. هوئتینگ [۲] استفاده از پیشین‌های مناسبی را پیشنهاد کرده است که در قسمت‌هایی از فضای پارامتر با درستنمایی بزرگ هموار باشند. در مدل

$$Y = \beta_0 + \sum_{j=1}^{p}\beta_j X_j + \varepsilon = X\beta + \varepsilon \qquad (۶)$$

که در آن $X_{n\times(p+1)}$ ماتریس مشاهدات، Y بردار n بعدی متغیرهای وابسته و ε بردار خطا است، فرض می‌شود ε ها برای مشاهدات مختلف، مستقل و دارای توزیع نرمال با میانگین صفر و واریانس σ^2 هستند. بردار $\beta = (\beta_0,\ldots,\beta_p)$ و σ^2

[۴] Integrated Likelihood
[۵] Improper Prior
[۶] Informative

پارامترهای نامعلوم هستند. توزیع‌های پیشین این پارامترها بایستی به گونه‌ای تعیین شوند که عدم حتمیت آن‌ها را به خوبی منعکس سازند. برای این منظور هوئتینگ [۲] ردهٔ پیشین‌های مزدوج نرمال—گاما را به‌صورت

$$\beta \sim N_{p+1}(\mu, \sigma^2 V) \quad , \quad \frac{\nu.\lambda}{\sigma^2} \sim \mathcal{X}_\nu^2 \qquad (۷)$$

در نظر گرفت، که در آن‌ها ν، λ، ماتریس $V_{(p+1)\times(p+1)}$ و بردار $p+1$ بعدی μ، ابرپارامترهایی هستند که باید براورد شوند.

ب — تعیین احتمال پیشین هر مدل.

وقتی هیچ اطلاع پیشینی در مورد مدل‌ها وجود ندارد یا میزان اطلاعات پیشین در مورد مدل‌ها اندک است، معمولاً فرض می‌شود توزیع مدل‌ها یکنواخت است. یعنی $Pr(M_j) = \frac{1}{T}$ در این‌صورت احتمال پسین مدل M_k بصورت

$$Pr(M_k|D) = \frac{Pr(D|M_k)}{\sum_{j=1}^{T} Pr(D|M_j)} \qquad (۸)$$

خواهد بود. هنگامی که فضای مدل بزرگ نیست، این انتخاب معقول و قابل دفاع است [۱۱]. اما وقتی تعداد مدل‌ها بسیار زیاد است، این نگرانی پیش می‌آید که چنین انتخابی به نتایج گمراه کننده‌ای منجر شود. مادیگان و رفتری [۹] و مادیگان و یورک [۱۰] ادعا کرده‌اند برای فضاهای مدل بسیار بزرگ، مثلاً فضای با بیش از 10^{12} مدل، چنین اثرات گمراه کننده‌ای را مشاهده نکرده‌اند.

اگر بعضی از مدل‌ها در مقایسه با سایرین دارای احتمال بیشتری باشند یا اطلاعات خوبی در مورد آنها در دسترس باشد، لازم است این اطلاعات برای تعدیل احتمال‌های پیشین مدل‌ها بکار گرفته شوند تا از پیشین‌های آگاهی بخش‌تر استفاده شود. معمولاً در مسائل مربوط به انتخاب متغیرهای پیش‌بین، اطلاعات پیشین به‌شکل شواهد قبلی برای در نظر گرفتن یک متغیر هستند. اگر تنها این نوع از اطلاعات پیشین در اختیار باشد و مدل M_k توسط بردار $(\delta_{k1}, \ldots, \delta_{kp})$ مشخص شود، که در آن

$$\delta_{ki} = \begin{cases} 1 & X_i \in M_i \\ 0 & X_i \notin M_i \end{cases} \qquad i = 1, \ldots, p$$

توابع نشانگر هستند. با در نظر گرفتن π_i به عنوان احتمال مؤثر بودن متغیر X_i بر Y و با فرض اینکه اطلاعات پیشین در مورد متغیرهای متفاوت تقریباً مستقل هستند، می‌توان

$$Pr(M_k) = \prod_{i=1}^{p} [\pi_i^{\delta_{ki}} \times (1-\pi_i)^{1-\delta_{ki}}] \qquad (۹)$$

را به عنوان احتمال پیشین صحیح بودن مدل M_k در نظر گرفت [۸]. چون پیشین (۹) به متغیرهای مهمتر احتمال بزرگتری نسبت می‌دهد، توزیع پیشین متغیر[۷] نامیده می‌شود.

ج — تعیین احتمال پسین هر مدل.

برای محاسبه احتمال پسین هر مدل لازم است $Pr(D|M_k)$ از رابطهٔ (۵) که یک انتگرال با بعد برابر با تعداد پارامترهای مدل M_k است، محاسبه

[۷] Variable Prior

شود. از اینرو محاسبهٔ دقیق این احتمال به دلیل پیچیده بودن انتگرال مربوطه تنها در حالات بسیار خاص و ساده امکان‌پذیر است و در سایر موارد از روش‌های تقریبی و محاسباتی استفاده می‌شود. یکی از این موارد خاص، رگرسیون خطی چند متغیره با فرض خطاهای نرمال و مستقل است. هوئتینگ [۲] نشان داد با استفاده از پیشین‌های نرمال–گاما می‌توان به این مهم دست یافت. در مدل (۶) توزیع حاشیه‌ای Y تحت مدل M_i، براساس پیشین‌های سره (۷)، یک توزیع t استودنت غیرمرکزی n بعدی با ν درجهٔ آزادی، میانگین $X\mu$ و واریانس $\frac{\nu}{\nu-2}\lambda(I+XVX^t)$ است، که در آن X ماتریس طرح، μ بردار میانگین β و V ماتریس واریانس کواریانس β است. در این مقاله از معیار اطلاع بیز[۸] برای تعیین احتمال پسین هر مدل استفاده می‌کنیم. معیار BIC برای رگرسیون خطی به‌صورت

$$BIC_j = n\log(1-R_j^2) + k_j \log n$$

است، که در آن R_j^2 ضریب تعیین تعدیل شده مدل jام، k_j تعداد متغیرهای پیش‌بین موجود در مدل jام و n تعداد مشاهدات را نشان می‌دهد. در این صورت درستنمایی جمع بستهٔ مدل jام، به‌صورت

$$\begin{aligned} Pr(D|M_j) &\propto \exp -0/5BIC_j) \\ &= \exp\{(-0/5(n\log(1-R_j^2) \\ &\quad +k_j\log n)\} \end{aligned}$$

[۸] Bayesian Information Criteria (BIC)

تقریب زده می‌شود. بر این اساس احتمال پسین مدل kام با استفاده از قاعدهٔ بیز به‌صورت

$$Pr(M_k|D) = \frac{\exp\{-0/5BIC_k\}\ .\ Pr(M_k)}{\sum_{j=1}^{T}\exp\{-0/5BIC_j\}\ .\ Pr(M_j)}$$

نوشته می‌شود. چون معیار BIC برای بسیاری از مدل‌ها دارای شکل بسته و ساده‌ای است، استفاده از این تقریب موجب سهولت محاسبه و افزایش سرعت می‌شود. بعلاوه این تقریب خصوصاً در کاربردهایی با تعداد متغیرهای پیش‌بین زیاد از دقت قابل قبولی برخوردار است.

د – تعیین توزیع پیش‌بین.

مؤلفهٔ لازم دیگر برای اجرای روش BMA، توزیع پیش‌بین (۳) است. این انتگرال از دو جزء تشکیل شده است که معمولاً برای جزء دوم آن از تقریب MLE استفاده می‌شود. بر این اساس رفتری و همکاران [۱۳] تقریب

$$Pr(\Delta|M_k,D) \approx Pr(\Delta|M_k,\hat{\theta}_k,D) \quad (10)$$

را برای (۳) مورد استفاده قرار دادند، که در آن $\hat{\theta}_k$ برآورد ML بردار پارامترهای مدل M_k است. بصورت شهودی می‌توان گفت وقتی تعداد متغیرهای پیش‌بین زیاد است، عدم حتمیّت مدل خیلی بیشتر از عدم حتمیّت پارامتر است و می‌توان از عدم حتمیت پارامتر در مقابل آن چشم پوشی کرد. تقریب (۱۰) از عدم حتمیت پارامتر $\hat{\theta}$ چشم‌پوشی می‌کند و آنرا همانند یک پارامتر

برآورد می‌کند. این امر هنگامی که تعداد متغیرهای پیش‌بین زیاد است، معقول می‌باشد و به همین دلیل عملکرد این تقریب بسیار خوب است. علاوه بر تقریب بالا برای توزیع پیش‌بین پسین، در موارد بسیار خاص این امکان وجود دارد که این توزیع را به‌صورت بسته محاسبه کرد [۲].

اگر همهٔ مولفه‌های لازم برای اجرای روش BMA مشخص باشد، محاسبهٔ مجموع (۲) بواسطه تعداد زیاد جملات آن عملاً امکان پذیر نیست. روشی معقول برای اجرای روش BMA یافتن زیر مجموعه‌ای از محتمل‌ترین مدل‌ها و استفاده از آن‌ها در محاسبهٔ مجموع (۲) است. مادیگان و رفتری [۹] روش پنجره اوکام[۹] را برای این منظور پیشنهاد کرده‌اند، که از دو اصل کلی پیروی می‌کند. بنابر اصل اول مدل‌هایی که در مقایسه با محتمل‌ترین مدل خیلی کم شانس هستند، کنار گذاشته می‌شوند. بنابر اصل دوم که تیغ اوکام نامیده می‌شود، مدل‌هایی که نسبت به زیرمدل‌های ساده‌تر خود کمتر از جانب داده‌ها حمایت می‌شوند، کنار گذاشته می‌شوند. با اجرای اصل اول مدل‌هایی که در مجموعهٔ

$$\mathcal{A}' = \left\{ M_k : \frac{Max_{M_k\in\mathcal{M}}\{Pr(M_l|D)\}}{Pr(M_k|D)} > C_1 \right\}$$

قرار دارند و همچنین بنابر اصل تیغ اوکام، مدل‌هایی که در مجموعه

$$\mathcal{B} = \left\{ M_k : \exists M_l \in \mathcal{A}\ s.t\ M_l \subset M_k, \frac{Pr(M_l|D)}{Pr(M_k|D)} > C_2 \right\}$$

باشند، از مجموع (۲) خارج می‌شوند، که در آن C_1 و C_2 توسط تحلیل‌گر تعیین می‌شوند. در این صورت مجموع (۲) را می‌توان بصورت

$$Pr(\Delta|D) = \frac{\sum_{M_k\in\mathcal{A}} Pr(\Delta|M_k, D).Pr(D|M_k).Pr(M_k)}{\sum_{M_k\in A} Pr(D|M_k).Pr(M_k)}$$

بازنویسی کرد، که در آن $\mathcal{A} = \mathcal{A}' - \mathcal{B} \in \mathcal{M}$ مجموعهٔ پذیرش است.

۳ ارزیابی روش BMA

در این بخش ابتدا کارایی مدل حاصل از روش BMA با کارایی سایر مدل‌ها بصورت نظری مورد مقایسه قرار می‌گیرد. سپس با استفاده از یک مثال کاربردی نشان داده می‌شود که روش BMA با لحاظ کردن عدم حتمیت، خطا را در تشخیص متغیرهای مؤثر و براورد پارامترها کاهش می‌دهد.

گود [۱] قاعدهٔ امتیازبندی لگاریتمی[۱۰] را برای هر پیشامد E به‌صورت $-\log Pr(E)$ معرفی نمود، از این قاعده می‌توان به عنوان یک تابع امتیاز استفاده نمود که مقادیر کوچک‌تر آن مطلوبیت بیشتر یا محتمل‌تر بودن E را نشان می‌دهد. با توجه به نامنفی بودن معیار اطلاع کولبک–لیبلر[۱۱] که برای دو اندازه احتمال P و Q به‌صورت

$$K(P;Q) = \int \log(\frac{dP}{dQ})\ dP \geq 0$$

تعریف می‌شود، می‌توان نوشت

$$0 \leq E\log\{\frac{Pr(\Delta|D)}{Pr(M_j|D)}\}$$

[۹] Occam's Window (OW)
[۱۰] Logarithmic Scoring Role (LPC)
[۱۱] Kullback-Liebler Information

$$= E\log\{\frac{\sum_{k=۱}^{T} Pr(\Delta|M_k,D).Pr(M_k|D)}{Pr(M_j|D)}\}$$
$$= E\log\{\sum_{k=۱}^{T} Pr(\Delta|M_k,D).Pr(M_k|D)\}$$
$$-E\log\{Pr(M_j|D)\}$$

که در آن Δ کمیّت قابل مشاهده‌ای است که باید پیش‌بینی شود. بنابراین نامساوی

$$-E[\log\{\sum_{k=۱}^{T}\{Pr(\Delta|M_k,D).Pr(M_k|D)\}] \leq$$
$$-E[\log\{Pr(\Delta|M_j,D)\}], \quad j=۱,\dots,T$$

که سمت چپ آن امتیاز لگاریتمی روش BMA و سمت راست آن نیز امتیاز لگاریتمی سایر مدل‌ها می‌باشد، نشان دهندهٔ قابلیت پیش‌بینی بیشتر روش BMA در مقایسه با سایر مدل‌ها است.

اکنون با ارائه یک مثال کاربردی نشان داده می‌شود که روش BMA در زمینهٔ تشخیص متغیرهای پیش‌بین مؤثر و برآورد ضرایب آنها نیز عملکرد مطلوبتری دارد. بعلاوه این روش عدم حتمیت واقعی هر پارامتر را نشان می‌دهد و اطلاعات هیچ متغیری را هر چند اندک باشد، نادیده نمی‌انگارد. جدول ۱ مشاهدات یک مطالعهٔ رگرسیونی را نشان می‌دهد که هدف آن بررسی تأثیر ۸ متغیر پیش‌بین شامل میزان بارندگی در فروردین ماه ($X_۱$)، دمای هوا در اردیبهشت ماه ($X_۲$)، میزان بارندگی در خرداد ماه ($X_۳$)، دمای هوا در خرداد ماه ($X_۴$)، میزان بارندگی در دی ماه ($X_۵$)، دمای هوا در دی ماه ($X_۶$)، میزان بارندگی در مهر ماه ($X_۷$) و دمای هوا در مهرماه ($X_۸$) بر میزان محصول ذرت (Y) می‌باشد. به دلیل وجود ۸ متغیر پیش‌بین، فضای مدل شامل $۲۵۶ = ۲^۸$ مدل است. روش پنجرهٔ اوکام ۳۱ مدل را بر اساس مقایسهٔ احتمال پسین آنها به عنوان بهترین مدل‌های موجود معرفی می‌کند و مدل حاصل از میانگین‌گیری وزنی روی این مدل‌ها را برای انجام استنباط و پیش‌بینی به‌کار می‌گیرد. مدل‌های منتخب روش OW به همراه احتمال پسین، $R^۲_{adj}$ و BIC برای هریک از آن‌ها در جدول ۲ ارائه شده‌اند. همانطور که ملاحظه می‌شود مقادیراحتمال پسین، $R^۲_{adj}$ و BIC برای مدل‌هایی که در پنجرهٔ اوکام قرار گرفته‌اند، کاملاً نزدیک به هم هستند. این امر انتخاب مدل مطلوب را دشوار می‌سازد. بیشترین مقدار $R^۲_{adj}$ مربوط به مدل شمارهٔ ۱۳ است، در حالی که براساس مقدار کمینهٔ BIC، مدل ۱ بهترین مدل است. انتخاب یک مدل و مبنا قرار دادن آن به این معنی است که احتمال صحیح بودن آن یک می‌باشد، در حالی که هیچ یک از مدل‌ها دارای چنین احتمال پسینی نمی‌باشند و بیشترین احتمال پسین مربوط به مدل شماره ۱ و کمتر از ۰/۲ است. از طرفی با توجه به برابر بودن احتمال پسین، $R^۲_{adj}$ و BIC برای بسیاری از مدل‌های جدول ۲، بر اساس این معیارها نمی‌توان تمایزی بین مدل‌های منتخب OW قائل شد. مطالب مطروحه از یک سو و تعداد زیاد مدل‌هایی که در پنجرهٔ اوکام قرار گرفته‌اند از سوی دیگر، حاکی از زیاد بودن عدم حتمیّت در فرآیند مدل‌سازی است و انتخاب یک مدل به عنوان مدل نهایی، چندان مطلوب به نظر نمی‌رسد. با توجه به جدول ۲، حضور متغیرهای $X_۶$ و $X_۵$

به ترتیب در ۲۵ و ۲۲ مدل از ۳۱ مدل منتخب روش OW نشان دهندهٔ تأثیرگذار بودن این دو متغیر بر متغیر پاسخ است و متغیرهای $X_۲$ و $X_۴$ از این نظر در رتبه‌های بعدی قرار دارند. براورد ضرایب این متغیرها که در جدول ۳ ارائه شده‌اند نیز به خوبی اهمیت آن‌ها را مورد تأئید قرار می‌دهد. در روش BMA احتمال تأثیر متغیر iام به‌صورت

$$Pr[\beta_i \neq ۰|D] = \sum_{\mathcal{A}_i} Pr(M_k|D)$$

محاسبه می‌شود، که در آن

$$\mathcal{A}_i = \{M_k;\ k = ۱, ۲, \ldots, T,\ \beta_i \neq ۰\}$$

مجموعه‌ی همهٔ مدل‌های شامل X_i است و آن‌را می‌توان به مقیاس بخت[۱۲] بیان و تعبیر نمود. جدول ۳ مدل منتخب روش گام به گام و مدل حاصل از روش BMA را به همراه براورد پارامترها و انحراف معیار آن‌ها، نشان می‌دهد. براساس معیار $R^۲_{adj}$ مدل حاصل از روش گام به گام نمی‌تواند بهترین مدل باشد. بعلاوه این مدل متغیرهای $X_۲$ و $X_۴$ را که در روش BMA احتمال مؤثر بودن بالایی دارند، کاملاً بی‌اثر تلقی می‌کند. یکی از نارسایی‌های روشهای کلاسیک کتمان عدم حتمیت واقعی برآوردها و کم براورد نمودن عدم حتمیت است [۱۱]. همانطور که ملاحظه می‌شود روش BMA این نارسایی را به‌صورت شایسته‌ای برطرف می‌سازد و خطای معیار براوردگرها در روش BMA که عدم حتمیت واقعی براوردها را نشان می‌دهند، به صورتی قابل ملاحظه از مقادیر مشابه در روش گام به گام بزرگ‌تر هستند. براورد ضرایب متغیرهای تقریباً نامربوط، در روش BMA کاملاً کوچک و تقریباً صفر هستند، با این وجود هیچ متغیری کاملاً نادیده گرفته نمی‌شود و از اطلاعات همهٔ متغیرها به نحوی مطلوب استفاده می‌شود. جدول ۴ کارایی پیش‌بینی روش BMA را بر اساس معیار LPC با مدل انتخابی روش گام به گام و سایر مدل‌های مقایسه می‌کند. معیار LPC برای روش BMA از حداقل مقدار ممکن آن برای سایر مدل‌ها کمتر است و این به معنی کارایی پیش‌بینی بهتر روش BMA است. کارایی این روش را می‌توان با افزایش پهنای پنجرهٔ اوکام، که منجر به میانگین‌گیری روی مدلهای بیشتر خواهد شد، بهبود بخشید. پهنای پنجرهٔ اوکام با انتخاب $C_۱$ و $C_۲$ تعیین می‌شود، که در این مثال به پیروی از مادیگان و رفتری [۹]

برای آن‌ها به ترتیب مقادیر $\frac{۱}{۲۰}$ و ۱ در نظر گرفته شده است. با توجه به اینکه افزایش پهنای پنجرهٔ اوکام، زمان تحلیل را افزایش می‌دهد، تعیین مقدار بهینه آن در سرعت محاسبات و دقت نتایج موًثر می‌باشد.

۴ بحث و نتیجه‌گیری

در مدل‌سازی انتخاب یک مدل به معنی از بین رفتن اطلاعات سایر مدل‌ها است. میانگین‌گیری

---
[۱۲] Odds Scale

بیزی مدل‌ها از اطلاعات همهٔ مدل‌ها، یا دسته‌ای از بهترین مدل‌ها استفاده می‌کند و برخلاف روش‌های کلاسیک عدم حتمیت را به خوبی منعکس می‌کند. از نقطه نظر کارایی پیش‌بین نیز نتیجهٔ حاصل از روش BMA از نتایج حاصل از هر یک از مدل‌های موجود در فضای مدل مطلوب‌تر است. برخی محققان میزان بهبود کارایی پیش‌بینی به روش BMA را معادل افزایش کارایی حاصل از ۴٪ افزایش حجم نمونه می‌دانند. عملکرد این روش با افزایش عدم حتمیت حاکم بر مدل‌سازی بهبود می‌یابد. به طور کلی استفاده از این روش در تحقیقاتی شامل نمونه‌های بزرگ و تعداد متغیرهای پیش‌بین زیاد، به‌صورت قابل توجهی بهتر است. در سایر حالاتی که عدم حتمیت اندک است، بهتر بودن روش BMA نسبت به سایر روش‌ها چشمگیر نمی‌باشد و بکارگیری روشهای مرسوم به‌دلیل سادگی مقرون به صرفه‌تر است.

۵ تشکر و قدردانی

از داوران محترم مجله به خاطر پیشنهادات اصلاحی که موجب ارائه بهتر مقاله گردید قدردانی می‌شود.

جدول ۱: میزان محصول ذرت و متغیرهای مؤثر آن.

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
۳۴/۰۰	۱۷/۷۵	۶۰/۲۰	۵/۸۳	۶۹/۰۰	۱/۴۹	۷۷/۹۰	۲/۴۲	۷۴/۴۰
۳۲/۹۰	۱۴/۷۶	۵۷/۵۰	۳/۸۳	۷۵/۰۰	۲/۷۲	۷۷/۲۰	۳/۳۰	۷۲/۶۰
۴۳/۰۰	۲۷/۹۹	۶۲/۳۰	۵/۱۷	۷۲/۰۰	۳/۱۲	۷۵/۸۰	۷/۱۰	۷۲/۲۰
۴۰/۰۰	۱۶/۷۶	۶۰/۵۰	۱/۶۴	۷۷/۸۰	۳/۴۵	۷۶/۱۰	۳/۰۱	۷۰/۵۰
۲۳/۰۰	۱۱/۳۶	۵۹/۵۰	۳/۴۹	۷۷/۲۰	۳/۸۵	۷۹/۷۰	۲/۸۴	۷۳/۴۰
۳۸/۴۰	۲۲/۷۱	۵۵/۰۰	۷/۰۰	۶۵/۹۰	۳/۳۵	۷۹/۴۰	۲/۴۲	۷۳/۶۰
۲۰/۰۰	۱۷/۹۱	۶۶/۲۰	۲/۸۵	۷۰/۱۰	۰/۵۱	۸۳/۴۰	۳/۴۸	۷۹/۲۰
۴۶/۳۰	۱۸/۵۳	۵۹/۵۰	۴/۶۷	۶۹/۲۰	۴/۲۴	۷۶/۵۰	۳/۹۹	۷۷/۸۰
۵۲/۲۰	۱۸/۵۶	۶۶/۴۰	۵/۳۲	۷۱/۴۰	۴/۵۷	۷۶/۷۰	۴/۷۲	۷۰/۷۰
۵۲/۳۰	۱۲/۴۵	۵۸/۴۰	۳/۵۶	۷۱/۳۰	۴/۵۷	۷۶/۷۰	۶/۴۴	۷۰/۷۰
۵۱/۰۰	۱۶/۰۵	۶۶/۰۰	۶/۲۰	۷۰/۰۰	۲/۲۴	۷۵/۱۰	۱/۹۴	۷۵/۱۰
۵۹/۹۰	۲۷/۱۰	۵۹/۳۰	۵/۹۳	۶۹/۷۰	۴/۸۹	۷۴/۳۰	۳/۱۷	۷۲/۲۰
۵۴/۷۰	۱۹/۰۵	۷۵/۵۰	۶/۱۶	۷۱/۶۰	۴/۵۶	۷۵/۴۰	۵/۰۷	۷۴/۰۰
۵۲/۰۰	۲۰/۷۹	۶۴/۶۰	۵/۸۸	۷۱/۷۰	۳/۷۳	۲۱/۶۰	۵/۸۸	۷۱/۸۰
۴۳/۵۰	۲۱/۸۸	۵۵/۱۰	۴/۷۰	۶۴/۱۰	۲/۹۶	۷۲/۱۰	۳/۴۳	۷۲/۵۰
۵۶/۷۰	۲۰/۰۲	۵۶/۵۰	۶/۴۱	۶۹/۸۰	۲/۴۵	۷۳/۸۰	۳/۵۶	۶۸/۹۰
۳۰/۵۰	۲۳/۱۷	۵۵/۶۰	۱۰/۳۹	۶۶/۳۰	۱/۷۲	۷۲/۸۰	۱/۴۹	۸۰/۶۰
۶۰/۵۰	۱۹/۱۵	۵۹/۲۰	۳/۴۲	۶۸/۶۰	۴/۱۴	۵۷/۰۰	۲/۵۴	۷۳/۹۰
۴۶/۱۰	۱۸/۲۸	۶۳/۵۰	۵/۵۱	۷۲/۴۰	۳/۴۷	۷۶/۲۰	۲/۳۴	۷۳/۰۰
۴۸/۲۰	۱۸/۴۵	۵۹/۸۰	۵/۷۰	۶۸/۴۰	۴/۶۵	۶۹/۷۰	۲/۳۹	۶۷/۷۰
۴۳/۱۰	۲۲/۰۰	۶۲/۲۰	۶/۱۱	۶۵/۲۰	۴/۴۵	۷۲/۱۰	۶/۲۱	۷۰/۵۰
۶۲/۲۰	۱۹/۰۵	۵۹/۶۰	۵/۴۰	۷۴/۲۰	۳/۸۴	۷۴/۷۰	۴/۷۸	۷۰/۰۰
۵۲/۹۰	۱۵/۶۷	۶۰/۰۰	۵/۳۱	۷۳/۲۰	۳/۲۸	۷۴/۶۰	۲/۳۳	۷۳/۲۰
۵۳/۹۰	۱۵/۹۲	۵۵/۶۰	۶/۳۶	۷۲/۹۰	۱/۷۹	۷۷/۴۰	۷/۱۰	۷۲/۱۰
۸۴/۴۰	۱۶/۷۵	۶۳/۶۰	۳/۰۷	۶۷/۲۰	۳/۲۹	۷۹/۸۰	۱/۷۹	۷۷/۲۰
۵۲/۸۰	۱۲/۳۴	۶۲/۴۰	۲/۵۶	۷۴/۷۰	۴/۵۱	۷۲/۷۰	۴/۴۲	۷۳/۰۰
۶۲/۱۰	۱۵/۸۲	۵۹/۰۰	۴/۸۴	۶۸/۹۰	۳/۵۴	۷۷/۹۰	۳/۷۶	۷۲/۹۰
۶۶/۰۰	۱۲/۲۴	۶۵/۵۰	۳/۸۰	۶۶/۴۰	۷/۵۵	۷۰/۵۰	۲/۵۵	۷۳/۰۰
۶۴/۲۰	۱۲/۷۲	۶۲/۸۰	۴/۱۱	۷۱/۵۰	۲/۲۹	۷۲/۳۰	۴/۹۲	۷۶/۳۰
۶۳/۲۰	۲۵/۰۸	۵۹/۷۰	۴/۴۳	۶۷/۴۰	۲/۷۶	۷۲/۶۰	۵/۳۶	۷۳/۲۰
۷۵/۴۰	۱۷/۷۹	۵۷/۴۰	۳/۳۶	۶۹/۴۰	۵/۵۱	۷۲/۶۰	۳/۰۴	۷۲/۴۰
۷۶/۰۰	۲۶/۶۱	۶۶/۶۰	۳/۱۲	۶۹/۱۰	۶/۲۷	۷۱/۶۰	۴/۳۱	۷۲/۵۰

جدول ۲: مدل‌های منتخب روش پنجرهٔ اوکام.

شماره مدل	مدل								$R^{۲}_{adj}$	BIC	احتمال پسین
۱	.	.	.	.	$X_۵$	$X_۶$	.	.	۰/۴۴۵۵	−۱۲/۴۶	۰/۱۹۱۴
۲	.	.	.	.	$X_۵$	$X_۶$	$X_۷$	.	۰/۴۷۳۲	−۱۰/۶۶	۰/۰۷۷۷
۳	$X_۱$	.	.	.	$X_۵$	$X_۶$	.	.	۰/۴۷۰۳	−۱۰/۶۶	۰/۰۷۱۰
۴	.	.	.	.	$X_۵$	.	.	.	۰/۳۳۸۰	−۱۰/۱۱۶۶	۰/۰۵۹۱
۵	.	.	.	.	.	$X_۶$	.	.	۰/۳۳۶۲	−۱۰/۰۲۶۰	۰/۰۵۶۵
۶	.	.	$X_۳$		$X_۵$	$X_۶$	.	.	۰/۴۵۷۹	−۱۰/۷۱۸۹	۰/۰۴۸۴
۷	.	.	$X_۳$	.	.	$X_۶$	.	.	۰/۳۹۵۰	−۹/۵۹۱۴	۰/۰۴۵۴
۸	.	$X_۲$	.	.	$X_۵$	$X_۶$	.	.	۰/۴۴۹۱	−۳/۱۸۸۱	۰/۰۳۷۱
۹	.	.	.	.	$X_۵$	$X_۶$		$X_۸$	۰/۴۴۷۶	−۹/۰۹۴۸	۰/۰۳۵۴
۱۰	.	.	.	$X_۴$	$X_۵$	$X_۶$	.	.	۰/۴۴۵۶	−۸/۹۷۸۵	۰/۰۳۳۴
۱۱	.	.	.	.	$X_۵$		$X_۷$	.	۰/۳۷۷۴	−۸/۶۴۲۰	۰/۰۲۸۳
۱۲	$X_۱$	.	.	.	$X_۵$		.	.	۰/۳۷۳۶	−۸/۱۴۴۱	۰/۰۲۵۶
۱۳	$X_۱$	.		.	$X_۵$	$X_۶$	$X_۷$	.	۰/۴۹۱۴	−۸/۳۲۳۱	۰/۰۲۴۱
۱۴	$X_۱$	.	$X_۳$	.	$X_۵$	$X_۶$	.	.	۰/۴۸۷۸	−۸/۰۹۱۶	۰/۰۲۴۱
۱۵	$X_۱$	.	$X_۳$	.	.	$X_۶$	.	.	۰/۴۲۷۶	−۸/۰۹۱۶	۰/۰۱۱۹
۱۶	.	.	$X_۳$	.	$X_۵$	$X_۶$	$X_۷$	.	۰/۴۸۲۶	−۷/۹۲۲۳	۰/۰۱۹۱
۱۷	.	.	.	.	.	$X_۶$	$X_۷$	.	۰/۳۶۰۱	−۷/۷۵۶۴	۰/۰۱۸۰
۱۸	$X_۱$	.	.	.	.	$X_۶$	.	.	۰/۳۵۸۷	−۷/۷۳۸۶	۰/۰۱۷۴
۱۹	.	.	.	.	.	$X_۶$		$X_۸$	۰/۳۵۵۵	−۷/۶۶۵۴	۰/۰۱۶۰
۲۰	.	$X_۲$	.	.	$X_۵$	$X_۶$	$X_۷$	.	۰/۴۷۵۷	−۷/۵۰۱۱	۰/۰۱۴۶
۲۱	.	.	.	$X_۴$	$X_۵$	$X_۶$	$X_۷$	.	۰/۴۷۵۶	−۷/۳۱۴۹	۰/۰۱۴۶
۲۲	$X_۱$	$X_۲$	.	.	$X_۵$	$X_۶$	.	.	۰/۴۷۴۷	−۷/۲۵۸۳	۰/۰۱۴۲
۲۳	.	.	.	.	$X_۵$	.	.	$X_۸$	۰/۳۵۰۳	−۷/۲۴۰۱	۰/۰۱۴۰
۲۴	.	.	.	.	$X_۵$	$X_۶$	$X_۷$	$X_۸$	۰/۴۷۳۳	−۷/۱۷۱۷	۰/۰۱۳۶
۲۵	.	.	$X_۳$	.	.	$X_۶$	$X_۷$	.	۰/۴۱۴۱	−۷/۱۵۶۴	۰/۰۱۳۴
۲۶	.	.	.	$X_۴$	$X_۵$	.	.	.	۰/۲۴۷۵	−۷/۰۹۶۱	۰/۰۱۳۱
۲۷	$X_۱$	.	.	$X_۴$	$X_۵$	$X_۶$	.	.	۰/۴۷۱۶	−۷/۰۶۲۸	۰/۰۱۲۸
۲۸	$X_۱$	.	.	.	$X_۵$	$X_۶$	.	$X_۸$	۰/۳۴۷۵	−۷/۰۹۶۱	۰/۰۱۳۱
۲۹	.	$X_۲$	.	.	.	$X_۶$	.	.	۰/۳۴۴۸	−۷/۹۵۷۹	۰/۰۱۲۲
۳۰	.	.	$X_۳$	.	.	$X_۶$	.	$X_۸$	۰/۴۰۵۵	−۶/۶۷۱۶	۰/۰۱۰۶
۳۱	$X_۱$	.	.	.	$X_۵$	.	$X_۷$	.	۰/۴۰۲۹	−۶/۵۲۸۱	۰/۰۰۹۸

جدول ۳: براورد پارامترها و احتمال مؤثر بودن در روش BMA و روش گام به گام.

روش BMA				روش گام به گام		
پارامترها	براورد پارامترها	انحراف معیار	احتمال تأثیر	برآورد پارامترها	انحراف معیار	$P-value$
$\beta_۰$	۱۶۱/۸۱۷	۸۵/۰۴۰	————	۱۶۳/۷۹۵	۵۵/۴۷	۰/۰۰۶۱
$\beta_۱$	۰/۰۲۷۵	۰/۰۷۰۶	۰/۲۲۹	————	————	————
$\beta_۲$	۰/۰۲۰	۰/۱۷۰	۰/۷۸۰	————	————	————
$\beta_۳$	−۰/۲۵۶	۰/۷۸۶	۰/۱۷۷	————	————	————
$\beta_۴$	−۰/۰۱۲	۰/۱۶۸	۰/۷۴	————	————	————
$\beta_۵$	۳/۰۹۵	۲/۱۸۳	۰/۷۹	۳/۶۳۳	۱/۴۹۴	۰/۰۲۱۲
$\beta_۶$	−۱/۶۱۳	۱/۰۰۶	۰/۸۵	−۱/۶۸۵	۰/۶۹۹	۰/۰۲۲۲
$\beta_۷$	۰/۳۳۲	۰/۸۳۷	۰/۲۳	————	————	————
$\beta_۸$	−۰/۰۳۵	۰/۲۷۵	۰/۱۰	————	————	————

جدول ۴: LPC مدل حاصل از روش BMA و سایر مدل‌ها.

مدل	LPC
روش BMA	۷۱/۶۸
روش گام به گام	۷۴/۵۶
سایر مدل‌ها	$\geq$ ۷۲/۶۶

مراجع

[1] Good, I. J. (1952), Rational decisions, *Journal of the Royal Statistical Society*. B, 14, 107-114.

[2] Hoeting, J. (1994), *Accounting for Uncertainty in Linear Regression Models*, Ph.D Disertation, Department of Statistics, University of Washington.

[3] Hoeting, J., Raftery, A. and Madigan, D. (1996), A Method for simultaneous variable selection and outlier identification in Linear Regression Models, *Journal of Statistical Computation ad Simulation*, 22, 251-271.

[4] Hoeting, J., Raftery, A. and Madigan, D., (1999), *Bayesian Simultaneous Variable Selection and Transformation selection in linear regression models*, Technical Report 9905, Dept. Statistics, Colorado State Univ. Avilable at www. Colostate. edu.

[5] Leamer, E. E. (1978), *Specification Searches*, New York: Wiley.

[6] Lipkovich, I, A. (2002), *Bayesian Model Averaging and Variable Selection in Multivariate Ecological Models*, Ph.D Disertatio, Faculty of The Virginia Polytechnic Institue and State University, Blacksburg, Virginia.

[7] Madigan, D., Gavrin, J. and Raftery, A. E. (1995), Elicting prior information to enhace yhe preformance of bayesian graphical models, *Communications in Statistics - Theory and Methods*, 24, 2271-2292.

[8] Madigan, D and Raftery, A. (1991), *Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window*, Technical Reports 21, Univ. Washington, Seattle.

[9] Madigan, D. and Raftery, A. (1994), Model selection and accounting for model uncertainty in graphical models using occam's window, *Journal of the American Statistical Association*, 89, 1535-1546.

[10] Madigan, D. and York, J. (1995), Bayesian Graphical Models for Discrete Data, *Internat. Statist. Rev.*, 63, 215- 232.

[11] Raftery, A. E. (1995), Bayesian Model Selection in Social Research(With Discussion), in *Sociological Methodology 1995(P. V. Marsden, ed.)*, 111-195. Blakwell. Cambridge, MA.

[12] Raftery, A. E., David, M. and Jenifer A. H. (1993), *Model Selection and Accounting for Model Uncertainty in Generalized Linear Models*, Technical Report, 262.

[13] Raftery, A, E., David, M. and Jenifer A. H. (1997), Bayesian model averaging in linear regressin models, *Journal of the American Statistical Association*, 97, 179-191.

[14] Robert, B. and Nobel, J. (2000), *Multivariate Applications of Bayesian Model Averaging*, Ph.D Disertation, Faculty of the Virginia Polytechnic Institue and State University.