م انگ نگ ر ز مدلها رگرس ون

افش ن فلاح محسن محمدزاده

گروه مار دانشگاه تر ت مدرس

چک ده در روشهای مدلسازی که سته ه مع ار انتخا مدل و ن ر تحل لگر مدلهای متفاوت انتخا می‌شوند عدم حتم ت در فر ند مدلسازی ناد ده گرفته می‌شود م انگ نگ ری زی مدلها ش وه‌ای توانمند در مدلسازی است که در ن ه هر مدل ه تناس م زان حما تی که از جان داده‌ها صورت می‌پذ رد وزنی اختصاص داده می‌شود و م انگ ن وزنی همه مدلها عنوان مدل نها ی کار گرفته می‌شود در ا ن مقاله نشان داده می‌شود که روش م انگ نگ ری زی خ ای مدلسازی را کاهش و کارا ی را افزا ش می‌دهد

واژه‌ها کل د عدم حتم ت توز پ ش ن م انگ نگ ری زی

مقدمه

معمولا در مدلسازی کلاسی از مدلها در ن رگرفته می‌شود سپس کی از نها ر اساس ک ا چند مع ار ارز ا ی عنوان هتر ن مدل انتخا می‌شود در حالی که ممکن است رقبای س ار خو ی رای مدل انتخا ی در ف ای مدل وجود داشته اشند ا ن ش وه مدلسازی دارای نارسا ی‌ها ی است که مهمتر ن نها در ن ر نگرفتن عدم حتم ت مدلهای انتخا ی در فر ند مدلسازی است م انگ نگ ری زی مدلها[1] BMA ش وه‌ای است که در ن از تمام مدلهای موجود در ف ای مدل رای دست ا ی ه مدلی مناس استفاده می‌شود در ا ن روش ر اساس م زان حما ت داده‌ها از هر مدل وزنی ه ن اختصاص داده می‌شود سپس مدل حاصل از م انگ ن وزنی همه مدلها رای انجام استنبا و پ ش نی ه کارگرفته می‌شود ر شه تار خی ا ن روش ه مقاله ل مر رمی‌گردد که ه دل ل محاسبات دشوار و پ چ ده در ن زمان چندان مورد توجه قرار نگرفت ا پ شرفت را انه‌ها و فنون محاسبات تقر بی در اوا ل دهه نود ا ن روش دو اره م رح و در کانون توجهات قرار گرفت ماد گان و رفتری و ماد گان و ورک دو روش پا ه‌ای رای اجرای ا ن ش وه مدلسازی ارا ه کردند هو ن گ و همکاران ه نحوه انجام محاسبات و اجرای ا ن روش پرداختند رو رت و ل پکوو چ م انگ نگ ری زی را در حالت چند متغ ره مورد توجه قرار دادند در ا ن مقاله روش م انگ نگ ری زی مدلها نحوه محاسبه مولفه‌های و روشهای اجرای ن در خش شرح داده

1) Bayesian Model Averaging

می‌شود و سپس در بخش ه ارز ابی روش BMA پرداخته و نهایتا بحث و نتیجه‌گیری در بخش ارائه خواهد شد

م انگ نگ ر ز مدلها

فرض کنید رای متغیر وابسته Y و مجموعه‌ای از متغیرهای پیش ن $X_1, \dots, X_k$ هدف یافتن بهترین مدل از ن همه مدلهای خطی

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

باشد که در ن $X_{i1}, \dots, X_{ip}$ زیرمجموعه‌ای از متغیرهای $X_1, \dots, X_k$ هستند ستیه ه حضور یا عدم حضور هر متغیر پیش ن در مدل ^{k} مدل رگرسیونی وجود دارد که فرض می‌شود همگی در فضای مدل $\mathcal{M}$ قرار دارند اگر $\mathcal{M} = \{M_1, \dots M_T\}$ نشان دهنده فضای مدل و Δ کمیتی مورد علاقه مانند مشاهده‌ای در ینده باشد در اینصورت استنباط زی ر اساس توزیع پسین Δ صورت می‌پذیرد که با فرض مشاهده مجموعه داده D بنا بر قاعده احتمال کل میخته‌ای از احتمالهای پسین همه مدلها صورت

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

است م انگین و واریانس پسین Δ ه ترتیب صورت

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

و

$$\begin{aligned} Var[\Delta|D] &= E_M[Var(\Delta|D, M)] + Var_M[E(\Delta|D, M)] \\ &= \sum_{k=1}^{T}(Var[\Delta|D, M_k] + \hat{\Delta}_k^2) \times Pr(M_k|D) - E[\Delta|D]^2 (\ ) \end{aligned}$$

هستند رفتری که در ن مولفه $Var_M[E(\Delta|D, M)]$ عدم حتمیت ن مدلها را نشان می‌دهد عبارت ک م انگین وزنی احتمال پسین است که در ن هر مدل با احتمال

پس ن متنا ر خود عنی $P(M_k|D)$ وزن‌دار شده است و توز ش ن Δ ه شر مدل M_k صورت

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

است که در ن θ_k ردار پارامترهای مدل M_k را نشان می‌دهد احتمال پس ن مدل M_k ا استفاده از قاعده ز صورت

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

دست می د که در ن $Pr(M_k)$ احتمال ش ن درست ودن مدل M_k و

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

درستنما ی جم سته[2] مدل M_k و $Pr(D|\theta_k, M_k)$ درستنما ی مدل M_k است رای محاسبه $Pr(\Delta|D)$ ا د مولفه‌های تشک ل دهنده ن را مشخص و جا گز ن نمود نا را ن لازم است توز ش ن پارامترها احتمالهای ش ن و پس ن هر مدل و توز ش ن کم ت مورد علاقه را رای همه مدلهای موجود در ف ای مدل مشخص کرد

الف تع ن احتمال ش ن پارامترها مدل کی از مسا ل دشوار در BMA تخص ص ش ن ه پارامترهای مدل است استفاده از ش ن‌های نامناس [3] منجر ه توز های پس ن نامناس می‌شود که در ا نصورت نمی‌توان ا ن احتمالها را ه عنوان احتمال مدل و نسبت نها را ه عنوان ز فاکتور تعب ر نمود ه هم ن دل ل س اری از محققان گونه‌های مختلفی از ش ن‌های گاهی خش[4] را شنهاد کرده‌اند هو نگ استفاده از ش ن‌های مناسبی[5] را شنهاد کرده است که در قسمتها ی از ف ای پارامتر ا درستنما ی زرگ هموار اشند مدل را می‌توان صورت

$$Y = X\beta + \varepsilon \qquad (\ )$$

نوشت که در ن $X_{n\times(p+1)}$ ماتر س مشاهدات Y ردار n عدی متغ رهای وا سته و ε ردار خ ا است فرض می‌شود εها رای مشاهدات مختلف مستقل و دارای توز نرمال ا م انگ ن صفر و وار انس σ^2 هستند ردار $\beta = (\beta_0, \ldots, \beta_p)$ و پارامتر σ^2 نامعلوم هستند توز عهای ش ن ا ن پارامترها ا ستی گونه‌ای تع ن شوند که عدم حتم ت نها را ه خو ی منعکس سازند هو نگ رده ش ن‌های مزدوج نرمال گاما را صورت

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

2) Integrated Likelihood 3) Improper Prior 4) Informative 5) Proper Prior

در ن رگرفت که در نها ν λ ماتر س $V_{(p+1)\times(p+1)}$ و ردار $p+$ عدی μ ا ر پارامترها ی هستند که ا د رورد شوند

تع ن احتمال پ شه ن هر مدل وقتی ه چ ا لا پ شه نی در مورد مدلها وجود ندارد ا م زان ا لاعات پ شه ن در مورد مدلها اندک است معمولا فرض می‌شود توز مدلها کنواخت است عنی $Pr(M_j)=\frac{1}{T}$ در ا نصورت احتمال پس ن مدل M_k صورت

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

خواهد ود اگر ع ی از مدلها در مقا سه ا سا ر ن دارای احتمال شتری اشند ا ا لاعات خو ی در مورد نها در دسترس اشد لازم است ا ن ا لاعات رای تعد ل احتمالهای پ شه ن مدلها کار گرفته شوند تا از پ شه ن‌های گاهی خش‌تر استفاده شود در مسا ل مر و ه انتخا متغ رهای پ‌ش ن ا لاعات پ شه ن شکل شواهد قبلی رای در ن رگرفتن ک متغ ر مورد استفاده قرار می‌گ رند فرض کن د تنها ا ن نو ازا لاعات پ شه ن در اخت ار اشد و مدل M_k توس ردار $(\delta_{k1},\ldots,\delta_{kp})$ مشخص شود که در ن

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

توا نشانگر هستند حال اگر π_i احتمال موثر ودن متغ ر X_i ر Y را نشان دهد و پذ ر م که ا لاعات پ شه ن در مورد متغ رهای متفاوت تقر با مستقل هستند می‌توان

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

را ه عنوان احتمال پ شه ن صح ح ودن مدل M_k در ن رگرفت چون ا ن توز ه متغ ری که مهمتر است احتمال زرگتری تخص ص می‌دهد توس ماد گان و رفتری توز پ شه ن متغ ر[6] نام ده شده است

ج تع ن احتمال پس ن هر مدل رای محاسبه احتمال پس ن هر مدل لازم است $Pr(D|M_k)$ از را ه که ک انتگرال ا عد را ر ا تعداد پارامترهای مدل M_k است محاسبه شود نا را ن محاسبه دق ق ا ن احتمال ه دل ل پ چ ده ودن انتگرال مر و ه تنها در حالات سه ار خاص و ساده امکان پذ ر است و در سا ر موارد از روشهای تقر بی و محاسباتی استفاده می‌شود در ا نجا از مع ار ا لا ز[7] BIC رای ا ن من ور هره می ر م مع ار BIC رای رگرس ون خ ی صورت

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

6) Variable Prior 7) Bayesian Information Criteria

تعر ف می‌شود که در ن n تعداد مشاهدات R_j^2 ر تع ن تعد ل شده مدل و k_j تعداد متغ رهای پ ش ن موجود در مدل jام و را نشان می‌دهد در ا نصورت درستنما ی جم سته مدل jام را می‌توان صورت تقر بی

$$Pr(D|M_j) \propto e^{-0/5 BIC_j} = e^{-0/5(n\log(1-r_j^2)+k_j\log n)} \qquad (\quad)$$

خواهد ود   ر ا ن اساس احتمال پس ن مدل kام ا استفاده از قاعده ز صورت

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

نوشت   چون مع ار BIC رای س اری از مدلها دارای شکل سته و ساده‌ای است استفاده از ا ن تقر   موج   سهولت محاسبه و افزا ش سرعت می‌شود

د   تع ن توز   پ ش ب ن انتگرال تشک ل دهنده توز   پ ش ن   از دو جز تشک ل شده است معمولا رای جز دوم ن از تقر

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

استفاده می شود رفتری و همکاران   که در ن $\hat{\theta}_k$ رورد حداکثر درستنما ی ردار پارامترهای مدل M_k است

اکنون ا فرض مشخص ودن همه مولفه‌های لازم رای اجرای روش BMA محاسبه مجمو

واس   ه تعداد ز اد جملات ن عملا امکان‌پذ ر ن ست و لازم است ز ر مجموعه‌ای از محتمل‌تر ن مدلها انتخا   شود   ماد گان و رفتری   روش پنجره اوکام[8] را رای ا ن من ور پ شنهاد کرده‌اند که از دو اصل کلی پ روی می‌کند   نا ر اصل اول مدلها ی که در مقا سه ا محتمل‌تر ن مدل خ لی کم شانس هستند کنارگذاشته می‌شوند   نا ر اصل دوم که ت غ اوکام[9] نام ده می‌شود مدلها ی که نسبت ه ز رمدلهای ساده‌تر خود کمتر از جان   داده‌ها حما ت می‌شوند کنارگذاشته می‌شوند   ا اجرای اصل اول مدلها ی که در مجموعه

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

قرار دارند و همچن ن نا ر اصل ت غ اوکام مدلها ی که در مجموعه

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

8) Occam's Window 9) (OW)Occam's Razor

اشند از مجمو خارج می‌شوند که در ن $C_۱$ و $C_۲$ توس تحا لگر تع ن می‌شوند در انصورت مجمو را می‌توان صورت

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

ازنو سی کرد که در ن $\mathcal{A} = \mathcal{A}' - \mathcal{B} \in \mathcal{M}$ مجموعه پذ رش است

مثال کار رد

درا ن خش کارا ی روش BMA در تشخ ص متغ رهای پ ش ن موثر در مدل و رورد پارامترهای مدل ا روش مدلسازی گام ه گام عنوان کی از روشهای مدلسازی مرسوم مورد مقا سه قرار می‌گ رد ا توجه ه ا نکه ور ـ عی کارا ی روش BMA ا افزا ش تعداد متغ رهای پ ش ن افزا ش می ا د رای نما ش هتر کارا ی روش BMA از مثالی ا تعداد متغ رهای پ ش ن کم استفاده شده تا حتی‌المکان از الا ودن کارا ی ا ن روش دل ل تعداد ز اد متغ ر پ ش ن اجتنا شود

جدول داده‌های مر و ه کارا ی شغلی پرستاران

Y															
$X_۱$ $X_۲$ $X_۳$ X X X															
Y															
$X_۱$ $X_۲$ $X_۳$ X X X															

جدول مشاهدات ک م العه رگرس ونی را نشان می‌دهد که هدف ن رسی تاث ر متغ ر مستقل شامل قا عـت $X_۱$ علاقه‌مندی $X_۲$ لند همتی $X_۳$ روا اجتماعی $X_۴$ هنرمندی درحل مشکلات $X_۵$ وا تکار $X_۶$ ر کارا ی شغلی پرستاران است دل ل وجود متغ ر پ ش ن ف ای مدل شامل $= {}^{۶}$ مدل است روش پنجره اوکام مدل را ر اساس مقا سه احتمالهای پس ن مدلها ه عنوان رتر ن مدلهای موجود مورد استفاده قرار می‌دهد مدلهای حاصل ه همراه احتمال پس ن مقدار $R^۲_{adj}$ ه درصد و مع ارا لا ز نها در جدول ارا ه شده‌اند تعداد ز اد مدلها ی که در پنجره اوکام قرار گرفته‌اند ش از $\frac{۱}{۴}$ مدلها نشان

جدول ‌ ‌ مدلهای انتخا ی توسـ ‌ روش پنجره اوکام

شماره مدل	مدل	احتمال پسـ ن	R^2_{adj}	BIC
	$X_3 \quad X_6$	———		———
	$X_2\ X_3 \quad X_6$			
	$X_1\ X_2\ X_3 \quad X_6$			
	$X_1 \quad X_3 \quad X_6$			
	X_3			
	$X_3 \quad X_5\ X_6$			
	$X_2\ X_3 \quad X_5\ X_6$			
	$X_3\ X_4 \quad X_6$			
	$X_2\ X_3$			
	$X_1 \quad X_3$			
	$X_2\ X_3\ X_4 \quad X_6$			
	$X_1\ X_2\ X_3$			
	$X_1\ X_2\ X_3 \quad X_5\ X_6$		———	
	$X_1\ X_2\ X_3\ X_4 \quad X_6$			
	$X_1 \quad X_3 \quad X_5\ X_6$			
	$X_1 \quad X_3\ X_4 \quad X_6$			
	$X_3 \quad X_5$			
	$X_3\ X_4$			

دهنده ز اد ودن عدم حتم ت است معنی ش از ک مدل صح ح وجود دارد نا را ن انتخا ک مدل ه عنوان مدل نها ی منـ قی نـ ر نمی‌رسد ملاحـ ه می‌شود که مقاد ر R^2_{adj} و مع ار ا لا ز BIC رای مدلها ی که در پنجره اوکام قرار گرفته‌اند نزد ک ه کد گر هستند شتر ن مقدار R^2_{adj} مر و ه مدل شماره است در حالی که ر اساس مقدار کمـ نه مع ار ا لا ز مدل هتر ن است انتخا ک مدل و مبنا قرار دادن ن ه معنی ن است که احتمال صح ح ودن ن ک می اشد در حالی که شتر ن احتمال پسـ ن مر و ه مدل شماره و حدود است و احتمال پسـ ن سا ر مدلها همگی از کمتر است جدول نشان می‌دهد که متغ ر X_3 در همه مدلها و متغ ر X_6 در مدل منتخ روش پنجره اوکام حـ ور دارند که نشان دهنده تاثـ رگذار ودن ا ن دو متغ ر است ـ رورد را متغ رهای مستقل که در جدول ارا ه شده نـ ز اهم ت دو متغ ر X_3 و X_6 را تا د می‌کند هما نـ ورکه در جدول ملاحـ ه می‌شود ـ روردهای ز را ا ن دو متغ ر که ا استفاده از روش BMA دست مده‌اند ه خو ی اهم ت ا ن دو متغ ر را منعکس می‌سازند جدول ـ رورد پارامترها و خـ ای مع ار نها را که ه روش گام ه گام دست مده‌اند را نشان می دهد هما نـ ورکه ملاحـ ه می شود خـ ای مع ار متنا ر ا ـ روردگرهای ا ن دو پارامتر در روش BMA صورتی قا ل ملاحـ ه از مقاد ر مشا ه در روش گام ه گام زرگتر هستند ا ن کی از رتر های روش BMA است که م زان واقعی عدم حتم ت را نشان می‌دهد م ا د گری که می‌توان ه ن اشاره کرد ا ن است

جدول روورد را در روش BMA

پارامتر رگرسیونی	$Pr[\beta_i \neq \quad \mid D]$	میانگین پسین	انحراف معیار پسین
β_1			
β_2			
β_3			
β_4			
β_5			
β_6			

جدول روورد را در مدل کامل $R^2_{adj} = \ /$ و روش گام به گام

مدل کامل	روورد پارامترها	خطای معیار	روش گام به گام	روورد پارامترها	خطای معیار
β_0			*		
β_1					
β_2					
β_3			*		
β_4					
β_5					
β_6			*		

که بخت حضور متغیر X_2 در مدل بیش از است که بر اساس قواعد سرانگشتی جفریز شواهد مثبتی برای حضور متغیر X_2 در مدل ارائه می‌کند در حالی که اطلاعات موجود در این متغیر با حذف آن در مدل حاصل از روش گام به گام نادیده گرفته می‌شوند روورد را متغیرهای در روش BMA در مقایسه با سایر روشها و خصوصا در مورد متغیرهای X_1 و X_4 کاملا کوچک و نزدیک صفر هستند با این وجود کاملا نادیده گرفته نمی‌شوند و از اطلاعات آنها استفاده می‌شود

بحث و نتیجه‌گیری

در هنگام مدلسازی غالبا نمی‌توان مدلی یافت که بصورت کامل به داده‌ها برازش داشته باشد در این حالات انتخاب یک مدل به معنی از بین رفتن اطلاعات سایر مدلها است میانگین‌گیری بیزی مدلها از اطلاعات همه مدلها یا دسته‌ای از بهترین مدلها استفاده می‌کند و برخلاف روشهای مرسوم عدم حتمیت را به خوبی منعکس می‌کند از نقطه نظر کارایی پیشبینی نیز نتیجه حاصل از روش BMA از نتایج حاصل از هر یک از مدلهای موجود در فضای مدل مطلوبتر است

عملکرد ا ن روش با افزا ش عدم حتم ت بهبود می ا د مثلا هر چه تعداد متغ رهای پ ش ن در مدل رگرس ونی شتر اشد کارا ی ا ن روش شتر می‌شود در د تر ن حالت که عدم حتم ت اندک اشد هتر ودن روش BMA نسبت ه سا ر روشها چشمگ ر نمی اشد و کارگ ری روشهای مرسوم د ل سادگی مقرون ه صرفه‌تر است

مراج

[1] Good, I. J., (1950). "Probability and The Weighing of Evidence", Griffin London.

[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". J. Comput, Statist. **22**, 251-271.

[4] Hoeting, J., Raftery, A. and Madigan, D., (1999). "Bayesian Simultaneous Variable Selection and Transformation Selection in Linear Regression Models". Tecnical Report 9905, Dept. Statistics, Colorado State Univ. Avalable at www. Colostate. edu.

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

[6] Madigan, D., Gavrin, J. and Raftery, A. E., (1995). "Elicting Prior Information To Enhace The Preformance of Bayesian Graphical Models", Comm. Statist. Thory Methods, **24**, 2271-2292.

[7] Madigan, D and Raftery, A., (1991). "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window", Technical Reports 213, Univ. Washington, Seattle.

[8] Madigan, D. and Raftery, A., (1994). "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window", Journal of American Statistical Association, **89**, 1535-1546.

[9] Miller A., (1990). "Subset Selection in Regression Variables", New York, Chpman-hal.

[10] 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.

[11] Raftery, A. E., (1996). "Approximate Bayes Factor and Accounting for Model Uncertainty in Generalized Linear Models", Technical Report, Biometrika, **83**, 351-266.

[12] Robert, B. Nobel, J. (2000). " Multivariate applications of Bayesian Model averaging", Ph.D disertation, Faculty of the virginia polytechnic institue and state university, Blacksburg, virginia.