تصحیح روش یادگیری آمیخته تقویت‌شده با استفاده از آزمون وونگ و کاربرد آن در مدل آمیخته گاما

صدیقه زمانی مهریان [۱]

تاریخ دریافت: ۱۴۰۱/۰۸/۱۲
تاریخ پذیرش: ۱۴۰۲/۰۱/۱۸

چکیده:
روش یادگیری آمیخته تقویت‌شده (BML)، روشي فزاینده برای یادگیری مدل‌های آمیخته در مسئله طبقه‌بندی است. در هر مرحله از روش یادگیری آمیخته تقویت‌شده، مؤلفه جدیدي با توجه به یک تابع هدف در جهت به حداکثر رساندن تابع هدف به مدل آمیخته اضافه مي‌شود. از جمله توابع هدف مورداستفاده در این روش، تابع درستنمایي و به‌طور معادل معیارهای اطلاع هستند. در این روش مؤلفه جدیدی به مدل آمیخته اضافه می‌شود که باعث بیشترین افزایش تابع درستنمایی شود. چون تابع درستنمایی و معیارهاي اطلاع توانایي تشخیص مدل‌هاي معادل را ندارد، بنابراین ممکن است مدل آمیخته جدید و مدل آمیخته فعلی معادل باشند و اضافه کردن مؤلفه جدید به مدل آمیخته فعلی باعث بهبود مدل نشود. در این مقاله روش یادگیری آمیخته تقویت‌شده با استفاده از آزمون انتخاب مدل وونگ که توانایي تشخیص مدل‌هاي معادل را دارد، تصحیح‌شده است. همچنین عملکرد دو روش یادگیری با استفاده از داده‌های شبیه‌سازی و مجموعه داده‌های واردات کالای ایالات‌متحده توسط گمرک ارزیابی‌شده است.

واژه‌های کلیدی: آزمون وونگ، انتخاب مدل، تابع درستنمایي ماکسیمم، روش یادگیری آمیخته تقویت‌شده.

۱ مقدمه

مفهوم تقویت کردن به‌طورگسترده در مسائل طبقه‌بندي الگوهاي مختلف در یادگیری ماشین استفاده می‌شود. روش تقویت کردن روش کلی برای ترکیب طبقه‌بندی کننده‌های چندگانه است که تقریباً برای هر نوع الگوریتم یادگیری به‌منظور بهبود دقت طبقه‌بندی کننده‌ها استفاده می‌شود [۱۰]. ایده اصلی روش تقویت کردن این است که به‌طور متوالی مجموعه‌ای از طبقه‌بندی کننده‌های ضعیف را آموزش داده و باهم ترکیب مي‌کنیم تا یک طبقه‌بندی قوی و قابل‌اعتماد ساخته شود [۱۱]. به‌عنوان‌مثال، در الگوریتم معروف تقویت‌کننده تطبیق‌پذیر [۲]، توزیع احتمالي برای همه نمونه‌های آموزشی در فضای ورودی معرفی می‌شود. در ابتدا، به تمام نمونه‌های آموزشي وزن‌های مساوی اختصاص داده می‌شوند، سپس در تکرارهای آموزشی، وزن آن نمونه‌هایی که طبقه‌بندی آن‌ها سخت‌تر و پیچیده‌تر هستند به‌تدریج افزایش می‌یابند و به‌طور خودکار این وزن‌ها در تمامي مجموعه آموزشی بر اساس نتایج طبقه‌بندی به‌روزرسانی می‌شوند. با استفاده از وزن‌های آموزشي به‌روزرسانی شده، طبقه‌بندی کننده جدید بر روی این نمونه‌ها آموزش داده می‌شود. در مرحله بعد، طبقه‌بندی کننده جدید با طبقه‌بندی کننده فعلي ترکیب می‌شود تا یک طبقه‌بندی پیشرفته بر اساس نظریه تقویت کردن به دست آید که به‌صورت تحلیلی، کران بالای نرخ خطای آموزشی را به حداقل می‌رساند. این فرآیند برای تولید مجموعه‌ای از طبقه‌بندی کننده‌های پایه تکرار می‌شود. الگوریتم تقویت کردن به‌طور موفقیت‌آمیزی برای انواع مختلفی از طبقه‌بندی‌ها استفاده می‌شود که می‌تواند دقت را نسبت به طبقه‌بندی کننده‌های پایه بهبود بخشد [۶].

کاربرد الگوریتم تقویت‌کننده برای برآورد تابع چگالی احتمال به‌طور گسترده موردمطالعه قرارگرفته است [۱۳]. اخیراً، الگوریتم‌های تقویت‌کننده در برخی از مسائل یادگیری مدل‌های آمیخته گسترش‌یافته است [۹] که الگوریتم یادگیری آمیخته تقویت‌شده (BML) نامیده می‌شود. ایده اصلی الگوریتم یادگیری آمیخته تقویت‌شده، یادگیری مدل‌های آمیخته به‌طور فزاینده و بازگشتی است. الگوریتم یادگیری آمیخته تقویت‌شده از یک مدل آمیخته شروع و به‌تدریج یک مؤلفه آمیخته جدید را به‌گونه‌ای اضافه می‌کند که همواره تابع هدف از پیش تعریف‌شده را بهینه کند. روش‌های مشابه دیگري برای یادگیری

[۱] گروه آمار، دانشکده علوم پایه، دانشگاه بین‌المللی امام خمینی (ره)، قزوین، ایران (نویسنده مسئول: zamani@sci.ikiu.ac.ir)

مدل‌های آمیخته وجود دارد [۱۴، ۵]، اما نکته اساسی روش یادگیری آمیخته تقویت‌شده این است که مؤلفه آمیخته جدید در هر مرحله با توجه به تابع هدف برآورد می‌شود، به‌طوری‌که هر مؤلفه جدید همیشه وقتي اضافه می‌شود که تابع هدف را بیش از همه افزایش می‌دهد. نیا و همکاران [۸] مدل تقویت گرادیان با مدل آمیخته گاوسی را معرفی کردند و به پیش‌بینی جریان ماهانه با استفاده از این مدل پرداختند.

دوو و همکاران [۱] تابع لگاریتم درستنمايي مدل آمیخته را به‌عنوان تابع هدف در نظر گرفتند و در هر تکرار مؤلفه جديدي را به مدل آمیخته اضافه مي‌کردند به‌طوری‌که تابع لگاریتم درستنمايي مدل آمیخته جدید نسبت به تابع لگاریتم درستنمايي مدل آمیخته فعلي افزایش یابد. در بعضي از موارد ممکن است این افزایش ناچیز باشد و مدل آمیخته جدید و مدل آمیخته فعلي معادل باشند. به‌عبارت‌دیگر روش یادگیری آمیخته تقویت‌شده توانايي تشخیص مدل‌هاي معادل را ندارد. در این مقاله، روش یادگیری آمیخته تقویت‌شده بر اساس آزمون انتخاب مدل وونگ تصحیح‌شده است. اشنایدر و همکاران [۱۲] برای مقایسه مدل‌های تک متغیره و چندمتغیره آشیانی از آزمون وونگ استفاده کردند و با استفاده از مطالعات شبیه‌سازی و داده‌های واقعی نشان دادند وقتی‌که مدل‌ها آشیانی هستند این آزمون معمولاً به‌خوبی یا گاهی بهتر از آزمون نسبت درستنمایی عمل می‌کند.

مقاله به شرح زیر سازمان‌دهی شده است: در بخش ۲، به بررسی الگوریتم یادگیری آمیخته تقویت‌شده برای مدل‌های آمیخته پرداخته‌شده است. در بخش ۳، الگوریتم یادگیری آمیخته تقویت‌شده برای مدل‌های آمیخته بر اساس آزمون انتخاب مدل تصحیح‌شده است و مثالی از مدل آمیخته گاما ارائه‌شده است. در بخش ۴، عملکرد دو روش یادگیری آمیخته تقویتی تصحیح‌شده و یادگیری آمیخته تقویت‌شده با استفاده از مطالعات شبیه‌سازی مقایسه شده است. آنالیز داده‌هاي واقعی واردات کالاهای ایالات‌متحده توسط گمرک در بخش ۵ ارائه‌شده است.

۲ مروری بر الگوریتم یادگیری آمیخته تقویت‌شده

در این بخش، به‌مرور الگوریتم یادگیری آمیخته تقویت‌شده (BML) پرداخته که کاربرد فراوانی در بسیاری از مسائل طبقه‌بندی الگوها دارد. مک لاکلان [۶]، مدل آمیخته $F_K(x)$،

$$F_K(x) = \sum_{k=1}^{K} c_k f_k(x), c_k \geq 0, \sum_{k=1}^{K} c_k = 1$$

را معرفی کرد که در آن K تعداد مدل‌های آمیخته، x بردار ویژگی‌ها، c_k و $f_k(x)$ به ترتیب نشان‌دهنده وزن‌ها و توزیع مؤلفه $k-$ام مدل آمیخته است.

یادگیری مدل‌های آمیخته به‌طور گسترده در یادگیری ماشین موردمطالعه قرارگرفته است. پاولوویچ [۹] و کیم و پاولوویچ [۳] روش یادگیری آمیخته تقویت‌شده را برای یادگیری مدل‌های آمیخته معرفی کردند. در این روش مؤلفه جدید $f_k(x)$ با وزن c_k به مدل آمیخته $F_{k-1}(x)$ با $k-1$ مؤلفه آمیخته اضافه می‌شود و به مدل آمیخته جدید $F_k(x)$،

$$F_k(x) = (1 - c_k)F_{k-1}(x) + c_k f_k(x)$$

با k مؤلفه آمیخته تبدیل می‌شود که در آن c_k نشان‌دهنده وزنی برای ترکیب مؤلفه آمیخته جدید با مدل آمیخته فعلی $F_{k-1}(x)$ است. این روش تا زمانی تکرار می‌شود که برخی از شرایط همگرایی ایجاد شود. هر مؤلفه آمیخته جدید $f_k(x)$ و وزن ترکیب آن c_k را می‌توان بر اساس تابع هدفی مانند $Q(F_k)$ که از پیش تعریف‌شده، تعیین کرد. الگوریتم یادگیری آمیخته تقویت‌شده (BML) در جدول ۱ توضیح داده‌شده است.

جدول ۱ توصیف الگوریتم یادگیری آمیخته تقویت‌شده

گام اول: برای $k = 1$ مدل اولیه $F_1(x_t)$ را انتخاب کنید.
گام دوم: برای $k = 2, 3, \cdots$ قرار دهید:
$$\{c_k^*, f_k^*\} = argmax_{c_k, f_k} Q(F_k)$$
اگر مؤلفه‌های بهینه c_k^*, f_k^* یافت شدند مدل آمیخته
$$F_k(x_t) = (1 - c_k^*)F_{k-1}(x_t) + c_k^* f_k^*(x_t)$$
را به‌عنوان مدل آمیخته جدید در نظر گرفته و گام دوم را تکرار کنید، در غیر این صورت به گام سوم بروید.
گام سوم: خروجی مدل آمیخته$F_k(x_t)$ است.

با در نظر گرفتن روش درستنمایی ماکسیمم (MLE) به‌عنوان روشی برای برآورد پارامترهای مدل آمیخته، تابع هدف را می‌توان تابعی از لگاریتم تابع درستنمایی مدل آمیخته $F_k(x)$ تعریف کرد. به‌عبارت‌دیگر $Q(F_k) = \sum_{t=1}^{T} \log F_k(x_t)$ که در آن T تعداد نمونه‌های یادگیری $\{x_1, x_2, \cdots, x_T\}$ است.

در گام دوم، مؤلفه جدید $f_k(x)$ با وزن c_k به مدل آمیخته $F_{k-1}(x)$ اضافه می‌شود که ماکسیمم‌کننده تابع هدف $Q(F_k)$ باشد. به‌عبارت‌دیگر

$$Q\left((1 - \epsilon) F_{k-1} + \epsilon f_k\right) > Q(F_{k-1}) \qquad (1)$$

که در آن ϵ نشان‌دهنده انحراف کوچک ثابت است. با استفاده از بسط تیلور مرتبه اول، می‌توان بسط جمله سمت چپ معادله ۱ را به‌صورت

$$\begin{aligned} Q\left((1-\epsilon)F_{k-1}+\epsilon f_k\right) &= Q(F_{k-1}+\epsilon(f_k-F_{k-1})) \\ &= Q(F_{k-1})+\epsilon\left\langle \bigtriangledown Q(F_{k-1}),(f_k-F_{k-1})\right\rangle \\ &+ O(\epsilon(f_k-F_{k-1})) \\ &\approx Q(F_{k-1})+\epsilon\left\langle \bigtriangledown Q(F_{k-1}),(f_k-F_(k-1))\right\rangle \end{aligned} \quad (2)$$

نوشت که در آن

$$\bigtriangledown Q(F_{k-1}) = \bigtriangledown Q(F)|_{F=F_{k-1}} = \frac{1}{F_{k-1}(x_t)}$$

و $\langle P_1, P_2\rangle$ نشان‌دهنده ضرب داخلی تابعی برای دو مدل آمیخته $P_1(x_t)$ و $P_2(x_t)$ است. به‌عبارت‌دیگر

$$\langle P_1, P_2\rangle = \frac{1}{T}\sum_{t=1}^{T} P_1(x_t)P_2(x_t).$$

اگر ϵ به‌اندازه کافی کوچک باشد، می‌توان همه جملات با مرتبه بالا را نادیده گرفت؛ بنابراین، تابع هدف بر اساس مدل آمیخته جدید را می‌توان با عبارات خطی ۲ تقریب زد، به دوو و همکاران [۱] مراجعه کنید.

مؤلفه جدید f_k باید به‌گونه‌ای برآورد شود که تابع هدف

$$Q((1-\epsilon)F_{k-1}+\epsilon f_k)$$

را به‌سرعت افزایش دهد. با اعمال تقریب خطی ۲ در معادله ۱، انتخاب مؤلفه جدید به مفهوم انتخاب f_kای است که ماکسیمم کننده $\langle \bigtriangledown Q(F_{k-1}),(f_k-F_{k-1})\rangle$ باشد. به‌عبارت‌دیگر

$$\begin{aligned} f_k^* &= argmax_{f_k}\left\langle \bigtriangledown Q(F_{k-1}),(f_k-F_{k-1})\right\rangle \\ &= argmax_{f_k}\frac{1}{T}\sum_{t=1}^{T}\frac{(f_k(x_t)-F_{k-1}(x_t))}{F_{k-1}(x_t)} \\ &= argmax_{f_k}\frac{1}{T}\sum_{t=1}^{T}\frac{f_k(x_t)}{F_{k-1}(x_t)}. \end{aligned} \quad (3)$$

در معادله ۳ انتخاب مؤلفه‌ی جدید $f_k(x_t)$ بر اساس روش درستنمایی ماکسیمم (ML) است. به‌شرط برآورد $f_k(x)$، وزن c_k^* به‌صورت

$$c_k^* = argmax_{c_k\in[0,1]}Q((1-c_k)F_{k-1}+c_kf_k^*) \quad (4)$$

قابل‌محاسبه است. در عمل، وزن بهینه c_k^* را می‌توان با استفاده از جستجوی شبکه‌ای در بازه‌ی $[0,1]$ پیدا کرد.

دوو و همکاران [۱] براي تعيين $f_k^*(x), k$ و وزن بهینه c_k^*، نامساوي $Q((1-\epsilon)F_{k-1}+\epsilon f_k) > Q(F_{k-1})$ را معرفي كردند كه در آن تابع هدف، تابع لگاریتم مدل آمیخته است. به‌عبارت‌دیگر تابع $f_k^*(x)$ با وزن c_k^* به مدل $F_{k-1}(x)$ اضافه مي‌شود اگر $Q(F_k) > Q(F_{k-1})$ باشد به‌طوری‌که $F_k(x_t) = (1-c_k^*)F_{k-1}(x_t) + c_k^*f_k^*(x_t)$ است. توجه کنید که افزایش اندک مقدار $Q(F_k)$ نسبت به $Q(F_{k-1})$ باعث انتخاب مدل آمیخته $F_k(x_t)$ به‌عنوان مدل بهینه مي‌شود، اما ممکن است که دو مدل $F_k(x_t)$ و $F_{k-1}(x_t)$ معادل باشند. به‌عبارت‌دیگر نامساوي معرفی‌شده توسط دوو و همکاران [۱] توانايي تشخيص مدل‌هاي معادل را ندارد. اگر دو مدل $F_k(x_t)$ و$F_{k-1}(x_t)$ معادل باشند مدل $F_{k-1}(x_t)$ به‌عنوان مدل بهینه انتخاب مي‌شود، درصورتي‌كه بر اساس نامساوي $Q((1-\epsilon)F_{k-1}+\epsilon f_k) > Q(F_{k-1})$ مؤلفه $f_k^*(x)$ با وزن c_k^* به مدل آمیخته $F_{k-1}(x_t)$ اضافه مي‌شود. در این مقاله به تصحیح الگوریتم معرفی‌شده توسط دوو و همکاران [۱] بر اساس آزمون انتخاب مدل که توانايي تشخيص مدل‌هاي معادل را دارند، پرداخته‌شده است.

۳ تصحیح الگوریتم یادگیری آمیخته تقویت‌شده

معمولاً توزيع درست پديده‌هاي تحت مطالعه معلوم نيست. به‌عبارت‌ديگر نمونه تصادفي $X_1,\cdots,X_T$ از توزيع مجهول $h(x_t)$ پيروي مي‌كند. براي انجام هر استنباط آماري نياز به توزيع درست جامعه داريم. از ديدگاه رياضي $h(x_t)$ تركيبي از منحني‌ها است كه نقاط مشاهده‌شده را به هم وصل مي‌كند. به دليل پيچيدگي اين نوع مدل‌ها، فرض مي‌شود كه نمونه تصادفي از يك توزيع (مدل) پارامتري پيروي مي‌كند. مطالعه و بررسي مدل يا ساختار داده‌هاي مشاهده‌شده، انتخاب مدل ناميده مي‌شود. براي اين منظور خانواده‌اي از مدل‌ها مانند $F = \{f^{\theta_i}(x_t), i = 1,2,\cdots,k\}$ در نظر گرفته مي‌شود و از بين مدل‌هاي پيشنهادي مدل بهينه تحت تابع هدف خاصی انتخاب مي‌شود.

بردار تصادفی$X(T\times 1)$ و فرضیه‌های

$$H_1 : X \sim f^{\theta_k}(x_t)$$

و

$$H_2 : X \sim f^{\theta_{k-1}}(x_t)$$

را در نظر بگیرید که در آن $f^{\theta_k}(x_t)$ مدل آمیخته $F_k(x_t)$ و $f^{\theta_{k-1}}(x_t)$ مدل آمیخته $F_{k-1}(x_t)$ هستند. توجه کنید که دو مدل پیشنهادی (رقیب) $F_k(x_t)$ و $F_{k-1}(x_t)$ مدل‌های آشیانی هستند. به این مفهوم که

مدل $F_{k-1}(x_t)$ با ایجاد محدودیت‌های پارامتری بر روی مدل $F_k(x_t)$ ایجاد می‌شود. همچنین فرض کنید $l(\hat{\theta}_i)$ تابع لگاریتم درستنمایی مدل ارائه‌شده تحت فرضیه‌های $H_i, i = 1, 2$ باشد. هدف از انتخاب مدل، انتخاب مدل آمیخته $F_k(x_t)$ یا مدل آمیخته $F_{k-1}(x_t)$ یا هر دو مدل به‌عنوان مدل‌های معادل است. با در نظر گرفتن دو مدل آمیخته $F_k(x_t)$ و $F_{k-1}(x_t)$ به‌عنوان دو مدل پیشنهادی (رقیب) آشیانی، مدلی به‌عنوان مدل بهینه انتخاب می‌شود که به مدل درست نزدیک‌تر باشد. به‌عبارت‌دیگر مدلی را انتخاب می‌کنیم که حداقل ریسک کولبک - لیبلر [۴] را داشته باشد. وونگ [۱۵] آزمون فرضیه

$$H_0 : E_h\left(\log \frac{f^{\theta_k}(x_t)}{f^{\theta_{k-1}}(x_t)}\right) = 0$$

در مقابل

$$H_f : E_h\left(log \frac{f^{\theta_k}(x_t)}{f^{\theta_{k-1}}(x_t)}\right) > 0$$

یا

$$H_g : E_h\left(log \frac{f^{\theta_k}(x_t)}{f^{\theta_{k-1}}(x_t)}\right) < 0$$

را براي انتخاب مدل در بين مدل‌هاي آشياني پيشنهادي بد-توصيف شده، معرفي كرد كه مبتني بر آماره نسبت درستنمايي است. پذيرش H_0 به مفهوم معادل بودن دو مدل $f^{\theta_k}(x_t)$ و $f^{\theta_{k-1}}(x_t)$ است. پذيرش H_f به اين مفهوم است كه مدل $f^{\theta_k}(x_t)$ براي برازش بر روي داده‌ها مناسب‌تر از $f^{\theta_{k-1}}(x_t)$ است، همچنين پذيرش H_g به اين مفهوم است كه مدل $f^{\theta_{k-1}}(x_t)$ براي برازش بر روي داده‌ها مناسب‌تر از $f^{\theta_k}(x_t)$ است. اين آزمون، زماني كه مدل خوب- توصيف‌شده باشد تبديل به آزمون نسبت درستنمايي نيمن -پيرسون مي‌شود. تعريف مي‌كنيم

$$\begin{aligned} LR(\hat{\theta}_k, \hat{\theta}_{k-1}) &= l(\hat{\theta}_k) - l(\hat{\theta}_{k-1}) \\ &= \sum_{t=1}^{T} \log f^{\hat{\theta}_k}(x_t) - \sum_{t=1}^{T} \log f^{\hat{\theta}_{k-1}}(x_t) \\ &= \sum_{t=1}^{T} \log\left(\frac{f^{\hat{\theta}_k}(x_t)}{f^{\hat{\theta}_{k-1}}(x_t)}\right). \end{aligned}$$

وونگ [۱۵] نشان داد كه تحت فرضيات $A_1 - A_6$ (ارائه‌شده در وونگ [۱۵]) و تحت فرضيه H_0 آماره $2LR(\hat{\theta}_k, \hat{\theta}_{k-1})$ داراي توزيع حدي مجموع وزني توزيع كاي اسكور است. اگر مدل‌هاي رقيب خوب- توصيف‌شده باشند اين توزيع حدي تبديل به توزيع كاي اسكور مي‌شود. به‌عبارت‌ديگر

$$2LR(\hat{\theta}_k, \hat{\theta}_{k-1}) \longrightarrow^d \chi^2(1).$$

با انتخاب سطح معنی‌داری α، اگر

$$\chi^2_{\frac{\alpha}{2}}(1) < 2LR(\hat{\theta}_k, \hat{\theta}_{k-1}) < \chi^2_{1-\frac{\alpha}{2}}(1)$$

باشد فرضيه معادل بودن دو مدل را مي‌پذيريم. با توجه به اين‌كه مدل‌هاي پيشنهادي آشياني هستند، بنابراين پذيرش فرضيه معادل بودن دو مدل منتج به پذيرش مدل $f^{\theta_{k-1}}(x_t)$ مي‌شود. اگر مقدار آماره $2LR(\hat{\theta}_k, \hat{\theta}_{k-1})$ از مقدار بحراني $\chi^2_{1-\frac{\alpha}{2}}(1)$ بيشتر باشد، فرضيه صفر را رد مي‌كنيم به‌طوري‌كه مدل $f^{\theta_k}(x_t)$ براي برازش بر روي داده‌ها مناسب‌تر از $f^{\theta_{k-1}}(x_t)$ است. اگر مقدار آماره $2LR(\hat{\theta}_k, \hat{\theta}_{k-1})$ از مقدار بحراني $\chi^2_{\frac{\alpha}{2}}(1)$ كم‌تر باشد، مدل $f^{\theta_{k-1}}(x_t)$ براي برازش بر روي داده‌ها مناسب‌تر از $f^{\theta_k}(x_t)$ است. با توجه به نواحي رد و پذيرش مي‌توان نتيجه گرفت وقتي‌كه $2LR(\hat{\theta}_k, \hat{\theta}_{k-1}) > \chi^2_{1-\frac{\alpha}{2}}(1)$ باشد مي‌توان مدل $f^{\theta_k}(x_t)$ را براي برازش بر روي داده‌ها انتخاب كرد. به‌عبارت‌ديگر مي‌توان مؤلفه جديد $f_k(x)$ با وزن c_k به مدل آميخته $F_{k-1}(x)$ اضافه كرد؛ بنابراين اگر تابع لگاريتم درستنمايي مدل آميخته به‌عنوان تابع هدف در نظر گرفته شود، هنگامی مؤلفه جديد $f_k(x)$ به مدل آميخته فعلی اضافه می‌شود كه

$$Q((1-\epsilon)F_{k-1} + \epsilon f_k) - Q(F_{k-1}) > \frac{1}{2}\chi^2_{1-\frac{\alpha}{2}}(1) \qquad (5)$$

الگوريتم يادگيری آميخته تقويتی تصحيح‌شده $(CBML)$ در جدول ۲ توضيح داده‌شده است.

جدول ۲. توصيف الگوريتم يادگيری آميخته تقويتی تصحيح‌شده

گام اول: برای $k = 1$ مدل اوليه $F_1(x_t)$ را بر اساس آزمون وونگ انتخاب كنيد.

گام دوم: برای $k = 2, 3, \cdots$ قرار دهيد:

$$\{c_k^*, f_k^*\} = argmax_{c_k, f_k} Q(F_k)$$

اگر مؤلفه‌های بهينه c_k^*, f_k^* يافت شدند مدل آميخته

$$F_k(x_t) = (1 - c_k^*)F_{k-1}(x_t) + c_k^* f_k^*(x_t)$$

را به‌عنوان مدل آميخته جديد در نظر گرفته و اگر

$$Q((1-\epsilon)F_{k-1} + \epsilon f_k) - Q(F_{k-1}) > \frac{1}{2}\chi^2_{1-\frac{\alpha}{2}}(1)$$

باشد گام دوم را تكرار كنيد، در غير اين صورت به گام سوم برويد.

گام سوم: خروجی مدل آميخته$F_k(x_t)$ است.

در گام دوم روش يادگيری آميخته تقويتی تصحيح‌شده، براي محاسبه مؤلفه‌های بهينه c_k^* و f_k^* بايد تابع لگاريتم درستنمايي مدل آميخته را ماكسيم كرد. چون برآوردگر ماكسيم درستنمايي پارامترهاي مدل آميخته جواب صريحي ندارد، بنابراين روش‌هاي ديگر از جمله الگوريتم

EM را می‌توان براي برآورد پارامترهاي مدل استفاده کرد. براي اين منظور در ادامه مدل آمیخته گاما به‌عنوان مثالی از مدل‌هاي آمیخته بیان می‌شود و پارامترهاي مجهول مدل با استفاده از الگوریتم EM برآورد می‌شوند.

مدل آمیخته گاما [۱۶] را در نظر بگیرید که هر مؤلفه آن از توزیع گاما با تابع چگالی

$$f_k^{\Phi_k}(x_t) = \frac{1}{B^{a_k}\gamma(a_k)} x_t^{a_k} e^{-\frac{x_t}{B_k}}$$

پیروی می‌کند که در آن $\Gamma(.)$ تابع گاما است و $\Phi_k = (a_k, B_k)$ نشان‌دهنده پارامترهای $k-$اُمین توزیع گاما است. همچنین پارامترهای مدل آمیخته $F_k(x_t)$ را با نماد $\theta_k = (\Phi_k, \theta_{k-1})$ نمایش می‌دهیم. با توجه به معادلات ۳ و ۴ داریم

$$\{f_k^*, c_k^*\} = argmax_{c_k, f_k} Q((1-c_k)F_{k-1} + c_k f_k).$$

چون برآوردگر ماکسیمم درستنمایي پارامترهاي مدل آمیخته جواب صریحي ندارد، بنابراین پارامترهاي مدل را با استفاده از الگوریتم EM برآورد می‌کنیم. فرض کنید $X = (x_1, \cdots, x_T)'$ نمونه‌اي تصادفی به‌اندازه T از توزیع آمیخته گاما

$$F_k(x_t) = \sum_{j=1}^{k} \lambda_j f_j^{\Phi_j}(x_t)$$

با تابع لگاریتم درستنمایي

$$Q(F_k) = \sum_{t=1}^{T} \log \sum_{j=1}^{k} \lambda_j f_j^{\Phi_j}(x_t)$$

باشد. چون ماکسیمم کردن تابع $Q(F_k)$ سخت و پیچیده هست، بنابراین متغیر نشانگر Z_{tj} را به‌منظور استفاده از الگوریتم EM تعریف می‌کنیم. اگر مشاهده $t-$ام متعلق به مؤلفه‌ی $j-$ام باشد آنگاه $Z_{tj} = 1$، در غیر این صورت $Z_{tj} = 0$ است. تابع لگاریتم درستنمایی کامل بر اساس داده‌هاي کامل $X_c = (X, Z)$ به‌صورت

$$Q_c(F_k) = \sum_{t=1}^{T} \sum_{j=1}^{k} Z_{tj} \log\left\{\lambda_j f_j^{\Phi_j}(x_t)\right\}$$

قابل‌محاسبه هست. برای الگوریتم EM، به‌طور مکرر تابع امید ریاضی شرطی $E(Q_c(F_k)|X, \theta^{(i)})$ را ماکسیمم می‌کنیم که در آن $\theta^{(i)}$ مقدار فعلی در $i-$اُمین تکرار و $\theta^{(0)}$ نقطه شروع مشخص‌شده توسط کاربر است.

در گام$-E$، برای $i-$اُمین تکرار $Z_{tj}^{(i)}, i = 1, 2, \cdots$ به‌صورت

$$Z_{tj}^{(i)} = P_{\theta^i}(Z_{tj} = 1|x_t) = \frac{\lambda_j f_j^{\Phi_j^{(i)}}(x_t)}{F_j(x_t)} \qquad (۶)$$

برآورد می‌شود.

در گام$-M$، ابتدا $\lambda_j^{(i+1)}$ و$a_j^{(i+1)}$ با ماکسیمم کردن $E(Q_c(F_k)|X, \theta^{(i)})$ به ترتیب نسبت به λ_j و a_j محاسبه می‌شود. براي این منظور از $E(Q_c(F_k)|X, \theta^{(i)})$ نسبت به λ_j و a_j مشتق گرفته و معادل با صفر قرار می‌دهیم. با حل این معادله داریم

$$\lambda_j^{(i+1)} = \frac{1}{T}\sum_{t=1}^{T} z_{tj}^{(i)} \qquad (۷)$$

و

$$\frac{E(Q_c(F_k)|X, \theta^{(i)})}{\partial a_j} = \sum_{t=1}^{T} z_{tj}^{(i)} \left[\log(x_t) + \log\left(B_j^{(i)}\right) - D(a_j)\right] = 0 \qquad (۸)$$

که در آن $D(a_j)$ نشان‌دهنده تابع دي-گاما است. سپس $B_j^{(i+1)}$ را با ماکسیمم کردن $E\left(Q_c(F_k)|X, \theta^{(i)}\right)$ نسبت به B_j اما در مقدار$a_j^{(i+1)}$ محاسبه می‌کنیم. به‌عبارت‌دیگر

$$B_j^{(i+1)} = \frac{a_j^{(i+1)} \sum_{t=1}^{T} z_{tj}^{(i)}}{\sum_{t=1}^{T} x_t z_{tj}}.$$

۴ مطالعه شبیه‌سازی

در این بخش عملکرد دو الگوریتم یادگیری ذکرشده با استفاده از مطالعات شبیه‌سازی مقایسه می‌شود. مدل آمیخته گاما به‌عنوان مدل درست در نظر گرفته‌شده و نمونه‌ای به‌اندازه $T = 400$ از مدل درست تولیدشده که در آن

$$k = 3, a_1 = 20, a_2 = 32, a_3 = 40, B_1 = 14, B_2 = 10, B_3 = 6,$$

$$c_1 = 0.5, c_2 = 0.3, c_3 = 0.2$$

است. هدف برازش مدل آمیخته گاما با استفاده از دو روش یادگیری آمیخته تقویت‌شده و یادگیری آمیخته تقویتی تصحیح‌شده بر اساس تابع درستنمایي ماکسیمم است. مقادیر برآورد پارامترهای مدل و تابع هدف تحت دو روش یادگیری آمیخته به ترتیب در جدول‌های ۳ و ۴ ارائه‌شده است. در جدول ۳، مقادیر برآورد پارامترهای مدل و تابع هدف تحت روش یادگیری آمیخته تقویتی تصحیح‌شده گزارش‌شده است. با توجه به جدول ۳ مشاهده می‌شود که مدل برآورد شده آمیخته با ۳ مؤلفه آمیخته

$$F_3(x_t) = 0.4980 G(23.1348, 15.9789) + 0.2493 G(32.2099, 7.1645) + 0.2527 G(35.0403, 4.7510)$$

مدلی مناسب برای برآورد بر روی داده‌های تولیدشده است. با در نظر گرفتن $\alpha = 0.05$، از جدول ۳ مشاهده می‌شود چون مقدار آماره $2LR(\hat{\theta}_k, \hat{\theta}_{k-1})$ برای $k = 2, 3$ بزرگ‌تر از $\chi^2_{0.975}(1) = 5.0238$ است، بنابراین مؤلفه‌های $f_3^{\Phi_3}(x_t)$ و $f_2^{\Phi_2}(x_t)$ به مدل اضافه می‌شود اما

ازآنجایی‌که مقدار آماره $2LR(\hat{\theta}_k, \hat{\theta}_{k-1})$ برای $k = 4$ کوچک‌تر از $\chi^2_{0.975}(1) = 5.0238$ است بنابراین دو مدل آمیخته $F_3(x_t)$ و $F_4(x_t)$ مدل‌های معادل هستند، پس مؤلفه $f_4^{\Phi_4}(x_t)$ به مدل اضافه نمی‌شود. مقادیر برآورد پارامترهای مدل و تابع هدف تحت روش یادگیری آمیخته تقویت‌شده در جدول ۴ ارائه‌شده است. با توجه به جدول ۴ مشاهده می‌شود که تحت روش یادگیری آمیخته تقویت‌شده، مدل آمیخته با ۵ مؤلفه آمیخته به‌عنوان مدل مناسب برای برازش بر روی داده‌ها انتخاب‌شده است. وقتی‌که $K = 4$ است اگرچه اختلاف تابع لگاریتم درستنمایی دو مدل آمیخته، ۰٫۰۰۷۳، بزرگ‌تر از ۰٫۰۰۰۰ است و این اختلاف باعث می‌شود که مدل $f_4^{\Phi_4}(x_t)$ بر اساس روش آمیخته تقویت‌شده به مدل آمیخته فعلی اضافه شود، اما این اختلاف بسیار ناچیز است؛ بنابراین مدل درست تحت روش یادگیری آمیخته تقویت‌شده به‌عنوان مدل بهینه انتخاب‌نشده است، به‌عبارت‌دیگر مدل آمیخته‌ای با تعداد مؤلفه‌های آمیخته بیشتر به‌عنوان مدل بهینه انتخاب‌شده است که برای برازش این مدل نیاز به برآورد تعداد بیشتری از مؤلفه‌های آمیخته (پارامتر) است.

جدول ۳. مقادیر برآورد پارامترهای مدل و تابع هدف تحت روش یادگیری آمیخته تقویتی تصحیح‌شده

	$K=1$	$K=2$	$K=3$	$K=4$
$\hat{c}_1$	۱٫۰۰۰۰	۰٫۶۶۶۴	۰٫۴۹۸۰	۰٫۴۹۷۹
$\hat{a}_1$	۲۳٫۱۳۴۸	۲۳٫۱۳۴۸	۲۳٫۱۳۴۸	۲۳٫۱۳۴۸
$\hat{B}_1$	۱۵٫۹۷۸۹	۱۵٫۹۷۸۹	۱۵٫۹۷۸۹	۱۵٫۹۷۸۹
$\hat{c}_2$		۰٫۳۳۳۶	۰٫۲۴۹۳	۰٫۲۴۹۳
$\hat{a}_2$		۳۲٫۲۰۹۹	۳۲٫۲۰۹۹	۳۲٫۲۰۹۹
$\hat{B}_2$		۷٫۱۶۴۵	۷٫۱۶۴۵	۷٫۱۶۴۵
$\hat{c}_3$			۰٫۲۵۲۷	۰٫۲۵۲۶
$\hat{a}_3$			۳۵٫۰۴۰۳	۳۵٫۰۴۰۳
$\hat{B}_3$			۴٫۷۵۱۰	۴٫۷۵۱۰
$\hat{c}_4$				۰٫۰۰۰۲
$\hat{a}_4$				۲۲٫۶۱۶۷
$\hat{B}_4$				۷٫۶۰۹۸
$2LR(\hat{\theta}_k, \hat{\theta}_{k-1})$		۲۰۰٫۶۹۳۹	۴۷٫۸۹۷۹	۰٫۰۱۴۶

نمودار تابع چگالی داده‌های تولیدشده و مدل‌های برازش داده‌شده بر روی این داده‌ها با استفاده از دو روش یادگیری ذکرشده در شکل ۱ نمایش داده‌شده است. این نمودار نشان می‌دهد که مدل برآورد شده با استفاده از روش یادگیری آمیخته تقویتی تصحیح‌شده برازش مناسب‌تری را ایجاد می‌کند.

۵ آنالیز داده‌های حقیقی

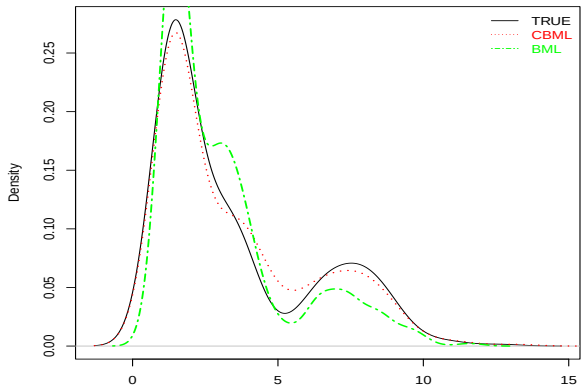
مجموعه داده‌های مربوط به واردات کالاهای ایالات‌متحده توسط گمرک (برحسب میلیون دلار) طی سال‌ها ۱۹۸۹ تا ۲۰۲۲ را در نظر بگیرید. این داده‌ها در

$https://fred.stlouisfed.org/series/IMP0004$

موجود است. نمودار تابع چگالی و نمودار چندک - چندک این داده‌ها در شکل ۲ نمایش داده‌شده است. شکل ۲ نشان می‌دهد که در نمودار چندک-چندک، تعداد زیادی از مشاهدات نزدیک به خط ۴۵ درجه (نیمساز اول و سوم) نیستند، به‌عبارت‌دیگر مشاهدات از توزیع نرمال پیروی نمی‌کنند. همچنین با توجه به مقادیر ضریب چولگی و ضریب کشیدگی این مشاهدات (که به ترتیب برابر ۰۷۳۶.۰ و ۱۸۴۱.۰۱- هستند) و $p-$مقدار آزمون شاپیرو که کوچک‌تر از ۰۵.۰ است این نتیجه نیز تائید می‌شود، پس فرض نرمال بودن مشاهدات رد می‌شود. همچنین نمودار تابع چگالی، شکل ۲، دو مدی بودن داده‌ها را نشان می‌دهد، بنابراین مدل آمیخته گاما برای برازش بر روی این مجموعه از داده‌ها پیشنهاد و پارامترهای مدل با استفاده از روش یادگیری آمیخته تقویتی تصحیح‌شده برآورد شده است.

جدول ۰۴. مقادیر برآورد پارامترهای مدل و تابع هدف تحت روش یادگیری آمیخته تقویت‌شده

	$K=۱$	$K=۲$	$K=۳$	$K=۴$	$K=۵$	$K=۶$
$\hat{c}_۱$	۱٫۰۰۰۰	۰٫۶۶۶۴	۰٫۴۹۸۰	۰٫۴۹۷۹	۰٫۴۸۷۶	۰٫۴۸۶۹
$\hat{a}_۱$	۲۳٫۱۳۴۸	۲۳٫۱۳۴۸	۲۳٫۱۳۴۸	۲۳٫۱۳۴۸	۲۳٫۱۳۴۸	۲۳٫۱۳۴۸
$\hat{B}_۱$	۱۵٫۹۷۸۹	۱۵٫۹۷۸۹	۱۵٫۹۷۸۹	۱۵٫۹۷۸۹	۱۵٫۹۷۸۹	۱۵٫۹۷۸۹
$\hat{c}_۲$		۰٫۳۳۳۶	۰٫۲۴۹۳	۰٫۲۴۹۳	۰٫۲۴۴۲	۰٫۲۴۳۷
$\hat{a}_۲$		۳۲٫۲۰۹۹	۳۲٫۲۰۹۹	۳۲٫۲۰۹۹	۳۲٫۲۰۹۹	۳۲٫۲۰۹۹
$\hat{B}_۲$		۷٫۱۶۴۵	۷٫۱۶۴۵	۷٫۱۶۴۵	۷٫۱۶۴۵	۷٫۱۶۴۵
$\hat{c}_۳$			۰٫۲۵۲۷	۰٫۲۵۲۶	۰٫۲۴۷۴	۰٫۲۴۷۰
$\hat{a}_۳$			۳۵٫۰۴۰۳	۳۵٫۰۴۰۳	۳۵٫۰۴۰۳	۳۵٫۰۴۰۳
$\hat{B}_۳$			۴٫۷۵۱۰	۴٫۷۵۱۰	۴٫۷۵۱۰	۴٫۷۵۱۰
$\hat{c}_۴$				۰٫۰۰۰۲	۰٫۰۰۰۲	۰٫۰۰۰۱
$\hat{a}_۴$				۲۲٫۶۱۶۷	۲۲٫۶۱۶۷	۲۲٫۶۱۶۷
$\hat{B}_۴$				۷٫۶۰۹۸	۷٫۶۰۹۸	۷٫۶۰۹۸
$\hat{c}_۵$					۰٫۰۲۰۶	۰٫۰۲۰۶
$\hat{a}_۵$					۲۴٫۷۰۹۸	۲۴٫۷۰۹۸
$\hat{B}_۵$					۷٫۵۱۱۸	۷٫۵۱۱۸
$\hat{c}_۶$						۰٫۰۰۱۷
$\hat{a}_۶$						۲۲٫۶۶۳۸
$\hat{B}_۶$						۱۶٫۰۱۶۶
$LR(\hat{\theta}_k, \hat{\theta}_{k-۱})$		۱۰۰٫۳۴۶۹	۲۳٫۹۴۸۶	۰٫۰۰۷۳	۰٫۰۲۲۷	−۱٫۶۲۴۴

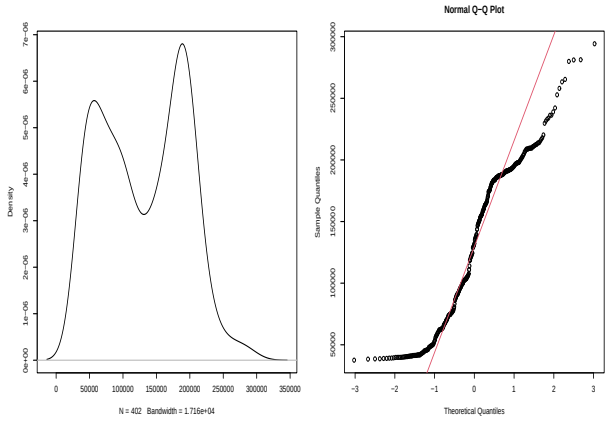


شکل ۱: نمودار تابع چگالی داده‌های تولیدشده و مدل‌های برازش داده‌شده .

پارامترهاي مدل پیشنهادی با استفاده از روش یادگیری آمیخته تقویتی تصحیح‌شده محاسبه و مقادیر پارامترهاي برآورد شده و تابع هدف در جدول ۵ ارائه‌شده است. با در نظر گرفتن $\alpha = ۰٫۰۵$ و مقایسه مقدار آماره $۲LR(\hat{\theta}_k, \hat{\theta}_{k-۱})$ برای $k = ۲, ۳$ با $\chi^۲_{۰٫۹۷۵}(۱)$ از جدول ۵ مشاهده می‌شود که مدل آمیخته برآورد شده

$$F_۲(x_t) = ۰٫۴۸۰۰G(۶٫۱۰۴۴, ۱۲۷۶۲٫۸۴۳۲۳) + ۰٫۵۲۰۰G(۴۳٫۳۳۳۱, ۴۳۴۵٫۷۵۸۹)$$

مدلی مناسب برای برآورد بر روی این مجموعه از داده‌ها است. برای $k = ۲$ مقدار آماره $۲LR(\hat{\theta}_k, \hat{\theta}_{k-۱}) = ۱۴۰/۹۷۴۲$ بزرگ‌تر از $\chi^۲_{۰/۹۷۵}(۱) = ۵/۰۲۳۸$ است بنابراین مؤلفه‌های $f^{\Phi_۲}_۲(x_t)$ به مدل اضافه می‌شود، اما ازآنجایی‌که مقدار آماره $۲LR(\hat{\theta}_k, \hat{\theta}_{k-۱}) = ۱/۶۸۴۹$ برای $k = ۳$ کوچک‌تر از $\chi^۲_{۰/۹۷۵}(۱) = ۵/۰۲۳۸$ است بنابراین دو مدل آمیخته $F_۲(x_t)$ و $F_۳(x_t)$ مدل‌های معادل هستند، پس مؤلفه $f^{\Phi_۳}_۳(x_t)$ به مدل اضافه نمی‌شود. تابع چگالی داده‌ها و مدل پیشنهادشده در شکل ۳ رسم شده است. شکل ۳ نشان می‌دهد که مدل آمیخته گاما با دو مؤلفه آمیخته براي برازش بر روي داده‌ها مدل مناسبی است.



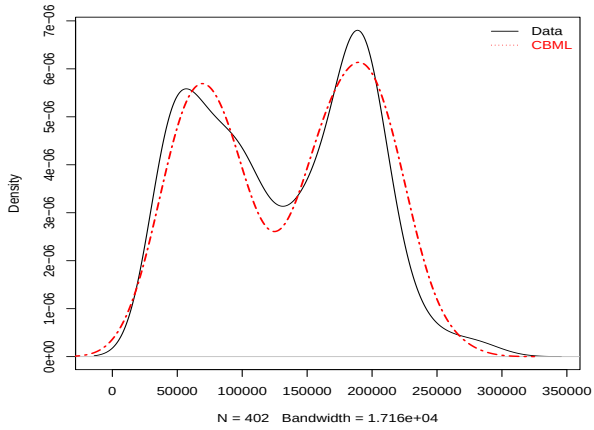
شکل ۲: نمودار تابع چگالی و نمودار چندک–چندک مشاهدات.

جدول ۵. مقادیر برآورد پارامترهای مدل و تابع هدف تحت روش یادگیری آمیخته تقویتی تصحیح‌شده

	$K = ۱$	$K = ۲$	$K = ۳$
$\hat{c}_۱$	۱/۰۰۰۰	۰/۴۸۰۰	۰/۴۷۹۳
$\hat{a}_۱$	۶/۱۰۴۴	۶/۱۰۴۴	۶/۱۰۴۴
$\hat{B}_۱$	۱۲۷۶۲/۸۴۳۲۳	۱۲۷۶۲/۸۴۳۲۳	۱۲۷۶۲/۸۴۳۲۳
$\hat{c}_۲$		۰/۵۲۰۰	۰/۵۱۹۳
$\hat{a}_۲$		۴۳/۳۳۳۱	۴۳/۳۳۳۱
$\hat{B}_۲$		۴۳۴۵/۷۵۸۹	۴۳۴۵/۷۵۸۹
$\hat{c}_۳$			۰/۰۰۱۳
$\hat{a}_۳$			۵۰۳/۸۲۷
$\hat{B}_۳$			۳۸۱/۷۹۰۶
$۲LR(\hat{\theta}_k, \hat{\theta}_{k-۱})$		۹۷۴۲.۱۴۰	۶۸۴۹.۱

۶ بحث و نتیجه‌گیری

تابع لگاریتم درستنمایي و معیارهای اطلاع از جمله توابع هدفی هستند که در روش یادگیری آمیخته تقویت‌شده استفاده می‌شوند. به‌عبارت‌دیگر در هر تکرار مؤلفه جدیدي به مدل آمیخته اضافه مي‌شود به‌طوری‌که تابع لگاریتم درستنمایي (معیار اطلاع) مدل آمیخته جدید نسبت به تابع لگاریتم درستنمایي (معیار اطلاع) مدل آمیخته فعلي افزایش (کاهش) می‌یابد. در بعضي از موارد ممکن است این افزایش ناچیز باشد و مدل آمیخته جدید و مدل آمیخته فعلي معادل باشند، به این مفهوم که اضافه کردن مؤلفه جدید باعث بهبود مدل فعلی نمی‌شود. به‌عبارت‌دیگر در این روش، اختلاف تابع درستنمایی (معیار اطلاع) دو مدل آمیخته فعلی و آمیخته جدید با صفر مقایسه می‌شود. مؤلفه جدید در صورتی به مدل



شکل ۳: نمودار تابع چگالی مشاهدات و مدل آمیخته گاما با دو مؤلفه $F_۲(x_t)$.

آمیخته فعلی اضافه می‌شود که اختلاف تابع درستنمایی (معیار اطلاع) مدل آمیخته فعلی از مدل آمیخته جدید بزرگ‌تر (کوچک‌تر) از صفر باشد. اگر مقدار این اختلاف ناچیز و به صفر نزدیک باشد اضافه کردن مؤلفه جدید، مدل آمیخته فعلی را بهبود نمی‌بخشد. ازاین‌جهت این روش نیاز به تصحیح دارد. به‌عبارت‌دیگر روش یادگیری آمیخته تقویت‌شده با تابع هدف لگاریتم درستنمایی (معیار اطلاع) مدل آمیخته توانایی تشخیص مدل‌های معادل را ندارد. در این مقاله، روش یادگیری آمیخته تقویت‌شده بر اساس آزمون انتخاب مدل وونگ تصحیح‌شده است. در روش آمیخته تقویتی تصحیح‌شده مؤلفه جدید با وزن بهینه c_k^* به مدل آمیخته فعلی اضافه می‌شود اگر

$$Q((۱-c_k^*)F_{k-۱}+c_k^* f_k)-Q(F_{k-۱})>\frac{۱}{۲}\chi^۲_{۱-\frac{\alpha}{۲}}(۱).$$

به‌عبارت‌دیگر اضافه کردن مؤلفه جدید به مدل آمیخته فعلی باعث بهبود مدل برازش می‌شود. عملکرد دو روش یادگیری با استفاده از داده‌های شبیه‌سازی و مجموعه داده‌های واردات کالای ایالات‌متحده توسط گمرک ارزیابی‌شده است. با استفاده از مطالعات شبیه‌سازی نشان داده‌شده که ممکن است روش یادگیری آمیخته تقویت‌شده مدل درست را انتخاب نکند. اگرچه افزایش ناچیز تابع لگاریتم درستنمایی مدل آمیخته جدید باعث اضافه شدن مؤلفه جدید به مدل آمیخته فعلی می‌شود اما باعث بهبود مدل نمی‌شود.

مراجع

[1] Du, J., Hu, Y., and Jiang, H. (2011). Boosted mixture learning of Gaussian mixture hidden markov models based on maximum likelihood for speech recognition. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, **19(7)**, 2091-2100.

[2] Freund, Y., and Schapire, R. (1997). A decision theoretic generalization of on line learning and an application to boosting. *Journal of Computer and System Sciences*, **55 (1)**, 119-139.

[3] Kim, M., and Pavlovic, V. (2007). A recursive method for discriminative mixture learning. *In Proc. ICML*, **24**, 409-416. https://doi.org/10.1145/1273496.1273548.

[4] Kullback, S., and Leibler, R. A. (1951). Information and sufficiency modelling. *Annals of Mathematical Statistics*, **22**, 79-86.

[5] Martins, I., Carvalho, P., Corte-Real, L., and Alba-Castro, J. L. (2018). BMOG: boosted Gaussian mixture model with controlled complexity for background subtraction. *Pattern Analysis & Applications*, **2(3)**, 641–654.

[6] Mason, L., Baxter, J., Bartlett, P., and Frean, M. (1999). Boosting algorithms as gradient descent in function space. *NIPS*, **11**, 512-518.

[7] McLachlan, G. (2001). *Finite Mixture Models*. New York: Wiley.

[8] Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., and Liu, J. (2020). Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *Journal of Hydrology*, **586(1)**, 124901.

[9] Pavlovic, V. (2004). Model based motion clustering using boosted mixture modeling. *In Proc. CVPR*, **1**, 811-818.

[10] Schapire, R. (1990). The strength of weaklearnability. *Mach. Learn*, **5**, 197–227.

[11] Schapire, R. (1999). Theoretical views of boosting and applications. *In: Watanabe, O., Yokomori, T. (eds) Algorithmic Learning Theory. ALT 1999. Lecture Notes in Computer Science, vol 1720*. Springer, Berlin, Heidelberg, 13-25.

[12] Schneider, L., Chalmers, R. P., Debelak, R., and Merkle, E. C. (2020). Model selection of nested and non-nested item response models using Vuong tests. *Multivariate Behavioral Research*, **55(5)**, 664-684.

[13] Tang, H., Chu, S., Hasegawa-Johnson, M., and Huang, T. (2009). Emotion recognition from speech via boosted Gaussian mixture models. *2009 IEEE International Conference on Multimedia and Expo*, New York, NY, US, 294-297.

[14] Vlassis, N., and Likas, A. (2002). A greedy EM algorithm for Gaussian mixture learning. *Neural Process. Lett*, **15**, 77-87.

[15] Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**, 307-333.

[16] Young, D. S., Chen, X., and Hewage, D. C. (2019). Finite mixture-of-Gamma distributions: Estimation, inference and model-based clustering. *Advances in Data Analysis and Classification*, **13**, 1053-1082.

Correcting Boosted Mixture Learning method using Vuong's test and its application in the Gamma Mixture Model

Sedigheh Zamani Mehreyan [1]

Abstract:

The boosted mixture learning method, BML, is an incremental method to learn mixture models for the classification problem. In each step of the boosted mixture learning method, a new component is added to the mixture model according to an objective function to ensure that the objective function is maximized. Sometimes the likelihood function or equivalently information criteria are defined as the objective function of BML. The mixture model is updated whenever a new component is added to the mixture model based on the maximum likelihood function and information criteria. Since the information criteria does not have the ability to identify equivalent models, therefore, it is possible that the new mixture model and the current mixture model are equivalent. In this paper, the boosted mixture learning method has been corrected using Vuong's model selection test, which has the ability to identify equivalent models. The performance of two learning methods is evaluated over simulation data and over the U.S. imports of goods by customs basis.

Keywords: Boosted mixture learning method, Maximum likelihood function, Model selection, Vuong's test.

[1] **Department of Statistics, Imam Khomeini International University, Qazvin, Iran, zamani@sci.ikiu.ac.ir**