طبقه بندی کننده مبتنی بر DEA فازی و کاربردهای آن در مدیریت مراقبت های بهداشتی

دانشجو: شایان رادی

استاد: دکتر محسن رستمی مال خلیفه

کد دانشجویی:۹۷۰۰۸٦٦۳۲

طبقه بندی کننده مبتنی بر DEA فازی و کاربردهای آن در مدیریت مراقبت های بهداشتی

Ai-bing Ji . Yanhua Qiao . Chang Liu

دریافت شده در: ۲۸ June ۲۰۱۸ / مورد پذیرش قرار گرفته در ۱۱ February ۲۰۱۹

چکیده

مدل‌های طبقه بندی فازی غیرخطی عملکرد طبقه‌بندی بهتری نسبت به طبقه بندی‌کننده‌های فازی خطی دارند. در بسیاری از مسائل طبقه‌بندی فازی غیرخطی، توابع تفکیک فازی تکه‌ای خطی می‌توانند توابع تفکیک فازی غیرخطی را تقریبی کنند. در این مقاله، ابتدا طبقه‌بندی‌کننده فازی بر اساس تحلیل پوششی داده‌ها (DEA) برای داده‌های آموزشی فازی قابل تفکیک افزایشی ساخته می‌شود که می‌تواند به طور گسترده در مدیریت مراقبت‌های بهداشتی با ویژگی‌های فازی اعمال شود، سپس طبقه‌بندی‌کننده مبتنی بر DEA فازی پیشنهادی را در تشخیص اعمال می‌کنیم. کرونری با علائم فازی و طبقه بندی مجموعه داده های سرطان پستان با اختلال فازی. آزمایش‌های عددی نشان می‌دهند که طبقه‌بندی‌کننده مبتنی بر DEA فازی دقیق و قوی است.

کلمات کلیدی تحلیل پوششی داده های فازی، دستگاه طبقه بندی DEA، تجزیه و تحلیل متمایز خطی تکه ای، مجموعه آموزشی فازی، مدیریت بهداشت و درمان

۱ مقدمه

داده کاوی یک روش مهم مدیریت دانش و مدیریت مراقبت های بهداشتی است[۱]، الگوریتم داده کاوی کلاسیک داده ها را دقیق فرض می کند، اما در برخی از برنامه ها، داده هایی که قرار است پردازش شوند به ندرت دقیقاً تعیین می شوند. عدم قطعیت با انواع مختلف در ساختار آنها کم و بیش پنهان است: عدم دقت یا تقریب، تصادفی و همچنین مبهم. آخرین نوع عدم قطعیت اغلب با ویژگی‌های زبان طبیعی مورد استفاده در فرآیند جمع‌آوری داده‌ها مرتبط است. در ارتباط بین مردم، میزان مشخصی از ابهام نه تنها قابل قبول است، بلکه حتی مفید است. در تشخیص بیماری، برخی از علائم یک بیماری نیز مبهم هستند. به نظر می رسد جهان نسبتاً نامطمئن تر از نتایجی است که قصد بیان آن را دارند. در نتیجه، توسعه ابزارهای ریاضی کافی برای پردازش ابهامات بسیار ضروری شد.

از زمان معرفی آن توسط لطفی زاده، تئوری مجموعه های فازی به طور پرباری در بسیاری از حوزه ها مورد استفاده قرار گرفته است. داده کاوی یکی از زمینه های مهم کاربرد آن است. در اوایل ظهور داده کاوی، چندین کار استفاده از تئوری مجموعه های فازی را در این حوزه پیشنهاد کرده اند.

مشارکت مجموعه های فازی در داده کاوی گوناگون است: افزایش قابلیت تفسیر، افزایش استحکام فرآیند و مدیریت اطلاعات نامشخص. هر دو با معرفی تئوری مجموعه‌های فازی برای ایجاد مدل داده‌کاوی فازی ارائه می‌شوند که به آن فرآیند ظرفیت استخراج اطلاعات پیچیده را ارائه می‌دهد که پردازش آن در یک محیط کلاسیک دشوار است. استحکام فرآیند آن را قادر می سازد نتایج مشابهی را هنگام مواجهه با داده ها با تغییرات کوچک (به عنوان نمونه در حضور نویز) ایجاد کند.

داده کاوی قصد دارد تا مدلی را بر اساس مجموعه داده هایی که دانش پس زمینه ای را ارائه می دهند راه اندازی کند. مدل را می توان دانش جدیدی دانست که از یادگیری تولید می شود، می تواند به اشکال مختلف باشد: به عنوان مثال. تابع ریاضی، شبکه عصبی، پایه قوانین، الگوها، قوانین تداعی.

تکنیک‌های داده کاوی مزایای متعددی را در مدیریت مراقبت‌های بهداشتی ارائه می‌دهند [۲، ۳] مانند تشخیص علت بیماری‌ها، شناسایی روش‌های درمانی پزشکی، ساختن سیستم‌های توصیه دارو، توسعه پروفایل سلامت افراد، سیاست‌های تحقیقاتی مراقبت‌های بهداشتی، حتی تشخیص تقلب در بیمه سلامت. فن‌آوری‌های داده‌کاوی می‌توانند با گروه‌بندی بیماران مبتلا به انواع بیماری‌های مشابه، مزایایی را برای مراقبت‌های بهداشتی فراهم کنند تا سازمان‌های مراقبت‌های بهداشتی درمان‌های بهتری برای آنها ارائه دهند.

داده کاوی فازی توسعه داده کاوی است که در آن مدل سازی مجموعه های فازی معرفی شده است. بسیاری از الگوریتم های کلاسیک داده کاوی به موارد فازی گسترش یافته اند [۴–۷]. بسط یک الگوریتم کلاسیک برای ایجاد یک الگوریتم یادگیری فازی کار جالبی است. مقالات زیادی در مورد درختان تصمیم فازی [۸–۱۰]، در ساخت پایه قوانین فازی [۱۱–۱۳] و روش هسته فازی [۱۴–۱۸] منتشر شده است. چالش در چنین موردی پیشنهاد الگوریتمی است که هم بتواند ورودی فازی را مدیریت کند و هم ویژگی های اصلی الگوریتم کلاسیک را برآورده کند.

طبقه بندی به عنوان یک روش داده کاوی مهم، یک سیستم سازماندهی مبتنی بر شباهت ها و تفاوت ها، یک روش مهم پیش بینی (تصمیم گیری) است و به طور گسترده در مدیریت مراقبت های بهداشتی [۲، ۱۹] و نظارت بر بیماری ها و هشدار اولیه [۲۰] استفاده می شود. . هدف طبقه‌بندی این است که با ارزیابی مجموعه‌ای از مقادیر مشخصه به گروه خاصی تعلق دارد یا خیر، تکنیک‌ها و روش‌های ماشین طبقه‌بندی مرسوم را می‌توان در آثار هان و کامبر [۲۱] یافت. در عمل، یک سیستم تصمیم گیری مراقبت های بهداشتی تعیین می کند که آیا یک بیمار با توجه به برخی از شاخص های علائم و شرایط فیزیکی مشخص، بیماری دارد یا خیر، در اصل، تشخیص بیماری ها یک مشکل طبقه بندی است.

تجزیه و تحلیل پوششی داده ها (DEA) برای ارزیابی کارایی نسبی در بین تعداد معینی از واحدهای تصمیم گیری (DMUs) با ورودی های متعدد و خروجی های متعدد، با حل مسائل برنامه ریزی خطی برای هر DMU استفاده می شود. مدل های شناخته شده DEA شامل مدل CCR توسط Charnes et al می باشد. [۲۲] و مدل BCC توسط با[۲۳] .Bunker et al.

DEA عمدتاً در ارزیابی عملکرد استفاده می شود [۲۴-۲۶]. علاوه بر ارزیابی عملکرد، DEA را می توان در طبقه بندی نیز به کار برد. [۲۷]. در کارهای آنها، حوزه های پذیرش ساخته شد و یک سیستم تصمیم گیری مبتنی بر نمونه ارائه شد که بر اساس نمونه های از پیش تعیین شده توسط کارشناسان، در مورد پذیرش یا رد ریسک اعتباری تصمیم گیری می کند. بر اساس مدل DEA، دو رویکرد برنامه‌ریزی ریاضی جدید برای به حداقل رساندن مجموع انحرافات و برای مفهوم کارایی نسبی DEA در حل دو مسئله طبقه‌بندی گروهی توسعه داده شده است [۲۸]. یان و وی [۲۹] یک رابطه هم ارزی بین ماشین طبقه بندی DEA و مدل DEA ایجاد کردند و سپس یک ماشین طبقه بندی مبتنی بر DEA ایجاد کردند که در آن طبقه بندی داده ها معادل آزمایش اینکه آیا یک DMU خاص در مجموعه امکان تولید قرار دارد یا خیر. در تمام طبقه‌بندی‌کننده‌های مبتنی بر DEA که در بالا ذکر شد، همه آنها بر اساس داده‌های دقیق هستند، وی [۳۰] طبقه‌بندی‌کننده‌هایی را برای مقابله با مشکلات طبقه‌بندی باینری با داده‌های بازه‌ای توسعه داد، اگرچه دارای برخی عدم قطعیت‌ها، همچنین بدون ابهام است، بنابراین ساختن آن بسیار ضروری است. طبقه بندی کننده بر اساس DEA فازی برای داده های آموزش فازی.

در این مقاله، یک طبقه‌بندی‌کننده فازی جدید مبتنی بر DEA ساخته شده است. ما ابتدا ماشین طبقه بندی تحلیل پوششی داده های فازی را برای داده های آموزش فازی یکنواختی شرطی معرفی می کنیم. سپس آزمایش‌هایی بر روی مجموعه داده‌های فازی واقعی و مجموعه داده‌های UCI مختل برای اثبات صحت و کارایی مدل پیشنهادی انجام می‌شود.

۲ پیشینه

در این بخش ابتدا مفاهیم و مدل‌های مرتبط DEA، DEA فازی و ماشین طبقه‌بندی DEA را بررسی می‌کنیم، سپس تعاریف پایه‌ای در مورد اندازه‌گیری امکان و برنامه‌ریزی محدود شانس فازی ارائه می‌کنیم، چند قضیه مهم اثبات می‌شوند.

۲.۱ تحلیل پوششی داده‌ها

فرض کنید n، DMU وجود دارد که هر کدام از s ورودی ها m خروجی تولید می کنند. DMU_0، DMU است که باید ارزیابی شود. DMU_k از بسته ورودی $x_k = (x_{1k}, x_{2k}, \dots, x_{sk})$ برای تولید بسته خروجی $y_k = (y_{1k}, y_{2k}, \dots, y_{mk})$ استفاده می کند. یک مدل استاندارد DEA برای ارزیابی DMU_0، معروف به مدل [۲۲] CCR، در مدل فرموله شده است:

$$\begin{cases} \max \sum_{i=1}^{m} \mu_i y_{i0} \\ \sum_{i=1}^{m} \mu_i y_{ik} - \sum_{j=1}^{s} \omega_j x_{jk} \le 0, (k = 1, 2, ..., n); \\ \sum_{j=1}^{s} \omega_j x_{j0} = 1; \\ \mu_i \ge 0, \omega_j \ge 0, (i = 1, 2, ...m; j = 1, 2, ..., s); \end{cases} \quad (1)$$

مدل های ارزیابی کارایی بالا مدل های CCR نامیده می شوند که فرم های خروجی گرا هستند.

شکل دوگانه این مدل (۱) به صورت زیر است:

$$\begin{cases} \min\theta \\ \sum_{k=1}^{n} \lambda_k x_{jk} \leq \theta x_{j0}, (j=1,2,...,s); \\ \sum_{k=1}^{n} \lambda_k y_{ik} \geq y_{i0}, (i=1,2,...,m); \\ \lambda_k \geq 0, (k=1,2,...n). \end{cases} \tag{2}$$

ماهیت مدل‌های CCR یافتن بردار وزن بهینه برای به حداکثر رساندن وزن آن برای ورودی‌ها و خروجی‌ها است که امتیاز کارایی DMU تحت ارزیابی را به حداکثر می‌رساند.

۲.۲ تحلیل پوششی داده های فازی

به منظور ارزیابی کارایی با داده های فازی، تئوری مجموعه های فازی به عنوان راهی برای کمی سازی داده های مبهم و مبهم در مدل های DEA پیشنهاد شده است. مدل های فازی DEA شکل مدل های برنامه ریزی خطی فازی را اتخاذ می کنند. مدل CCR با داده های فازی به صورت زیر ارائه می شود:

$$\begin{cases} \max \sum_{i=1}^{m} u_i \tilde{y}_{i0} \\ \sum_{i=1}^{m} u_i \tilde{y}_{ik} - \sum_{i=1}^{s} v_i \tilde{x}_{ik} \tilde{<} \tilde{0} \\ \sum_{i=1}^{s} v_i \tilde{x}_{i0} \tilde{=} 1 \\ u_i \geq 0, v_j \geq 0 (i=1,2,...,m; j=1,2,...,s); \end{cases} \tag{3}$$

جایی که ورودی $\tilde{x}_i(i = 1,2, \dots, m)$ و خروجی $\tilde{y}_i(i = 1,2, \dots, s)$ هر دو اعداد فازی هستند.

چندین روش فازی برای حل مدل DEA فازی [۳۱] پیشنهاد شده است، روش های DEA فازی را می توان به پنج گروه زیر طبقه بندی کرد: (۱) رویکرد تحمل [۳۲]؛ (۲) رویکرد مبتنی بر سطح (۳) .[۳۵-۳۳] – α رویکرد رتبه بندی فازی [۳۶، ۳۷]؛ (۴) رویکرد امکان [۳۸، ۳۹]؛ (۵) سایر تحولات [۴۰-۴۲]. به منظور بهبود قدرت تمایز مدل DEA فازی [۴۳]، DEA فازی متقاطع [۴۴] معرفی شد.

۲.۳ دستگاه طبقه بندی (DCM) DEA

Hong Yan و Quanling Wei [۲۹] روش DEA را به طبقه بندی داده های بزرگ گســترش دادند. آنها هر داده را به عنوان یک DMU ارزیابی شــده، با مقادیر مشــخصــه به عنوان ورودی و خروجی واحد از مقدار ۱ در نظر می گیرند. Toloo [۴۵، ۴۶] رویکرد DEA را برای یافتن کارآمدترین واحد مجموعه داده بدون ورودی (خروجی) صریح گسترش داد.

یک نمونه مجموعه داده آموزشی $\hat{T}$ (با کلاس تک) را در نظر بگیرید

$$\hat{T} = \{x_k \vee k = 1,2, \dots, n\}$$

جایی که $x_k \in E^m$، $x_k > 0$ و $k = 1,2, \dots, n$. با استفاده از این قانون در کاوش DEA، x_k یک $DMUs$ است با مشخصه های مشخصی که توسط $xk = (x_{1k}, x_{2k}, \dots, x_{sk})^T$ توصیف شده است که x_{ik} یکمین مقدار مشخصه DMU_k است.

این مشــکل را می توان با مدل/۲۲/ CCR با DMU_k ورودی داده شــده توســط x_k و خروجی $y_k = 1(k = 1,2,\ldots,n)$ توصــیف کرد. یعنی مجموعه داده های آموزشی نمونه در مدل DEA با $\{(x_k, 1) \vee k = 1,2,\ldots,n\}$ داده می شود.

سپس مدل CCR با مقدار ورودی-خروجی $(x_k, 1)$ برای DMU_k $(k = 1,2,\ldots,n)$ به صورت زیر ارائه می شود:

$$(P)\begin{cases} \max\mu_0 \\ \omega x_k - \mu_0 \geq 0, k = 1,2,\ldots,n; \\ \omega x_0 = 1; \\ \omega \geq 0, \mu_0 \geq 0. \end{cases} \qquad (4)$$

جایی که DMU_{j_0} واحد ارزیابی شده هست و $x_0 = x_{j_0}$ و $1 \leq j_0 \leq n$.

تعریف ۲.۱ [۲۸] اگر مقادیر بهینه (P)، ۱ باشد، DMU_0 ضعیف DEA کارآمد نامیده می شود.

برای DMU_0 با کارایی ضعیف DEA، x_0 برآورده می شود (توجه داشته باشید که $\omega_0 x_0 = 1$؛ $\mu_0^0 = 1$)

$$L: \omega^0 x - 1 = 0$$

L صفحه پشتیبان حوزه پذیرش است و ابر صفحه طبقه بندی نامیده می شود، x_0 در این صفحه قرار دارد.

دامنه پذیرش را می توان به شکل تقاطع آن ارائه کرد:

$$T = \{x \vee \omega^k x - 1 \geq 0, k = 1,2,\ldots,n\}$$

جایی که ω^k و $\mu_0^k = 1(k = 1,2,\ldots,n)$ جواب های بهینه (۴) هستند. آنگاه، برای $x \in \hat{T}$، طبقه بندی تابع به صورت زیر است:

$$d(x) = sign(\min_{1 \leq k \leq n}(\omega^k x - \mu_0^k))$$

اگر $d(x) = 1$، انگاه $x \in T$، $d(x) = 1$، پس $x \notin T$.

۲.٤ محاسبه احتمال و برنامه نویسی محدود بر شانس فازی

تعریف ۲.۲ X را یک مجموعه ناتهی فرض کنید، $P(X)$ کلاس همه زیر مجموعه های X خواهد بود، یک نگاشت.

Pos: $P(X) \to [0,1]$ یک محاسبه احتمال نام خواهد گرفت اگر شرط زیر را ارضا کند:

$$(1) Pos(\varphi) = 0; (2) Pos(X) = 1 (3) Pos(t \in TA_t) = Sup_{t \in T}\{Pos(A_t)\}.$$

تعریف۲.۳ فرض کنید $\tilde{a}$ یک عدد فازی باشد و تابع عضویت آن عبارت است از:

$$\mu_{\tilde{a}}(x) = \begin{cases} \dfrac{x - r_1}{r_2 - r_1}, r_1 \leq x < r_2 \\ 1, x = r_2 \\ \dfrac{x - r_3}{r_2 - r_3}, r_2 < x \leq r_3 \end{cases}$$

جایی که $r_1 \leq r_2 \leq r_3$ و r_1، r_2 و r_3 اعداد حقیقی هستند.

$\tilde{a}$یک عدد فازی مثلثی نامیده می شود که $(r1, r2, r3)$ نشان داده می شود.

کلاس های تمام اعداد فازی مثلثی با $T(R)$ نشان داده می شوند، اگر $\tilde{x}_i(i = 1,2, \dots, n)$ همه اعداد فازی باشند،

$\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ را بردار عدد فازی می نامند، کلاس های همه بردارهای اعداد فازی با $F^n(R)$ کلاس های تمام اعداد فازی نشـــان داده می شـوند، به خصـوص زمانی که $\tilde{x}_i(i = 1,2, \dots, n)$ هسـتند. همه اعداد فازی مثلثی، سـپس $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ را بردار اعداد فازی مثلثی می نامند، کلاس همه بردارهای اعداد فازی مثلثی با $T^n(R)$ نشان داده می شود.

تعریف ۲.٤ فرض کنید $\tilde{a}$ و $\tilde{b}$ دو عدد فازی باشد، سپس $\tilde{a} \vee \tilde{b}$ عدد فازی را با تابع عضویت

$\mu_{\tilde{a}\vee\tilde{b}}(x) = Sup_{s\vee t=x}\{\mu_{\tilde{a}}(s) \wedge \mu_{\tilde{b}}(t)\}$ نشان می دهد.

تعریف ۲.۵ فرض کنید $\tilde{a}, \tilde{b}$ دو عدد فازی باشند، آنگاه $\tilde{a} \gtrsim \tilde{b} \Leftrightarrow \tilde{a} \vee \tilde{b} = \tilde{a}$.

لم ۱ [۴۳] فرض کنید $\tilde{a}$ و $\tilde{b}$ دو عدد فازی باشد، سپس $\tilde{a} \gtrsim \tilde{b}$ اگر و فقط اگر برای $\forall h \in [0,1]$ دو عبارت زیر برقرار باشد:

$$inf\{s: \mu_{\tilde{a}}(s) \geq h\} \geq inf\{s: \mu_{\tilde{b}}(t) \geq h\},$$

$$sup\{s: \mu_{\tilde{a}}(s) \geq h\} \geq \sup\{s: \mu_{\tilde{b}}(t) \geq h\}.$$

طبق لم۱، ما به سادگی می‌توانیم لم بعدی را ثابت کنیم.

لم۲ فرض کنید $\tilde{a}$ ، $\tilde{b}$ دو عدد فازی باشد، (۱) برای $\lambda > 0$ ، $\lambda\tilde{a} \gtrsim \lambda\tilde{b}$ ؛ (۲) برای هر عدد فازی $\tilde{c}$، $\tilde{a} + \tilde{c} > \tilde{b} + \tilde{c}$

تعریف۲.٦ فرض کنید $\tilde{a}$ یک عدد فازی، $\tilde{b}$ یک عدد حقیقی باشد. سپس امکان اندازه گیری رویداد فازی $\tilde{a} < b$ به صورت زیر تعریف می‌شود:

$$Pos(\tilde{a} \leq b) = \{\mu_{\tilde{a}}(x) \vee x \in R. x \leq b\}$$

به طور مشابه داریم:

$$Pos(\tilde{a} < b) = \{\mu_{\tilde{a}}(x) \vee x \in R. x < b\}$$

$$Pos(\tilde{a} \geq b) = \{\mu_{\tilde{a}}(x) \vee x \in R. x \geq b\}$$

$$Pos(\tilde{a} = b) = \mu_{\tilde{a}}(b)$$

قضـــیه۱ $\tilde{a}$ و $\tilde{b}$ را دو عـدد فـازی، c را یـک عـدد حقیقی، برای یـک مرحلـه احتمـال $\alpha\epsilon[0,1]$ اگر $\boldsymbol{\tilde{a} \gtrsim \tilde{b}}$ و $\boldsymbol{Pos(\tilde{b} \geq c) \geq \alpha}$، آنگاه $\boldsymbol{Pos(\tilde{a} \geq c) \geq \alpha}$.

اثبات زیرا $\boldsymbol{\tilde{a} \gtrsim \tilde{b}}$، برای یک x داده شـده ، $t = x$، توسط لم۱، وجود دارد $s: s \leq t$ به طوری که $\mu_{\tilde{\boldsymbol{a}}}(s) \leq \mu_{\tilde{\boldsymbol{b}}}(x)$ ، بنابراین$\mu_{\tilde{a}\vee\tilde{\boldsymbol{b}}}(x) =$ $Sup_{s\vee t=x}\{\mu_{\tilde{\boldsymbol{a}}}(s)\wedge\mu_{\tilde{\boldsymbol{b}}}(t)\} \geq \mu_{\tilde{\boldsymbol{a}}}(s) \wedge \mu_{\tilde{\boldsymbol{b}}}(x) \geq \mu_{\tilde{\boldsymbol{b}}}(x)$، از ایـن رو $Pos(\tilde{a} \geq c) = \sup\{\mu_{\tilde{a}\vee\tilde{b}}(x)|x \geq c\} \geq \sup\{\mu_{\tilde{b}}(x)|x \geq c\} \geq \alpha$.

اگر $\tilde{x}_i(i = 1,2, \dots, n)$ همه اعداد فازی باشـند، آنگاه $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ یک بردار اعداد فازی نام دارد، کلاس همه بردار های اعداد فازی با $F^n(R)$ نشان داده شده است، بخصوص زمانی که $\tilde{x}_i(i = 1,2, \dots, n)$ همه اعداد فازی مثلثی باشند، آنگاه $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ یک بردار عدد فازی مثلثی نامیده می شود، تمام بردارهای اعداد فازی مثلثی با $T^n(R)$ نشان داده می شود.

با پیروی از اصل توسعه زاده، برای تابع $f: R^n \rightarrow R$ و $\tilde{y} = f(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ یک عدد فازی است و تابع عضویت آن عبارت است از:

$$\mu_{\tilde{y}}(v) = \sup_{u_1, u_2, \dots, u_n} \left\{ \min_{1 \le i \le n} \mu_{\tilde{x}_i}(u_i) | v = f(u_1, u_2, \dots, u_n) \right\}$$

قضیه ۲ فرض کنید $\tilde{x}_i (i = 1,2, \dots, n)$ اعداد فازی هستند، $\widetilde{\omega}_i (i = 1,2, \dots, n)$ ضریب واقعی مثبت هستند، $()_\alpha^L$ و $()_\alpha^U$ نشان دهنده کران پایین و بالای مجموعه سطح α یک عدد فازی هستند، سپس برای هر سطح احتمال معین $0 \le \alpha 1$, α۲ و $\alpha 3 \le 1$،
$Pos\{\omega_1\tilde{x}_1 + \omega_2\tilde{x}_2 + \cdots + \omega_n\tilde{x}_n \le b\} \ge a_1$ اگر $\omega_1(\tilde{x})_{a_1}^L + \omega_2(\tilde{x})_{a_1}^L + \cdots + \omega_n(\tilde{x})_{a_1}^L \le b$ ؛ $Pos\{\omega_1\tilde{x}_1 + \omega_2\tilde{x}_2 + \cdots + \omega_n\tilde{x}_n \ge b\} \ge a_2$ اگر $\omega_1(\tilde{x})_{a_2}^L + \omega_2(\tilde{x})_{a_2}^L + \cdots + \omega_n(\tilde{x})_{a_2}^L \ge b$ ؛

$Pos\{\omega_1\tilde{x}_1 + \omega_2\tilde{x}_2 + \cdots + \omega_n\tilde{x}_n = b\} \ge a_3$ اگر $\omega_1(\tilde{x})_{a_3}^L + \omega_2(\tilde{x})_{a_3}^L + \cdots + \omega_n(\tilde{x})_{a_3}^L \le b$ و $\omega_1(\tilde{x})_{a_3}^L + \omega_2(\tilde{x})_{a_3}^L + \cdots + \omega_n(\tilde{x})_{a_3}^L \ge b$.

اثبات فقط دلیل مورد اول ارائه خواهد شد. موارد دیگر را می توان با استفاده از استدلال های مشابه اثبات کرد.

$Pos\{\omega_1\tilde{x}_1 + \omega_2\tilde{x}_2 + \cdots + \omega_n\tilde{x}_n \le b\} \ge a_1$، فرض کنید

$(s_1^*, \dots, s_n^*) =^{\Delta} arg \sup_{s_1, \dots, s_n \in R} \{\min [\mu_{\tilde{x}_1}(s_1), \dots, \mu_{\tilde{x}_n}(s_n)] | \omega_1 s_1^* + \omega_2 s_2^* + \cdots + \omega_n s_n^* \le b\}$.

آنگاه $\min\{\mu_{\tilde{x}_1}(s_1), \dots, \mu_{\tilde{x}_n}(s_n)\} \ge a_1$ و $\omega_1 s_1^* + \omega_2 s_2^* + \cdots + \omega_n s_n^* \le b$.

$\min\{\mu_{\tilde{x}_1}(s_1), \dots, \mu_{\tilde{x}_n}(s_n)\} \ge a_1$ به $\mu_{\tilde{x}_1}(s_1^*) \ge a_1, \dots, \mu_{\tilde{x}_n}(s_n^*) \ge a_1$ اشاره دارد، بنابراین $s_1^* \in \left[(\tilde{x}_1)_{a_1}^L, (\tilde{x}_1)_{a_1}^U\right]$ ، $s_2^* \in \left[(\tilde{x}_2)_{a_1}^L, (\tilde{x}_2)_{a_1}^U\right]$ ، ... و $s_n^* \in \left[(\tilde{x}_n)_{a_1}^L, (\tilde{x}_n)_{a_1}^U\right]$ علاوه بر این $\omega_1 s_1^* + \omega_2 s_2^* + \cdots + \omega_n s_n^* \le b$ و نا منفی بودن از $\omega_i (i = 1,2, \dots, n)$، بدین ترتیب $\omega_1(\tilde{x})_{a_1}^L + \omega_2(\tilde{x})_{a_1}^L + \cdots + \omega_n(\tilde{x})_{a_1}^L \le b$.

متقابلا، اگر $\omega_1(\tilde{x})_{a_1}^L + \omega_2(\tilde{x})_{a_1}^L + \cdots + \omega_n(\tilde{x})_{a_1}^L \le b$، آنگاه a_1' ای وجود دارد که $a_1 \le a_1' \le 1$ به صورتی که $\omega_1(\tilde{x})_{a_1}^L + \omega_2(\tilde{x})_{a_1}^L + \cdots + \omega_n(\tilde{x})_{a_1}^L \le b$، بدیهی است که $\mu_{\tilde{x}_1}((\tilde{x}_1)_{a_1'}^L) \ge a_1, \dots, \mu_{\tilde{x}_n}((\tilde{x}_n)_{a_1'}^L) \ge a_1$ ، بنابراین $\min\left\{\mu_{\tilde{x}_1}((\tilde{x}_1)_{a_1'}^L\right) \ge a_1, \dots, \mu_{\tilde{x}_1}((\tilde{x}_1)_{a_1'}^L) \ge a_1 | \omega_1(\tilde{x}_1)_{a_1}^L + \omega_2(\tilde{x}_2)_{a_1}^L + \cdots + \omega_n(\tilde{x}_n)_{a_1}^L \le b\} \ge a_1$ در نتیجه

$$Pos\{\omega_1\tilde{x}_1 + \omega_2\tilde{x}_2 + \cdots + \omega_n\tilde{x}_n \le b\} = \sup_{s_1, \dots, s_n \in R} \min \{\mu_{\tilde{x}_1}(s_1), \dots, \mu_{\tilde{x}_n}(s_n) | s_1 + s_2 + \cdots + s_n \le b\} \ge a_1$$

۳ روش شناسی

طبقه بندی داده ها برای قضاوت در مورد اینکه آیا داده ها به یک گروه مشخص با توجه به برخی ویژگی های مشاهده شده تعلق دارند یا خیر. به عبارت دیگر، طبقه بندی داده ها برای ساختن یک تابع متمایز با استفاده از مجموعه ای از داده های از پیش انتخاب شده، به نام مجموعه آموزشی نمونه، و سپس آزمایش برچسب کلاس برای یک نمونه جدید است.

طبقه بندی کننده مبتنی بر DEA فازی استفاده از مدل DEA فازی برای ساخت چنین تابع متمایزی است.

۳. ۱ طبقه بندی کننده مبتنی بر DEA فازی افزایشی

در تجارت، مجموعه ویژگی های داده شده یک داده فازی افزایشی است، یعنی قاعده ((بزرگتر بهتر است)) اعمال می شود را برآورده می کند. این نشان می دهد که مقادیر ویژگی های بزرگتر اولویت شی بیشتری دارند.

مجموعه نمونه آموزش فازی $S = \left\{(\tilde{X}_1, y_1), (\tilde{X}_2, y_2), \dots, (\tilde{X}_l, y_l)\right\}$ را در نظر بگیرید، جایی که $\tilde{X}_j \in T^n(R)$

$j = 1,2, \dots, l$, $y_j \in \{-1,1\}$. اگر $y_i = 1$، آنگاه $(\tilde{X}_i, y_i)$ یک کلاس مثبت نامیده می شـــود. . اگر $y_i = -1$، آنگاه $(\tilde{X}_i, y_i)$ یک کلاس منفی نامیده می شود. بدون از دست دادن کلیت، مجموعه نمونه آموزشی فازی مثبت با

$S^+ = \{(\tilde{X}_1, y_1), (\tilde{X}_2, y_2), \dots, (\tilde{X}_l, y_l)\}$ نشان داده می شود و

$S^- = \{(\tilde{X}_{l_1+1}, y_{l_1+1}), (\tilde{X}_{l_1+2}, y_{l_1+2}), \dots, (\tilde{X}_l, y_l)\}$ مجموعه نمونه آموزشی فازی منفی است.

طبقه بندی بر اســاس مجموعه آموزشــی فازی $S = \{(\tilde{X}_1, y_1), (\tilde{X}_2, y_2), \dots, (\tilde{X}_l, y_l)\}$ برای یافتن یک تابع تصـمیم گیری $g(\tilde{X})$ اسـت، به طوری که کلاس مثبت و کلاس منفی را می توان با خطای طبقه بندی کم و عملکرد تعمیم خوب از هم جدا کرد.

برای ســادگی، فرض می کنیم که داده های آموزش فازی $\tilde{X}_k = (\tilde{x}_{1k}, \tilde{x}_{2k}, \dots, \tilde{x}_{nk})(k = 1,2, \dots, n)$ بردارهای اعداد فازی مثلثی هســتند، $\tilde{x}_{ik} = (l_{ik}, m_{ik}, r_{ik})$ و $i = 1,2 \dots, n$ ؛ $k = 1,2, \dots, l$.

تعریف ۱.۳ برای مجموعه آموزش فازی $S = S^+ \cup S^-$ و داده شـــده و یک ســـطح احتمال $a(0 \le a \le 1)$ ، در صـــورت وجود بردار مثبت $\omega \ge 0$ و عدد μ مثبت به طوری که $Pos\{\omega.\tilde{x}_i - \mu > 0\} \ge a$ به ازای $\tilde{x}_i \in S^+$، سپس مجموعه آموزشی فازی را با توجه به سطح امکان α افزایشی می گویند. ابر صفحه فازی $\omega.\tilde{x}_i - \mu \cong \tilde{0}$ یک ابر صفحه طبقه بندی فازی در سطح α نامیده می شود.

برای مجموعه آموزش فازی تفکیک پذیر افزایشی $S = S^+ \cup S^-$ دامنه پذیرش را می توان به صورت زیر نشان داد:

$$\Theta = \left\{ \tilde{x} \middle| \sum_{i=1}^{l_1} \lambda_i \tilde{x}_i \tilde{<} \tilde{x}, \tilde{x}_i \in S^+, \sum_{i=1}^{l_1} \lambda_i \ge 1, \lambda_i \ge 0 \right\}$$

مدل فازی CCR با ورودی-خروجی $(\tilde{X}_i, 1)$برای DMU_j، $j = 1,2, \dots, l_1$, برای ارزیابی DMU_{j_0} توسط زیر داده شده است:

$$\begin{cases} \max \mu_0 \\ \omega \tilde{x}_i - \mu_0 \tilde{>} 0, i = 1,2, \dots, l_1; \\ \omega \tilde{x}_0 \cong 1; \\ \omega \ge 0, \mu_0 \ge 0. \end{cases} \tag{5}$$

برای حل مدل فازی فوق (۵)، از رویکرد امکان [۳۰، ۳۱] اسـتفاده می کنیم. برای سـطح امکان داده شـده α، $\beta(0 \le \alpha, \beta \le 1)$ ، مدل امکان (۵) به شرح زیر است:

$$\begin{cases} \max \mu_0 \\ pos\left\{ \omega \tilde{x}_i - \mu_0 \ge 0 \right\} \ge \alpha, i = 1,2, \dots, l_1; \\ pos\left\{ \omega \tilde{x}_0 = 1 \right\} \ge \beta; \\ \omega \ge 0, \mu_0 \ge 0. \end{cases} \tag{6}$$

برای سادگی، $\tilde{x}_{j_0} = \tilde{x}_0$ فرض کنید $\alpha = \beta$. از لم ۱، مسئله بهینه سازی معادل مدل زیر است:

$$\begin{cases} \max \mu_0 \\ \omega(\tilde{x}_i)_\alpha^U - \mu_0 \geq 0, i = 1,2,...,l_1; \\ \omega(\tilde{x}_0)_\alpha^L \leq 1, \omega\left(\tilde{x}_0\right)_\alpha^U \geq 1; \\ \omega \geq 0, \mu_0 \geq 0. \end{cases} \tag{7}$$

تعریف ۳.۲ برای $DMU_{j_0}: (\tilde{x}_{j_0}, 1) \in S^+$ ارزیابی شده، اگر مقادیر بهینه (۶) کمتر از ۱ نباشد، $(\tilde{x}_{j_0}, 1)$ نقطه طبقه بندی کارآمد S^+ نامیده می شود.

برای سطح امکان داده شده α، $\beta(0 \leq \alpha, \beta \leq 1)$ و یک نقطه طبقه بندی کارآمد از S^+، با مدل (۷) (یا (۶) برای $\alpha = \beta$)، می توانیم بردار نرمال کارآمد را به دست آوریم. سطح مرزی دامنه پذیرش فازی Θ.

اگر مجموعه نمونه آموزشی $S = \{(\tilde{X}_1, y_1), (\tilde{X}_2, y_2), \dots, (\tilde{X}_l, y_l)\}$ یک مجموعه نمونه آموزشی کلاسیک باشد، به عبارتی $\tilde{X}_j \in R^n, j = 1,2, \dots, l$، طبقه‌بندی‌کننده مبتنی بر DEA فازی به ماشین طبقه‌بندی پوششی داده‌ها تبدیل می‌شود [۲۸].

قضیه ۳ اگر $(\tilde{x}_{j_0}, 1)$ یک نقطه طبقه بندی کارآمد S^+ است، $\omega = \omega_0$ ، μ_0 راه حل بهینه (۶) است، سپس برای هر $\tilde{x} \in \Theta$; $Pos\{\omega_0 \tilde{x} - \mu_0 \geq 0\} \geq \alpha$.

اثبات چون $\omega_0 > 0$، μ_0 جواب بهینه (۶) هستند، بنابراین $Pos\{\omega_0 \tilde{x}_i - \mu_0 \geq 0\} \geq \alpha$ ، $i = 1,2, \dots, l$ و $pos\{\omega_0 \tilde{x}_i - \mu_0 \geq 0\} \geq a$؛ توسط لم ۲، برای هر $\tilde{x} \in \Theta$ ،

$$\omega_0 \tilde{x}_i - \mu_0 \gtrsim \omega_0 \textstyle\sum_{i=1}^{l_1} \lambda_i \tilde{x}_i - \mu_0 = \sum_{i=1}^{l_1} \lambda_i (\omega_0 \tilde{x}_i - \mu_0) + \mu_0 (\sum_{i=1}^{l_1} \lambda_i - 1) \gtrsim \sum_{i=1}^{l_1} \lambda_i (\omega_0 \tilde{x}_i - \mu_0)$$

و توسط نظریه ۲، $Pos\{\omega_0 \tilde{x}_i - \mu_0 > 0\} \geq a$ اگر و فقط اگر $\omega_0 (\tilde{x}_i)_a^U - \mu_0 > 0$ برای $i = 1,2, \dots, l_1$. بنابراین $\sum_{i=1}^{l_1} \lambda_i [\omega_0 (\tilde{x}_i)_a^U - \mu_0] \geq 0$، که آن برابر است با، $\omega_0 \sum_{i=1}^{l_1} \lambda_i\, (\tilde{x}_i)_a^U - \mu_0 \sum_{i=1}^{l_1} \lambda_i] \geq 0$ آنگاه برای همه $\tilde{x} \in \Theta$ ، $(\omega_0 \tilde{x}_i - \mu_0)_a^U = \omega_0 (\tilde{x}_i)_a^U - \mu_0 \geq \omega_0 \sum_{i=1}^{l_1} \lambda_i\, (\tilde{x}_i)_a^U - \mu_0 \geq \omega_0 \sum_{i=1}^{l_1} \lambda_i\, (\tilde{x}_i)_a^U - \mu_0 \sum_{i=1}^{l_1} \lambda_i = \sum_{i=1}^{l_1} \lambda_i [\omega_0 (\tilde{x}_i)_a^U - \mu_0]$، همچنین توسط نظریه۲، $Pos\{\omega_0 (\tilde{x}_i)_a^U - \mu_0 \geq 0\} \geq a$.

برای در نظر گرفتن اختلال داده‌های آموزشی یا این واقعیت که برخی از نقاط داده ممکن است اشتباه طبقه‌بندی شوند، بردار متغیرهای ضعیف $\xi = (\xi_1, \dots, \xi_{l_1})^T$ را معرفی می‌کنیم. که میزان نقض محدودیت ها را اندازه گیری می کند، مدل CCR فازی مربوطه (۵) به شرح زیر است:

$$\begin{cases} \max \mu_0 - C \sum_{i=1}^{l_1} \xi_i \\ \omega \tilde{x}_i - \mu_0 \tilde{\geq} 0 + \xi_i, i = 1,2,...,l_1; \\ \omega \tilde{x}_0 \cong 1; \\ \xi_i \geq 0; \\ \omega \geq 0, \mu_0 \geq 0. \end{cases} \tag{8}$$

که در آن $C > 0$ از قبل مشخص شده است. برای سطح امکان داده شده α، $\beta(0 \leq \alpha, \beta \leq 1)$ ، مدل امکان (۸) به شرح زیر است:

$$\begin{cases} \max \mu_0 - C\sum_{i=1}^{l_1}\xi_i \\ pos\left\{\omega\tilde{x}_i - \mu_0 \geq 0 + \xi_i\right\} \geq \alpha, i = 1,2,\ldots,l_1; \\ pos\left\{\omega\tilde{x}_0 = 1\right\} \geq \beta; \\ \xi_i \geq 0 \\ \omega \geq 0, \mu_0 \geq 0. \end{cases} \tag{9}$$

همچنین از لم ۱، برای $\beta = \alpha$، مسئله بهینه سازی (۹) معادل مدل بهینه سازی کلاسیک زیر است:

$$\begin{cases} \max \mu_0 - C\sum_{i=1}^{l_1}\xi_i \\ \omega(\tilde{x}_i)_\alpha^U - \mu_0 \geq 0 + \xi_i, i = 1,2,\ldots,l_1; \\ \omega(\tilde{x}_0)_\alpha^L \leq 1, \omega(\tilde{x}_0)_\alpha^U \geq 1; \\ \xi_i \geq 0; \\ \omega \geq 0, \mu_0 \geq 0. \end{cases} \tag{10}$$

اگر $\omega_0 > 0$، $\mu_0 = 1$ جواب بهینه (۷) (یا (۱۰)) باشد و $\tilde{x}_0$ ، $pos\{\omega_0\tilde{x}_0 = 1\} \geq a$ را برآورده کند.

مشخص کن $L: \omega_0\tilde{x} - 1 \cong 0$.

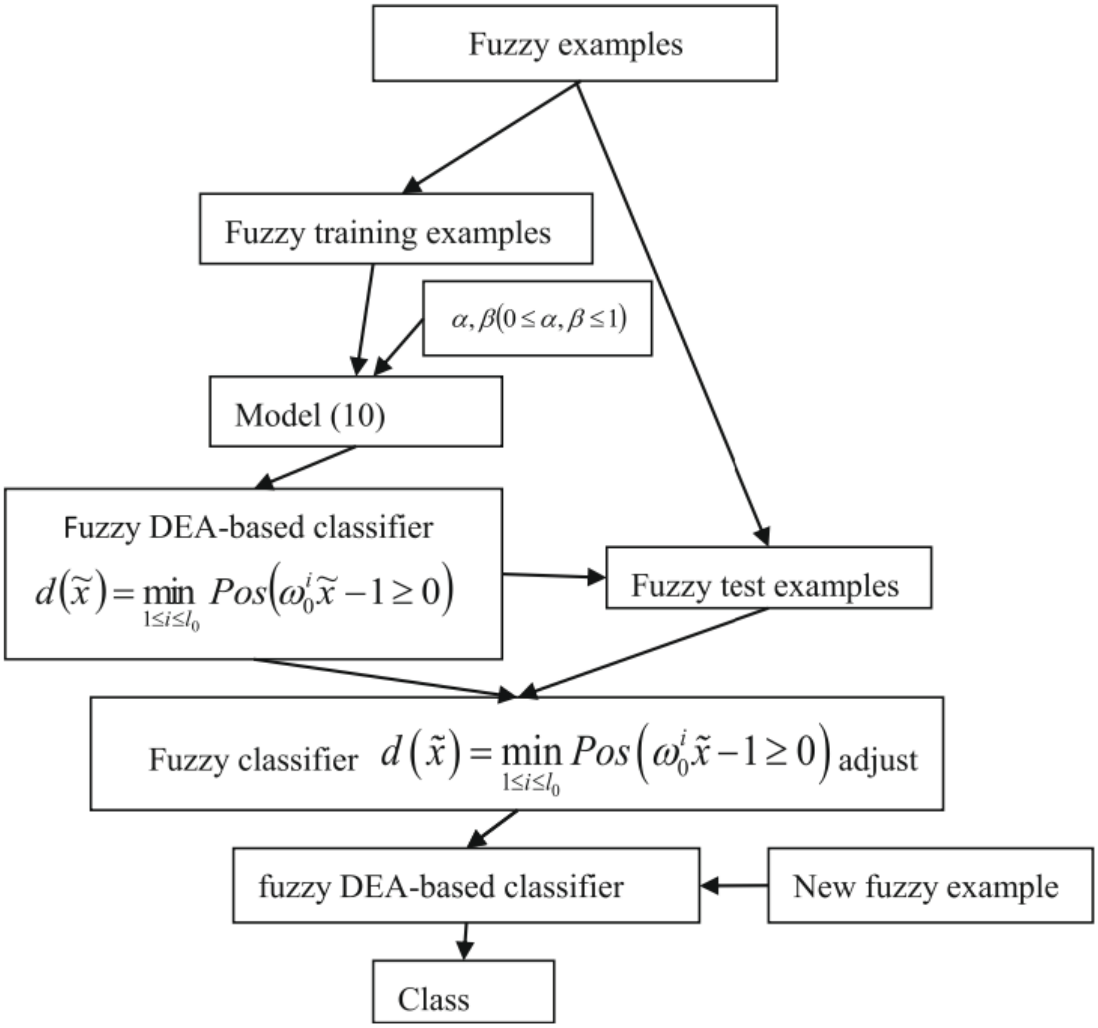
ابر صفحه L یک صفحه حمایت کننده از حوزه پذیرش است، ω_0 بردار نرمال ابر صفحه L و $\tilde{x}_0$ در این صفحه است.

فرض کنید تمام صـفحه پشـتیبان مجموعه آموزشـی فازی $S = S^+ \cup S^-$ با $\omega_0^j\tilde{x} - 1 \cong 0$ ، $j = 1,2,\ldots,l_0(l_0 < l_1)$ نشـان داده شـود، سپس تابع فازی طبقه بندی را می توان با $d(\tilde{x}) = \min_{1\leq i\leq l_0} Pos(\omega_0^i\tilde{x} - 1 \geq 0)$ ارائه داد.

قضـاوت در مورد اینکه آیا یک داده مفروض x به حوزه پذیرش فازی Θ تعلق دارد، معادل تشـخیص اینکه آیا x بالاتر از همه ابر صـفحه فازی اسـت (برای یک سطح اطمینان معین) است.

برای مثال فازی که کلاس ناشـناخته $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$، قانون تصـمیم گیری این اسـت: برای یک سـطح اطمینان معین $\alpha(0 < \alpha \leq 1)$ ، اگر $d(\tilde{x}) \geq a$، پس $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ یک مثال مثبت است، در غیر این صورت $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ یک مثال منفی است.

نمودار طبقه بندی کننده مبتنی بر DEA فازی افزایشی در شکل ۱ نشان داده شده است.



۳.۲ کاربردها در مدیریت مراقبت های بهداشتی

طبقه‌بندی‌کننده مبتنی بر DEA فازی پیشنهادی ما را می‌توان در مدیریت مراقبت‌های بهداشتی، مانند تشخیص بیماری‌ها، پیش‌بینی بیماری‌ها و نظارت بر سلامت افراد یا هشدار سلامت به کار برد.

در ادامه، طبقه‌بندی‌کننده فازی مبتنی بر DEA در تشخیص بیماری کرونر با علائم فازی و در پیش‌بینی سرطان پستان با داده‌های اختلال فازی استفاده می‌شود. برنامه ها همچنین دقت و کارایی طبقه بندی کننده مبتنی بر DEA فازی پیشنهادی را اثبات می کنند.

مثال ۱ تشخیص کرونری [۴۸]. ۳۰ نمونه به طور تصادفی به دو گروه ۲۴ نمونه تمرین و ۶ نمونه آزمایش تقسیم می شوند که داده های تمرینی جدول ۱ فشار دیاستولیک ($x^{\sim}\sim_{i1}$) و کلسترول پلاسما ($\tilde{x}_{i2}$) بیست و چهار نفر است که نیمی از آنها سالم هستند ($y_i = 1$)، بقیه بیماران کرونر($y_i = -1$)، $\tilde{x}_{i1}$ و $\tilde{x}_{i2}$ اعداد فازی مثلثی هستند.

نمونه ها را به طور تصادفی به داده های آموزشی (۲۴ نمونه) و داده های آزمون (۶) تقسیم می کنیم. بر اساس داده های آموزش فازی، برای پارامتر $C = 0.1$ ، $a = 0.95$ ، حل برنامه نویسی(۱۱) ، ما می توانیم چهار نقطه طبقه بندی کارآمد و صفحات پشتیبانی پاسخ دهنده آنها را به دست آوریم.

Table 1 The data of diastolic pressure and plasma cholesterol of the patient of Coronary and healthy people

i	$\tilde{x}_{i1}$ (KPa)	$\tilde{x}_{i2}$ (mmol/L)	y_i	i	$\tilde{x}_{i1}$	$\tilde{x}_{i2}$	y_i
1	(9.84,9.86,9.88)	(5.17,5.18,5.19)	1	16	(10.62,10.66,10.70)	(2.06,2.07,2.08)	−1
2	(13.31,13.33,13.35)	(3.72,3.73,3.74)	1	17	(12.51,12.53,12.55)	(4.44,4.45,4.46)	−1
3	(14.63,14.66,14.69)	(3.87,3.89,3.91)	1	18	(13.30,13.33,13.36)	(3.04,3.06,3.08)	−1
4	(9.32,9.33,9.34)	(7.08,7.10,7.12)	1	19	(9.32,9.33,9.34)	(3.90,3.94,3.98)	−1
5	(12.87,12.80,12.83)	(5.47,5.49,5.51)	1	20	(10.64,10.66,10.68)	(4.43,4.45,4.47)	−1
6	(10.64,10.66,10.68)	(4.06,4.09,4.12)	1	21	(10.64,10.66,10.68)	(4.89,4.92,4.95)	−1
7	(10.65,10.66,10.67)	(4.43,4.45,4.47)	1	22	(9.31,9.33,9.35)	(3.66,3.68,3.70)	−1
8	(13.31,13.33,13.35)	(3.60,3.63,3.66)	1	23	(10.64,10.66,10.68)	(3.20,3.21,3.22)	−1
9	(13.32,13.33,13.34)	(5.68,5.70,5.72)	1	24	(10.37,10.40,10.43)	(3.92,3.94,3.96)	−1
10	(11.97,12.00,12.03)	(6.17,6.19,6.21)	1	25	(9.31,9.33,9.35)	(4.90,4.92,4.94)	−1
11	(14.64,14.66,14.68)	(4.00,4.01,4.02)	1	26	(11.19,11.20,11.21)	(3.40,3.42,3.44)	−1
12	(13.31,13.33,13.35)	(3.99,4.01,4.03)	1	27	(9.31,9.33,9.35)	(3.62,3.63,3.64)	−1
13	(12.72,12.80,12.88)	(5.93,5.96,5.99)	1	28	(10.64,10.66,10.68)	(4.43,4.45,4.47)	−1
14	(13.3,13.33,13.36)	(5.88,5.96,6.04)	1	29	(10.64,10.66,10.68)	(2.65,2.69,2.73)	−1
15	(10.63,10.66,10.69)	(5,5.02,5.04)	−1	30	(10.61,10.66,10.71)	(2.71,2.77,2.83)	−1

$$L_1 : f_1\left(\tilde{x}\right) = 0.0886786\tilde{x}_1 + 0.02450144\tilde{x}_2 - 1 \cong 0;$$
$$L_2 : f_2\left(\tilde{x}\right) = 0.08856158\tilde{x}_1 + 0.02496741\tilde{x}_2 - 1 \cong 0;$$
$$L_3 : f_3\left(\tilde{x}\right) = 0.02909458\tilde{x}_1 + 0.1688751\tilde{x}_2 - 1 \cong 0;$$
$$L_4 : f_4\left(\tilde{x}\right) = 0.275824\tilde{x}_2 - 1 \cong 0.$$

سپس تابع تفکیک به صورت زیر داده می شود:

$$d(\tilde{x}) = \min_{1 \le i \le 4} Pos(f_i(\tilde{x}) - 1 \ge 0)$$

قانون تصـمیم گیری این اسـت: برای یک سـطح اطمینان معین $a = 0.95$، اگر $d(\tilde{x}) = \min_{1 \le i \le 4} Pos(f_i(\tilde{x}) - 1 \ge 0) \ge 0.95$ ، آنگاه $\tilde{X} = (\tilde{x}_1, \tilde{x}_2)$ یک مثال مثبت است. در غیر این صورت $\tilde{X} = (\tilde{x}_1, \tilde{x}_2)$ یک مثال منفی است.

برای همان مجموعه داده های فازی و سطح اطمینان، ما با ماشین بردار پشتیبان فازی خطی (LFSVM) [۴۷] مقایسه می کنیم، نتیجه آزمایش در جدول ۲ آورده شده است.

Table 2 Comparison with LFSVM

Classifiers	parameter	Training accuracies	Test accuracies	training times(seconds)
FDEACM	$C = 0.1$, $\alpha = 0.95$	95.8%	100%	240
LFSVM	$C = 0.1$, $\alpha = 0.95$	91.7%	83.3%	27

در مقایسـه با LFSVM، ماشـین طبقه‌بندی فازی DEA پیشـنهادی ما دقت آموزشـی و آموزش تسـت بهتری دارد، اما زمان محاسـبات طولانی‌تر اسـت. از آنجایی که مجموعه داده های فازی به صورت خطی قابل تفکیک نیستند، دقت آموزشی و دقت آزمون LFSVM بدترین هستند.

مثال۲ در این مثال، مجموعه داده را از مخزن دانشـگاه کالیفرنیا اروین (UCI) انتخاب می کنیم: مجموعه داده سـرطان پسـتان پیش آگهی ویسـکانسـین (Bennett و Mangasarian ۱۹۹۲) (http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%۲۸Original%۲۹) . مجموعه داده های سـرطان

پستان پیش آگهی ویسکانسین دارای ۶۹۹ رکورد و ۹ متغیر است (ضخامت توده، یکنواختی اندازه سلول، یکنواختی شکل سلول، چسبندگی حاشیه ای، اندازه سلول اپیتلیال منفرد، هسته های لخت، کروماتین ملایم، هسته طبیعی، میتوز). مجموعه داده شامل ۱۶ نمونه است که دارای ویژگی هایی با مقادیر گم شده و ۶۸۳ نمونه دارای داده های کامل هستند، این رکوردها به دو کلاس خوش خیم یا بدخیم تعلق دارند. به منظور تأیید استحکام طبقه‌بندی‌کننده مبتنی بر DEA فازی پیشنهادی ما، به‌طور تصادفی یک اختلال فازی را به هر مقدار مشخصه اضافه می‌کنیم، سپس می‌توانیم مجموعه نمونه‌های فازی را بدست آوریم.

طبقه بدخیم به عنوان کلاس مثبت در نظر گرفته می شود که از بین ۲۳۹ نمونه مثبت ۱۸۰ نمونه را به صورت تصادفی به عنوان نمونه آموزشی و بقیه نمونه های مثبت و تمامی نمونه های منفی را به عنوان نمونه آزمایش انتخاب می کنیم. هنگامی که پارامتر $\alpha = 0.85$ ، $C = 0.1$، حل برنامه نویسی (۱۱). ما می‌توانیم ۳۷ نقطه طبقه‌بندی کارآمد و صفحه‌های پشتیبانی پاسخگوی آنها را به دست آوریم، سپس تابع تفکیک را داریم:

$$d\left(\tilde{x}\right) = \min_{1 \le i \le 37} Pos\left(\omega_0^i \tilde{x} - 1 \ge 0\right)$$

مدل پیشنهادی، که بر روی مجموعه داده‌های اختلال فازی فوق آموزش داده شده است، با بردار پشتیبان آموزش‌دیده بر روی مجموعه داده سرطان پستان پیش‌آگهی اصلی ویسکانسین مقایسه می‌شود.

(http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+٪۲۸Original٪۲۹)، نتیجه در جدول ۳ آورده شده است.

نتایج آزمایش‌ها نشان می‌دهد که اگرچه مدل پیشنهادی دقت آموزشی کمتر و زمان‌های آموزشی طولانی‌تری دارد، اما دقت تست بهتری نسبت به SVM دارد. با این حال، نتایج مدل پیشنهادی تحت اختلال فازی به مجموعه داده ها به دست می آید. این آزمایش‌های عددی نشان می‌دهد که ماشین طبقه‌بندی فازی DEA پیشنهادی قوی است.

Table 3 Comparison result of Robust FDEACM with SVM

Classifiers	parameter	Training accuracies	Test accuracies	training times (minute)
FDEACM	$C = 0.1$, $\alpha = 0.85$	97.8%	96.8%	12
SVM	$C = 0.1$,	98.7%	96.3%	2.3

٤ نتیجه گیری و تحقیقات آتی

در این مقاله، در راستای ایده هونگ یان و همکاران. [۲۹]، ما یک ماشین طبقه بندی DEA را برای داده های آموزش فازی مورد بحث و تعریف قرار دادیم. ما داده های فازی تحت ارزیابی را به عنوان یک واحد تصمیم گیری با مقادیر مشخصه فازی داده شده به عنوان ورودی و یک خروجی واحد با مقدار ۱ در نظر می گیریم. سپس از مجموعه ای از واحدهای تصمیم گیری (DMUs) برای تشکیل یک مجموعه داده نمونه فازی استفاده کردیم. دامنه پذیرش فازی را بر اساس مجموعه نمونه فازی برای طبقه بندی بسازید. طبقه‌بندی‌کننده مبتنی بر DEA فازی از توابع تفکیک فازی تکه‌ای خطی برای تقریب توابع تفکیک فازی غیرخطی استفاده می‌کند. آزمایشات نشان می دهد که مدل پیشنهادی عملکرد خوبی دارد. کار نظری و نتایج محاسباتی نشان می‌دهد که روش توسعه‌یافته پتانسیل بالایی در عمل دارد. اما مدل پیشنهادی دارای اشکالاتی است، (۱) مدل پیشنهادی فقط برای طبقه‌بندی داده‌های فازی افزایشی قابل استفاده است. (۲) زمان آموزش مدل پیشنهادی طولانی تر است، که برای داده های بزرگ مناسب نیست. در آینده مدل طبقه بندی خود را در جهت های زیر تعمیم خواهیم داد. (۱) بردارهای نرمال ابرصفحه ها مثبت هستند و فقط از اطلاعات یک کلاس استفاده می کند [۲۹]، اطلاعات کلاس دیگر از بین می رود. همه اینها ممکن است منجر به عملکرد بد طبقه بندی شود. ممکن است یک طبقه‌بندی‌کننده جدید بر اساس DEA فازی تعمیم‌یافته بسازیم، که در آن شرایط غیرمنفی را می‌توان از دست داد و کابین اطلاعات طبقه‌بندی را به طور کامل مورد استفاده قرار داد. (۲) مدل پیشنهادی همه ویژگی‌ها را با اهمیت یکسان در نظر می‌گیرد، اما در برخی موارد، برخی از ویژگی‌ها از بقیه مهم‌تر هستند. و ما می توانیم ماشین طبقه بندی فازی DEA را با محدودیت های منطقه اطمینان مورد بحث قرار دهیم. (۳) داده های آموزشی مدل پیشنهادی داده های فازی هستند، ما می توانیم ماشین طبقه بندی پوششی داده های تصادفی را بر اساس تحلیل پوششی داده های تصادفی بسازیم [۴۹، ۵۰]. علاوه بر این، ماشین طبقه بندی پوششی داده ها را با نمونه آموزش فازی تصادفی بسازید.

سپاسگزاریها

این کار با کمک مالی از صندوق ملی علوم اجتماعی (۱۴BJY۰۱۰)، صندوق علوم اجتماعی استان هبی (HB۱۸GL۰۱۴) حمایت شد.

یادداشت ناشر

Springer Nature با توجه به ادعاهای قضایی در نقشه های منتشر شده و وابستگی های سازمانی بی طرف باقی می ماند.

منابع

۱. Kim S-P, Gupta D, Ajay K et al (۲۰۱۵) Accept/decline decision module for the liver simulated allocation model. Health Care Manag Sci ۱۸:۳۵–۵۷

۲. Chen H, Cheng B-C, Liao G-T, Kuo T-C (۲۰۱۴) Hybrid classification engine for cardiac arrhythmia cloud service in elderly healthcare management. J Vis Lang Comput ۲۵:۷۴۵–۷۵۳

۳. Mahamune M, Ingle S, Deo P, Chowhan S (۲۰۱۵) Healthcare knowledge management using data mining techniques. Advances in Computational Research ۱:۲۷۴–۲۷۸

۴. Luukka P (۲۰۱۱) A New Nonlinear Fuzzy Robust PCA Algorithm and Similarity Classifier in Classification of Medical Data Sets. International Journal of Fuzzy Systems ۳:۱۵۳–۱۶۳

۵. Ishibushi H, Yamamoto T, Nakashima T (۲۰۰۱) Fuzzy data mining: effect of fuzzy discretization. In: Proc. of the IEEE International Conference on Data Mining, ICDM, San Jose, pp. ۲۴۱–۲۴۸

۶. Guillaume S (۲۰۰۱) Designing fuzzy inference system from data: an interpretability-oriented review. IEEE Trans Fuzzy Syst ۹(۳): ۴۲۶–۴۴۳

۷. Wu K, Yap K-H (۲۰۰۶) Fuzzy SVM for content-based image retrieval. IEEE Comput Intell Mag ۱(۲):۱۰–۱۶

۸. Yuan Y, Shaw M (۱۹۹۵) Induction of fuzzy decision trees. FuzzySets Syst ۶۹:۱۲۵–۱۳۹

۹. Boyen X, Wehenkel L (۱۹۹۹) Automatic induction of fuzzy decision tree and its application to power system security assessment. Fuzzy Sets Syst ۱۰۲(۱):۳–۱۹

۱۰. Huang Y-P, Lai S-L, Sandnes FE, Liu S-I (۲۰۱۲) Improving Classifications of Medical Data Based on Fuzzy ART۲ DecisionTrees. International Journal of Fuzzy Systems ۳:۴۴۴–۴۵۴

۱۱. Wang L-X, Mendel J (۱۹۹۲) Generating fuzzy rules by learning from examples. IEEE Trans Syst Man Cybern ۲۲(۶):۱۴۱۴–۱۴۲۷

۱۲. Hong T-P, Chen J-B (۲۰۰۰) Processing individual fuzzy attributes for fuzzy rule induction. Fuzzy Sets Syst ۱۱۲(۱):۱۲۷–۱۴۰

۱۳. Hühn J, Hüllermeier E (۲۰۰۹) FR۳: a fuzzy rule learner for inducing reliable classifiers. IEEE Trans Fuzzy Syst ۱۷(۱):۱۳۸–۱۴۹

۱۴. Wu X-H, Zhou J-J (۲۰۰۶) Fuzzy discriminant analysis with kernel methods. Pattern Recogn ۳۹(۱۱):۲۲۳۶–۲۲۳۹

۱۵. Graves D, Pedrycz W (۲۰۱۰) Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study. Fuzzy Sets Syst ۱۶۱(۴):۵۲۲–۵۴۳

۱۶. Ji A-b, Pang J-h, Qiu H-j (۲۰۱۰) Support vector machine for classification based on fuzzy training data. Expert Syst Appl ۳۷(۴): ۳۴۹۵–۳۴۹۸

۱۷. Heo G, Gader P (۲۰۱۱) Robust kernel discriminant analysis using fuzzy memberships. Pattern Recogn ۴۴(۳):۷۱۶–۷۲۳

۱۸. Baklouti R, Mansouri M, Nounou M, Nounou H, Hamida AB (۲۰۱۵) Iterated Robust kernel Fuzzy Principal Component Analysis and application to fault detection. J Comput Sci ۲

۱۹. Lorence DP, Spink A (۲۰۰۳) Assessment of preferences for classification detail in medical information: is uniformity better? Inf Process Manag ۳۹:۴۶۵–۴۷۷

۲۰. Kulldorff M, Fang Z, Walsh SJ A Tree-Based Scan Statistic for Database Disease Surveillance. Biometrics ۵۹(۲۰۱۳):۳۲۳–۳۳۱

۲۱. J. Han, M. Kamber, Data Mining: Concepts and Techniques Morgan Kaufman Publishers, Inc. San Francisco, ۲۰۰۱

۲۲. Charnes A, Cooper WW, Rhodes E (۱۹۷۸) Measuring the efficiency of decision making units. Eur J Oper Res ۲:۴۲۹–۴۴۴

۲۳. Banker RD, Charnes A, Cooper WW (۱۹۸۴) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci ۳۰(۹):۱۰۷۸–۱۰۹۲

۲۴. Narci HO, Ozcan YA et al (۲۰۱۵) An Examination of Competition and Efficiency for Hospital industry in Turkey. Health Care Management Science (۴):۴۰۷–۴۱۸

۲۵. Nayar P, Ozcan YA, Yu F, Nguyen AT (۲۰۱۳) Data Envelopment Analysis: A Benchmarking Tool for Performance in Urban Acute Care Hospitals. Health Care Manag Rev ۲:۱۳۷–۱۴۵

۲۶. Ozcan YA (۲۰۱۴) Health Care Benchmarking and Performance Evaluation: An Assessment using Data Envelopment Analysis (DEA) ۲nd Edition. Springer, Newton

۲۷. Troutt MD, Rai A, Zhang A (۱۹۹۶) The potential use of DEA for credit applicant acceptance systems. Comput Oper Res ۲۳(۴):۴۰۵–۴۰۸

۲۸. Hasan Bal HHO (۲۰۰۷) Data envelopment analysis approach to two-group classification problem and experimental comparison with some classification models. Hacettepe Journal of Mathematics and Statistics ۳۶(۲):۱۶۹–۱۸۰

۲۹. Yan H, Wei Q (۲۰۱۱) Data envelopment analysis classification machine. Inf Sci ۱۸۱:۵۰۲۹–۵۰۴۱

۳۰. Wei Q, Chang T–S, Han S (۲۰۱۴) Quantile–DEA classifiers with interval data. Ann Oper Res ۲۱۷:۵۳۵–۵۶۳

۳۱. Hatami–Marbini A, Emrouznejad A, Tavana M (۲۰۱۱) A taxonomy and review of the fuzzy Data Envelopment Analysis literature: Two decades in the making. Eur J Oper Res ۲۱۴:۴۵۷–۴۷۲

۳۲. Sengupta JK (۱۹۹۲) A fuzzy systems approach in data envelopment analysis. Computers and Mathematics with Applications ۲۴(۸–۹): ۲۵۹–۲۶۶

۳۳. Hatami–Marbini A, Saati S, Tavana M (۲۰۱۰) An ideal–seeking fuzzy data envelopment analysis framework. Appl Soft Comput ۱۰(۴):۱۰۶۲–۱۰۷۰

۳۴. Kao C, Liu ST (۲۰۰۳) A mathematical programming approach to fuzzy efficiency ranking. Int J Prod Econ ۸۶(۲):۱۴۵–۱۵۴

۳۵. Puri J, Yadav SP (۲۰۱۳) A concept of fuzzy input mix–efficiency in fuzzy DEA and its application in banking. Expert Syst Appl ۴۰(۵): ۱۴۳۷–۱۴۵۰

۳۶. Guo P, Tanaka H (۲۰۰۱) Fuzzy DEA: a perceptual evaluation method. Fuzzy Sets Syst ۱۱۹(۱):۱۴۹–۱۶۰

۳۷. León T, Liern V, Ruiz JL, Sirvent I (۲۰۰۳) A fuzzy mathematical programming approach to the assessment of efficiency with DEA models. Fuzzy Sets Syst ۱۳۹(۲):۴۰۷–۴۱۹

۳۸. Guo P, Tanaka H, Inuiguchi M (۲۰۰۰) Self–organizing fuzzy aggregation models to rank the objects with multiple attributes. IEEETransactions on Systems, Man and Cybernetics, Part A – Systems and Humans ۳۰(۵):۵۷۳–۵۸۰

۳۹. Lertworasirikul S, Shu–Cherng F, Joines JA, Nuttle HLW (۲۰۰۳) Fuzzy data envelopment analysis (DEA): a possibility approach. Fuzzy Sets Syst ۱۳۹(۲):۳۷۹–۳۹۴

۴۰. Muren ZM, Cui W (۲۰۱۴) Generalized fuzzy data envelopment analysis methods. Appl Soft Comput ۱۹:۲۱۵–۲۲۵

۴۱. Ghasemi M–R, Ignatius J, Lozano S, Emrouznejad A, Hatami–Marbini A (۲۰۱۵) A fuzzy expected value approach under generalized data envelopment analysis. Knowl–Based Syst ۸۹:۱۴۸–۱۵۹

۴۲. Tavana M, Shiraz RK, Hatami–Marbini A, Agrell PJ, Paryab K(۲۰۱۳) Chance–constrained DEA models with random fuzzy inputs and outputs. Knowl–Based Syst ۵۲:۳۲–۵۲

۴۳. Angulo Meza L, Pereira Estellita M (۲۰۰۲) Lins, "Review of methods for increasing iscrimination in Data Envelopment Analysis". Ann Oper Res ۱۱۶(۱–۴):۲۲۵–۲۴۲

۴۴. Dotoli M, Epicoco N, Falagario M, Sciancalepore F (۲۰۱۵) A crossefficiency fuzzy Data Envelopment Analysis technique for performance evaluation of Decision Making Units under uncertainty. Comput Ind Eng ۷۹:۱۰۳–۱۱۴

۴۵. Toloo M, Kresta A (۲۰۱۴) Finding the best asset financing alternative: A DEA–WEO approach. Measurement ۵۵:۲۸۸–۲۹۴

۴۶. Toloo M (۲۰۱۳) The most efficient unit without explicit inputs: an extended MILP–DEA model. Measurement ۴۶:۳۶۲۸–۳۶۳۴

۴۷. Ramik J, Rimanek J (۱۹۸۵) Inequality relation between fuzzy numbers and its use in fuzzy optimization. Fuzzy Sets Syst ۱۶:۱۲۳–۱۳۸

۴۸. Ji A–b, Pang J–h, Qiu H–j (۲۰۱۰) Support vector machine for classification based on fuzzy training data. Expert Syst Appl ۳۷:۳۴۹۵–۳۴۹۸

۴۹. Olesen OB, Petersen NC (۲۰۱۶) Stochastic Data Envelopment Analysis– A review. Eur J Oper Res ۲۵۱(۱):۲–۲۱

۵۰. Dotoli M, Epicoco N, Falagario M, Sciancalepore F (۲۰۱۶) A stochastic cross–efficiency Data Envelopment Analysis approach for supplier selection under uncertainty. Int Trans Oper Res ۲۳(۴):۷۲۵–۷۴۸