مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی با نظارت در چارچوب کاهش بُعد بدون اتلاف برای شناسایی چهره

سمیه احمدخانی و پیمان ادیبی

گروه هوش مصنوعی، دانشکدهٔ مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران

چکیده

در این مقاله ابتدا مدل بانظارت روش ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی (SPPCAMM)[1] ارائه شده است؛ سپس با درنظرگرفتن جریمه نگاشت در یادگیری مدل پیش‌گو، روشی برای شناسایی چهره با استفاده از یک رویکرد کاهش بعد بدون اتلاف ارائه شده است. در روش پیشنهادی ابتدا یک خمینه زیربنایی محلی خطی با استفاده از مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی بانظارت از نمونهٔ داده‌ها به دست می‌آید؛ سپس دسته‌بند ماشین بردار پشتیبان با اعمال جریمهٔ نگاشت به‌عنوان مدل پیش‌گوی مذکور با استفاده از این خمینه محلی خطی آموزش داده می‌شود. بدین ترتیب از مزایای کاهش بعد در مدل پیش‌گو استفاده می‌شود؛ و در عین حال جلوی از دست رفتن اطلاعات مفید گرفته می‌شود. برای آموزش و ارزیابی روش پیشنهادی، از پایگاه داده‌های چهرهٔ شناخته‌شده استفاده شده است. روش استخراج ویژگی گابور بر روی تصاویر به کار گرفته شده است. نتایج آزمایش‌ها نشان می‌دهد که روش پیشنهادی نسبت به بسیاری از روش‌های معمول که کاهش بعد را انجام داده و سپس دسته‌بند را آموزش می‌دهند، و همچنین نسبت به روش جریمه نگاشت مبتنی بر مدل‌های کاهش بعد خطی و غیرخطی، دقت بیشتری دارد.

واژگان کلیدی: کاهش بعد بدون اتلاف، مدل ترکیبی، تحلیل مؤلفه اصلی احتمالی، بانظارت، جریمه نگاشت.

۱- مقدّمه

یادگیری یک مدل دسته‌بندی درفضایی با بُعد بالا و با وجود محدودیت در تعداد نمونه‌های آموزشی، کار دشواری است. ازاین‌رو روش‌های کاهش بعد فضای ویژگی به‌منظور به‌سازی یادگیری در این مدل‌ها مورد توجه پژوهش‌گران قرار گرفته‌اند. روش‌های کاهش بعد علاوه‌بر کوچک‌کردن بردار ویژگی، به‌نوعی باعث افزایش دقت دسته‌بندی نیز می‌شوند. از آنجا که بعد فضای ورودی در بسیاری از مسائل بیش از حد لازم، بزرگ است، استفاده از روش‌های کاهش بعد علاوه‌بر کم‌کردن طول بردار ویژگی و کاهش هزینهٔ محاسباتی مراحل بعدی، بالابردن کارایی از طریق مقابله با معضل بعد[1] را نیز در پی خواهد داشت. به همین دلیل این روش‌ها درعمل دارای اهمیت بوده و به‌طور وسیع مورد مطالعه قرار گرفته‌اند (بیشاپ، ۲۰۰۶).

یکی از روش‌های کاهش بُعد محبوب روش تحلیل مؤلفهٔ اصلی[2] (PCA) است، که تجزیهٔ مقادیر منفرد[3] (SVD) را روی ماتریس داده اجرا کرده و زیرفضایی با استفاده از مقادیر ویژه بزرگ ایجاد می‌کند (ون در ماتن و همکاران، ۲۰۰۹). بسیاری از روش‌های کاهش بعد بدون نظارت هستند؛ به‌طوری که تنها روی داده‌های بدون برچسب مشاهده شده ورودی تمرکز می‌کنند؛ اما در مواقعی که مقادیر برچسب‌ها یا خروجی در دسترس است، واردکردن این اطلاعات در نگاشت و رسیدن به یک افکنش بانظارت برای داده ورودی در کاربردهایی مانند دسته‌بندی می‌تواند مفیدتر باشد. بنابراین روش‌های کاهش بعد بانظارت به‌منظور به کارگیری اطلاعات خروجی ارائه شده‌اند (یو و همکاران، ۲۰۰۶).

[1] Curse of Dimensionality

[2] Principal Component Analysis

[3] Singular Value Decomposition (SVD)

روش تحلیل تمایز خطی[1] یا LDA روشی بانظارت است که به‌کمک یک بهینه‌سازی ساده بردار افکنش را طوری به‌دست می‌آورد که فاصله درون‌دسته‌ای نمونه‌های آموزشی کمینه و فاصله بین دسته‌ای آن‌ها بیشینه شود (فیشر، ۱۹۳۸). روش تحلیل تمایز خطی وقتی به‌خوبی کار می‌کند که نمونه داده‌ها جدایی‌پذیر خطی باشند. برای غلبه بر این محدودیت روش تحلیل تمایز هسته[2] (KDA) ارائه شده است (میکا و همکاران، ۱۹۹۹). فرض دیگری در روش تحلیل تمایز نخستین وجود دارد که توزیع احتمال هر دسته را محدود به توزیع گاسی تک‌قله‌ای می‌کند (فیشر، ۱۹۳۸).

روش تحلیل فیشر حاشیه‌ای[3] (MFA) با استفاده از ساختار جانشانی گراف[4] (یان و همکاران، ۲۰۰۷)، سعی در غلبه بر محدودیت‌های روش LDA دارد. در مقایسه با LDA، روش MFA مزایای زیر را دارد: ۱) جهت‌های افکنش در دسترس خیلی بیش‌تر از LDA است، و ۲) هیچ فرضی روی توزیع داده هر دسته وجود ندارد (یان و همکاران، ۲۰۰۷). به‌تازگی روش حفظ همسایگی و جانشانی تمایز حاشیه‌ای[5] (NP-MDE) بر اساس جانشانی خطی گراف و روش MFA پیشنهاد شده است؛ که می‌تواند پراکندگی درون‌دسته‌ای را کمینه و هم‌زمان حاشیهٔ بین دسته‌ای را بیشینه کرده، و به‌علاوه ساختار همسایگی با هر دسته را نیز حفظ کند (لان و همکاران، ۲۰۱۲)].

روش TRIMAP[6] روش بانظارت دیگری برای کاهش بعد است (چن وهمکاران، ۲۰۱۰). این روش که ترکیبی از روش Isomap (ژو و مارتینز، ۲۰۰۶) و تحلیل تمایز خطی است، ابتدا از Isomap برای پیداکردن زیرفضای مناسب استفاده می‌کند؛ اما این روش تابع فاصله اقلیدسی معمولی را برای تعیین فواصل به کار نمی‌برد؛ بلکه از تابع فاصله‌ای استفاده می‌کند که فاصله نمونه‌های متعلق به دسته‌های متفاوت را بیش‌تر از حد واقعی تخمین می‌زند. این کار باعث می‌شود تمایز بیش‌تری بین دسته‌های متفاوت ایجاد شود؛ سپس این مدل از تحلیل تمایز خطی برای ایجاد تمایز بیش‌تر میان دسته‌ها استفاده می‌کند.

برای رفع هم‌زمان مشکلات منتج از فرض‌های روش تحلیل تمایز خطی، روش تحلیل تمایز زیردسته مبتنی بر هسته[7] ارائه شده است (ژو و مارتینز، ۲۰۰۶). در این روش با استفاده از تقسیم هر دسته به تعدادی زیردسته سعی می‌شود تا بر فرض گاسی تک‌قله‌ای‌بودن تمام دسته‌ها غلبه شود؛ اما این روش مجبور است هر دسته را به‌کمک روش‌های خوشه‌بندی به زیرگروه‌هایی تقسیم کند. تحلیل تمایز زیردسته ترکیبی با گام کسری[8] روش بانظارت دیگری برای کاهش بعد محسوب می‌شود (گاللیس و همکاران، ۲۰۱۳)، که در آن مانند تحلیل تمایز زیردسته هر دسته به زیردسته‌هایی تقسیم می‌شود. ویژگی مهم این روش حل مشکل جدایی زیردسته‌ها است که زمانی رخ می‌دهد که رتبه ماتریس فاصلهٔ داخلی درون زیردسته‌ای از تعداد بعد کاهش‌یافته بیش‌تر باشد. نسخه مبتنی بر هسته این روش در (گاللیس و همکاران، ۲۰۱۳) ارائه شده است. در (هان و همکاران، ۲۰۱۳) نیز روشی برای حل مسئله کاهش بعد بانظارت ارائه شده است.

در میان روش‌های کاهش بعد ارائه‌شده تاکنون روش‌های بدون نظارت بسیاری وجود دارد؛ و به‌دلیل این‌که بانظارت‌کردن یک روش می‌تواند نتایج را در کاربردهایی مانند دسته‌بندی بهبود دهد، پژوهش‌هایی در زمینه بانظارت‌کردن برخی روش‌های کاهش بعد بدون نظارت صورت گرفته است. در (لی و گو، ۲۰۰۶) نسخه‌ای بانظارت برای روش Isomap به نام SE-Isomap پیشنهاد شده است. در این روش ماتریس فاصله ژئودسیک[9] هر دسته با توجه به اطلاعات مربوط به آن دسته به‌دست می‌آید؛ و سپس روش مقیاس‌سازی چندبعدی[10] (MDS) بر روی گرافی که با توجه به این ماتریس‌های فاصله به‌دست آمده، اعمال می‌شود. در (چن و لیو، ۲۰۱۱) نیز پیشنهادهایی برای بانظارت‌کردن روش جانشانی محلی خطی[11] (LLE) مطرح شده است.

در این مقاله ابتدا مدل بانظارت روش ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی[12] (PPCAMM) ارائه شده که SPPCAMM نامیده می‌شود. این مدل توسعه‌ای از مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی است که اطلاعات مربوط به برچسب داده را در نگاشت یادگیری‌شده منظور می‌کند؛ سپس این رویکرد بانظارت کاهش بعد محلی خطی SPPCAMM در چارچوبی به نام جریمه نگاشت[13] (ژانگ و

[1] Linear Discriminant Analysis (LDA)
[2] Kernel Discriminant Analysis (KDA)
[3] Marginal Fisher Analysis (MFA)
[4] Graph Embedding
[5] Neighborhood Preserving and Marginal Discriminant Embedding
[6] Tensor-based Riemannian Manifold Distance-Approximating Projection (TRIMAP)
[7] Kernel Sub-class Discriminant Analysis (KSDA)
[8] Fractional step Mixture Subclass Discriminant Analysis (FMSDA)
[9] Geodesic
[10] Multidimensional Scaling
[11] Locally Linear Embedding
[12] Probabilistic Principal Component Analysis
[13] Projection penalties

اشنایدر، ۲۰۱۰) به‌کار گرفته می‌شود. جریمه نگاشت این امکان را فراهم می‌آورد که از کاهش بعد برای بهبود دقت مدل پیش‌گو بدون خطر از دست‌دادن اطلاعات مرتبط استفاده کنیم (ژانگ و اشنایدر، ۲۰۱۰). چنان که در بخش ۴ خواهیم دید، نتایج آزمایش‌ها بر روی پایگاه داده‌های تصاویر چهره، نشان‌دهنده کارایی بالاتر روش پیشنهادی SPPCAMM در چهارچوب جریمه نگاشت نسبت به سایر روش‌های بررسی شده است.

ساختار مقاله به‌صورت زیر است: در بخش ۲ مروری بر کارهای مرتبط خواهیم داشت، و چند روش کاهش بعد بررسی می‌شود. در بخش ۳ روش پیشنهادی یعنی مدل با نظارت روش ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی و به‌کارگیری آن در چارچوب جریمه نگاشت معرفی می‌شود. نتایج پیاده‌سازی‌ها در بخش ۴ گزارش می‌شوند. درنهایت تحلیل نتایج در بخش ۵، و جمع‌بندی مقاله در بخش ۶ ارائه خواهد شد.

۲- مروری بر کارهای مرتبط

در این بخش روش‌های کاهش بعد مرتبط با روش پیشنهادی را مورد بررسی قرار می‌دهیم. این روش‌ها عبارت از روش تحلیل مؤلفهٔ اصلی[1] (PCA)، تحلیل مؤلفهٔ اصلی احتمالاتی[2] (PPCA)، تحلیل مؤلفهٔ اصلی احتمالاتی بانظارت[3] (SPPCA)، و مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی[4] (PPCAMM) (تیپینگ و بیشاپ، ۱۹۹۹-الف) هستند. در ادامه فرض می‌کنیم مجموعه‌ای از N نمونه آموزشی در اختیار است. ورودی nام که با $\mathbf{x}_n \in \mathrm{X}$ نمایش داده می‌شود، توسط یک بردار ویژگی M بعدی مشخص می‌شود ($\mathbf{x}_n \in \mathrm{X} \subseteq \mathrm{R}^M$). برای کاهش بعد، هدف یافتن نگاشتی به‌صورت $\Psi: \mathrm{X} \to \mathrm{Z}$ است که هر ورودی را به یک فضای K بعدی با شرط $K<M$، نگاشت می‌کند ($\mathrm{Z} \subseteq \mathrm{R}^K$).

روش PCA یک روش کاهش بعد بدون نظارت است که هدف آن یافتن مؤلفه‌های اصلی، که جهت‌هایی با بیش‌ترین واریانس را برای داده مشخص می‌کنند، است. اگر فرض کنیم $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]^{\mathrm{T}}$ ماتریس ورودی متمرکز شده را مشخص کند، که در آن میانگین نمونه‌ها از هر نمونه کم شده است، $\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{U}^{\mathrm{T}}$ تجزیه مقدار منفرد $\mathbf{X}$ را مشخص می‌کند که $\mathbf{V}$ و $\mathbf{U}$ ماتریس‌های متعامد ستونی با ابعاد $N \times N$ و $M \times M$ بوده و $\mathbf{D}$ ماتریسی قطری متشکل از مقادیر منفرد می‌باشد، که به‌ترتیب نزولی مرتب شده و در امتداد قطر قرار گرفته‌اند. ماتریس متشکل از K ستون نخست از ماتریس $\mathbf{U}$ که با $\mathbf{U}_K$ نشان داده می‌شود، به‌عنوان نگاشت Ψ در نظر گرفته می‌شود (پیرسون، ۱۹۰۱).

روش تحلیل مؤلفهٔ اصلی احتمالاتی (PPCA)، یک بیان احتمالاتی از روش PCA را ارائه می‌کند (تیپینگ و بیشاپ، ۱۹۹۹-ب). روش PPCA یک مدل متغیر نهان[5] است و فرآیند تولید هر ورودی $\mathbf{x}$ را به‌صورت $\mathbf{x} = \mathbf{W}_x\mathbf{z} + \boldsymbol{\mu}_x + \boldsymbol{\varepsilon}_x$ مدل می‌کند، که $\mathbf{z} \in \mathrm{R}^K$ متغیرهای نهان و $\mathbf{W}_x$ ماتریسی $M \times K$ است، که factor loading نامیده می‌شود. در این مدل احتمالاتی به‌طورمعمول فرض می‌شود که بردار متغیرهای نهان $\mathbf{z}$ دارای توزیع گاسین با میانگین صفر و ماتریس کواریانس واحد است (یعنی $\mathbf{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$)، و $\boldsymbol{\varepsilon}_x$ نوفه می‌باشد که دارای توزیع گاسین متقارن شعاعی به‌صورت $\boldsymbol{\varepsilon}_x \sim \mathrm{N}(\mathbf{0}, \delta_x^2\mathbf{I})$ فرض شده است. پارامترهای مدل با یک روش تکراری به نام الگوریتم میانگین‌گیری-بیشینه‌سازی[6] EM تعیین می‌شوند. لگاریتم شباهت[7] برای نمونه داده‌ها در این مدل به‌صورت زیر تعریف می‌شود (تیپینگ و بیشاپ، ۱۹۹۹-الف):

$$l = \sum_{n=1}^{N} \ln\{p(\mathbf{x}_n)\} \qquad (۱)$$

که در آن $p(\mathbf{x}_n)$ توزیع حاشیه‌ای $\mathbf{x}$ است، و به‌صورت $p(\mathbf{x}_n) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$ تعریف می‌شود. نتایج مرتبط می‌تواند در قالب قضیه ۱ به‌صورت زیر خلاصه شود. جزییات اثبات این قضیه در مرجع (یو و همکاران، ۲۰۰۶) و (تیپینگ و بیشاپ، ۱۹۹۹-الف) قابل دستیابی است:

قضیه ۱: اگر $\mathbf{S}_x = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu}_n)(\mathbf{x}_n - \boldsymbol{\mu}_n)^{\mathrm{T}}$ ماتریس کواریانس نمونه داده‌های $\{\mathbf{x}_n\}_{n=1}^{N}$ باشد و $\lambda_1 \geq ... \geq \lambda_M$ مقادیر ویژه متناظر با بردارهای ویژه $\mathbf{u}_1, ..., \mathbf{u}_M$ ماتریس $\mathbf{S}_x$ باشد، و فضای نهان در مدل PPCA، فضایی K بعدی باشد، آن‌گاه (یو و همکاران، ۲۰۰۶):

(الف) تخمین بیش‌ترین شباهت برای $\mathbf{W}_x$ به‌صورت زیر به‌دست می‌آید:

$$\mathbf{W}_x = \mathbf{U}_K(\boldsymbol{\Lambda}_K - \delta_x^2\mathbf{I})^{\frac{1}{2}}\mathbf{R} \qquad (۲)$$

[1] Principal Component Analysis
[2] Probabilistic Principal Component Analysis
[3] Supervised Probabilistic Principal Component Analysis
[4] PPCA Mixture Model
[5] Latent Variable
[6] Expectation Maximization
[7] Log-Likelihood

که $\mathbf{\Lambda}_K = \mathrm{diag}(\lambda_1,...,\lambda_K)$ و $\mathbf{U}_K = [\mathbf{u}_1,...,\mathbf{u}_K]$ و $\mathbf{R}$ یک ماتریس متعامد $K \times K$ است.

(ب) نتیجه نگاشت برای ورودی جدید $\mathbf{x}^*$، بردار $\mathbf{z}^*$ است که به‌صورت زیر به‌دست می‌آید:

$$\mathbf{z}^* = \mathbf{R}^{\mathrm{T}} (\mathbf{\Lambda}_K - \delta_x^2 \mathbf{I})^{\frac{1}{2}} \mathbf{\Lambda}_K^{-1} \mathbf{U}_K^{\mathrm{T}} (\mathbf{x}^* - \boldsymbol{\mu}) \qquad (۳)$$

مدل PPCA به‌دلیل بیان در چارچوب احتمالاتی، مزایایی نسبت به PCA دارد، که از آن جمله می‌توان رویه یادگیری سریع الگوریتم EM، ارائه روشی اصولی برای مدیریت ورودی‌های مفقود، و امکان درنظرگرفتن مدل ترکیبی برای PCA را برشمرد (تیپینگ و بیشاپ، ۱۹۹۹-الف) (تیپینگ و بیشاپ، ۱۹۹۹-ب).

در مسائل یادگیری بانظارت هر داده $\mathbf{x}$ با یک مقدار خروجی $\mathbf{y} \in \mathrm{Y}$ متناظر است. مدل SPPCA نسخه بانظارت مدل PPCA است، که در آن برای مسایل رگرسیون $\mathrm{y} \in \mathrm{R}$ و برای مسایل دسته‌بندی $\mathrm{y} \in \{+1,-1\}$ در نظر گرفته می‌شود (یو و همکاران، ۲۰۰۶). مدل SPPCA بر این باور است که بین فضای ورودی X و فضای خروجی Y کواریانس وجود دارد و این امکان وجود دارد که PPCA را به مدلی که این کواریانس را به خوبی نشان می‌دهد تعمیم داد. بیان ریاضی روش SPPCA به‌صورت زیر است: اگر بعد فضای خروجی را L فرض کنیم، هر ورودی $\mathbf{x}$ متناظر با یک بردار خروجی $\mathbf{y} = [\mathrm{y}_1,...,\mathrm{y}_L]^{\mathrm{T}} \in \mathrm{Y} \subseteq \mathrm{R}^L$ است. در SPPCA داده مشاهده شده $(\mathbf{x},\mathbf{y})$ توسط مدل متغیر نهان زیر تولید می‌شود:

$$\begin{aligned} \mathbf{x} &= \mathbf{W}_x \mathbf{z} + \boldsymbol{\mu}_x + \boldsymbol{\varepsilon}_x \\ \mathbf{y} &= \mathbf{W}_y \mathbf{z} + \boldsymbol{\mu}_y + \boldsymbol{\varepsilon}_y \end{aligned} \qquad (۴)$$

متغیرهای نهان $z \sim N(0, I)$ بین $\mathbf{x}$ و $\mathbf{y}$ مشترک می‌باشند، و دو مدل نوفه که از یکدیگر مستقل هستند و هر دو دارای توزیع ایزوتوپیک گاسین هستند، به‌صورت $\boldsymbol{\varepsilon}_x \sim \mathrm{N}(\mathbf{0}, \delta_x^2 \mathbf{I})$ و $\boldsymbol{\varepsilon}_y \sim \mathrm{N}(\mathbf{0}, \delta_y^2 \mathbf{I})$ در نظر گرفته می‌شوند. در این روش اگر قرار دهیم:

$$\mathbf{\Phi} = \begin{pmatrix} \delta_x^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \delta_y^2 \mathbf{I} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} \mathbf{W}_x \\ \mathbf{W}_y \end{pmatrix} \qquad (۵)$$

بر اساس فرضیاتی که برای این روش در نظر گرفته شده است، به‌راحتی دیده می‌شود که $(\mathbf{x},\mathbf{y})$ دارای توزیع توأم گاسین با میانگین $\boldsymbol{\mu}$ و ماتریس کواریانس $\mathbf{\Phi} + \mathbf{W}\mathbf{W}^{\mathrm{T}}$ هستند (یو و همکاران، ۲۰۰۶). نتایج مرتبط با این روش را می‌توان به‌طور خلاصه در قالب قضیه ۲ به شکل زیر بیان کرد (جزییات این قضیه در مرجع (یو و همکاران، ۲۰۰۶) موجود است):

قضیه ۲: ماتریس کواریانس نمونه داده‌های هنجارسازی شده $\{(\mathbf{x}_n,\mathbf{y}_n)\}_{n=1}^N$ را به‌صورت زیر تعریف می‌کنیم:

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{\Phi}^{-\frac{1}{2}} \begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{pmatrix} \begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{pmatrix}^{\mathrm{T}} \qquad \mathbf{\Phi}^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\delta_x^2} \mathbf{S}_x & \frac{1}{\delta_x \delta_y} \mathbf{S}_{xy} \\ \frac{1}{\delta_x \delta_y} \mathbf{S}_{yx} & \frac{1}{\delta_y^2} \mathbf{S}_y \end{pmatrix} \qquad (۶)$$

و $\lambda_1 \geq ... \geq \lambda_{(M+L)}$ مقادیر ویژه متناظر با بردارهای ویژه $\mathbf{u}_1,...,\mathbf{u}_{(M+L)}$ از ماتریس $\mathbf{S}$ می‌باشند. ماتریس $\mathbf{S}_x$ ماتریس کواریانس بین ورودی‌ها، $\mathbf{S}_y$ ماتریس کواریانس بین خروجی‌ها، و $\mathbf{S}_{xy}$ و $\mathbf{S}_{yx}$ ماتریس کواریانس بین ورودی‌ها و خروجی‌ها است. اگر فضای نهان را K بعدی در نظر بگیریم، نتایج زیر را خواهیم داشت (یو و همکاران، ۲۰۰۶):

(الف) در SPPCA، ماتریس‌های $\mathbf{W}_x$ و $\mathbf{W}_y$ به‌صورت زیر به‌دست می‌آیند:

$$\begin{aligned} \mathbf{W}_x &= \delta_x \mathbf{U}_x (\mathbf{\Lambda}_K - \mathbf{I})^{\frac{1}{2}} \mathbf{R} \\ \mathbf{W}_y &= \delta_y \mathbf{U}_y (\mathbf{\Lambda}_K - \mathbf{I})^{\frac{1}{2}} \mathbf{R} \end{aligned} \qquad (۷)$$

که $\mathbf{\Lambda}_K = \mathrm{diag}(\lambda_1,...,\lambda_K)$ و $\mathbf{U}_x$ شامل M سطر نخست از ماتریس $[\mathbf{u}_1,...,\mathbf{u}_K]$ بوده و $\mathbf{U}_y$ حاوی L سطر باقیمانده از آن است. ماتریس $\mathbf{R}$ یک ماتریس متعامد دلخواه $K \times K$ است.

(ب) نتیجه نگاشت برای ورودی متمرکزشده $\mathbf{x}^*$، بردار $\mathbf{z}^*$ است که به‌صورت زیر به‌دست می‌آید:

$$\mathbf{z}^* = \frac{1}{\delta_x} \mathbf{R}^{\mathrm{T}} (\mathbf{\Lambda}_K - \mathbf{I})^{-\frac{1}{2}} [\mathbf{U}_x^{\mathrm{T}} \mathbf{U}_x + (\mathbf{\Lambda}_K - \mathbf{I})^{-1}]^{-1} \mathbf{U}_x^{\mathrm{T}} \mathbf{x}^* \qquad (۸)$$

در حالت خاص اگر $L=0$ باشد، آن‌گاه مدل بدون نظارت است و $\mathbf{S} = \frac{1}{\delta_x^2} \mathbf{S}_x$ خواهد بود، که مشابه ماتریس مربوط به PPCA خواهد شد (یو و همکاران، ۲۰۰۶).

در مدل ترکیبی تحلیل مؤلفه‌های اصلی احتمالاتی (PPCAMM) از ترکیب خطی این مدل‌ها برروی مجموعه داده‌ها استفاده می‌شود. درواقع روش PPCAMM ترکیب چند PPCA محلی است که در این روش بخش‌بندی مناسب نمونه داده‌ها و تعیین مؤلفه‌های اصلی مدل برای هر بخش به‌طور هم‌زمان با استفاده از تخمین بیش‌ترین شباهت تعیین می‌شود. لگاریتم شباهت نمونه داده‌ها برای این مدل ترکیبی به‌صورت زیر تعیین می‌شود:

$$l = \sum_{n=1}^{N} \ln\{p(\mathbf{x}_n)\} = \sum_{n=1}^{N} \ln\{\sum_{i=1}^{M} \pi_i p(\mathbf{x}_n|i)\} \quad (۹)$$

که M تعداد مدل‌های مشارکت‌کننده در ترکیب، N تعداد نمونه داده‌ها، و $p(\mathbf{x}_n|i)$ یک مدل PPCA است. ضریب π_i نسبت مشارکت مؤلفه مدل iام در ترکیب مورد نظر می‌باشد، که برای آن محدودیت‌های احتمالاتی $\pi_i \geq 0$ و $\sum_i \pi_i = 1$ برقرار است. هر مؤلفه از ترکیب بردار میانگین $\boldsymbol{\mu}_i$ و همچنین پارامترهای $\mathbf{W}_{xi}$ و δ_{xi} مربوط به خود را دارد (تیپینگ و بیشاپ، ۱۹۹۹-الف).

می‌توان برای بهینه‌کردن پارامترهای هر مؤلفه از ترکیب الگوریتم تکراری EM را مورد استفاده قرار داد. نشان داده می‌شود $\mathbf{R}_{ni} = \mathrm{p}(i \mid \mathbf{x}_n)$، که پاسخ احتمال پسین مؤلفه iام برای تولید نقطه داده $\mathbf{x}_n$ را بیان می‌کند، به‌صورت زیر به‌دست می‌آید (تیپینگ و بیشاپ، ۱۹۹۹-الف):

$$\mathbf{R}_{ni} = \frac{\mathrm{p}(\mathbf{x}_n \mid i)\pi_i}{\mathrm{p}(\mathbf{x}_n)} \quad (۱۰)$$

همچنین می‌توان نشان داد که π_i و $\boldsymbol{\mu}_{xi}$ به‌صورت زیر به‌دست می‌آیند (تیپینگ و بیشاپ، ۱۹۹۹-الف):

$$\tilde{\pi}_i = \frac{1}{N}\sum_{n=1}^{N} \mathbf{R}_{ni} \quad (۱۱)$$

$$\tilde{\boldsymbol{\mu}}_{xi} = \frac{\sum_{n=1}^{N} \mathbf{R}_{ni}\mathbf{x}_n}{\sum_{n=1}^{N} \mathbf{R}_{ni}} \quad (۱۲)$$

به‌علاوه، $\mathbf{W}_{xi}$ و δ_{xi}^2 با استفاده از بردارهای ویژه و مقادیر ویژه ماتریس کواریانس وزن‌دار محلی مدل iام که به‌صورت زیر تعریف می‌شود، به‌دست می‌آیند (تیپینگ و بیشاپ، ۱۹۹۹-الف):

$$\mathbf{S}_{xi} = \frac{1}{\tilde{\pi}_i N}\sum_{n=1}^{N} \mathbf{R}_{ni}(\mathbf{x}_n - \tilde{\boldsymbol{\mu}}_{xi})(\mathbf{x}_n - \tilde{\boldsymbol{\mu}}_{xi})^{\mathrm{T}} \quad (۱۳)$$

۳- روش پیشنهادی

۳-۱- مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی بانظارت

روش PCA با وجود محدودیتی که به‌خاطر خطی‌بودن دارد، یکی از پرکاربردترین روش‌های کاهش بعد است. تعمیم‌های مختلفی از روش PCA برای غلبه بر محدودیت خطی‌بودن آن مطرح شده است. یک پیشنهاد مطرح این است که مجموعه داده‌های پیچیده به‌صورت ترکیبی خطی محلی از نگاشت‌های PCA بیان شود. با این وجود PCA متناظر با یک چگالی احتمالاتی نیست و لذا روش منحصربه‌فردی برای ترکیب مدل‌های PCA وجود ندارد. پیشنهادهای قبلی برای توسعه یک مدل ترکیبی[1] برای PCA تک‌منظوره بوده و عمومیت نداشته‌اند.

همان‌طورکه ذکر شد، تیپینگ و بیشاپ روشی مطرح کردند که در آن PCA در چارچوب بیش‌ترین شباهت[2] و براساس صورت خاصی از مدل متغیر نهان گاسی فرمول‌بندی شده است. این روش که به‌عنوان روش تحلیل مؤلفهٔ اصلی احتمالاتی (PPCA) معرفی شده است، منجربه ایجاد یک مدل ترکیبی خوش‌تعریف برای تحلیل مؤلفهٔ اصلی احتمالاتی محلی خطی می‌شود، که پارامترهای آن می‌توانند با استفاده از الگوریتم تکراری EM تعیین شوند (تیپینگ و بیشاپ، ۱۹۹۹-الف)؛ اما این روش هنگامی که خروجی‌های مربوط به نمونه‌داده‌های ورودی در اختیار باشد، مانند مسائل دسته‌بندی و رگرسیون، قادر به استفاده مفید از این اطلاعات نیست.

با توجه به اهمیت مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی در فضاهای غیر خطی و پیچیده و نیز غلبه بر محدودیت‌های روش PCA و نیز با توجه به این موضوع که روش‌های کاهش بعد بانظارت در مسائل مختلف به‌خصوص دسته‌بندی و پیش‌بینی، به‌دلیل استفاده از اطلاعات مسأله از کارایی بهتری برخوردار هستند، ارائه روش بانظارت مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی در این مقاله مورد توجه قرار گرفته است. روش ارائه‌شده تحت عنوان SPPCAMM معرفی می‌شود. در روش پیشنهادی از ترکیب خطی مدل‌ها بر روی مجموعه داده‌ها با درنظرگرفتن برچسب دسته‌ها استفاده می‌شود.

روش SPPCAMM را می‌توان ترکیب چند SPPCA محلی دانست، که در آن بخش‌بندی مناسب نمونه‌داده‌ها و تعیین مؤلفه‌های اصلی مدل برای هر بخش به‌طور هم‌زمان با استفاده از تخمین بیش‌ترین شباهت انجام می‌شود. لگاریتم شباهت نمونه‌داده‌ها برای این مدل مشابه PPCAMM (تیپینگ و بیشاپ، ۱۹۹۹-الف) بوده و به‌صورت زیر تعریف می‌شود، با این تفاوت که در این حالت هر نمونه به‌صورت $(\mathbf{x}_n, \mathbf{y}_n)$ است:

$$l = \sum_{n=1}^{N} \ln\{p(\mathbf{x}_n, \mathrm{y_n})\} = \sum_{n=1}^{N} \ln\{\sum_{i=1}^{M} \pi_i p(\mathbf{x}_n, \mathrm{y_n}|i)\} \quad (۱۴)$$

که M تعداد مدل‌های ترکیبی، N تعداد نمونه داده‌ها و $p(\mathbf{x}_n, \mathbf{y}_n|i)$ یک مدل SPPCA است. اگر پاسخ احتمال پسین مدل iام برای تولید نقطه داده $(\mathbf{x}_n, \mathbf{y}_n)$ را به‌صورت $\mathbf{R}_{ni} = p(i \mid \mathbf{x}_n, \mathbf{y}_n)$ بیان کنیم، از رابطه زیر می‌توان این احتمال را محاسبه کرد:

$$\mathbf{R}_{\mathrm{ni}} = \frac{p(\mathbf{x}_n, \mathbf{y}_n \mid i)\pi_i}{p(\mathbf{x}_n, \mathbf{y}_n)} \quad (۱۵)$$

[1] Mixture model

[2] Maximum likelihood

که π_i نسبت ترکیب مدل iام مورد نظر با محدودیت‌های $\pi_i \geq 0$ و $\sum_i \pi_i = 1$ است، و به‌صورت زیر به‌دست می‌آید:

$$\tilde{\pi}_i = \frac{1}{N}\sum_{n=1}^{N} \mathbf{R}_{ni} \qquad (۱۶)$$

رابطهٔ بالا در گام M از الگوریتم EM و با بیشینه‌کردن لگاریتم شباهت داده نسبت به π_i به‌دست می‌آید. همچنین با بیشینه‌کردن لگاریتم شباهت نسبت به پارامترهای میانگین $\mathbf{x}$ ها ($\boldsymbol{\mu}_{xi}$)، میانگین $\mathbf{y}$ ها ($\boldsymbol{\mu}_{yi}$)، δ^2_{xi}، δ^2_{yi}، $\mathbf{W}_{xi}$، و $\mathbf{W}_{yi}$ مقادیر جدید برای این پارامترها به‌دست می‌آید. به‌علاوه پارامترهای مجهول هر مؤلفه را می‌توان با تجزیهٔ مقدار ویژه ماتریس کواریانس محلی داده برای آن مؤلفه به‌دست آورد. ماتریس کواریانس محلی برای هر مدل به‌صورت زیر قابل تعریف است:

$$\mathbf{S}_i = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\Phi}_i^{-\frac{1}{2}} \begin{pmatrix} \mathbf{x}_n - \boldsymbol{\mu}_{xi} \\ \mathbf{y}_n - \boldsymbol{\mu}_{yi} \end{pmatrix} \begin{pmatrix} \mathbf{x}_n - \boldsymbol{\mu}_{xi} \\ \mathbf{y}_n - \boldsymbol{\mu}_{yi} \end{pmatrix}^{\mathrm{T}} \qquad (۱۷)$$

$$\boldsymbol{\Phi}_i^{\frac{1}{2}} = \begin{pmatrix} \frac{1}{\delta^2_{xi}}\mathbf{S}_{xi} & \frac{1}{\delta_{xi}\delta_{yi}}\mathbf{S}_{xyi} \\ \frac{1}{\delta_{xi}\delta_{yi}}\mathbf{S}_{yxi} & \frac{1}{\delta^2_{yi}}\mathbf{S}_{yi} \end{pmatrix} \qquad (۱۸)$$

که در این رابطه $\mathbf{S}_{xi}$ به‌صورت زیر تعریف می‌شود:

$$\mathbf{S}_{xi} = \frac{1}{\pi_i N}\sum_{n=1}^{N} \mathbf{R}_{ni}(\mathbf{x}_n - \tilde{\boldsymbol{\mu}}_{xi})(\mathbf{x}_n - \tilde{\boldsymbol{\mu}}_{xi})^{\mathrm{T}} \qquad (۱۹)$$

ماتریس‌های $\mathbf{S}_{yi}$، $\mathbf{S}_{xyi}$، و $\mathbf{S}_{yxi}$ نیز به‌صورت مشابه با رابطه بالا محاسبه می‌شوند.

با به‌دست‌آوردن مقادیر ویژه و بردارهای ویژه ماتریس کواریانس بالا می‌توان $\mathbf{W}_{xi}$ و $\mathbf{W}_{yi}$ برای هر مدل را با استفاده از روابط زیر به‌دست آورد:

$$\mathbf{W}_{xi} = \delta_{xi}\mathbf{U}_{xi}(\boldsymbol{\Lambda}_K - \mathbf{I})^{\frac{1}{2}}\mathbf{R}$$

$$\mathbf{W}_{yi} = \delta_{yi}\mathbf{U}_{yi}(\boldsymbol{\Lambda}_K - \mathbf{I})^{\frac{1}{2}}\mathbf{R} \qquad (۲۰)$$

که $\boldsymbol{\Lambda}_K = \mathrm{diag}(\lambda_1,...,\lambda_K)$ و $\mathbf{U}_{xi}$ شامل M سطر نخست از ماتریس $[\mathbf{u}_{1i},...,\mathbf{u}_{Ki}]$، و $\mathbf{U}_{yi}$، حاوی L سطر باقیمانده از آن است. ماتریس $\mathbf{R}$ یک ماتریس متعامد دلخواه $K \times K$ است.

نگاشت $\mathbf{z}^*$ برای ورودی جدید $\mathbf{x}^*$ در فضای کاهش بعد یافته به‌صورت زیر خواهد بود:

$$\mathbf{z}^* = \frac{1}{\delta_{xi}}\mathbf{R}^{\mathrm{T}}(\boldsymbol{\Lambda}_K - \mathbf{I})^{-\frac{1}{2}}[\mathbf{U}_{xi}^{\mathrm{T}}\mathbf{U}_{xi} + (\boldsymbol{\Lambda}_K - \mathbf{I})^{-1}]^{-1}\mathbf{U}_{xi}^{\mathrm{T}}(\mathbf{x}^* - \boldsymbol{\mu}_{xi}) \qquad (۲۱)$$

رابطه بالا بردار ویژگی کاهش بعدیافته برای ورودی $\mathbf{x}^*$ به‌کمک مؤلفه iام از ترکیب را نشان می‌دهد. بردار ویژگی کاهش‌بعدیافته نهایی به‌صورت زیر به‌دست می‌آید. ابتدا بردار ویژگی کاهش بعدیافته با هر مؤلفه از ترکیب را با استفاده از رابطه (۲۱) به‌دست می‌آوریم؛ سپس فاصله اقلیدسی بازیابی بردار کاهش‌بعدیافته از بردار اصلی را به‌دست می‌آوریم. مؤلفه‌ای که کم‌ترین فاصله اقلیدسی را به‌دست دهد پیدا شده و بردار ویژگی کاهش‌بعدیافته نهایی با افکنش بردار اصلی بر آن مؤلفه به دست می‌آید.

محاسبه مقادیر δ^2_{xi} و δ^2_{xi} با استفاده از روابط زیر انجام می‌شود (یو و همکاران، ۲۰۰۶) (تیپینگ و بیشاپ، ۱۹۹۹-الف):

$$\delta^2_{xi} = \frac{1}{M\sum_n \mathbf{R}_{ni}}\{\sum_{n=1}^{N} \mathbf{R}_{ni}\left\|\mathbf{x}_n - \boldsymbol{\mu}_{xi}\right\| + \mathbf{R}_{ni}\,\mathrm{tr}(\mathbf{W}_x^{\mathrm{T}}\mathbf{W}_x\mathbf{C}) - 2\mathbf{R}_{ni}\,\mathrm{tr}(\mathbf{X}\mathbf{W}_x\mathbf{Z}^{\mathrm{T}})\}$$

$$\delta^2_{yi} = \frac{1}{M\sum_n \mathbf{R}_{ni}}\{\sum_{n=1}^{N} \mathbf{R}_{ni}\left\|\mathbf{y}_n - \boldsymbol{\mu}_{yi}\right\| + \mathbf{R}_{ni}\,\mathrm{tr}(\mathbf{W}_y^{\mathrm{T}}\mathbf{W}_y\mathbf{C}) - 2\mathbf{R}_{ni}\,\mathrm{tr}(\mathbf{X}\mathbf{W}_y\mathbf{Z}^{\mathrm{T}})\}$$

(۲۲)

در روابط بالا $\mathbf{X} = [\mathbf{x}_1,..., \mathbf{x}_N]^{\mathrm{T}}$، $\mathbf{Z} = [<\mathbf{z}_1>,...,<\mathbf{z}_N>]^{\mathrm{T}}$، و $\mathbf{C} = \sum_{n=1}^{N} <\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}>$ است، که $<.>$ بیان‌کنندهٔ امید ریاضی بوده، و آماره‌های $<\mathbf{z}_n>$ و $<\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}>$ با تخمین توزیع $\mathbf{z}_n$ برای مشاهدات $(\mathbf{x}_n, \mathbf{y}_n)$ به‌دست می‌آیند. وقتی که ورودی $\mathbf{x}$ و خروجی $\mathbf{y}$ هر دو در دست باشند، توزیع پسین $\mathbf{z}$ برای $(\mathbf{x}, \mathbf{y})$ داده‌شده به‌صورت زیر تعریف می‌شود (یو و همکاران، ۲۰۰۶):

$$p(\mathbf{z}|\mathbf{x},\mathbf{y}) \propto p(\mathbf{x},\mathbf{y}|\mathbf{z})p(\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z})p(\mathbf{z}) \qquad (۲۳)$$

با توجه به این‌که سه جملهٔ سمت راست دارای توزیع گاسین هستند، توزیع پسین $\mathbf{z}$ نیز گاسین ($\mathrm{N}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$) خواهد بود، و داریم:

$$\boldsymbol{\mu}_Z = \mathbf{A}^{-1}[\frac{1}{\delta_x^2}\mathbf{W}_x^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}_x) + \frac{1}{\delta_y^2}\mathbf{W}_y^{\mathrm{T}}(\mathbf{y} - \boldsymbol{\mu}_y)]$$

$$\boldsymbol{\Sigma}_Z = \mathbf{A}^{-1}$$

$$\mathbf{A} = \frac{1}{\delta_x^2}\mathbf{W}_x^{\mathrm{T}}\mathbf{W}_x + \frac{1}{\delta_y^2}\mathbf{W}_y^{\mathrm{T}}\mathbf{W}_y + \mathbf{I} \qquad (۲۴)$$

بنابراین برای هر مدل در SPPCAMM نیز مشابه بالا داریم:

$$\mathbf{A}_i = \frac{1}{\delta_{xi}^2}\mathbf{W}_{xi}^{\mathrm{T}}\mathbf{W}_{xi} + \frac{1}{\delta_{yi}^2}\mathbf{W}_{yi}^{\mathrm{T}}\mathbf{W}_{yi} + \mathbf{I} \qquad (۲۵)$$

و آماره‌های $<\mathbf{z}_n>$ و $<\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}>$ با استفاده از روابط بالا به‌صورت زیر تعریف می‌شوند:

$$< \mathbf{z}_{ni} >= \mathbf{A}_i^{-1}[\frac{1}{\delta_{xi}^2}\mathbf{W}_{xi}^{\mathrm{T}}(\mathbf{x}_n - \boldsymbol{\mu}_{xi}) + \frac{1}{\delta_{yi}^2}\mathbf{W}_{yi}^{\mathrm{T}}(\mathbf{y}_n - \boldsymbol{\mu}_{yi})]$$

$$< \mathbf{z}_{ni}\mathbf{z}_{ni}^{\mathrm{T}} >= \mathbf{A}_i^{-1} + < \mathbf{z}_{ni} >< \mathbf{z}_{ni} >^{\mathrm{T}} \quad (۲۶)$$

در شکل (۲) کارنمای روش پیشنهادی نمایش داده شده است.

۳-۲- کاهش بعد محلی خطی بدون اتلاف با استفاده از جریمه نگاشت

همان‌طور که پیش‌تر گفته شد، در این مقاله روشی برای شناسایی چهره با استفاده از کاهش بعد بدون اتلاف ارائه می‌شود. یادگیری یک مدل پیش‌بینی خطی با پارامترهای $(\mathbf{w},b)$ در یک فضای p بعدی با N نمونه آزمایشی $\{\mathbf{x}_i, \mathrm{y}_i\}_{i=1}^N$ به‌صورت زیر قابل تعریف است:

$$\underset{\mathbf{w}\in\mathrm{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \mathrm{L}(\mathrm{y}_i, \mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \quad (۲۷)$$

که $\mathbf{w} \in R^p$ بردار پارامتر و b پارامتر مدل پیش‌گو را بازنمایی می‌کند. متغیرهای $\mathbf{x}_i$ و y_i به‌ترتیب بردار ویژگی p بعدی و خروجی مربوط به نمونه iام را مشخص می‌کنند. تابع L یک تابع اتلاف است، و به مدل پیش‌گو بستگی دارد که می‌تواند به‌صورت مربعات خطا در نظر گرفته شود.

اگر $\mathbf{P}$ را یک ماتریس با ابعاد $d\times p$ در نظر بگیریم، که یک کاهش بعد خطی به d بعد بر روی فضای ورودی p بعدی انجام می‌دهد، بازنمایی کاهش‌بعدیافته ورودی $\mathbf{x}$ به‌صورت $\mathbf{Px}$ خواهد بود. با اعمال کاهش بعد خطی مدل پیش‌گو در فضای کاهش بعد یافته به‌صورت زیر تعریف می‌شود:

$$\underset{\mathbf{v}\in\mathrm{R}^d, b}{\operatorname{argmin}} \sum_{i=1}^{N} \mathrm{L}(\mathrm{y}_i, \mathbf{v}^{\mathrm{T}}(\mathbf{P}\mathbf{x}_i) + b) \quad (۲۸)$$

که $\mathbf{v} \in R^d$ بردار پارامتر یادگرفته‌شده در فضای کاهش‌بعدیافته است.

از مقایسه رابطه (۲۷) با (۲۸) ارتباط بین کاهش بعد فضای ویژگی و محدودیت روی فضای پارامتر را می‌بینیم. استفاده از ماتریس کاهش بعد $\mathbf{P}$ در فضای ورودی، موجب می‌شود که بردار پارامتر $\mathbf{w}$ با بعد p بر زیرفضای زیر واقع شود:

$$M_p = \left\{\mathbf{w}\in R^p \mid \mathbf{w} = \mathbf{P}^T\mathbf{v}, \exists \mathbf{v}\in R^d\right\} \quad (۲۹)$$

بنابراین اعمال عمل‌گر کاهش بعد خطی $\mathbf{P}$ در فضای ورودی معادل با محدودکردن مدل جستجو به یک زیرفضای پارامتر M_p است، که در رابطه (۲۹) تعریف شده است. از دیدگاه مدل جستجو، همهٔ نامزدهایی که در M_p نیستند از دامنه جستجو حذف می‌شوند.

بدین ترتیب کاهش بعد باعث ازبین‌رفتن اطلاعات می‌شود، و هیچ تضمینی وجود ندارد که پارامترهای به‌دست آمده در فضای کاهش‌بعدیافته با پارامترهای بهینه در فضای اصلی یکسان باشند؛ لذا در روش جریمهٔ نگاشت پیشنهاد شده است که جستجوی پارامترها در فضای اصلی انجام شود، و فاصله آن از زیرفضای به‌دست آمده به‌عنوان جریمه در نظر گرفته شود (ژانگ و اشنایدر، ۲۰۱۰). با این روش مدل پیش‌گو به‌صورت زیر در خواهد آمد:

$$\operatorname{argmin} \sum_{i=1}^{N} \mathrm{L}(\mathrm{y}_i, \tilde{\mathbf{w}}^{\mathrm{T}}\mathbf{x}_i + \mathbf{v}^{\mathrm{T}}(\mathbf{P}\mathbf{x}_i) + b) + \lambda \mathrm{J}(\tilde{\mathbf{w}})$$

$$\tilde{\mathbf{w}} \in \mathrm{R}^p, \mathbf{v} \in \mathrm{R}^d, b \quad (۳۰)$$

که $\tilde{\mathbf{w}} = \mathbf{w} - \mathbf{P}^{\mathrm{T}}\mathbf{v}$ و J تابع جریمه است و می‌تواند به‌صورت $\|\|_1$ یا $\|\|_2^2$ در نظر گرفته شود. متغیر λ نیز پارامتر تنظیم است. اگر مدل پیش‌گوی مورد نظر و عمل‌گر کاهش بعد طوری طراحی شوند که روی فضای هسته کار کنند، مدل پیش‌گو با جریمه نگاشت به‌صورت زیر خواهد بود:

$$\underset{\mathbf{w}\in\mathrm{R}^p, b}{\operatorname{argmin}} \sum_{i=1}^{N} \mathrm{L}(\mathrm{y}_i, \mathbf{w}^{\mathrm{T}}\Phi(\mathbf{x}_i) + b) + \min_{\mathbf{v}\in R^d} \lambda J(\mathbf{w} - \mathbf{P}^T\mathbf{v}) \quad (۳۱)$$

که $\Phi(\mathbf{x}_i)$ بازنمایی نمونه $\mathbf{x}_i$ در فضای هسته است.

از آن‌جا که مسأله، دسته‌بندی است، در رابطهٔ بالا از SVM با متغیرهای سستی[1] $\{\xi_i\}_{i=1}^N$ به‌عنوان تابع L استفاده می‌شود. به این ترتیب مسأله برنامه‌ریزی درجهٔ دو[2]ی زیر را داریم:

$$\underset{\mathbf{w}\in\mathrm{R}^p, v\in R^d, b, \{\xi_i\}_{i=1}^N}{\operatorname{argmin}} \frac{1}{2}\left\|\mathbf{w} - \mathbf{P}^{\mathrm{T}}v\right\|_2^2 + C\sum_{i=1}^{n}\xi_i$$
$$s.t.\quad \mathrm{y}_i(\mathbf{w}^T\Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad ,\forall i$$
$$\xi_i \geq 0 \quad \forall i \quad (۳۲)$$

که $c \propto \frac{1}{\lambda}$ به‌جای λ در رابطهٔ (۳۱) قرار گرفته است. با قراردادن $\tilde{\mathbf{w}} = \mathbf{w} - \mathbf{P}^{\mathrm{T}}\mathbf{v}$ داریم [۱۶]:

$$\underset{\tilde{\mathbf{w}}\in\mathrm{R}^p, v\in R^d, b, \{\xi_i\}_{i=1}^N}{\operatorname{argmin}} \frac{1}{2}\left\|\tilde{\mathbf{w}}\right\|_2^2 + C\sum_{i=1}^{n}\xi_i$$
$$s.t.\quad \mathrm{y}_i(\tilde{\mathbf{w}}^T\Phi(\mathbf{x}_i) + \mathbf{v}^T\mathrm{P}\Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad ,\forall i$$
$$\xi_i \geq 0 \quad \forall i \quad (۳۳)$$

[1] Slack Variables
[2] Quadratic Programming

تا این‌جا کاهش بعد با درنظرگرفتن جریمهٔ نگاشت برای روش‌های کاهش بعد خطی و غیرخطی مبتنی بر هسته بیان شد. آنچه در این مقاله مورد بررسی قرار می‌گیرد، کاهش بعد محلی خطی با درنظرگرفتن جریمه نگاشت است. در روابط (۳۰) و (۳۳) عمل‌گر کاهش بعد $\mathbf{P}$ تنها از طریق جمله $\mathbf{P}\mathbf{x}_i$ (در رابطهٔ (۳۰)) و $\mathbf{P}\Phi(\mathbf{x}_i)$ (در رابطهٔ (۳۳)) ظاهر شده است. می‌توان با یک ترفند ساده این دو جمله را با یک تابع کاهش بعد عمومی مانند $\Psi(\mathbf{x}_i)$ جایگزین کرد [۱۶]. در این مقاله یک خمینه[1] زیربنایی محلی خطی با استفاده از روش ارائه‌شده برای مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی بانظارت (SPPCAMM) از نمونه داده‌ها به‌دست می‌آید، که همان تابع کاهش بعد $\Psi(\mathbf{x}_i)$ خواهد بود. با جای‌گذاری این تابع به‌جای $\mathbf{P}$ در رابطه (۳۳) مسألهٔ بهینه‌سازی زیر را خواهیم داشت:

$$\begin{aligned}&\underset{\tilde{\mathbf{w}}\in\mathrm{R}^{p},v\in R^{d},b,\{\xi_i\}_{i=1}^{N}}{\operatorname{argmin}} \frac{1}{2}\left\|\tilde{\mathbf{w}}\right\|_2^2+C\sum_{i=1}^{n}\xi_i\\ &s.t.\quad \mathrm{y}_i\left(\tilde{\mathbf{w}}^T\Phi(\mathbf{x}_i)+\mathbf{v}^T\Psi(\mathbf{x}_i)+b\right)\geq 1-\xi_i \quad ,\forall i\\ &\qquad \xi_i\geq 0 \quad \forall i \end{aligned} \qquad (۳۴)$$

درواقع در این روش به جای نگاشت داده بر خمینه زیربنایی و جستجوی پارامترهای مدل پیش‌گو در آن که موجب از دست‌رفتن اطلاعات می‌شود، همان‌طور که در روابط بالا بیان شد، جستجو را در فضای اصلی انجام می‌دهیم و فاصله پارامترهای به‌دست‌آمده در فضای اصلی از پارامترهای به‌دست‌آمده در خمینه محلی خطی حاصل از SPPCAMM را به‌عنوان یک جریمه نگاشت در یادگیری مدل پیش‌گو مورد استفاده قرار می‌دهیم.

۴- نتایج پیاده‌سازی

در این بخش به بررسی کارایی روش پیشنهادی پرداخته و آن را با چند روش کاهش بعد دیگر مورد مقایسه قرار می‌دهیم. برای این منظور مسئلهٔ شناسایی چهره را در نظر گرفته‌ایم، و آزمایش‌ها را بر روی پایگاه داده‌های چهره ORL و Yale انجام داده‌ایم. پایگاه داده Yale شامل ۱۶۵ تصویر سیاه‌سفید از پانزده نفر است. برای هر شخص یازده تصویر با حالت‌های مختلف چهره و شرایط نوری متفاوت وجود دارد. در این مقاله از تصاویر این پایگاه داده که بر روی سایت دانشگاه MIT قرار دارد[2] و شامل تصاویر هنجارسازی شده است، استفاده گردید. در این هنجارسازی، تصاویر بر اساس فاصله بین دو چشم تنظیم شده و سپس تصویر چهره از عکس استخراج می‌شود. پایگاه داده ORL شامل تصاویری از چهل نفر است. از هر نفر ده تصویر متفاوت وجود دارد. تصاویر این پایگاه داده، تغییراتی از قبیل شدت روشنایی، حالت‌های چهره (چشمان باز و بسته، و لبخند)، تصاویر با عینک و بدون عینک، مقیاس و فاصله و چرخش زاویهٔ سر را شامل می‌شود. در آزمایش بر روی این پایگاه داده‌ها ابتدا هر تصویر به ۳۲×۳۲ پیکسل تغییر اندازه داده می‌شود.

در آزمایش‌ها از روش استخراج ویژگی گابور[3] برای به‌دست آوردن بردار ویژگی مربوط به هر تصویر استفاده شده است. فیلترهای گابور ابزاری برای استخراج ویژگی هستند. این توابع دوبعدی، لبهٔ اشکال و همچنین گودی‌ها و برآمدگی‌های تصویر را تقویت می‌کنند. برای افزایش تمایز چشم‌ها، دهان و بینی که به‌عنوان اجزای مهم و اساسی چهره مطرح هستند، فیلترهای گابور می‌توانند مفید باشند (حقیقت و همکاران ۲۰۱۳). طول بردار ویژگی استخراج‌شده با این روش ۲۵۶۰ است.

در آزمایش‌ها روش‌های زیر با هم مقایسه شده‌اند: ۱) روش ماشین‌بردار پشتیبان (SVM) در فضای ویژگی کاهش‌بعدیافته با روش تحلیل مؤلفهٔ اصلی هسته (KPCA) برای کاهش بعد، که به‌اختصار آن را SVM-KPCA می‌نامیم. این روش ابتدا KPCA را اجرا کرده، و سپس SVM خطی را روی فضای کاهش‌بعدیافته اجرا می‌کند. ۲) روش ماشین‌بردار پشتیبان در فضای ویژگی کاهش‌بعدیافته با روش ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی که آن را به‌اختصار SVM-PPCAMM می‌نامیم. این روش ابتدا PPCAMM را اجرا کرده، و سپس SVM خطی را روی فضای کاهش‌بعدیافته اجرا می‌کند. ۳) روش ماشین‌بردار پشتیبان در فضای ویژگی کاهش‌بعدیافته با روش پیشنهادی مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی بانظارت که آن را به‌اختصار SVM-SPPCAMM می‌نامیم. در این روش ابتدا SPPCAMM را اجرا کرده، و سپس SVM خطی را روی فضای کاهش‌بعدیافته اجرا می‌کنیم. ۴) روش KPCA با درنظرگرفتن جریمه نگاشت که به‌اختصار آن را Proj-KPCA می‌نامیم. ۵) مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی با درنظرگرفتن جریمه نگاشت، که آن را به‌اختصار Proj-PPCAMM می‌نامیم. ۶) به‌کارگیری روش پیشنهادی

[1] Manifold

[2] http://vismod.media.mit.edu/vismod/classes/mas622-00/datasets/

[3] Gabor

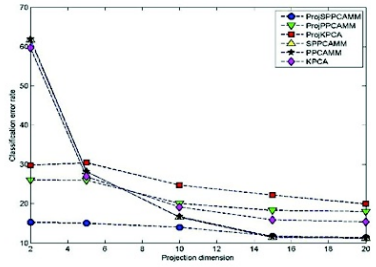
SPPCAMM با جریمه نگاشت، که آن را به‌اختصار Proj-SPPCAMM می‌نامیم.

در جدول (۱) تا (۴) در همه روش‌ها نمونه داده‌ها به فضایی با بعد ده کاهش بعد داده شده‌اند. در آزمایش‌های انجام‌شده هر بار به‌صورت تصادفی تعدادی از نمونه‌ها برای آموزش در نظر گرفته شده و مابقی برای آزمایش استفاده می‌شوند. ما از مجموعه‌های آموزشی با ۳، ۵، و ۷ تصویر از هر شخص برای مقایسه روش‌ها استفاده کرده‌ایم. مقدار نخستین پارامترهای δ_x^2 و δ_y^2 در روش‌هایی که این پارامتر را دارند، برابر 10^{-5} قرار داده شده است. هر آزمایش سی بار و به‌صورت مستقل تکرار شده است. در آزمایش‌هایی که در آن‌ها از روش استخراج ویژگی استفاده شده است، چون ویژگی‌های استخراج‌شده دارای مقادیر عددی کوچک هستند، برای عملکرد بهتر روش‌ها به‌خصوص روش‌های PPCAMM و SPPCAMM مقادیر با ضرب در یک مقدار عددی مناسب تغییر مقیاس داده شده‌اند.

جزییات مربوط به پیاده‌سازی جریمهٔ نگاشت به‌صورتی که در (ژانگ و اشنایدر، ۲۰۱۰) بیان شده، در نظر گرفته شده است. در استفاده از SVM با به‌کارگیری جریمه نگاشت از هسته چندجمله‌ای درجهٔ دو استفاده شده است. همچنین پارامتر C در آزمایش‌ها برابر 10^3 قرار داده شده است. در جدول (۱) و (۲) میانگین خطای دسته‌بندی برای هر روش بر روی پایگاه داده ORL گزارش شده است، و جدول (۳) و (۴) نتایج مربوط به پیاده‌سازی‌ها بر روی پایگاه داده Yale را نشان می‌دهند.

برای بررسی بیش‌تر روش‌های مورد مقایسه، به‌ازای تعداد ویژگی‌های استخراج شده مختلف، نه فقط ده ویژگی، آزمایش‌ها را تکرار کرده‌ایم. در آزمایش‌ها بر روی پایگاه داده Yale با مجموعه آموزشی شامل هفت تصویر از هر شخص، بعد داده به ۲، ۵، ۱۰، ۱۵، و ۲۰ کاهش داده شده، و متوسط نرخ خطای دسته‌بندی در مقابل تعداد ویژگی‌های استخراج‌شده برای سی بار اجرای تصادفی در قالب نمودار شکل (۱) رسم شده است. در این آزمایش‌ها خطای به‌دست آمده برای روش SPPCAMM در بعدهای مختلف به خطاهای روش PPCAMM بسیار نزدیک است. از این رو نمودار خطای این روش‌ها در این شکل بر روی هم قرار گرفته است.

در جدول (۵) معیار نرخ شناسایی درست[1] به معنای نسبت تعداد چهره‌هایی که درست تشخیص داده می‌شوند، به تعداد کل چهره‌ها برای ارزیابی گزارش شده است. آزمایش‌ها برای سی بار اجرای تصادفی بر روی پایگاه داده Yale با مجموعه آموزشی شامل هفت تصویر از هر شخص و کاهش بعدبه‌بعد ده انجام شده است.



(شکل- ۱): نمودار نرخ خطای دسته بندی (به درصد) در مقابل تعداد ویژگی‌های استخراج شده

در این جدول همچنین دو معیار ارزیابی دقت[2] و فراخوانی[3] برای یک‌بار اجرای تصادفی با همان تنظیمات قبلی گزارش شده است. این مقادیر از میانگین دقت و فراخوانی دسته‌های مختلف مسأله روی نمونه‌های آزمایشی، که در ادامه تعریف شده‌اند، محاسبه می‌شوند. دقت دستهٔ iام نسبت تعداد نمونه‌هایی است که الگوریتم به‌درستی آن‌ها را به دستهٔ iام منسوب کرده است، به تعداد کل نمونه‌هایی که الگوریتم آن‌ها را به دستهٔ iام منسوب کرده است. فراخوانی برای دستهٔ iام نیز نسبت تعداد نمونه‌هایی است که الگوریتم به‌درستی آن‌ها را به دستهٔ iام منسوب کرده است، به تعداد کل نمونه‌هایی که به‌درستی متعلق به دستهٔ iام هستند.

همچنین در جدول (۶) روش‌های مورد مقایسه از نظر زمان اجرا مورد بررسی قرار گرفته‌اند. زمان‌های اجرای به‌دست آمده برای کاهش بعدبه‌بعد ده بر روی پایگاه داده Yale و مجموعه آموزشی شامل سه تصویر از هر شخص بوده و بر حسب ثانیه گزارش شده‌اند.

از آن‌جا که روش پیشنهادی یک روش بانظارت است، با روش‌های کاهش بعد بانظارت LDA و SPPCA نیز مقایسه شده است. برای این منظور از پایگاه داده Yale استفاده شده، و کاهش بعدبه‌بعد ده بر روی نمونه داده‌ها انجام گرفته است. برای روش‌های مورد بررسی از مجموعه داده‌های آموزشی و آزمایشی یکسان استفاده شده است. نتایج این آزمایش‌ها در جدول شماره (۷) آمده است.

[1] Recognition rate

[2] Precision
[3] Recal

۵- تحلیل نتایج

همان‌طور که از نتایج به‌دست‌آمده از آزمایش‌ها مشاهده می‌شود، روش پیشنهادی با درنظرگرفتن جریمهٔ نگاشت (Proj-SPPCAMM) در مقایسه با سایر روش‌های کاهش بعد آزمایش‌شده، در تمام آزمایش‌ها خطای دسته‌بندی کمتری دارد. با مقایسهٔ روش PPCAMM با درنظرگرفتن جریمهٔ نگاشت و مدل بانظارت آن یعنی Proj-SPPCAMM برتری روش Proj-SPPCAMM به‌طورکامل مشهود است. این به بدان معناست که با استفاده از اطلاعات خروجی و اعمال جریمه نگاشت، قادر به یادگیری بهتر مدل در مسائل بانظارت هستیم. به‌علاوه همان‌طور که از نتایج آزمایش‌ها مشاهده می‌شود، روش PPCAMM در مقایسه با KPCA در بیش‌تر آزمایش‌ها خطای کمتری دارد. این امر می‌تواند نشان‌دهندهٔ مناسب‌تربودن روش‌های کاهش بعد محلی خطی نسبت به روش‌های کاهش بعد غیر خطی سراسری در ارائه مجموعه داده‌ها (به‌دلیل انعطاف‌پذیری بیش‌تر و ساده‌تربودن این مدل‌ها) باشد. از آن‌جا که مدل بانظارت این روش، یعنی SPPCAMM که در این مقاله ارائه شده است، هم دارای خاصیت محلی خطی است، و هم می‌تواند از اطلاعات مربوط به خروجی در مسائل پیش‌بینی و دسته‌بندی بهره گیرد، برتری این مدل نسبت به دیگر روش‌های کاهش بعد خطی و غیرخطی موجه بوده، و این ویژگی‌ها امکان استفاده از آن در مسائل پیچیده و بانظارت را فراهم می‌کند.

مقایسه روش پیشنهادی با روش LDA در جدول (۷) نیز برتری روش پیشنهادی به‌خصوص برای تعداد داده‌های آموزشی کم را نشان می‌دهد؛ مقایسه نتایج روش SPPCAMM و SPPCA برتری نسبی روش SPPCAMM را نشان می‌دهد؛ اما همان‌طور که از جدول (۷) مشاهده می‌شود، نتایج بسیار به هم نزدیک هستند. این امر را می‌توان به‌دلیل کوچک‌بودن مجموعه آموزشی و یا در حالت کلی کوچک‌بودن مجموعهٔ نمونه‌داده‌ها دانست. چون برای یک مجموعه‌دادهٔ کوچک چند مؤلفه در نظر گرفته می‌شود؛ درحالی‌که یک مؤلفه هم می‌تواند بازنمایی داده را به‌خوبی انجام دهد. اگر مجموعه‌داده بزرگ باشد استفاده از روش SPPCAMM و درنظرگرفتن چند مؤلفه برای بازنمایی داده، کارایی را به نحو قابل توجهی می‌تواند بهبود بخشد.

همان‌طورکه از نتایج به‌دست‌آمده از آزمایش‌ها مشاهده می‌شود، با افزایش تعداد داده‌های آموزشی متوسط خطای دسته‌بندی به‌طورتقریبی در همهٔ روش‌ها کاهش می‌یابد، که این ویژگی برای روش پیشنهادی نیز برقرار است. همچنین با توجه به نمودارهای شکل (۱) که خطای دسته‌بندی در مقابل تغییرات بعد را نشان می‌دهد، مشاهده می‌شود که با افزایش بعد، خطای دسته‌بندی کاهش می‌یابد.

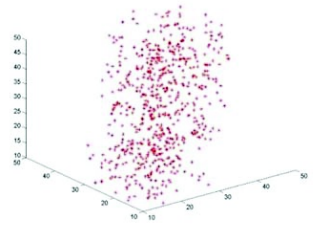
بررسی زمان اجرای روش‌های آزمایش‌شده در جدول (۶) نشان می‌دهد که زمان اجرا در روش پیشنهادی Proj-SPPCAMM نسبت به سایر روش‌ها بیشتر است. بعد از این روش و روش SPPCAMM زمان اجرای روش Proj-PPCAMM و PPCAMM زیاد است. در این روش‌ها قسمت بیش‌تر زمان اجرا مربوط به آموزش است؛ که دلیل آن استفاده از روش تکراری الگوریتم EM در آن‌ها است. زمان بالای آموزش را می‌توان ضعف روش پیشنهادی دانست؛ اما دقت به‌دست آمده از این روش به‌خصوص در بعدهای پایین قابل توجه است.

در روش پیشنهادی در بخش ۳-۱ روابط با فرض گاسی‌بودن توزیع داده به دست آمده است؛ لذا ممکن است این سئوال به ذهن آید که اگر توزیع داده غیر گاسی باشد آیا این روش همچنان عملکرد قابل قبولی خواهد داشت؟ همان‌طور که می‌دانیم در مسائل دنیای واقعی فرض گاسی‌بودن داده همیشه برقرار نیست. در مسایل بررسی‌شده در این مقاله، یعنی شناسایی چهره بر روی پایگاه داده Yale و ORL، نیز به الزام فرض گاسی‌بودن داده برقرار نیست. این در حالی است که نتایج آزمایش‌ها نشان می‌دهند روش پیشنهادی برای این مجموعه داده‌ها عملکرد خوبی داشته است؛ لذا مشاهده می‌شود که روش پیشنهادی برای دادهٔ غیر گاسی هم عملکرد قابل قبولی دارد.

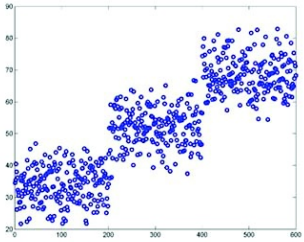
به‌منظور کسب بینش شهودی در این زمینه سه توزیع یکنواخت در فضای سه‌بعدی را که دامنهٔ آنها سه مکعب نزدیک به یکدیگر هستند، در نظر گرفته و از هر یک‌صد نمونه به تصادف برداشته‌ایم. بدین ترتیب چنان که در شکل (۲) نشان داده شده است، سیصد نمونه داده با بعد سه به‌دست آمده است. برای نمونه‌های تولیدشده از هر توزیع یک برچسب دسته در نظر گرفته‌ایم. بدین ترتیب نمونه‌ها از سه دسته متفاوت در نظر گرفته می‌شوند؛ سپس با استفاده از روش پیشنهادی SPPCAMM با سه مؤلفه، نمونه داده‌ها را به فضای با بعد دو نگاشت کرده‌ایم. شکل (۳) نمونه‌های کاهش‌بعدیافته بدین ترتیب را نشان می‌دهد. همان‌طورکه در شکل (۳) مشاهده می‌شود، نمونه داده‌ها به‌صورت سه دستهٔ به‌طورتقریبی جدا از هم در فضای

کاهش‌بعدیافته ظاهر شده‌اند، که بیان‌گر کارآیی مناسب روش پیشنهادی برای داده با توزیع غیر گاسی است.

در این مقاله ابتدا مدل بانظارت روش ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی (SPPCAMM) ارائه شده است. این مدل توسعه‌ای از مدل ترکیبی تحلیل مؤلفهٔ اصلی احتمالاتی است، که اطلاعات مربوط به برچسب داده را در نگاشت یادگیری‌شده منظور می‌کند. شکل (۴) روندنمای این مدل پیشنهادی را نشان می‌دهد؛ سپس این رویکرد بانظارت کاهش بعد محلی خطی SPPCAMM در چارچوبی به نام جریمهٔ نگاشت با استفاده از دسته‌بند SVM به‌کار گرفته شده است. جریمه نگاشت، این امکان را فراهم می‌آورد که از کاهش بعد برای بهبود دقت مدل پیش‌گو بدون خطر ازدست‌دادن اطلاعات مرتبط استفاده کنیم. نتایج آزمایش‌ها نشان‌دهندهٔ برتری روش پیشنهادی (SPPCAMM) با درنظرگرفتن جریمه نگاشت (Proj-SPPCAMM) در مقایسه با سایر روش‌های کاهش بعد بررسی‌شده در زمینهٔ دقت دسته‌بندی است.



(شکل- ۲): نمونه داده‌های تصادفی به دست آمده از سه توزیع یکنواخت نزدیک به هم



(شکل- ۳): نسخه کاهش بعد یافته نمونه داده‌های تصادفی شکل (۲) با روش SPPCAMM

۶- جمع‌بندی

پیشنهادهایی به‌شرح زیر برای ارتقای روش پیشنهادی در این مقاله مطرح هستند. همان‌طورکه بیان شد، در این مقاله از جریمهٔ نگاشت با دسته‌بند SVM استفاده شده است. به‌کارگیری ایدهٔ جریمه نگاشت با استفاده از سایر روش‌های دسته‌بندی، به‌عنوان یک موضوع پژوهشی قابل بررسی است. همچنین در روش کاهش بعد بدون اتلاف به‌کاررفته در این مقاله از هستهٔ چندجمله‌ای استفاده شده است. استفاده از هسته‌های دیگر و نیز یادگیری چندهسته‌ای در این مورد می‌تواند مفید باشد.

(جدول- ۱): متوسط خطای دسته‌بندی روش‌های مختلف بر روی پایگاه داده ORL برای ۳۰ بار اجرای تصادفی

ORL Dataset - Classification error rate (in %) ± standard deviation (*p*-value)						
Training faces(#)	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
3	18.39 ±3.27	18.01 ±3.40	16.83 ±2.88	15.50 ±2.54	16.63 ±3.47	**10.37 ±2.37**
5	9.08± 2.64	8.85± 2.27	8.37± 2.69	7.77± 2.01	9.71± 2.40	**4.25± 1.92**
7	5.61± 1.61	5.31± 2.42	5.58± 3.85	4.22± 1.69	4.86± 2.10	**2.14± 1.38**

(جدول- ۲): متوسط خطای دسته‌بندی روش‌های مختلف بر روی پایگاه داده ORL، استخراج ویژگی شده به روش گابور برای ۳۰ بار اجرای تصادفی

ORL Dataset – Gabor Feature Extraction- Classification error rate (in %) ± standard deviation (*p*-value)						
Training faces(#)	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
3	21.61 ±2.56	17.99 ±3.91	17.99 ±3.91	18.07 ±3.11	14.90 ±3.16	**7.23± 1.98**
5	11.00 ±2.23	9.91± 2.21	9.91± 2.21	9.32± 1.99	8.07± 1.76	**1.96± 0.84**
7	7.22± 1.54	5.95± 1.76	5.95± 1.76	6.31± 1.77	5.92± 1.70	**1.06± 0.90**

(جدول- ۳): متوسط خطای دسته‌بندی روش‌های مختلف بر روی پایگاه داده Yale برای ۳۰ بار اجرای تصادفی

Yale Dataset - Classification error rate (in %) ± standard deviation (*p*-value)						
Training faces(#)	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
3	25.91±3.85	26.52±3.11	26.52±3.11	25.96±3.46	26.17±2.63	**19.50±2.67**
5	22.06±4.10	21.62±3.38	21.62±3.38	21.97±4.16	21.70±2.85	**16.27±3.55**
7	20.11±3.63	18.62±3.97	18.01±4.27	21.39±4.91	20.84±3.55	**15.67±3.55**

(جدول- ۴): متوسط خطای دسته‌بندی روش‌های مختلف بر روی پایگاه داده Yale، استخراج ویژگی شده به روش گابور برای ۳۰ بار اجرای تصادفی

Yale Dataset – Gabor Feature Extraction- Classification error rate (in %) ± standard deviation (*p*-value)						
Training faces(#)	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
3	16.78±2.89	14.17±3.52	13.63±3.11	14.97±2.49	14.03±2.98	**10.46±2.65**
5	13.11±2.80	11.87±3.13	11.85±3.07	12.78±2.69	12.07±1.92	**8.11±2.80**
7	11.83±3.51	10.97±4.18	10.97±4.18	11.17±4.38	11.56±3.06	**7.74±3.41**

(جدول- ۵): بررسی معیارهای ارزیابی نرخ شناسایی درست، دقت، و فراخوانی بر روی پایگاه داده Yale و برای مجموعه آموزشی با ۷ تصویر از هر شخص

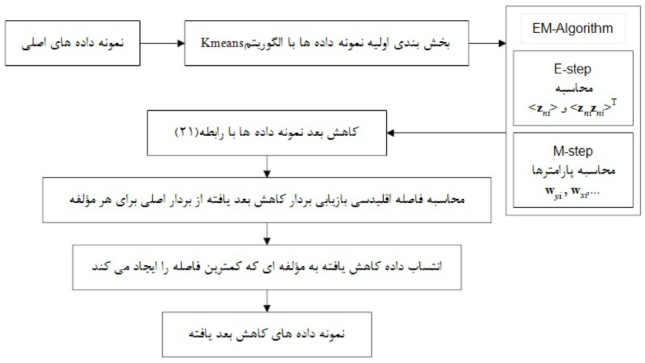
	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
Recognition rate	0.81	0.83	0.84	0.75	0.82	0.85
Precision	0.90	0.93	0.94	0.93	0.94	0.93
Recall	0.88	0.90	0.85	0.92	0.92	0.90

(جدول- ۶): بررسی زمان آموزش و آزمایش روش‌های مورد مقایسه بر حسب ثانیه

	KPCA	PPCAMM	SPPCAMM	Proj-KPCA	Proj-PPCAMM	Proj-SPPCAMM
زمان آزمایش	0.0960	0.1968	0.4242	3.8218	3.6171	4.8276
زمان آموزش	0.8156	43.5106	56.1545	3.0556	46.4498	72.4550

(جدول- ۷): متوسط خطای دسته‌بندی روش‌های بانظارت مختلف بر روی پایگاه داده Yale برای ۳۰ بار اجرای تصادفی

Yale Dataset - Classification error rate (in %) ± standard deviation (*p*-value)						
Training faces(#)	LDA	SPPCA	SPPCAMM	Proj- LDA	Proj- SPPCA	Proj-SPPCAMM
3	32.32±3.21	23.58±3.91	23.61±3.46	19.38±3.01	18.83±2.78	18.83±2.78
5	31.89±4.90	17.81±3.45	17.78±3.48	15.63±2.87	15.22±2.66	15.22±2.66
7	27.44±3.48	17.56±4.27	17.56±4.27	14.78±3.62	15.17±3.73	15.07±3.57



(شکل- ۴): کارنمای روش پیشنهادی SPPCAMM

۷- مراجع

C. Chen, J. Zhang, and R. Fleischer, "Distance approx-imating dimension reduction of Riemannian manifolds," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 40, pp. 208-217, 2010.

C.-G. Li and J. Guo, "Supervised isomap with explicit mapping," in Innovative Computing, Information and Control, 2006. ICICIC'06. First International Conference on, 2006, pp. 345-348.

C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

F. Han, T. Zhao, and H. Liu, "Coda: High dimensional copula discriminant analysis," Journal of Machine Learning Research, vol. 14, pp. 629-671, 2013.

J. Chen and Y. Liu, "Locally linear embedding: a survey," Artificial Intelligence Review, vol. 36, pp. 29-48, 2011.

K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 2, pp. 559-572, 1901.

L. Van der Maaten, E. Postma, and H. Van Den Herik, "Dimensionality reduction: A comparative review," Journal of Machine Learning Research, vol. 10, pp. 1-41, 2009.

M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," Neural computation, vol. 11, pp. 443-482, 1999.

M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 61, pp. 611-622, 1999.
http://vismod.media.mit.edu/vismod/classes/mas622-00/datasets/

M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb, "Identification Using Encrypted Biometrics," in Comp-uter Analysis of Images and Patterns, 2013, pp. 440-448.

M. Zhu and A. M. Martinez, "Subclass discriminant analysis," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 28, pp. 1274-1286, 2006.

N. Gkalelis, V. Mezaris, I. Kompatsiaris, and T. Stathaki, "Mixture subclass discriminant analysis link to restricted Gaussian model and other gener-alizations," Neural Networks and Learning Systems, IEEE Transactions on, vol. 24, pp. 8-21, 2013.

R.A. Fisher, “The Statistical Utilization of Multiple Measurements,” Annals of Eugenics, vol. 8, pp. 376-386, 1938.

S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller, “Fisher Discriminant Analysis with Kernels,” Proc. IEEE Signal Processing Soc. Workshop Neural Networks for Signal Processing IX, pp. 41-48, 1999.

S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, pp. 40-51, 2007.

S. Yu, K. Yu, V. Tresp, H.-P. Kriegel, and M. Wu, "Supervised probabilistic principal component analysis," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 464-473.

Y. Zhang and J. G. Schneider, "Projection penalties: Dimension reduction without loss," in Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 1223-1230.

Y.-D. Lan, H. Deng, and T. Chen, "Dimensionality Reduction Based on Neighborhood Preserving and Marginal Discriminant Embedding," Procedia Engineering, vol. 29, pp. 494-498, 2012.

سمیه احمدخانی کارشناسی خود را در رشت ه مهندسی کامپیوتر گرایش سخت‌افزار در سال ۱۳۸۹ از دانشگاه صنعتی همدان اخذ کرد. ایشان هم‌اکنون دانشجوی کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی در دانشگاه اصفهان است. زمینه‌های مورد علاقه ایشان یادگیری ماشین، شناسایی الگو و پردازش تصویر می‌باشد.
نشانی رایانامهٔ ایشان عبارت است از:
s.ahmadkhani@eng.ui.ac.ir

پیمان ادیبی مدرک کارشناسی خود را در رشتهٔ مهندسی کامپیوتر گرایش سخت‌افزار در سال ۱۳۷۶ از دانشگاه صنعتی اصفهان، و مدرک کارشناسی ارشد و دکترای خود را در رشته مهندسی کامپیوتر گرایش هوش مصنوعی از دانشگاه صنعتی امیرکبیر تهران به‌ترتیب در سال‌های ۱۳۸۰ و ۱۳۸۸ دریافت کرد. ایشان از سال ۱۳۸۹ تاکنون استادیار دانشکده مهندسی کامپیوتر دانشگاه اصفهان هستند. زمینه‌های پژوهشی مورد علاقهٔ ایشان در حال حاضر شامل یادگیری ماشین و شناسایی الگو، هوش محاسباتی و محاسبات نرم و بینایی کامپیوتر است.
نشانی رایانامهٔ ایشان عبارت است از:
adibi@eng.ui.ac.ir