اعمال تبدیل بر ویژگیها با استفاده از خطای کلاس‌بندی کمینه مبتنی بر هسته برای بازشناسی الگو و گفتار

بهزاد زمانی، احمد اکبری، بابک ناصرشریف

آزمایشگاه پردازش صدا و گفتار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران

نویسندهٔ عهده‌دار مکاتبات: بهزاد زمانی دهکردی،

آدرس پستی: تهران، دانشگاه علم و صنعت، دانشکده مهندسی کامپیوتر، آزمایشگاه پردازش صوت و گفتار

تلفن: ۰۹۱۳۳۸۳۹۲۴۹

رایانامک: bzamani@iust.ac.ir

چکیده

روشهای تبدیل ویژگی را می‌توان به دو دسته خطی و غیرخطی تقسیم کرد. روشهای تبدیل ویژگی مبتنی بر هسته از جمله روش‌های غیرخطی هستند که اخیراً مورد توجه بیشتری قرار گرفته‌اند. در این روشها، ایده اصلی نگاشت غیرخطی ویژگیها به فضایی با ابعاد بالاتر است. این نگاشت با هدفها و معیارهای متفاوتی صورت می‌گیرد. در آنالیز متمایزساز خطی مبتنی بر هسته (KLDA)، معیار تفکیک‌پذیری بیشتر ویژگیها در فضای جدید است، حال آنکه در آنالیز مولفه‌های اصلی مبتنی بر هسته (KPCA)، معیار متعامدسازی ویژگیها در فضای حاصل است. در مقاله حاضر یک روش جدید مبتنی بر هسته پیشنهاد و فرموله می‌شود که بر کمینه کردن خطای کلاس‌بندی در فضای ایجاد شده توسط هسته (KMCE) تکیه دارد. معیارهای بهینه‌سازی در روشهای KLDA و KPCA مستقل از خطای کلاس‌بندی می‌باشند در صورتی‌که در روش پیشنهادی علاوه بر بهره برداری از ایده‌ی نگاشت غیرخطی هسته، معیار کمینه‌سازی خطای کلاس‌بندی نیز مورد نظر قرار می‌گیرد. نتایج حاصل بر روی دادگان UCI و کلاس‌بندهای مختلف، نشان می‌دهند که روش پیشنهادی در مقایسه با روشهای تبدیل ویژگی خطی و روش‌های شناخته شده تبدیل ویژگی مبتنی بر هسته، در مورد کلاس‌بندهای مبتنی بر فاصله، نرخ بازشناسی بهتری دارد و در مورد کلاس‌بندهای آماری و مبتنی بر درخت تصمیم نیز کارآیی قابل قبولی دارد. همچنین آزمایشات انجام شده روی دادگان گفتاری Aurora2 عملکرد مطلوب روش پیشنهادی را نسبت به روشهای غیرخطی دیگر نشان می‌دهد.

کلمات کلیدی: تبدیل ویژگی، آنالیز تفکیک‌پذیری خطی، روش آنالیز مولفه اصلی، خطای کلاس‌بند کمینه، تابع هسته

۱. مقدمه

تبدیل ویژگی[1] اصل مهمی در شناسائی الگو[2]ست. هدف اصلی آن انتقال دادگان از فضای ویژگی اصلی به فضای جدیدی است که در این فضا توصیف ساختار داده‌ها بهتر انجام گیرد. به عنوان مثال کلاس‌های داده‌ای که بطور خطی جدایی‌پذیر نیستند، با اعمال تبدیلی بر داده‌ها یا همان ویژگیها به فضایی انتقال می‌یابند که بطور خطی جدایی‌پذیر باشند، عملاً کلاس‌بندی آسانتر می‌شود و تفکیک‌پذیری خطی کلاسها در فضای جدید افزایش می‌یابد. تبدیل ویژگی

[1] Feature Transformation
[2] Pattern Recognition

خود می‌تواند مراحل استخراج ویژگی[3]، کاهش ویژگی[4] و تبدیل ویژگی را در برداشته باشد که برای هر بخش روشهای مختلفی ارائه شده است. از سویی تبدیل ویژگی می‌تواند بصورت تبدیل خطی مانند آنالیز تفکیک‌پذیری خطی[5] (LDA) یا تبدیل غیرخطی مانند آنالیز تفکیک‌پذیری مبتنی بر هسته[6] (KDA) انجام گیرد.

تبدیل ویژگی فرایندی است که در طی آن مجموعه ویژگی جدیدی ایجاد می‌شود. از دیدگاهی می‌توان تولید ویژگی[7] و استخراج ویژگی را از انواع تبدیل ویژگی دانست. که معمولاً به هر دو روش کشف ویژگی[8] گفته می‌شود. در ایجاد ویژگی فرایند با تولید ویژگی‌های جدید همراه است که به این ترتیب علاوه بر ویژگی‌های اولیه، ویژگی‌های تولید شده نیز به مجموعه ویژگی‌ها اضافه می‌شود. ایجاد ویژگی تعداد ویژگی‌ها را زیاد می‌کند، در مقابل استخراج ویژگی تعداد ویژگی‌ها را کاهش می‌دهد. انتخاب ویژگی[9] زیرمجموعه‌ای از ویژگی‌های اصلی را بدون تغییر آن‌ها انتخاب می‌کند و به نوعی کاهش ویژگی را نیز در بردارد.

مطالعات زیادی در زمینه روش‌های تبدیل داده و ویژگی انجام شده است [4-1]. روش آنالیز مولفه اصلی[10](PCA) بر اساس مقادیر ویژه[11] و بردارهای ویژه[12] عمل می‌کند و بردارهای ویژه‌ای را بر می‌گرداند که دارای مقدار ویژه بزرگتری باشند [1]. روش PCA یک تبدیل خطی و بدون ناظر بوده و سعی در متعامد سازی داده‌های تبدیل یافته دارد [1]. از روشهای خطی دیگر می‌توان به روش MDS[13] اشاره کرد که در آن از ماتریس ضرب داخلی داده‌ها استفاده می‌شود و نشان داده شده است که هم‌ارز PCA می‌باشد [5].

روش آنالیز تفکیک‌پذیری خطی بر اساس معیار Fisher عمل می‌کند. این روش باناظر سعی در بدست آوردن تبدیلی دارد که واریانس درون کلاسی را کاهش و واریانس برون کلاسی را افزایش دهد [2]. عدم همبستگی دادگان در کلاس‌بندی مطلوب بوده، از اینرو روش آنالیز تفکیک‌پذیری خطی غیرهمبسته[14] (ULDA) مطرح گردیده است. روش ULDA تبدیلی بدست می‌آورد که علاوه بر افزایش معیار Fisher، مولفه‌های ویژگی تبدیل یافته بطور آماری همبستگی کمتری داشته باشند [3]. روش LDA برای مسائلی که فاصله کلاسها در آن کم می‌باشد عملکرد ضعیفی دارد که به همین جهت روش آنالیز تفکیک‌پذیری خطی وزن‌دهی[15] (WLDA) شده مطرح گردیده است[3]. عملکرد روش WLDA به نحوه انتخاب تابع وزن‌دهی کلاس‌ها وابسته است که Loog و همکارانش در [4] روش aPAC[16] را برای حل این مشکل پیشنهاد کردند. فرض یکسان بودن واریانس کلاس‌ها در روش LDA، عملکرد این روش را زمانی که واریانس درون کلاسها مشابه نیست، ضعیف می‌کند[33]. روشهای HDA[17] [6]، HLDA[18] [2]، UHLDA[19] [7] و PLDA[20] [8] این فرض را از LDA حذف کرده‌اند.

از دیگر روشهای تبدیل می‌توان به آنالیز مولفه‌های مستقل[21] (ICA) اشاره نمود. این روش همبستگی آماری مولفه‌های مختلف را کاهش می‌دهد [9]. از این تبدیل برای تفکیک-پذیری کور منابع[22] استفاده می‌شود. عملکرد ICA متناسب با درستی فرض مستقل بودن منابع می‌باشد [9]. روش خطی آنالیز تفکیک‌پذیری ارتجاعی[23] با ایجاد نیروی کشش بین نمونه‌های یک کلاس و نیروی رانش بین نمونه‌های دو کلاس

[3] Feature Extraction
[4] Feature Reduction
[5] Linear Discriminant Analysis (LDA)
[6] Kernel Discriminant Analysis (KDA)
[7] Feature Construction
[8] Feature Discovery
[9] Feature Selection
[10] Principal Component Analysis (PCA)
[11] Eigen Value
[12] Eigen Vector
[13] Metric Multi-Dimensional Scaling (MDS)
[14] Uncorrelated LDA (ULDA)
[15] Weighted LDA (WLDA)
[16] Approximate Pairwise Accuracy Criteria (aPAC)
[17] Heteroscedastic Discriminant Analysis (HDA)
[18] Heteroscedastic Linear Discriminant Analysis (HLDA)
[19] Uncorrelated Heteroscedastic LDA (UHLDA)
[20] Power Linear Discriminant Analysis (PLDA)
[21] Independent Component Analysis (ICA)
[22] Blind Source Separation (BSS)
[23] Springy Discriminant Analysis (SDA)

متفاوت سعی در تمایز نمونه‌ها دارد [10]. تکنیک‌های گفته شده با بهینه‌سازی معیاری سعی در بدست آوردن تبدیلی دارند، که اصولاً با معیار مبتنی بر خطای کلاس‌بندی متفاوت است. روش خطای کلاس‌بند کمینه[۲۴] (MCE) بر اساس کاهش خطای کلاس‌بندی عمل می‌کند. کاهش خطای کلاس‌بندی می‌تواند از طریق تغییر مدلها در مرحله یادگیری یا تبدیل ویژگی‌ها صورت پذیرد [11, 12]. روش تبدیل ویژگی MCE به صورت یک الگوریتم تکرار شونده است که در آن ماتریس تبدیل ویژگی با توجه به خطای کلاس‌بند بدست می آید. رابطه تکرار شونده این الگوریتم می تواند مبتنی بر گرادیان کاهشی [12] یا الگوریتم‌های تکاملی مانند استراتژی تکاملی[۲۵] [13] باشد.

دسته دیگری از روشهای تبدیل ویژگی، روشهای غیرخطی می‌باشند. تبدیلات خطی در متمایزسازی دادگانی که کلاسهای آن ذاتاً بطور خطی جدایی پذیر نیستند، عملکرد ضعیفی دارند. روش غیرخطی آنالیز مولفه‌های اصلی(NLPCA)[۲۶] مشابه با PCA برای تعیین و کاهش همبستگی داده‌ها به کار می‌رود. روش PCA همبستگی خطی بین ویژگیها را مشخص می‌کند، اما NLPCA همبستگی خطی و غیرخطی بین ویژگیها را بدون توجه به ماهیت غیرخطی داده‌ها مشخص می‌کند. در روش NLPCA یک شبکه عصبی برای تعیین نگاشت غیرخطی آموزش داده می‌شود. [14-16]

دیدگاه غیرخطی دیگر استفاده از تابع هسته می‌باشد، به این معنی که ابتدا با تابع هسته داده‌ها به فضای جدید نگاشت داده می‌شوند و در این فضا تبدیل خطی بر داده های نگاشت یافته اعمال می‌گردد. این ایده در ماشین بردار پشتیبان (SVM) غیرخطی بکار می‌رود [17, 18]. روش آنالیز تفکیک‌پذیری ارتجاعی مبتنی بر هسته[۲۷] [10]، آنالیز مولفه-های اصلی مبتنی بر هسته[۲۸] [19, 20]، و آنالیز تفکیک-پذیری مبتنی بر هسته[۲۹] [21] نمونه‌هایی از روشهای مبتنی بر هسته می‌باشند. Mika و همکارانش الگوریتم تفکیک-پذیری Fisher مبتنی بر هسته را بهمراه منظم‌سازی ماتریس واریانس درون کلاسی ارائه کردند [22]. Xiong و همکارانش الگوریتم KDA موثری پیشنهاد کردند که از آنالیز دو مرحله‌ای و تجزیه QR استفاده می‌کند [23]. Yang و همکارانش روش KPCA بهمراه LDA را پیشنهاد کردند که متمایزسازی را دو چندان می‌کند [24]. Xiong و همکارانش دو روش آنالیز تفکیک‌پذیری غیرهمبسته مبتنی بر هسته[۳۰] و آنالیز تفکیک‌پذیری متعامد مبتنی بر هسته[۳۱] ارائه داده‌اند که سعی در ناهمبستگی بیشتر دادگان و تعامد آنها دارند [25]. در روش‌های مبتنی بر هسته تعیین نوع هسته در عملکرد روش تبدیل ویژگی موثر است، در این راستا بررسی-هایی روی عملکرد توابع هسته مختلف در SVM و نیز سایر روشها صورت گرفته است [26]. انتخاب تابع هسته به روش تبدیل ویژگی مبتنی بر هسته و ماهیت دادگان بستگی دارد.

از منظر دیگری می‌توان روشهای تبدیل ویژگی را به دو دسته باناظر و بدون ناظر طبقه‌بندی نمود. در روش‌های باناظر اطلاعاتی از قبیل برچسب کلاس نمونه‌های آموزشی در اختیار قرار دارد، اما در روش‌های بدون ناظر هیچ اطلاعاتی از کلاس دادگان وجود ندارد. از روشهای باناظر می‌توان به LDA، MCE و از روشهای بدون ناظر می‌توان به PCA و KPCA اشاره نمود.

در این مقاله روش خطای کلاس‌بند کمینه مبتنی بر هسته[۳۲] (KMCE) ارائه شده است. همان گونه که گفته شد روش‌های تبدیل ویژگی عملکرد ضعیفی بر روی دادگانی که به طور خطی تفکیک‌پذیر نیستند، دارند. از اینرو روشهای غیرخطی مطرح شدند که شاخه‌ای از آنها روش‌های مبتنی بر هسته می‌باشند. از سویی دیگر روش خطای کلاس‌بند کمینه نسبت به روش‌های دیگر تبدیل ویژگی عملکرد

[24] Minimum Classification Error (MCE)
[25] Evolution Strategy (ES)
[26] Nonlinear Principal Component Analysis (NLPCA)
[27] Kernel Springy Discriminant Analysis (KSDA)
[28] Kernel Principal Component Analysis (KPCA)
[29] Kernel Discriminant Analysis (KDA)
[30] kernel Uncorrelated Discriminant Analysis (KUDA)
[31] Kernel Orthogonal Discriminant Analysis (KODA)
[32] Kernel Minimum Classification Error (KMCE)

بهتری دارد، چرا که معیار بهینه‌سازی آن کاهش نرخ خطای کلاس‌بندی است. با توجه به خواص مکانیزم نگاشت غیرخطی و روش خطای کلاس‌بند کمینه، در این مقاله روشی ارائه می شود تا با تلفیق این دو روش از خواص و مزیات هر دوِ آنها بهره جوید در روش پیشنهادی ابتدا ا یک نگاشت غیرخطی دادگان را به فضای با ابعاد بزرگتر می‌برد و سپس الگوریتم MCE در فضای جدید اعمال می‌شود. روش پیشنهادی یک روش باناظر و بدون کاهش ویژگی خواهد بود. عملکرد روش KMCE با روش‌های خطی MCE، LDA و PCA و روشهای غیرخطی KLDA و KPCA مقایسه شده است. آزمایشات بر روی دادگان مصنوعی، دادگان Iris، glass و vowel از مجموعه دادگان UCI [27] و دادگان گفتاری Aurora2 انجام شده است.

در ادامه مقاله، در بخش دوم روشهای تبدیل ویژگی خطی و غیرخطی بررسی می‌شوند. در بخش سوم روش پیشنهادی تبدیل ویژگی غیرخطی مبتنی بر هسته ارائه می‌شود. نتایج آزمایشات در بخش چهارم بیان می‌شوند. در نهایت بخش پنجم به نتیجه گیری اختصاص دارد.

۲. روشهای تبدیل ویژگی

مرحله استخراج ویژگی‌ها در سیستم‌های بازشناسی الگو یکی از قسمت‌های کلیدی و مؤثر در کارائی سیستم می‌باشد که دقت سیستم بازشناسی به عملکرد آن وابسته است. ویژگیها باید بتوانند میان کلاسها تمایز به وجود آورند . در مواردی که ویژگیها به تنهایی قادر به ایجاد تمایز کافی میان کلاسها نیستند، می‌توان از روش‌های تبدیل ویژگی برای ایجاد تمایز بیشتر میان کلاس‌ها استفاده کرد. در این بخش تعدادی از مهمترین روشهای تبدیل ویژگی خطی و غیرخطی مورد بررسی قرار می‌گیرند.

۱،۲. روشهای تبدیل ویژگی خطی

شاخه‌ای از روشهای تبدیلات ویژگی، روشهای خطی هستند که عملکرد قابل قبولی برای دادهایی با قابلیت تفکیک-پذیری خطی دارند. روش‌های PCA و LDA از شناخته شده ترین روشهایی این دسته هستند. روش MCE نیز روش تبدیل ویژگی دیگری است که خطی و باناظر می‌باشد و و هدف در آن بدست آوردن ماتریس تبدیلی است که خطای کلاس‌بندی را کاهش دهد. در این زیربخش سه روش PCA، LDA و MCE بررسی خواهند شد.

۱،۱،۲. آنالیز مولفه‌های اصلی

تبدیل خطی PCA داده ها را به فضایی نگاشت می کند که در آن بیشترین واریانس را داشته باشند. فرض کنید دادگان شامل N مشاهده، $X_k \in R^M$ ،k=1,…,N و $\sum_{k=1}^{N} X_k = 0$ باشد. ماتریس کوواریانس مجموعه دادگان از رابطه زیر بدست می‌آید: [28]

$$C = \frac{1}{N}\sum_{k=1}^{N} X_k X_k^T \qquad (۱)$$

با قطری‌سازی C مولفه‌های اصلی بدست می‌آیند که این مولفه‌ها تصویر متعامد روی بردارهای ویژه هستند که با حل معادله مقادیر ویژه محاسبه می‌شوند. [28]

$$\lambda V = CV \qquad (۲)$$

که $\lambda \geq 0$ مقادیر ویژه و $V \in R^M \backslash \{0\}$ (به معنی R^M به استثنای {0}) بردارهای ویژه می‌باشند. از طرفی $CV = \frac{1}{N}\sum_{k=1}^{N}(X_k.V)X_k$، همه پاسخها برای V بایستی ترکیب خطی از مشاهدات باشند، داریم: [28]

$$V = \sum_{k=1}^{N} \alpha_k X_k \qquad (۳)$$

که α_kها ضرایبی هستند که مقادیر آنها به گونه‌ای انتخاب می‌شوند که رابطه بالا برقرار باشد. [28]

۲،۱،۲. آنالیز تفکیک‌پذیری خطی

LDA یکی از روش‌های شناخته شده برای ایجاد تمایز میان ویژگیها و کاهش خطی ابعاد ویژگی[۳۳] باناظر می‌باشد. در LDA کلاسیک، ماتریس کوواریانس درون کلاسی و برون کلاسی بفرم زیر محاسبه می‌شوند. [29]

$$S_W = \frac{1}{N}\sum_{i=1}^{I}\sum_{n_i=1}^{N_i}\left(X_{n_i} - \mu_i\right)\left(X_{n_i} - \mu_i\right)^T \qquad (۴)$$

[33] Linear Dimensionality Reduction (LDR)

$$S_b = \frac{1}{N}\sum_{i=1}^{I} N_i(\mu_i - \mu)(\mu_i - \mu)^T \qquad (۵)$$

که N تعداد کل نمونه‌ها، N_i تعداد نمونه‌های کلاس iام، μ_i میانگین کلاس iام، I تعداد کلاسها و X_{n_i} نمونه n_iام از کلاس iام می‌باشند [30]. هدف LDA بدست آوردن ماتریس تبدیل، W است به نحوی که معیار Fisher را در معادله (۶) را بیشینه کند. [29]

$$J(W) = \frac{W^T trace(S_b) W}{W^T trace(S_w) W} \qquad (۶)$$

که در این رابطه تابع trace مجموع مقادیر روی قطر اصلی ماتریس مربعی در آرگمان ورودیش را محاسبه می نماید.

۲،۱،۳. خطای کلاس‌بند کمینه

روش کمینه نمودن خطای کلاس‌بندی (MCE)، روش متمایزسازی است که هم در گروه تبدیل ویژگی و هم در گروه آموزش کلاسبند بکار گرفته می‌شود [32 ,31]. هنگامی‌که روش MCE برای آموزش HMM (به عنوان یک کلاس‌بند) استفاده می‌شود، MCE پارامترهای مدل را بگونه‌ای تغییر می‌دهد که خطای کل کلاس‌بندی کاهش یابد [12]. در روش آموزش به کمک MCE، ابتدا تابع هدف (که برای یافتن پارامترهای HMM بکار می‌رود) با استفاده از یک تابع پیوسته مدل می‌شود. سپس، کمینه این تابع با یک روش مانند گرادیان احتمالی کاهشی[۳۴] بدست می‌آید [11].

ایده اصلی الگوریتم MCE، بهینه‌سازی نرخ خطای تجربی با استفاده از یک مجموعه داده آموزشی، جهت بهبود نرخ خطای بازشناسی می‌باشد. پس از آنکه نرخ خطای تجربی، بوسیله یک بازشناس یا کلاس‌بند، کمینه شد، تخمینی بایاس شده از نرخ خطای واقعی بدست می‌آید. یک راه مؤثر جهت کاهش این نرخ و بهبود کارآیی کلی سیستم، افزایش مرز میان کلاس‌ها در داده‌های آموزشی است. به این منظور از گرادیان کاهشی [11] یا ترکیب امتیازهای حاشیه و نرخ خطای تجربی استفاده شده است.

روش MCE برای یافتن ماتریس تبدیل ویژگی‌ها با هدف کمینه کردن خطای کلاس‌بندی نیز استفاده شده است [34]. Wang و Paliwal این روش را برای بازشناسی واکه‌ها بهبود بخشیدند [35]. در این رویکرد تابع هزینه برای کلاس kام بصورت زیر تعریف شده است:

$$d_k(O,F) = -g_k(O,F) + \frac{1}{\eta}\log\left[\frac{1}{M-1}\sum_{i=1,i\neq k}^{M}\exp\left(g_i(O,F)\eta\right)\right] \qquad (۷)$$

که در آن M تعداد کلاس‌ها، η مقداری مثبت (ارائه شده در [12])، و $g_i(O,F)$ لگاریتم احتمال تعلق O به کلاس λ_i است که بصورت زیر تعریف می‌شود:

$$g_i(O,F) = \log\, p(O \mid \lambda_i) \qquad (۸)$$

که F مجموعه پارامترهای مرتبط با ویژگیها و کلاس‌بندها می‌باشد. در روش MCE ، هدف کمینه نمودن تابع هزینه (۷) است. اما از آنجا که این توابع مشتق‌پذیر نیستند، از یک تابع پیوسته نرم مانند تابع سیگموئید بعنوان تابع هزینه استفاده می‌کنیم. بنابراین خواهیم داشت:

$$l_k(O,F) = \frac{1}{1+\exp\left(-\alpha d_k(O,F)\right)} \qquad (۹)$$

که $d_k(O,F)$ تابع هزینه (۷) می‌باشد و α پارامتر تنظیم بزرگتر از یک است. بنابراین اگر $d_k(O,F)$ خیلی کوچکتر از صفر باشد، در واقع یک کلاسبندی درست رخ داده و $l_k(O,F)$ به صفر نزدیک می‌شود. از طرفی مقدار مثبت برای $d_k(O,F)$ نشان‌دهنده جریمه برای کلاسبندی نادرست است و در اینصورت $l_k(O,F)$ به ۱ متمایل می‌شود. کل خطای کلاس‌بند برای مشاهده O از رابطه (۱۰) بدست می‌آید:

$$L = \sum_{k=1}^{M} l_k(O,F) \qquad (۱۰)$$

با استفاده از ماتریس تبدیل، بردارهای ویژگی از فضای ویژگی اولیه به فضای جدید برده می‌شود تا تمایز کلاس‌ها بیشتر گردد. تمایز بیشتر باعث کاهش خطا می‌شود. از اینرو با استفاده از معیار تابع هزینه سعی می‌شود ماتریس تبدیل بهینه بدست آید. مقادیر درایه‌های ماتریس تبدیل بروش

[34] Gradient Probabilistic Descent

تکراری با استفاده از گرادیان کاهشی برای تابع L، با رابطه (۱۱) قابل محاسبه خواهد بود.

$$w_{n,iter} = w_{n,iter-1} - \beta \frac{\partial L}{\partial w_n} \qquad (۱۱)$$

رابطه فوق را می‌توان بصورت ماتریسی بیان کرد:

$$\mathbf{W}_{iter} = \mathbf{W}_{iter-1} - \beta \frac{\partial L}{\partial \mathbf{W}} \qquad (۱۲)$$

که $\boldsymbol{W}$ماتریس تبدیل؛ n شاخص nامین درایه ماتریس تبدیل $\boldsymbol{W}$؛ $iter$ اندیس تکرار الگوریتم، و β پارامتر یادگیری می‌باشد. روابط (۱۱) و (۱۲) نشان‌دهنده رویه تکراری هستند. این رویه زمانی متوقف می‌شود که مقدار تابع هزینه از یک حد آستانه کمتر شود.

فرض کنید ماتریس تبدیل $\boldsymbol{W}$ بردار ویژگی‌های n بعدی x را به بردار y با بُعد d ($d \leq n$) تبدیل می‌کند. هدف، بدست آوردن ماتریس تبدیل $\boldsymbol{W}$ است که تابع هزینه را کمینه نماید. برای کمینه کردن خطا از رابطه (۹) نسبت به $\boldsymbol{W}$ مشتق گرفته می‌شود. به این ترتیب می‌توان نوشت:

$$\frac{\partial l_k(O,F)}{\partial W} = \frac{\partial l_k(O,F)}{\partial d_k(O,F)} \frac{\partial d_k(O,F)}{\partial W} \qquad (۱۳)$$

تابع هزینه $l_k(O,F)$ در رابطه (۹) تعریف شده است. بردار O رشته مشاهده یا رشته بردار ویژگی‌های تبدیل‌یافته است و بصورت $O=\{o_1,o_2,\ldots,o_T\}$ می‌باشد. همچنین بردار ویژگی‌های اصلی بصورت $X=\{x_1,x_2,\ldots,x_T\}$ نشان داده شده است. با توجه به رابطه (۹) خواهیم داشت[35] :

$$\frac{\partial l_k(O,F)}{\partial d_k(O,F)} = \alpha l_k(O,F)\big(1-l_k(O,F)\big) \qquad (۱۴)$$

برای محاسبه بخش دومِ رابطه (۱۳)، $\frac{\partial d_k(O,F)}{\partial W}$ ، از رابطه (۷) بر حسب W مشتق‌گیری می‌شود:

$$\frac{\partial d_k(O,F)}{\partial W} = -\frac{\partial g_k(O,F)}{\partial W} + \sum_{i=1,i\neq k}^{M} \left(\frac{\exp\big(g_i(O,F)\eta\big)}{\sum_{j=1,j\neq k}^{M} \exp\big(g_j(O,F)\eta\big)} \times \frac{\partial g_i(O,F)}{\partial W} \right) \qquad (۱۵)$$

در این رابطه $\frac{\partial g_j(O,F)}{\partial W}$ با توجه به نوع کلاس‌بند محاسبه می‌شود. به عنوان مثال نحوه محاسبه برای کلاس‌بند مبتنی بر فاصله بصورت زیر است.

$$\frac{\partial g_i(O,F)}{\partial w_d} = 2w_d\big(O(d) - y_i(d)\big)^2 \qquad (۱۶)$$

که $y_i(d)$ درایه dام مرکز کلاس iام می‌باشد. با جایگذاری (۱۶) در رابطه (۱۵) مقدار $\frac{\partial d_k(O,F)}{\partial W}$ بدست می‌آید. با جایگزین کردن رابطه محاسبه شده $\frac{\partial d_k(O,F)}{\partial W}$ و رابطه (۱۴) در (۱۳) مقدار $\frac{\partial l_k(O,F)}{\partial W}$ بدست می‌آید. با بکارگیری مقادیر $\frac{\partial l_k(O,F)}{\partial W}$ در رابطه (۱۲) ماتریس بهینه بدست می‌آید.

۲،۲. روشهای تبدیل ویژگی غیرخطی

همانطور که گفته شد می‌توان روش‌های تبدیل ویژگی را به دو دسته روشهای خطی و غیرخطی تقسیم نمود. علت پیدایش روشهای غیرخطی عملکرد ضعیف روشهای تبدیل ویژگی خطی برای دادگانی است که به طور خطی تفکیک‌پذیر نیستند. روشهای غیرخطی به این صورت عمل می‌کنند که به کمک یک نگاشت غیرخطی داده‌ها را به فضای جدید با ابعاد زیاد می برند که در این فضا دادگان به طور خطی تفکیک‌پذیر می‌شوند [37]. بعضی از روشهای غیرخطی با کمک شبکه عصبی نگاشت غیرخطی ایجاد می‌کنند و سپس تبدیلی خطی بکار می‌برند همانند روش NLPCA. دسته دیگر از روشهای غیرخطی روشهای مبتنی بر هسته می‌باشند. در روش‌های تبدیل غیرخطی، نگاشت غیرخطی در دست نمی‌باشد و با روشهایی آن را تخمین می‌زنند. روش‌های مبتنی بر هسته بر این ایده استوارند که ضرب داخلی دادگان نگاشت یافته با یک تابع غیرخطی نامعلوم را می‌توان با تابع هسته تخمین زد، که به آن ایده هسته[35] گفته می‌شود. در ادامه ابتدا ایده هسته و سپس دو روش KLDA و KPCA بررسی خواهند شد.

[35] Kernel Trick

۲،۲،۱. ایده تابع هسته

مجموعه X را در نظر بگیرید تابع هسته k به صورت زیر تعریف می شود.

$$k : X \times X \to \Re \qquad (۱۷)$$

از ویژگیهای تابع هسته می‌توان متقارن بودن و مثبت بودن را نام برد، که در روابط (۱۸) و (۱۹) این مطلب بیان شده است.

$$k(x,y) = k(y,x) \quad \forall x, y \in X \qquad (۱۸)$$

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0, \quad \forall c_1, c_2, ..., c_n \in \Re, x_1, x_2, ..., x_n \in X \qquad (۱۹)$$

می‌توان نشان داد که برای یک تابع هسته، k، فضای هیلبرت H و یک تبدیل $\phi : X \to H$ وجود دارد که رابطه زیر برقرار باشد:

$$k(x,y) = \langle \phi(x), \phi(y) \rangle = \phi(x).\phi(y) \qquad (۲۰)$$

۲،۲،۲. آنالیز مولفه‌های اصلی مبتنی بر هسته

تکنیک اصلی KPCA محاسبه تبدیل PCA در فضای نگاشت یافته توسط یک تابع نگاشت غیرخطی است که از ایده هسته برای تخمین این نگاشت استفاده می‌شود. داده-های نگاشت یافته $\varphi(X_1), \dots, \varphi(X_N)$ را در نظر بگیرید که میانگین آنها صفر نباشد. ابتدا با رابطه زیر میانگین داده‌های نگاشت یافته صفر می‌شود:

$$\widehat{\varphi}(X_k) = \varphi(X_k) - \frac{1}{N}\sum_{j=1}^{N} \varphi(X_j) \qquad (۲۱)$$

ماتریس کوواریانس از رابطه (۲۲) محاسبه می‌شود:

$$\widehat{\Sigma} = \frac{1}{N}\sum_{j=1}^{N} \widehat{\varphi}(X_j)\widehat{\varphi}(X_j)^T \qquad (۲۲)$$

معادله مقادیر ویژه برای ماتریس کوواریانس فوق بفرم $\widehat{\lambda}\widehat{V} = \widehat{\Sigma}\widehat{V}$ می‌باشد. که $\widehat{\lambda} \geq 0$ مقادیر ویژه و $\widehat{V} \in F\backslash\{0\}$ (به معنی F به استثنای $\{0\}$) بردارهای ویژه می‌باشند. معادل معادله مقادیر ویژه را می‌توان به صورت رابطه (۲۳) نوشت:

$$\widehat{\lambda}\left(\widehat{\varphi}(X_k).\widehat{V}\right) = \left(\widehat{\varphi}(X_k).\widehat{\Sigma}\widehat{V}\right), \mathrm{k} = 1, \dots, \mathrm{N} \qquad (۲۳)$$

$\hat{\alpha}_k$ها ضرایبی هستند که مقادیر آنها به گونه‌ای انتخاب می-شوند که رابطه (۲۴) برقرار باشد.

$$\hat{V} = \sum_{k=1}^{N} \hat{\alpha}_k \widehat{\varphi}(X_k) \qquad (۲۴)$$

که با جایگذاری رابطه (۲۴) در رابطه (۲۳) خواهیم داشت:

$$\widehat{\lambda}\alpha = K\alpha, \ (\alpha = (\alpha_1, \dots, \alpha_N)^T) \qquad (۲۵)$$

که K ماتریس هسته بوده و به فرم ماتریس مربعی $N \times N$ با عناصر $\mathrm{K_{i,j}} = \left(\varphi(\mathrm{X_i}), \varphi(\mathrm{X_j})\right) = \mathrm{k}(\mathrm{X_i}, \mathrm{X_j})$ می‌باشد. برای نرمال‌سازی راه‌حل‌های (λ_k, α^k)، بایستی رابطه $\lambda_k(\alpha^k.\alpha^k) = 1$ در فضای نگاشت یافته اعمال شود. همچنین، همانند هر الگوریتم PCA دیگری، دادگان در فضای نگاشت یافته باید متمرکز[۳۶] شوند. برای این کار ماتریس هسته با رابطه زیر جایگزین می‌شود،

$$\widehat{K} = K - 1_N K - K 1_N + 1_N K 1_N \qquad (۲۶)$$

که $(1_N)_{i,j} = 1/\mathrm{N}$ می‌باشد. [19]

۲،۲،۳. آنالیز تفکیک‌پذیری خطی مبتنی بر هسته

فرض کنید مشاهدات، X از طریق تابع $\Phi(.)$ به فضای جدید نگاشت داده شوند، ماتریس کوواریانس درون کلاسی و برون کلاسی در فضای نگاشت یافته بفرم زیر محاسبه می-شوند. [36]

$$S_W^{\varphi} = \frac{1}{N}\sum_{i=1}^{I}\sum_{n_i=1}^{N_i}\left(\varphi(X_{n_i}) - \mu_i^{\varphi}\right)\left(\varphi(X_{n_i}) - \mu_i^{\varphi}\right)^T \qquad (۲۷)$$

$$S_b^{\varphi} = \frac{1}{N}\sum_{i=1}^{I} N_i\left(\mu_i^{\varphi} - \mu^{\varphi}\right)\left(\mu_i^{\varphi} - \mu^{\varphi}\right)^T \qquad (۲۸)$$

که N تعداد کل نمونه‌ها، N_i تعداد نمونه‌های کلاس iام، μ_i^{φ} میانگین کلاس iام در فضای نگاشت یافته، I تعداد کلاسها و $\varphi(X_{n_i})$ نمونه n_iام از کلاس iام در فضای نگاشت یافته می-باشند. هدف KFDA ماکزیمم کردن رابطه زیر می‌باشد:

$$J(W) = \frac{W^T S_b^{\varphi} W}{W^T S_W^{\varphi} W} \qquad (۲۹)$$

اما ماتریس تبدیل W را می‌توان به صورت میانگین وزنی داده‌های آموزشی نگاشت یافته نوشت،

$$W = \sum_{i=1}^{I}\sum_{n_i=1}^{N_i} \alpha_{n_i}\varphi(X_{n_i}) = \Phi\alpha \qquad (۳۰)$$

[36] To be centered

بنابراین رابطه Fisher به صورت زیر خواهد شد:

$$J(W)=\frac{\alpha^T\Phi^T S_b^{\varphi}\Phi\alpha}{\alpha^T\Phi^T S_W^{\varphi}\Phi\alpha}=\frac{\alpha^T P\alpha}{\alpha^T Q\alpha}=J(\alpha) \qquad (۳۱)$$

که P و Q از طریق توابع هسته قابل محاسبه می‌باشند. روابط زیر نحوه محاسبه P و Q را نشان می‌دهند. [36]

$$P=K(R-\hat{1})K \qquad (۳۲)$$

$$Q=K(I-R)K \qquad (۳۳)$$

که K ماتریس هسته، $[\mathrm{K}]_{i,j}=\varphi(\mathrm{X_i}).\varphi(\mathrm{X_j})=\mathrm{k}(\mathrm{X_i},\mathrm{X_j})$ و تابع $k(.)$، تابع هسته می‌باشد. و I ماتریس همانی، $[\hat{1}]_{i,j}=1/\mathrm{N}$ که در آن N تعداد کل مشاهدات می‌باشد. و ماتریس R بفرم زیر محاسبه می‌شود.

$$[R]_{i,j}=\begin{cases}\frac{1}{\mathrm{N_j}}, & if\ i=j\\ 0, & \text{otherwise.}\end{cases} \qquad (۳۴)$$

که N_j تعداد مشاهدات مربوط به کلاس jام است.

۳. روش پیشنهادی - خطای کلاس‌بند کمینه مبتنی بر هسته (KMCE)

روش پیشنهادی، KMCE، غیرخطی سازی روش MCE از طریق ایده هسته است. این روش از سویی دارای خاصیت MCE یعنی انطباق معیار تخمین تبدیل با خطای کلاس-بندی می‌باشد. و از سوی دیگر دارای خاصیت نگاشت غیرخطی روشهای مبتنی بر هسته است. در این راستا فرض می‌شود که دادگان با یک تابع غیرخطی به فضای با ابعاد بزرگتر نگاشت داده می‌شوند سپس الگوریتم MCE روی دادگان نگاشت یافته اعمال می‌شود. با توجه به نامعلوم بودن تابع نگاشت بایستی تمامی روابط به گونه‌ای بازنویسی شوند که بتوان ضربهای نقطه‌ای نمونه‌های نگاشت یافته را با تابع هسته جایگزین نمود. در ادامه الگوریتم پیشنهادی فرموله می‌شود.

فرض کنید مشاهده X از طریق تابع $\Phi(.)$ به فضای جدید نگاشت داده شود که بفرم زیر نمایش داده می‌شود:

$$X^{\Phi}=\Phi(X) \qquad (۳۵)$$

همچنین فرض کنید که با ماتریس تبدیل W داده‌های نگاشت یافته به $X^{\Phi,W}$ تبدیل شوند.

$$X^{\Phi,W}=W\,\Phi(X)$$
$$W=\sum_{i=1}^{N}\alpha_i X_i^{\Phi} \qquad (۳۶)$$

از طرفی W بفرم $W=\sum_{i=1}^{N}\alpha_i X_i^{\Phi}$ تعریف می شود که N تعداد کل نمونه های آموزشی و X_i^{Φ} نمونه iام داده آموزشی نگاشت یافته می باشد. بنابراین رابطه (۳۶) به صورت زیر در می‌آید:

$$\begin{aligned}X^{\Phi,W}&=WX^{\Phi}=\left(\sum_{n=1}^{N}\alpha_n(X_n^{\Phi})^T\right)X^{\Phi}=\\ &\sum_{n=1}^{N}\alpha_n\left((X_n^{\Phi})^T X^{\Phi}\right)=\sum_{n=1}^{N}\alpha_n(X_n^{\Phi}.X^{\Phi})=\\ &\sum_{n=1}^{N}\alpha_n\left(\Phi(X_n).\Phi(X)\right)=\\ &\sum_{n=1}^{N}\alpha_n k(X_n,X)=X^{\Phi,\alpha}\end{aligned} \qquad (۳۷)$$

روش خطای کلاس‌بندی کمینه سعی در بدست آوردن تبدیلی دارد که خطای کلاس‌بندی را کاهش دهد. تابع هزینه در فضای نگاشت یافته بازنویسی شده است.

$$\begin{aligned}&d_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})=-g_k(X_{n_k}^{\Phi,\alpha},y_k^{\Phi,\alpha})+\\ &\frac{1}{\eta}\log\left[\frac{1}{I-1}\sum_{i=1,i\neq k}^{I}\exp\left(g_i(X_{n_k}^{\Phi,\alpha},y_i^{\Phi,\alpha})\eta\right)\right]\end{aligned} \qquad (۳۸)$$

که I تعداد کلاسها، η عدد مثبتی است. همچنین $X_{n_k}^{\Phi,\alpha}$ مشاهده‌ی نگاشت یافته و تبدیل یافته n_k از کلاس kام می-باشد. $g_i\left(X_{n_k}^{\Phi,\alpha},y_i^{\Phi,\alpha}\right)$ برابر با مربع فاصله $X_{n_k}^{\Phi,\alpha}$ با نماینده کلاس iام در فضای نگاشت شده و تبدیل یافته یعنی $y_i^{\Phi,\alpha}$ می‌باشد که به صورت رابطه زیر بیان می‌گردد.

$$\begin{aligned}y_i^{\Phi,\alpha}&=\frac{1}{N_i}\sum_{n_i=1}^{N_i}X_{n_i}^{\Phi,\alpha}=\\ &\frac{1}{N_i}\sum_{n_i=1}^{N_i}\sum_{n=1}^{N}\alpha_n k(X_n,X_{n_i})\end{aligned} \qquad (۳۹)$$

با توجه به استفاده نرم اقلیدسی و رابطه (۳۹) داریم:

$$\begin{aligned}&g_i\left(X_{n_k}^{\Phi,\alpha},y_i^{\Phi,\alpha}\right)=\left\|X_{n_k}^{\Phi,\alpha}-y_i^{\Phi,\alpha}\right\|^2=\\ &\left(\sum_{n=1}^{N}\alpha_n k(X_n,X_{n_k})-\frac{1}{N_i}\sum_{n_i=1}^{N_i}\sum_{n=1}^{N}\alpha_n k(X_n,X_{n_i})\right)\\ &\left(\sum_{n=1}^{N}\alpha_n k(X_n,X_{n_k})-\frac{1}{N_i}\sum_{n_i=1}^{N_i}\sum_{n=1}^{N}\alpha_n k(X_n,X_{n_i})\right)^T\end{aligned} \qquad (۴۰)$$

با بسط ضرب ماتریسی فوق بفرم عددی خواهیم داشت:

$$g_i\left(X_{n_k}^{\Phi,\alpha}, y_i^{\Phi,\alpha}\right)=\left\|X_{n_k}^{\Phi,\alpha}-y_i^{\Phi,\alpha}\right\|^2= \sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}k(X_n,X_{n_k})-\frac{1}{N_i}\sum_{n_i=1}^{N_i}\sum_{n=1}^{N}\alpha_{nm}k(X_n,X_{n_i})\right)^2 \quad (۴۱)$$

هدف روش خطای کلاس‌بندی کمینه این است که توابع هزینه مینیمم شوند. ولی این توابع مشتق‌پذیر نیستند، از اینرو از تابع پیوسته نرم مانند تابع سیگموئید بعنوان تابع هزینه استفاده می‌گردد، خواهیم داشت:

$$l_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})=\frac{1}{1+\exp\left(-\gamma d_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})\right)} \quad (۴۲)$$

که $d_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})$ تابع هزینه (۳۸) می‌باشد و γ پارامتر تنظیم بزرگتر از یک است. بنابراین اگر $d_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})$ خیلی کوچکتر از صفر باشد، در واقع یک کلاس‌بندی درست داریم و $l_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})$ به صفر نزدیک می‌شود. از طرفی مقدار مثبت برای $d_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})$ نشان دهنده جریمه برای کلاس‌بندی نادرست است که در اینصورت $l_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})$ به یک متمایل می‌شود. کل خطای کلاس‌بندی برای مشاهده از رابطه (۴۳) بدست می‌آید:

$$L=\sum_{k=1}^{I}\sum_{n_k=1}^{N_k}l_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha}) \quad (۴۳)$$

که I تعداد کلاسها، N_k تعداد مشاهدات کلاس kام و $N=\sum_{k=1}^{I}N_k$ کل مشاهدات است. با استفاده از ماتریس تبدیل بردارهای ویژگی را از فضای ویژگی اولیه به فضای جدید برده تا تمایز کلاسها بیشتر شود. تمایز بیشتر باعث کاهش خطا می‌شود. از همینرو با استفاده از معیار تابع هزینه سعی در بدست آوردن ماتریس تبدیل بهینه داریم. مقادیر درایه‌های ماتریس بر طبق رابطه (۳۶) وابسته به مقادیر α هستند که با تبدیل بروش تکراری با استفاده از گرادیان کاهشی با رابطه زیر محاسبه می‌شود:

$$\alpha_{pq,iter}=\alpha_{pq,iter-1}-\beta\frac{\partial L}{\partial\alpha_{pq}} \quad ;\ p=1,2,\ldots,N, q=1,2,\ldots,M \quad (۴۴)$$

هدف بدست آوردن ماتریس تبدیل بهینه‌ای است که تابع هزینه را کاهش دهد. از اینرو از تابع هزینه بر اساس α مشتق می‌گیریم.

$$\frac{\partial l_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})}{\partial\alpha_{pq}}= \frac{\partial l_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})}{\partial d_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})}\frac{\partial d_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})}{\partial\alpha_{pq}} \quad (۴۵)$$

با توجه به رابطه (۴۲) داریم:

$$\frac{\partial l_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})}{\partial d_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})}= \gamma l_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})\left(1-l_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})\right) \quad (۴۶)$$

برای محاسبه ترم دوم در رابطه (۴۵)، $\frac{\partial d_k(X_{n_k}^{\Phi,\alpha},y^{\Phi,\alpha})}{\partial\alpha_{pq}}$ ، از رابطه (۳۸) استفاده می‌شود. بعد از مشتق‌گیری از روابط فوق می‌توان نوشت:

$$\frac{\partial d_k(X_{n_k}^{\Phi,\alpha},Y^{\Phi,\alpha})}{\partial\alpha_{pq}}=-\frac{\partial g_k(X_{n_k}^{\Phi,\alpha},y_k^{\Phi,\alpha})}{\partial\alpha_{pq}}+ \sum_{i=1,i\neq k}^{I}\left\{\frac{\exp\left(g_i(X_{n_k}^{\Phi,\alpha},y_i^{\Phi,\alpha})\eta\right)}{\sum_{j=1,j\neq k}^{I}\exp\left(g_j(X_{n_k}^{\Phi,\alpha},y_j^{\Phi,\alpha})\eta\right)}\times\frac{\partial g_i(X_{n_k}^{\Phi,\alpha},y_i^{\Phi,\alpha})}{\partial\alpha_{pq}}\right\} \quad (۴۷)$$

مقدار $\frac{\partial g_i(X_{n_k}^{\Phi,\alpha},y_i^{\Phi,\alpha})}{\partial\alpha_{pq}}$ از رابطه زیر محاسبه می‌گردد.

$$\frac{\partial g_i(X_{n_k}^{\Phi,\alpha},y_i^{\Phi,\alpha})}{\partial\alpha_{pq}}= 2\left(k(X_p,X_{n_k})-\frac{1}{N_i}\sum_{n_i=1}^{N_i}k(X_p,X_{n_i})\right)\left(\sum_{n=1}^{N}\alpha_{nq}k(X_n,X_{n_k})-\frac{1}{N_i}\sum_{n_i=1}^{N_i}\sum_{n=1}^{N}\alpha_{nq}k(X_n,X_{n_i})\right) \quad (۴۸)$$

با جایگذاری (۴۸) در رابطه (۴۷) مقدار $\frac{\partial d_k(X_{n_k}^{\Phi,\alpha},y^{\Phi,\alpha})}{\partial\alpha}$ بدست می‌آید. با جانشینی رابطه محاسبه شده و رابطه (۴۶) در (۴۵) مقدار

$\frac{\partial l_k(X_{n_k}^{\Phi,\alpha}, y^{\Phi,\alpha})}{\partial \alpha}$ بدست می‌آید. با بکارگیری مقادیر $\frac{\partial l_k(X_{n_k}^{\Phi,\alpha}, y^{\Phi,\alpha})}{\partial \alpha}$ در رابطه (۴۴) ماتریس بهینه بدست می‌آید.

۴. ارزیابی روش پیشنهادی KMCE

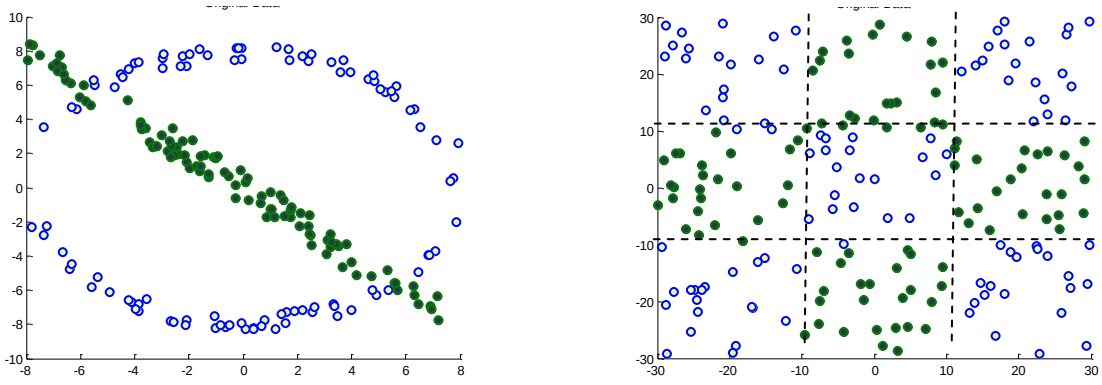
در این بخش به مقایسه و ارزیابی روش پیشنهادی با روش‌های MCE، LDA، PCA، KLDA و KPCA پرداخته خواهد شد. نتایج آزمایشات در سه زیر بخش ارائه می‌شود. در زیربخش اول عملکرد روشها در فضای دو بعدی بررسی می‌شود. و در زیربخش دوم نتایج کلاس‌بندی پس از اعمال هر یک از روشهای پیش پردازش مطرح شده در مقاله روی دادگان UCI گزارش می‌گردد. و در زیربخش سوم نتایج بازشناسی گفتار گسسته روی دادگان Aurora2 آورده شده است.

۱،۴. ارزیابی روش KMCE در فضای دو بعدی

ابتدا روش‌های تبدیل ویژگی بر روی دو مجموعه داده دو بعدی تعریف شده با نامهای CircleLine و Chess بفرم شکل‌های (۱-الف) و (۱-ب) اعمال می‌شوند. این مجموعه دادگان دارای ساختاری به فرم جدول (۱) می‌باشند. در این دو مجموعه داده میانگین و واریانس هر ویژگی برای هر دو کلاس یکسان می‌باشد. همانطور که در شکل‌های (۱-الف) و (۱-ب) مشاهده می‌شود، هر دو دادگان تفکیک‌پذیر خطی نیستند. از اینرو انتظار می‌رود که روشهای متمایزساز خطی عملکرد ضعیفی داشته باشند.

جدول ۱ – ساختار دادگان دوبعدی CircleLine و Chess

نام دادگان	تعداد کلاس	تعداد ویژگی	تعداد نمونه آموزشی	تعداد نمونه تست
CircleLine	۲	۲	۴۰۰	۲۰۰
Chess	۲	۲	۴۰۰	۲۰۰



الف – دادگان Chess

ب – دادگان CircleLine

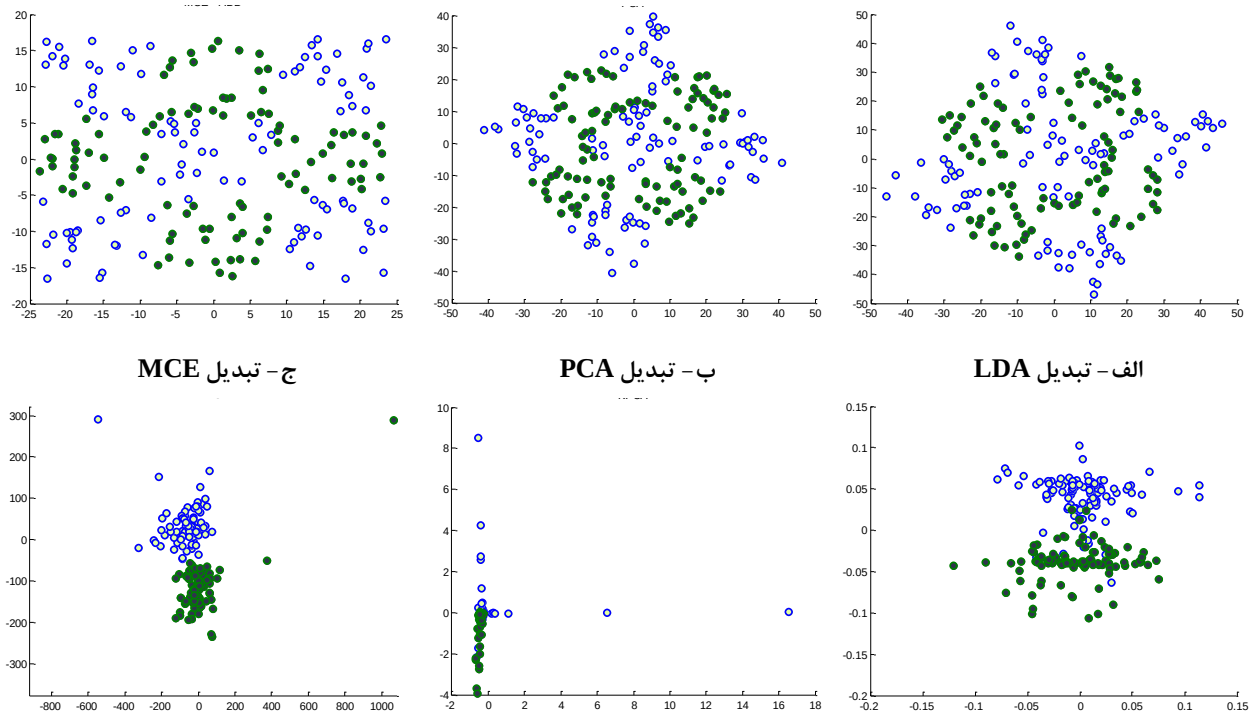
شکل ۱ – نمایش دوبعدی مجموعه تست دادگان Chess و CircleLine در فضای ویژگی

شکل (۲) تبدیل یافته‌های دادگان Chess را پس از اعمال الگوریتم‌های پیش‌پردازش نشان می‌دهد. در تمام روشها از دادگان آموزش برای بدست آوردن ماتریس تبدیل ویژگی استفاده شده است، سپس ماتریس تبدیل ویژگی بدست آمده روی دادگان تست اعمال می‌شود. شکل (۲-الف) فضای دوبعدی دادگان بعد از اعمال روش LDA را نشان می‌دهد. با توجه به فرض اولیه LDA مبنی بر تفکیک‌پذیری خطی دادگان و عدم تطبیق این فرض با دادگان Chess این روش عملکرد مطلوبی نداشته و صرفاً باعث دوران دادگان شده است.

شکل (۲-ب) فضای دوبعدی ویژگیها را بعد از اعمال الگوریتم PCA نمایش می دهد. روش PCA بدون توجه به برچسب کلاس نمونه‌های آموزشی سعی در بدست آوردن تبدیلی دارد که واریانس ویژگی‌ها در راستای آن ماکزیمم گردد، لذا با توجه به ساختار دادگان Chess واریانس در راستای قطر دادگان ماکزیمم است که انتظار می‌رود تبدیل PCA دادگان را دوران داده به نحوی که قطرها در راستای محورهای جدید ویژگی قرار گیرند. بنابراین مشاهده می-گردد که تبدیل PCA نیز عملکرد ضعیفی دارد.

شکل (۲-ج) پراکندگی نمونه‌ها پس از تبدیل MCE را نشان می‌دهد. این تبدیل سعی در کاهش خطای کلاس‌بندی دارد اما یک تبدیل خطی است و همانطور که در شکل دیده می-شود عملکرد ضعیفی برای دادگان Chess دارد. شکل (۲-د) تبدیل یافته از روش KLDA و شکل (۲-ه) تبدیل یافته از روش KPCA را نشان می‌دهند. در هر دو روش از تابع هسته گوسین با پارامتر ۲ استفاده شده است. همانطور که مشاهده می‌شود روش‌های مبتنی بر هسته نسبت به روشهای معمولی عملکرد بهتری داشته و پراکندگی دادگان به گونه است که دو کلاس توسط یک کلاس‌بند خطی تفکیک‌پذیر بوده و خطای کمتری نسبت به روشهای تبدیل معمولی حاصل می‌شود.

شکل (۲-و) نمونه‌های تبدیل یافته در فضای ویژگی‌های جدید پس از اعمال روش پیشنهادی، KMCE را نشان می-دهد. تابع هسته بکار رفته در این روش نیز تابع گوسین با پارامتر ۲ می‌باشد. با توجه به شکل می‌توان دید که عملکرد روش پیشنهادی نسبت به روش MCE بسیار بهتر بوده، همچنین نسبت به روش تبدیلات مبتنی بر هسته دیگر مثل KLDA و KPCA نیز عملکرد مناسبتری داشته است. این امر با توجه به الگوریتم روش پیشنهادی و معیاری که تبدیل بر اساس آن بدست می‌آید قابل انتظار است. از سوئی روش پیشنهادی یک روش غیرخطی بوده بنابراین برتری روشهای تبدیل ویژگی مبتنی بر هسته را در بردارد و از سوی دیگر معیاری که براساس آن ماتریس تبدیل تخمین زده می‌شود مستقیماً خطای کلاس‌بندی است.



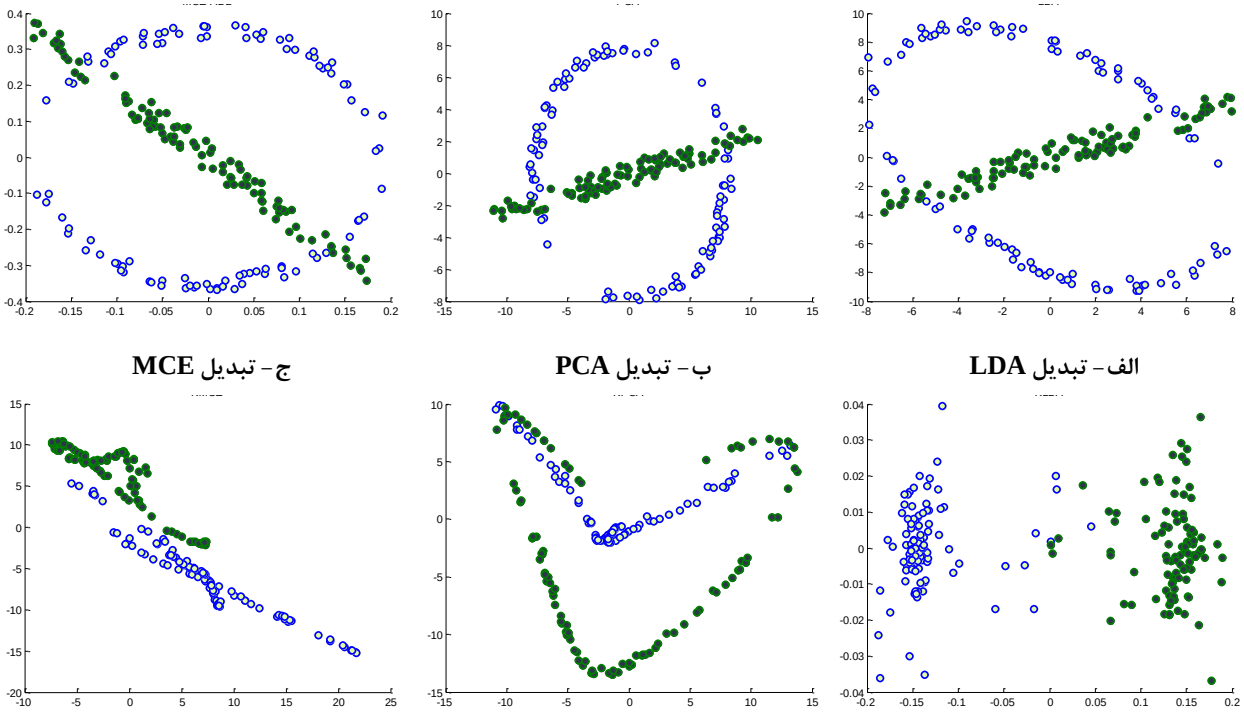
الف- تبدیل LDA ب- تبدیل PCA ج- تبدیل MCE

د- تبدیل KLDA ه- تبدیل KPCA و- تبدیل KMCE

شکل ۲ – نمایش دوبعدی مجموعه تست دادگان Chess بعد از اعمال تبدیلات ویژگی LDA، PCA، MCE، KLDA، KPCA و KMCE

شکل (۳) تبدیل یافته‌های دادگان CircleLine را پس از اعمال الگوریتم‌های پیش‌پردازش نشان می‌دهد. در تمامی روشهای تبدیل ویژگی از دادگان آموزش برای بدست آوردن ماتریس تبدیل ویژگی استفاده شده است و این ماتریس روی دادگان تست اعمال گردیده است. شکل‌های (۳-الف)، (۳-ب) و (۳-ج) بترتیب فضای دوبعدی دادگان بعد از اعمال روش‌های LDA، PCA و MCE را نشان می‌دهند. در LDA فرض بر تفکیک‌پذیر خطی بودن دادگان است و با توجه به ماهیت دادگان CircleLine یعنی تفکیک‌پذیر بودن غیرخطی، این روش عملکرد خوبی ندارد. روش PCA که یک روش بدون ناظر است تبدیل ویژگی‌ای بدست می‌آورد که واریانس ویژگی‌های جدید بیشتر شود، با توجه به پراکندگی دادگان در فضای ویژگی‌ها، واریانس در تمامی جهات یکسان است لذا PCA نمی‌تواند تبدیل بهینه‌ای بدست آورد. روش MCE ماتریس تبدیلی را بدست می‌آورد که خطای کلاس-بندی را کاهش دهد. روش MCE یک تبدیل خطی بدست می‌آورد و عملکرد ضعیفی برای دادگانی مانند CircleLine که تفکیک پذیر خطی نیستند، دارد.

شکل (۳-د) نمونه‌های تبدیل یافته از روش KLDA و شکل (۳-ه) نمونه‌های تبدیل یافته از روش KPCA را نشان می-دهند. همانطور که مشاهده می‌شود روش‌های مبتنی بر هسته نسبت به روشهای معمولی عملکرد بهتری داشته و پراکندگی دادگان به گونه است که دو کلاس توسط یک کلاس‌بند خطی تفکیک‌پذیر بوده و خطای کمتری نسبت به روشهای تبدیل معمولی حاصل می‌شود.



الف- تبدیل LDA ب- تبدیل PCA ج- تبدیل MCE

د- تبدیل KLDA ه- تبدیل KPCA و- تبدیل KMCE

شکل ۳ – نمایش دوبعدی مجموعه تست دادگان CircleLine بعد از اعمال تبدیلات ویژگی LDA، PCA، MCE، KLDA، KPCA و KMCE

۲,۴. ارزیابی روش KMCE به عنوان پیش پردازش در کلاس‌بندی

جهت بررسی بهتر عملکرد روش پیشنهادی نسبت به دیگر روش‌های تبدیل ویژگی، این تبدیلات را بعنوان پیش پردازش در کلاس‌بندی دادگان نگاشت یافته با هر یک از روشهای فوق بکار برده و نتایج کلاس‌بندی مورد ارزیابی قرار خواهد گرفت. برای انجام آزمایشات علاوه بر دو مجموعه دادگان Chess و CircleLine از مجموعه دادگان Iris، Glass

و Vowel مربوط به دادگان UCI [27] نیز استفاده شده است. مشخصات مجموعه دادگان در جدول (۲) آورده شده است. برای کلاس‌بندی از ابزار Weka استفاده شده است. کلاس‌بندهای مورد استفاده در این بخش در سه گروه طبقه بندی می‌شوند: گروه کلاس‌بندهای مبتنی بر فاصله مانند IB1، IBK و KStar، گروه کلاس‌بندهای آماری مانند BayesNet، گروه کلاس‌بندهای مبتنی بر درخت تصمیم-گیری مانند NBTree.

جدول (۳) نتایج کلاس‌بندی را روی کلاس‌بندهای مختلف و دادگان مختلف نشان می‌دهد. ستون‌های جدول نشان‌دهنده روش تبدیل ویژگی و سطرها نشان‌دهنده متد کلاس‌بند می-باشد. روش پیشنهادی یعنی KMCE در کنار روشهای LDA، PCA و MCE به عنوان روشهای تبدیل ویژگی خطی و KLDA و KPCA بعنوان روشهای تبدیل ویژگی غیرخطی مورد ارزیابی قرار گرفته است. ستون Normal نتایج کلاس-بندی را برای دادگانی که هیچ تبدیلی روی آنها انجام نگرفته است را نشان می‌دهد. اعداد مندرج در جدول نتایج کلاس-بندی دادگان نگاشت یافته را به درصد نشان می‌دهد. در هر سطر بیشترین مقدار بفرم ضخیم بهمراه خط زیر و مقدار دوم بفرم فقط ضخیم نمایش داده شده است.

جدول ۲ – مشخصات مجموعه دادگان UCI [27]

نام دادگان	تعداد کلاس	تعداد ویژگی	تعداد نمونه آموزشی	تعداد نمونه تست
Iris	۳	۴	۹۰	۶۰
Vowel	۱۱	۱۰	۵۲۸	۴۶۲
Glass	۶	۹	۱۲۷	۸۷

جدول ۳ – نتایج کلاس بندهای مختلف روی دادگان نگاشت یافته با تبدیلات مختلف

			Normal	LDA	PCA	MCE	KLDA	KPCA	KMCE
کلاس بند مبتنی بر فاصله	IB1	Chess	93.00	93.00	92.00	93.00	92.50	71.50	**97.00**
		CircleLine	98.00	98.00	97.50	98.00	97.00	96.50	**100.00**
		Iris	95.00	93.33	95.00	95.00	96.67	96.67	**98.33**
		Vowel	49.57	44.16	41.13	49.57	53.27	57.58	**58.44**
		Glass	66.67	67.82	72.41	66.67	68.97	71.26	**74.71**
	IBK,3	Chess	92.50	93.00	93.00	92.50	92.50	71.50	**97.50**
		CircleLine	96.50	97.00	96.50	97.00	97.00	96.50	**100.00**
		Iris	95.00	95.00	95.00	95.00	96.67	**98.33**	**98.33**
		Vowel	49.13	42.86	39.39	49.13	54.33	58.23	**59.74**
		Glass	65.52	65.52	71.26	65.52	74.71	65.52	**77.01**
	KStar	Chess	94.00	92.50	90.00	94.00	92.00	75.50	**99.00**
		CircleLine	97.00	95.50	95.50	97.00	97.50	94.50	**99.50**
		Iris	95.00	96.67	93.33	95.00	96.67	96.67	**98.33**
		Vowel	45.24	36.58	38.10	50.22	50.64	53.03	**57.80**
		Glass	**77.01**	63.22	68.97	**77.01**	66.67	72.41	**77.01**
کلاس بند آماری	BayesNet	Chess	50.00	73.50	82.50	50.00	**90.50**	68.50	**94.50**
		CircleLine	76.00	88.00	88.00	78.00	**97.00**	81.00	**88.50**
		Iris	91.67	**95.00**	**96.67**	**96.67**	**95.00**	**95.00**	**96.67**
		Vowel	36.58	41.56	37.01	36.36	**44.81**	39.61	**42.86**
		Glass	**68.97**	43.68	64.37	**68.97**	**66.67**	51.72	**66.67**
کلاس بند مبتنی بر درخت تصمیم	NBTree	Chess	**97.50**	88.50	82.00	**97.50**	90.50	68.50	**94.50**
		CircleLine	95.00	92.00	89.00	95.00	**97.00**	81.00	**99.00**
		Iris	**96.67**	**95.00**	91.67	**96.67**	**96.67**	91.67	**95.00**
		Vowel	33.33	29.44	33.55	**35.71**	35.04	35.50	**42.64**
		Glass	60.92	65.52	**77.01**	60.92	58.62	**67.82**	65.52

در ابتدای جدول (۳) نتایج کلاس‌بندهای مبتنی بر فاصله بیان شده است. در اینجا سه روش کلاس‌بندی K* (KStar)، نزدیکترین همسایه (IB1) و k نزدیکترین همسایه با شعاع همسایگی، k، سه (IBK,3) مد نظر قرار گرفته است. با توجه به نتایج، روشهای تبدیل ویژگی غیرخطی عملکرد بهتری نسبت به روشهای تبدیل ویژگی خطی دارند و این تفاوت در دادگان با تعداد ویژگی و تعداد کلاس بیشتر مانند Glass و Vowel مشهودتر است. روش پیشنهادی، KMCE، نسبت به روش MCE دارای نرخ کلاس بندی بالاتری است، این امر به دلیل ویژگی مبتنی بر هسته بودن روش پیشنهادی است. همان‌گونه که قبلاً گفته شد روش MCE یک روش تبدیل ویژگی خطی است و در دادگان با تعداد ویژگی و تعداد کلاس بیشتر عملکرد ضعیفی دارد. با بکارگیری MCE در قالب روش‌های مبتنی بر هسته این نقیصه برطرف گردید. روش پیشنهادی بر اساس کاهش خطای کلاس بندی عمل می‌کند، برای بدست آوردن خطا از کلاس‌بند فاصله استفاده شده است. بنابراین انتظار می‌رود که در کلاس بندهای مبتنی بر فاصله عملکرد بهتری داشته باشد و همانطور که در جدول (۳) مشاهده می شود در تمامی دادگان و روشهای کلاس‌بندی مبتنی بر فاصله، روش KMCE نرخ کلاس‌بندی بالاتری دارد.

باید توجه داشت که روش KMCE بر اساس کاهش خطای کلاس‌بند مبتنی بر فاصله عمل می‌کند و کارائی برتر این روش نسبت به دیگر روشهای تبدیل ویژگی در کلاس-بندهای مبتنی بر فاصله مشاهده شد. در ادامه عملکرد این روش در کلاس‌بندهایی که از معیار فاصله استفاده نمی‌کنند نیز ارزیابی خواهد شد. برای این منظور ابتدا روش پیشنهادی در کلاس‌بندهای آماری و سپس در کلاس‌بندهای مبتنی بر درخت تصمیم مورد بررسی قرار می‌گیرد.

بخشی از جدول (۳) نتایج کلاس‌بندی را روی کلاس‌بند آماری BayesNet نشان می‌دهد. چنانچه مشاهده می‌شود برای کلاس‌بند BayesNet روش KMCE نسبت به روشهای تبدیل ویژگی خطی نتایج بالاتری در کلاس بندی دارد. درضمن روش KMCE نسبت به روش KPCA عملکرد بالاتری داشته است، به عنوان مثال این بهبود در دادگان Glass از ۵۱,۷۲ درصد برای روش تبدیل ویژگی PCA به ۶۶,۶۷ درصد برای روش تبدیل ویژگی KMCE رسیده است. در این کلاس‌بند روش KLDA و روش پیشنهادی نتایج نزدیک به هم دارند.

نتایج کلاس‌بندهای مبتنی بر درخت تصمیم NBTree در انتهای جدول (۳) نشان داده شده است. با توجه به نتایج مندرج در جدول، روش KMCE در اغلب موارد بیشترین نتیجه بازشناسی را دارد و یا رتبه دوم را در بین روشهای تبدیل ویژگی به خود اختصاص داده است.

۴,۳. نتایج روی دادگان Aurora2

برای ارزیابی روش‌های پیشنهادی آزمایش‌هایی برای تشخیص کلمات مجزا بر روی دادگان Aurora2 انجام شده است [38]. بعد بردار ویژگی ۳۹ است که شامل ۱۲ ضریب MFCC به اضافه انرژی و مشتق اول و دوم این ضرایب می‌باشد. پس از استخراج ضرایب MFCC روی آنها هنجار سازی بهره ضرایب کپسترال (CGN[۳۷]) انجام می‌شود [39]. همچنین هر مدل مخفی مارکف که در اینجا برای بازشناسی گفتار بکار گرفته شده است دارای ۶ حالت و ۸ مخلوط گاوسی در هر حالت می‌باشد.

در جدول (۴) میانگین مقادیر دقت بازشناسی گفتار در هر مجموعه آزمایش دادگان Aurora2، بصورت جداگانه گزارش شده است. نتایج MFCC بدون اعمال CGN آورده شده است تا به عنوان مبنای کار باشد و نتایج قابل مقایسه با مقالات دیگر باشد. در تمامی موارد PCA، LDA، MCE، KPCA، KLDA و KMCE پس از استخراج ویژگی‌های MFCC و اعمال CGN این تبدیلات روی ویژگی‌ها اعمال شده است.

مشاهده می‌شود که روش پیشنهادی KMCE در هر سه مجموعه تست، نسبت به روش‌های تبدیل غیرخطی مبتنی بر هسته دیگر عملکرد بهتری داشته است. همانطور که

[37] Cepstral Gain Normalization

مشاهده می‌گردد پس از روش پیشنهادی روش KPCA عملکرد مطلوبی داشته و روش KLDA عملکرد ضعیفی از خود نشان داده است.

جدول ۴- میانگین مقادیر دقت بازشناسی کلمه Aurora2 روی مجموعه های A,B,C (انواع نویز) با نرخ سیگنال به نویز 0dB تا 20dB

	متوسط درصد دقت بازشناسی کلمه برای سه مجموعه آزمایش Aurora2		
	A	B	C
MFCC (baseline)	60.68	65.75	45.07
MFCC+CGN	79.19	80.91	70.68
PCA	80.74	81.67	74.47
KPCA	81.58	82.27	74.82
LDA	76.62	80.48	71.21
KLDA	78.95	81.19	74.76
MCE	79.03	80.57	71.78
KMCE	**81.96**	**83.27**	**76.71**

۵. نتیجه گیری

در این مقاله روش خطای کلاس‌بندی کمینه مبتنی بر هسته (KMCE) برای بازشناسی الگو پیشنهاد و فرموله گردیده است. این روش بر مبنای معیار کمینه کردن خطای کلاس‌بند در فضای ایجاد شده توسط هسته می‌باشد. معیار پیشنهادی با متعامدسازی ویژگی‌های در فضای هسته (مانند KPCA) و تفکیک‌پذیری بیشتر کلاس‌ها در فضای هسته (مانند KLDA) و دیگر تبدیلات خطی ویژگی‌ها مانند LDA، PCA و MCE مقایسه شده است. از سه گروه کلاس-بند مبتنی بر فاصله، آماری و مبتنی بر درخت تصمیم جهت ارزیابی تاثیر روش پیشنهادی نسبت به دیگر روشها بر نرخ بازشناسی استفاده شده است. نتایج نشان می‌دهند که روش KMCE برای تمامی دادگان و تمامی کلاس‌بندهای مبتنی بر فاصله عملکرد بهتری داشته است. این امر به معیار بدست آوردن تبدیل در روش پیشنهادی کاهش خطای کلاس‌بندی بر می‌گردد که روش برای کلاس‌بند فاصله فرموله شده است. بنابراین انتظار کارائی بهتر در کلاس‌بندهای مبتنی بر فاصله می‌رود. به علاوه عملکرد KMCE در کلاس‌بندهای آماری و مبتنی بر درخت تصمیم‌گیری نیز قابل قبول است. نتایج بازشناسی گفتار تحت نویز بر روی دادگان Aurora 2 نیز بیانگر عملکرد بهتر روش پیشنهادی یعنی KMCE است.

منابع

[1] I. T. Jolliffe, Principal component analysis, Springer-Verlag, New York, 1986.

[2] M. Loog, R. P. W. Duin, Linear dimensionality reduction via a heteroscedastic extension of LDA: the Chernoff criterion, IEEE Transaction Pattern Analysis and Machine Intelligence, Vol. 26, No. 6, pp. 732–739, 2004.

[3] Z. Jin, J. Y. Yang, Z. S. Hu, Z. Lou, Face recognition based on the uncorrelated discriminant transformation, Pattern Recognition, Vol. 34, No. 7, pp. 1405–1416, 2001.

[4] M. Loog, R. P. W. Duin, R. Haeb-Umbach, Multiclass linear dimension reduction by weighted pairwise Fisher criteria, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 7, 2001.

[5] H. Abdi, Metric multidimensional scaling, in Encyclopedia of Measurement and Statistics, pp. 598-605, 2007.

[6] I. Alphonso, Heteroscedastic discriminant analysis, Critical Review Paper, ECE 8990 - Special Topics in ECE-Pattern Recognition, Institute for Signal and Information Processing, Mississippi State University, 2001.

[7] A. K. Qin, P. N. Suganthan, M. Loog, Uncorrelated heteroscedastic LDA based on the weighted pairwise Chernoff criterion, Pattern Recognition, Vol. 38, pp. 613 – 616, 2005.

[8] M. Sakai, N. Kitaoka, S. Nakagawa, Power linear discrimant analysis, 9th International Symposium on Signal Processing and Its Applications (ISSPA), 2007.

[9] A. Hyvarinen, J. Karhunen, E. Oja, Independent component analysis, John Wiley & Sons, 2001.

[10] A. Kocsor, K. Kovacs, Kernel springy discriminant analysis and its application to a phonological awareness teaching system, in Proceedings of Text, Speech and Dialogue (TSD): 5th International conference, pp. 325-328, 2002.

[11] A. de la Torre, A. M. Peinado, A. J. Rubio, V. E. SGnchez, J. E. Diaz, An application of minimum

[24] J. Yang, A. Frangi, J. Yang, Z. Jin, KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition. IEEE TPAMI, Vol. 27, No. 2, pp. 230–244, 2005.

[25] T. Xiong, J. Ye, Kernel uncorrelated and orthogonal discriminant analysis: a unified approach, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), pp. 125–131, 2006.

[26] W. Cao, L. Li, X. Lv, Kernel function characteristic analysis based on support vector machine in face recognition, Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Vol. 5, pp. 2869-2873, 2007.

[27] C. Blake, E. Keogh, C. J. Merz, UCI repository of machine learning databases, 1998. Available: www.ics.uci.edu/ mlearn/MLRepository.html

[28] A. Lima, H. Zen, Y. Nankaku, C. Miyajima, K. Tokuda, T. Kitamura, On the use of kernel PCA for feature extraction in speech recognition, Proceeding of EuroSpeech, pp. 2625–2628, 2003.

[29] K. Fukunaga, Introduction to statistical pattern recognition, Academic Press, San Diego, California, USA, 1990.

[30] G. H. Golub, C. F. Van Loan, Matrix computations, The Johns Hopkins University Press, Baltimore, MD, USA, third edition, 1996.

[31] R. O. Duda, P. E. Hart, D. G. Stork, Pattern classification, 2nd edition, John Wiley & Sons, 2001.

[32] B. Zhang, S. Matsoukas, Minimum phoneme error based heteroscedastic linear discriminant analysis for speech recognition. In: Proceedings of ICASSP, Vol. 1, pp. 925-928, 2005.

[33] N. Kumar, A. G. Andreou, Heteroscedastic discriminant analysis and reduced rank hmms for improved speech recognition, Speech Communication, Vol. 25, No. 4, 1998.

[34] J. Hung, L. S. Lee, Data-driven temporal filters for robust features in speech recognition obtained via minimum classification error (MCE). In: Proceedings of ICASSP, pp. 373-376, 2002.

[35] X. Wang, K. K. Paliwal, Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. Pattern Recognition, Vol. 36, pp. 2429-2439, 2003.

[36] A. Kocsor, L. Toth, Kernel-based feature extraction with a speech technology application, classification error to feature space transformations for speech recognition, Speech Communication, Vol. 20, pp. 273-290, 1996.

[12] B. H. Juang, W. Chou, C. H. Lee, Minimum classification error rate methods for speech recognition, IEEE Transaction Speech Audio Processing, Vol. 5, No. 3, pp. 257–265, 1997.

[13] T. Rudolph, Minimum classification error optimization of word recognizers using evolution strategies, IEEE International Conference on Evolutionary Computation, Vol. 2, pp. 521-526, 1995.

[14] E. C. Malthouse, Limitations of nonlinear PCA as performed with generic neural networks, IEEE Transaction on Neural Networks, Vol. 9, No. 1, pp. 165-173, 1998.

[15] D. Tzovaras, M. G. Strintzis, Use of nonlinear principal component analysis and vector quantization for image coding, IEEE Transactions On Image Processing, Vol. 7, No. 8, pp 1218-1223, 1998.

[16] M. A. Kramer, Nonlinear principal component analysis using auto associative neural networks, AIChE J., Vol. 37, No. 2, pp. 233–243, 1991.

[17] V. N. Vapnik, Statistical learning theory, John Wiley & Sons Inc., 1998.

[18] N. Cristianini, J. Shawe-Taylor, An introduction to support vector machines and other kernel-based learning methods, Cambridge University Press, 2000.

[19] B. Scholkopf, A. Smola, K. R. Muller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput., Vol. 10, No. 5, pp. 1299–1319, 1998.

[20] C. Liu, Gabor-based kernel PCA with fractional power polynomial models for face recognition, IEEE Transaction Pattern Anal. Machine Intelligence, Vol. 26, No. 5, pp. 572–581, 2004.

[21] V. Roth and V. Steinhage, Nonlinear discriminant analysis using kernel functions, in Advances in Neural Information Processing Systems, pp. 568–574, 1999.

[22] S. Mika, G. Ratsch, J. Weston, B. Schokopf, K. R. Muller, Fisher discriminant analysis with kernels, In IEEE Neural Networks for Signal Processing Workshop, pp. 41– 48, 1999.

[23] T. Xiong, J. Ye, Q. Li, R. Janardan, V. Cherkassky, Efficient kernel discriminant analysis via QR decomposition, In NIPS, pp. 1529–1536, 2005.

IEEE Transactions on Signal Processing, Vol. 52, No. 8, pp. 2250-2263, 2004.

[37] X. Xie, K. M. Lam, Gabor-based kernel PCA with doubly nonlinear mapping for face recognition with a single face image, IEEE Transactions on Image Processing, Vol. 15, No. 9, pp. 2448-2492, 2006.

[38] H. G. Hirsch, D. Pearce, The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions, ISCA ITRW ASR, Paris, France, 2000.

[39] S. Yoshizawa, N. Hayasaka, W. Naoya, Y. Miyanaga, Cepstral gain normalization for noise robust speech recognition, Proc. ICASSP, pp. 209–212, 2004.