

بازشناسی واکه‌های انگلیسی با استفاده از روش خطای کلاس‌بند کمینه مبتنی بر هسته

بهزاد زمانی^۱، بابک ناصرشریف^۲ و احمد اکبری^۳

^۱آزمایشگاه پردازش صوت و گفتار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت، bzamani@iust.ac.ir

^۲گروه مهندسی کامپیوتر، دانشکده فنی، دانشگاه گیلان، Nasser_s@iust.ac.ir

^۳آزمایشگاه پردازش صوت و گفتار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت، akbari@iust.ac.ir

چکیده - خطای کلاس‌بند کمینه روشی متمایز ساز در حوزه مدل و ویژگی می‌باشد که نرخ بازشناسی را در سیستم‌های بازشناسی گفتار افزایش می‌دهد. روش خطای کلاس‌بند کمینه در حوزه ویژگی یک تبدیل خطی برای ویژگی‌ها بدست می‌آورد که سبب کاهش خطای کلاس‌بند در فضای تبدیل یافته می‌شود. این تبدیل در شرایطی کاراست که کلاس‌ها دارای خاصیت تفکیک‌پذیری خطی باشند. مقاله حاضر با بهره‌گیری از توابع هسته روشی ارائه می‌کند که روش خطای کلاس‌بند کمینه در جایی که کلاس‌ها بطور خطی تفکیک‌پذیر نیستند نیز به خوبی عمل کند. روش جدید را روش خطای کلاس‌بند کمینه مبتنی بر هسته می‌نامیم. عملکرد روش پیشنهادی برای بازشناسی واکه‌های TIMIT با استفاده از کلاس‌بند نزدیکترین همسایگی ارزیابی شده است. نتایج آزمایشات نشان می‌دهد که نرخ بازشناسی با استفاده از روش پیشنهادی نسبت به روش خطای کلاس‌بندی کمینه افزایش یافته است.

کلید واژه- خطای کلاس‌بندی کمینه، متمایزسازی، بازشناسی گفتار، توابع هسته.

محیط‌های دارای نویز جمع‌پذیر کاهش می‌یابد. بنابراین روش‌های مختلفی پیشنهاد شده‌اند که با اعمال تبدیل بر روی این ویژگی‌ها موجب مقاومت بیشتر آنها در برابر نویز و تمایز بیشتر واحدهای گفتاری می‌گردند [5, 6].

از جمله روش‌های تبدیل می‌توان به PCA که روشی بدون ناظر است اشاره کرد. هدف PCA، تصویر داده‌ها در فضای با ابعاد کوچکتر است بطوری که خطای مجذور مربعات به کمترین مقدار خود برسد [7]. یکی دیگر از روش‌ها تبدیل LDA است که داده‌ها را تا حد امکان به صورت خطی تفکیک‌پذیر می‌کند [15]. Heteroscedastic LDA روشی است که فرض یکسان بودن واریانس کلاس‌ها در LDA را حذف می‌کند [8]، LDA می‌تواند جایگزین DCT نیز شود [9].

مشکل تبدیلات فوق این است که معیار بهینه‌سازی آنها با معیار کمینه‌سازی خطای کلاس‌بند نهایی تفاوت دارد، که باعث کاهش کارایی سیستم کلاس‌بند می‌شود. روش MCE سعی می‌کند با استفاده از یک معیار همسان، پیوندی میان مراحل استخراج ویژگی و دسته‌بندی ایجاد کند. در [6] چهارچوبی برای بهینه‌سازی ماتریس‌های تبدیل (مانند PCA و LDA) با استفاده از MCE پیشنهاد شده است تا به این ترتیب ضعف این دو روش در عدم توجه به نوع دسته‌بند جبران گردد و کارایی این دو روش افزایش یابد.

برای بهبود عملکرد روش‌های متمایزساز از توابع هسته استفاده شده است. دیدگاه کلی این روش‌ها متمایزسازی این است که ابتدا با یک نگاشت غیرخطی ویژگی‌ها را به فضایی با ابعاد بزرگتر انتقال داده

۱- مقدمه

سیستم‌های بازشناسی گفتار عموماً شامل دو مرحله اصلی می‌باشند: استخراج ویژگی و دسته‌بندی. مرحله دسته‌بندی براساس اطلاعات آماری طراحی می‌شود. مدل مخفی مارکوف (HMM) از پرکاربردترین روش‌های کلاس‌بندی در بازشناسی گفتار است. پارامترهای HMM با استفاده از مجموعه داده‌ها و الگوریتم آموزش مانند الگوریتم بیشینه درست‌نمایی (Maximum Likelihood) مبتنی بر الگوریتم EM (Expectation Maximization) تخمین زده می‌شوند. [1]

الگوریتم فوق، درست‌نمایی یک کلاس را بدون توجه به تمایز کلاس‌ها بیشینه می‌کند. روش‌های یادگیری تمایزی (Discriminant Training) متعددی از قبیل روش بیشینه اطلاعات متقابل (Maximum Mutual Information) [4]، روش بیشینه فاصله مدل‌ها (Maximum Model Distance) و روش خطای کلاس‌بند کمینه (Minimum Classification Error (MCE)) [2,3] برای جایگزینی روش بیشینه درست‌نمایی مطرح گردیده‌اند.

در مرحله استخراج ویژگی، اطلاعات متمایز کننده مفیدی از سیگنال گفتار استخراج می‌شود. اکثر سیستم‌های بازشناسی از ویژگی‌های مل کپستروم استفاده می‌نمایند که نحوه عملکرد گوش و سیستم شنوایی انسان را شبیه‌سازی می‌کند. با این وجود، این ویژگی‌ها برای ایجاد تمایز میان واحدهای گفتاری بهینه نیستند و به علاوه کارایی آنها در

می‌باشد. پس از آنکه نرخ خطای تجربی، بوسیله یک بازشناسی یا دسته‌بند، کمینه شد، تخمینی بایاس شده از نرخ خطای واقعی بدست می‌آید. یک راه مؤثر جهت کاهش این نرخ و بهبود کارایی کلی سیستم، افزایش مرز میان کلاس‌ها در داده‌های آموزشی است. به این منظور از گرادبان کاهش [21] یا ترکیب امتیازهای حاشیه و نرخ خطای تجربی استفاده شده است.

روش MCE برای بدست آوردن تبدیل فضای ویژگی نیز استفاده می‌شود. در [22]، MCE برای یافتن ماتریس تبدیل ویژگی‌ها، که بتواند خطای کلاس‌بندی را کمینه کند، استفاده شده است. Wang و Paliwal این روش را برای بازشناسی واژه‌ها بهبود بخشیدند [23]. در همه این رویکردها از تابع هزینه‌ای بصورت رابطه (۱) استفاده شده است.

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

که در آن M تعداد کلاس‌ها، η مقداری مثبت (ارائه شده در [2])، و $g_i(O, F)$ لگاریتم احتمال تعلق O به کلاس λ_i که بصورت رابطه (۲) تعریف می‌شود:

$$g_i(O, F) = \log p(O | \lambda_i) \quad (2)$$

که F مجموعه پارامترهای مرتبط با ویژگی‌ها و کلاس‌بندها می‌باشد. در روش MCE، هدف کمینه نمودن تابع هزینه (۱) 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:

$$I_k(O, F) = \frac{1}{1 + \exp(-\alpha d_k(O, F))} \quad (3)$$

که $d_k(O, F)$ تابع هزینه (۱) می‌باشد و α پارامتر تنظیم بزرگتر از یک است. بنابراین اگر $d_k(O, F)$ خیلی کوچکتر از صفر باشد، در واقع یک دسته‌بندی درست رخ داده و $I_k(O, F)$ به صفر نزدیک می‌شود. از طرفی مقدار مثبت برای $d_k(O, F)$ نشان دهنده جریمه برای دسته‌بندی نادرست است و در اینصورت $I_k(O, F)$ به ۱ متمایل می‌شود. کل خطای کلاس‌بند برای مشاهده O از رابطه (۴) بدست می‌آید:

$$L = \sum_{k=1}^M I_k(O, F) \quad (4)$$

با استفاده از ماتریس تبدیل، بردارهای ویژگی از فضای ویژگی اصلی 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 تکراری با استفاده از گرادبان کاهش برای تابع L ، با رابطه (۵) قابل محاسبه خواهد بود.

سپس ادعا می‌شود که در این فضای ویژگی نمونه‌ها با یک کلاس‌بند خطی براحتی قابل تفکیک هستند. در روشهای مبتنی بر هسته نیازی به داشتن نگاشت غیرخطی نیست، چراکه ضرب نقطه‌ای زوج دادگان در فضای نگاشت یافته با تابع هسته جایگزین می‌شود. [13, 14]

روشهای مبتنی بر هسته برای معیار Fisher مورد توجه محققان قرار گرفته است [10, 11]. این روشها آنالیز تفکیک‌پذیر Fisher مبتنی بر هسته (kernel Fisher discriminant analysis (KFD)) یا آنالیز تفکیک‌پذیر مبتنی بر هسته (kernel discriminant analysis (KDA)) نامیده می‌شوند [11]. در روش KDA ابتدا داده‌ها را به فضای ویژگی با ابعاد زیاد نگاشت می‌دهد، سپس الگوریتم LDA را بر روی دادگان نگاشت یافته اعمال می‌کنند [12]. از دیگر روشهای مبتنی بر هسته می‌توان روش آنالیز مؤلفه‌های اصلی مبتنی بر هسته (kernel principal component analysis (KPCA) [18])، روش آنالیز مؤلفه‌های مستقل مبتنی بر هسته (kernel independent component analysis (KICA) [17]) و روش آنالیز تفکیک‌پذیری ارتجاعی (kernel springy discriminant analysis (KSDA) [16]) را نام برد.

در این مقاله روش خطای کلاس‌بند کمینه مبتنی بر هسته ارائه شده است. در روش پیشنهادی ابتدا نمونه‌ها را از فضای ویژگی فعلی به فضایی با ابعاد به اندازه کافی بزرگ نگاشت داده (نگاشت غیرخطی)، سپس در این فضا تبدیل ویژگی خطای کمینه بدست می‌آید. تمام روابط بر اساس نمونه‌های نگاشت یافته در فضای جدید بازنویسی شده و ضربهای نقطه‌ای زوج نمونه‌های نگاشت یافته ایجاد شده در روابط با توابع هسته جایگزین می‌شوند.

ساختار ادامه مقاله به این صورت است. بخش دوم، به تشریح الگوریتم MCE اختصاص دارد. در بخش سوم روش پیشنهادی یعنی الگوریتم MCE مبتنی بر هسته (Kernel Minimum Classification Error (KMCE)) توضیح داده شده است. بخش چهارم حاوی نتایج و ارزیابی روش‌های آزمایش شده است، و در ادامه، بخش پنجم نتیجه‌گیری نهایی را در بر دارد.

۲- خطای کلاس‌بند کمینه

روش کمینه نمودن خطای کلاس‌بندی (MCE) متمایزسازی است که هم در گروه تبدیل ویژگی و هم در گروه آموزش دسته‌بند بکار گرفته می‌شود [19, 20]. هنگامی که روش MCE برای آموزش HMM (به عنوان یک دسته‌بند) استفاده می‌شود، MCE پارامترهای مدل را بگونه‌ای تغییر می‌دهد که خطای کل کلاس‌بندی کاهش یابد [2]. 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 جستجوی گرادبان، مانند Gradient Probabilistic Descent، بدست می‌آید [21].

ایده اصلی الگوریتم MCE، بهینه‌سازی نرخ خطای تجربی با استفاده از یک مجموعه داده آموزشی، جهت بهبود نرخ خطای بازشناسی

بدست می‌آید. با بکارگیری مقادیر $\frac{\partial I_k(O, F)}{\partial W}$ در رابطه (۶) ماتریس بهینه بدست می‌آید.

$$w_{n,iter} = w_{n,iter-1} - \beta \frac{\partial L}{\partial w_n} \quad (5)$$

رابطه فوق را می‌توان بصورت ماتریسی بیان کرد:

$$W_{iter} = W_{iter-1} - \beta \frac{\partial L}{\partial W} \quad (6)$$

۳- خطای کلاس‌بند کمینه مبتنی بر هسته

این بخش دارای دو زیر بخش می‌باشد. ابتدا ایده اصلی تابع هسته بیان می‌شود در ادامه روش پیشنهادی عنوان می‌شود.

۳-۱- ایده تابع هسته

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

$$k: X \times X \rightarrow \mathfrak{R} \quad (11)$$

ویژگیهای تابع هسته بفرم زیر است:
متقارن است یعنی

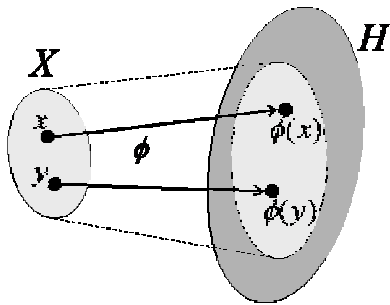
$$k(x, y) = k(y, x) \quad \forall x, y \in X$$

▪ K مثبت تعریف می‌شود.

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

می‌توان نشان داد که برای یک تابع هسته، k ، فضای هیلبرت H و یک تبدیل $\phi: X \rightarrow H$ وجود دارد (شکل (۱)) که رابطه زیر برقرار باشد:

$$k(x, y) = \langle \phi(x), \phi(y) \rangle = \phi(x) \cdot \phi(y) \quad (12)$$



$$k(x, y) = \phi(x) \cdot \phi(y)$$

شکل (۱) - ایده هسته: ضرب نقطه‌ای دو نمونه نگاشت یافته معادل تابع هسته دو نمونه می‌باشد

۳-۲- روش پیشنهادی

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

$$X^\Phi = \Phi(X) \quad (13)$$

که W ماتریس تبدیل؛ r شاخص r امین درایه ماتریس تبدیل W ؛ $iter$ اندیس تکرار الگوریتم، و β پارامتر یادگیری می‌باشد. رابطه (۷) نشان‌دهنده یک رویه تکراری است. این رویه زمانی متوقف می‌شود که مقدار تابع هزینه از یک حد آستانه کمتر شود.

فرض کنید ماتریس تبدیل W بردار ویژگی‌های n بُعدی x را به بردار y با بُعد d ($d \leq n$) تبدیل می‌کند. هدف، بدست آوردن ماتریس تبدیل W است که تابع هزینه را کمینه نماید. برای کمینه کردن خطا از رابطه (۳) نسبت به W مشتق گرفته می‌شود، می‌توان نوشت:

$$\frac{\partial I_k(O, F)}{\partial W} = \frac{\partial I_k(O, F)}{\partial d_k(O, F)} \frac{\partial d_k(O, F)}{\partial W} \quad (7)$$

تابع هزینه $I_k(O, F)$ همان رابطه (۳) می‌باشد. بردار O رشته مشاهده یا رشته بردار ویژگی‌های تبدیل یافته است و بصورت $O = \{o_1, o_2, \dots, o_T\}$ می‌باشد. همچنین بردار ویژگی‌های اصلی بصورت $X = \{x_1, x_2, \dots, x_T\}$ نشان داده شده است. با توجه به رابطه (۳) خواهیم داشت [8]:

$$\frac{\partial I_k(O, F)}{\partial d_k(O, F)} = \alpha I_k(O, F) (1 - I_k(O, F)) \quad (8)$$

برای محاسبه بخش دوم رابطه (۷)، $\frac{\partial d_k(O, F)}{\partial W}$ ، از رابطه (۱) بر حسب W مشتق گیری می‌شود:

$$\frac{\partial d_{k,ADD}(O, F)}{\partial W} = -\frac{\partial g_k(O, F)}{\partial W} + \sum_{i=1, i \neq k}^M \left(\frac{\exp(g_i(O, F)\eta)}{\sum_{j=1, j \neq k}^M \exp(g_j(O, F)\eta)} \times \frac{\partial g_i(O, F)}{\partial W} \right) \quad (9)$$

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.

$$\frac{\partial g_i(O, F)}{\partial w_d} = 2w_d (O(d) - y_i(d))^2 \quad (10)$$

که $y_i(d)$ درایه d ام مرکز کلاس i ام می‌باشد. با جایگذاری (۱۰) در معادله (۹) مقدار $\frac{\partial d_k(O, F)}{\partial W}$ بدست می‌آید. با جانشینی رابطه

محاسبه شده $\frac{\partial d_k(O, F)}{\partial W}$ و رابطه (۸) در (۷) مقدار $\frac{\partial I_k(O, F)}{\partial W}$

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

$$I_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha}) = \frac{1}{1 + \exp(-\gamma d_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha}))} \quad (20)$$

که $d_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})$ تابع هزینه (۱۶) 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 $I_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})$ 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 $I_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})$ به یک متمایل می‌شود. کل خطای کلاس‌بندی برای مشاهده از رابطه (۲۱) بدست می‌آید:

$$L = \sum_{k=1}^I \sum_{n_k=1}^{N_k} I_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha}) \quad (21)$$

که N_k تعداد مشاهدات کلاس k ام و $N = \sum_{k=1}^I N_k$ کل مشاهدات است. با استفاده از ماتریس تبدیل بردارهای ویژگی را از فضای ویژگی اصلی 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 تبدیل بروش تکراری با استفاده از گرادینان کاهش با رابطه زیر محاسبه می‌شود:

$$\alpha_{pq,iter} = \alpha_{pq,iter-1} - \beta \frac{\partial L}{\partial \alpha_{pq}}; p=1,2,\dots,N, q=1,2,\dots,M \quad (22)$$

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

$$\frac{\partial I_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})}{\partial \alpha_{pq}} = \frac{\partial I_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})}{\partial d_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})} \frac{\partial d_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})}{\partial \alpha_{pq}} \quad (23)$$

با توجه به رابطه (۲۰) داریم:

$$\frac{\partial I_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})}{\partial d_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})} = \gamma I_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha}) (1 - I_k(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha})) \quad (24)$$

برای محاسبه ترم دوم در رابطه (۲۳)، از رابطه

(۱۶) استفاده می‌شود. بعد از مشتق‌گیری از رابطه فوق می‌توان نوشت:

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

$$X^{\Phi,W} = W\Phi(X) \quad (14)$$

از طرفی W بفرم $W = \sum_{i=1}^N \alpha_i X_i^{\Phi}$ تعریف می‌شود که N تعداد کل نمونه‌های آموزشی و X_i^{Φ} نمونه i ام داده آموزشی نگاشت یافته می‌باشد. بنابراین رابطه بالا به صورت زیر در می‌آید:

$$\begin{aligned} X^{\Phi,W} &= WX^{\Phi} = \left(\sum_{n=1}^N \alpha_n (X_n^{\Phi})^T \right) X^{\Phi} = \\ &= \sum_{n=1}^N \alpha_n ((X_n^{\Phi})^T X^{\Phi}) = \sum_{n=1}^N \alpha_n (X_n^{\Phi} \cdot X^{\Phi}) = \\ &= \sum_{n=1}^N \alpha_n (\Phi(X_n) \cdot \Phi(X)) = \sum_{n=1}^N \alpha_n k(X_n, X) = X^{\Phi,\alpha} \end{aligned} \quad (15)$$

روش خطای کلاس‌بندی کمینه سعی در بدست آوردن تبدیلی دارد که خطای کلاس‌بندی را کاهش دهد. دو تابع هزینه برای این روش در مقالات ارائه شده است که آنها در فضای نگاشت یافته بازنویسی شده‌اند.

$$\begin{aligned} d_{k,Add}(X_{n_k}^{\Phi,\alpha}, Y^{\Phi,\alpha}) &= -g_k(X_{n_k}^{\Phi,\alpha}, y_k^{\Phi,\alpha}) + \\ &= \frac{1}{\eta} \log \left[\frac{1}{I-1} \sum_{i=1, i \neq k}^I \exp(g_i(X_{n_k}^{\Phi,\alpha}, y_i^{\Phi,\alpha})) \eta \right] \end{aligned} \quad (16)$$

که I تعداد کلاسها، η عدد مثبتی است. همچنین $X_{n_k}^{\Phi,\alpha}$ مشاهداتی نگاشت یافته و تبدیل یافته n_k از کلاس k ام می‌باشد. $g_i(X_{n_k}^{\Phi,\alpha}, y_i^{\Phi,\alpha})$ برابر با مربع فاصله $X_{n_k}^{\Phi,\alpha}$ با نماینده کلاس i ام در فضای نگاشت شده و تبدیل یافته یعنی $y_i^{\Phi,\alpha}$ می‌باشد که به صورت رابطه (۱۷) بیان می‌گردد.

$$y_i^{\Phi,\alpha} = \frac{1}{N_i} \sum_{n_i=1}^{N_i} X_{n_i}^{\Phi,\alpha} = \frac{1}{N_i} \sum_{n_i=1}^{N_i} \sum_{n=1}^N \alpha_n k(X_n, X_{n_i}) \quad (17)$$

با توجه به استفاده نرم اقلیدوسی و روابط (۱۵) و (۱۷) داریم:

$$\begin{aligned} g_i(X_{n_k}^{\Phi,\alpha}, y_i^{\Phi,\alpha}) &= \|X_{n_k}^{\Phi,\alpha} - y_i^{\Phi,\alpha}\|^2 = \\ &= \left(\sum_{n=1}^N \alpha_n k(X_n, X_{n_k}) - \frac{1}{N_i} \sum_{n_i=1}^{N_i} \sum_{n=1}^N \alpha_n k(X_n, X_{n_i}) \right) \\ &= \left(\sum_{n=1}^N \alpha_n k(X_n, X_{n_k}) - \frac{1}{N_i} \sum_{n_i=1}^{N_i} \sum_{n=1}^N \alpha_n k(X_n, X_{n_i}) \right)^T \end{aligned} \quad (18)$$

با بسط ضرب ماتریسی فوق بفرم عددی خواهیم داشت:

$$\begin{aligned} g_i(X_{n_k}^{\Phi,\alpha}, y_i^{\Phi,\alpha}) &= \|X_{n_k}^{\Phi,\alpha} - y_i^{\Phi,\alpha}\|^2 = \\ &= \sum_{m=1}^M \left(\sum_{n=1}^N \alpha_{nm} k(X_n, X_{n_k}) - \frac{1}{N_i} \sum_{n_i=1}^{N_i} \sum_{n=1}^N \alpha_{nm} k(X_n, X_{n_i}) \right)^2 \end{aligned} \quad (19)$$

جدول (۱) نحوه نامگذاری روشها را نشان می دهد، که از این نامها در گزارش نتایج استفاده شده است.

کلاس بند KNN با دو حالت k برابر ۳ و ۴ می باشد، بکار برده شده است. k شعاع همسایگی است که نمونه جدید فاصله اش با آنها بررسی می شود. همچنین برای بخش آموزش و تست دو حالت ۶۶ درصد و ۱۰ folds استفاده شد. در ضمن در تمام حالات برای هر کلاس یک نمونه به عنوان معرف مرکز کلاس در نظر گرفته شده است.

جدول (۲) نتایج بازشناسی واکه های TIMIT را برای کلاس بند KNN با آموزش و تست ۶۶ درصدی نشان می دهد. نتایج بازشناسی برای KNN با شعاع همسایگی ۳ و ۴ گزارش شده است. نتایج جدول (۲) برای حالت بدون پیش پردازش و ویژگی، بکارگیری تبدیل ویژگی MCE و MCE مبتنی بر هسته آورده شده است.

جدول (۲) - نتایج بازشناسی واکه های TIMIT برای کلاس بند KNN با آموزش و تست ۶۶ درصدی

	KNN	MCE-KNN	KMCE-KNN
K=3	۸۷.۴۶	۸۸.۲۱	۸۸.۷۵
K=4	۸۳.۷۵	۸۴.۵۷	۸۵.۱۸

مطابق نتایج بازشناسی نشان داده شده در جدول (۲) روش پیشنهادی، KMCE-KNN، موجب افزایش بیشتر نرخ بازشناسی کلاس بند نسبت به MCE شده است. همچنین نتایج برای شعاع همسایگی ۳ بیشتر از شعاع همسایگی ۴ می باشد. چنانچه مشاهده می شود، روش MCE باعث بهبود ۰.۷۵ درصدی برای شعاع همسایگی ۳ و بهبود ۰.۸۲ درصدی برای شعاع همسایگی ۴ شده است. در ضمن روش پیشنهادی نیز باعث افزایش ۱.۳۰ درصدی برای k، ۳ و ۱.۴۳ درصدی برای k، ۴ شده است.

جدول (۳) نتایج بازشناسی واکه های TIMIT را برای کلاس بند KNN با آموزش و تست ۱۰ folds نشان می دهد. نتایج بازشناسی برای KNN با شعاع همسایگی ۳ و ۴ گزارش شده است. نتایج جدول (۳) برای حالت بدون پیش پردازش و ویژگی، بکارگیری تبدیل ویژگی MCE و MCE مبتنی بر هسته آورده شده است.

جدول (۳) - نتایج بازشناسی واکه های TIMIT برای کلاس بند KNN با آموزش و تست ۱۰ folds

	KNN	MCE-KNN	KMCE-KNN
K=3	۹۳.۹۴	۹۴.۶۰	۹۵.۲۷
K=4	۹۰.۸۸	۹۱.۶۵	۹۲.۲۱

مطابق نتایج بازشناسی نشان داده شده در جدول (۳) روش پیشنهادی، KMCE-KNN، موجب افزایش بیشتر نرخ بازشناسی کلاس بند نسبت

$$\frac{\partial d_{k, \text{Add}}(X_{n_k}^{\Phi, \alpha}, Y^{\Phi, \alpha})}{\partial \alpha_{pq}} = -\frac{\partial g_k(X_{n_k}^{\Phi, \alpha}, Y_k^{\Phi, \alpha})}{\partial \alpha_{pq}} + \sum_{i=1, i \neq k}^I \left\{ \frac{\exp(g_i(X_{n_k}^{\Phi, \alpha}, Y_i^{\Phi, \alpha})\eta)}{\sum_{j=1, j \neq k}^I \exp(g_j(X_{n_k}^{\Phi, \alpha}, Y_j^{\Phi, \alpha})\eta)} \times \frac{\partial g_i(X_{n_k}^{\Phi, \alpha}, Y_i^{\Phi, \alpha})}{\partial \alpha_{pq}} \right\} \quad (25)$$

مقدار $\frac{\partial g_i(X_{n_k}^{\Phi, \alpha}, Y_i^{\Phi, \alpha})}{\partial \alpha_{pq}}$ از رابطه (۲۶) محاسبه می گردد.

$$\frac{\partial g_i(X_{n_k}^{\Phi, \alpha}, Y_i^{\Phi, \alpha})}{\partial \alpha_{pq}} = 2 \left(k(X_p, X_{n_k}) - \frac{1}{N_i} \sum_{n_i=1}^{N_i} k(X_p, X_{n_i}) \right) \left(\sum_{n=1}^N \alpha_{nq} k(X_n, X_{n_k}) - \frac{1}{N_i} \sum_{n_i=1}^{N_i} \sum_{n=1}^N \alpha_{nq} k(X_n, X_{n_i}) \right) \quad (26)$$

با جایگزینی (۲۶) در (۲۵) مقدار $\frac{\partial d_k(X_{n_k}^{\Phi, \alpha}, Y^{\Phi, \alpha})}{\partial \alpha_{pq}}$ بدست می آید.

با جانشینی رابطه محاسبه شده $\frac{\partial d_k(X_{n_k}^{\Phi, \alpha}, Y^{\Phi, \alpha})}{\partial \alpha_{pq}}$ و رابطه (۲۴) در

$$(23) \text{ مقدار } \frac{\partial l_k(X_{n_k}^{\Phi, \alpha}, Y^{\Phi, \alpha})}{\partial \alpha_{pq}} \text{ بدست می آید. با بکارگیری مقادیر در رابطه (22) ماتریس بهینه بدست می آید.}$$

۴- نتایج آزمایشات

برای ارزیابی عملکرد روش پیشنهادی، تاثیر روش فوق و روش MCE معمولی بر روی دقت کلاس بند بررسی شد. برای ارزیابی از واکه های دادگان TIMIT استفاده شده است که به همین منظور ۱۵۰ نمونه برای هر یک از ۱۳ واکه در نظر گرفته شد. هر نمونه را به دنباله ای از فریمهای ۲۵ میلی ثانیه به شیف ۱۰ میلی ثانیه تقسیم نموده و ۳۶ ویژگی برای هر فریم استخراج گردید. ویژگیها شامل ۱۲ ضریب mfcc، مشتق اول و دوم آنها می باشند. برای کلاس بندی از کلاس بند نزدیکترین همسایگی (K Nearest Neighborhood (KNN)) پیاده سازی شده در Weka بهره گرفته شده است. [24]

جدول (۱) - نحوه نامگذاری روشها

اختصار	توضیح
KNN	کلاس بند KNN بدون هیچگونه تبدیل ویژگی
MCE-KNN	کلاس بند KNN با تبدیل ویژگی MCE
KMCE-KNN	روش پیشنهادی - کلاس بند KNN با تبدیل ویژگی MCE مبتنی بر هسته

- Communications, Control and Signal Processing, pp. 819-824, 2008.
- [10] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, no. 10, pp. 2385-2404, 2000.
- [11] T. Gestel, J. Suyken, G. Lanckriet, A. Lambrechts, B. De Moor, and J. Vanderwalle, "Bayesian framework for least squares support vector machine classifiers, Gaussian process and kernel Fisher discriminant analysis," *Neural Comput.*, vol. 15, no. 5, pp. 1115-1148, May 2002.
- [12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York: Springer-Verlag, 2001.
- [13] J. Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [14] B. Scholkopf and A. Smola, *Learning With Kernels*. Cambridge, MA: MIT Press, 2002.
- [15] X. B. Li, J. Y. Li and R. H. Wang, Dimensionality reduction using MCE-optimized LDA transformation. In: *Proceedings of ICASSP*, pp. 137-141, 2004.
- [16] A. Kocsor and K. Kovács., Kernel springy discriminant analysis and its application to a phonological awareness teaching system, In *Proceeding Text Speech Dialogue*, vol. 2448, pp. 325-328, 2002.
- [17] F. R. Bach and M. I. Jordan, Kernel independent component analysis," *Journal of Machine Learning Res.*, vol. 3, pp. 1-48, 2002.
- [18] B. Schölkopf, A. J. Smola, and K.-R. Muller, *Kernel Principal Component Analysis in Advances in Kernel Methods-Support Vector Learning*, Cambridge, MA: MIT Press, pp. 327-352, 1999.
- [19] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern classification*, 2nd edition, John Wiley & Sons, 2001.
- [20] B. Zhang and S. Matsoukas, Minimum phoneme error based heteroscedastic linear discriminant analysis for speech recognition. In: *Proceedings of ICASSP*, vol. 1. Philadelphia, PA, pp. 925-928, 2005.
- [21] M. Loog and R. P. W. Duin, Linear Dimensionality Reduction via a Heteroscedastic Extension of LDA: The Chernoff Criterion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 732-739, 2004.
- [22] J. Hung and 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.
- [23] X. Wang and K. K. Paliwal, Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition*, vol. 36, pp. 2429-2439, 2003.
- [24] H. I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Second Edition, 2005.

به MCE شده است. همچنین نتایج برای شعاع همسایگی ۳ بیشتر از شعاع همسایگی ۴ می باشد چنانچه مشاهده شد، روش MCE باعث بهبود ۰.۶۶ درصدی برای شعاع همسایگی ۳ و بهبود ۰.۷۷ درصدی برای شعاع همسایگی ۴ شده است. در ضمن روش پیشنهادی باعث افزایش ۱.۳۳ درصدی برای شعاع همسایگی ۳ و ۴ شده است.

۵- نتیجه گیری

چنانچه گفته شد روشهای مبتنی بر هسته باعث عملکرد بهتر کلاس-بندهای خطی روی دادگان غیرخطی می شوند. از سوی دیگر روش خطای کلاس بند کمینه در حوزه ویژگی تبدیلی خطی بدست می آورد که برای کلاس بندها در فضای خطی موثر هستند. از اینرو در این مقاله روش خطای کلاس بند کمینه مبتنی بر هسته پیشنهاد شده است که یک تبدیل غیرخطی برای متمایزسازی کلاسهای که بطور خطی جدائی پذیر نیستند، بدست می آورد. عملکرد روش پیشنهادی بر روی واژه های TIMIT با استفاده از کلاس-بند نزدیکترین همسایگی ارزیابی شد. دادگان دارای ۱۳ کلاس و ۱۵۰ نمونه برای هر کلاس می باشد. روش پیشنهادی بطور میانگین باعث افزایش ۱.۳۱ درصدی برای شعاع همسایگی ۳ و ۱.۳۸ درصدی برای شعاع همسایگی ۴ نسبت به کلاس بند اصلی شده است و عملکرد بهتری نسبت به MCE دارد.

مراجع

- [1] L. Rabiner and B. H. Juang, *Fundamentals of speech recognition*. Prentice Hall. Englewood Cliffs, New Jersey. 1993.
- [2] B. H. Juang, W. Chou and C. H. Lee, Minimum classification error rate methods for speech recognition. *IEEE Transaction on Speech and Audio Processing*, Vol. 5, No. 3, pp. 257-265, 1997.
- [3] X. He and L. Deng, A new look at discriminative training of hidden markov models. *Pattern Recognition Letters*. Vo. 28, pp.1285-1294, 2007.
- [4] S. Kapadia, V. Valtchev and S. J. Young, MMI training for continuous phoneme recognition on the TIMIT database. *IEEE International conference on Acoustics, Speech and Signal Processing*, Vol. 2. pp. 491-494, 1993.
- [5] B. Nasersharif and A. Akbari, SNR-dependent compression of enhanced Mel sub-band energies for compensation of noise effects on MFCC features. *Pattern recognition letters*, Vol. 28, Iss. 11, pp. 1320-1326, 2007.
- [6] آذرخش جلالوند، بهزاد زمانی، بابک ناصرشریف، احمد اکبری، "بهبود روشهای متمایزسازی ویژگی براساس خطای دسته بندی کمینه برای بازشناسی گفتار"، چهاردهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، اسفند ۱۳۸۷.
- [7] P. Geladi, H. Isaksson, L. Lindqvist, S. Wold and K. Esbensen, "Principal components analysis of multivariate images," *Chemometrics and Intelligent Laboratory Systems*, Vo. 5, No. 3, pp. 209-220, 1989.
- [8] B. Zhang and S. Matsoukas, "Minimum phoneme error based heteroscedastic linear discriminant analysis for speech recognition," In: *Proceedings of ICASSP*, vol. 1. Philadelphia, PA, pp. 925-928, 2005.
- [9] H. Abbasian, B. Nasersharif, A. Akbari, M. Rahmani and M. S. Moin, "Optimized linear discriminant analysis for extracting robust speech features," *3rd International Symposium on*