بهبود روشهاي متمايزساز ويژگي براساس خطاي دسته‌بندي كمينه براي بازشناسي گفتار

آذرخش جلالوند
دانشكده مهندسي كامپيوتر
دانشگاه علم و صنعت ايران
az_jalalvand@comp.iust.ac.ir

بهزاد زماني دهكردي
دانشكده مهندسي كامپيوتر
دانشگاه علم و صنعت ايران
bzamani@iust.ac.ir

احمد اكبري
دانشكده مهندسي كامپيوتر
دانشگاه علم و صنعت ايران
akbari@iust.ac.ir

بابک ناصرشريف
گروه مهندسي كامپيوتر
دانشكده فني، دانشگاه گيلان
nasser_s@iust.ac.ir

چكيده: يكي از جنبه‌هاي مهم دسته‌بندي الگوها و بازشناسي گفتار، طراحي بخش استخراج ويژگي‌ها مي‌باشد. براي استخراج مؤثرتر ويژگي‌ها، گاه تبديلاتي بر آنها اعمال مي‌شود تا با نگاشت ويژگي‌ها به فضايي جديد، تفكيک ميان كلاس‌هاي ويژگي، استقلال آماري ويژگي‌ها يا مقاومت آنها نسبت به نويز بيشتر گردد. مثال‌هايي از اين نوع تبديل، آناليز مؤلفه‌هاي اصلي (Principal Components Analysis) و آناليز تفكيک‌پذير خطي (Linear Discriminant Analysis) هستند كه در تعيين تبديل تنها به داده‌ها توجه مي‌كنند و به نوع دسته‌بند توجهي ندارند كه اين سبب ضعف آنها در بكارگيري به عنوان پيش پردازش ويژگي‌ها مي‌گردد. در اين مقاله، چهارچوبي براي رفع اين ايراد روش‌هاي مذكور، با استفاده از معيار كمينه خطاي دسته‌بندي (Minimum Classification Error) براي مدل مخفي ماركوف ارائه شده است. نتايج آزمايش‌ها روي مجموعه داده Aurora2 نشان مي‌دهد كه چهارچوب پيشنهادي كارآيي تبديل‌هاي PCA و LDA را براي ويژگي‌هاي گفتاري MFCC بهبود مي‌بخشد.

واژه هاي كليدي: كمينه خطاي دسته‌بندي، تبديل ويژگي‌ها، مدل مخفي ماركوف، آناليز تفكيک‌پذير خطي، آناليز مؤلفه‌هاي اصلي

۱- مقدمه

سيستم‌هاي بازشناسي گفتار عموماً شامل دو مرحله اصلي مي‌باشند: استخراج ويژگي و دسته‌بندي. مرحله دسته‌بندي براساس اطلاعات آماري طراحي مي‌شود. يک روش دسته‌بندي معروف براي بازشناسي گفتار، مدل مخفي ماركوف (HMM) مي‌باشد. پارامترهاي HMM با استفاده از مجموعه داده‌ها و الگوريتمِ آموزش، بصورت خودكار تنظيم مي‌گردد. بعنوان يک الگوريتم آموزش معروف، مي‌توان از الگوريتم بيشينه درست‌نمايي (Maximum Likelihood) مبتني بر الگوريتم EM (Expectation Maximization) نام برد[1].

الگوريتم مذكور، درست‌نمائي يک كلاس را بيشينه مي‌كند بدون اينكه توجهي به تمايز كلاس‌ها داشته باشد. از اين جهت روش‌هاي يادگيري تمايزي (Discriminant Training) متعددي مانند بيشينه اطلاعات متقابل (Maximum Mutual Information) ، بيشينه فاصله مدل‌ها (Maximum Model Distance) و MCE [2,3] براي جايگزيني روش بيشينه درست‌نمائي مطرح گرديده‌اند [4].

در مرحله استخراج ويژگي، اطلاعات متمايز كننده مفيدي از سيگنال گفتار استخراج مي‌شود كه به كمک اين اطلاعات، دسته‌بندِ HMM مي‌تواند واحدهاي مختلف گفتار مانند واج‌ها، هجاها يا كلمات را بازشناسي نمايد. در اين راستا اكثر سيستم‌هاي بازشناسي از ويژگي‌هاي مل كپستروم استفاده مي‌نمايند كه نحوه عملكرد گوش و سيستم شنوايي انسان را شبيه‌سازي مي‌كند. با اين وجود، اين ويژگي‌ها براي ايجاد تمايز ميان واحدهاي گفتاري بهينه نيستند و به علاوه كارآيي آنها در محيط‌هاي داراي نويزِ جمع‌پذير كاهش مي‌يابد. بنابراين روش‌هاي مختلفي پيشنهاد شده‌اند كه با اعمال تبديل بر روي اين ويژگي‌ها موجب مقاومت بيشتر آنها در برابر نويز و تمايز بيشتر واحدهاي گفتاري مي‌گردند[7-5].

از جمله اين روش‌ها مي‌توان به PCA اشاره نمود. تبديل PCA يك روش بدون ناظر براي استخراج ويژگي ها مي‌باشد. هدف اصلي در اين روش، تصوير داده‌ها در يك فضا با ابعاد كوچكتر بطوري كه خطاي مجذور مربعات به كمترين مقدار خود برسد. اين روش، ابتدا بردارهاي ويژه مربوط به داده‌ها را بدست مي‌آورد و سپس با نگاشت داده‌ها روي بردارهاي ويژه، اجزاء اصلي داده‌ها را استخراج مي‌كند.

با توجه به اينكه هر كدام از بردارهاي ويژه، ميزان واريانس خاصي از كل واريانس داده‌ها را دارا هستند و اين ميزان واريانس متناسب با قدرت هر يك از بردارهاي ويژه مي‌باشد، مي‌توان براي حفظ يا حذف بعضي از مؤلفه‌هاي بردارهاي ويژگي تصميم‌گيري نمود. از اينرو ابتدا ماتريس تبديل حاصل محاسبه مي‌شود. سپس سطرهايي از ماتريس تبديل كه مربوط به بردارهاي ويژه با اهميت كمتر مي‌باشند، حذف مي‌گردند و در انتها، بردار ويژگي‌ها به فضاي جديد نگاشته مي‌شود.

يكي ديگر از روش‌هاي مطرح، تبديل LDA [8] مي‌باشد. اين روش سعي دارد كه داده‌ها را تا حدامكان به صورت خطي تفكيك‌پذير نمايد. براي اين منظور، لازم است كه نسبت واريانس بين كلاس به واريانس درون كلاس بيشينه گردد[7]. تفاوت اصلي بين PCA و LDA اين است كه PCA سعي بر متعامد كردن ويژگي‌هاي گفتار دارد، در

حالي كه LDA موجب تفكيك كلاس‌هاي مختلف گفتار مي‌گردد. تبديل‌هاي LDA و PCA، و تبديل‌هاي هم‌خانواده با آنها مانند Heteroscedastic LDA [9-11]، Kernel LDA [12] و Kernel PCA [13,14] مي‌توانند جايگزين DCT شوند [6] و يا همراه با آن استفاده گردند[15].

مشكل اصلي تبديل‌هاي مذكور اين است كه معيار بهينه‌سازي آنها با معيار كمينه‌سازي خطاي دسته‌بندي تفاوت دارد [16]، و اين موضوع مي‌تواند منجر به كاهش كارآيي دسته‌بند گردد. روش‌هاي مختلفي براي غلبه بر اين مشكل ارائه شده است. در برخي رويكردها، سعي شده با استفاده از يک معيار همسان، پيوندي ميان مراحل استخراج ويژگي و دسته‌بندي ايجاد شود. بعنوان مثالي از اينگونه روش‌ها مي‌توان به روش آموزش MCE اشاره نمود.

در اين مقاله، چهارچوبي براي بهينه‌سازي ماتريس‌هاي تبديل (مانند PCA و LDA) با استفاده از MCE پيشنهاد شده است تا به اين ترتيب ضعف اين دو روش در عدم توجه به نوع دسته‌بند جبران گردد و كارآيي اين دو روش افزايش يابد.

ساختار ادامه مقاله به اين صورت است. بخش دوم، به تشريح الگوريتم MCE اختصاص دارد. در بخش سوم چهارچوب پيشنهادي براي بهبود ماتريس‌هاي تبديل ويژگي توضيح داده شده است. بخش چهارم حاوي نتايج و ارزيابي روش‌هاي آزمايش شده است، و در ادامه، بخش پنجم نتيجه‌گيري نهايي را در بر دارد.

۲- الگوريتم MCE

روش كمينه نمودن خطاي دسته‌بندي (MCE)، يكي از شناخته‌شده‌ترين روش‌هاي متمايزسازي است كه هم در گروه تبديل ويژگي و هم در گروه آموزش دسته‌بند بكار گرفته مي‌شود [17,18]. هنگامي‌كه روش MCE براي آموزش HMM (به عنوان يک دسته‌بند) استفاده مي‌شود، MCE پارامترهاي مدل را بگونه‌اي تغيير مي‌دهد كه خطاي كل دسته‌بندي كاهش يابد [2]. در روش آموزش به كمک MCE، ابتدا تابع هدف (كه براي يافتن پارامترهاي HMM بكار مي‌رود) با استفاده از يک تابع پيوسته مدل مي‌شود. سپس، كمينه اين تابع با يک روش جستجوي گراديان، مانند Gradient Probabilistic Descent، بدست مي‌آيد [19]. اما از آنجا كه يكي از معايب روش‌هاي جستجوي گراديان، احتمال قرار گرفتن در كمينه‌هاي محلي است، يک روش MCE مبتني بر الگوريتم‌هاي ژنتيک، جهت حل اين مشكل، ارائه شده است[20].

ايده اصلي الگوريتم MCE، بهينه‌سازي نرخ خطاي تجربي با استفاده از يک مجموعه داده آموزشي، جهت بهبود نرخ خطاي بازشناسي مي‌باشد. پس از آنكه نرخ خطاي تجربي، بوسيله يک بازشناس يا دسته‌بند، كمينه شد، تخميني بااياس شده از نرخ خطاي واقعي بدست مي‌آيد. يک راه مؤثر جهت كاهش اين نرخ و بهبود كارآيي كلي سيستم، افزايش مرز ميان كلاس‌ها در داده‌هاي آموزشي است. به اين منظور از گراديان كاهشي [19] يا تركيب امتيازهاي حاشيه و نرخ خطاي تجربي استفاده شده است. مطرح‌ترين رويكرد در اين زمينه روش LM-MCE (Large-Margin MCE) است كه از روش تفكيک‌پذيري نواحي براساس حاشيه بزرگتر نسبت به روش MCE سنتي استفاده مي‌كند [21].

روش MCE براي بدست آوردن تبديل فضاي ويژگي نيز استفاده مي‌شود. در [22]، MCE براي يافتن ماتريس تبديل ويژگي‌ها، كه بتواند خطاي دسته‌بندي را كمينه كند، استفاده شده است. در ادامه و پس از انتشار مرجع [22] ، Wang و Paliwal اين روش را براي بازشناسي واكه‌ها بهبود بخشيدند [23]. در همه اين رويكردها از تابع هزينه‌اي بصورت رابطه (۱) يا (۲) استفاده شده است.

$$d_{k,Add}(O,F) = -g_k(O,F) + \frac{1}{\eta}\log\left[\frac{1}{M-1}\sum_{i=1,i\neq k}^{M}\exp\left(g_i(O,F)\eta\right)\right] \qquad (۱)$$

$$d_{k,Div}(O,F) = \frac{\log\left[\frac{1}{M-1}\sum_{i=1,i\neq k}^{M}\exp\left(g_i(O,F)\eta\right)\right]}{\eta g_k(O,F)} \qquad (۲)$$

كه در آن M تعداد كلاس‌ها، η مقداري مثبت (ارائه شده در [2])، و $g_i(O,\lambda_i)$ لگاريتم احتمال توليد شده براي مشاهده O، به شرط مدل λ_i، بوسيله HMM مي‌باشد و بصورت رابطه (۳) تعريف مي‌شود:

$$g_i(O,\lambda_i) = \log p(O\mid\lambda_i) \qquad (۳)$$

در روش MCE ، هدف كمينه نمودن تابع هزينه (۱) يا (۲) است. اما از آنجا كه اين توابع مشتق‌پذير نيستند، از يک تابع پيوسته نرم مانند تابع سيگموييد بعنوان تابع هزينه استفاده مي‌كنيم. بنابراين خواهيم داشت:

$$l_k(O,F) = \frac{1}{1+\exp\left(-\alpha d_k(O,F)\right)} \qquad (٤)$$

كه $d_k(O,F)$ يكي از دو تابع هزينه (۱) يا (۲) مي‌باشد و α پارامتر تنظيم بزرگتر از يک است. بنابراين اگر $d_k(O,F)$ خيلي كوچكتر از صفر باشد، در واقع يک دسته‌بندي درست رخ داده و $l_k(O,F)$ به صفر نزديک مي‌شود. از طرفي مقدار مثبت براي $d_k(O,F)$ نشان دهنده جريمه براي دسته‌بندي نادرست است و در اينصورت $l_k(O,F)$ به ۱ متمايل مي‌شود. كل خطاي دسته‌بندي براي مشاهده O از رابطه (۵) بدست مي‌آيد:

$$L = \sum_{k=1}^{M} l_k(O,F) \qquad (۵)$$

با استفاده از ماتريس تبـديل، بردارهـاي ويژگـي از فـضاي ويژگـي اصلي به فضاي جديد برده مي‌شود تا تمايز كلاس‌ها بيشتر گردد. تمـايز بيشتر باعث كاهش خطا مي‌شود. از اينرو با استفاده از معيار تابع هزينـه سعي مي‌شود مـاتريس تبـديل بهينـه بدسـت آيـد. مقـادير درايـه‌هـاي ماتريس تبديل بروش تكراري با استفاده از گراديان كاهـشي بـراي تـابع L، با رابطه (٦) قابل محاسبه خواهد بود.

$$w_{n,iter} = w_{n,iter-1} - \beta \frac{\partial L}{\partial w_n} \qquad (٦)$$

رابطه فوق را مي‌توان بصورت ماتريسي بيان كرد:

$$\mathbf{W}_{iter} = \mathbf{W}_{iter-1} - \beta \frac{\partial L}{\partial \mathbf{W}} \qquad (٧)$$

كه $\boldsymbol{W}$ماتريس تبديل؛ r شاخص rامين درايه ماتريس تبـديل $\boldsymbol{W}$؛ $iter$ انديس تكرار الگوريتم، و β پارامتر يادگيري مـي‌باشـد. رابطـه (٧) نشان‌دهنده يک رويه تكراري است. اين رويه زماني متوقف مي‌شـود كـه مقدار تابع هزينه از يک حد آستانه كمتر شود. در اين مقاله از اين روش براي بهبود ماتريس‌هاي تبديل استفاده شده است.

٣- بهينه‌سازي ماتريس تبديل ويژگي‌ها با استفاده از روش MCE (روش پيشنهادي)

همانطور كه اشاره شد، يكي از معايب تبديل‌هـاي LDA و PCA، تفاوت معيار بهينه‌سازي آنها با معيار كمينه‌سازي خطاي دسته‌بندي در دسته‌بندها مي‌باشد؛ كه تأثير منفي روي كارآيي دسته‌بند مـي‌گـذارد. از اين رو در اين بخش چهارچوبي براي بهينه‌سازي اينگونـه تبـديل‌هـا بـا معيار MCE پيشنهاد مي‌شود. فـرض كنيـد مـاتريس تبـديل $\boldsymbol{W}$ بـردار ويژگي‌هاي n بُعدي x را به بردار y بـا بُعـد d $(d \le n)$ تبـديل مـي‌كنـد. هدف، بدست آوردن ماتريس تبديل $\boldsymbol{W}$ است كه تـابع هزينـه را كمينـه نمايد. براي مشتق گرفتن از رابطه (٤) نسبت به $\boldsymbol{W}$ مي‌توان نوشت:

$$\frac{\partial l_k(O,F)}{\partial W} = \frac{\partial l_k(O,F)}{\partial d_k(O,F)} \frac{\partial d_k(O,F)}{\partial W} \qquad (٨)$$

تابع هزينه $l_k(O,F)$ همان رابطه (٤) مـي‌باشـد. بـردار O رشـته مـشاهده يـا رشـته بـردار ويژگـي‌هـاي تبـديل‌يافتـه اسـت و بـصورت $O=\{o_1,o_2,\ldots,o_T\}$ مي‌باشد. همچنين بردار ويژگي‌هاي اصلي بصورت $X=\{x_1,x_2,\ldots x_T\}$ نشان داده شده اسـت. بـا توجـه بـه رابطـه (٤) خواهيم داشت[8] :

$$\frac{\partial l_k(O,F)}{\partial d_k(O,F)} = \alpha l_k(O,F)\left(1-l_k(O,F)\right) \qquad (٩)$$

براي محاسبه بخش دومِ سمت راستِ رابطـه (٨)، ، از روابـط (١) و (٢) استفاده مي‌شود. بعد از مشتق‌گيري از روابط مذكور مي‌توان نوشت:

$$\frac{\partial d_{k,ADD}(O,F)}{\partial W} = -\frac{\partial g_k(O,F)}{\partial W} + \sum_{i=1,i\neq k}^{M}\left(\frac{\exp(g_i(O,F)\eta)}{\sum_{j=1,j\neq k}^{M}\exp(g_j(O,F)\eta)} \times \frac{\partial g_i(O,F)}{\partial W}\right) \qquad (١٠)$$

$$\frac{\partial d_{k,Div}(O,F)}{\partial W} = g_k^{-1}(O,F)\sum_{i=1,i\neq k}^{M}\left(\frac{\exp(g_i(O,F)\eta)}{\sum_{j=1,j\neq k}^{M}\exp(g_j(O,F)\eta)} \times \frac{\partial g_i(O,F)}{\partial W}\right) - \frac{\partial g_k(O,F)}{\partial W} g_k^{-2}(O,F)\log\left[\frac{1}{M-1}\sum_{i=1,i\neq k}^{M}\exp(g_i(O,F)\eta)\right]^{1/\eta} \qquad (١١)$$

در روابط (١٠) و (١١) لازم است كه $\frac{\partial g_j(O,F)}{\partial W}$ محاسبه شـود. از آنجا كه $O=WX$ ، مي‌توان اين عبارت را به صورت زير محاسبه نمـود [8]:

$$\frac{\partial g_j(O,F)}{\partial W} = \sum_{t=1}^{T}\delta(q_t-j)b_j^{-1}(o_t)\frac{\partial b_j(o_t)}{\partial W} = -\sum_{t=1}^{T}\delta(q_t-j)\sum_{m=1}^{M}\gamma_{jm}(o_t)\left[C_{jm}^{-1}(o_t-\mu_{jm})x_t^T\right] \qquad (١٢)$$

كه $j=1,2,\ldots,M$ نشان دهنده M مدل بدست آمـده؛ $\delta(.)$ تـابع دلتاي Kronecker؛ و q_t حالت HMM در زمان t مي‌باشـد. x_t، tامـين بـردار ويژگـي اصـلي و o_t بـردار تبـديل يافتـه x_t بـر اسـاس رابطـه $o_t = Wx_t$ است. همچنين T تعداد مشاهدات و $b_j(o_t)$ احتمال توليد شده براي بردار مشاهده o_t در jامين حالت HMM را نشان مي‌دهد.

$$b_j(o_t) = \sum_{m=1}^{M} c_{jm} b_j(o_t) = \frac{1}{(2\pi)^{n/2}}\sum_{m=1}^{M}\frac{c_{jm}}{|C_{jm}|^{1/2}}\exp\left(-\frac{1}{2}(o_t-\mu_{jm})^T C_{jm}^{-1}(o_t-\mu_{jm})\right) \qquad (١٣)$$

كه در اين رابطه، o_t بردار مشاهده براي فريم tام؛ μ_{jm} بردار ميانگين jامين مخلوط گاوسي در حالت m؛ n ابعاد بردار مشاهده؛ C_{jm} مـاتريس كوواريانس jامين مخلوط گاوسي در حالت m، و c_{jm} وزن jامين مخلوط گاوسي در حالت m مي‌باشد. براى محاسبه $\gamma_{jm}(o_t)$ داريم:

$$\gamma_{jm}(o_t) = \frac{c_{jm} b_{jm}(o_t)}{b_j(o_t)} \qquad (١٤)$$

که $b_{j,m}(o_t)$ احتمال تولید شده با jامین مخلوط گاوسي بـراي بـردار مشاهده o_t در حالت mام مي‌باشد. با جايگزيني (۱۳) و (۱۴) در (۱۲) و بـا جايگـذاري (۱۲) در هـر يـک از معـادلات (۱۱) يـا (۱۰) مقـدار $\frac{\partial d_k(O,F)}{\partial W}$ بدست مي‌آيد. با جانشيني رابطه محاسبه شـده $\frac{\partial d_k(O,F)}{\partial W}$ و رابطه (۹) در (۸) و سپس در رابطه (۷) ماتريس بهينه بدست مي‌آيـد. در روش پيشنهادي ماتريس‌هاي تبديل PCA و LDA به عنوان W در نظر گرفته شده‌اند و ايـن مـاتريس‌هـا بـا اسـتفاده از رابطـه (۷) بهينـه مي‌گردند. در بخش بعدی نتايج حاصل از اين بهينه‌سازی آمده است.

٤- پياده‌سازي و نتايج بدست آمده

براي ارزيابي روش‌هاي پيـشنهادي آزمـايش‌هـايي بـراي تـشخيص کلمات مجزا بر روي دادگان Aurora2 انجام شده است.بعد بردار ويژگي ۳۶ است که شامل ۱۲ ضريب MFCC و مشتق اول و دوم اين ضـرايب مي‌باشد. همچنين هر مدل داراي ۱۶ حالت و ۳ مخلوط گاوسي براي هر حالت مي‌باشد. نام اختصاري روش‌هاي متمايزسـاز موجـود و روش‌هـاي پيشنهادي در جدول (۱) آورده شده است.

٤-١ ارزيابي بهبود PCA با استفاده از MCE

در اين زيربخش، به ارزيـابي و مقايـسه روش‌هـاي مختلـف PCA مبتني بر MCE پرداخته شده است. در اين راستا شش روش با نام‌هاي MFCC، PCA، PCAMCE-D-opt، PCAMCE-A-opt، PCAMCE-D-Init و PCAMCE-A-Init با يکديگر مقايسه شـده‌اند. نام‌هاي مذکور در جدول (۱) توضيح داده شده‌اند.

جدول ۱: نام اختصاري روش‌هاي متمايزساز موجود و روش‌هاي پيشنهادي (روش‌هاي پيشنهادي با فونت ضخيم مشخص شده‌اند)

MCE-A	الگوريتم MCE با استفاده از رابطه (۱)
MCE-D	الگوريتم MCE با استفاده از رابطه (۲)
PCAMCE-A-Init	ماتريس تبديل PCA استفاده شده بعنوان مقدار اوليه ماتريس W در الگوريتم MCE-A
PCAMCE-D-Init	ماتريس تبديل PCA استفاده شده بعنوان مقدار اوليه ماتريس W در الگوريتم MCE-D
PCAMCE-A-opt	ماتريس تبديل PCA بهينه با استفاده از الگوريتم MCE-A
PCAMCE-D-opt	ماتريس تبديل PCA بهينه با استفاده از الگوريتم MCE-D
LDAMCE-A-Init	ماتريس تبديل LDA استفاده شده بعنوان مقدار اوليه ماتريس W در الگوريتم MCE-A
LDAMCE-D-Init	ماتريس تبديل LDA استفاده شده بعنوان مقدار اوليه ماتريس W در الگوريتم MCE-D
LDAMCE-A-opt	ماتريس تبديل LDA بهينه با استفاده از الگوريتم MCE-A
LDAMCE-D-opt	ماتريس تبديل LDA بهينه با استفاده از الگوريتم MCE-D

در جدول (۲) ميـانگين مقـادير خطـاي بازشناسـي گفتـار در هـر مجموعه آزمايش، بصورت جداگانه گزارش شده است. مشاهده مـي‌شـود که روش پيشنهادي PCAMCE-D-opt در هر سـه مجموعـه تـست، نسبت به روش‌هـاي PCA، PCAMC-A-Init و PCAMCE-D-Init عملکرد بهتري داشته‌است.پس از آن، روش‌هاي پيشنهادي ديگـر يعنـي PCAMCE-A-opt مناسب‌ترين کارآيي را نشان مي‌دهند. در مقايـسه با MFCC، روش‌هاي پيشنهادي در مجموعه‌هاي TEST A و TEST B مقـادير WER بيـشتري دارنـد. در حاليکـه در TEST C هـر دو از MFCC بهتر عمل کرده‌اند.

جدول۲:ميانگين مقادير خطای بازشناسی در روش‌هاي PCA-based

	متوسط خطاي بازشناسي کلمه (WER) براي سه مجموعه آزمايش		
	A	**B**	**C**
MFCC	36.50	26.49	59.18
PCA	39.80	33.29	69.21
PCAMCE-A-Init	45.55	38.38	72.47
PCAMCE-D-Init	43.85	35.03	69.98
PCAMCE-A-opt	38.25	27.42	57.89
PCAMCE-D-opt	36.53	27.50	57.48

٤-٢ ارزيابي بهبود LDA با استفاده از MCE

در اين زيربخش، به ارزيـابي و مقايـسه روش‌هـاي مختلـف LDA مبتني بر MCE پرداخته شده اسـت. ايـن روش بـا نـام‌هـاي MFCC، LDA، LDAMCE-D-opt، LDAMCE-A-opt، LDAMCE-D-Init و LDAMCE-A-Init ، که توضيحات آنهـا در جـدول (۱) وجـود دارد، با يکديگر مقايسه شده‌اند.

جدول۳:ميانگين مقادير خطای بازشناسی در روش‌هاي LDA-based

	متوسط خطاي بازشناسي کلمه (WER) براي سه مجموعه آزمايش		
	A	**B**	**C**
MFCC	36.50	26.49	59.18
LDA	46.90	34.45	73.01
LDAMCE-A-Init	45.42	35.69	73.50
LDAMCE-D-Init	46.48	35.41	73.66
LDAMCE-A-opt	38.56	29.44	57.72
LDAMCE-D-opt	37.94	29.10	59.11

با توجه به جدول (۳)، که مقادير خطاي بازشناسي گفتار را در هـر مجموعه آزمايش نشان مي‌دهد، هر دو روش پيشنهادي LDAMCE-A-opt و LDAMCE-D-opt، از روش‌هـاي LDA، LDAMCE-A-Init، و LDAMCE-D-Init کاملاً بهتر عمل نموده‌اند. اين افزايش نرخ کـارآيي در Test A حدود ۸%، در Test B حـدود ٦%، و در Test C حـدود ۱۵% مي‌باشد. همچنين در Test C، نتـايج روش‌هـاي پيـشنهادي از MFCC

نیز بهتـر بـوده اسـت؛ در حالیکـه در Test A و Test B، MFCC نتـایج بهتري را نشان مي‌دهد.

۵- نتيجه‌گيري

چنانکه گفته شـد، تبـدیلات وابـسته بـه داده‌اي همچـون تبـدیل متمایزسازي خطي و آنالیز مؤلفه‌هاي اصلي داراي این ضعف هستند کـه تنها به ماهیت داده‌ها توجه مي‌کنند و به دسته‌بند و خطـاي دسـته‌بنـدِ مورد استفاده براي داده‌ها توجهي ندارنـد. در ایـن مقالـه، روشـي بـراي جبران این ضعف بر اساس معیار خطاي دسته‌بندي کمینه ارائـه شـد و چارچوبي ریاضي براي بهبود ماتریس‌هاي تبدیل وابسته به داده بـر روي الگوریتم‌هاي دسته‌بندي همچون مدل مخفي مـارکف پیـشنهاد گردیـد. نتایج آزمایش‌ها در بازشناسي کلمات مجزا بـر روي دادگـان Aurora2 نشان داد که ماتریس‌هاي تبدیل بهبودیافته‌اي که هم به ماهیت داده‌هـا و هم به کار دسته‌بند توجه دارند، عملکرد بهتري نسبت به ماتریس‌هاي تبدیل اصلي دارند.

مراجع

[1] Rabiner, L., Juang, B.H., *Fundamentals of speech recognition*. Prentice Hall. Englewood Clifs, New Jersey. 1993.

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

[3] He, X. and Deng, L, A new look at discriminative training of hidden markov models. *Pattern Recognition Letters*. Vo. 28, pp.1285-1294, 2007.

[4] Kapadia, S., Valtchev, V., Young, S.J., 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] Nasersharif, B., Akbari, A. *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] Rabiner, L., Juang, B.H., *Fundamentals of speech recognition*. Prentice Hall. Englewood Clifs, New Jersey. 1993.

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

[8] Li, X.B., Li, J.Y. and Wang, R.H., Dimensionality reduction using MCE-optimized LDA. Transformation. *ICASSP*, pp. 137-141, 2004.

[9] Kapadia, S., Valtchev, V., Young, S.J., 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.

[10] Rabiner, L., Juang, B.H., *Fundamentals of speech recognition*. Prentice Hall. Englewood Clifs, New Jersey. 1993.

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

[12] Mika, S., Ratsch, G., Weston J., Scholkopf, B. and Mullers, K.R., Fisher discriminant analysis with kernels. *Proceedings of the IEEE Workshop on Neural Network for Signal Processing*, pp. 41-48, 1999.

[13] Kapadia, S., Valtchev, V., Young, S.J., 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.

[14] Nasersharif, B., Akbari, A. *SNR-dependent compression of enhanced Mel sub-band energies for compensation of noise effects on MFCC features*. Vol. 28, Iss. 11, pp. 1320-1326, 2007.

[15] Abbasian, H., Nasersharif, B., Akbari, A., Rahmani, M., Moin, M.S. Optimized linear discriminant analysis for extracting robust speech features. *3rd International Symposium on Communications, Control and Signal Processing*, pp. 819-824, 2008

[16] Somervuo, P. Experiments with Linear and Nonlinear Feature Transformations in HMM Based Phone Recognition .*IEEE International conference on Acoustics, Speech and Signal Processing*, vol. I, pp. 52-55, 2003.

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

[18] Zhang, B. and Matsoukas, S., Minimum phoneme error based heteroscedastic linear discriminant analysis for speech recognition. *In: Proceedings of ICASSP*, vol. 1. Philadelphia, PA, pp. 925-928, 2005.

[19] Loog, M., Duin, R. P.W., 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.

[20] Gales, M. J. F., Maximum likelihood multiple subspace projections for hidden markov models. *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 2, pp. 37-47, 2002.

[21] Yu, D., Deng, L., He, X., and Acero, A., Use of incrementally regulated discriminative margins in MCE training for speech recognition. *In: Proceedings of Interspeech 2006*, pp. 2418-2421, 2006.

[22] Hung, J. and Lee, L. S., 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] Wang, X. and Paliwal, K. K., Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition*, vol. 36, pp. 2429-2439, 2003.