

ارائه یک روش تبدیل ویژگی مبتنی بر ماکزیمم اطلاعات متقابل برای بازشناسی الگو و گفتار

سید میلاد بصیر^۱، بهزاد زمانی^۱، احمد اکبری^۱، بابک ناصرشریف^{۲،۱}
^۱ آزمایشگاه پردازش صدا و گفتار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
^۲ گروه مهندسی کامپیوتر، دانشکده فنی، دانشگاه گیلان

چکیده

تبدیل ویژگی‌ها یک روش متداول در بازشناسی الگو است که می‌تواند منجر به افزایش دقت رده بندی یا کاهش بار محاسباتی گردد. در این مقاله یک روش تبدیل ویژگی مبتنی بر ماکزیمم اطلاعات متقابل معرفی و ارائه می‌شود که اطلاعات متقابل میان ویژگی‌ها و کلاسها را ماکزیمم می‌کند. تبدیل حاضر با یک روش تکراری و بر اساس افزایش گرادیان نقطه‌ای تخمین زده می‌شود. نتایج آزمایشات انجام گرفته روی دادگان UCI و دادگان TIMIT نشان‌دهنده عملکرد بهتر روش پیشنهادی نسبت به دیگر روشها است. در بین دادگان‌های UCI در دادگان Glass افزایش دقت ۷.۹۴ درصدی، نسبت به بهترین حالت تبدیلات دیگر بدست آورده شده و در دادگان TIMIT این افزایش ۰.۰۷ درصد می‌باشد.

کلمات کلیدی: اطلاعات متقابل، افزایش گرادیان، تبدیل ویژگی، بازشناسی الگو، بازشناسی گفتار

۱- مقدمه

تبدیل ویژگی^۱ اصل مهمی در بازشناسی الگو است. هدف اصلی آن انتقال دادگان با ابعاد زیاد به فضای جدید با ابعاد کمتر می‌باشد، به گونه‌ای که در این فضا توصیف ساختار داده‌ها بهتر انجام گیرد. به عنوان مثال اگر ساختار کلاسهایی که تفکیک‌پذیری غیرخطی‌اند با تبدیلی به خطی تبدیل شود، عملاً کلاس‌بندی آسانتر می‌شود و تفکیک‌پذیری خطی کلاسها افزایش می‌یابد. تبدیل ویژگی خود می‌تواند مراحل استخراج ویژگی^۲، کاهش ویژگی^۳ و تبدیل ویژگی را در برداشته باشد، که برای هر بخش روشهای مختلفی ارائه شده است.

تبدیل ویژگی فرایندی است که در طی آن مجموعه ویژگی جدیدی ایجاد می‌شود. فرایند استخراج ویژگی به صورت یک تابع نگاشت از فضای ویژگی‌های اصلی به فضای ویژگی‌های جدید می‌باشد. ایجاد ویژگی تعداد ویژگی‌ها را زیاد می‌کند، در مقابل استخراج ویژگی تعداد ویژگی‌ها را کاهش می‌دهد. انتخاب ویژگی^۴ زیرمجموعه‌ای از ویژگی‌های اصلی را بدون تغییر آن‌ها انتخاب می‌کند و به نوعی کاهش ویژگی را نیز در بردارد.

در زمینه تبدیل ویژگی می‌توان به روش‌های PCA^۵ و LDA^۶ به عنوان روشهای تبدیل ویژگی خطی و KPCA^۷ و KLDA^۸ به عنوان

تبدیل ویژگی غیرخطی اشاره کرد[۱]. روش آنالیز مولفه اصلی (PCA) بر اساس مقادیر ویژه^{۱۱} و بردارهای ویژه^{۱۱} عمل می‌کند و بردارهای ویژه‌ای را بر می‌گرداند که دارای مقدار ویژه بزرگتری باشند[۲]. لذا می‌تواند از طریق تصویر نمونه‌ها روی بردارهای ویژه مهمتر برای حذف نویز نیز بکار رود[۲]. روش LDA نیز یک روش تبدیل ویژگی باناظر است. در LDA ماتریس تبدیل بر اساس معیار Fisher بدست می‌آید. یعنی تبدیلی را بدست می‌آورد که نسبت کوواریانس بین کلاسی به کوواریانس درون کلاسی را بیشینه کند[۳]. روشهای فوق دارای معایبی می‌باشند[۲]. به عنوان نمونه، PCA به دلیل اینکه مستقل از کلاس‌بندی می‌باشد و بدون ناظر است، تبدیلی را که می‌یابد الزاماً در جهت افزایش دقت کلاس‌بندی نمی‌باشد. LDA نیز فرض می‌کند که ماتریس کوواریانس درون کلاسی برای همه کلاسها برابر باشد، که در شرایط واقعی چنین فرضی نادرست می‌باشد. روش اطلاعات متقابل می‌تواند بر این محدودیت‌ها غلبه کند. نتایج فانو نشان می‌دهد که بیشینه کردن اطلاعات متقابل بین فضای تبدیل یافته و برچسب کلاسها می‌تواند حداقل خطای ممکن را برای کلاس‌بندی بدست آورد. یکی از دلایلی که از اطلاعات متقابل کمتر استفاده می‌شود این است که تابع توزیع نمونه‌ها باید تخمین زده شود که فرآیند زمانبری است. در [۴] و [۵] روشی با نام MRMI مطرح شده که با

⁷ Linear Discriminate Analysis

⁸ Kernel Principle Component Analysis

⁹ Kernel Linear Discriminate Analysis

¹⁰ Eigen value

¹¹ Eigenvector

¹ Feature Transformation

² Pattern Recognition

³ Feature Extraction

⁴ Feature Reduction

⁵ Feature Selection

⁶ Principle Component Analysis

نمی‌تواند به صورت فرم بسته ارزیابی شود لذا حد بالای آن به صورت فرم بسته زیر محاسبه می‌شود [۶]:

$$H_{upp}(F) = - \int_F p(F) \log p_{upp}(F) dF \quad (۴)$$

$$= \frac{n}{2} \log(2\pi) + \frac{1}{2} \log|S| + \frac{n}{2}$$

با توجه به فرم پارامتریک فوق اطلاعات متقابل به صورت رابطه (۵) بدست می‌آید:

$$I_{upp}(F; C) = H_{upp}(F) - H(F|C) (\geq I(F; C)) \quad (۵)$$

$$= \frac{1}{2} \log|S| - \frac{1}{2} \sum_C P_C \log|S_C|$$

فرض کنید ماتریس تبدیل W روی دادگان بفرم $Y = WF$ اعمال شود، که در آن W ماتریس تبدیل با ابعاد $n' \times n$ و بردار ویژگی تبدیل یافته با n' بعد می‌باشد. در فضای تبدیل یافته میانگین و ماتریس کواریانس به صورت $M' = WM$ و $S' = WSW^T$ خواهند بود.

در نتیجه حد بالای اطلاعات متقابل بین نمونه های تبدیل یافته و بر چسب کلاسی در فضای تبدیل یافته از رابطه زیر بدست می‌آید:

$$I_{upp}(Y; C) = H_{upp}(Y) - H(Y|C) \quad (۶)$$

$$= \frac{1}{2} \log|WSW^T| - \frac{1}{2} \sum_C P_C \log|WS_C W^T|$$

در بعضی از کاربردهای بازشناسی الگو مانند سیستم تشخیص گفتار فرض می‌شود تابع چگالی احتمالی برای کلاس‌های مختلف با ماتریس کواریانس گوسی قطری مدل می‌گردد. برای رسیدن به این شرایط سعی می‌شود فاصله بین کواریانس کلی چگالی شرطی کلاسها و کواریانس قطری تخمین زده شده از کواریانس کلی مینیمم شود در حالی که اطلاعات متقابل در فضای جدید ماکزیمم شود. اگر تقریب قطری $p(Y|C)$ با $p_{diag}(Y|C)$ نشان داده شود، داریم:

$$p_{diag}(Y|C) = \frac{1}{(2\pi)^{n/2} |S_C|^{n/2}} \quad (۷)$$

$\exp\left(-\frac{1}{2}[Y - M_C]^T S_{C,diag}^{-1} [Y - M_C]\right)$ که $M_C = WM_C$ و $S_{C,diag} = \text{diag}(WS_C W^T)$ می‌باشند. به منظور محاسبه فاصله بین $p(Y|C)$ و $p_{diag}(Y|C)$ از معیار واگرایی^{۱۲} بین دو چگالی استفاده می‌شود:

$$\text{div}[p(Y|C), p_{diag}(Y|C)] = \int_Y p(Y|C) \log \frac{p_{diag}(Y|C)}{p(Y|C)} dY \quad (۸)$$

و سپس میانگین وزنی آن برای کل کلاسها به صورت زیر محاسبه می‌شود:

$$\text{diagpenalty}(W) = \sum_C P_C \int_Y p(Y|C) \log \frac{p_{diag}(Y|C)}{p(Y|C)} dY \quad (۹)$$

استفاده از بیشینه‌سازی اطلاعات متقابل بین ویژگی‌های تبدیل یافته و کلاس، تبدیل جداساز خطی را یاد می‌گیرد. در این مقالات تخمین اطلاعات متقابل از طریق آنتروپی رینی درجه دوم استفاده شده است. در مقاله [۶] روشی ارائه شده که با فرض اولیه گوسی بودن توزیع کلاسها توانسته رابطه‌ای برای یافتن تبدیلی که اطلاعات متقابل را بیشینه کند، بیابد.

ساختار مقاله به این صورت است که در بخش ۲ به معرفی اطلاعات متقابل و چگونگی رسیدن به تابع هدف پرداخته می‌شود. در بخش ۳ روش پیشنهادی ماکزیمم سازی اطلاعات متقابل بر اساس افزایش گرادیان جهت یافتن ماتریس تبدیل ویژگی بهتر بررسی خواهد شد. در بخش ۴ نتایج آزمایشات و پیاده‌سازی‌ها آورده شده است. و در نهایت نیز جمع‌بندی کلی آورده می‌شود.

۲- اطلاعات متقابل

یک روش تخمین مفید بودن یک ویژگی اندازه اطلاعاتی است که این ویژگی برای جداسازی کلاسها دارد [۶]. در این حالت معیار اطلاعات متقابل بین بردار ویژگی و کلاسها می‌تواند معیار مناسبی در این زمینه باشد. اگر فرض کنیم که C یک متغیر گسسته و نشان دهنده برچسب کلاس، و F متغیری پیوسته با n بعد و بردار ویژگی باشد، اطلاعات متقابل بین F و C از رابطه زیر بدست می‌آید:

$$I(F; C) = H(F) - H(F|C) \quad (۱)$$

که در این رابطه $H(F)$ نشان دهنده آنتروپی متغیر تصادفی F و $H(F|C)$ نشان دهنده آنتروپی F است زمانی که C شناخته شده باشد. لذا اطلاعات متقابل عملاً کاهش آنتروپی متغیر F است، زمانی که متغیر C معلوم باشد. با توجه به روابط $H(F)$ و $H(F|C)$ رابطه (۱) بفرم زیر بانویسی می‌شود:

$$I(F; C) = \sum_C P_C \int_F p(F|C) \log \left[\frac{p(F|C)}{p(F)} \right] dF \quad (۲)$$

که در آن P_C احتمال اولیه کلاس C و $p(F|C)$ نشان دهنده توزیع نمونه‌ها در کلاس C و $p(F)$ نشان دهنده توزیع کلی نمونه‌ها می‌باشد. از آنجایی که هیچکدام از این دو توزیع به طور صریح شناخته شده نیستند، بایستی تخمین زده شوند. بنابراین فرم پارامتریکی در نظر گرفته شده و پارامترهای آن از روی نمونه‌های دادگان تخمین زده می‌شود. فرض کنید نمونه‌ها در فضای ویژگی F در هر کلاس C توزیع گوسین با میانگین M_C و کواریانس S_C را دارا باشند، آنگاه آنتروپی شرطی به صورت زیر محاسبه می‌شود:

$$H(F|C) = \sum_C P_C \left[\frac{n}{2} \log(2\pi) + \frac{1}{2} \log|S_C| + \frac{n}{2} \right] \quad (۳)$$

که n تعداد ابعاد فضای ویژگی، $M = \sum_C P_C M_C$ میانگین کلی نمونه‌ها، $S = \sum_C P_C [S_C + S_B]$ کواریانس کلی نمونه‌ها و $H(F)$ وجود $S_B = (M - M_C)(M - M_C)^T$ می‌باشد. با این وجود

¹² Divergence metric

که در آن $\nabla(f(X)) = \frac{\partial f}{\partial X}$ و مقدار پارامتر λ ضریب یادگیری است و مقدار آن سرعت همگرایی را تعیین می‌کند. این الگوریتم با یک نقطه دلخواه X_0 شروع کرده و در هر مرحله به روز می‌شود. به منظور استفاده از روش افزایش گرادینان می‌بایست مشتق تابع هدف (۱۲) را نسبت به متغیر W محاسبه می‌شود که به صورت زیر می‌باشد [۹].

$$\frac{\partial f(W)}{\partial W} = 2 \sum_C P_C [D_C W S_C] - [(W S W^T)^{-1}]^T W S^T \quad (14)$$

که D_C یک ماتریس قطری می‌باشد که تأمین درایه قطری که با $D_{C,i}$ نشان می‌دهند به صورت زیر به دست می‌آید:

$$D_{C,i} = \frac{1}{w_i S_C w_i^T} \quad (15)$$

با استفاده از رابطه (۱۵) ماتریس D_C محاسبه شده سپس با توجه به روابط (۱۳) و (۱۴) می‌توان تابع هدف (۱۲) را محاسبه کرد.

۴. نتایج آزمایشات

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 از مجموعه دادگان TIMIT و نیز مجموعه دادگان Glass Iris و CMC مربوط به دادگان UCI [۱۰] نیز استفاده شده است.

۴.۱. نتایج دادگان UCI

در این زیر بخش نتایج مربوط به دادگان UCI آورده می‌شود. مشخصات مجموعه دادگان در جدول (۱) آورده شده است. برای کلاس‌بندی از ابزار Weka [۱۱] استفاده شده است. کلاس‌بندی مورد استفاده در این بخش در سه گروه طبقه بندی می‌شوند: گروه کلاس‌بندی مبتنی بر فاصله مانند IBK، J1B و KStar، گروه کلاس‌بندی آماری مانند NaiveBayes، SMO و RBF، گروه کلاس‌بندی مبتنی بر درخت تصمیم‌گیری مانند NBTREE و Random Tree، J48 و LMT در ادامه نتایج کلاس بندی حاصل از هر یک از گروه‌ها آورده شده است. در این جا مقدار ۰.۰۱ برای پارامتر λ تنظیم می‌شود.

جدول ۱ - مشخصات مجموعه دادگان UCI [13]

نام دادگان	تعداد کلاس	تعداد ویژگی	تعداد نمونه
Iris	۳	۴	۱۵۰
CMC	۶	۱۰	۱۴۷۳
Glass	۶	۹	۲۱۴

جدول (۲) نتایج کلاس‌بندی را روی کلاس‌بندی آماری و دادگان مختلف نشان می‌دهد. ستون‌های جدول نشان‌دهنده روش تبدیل ویژگی و سطرها نشان‌دهنده متد کلاس‌بندی می‌باشد. روش پیشنهادی یعنی MMI گرادینان افزایشی در کنار روشهای LDA و PCA مورد ارزیابی قرار گرفته است. ستون Normal نتایج کلاس‌بندی را برای

چون مقادیر *diagpenalty* منفی می‌باشند بنابراین مینیمم فاصله بین $p(Y|C)$ و $p_{diag}(Y|C)$ معادل ماکزیمم کردن *diagpenalty* است [۸]. برای ماکزیمم کردن *diagpenalty* از اطلاعات متقابل استفاده می‌شود. تابع هدف جدید به صورت زیر خواهد بود:

$$I_{diagpenalty}(Y; C) = H(Y) - H(Y|C) + diagpenalty(W) \\ = - \int_Y p(Y) \log p(Y) dY \\ + \sum_C P_C \int_Y p(Y|C) \log p_{diag}(Y|C) dY \quad (10)$$

با توجه به اینکه $H(Y)$ را نمی‌توان به فرم بسته نشان داد، لذا حد بالای آن از طریق رابطه $\int_Y p(Y) \log p_{app}(Y) dY$ محاسبه می‌شود. از اینرو رابطه (۱۰) به صورت زیر بازنویسی می‌شود:

$$I_{diagpenalty,app}(Y; C) = - \log |W S W^T| \\ + \sum_C P_C \sum_{i=1}^{n'} \log [w_i S_C w_i^T] \quad (11)$$

ماتریس تبدیل W بگونه‌ای محاسبه می‌شود که $I_{diagpenalty,app}(Y; C)$ ماکزیمم شود. یعنی:

$$f(W) = \arg \max_W \sum_C P_C \left[\sum_{i=1}^{n'} \log (w_i S_C w_i^T) \right] - \log |W S W^T| \quad (12)$$

۳. ماکزیمم سازی تابع هدف مبتنی بر اطلاعات

متقابل با افزایش گرادینان

در این مقاله از روش افزایش گرادینان به منظور یافتن نقطه ماکزیمم تابع هدف استفاده شده است. این روش از آنجایی که در بهینه‌های محلی به خوبی عمل می‌کند و ممکن است در بهینه‌های محلی قرار گیرد، لذا نقطه شروع را یکی از نقاط PCA یا LDA در نظر گرفته می‌شود. نقطه برتری روش ما نسبت به دو روش دیگر آن می‌باشد که روش PCA مستقل از کلاس نمونه‌ها عمل می‌کند و لذا بیشتر در شرایط نویزی خوب عمل می‌کند. از طرفی در روش LDA محدودیت برابری ماتریس کوواریانس درون کلاسی را داریم و رابطه (۶) همان تابع هدف LDA می‌باشد اگر تمامی P_C ها مساوی و برابر P_W باشد. اما در روش پیشنهادی علاوه بر برطرف کردن این محدودیت با وارد کردن فرض کوواریانس قطری باعث استقلال بیشتر مجموعه ویژگی حاصل از تبدیل می‌شویم.

در مبحث گرادینان معمولاً از کاهش گرادینان استفاده می‌شود و معمولاً بر روی خطا تعریف می‌گردد تا بتوان به حداقل خطای ممکن رسید. (مانند الگوریتم انتشار رو به عقب در شبکه عصبی MLP) ولی در اینجا چون تابع هدف قرار است ماکزیمم گردد لذا از افزایش گرادینان استفاده می‌شود. با فرض داشتن تابع $f(x)$ به منظور یافتن نقطه ماکزیمم آن می‌توان از رابطه زیر استفاده کرد:

$$X_{k+1} = X_k + \eta \nabla(f(X_k)) \quad (13)$$

نتایج کلاس‌بندهای مبتنی بر درخت تصمیم در جدول (۴) نشان داده شده است. با توجه به نتایج مندرج در جدول، روش MMI در اغلب موارد بیشترین نتیجه بازشناسی را به خود اختصاص داده است.

جدول ۴ - نتایج کلاس‌بندهای مبتنی بر درخت تصمیم روی دادگان

نگاشت یافته با تبدیلات مختلف

		Normal	LDA	PCA	MMI
NBTree	Iris	94	<u>96.67</u>	90	<u>96.67</u>
	Glass	<u>95.32</u>	89.71	92.99	<u>98.59</u>
	CMC	42.7	33.33	42.7	<u>43.92</u>
J48	Iris	96	<u>96.67</u>	95.33	94
	Glass	<u>96.72</u>	92.05	<u>97.66</u>	<u>96.72</u>
	CMC	38.36	33.33	42.56	<u>43.92</u>
RandomTree	Iris	91.33	82	70	<u>96.67</u>
	Glass	<u>81.77</u>	67.28	80.37	<u>99.06</u>
	CMC	38.15	30.67	37.13	<u>40.32</u>
LMT	Iris	94	<u>98</u>	<u>96.67</u>	95.33
	Glass	<u>97.66</u>	<u>94.85</u>	<u>97.66</u>	<u>97.66</u>
	CMC	39.51	28.67	<u>42.7</u>	<u>40.86</u>

روش پیشنهادی بر اساس ماکزیمم کردن اطلاعات متقابل عمل می‌کند. این روش سعی دارد تبدیلی بدست آورد که اطلاعات متقابل ویژگی و کلاس‌ها را افزایش دهد. در صورتیکه دیگر روشها با کلاس‌بند ارتباطی ندارند. از اینرو عملکرد بهتر روش پیشنهادی مورد انتظار بود. این عملکرد در کلاس‌بندهای مختلف و دادگان مختلف قابل مشاهده می‌باشد.

۴.۲. نتایج دادگان TIMIT

دادگان آکوستیک/فونتیک TIMIT شامل ۶۳۰ گوینده بوده و هر گوینده ۱۰ جمله را ادا کرده است [۱۲]. در TIMIT ۶۱ فونم وجود دارد، که در این پژوهش به ۳۹ فونم کاهش یافته است. مدل مخفی مارکوف (HMM) به عنوان کلاس‌بند مورد استفاده قرار گرفته است. برای پیاده سازی HMM از ابزار HTK استفاده شده است. به همین منظور برای هر فونم یک مدل با ۳ حالت و ۱۶ مخلوط گوسی در نظر گرفته شده است. بردار ویژگی ۳۹ بعدی است که شامل ۱۲ ضریب mfcc و انرژی و مشتقات اول و دوم آنها می‌باشد. نتایج برای سه روش تبدیل ویژگی PCA، LDA و MMI و بدون تبدیل ویژگی در جدول (۵) آورده شده است.

روش	نرخ صحیح بازشناسی فونم (درصد)
Normal	71.90
PCA	71.66
LDA	70.50
MMI	71.73

همانطور که مشاهده می‌شود روش پیشنهادی نسبت به دو روش تبدیل ویژگی دیگر باعث افزایش ۱.۲۳ درصدی نسبت به روش LDA

دادگانی که هیچ تبدیلی روی آنها انجام نگرفته است را نشان می‌دهد. اعداد مندرج در جدول نتایج کلاس‌بندی 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.

چنانچه در جدول (۲) مشاهده می‌شود برای هر سه کلاس‌بند آماری روش پیشنهادی یعنی MMI نسبت به روشهای تبدیل ویژگی نتایج بالاتری در کلاس بندی دارد. به عنوان مثال این بهبود در دادگان Glass از ۸۹.۲۵ درصد برای روش تبدیل ویژگی PCA به ۹۷.۱۹ درصد برای روش تبدیل ویژگی MMI رسیده است.

جدول ۲ - نتایج کلاس‌بندهای آماری روی دادگان نگاشت یافته با

تبدیلات مختلف

		Normal	LDA	PCA	MMI
NaveBays	Iris	96	<u>98</u>	93.33	<u>97.33</u>
	Glass	84.11	<u>89.25</u>	84.11	<u>97.19</u>
	CMC	<u>42.49</u>	27.33	38.96	<u>41.01</u>
SMO	Iris	96	96	95.33	<u>97.33</u>
	Glass	80.84	79.43	<u>81.3</u>	<u>89.25</u>
	CMC	<u>42.7</u>	34	<u>42.7</u>	<u>42.7</u>
RBF	Iris	95.33	<u>97.33</u>	94.67	96
	Glass	93.45	89.25	<u>94.39</u>	<u>95.79</u>
	CMC	<u>43.38</u>	20.67	40.46	<u>45.14</u>

در جدول (۳) نتایج کلاس‌بندهای مبتنی بر فاصله بیان شده است. در اینجا سه روش کلاس‌بندی KStar، k و برابر یک (KNN-1) و سه (KNN-3) مد نظر قرار همسایگی، k، برابر یک (KNN-1) و سه (KNN-3) مد نظر قرار گرفته است. با توجه به نتایج، روش تبدیل ویژگی پیشنهادی برای تمامی حالات دارای نتیجه بهتری است. این تفاوت در دادگان با تعداد ویژگی و تعداد کلاس بیشتر مانند Glass و CMC مشهودتر است.

جدول ۳ - نتایج کلاس‌بندهای مبتنی بر فاصله روی دادگان نگاشت

یافته با تبدیلات مختلف

		Normal	LDA	PCA	MMI
Kstar	Iris	94.67	<u>95.33</u>	90.67	<u>98.67</u>
	Glass	89.71	85.04	<u>91.12</u>	<u>98.13</u>
	CMC	<u>39.71</u>	30.67	35.91	<u>46.23</u>
KNN-1	Iris	<u>95.33</u>	93.33	92.67	<u>96.67</u>
	Glass	91.12	92.05	<u>93.45</u>	<u>99.53</u>
	CMC	<u>35.64</u>	32	35.09	<u>38.42</u>
KNN-3	Iris	<u>95.33</u>	93.33	94	<u>97.33</u>
	Glass	<u>92.05</u>	91.12	89.25	<u>97.66</u>
	CMC	<u>38.64</u>	21.33	37.74	<u>42.7</u>

[10] C. Blake, E. Keogh, C. J. Merz, "UCI Repository of Machine Learning Databases," 1998. Available: www.ics.uci.edu/mlearn/MLRepository.html
Processing, vol. 37. NO. 11. 1989

[11] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten, "The WEKA Data Mining Software: An Update"

[12] KAI-FU LEE and HSIAO-WUEN HON, "Speaker-Independent Phone Recognition Using Hidden Markov Model," IEEE Transactions on Acoustics, Speech, and Signal

و ۰۰۷ در صدی نسبت به روش PCA در نرخ صحیح بازشناسی فونم شده است.

۵. نتیجه‌گیری

در این مقاله روش ماکزیمم سازی اطلاعات متقابل مبتنی بر افزایش گرادیان برای تبدیل ویژگی‌ها در بازشناسی الگو پیشنهاد شده و با روابط بیان شده است. در این روش ماتریس تبدیلی بدست می‌آید که اطلاعات متقابل ویژگی‌ها و کلاسها را افزایش می‌دهد. روش فوق به نقطه شروع حساس بوده و ممکن است در مینیمم‌های محلی قرار بگیرد. از اینرو برای شروع از ماتریسهای تبدیل LDA و یا PCA استفاده می‌شود که حداقل برتری آن بهتر بودن نتایج عملکرد ماتریس تبدیل نسبت به LDA و PCA است. از سه گروه کلاس‌بند مبتنی بر فاصله، آماری و مبتنی بر درخت تصمیم جهت ارزیابی تاثیر روش پیشنهادی نسبت به دیگر روشها بر نرخ کلاس‌بندی استفاده شده است. نتایج نشان می‌دهند که روش پیشنهادی برای تمامی دادگان و تمامی کلاس‌بندهای مبتنی بر فاصله عملکرد بهتری داشته است. به علاوه عملکرد MMI در کلاس‌بندی دادگان گفتار TIMIT مورد ارزیابی قرار گرفت که در بازشناسی فونم نیز نتایج بهتری حاصل شد.

مراجع

- [1] A. Kocsor, L. Toth, "Kernel-Based Feature Extraction with a Speech Technology Application, IEEE Transactions On Signal Processing," Vol. 52, No. 8, pp. 2250-2263, 2004.
- [2] Kari Torkkol, William M. Campbell, "Mutual Information in Learning Feature Transformations," in Proceedings of the 17th International Conference on Machine Learning, Stanford, CA, USA, June 29 - July 2 2000, pp. 1015-1022.
- [3] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, "Wiley-Interscience, second edition," 2001
- [4] K. Torkkola, "Feature extraction by non-parametric mutual information maximization," The Journal of Machine Learning Research, vol. 3, no. 7-8, pp. 1415-1438, 2003.
- [5] K. E. Hild II, D. Erdogmus, K. Torkkola, and J. C. Principe, "Feature extraction using information-theoretic learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 9, pp. 1385-1392, 2006.
- [6] Mukund Padmanabhan, Satya Dharanipragada, "Maximizing Information Content in Feature Extraction," IEEE Transactions on Speech and Audio Processing, Vol. 13, No. 4, July 2005.
- [7] N. Kumar, and A. G. Andreou, "Heteroscedastic discriminant analysis and reduced rank HMM's for improved speech recognition," Speech Commun., vol. 26, pp. 283-297, 1998.
- [8] M. J. F. Gales, "Semi-tied covariance matrices," in Proc. ICASSP, 1999.
- [9] K. Fukunaga, Introduction to Statistical Pattern Recognition. New York: Academic Press, 1990.