تحلیل صدای گریه نوزاد با استفاده از طبقه‌بند بازنمایی تنک مبتنی بر هسته

امیر اخوان، محمد حسن مرادی*

گروه مهندسی بیوالکتریک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر

چکیده

پردازش صدای گریه نوزاد اطلاعات مفیدی در مورد وضعیت نوزاد در اختیار قرار می‌دهد. این اطلاعات می‌تواند به منظور تشخیص بیماری و یا درک نیاز نوزاد استفاده شود. این مقاله به تحلیل صدای گریه نوزاد با روی‌کرد تفکیک دو نوع منشاء درد و گرسنگی در صدای گریه پرداخته است. الگوهای بازنمایی تنک علامت (سیگنال) یکی از جدیدترین ابزارهای پردازش در حوزه بازشناختی الگو است. از این‌رو، در مقاله جاری چارچوبی جدید برای استفاده از این الگوها در طبقه‌بندی انواع صدای گریه نوزاد ارائه می‌شود. به منظور طراحی دیکشنری در الگوی تنک پیشنهادی از اطلاعات طیفی با تفکیک‌پذیری (رزولوشنی) مشابه سامانه شنوایی انسان (ضرایب کِپستروم بسامد مِل) استفاده شده است. دیکشنری نهایی از انتقال این اطلاعات به فضای هسته تشکیل می‌شود. بررسی‌های انجام شده نشان می‌دهند که طبقه‌بند بازنمایی تنک مبتنی بر هسته عملکرد قابل قبولی در تفکیک دو نوع صدای گریه نوزاد دارد. به منظور مقایسه، خروجی روش پیشنهادی به همراه نتایج تعدادی از طبقه‌بندهای معروف این حوزه و طبقه‌بند بازنمایی تنک متداول ارائه شده است. نتایج نشان می‌دهند که الگوی بازنمایی تنک مبتنی بر هسته به طور کلی عملکرد بهتری نسبت به سایر طبقه‌بندهای ارائه شده دارد. الگوی تنک پیشنهادی علامت‌های گریه دو رده داده‌ها را به ازای روش اعتبارسنجی ۶- لایه با دقتی بیش از ۹۳ درصد تفکیک می‌نماید. علامت‌های بکار برده شده در این مقاله در مجموع از ۵۱ نوزاد سالم (۱۹ نوزاد پسر و ۳۲ نوزاد دختر) ثبت گردیده‌اند.

کلید واژه‌ها: الگوی بازنمایی تنک، طبقه‌بند بازنمایی تنک، الگوی هسته-پایه، ضرایب کپستروم، صدای گریه نوزاد

۱. مقدمه

گریه کردن یکی از مهم‌ترین افعال ارتباطی نوزاد است. از این‌رو بررسی موجدیس[1] علامت (سیگنال) صوتی ثبت شده در هنگام گریه نوزاد می‌تواند حاوی اطلاعات مفیدی از جسم و روان وی باشد [۱]. تحقیقات موجود در این حوزه نشان می‌دهد که از مشخصه‌های صوتی گریه نوزاد می‌توان در راستای درک نیاز و یا تشخیص یک بیماری خاص در او استفاده نمود. به طور کلی تلاش‌های صورت گرفته در حوزه پردازش صدای گریه نوزاد را می‌توان به دو گروه تقسیم نمود. در گروه اول هدف از پردازش صدای گریه تشخیص یک بیماری خاص در نوزاد است. در این میان می‌توان به تشخیص مشکلات شنوایی [۲]، اوتیسم[2] [۳]، بیماری‌های سامانه عصبی مرکزی [۴] و بیماری‌های تنفسی [۵] اشاره نمود. در بسیاری از این موارد تشخیص زود هنگام مشکل نوزاد، به روند درمان او کمک کرده و احتمال بهبودی را افزایش می‌دهد. گروه دوم به شناسایی وضعیت نوزاد از جمله درد [۶-۷]، گرسنگی [۸-۹]، ترس [۱۰] و خواب‌آلودگی [۱۱] می‌پردازد.

انتخاب و استخراج بردار ویژگی از صدای گریه یکی از مراحل مهم و تأثیرگذار در نتیجه تحلیل صدای گریه نوزاد است. با توجه به تحقیقات موجود در این حوزه مهم‌ترین ویژگی‌های بکار برده شده در بازشناختی و طبقه‌بندی صدای گریه نوزاد عبارتند از: ویژگی زمان- بسامد [۱۲،۵]، ضرایب پیش‌بینی خطی[3] [۱] و ضرایب کپستروم بسامد مِل[4] [۱۳]. از این‌رو در مقاله جاری از ضرایب کپستروم به عنوان بردار ویژگی استفاده می‌شود.

یکی از ابزارهای جدید پردازش علامت، الگوهای‌های بازنمایی تنک[5] است. این الگوها در کاربردهای مختلفی از جمله کاهش نوفه (نویز) [۱۴]، فشرده‌سازی [۱۵] و بازشناختی الگو [۱۶] بکار برده شده‌اند. تنک بودن نقش بسیار مهمی در

* نویسنده پاسخگو: mhmoradi@aut.ac.ir

[1] Waveform
[2] Autism
[3] LPC; Linear Prediction Coefficient
[4] MFCC; Mel Frequency Cepstrum Coefficient
[5] Sparse Representation Models

[DOR: 20.1001.1.23455748.1394.3.1.7.7]

بسیاری از حوزه‌های پردازش علامت و تصویر دارد. به عبارت دیگر ویژگی تنک بودن یک قید مناسب برای حل بسیاری از مسائل معکوس است [۱۷]. در الگوهای تنک تلاش می‌شود که علامت مورد بررسی براساس ترکیب خطی تعداد کمی پایه نمایش داده شود. به هر یک از این پایه‌ها، اتم و به مجموعه اتم‌ها، دیکشنری گفته می‌شود (دیکشنری ماتریسی دو بعدی است و اتم‌ها ستون‌های آن را تشکیل می‌دهند). در حالت کلی ابعاد هر اتم بسیار کم‌تر از تعداد اتم‌های موجود در دیکشنری است، بنابراین، با یک دستگاه معادلات زیر-معین[1] با بینهایت پاسخ روبرو هستیم. در این شرایط قید تنک بودن، انتخاب یک پاسخ مناسب را از بین تعداد زیادی پاسخ، ممکن می‌سازد. اخیراً، تلاش‌های موجود در این حوزه نشان داده‌اند که ضرایب بردار تنک می‌تواند به عنوان یک معیار متمایزکننده در کاربردهای بازشناختی الگو و طبقه‌بندی استفاده شود [۱۸]. اساس طبقه‌بندی با استفاده از الگوهای تنک بر این مبنا است که بازنمایی تنک داده‌های آزمون بر حسب علامت‌های آموزش می‌تواند ماهیت داده‌های آزمون را با توجه به بردار ضرایب تنک تعیین نماید. لازم به ذکر است که انتخاب دیکشنری مناسب یکی از عوامل تأثیرگذار در نتیجه این‌گونه الگوها می‌باشد. از این رو تلاش‌های زیادی به منظور یادگیری و تولید یک دیکشنری متمایزکننده انجام شده‌اند [۱۹-۲۰].

در مطالعات موجود در حوزه طبقه‌بندی، از الگوهای تنک از داده‌های آموزش به طور مستقیم در طراحی دیکشنری مورد نیاز استفاده می‌شود [۲۱-۲۲]، به عبارت دیگر فضای اتم‌های دیکشنری دقیقاً با فضای داده‌های آموزش و آزمون یکسان است. این روی‌کرد در کاربردهایی مناسب است که در فضای داده‌های مسئله اطلاعات متمایزکننده به منظور تفکیک طبقه‌ها از یکدیگر موجود باشند. چنین شرایطی در کاربرد جاری (طبقه‌بندی صدای گریه نوزاد) صادق نیستند. بنابراین در مقاله پیش‌رو مجموعه داده‌ها ابتدا به یک فضای ویژگی متمایزکننده منتقل شده و در ادامه دیکشنری الگوی تنک در فضای جدید ساخته می‌شود.

در این مقاله برای اولین بار از الگوهای بازنمایی تنک به منظور طبقه‌بندی صدای گریه نوزاد استفاده شده است. با توجه به این که ضرایب کپستروم بسامد مل یکی از ویژگی‌های مناسب در پردازش صدای گریه نوزاد است، نحوه طراحی دیکشنری با استفاده از این ضرایب ارائه می‌شود. به منظور بهبود دقت تفکیک، یکی از انواع الگوهای بازنمایی تنک با عنوان طبقه‌بند بازنمایی تنک مبتنی بر هسته[2] (هسته- پایه) پیشنهاد می‌شود. به منظور بررسی کارایی روش پیشنهادی، نتایج با خروجی تحلیل متمایزکننده خطی[3]، طبقه‌بند بازنمایی تنک متداول، ماشین بردار پشتیبان[4] و الگوی مخفی مارکوف[5] مقایسه شده‌اند. در ادامه به جزئیات مربوط به داده‌های مورد استفاده در این مقاله و نحوه استخراج بردار ویژگی به منظور ساخت دیکشنری پرداخته می‌شود. سپس طبقه‌بند بازنمایی تنک و طبقه‌بند بازنمایی تنک مبتنی بر هسته (هسته- پایه) به طور مختصر بیان می‌گردند. در انتها نتایج بکارگیری الگوی پیشنهادی بر روی داده‌های ثبت و نتیجه‌گیری ارائه می‌شوند.

۲. روش

در این بخش ابتدا مجموعه علامت‌ها و شرایط ثبت داده‌های مورد استفاده در این مقاله توصیف می‌شوند و در ادامه نحوه استخراج بردار ویژگی از آن‌ها ارائه شده است.

۲-۱. ثبت داده

داده‌های مورد استفاده در این مقاله به دو گروه علامت‌های (سیگنال‌های) مربوط به درد و علامت‌های مربوط به گرسنگی نوزاد تقسیم می‌شوند. علامت‌ها توسط ذخیره‌کننده سونی الگوی اِس‌ایکس-۷۵۰[6] با بسامد ۱۶ کیلوهرتز نمونه‌برداری و با دقت ۱۶ بیت کوانتیزه (کمیت‌دهی) شده‌اند. به منظور ثبت، میکروفون در فاصله ۱۰ سانتی‌متری دهان نوزاد قرار داده شد. در داده‌برداری مربوط به هر دو گروه (درد و گرسنگی) نوزادان ۲ تا ۳ روزه حضور داشتند. داده‌های گروه درد از نوزادانی که به واسطه ضربه لانست به پاشنه پا گریه می‌کنند، ثبت گردیدند. این داده‌ها در اتاقی آرام از مجموع ۳۲ نوزاد (۱۴ نوزاد پسر و ۱۸ نوزاد دختر) به دست آمده‌اند.

داده‌های مربوط به گروه گرسنگی قبل از شیردهی مادر

[1] Under-determined

[2] KSRC; Kernel Sparse Representation Classification
[3] LD; Linear Discriminative Analysis
[4] SVM; Support Vector Machine
[5] HMM; Hidden Markov Model
[6] Sx-750

[DOR: 20.1001.1.23455748.1394.3.1.7.7]

نوزادها ثبت گردیدند. در داده‌های این گروه ۱۹ نوزاد (۵ نوزاد پسر و ۱۴ نوزاد دختر) شرکت داشتند [۲۳].

۲-۲. استخراج بردار ویژگی با استفاده از ضرایب کپستروم بسامد مل

ضرایب کپستروم بسامد مل یکی از مهم‌ترین ویژگی بکاربرده شده در حوزه پردازش گفتار می‌باشند [۲۴]. اگر چه الگوهای موجود در علامت‌های (سیگنال‌های) گفتار با علامت مربوط به گریه نوزاد متفاوت هستند؛ ولی این ویژگی‌ها در بازشناختی و ایجاد تمایز انواع گریه نوزاد مؤثر می‌باشند. این ضرایب با الهام گرفتن از حساسیت غیرخطی گوش انسان نسبت به بسامدهای مختلف محاسبه می‌شوند. در ادامه روند محاسبه ضرایب کپستروم مل از داده‌های بکاربرده‌شده در این مقاله به اختصار بیان می‌شود.

ابتدا هر داده یک ثانیه‌ای به قطعات ۱۰۰ میلی‌ثانیه با هم‌پوشانی ۵۰٪ تقسیم می‌شود. به منظور ایجاد پیوستگی و کاهش اثر مولفه‌های بسامد بالا (ناشی از قطعه کردن علامت) از پنجره‌گذاری همینگ[1] در حوزه زمان استفاده می‌شود. داده‌های پنجره‌گذاری شده با استفاده از تبدیل فوریه سریع مطابق رابطه زیر به حوزه بسامد منتقل می‌شوند.

$$X(k)=\sum_{n=1}^{N} x(n)e^{-jk\frac{2\pi}{N}n}, \qquad (1)$$

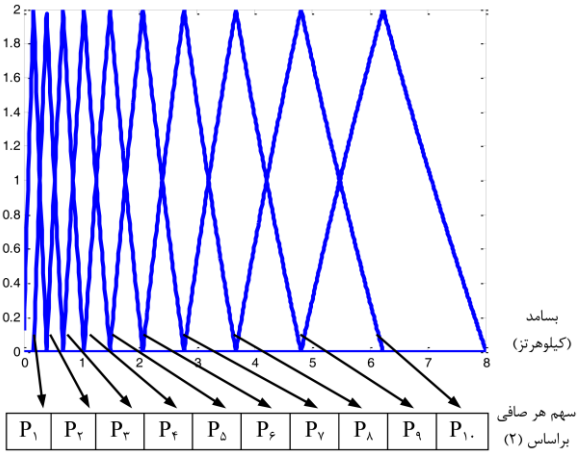
که در آن، $x(n)$ علامت ۱۰۰ میلی‌ثانیه‌ای در هر قطعه می‌باشد. در ادامه سهم هر یک از صافی‌های موجود در بانک صافی مل از طیف علامت به کمک رابطه ۲ محاسبه می‌شود.

$$P_i=\log\left(\sum_{k=1}^{N}(|X(k)|H_i(k))\right) \qquad (2)$$

در این رابطه H_i، صافی i-ام بانک صافی مل و $X(k)$ طیف علامت را نشان می‌دهند. بانک صافی مل از تعدادی صافی با پهنای باند یکسان در حوزه مل تشکیل شده است. نمونه‌ای از بانک صافی مل با دامنه مثلثی به ازای بسامد نمونه‌برداری ۱۶ کیلوهرتز به همراه سهم هر یک از صافی‌ها از طیف علامت مطابق رابطه ۲ در شکل ۱ نمایش داده شده‌اند. رابطه بین بسامد مرکزی هر یک از این صافی‌ها با مقیاس مل به صورت زیر است:

$$mel=\frac{1000}{ln(1+\frac{1000}{700})}ln(1+\frac{f}{700}), \qquad (3)$$

که در آن، f و mel به ترتیب مقیاس بسامد و مل می‌باشند.



شکل ۱ نمایش بانک صافی مل به ازای ۱۰ صافی و سهم هر صافی از طیف علامت با استفاده از رابطه ۲.

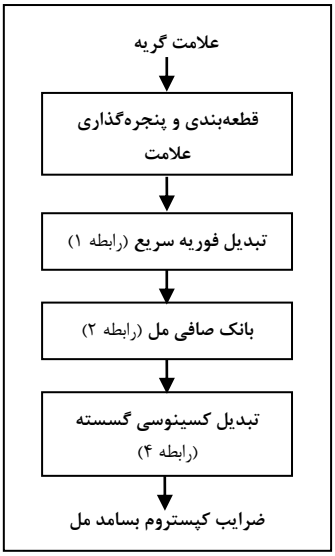
همان‌طور که در شکل ۱ مشخص است پهنای باند این صافی‌ها بر حسب هرتز متفاوت است، این درحالی است که با استفاده از رابطه ۳ می‌توان نشان داد که در مقیاس مل همه صافی‌ها دارای پهنای باند یکسان هستند. با توجه به این رابطه به ازای بسامدهای کم‌تر از ۷۰۰ هرتز، مقیاس مل به صورت خطی با بسامد تغییر می‌کند و در خارج از این محدوده رابطه موجود به صورت لگاریتمی در می‌آید. این رابطه بگونه‌ای ارائه شده است که با حساسیت غیرخطی سامانه شنوایی انسان نسبت به بسامدهای مختلف سازگار باشد. در انتها پس از محاسبه سهم هر یک از صافی‌های مل، ضرایب کپستروم مل با استفاده از تبدیل کسینوسی گسسته بر روی P_i مطابق زیر به دست می‌آیند.

$$coef_j=\sum_{i=1}^{M} P_i cos\left(j(i-\frac{1}{2})\frac{\pi}{M}\right), \qquad (4)$$

که در آن، j و M به ترتیب تعداد ضرایب مل و تعداد صافی‌های موجود در بانک صافی مل می‌باشند. پس از محاسبه این ضرایب به ازای همه قطعات ۱۰۰ میلی‌ثانیه‌ای، متوسط ضرایب در همه قطعات به عنوان بردار ویژگی نهایی استفاده می‌شود. با توجه به آنچه ذکر شد، نمودار بلوکی مرحله استخراج ویژگی در شکل ۲ نمایش داده شده است.

[1] Hamming windowing

[DOR: 20.1001.1.23455748.1394.3.1.7.7]



شکل ۲ نمودار بلوکی استخراج ویژگی ضرایب کپستروم بسامد مل.

۲-۳. طبقه‌بند بازنمایی تنک

تفکر اولیه بازنمایی تنک علامت (سیگنال) در نظریه حسگری فشرده[1] [۲۵] مطرح گردید. در این حوزه نحوه نمونه‌برداری از علامت با تعداد نمونه کم‌تر از نرخ نایکوئیست (کم‌تر از دو برابر بیشینه بسامد موجود در علامت) و با شرط بازسازی کامل بیان می‌شود. طبقه‌بند بازنمایی تنک (اِس‌آرسی)[2] یکی از روش‌های یادگیری بدون پارامتر است. این طبقه‌بندی نیازی به مرحله آموزش ندارد ولی در طراحی دیکشنری از داده‌های آموزش استفاده می‌کند. فرض کنید مجموعه دادگان آموزش و برچسب طبقه‌ها به صورت $\{(d_i, y_i)|d_i \in \vartheta \subset R^m, y_i \in \{1,2,\dots,c\}, i = 1,2,\dots,n\}$ نمایش داده شود. که در آن c تعداد طبقه‌ها، m بُعد فضای ورودی ϑ، n تعداد داده آموزش و y_i برچسب طبقه متناظر با داده d_i هستند. زیردیکشنری[3] D_j متعلق به طبقه j- ام به صورت $D_j = \left[d_{j1}, d_{j2}, \dots, d_{jn_j}\right]$ تعریف شده که در آن $j = 1,2,\dots,c$ و d_{ji}، i- امین داده طبقه j- ام و n_j تعداد داده‌های آموزشی موجود در این طبقه می‌باشند. دیکشنری نهایی با کنار هم قرار گرفتن زیردیکشنری طبقه‌های مختلف به‌صورت زیر ساخته می‌شود.

$$D = \left[D_1, D_2, \dots, D_c\right] \in R^{m\times n}, \quad (۵)$$

که در آن، $n = \sum_{j=1}^{c} n_j$. در طبقه‌بند بازنمایی تنک هر داده آزمون براساس ترکیب خطی داده‌های آموزش الگو می‌شود. به عبارت دیگر

$$y_t = D\alpha, \quad (۶)$$

که در آن، $\alpha \in R^n$ بردار ضرایب تنک و y_t داده‌های آزمون می‌باشند. در حالت آرمانی در صورتی که y_t به طبقه j تعلق داشته باشد، تنها درایه‌های متناظر با داده‌های آموزشی این طبقه در بردار ضرایب دارای مقدار غیر صفر هستند. بنابراین، بردار α ذاتاً تنک است؛ از این‌رو در طبقه‌بند بازنمایی تنک، بردار ضرایب توسط رابطه زیر تخمین زده می‌شود.

$$\alpha = arg\,min\,|\alpha|_0 \text{ subject to } y_t = D\alpha \quad (۷)$$

که در آن، $|\alpha|_0$ نُرم صفر (هنجار- تعداد درایه‌های غیر صفر) بردار ضرایب تنک می‌باشد. با توجه به غیرمحدب[4] بودن و پیچیدگی زیاد حل مسئله بهینه‌سازی بالا، در عمل از تقریب محدب آن (تبدیل نرم صفر به نرم یک) استفاده می‌شود. از طرفی با توجه به نوفه‌ای (نویزی) بودن داده‌های آموزش و آزمون، رابطه ۷ به صورت رابطه ۸ در می‌آید.

$$\alpha = arg\,min|\alpha|_1 \text{ subject to } \| y_t - D\alpha\|_2 \le \epsilon \quad (۸)$$

در این رابطه، ϵ بیشینه خطای قابل قبول در بازنمایی علامت آزمون را نشان می‌دهد. پس از تخمین بردار ضرایب تنک، طبقه‌ایی که دارای کمینه خطای بازسازی علامت آزمون باشد، به عنوان نتیجه طبقه‌بندی انتخاب می‌شود. به عبارت دیگر:

$$i = arg\,min_i\, E_i(y_t) = \|y_t - D_i\alpha_i\|_2, \quad (۹)$$

که در آن α_i و D_i به ترتیب بردار ضرایب تنک و زیردیکشنری متناظر با کلاس i- ام می‌باشند. در این مقاله از متوسط ضرایب کپستروم بسامد مل در قطعات مختلف علامت (سیگنال) یک ثانیه‌ای به عنوان اتم‌های دیکشنری استفاده می‌شود. اتم‌های موجود در دیکشنری همگی بهنجار (نرمالیزه) می‌شوند. این مسئله باعث می‌شود که طبقه‌بند بازنمایی تنک قادر نباشد بردارهای ویژگی متعلق به دو طبقه را در شرایطی که در یک راستا هستند، تفکیک نماید [۲۶]. بنابراین برای حل این مسئله در مقاله پیش‌رو استفاده از طبقه‌بند بازنمایی تنک مبتنی بر هسته (هسته- پایه) پیشنهاد می‌شود.

[1] CS; Compressive Sensing
[2] SRC; Sparse Representation Classification
[3] Sub-dictionary
[4] Nonconvex

[DOR: 20.1001.1.23455748.1394.3.1.7.7]

۲-۴. طبقه‌بند بازنمایی تنک مبتنی بر هسته (هسته-پایه)

با استفاده از روش‌های مبتنی بر هسته به سادگی می‌توان یک الگوریتم خطی را به معادل غیرخطی آن تعمیم داد. معمولاً توابع هسته $k(.,.)$ به صورت زیر تعریف می‌شوند:

$$k: \vartheta \times \vartheta \to R, \quad k(d,d') = \varphi(d)^T \varphi(d'), \tag{۱۰}$$

که در آن، d و d' دو نمونه از فضای ورودی ϑ و φ نگاشت غیرخطی مربوط به هسته $k(.,.)$ هستند. تعدادی از توابع هسته متداول عبارتند از هسته خطی[1]، هسته چندجمله‌ای[2] و هسته تابع پایه شعاعی[3]. به منظور طراحی طبقه‌بند بازنمایی تنک هسته- پایه، ابتدا داده‌ها با استفاده از نگاشت غیرخطی φ از فضای ورودی به فضای هسته به صورت زیر منتقل می‌شوند.

$$\varphi: d \in \vartheta \to \varphi(d) = \left[\varphi_1(d), \varphi_2(d), \dots, \varphi_D(d)\right]^T \in \Lambda, \tag{۱۱}$$

که در آن، $\varphi(d) \in R^D$ تصویر بردار d در Λ و $D \gg m$ بعد فضای هسته است. در اِس‌آرسی هر داده آزمون بر حسب ترکیب خطی داده‌های آموزش در فضای ورودی الگو می‌شود. به طور مشابه در کِی‌اِس‌آرسی تصویر هر داده آزمون در فضای هسته بر حسب تصویر داده‌های آموزش در Λ بیان می‌گردد. بنابراین، رابطه ۸ به صورت ۱۲ تغییر می‌کند.

$$\alpha = arg\,min\,|\alpha|_1 \quad \text{subject to} \| \varphi(y_t) - \Phi\alpha\|_2 \le \epsilon, \tag{۱۲}$$

که در آن α بردار ضرایب تنک در فضای هسته است. Φ ماتریس دیکشنری در Λ است و به صورت ۱۳ تعریف می‌گردد.

$$\Phi = \left[\varphi\left(d_1\right), \varphi\left(d_2\right), \dots, \varphi(d_n)\right] \in R^{D \times n}, \tag{۱۳}$$

با توجه به اینکه بعد فضای هسته D، بزرگ‌تر از بعد فضای ورودی است، پیچیدگی محاسباتی تعیین بردار تنک با استفاده از رابطه ۱۲ بیش‌تر از معادل آن در رابطه ۸ می‌باشد. از این رو معمولاً از روش‌های کاهش بعد در فضای هسته استفاده می‌گردد. می‌توان نشان داد در صورتی که ماتریس کاهش بعد (ماتریس شبه- ترادیسی[4]) به صورت $P = \Phi B$ و $B = \left[\beta_1, \beta_2, \dots, \beta_d\right] \subset R^{D \times d}$ انتخاب شود، رابطه ۱۲ به صورت رابطه ۱۴ تغییر می‌کند [۲۶]. لازم به ذکر است که در عمل بعد زیرفضای[5] کاهش یافته، d، به گونه‌ای انتخاب می‌شود که $d < n$ باشد.

$$\alpha = arg\,min\,|\alpha|_1 \text{ subject to } \left\|B^T k(\cdot, y_t) - B^T K\alpha\right\|_2 \le \epsilon, \tag{۱۴}$$

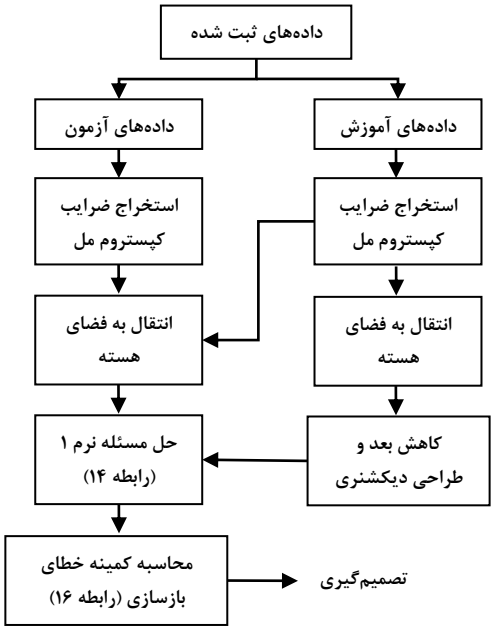
در این رابطه $K = \Phi^T \Phi$ است و $k(\cdot, y_t)$ به صورت زیر تعریف می‌شود.

$$k\left(\cdot, y_t\right) = \left[k\left(y_t, d_1\right), k\left(y_t, d_2\right), \dots, k(y_t, d_n)\right]^T, \tag{۱۵}$$

پس از محاسبه بردار ضرایب تنک در فضای هسته از تفکری مشابه اِس‌آرسی به منظور طبقه‌بندی استفاده می‌شود. به عبارت دیگر:

$$i = arg\,min_i\, E_i(y_t) = \left\|B^T k(\cdot, y_t) - B^T K\delta_i(\alpha)\right\|_2, \tag{۱۶}$$

که در آن، $\boldsymbol{\delta_i(\alpha)}$ عملگری است که تنها درایه‌هایی متناظر با طبقه i را از بردار $\boldsymbol{\alpha}$ حفظ و بقیه را صفر می‌نماید. با توجه به آن چه بیان شد، نمودار بلوکی نهایی طبقه‌بند بازنمایی تنک مبتنی بر هسته با استفاده از ضرایب کپستروم بسامد مل به صورت شکل ۳ پیشنهاد می‌شود.



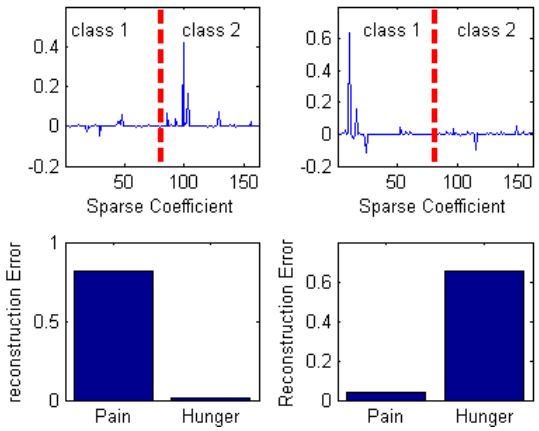
شکل ۳ نمودار بلوکی فرآیند بازشناختی گریه نوزاد با استفاده از روش پیشنهادی.

[1] Linear kernel
[2] Polynomial kernel
[3] RBF; Radial Basis Function kernel
[4] Pseudo-transformation matrix
[5] subspace

۳. نتایج

در این بخش به بررسی عملکرد الگوهای بازنمایی تنک در تفکیک دو نوع صدای گریه نوزاد (گریه ناشی از درد و گریه ناشی از گرسنگی) پرداخته می‌شود. ماشین بردار پشتیبان [۲۷-۲۸] و الگوی مخفی مارکوف [۲۹-۳۰] دو طبقه‌بند معروف در این حوزه می‌باشند؛ از این‌رو به منظور مقایسه و بررسی اثرگذاری نتایج حاصل از روش پیشنهادی علاوه بر تحلیل متمایزکننده خطی فیشر و طبقه‌بند بازنمایی تنک از ماشین بردار پشتیبان و الگوی مخفی مارکوف نیز استفاده شده است. در این قسمت ۲۴۲ علامت (سیگنال) یک ثانیه‌ای بهنجارشده (نرمالیزه شده) (نیمی از آن متعلق به طبقه گرسنگی و نیمی متعلق به طبقه درد) به منظور اعتبارسنجی الگوریتم‌ها بکاربرده شده‌اند. هر علامت به ۱۹ قطعه ۱۰۰ میلی‌ثانیه‌ای با هم‌پوشانی ۵۰ درصد تقسیم شده و متوسط ضرایب کپستروم مل هر قطعه به عنوان اتم‌های طبقه‌بند بازنمایی تنک و بردار ویژگی ورودی طبقه‌بندهای تحلیل متمایزکننده خطی و ماشین بردار پشتیبان استفاده می‌گردد. در ماشین بردار پشتیبان از هسته گوسی استفاده شده است. در حالت کلی به منظور محاسبه بردار ویژگی از ۲۱ صافی مثلثی به صورت یکنواخت در مقیاس مل با ۲۰ ضریب کپستروم استفاده گردیده است. نتایج موجود با متوسط‌گیری[1] بر روی ۲۰ تکرار با روش اعتبارسنجی K-لایه حاصل شده‌اند. در روش K- لایه مجموعه دادگان به صورت تصادفی به K قسمت مساوی تقسیم می‌شود. هر بار یک قسمت از دادگان به عنوان داده آزمون و K-۱ قسمت باقی‌مانده به عنوان داده آموزش استفاده می‌شود. به منظور کاهش بعد از روش تحلیل مولفه‌های اساسی استفاده گردیده و بعد فضای هسته به ۳۲ کاهش یافته است. شکل ۴ نمونه‌ای از نتیجه کِی‌اِس‌آرسی را بر روی دو داده از طبقه‌های درد و گرسنگی نمایش می‌دهد. منحنی بالا چپ بردار ضرایب تنک داده نمونه متعلق به طبقه گریه گرسنگی را نشان می‌دهد. در محاسبه این ضرایب از کرنل گوسی (تابع پایه شعاعی) استفاده شده است. در این منحنی مرز بین ضرایب تنک متناظر با دو طبقه توسط خط چین رنگ مشاهده می‌شود. ضرایب تنک سهم هر یک اتم‌های دیکشنری را در بازنمایی داده آزمون نشان می‌دهند. به عنوان مثال همان‌طور که مشاهده می‌شود در این مورد خاص درایه صدم بردار ضرایب تنک بیشترین دامنه را داشته از این‌رو اتم شماره صد از دیکشنری بیشترین سهم را در بازنمایی داده آزمون دارا می‌باشد. نکته قابل توجه این است که اکثر درایه‌های موجود در بردار ضرایب دارای مقدار صفر و یا نزدیک به صفر هستند به عبارت دیگر تنها تعداد معدودی از اتم‌های دیکشنری در بازنمایی داده‌ها مورد نظر شرکت یافته‌اند.

نمودار استوانه‌ای پایین چپ خطای بازسازی همان داده‌ها را به ازای هر دو طبقه نمایش می‌دهد. همان‌طور که قابل مشاهده است، خطای بازسازی طبقه گرسنگی بسیار کم‌تر از خطای بازسازی طبقه درد می‌باشد. در سمت راست شکل ۴ نیز بردار ضرایب تنک و خطای بازسازی یک داده‌ها نمونه متعلق به طبقه درد نمایش داده شده‌اند. روشن است که کِی‌اِس‌آرسی این دو داده‌ها نمونه را به درستی طبقه‌بندی می‌کند.



شکل ۴ نمونه‌ای از اندازه ضرایب بردار تنک در طبقه‌بندی با استفاده از کِی‌اِس‌آرسی برای دو داده آزمون (چپ) طبقه ۱ و (راست) طبقه ۲. خط‌چین مرز بین ضرایب تنک دو طبقه را نشان می‌دهد.

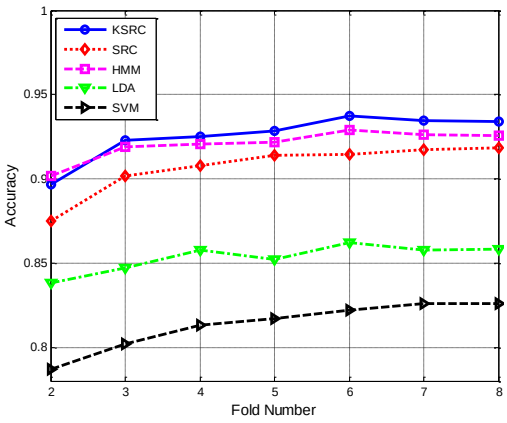
جدول ۱ مقایسه عملکرد توابع هسته مختلف در دقت کِی‌اِس‌آرسی (میانگین ± انحراف معیار).

نوع هسته	دقت طبقه‌بندی (٪)
چندجمله‌ای	۹۰٫۵±۱٫۳
گوسی	۹۱٫۴±۱٫۴
خطی	۹۰٫۷±۱٫۳

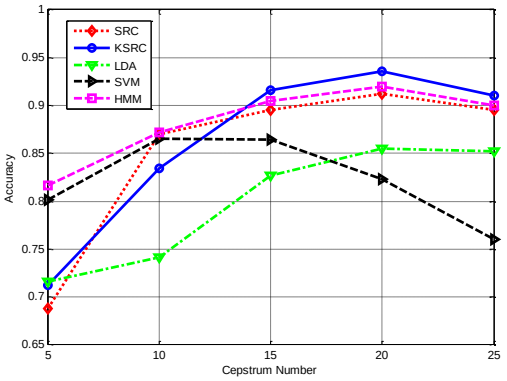
به منظور بررسی تاثیر انتخاب نوع تابع هسته، در جدول ۱ نتایج طبقه‌بند کِی‌اِس‌آرسی به ازای سه نوع تابع هسته ارائه

[1] Averaging

شده‌اند. این نتایج به ازای K=۳، در اعتبارسنجی K- لایه به دست آمده‌اند. همان‌طور که مشاهده می‌شود، تابع هسته گوسی کمی بهتر از دو تابع هسته خطی و چندجمله‌ای عمل می‌کند. بنابراین در ادامه به منظور انتقال داده‌ها از فضای ورودی به فضای هسته از تابع گوسی استفاده می‌شود. در شکل ۵ عملکرد روش پیشنهادی با طبقه‌بندهای اِل‌دی‌اِی و اِس‌آرسی، اِس‌وی‌اِم و اِچ‌اِم‌اِم[1] به ازای تعداد داده‌های آموزش مختلف مقایسه شده است. محور افقی مقدار K در روش K- لایه و محور عمودی دقت طبقه‌بندی را به ازای ۲۰ ضریب کپستروم نشان می‌دهد. همان‌طور که مشخص است با افزایش تعداد داده‌های آموزش (افزایش K) به طور کلی دقت طبقه‌بندها بهبود کمی داشته و در این بین طبقه‌بند بازنمایی تنک مبتنی بر هسته (هسته- پایه) عملکرد بهتری نسبت به سایر طبقه‌بندها دیگر دارد؛ به طوری که به ازای ۶- لایه دقت به ۹۳/۵ درصد می‌رسد.



شکل ۵ مقایسه دقت طبقه‌بندی روش‌های مختلف به ازای تعداد داده‌های آموزش متفاوت.



شکل ۶ مقایسه دقت طبقه‌بندی روش‌های مختلف به ازای تعداد ضرایب کپستروم متفاوت.

به منظور بررسی تأثیر تعداد ضرایب کپستروم بسامد مل در عملکرد طبقه‌بندها، شکل ۶ ارائه شده است. در این شکل دقت طبقه‌بندها به ازای ابعاد مختلف ورودی (تعداد ضرایب کپستروم به ازای هر قطعه علامت) مشاهده می‌شود. نتایج نشان می‌دهند که به ازای انتخاب ۲۰ ضریب کپستروم همه طبقه‌بندها به غیر از ماشین بردار پشتیبان بهترین عملکرد خود را دارند. در این شرایط روش‌های کِی‌اِس‌آرسی، اِس‌آرسی، اِل‌دی‌اِی، اِس‌وی‌اِم و اِچ‌اِم‌اِم قادرند داده‌های دو طبقه را به ترتیب با دقت‌های ۹۳/۵، ۹۱/۲، ۸۵/۵، ۸۲/۳ و ۹۲/۲ درصد از یکدیگر تفکیک کنند. همان‌طور که مشاهده می‌شود روش پیشنهادی عملکرد قابل قبولی در تفکیک دو نوع صدای گریه مربوط به درد و گرسنگی دارد و به طور کلی بهتر از سایر طبقه‌بندها عمل می‌کند.

۴. بحث

در این مطالعه، نظریه الگوهای تنک به منظور طبقه‌بندی صدای گریه نوزاد با دو منشاء درد و گرسنگی مورد بررسی قرار گرفت. به دلیل این که فضای داده‌های ثبت‌شده از گریه نوزاد، اطلاعات مناسبی به منظور تفکیک دو نوع صدای گریه (ناشی از درد و گرسنگی) در اختیار قرار نمی‌دهد، از این‌رو الگوی بازنمایی تنک در فضای بردار ویژگی مناسب پیشنهاد داده شد.

به طور کلی، ویژگی‌های مرتبط با طیف بسامدی علامت (سیگنال) اطلاعات مناسبی به منظور تفکیک و بازشناختی گفتار و صدای گریه در اختیار قرار می‌دهند. از این‌رو ضرایب کپستروم و ضرایب پیش‌بینی خطی در بسیاری از مطالعات موجود در این حوزه استفاده می‌شوند [۱، ۱۳]. در مقاله جاری ضرایب کپستروم به منظور طراحی دیکشنری موجود در الگوهای تنک بکار برده شدند. دلیل این انتخاب از دو منظر بیان می‌شود. اول اینکه ضرایب کپستروم سازگاری بسیار خوبی با حساسیت غیرخطی سامانه شنوایی انسان در بسامدهای مختلف دارند. به عبارت دیگر این ضرایب در بسامدهای پایین اطلاعات بسامدی را با تفکیک‌پذیری (رزولوشن) بیشتری نسبت به بسامدهای بالا در اختیار قرار می‌دهند. دوم اینکه یکی از پارامترهای مهم در انتخاب دیکشنری الگوی تنک، هم‌دوسی متقابل[2] بین زیردیکشنری

[1] HMM; hidden markov model

[2] MC; Mutual Coherence

طبقه‌بندی‌های مختلف داده‌ها است [۳۱]. این پارامتر به صورت زیر تعریف می‌شود.

$$\mu\left(D_1, D_2\right) = \max\left\{ \left|\langle d_{1i}, d_{2j}\rangle\right| : i = 1,\dots,n_1;\ j = 1,\dots,n_2 \right\}, \quad (17)$$

که در آن D_1، D_2، d_{1i} و d_{2j} به ترتیب زیردیکشنری‌های کلاس یک، دو، i- امین و j- امین اتم از زیردیکشنری‌های کلاس یک و دو هستند. مقدار کم این پارامتر نشان‌دهنده ناهمدوس[1] بودن دو زیردیکشنری است. به عبارت دیگر هر چه مقدار همدوسی متقابل بین دو زیردیکشنری انتخاب شده کمتر باشد، طبقه‌بندی در الگوی تنک با دقت بالاتری انجام خواهد شد. مقایسه هم‌دوسی متقابل دیکشنری ساخته شده توسط ضرایب کپستروم و ضرایب پیش‌بینی خطی نشان می‌دهد که زیردیکشنری‌های حاصل از ضرایب کپستروم همدوسی کم‌تری نسبت به زیردیکشنری‌های ضرایب پیش‌بینی خطی دارند. بنابراین در طراحی دیکشنری در کاربرد جاری از ضرایب کپستروم استفاده شده است.

همان‌طور که در شکل ۵ مشاهده می‌شود، به ازای ۲۰ ضریب کپستروم و تعداد متفاوت داده‌های آموزش، طبقه‌بندی با استفاده از الگوی بازنمایی تنک مبتنی بر هسته (هسته- پایه) نسبت به چهار طبقه‌بند دیگر، عملکرد بهتری دارد. انتقال اتم‌های دیکشنری به فضای هسته در روش کِی‌اِس‌آرسی دلیل اصلی این برتری است. به عبارت دیگر هر یک از اتم‌های دیکشنری در فضای هسته حاوی اطلاعاتی از شباهت اتم مورد نظر با سایر اتم‌های موجود در الگوی تنک است. در شکل ۶ حساسیت طبقه‌بندها به تعداد ضرایب کپستروم نمایش داده شده است. نکته قابل توجه این است که به ازای تعداد کم ضرایب کپستروم (۵ ضریب در مسئله جاری) عملکرد ماشین بردار پشتیبان و الگوی مخفی مارکوف بهتر از سایر طبقه‌بندها بوده ولی با افزایش تعداد ضرایب (۲۰ ضریب) روش کِی‌اِس‌آرسی به عملکرد بهینه خود نزدیک شده و در این شرایط بهتر از اِس‌ویی‌اِم و اِچ‌اِم‌اِچ عمل می‌کند. لازم به ذکر است که عملکرد طبقه‌بند بازنمایی تنک مبتنی بر هسته (هسته- پایه) با اعمال تمهیداتی قابل ارتقا می‌باشد. به عنوان مثال استفاده از روش‌های کاهش بعد با نظارت به جای روش تحلیل مولفه‌های اساسی در ساختار این طبقه‌بند می‌تواند به بهبود نتایج این الگو منجر شود.

۵. نتیجه‌گیری

در این مقاله به بررسی امکان استفاده از الگوهای بازنمایی تنک در طبقه‌بندی دو نوع صدای گریه نوزاد پرداخته شد. بدین منظور چارچوب مورد نیاز برای استفاده از این الگوها در طبقه‌بندی صدای گریه نوزاد ارائه شد. نتایج نشان دادند که طبقه‌بند بازنمایی تنک مبتنی بر هسته (هسته- پایه)، عملکرد بهتری نسبت به سایر طبقه‌بندهای ذکر شده در تفکیک صدای گریه نوزاد با دو منشاء گرسنگی و درد دارد. دلیل این امر این است که در تشکیل هر یک از اتم‌های دیکشنری در فضای هسته از اطلاعات همه داده‌های آموزش استفاده می‌گردد؛ این در حالی است که در سایر روش‌ها، بردار ویژگی و اتم‌های دیکشنری تنها از داده‌های مربوط به یک داده آموزش استخراج می‌شوند.

انتخاب مناسب دیکشنری در عملکرد الگوهای بازنمایی تنک بسیار تأثیرگذار است. از این‌رو استفاده از الگوریتم‌های یادگیری دیکشنری به منظور انطباق دیکشنری و مجموعه داده‌ها می‌تواند میزان قدرت تفکیک داده‌های دو طبقه را افزایش دهد. این مسئله در کارهای آتی بررسی می‌گردد.

۶. فهرست منابع

[1] A.R. Perez, C.A. Reyes-García, J.A. Gonzalez, O.F. Reyes-Galaviz, H.J. Escalante, S. Orlandi, "Classifying infant cry patterns by the genetic selection of a fuzzy model," Biomedical Signal Processing and Control, vol. 17, pp. 38-46, 2015.

[2] G. Varallyay, Z. Benyo, A. Illenyi, Z. Farkas, L. Kovacs, "Acoustic analysis of the infant cry: classical and new methods," International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1, pp. 313–316, 2004.

[3] S.J. Sheinkopf, J.M. Iverson, M.L. Rinaldi, B.M. Lester, "Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder," Autism Research, vol. 5, pp. 331–339, 2012.

[4] S.C. Ortiz, D.E. Beceiro, T. Ekkel, "A radial basis function network oriented for infant cry classification," Progress in Pattern Recognition, pp. 15–36, 2004.

[5] M. Hariharana, J. Saraswathy, R. Sindhu, W. Khairunizam, S. Yaacob, "Infant cry classification to identify asphyxia using time-

[1] Incoherent

Learning Systems, IEEE Transactions on , vol. 24, no. 7, 2013.

[17] M. Elad, M.A.T. Figueiredo, Yi Ma, "On the role of sparse and redundant representations in image processing," Proceedings of the IEEE, vol. 98, pp. 972–982, 2010.

[18] H. Zhang, Y. Zhang, T.S. Huang, "Simultaneous discriminative projection and dictionary learning for sparse representation based classification," Pattern Recognition, pp. 346–354, 2013.

[19] M. Jian, C. Jung, "Class-discriminative kernel sparse representation-based classification using multi-objective optimization," Signal Processing, IEEE Transactions on, vol. 61, no. 18, 2013.

[20] M. Yang, L. Zhang, X. Feng, D. Zhang, "Sparse representation based fisher discrimination dictionary learning for image classification," Springer, pp. 209–232, 2014.

[21] J. Liu, Z. Wu, Z. Wei, L. Xiao, L. Sun, "Spatial-spectral kernel sparse representation for hyperspectral image classification," in Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of , vol. 6, no. 6, pp. 2462-2471, Dec. 2013.

[22] M. Jian, C. Jung, "Class-discriminative kernel sparse representation-based classification using multi-objective optimization," in Signal Processing, IEEE Transactions on, vol. 61, no. 18, pp. 4416-4427, Sept.15, 2013.

[23] M. Salarian, "A system for the processing of infant cry to evalute its causes," in Biomedical Engineering. MSc thesis, Amirkabir University of Technology, 2011.

[24] M. Cutajar, E. Gatt, I. Grech, O. Casha, J. Micallef, "Comparative study of automatic speech recognition techniques," IET Signal Process, pp. 25–46, 2013.

[25] E. Candes, M. Wakin, "An introduction to compressed sampling," IEEE Signal Process Magazine, vol. 25, no. 2, pp. 21–30, 2008.

[26] L. Zhang, W.D Zhou, P.C. Chang, J. Liu, Z. Yan, T. Wang, F.Z. Li, "Kernel sparse representation-based classifier," Signal Processing, IEEE Transactions on, vol. 60, no. 4, 2012

[27] R. Sahak, W. Mansor, Y. Lee, A. Yassin, A. Zabidi, "Performance of combined support vector machine and principal component analysis in recognizing infant cry with asphyxia," Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp. 6292–6295, 2010.

[28] S.B. Montiel, C. Garcia, E. Tirado, M. Mandujano, "Improving baby caring with automatic infant cry recognition," Springer, vol. 4061 of Lecture Notes in Computer Science, pp. 691–698, 2006.

frequency analysis and radial basis neural networks," Expert Systems with Applications, vol. 39, pp. 9515–9523, 2012.

[6] S.E. Montiel, C.A. Garcia, "Fuzzy support vector machines for automatic infant cry recognition," Intelligent Computing in Signal Processing and Pattern Recognition, vol. 345, pp. 876-881, 2006.

[7] J.O. Garcia, C.A. Garcia, "Applying scaled conjugate gradient for the classification of infant cry with neural networks," The European Symposium on Artificial Neural Networks, 2003.

[8] S.E. Montiel, C.A. Garcia, "Identifying pain and hunger in infant cry with classifiers ensembles," International Conference on Computational Intelligence for Modelling, pp. 770-775, 2005.

[9] S.F. Molaeezadeh, M. Salarian, M.H. moradi, "Type-2 fuzzy pattern matching for classifying hunger and pain cries of healthy full-term infants," International Symposium on Artificial Intelligence and Signal Processing, pp. 233-237, 2012.

[10] M. Petroni, A.S. Malowany, C.C. Johnston, B.J. Stevens, "A comparison of neural network architectures for the classification of three types of infant cry vocalizations," IEEE 17th Annual Conference Engineering in Medicine and Biology Society, vol. 1, pp. 821-822, 1995.

[11] O.W. Hockert, T. Partanen, V. Vuorenkoski, E. Valanne, K. Michelsson, "The identification of some specific meanings in infant vocalization," Experientia, vol. 20, pp. 154-156, 1964.

[12] M. Hariharan, S. Yaacob, S.A. Awang, "Pathological infant cry analysis using wavelet packet transform and probabilistic neural network," Expert Systems with Applications, pp. 15377–15382, 2011.

[13] O.F.R. Galaviz, S.C. Ortiz, C.R. Garca, "Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies," 8th Mexican International Conference on Artificial Intelligence, pp. 330-335, 2009.

[14] J.S. Turek, I. Yavneh, M. Elad, "On MMSE and MAP denoising under sparse representation modeling over a unitary dictionary," Signal Processing, IEEE Transactions on, vol. 59, no. 8, pp. 3526-3535, 2011.

[15] A. Amini, M. Unser, F. Marvasti, "Compressibility of deterministic and random infinite sequences," Signal Processing, IEEE Transactions on, vol. 59, no. 11, pp. 5193-5201, 2011.

[16] J. Yang, D. Chu, L. Zhang, Y. Xu, J. Yang, "Sparse representation classifier steered discriminative projection with applications to face recognition," Neural Networks and

[29] D. Lederman, A. Cohen, E. Zmora, K. Wermke, S. Hauschildt, A. Stellzig, "On the use of hidden markov models in infants cry classification," The 22nd Convention of Electrical and Electronics Engineers in Israel, IEEE, pp. 350-352, 2002.

[30] D. Lederman, E. Zmora, S. Hauschildt, A. Stellzig, K. Wermke, "Classification of cries of infants with cleft-palate using parallel hidden Markov models," pp. 965–975, 2008.

[31] Y. Li, Z. Liang, N. Bi, Y. Xu, Z. Gu, "Sparse representation for brain signal processing: A tutorial on methods and applications," in Signal Processing Magazine, IEEE, vol. 31, no. 3, pp. 96-106, May 2014.

Journal of Acoustical Engineering Society of Iran, vol. 3, no. 1, 2015

Analysis of infants cry sound using kernel sparse representation-based classifier

A. Akhavan, M.H. Moradi*

Bioelectrical Engineering Department, Faculty of Biomedical Engineering, Amirkabir University

Abstract

Processing of infant cry sound provides useful information about his/her condition. This information can be used to establish a diagnostic method to determine the infant's needs. This paper addresses the analysis of newborn babies cry sound in order to discriminate crying associated with hunger from that originating from pain. Sparse representation models are one of the state of the art processing tools in pattern recognition and machine learning. In this work a novel framework is proposed in order to deal with sparsity-based approach in a classification task. The dictionary atoms of the sparse model are designed using Mel Frequency Cepstrum Coefficient in kernel space. Performance assessment of kernel sparse representation model shows the discriminative power of this model in classifying different types of infant cry sound. In order to compare, the results of conventional sparse representation model and some other well-known classifiers (Hidden Markov Model and Support Vector Machine) are also presented. The results show that the proposed model has better performance in comparison with the other presented classifiers. Using 6-fold cross validation the kernel sparse model can distinguish two types of infant cry with more than 93% accuracy. The pain and hunger databases are recorded from 51 (19 male and 32 female) 2-3 day old healthy infants.

Keywords: Sparse Representation Model, Sparse Representation Classifier, Kernel-based model, Mel Frequency Cepstrum Coefficient, Infant cry sound.

pp. 56-65 (In Persian)

[DOR: 20.1001.1.23455748.1394.3.1.7.7]

* Corresponding author E-mail: mhmoradi@aut.ac.ir