ارائه روشی جدید در بهینه سازی سیستم های پردازش گفتار با استفاده ازالگوریتم ژنتیک

علی اکبر برنگی، ایمان اسمعیل زاده و هومن نبوتی

مرکز تحقیقات سجاد

aliakbar_berangi@yahoo.com

imanesmaailzadeh@yahoo.com

hoomannabovati@gmail.com

چکیده – روش *DTW*[1] یکی از روش های متداول در باز شناسی گفتار گسسته [2] می با شد . یکی از مزایای این روش مرتفع کردن مشکل عدم همزمانی سیگنال هادر قسمت مقایسه نمونه ها می باشد .از مشکلا ت روش مذکور حجم محاسباتیٔ نسبتا بالای آن می یاشد. بکار گیری ویژگی های بیشتر جهت مقایسه سیگنال ها یاعث افرایش بازدهی الگوریتم می گردد، ولی به همان نسبت میزان محاسبات را افزایش می دهد. ضمنا همه ویژگی ها ارزش یکسانی ندارند لذا ارزش دهی به ویژگی ها ضروری می نماید.این مقاله به بهینه سازی یک سیستم پردازش گفتار با بهره گیری از الگوریتم ژنتیک اختصاص یافته است. البته روش ارائه شده برای تمامی شیوه هایی که با استخراج ویژگی[3] در پردازش سیگنال سر و کار دارند قایل پیاده ساری می باشد.

کلید واژه- DTW,Genetic Algorithm,Feature Vector,Isolated Word Recognition

۱- مقدمه

روش DTW بر مبنای تطبیق کلمه بیان شده با الگو های از پیش ذخیره شده عمل می نماید. در هر بار مقایسه، کلمه ادا شده با همه الگو ها تطبیق داده می شود و به کمک یک تابع هزینه[4] میزان تطابق کلمه با الگو مقایسه می شود و الگویی که کمترین هزینه را ایجاد کند به عنوان الگوی بازشناسی شده معرفی می گردد.[1]

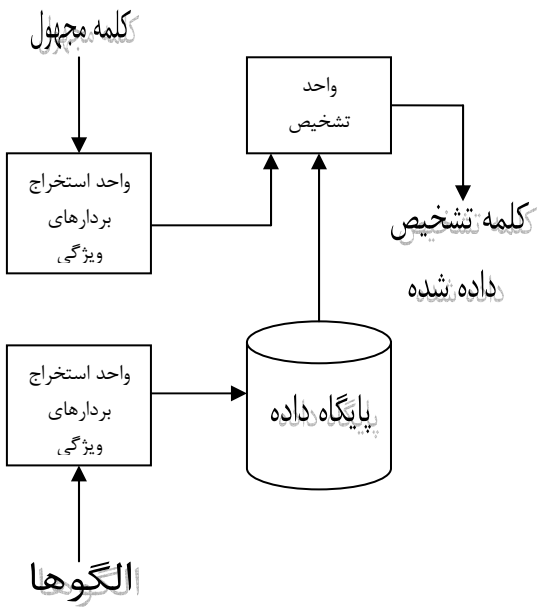
در این مقاله برای وزن دهی به ویژگی ها از الگوریتم ژنتیک استفاده شده است. الگوریتم ژنتیک روشی جهت یافتن نقاط اکسترمم توابع می باشد. در این شیوه نقاط بهینه به کمک چند عملگر در طول چند نسل بدست می آیند.

یکی از مزایای این روش این است که بهینه سازی مستقل از ضابطه تابع صورت می گیرد. از آنجا که ما قادر به ارائه تابعی مشخص برای یافتن بهینه ها نیستیم استفاده از روش مذکور می تواند راهکار مناسبی برای حل این مساله ارائه دهد.

در ادامه مقاله به طور خلاصه به بررسی یک سیستم بازشناسی گفتار پرداخته خواهد شد. و در انتها روشی جدید برای بهینه سازی سیستم معرفی خواهد شد.

۲- مبانی روش DTW

به طور کلی یک سیستم تشخیص کلمه توسط روش DTW را می توان توسط بلوک دیاگرام شکل ۱ نشان داد.



شکل ۱:سیستم تشخیص یک کلمه

در این سیستم ابتدا یک مجموعه ثابت از تعدادی کلمات مورد نظر انتخاب شده و سپس عملیات استخراج بردار های ویژگی بر روی این کلمات صورت می پذیرد و در پایگاه داده ذخیره می شود.

در زمان تشخیص، مرحله استخراج بردارهای ویژگی بر روی کلمه مجهول انجام گرفته و سپس در واحد تشخیص با بردار های ویژگی کلمات معلوم، مقایسه گشته و شبیه ترین کلمه به عنوان کلمه مورد نظر معرفی می شود. در سیستم مذکور برای رسیدن به زمان بلا درنگ[5] لازم است که سرعت محاسبات مراحل "استخراج بردارهای ویژگی" و"تشخیص" را افزایش دهیم. سرعت محاسبات در عملیات تعلیم دراین سیستم اهمیتی ندارد زیرا این عمل یک بار و به صورت آفلاین انجام می پذ یرد.[6]

۲-۱- مراحل تشخیص گفتار در روش DTW

در این قسمت دو واحد استخراج بردارهای ویژگی وتشخیص، بر مبنای روش DTW توضیح داده خواهد شد.

۲-۱-۱ – واحد استخراج بردارهای ویژگی

برای حذف نویز اضافی از تابع انرژی سیگنال استفاده می شود در این مقاله از یک پنجره کوچک در حدود ۵ میلی ثانیه استفاده شده که از ابتدای سیگنال تا نقطه ای که انرژی از یک میزان تعیین شده بیشترشود پیشروی می کند این کار به طور مشابه برای انتهای سیگنال نبز انجام می شود. در پایان، محل هایی که پنجره از آنها عبور کرده حذف می گردد.

سیگنال صحبت قبل از ورود به مرحله تشخیص باید تحت پردازش های اولیه برای استخراج ویژگی های مهم آن قرار گیرد. این ویژگی ها باید حداقل حساسیت را به اداهای مختلف کلام وبیشترین وابستگی را به خود کلام داشته باشند. البته قبل از استخراج ویژگی ها ابتدا باید قسمت های بی کلام سیگنال را حذف کرده و خروجی را به قابهای کوچک در حدود ۵ تا ۴۰ میلی ثانیه تقسیم نمود.

بهتر است به خاطر اینکه در ویژگی های استخراجی گسستگی وجود نداشته باشد این قابها همپوشانی[6] داشته باشند. در این مقاله هر قاب ۷٫۸ میلی ثانیه با همپوشانی ۵۰٪ به صورت ۲۵٪ در هر طرف و در فرکانس نمونه برداری ۸ کیلو هرتز در نظر گرفته شده است. نمونه ها ۱۶ بیتی هستند و هر قاب شامل ۶۲٫۸ نمونه می باشد.[6]

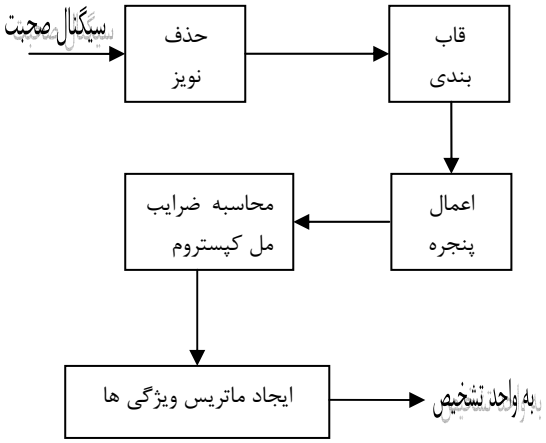
برای اینکه سیستم عملکرد مطلوب تری داشته باشد قبل از استخراج ویژگی به سیگنال پنجره اعمال شده است. در این مقاله از "پنجره همینگ" استفاده شده است که رابطه آن در زیر آمده است :

$$\omega(\kappa+1)=0.54-0.46\cos\left(2\pi\frac{\kappa}{n-1}\right)\qquad \kappa=0,\ldots,n-1 \qquad (۱)$$

خروجی این قسمت به واحد استخراج ویژگی داده می شود که به ازای هر قاب یک بردار ۱۲ تایی از ضرایب مل کپستروم[7] تولید می کند. لازم به ذکر است که می توان از ویژگی های دیگر نظیر ضرایب LPC[8]، انرژی و... نیز استفاده کرد که بهبود عملکرد سیستم را به بهای پایین آمدن سرعت به دنبال دارد. در واقع موضوع این مقاله استفاده از ویژگی کمترو تولید قدرت تشخیص بهتر برای کلمات می باشد که این کار به وسیله الگوریتم ژنتیک انجام می شود.

در نهایت این بردار ها در یک ماتریس قرار گرفته و به واحد تشخیص سپرده می شود. بلوک دیاگرام عملیات فوق در

شکل ۲ مشاهده می شود.



شکل ۲:سیستم استخراج ویژگی ها

۲-۱-۲ –واحد تشخیص

در این مقاله برای تشخیص کلمات به کمک ویژگی ها از الگوریتم DTW استفاده شده است.

در این روش بردار های ویژگی مربوط به هر قاب، در یکی ازستون های ماتریس مقایسه قرار می گیرد.تعدادستونهای ماتریس متناسب با زمان ادای کلمه می باشد اما تعداد سطرها بستگی به تعداد قابهای کلمات الگو دارد.

در روش DTW دو نوع فاصله محاسبه می شود فاصله محلی[۹] و فاصله سراسری[۱۰].

فاصله محلی فاصله محاسبه شده بین بردار ویژگی یک سیگنال و بردار ویژگی سیگنال دیگر است. با فرض این که X بردار ویژگی های قابی از سیگنال اول (الگو) و Y بردار ویژگی های قابی از سیگنال دوم (کلمه مجهول) باشد و هر یک از بردار ها دارای N عضو باشند و همچنین شماره سطرها با i و شماره ستون ها با j مشخص شوند فاصله محلی طبق رابطه زیر محاسبه می گردد.

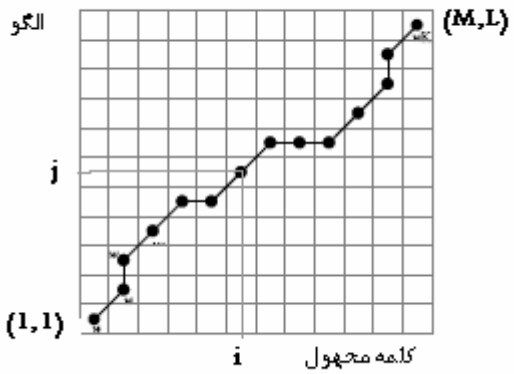
$$d(X,Y)=\sqrt{\sum_{n=1}^{N}\left(X_n^{i}\,\hat{}\,2-Y_n^{j}\,\hat{}\,2\right)} \qquad (۲)$$

تابع فوق برای تمام i ها و j ها محاسبه می شود و در عنصر متناظر با خود در ماتریس قرار می گیرد.سپس از خانه اول ماتریس ([۱و۱]) شروع کرده و از کم هزینه ترین مسیر به خانه ([M,L]) می رسد که در آن LوM به ترتیب تعداد عناصر سطر ها و ستون های ماتریس مذکور می باشد.

سپس به صورتی که در رابطه زیر مشاهده می شود فاصله سراسری تا خانه (i,j)، که برابر حاصل جمع فاصله های محلی طی شده می باشد محاسبه می گردد . .[2].

$$D(i,j)=\min\begin{Bmatrix}D(i-1,j-1)+D(i,j-1)\\+D(i-1,j)\end{Bmatrix}+d(i,j) \qquad (۳)$$

در مقایسه کلمه مجهول با هر یک از الگو ها این فاصله محاسبه شده و الگویی که کمترین هزینه (فاصله سراسری) را ایجاد کند به عنوان کلمه بازشناسی شده معرفی می گردد. در شکل ۳ نمای کلی این الگوریتم مشاهده می شود.



شکل ۳: شمای کلی مقایسه کلمه با الگو بوسیله ماتریس DTW

۳- بهینه سازی سیستم تشخیص

در این قسمت از مقاله به مساله بهینه سازی سیستم بیان شده در قسمت قبل پرداخته شده است.

این بهینه سازی به کمک الگوریتم ژنتیک صورت گرفته

است. ابتدا باید مغیاری برای سنجش مبزان عملکرد صحیح سیستم تعریف شود. همان گونه که پیشتر گفته شد کلمه مجهول باید با تمام الگوها مقایسه شود و الگویی که کم تربن هزینه در تطابق با کلمه را ایجاد کند به عنوان کلمه بازشناسی شده معرفی می شود. هر چه هزینه کلمه بازشناسی شده در مقایسه با هزینه تطابق سایر الگوها کمتر باشد (الیته اگر کلمه را درست تشخیص داده باشیم) شناسایی بهتر صورت گرفته است.

پیش از این بیان شد که ویژگی های یک قاب ارزش یکسانی ندارند و برای ارزش دهی به این ویژگی ها می توان برای هر ویژگی ضریبی در نظر گرفت. این ضرایب به این صورتند که ویژگی دارای اهمیت بالاتر ضریب بزرگتری دارد.

حال مساله به این شکل مطرح می شود که ضرایب چگونه انتخاب شوند تا معیار عملکرد بهینه شود. در این مقاله به کمک الگوریتم ژنتیک و یک تابع که برای نمونه های صدای یک شخص ، معیار ذکر شده را محاسبه می کند سعی شده است در نهایت ضرایبی بدست آیند که در صورت اعمال به بردار ویژگی ها معیار مقایسه بهترین برازندگی راداشته باشد.

در ادامه به طور مختصر در مورد الگوریتم بهینه سازی استفاده شده در مقاله توضیح داده شده است. ابتدا الگوریتم ژنتیک به انضمام تنظیمات خاصی که برای این کاربرد لازم است بررسی می شود.

۳-۱-کاربرد الگوریتم ژنتیک در واحد تشخیص

بهینه سازی در این مقاله به کمک الگوریتم ژنتیک صورت پذیرفته است. روش کار این گونه می باشد که ابتدا تابعی جهت بهینه سازی بوسیله الگوریتم تشکیل شده سپس پارامتر های مربوط به الگوریتم نظیر نوع عملگر انتخاب[11]، احتمال و نوع عملگر برش(پیوند)[12]، عملگر جهش[13]، نسل اولیه [14]و... مشخص می گردد. هر یک از این پارامترها می توانند در نتیجه الگوریتم نقشی تعیین کننده داشته باشند لذا به صورت مختصر به توضیح هر یک پرداخته می شود.

۳-۱-۱-عملگر انتخاب

این عملگر با اعمال نوعی فشار سعی می کند اعضای قویتر جامعه (اعضایی که در یک نسل برازندگی بیشتری دارند) را انتخاب کند تا بتوانند تولید مثلی با برازندگی بالاتر داشته باشند. در این مقاله از عملگر"stochastic uniform" استفاده شده است.

در نهایت همه اعضای جامعه به صورت رشته های باینری در آمده و به عملگر برش سپرده می شوند.[3]

۳-۱-۲- عملگر برش

به کمک این عملگر عمل تولید مثل صورت می گیرد. این عملگر دو رشته را با یک احتمال مشخص از نقطه ای تصادفی بریده و سپس دو رشته دم های خود را با یکدیگر معاوضه می کنند. اگر برش رخ ندهد خود رشته ها به نسل بعد منتقل می شوند. برش ممکن است از یک یا چند نقطه صورت پذیرد. در این مقاله از برش یک نقطه ای استفاده شده است.[3,4]

۳-۱-۳- عملگر جهش

این عملگر به این دلیل به کار گرفته می شود که تنوع ژنتیکی نسل ها حفظ شده تا الگوریتم به سمت مینیمم های محلی سرازیر نشود.

این عملگر با یک احتمال کوچک روی هر بیت اعضای نسل تاثیر گذاشته و آن بیت را تغییر وضعیت می دهد.[3,4]

۳-۱-۴-جمعیت اولیه

از مسائل مهم در بهینه سازی به کمک الگوریتم ژنتیک مساله جمعیت اولیه می باشد. اگر جمعیت اولیه بزرگ باشد عملکرد بهتر است اما بهای آن از دست دادن سرعت همگرایی الگوریتم می باشد در این مقاله نسل اولیه تنها ۲۰ عضو دارد چون به علت حجم بالای محاسبات(زیرا برای هر بار انجام الگوریتم بارها الگوریتم DTW اجرا می شود.) سرعت پایین است. اما همان گونه که در قسمت نتایج مشاهده خواهد شد همین تعداد عضو نیز کارایی بالایی در بهبود عملکرد سیستم نشان می دهند. [5]

۳-۲- نحوه اعمال الگوریتم ژنتیک به سیستم

در این مقاله برای بهینه سازی سیستم تشخیص از جعبه ابزار الگوریتم ژنتیک MATLAB استفاده شده است. به کمک این جعبه ابزار تابع بحث شده(تابعی که شامل معیار عملکرد است) به عنوان ورودی تعیین می گردد سپس تعداد ورودی ها برابر ۱۲ عدد مشخص می شود زیرا بردار ویژگی که باید وزن دار شود ۱۲ عضو به ازای هر قاب دارد.سپس پارامتر هایی که در قسمت های قبل بحث شد تنظیم گشته تا با توجه به خواسته کابر تابع بهینه گردد.

تنها مساله باقی مانده شرط پایان الگوریتم است که چون حجم محاسبات بالاست بهتر است که زمان کوتاه برای الگوریتم در نظر گرفته نشود و معیار پایان از روی تعداد نسل های سپری شده مشخص گردد. در این مقاله تعداد نسل ها سپری شده تنها۱۱ نسل بوده است اما همانطور که در قسمت نتایج مشاهده می گردد طی شدن همین تعداد نسل هم بهبود محسوسی بر عملکرد سیستم خواهد داشت.

۴-نتیجه‌گیری

در این مقاله یک سیستم تشخیص وابسته به گوینده جهت بهینه سازی انتخاب شده است. سیستم مذکور به بازشناسی کلمات یک تا چهار می پردازد، سپس الگوریتم ژنتیک برای بهینه سازی این سیستم بکار گرفته شد.

در این روش ابتدا در پایگاه داده چهار واژه نمونه که شامل ادای کلمات یک تا چهار بودند ذخیره شد سپس برای مقایسه عملکرد سیستم سه گروه از کلمات به سیستم اعمال شد(هر گروه شامل کلمات یک تا چهار) هر کدام از این کلمات جداگانه وارد الگوریتم تشخیص شده است، این کار در ابتدا بدون اعمال هیچ گونه ضریبی(برای بردار ویژگی ها) صورت گرفته و سپس به کمک الگوریتم ژنتیک ضرایب بهینه محاسبه شده اند و به واحد شناسایی اعمال گشته اند.در این قسمت چگونگی محاسبه ی معیار توضیح داده خواهد شد.روال کار به این گونه است که در مقایسه ی هر کلمه ی مجهول با الگوها چهار فاصله ی سراسری محاسبه می شود (مثلا D1,D2,D3,D4) در این صورت برای هر کلمه ادا شده(مجهول) تفاضل فاصله ی کلمه صحیح ازهریک ازفاصله ی کلمات دیگرمحاسبه می شود بدیهی است که هرچه این مقادیرتفاضل منفی ترباشد(یعنی هزینه کلمه صحیح کوچک و هزینه کلمات دیگر بزرگ باشند) تشخیص صحیحتر صورت گرفته است.

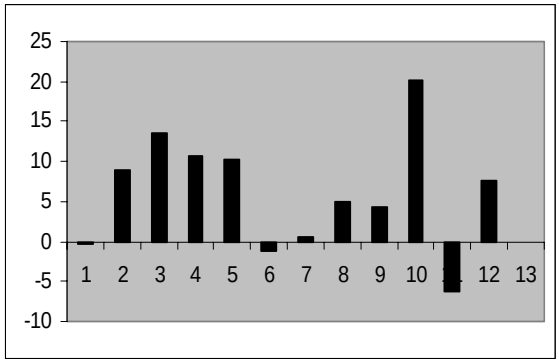
لازم به ذکر است که همه فواصل نرمالیزه شده اند یعنی بر مقدار بیشترین D تقسیم شده اندسپس مجموع مقادیر تفاضل به کمک یک سری توابع جریمه[15] معیار مورد نظر را ایجاد کرده اند.

نتایج این دو آزمایش در جدول ۱ و شکل ۴ ارائه شده است.

جدول ۱-نتایج آزمایش عملکرد دو روش تشخیص

ردیف	کلمه مورد آزمون	معیار قبل ازاعمال ضریب	معیار بعد از اعمال ضریب	میزان بهبود عملکرد ٪
۱	یک	-۱,۰۳۶۷	-۱,۰۳۴۲	-۰,۲۴
۲	دو	-۰,۵۵۸۹	-۰,۶۰۹۲	۹
۳	سه	-۱,۱۱۴۶	-۱,۲۶۶۵	۱۳,۶۳
۴	چهار	-۱,۵۸۷۷	-۱,۷۵۶۴	۱۰,۶۳
۵	یک	-۰,۸۴۶۴	-۰,۹۳۳۶	۱۰,۳۰
۶	دو	-۰,۸۳۵۸	-۰,۸۲۵۲	-۱,۲۷
۷	سه	-۱,۱۱۳۵	-۱,۱۲۰۹	۰,۶۶
۸	چهار	-۱,۲۶۴۸	-۱,۳۲۶۳	۴,۸۶
۹	یک	-۱,۰۵۵۹	-۱,۱۰۱۴	۴,۳۱
۱۰	دو	-۱,۳۰۸۴	-۱,۵۷۲۷	۲۰,۲
۱۱	سه	-۱,۰۸۵۲	-۱,۰۱۶۷	-۶,۳۱
۱۲	چهار	-۱,۳۷۹۳	-۱,۴۸۳۸	۷,۵۸

شکل۴-نمودار درصد بهبود تشخیص



همان گونه که در جدول مشاهده می شود در بیشتر موارد، قدرت شناسایی به کمک الگوریتم ژنتیک بالا تر رفته است.

از جدول می توان میانگین میزان بهبود را برابر ۶,۱۱ ٪ محاسبه کرد. بدیهی است هر چه مجموعه آموزشی کامل تر شود نتایج دقیقتر خواهد شد.

سپاسگزاری

در اینجا لازم است از راهنمایی های جناب آقای دکتر مافی نژاد و جناب آقای مهندس معروضی قدردانی شود.

مراجع

[1] L. R. RABINER and R. W. SCHAFER, "*Digital Processing of Speech Signals*," Prentice Hall, New Jersey, 1978

[2] Dr.John G.Harris,"Isolated Word ,Speech Recognition Using Dynamic Time Warping towards smart appliances",project towards eel 6825:pattern recognition

[3]Coley D.A.,"An Introduction to genetic Algorithms for scientists and engineers " , word scientific,2000

[4]Goldberg D.E,"Genetic Algorithms in Search Optimization and Machine Learning" ,Adisson Wesley Longman Inc.,1997

[5] Homaifar, A., Lai, SH, & Qi, X.. "Constrained optimization via genetic algorithms". Simulation, 62 (4), 242-254,1994

[6] محمد تاجمیری و مهرداد نورانی,"موازی سازی الگوریتم های تشخیص گفتار",پنجمین کنفرانس مهندسی برق ایران, دانشگاه صنعتی شریف ,۱۳۷۶

پی نوشت ها

1-Dynamic Time Warping
2-Isolated word recognition
3-Feature extraction
4-Cost function
5-Real time
6-Overlap
7-Mel-Cepstrum
8-Linear Prediction Coding
9-Local distance
10-Global distance
11-Selection operator
12-Cross over operator
13-Mutation operator
14-Initial population
15-penalty functions