بازشناسی کد شناسائی شخصی و تصدیق هویت گوینده به منظور کنترل دسترسی از راه دور توسط تلفن

محمد مهدی همایونپور — استادیار
امیر نجاری — کارشناسی ارشد
دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیر کبیر

چکیده

این مقاله به بیان تحقیقات صورت گرفته در راستای طراحی و پیاده سازی یک سیستم تصدیق هویت گوینده از راه دور و از طریق خطوط تلفنی می پردازد. در این سیستم، گوینده هویت خود را توسط یک کد شناسائی ۷ رقمی اعلام می نماید. یک الگوریتم بازشناسی ارقام مبتنی بر شبکه عصبی پیشگو که تلفیقی از شبکه عصبی چند لایه پرسپترون و الگوریتم برنامه‌ریزی پویا می باشد اقدام به بازشناسی ارقام کد شناسائی کاربر می نماید. با اعلام هویت گوینده، مدل مرجع آن گوینده که با استفاده از الگوریتم تلفیقی درخت برآمدگی و الگوریتم ژنتیکی آموزش دیده استخراج و با مقایسه همان گفتار گوینده در بیان کد شناسائی وی با این مدل مرجع، عمل تصدیق هویت صورت می گیرد. در آزمایشات صورت گرفته برای بازشناسی ارقام تاثیر تعداد دفعات ارائه داده‌های آموزشی به شبکه عصبی، تاثیر برش یا عدم برش خروجی گره ها برای جلوگیری از حالت اشباع، تاثیر تعداد ویژگیها و نوع ویژگی از نقطه نظر ایستا و گذرا بودن و نیز تاثیر ایجاد تغییرات در پارامترهای یادگیری شبکه مورد بررسی قرار گرفت. همچنین طی آزمایشهائی کارائی سیستم تصدیق هویت پیشنهاد شده ارزیابی و روشهای مختلفی برای تعیین سطح آستانه تصمیم‌گیری مورد بررسی و آزمایش قرار گرفت و نتایج حاصل از بازشناسی ارقام و تصدیق هویت با استفاده از روشهای گفته شده با نتایج حاصل از روش کلاسیک چندی سازی برداری مقایسه گردید.

کلمات کلیدی

تصدیق هویت گوینده، بازشناسی گفتار، شبکه عصبی پیشگو، برنامه ریزی پویا، شبکه عصبی درخت برآمدگی، چندی سازی برداری، الگوریتم های ژنتیکی، نرمال سازی گروهی

Personal Identification Number (PIN) Recognition and Speaker Verification for Access Control Over PSTN

M. Mehdi Homayounpour — Assistant Professor
Amir Najari — M.Sc. Student

Computer Engineering Department
Amirkabir University of Technology

Abstract

An implementation of a speaker verification system is presented in which the speaker claims his/her identity by a seven digits PIN over telephone. PIN digits are recognized by a hybrid predictive neural network system using multi-layer perceptron and dynamic programming algorithm. The speaker verification was done by comparing a speaker's model to his/her speech when uttering the seven digits PIN. The speaker's reference model was trained using Bumb tree and Genetic algorithm. The total number of epochs used to train the neural network, feature types and the dimension of feature vectors, and the effect of different neural net parameters were studied on Farsi digit recognition performance. Some other experiments were conducted to assess the performance of the proposed speaker verification method and to evaluate the effect of different types of decision threshold calculation on the speaker verification error rate.

Keywords

Speech Recognition, Speaker Recognition, Speaker Verification, Predictive Neural Network, Bumb Tree, Genetic Algorithm, Dynamic Programming, Cohort Normalization

مقدمه

امروزه شاهد فراگیر شدن فناوری پردازش خودکار گفتار در امور صنعتی، پزشکی، اداری و عمومی جامعه می‌باشیم. یکی از شاخه های مهم این فناوری تصدیق[1] هویت گوینده می باشد. تصدیق هویت گوینده چنانچه از راه دور و از طریق سیستم های تلفنی، شبکه های مخابراتی، کامپیوتری و اینترنت صورت گیرد ضمن ایجاد تسهیلات فراوان، از جاذبه بالائی برخوردار می‌باشد. در یک مثال میتوان به موردی اشاره نمود که در آن فردی از محل کار خود یا از منزل با بانک محل حساب خود تماس گرفته و پس از آنکه هویت وی از طریق صدای وی تائید شد، امکان انجام عملیات بانکی به وی داده شود. مثالی دیگر از اینگونه موارد، دسترسی به اطلاعات و بانکهای اطلاعاتی محرمانه از راه دور می باشد. تحقیقات بین المللی گسترده‌ای در حال حاضر در این زمینه در حال انجام میباشد که هدف از این تحقیقات عملیاتی و کاربردی نمودن این سیستم هاست. در مواردی نیز این سیستم ها به مرحله عملیاتی رسیده و در حال حاضر در مراکزی از قبیل بانکها و شرکتهای بیمه، مراکز کنترل پالایشگاه ها و نیروگاه‌ها، آزمایشگاه های تحقیقات استراتژیک و حتی بیمارستانها از سیستم های تصدیق هویت در امن نمودن دسترسی افراد مجاز به اطلاعات و اماکن خاص استفاده می شود. یقینا امن نمودن دسترسی به اطلاعات و اماکن به گونه‌ای که نیاز به استفاده از کارت های مغناطیسی، کدهای محرمانه و استفاده از کلمات رمز که امکان گم شدن، جعل شدن و حتی فراموشی آنها وجود دارد را مرتفع نماید، در کشور ما نیز مد نظر بوده و خیلی زود میتواند همگانی شده و در فرهنگ مردم کشور ما وارد گردیده و موجبات راحتی و تسهیلات بیشتر در زندگی مردم را فراهم آورد.

تصدیق هویت گوینده شاخه ای از مبحث بازشناسی گوینده است که هدف از آن شناسائی افراد از طریق صدای آنها می باشد. سابقه تحقیقات در این زمینه به بیش از دو دهه قبل میرسد. بازشناسی گوینده میتواند با شنیدن، مشاهده اسپکتروگرام توسط افرادمتخصص و نیز بصورت خودکار صورت گیرد. روش اول و دوم بیشتر جنبه های پلیسی و قانونی داشته و به منظور شناخت مجرمین و یا رفع سوء ظن از متهمین مورد استفاده قرار می گیرد. آنچه در اینجا مد نظر است استفاده از روشهای خودکار در بازشناسی گوینده می باشد. در این روشها با دریافت صدای یک فرد از طریق میکروفون و رقمی نمودن آن اقدام به استخراج ویژگیهائی[2] از گفتار می نمائیم که می توانند در متمایز نمودن این فرد از سایر گویندگان مفید واقع شود. به کمک این ویژگیها، در مرحله آموزش مدلی از گوینده ساخته شده و در مراحل بازشناسی با مقایسه صدای گوینده با این مدل و تعیین میزان شباهت بین آندو، عمل بازشناسی صورت می‌گیرد. ویژگیهای متعددی چون چگونگی تغییرات فرکانس ارتعاش تارهای صوتی که اصطلاحا" گام[3] نامیده میشود، فرکانسهای تشدید مجرای گفتار یا باصطلاح فرمانتها[4]، ضرائب طیف فوریه سیگنال گفتار، ضرائب پیشگوئی خطی[5] و مشتقات آن چون ضرائب انعکاسی[6]، ضرائب نسبت سطوح مقطع[7] و ضرائب زوج خطوط طیفی[8]، ضرائب کپستروم حاصل از آنالیز فوریه در معیار مل[9]، ضرائب کپستروم حاصل از آنالیز پیشگوئی خطی[10] و بسیاری ویژگیهای دیگر برای مشخص نمودن و بیان خصوصیات وابسته به گوینده موجود در گفتار مورد استفاده قرار گرفته و کارائی آنها ارزیابی شده است. این ویژگیها عموما سعی در مدل نمودن خصوصیات مجرای گفتار و یا خصوصیات سیگنال تحریک ناشی از ارتعاش تارهای صوتی گوینده نموده و به تنهائی یا ترکیبی از آنها برای بازشناسی گوینده بکار رفته اند. کارائی این ویژگیها می‌تواند در محیط های نویزی و در مواردی که صدا از طریق خطوط تلفنی و مخابراتی منتقل میشود تحت الشعاع قرار گیرد[1] . روشهائی چون لیفتر[11] نمودن میانگذر و نیز ویژگیهای دیگری چون RASTA و PLP پیشنهاد شده اند که تا اندازه ای با تاثیرات نویز و محدودیتهای حاصل از خطوط مخابراتی چون محدودیت پهنای باند، اضافه شدن نویز و مانند آن مقابله می نمایند. به منظور مدل نمودن گویندگان نیز روشهای متعددی پیشنهاد گردیده است که از آن جمله میتوان به روشهای در هم پیچیدن زمانی[12] ، چندی سازی برداری[13]، مدل مخفی مارکوف[14]، روشهای آماری مرتبه د و[15] روشهای شبکه عصبی[16] [2] و سیستم‌های فازی[3] اشاره نمود.

در این مقاله سیستمی برای تصدیق هویت گویندگان از طریق خطوط تلفن پیشنهاد گردیده است. در سیستم پیشنهادی قبل از مرحله تصدیق هویت گوینده، ارقام کد شناسائی شخصی گوینده بازشناسی میشوند.

بازشناسی ارقام و در حالت کلی تر، بازشناسی کلمات، یکی دیگر از شاخه‌ای پردازش خودکار گفتار می‌باشد که کاربردهای متعددی چون ارائه فرمانهای صوتی برای آن متصور می‌باشد. در بازشناسی ارقام نیز به جای استفاده مستقیم از سیگنال گفتار ویژگیهائی از آن استخراج میگردد که حاوی اطلاعات مفهومی گفتار می‌باشد. اصولا کلیه ویژگیهائی که برای بازشناسی گوینده در بالا بر شمردیم در بازشناسی ارقام مورد استفاده قرار میگیرند. لیکن ویژگی که بیش از همه تا کنون برای این منظور بکار گرفته شده است ضرائب کپستروم حاصل از آنالیز فوریه در معیار مل و مشتقات اول و دوم آنها میباشد که حتی در محیط های نویزی و تلفنی نیز کارائی خوبی از خود نشان داده‌اند. در بازشناسی ارقام نیز نیاز به استفاده از تکنیکهائی برای مدل نمودن هر رقم می‌باشد که کلیه تکنیکهای مدل سازی که برای تصدیق هویت برشمردیم برای این منظور نیز بکار گرفته شده‌اند.

اصولا پیاده سازی یک سیستم بازشناسی ارقام و بازشناسی گوینده و نیز آموزش، تست و ارزیابی آن نیازمند استفاده از یک دادگان گفتار مناسب می باشد. از آنجا که ما در سیستم طراحی شده، بازشناسی ارقام فارسی و نیز تصدیق هویت با استفاده از گفتار مربوط به همین ارقام را صورت میدهیم و از طرفی مایل به استفاده از سیستم طراحی شده از طریق تلفن می باشیم، بدلیل آنکه یک دادگان تلفنی از ارقام فارسی در اختیار نداشتیم، اقدام به ضبط یک دادگان صدای تلفنی نمودیم.

در این مقاله از شبکه عصبی پیشگو در بازشناسی ارقام و از تلفیق درخت عصبی بر آمدگی و الگوریتم ژنتیکی در تصدیق هویت گوینده استفاده شده است. ویژگی مورد استفاده در بازشناسی ارقام ضرائب کپستروم حاصل از آنالیز فوریه بر مبنای معیار مل و ویژگی مورد استفاده در تصدیق هویت ضرائب کپستروم حاصل از آنالیز پیشگوئی خطی می باشند. روشهای مختلف تعیین سطح آستانه و تصمیم گیری مورد ارزیابی قرار گرفته‌اند. روش چندی سازی برداری نیز، هم برای بازشناسی ارقام و هم برای تصدیق هویت گوینده مورد ارزیابی و نتایج حاصل از این روش با نتایج روشهائی که قبلا بیان شد، مقایسه و مورد تحلیل قرار گرفته اند.

در بخش ۱ این مقاله ابتدا به توصیف دادگان گفتار ضبط شده و در بخش ۲ به شرح روش پیشنهادی در بازشناسی ارقام می پردازیم. بخش ۳ و ۴ نیز به ترتیب به تصدیق هویت گوینده و بیان نتیجه‌گیری اختصاص دارند.

۱- دادگان صدای FARSDIGITS1

در یک سیستم بازشناسی گفتار و گوینده، داده‌های مختلف بسته به کاربرد سیستم در غالب دادگان گفتار یا بعبارتی پایگاههای داده صدا جمع‌آوری و به منظور آموزش و ارزیابی آن سیستم بکار گرفته میشوند. در این کار تحقیقاتی نیز به منظور ارزیابی سیستم تصدیق هویت و بخش بازشناسی کد شناسائی شخص گوینده و برآورد کارائی تکنیکهای مورد استفاده در طراحی آن و بدلیل در اختیار نبودن یک دادگان تلفنی استاندارد در زبان فارسی اقدام به طراحی و ضبط یک پایگاه داده بنام FARSDIGITS1 نمودیم. از آنجا که اهمیت سیستم های تصدیق هویت گوینده هنگامی بارز میشود که از آنها از راه دور و از طریق خطوط مخابراتی و تلفنی استفاده گردد، لذا دادگان مورد استفاده در این مقاله از طریق خطوط تلفن ضبط شده است. بدین معنا که گویندگان گفتار خودر را در یک طرف ارتباط تلفنی بیان و در طرف دیگر دستگاه ضبط صدا، گفتار آنها را ضبط نموده است. صدای ضبط شده از طریق خطوط تلفن دارای محدودیت پهنای باند و نویز بیشتری بوده و تحت تاثیر خصوصیات کابل و کانالهای مخابراتی، اکو و مانند آن قرار میگیرد و در نتیجه با صدائی که مستقیما و از طریق یک میکروفون و در محیط عاری از نویز و سرو صدا ضبط میشود کاملا متفاوت است. خصوصیات صدای تلفنی باعث میشود که بازشناسی گفتار و گوینده به مراتب مشکلتر گردد.

برای تشکیل این دادگان از ۱۰۰ گوینده شامل ۶۱ گوینده مذکر و ۳۹ گوینده مؤنث با متوسط سنی ۳۰ سال خواسته شد تا ارقام صفر الی نه فارسی را طی ۱ الی ۲ جلسه با فاصله زمانی ۷ الی ۳۰ روز بیان نمایند. گفتار این گویندگان با استفاده از یک مدار واسط تلفنی و کارت صدا با فرکانس نمونه‌برداری ۱۱۰۲۵ هرتز و دقت ۱۶ بیت بازاء هر نمونه، نمونه‌برداری شد. گویندگان عمدتا دارای لهجه تهرانی بودند. هر گوینده حداقل ده بار ارقام ۰ الی ۹ را بیان نموده است و هر تکرار این ارقام در یک فایل ذخیره شده است. ابتدا و انتهای ارقام موجود در هر فایل نیز با استفاده از اسپکتروگرام آن بصورت دستی تعیین و در نتیجه بازاء هر فایل صدا یک فایل برچسب ایجاد گردید که در این

فایل علاوه بـر مشخصات کلی مربوط بـه ضبط صدای گوینده، اطلاعات مربوط به شروع و انتـهای موقعیت ارقام نیز قرار داده شـد. بـه منظـور اسـتفاده از ایـن دادگـان در سیستم تصدیق هویت گوینده، به هر گوینده یـک رشـته ۷ رقمی تصادفی از ارقام به عنـوان کـد شناسـائی شـخصی او تخصیص داده شده است که با استفاده از فایلهای برچسب و فایلهای صدای متناظر آنها و با انتخاب و کنار هم گذاشـتن ارقام مربوط به کد ۷ رقمی مورد نظر از فایلهای حاوی ارقام ۰ تا ۹، این کد ساخته می‌شود. یعنی مثل اینکه گوینده این کد را دقیقا به همان صورت بیان نموده باشد. به این ترتیب این امکان فراهم گردید که برای هر گوینده کد شناسائی او که توسط خود او بیان شده باشد و نیز کد شناسـائی او کـه توسط سایر گوینده ها بیان شده باشد را تولید نمائیم.

۲- بازشناسی ارقام مسـتقل از گوینـده توسط شبکه عصبی پیشگو

قبل از این گفتیم که هر گوینده به منظـور معرفـی خود، کد شناسائی شخصی خـود را کـه یـک کـد ۷ رقمـی است بیان می نماید. یکی از وظایف ایـن سیسـتم آن اسـت که این کد شناسائی شخصی را بازشناسی نموده و ارقام آنرا تعیین نماید. لذا برای این منظور به یک سیستم بازشناسـی ارقام فارسی ۰ تا ۹ نیاز خواهیم داشت که در این بخش بـه بیان الگوریتم های بکار رفته بـرای ایـن منظـور پرداختـه و نتایج حاصله را بیان خواهیـم نمـود. لازم بذکـر اسـت کـه بازشناسی ارقام در اینجا بصورت مستقل از گوینـده صـورت گرفته و فرض بر آن است که گوینده ارقام را بصورت مجـزا و جدا از هم و از طریق تلفن بیان مـی نمـاید. بـاز شناسـی ارقام بیان شده بصورت ناپیوسـته شـامل مراحـل متعـددی است که عمده ترین آنها تعییـن محـدوده ارقـام، اسـتخراج ویژگی ها و مدل نمـودن آنـها مـی باشـد. تعییـن محـدوده کلمات (ارقام) و الگوریتـم مـورد اسـتفاده بـرای ایـن روش می‌بایست به صورت خودکار صورت گیرد. از آنجا که در این بخش هدف ارزیابی الگوریتم ارائه شده در بازشناسـی ارقـام می‌باشد. لذا تعیین محدوده کلمات بصورت دسـتی صـورت گرفته تا بدین ترتیـب اشـتباهات ناشـی از تعییـن خودکـار محدوده ارقام، به حساب الگوریتم بازشناسی گذاشته نشـود. در زیر به بیان مدل شبکه عصبی پیشگو در بازشناسی ارقام فارسی می‌پردازیم. شبکه عصبـی پیشـگو تلفیقـی از شـبکه عصبی چند لایه پرسپترونی و برنامه ریـزی پویـا بـروش در هم پیچیدن زمانی می باشـد. در ادامـه ایـن بخـش پـس از پیاده‌سازی شبکه عصبـی پیشـگو مـیزان کـارائی آنـرا بـا استفاده از دادگان ارقام تلفنی ضبط شده، مورد ارزیابی قرار می‌دهیم.

۲-۱- مدل شبکه عصبی پیشگو

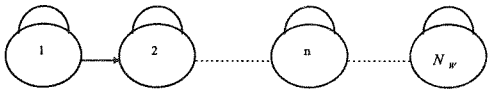
در سالهای اخیر شـبکه عصبـی و بخصـوص شـبکه عصبی چند لایه پرسپترون در زمینه های مختلف پـردازش اطلاعات بویژه بازشناسی گفتار مـورد اسـتفاده قـرار گرفتـه است[۴] و [۵]. از خصوصیات مهم اینگونه شبکه ها می‌توان به الگوریتم یادگیری مؤثر بنام انتشار خطا به عقب[17] [۴] و همچنیـن قـابلیت آنـها در نگاشـت ورودی بـه خروجـی در فضای مسئله اشاره نمود. شبکه عصبی چند لایه پرسپترون که در اینجا به اختصار شبکه عصبی نامیده می‌شود در ایـن مقالـه بـه عنـوان طبقـه بنـدی کننـده کلمـات موجـود در مجموعه کلمات یا لغتنامه[18] که شامل ارقـام ۰ الـی ۹ مـی باشد بکار رفته است. روشی که در اینجا بررسـی مـی‌گـردد نگرش دیگـری بـه مسـئله بازشناسـی ارقـام بـا اسـتفاده از شبکه‌های عصبـی چندلایـه اسـت. در ایـن روش از مـدل پیشگوی عصبی استفاده می‌شود. در این مدل شـبکه هـای عصبی چندلایه بعنوان پیشگویی کننده نمونه ها بکار گرفته می‌شوند. این مدل شامل یک دنباله شبکه عصبی چند لایـه پرسپترون جهت پیشگویی غیرخطی هر کـلاس مـی باشـد. در واقع این مدل از ساختار زمانی گفتـار جـهت بازشناسـی آن بهره می‌گیرد، چـرا کـه ارتبـاط زمـانی بیـن بردارهـای ویژگی متوالی در گفتار از آنجا که شامل اطلاعات مهمی در بازشناسی گفتار است حائز اهمیت زیادی می باشد. در ایـن روش تغییرات سرعت در بیان گفتار با بکارگیری الگوریتـم برنامه‌ریزی پویا نرمالیزه [۶] می‌شود. با توجه به شکل زیـر، شبکه پیشگوی عصبی با استفاده از بردارهای ویـژگـی a_{t-T} a_{t-1} ..... که قبل از بردار ویژگی a_t قرار دارند، بردار ویژگـی a_t را تخمین می‌زنـد. در اینجـا T معـرف تعـداد بردارهـای ویژگی قبلی است که در تخمین بردار a_t مورد استفاده قرار گرفته‌اند. می توان مقدار واقعی a_t را با مقدار تخمینی آن یعنی $\hat{a}_t$ مقایسه نمود و اختلاف آنها که خطای پیشـگویی می‌باشد را بدست آورد یعنی:

$$e = \left\| \hat{a}_t - a_t \right\|^2 \qquad (۱)$$

که e بنام مانده پیشگوئی معروف است. از مانده پیشگویی می‌توان بعنوان تابع خطا جهت اصلاح وزنهای عصبی پیشگو استفاده نمود. این موضوع در قسمت الگوریتم آموزش بیشتر توضیح داده خواهد شد.

۲-۲- الگوریتم بازشناسی کلمات

در روش پیشگویی عصبی چند لایه پرسپترون، مدل هر لغت شامل دنباله ای از پیشگوهای عصبی می‌باشد. برای مثال لغت w با این روش بصورت زیر مدل می‌شود:



شکل (۱) مدل هر لغت در روش ارائه شده.

هر گره خود یک پیشگوی عصبی بوده و تعداد این گره‌ها برای این لغت N_w می‌باشد. این شکل در واقع نحوه تخصیص بردارهای ویژگی یک گفتار را به هر پیشگوی عصبی نشان می‌دهد. گفتار ورودی توسط مدل لغت w به N_w بخش تقسیم بندی و هر بخش از آن به یک پیشگوی عصبی داده می‌شود و بعنوان مثال قسمت n ام گفتار توسط پیشگوی n ام تخمین زده می‌شود. چنانچه n(t) بیانگر این باشد که کدام پیشگوی عصبی مدل لغت w برای تخمین بردار t بکار میرود، دنباله {n(1),n(2),...n(T)} یک مسیر را روی صفحه محاسبات برنامه ریزی پویا نشان می‌دهد. شروط:

$n(1)=1$ و $n(N)=N_w$ و $n(t)=n(t-1)$ or $n(t-1)+1$

در سایر نقاط مسیر می بایست ارضاء گردند. مسئله بازشناسی یک کلمه را می‌توان بصورت یافتن دنباله‌ای از {n*(t)} تعریف نمود، بقسمی که مانده پیشگویی کل حداقل شود. با توجه به شرایط و محدودیتهای مذکور و الگوریتم برنامه ریزی پویا میتوان فرمول بازگشتی زیر را جهت حداقل سازی مانده پیشگویی کل ارائه نمود:

$$g_w(t,n) = d_w(t,n) + \min\left[g_w(t-1,n), g_w(t-1,n-1)\right] \qquad (۲)$$

$d_w(t, n)$ اندازه فاصله محلی (یا مانده پیشگوی) است که بصورت زیر تعریف می‌شود:

$$d_w(t,n) = \left\|\hat{a}_t(w,n) - a_t\right\|^2 \qquad (۳)$$

$g_w(t, n)$ در واقع هزینه رسیدن به نقطه (t,n) از نقطه شروع (1,1) می‌باشد بنابر این می‌توان مقدار $D(w) = g_w(T, N_w)$ که در انتهای فرمول بازگشتی فوق محاسبه می‌شود را بعنوان اختلاف کلی گفتار ورودی با مدل لغت w درنظر گرفت. از این نکته در بازشناسی ارقام استفاده می‌شود، یعنی برای بازشناسی یک گفتار (لغت) آنرا با مدلهای لغت های موجود در لغت نامه مقایسه می‌کنیم. مدلی که کمترین اختلاف کلی با گفتار ورودی را داراست، بعنوان لغت بازشناسی شده اعلام می‌گردد. بعد از یافتن D(w)، با استفاده از تکنیک برگشت به عقب[۱۹] می‌توان بر روی صفحه محاسبات برنامه ریزی پویا مسیر بهینه {n*(t)} را پیدا نمود. این مسیر بیانگر بخش بندی بهینه گفتار ورودی به N_w قسمت بنحوی که مانده پیشگویی کل حداقل گردد، می باشد. از این مسیر بهینه در قسمت بعدی برای آموزش پیشگوهای عصبی استفاده می‌شود.

۲-۳- الگوریتم آموزشی مدل کلمات

منظور از آموزش مدلها یافتن وزنهای پیشگوهای عصبی است بقسمی که مانده پیشگویی کل بازای مجموعه نمونه‌های آموزشی حداقل شود. تابع هدف بصورت متوسط مانده پیشگویی کل تمامی تکرارهای لغت w بصورت زیر:

$$D(w) = \frac{1}{M_w}\sum_{m=1}^{M_w} D(w,m) \qquad (۴)$$

تعریف می‌شود به گونه ای که M_w تعداد دفعات تکرار لغت w و D(w, m) مانده پیشگویی کل به ازای تکرار m ام لغت w می‌باشد. الگوریتم آموزش برای بهینه‌سازی تابع هدف، ترکیبی از الگوریتم برنامه ریزی پویا و الگوریتم آموزش شبکه عصبی چند لایه یعنی انتشار خطا به عقب می‌باشد که آنرا بصورت زیر بیان می‌نمائیم:

۱- وزنهای پیشگوهای عصبی را بصورت تصادفی انتخاب می‌کنیم.

۲- مراحل زیر را برای تمامی دفعات تکرار لغت w یعنی M_w به گونه‌ای که ($1 \le m \le M_w$) اجرا می‌کنیم.

۳- مانده پیشگویی کل یعنی D(w,m) را با استفاده از الگوریتم برنامه‌ریزی پویا محاسبه و نیز مسیر بهینه {n*(t)} را با تکنیک بازگشت به عقب معین می‌نمائیم.

۴- وزنهای هر پیشگوی عصبی را توسط الگوریتم انتشار خطا به عقب بر روی مسیر بهینه {n*(t)} اصلاح می‌کنیم. در اینجا خروجی مطلوب a_t به خروجی واقعی a_t (w, n (t)) در n (t) امین پیشگوی عصبی تخصیص داده می‌شود.

معیار تصحیح وزنها، کاهش مانده پیشگویی کل D (w,m) در امتداد مسیر بهینه {n*(t)} درصفحه محاسبات برنامه‌ریزی پویا میباشد. در نقطه {t, n*(t)} در صفحه مذکور مانده پیشگویی یعنی:

$$e = \left\| \hat{a}_t(w, n^*(t)) - a_t \right\|^2 \qquad (۵)$$

را می‌توان بعنوان تابع خطای انتشار به عقب درنظر گرفت که باید کاهش یابد و بنابراین اصلاح وزنها توسط قانون انتشار خطا به عقب، که در آن خروجی مطلوب a_t متناظر خروجی واقعی a_t (w, n*(t)) در n*(t) امین پیشگوی عصبی می‌باشد، انجام می پذیرد.

۲-٤- استخراج ویژگی و آموزش مدل‌های ارقام

گفتار ۵۸ نفر از گویندگان (۳۶ نفر مذکر و ۲۲ نفر مونث) از دادگان ارقام فارسی FARSDIGITS1 که در بخش ۱ شرح داده شد، برای انجام بررسیهای مربوط به بازشناسی ارقام استفاده شده است. این مجموعه ۵۸ گوینده‌ای به دو مجموعه یکی شامل ۵۰ گوینده برای آموزش و دیگری شامل ۸ گوینده برای ارزیابی تقسیم گردید. به منظور ساختن مدل ارقام، گفتار مربوط به بیان هر رقم به ۲۰ فریم مساوی تقسیم و از هر فریم پس از پیش تاکید[۲۰] و ضرب در پنجره همینگ[۲۱]، یک بردار ویژگی شامل ۱۲ پارامتر کپستروم بر مبنای معیار مل MFCC استخراج گردید. در هر بردار، ضرائب کپستروم به ۱/۱ برابر بزرگترین آنها نرمالیزه شدند.

همانگونه که قبلا اشاره شد در این مقاله مدل نمودن ارقام توسط مدل پیشگوی عصبی صورت گرفته است. تعداد پیشگوهای عصبی در مدل ارائه شده برای تمامی ارقام برابر ۱۰ انتخاب گردید. ساختار شبکه عصبی مورد استفاده در مدل پیشگوی فوق دارای سه لایه با ۱۲×۲ گره در لایه ورودی، ۴ گره در لایه مخفی و ۱۲ گره در لایه خروجی تعیین گردید. تعداد گره‌های لایه ورودی و خروجی به اندازه‌ای انتخاب شده اند که با قرار دادن بردارهای ویژگی خروجی دو فریم در ورودی شبکه بتوان بردار ویژگی فریم بعدی را پیشگوئی نمود. مدل هر رقم با استفاده از ویژگیهای بدست آمده از ۲ تکرار از هر گوینده که جمعا ۱۰۰ تکرار را شامل میشود آموزش داده شد. داده های آموزشی فوق طی ۱۰۰ تکرار در اختیار شبکه عصبی پیشگو قرار گرفت و بدین ترتیب مدل هر رقم بدست آمد.

۲-۵- ارزیابی روش بازشناسی کد شناسائی شخصی

بعد از آموزش مدل کلیه ارقام ۰ تا ۹، اقدام به ارزیابی میزان کارائی روش ارائه شده در بازشناسی ارقام نمودیم. راندمان بازشناسی برای نمونه‌های آموزشی ٪۹۵/۶ درصد بدست آمد. از طرفی طی آزمایش صورت گرفته بر روی داده‌های آزمایشی شامل ۱۰ تکرار از ارقام ۰ تا ۹ که توسط ۸ گوینده بیان شده بودند و سیستم هیچ آشنائی قبلی با آنها نداشت، راندمان ٪۸۳/۷ بدست آمد. طی آزمایشات دیگری اقدام به افزایش تعدا گره های لایه مخفی، تغییر تعداد تکرارها، برش زدن خروجی شبکه و تغییر پارامترهای شبکه عصبی شامل η و ضریب مومنتم α نمودیم. آزمایشات صورت گرفته گویای نکات زیر می باشد :

الف- با افزایش تعداد گره‌های لایه مخفی اگرچه خطای بازشناسی نمونه های آموزشی افزایش می‌یابد اما در عین حال خطای بازشناسی نمونه‌های آزمایشی کاهش می‌یابد. علت این امر تنظیم بیشتر وزنهای شبکه متناسب با نمونه‌های آموزشی و در عوض کاهش تعمیم‌پذیری آن می باشد. جدول ۱ نتایج آزمایش با تعداد لایه مخفی مختلف را نشان می‌دهد.

جدول (۱) نتایج بازشناسی با توجه به تعداد لایه‌های مخفی.

تعداد لایه مخفی	داده های آموزشی	داده های آزمایشی
۲	٪۹۰/۴	٪۷۷/۵
۴	٪۹۵/۶	٪۸۳/۷
۶	٪۹۶/۲	٪۷۱/۲
۹	٪۹۶/۸	٪۷۵/۰

ب- افزایش بیش از حد تعداد تکرار آموزش اگرچه باعث دقیقتر شدن پیشگویی‌ها می‌شود اما بدلیل آموزش بیش از

حد شبکه، تعمیم پذیری مدل کاهش می‌یابد. جدول ۲ نتایج بازشناسی را با توجه به تعداد تکرار آموزش نشان می‌دهد.

جدول (۲) نتایج بازشناسی با توجه به تعداد تکرار آموزش.

تعداد تکرار ها	داده های آموزشی	داده های آزمایشی
۵۰	۹۰/۴٪	۷۷/۵٪
۱۰۰	۹۵/۶٪	۸۳/۷٪
۲۰۰	۹۶/۲٪	۷۱/۲٪

ج- جهت افزایش یادگیری سیستم و جلوگیری از حالت اشباع، خروجی گره‌ها را برش زدیم [۷]. میدانیم که وقتی تابع سیگموئید برای تابع فعالیت انتخاب شود داریم:

$$f(x) = 1/(1+\exp(-x)) \; ; \; f'(x) = f(x)\,[1- f(x)] \qquad (۶)$$

بنابراین وقتی f(x) نزدیک 0 و یا 1 شود، مشتق آن بسمت صفر میل می‌کند در نتیجه یادگیری کاهش می‌یابد. برای جلوگیری از این مسئله مقدار تابع فعالیت گره های لایه مخفی به فاصله [۰/۱, ۰/۹] محدود شدند. جدول ۳ نتایج بازشناسی را با توجه به برش و عدم برش خروجی گره‌ها نشان می‌دهد. چنانچه در این جدول دیده می‌شود نتایج تقریبا مانند هم می‌باشد که این امر به این خاطر است که در شرایط آموزش با این داده های آموزشی، f(x) به 0 و یا 1 نزدیک نشده است. این نتایج بازاء ۱۰۰ بار تکرار بدست آمده است.

جدول (۳) نتایج بازشناسی با توجه به برش خروجی گره‌ها.

برش	داده های آموزشی	داده های آزمایشی
خیر	۹۴/۴٪	۸۳/۳٪
بلی	۹۵/۶٪	۸۳/۷٪

د- برای بررسی نقش تعداد ویژگیهای استفاده شده، آزمایشات دیگری انجام شد. در این آزمایشات، تعداد ویژگیهای برابر ۱۰ و ۱۲ انتخاب شدند. نتایج حاصل در جدول ۴ برای ۱۰۰ تکرار آورده شده است. این جدول نشان میدهد که افزایش ویژگیها در افزایش راندمان بازشناسی مؤثر است.

جدول (٤) نتایج بازشناسی با توجه به تعداد ویژگیها.

تعداد ویژگیها	داده های آموزشی	داده های آزمایشی
۱۰	۹۴/۱٪	۸۳/۳٪
۱۲	۹۵/۶٪	۸۳/۷٪

ه- ارزیابی دیگری نیز به منظور بررسی نقش ویژگیها و مشتق اول آنها که گویای اطلاعات گذرا و دینامیک گفتار می باشند صورت گرفت. برای این منظور دو آزمایش انجام شد که در آنها تعداد ویژگیها مساوی اختیار شدند، با این تفاوت که در آزمایش اول از ۱۶ ضریب کپستروم و در آزمایش دوم ۸ ضریب کپستروم بعلاوه ۸ مشتق اول آنها استفاده گردید. این آزمایشات یک بار بازاء ۲۰۰ تکرار و بار دیگر بازاء ۵۰۰ تکرار صورت گرفتند. جدول ۵ نتایج این بررسی را نشان می‌دهد. این نتایج گویای آن است که استفاده از مشتق اول ضرائب کپستروم یا بعبارتی استفاده از اطلاعات دینامیک موجود در گفتار موجب بهبود بازشناسی ارقام میگردد.

جدول (۵) نتایج بازشناسی با توجه ویژگیهای ایستا و گذرا.

ویژگی	تعداد تکرار	داده های آموزشی	داده های آزمایشی
8MFCC+8 ΔMFCC	۲۰۰	۹۵/۷٪	۸۲/۹٪
8MFCC+8 ΔMFCC	۵۰۰	۹۶/۱٪	۸۰/۸٪
16 MFCC	۲۰۰	۹۵/۰٪	۷۸/۷٪
16 MFCC	۵۰۰	۹۷/۹٪	۷۵/۸٪

و- مسئله مهم دیگر پارامترهای ثابت یادگیری η و ضریب مومنتم α می‌باشد که در یادگیری موثر می‌باشند. مقادیر مختلفی را برای η و α بصورت سعی و خطا آزمایش کردیم و نهایتا مناسب ترین راندمان بازاء $\eta = 0.01$ و $\alpha = 0.6$ بدست آمد.

۳- تصدیق هویت گوینده

قبلا اشاره شد که در یک سیستم تصدیق هویت، پس از اعلام هویت توسط گوینده، میزان شباهت صدای این گوینده و مدل او تعیین و سپس با یک سطح آستانه مقایسه شده و نهایتا نسبت به رد یا قبول او اقدام میگردد. این بخش اختصاص دارد به ارائه روشی برای مدل کردن گویندگان که از روش چندی سازی برداری و شبکه های عصبی استفاده می‌نماید. لازم بذکر است که آموزش این مدل توسط الگوریتم های ژنتیکی صورت میگیرد. همچنین در این بخش به انواع خطاها و نیز موضوع تعیین سطح آستانه تصمیم‌گیری و نیز ارزیابی سیستم های تصدیق

هویت گوینده خواهیم پرداخت و نتایج حاصل از آزمایشات صورت گرفته را ارائه خواهیم کرد.

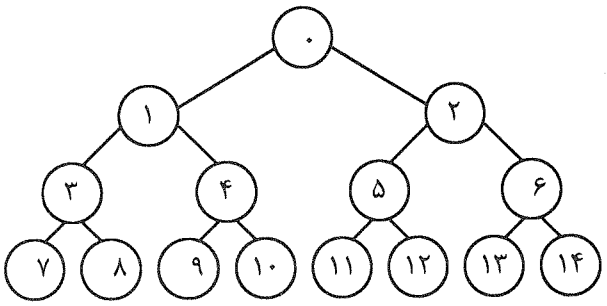
۳-۱- مدل شبکه عصبی درختی

در زمینه بازشناسی الگو، الگوریتمهای مختلفی جهت خوشه‌بندی داده‌ها وجود دارد که عمدتاً در یک سطح داده‌ها را به خوشه‌های مختلف تقسیم بندی می‌نمایند. در این مقاله روشی مورد استفاده قرار می‌گیرد که در آن خوشه‌بندی و طبقه‌بندی داده‌ها بصورت سلسله مراتبی صورت می‌گیرد. تفاوت عمده این روش سلسله مراتبی با سایر الگوریتمهای خوشه‌بندی، سرعت سریع آن جهت یافتن خوشه دربرگیرنده داده ورودی می‌باشد. آموزش این روش بر اساس الگوریتمهای ژنتیکی و شبکه عصبی بنا شده است.

یادگیری بخش مهمی از سیستم‌های هوشمند را تشکیل می دهد. یادگیری انواع مختلفی دارد که از آن جمله می‌توان به یادگیری بدون نظارت و یادگیری با نظارت اشاره نمود. در یادگیری بدون نظارت، در فاز آموزش داده‌های آموزشی به خوشه‌هایی تقسیم می‌شوند بنحوی که داده‌های موجود در هر یک از این خوشه‌ها بیشترین شباهت را به هم داشته باشند. البته باید توجه داشت که تعریف شباهت بسته به کاربرد مشخص می‌شود. در فاز آزمایش نیز می‌بایست تعلق داده ورودی را به یکی از این خوشه‌ها مورد بررسی قرار داد. الگوریتمهای مختلفی جهت خوشه‌بندی داده‌ها وجود دارد که عموماً در یک سطح داده‌ها را به خوشه‌های مختلف طبقه‌بندی می‌نمایند. یک سطحی بودن خوشه‌بندی به این معنا می‌باشد که برای یافتن خوشه در برگیرنده داده ورودی لازم است تمامی خوشه‌های موجود در سیستم بررسی شوند. بنابراین جستجو برای یافتن خوشه در این سیستم از مرتبه $O(n)$ می‌باشد. از این الگوریتمها می‌توان به الگوریتم LBG ، الگوریتم K-means و الگوریتم حداقل واریانس[۲۲] اشاره نمود. تک سطحی بودن خوشه‌بندی موجب اتلاف وقت در هنگام بازیابی می‌شود. برای افزایش سرعت دسترسی به خوشه‌های موجود در سیستم می‌توان آنها را بصورت سلسله مراتبی بنحوی سازماندهی نمود که ساختار حاصل بشکل یک درخت دودویی تبدیل گردد. در این صورت زمان بازیابی و یافتن خوشه‌ها بصورت لگاریتمی و درجه سیستم از مرتبه $O(n)$ به $O(\log_2 n)$ کاهش خواهد یافت. روش خوشه‌بندی که در اینجا مورد استفاده قرار میگیرد بر اساس ایده فوق و درخت برآمدگی[۲۳] که در آن از توزیع‌های نرمال جهت خوشه‌بندی سلسله مراتبی استفاده می‌شود بنا شده است. فاز آموزش این روش در واقع یافتن پارامترهای توزیع‌های نرمال این درخت می‌باشد. در این بررسی جهت یافتن پارامترهای بهینه از الگوریتمهای ژنتیکی استفاده شده است.

۳-۱-۱- شبکه درختی برآمدگی

ساختار درخت دودویی زیر را در نظر بگیرید:



شکل(۲) یک درخت باینری کامل. شماره‌گذاری گره‌های این درخت به این صورت است که ریشه دارای شماره صفر و فرزندان چپ و راست هر گره با شماره n به ترتیب 2n+1و 2n+2 می‌باشند.

این درخت از یک ریشه، تعدادی گره میانی و تعدادی برگ در آخرین سطح تشکیل شده است. متناظر هر گره میانی تابعی به شکل زیر اعمال میشود [۸]:

$$y = \prod_j \left[\frac{1}{\sqrt{2\pi\sigma_j}} e^{-\left(\frac{(x_j-\mu_j)^2}{2\sigma_j^2}\right)} \right] \qquad (۷)$$

در این تابع y حاصلضرب توابع نرمال با میانگین μ_j و واریانس σ_j متناظر عضو jام بردار ورودی می‌باشد. فضای تصمیم گیری چنین گره‌ای یک ابر بیضی می‌باشد که قسمتی از فضای ورودی گره را محصور می‌نماید. درخت برآمدگی یک ساختار مشابه درخت باینری فوق است که شرط زیر را نیز دارا می‌باشد [۹].

"مقدار تابع هر گره میانی باید در تمامی نقاط روی فضای ورودی از توابع متناظر شاخه‌های زیرین خود بزرگتر باشد."

دو گره فرزند گره ریشه باعث می‌شوند که منطقه محصور توسط گره ریشه به دو بخش تقسیم گردد. این عمل تقسیم بصورت بازگشتی برای هر گره تکرار می‌شود تا منطقه داده های ورودی به زیر مناطق کوچکتر تقسیم گردند. در واقع هدف درخت برآمدگی این است که فضای ورودی بطور مکرر به مناطق کوچکتر تقسیم بندی شود تا اینکه نهایتا در سطح برگها عمل طبقه بندی داده‌ها بصورت جداپذیر خطی قابل انجام باشد. در سطح برگها، متناظر هر برگ یک شبکه عصبی خطی وجود دارد که وظیفه طبقه بندی نقاطی که در منطقه آن برگ قرار می‌گیرند را بر عهده دارد. وقتی داده ورودی X به شبکه درختی اعمال شود، شاخه‌های قویتر و در نهایت برگ فعالتر انتخاب می‌شود و در این گره است که عمل طبقه بندی با توجه به خروجی شبکه خطی صورت می‌گیرد. این خروجی توسط رابطه $z^{leaf} = \sum_j w_j^{leaf} xj$ بدست می‌آید که در آن z^{leaf} خروجی ilم برگ شماره leaf، w_j^{leaf} jام وزن این برگ و x_j عضو jام بردار ورودی می‌باشد. این روند بصورت شبه کد در روابط ۸ آورده شده است. این الگوریتم در واقع مرحله بازشناسی و بازیابی می‌باشد. نکته مهم در انتخاب تابع گره‌های میانی در این است که با انتخاب توابع گره‌های میانی دیگر، شبکه‌هایی با خصوصیات و کاربردهای دیگر حاصل می‌شود. برای مثال، شکل۳ توابع دیگر موجود برای گره‌های میانی را نشان می‌دهد. انتخاب هر یک از این توابع، شبکه‌های درختی دیگر را باعث می‌شود. از انواع دیگر این نوع درختها می‌توان به Oct-Tree، Kd-Tree، BoxTree و BallTree اشاره نمود[۹]. تابع قسمت (B) در شکل ۳ در واقع همانی است که در شبکه درختی برآمدگی مورد استفاده قرار می‌گیرد. همانطور که دیده می‌شود این تابع بصورت یک برآمدگی[۲۴] است و به این دلیل نام این نوع شبکه‌های به نام شبکه درخت برآمدگی مشهور است.

شکل(۳) توابع مختلف برای انتخاب بعنوان توابع گره‌های میانی. انتخاب تابع قسمت (A) باعث بوجود آمدن درختهای Oct_Tree، Kd-Tree و BoxTree می‌شود. توابع قسمت (B) و (C) در درختهای BumpTree و BallTree (C) مورد استفاده قرار می‌گیرد. تابع قسمت (D) در درخت Kd-Tree (D) کارایی بالائی ایجاد می کند[۹].

$$Index = 0 \qquad /* this\ is\ root\ index */$$
$$While\ (Not\ isaLeaf(Index))$$
$$Begin$$

$$y_{left}^{Index} = \prod_j \left[\frac{1}{\sqrt{2\pi}\sigma_{left,j}^{Index}} e^{-\left(\frac{(x_j - \mu_{left,j}^{Index})^2}{2\sigma_{left,j}^{Index\ 2}} \right)} \right]$$

$$y_{right}^{Index} = \prod_j \left[\frac{1}{\sqrt{2\pi}\sigma_{right,j}^{Index}} e^{-\left(\frac{(x_j - \mu_{right,j}^{Index})^2}{2\sigma_{right,j}^{Index\ 2}} \right)} \right]$$

$$If\ (y_{left}^{Index} > y_{right}^{Index})\ /* Select\ Left\ Branch */$$
$$Index = Index = 2 * Index + 1$$
$$Else \qquad /* Select\ Right\ Branch */$$
$$Index = Index = 2 * Index + 2$$
$$EndIf$$
$$End\ While$$
$$Leaf = Index$$
$$z^{Leaf} = \sum_j w_j^{Leaf} x_j$$
$$If\ (z^{Leaf} > Threshold)$$
$$x \in Leaf \qquad (۸)$$
$$Else$$
$$x \notin Leaf$$
$$EndIf$$

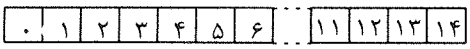
۳-۱-۲- آموزش شبکه درختی

برای آموزش شبکه درختی برآمدگی اولین قدم تعیین ساختار درخت یعنی تعداد لایه‌ها، میانگین و واریانس گره‌های میانی می‌باشد. در مرحله دوم لازم است که شبکه‌های خطی متناظر هر برگ آموزش ببینند. روشهای مختلفی برای تعیین ساختار درخت وجود دارد که در اینجا از الگوریتمهای ژنتیک جهت یافتن ساختار بهینه استفاده شده است. روش دیگر خوشه‌بندی بازگشتی است که در آن داده‌های آموزشی با استفاده از الگوریتمهای خوشه بندی معمولی بطور بازگشتی به دو قسمت تقسیم می‌گردند و آنگاه بعد از هر تقسیم بندی عمل خوشه‌بندی با توجه به اینکه باید شامل توزیع‌های نرمال باشند، اصلاح می‌گردند. جهت آموزش شبکه‌های خطی از قانون پرسپترون و تنها در

یک مرحله برای حداقل سازی خطای داده‌های آموزشی استفاده می‌کنیم.

۳-۱-۳- تعیین ساختار بهینه شبکه درختی برآمدگی توسط الگوریتم های ژنتیک

گام اساسی در بکارگیری الگوریتمهای ژنتیکی یافتن کدینگ مناسب جهت کد کردن اطلاعات می‌باشد زیرا الگوریتمهای ژنتیکی نه بر پارامترها بلکه بر کد شده آنها اعمال می‌شوند. در واقع می‌توان گفت میزان کارایی الگوریتمهای ژنتیکی بستگی زیادی به نحوه کد کردن دارد. درخت شکل ۲ را مجدداً در نظر بگیرید. این درخت را می‌توان با یک آرایه با طول ثابت ۱۵ بصورت شکل ۴ نشان داد.



شکل (٤) نمایش بلاکی درخت باینری شکل۲. هر بلاک معرف یک گره درخت است. هر بلاک شامل اطلاعات گره متناظر در درخت باینری نیز می‌باشد. ترتیب شماره‌گذاری هر یک از این بلاکها مانند شماره‌گذاریهای گره‌ها در درخت باینری می‌باشد. در علم ژنتیک به کل بلاکها یک کروموزم و به هر یک از اطلاعات موجود در هر بلاک یک ژن گفته می‌شود.

هر بلاک (خانه) در شکل فوق معرف یک گره درخت و هر ژن (مقدار) در هر بلاک معرف یک پارامتر آن گره مثلا میانگین و یا واریانس می‌باشد. بلاک شماره صفر بیانگر ریشه درخت می‌باشد. فرزندان یک گره شماره n در بلاکهای شماره 2n+1 و 2n+2 قرار می‌گیرند. مسئله مهمی که در کدینگ شبکه درختی برآمدگی وجود دارد وجود وابستگی بین پارامترهای گره‌های مختلف می‌باشد. برای مثال با توجه به شرط اصلی این درختها باید پارامترهای گره‌های فرزند در فضای تعریف شده توسط گره پدر قرار گیرند. بنابراین پارامترهای گره‌های فرزند بایستی با توجه به گره‌های پدر انتخاب شوند. استفاده از یک کدینگ معمولی که در آن باید هر پارامتر مستقل از پارامترهای دیگر ارائه‌گردد، موجب می‌شود که با اعمال عملگرهای ژنتیکی (جهش و ترکیب) ساختارهایی بوجود آیند که شرط اصلی شبکه‌های درختی برآمدگی را دارا نباشند. این امر حتی با انتخاب نسل اول از شبکه‌های معتبر با گذشت زمان در نسلهای آتی بوقوع خواهد پیوست. برای رفع این مشکل روش کدینگ زیر معرفی می‌شود. قبل از معرفی این روش لازم است که نحوه تعریف فضای هر گره مشخص شود. تابع میانی درخت در معادله ۷ را می‌توان بصورت زیر نیز نشان داد:

$$y = \frac{1}{\prod_j \sqrt{2\pi}\sigma_j} e^{-\left(\sum_j \frac{(x_j-\mu_j)^2}{2\sigma_j^2}\right)} \qquad (۹)$$

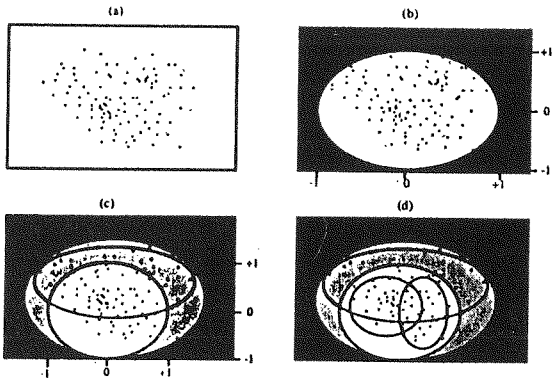
برای معرفی فضای تعریف شده توسط این گره از توان قسمت نمایی استفاده می‌کنیم یعنی:

$$S = \frac{(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{(x_2-\mu_2)^2}{2\sigma_2^2} + \frac{(x_3-\mu_3)^2}{2\sigma_3^2} + \ldots\ldots + \frac{(x_d-\mu_d)^2}{2\sigma_d^2} \qquad (۱۰)$$

مشاهده می‌شود که فضای S یک ابربیضی به مختصات $(\mu_1, \mu_2, \mu_3, \ldots, \mu_d)$ و شعاع $(\sqrt{2}\sigma_1, \sqrt{2}\sigma_2, \sqrt{2}\sigma_3, \ldots, \sqrt{2}\sigma_d)$ می‌باشد که d ابعاد فضا فرض شده است. بنابراین بین شعاع این ابر بیضی و واریانسهای ورودی رابطه مستقیم وجود دارد. بهمین دلیل شعاعهای ابربیضی و واریانسها گاهاً بجای هم مورد استفاده قرار می‌گیرند. در کدینگ ارائه شده، هر ژن یک مقدار حقیقی در فاصله (+۱ و -۱) می‌باشد. ژن در علم ژنتیک در واقع یک فیلد اطلاعاتی مثلاً میانگین می‌باشد. نحوه تبدیل یک کروموزم به یک ساختار درخت برآمدگی به شرح زیر است [۸]:

قبل از هر چیز لازم است فضایی که شبکه در آن تعریف می‌شود مشخص گردد. این فضا بصورت یک ابر بیضی به مرکز میانگین داده‌های ورودی و شعاعی که تمامی داده‌ها را در ابعاد مختلف شامل گردد تعریف می‌شود. حال مرکز سیستم مختصات خود را به مرکز این ابر بیضی انتقال داده و در هر بعد، بردار یکه را معادل شعاع در آن بعد انتخاب می‌کنیم. فضا و سیستم مختصات فوق توسط ریشه درخت معین می‌گردند و ژنهایی که مرکز دو گره فرزند ریشه را کد می‌نمایند برحسب این مختصات بیان می شوند و همچنین ژنهایی که شعاعهای این دو گره را کد می‌کنند بر روی مرکز مختصات پدر و قطر بیضی فضای تعریف

توسط آن نگاشت می‌شوند و این روند بصورت بازگشتی برای تمامی گره‌ها انجام می‌پذیرد. این روند بصورت گرافیکی در زیر ترسیم شده است:

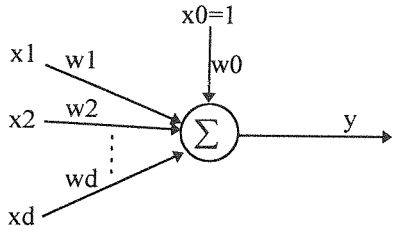


شکل (۵) نحوه تبدیل یک کروموزم به ساختار درخت برآمدگی در فضای دو بعدی. a)-فضای اولیه داده‌های آموزشی.b) - فضای نرمال شده و همچنین سیستم مختصات تعریف شده توسط گره ریشه. c)-و d)- نحوه تقسیم فضای تعریف شده هرگره توسط گره‌های فرزند آن.

با توجه به این کدینگ عملگرهای الگوریتمهای ژنتیکی، دیگر ساختارهای نامعتبر درست نمی‌کنند و همچنین عملگرهای جهش و ترکیب براحتی قابل تعریف می‌باشند. عملگر ترکیب به این صورت تعریف می‌شود که برای ترکیب دو ساختار یک گره بصورت تصادفی اختیار شده آنگاه زیرشاخه‌های متناظر آن گره در دو ساختار درختی با هم جابجا می‌شوند. در عمل جهش یک ساختار درختی یک گره بصورت تصادفی انتخاب شده و پارامترهای آن با احتمال کم با عدد تصادفی در فاصله (۰/۲+ و ۰/۲-) جمع می‌شود. بعد از اعمال عملگر جهش مقادیر پارامترها بین (۱+ و ۱-) برش می‌شوند تا کدینگ دچار مشکل نشود. نشان داده شده است که این روش کدینگ دارای خواص پیوستگی، هم ریختی، کامل بودن، بسته بودن و در برداشتن حداقل اطلاعات تکراری می‌باشد.

۳-۱-٤- آموزش شبکه‌های خطی متناظر برگها

در هر برگ یک پرسپترون ساده وجود دارد که بصورت زیر نمایش داده می‌شود:



شکل (٦) نمایش ورودی‌ها و خروجی یک شبکه پرسپترون تک لایه.

بنابراین $y = \sum_{i=1} w_i x_i + w_0$ که با فرض $x_0 = 1$ بطور خلاصه خواهیم داشت:

$$y = \sum_{i=0} w_i x_i \qquad (۱۱)$$

فرض می‌کنیم نمونه‌های آموزشی $\{x^p\}$ و خروجی‌های مطلوب $\{Y^p\}$ متناظر این نمونه‌ها باشد. خطای موجود در شبکه برابر است با:

$$E = 0.5\sum_p \left(y^p - Y^p\right)^2 \qquad (۱۲)$$

که y^p خروجی واقعی نمونه آموزشی x^p می‌باشد.

$$E = 0.5\sum_p \left(\sum_i w_i x_i^p - Y^p\right)^2 \qquad (۱۳)$$

برای حداقل نمودن خطا باید داشته باشیم:

$$\frac{\partial E}{\partial w_j} = 0 \qquad \forall j, j = 0,1,\ldots,d$$

$$\frac{\partial E}{\partial w_j} = \frac{\partial}{\partial w_j}\left[E = 0.5\sum_p \left(\sum_i w_i x_i^p - Y^p\right)^2\right]$$

$$= \sum_p \left[\left(\sum_i w_i x_i^p - Y^p\right) x_j^p\right]$$

$$\frac{\partial E}{\partial w_j} = \sum_i \left[w_i \left(\sum_p x_i^p x_j^p\right)\right] - \sum_p x_j^p Y^p$$

$$\frac{\partial E}{\partial w_j} = 0 \quad \Rightarrow \sum_i \left[w_i \left(\sum_p x_i^p x_j^p\right)\right] = \sum_p x_j^p Y^p$$

(۱۴)

برای حل معادلات خطی ۱۴ متغیرهای زیر تعریف می‌شوند:

$$M_{i,j} = \sum_p x_i^p x_j^p \qquad (۱۵)$$

با این متغیرها داریم:

$$B_j = \sum_p x_i^p Y^p \qquad (۱۶)$$

و یا بصورت ماتریسی داریم:

$$\sum_i w_i M_{i,j} = B_j \qquad (۱۷)$$

$$WM = B \qquad (۱۸)$$

که W^* بصورت زیر محاسبه می‌شود:

$$W^* = BM^+ \qquad (۱۹)$$

M^+ شبه معکوس M می‌باشد زیرا ممکن است M معکوس‌پذیر نباشد. روش ارائه شده در اینجا اولا قادر به آموزش یک مرحله‌ای شبکه خطی می‌باشد و ثانیا" به حافظه کمی نیاز دارد[۱۰].

۳-۲- انواع خطاها، تصمیم‌گیری و ارزیابی سیستم های تصدیق هویت گوینده

قبل از آنکه به بیان توضیحات بیشتردر خصوص اقدامات صورت گرفته برای تصدیق هویت با استفاده از شبکه عصبی و الگوریتم های ژنتیکی بپردازیم لازم است ابتدا به شرح مختصری از چگونگی ارزیابی سیستم های تصدیق هویت و نحوه تعیین راندمان و کارائی آنها بپردازیم. برای ارزیابی هر سیستم لازم است ماهیت خطاها و اشتباهات موجود در آن سیستم بررسی گردد. در فناوری تصدیق هویت گوینده نیز این ضرورت بدیهی بنظر می‌رسد. همچنانکه در تعریف مسئله تصدیق هویت گوینده لازم است که میزان نزدیکی گفتار ورودی به مدل گوینده ادعا شده معین گردد. اصولا" در تصدیق هویت گوینده یک سطح آستانه جهت اخذ تصمیم در رابطه با پذیرش یا رد ادعای گوینده لازم است که می بایست آنرا در مرحله آموزش سیستم تعیین نمائیم. در این بخش ضمن توضیح انواع خطاهای ممکن در سیستم های تصدیق هویت گوینده، به روشهای بدست آوردن سطح آستانه تصمیم گیری و چگونگی تصمیم گیری جهت رد یا قبول گوینده و تعیین میزان کارائی سیستم در تصدیق هویت گوینده ها اشاره خواهیم کرد.

۳-۲-۱- انواع خطاها

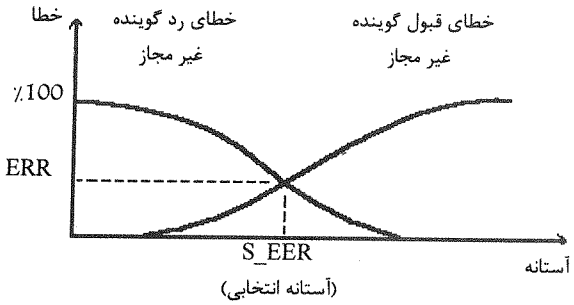
در فناوری تصدیق هویت گوینده دو نوع خطا ممکن است رخ دهد که عبارتند از:

- رد اشتباه[۲۵] یا باختصار FR: رد یک گوینده مجاز یا رد نادرست
- پذیرش اشتباه[۲۶] یا باختصار FA: قبول یک گوینده غیر مجاز یا تصدیق نادرست

۳-۲-۲- تعیین سطح آستانه

یک مرحله اساسی در پیاده سازی سیستمهای تصدیق هویت گوینده تعین مقدار آستانه جهت تصمیم‌گیری می‌باشد. یک مقدار آستانه پائین باعث خطای رد اشتباه بالاتر و درعین‌حال خطای پذیرش اشتباه پائینر می‌شود. مسلما" بسته به نوع کاربرد مقدار این سطح آستانه متفاوت می‌باشد. مثلا درکاربردهای نظامی بسیار مطلوب است که نرخ خطای رد بالا باشد یعنی گرچه تعدادی گویندگان مجاز اشتباها رد می‌شوند اما درعین‌حال گویندگان غیر مجاز به سختی ممکن است پذیرفته شوند که این خود باعث افزایش امنیت سیستم می‌شود. اما درکاربردهای دیگر مثلا یک سیستم بانکی مطلوب است که درصد تصدیق هویت گویندگان مجاز بالا باشد، هر چند ممکن است که درصد قبولی گویندگان غیر مجاز نیز زیاد شود.

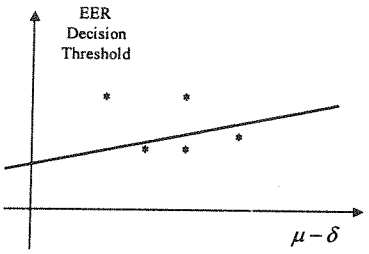
روشهای مختلفی جهت پیدا نمودن سطح آستانه‌ای ارائه شده است. یک روش معمول برای انجام اینکار، روش نرخ خطای برابر[۲۷] و یا بطور مختصر EER می‌باشد. در این روش مقدار آستانه‌ای طوری انتخاب می‌شود که درصد پذیرش اشتباه برابر با درصد رد اشتباه گردد. شکل زیر نحوه یافتن این سطح آستانه را نشان می‌دهد:



شکل (۷) یافتن آستانه تصمیم‌گیری به روش نرخ‌خطای برابر.

بنابراین برای یافتن سطح آستانه، به ازای هر گوینده مجاز نمودار شکل ۷ تشکیل و از روی آن سطح آستانه تصمیم‌گیری برای آن گوینده معین می‌گردد. یک اصلاحیه بر این روش این است که برای پیشی‌بینی وضعیتهای آینده، بجای استفاده از EER، آن را در یک ضریب که معمولا بزرگتر از ۱ است ضرب می‌کنند [۱۱]. اگر این ضریب ۱ باشد، مقدار نهایی همان EER خواهد بود. اگر این مقدار از ۱ بزرگتر باشد، سطح آستانه بطرف راست انتقال پیدا می‌کند که در اینصورت خطای قبول گوینده غیر مجاز زیاد ولی خطای رد گوینده مجاز کم می‌شود. این عمل برای مواردی که شرایط محیط در هنگام بکارگیری سیستم تغییرات زیادی می‌کند، مثلا روی خطوط تلفن بسیار مفید می‌باشد.

در روش ارائه شده لازم است که نمودار شکل ۷ برای بدست آوردن سطح آستانه تک تک گویندگان رسم شود و اگر به سیستم گوینده جدیدی اضافه گردد این روند برای بدست آوردن سطح آستانه تصمیم گیری وی نیز باید تکرار شود. در ادامه روشی ارائه می‌شود که در آن با توجه به اطلاعاتی که از تعداد محدودی گوینده آموزشی بدست می آوریم بتوانیم مقادیر سطوح آستانه را برای گویندگان جدید بدست آوریم [۱۲]. در این روش بازاء هر گوینده آموزشی، مدل وی با تکرارهائی از گفتار او و نیز تکرار هائی از گفتار سایر گوینده های آموزشی مقایسه و فواصل درون گوینده ای[۲۸] و بین گوینده ای[۲۹] وی بدست می آیند. به کمک این فواصل EER و نیز تفاضل میانگین و انحراف استاندارد فواصل بین گوینده ای یعنی $\mu - \delta$ را بدست می آوریم. بازاء مقدار EER و $\mu - \delta$ هر گوینده، یک نقطه که توسط * مشخص شده در نمودار شکل ۸ بدست می‌آید. این کار را برای کلیه گوینده های آموزشی صورت می دهیم.



شکل (۸) مقادیر آستانه تصمیم‌گیری EER برحسب میانگین منهای واریانس فواصل برون گوینده ای نظیر آنها.

حال با استفاده از روشهای عددی برازش خط، خط زیر را بگونه‌ای از بین نقاط فوق عبور می دهیم که خطای مربوط به نمایش این نقاط توسط این خط حداقل شود. به همین دلیل منبعد این روش را روش برازش خط می نامیم:

$$thresh = c1 * (\mu - \delta) + c2 \qquad (۲۰)$$

در واقع با استفاده از روشهای عددی ضرایب c1 و c2 بدست می‌آیند. با این روش برای بدست آوردن سطح آستانه تصمیم‌گیری یک گوینده نامشخص کافی است که ابتدا فواصل بین‌گویندگی او محاسبه شده و μ، δ و در نتیجه $\mu - \delta$ این فواصل بدست آید. آنگاه با استفاده از رابطه ۲۰، مقدار آستانه تصمیم‌گیری برای آن گوینده بدست می‌آید. نکته قابل ذکر این است که می‌توان اصلاحیه‌ای که بر روش EER اعمال شد، برای این روش نیز بکار برد یعنی اینکه مقدار آستانه حاصل را در ضریبی ضرب نمود تا کارایی سیستم افزایش یابد.

۳-۲-۳- نرمالسازی گروهی فواصل

همانگونه که قبلا گفته شد روش کلاسیک برای تصدیق هویت یک گوینده آن است که فاصله بین گفتار وی با مدل گفتار گوینده ادعا شده اندازه‌گیری و سپس با مقایسه این فاصله با یک سطح آستانه تصمیم گیری گوینده مورد نظر پذیرفته و یا رد شود. مزیت استفاده از چنین اندازه‌گیری جهت تنظیم سطح آستانه برای تصمیم‌گیری راجع به رد یا قبول هویت گوینده این است که این امتیازدهی انحراف و اختلاف واقعی بین مدل گوینده مجاز و گفتار ورودی را نشان می‌دهد. اما این امتیازدهی خام برای هم گوینده مجاز و هم گویندگان غیر مجاز بخاطر تغییرات درون‌گویندگی، تغییرات محیط ضبط و یا تغییرات فونتیکی ممکن است تغییرات بسیار شدیدی نماید و بنابراین بصورت توزیعهایی

با هم پوشانی زیاد مثل آنچه در شکل۷ روی داده است در آیند. در این شرایط، نرخ خطای یکسان و یا EER زیاد شده و باعث می‌شود که کارایی تصدیق هویت گوینده کاهش یابد.

یک روش دیگر نگرش به مشکل فوق استفاده از سطح آستانه پویا می‌باشد که در آن از شباهت نرمال شده گروهی[۳۰] و یا امتیازدهی پراکندگی استفاده می‌شود. در این روش یک گروه از گویندگان که به گوینده مجاز بسیار نزدیک می‌باشند، بعنوان اعضای گروه وابسته به آن گوینده در نظر گرفته می‌شوند. تشکیل گروه را می‌توان به معنی ایجاد یک محیط محلی در فضای تمامی گویندگان دانست که به کمک آن اندازه‌های فاصله یا شباهت را تا اندازه‌ای نرمالیزه می نمائیم. ساده‌ترین شکل این نرمال سازی محاسبه تفاضل زیر می باشد:

A=فاصله بین گفتار ورودی و مدل گفتار گوینده مورد نظر

B=تابعی از فاصله بین گفتار ورودی و مدلهای گویندگان موجود در محیط محلی گوینده مورد نظر

A-B=فاصله نرمالیزه شده

این روش هم در سیستم مستقل از متن و هم وابسته به متن بکار گرفته شده است. برای باز هم بهتر شدن روند تصمیم‌گیری میتوان ترکیبی از دو روش سطح آستانه پویا (روش نرمال سازی گروهی) و روش معمول استفاده نمود[۱۱]. بدین صورت که قبل از عمل نرمال سازی گروهی ابتدا از یک سطح آستانه مطلق استفاده می‌شود که سطح آستانه با استفاده از روش EER محاسبه می‌گردد. ابتدا فاصله واقعی بین گفتار ورودی و گفتار مدل مرجع بدست آمده و اگر این مقدار از سطح آستانه کوچکتر بود، آنگاه از نرمالسازی گروهی جهت بدست آوردن فاصله جدید استفاده می‌شود و در غیر اینصورت گوینده گفتار ورودی سریعا رد هویت می‌شود. بعد از نرمالسازی گروهی، فاصله جدید با سطح آستانه دیگر جهت تصدیق و یا رد نهایی مقایسه می‌شود.

در اینجا مثالی برای استفاده از الگوریتم فوق را برای حالتی که هر گوینده توسط یک کتابچه کد[۳۱] حاصل از تکنیک چندی کردن برداری مدل شده است را ارائه می‌کنیم. لازم به ذکر است که از اصلاحیه روش EER یعنی ضرب یک مقدار ثابت در مقدار نرخ خطای برابر نیز استفاده شده است و همچنین تابع نرمالسازی گروهی یک تابع آماری مثلا میانگین و یا حداقل مقادیر در نظر گرفته شده است [۱۱].

$$
\begin{aligned}
&if \quad S_c < kS_{\max} \\
&\qquad Score_N = S_c - stat[S_i, i = 1,...n] \\
&else \\
&\qquad \mathrm{Score_N} = \infty
\end{aligned}
\qquad (۲۱)
$$

که S_c میزان پراکندگی گفتار ورودی با مدل گوینده مجاز، k ضریب سطح آستانه مطلق، $S_{\max}$ حداکثر میزان پراکندگی حاصل از مقایسه گفتار آموزشی گوینده مجاز با مدل خودش، $Score_N$ امتیاز نرمال شده گروهی، Stat تابع نرمالسازی گروهی مثلا میانگین یا حداقل، S_i میزان پراکندگی حاصل از مقایسه گفتار ورودی با مدل (کتابچه کد) iامین عضو گروه این گوینده و n تعداد اعضای گروه گوینده می‌باشد.

۳-۲-٤- ارزیابی سیستم

برای ارزیابی سیستمهای تصدیق هویت گوینده لازم است کارایی سیستم بصورتی بیان شود تا بتوان سیستمهای مختلف را تحت یک معیار واحد با هم مقایسه نمود. آقایان Gong وHaton نرخ متوسط خطای تصدیق و نرخ خطای تصدیق کل را بصورت زیر پیشنهاد نمودند [۱۳].

$$
E_M = \frac{FA + FR}{2} \qquad (۲۲)
$$

که M تعداد گویندگان موجود در سیستم می‌باشد[۱۲]. E_M که در رابطه فوق تعریف شده متوسط حسابی می‌باشد. تعریف E_G بصورت متوسط هندسی نیز بصورت زیر پیشنهاد گردیده است:

$$
E_G = \sqrt{FA * FR} \qquad (۲۳)
$$

۳-۳- تصدیق هویت گوینده توسط تلفیق درخت برآمدگی و الگوریتم ژنتیکی

در این بخش به تشریح روش تلفیقی حاصل از ترکیب درخت برآمدگی و الگوریتم ژنتیکی برای انجام

تصدیق هویت گوینده و توصیف نتایج حاصل از این روش می‌پردازیم. همانطور که در ابتدای مقاله آوردیم پس از آنکه یک گوینده کد شناسائی شخصی خود که یک کد هفت رقمی است را بیان نمود ابتدا با استفاده از تکنیک گفته شده در قسمت بازشناسی ارقام، کد شناسائی شخصی وی بازشناسی و مشخص میشود. قدم بعدی آن است که هویت این گوینده که خود را به گونه ای که بیان شد معرفی نموده است، بررسی گردد. این عمل بدون آنکه گوینده صحبت اضافه دیگری را بیان نماید، با استفاده از همان گفتار اولیه وی در بیان کد شناسائی شخصی، صورت میگردد.

انجام تصدیق هویت گوینده به مراحل متعددی نیاز دارد که شامل استخراج ویژگیهای مناسب از گفتار گوینده، مدل نمودن گوینده، تعیین سطح آستانه تصمیم گیری جهت پذیرش یا رد گوینده و نهایتا مقایسه گفتار گوینده با مدل فرد ادعا شده و قبول یا رد وی می باشد. به منظور آموزش و ارزیابی روش ارائه شده از دادگان گفتاری که طی این تحقیق ضبط گردیده و توضیحات مربوط به آن قبل از این ارائه شده است استفاده نمودیم. از گفتار ۵۸ گوینده (۳۶ نفر مرد و ۲۲ نفر زن) این دادگان استفاده گردید. گویندگان ارقام فارسی صفر الی نه را ده بار تکرار و از طریق تلفن ضبط شده اند. با استفاده از این ارقام به هر گوینده یک کد ۷ رقمی منحصر به فرد اختصاص داده شد و بدین ترتیب کد شناسائی شخصی وی ایجاد گردید. مجموعه گفتارهای هر گوینده به دوبخش ۶ تایی برای آموزش و ۴ تایی برای تست تقسیم شد. برای آموزش مدل هر گوینده علاوه بر گفتار وی از اطلاعات گفتار گویندگان همجنس او نیز استفاده شد. بدین منظور دو- سوم گویندگان همجنس وی بصورت اتفاقی انتخاب و همراه با اطلاعات وی جهت آموزش سیستم بکار گرفته شدند. در مرحله آزمایش، ۴ تکرار باقیمانده هر گوینده و همچنین یک- سوم گوینده های همجنس باقیمانده و همین تعداد گوینده غیر همجنس مورد استفاده قرار گرفتند. برای آموزش مدل یک گوینده و نیز انجام تستهای تصدیق هویت به کمک این مدل، از گفتار مربوط به بیان کد هفت رقمی توسط این گوینده و سایر گویندگان استفاده گردید.

سیگنال گفتار ورودی مربوط به بیان کد شناسائی گوینده به فریم های ۳۰ میلی‌ثانیه‌ای که ۲۰ میلی ثانیه همپوشانی دارند تقسیم شده و هر فریم تحت آنالیز LPC مرتبه ۱۲ قرار گرفت و از آن ۱۲ ویژگی LPCC استخراج گردید. هر گوینده توسط یک شبکه درختی ۵ لایه کامل شامل ۳۱ گره مدل گردید. تعداد لایه های فوق بدین دلیل انتخاب شدند که اولا تعداد لایه های بیشتر موجب می‌گردید که بدلیل محدودیت داده های آموزشی، مدل درختی هر گوینده شامل برگهائی تهی از داده آموزشی گردیده و در نتیجه درختهای ناقص ایجاد شود، ثانیاً تعداد لایه های کمتر پوشش ضعیفتری از اطلاعات یادگیری در فضای داده های آموزشی ایجاد می‌نمود.

آموزش درخت برآمدگی به روشهای مختلفی میسر است که یک روش آن توسط Bostock شرح داده شده است[۱۴]. در این روش درخت بصورت بالا به پایین و بازگشتی با استفاده از الگوریتمهای خوشه‌بندی ساخته و آموزش داده می‌شود. روشهای دیگری نیز وجود دارند که بتوان با استفاده از آنها درخت برآمدگی را بنحو بهتری آموزش داد که یکی از آنها استفاده از الگوریتم ژنتیکی می‌باشد که در این تحقیق بمنظور بهینه سازی فرآیند یادگیری درخت برآمدگی مورد استفاده قرار گرفته است. همچنانکه در بخش الگوریتم های ژنتیک گفته شد این الگوریتم ها بعنوان تکنیکی برای بهینه سازی مورد استفاده قرار میگیرند و سعی در یافتن حداقل مطلق در فضای مسئله می‌نمایند. و این در حالی است که در روش ارائه شده توسط بوستاک یافتن یک حداقل محلی مورد نظر است، گرچه هیچ تضمینی در یافتن حداقل مطلق توسط این روش وجود ندارد. مشخصات الگوریتم ژنتیکی مورد استفاده در این تحقیق عبارت است از:

- تعداد تکرار (نسل) : ۵۰ تکرار
- جمعیت هر نسل: ده درخت
- نرخ ترکیب: ۷۰٪
- نرخ جهش: ۱۰٪
- تابع معیار: معکوس میانگین مربعات خطاهای شبکه‌های خطی

بدین ترتیب با استفاده از درخت برآمدگی و الگوریتم ژنتیکی با خصوصیات گفته شده مدلهای گویندگان ساخته شد. در مرحله تست هویت یک گوینده، بردار ویژگی حاصل از هر فریم گفتار ورودی به درخت اعمال شده و پس از پیمایش از ریشه به برگ در بر گیرنده آن، خروجی متناظر با آن برگ محاسبه گردید. این خروجی برای کلیه فریم‌های گفتار کد شناسائی ۷ رقمی گوینده محاسبه و

میانگین‌گیری شد. این میانگین در واقع بیانگرمیزان اختلاف بین مدل و گفتار ورودی می‌باشد. حال با در اختیار داشتن میانگین فوق و با توجه به آنچه که در بخش تصمیم گیری تصدیق هویت گوینده بیان نمودیم، با مقایسه این میانگین با سطح آستانه تصمیم گیری گوینده مورد نظر میتوان نسبت به رد یا قبول وی اقدام نمود.

برای مقایسه میزان کارائی روش فوق با کارائی روشهای متداول در تصدیق هویت گوینده، آزمایش دیگری صورت دادیم. در این آزمایش، از تکنیک چندی سازی برداری-k means برای مدل نمودن گویندگان استفاده کردیم. هر گوینده با استفاده از ۳۲و ۶۴ خوشه مدل گردید. در این روش به منظور تصدیق هویت یک گوینده گفتار وی با مدل گوینده ادعا شده مقایسه و فاصله بین بردار ویژگی هر فریم با مرکز ثقل نزدیکترین خوشه در مدل فوق بدست آمد و نهایتا میانگین فواصل بدست آمده بازاء کلیه فریمها در گفتار تست محاسبه گردید. این میانگین در واقع میزان اختلاف بین مدل گوینده ادعا شده و گفتار ورودی می‌باشد. با مقایسه این میانگین با سطح آستانه تصمیم گیری، میتوان نسبت به رد یا قبول گوینده تصمیم‌گیری نمود. نکته مهمی که در تصدیق هویت گوینده وجود دارد تعیین مناسب سطح آستانه تصمیم‌گیری می‌باشد. در آزمایشات فوق از روش نرخ خطای برابر و یا EER استفاده شده که این خطا نظیر محل تلاقی منحنی تغییرات خطاهای FR و FA می باشد. همانطور که در بخش بررسی خطاهای تصدیق هویت بیان نمودیم به منظور در نظر گرفتن تغییرات در محیط ضبط صدای گوینده در هنگام تست و همچنین تغییرات تدریجی صدای گوینده با گذشت زمان، در عمل بجای استفاده مستقیم از EER آنرا در یک ضریب بزرگتر ولی نردیک به یک ضرب می نمایند. در جدول ۶ بازاء مقادیر مختلفی برای این ضریب نتایج حاصل از تصدیق هویت با استفاده از درخت برآمدگی و نیز روش چندی کردن برداری با تعداد ۳۲ و ۶۴ برای حجم کتابچه کد، آورده شده است. در این جدول مشاهده می‌شود که راندمان روش شبکه عصبی درختی در بهترین حالت بازاء ضریب ۱/۱ برابر ٪۹۳/۹ و راندمان روش چندی سازی برداری بازاء همین ضریب ٪۹۶/۶و ٪۹۷/۸ بترتیب بازاء ۳۲ و ۶۴ خوشه بدست می‌آید. افزایش تعداد خوشه‌ها از ۳۲ خوشه به ۶۴ خوشه تغییر قابل توجهی در نتایج حاصل ایجاد نمی‌نماید.

جدول (٦) مقایسه راندمان روش شبکه عصبی درختی و روش چندی سازی برداری با تعیین آستانه بروش نرخ خطای برابر.

الف)- راندمان شبکه عصبی.

مقدار ضریب	شبکه عصبی درختی	
	آموزشی	آزمایشی
۰/۹	٪۹۵/۹	٪۸۰/۷
۱	٪۹۹/۵	٪۸۹/۲
۱/۱	٪۹۸/۷	٪۹۳/۹
۱/۲	٪۹۶/۰	٪۹۲/۱

ب)- راندمان چندی سازی برداری با کتاب کد ۳۲ عضوی.

مقدار ضریب	چندی سازی برداری(۳۲)	
	آموزشی	آزمایشی
۰/۹	٪۹۰/۷	٪۷۲/۵
۱	٪۹۹/۷	٪۹۰/۴
۱/۱	٪۹۸/۲	٪۹۶/۶
۱/۲	٪۹۳/۸	٪۹۵/۱

ج)- راندمان چندی سازی برداری با کتاب کد ٦٤ عضوی.

مقدار ضریب	چندی سازی برداری(۶۴)	
	آموزشی	آزمایشی
۰/۹	٪۹۷/۲	٪۷۲/۳
۱	٪۹۹/۶	٪۸۹/۰
۱/۱	٪۹۹/۱	٪۹۷/۸
۱/۲	٪۹۶/۰	٪۹۶/۳

در آزمایش دیگر بجای استفاده از EER از روش یافتن خط برازش $y = c1*(\mu-\sigma)+c2$ جهت تعیین آستانه تصمیم‌گیری استفاده شد. نتایج حاصل از تصدیق هویت گوینده با استفاده از درخت برآمدگی و چندی سازی برداری در جدول ۷ آورده شده است. در این جدول مشاهده می‌شود که راندمان روش شبکه عصبی درختی در بهترین حالت بازاء ضریب ۱/۲ برابر ٪۹۱/۰ و راندمان روش چندی‌سازی برداری بازاء همین ضریب ٪۹۶/۰ و ٪۹۶/۹ بترتیب بازاء ۳۲ و ۶۴ خوشه بدست می آید. در آزمایش دیگری سعی شد تصمیم گیری نسبت به رد یا قبول گوینده با استفاده از روش نرمالیزه نمودن گروهی صورت گیرد. برای این منظور ۵ گوینده که در فاز آموزش نزدیکترین فاصله را با هر گوی.وم.ومنده داشتند بعنوان گروه آن گوینده انتخاب شدند. برای تعیین فاصله نهایی یک گفتار با مدل هر گوینده، از فواصل گفتار ورودی با تک

تک گویندگان هم گروه گوینده مرجع میانگین‌گیری شد و این میانگین بعنوان فاصله نهایی در نظر گرفته شد. با توجه به این فرضها، نتایج تصدیق هویت با استفاده از روش نرمالیزه نمودن گروهی برای روش درخت برآمدگی و روشهای تعیین سطح آستانه نرخ خطای برابر و برازش خط در جداول ۸ و ۹ آورده شده است. در جدول ۸ مشاهده می‌شود که با استفاده از نرمال سازی گروهی و روش تعیین سطح خطای نرخ برابر در بهترین حالت بازاء ضریب ۱/۱ کارایی برابر ٪۷۱/۶ بدست می آید. همچنین با توجه به جدول ۹ نیز مشاهده می‌شود که با استفاده از نرمال سازی نتایج گروهی و روش تعیین سطح آستانه بکمک برازش خط، در بهترین حالت بازاء ضریب ۰/۹ راندمان ٪۶۷/۳ بدست می‌آید. این نتایج نشان میدهد که استفاده از روش نرخ خطای برابر یا EER در مقایسه با روش برازش خط راندمان بالاتری را نتیجه می‌دهد. این نتیجه میتواند بدین صورت قابل توضیح باشد که سطح آستانه بدست آمده در روش برازش خط با توجه به شکل ۸ مناسبترین سطح آستانه تصمیم گیری را برای آنکه خطای تصدیق هویت حداقل باشد، نتیجه نمی دهد.

جدول (۷) مقایسه روش شبکه عصبی درختی و روش چندی سازی برداری با تعیین آستانه بروش برازش خط.

الف- راندمان شبکه عصبی

مقدار ضریب	شبکه عصبی درختی	
	آموزشی	آزمایشی
۰/۹	٪۹۳/۰	٪۷۹/۲
۱	٪۹۷/۳	٪۸۴/۱
۱/۱	٪۹۸/۹	٪۸۹/۷
۱/۲	٪۹۸/۲	٪۹۱/۰

ب- راندمان چندی سازی برداری با کتاب کد ۳۲ عضوی

مقدار ضریب	چندی سازی برداری(۳۲)	
	آموزشی	آزمایشی
۰/۹	٪۸۸/۶	٪۷۵/۴
۱	٪۹۸/۸	٪۸۷/۵
۱/۱	٪۹۸/۸	٪۹۵/۱
۱/۲	٪۹۵/۲	٪۹۶/۰

ج)- راندمان چندی سازی برداری با کتاب کد ٦٤ عضوی

مقدار ضریب	چندی سازی برداری(۶۴)	
	آموزشی	آزمایشی
۰/۹	٪۹۴/۳	٪۷۵/۱
۱	٪۹۹/۴	٪۸۸/۵
۱/۱	٪۹۹/۲	٪۹۴/۶
۱/۲	٪۹۶/۹	٪۹۶/۹

جدول (۸) نتایج تصدیق هویت گوینده با استفاده از نرمال سازی گروهی و نرخ خطای برابر.

مقدار ضریب	داده های آموزشی	داده های آزمایشی
۰/۹	٪۶۴/۱	٪۶۵/۰
۱	٪۷۳/۴	٪۶۸/۳
۱/۱	٪۷۳/۲	٪۷۱/۶
۱/۲	٪۶۸/۹	٪۶۹/۰

جدول (۹) نتایج تصدیق هویت گوینده با استفاده از نرمال سازی گروهی و برازش خط.

مقدار ضریب	داده های آموزشی	داده های آزمایشی
۰/۹	٪۷۰/۷	٪۶۷/۳
۱	٪۷۲/۸	٪۶۷/۲
۱/۱	٪۶۷/۵	٪۶۷/۰
۱/۲	٪۶۰/۴	٪۶۴/۰

٤- نتیجه‌گیری

آزمایشات صورت گرفته جهت بازشناسی ارقام با تلفیق شبکه عصبی پیشگو و برنامه ریزی پویا نشان داد که افزایش بیش از حد تعداد گره ها در لایه مخفی و نیز ارائه بیش از حد داده های آموزشی به مدل در هنگام آموزش موجب گرایش بیش از حد مدل به داده های آموزشی و کاهش قدرت تعمیم پذیری آن و در نتیجه کاهش راندمان بازشناسی میگردد. در خصوص نوع ویژگیها و تعداد آنها نیز معلوم گردیدکه که افزایش تعداد ویژگیها تا حد معقول در بهبود راندمان مؤثر میباشد، ضمن آنکه اطلاعات دینامیک گفتار موجود در مشتق ضرائب کپسترال میتواند در بهبود کارائی تاثیر مثبت داشته باشد.

روش سلسله مراتبی جهت خوشه‌بندی داده‌ها در تصدیق هویت گوینده از مزیت سرعت بازیابی نسبتاً زیاد در مقایسه با دیگر روشهای خوشه‌بندی برخوردار است. مزیت دیگر آن سازگاری جهت ادغام با روشهای ژنتیکی می‌باشد. اگرچه شبکه‌های عصبی چند لایه نیز این قابلیت را دارند اما کندی سرعت آنها باعث کاهش توجه به این قابلیت آنها می‌شود. لازم به ذکر است که شبکه‌های عصبی چند لایه برای آموزش از قاعده انتشار به عقب استفاده می‌کنند که در این الگوریتم لازم است تکرارهای بسیاری جهت یافتن پارامترهای بهینه صورت گیرد، در صورتیکه همچنان که گفته شد در روش شبکه های عصبی درختی تنها در یک مرحله و با سرعتی بالا این امر میسر می‌باشد.

نتایج حاصل از این بررسی بنوعی روشنگر نکات دیگری نیز میباشد که مهمترین آنها عبارتند از:

- سرعت بازیابی خوشه‌ها در این روش سریعتر از دیگر روشهای خوشه‌بندی می‌باشد.
- زمان آموزش این روش بدلیل بکارگیری الگوریتمهای ژنتیکی که عموما زمانبر می‌باشند نسبت به سایر روشها طولانی‌تر می‌باشد.
- از مقایسه آزمایش این روش با آزمایش چندی سازی برداری معمولی چنین برمی‌آید که کارایی این روش کمتر از روش چندی سازی برداری بروش k-means می‌باشد اما مزیت آن همچنانکه پیشتر شرح داده شد سرعت بسیار بالاتر این روش در مقایسه با روشهای معمول مثل روش k-means می‌باشد. با فرض آنکه ۳۲ خوشه داشته باشیم، در خوشه‌بندیهای معمولی لازم است که ۳۲ خوشه بازای هر بردار ورودی مقایسه شوند. اما در درخت برآمدگی با توجه به شکل کامل آن لازم است بازاء هر سطح ۲ گره میانی و و چون ۳ سطح داریم در مجموع ۲×۳ گره برای یافتن برگ فعال مورد بررسی قرار گیرد و در نهایت تنها یک برگ با بردار ورودی مقایسه شود. در حالات کلی رتبه محاسبات لازم برای بازیابی خوشه در برگیرنده یک بردار ورودی در الگوریتمهای خوشه‌بندی معمولی $O(n)$ و در درخت برآمدگی $O(\log_2 n)$ می‌باشد.

در خصوص تعیین سطح آستانه نیز مشاهده شد که استفاده از روش نرخ خطای برابر یا EER در مقایسه با روش برازش خط راندمان بالاتری را نتیجه می‌دهد. لیکن روش نرخ خطای برابر نیاز به تعداد کافی و زیاد تستهای درون گویندگی با استفاده از داده های آموزشی دارد که این امر موجب می‌گردد که لازم شود گویندگان در تعداد دفعات بیشتری کد شناسایی شخصی خود را جهت آموزش سیستم بیان نمایند. از طرف دیگر روش تعیین سطح آستانه تصمیم گیری به روش برازش خط گرچه راندمان پایین تری را سبب می‌شوند، لیکن چون تعیین سطح آستانه به کمک آن تنها به آزمایشات برون گویندگی نیاز دارد، بنابراین این امکان فراهم می‌شود که گویندگان در تعداد دفعات کمتری کد شناسایی شخصی خود را بیان نمایند که این موضوع بدلیل سهولتی که برای گویندگان در مرحله آموزش فراهم می‌نماید در سیستمهای واقعی تصدیق هویت گوینده از اهمیت بالایی برخودار می‌باشد.

در خاتمه ابراز میدارد که آنچه در این مقاله ارائه گردید ماحصل فعالیتهای تحقیقاتی است که برای ایجاد یک سیستم تصدیق هویت از طریق تلفن صورت گرفته است. روشهای استفاده شده با هدف بهبود کارائی، سرعت و سهولت در انجام کار انتخاب شده و آزمایشاتی برای تعیین توانمندی این روشها صورت گرفته است. لزوما این روشها، بهترین روشها نبوده و البته این افق برای کلیه محققین در این زمینه باز است که روشها و تکنیکهای دیگر را مورد بررسی قرار داده و با تلاش خود کارائی سیستمهای تصدیق هویت گوینده را بهبود بخشند و از این طریق به امکان استفاده کارآمد تر از این سیستم ها در کاربردهائی که برای آنها متصور است یاری نمایند.

زیر نویسها

1-Speaker Verification
2-Features
3-Pitch
4-Formants
5-Linear Predictive coefficients
6-Reflection coefficients
7-Log Area Ratio Coefficients
8-Line Spectrum Pair Frequencies
9-MFCC: Mel Frequency Cepstrum Coefficients
10-LPCC: Linear Frequency Cepstrum coefficients
11-Liftering
12-Dynamic Time Warping
13-Vector Quantization
14-HMM: Hidden Markov Model
15-Second Order Statistical Measures
16-Neural Networks
17-Back Propagation
18-Vocabulary
19-Back Tracking
20-Preemphasis
21-Hamming Window
22-Minimum Variance
23-BumpTree
24-Bump
25-False Rejection
26-False Acceptance
27-Equal Error Rate
28-Intra Speaker Distance
29-Inter Speaker Distance
30-Cohort Normalization
31-Code Book

قدردانی

این پروژه تحقیقاتی در راستای طرح ملی تحقیقات به شماره NRCI357 انجام و از طرف شورای پژوهشهای علمی کشور حمایت گردیده است.

مراجع

[1] L. F. Lamel, I. L. Gauvain, "Speaker Verification Over Telephone", Speech communication, 31 (2-3), pp. 141-154, 2000.

[2] C. P. Lim, et al., "Speech Recognition using Artificial Neural Networks", Proceedings of First Conf. on Web Information Engineering, Vol. 1, pp. 419-423, 2000.

[3] W. Siew Chan, L. Chee Peng, R. Osman, "Text-Dependent Speaker Recognition using the Fuzzy ARTMAP Neural Network", Proceeding of Intelligent System Technology" Vol. 1, pp. 33-38, 2000.

[4] J. R. Deller, J. G. Proakis, and J. H. L. Hanson, Discrete-Time Processing of Speech Signals, Macmilan, NewYork, 1993.

[5] M. G. Rahim, Artificial Neural Networks for Speech Analysis/Synthesis, Chapman & Hall, University Press, Cambridge , 1994.

[6] H. Sakoe, R. Isotani, K.Yoshida, K. Iso, and T. Watanabe , "Speaker-Independent Word Recognition Using Dynamic Programming Neural Predication Model", ICASSP-89, pp. 29-32, 1989.

[7] C. G. Looney, Pattern Recognition Using Neural Networks, Oxford University Press, 1997.

[8] B. V. Williams, R. T. J. Bostock, D. Bounds, and A. Harget, "Improving Classification Performance in the Bumptree Network by Optimizing Topology with a Genetic Algorithm", First Conference on Genetic Algorithms, pp. 490-495, 1994.

[9] S. M. Omahandro, "Bumptree for Efficient Function, Constraint, and Classification Learning", Advances in Neural Information Processing System, 3, Morgan Kaufmann, 1991.

[10] S. Renals, and R. Rohwer, "Phoneme Classification Experiments Using Radial Basis Functions", International Joint Conference on Neural Network, I, pp. 461-467, Washington D. C., 1989.

[11] F. Chen, B. Miller, and M. Wagner, "Hybrid Threshold Approach in Text-Independent Speaker Verification", ICSLP-94, Yokohoma, pp. 1855-1858, 1994.

[12] S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification", IEEE Trans. on ASSP, ASSP-29, pp. 254-272, 1981.

[13] Y. Gong, and J. P. Haton, "Non-Linear Interpolation Methods for Speaker Recognition and Verification", ESCA Workshop on Automatic Speaker Recognition, Identification, and Verification, pp. 23-26, 1994.

[14] R. T. J. Bostock, Doctoral Dissertation, Dept. of Computer Science and Applied Mathematics, Aston University, Uk, 1993.