
بسم الله الرحمن الرحيم

دانشکده مهندسی برق و رباتیک

گروه الکترونیک

شناسایی گوینده مبتنی بر انتخاب ویژگی مناسب با استفاده از الگوریتم رقابت استعماری

دانشجو: محمد سلیمان پور

استاد راهنما:

دکتر حسین مروی

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ۱۳۹۳

شماره : ۱۲۴۹/آ.ت.ب
تاریخ : ۹۳/۱۱/۲۰
ویرایش : ------

بسمه تعالی

مدیریت تحصیلات تکمیلی
فرم شماره (۶)

فرم صورتجلسه دفاع پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای :

محمد سلیمان پور رشته : **برق - الکترونیک** گرایش : **سیستم**

تحت عنوان : **تشخیص گوینده بر مبنای انتخاب ویژگی های مناسب با استفاده از الگوریتم رقابت استعماری**

که در تاریخ **۹۳/۱۱/۲۰** با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح زیر است :

قبول (با درجه : بسیار خوب امتیاز ۱۸/۱۵) ☑ دفاع مجدد ☐ مردود ☐

۱ـ عالی (۲۰ ـ ۱۹)
۲ـ بسیار خوب (۱۸/۹۹ ـ ۱۸)
۳ـ خوب (۱۷/۹۹ ـ ۱۶)
۴ـ قابل قبول (۱۵/۹۹ ـ ۱۴)
۵- نمره کمتر از ۱۴ غیر قابل قبول

عضو هیأت داوران	نام ونام خانوادگی	مرتبهٔ علمی	امضـــاء
۱ـ استاد راهنما	حسین کاران	استادیار	
۲ـ استاد مشاور	—	—	
۳ـ نماینده شورای تحصیلات تکمیلی	سا سان ناصح	استادیار	
۴ـ استاد ممتحن	امیررضا مهرورزی	استادیار	
۵ ـ استاد ممتحن	سید مهدی مستوفی نائینی	دانشیار	

رئیس دانشکده :

تقدیم به:

به پدر و مادرم و مادربزرگ مهربانم

و

و همه ی آنان که شمع امید را در دل دیگران روشن نگاه داشته اند.

تشکر و قدردانی:

حمد و سپاس خداوند را که بی اذن و اراده او کلمه ای نگارده نخواهد شد. در اینجا فرصت را غنیمت می شمارم و از دکتر حسین مروی به خاطر همه ی راهنمایی ها و زحمات ایشان تشکر می کنم و همچنین از دکتر محمدرضا کرمی که بی هیچ چشم داشتی مرا در این امر یاری نمودن هم تشکر می کنم.

همچنین از برادر عزیزم تشکر می کنم که همواره حامی و الگوی من در عرصه های مختلف زندگی می باشد.

تعهد نامه

اینجانب **محمد سلیمان‌پور** دانشجوی دوره کارشناسی ارشد رشته **الکترونیک – سیستم** دانشکده **مهندسی برق و رباتیک** دانشگاه صنعتی شاهرود نویسنده پایان نامه **شناسایی گوینده مبتنی بر انتخاب ویژگی مناسب با استفاده از الگوریتم رقابت استعماری** تحت راهنمائی **دکتر حسین مروی** متعهد می‌شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافتهای آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می‌باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده:

شناسایی گوینده یکی از موضوع‌های مورد علاقه محققان در زمینه پردازش گفتار و هوش مصنوعی در چند دهه گذشته می‌باشد. در سالیان متوالی، تلاش‌های فراوانی در قسمت‌های مختلف این نوع سیستم‌ها برای بهبود کارایی آن شده است. یکی از این قسمت‌ها که می‌تواند باعث بهبود سیستم‌های شناسایی گوینده شود، قسمت انتخاب ویژگی می‌باشد. معمولاً برای برگزیدن ویژگی‌های مطلوب‌تر که باعث بهبود عملکرد سیستم می‌شود، بعد از مرحله استخراج ویژگی از انتخاب ویژگی استفاده می‌کنند. برای انجام این کار، روش‌ها و ابزارهای مختلفی وجود دارد که در این پایان نامه از خوشه‌بندی برای یافتن بردارهای ویژگی که بیشترین تشابه را دارند استفاده شده است و در نتیجه این بردارهای مشابه، بیشترین خصوصیات مربوط به مجرای صوتی فرد را مشخص می‌کند. این کار باعث همگرایی بیشتر برای ساخت مدل هر فرد و یا مرزهای تصمیم‌گیری بین افراد می‌شود.

در این پایان نامه دو روش پیشنهاد شده است که در هر دو روش از الگوریتم رقابت استعماری جهت خوشه‌بندی بردارهای ویژگی استفاده شده است و همچنین دو الگوریتم دیگر خوشه‌بندی K-Means و خوشه‌بندی PSO برای مقایسه کارایی الگوریتم رقابت استعماری بکار گرفته شده است. در نهایت، نتایج دو روش پیشنهادی را با دو روش SVM و ELM بهینه سازی شده را از نظر نرخ بازشناسی و مدت زمان استفاده شده از داده‌گان مقایسه کرده‌ایم. بررسی نتایج نشان می‌دهد که نرخ بازشناسی در روش‌های پیشنهادی بهبود یافته است.

در این پروژه ما از داده‌گان ELSDSR استفاده کرده‌ایم. به دلیل ماهیت داده‌گان موجود، در این پایان نامه یک سیستم شناسایی گوینده مستقل از متن خواهیم داشت.

کلیدواژه: شناسایی گوینده، MFCC، انتخاب ویژگی، الگوریتم رقابت استعماری، خوشه بندی

فهرست مطالب

فهرست اشکال

فهرست نمودار

فهرست جداول

فصل اول :

مقدمه‌ای بر سیستم‌های پردازش گفتار

و

شناسایی گوینده

۱-۱- مقدمه

امروزه اهمیت سیستم‌های تشخیص گوینده و بازشناسایی گفتار بر کسی پوشیده نیست و تحقیقات بسیار زیادی در این زمینه توسط محققان به انجام رسیده است. همین امر گواه بر اهمیت بالا و کاربرد وسیع آن در زمینه‌های گوناگون نظیر امنیت، جرم شناسی، طبقه‌بندی فایل‌های صوتی و غیره می‌باشد. امنیت یکی از بارزترین و مهم‌ترین پارامترها در هر موسسه و سازمان می‌باشد که هر یک به اندازه وسعت کارشان برای آن هزینه می‌کنند. شناسه‌های کاربری، رمزهای عبور، کارت‌های مغناطیسی از جمله روش‌های معمول برای شناسایی افراد در سازمان‌ها می‌باشد. همچنین امنیت پایین این سیستم-ها باعث شده است که در دهه‌های اخیر، موسسات و سازمان‌های بزرگ در پی یافتن راه بهتری برای جایگزین آن بجای روش‌های معمول ذکر شده باشند. روش‌های زیست سنجی[۱] یکی از این راه‌ها می‌باشد که شامل بازشناسایی افراد از طریق صوت، اثرانگشت، چهره، عنبیه و شبکیه چشم و غیره می‌باشد. بازشناسی افراد از طریق صوت به علت هزینه پایین و پیچیدگی کمتر می‌تواند راه مناسبی برای این کار باشد.

سیگنال‌های صوتی به دلیل داشتن ویژگی‌های منحصربه‌فرد می‌تواند به بازشناسی افراد کمک کند. به‌طور کلی ویژگی‌های گفتار افراد به دو دسته تقسیم می‌شود. دسته اول شامل ویژگی‌های مانند ماهیچه‌های صوتی، اندازه و شکل گلو، تارهای صوتی، مجرای دهان، مجرای بینی و غیره می‌باشد و از طرف دیگر در دسته دوم می‌توان به ویژگی‌های نظیر سطح تحصیلات گوینده، نحوه ادای صحبت، حالت صحبت و غیره اشاره کرد. محققان و دانشمندان معمولاً به دنبال ویژگی‌هایی هستند که علی رغم داشتن کارایی خوب، سادگی در استفاده از آن را هم دارا باشد. از این رو، در اکثر سیستم‌های تشخیص افراد از

---

[۱] Biometric

طریق سیگنال‌های گفتار از ویژگی دسته اول استفاده می‌کنند؛ چون هم این ویژگی‌ها از قابلیت اندازه-گیری بیشتری برخوردار هستند و هم در پیاده‌سازی راحت‌تر می‌باشند. بدیهی است استفاده از ویژگی-های دسته دوم به پیچیدگی سیستم افزوده، اما کارایی آن را افزایش می‌دهد.

۱-۲- تاریخچه

برای اینکه یک نگاه اجمالی بر روی تلاش‌های که برای پدید آمدن سیستم‌های پردازش گفتار انجام شد، داشته باشیم مروری بر تاریخچه این نوع سیستم‌ها خواهیم داشت. اولین مطالعات بر روی سیستم‌های پردازش گفتار اعم از سیستم‌های تشخیص گوینده و بازشناسی گفتار در اویل دهه ۱۹۵۰ میلادی شروع شد و تا به امروز تحقیقات زیادی بر روی این نوع سیستم‌ها انجام شده است. جرقه طراحی و ساخت اولین سیستم تشخیص گوینده در سال ۱۹۶۰ توسط پروزونسکی[۱] در آزمایشگاه بل با استفاده از طیف نگاشت[۲] و بانک‌های فیلتر زده شد[1]. در سیستم تشخیص گوینده مبتنی بر مدل HMM-GMM، فرد مورد نظر می‌تواند بدون هیچ محدودیتی از یک گفتار به عنوان رمز عبور استفاده کند؛ البته این رمز قبلاً در مرحله آموزش چندین بار ادا می‌شود تا مدل مورد نظر آن فرد ساخته شود[2]. مولاو و همکارانش در سال ۲۰۰۱ روشی را ارائه کردن که در آن ضرایب کپسترال مل[۳] را به‌طور مستقیم از طیف توان یک سیگنال صحبت استنتاج می‌کند[3]. این روش هم اکنون یکی از مهم‌ترین روش‌ها می‌باشد که در استخراج ویژگی از سیگنال‌های گفتار از آن استفاده می‌کنند.

شبکه‌های عصبی مصنوعی تاریخ طولانی مدتی در سیستم‌های شناسایی گفتار دارند و معمولاً در ترکیب با مدل مخفی مارکوف بکار رفته‌اند[4]. این فنّاوری در سال ۱۹۵۰ معرفی شده است اما تا سال ۱۹۸۰ نتوانسته بود کارایی خود را در سیستم‌های بازشناسی گفتار نشان دهد. در این دهه با

---

۱ Pruzanky

۲ Spectrogram

۳ Mel-Frequency Cepstral Coefficient or MFCC

پیدایش پردازنده‌های موازی، استفاده از شبکه‌های عصبی را که الهام گرفته از سیستم‌های زیستی انسان می‌باشد را ممکن کرده است. در این دوره اغلب از شبکه‌های عصبی برای شناسایی چند کلمه و یا چند آوا استفاده می‌شده است و برای اجتناب کردن از تغییرات ناگهانی سیگنال‌های گفتار از ترکیب مدل مخفی مارکوف و شبکه‌های عصبی استفاده کرده‌اند[5].

۱-۳- عوامل موثر در سیستم‌های پردازش گفتار

اگر چه در دهه‌های اخیر دانشمندان و محققان به نتایج مطلوبی در زمینه پردازش گفتار رسیده-اند، اما به دلیل پیچیدگی‌های موجود در قسمت شبیه‌سازی مدل و ویژگی صوتی انسان و یا استخراج ویژگی کاملا منحصربه‌فرد، همواره احتمال بروز خطا وجود داشته است. عوامل زیادی در به وجود آمدن سیستم‌ها و الگوریتم‌های مختلف پردازش صوت اعم از سیستم‌های تشخیص گوینده و بازشناسی گفتار موثر می‌باشد که هر کدام از الگوریتم‌های ابداع شده در یک شرایط خاص بهترین عملکرد را دارا می-باشند. به دلیل غیرخطی بودن خصوصیات صوتی افراد و دلایل نظیر جنسیت، حالت صحبت، بلندی گفتار، میزان شمرده بودن و به تناسب سرعت ادا گفتار نتایج مختلفی در آزمایش‌های مختلف بدست آمده است[6]. در این قسمت به طور مختصر به ابعاد مختلف شرایط تست می‌پردازیم.

۱-۳-۱- گفتار نویزی و غیر نویزی

یکی از جنبه‌های مهم در دقت سیستم‌های پردازش صوت وجود نویزهای محیطی می‌باشد که در هنگام صحبت کردن می‌تواند سیگنال اصلی را تضعیف کند. در اینجا می‌توان نویزهای زیادی را نام برد که می‌تواند در دقت تشخیص سیستم‌های پردازش گفتار تأثیر منفی داشته باشد نظیر صدای صحبت کردن دیگر افراد، راه رفتن، عبور وسایل نقلیه، اصوات موجودات مختلف و در مجموع صدای هر آنچه به غیر از گفتار اصلی. واضح است که وقتی ما این عوامل را به سیگنال اصلی اضافه کنیم دقت سیستم

کاهش پیدا می‌کند؛ اما گفتارهای غیر نویزی در محیط آزمایشگاهی و بدون هیچ سیگنال مزاحمی آماده و تهیه شده است و به همین خاطر کار تشخیص را راحت‌تر می‌کند.

۱-۳-۲- گفتار تک کلمه‌ای و جمله‌ای

چگونگی ادا کردن گفتار می‌تواند امری مهم در عملکرد یک سیستم پردازش صوت باشد. گفتار می‌تواند به صورت یک کلمه بیان شود، در این روش پیچیدگی‌های آن کمتر و دقت سیستم بیشتر می‌باشد. از طرف دیگر، گفتارهایی که به صورت جمله ادا شدن یا همان گفتارهای پیوسته مثل صحبت کردن عادی بین دو فرد به دلیل حذف برخی اصوات و نامشخص بودن مرز بین کلمات کار تشخیص را سخت می‌کنند. وقتی که از این نوع گفتار برای بازشناسی استفاده می‌شود، دقت این سیستم‌ها را پایین می‌آورد. البته برای کم کردن پیچیدگی در گفتارهای پیوسته، از سکوت‌های مصنوعی بین کلمات در هر جمله استفاده می‌کنند[7].

۱-۳-۳- اندازه و نوع پایگاه داده[۱]

یکی دیگر از عوامل موثر در دقت سیستم‌های تشخیص صوتی، اندازه و نوع پایگاه داده‌ها می‌باشد. وقتی می‌خواهیم یک کلمه را از بین یک مجموعه کوچک بازشناسی کنیم، کار راحت‌تری پیش رو خواهیم داشت. در مقابل وقتی ما از مجموعه بزرگ‌تری در امر بازشناسی استفاده می‌کنیم، به دلیل افزایش شباهت‌ها بین گفتارها، در دقت عملکرد سیستم‌ها تأثیر منفی خواهد گذاشت. به طور مثال، وجود گفتارهای مشابه مثل سنا، ثنا و صناع در بازشناسی گفتار و همچنین استفاده افراد با لهجه‌های مختلف از یک زبان یا همگی با یک لهجه واحد می‌تواند از دیگر عواملی باشد که در عملکرد سیستم تأثیرگذار است[7].

---

[۱] Database

۱-۳-۴- گفتار پیش آماده یا فی‌البداهه

گفتار پیش آماده به گفتاری اطلاق می‌شود که در آن افراد متنی که از قبل تهیه شده است را با یک نماهنگ ثابت ادا می‌کنند در این بین کمتر فرد از اصوات حشو نظیر آ، اوم، اِ و غیره استفاده می‌کند. در صورتی که در متن‌های فی‌البداهه افراد یک سری اصوات حشو یا سکوت برای ادا کردن گفتار از آن استفاده می‌کنند که معمولاً در زمان فکر کردن افراد درباره چیزی که می‌خواهند ادا کنند رخ می‌دهد.

۱-۳-۵- وابسته بودن یا مستقل بودن از متن

سیستم‌های وابسته به متن، سیستم‌هایی هستند که در آن باید گفتارهایی در قسمت تست ادا شوند که از قبل به سیستم آموزش داده شده است. به همین خاطر، دقت در این نوع سیستم‌ها افزایش قابل قبولی دارد و دیگر مزیت این سیستم نیاز کم آن به داده‌های آموزش می‌باشد. اگر چه در حال حاضر این نوع سیستم‌ها کارایی خاصی ندارند ولی در بعضی از سیستم‌های تصدیق گوینده استفاده می‌شود. در سیستم‌های وابسته به متن معمولاً از ویژگی نظیر انرژی، تعداد عبور از صفر، دامنه و غیره استفاده می‌شود که خصوصیات همان سیگنال صوتی را نشان می‌دهد. از طرف دیگر سیستم‌هایی وجود دارند که به متن گوینده وابسته نمی‌باشد یعنی در زمان تست شما می‌توانید از هر نوع گفتار با رعایت بعضی از شرایط در هنگام آموزش، استفاده کنید. این نوع سیستم‌ها معمولاً از ویژگی‌هایی که خصوصیات افراد را در نظر می‌گیرد همانند MFCC استفاده می‌شود. این ویژگی نه تنها توزیع فرکانسی مشخص کننده صدا را حمل می‌کند، بلکه عملکرد اندازه و شکل مجرای گلو[۱] و دهانه نای[۲] هر فرد را مشخص می‌کند[۵]. در این نوع سیستم علاوه بر نیاز بیشتر به داده‌های آموزش به داده‌های بیشتری برای تست

---

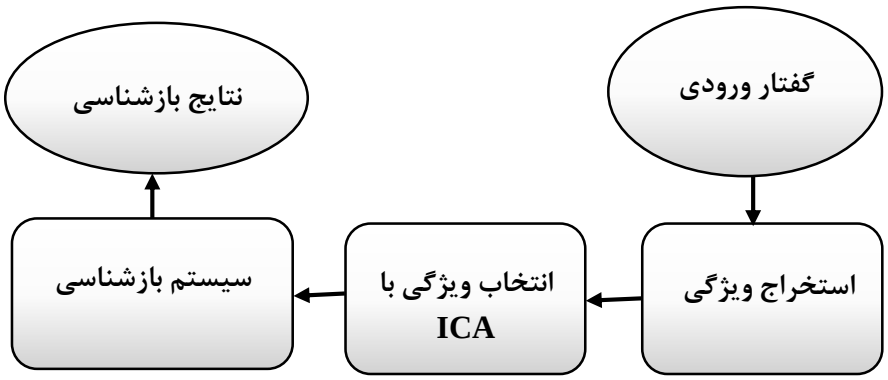
[۱] Vocal Tract

[۲] Glottal Source

هم نیاز می‌باشد و همچنین از دقت بازشناسی کمتری نسبت به سیستم وابسته به متن برخوردار می‌باشد. اگر چه وابسته بودن و یا مستقل بودن را می‌توان به عنوان انواع سیستم‌های پردازش گفتار هم اطلاق کرد.

۱-۴- هدف پایان‌نامه

سیگنال‌های گفتار دارای ویژگی‌های مختلفی هستند که با استخراج دقیق این ویژگی‌ها می‌توان باعث بهبود سیستم‌های پردازش گفتار شد. روش‌های مرسوم نظیر MFCC، انرژی، فرکانس‌های فرمنت، فرکانس‌های گام و اخیرا اطلاعات فاز را می‌توان به عنوان ویژگی در سیستم‌های پردازش گفتار شناخت[8]. بنابراین پس از استخراج ویژگی، یکی از مهم‌ترین و پیچیده‌ترین مسائل در حوزه بازشناسی الگو مسئله انتخاب ویژگی است. هدف از انتخاب ویژگی، انتخاب یک زیرمجموعه از مجموعه ویژگی‌ها است، به‌طوری که تعداد ویژگی‌های انتخاب شده کمترین و دقت بازشناسی الگو بیشترین مقدار را داشته باشد. در شکل (۱-۱) محل قرارگیری بلوک دیاگرام انتخاب ویژگی نشان داده شده است.



شکل (۳-۱) بلوک دیاگرام کلی سیستم‌های پردازش گفتار

از ویژگی‌های ذکر شده، ویژگی‌های همانند MFCC، LPCC و غیره خصوصیات مجرای صوتی هر فرد را نشان می‌دهد. بنابراین در این پایان نامه قصد داریم روشی را ارائه کنیم که در آن بتوانیم از بردارهای ویژگی مشابه مجرای صوتی هر فرد برای مدل‌سازی افراد یا آموزش جهت تشخیص افراد،

استفاده کنیم؛ که این خصوصیات مشابه (پر تکرار) با کمک الگوریتم رقابت استعماری بدست می‌آید. به طور کلی این کار باعث حذف یک سری از بردارهای ویژگی کم تکرار که توسط مجرای صوتی و دیگر ماهیچه‌های صوتی تولید می‌شود، می‌گردد. این نوع انتخاب ویژگی باعث همگرایی بیشتر در مدل‌سازی در جهت شناسایی افراد می‌شود.

۱-۵- ساختار پایان‌نامه

همان‌طور که قبلا هم بیان شده است، این فصل مقدمه‌ای در مورد سیستم‌های پردازش گفتار و شناسایی گوینده ارائه شده است. در فصل دوم، انواع سیستم‌های شناسایی گوینده معرفی کرده‌ایم و سپس به معرفی بلوک‌های مختلف سیستم‌های تشخیص هویت پرداخته‌ایم. همچنین در این فصل انواع روش‌های استخراج ویژگی را توضیح داده‌ایم.

در فصل سوم، دو نمونه از الگوریتم‌های تکاملی را که در این پایان‌نامه از آن استفاده کرده‌ایم به‌طور کامل شرح داده‌ایم. یکی از این الگوریتم‌های الگوریتم رقابت استعماری می‌باشد که الگوریتم اصلی این پروژه می‌باشد و الگوریتم دیگری که در این فصل به معرفی آن خواهیم پرداخت، الگوریتم بهینه‌سازی گروه ذرات می‌باشد. الگوریتم بهینه‌سازی ذرات، برای بررسی عملکرد الگوریتم اصلی این پروژه استفاده شده است. فصل چهارم به معرفی اجمالی روش‌های خوشه‌بندی می‌پردازیم و سپس به چند نمونه از الگوریتم‌های خوشه‌بندی اشاره خواهیم کرد.

در فصل پنجم، دو روش پیشنهادی که در این پایان‌نامه از استفاده شده را شرح خواهیم داد و در فصل آخر به نتایج الگوریتم‌های پیشنهادی کارهای آینده خواهیم پرداخت.

فصل دوم:

انواع سیستم‌های شناسایی گوینده

۲-۱- مقدمه

سیستم بازشناسی گوینده به سیستمی اطلاق می‌شود که از طریق ویژگی صوتی بتواند افراد را از یک دیگر تشخیص بدهد. سامانه شناسایی گوینده می‌تواند در کاربردهای مختلف امنیتی و کنترل دسترسی، به تنهایی یا در کنار دیگر روش‌های امنیتی مورد استفاده قرار گیرند. این نوع سیستم‌ها از روی صدا با توجه به اینکه صدای فرد همواره همراه وی بوده و معایبی مانند گم شدن و دزدیده شدن را ندارد، می‌توان بکار برد و همچنین می‌تواند بدون حضور فیزیکی و از راه دور مورد استفاده قرار گیرند. تشخیص یا شناسایی گوینده[۱] و تصدیق گوینده[۲] دو دسته از سیستم‌های بازشناسی گوینده محسوب می‌شوند که هر کدام از این دو دسته کاربردهای خاص خود را دارند[9].

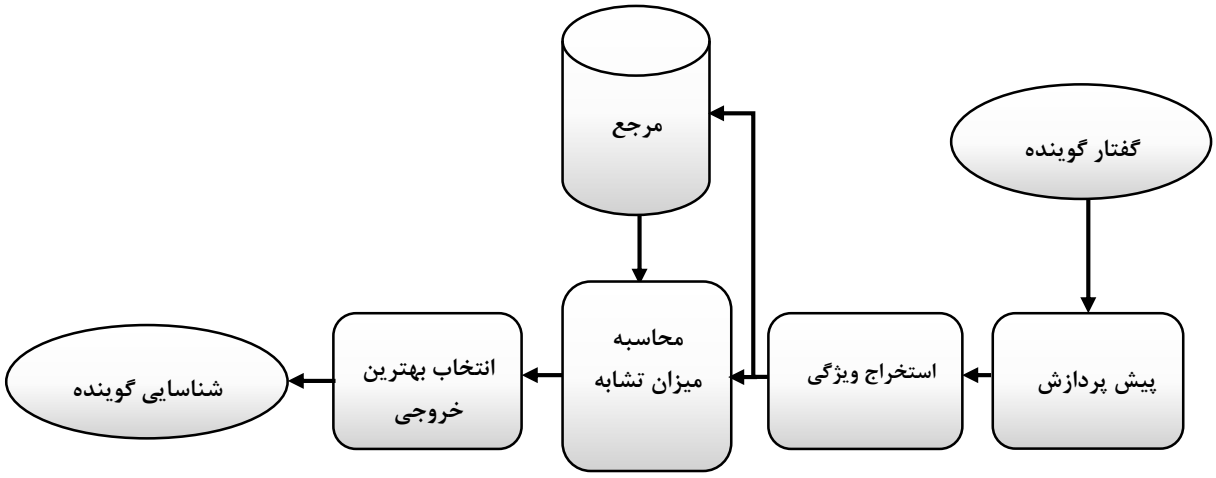
۲-۲- سیستم شناسایی گوینده

در شناسایی گوینده، سیستم تلاش می‌کند فرد مورد نظر را از بین مجموعه‌ای از افراد پیدا کند. همان طور که در شکل (۲-۱) مشخص می‌باشد در این نوع سیستم‌ها گفتار ورودی دریافت می‌شود و پس از استخراج ویژگی میزان شباهت بین گفتار بیان شده و مدل‌های مرجع محاسبه می‌شود و در نتیجه هر کدام از مدل‌های مرجع که بیشترین شباهت را با گفتار ادا شده دارا می‌باشد به عنوان فرد مورد نظر شناخته می‌شود[9].

---

[۱] Speaker Recognition or Identification

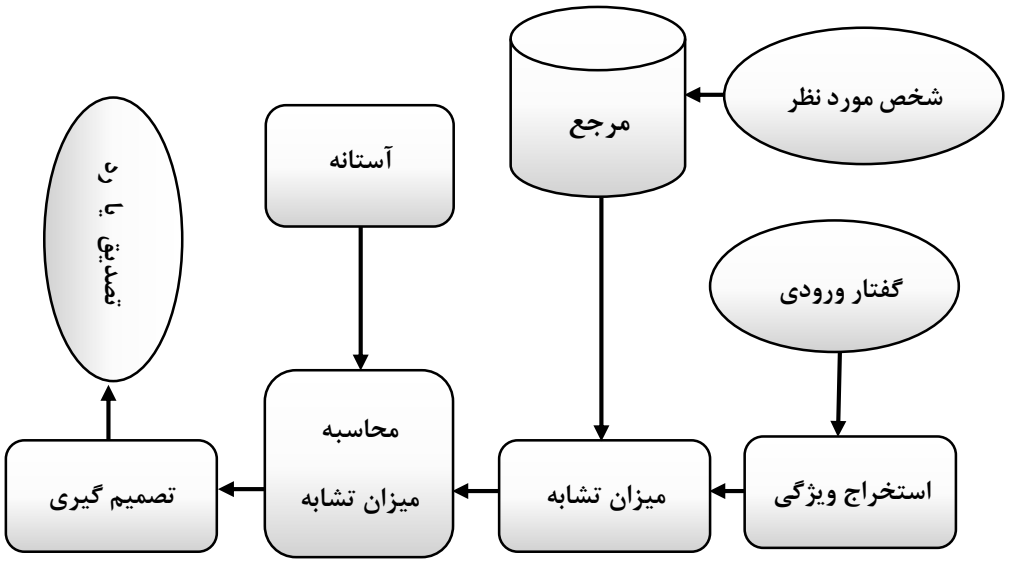
[۲] Speaker Authentication



شکل (۲-۱) بلوک دیاگرام سیستم‌های شناسایی گوینده

۲-۳- سیستم تصدیق گوینده

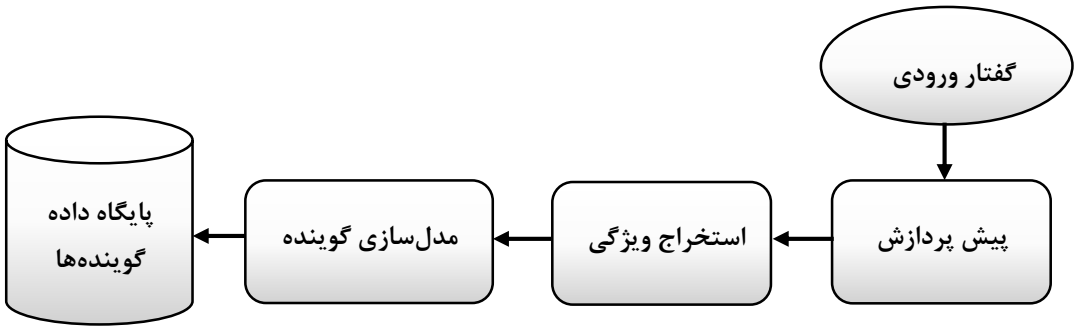
سیستم‌های تصدیق گوینده، درستی یا نادرستی ادعای فرد مورد نظر را مبنی بر معرفی خود به عنوان یکی از اعضا مرجع را مشخص می‌کند. اما بر خلاف سیستم‌های شناسایی گوینده که تنها یک ورودی دارد در این سیستم علاوه بر گفتار ورودی، شخص مورد نظر خود را به عنوان یکی از افراد مرجع به سیستم معرفی می‌کند. حال سیستم میزان تشابه سیگنال صوتی ورودی و ویژگی فرد مورد ادعا شده را مورد بررسی قرار می‌دهد و سپس در مورد آن تصمیم گیری خواهد شد. بلوک دیاگرام سیستم تصدیق گوینده را می‌توان در شکل (۲-۲) مشاهده کرد.



شکل (۲-۲) بلوک دیاگرام سیستم‌های تصدیق گوینده

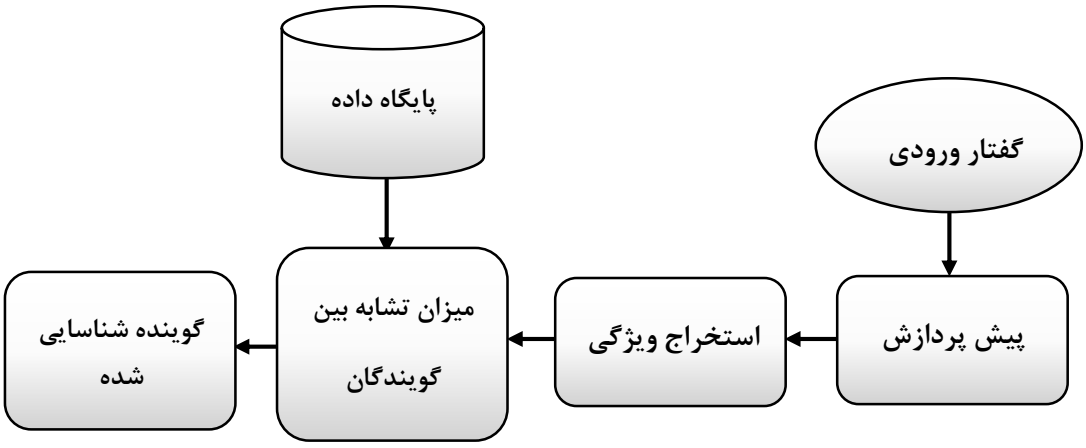
۲-۴- قسمت‌های مختلف سیستم‌های بازشناسایی گوینده

هر سیستم بازشناسایی گوینده به طور کلی از دو قسمت آموزش و تست تشکیل شده است. معمولاً در مرحله آموزش از گفتار ورودی پس از انجام پیش پردازش‌های لازم، ویژگی استخراج می‌کنند. بعد از بدست آمدن ویژگی‌های مناسب شروع به ساخت مدل هر یک از گوینده‌گان می‌کنیم، به‌طوری که بتوان گوینده‌های مختلف را از هم تمایز داد. در نهایت این مدل‌ها را در جایی به نام پایگاه داده یا مرجع ذخیره می‌کنیم. تقریبا این روند برای تمام سیستم‌های بازشناسایی گوینده یکسان می‌باشد که با روش‌های مختلف این کار انجام می‌دهند. همان طور که ملاحظه می‌کنید، شکل (۲-۳) نمودار بلوکی فرایند آموزش را نشان می‌دهد. البته در روش‌هایی همانند شبکه‌های عصبی، قسمتی به عنوان مدل سازی وجود ندارد و مجزا سازی بین افراد با استفاده از وزن‌ها صورت می‌گیرد.



شکل (۲-۳) فرایند آموزش در غالب سیستم‌های بازشناسایی گوینده

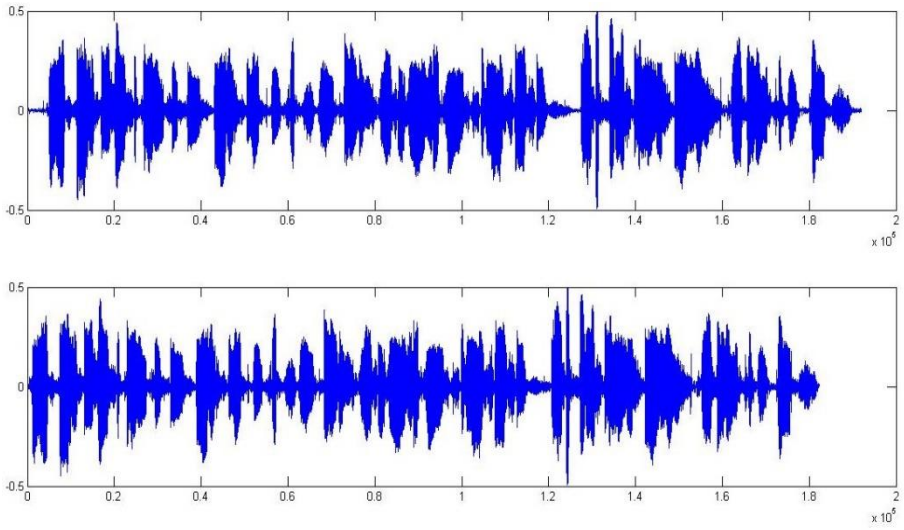
در قسمت تست هم همانند قسمت آموزش بعد از پیش پردازش‌های لازم شروع به استخراج ویژگی می‌کنیم و سپس ویژگی استخراج شده را با مدل‌های پایگاه داده مقایسه می‌کنیم. پس از حصول نتیجه، گوینده‌ای که مدل آن بیشترین شباهت را با ویژگی‌های استخراج شده را داشته باشد به عنوان گوینده مورد نظر انتخاب می‌کنیم. همچنین می‌توانیم نمودار بلوکی مورد نظر را در شکل (۲-۴) مشاهده کنید.



شکل (۲-۴) نمودار بلوکی قسمت آزمایش یا تست

۲-۴-۱- پیش پردازش[۱]

در اغلب سیستم‌های تطبیق الگو قبل از استخراج ویژگی یک سری فرایند به عنوان پیش پردازش انجام می‌شود که باعث بهبود در کارایی سیستم می‌شود. در سیستم‌های پردازش گفتار، فرایندی مانند نمونه برداری[۲]، حذف قسمت‌های سکوت، پنجره گذاری، حذف نویز و غیره از جمله کارهایی است که در قسمت پیش پردازش انجام می‌شود. در شکل (۲-۵) حالت حذف سکوت را که یکی از این فرایند می‌باشد را نشان می‌دهد. همان طور که در شکل زیر مشاهده می‌کنید با استفاده از محاسبه انرژی هر پنجره بر روی سیگنال اصلی و تعریف یک آستانه انرژی برای بیان حالت سکوت و غیر سکوت، می‌توان قسمت‌های سکوت را حذف کرد. همان طور که در شکل ملاحظه می‌کنید، حذف قسمت‌های سکوت در شروع و پایان سیگنال گفتار زیر به وضوح مشخص می‌باشد. به طور کلی، پیش پردازش باعث کمتر شدن نمونه‌ها و پیچیدگی عملیات محاسباتی هم می‌شود.



شکل (۲-۵) نمایش سیگنال اصلی و سیگنال پس از حذف سکوت

[۱] Per-Processing

[۲] Down Sampling

۲-۴-۲- استخراج ویژگی

استخراج ویژگی یکی از قسمت‌های است که در سال‌های بعد از پیدایش سیستم‌های پردازش گفتار بیشتر مورد توجه قرار گرفته است و محققان زیادی در این زمینه فعالیت کرده‌اند[10]. با وجود همه پیشرفت‌های خوبی که در این حیطه کسب شده است، اما هنوز ویژگی معرفی نشده است که تمام خصوصیات گفتاری یک فرد را باهم در بر بگیرد. هدف اصلی استخراج ویژگی، نمایش مجموعه‌ای از ویژگی‌ها سیگنال گفتار که بیان کننده خصوصیات گفتاری هر فرد می‌باشد. ویژگی‌ها در پردازش گفتار به دو دسته زمان کوتاه[۱] و زمان بلند[۲] تقسیم بندی شده است که در قسمت زمان کوتاه هموار سیگنال اصلی را با استفاده از قاب‌های مختلف، به قسمت‌های کوچک‌تر تبدیل می‌کنند. معمولاً این قاب‌های بین ۲۰ تا ۳۰ میلی ثانیه می‌باشد و با فاصله ۱۰ تا ۱۵ میلی ثانیه آن را بر روی سیگنال اصلی جا به‌جا می‌کنند. شکل (۲-۶) نمونه‌ای از قاب‌بندی توضیح داده شده را به نمایش گذاشته است. این نوع ویژگی-ها، خصوصیات ساختار فیزیکی مسیر گفتاری هر فرد را نشان می‌دهد که در بیشتر سیستم‌های بازشناسایی گفتار از این نوع ویژگی استفاده شده است. برای مثال، ویژگی‌های نظیر ضرایب کپسترال فرکانسی مل، ضرایب کپسترال پیش‌گویی خطی[۳]، ضرایب پیشگویی ادراکی خطی [۴]، انرژی و غیره می‌باشد. در حالی که ویژگی‌های زمان کوتاه خصوصیات ساختار فیزیکی مسیر گفتاری هر فرد را بیان می‌کند، ویژگی‌های زمان بلند دنبال خصوصیات ادراکی، آوایی، لغوی و غیره می‌باشند. تکرار بعضی کلمات و تیکه کلام، ادا کردن نوع خاصی از کلمات و تغییر لحن می‌تواند از جمله مثال‌های بارز این دسته باشد. برای داشتن دقت بازشناسی خوب باید هر دو نوع ویژگی را در نظر بگیریم اما به دلیل یک

---

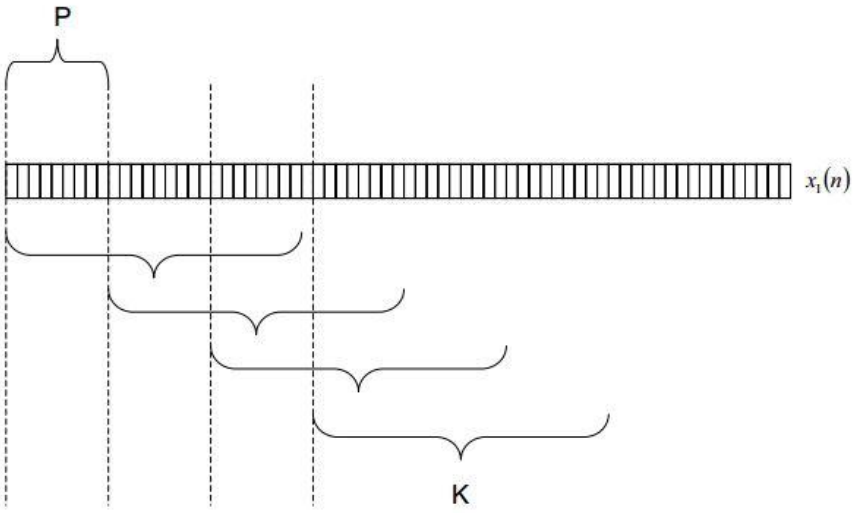
[۱] Short Term

[۲] Long Term

[۱] Linear Prediction Cepstral Coefficient

[۲] Perceptual Linear Predictive

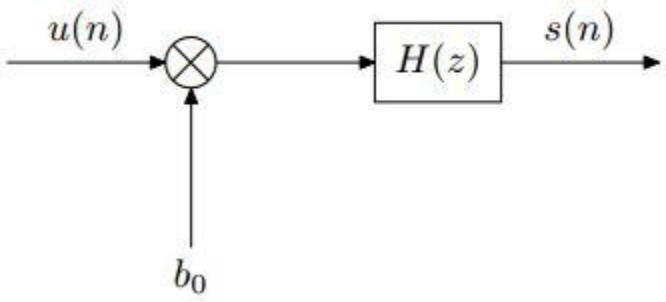
سری پیچیدگی‌ها که در قسمت استخراج ویژگی‌های زمان بلند وجود دارد و همچنین نیاز به آموزش زیاد این نوع ویژگی‌ها، فقط از ویژگی استخراج شده زمان کوتاه استفاده می‌کنیم.



شکل (۲-۶) نمایش نحوه قاب‌بندی هم‌پوشان

۲-۴-۲-۱- ضرایب پیش‌گویی خطی

ایده اصلی پنهان شده در ویژگی پیش‌گویی خطی، استخراج پارامترهای مجرای صوتی انسان که مدلی از این مجرای صوتی در شکل (۲-۷) نمایش داده شده است، می‌باشد. از این رو، برای محاسبه پارامترهایی نظیر دوره تناوب اصلی و فرکانس‌های فرمنت از این نوع ویژگی استفاده می‌شود. اگر بخواهیم به نقطه قوت این نوع استخراج ویژگی بپردازیم، توان محاسبه نسبتاً دقیق پارامترها و همچنین سرعت بالای آن در استخراج ویژگی می‌توان اشاره کرد و در مقابل نقطه ضعف این نوع استخراج ویژگی‌ها، تعیین مناسب مجموع ضرایب پیشگویی a_k از روی سیگنال صحبت به طور مستقیم برای داشتن یک تخمین مناسب می‌باشد[12][11].



شکل (۲-۷) مدل مجرای صوتی انسان

اگر فرض کنیم $s(n)$، سیگنال گفتار در نمونه n ام باشد؛ می‌توان با استفاده از ترکیب خطی P تا نمونه قبلی گفتار $s(n)$ را به صورت زیر مدل سازی کرد:

$$s(n)=b_0u(n)+a_1s(n-1)+a_2s(n-2)+\cdots+a_ps(n-p) \qquad (۱-۲)$$

همچنین در شکل دیگر خواهیم داشت:

$$s(n)=b_0\,u(n)+\sum_{i=1}^{p}a_i s(n-i) \qquad (۲-۲)$$

در جایی که $u(n)$ سیگنال تحریک نرمالیزه شده، b_0 بهره سیگنال تحریک[۱] و ضرایب $a_1,a_2,\ldots,a_p$ وزن‌هایی برای نمونه گفتار قبلی می‌باشد. البته تمام این ضرایب در همه قاب‌ها یکسان در نظر گرفته شده است. همچنین یکی دیگر از راه‌ها برای نمایش این معادلات، بردن فرمول‌های بالا به حوزه Z می‌باشد:

$$S(z)=b_0U(z)+\sum_{i=1}^{p}a_iS(z)z^{i} \qquad (۳-۲)$$

و همچنین برای تابع تبدیل خواهیم داشت:

[۱] Excitation Signal

$$H(z)=\frac{S(z)}{U(z)}=\frac{b_0}{1-\sum_{i=1}^{p}a_i z^{-i}} \qquad (۴-۲)$$

$H(z)$ در معادله (۲-۴) یک فیلتر تمام قطب می‌باشد که می‌تواند یک تقریب خوبی از مجرای صوتی انسان باشد. هدف از LPC بدست آوردن ضرایب فیلتر $a_1, a_2, \ldots, a_p$ برای پیش‌گویی سیگنال گفتار می‌باشد که در این رابطه هر قدر P بیشتر باشد، عمل پیش‌گویی دقیق‌تر انجام می‌پذیرد. البته عدد مناسب و معمول برای P بین ۴ تا ۱۶ می‌باشد که دو تا چهار فرمنت را در این بین در بر می‌گیرد؛ و همچنین عددی خارج از این بازه توصیه نمی‌شود زیرا عددی کمتر از این بازه باعث پیش‌گویی ضعیف و بیشتر از این بازه، باعث پیچیدگی بالا و افزایش هزینه محاسباتی بدون هیچ نتیجه بهتری می‌شود[12].

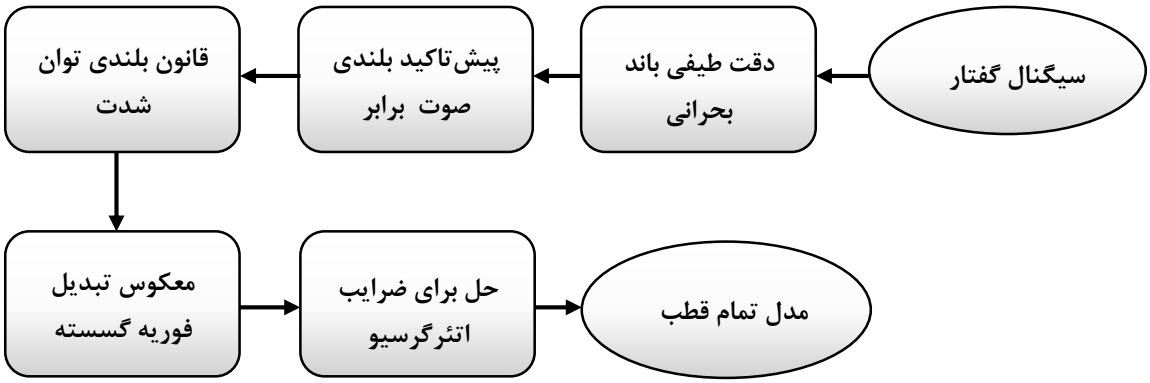
همه این محاسبات زمانی مورد استفاده قرار می‌گیرد که فرض ایستان بودن برای سیگنال گفتار در نظر گرفته شده باشد که برای این کار از فن قاب گذاری، همان طور که در قسمت قبل توضیح داده شده است، استفاده می‌شود.

۲-۴-۲-۲- ضرایب پیش‌گویی خطی ادراکی

همان طور که از قبل بیان شده است روش‌های LPC و PLP، از مدل تمام قطب اتورگرسیو[۱] برای تخمین طیف توان ترم کوتاه گفتار استفاده می‌کنند. اگرچه همان طور که Hermansky اشاره کرده است، مدل تمام قطب LPC با ادراک شنوایی انسان به خاطر در نظر نگرفتن دقت فرکانسی غیر یکنواخت و همچنین دقت شدت شنوایی انسان، سازگار نیست. از این رو، PLP این مشکل را با بکار بردن مدل تمام قطب طیف شنوایی مرتفع کرده است. طیف شنوایی طوری طراحی شده است که بتواند یک تخمینی از نرخ میانگین تحریک شنوایی بافت عصبی انسان باشد[11].

[۱] Autoregressive

در روش PLP، چندین مشخصات شناخته شده شنوایی انسان توسط تقریب‌های عملی و کاربردی شبیه سازی شده است و همچنین نتایج طیف شنوایی توسط مدل اتورگرسیو تمام قطب تقریب زده شده است. یک بلوک دیاگرام از چگونگی بدست آوردن PLP در شکل (۲-۸) نمایش داده شده است.



شکل (۲-۸) بلوک دیاگرام چگونگی بدست آوردن PLP [11]

- **آنالیز طیفی**

بخش‌های گفتار توسط پنجره همینگ زیر وزن دار شده است جایی که N طول پنجره را مشخص می‌کند. معمولاً طول نوعی پنجره‌های در فرانید بدست آوردن PLP ۲۰ میلی ثانیه می‌باشد. به‌وسیله DFT، بخش‌های گفتار قسمت بندی شده را به حوزه فرکانس انتقال می‌دهیم و سپس اجزای حقیقی و موهومی ترم کوتاه طیف گفتار به توان دو رسیده و باهم جمع می‌کنیم تا طیف توان زمان کوتاه همانند فرمول زیر بدست بیاید.

$$P(w) = \mathrm{Re}[S(w)]^2 + \mathrm{Im}[S(w)]^2 \qquad (۵-۲)$$

- **دقت طیفی باند بحرانی**

$P(w)$ از فرمول زیر بدست آمده است.

$$\Omega(w) = 6\ln\{w/1200\pi + [(w/1200)^2 + 1]^{0.5}\} \qquad (۶-۲)$$

و نتیجه طیف توان با طیف توان منحنی باند بحرانی شبیه سازی[1] شده، $\Psi(\Omega)$، کانولوشن می‌شود. این قسمت شبیه پردازش طیفی در تجزیه و تحلیل کپسترال مل بجز برای شکل خاص منحنی باند بحرانی می‌باشد. در این روش، محنی باند بحرانی به صورت زیر محاسبه می‌شود.

$$\Psi(\Omega)=\begin{cases}0 & \Omega<-1.3\\ 10^{2.5(\Omega+0.5)} & -1.3\le\Omega\le-0.5\\ 1 & -0.5\le\Omega\le0.5\\ 10^{-(\Omega-0.5)} & 0.5\le\Omega\le2.5\\ 0 & \Omega>2.5\end{cases} \qquad (۷-۲)$$

که حاصل کانولوشن گسسته $\Psi(\Omega)$ با $P(w)$ نمونه‌های طیفی توان باند بحرانی می‌شود:

$$\theta(\Omega_i)=\sum_{\Omega=-1.3}^{2.5}P(\Omega-\Omega_i)\Psi(\Omega) \qquad (۸-۲)$$

- **پیش‌تاکید بلندی صوت برابر[2]**

$\Theta[\Omega(w)]$ نمونه‌گیری شده توسط منحنی بلندی برابر شبیه‌سازی شده همان طور که در شکل زیر مشاهده می‌کنید، پیش‌تاکید شده است

$$\Theta[\Omega(w)]=E(w)\Theta(w) \qquad (۹-۲)$$

تابع $E(w)$ یک تقریبی از حساسیت شنوایی انسان در فرکانس‌های مختلف می‌باشد و حساسیت شنوایی انسان در سطح ۴۰ دسی‌بل شبیه‌سازی می‌کند. یک تقریب خاصی از طرف Makhoul و Cosell معرفی شده است که به صورت زیر می‌باشد[12].

$$E(w)=\left[(w^2+56.8*10^6)\,w^4\right]\Big/\left[\begin{matrix}(w^2+6.3*10^6)^{2*}\\ (w^2+0.38*10^9)\end{matrix}\right] \qquad (۱۰-۲)\ [۱۲]$$

[1] Power Spectrum of Simulated Critical-Band Curve

[1] Equal-Loudness Pre-emphasis

مقادیر اولین و آخرین نمونه‌ها به طور برابر در مقادیر نزدیک‌ترین همسایگی خود ساخته شده است؛ بنابراین $\Theta[\Omega(w)]$ با دو نمونه برابر شروع و به تمام می‌شود.

- **قانون توان شدت بلندی صوت[۱]**

آخرین عملگر پیش از مدل‌سازی کردن تمام قطب، متراکم سازی دامنه ریشه مکعبی می‌باشد:

$$\Phi(\Omega) = \Theta(\Omega)^{0.33} \qquad (۱۱-۲)$$

این عملگر یک تقریب از قانون توان شنوایی است و نسبت غیر خطی بین شدت صدا و بلندی صوت درک شده را شبیه‌سازی می‌کند. همچنین این عملگر اختلاف دامنه طیفی باند بحرانی را تا جایی که مدل کردن تمام قطب توسط مدل مرتبه پایین به طور نسبی انجام شود، کاهش می‌دهد.

- **مدل سازی اتورگرسیو[۲]**

در آخرین عملگر فرایندPLP ، $\Phi(\Omega)$ توسط طیف یک مدل تمام قطب با استفاده از روش مدل‌سازی خودهمبستگی طیفی تمام قطب تقریب زده شده است. در اینجا ما خلاصه‌ای از این روند را خواهیم داشت: معکوس تبدیل فوریه گسسته برای بدست آوردن تابع خودهمبستگی بر مبنای $\Psi(\Omega)$ در $\Phi(\Omega)$ بکار گرفته شده است. M+1 تای اول مقدار خودهمبستگی برای حل معادله Yule-Walker استفاده شده است تا ضرایب اتورگرسیو مدل تمام قطب مرتبه M ام بدست آید.

۲-۴-۲-۳- ضرایب کپسترال فرکانسی مل

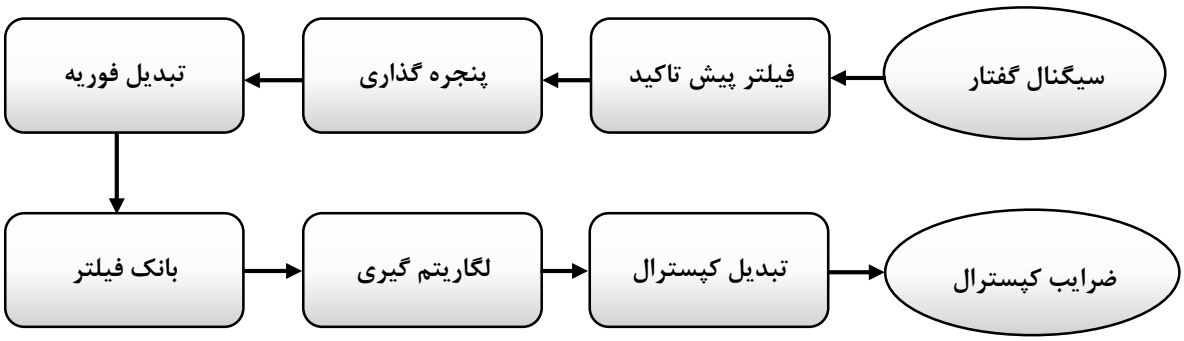
از آنجایی که ما در این پروژه از ویژگی MFCC استفاده کرده‌ایم، به طور مفصل چگونگی استخراج این ویژگی را شرح خواهیم داد. همان طور که در قسمت بالا بیان شده است، پس از انجام پیش پردازش‌های مختلف، شروع به استخراج ویژگی می‌کنیم. برای استخراج ویژگی MFCC طبق آنچه که

---

[۲] Intensity-Loudness Power Law

[۳] Autoregressive Modeling

در شکل (۲-۹) نمایش داده شده است عمل می‌کنیم. در ابتدا با استفاده از قاب مناسب هم‌پوشان سیگنال اصلی را به قسمت‌های کوچک‌تر تقسیم‌بندی می‌کنیم. معمولاً طول این قاب‌ها بین ۲۰ تا ۳۰



شکل (۲-۹) مراحل بدست آوردن MFCC

میلی ثانیه می‌باشد و با فاصله ۱۰ تا ۱۵ میلی ثانیه آن را بر روی سیگنال اصلی جا به‌جا می‌کنند. عملیات پیش تأکید بعد از مرحله قاب گذاری برای افزایش تأثیر توان در فرکانس‌های بالا و ایجاد یک حالت متوازن این توان در فرکانس‌های بالا و پایین، استفاده می‌شود. معادله تفاضلی آن در (۱-۳) آمده است. که در آن K یک عدد بین ۰,۹۵ تا ۰,۹۸ می‌باشد.

$$S' = S_n - kS_{n-1} \qquad (۱۲-۲)$$

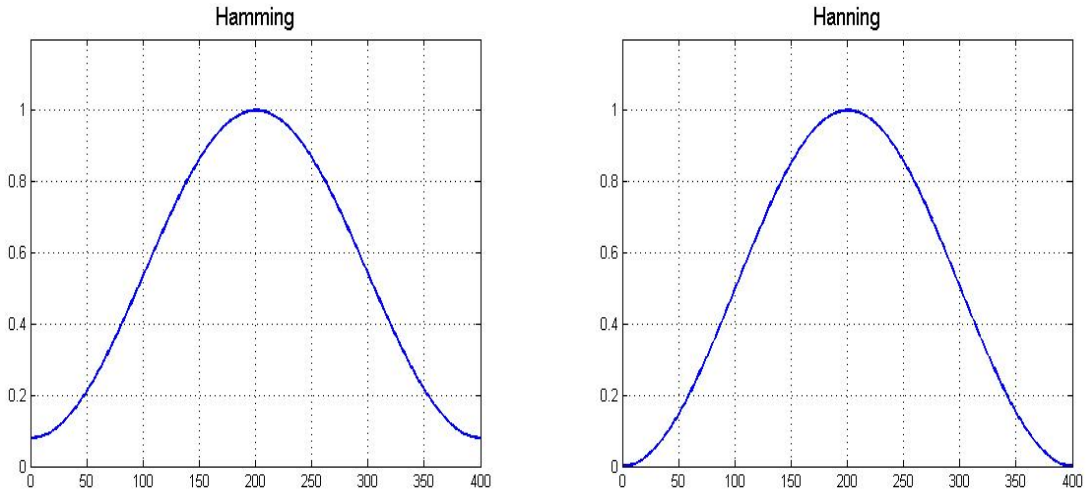
در قسمت بعد برای کم کردن اثرات ناپیوستگی در لبه‌های قاب از یک سری پنجره‌ها که معمولاً همینگ[۱] و هنینگ[۲] هستند استفاده می‌کنیم. در فرمول‌های (۲-۱۳) و (۲-۱۴) به ترتیب معادلات پنجره همینگ و هنینگ در حوزه زمان نمایش داده شده است. همچنین در شکل (۲-۱۰) به ازای N=400 دو پنجره فوق را مشاهده می‌کنید. در معادله‌های پایین N نشان دهنده طول پنجره و n هم نشان دهنده عنصر n ام پنجره‌ها می‌باشد.

[۱] Hamming

[۲] Hanning

$$W(\mathrm{n}) = 0.5 - 0.5\cos(\frac{2\pi n}{N-1}) \qquad (۱۳-۲)$$

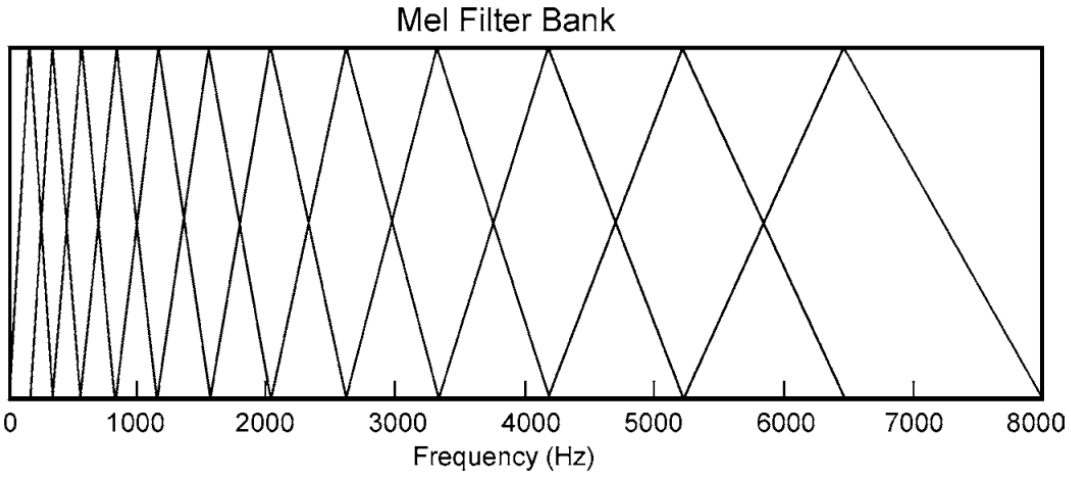
$$W(n) = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}) \qquad (۱۴-۲)$$



شکل (۲-۱۰) نمایش پنجره همینگ و هنینگ در حوزه زمان به طول ۴۰۰

بعد از اعمال پنجره، از تمام قاب‌ها تبدیل فوریه گسسته می‌گیریم و سپس طیف دامنه را محاسبه می‌کنیم. در قسمت بعد طیف دامنه هر یک از قاب‌ها را در بانک فیلتر مل غیر یکنواخت ضرب می‌کنیم. یک نمونه فیلتر بانک مل را در شکل (۲-۱۱) مشاهده می‌کنید. همان طور که در شکل زیر مشاهده می‌کنید، بانک فیلتر دارای توزیع یکنواختی در راستای محور فرکانسی نمی‌باشد اما به‌صورت متساوی الفاصله طبق فرمول (۲-۱۵) توزیع شده است.

$$mel(f) = 2595\log(f/700) \qquad (۱۵-۲)$$



شکل (۲-۱۱) یک نمونه بانک فیلتر مل

حال ضرایب بانک فیلتر را که جمع وزن‌دار از دامنه طیف در کانال i ام بانک فیلتر می‌باشد را با فرمول زیر محاسبه می‌کنیم.

$$m_i = \sum_k W_{k,i} X_k \qquad (۱۶-۲)$$

که X_k دامنه طیف و $W_{k,i}$ وزن فیلتر در کانال i ام نشان می‌دهد. در آخر با استفاده از فرمول (۲-۱۷) ضرایب نهایی بدست می‌آید که در آن N تعداد کانال‌های بانک فیلتر می‌باشد[14].

$$C_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} \log\left(m_j \cos\left(\frac{\pi i}{N} (J - 0.5) \right) \right) \qquad (۱۷-۲)$$

۲-۴-۲-۴- پویایی

در سیستم‌های بازشناسی گفتار معمولاً علاوه بر ضرایب کپسترال مشتقات آن را نیز برای بالا بردن دقت بازشناسی محاسبه می‌کنند. ضرایب کپسترال حاوی اطلاعات استاتیکی سیگنال صحبت می‌باشد و به حالت‌های گویش و تغییرات آن حساسیت دارد، در صورتی که مشتقات آن حاوی اطلاعات پویا و انتقال میانی حالات گوناگون لهجه می‌باشد. از این رو ترکیبی از ضرایب کپسترال و مشتقات آن می‌تواند ویژگی‌های بهتری از سیگنال صحبت را ایجاد کند. با استفاده از فرمول (۲-۱۸) مشتقات ضرایب

کپسترال را می‌توان محاسبه کرد. معمولاً از دو مرتبه اول مشتقات ضرایب کپسترال استفاده می‌کنند چون مشتق‌های مرتبه دو به بعد تأثیر چندانی در دقت بازشناسی ندارند.

$$d_t(i)=\sum_{T=1}^{N}\left[\frac{T\left(C_{t+T}(i)-C_{t-T}(i)\right)}{2\sum_{T=1}^{N}T^2}\right] \qquad N\le t\le T-N \qquad (۱۸-۲)$$

در رابطه بالا C_t ضرایب مشتق اول کپسترال در زمان t، T تعداد کل قاب‌ها و N نمایانگر طول قاب‌ها می‌باشد. همچنین برای محاسبه مشتق اول، اولین و آخرین قاب از معادلات زیر استفاده می‌کنیم.

$$C_t(i)=C_t(i)-C_{t-1}(i) \qquad t<N \qquad (۱۹-۲)$$

$$C_t(i)=C_t(i)-C_{t-1}(i) \qquad t\ge T-N \qquad (۲۰-۲)$$

در رابطه بالا N عددی اختیاری می‌باشد ولی اعداد ۲ و ۳ می‌تواند مقداری مناسب باشد. حال برای بدست آوردن مشتق دوم و بالاتر ضرایب کپسترال همین روند را از روابط (۲-۱۸) تا (۲-۲۰) استفاده می‌کنیم[15].

۲-۴-۲-۵- انرژی

انرژی یکی دیگر از ویژگی‌ها می‌باشد که در سیستم‌های پردازش گفتار استفاده می‌کنند. اگر چه استفاده تنها از ویژگی در سیستم‌های امروزه مرسوم نیست ولی به عنوان یک ویژگی مکمل می‌توان از آن استفاده کرد. قابلیت بارز خصوصیات انرژی، توان تمایز بین واج‌ها و گوینده‌ها را دارا می‌باشد. انرژی هر قاب توسط فرمول زیر محاسبه می‌گردد.

$$\mathrm{E}=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\mathrm{S}_i \qquad (۲۱-۲)$$

که در آن i نشان دهنده شماره قاب مورد نظر، S_i انرژی قاب i ام و N تعداد کل قاب می‌باشد که البته در طراحی‌های واقعی از الگوریتم این ویژگی استفاده می‌کنند.

۲-۴-۳- طبقه‌بندی

طبقه‌بندی، یکی دیگر از قسمت‌های سیستم‌های بازشناسایی گوینده می‌باشد که نقش اساسی در دقت این نوع سیستم‌ها ایفا می‌کند. به‌طورکلی، طبقه بندی کردن و یا ساخت مدل گوینده دو روش پیش رو برای دسته بندی در سیستم‌های بازشناسی مورد استفاده قرار می‌گیرد که به ترتیب به دسته تولیدی و تمایزی معروف هستند. در روش مدل سازی گوینده می‌توان به VQ، GMM، HMM و غیره اشاره کرد که معمولاً از تابع چگالی برای نشان دادن مدل گوینده‌گان یا کلمات مختلف استفاده می‌شود. همچنین در روش تمایزی معمولاً از مرزهای تصمیم گیری برای تفکیک گوینده‌ها و کلمات مختلف استفاده می‌شود. شبکه‌های عصبی مصنوعی و ماشین بردار پشتیبان[۱] را می‌توان نمونه‌ای از این دسته نام برد. در این قسمت به طور مختصر چند نمونه از این دسته‌بندی را شرح می‌دهیم.

۲-۴-۳-۱- شبکه‌های عصبی

شبکه‌های عصبی یکی دیگر از تکنولوژی‌هایی است که در دهه ۱۹۵۰ معرفی گردیده است ولی به دلیل نداشتن نتایج مناسب تا دهه ۱۹۸۰ کارایی چندانی نداشته است[16]. پس از پیدایش پردازنده‌های موازی و استفاده از سیستم‌های آموزش پذیر، استفاده از شبکه‌های عصبی که الهام گرفته از دستگاه پردازشی عصبی انسان می‌باشد، احیا گردیده است. شبکه‌های عصبی دارای مزیت‌های همچون تطبیق پذیری[۲]، غیرخطی بودن، خود یادگیری[۳] و پیاده سازی راحت‌تر آن نسبت به روش‌های دیگر می‌باشد. همچنین از دیگر مزیت‌های شبکه عصبی تخمین توابع غیر خطی می‌باشد، از این رو شبکه عصبی می‌تواند یک روش جدید برای حل مسائل بازشناسی گفتار که جز مسائل سخت طبقه بندی الگو می‌باشد، استفاده کرد. در اویل دهه ۱۹۹۰، تلاش‌ها برای استفاده از شبکه‌های عصبی بیشتر در تشخیص

---

[۱] Support Vector Machine

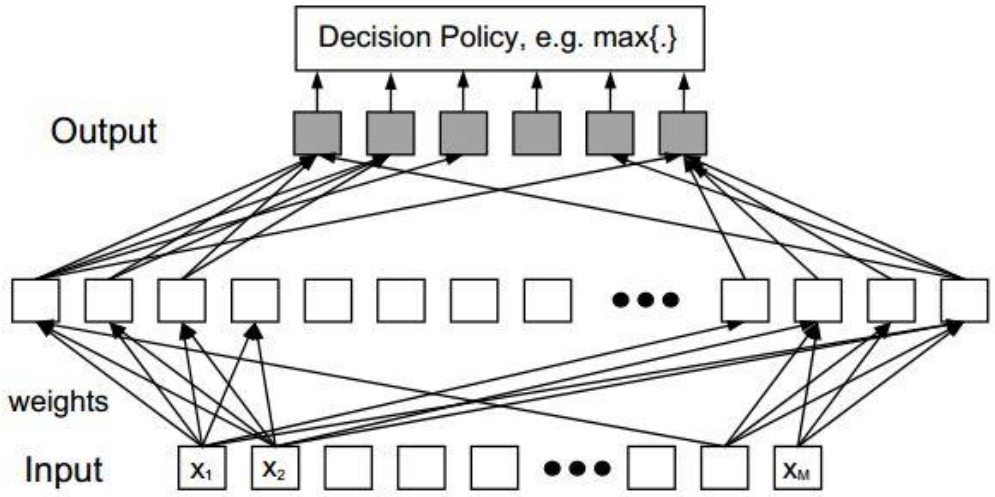
[۲] Adaptive

[۳] Self-Learning

کلمه‌های مجزا[۱] و واج‌ها[۲] متمرکز شده بود و با گذشت زمان این پیشرفت ادامه داشت تا این که در همه زمینه‌ها از شبکه عصبی برای بهبود روند بازشناسایی استفاده گردیده است[17]. شبکه MLP[۳] و RBF[۴] دو تا از شناخته‌ترین و پرکاربردترین از انواع شبکه‌های عصبی می‌باشند.

- MLP

همان در شکل (۲-۱۲) نمونه‌ای از شبکه MLP را مشاهده می‌کنید این شبکه از یک لایه ورودی، یک لایه خروجی و یک قسمت به نام لایه مخفی تشکیل شده است. هر کدام از لایه‌ها از چندین سلول[۵] تشکیل شده است. معمولاً در شناسایی الگو تعداد نرون‌های لایه ورودی را تعداد ویژگی‌ها و تعداد نرون‌های لایه خروجی را تعداد کلاس‌های موجود تعیین می‌کند.



شکل (۲-۱۲) شبکه عصبی پرسپترون

[۴] Isolated

[۵] Phonemes

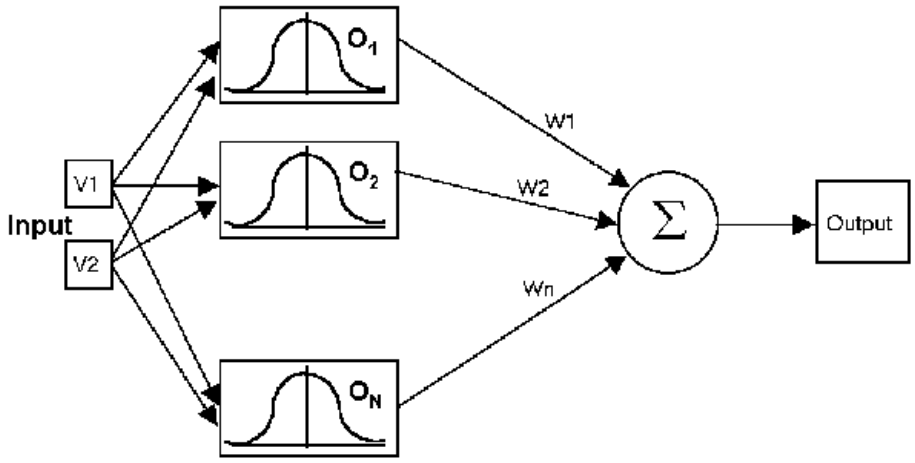
[۶] Multi-Layer Perceptron

[۷] Radial Basis Function

[۷] Neuron

- RBF

همان طور که قبلاً هم ذکر شد، RBF یکی دیگر از انواع شبکه عصبی می‌باشد که پرکاربرد می‌باشد. این شبکه در شکل (۲-۱۳) نشان داده شده است و همان طور که ملاحظه می‌کنید همانند MLP از یک لایه ورودی و خروجی و یک لایه میانی هم وجود دارد.

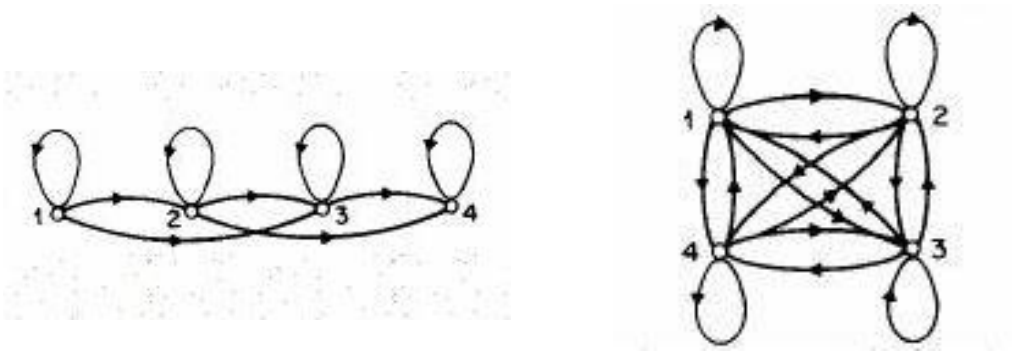


شکل (۲-۱۳) نمونه‌ای از شبکه RBF

۲-۴-۳-۲- مدل مخفی مارکوف

مدل‌سازی سیگنال یکی از زمینه‌ها در پردازش سیگنال می‌باشد که در آن خیلی نگاه‌ها را به خود معطوف کرده است. روش‌های زیادی برای مدل‌سازی خصوصیات سیگنال وجود دارد که یکی از این روش‌های معمول روش آماری می‌باشد[18]. از انواع این روش‌ها می‌توان مدل مخلوط گوسی، مدل مخفی مارکوف و زنجیره مارکوف اشاره کرد. HMM در دهه ۱۹۶۰ معرفی گردیده است و ابزاری قدرتمند برای مدل سازی سیگنال بر اساس فرایند آماری می‌باشد. اساس و پایه سیستم‌های بازشناسی گفتار پیوسته امروزی که بر مبنای HMM هستند در اویل دهه ۱۹۷۰ توسط گروه Carnegie Mellon و IBMکه HMM های گسسته را معرفی کرده‌اند، شکل گرفته است و بعد از آن در آزمایشگاه Bell، HMM هایی پیوسته معرفی شد. اولین تلاش‌ها برای ساخت سیستم‌های بازشناسی پیوسته مستقل از

متن در اویل دهه ۱۹۹۰ شروع شد که در اواسط همین دهه دقت قابل قبولی در این نوع سیستم‌ها کسب شده بود[19]. انواع مختلفی از مدل مخفی مارکوف هم در حوزه گسسته و هم در حوزه پیوسته وجود دارد که در علوم گوناگون کاربردهای خاص خود را دارند. در شکل (۲-۱۴) دو مدل آرگودیک و چپ به راست را مشاهده می‌کنید که مدل‌های پایه HMM هستند و در پردازش گفتار مورد استفاده قرار می‌گیرد.



(ب) (الف)

شکل (۲-۱۴) الف-مدل ارگودیگ HMM ب-مدل چپ به راست HMM [21][20]

۲-۴-۳-۳-مدل مخلوطی گوسی[1]

یکی دیگر از پرکاربردترین روش‌های مدل سازی آماری، مدل مخلوطی گوسی می‌باشد. مدل مخلوطی گوسی یک تابع چگال احتمالی پارامتری می‌باشد که مجموع وزن‌دار توزیع چگالی گوسی را نمایش می‌دهد. در این روش هر یک از کلاس‌ها یا الگوها توسط تابع چگال احتمال مدل سازی می‌شود که با ترکیب کردن خطی تعداد از این توابع بدست می‌آید. همچنین تحقیقات در این زمینه در دو قالب متفاوت برای بهبود کارایی این روش انجام می‌گیرد، در قالب اول محققان سعی در استفاده از دیگر روش‌ها نظیر شبکه‌های عصبی، ماشین بردار پشتیبان و غیره در مدل مخلوطی گوسی می‌باشند که

[1] Gaussian Mixture Model or GMM

باعث بهبود هر چه بیشتر سیستم‌های شناسایی گفتار می‌شود. از سوی دیگر، یک سری از تحقیقات بر روی خود الگوریتم و پارامترهای این الگوریتم انجام می‌شود که باعث به وجود آمدن یک مجموعه الگوریتم‌های بهبود یافته نظیر آنچه که در[23]، [24] و[25]مشاهده می‌شود، انجام می‌پذیرد.

مدل مخلوطی گوسی به طور معمول در مدل‌سازی پارامترهای توزیع احتمالاتی پیوسته و یا ویژگی سیستم‌های زیستی استفاده شده است. پارامترهای GMM با آموزش داده‌ها تخمین زده می‌شود که این کار توسط الگوریتم EM[۱] و یا تخمین MAP[۲] انجام می‌پذیرد. GMM مجموع وزن‌دار m تا توزیع چگالی گوسی با ابعاد D، به صورت شکل زیر می‌باشد.

$$N(x \mid \mu, \Sigma) = \frac{1}{(2\pi\sigma^2)^{D/2} \mid \sum \mid^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \mid \Sigma^{-1}(x-\mu)\right\} \qquad (۲۲-۲)$$

که در آن x برداری با D بعد ویژگی، μ بردار میانگین و $|\Sigma|$ ماتریس کوواریانس می‌باشد. و همان طور که از نام این مدل مشخص می‌باشد مخلوط گوسی از مجموع چندین توزیع گوسی بدست می‌آید.

$$P(x \mid \lambda) = \sum_{i=1}^{m} P_i b_i(x) \qquad (۲۳-۲)$$

در سیستم‌های شناسایی گوینده، P وزن‌های اجزای مخلوطی، b تابع چگالی ترکیبی گوینده‌ها یا کلاس‌ها و m هم تعداد گوینده می‌باشد. همچنین برای برقراری شرط مخلوط باید عبارت زیر هم برقرار باشد.

$$\sum_{i=1}^{m} P_i = 1 \qquad (۲۴-۲)$$

---

[۱] Expectation Maximization

[۲] Maximum A Posteriori

فصل سوم:

الگوریتم رقابت استعماری

۳-۱- مروری بر معرفی الگوریتم‌های تکاملی

بهینه‌سازی اهمیت ویژه‌ای در همه علوم مثل فیزیک، شیمی، مهندسی و غیره دارد. برای مثال پارامترها و متغیرهای زیادی در نتیجه یک فرایند موثر هستند، از این رو به ازای هریک از این پارامترها می‌تواند چندین جواب وجود داشته باشد. برای بدست آوردن بهینه‌ترین جواب بر اساس یک تابع هزینه[۱] و یا تابع برازندگی[۲] می‌توان از روش‌های بهینه سازی استفاده کرد. به طور کل روش‌های بهینه‌سازی در دو دسته خلاصه شده است؛ بهینه‌سازی محلی و بهینه‌سازی سرتاسری یا عام. برای بهینه‌سازی سراسر یا عام معمولاً از الگوریتم‌های تکاملی استفاده می‌کنند، نظیر الگوریتم ژنتیک، بهینه-سازی ذرات، باز پخت شبیه سازی شده و غیره. در این بین الگوریتم ژنتیک به عنوان یکی از معرف‌ترین الگوریتم تکاملی از این نوع می‌باشد که در سال ۱۹۶۲ توسط هالن معرفی شده است. همان گونه که از اسم این نوع از الگوریتم‌های تکاملی مشخص می‌باشد، این‌ها از تکامل زیستی و ژنتیکی انسان‌ها الهام گرفته است. از طرف دیگر، آنچه که مشخص است سرعت تکامل فرایند اجتماعی و سیاسی انسان‌ها بیشتر از سرعت تکامل زیستی آن‌ها می‌باشد. الگوریتم رقابت استعماری به عنوان یکی از الگوریتم‌های تکاملی نو ظهور شناخته می‌شود و از سیر تکامل سیاسی اجتماعی بشر الهام گرفته است و از سرعت همگرایی بالاتری نسبت به الگوریتم‌های یاد شده برخوردار است.

۳-۲- الگوریتم رقابت استعماری

همواره استعمار به عنوان یک پدیده تاریخی و موضوعی غیرقابل اجتناب ناپذیر مطرح بوده است و کشورهای استعمارگر برای بدست بعضی از منابع که در اختیار نداشتن کشورهای ضعیف را که همان منابع را در اختیار داشتن را به عنوان مستعمره خود قرار می‌دادند. الگوریتم رقابت استعماری از تاریخ

---

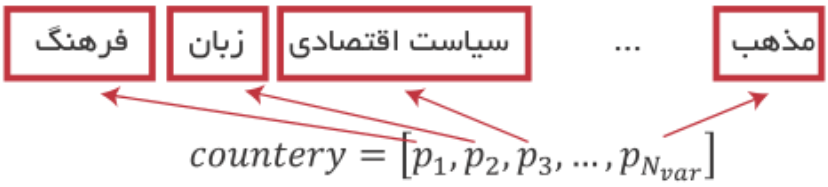
[۱] Cost Function

[۲] Fitness Function

رقابتی بین کشورهای استعمارگر شکل گرفته است. این الگوریتم در سال ۲۰۰۷ توسط اسماعیل آتش پز و همکارانش در دانشگاه تهران ارائه شده است. حال در چندین زیر بخش به معرفی این الگوریتم خواهیم پرداخت تا با نحوه عملکرد این الگوریتم بیشتر آشنا شویم.

۳-۲-۱- ایجاد جمعیت اولیه

یک امر مشترک در همه الگوریتم‌های تکاملی، بهینه سازی بر اساس یک سری متغیرهای داده شده مسئله می‌باشد. پس برای شروع کار در این نوع الگوریتم‌ها نیاز به تعریف یک سری آرایه‌ها از متغیرها داریم. این متغیرها در الگوریتم‌های مختلف، متفاوت می‌باشند و به‌طور مثال در الگوریتم ژنتیک، کوروموزم و در الگوریتم رقابت استعماری، کشور نامیده می‌شود. همان طور که در شکل (۳-۱) مشاهده می‌شود در اینجا هر کشور می‌تواند با پارامترهای مختلف نظیر موقعیت سیاسی و اجتماعی، زبان، فرهنگ، اقتصادی و غیره تعریف شود.

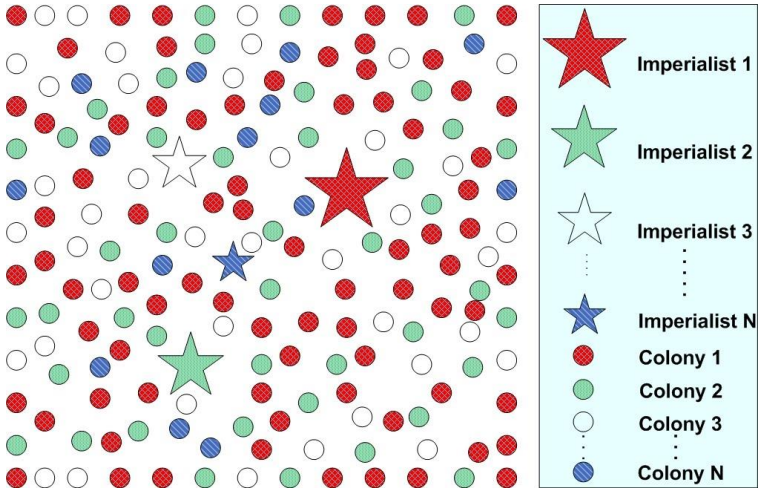


شکل (۳-۱) تعریف یک کشور بر اساس پارامترها[10]

پس از تعریف N کشور با پارامترهای ثابت، تعداد N_{Imp} کشور را که دارای کمترین تابع هزینه و یا بیشترین برازندگی می‌باشند را به عنوان کشورهای استعمارگر انتخاب می‌کنیم. پس از انتخاب کشورهای استعمارگر، بقیه کشورها یعنی $N_{Contery} - N_{Imp}$ تعداد کشور که به عنوان مستعمره شناخته می‌شوند را بر اساس قدر هر استعمار گر به آن‌ها تخصیص می‌دهیم. قدرت هر استعمارگر جهت تخصیص مستعمره‌ها، از فرمول (۳-۱) بدست می‌آید[10].

$$P_n = \left| \frac{C_n}{\sum_{i=1}^{N_{\text{Imp}}} C_i} \right| \qquad (۱-۳)$$

پس از مشخص شدن استعمارگرها و مستعمرات مربوط به آن‌ها، امپراطوری‌های اولیه همانند شکل(۳-۲) تشکیل می‌شود.

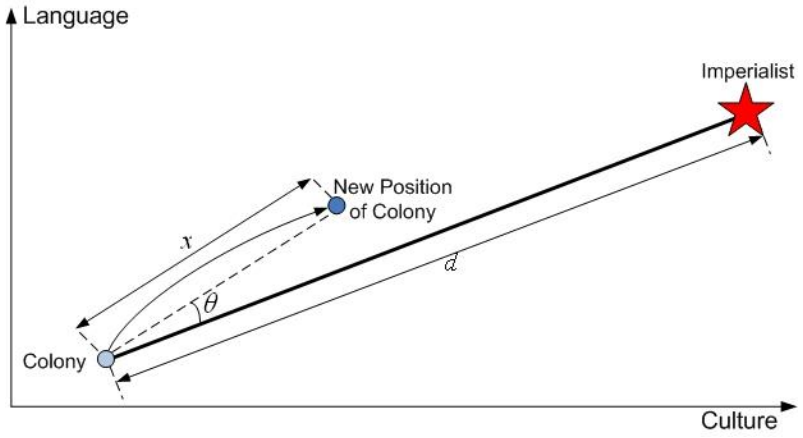


شکل (۳-۲)) نحوه تشکیل امپراطوری اولیه[10]

همواره کشورهای استعمارگر در تلاش هستند با تغییر در بخش‌های مختلف سیاسی، اجتماعی، فرهنگی، آموزشی مستعمرات خود، در آن‌ها طوری تغییر ایجاد کنند که به سمت آن‌ها گراییده شوند. به این سیاست در اصطلاح، جذب و ترکیب یا همگون سازی[۱] گفته می‌شود. البته این قبیل کارها در راستای منافع کشورهای استعمارگر انجام می‌پذیرد و باعث رونق هر چه بیشتر این کشورها می‌شود. اگر چه بعضی از این استعمارگران به ظاهر کارهای عمرانی و زیر بنایی همانند ساخت دانشگاه‌ها، بیمارستان-ها و غیره در مستعمرات خود انجام می‌دهند. به طور مثال کشور انگلستان در هند که به عنوان یکی از مستعمرات این کشور محسوب می‌شود تلاش‌های زیادی انجام داده تا این کشور را از طریق سیاست

[۱] Assimilation

همگون سازی به سمت خود بکشاند. در اینجا، تلاش این کشور را در دو زمینه فرهنگ و زبان نشان دادیم که باعث سوق دادن کشور هند به سمت این کشور می‌شود. همان گونه که در شکل (۳-۳) مشخص است، کشور مستعمره، به اندازه x واحد در جهت خط واصل مستعمره به استعمارگر و با زاویه θ، حرکت کرده و به موقعیت جدید، می‌رسد. X و θ عددی تصادفی با توزیع نرمال (و یا دیگر توزیع مناسب) می‌باشد. البته X و θ طوری انتخاب می‌شوند که از استعمارگر زیاد منحرف نشود، حداکثر طول X را دو برابر فاصله بین استعمارگر و مستعمره در نظر گرفته می‌شود و θ معمولاً بر حسب تجربه ۴۵ درجه برای حداکثر انحراف مناسب می‌باشد. زاویه θ به این منظور اضافه شده است که ما جستجوی قوی‌تری برای بدست آوردن بهترین جواب داشته باشیم[10].



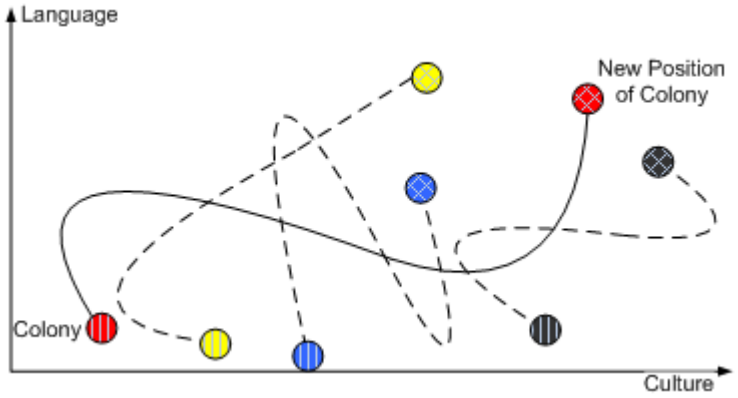
شکل (۳-۳) سیاست جذب کشورهای مستعمره توسط استعمارگرها[10]

۳-۲-۲- انقلاب[1]

در تاریخ کشورها، همیشه شاهد تلاش کشورها برای بهبود وضعیت موجود هستیم و در بعضی اوقات بهبود وضعیت یک کشور در زمینه‌های مختلف به شکل انقلاب نمو پیدا می‌کند و در وضعیت به

[1] Revolution

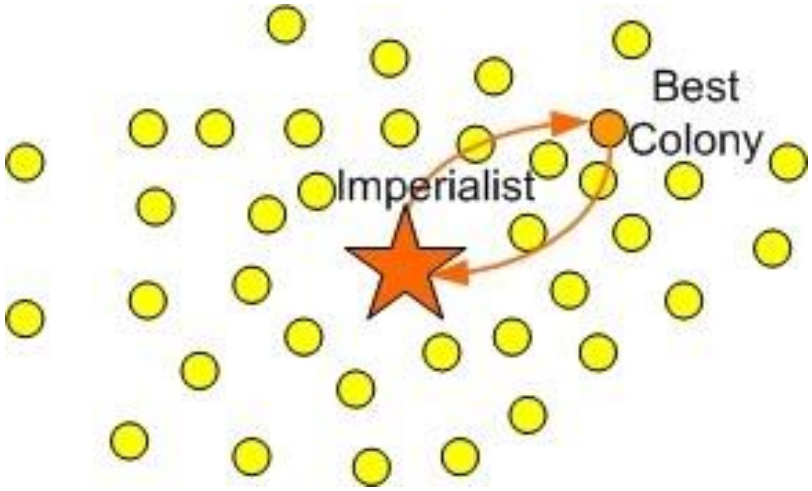
مراتب بهتری نسبت به قبل پیدا می‌کند. در این الگوریتم از حرکت‌های تصادفی برای شبیه سازی این امر استفاده شده است که در بعضی از زمان‌ها باعث می‌شود کشوری از موقعیت مینیم محلی خارج شده و در مینیم بهتری نسبت به قبل قرار گیرد. نحوه ایجاد انقلاب در دو زمینه زبان و فرهنگ در کشورهای مستعمره را در شکل زیر مشاهده می‌کنید. چه بسا در این بین، موقعیت کشور مستعمره از استعمارگر هم بهتر شود.



شکل (۳-۴) نحوه رخ دادن انقلاب در کشورها[10]

۳-۲-۳- تغییر موقعیت مستعمره و استعمارگر

گاهی در تغییر و تحولات سیاسی کشورها، در حین همگون سازی کشورها توسط کشورهای استعمارگر، کشورهای مستعمره در وضعیت بهتری نسبت به استعمارگر خود قرار می‌گیرند، یعنی این که از تابع هزینه کمتری یا شاخصه بهتری نسبت به استعمارگر برخوردار می‌شوند. هر گاه شاهد چنین اتفاقی در یک امپراطوری باشیم، جای مستعمره و استعمارگر در آن امپراطوری عوض می‌شود و از این پس دیگر مستعمرات در آن امپراطوری به سمت استعمارگر جدید گراییده می‌شوند. استعمارگر قبلی هم همانند شکل (۳-۵) به عنوان مستعمره جدید شناخته می‌شود.



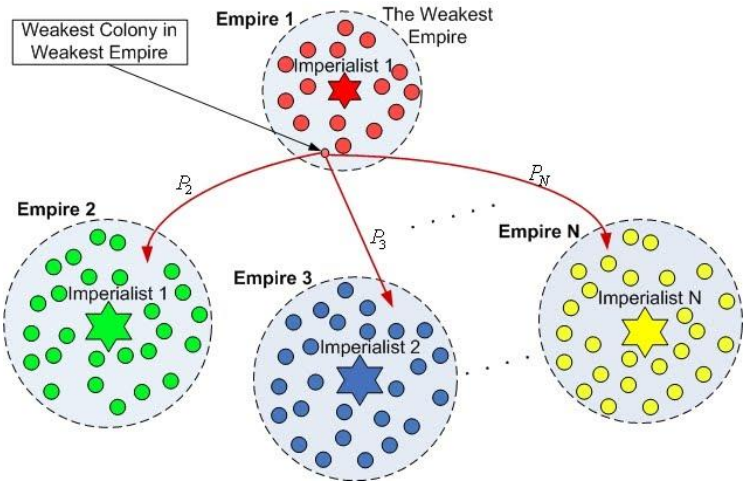
شکل (۳-۵) تغییر موقعیت بین مستعمره و استعمارگر[10]

۳-۲-۴- رقابت استعمارگرانه

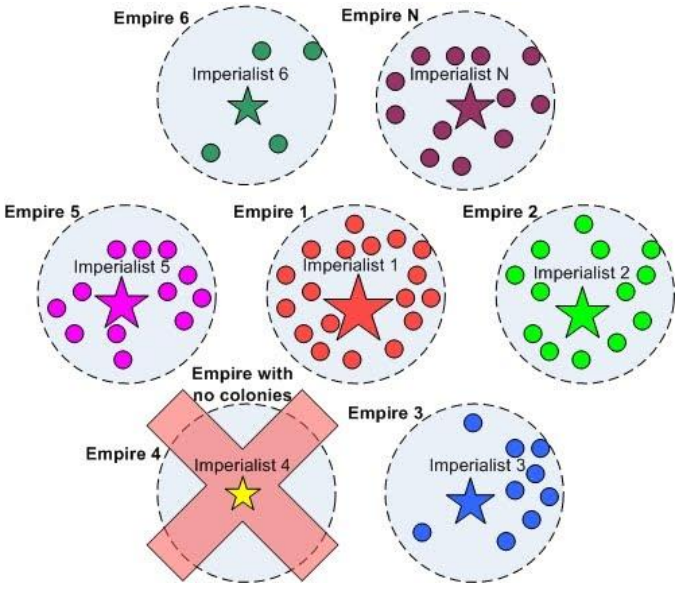
هر استعمارگر در جهان تلاش می‌کند به واسطه نفوذ در کشورهای مستعمره، قدرت و قلمرو حکمرانی خود را افزایش دهد؛ زیرا قدرت هر امپراطوری به قدرت مستعمرات آن نیز بستگی دارد. به همین خاطر قدرت هر امپراطوری در الگوریتم رقابت استعماری برابر با قدرت و توان استعمارگر و ضریبی از میانگین قدرت و توان مستعمرات آن تعریف شده است. در این الگوریتم هرگاه هر یک از امپراطوری‌ها نتواند قدرت خود را در هر تکرار حفظ کند و یا بهبود ببخشد، اگر به عنوان ضعیف‌ترین امپراطوری شناخته شود، ضعیف‌ترین مستعمره خود را از دست می‌دهد و دیگر امپراطوری‌ها بر سر بدست آوردن این مستعمره به رقابت خواهند پرداخت. هر چه قدرت یک امپراطوری بیشتر باشد احتمال تصاحب کردن این مستعمره، برای این امپراطوری بیشتر است. همان طور که مشخص است امپراطوری‌های ضعیف، ضعیف‌تر خواهند شد و امپراطوری‌های قوی، قوی‌تر خواهند شد. این روند رقابتی را می‌توان در شکل (۳-۶) در صفحه بعد مشاهده کرد.

۳-۲-۵- حذف امپراطوری بدون مستعمره

در روند رقابت بر سر ضعیف‌ترین مستعمره، ضعیف‌ترین امپراطوری به نقطه‌ای خواهیم رسید که ضعیف‌ترین امپراطوری دیگر هیچ مستعمره‌ای در اختیار ندارد و خود استعمارگر به عنوان مستعمره به دیگر امپراطوری‌ها واگذار می‌شود. البته می‌توان شرط‌های دیگری هم برای این الگوریتم در نظر گرفت تا منجر به حذف یک امپراطوری گردد. شکل (۳-۷) حذف امپراطوری بدون مستعمره را نشان می‌دهد.



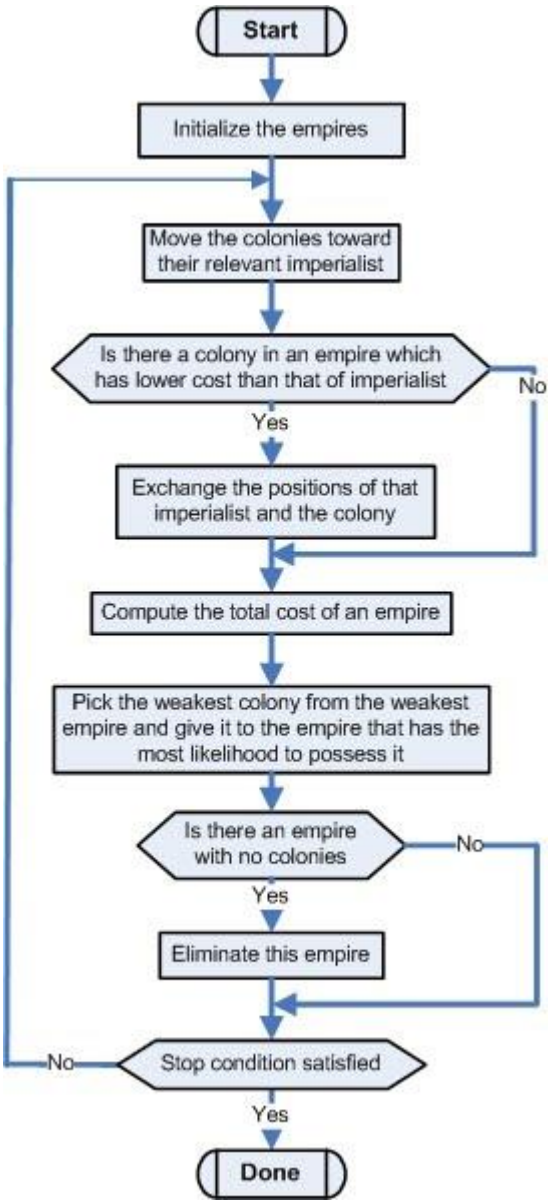
شکل(۳-۶) رقابت استعماری بین استعمارگران[10]



شکل (۳-۷) سقوط ضعیف‌ترین امپراطوری[10]

۳-۲-۶- همگرایی و فلوچارت الگوریتم رقابت استعماری

ما می‌توانیم شرطی را برای خاتمه این الگوریتم انتخاب کنیم و یا اینکه کار را تا آنجا پیش ببریم که یک امپراطوری واحد تشکیل شود و در اینجاست که دیگر رقابتی وجود ندارد. وقتی که الگوریتم به پایان رسید، متغیرهای استعمارگر موجود به عنوان بهترین و یا بهینه‌ترین جواب مسئله شناخته می‌شود. در شکل (۳-۸) فلوچارت این الگوریتم را ارائه شده است.



شکل (۳-۸) فلوچارت الگوریتم رقابت استعماری[10]

۳-۳- الگوریتم بهینه‌سازی گروه ذرات[۱]

الگوریتم فرا ابتکاری بهینه‌سازی گروه ذرات روش محاسباتی تکاملی مبتنی بر جمعیت جواب‌ها است. مانند سایر الگوریتم‌های فرا ابتکاری، الگوریتم مذکور ابزار بهینه سازی است که می‌تواند برای حل انواع مختلفی از مسائل بهینه سازی به کار گرفته شود. این الگوریتم یک روش‌های فرا ابتکاری است که با الهام گیری از رفتار اجتماعی گروهی از پرندگان مهاجر که در تلاش برای دستیابی به مقصد ناشناخته-ای هستند، توسط کندی و ابرهارت[۲] در سال ۱۹۹۵ میلادی توسعه داده شده است.

در الگوریتم PSO، جمعیت جواب‌ها، ذرات[۳] نامیده می‌شود و هر جواب مانند یک پرنده در گروهی از پرندگان است و ذره[۴] نام دارد و شبیه کرموزوم در الگوریتم ژنتیک است. تمامی ذرات دارای مقدار شایستگی هستند که با استفاده از تابع شایستگی[۵] محاسبه می‌گردند و تابع شایستگی ذرات باید بهینه گردد. جهت حرکت هر ذره توسط بردار سرعت[۶] آن ذره معین می‌شود. برخلاف الگوریتم ژنتیک، در فرآیند تکاملی الگوریتم مذکور، پرندگان جدیدی از نسل قبل (جواب‌های جدید از جواب‌های قبلی) ایجاد نمی‌گردد، بلکه هر پرنده رفتار اجتماعی خود را با توجه به تجربیاتش و رفتار سایر پرندگان گروه تکامل بخشیده و مطابق آن حرکت خود را به سوی مقصد بهبود می‌دهد. به عبارت دیگر، در این الگوریتم

---

۱ Particle Swarm Optimization

۲ Kennedy and Eberhart

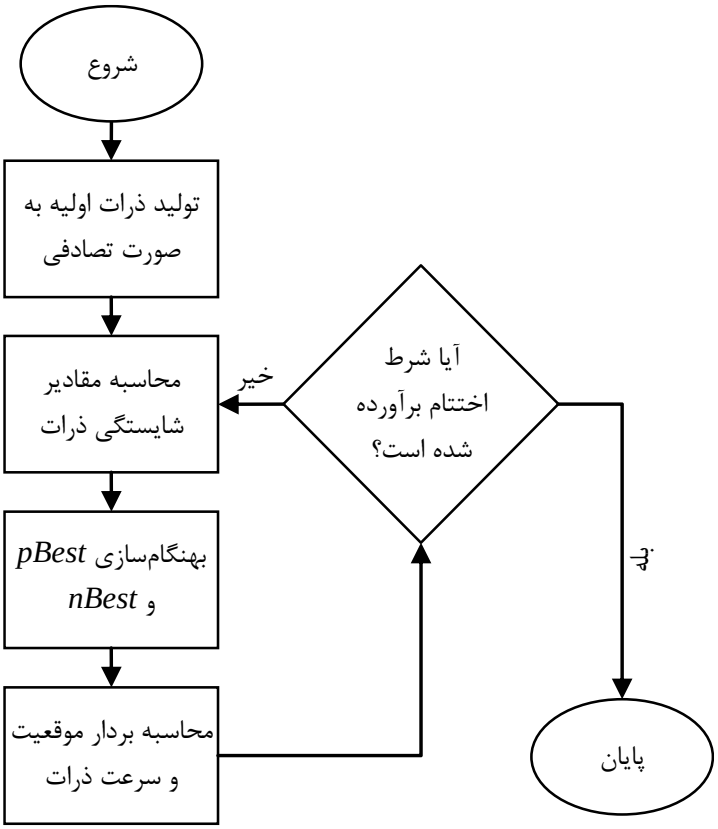
۳ Swarm

۴ Particle

۵ Fitness function

۶ Velocity

عملگرهای تکاملی چون تقاطع[1] و جهش[2] وجود ندارد[25] . ساختار کلی الگوریتم بهینه‌سازی گروه ذرات در شکل (۳-۹) ارائه شده است.



شکل (۳-۹) فلوچارت الگوریتم بهینه‌سازی گروه ذرات[26]

الگوریتم PSO با گروهی از جواب‌های تصادفی آغاز و سپس با به هنگام سازی ذرات در هر تکرار به دنبال جواب بهینه می‌گردد[27]. اگر متغیرهای تصمیم و به نوبه آن موقعیت ذرات، از نوع صفر و یک باشند؛ بردارهای سرعت و موقعیت هریک از ذرات در هر تکرار الگوریتم، طبق روابط (۳-۲) الی (۳-۴) محاسبه می‌شوند[28].

$$V_{it} = w.V_{it-1} + c_1.r_1.(pBest_i - x_{it}) + c_2.r_2.(nBest_i - x_{it}) \qquad (۳-۲)$$

[1] Crossover

[2] Mutation

$$-V_{\max} \le V_{it} \le V_{\max} \qquad (۳-۳)$$

$$s_i = 1/1 + e^{-V_{it}} \qquad (۳-۴)$$

$$x_{it} = \begin{cases} 1 & \rho \le s_i \\ 0 & otherwise \end{cases} \qquad (۳-۵)$$

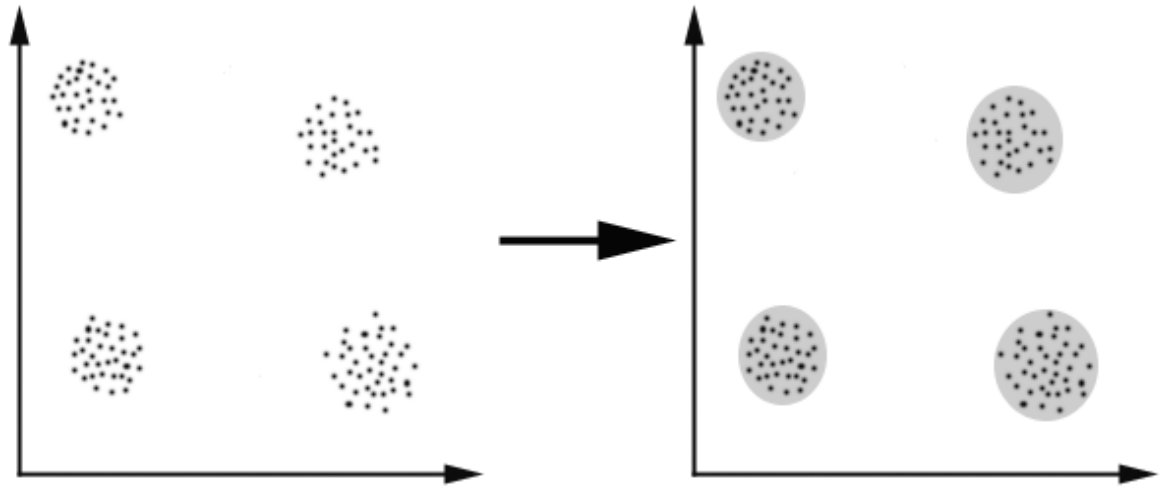
طبق رابطه (۳-۲) بردار سرعت جدید هر ذره بر اساس سرعت قبلی خود ذره ($V_{i(t-1)}$)، بهترین موقعیتی که ذره تاکنون به آن دست یافته است ($pBest_i$) و موقعیت بهترین ذره در همسایگی ذره که تابه‌حال بدست آمده است ($nBest_i$)، محاسبه می‌گردد. در صورتی که همسایگی هر ذره شامل تمام ذرات گروه باشد، آنگاه $nBest_i$ بیانگر موقعیت بهترین ذره در میان گروه است که با $gBest$ به آن اشاره می‌شود. r_1 و r_2 دو عدد تصادفی با توزیع یکنواخت بین [۰و۱] هستند که مستقل از یکدیگر تولید می‌شوند. c_1 و c_2 که با نام ضرایب یادگیری به آنان اشاره شده است، تأثیر $pBest$ و $nBest$ را بر فرآیند جستجو کنترل می‌نمایند. w بیانگر ضریب وزنی است. بردار سرعت ذرات با مقدار V_{max} محدود شده است. V_{max} به عنوان محدودیتی است که قابلیت جستجوی سرتاسری گروه ذرات را کنترل می‌کند. با استفاده از رابطه (۳-۴) بردار سرعت هریک از ذرات به بردار احتمال تغییر، تبدیل می‌شود. در رابطه فوق، s_i بیانگر احتمال آن است که x_{it} برابر با ۱ شود. سپس با استفاده از رابطه (۳-۵) بردار موقعیت هریک از ذرات بروز رسانی می‌گردد. در رابطه فوق، ρ عددی تصادفی با توزیع یکنواخت بین صفر و یک است.

فصل چهارم:

مروری بر الگوریتم‌های خوشه‌بندی

۴-۱- مقدمه

خوشه‌بندی، ابزاری غیر نظارتی در تحلیل داده‌ها می‌باشد تا به وسیله آن مجموعه داده‌ها را با شرایطی که مورد نظر ماست دسته‌بندی نماییم. خوشه یابی معمولاً با معیار شباهت‌ها شکل می‌گیرد به‌طوری که داده‌هایی که در یک خوشه قرار می‌گیرند بیشترین شباهت را باهم داشته باشند و داده‌هایی که در خوشه‌های متفاوت قرار می‌گیرند باهم کمترین شباهت را دارا باشند. همان طور که مشخص می‌باشد خوشه‌بندی با طبقه‌بندی متفاوت می‌باشد، زیرا در فرایند خوشه یابی ما هیچ نوع مرحله آموزش نداریم و فقط بر اساس معیار شباهت‌ها، داده‌های مشابه را از داده‌های غیر مشابه جدا می‌کنیم. در واقع آنچه که مشخص می‌باشد قبل از خوشه‌بندی باید ویژگی‌هایی را انتخاب کنیم که برای رسیدن به اهداف در خوشه‌بندی موثر باشد و داده‌ها مشابه و غیر مشابه را از هم جدا کند. در شکل(۴-۱) یک نمونه ساده خوشه‌بندی را نمایش می‌دهد.



شکل (۴-۱) یک نمونه ساده از خوشه‌بندی

۴-۲- انواع خوشه‌بندی

تا به امروز گروه‌های مختلفی از الگوریتم‌های خوشه‌بندی معرفی شده است که در شرایط داده‌ای مختلف کاربرد دارند. روش‌های همانند سلسله مراتبی[۱]، افرازی[۲]، مبتنی بر چگالی[۳]، توری[۴] و فازی را می‌توان به عنوان گروه‌های مختلفی از الگوریتم‌های خوشه‌بندی نام برد که در زیر به شرح مختصری از آن‌ها خواهیم پرداخت.

۴-۲-۱- روش سلسله مراتبی

این دسته الگوریتم‌ها به صورت تدریجی شروع به خوشه‌بندی می‌کنند؛ بدین صورت که با در نظر گرفتن بعضی از معیارها به تجزیه‌ی سلسله مراتبی داده خواهند پرداخت و سپس با روش‌هایی همچون اجماع و یا تقسیم تغییراتی در خوشه‌بندی اولیه ایجاد می‌کنند. الگوریتم‌های سلسله مراتبی به دو دسته پایین به بالا و بالا به پایین دسته‌بندی می‌شود. در دسته پایین به بالا در شروع کار هر داده به عنوان یک خوشه محسوب می‌شوند و در روند تکرار الگوریتم با ارضا شدن شرط‌هایی که کاربر برای ایجاد خوشه‌ها تعیین کرده است، شروع به اجماع شدن خوشه‌های کوچک می‌شود و این روند تا جایی ادامه پیدا می‌کند تا بالاخره شرط خاتمه ارضا شود. در این نوع دسته شرط‌های مختلفی نظیر تعداد داده‌های هر خوشه، تعداد تکرار، تعداد خوشه‌ها و غیره می‌توان برگزید. از سوی دیگر در دسته بالا به پایین همه نقاط موجود به عنوان یک خوشه کلی در نظر گرفته می‌شود و سپس بر اساس معیارهایی همانند فاصله، خوشه‌های بزرگ را به دو خوشه کوچک‌تر تقسیم‌بندی می‌کند و این عمل تا جایی ادامه پیدا می‌کند که شرط خاتمه محقق گردد. یکی از مزیت‌های روش‌های سلسله مراتبی این است که در

---

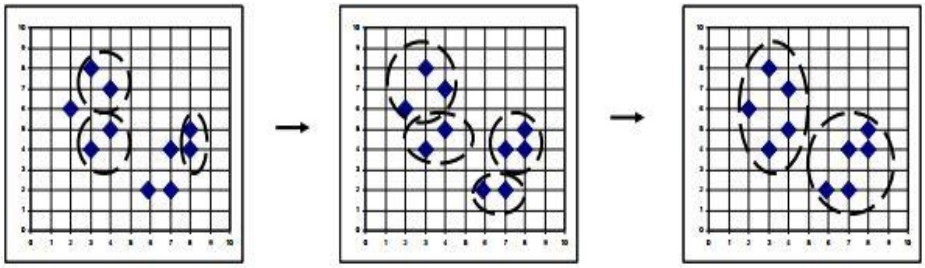
[۱] Hierarchical Method

[۲] Partitioning Method

[۳] Density-Based Method

[۴] Grid-Based Method

هر سطح از سلسله مراتب می‌توان اطلاعات مشخصی را استخراج کرد. همان گونه که در شکل (۴-۲) مشاهده می‌کنید، در این روش خوشه‌بندی به صورت مرحله‌ای انجام می‌پذیرد.



شکل (۴-۲) نمونه‌ای از خوشه‌بندی سلسله مراتبی نوع پایین به بالا

۴-۲-۲- روش افرازی

یکی دیگر از مشهورترین و رایج‌ترین روش‌های خوشه‌بندی، الگوریتم‌هایی هستند که مبتنی بر یافتن نقاط اولیه می‌باشند و به روش‌های افزاری معروف هستند. این نوع الگوریتم‌ها برخلاف الگوریتم-های سلسله مراتبی، مستقیماً عمل خوشه‌بندی را انجام می‌دهند. K-mean یکی از معروف‌ترین الگوریتم‌ها در این دسته می‌باشد که در شروع تعدادی نقطه به عنوان نماینده انتخاب می‌کند و دیگر نقاط را به آن انتساب می‌دهد و سپس با توجه به اعضای خوشه‌ها مراکز هر خوشه را بروز رسانی می‌کند. این بروز رسانی تا جایی ادامه پیدا می‌کند که شرط خاتمه‌ای که کاربر تعیین کرده را برآورده سازد.

۴-۲-۳- روش مبتنی بر چگالی

روش مبتنی بر چگالی، نمونه دیگری از روش‌های خوشه‌بندی می‌باشد که امروزه کاربرد فراوانی دارند. پیدا کردن نواحی پر تراکم و چگال و همچنین توسعه نواحی آن، ایده کلی می‌باشد که تمام الگوریتم‌های مبتنی بر چگالی از آن پیروی می‌کنند. چگونگی تشخیص نواحی چگال و توسعه نواحی آن بعد از یافتن ناحیه چگال فعلی، از عمده تفاوت‌های موجود در الگوریتم‌های معرفی شده در این

زمینه می‌باشد. یکی از اولین الگوریتم‌هایی که در این زمینه معرفی شده است، الگوریتم DBSCAN[1] می‌باشد که یک ساختار داده‌ای به نام R-tree را مورد استفاده قرار می‌دهد تا همسایگی یک نقطه را در مجموعه داده بدست آورد[29].

۴-۲-۴- روش مبتنی بر توری

یکی دیگر از انواع روش‌های خوشه‌بندی، روش مبتنی بر توری می‌باشد. این درسته از الگوریتم‌ها، برای انجام محاسبات از روش‌های توری الهام می‌گیرد؛ به‌طوری که فضای داده‌ها را به فضاهای کوچک‌تر تقسیم‌بندی می‌کنند و تمامی محاسبات لازم به همین شکل تا جایی که شرط خاتمه ارضا شود، ادامه پیدا می‌کند. به عبارت دیگر، برای انجام خوشه‌بندی به روش مبتنی بر توری، هر فضای داده به مجموعه-ای از سلول‌ها تقسیم می‌شود و یک شبکه سلولی ایجاد می‌کند و در اینجا هر سلول شامل اطلاعات آماری در مورد داده‌های متناظر می‌باشد. حال با ادغام سلول‌های متراکم که در همسایگی یک دیگر قرار دارند، خوشه‌بندی انجام می‌شود؛ از این رو مدت پردازش برای خوشه‌بندی مستقل از تعداد داده و متناسب با تعداد سلول‌های این شبکه می‌باشد[30].

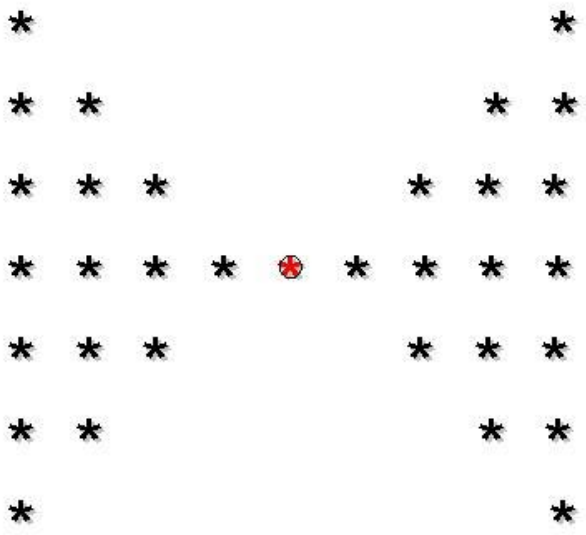
در این روش کیفیت و میزان مصرفی حافظه‌های خوشه‌بندی به اندازه‌ی سلول‌ها وابسته می‌باشد. هر چه سلول‌ها کوچک‌تر باشند کیفیت خوشه‌بندی بیشتر ولی از طرف دیگر پیچیدگی محاسبات و میزان مصرفی حافظه افزایش می‌یابد؛ در نتیجه استفاده از سلول‌هایی با اندازه متغیر و منطبق با تراکم داده‌ها پیشنهاد می‌شود. همچنین سرعت بالای پردازش، عدم وابستگی به ترتیب ورود داده‌ها، امکان تولید خوشه‌هایی با اشکال متفاوت، شناسایی خطاها و عدم نیاز به پیش فرض درباره تعداد خوشه‌ها را می‌توان به عنوان مزیت‌های این نوع روش خوشه‌بندی نام برد[30].

[1] A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise

۴-۲-۵- روش فازی

اساس و پایه خوشه‌بندی فازی بر مبنای نظریه فازی می‌باشد که توسط پرفسور لطفی زاده در سال ۱۹۶۴ به هدف مدل‌سازی عدم قطعیت مطرح گردیده است. بر خلاف خوشه‌بندی افرازی که هر داده ورودی متعلق به یک خوشه می‌باشد و نمی‌تواند عضو خوشه‌های بیشتری باشد، در روش‌های فازی یک نمونه می‌تواند متعلق به بیش از یک خوشه باشد. شکل (۴-۳) یک نمونه خوشه‌بندی فازی را به نمایش گذاشته است.

اگر نمونه‌های ورودی مطابق شکل (۴-۳) باشند مشخص است که می‌توان داده‌ها را به دو خوشه تقسیم کرد اما مشکلی که پیش می‌آید این است که داده مشخص شده در وسط می‌تواند عضو هر دو خوشه باشد بنابراین باید تصمیم گرفت که داده مورد نظر متعلق به کدام خوشه است، خوشه سمت راست یا خوشه سمت چپ. اما اگر از خوشه بندی فازی استفاده کنیم داده مورد نظر با تعلق نیم عضو خوشه سمت راست و با تعلق مشابه عضو خوشه سمت چپ است. تفاوت دیگر در این است که مثلاً داده‌های ورودی در سمت راست این شکل می‌توانند با یک درجه تعلق خیلی کم عضو خوشه سمت چپ نیز باشند که همین موضوع برای نمونه‌های سمت چپ نیز صادق است.



شکل (۴-۳) یک نمونه از خوشه‌بندی فاز

۴-۲-۶- خوشه‌بندی با استفاده از الگوریتم‌های تکاملی

از دید بهینه‌سازی، خوشه‌بندی می‌تواند به عنوان یک نوع خاصی از مسئله گروه‌بندی NP-hard بررسی شود[43]. این الگوریتم‌های فرا ابتکاری می‌توانند در مسائل NP-hard موثر باشند و همچنین قادر هستند از نظر زمانی راه‌حلی بهینه را در این نوع مسائل ارائه بدهند. تحت این فرض، تعدادی زیادی از الگوریتم‌ها تکاملی نظیر الگوریتم ژنتیک، رقابت استعماری، بهینه‌سازی گروه ذرات، پرندگان و غیره را می‌توان برای حل مسئله خوشه‌بندی پیشنهاد کرد. تمام این الگوریتم‌ها بر اساس بهینه‌سازی بعضی از توابع هدف که در اصطلاح تابع شایستگی و یا تابع هزینه نامیده می‌شوند، عمل می‌کنند. در زمینه خوشه‌بندی با استفاده از الگوریتم‌های تکاملی این توابع هزینه، معمولاً معیاری از فاصله می‌باشد.

۴-۳- خوشه‌بندی K-Means

از آنجا که الگوریتم خوشه‌بندی K-Means به عنوان اساس و پایه سایر الگوریتم‌های خوشه‌بندی می‌باشد و همچنین به عنوان یکی از الگوریتم‌های که در ازمایشات از آن برای مقایسه با روش پیشنهادی استفاده شده است، ما را بر آن داشت که این الگوریتم را به طور مختصر شرح دهیم.

یکی از مهم‌ترین اجزای الگوریتم‌های خوشه‌بندی، اندازه‌گیری شباهت‌ها برای جدا کردن دو الگو متفاوت از هم می‌باشد. الگوریتم خوشه‌بندی K-Means داده‌ها را در تعداد خوشه‌های که از قبل تعیین شده است و بر اساس فاصله اقلیدسی که معیار اندازه‌گیری شباهت‌ها در این الگوریتم می‌باشد، گروه‌بندی می‌کند. داده در داخل یک خوشه، فاصله اقلیدسی کمتری نسبت به مرکز خود در مقابل داده دیگر خوشه‌ها دارند و همچنین مراکز خوشه‌ها، میانگین داده همان خوشه می‌باشد.

فلوچارت الگوریتم خوشه بندیK-Means را می‌توان به طور مختصر همانند زیر تشریح کرد:

- تعیین N_c تا مراکز خوشه‌ها به‌صورت تصادفی
- تکرار فرایند پایین

- اختصاص دادن داده به یک کلاس (خوشه) با نزدیک‌ترین داده، جایی که فاصله هر داده با مرکز خوشه خود از فرمول زیر محاسبه می‌شود.

$$d(z_p, m_j) = \sqrt{\sum_{K=1}^{X_d} (E_{pK} - m_{jK})^2} \qquad (۱-۴)$$

که در آن K ابعاد را نشان می‌دهد.

- محاسبه دوباره بردار مرکزی خوشه‌ها با استفاده از فرمول زیر.

$$m_j = \frac{1}{n_j} \sum_{z_p \in C_j} z_p \qquad (۲-۴)$$

- چک کردن شرط توقف

در فرمول‌های بالا، z_p به p امین داده و m_j به بردار مرکزی خوشه j ام اختصاص داده شده است. فرایند این الگوریتم می‌تواند جایی متوقف گردد که یکی از شرط‌های زیر ارضا گردد:

- وقتی ماکزیمم تعداد تکرار تعریف شده با موفقیت به انجام رسیده باشد.
- هنگامی که میزان جابه‌جایی مراکز در همه خوشه‌های از حد آستانه کمتر باشد.
- زمانی که هیچ جابه‌جایی اعضا بین خوشه‌های مختلف رخ ندهد.

الگوریتم‌های خوشه‌بندی K-Means را می‌توان پایگاه‌های داده‌ای بزرگ بکار برد چون پیچیدگی محاسباتی آن نسبت با الگوریتم‌های خوشه‌بندی دیگر کمتر می‌باشد. از معایب این روش یکی این می‌باشد که ما زمانی از این روش می‌توانیم استفاده کنیم که تعداد خوشه‌ها مشخص باشد و همچنین راه خاصی برای پیدا کردن تعداد مناسب خوشه پیدا نشده است. از دیگر معایبی که این روش دارد می‌توان به کارا نبودن آن در زمانی که شکل خوشه‌ها از پیچیدگی خاصی برخوردار هستند. البته یکی از مهم‌ترین نقطه ضعف این روش این است که در برابر داده‌های پرت و به اصطلاح نویزی زیاد حساس می‌باشد چون‌که این داده‌ها به راحتی مراکز خوشه‌ها را تحت تأثیر قرار می‌دهند و ممکن است اثرات نامطلوبی را بر روی عملکرد سیستم داشته باشد.

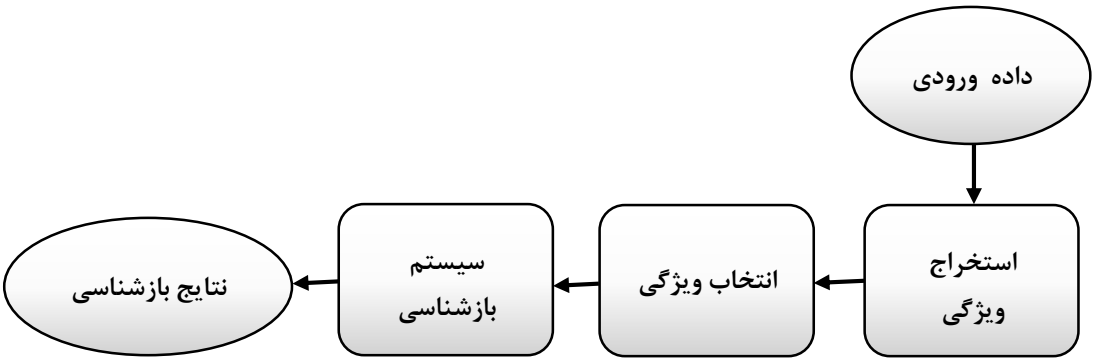
فصل پنجم:

معرفی روش‌های پیشنهادی

۵-۱- مقدمه

پس از آن که در فصول قبل به مقدمات و معرفی ابزارهای مورد نیاز پرداختیم، در این فصل به معرفی روش‌های پیشنهادی خواهیم پرداخت. استخراج و انتخاب ویژگی همیشه به عنوان یکی از زمینه‌های تحقیقاتی محققان در علوم گوناگون برای طبقه‌بندی و دسته‌بندی داده‌ها مورد توجه بوده است. همچنین، هریک از روش‌های پیشنهادی توسط محققان کاربرد خاص خود را دارا می‌باشد و در نوع خود باعث بهبود سیستم‌ها گردیده است. در سیستم‌های تشخیص گوینده و یا به طور کلی‌تر سیستم‌های تشخیص گفتار روش‌هایی همانند ضرایب کپسترال فرکانسی مل، ضرایب کپسترال پیش-گویی خطی، پیش‌گویی ادراک خطی، انرژی و غیره به عنوان استخراج ویژگی شناخته می‌شوند. این ویژگی‌ها در گذر زمان باعث بهبود عملکرد سیستم‌های پردازش گفتار شده‌اند و توانسته‌اند محدودیت-های بیشتری را از بین ببرند تا بتوانیم از این سیستم‌ها در محیط‌های طبیعی و با داده‌های بیشتر استفاده کنیم.

واضح است که همه ویژگی‌های استخراج شده برای شناسایی الگو در سیستم‌های تشخیص مفید واقع نمی‌شود و یا از ضریب یکسانی برخوردار نمی‌باشند؛ به همین خاطر بعد از استخراج ویژگی از بلوک انتخاب ویژگی برای برگزیدن ویژگی مناسب استفاده می‌کنند. پس هدف انتخاب ویژگی، بهبود عملکرد سیستم و یا حفظ دقت عملکرد سیستم با استفاده از کمترین داده، می‌باشد. بنابراین، استفاده از کمترین داده‌ای که باعث بهبود عملکرد سیستم می‌باشد مهم و ضروری می‌باشد، زیرا باعث کم شدن پیچیدگی سیستم‌های تشخیص الگو می‌شود و از طرفی سرعت روند تشخیص را هم افزایش می‌دهد. روند توضیح داده شده در شکل (۵-۱) آمده است.



شکل (۵-۱) بلوک دیاگرام تشخیص الگو

برای داشتن یک سیستم تشخیص گوینده، باید به دنبال ویژگی‌هایی باشیم که بیشتر خصوصیات مجرای صوتی فرد را شبیه سازی کرده و بجای وابسته بودن به جملات ادا شده به خصوصیات فیزیکی دستگاه‌ها صوتی هر فرد وابسته باشد. به همین خاطر، در این پروژه به دنبال این موضوع بوده‌ایم که خصوصیات و بردارهای که بیشتر از همه تکرار شده را برای مدل سازی و یا آموزش شبکه از آن استفاده کنیم. که در اینجا دو روش برای انجام این کار پیشنهاد شده است.

۵-۲- روش پیشنهادی اول

برای دست یافتن به اهداف بالا، قبل از قسمت انتخاب ویژگی، در قسمت استخراج ویژگی باید به دنبال ویژگی باشیم که خصوصیات فیزیکی و ویژگی‌های مجرای صوتی را شبیه سازی می‌کند. ما ضرایب کپسترال فرکانسی مل را به عنوان ویژگی مورد نظرمان استفاده کرده‌ایم، چون همان طور که در [5] و [19] آمده است، این ویژگی نه تنها توزیع فرکانسی مشخص کننده صدا را حمل می‌کند، بلکه عملکرد اندازه و شکل مجرای گلو[۱] و دهانه نای[۲] هر فرد را مشخص می‌کند. بنابراین این ویژگی به نوعی خصوصیات فردی مجرای صوتی هر فرد را مشخص می‌کند. بنابراین بعد از پیش پردازش‌های ذکر شده

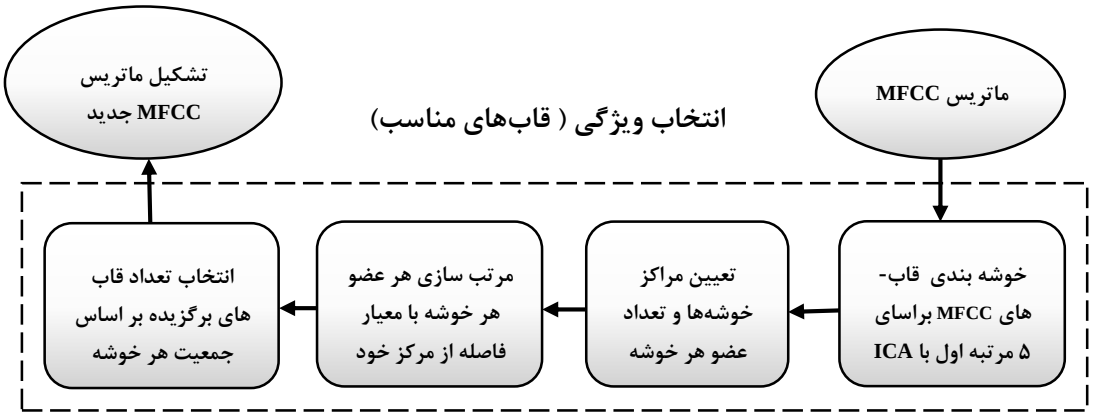
[۱] Vocal Tract

[۲] Glottal Source

در فصل‌های قبل، اقدام به قاب‌بندی گفتار ورودی می‌کنیم و سپس از قاب‌های بدست آمده ۱۳ مرتبه اول ضرایب کپسترال فرکانسی مل را به همراه دلتا و دلتا دو را محاسبه می‌کنیم. در آخر این بردارهای بدست آمده از هر قاب را در کنار هم قرار می‌دهیم تا برای هر جمله‌ای که توسط گوینده ادا می‌شود ما یک ماتریس MFCC، همانند ماتریس (۵-۱) داشته باشیم. در این ماتریس m نشان دهنده بعد هر قاب می‌باشد و n نشان دهنده‌ی تعداد قاب‌های MFCC هر جمله را مشخص می‌کند.

$$\begin{bmatrix} M_{1,1} & \cdots & M_{1,n} \\ \vdots & \ddots & \vdots \\ M_{m,1} & \cdots & M_{m,n} \end{bmatrix} \qquad (۱-۵)$$

در شکل(۶-۲) بلوک دیاگرام روش پیشنهادی اول را مشاهده می‌کنید. همان‌طور که در این شکل آمده است، بعد از تشکیل ماتریس ضرایب کپسترال فرکانسی مل یا MFCC، این ماتریس جهت انتخاب قاب‌هایی که بیشترین شباهت (تکرار)را باهم دارند به قسمت انتخاب ویژگی داده می‌شود. انتخاب ویژگی در این روش از ۴ قسمت اصلی تشکیل می‌شود که در ادامه به توضیح هر یک از بلوک خواهیم پرداخت.



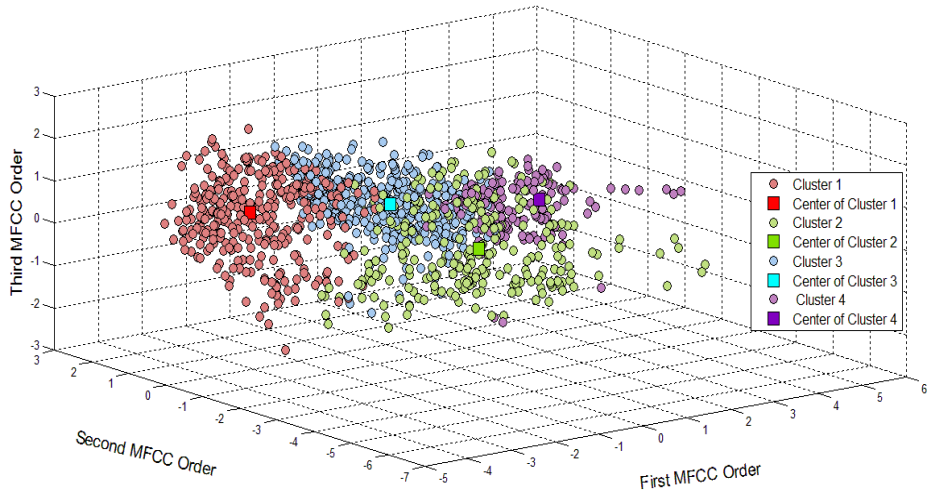
شکل (۵-۲) نمودار بلوکی روش پیشنهادی اول

۵-۲-۱- خوشه‌بندی قاب‌های MFCC با استفاده از ICA

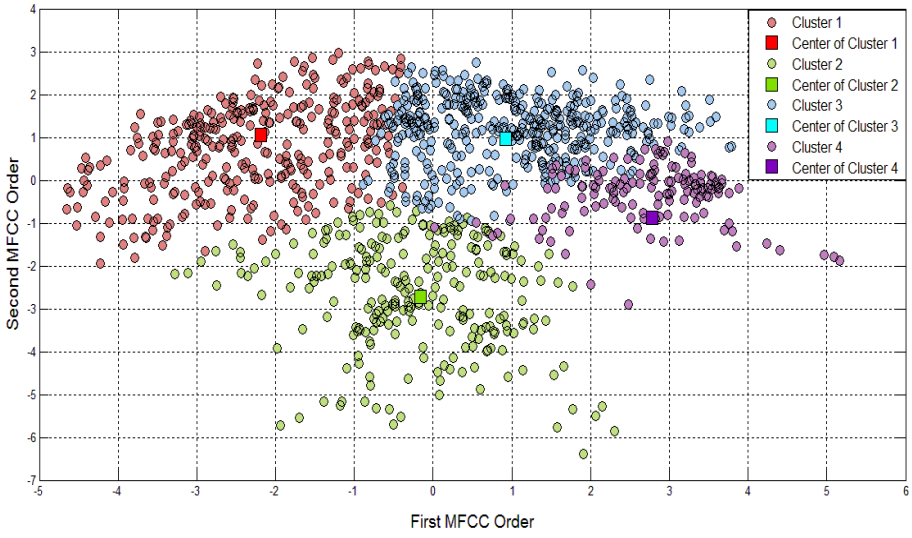
در این قسمت، هر یک از قاب‌های MFCC را به عنوان یک نمونه از عملکرد مجرای صوتی هر فرد می‌شناسیم. پس برای انتخاب بردارهای ویژگی مشابه یا پر تکرار از خوشه یابی با الگوریتم رقابت استعماری استفاده می‌کنیم. همان‌طور که در فصل قبل بیان شده به علت بالا بودن ابعاد هر داده یا همان قاب MFCC از این الگوریتم برای خوشه‌بندی استفاده می‌شود. در این مرحله هر یک از مرتبه‌های قاب MFCC به عنوان یک بعد در نظر گرفته می‌شود، پس اگر از هر قاب در قسمت استخراج ویژگی ۱۳ مرتبه اول MFCC، دلتا و دلتا دو را حساب کرده باشیم، ما یک داده یا بردار MFCC ۳۹ بعدی خواهیم داشت. همچنین اگر هم ما به ازای یک جمله به طور مثال ۱۰۰۰ قاب MFCC داشته باشیم، باید این ۱۰۰۰ قاب MFCC را که هر داده دارای ۳۹ بعد می‌باشد با استفاده از الگوریتم رقابت استعماری خوشه‌بندی کنیم. البته به خاطر این که در مرتبه‌های بالاتر MFCC، خصوصاً در دلتا و دلتا دو دامنه کاهش می‌یابد تأثیر زیادی در روند خوشه یابی ندارد و برای همین می‌توان فقط از چند مرتبه اول برای خوشه یابی استفاده کرد تا حجم و پیچیدگی محاسباتی کاهش پیدا کند.

همان‌طور که در شکل (۵-۳) و شکل (۵-۴) مشخص می‌باشد، به ترتیب نمایش سه بعدی و دو بعدی خوشه‌بندی پنج مرتبه اول قاب‌های MFCC را برای یک جمله در چهار خوشه متفاوت را نمایش می‌دهد. همان‌طور که مشخص می‌باشد، تعداد خوشه‌های مناسب می‌تواند از دیگر عوامل موثر در تعیین خصوصیات(قاب‌های MFCC مشابه) پر تکرار ماهیچه‌های فیزیکی صوت باشد که در فصل بعد به چگونگی تاثیر آن خواهیم پرداخت. این کار برای تمام جملات آموزش و تست انجام می‌شود. در شکل (۵-۳) و (۵-۴) به‌خوبی مشخص است که تعداد زیادی از داده‌ها از مراکز خود به طور قابل ملاحظه‌ای دور می‌باشند و حتی اگر ما تعداد خوشه‌های موجود را هم افزایش دهیم دوباره این داده‌های به اصطلاح پرت را مشاهده خواهیم کرد. استفاده از داده‌هایی که به مراکز خوشه‌ها نزدیک هستند، به نوعی یک

همگرایی بهتری برای مدل کردن افراد در سیستم‌های مدل‌سازی و یا وزن‌های شبکه در قسمت آموزش ایجاد می‌کنند.



شکل (۵-۳) نمایش سه بعدی خوشه بندی پنج مرتبه اول(بعد) قاب‌های MFCC با استفاده از ICA در چهار خوشه



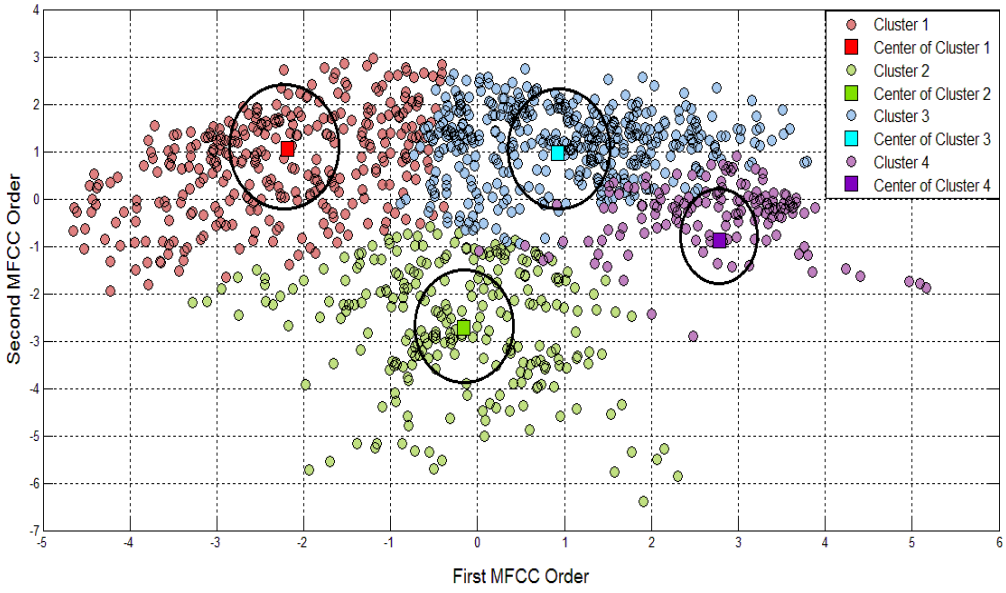
شکل (۵-۴) نمایش دو بعدی خوشه بندی پنج مرتبه اول قاب‌های MFCC با استفاده از ICA در چهار خوشه

۵-۲-۲-تعیین مراکز و تعداد اعضای خوشه‌ها

پس از خوشه‌بندی به تعیین مراکز خوشه‌ها می‌پردازیم که این کار از میانگین گرفتن از ابعاد مختلف تمام داده‌های موجود در هر خوشه بدست می‌آید. و بعد از این کار تعداد کل اعضای هر خوشه را مشخص می‌کنیم که این کار مشخص می‌کند که چه بردارهای ویژگی صوتی به طور نسبی بیشتر از بقیه بردارهای ویژگی تکرار شده است.

۵-۲-۳-مرتب سازی داده‌ها و یا بردار MFCC

با بدست آوردن مراکز خوشه‌ها فاصله هریک از اعضای خوشه‌ها را با مرکز آن را محاسبه می‌کنیم و با استفاده از معیار کمترین فاصله هر داده یا قاب‌های MFCC از مرکز خود در هر خوشه را مرتب‌سازی می‌کنیم. وقتی که در مرحله بعد سهم هر خوشه مشخص شد، بر اساس همین مرتب سازی شروع به انتخاب قاب MFCC خواهیم کرد. به طور مثال اگر سهم خوشه اول N تا باشد، ما N تا داده یا قاب MFCC را که به مرکز خوشه اول نزدیک‌تر می‌باشند را انتخاب می‌کنیم و در ماتریس MFCC جدید جای خواهیم داد. در شکل (۵-۵) چگونگی روند بالا را نشان می‌دهد و همچنان که مشخص می‌باشد تعداد مناسب خوشه‌ها در تعیین بردارهای ویژگی مشابه و پر تکرار خیلی موثر می‌باشد. البته نا گفته نماند این یک نمایش ۲ بعد اول از ۳۹ بعد هر قاب MFCC می‌باشد ولی برای خوشه بندی از ۵ بعد این بردار استفاده کردیم



شکل (۵-۵) نحوه انتخاب قاب MFCC از هر خوشه

۵-۲-۴-نحوه و تعداد انتخاب قاب MFCC از هر خوشه

روش‌های متفاوتی نظیر تعداد اعضای هر خوشه، معیار پراکندگی، موقعیت خوشه و غیره را می‌توان برای محاسبه سهم هر یک از خوشه‌ها در ماتریس MFCC جدید لحاظ کرد. در اینجا ما سهم هر یک از خوشه‌ها را در ماتریس جدید بر اساس تعداد اعضای خوشه‌ها مشخص کرده‌ایم؛ بدین ترتیب که هر چه تعداد اعضای یکه خوشه بیشتر باشد در تشکیل ماتریس جدید سهم بیشتری را به خود اختصاص می‌دهد. در واقع با این کار قصد داشتیم بردارهای ویژگی مجرای صوتی که بیشتر توسط افراد تکرار شده، دارای وزن بیشتری در مدل سازی آن داشته باشد. همچنان که در فرمول (۵-۱) آمده است، نحوه محاسبه ماتریس MFCC جدید را مشاهده می‌کنید.

$$\text{ماتریس MFCC جدید} = [\ \alpha_1\ (\text{تعداد اعضای خوشه اول}) + \cdots + \alpha_n\ (\text{تعداد اعضای خوشه n ام})\ ] \times N \qquad (۵-۲)$$

در فرمول بالا N تعداد کل قاب MFCC جدید که مورد نظر ما می‌باشد، خوشه اول بیشترین اعضا را دارا می‌باشد، α_1 ضریب خوشه یک می‌باشد که سهم خوشه یک را از کل ماتریس MFCC جدید

را مشخص می‌کند. در این بین α_1 بیشترین مقدار و α_n کمترین مقدار را دارا می‌باشد و همچنین شرط‌های زیر هم باید رعایت شود.

$$\alpha_m = \frac{\text{تعداد اعضای خوشه } m}{\text{کل اعضای}} \qquad m = 1,2,3,\ldots,n \qquad (۳-۵)$$

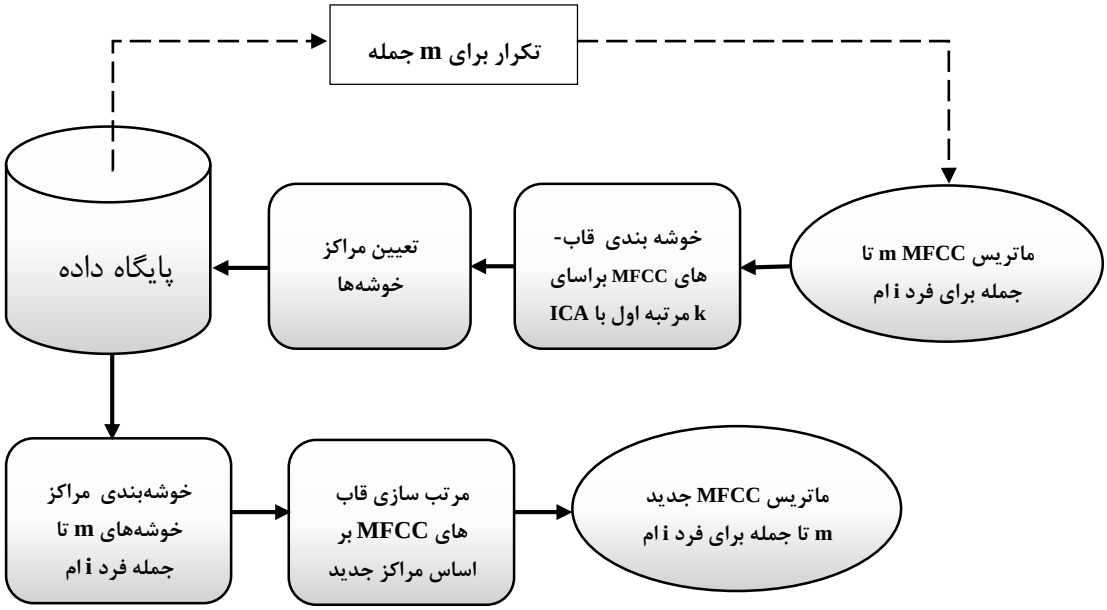
$$\alpha_1 > \alpha_2 > \cdots > \alpha_n \qquad (۴-۵)$$

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1 \qquad (۵-۵)$$

در آخر، در ماتریس جدید همه مرتبه‌های MFCC، دلتا و دلتا دو لحاظ شده است یعنی ما یک ماتریس $39 \times N$ خواهیم داشت. در واقع ما برای این که از حجم محاسباتی بکاهیم از ۵ مرتبه اول جهت خوشه یابی استفاده کرده‌ایم. بعد از اینکه ماتریس MFCC جدید را تشکیل داده‌ایم، این ماتریس جهت مدل‌سازی گفتار افراد و یا آموزش شبکه جهت تشخیص افراد آمده می‌باشد. این روند برای قسمت تست هم انجام می‌شود با این تفاوت که در قسمت تست سهم هر خوشه را برابر در نظر گرفته شده است. در فصل بعدی نتایج شبیه‌سازی این روش را با تعداد مختلف قاب‌های MFCC در قسمت تست بررسی خواهیم کرد.

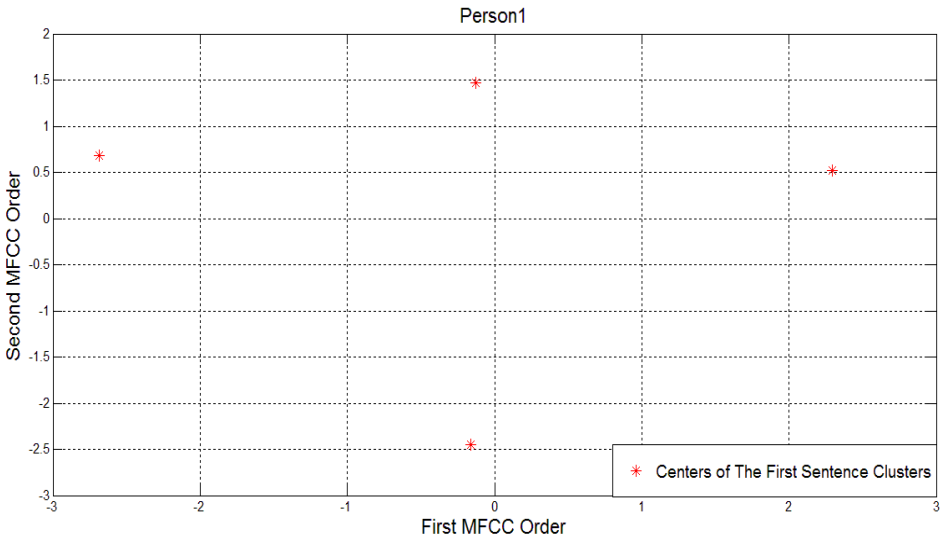
۵-۳- روش پیشنهادی دوم

در روش دوم هم همانند روش اول جهت یافتن قاب‌های MFCC که باعث همگرایی بیشتر در مدل‌سازی خصوصیات مجرای صوتی افراد می‌شود، قدم برمی‌داریم. در این روش مطابق آنچه که شکل (۵-۶) نشان داده شده است قسمت‌های اولیه همانند قسمت قبل می‌باشد؛ بنابراین به طور مختصر این قسمت‌ها را توضیح خواهیم داد و سپس قسمت‌هایی را توضیح خواهیم داد که این روش را با روش پیشنهادی اول متفاوت می‌سازد.

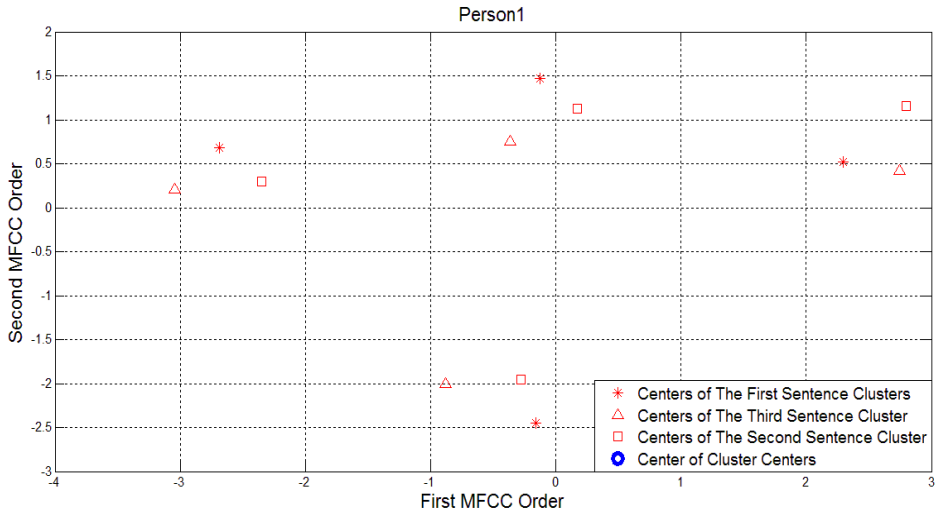


شکل (۵-۶) نمودار بلوکی روش پیشنهادی دوم

در ابتدا ماتریس MFCC جمله اول فرد i ام را برای خوشه یابی به الگوریتم رقابت استعماری می‌دهیم تا خوشه‌ها و مراکز آن تعیین شود. طبق آنچه که از قبل گفته شده است می‌توانیم از k مرتبه اول قاب MFCC استفاده کنیم که این k مرتبه MFCC می‌تواند متغیر باشد. در اینجا ما از ۱۳ مرتبه اول MFCC، دلتا و دلتا دو استفاده کردیم، پس k=۳۹ می‌باشد. k مناسب که هم حجم محاسباتی کمتری را در بر می‌گیرد و نتایج مطلوبی را از خود در آزمایشات نشان داده است، $2 \le k \le 7$ می‌باشد. بعد از خوشه بندی و مشخص شدن مراکز خوشه‌ها هر جمله فرد iام، اقدام به ذخیره سازی مراکز خوشه‌ها خواهیم کرد و این کار را برای تمام افراد انجام خواهیم داد. شکل (۵-۷) مراکز خوشه‌ها را برای یک جمله نشان می‌دهد.



شکل(۵-۷) مراکز خوشه‌ها برای یک جمله

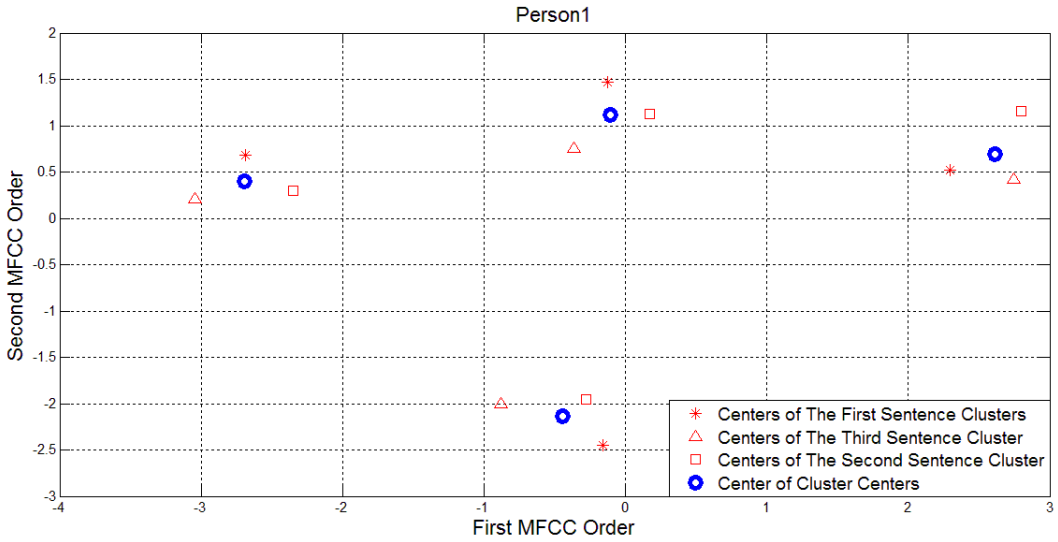


شکل (۵-۸) مراکز خوشه‌های مختلف برای سه جمله

همان طور که از شکل (۵-۸) پیداست، سه جمله را برای مدل‌سازی و یا آموزش انتخاب کرده‌ایم که به ترتیب با شکل‌های ستاره، مربع و مثلث مشخص می‌باشد و هر یک از جمله‌ها به ۴ خوشه مختلف تقسیم شده است. بعد از این که عملیات خوشه‌بندی m تا جمله فرد i ام تمام شد، به سراغ مراحل زیر خواهیم رفت.

۵-۳-۱-خوشه‌بندی مراکز خوشه‌ها

در اینجا دوباره با استفاده از الگوریتم رقابت استعماری خوشه یابی را بر روی مراکز خوشه‌های تمام جملات فرد i ام انجام می‌دهیم. برای نمونه، همان‌طور که در شکل (۵-۹) مشاهده می‌کنید، میانگین مراکز خوشه‌های متناظر تعیین شده است. همچنین، می‌توان به این نتیجه پی برد که وقتی ما چند جمله متفاوت که توسط یک شخص بیان شده است را با تعداد خوشه‌های برابر تقسیم می‌کنیم، مرکز خوشه‌ها متناظر به هم نزدیک می‌باشند. به معنی دیگر بردارهای ویژگی مجرای صوتی هر فرد با بیان جمله‌های مختلف تغییرات محسوسی نمی‌کند و مراکز خوشه‌های متناظر جملات مختلف در کنار هم قرار می‌گیرند. این نشان دهنده این است که وقتی جمله‌های مختلفی توسط یک فرد بیان می‌شود، بیشتر مؤلفه‌های مختلف قاب MFCC، بین یک دامنه‌های خاصی می‌باشند.



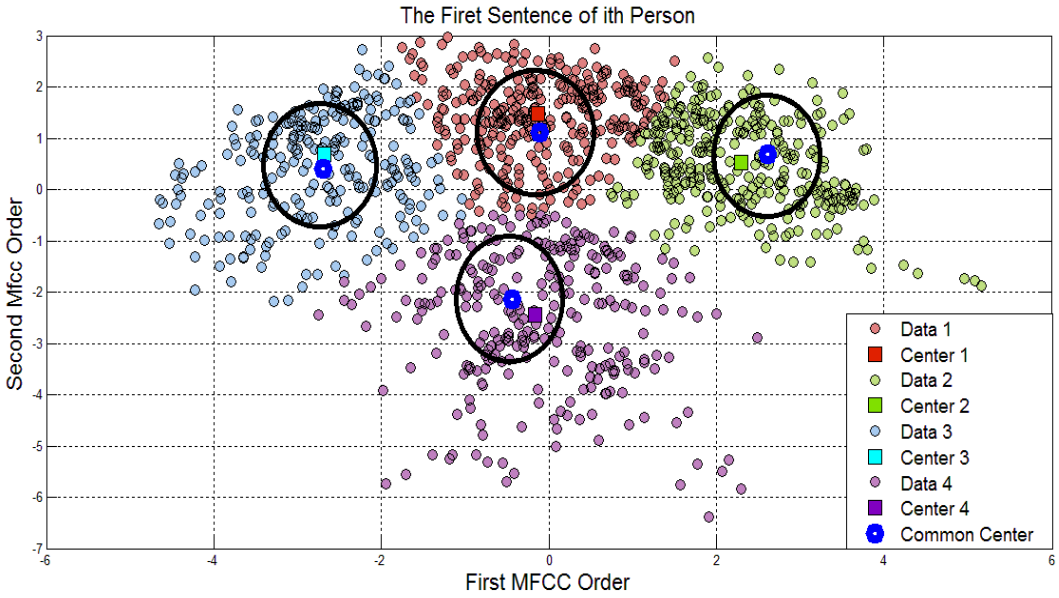
شکل (۵-۹) نمایش دوبعدی نحوه خوشه‌بندی مراکز جملات و تعیین مراکز جدید

۵-۳-۲-مرتب‌سازی بر اساس مراکز جدید

حال با بدست آوردن مراکز جدید، داده و یا همان قاب MFCC تک‌تک m تا جمله که توسط شخص i ام ادا شده را با استفاده از این مراکز و بر اساس معیار کمترین فاصله مرتب می‌کنیم تا

نزدیک‌ترین داده‌ها به مراکز جدید را تعیین کرده باشیم. ماتریس جدید MFCC با استفاده N تا داده که نزدیک‌تر به مراکز جدید می‌باشند، تشکیل می‌شود و همچنین این ماتریس برای هر یک از جمله‌ها با مراکز مشترک ساخته خواهد شد؛ در واقع به ازای هر یک از جمله‌ها یک ماتریس MFCC خواهیم داشت.

در شکل (۵-۱۰) یک نمونه از نحوه انتخاب داده‌ها در ماتریس MFCC جدید را مشاهده می‌کنید و همان‌طور که در این شکل مشخص کرده‌ایم و در بالا ذکر شده است ما بر طبق مراکز مشترک که در روند بالا بدست آمده است نزدیک‌ترین داده‌ها را انتخاب می‌کنیم.



شکل (۵-۱۰) نحوه انتخاب داده‌ها در ماتریس MFCC جدید بر اساس مراکز جدید (آبی) را نشان می‌دهد

در اینجا هم مثل روش پیشنهادی اول به نسبت اعضای خوشه‌ها جدید سهم هر یک از مراکز در ماتریس MFCC را مشخص کرده‌ایم. همچنان که در شکل (۵-۹) مشاهده می‌کنید مراکز جدید دارای ازایی برابر می‌باشند و لذا سهم هر یک از خوشه‌ها در ماتریس MFCC جدید برابر می‌باشد. حال این ماتریس برای مدل‌سازی افراد و یا آموزش شبکه که بازشناسی افراد آماده می‌باشد.

۵-۳-۳- نحوه انتخاب قاب‌های MFCC در مرحله تست

انتخاب قاب‌های MFCC در مرحله تست در دو روش یکسان می‌باشد؛ بدین صورت که بعد از خوشه بندی و تعیین مراکز در جمله تست، درصدی از قاب‌های MFCC نزدیک‌تر به مرکز در هر خوشه برای تست استفاده می‌شود. این کار دو مزیت دارد، اولین مزیت این کار این است که ما بیشتر روی قاب‌های MFCC تمرکز می‌کنیم که بیشتر آموزش دیده‌اند و دوم این که با هرس کردن داده‌هایی که از مرکز خوشه‌ها فاصله دارند از پیچیدگی محاسباتی و هزینه زمانی کم کرد. در آزمایش‌هایی که بر روی این قضیه انجام شد نتایج نشان می‌دهد که استفاده از ۰,۴۵ تا ۰,۷ از قاب‌های نزدیک‌تر به مراکز خوشه‌ها در روش پیشنهادی اول و ۰,۶ تا ۰,۹ از قاب‌های نزدیک‌تر به مراکز خوشه‌ها در روش پیشنهادی دوم می‌تواند ضریب مناسبی برای انتخاب قاب‌ها جهت تست باشد.

فصل ششم:

نتایج و جمع‌بندی

۶-۱- مقدمه

بعد از شرح کامل دو الگوریتم پیشنهادی در فصل قبل، در این فصل نتایج این الگوریتم‌ها را مشاهده می‌کنید و عملکرد آن‌ها را با دیگر روش‌ها بررسی می‌کنیم. قبل از بررسی این نتایج، به معرفی داده‌گانی که از آن استفاده کرده‌ایم خواهیم پرداخت. در آخر، جمع‌بندی و پیشنهادات را ارائه خواهیم کرد.

۶-۲- داده‌گان

یکی از عواملی که در دقت تشخیص گفتار و همچنین بازشناسی گوینده می‌تواند خیلی موثر باشد، استفاده از یک داده‌گان مناسب می‌باشد. در این زمینه، می‌توان داده‌گان‌های مختلفی را نام ببرد که عمدتاً در سه دسته گفتار مجزا، گفتار منقطع و پیوسته تقسیم‌بندی می‌شود. در داده‌گان نوع اول، یک سری کلمات جدا از هم تشکیل شده است. این نوع داده‌گان در اویل پیدایش سیستم‌های تشخیص گفتار و به خاطر اینکه از پیچیدگی‌های سیستم تشخیص بکاهند، مورد استفاده قرار می‌گرفت. همان طور که مشخص می‌باشد سیستم‌هایی که از این نوع داده‌گان استفاده می‌کنند از درصد بازشناسی خوبی برخوردار هستند. داده‌گان نوع دوم، از جمله‌هایی تشکیل که در آن بین کلمات از سکوت‌های مصنوعی برای جدا کردن کلمات تشکیل دهنده جمله استفاده شده است. اما یک سری داده‌گان وجود دارد که از جملات پیوسته و صحبت طبیعی انسان‌ها تشکیل شده است. در این نوع داده‌گان تشخیص به مراتب سخت‌تر می‌باشد، چون مرز بین کلمات ادا شده در آن مشخص نبوده و در این بین یک سری اصوات می‌تواند قرار بگیرد. به طور کلی، در داده‌گان نوع سوم، می‌توان حالت گفتار طبیعی انسان را مشاهده کرد. داده‌گانی که بتواند شرایط لازم برای هریک نوع از داده‌گان را ایجاد کند را می‌توان به عنوان یک داده‌گان مناسب و خوب اطلاق کرد.

در اینجا ما از داده‌گان ELSDSR[1] استفاده کرده‌ایم که از ۲۲ گوینده تشکیل شده است که ۱۰ نفر از آن‌ها زن و ۱۲ نفر مرد می‌باشند. افراد تشکیل دهنده این مرجع در بازه سنی ۲۴ تا ۶۳ سال می‌باشند که متوسط سنی ۴۰,۶ برای افراد زن و ۳۰,۴ برای مردها ثبت شده است. در این داده‌گان هر گوینده ۷ جمله (عبارت) را برای قسمت آموزش و ۲ جمله (عبارت) را برای قسمت تست ادا می‌کند. ۷ جمله‌ای که در قسمت آموزش قرار دارد برای همه افراد یکسان می‌باشد ولی در قسمت تست جملات ادا شده توسط همه گوینده‌ها متفاوت می‌باشد. بنابراین ما در اینجا یک سیستم شناسایی گوینده مستقل از متن داریم؛ زیرا در قسمت تست از جملاتی استفاده شده است که قبلا در قسمت آموزش استفاده نشده است. فرکانس نمونه برداری داده‌گان ELSDSR، ۱۶ KH می‌باشد و در دانشگاه TUD[2] طراحی و ساخته شده است. در شکل (۶-۱) مدت زمان کل ادا کل جملات در قسمت آموزش و تست برای هر گوینده مشخص شده است.

جدول (۶-۱) نمایش مشخصات هر گوینده در داده‌گان ELSDS [31]

No.	Male		Train(s)	Test(s)	Female		Train(s)	Test(s)
1		MASM	81.2	20.9		FAML	99.1	18.7
2		MCBR	68.4	13.1		FDHH	77.3	12.7
3		MFKC	91.6	15.8		FEAB	92.8	24.0
4		MKBP	69.9	15.8		FHRO	86.6	21.2
5		MLKH	76.8	14.7		FJAZ	79.2	18.0
6		MMLP	79.6	13.3		FMEL	76.3	18.2
7		MMNA	73.1	10.9		FMEV	99.1	24.1
8		MNHP	82.9	20.3		FSLJ	80.2	18.4
9		MOEW	88.0	23.4		FTEJ	102.9	15.8
10		MPRA	86.8	9.3		FUAN	89.5	25.1
11		MREM	79.1	21.8				
12		MTLS	66.2	14.05				

[1] English Language Speech Database For Speaker Recognition

[1] Technical University of Demark

۶-۳- ارائه نتایج روش‌های پیشنهادی

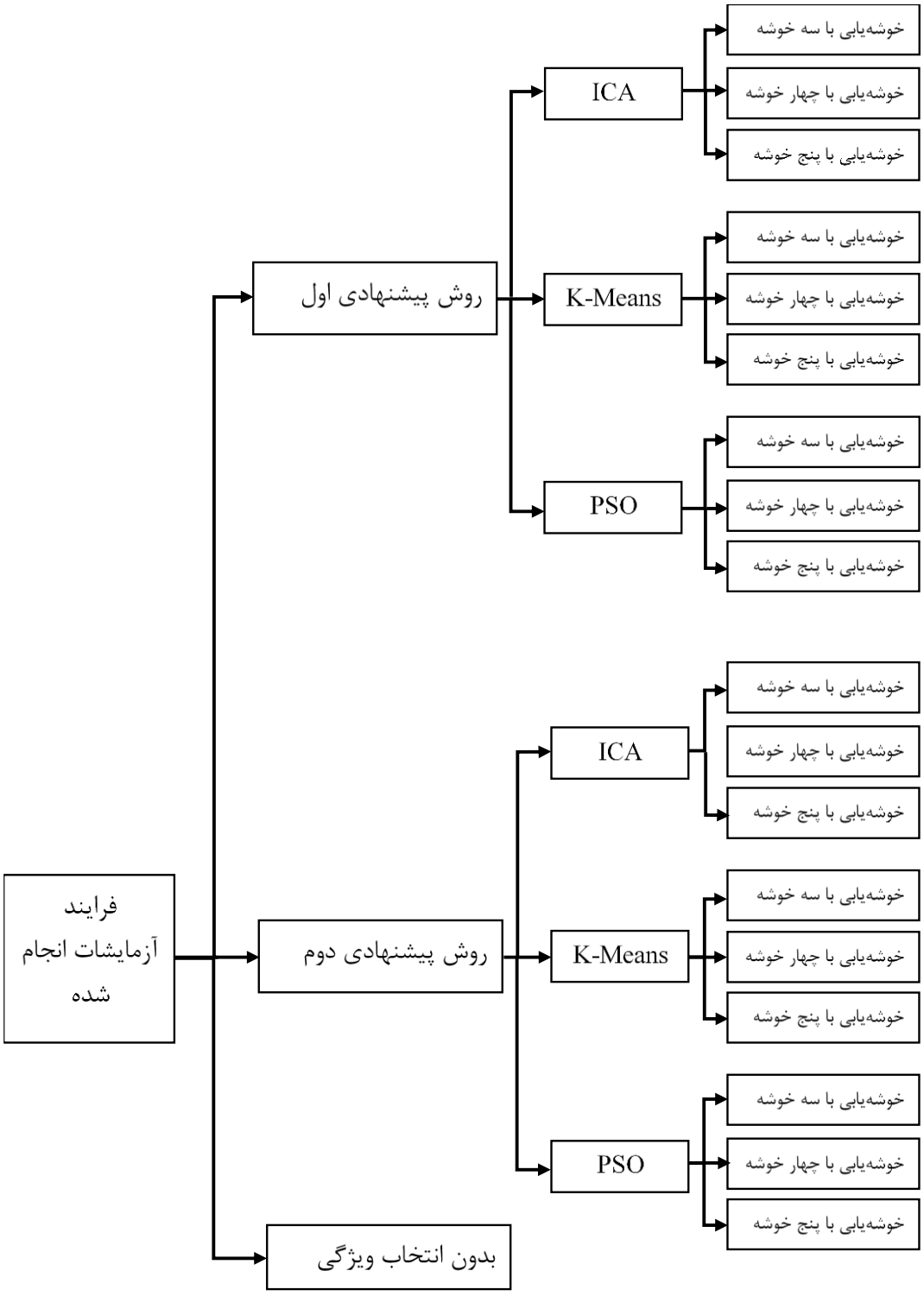
جهت بررسی عملکرد روش‌های پیشنهادی نیاز داریم داده‌گان حاضر را بروی یک سیستم شناسایی گوینده با اعمال روش‌های پیشنهادی ذکر شده، آزمایش کنیم. در اینجا ما از یک شبکه عصبی دو لایه پرسپترون پس انتشار خطا[۱] به عنوان طبقه بند استفاده کرده‌ایم. لایه ورودی این شبکه از ۳۹ ورودی به همراه یک ورودی به عنوان بایاس تشکیل شده است و همچنین در لایه خروجی به ازای هر یک از از افراد یک خروجی در نظر گرفته شده است. نهایتا از ۵۰ نرون در لایه مخفی استفاده شده است که این عدد به طور آزمون و خطا بدست امده است. همان طور که در فصل قبل اشاره شده است بعد از تشکیل ماتریس MFCC جدید برای هر فرد، این ماتریس جهت آموزش، به شبکه داده می‌شود. در این سیستم زمانی یک فرد به عنوان برنده انتخاب می‌شود که بیشترین تعداد قاب‌های MFCC را در خروجی این شبکه به خود اختصاص دهد.

فرکانس نمونه‌برداری جمله‌های ادا شده در داده‌گان حاضر ۱۶ کیلو هرتز می‌باشد و هر جمله با استفاده از یک قاب ۲۵ میلی ثانیه، قاب‌بندی شده است که در هر بار، قاب ۱۲,۵ میلی ثانیه جابه‌جا می‌شود. بعد از مرحله قاب گذاری، از تمام قاب‌های ایجاد شده ویژگی MFCC را استخراج می‌کنیم و سپس با استفاده از روش‌های پیشنهادی اقدام به ایجاد ماتریس MFCC جدید خواهیم کرد.

در شکل (۱-۱) روند آزمایشات انجام شده را مشاهده می‌کنید؛ چنانچه که در این شکل مشاهده می‌کنید الگوریتم‌های پیشنهادی با استفاده از ICA با الگوریتم‌های خوشه‌بندی k-Means و الگوریتم خوشه‌بندی PSO مقایسه شده است و همچنین این دو روش پیشنهادی با استفاده از ICA با روش‌های SVM، optimaized ELM و روش بدون بلوک انتخاب ویژگی مقایسه شده است.

---

[۱] Back-Propagation



شکل (۶-۱) فرایند آزمایش‌های انجام شده

۶-۳-۱- نتایج روش پیشنهادی اول

قبل از ارائه نتایج و مشاهدات روش پیشنهادی اول، به بررسی نحوه ارزیابی می‌پردازیم. در این آزمایش، ۲۲ گوینده وجود دارد که سه جمله از هر گوینده برای آموزش شبکه عصبی استفاده می‌شود و هر کدام از گوینده‌گان ۲ جمله را در قسمت تست ادا می‌کنند و هر فرد در صورتی به عنوان برنده معرفی می‌شود که بیشترین تعداد قاب‌های MFCC در قسمت تست به خود اختصاص دهد. به طور مثال، جدول (۶-۲) تعداد قاب‌های MFCC اختصاص داده شده به هر فرد زمانی که دو جمله تست مربوط به فرد اول به سیستم داده شده است را نشان می‌دهد. طبق جدول (۶-۲) تعداد قاب‌هایی که گوینده اول به خود اختصاص داده است در هر دو جمله تست نسبت به سایر گوینده‌گان بیشتر می‌باشد؛ از این رو گوینده اول در هر دو بار توسط سیستم درست تشخیص داده شده است. این کار برای تمام ۲۲ گوینده تکرار می‌شود تا جمله‌هایی که درست تشخیص داده شده است مشخص شود.

در این جدول علاوه بر تعداد کل قاب‌های اختصاص داده شده به هر فرد، تعداد قاب‌های اختصاص داده شده به هر فرد به تفکیک خوشه‌ها هم نشان داده شده است که در قسمت‌های بعد در مورد این موضوع هم توضیحاتی ارائه خواهیم کرد.

جدول (۶-۲) تعداد قاب‌های اختصاص داده شده به هر یک از گوینده‌ها، زمانی که دو جمله تست توسط گوینده اول ادا می‌شود

اسامی گوینده‌گان	جمله تست اول					جمله تست دوم				
	خوشه اول	خوشه دوم	خوشه سوم	خوشه چهارم	تعداد کل	خوشه اول	خوشه دوم	خوشه سوم	خوشه چهارم	تعداد کل
گوینده ۱	۵۳	۴۰	۱۷	۴۶	**۱۵۶**	۲۳	۲۸	۰	۳۸	**۸۹**
گوینده ۲	۱۳	۲۴	۱۴	۴۰	**۹۱**	۳۳	۲۷	۲	۸	**۷۰**
گوینده ۳	۱۶	۴۶	۴	۳	**۶۹**	۱۰	۲	۲	۱۰	**۲۴**
گوینده ۴	۴	۱	۲	۰	**۷**	۰	۶	۰	۴	**۱۰**
گوینده ۵	۲۷	۳۷	۲۰	۲۰	**۱۰۴**	۴۳	۵	۱	۲۵	**۷۴**
گوینده ۶	۵	۱۲	۴	۳۰	**۵۱**	۱۱	۲۵	۰	۵	**۴۱**
گوینده ۷	۱۵	۱۹	۳	۴	**۴۱**	۹	۴	۰	۱۹	**۳۲**
گوینده ۸	۰	۰	۵	۳	**۸**	۱	۱	۰	۱	**۳**
گوینده ۹	۱۱	۴۲	۲	۱۴	**۶۹**	۱۱	۵	۰	۵	**۲۱**
گوینده ۱۰	۷	۷	۳	۱۲	**۲۹**	۶	۵	۱	۵	**۱۷**
گوینده ۱۱	۵۰	۴	۵	۸	**۶۷**	۴	۰	۲	۲۳	**۲۹**
گوینده ۱۲	۸	۷	۴	۶	**۲۵**	۴	۵	۰	۸	**۱۷**
گوینده ۱۳	۲	۳	۱	۱۰	**۱۶**	۶	۴	۰	۴	**۱۴**
گوینده ۱۴	۲۳	۷	۵۳	۳۳	**۱۱۶**	۲۳	۱۱	۰	۲۰	**۵۴**
گوینده ۱۵	۴۰	۱۱	۱	۵	**۵۷**	۴	۳	۰	۱۰	**۱۷**
گوینده ۱۶	۲۲	۱۳	۹	۹	**۵۳**	۵	۱	۰	۱۲	**۱۸**
گوینده ۱۷	۵	۸	۳	۲	**۱۸**	۰	۶	۰	۱	**۷**
گوینده ۱۸	۹	۲۷	۱۴	۱۲	**۶۲**	۷	۱	۶	۱۱	**۲۵**
گوینده ۱۹	۱۱	۷	۸	۱۶	**۴۲**	۹	۳	۰	۸	**۲۰**
گوینده ۲۰	۱	۸	۱	۳	**۱۳**	۱	۰	۰	۰	**۱**
گوینده ۲۱	۱۵	۱۷	۱	۵	**۳۸**	۵	۰	۰	۱۶	**۲۱**
گوینده ۲۲	۰	۵	۰	۰	**۵**	۲	۲	۰	۲	**۶**

نرخ بازشناسی با استفاده از فرمول زیر بدست آمده است.

$$\text{SR rate} = Mean\left[\frac{\left(N_{Total} - N_{lost}\right)}{N_{Total}}\right]_{N_{Iteration}} \qquad (۱-۶)$$

که در آن $N_{Itration}$ تعداد تکرار دوره آزمایش می‌باشد که در اینجا سه می‌باشد، N_{Total} تعداد کل جملات تست که ۴۴ می‌باشد و هم چنین N_{Lost} تعداد جملاتی که نادرست تشخیص داده شده‌اند را نشان می‌دهد. نرخ بازشناسی برای هر دو روش پیشنهادی که با الگوریتم‌های مختلف در قسمت خوشه‌بندی استفاده شده است با استفاده از فرمول بالا بدست آمده است. جدول‌های (۶-۴) و (۶-۵) نرخ بازشناسی روش پیشنهادی اول را به ازای ۴۰۰ و ۶۰۰ قاب MFCC در مرحله آموزش و همچنین به ازای ضرایب مشارکت مختلف که نشان دهنده استفاده از تعداد قاب‌ها نزدیک به مراکز خوشه‌ها به اندازه ضریبی از کل قاب‌های موجود در هر خوشه را نشان می‌دهد

جدول (۶-۳) نرخ بازشناسی ۴۰۰ قاب از هر جمله آموزش برای روش پیشنهادی اول

مشارکت قاب‌های جمله تست هر خوشه	۳ خوشه			۴ خوشه			۵ خوشه		
	ICA	K-Means	PSO	ICA	K-Means	PSO	ICA	K-Means	PSO
۳۰ درصد	۶۵,۹	۶۸,۱۸	۶۵,۹	۶۵,۹	۷۹,۵۴	۷۹,۵۴	۷۷,۲۷	۸۱,۸۱	۸۶,۳۶
۴۰ درصد	۶۵,۹	۷۰,۴۵	۶۸,۱۸	۶۵,۹	۷۹,۵۴	۷۷,۲۷	۷۹,۵۴	۸۶,۳۶	۸۴,۰۹
۵۰ درصد	۷۲,۷۲	۷۲,۷۲	۶۸,۲	۶۸,۲	۷۰,۴۵	۷۹,۵۴	۷۹,۵۴	۸۴,۰۹	۸۶,۳۶
۶۰ درصد	۷۲,۷۲	۷۵	۶۸,۲	۷۰,۴۵	۶۸,۱۸	۷۷,۲۷	۸۴,۸۴	۸۶,۳۶	۸۵,۶
۷۰ درصد	۷۵	۷۲,۷۲	۷۶,۵۱	۷۰,۴۵	۶۸,۱۸	۷۹,۵۴	۸۴,۸۴	۸۴,۸۴	۸۶,۳۶
۸۰ درصد	۷۰,۴۵	۷۲,۷۲	۷۰,۴۵	۷۰,۴۵	۶۸,۱۸	۷۵	۸۱,۰۶	۸۱,۸۱	۸۶,۳۶
۹۰ درصد	۷۲,۹	۶۳,۶۳	۷۰,۴۵	۷۰,۴۵	۶۳,۶۳	۷۲,۷۲	۷۷	۷۷,۲۷	۸۸,۶۳

جدول (۶-۴) نرخ بازشناسی ۶۰۰ قاب از هر جمله آموزش برای روش پیشنهادی اول

مشارکت قاب‌های جمله تست هر خوشه	۳ خوشه			۴ خوشه			۵ خوشه		
	ICA	**K-Means**	**PSO**	**ICA**	K-Means	PSO	**ICA**	**K-Means**	**PSO**
۳۰ درصد	۷۰,۴۵	۷۷,۲۷	۷۷,۲۷	۷۲,۲۷	۸۴,۰۹	۷۷,۲۷	۸۸,۶۳	۸۴,۰۹	۸۴,۰۹
۴۰ درصد	۷۷,۲۷	۷۷,۲۷	۷۹,۵۴	۷۷,۲۷	۸۶,۳۶	۸۱,۸۱	۸۶,۳۶	۸۴,۰۹	۸۴,۰۹
۵۰ درصد	۸۱,۸۱	۸۱,۸۱	۷۹,۵۴	۸۷,۸۷	۸۸,۶۳	۸۷,۸۷	۸۹,۳۹	۸۷,۸۷	۸۹,۳۹
۶۰ درصد	۸۴,۰۹	۸۳,۳۳	۸۱,۸۱	۸۷,۸۷	۸۸,۶۳	۸۷,۸۷	۹۰,۱۵	۹۰,۱۵	۸۹,۳۹
۷۰ درصد	۸۴,۰۹	۷۷,۲۷	۸۱,۸۱	۸۷,۸۷	۸۱,۸۱	۸۴,۹	۹۰,۱۵	۸۹,۳۹	۸۹,۳۹
۸۰ درصد	۸۴,۹	۷۷,۲۷	۷۹,۵۴	۷۲,۷۲	۸۱,۸۱	۷۵,۷۵	۸۵,۶	۸۹,۳۹	۸۴,۸۴
۹۰ درصد	۸۶,۹	۷۷,۲۷	۷۹,۵۴	۷۵	۸۱,۸۱	۷۵,۷۵	۸۴,۰۹	۸۷,۸۷	۸۴,۸۴

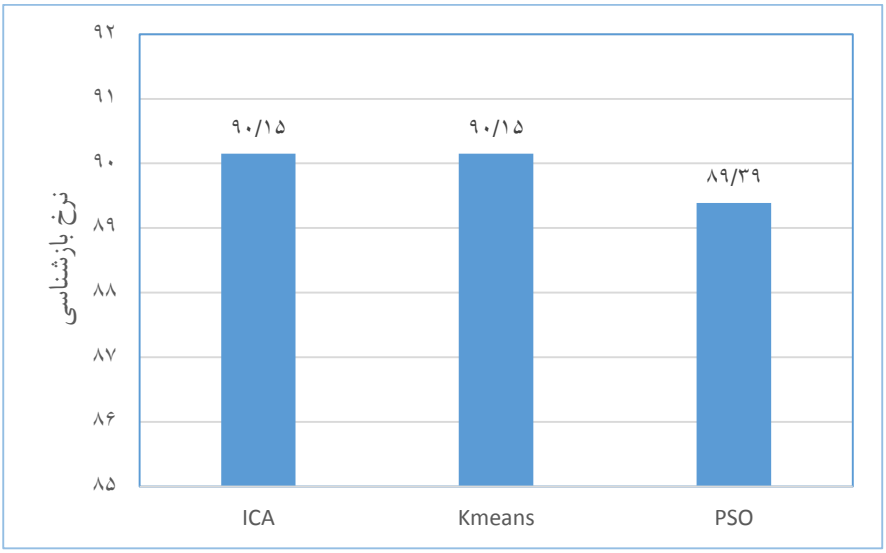
درکل هدف از ارائه جدول (۶-۳) و (۶-۴)، مشخص کردن پارامترهای تأثیر گذار در نرخ بازشناسی در روش‌های پیشنهادی می‌باشد. با بررسی دقیق‌تر جداول ارائه شده در بالا می‌توان به تأثیر مواردی نظیر تعداد خوشه‌ها، تعداد قاب‌ها در مرحله آموزش، انتخاب چه تعداد قاب نزدیک به هر خوشه در جمله تست و همچنین الگوریتم‌های استفاده شده برای خوشه‌بندی قاب‌ها پی برد.

با بررسی جداول حاضر می‌توان به این نتیجه رسید که تعداد خوشه مناسب به ازای هر جمله تعداد ۵ می‌باشد؛ زیرا وقتی هر دو جدول بالا را مشاهده می‌کنیم نرخ بازشناسی در این تعداد خوشه بیش از بقیه موارد می‌باشد. همان‌طور که در بالا ذکر شده است، تعداد مشارکت قاب‌های نزدیک به خوشه می‌توان تأثیرگذار باشد و با مشاهده این جداول می‌توان ضرایب بین ۰,۵ تا ۰,۷ را به عنوان ضرایب مناسب معرفی کرد که البته در این بین ما ضریب ۰,۶ را به عنوان مناسب‌ترین ضریب انتخاب کرده‌ایم.

تعداد قاب‌های آموزش یکی از مهم‌ترین پارامترها در نرخ بازشناسی می‌باشد که همان گونه که از مقایسه بین جدول (۶-۳) و جدول (۶-۴) مشخص می‌باشد تعداد ۴۰۰ قاب به ازای هر جمله در قسمت آموزش کافی نمی‌باشد و نیاز به قاب‌های بیشتری برای آموزش هر فرد نیازمندیم؛ از این رو ما در قسمت آموزش از تعداد ۶۰۰ قاب به ازای هر جمله در قسمت آموزش استفاده کرده‌ایم. البته تعداد قاب‌های بیش از ۶۰۰ هم مورد آزمایش قرار گرفته است که باعث بهبود نرخ بازشناسی نشده است.

از انواع الگوریتم‌هایی که برای خوشه‌بندی استفاده شده است، می‌توان به عنوان آخرین پارامتر که در این جداول به آن اشاره شده است نام برد. درکل، همان‌طور که از مقایسه دو جدول فوق و پارامترهای مختلف موجود در آن‌ها مشاهده می‌شود نرخ بازشناسی ۹۰,۱۵ که به ازای ضریب مشارکت ۰,۶ و خوشه‌بندی با تعداد ۵ خوشه که با تعداد ۶۰۰ قاب به ازای هر جمله آموزش داده شده است را می‌توان به عنوان پارامترهای برتر انتخاب کرد. همان طور که در نمودار (۶-۱) مشخص می‌باشد نرخ بازشناسی در خوشه‌بندی با استفاده ازICA و K-Means باهم برابر می‌باشد و نرخ بازشناسی با استفاده از PSO ۸۹,۳۹ می‌باشد. همان‌طور که مشخص می‌باشد الگوریتم ICA در روش پیشنهادی اول برتری محسوسی نسبت به دیگر الگوریتم‌هایی که قسمت خوشه‌بندی استفاده شده است ندارد.

این نتیجه از آنجا حاصل می‌شود که در قسمت خوشه‌بندی به علت وجود داده‌های زیاد و مناسب هیچ تغییر محسوسی در انتخاب بردارهای ویژگی پرتکرار ایجاد نمی‌شود و به نوعی هر سه الگوریتم در امر خوشه‌بندی تقریبا یکسان عمل می‌کنند.



نمودار (۶-۱) نرخ بازشناسی روش پیشنهادی اول به ازای الگوریتم‌های مختلف و با ضریب مشارکت ۰,۶ جمله تست، ۵ خوشه و ۶۰۰ قاب برای آموزش

جدول (۶-۶) مدت زمان استفاده شده به ازای هر فرد در قسمت آموزش را نشان می‌دهد که با استفاده از فرمول زیر بدست آمده است.

(۶-۱) ضریب جابه‌جایی × تعداد قاب‌های استفاده شده = مدت زمان استفاده شده به ازای هر فرد

(۶-۲) تعداد قاب برای هر جمله × تعداد جمله = تعداد قاب‌های استفاده شده در آموزش

با استفاده از فرمول بالا مدت زمان استفاده شده به ازای هر فرد در قسمت آموزش در تعداد قاب‌های متفاوت بدست می‌آید. جدول (۶-۵) مدت زمان استفاده شده به ازای هر فرد در قسمت آموزش را نشان می‌دهد و همچنان که مشخص می‌باشد این مدت زمان در دو روش پیشنهادی برابر می‌باشد.

جدول (۶-۵) مدت زمان استفاده شده به ازای هر فرد برای قسمت آموزش

	۴۰۰ قاب استفاده شده در آموزش	**۶۰۰ قاب استفاده شده در آموزش**
روش پیشنهادی اول	۱۵ ثانیه	۲۲,۵ ثانیه
روش پیشنهادی دوم	۱۵ ثانیه	۲۲,۵ ثانیه

۶-۳-۲- نتایج روش پیشنهادی دوم

همانند نتایج روش پیشنهادی اول در روش دوم هم دو جدول نرخ بازشناسی را ارائه کرده‌ایم. در جدول (۶-۶) نرخ بازشناسی به ازای ۴۰۰ قاب برای هر جمله در آموزش را نشان می‌دهد و جدول (۶-۷) نرخ بازشناسی را به ازای ۶۰۰ قاب برای هر جمله در آموزش را نشان می‌دهد.

جدول (۶-۶) نرخ بازشناسی به ازای ۴۰۰ قاب از هر جمله آموزش برای روش پیشنهادی دوم

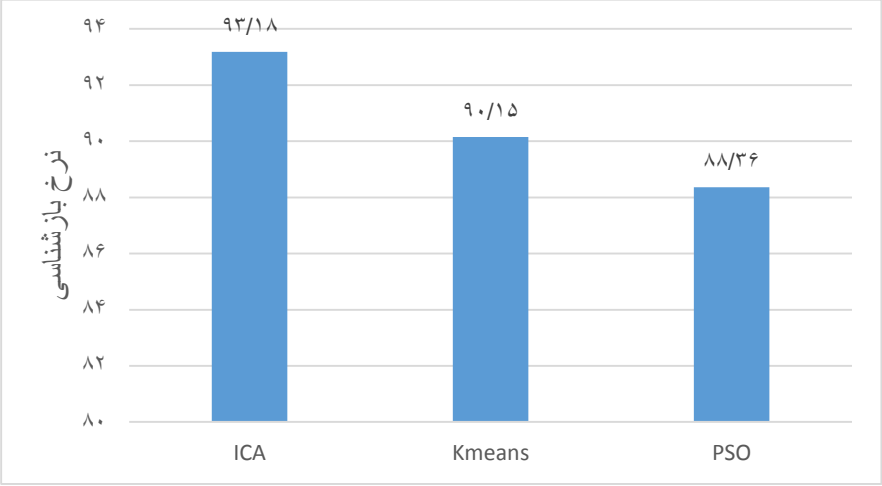
مشارکت قاب‌های هر خوشه از جمله تست	۳ خوشه			۴ خوشه			۵ خوشه		
	ICA	K-Means	PSO	ICA	K-Means	PSO	ICA	K-Means	PSO
۳۰ درصد	۶۵,۹	۶۳,۶۳	۶۱,۳۶	۷۰,۴۵	۶۵,۹	۷۲,۷۲	۷۵	۷۷,۲۷	۷۵
۴۰ درصد	۶۸,۱۸	۵۹,۰۹	۵۹,۰۹	۶۸,۱۸	۶۵,۹	۷۵	۷۲,۷۲	۷۷,۲۷	۷۲,۷۲
۵۰ درصد	۶۵,۹	۵۹,۰۹	۵۹,۰۹	۷۵	۶۸,۱۸	۷۵	۷۰,۴۵	۷۷,۲۷	۷۵
۶۰ درصد	۶۸,۱۸	۵۴,۵۴	۵۹,۰۹	۷۵	۷۰,۴۵	۷۹,۵۴	۷۰,۴۵	۷۹,۵۴	۷۷,۲۷
۷۰ درصد	۶۳,۶۳	۵۰	۶۳,۶۳	۷۲,۷۲	۶۸,۱۸	۷۹,۵۴	۶۵,۹	۸۴,۰۹	۷۷,۲۷
۸۰ درصد	۵۴,۵۴	۴۳,۱۸	۶۳,۶۳	۶۳,۶۳	۶۸,۱۸	۷۹,۵۴	۶۳,۶۳	۸۱,۸۱	۷۷,۲۷
۹۰ درصد	۵۲,۲۷	۳۴,۰۹	۶۵,۹	۶۱,۳۶	۶۵,۹	۷۲,۷۲	۶۵,۹	۷۹,۵۴	۷۵,۷۵

جدول (۶-۷) نرخ بازشناسی به ازای ۶۰۰ فریم از هر جمله آموزش برای روش پیشنهادی دوم

مشارکت قاب‌های هر خوشه از جمله تست	۳ خوشه			۴ خوشه			۵ خوشه		
	ICA	K-Means	PSO	ICA	K-Means	PSO	ICA	K-Means	PSO
۳۰ درصد	۶۵,۹	۷۹,۵۴	۷۹,۵۴	۷۹,۵۴	۸۶,۳۶	۷۵	۸۴,۰۹	۸۸,۶۳	۷۰,۴۵
۴۰ درصد	۷۲,۷۲	۷۷,۲۷	۸۱,۸۱	۸۶,۳۶	۸۸,۶۳	۸۱,۸۱	۸۸,۶۳	۸۸,۶۳	۷۲,۷۲
۵۰ درصد	۷۶,۵۱	۸۱,۸۱	۷۹,۵۴	۸۶,۳۶	۸۶,۳۶	۸۶,۳۶	.۹۰,۱۵	۸۸,۶۳	۸۳,۳۳
۶۰ درصد	۷۸,۷۸	۷۹,۵۴	۷۷,۲۷	۸۸,۶۳	۸۶,۳۶	۸۶,۳۶	۹۲,۴۲	۸۸,۶۳	۸۸,۳۶
۷۰ درصد	۷۸,۷۸	۷۹,۵۴	۷۹,۵۴	۸۸,۶۳	۸۹,۳۹	۸۶,۳۶	۹۲,۴۲	۹۰,۱۵	۸۸,۳۶
۸۰ درصد	۷۸,۷۸	۷۹,۵۴	۸۱,۸۱	۸۸,۶۳	۹۰,۹	۸۶,۳۶	۹۳,۱۸	۹۰,۱۵	۸۸,۳۶
۹۰ درصد	۷۶,۵۱	۷۹,۵۴	۸۱,۸۱	۸۴,۰۹	۸۶,۳۶	۸۶,۳۶	۸۴,۰۹	۸۸,۶۳	۸۷,۱۲

در اینجا هم همانند آنچه که در قسمت قبل بیان شده است؛ نرخ بازشناسی به ازای خوشه‌های مختلف، ضریب مشارکت نزدیک‌ترین قاب‌ها به مراکز خوشه‌ها در جمله تست، تعداد قاب‌های استفاده شده در قسمت آموزش به ازای هر جمله و برای الگوریتم‌های مختلف ارائه شده است.

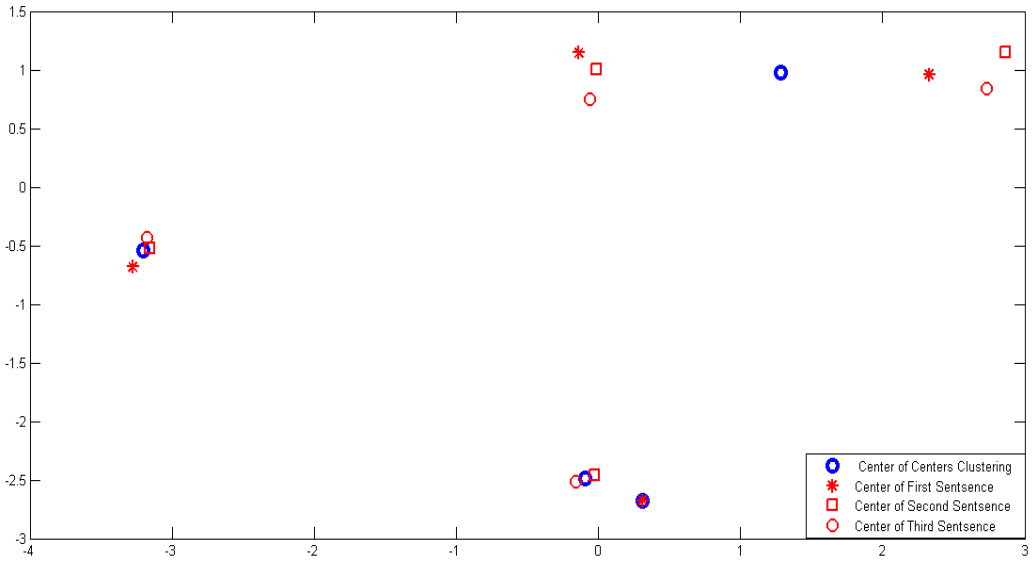
در جداول(۶-۶) و (۶-۷) نرخ بازشناسی به ازای مقادیر مختلف ارائه شده است. با بررسی این جداول به این نتیجه خواهیم رسید که در این روش هم همانند روش قبل ۴۰۰ قاب برای آموزش کافی نمی‌باشد و باید از قاب‌های بیشتری به ازای هر جمله استفاده کرد و همچنین خوشه‌بندی با ۵ خوشه نتایج بهتری را رقم زده و باعث بهبود نرخ باز شناسی می‌شود. در روش دوم ضریب مشارکت قاب‌های جمله تست را با توجه به نتایج بدست آمده از جدول بالا ۰,۸ انتخاب می‌کنیم. همانند چیزی که با مشخصات بالا در نمودار (۶-۲) آمده است. نرخ بازشناسی در این روش که با استفاده از الگوریتم رقابت استعماری انجام شده است ۹۳,۱۸ می‌باشد که ۳,۰۳ در صد از K-Means و ۴,۸۲ درصد PSO بیشتر می‌باشد.



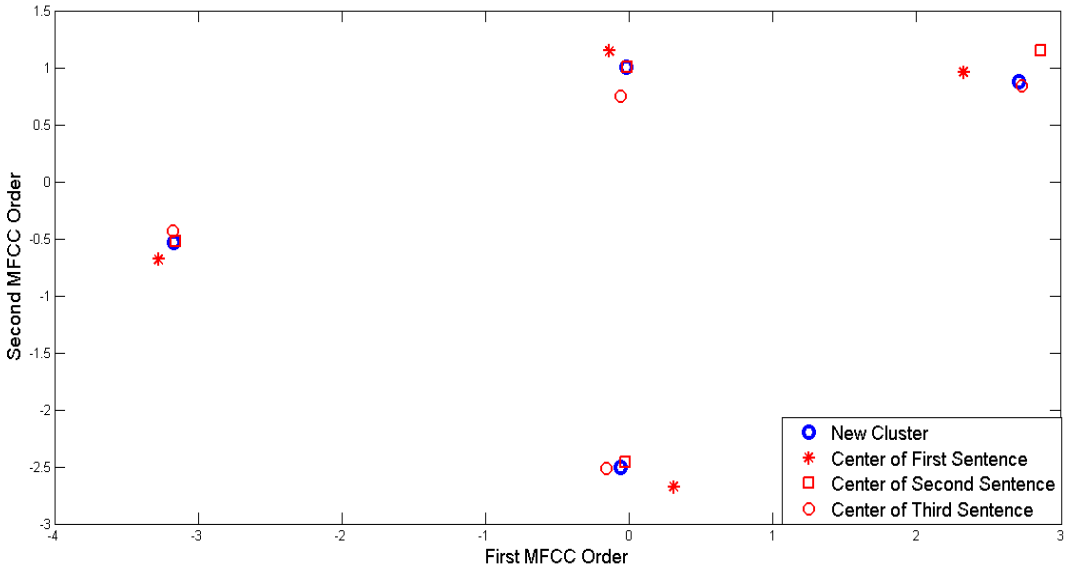
نمودار (۶-۲) نرخ بازشناسی به ازای الگوریتم‌های مختلف و با ضریب مشارکت ۰,۸ جمله تست، ۵ خوشه و ۶۰۰ قاب برای آموزش روش پیشنهادی دوم

در روش پیشنهادی اول نتایج سه الگوریتم ICA، k-Means و PSO تقریبا برابر بوده و برتری محسوسی نسبت به هم نداشته‌اند، اما در روش دوم بنا به آنچه که در زیر به آن اشاره خواهیم کرد

الگوریتم ICA نسبت به دو الگوریتم ذکر شده برتری یافته است. طبق آنچه که در فصل قبل و در قسمت روش دوم پیشنهادی ذکر شده است ما بعد از بدست آوردن مراکز هر جمله برای آنچه که آن را همگرا کردن مدل‌سازی و یا وزن شبکه نامیده می‌شود، مراکز خوشه‌های تمام جملات یک فرد را که در آموزش استفاده شده است را دوباره خوشه‌بندی کرده‌ایم . در نهایت، N تا قاب را که به مراکز جدید نزدیک‌تر هستند انتخاب می‌کنیم. به همین خاطر نیاز به الگوریتم داشته‌ایم که عمل خوشه‌بندی را در داده‌های کم و با داده‌های پرت به درستی انجام دهد. در شکل (۶-۲) خوشه بندی با استفاده از الگوریتم k-Means انجام شده است که این عمل به طور مناسب انجام نشده است، همان‌گونه که مشخص می‌باشد در یک خوشه دو مرکز جدید ایجاد شده است و از طرفی دو خوشه را یک خوشه بزرگ تشخیص داده است. در نتیجه این کار باعث می‌شود که وزن‌های شبکه به درستی همگرا نشوند. اما همان‌طور در شکل (۶-۳) آمده است، شکل (۶-۲) با استفاده از الگوریتم رقابت استعماری خوشه‌بندی شده است که همان‌گونه که مشخص می‌باشد این شکل به صورت کاملا مناسب خوشه‌بندی شده است که به انتخاب قاب‌های مناسب برای همگرایی بیشتر کمک می‌کند.



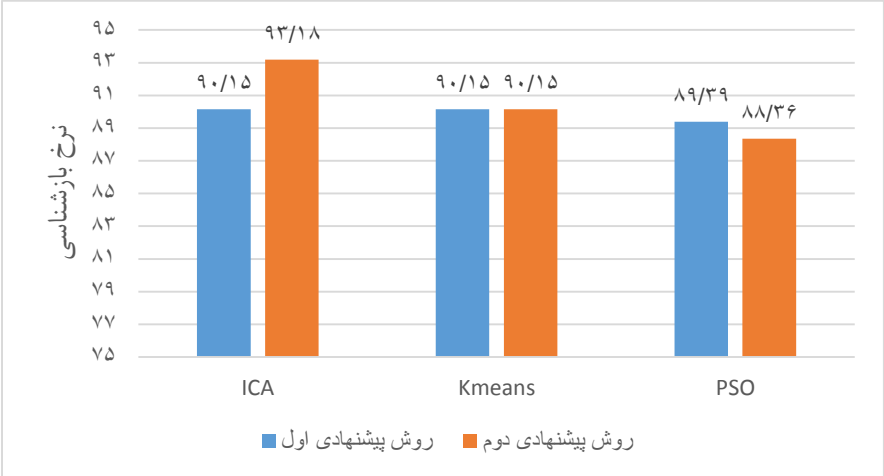
شکل (۶-۲) خوشه‌بندی مراکز سه جمله آموزش برای گوینده نهم توسط الگوریتم k-Means



شکل (۶-۳) خوشه‌بندی مراکز سه جمله آموزش برای گوینده نهم توسط الگوریتم ICA

۶-۳-۳- مقایسه عملکرد دو روش پیشنهادی با هم

در این قسمت دو روش پیشنهادی را از دو جنبه نرخ بازشناسی و مدت زمان استفاده شده به ازای هر فرد در قسمت آموزش را با یک دیگر مقایسه کنیم.



نمودار (۶-۳) نرخ بازشناسی برای هر دو روش پیشنهادی با الگوریتم‌های مختلف خوشه‌بندی

طبق نمودار (۶-۳) روش پیشنهادی دوم با ICA عملکرد بهتری نسبت به عملکرد روش پیشنهادی اول با استفاده از ICA داشته است که به‌طور متوسط باعث افزایش ۲,۵۱ نرخ بازشناسی شده

است و در این بین، روش پیشنهادی دوم با استفاده از ICA با نرخ بازشناسی ۹۳,۱۸بهترین عملکرد را بین سایر الگوریتم‌های خوشه‌بندی داشته است.

یکی دیگر از جوانبی که مورد بررسی قرار گرفته است، و مدت زمان استفاده شده به ازای هر فرد در قسمت آموزش در دو روش پیشنهادی می‌باشد. همان‌طور که از جدول (۶-۶) می‌تواند نتیجه گرفت، و مدت زمان استفاده شده به ازای هر فرد در قسمت آموزش در هر دو روش یکسان می‌باشد. فرمول بدست آوردن مدت زمان مصرفی از داده‌گان در قسمت آموزش برای هر دو روش یکسان بوده و این نرخ به طور متوسط برابر است ۲۲,۵ ثانیه به ازای هر فرد می‌باشد. بنابراین دو روش پیشنهادی از این نظر هیچ برتری نسبت به هم ندارند. اگر چه روش دوم به خاطر خوشه‌بندی مجدد از پیچیدگی بیشتری نسبت به روش اول برخوردار می‌باشد.

۶-۴- نتایج سیستم بدون استفاده از بلوک انتخاب ویژگی

نتایج این قسمت بر اساس تعداد قاب‌های MFCC که به شبکه داده شده است و بدون بکار بردن هیچ انتخاب ویژگی ثبت شده است. طبق آنچه که در جدول (۱-۱۰) آورده شده است تعداد قاب‌های آموزش در نرخ بازشناسی تا یک مقدار خاصی تأثیر دارد و بعد از آن، با اضافه کردن به قاب‌های آموزش تأثیر چندانی در نرخ بازشناسی مشاهده نمی‌شود. نا گفته نماند در این روش در قسمت تست تمام قاب‌های جمله آموزش مورد آزمایش قرار می‌گیرد

جدول (۶-۸) نرخ بازشناسی به ازای قاب‌های مختلف و بدون بلوک انتخاب ویژگی

	۴۰۰ قاب	۶۰۰ قاب	۸۰۰ قاب	۹۰۰ قاب
نرخ بازشناسی	۳۶,۳۶	۳۴,۱	۴۵,۴۵	۴۵,۴۵

۶-۵- مقایسه عملکرد روش‌های پیشنهادی با دیگر روش‌ها

برای مقایسه عملکرد دو روش پیشنهادی از مقاله [32] که در سال ۲۰۱۳ در مجله Springer به چاپ رسیده است استفاده شده است. در این مقاله از داده‌گان ELSDSR استفاده شده است و همان طور که در جدول (۶-۱۰) مشاهده می‌کنید درصد بازشناسی و مدت زمان آموزش برای ۱۰ کلاس (۱۰ نفر) در دو روش SVM و ELM[۱] بهینه به نمایش درآمده است. به همین خاطر نتایج مقاله مورد نظر را برای بررسی و مقایسه روش‌های پیشنهادی مناسب دیده شده است. با استفاده از میانگین‌گیری نرخ بازشناسی تمام کلاس‌ها نرخ بازشناسی برای روش SVM ۹۱,۴ و برای روش ELM بهینه ۹۱,۵ بدست آمده است. هم اکنون از دو جنبه نرخ بازشناسی و مدت زمان استفاده شده از داده‌گان برای آموزش به ازای هر فرد، روش‌های پیشنهادی و دیگر روش‌ها را مورد بررسی و مقایسه قرار خواهیم داد.

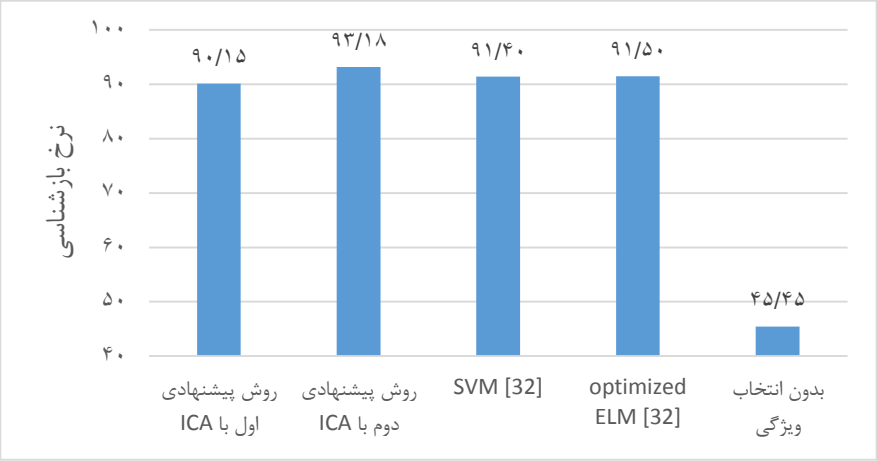
جدول (۶-۹)نتایج بدست آمده در شناسایی گوینده با روش SVM و optimaized ELM [32].

Class	Optimized ELM		SVM	
	Training time (s)	Testing rate (%)	Training time (s)	Testing rate (%)
2	**9.57**	85.54	9.94	**88.67**
4	**9.66**	**96.03**	10.28	94.81
6	12.18	82.93	**10.72**	**84.04**
7	**8.48**	83.92	9.20	**87.63**
8	**10.04**	**89.14**	10.94	87.98
9	11.41	**92.61**	11.43	91.59
10	**9.69**	**94.75**	9.90	92.63
13	9.32	98.82	9.54	98.10
18	9.38	**95.89**	9.39	93.31
20	8.36	**96.18**	9.00	95.68

نمودار (۶-۵) نرخ بازشناسی که دو روش پیشنهادی با استفاده از الگوریتم رقابت استعماری بدست آمده است را با روش دیگر مقایسه می‌کند. همان‌طور که در این نمودار نمایان می‌باشد روش

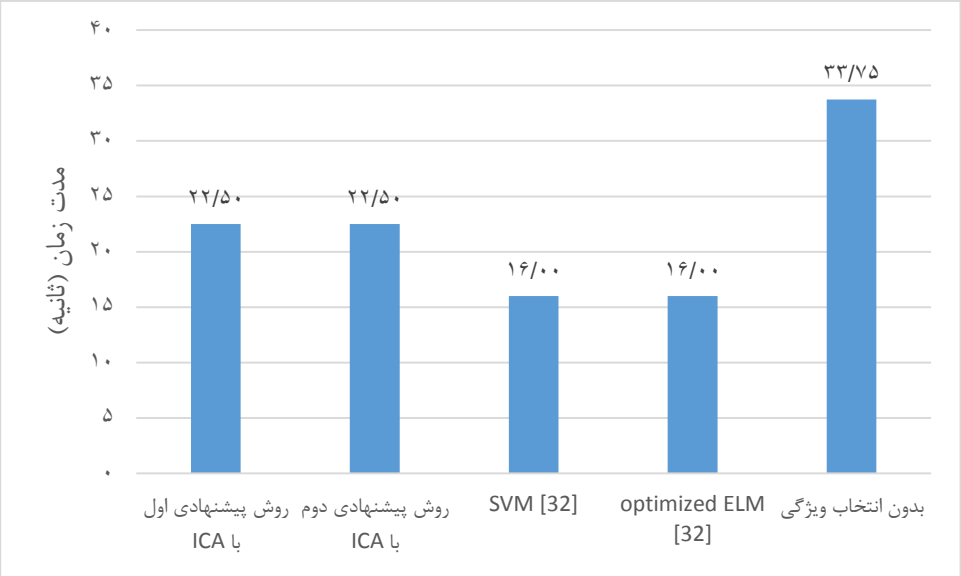
[۱] Extreme Learning Machine

پیشنهادی اول از نرخ بازشناسی تقریبا کمتری نسبت به روش SVM و ELM بهینه برخوردار می‌باشد. اما روش پیشنهادی دوم با نرخ بازشناسایی ۹۳,۱۸ درصد از نرخ بالاتری نسبت به دیگر روش‌ها برخوردار می‌باشد که نرخ بازشناسی را به اندازه ۱,۶۸ درصد نرخ بازشناسی را بهبود بخشیده است.



نمودار (۶-۴) مقایسه نرخ بازشناسی روش‌های پیشنهادی با استفاده از ICA با روش‌های دیگر

مدت زمان استفاده شده از داده‌گان به ازای هر فرد در قسمت آموزش هم به عنوان پارامتر بعدی مورد بررسی قرار خواهیم داد. در مقاله [32] آمده است که به ازای هر فرد ۲۰۰۰ قاب MFCC استفاده شده است و طول هر قاب ۱۶ میلی ثانیه و جابه‌جایی آن ۸ میلی ثانیه می‌باشد. بنابراین طبق فرمول‌های (۶-۱) و (۶-۲) به ازای هر فرد ۱۶ ثانیه از داده‌گان حاضر در قسمت آموزش استفاده کرده است. نمودار (۶-۵) مدت زمان استفاده شده به ازای هر فرد در قسمت آموزش را در روش‌های مختلف را نشان می‌دهد. همان‌گونه که در این نمودار مشخص می‌باشد مدت زمان مصرفی به ازای هر فرد بیشتر از دو روش SVM و ELM بهینه می‌باشد. در واقع این زمان در روش‌های پیشنهادی ۱,۴برابر روش SVM و ELM بهینه می‌باشد.



نمودار (۶-۵) مدت زمان استفاده شده از داده‌گان به ازای هر فرد در قسمت آموزش

۶-۶- جمع‌بندی

طبق جداول (۶-۴) و (۶-۷) نرخ بازشناسی روش پیشنهادی اول با الگوریتم رقابت استعماری ۹۰,۱۵٪ و در روش دوم پیشنهادی با الگوریتم رقابت استعماری ۹۳,۱۸٪ می‌باشد. همان‌طور که قبلا هم ذکر شده است الگوریتم روش پیشنهادی اول نسبت به سایر الگوریتم‌های خوشه‌بندی برتری خاصی ندارد، اما در روش دوم الگوریتم رقابت استعماری بدلیل آنچه که ذکر شده است نسبت به دو الگوریتم دیگر عملکر بهتری داشته است.

با بررسی نمودارها و جداول ارائه شده در این فصل و مقایسه آن با یکدیگر می‌توان به این نتیجه رسید که با استفاده از ۱,۴داده بیشتردر روش پیشنهادی دوم که با استفاده از الگوریتم رقابت استعماری انجام پذیرفته است نسبت به روش‌های SVM و optimaized ELM می‌توان نرخ بازشناسی را به اندازه ۱,۶۸٪ بهبود داد. همان‌طور که در قسمت نتایج هم ذکر شده است روش پیشنهادی دوم به خوشه‌بندی مراکز خوشه‌ها حساس می‌باشد، از این رو به الگوریتم خوشه‌بندی نیازمندیم که بتواند به طور مناسب

خوشه‌بندی مراکز را انجام دهد. از طرفی مزیت روش پیشنهادی اول این است که حساسیت آن به خوشه‌بندی کمتر می‌باشد و از پیچیدگی محاسباتی کمتری برخوردار می‌باشد.

۶-۷- پیشنهادات

از آنجایی که ما روش‌های پیشنهادی را بر روی بردارهای ویژگی پر تکرار مجرای صوتی هر فرد در نظر گرفته‌ایم، می‌توان دیگر ویژگی‌های که نشان دهنده‌ی بردارهای ویژگی مجرای صوتی و شبیه‌ساز ماهیچه‌های فیزیکی صوتی می‌باشند را به نوعی مورد آزمایش قرار دهند. LPCC، PLP و غیره می‌تواند همانند ویژگی استفاده شده در این پایان نامه باشد.

از آنجایی که این دو روش پیشنهادی به طور محسوسی به تعداد خوشه و نوع خوشه‌بندی وابسته می‌باشد، ارائه روشی برای انتخاب تعداد خوشه‌های مناسب و نوع مناسبی از خوشه‌بندی می‌تواند بهبود نرخ بازشناسی کمک کند.

با بررسی این پروژه می‌توان به این قضیه پی برد که پارامترهای زیادی در قسمت‌های مختلف این نوع از سیستم‌های شناسایی گوینده دخیل می‌باشد که با انتخاب درست آن می‌توان باعث بهبود نرخ بازشناسی شوند.

۶-۸- منابع و مراجع

[1] S. Pruzansky, (1963), "**Pattern-Matching Procedure for Automatic Talker Recognition**", J. Acoustic. Soc. Am., vol. 35, no. 3, pp. 354–358.

[2] M. F. BenZeghiba and H. Bourlard, (2003), "**User-customized password speaker verification using multiple reference and background models**", Speech Communication., vol. 48, no. 9, pp. 1200–1213.

[3] S. Molau, M. Pitz, R. Schluter, and H. Ney, (2001), "**Computing mel-frequency cepstral coefficients on the power spectrum**", in Acoustics, Speech and Signal Processin (ICASSP'01). 2001 IEEE International Conference on, vol. 1, pp. 73–76.

[4] Q. Zhu, B. Chen, N. Morgan, and A. Stolcke, (2005), "**Tandem connectionist feature extraction for conversational speech recognition**", in Machine Learning for Multimodal Interaction, Springer, pp. 223–231.

[5] L. R. Rabiner and B.-H. Juang, (1993), "**Fundamentals of speech recognition**", vol. 14. PTR Prentice Hall Englewood Cliffs.

[6] A. Waibel, (1990), "**Readings in speech recognition**", Morgan Kaufmann.

[7] J. R. Deller, J. G. Proakis, and J. H. Hansen, (2000), "**Discrete-time processing of speech signals**", IEEE New York, USA.

[8] S. Nakagawa, L. Wang, and S. Ohtsuka, (2012), "**Speaker identification and verification by combining MFCC and phase information**", Audio Speech Languastic Processing, IEEE, vol. 20, no. 4, pp. 1085–1095.

[9] D. A. Reynolds, (1995), "**Speaker identification and verification using Gaussian mixture speaker models**", Speech Communication, vol. 17, no. 1, pp. 91–108.

[10] E. Atashpaz-Gargari and C. Lucas, (2007), "**Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition**", in Evolutionary Computation, 2007. CEC 2007. IEEE Congress on, pp. 4661–4667.

[11] H. Hermansky, (1990), "**Perceptual linear predictive (PLP) analysis of speech**", J. Acoust. Soc. Am., vol. 87, no. 4, pp. 1738–1752.

[12] J. Makhoul, (1975), "**Linear prediction: A tutorial review**", Proc. IEEE, vol. 63, no. 4, pp. 561–580.

[13] F. zohra Chelali, A. Djeradi, and R. Djeradi, (2011), "**Speaker identification system based on PLP coefficients and artificial neural network**", environments, vol. 1, p. 2.

[14] M. Rosell, (2006), "**An introduction to front-end processing and acoustic features for automatic speech recognition**", Jan, vol. 17, pp. 1–10.

[15] M. A. Zissman and others, (1996), "**Comparison of four approaches to automatic language identification of telephone speech**", IEEE Trans. Speech Audio Processing, vol. 4, no. 1, p. 31.

[16] W. S. McCulloch and W. Pitts, (1943), "**A logical calculus of the ideas immanent in nervous activity**", Bull. Math. Biophysic, vol. 5, no. 4, pp. 115–133.

[17] R. Lippmann, (1989), "**Review of Neural Networks for Speech Recognition**", Neural Computer, vol. 1, no. 1, pp. 1–38, Mar.

[18] K.-F. Lee and H.-W. Hon, (1989), "**Speaker-independent phone recognition using hidden Markov models**", Acoustic Speech Signal Processing IEEE Transaction On, vol. 37, no. 11, pp. 1641–1648.

[19] M. Gales and S. Young, (2008), "**The application of hidden Markov models in speech recognition**", Found. Trends Signal Processing, vol. 1, no. 3, pp. 195–304.

[20] X. D. Huang, Y. Ariki, and M. A. Jack, (1990), "**Hidden Markov models for speech recognition**", vol. 2004. Edinburgh University press Edinburgh.

[21] R. O. Duda, P. E. Hart, and D. G. Stork, (1999), "**Pattern classification**", John Wiley & Sons.

[22] P. KaewTraKulPong and R. Bowden, (2002), "**An improved adaptive background mixture model for real-time tracking with shadow detection**", in Video-Based Surveillance Systems, Springer, pp. 135–144.

[23] Z. Zivkovic, (2004), "**Improved adaptive Gaussian mixture model for background subtraction**" in Pattern Recognition ICPR 2004. Proceedings of the 17th International Conference on, 2004, vol. 2, pp. 28–31.

[24] D. A. Reynolds and R. C. Rose, (1995), "**Robust text-independent speaker identification using Gaussian mixture speaker models**", Speech Audio Processing IEEE Transaction On, vol. 3, no. 1, pp. 72–83.

[25] B. Jiao, Z. Lian, and X. Gu, (2008), "**A dynamic inertia weight particle swarm optimization algorithm**", Chaos Solitons Fractals, vol. 37, no. 3, pp. 698–705.

[26] Furui S (1997) "**Recent advances in speaker recognition**", Pattern Recognition Letter 18:859–872.

[27] X. Hu, Y. Shi, and R. C. Eberhart, (2004), "**Recent advances in particle swarm", in IEEE congress on evolutionary computation**", vol. 1, pp. 90–97.

[28] A. P. Engelbrecht, (2006), "**Fundamentals of computational swarm intelligence**", John Wiley & Sons.

[29] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, (2012), "**A density-based algorithm for discovering clusters in large spatial databases with noise**", in Kdd, 1996, vol. 96, pp. 226–231.

[۳۰] خ. مهرداد و ز. علیرضا, پایان نامه ارشد، (۱۳۹۲) ، "**تشخیص الگو در جریان داده به روش یادگیری غیرنظارتی**"، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف.

[31] L. Feng and L. K. Hansen, (2005), "**A new database for speaker recognition**".

[32] Y. Lan, Z. Hu, Y. C. Soh, and G.-B. Huang, (2013), "**An extreme learning machine approach for speaker recognition**", Neural Computer Application, vol. 22, no. 3–4, pp. 417–425.

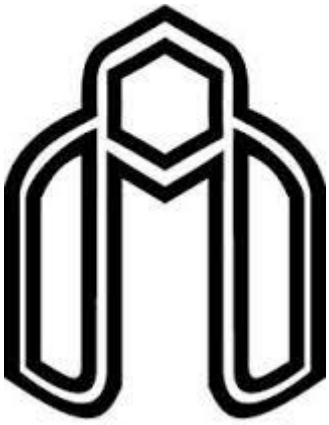
Abstract

The speaker recognition has been one of the interesting issues in signal and speech processing over the last few decades. A lot of efforts have been applied to improve the efficiency of speaker recognition system. Feature extraction is one of the main part of a speaker recognition system which can improve the performance of system. There are some methods in order to choose better feature, but clustering is used in this project to find feature vectors that have the most similarity. Therefore, these similar vectors specify the most properties in vocal tract of each person that they help us to build acoustic model of each person or obtain better decision boundary between different individuals.

We've purposed two methods that ICA is used in both and then use k-mean and PSO algorithm have been used in order to compare ICA's performance. Finally, two purposed methods using ICA have been compare to SVM and ELM optimaized in recognition rate and period of time used database. Experimental results show that recognition accuracy has been improved in perposed methods by using ICA.

In the project, we use English Language Speech Database for Speaker Recognition as a database. Therefore, our system is text independent speaker identification.

Keywords: Identification system, MFCC feature, feature selection, Imperialist competitive algorithm, clustering



Shahrood University of Technology

Faculty of Electrical and Robotic Engineering

Speaker identification based on appropriate features selection using ICA

Mohammad Soleymanpour

Supervisor**:**

Dr. Hossein Marvi

Febraury 2015