دانشگاه قم

دانشکده فنی و مهندسی

پایان‌نامه کارشناسی ارشد رشته مهندسی فناوری اطلاعات

عنوان:

ارائه یک ساختار جدید وابسته به بافت برای بازشناسی گفتار پیوسته

استاد راهنمای اول:

دکتر روح الله دیانت

استاد راهنمای دوم:

دکتر باقر باباعلی

نگارنده:

سید محمد احمدی

زمستان ۱۳۹۳

بسم الله الرحمن الرحيم

دانشگاه قم

دانشکده فنی و مهندسی

پایان‌نامه کارشناسی ارشد رشته مهندسی فناوری اطلاعات

عنوان:

ارائه یک ساختار جدید وابسته به بافت برای بازشناسی گفتار پیوسته

استاد راهنمای اول:

دکتر روح الله دیانت

استاد راهنمای دوم:

دکتر باقر باباعلی

نگارنده:

سید محمد احمدی

زمستان ۱۳۹۳

در راه تو دل خون شد و جانم به لب آمد

چـــیز دگری لایق تقـــدیم نـداریم

تسلیم تو گشـــتیم ســـراپا که نگـــویند

در پیـــش محبان، سر تسلیم نداریم

فرخی یزدی

تقدیر و سپاس

از راهنمایی‌های اساتید محترم راهنما، جناب آقای دکتر روح الله دیانت که تلاش‌های بی‌دریغشان هموارکننده ناهمواری‌های مسیر این پایان‌نامه بوده است؛ و نیز جناب آقای دکتر باقر باباعلی که مرا در این رساله، یاری نمودند، بسیار سپاس‌گزارم.

همچنین از همه کسانی که در نگارش این تحقیق مؤثر بودند، کمال تشکر را دارم.

چکیده

این پژوهش، کوششی است برای ارائه یک ساختار وابسته به بافت برای بازشناسی گفتار پیوسته. مدل‌سازی مستقل از بافت، اثر واج‌های پیرامون را بر تلفظ هر واج، نادیده می‌گیرد و به همین دلیل، برای مدل‌سازی هر واج، همه نمونه‌های آموزشی مربوط به آن واج را در نظر می‌گیرد. در نظر گرفتن اثر بافت در مدل‌سازی، می‌تواند خطای بازشناسی را تا حد زیادی کاهش دهد. این پایان‌نامه با در نظر گرفتن واحد آوایی سه‌واجی، واج‌های پیشین و پسین هر واج را نیز در مدل‌سازی دخالت می‌دهد. به این صورت که سه‌واجی‌های مشابه با هم را از طریق الگوریتم خوشه‌بندی K-Means تشخیص داده، یعنی سه‌واجی‌های مشابه را با هم در یک خوشه قرار می‌دهد. این کار موجب می‌شود که تنها نمونه‌هایی از یک واج، با هم آموزش داده شوند که بافت مشابهی داشته باشند. مهم‌ترین مشکلی که برای خوشه‌بندی وجود دارد، ناهمسانی ابعاد داده‌ها (ویژگی‌های مربوط به هر واج) است. برای همسان‌سازی ابعاد داده‌ها، از سه روش DTW ، تبدیل فوریه و PCA استفاده شد. مشاهده شد که روش‌های تبدیل فوریه و نیز PCA هم از نظر سرعت و هم از نظر دقت، بهتر از DTW عمل می‌کنند. همچنین در مقایسه دو روش تبدیل فوریه و PCA، روش تبدیل فوریه موفق‌تر عمل کرد. پس از خوشه‌بندی، خوشه‌هایی که داده‌های آموزشی کمی داشتند، با یکدیگر ادغام گردیدند و نیز واج‌هایی که داده‌های آموزشی زیادی داشتند، به عنوان سه‌واجی‌های ویژه، در یک خوشه مستقل قرار گرفتند. در نهایت، مدل نهایی وابسته به بافت ارائه شده، که از روش تبدیل فوریه برای همسان‌سازی ابعاد استفاده می‌کند و از روش K-Means، هر واج را ابتدا به سه خوشه تقسیم می‌کند و نیز پس از إعمال پس‌پردازش‌های یادشده، نرخ خطای واجی را از ۳۴,۷ (که مربوط به حالت مستقل از بافت بود) به ۳۰,۵۵ کاهش داد. بستر انجام آزمایش‌ها، محیط Kaldi بوده است. نرخ خطای واجی روش پیش‌فرض Kaldi در مرحله سه‌واجی، ۳۰,۹۸ است که نسبت به روش پیشنهادی، ۰,۴۳ درصد بیشتر است.

واژگان کلیدی: بازشناسی گفتار، گفتار پیوسته، بافت، خوشه‌بندی، سه‌واجی.

فهرست مطالب

فصل اول: مقدمات و کلیات

۱- تبیین موضوع ۲
۲- پرسش‌های تحقیق ۳
۲-۱- پرسش اصلی ۳
۲-۲- پرسش‌های فرعی ٤
۳- ضرورت و فایده انجام تحقیق ٤
٤- پیش‌فرض‌ها و فرضیه‌ها ٥
٥- محدودیت‌های تحقیق ٥
٦- نوآوری تحقیق ٥
۷- روش تحقیق ٦
۸- ساختار تحقیق ٦

فصل دوم: بخش‌های یک سیستم بازشناسی گفتار پیوسته

مقدمه ۹
۱- نمونه‌برداری از سیگنال صوتی ۹
۲- پردازش‌های قبل از استخراج ویژگی ۹
۲-۱- پیش‌تأکید (پیش‌پرداز سیگنال) ۱۰
۲-۲- قاب‌بندی ۱۰
۲-۳- پنجره‌گذاری ۱۰
۳- استخراج بردارهای ویژگی ۱۱
۳-۱- تحلیل مل کپستروم ۱۱
۳-۲- به‌کارگیری مشتق زمانی ضرایب بردار ویژگی ۱۳
٤- مدل‌سازی آکوستیک گفتار ۱۳
٤-۱- واحد آوایی ۱۳
٤-۲- مدل مخفی مارکوف (HMM) ۱٤
٤-۲-۱- ایده HMM ۱٤
٤-۲-۲- مدل‌سازی با استفاده از HMM ۱٥
٤-۲-۳- بیان مشکل ساخت مشاهده مربوط به HMM از سیگنال گفتار ۱٦
٥- رمزگشای آکوستیکی (رویه جستجو) ۱٦

٦- رمزگشای زبانی (مبدل واج به کلمه یا مدل زبانی) ١٧

فصل سوم: شرح برخی از الگوریتم‌های مرتبط با پروژه

١- خوشه‌بندی ٢٠
١-١- مقدمه ٢٠
١-٢- الگوریتم‌های خوشه‌بندی ٢١
١-٢-١- تکنیک‌های خوشه‌بندی مبتنی بر افراز داده‌ها ٢٣
١-٢-١-١- الگوریتم K-Means ٢٣
١-٢-٢- تکنیک‌های خوشه‌بندی سلسله‌مراتبی ٢٥
١-٢-٢-١- بالا به پایین یا تقسیم‌کننده ٢٥
١-٢-٢-٢- پایین به بالا یا متراکم شونده ٢٥
١-٢-٣- الگوریتم MKM ٢٦
٢- روش‌های یکسان‌سازی ابعاد داده‌ها ٢٧
٢-١- مقدمه ٢٧
٢-٢- پیچش زمانی پویا (DTW) ٢٨
٢-٢-١- محاسبه مسیر بهینه در DTW ٢٩
٢-٣- تبدیل فوریه و کاربردها ٣٠
٢-٣-١- کاربردهای تبدیل فوریه ٣٢
٢-٤- الگوریتم کاهش بعد PCA ٣٤

فصل چهارم: ساختارهای وابسته به بافت در بازشناسی گفتار پیوسته

١- مقدمه: بافت (زمینه یا سیاق) ٣٨
٢- روش‌های گوناگون به کارگیری اثر بافت در گفتار ٤٠
٢-١- روش‌های فونتیک ٤٠
٢-١-١- دوواجی ٤١
٢-١-٢- سه‌واجی ٤١
٢-١-٢-١- دسته‌بندی و گره زدن سه‌واجی‌ها ٤٢
٢-١-٢-٢- به کار بردن مدل‌های backed off ٤٣
٢-١-٢-٣- سه‌واجی مصنوعی ٤٣
٢-١-٣- هجا و نیم‌هجا ٤٣
٢-١-٤- واحدهای آوایی چندگانه ٤٤
٢-١-٤-١- استفاده از درخت تصمیم‌گیری ٤٤
٢-١-٤-٢- دسته‌بندی واج‌ها بر اساس شباهت آنها با مدل ٤٤
٢-١-٥- واحدهای آوایی وابسته به کلمه ٤٥
٢-٢- روش‌های آکوستیک ٤٥
٢-٢-١- استفاده از مشتق ضرایب بازنمایی ٤٥
٢-٢-٢- به‌کارگیری مدل‌های مارکوف مرتبه دوم و بالاتر ٤٦

۲-۲-۳- تخمین و به‌کارگیری پارامترهای آکوستیک وابسته به بافت ٤٦

فصل پنجم: شرح روش پیشنهادی

۱- مقدمه ٤٨
۲- مراحل مدل‌سازی ٤٨
۲-۱- مرحله اول: مدل مستقل از بافت ٤٨
۲-۲- مرحله دوم: مدل وابسته به بافت ٤٩
۳- دلیل استفاده از مدل وابسته به بافت ٤٩
٤- دلیل استفاده از واحد آوایی سه‌واجی در مدل وابسته به بافت ٤٩
٥- مشکلات استفاده از همه سه‌واجی‌های بالقوه ٥٠
٥-۱- راه حل مشکل اول ٥٠
٥-۱-۱- مراحل ادغام و گره زدن سه‌واجی‌ها ٥٠
٥-۲- راه حل مشکل دوم ٥١
٥-۳- راه حل مشکل سوم ٥١
٦- پیش از خوشه‌بندی (پیش‌پردازش‌ها) ٥٢
٦-۱- حل مشکل ناهمسانی طول ابعاد داده‌های خوشه‌بندی ٥٢
٦-۱-۱- راه حل اول (خام و ابتدایی) ٥٣
٦-۱-۲- راه حل دوم: روش DTW ٥٣
٦-۱-۳- راه حل سوم: تبدیل فوریه ٥٤
٦-۱-٤- راه حل چهارم: استفاده از الگوریتم PCA ٥٥
۷- خوشه‌بندی ٥٥
۷-۱- تعداد خوشه‌ها ٥٦
۷-۲- معیار شباهت در الگوریتم‌های خوشه‌بندی ٥٦
۷-۳- شرط توقف در الگوریتم‌های خوشه‌بندی ٥٦
۸- مقایسه دو الگوریتم خوشه‌بندی MKM و K-Means ٥٦
۹- پس از خوشه‌بندی (پس‌پردازش‌ها) ٥٧

فصل ششم: آزمایش‌ها و نتایج

۱- مقدمه ٥٩
۲- Kaldi ٥٩
۲-۱- کلدی در مقابل سایر ابزارها ٦٠
۳- پایگاه دادگان گفتاری ٦٠
٤- توضیح مختصری در مورد ساختار درختی تعریف شده در Kaldi ٦٢
٥- نتایج آزمایش‌ها ٦٥
٥-۱- نتایج مربوط به مدل مستقل از بافت ٦٥

۵-۲- مقایسه نتایج الگوریتم‌های یکسان‌سازی ابعاد ٦٦
۵-۳- انتخاب پارامتر مناسب برای تبدیل فوریه ٦٧
۵-٤- مقایسه K-Means و MKM ٦٧
۵-۵- إعمال پس‌پردازش..... ٦٩
۵-٦- مقایسه نتیجه روش پیشنهادی با روش به کار رفته در Kaldi..... ٦٩

فصل هفتم: نتیجه‌گیری و پیشنهادها

۱- جمع‌بندی و نتیجه‌گیری..... ۷۲
۲- محدودیت‌ها ۷۳
۳- پیشنهادها ۷٤
فهرست منابع و مآخذ ۷٤
Abstract 78

فهرست شکل‌ها و جدول‌ها

شکل ۲- ۱- مراحل استخراج ضرایب کپستروم..... ۱۱
شکل ۳- ۱- نمایش گرافیکی عملیات الگوریتم DTW ۲۹
شکل ۳- ۲- انتخاب محورهاي جديد براي دادههاي دو بعدي..... ۳٤
جدول ٦- ۱- مشخصات پایگاه دادگان TIMIT..... ٦۰
جدول ٦- ۲- مجموعه واج‌های اولیه (اصلی) TIMIT ٦۱
جدول ٦- ۳- نگاشت ٦۱ رده به ۳۹ رده..... ٦۲
جدول ٦- ٤- نتایج مربوط به مدل مستقل از بافت..... ٦٥
جدول ٦- ٥- مقایسه نتایج الگوریتم‌های یکسان‌سازی ابعاد..... ٦٦
جدول ٦- ٦- مقایسه نتایج الگوریتم DFT با پارامترهای مختلف..... ٦۷
جدول ٦- ۷- مقایسه نتایج الگوریتم‌های K-Means , MKM..... ٦۹
جدول ٦- ۸- مقایسه نتایج پیش و پس از اعمال پس‌پردازش‌ها..... ٦۹
جدول ٦- ۹- مقایسه نتایج روش پیشنهادی با روش به کار رفته در Kaldi ۷۰

فصل اول

مقدمات و کلیات

۱ - تبیین موضوع

یکی از پردازش‌های مهم و مطرح در حوزه پردازش صوت، بازشناسی گفتار[۱] می‌باشد. بازشناسی گفتار، به معنی تشخیص گفتار و تبدیل آن به یک سری پرونده‌های متنی می‌باشد. به الگوریتم‌های بازشناسي گفتار، اصطلاحاً روش‌هاي خودکار بازشناسي گفتار (ASR)[۲] گفته می‌شود.

واحد شناسایی در الگوریتم‌های بازشناسی گفتار، واج[۳] یا کلمه[۴] است. یعنی بعضی الگوریتم‌ها واج‌ها را تشخیص می‌دهند، اما بعضی دیگر مبتنی بر تشخیص کلمه می‌باشند.

سیستم‌های بازشناسی گفتار از لحاظ گفتار ورودی به چهار دسته تقسیم می‌شوند:

- سیستم بازشناسی کلمات مجزا:[۵] در این سیستم‌ها، در مرحله بازشناسی، گوینده هنگام بیان جمله، کلمات را با فاصله از یکدیگر بیان می‌کند. مکث بین هر دو کلمه متوالی سبب می‌شود که مرز بین کلمات به راحتی قابل شناسایی شود.
- سیستم بازشناسی گفتار متصل:[۶] در این سیستم‌ها، گوینده هر کلمه را با تأکید کافی بیان می‌کند و به دلیل وضوح الگوهای استرس، شناسایی مرزها آسان‌تر می‌شود.
- سیستم بازشناسی گفتار پیوسته:[۷] در این گونه سیستم‌ها، گوینده جملات را به طور طبیعی بیان می‌کند.
- سیستم‌های بازشناسی گفتار فی البداهه:[۸] در این گونه سیستم‌ها، گوینده، گفتاری با جملاتی ناقص، شروع‌های مجدد، خنده و سرفه بیان می‌کند که گفتار را از حالت سلیس بودن خارج می‌کند و اندازه مجموعه واژگان را عملاً نامحدود می‌کند [1].

---

1 . Speech Recognition
2 . Automatic Speech Recognition
3 . Phone
4 . Word
5 . Isolated Word Recognition
6 . Connected Speech Recognition
7 . Continuous Speech Recognition
8 . Spontaneous Speech Recognition

پیچیدگی مسئله بازشناسی گفتار در بازشناسی پیوسته، نسبت به سیستم‌های مجزا و متصل بیشتر است؛ زیرا مشکلاتی مانند مشخص نبودن مرز کلمات متوالی و ادغام شدن ابتدای یک کلمه با انتهای کلمه قبلی در این حالت وجود دارد [2].

الگوریتم‌های ASR را می‌توان از جنبه‌های مختلفی تقسیم‌بندی کرد. از یک جنبه، آنها را می‌توان به سه دسته زیر تقسیم‌بندی نمود.

- روش‌هاي مبتني بر قاعده: این روش‌ها در اوایل پیدایش الگوریتم‌های ASR استفاده می‌شدند. در اینجا هدف، تبدیل مستقیم بخش‌هاي گفتار به واج‌ها مي‌باشد.

- روش‌هاي مبتنی بر تطابق الگو: در این روش‌ها ابتدا برای هر واحد صوتی، الگوهایی استخراج می‌شد و سپس داده‌ای که می‌بایست بازشناسی گردد، با هر یک از الگوها مقایسه می‌شود. در نهایت داده به الگویی منتسب می‌گردد که بیشترین شباهت را با آن داشته باشد.

- روش‌های آماری: این دسته از روش‌ها در اصل زیرمجموعه‌ای از الگوریتم‌های انطباق الگو می‌باشند با این تفاوت که، معیار شباهت در اینجا، به صورت احتمالی است و داده به الگویی نسبت داده می‌شود که بیشترین احتمال تعلق را داشته باشد. مهم‌ترین روش در این دسته، بازشناسی گفتار با استفاده از مدل مخفی مارکف (HMM) می‌باشد که توسط Rabiner در اوائل دهه ۱۹۹۰ ارائه گردید [3] و تأثیر شگرفی در کارایی روش‌های بازشناسی گفتار داشت.

همچنین مدل‌سازی، می‌تواند بدون توجه به بافت (مستقل از بافت)[1]، و یا وابسته به بافت[2] باشد.

در ساختارهای مستقل از بافت، هر واحد آوایی (مثلاً هر واج) به طور مستقل مدل می‌شود و واحدهای آوایی اطراف آن مورد توجه قرار نمی‌گیرند. در حالی که واحدهای آوایی در هنگام مجاورت با هم، اثر متقابل بر یکدیگر گذاشته و تغییراتی در طرز ادای آنها ایجاد می‌شود. بخشی از مسئله مدل‌سازی اثر بافت، با انتخاب داده‌های آموزشی مناسب برای مدل تصادفی مورد استفاده حل می‌شود، ولی لازم است، روش‌های خاص برخورد با این مسئله نیز به‌کارگیری شود تا دقت بازشناسی تا حد ممکن بالا رود. این روش‌ها استفاده از ساختارهای وابسته به بافت در بازشناسی می‌باشند.

۲ - پرسش‌های تحقیق

۲-۱ - پرسش اصلی

چگونه می‌توان یک مدل بازشناسی گفتار وابسته به بافت جدید و کارا، برای گفتار پیوسته طراحی کرد؟

---

1 . Context-independent
2 . Context-dependent

۲-۲- پرسش‌های فرعی

- بازشناسی گفتار پیوسته چیست؟
- مدل بازشناسی گفتار وابسته به بافت، چه تفاوتی با مدل مستقل از بافت دارد؟
- واحد آوایی مورد استفاده در این پروژه چیست؟
- برای مدل‌سازی مدل وابسته به بافت این پروژه از چه الگوریتم‌هایی استفاده می‌شود؟
- شباهت‌ها و تفاوت‌های روش پیشنهادی با روش‌های پیشین چیست؟

۳- ضرورت و فایده انجام تحقیق

هدف نهایی از چنین پروژه‌ای، آشنایی کامل و عمیق با مدل HMM و Kaldi toolbox برای کاربرد بازشناسی گفتار پیوسته می‌باشد. این toolbox، با وجود این که بسیار قوی است، اما تاکنون در ایران، کمتر مورد توجه جدی قرار گرفته است؛ زیرا بسیار نوظهور و تازه است.

همچنین پیوستگی گفتار طبیعی، منجر به بروز مشکلاتی از قبیل عدم شناسایی واج‌ها به طور صحیح و دقیق می‌گردد که کماکان، جای پژوهش و تحقیق بیشتر در آن وجود دارد. این پژوهش نیز در تلاش است تا این معضل را تا حدودی حل کند.

فواید انجام چنین پژوهش‌هایی، با توجه به گسترش روزافزون کاربردهای بازشناسی گفتار، به سادگی قابل توجیه است.

از جمله کاربردهای بازشناسی گفتار می‌توان به استفاده از آن در سیستم‌های تجاری و اداری جهت دستیابی به بانک‌های اطلاعاتی، ورود داده و کنترل و مدیریت خودکار، ترجمه گفتار، کمک به معلولین جسمی اشاره کرد. کاربرد بازشناسی گفتار فقط به این موارد محدود نمی‌شود. از این سیستم‌ها می‌توان جهت کمک به ناشنوایان نیز بهره جست. با حل مسئله بازشناسی گفتار و پیاده‌سازی سخت‌افزاری آن به صورت قابل حمل، فرد ناشنوا می‌تواند صورت متنی پیام را بر روی صفحه نمایش ببیند که کمک بسیاری به این افراد خواهد کرد. همچنین کاربرد آن در مخابرات نیز می‌تواند زندگی بشر را آسان‌تر از گذشته کند. به عنوان مثال، در ساخت تلفن‌هایی با شماره‌گیر گفتاری و یا ساخت مراکز اطلاعات تلفن با اپراتور کامپیوتری از جمله کاربردهای آن در مخابرات است. سیستم‌های بازشناسی گفتار در صنعت نیز می‌تواند کاربرد داشته باشد. متداول‌ترین آنها، پیاده‌سازی سیستم‌های به اصطلاح «چشم آزاد، دست آزاد[1]» در کارخانه‌ها است. در موارد دیگر، از جمله ساخت اسباب‌بازی‌ها، کنترل بخش‌هایی از وسایل نقلیه نیز سیستم‌های بازشناسی گفتار می‌توانند به انسان‌ها خدمت کنند [4] [1].

1. Eye free-Hand free

۴- پیش‌فرض‌ها و فرضیه‌ها

Toolbox های معروفی چون HTK و Kaldi برای انجام بازشناسی گفتار در دسترس هستند.

برخی از الگوریتم‌های بازشناسی شناخته شده هستند و پیاده سازی برخی از آنها در دسترس است.

ما امیدواریم که بتوانیم مشکل ناهمسانی ابعاد مشاهدات را توسط الگوریتم DTW و یا تبدیل فوریه حل کنیم و از این طریق، عملکرد بازشناسی گفتار وابسته به بافت را بهبود بخشیم.

همچنین احتمالاً می‌توانیم از الگوریتم‌های بهتری برای دسته‌بندی و ادغام واحدهای آوایی (در پروژه ما، سه‌واجی‌ها) جهت حل مشکل وابستگی واج‌ها به بافت، بهره ببریم. نظیر استفاده از الگوریتم‌های K-Means یا MKM[1] برای خوشه‌بندی.

پیش‌بینی ما این است که الگوریتم MKM، بهتر از الگوریتم K-Means پاسخ دهد.

۵- محدودیت‌های تحقیق

از آن جا که محیط کاری Kaldi جدید است و کمتر مورد استفاده قرار گرفته است، و نیز تحقیقات زیادی راجع به آن صورت نگرفته است، لذا کار با آن، بسیار دشوار بوده و یادگیری و به‌کارگیری آن زمان‌بر است.

همچنین نتیجه بازشناسی گفتار، وابستگی شدیدی به تکرار آزمون و خطا و نیز تغییر پارامترهای گوناگون دارد.

به علاوه، بسیاری از الگوریتم‌های به‌کاررفته در واحدهای آوایی سه‌واجی پیچیده هستند و اجرای آنها بسیار زمان‌بر است.

ضمن این که به دلیل ناهمسانی ابعاد مشاهدات، الگوریتم‌های مرسوم خوشه‌بندی پاسخ‌گوی این داده‌ها نیستند و باید الگوریتم جدیدی برای این امر طراحی نمود.

نداشتن یک سیستم پردازشگر قوی و سریع، با وجود همه مشکلات فوق، موجب افزونی محدودیت‌ها و مشکلات می‌گردد.

۶- نوآوری تحقیق

بازشناسی گفتار مبتنی بر مدل‌های مارکوف، با ابزار Kaldi تاکنون کمتر در ایران مورد توجه قرار گرفته است.

به علاوه، این پژوهش در صدد آن است که پارامتر وابستگی به بافت را در گفتار پیوسته مورد توجه جدی قرار دهد. و ساختاری جدید ارائه دهد که نسبت به ساختارهای ارائه شده مشابه، بهتر عمل کنند.

---

1 . Modified K-Means

واحد آوایی به کار رفته در این پژوهش، سه‌واجی است. همچنین از الگوریتم‌های متعددی برای پیش‌پردازش و نیز خوشه‌بندی استفاده شده است.

به‌کارگیری برخی از این‌ها، پیشتر، به صورت مستقل و جداگانه، مورد پژوهش قرار گرفته است، اما به صورتی جامع به‌کار گرفته نشده است. علاوه بر این که ایده استفاده از الگوریتم‌های PCA[1] و DFT[2] (تبدیل فوریه گسسته) برای همسان‌سازی ابعاد مشاهدات، بدیع به نظر می‌رسد.

۷- روش تحقیق

پروژه شامل دو بخش است: بخش اول، عملی و کاربردی است. بخش دوم، خود شامل دو قسمت است: قسمت اول، شامل معرفی تئوری و پیش‌نیازهای لازم برای فهم پروژه؛ و قسمت دوم، شامل گزارش بخش عملی است.

روش گردآوری بخش تئوری و پیش‌نیازهای لازم، کتابخانه‌ای و استفاده از منابع اینترنتی بوده است.

اما بخش اول، کاملاً عملی و کاربردی است که در آن از ابزارهایی همچون Kaldi (که ابزاری برای بازشناسی گفتار است)، Matlab و برخی زبان‌های برنامه‌نویسی دیگر همچون C++، Shell-Programming و ... استفاده شده است.

اما برای آن که جزئیات بیشتری از روش به کار رفته در این پروژه ارائه کرده باشیم، باید توجه داشته باشیم که در ساختار وابسته به بافت ارائه شده، ما هر واجی را در وابستگی واج پیشین و پسین خود در نظر می‌گیریم. بنابراین هر واحد آوایی ما در واقع، سه‌واجی خواهد بود. تعداد این سه‌واجی‌ها بالقوه بسیار زیاد است. ما نیاز داریم که تعداد آنها را کاهش دهیم و سپس آنها را خوشه‌بندی کنیم. روش‌های متفاوتی برای خوشه‌بندی ارائه شده است. اما در همه این روش‌ها، لازم است که داده‌های ورودی، از نظر ابعاد یکسان باشند. ما از روش PCA که ذاتاً برای کاهش بعد استفاده می‌شود، و نیز از تبدیل فوریه، که ذاتاً برای بهسازی گفتار و فشرده‌سازی گفتار استفاده می‌شود، و نیز از روش DTW، با برخی اصلاحات و تغییرات، برای آماده‌سازی داده‌های خوشه‌بندی استفاده خواهیم کرد. برای خوشه‌بندی نیز از هر دو روش K-Means و MKM استفاده خواهیم کرد.

۸- ساختار تحقیق

این پژوهش، علاوه بر این فصل (فصل اول: کلیات) شامل فصل‌های زیر نیز می‌باشد:

مبانی تئوری پژوهش را به دلیل طولانی بودن، به دو فصل دوم و سوم تقسیم کردیم.

---

1 . Principle Component Analysis
2 . Discrete Furriere Transform

فصل دوم: شامل معرفی اجزای مختلف یک سیستم بازشناسی گفتار پیوسته.

فصل سوم: بررسی الگوریتم‌های مرتبط با پروژه، شامل الگوریتم‌های PCA, DTW, DFT و نیز الگوریتم‌های مختلف خوشه‌بندی.

فصل چهارم: پیشینه پژوهش: ساختارهای مختلف وابسته به بافت؛

فصل پنجم: شرح روش پیشنهادی؛

فصل ششم: گزارش آزمایش‌های پروژه؛

فصل هفتم: نتیجه‌گیری و پیشنهادها.

فصل دوم

بخش‌های یک سیستم

بازشناسی گفتار پیوسته

مقدمه

سیستم بازشناسی گفتار، شامل اجزای مختلفی می‌شود. همان گونه که در فصل گذشته توضیح داده شد، می‌توان گفتار را به چهار نوع مجزّا، متصل، پیوسته و فی‌البداهه تقسیم کرد.

بسته به نوع بازشناسی برای گفتار، ممکن است، برخی از این اجزا در سیستم موجود نباشند یا با جزئیات بیشتری در آن سیستم در نظر گرفته شده باشند [5].

۱ - نمونه‌برداری از سیگنال صوتی

برای استفاده از الگوریتم‌های پردازش سیگنال، باید موج پیوسته سیگنال گفتار ورودی را به شکل گسسته تبدیل نمود. به این منظور باید از موج گفتار ورودی نمونه‌برداری کرد. طبق قضیه نایکوئیست[1] (قضیه نمونه‌برداری)[2]، فرکانس نمونه‌برداری موج پیوسته، باید حداقل دو برابر بزرگ‌ترین مؤلفه فرکانسی موجود در موج پیوسته باشد [5].

قضیه نایکوئیست بیان می‌دارد که اگر بزرگ‌ترین مؤلفه فرکانسی موجود در $X(f)$ برابر با f_{max} باشد، و از سیگنال با نرخی بیشتر از $2f_{max}$ نمونه‌برداری شود، آن‌گاه سیگنال پیوسته $x(t)$ به طور کامل از روی سیگنال گسسته $x[n]$ قابل بازیابی می‌باشد. به $2f_{max}$ نرخ یا فرکانس نایکوئیست گفته می‌شود.

۲ - پردازش‌های قبل از استخراج ویژگی[3]

پیش از استخراج ویژگی، یک سری پیش‌پردازش روی سیگنال انجام می‌شود، مانند قاب‌بندی[4]، پیش‌تأکید[5]، اعمال پنجره، تبدیل فوریه زمان کوتاه و ... [3] [6].

1 . Nyquist theorem
2 . Sampling theorem
3 . Feature Extraction
4 . framing
5. Pre-Emphasizing