

Conceptual Design

PATTERN RECOGNITION FINAL PROJECT – PHASE III

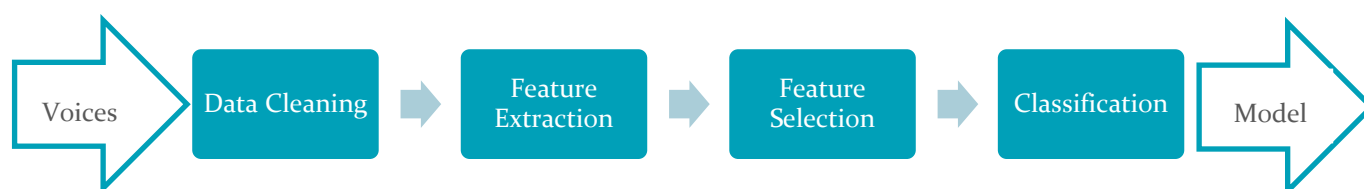
Fatemeh Ghaffari | 810198317

مسائل انتخاب شده

از بین مسائل ذکر شده، دو مسئله ی « ۱ - تشخیص جنسیت از روی صدا ، ۲ - خوشه بندی صداها » انتخاب می شود.

روند کلی مسائل یادگیری ماشین

همان طور که در درس ارائه شد، هر مسئله یادگیری ماشین برای تبدیل داده های خام به مدل باید مراحل شکل زیر را پشت سر بگذارد.



در ادامه به هر یک از این مراحل و ابزار انتخابی در هر مرحله برای مسائل مطرح شده پرداخته میشود.

تمیز سازی داده ها

برای داده های ما که صدای انسان هستند، لازم است مراحل زیر صورت بگیرد:

۱ - یکسان سازی نرخ نمونه برداری

۲ - حذف بخش های سکوت

۳ - حذف نویز

که این مراحل در قالب فرایند VAD انجام می شود.

استخراج ویژگی ها

برای حل مسائلی که درمورد شناسایی و تشخیص صدای انسان هستند، ویژگی های مربوط به فرکانس بیشترین کاربرد را دارند. تعداد زیادی از این ویژگی ها وجود دارند که تعدادی از معروفترین آنها که بیشتر در پروژه ها

استفاده شده اند عبارتند از: MFCC (Mel-Frequencies Cepstral Coefficients) Filter Banks, Spectral roll-off, Spectral centroid و ...

با توجه به بررسی پروژه های مشابه صورت گرفته به این نتیجه رسیدم که قریب به اتفاق آنها از ویژگی های MFCC استفاده کرده اند و طبقه بندی که بیشتر از سایرین برای این نوع ویژگی مناسب است، GMM یا HMM است.

روش MFCC به صورت کلی به این صورت است: سیگنال به بازه های زمانی کوچکتر تقسیم شده و طیف توان هر قسمت تخمین زده می شود و فیلتر MEL به آن اعمال می شود. جمع انرژی هر فیلتر محاسبه شده، لگاریتم آن محاسبه شده و ضرایب DCT آن محاسبه می شود و ضرایب ۲ تا ۱۳ نگه داری می شود.

انتخاب ویژگی ها

برای انتخاب ویژگی ها روش های مختلفی وجود دارد که یکی از کارا ترین و معمولترین آنها PCA است. همچنین از ۲۰ فیچر ابتدایی دسته فیچرهای MFCC برای مصارف تبدیل صدا به متن استفاده می شود. از این مسئله می توان نتیجه گرفت که این دسته ویژگی ها نباید ارتباط چندانی با هویت صاحب صدا داشته باشند. بنابراین از ابتدا می توان این دسته از ویژگی ها را کنار گذاشت.

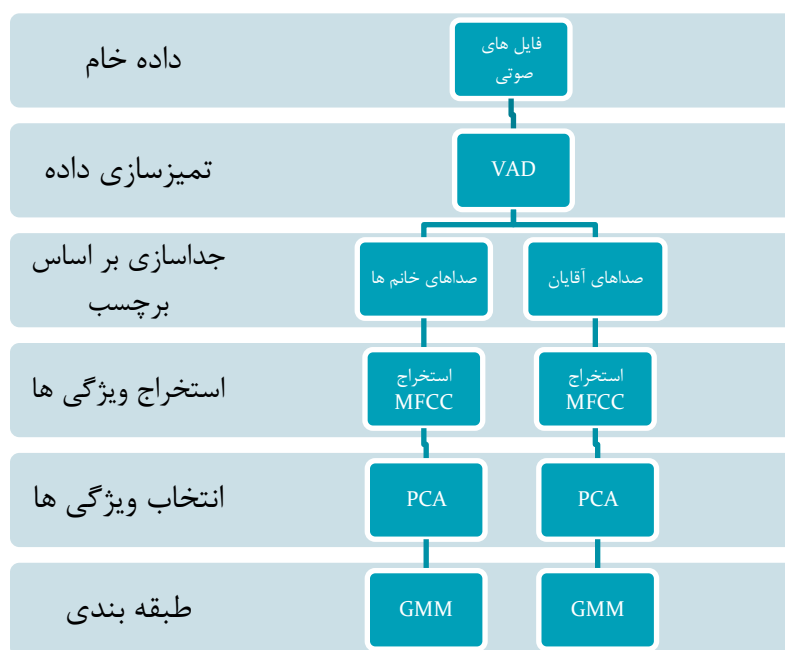
طبقه بندی

همانطور که قبل تر اشاره شد در بیشتر پروژه های مشابهی که بررسی شدند. از طبقه بندهای GMM یا HMM در کنار ویژگی های MFCC استفاده شده بود. در نتیجه ظاهراً این دو طبقه بند نسبت به سایرین نتیجه بهتری برای این مصرف داشته باشند.

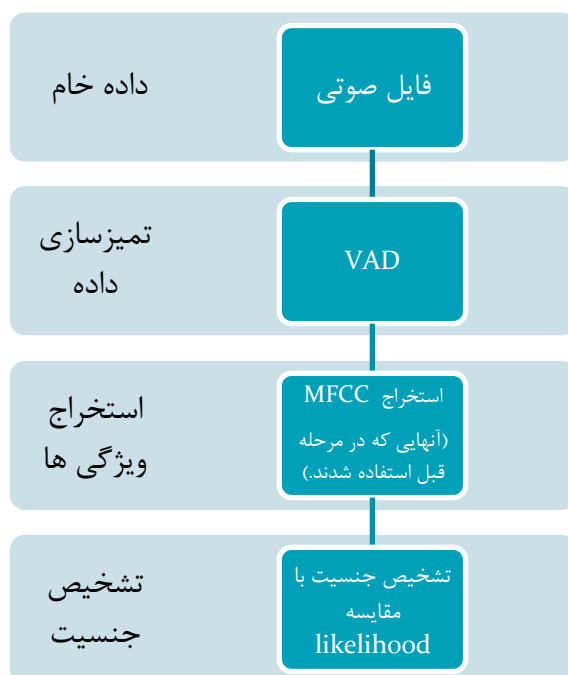
بلوک دیاگرام راه حل

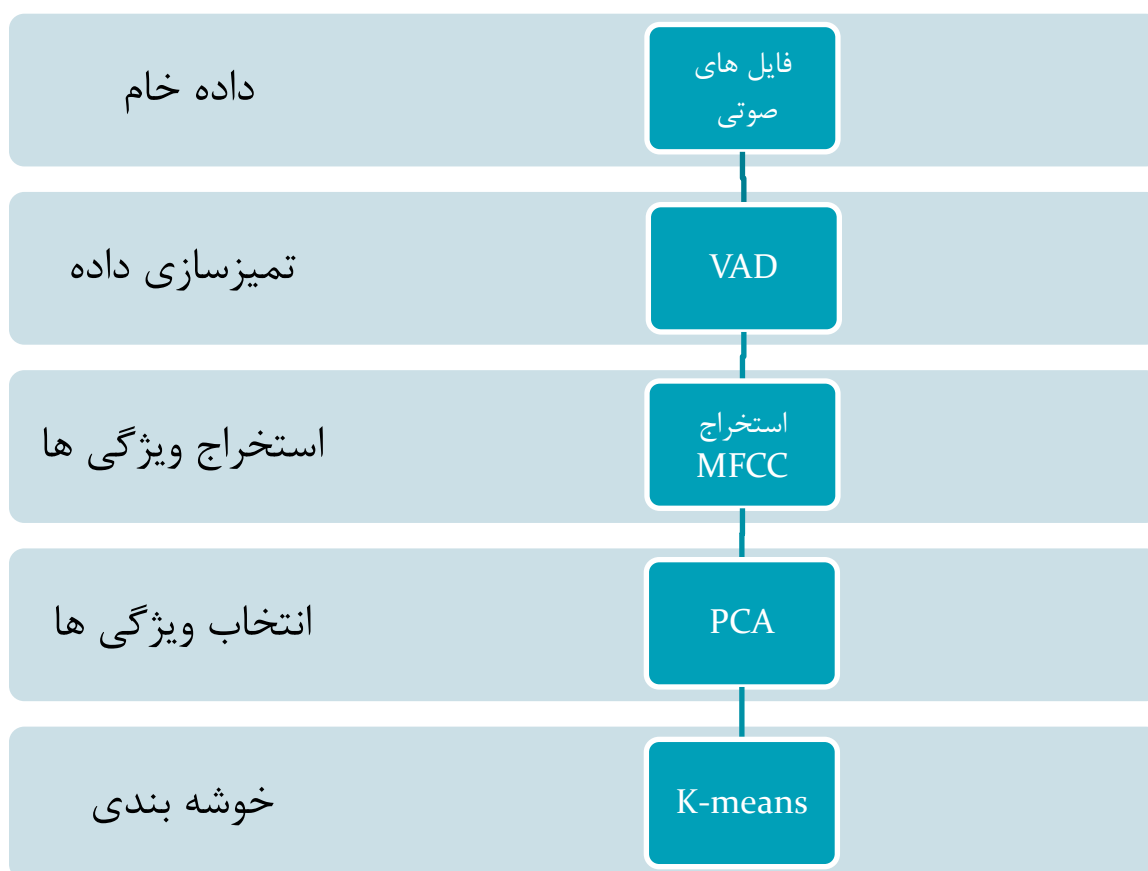
تشخیص جنسیت

Train



Test





چالش ها

در هر یک از بخش های بالا ممکن است با چالش هایی مواجه شویم. از قبیل یکسان نبودن فرمت ها صداها یا فرمت برچسب داده ها. همچنین ممکن است با خطای بالایی مواجه شویم و به این نتیجه برسیم که باید طبقه بند را عوض کنیم.

نحوه ارزیابی خروجی

خروجی را می توان به وسیله متریک های عددی، ماتریسی و نموداری ارزیابی نمود. استفاده از متریک های confusion matrix, accuracy و نمودار ROC انتخاب مناسبی به نظر می رسد. همچنین سرعت تشخیص درست برای یک داده تست هم معیار مناسبی به نظر می رسد.

عملکرد موردنظر

سایر پروژه هایی که در شرایط مشابه و با ابزارهای مشابه انجام شده بودند، خطایی بین ۵ تا ۱۰ درصد داشتند. با توجه به چالش های احتمالی انتظار می رود برای این پروژه خطایی بین ۲۰-۱۰ درصد داشته باشیم.