دانشگاه صنعتی امیرکبیر
(پلی‌تکنیک تهران)
دانشکده مهندسی کامپیوتر

گزارش کارآموزی

محل کارآموزی: شرکت عصر گویش‌پرداز

بازشناسی خودکار گفتار در ارتباطات کنترل ترافیک هوایی

نگارش: زهرا رحیمی

نام استاد کارآموزی: دکتر محمد رحمتی

مرداد و شهریور ۱۴۰۱

دانشگاه صنعتی امیرکبیر
(پلی‌تکنیک تهران)
دانشکده مهندسی کامپیوتر

گزارش کارآموزی

محل کارآموزی: شرکت عصر گویش‌پرداز

بازشناسی خودکار گفتار در ارتباطات کنترل ترافیک هوایی

نگارش: زهرا رحیمی

نام استاد کارآموزی: دکتر محمد رحمتی

مرداد و شهریور ۱۴۰۱

سپاس‌گزاری

بدینوسیله مراتب قدردانی و امتنان خود را خدمت

جناب آقای گوران مدیر فنی شرکت عصر گویش، بابت هماهنگی این دوره کارآموزی، آموزش‌ها و کمک‌های بی‌دریغشان که بدون راهنمایی‌های ایشان، تهیه این گزارش امکانپذیر نبود،

جناب آقای دکتر رحمتی استاد راهنمای گرانقدر، بابت راهنمایی‌های همیشگی‌شان،

جناب آقای دکتر صامتی استاد گرانقدر، بابت گردآوری کارآموزان و فراهم آوردن این فرصت،

سرکار خانم صادقی بابت کمک‌های دلسوزانه‌شان در طول دوران کارآموزی،

تمامی دست‌اندرکاران شرکت عصر گویش که باعث شدند حضور در کنارشان علاوه بر یادگیری‌های بسیار، تجربه زیسته و کاری پرباری را هم برای بنده داشته باشد،

ابراز و از زحماتشان صمیمانه سپاسگزارم و برایشان آرزوی توفیق روزافزون دارم.

زهرا رحیمی

بهار ۱۴۰۱

چکیده

مقدار قابل توجهی از مکالمات بین کنترلر ها و خلبانان از طریق کانال های رادیویی است. لذا رونویسی خودکار این مکالمات نه تنها باعث بهبود امنیت سیستم بلکه باعث پیشرفت عملکرد های عملیاتی می‌شود. با این حال سیستم های بازشناسی گفتار خودکاری که تا به امروز پیشنهاد شده‌اند دقت لازم برای استفاده های عملی را دارا نبوده‌اند. عواملی مانند کانال های رادیویی نویز دار، سرعت تکلم بالا و لهجه‌های متنوع چالش‌هایی را برای توسعه بازشناسی گفتار برای کنتلرهای ترافیک هوایی به وجود می‌آورند کما اینکه فقدان داده‌های گفتاری در مقیاس بزرگ نیز کار را برای سیستم‌های تشخیص خودکار سخت‌تر می‌کرد، اما از سوی دیگر این مکالمات دارای واژگان خاص و مشخص و همینطور عبارت های استانداردی هستند که می‌توان از آن ها برای جهت دهی به الگوریتم‌ها و تقویت آنها در این زمینه استفاده کرد. در این پژوهش سعی داریم قدم به قدم مراحل توسعه یک سیستم تشخیص گفتار خودکار را برای داده‌های خلبانی و ترافیک هوایی را با استفاده از مدل wav2vec 2.0 که در ادامه بیش‌تر به آن می‌پردازیم، معرفی کنیم.

واژه‌های کلیدی:

بازشناسی گفتار خودکار، تشخیص گفتار خودکار، مدل wav2vec 2.0، یادگیری خود نظارتی، نرخ خطای کلمه، تنظیم دقیق کردن مدل wav2vec 2.0

فهرست مطالب

عنوان صفحه

فهرست اشکال و جداول

فصل اول

معرفی موضوع و آشنایی با شرکت عصر گویش‌پرداز

۱- معرفی موضوع و آشنایی با شرکت عصر گویش‌پرداز

در این بخش به بیان معرفی شرکت عصر گویش‌پرداز می‌پردازیم که تاریخچه تاسیس شرکت، اهداف و محصولات آن را شامل می‌شود. سپس به موضوع و هدف کارآموزی خواهیم پرداخت که در آن خلاصه‌ای از موضوع پروژه و مقاصدی که در طول مدت کارآموزی بایست به آن پرداخت، را مرور می‌کنیم.

۱- ۱ آشنایی با شرکت

شرکت عصرگویش پرداز (سهامی خاص) به طور تخصصی فعالیت خود را از سال ۱۳۸۲ در زمینه پردازش و تشخیص گفتار به سرپرستی دکتر حسین صامتی شروع کرده است. سابقه و تجربه تخصصی شرکت به تحقیقات چندین ساله متخصصان این زمینه از دانشگاه صنعتی شریف برمی‌گردد که قبل از تاسیس رسمی شرکت کارهای تحقیقی را به منظور توسعه تعدادی از سیستم‌ها و موتورهای نرم‌افزاری شروع کرده‌اند. عمده محصولات و خدمات ارائه شده توسط این شرکت برای نخستین بار در کشور و به صورت حرفه‌ای در زمینه‌های پردازش و تشخیص گفتار بوده است. عصرگویش پرداز به عنوان اولین شرکت پیشرو در ارایه سیستم‌های مبتنی بر ساده‌ترین وسیله ارتباطی انسان برای زبان فارسی، علاوه بر توسعه تعدادی از سیستم‌ها و راه‌حل‌های مبتنی بر گفتار مانند سیستم دیکته زبان فارسی، سیستم تشخیص گفتار تلفنی، جستجوگر کلمات در گفتار ، تبدیل متن به گفتار و ... برای زبانهای فارسی و انگلیسی، توانایی انجام کلیه فعالیت‌های دیگر مبتنی بر گفتار را دارد. از آنجا که ارتباط کلامی راحت‌ترین، ساده‌ترین و سریع‌ترین راه ارتباطی می‌باشد با کمک سیستم‌های تشخیص گفتار عصر گویش پرداز می‌توان با رایانه‌ها از طریق صحبت ارتباط برقرار نمود، با آنها حرف زد، دستور داد یا از پشت تلفن و از راه دور بتوان سیستم‌های خانگی را کنترل نمود. با کمک این محصولات، بسیاری از افراد معلول و یا افرادی با آشنایی محدود با کامپیوتر و زبان‌های خارجی نیز می‌توانند تنها از طریق صحبت‌کردن با کامپیوتر ارتباط برقرار نمایند. در حال حاضر موتور تشخیص گفتار در این شرکت طراحی و پیاده‌سازی شده است که پایه و هسته اصلی سیستم‌های تشخیص گفتار فارسی است. این سیستم بر اساس آخرین تکنولوژی و استفاده از منابع علمی روز طراحی شده و دقتی بسیار قابل قبول در مقایسه با سیستم‌های معروف خارجی دارد.

برخی از محصولات:

- نویسا: نخستین سامانه تایپ گفتاری فارسی
- نیوشا: نخستین سامانه تلفن گویای هوشمند مبتنی بر گفتار
- آریانا: سامانه متن به گفتار فارسی با صدای طبیعی

- شناسا: تعیین هویت گوینده
- رمزآوا: احراز هویت گوینده
- بینا: تصویر خوان هوشمند
- رومند: چت بات هوشمند
- جویا: سامانه جستجوی عبارات و کلمات در گفتار
- پوشا: سامانه پنهان سازی اطلاعات در تصویر (استگانوگرافی)
- پدیدا: سامانه کشف تصاویر نهان نگاری شده
- پارسیا: اولین نرم‌افزار متـرجم گفتار به گفتار فارسی به انگلیسی/عربی
- نویسیار: اولین نرم‌افزار تایپ هوشمند فارسی
- کارا: نخستین سامانه تشخیص فرمان صوتی برای ویندوز

۱- ۲ محصولات مرتبط با تشخیص و سنتز گفتار

تعدادی از محصولات شرکت که بر اساس موتور تشخیص گفتار و سنتز گفتار توسعه داده شده‌اند، شامل موارد زیر می‌باشد:

- سیستم متن خوان فارسی برای خواندن متون با صدای طبیعی
- سیستم‌های تلفن گویا برای ارتباط تلفنی از راه دستورات صوتی
- سیستم‌های تشخیص دستورات صوتی مانند کنترل برنامه‌ها یا فرم‌های صوتی
- منشی تلفنی خودکار با قابلیت فهم گفتار تماس گیرنده
- جستجوگر واژه‌های کلیدی برای جستوگرکلامی در سیستم‌های امنیتی
- فهم گفتار در خودروها یا ساختمان‌های هوشمند
- تایپ هوشمند فارسی با قابلیت فعال شدن در همه محیط‌های تایپ جهت افزایش موثر سرعت تایپ مترجم کلامی فارسی-انگلیسی با امکانات محدود

علاوه بر زمینه‌های پردازش سیگنال‌ها و بویژه سیگنال‌های صوتی و تشخیص اتوماتیک گفتار، محققان این شرکت در زمینه‌های دیگری چون افزایش کیفیت گفتار، ، تبدیل گفتار به متن، پردازش زبانهای طبیعی شامل روش‌های آماری، دستوری و معنایی زبان، پردازش تصویر درمرحله تحقیق و توسعه سیستمها می‌باشند که هم اکنون برخی از این محصولات در اختیار کاربران قرار گرفته است. به علاوه این محصولات می‌تواند به زبان‌های دیگر و از جمله زبان انگلیسی نیز توسعه داده شود. این شرکت افتخار دارد با تلاش محققان وطن دوست توانسته است به یکی از تکنولوژی روز دنیا دست یابد و در حال حاضرآماده همکاری با

شرکت‌ها، موسسات و سازمان‌هایی است که خواهان استفاده از محصولات عصر گویش پرداز جهت تسریع بخشیدن در کار مدیران یا تکریم ارباب رجوع می باشد.

۱- ۳ توانمندی‌ها و خدمات

هدف این سامانه ایجاد ارتباط بین انسان و ماشین از طریق گفتار است. بدین معنی که انسان برای انجام کارها به جای استفاده از کلید و دکمه، با صحبت کردن درخواست خود را به رایانه یا دستگاه منتقل نماید.

۱- ۳- ۱ اجرا و کنترل برنامه‌های رایانه به کمک گفتار

این قابلیت کاربران را قادر می‌سازد تا بتوانند با استفاده از گفتار، کارهای کامپیوتری را انجام داده و یا نرم‌افزارها را کنترل نمایند. به عنوان مثال، کاربر می تواند با گفتن "به اینترنت وصل شو" مرورگر اینترنت را باز نماید. یا با گفتن "اندازه نوشته را بزرگ‌تر کن" اندازه متن نوشته شده در ویرایشگر را بزرگ‌تر نماید. به صورت مشابهی، کاربر می‌تواند فرمان‌های صوتی مختلفی را در نرم‌افزارهای نصب شده در رایانه تعریف نموده و با بیان آنها، نرم‌افزارها را کنترل کند. از فرمان‌های صوتی می‌توان برای افزایش قابلیت‌های جدید به نرم افزارهای مختلف مانند بازی‌ها و نرم‌افزارهای آموزشی استفاده نمود.

۱- ۳- ۲ اتوماسیون خانگی و صنعتی با به کارگیری تشخیص گفتار

هدف این سیستم، ارایه راه حلی برای تشخیص گفتار از راه دور جهت کنترل وسایل و ابزارهای مورد استفاده می‌باشد. از کاربردهای این سیستم، استفاده از گفتار در خودرو، منزل و یا کارخانه برای اجرای فرمان‌های متنوعی مانند روشن یا خاموش کردن یک دستگاه، کنترل کردن ربات‌ها و موارد مشابه می‌باشد. این سیستم می‌تواند از پشت خط تلفن نیز به منظور کنترل از راه دور در ساختمان‌های هوشمند مورد استفاده قرار گیرد

۱- ۳- ۳ پردازش گفتار در تلفن‌های همراه وDSP ها

هرچند استفاده از پردازشگرهای قوی در تلفن‌های همراه رو به افزایش است ولی توسعه نرم‌افزارها در این بسترها با توجه به میزان پردازش موردنیاز کار دشواری است. شرکت عصر گویش پرداز آمادگی دارد سامانه تشخیص گفتار، متن به گفتار و تشخیص هویت گوینده را با کارایی بالا و سرعت پردازش بهینه برای گوشی‌ها و DSPها توسعه دهد. برخی از کاربردهای این سیستم‌ها به صورت زیر است:

- اجرای فرمان‌های صوتی برروی تلفن همراه یا سخت‌افزارها

- شماره‌گیری یا تایپ گفتاری پیامک در تلفن همراه
- مترجم صوتی گفتار به گفتار (به صورت همراه)
- مجوز دسترسی با دستگاه‌های تایید هویت با صدا
- سخن‌گو کردن دستگاه‌ها (مانند ربات)

۱- ۳- ۴ پردازش زبان طبیعی(NLP)

یکی از پیش‌نیازهای سیستم‌های هوش مصنوعی مانند تشخیص گفتار، تبدیل متن به گفتار، ترجمه ماشینی، بازشناسی نویسه های نوری و تصحیح خطاهایی تایپی، برخورداری از اطلاعات زبانی است. شرکت عصر گویش پرداز جهت گردآوری، استخراج و به‌کارگیری اطلاعات زبانی در سیستم‌های خود از آخرین روش‌های موجود در زمینه پردازش زبان‌های طبیعی استفاده کرده است که نتیجه آن استخراج حجم وسیعی از اطلاعات زبان فارسی برای نخستین بار بوده است. از جمله این اطلاعات که در سیستم‌های تشخیص گفتار این شرکت مورد استفاده قرار گرفته است، پیکره‌های بزرگ متنی، مدل‌های زبانی آماری فارسی، مدل گرامری فارسی و مجموعه واژگان‌های مختلف برای زبان فارسی می‌باشد. این اطلاعات می‌تواند به صورت‌های مختلفی در نرم‌افزارهای کاربردی و فعالیت‌های پژوهشی مورد استفاده قرار گیرد.

۱- ۳- ۵ نرم‌افزارهای آموزشی و چندرسانه‌ای

ندر بسیاری از نرم‌افزارهای آموزشی مانند آموزش زبان خارجی، آموزش قرآن نیاز به بخش هوشمندی است که کاربران بتوانند میزان یادگیری خود در بیان جملات را ارزیابی کنند. بررسی میزان صحت تلفظ کلمات و عبارات در نرم‌افزارهای مختلف قابل استفاده بوده و بر اساس تکنیک‌های بازشناسی الگو و مدل‌سازی آماری، شباهت میان کلمه/عبارت تلفظ شده توسط کاربر و کلمه/عبارت مرجع را محاسبه می‌کند. قابلیت متن به گفتار نیز در نرم‌افزارهایی مانند کتاب صوتی و هر نرم‌افزاری که نیاز دارد اطلاعات مختلفی را به کاربر اعلام کند توسط ماژول آریانا قابل انجام است.

۱- ۳- ۶ بهبود کیفیت گفتار

در بسیاری از کاربردها بهبود کیفیت شنیداری صوت یا گفتار و یا قابل فهم کردن آن موردنیاز است. مثلاً حذف صداها و خش‌های اضافی از نوارهای قدیمی یا بهبود فایل‌های ضبط شده در یک سخنرانی باعث بهتر شدن کیفیت آرشیوهای صوتی می‌شود. بر اساس تحقیقات انجام شده شرکت عصرگویش پرداز با بهره‌گیری از آخرین روش‌های موجود در این زمینه قادر به توسعه محصولی برای انجام این کار می‌باشد که می‌تواند

هم به صورت یک نرم‌افزار مستقل مورد استفاده قرار گیرد و هم به صورت یک واحد مجزا در نرم‌افزارهای دیگر به کار گرفته شود. به عنوان مثال استفاده از این واحد در سیستم‌های بازشناسی گفتار در محیط های نویزی مانند محیط نمایشگاه یا داخل ماشین کارایی و دقت این سیستم‌ها را بهبود می‌دهد.

۱- ۴ معرفی پروژه

در ابتدا طرح پروژه را داریم که به شرح زیر است:

بازشناسی گفتار در محیط های پرنویز و با صدای کم کیفیت یکی از موارد چالشی حوزه پردازش گفتار است. یکی از مواردی که با تعداد زیادی از این نوع داده سر و کار دارد، داده های مکالمات کنترل هوایی است. در این پروژه قصد داریم با بررسی مقدار زیادی از این نوع داده مدل بازشناسی گفتاری تهیه کنیم که قابلیت بازشناسی این نوع داده را داشته باشد. انتظار می رود یک گزارش اولیه پس از تحقیقات انجام شده در وبلاگ شرکت قرار گیرد و اجتماع این گزارش و گزارش های هفتگی نیز گزارش نهایی پروژه را بسازد.

بعد از مطالعه مقالات مختلف و اولویت‌بندی آنها (لینک گوگل‌شیت) که در این موضوع بحث کرده‌بودند و مشورت با سرپرست کارآموزی، مقاله [۱] به عنوان مقاله مرجع انتخاب شد تا ادامه فعالیت های من با استناد به این مقاله پیش رود. رویکردی که در این مقاله در پیش گرفته شده بود، بر اساس روش یادگیری خود نظارتی بود که با استفاده از مدل wav2vec2 و مقایسه آن با مدل هیبریدی و نتایج بسیار کارآتر آن، به این نتیجه رسیدند که حتی در حوزه‌های جدیدی مانند ترافیک هوایی هم می‌توان با مدل wav2vec2 نتایج بهتری نسبت به مدل‌های پیشین بدست آورد.

۱- ۵ خلاصه و ساختار گزارش

مقصود از این سند، تهیه گزارشی از آموزش‌ها و کارهای انجام گرفته در دوره کارآموزی، به‌همراه توضیحات کامل درمورد ابزارها و مفاهیم علمی مورد استفاده است. در ابتدا به معرفی شرکت، مقدمات این دوره یعنی اهداف کارآموزی و لزوم انجام کارهای انجام گرفته ارائه شده است. در فصول آینده، به صورت عمیقتر به هر یک از موضوعات پرداخته خواهد شد.

فصل دوم

ابزارها و مفاهیم علمی

۲- ابزارها و مفاهیم علمی

در این بخش می‌خواهیم با ابزارها و مفاهیمی که لازم بود تا در ابتدای دوره با آن آشنا شویم، بپردازیم:

۲- ۱ محیط گوگل‌کولب[۱]

به طور دقیق،کولب یک محیط نوت بوک رایگان ژوپیتر[۲] است که به طور کامل در فضای ابری اجرا می‌شود. پروژه ژوپیتر پروژه‌ای با اهداف توسعه نرم‌افزار منبع‌باز و استانداردهای باز برای محاسبات تعاملی در چندین زبان برنامه نویسی است. نکته مهم اینکه نوت‌بوک‌هایی که ایجاد می‌شود این قابلیت را دارند که به طور همزمان توسط اعضای تیم ویرایش شوند. کولب از بسیاری از کتابخانه‌های معروف یادگیری ماشین پشتیبانی می‌کند که می توانند به راحتی در نوت بوک بارگیری شوند.

کارهایی که می‌شود با استفاده از گوگل‌کولب انجام داد:

- نوشتن کد در پایتون و اجرای آن
- مستندسازی کد
- ایجاد/آپلود/اشتراک‌گذاری نوت‌بوک
- وارد کردن/ذخیره نوت‌بوک از/به گوگل‌درایو
- وارد کردن/انتشار نوت‌بوک‌ها از گیت‌هاب
- وارد کردن مجموعه‌داده‌های خارجی به عنوان مثال از کگل[۳]
- سرویس ابری رایگان با جی‌پی‌یو[۴] رایگان [۹]

۲- ۲ پلتفرم هاگینگ‌فیس 🤗

هاگینگ‌فیس فقط یک ایموجی خندان در فضای مجازی نیست! بلکه یک شرکت آمریکایی است که ابزارهایی را برای ساخت برنامه‌های کاربردی با استفاده از یادگیری ماشین توسعه می‌دهد. این پلتفرم به دلیل کتابخانه ترنسفورمر[۵] مورد توجه قرار گرفته است زیرا برای برنامه‌های پردازش زبان طبیعی ساخته شده است و به کاربران این امکان را می‌دهد که مدل‌ها و مجموعه داده‌های یادگیری ماشین را به اشتراک بگذارند.

---

۱ Google Colab

۲ Jupyter

۳ Kaggle

۴ GPU

۵ transformer

۲- ۲- ۱ تاریخچه مختصر

این شرکت در سال ۲۰۱۶ تأسیس شد که در ابتدا یک چت‌بات را با هدف نوجوانان توسعه می‌داد. پس از اینکه مدل توسعه‌داده شده ربات‌چت در دسترس همگان قرار گرفت، شرکت تمرکز خود را بر روی پلتفرمی برای دموکراتیک کردن یادگیری ماشین متمرکز کرد. سپس در سال ۲۰۲۱، این شرکت کارگاه تحقیقاتی بیگ‌ساینس[۶] را با همکاری چندین گروه تحقیقاتی دیگر برای انتشار یک مدل زبان بزرگ باز راه‌اندازی کرد و اقدام به خرید Gradio کرد که یک کتابخانه نرم افزاری با هدف ایجاد تعامل در مدل های یادگیری ماشین است. در سال ۲۰۲۲ یک مدل زبان بزرگ چندزبانه با ۱۷۶ میلیارد پارامتر به پایان رسید و اعلام داشت که برنامه "سفیر دانشجویی" تا آخر ۲۰۲۳ ماموریت خود جهت آموزش یادگیری ماشین به ۵ میلیون نفر را احقاق خواهد کرد.

۲- ۲- ۲ محصولات

محصولات اصلی این شرکت برنامه‌های چت‌بات هستند که به کاربران اجازه می دهد با هوش مصنوعی توسعه یافته این شرکت تعامل داشته باشند. برای انجام این کار، هاگینگ‌فیس مدل پردازش زبان طبیعی خود[۷] به نام یادگیری چند وظیفه‌ای سلسله مراتبی[۸] را توسعه داد و کتابخانه‌ای از مدل‌های ان‌ال‌پی از پیش آموزش‌دیده‌شده را تحت پایتورچ-ترنسفورمر[۹] مدیریت کرد. برنامه‌های چت بات از سپتامبر ۲۰۱۹، فقط در آی‌اواس[۱۰] موجود است. این برنامه ها چتی[۱۱]، تاکینگ‌داگ[۱۲]، تاکینگ‌اگ[۱۳] و بالاس[۱۴] هستند.

۲- ۲- ۳ خدمات و تکنولوژی‌ها

در این بخش به بیان دو خدمت و تکنولوژی اساسی هاگینگ‌فیس می‌پردازیم تا بتوانیم در ادامه با شناخت بیشتری از آنها بهره کافی را ببریم:

---

[۶] BigScience
[۷] NLP
[۸] HMTL
[۹] PyTorch-Transformers
[۱۰] iOS
[۱۱] Chatty
[۱۲] Talking Dog
[۱۳] Talking Egg
[۱۴] Boloss

۲- ۲- ۳- ۱ کتابخانه ترنسفورمر[۱۵]

بسته ای است که شامل پیاده سازی‌های منبع‌باز مدل‌های ترنسفورمر برای کارهای متنی، تصویری و صوتی است. این با کتابخانه‌های یادگیری عمیق پایتورچ و تنسرفلو سازگار است و شامل پیاده‌سازی مدل‌های قابل‌توجهی مانند برت[۱۶] و جی‌پی‌تی[۱۷] است.

۲- ۲- ۳ – ۲ هاگینگ‌فیس هاب[۱۸]

پلتفرمی است که در آن کاربران می‌توانند مجموعه داده‌های از پیش‌آموزش‌دیده، مدل‌ها و دموهای پروژه-های یادگیری‌ماشین را به اشتراک بگذارند. هاب دارای ویژگی های الهام گرفته شده از گیت‌هاب برای به اشتراک‌گذاری کد و همکاری است، از جمله بحث و گفتگو و درخواست برای پروژه‌ها. همچنین میزبان هاگینگ‌فیس‌اسپیس است که سرویسی است که به کاربران اجازه می‌دهد تا دموهای مبتنی بر وب برنامه های یادگیری ماشین را با استفاده از گریدیو[۱۹] یا استریملیت[۲۰] بسازند. [۱۰]

۲- ۳ یادآوری و یادگیری بعضی مفاهیم پرکاربرد در یادگیری ماشین و بطور خاص پردازش گفتار

همانطور که می‌دانیم یادگیری ماشین یک تکنیک تجزیه و تحلیل داده است که به رایانه‌ها می آموزد تا کارهایی را انجام دهند که به طور طبیعی برای انسان ها و حیوانات اتفاق می افتد:«از تجربه بیاموزند.»
الگوریتم‌های یادگیری ماشین از روش‌های محاسباتی برای یادگیری مستقیم از داده‌ها بدون تکیه بر یک معادله از پیش تعیین‌شده استفاده می‌کنند که به عنوان مدل از آنها یاد می‌شود.

۲ – ۳- ۱ انواع روش‌های یادگیری ماشین

در اینجا لازم می‌دانم روش‌های یادگیری ماشین را باهم مرور کنیم:

۲- ۳- ۱- ۱ یادگیری نظارت‌شده

یادگیری نظارت‌شده مقوله‌ای است که در آن مدل یادگیری ماشینی را با داده‌های برچسب‌گذاری شده تغذیه می‌کنیم. مقادیرورودی و خروجی از قبل شناخته شده است و الگوریتم یادگیری ماشین تابع نگاشت

[۱۵] Transformer Library
[۱۶] Bert
[۱۷] GPT
[۱۸] Hugging Face Hub
[۱۹] Gradio
[۲۰] Streamlit

را یاد می‌گیرد. از نظر ریاضی، اگر Y خروجی و X ورودی باشد، الگوریتم‌های یادگیری ماشین سعی می‌کنند بهترین تابع نگاشت f را پیدا کنند به طوری که:

$$Y = f(X) \quad (۱)$$

در این روش، یادگیری به گونه‌ای اتفاق می‌افتد که یک سرپرست بر روند یادگیری نظارت می‌کند. ما از قبل پاسخ‌ها را می‌دانیم. از این رو الگوریتم ها سعی می کنند تابع را به گونه ای ترسیم کنند که پاسخ های پیش بینی شده، نزدیک به پاسخ های واقعی باشد. فرض کنید ماشین تابع نگاشت f را یاد گرفته است که مقادیر Y' را برای هر X پیش‌بینی می‌کند. زمانی یادگیری متوقف می‌شود که اختلاف بین (Y') پیش‌بینی‌شده و (Y) از یک مقدار حد آستانه مشخص پایین‌تر برود. یادگیری تحت نظارت را می توان بیشتر دسته بندی کرد:

- **طبقه‌بندی**[۲۱]:

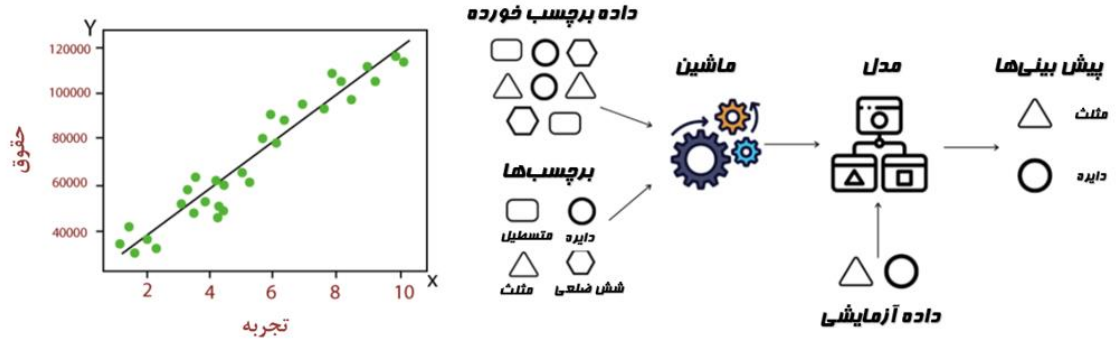
با در نظر گرفتن مثال تصویر ۲- ۱- الف، ورودی به مدل یادگیری ماشین، تصاویر اشکال هستند و خروجی، برچسب گذاری آن تصاویر به عنوان نام شکل است. بر اساس این داده های ورودی و خروجی، مدل یاد می گیرد که نوع دسته داده های تصویر دیده نشده را پیش بینی کند، خواه مستطیل، دایره، مثلث یا شش ضلعی باشد.

- **رگرسیون**[۲۲]:

با مثالی از تصویر ۲- ۱- ب، تجربه در محور X وجود دارد. برای هر تجربه، یک حقوق در محور Y وجود دارد. نقاط سبز مختصات (X, Y) در قالب داده های ورودی و خروجی هستند. مسئله رگرسیون سعی می کند تابع نگاشت پیوسته را از متغیرهای ورودی به خروجی پیدا کند.

برخی از الگوریتم های پرکاربرد در یادگیری تحت نظارت عبارتند از:

- رگرسیون خطی و لجستیک
- ماشین‌های بردار پشتیبان[۲۳]
- جنگل تصادفی



شکل ۲- ۱یادگیری نظارتی: الف) تصویر سمت راست: روش طبقه‌بندی و ب) تصویر سمت چپ: در روش رگرسیون اگر شیب تابع نگاشت یک و در نتیجه تابع نگاشت یک تابع خطی باشد، مدل، خط سیاه نشان داده شده در تصویر را یاد می گیرد.

[۲۱] classification

[۲۲] regression

[۲۳] SVM

۲- ۳- ۱- ۲ یادگیری بدون نظارت

یادگیری بدون نظارت مقوله ای از یادگیری ماشینی است که در آن مقدار داده های ورودی را می دانیم اما خروجی و عملکرد، هر دو ناشناخته هستند. در چنین سناریوهایی، الگوریتم‌های یادگیری ماشین تابعی را پیدا می‌کنند که شباهت را بین نمونه‌های داده ورودی مختلف پیدا می‌کند و آنها را بر اساس شاخص شباهت، که خروجی یادگیری بدون نظارت است، گروه‌بندی می‌کند. در چنین یادگیری، از آنجا که هیچ نظارتی وجود ندارد و داده های خروجی وجود ندارد، به آنها یادگیری بدون نظارت می‌گویند.

یادگیری بدون نظارت را می توان به صورت زیر دسته بندی کرد:

- **خوشه بندی (یا همان طبقه بندی بدون نظارت):**

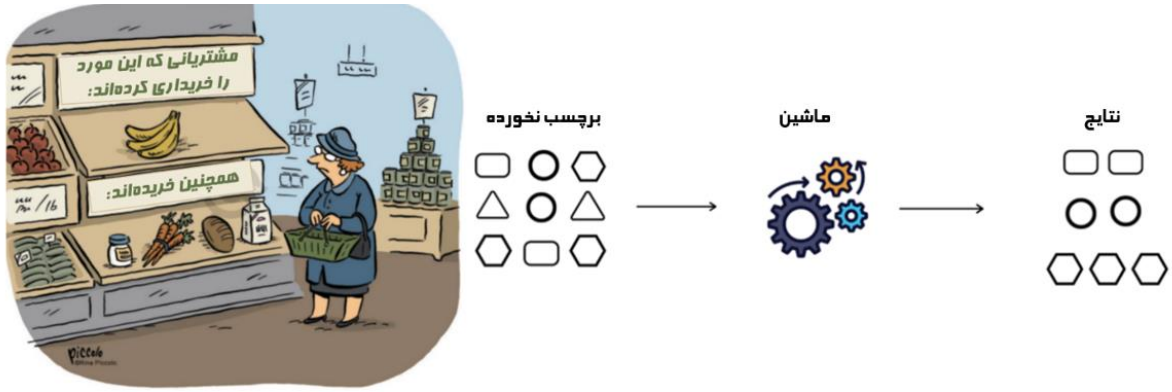
با مثالی از تصویر ۲- ۲- الف، داده های ورودی متشکل از تصاویر با اشکال مختلف داریم. الگوریتم‌های یادگیری ماشین سعی می‌کنند شباهت بین تصاویر مختلف را بر اساس مقدار پیکسل، اندازه و شکل تصاویر را پیدا کنند و گروه‌هایی را به عنوان خروجی‌هایی تشکیل دهند که نمونه‌های ورودی مشابه در آنها قرار دارند.

الگوریتم های خوشه بندی عبارتند از:

- خوشه بندی سلسله مراتبی
- خوشه‌بندی کی-میانگین[۲۴]
- الگوریتم‌های کاهش ابعاد مانند آنالیز مولفه اصلی[۲۵] و تجزیه بردار منفرد[۲۶].

- **پیوستگی:**

یادگیری در مورد کشف قوانینی است که بخش بزرگی از داده ها را توصیف می کند. اگر بخواهیم این مورد را با یک مثال شرح دهیم، در تصویر ۲- ۲- ب، مشتریانی که یک موز خریدند هویج نیز خریدند، یا مشتریانی که خانه جدید خریداری کردند نیز مبلمان جدید خریداری کردند.



شکل ۲- ۲یادگیری بدون نظارت: الف) تصویر سمت راست و ب) تصویر سمت چپ

[۲۴] K-means clustering
[۲۵] PCA
[۲۶] SVD

۲- ۳- ۱- ۳ یادگیری نیمه‌نظارتی

در یادگیری نیمه‌نظارتی، داده‌های ورودی داریم و فقط برخی از آن داده‌های ورودی به عنوان خروجی برچسب گذاری می شوند. درو اقع می‌توان گفت یادگیری نیمه‌نظارتی تا حدی تحت نظارت و تا حدی بدون نظارت است.

امروزه بسیاری از شرکت‌های بزرگ میلیون‌ها گیگابایت داده را جمع‌آوری کرده‌اند و هنوز در حال جمع‌آوری هستند. اما برچسب گذاری داده‌های جمع آوری شده به نیروی کار و منابع نیاز دارد و از این رو بسیار گران است.

برخی از موارد استفاده معروف از یادگیری نیمه‌نظارت عبارتند از:

- یادگیری نظارت‌شده بر روی داده های بدون برچسب و استفاده از خروجی پیش بینی شده به عنوان ورودی برای آموزش مجدد سایر مدل های یادگیری نظارت شده و آزمایش آن بر روی سایر داده های بدون برچسب.

به عنوان مثال، فرض کنید یک تکه بزرگ از داده ها در تصویر بالا وجود دارد، و مقدار کمی از مجموعه داده برچسب گذاری شده وجود دارد. ما می‌توانیم مدل را با استفاده از آن مقدار کمی از داده‌های برچسب‌گذاری‌شده آموزش دهیم و سپس مجموعه‌داده بدون برچسب را پیش‌بینی کنیم. پیش‌بینی بر روی یک مجموعه داده بدون برچسب، با دقت کمی برچسب را به هر داده متصل می‌کند که به عنوان مجموعه داده‌های شبه برچسب‌گذاری‌شده[۲۷] نامیده می‌شود. اکنون می توان یک مدل جدید با ترکیب مجموعه داده با برچسب واقعی و مجموعه داده با برچسب شبه آموزش داد.

- یادگیری بدون نظارت برای یادگیری ساختار موجود در داده‌ها.

۲- ۳- ۱- ۴ یادگیری تقویتی

در این روش، الگوریتم‌های یادگیری ماشین به‌عنوان عامل در محیطی عمل می‌کنند که این عوامل، اقدامات احتمالی را انتخاب می‌کنند. عامل بهترین اقدام را از بین تمامی گزینه های موجود در آن محیط انتخاب می‌کند و بر اساس آن انتخاب، پاداش یا ضرر دریافت می‌کند. الگوریتم‌ها به حداکثر رساندن پاداش و کاهش ضرر توجه می‌کنند تا اینکه در نهایت یاد بگیرند. الگوریتم مورد استفاده در یادگیری تقویتی، یادگیری-Q[۲۸] می‌باشد.

اگر به تاریخچه یادگیری ماشین نگاهی بیندازیم، متوجه می‌شویم که RL بسیار قدیمی است و برای مدت طولانی در صنعت است. اما به دلیل نیاز به آگاهی کامل از محیط، معمولا در محیط های شبیه سازی شده استفاده می شود. برخی از رایج ترین موارد استفاده در صنعت عبارتند از:

- عاملی که می تواند وسیله نقلیه را در داخل محیط شبیه سازی شده هدایت کند.
- پیش بینی قیمت سهام در بازار سهام

[۲۷] Pseudo-labeled dataset

[۲۸] Q-Learning

- عامل در محیط‌های بازی

۲- ۳- ۱- ۵ یادگیری خود نظارتی[۲۹]

در این روش، مدل، خود را آموزش می دهد تا بخشی از ورودی را از قسمت دیگری از ورودی یاد بگیرد. همچنین به عنوان یادگیری پیشگویی شناخته می شود. در اینجا، مسئله بدون نظارت با تولید خودکار برچسب‌ها به یک مشکل نظارت‌شده تبدیل می شود و هدف، شناسایی بخش‌های پنهان ورودی از بخش‌های غیر پنهان ورودی است.

به عنوان مثال، در پردازش زبان طبیعی، اگر چند کلمه داشته باشیم، با استفاده از یادگیری خود نظارتی می‌توانیم بقیه جمله را کامل کنیم. به طور مشابه، در یک ویدیو، می‌توانیم فریم‌های گذشته یا آینده را بر اساس داده‌های ویدیویی موجود پیش‌بینی کنیم. یادگیری خود نظارتی از ساختار داده‌ها برای استفاده از انواع سیگنال‌های نظارتی در مجموعه داده‌های بزرگ استفاده می‌کند (همه بدون تکیه بر برچسب‌ها).

از آنجا که مدل ما (مدل wav2vec) بر اساس همین مدل کار می‌کند، این روش یادگیری را بیشتر بسط داده تا با دانش بیشتری نسبت به آن به سمت پروژه قدم برداریم:

۲- ۳- ۱- ۵- ۱ یادگیری خود نظارتی در پردازش گفتار

قابلیت‌های یادگیری این مدل‌ها در سال ۲۰۱۳ و زمان انتشار مقاله Word2vec که دنیای پردازش زبان طبیعی را متحول کرد، تکامل یافته است. رویکردهای تعبیه کلمه[۳۰] ساده بود: به جای درخواست مدلی برای پیش‌بینی کلمه بعدی، می‌توانیم از آن بخواهیم کلمه بعدی را بر اساس محتوای قبلی پیش‌بینی کند. به دلیل چنین پیشرفت‌هایی، ما توانستیم نمایش معنی‌داری را از طریق توزیع تعبیه کلمه به دست آوریم که می‌تواند در بسیاری از سناریوها مانند تکمیل جمله، پیش‌بینی کلمات و ... استفاده شود.

در دهه گذشته، جریان تحقیقات و توسعه شگفت انگیزی در زمینه پردازش زبان طبیعی وجود داشته است. اجازه دهید برخی از موارد مهم را در زیر به طور خلاصه بیان کنیم:

- **پیش بینی جمله بعدی**[۳۱]

دو جمله همزمان از یک سند و یک جمله تصادفی از یک سند (خواه همان سند یا یک سند متفاوت) انتخاب می کنیم، جمله «آ»، جمله «ب» و جمله «س». سپس موقعیت نسبی جمله A را نسبت به جمله B را از مدل می‌پرسیم که خروجی مدل «کنار_جمله_هست» یا «کنار_جمله_نیست» خواهد بود. ما این کار را برای همه ترکیب ها انجام می دهیم.

سناریوی زیر را در نظر بگیرید:

۱. زهرا پس از اتمام ساعات مدرسه به خانه رفت.

۲. پس از تقریبا ۵۰ سال، سرانجام ماموریت فضاپیمای سرنشین‌دار به ماه در حال انجام است.

[۲۹] Self-Supervised-Learning (SSL)

[۳۰] Word embedding

[۳۱] Next Sentence Prediction

۳. زمانی که زهرا به خانه رفت، صدا و سیما تماشا کرد تا استراحت کند:)

هدف اصلی مدل در اینجا پیش‌بینی جملات بر اساس وابستگی‌های محتوایی بلندمدت است. اگر از شخصی بخواهیم هر دو جمله را که با درک منطقی ما مطابقت دارد ترتیب دهد، به احتمال زیاد جمله ۱ و سپس جمله ۳ را انتخاب می کند.

- **مدل‌سازی زبان خود-رگرسیون**[۳۲]:

در حالی که مدل‌های رمزگذاری خودکار مانند برت[۳۳] از یادگیری خود نظارتی برای کارهایی مانند طبقه‌بندی جملات استفاده می‌کنند، رویکردهای خود نظارتی در حوزه تولید متن کاربرد دارند.

مدل‌های خودرگرسیون مانند جی‌پی‌تی [۳۴](ترانسفورمر تولیدگر از پیش‌آموزش‌دیده) برای کار مدل‌سازی زبان کلاسیک از پیش‌آموزش دیده‌اند (با خواندن تمام کلمات قبلی، کلمه بعدی را پیش‌بینی می‌کند). چنین مدل‌هایی با قسمت رمزگشای ترانسفورمر مطابقت دارند و از یک پوشاننده[۳۵] در بالای جمله کامل استفاده می‌شود، به طوری که سر پوینتر فقط می‌تواند تا قبل آن را بخوانند، و نه آنچه را که بعد از آن است. فریم‌ورک جی‌پی‌تی شامل دو مرحله پیش‌آموزشی بدون نظارت و تنظیم‌دقیق[۳۶] نظارت‌شده می‌باشد که اطلاعات بیشتر در منبع [۳] قرار دارد و برای گزارش ما نیازی به آوردن مطالب نبود.

۲- ۳- ۱- ۵- ۲ چالش‌های یادگیری خود نظارتی

یادگیری خود نظارتی چگونه تقریباً در همه حوزه‌های جامعه یادگیری ماشین کاربرد دارد، اما اشکالاتی نیز دارد. یادگیری خود نظارتی در تلاش برای دستیابی به رویکرد «یک روش همه را حل می‌کند» است اما با تحقق این موضوع فاصله زیادی دارد. برخی از چالش های کلیدی عبارتند از:

- **دقت:**

اگرچه پیش‌فرض تکنیک یادگیری خود نظارتی استفاده نکردن از داده‌های برچسب‌گذاری‌شده است، اما نقطه ضعف این رویکرد این است که شما یا به مقادیر زیادی داده برای تولید شبه‌برچسب (یا سودو برچسب) دقیق نیاز دارید یا در مورد دقت به خطر می‌افتید. توجه به این نکته مهم است که برچسب های نادرست تولید شده در حین آموزش در مراحل اولیه نتیجه معکوس خواهند داشت.

- **کارایی محاسباتی:**

به دلیل مراحل متعدد آموزش (ابتدا تولید شبه‌برچسب‌ها و سپس آموزش بر روی آنها) زمان صرف شده برای آموزش یک مدل در مقایسه با یادگیری تحت نظارت زیاد است. همچنین، رویکردهای فعلی یادگیری خود نظارتی به حجم عظیمی از داده‌ها برای دستیابی به دقت نزدیک به همتای خودش در یادگیری تحت نظارت نیاز دارند.

- **تسک پیشگو**[۳۷]:

---

[۳۲] Auto-regression language modeling

[۳۳] BERT: Bidirectional Encoder Representation from Transformers

[۳۴] GPT: Generative Pre-trained Transformer

[۳۵] Mask

[۳۶] Fine-tune

انتخاب تسک پیشگو مورد استفاده شما بسیار مهم است. به عنوان مثال، اگر یک رمزگذار خودکار را به عنوان تسک پیشگو خود انتخاب کنید که در آن تصویر فشرده شده و سپس بازسازی می شود، همچنین سعی می‌کند نویز تصویر اصلی را تقلید کند و اگر وظیفه شما تولید تصاویر با کیفیت بالا باشد، این تسک پیشگو بیشتر ضرر می زد تا اینکه مفید باشد. [۲]

۲- ۳- ۲ برخی اصطلاحات پرکاربرد در پردازش گفتار

در این قسمت به شرح مختصری از برخی مفاهیمی که در پردازش زبان و گفتار پرتکرار هستند و علاوه برخواندن مقاله در بخش عملی هم به چشم می‌خورند می‌پردازیم:

۲- ۳- ۲- ۱ تنظیم دقیق مدل[۳۸]

تنظیم دقیق راهی برای اعمال یادگیری انتقالی[۳۹] است. درواقع، تنظیم دقیق فرآیندی است که از مدلی استفاده می‌کند که قبلاً برای یک کار خاص آموزش داده شده است و سپس مدل را تغییر می دهد تا برای یک کار مشابه دیگر قابل استفاده بشود. این چیزی است که رویکرد تنظیم دقیق را بسیار جذاب می کند. اگر بتوانیم یک مدل آموزش‌دیده پیدا کنیم که قبلاً یک کار را به خوبی انجام می‌داده، و آن کار حداقل از بیرون شبیه به کار ما باشد، می‌توانیم از همه چیزهایی که مدل قبلاً یاد گرفته است استفاده کنیم و آن را برای کار خاص خود به کار ببریم. البته، باید به این نکته توجه داشت که اطلاعاتی وجود دارند که مدل یاد گرفته است که ممکن است نیازمان نباشد یا حتی برعکس برای کار ما صدق نکند، یا ممکن است اطلاعات جدیدی وجود داشته باشد که مدل باید از داده‌های مربوط به همین کار یاد بگیرد.

برای مثال، مدلی که روی خودروها آموزش دیده‌است، هرگز پشت کامیون را ندیده است، بنابراین این ویژگی چیز جدیدی است که مدل باید درباره آن بیاموزد. با این حال، مدل ما برای تشخیص کامیون‌ها می‌تواند از مدلی که در ابتدا روی خودروها آموزش داده شده بود استفاده کند که این بسیار برای ما پسندیده است.

حال ببینیم که چگونه باید یک مدل را تنظیم دقیق کرد؟ برگردیم به مثالی که ذکر کردیم، پس مدلی داریم که پیش از این برای تشخیص خودروها آموزش دیده است و می‌خواهیم این مدل را برای تشخیص کامیون‌ها تنظیم کنیم. برای سادگی، فرض کنید آخرین لایه این مدل را برداریم. آخرین لایه قبلاً وظیفه طبقه بندی تصاویر را با عنوان «ماشین/ غیر ماشین» می کرد. پس از حذف آن، می خواهیم یک لایه جدید اضافه کنیم که هدف آن طبقه بندی این باشد که آیا یک تصویر کامیون است یا خیر. در برخی از مسائل، ممکن است بخواهیم بیش از آخرین لایه را حذف کنیم، و ممکن است بخواهیم بیش از یک لایه اضافه کنیم. این بستگی به شباهت کار برای هر یک از مدل ها دارد. لایه‌های انتهای مدل ما ممکن است ویژگی‌های آموخته‌ای داشته باشند که برای کار اصلی بسیار خاص هستند، در حالی که لایه‌ها در ابتدای مدل معمولاً ویژگی‌های عمومی‌تری مانند لبه‌ها، شکل‌ها و بافت‌ها را یاد می‌گیرند. بعد از اینکه ساختار مدل

---

[۳۷] Pretext task

[۳۸] Fine-tuning model

[۳۹] Transfer learning

موجود را اصلاح کردیم، می‌خواهیم لایه‌ها را در مدل جدیدمان فریز[۴۰] یا منجمد کنیم. پس از انجام این کار، فقط آموزش مدل بر روی داده های جدیدمان باقی می ماند. [۴]

۲- ۳- ۲- ۲ مدل زبانی[۴۱]

مدل زبانی استفاده از تکنیک‌های مختلف آماری و احتمالی برای تعیین احتمال وجود یک توالی معین از کلمات در یک جمله است. مدل‌های زبان بدنه داده‌های متنی را تجزیه و تحلیل می‌کنند تا مبنایی برای پیش‌بینی‌های کلمه‌شان فراهم کنند. سپس، مدل این قوانین را در داده‌های زبانی به کار می‌برد تا به طور دقیق جملات جدید را پیش‌بینی یا تولید کند. این مدل اساساً ویژگی‌های زبان پایه را می‌آموزد و از آن ویژگی‌ها برای درک عبارات جدید استفاده می‌کند. در اصل دو نوع مدل زبان وجود دارد:

- مدل‌های زبانی آماری[۴۲]:

مدل های آماری شامل توسعه مدل های احتمالی است که قادر به پیش بینی کلمه بعدی در دنباله با توجه به کلمات قبل از آن هستند مانند ان-گرم[۴۳]، یونی‌گرم[۴۴]، دوجهته[۴۵]، نمایی[۴۶] و فضای پیوسته.

- مدل‌های زبانی عصبی[۴۷]:

این مدل های زبانی مبتنی بر شبکه های عصبی هستند و بر کاستی‌های مدل‌های کلاسیک مانند ان-گرم غلبه می‌کنند و برای کارهای پیچیده‌ای مانند تشخیص گفتار یا ترجمه ماشینی استفاده می‌شوند. [۵]

۲- ۲- ۲- ۳ وظایف پایین دستی[۴۸]

کارهای پایین دستی در زمینه یادگیری خود نظارتی، مسئله‌ایست که در واقع می خواهیم آن را حل کنیم. به طور خاص، در یادگیری انتقالی، ابتدا یک مدل را با مجموعه داده‌های عمومی از پیش‌آموزش می‌دهیم[۴۹]، که این مسئله اصلی نیست، اما به مدل اجازه می‌دهد تا برخی از ویژگی‌های عمومی را بیاموزد. سپس این مدل از پیش‌آموزش داده شده را روی مجموعه داده خودمان تنظیم می کنیم که نشان دهنده کار اصلی است. این کار اخیر چیزی است که در زمینه یادگیری خود نظارتی، وظیفه پایین دستی نامیده می‌شود. [۶]

۲- ۲- ۲- ۴ نسبت سیگنال به نویز[۵۰]

اس‌ان‌آر نسبت بین اطلاعات مورد نظر یا قدرت یک سیگنال با سیگنال نامطلوب یا نویز پس‌زمینه است که واحد بیان آن دسی‌بل است. برای تعیین اس‌ان‌آر، باید با کم کردن مقدار نویز از مقدار قدرت سیگنال، تفاوت بین قدرت

---

[۴۰] freeze
[۴۱] Language model
[۴۲] Statistical Language Models
[۴۳] N-Gram
[۴۴] Unigram
[۴۵] Bidirectional
[۴۶] Exponential
[۴۷] Neural Language Models
[۴۸] downstream task
[۴۹] Pre-train
[۵۰] SNR

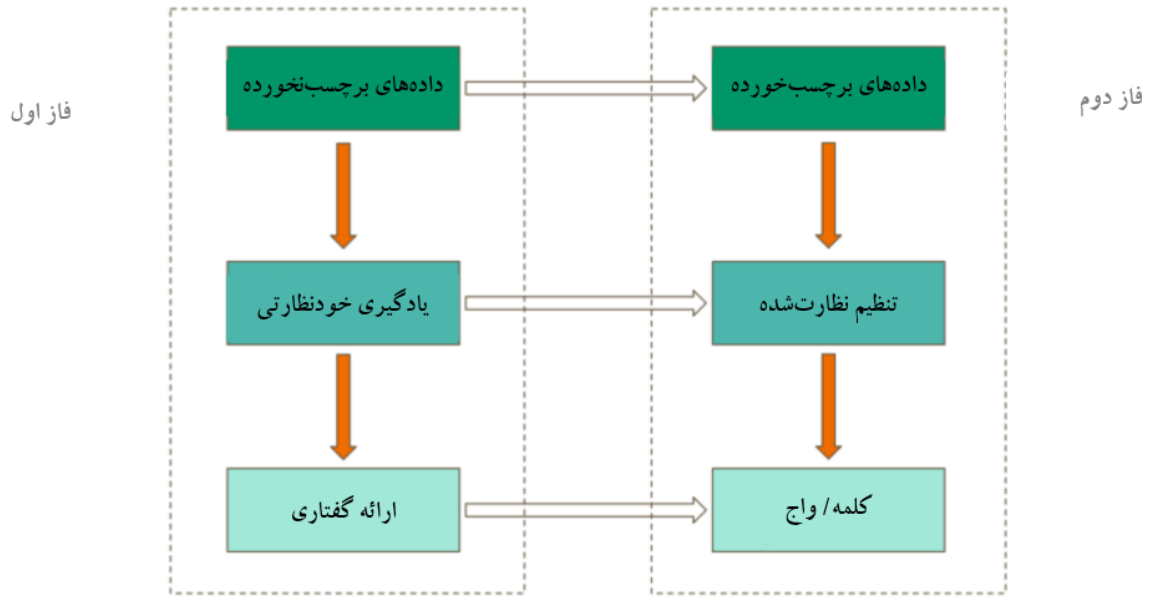
سیگنال مورد نظر و نویز ناخواسته را پیدا کرد. همچنین از نظر اتصال در شبکه‌های بی‌سیم، کارشناسان نیاز به اس-ان‌آر برابر حداقل ۲۰ دسی‌بل برای جستجو در وب دارند. در نتیجه به چهار دسته از مقادیر اس‌ان‌آر دست پیدا می-کنیم:

- ۵ الی ۱۰ دسی‌بل: سطح نویز تقریباً از سیگنال مورد نظر قابل تشخیص نیست.
- ۱۰ الی ۱۵ دسی‌بل: حداقل مورد قبول برای ایجاد یک اتصال غیرقابل اطمینان است.
- ۱۵ الی ۲۵ دسی‌بل: یک اتصال قابل قبول را فراهم می‌کند.
- ۲۵ الی ۴۰ دسی‌بل: خوب تلقی می‌شود.
- ۴۰ به بالا: سیگنال بسیار تمیز است.

۲- ۴ مدل wav2vec 2.0

Wav2vec 2.0 یکی از پیشرفته ترین مدل های فعلی برای تشخیص خودکار گفتار با یادگیری خود نظارتی است. این روش به ما امکان را می‌دهد تا یک مدل را روی داده‌های بدون برچسب که معمولا در دسترس‌تر است، آموزش دهیم. سپس، مدل را می‌توان بر روی یک مجموعه داده خاص برای یک هدف خاص تنظیم دقیق[۵۱] کرد.

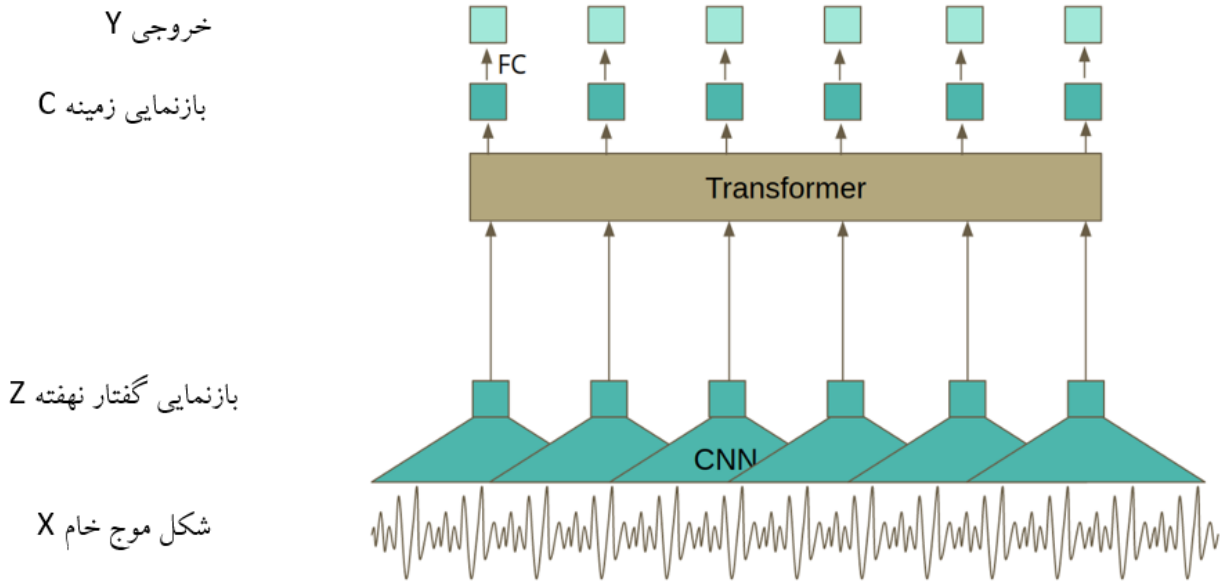
همانطور که در تصویر ۲- ۳ ارائه شده است، مدل در دو فاز آموزش داده می‌شود. مرحله اول در حالت خود نظارتی است که با استفاده از داده‌های بدون برچسب انجام می‌شود و هدف آن دستیابی به بهترین نمایش گفتار ممکن است. هدف تعبیه کلمه نیز دستیابی به بهترین نمایش زبان طبیعی است، با این تفاوت اصلی که wav2vec2.0 صدا را به جای متن پردازش می‌کند. مرحله دوم، تنظیم دقیق کردن نظارت شده است که در طی آن از داده‌های برچسب‌دار برای آموزش مدل برای پیش‌بینی کلمات یا واج‌های خاص استفاده می-شود.



شکل ۲- ۳ مراحل یادگیری wav2vec2.0

[۵۱] Fine-tune

در واقع در مدل wav2vec2 کاری که می‌خواهیم انجام دهیم یادگیری صرفا از طریق شنیدن است (مثل کاری که بچه انسان انجام می‌دهد اول یاد می‌گیرد بشنود و حرف بزند و بعد می‌تواند بخواند و بنویسد) ولی اگه بخواهیم یک ترنسفورمر بر روی یک سیگنال صوتی قرار دهیم هزینه‌اش برابر با توان دو طول آن سیگنال می‌شود، پس اول با گذاشتن چندین کانولوشن یک ارائه ساده‌تر بدست می‌آوریم در شکل ۲- ۴ با کانولوشن از X به Z می‌رویم. حال مقدار بالای هرکدام از این بردارهای حاصل رو می‌گیریم و وارد یک ترنسفورمر می‌کنیم که خروجی آن نیز بردار هایی است که با یک نگاشت خطی خروجی نهایی را می‌سازند. ایده اصلی پیش‌آموزش، مشابه برت است: بخشی از ورودی ترنسفورمر پوشانده شده‌است و هدف حدس زدن نمایش بردار ویژگی پنهان Z است. [۸]



شکل ۲- ۴ معماری مدل فاین‌تیون شده‌ی wav2vec 2.0

فصل سوم

کارهای انجام‌شده در دوران کارآموزی

۳- کارهای انجام‌شده در دوران کارآموزی

در این بخش، قصد داریم به مرور فازهایی که در طول دوران کارآموزی طی کردم، بپردازیم.

۳- ۱ خواندن مقاله مرتبط و ارائه آن

در ابتدای شروع دوره کارآموزی، هرکدام از کارآموزان می‌بایست پس از انتخاب پروژه‌ای که می‌خواستند روی آن کار کنند، مقاله‌ای را انتخاب کرده و پس از یک الی دو هفته ارائه‌ای از آن داشته‌باشند. مقاله [۱] با توجه به مقیاس‌های تعداد ارجاعات[۱]، سال انتشار مقاله، شاخص اچ[۲] نویسندگان و مدل‌های استفاده شده انتخاب شد. لینک گوگل‌شیت انتخاب مقاله در اینجا است. شروع به خواندن مقاله کردم و از آنجا که اصطلاحات زیادی را نفهمیدم بعد از یک دور روزنامه‌وار خواندن، یک دور تخصصی خواندم و هرجا که به کلمه‌ای جدید برمی‌خوردم با جستجو، یا با خواندن سایت یا دیدن ویدئو به مفهوم آن پی می‌بردم. بعد از این مرحله ساخت پاورپوینت برای ارائه را آغاز کردم که در اینجا می‌توانید اسلایدها را مشاهده نمایید. پس بعد از آمادگی جهت ارائه و ارائه پاورپوینت، یک ارائه نوشتاری نیز در قالب گزارش از مقاله و ارائه‌ای که داشتیم هم از ما خواسته شد که در فاز بعدی توضیح خواهم داد.

۳- ۲ تهیه گزارش از مقاله

لینک گزارش در اینجا قرار داده شده‌است. بعد از تحویل گزارش، آغاز فاز بعدی از کارآموزی یعنی تنظیم دقیق کردن مدل wav2vec2 روی دیتای فارسی آغاز شد.

۳- ۳ تنظیم دقیق مدل wav2vec2-large-XLSR روی پایگاه‌داده شمو[۳] فارسی

از آنجا که تاکید مقاله بر روی مدل wav2vec2 بود و من هم با تنظیم دقیق کردن مدل آشنایی نداشتم، این فاز صرفا آشنایی با این کار بود تا بتوانم در نهایت روی دیتای خلبانی هم یک سیستم تشخیص گفتار خودکار راه‌اندازی کنم. پس کار را با خواندن کد تنظیم دقیق کردن مدل Wav2vec2-XLS-R برای تشخیص گفتار خودکار چندزبانه در هاگینگ‌فیس شروع کردم که این سایت مرجع من برای این کار بود. کار را شروع کردم و کد ها را در کولب ران می کردم و سعی می‌کردم با تغییر جزئی کد را با دیتاست زبان فارسی تطبیق دهم (برای مثال در حذف یک‌سری حروف یا لود کردن زبان فارسی از کامن‌وویس)، اما با خطای «عدم

[۱] citations

[۲] h-index

[۳] ShEMO: Sharif Emotional Speech Database

دسترسی کولب به دیتاست» روبرو شدم که ادامه کار عملا غیرممکن بود. در اینجا با مشورت با منتورها، تصمیم بر این شد که کار را با استفاده از دیتابیس شمو موجود در کگل پیش ببریم. با اجرای کد به WER ۴۷ درصد رسیدیم که در اینجا لینک آن قرار داده شده‌است.

```
results = {}
metrics = trainer.evaluate()
max_val_samples = len(_common_voice_test)
metrics["eval_samples"] = min(max_val_samples, len(_common_voice_test))

trainer.log_metrics("eval", metrics)
trainer.save_metrics("eval", metrics)
```

```
***** Running Evaluation *****
  Num examples = 284
  Batch size = 10
[29/29 00:23]
reference: "آقای هارترایت به هر حال امیدوارم که از زندگی با ما لذت ببرید"
predicted: "آقای حارت رایت برحال امیدوارم که از زندگی با مالذت ببرین"
reference: "من دویست صفحه دیگه هم خوندم"
predicted: "من دیوش سفایه دیگم خوندهم"
reference: "سوءاستفاده از اینکه می دونه من چقدر آدم احساساتی ای هستم"
predicted: "سو استفاده از اینکه می دون من چقدر آدم احساساتی ای هستم"
***** eval metrics *****
  epoch                   =        0.5
  eval_loss               =     1.5622
  eval_runtime            = 0:00:25.10
  eval_samples            =        284
  eval_samples_per_second =     11.314
  eval_steps_per_second   =      1.155
  eval_wer                =     0.4717
```

شکل ۳- ۱ تصویر نرخ خطای کلمه مدل wav2vec2-large-XLSR روی دیتابیس شمو فارسی

```
[ ] print("Prediction:")
    print(processor.decode(pred_ids))

    print("\nReference:")
    print(common_voice_test_transcription["sentence"][0].lower())
```

```
Prediction:
اوبه والا به قول قدیمیها گفتنی مسجد چها ج نفوری بچه گدا فراوونی

Reference:
خوبه والا به قول قدیمی ها گفتنی مسجد شاه چراغونه بچه گدا فراوونه
```

شکل ۳- ۲ تصویر خروجی مدل wav2vec2-large-XLSR روی دیتابیس شمو فارسی

در تصویر ۳- ۱ evel_wer یا همان نرخ خطای کلمه ارزیابی شده که برابر با ۴۷ درصد است، مشاهده می‌شود. برای مثال در جمله اول «هارترایت»، «حارت رایت» و «ببرید»، «ببرین» پیش‌بینی شده‌است. درست است که برخی کلمات اشتباه رونویسی شده‌اند اما این درصد قابل قبول می‌باشد. همینطور در تصویر ۳- ۲ هم می‌بینیم که واژه «خوبه» در یک حرف، «شاه چراغونه» در تمام حروف و «فراوونه» در یک حرف دچار خطا شده‌اند.

۳- ۴ تنظیم دقیق کردن مدل wav2vec2-base روی مجموعه‌داده انگلیسی تیمیت[۴]

در ادامه برای نزدیک‌تر شدن به اصل پروژه که مربوط به مکالمات خلبانی انگلیسی بود، به تنظیم دقیق کردن مدل wav2vec2-base روی دیتاست تیمیت[۵] که به زبان انگلیسی است، پرداختم. همانطور که در تصویر ۳- ۳ مشخص است، در این جمله دارای دو خطا در «bungalow»، یک خطا در «pleasantly» و یک خطا در «shore» می‌باشد. کد این بخش هم در اینجا قرار داده‌شده است.

```
[87] print("Prediction:")
     print(processor.decode(pred_ids))

     print("\nReference:")
     print(timit_test_transcription["test"]["text"][0].lower())

     Prediction:
     the bunglo was plesntly situated near the shor

     Reference:
     the bungalow was pleasantly situated near the shore.
```

شکل ۳- ۳ تصویر خروجی مدل wav2vec2-base روی دیتاست تیمیت

۳- ۵ تنظیم دقیق کردن مدل wav2vec2-large-robust روی مجموعه‌داده ترافیک هوایی[۶]

این بخش شامل چندین فاز می‌باشد که در ادامه بحث می‌کنیم:

۳- ۵- ۱ بارگذاری و جدا کردن مجموعه‌داده به مجموعه‌داده‌های آموزشی و آزمایشی

در این فاز که فاز اصلی پروژه نیز می‌باشد، ابتدا به بارگذاری دیتاست هوایی پرداختم. که در این بخش می‌بایست فایل سی‌اس‌وی[۷] حاوی اطلاعات دیتا (شامل دایرکتوری، زیر شاخه، نام فایل، شناسه گوینده، شناسه جلسه، شناسه گفتار، رونویسی، مدت ضبط به ثانیه، شناسه ضبط و...) می‌باشد را به دو ستون حاوی مسیر فایل صوتی و رونویسی آن کاهش داد(شکل ۳- ۵).

بعد از بارگذاری دیتاست به شکل صحیح و حذف ستون‌های اضافی، به جدا کردن مجموعه داده آموزشی از مجموعه داده آزمایشی پرداختم. سپس به دلیل عدم استفاده از مدل زبانی به نرمال‌سازی[۸] داده با کوچک کردن تمام حروف و حذف برخی حروف خاص مثل «. ، ؟ ! : ؛» (این حروف معمولا کمکی به درک معنای گفتار نمی‌کنند) می‌پردازیم(شکل ۳- ۴) و سپس لغت‌نامه منحصربفرد از حروف را از رونویسی‌ها

[۴] Timit-asr dataset:

[۵] یک مجموعه گفتار پیوسته آکوستیک-آوایی استاندارد است که برای ارزیابی سیستم‌های تشخیص خودکار گفتار استفاده می‌شود.

[۶] ATCOSIM

[۷] csv

[۸] normalize

استخراج می‌کنیم و در آخر این مرحله با استفاده از کتابخانه Wav2vec2CTCTokenizer و همان لغت‌نامه به عنوان ورودی یک واحدساز[۹] تولید می‌کنیم.

	directory	subdirectory	filename	speaker_id	session_id	utterance_id	transcription	recording_corrupt	comment_transcriptionist	length_sec	recording_id	recording_startpos_sec
0	sm1	sm1_01	sm1_01_001	sm1	1	1	~p ~s ~a eight one zero turn right to trasadin...	0	NaN	3.299750	011_0001	133.504406
1	sm1	sm1_01	sm1_01_002	sm1	1	2	lufthansa five three one eight contact zurich ...	0	NaN	4.663094	011_0002	180.687594
2	sm1	sm1_01	sm1_01_003	sm1	1	3	~p ~s ~a eight one zero contact zurich one thr...	0	NaN	4.229750	011_0003	370.245000
3	sm1	sm1_01	sm1_01_004	sm1	1	4	sabena four eight one rhein identified	0	NaN	2.641969	011_0004	490.911656
4	sm1	sm1_01	sm1_01_005	sm1	1	5	transwede one zero one rhein identified set co...	0	NaN	3.911500	011_0005	511.825625
...	...	...	...	...	...	...	...	...	...	...	...	...
10073	sm2	sm2_09	sm2_09_204	sm2	9	204	nine six zero one squawk two seven six four	0	NaN	2.837000	111_0809	12885.187281
10074	sm2	sm2_09	sm2_09_205	sm2	9	205	lufthansa five five zero four yes i'll call yo...	0	NaN	6.694937	111_0810	12897.565094
10075	sm2	sm2_09	sm2_09_206	sm2	9	206	cross air five one eight is identified climb t...	0	NaN	5.286531	111_0811	12923.872625
10076	sm2	sm2_09	sm2_09_207	sm2	9	207	sata nine six zero one is identified <OT> oh i...	0	NaN	3.378281	111_0812	12934.923875
10077	sm2	sm2_09	sm2_09_208	sm2	9	208	<OT> thank you very much thank you very much [...	0	NaN	3.067031	111_0813	12940.636375

[10078 rows x 12 columns]

شکل ۳- ۴ تصویر ستون‌های فایل اطلاعات مجموعه داده به همراه محتوی آنها

	transcription
0	british midland seven two zero turn right by one five degrees
1	lufthansa triple five zero turn ah left heading two five zero
2	iberia three four two two right ten degrees
3	hapag lloyd one one two is identified
4	roger
5	merair six nine five two contact zurich one three four decimal six
6	lufthansa four five eight two climb to flight level three hundred
7	hamburg air two five four six contact marseille one two five eight five au revoir
8	transwede one zero seven call zurich on one three four six good day
9	portugalia five four three geneva one three three decimal one five bye bye
10	alitalia four zero one climb flight level two nine zero
11	lufthansa five two one seven turn to delta kilo bravo
12	lufthansa eight two two one contact ah milan one three four five two good bye
13	gulf air zero three two good morning radar contact maintain flight level three four zero proceed trasadingen zurich east fusse
14	netherlands air force four one four contact zurich on one three three decimal four
15	gulf air zero three two set course direct fusse further contact munich on one three four milan one three four five two bye bye
16	jet set four nine seven rhein radar identified
17	airfrans six seven zero good afternoon descend flight level three three zero on radar heading of one nine zero
18	aero lloyd confirm europa europa three six one bonjour identified cleared kines st prex willisau flight level three four zero
19	lufthansa four three five six what is your rate of climb ah if cleared higher

	transcription
0	delta mike echo fly heading one eight zero
1	~l ~t ~u seven seven five six geneva one three three one five good bye
2	~u ~s ~a ~u ~s air one four descend flight level three three zero
3	merair six nine five two climb initially to flight level three zero zero
4	[EMPTY]
5	lufthansa four six five two contact zurich on one three four six good day
6	alitalia three seven zero identified cleared aosta st prex dijon flight level three two zero
7	speedbird one two nine contact rhein radar one three two decimal four
8	alitalia four zero one zurich one three four decimal six
9	good morning swissair six five two zero radar contact continue climb flight level three two zero proceed direct trasadingen karlsruhe
10	sabena nine three six nine rhein radar identified
11	sabena seven eight one six radar contact direct karlsruhe
12	alitalia two nine two good afternoon squawk two seven six one
13	belgian airforce three three four due traffic continue present heading
14	roger delta mike echo maintain heading one seven one
15	air malta zero zero four turn right one zero degrees due traffic
16	merair six nine five two climb to flight level three three zero
17	fox oscar kilo sierra india good morning maintain two nine zero trasadingen saronno
18	lufthansa four seven zero zero zurich one three four six tschuss
19	lufthansa four six five two climb to flight level three four zero

شکل ۳- ۵ تصویر نمایش خروجی رندوم از بازنویسی: الف) قبل از نرمالایز کردن و ب) بعد از نرمالایز کردن

[۹] tokenizer

۳- ۵- ۲ ساخت تشخیص‌دهنده ویژگی wav2vec2

ابتدا لازم است یادآوری کنیم که گفتار یک سیگنال پیوسته است و برای پردازش توسط کامپیوتر، ابتدا باید گسسته شود که معمولاً به این کار نمونه‌برداری می‌گویند. نرخ نمونه‌گیری در اینجا نقش مهمی ایفا می‌کند، زیرا تعیین می‌کند که در هر ثانیه چند داده سیگنال گفتار اندازه‌گیری شود. بنابراین، نمونه‌برداری با نرخ نمونه‌برداری بالاتر منجر به تقریب بهتر سیگنال گفتار واقعی می‌شود، اما مقادیر بیشتری در هر ثانیه را نیاز دارد(و بالعکس با نرخ نمونه‌برداری پایین‌تر مقادیر کمتری در هر ثانیه اندازه‌گیری شده ولی به تبع در نهایت شهود خوبی از سیگنال را نمی‌دهد). و اینکه نرخ نمونه‌برداری داده‌هایی که برای پیش‌آموزش مدل استفاده شده‌است باید با نرخ نمونه‌برداری مجموعه داده مورد استفاده برای تنظیم دقیق کردن مدل مطابقت داشته باشد. نرخ نمونه‌برداری مدل wav2vec2، ۱۶ کیلوهرتز و نرخ نمونه‌برداری مجموعه‌داده‌ی ترافیک هوایی ما، ۳۲ کیلوهرترز است پس می‌بایست سیگنال گفتار مجموعه‌داده هوایی را روی ۱۶ کیلوهرتز نمونه‌برداری کاهشی[۱۰] کنیم. سپس با استفاده از کتابخانه‌های Wav2vec2FeatureExtractor و Wav2vec2Processor یک استخراج‌کننده ویژگی و یک پراسسور تولید می‌کنیم.

۳- ۵- ۳ پیش‌پردازش داده

تا کنون، ما به مقادیر واقعی سیگنال گفتار نگاه نکرده‌ایم، بلکه فقط رونویسی را بررسی کرده‌ایم. علاوه بر «متن»، مجموعه داده‌های ما شامل ستون «فایل» است اما ما به ستونی دیگر حاوی اطلاعات صوتی نیز نیازمندیم. پس در این مرحله با استفاده از کتابخانه librosa مجموعه‌داده را با مجموعه‌داده جدید حاوی این اطلاعات، بازسازی می‌کنیم. در آخر مجددا مجموعه‌داده را با دو ستون «مقادیر ورودی» و «برچسب‌ها» به وسیله پراسسور و نرخ نمونه‌برداری بازنویسی می‌کنیم.

۳- ۵- ۴ آموزش و ارزیابی

در این قسمت یک جمع‌کننده داده تعریف می‌کنیم. زیرا برخلاف اکثر مدل‌های پردازش زبان طبیعی، wav2vec2 طول ورودی بسیار بیشتری نسبت به طول خروجی دارد. سپس معیار ارزیابی در طول آموزش مدل باید بر اساس میزان خطای کلمه ارزیابی[۱۱] شود. پس از تنظیم دقیق مدل، آن را به درستی بر روی داده های آزمایش ارزیابی می کنیم و تأیید می کنیم که آیا مدل واقعاً یاد گرفته است که گفتار را به درستی رونویسی کند.

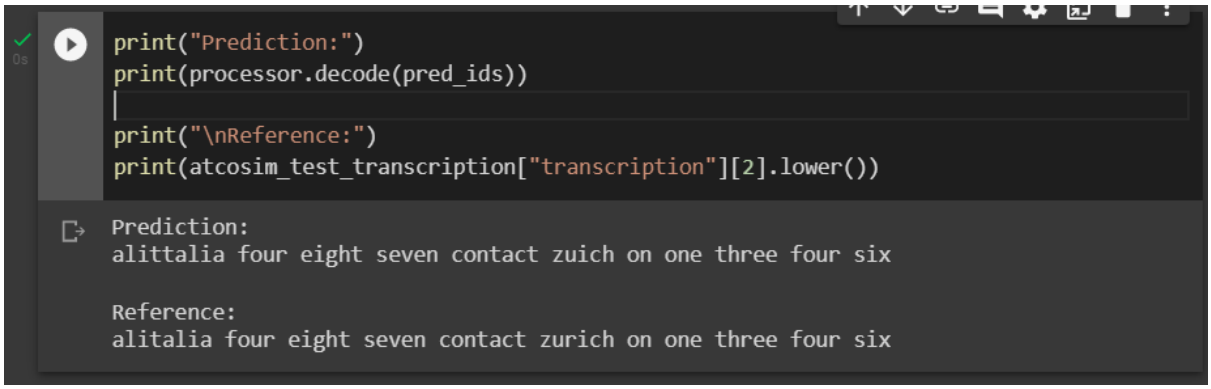
۳- ۵- ۵ نتایج

در این قسمت می‌خواهیم به تحلیل نتایج بپردازیم. همانطور که در تصویر ۳- ۶ می‌بینیم فقط در دو کلمه «Alitalia» و «Zurich» و فقط در یک حرف خطا وجود دارد. در شکل ۳- ۷ WER یا نرخ خطای کلمه‌ی

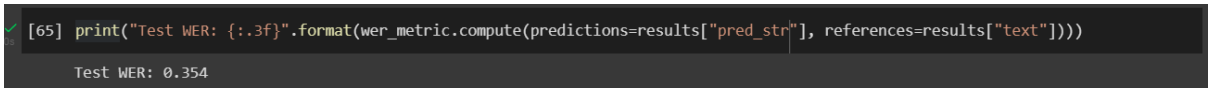
[۱۰] Down sample

[۱۱] WER: Word Error Rate

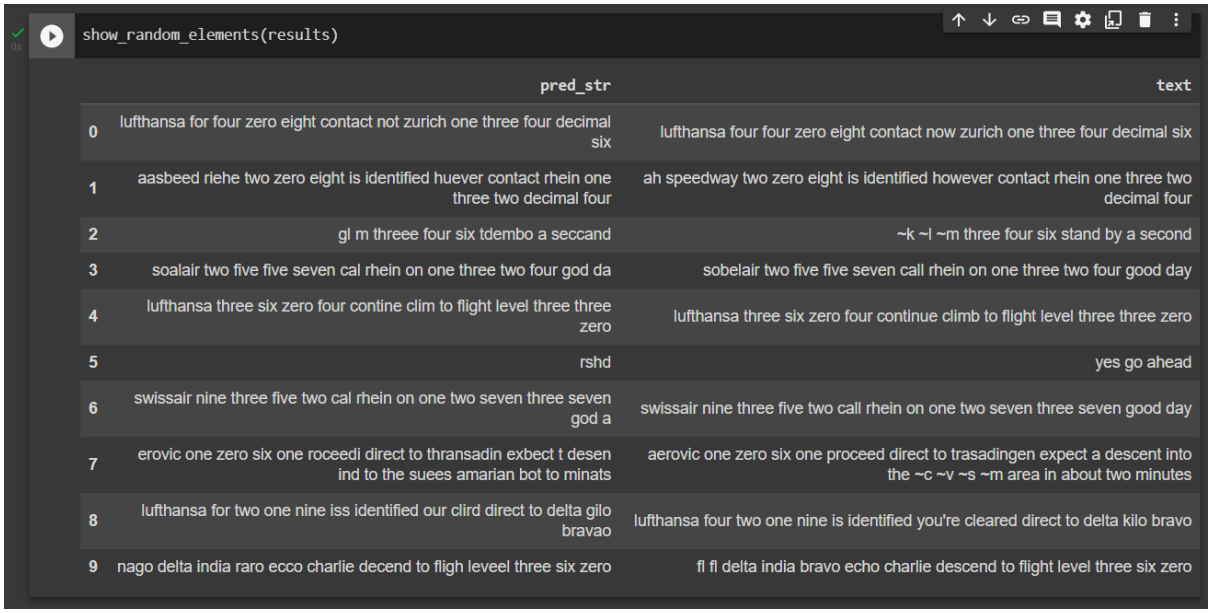
مدل wav2vec2-robust را مشاهده می‌کنیم که برابر ۳۵ درصد است و این WER برای مکالمات خلبانی که حاوی نویز و کلمات خاص و لهجه‌های متفاوت است قابل قبول می‌باشد. و در شکل ۳- ۸، ده نتیجه از رونویسی‌های پیش‌بینی شده را آورده‌ایم تا بتوانیم مقایسه خوبی از نتایج داشته باشیم. کد این بخش نیز در اینجا موجود است.



شکل ۳- ۶ تصویر یکی از داده‌های پیش‌بینی شده به همراه خود متن در دیتاست ترافیک هوایی



شکل ۳- ۷ نرخ خطای کلمه‌ی مدل wav2vec2-large-robust روی دیتاست ترافیک هوایی



	pred_str	text
0	lufthansa for four zero eight contact not zurich one three four decimal six	lufthansa four four zero eight contact now zurich one three four decimal six
1	aasbeed riehe two zero eight is identified huever contact rhein one three two decimal four	ah speedway two zero eight is identified however contact rhein one three two decimal four
2	gl m threee four six tdembo a seccand	~k ~l ~m three four six stand by a second
3	soalair two five five seven cal rhein on one three two four god da	sobelair two five five seven call rhein on one three two four good day
4	lufthansa three six zero four contine clim to flight level three three zero	lufthansa three six zero four continue climb to flight level three three zero
5	rshd	yes go ahead
6	swissair nine three five two cal rhein on one two seven three seven god a	swissair nine three five two call rhein on one two seven three seven good day
7	erovic one zero six one roceedi direct to thransadin exbect t desen ind to the suees amarian bot to minats	aerovic one zero six one proceed direct to trasadingen expect a descent into the ~c ~v ~s ~m area in about two minutes
8	lufthansa for two one nine iss identified our clird direct to delta gilo bravao	lufthansa four two one nine is identified you're cleared direct to delta kilo bravo
9	nago delta india raro ecco charlie decend to fligh leveel three six zero	fl fl delta india bravo echo charlie descend to flight level three six zero

شکل ۳- ۸ خروجی مدل wav2vec2-large-robust رندوم از خود متن به همراه رونویسی ها از دیتاست ترافیک هوایی

فصل چهارم

نتیجه گیری

۴- جمع‌بندی و نتیجه‌گیری

در این گزارش ابتدا به مرور برخی روش‌های یادگیری علی‌الخصوص یادگیری خودنظارتی پرداختیم و مدل wav2vec2 را معرفی کردیم که از در فاز دومش از این روش استفاده می‌کند. سپس با تنظیم دقیق کردن همین مدل روی زبان‌های فارسی و انگلیسی توانستیم با مراحل تنظیم دقیق کردن مدل آشنا شویم تا بتوانیم برای پروژه اصلی که تنظیم دقیق کردن مدل wav2vec2 روی دیتای ترافیک هوایی است آماده باشیم. در نهایت نتایج و نرخ خطای کلمه‌ی هر فاز را نیز گزارش کردیم.

۴- ۱ نتیجه‌گیری

حوزه پردازش زبان طبیعی و گفتار دنیایی وسیع است که با این پروژه و با این کارآموزی فهمیدم همچنان عطش سیری ناپذیری نسبت به آن دارم. علاوه بر آن، آموخته‌هایی که در تنظیم دقیق کردن مدل داشتم شامل برخوردن به خطاهای متفاوت، تجربه‌های گوناگون از شرکت در جلسات آزمایشگاه پردازش گفتار دانشکده کامپیوتر شریف و محیط کاری و چالش‌های آن، همه و همه این چند ماه اخیر من را سرشار از یادگیری و لذت کرد.

۴- ۲ کارهای آینده

امید است در آینده بتوانیم تشخیص گفتار خودکار را برای داده های ترافیک هوایی فارسی هم داشته باشیم تا امنیت جان مسافران هم‌وطنمان تامین شود.

منابع و مراجع

[1] J. Zuluaga-Gomez *et al.*, "How Does Pre-trained Wav2Vec2.0 Perform on Domain Shifted ASR? An Extensive Benchmark on Air Traffic Control Communications." arXiv, Mar. 31, 2022. doi: 10.48550/arXiv.2203.16822.

[2] "Supervised, Unsupervised, and Semi-Supervised Learning | EnjoyAlgorithms." https://medium.com/enjoy-algorithm/supervised-unsupervised-and-semi-supervised-learning-64ee79b17d10 (accessed Oct. 27, 2022).

[3] "Self-Supervised Learning and Its Applications - neptune.ai." https://neptune.ai/blog/self-supervised-learning (accessed Oct. 27, 2022).

[4] "Fine-tuning a Neural Network explained - deeplizard." https://deeplizard.com/learn/video/5T-iXNNiwIs (accessed Oct. 27, 2022).

[5] "What are Language Models in NLP?" https://insights.daffodilsw.com/blog/what-are-language-models-in-nlp (accessed Oct. 27, 2022).

[6] nbro, "Answer to 'Which tasks are called as downstream tasks?,'" *Artificial Intelligence Stack Exchange*, Jun. 27, 2021. https://ai.stackexchange.com/a/28424 (accessed Oct. 27, 2022).

[7] "Wav2Vec 2.0: Self-Supervised Learning for ASR | Towards Data Science." https://towardsdatascience.com/wav2vec-2-0-a-framework-for-self-supervised-learning-of-speech-representations-7d3728688cae (accessed Oct. 27, 2022).

[8] "Wav2Vec 2.0: Self-Supervised Learning for ASR | Towards Data Science." https://towardsdatascience.com/wav2vec-2-0-a-framework-for-self-supervised-learning-of-speech-representations-7d3728688cae (accessed Oct. 27, 2022).

[9] "Google Colab - What is Google Colab?" https://www.tutorialspoint.com/google_colab/what_is_google_colab.htm (accessed Oct. 27, 2022).

[10] K. Amirou *et al.*, "Hugging Face - Wiki," *Golden*. https://golden.com/wiki/Hugging_Face-39P6RJJ (accessed Oct. 05, 2022).

[11] ZahraRahimii, "ZahraRahimii/Internship-at-Asr-Gooyesh." Oct. 27, 2022. Accessed: Oct. 27, 2022. [Online]. Available: https://github.com/ZahraRahimii/Internship-at-Asr-Gooyesh

واژه‌نامه‌ی فارسی به انگلیسی

پ

پایگاه‌داده ..... database

پردازش زبان طبیعی ..... natural language processing

پیوستگی یادگیری بدون نظارت ..... association in unsupervised learning

ت

ترنسفورمر ..... transformer

تشخیص گفتار ..... speech recognition

تشخیص گفتار خودکار ..... automatic speech recognition

تنظیم دقیق ..... fine-tune

خ

خوشه‌بندی یادگیری بدون نظارت ..... clustering in unsupervised learning

ر

رگرسیون یادگیری نظارت شده ..... regression in supervised learning

ط

طبقه‌بندی در یادگیری نظارت شده ..... classification in supervised learning

گ

گوگل کولب ..... google colab

م

مجموعه‌داده ..... dataset

مدل زبانی ..... language model

ن

نرخ خطای کلمه ..... word error rate

نسبت سیگنال به نویز ..... signal-to-noise ratio

ه

هاگینگ فیس ..... Hugging Face

ک

کار پایین دستی ..... downstream task

ی

یادگیری بدون نظارت ..... unsupervised learning

یادگیری تقویتی ..... reinforcement learning

یادگیری خود نظارتی ..... self-supervised learning

یادگیری نظارت شده ..... supervised learning

یادگیری نیمه نظارتی ..... semi supervised learning

واژه‌نامه‌ی انگلیسی به فارسی

A

association in unsupervised learning…. پیوستگی یادگیری بدون نظارت

automatic speech recognition……. تشخیص گفتار خودکار

C

classification in supervised learning… طبقه‌بندی در یادگیری نظارت شده

clustering in unsupervised learning… خوشه‌بندی یادگیری بدون نظارت

D

database…… پایگاه‌داده

dataset…….... مجموعه داده

downstream task…… کار پایین دستی

G

fine-tune…… تنظیم دقیق

G

google colab……... گوگل کولب

H

Hugging Face…….. هاگینگ فیس

L

language model…… مدل زبانی

N

natural language processing… پردازش زبان طبیعی

R

regression in supervised learning….. رگرسیون در یادگیری نظارت شده

reinforcement learning…… یادگیری تقویتی

S

self-supervised learning……. یادگیری خود نظارتی

semi supervised learning…... یادگیری نیمه نظارتی

signal-to-noise ratio…….. نسبت سیگنال به نویز

speech recognition…….. تشخیص گفتار

supervised learning……..... یادگیری نظارت شده

T

Transformer…….. ترنسفورمر

U

unsupervised learning…… یادگیری بدون نظارت

W

word error rate…… نرخ خطای کلمه