دانشگاه صنعتی امیرکبیر
(پلی‌تکنیک تهران)
دانشکده مهندسی برق

گزارش کارآموزی
گرایش الکترونیک

محل کارآموزی

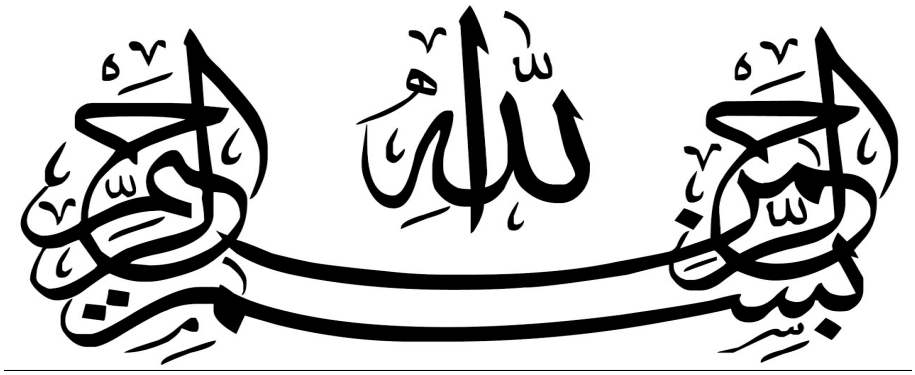
شرکت عصر گویش پرداز

نگارش
پارسا محمدی

استاد راهنما
دکتر ساناز سیدین

سرپرست کارآموزی
دکتر حمزه بیراوند

شهریور ۱۴۰۲


بسم الله الرحمن الرحيم

سپاس گزاری

اینجانب پارسا محمدی مراتب امتنان و تشکر خود را نسبت به استاد کارآموزی دکتر سانازسیدین و سرپرست محترم کارآموزی دکتر حمزه بیراوند که در گذراندن این دوره کارآموزی همواره مرا یاری نموده اند، ابراز می دارم.

پارسا محمدی

شهریور ۱۴۰۲

چکیده

مدل‌های بازشناسی خودکار گفتار دسته‌ای از سیستم‌های هوش مصنوعی هستند[۱] که برای تبدیل زبان گفتاری به متن نوشتاری طراحی شده‌اند. این مدل‌ها در طیف گسترده‌ای از کاربردها، از جمله خدمات رونویسی، دستیارهای صوتی و دستگاه‌های کنترل‌شده با صدا، نقش مهمی دارند.

در این پروژه مقاله یکی از جدیدترین و بهترین مدل های بازشناسی گفتار خوانده شده و مدل آن پیاده سازی شده است. این معماری جدید E-Branchformer نام دارد که توانسته است به نتایج بهتری در مقایسه با مدل های Conformer بدست بیاورد. در این پروژه کارآموزی به منظور ایجاد یک مدل بازشناسی گفتار برای زبان فارسی، مدل E-Branchformer بر روی داده های فارسی مجموعه دادگان کامن ویس[۲] به عنوان مدل صوت شناسی[۳] آموزش داده شده است و همچنین یک مدل زبانی ترانسفرمری نیز بر روی داده های متنی این پایگاه داده آموزش داده شده است. این مدل نهایی که ترکیب مدل صورت شانسی و زبانی می‌باشد موفق به کسب نرخ خطای کلمات[۴] ۳ درصد بر روی داده های تست کامن ویس شده است که با توجه به حجم کم دادگان نتیجه بسیار خوبی می‌باشد. در مرحله بعد این مدل بر روی سرور های هاگینگ فیس[۵] پیاده سازی شده است و مدل به صورت برخط[۶] برای عموم قابل دسترس می‌باشد.

واژه‌های کلیدی:

پردازش گفتار ، بازشناسی گفتار فارسی، پردازش زبان طبیعی

[1] Automatic Speech Recognition
[2] Common Voice
[3] Acoustics
[4] WER: Word Error Rate
[5] Huggingface
[6] Online

فهرست مطالب

۲ معرفی محل کارآموزی و پروژه کارآموزی ۴

۲-۱ معرفی شرکت ۵

۲-۱-۱ محصولات شرکت ۵

۲-۱-۲ زمینه‌های فعالیت ۶

۲-۲ پروژه کارآموزی ۷

۳ مدل شبکه عصبی و دادگان ۸

۳-۱ مطاله مقالات جدید ۹

۳-۲ معماری کانفرمر ۱۰

۳-۳ معماری ای-برانچفرمر ۱۲

۳-۴ مقایسه ای-برانچفرمر و کانفرمر ۱۴

۳-۵ دادگان ۱۵

۴ پیاده سازی عملی پروژه کارآموزی ۱۷

۴-۱ ارائه مباحث نظری پروژه در جلسه آزمایشگاه ۱۸

۴-۲ فراگیری ابزار مورد نیاز پروژه ۱۹

۴-۳ آموزش مدل بر روی دادگان ۲۱

۴-۴ بررسی خروجی و رفع ایرادات ۲۳

۴-۴-۱ اشتباه در انتخاب درست تعداد مجموعه توکن ها ۲۴

۴-۴-۲ عدم استفاده از کل داده ها ۲۵

۴-۴-۳ عدم استفاده از مدل زبانی ۲۵

۴-۵ فراگیری مباحث مورد نیاز برای پیاده سازی بر روی سرور ۲۶

۴-۶ پیاده سازی بر روی سرور ۲۷

۴-۷ مستندسازی و ارائه خروجی ۲۸

۵ جمع‌بندی و نتیجه‌گیری و پیشنهادات ۲۹

۵-۱ نتیجه‌گیر و جمع‌بندی ۳۰

۵-۲ پیشنهادات ۳۰

منابع و مراجع ۳۲

پیوست ۳۵

واژه‌نامه‌ی انگلیسی به فارسی ۴۱

فهرست تصاویر

شکل صفحه

۳-۱ تصویر معماری مدل ترانسفورمر. این مدل از یک بخش رمزگذار و رمزگشا تشکیل شده است و بخش های آن توسط فلش در تصویر مشخص است. ۱۱

۳-۲ معماری رمزگذار کانفرمر. ۱۲

۳-۳ معماری رمزگذار برانچفرمر. ۱۳

۳-۴ جدول مقایسه کننده دو مدل کانفرمر و ای-برانچفرمر. این دو مدل در ۱۲ دیتاست با هم مقایسه شده اند. ۱۴

۴-۱ تصاویر مربوط به روند آموزش مدل صوت شنانسی بازشناسی گفتار فارسی. این تصاویر خروجی تنسوربرد می‌باشند. ۲۲

۴-۲ نتایج دیکد مدل صوت شنانسی بدون مدل زبانی بر روی دادگان تست کامن ویس . . ۲۳

۴-۳ نمودار های روند آموزش و خروجی آنها بر روی مجموعه ارزیابی با ۱۵۰ توکن ۲۵

۴-۴ نتایج دیکد مدل صوت شنانسی به همراه مدل زبانی بر روی دادگان تست کامن ویس فارسی ۲۶

۴-۵ تصویر رابط کاربری نرم افزار بازشناسی گفتار فارسی ۲۸

۱ نامه ارسال شده توسط دکتر ریاحی به صنعت جهت معرفی کارآموز به شرکت ۳۵

۲ پاسخ شرکت به نامه دکتر ریاحی و علام موافقت شرکت ۳۶

۳ فرم تایید صنعت مبنی بر پذیرش کارآموز ۳۷

۴ فرم ارزیابی صنعت از دانشجو ۳۸

۵ نامه پایان ۲۴۰ ساعت کارآموزی در شرکت عصر گویش پرداز ۳۹

۶ گواهی شرکت در دوره کارآموزی شرکت عصر گویش پرداز ۴۰

فصل اول

مقدمه

صوت یکی از مهم‌ترین حالات انرژی در جهان ما می‌باشد و راه ارتباطی اصلی بسیاری از انسان‌ها و دیگر موجودات، از طریق سیگنال های صوتی می‌باشد. به همین دلیل، درک و پردازش این نوع از داده‌ها، اهمیت بسیاری در عصر حاضر برای ما دارا می‌باشد.

مدل‌های تشخیص خودکار گفتار دسته‌ای از سیستم‌های هوش مصنوعی هستند که برای تبدیل زبان گفتاری به متن نوشتاری طراحی شده‌اند. این مدل‌ها در طیف گسترده‌ای از کاربردها، از جمله خدمات رونویسی، دستیارهای صوتی و دستگاه‌های کنترل‌شده با صدا، نقش مهمی دارند. مدل‌های از تکنیک‌های یادگیری عمیق، مانند شبکه‌های عصبی ترانسفورمر برای پردازش داده‌های صوتی و تولید رونویسی دقیق از کلمات و عبارات گفتاری استفاده می‌کنند.

تشخیص خودکار گفتار سر به سر [1] یک روش در تشخیص خودکار گفتار است که در آن از یک مدل شبکه عصبی واحد برای تبدیل مستقیم زبان گفتاری به متن نوشتاری بدون نیاز به اجزای میانی مانند واج یا واحدهای زبانی استفاده می‌شود. هدف مدل‌های بازشناسی گفتار سر به سر ساده‌سازی مسیر[2] بازشناسی گفتار است و به دلیل توانایی‌شان در یادگیری نگاشت‌های پیچیده گفتار به متن محبوبیت پیدا کرده‌اند. آنها اغلب مبتنی بر معماری‌های یادگیری عمیق، مانند شبکه‌های عصبی مکرر [3] یا ترانسفورماتورها هستند، و در دستیابی به نتایج رقابتی در تسک های تشخیص گفتار، نویدبخش نشان داده‌اند، و آنها را برای برنامه‌هایی مانند خدمات رونویسی، دستیارهای صوتی و غیره ارزشمند می‌سازند.

در شرکت عصر گویش پرداز کار مشابه‌ای در زمینه بازشناسی گفتار قبلا انجام شده است. در حال حاظر نرم افزار نویسا شرکت عصر گویش پرداز برای کاربرد های تجاری بازشناسی گفتار استفاده می‌شود. اما محدودیت نرم افزار فعلی نویسا این است که این نرم افزار با مدل های ان گرام[4] آموزش داده شده است و روش های تعبیه‌سازی کلمات[5] برای آموش این مدل استفاده شده است؛ مدل فعلی در درک اسامی خاص و کلمات پیچیده ضعیف عمل می‌کند زیرا در دیکشینری آن کلمه خاص ممکن است موجود نباشد. اما مدل های خودنگرش[6] به روش های تعبیه سازی نیاز ندارند و خودشان می‌توانند ارزش کلمات را در جملات درک کنند. پروژه من آموزش یک مدل سر به سر بازشناسی گفتار فارسی می‌باشد که دارای معماری خودنگرش باشد.

[1] End to End ASR
[2] Pipeline
[3] RNN: Recurrent Neural Network
[4] Ngram
[5] Word Embedding
[6] Self Attention

در مطالعاتی که در این دوره کارآموزی انجام گرفت چند مقاله به منظور یافتن بهترین معماری برای تشخیص گفتار فارسی بررسی شد؛ یکی از بهترین کاندیدا ها مدل Whisper شرکت Open AI می‌باشد. این مدل توانایی بازشناسی گفتار به زبان های مختلف دارد، اما دقت خروجی آن برای زبان فارسی کم است. شرکت عصرگویش پرداز قبلا برای فاین تیون[۷] کردن این مدل اقدام کرده است اما با توجه به ماهیت نیمه نظارتی [۸] این مدل خروجی مدل فاین توین شده مناسب نبود. با همفکری که در شرکت صورت گرفت تصمیم بر این شد که از مدل های Conformer یا Branchformer برای آموزش مدل بازشناسی گفتار فارسی استفاده شود.

در ادامه، در فصل دوم به معرفی شرکت عصرگویش پرداز پرداخته و بخشی از مهم ترین محصولات و زمینه های فعالیت این شرکت بررسی خواهند شد همچنین به معرفی این پروژه کارآموزی خواهیم پرداخت. در فصل سوم و چهارم، تجربیات کسب شده در این دوره کارآموزی سه ماهه، بیان خواهد شد و برخی از چالش ها و راه حل هایی که در این دوره ارائه شدند، بررسی خواهند شد. در فصل سوم به مباحث تئوری پروژه از جمله مدل انتخابی و دادگان اشاره خواهد شد و در فصل چهارم به تجربیات عملی و چالش های آموزش و پیاده سازی اشاره مدل اشاره خواهد شد. در نهایت در فصل پنجم، نتیجه گیری مربوط به این دوره کارآموزی بیان خواهد شد و پیشنهاد هایی در جهت بهبود خروجی مدل ارائه شده، ذکر خواهد شد.

[7] Fine Tune

[8] Semi-Supervised

فصل دوم

معرفی محل کارآموزی و پروژه کارآموزی

در این قسمت، به طور مختصر، شرکت عصرگویش‌پرداز معرفی شده و در ادامه محصولات اصلی شرکت و همچنین زمینه‌های فعالیت این شرکت ذکر خواهند شد.

۲-۱ معرفی شرکت

عصر گویش پرداز (سهامی خاص) فعال‌ترین شرکت در زمینه هوش مصنوعی و پردازش سیگنال گفتار بوده که فعالیت خود را از ابتدای سال ۱۳۸۲ شروع کرده است. 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. این شرکت با پشتوانه فنی گروهی از متخصصان کشور از دانشگاه صنعتی شریف تأسیس شد که سابقه و تجربه پژوهشی آنها در زمینه‌های مرتبط با پردازش سیگنال به چندین سال قبل از شروع رسمی فعالیت شرکت برمی‌گردد.

۲-۱-۱ محصولات شرکت

عصرگویش پرداز پیشرو در ارائه سیستم های مبتنی بر گفتار برای زبان فارسی، محصولات مختلفی را توسعه داده است که بیشتر آنها برای نخستین بار برای زبان فارسی انجام شده و منحصراً توسط این شرکت تولید می‌شوند. برخی از محصولات این شرکت عبارتند از:

- 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

۲-۱-۲ زمینه‌های فعالیت

این شرکت امروزه دارای گروهی متخصص و منسجم از افرادی با تخصص و تجربه بالا بوده و سابقه طولانی و موفق در زمینه تحقیق و توسعه و کاربردی کردن توانمندی های پژوهشی دارد و علاوه بر ارائه محصولات مختلف در زمینه‌های هوش مصنوعی، پردازش گفتار فارسی و انگلیسی و پردازش تصویر، قادر به انجام پروژه های مختلف و ارائه خدمات در زمینه‌های مختلف نرم‌افزاری می‌باشد. از جمله زمینه‌های فعالیت این شرکت:

- تولید نرم افزارها و سخت افزارهای هوشمند
- هوش مصنوعی و شناسایی الگو
- پردازش سیگنال (گفتار و تصویر)
- تشخیص گفتار و تایپ گفتاری (تبدیل گفتار به متن)
- سنتز گفتار و متن خوان (تبدیل متن به گفتار)
- شناسایی افراد از روی صدا
- پردازش زبان طبیعی
- بهبود کیفیت گفتار
- طراحی دادگان‌های گفتاری و متنی

- طراحی، توسعه و پشتیبانی نرم افزارهای کاربردی مرتبط
- سیستم‌های تلفن گویا (با قابلیت تشخیص گفتار)
- سامانه‌های تلفنی مبتنی بر ویپ (استریسک، الستیکس و ...)
- برنامه نویسی روی ریزکامپیوترها (DSP، تلفن همراه و ...)

با توجه به نوآوری های انجام گرفته در شرکت عصرگویش پرداز، این شرکت علاوه بر انتشار مقاله‌های مختلف در نشریات و کنفرانس‌های علمی ملی و بین‌المللی، دارای افتخارات و تأییدیه‌های متعددی می‌باشد.

۲-۲ پروژه کارآموزی

شرکت عصر گویش پرداز هرساله در فصل تابستان تعداد محدودی کارآموز از دانشجویان بهترین دانشگاه های ایران جذب می‌کند. دانشجویانی که بعد از ارسال رزومه و قبولی در مصاحبه انتخاب می‌شوند در دوره سه ماه کارآموزی شرکت عصر گویش پرداز مشغول می‌شوند. هر کارآموز باید یک یا دو پروژه از پروژه های فعال شرکت را تکمیل کند و پس از پیاده سازی و ارائه خروجی پروژه به مسئول کارآموزی گواهی اتمام کارآموزی را دریافت می‌کند. پروژه های کارآموزی از پروژه های فعال و حل نشده شرکت می‌باشد. دانشجویان بعد از اینکه به گروه های چند نفری تقسیم شدند بر اساس دانش و علاقه آنها یک پروژه به آنها داده می‌شود و یکی از دانشجویان ارشد هوش مصنوعی دکتر صامتی (بنیان گذار شرکت) به عنوان رییس گروه مشخص شده و مسئول هدایت دانشجویان در طول مدت کارآموزی می‌باشد.

بعد از همفکری با استاد محترم کارآموزی دکتر سیدین و همچنین مشورت با مسئول کارآموزی شرکت پروژه بازشناسی گفتار فارسی انتخاب شد. من در طول دوره کارآموزی موظف به پیدا کردن بهترین مدل ترانسفرمری برای این امر و پیاده سازی آن بر روی سرور های شرکت بودم. مدل جدید قرار است بجای مدل قدیمی شرکت در نرم افزار نویسا قرار گیرد. در فصل بعدی به مدل انتخاب شده، معماری آن، دادگان استفاده شده و چالش های آن بیان خواهد شد.

فصل سوم

مدل شبکه عصبی و دادگان

۳-۱ مطاله مقالات جدید

در این فصل به مباحث نظری این پروژه کارآموزی اشاره می کنیم. در این فصل بیان می شود که چرا برای این پروژه نیاز به مطاله جدید ترین مقالات حوزه بازشناسی گفتار می‌باشد. در ادامه بهترین مدل های فعلی بازشناسی گفتار اشاره می شود و خلاصه ای از مقالات آنها بیان می شود ؛ و بعد از مقایسه بهترین مدل های این حوزه بیان می شود که مدل نهایی پروژه E-Branchformer می‌باشد. با استفاده این مدل و دادگان کامن ویس[۱] مدل بازشناسی فارسی آموزش داده شده است. در ادامه به جزئیات دادگان استفاده شده در این پروژه اشاره خواهد شد. در فصل بعدی به جزئیات پیاده سازی و چالش های آن اشاره خواهد شد.

با توجه به اینکه هدف اصلی این پروژه کارآموزی بروزرسانی مدل فعلی استفاده شده در نرم افزار بازشناسی گفتار فارسی(نویسا) می‌باشد[۲]؛ باید جدید ترین مقالات در حوزه بازشناسی گفتار مطالعه شود تا با جدیدترین مدل ها و روش های بازشناسی گفتار آشنا شد. پس از انجام تحقیقات و مشورت با ارشد پروژه نتیجه این شد که با توجه به اینکه امروزه مدل های سر به سر[۳] بهترین نتایج را دارند از این مدل های استفاده شود. همچنین مدل های ترانسفورمری با دارای بودن خاصیت خودنظارتی[۴] در حجم داده های زیاد نتایج بسیار خوبی را خروجی می‌دهند. اخیرا مدل های ترانسفورمری با شبکه های کانولوشنی ترکیب شده اند این امر باعث بهبود قابل توجهی در خروجی آنها شده است.

در مطالعاتی که در این دوره کارآموزی انجام گرفت چند مقاله به منظور یافتن بهترین معماری برای تشخیص گفتار فارسی بررسی شد؛ یکی از بهترین کاندیدا ها مدل Whisper شرکت Open AI می‌باشد. این مدل توانایی بازشناسی گفتار به زبان های مختلف دارد، اما دقت خروجی آن برای زبان فارسی کم است. شرکت عصرگویش پرداز قبلا برای فاین تیون کردن این مدل اقدام کرده است اما با توجه به ماهیت نیمه نظارتی این مدل خروجی مدل فاین توین شده مناسب نبود. با مطالعه و همفکری که در شرکت صورت گرفت تصمیم بر این شد که از مدل های Conformer یا Branchformer برای آموزش مدل بازشناسی گفتار فارسی استفاده شود.

[1] Common Voice
[2] https://nevisalive.com/
[3] End to End
[4] Self-attention

۳-۲ معماری کانفرمر

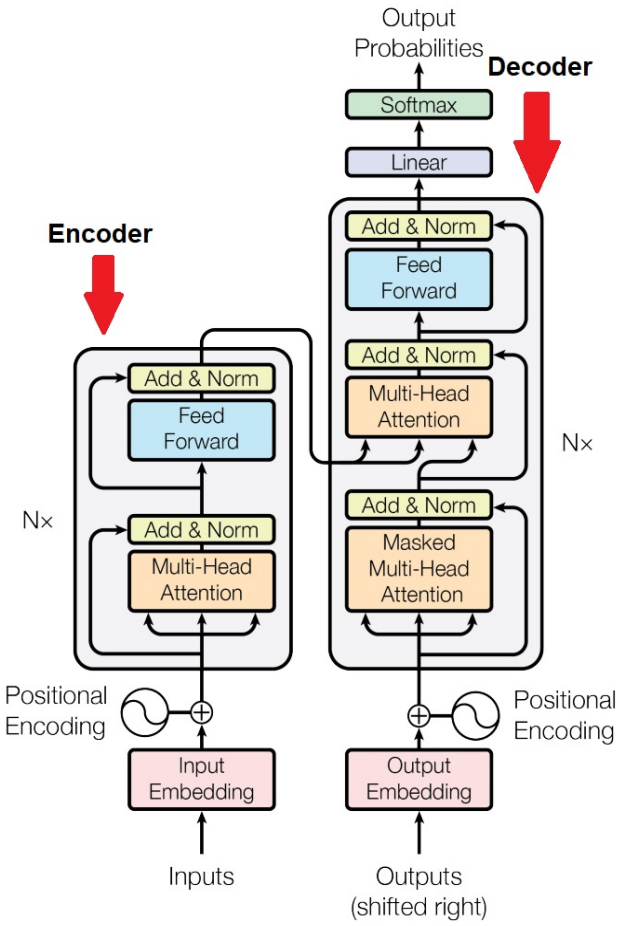
باتوجه به ویژگی های ارزشمند معماری کانفرمر[۵] در این دوره کارآموزی ابتدا مقاله این مدل خوانده شد و گزارشی از آن تهیه شد. در ادامه ویژگی های این مدل و معماری آن را مورد بررسی قرار می دهیم.

اخیراً مدل‌های مبتنی بر شبکه عصبی ترانسفورماتور و کانولوشن [۶] نتایج امیدوارکننده‌ای را در تشخیص خودکار گفتار بازشناسی گفتار نشان داده‌اند که عملکرد بهتری از شبکه‌های عصبی بازگشتی دارد. مدل‌های ترانسفورماتور در ثبت تعاملات جهانی [۷] مبتنی بر محتوا خوب هستند، در حالی که CNN ها از ویژگی‌های محلی به طور موثر بهره‌برداری می‌کنند. ترکیب این دو مدل باعث می‌شود که هم ویژگی های محلی و هم جهانی به خوبی استخراج شود. مدل کانفرمر به واسطه معماری خاص خود توانسته است که ویژگی ترانسفورماتور ها و کانولوشن ها را با هم ترکیب کند، و هم در دنباله های کوتاه و بلند داده ها به خوبی عمل کنند. [۱]

معماری مدل کانفورمر مانند دیگر مدل های سر به سر ترانسفورمری می‌باشد. این مدل از یک بخش رمزگزار[۸] و یک بخش رمزگشا[۹] تشکیل شده است. این دو بخش توسط جوینت های CTC به هم متصل شده اند. مدل ترانسفورمر در سال ۲۰۱۷ توسط مقاله `attention all you need` منتشر شد. [۲]

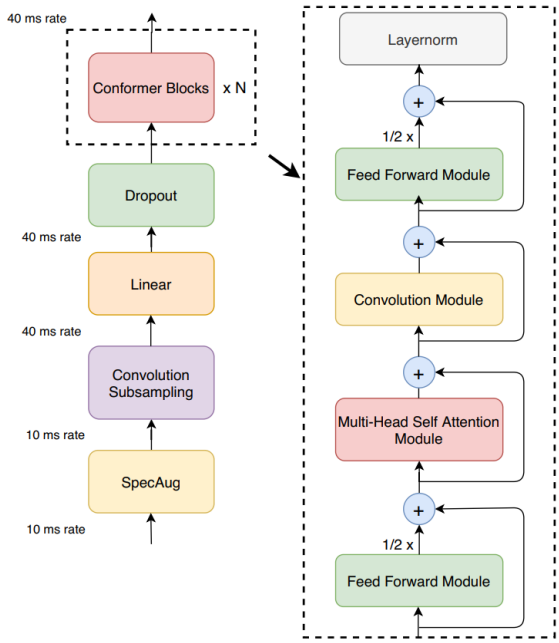
همین طور که در تصویر ۳-۱ مشخص است معماری ترانسفورمر ها از بخش های رمزگذار و رمزگشا تشکیل شده است و تفاوت معماری کانفرمر و ترانسفرمر عادی در ورودی مدل و بخش رمزگذار می‌باشد. در رمزگذار از ترکیب شبکه های کانولوشنی و ترانسفورمری استفاده شده است که این امر این مدل را قادر می سازد درک بهتری از ویژگی های متن ورودی بدست بیاورد.

[5] Conformer
[6] CNN: Convolutional Neural Network
[7] Global Optima
[8] Encoder
[9] Decoder



شکل ۳-۱: تصویر معماری مدل ترانسفورمر. این مدل از یک بخش رمزگذار و رمزگشا تشکیل شده است و بخش های آن توسط فلش در تصویر مشخص است.

همان طور که از تصویر ۳-۲ مشخص است. بلاک کانقرمر یک ماژول `Feed Forward` و بعد از آن ماژول خود نظارتی آمده است. سپس خروجی این ماژول وارد ماژول کانولوشن می شوند در نهایت نیمی از ابعاد وارد `Feed Forward` می شود. در مرحله آخر یک بلوک `Layernorm` قرار داده شده است.



شکل ۳-۲: معماری رمزگذار کانفرمر.

همان طور که پیش تر گفته شد این امر معماری کانفرمر را قادر می سازد که هم ویژگی های جهانی و هم ویژگی های محلی را به خوبی تشخیص بدهد. این معماری عالی معماری کانفرمر را به یکی از بهترین مدل های موجود برای کار های بازشناسی گفتار تبدیل کرده است. این مدل یکی از بهترین کاندیدا های مدل بازشناسی گفتار فارسی بود. من در این دوره کارآموزی مقاله کانفرمر را مطاله کردم و چند پیاده سازی از آن را بررسی کردم. مقاله خوانده شده در یک جلسه برای اعضای بخش تحقیقات و توسعه[۱۰] شرکت معرفی شد و نظرات آنها در مورد این مدل دریافت شد. این مدل با تمام ویژگی های خوب آن اندکی قدیمی شده است و مدل های جدید تر وجود دارند که ادعا می کنند به نتایج بهتری دست یافته اند. در بخش بعدی به یکی از بهترین آنها اشاره می شود.

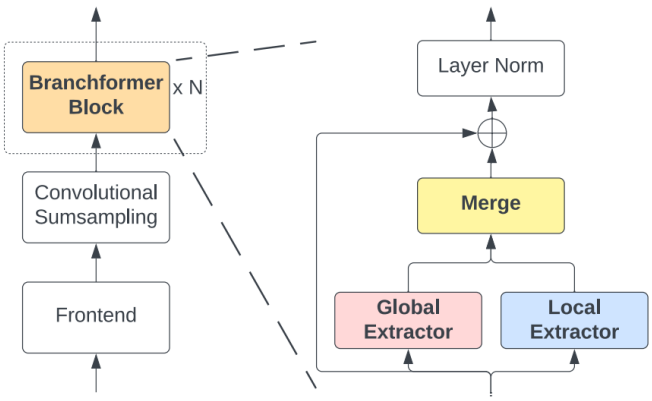
۳-۳ معماری ای-برانچفرمر

یکی از جدید ترین معماری های ارائه شده در حوزه بازشناسی گفتار معماری ای-برانچفرمر[۱۱] می‌باشد. این مدل که در کنفراس Interspeech سال ۲۰۲۳ معرفی شد و نتایج خوبی که بر روی دیتاست های مختلف بدست آورد توجه های زیادی را بخودش جلب کرده است.

[10] Research and Development
[11] E-Branchformer

مدل E-Branchformer بهبود یافته مدل Branchformer می‌باشد.[۳] حرف E مخفف کلمه Enhanced می‌باشد که به معنای تقویت شده می‌باشد و ای-برانچفرمر تقویت شده مدل برانچفرمر می‌باشد. برانچفرمر های ساختاری بسیار مشابه به کانفرمر ها دارند. در برانچفرمر ها نیز فقط معماری بخش رمزگذار با معماری ترانسفر ها متفاوت است، در رمزگذار آن مانند کانفرمر ترکیبی از بلوک های کانفرمری و خود نظارتی استفاده می شود تا هم ویژگی های محلی و هم جهانی داده ها به خوبی توسط مدل درک شود.

تفاوت اصلی کانفرمر و برانچفرمر در نوع ترکیب بلوک های کانفرمری و خود نظارتی می‌باشد. در برانچفمر ها تلاش شده است که این ترکیب به صورت موازی انجام شود، در حالی که در کانفرمر این ترکیب به صورت سری انجام شده است این امر در شکل ۳-۲ مشخص است. موازی سازی انجام شده در برانچفرمر ها باعث می شود که عمق شبکه عصبی کمتر شود و راحت تر به نقطه بهینه همگرا شود. همچنین تعداد پارامتر های استفاده شده در برانچفرمر در تعداد لایه های برابر کمتر از کانفرمر ها می‌باشد، که امر باعث کاهش هزنیه محاسباتی آموزش این مدل می شود.



شکل ۳-۳: معماری رمزگذار برانچفرمر.

شکل ۳-۳ معماری رمزگذار برانچفرمر را نمایش می دهد. همین طور که مشخص است نیمی از ابعاد داده ها به استخراج کننده محلی[۱۲] و نیم دیگر به استخراج کننده جهانی داده می شود. این نوع معماری خاص مدل را قادر می سازد که هم ویژگی های محلی و هم ویژگی های جهانی[۱۳] را به خوبی تشخیص دهد و مدل هم در دنباله[۱۴] داده های کوتاه و هم در دنباله داده های بلند خوب عمل کند.

[12] Local
[13] Global
[14] Sequence

تفاوت برانچفرمر و ای-برانچفرمر این نوع ترکیب این دو بخش موازی می‌باشد. ای-برانچفرمر بسیار بهینه تر با این دو بخش را باهم ترکیب می‌کند. در ادامه برای یافتن بهترین مدل برای بازشناسی گفتار باید مطالعه ای صورت بپذیرد که بهترین مدل ممکن برای آموزش استفاده شود. در بخش بعدی به مطالعه مقاله ای پرداخته می شود که این دو مدل برتر یعنی ای-برانچفرمر و کانفرمر را با هم مقایسه کرده است و بهترین مدل را معرفی کرده است.

۳-۴ مقایسه ای-برانچفرمر و کانفرمر

اخیرا مقاله ای چاپ در چاپ شده است که دو مدل برتر ای-برانچفرمر و کانفرمر را باهم مقایسه می‌کند.[۴] این مقاله این دو مدل را در سه تا از تسک های گفتار بازشناسی گفتار، ترجمه گفتار و درک گفتار بررسی می‌کند. مقایسه در تعدادی از بزرگترین و معروف ترین دیتاست های منبع باز صورت گرفته است؛ و خروجی های این مدل ها بر روی این دیتاست ها با هم مقایسه شده اند.

Dataset	Token	Metric	Evaluation Sets	Conformer			E-Branchformer		
				Params	MACs	Results ↓	Params	MACs	Results ↓
AIDATATANG [38]	Char	CER	dev / test	46.0	14.7	‡ 3.6 / 4.3	45.4	15.5	‡ **3.4 / 4.1**
AISHELL [39]	Char	CER	dev / test	46.3	15.3	4.3 / 4.6	45.7	15.5	**4.2 / 4.4**
AphasiaBank [30]	Char	WER	patients / control	44.2	30.1	40.3 / 35.3	45.7	32.0	**36.2 / 31.2**
CHiME4 [40]	Char	WER	{dt05,et05}_{simu,real}	30.4	8.8	‡ 7.8 / 9.5 / 12.5 / 14.8	30.8	8.8	‡ **6.8 / 8.4 / 10.8 / 13.0**
Fisher-Callhome [41]	BPE	WER	dev / dev2 / test / devtest / evltest	43.8	11.6	20.7 / 20.9 / 19.4 / 38.3 / 38.8	43.2	12.1	**20.5 / 20.2 / 18.7 / 37.8 / 37.6**
FLEURS [42]	BPE	CER	dev / test	†126.6	48.8	‡ 10.1 / 10.4	†127.4	50.1	‡ **9.3 / 9.2**
GigaSpeech [43]	BPE	WER	dev / test	116.2	20.0	10.9 / 10.8	148.9	26.1	**10.6 / 10.5**
JSUT [44]	Char	CER	dev / eval1	45.1	11.6	‡ 12.3 / 13.6	44.2	12.1	‡ **11.8 / 13.0**
LibriSpeech 100h [14]	BPE	WER	{dev,test}_{clean,other}	39.0	10.3	6.3 / 17.0 / 6.6 / 17.2	38.5	9.9	**6.1 / 16.7 / 6.3 / 17.0**
LibriSpeech 960h [14]	BPE	WER	{dev,test}_{clean,other}	147.8	42.5	‡ 1.72 / 3.65 / **1.85** / 3.95	148.9	42.7	‡ **1.67 / 3.64 / 1.85 / 3.71**
MuST-C [45]	BPE	WER	tst-{COMMON, HE}.en-de	46.1	12.0	7.7 / 6.7	37.7	9.9	**7.3 / 6.0**
Switchboard [46]	BPE	WER	eval2000 (callhm / swbd)	36.7	10.3	13.5 / 7.4	36.2	9.9	**13.4 / 7.3**
TEDLIUM2 [47]	BPE	WER	dev / test	35.5	10.3	7.5 / 7.6	35.0	9.9	**7.3 / 7.1**
VoxForge [48]	Char	CER	dt_it / et_it	35.2	13.2	9.0 / 8.1	34.7	12.6	**8.8 / 8.0**
WSJ [49]	Char	WER	dev93 / eval92	35.2	13.2	‡ **6.5 / 4.1**	34.7	12.6	‡ **6.5** / 4.3

شکل ۳-۴: جدول مقایسه کننده دو مدل کانفرمر و ای-برانچفرمر. این دو مدل در ۱۲ دیتاست با هم مقایسه شده اند.

همین طور که در تصویر ۳-۴ مشخص است مدل ای-برانچفرمر نتایج بسیار بهتری نسبت به کانفرمر در اکثر دادگان بدست آورده است. این جدول و جداول دیگری که در این دو مدل را در تسک های مختلف مقایسه می کنند نشان می‌دهند که مدل ای-برانچفرمر نتایج بسیار بهتری نسبت به کانفرمر در هر سه تسک گفتار بدست آورده است. این مقاله بر برتری ای-برانچفرمر تأکید می‌کند زیرا معماری موازی شده رمزگذار آن باعث بهبود نتایج و کاهش هزینه محاسباتی شده است.

با توجه به نتایج بهتر مدل ای-برانچفرمر بعد از مشورت هایی که در گروه انجام شد تصمیم بر این

شد که این مدل برای بازشناسی گفتار فارسی استفاده شود و این مدل بر روی دادگان فارسی آموزش داده شود و خروجی آن بررسی شود. این مقاله همچنین اشاره شده است که از ابزار ESPnet برای آموزش دو مدل کانفرمر و ای-برانچفرمر بر روی دادگان مختلف استفاده شده است و دستورعمل های [15] دستور عمل ها پیشفرض آن در آموزش استفاده شده است. جعبه ابزار[16] یی-اس-پی نت[17] ابزاری بسیار قدرتمند در ضمیه بازشناسی گفتار می‌باشد و کاربرد گسترده آن در تسک های مختلف گفتار و ارجاعات زیاد در مقالات جدید، فراگیری این ابزار برای ادامه کار الزامی می‌باشد. در فصل بعدی به یادگیری این ابزار و چالش های پیاده سازی با آن اشاره شده است.

۳-۵ دادگان

بر اساس پیشنهاد ارشد پروژه تصمیم بر این شد که که مدل ابتدا بر رو دادگان پایگاه داده Common Voice آموزش داده شود و خروجی مدل بررسی شود. در شرکت عصر گویش پرداز مدل های قبلی بازشناسی گفتار ابتدا بر روی این دادگان آموزش داده شده اند؛ آموزش بر روی دادگان این امکان را فراهم می‌کند که نتایج خروجی این مدل با مدل های قدیمی شرکت مقایسه شود. این پایگاه داده زبان فارسی را هم پشتیبانی می‌کند و داده های مورد نیاز برای بازشناسی گفتار فارسی را دارا می‌باشد. در ادامه جزئیات این پایگاه داده بیان می شود.

کامن ویس[18] یک پروژه جمع سپاری است که توسط شرکت موزیلا برای ایجاد یک پایگاه داده رایگان برای نرم افزار تشخیص گفتار آغاز شده است. این پروژه توسط داوطلبانی پشتیبانی می شود که جملات نمونه را با میکروفون ضبط می کنند و ضبط های دیگر کاربران را بررسی می کنند. جملات بازشناسی شده در یک پایگاه داده صوتی که تحت مجوز مالکیت عمومی CC0 در دسترس است، جمع آوری می شود. این مجوز تضمین می‌کند که توسعه دهندگان می توانند از پایگاه داده برای برنامه های صوتی به متن بدون محدودیت یا هزینه استفاده کنند.

این پایگاه داده دارای داده صوتی و متنی از ۱۱۲ تا زبان های دینا می‌باشد و مجموعاً ۲۸ هزار ساعت داده در این پایگاه داده موجود است. داده های این پایگاه داده برای عموم مردم به راحتی در سایت این

[15] Recipe
[16] Toolkit
[17] ESPnet
[18] Common Voice

پایگاه داده در دسترس است.[19] در آخرین نسخه حال حاظر این دادگان برای زبان فارسی ۳۹۷ ساعت داده صوتی به همراه متن متناظر با آن موجود می‌باشد. البته فرمت فایل های صوتی mp3 می‌باشد که این فرمت برای پردازش در مدل های گفتاری مناسب نمی‌باشد و باید فرمت آن تغییر کند این خود یکی از چالش هایی بود که من در این پروژه با آن مواجه شدم که در فصل بعد به جزئیات آن اشاره خواهد شد. یکی از مشکلات این پایگاه داده این است که درستی تطابق همه داده های صوت و متن بررسی نشده است و فقط بخشی از داده ها توسط کابران بررسی شده است. با این حال بعد از بررسی که از این پایگاه داده انجام گرفت به این نتیجه رسیدیم که اکثر داده ها تطابق خوبی دارند و این پایگاه داده برای آموزش نسخه های اولیه مدل بازشناسی گفتار فارسی مناسب می‌باشد.

در ادامه از این پایگاه داده برای آموزش دادن مدل بازشناسی گفتار فارسی با معماری ای-برانچفرمر استفاده شده است و از ابزار ESPnet به این منظور مورد استفاده قرار گرفته شده است. در فصل بعدی به پیاده سازی عملی این پروژه کارآموزی و چالش های آن اشاره خواهیم کرد.

[19] https://commonvoice.mozilla.org/

فصل چهارم

پیاده سازی عملی پروژه کارآموزی

در این فصل به پیاده سازی عملی این پروژه و چالش های آن اشاره خواهد شد. با توجه به نتیجه گیری هایی که در مطالعات نظری فصل قبل بدست آمد تصمیم بر این شد که معماری ای-برانچفرمر[1] با استفاده ابزار یی-اس-پی نت و دادگان Common Voice برای زبان فارسی آموزش داده شود. مدل حاصل اگر نتایج خوبی داشته باشد می تواند بجای مدل فعلی نرم افزار نویسا شرکت عصر گویش پرداز استفاده شود.

برای پیاده سازی عملی پروژه بر اساس دستور مسئول کارآموزی باید مراحل زیر انجام شود.

۱. ارئه گزارش از مباحث تئوری پروژه در جلسه آزمایشگاه دکتر صامتی
۲. فراگیری ابزار مورد نیاز پروژه
۳. آموزش مدل بر روی دادگان
۴. بررسی خروجی و رفع ایرادات
۵. فراگیری مباحث مورد نیاز برای پیاده سازی بر روی سرور
۶. پیاده سازی بر روی سرور
۷. مستندسازی و ارائه خروجی

در ادامه این فصل هر یک از مراحل پیاده سازی عملی پروژه توضیح داده خواهد شد.

۴-۱ ارائه مباحث نظری پروژه در جلسه آزمایشگاه

بعد از انجام مطالعات نظری که در فصل قبل بیان شد گزارشی از اقدامات انجام شده و مباحث آموخته شده آماده شد و در جلسه آزماشگاه دکتر صامتی خدمت خود دکتر و دستیاران ایشان ارائه شد. بعد از ارائه، از نظرات ایشان و دیگر اعضای آزمایشگاه برای ادامه کار استفاده شد. طبق نظر دکتر باید خروجی مدل بر روی پایگاه داده خود شرکت بررسی شود تا عملکرد آن را با مدل های قبلی شرکت مقایسه کرد. همچنین با توجه به اینکه در بازشناسی گفتار به دو مدل صوت شنانسی[2] و زبانی برای دریافت خروجی نیاز می باشد؛ باید مدل زبانی بر روی دادگان پایگاه داده ناب آموزش داده شود تا مدل زبانی قوی برای این پروژه تهیه شود.

[1] E-Branchformer
[2] Acoustic

پیکره متنی ناب یکی از پروژه های شرکت عصر گویش پرداز می‌باشد که در آن ۲۲۵۸۹۲۹۲۵ جمله فارسی موجود می‌باشد که این جملات شامل متن های رسمی،غیر رسمی و حتی انواع اشعار فارسی می باشد. حجم زیاد داده های این مدل و تنوع خوب آن، این پایگاه داده منبع باز را به مرجع کاملی برای زبان فارسی کرده است.[۵]

۲-۴ فراگیری ابزار مورد نیاز پروژه

یکی از طولانی ترین و سخت ترین بخش های پروژه در یادگیری جعبه ابزار[۳] مورد نیاز این پروژه بود. بعد از مطالعات تئوری نظر بر این شد که از ابزار یی-اس-پی نت[۴] در این پروژه استفاده شود. یی-اس-پی نت ابزاری بسیار قدرتمند در پردازش گفتار است و همه تسک های پردازش گفتار را به خوبی پوشش می دهد. این ابزار اخیرا بسیار توسط محققین استفاده میشود و پیاده سازی پردازش گفتار به این ابزار انجام می دهند. با توجه به این که در مقاله ای-برانچفرمر این معماری در یی-اس-پی نت پیاده سازی شده است پس یادگیری این ابزار کار آمد اجتناب ناپذیر می‌باشد.

یی-اس-پی نت عمدتاً بر روی بازشناسی خودکار گفتار سرتاسر [۵] تمرکز دارد و از ابزارهای شبکه عصبی پویا پرکاربرد، `Chainer` و `PyTorch` ، به عنوان یک موتور یادگیری عمیق اصلی استفاده می‌کند. یی-اس-پی نت همچنین از سبک جعبه ابزار `Kaldi ASR` برای پردازش داده ها، استخراج ویژگی/قالب، و دستور العمل ها پیروی می کند تا یک راه اندازی کامل برای تشخیص گفتار و سایر آزمایش های پردازش گفتار ارائه دهد. یی-اس-پی نت به طور کامل از مزایای دو پیاده سازی `ASR` سر به سر بر اساس طبقه بندی زمانی اتصالگرا [۶] و شبکه انکدر-دیکدر مبتنی بر توجه[۷] استفاده می کند. روش‌های مبتنی بر توجه از مکانیزم توجه برای انجام هم‌ترازی بین فریم‌های صوتی و نمادهای شناسایی استفاده می‌کنند، در حالی که `CTC` از مفروضات مارکوف برای حل مؤثر مسائل متوالی توسط برنامه‌نویسی پویا استفاده می‌کند. یی-اس-پی نت ترکیبی `CTC` و `Attention` سر به سر را اتخاذ می کند که به طور مؤثر از مزایای هر دو معماری در آموزش و رمزگشایی استفاده می کند. در طول آموزش، از چارچوب یادگیری چندهدفه برای بهبود استحکام در ترازهای نامنظم و دستیابی به همگرایی سریع استفاده می‌شود. در طول دیکد

[3] Toolkit
[4] ESPnet
[5] End to End
[6] CTC: Connectionist Temporal Classification
[7] Attention

کردن، رمزگشایی مشترک را با ترکیب امتیازات مبتنی بر توجه و CTC در یک الگوریتم جستجوی پرتوی یک‌گذر انجام می‌دهد تا ترازهای نامنظم را حذف کند.[۶] [8]

یکی از بزرگ ترین مشکلات این ابزار نبود مستندات کافی برای آموزش کار با این ابزار می‌باشد. با توجه به اینکه هیچ کس در شرکت به این ابزار تسلط نداشت و آموزشی هم در اینترنت برای کار با این مدل موجود نبود؛ فراگیری کار با این ابزار بسیار سخت و زمان بر بود. برای یادگیری کا با ابزار من جبور شدم که تمام کد های موجود در این صفحه این ابزار را مطالعه کنم. باگ ها پیام های مناسبی را نمایش نمی‌دادند و دیباگ کردن حل مشکلات با این ابزار بسیار زمان بر بود.

همچنین وابستگی این ابزار به ابزار Kaldi من را وادار کرد که مستندات این ابزار را هم مطالعه کنم تا بتوانم مدل خود را آموزش دهم. این ابزار سنگین است و نصب آن معمولا ۱۰ الی ۱۵ دقیقه طول می‌کشد و با توجه اینکه این ابزار بر روی گوگل کلب[9] نصب نمی‌باشد؛ برای هربار استفاده از کد باید یک بار آن را نصب کنیم که فرآینده آموزش مدل را بسیار طولانی می‌کند. با توجه به این مشکلات حجم زیاد دادگان تصمیم بر این شد که من یی-اس-پی نت را بر روی یکی از سرور های شرکت نصب کنم و آموزش را با آن انجام دهم. اینکار زمان نصب و دانلود داده ها را نسبت به گوگل کلب بسیار کمتر می‌کند اما کار با این ابزار یکی از چالش هایی بود که من همچنان باید با آن در طول این کارآموزی دست و پنجه نرم می‌کردم.

در نهایت بعد از دو هفته تلاش من موفق شدم که مدل صوت شنانسی را بر روی دادگان کامن ویس آموزش دهم. برای آموزش مدل های بازشناسی گفتار در ابزار یی-اس-پی نت اقدامات زیر انجام شد.

۱. دانلود داده ها

۲. انجام مراحل پیش پردازش (حذف داده های بلند و کوتاه، تبدیل داده ها به فرمت کلدی، تغییر فرمت داده به WAV و ...)

۳. افزایش داده ها برای جلوگیری از اورفیت شدن (افزایش و کاهش سرعت گفتار و اضافه کردن نویز به داده های صوتی)

۴. ایجاد توکن لیست[10] با استفاده از داده ها متنی

[8] https://github.com/espnet/espnet
[9] Google Colab
[10] Token List

۵. ایجاد پیکربندی[11] مدل (مشخص کردن معماری، مشخص کردن انجام مینی-بچ[12] ها با استفاده واریانس داده ها و ...)

۶. آموزش مدل (این مرحله دو روز به طول انجامید)

۷. دیکد کردن مدل بر روی دادگان آزمایش کامن ویس

۸. محاسبه متریک های خطا

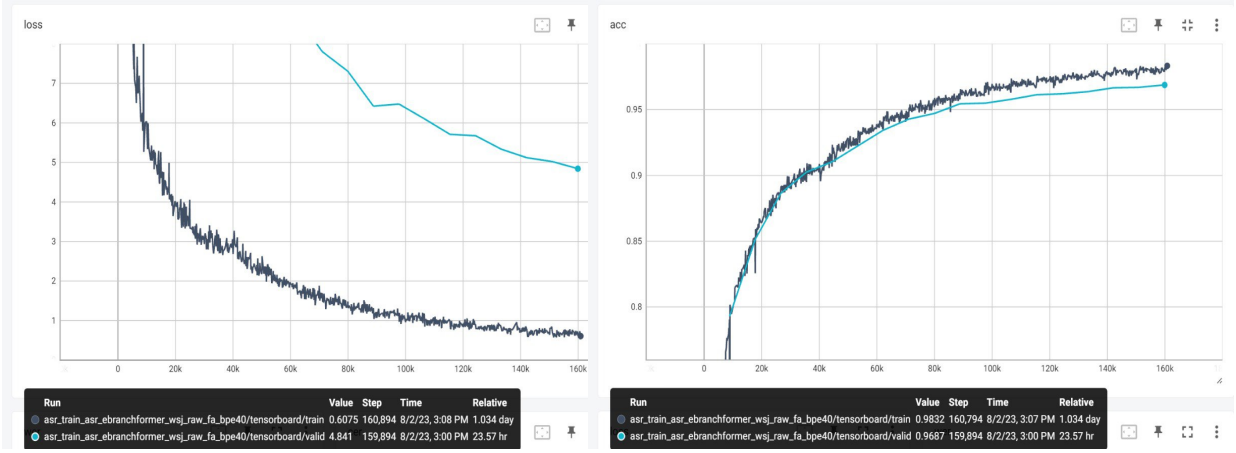
۴-۳ آموزش مدل بر روی دادگان

کدهای پیاده سازی اولیه موجود می‌باشد البته این کد ها در گوگل کلب زده شده اند ولی بر روی سرور شرکت اجرا شده اند. جزیئات معماری و هایپر پارامتر های آن در کد زیر قابل دریافت است [13]

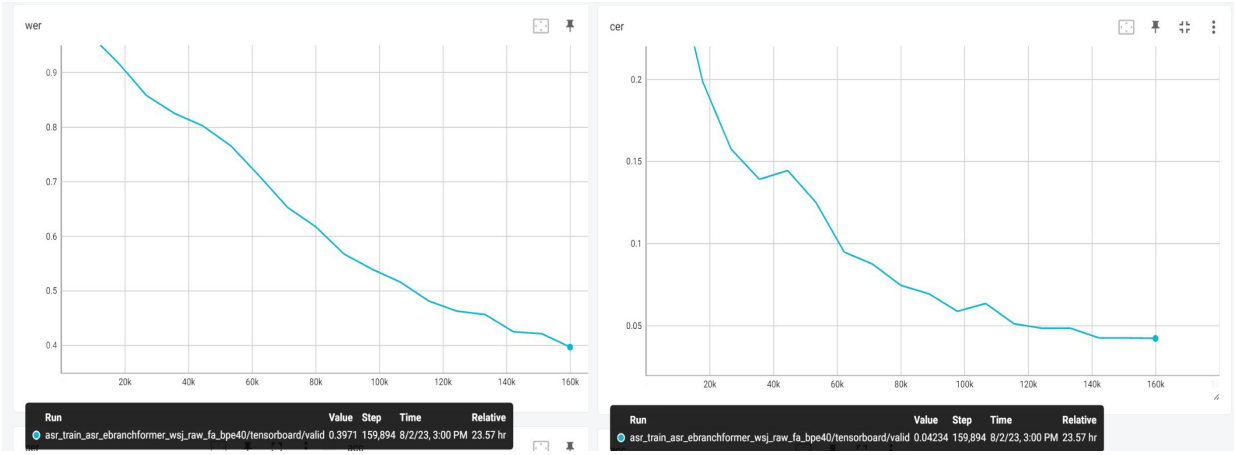
یکی دیگر از مشکلاتی که با آن در این دوره درگیر شدم کمبود حافظه واحد پردازش گرافیکی بود. به علت بزرگ بودن مدل و زیاد بودن حجم داده سرور شرکت قادر نبود مدل را آموزش دهد تا زمانی که در مرحله ۵ آموزش مدل مقدار بچ سایز ها را کوچک کردیم تا بتوانیم مدل را آموزش دهیم. این کار باعث شد روند آموزش پایدار شود اما زمان مورد نیاز برای آموزش مدل را افزایش داد به طوری که یک روز برای آموزش و نصف روز برای دیکد کردن مدل بر روی داده های تست استفاده شد.

شکل ۴-۱ روند آموزش مدل صوت شنانسی را نمایش می‌دهد. این تصاویر با استفاده تنسوربرد[14] ایجاد شده اند که درک روند آموزش مدل را با تصاویر و گزارشاتی که ارائه می‌کند بسیار راحت می‌کند.[۷] همین طور که از تصاویر مشخص است مدل روند بسیار پایداری در طول آموزش داشته است و خروجی بسیار رضایت بخش می‌باشد. برای نرخ خطای حروف[15] مقدار ۴ ٪۰ بر روی دیتای های ولیدیشن[16] بدست آمده است. و همچنین مقدار ۳۹۷٪۰ درصد برای نرخ خطای کلمه[17] بدست آمده است. نرخ خطا

[11] Config
[12] Mini-Batch
[13] https://colab.research.google.com/drive/15FfGd7uUQ-m8NUEn-wJLvjVJ62kQ9_T6?usp=sharing
[14] Tensor Board
[15] CER: Character Error Rate
[16] Validation
[17] WER: Word Error Rate



(ب) مقدار تابع هزینه مدل (آ) دقت مدل



(د) مقدار خطای کلمه مدل (ج) مقدار خطای حروف مدل

شکل ۴-۱: تصاویر مربوط به روند آموزش مدل صوت شنانسی بازشناسی گفتار فارسی. این تصاویر خروجی تنسوربرد می‌باشند.

کلمه و حروف از مهمترین متریک های ارزیابی مدل های بازشناسی گفتار می‌باشند که نتایج حاصله با توجه به حجم دادگان و نوع آنها بسیار رضایت بخش می‌باشد.

برای آموزش این مدل همان طور که مقاله ای-برانچفرمر تاکید کرده است [۳] از دستورعمل های پیش فرض یی-اس-پی نت استفاده شده است اما در مواردی که نیاز به تغییر جزئیات به منظور مناسب سازی برای زبان فارسی بوده است کد ها به صورت دستی تغییر داده شده‌اند. در با توجه به نبود هیچ دوره و آموشی در مورد آموزش دادن مدل زبانی با یی-اس-پی نت وجود نداشت به استفاده از مدل صوت شنانسی اکتفا شده اما در مراحل بعد یک مدل زبانی برای مدل آموزش داده شده است که خروجی آن را بهبود ببخشد.

۴-۴ بررسی خروجی و رفع ایرادات

همانطور که پیشتر اشاره شد در بازشناسی گفتار یک مدل زبانی به کنار یک مدل صوت شنانسی قرار می‌گیرد که درک گفتار و نوشتار آن به بهترین حالت ممکن صورت بپذیرد با این حال می‌توان از مدل صوت شنانسی به تنهایی برای این منظور استفاده کرد که باعث کاهش دقت زبانی گفتار رونویسی شده می‌شود. با این حال در این مرحله مدل صوت شنانسی را بر روی دادگان تست دیکد کردم تا خروجی تست آن را بررسی کینم.

```
# RESULTS
## Environments
- date: `Wed Aug  2 23:29:17 +0330 2023`
- python version: `3.10.12 (main, Jul  5 2023, 18:54:27) [GCC 11.2.0]`
- espnet version: `espnet 202304`
- pytorch version: `pytorch 1.13.1`
- Git hash: `a719135834e382e84b560e4ad869eaa3ef37ef09`
  - Commit date: `Sat Jul 29 21:06:06 2023 +0900`

## exp/asr_train_asr_ebranchformer_wsj_raw_fa_bpe40
### WER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|
|---|---|---|---|---|---|---|---|---|
|decode_asr_asr_model_valid.acc.ave/test_fa|3989|30347|88.0|10.9|1.1|5.0|16.9|45.8|

### CER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|
|---|---|---|---|---|---|---|---|---|
|decode_asr_asr_model_valid.acc.ave/test_fa|3989|145161|96.3|2.3|1.3|2.1|5.7|45.8|

### TER

|dataset|Snt|Wrd|Corr|Sub|Del|Ins|Err|S.Err|
|---|---|---|---|---|---|---|---|---|
|decode_asr_asr_model_valid.acc.ave/test_fa|3989|145253|96.4|2.3|1.3|2.0|5.7|45.8|
```

شکل ۴-۲: نتایج دیکد مدل صوت شنانسی بدون مدل زبانی بر روی دادگان تست کامن ویس

شکل ۴-۲ خروجی مدل را بر روی دادگان تست کامل ویس نشان می‌دهد. مقدار نرخ خطای کلمه ۱۶٫۹ برای کاربرد های امروز نرخ بالایی است اگرچه وجود یک مدل زبانی کنار مدل صوت شنانسی می‌تواند این نرخ خطا را تا نصف کاهش دهد با این حال در کاربرد های امروزی این نرخ بالایی می‌باشد.

در ادامه روند آموزش توسط ارشد پروژه مورد بررسی قرار گرفت و تعدادی از ایرادات که مرتبه اول آموزش مرتکب شدم شناسایی شد و برای اصلاح آنها تلاش شد. در ادامه به تعدادی از مهم ترین این خطاها و راهکار هایی که برای اصلاح آنها انجام دادم اشاره خواهد شد؛ پس از اصلاح این ایرادات مدل دوباره آموزش داده شد و نتایج بسیار ارزشمندی حاصل شد.

۴-۴-۱ اشتباه در انتخاب درست تعداد مجموعه توکن ها

بعد از بررسی هایی که توسط تیم انجام شد یکی از اشتباهات من در طول آموزش انتخاب اشتباه تعداد اعضای مجموعه توکن ها بود. در ابزار یی-اس-پی نت استیج پنجم آموزش در بازشناسی گفتار باید مجموعه توکن های مشخص شود. زمانی که از مدل های توجه استفاده می‌شود باید توکن لیستی از حروف زبان فارسی ایجاد شود که مدل هر آوایی را به یک توکن متناظر[۱۸] کند و رونویسی زبان فارسی را یادبگیرد. این کار می‌تواند به سه نوع در یی-اس-پی نت انجام شود.

۱. بر اساس حروف: در این حالت مجموعه توکن ها حروف های زبان فارسی می‌باشد

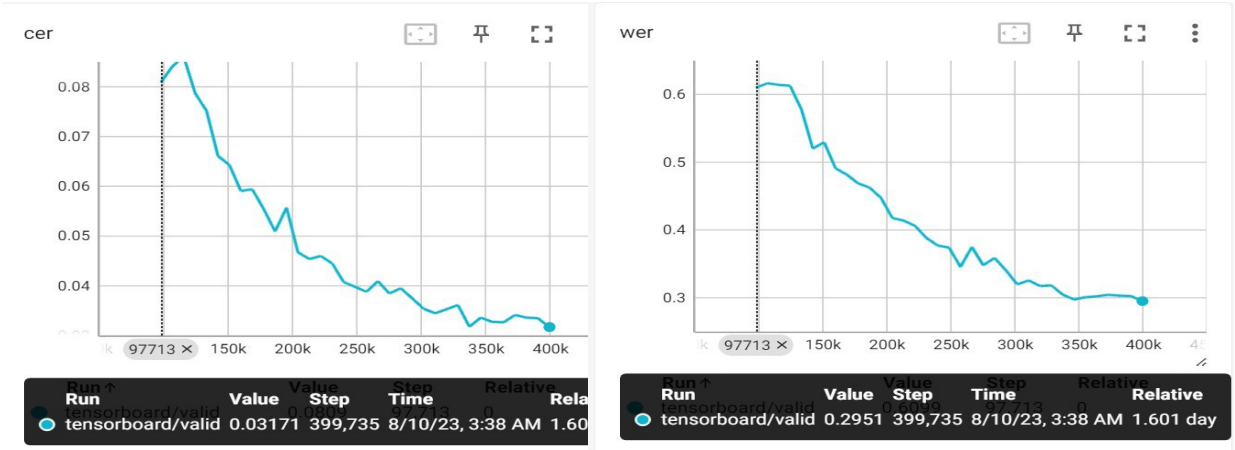
۲. بر اساس کلمات: در این حالت مجموعه توکن ها بر اساس کلمات زبان فارسی می‌باشد

۳. براساس مدل bpe: در حالت ترکیبی از کلمات و حروف ها برای زبان فارسی ایجاد می‌شود.

برای زبان فارسی با توجه به تکراری بودن بعضی از ترکیبات کلمات در اکثر جملات فارسی بهتر است که از bpe استفاده شود. برای استفاده از این نوع ایجاد کننده توکن باید به ابزار یه سری اطلاعات بدهیم. برای مثال باید تعداد اعضای مجموعه توکن ها را با مشخص کردن مقدار متغییر nbpe انجام دهیم. برای مثال اگر مقدار این متغییر را ۱۵۰ بگذاریم خود ابزار ای-ای-پی نت برای ما لیستی از ۱۵۰ توکن (که تعدادی از آنها حروف و تعدادی دیگر کلمه می‌باشند) ایجاد می‌کند و آن را در یک فایل به نام `tokens.txt` ذخیره می‌کند. این فایل بعدا توسط مدل برای آموزش استفاده می‌شود.

من در بار اول آموزش اشتباها مقدار nbpe را ۴۰ گذاشتم ولی در واقع مقدار استاندارد آن ۱۵۰ می‌باشد و این کار باعث شده بود بعضی از حروف مانند "ژ" در مجموعه توکن های نباشد. این امر باعث کاهش شدید دقت مدل شده بود و پس از اصلاح آموزش را دوباره آغاز کردم و نتایج جدید بدست آمد.

[18] Map



(آ) نرخ خطا کلمه بر روی ولیدیشن (ب) نرخ خطا حروف بر روی ولیدیشن

شکل ۴-۳: نمودار های روند آموزش و خروجی آنها بر روی مجموعه ارزیابی با ۱۵۰ توکن

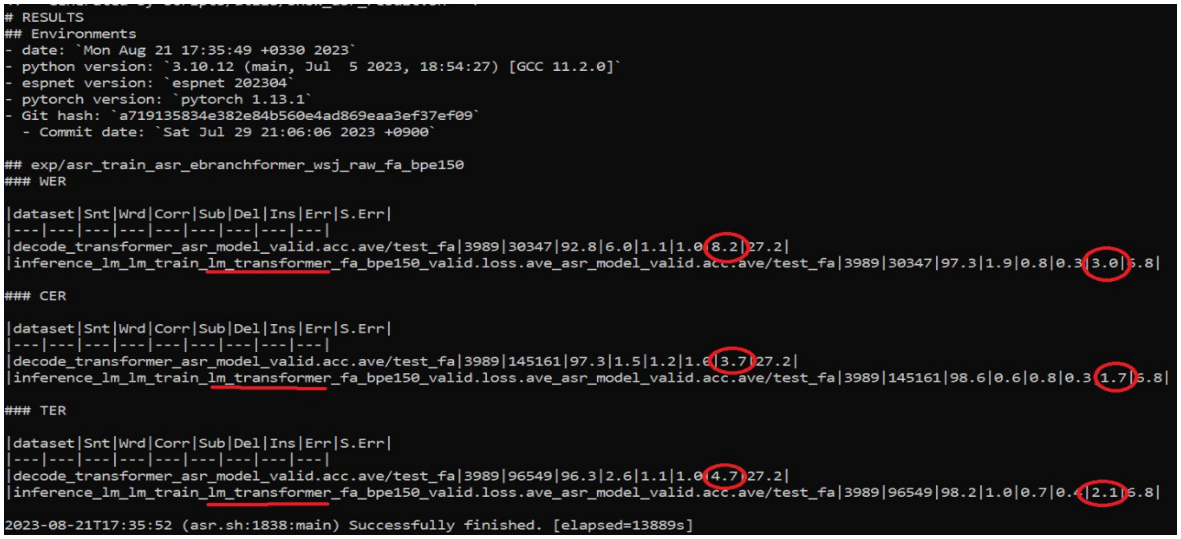
همین طور که در تصویر ۴-۳ب و ۴-۳آمشخص است بعد از اعمال تغییرات متریک های مدل بر روی ولیدیشن بهتر شده است. همچنین بر روی تست ست هم خروجی دوبرابر بهتر شده و نرخ خطای کلمات به ۸ درصد رسیده است.

۴-۴-۲ عدم استفاده از کل داده ها

با توجه به اینکه روند پیش پردازش داده ها بر روی CPU انجام می‌شود و باید تمام فایل های صوتی تغییر فرمت داده شوند من نتوانستم در اولین مرتبه آموزش از کل داده ها استفاده کنم زیرا پیش پردازش کل داده ها روی آن سرور بسیار ارزشمند بود و از بخشی از داده های استفاده کردم. تعداد کم داده ها باعث کاهش دقت مدل شده است و در آموزش جدید از یک سرور با CPU قدرتمند تر استفاده شد و خروجی ها بسیار بهبود یافت.

۴-۴-۳ عدم استفاده از مدل زبانی

همانطور که گفته شده برای بهرمندی از بهترین حالت ممکن خروجی باید مدل صوت شنانسی و زبانی کنار هم آموزش داده شوند و استفاده شوند. به همین منظور در این مرحله داده های یک مدل زبانی ترانسفرمری (که در نقش دیکدر در بازشناسی گفتار استفاده می‌شود) بر روی دادگان کامن ویس فارسی آموزش داده شد.آموزش این مدل نصف روز زمان برد. و در نهایت مدل نهای (که شامل مدل صوت شناسی و زبانی است) بر روی دادگان تست کامن ویس فارسی دیکد شده‌اند و نتایج زیر حاصل شده‌اند.



شکل ۴-۴: نتایج دیکد مدل صوت شنانسی به همراه مدل زبانی بر روی دادگان تست کامن ویس فارسی

نتایجی که تصویر ۴-۴ نمایش می‌دهد بسیار عالی و رضایت بخش است. من موفق شدم به نرخ خطای ۳ درصد در دادگان تست کامن ویس برسم که این نتیجه در کاربرد فعلی بازشناسی گفتار بسیار عالی می‌باشد. بعد از رسیدن به این نتیاج عالی خروجی مدل در یک جلسه خدمت دکتر صامتی و بقیه اعضای آزمایشگاه نمایش داده شد و مورد قدردانی و تشویق همه اعضا قرار گرفت.

در مرحله بعد کار باید مدل بر روی سرورها پیاده سازی شود که جزئیات مربوط به آن در بخش بعدی بیان خواهد شد.

۴-۵ فراگیری مباحث مورد نیاز برای پیاده سازی بر روی سرور

در این مرحله کار باید مدل ساخته شده بر روی یک سرور پیاده سازی شود تا بتوان از آن به عنوان یک سرویس جدید برای شرکت استفاده کرد. برای مرحله اول ابتدا باید یک نمونه ساده پیاده سازی شود تا خروجی کار و سرعت عملکرد آن بررسی شود. در ادامه به بررسی چند روش برای پیاده سازی مدل بر روی بک‌اند سرور پرداختم که در نهایت به این نتیجه رسیدم که استفاده از فریم ورک گرادیو[۱۹] آسان ترین و سریع ترین راه پیاده سازی مدل های هوش مصنوعی بر روی بک‌اند سرور می‌باشد.[۲۰]

گرادیو یک کتابخانه پایتون منبع باز است که برای ساخت دموهای یادگیری ماشین و علوم داده و برنامه های کاربردی وب استفاده می شود. با گرادیو ، می توانید به سرعت یک رابط کاربری زیبا در اطراف

[19] Gradio

[20] https://github.com/gradio-app/gradio

مدل های یادگیری ماشین یا گردش کار علم داده خود ایجاد کنید و به افراد اجازه دهید با کشیدن و رها کردن در تصاویر خود، چسباندن متن، ضبط صدای خود و تعامل با آن، آن را امتحان کنند.[۸]

۶-۴ پیاده سازی بر روی سرور

در این مرحله از پروژه من اقدام به پیاده سازی مدل بر روی گرادیو کردم. یکی از بزرگترین چالش های این بخش فهمیدن این مسئله بود که چگونه می‌توانیم از مدل آموزش دیده در گرادیو استفاده کنم. با نبود هیچ آموزش و داکیومنتشن در این رابطه من مجبور شدم که دوباره تمام کد های پی-اس-پی نت را بررسی کنم تا در نهایت موفق به پیدا کردن و استفاده از فایل های مربوط به این زمینه در پی-اس-پی نت شدم. گرادیو را می‌توان به راحتی در هر سروری پیاده سازی کرد ولی برای استفاده عمومی باید سرور قابلیت هاستینگ را داشته باشد. به همین منظور از سرور های هاگینگ فیس[۲۱] برای پیاده سازی مدل استفاده شده است.[۲۲]

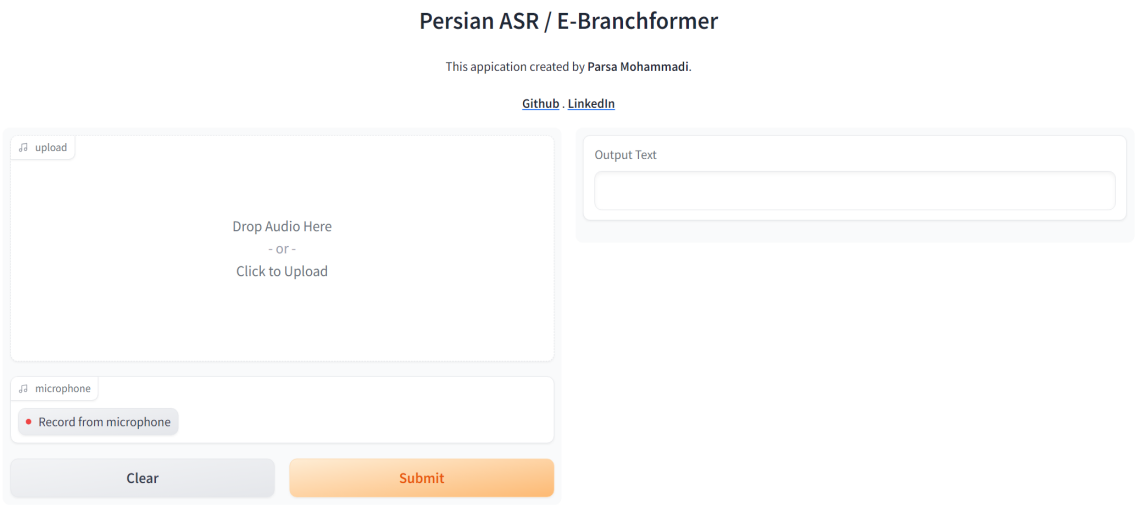
هاگینگ فیس یک شرکت فرانسوی-آمریکایی است که ابزارهایی را برای ساخت برنامه های کاربردی با استفاده از یادگیری ماشین، مستقر در شهر نیویورک توسعه می دهد. این به خاطر کتابخانه ترانسفورماتورهای خود که برای برنامه های کاربردی پردازش زبان طبیعی ساخته شده است و پلتفرم آن که به کاربران اجازه می دهد مدل ها و مجموعه داده های یادگیری ماشین را به اشتراک بگذارند و کار خود را در یک فضا به نمایش بگذارند قابل توجه است.[۹]

بعد از یادگیری کار با هاگینگ فیس و گرادیو موفق شدم نسخه اولیه نرم سرویسی باز شناسی گفتار فارسی را پیاده سازی کنیم. برای دسترسی به این نرم افزار بررسی آن کافی است بر روی این لینک کلیک کنید. ر این نرم افزار امکان ضبط فایل صدا و همچینین ارسال فایل صوتی وجود دارد و پس از چند ثانیه پردازش نرم افزار متن گفته شده در فایل صوتی را بازنویسی می‌کند.[۲۳] با تمام تلاش های صورت گرفته این نرم افزار همچنان در مرحله توسعه می‌باشد و هنوز آماده کار های تجاری نشده است برای کار های حرفه ای همچنان به کار های بیشتر نیاز دارد. با توجه به اینکه سرور های رایگان هاگینگ فیس فقط اجازه استفاده از CPU می‌دهد ممکن زمان پردازش مقداری طولانی تر از GPU باشد.

[21] Hugging Face

[22] https://github.com/huggingface , https://huggingface.co/

[23] https://huggingface.co/spaces/parsa-mhmdi/persian-asr



شکل ۴-۵: تصویر رابط کاربری نرم افزار بازشناسی گفتار فارسی

۴-۷ مستندسازی و ارائه خروجی

در نهایت کار خلاصه از تمام انجام شده و نتایج حاصل شده مستند سازی شده و در جلسه آزماشیگاه ارئه شد. دکتر صامتی ضمن قدردانی پیشنهاد پیاده سازی این مدل بازشناسی گفتار برای زبان عربی را دادند که در صورت ادامه همکاری با شرکت این پروژه برای کارفرمای مربوطه انجام شود. در ادامه تمام کد ها گزاش ها و مدل های آموزش داده شده خدمت شرکت تسلیم شد و بر روی ریپازیتوری کارآموزش قرار گرفت.

بعد از اتمام این تجربه دلنشین و بسیار آموزنده کارآموزی نامه های مربوط به گذراندن ۲۴۰ ساعت کارآموزی با امضای مدیرعامل شرکت دریافت شد و یک لوح تقدیر به من داده شد. تمام این مدارک در انتهای این گزارش کارآموزی پیوست شده است.

فصل پنجم

جمع‌بندی و نتیجه‌گیری و پیشنهادات

در این فصل با توجه مباحثی که در فصل های قبل بیان شد و تجارب ارزشمندی که کسب شد به نتیجه‌گیری و پیشنهادات این پروژه کارآموزی اشاره خواهد شد.

۵-۱ نتیجه‌گیر و جمع‌بندی

همانطور که در فصل های گذشته بیان شد مدل های بازشناسی گفتار امروز کاربرد های زیادی دارند اما با اینکه چندین سرویس بازشناسی گفتار فارسی موجود است امّا هیچ کدام از آنها از مدل های جدید خودنگرش[1] استفاده نمی‌کنند؛ این امر باعث شده است در بعضی موارد ضعیف عمل کنند. من در این دوره کارآموزی با راهنمایی ارشد پروژه و کمک های استاد راهنمای کارآموزی دکتر سیدین و سرپرست کارآموزی دکتر بیراوند موفق و شدم یک مدل بازشناسی گفتار فارسی با دقت بسیار بالا درست کنم.

در نهایت این سرویس بازشناسی گفتار فارسی بر روی سرور های شرکت و سرور های هاگینگ فیس پیاده سازی شد و در حال حاظر نسخه اولیه آن در دسترس عموم می‌باشد. این مدل جدید توانایی درک کلمات جدید و پیچیده را دارد و می‌تواند به درستی رو نویسی کند؛ امکانی که در مدل های قبلی فراهم نبود.

با این حال برای رسیدن به نسخه نهایی و استفاده در صنعت نیازمند استفاده دیتاست های بزرگتر و با دامنه گفتار گسترده تر می‌باشد. همچنین استفاده از مدل زبانی بزرگ هم میتواند در بهبود عملکرد این مدل نقش به سزایی ایفا کند.

۵-۲ پیشنهادات

همان طور پیشتر گفته شد تعداد بیشتر دادگان می‌تواند باعث افزایش هرچه بیشتر دقت سرویس شود. به این منظور استفاده از داده های گفتاری که شامل لهجه های مختلف، صدا های پس زمینه مختلف و گوینده هایی با بازه سنی بیشتر است می‌تواند دقت سرویس بازشناسی گفتار فارسی را بسیار افزایش دهد.

پیشنهاد دیگری که می‌توان برای این پروژه ارائه کرد استفاده مدل هایی با معماری بزرگتر و پارامتر های بیشتر می‌باشد. برای این پروژه بخاطر محدودیت در امکانات محسباتی من مجبور شدم که از معماری متوسط ای-برانچفرمر متوسط استفاده کنم که بتوان با وجود حافظه کم واحد پردازش گرافیکی

[1] Self-attention

مدل را آموزش دهم. اما اگر سخت افزار های قوی تر با توان محاسباتی و حافظه بیشتر موجود باشد می‌توان مدل ای-برانچفرمر بزرگ را آموزش داد و دقت خروجی را بالا برد.

به عنوان آخرین پیشنهاد می‌توان استفاده مدل های زبانی بزرگ[۲] را مطرح کرد. تحقیقات جدید نشان می‌دهند که می‌توان یک مدل زبانی بزرگ را بر روی حجم زیادی از دادگان آموزش داد و از آن برای تمام کار های پردازش گفتار و متن استفاده کرد و نتایج خوبی دریافت کرد.[۱۰] باتوجه به اینکه در بازشناسی گفتار از مدل زبانی و مدل صوت شناسی کنار هم استفاده می‌شود بنظر می‌آید که مدل زبانی قوی می‌تواند ضعف مدل صوت شناسی را جبران کند و خروجی هایی با دقت بالا تولید کند.

[2]Large Language Model

منابع و مراجع

[1] Gulati, Anmol, Qin, James, Chiu, Chung-Cheng, Parmar, Niki, Zhang, Yu, Yu, Jiahui, Han, Wei, Wang, Shibo, Zhang, Zhengdong, Wu, Yonghui, and Pang, Ruoming. Conformer: Convolution-augmented transformer for speech recognition, 2020.

[2] Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N., Kaiser, Lukasz, and Polosukhin, Illia. Attention is all you need. CoRR, abs/1706.03762, 2017.

[3] Kim, Kwangyoun, Wu, Felix, Peng, Yifan, Pan, Jing, Sridhar, Prashant, Han, Kyu J., and Watanabe, Shinji. E-branchformer: Branchformer with enhanced merging for speech recognition, 2022.

[4] Peng, Yifan, Kim, Kwangyoun, Wu, Felix, Yan, Brian, Arora, Siddhant, Chen, William, Tang, Jiyang, Shon, Suwon, Sridhar, Prashant, and Watanabe, Shinji. A comparative study on e-branchformer vs conformer in speech recognition, translation, and understanding tasks, 2023.

[5] Sabouri, Sadra, Rahmati, Elnaz, Gooran, Soroush, and Sameti, Hossein. naab: A ready-to-use plug-and-play corpus for farsi. arXiv preprint arXiv:2208.13486, 2022.

[6] Watanabe, Shinji, Hori, Takaaki, Karita, Shigeki, Hayashi, Tomoki, Nishitoba, Jiro, Unno, Yuya, Soplin, Nelson Enrique Yalta, Heymann, Jahn, Wiesner, Matthew, Chen, Nanxin, Renduchintala, Adithya, and Ochiai, Tsubasa. Espnet: End-to-end speech processing toolkit, 2018.

[7] Abadi, Martin, Agarwal, Ashish, Barham, Paul, Brevdo, Eugene, Chen, Zhifeng, Citro, Craig, Corrado, Greg S., Davis, Andy, Dean, Jeffrey, Devin, Matthieu, Ghemawat, Sanjay, Goodfellow, Ian, Harp, Andrew, Irving, Geoffrey, Isard, Michael, Jia, Yangqing, Jozefowicz, Rafal, Kaiser, Lukasz, Kudlur, Manjunath, Levenberg, Josh, Mane, Dandelion, Monga, Rajat, Moore, Sherry, Murray, Derek, Olah, Chris, Schuster, Mike, Shlens, Jonathon, Steiner, Benoit, Sutskever, Ilya, Talwar, Kunal, Tucker, Paul, Vanhoucke, Vincent, Vasudevan, Vijay, Viegas, Fernanda, Vinyals, Oriol, Warden, Pete, Wattenberg, Martin, Wicke, Martin, Yu, Yuan, and Zheng, Xiaoqiang. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[8] Abid, Abubakar, Abdalla, Ali, Abid, Ali, Khan, Dawood, Alfozan, Abdulrahman, and Zou, James. Gradio: Hassle-free sharing and testing of ml models in the wild. arXiv preprint arXiv:1906.02569, 2019.

[9] Wolf, Thomas, Debut, Lysandre, Sanh, Victor, Chaumond, Julien, Delangue, Clement, Moi, Anthony, Cistac, Pierric, Rault, Tim, Louf, R mi, Funtowicz, Morgan, Davison, Joe, Shleifer, Sam, von Platen, Patrick, Ma, Clara, Jernite, Yacine, Plu, Julien, Xu, Canwen, Scao, Teven Le, Gugger, Sylvain, Drame, Mariama, Lhoest, Quentin, and Rush, Alexander M. Huggingface's transformers: State-of-the-art natural language processing, 2020.

[10] Huang, Shaohan, Dong, Li, Wang, Wenhui, Hao, Yaru, Singhal, Saksham, Ma, Shuming, Lv, Tengchao, Cui, Lei, Mohammed, Owais Khan, Patra, Barun, Liu,

Qiang, Aggarwal, Kriti, Chi, Zewen, Bjorck, Johan, Chaudhary, Vishrav, Som, Subhojit, Song, Xia, and Wei, Furu. Language is not all you need: Aligning perception with language models, 2023.

پیوست

در این بخش نامه های مربوط کارآموزی پیوست شده است.

بسمه تعالی

تاریخ: ...
شماره: ...
پیوست: ...

دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

۰۲/ب/۱۲۷۴

ریاست محترم شرکت عصر گویش پرداز

با سلام و احترام

بدینوسیله پارسا محمدی (۹۹۲۳۱۲۱) دانشجوی مهندسی الکترونیک جهت گذراندن کارآموزی ۱ به مدت ۲۴۰ ساعت معرفی میگردند. ضمن تشکر از همکاری آن مدیریت در ارتقاء توانایی علمی- اجرایی مهندسین آینده این کشور، خواهشمند است مراتب موافقت خود را کتباً اعلام فرمائید.

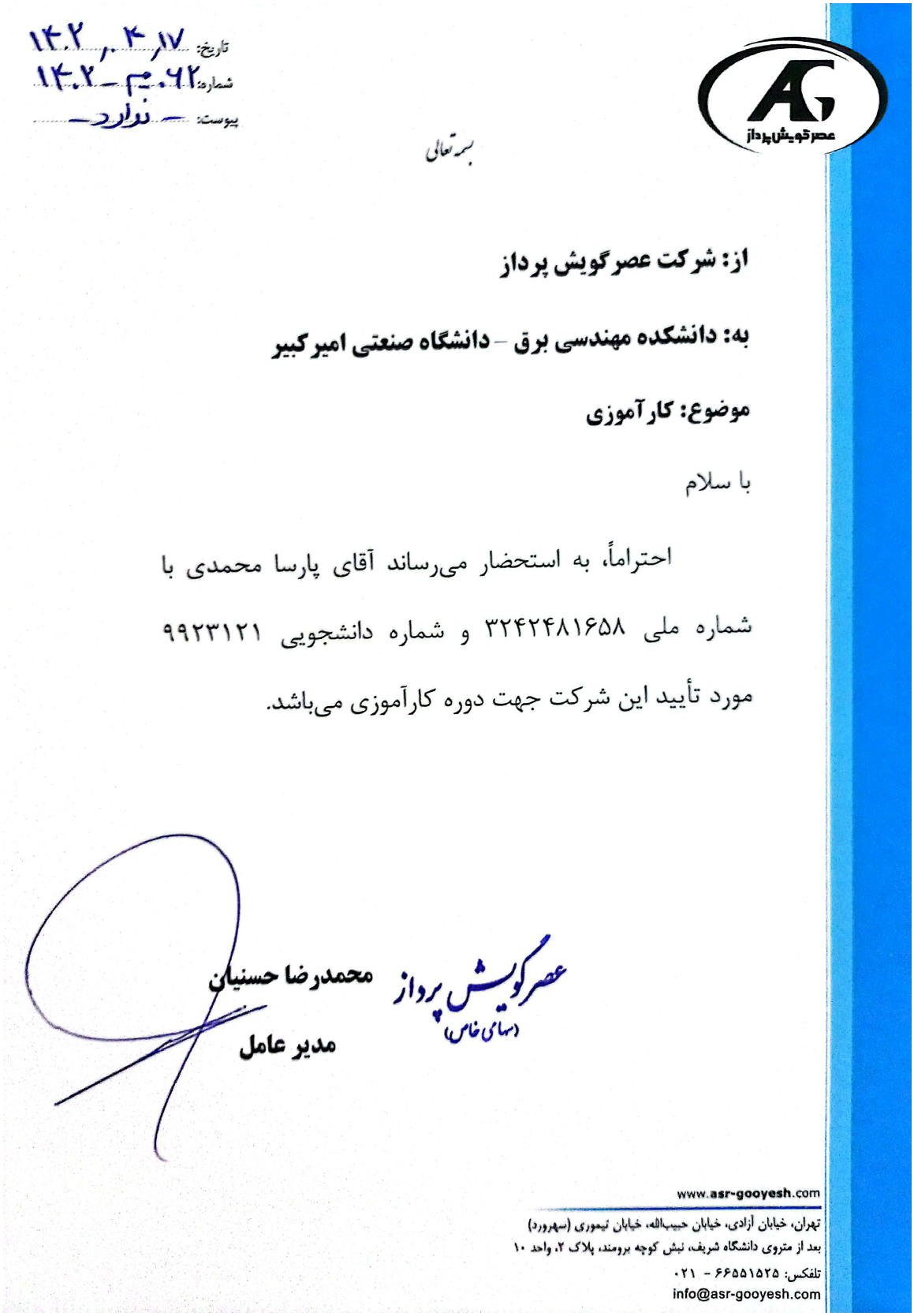
با تشکر
غلامحسین رباحی دهکردی
معاون تحصیلات تکمیلی و پژوهشی دانشکده مهندسی برق

نمابر: ۶۶۴۰۶۴۶۹
پست الکترونیک: beyranvand@aut.ac.ir

استاد کارآموزی
دکتر سیدین ساناز

تهران، خیابان حافظ، روبروی خیابان سمیه، شماره ۴۲۴، صندوق پستی ۴۴۱۳-۱۵۸۷۵
تلفن: ۶۴۵۴۰۰-۱ http://www.aut.ac.ir

شکل ۱: نامه ارسال شده توسط دکتر ریاحی به صنعت جهت معرفی کارآموز به شرکت



تاریخ: ۱۴۰۲/۰۴/۱۷
شماره: ۱۴۰۲-م.۰۶۲
پیوست: ندارد

بسمه تعالی

از: شرکت عصرگویش پرداز

به: دانشکده مهندسی برق – دانشگاه صنعتی امیرکبیر

موضوع: کارآموزی

با سلام

احتراماً، به استحضار می‌رساند آقای پارسا محمدی با شماره ملی ۳۲۴۲۴۸۱۶۵۸ و شماره دانشجویی ۹۹۲۳۱۲۱ مورد تأیید این شرکت جهت دوره کارآموزی می‌باشد.

عصرگویش پرداز
(سهامی خاص)

محمدرضا حسنیان
مدیر عامل

www.asr-gooyesh.com
تهران، خیابان آزادی، خیابان حبیب‌الله، خیابان تیموری (سهروردی)
بعد از متروی دانشگاه شریف، نبش کوچه برومند، پلاک ۲، واحد ۱۰
تلفکس: ۶۶۵۵۱۵۲۵ - ۰۲۱
info@asr-gooyesh.com

شکل ۲: پاسخ شرکت به نامه دکتر ریاحی و علام موافقت شرکت

فرم شماره ۰۳-۲۳۰۳-AUT-PR

بسمه تعالی

شماره: _____

تاریخ: ۱۴۰۲/۰۴/۱۷

تاییدیه صنعت مبنی بر پذیرش کارآموز

دانشگاه صنعتی امیرکبیر

مشخصات دانشجو:

نام و نام خانوادگی: پارسا محمدی شماره دانشجویی: 9923121

رشته تحصیلی: مهندسی برق میزان ساعت کارآموزی: 240 ساعت

مشخصات محل کارآموزی:

نام محل کارآموزی: شرکت عصرگویش پرواز کد:

آدرس: تهران- خیابان آزادی- خیابان حبیب الهی وب سایت: asr-gooyesh.com

بلوار تیموری- نبش کوچه برومند- پلاک ۲- واحد ۱۰

بدینوسیله به اطلاع می‌رساند این شرکت/سازمان با انجام کارآموزی آقا/خانم پارسا محمدی به شماره دانشجویی 9923121 به مدت ۲۴۰ ساعت موافقت می‌نماید. در پایان دوره، تاییدیه و فرم ارزشیابی به دانشگاه ارسال می‌گردد.

نام و نام خانوادگی ریاست آموزش در صنعت: [illegible]

شماره تلفن: ۶۱۹۴۱۰۰۰ شماره فاکس: ۶۱۹۴۱۰۰۰ پست الکترونیکی: info@asr-gooyesh.com

مهر و امضا تاریخ: ۱۴۰۲/۰۴/۱۷

عصرگویش پرواز (سهامی خاص)

شکل ۳: فرم تایید صنعت مبنی بر پذیرش کارآموز

بسمه تعالی

تاریخ :

شماره :

ارزشیابی صنعت از دانشجو

دانشگاه صنعتی امیر کبیر
(پلی تکنیک تهران)

معاونت پژوهشی
(فرم -603)

مشخصات دانشجو :

شرح مختصر :

نام و نام خانوادگی : پارسا محمدی شماره دانشجویی: 9923121 رشته تحصیلی : مهندسی برق

تاریخ شروع کارآموزی : 1402.04.17 تاریخ خاتمه کارآموزی : 1402.05.31 محل کارآموزی وابسته به

موضوع کارآموزی : هوش مصنوعی - توسعه مدل ASR

ردیف	خصوصیات کارآموز	عالی (4 امتیاز)	خوب (3 امتیاز)	متوسط (2 امتیاز)	ضعیف (1 امتیاز)
1	استعداد و قدرت فراگیری	۴			
2	نحوه انجام کار و میزان علاقه	۴			
3	پیگیری وظایف و میزان پشتکار	۴			
4	ارزش پیشنهادات کارآموزی جهت بهبود کار	۴			
5	حضور به موقع و احترام به قوانین جاری محیط کار	۴			-
6	نحوه رفتار اجتماعی و رعایت رفتار متین و احترام آمیز با دیگران در محیط کار	۴			

نظر کلی سرپرست کارآموزی راجع به نکات برجسته و ضعیف و کیفیت کارکارآموزی

۴ تای پارسا محمدی از ۶ کارآموز خوب و برتر ما بودن دوره بودند. ایشان روی موضوع آموزش مدل های باز شناسی گفتار کار کردند و نتایج بسیار خوبی گرفتند، کدها و مدل ها را ایشان تحویل داده شده است

نام ونام خانوادگی سرپرست کارآموز در صنعت : سروش گودرزی سمت :

شماره تلفن : ۶۶۹۶۱۰۰۰ شماره نمابر :

عصر گویش پرواز (سهامی خاص)

مهر و امضاء تاریخ :

نام و نام خانوادگی ریاست آموزش در صنعت : شماره تلفن : ۶۶۹۶۱۰۰۰

شماره نمابر : ۶۶۹۶۱۰۰۰

عصر گویش پرواز (سهامی خاص)

مهر امضاء تاریخ : ۱۴۰۲/۹/۱

شکل ۴: فرم ارزیابی صنعت از دانشجو

تاریخ: ۱۴۰۲/۰۵/۳۱
شماره: ۱۴۰۲-۰۵-۰۸۵
پیوست: ندارد

بسمه تعالی

از: شرکت عصرگویش پرداز

به: دانشکده مهندسی برق – دانشگاه صنعتی امیرکبیر

موضوع: کارآموزی

با سلام

احتراماً، به استحضار می‌رساند آقای پارسا محمدی با شماره ملی ۳۲۴۲۴۸۱۶۵۸ و شماره دانشجویی ۹۹۲۳۱۲۱ از تاریخ ۱۴۰۲/۰۴/۱۰ لغایت ۱۴۰۲/۰۵/۳۰ به مدت ۲۴۰ ساعت در بخش هوش مصنوعی کارآموزی خود را در این شرکت با موفقیت گذرانده است.

محمدرضا حسنیان
مدیر عامل

عصرگویش پرداز
(سهامی خاص)

www.asr-gooyesh.com
تهران، خیابان آزادی، خیابان حبیب‌الله، خیابان تیموری (سهروردی)
بعد از متروی دانشگاه شریف، نبش کوچه برومند، پلاک ۲، واحد ۱۰
تلفکس: ۶۶۵۵۱۵۲۵ - ۰۲۱
info@asr-gooyesh.com

شکل ۵: نامه پایان ۲۴۰ ساعت کارآموزی در شرکت عصر گویش پرداز


گواهی دوره کارآموزی

بدین‌وسیله گواهی می‌شود :

آقای / خانم پارسا محمدی به شماره‌ملی ۳۲۴۲۴۸۱۶۵۸

به مدت ۲۹۵ ساعت دوره کارآموزی خود را با موضوع مرتبط با سامانه‌های هوش مصنوعی

در مرکز نوآوری این شرکت از تاریخ ۱۴۰۲/۴/۱۷ تا تاریخ ۱۴۰۲/۵/۳۱ با موفقیت گذرانده است .

مدیرعامل

محمدرضا حسینیان

شکل ۶: گواهی شرکت در دوره کارآموزی شرکت عصر گویش پرداز

واژه‌نامه‌ی انگلیسی به فارسی

A

Automatic Speech Recognition بازشناسی خودکار گفتار

Acoustics . . . . . صوت شناسی

B

Batch . . . . . دسته

C

Convolutional Neural Network . . شبکه‌های عصبی پیچشی

Connectionist Temporal Classification طبقه بندی زمانی ارتباط گرایانه

Character Error Rate . . نرخ خطای حرف

D

Decoder . . . . . رمز گشا

E

End to End . . . . . سر به سر، سرهم پیوسته

Encoder . . . . . رمز گذار

Enhanced . . . . . تقویت شده

G

Global Optima . . . . نتقطه بهینه سراسری

Global . . . . . جهانی، سراسر

L

Local . . . . . محلی

Large Language Model . مدل زبانی بزرگ

M

Mini-Batch . . . . . کوچک دسته

Map . . . . . نگاشت

N

بازشناسی گفتار . . . . Speech Recognition

خود نگرش . . . . . Self Attention

نیمه نظارتی . . . . . Semi-Supervised

دنباله . . . . . Sequence

T

جعبه ابزار . . . . . Toolkit

مجموعه نماد ها، نشان ها . . . . . Token List

V

ارزیابی . . . . . Validation

W

نرخ خطای کلمه . . . . . Word Error Rate

تعبیه کلمات . . . . . Word Embedding

پردازش زبان طبیعی . . Natural Language Processing

O

بر خط . . . . . Online

P

خط لوله . . . . . Pipeline

R

شبکه عصبی مکرر . . . . Recurrent Neural Network

تحقیق و توسعه . . . . . Research and Development

دستور عمل . . . . . Recipe

S