

بسم الله الرحمن الرحيم

گزارش پنجم پروژه "استخراج کلمات
کلیدی از منابع صوتی و ویدئویی
با استفاده از هوش مصنوعی"
فاطمه وحیدیونسی

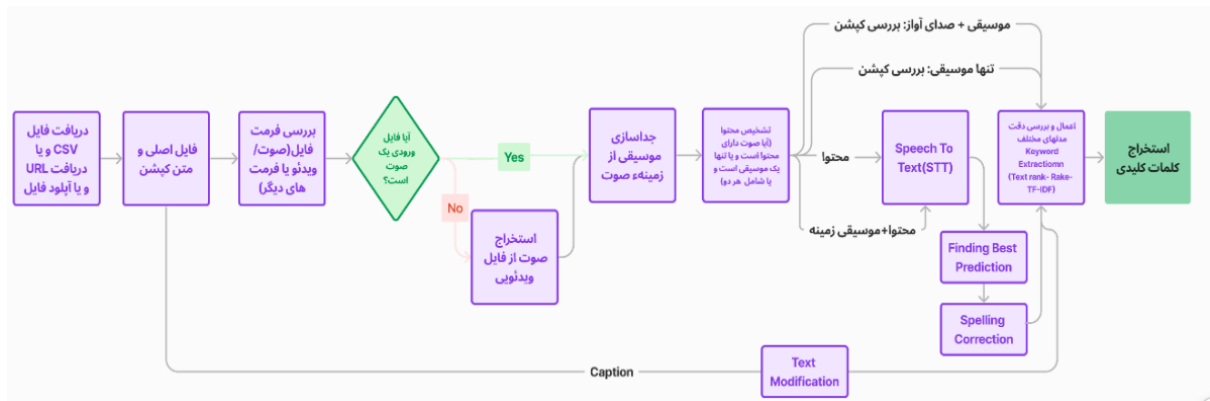


مرکز نوآوری علوم و فناوری های شناختی
دانشگاه علم و صنعت ایران

تاریخ تدوین: 1403/2/29

❖ تشریح پیشرفت پروژه:

همانطور که در گزارش چهارم، مورخ 23 فروردین ماه و جلسه ارائه پنجم اردیبهشت ماه بیان شد، تاکنون یک نمونه اولیه از تمامی مراحل مربوط به فرآیند استخراج کلمات کلیدی از منابع ویدئویی و صوتی انجام شده و اکنون به دنبال افزایش سرعت و دقت مراحل مختلف هستیم. چند راه حل جهت نیل به این اهداف در گزارش قبل بیان شد که در این مدت مورد بررسی قرار گرفته و نتایج آن در این گزارش بیان می شود.



تصویر 1- نمودار مراحل مختلف پروژه "استخراج کلمات کلیدی از منابع ویدئویی و صوتی"

*** تست دقت و سرعت اجرای تمامی مراحل پروژه بر روی چند نمونه محتوای ویدئویی (از منبع اینستاگرام) و صوتی (از منبع تلگرام) انجام شده و نتایج خروجی آنها در قالب یک فایل Google Sheet در اختیار قرار گرفته است.

❖ راه حل های پیشنهاد شده جهت افزایش سرعت و دقت مدل، در گزارش قبل:

جهت افزایش دقت:

1. بررسی تغییر تنظیمات مدل "کاهش نویز" بر روی دقت خروجی مدل STT
2. پیاده سازی مدل استخراج کلمات کلیدی

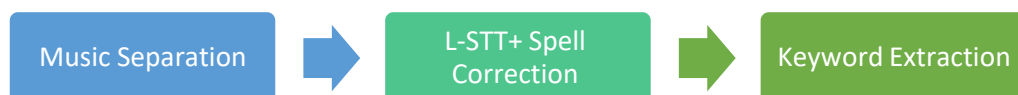
جهت افزایش سرعت اجرا:

1. بررسی تاثیر کاهش حجم فایل و یا فشرده سازی محتوا (افزایش سرعت صوت) بر روی زمان اجرای هر یک از مراحل. درمورد مرحله تبدیل صوت به متن (STT) باید بررسی شود که این عملیات باعث کاهش کیفیت و دقت این مرحله نگردد.
2. بررسی تکنیک های اکوستیکی جهت افزایش سرعت مرحله "تشخیص محتوا"
3. حذف فرآیندهای کم تاثیر روی دقت (مانند ادغام مدل های STT). همچنین درمورد مدیاهایی با محتوای مشخص (مانند یک کانال سخنرانی در تلگرام) مرحله جداسازی و تشخیص محتوای موسیقایی لازم نبوده و قابل حذف است. به مثال زیر توجه کنید، با توجه به تفاوت محتوا در شبکه های اجتماعی و مدیاهای مختلف، معماری زیر پیشنهاد می شود:

حالت اول (بررسی محتوای اینستاگرام/آپارات: محتوای دارای caption):



حالت دوم (بررسی محتوای یک کانال سخنرانی در تلگرام، محتوای صوتی و فاقد caption):



همانگونه که در نمودار بالا نمایش داده شده است، بسته به نوع محتوا و سورس آن، می توان برخی مراحل اجرا را حذف نمود و سرعت اجرا را بالاتر برد.

4. برای افزایش سرعت اجرا، می بایستی سخت افزار (GPU) قوی تری را به کار برد. (متعاقب آن، امکان تست بر روی مجموعه دادگان بیشتر را نیز ممکن می سازد).

❖ نتیجه بررسی راه حل های توسعه ای فوق:

جهت افزایش دقت:

تاثیر مدل "کاهش نویز" بر روی دقت خروجی مدل STT مودر بررسی قرار گرفت. جهت کاهش نویز صوت، از مدل VoiceFixer استفاده میکنیم که تنظیم میزان تاثیر مدل بر روی صدا در آن تعبیه نشده است. بنابراین تابعی جهت این مهم کدنویسی شد که در آن از ادغام صوت اصلی (Original- قبل از Denoising) و صوت Denoise شده، با ضریب بتا، می توان تاثیر مدل VoiceFixer را بر روی روند Denoising مدیریت کرد. تست های انجام شده بر روی خروجی این تابع نشان داد که صوت حاصل از ادغام افزایش 12 دسی بل تراز صوتی مربوط به صوت اصلی و صوت Denoise شده دقت فرآیند STT را تا 10٪ در تشخیص کلمات افزایش می دهد.

نکته اینکه اجرای فرآیند Denoising خود زمان بر است و حدود نصف زمان اجرای فرآیند STT نیاز به زمان دارد. بنابراین توصیه می شود که درمورد صوت هایی که از قبل پیش بینی می شود که دارای نویز باشند (مانند صوت های ضبطی قدیمی- سخنرانی ها و صداهای ضبط شده در مکان های شلوغ) این مرحله اجرا شده و درمورد صوت های دیگر که معمولاً فاقد نویز هستند، نیازی به اجرای این مرحله نیست.

همچنین تا لحظه نگارش این گزارش، پژوهشی در Literature جهت ساخت مدلی بدیع از استخراج کلمات کلیدی با دقتی بیشتر از مدل فعلی انجام شده و یک مدل Hybrid با استفاده از مفاهیم TF و Word2vec در حال پیاده سازی است.

مقالات مرور شده جهت این امر در انتهای این گزارش لیست شده اند.

جهت افزایش سرعت اجرا:

مرحله تشخیص محتوا: جهت افزایش سرعت این مرحله، با بررسی پارامترهای آکوستیکی مختلف، الگوریتم جدیدی جهت تشخیص طول زمانی موسیقی در صوت، کدنویسی شد که سرعت تشخیص محتوای فایل های دارای موسیقی زمینه را حداقل 80٪ درصد افزایش می دهد و حتی دقت را نیز بهبود می دهد.

مرحله STT: بررسی تاثیر کاهش حجم فایل و یا فشرده سازی محتوا (افزایش سرعت صوت) بر روی زمان اجرای مرحله STT نشان داد که افزایش سرعت صوت، از آنجا که بر روی کاهش حجم نیز موثر است باعث افزایش سرعت مرحله STT می شود اما به طور میانگین دقت این مرحله را تا 15٪ کاهش می دهد باتوجه به اهمیت این مرحله و تاثیر دقت آن بر روی مراحل بعدی، بنابراین نهایتاً فشرده سازی و افزایش سرعت صوت برای این مرحله پیشنهاد نمی شود.

مرحله تبدیل ویدئو به صوت: درمورد مرحله تبدیل ویدئو به صوت، افزایش سرعت و کاهش حجم، باعث کاهش زمان تبدیل به صوت می شود اما خود عملیات افزایش سرعت و Compressing، زمان بر است و پیشنهاد نمی شود.

مرحله Music Separation: جهت افزایش سرعت این مرحله راه حل های زیر مورد بررسی قرار گرفت:

1. به جز مدل Demucs مدل Spleeter قابل توصیه است. این مدل در صوت های کوتاه دقتی مشابه Demucs دارد و زمان اجرایی برابری می طلبد اما در صوت های بلند، میانگین 35 درصد سریعتر عمل می کند در مقابل دقتی 5 درصد کمتر در خروجی STT (به علت تغییر در کیفیت Vocal) ایجاد میکند که در پروژه هایی که سرعت دارای اهمیت است، این میزان کاهش دقت در STT قابل چشم پوشی است. Spleeter به جای جداسازی سه ساز موسیقیایی، دو فایل vocal و موسیقی را در خروجی می دهد که نیاز به پردازش کمتر دارد. همچنین نیاز به عملیات ادغام مجدد فایل های موسیقیایی پس از این مرحله را از بین می برد. در نتیجه سرعتی بالاتر نسبت به Demucs ارائه می دهد.
2. درمورد محتوایی که از قبل می دانیم محتوای آنها یا فاقد موسیقی است و یا دارای موسیقی "زمینه و کوتاه" است (مانند محتوای یک کانال سخنرانی در تلگرام) می توان از این مرحله عبور کرد و اینگونه سرعت اجرا را افزایش داد.

❖ پیشبرد و توسعه آتی در پروژه:

تاکنون این پروژه به یک نمونه اولیه قابل اجرا دست یافته است، جهت توسعه بیشتر (بهبود بیشتر دقت و سرعت اجرا) مراحل زیر پیشنهاد می شود:

- 1- بررسی خروجی و نتایج مدل Fine-Tune شده STT
- 2- بررسی امکان افزایش بیشتر سرعت مرحله تشخیص محتوا با بررسی پارامترهای آکوستیکی دیگر
- 3- پیاده سازی مدل جدید جهت افزایش دقت مرحله Keyword Extraction
- 4- اضافه کردن حالت پنجم "محتوا فاقد صوت" برای ویدئو هایی که سایلنت هستند در مرحله تشخیص محتوا.

مقالات مروری جهت پیاده سازی مدل استخراج کلمات کلیدی:

- Comparison of the Performance of approaches in discovering and extracting e-book topics, Fatemeh Zarmehr, 2022
- Improved Keyword Extraction for persian texts using Rake algorithm, Elaheh Mehrabi, 2020
- Keyphrase Extraction from persian texts: A review of the Literature, Atefeh Kalantari, 2020
- An approach for extraction of keywords and weighting words for improvement farsi documents classification, Vahideh Rezaei, 2018
- Keyword Extraction a review of methods and approaches, Slobodan Belgia
- Applying Deep Learning for Arabic Keyphrase extraction, Muhammad Helmy, 2018
- بهبود دقت واژگان کلیدی استخراج شده از متن فارسی با استفاده از الگوریتم word2vec. محمدرضا حسینی آهنگر. 1400
- استخراج خودکار کلمات کلیدی از متون کوتاه فارسی با استفاده از word2vec. امید حاجی زاده. 1398