



دانشکده مهندسی کامپیوتر

استفاده از یادگیری عمیق در تشخیص تکنیک‌های متقاعدسازی به کاررفته در میم‌ها

پروژه برای دریافت درجه کارشناسی

در رشته مهندسی کامپیوتر

مهدیه نادری

استاد راهنما:

دکتر سید صالح اعتمادی

مهر ۱۴۰۳

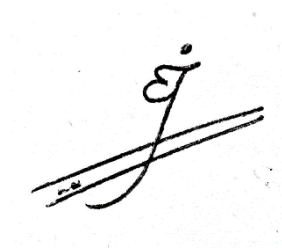
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

تأییدیه‌ی صحت و اصالت نتایج

باسمه تعالی

اینجانب مهدیه نادری به شماره دانشجویی ۹۸۵۲۲۰۷۶ دانشجوی رشته مهندسی کامپیوتر مقطع تحصیلی کارشناسی تأیید می‌نمایم که کلیه‌ی نتایج این پایان نامه حاصل کار اینجانب و بدون هرگونه دخل و تصرف است و موارد نسخه‌برداری شده از آثار دیگران را با ذکر کامل مشخصات منبع ذکر کرده‌ام. در صورت اثبات خلاف مندرجات فوق، به تشخیص دانشگاه مطابق با ضوابط و مقررات حاکم (قانون حمایت از حقوق مؤلفان و مصنفان و قانون ترجمه و تکثیر کتب و نشریات و آثار صوتی، ضوابط و مقررات آموزشی، پژوهشی و انضباطی ...) با اینجانب رفتار خواهد شد و حق هرگونه اعتراض در خصوص احقاق حقوق مکتسب و تشخیص و تعیین تخلف و مجازات را از خویش سلب می‌نمایم. در ضمن، مسئولیت هرگونه پاسخگویی به اشخاص اعم از حقیقی و حقوقی و مراجع ذیصلاح (اعم از اداری و قضایی) به عهده‌ی اینجانب خواهد بود و دانشگاه هیچ‌گونه مسئولیتی در این خصوص نخواهد داشت.

نام و نام خانوادگی: مهدیه نادری



امضا و تاریخ:

مجوز بهره‌برداری از پایان‌نامه

بهره‌برداری از این پایان‌نامه در چهارچوب مقررات کتابخانه و با توجه به محدودیتی که توسط استاد راهنما به شرح زیر تعیین می‌شود، بلامانع است:

- ☐ بهره‌برداری از این پایان‌نامه/ رساله برای همگان بلامانع است.
- ☐ بهره‌برداری از این پایان‌نامه/ رساله با اخذ مجوز از استاد راهنما، بلامانع است.
- ☐ بهره‌برداری از این پایان‌نامه/ رساله تا تاریخ ممنوع است.

نام استاد یا اساتید راهنما:

تاریخ:

امضا:

تقدیم به:

با قلبی سرشار از عشق و قدردانی، این پایان نامه را به پدر و مادر عزیزم تقدیم می کنم. شما تنها پناه من در تمامی فراز و نشیب های زندگی بودید، هر آنچه هست از برکت فداکاری های بی دریغ و محبت های بی پایان شماست. هر قدمی که در این مسیر برداشتم، با یاد شما و برای سربلندی تان بوده است. این پایان نامه همچنین تقدیم می شود به تمامی کسانی که در مسیر علم، با عشق و از خودگذشتگی گام نهادند؛ به همه ی آموزگارانی که شعله های دانش را زنده نگه داشتند و به آنان که در مسیر حقیقت و آگاهی، بی ادعا کوشیدند.

تشکر و قدردانی:

بدین وسیله مراتب سپاس و قدردانی خود را به تمامی کسانی که در این مسیر همراه و یاریگر من بوده‌اند، ابراز می‌دارم.

با کمال احترام و سپاس از اساتید ارجمندم، به ویژه جناب آقای دکتر سید صالح اعتمادی، که با دانش، دلسوزی و راهنمایی‌های ارزشمندشان، نقشی بی‌بدیل در این مسیر داشتند.

از خانم غزل زمانی‌نژاد و خانم فاطمه‌زهررا بخشنده نیز به خاطر همراهی‌های بی‌دریغ و حمایت‌های ارزشمندشان در طول انجام این تحقیق نهایت سپاس را دارم. حضور و همراهی شما در لحظات دشوار، نیروی مضاعفی به من بخشید.

چکیده

در این پایان‌نامه، به بررسی تکنیک‌های متقاعدسازی در میم‌ها پرداخته شده‌است. این پژوهش در چارچوب شرکت تیم CVcoders در تسک فرعی ۱ و ۲ از تسک ۴ مسابقه SemEval 2024 انجام گرفته که موضوع آن شناسایی روش‌های متقاعدسازی روان‌شناختی و بلاغی در محتوای چندزبانه و چندرسانه‌ای است. برای هر دو بخش، از داده‌های ارائه‌شده توسط مسابقه SemEval استفاده شد و به‌منظور بهبود عملکرد مدل‌ها در مواجهه با عدم تعادل طبقات، از تکنیک‌های پیشرفته‌ای مانند Focal Loss بهره بردیم. مدل‌ها تنها با استفاده از داده‌های انگلیسی آموزش داده شدند و در نهایت، با داده‌های به زبان مقدونیه شمالی، بلغاری و انگلیسی آزمایش شدند.

در تسک فرعی ۱، که تنها شامل داده‌های متنی و ۲۰ طبقه‌بندی مختلف بود، از مدل‌های پیش‌آمोخته XLM-RoBERTa و GPT-2 استفاده کردیم. نتایج نشان می‌دهند که مدل GPT-2 طبق معیار ارزیابی Hierarchical F1 در زبان انگلیسی عملکرد بهتری داشته است. در تسک فرعی ۲، داده‌های متنی و تصویری با ۲ کلاس مختلف مورد بررسی قرار گرفتند. برای این منظور، از ترکیب مدل‌های پیش‌آموخته متن و تصویر شامل متون و تصاویر تحلیل شدند تا مشخص شود آیا در هر میم از تکنیک‌های متقاعدسازی استفاده شده است یا خیر. ترکیب مدل‌های VGG و GPT-2 بهترین عملکرد را در این زمینه ارائه داد.

واژه‌های کلیدی: تکنیک‌های متقاعدسازی، پروپاگاندا، پردازش زبان طبیعی، یادگیری عمیق، پردازش تصویر،

Focal Loss, GPT-2, XLM-RoBERTa, SemEval 2024

فهرست مطالب

ط	فهرست شکل‌ها
ي	فهرست جدول‌ها
۱	فصل ۱: مقدمه
۲	۱-۱- تعریف مسئله
۳	۱-۲- اهمیت موضوع
۳	۱-۳- اهداف پژوهش
۴	۱-۴- ساختار پایان‌نامه
۵	فصل ۲: کارهای پیشین و مرتبط
۶	۲-۱- مقدمه
۶	۲-۲- مفاهیم پروپاگاندا و متقاعدسازی
۶	۱-۱- معرفی تسک فرعی ۱
۷	۲-۳- طبقه‌بندی چندبرچسبی سلسله‌مراتبی
۸	۱-۳-۲- مفهوم طبقه‌بندی چند برچسبی
۸	۲-۳-۲- مفهوم سلسله‌مراتب در طبقه‌بندی
۸	۲-۴- معرفی تسک فرعی ۲
۹	۱-۴-۲- مدل‌های چندوجهی
۹	۲-۴-۲- پردازش موازی و فیوژن
۱۰	۲-۵- کارهای پیشین
۱۱	۲-۶- مدل‌های زبانی
۱۲	۱-۶-۲- شبکه عصبی عمیق GPT-2
۱۳	۱-۶-۱-۱- آماده‌سازی ورودی متنی GPT-2
۱۵	۲-۶-۱-۲- دیاگرام معماری GPT-2
۱۶	۲-۶-۲- شبکه عصبی XLM
۱۷	۲-۶-۳- شبکه عصبی XLM-RoBERTa
۱۹	۱-۶-۳-۱- معماری XLM-RoBERTa
۲۰	۲-۷- مدل‌های بینایی عمیق
۲۱	۱-۷-۲- شبکه عصبی VGG-16
۲۱	۱-۷-۱-۱- معماری VGG-16
۲۲	۱-۷-۲- چگونگی استفاده از VGG-16 برای طبقه‌بندی تصویر
۲۳	۱-۷-۱-۳- مزایا و معایب VGG-16
۲۴	۲-۷-۲- شبکه عصبی ViT

۲۴ ViT معماری ۲-۷-۲-۱-
۲۵ ViT نحوه عملکرد ۲-۷-۲-۲-
۲۶ ViT چگونگی استفاده از ViT برای طبقه‌بندی تصاویر ۲-۷-۲-۳-
۲۷ ViT مزایا و معایب ۲-۷-۲-۴-
۲۷ معیار ارزیابی ۲-۸-
۲۸ دقت (Accuracy) ۲-۸-۱-
۲۸ F1-Score ۲-۸-۲-
۳۰ معیار ارزیابی سلسله مراتبی ۲-۸-۳-

۳۲ فصل ۳: روش‌های پیشنهادی

۳۳ مقدمه ۳-۱-
۳۳ جمع آوری داده‌ها ۳-۱-۱-
۳۳ روش‌های پیشنهادی تسک فرعی ۱ ۳-۲-
۳۳ مجموعه داده‌ها ۳-۲-۱-
۳۷ پیش‌پردازش داده‌ها ۳-۲-۲-
۳۷ پردازش متن با استفاده از NLTK ۳-۲-۲-۱-
۳۸ پیاده‌سازی مدل‌ها ۳-۲-۳-
۳۸ چالش‌های روش پیشنهادی ۳-۲-۴-
۳۸ عدم توازن داده‌ها ۳-۲-۴-۱-
۳۹ طبقه‌بندی چندبرچسبی سلسله مراتبی ۳-۲-۴-۲-
۴۰ فرایند آموزش مدل ۳-۲-۵-
۴۰ ارزیابی نتایج ۳-۲-۶-
۴۰ روش‌های پیشنهادی تسک فرعی ۲ ۳-۳-
۴۱ مجموعه داده‌ها ۳-۳-۱-
۴۱ پیش‌پردازش داده‌ها ۳-۳-۲-
۴۲ API OpenAI برای پیش‌پردازش اولیه ۳-۳-۲-۱-
۴۲ پردازش بیشتر متن با استفاده از NLTK ۳-۳-۲-۲-
۴۳ اصلاح دستی داده‌ها ۳-۳-۲-۳-
۴۳ پیش‌پردازش تصویر ۳-۳-۲-۴-
۴۴ استخراج ویژگی‌ها ۳-۳-۳-
۴۴ پیاده‌سازی مدل ۳-۳-۴-
۴۵ چالش‌های روش پیشنهادی ۳-۳-۵-
۴۵ چالش داده‌ها ۳-۳-۵-۱-
۴۶ بیش‌برازش ۳-۳-۵-۲-
۴۶ فرایند آموزش مدل ۳-۳-۶-
۴۶ ارزیابی نتایج ۳-۳-۷-

فصل ۴: نتایج و تفسیر آنها

۴۷

- ۴۸ ۱-۴- نتایج تسک فرعی ۱
- ۴۸ ۱-۱-۴ نتایج
- ۴۹ ۲-۱-۴- تحلیل نتایج
- ۴۹ ۲-۴- نتایج تسک فرعی ۲ب
- ۴۹ ۱-۲-۴- نتایج
- ۵۱ ۲-۲-۴- تحلیل نتایج

فصل ۵: جمع‌بندی و پیشنهادها

۵۲

- ۵۳ ۱-۵- جمع‌بندی
- ۵۳ ۲-۵- پیشنهادها

۵۵

مراجع

فهرست شکل‌ها

شکل (۱-۲) نمونه سلسله مراتب تکنیک‌ها برای تسک فرعی ۲ [۱]	۷
شکل (۲-۲) نمونه استفاده از توکنایزر GPT-2 [۹]	۱۳
شکل (۲-۳) نمونه فرمت قابل قبول داده در GPT-2 [۹]	۱۴
شکل (۴-۲) معماری مدل GPT-2 [۹]	۱۶
شکل (۵-۲) پیش‌آموزش مدل زبانی چندزبانه [۱۳]	۲۰
شکل (۶-۲) نمای کلی مدل ViT [۱۸]	۲۶
شکل (۱-۳) نمودار توزیع داده‌ها در مجموعه داده آموزشی	۳۴
شکل (۲-۳) نمودار توزیع داده‌ها در مجموعه داده اعتبارسنجی	۳۵
شکل (۳-۳) نمودار توزیع داده‌ها در مجموعه داده توسعه	۳۵
شکل (۴-۳) ساختار داده و مراحل پیش پردازش به کاررفته [۲]	۴۴
شکل (۵-۳) معماری مدل ترکیب شده از VGG-16 و GPT-2 برای تسک فرعی ۲ [۲]	۴۵
شکل (۱-۴) دقت، فراخوانی، نمره F-score و F-macro در مقابل آستانه در مجموعه توسعه [۲]	۵۱

فهرست جدول‌ها

جدول (۱-۳) توزیع مجموعه داده‌ها تسک فرعی ۱	۳۹
جدول (۲-۳) توزیع مجموعه داده‌ها تسک فرعی ۲	۴۱
جدول (۱-۴) ابرپارامترهای بهترین مدل در تسک فرعی ۱	۴۸
جدول (۲-۴) نتایج مجموعه تست به زبان انگلیسی در تسک فرعی ۱	۴۸
جدول (۳-۴) نتایج مجموعه تست به زبان‌های بلغاری و مقدونیه شمالی در تسک فرعی ۱	۴۹
جدول (۴-۴) نتایج مجموعه اعتبارسنجی در تسک فرعی ۲	۵۰
جدول (۵-۴) نتایج خروجی بهترین مدل روی مجموعه تست زبان‌های انگلیسی، بلغاری و مقدونیه‌ای در تسک فرعی ۲	۵۰
جدول (۶-۴) ابرپارامترهای بهترین مدل در تسک فرعی ۲	۵۰

فصل ۱:

مقدمه

۱-۱- تعریف مسئله

با گسترش روزافزون شبکه‌های اجتماعی و حجم بالای محتوای تولید شده توسط کاربران، شناسایی و تحلیل روش‌های متقاعدسازی و پروپاگاندا به یکی از چالش‌های اساسی در حوزه‌های تحلیل داده و پردازش زبان طبیعی تبدیل شده است. این روش‌ها، که از استراتژی‌های روانشناختی و بلاغی برای تأثیرگذاری بر افکار عمومی استفاده می‌کنند، نقش عمده‌ای در انتشار اطلاعات نادرست، جهت‌دهی به افکار عمومی، و حتی دستکاری اجتماعی ایفا می‌کنند. از این رو، نیاز به روش‌های خودکار برای تشخیص این تکنیک‌ها در متون و تصاویر، به‌ویژه در قالب میم‌ها، ضروری به نظر می‌رسد [۱].

مسئله اصلی این پژوهش، توسعه روشی است که قادر به شناسایی تکنیک‌های متقاعدسازی در داده‌های چندرسانه‌ای و چندزبانه باشد. داده‌های مورد استفاده شامل متون و تصاویر (میم‌ها) است که در آن‌ها از روش‌های متقاعدسازی و پروپاگاندا استفاده شده است. هدف این پروژه، ارائه مدلی مبتنی بر یادگیری عمیق است که بتواند این تکنیک‌ها را به‌صورت خودکار شناسایی و دسته‌بندی کند.

این پروژه با تمرکز بر دو زیرمسئله تعریف شده در تسک ۴ مسابقه SemEval 2024 انجام شده است. زیرمسئله اول شامل تشخیص تکنیک‌های متقاعدسازی در داده‌های متنی با ۲۰ کلاس مختلف است. زیرمسئله دوم شامل تشخیص این تکنیک‌ها در داده‌های چندرسانه‌ای (شامل متن و تصویر) است که در آن از دو طبقه‌بندی "وجود" یا "عدم وجود" تکنیک متقاعدسازی در هر میم استفاده می‌شود [۱].

با توجه به چالش‌های موجود در تشخیص این تکنیک‌ها و تنوع بالای داده‌های متنی و تصویری، این پژوهش تلاش دارد با استفاده از مدل‌های پیش‌آموزش‌دیده مانند GPT-2 و XLM-RoBERTa و به‌کارگیری تکنیک‌های بهبود عملکرد مانند Focal Loss، راهکاری کارآمد برای حل این مسئله ارائه دهد.

شبکه‌های اجتماعی امروزه به بستری برای انتشار اطلاعات و محتواهایی تبدیل شده‌اند که تأثیر زیادی بر افکار عمومی دارند. یکی از مهم‌ترین چالش‌ها در این فضا، استفاده از تکنیک‌های متقاعدسازی و پروپاگاندا به‌منظور تغییر نگرش کاربران است. این تکنیک‌ها که شامل روش‌های روانشناختی و بلاغی نظیر ساده‌سازی بیش از حد علت و معلول، برچسب‌زنی و تخریب شخصیت هستند، به‌طور خاص در میم‌ها به‌عنوان یکی از پرطرفدارترین انواع محتوا در کمپین‌های اطلاعات نادرست به‌کار می‌روند [۱]. این میم‌ها به‌سرعت در شبکه‌های اجتماعی منتشر شده و از طریق این تکنیک‌ها تأثیر قابل‌توجهی بر مخاطبان می‌گذارند.

مسئله اصلی در این پژوهش، توسعه‌ی مدلی است که بتواند این تکنیک‌های متقاعدسازی را در داده‌های متنی و تصویری شناسایی کند. این تحقیق بخشی از تسک ۴ مسابقه SemEval 2024 را شامل می‌شود که

به بررسی این تکنیک‌ها در دو زیرمسئله می‌پردازد: تسک فرعی ۱ که تنها شامل محتوای متنی است و تسک فرعی ۲ که تحلیل داده‌های چندرسانه‌ای شامل متون و تصاویر را انجام می‌دهد [۱].

۲-۱- اهمیت موضوع

در دوران حاضر، که اطلاعات به راحتی در اینترنت و شبکه‌های اجتماعی منتشر می‌شود، شناسایی تکنیک‌های متقاعدسازی و پروپاگاندا اهمیت ویژه‌ای پیدا کرده است. این تکنیک‌ها نقش عمده‌ای در پخش اطلاعات نادرست و کمپین‌های تبلیغاتی جهت‌دار دارند. از آنجا که میم‌ها به عنوان یکی از انواع محبوب محتوا به سرعت میان کاربران دست به دست می‌شوند، شناخت و مقابله با تأثیرات منفی آن‌ها بر جامعه ضرورت دارد. این پژوهش با هدف شناسایی خودکار این تکنیک‌ها، گامی مؤثر در جهت جلوگیری از انتشار بی‌رویه اطلاعات نادرست برمی‌دارد و به تقویت آگاهی کاربران کمک می‌کند.

۳-۱- اهداف پژوهش

اهداف اصلی این پژوهش عبارتند از:

- توسعه‌ی مدلی برای شناسایی تکنیک‌های متقاعدسازی در داده‌های متنی، با استفاده از مدل‌های پیش‌آموزش‌دیده نظیر GPT-2 و XLM-RoBERTa.
- بررسی و تحلیل تکنیک‌های متقاعدسازی در داده‌های چندرسانه‌ای شامل میم‌ها، که در آن هر دو عنصر تصویری و متنی مورد بررسی قرار می‌گیرند.
- ارزیابی مدل‌های توسعه داده شده در دو زیرمسئله‌ی Subtask 1 و Subtask 2b از مسابقه SemEval2024.
- ارائه روشی کارآمد برای مقابله با عدم تعادل طبقات در مجموعه داده‌ها، با استفاده از روش‌های پیشرفته نظیر Focal Loss.

۴-۱- ساختار پایان‌نامه

این پایان‌نامه در چند فصل به شرح زیر تنظیم شده است:

- فصل دوم: کارهای پیشین و مرتبط
در این فصل به بررسی پژوهش‌های مرتبط با تکنیک‌های متقاعدسازی و پروپاگاندا و مدل‌های یادگیری عمیق در این زمینه پرداخته می‌شود. همچنین مباحثی پیرامون پردازش زبان طبیعی و تحلیل داده‌های چندرسانه‌ای مورد بررسی قرار می‌گیرد.
- فصل سوم: روش‌های پیشنهادی
این فصل شامل توضیحات مربوط به مجموعه داده‌های استفاده‌شده در پروژه، مدل‌های به‌کارگرفته‌شده فرآیندهای پیش‌پردازش داده است. همچنین جزئیات مربوط به ارزیابی مدل‌ها و بهبود عملکرد آن‌ها از طریق روش‌هایی نظیر Focal Loss مطرح خواهد شد.
- فصل چهارم: نتایج و تفسیر آن‌ها
در این فصل به توضیح مراحل پیاده‌سازی مدل‌های پیشنهادی پرداخته شده و نتایج به‌دست‌آمده از ارزیابی مدل‌ها در هر دو زیرمسئله ارائه خواهد شد. همچنین چالش‌ها و مسائل پیش‌آمده در طول پیاده‌سازی مورد بحث قرار می‌گیرد.
- فصل پنجم: جمع‌بندی و پیشنهادات
این فصل به جمع‌بندی نتایج پژوهش پرداخته و پیشنهاداتی برای پژوهش‌های آینده در زمینه‌ی شناسایی تکنیک‌های متقاعدسازی و مقابله با پروپاگاندا ارائه خواهد شد.

فصل ۲:

کارهای پیشین و مرتبط

۱-۲- مقدمه

در این فصل، به بررسی تحقیقات پیشین در زمینه تکنیک‌های متقاعدسازی در میم‌ها و تسک‌های مرتبط با آن می‌پردازیم. ابتدا مفاهیم اولیه و توضیحات تسک‌ها بیان می‌شود. سپس زمینه کاری هر کدام و بعد توابع ضرر و معیارهای ارزیابی شرح داده می‌شوند.

۲-۲- مفاهیم پروپاگاندا و متقاعدسازی

تکنیک‌های متقاعدسازی و پروپاگاندا به عنوان ابزارهایی برای تأثیرگذاری بر افکار و رفتارهای عمومی، در زمینه‌های مختلفی نظیر تبلیغات، سیاست و رسانه‌ها به کار می‌روند. این تکنیک‌ها می‌توانند شامل استدلال‌های منطقی، بازی با احساسات و ایجاد ارتباط عاطفی با مخاطب باشند.

تکنیک‌های متقاعدسازی عمدتاً بر روی مخاطب تأثیر می‌گذارند تا او را به پذیرش یک ایده یا تغییر رفتار ترغیب کنند. از جمله این تکنیک‌ها می‌توان به استدلال‌های منطقی، احساس‌گرایی، برجسب‌زنی، تخریب شخصیت و ساده‌سازی بیش از حد اشاره کرد. هدف این تکنیک‌ها ایجاد ارتباط مؤثر با مخاطب به منظور القای نظر یا تصمیم‌گیری خاص است و در واقع سعی دارند تا منطق و احساسات مخاطب را به نحو کارآمدی به چالش بکشند [۴].

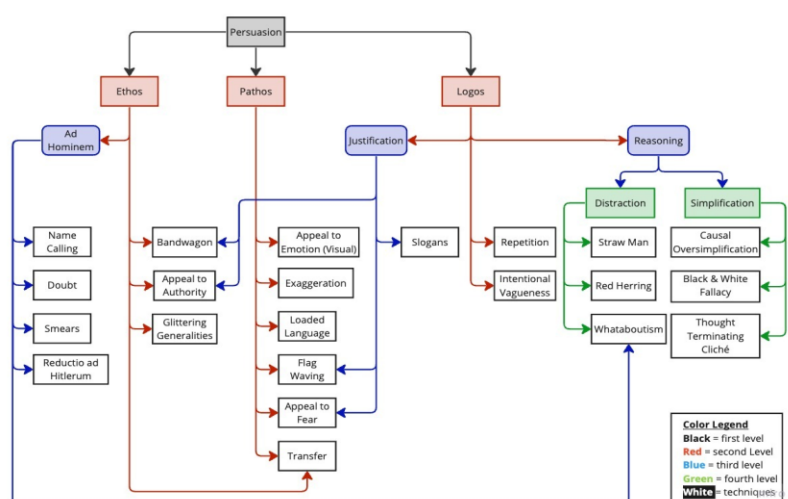
پروپاگاندا به نوعی از اطلاعات اطلاق می‌شود که به طور هدفمند برای ترویج یک ایده یا ایدئولوژی خاص شکل‌گیری می‌شود [۴]. این اطلاعات معمولاً با استفاده از تکنیک‌های روان‌شناختی و بلاغی تهیه می‌شوند تا تأثیر عمیقی بر افکار عمومی بگذارند. پروپاگاندا به‌ویژه در زمینه‌های سیاسی و اجتماعی کاربرد فراوانی دارد و می‌تواند به شکل‌های مختلفی از جمله متون، تصاویر و میم‌ها به کار رود.

۱-۱- معرفی تسک فرعی ۱

در این تسک، هدف شناسایی تکنیک‌های متقاعدسازی موجود در محتوای متنی یک میم است. تنها با در نظر گرفتن «محتوای متنی»، لازم است که تعیین شود کدام یک از ۲۰ تکنیک متقاعدسازی، که به صورت سلسله‌مراتبی سازمان‌دهی شده‌اند، در میم استفاده شده است. در این فرآیند، اگر گره والد یک تکنیک انتخاب

شود، تنها یک پاداش جزئی به دست می‌آید. به عبارت دیگر، این مسئله یک چالش در زمینه طبقه‌بندی چندبرچسبی سلسله‌مراتبی است [۱][۲].

شکل (۲-۱) زیر نمایی از سلسله‌مراتب تکنیک‌های متقاعدسازی را نشان می‌دهد. لازم به ذکر است که در این تصویر ۲۲ تکنیک وجود دارد، اما در تسک فرعی ۱، تکنیک‌های «انتقال»^۱ و «استدلال به احساسات قوی»^۲ لحاظ نشده‌اند. بنابراین، باید سلسله‌مراتب را بدون این دو تکنیک تصور کرد.



شکل (۲-۱) نمونه سلسله‌مراتب تکنیک‌ها برای تسک فرعی ۱^۱

اگر برای حل این تسک به داده‌های برچسب‌گذاری شده اضافی نیاز باشد، می‌توان به مجموعه داده PTC و همچنین داده‌های تسک ۳ مسابقه SemEval 2023 مراجعه کرد.

۳-۲- طبقه‌بندی چندبرچسبی سلسله‌مراتبی

طبقه‌بندی چندبرچسبی سلسله‌مراتبی یکی از رویکردهای پیشرفته در یادگیری ماشین و تحلیل داده‌ها است که به منظور شناسایی و طبقه‌بندی داده‌ها به مجموعه‌ای از برچسب‌ها (برچسب‌های چندگانه) و در عین حال در قالب یک ساختار سلسله‌مراتبی طراحی شده است. این نوع طبقه‌بندی در بسیاری از حوزه‌ها از جمله تحلیل متن، پردازش تصویر و شناسایی تکنیک‌های متقاعدسازی در میم‌ها کاربرد دارد.

^۱ Transfer

^۲ Appeal to Strong

۱-۳-۲- مفهوم طبقه بندی چند برچسبی

در طبقه بندی چندبرچسبی، هر نمونه می تواند به چندین کلاس یا برچسب مختلف تعلق داشته باشد. به عنوان مثال، یک متن می تواند هم به عنوان یک متن آموزشی و هم به عنوان یک متن علمی طبقه بندی شود. این نوع طبقه بندی به خصوص زمانی مهم است که داده ها دارای ویژگی های چندگانه باشند و به سادگی نتوان آن ها را به یک دسته خاص نسبت داد.

۲-۳-۲- مفهوم سلسله مراتب در طبقه بندی

در طبقه بندی سلسله مراتبی، برچسب ها به صورت ساختاری سازمان دهی می شوند و می توانند شامل گره های والد و فرزند باشند. این ساختار به این معنی است که برچسب های خاص می توانند زیرمجموعه ای از برچسب های عمومی تر باشند. به عنوان مثال، برچسب «تکنیک های متقاعدسازی» ممکن است شامل زیرمجموعه های «استدلال منطقی» و «احساس گرایی» باشد. این نوع سازمان دهی می تواند به درک بهتر و ساختارمندتری از داده ها کمک کند و همچنین به الگوریتم ها اجازه می دهد تا روابط پیچیده تری بین برچسب ها را شناسایی کنند.

۴-۲- معرفی تسک فرعی ۲

برای زیرمسئله ۲، هدف شناسایی تکنیک های متقاعدسازی در یک میم است، با استفاده از هر دو محتوای متنی و تصویری. این مسئله یک طبقه بندی دوتایی است، یعنی باید تعیین شود که آیا میم حاوی حداقل یکی از ۲۲ تکنیک متقاعدسازی است یا خیر. برخلاف زیرمسئله ۱، در اینجا سلسله مراتب تکنیک ها تا دو سطح اول از گره ریشه قطع شده است [۱]. رویکرد ما با ترکیب سه تکنیک پیش پردازش پیشرفته و استفاده از مدل های پیشرفته در زمینه پردازش زبان طبیعی و بینایی کامپیوتری، به طور مؤثر این وظیفه پیچیده را انجام می دهد. با الهام از کارهای مرتبط در این زمینه ها، از مدل های پیش آموزش دیده و تکنیک های مدرن بهره می بریم تا دقت و کارایی سیستم خود را بهبود دهیم.

برای انجام این تسک به طور کلی، دو روش اصلی وجود دارد: استفاده از مدل‌های چندوجهی^۱ و پردازش موازی مدل‌های متن و تصویر با ترکیب نتایج آن‌ها در مرحله نهایی.

انتخاب بین این دو روش بستگی به نیازهای خاص پروژه و نوع داده‌هایی که در دسترس است، دارد. مدل‌های چندوجهی معمولاً برای کاربردهایی که متن و تصویر به‌طور نزدیک با یکدیگر ارتباط دارند، مناسب‌تر هستند. در حالی که پردازش موازی ممکن است در موقعیت‌هایی که هر یک از نوع داده‌ها نیاز به پردازش خاص خود دارند، انتخاب بهتری باشد.

۱- ۴- ۲- مدل‌های چندوجهی

مدل‌های چندوجهی به‌طور هم‌زمان به تحلیل و پردازش داده‌های متنی و تصویری می‌پردازند. این مدل‌ها معمولاً به‌صورت یکپارچه طراحی می‌شوند و می‌توانند از ساختارهای عمیق یادگیری مانند شبکه‌های عصبی کانولوشنی^۲ برای تحلیل تصویر و شبکه‌های عصبی بازگشتی^۳ یا ترنسفورمرها^۴ برای تحلیل متن استفاده کنند. در این روش، مدل به‌طور هم‌زمان ویژگی‌های متنی و تصویری را استخراج کرده و از آن‌ها برای شناسایی تکنیک‌های متقاعدسازی استفاده می‌کند. به‌این ترتیب، اطلاعات غنی‌تری از هر دو منبع در اختیار مدل قرار می‌گیرد. این رویکرد به دلیل ارتباط نزدیک بین متن و تصویر در میم‌ها، معمولاً نتایج بهتری به همراه دارد، زیرا مدل می‌تواند از زمینه‌های متنی و بصری به‌عنوان مکمل‌های یکدیگر استفاده کند [۴].

۲- ۴- ۲- پردازش موازی و فیوژن

روش دوم شامل پردازش جداگانه داده‌های متنی و تصویری است. در این حالت، ابتدا یک مدل برای تحلیل متن و یک مدل دیگر برای تحلیل تصویر طراحی می‌شود. این دو مدل به‌صورت مستقل عمل می‌کنند و نتایج هر کدام در مرحله نهایی با یکدیگر ترکیب می‌شوند.

^۱ multimodal models

^۲ CNNs

^۳ RNNs

^۴ Transformers

تکنیک‌های مختلفی برای این فیوژن وجود دارد، مانند ترکیب خطی یا استفاده از لایه‌های طبقه‌بندی مشترک که ورودی‌های دو مدل را به‌عنوان ویژگی‌ها دریافت می‌کند. هدف این روش این است که هر مدل به‌طور خاص بر روی نوع داده‌ای که با آن کار می‌کند، تمرکز کند و در نهایت نتایج را به شکلی مؤثر ترکیب کند. این روش می‌تواند از مزیت‌های خاصی برخوردار باشد، به‌ویژه زمانی که نیاز به تجزیه و تحلیل دقیق‌تری از ویژگی‌های هر دو نوع داده وجود دارد. با این حال، این رویکرد ممکن است به‌عنوان یک فرآیند زمان‌برتر در مقایسه با مدل‌های چندوجهی در نظر گرفته شود، زیرا نیاز به ترکیب دو مجموعه از داده‌ها دارد.

۵-۲- کارهای پیشین

در تسک ۳ مسابقه SemEval 2023 که به شناسایی تکنیک‌های اقناعی در متن‌های خبری پرداخته می‌شود، نتایج بهترین مدل‌ها برای هر تسک فرعی به شرح زیر است:

تسک فرعی ۱: طبقه‌بندی نوع اخبار

داده‌ها و زبان‌ها: داده‌ها شامل مقالات خبری به زبان‌های انگلیسی، فرانسه، آلمانی، ایتالیایی، لهستانی و روسی هستند.

نوع تسک: طبقه‌بندی چندکلاسه برای شناسایی نوع مقاله (نظر، گزارش، یا طنز).

لیبل‌ها: سه نوع ژانر: «گزارش»، «نظر»، «طنز».

معیار ارزیابی: معیار F1-macro

بهترین نتایج در این تسک با استفاده از مدل XLM-RoBERTa به دست آمد. این مدل با تمرکز بر ویژگی‌های زبانی و ساختاری متن، توانست دقت بالایی در طبقه‌بندی مقالات خبری به سه نوع (خبر، نظر و طنز) ارائه دهد [۲۴][۲۳].

تسک فرعی ۲: شناسایی فریم‌ها

داده‌ها و زبان‌ها: مشابه تسک فرعی ۱.

نوع تسک: شناسایی چارچوب‌ها به صورت چندلیلی.

لیبل‌ها: ۱۴ چارچوب مختلف مانند «اقتصادی»، «اخلاقی»، و «سیاسی».

معیار ارزیابی: معیار F1-micro

در این تسک، برخی تیم‌ها از مدل‌های ALBERT و mBERT استفاده کردند که عملکرد خوبی در شناسایی

استفاده از یادگیری عمیق در تشخیص تکنیک‌های متقاعدسازی به کاررفته در میم‌ها کارهای پیشین و مرتبط

فریم‌های خبری داشتند. این مدل‌ها توانستند فریم‌های مختلف را با دقت شناسایی کنند و به تجزیه و تحلیل جنبه‌های متفاوت اخبار کمک کنند [۲۳].

تسک فرعی ۳: شناسایی تکنیک‌های اقناعی

داده‌ها و زبان‌ها: مشابه تسک فرعی ۱

نوع تسک: شناسایی تکنیک‌های متقاعدسازی در هر پاراگراف به صورت چندلیلی.

لیبل‌ها: ۲۳ تکنیک متقاعدسازی مختلف.

معیار ارزیابی: معیار F1-micro

تیم KInITVeraAI در این زیرتسک با استفاده از XLM-RoBERTa large نتایج برجسته‌ای کسب کرد.

این تیم علاوه بر آموزش مدل، به بررسی استراتژی‌های پیش‌پردازش و تنظیم آستانه اطمینان پرداختند که تأثیر زیادی بر بهبود عملکرد داشت [۲۴].

مدل‌های ترکیبی و روش‌های تجمعی نیز در این تسک به کار گرفته شدند و نشان دادند که ترکیب مدل‌ها می‌تواند به بهبود عملکرد کلی کمک کند.

۶-۲- مدل‌های زبانی

مدل‌های زبانی ابزارهای قدرتمندی در پردازش زبان طبیعی^۱ هستند که با استفاده از الگوریتم‌های یادگیری ماشین و یادگیری عمیق به تحلیل و تولید متن می‌پردازند. این مدل‌ها توانایی درک و تولید زبان انسانی را دارند و می‌توانند در کاربردهای متنوعی مانند ترجمه ماشینی، خلاصه‌سازی متن، پاسخ به سوالات و تولید متن خلاقانه به کار روند.

مدل‌های زبانی بر اساس الگوهای زبانی موجود در داده‌های متنی آموزش می‌بینند. این مدل‌ها معمولاً به دو دسته تقسیم می‌شوند: مدل‌های زبانی سنتی و مدل‌های زبانی پیشرفته‌تر مبتنی بر ترنسفورمرها. مدل‌های سنتی مانند مدل‌های n-gram، با استفاده از احتمالات شرطی برای پیش‌بینی کلمات بعدی در یک جمله کار می‌کنند. در این مدل‌ها، توالی کلمات با استفاده از تعدادی از کلمات قبلی (n) مورد بررسی قرار می‌گیرد. اما این مدل‌ها به دلیل وابستگی‌های بلندمدت در زبان و مشکلات مربوط به ذخیره‌سازی و محاسبات، محدودیت‌هایی دارند.

^۱ Natural Language Processing

با ظهور مدل‌های مبتنی بر ترنسفورمر، مانند^۱ BERT، کیفیت پردازش زبان به طور چشمگیری افزایش یافته است. ترنسفورمرها از ساختار خاصی به نام توجه استفاده می‌کنند که به آن‌ها این امکان را می‌دهد تا وابستگی‌های بلندمدت را به‌طور مؤثر مدیریت کنند. BERT به‌طور خاص برای درک متن و تولید نمایندگی‌های معنایی با دقت بالا طراحی شده است، در حالی که GPT برای تولید متن و پیش‌بینی توالی کلمات بهینه شده است [۵][۶].

مدل‌های زبانی بزرگ، مانند GPT-2 و GPT-3، با استفاده از مقادیر زیادی از داده‌های متنی آموزش می‌بینند و توانایی‌های پیشرفته‌ای در زمینه‌های مختلف زبان دارند. این مدل‌ها می‌توانند به‌طور خودکار متن‌هایی با کیفیت بالا تولید کنند، جملات را کامل کنند و حتی به سوالات پیچیده پاسخ دهند. این قابلیت‌ها آن‌ها را به ابزارهای ارزشمندی در زمینه‌های مختلف تبدیل کرده است، از جمله تولید محتوا، خدمات مشتری، و سیستم‌های توصیه‌گر [۶].

با این حال، مدل‌های زبانی چالش‌هایی نیز دارند. یکی از این چالش‌ها، تمایل به تولید اطلاعات نادرست یا تحریف شده است، که می‌تواند به دلیل داده‌های آموزشی نادرست یا تعصبات موجود در داده‌ها باشد. همچنین، مصرف بالای منابع محاسباتی برای آموزش و اجرای این مدل‌ها، نگرانی‌های زیست‌محیطی و اقتصادی را به همراه دارد [۵].

به‌طور کلی، مدل‌های زبانی ابزارهای حیاتی در پردازش زبان طبیعی به شمار می‌روند که با پیشرفت‌های تکنولوژیکی و تحقیقاتی، به بهبود توانایی‌های خود ادامه می‌دهند و کاربردهای گسترده‌تری در آینده پیدا خواهند کرد [۶].

یکی از مهم‌ترین مزیت‌های LLM‌ها، پیش‌آمورخته بودن آن‌هاست. یعنی این مدل‌ها ابتدا بر روی داده‌های عظیم بی‌نظارت آموزش دیده و سپس می‌توانند برای وظایف خاص زبانی تنظیم مجدد (fine-tune) شوند. با این روش، مدل قادر است دانش عمومی و گسترده‌ای از زبان را کسب کند و سپس به‌طور خاص بر روی داده‌های محدودتر تنظیم شود تا وظایف مشخص‌تری را انجام دهد [۵].

۱-۶-۲- شبکه عصبی عمیق GPT-2

شبکه عصبی GPT-2^۲ (ترنسفورمر تولیدگر از پیش آموزش‌دیده ۲) یکی از پیشرفته‌ترین مدل‌های زبان

^۱ Bidirectional Encoder Representations from Transformers

^۲ Generative Pretrained Transformer-2

پردازشی مبتنی بر یادگیری عمیق است که توسط شرکت OpenAI توسعه یافته است. این مدل در سال ۲۰۱۹ معرفی شد و توجه زیادی به خود جلب کرد، زیرا توانایی‌های بی‌سابقه‌ای در درک و تولید متن انسانی از خود نشان داد. GPT-2 بر پایه معماری Transformer طراحی شده که به طور خاص برای پردازش داده‌های متنی در مقیاس بزرگ به کار می‌رود. این مدل توانایی تولید متونی دارد که نه تنها از لحاظ زبانی درست هستند، بلکه از نظر منطقی و موضوعی نیز به شکل قابل قبولی دنبال می‌شوند [۷].

۱-۶-۲- آماده سازی ورودی متنی GPT-2

در ادامه به بررسی معماری این مدل می‌پردازیم. قبل از بررسی دیاگرام معماری، ضروری است که نحوه آماده‌سازی ورودی متنی را توضیح دهیم. متن خام ابتدا باید توکنایز شود، به این معنی که کلمات به اعداد صحیحی تبدیل می‌شوند که به ایندکس‌های موجود در واژگان مدل ارتباط دارند. در این راستا، مدل GPT-2 از روش خاصی به نام BPE^۱ برای توکن‌سازی استفاده می‌کند که متن را به زیرکلمات تقسیم می‌نماید [۸]. در این مطالعه، از کتابخانه Tiktoken که یک توکنایزر BPE سریع است و در مدل‌های OpenAI به کار می‌رود، استفاده خواهد شد. به عنوان مثال، شکل (۲-۲) نشان‌دهنده نحوه توکنایز کردن متن است:

```
import tiktoken

enc = tiktoken.get_encoding("gpt2") # Load the GPT-2 tokenizer

text = """In a remote mountain village, nestled among the towering peaks and pin
a solitary storyteller sat by a crackling bonfire. Their voice rose and fell lik
weaving tales of ancient legends and forgotten heroes that captivated the hearts
The stars above shone brightly, their twinkling adding a celestial backdrop to t
as the storyteller transported their audience to worlds of wonder and imaginatio

tokens = enc.encode(text) # Tokenize the text
tokens.append(enc.eot_token) # Add the end of text token

print(tokens[:10])
```



```
[818, 257, 6569, 8598, 7404, 11, 16343, 992, 1871, 262]
```

شکل (۲-۲) نمونه استفاده از توکنایزر GPT-2 [۹]

^۱ Byte-Pair Encoding

استفاده از یادگیری عمیق در تشخیص تکنیک‌های متقاعدسازی به کاررفته در میم‌ها کارهای پیشین و مرتبط

پس از توکنایز کردن متن با استفاده از تابع encode، توکن‌ها به صورت لیستی از اعداد صحیح در می‌آیند که نشان‌دهنده زیرکلمات یا کلمات واژگان مدل GPT-2 هستند. همچنین، یک توکن خاص `<|endoftext|>` به انتهای توکن‌ها اضافه می‌شود تا پایان توالی متن را مشخص کند [۹].

سپس، توکن‌ها باید به فرمتی تبدیل شوند که مدل GPT بتواند آن‌ها را پردازش کند. این مدل انتظار دارد که داده ورودی به صورت یک دسته از توالی‌ها (batch of sequences) باشد [۱۰]. شکل تانسور ورودی باید به صورت (B, T) باشد که در آن:

- B اندازه دسته (تعداد توالی‌هایی که به صورت همزمان پردازش می‌شوند) و
- T طول توالی (تعداد توکن‌ها در هر توالی) است [۱۰].

برای نمونه، یک دسته شامل ۵ توالی که هر کدام شامل ۱۰ توکن هستند، با استفاده از ۵۰ توکن اول ورودی به صورت زیر ایجاد می‌شود:

```
import torch

B, T = 5, 10
data = torch.tensor(tokens[:50+1])

x = data[:-1].view(B, T) # Input tensor
y = data[1:].view(B, T) # Target tensor for next token prediction

print(x)
print(y)
```



```
tensor([[ 818, 257, 6569, 8598, 7404, 11, 16343, 992, 1871, 262],
        [38879, 25740, 290, 20161, 17039, 11, 257, 25565, 1621, 660],
        [ 6051, 3332, 416, 257, 8469, 1359, 5351, 6495, 13, 5334],
        [ 3809, 8278, 290, 3214, 588, 257, 7758, 29512, 7850, 11],
        [44889, 19490, 286, 6156, 24901, 290, 11564, 10281, 326, 3144]])

tensor([[ 257, 6569, 8598, 7404, 11, 16343, 992, 1871, 262, 38879],
        [25740, 290, 20161, 17039, 11, 257, 25565, 1621, 660, 6051],
        [ 3332, 416, 257, 8469, 1359, 5351, 6495, 13, 5334, 3809],
        [ 8278, 290, 3214, 588, 257, 7758, 29512, 7850, 11, 44889],
        [19490, 286, 6156, 24901, 290, 11564, 10281, 326, 3144, 30829]])
```

شکل (۲-۳) نمونه فرمت قابل قبول داده در GPT-2 [۹]

در اینجا، x شامل توکن‌های ورودی و y شامل توکن‌های هدف است که به صورت یک موقعیت جابجا شده‌اند تا مدل بتواند یاد بگیرد که توکن بعدی در توالی چیست. در طول فرایند آموزش، این تانسورها به مدل ورودی

داده می‌شوند و تنسور هدف y برای محاسبه ضرر با استفاده از روش ضرر متقاطع^۱ به کار می‌رود [۹][۱۰].

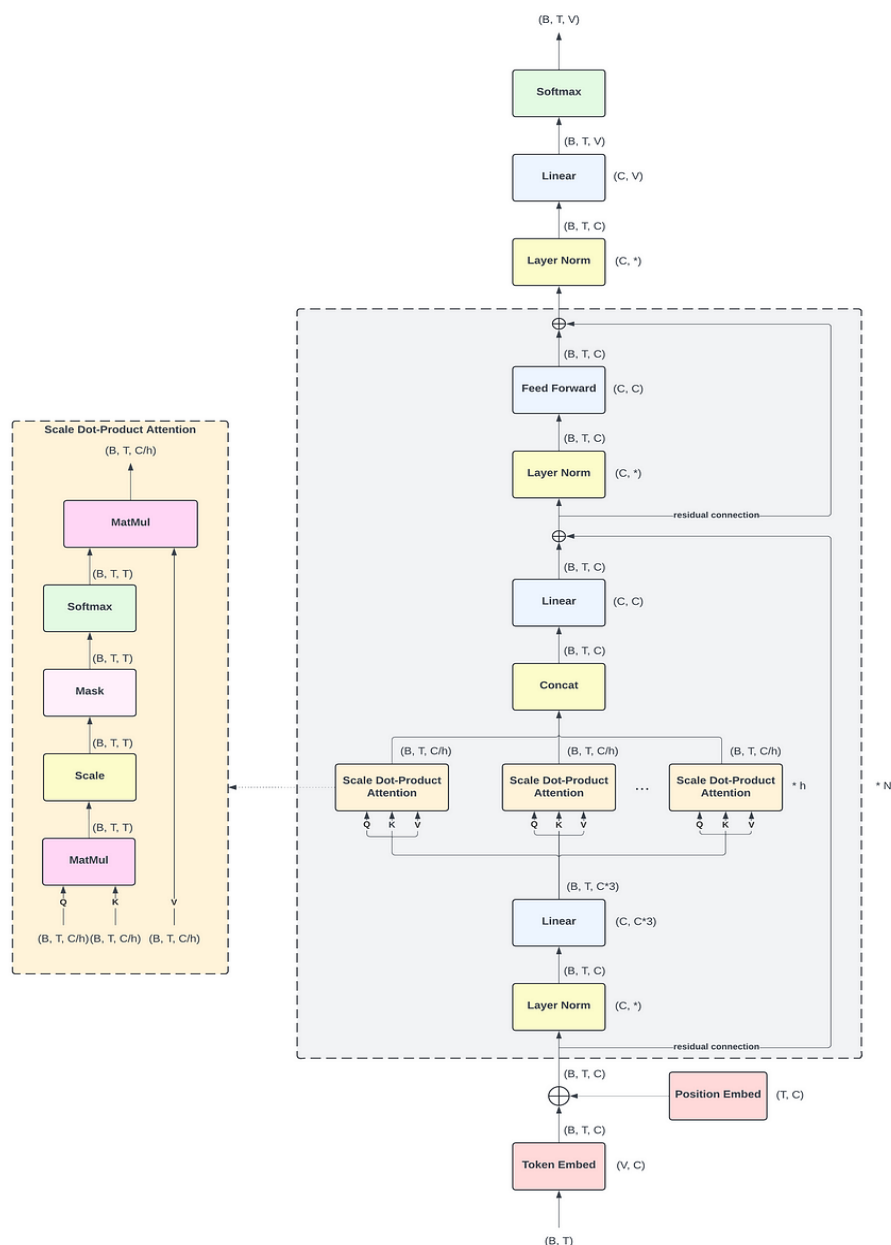
۲-۱-۶-۲- دیاگرام معماری GPT-2

در نهایت، به بررسی دیاگرام معماری GPT-2 پرداخته می‌شود تا نحوه حرکت داده‌های ورودی در مدل به‌طور دقیق‌تری مورد بررسی قرار گیرد. برخی از پارامترهای کلیدی تنظیمات برای GPT-2 به شرح زیر است:

- اندازه واژگان (V): ۵۰,۲۵۷
- حداکثر طول توالی (T): ۱,۰۲۴
- ابعاد جاسازی (C): ۷۶۸
- تعداد سرها (h): ۱۲
- تعداد لایه‌ها (N): ۱۲
- اندازه دسته (B): ۵۱۲

این دیاگرام و تنظیمات به درک بهتر فرآیند تغییر داده‌ها در هر مرحله، از ورودی تا خروجی، کمک خواهند کرد [۹].

^۱ cross-entropy loss



شکل (۲-۴) معماری مدل GPT-2 [۹]

۲-۶-۲- شبکه عصبی XLM

XLM^۱ یکی از اولین مدل‌های زبانی چندزبانه است که توسط تیم تحقیقاتی فیسبوک (Meta AI) در سال ۲۰۱۹ معرفی شد. این مدل برای انجام وظایف زبانی چندزبانه مانند ترجمه ماشین، تشخیص زبان و انتقال

^۱ Cross-lingual Language Model

یادگیری بین زبان‌های مختلف طراحی شده است [۱۲].

XLM یک مدل بر اساس معماری ترنسفورمر است و مشابه مدل‌های زبانی تک‌زبانه مانند BERT عمل می‌کند، اما تفاوت اصلی آن این است که این مدل برای پردازش چند زبان آموزش داده شده است. XLM از داده‌های چندزبانه برای آموزش استفاده می‌کند و می‌تواند یادگیری را از یک زبان به زبان دیگر انتقال دهد، به این معنی که اگر مدلی در یک زبان (مثلاً انگلیسی) آموزش ببیند، می‌تواند عملکرد قابل قبولی در زبان‌های دیگر (مانند فرانسوی یا چینی) نیز داشته باشد، حتی اگر داده‌های آموزشی کمتری برای آن زبان‌ها موجود باشد [۱۳].

ویژگی‌های اصلی XLM

۱. یادگیری چندزبانه XLM: با داده‌هایی از زبان‌های مختلف آموزش داده شده و قادر به انجام وظایف چندزبانه است.

۲. پیش‌تربیتی^۱: مانند BERT، XLM نیز با استفاده از پیش‌تربیتی به روش masked language modeling^۲ آموزش داده شده است. در این روش، برخی از کلمات در جمله مخفی می‌شوند و مدل باید این کلمات را پیش‌بینی کند.

۳. ترجمه خودکار (TLM): علاوه بر MLM، XLM از یک روش پیش‌تربیتی جدید به نام TLM نیز استفاده می‌کند. در این روش، جملات ترجمه شده در زبان‌های مختلف به مدل داده می‌شوند و مدل باید با استفاده از اطلاعات از هر دو زبان، پیش‌بینی‌های خود را انجام دهد. این روش باعث می‌شود مدل به طور مؤثرتری یاد بگیرد که چگونه بین زبان‌ها ارتباط برقرار کند.

۳-۶-۲- شبکه عصبی XLM-RoBERTa

XLM-RoBERTa نسخه بهبود یافته‌ای از XLM است که در سال ۲۰۱۹ توسط فیسبوک معرفی شد. این مدل بر پایه RoBERTa (یک نسخه بهینه‌شده از BERT) ساخته شده و برای یادگیری عمیق چندزبانه طراحی شده است. XLM-RoBERTa با استفاده از داده‌های عظیم چندزبانه آموزش دیده و توانایی بالایی در پردازش و تحلیل متون چندزبانه دارد [۱۲].

^۱ Pretraining

^۲ MLM

^۳ Translation Language Modeling

ویژگی‌های اصلی XLM-RoBERTa [۱۴]

۱. بهره‌گیری از RoBERTa: RoBERTa یکی از نسخه‌های پیشرفته BERT است که با استفاده از بهینه‌سازی‌های مختلف (مانند استفاده از داده‌های بیشتر و تغییرات در فرآیند پیش‌تربیتی) عملکرد بهتری نسبت به BERT دارد. XLM-RoBERTa با بهره‌گیری از این بهینه‌سازی‌ها، دقت و کارایی بالاتری در پردازش زبان‌های مختلف به دست آورده است.
۲. داده‌های عظیم‌تر XLM-RoBERTa: بر روی داده‌های بسیار بزرگ‌تری نسبت به XLM اولیه آموزش داده شده است. این مدل از ۲,۵ ترابایت داده متنی از ۱۰۰ زبان مختلف استفاده کرده است که باعث بهبود عملکرد آن در تمامی وظایف چندزبانه شده است.
۳. یادگیری بی‌نظارت چندزبانه: این مدل به صورت بی‌نظارت و بدون نیاز به داده‌های برچسب‌دار زبان‌های مختلف آموزش دیده است، که آن را قادر می‌سازد در بسیاری از زبان‌ها به خوبی عمل کند حتی اگر داده‌های کمی برای آن زبان‌ها موجود باشد.
۴. تطبیق‌پذیری بالا XLM-RoBERTa: توانایی بالایی در انجام وظایف مختلف زبان‌شناسی مانند ترجمه، طبقه‌بندی متن، پاسخ به سوالات و تحلیل احساسات دارد. این مدل همچنین در بسیاری از رقابت‌های معتبری مانند GLUE و XNLI عملکرد برجسته‌ای از خود نشان داده است.

کاربردهای XLM-RoBERTa [۱۴]

۱. ترجمه خودکار: این مدل در وظایف مربوط به ترجمه چندزبانه عملکرد بسیار خوبی دارد و می‌تواند بین زبان‌های مختلف ارتباط برقرار کند.
 ۲. تحلیل متون چندزبانه XLM-RoBERTa: می‌تواند در تحلیل متون در زبان‌های مختلف، از جمله تحلیل احساسات، طبقه‌بندی موضوعی و استخراج اطلاعات، به کار گرفته شود.
 ۳. تشخیص زبان: این مدل در تشخیص زبان متن‌ها، حتی در مواردی که متن‌های کوتاه یا مبهم باشند، عملکرد بسیار خوبی دارد.
- XLM-RoBERTa یکی از قدرتمندترین مدل‌های زبانی چندزبانه است که توانسته است با استفاده از بهینه‌سازی‌های مختلف، عملکرد بسیار بهتری نسبت به نسخه‌های پیشین خود نشان دهد. این مدل نقش مهمی در پیشرفت تکنولوژی پردازش زبان طبیعی در سطح جهانی داشته است و می‌تواند به عنوان یکی از پایه‌های اصلی برای توسعه سیستم‌های چندزبانه به کار رود.

تفاوت‌های XLM و XLM-RoBERTa [۱۲][۱۴]

۱. حجم داده‌های آموزشی XLM-RoBERTa: از داده‌های بسیار بیشتری نسبت به XLM اولیه استفاده

کرده است، که باعث بهبود چشمگیر عملکرد آن شده است.

۲. روش‌های پیش‌تربیتی: در حالی که XLM از هر دو روش MLM و TLM برای پیش‌تربیتی استفاده

کرده، XLM-RoBERTa فقط از MLM بهره می‌برد. اما به دلیل بهینه‌سازی‌های انجام شده در

معماری RoBERTa، این مدل بدون نیاز به TLM توانسته است به عملکرد بهتری دست یابد.

۳. دامنه زبان‌ها XLM-RoBERTa: بر روی ۱۰۰ زبان آموزش دیده است، در حالی که XLM بر روی

تعداد کمتری از زبان‌ها آموزش دیده بود.

۱-۳-۶-۲- معماری XLM-RoBERTa

XLM-R^۱ از مدل ترنسفورمر استفاده می‌کند که با هدف مدل زبان ماسک‌شده چندزبانه آموزش داده شده و

تنها از داده‌های تک‌زبانه بهره می‌برد. در این مدل، متن‌هایی از هر زبان به صورت جریان‌های متنی انتخاب

می‌شوند و مدل برای پیش‌بینی توکن‌های ماسک‌شده در ورودی آموزش داده می‌شود. این مدل از sentence

piece با مدل زبانی unigram برای توکنیزه کردن زیرواحد‌های متنی در متن خام استفاده می‌کند و دسته‌هایی

از زبان‌های مختلف را بر اساس همان توزیع نمونه‌گیری با $(\alpha = 0.3)$ نمونه‌برداری می‌کند [۱۳].

هدف اصلی مدل زبان ماسک‌شده این است که یک یا چند کلمه در جمله را ماسک کند و مدل باید توکن‌های

ماسک‌شده را با توجه به دیگر کلمات موجود در جمله پیش‌بینی کند. هدف مدل زبان ترجمه‌شده (TLM)،

گسترش MLM به جفت جملات موازی است [۱۳]. به عنوان مثال، برای پیش‌بینی یک کلمه انگلیسی

ماسک‌شده، مدل می‌تواند هم به جمله انگلیسی و هم به ترجمه فرانسوی آن توجه کند، به این ترتیب مدل

تشویق می‌شود که بازنمایی‌های انگلیسی و فرانسوی را هم‌تراز کند. اکنون مدل می‌تواند از بستر فرانسوی

استفاده کند اگر بستر انگلیسی به تنهایی برای استنباط کلمات ماسک‌شده کافی نباشد [۱۳].

شکل (۵-۲) پیش‌آموزش مدل زبانی چند زبانه را نشان می‌دهد. هدف MLM (مدل‌سازی زبانی ماسک‌شده)

مشابه هدف معرفی‌شده توسط دولین و همکاران (۲۰۱۸) است، اما با استفاده از جریان‌های پیوسته‌ای از متن

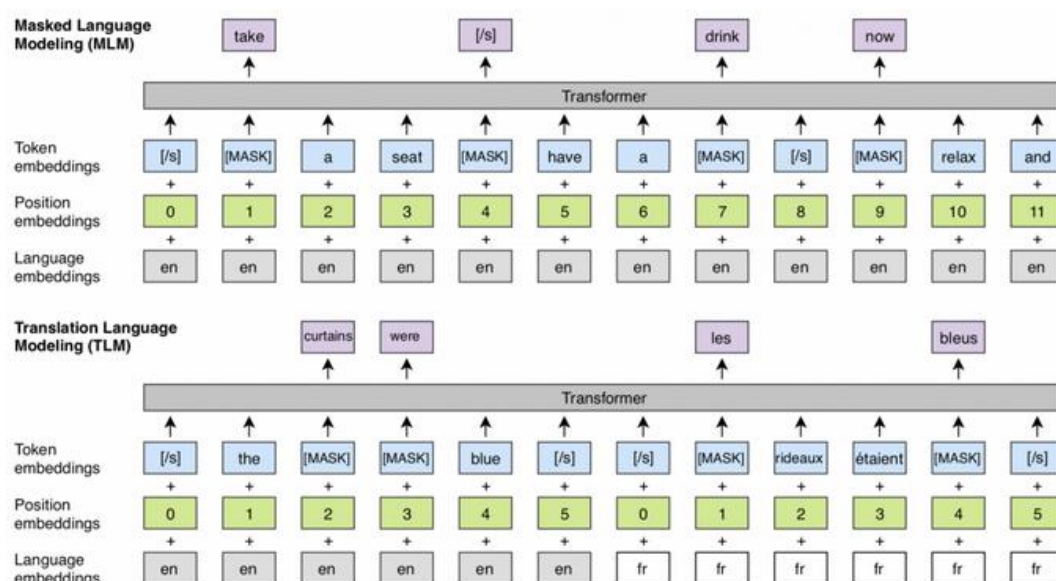
به جای جفت جملات. هدف TLM (مدل‌سازی زبانی ترجمه) هدف MLM را به جفت جملات موازی گسترش

می‌دهد. برای پیش‌بینی یک کلمه ماسک‌شده در انگلیسی، مدل می‌تواند هم به جمله انگلیسی و هم به ترجمه

فرانسوی آن توجه کند و تشویق می‌شود که نمایش‌های انگلیسی و فرانسوی را با هم تطبیق دهد. جاسازی‌های

^۱ XLM-RoBERTa

موقعیتی جمله هدف مجدداً تنظیم می‌شوند تا این هم‌ترازی را تسهیل کنند. [۱۳]



شکل (۵-۲) پیش‌آموزش مدل زبانی چندزبانه [۱۳]

۷-۲- مدل‌های بینایی عمیق

در سال‌های اخیر، توجه زیادی به بررسی و تحلیل تکنیک‌های متقاعدسازی در متون و محتوای چندرسانه‌ای شده است. پژوهشگران تلاش کرده‌اند با استفاده از روش‌های مختلف، این تکنیک‌ها را شناسایی و تحلیل کنند [۱۷].

در سال‌های اخیر، مدل‌های بینایی عمیق نقش مهمی در پیشرفت‌های تکنولوژی بینایی ماشین و پردازش تصویر ایفا کرده‌اند. این مدل‌ها با استفاده از شبکه‌های عصبی پیچیده قادر به تحلیل و استخراج ویژگی‌های بصری از داده‌های تصویری هستند. دو مدل مطرح و پیشرو در این حوزه، VGG-16 و ViT هستند که به صورت گسترده در مسائل مختلف بینایی مورد استفاده قرار گرفته‌اند [۱۵].

مدل‌های بینایی عمیق مانند VGG-16 و ViT هر یک نقاط قوت خود را دارند و انتخاب آن‌ها بسته به نوع مسئله و داده متفاوت است. VGG-16 با وجود ساختار ساده‌تر و عملکرد قابل اعتماد در مسائل مختلف بینایی، همچنان مورد استفاده قرار می‌گیرد، در حالی که ViT با ارائه رویکردی نوین در پردازش تصویر، توانسته است مرزهای جدیدی در کارایی و دقت به وجود آورد [۱۵][۱۸].

۱-۷-۲- شبکه عصبی VGG-16

مدل VGG-16 یکی از شناخته‌شده‌ترین معماری‌های شبکه عصبی پیچشی^۱ است که توسط Karen Simonyan و Andrew Zisserman از دانشگاه آکسفورد در سال ۲۰۱۴ در مقاله‌ای با عنوان Very Deep Convolutional Networks for Large-Scale Image Recognition معرفی شد. این مدل یکی از موفق‌ترین معماری‌ها در رقابت‌های طبقه‌بندی تصویر، به‌ویژه در مسابقات ImageNet، بوده و به دلیل سادگی و کارایی بالای آن در بسیاری از مسائل پردازش تصویر همچنان مورد استفاده قرار می‌گیرد [۱۶].

ویژگی بارز این مدل، استفاده از فیلترهای کوچک ۳ در ۳ در تمامی لایه‌های کانولوشن است که با عمق بیشتر به مدل امکان می‌دهد تا ویژگی‌های پیچیده‌تری از تصاویر را استخراج کند. اگرچه این مدل از لحاظ تعداد پارامترها بسیار حجیم است، اما به دلیل دقت بالای آن در بسیاری از مسائل مانند طبقه‌بندی تصویر و تشخیص اشیا کاربرد گسترده‌ای پیدا کرده است.

۱-۷-۲-۱- معماری VGG-16

VGG-16 یک شبکه عصبی عمیق است که از ۱۶ لایه قابل آموزش تشکیل شده است (۱۳ لایه کانولوشن و ۳ لایه تمام‌متصل). ساختار این مدل به گونه‌ای طراحی شده که از ترکیب چندین لایه کانولوشن با اندازه کوچک و لایه‌های Pooling استفاده می‌کند تا ویژگی‌های تصویری را استخراج و فشرده کند. در نهایت، داده‌های استخراج شده از این لایه‌ها وارد لایه‌های تمام‌متصل شده و در خروجی طبقه‌بندی انجام می‌شود [۱۵][۱۶].

لایه‌های اصلی VGG-16:

۱. ورودی: این مدل تصاویری با اندازه ثابت ۲۲۴ در ۲۲۴ پیکسل را به عنوان ورودی دریافت می‌کند.

۲. لایه‌های کانولوشن:

○ VGG-16 از چندین لایه کانولوشن ۳ در ۳ استفاده می‌کند. فیلترهای کوچک ۳ در ۳ به این

مدل امکان می‌دهند تا به تدریج ویژگی‌های پیچیده‌تری از تصاویر استخراج کند، در حالی

که تعداد پارامترها به طور کنترل‌شده‌ای افزایش می‌یابد.

- بعد از هر دو یا سه لایه کانولوشن، یک لایه Max-Pooling قرار می‌گیرد که اندازه تصویر را کاهش می‌دهد و ویژگی‌های مهم را از هر ناحیه انتخاب می‌کند.
- تعداد فیلترها در هر بلوک کانولوشنی به تدریج افزایش می‌یابد (از ۶۴ فیلتر در اولین لایه به ۵۱۲ فیلتر در لایه‌های پایانی).

۳. لایه‌های Pooling :

بعد از هر گروه از لایه‌های کانولوشن، یک لایه Max-Pooling با اندازه 2×2 قرار دارد که اندازه ویژگی‌های استخراج شده را نصف می‌کند، و در عین حال مهم‌ترین ویژگی‌ها را حفظ می‌کند. این لایه‌ها به کاهش ابعاد و جلوگیری از بیش‌برازش کمک می‌کنند.

۴. لایه‌های تمام‌متصل (Fully Connected) :

در پایان شبکه، سه لایه تمام‌متصل قرار دارد:

- دو لایه با ۴۰۹۶ نورون.
- یک لایه خروجی با تعداد نورون‌هایی که به تعداد کلاس‌های طبقه‌بندی بستگی دارد (به‌عنوان مثال، در ImageNet که ۱۰۰۰ کلاس دارد، این لایه ۱۰۰۰ نورون خواهد داشت)

۵. لایه Softmax :

در انتهای مدل، یک لایه Softmax قرار دارد که احتمالات هر کلاس را محاسبه می‌کند و کلاس نهایی تصویر را بر اساس بالاترین احتمال پیش‌بینی می‌کند.

۲-۷-۱-۲- چگونگی استفاده از VGG-16 برای طبقه‌بندی تصویر

مدل VGG-16 به‌طور ویژه برای مسائل طبقه‌بندی تصویر طراحی شده است و می‌تواند برای طبقه‌بندی تصاویر در دسته‌های مختلف استفاده شود [۱۷]. برای استفاده از این مدل، مراحل زیر باید انجام شود:

۱. آموزش از ابتدا یا استفاده از مدل از پیش‌آمोخته:
می‌توان مدل VGG-16 را از ابتدا با مجموعه داده مورد نظر خود آموزش داد، اما از آنجایی که این مدل برای دیتاست ImageNet آموزش داده شده، بسیاری از کاربران از نسخه پیش‌آموزش‌یافته آن استفاده می‌کنند و لایه‌های آخر را برای مجموعه داده خاص خود تنظیم می‌کنند [۱۷].
۲. پیش‌پردازش داده‌ها:

تصاویر ورودی باید به اندازه ثابت ۲۲۴ در ۲۲۴ تغییر اندازه داده شوند و مقادیر پیکسل‌ها نرمال‌سازی

شوند تا عملکرد مدل بهبود یابد [۱۷].

۳. استخراج ویژگی‌ها:

لایه‌های کانولوشنی VGG-16 به عنوان استخراج کننده ویژگی‌های پیچیده از تصویر عمل می کنند. این ویژگی‌ها سپس توسط لایه‌های تمام متصل پردازش شده و کلاس‌های احتمالی پیش‌بینی می شوند [۱۷].

۴. فاین تیونینگ (Fine-tuning):

در صورت استفاده از یک مدل پیش آموزش یافته، می توان تنها لایه‌های نهایی را با توجه به مجموعه داده ویژه مجدداً آموزش داد یا از تمامی لایه‌های مدل برای یادگیری ویژگی‌های جدید استفاده کرد [۱۶].

۵. پیش‌بینی و طبقه‌بندی:

پس از آموزش مدل، می توان از آن برای پیش‌بینی کلاس هر تصویر جدید استفاده کرد. با دریافت تصویر ورودی و اعمال لایه‌های کانولوشنی و تمام متصل، مدل کلاس نهایی را پیش‌بینی می کند.

۳-۱-۷-۲- مزایا و معایب VGG-16

مزایا:

۱. سادگی و کارایی: ساختار ساده و یکنواخت فیلترهای ۳ در ۳ باعث شده که VGG-16 به یکی از مدل‌های پایه و پرکاربرد در مسائل بینایی تبدیل شود.
۲. دقت بالا: این مدل به دلیل عمق زیاد و توانایی استخراج ویژگی‌های پیچیده، در بسیاری از مسائل پردازش تصویر عملکردی بسیار خوب داشته است.

معایب:

۱. تعداد بالای پارامترها: یکی از مشکلات اصلی VGG-16 تعداد زیاد پارامترها (حدود ۱۳۸ میلیون پارامتر) است که باعث افزایش حجم مدل و نیاز به منابع محاسباتی زیاد برای آموزش و استفاده از آن می شود [۱۷].
۲. حافظه و زمان پردازش: به دلیل تعداد پارامترهای زیاد، VGG-16 نیازمند حافظه بیشتر و زمان پردازش طولانی تر نسبت به مدل‌های جدیدتر است [۱۷].

۲-۷-۲- شبکه عصبی ViT^۱

مدل ViT که توسط Dosovitskiy و همکاران در سال ۲۰۲۰ معرفی شد، رویکردی کاملاً متفاوت با CNN ها دارد و بر اساس معماری ترنسفورمر (Transformer) عمل می‌کند. ترنسفورمرها ابتدا در پردازش زبان طبیعی به کار گرفته شدند، اما ViT نشان داد که این معماری می‌تواند در مسائل بینایی نیز به طور موثر عمل کند [۱۸].

ViT تصاویر را به بخش‌های کوچکی (پچ‌ها) تقسیم می‌کند و هر پچ را به عنوان یک توکن ورودی برای ترنسفورمر در نظر می‌گیرد. این مدل برخلاف CNN ها که از کانولوشن برای استخراج ویژگی‌ها استفاده می‌کنند، به کمک مکانیزم توجه^۲ در ترنسفورمر، می‌تواند ویژگی‌های مهم هر بخش از تصویر را شناسایی و پردازش کند [۱۹].

از مزایای اصلی ViT، توانایی آن در کار با مجموعه داده‌های بزرگ و کاهش نیاز به معماری‌های پیچیده کانولوشن است. همچنین، ViT در برخی موارد توانسته است دقت بهتری نسبت به CNN ها از خود نشان دهد و به خصوص در مسائل با داده‌های بزرگتر، کارایی بالاتری دارد [۱۹].

۱-۲-۷-۲- معماری ViT

معماری ViT بر مبنای ساختار ترنسفورمر است که از اجزای اصلی زیر تشکیل شده است:

۱. ورودی تصویر و تقسیم‌بندی:

○ به جای پردازش مستقیم تصاویر با ابعاد بزرگ، ViT تصاویر را به بلوک‌های کوچک‌تر (پچ‌ها) تقسیم می‌کند. هر تصویر به 16×16 پیکسل تقسیم می‌شود و این پچ‌ها به عنوان ورودی به مدل داده می‌شوند [۲۰].

○ هر پچ به یک وکتور تبدیل می‌شود، به طوری که با استفاده از یک لایه Flattening، هر پچ به یک وکتور سطری تبدیل می‌شود [۲۰].

۲. وکتورهای موقعیت:

○ به منظور حفظ اطلاعات مکانی هر پچ، یک وکتور موقعیت (Positional Encoding) به هر

^۱ Vision Transformer

^۲ Attention

وکتور پچ افزوده می‌شود. این کار به مدل اجازه می‌دهد که موقعیت هر پچ در تصویر را تشخیص دهد [۲۰].

۳. لایه‌های ترنسفورمر:

- ViT از چندین لایه ترنسفورمر تشکیل شده است که هر لایه شامل دو بخش اصلی است:
 - Self-Attention Mechanism: این بخش به مدل اجازه می‌دهد که توجه خود را بر روی پچ‌های مختلف تصویر متمرکز کند و وابستگی‌ها را بین آنها شناسایی کند [۲۰].
 - Feed-Forward Neural Network: بعد از محاسبات attention، داده‌ها از طریق یک شبکه عصبی پیش‌خور (Feed-Forward) پردازش می‌شوند [۲۰].

۴. لایه‌های نرمال‌سازی و Dropout:

- برای بهبود عملکرد و جلوگیری از بیش‌برازش، در بین لایه‌های attention و feed-forward از تکنیک‌های نرمال‌سازی و Dropout استفاده می‌شود [۲۰].

۵. لایه خروجی:

- در انتهای مدل، یک لایه کلاس‌بندی وجود دارد که از یک لایه تمام‌متصل (Fully Connected Layer) برای پیش‌بینی کلاس نهایی استفاده می‌کند [۲۰].

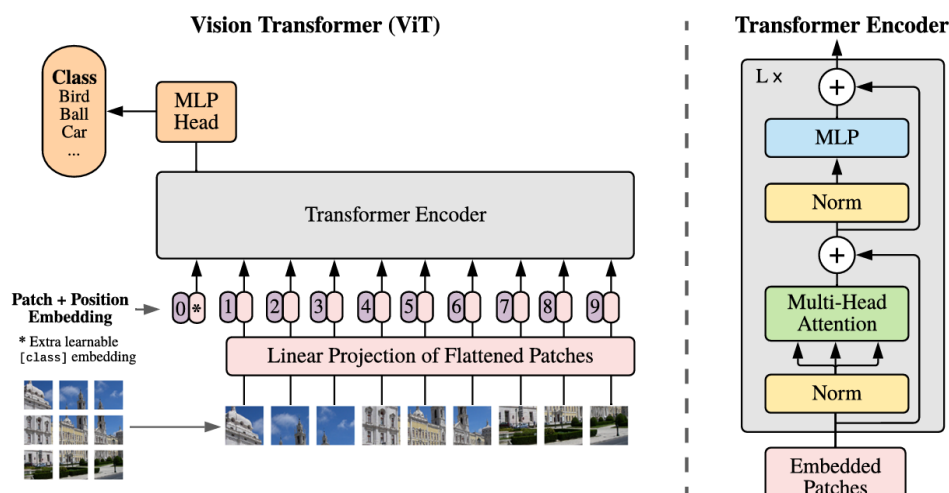
۲-۲-۷-۲- نحوه عملکرد ViT

عملکرد ViT به صورت زیر است:

۱. تقسیم تصویر: ابتدا تصویر ورودی به پچ‌های کوچک‌تر تقسیم می‌شود و هر پچ به یک وکتور تبدیل می‌شود [۱۸].
۲. افزودن وکتورهای موقعیت: وکتورهای موقعیت به وکتورهای پچ اضافه می‌شوند تا اطلاعات مکانی حفظ شود.
۳. پردازش در لایه‌های ترنسفورمر: وکتورهای پچ به لایه‌های ترنسفورمر ارسال می‌شوند، جایی که self-attention و feed-forward به ترتیب وابستگی‌ها را شناسایی و پردازش می‌کنند.
۴. پیش‌بینی کلاس: در انتها، خروجی نهایی به لایه تمام‌متصل منتقل می‌شود که کلاس نهایی را پیش‌بینی می‌کند.

شکل (۶-۲) نمای کلی مدل ViT را نشان می‌دهد. در این مثال ما یک تصویر را به بخش‌های ثابت تقسیم

استفاده از یادگیری عمیق در تشخیص تکنیک‌های متقاعدسازی به کاررفته در میم‌ها کارهای پیشین و مرتبط می‌کنیم، هر یک از آن‌ها را به صورت خطی تعبیه می‌کنیم، به آن‌ها تعبیه موقعیت اضافه می‌کنیم و دنباله‌ی حاصل از وکتورها را به یک کدگذار ترنسفورمر استاندارد وارد می‌کنیم. برای انجام طبقه‌بندی، از رویکرد استاندارد افزودن یک «توکن طبقه‌بندی» یادگیرنده اضافی به دنباله استفاده می‌کنیم. تصویر کدگذار ترنسفورمر از کار واسیانی و همکاران (۲۰۱۷) الهام گرفته شده است [۱۸].



شکل (۲-۶) نمای کلی مدل ViT [۱۸]

۳-۲-۷-۲- چگونگی استفاده از ViT برای طبقه‌بندی تصاویر

برای استفاده از مدل ViT در طبقه‌بندی تصاویر، مراحل زیر دنبال می‌شود:

۱. آموزش مدل:
 - می‌توان مدل ViT را از ابتدا آموزش داد یا از مدل‌های پیش‌آموزش‌یافته استفاده کرد. برای آموزش، مجموعه داده‌های بزرگ مانند ImageNet معمولاً استفاده می‌شود.
۲. پیش‌پردازش داده‌ها:
 - تصاویری که به مدل داده می‌شوند باید به ابعاد ثابت (به عنوان مثال، ۲۲۴ در ۲۲۴ پیکسل) تغییر اندازه شوند و همچنین ممکن است نرمال‌سازی شوند.
۳. تبدیل تصویر به پچ‌ها:
 - پس از پیش‌پردازش، تصویر به پچ‌های کوچک تقسیم می‌شود و هر پچ به وکتور تبدیل می‌شود.

- پس از آموزش، مدل می‌تواند بر روی تصاویر جدید اجرا شود و کلاس‌های مربوطه را پیش‌بینی کند. مدل ViT به ویژه در شناسایی الگوهای پیچیده و وابستگی‌های درون تصویر موفق عمل می‌کند.

۴-۲-۷-۲- مزایا و معایب ViT

مزایا:

- عملکرد بالا: ViT در مقایسه با مدل‌های CNN سنتی، در بسیاری از وظایف طبقه‌بندی تصویر عملکرد بالاتری نشان می‌دهد [۲۰].
- انعطاف‌پذیری: این مدل قابلیت پردازش انواع مختلف داده‌ها را دارد و می‌تواند برای وظایف مختلفی مانند تشخیص اشیا و تشخیص تصویر استفاده شود [۱۸].

معایب:

- نیاز به داده‌های زیاد: ViT به مقدار زیادی داده برای آموزش نیاز دارد، و این ممکن است در شرایطی که داده‌های آموزشی محدود است، مشکل‌ساز باشد [۱۹].
- پیچیدگی محاسباتی: از آنجایی که مدل ViT بر مبنای self-attention است، نیاز به منابع محاسباتی بالایی دارد که ممکن است در کاربردهای زمان واقعی مشکل ایجاد کند [۱۸].

۸-۲- معیار ارزیابی

در فرآیند توسعه و ارزیابی مدل‌های یادگیری ماشین، استفاده از معیارهای ارزیابی مناسب برای سنجش عملکرد مدل اهمیت بسیاری دارد. این معیارها به ما کمک می‌کنند تا کیفیت پیش‌بینی‌های مدل را اندازه‌گیری کرده و نقاط قوت و ضعف آن را شناسایی کنیم.

۱-۸-۲- دقت (Accuracy)

دقت یکی از ساده‌ترین و متداول‌ترین معیارهای ارزیابی در یادگیری ماشین است. دقت به نسبت تعداد پیش‌بینی‌های درست به کل تعداد نمونه‌ها اشاره دارد و به صورت زیر محاسبه می‌شود [۲۱]:

(۲-۱)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

که در آن:

- TP (True Positives): تعداد پیش‌بینی‌های درست مثبت
- TN (True Negatives): تعداد پیش‌بینی‌های درست منفی
- FP (False Positives): تعداد پیش‌بینی‌های نادرست مثبت
- FN (False Negatives): تعداد پیش‌بینی‌های نادرست منفی

مزایا:

- سادگی: دقت به راحتی قابل فهم و محاسبه است.
- استفاده عمومی: در بسیاری از مسائل، دقت یک معیار اولیه برای ارزیابی عملکرد مدل محسوب می‌شود.

معایب:

- عدم کارایی در داده‌های نامتعادل: در شرایطی که توزیع کلاس‌ها نامتعادل باشد، دقت می‌تواند گمراه‌کننده باشد. به عنوان مثال، اگر ۹۵٪ از نمونه‌ها از یک کلاس باشند، مدل می‌تواند با پیش‌بینی همه نمونه‌ها به عنوان کلاس غالب، دقت ۹۵٪ کسب کند، در حالی که عملاً هیچ کارایی نداشته است.

۲-۸-۲- F1-Score

F1-Score یک معیار ارزیابی جامع‌تر است که به تعادل بین دقت و فراخوانی توجه می‌کند F1-Score. به‌ویژه در شرایط نامتعادل اهمیت دارد و به صورت زیر محاسبه می‌شود:

$$F1 = \frac{Precision \times Recall}{Precision + Recall} \times 2 \quad (2-2)$$

که در آن:

دقت (Precision): نسبت پیش‌بینی‌های درست مثبت به کل پیش‌بینی‌های مثبت:

$$Precision = \frac{TP}{TP + FP} \quad (2-3)$$

فراخوانی (Recall): نسبت پیش‌بینی‌های درست مثبت به کل نمونه‌های واقعی مثبت:

$$Recall = \frac{TP}{TP + FN} \quad (2-4)$$

در ارزیابی مدل‌های طبقه‌بندی، F1-Macro و F1-Micro به دو نوع مختلف F1-Score اشاره دارند.

F1-Macro

این معیار به‌طور مستقل برای هر کلاس محاسبه می‌شود و سپس میانگین آن‌ها گرفته می‌شود.

F1-Macro میانگین دقت و فراخوانی برای همه کلاس‌ها را در نظر می‌گیرد:

$$F1-Macro = \frac{1}{N} \sum_{i=1}^N F1_i \quad (2-5)$$

ویژگی‌ها:

- عدم تأثیر از توزیع کلاس‌ها: F1-Macro به‌طور مساوی به همه کلاس‌ها اهمیت می‌دهد.
- استفاده در مسائل چندکلاسی: معمولاً برای ارزیابی مدل‌های چندکلاسی مناسب است.

F1-Micro

این معیار به‌طور کلی از طریق جمع‌آوری تمام پیش‌بینی‌های درست و نادرست در همه کلاس‌ها محاسبه می‌شود:

$$F1-Micro = \frac{2 \times (Precision_{micro} \times Recall_{micro})}{(Precision_{micro} + Recall_{micro})} \quad (2-6)$$

که دقت و فراخوانی میکرو به صورت زیر محاسبه می‌شوند:

$$Precision_{micro} = \frac{TP_{micro}}{TP_{micro} + FP_{micro}} \quad (2-7)$$

$$Recall_{micro} = \frac{TP_{micro}}{TP_{micro} + FN_{micro}} \quad (۸-۲)$$

که در آن:

- TP_{micro} تعداد کل مثبت‌های واقعی در تمام کلاس‌ها
- FP_{micro} تعداد کل مثبت‌های کاذب در تمام کلاس‌ها
- FN_{micro} تعداد کل منفی‌های کاذب در تمام کلاس‌ها

ویژگی‌ها:

- تأثیر از توزیع کلاس‌ها: F1-Micro به کلاس‌هایی که تعداد بیشتری نمونه دارند، اهمیت بیشتری می‌دهد.
- استفاده در مسائل با تمرکز بر دقت کلی: معمولاً در مواردی که به دقت کلی طبقه‌بندی در تمام کلاس‌ها اهمیت بیشتری داده می‌شود، استفاده می‌شود.

۳-۸-۲- معیار ارزیابی سلسله مراتبی

در ارزیابی سیستم‌های طبقه‌بندی، معمولاً از معیارهای دقت و فراخوانی استفاده می‌شود. اما این معیارها برای طبقه‌بندی سلسله‌مراتبی مناسب نیستند زیرا تفاوتی بین انواع خطاهای طبقه‌بندی قائل نمی‌شوند. برای مثال، طبقه‌بندی نادرست به یک نود هم‌سطح یا والد، بهتر از طبقه‌بندی نادرست به یک نود دورتر است. به همین منظور، معیاری بر اساس فاصله معرفی شده است که به محاسبه خطاها در یک درخت سلسله‌مراتبی می‌پردازد. این معیار نه تنها به خطاهای نزدیک‌تر امتیاز کمتری می‌دهد، بلکه از معیارهای دقیق‌تری برای ارزیابی استفاده می‌کند [۲۱].

معیار جدید شامل دقت سلسله‌مراتبی (hP) و یادآوری سلسله‌مراتبی (hR) است که به وسیله افزودن برچسب‌های والد به نتایج محاسبه می‌شود. با این رویکرد، خطاهای سطح بالاتر در سلسله‌مراتب از خطاهای سطح پایین‌تر شدیدتر مجازات می‌شوند [۲۱].

این معیار جدید به راحتی قابل محاسبه است و برای طبقه‌بندی چندبرچسبی در سلسله‌مراتب DAG^۱ فرموله شده است و در مقایسه با معیارهای استاندارد، دقت و تفکیک‌پذیری بالاتری دارد.

^۱ Directed Acyclic Graph

$$hP = \frac{\sum_i |\hat{C}_i \cap \hat{C}'_i|}{\sum_i |\hat{C}'_i|} \quad (2-9)$$

$$hR = \frac{\sum_i |\hat{C}_i \cap \hat{C}'_i|}{\sum_i |\hat{C}_i|} \quad (2-10)$$

$$hF_\beta = \frac{(1+\beta^r).hP.hR}{(\beta^r.hP+hR)} \cdot \beta \in [0, +\infty] \quad (2-11)$$

در این روابط \hat{C}_i مجموعه برچسب‌های پیش‌بینی شده و \hat{C}'_i مجموعه برچسب‌های واقعی با برچسب‌های والد مربوطه هستند [۲۱]. β به عنوان یک پارامتر تعادل استفاده می‌شود که اهمیت دقت و یادآوری را در محاسبه اندازه‌گیری ترکیبی hF تعیین می‌کند [۲۱]. با تنظیم β ، می‌توان تأکید بیشتری بر یکی از این دو معیار گذاشت [۲۱]. برای مثال، اگر β برابر با ۱ باشد، دقت و یادآوری برابر اهمیت دارند، در حالی که اگر β بزرگتر از ۱ باشد، تأکید بیشتری بر یادآوری خواهد بود و اگر کوچکتر از ۱ باشد، دقت اهمیت بیشتری خواهد داشت [۲۱].

فصل ۳:

روش‌های پیشنهادی

۱-۳- مقدمه

این فصل به شرح روش‌های تحقیق استفاده‌شده در این پژوهش می‌پردازد. هدف اصلی این تحقیق، شناسایی تکنیک‌های متقاعدسازی در محتواهای چندرسانه‌ای، به‌ویژه در میم‌ها، با استفاده از مدل‌های یادگیری عمیق است. در این راستا، ما دو زیرمسئله را مورد بررسی قرار می‌دهیم. در هر مسئله داده‌ها، مدل‌های استفاده شده، چالش‌ها و روش‌ها، نحوه‌ی فرایند آموزش مدل بیان می‌شوند.

۱-۱-۳- جمع‌آوری داده‌ها

برای این پژوهش، از داده‌های ارائه‌شده توسط مسابقه SemEval 2024 استفاده شده است. برای دسترسی به داده‌ها لازم است در این مسابقه ثبت نام کنید. استفاده از تصاویر میم‌ها در مستندات علمی مجاز نیست. مجموعه داده شامل سه دسته اصلی است:

۱. مجموعه آموزش: شامل داده‌هایی برای آموزش مدل‌ها.
۲. مجموعه اعتبارسنجی (Dev set): برای ارزیابی مدل‌ها در طول فرآیند آموزش.
۳. مجموعه تست: شامل داده‌هایی که برای ارزیابی نهایی استفاده می‌شود و برچسب‌گذاری نشده است.

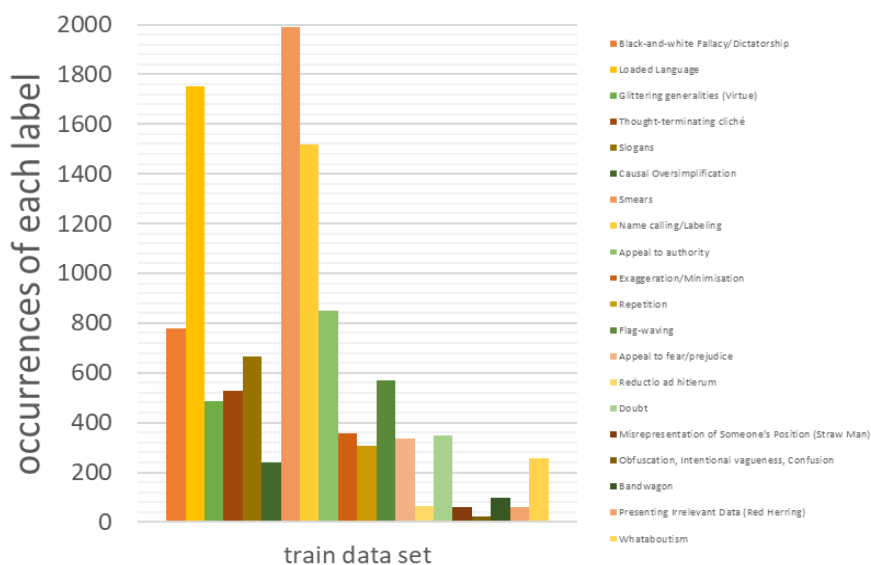
۲-۳- روش‌های پیشنهادی تسک فرعی ۱

تسک فرعی ۱ که به شناسایی تکنیک‌های متقاعدسازی در داده‌های متنی می‌پردازد و مسئله‌ی طبقه‌بندی متنی است. برای دستیابی به اهداف پژوهش، از یک رویکرد تجربی استفاده شده است که در آن داده‌ها جمع‌آوری، پردازش، و مدل‌ها پیاده‌سازی می‌شوند.

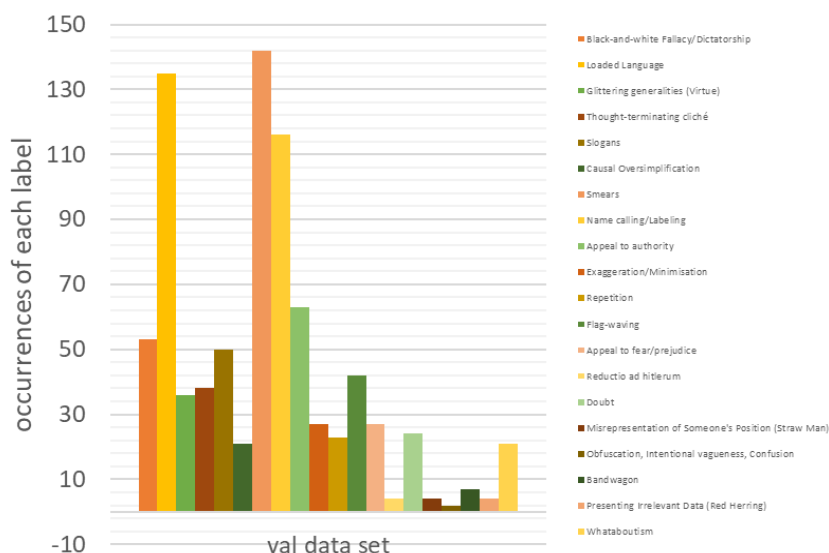
۱-۲-۳- مجموعه داده‌ها

مجموعه داده‌ها در قالب فایل JSON ارائه می‌شود و شامل اطلاعات زیر هستند:

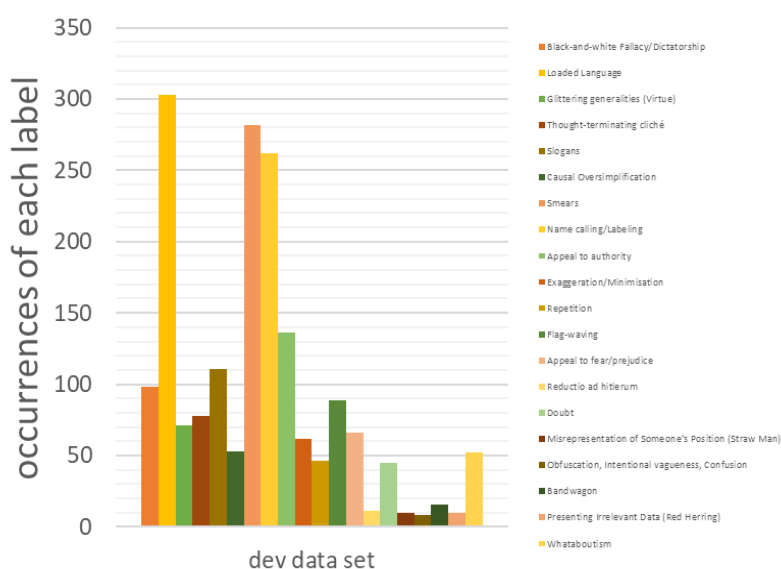
- ID: شناسه منحصر به فرد هر میم.
 - متن: محتوای متنی موجود در میم.
 - برچسب‌ها: لیستی از تکنیک‌های متقاعدسازی موجود در متن و تصویر.
- در این تسک ۷۰۰۰ نمونه در مجموعه داده آموزشی، ۵۰۰ نمونه در مجموعه داده اعتبارسنجی و ۱۰۰۰ نمونه در مجموعه داده توسعه بدون برچسب وجود دارد. در ابتدا، برچسب‌های مجموعه داده توسعه برای استفاده در اهداف آزمون در دسترس نبودند. با این حال، در نهایت، این برچسب‌ها تحت عنوان `dev_gold_labels` به شرکت‌کنندگان به‌طور کامل قابل دسترسی شدند.
- تمام مجموعه داده‌های برچسب‌دار به زبان انگلیسی هستند. برای بهبود روند آموزش همان ابتدا تمام داده‌های هر سه مجموعه را برای آموزش مدل‌ها استفاده کردیم.
- مجموعه داده تست زبان انگلیسی شامل ۱۵۰۰ داده، مجموعه داده تست زبان بلغاری ۴۳۶ داده و مجموعه داده تست زبان مقدونیه شمالی ۲۵۹ داده دارند.
- نمودارهای توزیع داده‌ها به ازای هر برچسب در مجموعه داده‌ی آموزشی در شکل (۳-۱)، مجموعه‌ی اعتبارسنجی در شکل (۳-۲) و مجموعه‌ی توسعه دهنده در شکل (۳-۳) قابل مشاهده هستند.



شکل (۳-۱) نمودار توزیع داده‌ها در مجموعه داده آموزشی



شکل (۲-۳) نمودار توزیع داده‌ها در مجموعه داده اعتبارسنجی



شکل (۳-۳) نمودار توزیع داده‌ها در مجموعه داده توسعه

مجموعه داده‌ها به طور قابل توجهی شامل برچسب‌های «تحقیر» و «زبان مغرضانه» است، که نشان‌دهنده تمرکز بر محتوای احساسی و یا دارای سوگیری منفی است. همچنین، توزیع برچسب‌های مربوط به «نام‌گذاری

^۱ Smears

^۲ Loaded Language

تحقیق‌آمیز» و «استدلال ترس‌زا» نشان می‌دهد که استفاده از تکنیک‌های تحلیل احساسات در تحلیل این داده‌ها می‌تواند ارزشمند باشد. در ادامه، نمونه‌هایی از داده‌ها مورد بررسی قرار می‌گیرد:

شناسه: ۶۳۱۳۵^۲

متن:

"Critical Thinking Essentials: Are my biases affecting how I examine the issue? Am I using information that can be verified with reliable data? Am I basing my position on what I KNOW to be the truth, or what I WANT to be the truth? I might be wrong. (A little humility goes a long way.)"

برچسب‌ها: «تردید»، «شعار»

ترجمه: «اصول تفکر انتقادی: آیا تعصبات من بر نحوه بررسی مسئله تأثیر می‌گذارد؟ آیا از اطلاعاتی استفاده می‌کنم که با داده‌های معتبر قابل تأیید است؟ آیا موضع من بر اساس چیزی است که می‌دانم حقیقت است یا چیزی که می‌خواهم حقیقت باشد؟ شاید اشتباه کنم. (کمی تواضع کمک زیادی می‌کند.)»

این متن به وضوح در حال طرح شک و سوالات پیرامون نحوه پردازش اطلاعات است، که به تکنیک «تردید» مرتبط است. همچنین جمله آخر ("شاید اشتباه کنم.") به عنوان شعاری برای تشویق به تفکر انتقادی محسوب می‌شود که برچسب «شعار» را توجیه می‌کند.

شناسه: ۶۷۳۹۴

متن: "KYLE RITTENHOUSE ALL CHARGES NOT GUILTY"

برچسب‌ها: «کلی‌گویی‌های درخشان (فضیلت)»

ترجمه: «کایل ریتنهاوس: تمام اتهامات بی‌گناه.»

استفاده از جملات کلی و مثبت درباره کسی برای ترویج یک ایده خاص (فضیلت) در اینجا به وضوح مشاهده می‌شود. این متن درباره تبرئه شدن یک فرد است و از زبان مثبت و کلی استفاده می‌کند. این نوع جملات که به تبلیغ فضیلت می‌پردازند، با برچسب کلی‌گویی‌های درخشان مرتبط هستند. استفاده از عمومیات زیبا برای برانگیختن احساسات مثبت و ترویج ایده خاصی در این جمله واضح است.

^۱ Name calling/Labeling

^۲ Fear-Based Appeals

^۳ ID

^۴ Doubt

^۵ Slogans

^۶ Glittering generalities (Virtue)

شناسه : ۶۳۲۹۲

متن:

"This is why we're free, This is why we're safe "

برچسب‌ها: «ساده‌سازی علی»

ترجمه: «این دلیل آزادی ماست، این دلیل امنیت ماست.»

این متن به وضوح دلایل پیچیده‌ای مانند آزادی و امنیت را به یک دلیل ساده تقلیل می‌دهد. ساده‌سازی بیش از حد دلایل پیچیده، دقیقاً همان چیزی است که برچسب «ساده‌سازی علی» نشان می‌دهد.

شناسه: ۷۰۴۱۹

متن:

"if you say we're in the middle of a deadly pandemic but you still support open borders
you're either a liar or a complete moron"

برچسب‌ها: «زبان مغرضانه»، نام‌گذاری تحقیرآمیز، «مغالطه سیاه‌وسفید/دیکتاتوری»، تحقیر

ترجمه: «اگر بگویید ما در میانه یک همه‌گیری مرگبار هستیم، ولی همچنان از مرزهای باز حمایت می‌کنید،
یا دروغ‌گو هستید یا کاملاً احمق.»

این متن از زبان قوی و احساسی استفاده می‌کند که به وضوح با برچسب «زبان مغرضانه» مرتبط است. همچنین، از نام‌گذاری تحقیرآمیز برای توهین به افراد استفاده می‌شود که با برچسب زدن به دیگران همخوانی دارد. استدلال دو قطبی که انتخاب‌ها را به دو گزینه محدود می‌کند (یا دروغ‌گو یا احمق) با «مغالطه سیاه‌وسفید/دیکتاتوری» مرتبط است. در نهایت، برچسب «تحقیر» به دلیل استفاده از زبان توهین‌آمیز برای تحقیر مناسب است.

۲-۲-۳- پیش‌پردازش داده‌ها

۱-۲-۲-۳- پردازش متن با استفاده از NLTK

پس از پیش‌پردازش اولیه، از ابزار NLTK برای انجام پاک‌سازی‌های بیشتر بر روی داده‌های متنی استفاده می‌کنیم. این فرآیند شامل حذف علائم نگارشی، تبدیل تمام متن به حروف کوچک، و انجام مراحل مختلف

^۱ Causal Oversimplification

^۲ Black-and-white Fallacy/Dictatorship

توکن‌سازی و لماتیزه کردن برای استانداردسازی بیشتر متن‌ها است. با استفاده از NLTK، اطمینان حاصل می‌شود که داده‌های متنی تا حد امکان بهینه و آماده برای استفاده در مدل‌های یادگیری عمیق هستند.

۳-۲-۳- پیاده‌سازی مدل‌ها

برای استخراج ویژگی‌های متنی، مدل‌های زبانی پیش‌آموزش داده‌شده bert، XLM-RoBERTa و GPT-2 روی داده‌های متنی میم‌ها fine-tune شد. در هر دو مدل در انتها یک لایه ی دراپ اوت و بعد از آن لایه کاملاً متصل با ۲۰ نود و تابع فعال‌سازی سیگموئید برای طبقه‌بندی استفاده شد.

۳-۲-۴- چالش‌های روش پیشنهادی

۱-۴-۲-۳- عدم توازن داده‌ها

در این تسک هر داده می‌تواند برچسبی نداشته باشد و یا یک یا چند برچسب داشته باشد. همانطور که در جدول (۱-۳) قابل مشاهده است توزیع داده‌ها در مجموعه داده‌های استفاده شده (اجتماعی از داده‌های آموزشی، اعتبارسنجی و توسعه) یکسان نیست و این موضوع موجب ایجاد بایاس در خروجی مدل خواهد شد.

جدول (۳-۱) توزیع مجموعه داده‌ها تسک فرعی ۱

تعداد تکرار	شماره برچسب
۳۷۴	۰
۳۱	۱
۸۲۸	۲
۱۲۰	۳
۱۰۴۹	۴
۷۰۲	۵
۴۳۰	۶
۳۱۴	۷
۹۳۱	۸
۶۴۴	۹
۷۶	۱۰
۷۳	۱۱
۳۳۱	۱۲
۵۹۵	۱۳
۴۱۹	۱۴
۱۸۹۶	۱۵
۲۴۱۴	۱۶
۷۸	۱۷
۴۴۵	۱۸
۲۱۸۸	۱۹

برای حل این مشکل از `class_weight_dict` استفاده شده‌است. با توجه به تعداد برای هر لیبل احتمال وجود و عدم وجود در نظر گرفته می‌شود و این دیکشنری هنگام آموزش مدل‌ها در محاسبه تابع ضرر استفاده می‌شود.

۲-۴-۳- طبقه بندی چندبرچسبی سلسله مراتبی

طبقه‌بندی چندبرچسبی سلسله‌مراتبی با چالش‌های خاص خود همراه است. یکی از این چالش‌ها، تعیین اینکه چگونه می‌توان به‌درستی برچسب‌های والد و فرزند را شناسایی کرد و چه زمانی یک نمونه باید به یک گره والد یا فرزند نسبت داده شود. به‌علاوه، وجود عدم تعادل در تعداد نمونه‌های هر کلاس می‌تواند به پیچیدگی‌های بیشتری در فرآیند یادگیری منجر شود [۲۱].

برای حل این چالش‌ها، از تکنیک‌های مختلفی مانند روش‌های یادگیری عمیق، الگوریتم‌های درخت تصمیم،

و روش‌های مبتنی بر فیلتر استفاده می‌شود. به عنوان مثال، الگوریتم‌های درخت تصمیم می‌توانند با استفاده از ویژگی‌های داده‌ها، به صورت خودکار گره‌های سلسله‌مراتبی را ایجاد کنند و سپس بر اساس این ساختار به طبقه‌بندی نمونه‌ها بپردازند.

۵-۲-۳- فرایند آموزش مدل

ما مدل خود را با استفاده از ترکیبی از تکنیک‌های یادگیری نظارت‌شده و تنظیم دقیق آموزش دادیم. مدل‌ها را با داده‌های آموزشی که با داده‌های اعتبارسنجی ترکیب شده است، آموزش داده‌ایم. برای مقابله با عدم توازن کلاس‌ها، از class-weight در کنار BCE^۱ استفاده کرده و از بهینه‌ساز AdamW برای بهینه‌سازی نزول گرادیان بهره می‌بریم. ابرپارامترهایی مانند نرخ یادگیری، اندازه دسته‌ها و نرخ دراپ‌آوت با استفاده از جستجوی شبکه‌ای و اعتبارسنجی متقابل روی مجموعه توسعه تنظیم می‌شوند.

۶-۲-۳- ارزیابی نتایج

عملکرد سیستم ما با استفاده از معیارهای ارزیابی استاندارد مانند امتیاز F1-macro در مجموعه آزمایش توسط در وبگاه مسابقه ارزیابی می‌شود. ما نتایج خود را با مدل‌های پایه مقایسه می‌کنیم تا اثربخشی رویکرد خود را ارزیابی کنیم. بهترین نتیجه را مدل GPT-2 کسب کرد.

۳-۳- روش‌های پیشنهادی تسک فرعی ۲ب

تسک فرعی ۲ب، به تحلیل داده‌های تصویری و متنی در قالب میم‌ها و شناسایی وجود یا عدم وجود تکنیک‌های متقاعدسازی می‌پردازد. این مسئله طبقه‌بندی چند رسانه‌ای محسوب می‌شود.

^۱ Binary Cross-Entropy

۱- ۳-۳- مجموعه داده‌ها

مجموعه داده‌ها در قالب فایل JSON ارائه می‌شود و شامل اطلاعات زیر هستند:

- ID: شناسه منحصر به فرد هر میم.
 - متن: محتوای متنی موجود در میم.
 - تصویر: نام فایل تصویر مربوط به میم.
 - برچسب: Propagandistic یا Non-Propagandistic
- همان‌طور که در جدول (۳-۲) نشان داده شده است، در این تسک ۱۲۰۰ نمونه در مجموعه داده آموزشی، ۱۵۰ نمونه در مجموعه داده اعتبارسنجی و ۳۰۰ نمونه در مجموعه داده توسعه یا dev_gold_labels وجود دارد. تعداد نمونه‌های هر برچسب، در هر سه مجموعه توزیع مشابهی دارند.

جدول (۳-۲) توزیع مجموعه داده‌ها تسک فرعی ۲ب

Data Set/Label	Propagandistic	Non-Propagandistic
Train	۸۰۰	۴۰۰
Validation	۱۰۰	۵۰
Development	۲۰۰	۱۰۰

تمام مجموعه داده‌های برچسب‌دار به زبان انگلیسی هستند. مجموعه داده تست زبان انگلیسی شامل ۶۰۰ داده، مجموعه داده تست زبان بلغاری ۱۰۰ داده و مجموعه داده تست زبان مقدونیه شمالی ۱۰۰ داده دارند.

۲- ۳-۳- پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها مرحله‌ای حیاتی در فرآیند تحقیق است. به طور کلی از روش‌های زیر استفاده کردیم:

۱. تصاویر: پردازش تصاویر شامل تغییر اندازه به ۲۲۴ در ۲۲۴ در حالت RGB و نرمال‌سازی داده‌های بصری برای استفاده در مدل‌های بینایی ماشین است.

۲. تجزیه و تحلیل زبانی: استفاده از ابزارهای^۱ NLTK برای پردازش و نرمال‌سازی داده‌های متنی. این شامل توکن‌سازی، حذف کلمات توقف و ریشه‌یابی است.
۳. پاکسازی متن: حذف نویز و اطلاعات غیرضروری از محتوای متنی. در این مرحله، از API OpenAI برای حذف اطلاعات اضافی مانند تاریخ‌ها، نام‌های کاربری و سایر جزئیات استفاده می‌شود.

۱-۲-۳-۳ API OpenAI برای پیش‌پردازش اولیه

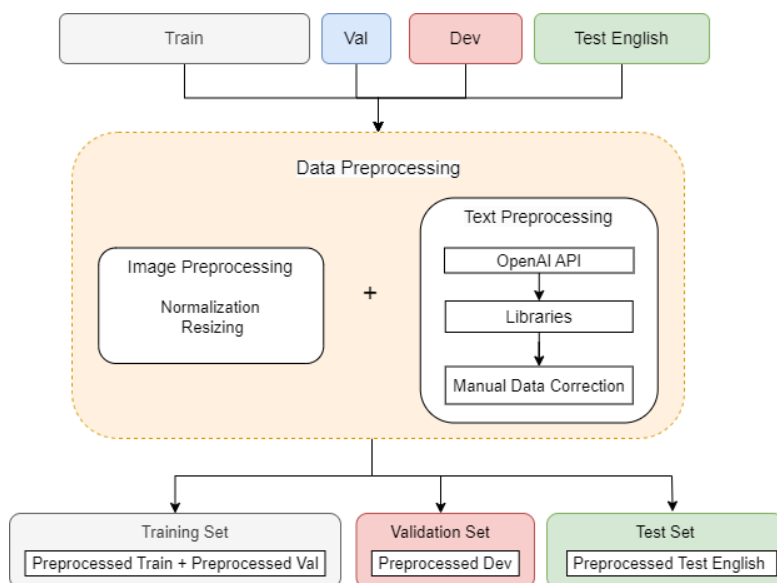
ما از API OpenAI برای پیش‌پردازش اولیه متن استفاده می‌کنیم و از قابلیت‌های پیشرفته پردازش زبان طبیعی آن برای مدیریت چالش‌های رایج در استخراج متن میم‌ها بهره می‌بریم. این API به‌طور مؤثر اطلاعات اضافی مانند تاریخ‌ها، نام کاربری‌ها و متن‌های اضافی که ممکن است همراه با محتوای اصلی میم باشد را شناسایی و حذف می‌کند. با بهره‌گیری از قدرت API OpenAI، اطمینان حاصل می‌کنیم که داده‌های متنی وارد شده به سیستم ما تمیز و عاری از نویزهای غیرضروری باشد.

در پیاده‌سازی، ابتدا دستورالعمل خاصی برای استخراج متن اصلی و حذف اطلاعات اضافی مانند نام کاربری و کاراکترهای اضافی تعریف می‌شود. سپس از طریق یک درخواست HTTP به API OpenAI، درخواست ارسال شده و متن پیش‌پردازش‌شده از پاسخ دریافت می‌گردد. این متن پردازش‌شده سپس در یک فایل ذخیره می‌شود تا به‌عنوان داده تمیزشده برای مراحل بعدی مدل‌سازی و تحلیل استفاده شود.

۲-۲-۳-۳ پردازش بیشتر متن با استفاده از NLTK

پس از پیش‌پردازش اولیه، از ابزار NLTK برای انجام پاک‌سازی‌های بیشتر بر روی داده‌های متنی استفاده می‌کنیم. این فرآیند شامل حذف علائم نگارشی، تبدیل تمام متن به حروف کوچک، و انجام مراحل مختلف توکن‌سازی و لماتیزه کردن برای استانداردسازی بیشتر متن‌ها است. با استفاده از NLTK، اطمینان حاصل می‌شود که داده‌های متنی تا حد امکان بهینه و آماده برای استفاده در مدل‌های یادگیری عمیق هستند.

در شکل (۳-۱)، ساختار داده و مراحل پیش‌پردازش به کاررفته در روش پژوهش گفته شده نشان داده شده است.



شکل (۳-۴) ساختار داده و مراحل پیش‌پردازش به کاررفته [۲]

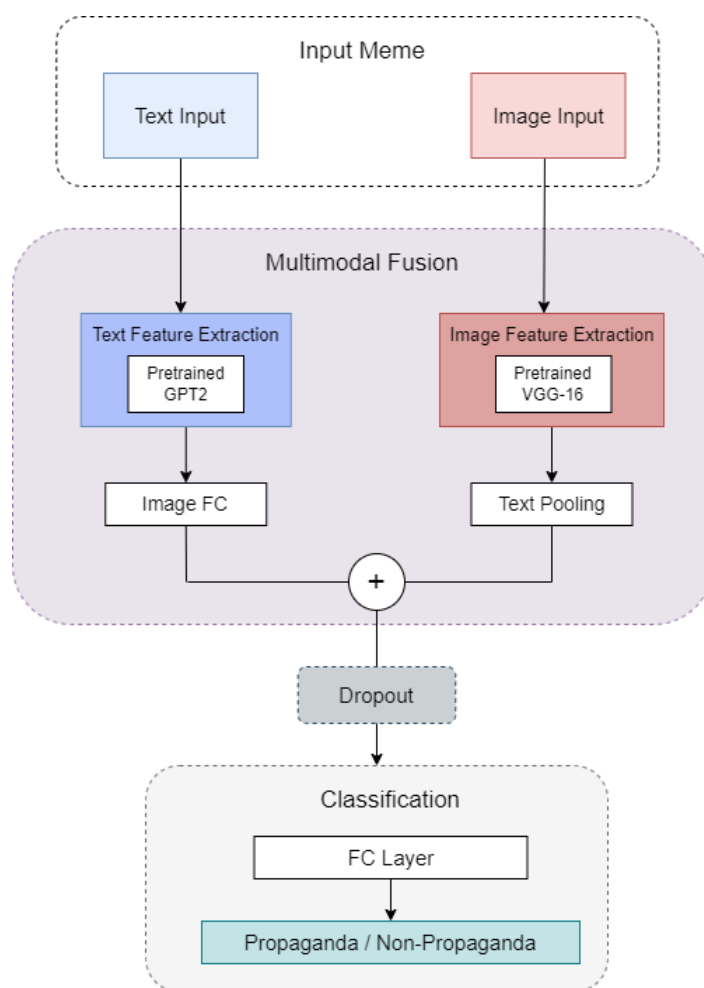
۳-۳-۳- استخراج ویژگی‌ها

برای استخراج ویژگی‌های متنی، مدل‌های زبانی پیش‌آموزش داده شده GPT-2 و XLM-RoBERTa روی داده‌های متنی میم‌ها fine-tune شد. این مدل‌ها اطلاعات معنایی و نحوی موجود در محتوای متنی را دریافت می‌کنند و به یادگیری نمایشی مؤثر برای وظایف بعدی کمک می‌کنند. برای استخراج ویژگی‌های تصویری، از شبکه‌ی عصبی کانولوشنی VGG و ترانسفورمر بصری ViT استفاده شد تا ویژگی‌های بصری از میم‌ها استخراج شود. ویژگی‌های استخراج شده از هر دو بخش به هم متصل می‌شوند تا یک نمایه چندوجهی از میم‌ها ایجاد شود.

۳-۳-۴- پیاده‌سازی مدل

معماری مدل ما شامل یک لایه فیوژن چندوجهی و پس از آن یک لایه طبقه‌بندی است. لایه ادغام چندوجهی ویژگی‌های متنی و تصویری را از طریق اتصال ترکیب می‌کند تا اطلاعات هر دو بخش را یکپارچه سازد. لایه طبقه‌بندی از یک رویکرد طبقه‌بندی دودویی برای پیش‌بینی وجود یا عدم وجود تکنیک‌های متقاعدسازی در میم‌ها استفاده می‌کند. در شکل (۳-۲)، معماری بهترین مدل ما که ترکیبی از VGG-16 و GPT-2 برای

تسک فرعی ۲ است، ارائه شده است.



شکل (۳-۵) معماری مدل ترکیب شده از VGG-16 و GPT-2 برای تسک فرعی ۲ [۲]

۵-۳-۳- چالش‌های روش پیشنهادی

۱-۵-۳-۳- چالش داده‌ها

در ابتدای کار، ما از همان مجموعه داده آموزشی ارائه‌شده برای آموزش مدل اولیه خود استفاده کردیم. با این حال، متوجه شدیم که دقت مدل بر روی داده‌های آموزشی پس از ۲ دوره به ۹۰٪ رسید، اما دقت بر روی داده‌های اعتبارسنجی چندان امیدوارکننده نبود. با وجود تنظیم ابر پارامترها، این تفاوت در دقت بهبود نیافت. بنابراین، تصمیم گرفتیم با استفاده از کل مجموعه داده به آموزش مدل‌های خود ادامه دهیم. این مجموعه داده گسترش‌یافته به‌طور قابل‌توجهی عملکرد مدل را بر روی داده‌های آزمایش بهبود بخشید.

۲-۵-۳-۳- بیش برآزش

در مرحله توسعه، با چالش‌هایی مرتبط با بیش‌برآزش مدل مواجه شدیم، به‌ویژه زمانی که از معماری‌های پیچیده مانند ترکیب XLM-RoBERTa برای پردازش متن و VGG برای تحلیل تصاویر استفاده می‌کردیم. بدون به‌کارگیری نرمال‌سازی مناسب، مدل اولیه ما نشانه‌هایی از بیش‌برآزش را نشان داد که قابلیت تعمیم‌دهی آن را کاهش می‌داد. برای حل این مشکل، تکنیک‌های منظم‌سازی مانند لایه‌های دراپ‌آوت را پیاده‌سازی کردیم تا از بیش‌برآزش جلوگیری کنیم و به استحکام مدل بیفزاییم. این اقدامات نقش مهمی در تثبیت فرایند آموزش و بهبود عملکرد کلی سیستم ما داشتند.

۶-۳-۳- فرایند آموزش مدل

ما مدل خود را با استفاده از ترکیبی از تکنیک‌های یادگیری نظارت‌شده و تنظیم دقیق آموزش دادیم. مدل‌ها را با داده‌های آموزشی که با داده‌های اعتبارسنجی ترکیب شده است، آموزش داده‌ایم. برای مقابله با عدم توازن کلاس‌ها، از تابع زیان Focal Loss در کنار BCE^۱ استفاده می‌کنیم و از بهینه‌ساز AdamW برای بهینه‌سازی نزول گرادیان بهره می‌بریم. ابرپارامترهایی مانند نرخ یادگیری، اندازه دسته‌ها و نرخ دراپ‌آوت با استفاده از جستجوی شبکه‌ای و اعتبارسنجی متقابل روی مجموعه توسعه تنظیم می‌شوند.

۷-۳-۳- ارزیابی نتایج

عملکرد سیستم ما با استفاده از معیارهای ارزیابی استاندارد مانند امتیاز F1-macro در مجموعه آزمایش توسط در وبگاه مسابقه ارزیابی می‌شود. ما نتایج خود را با مدل‌های پایه مقایسه می‌کنیم تا اثربخشی رویکرد خود را ارزیابی کنیم.

^۱ Binary Cross-Entropy

فصل ۴:

نتایج و تفسیر آنها

۱-۴- نتایج تسک فرعی ۱

۱-۱-۴- نتایج

جدول (۱-۴) هایپرپارامترهای استفاده شده در آموزش مدل انتخاب شده را نشان می‌دهد.

جدول (۱-۴) ابرپارامترهای بهترین مدل در تسک فرعی ۱

پارامتر	مقدار
تعداد دوره آموزشی	۵
اندازه دسته آموزشی	۱۶
اندازه دسته اعتبارسنجی	۱۶
نرخ یادگیری	۰,۰۰۱

در جدول (۲-۴) نتایج مدل‌های مختلف نشان داده شده است. baseline حالتی را در نظر می‌گیرد که مدل برای همه داده‌ها برچسب "Smears" که بیشترین داده‌های آموزشی به از آن نوع هستند را خروجی داده باشد.

جدول (۲-۴) نتایج مجموعه تست به زبان انگلیسی در تسک فرعی ۱

Model	Hierarchical F1	Hierarchical Precision	Hierarchical Recall
XLM-RoBERTa	۰/۴۲۶۶۷	۰/۲۹۲۴۷	۰/۷۷۷۹۸
XLM-RoBERTa with best threshold	۰/۵۰۵۷۳	۰/۴۷۵۵۵	۰/۵۴۰۰۱
GPT-2	۰/۴۰۲۶۶	۰/۲۶۹۸۶	۰/۷۹۲۸۱
GPT-2 with best threshold	۰/۵۹۷۲۷	۰/۵۲۵۶۳	۰/۶۹۱۵۲
Baseline	۰/۳۶۸۶۵	۰/۴۷۷۱۱	۰/۳۰۰۳۶

بهترین مدل XLM-RoBERTa و GPT-2 بر اساس Hierarchical F1 روی داده‌های تست به زبان‌های بلغاری و مقدونیه شمال آزمایش کردیم. نتایج در جدول (۳-۴) ثبت شدند.

جدول (۴-۳) نتایج مجموعه تست به زبان‌های بلغاری و مقدونیه شمالی در تسک فرعی ۱

Language	Model	Hierarchical F1	Hierarchical Precision	Hierarchical Recall
Bulgarian	XLM-RoBERTa	۰/۳۸۴۳۳	۰/۳۸۶۷۲	۰/۳۸۱۹۷
	GPT-2	۰/۳۳۰۷۰	۰/۲۰۹۵۳	۰/۷۸۴۱۸
	Baseline	۰/۲۸۳۷۷	۰/۳۱۸۸۱	۰/۲۵۵۶۷
North Macedonian	XLM-RoBERTa	۰/۳۰۸۶۳	۰/۱۸۸۹۱	۰/۸۴۲۵۶
	GPT-2	۰/۳۱۲۲۲	۰/۲۸۹۰۱	۰/۳۳۹۴۸
	Baseline	۰/۳۰۶۹۲	۰/۳۱۴۰۳	۰/۳۰۰۱۲

۲-۱-۴- تحلیل نتایج

در زبان انگلیسی مدل GPT-2 با تکنیک انتخاب بهترین آستانه برای هر برچسب توانست نتیجه‌ی بهتری کسب نماید. هیچ کدام از مدل‌ها در زبان مقدونیه‌ای نتیجه‌ی مناسبی کسب نکردند. بر اساس معیار Hierarchical F-1 مدل GPT-2 با مقدار ۰/۳۱۲۲۲ عملکرد بهتری داشته است. در زبان بلغاری مدل XLM-RoBERTa با مقدار دقت ۰/۳۸۴۳۳ نتیجه‌ی بهتری داشته است.

۲-۴- نتایج تسک فرعی ۲ب

۱-۲-۴- نتایج

در مجموعه‌داده آزمایشی زبان انگلیسی، امتیاز F1-macro برابر با ۰/۶۷ و امتیاز F1-micro برابر با ۰/۷۴ حاصل شد. نتایج ارزیابی چهار ترکیب مختلف از مدل‌ها، با استفاده از بهترین آستانه ممکن بر اساس F1-macro در مجموعه‌داده توسعه زبان انگلیسی، در جدول (۴-۴) ارائه شده است. این ترکیب‌ها شامل VIT + GPT-2 و VIT + XLM-RoBERTa، VGG+GPT-2، VGG+XLM-RoBERTa هستند.

جدول (۴-۴) نتایج مجموعه اعتبارسنجی در تسک فرعی ۲ ب

Model	F1-macro	F1-macro Best Treshold
VGG+XLM-RoBERTa	۰/۵۸	۰/۶۳
VGG+GPT-2	۰/۷۱	۰/۷۶
ViT + XLM-RoBERTa	۰/۴۰	۰/۵۳
ViT + GPT-2	۰/۳۵	۰/۵۱

جدول (۴-۵) نتیجه ی آزمایش مدل آموزش دیده ی VGG+GP-2 روی داده‌های تست زبان‌های انگلیسی، بلغاری و مقدونیه‌ای را نشان می‌دهد.

جدول (۴-۵) نتایج خروجی بهترین مدل روی مجموعه تست زبان‌های انگلیسی، بلغاری و مقدونیه‌ای در تسک فرعی ۲ ب

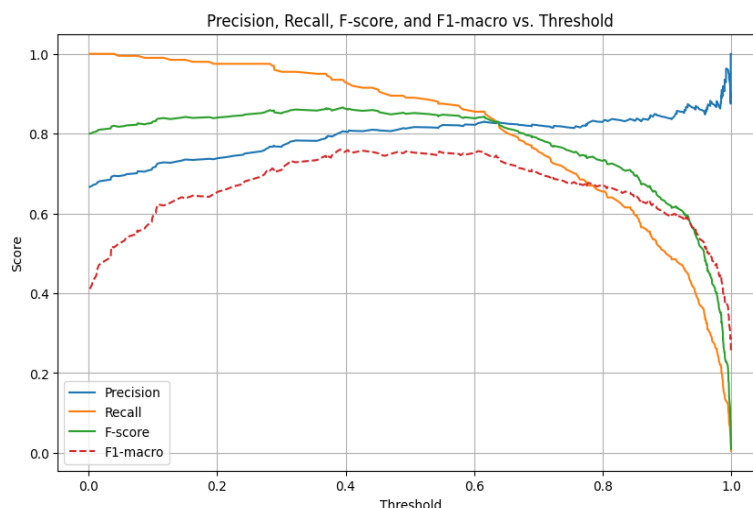
Language	F1-macro	Baseline F1-macro	F1-micro	Baseline F1-micro
English	۰/۶۷۳۹۸	۰/۲۵۰۰۰	۰/۷۴۰۰۰	۰/۳۳۳۳۳
Bulgarian	۰/۵۱۶۳۷	۰/۱۶۶۶۷	۰/۷۴۰۰۰	۰/۲۰۰۰۰
North Macedonian	۰/۵۷۶۵۳	۰/۰۹۰۹۱	۰/۷۹۰۰۰	۰/۱۰۰۰۰

تنظیم ابرپارامترها نقش حیاتی در بهینه‌سازی عملکرد مدل ایفا کرد. ما با پارامترهای مختلفی از جمله نرخ‌های یادگیری، اندازه‌های دسته‌ای و آستانه‌ها آزمایش کردیم تا بهترین پیکربندی را پیدا کنیم. پارامترهای آموزشی بیشتری در جدول (۴-۶)، علاوه بر پارامترهای ذکر شده در بالا، مشخص شده‌اند.

جدول (۴-۶) ابرپارامترهای بهترین مدل در تسک فرعی ۲ ب

پارامتر	مقدار
تعداد دوره آموزشی	۱۰
اندازه دسته آموزشی	۳۲
اندازه دسته اعتبارسنجی	۳۲
وزن کاهشی	۰,۰۰۱
نرخ یادگیری	۰,۰۰۱
بهترین آستانه	۰,۳۹

علاوه بر این، در شکل (۴-۱) تأثیر آستانه‌ها بر معیارهای دقت، فراخوانی، F-score و F-macro به‌صورت تصویری نمایش داده شده است. این تحلیل بینش‌هایی در مورد تبادلهای بین این معیارها فراهم می‌کند و انتخاب یک آستانه بهینه برای ارزیابی مدل و تصمیم‌گیری را راهنمایی می‌کند.



شکل (۴-۱) دقت، فراخوانی، نمره F-score و F-macro در مقابل آستانه در مجموعه توسعه [۲]

۲-۲-۴- تحلیل نتایج

برای به دست آوردن درک عمیق‌تری از عملکرد سیستم خود، مطالعات حذف انجام دادیم و نتایج طراحی مختلف را مقایسه کردیم تا بهترین پیکربندی‌ها را شناسایی کنیم. ما از کل مجموعه داده آموزشی برای این تحلیل‌ها استفاده کردیم و از ترکیبی از داده‌های آموزشی، اعتبارسنجی و داده‌های بدون برچسب برای اهداف آموزش و اعتبارسنجی بهره بردیم.

از طریق آزمایش‌های سیستماتیک، مشاهده کردیم که استفاده از فوکال لاس^۱ همراه با فعال‌سازی دوتایی سیگموئید، به طرز قابل توجهی عملکرد مدل را بهبود بخشید. علاوه بر این، آموزش مدل با استفاده از داده‌های اعتبارسنجی و توسعه به عنوان یک مجموعه اعتبارسنجی اضافی، منجر به افزایش قابل توجهی در دقت شد.

^۱ ablation studies

^۲ gold_unlabeled

^۳ focal loss

فصل ۵:

جمع‌بندی و پیشنهادها

۱-۵- جمع‌بندی

در تسک ۱، در زبان انگلیسی مدل GPT-2 با تکنیک انتخاب بهترین آستانه برای هر برچسب توانست نتیجه‌ی بهتری کسب نماید. هیچ کدام از مدل‌ها در زبان مقدونیه‌ای نتیجه‌ی مناسبی کسب نکردند. بر اساس معیار Hierarchical F-1 مدل GPT-2 با مقدار ۰/۳۱۲۲۲ عملکرد بهتری داشته است. همچنین در زبان بلغاری مدل XLM-RoBERTa با مقدار دقت ۰/۳۸۴۳۳ نتیجه‌ی بهتری داشته است.

نتایج تسک ۲ نشان می‌دهد که ترکیب‌های مختلف از مدل‌ها در زبان انگلیسی عملکردهای متفاوتی داشته‌اند. بهترین نتیجه در مجموعه داده انگلیسی با ترکیب VGG و GPT-2 حاصل شده است. این نتایج نشان می‌دهد که این ترکیب برای داده‌های انگلیسی توانایی بهتری در تشخیص و طبقه‌بندی دقیق تکنیک‌های متقاعدسازی داشته است.

در مقابل، ترکیب‌های دیگری مانند ViT و XLM-RoBERTa و ViT و GPT-2 عملکرد کمتری داشته‌اند، که ممکن است به دلیل عدم هماهنگی مناسب بین ویژگی‌های استخراج شده از تصویر و متن باشد.

برای داده‌های بلغاری و مقدونیه‌ای، با وجود اینکه ترکیب VGG و GPT-2 عملکرد خوبی در داده‌های انگلیسی داشت، نتایج در این زبان‌ها به‌طور کلی پایین‌تر بود. این کاهش عملکرد در زبان‌های کمتر رایج نشان می‌دهد که مدل‌ها ممکن است نیاز به تنظیمات یا داده‌های آموزشی بیشتری برای مقابله با تفاوت‌های زبانی و فرهنگی داشته باشند. به‌طور کلی، استفاده از GPT-2 بهبود قابل توجهی در عملکرد مدل‌ها در زبان انگلیسی و سایر زبان‌ها ارائه داده است، اما چالش‌های مرتبط با داده‌های چندزبانه همچنان وجود دارد.

۲-۵- پیشنهادها

در راستای بهبود نتایج به‌دست‌آمده، کارهای آینده می‌تواند بر پیش‌آموزش مدل‌ها بر روی داده‌های خاص میم و اصلاح تکنیک‌های پیش‌پردازش برای متن‌های استخراج‌شده متمرکز شود. به‌ویژه، پیشنهاد می‌شود که:

- پیش‌آموزش بر روی داده‌های میم خاص: جمع‌آوری و استفاده از مجموعه داده‌های بزرگ‌تر و متنوع‌تر از میم‌ها که شامل انواع مختلف متون و تصاویر باشد، می‌تواند به بهبود کیفیت ویژگی‌های استخراج‌شده کمک کند.
- توسعه تکنیک‌های پیش‌پردازش: بهبود تکنیک‌های پیش‌پردازش به‌خصوص در حذف نویز و اطلاعات

اضافی از متن‌ها می‌تواند دقت مدل را افزایش دهد. پیشنهاد می‌شود از تکنیک‌های پردازش زبان طبیعی پیشرفته‌تری برای بهبود کیفیت متن‌های استخراج‌شده استفاده شود.

- بهینه‌سازی معماری مدل: آزمایش با معماری‌های مختلف مدل و ترکیب‌های جدید می‌تواند به شناسایی ساختار بهینه برای پردازش چندرسانه‌ای کمک کند. به عنوان مثال، بررسی تأثیر مدل‌های جدیدتر و پیشرفته‌تر در پردازش تصویر و متن می‌تواند مفید باشد.
- تحلیل تأثیر پارامترها: انجام مطالعات عمیق‌تر بر روی تأثیر هایپرپارامترها بر روی عملکرد مدل و استفاده از روش‌های بهینه‌سازی پیشرفته مانند جستجوی شبکه‌ای و بهینه‌سازی بی‌قاعده برای تنظیم بهینه پارامترها.
- اصلاح مجموعه داده‌ها: استفاده از روش‌های دیگر برای بدست آوردن متن میم‌ها
- تحقیقات میان‌رشته‌ای: گنجاندن دیدگاه‌های میان‌رشته‌ای از حوزه‌های مختلف مانند روانشناسی و علوم اجتماعی در طراحی مدل و استراتژی‌های تحلیل داده، می‌تواند به درک بهتر از تأثیرات تکنیک‌های متقاعدسازی و بهبود دقت تشخیص کمک کند.

مراجع

- [1] Propaganda, "SemEval 2024 Task 4: Multilingual Detection of Persuasion Techniques in Memes," 2024.
- [2] Bakhshande, F., Naderi, M. "CVcoders on SemEval-2024 Task 4." In Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), June 2024, Mexico City, Mexico, pp. 1912–1918.
- [3] Dimitrov, Dimitar, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. "SemEval-2021 Task 6: Detection of Persuasion Techniques in Texts and Images." Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2021), 2021, pp. 70–98.
- [4] Jowett, Garth S., and Victoria O'Donnell. *Propaganda & Persuasion*. 6th ed., SAGE Publications, 2019.
- [5] Vaswani, Ashish, et al. "Attention Is All You Need. ArXiv, 2017.
- [6] Brown, Tom B., et al. "Language Models Are Few-Shot Learners (GPT-3)." ArXiv, 2020.
- [7] Better Language Models and Their Implications, OpenAI Blog, 2019.
- [8] Karpathy, Andrej. "Let's Reproduce GPT-2 (124M)." YouTube.
- [9] Karpathy, Andrej. "nanoGPT" repo.
- [10] OpenAI. (2023). GPT-2. GitHub. Archived from the original on March 11, 2023.
- [11] Hsinhung, W. (n.d.). *GPT-2 detailed model architecture*. Medium.
- [12] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2019). *Unsupervised Cross-lingual Representation Learning at Scale*.
- [13] Lample, G., & Conneau, A. (2019). "Cross-lingual Language Model Pretraining."
- [14] Conneau, A., et al. (2020). "Unsupervised Cross-lingual Representation Learning at Scale (XLM-R)."
- [15] Simonyan, K., & Zisserman, A. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv preprint arXiv:1409.1556*.
- [16] Deng, J., et al. (2009). "ImageNet: A large-scale hierarchical image database." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 248–255.
- [17] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. [Chapter 9: Convolutional Networks].
- [18] Dosovitskiy, A., et al. (2020). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *arXiv preprint arXiv:2010.11929*.
- [19] Carion, N., et al. (2020). "End-to-End Object Detection with Transformers." European Conference on Computer Vision (ECCV), 2020.
- [20] Touvron, H., et al. (2021). "Training data-efficient image transformers & distillation through attention." *arXiv preprint arXiv:2012.12877*.
- [21] S. Kiritchenko, S. Matwin, R. Nock, and A. F. Famili, "Learning and Evaluation in the Presence of Class Hierarchies: Application to Text Categorization," in *Proceedings of the National Research Council Canada*, University of Ottawa, Canada, 2004.
- [22] Propaganda, "SemEval 2023 Task 3: Persuasion Techniques in Texts and Images," 2023.
- [23] Holada, M., Potapcova, T., Tamasi, K., Mikolov, S., Yang, X., & Blashkova, V. (2023). KInITVeraAI at SemEval-2023 Task 3: Simple yet Powerful Multilingual Fine-Tuning for Persuasion Techniques Detection. *arXiv:2304.11924*.
- [24] M. Bach, T. Minervini, S. Pradhan, and E. Hovy, "Evolving Multimodal Models: Detection of Persuasion Techniques with Limited Labeled Data," ArXiv, 2023.

Abstract:

This thesis examines persuasion techniques in memes. This research was conducted within the framework of the CVcoders team participating in Subtask 1 and 2b of Task 4 in the SemEval2024 competition, which focuses on identifying psychological and rhetorical persuasion methods in multilingual and multimodal content. For both tasks, we utilized the datasets provided by the SemEval competition, and to enhance model performance in the face of class imbalance, we employed advanced techniques such as Focal Loss. The models were trained solely on English data and were ultimately tested on data in North Macedonian, Bulgarian, and English.

In Subtask 1, which consisted only of textual data and 20 different classifications, we utilized pre-trained models XLM-RoBERTa and GPT-2. The results indicate that the GPT-2 model performed better according to the Hierarchical F1 evaluation metric in the English language. In Subtask 2b, both textual and visual data were examined with two different classes. For this purpose, we used a combination of pre-trained text and image models, including XLM-RoBERTa, GPT-2, VGG, and ViT, for classification. In this section, multimodal data consisting of texts and images were analyzed to determine whether persuasion techniques were used in each meme. The combination of the VGG and GPT-2 models yielded the best performance in this context.

Keywords: persuasion techniques, propaganda, natural language processing, deep learning, image processing, SemEval 2024, XLM-RoBERTa, GPT-2, Focal Loss.



Iran University of Science and Technology
School of Computer Engineering

Detecting Persuasion Techniques in Memes Using Deep Learning

Bachelor of Science Thesis in Computer Engineering

By:
Mahdieh Naderi

Supervisor:
Dr. Sayyed Sauleh Eetemadi

October 2024