



دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

پایان نامه کارشناسی
مهندسی کامپیوتر

تشخیص اهمیت اخبار فارسی با استفاده از مدل‌های زبانی بزرگ

نگارش

شایان صالحی

استاد راهنما

دکتر مهدی جعفری

بهمن ۱۴۰۴

سپاس

از استاد بزرگوارم، دکتر جعفری به خاطر زحمات و راهنمایی‌هایی که در طول این پروژه داشته‌اند
متشکرم و همچنین از دانشجوی دکترا ایشان، آقای معین سلیمی به خاطر زمان و راهنمایی‌هایی که برای
پیش‌بردن پروژه انجام داده‌اند قدردانم.

چکیده

این پروژه به بررسی قدرت تشخیص اهمیت یک خبر فارسی توسط مدل‌های زبانی بزرگ پرداخته و قدرت یادگیری از محتوا، قدرت استدلال و قدرت تفکر آن را ارزیابی کرده است. در ابتدا، از دادگان علائم‌گذاری‌شده توسط افراد در حوزه‌های مختلف از جمله ورزشی، سیاسی، اجتماعی، پزشکی و فرهنگی استفاده و محیطی برای ارزیابی مدل‌های زبانی بزرگ توسعه داده شده است. در این محیط مدل‌های مختلف موجود بررسی و ارزیابی شده و در نهایت با تمام حالات مختلف و شرایط مختلف، قدرت تحلیل آنها در زبان فارسی و انگلیسی بررسی شده است. این پروژه نشان‌دهنده که دستورهای^۱ شامل زنجیره تفکر^۲ و درخت تفکر^۳ باعث بهبود کارایی مدل‌ها و همچنین روش تنظیم نمادها^۴ باعث حساسیت بسیار زیاد به پرسش داده شده و محتوای آن می‌شود.

کلیدواژه‌ها: مدل‌های زبانی بزرگ، پردازش زبان‌های طبیعی، یادگیری ماشین، تشخیص اهمیت اخبار

¹Prompt

²Chain-of-Thoughts

³Tree-of-Thoughts

⁴Symbol Tuning

فهرست مطالب

۱	مقدمه	۱
۱-۱	تعریف مسئله چگونگی بررسی مهم بودن یک خبر	۱
۲-۱	اهمیت موضوع تشخیص اهمیت اخبار	۲
۳-۱	ادبیات به کار رفته در این پژوهش	۲
۴-۱	اهداف پژوهش	۲
۲	مفاهیم اولیه	۴
۱-۲	مدل‌های زبانی بزرگ	۴
۲-۲	یادگیری درونی	۴
۳-۲	تنظیم بر اساس دستورالعمل	۵
۴-۲	درخواست‌های سامانه و کاربر	۵
۵-۲	مهندسی درخواست	۶
۳	کارهای پیشین	۷
۱-۳	تشخیص اهمیت اخبار	۷
۱-۱-۳	رویکردهای کلاسیک	۷
۲-۱-۳	رویکردهای تشخیص اهمیت با استفاده از یادگیری ماشین	۸
۳-۱-۳	رویکرد یادگیری عمیق	۸
۴-۱-۳	استفاده از مدل‌های زبانی بزرگ	۸

۲-۳	تنظیم بر اساس دستورالعمل و تنظیم نمادین	۹
۳-۳	یادگیری چندنمونه‌ای در تشخیص اهمیت اخبار	۹
۴-۳	تشخیص اهمیت اخبار فارسی	۹
۵-۳	سامانه‌های درخواست وابسته به پرسش	۱۰

۴ روش پیشنهادی

۱-۴	انواع دستورالعمل‌های توسعه‌داده شده	۱۱
۱-۱-۴	دستورالعمل‌های وانیلا	۱۱
۲-۱-۴	دستورالعمل‌های چندنمونه‌ای	۱۲
۳-۱-۴	دستورالعمل‌های زنجیره تفکر	۱۳
۴-۱-۴	دستورالعمل‌ها درخت تفکر	۱۴
۲-۴	رویکرد دستورالعمل سیستمی و کاربر	۱۴
۳-۴	تحلیل قدرت استدلال در حالت چند زبانی	۱۵
۴-۴	رویکرد تنظیم نمادها	۱۵
۵-۴	تنظیم براساس دستورالعمل و حالت‌های مختلف آن	۱۶
۶-۴	انواع متن بررسی شده از اخبار	۱۶

۵ نتایج جدید

۶ نتیجه‌گیری

مراجع

واژه‌نامه

آ مطالب تکمیلی

۱-آ	دستورالعمل‌های به کار گرفته شده	۲۳
۱-۱-آ	دستورهای وانیلا یا خام	۲۳

- آ-۱-۲ دستوره‌ای نمادین ۲۵
- آ-۱-۳ دستورها با رویکرد یادگیری چند نمونه‌ای ۲۶
- آ-۲ نتایج اضافی‌تر ۲۷

فهرست جداول

فهرست تصاویر

۱-۴ حالت‌های مختلف یک دستورالعمل و برچسب‌های آن ۱۶

فصل ۱

مقدمه

در دنیای رو به پیش رفت روزمره، حجم عظیمی از اخبار شبانه‌روز به سمت کاربران روانده می‌شود. در این حین می‌دانیم که بسیاری از این اخبار مبنای درستی نداشته و بسیاری نیز برای کاربران بسیار اهمیت کمی دارد. با معرفی یک بستر که بتوان به وسیله آن اخبار مهم به خصوص با توجه به فرهنگ ایرانیان تشخیص داده خود یک چالش بزرگ اما بسیار کاربردی است. در اینجا با استفاده و بهره‌گیری از مدل‌های زبانی بزرگ و دانش که توسط آنها جمع‌آوری شده است به انجام این امر پرداختیم. در ادامه همچنین چالش‌های این مدل‌ها و منطبق نبودن آن طبق فرهنگ و عادات ایرانیان بررسی می‌کنیم و با ارائه روش یادگیری چند نمونه^۱، این مشکل را برای طرف می‌کنیم.

۱-۱ تعریف مسئله چگونگی بررسی مهم بودن یک خبر

مسئله به این شکل تعریف می‌شود که یک خبر در هر دسته‌ای که قرار داشته باشد یا دارای اهمیت بالا یا برچسب ۱ و یا دارای اهمیت پایین و برچسب ۰ است. با دادگان جمع‌آوری شده و برچسب‌گذاری‌های انسانی روی آنها، به ۵۵۰۹ داده آموزش و ۱۱۸۰ داده تست و ارزیابی رسیده، که با استفاده از آنها مدل‌ها توصیه نمونه براساس شباهت تعریف شده است و هدف آن است که مدل بتواند اهمیت خبر (۰ یا ۱) را تشخیص دهد و به کاربر اعلام کند.

¹Few-Shot Learning

۲-۱ اهمیت موضوع تشخیص اهمیت اخبار

از اهمیت این کار و محیط توسعه داده شده می توان به موارد زیر اشاره کرد:

- تسهیل پیگیری اخبار برای کاربران، از آنجایی که این محیط توان تشخیص اخبار مهم در دسته ها مختلف را داشته، می توان برای کاربران صرفا اخبار مهم را دسته بندی کرده و آنها با خواندن این اخبار در وقت خود نسبت به خواندن مطالب بی اهمیت صرفه جویی خواهند کرد.
- بررسی قدرت استدلال و تفکر مدل های زبانی بزرگ، از آنجایی تشخیص اهمیت یک خبر کار نسبتا پیچیده ای برای این مدل ها احتساب می شود، این بستر فراهم شده است که قدرت استدلال و تحلیل مدل های مختلف در شرایط گوناگون ارزیابی و اعلام شود.
- در این کار، روش هایی برای بهبود و بهینه کردن دقت این مدل ها پیشنهاد و بررسی شده که در جنبه های دیگری غیر از تشخیص اخبار مهم می توان کمک کننده باشد و به کار گرفته شود. از جمله اینها مسئله طبقه بندی و یادگیری محتوای دستور یا درخواست داده شده به مدل های زبانی بزرگ است.

۳-۱ ادبیات به کار رفته در این پژوهش

از آنجایی که بخش هایی از این پروژه الهام گرفته و ادامه کار تنظیم نمادها [۱] بوده از ادبیات این کار نیز در اینجا استفاده شده است. تنظیم نمادها عبارت است از روشی که به جای برچسب های اصلی که در اینجا همان ۰ یا ۱ هستند، یک رشته از نمادها همانند !، #، & و کاراکترهای دیگر جایگزین شود و مدل نتواند به دانش پیشینه خود اتکا کند.

۴-۱ اهداف پژوهش

اهداف این پژوهش صورت گرفته به دو قسمت کلی تقسیم می شود:

- ابتدا با ساختار و تعریف اخبار مهم پرداخته شده است، در این پژوهش نتیجه ها و بررسی های انجام شده حاکی این موضوع است که اخبار در دسته های گوناگون و برای اشخاص با فرهنگ های مختلف اهمیت متفاوتی دارد. بنابراین انجام یک مسئله طبقه بندی روی آنها کار آسانی نبوده و با بهبودهای

انجام شده در این پژوهش، مسیری برای پژوهش‌های بعدی در جهت رسیدن که دقت بالا با در نظر گرفتن تمام این شرایط فراهم کند.

- مدل‌های زبان بزرگ که کانون اصلی توجه این پژوهش بوده است در این مسئله خاص به طور کامل بررسی شده و تمامی نقاط ضعف و قوت این مدل‌ها در تشخیص اهمیت اخبار بررسی شده است. همچنین تفاوت قدرت استدلال این مدل‌ها در زبان فارسی با انگلیسی مورد مقایسه قرار گرفته که خود می‌تواند مورد استناد برای پژوهش‌های آینده در این زمینه قرار گیرد.

فصل ۲

مفاهیم اولیه

در اینجا به مفاهیم اصلی به کار برده شده در این پروژه، و بررسی کاربرد و پیشینه آن می‌پردازیم.

۱-۲ مدل‌های زبانی بزرگ

مدل‌های زبانی بزرگ^۱ به سامانه‌های هوش مصنوعی گفته می‌شوند که بر اساس پردازش زبان طبیعی طراحی شده‌اند و قادر به تولید و درک متن‌های انسانی در مقیاس وسیع هستند. این مدل‌ها با استفاده از حجم بسیار زیادی از داده‌های متنی آموزش می‌بینند و می‌توانند وظایف متنوع زبانی، از جمله ترجمه، خلاصه‌سازی و پاسخ به پرسش‌ها را انجام دهند.

مفهوم مدل‌های زبانی از دهه ۱۹۸۰ با ظهور الگوریتم‌های احتمالاتی ساده آغاز شد. با معرفی شبکه‌های عصبی در دهه ۱۹۹۰ و توسعه یادگیری عمیق در دهه ۲۰۱۰، مدل‌هایی مانند ترانسفورمرها و سیستم‌هایی نظیر جی‌پی‌تی (نسل اول تا سوم) و مدل‌های مشابه توانستند به کارایی فوق‌العاده‌ای دست یابند. [۲] افزایش قدرت محاسباتی و دسترسی به داده‌های بیشتر، این پیشرفت‌ها را تسهیل کرد.

۲-۲ یادگیری درونی

یادگیری درون‌متنی^۲ به توانایی یک مدل زبانی اشاره دارد که بتواند بر اساس نمونه‌هایی که در همان متن ورودی ارائه می‌شود، وظایف جدیدی را یاد بگیرد. در این روش، نیاز به آموزش دوباره مدل وجود ندارد،

^۱Large Language Models

^۲In-Context Learning

بلکه مدل از اطلاعات داده شده در همان لحظه استفاده می‌کند. [۳]

این مفهوم در اوایل دهه ۲۰۲۰ با توسعه مدل‌هایی مانند جی‌پی‌تی ۳ به وضوح مطرح شد. این مدل‌ها نشان دادند که بدون نیاز به آموزش دوباره، می‌توانند تنها با ارائه نمونه‌هایی در ورودی، وظایف مختلفی را انجام دهند. این پیشرفت‌ها نقطه عطفی در ساده‌سازی استفاده از مدل‌های زبانی محسوب می‌شوند.

۳-۲ تنظیم بر اساس دستورالعمل

تنظیم بر اساس دستورالعمل^۳ فرآیندی است که در آن یک مدل هوش مصنوعی با استفاده از داده‌هایی آموزش می‌بیند که حاوی دستورالعمل‌های خاصی برای انجام وظایف مختلف هستند. [۴] هدف این روش بهبود عملکرد مدل در درک و اجرای دستورالعمل‌هاست.

ایده این روش از مفاهیم یادگیری انتقالی نشأت گرفته است. در سال‌های اخیر، با توجه به توانایی مدل‌های بزرگ زبانی در تعمیم وظایف، محققان تلاش کردند تا این مدل‌ها را با داده‌های حاوی دستورالعمل بهبود دهند. پروژه‌هایی مانند اجرای دستور عمل در مدل‌های جی‌پی‌تی [۵] نشان‌دهنده موفقیت این رویکرد هستند.

۴-۲ درخواست‌های سامانه و کاربر

درخواست‌های سامانه^۴ و کاربر به متونی اطلاق می‌شود که برای هدایت مدل زبانی به سمت تولید پاسخ مناسب استفاده می‌شوند. درخواست سامانه معمولاً وظیفه مشخص کردن قواعد کلی را دارد، در حالی که درخواست کاربر هدف یا سؤال خاصی را بیان می‌کند. [۶]

این مفهوم با گسترش استفاده از مدل‌های زبانی در تعاملات انسانی به وجود آمد. اولین تلاش‌ها برای تعریف و تمایز این دو نوع درخواست در توسعه رابط‌های کاربری تعاملی و چت‌بات‌ها مشاهده شد. این ایده در مدل‌های زبانی بزرگ تکامل یافت.

³Instruction Tuning

⁴System Prompt

۵-۲ مهندسی درخواست

مهندسی درخواست^۵ به هنر و دانش طراحی درخواست‌ها برای هدایت مدل‌های زبانی جهت تولید پاسخ‌های دقیق و مفید اشاره دارد. این فرآیند شامل ایجاد ورودی‌هایی است که بتوانند بهترین نتیجه ممکن را از مدل دریافت کنند.

این مفهوم با ظهور مدل‌های زبانی پیچیده و نیاز به بهره‌برداری بهتر از توانایی‌های آن‌ها مطرح شد. در سال‌های اخیر، مقالات و ابزارهای بسیاری برای استانداردسازی و بهبود این فرآیند ارائه شده است. [۷] مهندسی درخواست در زمینه‌های مختلف، از پژوهش گرفته تا صنعت، نقش کلیدی ایفا می‌کند.

^۵Prompt Engineering

فصل ۳

کارهای پیشین

همواره در طول زمان بررسی اهمیت اخبار چه در زبان فارسی و چه در زبان انگلیسی یک دغدغه و یک کار مبهم بوده است. از آنجایی که اهمیت یک خبر وابسته به عوامل مختلف همانند فرهنگ، موقعیت جغرافیایی، سلائق شخصی و دیدگاه‌های کاربران بوده در نگاه اول به نظر این کار، ناممکن می‌رسد. اما پژوهش‌های اخیر نشان داده است که با استفاده از دادگان‌های برچسب‌گذاری شده و استفاده از یادگیری چند نمونه‌ای می‌توان به نتایج قابل قبولی برای این قسمت رسید.

در اینجا به روش‌های مختلف که در گذشته برای بررسی اهمیت اخبار توسعه داده شده است پرداخته شده است و سپس مسیرهای مختلف بررسی و آنالیز این طبقه‌بندی را در مدل‌های زبانی بزرگ بیان شده است.

۳-۱ تشخیص اهمیت اخبار

این بخش به بررسی روش‌های مختلفی می‌پردازد که در طول زمان برای تشخیص اهمیت اخبار استفاده شده‌اند. این روش‌ها شامل رویکردهای کلاسیک، یادگیری ماشین و هوش مصنوعی، یادگیری عمیق و در نهایت مدل‌های زبانی بزرگ هستند.

۳-۱-۱ رویکردهای کلاسیک

در روش‌های کلاسیک، تشخیص اهمیت اخبار بیشتر بر اساس معیارهای دستی انجام می‌شد. از معیارهایی مانند طول خبر، تعداد دفعات ذکر شدن یک موضوع در منابع مختلف، یا تحلیل‌های آماری ساده برای این

کار استفاده می‌شد. [۸] این روش‌ها به دلیل محدودیت در قابلیت درک معنایی متون، کارایی پایینی در مسائل پیچیده داشتند.

۳-۱-۲ رویکردهای تشخیص اهمیت با استفاده از یادگیری ماشین

با ظهور الگوریتم‌های یادگیری ماشین^۱، از مدل‌هایی مانند ماشین بردار پشتیبان^۲، دسته‌بند بیزین ساده و جنگل‌های تصادفی^۳ [۹] برای تحلیل اخبار و تشخیص اهمیت آن‌ها استفاده شد. [۱۰] این روش‌ها از ویژگی‌های استخراج‌شده مانند تعداد کلمات کلیدی، میزان تعامل خبرها و تعداد نقل‌قول‌ها بهره می‌بردند.

۳-۱-۳ رویکرد یادگیری عمیق

با توسعه یادگیری عمیق^۴، استفاده از شبکه‌های عصبی مانند شبکه‌های بازگشتی و شبکه‌های توجه‌محور برای درک معنایی متون و تحلیل اخبار رواج یافت. مدل‌هایی نظیر LSTM و GRU توانستند با درک وابستگی‌های طولانی‌مدت در متن [۱۱]، عملکرد چشمگیری ارائه دهند.

۳-۱-۴ استفاده از مدل‌های زبانی بزرگ

مدل‌های زبانی بزرگ با پیش‌آموزش بر داده‌های گسترده، توانایی تحلیل متون خبری را با دقت بالا فراهم کرده‌اند. [۲] این مدل‌ها با توجه به پیکربندی و حجم داده‌های آموزشی، قادرند وظایف مختلف را به صورت چندمنظوره انجام دهند.

در کارهای پیشین انجام شده بررسی شده که رفتار مدل‌های زبانی بزرگ در تشخیص اخبار جعلی چگونه بوده است. به خصوص در پژوهش‌های قبلی [۱۲] با تعریف بازیگر خوب و بد، سعی بر ارزیابی این نوع اخبار داشته و نشان می‌دهد که این مدل‌ها توانایی مناسب جهت تشخیص اخبار جعلی در شرایط از پیش تعریف شده مناسب خواهند داشت.

همچنین کارهای فراتری نسبت به صرفاً اتکا کردن به پردازش متن انجام شده است، به طوری که با بهره‌گیری همزمان از مدل‌های تصویری مانند CLIP اهمیت اخبار براساس محتوای تصویری، ویدیوی و صوتی به همراه متن آنها نیز بررسی شود. [۱۳]

¹Machine Learning

²SVM

³Random Forest Tree

⁴Deep Learning

۲-۳ تنظیم بر اساس دستورالعمل و تنظیم نمادین

تنظیم بر اساس دستورالعمل به آموزش مدل‌های زبانی بزرگ با داده‌هایی اشاره دارد که شامل دستورالعمل‌های دقیق برای انجام وظایف هستند. [۱۴] این روش باعث می‌شود مدل‌ها بتوانند وظایف مشخصی مانند دسته‌بندی اهمیت اخبار را با دقت بیشتری انجام دهند. از سوی دیگر، تنظیم نمادین^۵ شامل استفاده از اطلاعات ساختاریافته مانند نمودارهای دانش یا نمایش‌های معنایی برای تقویت عملکرد مدل‌ها است.

۳-۳ یادگیری چندنمونه‌ای در تشخیص اهمیت اخبار

در این روش، مدل‌ها با تعداد بسیار کمی از نمونه‌های آموزشی، وظایف خود را یاد می‌گیرند. این ویژگی در تحلیل اخبار و تشخیص اهمیت آن‌ها، به‌ویژه در مواقعی که داده‌های آموزشی محدود است، کاربرد دارد. مدل‌هایی مانند Aya توانسته‌اند نشان دهند که تنها با چند نمونه ورودی می‌توانند وظایف پیچیده را انجام دهند. [۱۵]

۴-۳ تشخیص اهمیت اخبار فارسی

در گذشته، تلاش‌هایی برای توسعه مدل‌های تشخیص اهمیت اخبار فارسی صورت گرفته است. این تلاش‌ها بیشتر بر اساس روش‌های یادگیری ماشین کلاسیک بوده و از ویژگی‌های زبانی خاص فارسی مانند ریشه‌یابی و تحلیل صرفی بهره گرفته‌اند. [۱۶] با این حال، استفاده از مدل‌های زبانی بزرگ برای زبان فارسی هنوز در مراحل ابتدایی قرار دارد.

همچنین این پژوهش، ادامه مسیر کار خبرچین [۱۷] بوده که با استفاده از مدل‌های مبنی بر معماری ترانسفورمر سعی داشته که به بررسی اهمیت اخبار بپردازد. در این پژوهش سعی شده با استفاده از مدل‌های زبانی بزرگ رویکرد کلی‌تری نسبت به بررسی اهمیت اخبار ارائه شود که بتواند بستر جامع‌تری برای طبقه‌بندی این حوزه فراهم کند.

^۵Symbol Tuning

۵-۳ سامانه‌های درخواست وابسته به پرسش

این سامانه‌ها با طراحی درخواست‌هایی که وابسته به موضوع پرسش هستند، قادرند نتایج بهینه‌ای در تشخیص اهمیت اخبار ارائه دهند. برای این منظور، از مهندسی درخواست استفاده می‌شود تا مدل‌های زبانی بزرگ بتوانند بر اساس متن ورودی و هدف پرسش، خروجی مطلوبی تولید کنند. [۱۸]

پژوهش‌های اخیر در این زمینه انجام شده است که نشان می‌دهد استفاده از دستورهای مختلف بسته به ورودی کاربر می‌تواند نتایج ثمربخش‌تر به ارقام بیاورد. اگرچه در یکسری پژوهش‌های به این رویکرد یادگیری تقویتی^۶ الحاق می‌شود [۱۹] اما بیشتر یک سیستم در پس‌زمینه بوده که بتواند بهترین دستور را با توجه به ورودی و محتوا تشخیص دهد.

⁶Reinforcement Learning

فصل ۴

روش پیشنهادی

در این قسمت به روش‌های توسعه داده‌شده و نحوه به دست آمدن نتایج و خروجی‌ها می‌پردازیم. فرآیند به این صورت طی می‌شود که ابتدا دستورالعمل مناسب براساس شرط‌های مشخص شده انتخاب می‌شود و سپس در صورت نیاز نمونه‌های مشابه به خبر مورد نظر در دستور آمده و سپس مدل‌های زبانی بزرگ ۸ یا ۹ میلیارد پارامتر روی کارت گرافیکی بالا آمده و به صورت دسته‌های ۱۰ تایی نتایج از خروجی این مدل‌ها ذخیره شده و تحلیل‌ها روی آن انجام می‌شود.

۴-۱ انواع دستورالعمل‌های توسعه داده شده

در اینجا به چهار حالت پرامپت‌ها یا همان دستورالعمل‌ها را که به کار گرفته شده است پرداخته می‌شود و کاربرد و اهداف هر کدام بررسی می‌شود. یک دستورالعمل p_i از یک مجموعه $P = \{p_1, p_2, \dots, p_n\}$ انتخاب می‌شود و براساس آن نتایج خروجی که به صورت برچسب‌های $l_j = \langle 0, 1 \rangle$ مشخص می‌شود خروجی گرفته می‌شود.

۴-۱-۱ دستورالعمل‌های وانیلا

دستورالعمل‌های وانیلا^۱ که همان دستورالعمل‌های خام بوده صرفاً اطلاعات مورد نیاز را برای مدل‌های زبانی بزرگ تهیه و توضیح می‌دهد. در ابتدا وظیفه طبقه‌بندی انجام شده توضیح داده می‌شود و سپس مشخص می‌شود که حتماً خروجی به حالت $l_j = \langle 0, 1 \rangle$ می‌بایست باشد و چندین بار روی این موضوع

^۱Vanila

تاکید می‌شود تا مطمئن شویم خروجی این مدل‌ها صرفاً یک برجسب باشد. سپس توضیح اخبار مهم براساس دسته‌های مختلف تهیه شده و درنهایت خبری که می‌خواهیم طبقه‌بندی شود در پایان این دستور می‌آید. برای نمونه می‌توان یک دستور را به شکل زیر مشاهده کنیم:

- برای این وظیفه‌ی طبقه‌بندی، از شاخه‌های فکری زیر استفاده کنید تا تصمیم بگیرید که آیا خبر «مهم» (۱) است یا «غیر مهم» (۰):
۱. ابتدا بررسی کنید که آیا موضوع خبر می‌تواند بخش بزرگی از کاربران فارسی‌زبان را تحت تأثیر قرار دهد و به گستردگی احتمالی آن توجه کنید. ۲. سپس محتوای موضوع را از نظر اهمیت اقتصادی، سیاسی و اجتماعی تحلیل کنید.
 - برای اخبار اقتصادی: عواملی مانند تورم، وضعیت مسکن و روندهای بورس که برای کاربران عادی اهمیت دارند را مدنظر قرار دهید.
 - برای اخبار سیاسی: ارزیابی کنید که آیا محتوا به سیاست‌های کلان ایران، تغییرات مهم در دولت، یا تعاملات جهانی مربوط می‌شود.
 - برای اهمیت اجتماعی: بررسی کنید که آیا خبر شامل رویدادهای ورزشی محبوب یا موضوعاتی با جذابیت گسترده است.
 ۳. بررسی کنید که آیا جذابیت خبر عمومی است یا فقط برای مخاطبان خاصی جذابیت دارد.
 - اگر خبر از نظر گسترده‌ای مهم است، آن را با «۱» برجسب بزنید. در غیر این صورت، اگر بیشتر برای مخاطبان خاص جذاب است، «۰» را انتخاب کنید.
- تحلیل این سناریوها را به پایان رسانده و طبقه‌بندی نهایی را به صورت زیر ارائه دهید:

طبقه بندی نهایی: «۰ یا ۱»

که این نمونه آورده شده شامل مفهوم درخت‌های تفکر نیز می‌شود. نمونه‌های بیشتر این نوع دستورات را می‌توانید در قسمت مطالب تکمیلی مشاهده کنید.

۴-۱-۲ دستورالعمل‌های چندنمونه‌ای

در این نوع دستورالعمل‌ها علاوه برای بر محتوای خود پرامپت، چندین نمونه براساس خبر خواسته شده نیز آورده می‌شود. فرض کنید خبری که اکنون می‌خواهیم طبقه بندی کنیم به صورت t_i از مجموعه دادگان تست یعنی $T = \{t_1, t_2, \dots, t_n\}$ باشد. آنگاه به ازای هر هر تایتل یا عنوان خبر در دادگان آموزش خود متن و همچنین نسخه برداری آن را به وسیله مدل TF-IDF خواهیم داشت که به این صورت تعریف می‌شود:

$$S_{<Title>} = \{s_1, s_2, \dots, s_n\} \xrightarrow{TF-IDF} V_{<Title, TF-IDF>} = \{v_1, v_2, \dots, v_n\} \quad (1-4)$$

که در آن s_i مشخص کننده یک نمونه و v_j نشان‌دهنده بردار شده است. پس از این اقدام با روش *cosine similarity* از تمامی بردارها، براساس k انتخابی یا همان تعداد نمونه‌ها، شبیه‌ترین عناوین خبرها را همراه برجسب اهمیت آنها قرار می‌دهیم.

$$E_{<samples>} = \{< s_i, l_i > | s_i \in \operatorname{argmax}_k (\cosine(V_{<Title, TF-IDF>}, t_i))\} \quad (2-4)$$

این نمونه‌های کمک می‌کند که مدل مشاهده کند که هرکدام از اخبار شبیه به خبر داده شده به چه صورت از نظر اهمیت ارزیابی شده است و سپس تصمیم نهایی خود را براساس آن تغییر دهد. نمونه‌های در دستور

العمل به این صورت قرار می‌گیرد:

نمونه‌ها: به نمونه‌های زیر نگاه کنید و بر اساس آنها تشخیص دهید که کدام خبرها مهم هستند و با عدد '۰' نمایش داده می‌شوند و کدام خبرها غیرمهم هستند و با عدد '۱' نمایش داده می‌شوند.

- نمونه: هشدار؛ بارش باران و برف در این استان‌ها طبقه بندی نهایی: ۱
- نمونه: بارش برف و باران در جاده‌های ۱۴ استان طبقه بندی نهایی: ۱
- نمونه: پیش‌بینی بارش برف و باران در این استان‌ها/ گرم‌ترین و سردترین شهرها کدامند؟ طبقه بندی نهایی: ۱
- نمونه: بارش برف و باران از فردا در این استان‌ها شروع می‌شود طبقه بندی نهایی: ۱
- نمونه: پیش‌بینی بارش ۵ روزه در ۸ استان طبقه بندی نهایی: ۱
- نمونه: پیش‌بینی رگبار و رعدوبرق طی ۵ روز آتی در برخی نقاط کشور/ وزش باد شدید در شرق طبقه بندی نهایی: ۱
- نمونه: بارش باران در نقاط مختلف کشور طبقه بندی نهایی: ۱
- نمونه: هشدار! دیگر بارش برف و باران هم تاثیر چندانی در آلودگی هوا ندارد طبقه بندی نهایی: ۰
- نمونه: ثبت بیش از ۲۳ میلیون سفر نوروزی فقط در روز گذشته | بارش برف و باران در جاده‌های ۴ استان طبقه بندی نهایی: ۰
- نمونه: بارش برف و باران در جاده‌های ۵ استان | رانندگان قبل از سفر حتما با این سامانه تماس بگیرند طبقه بندی نهایی: ۰
- نمونه: هوای تهران در آخر هفته | چه خبر از بارش برف و باران؟ طبقه بندی نهایی: ۰
- نمونه: ورود سامانه بارشی؛ بارش باران در ۸ استان طی ۲۴ ساعت آینده | تهران و ۳ استان دیگر منتظر وزش باد شدید باشند طبقه بندی نهایی: ۱
- نمونه: بارش باران و تگرگ در آذربایجان غربی طبقه بندی نهایی: ۰
- نمونه: تشدید بارش‌ها از فردا در کشور؛ بارش برف در تهران | کاهش ۴ تا ۸ درجه‌ای دما در نوار شمالی طبقه بندی نهایی: ۱
- نمونه: بارش برف و باران در این ۱۱ استان | آماده‌باش امدادگران هلال احمر | ورود سامانه بارشی جدید به کشور از امروز طبقه بندی نهایی: ۱
- نمونه: تداوم بارش در برخی استان‌ها / وزش باد شدید و خیزش گرد و خاک در شرق کشور طبقه بندی نهایی: ۱
- نمونه: هشدار هواشناسی | افزایش آلودگی هوا در ۲ کلانشهر | وزش باد شدید و ارتفاع امواج در ۳ استان طبقه بندی نهایی: ۰
- نمونه: پیش‌بینی بارش‌ها در کشور | امروز پنجشنبه فقط ۲ استان بارندگی ندارد | گرم‌ترین و سردترین شهرهای کشور طبقه بندی نهایی: ۱
- نمونه: عکس اقدام عجیب عراق در جعل نام خلیج فارس طبقه بندی نهایی: ۱
- نمونه: باخت برانکو به عراق در فینال جام خلیج فارس طبقه بندی نهایی: ۰

۳-۱-۴ دستورالعمل‌های زنجیره تفکر

در این نوع دستورها، از مدل زبانی بزرگ خواسته می‌شود که صرفاً به یک زاویه دید اتکا نکند و موقعیت‌های مختلف را بررسی کند. بعضاً در این حالت به مدل اجازه می‌دهیم که خروجی خود را فراتر از برچسب $\langle 0, 1 \rangle$ ببرد و سعی کند تفکرات خود را بسازد و سپس طبقه‌بندی نهایی را در یک قالب مشخص بیان کند که سپس بتوان آن را استخراج کرد. برای نمونه یک نوع از این پرامپت‌های استفاده در زبان انگلیسی را می‌توانید به صورت زیر مشاهده کنید:

The goal is to classify news items into 'important' (1) or 'not important' (0). To classify accurately, follow these steps:

1. Identify if the news topic could be relevant to a large Persian-speaking audience.
2. Assess if it pertains to significant economic events (currency or inflation changes, housing updates, etc.), critical political events (government actions, international relations), or socially impactful themes that could affect many people.
3. Finally, determine if the story has widespread appeal or is only relevant to a niche audience.

If the news is of broad interest and covers the themes above, label it as '1'. Otherwise, label it as '0'. Upon completing the scenario analysis, output the final classification using the format below:

Final Classification: [1 or 0]

۴-۱-۴ دستورالعمل‌ها درخت تفکر

در این حالت از دستورها، علاوه بر زنجیره تفکر از مدل خواسته می‌شود که در یک یا دو تا از زنجیره‌ها، حالت‌های دیگر را به صورت درختی بررسی کند. در اینجا، از آنجایی که هدف ما طبقه‌بندی اخبار است خواسته شده که در انواع دسته‌بندی خبرها همانند اقتصادی، فرهنگی، سلامت و بهداشت و ورزشی مدل در خصوص اهمیت خبر فکر کرده و سپس اظهار نظر کند.

۴-۲ رویکرد دستورالعمل سیستمی و کاربر

یکی از رویکردهایی که در این پژوهش مورد بررسی قرار گرفته بررسی تاثیر تعریف قاعده‌مند دستور سیستمی و کاربری و جداسازی آن و تاثیر آن برای دقت خروجی مدل‌های زبانی بزرگ است. دستورالعمل سیستمی، دستوری است که معمولاً ثابت است و وظایف اصلی را برای مدل شرح می‌دهد که طبق آن به دستور کاربر پاسخ دهد.

در این پژوهش قواعد اصلی، تعریف مهم و غیرمهم بودن خبر، نمونه‌ها و شرح اصلی به صورت دستورالعمل سیستمی تعریف شده و خود خبری که می‌خواهیم مورد طبقه‌بندی قرار دهیم به صورت دستورالعمل کاربر^۲ تعریف شده است. این رویکرد با رویکرد دستورالعمل یکپارچه مقایسه شده و در قسمت نتایج مشاهده شده که تعریف جداگانه آنها به بهبود کارایی مدل‌ها کمک شایانی می‌کند.

²User Prompt

۳-۴ تحلیل قدرت استدلال در حالت چند زبانی

در این پژوهش این سوال مورد هدف قرار گرفته است که تفاوتی بین دستورالعمل به زبان انگلیسی و فارسی وجود دارد؟ آیا مدل‌های زبان بزرگ فرهنگ و موقعیت جغرافیای کشور ایران را بیشتر در زبان فارسی ذخیره و استدلال می‌توانند بکنند و یا اینکه در حالتی که به صورت زبان انگلیسی به آنها دستور دهیم قدرت استدلال بیشتری خواهند داشت؟ با بررسی دستورهای سیستمی در فصل نتایج به هر دو زبان، نتایج به دست آمده حاکی است که در یکسری شرایط خاص و وابسته به مدل استفاده شده، تفاوت عملکرد متفاوت است به طوری که در دستورهای وانیا در زبان انگلیسی قدرت استدلال بیشتر و در حالت‌های درخت تفکر و زنجیره تفکر به نظر می‌آید در زبان فارسی قدرت تفکر بیشتری داریم.

۴-۴ رویکرد تنظیم نمادها

این رویکرد که الهام گرفته از روش معرفی شده در مقاله [۱] هست بر این عنوان تمرکز می‌کند که به جای قرار دادن تعریف برجسب‌های اصلی برای یک مدل زبانی بزرگ یا همان $l_j = < 0, 1 >$ به جای برجسب‌های اصلی، برجسب‌های نمادین و بی‌ارتباط استفاده شود و قدرت مدل در این سناریو بررسی شود.

در نگاه اول شاید سوال پیش بیاید که چرا اصلاً این رویکرد موثر واقع خواهد شد؟ و چرا قرار ندادن واژه‌های «مهم» و «غیرمهم» و گمراه کردن مدل با سمبل‌های غیر مرتبط در نهایت می‌تواند به دقت مدل کمک کند؟ جواب این سوال‌ها را باید اینطور داد که تنظیم نمادها این هدف را دنبال می‌کند که مدل زبانی بزرگ نتواند به دانش پیشینه خود که روی آن آموزش دیده تکیه کند و تمامی خروجی‌ها را براساس نمونه‌های قرار داده شده در دستور یادبگیرد و پیش‌بینی کند.

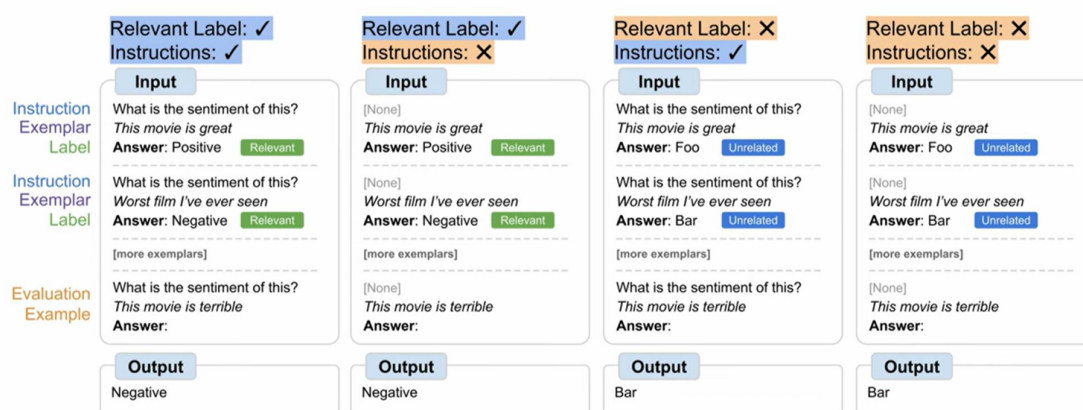
در سناریو تشخیص اهمیت اخبار، بسیاری از اوقات مشاهده می‌شود که این مدل‌های رویکردی غیر از فرهنگ مربوط به کشور ایران در پیش‌بینی اهمیت اخبار به کار می‌برد زیرا که بیشتر داده‌های آموزش خود به زبان انگلیسی بوده و شاید یک چیز بی‌اهمیت برای یک فرد ایرانی، در فرهنگ غرب بسیار مهم واقع شود. در این شرایط نیاز داریم که مدل بسیار بیشتر به مثال‌ها و نمونه‌های قرار داده شده توجه کند. در این پژوهش به جای برجسب‌های «مهم» و «غیرمهم» از برجسب‌های «۵۸» و «۴۷» استفاده شده است تا مدل را از دانش پیشینه خود محروم کند. یعنی در پرامپت قرار گرفته برای مدل هیچ واژه «مهم» یا «غیرمهم» وجود نداشته و مدل براساس مثال‌ها و تعاریف‌ها باید مفهوم دو برجسب «۵۸» و «۴۷» را یادبگیرد و سپس برجسب نهایی را براساس این دو پیش‌بینی کند.

$$l_j = < 0, 1 > \xrightarrow{\text{SymbolTuning}} l_{<\text{Symbolic}>} = < 47, 58 > \quad (3-4)$$

یکی از کاربردهای دیگر این روش نیز سنجیدن حساسیت مدل‌ها به دستور داده شده است، از آنجایی که مدل نمی‌تواند به دانش پیشینه خود تکیه کند بسیار بیشتر به محتوای دستور حساس‌تر شده و با سنجیدن آنها به وسیله K های مختلف نمونه‌ها می‌توان رویکردهای جالبی از عملکرد این مدل‌ها و حساسیتی که به پراپمت نشان می‌دهد بررسی کرد.

۵-۴ تنظیم براساس دستورالعمل و حالت‌های مختلف آن

در رویکردی که در پراپمت یک شرح دستور و نمونه می‌آوریم چهار حالت ممکن طبق ۴-۱ داریم.



شکل ۴-۱: حالت‌های مختلف یک دستورالعمل و برجسب‌های آن

که همانطور که مشخص است چهار حالت داریم که در یکسری از حالت شرح خود وظیفه کامل مشخص است و در دیگر حالت‌ها صرفاً به آوردن نمونه اتکا شده است. و همچنین حالت‌های دیگر به وسیله آوردن برجسب‌های مرتبط و غیرمرتبط شکل می‌گیرد.

در این کار، دو حالت شرح وظیفه به همراه برجسب مرتبط و شرح وظیفه به همراه برجسب غیرمرتبط بررسی شده است. از آنجایی که بدون شرح وظیفه درک اینکه برجسب‌ها چه مفهومی برای مدل زبانی بزرگ در تشخیص اهمیت خبر دارد بسیار پیچیده و مبهم می‌شود، به بررسی اینگونه حالات در این پژوهش پرداخته شده است.

۶-۴ انواع متن بررسی شده از اخبار

فصل ۵

نتایج جدید

در این فصل نتایج جدید به دست آمده در پایان نامه توضیح داده می شود. در صورت نیاز می توان نتایج جدید را در قالب چند فصل ارائه نمود. همچنین در صورت وجود پیاده سازی، بهتر است نتایج پیاده سازی را در فصل مستقلی پس از این فصل قرار داد.

فصل ۶

نتیجه‌گیری

در این فصل، ضمن جمع‌بندی نتایج جدید ارائه‌شده در پایان‌نامه یا رساله، مسائل باز باقی‌مانده و همچنین پیشنهادهایی برای ادامه‌ی کار ارائه می‌شوند.

Bibliography

- [1] J. Wei, L. Hou, A. Lampinen, X. Chen, D. Huang, Y. Tay, X. Chen, Y. Lu, D. Zhou, T. Ma, and Q. V. Le. Symbol tuning improves in-context learning in language models, 2023.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.
- [3] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners, 2020.
- [4] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le. Finetuned language models are zero-shot learners, 2022.
- [5] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe. Training language models to follow instructions with human feedback, 2022.
- [6] S. Gao, A. Sethi, S. Agarwal, T. Chung, and D. Hakkani-Tur. Dialog state tracking: A neural reading comprehension approach, 2019.
- [7] L. Reynolds and K. McDonell. Prompt programming for large language models: Beyond the few-shot paradigm, 2021.
- [8] H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2):159–165, 1958.

- [9] M. Felicilda, A. Geriane, V. Agustin, M. C. Blanco, J. Morano, L. Mahusay, and J. Guialil. Enhancement of random forest algorithm applied in fake news detection. *World Journal of Advanced Research and Reviews*, 22:1075–1079, 05 2024.
- [10] Y. Yang and J. O. Pedersen. A comparative study on feature selection in text categorization. In *Proceedings of the Fourteenth International Conference on Machine Learning, ICML '97*, page 412–420, San Francisco, CA, USA, 1997. Morgan Kaufmann Publishers Inc.
- [11] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, Nov. 1997.
- [12] B. Hu, Q. Sheng, J. Cao, Y. Shi, Y. Li, D. Wang, and P. Qi. Bad actor, good advisor: Exploring the role of large language models in fake news detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20):22105–22113, Mar. 2024.
- [13] J. Wang, Z. Zhu, C. Liu, R. Li, and X. Wu. Llm-enhanced multimodal detection of fake news. *PLOS ONE*, 19, 10 2024.
- [14] V. Sanh, A. Webson, C. Raffel, S. H. Bach, L. Sutawika, Z. Alyafeai, A. Chaffin, A. Stiegler, T. L. Scao, A. Raja, M. Dey, M. S. Bari, C. Xu, U. Thakker, S. S. Sharma, E. Szczechla, T. Kim, G. Chhablani, N. Nayak, D. Datta, J. Chang, M. T.-J. Jiang, H. Wang, M. Manica, S. Shen, Z. X. Yong, H. Pandey, R. Bawden, T. Wang, T. Neeraj, J. Rozen, A. Sharma, A. Santilli, T. Fevry, J. A. Fries, R. Teehan, T. Bers, S. Biderman, L. Gao, T. Wolf, and A. M. Rush. Multitask prompted training enables zero-shot task generalization, 2022.
- [15] P. Zheng, H. Chen, S. Hu, B. Zhu, J. Hu, C.-S. Lin, X. Wu, S. Lyu, G. Huang, and X. Wang. Few-shot learning for misinformation detection based on contrastive models. *Electronics*, 13(4), 2024.
- [16] M. Heydari, M. Khazeni, and M. A. Soltanshahi. Deep learning-based sentiment analysis in persian language. In *2021 7th International Conference on Web Research (ICWR)*, page 287–291. IEEE, May 2021.
- [17] H. H. Hemati, A. Lagzian, M. S. Sartakhti, H. Beigy, and E. Asgari. Khabarchin: Automatic detection of important news in the persian language, 2023.
- [18] L. Reynolds and K. McDonell. Prompt programming for large language models: Beyond the few-shot paradigm, 2021.

- [19] H. Sun, A. Hüyük, and M. van der Schaar. Query-dependent prompt evaluation and optimization with offline inverse rl, 2024.

واژه‌نامه

ت

تنظیم براساس دستورالعمل Insturction Tuning....
تنظیم نمادین Symbol Tuning

ق

قطعی Deterministic.....

ج

جنگل‌های تصادفی Random Forest Tree.....

م

مدل‌های زبانی بزرگ Large Language Models....

مهندسی درخواست Prompt Engineering.....

ماشین بردار پشتیبان SVM.....

د

دستور Prompt

درخت تفکر Tree-of-Thoughts.....

درخواست سامانه System Prompt.....

و

دستورالعمل خام Vanila Prompt.....

دستورالعمل کاربر User Prompt.....

ز

زنجیره تفکر Chain-of-Thoughts.....

ی

یادگیری چند نمونه Few-Shot Learning.....

یادگیری درون‌متنی In-Context Learning.....

یادگیری ماشین Machine Learning

یادگیری عمیق Deep Learning.....

یادگیری تقویتی Reinforcement Learning.....

ص

صدق‌پذیری Satisfiability.....

پیوست آ

مطالب تکمیلی

در اینجا تمام محتوای تکمیلی پژوهش از جمله دستورالعمل‌های استفاده شده در مراحل مختلف برای مدل‌های زبانی بزرگ و نتایج دسته‌بندی شده قرار گرفته شده است.

آ-۱ دستورالعمل‌های به کار گرفته شده

در این بخش انواع دستورالعمل‌های نوشته شده در این پژوهش آمده و شرح داده می‌شود.

آ-۱-۱ دستورالعمل‌های وانیلا یا خام

ابتدایی‌ترین دستورالعمل‌های نوشته شده برای مدل‌های زبانی بزرگ به طوری که شامل تعریف اخبار مهم و شرح وظیفه است.

هدف، داشتن یک دسته‌بند دودویی است که با گرفتن هر متن ورودی، کلاس آن را در خروجی مشخص می‌کند. کلاس‌ها شامل دو دسته‌ی غیرمهم و مهم هستند. یعنی خبر نوع غیرمهم و خبر نوع مهم.

شرح تسک:

متن یا خبری را مهم می‌گوییم اگر که برای بیش‌تر کاربران فارسی‌زبان اهمیت بالایی داشته باشد. یا به عبارت دیگر، جمعیت زیاد و بزرگی از ایرانیان مایل باشند که آن متن یا خبر را بخوانند و یا برای یکدیگر بفرستند.

اگر خبری مربوط به یک قشر کوچک یا جامعه‌ی خاصی از کاربران باشد و یا ارزش خواندن کمی داشته باشد و یا خاص نباشد، آن خبر از نوع غیرمهم است.

در صورتی که متن ورودی از نوع مهم باشد، کلاس مهم خواهد بود و در صورتی که غیرمهم باشد، کلاس غیرمهم خواهد بود.

برخی از مفاهیم از نوع مهم عبارت‌اند از: یارانه و سهام و مواردی که قرار است پول به مردم برسد مهم هستند ثبت نام مسکن و خانه و اخبار مربوط به وام‌ها و... ثبت نام خودرو افزایش و کاهش‌های شدید و زیاد قیمت ارز یا طلا و سکه و یا تورم

سیاسی: اخبار جنگ، برجام، توافق های ایران، تحریم های ایران، خبرهای جنگ های بزرگ منطقه ای، عزل و نصب مقامات بلندپایه ایرانی، این ها همگی مهم هستند

ورزشی: اخبار مربوط به تیم های معروف و پرطرفدار ایرانی و همین طور اروپایی مهم است

تمام اخبار بالا از نوع مهم بوده و اخبار دسته های دیگر که کمتر خواننده دارند را از نوع غیرمهم در نظر می گیریم.

با توجه به متن زیر تنها در یک عدد پاسخ بده که باتوجه به مفاهیمی که در بالا مطرح شد و قدرت استنتاجی که خودت داری، آیا متن مهم حساب می شود یا غیرمهم. (مهم یا غیرمهم):

در خروجی فقط مجاز هستی مهم یا عدد غیرمهم بنویسی. بدون هیچ توضیح اضافی.

و نسخه انگلیسی آن به این صورت نوشته و مورد بررسی قرار گرفته است:

The goal is to have a binary classifier that, by receiving any input text, determines its class in the output. The classes include two categories: 'not important' and 'important', meaning news type 'not important' and news type 'important'.

Task description:

We label a text or news as 'important' if it is of high importance to most Persian-speaking users. In other words, if a large population of Iranians are likely to read, share, or be interested in it, it is classified as 'important'.

If the news pertains to a small group or a specific community of users, has little reading value, or is not significant, it is classified as 'not important'.

If the input text is of type 'important', the output class will be 'important'; if it is 'not important', the output class will be 'not important'.

Some concepts that fall under type 'important' are: Subsidies, stocks, and matters that involve receiving money are important. Housing and home registrations, news related to loans, etc. Car registrations Significant fluctuations in currency, gold, coins, or inflation rates

Politics: News about war, the JCPOA, Iran's agreements, Sanctions on Iran, News of major regional wars, Dismissal and appointment of high-ranking Iranian officials, These are all important.

Sports: News about famous and popular Iranian teams as well as European teams is important.

All the above news are classified as type 'important', and other news categories that have fewer readers are considered as type 'not important'.

A text or news is classified as 'not important' if it pertains to a specific small section of the society. News that does not engage a broad spectrum of the community is type 'not important'. For example: Sports: News about non-famous clubs and small events are of type 'not important'. Politics: News about non-prominent figures that do not affect the Iranian society is of type 'not important'. Social: News that does not engage a large section of society is type 'not important'.

Based on the following text, respond with only a single label that, considering the concepts discussed above and your own inferential ability, indicates whether the text should be classified as 'important' or 'not important'. ('important' or 'not important'):

You are only allowed to write the label 'important' or 'not important' in the output, without any additional explanation.

که شامل تعاریف اخبار مهم برای دسته های مختلف خبری به صورت اضافه تر است.

آ-۱-۲ دستوره‌های نمادین

در اینجا در دستور نوشته شده، هیچ اسمی از «مهم» و «غیرمهم» بودن و تعاریف آنها برده نشده و صرفاً از برچسب‌های نمادین «۵۸» و «۴۷» استفاده شده است.

هدف، داشتن یک دسته‌بند دودویی است که با گرفتن هر متن ورودی، کلاس آن را در خروجی مشخص می‌کند. کلاس‌ها شامل دو دسته‌ی ۴۷ و ۵۸ هستند. یعنی خبر نوع ۴۷ و خبر نوع ۵۸.

شرح تسک:

متن یا خبری را ۵۸ می‌گوییم اگر که برای بیش‌تر کاربران فارسی‌زبان اهمیت بالایی داشته باشد. یا به عبارت دیگر، جمعیت زیاد و بزرگی از ایرانیان مایل باشند که آن متن یا خبر را بخوانند و یا برای یکدیگر بفرستند.

اگر خبری مربوط به یک قشر کوچک یا جامعه‌ی خاصی از کاربران باشد و یا ارزش خواندن کمی داشته باشد و یا خاص نباشد، آن خبر از نوع ۴۷ است. در صورتی که متن ورودی از نوع ۵۸ باشد، کلاس ۵۸ خواهد بود و در صورتی که ۴۷ باشد، کلاس ۴۷ خواهد بود.

برخی از مفاهیم از نوع ۵۸ عبارت‌اند از: یارانه و سهام و مواردی که قرار است پول به مردم برسد مهم هستند ثبت نام مسکن و خانه و اخبار مربوط به وام‌ها و... ثبت نام خودرو افزایش و کاهش‌های شدید و زیاد قیمت ارز یا طلا و سکه و یا تورم

سیاسی: اخبار جنگ، برجام، توافق‌های ایران، تحریم‌های ایران، خبرهای جنگ‌های بزرگ منطقه‌ای، عزل و نصب مقامات بلندپایه ایرانی، این‌ها همگی مهم هستند

ورزشی: اخبار مربوط به تیم‌های معروف و پرطرفدار ایرانی و همین‌طور اروپایی مهم است

تمام اخبار بالا از نوع ۵۸ بوده و اخبار دسته‌های دیگر که کمتر خواننده دارند را از نوع ۴۷ در نظر می‌گیریم.

با توجه به متن زیر تنها در یک عدد پاسخ‌ده که با توجه به مفاهیمی که در بالا مطرح شد و قدرت استنتاجی که خودت داری، آیا متن ۵۸ حساب می‌شود یا ۴۷. (۵۸ یا ۴۷):

در خروجی فقط مجاز هستی عدد ۵۸ یا عدد ۴۷ بنویسی. بدون هیچ توضیح اضافه‌ای.

در نتایج به دست آمده، متوجه شدیم که در این دستور صرفاً تعریف اخبار «مهم» یا همان «۵۸» آمده است و برای همین مدل در رویکر بدون نمونه صرفاً تمامی اخبار را مهم پیش‌بینی می‌کند. برای حل این مشکل دستور زیر با اضافه شدن تعاریف اخبار غیرمهم یا همان «۴۷» نوشته و به کار گرفته شد.

هدف، داشتن یک دسته‌بند دودویی است که با گرفتن هر متن ورودی، کلاس آن را در خروجی مشخص می‌کند. کلاس‌ها شامل دو دسته‌ی ۴۷ و ۵۸ هستند. یعنی خبر نوع ۴۷ و خبر نوع ۵۸.

شرح تسک:

متن یا خبری را ۵۸ می‌گوییم اگر که برای بیش‌تر کاربران فارسی‌زبان اهمیت بالایی داشته باشد. یا به عبارت دیگر، جمعیت زیاد و بزرگی از ایرانیان مایل باشند که آن متن یا خبر را بخوانند و یا برای یکدیگر بفرستند.

اگر خبری مربوط به یک قشر کوچک یا جامعه‌ی خاصی از کاربران باشد و یا ارزش خواندن کمی داشته باشد و یا خاص نباشد، آن خبر از نوع ۴۷ است. در صورتی که متن ورودی از نوع ۵۸ باشد، کلاس ۵۸ خواهد بود و در صورتی که ۴۷ باشد، کلاس ۴۷ خواهد بود.

برخی از مفاهیم از نوع ۵۸ عبارت‌اند از: یارانه و سهام و مواردی که قرار است پول به مردم برسد مهم هستند ثبت نام مسکن و خانه و اخبار مربوط به وام‌ها و... ثبت نام خودرو افزایش و کاهش‌های شدید و زیاد قیمت ارز یا طلا و سکه و یا تورم

سیاسی: اخبار جنگ، برجام، توافق‌های ایران، تحریم‌های ایران، خبرهای جنگ‌های بزرگ منطقه‌ای، عزل و نصب مقامات بلندپایه ایرانی، این‌ها همگی مهم هستند

ورزشی: اخبار مربوط به تیم‌های معروف و پرتعداد ایرانی و همین‌طور اروپایی مهم است

تمام اخبار بالا از نوع ۵۸ بوده و اخبار دسته‌های دیگر که کمتر خواننده دارند را از نوع ۴۷ در نظر می‌گیریم.

متن یا خبری را ۴۷ می‌گویند که مربوط به بخش خاص و کوچکی از جامعه باشد. اخباری که گستره‌ی وسیعی از جامعه را درگیر نکند، اخبار از نوع ۴۷ هستند. برای نمونه: ورزشی: اخبار مربوط به باشگاه‌های غیر معروف و رخدادهای کوچک از نوع ۴۷ هستند. سیاسی: اخبار مربوط به شخصیت‌های غیرمشهور که تأثیری روی جامعه‌ی ایران ندارد از نوع ۴۷ هستند. اجتماعی: اخباری که گستره‌ی وسیعی از جامعه را درگیر نمی‌کند از نوع ۴۷ هستند.

با توجه به متن زیر تنها در یک عدد پاسخ بده که باتوجه به مفاهیمی که در بالا مطرح شد و قدرت استنتاجی که خودت داری، آیا متن ۵۸ حساب می‌شود یا ۴۷. (۵۸ یا ۴۷):

در خروجی فقط مجاز هستی عدد ۵۸ یا عدد ۴۷ بنویسی. بدون هیچ توضیح اضافه‌ای.

آ-۱-۳ دستورها با رویکرد یادگیری چند نمونه‌ای

در این نوع پرامپت‌ها، چندین نمونه و مثال برای یادگیری و شباهت‌سنجی در اختیار مدل زبانی بزرگ قرار می‌گیرد که یک نمونه از این نوع دستورها در اینجا قرار داده شده است.

هدف، داشتن یک دسته‌بند دودویی است که با گرفتن هر متن ورودی، کلاس آن را در خروجی مشخص می‌کند. کلاس‌ها شامل دو دسته‌ی ۱ یا ۰ هستند. ۱ یعنی خبر مهم است و ۰ یعنی خبر مهم نیست.

شرح تسک:

متن یا خبری را مهم یا تاثیرگذار می‌گوییم اگر که برای بیش‌تر کاربران فارسی‌زبان اهمیت بالایی داشته باشد. یا به عبارت دیگر، جمعیت زیاد و بزرگی از ایرانیان مایل باشند که آن متن یا خبر را بخوانند و یا برای یکدیگر بفرستند. اگر خبری مربوط به یک قشر کوچک یا جامعه‌ی خاصی از کاربران باشد، آن خبر مهم نیست. در صورتی که متن ورودی مهم باشد، کلاس ۱ خواهد بود و در صورتی که مهم نباشد، کلاس ۰ خواهد بود.

برخی از مفاهیم مهم و از کلاس ۱ عبارت‌اند از: یارانه و سهام و مواردی که قرار است پول به مردم برسد مهم هستند ثبت نام مسکن و خانه و اخبار مربوط به وام‌ها و... ثبت نام خودرو افزایش و کاهش های شدید و زیاد قیمت ارز یا طلا و سکه و یا تورم

سیاسی: اخبار جنگ، برجام، توافق های ایران، تحریم های ایران، خبرهای جنگ‌های بزرگ منطقه‌ای، عزل و نصب مقامات بلندپایه ایرانی، این‌ها همگی مهم هستند

ورزشی: اخبار مربوط به تیم‌های معروف و پرتعداد ایرانی و همین‌طور اروپایی مهم است

نمونه‌ها: چند نمونه پایین را ببین و باتوجه به آن‌ها به سوال پایین پاسخ بده

SAMPLES

از روی نمونه‌های بالایی یاد بگیر و خروجی را مشخص کن (فقط ۰ یا ۱). حال با توجه به «نمونه‌های بالا»، برای متن زیر تنها در یک واژه پاسخ بده که باتوجه به مفاهیمی که در بالا مطرح شد و قدرت استنتاجی که خودت داری، آیا متن مهم (تاثیرگذاری) حساب می‌شود یا خیر. (۱ یا ۰):

در خروجی فقط مجاز هستی عدد ۱ یا عدد ۰ بنویسی. بدون هیچ توضیح اضافه‌ای.

آ-۲ نتایج اضافی‌تر

در این بخش به نتایج اضافه‌تر در دسته‌های گوناگون خبری از جمله سیاسی، اقتصادی، ورزشی و فرهنگی اشاره شده است.

Abstract

This work examines the capability of large language models (LLMs) to measure the importance of Persian news, evaluating their learning ability from content, reasoning skills, and overall cognitive capacities. Initially, annotated datasets were collected from various domains, including sports, politics, social issues, medicine, and culture, to develop an evaluation framework for LLMs. Within this framework, various existing models were analyzed and assessed under different scenarios and conditions to evaluate their analytical performance in both Persian and English. The findings indicate that prompts incorporating Chain-of-Thoughts and Tree-of-Thoughts significantly improve the models' performance. Additionally, the Symbol Tuning method enhances sensitivity to the input queries and their content.

Keywords: Large Language Models, Natural Language Processing, Machine Learning, News Importance Detection



Sharif University of Technology
Department of Computer Engineering

B.Sc. Thesis

News Importance Detection With the Use of Large Language Models

By:

Shayan Salehi

Supervisor:

Dr. Mahdi Jafari

January 2025