تشخیص طنز در زبان فارسی با رویکرد یادگیری عمیق

فاطمه نجفی لپوندانی[۱]، محمدحسن شیرعلی شهرضا[۱]

[۱]دانشکده ریاضی و علوم کامپیوتر، دانشگاه صنعتی امیر کبیر (پلی‌تکنیک تهران)، تهران، ایران
{fatemenajafi135,hshirali}@aut.ac.ir

چکیده

طنز و طعنه روشی خلاقانه برای بیان احساسات است که انسان‌ها با به کار بردن آن در گفتار یا نوشتار، ممکن است هدف و منظوری متفاوت با آنچه بیان می‌شود داشته باشند. با گسترش استفاده از اینترنت، استفاده از شبکه‌های اجتماعی و وب‌سایت‌های فروش اینترنتی افزایش پیدا کرده است. با مرسوم‌شدن اعلام دیدگاه‌ها در این شبکه‌ها یا بیان نظرات در مورد کالاهایی که در وب‌سایت‌ها به فروش گذاشته شده‌اند، بررسی نظرات و احساسات کاربران در این موارد برای شرکت‌ها و سازمان‌ها اهمیت شایانی پیدا می‌کند. چون معمولا این نظرات با زبان رسمی و به صراحت اعلام نمی‌شوند و گاهی آغشته به شوخی، طعنه و کنایه هستند، کارایی تحلیل احساسات تحت تأثیر قرار می‌گیرد. در سال‌های اخیر، توییتر به منبع بزرگی از ابراز ایده‌ها و دیدگاه‌های کاربران در تقریبا تمامی زمینه‌ها تبدیل شده است. این مسئله باعث جلب توجه شرکت‌ها و پژوهشگرانی به توییتر است که در زمینه تحلیل نظرات کاربران فعالیت می‌کنند. درصد بالایی از توییت‌ها شامل طنز هستند و کاربران فارسی‌زبان نیز از این قاعده مستثنی نمی‌شوند. تحلیل‌هایی که به وجود شوخی در متن آگاه باشند، می‌توانند با دقت بهتری احساسات را پیش‌بینی کنند. در این پژوهش مجموعه‌داده‌ای از توییتر فارسی معرفی می‌شود که برچسب نمونه‌ها نشان‌گر وجود طنز و یا عدم وجود آن است. با استفاده از تطبیق دقیق مدل‌های زبانی از پیش آموزش داده شده بر روی مجموعه داده به دست آمده، مدلی برای تشخیص وجود یا عدم وجود طنز در زبان فارسی ارائه می‌دهیم.

کلمات کلیدی: تشخیص طنز، تحلیل احساسات، پردازش زبان طبیعی، یاد گیری عمیق، ترنسفرمرها، زبان فارسی.

۱ مقدمه

طنز و طعنه روشی خلاقانه برای بیان احساسات است که انسان‌ها با به کار بردن آن در گفتار یا نوشتار، ممکن است هدف و منظوری متفاوت با آنچه بیان می‌شود داشته باشند زیرا این ماهیت خلاقانه مانع از درک صحیح متن می‌شود.

با گسترش استفاده از اینترنت، استفاده از شبکه‌های اجتماعی و وب‌سایت‌های فروش اینترنتی افزایش پیدا کرده است. با مرسوم شدن اعلام دیدگاه‌ها در این شبکه‌ها یا بیان نظرات در مورد کالاهایی که در

وب‌سایت‌ها به فروش گذاشته شده‌اند، بررسی نظرات و احساسات کاربران در این موارد برای شرکت‌ها و سازمان‌ها اهمیت شایانی پیدا می‌کند. چون معمولا این نظرات با زبان رسمی و به صراحت اعلام نمی‌شوند و گاهی آغشته به شوخی، طعنه و کنایه هستند، کارایی تحلیل احساسات تحت تاثیر قرار می‌گیرد. برای مثال جمله «مرسی برای این همه بدبختی، خدایا شکرت!» حاوی طنز است درحالی‌که جمله «من امروز خیلی خوشحالم. دمت گرم روزم رو ساختی» بدون طنز و جدی است.

در سال‌های اخیر، توییتر به منبع بزرگی از ابراز ایده‌ها و دیدگاه‌های کاربران در تقریبا تمامی زمینه‌ها تبدیل شده است. این مسئله باعث جلب توجه شرکت‌ها و پژوهشگرانی به توییتر است که در زمینه تحلیل نظرات کاربران فعالیت می‌کنند. درصد بالایی از توییت‌ها شامل طنز هستند و کاربران فارسی‌زبان نیز از این قاعده مستثنی نمی‌شوند. تحلیل‌هایی که به وجود شوخی در متن آگاه باشند، ممکن است بتوانند با دقت بهتری احساسات و قطبیت متن را پیش‌بینی کنند. مسئله تشخیص طنز، مسئله‌ی دسته‌بندی است و نمونه‌ها باید با یکی از دو برچسب ۱ و ۰ که به ترتیب نشان‌دهنده متن طنز و متن جدی است دسته‌بندی شوند.

در این پژوهش نوآوری‌هایی برای حل مسئله تشخیص طنز در زبان فارسی ارائه می‌شود:

- معرفی دادگان جدید برای تشخیص طنز در زبان فارسی که به صورت خودکار برچسب‌زده شده‌اند،
- ارائه مدل‌هایی که از تطبیق دقیق[1] مدل‌های زبانی از پیش‌آماده‌شده بر روی دادگان به دست آمدند،
- تحلیل و ارزیابی روش‌های پیشنهادی

فصل بعدی، بررسی پیشینه پژوهش مسئله را شامل می‌شود. در فصل سوم راهکارهای پیشنهادی برای حل مسئله معرفی می‌شوند. در فصل چهارم به معرفی دادگان استفاده شده در پژوهش و چگونگی جمع‌آوری آن می‌پردازیم. آزمایش‌های انجام شده و نتایج آن‌ها در فصل پنجم تحلیل شده و نتیجه‌گیری و پیشنهادهایی برای کارهای آینده در فصل ششم قرار گرفته است.

۲ مروری بر کارهای دیگران

نظر به این که پژوهش‌های انجام شده برای تشخیص طنز در زبان فارسی محدود است، برای شناخت ایده‌های مشابه ابتدا به بررسی تشخیص طنز در سایر زبان‌ها می‌پردازیم و سپس کارهای انجام شده بر روی زبان فارسی را بررسی می‌کنیم.

پژوهش‌های قدیمی‌تر از مجموعه ویژگی‌های حاصله از متن به کمک الگوریتم‌های یادگیری ماشین برای دسته‌بندی متون به طنز و جدی استفاده کرده‌اند. از این دسته راهکارها می‌توان به [۲۱] و [۳] اشاره کرد که تنها با ۴ و ۷ ویژگی اقدام به تشخیص طنز کرده‌اند. این راهکارها عموما متن را در نظر نمی‌گیرند.

[1] Fine tuning

بعضی پژوهش‌ها با عنایت به زمینه‌ی متن به تشخیص طنز موجود در متن پرداختند. برای این کار ویژگی‌هایی از نویسنده متن (یا توییت) و یا نوشته‌های پیشین وی استخراج می‌کنند. [۱۹] در کنار این ویژگی‌ها، به تحلیل ویژگی‌های روان‌شناسی و رفتاری نویسنده می‌پردازد. مانند آنچه در [۲] و [۲۲] استفاده شد، این ویژگی‌ها می‌توانند از تعاملی که سایر کاربران با نوشته دارند، باشد. محدودیتی که این روش‌ها ایجاد می‌کنند به دلیل اطلاعات بیشتری است که برای پردازش نیاز دارند و جمع‌آوری دادگانی با این ویژگی‌ها آسان نیست.

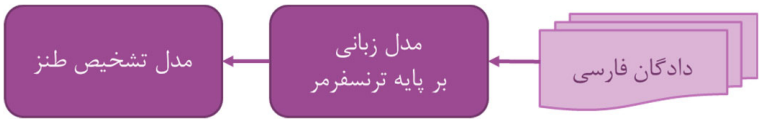
با گسترش استفاده از تکنیک‌های یادگیری عمیق برای حل مسائل پردازش زبان طبیعی، بیشتر پژوهش‌های اخیر از این روش‌ها برای تشخیص طنز در متن استفاده می‌کنند. پژوهش [۱۱] با یادگیری انتقالی از شبکه‌ای عمیق که برای مسئله دسته‌بندی صورتک از پیش آموزش دیده شده بود، به تشخیص طعنه پرداخت. امیر و همکاران [۱] ویژگی‌های نویسنده متن را بر اساس فعالیت‌های قبلی استخراج می‌کرد. در مقاله [۱۲] شبکه‌ای تشکیل‌شده از CNN و LSTM که در دنبال آن قرار می‌گرفت معرفی شد تا تنها بر اساس متن وجود طنز را تشخیص دهد. ژانگ و همکاران به صورت خودکار با استفاده از شبکه عمیق GRU دو سویه، ویژگی‌های معنایی و محتوایی را استخراج می‌کرد. [۲۳] با در نظر گرفتن ویژگی‌های مربوط به نویسنده توییت و متن توییت، مقاله [۱۴] روشی ترکیبی بر اساس محتوا و زمینه متن به کمک شبکه CNN ارائه کردند. کومار و همکاران در [۱۵] به کمک بازنمایی برداری GloVe [۱۸] و استخراج ویژگی‌هایی چون علائم نگارشی استفاده شده، مدل ترکیبی از BiLSTM بر اساس توجه و CNN معرفی کردند.

بدین جهت که شبکه‌های یاد شده برای عملکرد مناسب از دادگانی در مرتبه میلیون توییت استفاده کرده‌اند، روش پیشنهادی محدودیت‌های این شبکه‌ها را در نظر می‌گیرد. مورد دیگری که در این پژوهش به آن توجه می‌شود، استفاده از مدل‌های زبانی که اخیرا معرفی شده‌اند و مرز دانش هستند، به جای روش‌های تعبیه کلمات قدیمی استفاده شده در پژوهش‌های پیشین است.

SemEval مجموعه کارگاه‌های پژوهشی و بین‌المللی در حوزه پردازش زبان طبیعی است که هر ساله به چالش‌های تعدادی از مسئله‌های پردازش زبان طبیعی می‌پردازد. یکی از مسئله‌هایی که تاکنون در این محفل بررسی شده است، مسئله‌ی تشخیص طعنه در زبان انگلیسی و عربی است. بیشتر شرکت‌کنندگان در این پژوهش راهکارهای خود را به کمک مدل‌های از پیش آموزش دیده بر پایه ترنسفرمر ارائه نموده‌اند. [۱۰] ما نیز در این پژوهش از رویکردهای مشابه برای زبان فارسی استفاده می‌کنیم.

تلاش‌های انجام شده بر روی زبان فارسی در پژوهش‌های [۱۷] و [۱۳] دیده می‌شود. بکایی‌نژاد و همکاران در مقاله [۱۷] مدلی بر روی دادگانی از ۲۵۰۰ توییت ارائه دادند. پس از پیش‌پردازش و آماده‌سازی دادگان، ویژگی‌های قطبیت متن، استفاده از علائم نگارشی و ... را استخراج کردند و از آن‌ها برای آموزش مدل‌ها استفاده کردند. در ارزیابی مدل‌های بررسی شده، مدلی بر اساس ماشین بردار پشتیبان با $F1 = ٪۸۰$ بالاترین مقدار را داشت. در این پژوهش متن به عنوان ویژگی در الگوریتم‌های دسته‌بندی استفاده نمی‌شود و برای دسته‌بندی تنها به چند ویژگی محدودی که از متن استخراج کرده‌اند اکتفا می‌کند اما در این پژوهش قصد داریم از متن توییت در الگوریتم تشخیص استفاده کنیم.

گل‌عزیزیان و همکاران در مقاله [۱۳] با استفاده از دو شبکه عمیق و یادگیری انتقالی، تشخیص شوخی



شکل ۱: راهکار پیشنهادی برای مسئله تشخیص طنز در زبان فارسی؛ تطبیق دقیق مدل زبانی بر پایه ترنسفرمر بر دادگان تشخیص طنز به زبان فارسی مدلی برای تشخیص طنز زبان فارسی ارائه می‌دهد.

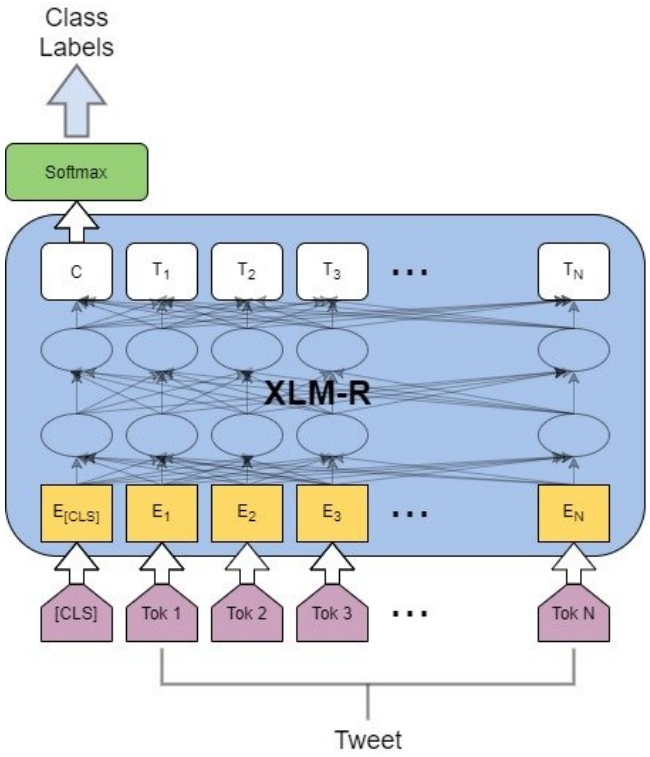
در زبان فارسی را انجام دادند که نتیجه آن $F1 = ٪۸۳$ بود. شبکه اول برای مسئله‌ی پیش‌بینی صورتک و شبکه دوم برای تشخیص طنز ارائه شده بود. در این پژوهش از فست‌تکست [۴] برای تعبیه جملات استفاده شده است. با توجه به این که شبکه اول شخصی‌سازی شده است، در این پژوهش سعی داریم با راهکار عمومی‌تری مسئله را بررسی کنیم همچنین به عنوان جایگزینی برای روش فست‌تکست از تکنیکی با کارایی بیشتر برای تبدیل متن به بردار استفاده کنیم.

۳ راهکار پیشنهادی

پس از معرفی مکانیزم توجه و معماری ترنسفرمر [۷] و مدل زبانی برت [۵] که بر پایه ترنسفرمر بود، مدل‌های زبانی دیگری با معماری مشابهی نسبت به برت توسعه‌داده شدند که در بسیاری از مسائل پردازش زبان طبیعی عملکرد بهتری از خود نشان می‌دادند. پژوهش‌های انجام‌شده برای تشخیص طنز و طعنه در زبان‌های دیگر [۱۰] نشان از کارایی مدل‌های زبانی بر اساس ترنسفرمرها دارد به همین دلیل در ادامه راهکارهایی را با استفاده از مدل‌های زبانی مختلفی که بر پایه ترنسفرمر هستند برای حل مسئله تشخیص طنز و طعنه در زبان فارسی بررسی می‌کنیم.

به عنوان راهکار اول، تطبیق دقیق مدل‌های زبانی مختلف را برای مسئله‌ی دسته‌بندی بررسی می‌کنیم. از میان مدل‌های موجود، به بررسی مدل زبانی پارس‌برت [۹] که تنها بر روی پیکره‌ای از متن رسمی زبان فارسی و با معماری مدل BERTBase [۵] از پیش آموزش‌دیده شده است و دو مدل زبانی XLM-RoBERTaBase و XLM-RoBERTaLarge که چندزبانه هستند می‌پردازیم. این دو مدل به ترتیب با معماری BERTBase و BERTLarge بر روی پیکره‌ای متشکل از ۱۰۰ زبان که هر کدام شامل متون‌هایی به زبان رسمی و محاوره بودند، پیش‌آموزش دیده‌شده است [۶] [۱۶]. در شکل ۱ شمایی از راهکار پیشنهادی دیده می‌شود که نشان می‌دهد برای استفاده از مدل‌های زبانی بر پایه ترنسفرمر مدل از پیش‌آموزش‌دیده‌شده را با دادگان توییت‌های فارسی تطبیق دقیق می‌دهیم تا مدلی برای تشخیص طنز در زبان فارسی حاصل شود. در شکل ۲ معماری دسته‌بندی برای مسئله تشخیص طنز به کمک مدل زبانی برپایه ترنسفرمر مشاهده می‌شود.

مقاله [۸] نشان می‌دهد، تطبیق دقیق مدل‌های زبانی بر پایه ترنسفرمرها بر روی مسئله‌های متفاوت، عملکردی یادگیری بدون نمونه را افزایش می‌دهد. به عنوان راهکار دوم برای تشخیص متن حاوی طنز، تاثیر دادگان با نمونه‌هایی از زبان‌های دیگر را بررسی می‌کنیم. برای این کار پیکره‌ای که از چند دادگان تشخیص



شکل ۲: معماری روش پیشنهادی برای دسته‌بندی بر پایه ترنسفرمر [۲۰]

طنز و طعنه به زبان‌هایی غیر از زبان فارسی تشکیل شده‌است، همراه با مدل زبانی XLM-RoBERTa استفاده می‌کنیم و نتایج مدل به‌دست آمده را بر دادگان فارسی بررسی می‌کنیم. این روش مشابه یادگیری بدون نمونه است. چون در مرحله‌ی تطبیق دقیق اول، مدل نمونه‌ای از دسته‌های حاوی طنز و بدون طنز از زبان فارسی نمی‌بیند. بعد از این کار، بررسی نتایج را با تطبیق دقیق روی دادگان فارسی ادامه می‌دهیم و نتایج به دست آمده را مقایسه می‌کنیم.

۴ دادگان

دو پژوهش انجام شده در حوزه تشخیص طنز و طعنه فارسی هر کدام از دادگانی فارسی برای کار خود استفاده کرده‌اند. نمونه‌های دادگان یاد شده در [۱۷] که به واسطه هشتگ‌های استفاده‌شده در متن توییت برچسب‌دار بودند، به کمک نیروی انسانی نیز بررسی شده‌اند اما متاسفانه این دادگان در دسترس نیست. نمونه‌ها در دادگان تشخیص طنز فارسی که در [۱۳] معرفی شده‌است، به کمک نیروی انسانی برچسب‌گذاری شده اند. این دادگان کمتر از ۳,۰۰۰ توییت را شامل می‌شود و همین‌طور به دلیل مشخص نبودن دادگان آموزش و آزمایش، نمی‌توانیم روش معرفی شده بر روی این دادگان را با روش‌های پیشنهادی این پژوهش

جدول ۱: جدول جزئیات دادگان معرفی‌شده

ویژگی	طنز	بدون طنز
تعداد توییت	۷،۰۱۴	۷،۹۳۲
میانگین طول توییت	۳۰	۴۵
طول بیشینه توییت	۲۶۰	۴۳۰

مقایسه کنیم. در نهایت به دلیل این کمبودها بر آن شدیم دادگانی مناسب و غنی با نمونه‌های بیشتر برای حل مسئله تشخیص طنز و طعنه در زبان فارسی و به طور اختصاصی از توییتر ارائه دهیم.

تجربه [۱۳] برای جمع‌آوری و برچسب زدن داده، بیان می‌کرد نرخ پیام‌های طنز در کانال تلگرامی «توییتر فارسی» بیشتر از توییت‌های روزمره است. نظر به این که به‌روزرسانی پیام‌رسان تلگرام از تاریخ ۳۰ دسامبر ۲۰۲۱ برای هر پیام اجازه ثبت واکنش را از طریق انتخاب یکی از صورتک‌های « 😁👍👎❤️😢 » به کاربران تلگرامی می‌دهد، بر آن شدیم که دادگانی از توییت‌ها را از این کانال در پیام‌رسان تلگرام جمع‌آوری کنیم.

به دلیل کمبود تعداد واکنش‌های ثبت‌شده برای پیام‌هایی که زمان ارسال آن‌ها با به‌روزرسانی اخیر تلگرام فاصله زیادی دارد، از مجموع ۱۲۹,۵۳۶پیام یکتا که از کانال توییتر فارسی دریافت شد، تنها به ۳۱,۹۸۳ پیام که در بازه‌ی زمانی ابتدای دسامبر ۲۰۲۱ تا ۲۵ آگوست ۲۰۲۲ ارسال شده‌بودند اکتفا کردیم. از میان این پیام‌ها، از درنظر گرفتن آن‌هایی که ویدئو یا تصویر همراه خود دارند نیز صرفه‌نظر کردیم.

برای دادگانی که به برچسب‌زنی دستی نیاز دارند، برچسب‌زن‌ها طبق مجموعه‌ای از قوانین که معمولا زبان‌شناسان مشخص می‌کنند، اقدام به اختصاص دسته به هر نمونه می‌کنند. در انتها میانگینی از برچسب‌های اختصاص داده شده، دسته هر نمونه را مشخص می‌کند. نظر به ماهیت خلاقانه طنز، زبان‌شناسان قادر نیستند قوانینی کلی برای تشخیص آن وضع کنند و همین مسئله‌ی برچسب‌زدن را دشوارتر می‌سازد؛ ضمن این که برچسب زدن به خودی خود فرایندی زمان‌گیر است. واکنش‌های ثبت‌شده برای هر پیام می‌تواند نشان‌دهنده برچسب‌هایی باشد که کاربران تلگرام به هر پیام اختصاص داده‌اند. می‌توانیم این مسئله را برچسب‌گذاری با تعداد زیادی برچسب‌زن ببینیم که پیام‌ها را خوانده‌اند و در صورتی که متوجه طنز آشکار یا پنهان پیام شده‌اند از واکنش « 😁 » برای بیان نظرشان استفاده کرده‌اند.

برای ۱۸,۱۳۳ پیام باقی‌مانده تا این مرحله، راهکاری خودکار برای اختصاص برچسب اعمال کردیم. این فرایند که برای هر نمونه بر اساس دو واکنشی که تکرار بیشینه داشتند انجام می‌شد، به پیام‌هایی که بیشترین واکنش آن‌ها « 😁 » بود برچسب ۱ به نشانه طنز اختصاص می‌داد. اگر بیشترین واکنش ثبت‌شده برای نمونه‌ای « 👍 » باشد درحالی که واکنش دوم از نظر تعداد صورتک « 😁 » است، مرز تشخیص طنز و جدی ساده نیست و این دسته پیام‌ها را نادیده می‌گیریم. سایر پیام‌ها برچسب ۰ به نشانه پیامی جدی دریافت می‌کنند. در جدول ۱ جزئیات مربوط به دادگان معرفی شده و در جدول ۲ نمونه‌هایی از این دادگان مشاهده می‌شود.

جدول ۲: جدول نمونه‌هایی از دادگان معرفی شده همراه با برچسب

توییت	برچسب
پشت یه کامیونه نوشته بود: سلطان خیانت هیدروژن! هم پیوند کوالانسی میگیره هم هیدروژنی! فکر کنم رانندش لیسانس شیمی داشته 😕😂🤦	۱
آره مهاجرت خوبه ولی قشنگترش این بود که همینجا کنار خانواده و دوستامون به خواسته‌هایی که داشتیم برسیم :(	۰
تاس کباب داشتیم بابام جفت شیش آورد همه‌شو خورد	۱
مدیون تاول های پامون تو راه اشتباه نباشیم! هر جا که فهمیدیم مسیر درست را انتخاب نکردیم،بدون تردید دور بزنیم و برگردیم!	۰

برای اجرای راهکار پیشنهادی دوم، پیکره تشخیص طنز و طعنه را از دادگان تشخیص طنز یا طعنه به زبان‌های انگلیسی، پرتغالی برزیل، عربی و هندی را گردهم آوردیم که شامل ۹۴۴۲۸ توییت یکتا است که از این تعداد، ۶۵۷۱۱ توییت طنز هستند.

۵ آزمایش‌ها

در این بخش به آزمایش‌هایی که برای بررسی راهکارهای پیشنهادی انجام شده‌اند، می‌پردازیم. برای اجرای آزمایش‌های از پیکره چند زبانه تشخیص طنز و دادگان فارسی معرفی شده که هر دو در فصل چهارم بررسی شدند استفاده می‌شود. برای آموزش و ارزیابی مدل‌ها، دادگان تشخیص طنز در زبان فارسی را به تصادف با نسبت ۸ به ۲ به قسمت‌های آموزش و آزمایش تقسیم شدند. معیارهای دقت، صحت ، فراخوانی و $F1$ برای ارزیابی و مقایسه مدل‌ها به ما کمک می‌کنند.

۱.۵ مدل‌های پایه

برای مقایسه عملکرد روش‌های پیشنهادی، به مدل‌های پایه نیاز داریم. از الگوریتم‌های یادگیری ماشین به عنوان مدل پایه استفاده می‌کنیم.

الگوریتم نایو بیز، ماشین بردار پشتیبان و مدل شبکه عصبی چندلایه را همراه با روش tfidf به عنوان راهکاری برای تبدیل متن به بردار استفاده می‌کنیم.

برای الگوریتم نایو بیز از روش گاوسی بهره‌بردیم و برای الگوریتم ماشین بردار پشتیبان از هسته تابع پایه‌ای شعاعی استفاده کردیم. در الگوریتم پرسپترون چندلایه ، شبکه‌ای با دو لایه پنهان که هر لایه شامل ۵۰ گره است مورد استفاده قرار گرفت. این شبکه با تابع تانژانت هیپربولیک فعال‌سازی می‌شد و با تابع آدام بهینه‌سازی صورت می‌گرفت.

در میان سه الگوریتم یاد شده، الگوریتم ماشین بردار پشتیبان با $F1 = ٪۷۹$ عملکرد بهتری از خود نشان داد.

۲.۵ آزمایش‌های روش تطبیق دقیق مدل‌های زبانی بر پایه ترنسفرمر

به عنوان روش پیشنهادی اول، از مدل‌های زبانی بر پایه ترنسفرمر استفاده شد. مدل زبانی پارس‌برت و دو نسخه مدل زبانی XLM-RoBERTa، ترنسفرمرهای انتخابی هستند. هر کدام از این ترنسفرمرهای مورد نظر، بر روی قسمت آموزش دادگان تطبیق دقیق داده شدند. از میان سه مدل زبانی انتخابی، نسخه بزرگ XLM-RoBERTa با معیار $F1$ برابر با ۸۴٫۶٪ کارایی بهتری را نمایش می‌دهد.

۳.۵ آزمایش‌های روش آموزش بدون نمونه

روش پیشنهادی دیگر، استفاده از پیکره متشکل از چندین زبان غیر از زبان فارسی است. برای انجام این آزمایش، تطبیق دقیق مدل زبانی XLM-RoBERTa بزرگ را بر روی پیکره چندزبانه انجام می‌دهیم. این مدل زبانی به دلیل چندزبانه بودن و پیش‌آموزش بر روی صد زبان که هر چهار زبان پیکره طنز و طعنه چندزبانه گردآوری شده را شامل می‌شود، انتخاب شده است. مدل حاصله را بر روی بخش آزمایش دادگان فارسی معرفی‌شده ارزیابی شد. مدلی که هیچ نمونه‌ای حاوی شوخی یا بدون شوخی از زبان فارسی را ندیده است، می‌تواند با دقت ۷۰٫۶٪ مقدار $F1$ برابر با ۷۰٫۶٪ کسب کند. مدل زبانی که در این مرحله به دست آمده را بر روی دادگان فارسی معرفی شده نیز تطبیق دقیق داده شد. مدل نهایی می‌تواند وجود طنز و طعنه در قسمت آزمایش دادگان را با دقت و $F1$ برابر با ۸۴٫۵٪ پیش‌بینی کند.

۴.۵ مقایسه نتایج

در روش‌هایی که مدل از تطبیق دقیق مدل‌های زبانی بر روی دادگان فارسی به‌دست می‌آید، می‌توانیم مدل‌ها را از دو دیدگاه بررسی کنیم؛ دیدگاه پیکره‌ای که مدل بر آن از پیش‌آموزش دیده شده است و دیدگاه معماری. در دیدگاهی که به تعداد زبان و حجم پیکره برای پیش‌آموزش توجه می‌شود، نتایج آزمایش‌ها نشان می‌دهد استفاده از مدل چندزبانه عملکرد بهتری نسبت به مدل زبانی که تنها بر روی زبان فارسی از پیش‌آموزش‌دیده‌شده است دارد. کارایی دو مدل چندزبانه می‌تواند به دلیل حجم بزرگ‌تر پیکره‌ای باشد که بر روی آن آموزش‌دیده‌شده‌اند. با توجه به این که پیکره این مدل زبانی چندزبانه شامل نمونه‌هایی از متون با زبان محاوره و عامیانه می‌شود، دقت بالاتر این دو مدل زبانی برای مسئله‌ای که نمونه‌های آن لزوما به زبان فارسی رسمی نوشته نشده‌اند نسبت به پارس‌برت با پیکره‌ای که غالبا به زبان فارسی رسمی است، توجیه می‌شود. از دیدگاه معماری، مدل زبانی XLM-RoBERTa بزرگ با داشتن تعداد پارامترهای بیشتر، کارایی بهتری از خود نشان می‌دهند. به همین دلیل از میان آزمایش‌های تطبیق دقیق مدل‌های زبانی بر پایه ترنسفرمر بر دادگان فارسی، استفاده از نسخه بزرگ مدل زبانی XLM-RoBERTa عملکرد بهتری برای تشخیص طنز بر روی قسمت آزمایش دادگان فارسی معرفی‌شده دارد.

در آزمایشی که مدل زبانی XLM-RoBERTa بزرگ در دو مرحله بر روی دو پیکره تشخیص طنز چندزبانه و دادگان تشخیص طنز فارسی تطبیق دقیق داده می‌شود، می‌بینیم نتیجه معیارهای ارزیابی با حالتی که تنها بر روی دادگان زبان فارسی تطبیق دقیق انجام می‌شود تفاوتی کمتر از ۰٫۳٪ دارد و به نظر

جدول ۳: نتایج آزمایش‌های انجام شده بر روی دادگان فارسی معرفی شده

آزمایش	دقت	فراخوانی	صحت	F1
نایو بیز	۶۶٪	۶۸٪	۶۷٪	۶۵٪
ماشین بردار پشتیبان	۷۹٪	۷۸٪	۷۹٪	۷۹٪
شبکه پرسپترون چندلایه	۷۳٪	۷۳٪	۷۳٪	۷۳٪
پارس‌برت، نسخه ۳	۸۱.۳٪	۸۱.۴٪	۸۱.۳٪	۸۱.۳٪
$\text{XLM-RoBERTa}_{\text{Base}}$	۸۲.۶٪	۸۲.۸٪	۸۲.۶٪	۸۲.۵٪
$\text{XLM-RoBERTa}_{\text{Large}}$	**۸۴.۷٪**	**۸۴.۷٪**	**۸۴.۶٪**	**۸۴.۶٪**
پیش‌آموزش‌دیده بر روی سایر زبان‌ها	۷۰.۶٪	۷۰.۷٪	۷۰.۶٪	۷۰.۶٪
تطبیق دقیق مدل پیش‌آموزش‌دیده بر روی سایر زبان‌ها بر روی دادگان فارسی	۸۴.۵٪	۸۴.۶٪	۸۴.۴٪	۸۴.۵٪

می‌رسد استفاده از دادگان زبان‌های دیگر برای تطبیق دقیق هنگامی که از مدل زبانی چندزبانه استفاده می‌شود، تاثیر مثبتی در عملکرد تشخیص طنز زبان فارسی ندارد.

جزئیات ارزیابی هر یک از آزمایش‌های بیان شده در قسمت‌های ۱.۵ تا ۳.۵ در جدول ۳ دیده می‌شود. از میان آزمایش‌ها بر روی مدل پایه، مدل‌های زبانی بر پایه ترنسفرمر و روش استفاده از پیکره تشخیص طنز زبان‌های دیگر، تطبیق دقیق نسخه بزرگ مدل زبانی XLM-RoBERTa بر روی دادگان فارسی با کسب مقدار ۸۴٫۶٪ برای معیار $F1$ عملکرد بهتری برای تشخیص طنز موجود در قسمت آزمایش دادگان فارسی معرفی‌شده داشته است.

۶ نتیجه‌گیری و کارهای آینده

در این پژوهش ضمن معرفی دادگان جدیدی برای تشخیص طنز در زبان فارسی، دو راهکار کلی برای تشخیص طنز در زبان فارسی ارائه شد. هر دو این روش‌ها با تطبیق دقیق مدل‌های زبانی بر پایه ترنسفرمرها بر روی دادگان معرفی شده و پیکره گردآوری شده از چهار زبان دیگر، مدل‌هایی برای تشخیص طنز در زبان فارسی ارائه می‌دهند. از میان مدل‌های پیشنهادی و آزمایش‌شده، مدل حاصله از تطبیق دقیق مدل زبانی XLM-RoBERTaLarge بر روی دادگان معرفی شده، بهترین عملکرد را داراست. این مدل بر روی قسمت آزمایش دادگان فارسی معرفی شده می‌تواند با دقت ۸۴٫۷٪ و ۸۴٫۶٪ وجود طنز در متن فارسی را تشخیص دهد.

از آنجایی که یکی از اهداف تشخیص طنز در متن، افزایش عملکرد مدل‌های تحلیل احساسات است، بررسی تاثیر وجود ویژگی که نشان دهد نمونه طنز است یا خیر در کنار متن و سایر ویژگی‌هایی که برای تشخیص قطبیت متن در مسئله خاصی استفاده می‌شود، می‌تواند یکی از کارهای آینده در این زمینه باشد. روش ارائه شده برای جمع‌آوری دادگان می‌تواند به صورت خودکار فرایند برچسب‌زنی را نیز انجام دهد. این روش راه‌اندازی سیستم‌هایی که نیاز به بروزرسانی مداوم مدل را دارند تسهیل می‌بخشد و باعث می‌شود مدل با نمونه‌های طنز تازه‌ای آشنا شود که به خاطر ذات خلاقانه طنز به مرور زمان ایجاد شده‌اند.

سپاس‌گزاری

اجرای آزمایش‌های مربوط به مدل‌های زبانی بر پایه ترنسفرمر بر روی واحد پردازش گرافیکی به زمان کمتری نیاز دارد. از شرکت دادماتک که سخت‌افزار مورد نیاز اجرای آزمایش‌های این پژوهش را در اختیارمان قرار داد سپاس‌گزاریم.

مراجع

[1] Amir, S., Wallace, B. C., Lyu, H., and Silva, P. C. M. J. "Modelling context with user embeddings for sarcasm detection in social media". *arXiv preprint arXiv:1607.00976 20*, 5 (2016), 1924–1939.

[2] Bamman, D., and Smith, N. "Contextualized sarcasm detection on twitter". *proceedings of the international AAAI conference on web and social media 9*, 1 (2015), 574–577.

[3] Barbieri, F., Saggion, H., and Ronzano, F. "Modelling sarcasm in twitter, a novel approach". *proceedings of the 5th workshop on computational approaches to subjectivity, sentiment and social media analysis* (2014), 50–58.

[4] Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. "Enriching word vectors with subword information". *Transactions of the association for computational linguistics 5* (2017), 135–146.

[5] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. "Enriching word vectors with subword information". *arXiv preprint arXiv:1810.04805* (2018).

[6] et al., A. C. "Unsupervised cross-lingual representation learning at scale". *arXiv preprint arXiv:1911.02116* (2019).

[7] et al., A. V. "Attention is all you need". *Advances in neural information processing systems 30* (2017).

[8] et al., J. W. "Finetuned language models are zero-shot learners". *arXiv preprint arXiv:2109.01652* (2021).

[9] Farahani, M., Gharachorloo, M., Farahani, M., and Manthouri, M. "Parsbert: Transformer-based model for persian language understanding". *Neural Processing Letters 53*, 6 (2021), 3831–3847.

[10] Farha, I. A., Oprea, S. V., Wilson, S., and Magdy, W. "SemEval-2022 Task 6: iSarcasmEval, Intended Sarcasm Detection in English and Arabic". *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)* (2022), 802–814.

[11] Felbo, B., Mislove, A., Sogaard, A., Rahwan, I., and Lehmann, S. "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm". *arXiv preprint arXiv:1708.00524* (2017).

[12] Ghosh, A., and Veale, T. "Fracking sarcasm using neural network". *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis* (2016), 161–169.

[13] Golazizian, P., Sabeti, B., Asli, S. A. A., Majdabadi, Z., Momenzadeh, O., and Fahmi, R. "Irony detection in Persian language: A transfer learning approach using emoji prediction". *Proceedings of The 12th Language Resources and Evaluation Conference* (2020), 2839–2845.

[14] Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., and Mihalcea, R. "Cascade: Contextual sarcasm detection in online discussion forums". *arXiv preprint arXiv:1805.06413* (2018), 1490–1498.

[15] Kumar, A., Sangwan, S. R., Arora, A., Nayyar, A., and Abdel-Basset, M. "Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network". *IEEE access 7* (2019), 23319–23328.

[16] Lample, G., and Conneau, A. "Cross-lingual language model pretraining". *arXiv preprint arXiv:1901.07291* (2019).

[17] Nezhad, Z. B., and Deihimi, M. A. "Sarcasm detection in Persian". *Journal of Information and Communication Technology 20*, 1 (2021), 1–20.

[18] Pennington, J., Socher, R., and Manning, C. D. "Glove: Global vectors for word representation". *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (2014), 1532–1543.

[19] Rajadesingan, A., Zafarani, R., and Liu, H. "Sarcasm detection on twitter: A behavioral modeling approach". *Proceedings of the eighth ACM international conference on web search and data mining* (2015), 97–106.

[20] Ranasinghe, T., Gupte, S., Zampieri, M., and Nwogu, I. "Wlv-rit at hasoc-dravidiancodemix-fire2020: Offensive language identification in code-switched youtube comments". *arXiv preprint arXiv:2011.00559* (2020).

[21] Reyes, A., Rosso, P., and Veale, T. "A multidimensional approach for detecting irony in twitter". *Language resources and evaluation 47*, 1 (2013), 239–268.

[22] Wallace, B. C., and Charniak, E. "Sparse, contextually informed models for irony detection: Exploiting user communities, entities and sentiment". *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (2015), 1035–1044.

[23] Zhang, M., Zhang, Y., and Fu, G. "Tweet sarcasm detection using deep neural network". *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: technical papers* (2016), 2449–2460.