



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

پروژه کارشناسی
گرایش هوش مصنوعی

پیاده سازی ابزار جمع آوری اخبار جعلی فارسی
و دسته بندی آن

نگارش

محمد رضا صمدی

استاد راهنما

دکتر سعیده ممتازی

بهمن ۱۳۹۹

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

صفحه فرم ارزیابی و تصویب پایان نامه- فرم تأیید اعضاء کمیته دفاع

در این صفحه فرم دفاع یا تأیید و تصویب پایان نامه موسوم به فرم کمیته دفاع- موجود در پرونده آموزشی- را قرار دهید.

نکات مهم:

- نگارش پایان نامه/رساله باید به **زبان فارسی** و بر اساس آخرین نسخه دستورالعمل و راهنمای تدوین پایان نامه های دانشگاه صنعتی امیرکبیر باشد.(دستورالعمل و راهنمای حاضر)
- رنگ جلد پایان نامه/رساله چاپی کارشناسی، کارشناسی ارشد و دکترا باید به ترتیب مشکی، طوسی و سفید رنگ باشد.
- چاپ و صحافی پایان نامه/رساله بصورت **پشت و رو(دورو)** بلامانع است و انجام آن توصیه می شود.

به نام خدا



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

تعهدنامه اصالت اثر

تاریخ: بهمن ۱۳۹۹

اینجانب محمدرضا صمدی متعهد می‌شوم که مطالب مندرج در این پایان‌نامه حاصل کار پژوهشی اینجانب تحت نظارت و راهنمایی اساتید دانشگاه صنعتی امیرکبیر بوده و به دستاوردهای دیگران که در این پژوهش از آنها استفاده شده است مطابق مقررات و روال متعارف ارجاع و در فهرست منابع و مآخذ ذکر گردیده است. این پایان‌نامه قبلاً برای احراز هیچ مدرک هم‌سطح یا بالاتر ارائه نگردیده است. در صورت اثبات تخلف در هر زمان، مدرک تحصیلی صادر شده توسط دانشگاه از درجه اعتبار ساقط بوده و دانشگاه حق پیگیری قانونی خواهد داشت.

کلیه نتایج و حقوق حاصل از این پایان‌نامه متعلق به دانشگاه صنعتی امیرکبیر می‌باشد. هرگونه استفاده از نتایج علمی و عملی، واگذاری اطلاعات به دیگران یا چاپ و تکثیر، نسخه‌برداری، ترجمه و اقتباس از این پایان‌نامه بدون موافقت کتبی دانشگاه صنعتی امیرکبیر ممنوع است. نقل مطالب با ذکر مآخذ بلامانع است.

محمدرضا صمدی

امضا

سپاس‌گزاری

اکنون که مراحل پژوهش، تدوین و نگارش پایان نامه به پایان رسیده است، از مادر و پدر عزیزتر از جانم متشکرم که درطول زندگی و دوران تحصیل همراه و مشوقم بوده‌اند و با ایثار و از خودگذشتگی و تحمل زحمات، مرا در این راه یاری نمودند. از سرکار خانم دکتر ممتازی که به عنوان استاد راهنما با سعه صدر در مسیر این پژوهش، همواره راهنما و راهگشای اینجانب بوده‌اند تقدیر و تشکر می‌نمایم.

محمد رضا صدیقی

بهمن ۱۳۹۹

چکیده

امروزه نرخ انتشار اخبار در بستر شبکه‌های اجتماعی و یا وبسایت‌های خبری با سرعت بسیار بیشتری نسبت به گذشته در حال رشد است. وجود برخی از اخبار و یا مطالب تأیید نشده در میان این جریان گسترده خبری، آثار منفی بسیاری را بر روی افکار عمومی و حتی تصمیمات کلان دولت‌ها دارد. به منظور مقابله با انتشار اخبار جعلی و جلوگیری از تخریب اعتماد عمومی، در این پروژه قصد داریم تا سامانه هوشمندی را به منظور تشخیص صحت اخبار منتشر شده در بستر وبسایت‌های خبری پیاده‌سازی کنیم. با توجه به عدم وجود یک مجموعه داده جامع اخبار جعلی در زبان فارسی، ابتدا باید روشی را به منظور استخراج دادگان مناسب برای این فعالیت ارائه و پیاده‌سازی کنیم. در این پروژه ما با استفاده از روش ارائه شده مجموعه دادگانی را برای تشخیص اخبار جعلی معرفی کرده‌ایم. علاوه بر این مجموعه داده، ما مدل‌های نوین عمیقی را معرفی کردیم تا با استفاده از آن‌ها بتوانیم اخبار جعلی منتشر شده در وبسایت‌های خبری را تشخیص دهیم. تمامی مدل‌های معرفی شده در این پروژه از مدل‌های از پیش‌آموزش‌دیده مبتنی بر معماری انتقال‌دهنده برای استخراج ویژگی‌های مبتنی بر بافت استفاده کرده‌اند. همچنین به منظور ارزیابی کارایی مدل‌های معرفی شده در این پروژه، ما از دو مجموعه داده تشخیص اخبار جعلی در زبان انگلیسی و دو مجموعه دادگان دیگر فارسی که مرتبط با تشخیص شایعات فضای مجازی بودند نیز استفاده کردیم.

واژه‌های کلیدی:

تشخیص اخبار جعلی - شبکه عصبی عمیق - بازنمایی مبتنی بر بافت - ابزار جمع‌آوری اخبار

فهرست مطالب

صفحه

عنوان

۱	مقدمه و طرح مسئله	۱
۲	۱-۱ مقدمه	۲
۲	۲-۱ طرح مسئله	۲
۳	۳-۱ راه حل پیشنهادی	۳
۴	۲ پیشینه تشخیص اخبار جعلی	۴
۵	۱-۲ کارهای انجام شده در زبان انگلیسی	۵
۸	۲-۲ دادگان موجود در انگلیسی و آمار دادگان	۸
۸	۱-۲-۲ دادگان لیار	۸
۱۰	۲-۲-۲ دادگان آی.اس.اُ.تی	۱۰
۱۰	۳-۲-۲ دادگان بازفید	۱۰
۱۰	۴-۲-۲ توئیتز ۱۵	۱۰
۱۱	۳-۲ دادگان موجود در سایر زبان‌ها	۱۱
۱۲	۴-۲ کارهای انجام شده و دادگان موجود در زبان فارسی	۱۲
۱۴	۳ تهیه دادگان فارسی «تاج»	۱۴
۱۵	۱-۳ مقدمه‌ای بر جمع‌آوری داده‌ها	۱۵
۱۵	۲-۳ مراحل جمع‌آوری داده‌ها	۱۵
۱۶	۳-۳ نشانه‌گذاری داده	۱۶
۱۹	۴-۳ آمار دادگان	۱۹
۲۱	۴ معماری مدل‌های ارائه شده	۲۱
۲۲	۱-۴ مقدمه	۲۲
۲۲	۲-۴ مدل‌های دسته‌بندی متن	۲۲
۲۲	۱-۲-۴ شبکه عصبی پرسپترون ساده	۲۲
۲۳	۲-۲-۴ شبکه عصبی پیچشی	۲۳
۲۵	۳-۴ بازنمایی برداری اخبار با رویکردهای مختلف	۲۵

۲۶	۱-۳-۴ بازنمایی مبتنی بر واژه
۲۷	۲-۳-۴ بازنمایی مبتنی بر تعبیه ایستا (ورد۲وک)
۲۸	۳-۳-۴ بازنمایی مبتنی بر موضوع (تخصیص نهان دیریکله)
۲۹	۴-۳-۴ مدل‌های بازنمایی متن مبتنی بر بافت
۳۳	۴-۴ معماری‌های پیشنهادی برای تشخیص اخبار جعلی
۳۳	۱-۴-۴ مدل زبانی مبتنی بر بافت + پرسپترون ساده
۳۵	۲-۴-۴ مدل زبانی مبتنی بر بافت + پیچشی
۳۷	۵ بررسی نتایج و ارزیابی مدل‌های ارائه شده
۳۸	۱-۵ آزمایش‌های انجام‌شده بر روی پیکره «تاج»
۳۸	۱-۱-۵ پیش‌پردازش داده‌های خبری
۳۹	۲-۱-۵ تنظیمات آزمایش‌ها
۴۰	۳-۱-۵ نتایج
۴۱	۲-۵ آزمایش‌های انجام‌شده بر روی سایر پیکره‌های فارسی
۴۲	۳-۵ آزمایش‌های انجام‌شده بر روی پیکره‌های زبان انگلیسی
۴۸	۶ بهبود تشخیص اخبار جعلی با فراداده‌های حاصل از پردازش زبان
۴۹	۱-۶ انگیزه
۴۹	۲-۶ استخراج ویژگی‌های موضوعی مستندات
۵۰	۱-۲-۶ دسته‌بندی بانظارت براساس ۶ دسته مختلف
۵۰	۲-۲-۶ دسته‌بندی بدون نظارت براساس روش تخصیص نهان دیریکله
۵۱	۳-۶ تشخیص موجودیت‌های نامدار
۵۲	۴-۶ تحلیل احساسات
۵۴	۵-۶ نتایج حاصل از استفاده از پردازش‌های متنی در تشخیص اخبار
۵۸	۷ جمع‌بندی و کارهای آتی
۵۹	۱-۷ جمع‌بندی و نتیجه‌گیری
۵۹	۲-۷ کارهای آتی
۶۰	منابع و مراجع

شکل	فهرست اشکال	صفحه
۱-۴	نمایی از شبکه عصبی دارای نورون‌های مدل پرسپترون ساده	۲۲
۲-۴	نمای کلی مدل پیچشی الکسنت (کریژفسکی و دیگران، ۲۰۱۲)	۲۴
۳-۴	نمای کلی یک شبکه عصبی عمیق پیچشی (له و دیگران، ۲۰۱۷)	۲۵
۴-۴	نمای کلی دو مدل پرش‌نگاشت و کیسه‌واژه پیوسته (سلیمان و دیگران، ۲۰۱۷)	۲۸
۵-۴	نحوه ترکیب تعبیه‌های سه‌گانه در مدل زبانی برت (دولین و دیگران، ۲۰۱۸)	۳۰
۶-۴	نحوه عملکرد یادگیری مدل برت (دولین و دیگران، ۲۰۱۸)	۳۱
۷-۴	نحوه استفاده از یک لایه پرسپترون ساده در کنار مدل برت	۳۴
۸-۴	نحوه ارتباط لایه پیچشی با بردارهای خروجی مدل برت	۳۵
۱-۵	ماتریس درهم‌ریختگی نتایج آزمایش‌های دادگان لیار	۴۴
۲-۵	ماتریس درهم‌ریختگی نتایج آزمایش‌های دادگان آی.اس.اُ.تی.	۴۶
۱-۶	شمای کلی مدل ارائه شده برای به‌کار بردن تمام فراداده‌ها و دسته‌بندی اخبار با	
۵۶	استفاده از دسته‌بند پیچشی	

فهرست جداول

صفحه

جدول

۱-۲	آمار و اطلاعات مرتبط با دادگان شایعات در شبکه‌های اجتماعی (لی و دیگران، ۲۰۱۹) .	۹
۲-۲	آمار و اطلاعات مربوط به دادگان لیار	۹
۳-۲	آمار و اطلاعات مربوط به دادگان آی.اس.اُ.تی	۱۰
۴-۲	آمار و اطلاعات مربوط به دادگان بازفید	۱۱
۵-۲	آمار و اطلاعات مربوط به دادگان توییتر ۱۵	۱۱
۶-۲	لیست دادگان موجود در سایر زبان‌ها	۱۱
۷-۲	آمار دادگان موجود فارسی	۱۳
۱-۳	آمار دادگان «تاج»	۲۰
۱-۴	توابع فعال‌ساز پرکاربرد به‌همراه روابط و نمودار آنها	۲۳
۲-۴	روابط قابل استفاده برای محاسبه بسامد واژه در یک سند	۲۶
۳-۴	روابط قابل استفاده جهت محاسبه معکوس بسامد سند	۲۷
۱-۵	نتایج تشخیص اخبار جعلی فارسی با استفاده از بازنمایی‌های مختلف و دسته‌بندی با شبکه عصبی پیچشی	۴۱
۲-۵	نتایج تشخیص اخبار جعلی فارسی با استفاده از بازنمایی پارس‌برت و مقایسه دسته‌بندها	۴۱
۳-۵	ارزیابی و مقایسه مدل‌های ارائه‌شده با مدل‌های پیشین بر روی دادگان فارسی شبکه‌های اجتماعی	۴۲
۴-۵	نتایج آزمایش مدل‌های ارائه‌شده بر روی دادگان لیار (۲ برچسب)	۴۳
۵-۵	مقایسه مدل‌های ارائه شده با مدل‌های پایه براساس دادگان لیار (۶ برچسب)	۴۳
۶-۵	آمار تاریخچه گوینده دو نمونه خبر ۱ و ۲	۴۵
۷-۵	آمار واژگان یک نمونه ۱ (خبر جعلی دسته‌بندی‌شده در دسته خبر اصیل)	۴۵
۸-۵	آمار واژگان نمونه ۲ (خبر اصیل دسته‌بندی‌شده در دسته خبر جعلی)	۴۵
۹-۵	مقایسه مدل‌های ارائه‌شده با مدل‌های پیشین بر روی دادگان آی.اس.اُ.تی.	۴۶
۱-۶	آمار و اطلاعات مربوط به دادگان همشهری	۵۰
۲-۶	راهنمای برچسب واژگان	۵۳

- ۳-۶ مثال از یک جمله برچسب‌خورده توسط مدل تشخیص موجودیت نامدار ۵۵
- ۴-۶ نتایج تشخیص اخبار جعلی فارسی با استفاده از بازنمایی پارس‌برت و دسته‌بندهای
- مختلف ۵۷

فصل اول

مقدمه و طرح مسئله

۱-۱ مقدمه

در گذشته انتشار اخبار تنها از طریق روزنامه و تلویزیون انجام می‌شد؛ اما امروزه با گسترش چشمگیر رسانه‌های اجتماعی و وبگاه‌های خبری، حجم بالایی از اطلاعات از جمله اخبار، به‌راحتی در میان کاربران مبادله می‌شود. با افزایش روزافزون تعداد اخبار منتشرشده در این رسانه‌ها، تشخیص درستی و صحت این اخبار از اهمیت ویژه‌ای برخوردار است؛ چراکه اخبار منتشرشده توسط رسانه‌ها یا افراد در رسانه‌های اجتماعی ممکن است جعلی باشد و به‌سرعت میان افراد دست‌به‌دست شود. به‌عنوان مثال، توییتر یکی از محبوب‌ترین رسانه‌های اجتماعی برای به اشتراک گذاشتن اخبار و نظرات در مورد رخداد‌های مختلف توسط کاربران و اصحاب رسانه است. طبق آمار، روزانه حدود ۵۳۳ میلیون توییتر در توییتر به اشتراک گذاشته می‌شود که این رقم می‌تواند نشان‌دهنده مناسب‌بودن این بستر برای انتشار اخبار جعلی در میان کاربران باشد. به منظور ایجاد یک چهارچوب مناسب برای طرح و حل این مسئله، پژوهشگران بسیاری از دیدگاه‌های مختلفی همچون جامعه‌شناسی، سیاسی و غیره، تعریف‌هایی از عبارت “خبر جعلی” ارائه کرده‌اند. [لیزر و دیگران \(۲۰۱۸\)](#) با دیدگاه سیاسی در ارتباط با انتخابات سال ۲۰۱۶ آمریکا، به تعریف خبر جعلی پرداخته است. به تعبیر آن‌ها، خبر جعلی اطلاعاتی ساختگی است که از لحاظ ساختاری تقلیدی از خبر اصیل است اما از لحاظ نیت و فرایند انتشار کاملاً متفاوت با آن است. [رابین و دیگران \(۲۰۱۵\)](#) خبر جعلی را خبری کذب یا دروغ عنوان می‌کنند که می‌تواند باعث فریب مردم شود. این گونه اخبار شامل سه دسته جعل جدی^۱، کلاهبرداری بزرگ^۲ و جعل طنزآمیز^۳ است. این اخبار عمدتاً با هدف گمراه کردن و به‌منظور آسیب‌رساندن به یک گروه یا افرادی با اهداف مالی و سیاسی منتشر می‌شود و می‌تواند بر عقاید افراد و تصمیم‌های شخصی آنها تأثیر مستقیم داشته باشد. براساس پژوهش‌های انجام‌شده، میزان اثرگذاری خبر جعلی پنج برابر اخبار موثق و واقعی است.

۲-۱ طرح مسئله

اخبار جعلی بر روی افکار عمومی کشورها و حتی اقدامات سیاسی دولت‌ها و نهادهای بین‌المللی تأثیر قابل‌توجه‌ای دارد. نمونه آن، انتخابات سال ۲۰۱۶ ریاست جمهوری آمریکا است که حجم زیادی از اخبار جعلی در ارتباط با کاندیداها در رسانه‌های اجتماعی منتشر شد و تأثیر بسیار زیادی بر روی نتیجه این

¹Serious fabrication

²Large-scale hoaxes

³Humorous fakes

انتخابات داشت. در سال‌های اخیر در ایران نیز شاهد انتشار گسترده اخبار جعلی در مورد مسائل سیاسی و اجتماعی در شبکه‌های مجازی به‌منظور کنترل افکار جامعه و ایجاد حس بی‌اعتمادی بوده‌ایم. از این‌رو، بسیاری از رسانه‌های اجتماعی برای جلوگیری از انتشار این اخبار راهکارهایی را ارائه کرده‌است. برای مثال، فیسبوک برای جلوگیری از انتشار اخبار جعلی در این رسانه اجتماعی توسط کاربران، از هوش مصنوعی و بررسی محتوا توسط انسان بهره برده است و تلاش کرده است تا انتشار اینگونه اخبار را کاهش دهد. اهمیت این موضوع موجب شده‌است تا کشورها و شرکت‌های بزرگ حوزه فناوری، سرمایه‌گذاری‌های چشمگیری برای مقابله با انتشار اخبار جعلی شروع کنند و شرکت‌های نوپای بسیاری در این زمینه شکل بگیرد. یکی از محصولات این شرکت‌های نوپا در سطح جهان، ابزار لاجیکالی^۴ است که با کمک هوش مصنوعی و مدل‌های یادگیری ماشین به مبارزه با اخبار جعلی پرداخته است.

۳-۱ راه حل پیشنهادی

علی‌رغم فعالیت‌های انجام‌شده در این حوزه برای زبان‌های مختلف و به‌طور خاص برای زبان انگلیسی، زبان فارسی در این زمینه رشد چشم‌گیری نداشته است؛ بنابراین، تهیه یک ابزار جهت تشخیص اخبار جعلی فارسی از اهمیت به‌سزایی برخوردار است. در این پروژه، ما قصد داریم تا با استفاده از مفاهیم به‌روز یادگیری عمیق در حوزه پردازش زبان طبیعی و تنها با استفاده از ویژگی‌های مرتبط با متن اخبار، سامانه هوشمندی را به منظور تشخیص اخبار جعلی پیاده‌سازی کنیم. با توجه به عدم وجود دادگان مناسب و جامع به زبان فارسی، بخش اصلی از این پروژه شامل استخراج اخبار جعلی و برچسب‌گذاری خبرهای خزش‌شده خواهد بود. پس از آموزش و ارزیابی مدل به پیاده‌سازی یک کتابخانه در زبان پایتون و ابزار تحت وب می‌پردازیم تا با دریافت یک خبر بتواند در مورد میزان احتمال جعلی بودن آن خبر تصمیم بگیرد. این ابزار می‌تواند در آینده به‌عنوان یک افزونه به وبگاه‌های خبری و یا وبگاه‌های رصد اخبار و رسانه‌های اجتماعی اضافه شود و کاربر را در راستای اطمینان از صحت اخبار راهنمایی کند.

^۴Logically

فصل دوم

پیشینه تشخیص اخبار جعلی

۱-۲ کارهای انجام شده در زبان انگلیسی

با رشد چشم‌گیر استفاده از الگوریتم‌های یادگیری ماشین و شبکه‌های عصبی عمیق در پردازش زبان طبیعی، تحقیقات و طرح‌های بسیاری در زمینه تشخیص اخبار کذب انجام شده است. در ادامه به پژوهش‌های مرتبط انجام‌شده اشاره کرده و سپس در بخش بعد، مجموعه داده‌های موجود در زبان انگلیسی را مرور می‌کنیم.

ایساتراور و دیگران (۲۰۱۷) با استفاده از تحلیل و بررسی متن به وسیله روش مبتنی بر چندتایی^۱ و نمایش برداری مبتنی بر بسامد واژه-معکوس بسامد سند^۲ از الگوریتم‌های متداول یادگیری ماشین مانند ماشین بردار پشتیبان^۳، ماشین بردار پشتیبان خطی^۴، نزدیک‌ترین همسایه^۵، درخت تصمیم^۶، گرادیان کاهشی تصادفی^۷ و رگرسیون لجستیک^۸ برای تشخیص اخبار جعلی استفاده کرده‌اند. با ایجاد ترکیب‌های متفاوت از توالی واژه‌ها به صورت تکی و دوتایی و غیره، اطلاعات آماری هر ترکیب را با استفاده از روش مبتنی بر واژه برای اخبار جعلی مجموعه یادگیری شامل اخبار جعلی و اصیل بازنمایی کردند.

ژانگ و دیگران (۲۰۲۰) یک سامانه با عنوان «تشخیص‌دهنده جعل»^۹ را پیاده‌سازی کردند که از دو بخش اصلی تشکیل شده است: «یادگیری ویژگی بازنمایی»^{۱۰} و «استخراج ویژگی صریح»^{۱۱}. در کنار هر خبر، اطلاعات اجتماعی متنوعی مرتبط با آن خبر وجود دارد که مدل‌سازی در آن توسط یک واحد یادگیرنده ویژگی انجام می‌پذیرد. این قسمت، علاوه بر یادگیری ویژگی‌های آشکار مرتبط با عنوان یک خبر، مانند گروهی از واژه‌های استفاده‌شده، دارای بخش دیگری برای یادگیری ویژگی‌های نهان یک خبر، مانند ویژگی‌های مرتبط با موضوع و یا نویسنده آن خبر، است. علاوه بر این واحد یادگیرنده ویژگی، یک واحد پردازشی برای ایجاد ارتباط مناسب میان بردارهای مختلف متن خبر، عنوان خبر و موضوع خبر به نام «واحد انتشار دروازه‌ای»^{۱۲} نیز در معماری این مدل استفاده شده است. این واحد پردازشی، امکان استفاده از چندین ورودی متنوع را به صورت همزمان ایجاد می‌نماید.

¹N-gram

²Term Frequency - Inverse Document Frequency (tf-idf)

³Support Vector Machine (SVM)

⁴Linear Support Vector Machine (LSVM)

⁵Nearest Neighbor

⁶Decision Tree

⁷Stochastic Gradient Descent (SGD)

⁸Logistic Regression

⁹Fake Detector

¹⁰Representation Feature Learning

¹¹Explicit Feature Extraction

¹²Gated Diffusive Unit (GDU)

در پژوهشی که توسط **رمضانی و دیگران** (۲۰۱۹) انجام شده است، آنان با استفاده از «شبکه‌های عصبی بازگشتی»^{۱۳} و معرفی یک «تابع هزینه»^{۱۴} جدید، سعی کرده‌اند تا دنباله اخبار را به صورت یک پیوستار زمانی بررسی کنند و در هر مقطع زمانی با یک احتمال، برچسب خبر را مشخص کنند. با گذر زمان و دریافت اطلاعات بیشتر، دقت برچسب یک خبر دقیق‌تر خواهد شد. همچنین در تابع هزینه معرفی شده، پارامتر زمان نیز مؤثر است؛ به این دلیل که هدف، برچسب‌گذاری دقیق در سریع‌ترین زمان ممکن است تا از انتشار گسترده خبر جعلی به سرعت جلوگیری شود. این پژوهش بر روی دادگان جمع‌آوری شده از دو شبکه اجتماعی توییتر و «سینا ویبو»^{۱۵} آزمایش شده است.

خطر و دیگران (۲۰۱۹) از «خودکدگذار وردشی»^{۱۶} استفاده کردند تا اخبار جعلی را تشخیص دهند. برای این کار از مجموعه داده توییتر استفاده کردند که شامل یک بخش متن و یک عکس به همراه آن است. مدل آنها از سه بخش کدگذار^{۱۷} متن و تصویر، کدگشای^{۱۸} متن و تصویر و تشخیص‌دهنده خبر جعلی تشکیل شده است. بخش کدگذار، خود شامل یک شبکه عصبی بازگشتی برای استخراج ویژگی‌های متنی و یک «شبکه عصبی پیچشی»^{۱۹} برای استخراج ویژگی‌های مرتبط با عکس است. سپس، ویژگی‌های نهان استخراج شده از دو منبع عکس و متن، «بردار نهان»^{۲۰} اصلی را شکل می‌دهد. مجموعه کدگذار و کدگشا در فرایند یادگیری، ویژگی‌های نهان اخبار جعلی را با توجه به متن و عکس‌ها یاد می‌گیرد؛ و پس از آن با استفاده از این ویژگی‌های نهان توسط بخش تشخیص‌دهنده، جعلی بودن آن را تشخیص می‌دهد. **یانگ و دیگران** (۲۰۱۹) برای تشخیص اخبار جعلی از روش یادگیری «بدون نظارت»^{۲۱} استفاده کرده‌اند که در آن با استفاده از یک مدل احتمالاتی گرافیکی، صحت اخبار و اعتبار کاربران مدل شده است. آنها با استفاده از ویژگی‌های نهان به دست آمده برای کاربر و خبر مورد نظر توسط روش نمونه‌برداری گیبز^{۲۲} کار دسته‌بندی خبر جعلی را انجام می‌دهند. این روش بر روی دو مجموعه داده لیار^{۲۳} (ونگ، ۲۰۱۷) و بازفیدنیز^{۲۴} آزمایش شده است.

¹³Recurrent Neural Network (RNN)

¹⁴Loss function

¹⁵Sina Weibo

¹⁶Variational Autoencoder (VA)

¹⁷Encoder

¹⁸Decoder

¹⁹Convolutional Neural Network (CNN)

²⁰Latent Vector

²¹Unsupervised learning

²²Gibbs sampling

²³Liar

²⁴<https://github.com/BuzzFeedNews/2016-10-facebook-fact-check>

لیو و دیگران (۲۰۱۹) یک مدل دومرحله‌ای براساس مدل برت^{۲۵} ارائه کرده‌اند که با استفاده از تعبیه اطلاعات فراداده مانند نام گوینده، شغل و غیره به همراه متن اصلی، خبر جعلی تشخیص داده می‌شود. در این مدل، به جای استفاده از اولین بردار خروجی مدل برت که نماینده تمام جمله است، از تمامی بردارهای خروجی برای کلمات استفاده می‌شود تا با به کارگیری روش توجه، برای هر کدام یک وزن محاسبه شود. مرحله اول این مدل، به طبقه‌بندی کلی اخبار جعلی و درست می‌پردازد و با استفاده از بردارهای خروجی این مرحله، در مرحله دوم با دسته‌بندی جزئی‌تر، اخبار دسته‌بندی می‌شود. این مدل بر روی دادگان لیار به دو صورت دسته‌بندی دوتایی و دسته‌بندی چندتایی آزمایش شده است.

ژوا و دیگران (۲۰۱۹) با استفاده از پیکره اخبار از خبرگزاری سی.ان.ان^{۲۶} و «دیلی میل»^{۲۷} مدل برت را آموزش داده‌اند و از دادگان مرحله اول مسابقه اخبار جعلی^{۲۸} (اف.ان.سی.وان^{۲۹}) برای «تنظیم دقیق»^{۳۰} مدل استفاده کرده‌اند. در این مدل از تابع «آنتروپی متقاطع وزن دار»^{۳۱} برای محاسبه خطا استفاده شده است تا مشکل متوازن نبودن داده‌ها حل شود. در این پژوهش، دو مدل برای تشخیص خبر جعلی ارائه شده است که در یک مدل از مدل آماده برت به همراه تابع آنتروپی متقاطع وزن دار استفاده شده و در مدل دیگر با استفاده از پیکره اخبار، مدل برت را آموزش داده‌اند.

گلدانی و دیگران (۲۰۲۰) دو شبکه کپسولی^{۳۲} متفاوت برای تشخیص اخبار جعلی با طول‌های متفاوت ارائه کرده‌اند. به منظور دسته‌بندی اخبار جعلی متوسط و یا طولانی از ۴ شبکه موازی و برای اخبار کوتاه از ۲ شبکه موازی با اندازه صافی^{۳۳} متفاوت به منظور استخراج ویژگی‌های سطح بالا در متن استفاده کردند. پس از استخراج ویژگی‌های سطح بالا، هر شبکه به یک لایه چگال^{۳۴} متصل شده است و در نهایت با استفاده از یک لایه میانگین ادغام^{۳۵} بردار نهایی و احتمال جعلی بودن خبر را مشخص می‌کنند.

کالیار و دیگران (۲۰۲۰) یک شبکه عصبی عمیق پیچشی را با نام اف.ان.دی.نت^{۳۶} برای تشخیص

²⁵BERT

²⁶CNN

²⁷Daily Mail

²⁸fakenewschallenge.org

²⁹FNC-1

³⁰Fine tune

³¹Weighted Cross-entropy

³²Capsules Network

³³Filter

³⁴Dense

³⁵Average Pooling

³⁶FNDNET

اخبار جعلی ارائه کردند. در این مدل برای بازنمایی اخبار از مدل بازنمایی «بردار سراسری»^{۳۷} ارائه شده توسط پنینگتون و دیگران (۲۰۱۴)، استفاده کردند. معماری این مدل شامل لایه‌های پیچشی که به صورت آبشاری قرار گرفته است، می‌شود تا ویژگی‌های مناسبی را برای اخبار تولید کند. درنهایت با استفاده از لایه‌های چگال، احتمال تعلق هر خبر به دسته جعلی یا اصیل مشخص می‌شود. به منظور ارزیابی مدل ارائه شده، آنها از دادگان وبگاه کگل^{۳۸} که مربوط به انتخابات سال ۲۰۱۶ آمریکا است استفاده کردند.

۲-۲ دادگان موجود در انگلیسی و آمار دادگان

در زبان انگلیسی دادگان زیادی برای تشخیص اخبار جعلی معرفی شده که از رسانه‌های اجتماعی یا وبگاه‌های خبری استخراج شده است. جدول ۱-۲ به صورت مختصر فهرستی از دادگان موجود برای زبان انگلیسی را ارائه می‌دهد. در این جدول، ستون «داده و فراداده‌ها» نمایانگر اطلاعات و فراداده‌های اصلی موجود در هر یک از این دادگان است. عبارات TS، UI، T به ترتیب به معنای اطلاعات مرتبط با انتشار خبر^{۳۹}، برچسب زمانی^{۴۰}، ویژگی‌های مرتبط با اطلاعات کاربر^{۴۱} و متن خبر^{۴۲} است.

با توجه به گسترش پژوهش فعالیت‌های تشخیص اخبار جعلی در سال‌های اخیر و تعداد زیاد دادگان این حوزه، در این پژوهش مجموعه دادگان لیار و آی.اس.ا.تی.^{۴۳} به عنوان نمونه‌های مطرح دادگان اخبار جعلی مورد مطالعه قرار گرفته است که ویژگی‌های این دو مجموعه داده اصلی به همراه دو مجموعه پرکاربرد دیگر در این حوزه در ادامه توضیح داده می‌شود.

۱-۲-۲ دادگان لیار

ونگ (۲۰۱۷) دادگانی با نام «لیار» برای تشخیص خبر کذب به زبان انگلیسی معرفی کردند. در این مجموعه هر خبر شامل ۱۴ فراداده است، از جمله شماره خبر، متن خبر، عنوان خبر، گوینده خبر، شغل گوینده، شهر، حزب و اطلاعات زمینه‌ای و ۵ ویژگی مربوط به اعتبار گزاره‌های قبلی. هر خبر نیز دارای یک برچسب از مجموعه ۶ تایی است که این برچسب‌ها عبارتند از: درست، نیمه درست، عمدتاً درست،

³⁷Global Vectors (GloVe)

³⁸Kaggle

³⁹Propagation Information

⁴⁰Timestamp

⁴¹User Information

⁴²Text

⁴³ISOT

جدول ۲-۱: آمار و اطلاعات مرتبط با دادگان شایعات در شبکه‌های اجتماعی (لی و دیگران، ۲۰۱۹)

مرجع	منبع	داده و فراداده‌ها	تعداد خبر	عنوان مجموعه داده
زوبیگا و دیگران (۲۰۱۶)	توییتر	PI/TS/UI/T	۳۳۰	PHEME-R
کوکینا و دیگران (۲۰۱۸)	توییتر	PI/TS/UI/T	۶, ۴۲۵	PHEME
ما و دیگران (۲۰۱۶)	توییتر	TS/UI/T	۹۹۲	Ma-Twitter
ما و دیگران (۲۰۱۶)	ویبو	TS/UI/T	۴۶۶۴	Ma-Weibo
ما و دیگران (۲۰۱۷)	توییتر	PI/TS/UI/T	۱۴, ۹۰	Twitter15
ما و دیگران (۲۰۱۷)	توییتر	PI/TS/UI/T	۸۱۸	Twitter16
دادگان فعالیت ۷ مسابقه «ارزیابی معنایی» سال ۲۰۱۹	توییتر، ردیت	PI/TS/UI/T	۳۲۵	SemEval19
شایعات کگل براساس سایت اسناپس	توییتر، فیسبوک	T	۱۶, ۹۰۰	Kaggle Snopes
تاجینی و دیگران (۲۰۱۷)	فیسبوک	TS/UI/T	۱۵, ۵۰۰	Facebook Hoax
شایعات کگل براساس سایت پولیتی‌فکت	توییتر	PI/TS/UI/T	۲, ۹۲۳	Kaggle PolitiFact
شو و دیگران (۲۰۲۰)	توییتر	PI/TS/UI/T	۲۳, ۱۹۶	FakeNewsNet

جعلی، دروغ و به‌سختی درست. تعداد کل اخبار ۱۲, ۷۸۸ خبر می‌باشد که در ۳ دسته داده آموزشی^{۴۴}، داده اعتبارسنجی^{۴۵}، داده ارزیابی^{۴۶} تقسیم شده است. همچنین باتوجه‌به این که «حزب» خبر یکی از فراداده‌های موجود در این مجموعه دادگان است، فراوانی ۳ «حزب» دموکرات، جمهوری خواه و بی‌طرف به‌ترتیب ۴, ۱۵۰، ۵, ۶۸۷ و ۲, ۱۸۵ خبر است. جدول ۲-۲ آمار کلی این مجموعه داده را نمایش می‌دهد.

جدول ۲-۲: آمار و اطلاعات مربوط به دادگان لیار

آماره	مقدار
اندازه مجموعه آموزش	۱۰, ۲۶۹
اندازی مجموعه اعتبارسنجی	۱, ۲۸۴
اندازه مجموعه آزمون	۱, ۲۸۳
میانگین طول خبر	۱۷/۹

^{۴۴}Train set

^{۴۵}Validation set

^{۴۶}Test set

۲-۲-۲ دادگان آی.اس.ا.تی

مجموعه داده «آی.اس.ا.تی.» شامل اخبار اصیل و جعلی است که توسط ایساتراور و دیگران (۲۰۱۷) تهیه شده است. ۲۱,۴۱۷ خبر اصیل آن از وبگاه خبری رویترز و ۲۳,۴۸۱ خبر جعلی از وبگاه‌های پولیتی‌فکت و ویکی‌پدیا جمع‌آوری و برچسب‌گذاری شده است. در این مجموعه، هر خبر علاوه بر عنوان خبر و متن خبر، دارای ۲ فراداده تاریخ خبر و موضوع خبر نیز می‌باشد. جدول ۲-۳ آمار اخبار موجود در این مجموعه داده را به تفکیک اخبار اصیل و جعلی نمایش می‌دهد.

جدول ۲-۳: آمار و اطلاعات مربوط به دادگان آی.اس.ا.تی

خبر	تعداد خبر	نوع	تعداد خبر
اخبار اصیل	۲۱,۴۱۷	اخبار جهان ^{۴۷} اخبار سیاسی ^{۴۸}	۱۰,۱۴۵ ۱۱,۲۷۲
اخبار جعلی	۲۳,۴۸۱	اخبار دولت خاورمیانه اخبار آمریکا اخبار چپ سیاسی دیگر	۱,۵۷۰ ۷۷۸ ۷۸۳ ۴,۴۵۹ ۶,۸۴۱ ۹,۰۵۰

۳-۲-۲ دادگان بازفید

این دادگان شامل اخبار مرتبط با انتخابات آمریکا است که در فیسبوک^{۴۹} منتشر شده بود.^{۵۰} این مجموعه شامل ۲,۲۸۳ عنوان پست فیسبوک است که هر خبر از ۱۲ فراداده مانند صفحه خبر، تاریخ، نوع محتوا، برچسب و اطلاعات مرتبط با میزان تأثیرگذاری بر کاربران مانند تعداد دفعات به اشتراک گذاشتن و تعداد نظرات و غیره تشکیل شده است. جدول ۲-۴ آمار اخبار موجود در این مجموعه داده را نمایش می‌دهد.

۴-۲-۲ توئیتر ۱۵

ما و دیگران (۲۰۱۷) دادگانی از شبکه اجتماعی توئیتر منتشر کردند که دارای ۴ برچسب درست، نادرست، نامعتبر و غیرشایعه است. این پیام‌ها براساس سایت‌های اعتبارسنجی برچسب‌گذاری شده‌اند. در این

⁵⁶World-News

⁵⁷Politics-News

⁴⁹Facebook

⁵⁰<https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis>

جدول ۲-۴: آمار و اطلاعات مربوط به دادگان بازفید

حزب	بدون محتوای واقعی	عمدتاً درست	ترکیب درست و نادرست	عمدتاً نادرست	مجموع
چپ	۱۱۶	۲۶۵	۶۸	۲۲	۴۷۱
جریان اصلی	۵۲	۱,۰۸۵	۸	۰	۱,۱۴۵
راست	۹۶	۳۱۹	۱۶۹	۸۲	۶۶۶

دادگان علاوه بر اطلاعات متن توییت‌ها، اطلاعات ۲۷۶ هزار کاربر و شبکه ارتباط آن‌ها و همچنین نحوه بازنشر هر توییت مشخص شده است. جزئیات توزیع این برچسب‌ها در جدول ۲-۵ آورده شده است.

جدول ۲-۵: آمار و اطلاعات مربوط به دادگان توییت‌ر ۱۵

برچسب	تعداد
غیرشایعه	۳۷۴
نادرست	۳۷۰
درست	۳۷۲
نامعتبر	۳۷۴

۲-۳ دادگان موجود در سایر زبان‌ها

با توجه به اهمیت موضوع تشخیص اخبار جعلی، در زبان‌هایی با منابع کم نیز پژوهش‌های زیادی در سال‌های اخیر بر روی اخبار جعلی انجام شده است و مجموعه داده‌هایی برای اخبار جعلی در این زبان‌ها هم جمع‌آوری شده است. جدول ۲-۶ به صورت مختصر لیستی از دادگان موجود برای اخبار جعلی برای سایر زبان‌ها را نمایش می‌دهد.

جدول ۲-۶: لیست دادگان موجود در سایر زبان‌ها

سال ارائه	داده	زبان ارائه
۲۰۱۹	(وگل و ژیانگ، ۲۰۱۹)	آلمانی
	(ریس و پالفاکس، ۲۰۱۹)	اسپانیایی
	(ویکاریو و دیگران، ۲۰۱۹)	ایتالیایی
	(لیو و دیگران، ۲۰۱۹)	فرانسوی
	(الخیر و دیگران، ۲۰۱۹)	عربی
۲۰۱۸	(لوژینکوف و دیگران، ۲۰۱۸)	روسی

۴-۲ کارهای انجام شده و دادگان موجود در زبان فارسی

باتوجه به اینکه تمرکز اصلی این پروژه بر روی تهیه دادگان فارسی و استانداردسازی داده به دست آمده برای کاربرد در سیستم تشخیص اخبار جعلی فارسی است، در این فصل مروری بر کارهای انجام شده در حوزه تشخیص اخبار جعلی فارسی ارائه می گردد و دادگان موجود در این حوزه معرفی می شود. در زبان فارسی پژوهشگرانی به صورت محدود در این حوزه فعالیت داشته اند و مجموعه دادگان انگشت شماری موجود است که در ادامه به معرفی آنها می پردازیم.

زمانی و دیگران (۲۰۱۷) در دانشگاه تهران با تمرکز بر روی شبکه های اجتماعی به تحلیل و بررسی شایعات در توییتر پرداختند. در این پژوهش، با بهره گیری از اطلاعات مرتبط با شبکه ارتباطی کاربران و ویژگی های مرتبط با هر توییت به دسته بندی شایعات با استفاده از مدل های سنتی یادگیری ماشین پرداختند. علاوه بر تحلیل ویژگی های پراهمیت برای تشخیص شایعات در شبکه های اجتماعی، یک مجموعه داده در حوزه تشخیص شایعات توییتر ارائه دادند که شامل ۷۸۳ توییت جعلی و ۷۸۳ توییت اصیل است. به منظور استخراج شایعات از دو وبسایت ایرانی گمانه^{۵۱} و شایعات^{۵۲} استفاده کردند. در این پژوهش با استفاده از مدل های یادگیری ماشین مانند بیز ساده^{۵۳}، ماشین بردار پشتیبان^{۵۴}، ک-نزدیک ترین همسایه^{۵۵} و درخت تصمیم^{۵۶} به دسته بندی این توییت ها براساس ویژگی های مرتبط با گراف کاربران و یا آماره های خود توییت ها پرداختند.

محمودآباد و دیگران (۲۰۱۸) در پژوهشی مجموعه دادگانی را با استفاده از بررسی پست های ۱۱,۹۸۱ کاربر فارسی زبان در شبکه اجتماعی توییتر استخراج کردند. این مجموعه داده شامل ۷۰۴, ۵۹۳, ۳ توییت فارسی است که عمدتاً در مورد زلزله کرمانشاه بوده و با استفاده از سایت شایعات، برچسب گذاری کردند. مجموعه داده از نوامبر تا دسامبر ۲۰۱۷ جمع آوری شده و شامل ۴,۳۴۵ توییت شایعه است. به دلیل نامتوازن بودن اخبار شایعه و واقعی با استفاده از الگوریتم بیش نمونه برداری SMOTE، آنها داده ها را متوازن کرده و سپس هر توییت را با برداری شامل اطلاعات زمینه ای، اطلاعات ساختاری و اطلاعات جمعیتی بازنمایی کردند. در نهایت داده ها با استفاده از الگوریتم های متداول یادگیری ماشین همچون بیز ساده، ماشین بردار پشتیبان، درخت تصمیم و جنگل تصادفی^{۵۷} دسته بندی شدند.

⁵¹<https://gomaneh.net/>

⁵²<http://shayeaat.ir/>

⁵³Naive Bayes

⁵⁴Support Vector Machine

⁵⁵K Nearest Neighbor

⁵⁶Decision Tree

⁵⁷Random Forest

زهران و دیگران (۲۰۱۹) با دیدگاه مبتنی بر روش دسته‌بندی موضع^{۵۸} به تشخیص اخبار جعلی پرداخته‌اند. برای این منظور ۵۳۴ ادعا به همراه ۲,۱۲۴ متن خبر در مورد آن ادعاها جمع‌آوری شد و سپس با استفاده از معماری یادگیری عمیق حافظه کوتاه-مدت طولانی پشته‌ای^{۵۹} به دسته‌بندی رابطه میان ادعاها و متن‌های مرتبط با آن در ۴ کلاس موافق^{۶۰}، ناموافق^{۶۱}، نامرتبط^{۶۲}، بحث‌شده^{۶۳} پرداخته‌اند. ساختار این پژوهش مشابه پژوهش‌های انجام‌شده در زمینه تشخیص موضع^{۶۴} در زبان انگلیسی بوده و به دسته‌بندی اخبار براساس جعلی یا اصیل بودن پرداخته است.

جهان‌بخش و دیگران (۲۰۲۰) مجموعه دادگان دیگری با تمرکز برروی اخبار منتشر شده در شبکه اجتماعی تلگرام منتشر کردند. این مجموعه داده شامل ۸۸۲ شایعه و ۸۸۲ پست اصیل است که از کانال خبرگزاری‌هایی مانند: خبرگزاری فارس، دانشجویان ایران (ایسنا)، تسنیم، مهر و خبرگزاری جمهوری اسلامی (ایرنا) و همچنین سه وبسایت گمانه، ویکی‌هواکس^{۶۵} و شایعات خزش شده‌اند. در پژوهش انجام شده توسط جهان‌بخش و دیگران (۲۰۲۰) از مدل ارائه‌شده توسط آلپورت و پوستمن (۱۹۴۷) الهام گرفته شده‌است. در این مدل یک رابطه به‌عنوان قدرت شایعه معرفی شده است که ارتباط مستقیمی با اهمیت خبر و ابهام آن دارد. با توجه به این مقاله، جهان‌بخش و دیگران (۲۰۲۰) به پیاده‌سازی مدلی برای محاسبه این ضریب، تحت عنوان قدرت انتشار شایعه در زبان فارسی پرداختند. برای محاسبه این ضریب از ویژگی‌های مانند احساس خبر، اهمیت خبر و ابهام خبر استفاده شده است.

جدول ۲-۷ خلاصه‌ای از دادگان موجود زبان فارسی را نمایش می‌دهد.

جدول ۲-۷: آمار دادگان موجود فارسی

مقاله	تعداد داده جعلی	تعداد کل داده	منبع اخبار جعلی	منبع اخبار
جهان‌بخش و دیگران (۲۰۲۰)	۸۸۲	۱,۷۶۴	ویکی‌هواکس / گمانه / شایعات	تلگرام
زهران و دیگران (۲۰۱۹)	۶۰۰	۲,۱۲۴	شایعات / فیک‌نیوز	وبسایت‌ها
محمودآباد و دیگران (۲۰۱۸)	۴,۳۴۵	۳,۵۹۸,۰۴۹	شایعات	توییتر
زمانی و دیگران (۲۰۱۷)	۷۸۳	۱,۵۶۶	شایعات / گمانه	توییتر

⁵⁸ Stance Classification

⁵⁹ Stack LSTM

⁶⁰ Agree

⁶¹ Disagree

⁶² Unrelated

⁶³ Discussed

⁶⁴ Stance Detection

⁶⁵ <https://wikihoax.org/>

فصل سوم

تهیه دادگان فارسی «تاج»

۱-۳ مقدمه‌ای بر جمع‌آوری داده‌ها

در زبان انگلیسی، مجموعه دادگان متنوعی برای اخبار جعلی تهیه شده است که برخی از آن‌ها مربوط به اخبار خبرگزاری‌ها و برخی دیگر در مورد اخبار منتشرشده در شبکه‌های اجتماعی است. با توجه به کمبود دادگان فارسی در حوزه تشخیص اخبار جعلی و ذات روش‌های مبتنی بر یادگیری عمیق که کاملاً وابسته به دادگان یادگیری است، به‌منظور پیاده‌سازی یک سامانه تشخیص اخبار جعلی کارآمد نیاز است هدف ما در گام اول به تهیه یک مجموعه داده جامع در این حوزه متمرکز گردد. استخراج اخبار جعلی بدون آگاهی از محتوای این اخبار چالش‌های فراوانی دارد. به‌عنوان مثال با قراردادن متن یک خبر در مقابل شخصی که مسئولیت برچسب‌زنی را دارد نمی‌توان انتظار داشت در مورد جعلی بودن آن خبر نظر دهد. در نتیجه باید روشی برای استخراج اخبار جعلی ارائه دهیم که از یکسو طیف وسیعی از موضوعات را در بر گیرد و از سوی دیگر قضاوت در مورد جعلی بودن اخبار در فرایند نشانه‌گذاری داده توسط برچسب‌زن امکان‌پذیر باشد. علاوه بر این، منبع رسمی‌ای برای این اخبار وجود ندارد که بتوان به‌صورت مستقیم آن‌ها را استخراج کرد. برای همین منظور، در این پروژه ما یک روش مبتنی بر ۵ مرحله را ارائه دادیم تا علاوه بر جمع‌آوری حجم قابل قبولی داده، این چالش‌ها را حل نماییم.

۲-۳ مراحل جمع‌آوری داده‌ها

باتوجه به چالش‌های ذکرشده برای جمع‌آوری اخبار جعلی، در این بخش یک روش ابداعی ۵ مرحله‌ای برای استخراج اخبار جعلی را ارائه می‌دهیم:

۱. باتوجه به نبود یک منبع اختصاصی و مطمئن برای ذخیره اخبار جعلی منتشرشده در فضای مجازی، امکان دسترسی مستقیم به این اخبار وجود ندارد. بنابراین ما از یک فرض عموماً درست در اخبار فارسی استفاده کردیم تا بتوانیم اخبار جعلی را به‌صورت عمومی و در گستره وسیعی از موضوعات استخراج کنیم. در اخبار فارسی عمدتاً اخبار جعلی پس از مدتی توسط نهادهای رسمی و یا وبسایت‌های خبری تکذیب می‌شوند. بنابراین برای هر اخبار جعلی انتظار می‌رود که یک تکذیبیه‌ای وجود داشته باشد. بر همین اساس از رویکرد معکوس برای رسیدن به اخبار جعلی استفاده نموده‌ایم. در گام اول با جستجو یک پرس‌وجو کلی مانند "تکذیب خبر" با استفاده از رابط کاربری برنامه گوگل، فهرستی از اخبار تکذیب‌شده را استخراج کردیم.

۲. با داشتن عنوان اخبار تکذیب‌شده که در مرحله قبل استخراج شده است می‌توانیم با حذف کلمات

خاص مانند "تکذیب" و "شایعه" به عنوانی برسیم که از لحاظ متنی به عنوان خبر جعلی اولیه شباهت دارد.

۳. پس از استخراج عنوان احتمالی خبر کذب، با استفاده از موتور جست و جوی گوگل تلاش می کنیم تا اصل خبر جعلی را در وبسایتها بیابیم. در این مرحله برای هر خبر تکذیب شده فهرستی از اخبار جعلی کاندید وجود خواهد داشت.

۴. پس از استخراج اخبار جعلی کاندید به ازای هر خبر تکذیب شده، تمامی اخبار به صورت دستی توسط عوامل انسانی برچسب گذاری می شود.

۵. پس از مشخص شدن تعداد اخبار جعلی در گام قبلی، به همان تعداد خبر اصیل به صورت تصادفی از خبرگزاری های متفاوت استخراج می کنیم و در صورت تائید به مجموعه دادگان اضافه می شوند.

۳-۳ نشانه گذاری داده

فرایند برچسب زنی اخبار توسط عامل انسانی انجام شده است. در هنگام نشانه گذاری داده، به هر خبر کاندیدای جعلی یکی از برچسب های زیر تعلق می گیرد: جعلی، جعلی چندرسانه ای، پرسشی، اطلاعات بیشتر، تکذیبی و نامربوط. به منظور دسته بندی اخبار جعلی در این پروژه تنها اخباری که برچسب جعلی داشته اند در مجموعه دادگان قرار داده شده اند، اما در آینده می توان از برچسب های دیگر نیز استفاده کرد. توضیح این برچسب ها در ادامه آمده است:

• جعلی

اخباری که پس از مدتی توسط نهادهای رسمی و یا افراد تکذیب شده است. به عنوان مثال "سریال نون خ با کمک ۳ میلیارد تومانی استانداری در کرمانشاه ساخته می شود" یک عنوان جعلی است که توسط وبسایت های خبری رسمی تکذیب شده است.

• جعلی چندرسانه ای

اخباری است که اطلاعات نادرستی را که عمدتاً شامل تصاویر و یا فیلم هستند دربردارد. این اخبار نیز از لحاظ محتوایی اخباری هستند که پس از مدتی توسط نهادهای رسمی و یا افراد تکذیب شده است. به عنوان مثال: "(ویدئو) جشن و پایکوبی پرستاران بیمارستان رازی رشت - برخط نیوز" نمونه ای از این دسته اخبار است.

• پرسشی

اخبار پرسشی عموماً اخباری هستند که در عنوان آن‌ها یک پرسشی مرتبط با خبر تکذیب شده اولیه مطرح شده است، مثال: «ماجرای کمک مالی استانداری کرمانشاه برای ساخت «نون.خ» چیست؟ - شهرخبر». این اخبار عموماً محتوای با ارزشی ندارند و در دادگان لحاظ نمی‌شوند.

• اطلاعات بیشتر

اخباری که در مورد خبر تکذیب‌شده اولیه جزئیات بیشتری را مطرح می‌کنند برچسب اطلاعات بیشتر را دریافت می‌کنند. به عنوان مثال عنوان «تازه‌ترین جزئیات از فصل دوم سریال «نون.خ»» یک نمونه از اخباری است که به جزئیات بیشتری در مورد یک موضوع تکذیب شده می‌پردازند.

• تکذیبی

در میان کاندیدهای جعلی، به صورت محدود اخبار تکذیب‌شده دیگری نیز یافت شده است که اگرچه دارای کلمات «تکذیب»، «شایعه» نیستند اما همچنان محتوای تکذیبی را در بر دارند. این اخبار با برچسب تکذیبی مشخص شده اند. مثال: «عدم صحت خبر ملاقات ابتکار با زنان شاغل در سفارتخانه های خارجی»

• نامربوط

در نهایت با توجه به آنکه از یک موتور جست‌وجوگر عمومی استفاده کرده‌ایم بسیاری از اخبار کاندید ارتباط مستقیمی به عنوان جست‌جو شده تکذیبی ندارند. این اخبار با برچسب نامربوط مشخص شده‌اند. مثال: «تمجید بهروز شعبی از طراحی و اجرای شوخی‌های «نون.خ ۲»»

تمامی مثال‌های مطرح شده در توضیحات این بخش مربوط به سه متن تکذیبی می‌باشد: (۱) تکذیبی مربوط به سریال نون.خ، (۲) تکذیبی مربوط به جشن و پایکوبی پرستاران و (۳) تکذیبی مربوط به دیدار خانم ابتکار. متن تکذیبی این سه موضوع به شرح زیر می‌باشد:

۱. تکذیبی مربوط به سریال نون.خ:

به گزارش تابناک به نقل از خبرآنلاین، مهدی فرجی تهیه‌کننده سریال «نون.خ» درباره انتشار اخباری پیرامون کمک‌های استانداری کرمانشاه به این سریال گفت: «از ابتدای ساخت این سریال، گفتگوهای با مسؤولان استان کرمانشاه و استاندار داشتیم و حتی استاندار در مراسم آغاز تصویربرداری سریال نیز حضور پیدا کرد.»

فرجی در ادامه گفت: «استانداری کرمانشاه فقط در بخش لجستیک و پشتیبانی کمک‌هایی به

ما کردند که تشکر و قدردانی می‌کنم، اما هیچ‌گونه پشتیبانی مالی برای ساخت این سریال انجام ندادند.»

وی ادامه داد: «تمامی اخبار درباره کمک مالی و یا صرف بودجه‌ای از طرف سازمان‌های خارج از صداوسیما برای ساخت این سریال صحت ندارد و هیچ نهاد و سازمانی در استان کرمانشاه برای ساخت سریال «نون. خ» به سازندگان این سریال کمک مالی نکرده است.»

سریال «نون. خ» به کارگردانی سعید آقاخانی و تهیه‌کنندگی مهدی فرجی تولید شده و این شب‌ها ساعت ۲۲ از شبکه یک سیما پخش می‌شود.

۲. تکذیبیه مربوط به جشن و پایکوبی پرستاران:

به گزارش همشهری آنلاین به نقل از ایسنا، دکتر محمدرضا نقی‌پور در جمع خبرنگاران گفت: تصاویری که از برگزاری جشن پایان کرونا در بیمارستان رازی رشت منتشر شده کذب است.

سخنگوی دانشگاه علوم پزشکی گیلان توضیح داد: جشنی از سوی یک گروه هنری خصوصی برای قدردانی از کادر درمانی استان در محوطه بیمارستان رازی برگزار شده بود و به معنای جشن پایان کرونا در گیلان نیست.

نقی‌پور ادامه داد: تاکید این جشن به هیچ وجه برای اعلام موفقیت در مهار کرونا در استان نیست به همین دلیل مردم باید کرونا را جدی بگیرند و از خانه خارج نشوند.

افزود: کرونا در کشور هنوز وارد مرحله کنترل نشده است، به همین دلیل مردم باید خود مراقبتی و دیگر مراقبتی را جدی بگیرند.

۳. تکذیبیه مربوط به دیدار خانم ابتکار:

دیدارنیوز - در این اطلاعیه آمده است: این خبر که به نظر می‌رسد منشاء آن ترورهای منافقین در آلبانی باشد، از اساس دروغ بوده و ابتکار هیچ‌گونه دیداری با بانوان شاغل در سفارتخانه‌های خارجی در هیچ محلی نداشته است. همچنین گفت و شنوهای مورد اشاره نیز زاده توهّمات معمول منتشرکنندگان اینگونه اخبار جعلی است.

این اطلاعیه می‌افزاید: دیدارهای خارجی معاون رئیس‌جمهور در محل معاونت با حضور نماینده وزارت امور خارجه و فقط در سطح سفرای کشورها صورت می‌گیرد. همچنین لازم به ذکر است که نامه به رهبران سیاسی زن جهان فقط به صورت مکاتبه و از طریق وزارت امور خارجه بوده و برای ارسال این نامه هیچ‌گونه ملاقات خارجی صورت نگرفته است. البته جای تعجب از رسانه‌های

داخلی نیز هست که اینچنین فریب منافقین را می‌خورند و در میدان آنها بازی می‌کنند.

در طول فرایند برچسب‌زنی اخبار، چالش‌های بسیاری وجود داشت. با ظهور و گسترش شبکه‌های اجتماعی و سهولت انتشار اخبار در این رسانه‌ها نرخ تولید اخبار جعلی در این شبکه‌ها بسیار بیشتر از وب‌سایت‌های خبری است. باتوجه‌به این موضوع، ما در طول فرایند استخراج اخبار جعلی با اخبار تکذیبی مواجه شدیم که ریشه اولیه خبر جعلی آن در هیچ وب‌سایت خبری فارسی یافت نشده است. علاوه بر این، بخش اندکی از اخبار جعلی فارسی، پس از مدتی توسط رسانه‌ها دوباره تأیید شده و در واقع این اخبار نمی‌تواند دیگر جعلی به حساب بیاید.

برای دسته اول بالاترین احتمال، وجود اخبار جعلی اولیه در شبکه‌های اجتماعی یا پیام‌رسان‌ها هست که توسط موتور جستجو در دسترس نمی‌باشد. برای مثال، ممکن است یک خبر جعلی در تلگرام منتشر شود و سپس تکذیبیه آن در سایت خبری منتشر شود. علی‌رغم دسترسی به تکذیبیه، اصل خبر با استفاده از موتور جستجو در دسترسی نمی‌باشد و نمی‌توانیم آن خبر را در دادگان خود داشته باشیم. این امر انگیزه مهمی برای گسترش این پروژه و تهیه دادگان بیشتر در بستر شبکه‌های اجتماعی و پیام‌رسان‌ها می‌باشد که جزء کارهای آتی این پژوهش لحاظ می‌گردد.

۴-۳ آمار دادگان

مجموعه دادگان تاج شامل ۱,۸۶۰ خبر جعلی و ۱,۸۶۰ خبر اصیل درمورد طیف وسیعی از موضوعات خبری است. اخبار اصیل از ۵ وب‌سایت خبرگزاری معتبر فارسی از جمله ایرنا^۱، ایسنا^۲، همشهری آنلاین^۳، فارس^۴ و مهر^۵ استخراج شده‌است. همچنین اخبار جعلی عمدتاً از تعداد زیادی وب‌سایت‌های غیررسمی خزش شده‌است که اخبار را از منابع رسمی منتشر نمی‌کند؛ به همین علت، منبع برخی از این اخبار شبکه‌های اجتماعی یا منابع غیرقابل اعتماد است. جدول ۳-۱ آمار مجموعه داده استخراج‌شده برای سامانه «تشخیص اخبار جعلی» که به اختصار «تاج» می‌نامیم را نشان می‌دهد. این دادگان در مجموع شامل بیش از ۱,۱۲۷,۰۰۰ واژه است که شامل اخبار منتشرشده در بازه زمانی دی ۱۳۸۸ تا مهر ۱۳۹۹ می‌باشد.

^۱<https://www.irna.ir>

^۲<https://www.isna.ir>

^۳<https://www.hamshahrionline.ir>

^۴<https://www.farsnews.ir>

^۵<https://www.mehrnews.com>

جدول ۳-۱: آمار دادگان «تاج»

آماره	مقدار
تعداد اخبار جعلی	۱,۸۶۰
تعداد اخبار اصیل	۱,۸۶۰
تعداد وبسایت متمایز	۵۹۲
میانگین طول اخبار	۳۰۳
کمترین طول خبر	۹
بیشترین طول خبر	۷,۱۷۲

فصل چهارم

معماری مدل‌های ارائه شده

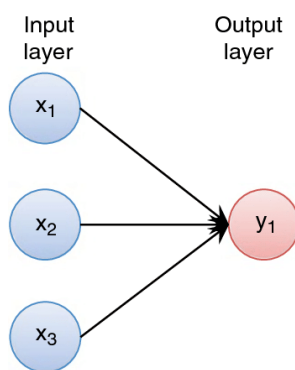
۱-۴ مقدمه

تاکنون در زبان فارسی، پژوهشگران عمدتاً از مدل‌های مرسوم یادگیری ماشین برای تشخیص اخبار جعلی استفاده کرده‌اند. یکی از اصلی‌ترین دلایل استفاده از این روش‌ها کافی نبودن داده‌های یادگیری برای آموزش مدل‌های به روز و پیچیده است. در این پروژه، پس از جمع‌آوری یک مجموعه داده از اخبار جعلی منتشرشده در پایگاه‌های خبری فارسی، استفاده از مدل‌های عمیق امکان‌پذیر می‌شود. به‌طور کلی پردازش داده متنی در سیستم تشخیص خبر جعلی از دو بخش اصلی بازنمایی متن و دسته‌بندی متن تشکیل شده است که در ادامه توضیح هر یک از آن‌ها ارائه می‌گردد.

۲-۴ مدل‌های دسته‌بندی متن

۱-۲-۴ شبکه عصبی پرسپترون ساده

شبکه عصبی شامل شبکه‌ای از عناصر پردازش ساده (نورون‌ها) است، که می‌تواند رفتار پیچیده کلی تعیین‌شده‌ای از ارتباط بین عناصر پردازش و پارامترهای عنصر را نمایش دهد. این نورون‌ها مجموعه‌ای از ویژگی‌های ورودی را گرفته و با توجه به ماتریس وزن، تمایل^۱ و تابع فعال‌سازی^۲ خروجی را به دست می‌آورد. شکل ۱-۴ نمایی از شبکه عصبی ساده را نمایش می‌دهد.



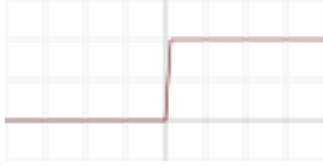
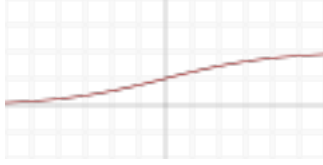
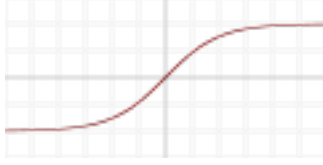
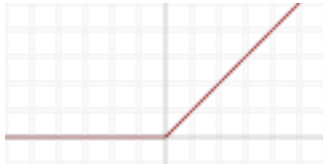
شکل ۱-۴: نمایی از شبکه عصبی دارای نورون‌های مدل پرسپترون ساده

توابع فعال‌ساز با استفاده از ایجاد روابط غیرخطی میان ورودی و خروجی هر نورون سعی می‌کند تا ارتباطات پیچیده‌تری را یاد بگیرد. در جدول ۱-۴ چند نمونه از این توابع فعال‌ساز با نمودار مربوط به آنها آورده شده است.

¹Bias

²Activation function

جدول ۴-۱: توابع فعال‌ساز پرکاربرد به همراه روابط و نمودار آنها

نام	نمودار	تعریف ریاضی
Binary step		$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$
Sigmoid		$f(x) = \frac{1}{1 + \exp(-x)}$
Tanh		$f(x) = \tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$
Relu		$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$

۲-۲-۴ شبکه عصبی پیچشی

این دسته از شبکه‌ها که در ابتدا برای حل مسائل بینایی ماشین ارائه شده بود با به کار بستن هسته‌هایی^۳ با وزن‌های مشترک روی ناحیه‌هایی از ورودی عمل پیچش را انجام می‌دهد.

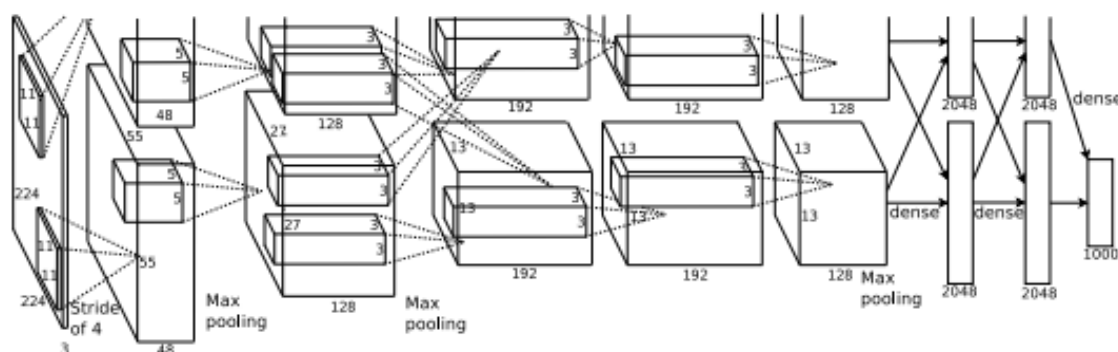
ایده استفاده از وزن‌های مشترک زمانی مطرح شد که تصاویر خاصیت ایستا دارد. این بدان معناست که آماره‌های بخش‌های مختلف یک تصویر و الگوی کلی آنچه که قرار است در تصویر تشخیص داده شود ثابت است و تغییری نمی‌کند. اما استفاده از شبکه‌های پیچشی تنها منحصر به حوزه پردازش تصویر نیست و به مرور وارد سایر حوزه‌ها نظیر پردازش متن نیز شده‌است. این نوع شبکه‌ها عمدتاً از لایه‌هایی مانند ادغام^۴ و پیچش^۵ تشکیل شده‌است. لایه پیچش از فیلترهایی استفاده می‌کند که با انجام عملیات پیچش بر روی داده ورودی، یک نگاشت ویژگی جدید تولید می‌کند. لایه ادغام نیز معمولاً پس از لایه

^۳Kernel

^۴Pool

^۵Convolution

پیچش، برای نمونه‌گیری از نگاشت ویژگی استفاده می‌کند. دو نوع کاربرد این نمونه‌گیری، «میانگین مقادیر»^۶ و «بیشترین مقدار»^۷ است. در نهایت پس از به‌دست‌آمدن نگاشت ویژگی نهایی و بردارسازی آن، از یک شبکه عصبی چند لایه استفاده می‌شود تا خروجی نهایی باتوجه به ورودی به‌دست آید. از شبکه‌های عصبی پیچشی دو بعدی، عمدتاً برای عکس و از شبکه‌های عصبی پیچشی یک‌بعدی بیشتر برای متن استفاده می‌شود. در شکل ۲-۴ یک نمونه از شبکه عمیق پیچشی که برای دسته‌بندی عکس استفاده می‌شود نشان داده شده است.



شکل ۲-۴: نمای کلی مدل پیچشی الکسنت (کریژفسکی و دیگران، ۲۰۱۲)

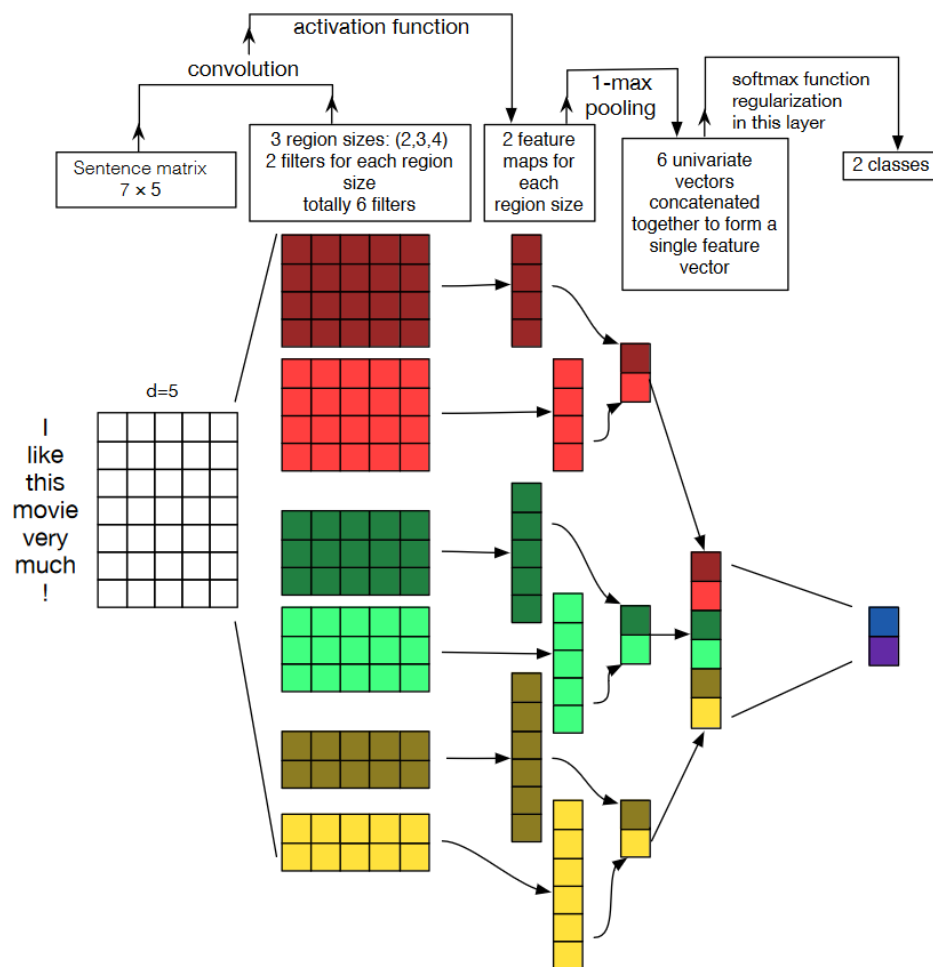
در ادامه یک مثال کاربردی از شبکه پیچشی در زمینه پردازش متن توضیح داده می‌شود. شکل ۳-۴ را در نظر بگیرید. یکی از مسائل کاربرد در حوزه پردازش زبان طبیعی، «تحلیل احساسات»^۸ است که در این مثال ما به‌طور خاص به‌دنبال تشخیص احساس و تمایل در جمله I like this movie very much هستیم. در این مثال کاربردی، داده ورودی به جای یک تصویر، یک ورودی به ابعاد $n \times d$ است که n طول جمله بر حسب واژه‌ها و d ابعاد تعبیه^۹ واژه‌ها می‌باشد. سپس در لایه پیچش، روی این ورودی هسته‌هایی با ابعاد مختلف حلقه‌ای زده می‌شود. معمولاً ابعاد این هسته‌ها برای متن ورودی یک بعدی است که این به‌معنای داشتن یک پنجره لغزان روی سطرهای ماتریس می‌باشد. به بیان دیگر، می‌توان این هسته‌ها را بازنمایی سطح بالا از چندتایی‌های واژه‌های ورودی دانست. در لایه بعد، ادغام بیشینه روی خروجی مرحله قبل انجام شده و خروجی آن‌ها نیز برای دسته‌بندی به لایه تماماً متصل داده می‌شود. در مسئله تحلیل احساس می‌توان فرض کرد که عبارت like this movie (که مهم‌ترین عامل

^۶Average pooling

^۷Max pooling

^۸Sentiment Analysis

^۹Embedding



شکل ۴-۳: نمای کلی یک شبکه عصبی عمیق پیچشی (له و دیگران، ۲۰۱۷)

نشان‌دهنده احساس است) در هر جای متن به عنوان احساس مثبت در نظر گرفته شده و مکان این عبارت برای تصمیم‌گیری در مورد دسته‌بندی آن اهمیتی ندارد. بنابراین، با حرکت آن پنجره لغزان می‌توان عبارات مثبت را شناخت؛ و در نهایت می‌توان با استفاده از این اطلاعات، مثبت یا منفی بودن احساس کلی عبارت را مشخص نمود.

۴-۳ بازنمایی برداری اخبار با رویکردهای مختلف

یکی از چالش‌های پردازش زبان طبیعی یافتن یک بازنمایی مناسب برای جملات و یا اجزای کوچکتر آن مانند واژه‌ها است. واژه علاوه بر صورت‌واژه که از طریق خط تجلی عینی پیدا می‌کند، دارای معنا است که معنای واژه با توجه به بافتی که در آن ظاهر می‌شود تعیین می‌شود. مهم‌ترین عامل تأثیرگذار بر کیفیت یک بازنمایی، شیوه مدل‌سازی معنایی واژه‌ها و یا جملات است. هرچقدر حجم اطلاعات در این بازنمایی

بیشتر باشد، بازنمایی دقیق‌تری از واژه یا جمله به دست می‌آید. در سال‌های اخیر بازنمایی‌های متفاوتی در حوزه پردازش زبان طبیعی معرفی شده است. در ادامه به توضیح چهار نوع بازنمایی مختلف که به چهار حیطه مختلف اشاره دارد می‌پردازیم. این بازنمایی‌ها عبارت است از: بازنمایی بسامد واژه-معکوس بسامد سند که نوعی بازنمایی پایه مبتنی بر واژه است، بازنمایی تعبیه ایستا ورد۲وک که نوعی بازنمایی مبتنی بر معنا براساس هم‌نشینی کلمات است، بازنمایی تخصیص نهان دیریکله که نوعی بازنمایی مبتنی بر موضوع است و مدل‌هایی مانند برت که تحت عنوان بازنمایی‌های مبتنی بر معنای بافت‌محور مطرح می‌شوند.

۱-۳-۴ بازنمایی مبتنی بر واژه

یکی از ساده‌ترین روش‌ها برای بازنمایی اخبار روش مبتنی بر واژه است. در این روش با استفاده از شمارش واژه‌های پراهمیت در یک پیکره متنی، یک بازنمایی از آن پیکره ساخته می‌شود. این روش از دو بخش بسامد واژه و معکوس بسامد سند تشکیل شده است. بسامد واژه برابر است با تعداد دفعاتی که یک واژه در یک متن تکرار شده است. البته برای محاسبه این بسامد، فرمول‌های متنوعی وجود دارد. همچنین معکوس بسامد سند به معنای تعداد دفعاتی است که یک اصطلاح در اسناد دیگر به کار رفته است که همانند بسامد واژه، روش‌های متنوعی برای محاسبه این معیار وجود دارد که در جدول ۲-۴ و ۳-۴ آورده شده است. N تعداد تمام اسناد، n_t تعداد اسنادی که واژه t در آنها وجود دارد.

جدول ۲-۴: روابط قابل استفاده برای محاسبه بسامد واژه در یک سند

رویه وزن‌دهی ^{۱۰}	وزن بسامد واژه
دودویی ^{۱۱}	۰, ۱
تعداد	$f_{t,d}$
بسامد واژه	$\frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$
هنجارسازی لگاریتمی ^{۱۲}	$\log(1 + f_{t,d})$
هنجارسازی دو نیم ^{۱۳}	$0.5 + 0.5 \cdot \frac{f_{t,d}}{\max_{t' \in d} f_{t',d}}$
هنجارسازی دو کا ^{۱۴}	$K + (1 - K) \cdot \frac{f_{t,d}}{\max_{t' \in d} f_{t',d}}$

⁷Weighting scheme

⁸Binary

⁹Log normalization

¹⁰Double normalization 0.5

¹¹Double normalization k

جدول ۴-۳: روابط قابل استفاده جهت محاسبه معکوس بسامد سند

وزن معکوس بسامد سند	رویه وزن‌دهی
۱	یگانه
$-\log \frac{n_t}{N}$	معکوس بسامد سند
$\log\left(\frac{N}{1+n_t}\right) + 1$	معکوس بسامد سند هموار
$\log\left(\frac{\max_{t' \in d} n_{t'}}{1+n_t}\right)$	معکوس بسامد سند بیش‌ترین
$\frac{N-n_t}{n_t}$	معکوس بسامد سند احتمالاتی

۴-۳-۲ بازنمایی مبتنی بر تعبیه ایستا (ورد۲وک)

در روش مبتنی بر واژه، بر اساس فراوانی و توالی واژه‌های استفاده‌شده در یک پیکره سعی می‌کنیم تا یک بازنمایی مناسب ایجاد کنیم. اما این بازنمایی ارتباط میان اجزای جمله را به‌صورت محدود مدل‌سازی می‌کند. در روش مبتنی بر تعبیه ایستا، با استفاده از یک پیکره بزرگ متنی، برای هر واژه یک بازنمایی در اندازه پنجره مشخص ساخته می‌شود. این بازنمایی تا حد قابل قبولی مفاهیم هر واژه را در ساخت بردار در نظر می‌گیرد. خروجی این شیوه بردارسازی به این صورت است که اگر با استفاده از ابزار مصورسازی و کاهش بُعد، یک تصویر از مجموعه واژه‌های یک پیکره را داشته باشیم، اسامی شهر و یا نام ورزش‌های مختلف در نزدیکی هم دیده خواهد شد. علاوه بر این، روابط منطقی نیز میان بردارهای واژه‌ها برقرار است، به‌عنوان مثال، تفاضل بردار واژه‌های «مرد» و «زن» با تفاضل بردار واژه‌های «پادشاه» و «ملکه» برابر است. در بازنمایی ورد۲وک، واژه‌ها به فضای برداری نگاشت می‌شود. اگر نیاز به نگاشت اسناد باشد روش داک۲وک^{۱۵} و یا میانگین برداری ورد۲وک مورد استفاده قرار می‌گیرد. الگوریتم مبتنی بر تعبیه ایستا از دو مدل کیسه‌واژه پیوسته^{۱۶} و پرش‌نگاشت^{۱۷} (میکلوف و دیگران، ۲۰۱۳) استفاده می‌گردد که در ادامه به‌صورت خلاصه آنها را مرور می‌کنیم.

مدل کیسه‌واژه پیوسته

برای یادگیری بردار واژه‌ها از همنشینی واژگان در یک پنجره لغزان استفاده می‌شود. در مدل کیسه‌واژه پیوسته، یک واژه به‌عنوان واژه هدف انتخاب شده و به اندازه پنجره لغزان واژه‌های قبل و بعد آن در فرایند یادگیری بردار واژه هدف شرکت می‌کنند. در این مدل هر واژه با استفاده از بازنمایی تک‌روشن به یک

¹⁵Doc2vec

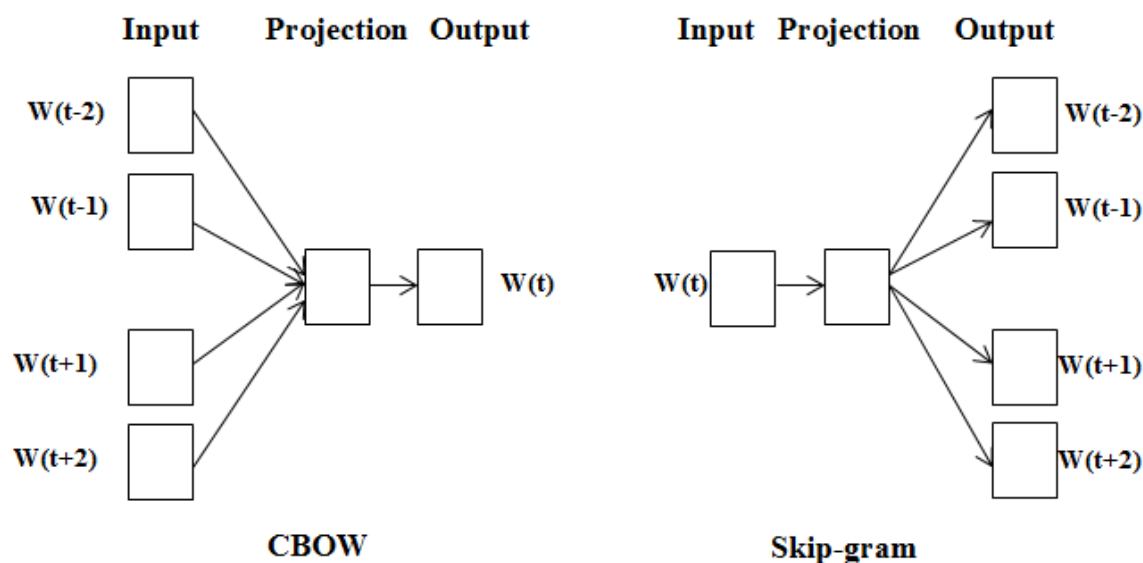
¹⁶Continuous Bag of words

¹⁷Skip-gram

شبکه عصبی با یک لایه میانی داده می‌شود و در خروجی انتظار داریم تا احتمال واژه هدف بیشتر از بقیه واژه‌ها باشد. به این صورت مدل تلاش می‌کند تا واژه هدف را با استفاده از واژه‌های دیگر حدس بزند.

مدل پرش‌نگاشت

در مدل کیسه‌واژه پیوسته معنای برخی از واژه‌ها به درستی یادگرفته نمی‌شود؛ چراکه در این میان، تعداد بسیار زیادی از واژه‌های هم‌معنا محو خواهند شد. به همین دلیل، این مدل دقیقاً برعکس مدل کیسه‌واژه پیوسته عمل می‌کند تا مشکل معنایی واژه‌های کم تکرار حل شود؛ به این شکل که مدل شبکه عصبی با یک لایه میانی، بردار واژه میانی را دریافت می‌کند و در خروجی، انتظار داریم تا واژه‌هایی که در بافت جایگاهی^{۱۸} اطراف واژه هدف در پنجره لغزان وجود دارد، پیش‌بینی گردد. بنابراین با کاهش هزینه در هر مرحله، مدل تلاش می‌کند تا با توجه به روابط میان واژه‌ها، برداری را برای بازنمایی آنها در لایه میانی یاد بگیرد. ساختار این مدل در شکل ۴-۴ آورده شده است.



شکل ۴-۴: نمای کلی دو مدل پرش‌نگاشت و کیسه‌واژه پیوسته (سلیمان و دیگران، ۲۰۱۷)

۴-۳-۳ بازنمایی مبتنی بر موضوع (تخصیص نهان دیریکله)

الگوریتم تخصیص نهان دیریکله^{۱۹} یک روش مدل‌سازی بدون نظارت برای تعیین موضوعات پیکره متنی است. روش عملکرد مدل‌سازی موضوع به این صورت است که پس از ورود پیکره به این الگوریتم، در

¹⁸Local context

¹⁹Latent Dirichlet allocation (LDA)

نهایت ماتریس سند-موضوع و ماتریس واژه-موضوع ساخته می‌شود. به‌منظور استفاده از این روش برای یافتن یک بازنمایی مناسب برای هر سند می‌توانیم از ماتریس اول استفاده کنیم. هر سطر ماتریس سند-موضوع مرتبط با یک سند است و ستون آن احتمال هر موضوع را نشان می‌دهد. همچنین در ماتریس واژه-موضوع هر ستون یک موضوع است و هر سطر آن نیز امتیاز واژه در آن موضوع است.

۴-۳-۴ مدل‌های بازنمایی متن مبتنی بر بافت

با معرفی معماری انتقال‌دهنده‌ها (واسوانی و دیگران، ۲۰۱۷) طیف وسیعی از مدل‌های زبانی مبتنی بر بافت معرفی شدند. در این بخش به بررسی مدل‌های برت (دولین و دیگران، ۲۰۱۸) و مدل‌هایی که معماری مشابه برت را دارند مانند برت چند زبانه^{۲۰}، پارس‌برت^{۲۱} (فراهانی و دیگران، ۲۰۲۰)، آلبرت-فارسی^{۲۲} (فراهانی، ۲۰۲۰) و ایکس.ال.ام-روبرتا^{۲۳} (کنو و دیگران، ۲۰۱۹) می‌پردازیم.

برت

دولین و دیگران (۲۰۱۸) یک مدل زبانی جدید با نام برت را معرفی کردند. ایده اصلی این مدل استفاده از انتقال‌دهنده‌های^{۲۴} دوطرفه برای یادگیری معنا و ساختار واژه‌های موجود در متن است. مدل برت به‌صورت دو نسخه «برت پایه»^{۲۵} و «برت بزرگ»^{۲۶} معرفی شده است. برت پایه دارای ۱۲ لایه انتقال‌دهنده و ۱۱۰ میلیون پارامتر و برت بزرگ دارای ۲۴ لایه و ۳۴۰ میلیون پارامتر است. انتقال‌دهنده‌ها از «مکانیزم توجه»^{۲۷} در فرایند یادگیری استفاده می‌کند و سعی می‌کند تا ارتباط مفهومی میان واژه‌های موجود در یک جمله را به‌درستی یاد بگیرد. این مدل برای ایجاد یک بازنمایی مناسب از جملات از ترکیب سه تعبیه استفاده می‌کند که در ادامه توضیحات آنها ارائه می‌شود:

- تعبیه واژه^{۲۸}: منظور از تعبیه واژه تبدیل هر واژه به فضای برداری یکتا است.

²⁰Multilingual BERT

²¹ParsBERT

²²ALBERT-Persian

²³XLNet-RoBERTa

²⁴Transformer

²⁵BERT-Base

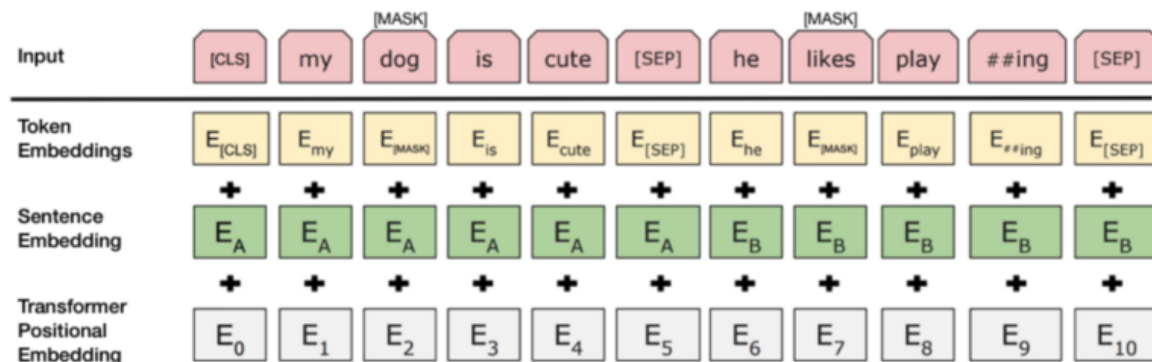
²⁶BERT-Large

²⁷Attention mechanism

²⁸Token Embedding

- تعبیه جمله ^{۲۹}: هر داده ورودی مدل برت از دو جمله “آ” و “ب” تشکیل شده‌است که در یکی از مراحل یادگیری، این مدل برای پیش‌بینی جمله بعدی کاربرد دارد. به همین منظور، این مدل یک تعبیه برای مشخص کردن جمله‌ای که واژه مورد نظر در آن قرار دارد، ایجاد می‌کند.
- تعبیه موقعیت انتقال‌دهنده‌ها ^{۳۰}: تعبیه موقعیت انتقال‌دهنده‌ها یک تعبیه از موقعیت هر واژه در جمله است که باعث می‌شود بردار حضور هر واژه در جایگاه‌های مختلف در یک جمله متفاوت باشد.

درنهایت با ترکیب این سه بخش، یک بردار ورودی برای مدل برت ساخته می‌شود که پس از مراحل یادگیری گفته‌شده در بخش قبل، برای هر جمله یک بازنمایی مناسب ارائه می‌یابد. شکل ۴-۵ ساختار این روش را نمایش می‌دهد.



شکل ۴-۵: نحوه ترکیب تعبیه‌های سه‌گانه در مدل زبانی برت (دولین و دیگران، ۲۰۱۸)

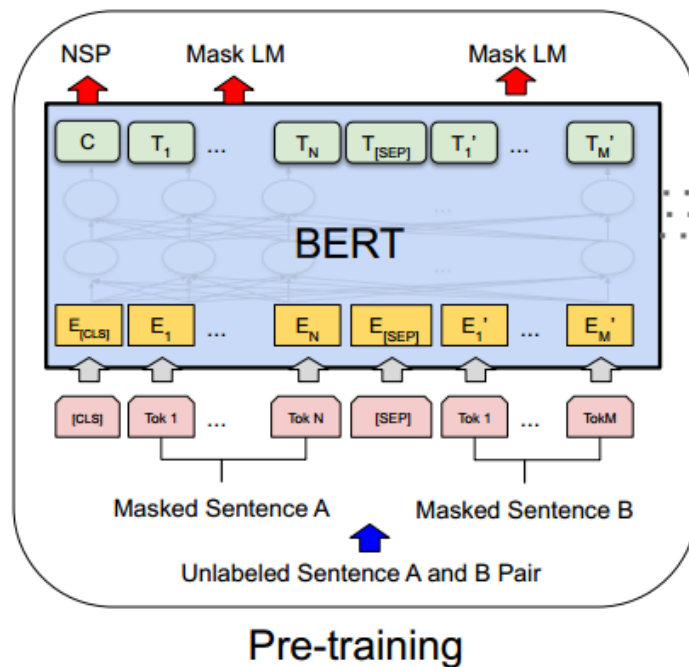
فرایند یادگیری مدل‌های زبانی عموماً به این صورت است که تلاش می‌کند تا واژه بعدی یک دنباله را حدس بزند و بر این اساس، ارتباط میان واژه‌ها را تشخیص دهند. اما در مدل زبانی برت که به‌صورت دوطرفه (چپ‌به‌راست و راست‌به‌چپ) متن را بررسی می‌کند نمی‌توان فقط از این روش استفاده کرد. همچنین برای آموزش مدل برت دو مرحله اصلی برروی پیکره‌های متنی ویکی‌پدیا و کتاب‌ها اعمال شده‌است که در ادامه با آن‌ها آشنا می‌شویم. در شکل ۴-۶ نحوه آموزش در مدل برت نمایش داده شده‌است.

- پوشش واژه‌ها

قبل از استفاده از داده‌ها در شروع فرایند یادگیری، ۱۵ درصد از واژه‌های داخل متن به‌صورت

²⁹Sequence Embedding

³⁰Transformer Positional Embedding



شکل ۴-۶: نحوه عملکرد یادگیری مدل برت (دولین و دیگران، ۲۰۱۸)

تصادفی انتخاب می‌شود. از این حجم واژه‌ها ۸۰ درصد واژه‌ها با عبارت [MASK] جایگزین شده و ۱۰ درصد با یک واژه تصادفی جایگزین می‌شود و ۱۰ درصد دیگر بدون تغییر باقی می‌ماند. در ادامه، مدل سعی می‌کند تا با استفاده از اطلاعات زمینه‌ای، این واژه‌ها را حدس بزند. این کار با اضافه کردن یک لایه پس از لایه کدگذار مدل برت انجام می‌شود تا با استفاده از آن، احتمال وجود هریک از واژه‌ها مشخص شود و با داشتن واژه اصلی، خطای مدل حساب شده و تمامی وزن‌ها به‌روزرسانی می‌شود.

• پیش‌بینی جمله بعدی

در این حالت، مدل برت یک جفت دنباله مانند "آ" و "ب" را دریافت می‌کند. در ابتدا و انتها جملات عبارات [CLS] و [SEP] قرار می‌گیرد. پس از اضافه کردن نوع جمله ("آ" یا "ب") به ویژگی‌های نهفته هر جمله، تمامی ورودی‌ها وارد انتقال‌دهنده‌ها شده و در نهایت با استفاده از یک لایه، مقدار احتمال ظهور دنباله "ب" به‌عنوان جمله بعدی دنباله "آ" مشخص می‌شود. با این روش مدل برت تلاش می‌کند تا مفاهیم زمینه‌ای را به‌طور کامل میان واژه‌ها و عبارات یاد بگیرد.

در نهایت پس از آماده‌شدن مدل زبانی، می‌توانیم از آن در مسائل مختلف مانند دسته‌بندی متن،

پرسش و پاسخ^{۳۱} و تشخیص موجودیت‌های نامدار^{۳۲} استفاده کنیم.

برت چند زبانه

مدل زبانی برت چند زبانه (دولین و دیگران، ۲۰۱۸) با استفاده از داده‌هایی از ۱۰۴ زبان مختلف شامل زبان فارسی آموزش داده شده است. مدل زبانی برت دارای دو نوع پایه و بزرگ است که ما در این پروژه از مدل برت پایه، شامل ۱۲ لایه انتقال‌دهنده و ۱۱۰ میلیون پارامتر استفاده کرده‌ایم. این مدل با استفاده از پیکره متنی ویکی‌پدیا^{۳۳} که به زبان‌های مختلف موجود است، آموزش داده شده است.

پارس‌برت

فراهانی و دیگران (۲۰۲۰) مدل زبانی پارس‌برت که مبتنی بر معماری مدل برت است، انحصاراً برای زبان فارسی آموزش دادند. این مدل عملکرد بهتری نسبت به مدل‌های چند زبانی مانند برت داشته است. همچنین به دلیل آنکه این مدل تنها برای زبان فارسی ارائه شده است، سبک‌تر از مدل زبانی برت چندزبانه است. به منظور آموزش مدل زبانی پارس‌برت از منابع متنی موجود در وبسایت‌های ویکی‌پدیا، بیگ‌بنگ، چطور، الی‌گشت، دیجی‌کالا، تداک، کتاب‌ها و پیکره میراث استفاده شده است. همچنین برای ارزیابی مدل ارائه شده در پارس‌برت از تحلیل احساسات، تشخیص موجودیت‌های نامدار و دسته‌بندی متن استفاده شده است.

آلبرت-فارسی

مدل زبانی آلبرت یک نسخه سبک‌تر از برت است که معماری کاملاً مشابهی با آن دارد که در پیاده‌سازی آن تغییر اندکی نسبت به مدل برت وجود دارد (فراهانی، ۲۰۲۰). آلبرت-فارسی، مانند مدل زبانی پارس‌برت بر روی پیکره‌های متنی فارسی آموزش داده شده است. علیرغم اینکه منابع این پیکره‌های فارسی مشابه مدل زبانی پارس‌برت است، به مراتب حجم کمتری نسبت به پارس‌برت دارد.

ایکس.ال.ام-روبرتا

با توجه به موفقیت‌های چشمگیر مدل‌های ازپیش آموزش دیده، کنو و دیگران (۲۰۱۹) یک مدل بین‌زبانی ارائه کردند که توانست دقت بهتری را در مقایسه با مدل‌های زبانی رایج مانند برت ثبت کند. این مدل

³¹Question Answering (QA)

³²Name Entity Recognition (NER)

³³Wikipedia

که بر روی دادگان ۱۰۰ زبان مختلف از جمله فارسی آموزش داده شده است با استفاده از ۳ روش آموزش داده می‌شود: (۱) پیش‌بینی کلمه بعدی، (۲) پیش‌بینی کلمه پوشیده شده و (۳) مدل زبانی مبتنی بر ترجمه.

۴-۴ معماری‌های پیشنهادی برای تشخیص اخبار جعلی

به طور کلی، مدل‌های بازنمایی متن مبتنی بر بافت خروجی‌های متفاوتی را می‌توانند ایجاد کنند. مدل‌هایی با ساختار مشابه برت مانند پارس‌برت، آلبرت-فارسی و برت چند زبانه ۲ نوع بازنمایی برای متن ورودی ایجاد می‌کنند:

- بازنمایی تجمعی^{۳۴} به ازای هر دنباله از واژه‌های ورودی، یک بازنمایی کلی برای کل دنباله ارائه می‌کند. این بازنمایی تنها توسط مدل‌های مبتنی بر ساختار برت ایجاد می‌شود.

- بازنمایی دنباله‌ای^{۳۵} که در آن برای هر واژه یک بردار بازنمایی مبتنی بر بافت ارائه می‌گردد. برخلاف بازنمایی تجمعی، بازنمایی دنباله‌ای را تمامی مدل‌های بازنمایی متن مبتنی بر بافت تولید می‌کنند.

در این پروژه، دو معماری مختلف مورد استفاده قرار گرفته‌است که در هر یک از این معماری‌ها، یکی از دو حالت فوق به کار رفته‌است. در روش اول با استفاده از یک مدل پرسپترون ساده به دسته‌بندی اخبار بازنمایی شده توسط بردارهای تجمعی می‌پردازیم و در روش دوم با استفاده از بازنمایی دنباله‌ای و یک شبکه عصبی پیچشی به استخراج ویژگی‌های سطح بالاتر از هم‌نشینی واژگان یک خبر می‌پردازیم و در نهایت با استفاده از یک لایه چگال، اخبار دسته‌بندی می‌شوند.

۴-۴-۱ مدل زبانی مبتنی بر بافت + پرسپترون ساده

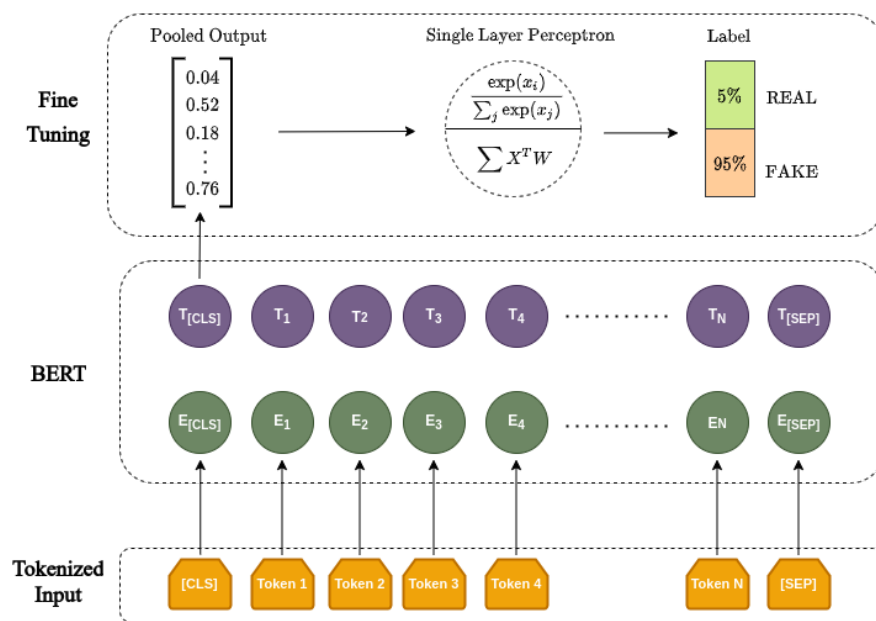
برای استفاده از مدل‌های مبتنی بر بافت در مسئله دسته‌بندی کافی است تا از بردار خروجی متناظر با توکن [CLS] به عنوان بردار نماینده آن دنباله استفاده کنیم. دلیل استفاده از این بردار این است که شامل تمامی بردارهای توکن‌های موجود در متن است، درحالی‌که بردارهای متناظر با توکن‌های دیگر تنها نماینده آن توکن در عبارت ورودی است. بنابراین کافی است تا بردار نماینده دنباله متنی ورودی

³⁴Pooled

³⁵Sequence

را وارد یک لایه تکی پرسپترون کنیم که می‌تواند عمل تشخیص برچسب دنباله ورودی را برای ما انجام دهد. این ساختار با استفاده از مدل زبانی برت در شکل ۴-۷ ترسیم شده است.

به عبارت دیگر، با استفاده از بازنمایی به دست آمده از مدل برت برای یک جمله و داشتن برچسب آن سعی می‌کنیم تا با استفاده از دادگان موجود، یک مدل کامل برای تشخیص اخبار جعلی بسازیم. برای این هدف کافی است تا وزن‌های ورودی به لایه پرسپترون ساده آموزش داده شود به صورتی که بتواند اخبار را به درستی دسته‌بندی کند.



شکل ۴-۷: نحوه استفاده از یک لایه پرسپترون ساده در کنار مدل برت

به منظور جلوگیری از «بیش‌برازش»^{۳۶} مدل، از روش «حذف تصادفی نورون‌ها»^{۳۷} استفاده می‌کنیم تا مدل نهایی تا حد ممکن ساده باشد و به درستی ویژگی‌های مهم را یاد بگیرد. در انتهای ترین بخش مدل هم یک لایه نورون به تعداد دسته‌های دادگان ورودی وجود دارد که تابع فعال‌ساز آنها «بیشینه نرم»^{۳۸} است چراکه نیاز داریم تا احتمال تعلق داشتن خبر به هر دسته را بدانیم و بیشترین آن را به عنوان برچسب پیش‌بینی شده اعلام نماییم.

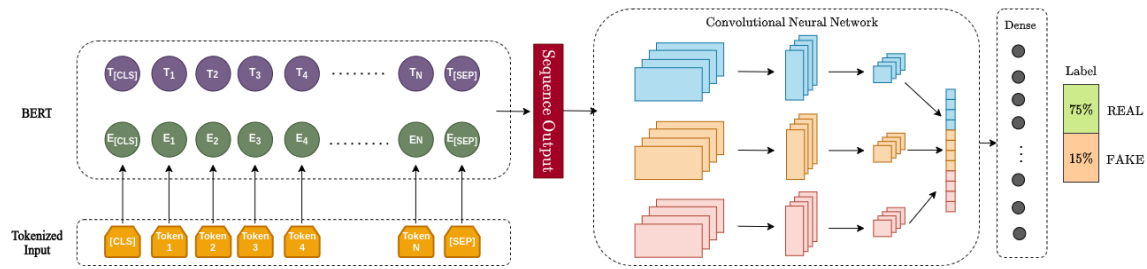
تابع «خطا آنتروپی متقاطع طبقه‌ای»^{۳۹} نیز برای محاسبه میزان خطای پیش‌بینی مدل دسته‌بند خبر جعلی استفاده شده است که رابطه ریاضی آن در ادامه آورده شده است.

³⁶Overfitting

³⁷Dropout

³⁸Softmax

³⁹ Categorical Cross-entropy



شکل ۴-۸: نحوه ارتباط لایه پیچشی با بردارهای خروجی مدل برت

$$Loss = - \sum_{c=1}^M y_{o,c} \cdot \log(p_{o,c}) \quad (4-1)$$

در این رابطه، M تعداد دسته‌ها و y_o بردار تک‌روشن از برچسب خبر o است و p_o نیز بردار احتمال پیش‌بینی‌شده توسط مدل است که نشان‌دهنده احتمال تعلق به هر دسته است.

۴-۴-۲ مدل زبانی مبتنی بر بافت + پیچشی

با استفاده از خروجی دنباله‌ای یک مدل بازنمایی مبتنی بر بافت و اتصال آن به یک شبکه عصبی پیچشی، عملیات نگاشت ویژگی انجام داده می‌شود. پس از عملیات یادگیری، این مدل سعی می‌کند تا با تبدیل بازنمایی تولیدشده به یک بازنمایی دیگر، عملیات پیش‌بینی کذب‌بودن یا اصیل‌بودن یک خبر را انجام دهد. در این مدل، به جای بردار خروجی تجمعی، از بردار خروجی دنباله‌ای که شامل دنباله‌ای از بردارهای تک‌تک واژه‌های داخل خبر است استفاده می‌گردد؛ چراکه با استفاده از لایه‌های پیچشی پس از آن می‌توان هم‌نشینی‌هایی از واژه‌ها که نشان‌دهنده اخبار جعلی است را شناسایی کرد. برای این منظور، از سه لایه پیچشی موازی شامل ۳۰ فیلتر با اندازه‌های ۳، ۴ و ۵ استفاده می‌کنیم. همچنین تابع فعال‌ساز این لایه، تابع «یک‌سوساز خطی»^{۴۰} است. پس از این لایه، یک لایه ادغام بیشترین مقدار قرار دارد تا با استفاده از عملیات «نمونه‌کاهی»^{۴۱}، ویژگی‌های مناسب را استخراج کند. درنهایت، با استفاده از یک «لایه مسطح‌شده»^{۴۲}، ماتریس مرحله قبل به یک بردار یک بعدی تبدیل می‌شود تا با استفاده از یک لایه چگال^{۴۳} با استفاده از تابع فعال‌ساز «بیشینه نرم»، احتمال هر دسته مشخص شود. شکل ۴-۸ نحوه

⁴⁰ReLU

⁴¹Downsampling

⁴²Flatten layer

⁴³Dense

اتصالات و به‌دست‌آوردن ویژگی‌های سطح بالاتر با استفاده از مدل زبانی برت را نشان می‌دهد.

فصل پنجم

بررسی نتایج و ارزیابی مدل‌های ارائه شده

۱-۵ آزمایش‌های انجام‌شده بر روی پیکره «تاج»

پس از استخراج یک مجموعه داده در حوزه اخبار جعلی، به ارزیابی مدل‌های ارائه شده می‌پردازیم. همانطور که در بخش ۴-۳-۴ عنوان شد، چندین مدل از پیش آموزش دیده برای زبان فارسی موجود است. در این بخش ما با استفاده از دسته‌بندهای ذکرشده، پیچشی و پرسپترون، یک مقایسه میان این مدل‌ها انجام می‌دهیم تا بهترین مدل را برای تشخیص اخبار جعلی فارسی انتخاب کنیم. به منظور اثبات کارایی مدل‌های ارائه‌شده، ما از دادگان دیگر فارسی و انگلیسی نیز برای ارزیابی استفاده می‌کنیم.

۱-۱-۵ پیش‌پردازش داده‌های خبری

یکی از مهم‌ترین بخش‌های تشخیص اخبار جعلی پیش‌پردازش داده‌های خبری است. پیش‌پردازش داده‌ها به معنی یک‌دست‌سازی واژه‌ها برای استفاده در الگوریتم‌های پردازش متن است که در ادامه مراحل آن را مرور می‌کنیم. در این پروژه، از ابزار پیش‌پردازش ارائه شده توسط **عبدالله‌پور (۲۰۲۰)** استفاده شده است.

هنجارسازی^۱ واژه‌ها

قبل از این که مجموعه‌ای از متن‌ها به‌عنوان داده ورودی مورد استفاده الگوریتم‌های یادگیری قرار گیرد، ابتدا باید پیش‌پردازش‌هایی روی آنها انجام گیرد تا صورت‌های غیرمعیار به شکل معیار تبدیل گردد. اگر حروف، نشانه‌های نگارشی و واژه‌های فارسی به شکل یکسانی نوشته نشود، مجموعه داده استفاده‌شده قابل تحلیل توسط سامانه‌های رایانه‌ای نخواهند بود. طی فرایند هنجارسازی، علایم نگارشی، حروف، فاصله‌های بین واژه‌ها و اختصارات بدون ایجاد تغییرات معنایی در متن، به شکل استاندارد تبدیل می‌گردد. برای عنوان مثال، برخی از حروف در زبان فارسی به کدهای مختلف به چند شکل متفاوت نوشته می‌شود که باید به یک صورت یکسان تبدیل گردد، مانند انواع حروف «ک» در زبان فارسی و عربی و یا شکل‌های متفاوت نوشتاری حرف «ی» در فارسی، عربی و پشتو. همچنین در این مرحله تمامی اعراب‌ها مانند فتحه، کسره، ضمه و تشدید از واژه‌ها حذف و یا به صورت استاندارد تبدیل می‌شود. ممکن است علامت تنوین^۲ در واژه‌ای مانند «حتماً» یا همزه آخر بعضی از واژه‌ها مانند «انشاء» حذف شده باشد که برای کاهش تنوع نگارشی واژه‌ها و کوچک‌سازی واژگان هنجارسازی واژه‌ها انجام می‌پذیرد.

^۱Text normalization

واحدسازی^۲ واژه‌ها

در زبان فارسی بعضی از حروف چهار شکل دارد که سه شکل آن به حرف مجاور چسبیده است و شکل دیگر منفصل است. مجموعه محدودی از حروف مانند «ا»، «د»، «ذ»، «ر»، «ز»، «ژ» و «و» فقط دارای دو شکل منفصل یا چسبیده به حرف قبلی را دارد. عدم چسبندگی این مجموعه محدود به حرف بعدی سبب می‌شود در هنگام نگارش متن فاصله کامل به‌عنوان مرزنامی واژه وارد نشود و یک واحد به‌صورت یکپارچه نوشته شود، مانند ویابتر است. اگرچه ممکن است زنجیره حروف در این مثال خوانش توسط انسان را دچار اختلال نکند، تمام زنجیره به‌عنوان یک واژه توسط رایانه تشخیص داده می‌شود. نادیده گرفتن درج فاصله کامل در متن به‌خصوص متن‌های خبری که با سرعت تهیه و منتشر می‌شود اجتناب‌ناپذیر است. هر خبر شامل مجموعه از جملات است که هر جمله نیز از مجموعه‌ای از واژه تشکیل شده است. به‌منظور قابل‌فهم کردن اخبار برای الگوریتم‌های یادگیری ماشین و شبکه‌های عمیق باید هر جمله خبر به کوچک‌ترین جزء تشکیل‌دهنده که همان واژه‌ها است تجزیه شود. به این عمل قطعه‌بندی جملات می‌گویند. در زبان فارسی بسیاری از واژه‌های مرکب دارای نیم‌فاصله در درون واژه است که مانع چسبندگی حروف در درون یک واژه می‌شود که به خوانش راحت‌تر آن توسط گویشور می‌انجامد، مانند فروشگاه‌ها. اما بسیاری از نوشته‌های کاربران و خبرگزاری‌ها از این قانون پیروی نمی‌کند. به‌عنوان مثال واژه‌هایی مانند «می‌نوشد»، «کتاب‌ها»، «بین‌الملل» که باید به ترتیب به‌صورت «می‌نوشد»، «کتاب‌ها» و «بین‌الملل» نوشته شود با فاصله کامل نوشته می‌شود. گاهی وجود فاصله کامل یا نیم‌فاصله به ابهام منجر می‌شود مانند «می‌نوشد» و «می‌نوشد» که رفع ابهام این موارد چندان ساده نیست. لازم به ذکر است ممکن است فاصله بین دو تکواژ حذف شود و صورت دیگری از نوشتار به‌دست آید، مانند علاقمند. بر اساس آنچه ذکر شد چنین می‌توان نتیجه گرفت که پیش از انجام هر عملیاتی برای تقطیع جملات به واژه‌ها، باید عملیات اصلاح این نگارش‌ها انجام شود.

۲-۱-۵ تنظیمات آزمایش‌ها

مدل‌های برت چندزبانه و ایکس.ال.ام-روبرتا دارای نسخه‌های متفاوتی است. در این پژوهش، ما از نسخه پایه هر دو مدل استفاده کردیم. مدل برت چندزبانه پایه دارای ۱۷۹ میلیون پارامتر قابل آموزش، ۱۲ لایه انتقال‌دهنده و لایه میانی با اندازه ۷۶۸ است. مدل ایکس.ال.ام-روبرتا پایه نیز دارای ۲۷۰ میلیون پارامتر، ۱۲ لایه انتقال‌دهنده و لایه میانی با اندازه ۷۶۸ است. مدل‌های دیگر مانند پارس‌برت و آلبرتا-فارسی

²Tokenization

همگی بر پایه مدل برت است.

معماری دسته‌بند پیچشی شامل ۳ لایه پیچشی موازی شامل ۳۰ فیلتر با ابعاد متفاوت ۳ و ۴ و ۵ است که ویژگی‌های سطح بالاتر را از ماتریس خروجی مدل ازپیش آموزش‌دیده که شامل یک بردار مجزا برای هر کلمه ورودی است، استخراج می‌کند. لایه آخر این دسته‌بند نیز یک پرسپترون ساده است که با استفاده از تابع بیشینه‌هموار^۳، برچسب هر خبر را مشخص می‌کند. به‌منظور جلوگیری از مشکل بیش‌برازش^۴ در مدل نیز از پارامترهای تنظیم‌کننده در فیلترها استفاده شده است.

شبکه عصبی پیشرو پرسپترون نیز شامل یک لایه پرسپترون با تابع فعال‌ساز بیشینه‌هموار است که با استفاده از بردار خروجی مدل از پیش آموزش‌دیده به محاسبه احتمال تعلق یک خبر به هر کدام از دسته‌ها می‌پردازد.

در تمامی مدل‌ها نیز یک لایه بیرون‌انداز^۵ قبل از لایه نهایی قرار گرفته است. همچنین در آموزش مدل‌ها نرخ یادگیری^۶ در آموزش برابر با $5 - 5e$ ، تعداد دوره‌ها^۷ برابر با ۴ و مقدار اندازه دسته^۸ ۶۴ است.

۳-۱-۵ نتایج

جدول ۱-۵ نتایج ارزیابی مدل‌های ارائه‌شده با استفاده از مدل‌های از پیش آموزش‌دیده متفاوت و با استفاده از شبکه عصبی پیچشی را نمایش می‌دهد. طبق نتایج به‌دست‌آمده می‌توان نتیجه گرفت که مدل پارس‌برت نسبت به بقیه مدل‌ها توانسته است متن‌های ورودی را دقیق‌تر بازنمایی کند و ما با استفاده از این مدل می‌توانیم اخبار جعلی را با دقت بالاتری نسبت به سایر مدل‌ها تشخیص دهیم.

در گام بعدی، در مدل پارس برت علاوه بر دسته‌بندی با شبکه پیچشی، دسته‌بندی با شبکه پیشرو پرسپترون تک‌لایه نیز مورد بررسی قرار گرفت که نتایج آن در جدول ۲-۵ گزارش شده است. همان‌طور که نتایج این جدول نشان می‌دهد شبکه پرسپترون تک‌لایه به نتایج نزدیکی در مقایسه با شبکه پیچشی رسیده است و با تفاوت اندک به دقت و معیار اف بالاتری دست یافته است. بر اساس این نتایج، بازنمایی پارس‌برت در سطح واژه به همراه شبکه پیچشی و بازنمایی پارس‌برت در سطح متن به همراه شبکه پیشرو پرسپترون در گام‌های بعدی این پروژه مورد استفاده قرار خواهد گرفت.

³Softmax

⁴Overfit

⁵Dropout Layer

⁶Learning Rate

⁷Epoch

⁸Batch Size

جدول ۵-۱: نتایج تشخیص اخبار جعلی فارسی با استفاده از بازنمایی‌های مختلف و دسته‌بندی با شبکه عصبی پیچشی

مدل	فراخوانی	صحت	معیار اف	دقت
برت چندزبانه	۸۸/۵۵	۸۴	۸۶/۲۱	۸۷/۳۳
آلبرت-فارسی	۸۷/۰۲	۹۲	۸۹/۴۴	۸۹/۷۵
پارس برت	۸۹/۱۳	۹۳/۷۱	۹۱/۳۶	۹۱/۶۴
ایکس.ال.ام-روبرتا	۸۷/۸۴	۹۰/۸۵	۸۹/۳۲	۸۹/۷۵
ورد۲وک	۷۷/۳۴	۸۰	۷۸/۶۵	۷۹/۵۱

جدول ۵-۲: نتایج تشخیص اخبار جعلی فارسی با استفاده از بازنمایی پارس برت و مقایسه دسته‌بندها

مدل	دسته‌بند	فراخوانی	صحت	معیار اف	دقت
پارس برت	شبکه عصبی پیچشی	۸۹/۱۳	۹۳/۷۱	۹۱/۳۶	۹۱/۶۴
پارس برت	شبکه پیشرو پرسپترون تک‌لایه	۹۲/۴۴	۹۰/۸۵	۹۱/۶۴	۹۲/۱۸

۵-۲ آزمایش‌های انجام‌شده بر روی سایر پیکره‌های فارسی

همانطور که ذکر شد، از آنجا که دادگان تهیه‌شده در این پروژه برای اولین بار در زمینه تشخیص خبر جعلی مورد استفاده قرار می‌گیرد، نتایج پایه دیگری بر روی این دادگان به‌منظور مقایسه وجود ندارد. بر همین اساس، به‌منظور ارزیابی مدل‌های ارائه‌شده برای زبان فارسی و مقایسه نتایج این پروژه با کارهای قبلی انجام‌شده در زبان فارسی، ما از دو مجموعه داده موجود برای تشخیص شایعات در شبکه‌های اجتماعی توییتر و تلگرام استفاده کردیم. مدل‌های پایه‌ای که برای مقایسه استفاده شده است در ادامه مرور می‌شوند:

- **زمانی و دیگران (۲۰۱۷)** یک روش مبتنی بر مدل «بهینه‌سازی کمینه متوالی»^۹ که یک روش آموزش ماشین بردار پشتیبان است برای تشخیص شایعات فارسی در شبکه اجتماعی تلگرام ارائه داده‌اند. به‌منظور مقایسه نتایج، ما در این بخش نتایج آزمایش مدل‌های ارائه‌شده خود را با مدل ارائه‌شده توسط **زمانی و دیگران (۲۰۱۷)** مقایسه نموده‌ایم. در این مقایسه از دادگانی که توسط **زمانی و دیگران (۲۰۱۷)** ارائه گردیده است استفاده کرده‌ایم. شایان ذکر است در مقاله ارائه‌شده توسط ایشان، علاوه بر اطلاعات متنی، از ویژگی‌های مرتبط با گراف کاربران توییتر نیز استفاده کرده‌اند، اما به‌دلیل عدم دسترسی به این اطلاعات، مقایسه ما تنها با بخشی از مدل‌های ارائه‌شده

^۹Sequential Minimal Optimization (SMO)

جدول ۵-۳: ارزیابی و مقایسه مدل‌های ارائه‌شده با مدل‌های پیشین بر روی دادگان فارسی شبکه‌های اجتماعی

مدل	دادگان	فراخوانی	صحت	معیار اف	دقت
ماشین بردار پشتیبان (زمانی و دیگران، ۲۰۱۷)	توییت‌ر	۷۱/۴۴	۸۱/۲۴	۷۶/۰۲	۷۴/۲۶
ماشین بردار پشتیبان (جهان‌بخش و دیگران، ۲۰۲۰)		۷۶/۳	۷۶/۳	۷۶/۳	-
پارس‌برت - پیچشی		۹۵/۷۱	۹۷/۱۰	۹۶/۴۰	۹۶/۱۲
پارس‌برت - پرسپترون		۹۴/۲۰	۹۴/۲۰	۹۴/۲۰	۹۳/۷۹
ماشین بردار پشتیبان (جهان‌بخش و دیگران، ۲۰۲۰)	تلگرام	۸۲/۸	۸۲/۹	۸۲/۸	-
پارس‌برت - پیچشی		۹۳/۴۹	۹۵/۰۴	۹۴/۲۶	۹۲/۵۱
پارس‌برت - پرسپترون		۹۲/۰۶	۹۵/۸۶	۹۳/۹۲	۹۱/۹۷
پارس‌برت - پرسپترون					

توسط زمانی و دیگران (۲۰۱۷) که بر روی دادگان متنی پیاده شده، انجام پذیرفته است.

- **جهان‌بخش و دیگران (۲۰۲۰)** با ارائه مدل مبتنی بر ماشین بردار پشتیبان به دسته‌بندی شایعات در شبکه اجتماعی توییت‌ر پرداختند. ما در بخش دیگری از آزمایش‌ها از دادگان توییت‌ر ارائه‌شده توسط **جهان‌بخش و دیگران (۲۰۲۰)** استفاده کردیم و مدل خود را با نتایج آن‌ها مقایسه نمودیم.
- جدول ۵-۳ نتایج حاصل از مدل‌های پیاده‌شده در این پژوهش را بر روی دو مجموعه داده حاصل از تلگرام و توییت‌ر نشان می‌دهد. در این آزمایش‌ها از مدل زبانی پارس‌برت استفاده شده است.

همانطور که از نتایج مشخص است، مدل‌های ارائه شده با استفاده از بازنمایی مبتنی بر مفهوم به صورت چشمگیری معیارهای ارزیابی را بهبود دادند. همچنین، دسته‌بند پیچشی درمقایسه با مدل شبکه پرسپترون پیشرو عملکرد بهتری داشته است چراکه در این حالت مدل پیچشی ویژگی‌های سطح بالاتری را براساس هم‌نشینی واژگان ایجاد می‌کند. با توجه به آزمایش‌های انجام شده بر روی دادگان توییت‌ر، مدل پارس‌برت-پیچشی ۲۰ درصد معیار اف را در مقایسه با مدل‌های ارائه شده توسط زمانی و دیگران (۲۰۱۷) و **جهان‌بخش و دیگران (۲۰۲۰)** بهبود داده است. علاوه بر این، این مدل توانست بر روی دادگان تلگرام نیز معیار اف را در مقایسه با نتایج بدست‌آمده توسط **جهان‌بخش و دیگران (۲۰۲۰)** ۱۱ درصد ارتقا دهد.

۵-۳ آزمایش‌های انجام‌شده بر روی پیکره‌های زبان انگلیسی

همان‌طور که پیش‌تر مطرح شد، برای تشخیص خبر جعلی فارسی ابتدا شاکله‌ی یک مدل را براساس داده انگلیسی ساختیم و سپس با تغییر داده به فارسی، مدل را با داده فارسی آموزش داده و آن را بهینه کردیم. از این‌رو، در ادامه، مدل آموزش داده شده با داده انگلیسی توضیح داده می‌شود. برای استفاده از دادگان

لیار از ویژگی متن خبر و ۵ ویژگی مرتبط با تاریخچه اعتبار هر عنوان خبر استفاده می‌گردد. تاریخچه اعتبار هر گوینده شامل تعداد اخبار «به‌سختی درست»، «جعلی»، «نیمه درست»، «عمدتاً درست» و «دروغ» که از این گوینده تا کنون منتشر شده‌است می‌باشد. علاوه بر آنها، تمامی ۶ برجسب که پیش‌تر معرفی شده بود به ۲ دسته «جعلی» و «اصیل» نگاشت می‌شود و برجسب زده می‌شوند؛ به این صورت که برجسب‌های «درست»، «نیمه درست» و «عمدتاً درست» به دسته اصیل و برجسب‌های «جعلی»، «دروغ» و «به‌سختی درست» به دسته جعلی نگاشت می‌شوند. در جدول ۴-۵ نتایج حاصل از ارزیابی این دادگان در حالت ۲ برجسب «جعلی» و «اصیل» با استفاده از دو مدل برت و شبکه عصبی پیچشی نمایش داده شده‌است. در این جدول، مدل حاصل از شبکه پیچشی که از ورد۲وک^{۱۰} (میکلوف و دیگران، ۲۰۱۳) برای تعبیه واژه‌ها استفاده شده‌است به‌عنوان مدل پایه مورد استفاده و مقایسه قرار گرفته‌است. همچنین در شکل ۱-۵ ماتریس درهم‌ریختگی دو مدل «برت + پرسپترون ساده» و «برت + پیچشی» و مدل پایه نمایش داده شده‌است.

جدول ۴-۵: نتایج آزمایش مدل‌های ارائه‌شده بر روی دادگان لیار (۲ برجسب)

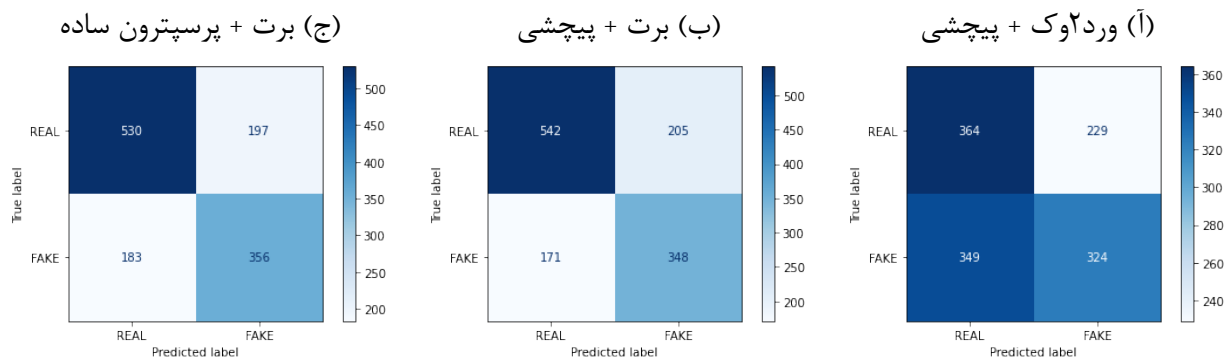
مدل	بازنمایی	فراخوانی	صحت	دقت	معیار اف
شبکه عصبی پیچشی	ورد۲وک	۵۸/۵۸	۴۸/۱۴	۵۴/۳۴	۵۲/۸۴
شبکه عصبی پیچشی	برت	۶۲/۹۲	۶۷/۰۵	۷۰/۳۰	۶۴/۹۱
پرسپترون ساده	برت	۶۴/۳۷	۶۶/۰۴	۶۹/۹۸	۶۵/۱۹

جدول ۵-۵: مقایسه مدل‌های ارائه‌شده با مدل‌های پایه براساس دادگان لیار (۶ برجسب)

مدل	بازنمایی	دقت تست	دقت اعتبارسنجی
ال.اس.تی.ام مکانیزم توجه (لانگ و دیگران، ۲۰۱۷)	ورد۲وک	۳۸/۵	۳۷/۸
شبکه کپسولی (گلدانی و دیگران، ۲۰۲۰)	ورد۲وک	۳۹/۵	۴۰/۹
شبکه عصبی پیچشی	ورد۲وک	۳۹/۳	۳۹/۸
	برت	۴۴/۱	۴۵/۵
	برت	۴۵/۱	۴۸/۴
پرسپترون ساده			

همان‌طور که در جدول ۴-۵ مشاهده می‌شود، مدل پرسپترون ساده با استفاده از بازنمایی تولیدشده توسط مدل برت نتیجه بهتری را برای این مجموعه دادگان داشته و با دقت بالاتری اخبار جعلی را

¹⁰Word2vec



شکل ۵-۱: ماتریس درهم‌ریختگی نتایج آزمایش‌های دادگان لیار

تشخیص داده‌است. در این مدل، میزان خطای مثبت کاذب که نشانگر تعداد اخبار جعلی است که توسط سیستم به خطا اصیل تشخیص داده شده‌است کمتر از سایر مدل پیشنهادی ما می‌باشد. برای مقایسه مدل پیشنهادی با سایر مدل‌های مطرح و به روز، نتایج آزمایش مدل‌های این طرح با مدل‌هایی که پیش‌تر برای دسته‌بندی اخبار دادگان لیار با ۶ برچسب ارائه شده است در جدول ۵-۷ نمایش داده شده است. با توجه به نتایج، مدل پرسپترون ساده با استفاده از بازنمایی برت نسبت به دیگر مدل‌ها با دقت بالاتری اخبار لیار را دسته‌بندی کرده است.

برای بررسی خطا در مدل‌های ارائه شده، در ادامه دو نمونه از اخباری که به اشتباه در آزمایش مدل «برت + پرسپترون» دسته‌بندی شده‌است، مورد بررسی قرار می‌گیرد. ابتدا سابقه گوینده‌ی هر دو خبر در جدول ۵-۶ نمایش داده شده است. در ادامه جدول ۵-۷ آمار واژگان نمونه (۱) را نمایش می‌دهد که یک خبر جعلی است و به اشتباه در دسته اخبار اصیل قرار گرفته‌است. همچنین جدول ۵-۸ آمار واژگان نمونه (۲) را نشان می‌دهد که یک خبر اصیل است و به اشتباه در دسته جعلی قرار گرفته‌است.

(۱)

Citizens Property Insurance has over \$500 billion worth of risk, with less than \$10 billion worth of surplus.

(۲)

Says the unemployment rate for college graduates is 4.4 percent and over 10 percent for noncollege-educated.

جدول ۵-۶: آمار تاریخیچه گوینده دو نمونه خبر ۱ و ۲

نمونه	به‌سختی درست	جعلی	دروغ	عمدتاً درست	نیمه درست
نمونه ۱	۲۸	۲۳	۷	۳۴	۳۸
نمونه ۲	۱۲	۱۶	۵	۷	۱۳

جدول ۵-۷: آمار واژگان یک نمونه ۱ (خبر جعلی دسته‌بندی شده در دسته خبر اصیل)

واژه هدف	تعداد واژه در اخبار جعلی	تعداد واژه در اخبار اصیل
citizen	۳۵	۴۰
properti	۳۷	۴۳
insur	۹۸	۱۲۱
billion	۱۹۳	۲۲۹
worth	۳	۲۵
risk	۱۳	۱۵
less	۵۳	۱۹۷
surplu	۱۳	۱۹

جدول ۵-۸: آمار واژگان نمونه ۲ (خبر اصیل دسته‌بندی شده در دسته خبر جعلی)

واژه هدف	تعداد واژه در اخبار جعلی	تعداد واژه در اخبار اصیل
say	۱۲۲۲	۱۲۸۱
unemploy	۶۵	۱۲۲
rate	۱۲۱	۲۸۴
colleg	۳۵	۹۱
graduat	۱۳	۵۱
percent	۳۵۱	۸۳۶
noncollegeeduc	۰	۱

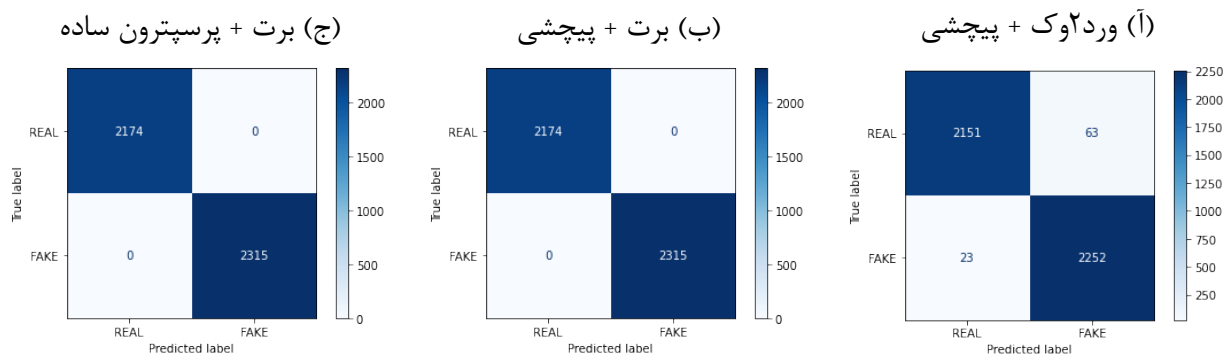
باتوجه به آمار واژگان و همچنین سابقه گوینده می‌توان نتیجه گرفت که هم سابقه فرد و هم واژگان به‌کاررفته در این دو نمونه خبر ماهیتی متفاوت از برجسب صحیح را دارد و شباهت اطلاعات موجود در واژگان و سابقه گوینده به دسته دیگر سبب خطای مدل پیشنهادشده ما شده‌است.

در آزمایش با دادگان ای.آس.ا.تی. نیز از ویژگی متن خبر به تنهایی استفاده شده‌است که نتایج مقایسه آن با مدل های پیشین [ایساتراور و دیگران \(۲۰۱۷\)](#) و مدل شبکه کپسولی ([گلدانی و دیگران، ۲۰۲۰](#)) در جدول ۵-۹ گزارش شده‌است. در این دادگان نیز مدل شبکه پیچشی با استفاده از ورد۲وک به‌عنوان مدل پایه به‌کار رفته‌است. طبق نتایج گزارش شده در این جدول، دو مدل پرسپترون ساده و شبکه عصبی

پیچشی که از بازنمایی مدل برت استفاده کردند دقت بالاتری را نسبت به مدل‌های پیشین ثبت کردند. در شکل ۲-۵ نیز ماتریس درهم‌ریختگی حاصل از ارزیابی مدل‌های ارائه شده بر روی دادگان آی.اس.اُ.تی نمایش داده شده است. همان‌طور که در نتایج مشاهده می‌شود دقت هر دو مدل مبتنی بر برت به ۱۰۰٪ رسیده است و هیچ خطای مثبت کاذب و منفی کاذب در نتایج وجود ندارد.

جدول ۵-۹: مقایسه مدل‌های ارائه شده با مدل‌های پیشین بر روی دادگان آی.اس.اُ.تی.

مدل	بازنمایی	دقت
بردار ماشین پشتیبان بردار ماشین پشتیبان خطی کا نزدیک‌ترین همسایه درخت تصمیم گرادین کاهشی تصادفی رگرسیون خطی شبکه کپسولی	ورد ۲وک	۸۶ ۹۲ ۸۳ ۸۹ ۸۹ ۸۹ ۹۹/۸
شبکه عصبی پیچشی	ورد ۲وک	۹۸/۸
	برت	۱۰۰
پرسپترون ساده	برت	۱۰۰



شکل ۲-۵: ماتریس درهم‌ریختگی نتایج آزمایش‌های دادگان آی.اس.اُ.تی.

همان‌طور که از نتایج ارائه شده بر روی دادگان لیار و آی.اس.اُ.تی مشخص است، هر دو مدل «برت + پرسپترون ساده» و «برت + پیچشی» نتایج قابل قبولی نسبت به روش‌های پیشین ارائه داده است. استفاده از لایه‌های پیچشی پس از بازنمایی ارائه شده توسط مدل برت به منظور به دست آوردن بازنمایی سطح بالاتر با توجه به هم‌نشینی واژه‌ها در جمله بوده است؛ اما طبق نتایج به دست آمده توسط مدل پرسپترون ساده، بازنمایی‌های حاصل از مدل برت به میزان قابل قبولی دارای ویژگی‌های مناسبی برای نمایندگی واژه‌ها و جملات است. علاوه بر این، از مقایسه مدل «برت + پیچشی» و مدل «ورد ۲وک + پیچشی» که تنها از

بازنمایی حاصل از همنشینی واژگان استفاده می‌کند، می‌توان نتیجه گرفت که قدرت مدل‌های ارائه‌شده در این بخش از پروژه حاصل از توانایی زیاد مدل برت برای بازنمایی جملات و کلمات است.

فصل ششم

بهبود تشخیص اخبار جعلی با فراداده‌های حاصل از پردازش زبان

۱-۶ انگیزه

در بسیاری از موارد صحت‌سنجی یک خبر با استفاده از متن خبر کار بسیار دشوار و پیچیده‌ای است. اما بعضی از فراداده‌های مرتبط با اخبار می‌تواند در تشخیص موثق بودن این دسته از اخبار به ما کمک کند. در این فاز، به بررسی تأثیر استفاده از سه فراداده موضوع خبر، احساس خبر و موجودیت‌های نامدار موجود در خبر می‌پردازیم. با تحلیل و بررسی اخبار جعلی می‌توانیم دریابیم که آیا بعضی از موضوعات مثلاً سیاسی یا اجتماعی نسبت به موضوعات دیگر مانند ورزشی بیشتر در معرض جعل خبر قرار دارد و خبر جعلی بیشتر در این موضوع مشخص تولید می‌شود یا خیر. دلیل این موضوع را می‌توان به دلیل تأثیرگذاری زیاد اخبار سیاسی و اجتماعی مطرح کرد. در صورت تأیید این فرضیه، انتظار می‌رود با دانستن موضوع هر خبر، بتوانیم در کنار متن خبر، با دقت بیشتری میزان جعلی بودن آن خبر را مشخص کنیم. علاوه بر این، یافتن موجودیت‌های نامدار و حس متن در هر خبر نیز می‌تواند اطلاعات نسبتاً مناسبی در مورد میزان صحت آن خبر ارائه دهد. در صورت تأیید این فرضیه نیز می‌توان حدس زد وجود اسامی خاص افراد، سازمان‌ها، مکان‌ها و ... در متن و یا حس مثبت یا منفی در متن می‌تواند به تشخیص جعلی بودن آن کمک کند. در ادامه این فصل به پردازش‌های متنی داده‌های خبری شامل دسته‌بندی موضوعی، تشخیص موجودیت‌های نامدار و تحلیل احساسات می‌پردازیم و در نهایت تأثیر آن‌ها را بررسی می‌نماییم.

۲-۶ استخراج ویژگی‌های موضوعی مستندات

یکی از ویژگی‌های مهم هر خبر، موضوع آن خبر است که کمک زیادی در تشخیص جعلی یا اصیل بودن آن می‌کند. برای مثال، عمده اخبار جعلی در اخبار سیاسی است که با استفاده از ویژگی موضوع خبر می‌توانیم با دقت بیشتر جعلی بودن خبر را تشخیص دهیم. دو رویکرد برای بهره‌گیری از اطلاعات مربوط به موضوع اخبار فارسی در این پروژه مورد بررسی قرار گرفته است:

- دسته‌بندی موضوعی با استفاده از روش‌های یادگیری با نظارت
- مدل‌سازی موضوعی با استفاده از روش‌های یادگیری بدون نظارت

۶-۲-۱ دسته‌بندی بانظارت براساس ۶ دسته مختلف

موضوعات اخبار به دو صورت مشخص می‌شود. در یک حالت اخبار به صورت کلی به چند دسته اصلی تقسیم می‌شود و سپس در سطح دوم هر یک از این دسته‌ها به چند زیر دسته تقسیم می‌گردد. یکی از دسته‌بندی‌های استاندارد که در این زمینه وجود دارد، دسته‌بندی ارائه‌شده در پیکره همشهری است (آل‌احمد و دیگران، ۲۰۰۹) که حاوی ۶ دسته کلی اخبار علمی و دانشگاهی، فرهنگی و هنری، سیاسی و اقتصادی، اجتماعی، بین‌المللی و ورزشی در این داده موجود است. برای دسته‌بندی اخبار بر اساس این ۶ دسته می‌توانیم از یک روش یادگیری بانظارت استفاده کنیم. همانند هر مسئله دسته‌بندی، مجموعه‌ای از اخبار شامل متن خبر و موضوع آن را به الگوریتم یادگیری می‌دهیم و پس از اتمام یادگیری، با استفاده از این مدل می‌توانیم موضوع اخبار جدید را تشخیص دهیم و از آن به‌عنوان ویژگی‌ای اضافه‌شده به بردار تعبیه متن خبر، استفاده کنیم. برای این منظور، در این پژوهش از دسته‌بندی با نظارت با استفاده از مدل زبانی برت و برای آموزش مدل دسته‌بند از مجموعه اخبار همشهری که آمار آن در جدول ۶-۱ آمده، استفاده کرده‌ایم. پس از آموزش دسته‌بند، عملیات دسته‌بندی می‌تواند بر روی هر خبری انجام شود و حاصل این دسته‌بندی به‌صورت یک بردار تک‌روشن‌شش‌بُعدی نمایش داده شود تا از آن به‌عنوان ویژگی در آموزش مدل‌های تشخیص اخبار جعلی بهره برده شود.

جدول ۶-۱: آمار و اطلاعات مربوط به دادگان همشهری

ویژگی	مقدار
حجم دادگان	۵۶۴ مگابایت
نوع اسناد	متن
تعداد اسناد	۱۶۶,۷۷۴
تعداد واژه‌های یکتا	۴۱۷,۳۳۹
میانگین طول سند	۳۸۰ واژه
تعداد دسته‌ها	۸۲
تعداد موضوعات	۶۵

۶-۲-۲ دسته‌بندی بدون نظارت براساس روش تخصیص نهان دیریکله

همانطور که پیش‌تر در بخش ۴-۳-۳ اشاره شد، روش تخصیص نهان دیریکله (بلی و دیگران، ۲۰۰۳) یک روش بدون نظارت بر مبنای مدل‌های احتمالاتی است که در آن با دریافت پیکره اخبار، دسته‌های موضوعات مستندات را انتخاب می‌کند. هر دسته باتوجه به هم‌نشینی واژه‌ها در پیکره متن مشخص

می‌شود. بنابراین نماینده هر موضوع مجموعه‌ای از واژه‌ها خواهد بود که باید برچسب آن‌ها مشخص گردد. این روش ۲ ماتریس خروجی تولید می‌کند:

- ماتریس واژه-موضوع، که نشان می‌دهد در هر موضوع استنتاج شده چه واژه‌هایی قرار دارد.
 - ماتریس سند-موضوع، که نشان‌دهنده آن است که هر سند چه سهمی از موضوعات مختلف را دارد. بیشترین سهم نشان‌دهنده موضوع اصلی آن سند خواهد بود.
- یکی از چالش‌های اصلی این روش پیدا کردن مقدار بهینه برای تعداد موضوعات است؛ چراکه اگر این مقدار خیلی زیاد شود، موضوعات نسبتاً مرتبط از یکدیگر جدا خواهند شد و خطای احتمالی برای پیش‌بینی یک خبر جدید بیشتر می‌شود. از طرف دیگر، اگر تعداد موضوعات خیلی کم انتخاب شود، موضوعات نامرتب دارای یک برچسب موضوعی یکسان خواهد شد. یکی از راه‌های پیدا کردن مقدار بهینه برای موضوعات اخبار در کنار داشتن آگاهی از مقدار حدودی تعداد موضوعات معمول در خبرگزاری‌ها، استفاده از تعدادی است که خطای کلی مدل به کمترین مقدار برسد. برای این منظور، چند عدد متفاوت برای مدل‌سازی موضوعی انتخاب خواهد شد و با بررسی خروجی مدل و ارزیابی کیفی، به‌صورت تجربی تعداد بهینه موضوعات مشخص می‌گردد.

۳-۶ تشخیص موجودیت‌های نامدار

یکی از عملیات‌های پایه در پردازش زبان طبیعی شناسایی موجودیت‌های نامدار است که در آن تمامی اسامی خاص موجود در متن استخراج می‌گردد و به گروه‌های از پیش تعریف‌شده‌ای مانند اسم افراد، سازمان‌ها، مکان‌ها و ... دسته‌بندی می‌شود. شناسایی موجودیت‌های نامدار در بسیاری از عملیات‌ها مانند جستجوهای معنادار^۱، ترجمه ماشینی^۲، استخراج اطلاعات از متن^۳، مرجع‌یابی در متن^۴، سیستم‌های پرسش و پاسخ^۵، سیستم‌های خبره^۶، کشف دانش^۷، مدیریت دانش^۸، تحلیل احساسات^۹، بازیابی

¹Semantic Search

²Machine Translation

³Information Extraction

⁴Co-reference Resolution

⁵Question Answering Systems

⁶Expert Systems

⁷Knowledge Discovery

⁸Knowledge Management

⁹Sentiment Analysis

اطلاعات^{۱۰}، تحلیل خبر^{۱۱} و بسیاری دیگر از شاخه‌های مرتبط با پردازش زبان طبیعی کاربرد دارد. به منظور استخراج موجودیت‌های نامدار دو روش قاعده‌مند و آماری وجود دارد (جورافسکی و مارتین، ۲۰۰۹). در روش‌های قاعده‌مند، براساس قوانین ازپیش مشخص‌شده موجودیت‌های نامدار تشخیص داده می‌شوند. در روش آماری، از با استفاده از مفاهیم یادگیری ماشین به دسته‌بندی موجودیت‌های نامدار پرداخته می‌شود. در روش دسته‌بندی بانظارت، از پیکره‌هایی که موجودیت‌های نامدار مشخص دارند استفاده می‌شود و مدل یاد می‌گیرد تا چگونه دسته‌بندی موجودیت‌های نامدار جدید را انجام دهد. موجودیت‌های نامدار می‌توانند نقش مهمی را در تشخیص اخبار جعلی ایفا کنند چراکه اخبار جعلی توزیع متفاوتی از اسامی خاص مانند افراد را نسبت به اخبار اصیل می‌توانند داشته باشند. در این پروژه ما از یک مدل آموزش دیده که به صورت بانظارت آموزش داده شده است، استفاده کردیم.

عبدالله پور (۲۰۲۰) با استفاده از مدل ایکس.ال.ام-روبرتا به نتایج قابل قبولی در تشخیص موجودیت‌های نامدار رسیده است و ما برای استخراج موجودیت‌های نامدار هر خبر از این مدل استفاده می‌کنیم. مسئله پر اهمیت در آموزش مدل، داده‌های یادگیری هستند. بسیاری از پیکره‌های معرفی شده در زبان فارسی تنها موجودیت‌های نامدار مربوط به نام اشخاص، موقعیت‌های مکانی و اسامی سازمان‌ها را شامل می‌شوند. به منظور استخراج موجودیت‌های بیشتر در متن ورودی، برای آموزش این مدل از پیکره موجودیت‌های نامدار تهیه‌شده در آزمایشگاه پردازش زبان طبیعی دانشگاه صنعتی امیرکبیر که شامل ۱۵ موجودیت نامدار است، استفاده شده است (ممتازی و ترابی، ۲۰۲۰). این دادگان شامل ۳۱ برچسب برای مشخص کردن ۱۵ نوع موجودیت متفاوت است که شامل اسم شخص مفرد، اسم شخص جمع، موقعیت مکانی، اسم سازمان، زبان، ملیت، رخداد، شغل، کتاب، اسم فیلم، تاریخ، مذهب، عنوان علمی و دانش، روزنامه و سایر اسامی خاص ذکر نشده در زبان فارسی می‌باشد. جدول ۶-۲ به تفصیل به شرح و بررسی برچسب‌های استفاده‌شده برای دادگان می‌پردازد.

۴-۶ تحلیل احساسات

نظرکاوی یا تحلیل احساسات، مطالعه محاسباتی نظرات کاربران حول محصولات، خدمات، سازمان‌ها، اشخاص، رویدادها و جنبه‌های مختلف آن است. در سال‌های اخیر این حوزه یکی از حوزه‌های پژوهشی فعال در پردازش زبان‌های طبیعی بوده است. نظرکاوی در سطوح مختلف سند، جمله و جنبه مورد

¹⁰Information Retrieval

¹¹News Analysis

جدول ۶-۲: راهنمای برچسب واژگان

برچسب	تعریف انگلیسی برچسب	تعریف فارسی برچسب	توضیحات
O	Out	هیچ	واژه مورد نظر از موجودیت‌های نامدار نیست.
PEI	Person Individual	شخص مفرد	این علامت به نام شخص اشاره دارد.
PEG	Person Group	اسم خاص گروه	این علامت به موجودیت نامداری اشاره دارد که به صورت جمع آمده است.
LOC	Location	موقعیت مکانی	واژه مورد نظر این برچسب به موقعیت مکانی خاصی اشاره دارد.
ORG	Organization	سازمان	اسامی سازمان‌ها و مؤسسات مختلف با این برچسب نشان داده می‌شود.
LAN	Language	زبان	این برچسب نشان‌دهنده زبان‌های مختلف است.
NAT	Nationality	ملیت	ملیت‌های مختلف با این علامت تعیین می‌شود.
EVN	Events	رخدادها	رخدادها و وقایع خاص را با این برچسب نشان دادیم.
JOB	Job	شغل	این برچسب نمایانگر مشاغل است.
BOK	Book	کتاب	اسامی کتاب‌ها با این برچسب مشخص می‌شود.
FLM	Film	فیلم	نام فیلم‌های مختلف با این برچسب تعیین می‌شود.
DTE	Date	تاریخ	برای نشان دادن تاریخ و دوره‌های مختلف از برچسب استفاده شده است.
REL	Religion	مذهب	این برچسب تعیین‌کننده مذاهب مختلف است.
FLD	Field	زمینه	زمینه‌ها و دانش‌های مختلف با این برچسب تعیین گردیده است.
MAG	Magazine	روزنامه و مجله	اسامی روزنامه‌ها و مجلات با این برچسب آمده است.
OTH	Other	سایرین	اگر واژه‌ای به عنوان موجودیت نامدار باشد ولی در بین موجودیت‌های معرفی شده در بالا نباشد با این برچسب مشخص می‌شود.

مطالعه قرار گرفته‌است. در حوزه کالا و خدمات هنگام تصمیم‌گیری عقاید دیگران تأثیر به‌سزایی در تصمیم نهایی دارد. در دنیای واقعی، شرکت‌ها همواره نیازمند عقاید مردم برای بهبود کیفیت و توسعه خدمات خود است. در دنیای نشر خبر، سوگیری احساسی اخبار یکی از جنبه‌های مؤثر در تحلیل خبر می‌باشد. مثبت یا منفی بودن یک خبر می‌تواند جنبه‌های مختلفی از تحلیل آن خبر را فراهم آورد. در حال حاضر داده‌های متنی بخش عظیمی از شبکه‌های اجتماعی، فروشگاه‌های اینترنتی و انواع مختلف سامانه‌های ارتباطی را در بر گرفته‌است. از این داده‌ها می‌توان در راستای درک احساسات کاربران به مطالب مختلف مانند یک محصول، یک نوشته موضوعی و ... استفاده کرد.

نظرکاوی در سال ۲۰۰۲ توسط **پنگ و لی (۲۰۰۲)** معرفی شد. با روی کارآمدن شبکه‌های اجتماعی و تأثیر روزافزون آن‌ها بر زندگی مردم، اهمیت نظرکاوی و کاربردهای آن بیشتر مشخص گردید. با توجه به این که افراد با حضور در این شبکه‌ها به بیان نظرات، عقاید و دیدگاه‌های خود درباره کلیه مسائل سیاسی، اجتماعی، فرهنگی، اقتصادی و حتی فردی می‌پردازد، تحلیل احساسات بیشتر از هر چیز در شبکه‌های اجتماعی مورد توجه قرار گرفته‌است. در کنار بستری که در شبکه‌های اجتماعی فراهم شده‌است، متون موجود در داده‌های خبری نیز می‌توانند دارای حس مثبت یا منفی باشد که به تحلیل آن کمک می‌نماید. برای استفاده از تحلیل احساسات در تشخیص اخبار جعلی، تحلیل احساس در سطح سند مورد توجه این پژوهش قرار گرفته‌است و با بهره‌گیری از یک مدل مبتنی بر یادگیری عمیق حس موجود در متن به صورت یک بردار دو بعدی برای حس مثبت و منفی مشخص می‌گردد. خروجی این مدل به عنوان یکی از ویژگی‌های ورودی در مدل تشخیص اخبار جعلی استفاده خواهد شد.

برای پیاده‌سازی بخش تشخیص احساسات از شبکه پیچشی استفاده شده‌است. با توجه به این که پایه اصلی تحلیل احساسات دسته‌بندی متون است، ساختار شبکه پیچشی مورد استفاده در این بخش مشابه شبکه پیچشی‌ای است که در تشخیص خبر جعلی مورد استفاده قرار گرفته است و توضیحات آن در بخش ۴-۴-۲ آمده است. بازنمایی مورد استفاده در این شبکه نیز بازنمایی ایکس.ال.ام-روبرتا می‌باشد.

۵-۶ نتایج حاصل از استفاده از پردازش‌های متنی در تشخیص اخبار

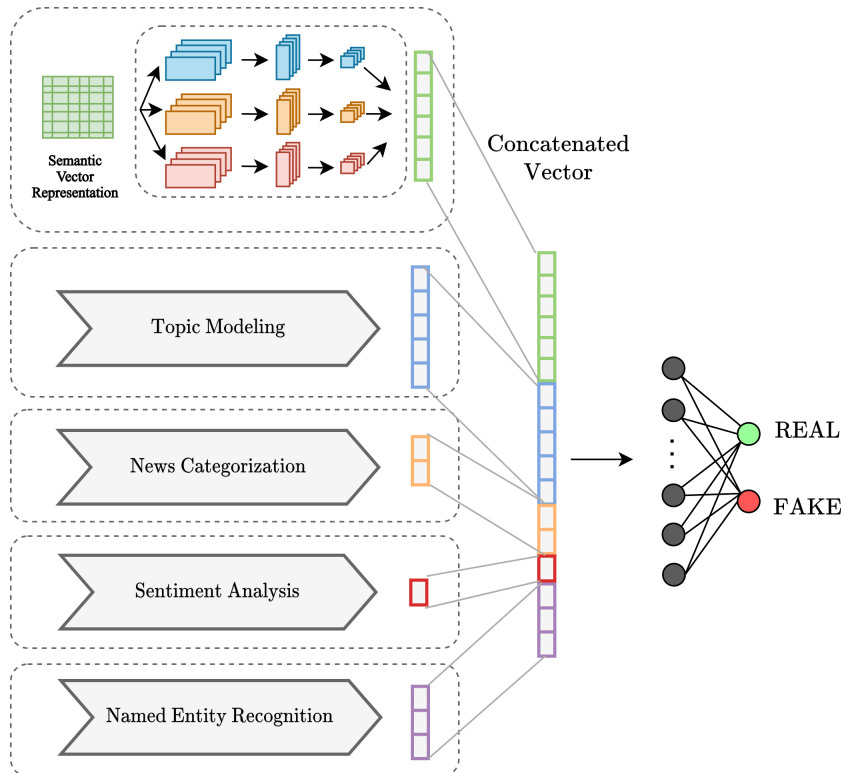
در فصل گذشته، با بررسی و مقایسه مدل‌های از پیش آموزش دیده، بهترین مدل بازنمایی اخبار ورودی را انتخاب کردیم. در این بخش می‌خواهیم با استفاده از هر دو مدل دسته‌بند شبکه پیشرو پرسپترون و شبکه عصبی پیچشی و استفاده از مدل پارس‌برت برای بازنمایی اخبار، از فراداده‌هایی مانند موضوع خبر، موجودیت‌های نامدار و یا حس متن در کنار بردار بازنمایی متن خبر استفاده کنیم.

برای استفاده از ویژگی موضوع خبر در رویکرد اول یک بردار ویژگی ۶ بعدی حاصل از دسته‌بندی موضوعی اخبار و در رویکرد دوم یک بردار ویژگی ۵۰ بعدی حاصل از تخصیص نهان دیریکله خواهیم داشت. بازنمایی خروجی تحلیل احساس نیز یک بردار دو بعدی است که هر درایه آن نشان‌دهنده مثبت و یا منفی بودن آن خبر است. همچنین، اطلاعات مربوط به موجودیت‌های نامدار موجود در یک خبر با استفاده از یک بردار شامل ۱۵ ویژگی بازنمایی می‌شود. برای نمونه، در جدول ۳-۶ نمونه‌ای از یک جمله برچسب‌گذاری شده نمایش داده شده‌است. در بردار ۱۵ تایی حاصل از این جمله، درایه مربوط به موجودیت کتاب مقدار ۲، درایه‌های مربوط به موجودیت‌های زبان، شخص مفرد، شغل و تاریخ مقدار ۱ و سایر درایه‌ها مقدار صفر خواهد داشت.

جدول ۳-۶: مثال از یک جمله برچسب‌خورده توسط مدل تشخیص موجودیت نامدار

برچسب	واژه
b-BOK	کتاب
i-BOK	گلستان
O	و
b-BOK	کتاب
i-BOK	بوستان
O	از
O	شاهکارهای
O	ادب
b-LAN	فارسی
O	و
O	اثر
b-PEI	سعدی
b-JOB	شاعر
b-DTE	قرن
i-DTE	ششم
i-DTE	هجری
O	است

به‌منظور استفاده از ویژگی‌های فراداده‌های مختلف، بردارهای این ویژگی‌ها را ابتدا به یک لایه چگال با همان تعداد نورون وارد می‌کنیم و پس از آن بردار خروجی لایه چگال را به انتهای بردار ویژگی‌های نهایی استخراج‌شده از متن اخبار الحاق می‌کنیم. درنهایت پس از ساخت بردار ویژگی جدید، شامل ویژگی‌های متن و فراداده‌ها، آن را وارد یک لایه پرسپترون می‌کنیم تا برچسب اخبار مشخص شود. شمای کلی مدل ارائه‌شده برای استفاده از تمام فراداده‌ها در کنار ویژگی‌های مبتنی بر بافت اخبار در شکل ۱-۶ ارائه شده است.



شکل ۶-۱: شمای کلی مدل ارائه شده برای به کار بردن تمام فراداده‌ها و دسته‌بندی اخبار با استفاده از دسته‌بند پیچشی

جدول ۶-۴ نتایج حاصل از استفاده از هریک از فراداده‌های ذکر شده در کنار بازنمایی متنی در دو دسته‌بند پیچشی و پیشرو را نمایش می‌دهد. دسته‌بند پیچشی ابتدا با استفاده از ماتریس دنباله‌ای مدل پارس‌برت برای متن ورودی، ویژگی‌های سطح بالای مبتنی بر متن را استخراج می‌کند و پس از بهم پیوستن فراداده‌ها به یکدیگر، یک بردار نهایی وارد یک لایه پرسپترون ساده می‌شود تا پیش‌بینی صحت خبر انجام شود. در حالت استفاده از دسته‌بند پرسپترون ساده، خروجی تجمعی مدل پارس‌برت استفاده شد و فراداده‌ها به انتهای بردار نماینده متن ورودی اضافه می‌شوند و بردار نهایی به یک لایه چگال وارد می‌شود تا عمل دسته‌بندی انجام شود. برای مقایسه راحت‌تر، نتایج حاصل از تشخیص اخبار بدون فراداده که در فصل قبل گزارش شده بود در این جدول تکرار شده است. براساس جدول ۶-۴، با استفاده از پردازش‌های متنی توانستیم دقت تشخیص اخبار جعلی را بهبود دهیم. همانطور که انتظار داشتیم دسته‌بند پیچشی توانست عملکرد بهتری را نسبت به مدل شبکه عصبی پرسپترون پیشرو داشته باشد. با توجه به نتایج، ارزشمندترین فراداده که توانست دقت تشخیص مدل ما را بیشتر بهبود دهد، موضوع هر خبر (دسته‌بندی) بود که باعث افزایش ۱/۲٪ معیار اف و ۱/۶٪ دقت دسته‌بند پیچشی شد. این نتیجه دور از انتظار هم نبود چراکه موضوع هر خبر تاثیر بسیار زیادی در تشخیص اخبار جعلی دارد.

جدول ۴-۶: نتایج تشخیص اخبار جعلی فارسی با استفاده از بازنمایی پارس‌برت و دسته‌بندی‌های مختلف

بازنمایی	دسته‌بند	فراداده	فراخوانی	صحت	معیار اف	دقت
پارس‌برت	پیچشی	-	۸۹/۱۳	۹۳/۷۱	۹۱/۳۶	۹۱/۶۴
		موجودیت‌های نامدار	۹۱/۱۶	۹۴/۲۸	۹۲/۶۹	۹۲/۹۹
		احساس خبر	۹۱/۵۲	۹۲/۵۷	۹۲/۰۴	۹۲/۴۵
		موضوع (دسته‌بندی)	۹۶/۸۷	۸۸/۵۷	۹۲/۵۳	۹۳/۲۶
		موضوع (تخصیص نهان دیریکله)	۹۱/۴۲	۹۱/۴۲	۹۱/۴۲	۹۱/۹۱
	پرسپترون	تمام ویژگی‌ها	۹۳/۶۴	۹۲/۵۷	۹۳/۱۰	۹۳/۵۳
		-	۹۲/۴۴	۹۰/۸۵	۹۱/۶۴	۹۲/۱۸
		موجودیت‌های نامدار	۹۳/۱۶	۸۵/۷۱	۸۹/۲۸	۹۰/۲۹
		احساس خبر	۸۳/۹۰	۹۸/۲۸	۹۰/۵۲	۹۰/۲۹
		موضوع (دسته‌بندی)	۸۹/۱۳	۹۳/۷۱	۹۱/۳۶	۹۱/۶۴
		موضوع (تخصیص نهان دیریکله)	۸۷/۸۹	۹۵/۴۲	۹۱/۵۰	۹۱/۶۴

همانطور که پیش از این هم اشاره شده، عمده اخبار جعلی در حوزه‌های سیاسی و اجتماعی منتشر می‌شوند. همچنین فراداده‌های مربوط به احساس خبر، بردار موجودیت‌های نام‌دار و موضوع (تخصیص نهان دیریکله) توانستند به طور مجزا دقت نهایی مدل را ۵٪، ۳۵٪ و ۳٪ افزایش دهند. در نهایت با استفاده از تمامی فراداده‌های استخراج شده از اخبار که با پیوستن به یکدیگر یک بردار ۷۳ بعدی را تشکیل دادند، مدل پارس‌برت-پیچشی ارائه شده را ارزیابی کردیم که باعث افزایش ۱/۷۴ درصدی معیار اف و ۱/۸۹ درصدی دقت شد.

فصل هفتم

جمع‌بندی و کارهای آتی

۱-۷ جمع‌بندی و نتیجه‌گیری

با گسترش روزافزون شبکه‌های اجتماعی و استفاده همگانی از آن‌ها، حجم قابل توجهی از اطلاعات در میان مردم منتقل می‌شود. این انتقال سریع و گسترده اطلاعات در فضای مجازی، عموم مردم را در معرض دریافت اطلاعات غلط و اخبار جعلی قرار می‌دهد. انتشار اخبار جعلی اثرات منفی جبران‌ناپذیری را بر روی افکار عمومی و حتی سیاست‌های کلان دولت‌ها دارد. امروزه با گسترش استفاده از هوش مصنوعی و پردازش زبان طبیعی، محققان پژوهش‌های بسیاری را در این حوزه با روش‌های گوناگون انجام داده‌اند. در این پژوهش قدم‌های اولیه برای ایجاد یک سامانه جامع تشخیص اخبار جعلی با استفاده از مفاهیم به‌روز پردازش زبان طبیعی برداشته شد. با توجه به کمبود مجموعه دادگان منسجم اخبار جعلی به زبان فارسی، در گام اول ما یک روش ۵ مرحله‌ای برای استخراج این اخبار ارائه کردیم. سپس با استفاده از این روش، مجموعه دادگانی به نام «تاج» برای پژوهش‌های آتی و استفاده از آن برای آموزش مدل‌های اولیه ارائه کردیم.

تشخیص اخبار جعلی یک مسئله پیچیده است که حتی انسان به طور معمول با دیدن متن یک خبر به دشواری می‌تواند در مورد جعلی بودن آن اظهار نظر کند. به همین دلیل، برای پیاده‌سازی یک سامانه مقاوم در برابر خطا، باید از فراداده‌های زیادی در کنار متن خبر استفاده کنیم. در این پژوهش ما از فراداده‌های سطح بالای مبتنی بر متن اخبار مانند احساس خبر، موضوعات خبر، دسته خبر و موجودیت‌های نامدار خبر استفاده کردیم و توانستیم بهتر از قبل اخبار جعلی را تشخیص دهیم.

۲-۷ کارهای آتی

همانطور که اشاره شد، تشخیص اخبار جعلی نیازمند اطلاعات بسیار زیادی می‌باشد. این اطلاعات به متن اخبار محدود نمی‌شود بلکه شامل ویژگی‌های زمانی، پروفایل کاربری و تاریخچه انتشار خبر می‌باشد. یکی از منابع غنی و قابل استفاده برای آموزش یک مدل هوشمند برای تشخیص اخبار جعلی، شبکه‌های اجتماعی مانند توئیتر می‌باشد چراکه این شبکه‌ها فراداده‌های بیشتری نسبت به متن خبر در اختیار مدل ما می‌گذارد که باعث می‌شود مدل هوشمند، ماهیت خبر جعلی را به درستی بشناسد و بتواند روی اخبار جدید عملکرد بهتری داشته باشد.

منابع و مراجع

- Abdallah, Z. S., M. Carman, and G. Haffari (2017). Multi-domain evaluation framework for named entity recognition tools. *Computer Speech Language* 43, 34 – 55.
- Abdolahpour, M. M. (2020). Implementation of basic text processing for persian language.
- Ahmed, H., I. Traore, and S. Saad (2017). Detection of online fake news using n-gram analysis and machine learning techniques. In *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*, pp. 127–138. Springer.
- AleAhmad, A., H. Amiri, E. Darrudi, M. Rahgozar, and F. Oroumchian (2009). Hamshahri: A standard persian text collection. *Knowledge-Based Systems* 22(5), 382–387.
- Alkhair, M., K. Meftouh, K. Smaïli, and N. Othman (2019). An arabic corpus of fake news: Collection, analysis and classification. In *International Conference on Arabic Language Processing*, pp. 292–302. Springer.
- Allport, G. W. and L. Postman (1947). *The psychology of rumor*. Henry Holt.
- Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan), 993–1022.
- Chen, Y., T. Lasko, Q. Mei, J. Denny, and H. Xu (2015, 09). A study of active learning methods for named entity recognition in clinical text. *Journal of biomedical informatics* 58.
- Conneau, A., K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave,

- M. Ott, L. Zettlemoyer, and V. Stoyanov (2019). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.
- Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Farahani, M. (2020). Albert-persian: A lite bert for self-supervised learning of language representations for the persian language. <https://github.com/m3hrdadfi/albert-persian>.
- Farahani, M., M. Gharachorloo, M. Farahani, and M. Manthouri (2020). Parsbert: Transformer-based model for persian language understanding. ArXiv abs/2005.12515.
- Ghayoomi, M. (2012, 01). Bootstrapping the development of an hpsg-based treebank for persian. Linguistic Issues in Language Technology 7(1).
- Goldani, M. H., S. Momtazi, and R. Safabakhsh (2020). Detecting fake news with capsule neural networks. arXiv preprint arXiv:2002.01030.
- Huang, Z., W. Xu, and K. Yu (2015). Bidirectional lstm-crf models for sequence tagging. arXiv preprint arXiv:1508.01991.
- Hussain, S., J. Kuli, and G. Hazarika (2016, 03). The first step towards named entity recognition in missing language. In Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT).
- Jahanbakhsh-Nagadeh, Z., M.-R. Feizi-Derakhshi, M. Ramezani, T. Rahkar-Farshi, M. Asgari-Chenaghlu, N. Nikzad-Khasmakhi, A.-R. Feizi-Derakhshi, M. Ranjbar-Khadivi, E. Zafarani-Moattar, and M.-A. Balafar (2020). A model to measure the spread power of rumors. arXiv preprint arXiv:2002.07563.
- Jurafsky, D. and J. H. Martin (2009). Speech and Language Processing (2nd Edition). USA: Prentice-Hall, Inc.

- Jwa, H., D. Oh, K. Park, J. M. Kang, and H. Lim (2019). exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). *Applied Sciences* 9(19).
- Kaliyar, R. K., A. Goswami, P. Narang, and S. Sinha (2020). Fndnet—a deep convolutional neural network for fake news detection. *Cognitive Systems Research* 61, 32–44.
- Khattar, D., J. S. Goud, M. Gupta, and V. Varma (2019). Mvae: Multimodal variational autoencoder for fake news detection. In *The World Wide Web Conference*, pp. 2915–2921.
- Kochkina, E., M. Liakata, and A. Zubiaga (2018). PHEME dataset for rumour detection and veracity classification.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105.
- Lazer, D. M., M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, et al. (2018). The science of fake news. *Science* 359(6380), 1094–1096.
- Le, H. T., C. Cerisara, and A. Denis (2017). Do convolutional networks need to be deep for text classification? *arXiv preprint arXiv:1707.04108*.
- Li, Q., Q. Zhang, L. Si, and Y. Liu (2019). Rumor detection on social media: Datasets, methods and opportunities. *arXiv preprint arXiv:1911.07199*.
- Liu, C., X. Wu, M. Yu, G. Li, J. Jiang, W. Huang, and X. Lu (2019). A two-stage model based on bert for short fake news detection. In *International Conference on Knowledge Science, Engineering and Management*, pp. 172–183. Springer.

- Liu, Z., S. Shabani, N. G. Balet, and M. Sokhn (2019). Detection of satiric news on social media: Analysis of the phenomenon with a french dataset. In 28th International Conference on Computer Communication and Networks, pp. 1–6.
- Long, Y., Q. Lu, R. Xiang, M. Li, and C.-R. Huang (2017, November). Fake news detection through multi-perspective speaker profiles. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Taipei, Taiwan, pp. 252–256. Asian Federation of Natural Language Processing.
- Lozhnikov, N., L. Derczynski, and M. Mazzara (2018). Stance prediction for russian: data and analysis. In International Conference in Software Engineering for Defence Applications, pp. 176–186. Springer.
- Ma, J., W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha (2016). Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, p. 3818–3824. AAAI Press.
- Ma, J., W. Gao, and K.-F. Wong (2017, July). Detect rumors in microblog posts using propagation structure via kernel learning. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, pp. 708–717. Association for Computational Linguistics.
- Mahmoodabad, S. D., S. Farzi, and D. B. Bakhtiarvand (2018). Persian rumor detection on twitter. In 2018 9th International Symposium on Telecommunications (IST), pp. 597–602. IEEE.
- Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pp. 3111–3119.
- Momtazi, S. and F. Torabi (2020). Named entity recognition in persian text using deep learning. Signal and Data Processing 16(4), 93–112.

- Pang, B., L. Lee, and S. Vaithyanathan (2002, July). Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pp. 79–86. Association for Computational Linguistics.
- Pennington, J., R. Socher, and C. D. Manning (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.
- Poostchi, H., E. Z. Borzeshi, M. Abdous, and M. Piccardi (2016). Personer: Persian named-entity recognition. In *COLING 2016-26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers*.
- Ramezani, M., M. Rafiei, S. Omranpour, and H. R. Rabiee (2019). News labeling as early as possible: real or fake? In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 536–537. IEEE.
- Reyes, J. and L. Palafox (2019). Detection of fake news based on readability. *Reunion Internacional de Inteligencia Artificial y sus Aplicaciones*.
- Rubin, V. L., Y. Chen, and N. K. Conroy (2015). Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology* 52(1), 1–4.
- Shahshahani, M. S., M. Mohseni, A. Shakery, and H. Faili (2019). Payma: A tagged corpus of persian named entities. *Signal and Data Processing* 16(1), 91–110.
- Shu, K., D. Mahudeswaran, S. Wang, D. Lee, and H. Liu (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data* 8(3), 171–188.

- Suleiman, D., A. Awajan, and N. Al-Madi (2017). Deep learning based technique for plagiarism detection in arabic texts. In 2017 International Conference on New Trends in Computing Sciences (ICTCS), pp. 216–222. IEEE.
- Tacchini, E., G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro (2017). Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506.
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, and I. Polosukhin (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17, Red Hook, NY, USA, p. 6000–6010. Curran Associates Inc.
- Vicario, M. D., W. Quattrociocchi, A. Scala, and F. Zollo (2019). Polarization and fake news: Early warning of potential misinformation targets. ACM Transactions on the Web (TWEB) 13(2), 1–22.
- Vogel, I. and P. Jiang (2019). Fake news detection with the new german dataset “germanfake-enc”. In International Conference on Theory and Practice of Digital Libraries, pp. 288–295. Springer.
- Wang, W. Y. (2017). ” liar, liar pants on fire”: A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648.
- Yang, S., K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu (2019). Unsupervised fake news detection on social media: A generative approach. In Proceedings of the AAAI Conference on Artificial Intelligence, Volume 33, pp. 5644–5651.
- Zamani, S., M. Asadpour, and D. Moazzami (2017). Rumor detection for persian tweets. In 2017 Iranian Conference on Electrical Engineering (ICEE), pp. 1532–1536. IEEE.
- Zarharan, M., S. Ahangar, F. S. Rezvaninejad, M. L. Bidhendi, S. S. Jalali, S. Eetemadi,

- M. T. Pilehvar, and B. Minaei-Bidgoli (2019). Persian stance classification dataset. In Proceedings of the Conference for Truth and Trust Online.
- Zhang, J., B. Dong, and S. Y. Philip (2020). Fakedetector: Effective fake news detection with deep diffusive neural network. In 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 1826–1829. IEEE.
- Zubiaga, A., M. Liakata, R. Procter, G. W. S. Hoi, and P. Tolmie (2016). PHEME rumour scheme dataset: journalism use case. PHEME rumour scheme dataset: journalism use case.