



دانشگاه کاشان
University of Kashan

مجله محاسبات نرم

SOFT COMPUTING JOURNAL

تارنمای مجله: scj.kashanu.ac.ir



تحلیل احساسات رمزارزها با یادگیری انتقالی شات صفر

کیا جهان‌بین¹، کاندید دکتری، محمد علی زارع چاهوکی^{*}، دانشیار

¹ گروه کامپیوتر و نرم‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران.

اطلاعات مقاله

چکیده

تاریخچه مقاله:

دریافت 10 مرداد ماه 1403

پذیرش 20 مهر ماه 1403

کلمات کلیدی:

تحلیل احساسات جنبه‌محور رمزارز

یادگیری عمیق شات صفر

شبکه‌های عصبی عمیق پیش‌آموزش دیده

لایه توجه

روش گریدی

نرخ یادگیری وفق‌پذیر

تحلیل احساسات جنبه‌محور توئیت‌ها با شبکه‌های عصبی عمیق بسیار مورد توجه است. با این حال، این شبکه‌ها برای کارایی مطلوب نیازمند دادگان غنی می‌باشند. بنابراین، در این مطالعه از یادگیری انتقالی عمیق شات صفر و دادگان میان‌دامنه‌ای برای افزایش غنای منابع آموزشی در دامنه هدف استفاده شده است. در این پژوهش، ابتدا توئیت‌های تاثیرگذاران رمزارزها جمع‌آوری شده و سپس، از ترکیب شبکه عصبی عمیق، لایه توجه و شبکه پیش‌آموزش دیده دستیلیرت برای تحلیل احساسات جنبه‌محور استفاده شده است. نقشه حرارتی لایه توجه نشان می‌دهد استفاده از این لایه بعد از مدل‌های عمیق توانسته به برجسته‌سازی کلمات جنبه کمک کند. در یادگیری انتقالی شات صفر، مدل هیچ نمونه‌ای از داده‌های هدف دارای برچسب را در اختیار ندارد و آموزش با دادگان میان‌دامنه‌ای انجام می‌شود. علاوه بر این، استفاده از دادگان میان‌دامنه‌ای و انتخاب بهینه آنها با از ضریب شباهت پیرسون در شرایط کمبود داده غنی سبب کاهش انتقال منفی شده و همچنین حساسیت مدل را به محتوی زمینه متن کمک می‌کند. همچنین، برای حل مشکل دادگان نامتوازن از رویکرد نمونه‌برداری بیش از حد اقلیت مصنوعی استفاده شده است. یافته‌های تجربی نشان می‌دهد، مدل در مقایسه با پژوهش‌های پیشین، به صورت میانگین دقت و F1 را بر روی مجموعه داده آزمون SemEval در حدود 2 درصد افزایش داده و میانگین نرخ ROC-AUC به 86/35 درصد رسیده است. همچنین، در این پژوهش از نرخ یادگیری وفق‌پذیر استفاده شده که سبب همگرایی سریع‌تر مدل می‌شود. انتخاب ابرپارامترهای شبکه نیز با روش گریدی صورت گرفته که موجب وفق‌پذیری بیشتر مدل با دامنه هدف شده است.

© 1403 نویسندگان. مقاله با دسترسی آزاد تحت مجوز CC-BY

1. مقدمه

پردازش زبان طبیعی مورد توجه قرار گیرد. با این حال، عملکرد مدل‌های یادگیری عمیق در پردازش زبان طبیعی یا در سایر زمینه‌های داده‌کاوی بسیار وابسته به وجود دادگان غنی برچسب‌دار است. از سوی دیگر، در بسیاری از دامنه‌های مرتبط با پردازش زبان طبیعی مانند تحلیل احساسات رمزارزها یا بررسی واکنش مخرب داروها دادگان غنی برچسب‌دار برای آموزش شبکه‌های عصبی عمیق در دسترس نمی‌باشد. بنابراین،

پیشرفت یادگیری عمیق سبب شد کاربرد شبکه‌های عصبی مانند شبکه‌های عصبی بازگشتی¹ و پیچشی در وظایف مختلف

✧ نوع مقاله: پژوهشی

* نویسنده مسئول

پست(های) الکترونیک: kia.jahanbin@stu.yazd.ac.ir (جهان‌بین)

chahooki@yazd.ac.ir (زارع چاهوکی)

¹ Recurrent Neural Networks

می‌کردند. با این حال، در مدل‌های مبتنی بر کیسه کلمات، معنای کلی متن نادیده گرفته می‌شود. از این رو، روش‌های مبتنی بر تعبیه کلمات ارائه گردید. این روش‌ها، کلمات را از نظر تشابه معنای، دارای زمینه‌های مشابه در نظر می‌گیرند. نقطه ضعف این روش‌ها نیازمندی‌های آنها به مجموعه واژگان گسترده می‌باشد [4]. معرفی ترانسفورمرها از اوایل سال 2018 کمک شایانی به حل مسائل پردازش زبان طبیعی کرده است. یکی از معروف‌ترین ترانسفورمرها، نمایش‌های رمزگذار دوطرفه از ترانسفورمرها (برت) می‌باشد که بر روی مجموعه بزرگی از داده‌های انگلیسی آموزش دیده است. برت دارای لایه‌های رمزگذاری می‌باشد که با مکانیزم خود توجهی تقویت شده‌اند [3].

با این حال، همچنان یکی از چالش‌های اصلی عدم وجود دادگان در دامنه‌های خاص برای آموزش مدل‌های عمیق پایه و پیش‌آموزش دیده می‌باشد. بلو و همکاران [4]، برای افزایش غنای دادگان هدف ستون‌های مربوط به متن‌ها از دادگان مختلف را با یکدیگر الحاق کرده‌اند. در این مقاله نویسندگان، توجهی به مشابهت دادگان هدف و مبدا نکرده و صرفاً با الحاق داده‌ها سعی در افزایش غنای دادگان مبدا داشته‌اند. در این پژوهش به دنبال ارائه پاسخی به سوالات پژوهشی زیر می‌باشیم:

- 1- چگونه می‌توان دادگان میان دامنه‌ای مناسبی در دامنه مبدا انتخاب کرد که انتقال ویژگی‌های آنها به دامنه هدف موجب بروز پدیده انتقال منفی نگردد؟
- 2- چگونه می‌توان با ترکیب شبکه‌های یادگیری عمیق، لایه توجه و آموزش آنها با دادگان میان‌دامنه‌ای مدلی مطلوب با کمترین افت کارایی برای وظیفه تحلیل احساسات جنبه‌محور ارائه داد؟
- 3- با توجه به اینکه استفاده از مدل‌های یادگیری عمیق پایه و پیش‌آموزش دیده برای استخراج ویژگی و انتقال دانش از دامنه مبدا به هدف موثر می‌باشد. کدامیک از این مدل‌ها برای تحلیل احساسات رمزارزها اثربخشی بیشتری دارد؟

آموزش مدل‌های عمیق با مجموع دادگان کوچک سبب افت کارایی این مدل‌ها می‌شود [1].

یکی از پیشرفت‌های موثر در حوزه پردازش زبان طبیعی و بینایی ماشین یادگیری انتقالی می‌باشد. یادگیری انتقالی در بالابردن کارایی وظایف بینایی ماشین بسیار موثر بوده است. دلیل این موفقیت در دسترس بودن مجموعه داده‌های بزرگ برچسب‌گذاری شده مانند ایمپجنت و کوکو می‌باشد. در نتیجه شبکه‌های عصبی کانولوشن محور بر روی این مجموع داده‌ها به عنوان ستون فقرات شبکه (لایه استخراج‌گر ویژگی) آموزش می‌بینند. یادگیری انتقالی در پردازش زبان طبیعی تا سال 2018 که ترانسفورمرها توسط گوگل معرفی شدند، چندان مورد توجه قرار نگرفته بود [2]. یکی از وظایف پردازش زبان طبیعی که استفاده از یادگیری انتقالی در آن بسیار مورد توجه قرار گرفته است، تحلیل احساسات می‌باشد. با ظهور شبکه‌های اجتماعی بلادرنگ مانند توییتر، مجموعه‌های بی‌نظیری از دیدگاه‌های عمومی و تخصصی در دسترس قرار گرفته است [3]. با این حال فقدان داده‌های برچسب‌دار در حوزه‌های خاص مانند رمزارزها سبب شده تحلیل احساسات در این دامنه‌ها چالش برانگیز باشد. همچنین، مشکل مذکور در وظیفه تحلیل احساسات جنبه‌محور، تشدید می‌شود. زیرا، دادگان بسیار کمیابی با محتوی سهام و یا رمزارزها برای آموزش مدل‌های عمیق حاوی برچسب جنبه‌ها و احساسات هر جنبه وجود دارد. یادگیری انتقالی می‌تواند از دانش موجود در سایر دامنه‌ها برای انجام وظایف خاص در دامنه‌ای مشابه استفاده کند. با این حال استفاده از دادگان در دامنه‌های مختلف که به عنوان دادگان میان‌دامنه‌ای شناخته می‌شود، نیز چالش برانگیز است. نشان داده شده استفاده از دادگان نامشابه یا با مشابهت پایین با وظیفه یادگیری دامنه هدف، سبب بروز پدیده انتقال منفی¹ می‌شود. انتقالی منفی به دانش منتقل شده از دامنه مبدا گفته می‌شود که سبب افت کارایی مدل در دامنه هدف شود [2].

برای انجام وظایف پردازش زبان طبیعی در ابتدا تکنیک‌های مانند ایجاد کیسه‌ای از کلمات به وجود آمد که از N-gram استفاده

¹ Negative transfer

4- دادگان نامتوازن سبب جهت‌گیری مدل به سمت قطبیتی خاص می‌شود. چگونه می‌توان با این مشکل برخورد کرد؟

برای پاسخ به این سوالات در این پژوهش از مدل‌های مختلف یادگیری عمیق پایه، پیش‌آموزش‌دیده و ترکیب آنها با لایه توجه استفاده شده است. بخشی از این پژوهش را می‌توان مقایسه عملکرد مدل‌های متداول یادگیری عمیق (پایه، پیش‌آموزش‌دیده و ترکیبی) در تحلیل احساسات توئیت‌های مرتبط با رمازها دانست. با این حال صرفاً این پژوهش مقایسه مدل‌های پیشین نمی‌باشد، زیرا از روش‌های تنظیم دقیق برای پیاده‌سازی معماری ستون فقرات شبکه‌های عصبی عمیق ترکیبی استفاده شده است. در اینجا، دو روش تنظیم دقیق برای مدل‌های پیش‌آموزش‌دیده و شبکه‌های عصبی عمیق ترکیبی استفاده شده است. روش اول آموزش کل معماری عمیق می‌باشد که شامل بروزرسانی وزن‌های ستون فقرات شبکه و شبکه عصبی عمیق پایانی می‌باشد. روش دوم انجماد کل وزن‌های ستون فقرات شبکه و آموزش شبکه عصبی عمیق پایانی است. مقایسه روش‌های مختلف تنظیم دقیق شبکه‌های عصبی عمیق تا کنون کمتر مورد توجه قرار گرفته است [5]. لازم به توضیح است که در اینجا ستون فقرات شبکه که همان لایه استخراج‌گر ویژگی‌ها می‌باشد، شبکه‌ای پیش‌آموزش‌دیده از خانواده برت می‌باشد.

همچنین در این پژوهش تاثیر دادگان میان‌دامنه‌ای بر کارایی شبکه‌های عصبی عمیق ترکیبی در یادگیری انتقالی مورد توجه قرار گرفته است. لذا، در این پژوهش برای غنای دادگان رمازها در حوزه هدف از دادگانی میان‌دامنه‌ای مشابه در دامنه مبدا استفاده شده است. در این مقاله از یادگیری انتقالی شات صفر¹ استفاده شده، زیرا مدل در هنگام آموزش هیچ نمونه داده‌ای از دادگان دامنه هدف را نمی‌بیند. در این مقاله بر خلاف مقالات مشابه [4] که تنها دادگان مختلف را با یکدیگر الحاق می‌کنند، ابتدا شباهت دادگان با توئیت‌های رمازها با استفاده از معیار شباهت پیرسون محاسبه شده و سپس دادگان با شباهت بیشتر

در دامنه مبدا مورد استفاده قرار می‌گیرد [6]. در این پژوهش شباهت سه مجموعه دادگان تحلیل احساسات توئیت²، احساسات توئیت‌های بیت‌کوین³ و IMDB با دیتاست دامنه هدف که توئیت‌های جمع‌آوری رمازها می‌باشد مقایسه شده است. سپس متون دادگانی که دارای شباهت بیشتر از 40 درصد می‌باشد، با یکدیگر الحاق می‌شوند. همچنین، یکی از مشکلات دادگان در دنیای واقعی عدم توازن قطبیت‌ها می‌باشد. برای مقابله با این مشکل، در این پژوهش از تکنیک روش نمونه‌برداری بیش از حد اقلیت مصنوعی (SMOTE)⁴ [7] استفاده شده است. SMOTE تکنیکی آماری است که به طور مصنوعی نمونه‌های جدیدی از کلاس اقلیت را در فضای ویژگی ایجاد می‌کند تا مجموعه داده را متعادل کند.

به صورت کلی در این مقاله، روشی ترکیبی مبنی بر یادگیری عمیق انتقالی شات صفر برای تحلیل احساسات جنبه‌محور رمازها معرفی شده است. در این روش، ابتدا توئیت‌های افراد تاثیرگذار در حوزه رماز استخراج شده و عملیات پیش‌پردازش مانند یکسان‌سازی، حذف کلمات توقف، بن‌سازی و حذف کارکترهای زائد روی آنها اعمال می‌شود. سپس دادگان میان‌دامنه‌ای مشابه توئیت‌ها که دارای ضریب همبستگی پیرسون بیشتر از 40 درصد است ادغام شده و پس از عملیات پیش‌پردازش و استفاده از تکنیک SMOTE برای متوازن‌سازی داده‌ها به لایه استخراج‌گر ویژگی (ستون فقرات) یا همان دامنه مبدا برای استخراج دانش تزریق می‌شود. این لایه از شبکه پیش‌آموزش‌دیده دیستیل‌برت به همراه لایه توجه تشکیل شده است. لایه توجه به مدل کمک می‌کند ویژگی‌های مهم متن را شناسایی کرده و ویژگی‌های غیرضروری را نادیده بگیرد [8]. پس از استخراج دانش از دادگان میان‌دامنه‌ای این ویژگی‌ها به لایه پایانی یا همان دامنه هدف تزریق می‌شود. در دامنه هدف برای استخراج جنبه از دادگان توئیت‌ها از روش spaCy استفاده شده است [9]. همچنین، در این لایه برای تحلیل احساسات جنبه‌محور از شبکه عصبی عمیق حافظه کوتاه مدت بلند دو

² Twitter Sentiment Analysis (kaggle.com)

³ <https://data.world/mercal/btc-tweets-sentiment>

⁴ Synthetic Minority Over-sampling Technique

¹ Zero-shot transfer learning

می‌کند که بتوان بدون پیاده‌سازی مجدد شبکه‌های عصبی عمیق پایه و پیش‌آموزش دیده را با یکدیگر مقایسه کرد. در واقع با این لایه می‌توان به سرعت قسمت ستوان فقرات شبکه را با مدلی دیگر جایگزین کرد.

4- در این پژوهش، انتخاب برخی از ابرپارامترها و نرخ یادگیری شبکه‌های عصبی عمیق به صورت فوق‌پذیر و با کمک تکنیک جستجوی گریدی انجام شده است.

5- در این پژوهش با استفاده از Api توئیت‌نظرات تاثیرگذاران در حوزه رمزارزهای مختلف در بازه زمانی هشت ماه از سال 2021 تا 2022 جمع‌آوری شده است. این توئیت‌ها شامل تحلیل‌ها و نظرات تخصصی صاحب‌نظران در حوزه رمزارز می‌باشد و غنای محتوایی آن بسیار بیشتر از استخراج نظرات به وسیله هشتگ‌ها بدون در نظر گرفتن تخصص نویسنده است.

در بخش دوم به در ابتدا مفاهیم شبکه‌های عصبی عمیق پایه، پیش‌آموزش دیده و یادگیری عمیق انتقالی را تشریح می‌کنیم؛ سپس، به مرور مقالات پیشین در زمینه تحلیل احساسات جنبه‌محور با یادگیری عمیق انتقالی می‌پردازیم. در بخش سوم پیاده‌سازی شبکه‌های عصبی عمیق پیش‌آموزش دیده و ترکیبی را شرح می‌دهیم و در انتها نتایج آزمایشات تجربی، نتیجه‌گیری و پیشنهاداتی جهت تحقیقات آینده را بیان می‌کنیم.

2. مفاهیم و مروری بر تحقیقات پیشین

در این بخش در ابتدا مروری کوتاه بر مفاهیم تحلیل احساسات جنبه‌محور، شبکه‌های عصبی عمیق، پیش‌آموزش دیده و یادگیری عمیق انتقالی خواهیم داشت و سپس تحقیقات پیشین در زمینه تحلیل احساسات جنبه‌محور را بررسی می‌کنیم.

1.2. مفاهیم تحلیل احساسات جنبه‌محور، شبکه‌های

عصبی عمیق و یادگیری عمیق انتقالی

با گسترش شبکه‌های اجتماعی تقریباً بیش از نیمی از جمعیت دنیا در یک یا چند شبکه اجتماعی حضور دارند [5]. علاوه بر

طرفه (BiLSTM) با ترکیب لایه توجه بهره‌گیری شده است. در ادامه، بعد از تزریق دادگان هدف (توئیت‌های اثرگذاران) به مدل ترکیبی تنظیم دقیق شده BiLSTM و لایه توجه، ویژگی‌های اثرگذار بر قطبیت هر جنبه استخراج شده و بعد از عملیات برون‌ریزی¹ این ویژگی‌ها به لایه‌ای کاملاً متصل تزریق می‌شود. در این لایه، با استفاده از تابع سافت‌مکس قطبیت هر جنبه تعیین شده و در خروجی نمایش داده می‌شود. در زیر نوآوری‌های اصلی این پژوهش تشریح شده است.

1- در این پژوهش بر خلاف اغلب مقالات [10]- [12] که از تحلیل احساسات سنتی با روش‌های عمومی مانند ویدر² برای تحلیل احساسات رمزارزها بدون در نظر گرفتن جنبه‌های موجود در متن استفاده کرده‌اند، از یادگیری عمیق انتقالی با مدل‌های تنظیم دقیق شده برای تحلیل احساسات رمزارزها استفاده شده است. همچنین در این مطالعه، از یادگیری انتقالی شات صفر استفاده شده، زیرا مدل هیچ نمونه‌ای در دادگان دامنه هدف را در فرایند آموزش نمی‌بیند.

2- مدل‌های ترکیبی مورد استفاده توسط سه مجموعه دادگان با شباهت بیشتر از 40 درصد با دادگان رمزارزها آموزش دیده‌اند. برای سنجش شباهت دادگان مبدا و هدف از ضریب همبستگی پیرسون به دلیل اینکه توزیع آماری دادگان کاملاً نرمال نیست، استفاده شده است. این کار به صورت موثری انتقال منفی را کاهش می‌دهد. همچنین، از تکنیک SMOTE برای برخورد با پدیده عدم توازن دادگان استفاده شده است.

3- در این پژوهش دو روش مختلف تنظیم دقیق مدل‌های پیش‌آموزش دیده برای شناخت بهترین رویکرد جهت تحلیل احساسات جنبه‌محور رمزارزها مورد مقایسه قرار گرفته است. همچنین، در این مطالعه از لایه الحاق به عنوان یک سوئیچ استفاده شده است. این لایه کمک

¹ Dropout

² Vader

و چند وظیفه‌ای تقسیم می‌شود. لیو و همکاران¹ [17]، بیان می‌کنند جمله دارای چهار امان احساسی برای استخراج می‌باشد که عبارتند از: کلمه جنبه²، طبقه یا گروه جنبه³، کلمه عقیده⁴ و قطبیت احساسات⁵. برای مثال در جمله «بیت‌کوبین فوق‌العاده است»، کلمه جنبه برابر بیت‌کوبین، گروه جنبه برابر رمزارز، کلمه عقیده برابر فوق‌العاده و قطبیت احساسات برابر مثبت می‌باشد. با این چهار امان در تحلیل احساسات جنبه‌محور تک وظیفه‌ای، می‌توان چهار وظیفه تعریف کرد، به عبارت دیگر استخراج هر امان در جمله یک وظیفه محسوب می‌شود. همچنین، در تحلیل احساسات جنبه‌محور چندوظیفه‌ای، می‌توان ترکیب‌های دوتایی، سه‌تایی و یا استخراج کل امان‌ها را به عنوان (چهارتایی) یک وظیفه تعریف کرد، که در این پژوهش ترکیب‌های دوتایی استخراج جنبه و قطبیت آن مورد مطالعه قرار گرفته است.

1.1.2. شبکه‌های عصبی عمیق

یادگیری عمیق در واقع، زیرمجموعه‌ای از یادگیری ماشین می‌باشد. یادگیری عمیق نیازی به قوانین طراحی شده توسط انسان ندارد، بلکه از مقدار زیادی داده برای نگاشت ورودی به برجسب‌های از پیش تعریف‌شده (در مسائل طبقه‌بندی) استفاده می‌کند [18]. مدل‌های پایه یادگیری ماشین نیازمند چندین مرحله متوالی پیش‌پردازش، استخراج و انتخاب ویژگی‌های برتر، یادگیری و طبقه‌بندی می‌باشند. در حالی که، مدل‌های یادگیری عمیق، توانایی یادگیری مجموعه‌ای از ویژگی‌ها به صورت خودکار را دارا می‌باشند [19].

یکی از انواع خانواده مدل‌های عصبی که بسیار در حل مسائل پردازش زبان طبیعی مورد استفاده قرار گرفته، شبکه‌های عصبی بازگشتی (RNNs) می‌باشند. ویژگی برجسته این شبکه‌ها استفاده از «حافظه» می‌باشد. حافظه این امکان را به شبکه می‌دهد که اطلاعات ورودی‌های قبلی را برای اثرگذاری در ورودی فعلی استفاده کند [20]. به بیان رسمی اگر «x» لایه ورودی، «h» لایه

این، رسانه‌های اجتماعی بستری تعاملی جهت تبلیغات، ارائه نظرات تخصصی و کسب و کار فراهم آورده‌اند. این امر سبب شده بسیاری از افراد در تصمیم‌گیری جهت انتخاب کالا یا فرصت سرمایه‌گذاری به شدت به محتوای این رسانه‌ها متکی باشند [13]. توئیتر یکی از شبکه‌های اجتماعی است که به دلیل دسترسی آسان به توئیتهای به وسیله API و طول حداکثری 280 کاراکتر برای هر توئیتهای، یکی از کابردی‌ترین میکرو بلاگ‌ها برای تحلیل و بررسی نظرات می‌باشد. سینگ و همکاران [14]، نشان دادند نظرات کاربران در اغلب اوقات شامل «هدف» و «احساس» است. به عنوان مثال، «قیمت سهام A (هدف) در حال افزایش است (احساس)»، بنابراین تحلیل احساسات شامل شناسایی هدف و احساسات مرتبط با هدف می‌باشد. هدف می‌تواند سند، جمله (سطح جمله) یا حتی جنبه‌ای در یک جمله باشد.

تحلیل احساسات را می‌توان از نظر مدل‌های یادگیری ماشین برای حل مساله به سه رویکرد دسته طبقه‌بندی کرد: (1) مبتنی بر واژگان، که از فرهنگ لغتی از پیش تهیه شده برای تشخیص احساسات در سطح کلمه استفاده می‌کند. (2) یادگیری ماشین پایه، که از الگوریتم‌های طبقه‌بندی‌کننده‌ای که شبکه عصبی نیستند برای طبقه‌بندی کلمات به برجسب‌های از پیش تعیین شده استفاده می‌کند. (3) یادگیری عمیق، که از شبکه‌های عصبی عمیق (پایه یا پیش‌آموزش‌دیده) برای استخراج ویژگی‌های معنایی و تحلیل احساسات استفاده می‌کند [15]. از سوی دیگر، تحلیل احساسات از نظر عمق بررسی متن به چهار حالت تحلیل احساسات در سطح سند، جمله، عبارت و جنبه تقسیم می‌شود. در سطح سند، کل متن به عنوان موجودیتی واحد در نظر گرفته می‌شود و احساسات عمومی آن بررسی می‌شود؛ در سطح جمله، احساسات موجود در هر جمله به طور مجزا تحلیل می‌شود؛ در سطح عبارت، هر عبارت یا تکه‌ای از جمله مورد بررسی قرار می‌گیرد؛ و در سطح جنبه، جنبه‌های متن در سطح هر جمله یا عبارت شناسایی شده و قطبیت هر جمله تعیین می‌شود [16]. در این پژوهش هدف ما تحلیل احساسات جنبه‌محور می‌باشد. تحلیل احساسات جنبه‌محور از نظر نوع وظایف به دو دسته تک

¹ Liu et al.

² Aspect Term

³ Aspect Category

⁴ Opinion Term

⁵ Sentiment Polarity

شناخته می‌شوند. این دروازه‌ها، همانند دروازه‌های LSTM، کنترل می‌کنند چه میزان و کدام اطلاعات باید حفظ شود [26]. نوعی دیگری از LSTM که اجازه پردازش اطلاعات در دو جهت را می‌دهد BiLSTM نام دارد. برخلاف LSTM، در این شبکه ورودی در هر دو جهت (رو به عقب و رو به جلو) جریان دارد. به عبارت دیگر، BiLSTM یک لایه LSTM اضافه دارد که جهت جریان اطلاعات را معکوس می‌کند. به این معنی که، دنباله ورودی در لایه اضافی LSTM به عقب جریان می‌یابد؛ سپس، خروجی‌های هر دو لایه LSTM با روش‌های مانند میانگین، مجموع، ضرب یا الحاق ترکیب می‌شوند [20].

2.1.2. یادگیری انتقالی عمیق و شبکه پیش‌آموزش دیده برت

یادگیری انتقالی تکنیکی است که از دانش موجود در دامنه مبدا برای بهبود عملکرد وظیفه در دامنه هدف استفاده می‌کند. در یادگیری انتقالی دامنه مبدا را با D و دامنه هدف را با T نشان می‌دهند. دامنه مبدا (D) از دو جزء تشکیل شده است: داده که با X و توزیع آن که با $P(X)$ و به صورت رابطه (1) نشان می‌دهیم [21].

$$D = X, P(X) \quad (1)$$

دامنه هدف (T) نیز از دو جزء تشکیل شده است: داده برچسب y و تابع پیش‌بینی $f(\cdot)$ که به صورت رابطه (2) نشان می‌دهیم.

$$T = y, f(\cdot) \quad (2)$$

به بیان رسمی، یادگیری انتقالی رویکردی است برای بهبود وظیفه یادگیری (TT) در دامنه هدف (TD)، با دانش آموخته شده از وظیفه یادگیری (ST) در دامنه مبدا (SD)¹، که به طور معمول $SD \neq TD$ و $ST \neq TT$ می‌باشد. شبکه‌های عصبی از پیش‌آموزش دیده مدل‌های مناسبی جهت پیاده‌سازی تکنیک‌های یادگیری انتقالی می‌باشند [21]. یادگیری عمیق انتقالی را می‌توان به سه دسته زیر، تک و چند شات² تقسیم کرد. در یادگیری عمیق انتقالی زیروشات که در این مقاله نیز استفاده شده است،

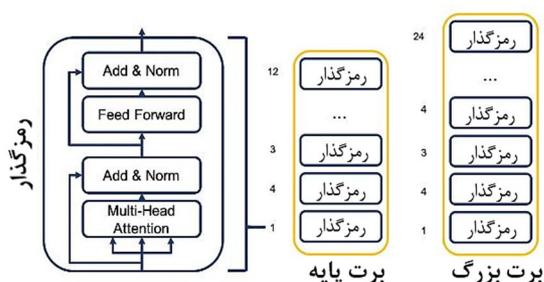
پنهان و « y » لایه خروجی شبکه باشد. در هر زمان t ، ورودی فعلی ترکیبی از ورودی $x(t)$ و $x(t-1)$ حاصل می‌شود. بنابراین، خروجی در هر زمان t برای بهبود خروجی بعدی ($t+1$) به شبکه برمی‌گردد. با این حال این شبکه‌ها دارای دو مشکل اساسی محوشدگی و انفجار گرادیان می‌باشند. گرادیان‌ها اطلاعات مورد استفاده در شبکه عصبی بازگشتی را حمل می‌کنند. زمانیکه گرادیان خیلی کوچک شود، بروزرسانی پارامترها غیر موثر می‌شود و به این پدیده محوشدگی گرادیان می‌گویند. از سوی دیگر، در حین آموزش شبکه عصبی، اگر شیب به جای کاهش، تمایل به رشد تصاعدی داشته باشد، پدیده انفجار گرادیان رخ می‌دهد. این مشکل زمانی ایجاد می‌شود که خطای بزرگ در گرادیان‌ها انباشته شده و سبب می‌گردد بروزرسانی بسیار بزرگی در وزن‌های شبکه عصبی در حین فرآیند آموزش رخ دهد [20].

برای رفع مشکل RNN، شبکه‌های عصبی عمیق حافظه بلند کوتاه مدت (LSTM)، واحد بازگشتی گیتی (GRU) و BiLSTM معرفی شده‌اند. در واقع می‌توان LSTM را مدل پایه دو شبکه دیگر نامید که با اعمال تغییراتی در LSTM به وجود آمده‌اند. LSTM نوع خاصی از شبکه عصبی بازگشتی بوده و قادر به یادگیری وابستگی‌های طولانی مدت می‌باشد. LSTM از سه قسمت تشکیل شده و هر قسمت عملکرد جداگانه‌ای دارد. بخش اول انتخاب می‌کند که آیا اطلاعاتی که از دوره زمانی $t-1$ به دست آمده باید به خاطر سپرده شود یا فراموش گردد. در قسمت دوم، LSTM سعی می‌کند اطلاعات ورودی جدید را بیاموزد و در نهایت در بخش سوم، LSTM اطلاعات بروز شده دوره زمانی فعلی (t) را به دوره زمانی بعدی ($t+1$) منتقل می‌کند. این سه قسمت در LSTM به ترتیب به عنوان دروازه‌های فراموشی، ورودی و خروجی شناخته می‌شوند [20]. نوعی دیگری از شبکه عصبی بازگشتی که تا حد زیادی مشابه LSTM می‌باشد و برای حل مشکل حافظه کوتاه مدت پیاده‌سازی شده، GRU نام دارد. در این شبکه جهت تسریع در فرآیند آموزش و سبک‌سازی مدل از دو دروازه بجای سه دروازه استفاده شده است. این دروازه‌ها که به نام‌های تنظیم مجدد و به‌روزرسانی

¹ TT: Target Task, TD: Target Domain, ST: Source Task, SD: Source Domain

² Zero/One/Few Shot transfer Learning

می‌کند. این کار سبب می‌شود هنگام تعیین زمینه‌ای برای یک کلمه، برت ورودی‌های قبلی و بعدی کلمه را نیز در نظر بگیرد. در حالیکه شبکه‌های مبتنی بر بازگشت مانند LSTM، ورودی فعلی را فقط با در نظر گرفتن ورودی‌های قبلی پردازش می‌کنند. برت به صورت استاندارد به دو شکل پایه و بزرگ ارائه شده است. برت پایه شامل 12 لایه رمزگذار و 768 لایه پنهان است. از سوی دیگر، برت بزرگ دارای 24 لایه رمزگذار با 1024 لایه پنهان می‌باشد. انواع دیگر مدل‌های برت که اغلب مدل‌های فشرده‌تر، سریع‌تر و موثرتر از برت پایه می‌باشند، عبارتند از: آلبرت، الکترا، دستیل برت، اسپن برت، برت سام³ و XLNet. در این پژوهش از سه مدل پایه برت، XLNet و دستیل برت استفاده شده است. در شکل (1) تفاوت برت پایه و بزرگ نشان داده شده است.



شکل (1): مقایسه دو نسخه برت پایه و بزرگ [22]

در ادامه دو نوع شبکه عصبی عمیق مبتنی بر برت یعنی برت تقطیره شده و XLNet بررسی می‌گردد.

دستیل برت نسخه سبک‌سازی شده برت می‌باشد. این مدل فشرده‌تر، کوچک‌تر، سریع‌تر و سبک‌تر از برت است. استفاده از برت به دلیل وجود میلیون‌ها پارامتر برای کاربردهای دنیای واقعی به خصوص در تلفن‌های همراه یا دستگاه‌ها با منابع پردازشی ضعیف، چالش برانگیز است. اگرچه ممکن است معمارهای پیچیده‌تر به دقت بالاتری دست یابند، اما می‌شود با روش‌های تنظیم دقیق به میزان قابل توجهی دقت مدل‌های سبک‌تر را نیز بهبود داد. در مقایسه با مدل‌های مانند برت بزرگ و XLNet، که عملکرد مناسبی دارند، هدف دستیل برت کاهش زمان محاسبه است. در فشرده‌سازی دستیل برت، از چارچوب

مدل هیچ نمونه‌ای از داده‌های هدف دارای برچسب را در اختیار ندارد و تنها با دادگان میان دامنه‌ای مشابه آموزش دیده است. در یادگیری انتقالی تک‌شات، مدل تنها یک نمونه از داده‌های هدف دارای برچسب را در اختیار دارد. از سوی دیگر، در یادگیری انتقالی با شات محدود مدل تعداد کمی از نمونه‌های داده‌های هدف دارای برچسب را در اختیار دارد و باید بتواند با استفاده از این تعداد کم نمونه و دانش قبلی خود، پیش‌بینی‌های دقیقی انجام دهد. برای انجام یادگیری انتقالی، در سال‌های اخیر شبکه‌های عصبی عمیق پیش‌آموزش دیده خانواده برت در پردازش زبان طبیعی استفاده شده است. برت مبنی بر یادگیری انتقالی بوده از ترانسفورمرها استفاده می‌کند. یادگیری انتقالی مستلزم آموزش مدل برای وظیفه‌ای با مجموعه دادگان غنی در دامنه مبدا و انتقال دانش به دست آمده برای وظیفه دیگری در دامنه هدف می‌باشد [22].

برت، برای دو وظیفه زبان پوشانده شده و بدون پوشش¹ آموزش دیده است. در وظیفه زبان پوشانده شده، جملات به شبکه داده می‌شود و برخی از کلمات برای مدل پنهان می‌شوند. سپس، مدل تلاش می‌کند این کلمات پنهان را پیش‌بینی کند. در وظیفه زبان بدون پوشش، دو جمله به مدل داده می‌شود و مدل باید پیش‌بینی کند یک جمله می‌تواند ادامه جمله دیگری باشد یا خیر. مدل برت به صورت پیش‌فرض بر روی دو مجموعه داده بزرگ و یکی پدیای انگلیسی و مجموعه دادگان کتاب‌ها که شامل ۱۱,۰۳۸ کتاب می‌باشد، آموزش دیده است. این مدل، از رمزگذاری مبتنی بر ترانسفورمرها که نوعی شبکه عصبی می‌باشند، استفاده می‌کند. برت یک جمله را به عنوان ورودی می‌گیرد و تمامی کلمات جمله را نشانه‌گذاری² می‌کند. سپس، کلمات نشانه‌گذاری شده به مدل برت باز می‌گردد. خروجی برت نمایش برداری برای هر کلمه نشانه‌گذاری شده است [22]. مدل‌های که از لایه‌های رمزگذارهای مانند برت استفاده می‌کنند، می‌توانند درک بهتری از زمینه متن داشته باشد. زیرا رمزگذار، کل جمله را به عنوان ورودی و به صورت همزمان پردازش

¹ Masked and Unmasked Language

² Tokenization

³ AIBERT, SpanBERT, ELECTRA, SumBERT

بر واژگان، یادگیری ماشین پایه و یادگیری عمیق با شبکه‌های عصبی عمیق (پایه، ترکیبی یا پیش‌آموزش دیده) تقسیم کرد. در ادامه پژوهش‌های مرتبط با این تقسیم‌بندی‌های یادشده را بررسی می‌کنیم.

1.2.2. تحلیل احساسات با یادگیری ماشین پایه

سه‌حسینی و همکاران [25]، از بیز ساده و K-همسایه نزدیک برای تعیین تطبیق احساسات توئیت‌ها استفاده کرده‌اند. در این مطالعه نشان داده شد، بیز ساده کارایی بالاتری نسبت به از K-همسایه نزدیک دارد. در پژوهش [3]، توئیت‌های مرتبط با بررسی محصولات آمازون جمع‌آوری شده و سپس سه مدل یادگیری ماشین‌های بردار پشتیبان، لجستیک رگرسیون و K-همسایه نزدیک، برای تحلیل احساسات داده‌ها با یکدیگر مقایسه شده‌اند. نتایج آزمایشات نشان می‌دهد مدل لجستیک رگرسیون با دقت بیش از 88 درصد بهترین کارایی را داشته است.

نویسندگان در پژوهش [26]، روشی برای پیش‌بینی تغییرات قیمت بیت‌کوین و اتریوم با استفاده از داده‌های توئیت و گوگل ترند ارائه کرده‌اند. با تجزیه و تحلیل توئیت‌ها، پژوهشگران متوجه شده‌اند که حجم توئیت‌ها، به جای احساسات توئیت، می‌تواند پیش‌بینی‌کننده جهت قیمت باشد. در این مقاله از مدلی خطی جهت پیش‌بینی تغییرات قیمت استفاده شده است. در پژوهش [27]، از مدل ترکیبی LSTM و GRU با ورودی‌های ارزش بازار سهام، حجم معاملات، ویژگی‌های قیمتی (OHLCV) و حجم نظرات گوگل ترند و توئیت جهت پیش‌بینی مونرو و لایت‌کوین، استفاده شده است.

تحقیقات در حوزه LSTM نشان داده، استفاده از این روش به این دلیل توانایی آن در به خاطر سپردن ویژگی‌های داده‌ها، بسیار موثر است. در همین راستا هوانگ و همکاران [28]، مدلی برای تحلیل احساسات مبتنی بر LSTM برای پیش‌بینی قیمت رمازرها ارائه کرده‌اند. این مدل با بهره‌گیری از قابلیت LSTM در یادگیری وابستگی‌های بلندمدت در داده‌های متنی، به تحلیل احساسات کاربران در پلتفرمی چینی پرداخته است. آنها با ایجاد یک دیکشنری اختصاصی برای واژگان مرتبط با رمازرها، دقت

«معلم-دانش‌آموز» استفاده شده که به آن تقطیر دانش نیز گفته می‌شود. در این تکنیک، مدل یا شبکه‌ای بزرگ به عنوان «معلم» آموزش داده می‌شود و دانش به مدل کوچک‌تر که «دانش‌آموز» نامیده می‌شود منتقل می‌گردد [23]. در دستیل‌برت، توکن‌های تعبیه شده حذف شده و همچنین تعداد پارامترهای شبکه کاهش تعداد یافته است. این کار سبب شده تا حد زیادی محاسبات شبکه کاهش یافته و سرعت یادگیری بهبود یابد. بر اساس استاندارد ارزیابی عمومی مدل‌های زبانی (GLUE)، کارایی دستیل‌برت بدون هیچ‌گونه تنظیم دقیق 97 درصد برت پایه می‌باشد. با این حال، 40 درصد پارامترهای شبکه عصبی کاهش یافته و زمان آموزش به میزان قابل توجهی سریعتر شده است. محققان در مرجع [23]، نشان داده‌اند با ترکیب روش تقطیر انتقالی و چارچوب یادگیری دومرحله‌ای، می‌توان به دقتی برابر و یا بالاتر از برت پایه دست یافت. در ادامه XLNet را معرفی می‌کنیم.

XLNet در واقع نوعی برت است که با ترکیب ترانسفورمرهای بزرگ دوجهته -XL و رمزگذاری خودکار برت، ایجاد شده است. در این شبکه از تکنیک جایگشت برای یافتن زمینه متن به صورت دوجهته استفاده شده، همچنین نشانه‌ها به طور تصادفی پیش‌بینی می‌شوند. در برت پایه 15% از توکن‌ها پوشانده شده و سایر نشانه‌ها به جای ترتیب تصادفی به صورت متوالی پیش‌بینی می‌شوند. لذا، پیش‌بینی تصادفی نشانه‌ها به XLNet امکان می‌دهد وابستگی‌ها و روابط طولانی بین کلمات را بهتر یاد بگیرد. تحقیقات نشان داده XLNet پایه در وظایف پردازش زبان طبیعی از جمله تحلیل احساسات، پاسخ به سوال و استنتاج زبان طبیعی بهتر از برت پایه عمل می‌کند. به عبارت دیگر، تکنیک زمینه دو جهته به کلمات قبل و بعد از نشانه نگاه می‌کند تا پیش‌بینی دقیق‌تری ارائه دهد [24].

2.2. پیشینه تحقیق در تحلیل احساسات شبکه‌های

اجتماعی با توجه به رمازرها

همان‌گونه که ذکر شد رویکردهای حل مساله در تحلیل احساسات از نظر مدل‌های پیاده‌سازی می‌توان به سه دسته مبتنی

این روش به ویژه در تحلیل احساسات رسانه‌های اجتماعی محبوبیت زیادی پیدا کرده و نتایج قابل اعتمادی را ارائه داده است [10]، [12]. در سوتیریوس و همکاران [30]، به بررسی امکان پیش‌بینی قیمت ارزهای دیجیتال با استفاده از تحلیل احساسات در شبکه‌های اجتماعی پرداخته شده است. نویسندگان داده‌های توئیتر و قیمت هفت ارز دیجیتال محبوب را جمع‌آوری و پس از پاکسازی، با استفاده از ابزار ویدر به تحلیل احساسات پرداختند. سپس، برای ارزیابی تأثیرگذاری احساسات بر قیمت، از آزمون‌های علیت گرنجر استفاده کردند. نتایج نشان داد که در برخی از رمزارزها مانند اتریوم و پولکادات، احساسات توئیتری می‌تواند منجر به پیش‌بینی بسیار دقیق‌تر شود. این پژوهش بر تأثیر قابل توجه احساسات شبکه‌های اجتماعی بر نوسانات بازار رمزارزها تأکید دارد. با این حال، بدیهی است درصد خطای روش‌های مانند ویدر می‌تواند در تحلیل احساسات دامنه‌های خاص بالا باشد، زیرا این روش‌ها اغلب برای توئیتهای عمومی ساخته شده‌اند. همچنین، مدل‌های پایه یادگیری ماشین، نیاز به دادگان غنی برچسب‌دار جهت آموزش دارند که در اکثر اوقات این دادگان موجود نمی‌باشد. علاوه بر این، استخراج ویژگی در مدل‌های پایه یادگیری ماشین برخلاف شبکه‌های عصبی عمیق به صورت دستی می‌باشد. بزرگ‌ترین مشکل در این مدل‌ها عدم امکان استفاده از یادگیری انتقالی می‌باشد.

2.2.2. تحلیل احساسات رمزارزها با مدل‌های یادگیری عمیق

سیهان و همکاران [31]، رویکردی برای تحلیل احساسات مرتبط با رمزارزها ارائه کرده‌اند. در این مطالعه، معماری ترکیبی عمیق شامل لایه‌های پیچشی، مکانیسم بهبود گروهی²، لایه‌های دو جهته، مکانیسم توجه و لایه‌های کاملاً ارائه شده است. مکانیسم بهبود گروهی به بهبود ویژگی‌های مفید کمک می‌کند، در حالی که مکانیسم توجه با اختصاص وزن‌های مختلف به ویژگی‌ها، اطلاعات زمینه‌ای را برجسته می‌سازد. نتایج نشان می‌دهد معماری پیشنهادی، عملکرد مطلوبی نسبت به معماری‌های پیشین به دست آورده است.

تحلیل احساسات را بهبود داده و نتایج نشان داد که این رویکرد توانسته است در مقایسه با مدل‌های سنتی خودرگرسیون، به طور قابل توجهی دقت و فراخوان بالاتری در پیش‌بینی نوسانات قیمت رمزارزها داشته باشد. همچنین نیر و همکاران [11]، به بررسی مدل‌های تحلیل احساسات برای توئیتهای مرتبط با رمزارزها با استفاده از تکنیک‌های مختلف یادگیری عمیق پرداخته‌اند. آنها از پنج مدل یادگیری عمیق شامل RNN، LSTM، GRU و ترکیب BiLSTM و CNN برای تحلیل احساسات توئیتهای مربوط به بیت‌کوین استفاده کرده‌اند. داده‌های مورد استفاده شامل بیش از 1,5 میلیون توئیتهای بوده که از وبگاه Kaggle استخراج شده است. در ارزیابی مدل‌ها از ویدر برای محاسبه نمره احساسات و از فست‌تکست¹ برای استخراج ویژگی‌ها استفاده شد. نتایج نشان داد که مدل LSTM با دقت 95/95% بهترین عملکرد را در پیش‌بینی و طبقه‌بندی احساسات توئیتهای به دست آورده است.

در مقالاتی که از داده‌های توئیتر جهت پیش‌بینی قیمت رمزارز استفاده شده یکی از خلاءهای تحقیقاتی عدم تنظیم دقیق مدل برای تحلیل احساسات رمزارز می‌باشد. در این مقالات اغلب پژوهشگران از روش‌های برپایه واژگان مانند ویدر [29] برای تشخیص خودکار احساسات استفاده می‌کنند. ویدر مدلی برای تحلیل احساسات مبتنی بر فرهنگ لغت و قواعد از پیش تعریف شده است که به طور خاص برای تحلیل احساسات در متون کوتاه و غیررسمی، مانند نظرات شبکه‌های اجتماعی و دیدگاه‌های برخط، توسعه یافته است. این مدل با استفاده از لغت‌نامه‌ای غنی از کلمات و عبارات احساسی، به هر کلمه یا عبارت یک نمره احساسی (مثبت، منفی، یا خنثی) اختصاص می‌دهد. یکی از ویژگی‌های برجسته ویدر توانایی آن در درک و تحلیل جزئیات احساسی خاص متون غیررسمی است، مانند استفاده از علائم نگارشی، شدت‌سنج‌ها (مانند «خیلی» یا «بسیار») و حتی اصطلاحات عامیانه است. ویدر به دلیل سادگی، دقت بالا و عدم نیاز به داده‌های آموزشی در بسیاری از کاربردهای پردازش زبان طبیعی در محیط‌های واقعی استفاده می‌شود [29].

² Group-wise Enhancement Mechanism

¹ FastText

ژائو و همکاران [35]، رویکردی را برای تحلیل احساسات جنبه‌محور پیشنهاد کردند که از شبکه‌های پیچشی مبتنی بر گراف⁴ برای دریافت وابستگی‌های احساسات بین جنبه‌های مختلف در یک جمله استفاده می‌کند. نویسندگان استدلال می‌کنند که رویکردهای موجود عمدتاً این وابستگی‌ها را نادیده می‌گیرند، که می‌تواند اطلاعات ارزشمندی را برای تحلیل احساسات ارائه دهد. مدل ارائه شده شبکه‌های پیچشی مبنی بر گراف وابستگی احساسات (SDGCN) ابتدا از مکانیسم توجه دو طرفه با رمزگذاری موقعیت برای به دست آوردن نمایش‌های خاص جنبه استفاده می‌کند؛ سپس، از GCN برای مدل‌سازی وابستگی‌های احساسات بین جنبه‌ها بهره می‌گیرد. زونیک و همکاران [36]، رویکردی برای تحلیل احساسات جنبه‌محور با استفاده از GCN بر روی نمودارهای وابستگی نحوی معرفی کرد. نویسندگان استدلال می‌کنند که ساختار نحوی جملات برای تجزیه و تحلیل دقیق احساسات بسیار مهم است، به ویژه برای حوزه‌های سلامت و رفاه که در آن احساسات اغلب به دلیل ماهیت محتوا به سمت منفی گرایش پیدا می‌کنند. روش آنها شامل ایجاد نموداری از جملات است که در آن گره‌ها کلمات و یال‌ها وابستگی‌های نحوی را نشان می‌دهند. در این مطالعه، به اندازه کافی عدم تعادل طبقاتی در مجموعه دادگان را در نظر نمی‌گیرد. این مشکل، می‌تواند توانایی مدل برای تعمیم در کلاس‌های احساسات مختلف را تغییر دهد. علاوه بر این، این مقاله ممکن است فاقد بحث مفصل در مورد کارایی محاسباتی مدل، که جنبه‌ای حیاتی در هنگام کاربرد برای سناریوهای دنیای واقعی است. علاوه بر این، این مطالعه با تمرکز بر یک حوزه خاص (سلامت و رفاه) محدود شده، در نتیجه سوالاتی در مورد عملکرد آن بر روی مجموعه دادگان با زمینه‌های متفاوت ایجاد می‌شود. لی و همکاران [37]، برای شناسایی موثر جفت‌های جنبه-نظر مدلی مبنی بر برچسب‌گذاری فاصله محور⁵ (QDSL) ارائه کرده‌اند. نویسندگان وظیفه استخراج جفت جنبه-نظر⁶

مودیج و همکاران [32]، به بررسی استفاده از تکنیک‌های تعبیه گراف برای تحلیل احساسات در زمینه رمزارزها پرداخته‌اند. نویسندگان با هدف بهبود دقت مدل‌های تحلیل احساسات، از روش تعبیه گراف برای نمایش و استخراج روابط معنایی و نحوی بین کلمات در داده‌های متنی استفاده کردند. در این پژوهش، داده‌های مورد استفاده به صورت گراف نمایش داده شد و با بهره‌گیری از تکنیک‌های تعبیه گراف، مانند DeepWalk و مدل Word2Vec، تبدیل به بردارهای کم‌بعد شدند که ویژگی‌های ساختاری متن را حفظ می‌کنند. سپس این بردارها به عنوان ورودی به یک مدل BiLSTM داده شد می‌شود. نتایج می‌دهد، مدل تعبیه گرافی به طور قابل توجهی عملکرد بهتری نسبت به مدل‌های سنتی تعبیه کلمات داشته و دقت بالاتری بدست آورده است. این پژوهش نشان می‌دهد استفاده از تعبیه گرافی می‌تواند منجر به بهبود درک و تحلیل دقیق‌تر احساسات در متون مرتبط با رمزارزها شود. همچنین اسالم و همکاران [33]، مدلی ترکیبی مبنی بر LSTM و GRU برای تحلیل احساسات توییت‌های مرتبط با رمزارزها ارائه کرده‌اند. آنها با جمع‌آوری ۴۰,۰۰۰ توییت رمزارز، این داده‌ها را پس از پیش‌پردازش با استفاده از مدل ترکیبی پیشنهادی تحلیل کردند. برای استخراج ویژگی‌ها از روش‌های TF-IDF، BoW و Word2Vec استفاده شده است و سپس این ویژگی‌ها به مدل ترکیبی LSTM-GRU داده شده‌اند. نتایج نشان این پژوهش بیان می‌کند که استفاده از مدل‌های ترکیبی می‌تواند به طور قابل توجهی دقت تحلیل احساسات رمزارزها را بهبود دهد. در رویکردی دیگر که مبنی بر مدل پیش‌آموزش‌دیده روبرتا است، داوچف و همکاران [34]، برای پیش‌بینی قیمت بیت‌کوین از مجموعه دادگان توییت‌های کگل³ و دادگان قیمتی پیشین استفاده کرده‌اند. در این پژوهش قطبیت توییت‌های مربوط به بیت‌کوین بعد از پیش‌پردازش، توسط شبکه روبرتا تعیین شده است. سپس، مقادیر حاصل از قطبیت توییت‌ها با مقادیر قیمت‌های گذشته بیت‌کوین ترکیب شده و برای پیش‌بینی قیمت به رگرسورهای FbProphet و XGBoost تزریق شده است.

⁴ Graph Convolutional Networks (GCNs)

⁵ Question-Driven Span Labeling (QDSL)

⁶ Aspect-Opinion Pair Extraction (AOPE)

³ Kaggle

دقیق شده است. تنظیم دقیق مدل دامنه هدف و انتخاب دادگان مشابه در دامنه مبدا کمک می‌کند کارایی روش پیشنهادی بهبود یافته و از انتقال منفی به شدت کاسته شود. همچنین، در این مقاله، داده‌ها شامل جملات متنی هستند که به منظور تحلیل احساسات جمع‌آوری شده‌اند. هر جمله به عنوان یک نمونه داده در نظر گرفته می‌شود و دارای یک برچسب است که بیانگر احساسات مرتبط با آن جمله (مثبت، منفی یا خنثی) می‌باشد. بنابراین، ماهیت داده‌ها متنی است و هر نمونه داده یک جمله کامل است، نه کلمات یا بخش‌های جداگانه از یک جمله. همان‌گونه که ذکر شد در اینجا از تکنیک SMOTE برای حل مشکل دادگان نامتوازن استفاده می‌شود. SMOTE به طور معمول برای رفع مشکل عدم توازن در داده‌های عددی و دسته‌بندی‌ها استفاده می‌شود. این تکنیک با تولید نمونه‌های جدید برای کلاس‌های اقلیت، توازن بین کلاس‌ها را برقرار می‌کند. در زمینه داده‌های متنی، SMOTE به طور مستقیم بر روی داده‌های متنی خام (جملات یا کلمات) اعمال نمی‌شود، بلکه ابتدا نیاز است که داده‌های متنی به یک نمایش عددی (نظیر بردارهای ویژگی) تبدیل شوند. در این مقاله، جملات متنی ابتدا با استفاده از توکنایزر روبرتا به توکن‌های کوچک‌تر تبدیل می‌شوند و سپس به بردارهای تعبیه‌ای⁸ که نمایانگر ویژگی‌های معنایی عمیق جملات هستند، تبدیل می‌گردند. این بردارها نمایش عددی جملات را تشکیل می‌دهند و هر بردار نماینده یک جمله است. پس از تبدیل داده‌های متنی به این فضای تعبیه‌ای، SMOTE بر روی این بردارهای عددی اعمال می‌شود تا نمونه‌های جدیدی از کلاس‌های اقلیت تولید شود. این نمونه‌های جدید به عنوان بردارهای تعبیه‌ای جدید به مجموعه داده اضافه می‌شوند [7].

کبیر و همکاران [39]، چارچوبی برای طبقه‌بندی احساسات مبتنی بر جنبه را معرفی کردند. نوآوری اصلی این تحقیق در روش ترکیبی آن نهفته است که عبارت است از ترکیبی از سینتکس و رویکردهای مبتنی بر وابستگی برای استخراج جنبه‌ها از متن انتقادات مرتبط با دادگان هتل. این مطالعه به دلیل تمرکز

را به وظایف فرعی استخراج کلمه جنبه⁷ (ATE) و استخراج نظر با جنبه مشخص (ASOE) تقسیم می‌کنند. این مدل از یک طرح برچسب‌گذاری مبتنی بر دهانه برای شناسایی جفت‌های جنبه-نظر استفاده می‌کند و پیشرفت‌های قابل توجهی نسبت به روش‌های پیشین در مجموعه داده‌های معیار نشان می‌دهد. در این تحقیق، علاوه بر در نظر نگرفتن مشکل عدم تعادل کلاس در مجموعه داده و عدم تنظیم دقیق مدل پیشنهادی، مدل پیشنهادی بر روی یک مجموعه داده با زمینه متفاوت، آزمایش نشده است. هبت و همکاران [38]، مدلی ترکیبی از شبکه‌های عصبی بازگشتی مانند LSTM، BiLSTM و GRU و شبکه‌های عصبی کانولوشنی با لایه‌های تعبیه‌شده مختلف (XLNet، MultiFiT و CamemBERT) ارائه کرده‌اند. نتایج نشان می‌دهد که ترکیب GRU و CNN با XLNet در سه مجموعه داده فرانسوی: نظرات مشتریان فرانسوی آمازون (FACR)، AlloCiné Dataset (AC) و تجزیه و تحلیل احساسات فرانسوی تویتر، به ترتیب با دقت 96/5، 90/1 و 89/6 درصد دست یافته است.

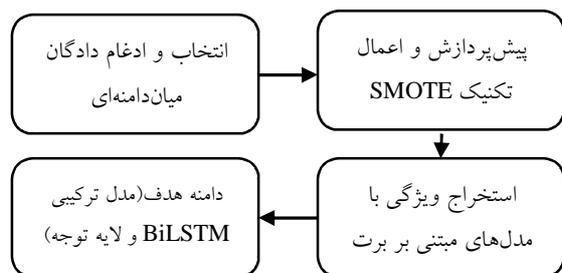
همان‌گونه که در مقالات مبتنی بر مدل‌های پیش‌آموزش دیده [4]، [38]، نشان داده شده، مدل‌های عصبی عمیق باید نسبت به وظیفه و نوع داده ورودی، تنظیم گردند. نشان داده شده [21] مدل‌های عصبی عمیقی که در تحلیل احساسات دادگان تویتهای عمومی و یا نظرات فیلم‌ها کارایی مناسبی دارند، کارایی خود را بر روی دادگانی مانند تویتهای اثرات مخرب واکسن و یا دارو از دست می‌دهند. از این رو، در این پژوهش به دلیل کمبود نظرات مرتبط با پیش‌بینی رمزارزهای مختلف از دادگان میان دامنه‌ای با درجه شباهت کسینوسی بیشتر از 40 درصد استفاده شده است. سپس متن‌های این دادگان با هم الحاق شده و به شبکه‌های عصبی عمیق پایه و پیش‌آموزش دیده دامنه مبدا جهت استخراج ویژگی تزریق شده است. در ادامه، دانش موجود در دامنه مبدا برای آموزش شبکه عصبی عمیق دامنه هدف استفاده شده است. معماری مورد استفاده این مقاله با دو روش مختلف که در بخش بعد مفصل تشریح می‌گردد، تنظیم

⁸ Embedding Vectors⁷ Aspect Term Extraction (ATE)

نمودار سند-کلمه برای به تصویر کشیدن وابستگی‌های گلوبال کلمات در متن می‌سازد. چوماکوف و همکاران [42]، روشی برای استخراج سه‌گانه احساسات جنبه¹¹ (ASTE) با استفاده از GPT معرفی کردند. این رویکرد از تکنیک‌های یادگیری و تنظیم دقیق با مدل‌های GPT برای پردازش و تجزیه و تحلیل موثر داده‌های متنی در حوزه‌های مختلف، از جمله موارد ناشناخته، استفاده می‌کند. این روش دارای توانایی کاهش وابستگی به داده‌های گسترده دامنه خاص و حفظ عملکرد بالا در برنامه‌های کاربردی دنیای واقعی را دارا می‌باشد. در جدول (1) خلاصه مقالات فوق به همراه نقاط ضعف آنها تشریح شده است. در جدول (2)، مدل پیشنهادی به صورت خلاصه و از نظر ساختار معماری و نوآوری با مقالات جدول (1) مقایسه شده است.

3. روش پیشنهادی

مدل‌های پیشنهادی در این پژوهش، ترکیبی از شبکه‌های پیش‌آموزش مبتنی بر برت و شبکه عصبی عمیق BiLSTM می‌باشد. چارچوب کلی این مدل‌ها در شکل (2) نشان داده شده است.



شکل (2): مدل عمومی شبکه‌های عصبی عمیق پیش‌آموزش دیده ترکیبی برای تحلیل احساسات رمزارها

همان‌گونه که در شکل (2) مشاهده می‌شود در اینجا برای بهبود دقت و کارایی مدل تحلیل احساسات جنبه‌محور، از ترکیب شبکه عصبی BiLSTM با یک لایه توجه استفاده شده است. ساختار BiLSTM به طور همزمان اطلاعات متنی را از دو جهت

بر یک حوزه واحد (بررسی هتل)، ابهاماتی در مورد کاربرد و اثربخشی مدل پیشنهادی در دامنه‌های مختلف ایجاد می‌کند. بو هوانگ و همکاران [40]، مدل CPA-SA را معرفی کردند که محتوای معنایی یک جمله را در نظر می‌گیرد و وزن کلمات متنی را بر اساس موقعیت مربوط به جنبه تنظیم می‌کند. هدف این روش کاهش تاثیر تفاوت تعداد کلمات در هر دو طرف کلمات جنبه هنگام تعیین قطبیت احساسات است. علاوه بر این، این مدل تاثیر ارتباطات متنی بین جملات در یک سند را به حساب می‌آورد و بازنمایی معنایی را تقویت می‌کند. مدل CPA-SA در برابر چندین مدل پیشرفته بر روی چهار مجموعه داده عمومی مانند توئیت‌های جریانات عمومی SemEval آزمایش شده و پیشرفت‌های قابل توجهی گزارش شده است. این مطالعه همچنین یک تابع زیان اصلاح شده برای حل مشکل عدم تعادل تعداد قطبیت‌ها در تحلیل احساسات پیشنهاد می‌کند. با این حال، در حالی که به مشکل عدم تقارن تا حدی پرداخته شده، تابع وزنی تعیین موقعیت خاص جنبه به اندازه کافی برای زمینه‌های متنی پیچیده‌تر و تخصصی انعطاف‌پذیر نیست. همچنین، اتکای این مطالعه به ابرپارامترهای ثابت و مرزهای آستانه ممکن است سازگاری آن را در سناریوهای مختلف دنیای واقعی محدود کند.

یونژیا و همکاران⁹ [37]، روش جدیدی را برای تحلیل احساسات جنبه‌محور به نام شبکه حافظه زمینه و جنبه (CAMN) پیشنهاد کرده‌اند. این روش از شبکه‌های حافظه عمیق، BiLSTM و مکانیسم‌های توجه چندسری برای گرفتن بهتر ویژگی‌های احساسی در متون، به ویژه متون کوتاه مانند توئیت‌ها یا بررسی‌های محصول، استفاده کرده است.

شیائوفی و همکاران¹⁰ [41]، مدلی را پیشنهاد کردند که از وابستگی‌های متنی گلوبال و محلی برای کاوش در ساختار متن با استفاده از شبکه‌های کانولوشن گراف (GCNs) استفاده می‌کند. مدل GL-GCN از وابستگی نحوی و اطلاعات ترتیبی برای کاوش ساختارهای محلی استفاده کرده، در حالی که یک

⁹ Yanxia et al.

¹⁰ Xiaofei et al.

¹¹ Aspect Sentiment Triplet Extraction (ASTE)

(جلو به عقب و عقب به جلو) پردازش می‌کند، که این ویژگی به مدل اجازه می‌دهد تا وابستگی‌های طولانی‌مدت را در هر دو جهت شناسایی کند. این قابلیت باعث می‌شود که مدل بتواند اطلاعات بیشتری از توالی جملات استخراج کرده و درک بهتری از ساختار معنایی متن‌ها داشته باشد.

جدول (1): تحقیقات پیشین در تحلیل احساسات جنبه‌محور رمزرها با مدل‌های پایه یادگیری عمیق و پیش‌آموزش دیده

| مقاله | نوآوری | چالش تحقیقاتی |
|--------------------------|--|--|
| سیهان و همکاران (2023) | استفاده از معماری CNN-RNN با مکانیسم‌های توجه و بهبود گروهی برای تحلیل احساسات | پیچیدگی محاسباتی بالا و نیاز به داده‌های برچسب‌گذاری شده بیشتر برای آموزش بهتر |
| مودیج و همکاران (2023) | استفاده از تکنیک‌های تعبیه گراف برای استخراج روابط معنایی و نحوی در تحلیل احساسات | نیاز به بررسی تعمیم‌پذیری تکنیک تعبیه گراف به سایر زبانها و دامنه‌ها |
| اسالم و همکاران (2023) | ارائه مدل ترکیبی LSTM-GRU برای تحلیل احساسات و تشخیص عواطف در توییت‌های ارزش‌های دیجیتال | تعادل بخشی به داده‌ها و نیاز به بهینه‌سازی بیشتر برای کاهش زمان محاسباتی |
| داوچف و همکاران (2020) | استفاده از مدل‌های انتقال یادگیری مانند روبرتا برای تحلیل احساسات و پیش‌بینی قیمت بیت‌کوین | چالش‌های مرتبط با عدم تعادل داده‌ها و نیاز به بهبود دقت پیش‌بینی با استفاده از داده‌های ترکیبی بیشتر |
| ژائو و همکاران (2020) | استفاده از GCN برای دریافت وابستگی‌های احساسات بین جنبه‌ای در جملات مختلف | تعمیم‌پذیری در دامنه‌های متفاوت و عدم بررسی استحکام مدل |
| زونیک و همکاران (2021) | استفاده از GCN بر روی نمودارهای وابستگی نحوی برای تحلیل احساسات جنبه‌محور | عدم تعادل طبقاتی در مجموعه دادگان، فاقد بحث مفصل در مورد کارایی محاسباتی |
| لی و همکاران (2021) | مدل برچسب‌گذاری فاصله‌محور (QDSL) برای شناسایی جفت‌های جنبه-نظر | عدم تعادل کلاس در مجموعه داده، تنظیم دقیق مدل پیشنهادی و آزمایش بر روی دادگان با زمینه متفاوت |
| هبت و همکاران (2022) | ترکیب GRU و CNN با XLNet برای تحلیل احساسات دادگان فرانسوی | عدم تنظیم مدل‌های عصبی عمیق نسبت به وظیفه و نوع داده ورودی |
| کبیر و همکاران (2023) | چارچوب ترکیبی شناسایی وابستگی نحوی با لغت‌نامه و رگرسیون لجستیک | ابهام در اثربخشی مدل پیشنهادی بر روی دامنه‌های مختلف |
| بوهوانگ و همکاران (2022) | مدل CPA-SA با استفاده از اطلاعات موقعیت خاص جنبه برای تحلیل احساسات جنبه‌محور | انعطاف‌پذیری ناکافی برای زمینه‌های متنی پیچیده‌تر و اتکا به ابرپارامترهای ثابت و مرزهای آستانه |
| یونژیا و همکاران (2021) | شبکه حافظه زمینه و جنبه (CAMN) برای تحلیل احساسات جنبه‌محور | عدم تعادل کلاس‌های قطبیت و تنظیم ابرپارامترها |
| شیائوفی و همکاران (2021) | مدل GL-GCN با استفاده از وابستگی‌های متنی گلوبال و محلی برای تحلیل احساسات جنبه‌محور | استحکام مدل در سراسر حوزه‌ها و پیچیدگی محاسباتی بررسی نشده |
| چوماکوف و همکاران (2023) | استفاده از GPT برای استخراج سه‌گانه احساسات جنبه (ASTE) | احتمال بالقوه بیش‌برازش در سناریوهای یادگیری با شات محدود |

جدول (2): مقایسه تحقیقات پیشین در تحلیل احساسات رمزارزها با مدل پیشنهادی

| مقاله | مقایسه |
|--------------------------|---|
| سیهان و همکاران (2023) | مدل پیشنهادی با بهره‌گیری از یادگیری انتقالی شات صفر، قابلیت تعمیم‌پذیری بیشتری نسبت به LSTM دارد و می‌تواند با استفاده از تکنیک شات صفر کارایی بیشتری در دامنه‌های مختلف به داشته باشد. |
| مودیج و همکاران (2023) | روش پیشنهادی با استفاده از مدل‌های عمیق و یادگیری انتقالی، در عین کاهش پیچیدگی محاسباتی، دقت بالاتری نسبت به تکنیک‌های تعبیه گراف ارائه می‌دهد و نیازی به برچسب‌گذاری دستی ندارد. |
| اسالم و همکاران (2023) | هر دو مقاله از یادگیری انتقالی شات صفر استفاده می‌کنند، اما مطالعه حاضر با بهره‌گیری از بهینه‌سازی داده‌ها، نرخ داده وفق‌پذیر و تنظیمات بهتر ابرپارامترها، می‌تواند دقت و کارایی بیشتری در تحلیل احساسات جنبه‌محور در حوزه رمزارزها ارائه می‌دهد. |
| داوچف و همکاران (2020) | روش پیشنهادی با استفاده از شبکه‌های عمیق و یادگیری شات صفر، تعمیم‌پذیری و دقت بالاتری نسبت به روش‌های مبتنی بر مدل‌های از پیش آموزش‌دیده بدون تکنیک شات صفر دارد. این رویکردها برای کارایی بهتر نیاز به داده‌های ترکیبی بیشتری در دامنه مبدا دارند. |
| ژائو و همکاران (2020) | مقاله حاضر با استفاده از یادگیری انتقالی شات صفر، قابلیت تعمیم‌پذیری بهتری نسبت به مدل GCN دارد و از آنجا که نیازی به داده‌های برچسب‌گذاری شده ندارد، در دامنه‌های متفاوت بهتر عمل می‌کند. همچنین در این پژوهش، بر روی استحکام مدل و تنظیم دقیق آن تمرکز بیشتری دارد. |
| زونیک و همکاران (2021) | روش پیشنهادی از روش‌های بهینه‌سازی و تعادل‌دهی داده‌ها استفاده می‌کند که مشکلات عدم تعادل طبقاتی را که در روش زونیک وجود دارد را مرتفع می‌کند. همچنین، کارایی محاسباتی مدل پیشنهادی به دلیل استفاده از مدل‌های سبک‌تر مانند دستیل‌برت بهبود یافته است. |
| لی و همکاران (2021) | در این پژوهش، از یادگیری انتقالی شات صفر استفاده شده که نیاز به تنظیم دقیق مدل و تعادل کلاس‌ها را کاهش می‌دهد. این روش همچنین در دامنه‌های مختلف بدون نیاز به داده‌های برچسب‌گذاری شده عملکرد بهتری ارائه می‌دهد. |
| هبت و همکاران (2022) | مقاله حاضر با استفاده از شبکه دیستیل‌برت و یادگیری انتقالی شات صفر، تنظیمات مدل را نسبت به وظیفه و نوع داده بهینه‌تر می‌کند. |
| کبیر و همکاران (2023) | روش پیشنهادی با استفاده از مدل‌های عمیق و یادگیری انتقالی، کارایی بیشتری در دامنه‌های مختلف دارد و وابستگی کمتری به تنظیمات دستی و لغت‌نامه‌های از پیش تعیین شده دارد. |
| بوهوانگ و همکاران (2022) | مقاله حاضر با انعطاف‌پذیری بیشتری در زمینه‌های متنی پیچیده عمل می‌کند و از ابرپارامترهای ثابت و مرزهای آستانه استفاده نمی‌کند، که باعث بهبود عملکرد در دامنه‌های مختلف می‌شود. |
| یونژیا و همکاران (2021) | مقاله حاضر از یادگیری انتقالی شات صفر استفاده می‌کند که مشکلات عدم تعادل کلاس‌های قطبیت و تنظیم ابرپارامترها را بهتر حل می‌کند و نیاز به داده‌های گسترده برچسب‌گذاری شده را کاهش می‌دهد. |
| شیائوفی و همکاران (2021) | این مقاله با تمرکز بر تعمیم‌پذیری و بهینه‌سازی مدل، استحکام بیشتری در دامنه‌های مختلف دارد و پیچیدگی محاسباتی کمتری نسبت به مدل GL-GCN دارد. |
| چوماکوف و همکاران (2023) | مقاله حاضر از یادگیری انتقالی شات صفر استفاده می‌کند که احتمال بیش‌برازش را در سناریوهای یادگیری با شات محدود کاهش می‌دهد و نیازی به تنظیمات پیچیده مدل GPT ندارد. |

می‌شود. برای شباهت سنجی از تابع شباهت پیرسون بر اساس الگوریتم (1) استفاده شده است. سپس دادگانی انتخاب می‌گردند که دارای شباهت کسینوسی بیشتر از 40 درصد با دادگان دامنه هدف باشند.

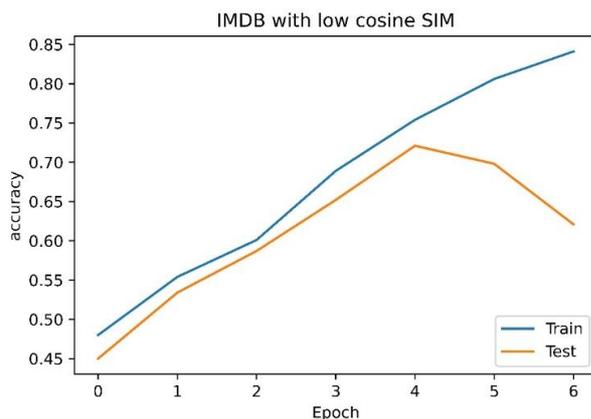
الگوریتم (1): انتخاب دادگان دامنه هدف با معیار شباهت پیرسون

```

Import Pearson similarity
Res= [],
Datasets= {IMDB, Crypto, Sentiment},
DatasetT= Tweet_Crypto
for i=1 to #Datasets
  Res[i] = Pea_Sim (Datasetsi[tweet], Tweet_Crypto)
  if Res[i]>0.4 then
    Datasetsi[flag]= True,
  else
    Datasetsi[flag]= False
  endif
endfor
DatasetsFinal= [x in Datasets[tweets] if Datasets[flag]==True]

```

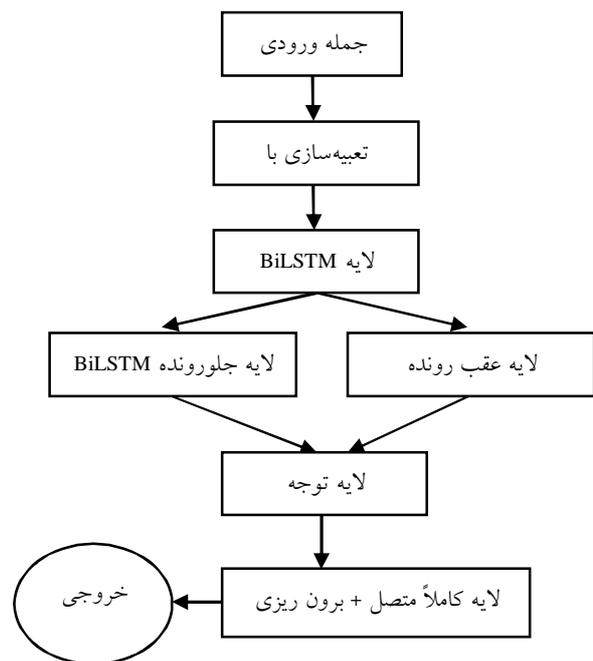
نتایج شکل (4) نشان می‌دهد، در صورتی که مدل با دیتاست‌های که دارای شباهت پیرسون کمتر 40 درصد باشند آموزش داده شود، انتقال دانش می‌تواند تاثیر منفی بر کارایی دامنه هدف داشته باشد.



شکل (4): بروز انتقال منفی با استفاده از دیتاست IMDB در دامنه مبدا

شکل (4) همچنین نشان‌دهنده بروز حالت انتقال منفی و بیش‌برازش شبکه عصبی عمیق می‌باشد. در اینجا افت کارایی شبکه به دلیل انتقال منفی دانش می‌باشد زیرا، مدل برت پایه که

پس از عبور داده‌های متنی از لایه BiLSTM، از لایه توجه برای تقویت بخش‌های مهم‌تر جملات استفاده می‌شود. به بیان دیگر، خروجی نهایی هر طرف BiLSTM به لایه توجه داده شده سپس، این بردارهای خروجی در کنار هم قرار می‌گیرد و مکانیزم توجه روی هر توکن خروجی BiLSTM اعمال شده و سبب غنی‌تر شدن توکن‌ها می‌شود. بنابراین، لایه توجه به مدل امکان می‌دهد که روی کلمات یا عبارات کلیدی که تاثیر بیشتری بر روی تحلیل احساسات دارند، تمرکز بیشتری داشته باشد. این امر به بهبود دقت در شناسایی احساسات مرتبط با جنبه‌های خاص جملات کمک می‌کند. خروجی لایه توجه سپس به لایه‌ای کاملاً متصل تزریق می‌شود تا فرآیند دسته‌بندی نهایی انجام گیرد و احساسات مرتبط با جنبه به یکی از قطبیت‌های مثبت، منفی یا خنثی تخصیص داده شود. فرآیند عملکرد BiLSTM و لایه توجه در شکل (3) نشان داده شده است.



شکل (3): فرآیند پردازش جملات در BiLSTM و لایه توجه

در شکل (3) فرآیند پردازش جملات بعد از ورود به لایه‌های BiLSTM و توجه نشان داده شده است. همچنین در این پژوهش، در ابتدا شباهت دادگان انتخاب شده جهت آموزش شبکه عصبی پیش‌آموزش‌دیده با دادگان رمزازها مقایسه

بهبترین ابرپارامترها برای شبکه عصبی عمیق با پایس کردن معیارهای کارایی شبکه انتخاب شود. معیاری که به طور معمول در مقالات برای مانتیور کردن انتخاب می‌شود، تابع نرخ خطا می‌باشد. در این مقاله، نرخ خطای داده اعتبارسنجی برای پایس کردن انتخاب شده است. در الگوریتم (2) نحوه انتخاب ابرپارامترها نشان داده شده است.

الگوریتم (2): استفاده از رویکرد گریدی برای انتخاب ابرپارامترهای شبکه عصبی عمیق

```
Optimizer = {'SGD', 'Adam', 'AdamW'}
Dropout = {0.1, 0.2, 0.3, 0.4}
Batch Size = {16, 32, 64, 128}
Def grid_search (Optimizer, Dropout, Batch Size)
  For i=0 to Number_of_Epochs
    Calc_Lossi = (Optimizeri, Dropouti, Batch Sizei)
  Return (Min (Calc_Loss))
endDef
```

در این پژوهش سه ابرپارامتر تابع بهینه‌سازی، نرخ برون‌ریزی شبکه و طول بچ با استفاده از روش گریدی تعیین شده است. این سه ابرپارامتر تاثیر بسیار زیادی بر کارایی شبکه‌های عصبی دارند [43]. نتایج آزمایشات نشان داده است انتخاب سه ابرپارامتر به شرح جدول (3) بهترین کارایی را داشته است. از این رو برای پیاده‌سازی کلیه مدل‌ها در این پژوهش از مقادیر ابرپارامترهای جدول (3) استفاده می‌شود.

جدول (3): ابرپارامترهای منتخب

| مقدار | ابروپارامتر |
|-------|----------------|
| 64 | طول بچ |
| AdamW | تابع بهینه‌ساز |
| 0/2 | نرخ برون ریز |

همچنین، نرخ یادگیری شبکه عصبی در این پژوهش برخلاف اغلب مقالات [4]، به صورت پویا و با استفاده تکنیک وارم‌آپ¹² تعیین می‌گردد. تکنیک وارم‌آپ [44]، رویکردی نوین است که

از دادگان احساسات توثیت‌های بیت‌کوین به عنوان دادگان مبدا استفاده کرده، دقت 85% بعد از 6 دور دست یافته است. پس می‌توان نتیجه گرفت، کاهش کارایی مدل به دلیل پارامترهای شبکه نبوده است. بنابراین انتخاب مجموعه دادگان غنی در دامنه مبدا اگرچه می‌تواند سبب افزایش کارایی مدل در دامنه هدف گردد، با این حال، انتخاب مجموعه دادگان نامشابه یا با شباهت کمتر می‌تواند سبب انتقال منفی دانش شود [15].

در ادامه مجموعه دادگان منتخب جهت استفاده در شبکه‌های پیش‌آموزش دیده دامنه مبدا، به یکدیگر الحاق می‌گردد. سپس عملیات پیش پردازش متن روی آنها انجام می‌شود. این عملیات شامل نشانه‌گذاری، حذف کلمات توقف، بن‌سازی، یکسازسازی متن و حذف کارکترهای زائد مانند پیوندها می‌باشد. سپس داده‌های پاکسازی شده با تکنیک SMOTE متوازن شده و در نهایت به شبکه عصبی عمیق پیش‌آموزش دیده دیستیل‌برت در دامنه مبدا تزریق می‌شود.

SMOTE تکنیکی آماری برای متعادل‌سازی دادگان است که سبب می‌شود حساسیت مدل به کلاسی خاص در دادگان کم شود. به بیان رسمی، برای یک نمونه کلاس اقلیت x SMOTE یک نمونه تصادفی \bar{x} را از k نزدیکترین همسایه x انتخاب می‌کند. سپس نمونه‌ای جدید به اسم x_{new} با درون‌یابی بین x و \bar{x} توسط رابطه (3) ایجاد می‌شود.

$$x_{new} = x + \lambda \times (\bar{x} - x) \quad (3)$$

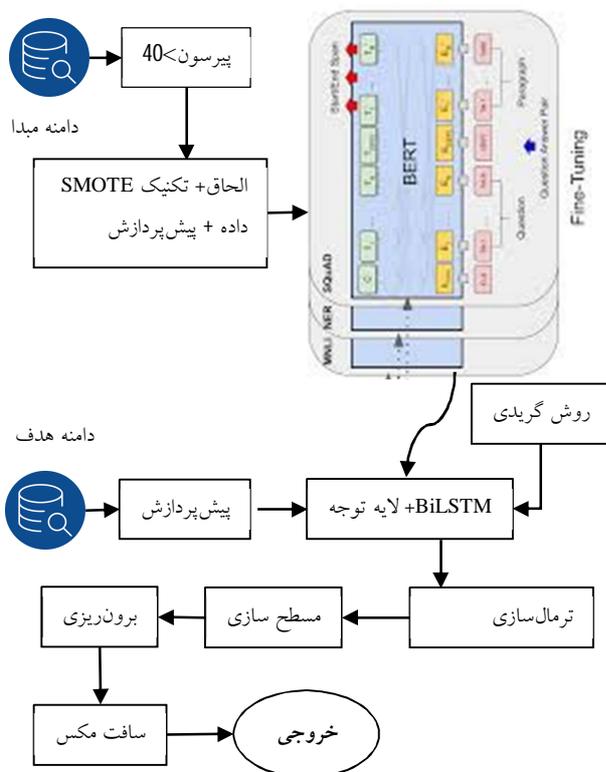
در رابطه (3)، λ عددی تصادفی بین 0 و 1 است. این فرآیند تا زمانی که کلاس‌ها متعادل شوند تکرار می‌شود. SMOTE منجر به ارائه مجموعه دادگانی متعادل‌تر می‌شود که برای آموزش مدل‌های یادگیری ماشین بسیار مهم است. زیرا، داده‌های نامتعادل می‌تواند منجر به مدل‌های مغرضانه شود که عملکرد ضعیفی در کلاس اقلیت دارند. این تکنیک توانایی مدل را برای تعمیم و عمل موثر بر روی داده‌های دیده نشده افزایش می‌دهد. همچنین در این پژوهش برای انتخاب ابرپارامترهای مدل عصبی عمیق پایه BiLSTM در دامنه هدف، از روش گریدی استفاده شده است. روش گریدی این امکان را می‌دهد که مجموعه‌ای از

¹² Warmup

2- انجماد برخی از لایه‌ها: روش دیگر تنظیم مدل‌های پیش‌آموزش دیده، آموزش جزئی آنها است. می‌توان وزن لایه‌های اولیه را ثابت نگه داریم و فقط لایه‌های بالاتر را دوباره آموزش دهیم. این روش پیشرفته‌تر است و می‌توان آزمایش کرد بسته به غنای دادگان دامنه هدف، چند لایه باید منجمد شود و چند لایه باید آموزش داده شود.

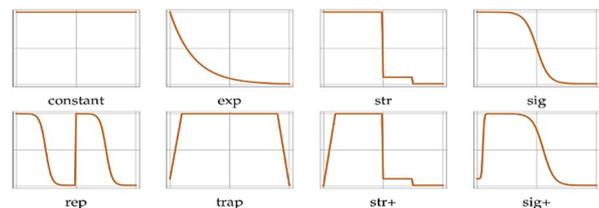
3- کل شبکه را منجمد کنیم: می‌توان تمام لایه‌های شبکه را منجمد کرد و چند لایه شبکه عصبی (پایه یا عمیق) به آن متصل کنیم و تنها شبکه پایانی را دوباره آموزش دهیم. در این روش تنها وزن لایه‌های نهایی پیوست شده در فرایند آموزش بروز رسانی می‌شود.

در این پژوهش دو روش تنظیم دقیق آموزش و انجماد کل شبکه با یکدیگر مقایسه شده است. چارچوب کلی مدل پیشنهادی در شکل (6) نشان داده شده است.



شکل (6): معماری پیشنهادی یادگیری انتقالی عمیق مبتنی بر شبکه‌های تنظیم دقیق پیش‌آموزش دیده برت

از اندازه گام در شیب نزولی مشتق‌های گرادینان را در در ابتدای آموزش شبکه کوچک انتخاب می‌کند. در ادامه، نرخ یادگیری به صورت خطی یا غیرخطی در چند دوره افزایش یافته و سپس به سمت صفر میل می‌کند. مدل‌های یادگیری ماشین در ابتدای آموزش از مدل ایده‌آل فاصله زیادی دارد. بنابراین، نرخ یادگیری بیش از حد بزرگ باعث بی‌ثباتی عددی می‌شود. در نتیجه، آموزش شبکه عصبی با دقت بالا در دوره‌های اول این امکان را می‌دهد در مراحل میانی نرخ یادگیری افزایش یافته در نتیجه همگرایی شبکه تسریع گردد [44]. در شکل (5) متدهای مختلف وارم‌آپ را نشان داده شده است.



شکل (5): متدهای مختلف وارم‌آپ [44]

در این پژوهش از روش sig+ استفاده شده است. تکنیک وارم‌آپ در واقع نوعی روش تبرید فلزات¹³ است که در ابتدا نرخ یادگیری به سمت صعودی میل کرده (افزایش دما)، در مراحل میانی این نرخ ثابت نگه داشته شده و سپس به سمت صفر (انجماد) میل می‌کند. در این پژوهش همچنین دو تکنیک تنظیم دقیق برای شبکه‌های پیش‌آموزش دیده مبتنی بر برت یعنی دستیل‌برت و XLNet با یکدیگر مقایسه شده است. به طور کلی سه روش برای تنظیم شبکه‌های عصبی پیش‌آموزش دیده به شرح زیر وجود دارد [5]:

1- آموزش کل شبکه: می‌توان کل شبکه پیش‌آموزش دیده را با دادگان دامنه هدف آموزش داد سپس خروجی را به لایه‌ای متراکم با فعال‌سازهای سیگموئید یا سافت‌مکس متصل کرد. در این حالت، خطا در شبکه منتشر شده و وزن‌های شبکه بر اساس مجموعه داده دامنه هدف بروز می‌شود.

¹³ Annealing

مقایسه شبکه‌های عصبی پایه و پیش‌آموزش دیده کاربرد دارد. برای مثال در مواردی که می‌خواهیم کارایی شبکه برت+BiLSTM را با BiLSTM مقایسه کنیم به راحتی می‌توان بدون کدنویسی مجدد داده‌های منتقل شده از برت را نادیده گرفت. در این پژوهش کارایی شبکه‌های عصبی عمیق پایه، پیش‌آموزش دیده و ترکیبی براساس معماری ارائه شده در شکل (4) با یکدیگر مقایسه شده است. شبکه‌های مورد مطالعه این پژوهش در جدول (4) ذکر شده است.

جدول (4): معماری‌های عصبی عمیق پیاده‌سازی شده

| نام مدل | نوع مدل |
|--------------------------|---------------------------------|
| LSTM BiLSTM CNN | شبکه عصبی عمیق پایه |
| برت پایه | شبکه عمیق پیش‌آموزش دیده |
| CNN+BiLSTM | شبکه عصبی عمیق پایه ترکیبی |
| BiLSTM+برت (پایه، دستیل) | شبکه عمیق پیش‌آموزش دیده ترکیبی |

4. نتایج آزمایشات تجربی

در این بخش در ابتدا معیارهای ارزیابی شبکه‌های عصبی عمیق برای تحلیل احساسات جنبه‌محور رمزارزها را توضیح داده شده؛ سپس، دادگان مورد استفاده در دامنه هدف و مبدا مورد بررسی قرار گرفته و در نهایت نتایج آزمایشات تجربی تشریح می‌شود.

1.4. معیارهای ارزیابی

اولین معیار مرسوم در ارزیابی روش‌های طبقه‌بندی متن مانند تحلیل احساسات معیار صحت می‌باشد [45]. این معیار با تقسیم تعداد پیش‌بینی‌های صحیح بر تعداد کل پیش‌بینی‌ها محاسبه می‌شود. معیار صحت از طریق رابطه (4) به دست می‌آید.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

در رابطه (4)، TP تعداد مثبت واقعی، TN منفی واقعی، FP مثبت کاذب و FN تعداد منفی کاذب پیش‌بینی شده می‌باشد.

همانگونه که در شکل (6) نشان داده شده است، در ابتدا دادگان دامنه هدف به وسیله معیار شباهت پیرسون انتخاب می‌شوند. سپس، توثیت و برچسب‌های دادگان منتخب که دارای شباهت بیشتر از 40 درصد با دادگان رمزارزها هستند، با یکدیگر الحاق شده و دادگان دامنه مبدا را تشکیل می‌دهند. در ادامه برخی از عملیات پیش‌پردازش معمول در پردازش زبان طبیعی مانند ریشه‌یابی، حذف کلمات توقف، یکسان‌سازی حروف، حذف عبارات زائد مانند پیوندها و نشانه‌گذاری بر روی آنها انجام شده و بعد از اعمال تکنیک متعادل‌سازی داده با SMOTE به شبکه پیش‌آموزش دیده از خانواده برت تزریق می‌شود. ذکر این نکته ضروری است که برای نشانه‌گذاری هم در دامنه مبدا و هم در دامنه هدف از نشانه‌گذار مبتنی بر برت استفاده می‌شود. برای مثال اگر شبکه پیش‌آموزش دیده، XLNet باشد از نشانه‌گذار XLNet استفاده می‌شود. در ادامه دادگان دامنه مبدا برای استخراج دانش به یکی از انواع برت تزریق می‌گردد. در این پژوهش از سه نسخه برت پایه، دستیل برت و XLNet استفاده شده است. سپس برای انتخاب ابرپارامترهای شبکه پیش‌آموزش دیده از تکنیک گریدی (الگوریتم (2)) استفاده شده است. ویژگی‌های استخراج شده از دامنه مبدا با دو روش تنظیم به شبکه BiLSTM با ترکیب لایه توجه در دامنه هدف تزریق می‌شود. در این پژوهش برای مقایسه و پاسخ به سوالات پژوهشی از دو روش تنظیم آموزش کل شبکه و انجمادکل وزن‌های شبکه دامنه مبدا، استفاده شده است.

در دامنه هدف نیز بعد از پیش‌پردازش دادگان رمزارز، داده‌ها نشانه‌گذاری شده و به شبکه عصبی عمیق BiLSTM تزریق شده است. سپس ویژگی‌های استخراج شده از شبکه پیش‌آموزش دیده دامنه هدف و دامنه مبدا با یکدیگر الحاق می‌شود. وجود لایه الحاق سبب می‌شود شبکه‌های که از نظر ساختار (برای مثال تعداد لایه‌های پنهان) با یکدیگر یکسان نیستند بتوانند به معماری یکسانی دست یابند. همچنین لایه الحاق در این پژوهش می‌تواند مانند یک سوئیچ عمل کند، یعنی هرگاه نیاز به دانش دامنه هدف نبود و یا شبکه پیش‌آموزش دیده تغییر کرد می‌توان این تغییرات را بدون دستکاری دامنه هدف اعمال کرد. این امر بیشتر در موارد

کلاس‌ها و t نشان‌دهنده برچسب واقعی کلاس i -ام می‌باشد. آخرین معیار، نمودار منحنی AUC-ROC است که نشان‌دهنده دقت مدل در تشخیص هر کلاس قطبیت می‌باشد. این معیار در صورت رابطه (9) تعریف شده است.

$$ROC - AUC = \int TP(FP) dFP \quad (9)$$

2.4. معرفی دادگان مورد ارزیابی

در این پژوهش از سه مجموعه دادگان مورد بررسی جهت استفاده در دامنه مبدا، در جدول (5) تشریح شده‌اند.

جدول (5): دادگان دامنه مبدا

| عنوان | مثبت | منفی | خنثی | مجموع |
|----------------------------|--------|--------|--------|--------|
| IMDB | ۱۲۵۰۰ | ۱۲۵۰۰ | 0 | ۵۰,۰۰۰ |
| تحلیل احساسات توییتر | ۲۲۸۰۷ | ۲۱,۱۰۹ | ۱۸,۳۰۶ | ۶۲,۲۲۲ |
| احساسات توییت‌های بیت‌کوین | ۲۲,۹۳۷ | ۵,۹۸۳ | ۲۱,۹۳۲ | ۵۰,۸۵۲ |

در زیر این دادگان معرفی شده است.

دادگان IMDB: مجموعه داده بزرگ نقد فیلم‌های وبگاه IMDB. برای طبقه‌بندی احساسات دو یا چند کلاس پیشنهاد شده است. این مجموعه دادگان شامل مجموعه‌های آموزشی و ارزیابی برچسب‌دار می‌باشد. یکی از ویرایش‌های دو کلاس این دادگان، شامل ۲۵,۰۰۰ نمونه آموزشی و ارزیابی است. این ویرایش، دارای ۱۲,۵۰۰ برچسب مثبت و منفی برای داده‌های ارزیابی و آموزشی می‌باشد. این دادگان به دلیل شباهت پایین‌تر از 40 درصد به عنوان داده میان‌دامنه‌ای برای آموزش مدل استفاده نشده است.

دادگان تحلیل احساسات توییتر: این مجموعه داده شامل بیش از 62 هزار توییت است که با استفاده از Api توییتر در سال 2022 استخراج شده است. توییت‌ها دارای برچسب مثبت، منفی و خنثی می‌باشد و از آن می‌توان در وظیفه تحلیل احساسات استفاده کرد. این مجموعه دادگان به‌طور مستقیم با موضوع تحقیق در ارتباط است زیرا همانند داده‌های اصلی مقاله شامل

یکی دیگر از معیارهای مرسوم، معیار دقت¹⁴ است. این معیار برای گزارش تعداد کل پیش‌بینی‌های صحیح شبکه می‌باشد. این معیار با تقسیم تعداد پیش‌بینی‌های مثبت توسط شبکه بر تعداد کل برچسب‌های مثبت در دادگان برای هر قطبیت مثبت، منفی و خنثی به دست می‌آید. معیار دقت به صورت رابطه (5) تعریف می‌شود.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

معیار بعدی فراخوان¹⁵ است. فراخوان نشان‌دهنده تمام برچسب‌های مثبتی است که توسط شبکه به درستی پیش‌بینی شده‌اند. این معیار با تقسیم برچسب‌های مثبت پیش‌بینی‌شده توسط شبکه با مجموع تعداد کل برچسب‌های مثبت و تعداد منفی‌های کاذب پیش‌بینی شده به دست می‌آید. معیار فراخوان به صورت رابطه (6) تعریف می‌شود.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

معیار دیگر F1-Score است. این معیار ترکیبی از دقت و فراخوان می‌باشد و به صورت رابطه (7) تعریف می‌گردد.

$$F1 - SCORE = \frac{TP}{TP + \frac{1}{2}(FN + FP)} \quad (7)$$

معیار بعدی، تابع ضرر شبکه می‌باشد که آنتروپی متقابل طبقه‌بندی¹⁶ نام دارد. هر میزان این معیار به سمت صفر میل کند، نشان می‌دهد شبکه از کارایی بالاتری برخوردار است. به این معیار خطای سافت مکس نیز می‌گویند و از طریق رابطه (8) به دست می‌آید.

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}}, \quad (8)$$

$$CE = -\sum_i^C t_i \log(f(s)_i)$$

در رابطه (8)، S امتیاز اکتسابی توسط تابع سافت مکس (f) برای کلاس i -ام است. در این رابطه، تابع ضرر آنتروپی متقابل طبقه‌بندی را با CE نشان داده که در آن C نشان‌دهنده تعداد

¹⁴ Precision

¹⁵ Recall

¹⁶ Categorical cross-entropy

محققان می‌توانند از این مجموعه داده در تحلیل احساسات مبتنی بر جنبه نیز استفاده کنند. یکی دیگر از کاربردهای این مجموعه داده، بررسی تاثیر توئیتهای تاثیرگذاران متخصص بر روند بازار رمزارزها است. زیرا بر اساس علم اقتصاد رفتاری [33]، نظرات معامله‌گران بر روند بازار سهام تاثیرگذار است. این دیتاست شامل 80 هزار سطر می‌باشد که به ترتیب دارای 39,526 هزار برچسب مثبت، 28,715 برچسب منفی و 11,759 برچسب خنثی است. در هر دو مجموعه دادگان در دامنه مبدا و هدف، 80 درصد برای آموزش شبکه و 20 درصد برای ارزیابی استفاده شده است. همچنین، برای ارزیابی مدل مانند هوانگ و همکاران [40] از مجموع دادگان SemEval که شامل توئیتهای عمومی در حوزه‌های مختلف مانند رستوران و لپ‌تاپ است، استفاده شده است. دیتاست‌های مورد ارزیابی در این پژوهش [50] SemEval-2016Task 5 و [49] SemEval-2015 Task 12 است که از این پس به صورت اختصاری Res2015 و Res2016 نامگذاری می‌شود. در جدول (6) این مجموعه دادگان معرفی شده است.

جدول (6): تنظیمات شبکه‌های عصبی عمیق

| دادگان | آموزش | | اعتبارسنجی | | | ارزیابی | | | |
|---------|-------|------|------------|------|------|---------|------|----|-----|
| | مثبت | منفی | مثبت | خنثی | منفی | مثبت | منفی | | |
| Res2015 | 808 | 29 | 228 | 147 | 5 | 44 | 340 | 25 | 159 |
| Res2016 | 706 | 454 | 406 | 191 | 9 | 60 | 474 | 29 | 127 |

3.4. نتایج آزمایشات مدل‌های عصبی عمیق پایه،

پیش‌آموزش دیده و ترکیبی

در ابتدا شباهت محتوایی داده‌های جمع‌آوری شده رمزارزها با سه مجموعه دادگان دامنه مبدا، ارزیابی شده که کمترین شباهت برابر با 36/5 درصد مربوط به مجموع دادگان IMDB می‌باشد. بنابراین این مجموعه داده به دلیل شباهت کمتر از 40 درصد به عنوان داده میان‌دامنه‌ای استفاده نمی‌شود. در جدول (7) نتایج این ارزیابی نشان داده شده است.

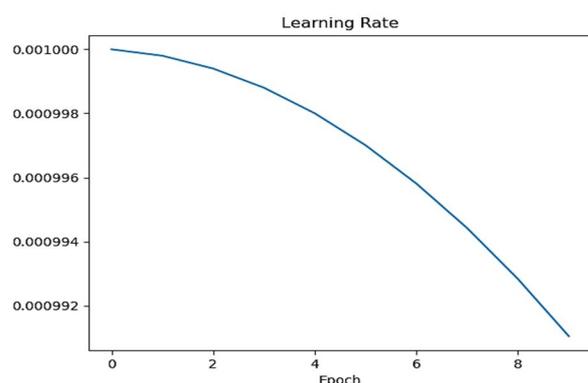
توئیتهای کاربران است. بنابراین، به بهبود دقت مدل در تحلیل احساسات توئیتهای مربوط به رمزارزها کمک می‌کند. در این دادگان متوسط طول جملات 18,99 کلمه و حداکثر طول بزرگترین جمله 198 کلمه است. همچنین انحراف معیار برابر 14,3 و تعداد کلمات یکتا برابر 34 هزار کلمه می‌باشد. این مجموع دادگان به مدل کمک می‌کند با توئیتهای تقریباً بزرگ نیز آشنا شود.

دادگان توئیتهای احساسات بیت‌کوین: این دادگان شامل توئیتهای با هشتگ بیت‌کوین بوده که از آپریل تا سپتامبر سال 2018 جمع‌آوری شده است. این مجموع دادگان شامل بیش از 50 هزار توئیتهای دارای برچسب مثبت، منفی و خنثی می‌باشد. در این دادگان متوسط طول جملات 17,16 و حداکثر طول بزرگترین جمله 39 کلمه است. همچنین انحراف معیار برابر 13,2 و تعداد کلمات یکتا تقریباً برابر 55 هزار کلمه می‌باشد. این دادگان به دلیل ارتباط مستقیم با موضوع اصلی تحقیق، یعنی تحلیل احساسات جنبه‌محور در حوزه رمزارزها، انتخاب شده و شامل توئیتهایی است که نظرات کاربران درباره بیت‌کوین را جمع‌آوری و برچسب‌گذاری کرده است. این مجموعه دادگان تاکنون در مقالات متعددی استفاده شده است [46]-[48].

در دامنه هدف نیز از مجموعه دادگان رمزارزها جمع‌آوری شده از توئیتر استفاده شده است [6]. مجموعه دادگان دامنه هدف مربوط به نظرات بیش از 50 تاثیرگذار (شخص یا شرکت) در مورد بیش از 40 رمزارز است. این مجموعه داده از طریق API توئیتر به مدت هشت ماه از سال 2021 تا 2022 جمع‌آوری شده است. هدف از این مجموعه داده جمع‌آوری توئیتهای تخصصی در مورد رمزارزهای مختلف است. در مجموعه داده‌های موجود در زمینه رمزارزها، تخصص کاربران توئیتهای زنده نادیده گرفته شده و توئیتهای تنها با استفاده از هشتگ استخراج شده‌اند. لذا، مجموعه داده استفاده شده در این مقاله کاربردهای زیادی در یادگیری ماشین و متن‌کاوی دارد. محققان می‌توانند از این مجموعه داده در تجزیه و تحلیل احساسات، یادگیری انتقالی در تحلیل احساسات و طبقه‌بندی متن استفاده کنند. همچنین، به دلیل ذخیره توئیتهای مربوط به هر رمزارز در جداگانه،

جدول (9): سخت افزار مورد استفاده

| مشخصات | سخت افزار |
|--------------|-------------------|
| Colab Google | محیط برنامه نویسی |
| 16.0 GB | حافظه داخلی |
| 3 پایتون | زبان برنامه نویسی |
| Telsa T4 | پردازنده گرافیکی |
| 11,2 | نسخه کودا |



شکل (7): تغییرات نرخ یادگیری

نتایج جدول (10)، نشان دهنده کارایی مدل‌های XLNet +RNN، +BiLSTM، دستیل برت (با انجماد و بدون انجماد لایه‌ها)، برت پایه و برت + BiLSTM می‌باشد.

جدول (10): دقت مدل‌های پیش‌آموزش دیده ساده و ترکیبی مبتنی بر برت بر روی دادگان حساسات توییت‌های بیت‌کوین

| نام مدل | صحت | دقت | فراخوان | F1-Score |
|-------------------------------|-------|-------|-----------------------|----------|
| LSTM | 0/651 | 0/623 | 0/643 | 0/646 |
| BiLSTM | 0/723 | 0/651 | 0/692 | 0/711 |
| CNN+ BiLSTM | 0/781 | 0/695 | 0/7 | 0/738 |
| XLNet+ CNN | 0/869 | 0/846 | 0/832 | 0/85 |
| | 0/751 | | انتقال منفی (با IMDB) | |
| BiLSTM+ (freeze) distilbert | 0/876 | 0/846 | 0/825 | 0/846 |
| BiLSTM +distilbert (unfreeze) | 0/882 | 0/857 | 0/841 | 0/861 |
| برت پایه | 0/806 | 0/786 | 0/774 | 0/78 |
| برت پایه+ | 0/817 | 0/792 | 0/785 | 0/80 |
| BiLSTM | | | | |

جدول (7): محاسبه معیار شباهت کسینوسی

| امتیاز (%) | دادگان |
|------------|---------------------------------|
| 53/2 | Pear (sentiment, My_Crypto) |
| 64/1 | Pear (Crypto Tweets, My_Crypto) |
| 36/5 | Pear (IMDB, My_Crypto) |

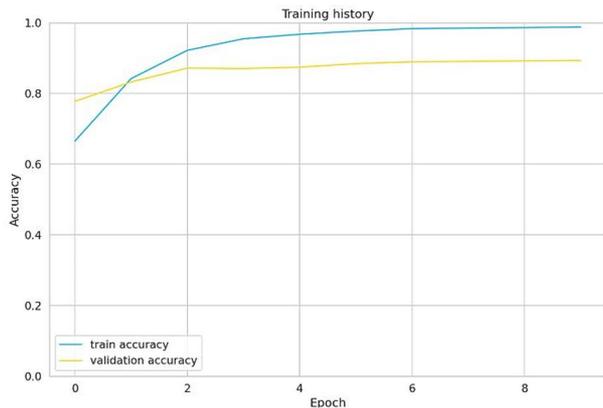
در این پژوهش برای مدل‌های ترکیبی پیش‌آموزش دیده مبتنی بر برت (XINet، پایه و دستیل)، شبکه‌های عصبی عمیق پایه (CNN, BiLSTM) و شبکه‌های عصبی عمیق پایه ترکیبی (CNN+BiLSTM) از تنظیمات ذکر شده در جدول (8) استفاده شده است.

جدول (8): تنظیمات شبکه‌های عصبی عمیق

| مقدار | نام پارامتر |
|-----------------------------------|----------------------------|
| 10 | تعداد دور |
| AdamW | تابع بهینه‌ساز |
| وقفی (Warmup) | نرخ یادگیری |
| Model 1 = True Model 2 = False | انجماد لایه‌ها |
| Softmax/ Sigmoid | توابع فعال‌ساز |
| ۳۲،۶۴،۱۲۸ | طول بچ |
| ۳۲،۶۴ | تعداد لایه‌های مخفی BiLSTM |
| 0/2 | نرخ برون‌ریزی |
| 2×۲،۳×۳،۵×۵ | طول کرنل CNN |
| آنزوی متقابل طبقه‌بندی | تابع ضرر |

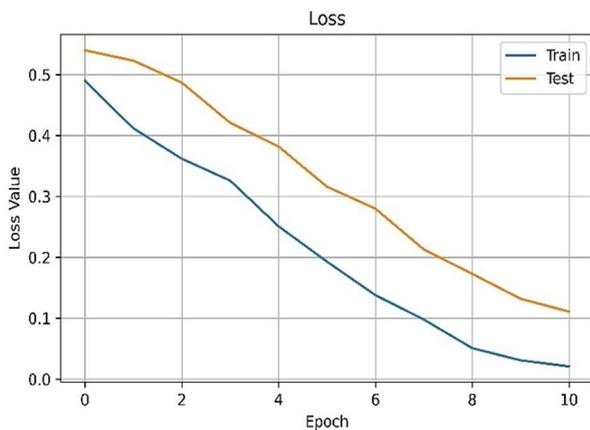
تنها تفاوت در پیاده‌سازی شبکه‌ها این است که شبکه‌های عصبی عمیق پایه و پایه-ترکیبی تنظیمات مرتبط با انجماد لایه‌ها وجود ندارد. همچنین تنظیمات مرتبط با CNN مربوط به معماری ترکیبی CNN+BiLSTM است. روش پیشنهادی، بر روی سخت‌افزار با مشخصات جدول (9)، پیاده‌سازی شده است.

همان‌گونه که ذکر شد در این پژوهش از نرخ یادگیری وفق‌پذیر وارم‌آپ استفاده شده است. در نمودار شکل (7)، میزان کاهش نرخ یادگیری در ده دور نشان داده شده است.



شکل (8): معیار دقت شبکه عصبی عمیق ترکیبی BiLSTM+ دستیل برت

در شکل (9)، نرخ از دست رفتگی داده شبکه دستیل برت + BiLSTM در حالت آموزش و ارزیابی در ده دور نشان داده شده است.



شکل (9): نرخ از دست رفتگی داده مدل دستیل برت + BiLSTM

همچنین، با اجرای مدل BiLSTM+دستیل برت بر روی مجموع دادگان Res2015 و Res2016 نتایج جدول (11) به دست آمده است. این نتایج در مقایسه با پژوهش‌های مشابه امیدوارکننده می‌باشد. ما مدل پیشنهادی را با پژوهش‌های زیر مقایسه کرده‌ایم.

- CPA-SA [40]: این مدل از دو تابع وزنی موقعیت متن، برای تنظیم اهمیت کلمات متنی بر اساس موقعیت آنها نسبت به کلمات جنبه در جمله‌ها استفاده می‌کند. این رویکرد به کاهش تاثیر عدم تعادل کلمات بر قضاوت احساسات ناشی از تعداد کلمات متفاوت در دو طرف کلمات جنبه کمک می‌کند. با استفاده از لایه‌های GRU

همان‌گونه که در جدول (10)، نشان داده شده، استفاده از دادگان IMDB به دلیل عدم شباهت دادگان مبدا به هدف، سبب بروز پدیده یادگیری انتقالی منفی شده است. همچنین، مدل‌های پیش‌آموزش دیده برت منجمد و بدون انجماد، تقریباً به نتایج مشابهی دست یافته‌اند. دلیل این امر، می‌تواند شباهت متن‌های مجموع دادگان دامنه مبدا و هدف باشد. دلیل دیگر، استفاده از مدل BiLSTM است. این مدل به دلیل دو جهت بودن و دسترسی به توالی‌های قبلی و بعدی [51] برای بالابردن کارایی در وظیفه تحلیل احساسات استفاده شده است. با این حال، در کاربردهای دنیای واقعی آموزش کل لایه‌های شبکه‌های عصبی عمیق بسیار هزینه‌بر می‌باشد. بنابراین می‌توان با تحمل درصد ناچیزی از خطا از تکنیک‌های تنظیم دقیق مبنی بر انجماد لایه‌های استخراج‌گر ویژگی استفاده کرد. همچنین، می‌توان به لایه‌های که باید منجمد شود به صورت یک ابرپارامتر نگاه کرد. به عبارت دیگر، با در نظر گرفتن داده‌های دامنه هدف به صورت وفق‌پذیر برخی لایه‌ها را منجمد و برخی دیگر را آموزش داد. این امر سبب ایجاد موازنه بین سرعت آموزش و دقت شبکه‌های عصبی عمیق می‌گردد که در کارهای آینده مورد توجه قرار می‌گیرد. از سوی دیگر، می‌توان متصور بود در شبکه XLNet نیز اگر شبکه عصبی بازگشتی با BiLSTM جایگزین شود به نتایج بهتری دست بیابیم. زیرا XLNet مشکل فرض استقلال در برت را حل کرده و همچنین مدل BiLSTM دارای مشکل وابستگی طولانی مدت مانند RNN نیست. فرض استقلال، تصور می‌کند هر توکن پوشانده‌شده به تمام توکن‌های ورودی وابسته است اما از سایر توکن‌های پوشانده‌شده مستقل می‌باشد. همچنین در موارد ترکیب شبکه‌های عصبی عمیق پایه می‌بینیم، شبکه عصبی ترکیبی CNN+BiLSTM بهترین عملکرد را نسبت به سایر شبکه‌ها بدست آورده است. این امر به دلیل ویژگی قدرتمند شبکه CNN در استخراج ویژگی‌ها می‌باشد. همچنین شبکه عصبی عمیق BiLSTM نیز به دلیل پردازش توالی‌ها در دو جهت سبب می‌گردد کارایی مدل افزایش یابد. شکل (8)، میزان صحت شبکه دستیل برت + BiLSTM در حالت آموزش و ارزیابی نشان داده شده است.

ASGCN، در یک مجموعه داده مطلوب و در مجموعه داده دیگری کاهش یافته است. در مجموع مدل پیشنهادی به صورت میانگین دقت و F1 را بر روی مجموعه داده Res2015 در حدود 4 درصد و F1 را بر روی مجموعه داده Res2016 در حدود 1 درصد افزایش داده است. در شکل (10) معیار ROC-AUC را برای درک بهتر کارایی مدل BiLSTM+دستیل برت نشان می‌دهد.

جدول (11): مقایسه مدل BiLSTM+دستیل برت و مقالات پیشین

| مدل‌ها | Res2015 | | Res2016 | |
|----------------------|---------|-------|---------|-------|
| | دقت | F1 | دقت | F1 |
| CPA-SA | 79/57 | 60/26 | 88/8 | 71/47 |
| TAS-BERT | - | 70/42 | - | 70/66 |
| BERT-pair-NLI-B | - | 70/78 | - | 80/25 |
| ASGCN | 79/34 | 60/78 | 88/69 | 66/64 |
| THA + STN | - | 71/46 | - | 73/61 |
| MSRL-Net | - | 92/68 | - | 86/58 |
| +BiLSTM دستیل برت | 87/61 | 93/1 | 85/14 | 86/86 |

در شکل (10)، کلاس «0» برابر با قطب مثبت، کلاس «1» برابر با قطب خنثی و کلاس «2» برابر با قطب منفی است. در مجموعه داده‌های Res2015 و Res2016، کلاس خنثی کمترین تعداد توییت را دارد. شکل (10) نشان می‌دهد که مدل هر سه قطب مثبت، منفی و خنثی را برای جنبه‌های مختلف متن تشخیص داده است. همچنین، نرخ ROC-AUC برای مجموعه داده Res2015 برابر با 85/4 و برای Res2016 برابر با 87/3 است.

ما همچنین از نمودار نقشه حرارتی در شکل (11) برای نشان دادن تاثیر لایه توجه بر روی متنی از دادگان SemEval استفاده کردیم. هر سلول وزن لایه‌های توجه مدل را نشان می‌دهد که به نشانه‌ها (کلمات) خاص طول پردازش متن اختصاص داده شده است. این وزن‌ها نشان می‌دهند که مدل کدام بخش از متن را هنگام تعیین احساسات بر اساس جنبه‌های خاص مهم‌تر می‌داند.

دوطرفه (BiGRU) در هر دو سطح تک و چند جمله‌ای، این مدل تاثیر متنی هر جمله در سند را بر قطبیت احساسات جملات فردی به تصویر می‌کشد. علاوه بر این، این مقاله با تجزیه و تحلیل ویژگی‌های توزیع نمونه‌های چالشی و معرفی یک تابع ضرر جدید، راه‌حلی برای مساله عدم تعادل قطبیت‌ها در تحلیل احساسات ارائه می‌دهد.

- TAS-BERT [52]: این روش از مدل زبانی از پیش آموزش دیده برتا برای نشان داده وابستگی بین اهداف، جنبه‌ها و احساسات استفاده می‌کند.
- ASGCN [40]: از شبکه پیچشی گراف محور (GCN) برای گرفتن اطلاعات معنایی و ایجاد ارتباط بین عناصر استفاده می‌کند.
- THA+STN [52]: THA + STN، مدلی عصبی برای تشخیص احساسات هدفمند (TD) است. این مدل دارای دو لایه توجه خطی و لایه‌های کاملاً متصل (FC) است.
- MSRL-Net [53]: نویسندگان مدلی بر پایه برت به نام شبکه یادگیری تقویت شده معنایی چندسطحی (MSRL-Net) برای تحلیل احساسات جنبه‌محور پیشنهاد داده‌اند. در MSRL-Net، برای انجام وظیفه تحلیل احساسات جنبه‌محور از روابط وابستگی کلمه و روابط کلمه-جمله برای تقویت نمایش معنایی سطح کلمه در تطبیق معنایی جمله استفاده شده است.

نتایج این مقایسه در جدول (11) ذکر شده است. در این جدول نشان داده شده که مدل پیشنهادی در هر دو مجموعه داده Res2015 و Res2016 عملکرد پایداری دارد و F1 و دقت بسیار نزدیک است. در حالی که در مدل‌های دیگر، این دو معیار فاصله به نسبت زیادی با یکدیگر دارند که نشان می‌دهد این مدل‌ها دارای ابی در معیارهای دقت یا یادآوری هستند. علاوه بر این، عملکرد مدل در هر دو مجموعه داده نزدیک به یکدیگر و در سطح بهینه است. در مقابل، عملکرد مدل‌های دیگر مانند

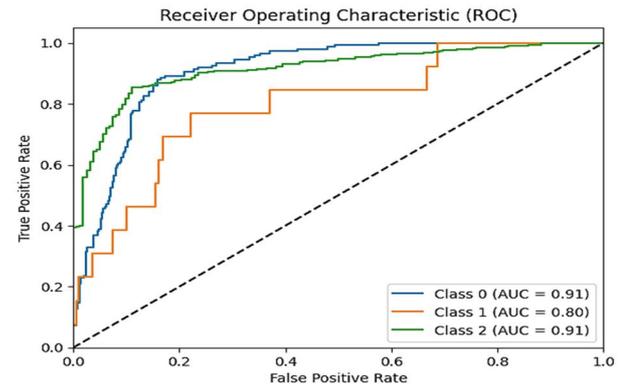
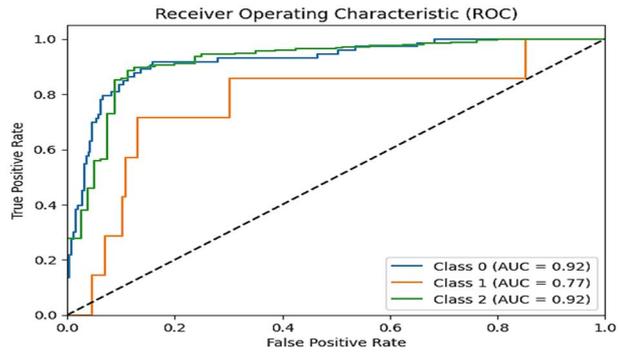
کلمات «portions» و «tiny» را نشان می‌دهد که نشان‌دهنده توانایی مدل برای گرفتن احساسات مربوط به جنبه‌های مختلف در یک جمله است. در لایه 2، که در ردیف پایین نشان داده شده است، افزایش قابل توجهی در وزن توجه برای این کلمات وجود دارد که نشان‌دهنده تجزیه و تحلیل عمیق و تایید ارتباط آنها است. این افزایش تمرکز از لایه 1 به لایه 2 نشان می‌دهد که این مدل نشانه‌های لازم را شناسایی کرده و تفسیر خود را با در دسترس قرار گرفتن اطلاعات زمینه‌ای بیشتر از طریق شبکه، اصلاح می‌کند. بنابراین می‌توان به سوالات پژوهشی ذکر شده در مقدمه اینگونه پاسخ داد:

1- چگونه می‌توان دادگان میان دامنه‌ای مناسبی در دامنه مبدا انتخاب کرد که انتقال ویژگی‌های آنها به دامنه هدف موجب بروز پدیده انتقالی منفی نشود؟

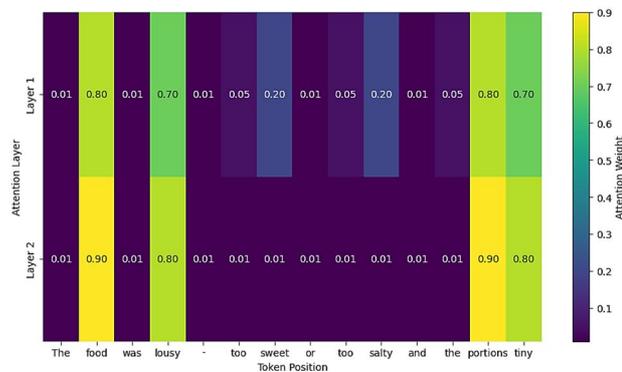
با توجه به افزایش ابزارهای مانند API ها و رصد گسترده داده‌های موجود در وب، اگر در دامنه‌ای خاص داده کمیاب باشد می‌توان از دانش موجود در دامنه‌های مشابه استفاده کرد. با این حال، اگر دامنه‌ها مشابه نبوده یا شباهت پائینی داشته باشند، استفاده از دادگان میان‌دامنه‌ای ممکن است موجب انتقال منفی گردد. یکی از روش‌های موثری که در این پژوهش پیشنهاد شده است استفاده از ضریب شباهت پیروسون است. نتایج جدول (6) و نمودار شکل (3) نشان می‌دهد اگر امتیاز مرتبط با شباهت کسینوسی داده مبدا و هدف کم باشد، انتقال دانش موجب بروز پدیده انتقال منفی می‌گردد.

2- چگونه می‌توان با ترکیب شبکه‌های یادگیری عمیق، لایه توجه و آموزش آنها با دادگان میان‌دامنه‌ای مدلی مطلوب با کمترین افت کارایی برای وظیفه تحلیل احساسات جنبه‌محور ارائه داد؟

شبکه‌های یادگیری عمیق هر یک برای وظیفه‌ای خاص طراحی شده است. برای مثال CNN برای پردازش تصویر به وجود آمد ولی به دلیل ویژگی قدرتمند آن در استخراج ویژگی در سایر وظایف مانند پردازش زبان طبیعی نیز بکار گرفته شد [51]. از همین رو، می‌توان بر ترکیب قابلیت‌های سودمند شبکه‌های



شکل (10): معیار ROC-AUC شبکه عصبی عمیق ترکیبی BiLSTM+ دستیل برت (شکل اول RES2015 و دوم RES2016)



شکل (11): نقشه حرارتی وزن‌های لایه توجه

همان‌طور که در شکل (11) مشاهده می‌شود، نقشه حرارتی به طور موثر نشان می‌دهد که چگونه مکانیسم لایه توجه در مدل جمله «غذا کثیف بود-خیلی شیرین یا خیلی شور و قسمت‌های کوچک» را پردازش می‌کند. ردیف بالا که نمایانگر لایه 1 است، تمرکز اولیه مدل را بر روی کلمات کلیدی مانند «food» و «lousy» نشان می‌دهد و شناسایی جنبه‌های حیاتی و احساسات مرتبط با آنها را برجسته می‌کند. این لایه همچنین توجه به

4- دادگان نامتوازن سبب جهت‌گیری مدل به سمت قطبیتی خاص می‌شود، چگونه می‌توان با این مشکل برخورد کرد؟

برای رفع مشکل دادگان نامتوازن در این پژوهش از تکنیک SMOTE استفاده شده است. SMOTE تکنیکی آماری است که به طور مصنوعی نمونه‌های جدیدی از کلاس اقلیت را در فضای ویژگی ایجاد می‌کند تا مجموعه داده را متعادل کند. این تکنیک سبب می‌شود توازن قطبیت‌ها در کل دادگان حفظ شده و مدل در فرآیند یادگیری به سمت قطبیتی خاص جهت‌گیری نداشته باشد.

در بخش نتایج مقاله، مدل پیشنهادی با چندین روش مقایسه شده و بهبودهای چشمگیری در دقت و عملکرد کلی مشاهده شده است. نتایج نشان می‌دهد مدل پیشنهادی توانسته به صورت میانگین دقت و F1 را بر روی مجموعه داده پنج‌مارک SemEval در حدود 2 درصد افزایش داده و میانگین نرخ ROC-AUC به 86,35 درصد رسیده است. این نتایج به دلیل استفاده از یادگیری انتقالی شات صفر، بهینه‌سازی داده‌های ورودی و ترکیب مدل‌های پیش‌آموزش دیده با شبکه‌های عصبی عمیق حاصل شده است که برخی از بهبودها در زیر ذکر شده است.

- **عملکرد بهبود یافته در مقایسه با مدل‌های سنتی:** نتایج نشان می‌دهند که مدل پیشنهادی به طور قابل توجهی عملکرد بهتری نسبت به مدل‌های سنتی مانند LSTM و GRU دارد. این بهبود به دلیل استفاده از داده‌های متنوع و مرتبط با همبستگی بالا با موضوع مورد بحث است که به مدل کمک می‌کند تا درک عمیق‌تری از داده‌های ورودی داشته باشد. تفاوت اصلی این است که مدل‌های سنتی به داده‌های برچسب‌گذاری شده بیشتر و تنظیمات پیچیده‌تری نیاز دارند، در حالی که مدل پیشنهادی با استفاده از یادگیری انتقالی و داده‌های کم‌برچسب‌گذاری شده یا بدون برچسب‌گذاری به نتایج مطلوب‌تری دست یافته است.
- **بهبود دقت در تحلیل احساسات جنبه‌محور:** یکی از مهم‌ترین نتایج به دست آمده، بهبود دقت در تحلیل

عصبی عمیق سبب پایداری بیشتر مدل یادگیری گردید. برای نمونه شبکه‌های عصبی عمیق بازگشتی مانند BiLSTM برای پردازش توالی‌ها بسیار مناسب هستند. با این حال، این شبکه‌ها نیازمند دادگان بسیار غنی جهت آموزش می‌باشند. از طرف دیگر، در دامنه‌های خاص مانند رمازرها دادگان غنی بسیار نادر می‌باشد. همچنین، این شبکه‌ها (مانند BiLSTM) قدرت بالایی در استخراج ویژگی‌ها ندارند. لذا، همان‌گونه که در معماری مدل پیشنهادی شکل (5) نشان داده شده می‌توان از شبکه‌های پیش‌آموزش دیده مانند دستیل‌برت به عنوان ستون فقرات شبکه، جهت استخراج ویژگی و انتقال دانش استفاده کرد. با دانش استخراج شده و تنظیم شبکه دستیل‌برت می‌توان این دانش را به BiLSTM انتقال داد و آن را جهت انجام وظیفه در دامنه هدف آموزش داد. کارایی این روش در جدول (9)، اثبات شده است.

3- با توجه به اینکه استفاده از مدل‌های یادگیری عمیق پایه و پیش‌آموزش دیده برای استخراج ویژگی و انتقال دانش از دامنه مبدا به هدف موثر می‌باشد. کدامیک از این مدل‌ها برای تحلیل احساسات رمازرها اثربخشی بیشتری دارد؟

همان‌گونه که ذکر شد شبکه‌های عصبی عمیق، مدل‌های حریص به داده هستند. این شبکه‌ها برای کارایی مطلوب نیازمند دادگان غنی برچسب‌دار می‌باشند. در تحلیل احساسات رمازرها دادگان غنی برچسب‌دار مخصوصاً همراه با جنبه‌های متن که مرتبط با انواع رمازرها باشد، کمیاب است. از این رو، آموزش شبکه‌های عصبی پایه در دامنه رمازرها امری چالش برانگیز می‌باشد. لذا، می‌توان از شبکه‌های عصبی پیش‌آموزش دیده مانند انواع برت که توسط مجموعه بزرگی از دادگان جهت وظایف مرتبط با پردازش زبان طبیعی تنظیم شده‌اند، استفاده کرد. بنابراین، همان‌گونه که در جدول (9)، نشان داده شده ترکیب شبکه‌های عصبی عمیق پیش‌آموزش دیده ساده و ترکیبی منجر به دست‌یابی به نتایج مطلوب‌تری نسبت به شبکه‌های عصبی عمیق پایه شده است.

و همچنین ورودی‌های بعدی را در نظر می‌گیرد. همچنین، نتایج نشان می‌دهد، استفاده از روش ضریب شباهت پیرسون جهت انتخاب مجموعه دادگان دامنه مبدا انتقال منفی دانش به دامنه هدف را به صورت چشم‌گیری کاهش می‌دهد. علاوه بر این، نرخ یادگیری و فوق‌پذیر مورد استفاده در این پژوهش و انتخاب ابرپارامترها به روش گریدی سبب گردیده شبکه‌های عصبی عمیق در تعداد تکرار پائین‌تری به همگرایی برسند که بسیار در کاربردهای تجاری، حیاتی می‌باشد. همچنین در این پژوهش نشان داده شد، استفاده از مدل‌های تنظیم‌شده مبنی بر برت با ترکیب شبکه‌های عصبی عمیق BiLSTM می‌تواند کارایی مطلوبی در تحلیل احساسات دامنه‌های نوظهور مانند رمزارزها داشته باشد. از سوی دیگر، این پژوهش با استفاده از یادگیری انتقالی با شات صفر و صرفاً استفاده از دادگان میان‌دامنه‌ای مشابه نتایج امیدوارکننده‌ای را برای تحلیل احساسات جنبه‌محور در دامنه‌های خاص به دست آورده است. در تحقیقات آینده، می‌توان به دنبال پاسخ به این سوال بود که چه میزان از انتقال دانش سبب بهبود مدل در دامنه هدف می‌شود. به عبارت دیگر، تا چه میزان افزایش مجموعه دادگان در دامنه مبدا کارایی مدل در دامنه هدف را بهبود می‌بخشد. همچنین می‌توان بررسی کرد آیا استفاده از لایه‌های توجه در شبکه‌های عصبی عمیق ترکیبی پایه مانند CNN+BiLSTM یا اشتراک‌گذاری وزن‌ها در شبکه‌های کانولوشنی می‌تواند سبب شود این مدل‌ها به دقت‌های مشابهی با شبکه‌های عصبی عمیق آموزش دیده دست یابند. علاوه بر این، می‌توان لایه‌های منجمد شده شبکه‌های پیش‌آموزش دیده را به عنوان یک ابرپارامتر برای بهینه‌سازی مورد بررسی قرارداد. همچنین می‌توان آزمایش کرد ترکیب ویژگی‌های استخراج شده از مدل‌های تعبیه کلمات مانند Word2Vec با وزن‌های استخراج شده از شبکه‌های پیش‌آموزش دیده چه تاثیری بر کارایی مدل‌های عمیق دامنه هدف دارد.

تعارض منافع: نویسندگان اعلام می‌کنند که هیچ تعارض منافعی ندارند.

احساسات جنبه‌محور است. مدل پیشنهادی با استفاده از مکانیسم‌های توجه و ترکیب داده‌های متنوع، قادر به شناسایی بهتر روابط بین جنبه‌ها و احساسات در متن‌ها شده است. این بهبود به دلیل استفاده از داده‌های همبسته و مرتبط با موضوع و همچنین استفاده از شبکه‌های عمیق پیش‌آموزش دیده بوده که توانایی مدل را در تحلیل دقیق‌تر افزایش داده است.

- **کاهش پدیده انتقال منفی:** یکی از مشکلات رایج در یادگیری انتقالی، انتقال منفی است که می‌تواند دقت مدل را کاهش دهد. در این تحقیق، با استفاده از داده‌هایی با همبستگی بالا و تنظیم ابرپارامترها، این پدیده به طور قابل توجهی کاهش یافته است. این انتخاب دقیق داده‌ها باعث شده است که مدل نه تنها دقت بالایی در دامنه‌های مختلف داشته باشد، بلکه توانایی تعمیم‌پذیری بیشتری را نیز نشان دهد.
- **مقایسه با روش‌های پیشرفته‌تر:** در مقایسه با روش‌های پیشرفته مانند GCN و مدل‌های ترکیبی، مدل پیشنهادی توانسته است با ساده‌سازی فرآیند آموزش، استفاده از داده‌های مرتبط، تنظیمی ابرپارامترها و نرخ یادگیری وفق‌پذیر، عملکردی مطلوب ارائه دهد. این تفاوت به دلیل استفاده از یادگیری انتقالی و بهینه‌سازی‌های انجام شده در مدل است که باعث شده است تا مدل نه تنها دقت بیشتری داشته باشد، بلکه کارایی بالاتری از نظر معیار FI را نیز نشان دهد.

5. نتیجه‌گیری و کارهای آینده

رویکردهای اولیه پردازش زبان طبیعی با استفاده از CNN، شبکه عصبی بازگشتی و BiLSTM دارای محدودیت‌هایی برای درک بافت عمیق‌تر جملات هستند. از این رو، شبکه پیش‌آموزش دیده برت درک بهتری از محتوای جملات نسبت به رویکردهای پایه فراهم می‌آورد. زیرا، در لایه‌های رمزگذار شبکه برت همه ورودی‌ها (کل جمله) به طور همزمان پردازش می‌شوند. بنابراین، هنگام ساختن زمینه برای یک کلمه، برت ورودی‌های قبل از آن

- [1] J. Chai and A. Li, "Deep learning in natural language processing: A state-of-the-art survey," in Proc. 2019 Int. Conf. Mach. Learn. Cybern. (ICMLC), 2019, doi: 10.1109/ICMLC48188.2019.8949185.
- [2] F. Zare Mehrjardi, M. Yazdian-Dehkordi, and A. Latif, "Evaluating classical machine learning and deep-learning methods in sentiment analysis of Persian telegram message," *Soft Comput. J.*, vol. 11, no. 1, pp. 88-105, 2022, doi: 10.22052/scj.2023.246553.1077 [In Persian].
- [3] O. Khalaf Beigi, S.A. Bashiri Mosavi, and S. Gharloghi, "Applying Character-Level Neural Network-Based Sentiment Analysis Model on Persian Comments of the Social Media-Online Store Platforms," *Soft Comput. J.*, vol. 11, no. 2, pp. 118-133, 2023, doi: 10.22052/scj.2023.248311.1094 [In Persian].
- [4] F. Pourgholamali, M. Kahani, and E. Asgarian, "Exploiting Big Data Technology for Opinion Mining," *Soft Comput. J.*, vol. 9, no. 1, pp. 26-39, 2020, doi: 10.22052/scj.2021.111450 [In Persian].
- [5] K.W. Church, Z. Chen, and Y. Ma, "Emerging trends: A gentle introduction to fine-tuning," *Nat. Lang. Eng.*, vol. 27, no. 6, pp. 763-778, 2021, doi: 10.1017/S1351324921000322.
- [6] K. Jahanbin and M.A.Z. Chahooki, "Database of Twitter influencers in cryptocurrency (2021-2023) with sentiments," *Research Square*, 2023, doi: 10.21203/rs.3.rs-3192598/v1.
- [7] D. Elreedy, A.F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4903-4923, 2024, doi: 10.1007/s10994-022-06296-4.
- [8] G. Lu, Y. Liu, J. Wang, and H. Wu, "CNN-BiLSTM-Attention: A multi-label neural classifier for short texts with a small set of labels," *Inf. Process. Manag.*, vol. 60, no. 3, p. 103320, 2023, doi: 10.1016/j.ipm.2023.103320.
- [9] H.N. Serere, B. Resch, and C.R. Havas, "Enhanced geocoding precision for location inference of tweet text using spaCy, Nominatim and Google Maps. A comparative analysis of the influence of data selection," *PLoS One*, vol. 18, no. 3, p. e0282942, 2023, doi: 10.1371/journal.pone.0282942.
- [10] M. Isnan, G.N. Elwirehardja, and B. Pardamean, "Sentiment analysis for TikTok review using VADER sentiment and SVM model," *Procedia Comput. Sci.*, vol. 227, pp. 168-175, 2023, doi: 10.1016/j.procs.2023.10.514.
- [11] M. Nair, L.A. Abd-Elmegid, and M.I. Marie, "Sentiment analysis model for cryptocurrency tweets using different deep learning techniques," *J. Intell. Syst.*, vol. 33, no. 1, 2024, doi: 10.1515/jisys-2023-0085.
- [12] T. Pano and R. Kashef, "A complete VADER-based sentiment analysis of bitcoin (BTC) tweets during the era of COVID-19," *Big Data Cogn. Comput.*, vol. 4, no. 4, p. 33, 2020, doi: 10.3390/bdcc4040033.
- [13] A.K. Jha and N.K. Verma, "Social media sustainability communication: An analysis of firm behaviour and stakeholder responses," *Inf. Syst. Front.*, vol. 25, no. 2, pp. 723-742, 2023, doi: 10.1007/s10796-022-10257-6.
- [14] N.K. Singh, D.S. Tomar, and A.K. Sangaiah, "Sentiment analysis: A review and comparative analysis over social media," *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 1, pp. 97-117, 2020, doi: 10.1007/s12652-018-0862-8.
- [15] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowl. Inf. Syst.*, vol. 60, no. 2, pp. 617-663, 2019, doi: 10.1007/s10115-018-1236-4.
- [16] Z. Nanli, Z. Ping, L. Weiguo, and C. Meng, "Sentiment analysis: A literature review," in Proc. 2012 Int. Symp. Manag. Technol. (ISMOT), 2012,

- pp. 572-576, doi: 10.1109/ISMOT.2012.6679538.
- [17] H. Liu, I. Chatterjee, M. Zhou, X.S. Lu, and A. Abusorrah, "Aspect-based sentiment analysis: A survey of deep learning methods," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 6, pp. 1358-1375, 2020, doi: 10.1109/TCSS.2020.3033302.
- [18] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 1, pp. 249-270, 2022, doi: 10.1109/TKDE.2020.2981333.
- [19] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040-53065, 2019, doi: 10.1109/ACCESS.2019.2912200.
- [20] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235-1270, 2019, doi: 10.1162/neco_a_01199.
- [21] N. Agarwal, A. Sondhi, K. Chopra, and G. Singh, "Transfer learning: Survey and classification," in *Smart Innovations in Communication and Computational Sciences*, Singapore: Springer, 2021, pp. 145-155, doi: 10.1007/978-981-15-5345-5_13.
- [22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional Transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018, doi: 10.48550/arXiv.1810.04805.
- [23] I. Staliunaite and I. Iacobacci, "Compositional and lexical semantics in RoBERTa, BERT and DistilBERT: A case study on CoQA," in *Proc. 2020 Conf. Empirical Methods Nat. Lang. Process. (EMNLP)*, 2020, doi: 10.48550/arXiv.2009.08257.
- [24] A.F. Adoma, N.-M. Henry, and W. Chen, "Comparative analyses of Bert, Roberta, distilbert, and xlnet for text-based emotion recognition," in *Proc. 2020 17th Int. Comput. Conf. Wavelet Active Media Technol. Inf. Process. (ICCWAMTIP)*, 2020, doi: 10.1109/ICCWAMTIP51612.2020.9317379.
- [25] M. Suhasini and B. Srinivasu, "Emotion detection framework for twitter data using supervised classifiers," in *Advances in Intelligent Systems and Computing*, Singapore: Springer, 2020, pp. 565-576, doi: 10.1007/978-981-15-1097-7_47.
- [26] J. Abraham, D.W. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis," *SMU Data Sci. Rev.*, vol. 1, no. 3, p. 1, 2018.
- [27] S. Biswas, M. Pawar, S. Badole, N. Galande, and S. Rathod, "Cryptocurrency price prediction using neural networks and deep learning," in *Proc. 2021 7th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, 2021, vol. 1, pp. 408-413, doi: 10.1109/ICACCS51430.2021.9441872.
- [28] X. Huang et al., "LSTM based sentiment analysis for cryptocurrency prediction," in *Proc. Database Syst. Adv. Appl. (DASFAA)*, 2021, pp. 617-621, doi: 10.1007/978-3-030-73200-4_47.
- [29] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Web Social Media*, 2014, vol. 8, no. 1, pp. 216-225, doi: 10.1609/icwsm.v8i1.14550.
- [30] S. Oikonomopoulos, K. Tzafilkou, D. Karapiperis, and V. Verykios, "Cryptocurrency price prediction using social media sentiment analysis," in *Proc. 2022 13th Int. Conf. Inf. Intell. Syst. Appl. (IISA)*, 2022, pp. 1-8, doi: 10.1109/IISA56318.2022.9904351.
- [31] G.C. Habek, M.A. Tocoglu, and A. Onan, "Bi-directional CNN-RNN architecture with group-wise enhancement and attention mechanisms for cryptocurrency sentiment analysis," *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2022.2145641.
- [32] I. Moudhich and A. Fennan, "Graph embedding approach to analyze sentiments on cryptocurrency," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 14, no. 1, p. 690, 2024, doi: 10.11591/ijece.v14i1.pp690-697.

- [33] N. Aslam, F. Rustam, E. Lee, P.B. Washington, and I. Ashraf, "Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble LSTM-GRU model," *IEEE Access*, vol. 10, pp. 39313-39324, 2022, doi: 10.1109/ACCESS.2022.3165621.
- [34] J. Davchev, K. Mishev, I. Vodenska, L. Chitkushev, and D. Trajanov, "Bitcoin price prediction using transfer learning on financial micro-blogs," in *Proc. 16th Annu. Int. Conf. Comput. Sci. Educ. Comput. Sci.*, 2020.
- [35] P. Zhao, L. Hou, and O. Wu, "Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification," *Knowl. Based Syst.*, vol. 193, p. 105443, 2020, doi: 10.1016/j.knosys.2019.105443.
- [36] A. Zunic, P. Corcoran, and I. Spasic, "Aspect-based sentiment analysis with graph convolution over syntactic dependencies," *Artif. Intell. Med.*, vol. 119, p. 102138, 2021, doi: 10.1016/j.artmed.2021.102138.
- [37] Y. Lv et al., "Aspect-level sentiment analysis using context and aspect memory network," *Neurocomputing*, vol. 428, pp. 195-205, 2021, doi: 10.1016/j.neucom.2020.11.049.
- [38] N. Habbat, H. Anoun, and L. Hassouni, "Combination of GRU and CNN deep learning models for sentiment analysis on French customer reviews using XLNet model," *IEEE Eng. Manag. Rev.*, vol. 51, no. 1, pp. 41-51, 2022, doi: 10.1109/EMR.2022.3208818.
- [39] M.M. Kabir, Z.A. Othman, M.R. Yaakub, and S. Tiun, "Hybrid syntax dependency with lexicon and logistic regression for aspect-based sentiment analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 10, 2023, doi: 10.14569/IJACSA.2023.0141085.
- [40] B. Huang et al., "Aspect-level sentiment analysis with aspect-specific context position information," *Knowl. Based Syst.*, vol. 243, p. 108473, 2022, doi: 10.1016/j.knosys.2022.108473.
- [41] X. Zhu, L. Zhu, J. Guo, S. Liang, and S. Dietze, "GL-GCN: Global and Local Dependency Guided Graph Convolutional Networks for aspect-based sentiment classification," *Expert Syst. Appl.*, vol. 186, p. 115712, 2021, doi: 10.1016/j.eswa.2021.115712.
- [42] S. Chumakov, A. Kovantsev, and A. Surikov, "Generative approach to aspect based sentiment analysis with GPT language models," *Procedia Comput. Sci.*, vol. 229, pp. 284-293, 2023, doi: 10.1016/j.procs.2023.12.030.
- [43] P. Liashchynskyi, "Grid search, random search, genetic algorithm: A big comparison for NAS," *arXiv preprint arXiv:1912.06059*, 2019, doi: 10.48550/arXiv.1912.06059.
- [44] K. Nakamura, B. Derbel, K.-J. Won, and B.-W. Hong, "Learning-rate annealing methods for deep neural networks," *Electronics*, vol. 10, no. 16, p. 2029, 2021, doi: 10.3390/electronics10162029.
- [45] G. D'Aniello, M. Gaeta, and I. La Rocca, "KnowMIS-ABSA: An overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5543-5574, 2022, doi: 10.1007/s10462-021-10134-9.
- [46] A.J. Amadeo, J.G. Siento, T.A. Eikwine, Diana, and I.H. Parmonangan, "Temporal Fusion Transformer for multi horizon bitcoin price forecasting," in *Proc. 2023 IEEE 9th Inf. Technol. Int. Seminar (ITIS)*, 2023, doi: 10.1109/ITIS59651.2023.10420330.
- [47] A.Y. Noura, M. Bouchakwa, and M. Amara, "Role of social networks and machine learning techniques in cryptocurrency price prediction: A survey," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, 2024, doi: 10.1007/s13278-024-01316-8.
- [48] G. Serafini et al., "Sentiment-driven price prediction of the bitcoin based on statistical and deep learning approaches," in *Proc. 2020 Int. Joint Conf. Neural Networks (IJCNN)*, 2020, doi: 10.1109/IJCNN48605.2020.9206704.
- [49] N. Elhadad et al., "SemEval-2015 Task 14: Analysis of clinical text," in *Proc. 9th Int. Worksh.*

- Semantic Eval. (SemEval), 2015, doi: 10.18653/v1/S15-2051.
- [50] M. Pontiki et al., "SemEval-2016 Task 5: Aspect based sentiment analysis," in Proc. 10th Int. Worksh. Semantic Eval. (SemEval), 2016, doi: 10.18653/v1/S16-1002.
- [51] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, "A CNN-BiLSTM model for document-level sentiment analysis," Mach. Learn. Knowl. Extr., vol. 1, no. 3, pp. 832-847, 2019, doi: 10.3390/make1030048.
- [52] H. Wan et al., "Target-aspect-sentiment joint detection for aspect-based sentiment analysis," in Proc. AAAI Conf. Artif. Intell., 2020, vol. 34, no. 5, pp. 9122-9129, doi: 10.1609/aaai.v34i05.6447.
- [53] Z. Hu, Z. Wang, Y. Wang, and A.-H. Tan, "MSRL-Net: A multi-level semantic relation-enhanced learning network for aspect-based sentiment analysis," Expert Syst. Appl., vol. 217, p. 119492, 2023, doi: 10.1016/j.eswa.2022.119492.