روشی کارا بر پایه ترکیب مدل‌های یادگیری ژرف برای تجزیه و تحلیل احساسات در متون

حسین صدر[1]، میرمحسن پدرام[2]* و محمد تشنه‌لب[3]

گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، مؤسسه آموزش عالی راهبرد شمال، رشت، ایران

گروه مهندسی برق و کامپیوتر، دانشکده فنی و مهندسی، دانشگاه خوارزمی، تهران، ایران

دانشکده مهندسی برق، گروه سیستم‌ها و کنترل، دانشگاه صنعتی خواجه نصیر طوسی، تهران، ایران

چکیده

یکی از مهم‌ترین داده‌های متنی موجود در سطح وب احساسات و دیدگاه‌های افراد نسبت به یک موضوع یا مفهوم مشخص است. با این حال، یافتن و نظارت بر وبگاه‌های حاوی این احساسات و استخراج اطلاعات موردنیاز از آن‌ها به‌علت گسترش وبگاه‌های گوناگون کاری دشوار محسوب می‌شود. در این راستا، توسعه سامانه‌های تجزیه و تحلیل خودکار احساسات که بتواند نظرات را استخراج کرده و روند فکری مرتبط با آن‌ها را بیان کند، در سال‌های اخیر توجه زیادی را به خود جلب کرده است و روش‌های بر پایه یادگیری ژرف، یکی از راه‌کارهایی هستند که توانسته‌اند به نتایج چشم‌گیری در کاربردهای مختلف پردازش زبان‌های طبیعی به‌خصوص تجزیه و تحلیل احساسات دست یابند؛ اما این روش‌ها برخلاف عملکرد قابل‌توجه هنوز با چالش‌هایی مواجه هستند و نیاز به پیشرفت در این حوزه همچنان وجود دارد؛ ازاین‌رو، هدف این مقاله ترکیب مدل‌های یادگیری ژرف به‌منظور ارائه یک روش جدید برای تجزیه و تحلیل احساسات متنی است که بتواند ضمن استفاده هم‌زمان از مزایای شبکه‌های عصبی ژرف بر مشکلات آن‌ها چیره شود. در این راستا، در این مقاله روشی بر پایه ترکیب شبکه عصبی پیچشی و شبکه عصبی هم‌گشتی معرفی شده است که در آن به‌منظور حفظ وابستگی‌های بلندمدت در جملات و کاهش ازدست‌رفتن داده‌های محلی که به‌عنوان چالش‌های شبکه عصبی پیچشی به شمار می‌آیند، از لایه هم‌گشتی تعمیم‌یافته که در آن از یک ویژگی میانی حاصل از ترکیب گره‌های فرزندان استفاده می‌شود، به‌عنوان جایگزین لایه ادغام در شبکه عصبی پیچشی بر پایه سازوکار توجه استفاده شده است. بر اساس نتایج آزمایش‌ها، روش پیشنهادی به‌ترتیب با دقت ۵۳/۹۲ و ۹۲/۸۹ درصد روی مجموعه‌داده‌های SST1 و SST2 و دارای دقت بالاتری نسبت به سایر روش‌های موجود است.

واژگان کلیدی: تجزیه و تحلیل احساسات، یادگیری ژرف، شبکه عصبی پیچشی، شبکه عصبی هم‌گشتی، سازوکار توجه

Efficient Method Based on Combination of Deep Learning Models for Sentiment Analysis of Text

Hossein Sadr[1], Mir Mohsen Pedram[2]* & Mohamad Teshnehlab[3]

Department of Computer Engineering, Rahbord Shomal Institute of Higher Education, Rasht, Iran

Department of Electrical and Computer Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran Industrial Control Center of Excellence, Faculty of Electrical and Computer Engineering, K. N. Toosi University, Tehran, Iran

Abstract

People's opinions about a specific concept are considered as one of the most important textual data that are available on the web. However, finding and monitoring web pages containing these comments and extracting valuable information from them is very difficult. In this regard, developing automatic sentiment analysis systems that can extract opinions and express their intellectual process has attracted considerable attention in recent years. Sentiment analysis is considered as one of the most active research areas in the field of natural language processing which tries to classify a piece of text containing

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۳۹۸/۵/۲۴ • تاریخ پذیرش: ۱۳۹۹/۱۰/۲۱ • تاریخ انتشار: ۱۴۰۱/۰۳/۳۱ • نوع مطالعه: پژوهشی

سال ۱۴۰۱ شمارهٔ ۱ پیاپی ۵۱

[DOI: 10.52547/jsdp.19.1.19]

opinions based on its polarity and determine whether an expressed opinion about a specific topic, event or product is positive or negative.
Since about a decade ago, many studies have been carried out to investigate the effects of traditional classification models, such as Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, etc. in the task of sentiment analysis. Although machine learning models have achieved great success in this filed, they are still confronted with some limitations, notably manual feature engineering requirements. In other words, the classification performance of machine learning models is highly dependent on the extracted features and they play an important role in obtaining higher classification accuracy. To deal with these problems, deep learning models have been extensively employed as an alternative to traditional machine learning models and have achieved impressive results. It is worth mentioning that despite the remarkable performance of these methods, they are still confronted with some limitations and they are on their first steps of progress.
Therefore, the goal of this paper is to propose a combinational deep learning model that can overcome their problems as well as utilizing their benefits. In this regard, an efficient method based on combination of convolutional and recursive neural networks is proposed in this paper that employs a generalized recursive neural network, where an intermediate feature is obtained by combining children's nodes, as an alternative of pooling layer in attention-based convolutional neural network with the aim of capturing long term dependencies and decreasing the loss of local information. Based on empirical results, the proposed method with the accuracy of 53.92% and 92.89% respectively on SST1 and SST2 datasets not only outperforms other existing models but also can be trained much faster.

Keywords: Sentiment analysis, Deep Leaning, Convolutional neural network, Recursive neural network, Attention mechanism

۱- مقدمه

نظرات[1] و احساسات، مرکز بیشتر فعالیت‌های بشر بوده و نقش کلیدی را در تصمیمات انسان‌ها ایفا می‌کنند. در واقع، احساسات و برداشت‌های هر شخص و هم‌چنین تصمیماتی که وی در زندگی می‌گیرد، به دیدگاه دیگران و ارزیابی پیشین از دنیای واقعی وابسته است [1,2]. پیش‌تر، افراد به‌منظور دانستن نظر دیگران تنها می‌توانستند با خانواده و دوستان خود مشورت کنند و سازمان‌ها و تولیدکنندگان هم برای این منظور از برگه‌های نظرسنجی استفاده می‌کردند. امروزه، توسعه شبکه‌های اجتماعی، بلاگ‌ها و انجمن‌های بحث و گفتگو باعث شده که حجم گسترده‌ای از احساسات و نظرات نسبت به یک موضوع خاص به‌صورت برخط[2] در دسترس باشد و در نتیجه افراد و سازمان‌ها برای تصمیم‌گیری‌های خود می‌توانند مبادرت به استفاده از این محتوا کند [3,4].

بااین‌حال، یافتن و نظارت بر وبگاه‌های حاوی این نظرات در وب و هم‌چنین استخراج اطلاعات مورد نیاز از آن‌ها به‌علت گسترش تارنماهای گوناگون، کاری دشوار محسوب می‌شود و برای یک فرد عادی شناسایی وبگاه‌های مرتبط و سپس استخراج و خلاصه‌سازی نظرات کاری دشوار خواهد بود. در همین راستا، توسعه سامانه‌های خودکار تجزیه و تحلیل نظرات و احساسات که بتوانند حجم انبوهی از آن‌ها را استخراج کرده و مفهوم مرتبط با آن‌ها را بیان سازد؛ در سال‌های اخیر توجه زیادی را به خود جلب کرده است [5,6].

درکل، روش‌های تجزیه و تحلیل احساسات به دو دسته روش‌های بر پایه منبع واژگان[3] و روش‌های بر پایه یادگیری ماشین[4] تقسیم می‌شوند [7,8]. روش‌های بر پایهِ منبع واژگان، از یک منبع واژه پیش‌ساخته همانند یک وا ژه‌نامه استفاده می‌کنند که با استفاده از این منبع و مقایسه واژگان موجود در احساسات، می‌تواند احساسات کاربران را تحلیل کنند. این روش‌ها به‌طور عمومی یک روش دستی هستند و فردی خبره باید تمامی واژگان موجود در منبع واژگان را برچسب‌گذاری کند. در مقابل، روش‌های یادگیری ماشین ترکیبی از روش‌های خودکار برای تشخیص و شناسایی الگو در مجموعه‌ای از داده‌ها هستند روش‌های یادگیری ماشین از مهم‌ترین روش‌های با ناظر برای برای پیش‌بینی داده‌های نامعلوم و تصمیم‌گیری درباره آن‌ها به حساب می‌آیند و شامل الگوریتم‌هایی از قبیل بیزی[5]، ماشین بردار پشتیبان[6] و آنتروپی بیشینه[7] هستند [9,10].

در جامعه پژوهشی کنونی، رویکردهای برجسته تجزیه و تحلیل احساسات به‌طور فزاینده‌ای بر پایه

[1] Opinion
[2] Online
[3] Lexicon based
[4] Machine learning based
[5] Naïve Bayes
[6] Support vector machine
[7] Maximum entropy

[DOI: 10.52547/jsdp.19.1.19]

روش‌های یادگیری ماشین هستند. مزایای این روش‌ها در مقابل رویکردهای رایج مهندسی دانش (شامل تعریف دستی دسته‌بند با استفاده از متخصصین)، اثربخشی خوب و صرفه‌جویی قابل‌توجه از لحاظ نیروی متخصص و قابلیت حمل[1] برای حوزه‌های مختلف است [9,10]؛ اما این رویکردها با وجود عملکرد به‌نسبه مناسب همچنان فاقد ریزبینی‌های ساختاری و معنایی در محتویات متنی هستند و نادیده‌گرفتن یک عبارت به‌طورکامل مثبت می‌تواند احساسات موجود در متن را به‌طورکامل متحول کند؛ درواقع، روش‌های تجزیه و تحلیل احساسات بر پایه یادگیری ماشین به‌شدت وابسته به مهندسی ویژگی‌ها[2] هستند و فقدان داده‌های برچسب‌گذاری‌شده به‌عنوان مهم‌ترین چالش آن‌ها به‌حساب می‌آید. از طرف دیگر، استخراج دستی ویژگی‌ها کار بسیار دشواری است و با توجه به خاصیت پویای زبان، ویژگی‌های استخراج‌شده ممکن است در یک بازه زمانی بسیار کوتاه منسوخ شوند؛ در نتیجه نیاز به روشی است که بتواند بر این مشکلات غلبه کند و ساختار جمله را به‌صورت مجموعه‌ای از ویژگی‌ها نشان دهد. برای حل این مشکلات، روش‌های تجزیه و تحلیل احساسات و یادگیری ژرف[3] باهم تلفیق شده‌اند؛ زیرا شبکه‌های یادگیری ژرف با توجه به توانایی یادگیری خودکار بسیار مؤثر بوده و نیاز به استخراج دستی وِیژگی‌ها را از بین برده‌اند [3,11].

با این‌که پژوهش‌های متعددی در حوزه تجزیه و تحلیل احساسات با استفاده از یادگیری ژرف صورت گرفته و نتایج قابل‌ملاحظه‌ای در این حوزه به‌دست آمده است، اما این روش‌ها برخلاف مزایای قابل‌توجهشان هنوز با چالش‌هایی مواجه هستند و نیاز به پیشرفت در این حوزه همچنان وجود دارد. با توجه به این‌که هرکدام از مدل‌های یادگیری ژرف دارای مزایا و معایب خود هستند، درنتیجه نیاز به یافتن راه‌کاری است که بتوان به‌کمک آن ضمن استفاده از مزایای شبکه‌های موجود بر مشکلات آن‌ها غلبه کرد.

شبکه‌های عصبی پیچشی[4] و شبکه‌های عصبی هم گشتی[5] از جمله مدل‌های یادگیری ژرف هستند که در سال‌های اخیر بسیار موردتوجه پژوهش‌گران حوزه پردازش زبان طبیعی به‌خصوص تجزیه و تحلیل احساسات و نظرکاوی قرار گرفته‌اند. از مهم‌ترین مزایای شبکه‌های عصبی پیچشی تعداد کم پارامترها و آموزش ساده است. از طرف دیگر، این شبکه‌ها توانایی استخراج ویژگی‌های محلی[6] را دارند که با افزایش تعداد لایه‌های شبکه‌های عصبی پیچشی نیز می‌توان ویژگی‌های با ارزش‌تری را از توالی ورودی استخراج کرد [12].

مهم‌ترین مشکل شبکه‌های پیچشی نیز مربوط به استخراج ویژگی‌های سطح پایین[7] و عدم توانایی در حفظ وابستگی‌های بلندمدت[8] است. از دیگر مشکلات شبکه‌های عصبی پیچشی نیز می‌توان به اختصاص ارزش یکسان به کلیه واژگان موجود در جملات اشاره کرد. در مقابل و برخلاف شبکه‌های عصبی پیچشی، شبکه‌های عصبی هم‌گشتی توانایی استخراج ویژگی‌های سطح بالا و معنادار را دارند. هم‌چنین این شبکه‌ها می‌توانند ساختار جملات را به‌صورت درختی نشان دهند که باعث حفظ ترتیب واژگان در جمله و وابستگی‌های بلندمدت می‌شود.

در این راستا، در این مقاله یک روش جدید بر پایه ترکیب شبکه عصبی پیچشی و همگشتی برای تجزیه و تحلیل احساسات موجود در متن معرفی شده است که بتواند ضمن استفاده بهینه از مزایای این شبکه‌ها بر مشکلاتشان غلبه کند. در روش پیشنهادی از شبکه عصبی هم‌گشتی به‌عنوان جایگزین لایه ادغام[9] در شبکه عصبی پیچشی استفاده می‌شود تا بتوان ضمن کاهش ازدست‌رفتن داده‌های محلی به‌وسیلهٔ لایه ادغام، وابستگی بلندمدت در جملات را حفظ کند. هم‌چنین در روش پیشنهادی به‌منظور تأکید روی واژگانی که دارای اهمیت بیشتری روی معنای جملات هستند، از سازوکار توجه روی شبکه عصبی پیچشی و برای بررسی تأثیر واژگان روی یکدیگر از ویژگی میانی حاصل از ترکیب گره‌های فرزند در شبکه عصبی هم‌گشتی استفاده شده است. نوآوری‌های این مقاله را می‌توان به‌صورت زیر بیان کرد:

۱) استفاده از شبکه عصبی هم‌گشتی به‌عنوان جایگزین لایه ادغام در شبکه عصبی پیچشی به‌منظور حفظ وابستگی‌های بلندمدت و کاهش ازدست‌رفتن ویژگی های محلی

۲) استفاده از سازوکار توجه روی شبکه عصبی پیچشی به منظور تأکید روی کلمات با اهمیت بالاتر

[1] Portable
[2] Feature engineering
[3] Deep learning
[4] Convolutional neural network
[5] Recursive neural network
[6] Local features
[7] Low level features
[8] Long-time dependency
[9] Pooling

[DOI: 10.52547/jsdp.19.1.19]

۳) استفاده از ویژگی میانی حاصل از ترکیب گره‌های فرزندان در شبکه عصبی هم‌گشتی به‌منظور درنظرگرفتن تأثیر کلمات روی یکدیگر

بخش‌های بعدی مقاله نیز به‌صورت زیر سازماندهی شده است: مرور مختصری از مهم‌ترین پژوهش‌های انجام‌شده در زمینه تجزیه و تحلیل احساسات با استفاده از یادگیری ژرف در بخش دوم ارائه شده است. در بخش سوم روش پیشنهادی و دلایل انتخاب آن به تفضیل بیان شده و بخش چهارم شامل جزئیات پیاده‌سازی و نتایج حاصل از آزمایش‌ها است. توضیحاتی در راستای نتایج حاصل از پژوهش و کارهایی که در ادامه می‌توان انجام داد، در بخش پنجم ارائه شده است.

۲- پیشینهٔ پژوهش

تجزیه و تحلیل احساسات یکی از موضوعات داغ پژوهشی در حوزه پردازش زبان طبیعی است که در سال‌های اخیر نظر پژوهش‌گران زیادی را به خود جلب کرده است. در طول زمان راه‌کارهای مختلفی در حوزه تجزیه و تحلیل احساسات با هدف کاهش خطاها و افزایش سطح دقت داده‌ها در شبکه‌های اجتماعی معرفی شده‌اند و به‌طورکلی می‌توان آن‌ها را به دو دسته بر پایه منبع واژگان و روش‌های یادگیری ماشین تقسیم کرد [13].

ازآن‌جایی‌که مهم‌ترین چالش پیش‌روی روش‌های یادگیری ماشین استخراج ویژگی‌های مناسب است و این الگوریتم‌ها برای استخراج ویژگی‌های مناسب وابسته به انسان بوده و دقت ویژگی‌های انتخابی روی دقت این روش‌ها تأثیرگذار است، پژوهش‌گران در چند سال اخیر تلاش زیادی برای ترکیب یادگیری ژرف و مفاهیم مرتبط با تجزیه و تحلیل احساسات کرده‌اند؛ زیرا برخلاف یادگیری ماشین، یادگیری ژرف اغلب نیازمند مهندسی دستی ویژگی‌ها نیست و هدف آن ایجاد شبکه‌های بزرگ عصبی است که علاوه‌بر توانایی یادگیری می‌توانند بدون دخالت انسان در مورد مسائل فکر کنند [14,9]. با توجه به اینکه تأکید این مقاله روی حل مسأله تجزیه و تحلیل احساسات با استفاده از ترکیب شبکه عصبی پیچشی و هم گشتی است، در ادامه برخی از مطالعات انجام‌شده در این حوزه با تأکید بر این نوع شبکه‌های یادگیری ژرف و مدل‌های ترکیبی مورد بررسی قرار می‌گیرند.

شبکه عصبی پیچشی به شبکه‌هایی گفته می‌شود که دارای یک یا چندلایه پیچش هستند. هر لایه پیچش مانند یک فیلتر محلی[1] عمل می‌کند. شبکه پیچشی به‌طورمعمول با یک لایه ادغام همراه است. هدف از این لایه استخراج اطلاعات مهم از بین اطلاعاتی است که لایه پیچش استخراج کرده است. در این راستا، *اسلام و همکارش* [15] یک شبکه عصبی پیچشی جدید برای تجزیه و تحلیل احساسات بصری پیشنهاد کردند که با استفاده از زبان‌های پایتون و کافی[2] در یک ماشین لینوکس اجرا شد. در این مدل، برای انتقال روش یادگیری پارامترها و آموزش وزن‌ها از گوگل‌نت استفاده شد که با افزایش اندازه، عملکرد آن افزایش می‌یابد. آن‌ها یک مدل شبکه‌های عصبی پیچشی با الهام از گوگل‌نت با ۲۲ لایه جهت تجزیه احساس پیشنهاد کردند که با استفاده از الگوریتم گرادیان نزولی تصادفی[3] بهینه‌سازی شده بود.

در ادامه *اویانگ و همکاران* [16] یک چارچوب هفت‌لایه‌ای را برای تجزیه احساسات در سطح عبارات پیشنهاد دادند که این چارچوب از شبکه عصبی پیچش و Word2vec برای پردازش احساس استفاده می‌کرد. در همین راستا، کیم و همکارانش نیز از یک شبکه عصبی پیچشی تک‌لایه روی بردارهای حاصل از مدل بازنمایی ویژگی Word2vec برای دسته‌بندی جملات با تأکید بر تجزیه و تحلیل احساسات استفاده کردند. *یین و همکارانش* [17] به‌منظور تحلیل احساسات در سطح جمله از شبکه‌های عصبی پیچشی تلفیق‌شده با واژگان احساسی[4] استفاده کردند. نشانش احساس بر پایه سنتی وردنت[5] برای هر واژه به کار رفت؛ سپس هر دو بردار نشانش واژه و نشانش احساس[6] به‌عنوان ورودی طبقه‌بندی‌کننده یک شبکه عصبی پیچشی در نظر گرفته شدند.

شبکه‌های عصبی هم‌گشتی نیز نوعی دیگر از شبکه‌های عصبی ژرف هستند که با اعمال مجموعه یکسانی از وزن‌ها به‌صورت هم‌گشتی روی ورودی ساخت‌یافته برای تولید ساختار پیش‌بینی‌شده بر روی ساختارهای ورودی با اندازه‌های مختلف استفاده می‌شوند.

[1] Local filter
[2] Caffe
[3] SGD
[4] Sentiment Lexical-Augmented Convolutional Neural Networks for Sentiment Analysis (SCNN)
[5] Senti-WordNet
[6] Sentiment embedding

[DOI: 10.52547/jsdp.19.1.19]

این نوع شبکه‌ها بیشتر برای ورودی‌هایی که دارای ساختار سلسله‌مراتبی[1] مانند جملات هستند، مناسب هستند. شبکه عصبی هم‌گشتی جزء روش‌های یادگیری با نظارت است که از ساختار درختی برای آموزش استفاده می‌کند و هرکدام از گره‌ها متریک‌های مربوط به خودشان را دارند.

ساچر و همکاران [18] مدل نیمه ناظر هم‌گشتی و خودکاری که شبکه عصبی هم‌گشتی[2] نامیده می‌شود پیشنهاد کردند که در آن به‌جای استفاده از کیسه‌ای از واژگان[3] از بردار واژگان متناوب به‌عنوان ورودی استفاده می‌شود. در این مدل از درخت دودویی که در آن گره‌های برگ با بردار واژگان مطابقت دارند، برای دسته‌بندی استفاده می‌شود.

ساچر و همکاران [19] در ادامه مدلی به نام بردار ماتریس شبکه عصبی هم‌گشتی ارائه کردند که شبیه شبکه عصبی هم‌گشتی است؛ با این تفاوت که برای نمایش؛ واژگان و عبارات با استفاده از درخت دودویی از دو بردار و یک ماتریس استفاده می‌کند. هم‌چنین آن‌ها [20] یک شبکه عصبی هم‌گشتی را بر پایه تنسور[4] و بانک درختی مربوط به احساسات معرفی کردند که اثرات ترکیبی موجود در سطوح مختلف عبارت مانند مثبت یا منفی‌بودن آن را به‌روشنی تعیین می‌کند. این مدل به‌جای استفاده از یک ماتریس بازنمایی[5] برای هر واژه و عبارت از توابع ترکیبی بر پایه تنسور برای تمام گره‌ها در درخت عبارت دودویی استفاده می‌کند.

با بررسی این شبکه‌ها می‌توان دریافت که هرکدام از آن‌ها دارای نقاط قوت و ضعف خود هستند (جدول ۱). در این راستا معرفی مدل‌های ترکیبی که بتواند به‌طور هم‌زمان با استفاده از نقاط قوت آن‌ها بر نقاط ضعفشان غلبه کند، نظر پژوهش‌گران زیادی را در سال‌های اخیر به خود جلب کرده است.

در این راستا، *چن و همکارانش* [21] برای بهبود عملکرد طبقه‌بندی احساسات مدلی ترکیبی بر پایه ترکیب شبکه عصبی پیچشی و شبکه عصبی حافظه طولانی کوتاه‌مدت پیشنهاد دادند. این مدل از بردارهای واژه پیش آموزش به‌عنوان ورودی استفاده می‌کند، سپس ویژگی‌ها به‌دو لایه حافظه طولانی کوتاه‌مدت هدایت می‌شوند که می‌تواند ویژگی‌های وابسته به متن را استخراج کرده و جملات نماینده را برای طبقه‌بندی احساسات ایجاد کند.

در این راستا، *چن و همکارانش* [21] برای بهبود عملکرد طبقه‌بندی احساسات مدلی ترکیبی بر پایه ترکیب شبکه عصبی پیچشی و شبکه عصبی حافظه طولانی کوتاه‌مدت پیشنهاد دادند. این مدل از بردارهای واژه پیش‌آموزش به‌عنوان ورودی و از شبکه عصبی پیچشی برای به‌دست‌آوردن ویژگی‌های محلی مهم متن استفاده می‌کند؛ سپس ویژگی‌ها به دو لایه حافظه طولانی کوتاه‌مدت هدایت می‌شوند که می‌تواند ویژگی‌های وابسته به متن را استخراج کرده و جملات نماینده را برای طبقه‌بندی احساسات ایجاد کند.

حسن و همکارش [22] برای غلبه بر کاستی‌های تحلیل احساسات متن‌های کوتاه در روش‌های سنتی و یادگیری عمیق، روش ConvLstm را پیشنهاد دادند که ترکیبی از شبکه عصبی پیچشی و حافظه طولانی کوتاه‌مدت روی بردارهای واژگان پیش آموزش‌دیده است تا وابستگی‌های بلندمدت را در متون کوتاه به‌طور مؤثر ثبت کند.

تیماراجو و همکارش [23] برای حل مشکل طبقه‌بندی احساسات در جملات و عبارات طولانی، ترکیبی از شبکه عصبی هم‌گشتی برای تحلیل سطح جمله و یک شبکه عصبی برگشتی برای تحلیل عبارت پیشنهاد دادند. در آزمایش‌های آن‌ها از مجموعه داده‌های آی‌ام‌دی بی[6] استفاده شد و آموزش بردارهای واژه روی این پیکره با استفاده از روش اسکیپ‌گرام صورت گرفته است.

وانگ و همکارانش [24] شبکه عصبی پیچشی و شبکه عصبی برگشتی و اشتراک بین آن‌ها را برای تحلیل احساساتی متون کوتاه ارائه کردند. تحلیل احساسات متون کوتاه به‌دلیل اطلاعات زمینه‌ای محدود، دشوار است. شبکه عصبی پیچشی در واژگان مجاور قادر به حفظ ویژگی‌های محلی و روابط متوالی آن‌ها در یک جمله است. *ون و همکارانش* [25] به مسأله تحلیل احساسات در سطح عبارت می‌پردازند. با توجه به عملکرد خوب دو شبکه عصبی پیچشی و برگشتی در تحلیل احساسات سطح عبارت و برای کاهش نقاط ضعف این دو مدل آن‌ها ترکیبی از شبکه عصبی پیچشی و برگشتی را پیشنهاد دادند. مدل دارای سه واحد لایه نشانش واژه، لایه پیچشی و حوزه انتخابی درخت حافظه طولانی کوتاه‌مدت است.

[1] Hierarchical structure
[2] Recursive Neural Network
[3] Bag-Of-Words
[4] Recursive Neural Tensor Network (RNTN)
[5] Representation Matrix
[6] IMDB

[DOI: 10.52547/jsdp.19.1.19]

اگرچه استفاده از روش‌های ترکیبی نیز باعث افزایش قابل توجه دقت در تجزیه و تحلیل احساسات شده است، اما این روش‌ها نیز هنوز در ابتدای مسیر قرار دارند و نیاز به پیشرفت در این حوزه همچنان محسوس است. در این راستا و به‌منظور بهبود و بهینه‌سازی روش‌های پیشین، هدف این مقاله ترکیب و استفاده هم‌زمان شبکه عصبی پیچشی و هم‌گشتی به‌منظور ایجاد هم‌افزایی و بهره گیری از مزایای آن‌ها است. گفتنی است که در بین روش های ترکیبی، این نخستین‌بار است که شبکه عصبی پیچشی و شبکه عصبی هم‌گشتی با هم ترکیب شده است. روش پیشنهادی این مقاله با توجه به این‌که از لایه هم گشتی به‌عنوان جایگزین لایه ادغام بهره می‌برد، در مقایسه با سایر روش‌های ترکیبی توانایی بیشتری در حفظ وابستگی‌های بلندمدت داشته و پایدارتر است. درواقع، روش پیشنهادی به‌کمک شبکه پیچشی ویژگی‌های سطح پایین را استخراج کرده و آن‌ها را به‌عنوان ورودی در اختیار شبکه همگشتی که توانایی استخراج ویژگی‌های سطح بالا را دارد، قرار می‌دهد. گفتنی است که استفاده از سازوکار توجه در شبکه عصبی پیچشی و ویژگی میانی در شبکه عصبی هم‌گشتی نیز به ترتیب باعث تأکید روی قسمت‌های مهم‌تر متن و افزایش انعطاف‌پذیری روش پیشنهادی در مقایسه با سایر مدل‌های ترکیبی می‌شود.

(جدول-۱): مقایسه مزایا و معایب شبکه‌های عصبی پیچشی و هم‌گشتی

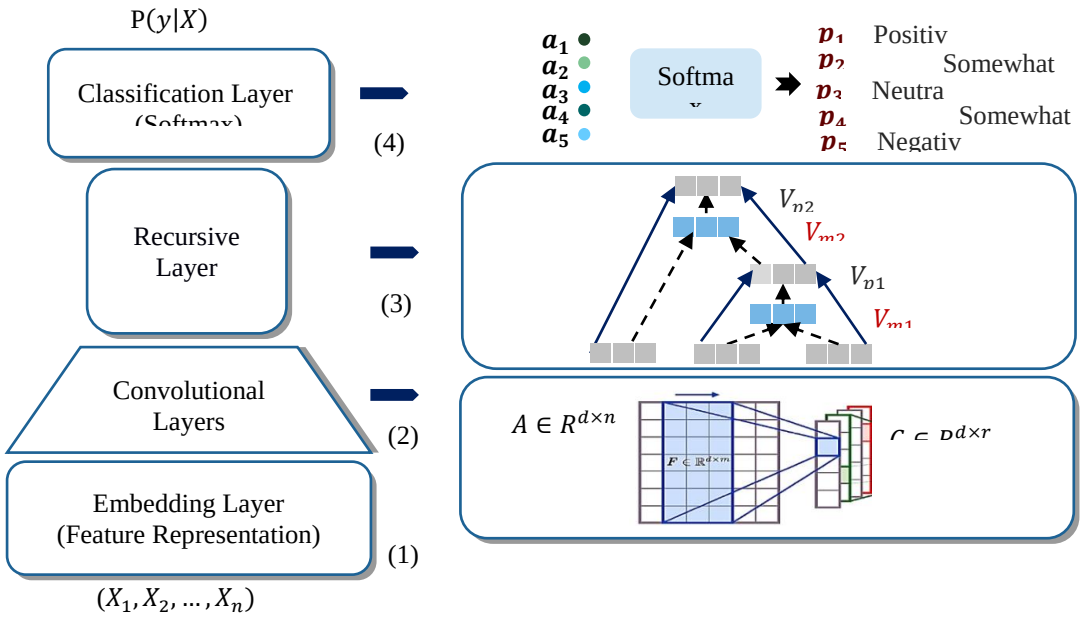
(Table-1): Comparison of the advantages and disadvantages of convolutional and recursive neural networks

نوع شبکه	مزایا	معایب
شبکه عصبی پیچشی	۱- توانایی بالا در استخراج ویژگی‌های محلی (عدم نیاز به مهندسی دستی ویژگی‌ها) ۲- توانایی یادگیری بردارهای بازنمایی ۳- سادگی الگوریتم ۴- افزایش دقت دسته‌بندی هم‌زمان با افزایش تعداد لایه‌ها ۵- مناسب برای جملات با طول کوتاه	۱-ناتوانی در استخراج ویژگی‌های سطح بالا ۲- ناتوانی در حفظ وابستگی‌های کلمات با فاصله طولانی و درک معنای جملات طولانی ۳- اختصاص ارزش یکسان به کلیه کلمات موجود در جملات ۴- از دست رفتن داده ها در اثر عملیات ادغام ۵- بار محاسباتی بالا در اثر افزایش تعداد لایه‌های میانی ۶- نیاز به تعداد زیاد داده‌های آموزشی
شبکه عصبی هم گشتی	۱- استخراج ویژگی‌های سطح بالا ۲- مناسب برای یادگیری دنباله‌های ترتیبی مانند متون ۳- توانایی حفظ وابستگی‌های کلمات با فاصله طولانی ۴- مناسب برای تحلیل جملات طولانی با توجه به ایجاد ساختار درختی	۱- بار محاسباتی بالا ۲- نیاز به داده‌های آموزشی با فرمت خاص ۳- مشکلات مربوط بیش برازش گرادیان ۴- تجزیه جمله به صورت نادرست با توجه به اینکه تأثیر کلمات روی یکدیگر در ایجاد درخت جمله مورد بررسی قرار نمی‌گیرد. ۵- عدم توانایی در استخراج ویژگی‌های محلی و سطح پایین

۳- روش ترکیبی پیشنهادی

در طول زمان روش‌های مختلفی برای پیاده‌سازی سیستم‌های تجزیه و تحلیل احساسات با استفاده از مفهوم یادگیری ژرف معرفی شدند که از مهم‌ترین آن‌ها می‌توان به شبکه‌های عصبی هم‌گشتی و شبکه‌های عصبی پیچشی اشاره کرد. با توجه به اینکه هرکدام از این روش‌ها ویژگی‌های منحصربه‌فرد خود را دارند، هدف این مقاله معرفی یک روش جدید بر پایه ترکیب شبکه عصبی پیچشی و شبکه عصبی هم‌گشتی برای تجزیه و تحلیل احساسات موجود در متن است. دیاگرام مربوط به رویکرد پیشنهادی در شکل (۱) نشان داده شده است. همان‌طور که در شکل مشخص است، رویکرد پیشنهادی شامل چهار لایه بهینه‌سازی شده که تشریح هر یک از آن‌ها و علت به‌کارگیری‌شان در ادامه آمده است.

[DOI: 10.52547/jsdp.19.1.19]

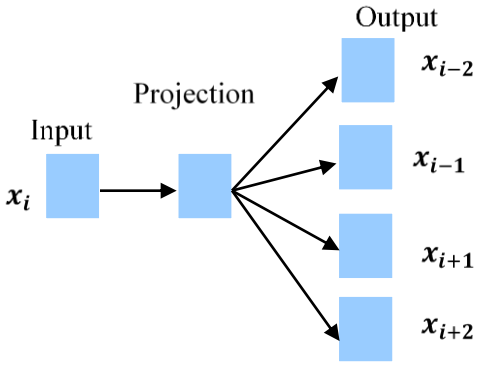


(شکل-۱): دیاگرام روش پیشنهادی شامل چهار لایه نشانش (بازنمایی ویژگی‌ها)، پیچش، بازگشتی و دسته‌بندی

(Figure-1): Diagram of the proposed model including embedding layer, convolutional layer, recursive layer and classification layer

۱-۳- لایه نشانش[1]

برای این‌که بتوانیم از واژگان به‌عنوان ورودی شبکه عصبی استفاده کنیم، باید هر واژه را به یک بردار تبدیل کنیم. ابعاد این بردار برابر تعداد واژگان موجود در واژه‌نامه است. در این بردار، درایه متناظر با هر واژه برابر یک و مابقی درایه‌های آن صفر است. این شیوه ساده‌ترین روش برای نمایش واژگان است که به آن بردار یک-داغ[2] می‌گویند. درنتیجه هر سند به‌صورت دنباله‌ای از بردارهایی $(x_1, x_2 \cdots x_n)$ با این ویژگی‌ها خواهد بود. با توجه به این‌که استفاده از این بردارها به‌تنهایی کارآمد نیست و توانایی نشان‌دادن واژگان مشابه را ندارد، در این لایه هرکدام از این بردارها به یک فضای برداری d-بعدی نگاشت می‌شود که به آن بازنمایی توزیع‌شده می‌گویند. بازنمایی توزیع‌شده واژگان در یک فضای برداری باعث می‌شود تا الگوریتم‌های یادگیری به کارایی بهتری دست یابند که ناشی از گروه‌بندی واژگان با معنای نزدیک به هم است [26]. با توجه به آزمایش‌های صورت‌گرفته روی مدل‌های مختلف برای تولید بردارهای بازنمایی، مدل اسکیپ‌گرام[3] در این مقاله مورد استفاده قرار گرفته است. ساختار مدل اسکیپ‌گرام در شکل (۲) نشان داده شده است [27].



(شکل-۲): ساختار مدل اسکیپ‌گرام [27]

(Figure-2): Skip-Gram model structure [27]

هدفی که در آموزش مدل اسکیپ‌گرام دنبال می‌شود، یافتن بردارهای بازنمایی برای واژگان است. این بردارها باید طوری آموزش ببینند که از روی بردار یک واژه بتوان بردار واژگانی را که پیرامون آن در یک جمله قرار می‌گیرند تخمین زد و به بیان دقیق‌تر درصورتی‌که توالی واژگان به‌صورت $x_1, x_2 \cdots x_n$ موجود باشد، هدفی که این مدل دنبال می‌کند، بیشینه‌کردن میانگین لگاریتم درست‌نمایی رابطه (۱) است [27]:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{-c\le j\le c, j\neq 0}\log p\left(x_{i+j}|x_i\right) \qquad (۱)$$

در این رابطه c نشان‌دهنده اندازه محتوایی[4] است که به‌ازای هر واژه باید تخمین زده شود. هرچه این اندازه بزرگ‌تر باشد، داده‌های آموزشی بیشتری مورد نیاز است و

[1] Embedding
[2] One-hot vector
[3] Skip-gram
[4] Content

* نویسندهٔ عهده‌دار مکاتبات
* Corresponding author

• تاریخ ارسال مقاله: ۱۳۹۸/۵/۲۴ • تاریخ پذیرش: ۱۳۹۹/۱۰/۲۱ • تاریخ انتشار: ۱۴۰۱/۰۳/۳۱ • نوع مطالعه: پژوهشی

[DOI: 10.52547/jsdp.19.1.19]

درنتیجه بردارهای بازنمایی با دقت بهتری محاسبه می‌شوند. در این روش، مقدار احتمال $p(x_o|x_I)$ با استفاده از تابع سافت مکس[1] به‌صورت رابطه (۲) محاسبه می‌شود [27]:

$$p(x_o|x_i) = \frac{\exp({v'_{x_o}}^T v_{x_I})}{\sum_{x=1}^{|v|}({v'_x}^T v_{x_I})} \quad (۲)$$

در این رابطه v'_x و v_x به‌ترتیب بردارهای بازنمایی ورودی و خروجی برای واژه x است و $|v|$ نیز نشان‌دهنده تعداد واژگان قرارگرفته در واژه‌نامه است. درواقع در این روش برای هر واژه دو بردار بازنمایی در نظر گرفته می‌شود که در ابتدا هرکدام به‌صورت تصادفی مقداردهی شده‌اند. سپس در حین آموزش این مقادیر به‌نحوی اصلاح می‌شوند که رابطه (۱) بیشینه شود. در نهایت، برداری که به‌عنوان بردار بازنمایی برای واژگان مورداستفاده قرار می‌گیرد، از میانگین یا الحاق این دو بردار حاصل می‌شود [27].

۲-۳- لایه پیچشی بر پایه ساز و کار توجه

شبکه‌های عصبی پیچشی از لایه‌هایی به همراه فیلترهای پیچشی استفاده می‌کنند که این فیلترها به ویژگی‌های محلی اعمال می‌شوند. درواقع، در این نوع شبکه‌ها به‌جای اینکه تمامی نرون‌های یک لایه به لایه بعد متصل باشند، تنها قسمتی از نورون‌های یک لایه به لایه بعد متصل هستند. علت انتخاب شبکه عصبی پیچشی در این لایه است که این نوع شبکه‌ها پارامترهای کمی دارند و آموزش آن‌ها ساده است و می‌توانند ویژگی‌های محلی را استخراج کنند و با افزایش تعداد لایه‌های شبکه‌های عصبی پیچشی نیز می‌توان ویژگی‌های با ارزش‌تری را از توالی ورودی استخراج کرد [28].

به‌طورکلی، یک لایه پیچشی از دو مرحله تشکیل شده است: در مرحله نخست برای انجام عملیات پیچش یک فیلتر $w \in R^{c\times d}$ روی مجموعه ورودی از لایه نشانش اعمال می‌شود. درصورتی‌که فرض شود $A[i:j]$ پنجره‌ای شامل بردار واژگانی باشد که میان i امین و j امین سطر ماتریس A قرار دارند، ویژگی $\bar{x_i}$ بر اساس پنجره‌ای از واژگان با اندازه c بر اساس رابطه (۳) محاسبه می‌شود [28].

$$\bar{x_i} = f(w \circ A[i:i+c-1] + b) \quad (۳)$$

در این رابطه متغیر i به‌صورت $i = 1 \dots n-c+1$ مقدار می‌گیرد و $w \in R^{c\times d}$ و $b \in R$ نیز به‌ترتیب نشان‌دهنده ماتریس فیلتر پیچش و عملیات بایاس هستند. f و $\circ$ نیز به ترتیب توابع فعال‌سازی (مانند Tanh و Relu) و عملگر پیچش هستند. با اعمال فیلتر w به تمامی پنجره‌های ممکن از واژگان و قراردادن ویژگی‌های حاصل از آن‌ها در یک بردار یک نگاشت ویژگی به‌صورت $\bar{x} = [\overline{x_1}, \overline{x_2}, \dots, \overline{x_{n-c+1}}]$ تشکیل خواهد شد که در آن $\bar{x} \in R^{n-c+1}$ است [29].

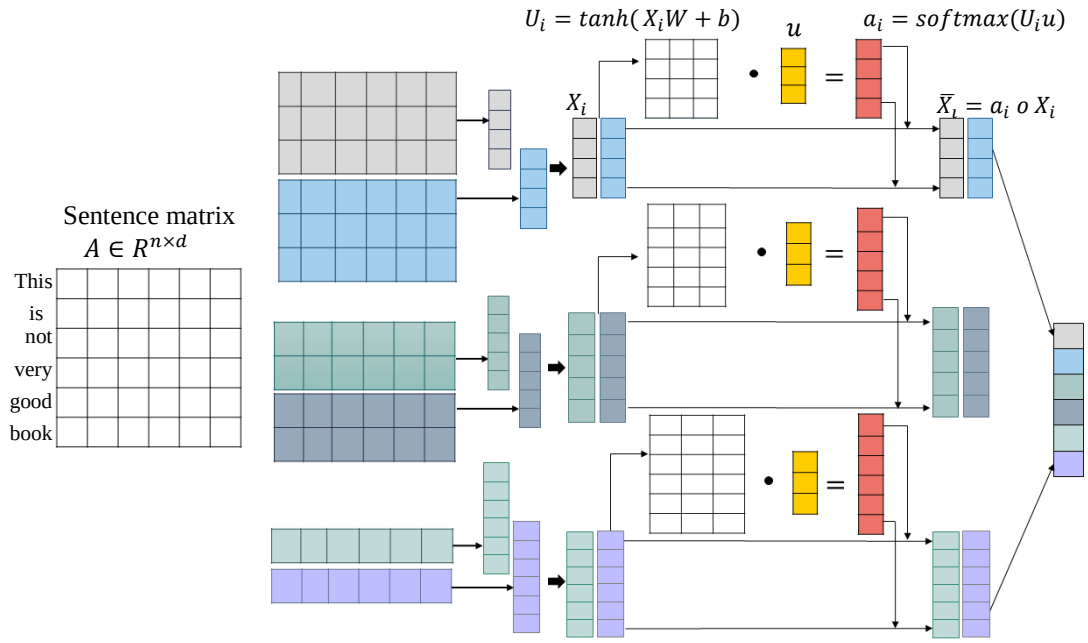
این باور وجود دارد که تمامی اطلاعاتی که در متن وجود دارند، ارزش یکسانی نداشته و به بیان دیگر برخی از واژگان موجود در جمله، تأثیری روی معنای جمله ندارند. یکی از چالش‌های شبکه عصبی پیچشی نیز ناتوانی آن برای تأکید روی واژگان با اهمیت بیشتر است. به همین دلیل نیاز به راه‌کاری است که بتوان به‌کمک آن روی قسمت‌های مهم‌تری ازجمله تأکید بیشتری داشت. برای دست‌یابی به این هدف در این مقاله در لایه پیچشی از سازوکار توجه استفاده شده است. ساختار شبکه پیچشی تلفیق شده با سازوکار توجه پیشنهادی در شکل (۳) نشان داده شده است.

همان‌طور که مشخص است، پس از اعمال فیلترهای پیچش به ماتریس حاصل از سند ورودی (A)، نگاشت‌های ویژگی متناظر با فیلترهای هم‌اندازه کنار یکدیگر قرار می‌گیرند و خود تشکیل یک ماتریس می‌دهند. هر سطر از این ماتریس متناظر با یک بردار ویژگی است که به‌وسیله فیلترهای هم‌اندازه مختلف از یک بخش خاص از سند ورودی استخراج شده‌اند. بر روی هر سطر از این ماتریس با استفاده از سازوکار توجه یک وزن اعمال می‌شود. این کار باعث ایجاد تغییر در مقادیر عناصر نگاشت‌های ویژگی استخراج‌شده می‌شود. به بیانی دقیق‌تر فرض می‌شود که در شبکه پیچشی تعداد M اندازه مختلف برای فیلترها در نظر گرفته شده است که به‌ازای هرکدام از این اندازه‌ها m فیلتر مختلف به‌کار رفته می‌شود. به‌این‌ترتیب پس از اعمال فیلترهای $w_{ij} \in R^{c_i\times d}$ (که در آن $i = 1,2, \dots M$ و $j = 1,2, \dots m$) به ماتریس سند ورودی تعداد $M \times m$ نگاشت ویژگی حاصل می‌شود. از کنار هم قرارگرفتن بردارهای نگاشت ویژگی مربوط به فیلترهایی با اندازه $c_i \times d$ یک ماتریس به‌صورت زیر (رابطه ۴) حاصل می‌شود:

$$\bar{X}_i = \begin{bmatrix} \bar{x}_{1,1} & \cdots & \bar{x}_{1,m} \\ \vdots & \ddots & \vdots \\ \bar{x}_{n-c_i+1,1} & \cdots & \bar{x}_{n-c_i+1,m} \end{bmatrix} \quad (۴)$$

[1] Softmax

[DOI: 10.52547/jsdp.19.1.19]



(شکل-۳): نمای کلی لایه پیچشی بر پایه سازوکار توجه

(Figure 3): Overall Diagram of the convolutional layer based on attention mechanism

در این رابطه n نشان‌دهنده تعداد واژگان به‌کار رفته در سند است. عناصر موجود در سطرهای این ماتریس ویژگی‌هایی هستند که به‌وسیلهٔ فیلترهای هم‌اندازه مختلف از بخش خاصی از سند ورودی استخراج شده‌اند. هدف تخصیص وزن‌هایی به هرکدام از این سطرها است تا بتوان به‌وسیله آن‌ها به بخش‌هایی از سند که حاوی اطلاعات مفیدتری برای دسته‌بندی هستند، اهمیت بیشتری داد. این وزن‌ها مطابق روابط (۵) و (۶) محاسبه می‌شوند:

$$U_i = \tanh(\bar{X}_i W + b) \qquad (۵)$$

$$a_i = \mathrm{softmax}(U_i u) \qquad (۶)$$

در این روابط ابتدا ماتریس $\bar{X}_i$ با استفاده از یک شبکه پرسپترون تک‌لایه و با استفاده از ماتریس وزن $W \in R^{m \times d}$ به ماتریس میانی $U_i \in R^{n-c_i+1 \times d}$ نگاشت می‌شود، سپس بر اساس میزان شباهت کسینوسی هرکدام از سطرهای ماتریس U_i با بردار محتوای $u \in R^{d \times 1}$ یک وزن محاسبه می‌شود که نشان‌دهنده میزان اهمیت ویژگی‌های استخراج‌شده در سطر متناظر ماتریس $\bar{X}_i$ است. d نشان‌دهنده بعد بردار محتوای u است. در رابطه (۶) عناصر بردار $a_i \in R^{n-c_i+1 \times 1}$ وزن‌های مربوط به هرکدام از سطرهای ماتریس $\bar{X}_i$ هستند. با توجه به تابع سافت‌مکس به کارگرفته‌شده در این رابطه، مجموع عناصر a_i برابر با یک خواهد بود. در ابتدای آموزش بردار u با صفر مقداردهی اولیه می‌شود. به‌این‌ترتیب در رابطه (۶) وزن یکسانی برای سطرهای مختلف ماتریس $\bar{X}_i$ محاسبه می‌شود. در ادامه این وزن‌ها با استفاده از نمونه‌های آموزشی یاد گرفته می‌شوند. برای استفاده از اطلاعات مربوط به وزن‌های به‌دست‌آمده برای دسته‌بندی هرکدام از این وزن‌ها در سطرهای متناظرشان در ماتریس $\bar{X}_i$ مطابق رابطه (۷) ضرب می‌شوند:

$$\hat{X}_i = a_i \, o \, \bar{X}_i \qquad (۷)$$

در این رابطه o نشان‌دهنده عملگری است که هرکدام از عناصر بردار a_i را در سطر متناظرش در $\bar{X}_i$ ضرب می‌کند. همان‌طور که گفته شد، ستون‌های ماتریس $\bar{X}_i$ نگاشت ویژگی هستند که با استفاده از اعمال فیلترهای مختلف روی ماتریس سند ورودی حاصل شده‌اند؛ بنابراین عناصر ستون‌های ماتریس $\hat{X}_i$ متناظر با همان نگاشت ویژگی‌هایی هستند که در یک بردار وزن ضرب شده‌اند.

با توجه به این‌که نگاشت ویژگی حاصل‌شده به‌اندازه سند و اندازه فیلتری که بر آن اعمال می‌شود، بستگی دارد، نیاز به یک بردار با طول ثابت است که بتوان دسته‌بندی را بر اساس آن انجام داد. این عملیات در مرحله دوم و به‌وسیلهٔ یک تابع ادغام صورت می‌گیرد که هدف اصلی آن کاهش ابعاد است؛ اما لایه ادغام با توجه

[DOI: 10.52547/jsdp.19.1.19]

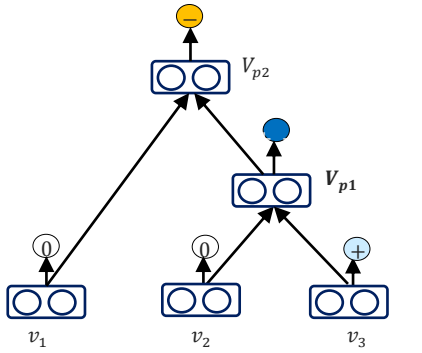
به اینکه از بین ویژگی‌های موجود میانگین، بیشینه و یا کمینه آن‌ها را انتخاب می‌کند، منجر به ازدست‌رفتن بسیاری از ویژگی‌های محلی می‌شود. از طرف دیگر، یکی دیگر از مشکلات شبکه‌های عصبی پیچشی عدم توانایی آن‌ها در حفظ وابستگی‌های بلندمدت است.

نتایج آزمایش‌ها نشان داده‌اند که شبکه‌های عصبی پیچشی برای حفظ وابستگی‌های بلندمدت به تعداد زیادی لایه پیچشی نیاز دارند. با توجه به اینکه افزایش تعداد لایه‌ها موجب افزایش بار محاسباتی سامانه می‌شود و امروزه همواره تلاش بر این است که با کاهش تعداد لایه و محاسبات، دقت سامانه افزایش یابد. در این راستا در روش پیشنهادی از شبکه عصبی هم‌گشتی به‌عنوان جایگزین لایه ادغام در شبکه عصبی پیچشی استفاده می شود. شبکه‌های عصبی هم‌گشتی برخلاف شبکه‌های عصبی پیچشی منجر به استخراج ویژگی‌هایی می‌شود که عبارات معناداری را نشان می‌دهند و می‌توانند ساختار جملات را به‌صورت درختی نشان دهند که باعث حفظ ترتیب واژگان در جمله و وابستگی‌های بلندمدت می‌شود.

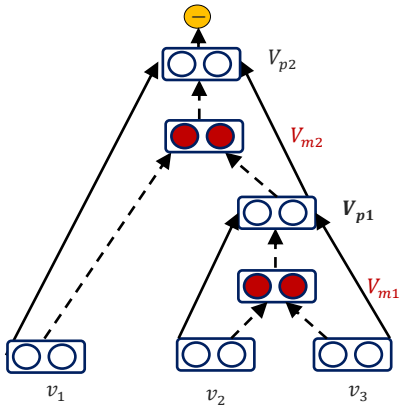
۳-۳- لایه هم‌گشتی تعمیم‌یافته

همان‌طور که پیش از این اشاره شد، مهم‌ترین مشکل لایه پیچشی عدم توانایی آن‌ها برای استخراج ویژگی‌های سطح بالا و حفظ وابستگی‌های طولانی‌مدت است. گفتنی است که ساختار درختی درک بهتری از جمله ایجاد می‌کند و از آن‌جایی‌که در فرآیند تجزیه و تحلیل احساسات نحوه قرارگیری واژه در جمله در مشخص‌کردن قطبیت آن تأثیر قابل‌توجهی دارد، در روش پیشنهادی این مقاله از شبکه عصبی هم‌گشتی به‌عنوان جایگزین لایه ادغام استفاده می‌شود که باعث می‌شود ضمن کاهش از دست رفتن ویژگی‌های محلی، ترتیب واژگان در جمله لحاظ شده و وابستگی‌های بلندمدت حفظ شود. از طرف دیگر، آزمایش‌ها نشان داده شده است که درخت‌های ایجادشده به‌وسیلهٔ شبکه‌های عصبی هم‌گشتی همواره درست نیستند و با افزایش تعداد ویژگی‌ها، دقت این شبکه‌ها نیز افزایش می‌یابد. در این راستا، روش‌های مختلفی در طول زمان معرفی شدند که سعی داشتند ویژگی‌هایی را به‌صورت دستی به ساختار درختی شبکه‌های عصبی هم‌گشتی اضافه کنند [19,20]. به‌بیان‌دیگر، ایده‌ای که پشت یادگیری ژرف وجود دارد و آن را جذاب می‌کند؛ عدم نیاز به مهندسی دستی ویژگی‌ها است؛ ازاین‌رو، در این مقاله برای رسیدن به این هدف در لایه هم‌گشتی، از یک ویژگی میانی جدید حاصل از ترکیب ویژگی‌های موجود با هدف افزایش دقت درخت تجزیه جملات و پیرو آن دقت دسته‌بندی استفاده شد.

در شکل (۴- الف) ساده‌ترین شبکه عصبی هم‌گشتی نشان داده شده است که در آن v_1، v_2 و v_3 بردار بازنمایی واژگان ورودی و V_{p_1} بردار خروجی لایه پنهان در گره والد است که به‌صورت $V_{p_1} = f(w\begin{bmatrix}v_2\\v_3\end{bmatrix} + b)$ محاسبه می‌شود (f یک تابع فعال‌ساز است). شکل (۴-ب) یک شبکه عصبی هم‌گشتی تعمیم‌یافته را نشان می‌دهد که برای ایجاد یک تعامل قوی‌تر بین بردارهای گره فرزندان و گره والد یک ویژگی جدید از طریق لایه عصبی میانی اضافه شده است.



الف



ب

(شکل-۴): (الف) درخت تولیدشده برای جملات با استفاده از (ب) درخت تولیدشده برای [16] شبکه عصبی بازگشتی جملات با استفاده از شبکه عصبی بازگشتی تعمیم‌یافته

(Figure-4): (a) Sentence tree generated using recursive neural network [16] (b) sentence tree generated using generalized recursive neural network

در این شبکه برای محاسبه بردار والد V_{p_1} باید V_{m_1} با استفاده از یک لایه عصبی تودرتو محاسبه شود و سپس بردار والد با استفاده از این ویژگی جدید محاسبه خواهد شد. درواقع یک ویژگی جدید به‌صورت

[DOI: 10.52547/jsdp.19.1.19]

$V_{m_1} = f(w\begin{bmatrix}v_2\\v_3\end{bmatrix} + b)$ به‌دست خواهد آمد که برای تولید گره والد به‌صورت $V_{p_1} = f(w\begin{bmatrix}v_2\\v_3\\v_{m_1}\end{bmatrix} + b)$ استفاده خواهد شد. به‌طورکلی در این لایه با توجه به داشتن یک گراف غیر چرخشی جهت‌دار موقعیت، از گره‌های با ترتیب توپولوژیکی و به‌صورت هم‌گشتی برای تولید بازنمایی‌های بیشتر با استفاده از بازنمایی‌های محاسبه‌شده برای فرزندان استفاده می‌شود.

۴-۳- لایه دسته‌بندی

هدف اصلی این لایه مشخص‌کردن قطبیت احساسات است. اساس لایه دسته‌بندی، دسته‌بند رگرسیون ترابری است که با داشتن ورودی با ابعاد مشخص از لایه‌های قبلی عملیات طبقه‌بندی را به‌کمک تابع فعال‌ساز سافت‌مکس (رابطه ۸) انجام می‌دهد. در این رابطه a_i وزن ورودی‌ها و P_i ردهٔ خروجی و k نشان‌دهنده تعداد دسته‌ها است [30].

$$Softmax: P_i = \text{p(class=i)} = \frac{e^{a_i}}{\sum_{c=1}^{c=k} e^{a_c}} \quad (۸)$$

گفتنی است طی آموزش این روش سعی خواهد شد که تابع زیان[1] رابطه (۹) کمینه شود. مقدار این تابع با استفاده از خروجی لایه سافت‌مکس شبکه به‌ازای نمونه‌های آموزشی و برچسب‌های متناظرشان محاسبه می‌شود [30].

$$loss(\hat{y}, y, W) = \frac{1}{n} cross\_entroy(\hat{y}, y) + \lambda \|W\|_2 \quad (۹)$$

در این رابطه $\hat{y}$ مجموعه حاوی خروجی‌های لایه سافت‌مکس به‌ازای نمونه‌های آموزشی و y مجموعه برچسب‌های متناظر با نمونه‌های آموزشی و n نشان‌دهنده تعداد نمونه‌های آموزشی است. W نشان‌دهنده ماتریس وزن و ضریب λ میان عبارت تنظیم ساز[2] و میانگین خطای کراس متقابل[3] تعادل برقرار می‌کند. برای آموزش روش نیز از گرادیان نزولی تصادفی استفاده خواهد شد [30].

به‌طورکلی، جنبه نوآوری این مقاله را می‌توان به دو بخش تقسیم کرد. ازآنجایی‌که هدف این مقاله تعیین میزان قطبیت احساسات است و در احساسات کاربران ترتیب واژگان در جمله در معنا تأثیرگذار است، به‌منظور استخراج بهینه ویژگی‌ها شبکه‌های عصبی پیچشی و عصبی هم‌گشتی با هم ترکیب شدند. علت استفاده از شبکه‌های عصبی پیچشی توانایی آن در استخراج ویژگی سطح پایین است که در حالت عادی استخراج آن‌ها غیرممکن است و می‌توانند به‌عنوان ورودی شبکه‌های عصبی هم‌گشتی بسیار مناسب باشند. علت استفاده از شبکه عصبی هم‌گشتی نیز توانایی آن در حفظ وابستگی‌های بلندمدت است که در جملات طولانی برای درک معنای جمله از اهمیت بالایی برخوردار است. گفتنی است که استفاده از شبکه عصبی هم‌گشتی به‌عنوان جایگزین لایه ادغام منجر به کاهش ازدست‌رفتن داده‌های محلی نیز می‌شود. هم‌چنین در روش پیشنهادی برای افزایش دقت شبکه عصبی پیچشی و تأکید بر واژگانی که دارای تأثیر بیشتری روی معنای جمله هستند از سازوکار توجه در لایه پیچشی و به‌منظور افزایش دقت درخت‌های تجزیه جملات تولیدشده به‌وسیلهٔ شبکه عصبی هم‌گشتی از شبکه‌های عصبی هم‌گشتی تعمیم‌یافته که از یک ویژگی میانی حاصل از ترکیب گره‌های فرزندان بهره می برد، استفاده شده است. انتظار داریم که روش پیشنهادی با توجه به موارد گفته‌شده نه‌تنها در مقایسه با روش‌های موجود از دقت بالاتری برخوردار باشد، بلکه بتواند وابستگی‌های بلندمدت را در جمله حفظ کند و معنایی جملات را به‌درستی تشخیص دهد.

۴- نتایج و آزمایش‌ها

۴-۱- مجموعه‌داده

در کلیه آزمایش‌های انجام‌شده این مقاله از مجموعه‌داده استنفورد[4] به‌عنوان مجموعه آموزشی استفاده شده است که خود دارای دو زیر بخش SST1 و SST2 است. جزئیات مرتبط با این دو مجموعه‌داده در جدول (۲) نشان داده شده است.

(جدول-۲): خلاصه آماری مجموعه‌داده‌های SST1 و SST2
(Table-2): Summary Statists of SST1 and SST2 datasets

مجموعه‌داده	تعداد کلاس	اندازه واژگان	طول میانگین جملات	مجموعه‌داده/ تعداد جمله
SST1	5	18	18K	8544 /Train 1101 /Dev 2210 /Test
SST2	2	19	15K	6920 /Train 873 /Dev 1821 /Text

[1] Loss
[2] Regularization term
[3] Cross-entropy
[4] Stanford Sentiment Treebank

[DOI: 10.52547/jsdp.19.1.19]

به‌طورکلی، این مجموعه شامل ۱۱۸۵۵ جمله (نظر) است که از تارنمای rottentomatoes.com استخراج شده‌اند و مرتبط با نظرات افراد درباره فیلم‌های مختلف هستند. SST1 دارای پنج رده و سه زیرمجموعه train، dev و text مجزا است که هرکدام از آن‌ها به‌ترتیب دارای ۱۱۰۱،۸۵۴۴ و ۲۲۱۰ جمله است. گفتنی است که مجموعه train شامل عبارات استخراج‌شده از جملات و خود جملات است درحالی‌که مجموعه test تنها خود جملات را در برمی‌گیرد.

در ادامه برای انجام ارزیابی‌های بیشتر، تمام جملات خنثی این مجموعه‌داده نادیده گرفته شده و جملات منفی و خیلی منفی و مثبت و خیلی مثبت در دو گروه مثبت و منفی ادغام شدند. این مجموعه دارای دو رده است که با نام SST2 در آزمایش‌ها مورد استفاده قرار می‌گیرد. همانند SST1، SST2 نیز دارای سه زیرمجموعه train، dev و text مجزا است که هرکدام از آن‌ها به‌ترتیب دارای ۸۷۳،۶۹۲۰ و ۱۸۲۱ جمله است.

۲-۴- روش ارزیابی

اهمیت الگوریتم‌های یادگیری ژرف زمانی مشخص می‌شود که دانش تولیدشده در مرحله یادگیری مدل، در مرحله ارزیابی مورد تحلیل قرار گیرد تا بتوان ارزش آن را تعیین نمود و در پی آن کارایی مدل یادگیرنده را نیز مشخص کرد. در این مقاله برای ارزیابی روش پیشنهادی از معیار دقت[1] دسته‌بندی برای داده‌های آزمایشی به‌عنوان معیار ارزیابی استفاده کرد. مقدار این معیاری از تقسیم تعداد نمونه‌هایی که درست دسته‌بندی شده‌اند بر تعداد کل نمونه‌های آزمایشی مطابق رابطه (۱۰) به‌دست می‌آید. گفتنی است که در مقالات این حوزه که آزمایش‌ها خود را روی مجموعه‌داده‌های SST1 و SST2 انجام داده و برای مقایسه در این مقاله نیز مورد استفاده قرار گرفته‌اند، از همین معیار برای ارزیابی استفاده شده است.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (۱۰)$$

در این رابطه، TN و TP به ترتیب تعداد نمونه‌هایی هستند که مدل به‌درستی در دسته‌های مثبت و منفی قرار داده است و FN نیز نشان‌دهنده تعداد نمونه‌هایی است که به‌اشتباه در دسته منفی و FP نشان‌دهنده تعداد نمونه‌هایی است که به‌اشتباه در دسته مثبت قرارگرفته‌اند [38].

[1] Accuracy

۳-۴- توصیف آزمایش‌ها

به‌منظور بررسی دقیق عملکرد روش پیشنهادی، آزمایش های مختلفی با چندین ترکیب مختلف از روش پیشنهادی صورت گرفته است که توضیحات آن در ادامه آمده است. گفتنی است که در آزمایش‌های این بخش، از بردارهای واژگانی که با استفاده از مدل بازنمایی ویژگی اسکیپ‌گرام آموزش دیده‌اند به‌عنوان ورودی روش پیشنهادی استفاده می‌شود.

- **CNN-Attention-RNN-Plus-Rand**:

ترکیب شبکه عصبی پیچشی بر پایه سازوکار توجه و شبکه عصبی هم‌گشتی با ویژگی میانی معرفی شده است که از بردارهایی واژگانی که به‌صورت تصادفی مقداردهی شده‌اند، به‌عنوان ورودی شبکه عصبی پیچشی استفاده می‌کند.

- **CNN-Attention-RNN-Plus-Static:**

ترکیب شبکه عصبی پیچشی بر پایه سازوکار توجه و شبکه عصبی هم‌گشتی با ویژگی میانی معرفی شده است که از بردارهای واژگانی که با استفاده از مدل اسکیپ‌گرام تولید شده‌اند، به‌عنوان ورودی استفاده می‌کند. گفتنی است که این بردارها در طی فرآیند آموزش ثابت بوده و وزن بردار واژگان در طی فرآیند آموزش به‌روزرسانی نمی‌شود.

- **CNN-Attention-RNN-Plus-nonStatic** :

ترکیب شبکه عصبی پیچشی بر پایه سازوکار توجه و شبکه عصبی هم‌گشتی با ویژگی میانی معرفی شده است که از بردارهای واژگانی که با استفاده از مدل اسکیپ‌گرام تولید شده‌اند، به‌عنوان ورودی استفاده می‌کند. گفتنی است که این بردارها در طی فرآیند آموزش ثابت نیستند و وزن آن‌ها در طی فرآیند آموزش به‌روزرسانی می‌شود.

- **CNN-Attention-RNN-Plus-2channel:**

ترکیب شبکه عصبی پیچشی بر پایه سازوکار توجه و شبکه عصبی هم‌گشتی با ویژگی میانی معرفی شده است که از ترکیب بردارهای واژگانی که با استفاده از مدل اسکیپ‌گرام تولید شده‌اند و بردارهایی که به‌صورت تصادفی مقداردهی شده‌اند، به‌عنوان ورودی استفاده می‌کند.

با توجه به این‌که تابع هزینه شبکه‌های عصبی غیرقابل انعطاف است، الگوریتم‌هایی که برای آموزش مورد استفاده قرار می‌گیرند، تنها قادر به یافتن بهینه مطلوب هستند، درنتیجه، با توجه به پارامترهای اولیه اختصاص

[DOI: 10.52547/jsdp.19.1.19]

داده‌شده، بهینه‌های مختلف محلی می‌توانند در طول اجراهای مختلف از یک مدل به‌دست آیند. بر اساس مدل ارزیابی استفاده‌شده توسط *ون و همکارانش* [12,25]، تمامی انواع مختلف آزمایش‌ها به‌کمک میانگین و انحراف معیار در پنج بار اجرا مورد ارزیابی قرار گرفتند. گفتنی است که بهترین دقت در میان پنج اجرا در بخش نتایج گزارش شده است و پارامترهای مورد استفاده انواع مختلف پیاده‌سازی‌های روش پیشنهادی در جدول (۳) نشان داده شده‌اند که در آن r نشان‌دهنده اندازه سلول حافظه است و |θ|به تعداد پارامترهای قابل‌آموزش روش پیشنهادی اشاره می‌کند.

(جدول-۳): مقایسه اندازه سلول حافظه و تعداد پارامترهای آموزش‌پذیر برای انواع پیاده‌سازی‌های روش پیشنهادی

(Table-3): Comparison of memory cell size and number of trainable parameters between variations of the proposed model

انواع مختلف پیاده‌سازی‌های روش پیشنهادی	r	$\|\theta\|$
CNN-Attention-RNN-Plus-Rand	151	483,652
CNN-Attention-RNN-Plus -Static	165	499,435
CNN-Attention-RNN-Plus -Static	172	835,257
CNN-Attention-RNN-Plus-2channel	175	744,161

۴-۴- آموزش و ابر پارامترها

در آزمایش‌های این مقاله از روش اسکیپ‌گرام برای آموزش بردارهای واژگان استفاده شد که در آن اندازه پنجره برابر پنج و ابعاد بردار واژگان برابر ۱۵۰ در نظر گرفته شد. گفتنی است که در طول آموزش مدل اسکیپ گرام نرخ یادگیری برای به‌روزرسانی بردار واژگان و کمینه‌کردن بردار هزینه برابر ۰/۰۲۵ بود. در شبکه عصبی پیچشی نیز اندازه و تعداد فیلترها به‌عنوان ابر پارامتر در نظر گرفته می‌شود. در شبکه پیچشی این مقاله اندازه فیلترها برابر ۳، ۴ و ۵ و تعداد آن‌ها برابر ۱۵۰ بوده است که با برابر با ابعاد بردار واژگان است. از تابع ReLu نیز به‌عنوان تابع فعال‌ساز استفاده شده و نرخ یادگیری برابر ۰/۰۰۱ بوده است. در شبکه عصبی هم‌گشتی نیز تعداد واحدهای پنهان به‌عنوان ابر پارامتر در نظر گرفته می‌شود که در آزمایش‌های این بخش بین ۳۰ تا ۴۰ بوده و نرخ یادگیری نیز برابر ۰/۰۱ است. جزئیات مرتبط با تنظیم ابر پارامترهای این بخش از آزمایش‌ها نیز در جدول (۴) نشان داده شده است.

از طرف دیگر، پایداری یکی از مهم‌ترین ویژگی‌های شبکه‌های یادگیری ژرف به‌حساب می‌آید که سعی می‌کند میزان خطای آموزش و آزمایش را به یک میزان کاهش دهد. در آزمایش‌های این مقاله نیز از روش از قلم انداختن[1] نیز برای افزایش پایداری و کاهش بیش برازش[2] استفاده شده است. از قلم انداختن یک روش تنظیم است که با نادیده‌گرفتن تصادفی برخی از نمونه‌ها در طول فرآیند آموزش و در هر تکرار از بیش برازش روش جلوگیری کرده و باعث پایداری آن می‌شود. در این راستا روش پیشنهادی با نرخ ازقلم‌انداختن ۰/۵ در لایه پیچشی و ۰/۲ در لایه هم‌گشتی تنظیم شد. گفتنی است استفاده از لایه هم‌گشتی با توجه به این واقعیت که ساختار درختی کمتر نسبت به داده‌های دور افتاده تحت تأثیر قرار می‌گیرد، به‌خودی‌خود نیز باعث افزایش پایداری روش پیشنهادی می‌شود.

(جدول-۴): مقدار ابر پارامترهایی ترکیب متوالی شبکه پیچشی و هم‌گشتی با آن‌ها تنظیم شد

(Table-4): Hyper parameters values used for regularizing embedding, convolutional and recursive layers

لایه	ابر پارامترهای شبکه	مقادیر تخصیص‌یافته
لایه نشانش	بعد بردارهای واژگان	150
	اندازه پنجره	5
	نرخ یادگیری	0.025
لایه پیچشی	اندازه فیلترها	3- 4- 5
	تعداد فیلترها	150
	نرخ از قم انداختن	0.5
	نرخ یادگیری	0.001
لایه هم گشتی	تعداد واحدهای مخفی	30- 40
	نرخ یادگیری	0.01
	نرخ از قلم انداختن	0.2

۴- نتایج و بحث

به‌منظور نشان دادن عملکرد روش پیشنهادی و اثبات این ادعا که استفاده از ساز و کار توجه در شبکه عصبی پیچشی، بهره‌بردن از شبکه عصبی هم‌گشتی به‌عنوان جایگزین لایه ادغام و استفاده از یک ویژگی میانی در شبکه عصبی هم‌گشتی منجر به افزایش دقت دسته‌بندی می‌شود، روش پیشنهادی با طیف وسیعی از روش‌های موجود مورد مقایسه قرار گرفته است که نتایج آن در جدول (۵) نشان داده شده است. این جدول دارای دو بخش است که در بخش نخست نتایج مرتبط با مدل‌های موجود و در بخش دوم نتایج حاصل از پیاده‌سازی انواع مختلف روش پیشنهادی روی مجموعه‌داده‌های SST1 و SST2 نشان داده شده است. با توجه به نتایج به‌دست‌آمده می‌توان گفت که روش‌های بر پایه یادگیری ژرف به‌طور میانگین از دقت بالاتری نیست به روش‌های

[1] Dropout
[2] Overfitting

[DOI: 10.52547/jsdp.19.1.19]

یادگیری ماشین برخوردار هستند. این مسأله هم‌چنین می‌تواند توانایی روش‌های بر پایه یادگیری ژرف را در ایجاد مفهوم نمایش معنایی متن به‌طور مؤثر ثابت کند. هم‌چنین دقت بالاتر روش‌های بر پایه یادگیری ژرف نیز تأییدکننده توانایی این روش‌ها در تشخیص ویژگی‌های مستخرج از اطلاعات متنی در زمانی است که مشکل کمبود و پراکندگی داده‌ها وجود ندارد. به‌طورکلی، با در نظرگرفتن نتایج به‌دست‌آمده در جدول (۵) می‌توان به مسائلی که در ادامه آمده است، پی برد:

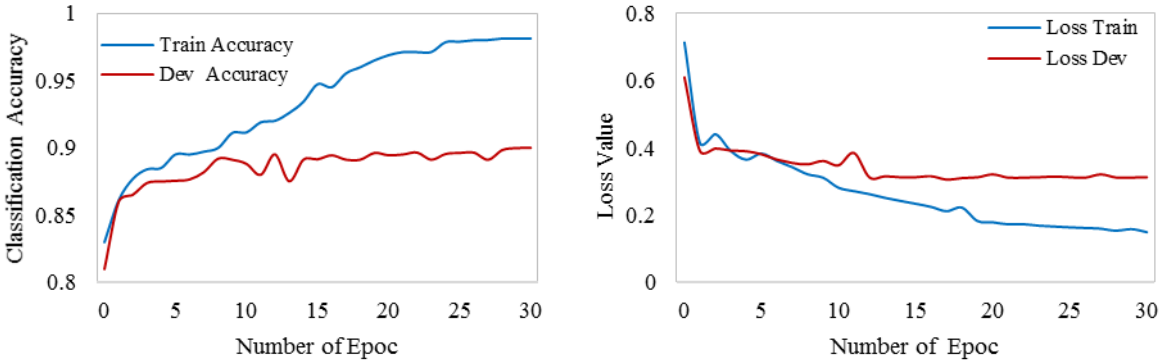
- واضح است که روش‌های ترکیبی از دقت بالاتری نسبت به روش‌های سنتی یادگیری ژرف برخوردار هستند که این مسأله به‌علت این است که روش‌های ترکیبی می توانند از ویژگی‌های ناهم‌گنی که از شبکه‌های عصبی ژرف با ساختار متفاوت به‌دست می‌آیند، به‌طور هم‌زمان استفاده کنند. درواقع روش‌های ترکیبی بر نقاط ضعف روش‌های تکی غلبه کرده و از نقاط قوت آن‌ها به‌صورت هم‌زمان بهره می‌برند که این مسأله باعث شده روش های ترکیبی از دقت بالاتری برخوردار باشند.
- با مقایسه روش ترکیبی پیشنهادی با سایر مدل‌های شبکه‌های عصبی پیچشی و هم‌گشتی می‌توان گفت که روش پیشنهادی از دقت بالاتری نسبت به کلیه روش های موجود برخوردار است. با توجه به استفاده از لایه پیچشی، روش پیشنهادی برای ساخت بازنمایی معنایی از متن و به‌دست‌آوردن اطلاعات متنی بسیار مناسب است. هم‌چنین استفاده از لایه هم‌گشتی به‌جای لایه ادغام باعث می‌شود که بتوان اطلاعات متنی را با استفاده از ترکیب‌های مختلف معنایی در درخت متنی بدون ساختار نشان داد که این مسأله باعث کاهش ازدست‌رفتن داده‌های محلی و حفظ وابستگی‌های طولانی‌مدت می‌شود. از طرف دیگر، روش پیشنهادی بسیار فشرده بوده و به تعداد کمتری پارامتر برای آموزش نیاز دارد.
- گفتنی است که تعداد کم پارامترها باعث می‌شود که این روش کمتر تحت تأثیر بیش برازش قرار بگیرد و روشی پایدار در برابر تغییرات باشد، هم‌چنین، در مقایسه با مدل هم‌گشتی روش پیشنهادی به زمان کمتری برای ساخت درخت بازنمایی جملات نیاز دارد و سریع‌تر نیز آموزش می‌بینند. علت این مسأله را می‌توان منوط به استفاده از ویژگی‌های مستخرج از لایه پیچشی به‌عنوان ورودی شبکه عصبی هم‌گشتی دانست که منجر به کاهش قابل‌توجه تعداد ویژگی‌های ورودی و استخراج ویژگی‌های منحصربه‌فرد می‌شود. نتایج حاصل از آزمایش‌ها نیز عملکرد روش پیشنهادی را تأیید می‌کند.
- هدف از طراحی این روش ترکیبی استفاده از مزایایی هرکدام از شبکه‌های عصبی پیچشی و هم‌گشتی و غلبه بر مشکلات آن‌ها و افزایش قدرت مدل دسته‌بندی است. از طرف دیگر، بر اساس دو دلیل که در ادامه بیان خواهند شد می‌توان ادعا کرد که روش ترکیبی پیشنهادی روش پایداری است. دلیل نخست مرتبط با استفاده از ساختار درختی شبکه عصبی هم‌گشتی به‌جای لایه ادغام است که می‌تواند با درنظرگرفتن روابط نحوی و معنایی نمایش سطح واژگان را نیز استخراج کند. دلیل دوم مرتبط با استفاده از روش از قلم‌انداختن برای کاهش بیش‌برازش و یادگیری ویژگی های قدرتمند است که در آموزش روش پیشنهادی استفاده شده است. بر اساس نتایج به‌دست‌آمده می‌توان گفت که استفاده از روش ازقلم‌انداختن می‌تواند به‌طور هم‌زمان باعث افزایش عملکرد و کاهش خطای دسته‌بندی شود
- در میان تمام پیاده‌سازی‌های مختلف روش پیشنهادی، روش CNN-Attention-RNN-Plus-Rand که از بردارهای از پیش آموزش‌دیده بهره نمی‌برد، دارای دقت پایین‌تری نسبت به سایر روش‌ها روی هر دو مجموعه‌داده است. درحالی‌که سایر پیاده‌سازی‌ها که از بردارهای از پیش آموزش‌دیده بهره می‌برند، دارای دقت به‌مراتب بالاتری هستند. درواقع، می‌توان ادعا کرد که استفاده از لایه نشانش برای آموزش بردارهای واژگان می‌تواند باعث افزایش دقت دسته‌بندی شود. پس از روش CNN-Attention-RNN-Plus-Rand، روش CNN-Attention-RNN-Plus -Static دارای کمترین دقت روی هر دو مجموعه‌داده است. بر این اساس می‌توان نتیجه گرفت که به‌روزرسانی بردار وزن واژگان در طول فرآیند آموزش بدون توجه به اینکه بردارها از پیش آموزش دیده‌اند یا خیر می‌تواند منجر به افزایش دقت شود. در میان انواع مختلف پیاده‌سازی‌های روش پیشنهادی، روش CNN- Attention-RNN-Plus-2channel دارای بالاترین دقت روی هر دو مجموعه‌داده SST1 و SST2 است که به‌ترتیب برابر ۵۳/۹۲٪ و ۹۲/۸۹٪ است.
- به‌منظور تأیید این ادعا که روش پیشنهادی می‌تواند در مقایسه با شبکه عصبی پیچشی و هم‌گشتی سریع‌تر آموزش داده شود، درحالی‌که به تعداد کمتری پارامتر نیاز دارد، دقت بهترین پیاده‌سازی انجام‌شده برای روش پیشنهادی (CNN-Attention-RNN-Plus-2channel) روی بخش آموزش و اعتبارسنجی مجموعه‌داده SST1 با استفاده از پارامترهای بهینه بر اساس تعداد ایپک در شکل (۴) نشان داده شده است. همان‌طور که مشخص است، پس از پانزده ایپک آموزش این روش به بالاترین دقت روی مجموعه ارزیابی دست می‌یابد درنتیجه می‌توان گفت که این روش قابلیت این را دارد که بتواند بسیار سریع‌تر از سایر روش‌ها آموزش ببیند.

[DOI: 10.52547/jsdp.19.1.19]

(جدول-۵): مقایسه روش پیشنهادی با سایر روش‌های موجود

(Table-5): Comparison of the proposed model with other existing models

گروه		روش	مجموعه داده			
			SST 1		SST2	
A	روش‌های بر پایه یادگیری ماشین	NB[20]	41		81.8	
		BiNB [20]	41.9		83.1	
		SVM [31]	40.7		79.4	
		WordVec-AVE [31]	32.7		80.1	
B	روش‌های بر پایه شبکه عصبی پیچشی	CNN-1 layer [32]	37.4		77.1	
		CNN-non static [33]	48		87.2	
		CNN-multichannel [33]	47.4		88.1	
		DCNN [32]	48.5		86.8	
		MVCNN [34]	49.6		89.4	
C	روش‌های بر پایه شبکه عصبی برگشتی	LSTM [35]	46.2		85.2	
		Bi-LSTM [35]	49.1		87.5	
		Tree-LSTM [35]	51		88	
		Tree-GRU [36]	50.5		88.6	
		Tree-GRU+attention [36]	51		89	
		LSTM+RNN+attention [37]	48		81.6	
D	روش‌های بر پایه شبکه عصبی هم گشتی	RecRNN [20]	43.2		82.4	
		RNTN [20]	45.7		85.4	
		MVRNN [19]	44.4		82.9	
E	روش‌های ترکیبی	CNN-GRU [24]	50.68		89.95	
		CNN-LSTM [24]	51.5		89.56	
		ConvLstm [22]	47.5		88.3	
		Recursive-recurrent [23]	45		85.1	
		RCNN [31]	45.8		85.8	
		2channel CNN-Tree-LSTM [25]	52.46		89.7	
		CNN-LSTM (Glove Amazon) [25]	50.84		90.39	
		CNN-RNN-2channel[1]	52.84*		91.88*	
روش پیشنهادی			Mean(std)	max	Mean(std)	max
		CNN-Attention-RNN-Plus-Rand	51.25 (0.93)	51.95	89.96 (0.65)	91.05
		CNN-Attention-RNN-Plus -Static	51.72 (0.68)	51.85	90.01(0.75)	90.3
		CNN-Attention-RNN-Plus -Static	52.65 (0.73)	52.93	91.38(0.45)	91.98
		CNN-Attention-RNN-Plus-2channel	**52.98 (0.67)**	**53.92**	**92.11(0.48)**	**92.89**



(شکل-۴): ارزیابی مدل CNN-Attention-RNN-Plus-2channel. شکل سمت چپ دقت دسته‌بندی را برای نمونه‌های آموزشی و ارزیابی در اپیاک‌های مختلف شکل سمت راست میزان خطای کراس آنتروپی را برای نمونه‌های آموزشی و ارزیابی در اپیاک‌های مختلف نشان می‌دهد.

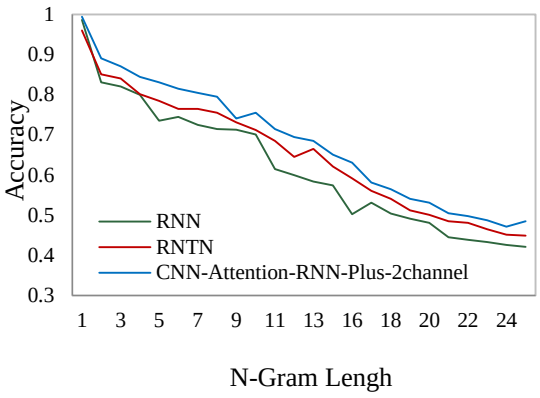
(Figure-4): Evaluation of CNN-Attention-RNN-Plu-2channel model. Classification accuracy for training and validation sets based on number of epochs is presented in left figure. The Loss value for training and validation sets based on number of epochs is presented in right figure.

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۳۹۸/۵/۲۴ • تاریخ پذیرش: ۱۳۹۹/۱۰/۲۱ • تاریخ انتشار: ۱۴۰۱/۰۳/۳۱ • نوع مطالعه: پژوهشی

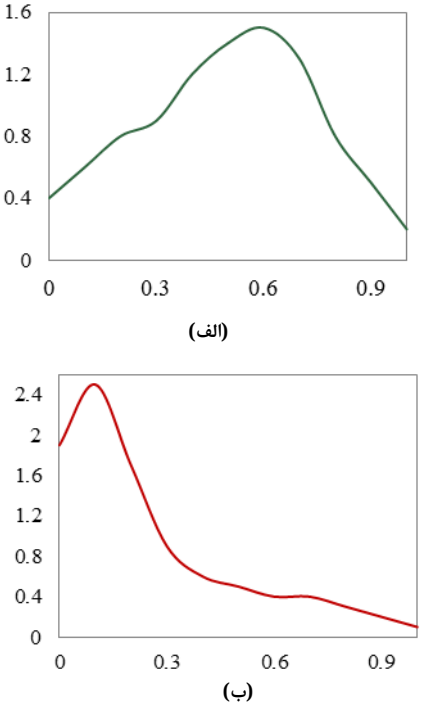
[DOI: 10.52547/jsdp.19.1.19]



(شکل-۵): مقایسه دقت مدل‌ها بر اساس اندازه طول n-گرام

(Figure-5): Comparison of the accuracy of various models based on n-gram length

- به‌منظور ارائه جزئیات کامل از عملکرد روش پیشنهادی، مقایسه دقت روش‌های مختلف بر اساس طول n-گرام در شکل (۵) نشان داده شده است. همان‌طور که مشخص است، روش پیشنهادی از دقت بالاتری روی مجموعه‌های کوچک برخوردار است و با افزایش طول n-گرام عملکرد آن کاهش می‌یابد. گفتنی است که روش پیشنهادی دارای عملکرد مشابهی با سایر روش‌ها برای طول n-گرام‌ها با طول بیشتر است. البته گفتنی است که روش پیشنهادی نسبت به سایر روش‌های موجود از دقت بالاتری در مواجه با جملات طولانی‌تر برخودار است. درواقع روش پیشنهادی که از ساختار درختی بهره می‌برد، برای دسته‌بندی جملات نیز کل جمله در در قالب یک بردار مدل‌سازی می‌کند و درنهایت بردار نهایی که بر اساس کلمه جمله ورودی است وارد لایه دسته‌بندی می‌شود.
- به‌منظور نشان‌دادن عملکرد سازوکار توجه و مشخص‌سازی عملکرد روش پیشنهادی در تشخیص اهمیت واژگان با توجه به متن، توزیع وزن توجه دو واژه good و bad روی یک تکه از بخش آزمایش مجموعه‌داده SST1 در شکل (۶) نشان داده شده است. بر اساس این توزیع، نمره توجه اختصاص داده شده در بازه صفر تا یک قرار دارد. این توزیع نشان‌دهنده قدرت روش پیشنهادی در مواجه با متون مختلف و اختصاص وزن با درنظرگرفتن متن است؛ زیرا همان‌طور که مشخص است، با افزایش نمره وزن واژه good نیز افزایش می‌یابد اما در مورد واژه bad به‌طورکامل برعکس است و با افزایش نمره، میزان وزن آن کاهش می‌یابد. این مسأله بدین معنی است که واژه good دارای تأثیر بیشتری در جملات مثبت و واژه bad دارای تأثیر بیشتری در جملات منفی است.



(شکل-۶): توزیع وزن توجه مرتبط با دو واژه good (الف) و bad (ب) بر اساس نمرات

(Figure-6): Attention weight distributions of words good (a) and bad (b)

گفتنی است که عملکرد روش‌های یادگیری ژرف تحت تأثیر پارامترهای مختلفی مانند اندازه مجموعه‌داده، انتخاب ویژگی‌ها و ناپدیدشدن و انفجار گرادیان قرار دارد و هزینه محاسباتی بالا و نیاز به تعداد زیاد داده‌های آموزشی همچنان از چالش‌های آن‌ها به‌شمار می‌آید. از طرف دیگر، علاوه‌بر این‌که تعداد زیاد داده‌های آموزشی باعث دشوارشدن تنظیم پارامترهای روش‌های یادگیری ژرف می‌شود، تنظیم پارامترهای آن‌ها به توجه به این‌که تئوری ریاضی ضعیفی دارند، باید به‌صورت آزمون‌وخطا انجام شود. درنتیجه آموزش بهترین دسته‌بندی بر پایه یادگیری ژرف همچنان یکی از مهم‌ترین موضوعات پژوهشی به‌شمار می‌آید و نمی‌توان تصور کرد که یک روش مشخص با پارامترهای تنظیم‌شده برای تمامی مجموعه‌داده‌ها مناسب باشد.

۵- نتیجه‌گیری و کارهای آینده

تجزیه و تحلیل احساسات زمینه‌ای است که در آن نظرات، تمایلات، نگرش، احساسات و ارزیابی افراد نسبت

[DOI: 10.52547/jsdp.19.1.19]

به یک مفهوم مشخص مانند یک محصول، رخداد، خدمت و یا نهاد موردبررسی قرار می‌گیرد و هدف اصلی آن تعیین جهت‌گیری نظرات افراد نسبت به آن مفهوم است به‌طوری‌که مشخص شود که نظر آن‌ها مثبت، منفی و یا حتی خنثی است. با اینکه در طول زمان مطالعات مختلفی در این حوزه صورت گرفته است و پژوهش‌گران به نتایج قابل‌توجهی دست یافته‌اند، اما این مفهوم با توجه به کاربرد گسترده‌اش در حوزه‌های مختلف هنوز یکی از مهم‌ترین موضوعات پژوهشی در حوزه پردازش زبان طبیعی به‌شمار می‌آید و نیاز به پیشرفت در این حوزه همچنان وجود دارد. از طرف دیگر، در سال‌های اخیر روش‌ها یادگیری ژرف به پیشرفت قابل‌توجهی در کاربردهای مختلف پردازش زبان طبیعی به‌خصوص تجزیه و تحلیل احساسات دست یافته‌اند. گفتنی است که با اینکه پژوهش‌های متعددی در راستای استفاده از روش‌های یادگیری ژرف برای تجزیه و تحلیل احساسات صورت گرفته است، اما این روش‌ها همچنان با چالش‌های فراوانی در این حوزه مواجه هستند و به‌نوعی در ابتدای مسیر پیشرفت قرار دارند. در همین راستا، در این مقاله یک روش ترکیبی حاصل از ترکیب شبکه عصبی پیچش و هم‌گشتی برای تجزیه و تحلیل احساسات موجود در متن معرفی شده است که بتواند ضمن غلبه بر مشکلات و چالش‌های الگوریتم‌های موجود از نقاط قوت آن‌ها به بهترین نحو استفاده کند. در همین راستا در روش پیشنهادی، برای ایجاد تأکید روی واژگانی که نقش مؤثری روی معنی واژگان دارند از سازوکار توجه روی شبکه عصبی پیچشی و برای کاهش ازدست‌رفتن داده‌های محلی توسط لایه ادغام از شبکه عصبی هم گشتی به‌عنوان جایگزین لایه ادغام استفاده شده است. از طرف دیگر، به‌منظور افزایش دقت درخت تجزیه جملات در شبکه عصبی هم‌گشتی، از یک ویژگی میانی حاصل از ترکیب دو گره فرزند استفاده شده است که می‌تواند با درنظرگرفتن تأثیر واژگان روی یکدیگر باعث افزایش دقت دسته‌بندی شود. درواقع در روش پیشنهادی، ویژگی‌های استخراج‌شده از شبکه عصبی پیچشی بر پایه بر سازوکار توجه به‌جای اینکه به لایه ادغام داده شوند، به‌طورمستقیم وارد شبکه عصبی هم‌گشتی تعمیم‌یافته می‌شوند. بر اساس نتایج آزمایش‌ها، روش پیشنهادی به‌ترتیب با دقت ۵۳/۹۲ و ۹۲/۸۹ درصد روی مجموعه‌داده‌های SST1 و SST2 نه‌تنها دارای دقت بالاتری نسبت به روش‌های پایه و سایر روش‌های ترکیبی است، بلکه بسیار سریع‌تر از آن‌ها نیز آموزش می‌یابد.

با توجه به این‌که مفاهیم مرتبط با تجزیه و تحلیل احساسات و یادگیری ژرف از مهم‌ترین موضوعات پژوهشی در حوزه پردازش زبان طبیعی و هوش مصنوعی به‌حساب می‌آیند که نتایج حاصل از آن‌ها می‌تواند تأثیر قابل‌توجهی در دنیای واقعی داشته باشد، توسعه و ادامه فرآیند پیشرفت آن‌ها از اهمیت فراوانی برخوردار است. در ادامه این پژوهش نیز می‌توان از روش‌های انطباق دامنه[1] و انتقال یادگیری[2] برای افزایش حجم داده‌های آموزشی و پس از آن آموزش روش به‌کمک مجموعه‌داده توسعه‌یافته استفاده کرد. هم‌چنین می‌توان کاربرد روش پیشنهادی را در سایر حوزه‌ها و زبان‌های دیگر، به‌ویژه زبان فارسی سنجید. استفاده از سایر روش‌های بازنمایی ویژگی‌ها برای تولید ورودی مناسب روش پیشنهادی نیز از کارهایی است که می‌توان در ادامه انجام داد.

۶-مراجع 6- References

[1] H. Sadr, M. M. Pedram, and M. Teshnehlab, "A Robust Sentiment Analysis Method Based on Sequential Combination of Convolutional and Recursive Neural Networks," *Neural Processing Letters*, pp. 1-17, 2019.

[2] H. Sadr, M. M. Pedram, and M. Teshnelab, "Improving the Performance of Text Sentiment Analysis using Deep Convolutional Neural Network Integrated with Hierarchical Attention Layer," *International Journal of Information and Communication Technology Research*, vol. 11, no. 3, pp. 57-67, 2019.

[3] Mohades Deilami, Fatemeh, Hossein Sadr, and Morteza Tarkhan. "Contextualized Multidimensional Personality Recognition using Combination of Deep Neural Network and Ensemble Learning." *Neural Processing Letters* 2022: 1-18.

[4] V. Vyas and V. Uma, "Approaches to sentiment analysis on product reviews," in *Sentiment Analysis and Knowledge Discovery in Contemporary Business*: IGI Global, 2019, pp. 15-30.

[5] Kalashami, Mahsa Pourhosein, Mir Mohsen Pedram, and Hossein Sadr. "EEG Feature Extraction and Data Augmentation in Emotion Recognition." *Computational Intelligence and Neuroscience* 2022.

[6] S. M. H. Chowdhury, S. Abujar, M. Saifuzzaman, P. Ghosh, and S. A. Hossain,

[1] Domain Adaptation
[2] Transfer learning

[DOI: 10.52547/jsdp.19.1.19]

[16] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," *Comput. Inf. Technol. Ubiquitous Comput. Commun. Dependable, Auton. Secur. Comput. Pervasive Intell. Comput. (CIT/IUCC/DASC/PICOM)*, 2015 IEEE Int. Conf., pp. 23592364, 2015.

[17] R. Yin, P. Li, and B. Wang, "Sentiment Lexical-Augmented Convolutional Neural Networks for Sentiment Analysis," *IEEE Second International Conference on Data Science in Cyberspace*, 2017.

[18] R. Socher, Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions," Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics., 2011.

[19] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng, "Semantic Compositionality through Recursive Matrix-Vector Spaces," Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics., 2012.

[20] R. Socher, A. Perelygin, and Wu, "Recursive deep models for semantic compositionality over a sentiment treebank," Proceedings of the conference on empirical methods in natural language processing (EMNLP), 2013.

[21] Q. Huang, X. Zheng, R. Chen, and Z. Dong, "Deep Sentiment Representation Based on CNN and LSTM " International Conference on Green Informatics, 2017.

[22] A. Hassan and A. Mahmood, "Deep Learning approach for sentiment analysis of short texts," in *Control, Automation and Robotics (ICCAR), 2017 3rd International Conference on*, 2.17 IEEE, pp. 705-710 .

[23] A. Timmaraju and V. Khanna, "Sentiment Analysis on Movie Reviews using Recursive and Recurrent Neural Network Architectures," 2017.

[24] X. Wang, W. Jiang, and Z. Luo, "Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts," 2016.

[25] V. D. Van, Œ. Thai, and M.-Q. o. Nghiem, "Combining Convolution and Recursive Neural Networks for Sentiment Analysis," 2018.

[26] S. M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings," *Expert Systems with Applications*, vol. 117, pp. 139-147, 2019.

"Sentiment Prediction Based on Lexical Analysis Using Deep Learning," in *Emerging Technologies in Data Mining and Information Security*: Springer, 2019, pp. 441-449.

[7] Soleymanpour, Shiva, Hossein Sadr, and Mojdeh Nazari Soleimandarabi. "CSCNN: cost-sensitive convolutional neural network for encrypted traffic classification." *Neural Processing Letters* , pp.3497-3523, 2021.

[8] Sadr, Hossein, and Mojdeh Nazari Soleimandarabi. "ACNN-TL: attention-based convolutional neural network coupling with transfer learning and contextualized word representation for enhancing the performance of sentiment classification." *The Journal of Supercomputing* 2022, pp. 1-27, 2022.

[9] H. Sadr, M. Nazari, M. M. Pedram, and M. Teshnehlab, "Exploring the Efficiency of Topic-Based Models in Computing Semantic Relatedness of Geographic Terms," International *Journal of Web Research*, vol. 2, no. 2, pp. 23-35, 2019.

[10] H. Sadr, M. M. Pedram, and M. Teshnehlab, "Multi-View Deep Network: A Deep Model Based on Learning Features From Heterogeneous Neural Networks for Sentiment Analysis," *IEEE Access*, vol. 8, pp. 86984-86997, 2020.

[11] H. Sadr, M. N. Soleimandarabi, M. Pedram, and M. Teshnelab, "Unified Topic-Based Semantic Models: A Study in Computing the Semantic Relatedness of Geographic Terms," in *2019 5th International Conference on Web Research (ICWR)*, 2019: IEEE, pp. 134-140 .

[12] V. D. Van, T. Thai, and M.-Q. Nghiem, "Combining convolution and recursive neural networks for sentiment analysis," in *Proceedings of the Eighth International Symposium on Information and Communication Technology*, 2017: ACM, pp. 151-158 .

[13] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment Analysis Based on Deep Learning: A Comparative Study," *Electronics*, vol. 9, no. 3, pp. 483, 2020.

[14] H. Sadr and M. Nazari Solimandarabi, "Presentation of an efficient automatic short answer grading model based on combination of pseudo relevance feedback and semantic relatedness measures," *Journal of Advances in Computer Research*, vol. 10, no. 2, pp. 1-10, 2019.

[15] J. Islam and Y. Zhang., "Visual Sentiment Analysis for Social Images Using Transfer Learning Approach," *2016 IEEE Int. Conf. Big Data Cloud Comput. (BDCloud), Soc. Comput. Netw. (SocialCom), Sustain. Comput. Commun.*, pp. 124130, 2016.

[DOI: 10.52547/jsdp.19.1.19]

[27] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality, Nips," 2013.

[28] O. Irsoy and C. Cardie, "Deep recursive neural networks for compositionality in language," in *Advances in neural information processing systems*, 2014, pp. 2096-2104 .

[29] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," *arXiv preprint arXiv:1510.03820*, 2015.

[30] Y. LeCun, Y. Bengio, and G. Hinton, " Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May, 2015.

[31] C. DU and L. HUANG, "Sentiment Classification Via Recurrent Convolutional Neural Networks," *DEStech Transactions on Computer Science and Engineering*, no. cii, 2017.

[32] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[33] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[34] W. Yin and H. Schütze, "Multichannel variable-size convolution for sentence classification," *arXiv preprint arXiv:1603.04513*, 2016.

[35] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *arXiv preprint arXiv:1503.00075*, 2015.

[36] F. Kokkinos and A. Potamianos, "Structural attention neural networks for improved sentiment analysis," *arXiv preprint arXiv:1701.01811*, 2017.

[37] Y. Wang, M. Huang, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, pp. 606-615.

[38] Sadr, Hossein, Mir M. Pedram, and Mohammad Teshnehlab. "Convolutional neural network equipped with attention mechanism and transfer learning for enhancing performance of sentiment analysis." Journal of AI and Data Mining 9.2, 2021 : 141-151.

حسین صدر مدرک کارشناسی خود را در رشته مهندسی رایانه در گرایش مهندسی نرم‌افزار دریافت کرد. وی مدرک کارشناسی ارشد خود را در همان رشته با احراز رتبه نخست از دانشگاه آزاد اسلامی واحد علوم و تحقیقات و مدرک دکترای تخصصی خود را در رشته مهندسی رایانه در گرایش سامانه‌های نرم‌افزاری از دانشگاه آزاد اسلامی، دریافت نمود. وی در حال حاضر پژوهش‌گر دوره پسادکتری مدل‌سازی شناختی است. نامبرده چهار دوره عضو باشگاه پژوهش‌گران و نخبگان بوده و عضو هیأت علمی گروه مهندسی رایانه و فناوری اطلاعات در موسسه آموزش عالی راهبرد شمال رشت است؛ همچنین وی عضو انجمن‌های سامانه‌های هوشمند ایران، سامانه‌های فازی ایران و مؤسسه مهندسان برق و الکترونیک امریکا (IEEE) و همچنین مدیر هسته فناور مستقر در پارک علم و فناوری گیلان است. زمینه‌های پژوهشی نامبرده شامل پردازش زبان‌های طبیعی، علوم داده، بازیابی اطلاعات، یادگیری ژرف و علوم شناختی است.

نشانی رایانامه ایشان عبارت است از:

Sadr@Rahbordshomal.ac.ir

میرمحسن پدرام مدرک کارشناسی خود را از دانشگاه صنعتی اصفهان و مدرک کارشناسی ارشد و دکترای خود را در رشته مهندسی برق از دانشگاه تربیت مدرس تهران دریافت کرد. ایشان در حال حاضر با مرتبه علمی دانشیار عضو هیأت علمی دانشگاه خوارزمی است. وی مدیرگروه مهندسی برق و کامپیوتر و هم‌چنین سرپرست آزمایشگاه داده‌کاوی و متن‌کاوی در این دانشگاه است. ایشان عضو هیأت مدیره انجمن سامانه‌های هوشمند نیز هستند. زمینه‌های پژوهشی وی شامل داده‌کاوی، متن‌کاوی، یادگیری ماشین و مدل‌سازی شناختی است.

نشانی رایانامه ایشان عبارت است از:

Pedram@Khu.ac.ir

محمد تشنه‌لب مدرک کارشناسی خود را از دانشگاه استونی‌بروک آمریکا و مدرک کارشناسی ارشد و دکترای خود را در رشته مهندسی برق از دانشگاه ساگا ژاپن دریافت کرد. وی در

[DOI: 10.52547/jsdp.19.1.19]

حال حاضر با مرتبه علمی استاد عضو هیأت علمی دانشگاه صنعتی خواجه نصیرالدین طوسی است. ایشان عضو هیأت مدیره انجمن سامانه‌های هوشمند و انجمن سامانه‌های‌های فازی ایران هستند. زمینه‌های پژوهشی وی شامل هوش مصنوعی و سامانه‌های خبره، محاسبات نرم، شبکه‌های عصبی و پردازش تکاملی است.

نشانی رایانامه ایشان عبارت است از:

Teshnehlab@Kntu.ac.ir

[DOI: 10.52547/jsdp.19.1.19]