Sadra Med. Sci. J. 2025; 13(1): 155-168.

Sadra Medical Sciences Journal

https://smsj.sums.ac.ir/

Original Article

Using FastText-BERT to Extract Semantic Relations and Improve Sentiment Analysis of Persian Healthcare Service Reviews

Faezeh Forootan[1], PhD Candidate; Raouf Khayami[2*], PhD; Pirooz Shamsinejad[3], PhD

[1]PhD Candidate, Department of Computer Engineering and Information Technology, Shiraz University of Technology, Shiraz, Iran
[2]Associate Professor, Department of Computer Engineering, and Information Technology Shiraz University of Technology, Shiraz, Iran
[3]Assistant Professor, Department of Computer Engineering, and Information Technology Shiraz University of Technology, Shiraz, Iran

Article Information

Article History:
Received: Dec. 18, 2023
Accepted: Feb. 22, 2024

***Corresponding Author:**
Raouf Khayami, PhD;
Associate Professor, Department of Computer Engineering, and Information Technology Shiraz University of Technology, Shiraz, Iran
Email: khayami@sutech.ac.ir

Abstract

Introduction: The analysis of patients' opinions is considered a valuable indicator for assessing the quality of healthcare services. The increasing volume of textual reviews about healthcare has made these reviews a critical factor in other patients' decision-making processes when selecting medical services. Consequently, researchers aimed to extract valuable insights, classify sentiments, and identify patient needs and behavioral patterns through sentiment analysis, thereby developing appropriate strategies to enhance patient satisfaction. However, patient reviews often contain a significant amount of specialized terminology, and existing sentiment analysis tools are typically trained on general-domain data. Therefore, to analyze these reviews accurately, it is essential to employ models and their combinations in a way that ensures reliable and valid results.

Methods: To improve the efficiency and accuracy of sentiment analysis for Persian healthcare reviews, this study utilized the FastText-BERT hybrid embedding model for semantic relation extraction and the CNN-BiLSTM model for sentence-level sentiment classification.

Results: The proposed framework achieved an accuracy of 86% and an F1-score of 84.99%.

Conclusion: The results demonstrated that combining embedding models leverages the strengths of both approaches, enabling the identification of specialized and out-of-domain expressions and the extraction of semantic relationships between them. This combination significantly enhances the efficiency and accuracy of sentiment analysis.

Keywords: Sentiment Analysis; Delivery of Health Care; Symbiosis; Semantic; Persian; Clinical Coding

Please cite this article as:
Forootan F, Khayami R, Shamsinejad P. Using FastText-BERT to Extract Semantic Relations and Improve Sentiment Analysis of Persian Healthcare Service Reviews. Sadra Med. Sci. J. 2025; 13(1): 155-168. doi: 10.30476/smsj.2025.100979.1469.

دوره ۱۳، شماره ۱، زمستان ۱۴۰۳، صفحات ۱۵۵-۱۶۸

مجله علوم پزشکی صدرا

https://smsj.sums.ac.ir/

مقاله پژوهشی

استفاده از FastText-BERT برای استخراج روابط معنایی و بهبود نتایج تحلیل احساسات نظرات فارسی خدمات مراقبت‌های بهداشتی

فائزه فروتن[۱]، سید رئوف خیامی[۲*]، پیروز شمسی‌نژاد[۳]

[۱]دانشجوی دکتری مهندسی فناوری اطلاعات، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز، ایران
[۲]دانشیار، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز، ایران
[۳]استادیار، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز، ایران

اطلاعات مقاله

تاریخچه مقاله:
تاریخ دریافت: ۱۴۰۲/۰۹/۲۷
تاریخ پذیرش: ۱۴۰۲/۱۲/۰۳

***نویسنده مسئول:**
سید رئوف خیامی،
دانشیار، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز، شیراز، ایران
پست الکترونیکی: khayami@sutech.ac.ir

چکیده

مقدمه: امروزه تحلیل نظرات بیماران به‌عنوان یک شاخص ارزشمند، برای سنجش کیفیت خدمات مراقبت‌های بهداشتی محسوب می‌گردد. افزایش نظرات متنی پیرامون مراقبت‌های بهداشتی، منجر گردیده تا این نظرات نقش مهمی در زمینهٔ تصمیم‌گیری سایر بیماران برای انتخاب خدمات پزشکی و درمانی داشته باشند. بر همین اساس پژوهشگران در تلاش‌اند تا ضمن استخراج اطلاعات ارزشمند و طبقه‌بندی احساسات، بتوانند نیازها و الگوی رفتاری بیماران را شناسایی و استراتژی‌های مناسبی را برای بهبود سطح رضایت آن‌ها در نظر بگیرند. اما نظرات بیماران شامل حجم بالایی از عبارات تخصصی است و ابزارهای پردازشی موجود در حوزه تحلیل احساسات بر اساس دامنه‌های عمومی آموزش دیده‌اند. بنابراین برای تحلیل دقیق این نظرات باید از مدل‌ها و ترکیب آن‌ها به نحوی استفاده گردد که عملکرد نهایی به دریافت نتایج معتبر بیانجامد.

مواد و روش‌ها: در این تحقیق با هدف بهبود کارایی و افزایش دقت تحلیل احساسات در حیطه نظرات فارسی مراقبت‌های بهداشتی، از مدل ترکیبی تعبیه سازی FastText-BERT برای استخراج روابط معنایی و از CNN-BiLSTM برای طبقه‌بندی احساسات در سطح جمله استفاده شد.

یافته‌ها: نتایج نهایی دقت ۸۶٪ و F1-score=۸۴/۹۹٪ را برای چارچوب پیشنهادی نشان داد.

نتیجه‌گیری: بر اساس نتایج می‌توان اذعان داشت، ترکیب مدل‌های تعبیه سازی به واسطه بهره‌مندی از نقاط قوت فراوان (در هر دو روش) در شناسایی عبارات تخصصی و خارج از دامنه، و همچنین استخراج روابط معنایی میان آن‌ها، باعث بهبود کارایی و افزایش دقت تحلیل احساسات می‌گردند.

کلمات کلیدی: تحلیل احساسات؛ ارائه مراقبت‌های بهداشتی؛ همزیستی؛ معنایی؛ فارسی؛ کدگذاری بالینی

لطفاً این مقاله را به این صورت استناد کنید:
فروتن ف، خیامی س ر، شمسی‌نژاد پ. استفاده از FastText-BERT برای استخراج روابط معنایی و بهبود نتایج تحلیل احساسات نظرات فارسی خدمات مراقبت‌های بهداشتی. مجله علوم پزشکی صدرا. دوره ۱۳، شماره ۱، زمستان ۱۴۰۳، صفحات ۱۵۵-۱۶۸.

مقدمه

۶۴/۵ درصد از کل جمعیت جهان در حال استفاده از اینترنت هستند و روزانه به‌طور میانگین ۶ ساعت و ۴۰ دقیقه از زمان خود را به‌صورت آنلاین سپری می‌کنند (۱). اما نکتهٔ حائز اهمیت، افزایش روند استفاده کاربران از اینترنت برای پیدا کردن خدمات مورد نیاز خود است. مراقبت‌های بهداشتی از جمله این خدمات هستند که ۷ درصد از کل جستجوهای روزانه کاربران اینترنت را به خود اختصاص داده‌اند. به‌گونه‌ای که امروزه، ۸۲/۸ درصد از بیماران از موتورهای جستجو برای یافتن پزشک و مراکز ارائه‌دهنده خدمات مراقبت‌های بهداشتی استفاده می‌کنند (۲). فضای اینترنت به بیماران اجازه داده تا ضمن تعامل با یکدیگر بتوانند به مقایسه خدمات ارائه شده از سوی پزشکان و مراکز بهداشتی و درمانی بپردازند و نظرات خود را ثبت کنند. بنابراین افزایش روزانه حجم نظرات منتشرشده در این زمینه باعث شده تا، ۷۴ درصد از بیماران پیش از آغاز درمان، نظرات را مطالعه و بر اساس آن‌ها پزشک یا بیمارستانی را برای درمان خود انتخاب نمایند (۲)؛ اما مسئله آن است که هر یک از این نظرات می‌توانند انتقال دهنده حس مثبت، منفی یا خنثی باشند، و همین می‌تواند باعث جذب یا دفع بیماران به سمت استفاده از خدمات مراقبت‌های بهداشتی موردنظر گردد. به‌گونه‌ای که ۶۹,۹ درصد از بیماران اذعان داشته‌اند، مطالعه نظرات مثبت نقش حیاتی را در انتخاب پزشک یا بیمارستان دارد (۲). در مقابل، نظرات منفی باعث شده‌اند تا آن‌ها از مراجعه به پزشک یا بیمارستان موردنظر صرف نظر کنند (۲). بنابراین مطالب، تحلیل نظرات مراقبت‌های بهداشتی بسیار حیاتی و ارزشمند است، زیرا نتایج حاصل از آن به مدیران کمک می‌کند تا الگوی رفتاری بیماران و همچنین نیازهای فعلی و آینده آن‌ها را شناسایی کنند و بتوانند بر اساس آن‌ها استراتژی‌های مناسبی برای بهبود سطح رضایت بیماران تعیین نمایند (۳).

تحلیل احساسات[1] روشی جدید در زمینهٔ پردازش نظرات است که به تخمین، تعیین و طبقه‌بندی قطبیت[2] احساسات افراد در رابطه با یک موضوع خاص می‌پردازد (۳). اما وظیفه اصلی تحلیل احساسات، تعیین قطبیت نظرات بر اساس کلمات موجود در متن و ارتباط معنایی[3] میان آن‌ها است (۳). بر همین اساس تا امروز روش‌ها و مدل‌های زبانی[4] مختلفی تحت عنوان تعبیه کلمات[5] و به‌منظور تشخیص و استخراج روابط معنایی[6] میان کلمات و درنهایت تبدیل این ارتباطات به بردارهای عددی، ارائه شده‌اند. این در شرایطی است که هر یک از این روش‌ها به‌ویژه FastText و BERT[7] بر اساس زمینه‌های عمومی مانند داده‌های ویکی‌پدیا و ... از پیش‌آموزش[8] داده شده‌اند (۴). ولی مسئله مهم آن است که نظرات مراقبت‌های بهداشتی شامل کلمات و عبارات تخصصی مانند عنوانِ بیماری‌ها، اصطلاحات پزشکی و ... هستند که معمولاً در دامنه و زمینه‌های عمومی موجود قرار ندارند (۴-۶). بنابراین برای تحلیل دقیق این دسته از نظرات، لازم است تا از مدل‌هایی برای شناسایی و استخراج روابط معنایی میان کلمات استفاده گردد که بر اساس داده‌های پزشکی از پیش‌آموزش داده شده‌اند. این در حالی است که امروزه فقدان چنین مدلی در تعبیه‌سازی، به‌عنوان یکی از مهم‌ترین چالش‌ها در حوزه تحلیل احساسات در نظرات مراقبت‌های بهداشتی به شمار می‌رود (۴). در این تحقیق با هدف غلبه بر این چالش و همچنین با هدف بهبود کارایی در مرحله استخراج روابط معنایی و افزایش دقت تحلیل احساساتِ نظرات فارسی، به ارائه یک چارچوب جدید بر اساسِ ترکیب مدل‌های تعبیه‌سازی پرداخته شده است. در طی این چارچوب پس از استخراج، جمع‌آوری و پیش‌پردازش نظرات فارسی مراقبت‌های بهداشتی، از FastText-BERT برای استخراج روابط معنایی و تبدیل نظرات به بردارهای عددی استفاده می‌گردد. سپس مدل CNN-BiLSTM[9] برای طبقه‌بندی احساسات در سطح جمله و در سه گروه مثبت، منفی و خنثی، روی مجموعه داده اعمال می‌شود. بنابراین مهم‌ترین نوآوری‌های این چارچوب عبارت‌اند از:

- استخراج و جمع‌آوری متون نظرات فارسی منتشرشده در زمینهٔ مراقبت‌های بهداشتی.
- ترکیب مدل‌های تعبیه‌سازی FastText و BERT با هدف بهبود کارایی در مرحله استخراج روابط معنایی.
- به‌کارگیری مدل ترکیبی CNN-BiLSTM برای طبقه‌بندی نظرات فارسی پیرامون خدمات مراقبت‌های بهداشتی.

تا امروز در هیچ‌یک از تحقیقات انجام شده در زمینهٔ تحلیل احساسات نظرات فارسی، از ساختاری همچون چارچوب پیشنهادی در این تحقیق، و به‌ویژه در تحلیل

1. Sentiment Analysis
2. Polarity
3. Semantic Relationship
4. Language Model
5. Word Embedding
6. Semantic Relation Extraction
7. Bidirectional Encoder Representations from Transformers (BERT)
8. Pre-trained
9. Convolutional Neural Network - Bidirectional Long-Short Term Memory

نظرات مراقبت‌های بهداشتی، استفاده نشده است.

تحقیقات انجام‌شده در حوزه تحلیل احساسات نظرات مراقبت‌های بهداشتی در زبان‌های خارجی

سابقه تحلیل احساسات نظرات مراقبت‌های بهداشتی در زبان‌های خارجی به‌ویژه در زبان انگلیسی به سال ۲۰۱۳ و تحقیق صورت گرفته توسط گریوز[10] و همکاران باز می‌گردد (۷). بااین‌حال در این بخش، به بررسی جدیدترین و معتبرترین تحقیقاتی پرداخته شده که از روش‌های مختلفی برای استخراج روابط معنایی و تعبیه‌سازی نظرات استفاده کرده‌اند.

جیمنز[11] و همکاران در ۲۰۱۹، بعد از پیش‌پردازش مجموعه داده‌های COPOS[12] و DOS[13]، از Word-2Vec، TF-IDF[14]، TF[15] و BTO[16] برای استخراج روابط معنایی و تعبیه‌سازی، و بعد برای طبقه‌بندی احساسات در سطح سند از الگوریتم ماشین بردار پشتیبان[17] استفاده کرده‌اند (۸). نتایج نشان می‌دهد، ماشین بردار پشتیبان بر اساس Word2Vec در شرایطی که روابط معنایی در نظر گرفته شده است، روی مجموعه‌داده DOS به دقت ۶۵/۳ درصد و بر اساس BTO روی مجموعه داده COPOS به دقت ۸۹/۲ درصد دست یافته است.

گارگ[18] در ۲۰۲۱، ۲۱۵۰۶۳ نظر موجود در سایت Drugs.com را به‌عنوان مجموعه داده در نظر گرفت (۹). سپس روش‌های TF-IDF، کیسه کلمات[19]، word۲vec را به‌صورت جداگانه و برای استخراج روابط معنایی و تعبیه‌سازی به کار برد و در ادامه برای طبقه‌بندی احساسات در سطح سند از رگرسیون لوجستیک[20]، نیوبیز[21]، گرادیان کاهشی تصادفی[22]، ماشین بردار پشتیبان خطی استفاده نمود. نتایج نشان می‌دهد، اگرچه TF-IDF در هنگام تبدیل متون به بردارهای عددی فقط به روابط نحوی توجه می‌کند و روابط معنایی را در نظر نمی‌گیرد، ولی ماشین بردار پشتیبان خطی بر اساس این روش توانسته است به دقت ۹۳ درصد دست یابد و نسبت به سایر روش‌ها عملکرد بهتری از خود نشان دهد.

10. Greaves et al.
11. Jiménez et al.
12. Corpus of Patient Opinions in Spanish (COPOS)
13. Drug Opinions Spanish (DOS)
14. Term Frequency - Inverse Document Frequency (TF-IDF)
15. Term Frequency (TF)
16. Binary Term Occurrences (BTO(
17. Support Vector Machines
18. Garg
19. Bag Of Words
20. Logistic Regression
21. Naïve Bayes
22. Stochastic Gradient Descent

همچنین سرانو-گوئررو[23] و همکاران در سال ۲۰۲۲، به بررسی ۲۱۲۹۴۰ نظر موجود در سایت Careopinion پرداختند (۱۰). آن‌ها هر یک از روش‌های GloVe، FastText و PubMed+PM را برای تعبیه‌سازی و سپس مدل BiGRU-MCCNN[24] را برای طبقه‌بندی احساسات به کار بردند. در ادامه برای مقایسه دقت و کارایی مدل پیشنهادی مدل‌های ماشین بردار پشتیبان، رگرسیون لوجستیک، شبکه عصبی پیچشی[25]، شبکه عصبی بازگشتی[26]، الگوریتم شبکه عصبی حافظه طولانی_کوتاه‌مدت[27]، شبکه عصبی خود بازگشتی دوطرفه[28]، شبکه عصبی واحد بازگشتی دروازه‌دار[29]، CNN-BiLSTM-CNN[30]، MCBiGRU[32]، BiLSTM[31] و MNB[33] را بر اساس TF-IDF پیاده‌سازی کرده‌اند. نتایج نشان می‌دهد BiGRU-MCCNN[34] بر اساس PubMed+PM (به‌دلیل در نظر گرفتن روابط معنایی میان کلمات)، به دقت ۹۵/۸۲ درصد دست یافته است. همچنین بر اساس مقایسه صورت گرفته مشخص گردید، ماشین بردار پشتیبان از میان الگوریتم‌های یادگیری ماشین و MCBiGRU از میان الگوریتم‌های یادگیری عمیق بر اساس TF-IDF به ترتیب به دقت‌های ۸۹/۴ درصد و ۹۳/۹ درصد دست یافته‌اند و به نسبت دیگر مدل‌ها عملکرد بهتری دارند.

تحقیقات انجام‌شده در حوزه تحلیل احساسات نظرات مراقبت‌های بهداشتی در زبان فارسی

با وجود اهمیت تحلیل احساسات نظرات پیرامون مراقبت‌های بهداشتی، تنها تعداد انگشت‌شماری از تحقیقات فارسی موجود در حوزه تحلیل احساسات روی این دسته از نظرات تمرکز کرده‌اند که عبارت‌اند از:

- بوکائی‌نژاد و دیهیمی در ۲۰۱۹، ۸۰۳۲۷۸ توییت فارسی مرتبط با واکسن‌های کرونا را جمع‌آوری و بعد از پیش‌پردازش، به‌صورت دستی و بر اساس سه گروه مثبت، منفی و خنثی برچسب‌گذاری کرده‌اند

23. Serrano-Guerrero et al.
24. Bidirectional Gated recurrent units
25 Convolutional Neural Network
26. Recurrent Neural Network
27. Long short-term memory (LSTM)
28. Bidirectional Long-Short Term Memory
29. Gated recurrent units (GRUs)
30. Bidirectional Long-Short Term Memory_Convolutional Neural Network
31. Convolutional Neural Network_Bidirectional Long-Short Term Memory
32. Multi-Channel Bidirectional Gated Recurrent Unit (MCBiGRU)
33. Multinomial Naive Bayes
34. Bidirectional Gated Recurrent Unit_Multi-model Cascaded Convolutional Neural Network

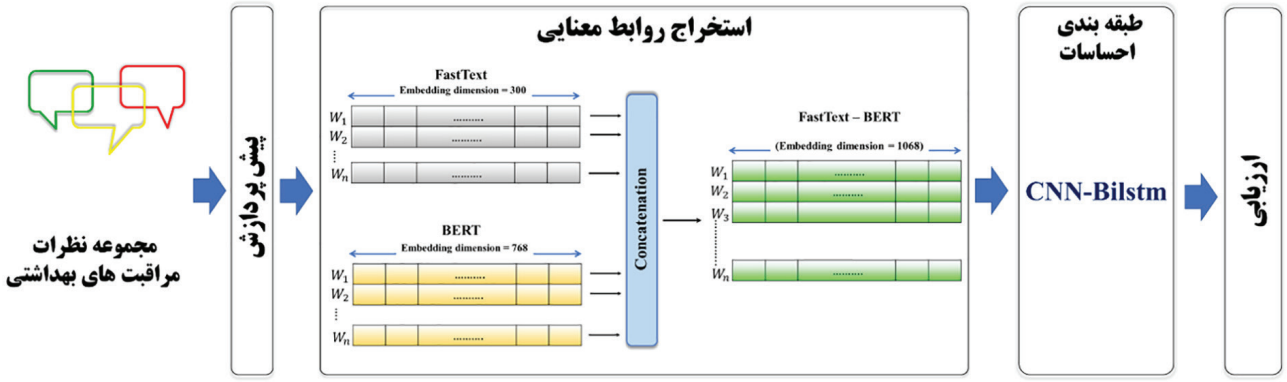
(۱۱). سپس از Word۲vec برای تعبیه‌سازی، از شبکه عصبی پیچشی برای استخراج ویژگی‌ها و از شبکه عصبی حافظه طولانی کوتاه‌مدت برای طبقه‌بندی احساسات استفاده کرده‌اند. نتایج دقت ۸۵ درصد را برای روش پیشنهادی نشان می‌دهد.

- تقی‌زاده و همکاران[35] در ۲۰۲۱، بر این باورند که برای دست‌یابی به نتایج دقیق در تحلیل احساسات فارسی، لازم است تا از ابزارهای مختص به این زبان استفاده گردد (۱۲). بر همین اساس بعد از جمع‌آوری ۲ میلیون نظر فارسی خدمات مراقبت‌های بهداشتی، به ارائه مدل تعبیه‌سازی SINA-BERT[36] پرداخته‌اند. بر اساس کار آن‌ها، SINA-BERT یک مدل بازنمایی زبان از پیش آموزش‌دیده شده بر اساس متون نظرات پزشکی و مبتنی بر BERT است که استفاده از آن برای تحلیل احساسات نظرات فارسی خدمات مراقبت‌های بهداشتی به نسبت ParsBERT[37] (از دیگر مدل‌های زبانی مختص زبان فارسی) باعث افزایش دقت تحلیل احساسات می‌گردد.

تحقیقات انجام‌شده در حوزه تحلیل احساسات بر اساس ترکیب روش‌های استخراج روابط معنایی و تعبیه سازی

- میتی و همکاران[38] در ۲۰۲۱، برای تشخیص و طبقه‌بندی نظرات توهین‌آمیز منتشرشده به زبان انگلیسی و هندی، از مدل ترکیبی BERT-FastText برای استخراج روابط معنایی و از شبکه عصبی واحد بازگشتی دروازه‌دار مبتنی بر توجه[39]، برای طبقه‌بندی استفاده کرده‌اند (۱۳). همچنین در ادامه هر یک از مدل‌های FastText-BERT-GRU، BERT-LSTM، GRU، FastText-LSTM، همچنین BERT-FastText را بر اساس شبکه عصبی حافظه طولانی کوتاه‌مدت، برای مقایسه با روش پیشنهادی خود، روی مجموعه داده موجود پیاده‌سازی کرده‌اند. نتایج نشان می‌دهد روش پیشنهادی با دقت ۶۹/۱۷ درصد و معیار F برابر با ۴۸/۶۸ درصد، روی زبان هندی و همچنین دقت ۷۶/۳۲ درصد و معیار F برابر با ۷۸/۷۵ درصد، روی زبان انگلیسی به نسبت سایر مدل‌ها عملکرد بهتری دارد.
- بدری و همکاران[40] در ۲۰۲۲، برای شناسایی و تشخیص توییت‌های توهین‌آمیز، از FastText+Glove برای استخراج روابط معنایی و از BiGRU برای طبقه‌بندی استفاده کرده‌اند (۱۴). سپس برای مقایسه، هر یک از مدل‌های شبکه عصبی بازگشتی دروازه‌دار دو طرفه[41] بر اساس Glove و شبکه عصبی بازگشتی دروازه‌دار دو طرفه بر اساس FastText را به‌صورت جداگانه روی مجموعه داده اعمال کرده‌اند. نتایج نشان می‌دهد روش پیشنهادی با معیار F برابر با ۷۹ درصد به نسبت هر یک از مدل‌های BiGRU-Glove و BiGRU-FastText و همچنین روش RoBERTa[42] عملکرد بهتری دارد. ازاین‌رو آن‌ها در پایان کار خود اذعان داشته‌اند که ترکیب مدل‌های تعبیه‌سازی باعث می‌گردد تا مدل عملکرد بهتری برای شناسایی عبارات خاص (عبارات توهین‌آمیز) و استخراج روابط معنایی میان آن‌ها داشته باشد. همچنین ترکیب این روش‌ها باعث بهبود و افزایش دقت مدل طبقه‌بندی نیز می‌گردد.
- آلوتایبی و گوپتا[43] در ۲۰۲۲ بر این باورند که اگرچه استفاده از روش‌های تعبیه‌سازی در تحلیل احساسات باعث افزایش دقت مدل می‌گردد، اما مهم‌ترین مشکل آن‌ها در این است که برای آموزش به مجموعه داده‌های بزرگ نیاز دارند (۱۵). به همین دلیل عملکرد این مدل‌ها روی مجموعه داده‌های کوچک ضعیف است. ازاین‌رو آن‌ها برای بهبود و افزایش دقت تحلیل احساسات روی مجموعه داده‌های کوچک، از ترکیب روش برچسب‌گذاری بخشی از گفتار[44]، FastText و Wordposition-2vec برای تعبیه‌سازی و از CRNN+BiLSTM[45] برای طبقه‌بندی استفاده کرده‌اند. نتایج نشان می‌دهد، روش پیشنهادی بر نظرات IMDB به دقت ۹۲/۲۱ درصد دست یافته است. این بدان معنی است که ترکیب روش‌های تعبیه‌سازی ضمن بهبود عملکرد مدل در مرحله استخراج روابط معنایی روی مجموعه داده‌های کوچک، باعث افزایش دقت مدل تحلیل احساسات نیز می‌گردد.
- دیدی و همکاران[46] در ۲۰۲۲ چارچوب جدیدی را برای تحلیل احساسات نظرات کرونا، پیشنهاد داده‌اند (۱۶). بدین‌صورت که از ترکیب TF-IDF برای استخراج ویژگی‌های نحوی و از FastText و Glove برای استخراج ویژگی‌های معنایی استفاده کرده‌اند. سپس الگوریتم‌های XGBoost، جنگل تصادفی[47]،

35. Taghizadeh et al.
36. SINA- Bidirectional encoder representations from transformers (BERT)
37. Monolingual BERT for the Persian language (ParsBERT)
38. Maity et al.
39. Attention-based
40. Badri et al.
41. Bidirectional Gated recurrent units
42. Robustly Optimized BERT Pre-training
43. Alotaibi and Gupta
44. Part of Speech (POS)
45. Convolutional Recurrent Neural Network (CRNN)
46. Didi
47. Random Forest



شکل ۱. چارچوب پیشنهادی تحقیق

آدابوست[48]، درخت تصمیم[49]، رگرسیون لوجستیک، نیوبیز و شبکه عصبی پیچشی را برای طبقه‌بندی احساسات در سطح جمله به کار برده‌اند. نتایج نشان می‌دهد SVM بر اساس روش TF-IDF+FastText به دقت ۸۸/۷۲ درصد و بر اساس TF-IDF+Glove به دقت ۸۶/۱۶ درصد دست یافته است و نسبت به سایر الگوریتم‌ها عملکرد بهتری دارد.

مواد و روش‌ها

باهدف بهبود کارایی در مرحله استخراج روابط معنایی و افزایش دقت تحلیل احساسات نظرات فارسیِ خدمات مراقبت‌های بهداشتی، در این تحقیق بر اساس تکنیک‌های پردازش متن، چارچوب جدیدی پیشنهاد شده است (شکل ۱).

جامعه آماری

۱/۶ درصد از وبسایت‌های موجود در فضای اینترنت بر اساس محتوای فارسی ایجاد شده و به همین دلیل روزانه حجم بالایی از نظرات فارسی (درزمینهٔ موضوعات مختلف) در اینترنت و انواع پلتفرم‌های آنلاین در حال ثبت و انتشار هستند (۱۷)، اما آنچه حائز اهمیت است، افزایش روزافزون تعداد نظرات متنی درزمینهٔ مراقبت‌های بهداشتی است. بااین‌حال اگرچه نظرات فارسی زیادی در این زمینه وجود دارند ولی تا امروز هیچ‌گونه مجموعه داده بزرگ و معتبری در رابطه با این حوزه در دسترس نیست. بنابراین در ابتدای این تحقیق مجموعه داده‌ای بر اساس نظرات مراقبت‌های بهداشتی و بر اساس سایت‌های مرتبط با این موضوع، استخراج و جمع‌آوری گردید.

پیش‌پردازش

زبان فارسی یک زبان چالش‌برانگیز در حوزه پردازش متن است. بر همین اساس پیش‌پردازش نظرات فارسی یکی از مهم‌ترین مراحل تحلیل احساسات است. زیرا کیفیت و ساختار صحیح عبارات و کلمات در زبان فارسی که به واسطه اجرای دقیق مراحل پیش‌پردازش حاصل می‌گردد، ازجمله مواردی است که باعث بهبود و افزایش دقت مدل طبقه‌بندی و تحلیل احساسات فارسی می‌شود (۱۸).

استخراج روابط معنایی

کلمات و عبارات بسته به نقش دستوری و زمینه‌ای که در آن قرار دارند، می‌توانند معانی متفاوتی داشته باشند (۱۹). بااین‌وجود طبق نظریه معناشناسی توزیعی[50] (یا نظریه توزیعی[51])، زمانی که این کلمات درزمینه‌های مشابه به کار می‌روند با سایر کلمات و عبارات موجود در متن، شباهت و ارتباط نحوی و معنایی دارند (۲۰). تعبیه کلمات روشی است که متون را بر اساس شباهت‌های معنایی، نحوی و ارتباط موجود بین کلمات به بردارهای عددی تبدیل می‌نماید. در این تحقیق به‌منظور بهبود عملکرد در مرحله استخراج روابط معنایی و با هدف غلبه بر چالش مطرح در حوزه تحلیل احساسات نظرات مراقبت‌های بهداشتی، یعنی «فقدان مدل تعبیه‌سازی برای تبدیل متون نظرات پزشکی به بردارهای عددی و استخراج روابط معنایی و نحوی میان آن‌ها» از ترکیب FastText و BERT استفاده گردید. زیرا FastText بر اساس اطلاعات داخلی زیر-کلمات[52]، می‌تواند ضمن ایجاد نمایش برداری برای کلمات عمومی، برای کلمات نادر و کلمات خارج از دامنه[53] نیز، بردارهای عددی ایجاد نماید (۴). اما مسئله آن است که FastText نیز مانند دیگر روش‌های تعبیه‌سازی، تک جهته

48. AdaBoost
49. Decision Tree
50. Distributional Semantics
51. Distributional Hypothesis
52. Sub-words information
53. Out-of-Vocabulary (OOV)

است و بردار عددی هر توکن را بر اساس زمینهٔ چپ یا راست تولید می‌کند (۴). بنابراین در ادامه و برای رفع این محدودیت از BERT استفاده شد. زیرا BERT بر اساس مکانیزم توجه[54] و رمزگذار دو جهته مبتنی بر ترانسفورماتور می‌تواند اطلاعات قبل و بعد کلمه (یعنی زمینه چپ و راست کلمه) را با هدف استخراج دقیق اطلاعات معنایی بین کلمات در نظر بگیرد و سپس بردار عددی مربوطه را ایجاد نماید(۲۰، ۲۱).

طبقه‌بندی احساسات

برای طبقه‌بندی احساسات، از رویکرد یادگیری عمیق و مدل ترکیبی CNN-BiLSTM استفاده می‌شود. زیرا ترکیب مدل‌های تعبیه‌سازی می‌توانند منجر به افزایش ابعاد بردارهای عددی و درنتیجه افزایش ابعاد مسئله شوند. به همین دلیل در ابتدای مرحلهٔ طبقه‌بندی احساسات، از CNN برای کاهش ابعاد مسئله و استخراج ویژگی‌های محلی[55] بین کلمات متوالی در یک جمله، استفاده شد، اما ازآنجاکه تمرکز این پژوهش روی تحلیل احساسات نظرات در سطح جمله بود، در نظر گرفتن توالی کلمات و ارتباط میان آن‌ها نقش مهمی در نتیجه نهایی مدل تجزیه و تحلیل احساسات ایفا کرد. این در شرایطی است که CNN نمی‌تواند توالی و وابستگی‌های خیلی دور بین کلمات و جملات را یاد بگیرد و به همین دلیل قادر به در نظر گرفتن ارتباط معنایی میان آن‌ها نیست. بنابراین برای غلبه بر این مشکل، در ادامه از Bi-LSTM استفاده گردید. زیرا Bi-LSTM توالی و ترتیب زمانی و همچنین وابستگی‌های بین کلمات موجود در یک متن را به شکل بهتری حفظ می‌کند (۲۳، ۲۴). همچنین Bi-LSTM قادر است تا هم‌زمان با پردازش داده‌ها در دو جهت، برای تحلیل دقیق‌تر کلمات، زمینهٔ متن[56] را نیز در نظر بگیرد (۲۳، ۲۴).

ملاحظات اخلاقی

تمامی اصول اخلاقی در مراحل مختلف این مقاله در نظر گرفته شده است.

پیاده‌سازی مراحل چارچوب پیشنهادی

در ادامه و در بخش یافته‌ها، به پیاده‌سازی مراحل چارچوب پیشنهادی برای این تحقیق بر روی نظرات استخراج شده از سایت نوبت دات آی آر، در محیط گوگل کولب[57] و تحت زبان برنامه‌نویسی پایتون[58] و تنظیمات GPU پرداخته شده است. همچنین به نتایج به دست آمده برای هر یک از این مراحل اشاره شده است.

یافته‌ها

پیاده‌سازی و نتایج مرحله جامعه آماری

سایت نوبت دات‌آی‌آر[59] به کاربران و بیماران فارسی‌زبان اجازه می‌دهد تا هم‌زمان با رزرو نوبت برای دریافت خدمات درمانی، بتوانند نظرات دیگر بیماران را مطالعه و با آن‌ها تعامل داشته باشند. به همین دلیل، سایت نوبت دات‌آی‌آر، هم درزمینهٔ رزرو نوبت برای مراقبت‌های بهداشتی و هم درزمینهٔ استخراج نظرات فارسی مرتبط با مراقبت‌های بهداشتی، یک سایت مرجع قلمداد می شود. بر همین اساس در ابتدای این تحقیق ۵۰۶۰ متن نظر از این سایت استخراج و به‌عنوان مجموعه داده در نظر گرفته شد. در بررسی اولیه مجموعه داده مشخص گردید که طول نظرات متفاوت است. یعنی کوتاه‌ترین نظر مانند «تشخیصش عالیه» از ۲ کلمه یا توکن تشکیل شده است. درحالی‌که بلندترین نظر مانند «من افسردگی و خود کم بینی داشتم و از ترس و خجالتی بودن و پیامدهاش رنج می‌بردم و با کمک خانم دکتر تا حد بسیار زیادی به شناخت بهتری از خودم رسیدم و تونستم خودم رو تغییر بدم و زندگی جدید و متفاوتی را تجربه کنم و نمی‌دونم واقعاً چه طوری ازشون تشکر کنم فقط آرزو می‌کنم که سلامت باشند و زندگی پرمعنا و پربرکتی داشته باشند» از ۶۹ کلمه یا توکن تشکیل گردیده است. همچنین مشخص شد که به همراه هر نظر، رتبه‌بندی ستاره‌ای مرتبط با آن نیز استخراج شده است. بنابراین در این تحقیق، از رتبه‌بندی‌های استخراج شده برای مشخص کردن قطبیت نظرات استفاده شد. بدین‌صورت که به نظرات که دارای رتبه‌بندی ۴ و ۵ ستاره‌ای، قطبیت و برچسب مثبت، و به نظرات دارای رتبه‌بندی ۳ ستاره‌ای قطبیت و برچسب خنثی اختصاص داده شد. همچنین به نظراتی که امتیاز ۲، ۱ داشتند یا فاقد رتبه‌بندی ستاره‌ای بودند، امتیاز منفی نسبت داده شد. بنابراین مجموعه داده نامتوازن و شامل ۲۱۳۴ نظر مثبت، ۱۲۰۸ نظر خنثی و ۱۷۱۸ نظر منفی است.

54. Attention Mechanism
55. Local features
56. Text Contexts

57. https://colab.research.google.com/
58. Python
59. https://nobat.ir/

پیاده‌سازی مرحله پیش‌پردازش

برای پیش‌پردازش از کتابخانه پارسی‌ور[60] که مختص زبان فارسی می‌باشد، استفاده شده است. بدین صورت که:

با توجه به متن نظرات استخراج شده مشخص گردید برخی از کاربران به علت شباهت‌های زیادی که بین رسم‌الخط زبان فارسی و عربی وجود دارد، در نگارش برخی از حروف فارسی به‌صورت سهوی یا عمدی، از معادل آن‌ها در زبان عربی استفاده کرده‌اند. همچنین برخی دیگر، اعداد و عبارات را تحت فرمت «فارسی-انگلیسی» یا همان FEnglish ثبت نموده‌اند. مانند نظر «Awli bode,behtarin doctor hastn». بنابراین در ابتدا و برای یکسان‌سازی و نرمال‌سازی نظرات، از ماژول Normalizer استفاده گردید. سپس ماژول Tokenizer برای تفکیک نظرات به واحدهای دستوری جملات و کلمات، پیاده‌سازی گردید. در ادامه و برای حذف ایست‌واژه‌ها با استفاده از زبان برنامه نویسی پایتون و بر اساس لیست ایست‌واژه فارسی[61]، تابع «remove-stopwords» ایجاد و روی نظرات اعمال گردید. دلیل این امر این بود که کتابخانه‌های مرتبط با پیش‌پردازش زبان فارسی، فاقد ماژول مناسب برای اجرای این مرحله هستند. همچنین درنهایت برای ریشه‌یابی، ماژول FindStems روی نظرات پیاده‌سازی شد. زیرا ریشه‌یابی نیز مانند مرحله حذف ایست واژه‌ها، در تلاش است تا با تبدیل شکل‌های مختلف کلمات به ریشه اصلی آن‌ها، ضمن کاهش ابعاد مجموعه داده عملکرد مدل را نیز بهبود ببخشد.

پیاده‌سازی مرحله استخراج روابط معنایی

در این پژوهش ترکیب FastText و BERT، بر اساس روش Concatenation پیاده‌سازی می‌گردد به این ترتیب که:

پس از پیش‌پردازش، بردارهای عددی نظرات به‌صورت هم‌زمان و جداگانه توسط هر یک از مدل‌های FastText و BERT ایجاد می‌شوند، سپس خروجی حاصل از FastText به خروجی حاصل از BERT، الحاق و بردارهای تعبیه نهایی ایجاد می‌گردند؛ اما نکته مهم این است که اگر تعبیه‌های حاصل از FastText دارای ابعاد d۱ و تعبیه‌های حاصل از BERT دارای ابعاد d۲ باشند، بردارهای الحاقی نهایی، ابعاد d1+d2 خواهند داشت، بنابراین در این مرحله، مقدار Embedding dimension که نشان‌دهنده ابعاد و اندازه بردارهای عددی برای هر یک از روش‌های FastText و BERT است، مانند پژوهش و بر اساس مقدار متداول آن‌ها یعنی ۳۰۰ بُعد برای FastText و ۷۶۸ بُعد برای BERT در نظر گرفته شدند. بنابراین اندازه بردارهای الحاقی FastText-BERT در این پژوهش برابر با ۱۰۶۸ بودند.

پیاده‌سازی مرحله تحلیل احساسات

در این تحقیق برای تحلیل احساسات در سطح جمله از مدل یادگیری عمیق CNN-BiLSTM استفاده شده است. بدین نحو که پیش از ایجاد مدل ۷۰ درصد نظرات برای آموزش (۳۵۴۲ نظر)، و ۳۰ درصد برای آزمایش (۱۵۱۸ نظر) در نظر گرفته شد و در ادامه مراحل کار به شرح زیر روی مجموعه داده پیاده‌سازی گردید:

- در ابتدا بردارهای تعبیه الحاقی (خروجی FastText-BERT) به‌عنوان ورودی به CNN داده می‌شوند. سپس CNN لایه‌های Convolution، Max-pooling و Dropout را با هدف کاهش ابعاد و شناسایی ویژگی‌های معنایی، روی بردارهای تعبیه الحاقی اعمال می‌نماید. در ادامه خروجی CNN به‌عنوان ورودی به BiLSTM داده می‌شود تا توالی کلمات و همچنین ارتباط میان کلمات و عبارات در فرآیند طبقه‌بندی و تحلیل احساسات در نظر گرفته شوند.
- در این مرحله نیز تکنیک Grid Search برای تنظیم پارامترهای CNN-BiLSTM و با هدف بهبود و افزایش دقت تحلیل احساسات، روی مجموعه داده پیاده‌سازی شد (جدول ۱). درنهایت مدل طبقه‌بندی بر اساس مقادیر بهینه پارامترها ایجاد و روی مجموعه داده اعمال شد.
- در آخر همانطور که در (جدول ۲) اشاره شده، روش پیشنهادی توانسته نظرات مرتبط با خدمات مراقبت‌های بهداشتی را با دقت ۸۶ درصد و F1-score برابر با ۸۴/۹۹ درصد در سه گروه احساسات منفی، مثبت و خنثی طبقه‌بندی نماید.

بحث

در این بخش با هدف بررسی عملکرد چارچوب پیشنهادی در مرحله استخراج روابط معنایی، به‌مقایسه عملکرد مدل ترکیبی پیشنهادی FastText-BERT با هریک از مدل‌های TF-IDF، FastText و همچنین BERT پرداخته شده است. لازم به ذکر است که مبنای مقایسه معیارهای ارزیابی دقت و معیار F1- هستند.

60. Parsivar
61. https://github.com/kharazi/persianstopwords/blob/master/persian

تأثیر در نظر گرفتن روابط معنایی روی تحلیل احساسات

TF-IDF ازجمله روش‌های متداول در حوزه پردازش متن است که متون را تنها بر اساس میزان فراوانی کلمات، یعنی «تعداد تکرار هر کلمه به ازای تمامی کلماتی که در متن وجود دارند» به بردارهای عددی تبدیل می‌کند. اما مسئله آن است که برخلاف سایر روش‌های تعبیه‌سازی (مانند BERT و FastText و. ..)، TF-IDF در طی فرآیند تبدیل نظرات به بردارهای عددی، معنی یا بافت معنایی کلمات را در نظر نمی‌گیرد. این در شرایطی است که تحلیل احساسات نظرات مراقبت‌های بهداشتی مستلزم درک اصطلاحات پیچیده پزشکی، زمینهٔ متن و همچنین ارتباط معنایی میان کلمات و عبارات موجود در این حوزه است. با توجه به (جدول ۳) و برتری نتایج برای CNN-BiLSTM بر اساس FastText-BERT به نسبت TF-IDF، می‌توان ادعا کرد که در نظر گرفتن روابط معنایی میان کلمات و عبارات، می‌تواند باعث بهبود و افزایش دقت تحلیل احساسات شود.

اگرچه به‌کارگیری TF-IDF به نسبت FastText-BERT، باعث کاهش پیچیدگی و کاهش مدت‌زمان پردازش و اجرا می‌گردد؛ ولی استفاده از آن به‌دلیل عدم توجه به روابط معنایی میان کلمات، باعث کاهش دقت مدل تحلیل احساسات می‌شود. این در حالی است که به‌کارگیری FastText-BERT به واسطه در نظر گرفتن

جدول ۱. مقدار بهینه پارامترها برای مدل CNN-BiLSTM

پارامترها	مقادیر بهینه پارامترها	
Convolution1D	Filters=۳۸, kernel_size=۳ ‘activation=‘relu	پارامترهای ایجاد مدل
MaxPooling	pool_size=۲	
Dropout	Rate=۰/۱	
Convolution1D	Filters=۶۴, kernel_size=۳, ‘activation=‘relu	
MaxPooling	pool_size=۲	
Dropout	Rate=۰/۱	
Convolution1D	Filters=۳۲, kernel_size=۳, ‘activation=‘relu	
MaxPooling	pool_size=۲	
Dropout	Rate=۰/۱	
Bidirectional (LSTM)	Units=۲۵۰	
Optimizer	‘adam’	پارامترهای ضروری اجرای مدل
Loss Function	categorical_crossentropy	
Epochs	۱۰۰	
Batch size	۲۵۰	

تمامی پارامترهای اشاره شده در جدول فوق، پارامترهای ضروری الگوریتم برای ایجاد مدل طبقه‌بندی و همچنین پارامترهای ضروری برای اجرای مدل هستند.

جدول ۲. نتایج روش پیشنهادی

معیارهای ارزیابی	دقت[1]	صحت[2]	فراخوانی[3]	معیار-F1[4]
نتایج	۸۶	۸۴/۲۱	۸۵/۷۹	۸۴/۹۹

1. Accuracy
2. Precision
3. Recall
4. F1-score

جدول ۳. مقایسه نتایج TF-IDF و چارچوب پیشنهادی

	مدل طبقه‌بندی		معیارهای ارزیابی			
	مدل استخراج روابط معنایی	مدل طبقه‌بندی	دقت	صحت	فراخوانی	معیار-F1
روش پیشنهادی برای مقایسه	TF-IDF	CNN-BiLSTM	۶۱/۱۷	۶۰/۰۴	۶۱/۲۹	۶۰/۶۵
روش پیشنهادی تحقیق	FastText-BERT Embedding dimension=۱۰۶۸	CNN-BiLSTM	۸۶	۸۴/۲۱	۸۵/۷۹	۸۴/۹۹

تمامی اختصارات انگلیسی اشاره شده در جدول فوق، عنوان مدل‌های زبانی و همچنین الگوریتم‌های مربوطه هستند.

جدول ۴. مقایسه نتایج FastText و BERT و چارچوب پیشنهادی

	مدل طبقه‌بندی		معیارهای ارزیابی			
	مدل استخراج روابط معنایی	مدل طبقه‌بندی	دقت	صحت	فراخوانی	معیار-F1
روش پیشنهادی برای مقایسه	FastText Embedding dimension=۳۰۰	CNN-BiLSTM	۷۹/۰۳	۷۸/۰۶	۷۸/۵۱	۷۸/۲۸
	BERT Embedding dimension=۷۶۸	CNN-BiLSTM	۸۱	۸۰/۱۳	۸۰/۶۹	۸۰/۴۰
روش پیشنهادی پژوهش	FastText+ BERT Embedding dimension=۱۰۶۸	CNN-BiLSTM	۸۶	۸۴/۲۱	۸۵/۷۹	۸۴/۹۹

*: تمامی اختصارات انگلیسی اشاره شده در جدول فوق، عنوان مدل‌های زبانی و همچنین الگوریتم‌ها می‌باشند که در متن مقاله به آنها اشاره شده و پاورقی شده‌اند.

روابط معنایی میان کلمات، درک زمینه، ایجاد بردار برای کلمات خارج از دامنه و همچنین کلمات نادر موجود در نظرات مراقب‌های بهداشتی، باعث افزایش دقت و کارایی تحلیل احساسات می‌گردد.

تأثیر به‌کارگیری مدل‌های مختلف استخراج روابط معنایی روی دقت تحلیل احساسات

به‌منظور بررسی تأثیر به‌کارگیری مدل‌های مختلف استخراج روابط معنایی روی کارایی و دقت تحلیل احساسات، به پیاده‌سازی جداگانه هر یک از روش‌های تعبیه‌سازی FastText و BERT روی مجموعه داده، پرداخته شد. درنهایت و با توجه به (جدول ۴) مشخص گردید که CNN-BiLSTM بر اساس FastText و با تمرکز بر اطلاعات داخلی زیر-کلمات توانسته به دقت ۷۹/۰۳ درصد و F1-score=۷۸/۲۸٪ دست یابد. علاوه بر این CNN-BiLSTM بر اساس BERT توانسته به واسطه فهم زمینه[۶۲] و تشخیص ارتباط بین کلمات موجود در مجموعه داده، به دقت ۸۱ درصد و همچنین F1-score=۸۰/۴۰٪ دست پیدا کند. این شرایطی است که دقت و F1-score چارچوب پیشنهادی بر اساس FastText-Bert به ترتیب برابر با ۸۶ درصد و ۸۴/۹۹ درصد است. بنابراین برتری نتایج به‌دست‌آمده برای روش پیشنهادی نشان می‌دهد که ترکیب مدل‌های تعبیه‌سازی به نسبت به کار بردن هر یک از این مدل‌ها به‌صورت جداگانه، نه‌تنها باعث بهبود کارایی و افزایش دقت تحلیل احساسات می‌گردند؛ بلکه در چنین شرایطی مدل ترکیبی با بهره‌مندی از نقاط قوت هر دو روش، در تلاش است تا عملکرد بهتری را درزمینهٔ شناسایی عبارات خاص، عبارات خارج از دامنه و استخراج روابط معنایی میان آن‌ها داشته باشد.

تأثیر اندازه بردار روابط معنایی روی تحلیل احساسات

همانطور که پیش‌تر اشاره شد، اگر اندازه بردارهای FastText و BERT به ترتیب برابر با d1 و d2 باشند، اندازه بردارهای نهایی برابر با d1+d2 هستند. بنابراین

62. Context

جدول ۵. مقایسه نتایج حاصل از پیاده‌سازی روش پیشنهادی بر اساس ابعاد مختلف بردارهای روابط معنایی

	مدل طبقه‌بندی		معیارهای ارزیابی			
	مدل استخراج روابط معنایی	مدل طبقه‌بندی	دقت	صحت	فراخوانی	معیار-F1
روش پیشنهادی برای مقایسه	FastText-BERT (۷۶۸+۱۰۰=۸۶۸)	CNN-BiLSTM	۸۱/۳۷	۷۸/۶۱	۷۹/۵۶	۷۹/۰۸
	FastText-BERT (۷۶۸+۲۰۰=۹۶۸)	CNN-BiLSTM	۸۳/۱۷	۸۰/۹۶	۸۱/۶۳	۸۱/۲۹
روش پیشنهادی پژوهش	FastText-BERT (۷۶۸+۳۰۰=۱۰۶۸)	CNN-BiLSTM	۸۶	۸۵/۷۹	۸۴/۲۱	۸۴/۹۹

*: تمامی اختصارات انگلیسی اشاره شده در جدول فوق، عنوان مدل‌های زبانی و همچنین الگوریتم‌ها می‌باشند که در متن مقاله به آنها اشاره شده و پاورقی شده‌اند.

در طی فرآیند ترکیب روش‌های تعبیه‌سازی به‌ویژه ترکیب بر اساس Concatenation، باید مقدار پارامتر Embedding dimension برای هر یک از روش‌ها به نحوی در نظر گرفته شود که ابعاد بردارهای نهایی بسیار بزرگ یا کوچک نشوند؛ زیرا اگر ابعاد این بردارها خیلی بزرگ باشند، هم‌زمان با افزایش پیچیدگی مسئله، مدت‌زمان پردازش نیز افزایش می‌یابد. درصورتی‌که اگر ابعاد بردارها خیلی کوچک در نظر گرفته شوند، اطلاعات معنایی استخراج شده فاقد اعتبار هستند. بنابراین با هدف بررسی تأثیر اندازه بردارهای روابط معنایی روی کارایی و دقت تحلیل احساسات، مقادیر مختلفی برای Embedding dimension در هر یک از روش‌های تعبیه‌سازی در نظر گرفته شد (جدول ۵).

سپس مدل CNN-BiLSTM بر اساس این مقادیر روی مجموعه داده پیاده‌سازی گردید. لازم است به این مهم توجه گردد که مقادیر اشاره شده بر اساس مقدار معمول این پارامتر در هر یک از روش‌های FastText و BERT الهام گرفته شده‌اند (یعنی مقادیر ۱۰۰ الی ۳۰۰ برای FastText و مقدار ۷۶۸ برای BERT). با توجه به نتایج (جدول ۵)، CNN-BiLSTM بر اساس روش ترکیبی FastText-BERT با Embedding dimension برابر با ۸۶۸ به دقت ٪۸۱/۳۷ و F1-score=٪۷۹/۰۸ دست یافته است. درحالی‌که اگر مقدار Embedding dimension برای FastText-BERT برابر با ۹۶۸ در نظر گرفته شود، دقت ٪۸۳/۱۷ و F1-score=٪۸۱/۲۹ برای CNN-BiLSTM حاصل می‌گردد. این در شرایطی است که در روش پیشنهادی، مقدار Embedding dimension برابر با ۱۰۶۸ در نظر گرفته شده و همین موجب گردیده تا CNN-BiLSTM بتواند نظرات مراقبت‌های بهداشتی را با دقت ٪۸۶ و F1-score=٪۸۴/۹۹ طبقه‌بندی نماید. بنابراین بر اساس نتایج می‌توان این‌گونه استنباط کرد که اندازه بردارهای روابط معنایی ضمن اینکه روی مدت‌زمان پردازش و پیچیدگی مسئله تأثیر می‌گذارند؛ روی دقت تحلیل احساسات نیز اثرگذار هستند. به‌گونه‌ای که اگرچه طولانی و بزرگ بودن مقدار بردارهای روابط معنایی الحاقی حاصل از FastText-BERT در چارچوب پیشنهادی، باعث افزایش مدت‌زمان پردازش و پیچیدگی مسئله شده‌اند. ولی طولانی بودن اندازه این بردارها (که به واسطه ترکیب مدل‌های تعبیه‌سازی رخ داده است)، باعث شده‌اند تا بردارهای نهایی که برای هر کلمه ایجاد می‌شوند، از لحاظ برخورداری از اطلاعات معنایی و احساسی دقیق‌تر، کامل‌تر و پیشرفته‌تر باشند. همین امر باعث شده که مدل بتواند بر اساس این بردارهای دقیق و پیشرفته، تفاوت‌ها و شباهت‌های ظریف معنایی و احساسی موجود در نظرات را به شکل بهتری درک نماید و کارایی و دقت تحلیل احساسات را بهبود و افزایش دهد. این در شرایطی است که Embedding dimension برابر با ۸۶۸ برای FastText-BERT ضمن اینکه باعث کاهش مدت‌زمان پردازش و پیچیدگی مسئله شده است، دقت و کارایی تحلیل احساسات را نیز کاهش داده است. زیرا در زمانی که اندازه بردارهای الحاقی کوچک در نظر گرفته می‌شوند، مدل ناچار است تا در طی فرآیند تبدیل عبارات و روابط معنایی آن‌ها به بردارهای عددی، برخی از اطلاعات را نادیده بگیرد و حذف نماید.

جدول ۶. مقایسه نتایج روش پیشنهادی روی مجموعه داده نظرات فارسی

مجموعه داده	مدل استخراج روابط معنایی	مدل طبقه‌بندی	دقت	صحت	فراخوانی	معیار-F1
نظرات دیجی‌کالا	FastText+ BERT Embedding) (dimension=۱۰۶۸	CNN-BiLSTM	۹۳	۸۹/۵	۹۱/۲۵	۹۰/۳۵
نظرات‌مراقبت‌های بهداشتی	FastText+ BERT Embedding) (dimension=۱۰۶۸	CNN-BiLSTM	۸۶	۸۵/۷۹	۸۴/۲۱	۸۴/۹۹

تمامی اختصارات انگلیسی اشاره شده در جدول فوق، عنوان مدل‌های زبانی و همچنین الگوریتم‌ها می‌باشند که در متن مقاله به آنها اشاره شده و پاورقی شده‌اند.

مقایسه نتایج روش پیشنهادی روی مجموعه داده‌های مختلف در زبان فارسی

در این مرحله به مقایسه عملکرد چارچوب پیشنهادی روی نظرات دیجی‌کالا[۶۳] که از دیگر مجموعه داده‌های معتبر در زبان فارسی است، پرداخته شده است. این مجموعه داده نامتوازن و شامل ۷۵۰۰ متن نظر (۲۸۰۰ نظر منفی، ۳۲۰۰ نظر مثبت و ۱۵۰۰ نظر خنثی) منتشرشده درباره محصولات دیجیتال سایت دیجی‌کالا است. طول نظرات در این مجموعه داده، متفاوت است. کوتاه‌ترین نظر از ۲ کلمه تشکیل گردیده، مانند «تقریباً خوبه» و بلندترین متن نظر از ۲۴ کلمه تشکیل شده است مانند «چند تا گوشی رو مقایسه کردم. هم از نظر کیفیت هم قیمت آخر اینو خریدم چون با توجه به پولم فقط این اوکی بود».

در این مرحله پس از پیش‌پردازش، مجموعه داده بر اساس رویکرد Hold-out به دو دسته ۷۰ درصد برای آموزش و ۳۰ درصد برای آزمایش تقسیم گردید. سپس چارچوب پیشنهادی که شامل FastText-BERT برای تعبیه‌سازی و CNN-BiLSTM برای طبقه‌بندی احساسات است، روی نظرات پیاده‌سازی گردید.

نتایج اشاره شده در (جدول ۶) نشان می‌دهند، چارچوب پیشنهادی توانسته نظرات دیجی‌کالا را با دقت ٪۹۳ و F1-score=۹۰/۳۵٪، به سه گروه احساسات مثبت، خنثی و منفی طبقه‌بندی نماید. درحالی‌که دقت و F1-score حاصل از پیاده‌سازی روش پیشنهادی روی نظرات خدمات مراقبت‌های بهداشتی به ترتیب برابر با ۸۶ درصد و ۸۴/۹۹ درصد است. بنابراین، با توجه به برتری نتایج به‌دست‌آمده برای مجموعه داده دیجی‌کالا می‌توان این‌گونه استنباط کرد که دامنه لغات و زمینهٔ متون نظرات، ازجمله عوامل تأثیرگذار روی دقت و کارایی مدل تحلیل احساسات هستند.

63. https://www.kaggle.com/datasets/soheiltehranipour/digikala-comments-persian-sentiment-analysis

به‌گونه‌ای که اگر دامنه لغات و زمینهٔ نظرات شامل اصطلاحات و عبارات عمومی و همچنین فاقد تعداد بالایی از عبارات و اصطلاحات خارج از دامنه باشند (مانند نظرات دیجی‌کالا)، می‌توانند باعث افزایش دقت تحلیل احساسات گردند. زیرا در حال حاضر اکثر ابزارها و مدل‌های زبانی موجود در حوزه پردازش متن و زبان طبیعی، به‌ویژه مدل‌های تعبیه‌سازی بر اساس این دسته از لغات و زمینه‌های متون آموزش‌دیده‌اند. این در شرایطی است که اگر دامنه لغات محدود یا گسترده باشند و زمینهٔ متون نظرات تخصصی و شامل تعداد بالایی از عبارات و اصطلاحات خارج از دامنه باشند (مانند نظرات مراقبت‌های بهداشتی)، می‌توانند باعث کاهش کارایی و دقت مدل تحلیل احساسات گردند.

نتیجه‌گیری

در این تحقیق با هدف بهبود کارایی استخراج روابط معنایی و افزایش دقت تحلیل احساساتِ نظرات فارسیِ مراقبت‌های بهداشتی، به ارائه یک چارچوب جدید پرداخته شده است. در طی آن، پس از جمع‌آوری نظرات فارسی و پیش‌پردازش، به‌منظور بهبود عملکرد در مرحله استخراج روابط معنایی و تبدیل نظرات به بردارهای عددی از FastText-BERT استفاده می‌گردد. سپس CNN-BiLSTM روی نظرات و برای طبقه‌بندی احساسات در سطح جمله اعمال می‌گردد. درنهایت نتایج دقت ۸۶ درصد و F1-score برابر با ۸۴/۹۹ درصد را برای چارچوب پیشنهادی این تحقیق نشان می‌دهد، اما در ادامه و با توجه به برتری نتایج حاصل از چارچوب پیشنهادی به نسبت پیاده‌سازی هر یک از روش‌های FastText، TF-IDF و BERT مشخص گردید که اگرچه ترکیب مدل‌های تعبیه‌سازی باعث افزایش زمان پردازش می‌گردند ولی باعث می‌شوند تا

مدل بتواند عملکرد دقیق‌تری را درزمینهٔ شناسایی عبارات تخصصی و خارج از دامنه و استخراج روابط معنایی میان آن‌ها داشته باشد. همچنین نتایج حاصل از در نظر گرفتن مقادیر مختلف برای ابعاد بردارهای الحاقی مدل FastText-BERT نشان می‌دهد، اندازه بردارهای روابط معنایی ازجمله عوامل اثرگذار بر کارایی و دقت مدل تحلیل احساسات هستند. به‌گونه‌ای که اگرچه طولانی بودن این بردارها باعث افزایش مدت‌زمان پردازش و پیچیدگی مسئله شده‌اند؛ ولی موجب گردیده‌اند تا بردارهای نهایی ایجاد شده برای هر کلمه، از لحاظ برخورداری از اطلاعات معنایی و احساسی دقیق‌تر، کامل‌تر و پیشرفته‌تر باشند که همین باعث افزایش دقت تحلیل احساسات می‌گردد.

درنهایت برتری نتایج حاصل از پیاده‌سازی چارچوب پیشنهادی روی نظرات دیجی‌کالا به نسبت نظرات مراقبت‌های بهداشتی نشان می‌دهد؛ وجود عبارات تخصصی و خارج از دامنه، در شرایطی که ابزارها و مدل‌های زبانی موجود در حوزه تحلیل احساسات بر اساس دامنه‌های عمومی مانند ویکی‌پدیا آموزش‌دیده‌اند؛ از دیگر عوامل اثرگذار روی کارایی و دقت تحلیل احساسات هستند.

تعارض منافع

هیچ‌گونه تضاد منافعی وجود ندارد.

منابع

1. DataReportal [Internet]. Digital Around the World, Global Digital Insights. [Accessed Sep. 06, 2022]. Available from: https://datareportal.com/global-digital-overview
2. Digitalis [Internet]. Healthcare Marketing Statistics Digitalis Medical. [Accessed Sep. 06, 2022]. Available from: https://digitalismedical.com/blog/healthcare-marketing-statistics/
3. Abualigah L, Alfar HE, Shehab M, Hussein AMA. Sentiment analysis in healthcare: a brief review. Recent advances in NLP: the case of Arabic language. 2020:129-41 doi: 10.1007/978-3-030-34614-0_7.
4. Khattak FK, Jeblee S, Pou-Prom C, Abdalla M, Meaney C, Rudzicz F. A survey of word embeddings for clinical text. J Biomed Inform. 2019;100S:100057.
5. Parvinnia E, Mohammadi M, BANANZADEH A, Khayami SP. Analysis of data on patients with colon cancer using the data mining techniques Case study: Patients at colorectal research center of Shaheed Faghihi hospital in Shiraz. RAZI JOURNAL OF MEDICAL SCIENCES (JOURNAL OF IRAN UNIVERSITY OF MEDICAL SCIENCES),[online]. 2018;25(9):46-56
6. Boroumandzadeh M, Parvinnia E. Automated classification of BI-RADS in textual mammography reports. Turkish Journal of Electrical Engineering and Computer Sciences. 2021;29(2):632-47. doi; 10.3906/elk-2002-31.
7. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res. 2013;15(11):e239.
8. Jimenez-Zafra SM, Martin-Valdivia MT, Molina-Gonzalez MD, Urena-Lopez LA. How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain. Artif Intell Med. 2019;93:50-7.
9. Garg S, editor Drug recommendation system based on sentiment analysis of drug reviews using machine learning. 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence); 2021: IEEE. doi10.1109/Confluence51648.2021.9377188:
10. Serrano-Guerrero J, Bani-Doumi M, Romero FP, Olivas JA. Understanding what patients think about hospitals: A deep learning approach for detecting emotions in patient opinions. Artif Intell Med. 2022;128:102298.
11. Bokaee Nezhad Z, Deihimi MA. Twitter sentiment analysis from Iran about COVID 19 vaccine. Diabetes Metab Syndr. 2022;16(1):102367.
12. Taghizadeh N, Doostmohammadi E, Seifossadat E, Rabiee HR, Tahaei MS. SINA-BERT: a pre-trained language model for analysis of medical texts in Persian. arXiv preprint arXiv:210407613. 2021.
13. Maity K, Kumar A, Saha S, editors. Attention Based BERT-FastText Model for Hate Speech and Offensive Content Identification

in English and Hindi Languages. FIRE (Working Notes); 2021.
14. Badri N, Kboubi F, Chaibi AH. Combining fasttext and glove word embedding for offensive and hate speech text detection. Procedia Computer Science. 2022;207:769-78.doi: 10.1016/j.procs.2022.09.132.
15. Alotaibi FS, Gupta V. Sentiment analysis system using hybrid word embeddings with convolutional recurrent neural network. Int Arab J Inf Technol. 2022;19(3):330-5 doi: 10.34028/iajit/19/3/6.
16. Didi Y, Walha A, Wali A. COVID-19 tweets classification based on a hybrid word embedding method. Big Data and Cognitive Computing. 2022;6(2):58.doi; 10.3390/bdcc6020058.
17. Kinsta [Internet]. Usage Statistics and Market Share of Persian for Websites. [Accessed Oct. 13, 2023]. Available from: https://w3techs.com/technologies/details/cl-fa-
18. Asgarian E, Kahani M, Sharifi S. The impact of sentiment features on the sentiment polarity classification in Persian reviews. Cognitive Computation. 2018;10:117-35.doi: 10.1007/s12559-017-9513-1.
19. Jbene M, Tigani S, Saadane R, Chehri A. Deep Neural Network and Boosting Based Hybrid Quality Ranking for e-Commerce Product Search. Big Data and Cognitive Computing. 2021;5(3):35.doi: 10.3390/bdcc5030035.
20. Jurafsky D, Martin JH [Internet]. Speech and Language Processing." [Accessed Oct. 14, 2023]. Available from: https://web.stanford.edu/~jurafsky/slp3/
21. Lil'Log [Internet]. Learning Word Embedding. [Accessed Nov. 23, 2020]. Available from: https://lilianweng.github.io/lil-log/2017/10/15/learning-word-embedding.html
22. Cheng Y, Yao L, Xiang G, Zhang G, Tang T, Zhong L. Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism. IEEE Access. 2020;8:134964-75.doi: 10.1109/ACCESS.2020.3005823.
23. Liang H, Sun X, Sun Y, Gao Y. Text feature extraction based on deep learning: a review. EURASIP J Wirel Commun Netw. 2017;2017(1):211.
24. Rhanoui M, Mikram M, Yousfi S, Barzali S. A CNN-BiLSTM model for document-level sentiment analysis. Machine Learning and Knowledge Extraction. 2019;1(3):832-47.doi: 10.3390/make1030048.