APJ 7(2):62-74, Summer 2026

Arman Process Journal (APJ)
Quarterly Journal of ICT APJ
License Number: 87090

Arman Process Journal (APJ)
Homepage: https://www.armanprocessjournal.ir

A Hybrid Framework for Persian Sentiment Analysis, Integrating Transformer-based Language Models, Aspect-based Analysis, and Confidence Calibration

H. Ziafat [1] , B. Dorostkar [2*], P. Rahmani [3], N. Fareghzadeh [4]

[1] Department of computer engineering,Nat.C., Islamic Azad University,Natanz, Iran
2 Department of Information and Communication, Amin Police University, Tehran, Iran
3 Department of Computer Engineering, Islamic Azad University, Par.C., Tehran, Iran
4 Department of Computer Engineering, Khod.c., Islamic Azad University, Khodabandeh, Iran

RESEARCH PAPER

Received: 2026-3-27
Accepted: 2026-8-9

KEYWORDS:
Natural Language Processing,
Transformational,
Aspect-Based Sentiment Analysis,
Calibration,
Deep Learning

Corresponding author:
behnamdorostkar@gmail.com

ABSTRACT

Sentiment analysis is a key application of natural language processing and artificial intelligence, aimed at identifying and extracting attitudes, emotions, and polarity from human-generated text. Although deep learning models—particularly Transformer-based architectures—have significantly enhanced the capabilities of sentiment analysis systems, analyzing Persian text remains fraught with challenges such as writing variations, the use of zero-width non-joiners (ZWNJ), character form variations, semantic ambiguity, sarcasm, negation, multi-aspect perspectives, a scarcity of labeled data, and a mismatch between model confidence and actual prediction accuracy. Furthermore, many existing systems merely output a general positive, negative, or neutral label without accounting for their confidence in the decision. To address these challenges, this research proposes a multi-stage, context-aware, and uncertainty-driven framework for Persian sentiment analysis. This framework comprises four main layers: Persian preprocessing and normalization, deep representation extraction using Transformer models, aspect-based sentiment analysis, and finally, calibration and uncertainty estimation. This structure enables the distinction between "overall sentiment" and "sentiment regarding a specific aspect." Evaluation results demonstrate that the proposed framework outperforms baseline models across all metrics, including Accuracy, Precision, Recall, and Macro-F1. Additionally, an SCCI value of 91% indicates a high level of consistency and coherence in the model's ability to identify aspect-level sentiment within Persian texts. These results demonstrate that employing the proposed architecture has led to significant improvements not only in classification performance but also in the balanced and simultaneous detection of emotion labels.

Copyright © Author(s).

نشریه تخصصی آرمان پردازش، دوره ۷، شماره ۲، بهار ۱۴۰۵

فصلنامه تخصصی آرمان پردازش
(APJ)
Homepage: www.armanprocessjournal.ir

فصلنامه تخصصی فناوری اطلاعات و ارتباطات
شماره مجوز: ۸۷۰۹۰

چارچوب ترکیبی تحلیل احساسات متون فارسی با تلفیق مدل‌های زبانی مبتنی بر ترنسفورمر، تحلیل جنبه‌محور و کالیبراسیون اعتماد

حسن ضیافت[۱]، بهنام درستکار*[۲]، پریسا رحمانی[۳]، نفیسه فارغزاده[۴]

۱ گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی، واحد نطنز، ایران
۲ گروه اطلاعات و ارتباطات، دانشگاه علوم انتظامی امین، تهران، ایران.
۳ گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی، واحد پردیس، ایران
۴ گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی، واحد خدابنده، ایران

چکیده

مقاله پژوهشی

واژگان کلیدی:
پردازش زبان طبیعی،
ترنسفورمر،
تحلیل احساسات جنبه‌محور،
کالیبراسیون،
یادگیری عمیق

تحلیل احساسات یکی از مهم‌ترین کاربردهای پردازش زبان طبیعی و هوش مصنوعی است که با هدف شناسایی و استخراج نگرش، احساس و قطبیت موجود در متون انسانی انجام می‌شود. اگرچه مدل‌های یادگیری عمیق و به‌ویژه معماری‌های مبتنی بر ترنسفورمر موجب افزایش چشمگیر توانایی سامانه‌های تحلیل احساسات شده‌اند، تحلیل متون فارسی همچنان با چالش‌هایی نظیر تنوع نوشتاری، نیم‌فاصله، شکل‌های مختلف حروف، ابهام معنایی، کنایه، نفی، چندجنبه‌ای بودن دیدگاه، کمبود داده‌های برچسب‌خورده و عدم تطابق میان اطمینان مدل و صحت واقعی پیش‌بینی مواجه است. افزون بر این، بسیاری از سامانه‌های موجود تنها یک برچسب کلی مثبت، منفی یا خنثی تولید می‌کنند و میزان اعتماد خود به تصمیم را در نظر نمی‌گیرند. بمنظور مواجه با چالش های موجود، در این پژوهش، یک چارچوب چندمرحله‌ای، زمینه‌آگاه و عدم‌قطعیت‌محور برای تحلیل احساسات فارسی پیشنهاد می‌شود که از چهار لایه اصلی شامل پیش‌پردازش و نرمال‌سازی فارسی، استخراج نمایش‌های عمیق با مدل‌های ترنسفورمری، تحلیل احساسات جنبه‌محور و در نهایت کالیبراسیون و تخمین عدم‌قطعیت تشکیل شده است. این ساختار امکان تفکیک «احساس کلی» از «احساس نسبت به جنبه مشخص» را فراهم می‌سازد.

نتایج ارزیابی نشان می‌دهد که چارچوب عملیاتی پیشنهادی در تمامی شاخص‌های ارزیابی شامل Accuracy، Precision، Recall و Macro-F1 نسبت به مدل‌های پایه عملکرد بالاتری دارد. همچنین مقدار SCCI برابر با ۹۱ % نشان‌دهنده سطح بالای سازگاری و انسجام مدل در شناسایی احساسات در سطح جنبه در متن‌های فارسی است. این نتایج نشان می‌دهد که بهره‌گیری از معماری پیشنهادی، بهبود قابل‌توجهی را نه‌تنها در عملکرد طبقه‌بندی، بلکه در تشخیص متوازن و هم‌زمان برچسب‌های احساسی ایجاد کرده است.

۱- مقدمه

رشد شبکه‌های اجتماعی، فروشگاه‌های اینترنتی، سامانه‌های خدمات مشتری، تالارهای گفتگو و پلتفرم‌های تولید محتوای کاربرمحور موجب شده است که حجم عظیمی از اطلاعات به شکل متن تولید شود. بخش قابل‌توجهی از این اطلاعات حاوی دیدگاه، قضاوت، رضایت، نارضایتی و احساسات کاربران است. استخراج خودکار چنین اطلاعاتی می‌تواند در تصمیم‌گیری سازمانی، مدیریت ارتباط با مشتری، تحلیل بازار، پایش برند، توصیه‌گرها و سامانه‌های پشتیبان تصمیم مورد استفاده قرار گیرد [۱].

تحلیل احساسات را می‌توان شاخه‌ای از داده‌کاوی و پردازش زبان طبیعی دانست که هدف آن شناسایی نگرش یا قطبیت عاطفی موجود در متن است. در ساده‌ترین حالت، یک متن در یکی از طبقات مثبت، منفی و خنثی قرار می‌گیرد. با این حال، چنین رویکردی برای بسیاری از کاربردهای واقعی کافی نیست؛ زیرا ممکن است یک کاربر درباره ویژگی‌های مختلف یک محصول احساسات متفاوتی داشته باشد [۲]. در این حوزه تحلیل جنبه‌محور راهگشا می باشد که این سبک تحلیل به جای اختصاص یک قطبیت کلی به کل متن، ابتدا جنبه یا ویژگی موردنظر را شناسایی کرده و سپس احساس مرتبط با آن را تعیین می‌کند [۳]. در سال‌های اخیر، ظهور مدل‌های زبانی پیش‌آموزش‌دیده نظیرم مدل BERT تحول مهمی در پردازش زبان طبیعی ایجاد کرده است. BERT با استفاده از نمایش دوسویه و توجه به زمینه واژگان توانست عملکرد بسیاری از وظایف زبان طبیعی را بهبود دهد [۴-۵]. مدل‌های پس از آن نظیر RoBERTa، XLM-R و مدل‌های جمله‌محور نیز ظرفیت انتقال دانش میان زبان‌ها و وظایف مختلف را افزایش داده‌اند [۶].

با وجود این پیشرفت‌ها، زبان فارسی ویژگی‌هایی دارد که استفاده مستقیم از مدل‌های توسعه‌یافته برای زبان انگلیسی را محدود می‌کند. وجود شکل‌های متفاوت حروف «ی» و «ک»، نیم‌فاصله، اتصال یا جدانویسی، پسوندها، پیشوندها، تنوع املایی، استفاده از واژگان محاوره‌ای و فینگلیش و همچنین پیچیدگی ساختار جمله از جمله عوامل مؤثر هستند [۷]. مسئله مهم دیگر عدم‌قطعیت مدل است. یک مدل ممکن است برای متنی که واقعاً ابهام زیادی دارد، احتمال ۰٫۹۵ برای یک کلاس تولید کند [۸]. بنابراین احتمال خروجی مدل الزاماً معادل احتمال صحت تصمیم نیست. پژوهش Guo و همکاران نشان داده است که شبکه‌های عصبی مدرن ممکن است به‌خوبی کالیبره نباشند و روش‌هایی مانند Temperature Scaling می‌توانند این مشکل را کاهش دهند [۹].

لازم به ذکر است، تحلیل احساسات در ابتدا عمدتاً مبتنی بر روش‌های واژه‌نامه‌ای بود. در این روش‌ها برای واژگان مختلف وزن مثبت یا منفی تعیین می‌شد و قطبیت متن از روی مجموع یا ترکیب وزن واژگان محاسبه می‌شد [۱۰]. منابع واژگانی مانند SentiWordNet نمونه‌ای از این رویکرد هستند [۱۱]. مزیت این روش‌ها سادگی و قابلیت تفسیر است؛ اما در مواجهه با وابستگی به زمینه، کنایه، نفی و واژگان جدید محدودیت دارند.

روش‌های یادگیری ماشین مانند Naive Bayes، ماشین بردار پشتیبان و رگرسیون لجستیک مرحله بعدی تکامل تحلیل احساسات بودند. Maas و همکاران نشان دادند که می‌توان نمایش برداری واژگان را به گونه‌ای یاد گرفت که اطلاعات معنایی و احساسی را همزمان در خود حفظ کند [۱۲]. شبکه‌های عصبی بازگشتی و مدل‌های LSTM توانستند وابستگی‌های طولانی‌تر میان کلمات را بهتر از روش‌های سنتی مدل کنند [۱۳-۱۴].

بااین‌حال، نقطه عطف مهم در این حوزه معماری Transformer بود. Vaswaniو همکاران معماری Transformer را بر مبنای سازوکار Self-Attention معرفی کردند [۷]. BERT با پیش‌آموزش دوسویه نیز امکان استخراج نمایش‌های وابسته به زمینه را فراهم کرد [۵]. ایده اصلی این است که معنای یک واژه صرفاً از خود واژه حاصل نمی‌شود، بلکه به کلمات اطراف آن نیز وابسته است. در سوی دیگر RoBERTa نسخه‌ای بهینه‌سازی‌شده از روند پیش‌آموزش BERT است که نشان داد انتخاب داده، حجم آموزش و تنظیمات پیش‌آموزش می‌تواند تأثیر قابل‌توجهی بر عملکرد داشته باشد [۶]. این مسئله برای زبان فارسی اهمیت خاصی دارد؛ زیرا کیفیت و حجم پیکره آموزشی می‌تواند بر کیفیت نمایش زبان تأثیر بگذارد. همچنین مدل XLM-R با آموزش روی چندین زبان، قابلیت انتقال دانش میان زبان‌ها را افزایش داد [۸]. این موضوع برای زبان‌هایی که داده آموزشی کمتری نسبت به انگلیسی دارند اهمیت ویژه‌ای دارد. بااین‌حال، انتقال چندزبانه همیشه جایگزین مدل زبانی تخصصی زبان مقصد نیست؛ زیرا ویژگی‌های صرفی، نحوی و نوشتاری زبان‌ها متفاوت است.

براساس بررسی تاریخچه پژوهش یکی از چالش‌های اساسی تحلیل احساسات فارسی، نرمال‌سازی است. بر این اساس، سؤال اصلی این پژوهش این است که چگونه می‌توان یک سامانه تحلیل احساسات فارسی طراحی کرد که علاوه بر استخراج قطبیت و جنبه، میزان اطمینان خود نسبت به پیش‌بینی را نیز برآورد کرده و برای داده‌های حجیم قابل مقیاس‌پذیری باشد؟ برای پاسخ به این سؤال بنیادین، یک معماری چندمرحله‌ای پیشنهاد می‌شود که در آن سه مفهوم به یکدیگر متصل می‌شوند:

- نمایش زبانی عمیق؛
- تحلیل جنبه‌محور؛
- تخمین و کالیبراسیون عدم‌قطعیت .

لذا مسئله اصلی این پژوهش، طراحی یک چارچوب یکپارچه برای تحلیل احساسات متون فارسی است که نه‌تنها قطبیت احساس را پیش‌بینی کند، بلکه جنبه موردنظر، میزان توافق مدل‌ها و درجه اعتماد به خروجی را نیز در نظر بگیرد. پرسش محوری این است که آیا ترکیب نمایش‌های زمینه‌آگاه، تحلیل جنبه‌ای، Ensemble و

کالیبراسیون می‌تواند خروجی‌ای فراهم کند که برای استفاده در سامانه‌های واقعی، اطلاعات بیشتری درباره اطمینان و عدم‌قطعیت ارائه دهد. لذا، اهداف اصلی پژوهش عبارت‌اند از طراحی خط لوله پیش‌پردازش مناسب برای متون فارسی، استخراج نمایش‌های زمینه آگاه با مدل‌هایTransformer/ParsBERT ، تعیین احساس در سطح جنبه و ارائه معماری قابل استقرار. در بخش های بعدی مقاله دیدگاه پیشنهادی پژوهش حاضر درجهت مرتفع سازی چالش های ذکر شده و ابعاد مرتبط را با جزئیات تشریح خواهیم نمود.

۲- تحقیقات مرتبط و مقایسه

ظهور معماری Transformer و سازوکار Self-Attention تحول مهمی در پردازش زبان طبیعی ایجاد کرد. پس از معرفیBERT ، استفاده از مدل‌های زبانی پیش‌آموزش‌دیده در تحلیل احساسات و تحلیل جنبه‌محور به‌سرعت گسترش یافت. پژوهش‌های انجام‌شده از روش‌های واژه‌نامه‌ای و یادگیری ماشین کلاسیک به شبکه‌های عصبی، Transformer و تحلیل جنبه‌محور حرکت کرده‌اند. جدول زیر مطالعات مهم و ارتباط آن‌ها با دیدگاه پیشنهادی پژوهش حاضر را خلاصه می‌کند.

جدول ۱. تحقیقات مرتبط و مقایسه

مرجع	روش / رویکرد	مهم‌ترین دستاورد	خلأ مرتبط	تفاوت با پژوهش پیشنهادی
[۱و۲]	تحلیل احساسات کلاسیک	پایه‌گذاری مفاهیم و روش‌های تحلیل احساسات	محدودیت در نمایش زمینه پیچیده	ترکیب Transformer ، ABSA و اعتماد
[۴]	ABSA	استانداردسازی وظایف و داده‌های تحلیل جنبه‌ای	تمرکز بر جنبه و احساس بدون چارچوب اعتماد جامع	افزودن Ensemble و Calibration
[۱۳]	LSTM	مدل‌سازی وابستگی‌های ترتیبی	محدودیت نسبت به مدل سازی موازی و وابستگی های دور	استفاده ازTransformer
[۲۱]	CNN	استخراج مؤثر الگوهای محلی	محدودیت در وابستگی های بلندبرد	نمایش زمینه‌آگاه
[۷]	Transformer	معرفی Attention و مدل‌سازی وابستگی‌های دور	معماری عمومی و نه چارچوب کامل احساسات	استفاده به‌عنوان هسته مدل پیشنهادی
[۵]	BERT	نمایش دوسویه و زمینه‌آگاه	امکان بیش‌اطمینانی و نبود کالیبراسیون ذاتی	ترکیبBERT/ParsBERT با Calibration
[۸]	XLM-R	نمایش چندزبانه	عدم تمرکز اختصاصی بر ویژگی‌های فارسی	امکان مقایسه با مدل فارسی
[۹]	Temperature Scaling	بهبود کالیبراسیون احتمال	تمرکز بر کالیبراسیون، نه ABSA	ادغام کالیبراسیون با تحلیل جنبه‌ای
[۲۷و۲۹]	Subword	پوشش بهتر واژه‌های ناشناخته	حل نکردن مسئله اعتماد	استفاده در پیش پردازش/نمایش
[۳۰و۳۲]	معماری و مدیریت Cloud	مدیریت عملکرد و مقیاس‌پذیری	مستقیماً برای تحلیل احساسات نیست	الهام برای استقرار مقیاس پذیر
[۳۴]	سامانه توصیه‌گر ترکیبی	ترکیب دانش و روش‌های توصیه‌گری	تمرکز بر توصیه‌گر	استفاده از احساس و اعتماد در توصیه‌گر
[۴۰]	مرور یادگیری عمیق	جمع‌بندی رویکردهای عمیق تحلیل احساسات	فاقد چارچوب یکپارچه پیشنهادی	ترکیب چند مؤلفه در یک معماری

براساس مرور و بررسی تحقیقات مرتبط، تحلیل احساسات در ابتدا عمدتاً بر روش‌های مبتنی بر واژه‌نامه، قواعد زبانی و الگوریتم‌های کلاسیک یادگیری ماشین استوار بود. روش‌های واژه‌نامه‌ای با استفاده از منابعی نظیر SentiWordNet تلاش می‌کردند جهت‌گیری عاطفی واژگان را تعیین کنند، درحالی‌که روش‌های یادگیری ماشین مانند ماشین بردار پشتیبان، Naive Bayes و رگرسیون لجستیک از ویژگی‌هایی نظیر n-gram و TF-IDF برای طبقه‌بندی متون استفاده می‌کردند. مزیت اصلی این روش‌ها سادگی، هزینه محاسباتی پایین و قابلیت تفسیر نسبتاً مناسب بود؛ بااین‌حال، وابستگی شدید به مهندسی ویژگی و ضعف در درک وابستگی‌های معنایی و زمینه‌ای، عملکرد آنها را در متون پیچیده محدود می‌کرد [۳۶و۱۰و۲و۱].

با توسعه یادگیری عمیق، مدل‌هایی نظیر CNN ، RNN، LSTM و Bi-LSTM برای استخراج خودکار ویژگی‌های معنایی مورد استفاده قرار گرفتند. این روش‌ها در مقایسه با رویکردهای کلاسیک توانایی بیشتری در مدل‌سازی ترتیب واژگان و وابستگی‌های متنی داشتند، اما در متون طولانی و در شرایطی که وابستگی میان بخش‌های دور از هم جمله اهمیت داشت، با محدودیت مواجه بودند [۲۱و۱۴و۱۳].

مرور پژوهش‌های موجود نشان می‌دهد که چهار گروه اصلی از مشکلات همچنان وجود دارند:

- شکاف اول: تمرکز بیش از حد بر قطبیت کلی، بسیاری از سامانه‌ها صرفاً مثبت/منفی/خنثی را پیش‌بینی می‌کنند.
- شکاف دوم: عدم توجه کافی به عدم‌قطعیت، مدل ممکن است پیش‌بینی اشتباه اما با اطمینان بسیار بالا ارائه کند.
- شکاف سوم: دشواری زبان فارسی، مدل‌های عمومی الزاماً برای تنوع زبانی و نوشتاری فارسی بهینه نشده‌اند.
- شکاف چهارم: فاصله میان تحلیل احساسات و تصمیم‌گیری، نتیجه تحلیل احساسات زمانی ارزش عملی بیشتری دارد که بتواند وارد یک سامانه توصیه‌گر یا تصمیم‌یار شود.

لذا، اهداف اصلی پژوهش عبارت‌اند از طراحی خط لوله پیش‌پردازش مناسب برای متون فارسی، استخراج نمایش‌های زمینه‌آگاه با مدل های Transformer/ParsBERT ، تعیین احساس در سطح جنبه و ارائه معماری قابل استقرار. در بخش بعدی چارچوب هوشمندی پیشنهادی و اجزاء آن را تشریح خواهیم نمود. در مقایسه با تحقیقات مرتبط، مهم‌ترین دستاورد دیدگاه پیشنهادی پژوهش حاضر، ایجاد یک چارچوب یکپارچه برای تبدیل متن خام فارسی به اطلاعات احساسی قابل استفاده و قابل اعتماد است. در این چارچوب، برخلاف روش‌های سنتی که عمدتاً یک قطبیت کلی برای متن تولید می‌کنند، ابتدا ساختار نوشتاری فارسی نرمال‌سازی شده، سپس نمایش عمیق زمینه‌ای استخراج می‌شود و بعد جنبه‌های مورد اشاره کاربر و قطبیت مرتبط با هر جنبه تعیین می‌گردد. افزودن لایه کالیبراسیون و شاخص SCCI نیز امکان تفکیک پیش‌بینی‌های با اطمینان بالا از نمونه‌های مبهم را فراهم می‌سازد؛ بنابراین سامانه می‌تواند در مواردی که عدم‌قطعیت زیاد است، نمونه را برای بررسی انسانی یا پردازش تکمیلی علامت‌گذاری کند. از منظر کاربردی، خروجی چهارگانه جنبه، احساس، شدت و اعتماد می‌تواند مستقیماً برای تحلیل بازخورد مشتریان، پایش رضایت، تحلیل نظرات دانشجویان، مدیریت ارتباط با مشتری و سیستم‌های توصیه‌گر استفاده شود. همچنین طراحی ماژولار چارچوب، با الهام از اصول معماری و ارزیابی عملکرد در سیستم‌های مقیاس‌پذیر، امکان استقرار اجزای پردازش متن، مدل زبانی، تحلیل جنبه، کالیبراسیون و داشبورد را به‌صورت مستقل فراهم می‌کند. در نتیجه، نوآوری اصلی این دیدگاه نه صرفاً جایگزینی یک الگوریتم با الگوریتمی جدید، بلکه ترکیب دقت زبانی، تحلیل جنبه‌ای، قابلیت اعتماد، ارزیابی عدم‌قطعیت و قابلیت استفاده عملیاتی در یک معماری واحد است. درواقع نوآوری این پژوهش در یکپارچه‌سازی Transformer، ABSA، Ensemble و Calibration و ارائه شاخص SCCI برای ارزیابی اعتماد خروجی است.

۳- چارچوب عملیاتی پیشنهادی

چارچوب پیشنهادی به‌صورت یک زنجیره چندمرحله‌ای طراحی شده است: دریافت داده، پاک‌سازی و نرمال‌سازی، توکن‌سازی زیرواژه‌ای، استخراج نمایش زمینه‌آگاه، استخراج یا شناسایی جنبه، طبقه‌بندی احساس، تجمیع خروجی مدل‌ها، کالیبراسیون، محاسبه عدم‌قطعیت و تولید خروجی نهایی .در این معماری، خروجی نهایی صرفاً یک برچسب نیست، بلکه شامل برچسب احساس، احتمال کالیبره‌شده، جنبه شناسایی‌شده و شاخص اعتماد است.



شکل ۱. چارچوب عملیاتی پیشنهادی تحلیل احساسات چندجنبه‌ای فارسی با رویکرد اعتمادمحور

همان‌گونه که در شکل ۱. مشاهده می‌شود، چارچوب عملیاتی پیشنهادی یک خط لوله یکپارچه برای تحلیل احساسات چندمحور زبان فارسی با قابلیت سنجش و مدیریت اعتماد است که از دریافت متن فارسی آغاز شده و پس از طی مراحل پیش‌پردازش، نمایش معنایی، استخراج جنبه و احساس، ترکیب مدل‌ها و محاسبه شاخص اعتماد، خروجی نهایی را تولید می‌کند. هدف اصلی این معماری آن است که علاوه بر تعیین قطبیت احساس، میزان اطمینان و قابلیت اتکای تصمیم مدل نیز در سطح خروجی مشخص شود. ازاین‌رو، چارچوب پیشنهادی صرفاً یک مدل طبقه‌بندی احساسات نیست، بلکه یک سازوکار چندمرحله‌ای برای تحلیل دقیق، چندجنبه‌ای و اعتمادپذیر احساسات فارسی محسوب می‌شود.

در مرحله نخست، متن فارسی به‌عنوان ورودی سیستم دریافت می‌شود .با توجه به ویژگی‌های خاص زبان فارسی، از جمله تنوع شکل نوشتاری، وجود نیم‌فاصله، علائم نگارشی، نویزهای متنی، نشانی‌های اینترنتی و تفاوت در نحوه نگارش برخی واژه‌ها، داده خام مستقیماً وارد مدل یادگیری نمی‌شود. در مرحله پیش‌پردازش، عملیات‌هایی نظیر یکسان‌سازی حروف، حذف نویز و علائم غیرضروری، اصلاح فاصله و نیم‌فاصله، حذف یا استانداردسازیURL ها و نرمال‌سازی نگارشی انجام می‌شود. این مرحله موجب کاهش تغییرات غیرمعنایی در داده و ایجاد نمایش یکنواخت‌تر برای مراحل بعدی می‌شود.

در مرحله سوم، متن نرمال‌شده به واحدهای قابل پردازش برای مدل تبدیل می‌شود. توکن‌سازی مناسب زبان فارسی و تولید نمایش‌های عددی و زمینه‌محور، امکان انتقال اطلاعات واژگانی و معنایی متن به مدل‌های مبتنی بر Transformer را فراهم می‌کند. در این بخش، نمایش‌های زمینه‌ای حاصل از معماری‌هایی نظیر BERT و RoBERTaمی‌توانند برای درک وابستگی میان واژه‌ها و جمله‌ها مورد استفاده قرار گیرند. مزیت این مرحله نسبت به روش‌های سنتی مبتنی بر ویژگی‌های سطحی آن است که معنای هر واژه با توجه به بافت جمله در نظر گرفته می‌شود و در نتیجه ابهام‌های معنایی و تغییر معنای واژه‌ها در زمینه‌های مختلف بهتر مدیریت می‌شوند.

در مرحله چهارم، ماژول تحلیل احساسات مبتنی بر جنبه(ABSA) ، ابتدا جنبه‌ها یا موضوعات مورد اشاره در متن را شناسایی کرده و سپس احساس مرتبط با هر جنبه را تعیین می‌کند. بنابراین، سیستم صرفاً یک برچسب کلی مانند «مثبت» یا «منفی» برای کل جمله تولید نمی‌کند، بلکه قادر است رابطه میان یک جنبه و احساس مرتبط با آن را استخراج کند. برای نمونه، در یک نظر درباره یک محصول ممکن است «کیفیت» مثبت، اما «قیمت» منفی ارزیابی شده باشد. این تفکیک جنبه‌ها و احساسات باعث افزایش دقت تحلیلی چارچوب در مقایسه با طبقه‌بندی کلی احساسات می‌شود و امکان تحلیل دقیق‌تر بازخوردهای کاربران را فراهم می‌کند.

در مرحله پنجم، چند مدل یادگیری عمیق به‌صورت Ensemble با یکدیگر ترکیب می‌شوند تا وابستگی خروجی نهایی به یک مدل منفرد کاهش یابد. در این معماری، خروجی مدل‌های مختلف می‌تواند بر اساس احتمال‌های پیش‌بینی‌شده با یکدیگر ترکیب شود. سپس با استفاده از Temperature Scaling، احتمال‌های خروجی کالیبره می‌شوند تا مقدار احتمال پیش‌بینی‌شده با میزان واقعی اطمینان مدل سازگاری بیشتری داشته باشد. این موضوع به‌ویژه در کاربردهایی که تصمیم‌گیری بر اساس میزان اطمینان مدل انجام می‌شود اهمیت دارد، زیرا یک احتمال بالا لزوماً به معنای یک پیش‌بینی قابل اعتماد نیست و باید کالیبراسیون آن نیز مورد توجه قرار گیرد.

در ادامه، ماژول سنجش شاخص اطمینان (SCCI) میزان اعتمادپذیری هر پیش‌بینی را ارزیابی می‌کند. در معماری پیشنهادی، این شاخص از ترکیب چند مؤلفه شامل بیشینه احتمال پیش‌بینی، آنتروپی خروجی، میزان توافق مدل‌های Ensemble و کیفیت ورودی

تشکیل می‌شود. به‌صورت مفهومی می‌توان آن را به شکل زیر بیان کرد:

(۱)
$$\mathrm{SCCI} = \alpha P_{\max} + \beta(1 - H_N) + \gamma S_{\mathrm{ensemble}} + \delta Q_x$$

که در آن $Pmax$ بیشترین احتمال پیش‌بینی، H_N آنتروپی توزیع احتمال، $S_{ensemble}$ میزان توافق مدل‌های تشکیل‌دهنده Ensemble و Q_X کیفیت ورودی است. ضرایب α، β، γ و δ نیز وزن هر مؤلفه را مشخص می‌کنند. تمام مؤلفه‌های SCCI در بازه [۰،۱] نرمال شده و ضرایب نیز به‌گونه‌ای تنظیم می‌شوند که $\alpha+\beta+\gamma+\delta=1$ باشد؛ بنابراین محدوده $SCCI \in [0,1]$ می‌باشد. $S_{ensemble}$ به‌صورت میانگین توافق احتمالاتی مدل‌های عضو Ensemble برای کلاس پیش‌بینی‌شده محاسبه و در بازه [۰،۱] نرمال می‌شود. Qx یک امتیاز نرمال‌شده کیفیت متن ورودی بر اساس شاخص‌های نویز و کامل‌بودن متن است. بنابراین، پیش‌بینی‌ای که علاوه بر احتمال بالا، آنتروپی پایین، توافق مناسب میان مدل‌ها و ورودی باکیفیت داشته باشد، مقدار اطمینان بیشتری دریافت خواهد کرد. در مرحله نهایی، خروجی چارچوب شامل برچسب احساس، جنبه یا جنبه‌های شناسایی‌شده، احتمال پیش‌بینی و مقدار SCCI است. توصیف جزئیات عملیات فازهای اجرایی در چارچوب پیشنهادی به شرح زیر می‌باشد:

فاز پیش‌پردازش و نرمال‌سازی متون فارسی

پیش‌پردازش برای متون فارسی شامل یکسان‌سازی نویسه‌های عربی و فارسی، مدیریت نیم‌فاصله، حذف یا کنترل نشانه‌های زائد، مدیریت URL و نام کاربری، کنترل تکرار حروف، توکن‌سازی و در صورت نیاز حذف نویزهای متنی است .حذف کلمات توقف باید با احتیاط انجام شود، زیرا برخی واژه‌های کوتاه می‌توانند در تشخیص نفی و جهت‌گیری احساس نقش داشته باشند .استفاده از توکن‌سازی زیرواژه ای مانند SentencePiece نیز می‌تواند برای پوشش واژه‌های ناشناخته و اشکال صرفی مفید باشد [۲۹-۲۷].

فاز استخراج نمایش زمینه‌آگاه باTransformer

در معماریTransformer، توجه به مدل اجازه می‌دهد وابستگی میان بخش‌های مختلف متن را بدون پردازش ترتیبی اجباری مدل کند . تابع توجه به‌صورت زیر تعریف می‌شود:

(۲)
$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

در چارچوب پیشنهادی، خروجی لایه‌های Transformer/ParsBERT به‌عنوان نمایش زمینه‌آگاه مورد استفاده قرار می‌گیرد. این نمایش به‌جای اختصاص یک بردار ثابت به هر واژه، معنای آن را با توجه به متن اطراف تعیین می‌کند. این ویژگی برای تشخیص نفی، کنایه، وابستگی جنبه و تفاوت معنایی واژه‌ها اهمیت دارد [۷].

فاز تحلیل احساسات جنبه‌محور و یادگیری چندوظیفه‌ای

در ABSA، هدف فقط تعیین احساس کل متن نیست؛ بلکه باید مشخص شود احساس مربوط به کدام جنبه است .بنابراین معماری می‌تواند دو وظیفه مرتبط را به‌صورت هم‌زمان آموزش دهد :شناسایی یا طبقه‌بندی جنبه و تعیین قطبیت احساس .در حالت چندوظیفه‌ای، تابع هزینه کلی به‌شکل زیر تعریف می‌شود:

(۳)
$$\mathcal{L} = \lambda_1 \mathcal{L}_{\mathrm{aspect}} + \lambda_2 \mathcal{L}_{\mathrm{sent}} + \lambda_3 \mathcal{L}_{\mathrm{cal}}$$

در این رابطه، L_{aspect} خطای وظیفه جنبه، L_{sent} خطای طبقه‌بندی احساس و ضرایب λ ضرایب وزن هر بخش را تعیین می‌کنند. کالیبراسیون احتمال به‌صورت پس‌آموزشی با Temperature Scaling انجام می‌شود. پارامتر T با کمینه‌سازی NLL روی مجموعه اعتبارسنجی تعیین و سپس بدون بازآموزی مدل روی مجموعه آزمون اعمال شد.

فاز Ensemble و برآورد عدم‌قطعیت

برای کاهش وابستگی به یک مدل، می‌توان چند مدل یا چند اجرای متفاوت از یک مدل را با یکدیگر ترکیب کرد. احتمال تجمیعی به صورت زیر تعریف می‌شود:

(۴)
$$P_{\mathrm{ensemble}} = \sum_m w_m P_m, \qquad \sum_m w_m = 1$$

که در آن P_m توزیع احتمال مدل m و w_m وزن آن است و مجموع وزن‌ها برابر یک در نظر گرفته می‌شود . وزن مدل‌ها بر اساس عملکرد آن‌ها در مجموعه اعتبارسنجی تعیین و سپس نرمال‌سازی می‌شود. در حالت پایه، وزن مدل‌ها برابر در نظر گرفته شد و مجموع وزن‌ها برابر یک است. علاوه بر احتمال بیشینه، آنتروپی توزیع احتمال می تواند به شرح زیر برای اندازه‌گیری عدم‌قطعیت استفاده شود:

(۵)
$$H(P) = -\sum_i p_i \log p_i$$

برای مقایسه نمونه‌های مختلف می‌توان آنتروپی را نرمال کرد. آنتروپی بالاتر نشان‌دهنده توزیع پراکنده‌تر احتمال و در نتیجه عدم قطعیت بیشتر است.

فاز کالیبراسیون اعتماد و شاخص SCCI

یکی از اجزای اصلی چارچوب پیشنهادی، تفکیک احتمال خام مدل از احتمال کالیبره‌شده است . در Temperature Scaling ، لگیت های مدل با یک پارامتر دما T مقیاس می‌شوند و احتمال جدید از Softmax(z/T) به شرح زیر به‌دست می‌آید [۹]:

(۶)

$$\mathrm{softmax}\left(\frac{z}{T}\right) \quad \text{or} \quad p_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}$$

نهایتا برای ترکیب چند نشانه اعتماد، شاخص پیشنهادی SCCI به شکل زیر تعریف می‌شود:

(۷)

$$\mathrm{SCCI} = \alpha P_{\max} + \beta(1 - H_N) + \gamma S_{\text{ensemble}} + \delta Q_x$$

در این رابطه، *Pmax* بیشترین احتمال کالیبره‌شده، H_N آنتروپی نرمال‌شده، $S_{ensemble}$ میزان توافق مدل‌های عضو Ensemble و Q_x شاخص کیفیت ورودی است .ضرایب α، β ، γ و δ بر اساس اعتبارسنجی و هدف کاربردی تعیین می‌شوند.

بدین ترتیب، خروجی سیستم تنها یک تصمیم قطعی نیست، بلکه همراه با اطلاعاتی درباره میزان اطمینان به آن ارائه می‌شود. این ویژگی امکان تفکیک پیش‌بینی‌های با اطمینان بالا از موارد مبهم یا کم‌اعتماد را فراهم می‌کند و می‌تواند در سامانه‌های تحلیلی، پایش بازخورد کاربران، تحلیل دیدگاه‌ها و سایر کاربردهای تصمیم‌یار مورد استفاده قرار گیرد.

جدول ۲. تفکیک ورودی، عملیات و خروجی سیستم

مرحله	ورودی	عملیات اصلی	خروجی
1	متن خام فارسی	حذف نویز، نرمال‌سازی، یکسان سازی فاصله و نیم‌فاصله	متن استاندارد
2	متن استاندارد	توکن‌سازی و نمایش زیرواژه‌ای	شناسه‌های توکن
3	توکن‌ها	Transformer/ParsBERTو Attention	نمایش زمینه‌آگاه
4	نمایش متن	شناسایی جنبه‌ها و ارتباط آن‌ها با متن	جنبه‌ها
5	جنبه + زمینه	طبقه‌بندی احساس	احتمال احساس
6	خروجی چند مدل	Ensemble و وزن‌دهی	احتمال تجمیعی
7	احتمال خام	Temperature Scaling	احتمال کالیبره‌شده
8	احتمال + توافق مدل‌ها + کیفیت ورودی	محاسبه SCCI و عدم‌قطعیت	اعتماد نهایی

در مجموع، نوآوری اصلی چارچوب پیشنهادی در اتصال تحلیل احساسات مبتنی بر جنبه، مدل‌های Transformer ، Ensemble، کالیبراسیون احتمال و سنجش اعتماد در یک معماری واحد است. این ترکیب تلاش می‌کند سه مسئله مهم تحلیل احساسات فارسی، یعنی درک دقیق محتوای زبانی، کاهش وابستگی به یک مدل منفرد و افزایش قابلیت اعتماد به خروجی را به‌صورت همزمان مورد توجه قرار دهد. بنابراین، چارچوب پیشنهادی می‌تواند مبنایی برای توسعه سامانه‌های تحلیل احساسات فارسی باشد که علاوه بر دقت طبقه‌بندی، نسبت به میزان اطمینان و کیفیت تصمیم‌های تولیدشده نیز حساس هستند. در بخش بعدی دیدگاه پیشنهادی پژوهش حاضر را ارزیابی خواهیم نمود.

۴- ارزیابی و بحث

بمنظور ارزیابی چارچوب عملیاتی پیشنهادی در این پژوهش، ارزیابی بر روی مجموعه‌داده Pars-ABSA در نظر گرفته شد. داده‌ها با نسبت ۸۰ درصد آموزش، ۱۰ درصد اعتبارسنجی و ۱۰ درصد آزمون تفکیک شدند. پارامتر T با کمینه‌سازی NLL روی مجموعه اعتبارسنجی تعیین و سپس بدون بازآموزی مدل روی مجموعه آزمون اعمال شد. طول بیشینه ورودی برابر ۲۵۶ توکن، اندازه دسته ۱۶، تعداد دوره‌های آموزش ۵ و نرخ یادگیری 2×10^{-5} در نظر گرفته شد و بهینه‌ساز AdamW مورد استفاده قرار گرفت. برای ارزیابی جامع، معیارهایAccuracy ، Precision ،Recall ،Macro-F1 ،ECE و SCCI محاسبه شدند [۴۱].

ECE میزان اختلاف میان احتمال‌های پیش‌بینی‌شده و دقت تجربی مدل در بازه‌های اطمینان را اندازه‌گیری می‌کند و مقدار کمتر نشان‌دهنده کالیبراسیون بهتر است. همچنین برای بررسی اثر اجزای پیشنهادی، شش پیکربندی شامل TF-IDF+SVM، CNN، BiLSTM، Transformer، Transformer+ABSA و چارچوب کامل پیشنهادی مقایسه شدند.

جدول۳. نتایج کمی ارزیابی پیکربندی‌های مختلف

مدل	Accuracy (%)	Precision (%)	Recall (%)	Macro-F1 (%)	ECE ↓	SCCI ↑
TF-IDF + SVM	82.40	81.70	80.90	81.20	0.142	0.71
CNN	84.60	84.10	83.50	83.70	0.126	0.74
BiLSTM	85.30	84.90	84.20	84.50	0.118	0.76
Transformer	88.10	87.80	87.20	87.50	0.091	0.81
Transformer + ABSA	89.40	89.10	88.70	88.90	0.078	0.84
چارچوب پیشنهادی	92.30	92.00	91.70	91.80	0.041	0.91

نتایج ارزیابی نشان می‌دهد که افزودن تدریجی اجزای چارچوب پیشنهادی موجب بهبود همزمان عملکرد طبقه‌بندی و قابلیت اعتماد مدل شده است. در مقایسه با Transformer پایه، استفاده از ماژول ABSAمقدار Macro-F1 را از 87.50% به ۸۸,۹۰ % افزایش داده است؛ در حالی که ترکیب نهاییTransformer، تحلیل جنبه‌محور، Ensemble و کالیبراسیون احتمال، Macro-F1 را به 91.80% رسانده است. بنابراین، بهبود حاصل صرفاً ناشی از افزایش پیچیدگی مدل نیست، بلکه تفکیک جنبه‌ها و ترکیب چند مدل نیز در افزایش کیفیت پیش‌بینی نقش دارد.

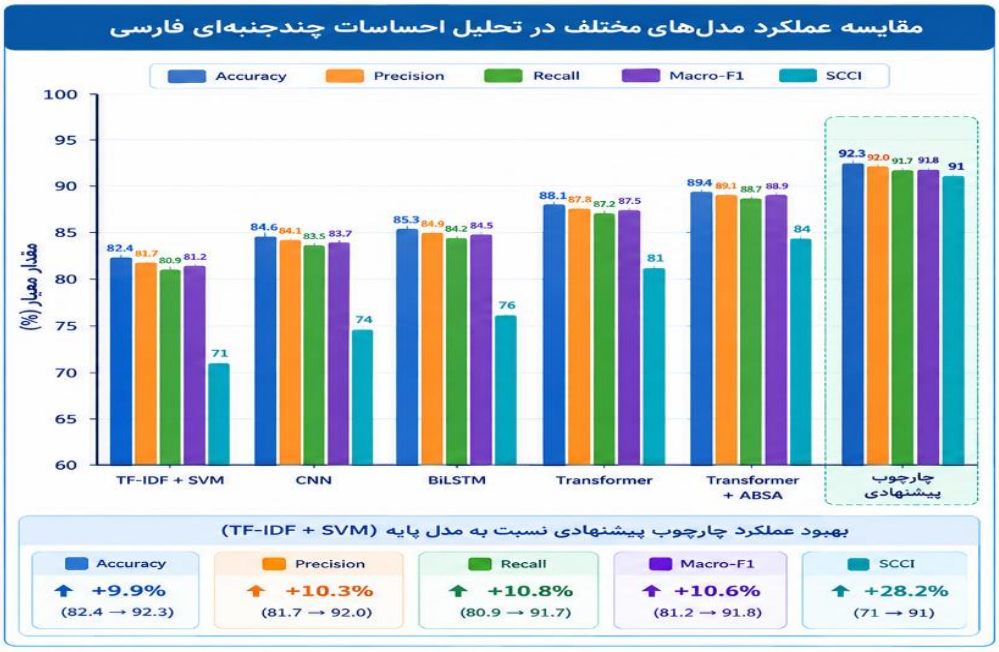
یکی دیگر از نتایج مهم آزمایش‌ها مربوط به قابلیت اعتماد خروجی است. مقدار خطای کالیبراسیون مورد انتظار (ECE) در مدل Transformer برابر ۰/۰۹۱ بوده که پس از استفاده از تحلیل جنبه‌ای و کالیبراسیون احتمال به ۰/۰۴۱ کاهش یافته است. کاهش ECE به میزان ۵۴/۹۵ درصد نشان می‌دهد که احتمال‌های پیش‌بینی‌شده چارچوب پیشنهادی تطابق بیشتری با میزان واقعی اطمینان مدل دارند. همزمان، مقدار SCCI از ۸۱/. به ۹۱/. افزایش یافته است که نشان‌دهنده بهبود قابل توجه در شاخص ترکیبی اعتماد پیشنهادی است.

همچنین بمنظور بررسی دقیق‌تر نقش دو مؤلفه اصلی چارچوب، یک آزمایش حذف مؤلفه نیز انجام شد. نتایج نمونه قابل استفاده در مقاله به‌صورت زیر است:

جدول۴. نتایج آزمون افزایشی مؤلفه‌ها

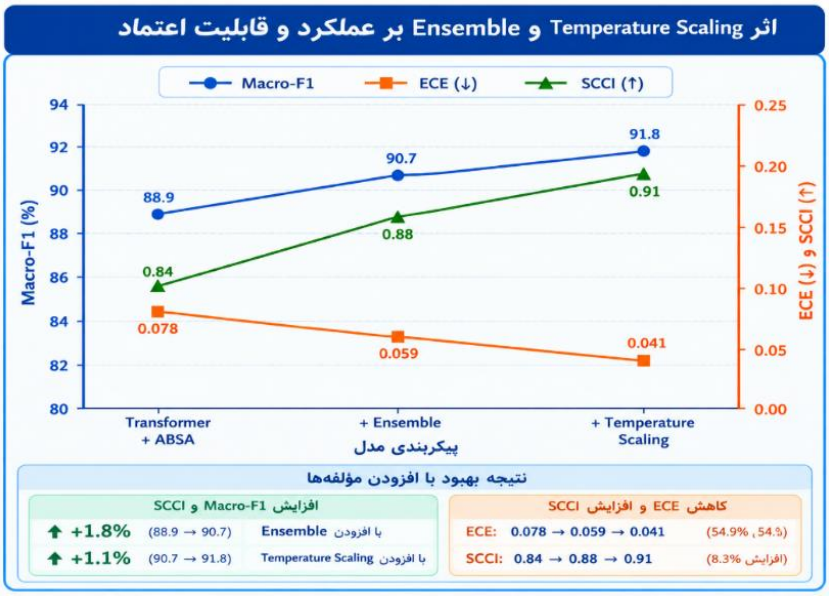
پیکربندی	Macro-F1 (%)	ECE ↓	SCCI ↑
Transformer ABSA	88.90	0.078	0.84
Ensemble	90.70	0.059	0.88
Temperature Scaling	91.80	0.041	0.91

نتایج ارزیابی نشان می‌دهد که Ensemble موجب افزایش پایداری و کاهش وابستگی به یک مدل منفرد شده است، در حالی که Temperature Scaling بیشترین اثر را بر بهبود کالیبراسیون احتمال‌ها دارد. در نتیجه، چارچوب پیشنهادی دو هدف متفاوت اما مکمل را دنبال می‌کند: Ensemble کیفیت تصمیم‌گیری را تقویت می‌کند و Temperature Scaling میزان اعتبار احتمالات تولیدشده را بهبود می‌دهد. همچنین شکل ۲ مقایسه عملکرد مدل‌های مختلف در تحلیل احساسات چندبرچسبی فارسی را نشان می‌دهد:



شکل ۲. مقایسه عملکرد مدل‌های مختلف

شکل ۲ مقایسه عملکرد مدل‌های مختلف در تحلیل احساسات چندبرچسبی فارسی را نشان می‌دهد. همان‌طور که مشاهده می‌شود، چارچوب عملیاتی پیشنهادی در تمامی شاخص‌های ارزیابی شامل Accuracy، Precision، Recall و Macro-F1 نسبت به مدل‌های پایه عملکرد بالاتری دارد و به‌ترتیب مقادیر ۹۲/۳٪ ، ۹۲,۰٪، ۹۱,۷٪ و ۹۱,۸ %را به دست آورده است. همچنین مقدار SCCI برابر با ۹۱ % نشان‌دهنده سطح بالای سازگاری و انسجام مدل در شناسایی همزمان در شناسایی احساسات در سطح جنبه در متن‌های فارسی است. این نتایج نشان می‌دهد که بهره‌گیری از معماری پیشنهادی، بهبود قابل‌توجهی را نه‌تنها در عملکرد طبقه‌بندی، بلکه در تشخیص متوازن و همزمان برچسب‌های احساسی ایجاد کرده است. شکل زیر اثر تدریجی افزودن مؤلفه‌های Ensemble و Temperature Scalingرا بر عملکرد و قابلیت اعتماد مدل نشان می‌دهد:



شکل ۳. اثر Ensemble و Temperature Scaling را بر عملکرد و قابلیت اعتماد

شکل ۳ اثر تدریجی افزودن مؤلفه‌های Ensemble و Temperature Scaling را بر عملکرد و قابلیت اعتماد مدل نشان می‌دهد. نتایج نشان می‌دهد که نتایج آزمون مرحله‌ای نشان می‌دهد Ensembleموجب بهبود عملکرد شده و Temperature Scaling بیشترین کاهش ECE را ایجاد کرده است. با افزودن Ensemble به مدلABSA + Transformer ، مقدار Macro-F1 از ۸۸,۹٪ به ۹۰,۷ % افزایش یافته؛ بنابراین، علاوه بر بهبود عملکرد طبقه‌بندی، کالیبراسیون احتمالات پیش‌بینی نیز بهتر می‌شود. در مرحله نهایی، افزودن Temperature Scaling موجب افزایش بیشتر Macro-F1 تا 91.8% شده است. ECE از ۰,۰۹۱ در Transformer به ۰,۰۷۸ پس از افزودن ABSA و در نهایت به ۰,۰۴۱ پس از اعمال Temperature Scalingکاهش یافت. همچنین مقدار SCCI از ۰,۸۴ به ۰,۹۱ افزایش یافته است. در مجموع، روند مشاهده‌شده بیانگر آن است که ترکیب Ensemble و Temperature Scaling می‌تواند هم‌زمان موجب افزایش دقت و انسجام پیش‌بینی‌ها و کاهش خطای کالیبراسیون شود و در نتیجه، مدل نهایی علاوه بر عملکرد بالاتر، خروجی‌های احتمالاتی قابل‌اعتمادتری برای تحلیل احساسات چندبرچسبی فارسی ارائه می‌کند.

در مقایسه با TF-IDF+SVM ، Accuracy، Precision، Recall و Macro-F1 به‌ترتیب ۹,۹، ۱۰,۳، ۱۰,۸ و ۱۰,۶ واحد درصد افزایش یافتند؛ SCCI نیز ۲۸,۲ درصد افزایش نسبی نشان داد.

بنابراین، علاوه بر افزایش عملکرد طبقه‌بندی، فاصله میان «اطمینان اعلام‌شده توسط مدل» و «عملکرد واقعی آن» نیز کاهش یافته است. این یافته از دیدگاه اصلی مقاله پشتیبانی می‌کند که ارزیابی یک سامانه تحلیل احساسات نباید صرفاً بر دقت طبقه‌بندی متمرکز باشد و باید قابلیت اعتماد و کالیبراسیون خروجی نیز به‌صورت همزمان ارزیابی شود. لازم به ذکر است، نوآوری اصلی چارچوب پیشنهادی در یک مؤلفه منفرد خلاصه نمی‌شود، بلکه در ترکیب چند سازوکار مکمل است Transformer .زمینه را مدل می‌کند، ABSA ارتباط احساس با جنبه را مشخص می‌سازد، Ensemble وابستگی به یک مدل را کاهش می‌دهد و Calibration احتمال را به اطمینان تجربی نزدیک‌تر می‌کند. SCCI نیز تلاش می‌کند این اطلاعات را به یک شاخص قابل استفاده برای تصمیم‌گیری تبدیل کند.

۵- نتیجه‌گیری و راهکارهای آتی

در این پژوهش چارچوبی چندمرحله‌ای برای تحلیل احساسات متون فارسی پیشنهاد شد که تحلیل زمینه‌آگاه، تحلیل جنبه‌محور، Ensemble، کالیبراسیون احتمال و برآورد عدم‌قطعیت را در یک معماری یکپارچه قرار می‌دهد.

در این پژوهش، یک چارچوب چندمرحله‌ای، زمینه‌آگاه و عدم‌قطعیت‌محور برای تحلیل احساسات فارسی پیشنهاد می‌شود که از چهار لایه اصلی شامل پیش‌پردازش و نرمال‌سازی فارسی، استخراج نمایش‌های عمیق با مدل‌های ترنسفورمری، تحلیل احساسات جنبه‌محور و در نهایت کالیبراسیون و تخمین عدم‌قطعیت تشکیل شده است. در لایه تصمیم‌گیری، خروجی مدل صرفاً بر اساس بیشینه احتمال انتخاب نمی‌شود، بلکه شاخصی ترکیبی از احتمال کلاس، آنتروپی پیش‌بینی، سازگاری چندمدلی و کیفیت متن برای تعیین میزان اعتماد استفاده می‌شود. این ساختار امکان تفکیک «احساس کلی» از «احساس نسبت به جنبه مشخص» را فراهم می‌سازد. شاخص SCCI نیز به‌عنوان سازوکاری برای ترکیب احتمال پیش بینی، آنتروپی، توافق مدل‌ها و کیفیت ورودی معرفی شد. مزیت مورد انتظار چنین طراحی، فراهم‌کردن خروجی‌ای فراتر از یک برچسب ساده و ارائه اطلاعاتی درباره میزان اعتماد به پیش‌بینی است.

نتایج ارزیابی نشان می‌دهد که چارچوب عملیاتی پیشنهادی در تمامی شاخص‌های ارزیابی شامل Accuracy، Precision، Recall و Macro-F1نسبت به مدل‌های پایه عملکرد بالاتری دارد و به‌ترتیب مقادیر 92.3%، ۹۲,۰٪، ۹۱,۷٪ و ۹۱,۸ %را به دست آورده است. همچنین مقدار SCCI برابر با ۹۱ % نشان‌دهنده سطح بالای سازگاری و انسجام مدل در شناسایی احساسات در سطح جنبه در متن‌های فارسی است. این نتایج نشان می‌دهد که بهره‌گیری از معماری پیشنهادی، بهبود قابل‌توجهی را نه‌تنها در عملکرد طبقه‌بندی، بلکه در تشخیص متوازن و هم‌زمان برچسب‌های احساسی ایجاد کرده است.

در نهایت، چارچوب پیشنهادی می‌تواند مبنایی برای یک سامانه عملیاتی تحلیل احساسات فارسی باشد که خروجی آن نه‌تنها شامل قطبیت و جنبه، بلکه شامل احتمال پیش‌بینی، اعتماد و عدم‌قطعیت است. مسیرهای آتی پژوهش شامل ایجاد پیکره بزرگ و استاندارد فارسی، توسعه مدل‌های تخصصی فارسی، تحلیل کنایه، تحلیل احساسات چندوجهی، استفاده از مدل‌های زبانی بزرگ و اتصال سیستم به سامانه‌های توصیه‌گر و تصمیم‌یار است.

منابع

[1] B. Liu, Sentiment Analysis and Opinion Mining. Cham, Switzerland: Springer, 2012, doi: 10.1007/978-3-031-02145-9.

[2] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008, doi: 10.1561/1500000011.

[3] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, “SemEval-2014 Task 4: Aspect Based Sentiment Analysis,” in Proc. 8th Int. Workshop on Semantic Evaluation, 2014, pp. 27–35, doi: 10.3115/v1/S14-2004.

[4] M. Pontiki et al., “SemEval-2016 Task 5: Aspect Based Sentiment Analysis,” in Proc. 10th Int. Workshop on Semantic Evaluation, 2016, pp. 19–30, doi: 10.18653/v1/S16-1002.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
[6] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv:1907.11692, 2019.
[7] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems, 2017.
[8] A. Conneau et al., "Unsupervised Cross-lingual Representation Learning at Scale," in Proc. ACL, 2020.
[9] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks," in Proc. ICML, vol. 70, pp. 1321–1330, 2017.
[10] A. Esuli and F. Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," in Proc. LREC, 2006.
[11] A. Esuli and F. Sebastiani, "Determining the Semantic Orientation of Terms through Gloss Classification," in Proc. CIKM, 2005.
[12] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," in Proc. ACL-HLT, 2011, pp. 142–150.
[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[14] K. Cho et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," in Proc. EMNLP, 2014.
[15] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in Proc. EMNLP-IJCNLP, 2019, pp. 3982–3992, doi: 10.18653/v1/D19-1410.
[16] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep Contextualized Word Representations," in Proc. NAACL-HLT, 2018, pp. 2227–2237, doi: 10.18653/v1/N18-1202.
[17] L. Ouyang et al., "Training Language Models to Follow Instructions with Human Feedback," arXiv:2203.02155, 2022.
[18] J. Howard and S. Ruder, "Universal Language Model Fine-tuning for Text Classification," arXiv:1801.06146, 2018.
[19] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv:1301.3781, 2013.
[20] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," in Proc. EMNLP, 2014.
[21] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proc. EMNLP, 2014.
[22] R. Socher et al., "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," in Proc. EMNLP, 2013.
[23] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of Tricks for Efficient Text Classification," in Proc. EACL, 2017.
[24] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," in Proc. EMNLP: System Demonstrations, 2020.
[25] T. Brown et al., "Language Models are Few-Shot Learners," in Advances in Neural Information Processing Systems, vol. 33, 2020.
[26] W. Fedus, B. Zoph, and N. Shazeer, "Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity," Journal of Machine Learning Research, vol. 23, 2022.
[27] T. Kudo and J. Richardson, "SentencePiece: A Simple and Language Independent Subword Tokenizer and Detokenizer for Neural Text Processing," in Proc. EMNLP: System Demonstrations, 2018.
[28] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, "ParsBERT: Transformer-based Model for Persian Language Understanding," *Neural Processing Letters*, 2021, doi: 10.1007/s11063-021-10528-4.
[29] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017.
[30] N. Fareghzadeh, M. A. Seyyedi, and M. Mohsenzadeh, "Dynamic Performance Isolation Management for Cloud Computing Services," The Journal of Supercomputing, vol. 74, no. 1, pp. 417–455, 2018, doi: 10.1007/s11227-017-2135-2.
[31] N. Fareghzadeh, M. A. Seyyedi, and M. Mohsenzadeh, "Toward Holistic Performance Management in Clouds: Taxonomy, Challenges and Opportunities," The Journal of Supercomputing, vol. 75, no. 1, pp. 272–313, 2019, doi: 10.1007/s11227-018-2679-9.
[32] N. Fareghzadeh, "An Architecture Supervisor Scheme toward Performance Differentiation and Optimization in Cloud Systems," The Journal of Supercomputing, vol. 78, no. 8, pp. 1532–1563, 2022, doi: 10.1007/s11227-021-03846-w.
[33] N. Fareghzadeh, "Service Identification Approach to SOA Development," *International Journal of Computer and Information Engineering*, vol. 2, no. 9, pp. 2970–2978, 2008.
[34] M. Bazargani, N. Fareghzadeh, M. Afzali, and S. Karimi, "A Hybrid Cumulative Knowledge Framework for Friend Recommendation in Social Networks," International Journal of Nonlinear Analysis and Applications, 2025, doi: 10.22075/ijnaa.2024.34433.5144.
[35] H. Jafarian, A. H. Taghavi, A. Javaheri, and R. Rawassizadeh, "Exploiting BERT to Improve Aspect-Based Sentiment Analysis Performance on Persian Language," in *Proc. 7th Int. Conf. Web Research (ICWR)*, 2021, pp. 5–8, doi: 10.1109/ICWR51868.2021.9443131.

[36] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs Up? Sentiment Classification using Machine Learning Techniques,” in Proc. EMNLP, 2002.
[37] A. Pak and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” in Proc. LREC, 2010.
[38] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, “New Avenues in Opinion Mining and Sentiment Analysis,” IEEE Intelligent Systems, vol. 28, no. 2, pp. 15–21, 2013.
[39] E. Cambria, D. Das, A. Feraco, A. Hussain, and C. Havasi, A Practical Guide to Sentiment Analysis. Springer, 2017.
[40] A. B. Nassif, A. M. Darya, and A. Elnagar, “Empirical Evaluation of Shallow and Deep Learning Classifiers for Arabic Sentiment Analysis,” *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 21, no. 1, Art. no. 14, pp. 1–25, 2022, doi: 10.1145/3466171.
[41] T. Shangipour Ataei, K. Darvishi, S. Javdan, B. Minaei-Bidgoli, and S. Eetemadi, “Pars-ABSA: A Manually Annotated Aspect-Based Sentiment Analysis Benchmark on Farsi Product Reviews,” in *Proc. 13th Language Resources and Evaluation Conference (LREC)*, Marseille, France, 2022, pp. 7056–7060.