تجزیه و تحلیل احساسات چندوجهی با استفاده از تکنیک‌های یادگیری انتقال چندگانه ترکیبی با به کارگیری شبکه‌های کانولوشنی وزن‌دار ترکیبی

علیرضا قربانعلی[۱]، محمدکریم سهرابی[۱،*] و فرزین یغمایی[۱،۲]

اطلاعات مقاله

نوع مقاله: پژوهشی
دریافت مقاله: ۱۴۰۰/۰۸/۲۳
بازنگری مقاله: ۱۴۰۱/۰۴/۰۶
پذیرش مقاله: ۱۴۰۱/۰۷/۰۴

واژگان کلیدی:
تجزیه و تحلیل احساسات چندوجهی،
یادگیری عمیق،
یادگیری انتقال،
مکانیزم توجه،
شبکه کانولوشنی وزن دار.

چکیده

تحلیل نظرات کاربران شبکه‌های اجتماعی در رابطه با موضوعات مختلف می‌تواند منجر به درک صحیح موقعیت، نگرش و نظر آنها نسبت به این موضوعات گردد. احساسات مستتر در نظرات، بازخوردها یا انتقادات، شاخص‌های مفیدی برای اهداف گوناگون هستند و آنها را در دسته‌های منفی، مثبت و خنثی طبقه‌بندی می‌کنند. تجزیه و تحلیل احساسات یکی از مسائل حوزه پردازش زبان طبیعی است که کاربردهای متنوعی دارد. دسته‌ای از نظرات کاربران شبکه‌های اجتماعی به صورت چندوجهی و در قالب ترکیبی از چند رسانه مانند متن، تصویر و صورتک به اشتراک گذاشته می‌شود که منجر به ارائه و درک بهتر احساسات می‌شوند. در این مقاله یک روش ترکیبی یادگیری انتقال با استفاده از پنج مدل از پیش آموزش دیده شده و شبکه‌های کانولوشنی ترکیبی برای تجزیه و تحلیل احساسات چندوجهی ارائه می‌شود. در این رویکرد برای استخراج ویژگی‌های تصاویر از دو مدل از پیش آموزش دیده مبتنی بر شبکه‌های کانولوشنی، و برای استخراج ویژگی‌های متون و تعبیه کلمات از سه مدل از پیش آموزش دیده دیگر استفاده می‌شود. ویژگی‌های استخراج شده توسط این مدل‌ها توسط شبکه‌های کانولوشنی مورد استفاده قرار گرفته و مکانیزم‌های توجه بصری و توجه چندراسی، به ترتیب برای تمرکز بر روی مهمترین نواحی احساسی تصاویر و برجسته‌سازی کلمات دارای احساس در متون به کار گرفته می‌شوند. دو نوع همجوشی زودهنگام (در سطح ویژگی‌ها) و دیرهنگام (در سطح تصمیم‌گیری) نیز در روش ارائه شده مورد استفاده قرار می‌گیرند. ویژگی‌های استخراج شده متون و تصاویر به صورت نرمال‌سازی شده در همجوشی زودهنگام ترکیب و طبقه‌بندی می‌شوند و قطبیت و برچسب نهایی نیز با ترکیب نتایج حاصل از طبقه‌بندی تصاویر و متون، در قالب همجوشی دیرهنگام و با استفاده از تکنیک رای‌گیری تعیین می‌شود. نتایج حاصل از آزمایشات تجربی مدل پیشنهادی بر روی مجموعه داده استاندارد، دقت مطلوب ۹۶٪ را نشان می‌دهد.

* پست الکترونیک نویسنده مسئول: Amir_sohraby@aut.ac.ir
۱. دانشکده مهندسی کامپیوتر، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران
۲. دانشکده مهندسی کامپیوتر ، دانشگاه سمنان

https://doi.org/10.22075/jme.2022.25099.2173

۱- مقدمه

تجزیه و تحلیل احساسات یکی از مسائل پردازش زبان طبیعی است که در حوزه‌های مختلفی مانند پیش‌بینی قیمت سهام [۱]، پزشکی [۲]، پیش‌بینی [۳]، گردشگری [۴] و صنعت [۵] کاربرد دارد. تجزیه و تحلیل احساسات می‌تواند یک ابزار قدرتمند برای تعیین افکار عمومی نیز باشد [۶]. اغلب مردم در شبکه‌های اجتماعی پیام‌های متنی خود را همراه با تصاویر یا ویدیوها به اشتراک می‌گذارند که این عمل باعث به وجود آمدن حجم عظیمی از داده‌های چندوجهی می‌گردد. به اشتراک‌گذاری نگرش‌ها و نظرات به صورت ترکیبی از متون و تصاویر معمولا به ارائه بهتر احساسات آنها می‌انجامد و تجزیه و تحلیل این داده‌های چندوجهی به درک کامل‌تر و دقیق‌تر این نگرش‌ها و نظرات مردم کمک می‌کند [۷]. به طور کلی داده‌های چندوجهی در مقایسه با داده‌های تک‌وجهی حاوی اطلاعات بیشتر و مفیدتری هستند و تشخیص واقعی‌تر احساس کاربران را در پی دارند. تحقیقات مختلفی با استفاده از روش‌ها و معماری‌های گوناگون برای تجزیه و تحلیل احساسات چندوجهی[1] انجام شده است. در کنار مطالعاتی که از ترکیب شبکه‌های عصبی عمیق استفاده نموده‌اند [۷ و ۸] و روش‌هایی که با استفاده از تکنیک‌های یادگیری ترکیبی[2] به تجزیه وتحلیل احساسات پرداخته‌اند [۹]، دسته‌ای دیگر از روش‌ها وجود دارند که از یادگیری انتقال[3] برای انجام طبقه بندی متون و تصاویر بهره برده‌اند [۱۰]. یک روش یادگیری انتقال ویژگی مبتنی بر خودرمزنگاری در [۱۱] ارائه شده است. هدف این مقاله، ارائه یک مدل ترکیبی یادگیری انتقال چندگانه مبتنی بر شبکه‌های کانولوشنی وزن‌دار تقویت شده[4] برای تجزیه و تحلیل احساسات چندوجهی است. در این روش از دو مرحله همجوشی زودهنگام و دیرهنگام برای ترکیب ویژگی‌های استخراج شده و ترکیب نتایج طبقه‌بندها استفاده گردیده است. در روش پیشنهادی، ابتدا ویژگی‌های تصاویر ورودی به صورت جداگانه توسط مدل‌های از پیش آموزش دیده VGG19 و MobileNetV3، برای استخراج بر روی مجموعه داده مورد نظر تنظیم دقیق[5] می‌گردند. برای استخراج ویژگی‌های متون و تعبیه کلمات نیز از مدل‌های از پیش آموزش دیده BERT، Glove و Cove به صورت همجوشی شده استفاده می‌شود. سپس ویژگی‌های استخراج شده به صورت نرمال شده (تسطیح شده) در یک بردار ویژگی ترکیب می‌شوند. از این ویژگی‌ها در شبکه‌های کانولوشنی وزن‌دار تقویت شده استفاده شده و سپس با استفاده از مکانیزم توجه بصری[6] به تمرکز بر روی مهمترین نواحی احساسی تصاویر، و مکانیزم توجه چندراسی[7] به برجسته‌سازی کلمات دارای احساس در متون پرداخته می‌شود. در مرحله بعد نتایج حاصل از طبقه‌بندی تصاویر و متون به صورت جداگانه و با استفاده از تکنیک رای‌گیری ترکیب شده، و همجوشی دیرهنگام[8] برای تعیین قطبیت نهایی به کار گرفته می‌شود.

به طور کلی، جنبه‌ها و ویژگی‌های اصلی به کار گرفته شده در رویکرد این مقاله به شرح زیر است:

- استفاده از چند تکنیک یادگیری انتقال تنظیم شده برای استخراج بهترین ویژگی‌های متون و تصاویر
- استفاده از مکانیزم‌های توجه بصری و توجه چندراسی برای تمرکز بر روی نواحی احساسی تصویر و کلمات احساسی متن
- استفاده از تکنیک‌های یادگیری ترکیبی مبتنی بر شبکه‌های کانولوشنی و تکنیک رای‌گیری برای انجام طبقه‌بندی
- استفاده از همجوشی زودهنگام برای ترکیب ویژگی‌های استخراج شده از متون و تصاویر
- استفاده از همجوشی دیرهنگام برای ترکیب نتایج و تشخیص برچسب نهایی احساسات
- آزمایش روش پیشنهادی بر روی مجموعه داده MVSA-Multiple و مقایسه و نمایش برتری دقت نتایج به دست آمده نسبت به سایر روش‌های موجود

ادامه مطالب مقاله به این صورت سازمان‌دهی شده‌اند: در بخش ۲ به مرور پیشینه و کارهای مرتبط پرداخته می‌شود. بخش ۳ روش انجام کار و بخش ۴ نتایج پیاده‌سازی و مقایسه با سایر روش ها را نشان می‌دهد. بخش ۵ به جمع‌بندی مقاله می‌پردازد.

[1] Multimodal sentiment analysis
[2] Ensemble learning
[3] Transfer learning
[4] Weighted convolutional neural networks ensemble (WCNNE)
[5] Fine-tuned
[6] Visual attention mechanism
[7] Multi-head attention
[8] Late fusion

۲- کارهای مرتبط

در این بخش به مرور کارهای انجام شده در زمینه تحلیل احساسات متنی، تصویری و چندوجهی می‌پردازیم. همچنین مروری کوتاه بر شبکه‌های عصبی کانولوشنی، یادگیری ترکیبی و تکنیک‌های یادگیری انتقال خواهیم داشت.

۲-۱- تحلیل احساسات متنی

تحلیل احساسات متنی یکی از مسائل حوزه پردازش زبان طبیعی است که به دو روش مبتنی بر واژگان و یادگیری ماشین انجام می‌شود. روش‌های یادگیری ماشین به دسته‌های نظارت شده، نیمه نظارت شده و نظارت نشده تقسیم می‌شوند. یک مدل ترکیبی از تکنیک‌های نظارت شده و نظارت نشده برای تحلیل احساسات در [۱۲] ارائه شده است. یک مدل مبتنی بر گراف در [۱۳] پیشنهاد شده که اطلاعات هشتگ‌ها را برای طبقه‌بندی احساسات ترکیب می‌کند. با توسعه شبکه-های عصبی عمیق، روش‌های مبتنی بر این نوع یادگیری نیز نتایج مطلوبی در طبقه‌بندی متون کسب کرده‌اند. به عنوان مثال پارامترهای شبکه‌های کانولوشنی در [۱۴] با استفاده از الگوریتم ژنتیک تنظیم شده و برای طبقه‌بندی متون به کار گرفته شده است. مطالعاتی نیز با استفاده از شبکه‌های عصبی یادآور[1] برای طبقه‌بندی متون انجام شده است [۱۵]. استفاده از مدل‌های از پیش آموزش دیده جهت استخراج ویژگی‌های متون و تعبیه کلمات نیز به دلیل بهبود دقت و افزایش سرعت طبقه‌بندی مورد استقبال محققین قرار گرفته است. به عنوان مثال در کارهای [۱۰ و ۱۶] از مدل از پیش آموزش دیده BERT استفاده شده است.

۲-۲- تحلیل احساسات بصری

تحلیل احساسات تصاویر نسبت به متون دارای پیچیدگی‌های بیشتری است زیرا تصاویر انتزاعی و ذهنی‌تر هستند. برای تجزیه و تحلیل احساسات تصاویر معمولاً سه نوع ویژگی بصری مورد استفاده قرار می‌گیرد که عبارت است از ویژگی های سطح پایین [۱۷]، سطح متوسط [۱۸] و سطح بالا [۱۹]. از مدل پیوسته کلمات[2] در [۲۰] برای استخراج اطلاعات متنی و از رمزگذار خودکار[3] برای به دست آوردن ویژگی‌های بصری قوی در پیام‌های کوتاه توییتر استفاده شده است. از مکانیزم‌های مختلف توجه نیز برای پوشش نواحی تصویر با هدف کشف احساسات استفاده شده است [۲۱]. در [۲۲] برخلاف برخی روش‌های موجود که احساسات یا عواطف را مستقیماً از ویژگی‌های سطح پایین بصری استنتاج می کنند، یک رویکرد جدید مبتنی بر درک مفاهیم بصری که به شدت با احساسات مرتبط هستند پیشنهاد شده است. این رویکرد به طور خودکار یک هستان‌شناسی احساسات بصری در مقیاس بزرگ[4] (VSO) متشکل از بیش از ۳۰۰۰ جفت اسم صفت (ANP) ایجاد کرد.

۲-۳- تحلیل احساسات چندوجهی

از آنجا که کاربران شبکه‌های اجتماعی معمولا نظرات خود را به صورت ترکیبی از متن، تصویر و یا صورتک به اشتراک می-گذارند، شبکه‌های اجتماعی به طور مداوم در حال تولید حجم بسیار بالایی از داده‌های چندوجهی هستند. تحلیل احساسات تک‌وجهی بر روی متن یا تصویر ممکن است برای درک دقیق و کامل احساسات کافی نباشد. برای درک بهتر احساسات، از تجزیه و تحلیل چندوجهی احساسات استفاده می‌شود. در [۷] از ترکیب شبکه‌های کانولوشنی و حافظه کوتاه مدت طولانی[5] به همراه مکانیزم توجه برای تحلیل احساسات چندوجهی استفاده شده است. تحلیل احساسات چندوجهی متن و عکس در [۲۳] با استفاده از conventional SentiBank و شبکه‌های از پیش آموزش دیده کانولوشنی انجام شده است. پنج مدل کانولوشنی از پیش آموزش دیده نیز برای استخراج ویژگی‌های تصاویر و طبقه‌بندی آنها در این مقاله مورد استفاده قرار گرفته و از مدل word2vec برای استخراج ویژگی‌های متنی و تعبیه کلمات استفاده شده است. یک مدل گرافیکی احتمالی برای به دست آوردن میزان همبستگی بین داده‌های متنی و تصویری در داده‌های فلیکر[6] در [۲۴] پیشنهاد شده است. در [۲۵] برای تجزیه و تحلیل احساسات چندوجهی از شبکه‌های کانولوشنی به صورت ترکیبی استفاده شده و در برای تعبیه کلمات و استخراج ویژگی‌های متن، مدل BERT به کار گرفته شده است. در روش پیشنهادی این مقاله، نتایج حاصل از

[1] Recurrent neural network (RNN)
[2] Continuous bag-of-words
[3] Auto-encoder
[4] Visual Sentiment Ontology
[5] Long short-term memory (LSTM)
[6] Flickr

طبقه‌بندی تصاویر و متون با استفاده از تکنیک رای‌گیری مشخص می‌شود و برچسب نهایی قطبیت با استفاده از مکانیزم همجوشی مشخص می‌گردد. در [۱۰] از تکنیک Adaboost و مدل تنظیم دقیق شده VGG16 برای تجزیه و تحلیل احساسات چندوجهی استفاده شده است. یک مدل شبکه عصبی ادغام شده[1] (MNN) در [۲۶] پیشنهاد شده است که از شبکه CNN برای استخراج ویژگی‌های متن و تصویر استفاده می‌کند و دو استراتژی ادغام‌شده، با نام‌های EarlyRMNN و Late-RMNN را برای دریافت ویژگی‌های عمیق‌تر و متمایزتر به کار می‌گیرد. یک چارچوب سلسله مراتبی مبتنی بر مکانیزم توجه برای تحلیل احساسات چندوجهی نیز در [۲۷] ارائه شده است. این ساختار سلسله مراتبی با به کارگیری LSTM از متن و تصویر برای استخراج ویژگی معنایی بصری به عنوان اطلاعات اضافی برای متن در تجزیه و تحلیل احساسات چندوجهی استفاده کرده است. رابطه متقابل اطلاعات بصری و متنی در [۲۸] در نظر گرفته شده و یک شبکه حافظه مشترک جدید برای مدل‌سازی تکراری تعاملات بین محتوای بصری و کلمات متنی برای تحلیل احساسات چندوجهی پیشنهاد شده است. در [۲۹] یک مکانیسم ترکیب اطلاعات تعاملی برای یادگیری تعاملی بازنمایی‌های متنی خاص بصری و بازنمایی‌های بصری خاص متنی پیشنهاد شده است. همچنین، یک مکانیزم استخراج اطلاعات برای استخراج اطلاعات معتبر و فیلتر کردن داده‌های اضافی برای نمایش‌های متنی و بصری خاص ارائه شده است. یک مدل تجزیه و تحلیل احساسات چندوجهی بر اساس شبکه توجه چندنمایی[2] (MVAN) نیز در [۳۰] پیشنهاد شده است.

۲-۴- همجوشی

همجوشی داده‌های چندوجهی، ویژگی‌های چندین منبع داده را برای پیش‌بینی مقدار کلاس نهایی با هم ترکیب می‌کند. روش‌های اصلی همجوشی برای تجزیه و تحلیل احساسات چندوجهی، همجوشی زودهنگام[3] [۳۱ و ۳۲]، همجوشی میان‌مدت[4] [۳۳ و ۳۴] و همجوشی دیرهنگام [۳۴] هستند. همجوشی زود هنگام یا اولیه منجر به تولید بردارهای بزرگ ورودی می‌شود و ممکن است افزونگی نیز داشته باشد. در این نوع همجوشی، ویژگی‌ها با استفاده از مدل‌های شبکه‌های عصبی عمیق استخراج می‌شوند. ویژگی‌های ورودی به هم متصل شده و وارد طبقه‌بند می‌شوند [۲۵]. در [۱۰] از نظریه احتمال توسعه یافته دمپستر-شفر برای همجوشی نتایج در سطح تصمیم‌گیری استفاده شده است، ولی در این مقاله، همجوشی را در دو سطح ویژگی‌ها و نتایج انجام شده است. همجوشی در سطح ویژگی‌ها کمک شایانی به تشخیص ویژگی‌های مهم و مشترک می‌نماید.

۲-۵- شبکه‌های عصبی کانولوشنی

شبکه‌های عصبی مصنوعی، به عنوان یکی از زیرمجموعه‌های هوش مصنوعی دارای مزایای زیادی می‌باشند که در کارهایی از قبیل افزایش سرعت فرایند مدل‌سازی، توانایی مدل‌سازی هم‌زمان بسیار مفید هستند [۳۵]. شبکه‌های عصبی کانولوشنی، شبکه‌های عصبی سلسله‌مراتبی هستند که در سال ۱۹۹۰ معرفی شدند. از این شبکه‌ها، که از پرکاربردترین معماری‌های یادگیری عمیق برای پردازش و تشخیص تصویر هستند، برای پردازش زبان طبیعی نیز می‌توان استفاده کرد [۳۶]. شبکه‌های کانولوشنی عمیق برای طبقه‌بندی متون نیز مورد استفاده قرار گرفته‌اند. در [۳۷] یک معماری ترکیبی با استفاده از شبکه‌های کانولوشنی و Bi-GRU برای تجزیه و تحلیل احساسات متنی ارائه شده است. در روش فوق از مکانیزم توجه برای کشف ویژگی‌های مناسب‌تر استفاده شده است. در [۳۸] یک روش ترکیبی برای تحلیل احساسات با شبکه‌های CNN-LSTM بر روی مجموعه داده‌های زبان عربی ارائه شده که در آن ویژگی‌های محلی با استفاده از شبکه کانولوشنی استخراج، و برای حفظ وابستگی‌های متون از ۲ لایه LSTM استفاده شده است. دراین مقاله از این شبکه‌های کانولوشنی به صورت ترکیبی استفاده شده تا دقت بالاتری به دست آید.

۲-۶- یادگیری انتقال

یادگیری انتقال به معنای انتقال پارامترهای یک شبکه عصبی آموزش دیده بر روی یک مجموعه داده و استفاده مجدد از آن برای مساله دیگر با مجموعه داده و وظیفه دیگر است. موارد

[1] Merged Neural Network
[2] Multi-view Attentional Network
[3] Early Fusion
[4] Intermediate Fusion

متعددی نیز وجود دارد که یادگیری انتقال در آنها برای حل مسائل دنیای واقعی با الگوریتم‌های دیگر ترکیب می‌شود. لایه‌های اولیه شبکه‌های عمیق ویژگی‌هایی را یاد می‌گیرند که مختص مجموعه داده یا کار خاصی نیست و بنابراین قابل استفاده برای بسیاری از مجموعه‌های داده و وظایف است [۱۰]. اخیرا استفاده از مدل‌های از پیش آموزش دیده برای استخراج ویژگی‌های تصاویر و متون مورد توجه قرار گرفته و مطالعات مختلفی نیز در این زمینه انجام شده است. در [۷] از مدل VGG19 برای استخراج ویژگی‌های تصاویر و در [۳۹] از مدل MobileNetV2 برای تشخیص قطبیت تصاویر استفاده کرده‌اند. در پژوهش‌های [۱۶ و ۴۰] از مدل BERT برای استخراج ویژگی‌ها و تعبیه کلمات استفاده شده است. در این مقاله از ترکیب سه مدل از پیش آموزش دیده و تنظیم دقیق شده برای استخراج ویژگی‌های تصاویر و سه مدل از پیش آموزش دیده برای استخراج ویژگی‌های متون و تعبیه کلمات در کنار شبکه‌های کانولوشنی ترکیبی و مکانیزم توجه استفاده می‌شود.

۲-۷- شبکه‌های عصبی ترکیبی یا تقویت شده

یکی از رویکردهای حل مسائل پیچیده استفاده از رویکرد یادگیری ترکیبی به جای استفاده از شبکه‌های عصبی واحد است. در این روش، به کارگیری ترکیبی از چندین مدل یادگیری به جای استفاده منفرد از یک مدل یادگیری، توان تخمین مدل نهایی را بالا می‌برد. در الگوریتم‌های یادگیری ترکیبی، یک نمونه توسط چندین طبقه‌بند مختلف طبقه‌بندی می‌شود و نتایج طبقه‌بندی‌ها به روش‌های مختلفی با یکدیگر ترکیب شده و نتیجه نهایی برای آن نمونه خاص تعیین می‌گردد. در حوزه یادگیری عمیق، پیش‌بینی‌های انجام شده توسط چند مدل شبکه عصبی عمیق برای کاهش واریانس[1] پیش‌بینی‌ها و کاهش خطای تعمیم[2] ترکیب می‌شوند و باعث بهبود دقت[3] پیش‌بینی می‌گردند. از تکنیک‌های معروف یادگیری ترکیبی می‌توان به بسته‌بندی[4]، تقویت[5] و رای گیری[6] [۴۱] اشاره کرد. یک معماری ترکیبی از شبکه های کانولوشنی برای طبقه‌بندی تصاویر در [۴۲] معرفی شده است که برای هر طبقه‌بند با توجه به نتایج تولید شده در زمان آموزش یک وزن در نظر می‌گیرد. این کار باعث می‌شود آن شبکه عملکرد بهتری در زمان طبقه‌بندی داشته باشد. تحقیقات دیگری نیز با استفاده از شبکه‌های کانولوشنی ترکیبی در حوزه‌های مختلف از جمله تشخیص اشیا [۴۳]، تخمین [۴۴]، تحلیل احساسات [۴۵]، صدا [۴۶]، تشخیص بیماری [۴۷]، و پیش‌بینی سهام [۴۸] انجام شده است. در روش پیشنهادی این مقاله برای ترکیب نتایج شبکه‌های کانولوشنی از تکنیک رای‌گیری استفاده شده و ویژگی‌های ترکیب شده نیز با یک لایه کاملا متصل طبقه‌بندی می‌شوند.

۲-۸- مکانیزم توجه[7]

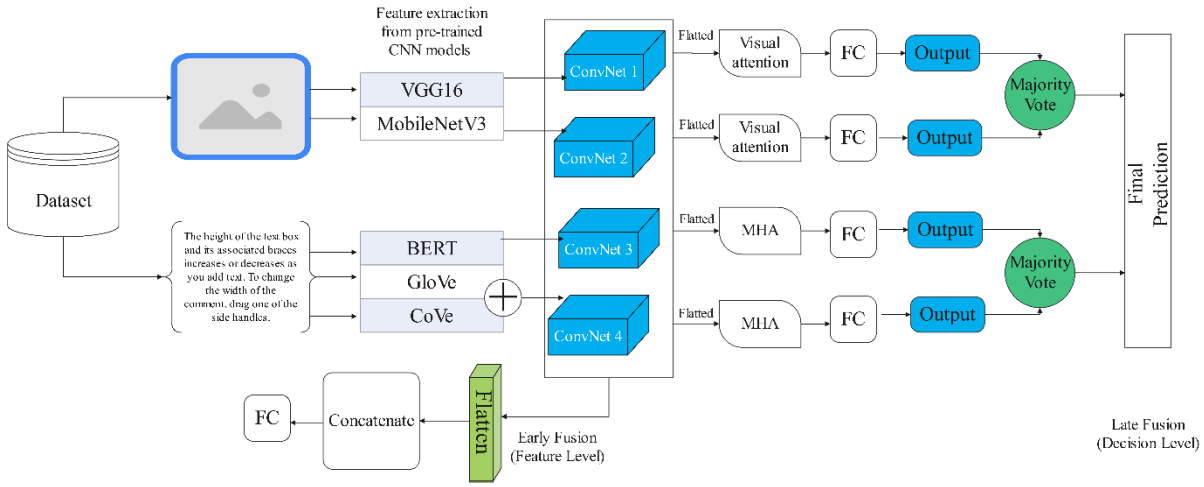
در سال‌های اخیر مکانیزم توجه برای حل چالش‌های حیطه پردازش زبان طبیعی و بینایی ماشین مورد استفاده قرار گرفته است. مکانیزم توجه قادر به یادگیری نگاشت بین ورودی‌های مختلف در یک زمینه معین است و می‌تواند فاصله بین وجه-های مختلف را پر کند. در [۲۳] از مکانیزم توجه برای کشف مهمترین ویژگی‌های احساسی استفاده شده است. مکانیزم توجه در مطالعات [۴۹] برای افزایش تأثیر جنبه‌های استخراج جهت پیش بینی مورد استفاده قرار گرفته است.

۳- روش پیشنهادی

رویکرد پیشنهادی این مقاله استفاده از چند مدل یادگیری انتقال به صورت ترکیبی مبتنی بر شبکه‌های کانولوشنی گروهی با استفاده از مکانیزم توجه است. در این رویکرد همجوشی زودهنگام در سطح ویژگی‌ها و همجوشی دیرهنگام در سطح تصمیم‌گیری و برای تعیین قطبیت نهایی استفاده شده است. شکل (۱) مدل پیشنهادی را نشان می‌دهد. مدل پیشنهادی، تصاویر و متون را به عنوان ورودی دریافت می‌کند. در این مدل تعبیه کلمات با استفاده از مدل‌های از پیش آموزش دیده BERT، Glove و Cove انجام گردیده است و برای استخراج ویژگی‌های تصاویر از مدل‌های از پیش آموزش دیده VGG16 و MobileNetV3، که بر روی مجموعه داده مورد نظر تنظیم دقیق شده‌اند، استفاده گردیده است. ویژگی‌های استخراج شده متون و تصاویر به صورت جداگانه در یک بردار

[1] Variance
[2] Generalization error
[3] Accuracy
[4] Bagging
[5] Baoosting
[6] Voting
[7] Attention

ویژگی ترکیب و طبقه‌بندی می‌شوند. از آنجا که مکانیزم توجه این امکان را برای شبکه‌های عصبی عمیق فراهم می‌کند که هم بر روی بخش‌های مهم متن و هم بر روی نواحی مهم تصاویر تمرکز کنند، از دو نوع مکانیزم توجه، به نام‌های توجه بصری و توجه چندراسی، به ترتیب برای تمرکز بر روی نواحی مهم متن و تصویر استفاده شده و لایه کاملا متصل[1] نیز به کارگیری شده است. خروجی‌های حاصل از طبقه‌بندهای گروهی با روش رای‌گیری برای تعیین قطبیت ترکیب می شوند و در آخر برای تعیین برچسب نهایی از همجوشی دیرهنگام استفاده می‌شود.



شکل ۱. مدل پیشنهادی ترکیبی برای تحلیل احساسات چندوجهی

۳-۱- آماده سازی داده‌ها

از آنجا که داده‌های مورد بررسی در این مقاله به صورت متن و تصویر هستند، آماده‌سازی آنها به دو بخش تقسیم می‌شود.

۳-۱-۱- آماده سازی داده های متنی

اولین مرحله از آماده‌سازی داده‌های متنی، پیش‌پردازش آنها و پاک‌سازی آن اشکالات و موارد نامطلوب است. مجموعه داده‌های مورد استفاده در این مطالعه، MVSA-Multiple [۲۳] است که دارای برخی داده‌های خام با نام‌های اختصاری، اشتباهات املایی و نمادهای نامطلوب (مانند !، @، &، #) است. داده‌های خام برای پاک‌سازی از نمادهای نامطلوب پیش‌پردازش می‌شوند و در فرایند پیش‌پردازش، نمادهای نامطلوب و اعداد حذف می‌شوند.

گام بعدی آماده‌سازی داده‌های متنی، استخراج ویژگی‌های[2] آنها است. در روش پیشنهای، ویژگی‌ها به طور خودکار با استفاده از مدل‌های از پیش آموزش دیده و شبکه‌های کانولوشنی استخراج می‌شوند. در این مرحله ویژگی‌های نامرتبط یا غیرضروری نادیده گرفته می‌شوند تا ابعاد داده‌ها به حداقل برسد و به بهبود دقت و کاهش زمان پردازش برای ورودی‌های بزرگ منجر شود.

تعبیه کلمات فاز بعدی آماده‌سازی داده‌های متنی است که توسط لایه‌های تعبیه[3] انجام می‌شود. لایه‌های تعبیه، متن کاربر را به بردارها تبدیل نموده و داده‌های متنی ورودی را به اعداد صحیح تبدیل می‌کنند. در این مقاله از روش‌های تعبیه کلمات Glove و Cove و مدل بازنمایی پویای BERT استفاده شده است.

- **مدل BERT**

مدل از پیش آموزش دیده BERT، که توسط گوگل در سال ۲۰۱۸ ارایه شده است، به عنوان یک مدل بازنمایی پویا قادر است بازنمایی متن را از هر دو سمت چپ و راست یاد بگیرد [۵۰]. مدل BERT را می‌توان با افزودن یک لایه خروجی تنظیم دقیق، برای مسایل پردازش زبان طبیعی و طبقه‌بندی استفاده کرد. به طور کلی با استفاده و تنظیم دقیق مدل‌های

[1] Fully-connected
[2] Feature extraction
[3] Embedding layer

از پیش آموزش دیده بر روی داده‌های متنی می‌توان در زمان کوتاه‌تر به دقت مناسب دست یافت. مدل اصلی BERT با ترکیبی از متون با ۳,۳ میلیارد کلمه بر روی ۱۶ عدد پردازشگر تنسور[1] یا TPU به مدت حدود ۴ روز آموزش دیده است. این در حالی است که اکثر مدل‌های از پیش آموزش دیده را می‌توان با تنظیم دقیق، در حدود یک تا چند ساعت با استفاده از یک GPU اجرا نمود.

برای پیاده‌سازی این فرآیند مراحل زیر مورد نیاز است:

۱. انتخاب یک مدل از پیش آموزش دیده BERT با توجه به نیازهای زبانی، که مدل انتخاب شده برای این مقاله، uncased_L-12_H-768_A-12 BERT است.
۲. اصلاح معماری مدل از پیش آموزش دیده متناسب با کاربرد.
۳. آماده‌سازی داده‌های آموزشی.
۴. تنظیم دقیق مدل اصلاح شده از پیش آموزش دیده BERT با آموزش بیشتر بر روی مجموعه داده آموزشی.

صرف‌نظر از لایه‌های خروجی، معماری‌های یکسانی در پیش‌آموزش و در تنظیم دقیق استفاده می‌شوند. از همان پارامترهای مدل از پیش آموزش دیده برای مقداردهی اولیه مدل‌ها برای وظایف مختلف پایین‌دستی استفاده می‌شود. در طول تنظیم دقیق، تمام پارامترها به دقت تنظیم می شوند. نماد [CLS] یک نماد ویژه است که جلوی هر نمونه ورودی اضافه می‌شود، و [SEP] یک نشانه جداکننده ویژه است که در موارد لزوم، مثلا برای جدا کردن سؤالات و پاسخ‌ها مورد استفاده قرار می‌گیرد [۵۱].

- **مدل Glove**

مدل Glove یک روش تعبیه کلمه از پیش آموزش داده شده است که در دسترس عموم و استفاده از آن آزاد است. در این مقاله از بسته glove.twitter.27B.zip با حجم 1.42 GB استفاده شده است. این مدل حاوی ۱,۲ میلیون کلمه برای آموزش کلمات است. لایه Glove embedding برای یافتن کلمات منحصر به فرد استفاده می‌شود.

- **مدل Cove**

مدل Cove نیز یک مدل از پیش آموزش دیده و برای استفاده عموم در دسترس است. این مدل توالی‌های ورودی کلمه w را به دنباله‌های بردارهای $\tilde{w}$ تبدیل می‌کند. در مدل پیشنهادی، ورودی‌های متنی داده شده به صورت جداگانه با استفاده از مکانیزم تعبیه دو مدل از پیش آموزش دیده Glove و Cove کدگذاری می‌شوند و کلمات متن آنها به بردارها تبدیل می‌شوند. در این فرآیند، شاخص قطبیت مثبت با مقدار ۰ به عنوان [۰ ۰ ۱]، شاخص قطبیت خنثی با مقدار ۱ به عنوان [۰ ۱ ۰] ، شاخص قطبیت منفی با مقدار ۲ به عنوان [۱ ۰ ۰] نشان داده می‌شوند

۳-۱-۲- آماده سازی داده‌های تصویری

مشابه داده‌های متنی، در بخش داده‌های تصویری نیز از مدل های از پیش آموزش دیده برای استخراج ویژگی‌های تصاویر استفاده می‌شود. شبکه‌های کانولوشنی نتایج مناسبی برای استخراج ویژگی‌های تصاویر به دست آورده‌اند. به دلیل محدودیت اندازه لایه کانولوشن، اندازه ۱۲۸×۱۲۸ برای تصاویر ورودی در نظر گرفته می‌شود. انتخاب درست اندازه تصاویر ورودی تأثیر زیادی در سرعت پردازش دارد. اندازه بزرگ باعث ایجاد فضای ویژگی بزرگتر و سرعت یادگیری کمتر می‌شود و اندازه کوچک تصاویر ممکن است مفهوم تصاویر را به صورت کلی از بین ببرد.

۳-۱-۲-۱ استخراج ویژگی‌های تصاویر با استفاده از مدل VGG16

شبکه VGG16 از ۱۳ لایه کانولوشن و سه لایه کاملا متصل تشکیل شده است و پس از هر مرحله به لایه pooling متصل می‌شود. لایه‌های کانولوشن از هسته‌های ۳×۳ با stride ۱ و padding ۱ استفاده می‌کنند تا اطمینان حاصل شود که هر نقشه فعال‌سازی[2] همان ابعاد فضایی لایه قبلی را حفظ می کند. یک فعال‌سازی واحد خطی اصلاح شده[3] یا ReLU پس از هر کانولوشن در پایان max pooling انجام می‌شود. این کار در انتهای هر بلوک برای کاهش بعد فضایی انجام

[1] Tensor processing unit
[2] Activation plan
[3] Rectified Linear Unit (ReLU)

می‌شود. برای آن که هر نقشه فعال‌سازی بعد فضایی لایه قبلی را نصف کند از لایه max pooling با هسته ۲×۲ و با stride ۲ و بدون padding استفاده می‌کنند. لایه‌های کاملا متصل با ۴۰۹۶ واحد فعال شده ReLU قبل از ۱۰۰۰ لایه softmax کاملا متصل استفاده می‌شوند [۵۲]. این مدل از دو بخش ویژگی‌ها و طبقه‌بندی کننده تشکیل شده است. برای استفاده از این مدل برای استخراج ویژگی‌ها و تنظیم دقیق از کتابخانه keras استفاده شده است. برای حالت استخراج ویژگی‌ها پارامترهای آموزش دیده شبکه فریز، و آخرین لایه قسمت کاملا متصل حذف، و طبقه‌بندی کننده خطی Log soft max به آن اضافه می‌شود و پارامترهای اخرین لایه با مجموعه داده آموزش می‌بینند.

۳-۱-۲-۲ استخراج ویژگی‌های تصاویر با استفاده از مدل MobileNetV3

شبکه‌های کانولوشنی به طور گسترده‌ای در طبقه‌بندی تصاویر استفاده می‌شوند و برای مقابله با بسیاری از مشکلات و بهبود عملکرد آنها از نظر سرعت و اندازه، روش‌های مختلفی پیشنهاد شده‌اند. با افزایش تعداد داده‌های تصویری پردازش شده توسط دستگاه‌های تلفن همراه، استفاده از شبکه‌های عصبی برای تلفن‌های همراه نیز مورد اقبال واقع شده است. این شبکه‌ها نیاز به محاسبات گسترده و پشتیبانی سخت افزاری پیشرفته دارند که سازگاری با دستگاه‌های تلفن همراه را دشوار می‌کنند. نتایج تحقیقات نشان می‌دهد که MobileNetV3 می‌تواند بین کارآیی و دقت برای وظایف طبقه بندی تصویر تعادل مناسبی برقرار نماید [۵۳]. مرحله استخراج ویژگی پس از تنظیم دقیق مدل برای ۱۵۰ دوره در طول ده اجرای اولیه تصادفی انجام می‌شود که در آن از مدل MobileNetV3 استفاده شده که منجر به بالاترین دقت طبقه‌بندی در هر مجموعه داده شود. دسته‌ای با اندازه ۳۲ و یک رویکرد گرادیان نزولی تصادفی به نام RMSprop برای تنظیم دقیق مدل با نرخ یادگیری 1×10^{-4} استفاده می‌شود [۵۴].

۳-۲- مکانیزم‌های توجه بصری و چندراسی

در این مقاله از دو مکانیزم توجه برای تمرکز بر روی نواحی تصویری دارای احساس عاطفی و کشف کلمات دارای احساس استفاده شده است. این بخش به طور خلاصه به تشریح این دو مکانیزم می‌پردازد.

۳-۲-۱ مکانیزم توجه بصری برای داده های تصویری

به طور معمول تمام بخش‌های یک تصویر حاوی احساسات نیست و تنها بخشی از نواحی یک تصویر دارای احساسات عاطفی می‌باشد. با برجسته‌تر کردن این نواحی و تمرکز بیشتر بر روی آنها می‌توان به طبقه‌بندی موثرتری دست یافت. نتایج تحقیقات نشان داده است که مکانیزم توجه برای بسیاری از وظایف مرتبط با بینایی مانند تجزیه و تحلیل احساسات بصری مفید بوده است [۵۵]. در [۵۶] از مکانیزم توجه برای کشف خودکار مناطق بصری دارای ویژگی احساسی برای تجزیه و تحلیل احساسات استفاده شده است. مشابه [۷]، مکانیزم توجه مورد استفاده در این مقاله نیز با استفاده از معادلات و روابط ۱ تا ۴ به صورت زیر فرموله می‌شود.

مجموعه $I = [I_1, I_2, I_3, \dots\dots, I_n]$، مجموعه‌ای از n تصویر را نشان می‌دهد. برای هر تصویر I_i، از شبکه‌های عصبی کانولوشنی برای به دست آوردن نقشه‌های ناحیه تصویر

$R_i = [r_i^1, r_i^2, r_i^3, \dots\dots\dots\dots\dots\dots, r_i^D] \in \mathbb{R}^{D*M}$

به شرح رابطه ۱ استفاده می‌شود:

$$R_i = fc(I_i; \Theta c), R_i \in \mathbb{R}^{D*M} \qquad (۱)$$

در رابطه ۱، θ_c نمایانگر پارامترهای لایه‌های شبکه کانولوشنی است، D تعداد نواحی تصویر را نشان می‌دهد و M بیانگر ابعاد نقشه است. در این مکانیزم توجه، امتیاز α_i^j به صورت یک عدد بین ۰ و ۱ به هر ناحیه تصویر I_i^j بر اساس میزان ارتباط آن با احساسات اختصاص می‌یابد و از یک تابع softmax برای محاسبه α_i^j بر اساس رابطه ۲ استفاده می‌شود:

$$\alpha_i^j = \frac{\exp(e_i^j)}{\sum_{j=1}^{D} \exp(e_i^j)} \qquad (۲)$$

ویژگی‌های بصری حضور یافته را می‌توان به عنوان میانگین وزنی هر یک از مناطق بر اساس رابطه ۳ محاسبه کرد:

$$W_i = \sum_{1 \le j \le D} \alpha_i^j I_i^j, W_i \in \mathbb{R}^M \qquad (۳)$$

مکانیزم توجه برای تولید خودکار ویژگی‌های تصویر مورد نظر بر اساس رابطه ۴ نشان داده می‌شود و $\Theta_a^{(v)}$ وزن پارامترهایی همچون بایاس است. در مقایسه با ویژگی‌های بصری مستقل اصلی، W_i نگاشت ویژگی‌های بصری وزنی برای نشان دادن

ویژگی‌های مربوط به احساسات موثرتر است. در مرحله بعد، این ویژگی می‌تواند به عنوان ورودی طبقه‌بندی کننده احساسات ارائه شود. و لایه کاملا متصل برای انجام طبقه‌بندی احساسات ساخته شده است.

$$W_i = fa\left(R_i; \Theta_a^{(v)}\right), W_i \in \mathbb{R}^M \quad (۴)$$

۲-۲-۳ مکانیزم توجه چندراسی برای داده های متنی

مشابه آنچه در مورد تفاوت اهمیت نواحی مختلف تصاویر در انتقال احساسات و عواطف بیان شد، برخی از کلمات متون نیز می‌توانند حاوی احساسات بیشتری در مقایسه با سایر کلمات باشند. در مطالعات اخیر اثبات شده است که استفاده از مکانیزم‌های توجه برای بسیاری از وظایف پردازش زبان طبیعی از جمله تجزیه و تحلیل احساسات متن [۴۹ و ۵۷] مفیده بوده است. مکانیزم توجه چندراسی (MHA) استفاده شده در روش پیشنهادی قادر است به صورت مستقیم مهمترین کلمات دارای احساسات را برجسته کند. مکانیزم توجه چندراسی یک ماژول برای مکانیزم‌های توجه است که چندین بار به طور موازی از طریق یک مکانیزم توجه اجرا می‌شود. از نظر بصری، توجه چندراسی اجازه می‌دهد تا به طور متفاوتی به بخش‌هایی از دنباله توجه شود (به عنوان مثال وابستگی‌های بلندمدت در مقابل وابستگی‌های کوتاه مدت). مکانیزم توجه چندراسی از طریق multiple scaled dot-product attention محاسبه و متصل می‌شود که دارای سه ماتریس ورودی است: Q(query)، K(Key) و V(value). در حوزه پردازش زبان طبیعی، کلید (K) و مقدار (V) معمولا برابر هستند[۵۸]. ساختار scaled dot-product attention و به صورت رابطه ۵ محاسبه می‌شود:

$$\text{Attention (Q, K, V)} = \text{softmax } (\frac{QK^T}{\sqrt{d_k}})\text{V} \quad (۵)$$

در این رابطه، $\frac{1}{\sqrt{d_k}}$ عامل مقیاس‌بندی است. توجه چندراسی (MHA) را می‌توانیم با معادلات ۶-۸ به دست آوریم [۵۹]:

$$hd_i = Attention\ (QW_i^Q, KW_i^k, VW_i^V) \quad (۶)$$

$$\text{MHA (Q, K, V)} = \text{Concat } (hd_i, \dots, hd_H\ )W^O \quad (۷)$$

$$W_i^Q \in \mathbb{R}^{d_{model} * d_k} \quad , W_i^K \in \mathbb{R}^{d_{model} * d_k}, W_i^V \in \mathbb{R}^{d_{model} * d_V}\ , and\ W^O \in \mathbb{R}^{Hd_v * d_{model}} \quad (۸)$$

در رابطه ۸ ماتریس‌های وزنی آورده شده است و d_{model} بعد نمایش کلمه پنهان پس از پردازش شبکه کانولوشنی را نشان می‌دهد. $d_k=d_V$ ابعاد راس توجه را نشان می‌دهد. با توجه به کاهش ابعاد هر راس، کل هزینه محاسباتی، مشابه تک‌راس با ابعاد کامل است. در محاسبه تعاملی بخش زمینه، همان‌طور که گفته شد توجه چندراسی دارای سه ورودی است که با Q، K و V نشان داده می‌شود، که Q معناشناسی زمینه‌ای را نشان می‌دهد و K و V اطلاعات گرایش احساسی را نشان می‌دهند. می‌توان بازنمایی‌های تعاملی $[h_1^c, h_2^c, h_3^c, \dots \dots h_n^c]$ معناشناسی زمینه‌ای و بازنمایی‌های تعاملی $[h_1^s, h_2^s, h_3^s, \dots \dots h_m^s]$ را از اطلاعات گرایش احساسی به دست آورد.

۳-۳- همجوشی

استراتژی همجوشی یکی از استراتژی‌های مورد استفاده برای ادغام نتایج در تحلیل احساسات چندوجهی است و به عنوان روشی مؤثر برای ترکیب ویژگی‌هایی با ماهیت‌های متفاوت در مسائل مختلف یادگیری ماشینی به کار می‌رود و می‌تواند باعث بهبود در تجزیه و تحلیل احساسات شود. در این مقاله از همجوشی‌های زودهنگام و دیرهنگام استفاده شده است. ابتدا ویژگی‌های استخراج شده توسط مدل‌های از پیش-آموزش دیده به صورت نرمال‌سازی شده در یک بردار از ویژگی‌ها ترکیب می‌شوند و برای پیش‌بینی نهایی از لایه کاملا متصل بر روی آنها استفاده می‌شود. همچنین نتایج حاصل از طبقه‌بندها در سطح تصمیم‌گیری ترکیب می‌شود تا قطبیت نهایی تعیین گردد. یعنی در این استراتژی نتایج رای‌گیری شده از خروجی‌های شبکه‌های ترکیبی در نهایت ادغام می-شوند. همچنین مشابه [۶۰]، برای الحاق ویژگی‌های متنی از Logistic regression به صورت همجوشی زودهنگام استفاده می‌شود.

۴- نتایج آزمایشات

در این بخش به ارائه نتایج آزمایشات عملی حاصل از پیاده سازی روش پیشنهادی و تحلیل آنها خواهیم پرداخت.

۴-۱- مجموعه داده

در این مطالعه از مجموعه داده MVSA-Multiple برای ارزیابی

استفاده شده است. MVSA-Multiple شامل ۱۹۶۰۰ دوگانه تصویر و متن است که توسط سه annotator برچسب‌گذاری شده‌اند. در این مجموعه داده احساس واقعی با اخذ اکثریت آرا از سه احساس مثبت، منفی و خنثی برای هر حالت به طور جداگانه محاسبه شده است. ۷۵٪ داده‌ها به عنوان داده‌های آموزشی ۱۰٪ به عنوان داده اعتبار سنجی و ۱۵٪ به عنوان داده آزمایشی در نظر گرفته شده است. برای ارزیابی بازدهی مدل‌های طبقه‌بندی، از چهار معیار صحت (ACC)، دقت (P)، امتیاز (F1) و فراخوانی (R) استفاده می‌شود که مقادیر آنها از روابط ۹ تا ۱۲ به دست می‌آیند.

$$P = \frac{TP}{TP+FP} \quad (۹)$$

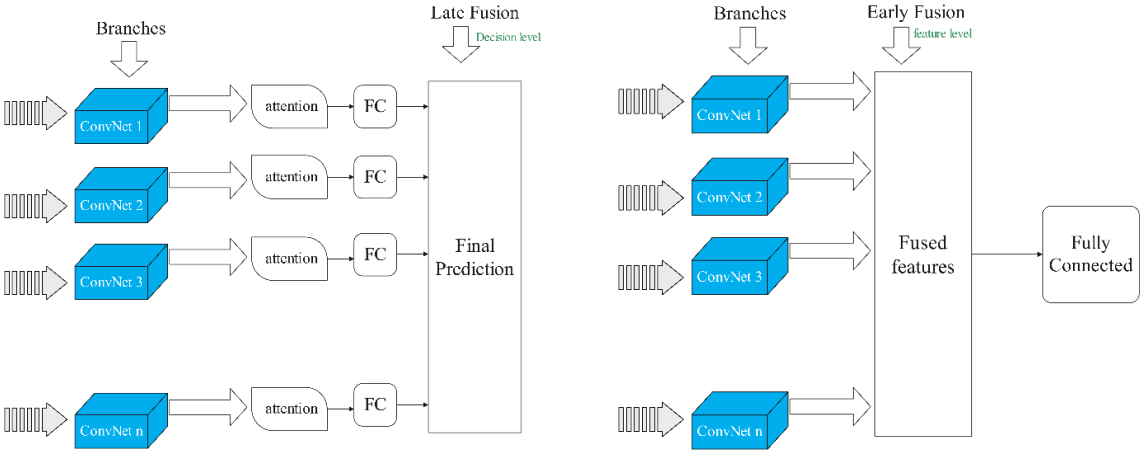
$$R = \frac{TP}{TP+FN} \quad (۱۰)$$

$$F1 = \frac{2PR}{P+R} \quad (۱۱)$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (۱۲)$$

۴-۲- نتایج ارزیابی

در جداول ۱ تا ۴ نتایج حاصل از پیاده‌سازی روش پیشنهادی بر روی مجموعه داده MVSA-Multiple نشان داده شده و با نتایج سایر روش‌های اجرا شده بر روی این مجموعه داده مقایسه شده است. در شکل (۳) میزان دقت و خطای روش پیشنهادی و در شکل (۴) ماتریس درهم‌ریختگی مدل پیشنهادی نمایش داده شده است.



الف) استراتژی‌های همجوشی زودهنگام (سطح ویژگی‌ها) ب) همجوشی دیرهنگام (سطح تصمیم گیری)

شکل ۲. استراتژی‌های همجوشی زودهنگام و دیرهنگام

جدول ۱. عملکرد روش پیشنهادی برای طبقه‌بندی متون

Accuracy (%)	F1 (%)	Recall (%)	Precision (%)	روش
۹۳/۸۸	۹۳/۹۳	۹۴/۱۴	۹۳/۷۲	BERT+CNN+MHA
۹۳/۱۲	۹۲/۵۹	۸۹/۸۰	۹۵/۵۵	Fusion cove and Glove+CNN+ MHA
۹۵/۴۱	۹۳/۸۲	۹۱/۶۳	۹۶/۱۲	Ensemble Transfer learning (ETL)

جدول ۲. عملکرد روش پیشنهادی برای طبقه‌بندی تصاویر

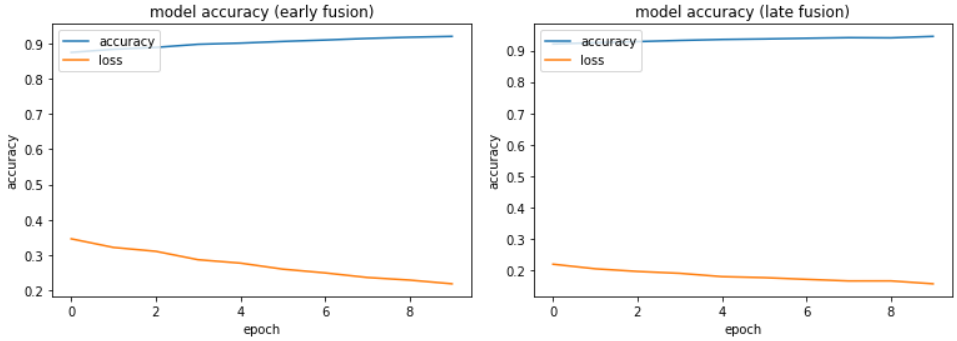
Accuracy (%)	F1 (%)	Recall (%)	Precision (%)	روش
۹۳/۸۱	۹۱/۴۰	۹۰/۶۹	۹۲/۱۲	VGG16+CNN+VA
۹۰/۶۲	۹۱/۷۶	۹۱/۵۰	۹۲/۰۳	MobileNet+CNN+ VA
۹۵/۱۸	۹۴/۵۹	۹۲/۳۶	۹۶/۹۲	Ensemble Transfer learning (ETL)
۹۶	۹۳/۴۷	۹۲	۹۵	Fusion ETL

جدول ۳. نتایج روش‌های همجوشی زودهنگام و دیرهنگام

Accuracy (%)	F1 (%)	Recall (%)	Precision (%)	روش
۹۲/۰۳	۹۱/۸۵	۹۰/۴۴	۹۳/۳۲	Early fusion (Features fusion)
۹۶/۵۲	۹۳/۷۱	۹۲/۳۱	۹۵/۱۷	Late fusion ETL (Decision fusion)

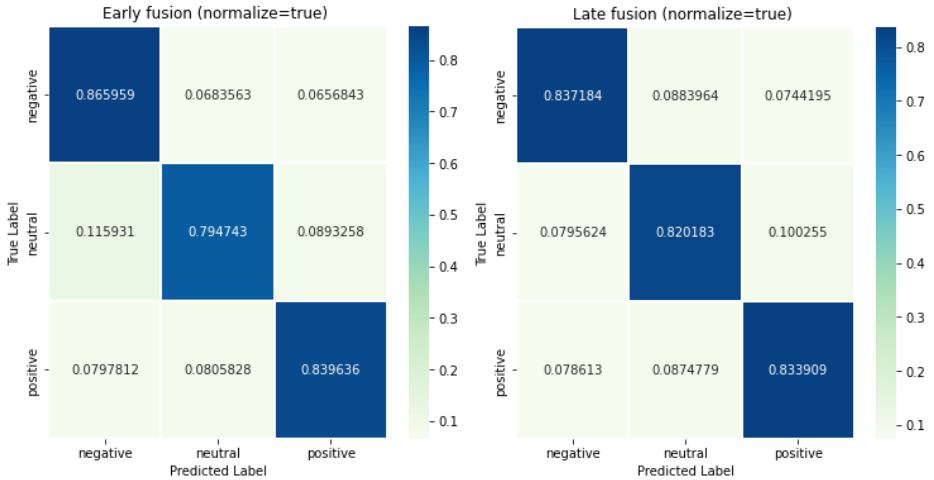
جدول ۴. مقایسه روش پیشنهادی با سایر روش‌ها

Accuracy (%)	F1 (%)	روش	مجموعه داده
۶۵/۶۲	۵۵/۳۶	[22] 2013 SentiBank & Borth SentiStrength	MVSA-Multiple
۶۷/۹۰	-	[26] 2017 Late-RMNN Xu	
۶۷/۷۶	-	[27] 2017 HSAN Xu	
۶۸/۹۲	۶۸/۸۳	[28] 2018 CoMN Xu	
۷۷/۸۹	۷۵/۲۶	[29] 2020 DMLANet Jiang	
۷۲/۹۸	۷۲/۹۸	[30] 2020 MVAN-M Xiaocui	
۹۶/۵۲	۹۴/۸۰	Fusion ETL	



الف) میزان دقت و خطا در همجوشی دیرهنگام ب) میزان دقت و خطا در همجوشی زودهنگام

شکل ۳. میزان دقت و خطای همجوشی



الف) ماتریس درهم‌ریختگی همجوشی دیرهنگام ب) ماتریس درهم‌ریختگی همجوشی زودهنگام

شکل ۴. ماتریس‌های درهم‌ریختگی همجوشی

۵-نتیجه‌گیری

در این مقاله، از مکانیزم‌های توجه بصری و چندراسی در یک چارچوب یادگیری ترکیبی انتقال چندگانه برای تحلیل احساسات چندوجهی استفاده شده است. روش‌های همجوشی زودهنگام و دیرهنگام به صورت مجزا برای الحاق متون، ترکیب ویژگی‌ها و تعیین قطبیت نهایی به کار گرفته شده‌اند. در همجوشی زودهنگام همبستگی داخلی بین مدل‌های از پیش آموزش دیده Glove و Cove به وجود آمده، و همجوشی دیرهنگام برای خروجی‌های طبقه‌بندها اعمال شده است تا برچسب نهایی تعیین گردد. به صورت جداگانه از مدل BERT نیز برای استخراج ویژگی‌ها و تعبیه کلمات استفاده شده است. زیرا این مدل می‌تواند متن را به صورت دوطرفه بخواند و به صورت هم‌زمان پردازش نماید و به همین دلیل در انجام مسائل پردازش زبان طبیعی، به نتایج بسیار خوبی دست یافته است. استفاده از مکانیزم‌های توجه در روش ارائه شده باعث گردیده تا بر روی نقاط دارای احساسات در متن و تصویر تمرکز شود که این کار باعث بهبود عملکرد طبقه‌بندی شده است. یادگیری ترکیبی نیز با ترکیب طبقه‌بندها منجر به دست‌یابی به نتایج بهتری شده است. به‌طورکلی استفاده ترکیبی از چندین تکنیک یادگیری ، یادگیری انتقال و مکانیزم‌های توجه در این مطالعه باعث گردیده است که زمان آموزش کوتاه‌تر شده و دقت نتایج افزایش یابد.

مراجع

[1] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction", Expert Systems with Applications, Vol. 178, 2021, 115019

[2] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and AE. Hassanien, "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media",Applied Soft Computing, Vol.97, 2020, 106754.

[۳] حمیدرضا میرشاهولد، رامین قاسمی اصل، ناهید رئوفی و مهرداد ملک زاده دیرین، مدل سازی و پیش بینی نقطه اشتعال ترکیبات هیدرو کربنی با استفاده از شبکه عصبی"، نشریه مدل سازی در مهندسی، دوره ۱۹، شماره ۶۴، بهار ۱۴۰۰، صفحه ۱۰۹-۱۱۶.

[4] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, " Tourism recommendation system based on semantic clustering and sentiment analysis", Expert Systems with Applications, Vol. 167, 2021, 114324

[۵] هادی تقی‌زاده، تاج بخش نوید چاخرلو، عادل علیزاده و آیدین شیخ عبداله‌زاده ممقانی، "مدل‌سازی عمر خستگی اتصالات دو لبه برشی با استفاده از شبکه عصبی مصنوعی"، نشریه مدل سازی در مهندسی، دوره ۱۵، شماره ۴۹، تابستان ۱۳۹۶، صفحه ۵۵-۶۳.

[6] H. Jafarian, AH. Taghavi, A. Javaheri, and R. Rawassizadeh ,"Exploiting BERT to improve aspect-based sentiment analysis performance on Persian language", In: 7th International Conference on Web Research (ICWR), IEEE, 2021, pp. 5-8.

[7] F. Huang, X. Zhang, Z. Zhao, J. Xu, and Z. Li, "Image–text sentiment analysis via deep multimodal attentive fusion", Knowledge-Based Systems, Vol. 167, 2019, pp. 26-37.

[8] W. Nie, Y. Yan, D. Song, and K. Wang," Multi-modal feature fusion based on multi-layers LSTM for video emotion recognition", Multimedia Tools and Applications , Vol. 80(11), 2021, pp. 16205-16214.

[9] V. Aiswaryadevi, S. Kiruthika, G. Priyanka, N. Nataraj, and M. Sruthi,"Effective Multimodal Opinion Mining Framework Using Ensemble Learning Technique for Disease Risk Prediction", In: Inventive Computation and Information Technologies. Springer,2021, pp. 925-933.

[10] A. Ghorbanali, MK. Sohrabi, and F. Yaghmaee, "Ensemble transfer learning-based multimodal sentiment analysis using weighted convolutional neural networks:, Information Processing & Management, 2022, Vol. 59(3), p. 102929.

[11] J. Deng, S. Frühholz, Z. Zhang, and B. Schuller, "Recognizing emotions from whispered speech based on acoustic feature transfer learning", IEEE Access, Vol. 5, 2017, pp. 235-246.

[12] A. Maas, RE. Daly, PT. Pham, D. Huang, AY. Ng, and C.Potts,"Learning word vectors for sentiment analysis", In: Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies, 2011, pp. 142-150.

[13] Y. Rao, J. Lei, L. Wenyin, Q. Li, and M. Chen , "Building emotional dictionary for sentiment analysis of online news", World Wide Web, Vol. 17 (4), 2014, pp. 723-742.

[14] A. Ishaq, S. Asghar, and SA. Gillani , " Aspect-based sentiment analysis using a hybridized approach based on CNN and GA", IEEE Access, Vol. 8,2020, pp. 499-512.

[15] Y. Ma, J. Yu, B. Ji, J. Chen, S. Zhao, and J. Chen, "Three-Way Decisions Based RNN Models for Sentiment Classification", In: International Joint Conference on Rough Sets. Springer, 2021, pp. 247-258.

[16] L. Zhao, L. Li, X. Zheng, and J. Zhang, "A BERT based sentiment analysis and key entity detection approach for online financial texts", In: 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, 2021, pp. 233-238.

[17] Y. Yang, J. Jia, S. Zhang, B. Wu, Q. Chen, J. Li, C. Xing, and J. Tang, "How do your friends on social media disclose your emotions? ",In: 28th AAAI conference on artificial intelligence, 2014, pp. 306-312.

[18] A. Yadav, DK. Vishwakarma A deep learning architecture of RA-DLNet for visual sentiment analysis. Multimedia Systems, Vol. 26, 2020, pp. 431-451.

[19] Q. You, J. Luo, H. Jin, and J. Yang, "Joint visual-textual sentiment analysis with deep neural networks", In: Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 1071-1074.

[20] C. Baecchi, T. Uricchio, M. Bertini, and A. Del Bimbo, " A multimodal feature learning approach for sentiment analysis of social network multimedia", Multimedia Tools and Applications , Vol. 75 (5) ,2016, pp. 507-525.

[21] X. Zhu, B. Cao, S. Xu, B. Liu, and J. Cao, "Joint visual-textual sentiment analysis based on cross-modality attention mechanism", In: International conference on multimedia modeling, Springer, 2019, pp. 264-276.

[22] D. Borth, R. Ji, T. Chen, T. Breuel, and S-F. Chang , "Large-scale visual sentiment ontology and detectors using adjective noun pairs", In: Proceedings of the 21st ACM international conference on Multimedia, 2013, pp. 223-232.

[23] Z. Zhao, H. Zhu, Z. Xue, Z. Liu, J. Tian, MCH. Chua, and M. Liu, "An image-text consistency driven multimodal sentiment analysis approach for social media", Information Processing & Management, Vol. 56 (6), 2019, p. 102097.

[24] Q. Fang, C. Xu, J. Sang, MS. Hossain, and G. Muhammad, " Word-of-mouth understanding: Entity-centric multimodal aspect-opinion mining in social media", IEEE Transactions on Multimedia 17, Vol. 12, 2015, pp. 281-296.

[۲۵] علیرضا قربانعلی، محمد کریم سهرابی و فرزین یغمایی، "طبقه‌بندی و تجزیه و تحلیل احساسات چندوجهی با استفاده از شبکه‌های کانولوشن وزن‌دار ترکیبی" ، نشریه فناوری اطلاعات در طراحی مهندسی، دوره ۱۴، شماره ۱، شهریور ۱۴۰۰، صفحه ۱-۱۰.

[26] N. Xu, W. Mao, "A residual merged neutral network for multimodal sentiment analysis", In: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), IEEE, 2017, pp. 6-10.

[27] N. Xu, "Analyzing multimodal public sentiment based on hierarchical semantic attentional network", In: 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), IEEE, 2017, pp. 152-154.

[28] N. Xu, W. Mao, and G. Chen, "A co-memory network for multimodal sentiment analysis ", In: The 41st international ACM SIGIR conference on research & development in information retrieval, 2018, pp. 929-932.

[29] T. Jiang, J. Wang, Z. Liu, and Y. Ling, "Fusion-extraction network for multimodal sentiment analysis", Advances in Knowledge Discovery and Data Mining , Vol. 12085,2020,785.

[30] X. Yang, S. Feng, D. Wang, and Y. Zhang, "Image-text Multimodal Emotion Classification via Multi-view Attentional Network", IEEE Transactions on Multimedia, Vol. 23, 2020, pp. 41014-4026

[31] D. Gkoumas, Q. Li, C. Lioma, Y. Yu, and D. Song, "What makes the difference? An empirical comparison of fusion strategies for multimodal language analysis", Information Fusion, Vol. 66, 2021, pp. 184-197.

[32] Y. Xiao, F. Codevilla, A. Gurram, O. Urfalioglu, and AM. López, "Multimodal end-to-end autonomous driving", IEEE Transactions on Intelligent Transportation Systems, Vol. 23, 2022, pp. 537-547.

[33] X. Zhang, J. Liu, J. Shen, S. Li, K. Hou, B. Hu, J. Gao, and T. Zhang, "Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine", IEEE transactions on cybernetics, Vol. 51(9), 2021, 4386-4399.

[34] J. Huang, J. Tao, B. Liu, Z. Lian, and M. Niu, "Multimodal transformer fusion for continuous emotion recognition", In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 507-511.

[۳۵] فاضل فصیحی، محمودرضا کی‌منش، سیدعلی صحاف و سهیل قره، "تعیین ضریب بار هم‌ارز مبتنی بر الگوریتم شبکه عصبی مصنوعی ، نشریه مدل سازی در مهندسی، دوره ۱۹، شماره ۶۵، تابستان ۱۴۰۰، صفحه ۱۴۹-۱۶۰.

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

[37] Y. Cheng, L. Yao, G. Xiang, G. Zhang, T. Tang, and L. Zhong , " Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism", IEEE Access , Vol. 8, 2020, pp. 964-975.

[38] AH. Ombabi, W. Ouarda, and AM. Alimi , "Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks", Social Network Analysis and Mining , Vol 10 (1) ,2020, pp.1-13.

[39] JP. Gujjar, HP. Kumar, and NN. Chiplunkar, Image classification and prediction using transfer learning in colab notebook. Global Transitions Proceedings, 2021. Vol. 2(2), p. 382-385.

[40] T. Tang, X. Tang, and T. Yuan ,"Fine-Tuning BERT for Multi-Label Sentiment Analysis in Unbalanced Code-Switching Text",IEEE Access , Vol.8,2020, pp. 248-256.

[41] TN. Rincy, R. Gupta,"Ensemble Learning Techniques and its Efficiency in Machine Learning: A Survey", In: 2nd International Conference on Data, Engineering and Applications (IDEA), IEEE, 2020, pp. 1-6.

[42] X. Frazao, LA. Alexandre, "Weighted convolutional neural network ensemble", in Iberoamerican Congress on Pattern Recognition. 2014. In: E. Bayro-Corrochano, E. Hancock, (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2014. Lecture Notes in Computer Science, vol 8827.

[43] Á. Casado-García, J. Heras," Ensemble methods for object detection", In: ECAI 2020. IOS Press, (2020) pp. 688-695.

[44] Y. Kawana, N. Ukita, J-B. Huang,and M-H. Yang,"Ensemble convolutional neural networks for pose estimation", Computer Vision and Image Understanding , Vol 169 ,2018, pp. 62-74.

[45] S. Poria, H. Peng, A. Hussain, N. Howard, and E. Cambria, "Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis", Neurocomputing, Vol. 261, 2017, pp. 217-230.

[46] L. Nanni, YM. Costa, RL. Aguiar, RB. Mangolin, S. Brahnam, and CN Silla ," Ensemble of convolutional neural networks to improve animal audio classification", EURASIP Journal on Audio, Speech, and Music Processing, 2020, https://doi.org/10.1186/s13636-020-00175-3.

[47] AK. Das, S. Ghosh, S. Thunder, R. Dutta, S. Agarwal, and A. Chakrabarti, "Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network", Pattern Analysis and Applications, Vol. 24, 2021, pp. 1111-1124.

[48] D. Alexandru, S. Stelian, NI. Alina, and F.Aschim, "Ensembles of Convolutional Neural Networks Trained Using Unconventional Data for Stock Predictions", In: Business Revolution in a Digital Era, Springer, 2021, pp. 241-250.

[49] Y. Wang, M. Huang, X. Zhu, 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.

[50] J. Briskilal, C. Subalalitha, "An ensemble model for classifying idioms and literal texts using BERT and RoBERTa", Information Processing & Management, Vol. 59(1), 2022, 102756.

[51] J. Devlin, M-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", 2018, arXiv preprint arXiv:181004805.

[52] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint ,2014,arXiv:14091556.

[53] S. Qian, C. Ning,and Y. Hu ,"MobileNetV3 for Image Classification", In: 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), IEEE, 2021, pp. 490-497.

[54] M. Abd Elaziz, A. Dahou, NA. Alsaleh, AH. Elsheikh, AI. Saba, and M. Ahmadein, " Boosting COVID-19 Image Classification Using MobileNetV3 and Aquila Optimizer Algorithm" Entropy, Vol. 23(11), 2021, 1383.

[55] Q .You, L. Cao, H. Jin, and J. Luo ,"Robust visual-textual sentiment analysis: When attention meets tree-structured recursive neural networks", In: proceedings of the 24th ACM international conference on multimedia, 2016, pp. 1008-1017.

[56] Q. You, H. Jin,and J. Luo ,"Visual sentiment analysis by attending on local image regions", In: Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 231-237.

[57] H. Chen, M. Sun, C. Tu, Y. Lin, and Z. Liu ,"Neural sentiment classification with user and product attention", In: Proceedings of the 2016 conference on empirical methods in natural language processing, 2016, pp. 1650-1659.

[58] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, AN. Gomez, Ł. Kaiser,and I. Polosukhin," Attention is all you need. In: Advances in neural information processing systems", 2017, pp. 5998-6008.

[59] Y. Zhu, W. Zheng, and H. Tang, "Interactive dual attention network for text sentiment classification", Computational Intelligence and Neuroscience 2020, p. 8858717.

[60] Q. Le, T. Mikolov ,"Distributed Representations of Sentences and Documents", In: Proceedings of the 31st International Conference on Machine Learning Research, PMLR, Vol. 32(2), 2014, pp. 1188-1196.