Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال اول، شماره دوم، پاییز و زمستان ۱۴۰۲، صفحات۱۱۷-۱۰۵

استفاده از مدل‌های یادگیری انتقالی برای بهبود تشخیص احساسات بصری در شبکه‌های اجتماعی

محمد روستایی [۱*]، میثم میرزایی [۲]

[۱] دانشکده و پژوهشکده هوش مصنوعی و علوم شناختی، دانشگاه جامع امام حسین (ع)، تهران، ایران

[۲] پژوهشگر دانشگاه جامع امام حسین (ع)، تهران، ایران

مقاله پژوهشی

تاریخ دریافت:
۱۴۰۲/۰۳/۲۹

تاریخ پذیرش:
۱۴۰۲/۰۸/۲۷

کلیدواژه‌ها:
تجزیه‌وتحلیل احساسات تصویر، انتقال یادگیری، یادگیری عمیق، شبکه‌های اجتماعی

نویسنده‌مسئول:
M.Roustaei@ihu.ac.ir

چکیده

تجزیه‌وتحلیل احساسات افراد از محتوای رسانه‌های اجتماعی از طریق متن، گفتار و تصاویر، در انواع مختلفی از برنامه‌ها و کاربردها موردنیاز است. اکثر مطالعات تحقیقاتی اخیر درزمینه تجزیه‌وتحلیل احساسات، بر داده‌های متنی تمرکز داشته‌اند. بااین‌حال، کاربران رسانه‌های اجتماعی، عکس‌ها و فیلم‌های مشابه بیشتری نسبت به متن به اشتراک می‌گذارند. به‌عبارت‌دیگر، تصاویر بهترین روش برای انتقال احساسات به دیگران هستند. ازاین‌رو، تمرکز بر توسعه یک مدل تحلیل احساسات بر اساس تصاویر در رسانه‌های اجتماعی اهمیت دارد. در این مقاله، از مدل یادگیری انتقال DenseNet-121 برای تحلیل احساسات بر اساس تصاویر استفاده خواهیم کرد. برای پیاده‌سازی این روش، از تصاویر موجود در مجموعه داده Image Sentiment استفاده خواهیم نمود. این مجموعه داده شامل آدرس‌های اینترنتی تصاویر به همراه قطبیت‌های احساسی آن‌ها است. بر اساس نتایج به‌دست‌آمده، دقت مدل پیشنهادی در این مقاله برابر با ۸۹٪ است که در مقایسه با کارهای پیشین درزمینه تجزیه‌وتحلیل احساسات بصری، مدل پیشنهادی، بهبود ۵ تا ۱۰ درصدی را نشان‌می‌دهد.

: 10.22034/ABMIR.2023.20213.1028

E-ISSN: 2821-2037

۱- مقدمه

در دنیای امروز، مفهوم تعامل انسان و کامپیوتر بسیار پرکاربرد است [۱]. این مفهوم شامل استفاده از ماشین‌ها / کامپیوترها برای تصمیم‌گیری و پیش‌بینی مسائل خاصی مانند احساسات افراد می‌شود. این مسئله از برخی تکنیک‌های هوش مصنوعی برای پیاده‌سازی تحلیل احساسات بر روی تصاویر استفاده می‌کند.

شبکه‌های اجتماعی مانند فیس‌بوک و یوتیوب در بین آمریکایی‌ها بسیار محبوب هستند و متن / تصاویر زیادی را منتشر می‌کنند [۲]. پژوهشگرانی که در این زمینه فعالیت می‌کنند، متوجه شده‌اند که استفاده از رسانه‌های اجتماعی مانند اینستاگرام، پین‌تر است و واتساپ در میان افراد بالای ۳۰ سال کمتر رایج است و اطلاعات بدون پردازش رسانه‌های اجتماعی را می‌توان به‌طور مفیدی تبدیل و برای برنامه‌های کاربردی مرتبط با کسب‌وکار پردازش کرد [۳].

تحلیل احساسات یک حوزه مهم از پردازش زبان طبیعی است و درک نظرات و عقاید افراد در بسیاری از کاربردها اساسی است [٤-٦]. به‌عنوان‌مثال، ارزیابی مشتریان [۷]، بررسی فیلم‌ها [۸]، بررسی هتل‌ها [۹] و نظارت بر رسانه‌های اجتماعی [۱۰] از کاربردهای رایج تحلیل احساسات هستند. تحلیل احساسات شامل دسته‌بندی احساسات افراد به دسته‌های مثبت، منفی و خنثی می‌شود [۱۱].

یکی از چالش‌های اساسی در تحلیل احساسات مبتنی بر متن، وجود داده‌های چندزبانه است و این باعث می‌شود که متون پیچیده برای همه قابل‌فهم نباشند. در مقابل، محتوای بصری نسبت به متن، جذابیت بیشتری برای کاربران شبکه‌های اجتماعی دارد. برخی از رسانه‌های اجتماعی مانند فلیکر و اینستاگرام، به پست‌های تصویری بیشتر اهمیت می‌دهند تا پست‌های متنی [۱۲]. برای تحلیل احساسات از تصاویر، نیازمند رویکردهای هوشمندانه هستیم و باید تلاش زیادی در این زمینه انجام شود. مدل‌های پیش‌بینی احساسات بصری شامل تشخیص دسته‌های احساسات (مثبت، منفی و خنثی) مرتبط با تصویر هستند. روش‌های مختلفی برای انجام تحلیل احساسات بصری وجود دارد، اما استفاده از مدل‌های یادگیری عمیق پیش‌آموزش‌دیده می‌تواند عملکرد وظایف طبقه‌بندی احساسات را حتی با مجموعه داده‌های نامتعادل بهبود بخشد [۱۳]. در مقایسه با مدل‌های یادگیری ماشین برای همان کار، مدل‌های یادگیری انتقالی بهترین نتایج را در طبقه‌بندی احساسات به‌دست می‌دهند [۱٤-۱٦]. در این مقاله، از مدل یادگیری انتقالی پیش‌آموزش‌دیده DenseNet-121 برای پیش‌بینی احساسات تصویر در مجموعه داده احساسات Image Sentiment استفاده خواهیم کرد.

ادامه این مقاله به بخش‌های زیر تقسیم می‌شود: بخش ۲ برخی از کارهای موجود در مورد تجزیه‌وتحلیل احساسات تصویر که در سال‌های اخیر معرفی‌شده‌اند را توضیح می‌دهد. بخش ۳ یک بلوک دیاگرام را نشان‌می‌دهد که مراحل تعیین دسته‌بندی تصویر را توضیح می‌دهد. در بخش ٤، روش موردنظر با توجه به مفهوم یادگیری انتقالی به‌همراه جزئیات معماری آن ارائه‌شده است. در بخش ٥، اطلاعات مربوط به مجموعه داده و برخی از تصاویر نمونه نشان داده‌شده است. در بخش ٦، نتایج به‌دست‌آمده از شبیه‌سازی برحسب برخی از معیارهای عملکرد، مانند دقت، صحت، معیار F1و غیره نشان داده‌شده است. در بخش ۷، به نتیجه‌گیری و معرفی راهکارهایی برای کارهای آینده می‌پردازیم.

۲- پیشینه تحقیق

در این قسمت، به بررسی برخی از مقالات موجود درزمینه پیش‌بینی احساسات تصویری می‌پردازیم. تحلیل احساس تصاویر با استفاده از رویکردهای مختلف انجام‌می‌شود که شامل استخراج ویژگی‌های سطح پایین، ویژگی‌های معنایی و استفاده از مدل‌های یادگیری ماشینی و عمیق است.

دریکی از مقالات [۱۷]، ویژگی‌های سطح پیکسل همراه با برخی از ویژگی‌های سطح پایین مانند رنگ و بافت برای پیش‌بینی احساسات تصویر استفاده‌شده است. نویسندگان در مقاله دیگری [۱۸] از تکنیک Bag-of-Visual-Words (BoVW) به‌همراه تحلیل معنایی پنهان (LSA) برای طبقه‌بندی احساسات تصویری استفاده کردند و به چالش ثبت انواع احساسات موجود در مناطق مختلف تصویر پرداختند.

در مقالات دیگری [۱۹،۲۰]، ویژگی‌های زیبایی‌شناختی از تصاویر مرتبط با زیبایی یا هنر برای تحلیل احساسات بصری استخراج شدند و نتایج قابل‌توجهی نسبت به ویژگی‌های سطح پایین به‌دست آمد.

مفهوم جفت صفت–اسم (ANPs) توسط نویسندگان در [۲۱] برای توصیف جزئیات احساسی / عاطفی تصاویر پیشنهادشده است و رویکردی به نام SentiBank برای شناسایی ۱۲۰۰ ANP در تصاویر توسعه داده‌شده است.

همچنین، در مقاله [۲۲]، ویژگی سطح میانی به نام Sentribute به‌عنوان جایگزینی برای ویژگی‌های سطح پایین در طبقه‌بندی احساسات بصری توسعه داده‌شده است و توانایی برقراری ارتباط بین این ویژگی‌ها و احساسات عاطفی توسط تصاویر را نشان داده است.

نویسندگان در [۲۳] نیز برای استخراج ویژگی‌های تصویر و طبقه‌بندی احساسات از تصاویر از ANP ها به‌عنوان توصیف‌گرهای سطح میانی با استفاده از ماشین بردار پشتیبان (SVM)استفاده کردند و در مجموعه داده هستی‌شناسی احساسات بصری (VSO) با دقت ۰،۸۶ در موضوعات مختلف نتایج قابل‌توجهی به‌دست آوردند.

مقاله [۲٤] نشان‌می‌دهد که ANP ها برای تشخیص خودکار احساسات و تحلیل احساسات توسط تصاویر استفاده می‌شوند. استفاده از ANP ها به تعیین اطلاعات احساسی و عاطفی کمک می‌کند و نسبت به توصیف‌گرهای تصویر سطح پایین، نتایج بهتری به‌دست می‌آید. محققان در [۲۵] از مفاهیم شبکه‌های عصبی عمیق و جفت صفت و اسم برای غلبه بر برخی از چالش‌های تجزیه‌وتحلیل احساسات تصاویر استفاده کردند. آن‌ها با ترکیب شبکه عصبی کانولوشنال (CNN) با شبکه‌های جفت صفت و اسم مجزا به نتایج بهتری رسیدند. در مقاله [۲٦]، محققان از برخی مناطق محلی و کل تصاویر که اطلاعات احساسات را با استفاده از شبکه عصبی کانولوشنال (CNN) دارند، برای به‌دست آوردن امتیازات احساسی استفاده کردند. معماری ترکیبی از CNN با یک مدل توجه بصری توسط نویسندگان در [۲۷] برای تشخیص احساسات تصویر در مجموعه داده‌های عکس توئیتر و ART پیشنهاد شد. در مقاله دیگری [۲۸]، تجزیه‌وتحلیل احساسات تصویر به‌وسیله مشتق کردن توصیفات متنی از تصاویر و استفاده از یک طبقه‌بندی کننده ماشین بردار پشتیبان (SVM) برای تعیین قطبیت احساسات آموزش داده شد. آن‌ها از یک مجموعه داده با ٤۷،۲۳۵ تصویر استفاده کردند و با ترکیب متن و ویژگی‌های بصری، به‌دقت ۷۳،۹٦ درصد رسیدند. همچنین، در [۲۹]، نویسندگاآن‌یک مدل حافظه کوتاه‌مدت مبتنی بر یادگیری عمیق (LSTM) برای تجزیه‌وتحلیل احساسات تصویر در تصاویر فلیکر و مجموعه داده‌های توئیتر طراحی کردند. آن‌ها به‌ترتیب به‌دقت ۰،۸٤ و ۰،۷۵ رسیدند. همچنین، مقاله [۳۰] ویژگی‌های معنایی و بصری را برای تجزیه‌وتحلیل احساسات تصویر معرفی کرد که باعث بهبود نتایج شد.

در مقاله [۳٤] هدف اصلی پژوهش، شناسایی و تحلیل احساسات ایجادشده توسط تصاویر است، به‌ویژه در شبکه‌های اجتماعی. مقاله از روش‌های یادگیری عمیق مانند AlexNet و ResNet50 استفاده می‌کند و کاربردهای آن را درزمینه‌هایی مانند بازاریابی هدفمند و علوم سیاسی مورد بررسی قرار می‌دهد.

۳- معرفی روش پیشنهادی

مراحل روش پیشنهادی برای پیش‌بینی احساسات با استفاده از یک مدل یادگیری انتقال در شکل ۱ نمایش داده‌شده است. برای این منظور، مدل DenseNet-121 که از پیش آموزش‌دیده‌است، باید به‌دقت تنظیم شود. برای بهبود عملکرد سیستم و کاهش نرخ خطا، از اتصال پرش استفاده می‌شود و از یک لایه نرمال‌سازی دسته‌ای قبل از لایه‌های وزن استفاده می‌شود. انتخاب مدل DenseNet-121به‌دلیل کاهش مشکل گرادیان ناپدید شدن، نیاز به پارامترهای کمتر و قابلیت استفاده مجدد از ویژگی‌ها صورت گرفته است. روش پیشنهادی شامل چهار مرحله است.

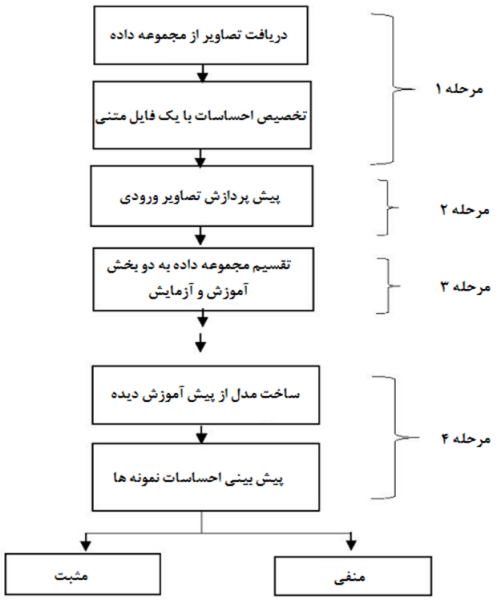
در مرحله اول، تصاویری که از مجموعه داده Image Sentiment استخراج‌شده‌اند، همراه با برچسب‌های احساسی آن‌ها از یک فایل متنی بارگذاری می‌شوند.

در مرحله دوم، پیش‌پردازش تصاویر با تبدیل آن‌ها به فرمت RGB انجام‌می‌شود و تغییر اندازه تصاویر در ابعاد مختلف بر اساس نیاز مدل از پیش آموزش‌دیده صورت‌می‌گیرد. مدل نیاز به تصاویر با

ابعاد ۲۲٤ * ۲۲٤ * ۳ دارد و در مرحله بعد، نرمال‌سازی پیکسل‌های تصویر انجام‌می‌شود.

در مرحله سوم، نمونه‌های آموزشی و آزمایشی با تقسیم تصاویر ورودی به دودسته آموزش و آزمایش تقسیم می‌شوند. حدود ۸۰ درصد از نمونه‌ها برای آموزش و ۲۰ درصد باقی‌مانده برای آزمایش استفاده می‌شود.

در مرحله چهارم، فرآیند آموزش با ساختار مدل از پیش آموزش‌دیده با لایه‌های اضافی آغاز می‌شود و در مرحله بعد، پیش‌بینی برای نمونه‌های ناشناخته انجام‌می‌شود. مدل به‌طور دقیق تنظیم می‌شود و درنهایت، معیارهای عملکرد مدل مانند دقت، صحت، فراخوانی و معیار F1 محاسبه می‌شوند و با نتایج مقالات دیگر مقایسه می‌شوند.



شکل (۱) : مراحل روش پیشنهادی در پیش‌بینی احساسات

٤- معماری روش پیشنهادی

در این بخش، نیاز به مفهوم یادگیری انتقالی موردبحث قرار می‌گیرد و معماری مدل از پیش آموزش‌دیده‌ای که برای پیش‌بینی احساسات تصویر به کار گرفته‌ایم، ارائه می‌شود.

٤-١ پیش‌پردازش تصاویر

قبل از ایجاد یک مدل یادگیری، لازم است تصاویر را پیش‌پردازش کرده و ابعاد آن‌ها را کاهش دهیم تا مدل یادگیری بهتری بسازیم. درروش پیشنهادی، ابتدا با استفاده از مدل رنگی RGBr ، اجزای رنگی تصاویر را کاهش می‌دهیم و ویژگی‌های رنگ فشرده را از مدل رنگ جدید استخراج می‌کنیم. یکی از روش‌های معمول برای کاهش ابعاد تصاویر، تجزیه‌وتحلیل اجزای اصلی (PCA) است که معیار خطای میانگین مربع را بهینه می‌کند. اما برای به‌دست آوردن بردارهای پایه وابسته به داده‌ها، نیاز به روشی زمان‌بر است. وقتی داده‌ها در یک فضای با ابعاد بالا قرار دارند، استفاده ازاین‌روش محاسباتی ناکارآمدتر می‌شود.

برای غلبه بر مشکلات PCA ، از تبدیل کسینوس معکوس (DCT) برای کاهش ابعاد تصاویر جزء رنگی مدل RGBr استفاده می‌کنیم. با توجه به اینکه بردارهای پایه DCT ثابت هستند، DCTقادر است به‌طور قابل‌توجهی بهبود کارایی محاسباتی در کاهش ابعاد ارائه دهد. روش DCT تصاویر جزء رنگی مدل RGBr را از حوزه مکانی به حوزه فرکانس تبدیل می‌دهد. با توجه به اینکه ویژگی‌های فرکانس پایین در حوزه فرکانس اطلاعات خوبی را نمایش می‌دهند، این روش برای تشکیل بردارهای الگوی کم بعدی جهت نمایش تصاویر جزء رنگی Rr ، Gr و Br انتخاب‌شده است. سپس سه بردار الگوی کم بعد به یکدیگر متصل می‌شوند تا یک بردار الگوی تقویت‌شده برای نمایش تصویر رنگی در مدل RGBr ایجاد شود. بردار الگوی تقویت‌شده حاوی اطلاعات فشرده تصویر RGBr است و برای استخراج ویژگی‌های رنگی تشخیصی بیشتر پردازش می‌شود. یکی از روش‌های رایج برای استخراج ویژگی، تجزیه‌وتحلیل تفکیک‌شده است که معیارهای تفکیک‌پذیری کلاس‌ها را بر اساس ماتریس‌های پراکندگی بهینه می‌کند. می‌توان ماتریس‌های پراکندگی درون کلاس و بین کلاس را به‌طور خوبی با رابطه (۱) برآورد کرد.

$$S_w = \sum_{i=1}^{L} \frac{N_i}{n} \frac{1}{N_i - 1} \sum_{j=1}^{N_j} (X_j^{(i)} - M_i)(X_j^{(i)} - M_i)^t$$

$$S_b = \sum_{i=1}^{L} \frac{N_i}{n}(M_i - M_0)(M_i - M_0)^t$$
$$M_i = \sum_{j=1}^{N_i} \frac{X_j^{(i)}}{N_i} \quad (۱)$$
$$M_0 = \sum_{i=1}^{L} \frac{N_i M_i}{n}$$

مقادیر ویژه ماتریس پراکندگی درون کلاس در مخرج ماتریس تبدیل کلی T قابل‌مشاهده هستند، که بردارهای طرح تجزیه‌وتحلیل تمایز را تعریف می‌کنند. درصورتی‌که بردار الگوی تقویت‌شده به طور مستقیم با تجزیه‌وتحلیل متمایز پردازش شود، مقادیر ویژه ماتریس پراکندگی درون کلاس شامل بسیاری از مقادیر ویژه دنباله‌ای می‌شوند. این مقادیر ویژه به‌دلیل کوچک بودن خود و ظاهر شدن در مخرج T، تمایل به جذب نویز و ایجاد برازش بیش‌ازحد در تجزیه‌وتحلیل دارند. برای حل این مشکل، تجزیه‌وتحلیل متمایز باید با استفاده از یک فرآیند کاهش ابعاد مانند PCA انجام شود. بنابراین، در هر مرحله، انتخاب ویژگی انجام می‌شود. برای این منظور، از الگوریتم SURF استفاده می‌شود. الگوریتم SURF دریافت فریم ورودی را انجام می‌دهد و از آن‌یک تصویر انتگرال می‌سازد و سپس نقاط کلیدی را شناسایی می‌کند. این نقاط بر اساس الگوریتم Hessian استخراج می‌شوند.

بعد از ورود نقاط ویژگی یا نقاط کلیدی به مرحله بعدی، هدف ما در این مرحله ایجاد یک فضای مقیاس است. ما قصد داریم نقاط بیشینه یا اکسترمم را در این فضا پیدا کنیم. این بخش از الگوریتم به‌شدت شبیه به الگوریتم SIFT است. یکی از بهبودهایی که در الگوریتم SURF صورت گرفته است، استفاده از انتگرال تصویر است. مزیت انتگرال تصویر در این است که سرعت عمل را افزایش می‌دهد. الگوریتم SURF سه تا پنج برابر سریع‌تر از الگوریتم SIFT عمل می‌کند. درنهایت، ما یک بردار ویژگی تشکیل می‌دهیم که بر اساس دوران‌ها حول نقاط ویژگی ما است. هر نقطه ویژه که به‌دست‌آمده است، باید دوران‌ها حول آن نقطه ویژه را به‌دست آوریم و درنهایت یک بردار ویژگی با ابعاد ٦٤ تشکیل دهیم. بردار ویژگی با ابعاد ٦٤ قابل‌توجه است. برخلاف بردار ویژگی SIFT که با ۱۲۸ عنصر کار می‌کند، در SURF این تعداد ٦٤ است و عملکرد الگوریتم SURF نسبت به SIFT بهتر است.

چگونگی محاسبه انتگرال تصویر:

برای محاسبه انتگرال تصویر، به‌عنوان‌مثال، می‌خواهیم شدت رنگ نقطه با مختصات y و x را به‌دست آوریم. اگر تصویر ما ۸ بیتی باشد، در نقطه I (x, y)، عددی بین ۰ تا ۲۵۵ وجود دارد که شدت رنگ را نشان‌می‌دهد. اما وقتی می‌خواهیم انتگرال تصویر این نقطه را محاسبه کنیم، مقدار داخل I (x, y) را نمی‌گیریم، بلکه مقدار جلوتر از آن را نمایش می‌دهد. به‌عبارت‌دیگر، مقادیری که در جلوی نقطه موردنظر قرار دارند را نشان‌می‌دهد.

$$\sum_{i \le x} \sum_{j \le y} I(i, j) \quad (۲)$$

که طبیعتاً این مقدار، مقدار بزرگی است. بعد از محاسبه انتگرال تصویر به مرحله دوم می‌رویم. در این مرحله با یک ماتریس ۲*۲ به نام ماتریس Hessian سروکار داریم. برای محاسبه ماتریس Hessian از فرمول ۳ استفاده می‌کنیم:

$$\begin{bmatrix} \frac{\partial I^2}{\partial x^2} & \frac{\partial I^2}{\partial x \partial y} \\ \frac{\partial I^2}{\partial y \partial x} & \frac{\partial I^2}{\partial y^2} \end{bmatrix} \quad (۳)$$

که I تصویر ما است و x و y مختصات پیکسل‌های آن بعدازاینکه این ماتریس را محاسبه کردیم، می‌توانیم محاسبه کنیم که این ماتریس چه اطلاعاتی را به ما نمایش دهد. در تصویر برای یک نقطه خاص ماتریس Hessian محاسبه‌شده است. چند حالت ممکن است برای آن وجود داشته‌باشد. گاهی امکان دارد در آن نقطه یک لبه افقی داشته‌باشیم در این حالت DYY تغییرات بیشتر است یعنی از یک سطح روشنایی به سطح روشنایی دیگر وارد می شود و مقدار تغییرات y خیلی بیشتر است. اگر محدوده موردنظر یک لبه عمودی باشد در این حالت DXX تغییرات بیشتر است و اگر در نقطه موردنظر یک‌گوشه یا corner داشته‌باشیم DXY بیشتر است.

یکی از بخش‌های اساسی تصویر پردازش، لبه‌ها هستند که به‌طور فراوان در تصاویر وجود دارند. لبه‌ها اطلاعات خاصی را به کاربر منتقل نمی‌کنند، به‌جز آنکه نقاطی را نشان می‌دهند که یک شکل را تشکیل می‌دهند، و هرگز ویژگی‌های خاصی از تصویر به ما ارائه نمی‌دهند. بنابراین، لبه‌ها به‌تنهایی نمی‌توانند عامل مفیدی برای ما

باشند. ماتریس حسین (Hessian) به ما مجموعه‌ای از نقاط کلیدی می‌دهد که بیشتر به‌عنوان کاندیدای گوشه‌ها شناخته می‌شوند. اما این گوشه‌ها به‌صورت دقیق‌ترین حالت ممکن به ما ارائه نمی‌دهند. درزمینه پردازش تصویر و بینایی ماشین، مفهومی به نام کالیبره کردن دوربین وجود دارد. درگذشته، برای کالیبره کردن دوربین، نقاط موردنیاز را به‌صورت دستی از یک تصویر استخراج می‌کردند. اما در حال حاضر الگوریتم‌هایی مانند الگوریتم هسین برای این کار استفاده می‌شوند که تنها با ارائه تصویر به الگوریتم، خود الگوریتم بر اساس گوشه‌های تصویر، گوشه‌های موردنظر را انتخاب می‌کند. با انتخاب گوشه‌ها، می‌توان فهمید کدام نقاط هم‌راستا هستند و بر اساس آن‌ها جدولی را تهیه‌کرده و فاصله بین آن‌ها را محاسبه کرد. بنابراین، الگوریتم هسین سعی می‌کند یک مجموعه ویژگی‌های کلیدی از تصویر به ما ارائه دهد که این ویژگی‌ها به‌طور قابل‌قبولی گوشه‌ها هستند. این همان الگوریتمی است که درروش sift نیز استفاده می‌شود تا نقاط ویژگی را پیدا کند. هنگام محاسبه مشتق نقاط بر اساس ماتریس حسین، مقدار دترمینان را با استفاده از فرمول ٤ محاسبه می‌کنیم.

$$\det(H) = D_{xx} \times D_{yy} - (0.9 \times D_{xy})^2 \qquad (٤)$$

ضریب ۰٫۹ عددی است که از طریق تجربه به‌دست‌آمده است.

تشکیل فضای مقیاس

ابتدا، قبل از هر چیز، باید یک فضای مقیاس را تشکیل دهیم. به این معنی که نیاز داریم یک سری تصاویر با ابعاد و اندازه‌های مختلف داشته‌باشیم. سپس از تابع گوسی برای اعمال اثر بلاورزی به تصویر استفاده می‌کنیم. هر دسته تصاویر، یک فضای مقیاس را شکل می‌دهد که ابعاد آن ممکن است متفاوت باشد. برای محاسبه الگوریتم، اهمیت دارد که تعداد تصاویر در هر دسته زیاد نباشد زیرا این موضوع می‌تواند سرعت اجرای الگوریتم را کاهش دهد. همچنین، زیاد کردن تعداد دسته‌های تصویر نیز به افزایش محاسبات منجر می‌شود. بنابراین، در تشکیل فضای مقیاس، باید به تعداد دسته‌ها و تعداد تصاویر در هر دسته توجه کنیم. در هر دسته، تصاویر باید سطوح تاری مختلفی داشته‌باشند. به‌عنوان‌مثال، یک تصویر را با ضریب x و تصویر دیگر را با ضریب kx و تصویر بعدی را با ضریب k2x و غیره به تابع گوسی می‌دهیم تا میزان تاری تصاویر متفاوت باشد.

$$L(x, y, \delta) = G(x, y, \delta) * I(x, y) \qquad (٥)$$

که در این فرمول علامت * نشانه کانولوشن است. G (x, y, δ) فیلتر گوسی است و I (x, y) تصویر ورودی است.

$$G(x, y, \delta) = \frac{1}{2\pi\delta^2} * e^{\frac{x^2+y^2}{2\delta^2}} \qquad (٦)$$

بعد از محاسبه مقادیر L با استفاده از فرمول ٥، باید هر ماتریس مجاور را از هم کم کنیم تا در مرحله بعد یک سری تصاویر را در اختیارمان قرار دهد که یکی کمتر از مرحله قبل است. این کاهش تصاویر باعث نمایش یک سری لبه می‌شود. وقتی یک تصویر تار را از یک تصویر کم‌تار کنیم، طبیعتاً یک تصویر تارتر به‌دست می‌آید که اکثر نقاط لبه را نشان‌می‌دهد. در مرحله بعد، هر سه تصویر را انتخاب و از هم کم می‌کنیم تا اکسترمم‌ها را محاسبه کنیم. برای دریافت نقطه اکسترمم از هر تصویر، یک نقطه را انتخاب کرده و این نقطه را با ۸ پیکسل اطراف آن و با ۹ پیکسل در همان موقعیت از تصاویر بالا و پایین خود مقایسه می‌کنیم. اگر این نقاط از همه نقاط کوچک‌تر یا بزرگ‌تر باشند، بعداً تشخیص می‌دهیم که آیا این نقطه، نقطه‌ای مهم است یا خیر. اما درSURF ، اگر نقطه موردنظر از تمام ٢٦ پیکسل اطراف بزرگ‌تر باشد، به‌عنوان نقطه اکسترمم در نظر گرفته‌می‌شود. این عملیات را برای تمام دسته‌ها انجام می‌دهیم، که در آن‌ها یک سری نقاط کاندید اکسترمم به‌دست می‌آید، سپس این نقاط را ترکیب کرده و درنهایت در یک تصویر نشان می‌دهیم. بعد از محاسبه تمام نقاط اکسترمم و داشتن تمام نقاط Hessian، دو مجموعه نقطه داریم: نقاط Hessian و نقاط اکسترمم. این هر دو مجموعه نقطه را در یک تصویر قرار می‌دهیم. سپس با استفاده از درون‌یابی، مختصات بین این نقاط را به‌دست می‌آوریم. نکته مهم این است که بهتر است این مرحله درون‌یابی انجام شود. برای درون‌یابی، از فرمول ۷ استفاده می‌کنیم:

$$H(x) = H + \frac{\partial H^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 H}{\partial x^2} x$$
$$\hat{x} = -\frac{\partial^2 H^{-1}}{\partial x^2} \frac{\partial H}{\partial x} \qquad (٧)$$
$$\frac{\partial^2 H}{\partial x^2} = \begin{bmatrix} d_{xx} & d_{yx} & d_{sx} \\ d_{xy} & d_{yy} & d_{sy} \\ d_{xs} & d_{ys} & d_{ss} \end{bmatrix}$$

برای تشکیل بردار ویژگی قبل از آن پارامتری به نام Rotation را محاسبه می‌کنیم. خود Rotation می‌تواند یک پارامتر مجزا کننده باشد. ناحیه‌ای که حول محور نقطه ویژگی انتخاب‌شده دوران داده‌شده است را برش می‌دهیم و ناحیه برش خورده را به ۱۶ مربع دسته‌بندی می‌کنیم و برای هر مربع ٤ مقدار را محاسبه می‌کنیم.

$$\sum \boldsymbol{dx}, \sum |\boldsymbol{dx}|, \sum \boldsymbol{dy}, \sum |\boldsymbol{dy}| \qquad (۸)$$

هرکدام از این‌ها یک مقدار را به ما می‌دهد. پس هر مربع ٤ مقدار دارد و درنهایت یک بردار به طول ٦٤ داریم که بردار ویژگی ما را تشکیل می‌دهد.

٤-۲ تقسیم‌بندی مجموعه‌های آموزش و آزمایش

تصاویر را به دو گروه، به‌منظور بخش‌های تست و آموزش (train)، تقسیم می‌کنیم. برای انجام این کار از روش تقسیم متقابل k-fold (k-fold cross validation) استفاده می‌کنیم. اعتبارسنجی متقابل، یک روش نمونه‌گیری مجدد است که برای ارزیابی مدل‌های یادگیری ماشین بر روی یک نمونه محدود از داده‌ها استفاده می‌شود. این روش دارای یک پارامتر واحد به نام k است که به تعداد گروه‌هایی اشاره دارد که یک نمونه داده معین باید به آن‌ها تقسیم شود. به‌این‌ترتیب، این روش اغلب با نام اعتبارسنجی متقاطع k-fold شناخته می‌شود. معیار عملکرد گزارش‌شده توسط اعتبارسنجی متقاطع k، برابر با میانگین مقادیر محاسبه‌شده در طول اجرای حلقه‌هاست. این رویکرد باعث هدر رفتن داده‌های زیادی نمی‌شود که این‌یک مزیت بزرگ در حل مسائلی مانند استنتاج معکوس است که در آن تعداد نمونه‌ها بسیار کم است.

٤-۳ یادگیری انتقالی

یادگیری انتقالی یکی از رویکردهای مهم در حوزه یادگیری عمیق است که به‌تازگی برای حل چالش‌های پیچیده در حوزه‌های پردازش زبان طبیعی (NLP) و بینایی کامپیوتری استفاده می‌شود. این رویکرد با کاهش زمان محاسبات و افزایش سرعت با استفاده از شبکه‌های عمیق، عملکرد مدل‌های طبقه‌بندی را بهبود می‌بخشد. در این روش، از یک مدلی استفاده می‌شود که در یک وظیفه شناخته‌شده و قابل‌مقایسه آموزش‌دیده‌است. این روش شامل یک فرآیند بهینه‌سازی است که اطلاعاتی که از حل یک مسئله قبلی به‌دست می‌آید، به یک چالش جدید منتقل می‌شود. بسیاری از مدل‌های مبتنی بر یادگیری انتقالی از پیش‌آموزش‌دیده در دسترس هستند که هر یک از آن‌ها بر روی مجموعه داده‌های بزرگی آموزش‌دیده‌شده‌اند که شامل میلیون‌ها تصویر آموزشی است. مدل‌های VGG-19 ، DenseNetو ResNet ازجمله مدل‌های محبوب هستند که بر روی مجموعه داده ImageNet آموزش‌دیده شده‌اند.

مدل یادگیری عمیق DenseNet به‌منظور غلبه بر مشکل ناپدید شدن گرادیان‌ها و بهبود دقت سیستم معرفی‌شده است. این مدل نسبت به مدل ResNet نیاز به تعداد کمتری پارامتر و فیلتر دارد (۱۲ فیلتر در هر لایه). اندازه تصویر ورودی پیش‌فرض برای این مدل ۲۲٤ * ۲۲٤ است و وزن‌های آن بر روی مجموعه داده ImageNet آموزش داده‌می‌شود. مدل DenseNet-121 شامل ۱۲۱ لایه، ٤ بلوک متراکم و لایه‌های انتقالی بین آن‌ها است. درون بلوک‌های متراکم، تعدادی لایه کانولوشن وجود دارد که نقشه‌های ویژگی متفاوتی را ایجاد می‌کنند. هدف لایه انتقالی در این مدل، نمونه‌برداری پایین‌تر از طریق یک روش عادی‌سازی دسته‌ای است. این کار به‌منظور حفظ نقشه‌های ویژگی تولیدشده توسط فرآیند Max-Pooling را به حداکثر امکان کوچک نگه‌می‌دارد. نقشه‌های ویژگی در هر بلوک متراکم در مدل DenseNet ابعاد یکسانی دارند و با نقشه‌های ویژگی مشتق‌شده از لایه‌های قبلی به‌هم متصل می‌شوند. این موضوع در رابطه ۹ نشان داده‌شده است.

$$\boldsymbol{x_l} = \boldsymbol{H_l}([\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_{l-1}}]) \qquad (۹)$$

به‌طوری‌که [x_1,x_2, …, x_ (l-1)] نشان‌دهنده نقشه‌های ویژگی(9)هم‌پیوسته به‌دست‌آمده از لایه‌های قبلی است، یعنی از لایه های ۰ تا L1. همچنین، H_l تابع تبدیل غیرخطی را نشان‌می‌دهد و "l" شماره لایه را نشان‌می‌دهد.

مدل DenseNet-121 شامل لایه‌های کانولوشن اولیه با اندازه خروجی ۱۱۲ * ۱۱۲ * ٦٤ و یک لایه حداکثر ادغام با اندازه خروجی ٥٦ * ٥٦ * ٦٤ است. این خروجی تولیدشده به‌عنوان ورودی برای اولین لایه متراکم استفاده می‌شود. یکی از پارامترهای مهم این مدل نرخ رشد "k" است که ابعاد کانال را در هر لایه تعیین می‌کند.

برای طبقه‌بندی مثبت و منفی، لایه‌های اضافی به معماری اضافه‌شده است. این لایه‌ها شامل یک لایه جمع‌آوری متوسط سراسری، یک لایه حذفی، یک لایه متراکم با ۵۱۲ واحد و تنظیم‌کننده L2، و یک لایه متراکم نهایی با ۲ واحد هستند. سپس با فریز کردن لایه‌های بلوک ۱ و بلوک ۲ شبکه و باز کردن لایه‌های باقی‌مانده (بلوک ۳ و بلوک ٤) که بر روی مجموعه داده احساسات تصویری ImageNet آموزش‌دیده‌شده‌اند، شبکه تنظیم می‌شود.

۵- مجموعه داده

در این پژوهش ما از مجموعه داده Image-Sentiment[1] استفاده کرده‌ایم.

"Image-Sentiment" یک مجموعه داده است که شامل تصاویر و برچسب‌های احساسی متناظر آن‌ها است. این دیتاست برای تحلیل و پیش‌بینی احساسات موجود در تصاویر استفاده می‌شود. تمرکز اصلی این دیتاست بر روی تفسیر و تحلیل احساسات مثبت، منفی و خنثی در تصاویر است.

این دیتاست شامل دو بخش اصلی است: تصاویر و برچسب‌های احساسی متناظر با هر تصویر. تصاویر موجود در این دیتاست از منابع مختلفی مانند وب، شبکه‌های اجتماعی و پایگاه‌های داده عمومی جمع‌آوری‌شده‌اند. برای هر تصویر، یک برچسب احساسی مشخص است که میزان احساس مثبت، منفی یا خنثی مرتبط با آن را نشان‌می‌دهد.

کلاس‌های احساسات مختلف شامل بسیار مثبت، مثبت، بسیار منفی، منفی و خنثی هستند. ما تصاویر را از این مجموعه داده دانلود می‌کنیم، که فقط شامل کلاس‌های مثبت و منفی است

بخشی از تصاویر موجود در این مجموعه داده در شکل ۲ نمایش داده‌شده است.

[1] https://datasets.simula.no/image-sentiment

شکل (۲): نمونه‌هایی از تصاویر موجود در مجموعه داده

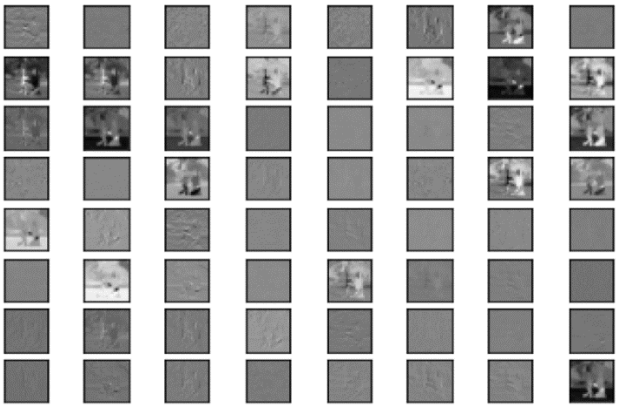
جدول (۱): مشخصات کلاس‌های موجود در مجموعه داده

نوع تگ	تعداد نمونه‌های موجود در مجموعه داده
Positive	803
Negative	2297
Neutral	579

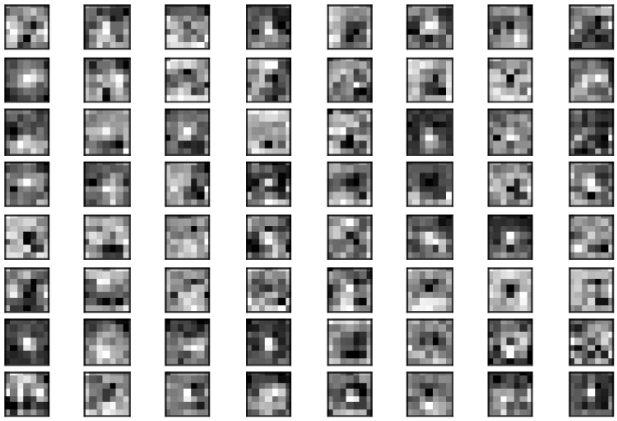
٦- نتایج ارزیابی

روش پیشنهادی با استفاده از زبان Python 3.5 پیاده‌سازی شده است و برای اجرای آن از یک GPU با سرعت ۲,۳۰ گیگاهرتز و ۱٦ گیگابایت رم بهره گرفته‌شده است. شکل ۳ و ٤ نمایش‌دهنده نقشه‌های ویژگی استخراج‌شده از لایه‌های کانولوشن اولیه و نهایی مدل DenseNet-121 است.

نتایج طبقه‌بندی مدل‌های مختلف از پیش آموزش‌دیده با استفاده از کتابخانه SciKit-learn و معیارهای مختلف عملکرد مانند دقت، صحت، فراخوانی و معیار F1 با استفاده از روابط زیر محاسبه می‌شود. این نتایج با مدل‌های ارائه‌شده در مراجع [۳۲] و [۳۳] که به‌ترتیب از روش VGGNet و ResNet استفاده کرده‌اند، مقایسه خواهد شد.



شکل (۳): نقشه ویژگی از Dense_Block-1 در DenseNet-121

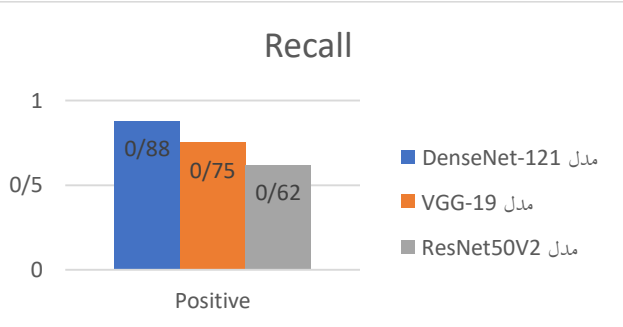


شکل (٤): نقشه ویژگی از Dense_Block-4 در DenseNet-121

معیار فراخوان

معیار فراخوان، از طریق رابطه ۱۰ محاسبه می‌شود.

$$\boldsymbol{Recall = \frac{TP}{TP + FN}} \qquad (۱۰)$$

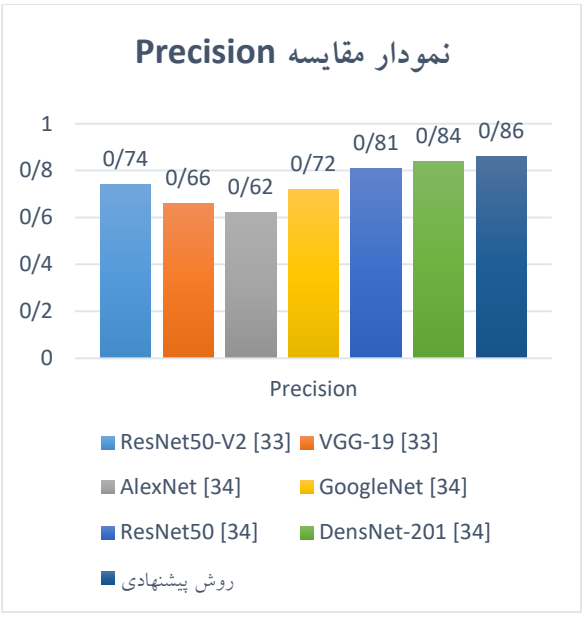


نمودار (۱): مقایسه مقدار فراخوانی در تشخیص داده‌های مثبت

معیار صحت

صحت یکی از مهم‌ترین معیارهای ارزیابی در الگوریتم‌های طبقه‌بندی است که از رابطه ۱۱ محاسبه می‌شود.

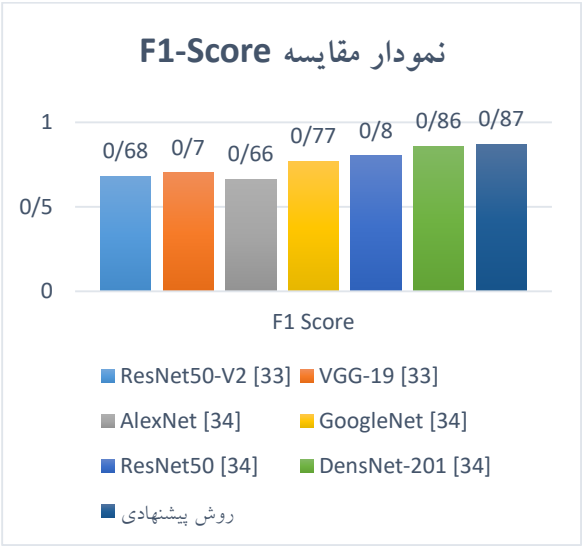
$$\boldsymbol{Precision = \frac{TP}{TP + FP}} \qquad (۱۱)$$



نمودار (۲): مقایسه مقدار صحت در تشخیص داده‌های مثبت

معیار F1-Score

این معیار از رابطه ۱۲ محاسبه می‌شود؛ که معیاری بین Recall و Precision است.



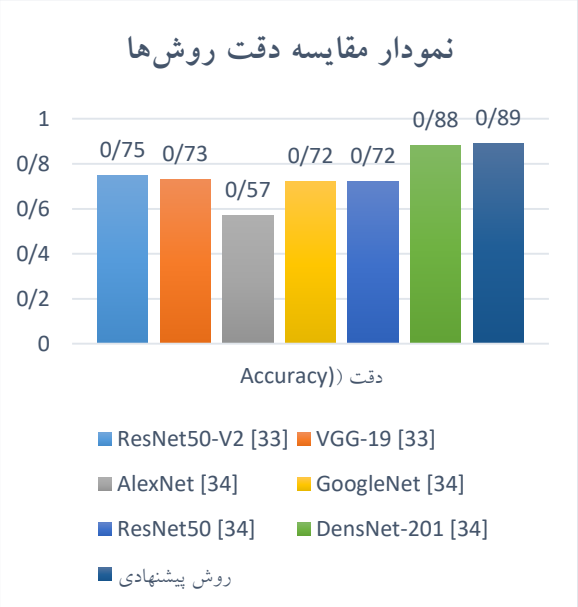
نمودار (۳): مقایسه مقدار F1 در تشخیص داده‌های مثبت

$$F_1 = 2\,\frac{Precision.recall}{Precision + recall} \qquad (۱۲)$$

معیار دقت

معیار دقت، از طریق رابطه ۱۳ محاسبه می‌شود.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (۱۳)$$



نمودار (٤): مقایسه مقدار دقت‌های مدل‌های مختلف در مقالات

با توجه به نتایج به‌دست‌آمده می‌توان نتیجه گرفت که روش پیشنهادی توانسته است نسبت به دو روش VGG-19 و ResNet50-V2 در دسته‌بندی داده‌ها به دو گروه مثبت و منفی بهتر عمل کند. همین‌طور با مقایسه روش پیشنهادی نسبت به روش DensNet-201 در مقاله [٣٤] می‌توان نتیجه گرفت که روش پیشنهادی ما، نیاز به محاسبات و حافظه کمتر است درعین‌حال دقت بیشتری نیز ارائه می‌دهد. علت این مسئله پیش‌پردازش مجموعه داده و نحوه پیاده‌سازی شبکه DensNet-121 است.

جدول (۲): مقایسه دقت روش پیشنهادی با سایر مقالات

روش مورداستفاده در مقاله	دقت
ResNet50-V2 [33]	0.75
VGG-19 [33]	0.73
AlexNet [34]	0.57
GoogleNet [34]	0.72
ResNet50 [34]	0.72
DensNet-201 [34]	0.88
روش پیشنهادی	0.89

در زیر نتایج جزئی‌تر روش پیشنهادی توسط معیارهای ارزیابی مختلف آورده شده است.

جدول (۳): نتایج جزئی به‌دست‌آمده از روش پیشنهادی

DenseNet-121			
Class	**Precision**	**Recall**	**F1 Score**
Positive	0.86	0.88	0.87
Negative	0.92	0.9	0.91
Accuracy	0.89		

جدول (٤): نتایج جزئی به‌دست‌آمده از روش [32] مدل VGG-19

VGG-19 در مقاله [32]			
Class	**Precision**	**Recall**	**F1 Score**
Positive	0.66	0.75	0.7
Negative	0.81	0.72	0.76
Accuracy	0.73		

جدول (٥): نتایج به‌دست‌آمده از روش [33] مدل ResNet50-V2

ResNet50-V2 در مقاله [33]			
Class	**Precision**	**Recall**	**F1 Score**
Positive	0.74	0.62	0.68
Negative	0.76	0.84	0.8
Accuracy	0.75		

۷- نتیجه‌گیری

در این مقاله، ما از مدل DenseNet-121 از پیش آموزش‌دیده برای پیش‌بینی احساسات از تصاویر استفاده کرده‌ایم. با اعمال بهبودهایی ازجمله اضافه کردن لایه‌های حذف، نرمال‌سازی دسته‌ای و تنظیم وزن، توانستیم اثر بیش برازش را کاهش داده و عملکرد مدل را در تجزیه‌وتحلیل احساسات بهبود دهیم.

یکی از نکات مهم تحقیق ما، مقایسه عملکرد مدلDenseNet-121 با سایر مدل‌های مطرح در این حوزه است. این مقایسه با استفاده از معیارهای مختلف انجام شد و نتایج نشان دادند که مدل DenseNet-121عملکرد بهتری نسبت به مدل‌های دیگر ارائه‌شده در مقالات مشابه داشته است. علاوه بر این، ما موفق به کنترل مشکل ناپدید شدن گرادیان‌ها با محاسبات کمتر شدیم که به عملکرد بهتر مدل کمک کرده است.

درنهایت، دقت مدل DenseNet-121 پیاده‌سازی شده در این مقاله در تجزیه‌وتحلیل احساسات تصویر در مجموعه داده ما به مقدار ۰٫۸۹ رسید. این نتیجه نشان از کیفیت بالا و قابل‌قبول مدل دارد. برای پژوهش‌های آتی، می‌توانیم به‌منظور بهبود عملکرد مدل، تعداد تصاویر استفاده‌شده برای آموزش مدل را افزایش داده و ترکیب چند روش برای بهبود نتایج را بررسی کنیم. این تحقیقات آینده ما را در راستای بهبود ادامه داده‌های ارائه‌شده در این مقاله هدایت خواهد کرد.

References

[1] Karray, F.; Alemzadeh, M.; Abou Saleh, J.; Nours Arab, M. Human-Computer Interaction: Overview on State of the Art. Int. J. Smart Sens. Intell. Syst. 2008, 1, 137–159.

[2] Auxier, B.; Anderson, M. Social media use in 2021. Pew Res. Cent. 2021.

[3] Sivarajah, U.; Kamal, M.M.; Irani, Z.; Weerakkody, V. Critical analysis of Big Data challenges and analytical methods. J. Bus. Res. 2017, 70, 263–286.

[4] Bansal, B.; Srivastava, S. On predicting elections with hybrid topic based sentiment analysis of tweets. Procedia Comput. Sci. 2018, 135, 346–353.

[5] El Alaoui, I.; Gahi, Y.; Messoussi, R.; Chaabi, Y.; Todoskoff, A.; Kobi, A. A novel adaptable approach for sentiment analysis on big social data. J. Big Data 2018, 5, 12.

[6] Drus, Z.; Khalid, H. Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. Procedia Comput. Sci. 2019, 161, 707–714.

[7] Zhao, H.; Liu, Z.; Yao, X.; Yang, Q. A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. Inf. Processing Manag. 2021, 58, 102656.

[8] Dashtipour, K.; Gogate, M.; Adeel, A.; Larijani, H.; Hussain, A. Sentiment Analysis of Persian Movie Reviews Using Deep Learning. Entropy 2021, 23, 596.

[9] Farisi, A.A.; Sibaroni, Y.; Faraby, S.A. Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. J. Phys. Conf. Ser. 2019, 1192, 012024.

[10] Melton, C.A.; Olusanya, O.A.; Ammar, N.; Shaban-Nejad, A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. J. Infect. Public Health 2021, 14, S1876034121002288.

[11] Mishra, N.; Jha, C.K. Classification of Opinion Mining Techniques. Int. J. Comput. Appl. 2012, 56, 1–6.

[12] Kim, M.; Lee, S.M.; Choi, S.; Kim, S.Y. Impact of visual information on online consumer review behavior: Evidence from a hotel booking website. J. Retail. Consum. Serv. 2021, 60, 102494.

[13] Xiao, Z.; Wang, L.; Du, J.Y. Improving the Performance of Sentiment Classification on Imbalanced DatasetsWith Transfer Learning. IEEE Access 2019, 7, 28281–28290.

[14] Praveen Gujjar, J.; Prasanna Kumar, H.R.; Chiplunkar, N.N. Image Classification and Prediction using Transfer Learning in Colab Notebook. Glob. Transit. Proc. 2021, 2, S2666285X21000960.

[15] Zhang, Q.; Yang, Q.; Zhang, X.; Bao, Q.; Su, J.; Liu, X.Waste image classification based on transfer learning and convolutional neural network. Waste Manag. 2021, 135, 150–157.

[16] Dilshad, S.; Singh, N.; Atif, M.; Hanif, A.; Yaqub, N.; Farooq, W.A.; Ahmad, H.; Chu, Y.; Masood, M.T. Automated image classification of chest X-rays of COVID-19 using deep transfer learning. Results Phys. 2021, 28, 104529.

[17] Siersdorfer, S.; Minack, E.; Deng, F.; Hare, J. Analyzing and predicting sentiment of images on the social web. In Proceedings of the International Conference on Multimedia MM '10, Firenze, Italy, 25–29 October 2010; pp. 715–718.

[18] Rao, T.; Xu, M.; Liu, H.; Wang, J.; Burnett, I. Multi-scale blocks based image emotion classification using multiple instance learning. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 634–638.

[19] Datta, R.; Joshi, D.; Li, J.; Wang, J.Z. Studying Aesthetics in Photographic Images Using a Computational Approach. In Computer Vision—ECCV 2006; Leonardis, A., Bischof, H., Pinz, A.,

[29] Xu, J.; Huang, F.; Zhang, X.; Wang, S.; Li, C.; Li, Z.; He, Y. Sentiment analysis of social images via hierarchical deep fusion of content and links. Appl. Soft Comput. 2019, 80, 387–399.
[30] Huang, F.; Zhang, X.; Zhao, Z.; Xu, J.; Li, Z. Image-text sentiment analysis via deep multimodal attentive fusion. Knowl. Based Syst. 2019, 167, 26–37.
[31] Deng, J.; Dong, W.; Socher, R.; Ki, L.; Li, K.; Fei-Fei, K. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, USA, 20–25 June 2009.
[32] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
[33] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
[34] Usha Kingsly Devi, K., & Gomathi, V. (2023). Deep Convolutional Neural Networks with Transfer Learning for Visual Sentiment Analysis. Neural Processing Letters, 55(4), 5087-5120.
[35] Chandrasekaran, G., Antoanela, N., Andrei, G., Monica, C., & Hemanth, J. (2022). Visual sentiment analysis using deep learning models with social media data. Applied Sciences, 12(3), 1030.
[36] An, J., Zainon, W. M. N. W., & Ding, B. (2023). Leveraging Vision-Language Pre-Trained Model and Contrastive Learning for Enhanced Multimodal Sentiment Analysis. Intelligent Automation & Soft Computing, 37(2).
Eds.; Springer: Berlin/Heidelberg, Germany, 2006; Volume 3953, pp. 288–301.
[20] Marchesotti, L.; Perronnin, F.; Larlus, D.; Csurka, G. Assessing the aesthetic quality of photographs using generic image descriptors. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 1784–1791.
[21] Borth, D.; Chen, T.; Ji, R.; Chang, S.-F. SentiBank: Large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In Proceedings of the 21st ACM International Conference on Multimedia—M '13, Barcelona, Spain, 21–25 October 2013; pp. 459–460.
[22] Yuan, J.; Mcdonough, S.; You, Q.; Luo, J. Sentribute: Image sentiment analysis from a mid-level perspective. In Proceedings of the Second InternationalWorkshop on Issues of Sentiment Discovery and Opinion Mining—WISDOM '13, Chicago, IL, USA, 11 August 2013; pp. 1–8.
[23] Zhao, Z.; Zhu, H.; Xue, Z.; Liu, Z.; Tian, J.; Chua, M.C.H.; Liu, M. An image-text consistency driven multimodal sentiment analysis approach for social media. Inf. Processing Manag. 2019, 56, 102097.
[24] Fernandez, D.; Woodward, A.; Campos, V.; Giro-i-Nieto, X.; Jou, B.; Chang, S.-F. More cat than cute? Interpretable Prediction of Adjective-Noun Pairs. In Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, Mountain View, CA, USA, 27 October 2017; pp. 61–69.
[25] Yang, J.; She, D.; Sun, M.; Cheng, M.-M.; Rosin, P.L.; Wang, L. Visual Sentiment Prediction Based on Automatic Discovery of Affective Regions. IEEE Trans. Multimed. 2018, 20, 2513–2525.
[26] Wang, J.; Fu, J.; Xu, Y.; Mei, T. Beyond Object Recognition: Visual Sentiment Analysis with Deep Coupled Adjective and Noun Neural Networks. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), New York, NY, USA, 9–15 July 2016; pp. 3484–3490.
[27] Song, K.; Yao, T.; Ling, Q.; Mei, T. Boosting image sentiment analysis with visual attention. Neurocomputing 2018, 312, 218–228.
[28] Ortis, A.; Farinella, G.M.; Torrisi, G.; Battiato, S. Visual Sentiment Analysis Based on on Objective Text Description of Images. In Proceedings of the 2018 International Conference on Content-Based Multimedia Indexing (CBMI), La Rochelle, France, 4–6 September 2018; pp. 1–6.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 1, No. 2, Autumn & Winter 2023, pp. 105-117

Improving Visual Sentiment Analysis in Social Networks using Transfer Learning Models

Mohammad Roustaei[1*], Meysam Mirzaee[2]

[1]Faculty and Research Institute of Artificial Intelligence and Cognitive Sciences, Imam Hossein Comprehensive University, Tehran, Iran

[2]Researcher at Imam Hossein Comprehensive University, Tehran, Iran

Article Information

Original Research Paper

Received:
2023 June 19

Accepted:
2023 November 18

Keywords:
Visual Sentiment Analysis, Transfer Learning, Deep Learning, Social Networks

Corresponding Author*:
M.Roustaei@ihu.ac.ir

Abstract

Analyzing individuals' emotions from the content of social media through text, speech, and images is necessary for various types of applications and purposes. Most recent research studies in the field of sentiment analysis have focused on textual data. However, social media users share more images and videos compared to text. In other words, images are the most effective way to convey emotions to others. Therefore, focusing on the development of a sentiment analysis model based on images in social media is important. In this article, we will use the DenseNet-121 transfer learning model to analyze emotions based on images. To implement this approach, we will utilize the images available in the Image Sentiment dataset. This dataset includes internet links to images along with their emotional polarities. Based on the obtained results, the accuracy of the proposed model in this article is 89%, which, compared to previous work in the field of visual sentiment analysis, shows a 5% to 10% improvement

: 10.22034/ABMIR.2023.20213.1028

E-ISSN: 2821-2037