دوره ۱۴، شماره بهار و تابستان ۱۴۰۰

مجله فناوری اطلاعات در طراحی مهندسی
Information Technology in Engineering Design
http://ited.sinaweb.net

طبقه‌بندی و تجزیه و تحلیل احساسات چند وجهی با استفاده از شبکه‌های کانولوشن وزن‌دار ترکیبی

علیرضا قربانعلی [۱] محمد کریم سهرابی *[۲] فرزین یغمایی [۳]

(۱) دانشکده مهندسی کامپیوتر، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران

(۲) دانشکده مهندسی کامپیوتر، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران*

(۳) دانشکده مهندسی کامپیوتر، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان

(تاریخ ارسال : ۱۳۹۹/۱۲/۱۴ تاریخ پذیرش : ۱۴۰۰/۰۴/۲۶)

چکیده

تجزیه و تحلیل احساسات در داده‌های رسانه‌های اجتماعی برای درک موقعیت، نگرش و نظر مردم نسبت به یک رویداد خاص ضروری است و کاربردهای بسیاری مانند پیش‌بینی انتخابات و ارزیابی محصولات را شامل می‌شود. تلاش‌های زیادی برای تحلیل احساسات در متون و تصاویر به صورت مجزا صورت گرفته است، اما در حوزه تحلیل احساسات اسناد مشتمل بر چند رسانه، که تحلیل احساسات چند وجهی نام دارد، پژوهش‌های کمتری صورت گرفته است. در این مقاله، استفاده از شبکه‌های کانولوشنی وزن‌دار ترکیبی برای طبقه‌بندی اسناد چند وجهی پیشنهاد شده است. شبکه‌های کانولوشنی وزن‌دار ترکیبی نوعی از شبکه‌های عصبی کانولوشنی هستند که در آنها خروجی شبکه‌های کانولوشنی مختلف با احتمالی با یکدیگر ترکیب می‌شوند. هر شبکه دارای وزن مخصوصی است که باعث می‌شود آن شبکه عملکرد مناسبی در زمان طبقه‌بندی داشته باشد. در برخی مسائل طبقه‌بندی، این شبکه‌ها نتایج بهتری را نسبت به رویکردهای معمولی، که از میانگین احتمال برای پیش‌بینی استفاده می‌کنند، به دست آورده‌اند. با توجه به استفاده از داده‌های چند وجهی، از همجوشی نیز در رویکرد پیشنهادی این مقاله استفاده شده است. به منظور تعبیه کلمات در تحلیل متنی احساسات نیز روش نمایش رمزگذار دو جهته از مبدل‌ها به کار گرفته شده است. مدل پیشنهادی بر روی مجموعه داده‌های دو پایگاه داده استاندارد ارزیابی شده است و نتایج این ارزیابی نشان داده است که استفاده از رویکرد پیشنهادی دقت مناسبی برای دسته بندی احساسات در داده‌های تک وجهی و چند وجهی خواهد داشت.

واژه‌های کلیدی: تجزیه و تحلیل احساسات، شبکه‌های کانولوشنی وزن‌دار ترکیبی، یادگیری عمیق، همجوشی، تعبیه کلمات

* عهده‌دار مکاتبات:

محمد کریم سهرابی

نشانی: دانشکده مهندسی کامپیوتر، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران

پست الکترونیکی: **Amir_sohraby@aut.ac.ir** تلفن : ۰۹۳۷۴۷۸۹۰۸۰

۱- مقدمه

تجزیه و تحلیل احساسات کاربردهای فراوانی در مسائل و چالش‌های گوناگون دارند و زمینه‌های استفاده از آن بسیار گسترده است. به عنوان مثال، شرکت‌ها علاقه‌مند به دانستن نظرات مشتریان نسبت به محصولات و تولیدات خود هستند و یا سهام‌داران می‌توانند از احساسات و نظرات مردم برای پیش‌بینی بازار سهام استفاده کنند. با افزایش محبوبیت رسانه‌های اجتماعی، داده‌های چند وجهی در پیام‌ها و اسناد حاوی چند رسانه، مانند متن‌های مشتمل بر تصاویر و صورتک‌ها، به طور فزاینده‌ای در شبکه‌های اجتماعی، مانند توییتر و اینستاگرام، در دسترس قرار گرفته است و تعداد افراد بیشتری نظرات خود را از طریق این نوع پیام‌ها بیان می‌کنند. بنابراین، تحلیل احساسات در داده‌های چند وجهی در مقیاس بزرگ می‌تواند به شناخت بهتر نگرش یا نظرات مردم نسبت به رویدادها یا موضوعات خاص کمک کند. چالش‌های متفاوتی در زمینه تحلیل احساسات چند وجهی وجود دارد. داده‌های چند وجهی در مقایسه با داده‌های تک وجهی دارای الگوهای متنوع‌تر و محتوای بصری و متون ناهمگن هستند و حتی ممکن است محتوای بصری و متون آنها حاوی اطلاعات معنایی متفاوتی باشند. برای تحلیل احساسات چند وجهی، مدل‌ها و روش‌های متفاوتی پیشنهاد شده است که هرکدام دارای مزایا و معایبی هستند. مدل‌ها و روش‌های گوناگونی از یادگیری عمیق نیز برای تحلیل احساسات در داده‌های یک وجهی و چند وجهی به کار گرفته شده است. شبکه‌های عصبی کانولوشنی، شبکه‌های عصبی بازگشتی، حافظه‌های کوتاه‌مدت طولانی، و واحدهای درگاهی بازگشتی نمونه‌هایی از این مدل‌ها و روش‌ها هستند. شبکه عصبی کانولوشنی یکی از معماری‌های مهم در یادگیری عمیق است که کاربردهای فراوانی به ویژه در پردازش تصاویر دارد. در این مقاله، استفاده از شبکه‌های کانولوشنی وزن‌دار ترکیبی منجر به دست‌یابی به کارایی بالاتری در تحلیل احساسات چند وجهی می‌گردد. شبکه‌های کانولوشنی وزن‌دار ترکیبی نوعی از شبکه‌های عصبی کانولوشنی هستند که در آنها خروجی شبکه‌های کانولوشنی با احتمال مشخصی با یکدیگر ترکیب می‌شوند [۱]. میانگین ترکیب چندین مدل، زمانی مفید واقع خواهد شد که مدل‌ها دارای ساختارهای متفاوت باشند و یا بر روی مجموعه داده‌های متفاوت آموزش ببینند. در روش ارائه شده در این مقاله، برای حل مساله ادغام داده‌های چند وجهی از روش همجوشی زودهنگام و پیش بینی دیرهنگام [۱] [۲] به کار گرفته شده، و نمایش رمزگذار دو جهته از مبدل‌ها [۳] برای تعبیه کلمات استفاده شده است.

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

- تجزیه و تحلیل احساسات چند وجهی با استفاده از شبکه‌های کانولوشنی وزن‌دار ترکیبی
- استفاده از همجوشی زودهنگام و پیش بینی دیرهنگام به منظور ادغام داده‌های چند وجهی
- استفاده از مدل نمایش رمزگذار دو جهته از مبدل‌ها برای تعبیه کلمات

ادامه مطالب مقاله به این صورت سازمان‌دهی شده‌اند: در بخش ۲ به مرور کارهای مرتبط پرداخته می‌شود. در بخش ۳ مساله و پیش‌نیازهای آن بیان می‌شوند، و علاوه بر ارائه تعاریف رسمی مرتبط با موضوع، پیش‌نیازهای لازم برای توضیح روش پیشنهادی نیز بررسی می‌گردند. روش پیشنهادی در بخش ۴ ارائه و توصیف می‌شود. در بخش ۵ نتایج پیاده‌سازی و اجرای روش پیشنهادی بر روی دو مجموعه داده متفاوت گزارش و ارزیابی می‌شوند، و در نهایت در بخش ۶، موارد مطرح شده در مقاله جمع‌بندی شده و نتیجه‌گیری نهایی ارائه می‌گردد.

۲- کارهای مرتبط

تحلیل احساسات مساله مهمی در نظرکاوی[2] است که در سال‌های اخیر به سرعت توسعه یافته است [۴-۹] و از آن برای مجموعه وسیعی از

[1] Early Fusion and Late Prediction (EFLP)
[2] Opinion mining

کاربردها، از جمله پیش‌بینی انتخابات، پیش‌بینی بازار سهام و بهابازار، ارزیابی محصولات، و عملکرد گیشه فروش فیلم، استفاده شده است. پژوهش‌های مرتبط عمدتا شامل تحلیل احساسات متنی، تحلیل احساسات از روی تصاویر، و تجزیه و تحلیل احساسات چند وجهی هستند.

۲-۱- تحلیل احساسات متنی

تحلیل احساسات متنی یک حوزه تحقیقاتی در پردازش زبان طبیعی[3] است که به دو روش اصلی مبتنی بر واژگان و مبتنی بر یادگیری ماشین انجام می‌شود. به عنوان مثال، سه استراتژی هرس به منظور ایجاد یک فرهنگ لغت عاطفی در سطح کلمه برای تشخیص احساسات در متن‌های اجتماعی در [۱۰] ارائه شده است. تحلیل احساسات متنی در [۱۱]، با استفاده از یک مدل جنبه-دسته و بر مبنای رویکردی گسترش یافته از نمایش رمزگذار دو جهته از مبدل‌ها بر اساس بهینه‌سازی قدرتمند آنها در مرحله پیش‌آموزش انجام می‌شود. با ترکیب روش تخصیص پنهان دیریکله برای نمایش متون با شبکه عصبی کانولوشنی به همراه واحد درگاهی بازگشتی برای طبقه‌بندی، یک روش تحلیل احساسات متنی کارآمد نیز در [۱۲] ارائه شده است. ترکیب اطلاعات متنی با الگوی انتشار احساسات نیز منجر به ارائه روش دیگری برای تحلیل احساسات متنی در توییتر در [۱۳] شد. تبدیل جملات به بردارهای عددی، تکنیک به کار گرفته شده در [۱۴] برای افزایش کارایی پیش پردازش در تحلیل احساسات متنی بود. چارچوبی بر اساس اصطلاحات زبانی احتمالی در [۱۵] ارائه شد تا با در نظر گرفتن تنوع کاربردهای هر کلمه به تحلیل دقیق‌تر متون کوتاه بپردازد. شناسایی و تحلیل احساسات و عواطف از متن‌های توییتر هدف مقاله [۱۶] بوده که با محاسبه امتیاز نفوذ کاربران و به کارگیری نتایج به منظور توصیه‌های شخصی‌سازی شده برای سایر کاربران محقق شده است. در مقاله [۱۷] از ویژگی دامنه برای تجزیه و تحلیل احساسات در متون مهندسی نرم افزار استفاده شده است. تحلیل احساسات در متون میکروبلاگ‌های چینی نیز با به کارگیری دیکشنری توسعه یافته احساسات در [۱۸] انجام شده است.

۲-۲- تحلیل احساسات بصری و تصویر

تنظیم دقیق شبکه‌های عصبی کانولوشنی برای پیش‌بینی احساسات بصری در [۱۹] انجام شده است. تاثیر معنای اشیا در تصاویر برای تحلیل احساسات بصری در [۲۰] مورد توجه قرار گرفت. مدل همبستگی احساسات معانی اشیا تصویر که در این مقاله ارائه شد، از مدل بیزین برای طبقه‌بندی احساسات تصویری استفاده می‌کرد. با ترکیب چارچوب طبقه‌بندی احساسات شبکه عصبی کانولوشنی با توجهات بصری، شبکه احساسات مبتنی بر توجهات بصری در [۲۱] ساخته و ارائه شد. یک شبکه ضعیف آموزش دیده در [۲۲] برای تحلیل احساسات بصری از روی تصاویر بدون برچسب ارائه شده است. برای به دست آوردن احساسات ضمنی در تصاویر میکروبلاگ، یک مدل موضوعی احساسات بصری در [۲۳] ارائه شده است. در این مقاله، ابتدا با استفاده از هستی‌شناسی احساسات بصری، ویژگی‌های احساسات بصری به دست می‌آید و سپس با استفاده از تمام تصاویر آن موضوع، یک مدل موضوعی احساس بصری ایجاد می‌شود. انتخاب ویژگی‌های متمایز هستی‌شناسانه بصری با توجه به مدل موضوعی احساسات به عنوان مزیت اصلی رویکرد این مقاله معرفی شده است. با این استدلال که برخی از احساسات موجود در یک تصویر از اشیا برجسته موجود در آن تصویر قابل شناسایی و استخراج است، روش ارائه شده در [۲۴] به ترکیب اطلاعات سراسری و محلی تصویر برای تحلیل احساسات تصاویر پرداخته است. احساسات سراسری از کل تصویر و احساسات محلی از زیرتصاویری که با برش اشیا برجسته موجود در تصویر ایجاد شده‌اند، استخراج می‌شوند و با همجوشی این اطلاعات، تحلیل نهایی احساسات انجام می‌شود.

[3] Natural Language Processing (NLP)

۲-۳- تحلیل احساسات چند وجهی

در مقاله [۲۵] از متن و تصویر برای تحلیل احساسات چند وجهی با استفاده از مکانیزم توجه[4] جهت استخراج ویژگی‌های مناسب متن و تصویر و همجوشی‌های زودهنگام و دیرهنگام استفاده شده است. یک رویکرد همجوشی سلسله‌مراتبی برای بهبود مکانیزم همجوشی داده‌های رسانه‌های گوناگون در [۲۶] ارائه شده که در آن وجه‌های مختلف رسانه‌ای ابتدا دو به دو و سپس همه با هم همجوشی و ادغام می‌شوند. شبکه ارائه شده در [۲۷] توجه خود را به همبستگی معنایی محتوای هر دو وجه تصویر و متن معطوف نموده است و یک نمایش جمله آگاه از کلاس را برای پردازش متن‌ها به منظور بهره‌برداری کارآمد از وابستگی محتوای تصویری به محتوای متنی به کار گرفته است. روش ارائه شده در مقاله [۲۸] با استخراج ویژگی‌های سطح میانی و ادغام آنها با سایر ویژگی‌های متنی، بصری، و اجتماعی به تحلیل چند وجهی احساسات با به کارگیری رویکرد یادگیری ماشین پرداخته است. علاوه بر ارائه یک روش جنبه محور برای تحلیل چند وجهی احساسات، یک پایگاه داده بزرگ مقیاس از زوج‌های متن-تصویر نیز برای تحلیل چند وجهی جنبه محور احساسات نیز در مقاله [۲۹] ساخته و ارائه شده است. مدل همجوشی عمیق سلسله‌مراتبی، نوآوری مقاله [۳۰] برای تحلیل احساسات چند وجهی بود که به کشف همبستگی میان تصاویر، متون، و لینک‌های اجتماعی می‌پرداخت.

۳- بیان مساله و پیش نیازها

با توجه به حجم روزافزون محتوای تولید شده توسط کاربران در فضای وب، تحلیل احساسات یک ابزار مهم برای استخراج اطلاعات در مورد نظرات افراد است [۳۱] و با استفاده از طبقه‌بندی احساسات می‌توان دیدگاه مثبت یا منفی یک متن (مثلا یک نظر در بررسی یک محصول یا یک پست در یک وبلاگ) را شناسایی نمود. داده‌های تک وجهی از نظراتی تشکیل می‌شوند که تنها مشتمل بر یک نوع رسانه مانند متن یا تصویر هستند. افزایش مداوم تعداد این نوع نظرات در شبکه‌های اجتماعی منجر به ایجاد مجموعه‌های داده بسیار بزرگ و در مقیاس کلان‌داده می‌شود. داده‌های چند وجهی نیز همانند داده‌های تک وجهی از حجم عظیمی برخوردار هستند، اما تنوع کلان‌داده‌های چند وجهی بیش از سایر مشخصات آنها برجسته است [۳۲]. به همین دلیل روش‌های تحلیل احساسات در داده‌های چند وجهی، علاوه بر غلبه بر مساله حجم بالای داده، باید بتوانند مکانیزم‌های مناسبی برای مدیریت این تنوع داده‌ای نیز اتخاذ نمایند. از مهم‌ترین رویکردهای مواجهه با حجم و تنوع بالای داده در تحلیل احساسات چند وجهی، مدیریت مجزای هر یک از رسانه‌ها به صورت موازی و در نهایت ادغام و ترکیب خروجی‌ها با به کارگیری استراتژی‌های همجوشی داده‌ها است. همجوشی داده‌ها روشی اساسی برای استخراج داده‌های چند وجهی است و هدف آن ادغام داده‌هایی با توزیع مختلف، از منابع مختلف و یا با انواع مختلف در یک فضای اصلی است که در آن حالت‌های مختلف را می‌توان به صورت یکنواخت نشان داد [۳۲].

سه نوع استراتژی همجوشی در تحلیل احساسات چند وجهی به کار گرفته می‌شود [۲۶] که شامل همجوشی زودهنگام، همجوشی میان مدت و همجوشی دیرهنگام است. در همجوشی زودهنگام تمام ورودی‌ها در یک بردار در نظر گرفته می‌شوند و ممکن است منجر به بردارهای ورودی بزرگ شود که شامل تکرارها نیز می‌باشد. همجوشی دیرهنگام به تجمیع تصمیمات از طبقه‌بندی کننده‌های احساسی چندگانه می‌پردازد که هر کدام در شرایط جداگانه آموزش دیده‌اند (یعنی ورودی‌ها در طبقه‌بندهای جداگانه قرار می‌گیرند و در نهایت همه خروجی‌های طبقه‌بندها با یکدیگر ادغام می‌شوند). در این مقاله از همجوشی زودهنگام برای ادغام‌های محلی استفاده شده است.

شبکه‌های عصبی کانولوشنی در انجام طبقه‌بندی کارآمد هستند، اما محاسبات سنگین باعث می‌شود که آموزش آنها کند باشد [۱]. با استفاده از تکنیک‌های برنامه‌نویسی موازی و بهره‌گیری از قدرت پردازش واحدهای پردازش گرافیکی[5]، زمان آموزش این شبکه‌ها به اندازه قابل توجهی

[4] Attention
[5] Graphics Processing Unit (GPU)

کاهش می‌یابد و امکان آموزش شبکه‌های بزرگ و آموزش هم‌زمان چندین شبکه به منظور ترکیب و همجوشی نتایج آنها را فراهم می‌شود. رویکرد مبتنی بر همجوشی می‌تواند دقت طبقه‌بندی را به میزان قابل توجهی افزایش دهد و عملکرد مدل‌های عصبی را بهبود بخشد. از این رو در این مقاله از شبکه‌های کانولوشنی وزن‌دار ترکیبی برای طبقه‌بندی چند وجهی استفاده شده است.

یکی دیگر از موضوعاتی که در پردازش متن وجود دارد تعبیه کلمات است. استفاده از روش‌های یادگیری عمیق در پردازش زبان طبیعی، با نمایش کلمات به عنوان بردار در یک فضای پیوسته با ابعاد کم امکان‌پذیر است. در روش‌های سنتی، تعبیه کلمات ایستا بودند و هر کلمه بدون در نظر گرفتن زمینه محتوایی، یک بردار واحد داشت. در کارهای اخیرتر مانند تعبیه کلمات متنی با نمایش رمزگذار دو جهته از مبدل‌ها، بردارهای کلمات نسبت به زمینه محتوایی که در آن قرار دارند، حساس هستند [۳۳]. در این مقاله نیز برای تعبیه کلمات از نمایش رمزگذار دو جهته از مبدل‌ها استفاده شده است.

در شبکه‌های کانولوشنی وزن‌دار ترکیبی، برای شبکه‌های کانولوشنی مختلف وزن‌های مخصوصی در نظر گرفته می‌شود که بر روی داده‌های اعتبار سنجی[6] آنها اعمال می‌شوند. شبکه‌هایی که در هنگام آموزش دارای خطای کمتری بر روی این داده‌ها باشند، در زمان همجوشی وزن‌های بیشتری خواهند داشت. معماری این مدل در شکل ۱ نشان داده شده است. نتایج خروجی شبکه کانولوشنی j در وزن‌های اختصاص داده شده به آن شبکه (α_j)، مطابق رابطه (۱) ضرب می‌شوند:

$$S_i = \sum_{j=1}^{n} \alpha_j r_j(i) \quad (1)$$

بر اساس رابطه‌های (۲) و (۳)، دو روش متفاوت برای محاسبه وزن‌های α_j وجود دارد. در روش اول مطابق رابطه (۲) از میانگین وزن‌ها استفاده می‌شود که در آن Aj دقت مدل j روی داده‌های اعتبارسنجی است.

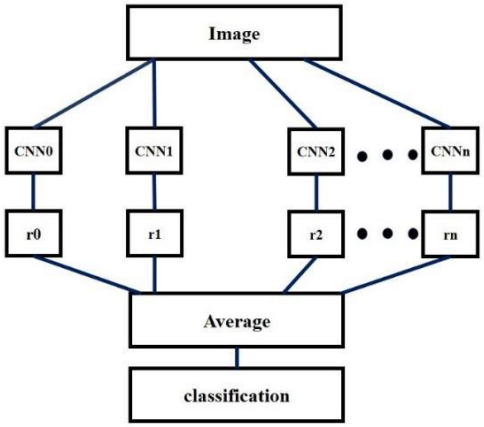
$$\alpha_j = \frac{A_j}{\sum_{i=1}^{n} A_i} \quad (2)$$

در روش دوم، مطابق رابطه (۳) وزن α_j توسط یک مکانیزم اهمیت‌دهی انتخاب می‌شود که در آن وزن‌ها براساس ترتیب دقتشان در مجموعه اعتبارسنجی انتخاب می‌شوند.

$$\alpha_j = \frac{R(A_j)}{\sum_{i=1}^{n} R(A_i)} \quad (3)$$

در رابطه (۳)، R تابعی است که موقعیت شبکه را براساس ترتیب افزایشی دقت نشان می‌دهد. مثلا مقدار R برای شبکه با بالاترین دقت برابر n، برای شبکه با دومین رتبه دقت برابر $n-1$ و برای شبکه با کمترین دقت برابر ۱ خواهد بود. در [۱] نشان داده شد که این شبکه‌ها بر روی مسائل طبقه‌بندی به دقت مطلوبی نسبت به روش‌های معمول دست یافته‌اند. شایان ذکر است که در [۱] شبکه‌های کانولوشنی وزن‌دار ترکیبی برای طبقه‌بندی تصاویر ارائه و مورد استفاده واقع شدند ولی رویکرد پیشنهادی در این مقاله، این شبکه‌ها را برای تحلیل احساسات چند وجهی مورد استفاده قرار خواهد داد. در این رویکرد خروجی هر شبکه را به صورت یک اسلاید[7] نشان می‌دهیم که این اسلایدها به صورت تسطیح شده به یک لایه همجوشی داده می‌شوند. لایه همجوشی بر روی ترکیب ویژگی‌های استخراج شده از شبکه‌های کانولوشنی آموزش داده می‌شود.

[6] Validation
[7] Slide



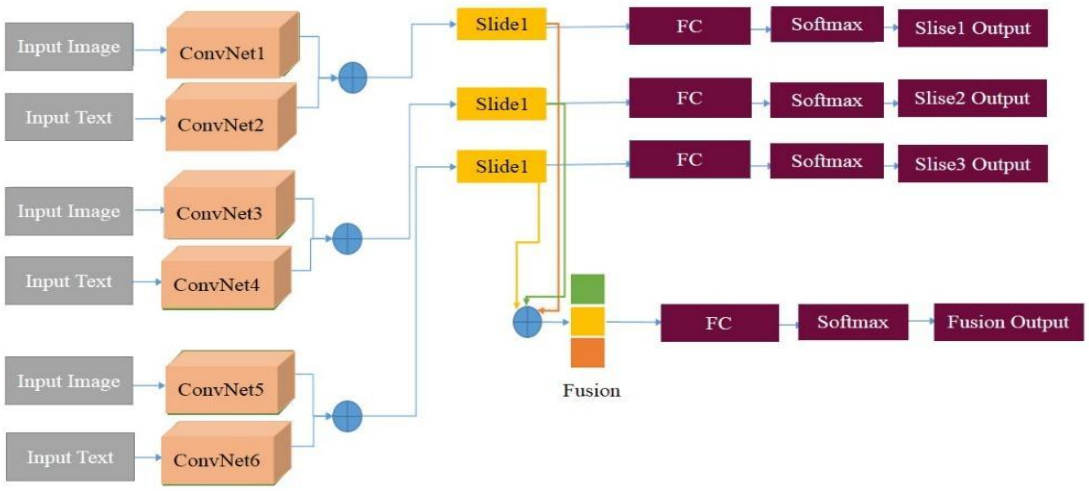
شکل ۱. معماری شبکه‌های کانولوشنی وزن‌دار ترکیبی

۴- روش پیشنهادی

رویکرد پیشنهادی این مقاله مبتنی بر شبکه‌های کانولوشنی وزن‌دار ترکیبی است و از سه گام اصلی، یعنی آماده‌سازی داده‌ها و آموزش شبکه کانولوشنی، محاسبه اسلایدها، و محاسبه همجوشی تشکیل شده است. در این بخش به توصیف هر یک از این گام‌ها خواهیم پرداخت.

۴-۱- آماده سازی داده‌ها و آموزش شبکه کانولوشنی

از آنجا که داده‌های مورد بررسی در مدل پیشنهادی این مقاله متنی و تصویری هستند، آماده‌سازی آنها به دو زیر بخش برای آماده‌سازی انشعاب‌های متنی و تصویری تقسیم شده و پس از این آماده‌سازی‌ها محاسبه اسلایدها و همجوشی انجام می‌شود. شکل ۲ نمای کلی از مدل پیشنهادی را نشان می‌دهد.



شکل ۲. مدل پیشنهادی برای تحلیل احساسات چند وجهی

۴-۱-۱- آماده‌سازی انشعاب متنی

به منظور استخراج ویژگی‌های نمایش رمزگذار دو جهته از مبدل‌ها، از کتابخانه Keras BERT Python استفاده می‌شود. در روش پیشنهادی این مقاله، از مدل قبلا آموزش دیده BERT (uncased_L-12_H-768_A-12) برای استخراج ویژگی‌های متنی استفاده شده است. پیش از این نیز در برخی مقالات اخیر از مدل قبلا آموزش دیده BERT برای استخراج ویژگی‌های متنی استفاده شده است. مثلا در [۳۴] که به بهبود تجزیه وتحلیل احساسات مبتنی بر جنبه[8]برای داده‌های متنی زبان فارسی می‌پردازد، از ورودی جفت جمله در مدل BERT از قبل آموزش‌دیده استفاده شده است.

۴-۱-۲- آماده‌سازی انشعاب تصویری

مشابه انشعاب متن، از شبکه کانولوشنی برای استخراج ویژگی‌های تصویری استفاده می‌شود. به دلیل محدودیت اندازه لایه کانولوشن، اندازه 128×128 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.

۴-۲- محاسبه اسلایدها

در این گام نتیجه تسطیح شده دو شبکه کانولوشنی آموزش داده شده بر روی انشعاب‌های مختلف ادغام می‌شوند و به صورت تسطیح شده در قالب یک اسلاید به یک شبکه کاملا متصل داده می‌شود. در واقع در این اسلاید فرآیند نرمال‌سازی دسته‌ای[9] روی خروجی‌های شبکه‌های کانولوشنی اتفاق می‌افتد. نرمال‌سازی دسته‌ای روشی برای بهبود سرعت، عملکرد و پایداری شبکه‌های عصبی مصنوعی است که در [۳۵]ارائه شده است. برای نرمال‌سازی، لایه ورودی با تمرکز مجدد و تغییر اندازه مجدد استفاده می‌شود. در مرحله آموزش شبکه‌ها، با تغییر پارامترهای لایه‌های قبلی، توزیع ورودی‌ها به لایه فعلی به ترتیب تغییر می‌کند، به گونه‌ای که لایه فعلی نیاز به تعدیل مداوم با توزیع‌های جدید دارد. این مشکل به ویژه برای شبکه‌های عمیق شدیدتر می‌شود زیرا تغییرات کوچک در لایه‌های پنهان کم‌عمق تقویت می‌شوند و در نتیجه تغییر قابل توجهی در لایه‌های پنهان عمیق‌تر ایجاد می‌شود. بنابراین، روش نرمال‌سازی دسته‌ای برای کاهش این شیفت‌های ناخواسته برای سرعت بخشیدن به آموزش و تولید مدل‌های قابل اطمینان‌تر پیشنهاد شده است.

۴-۳- محاسبه همجوشی

استراتژی پیش‌بینی زودهنگام و همجوشی دیرهنگام[10] یکی از استراتژی‌های مورد استفاده برای همجوشی در تحلیل احساسات چند وجهی است [۳۶-۳۷]. در این مقاله از استراتژی همجوشی زودهنگام و پیش‌بینی دیرهنگام استفاده می‌شود. این استراتژی ابتدا ویژگی‌های استخراجی را از شاخه‌های جانبی جدا می‌کند و یک ویژگی نرمال شده را به دست می‌آورد و سپس از یک لایه کاملاً متصل به دنبال ویژگی استخراج شده برای پیش‌بینی نهایی استفاده می‌کند.

[8] Aspect-based sentiment analysis
[9] Batch Normalization

الگوریتم عملکرد شبکه کانولوشنی در روش پیشنهادی در الگوریتم ۱ نشان داده شده است.

Algorithm 1. The proposed convolutional network model

Data: T input images$[I_1, I_2, \dots, I_n]$, T input text $[T_1, T_2, \dots, T_n]$, T target label $[l_1, l_2, \dots, l_3]$
Trained CNN weight matrices
While $epoch\ r: 1 \rightarrow R$ **do**
Begin
While $training\ image\ number\ i: 1 \rightarrow I\ and\ training\ text\ number\ t: 1 \rightarrow T$ **do**
Begin
Compute the J hidden activation matrices $Z1, \dots, ZJ$
Compute error $y_t - a$
Back-propagate error to compute
Cross entropy gradient $\nabla_\theta E(\theta)$
Update network parameters θ
Using gradient descent
Compute Slides
Compute Slides
End
End

الگوریتم ۱. الگوریتم شبکه کانولوشنی

۵- نتایج آزمایشات و ارزیابی‌ها

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

۵-۱- مجموعه داده

مدل ارائه شده بر روی دو مجموعه داده استاندارد Cifar و SemEval آزمایش شده است.

- مجموعه داده Cifar [۳۸]: نسخه ۱۰ این دیتاست شامل ۶۰۰۰۰ تصویر رنگی 32×32 پیکسل می‌باشد که در ۱۰ کلاس دسته‌بندی شده است. در آزمایش‌های این مقاله، ۵۰ هزار تصویر را برای آموزش و ۱۰ هزار تصویر را برای آزمایش استفاده کرده‌ایم.
- مجموعه داده SemEval [۳۹]: این مجموعه داده از چالش‌های نیمه تمام موسوم به تحلیل حافظه[11] در سال ۲۰۲۰ تشکیل شده است[12] که شامل ۷۰۰۰ هزار تصویر به همراه متن می‌باشد.

۵-۲- نتایج ارزیابی

جدول ۱ نتایج حاصل از پیاده سازی روش پیشنهادی بر روی مجموعه داده SemEval است. همان‌طور که در جدول ۱ مشاهده می‌شود، زمانی که عملیات شبکه‌های کانولوشنی وزن‌دار ترکیبی، همجوشی و اسلاید انجام می‌شود دقت پردازش بالاتر از زمانی است که داده‌های تصویری و متنی به صورت مجزا پردازش می‌شوند. جدول ۲ نتایج اجرای مدل پیشنهادی بر روی مجموعه داده تصاویر Cifar-10 می‌باشد. همان‌طور که

[11] Memotion Analysis
[12] https://competitions.codalab.org/competitions/20629

در جدول ۲ مشاهده می‌گردد، استفاده از عملیات شبکه‌های کانولوشنی وزن‌دار ترکیبی، همجوشی و اسلاید باعث بهبود دقت مدل نسبت به حالتی می‌شود که خروجی به صورت جداگانه از هریک از شبکه‌های کانولوشنی گرفته شود. در این رویکرد خروجی که به هر اسلاید داده می‌شود در واقع خروجی شبکه کانولوشنی است که کاملا مسطح گردیده است.

جدول ۱. نتایج اجرای مدل پیشنهادی بر روی مجموعه داده SemEval

SemEval	دقت
Image-convent	0.87
BERT-convent	0.90
Fusion	0.93

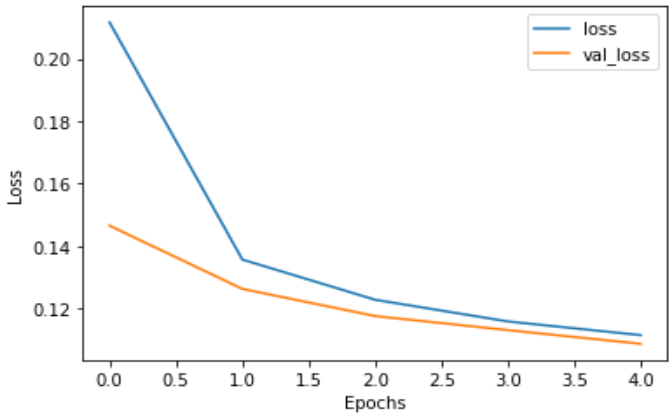
جدول ۲. نتایج اجرای مدل پیشنهادی بر روی مجموعه داده Cifar-10

Cifar-10	دقت
Slide1	0.9416
Slide2	0.9416
Slide3	0.9369
Fusion	0.9449

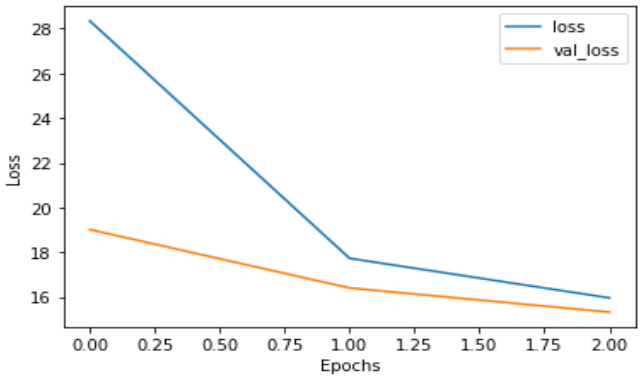
نمودار خطای مدل پیشنهادی برای مجموعه داده Cifar-10 در شکل ۳ آورده شده است. در این شکل، نمودارهای خطای داده‌های آموزش و اعتبار سنجی در دوره‌های آموزش مدل نشان داده شده است. مقدار خطا[۱۳] برای داده‌های آموزشی[۱۴] و اعتبارسنجی[۱۵] محاسبه می‌شود. مقدار خطا مانند مقدار دقت به صورت درصدی ارائه نمی‌شود و میزان خطای کلی نمونه‌ها خواهد بود. در شبکه‌های عصبی هدف اصلی کاهش مقدار خطا از طریق اصلاح وزن‌ها است که این اصلاح وزن از طریق بهینه سازها و به کمک الگوریتم‌های پس‌انتشار [۱۶] انجام می‌شود. در شکل ۳ نمودارهای خطای داده‌های آموزش و اعتبار سنجی در پنج دوره آموزش مدل نشان داده شده است، در این جا ۸۰ درصد نمونه‌ها به عنوان داده آموزشی و ۲۰ درصد آن به عنوان داده اعتبارسنجی مورد استفاده قرار گرفته است. همان‌طور که مشاهده می‌شود با افزایش دوره‌های آموزش، میزان خطای مدل کاهش یافته است و بیش‌برازش[۱۷] نشده است.

شکل ۴ خطای مدل همجوشی پیشنهادی بر روی مجموعه داده SemEval را نشان می‌دهد. در این شکل نمودارهای خطای داده‌های آموزش و اعتبار سنجی در سه دوره آموزش مدل نشان داده شده است، در این جا نیز ۸۰ درصد نمونه‌ها به عنوان داده آموزشی و ۲۰ درصد آن به عنوان داده اعتبارسنجی مورد استفاده قرار گرفته است. این شکل نیز دارای بیش‌برازش نبوده و با افزایش دوره‌های آموزش، خطای داده‌های آموزش اعتبارسنجی کاهش پیاده کرده است.

[13] Loss
[14] Train
[15] Validation
[16] Back propagation
[17] Overfit



شکل ۳. نمودار خطای داده‌های آموزشی و اعتبارسنجی مدل همجوشی بر روی مجموعه داده Cifar-10



شکل ۴. نمودار خطای داده‌های آموزشی و اعتبارسنجی مدل همجوشی بر روی مجموعه داده SemEval

۶- نتیجه‌گیری

تجزیه و تحلیل احساسات یکی از کاربردهای پردازش زبان طبیعی است که همواره مورد استفاده و توجه محققان قرار گرفته است. تحلیل احساسات چند وجهی، منجر به دست‌یابی به اطلاعات کامل‌تری می‌گردد. شبکه‌های یادگیری عمیق از مفیدترین ابزارهای به کار گرفته شده برای تحلیل احساسات چند وجهی بوده‌اند. استفاده از شبکه‌های ترکیبی باعث افزایش کارایی در شبکه‌های یادگیری عمیق می‌شوند. در این مقاله، با استفاده از شبکه‌های کانولوشنی وزن‌دار ترکیبی در دیتاست‌های تک وجهی و چند وجهی، به همراه استفاده از همجوشی و تعبیه کلمات با به کارگیری نمایش رمزگذار دو جهته از مبدل‌ها، به دقت ۹۴ درصد و بر روی داده‌های تصویری، و ۹۳ درصد و بر روی داده‌های چند وجهی متن و تصویر دست یافتیم که نشان می‌دهد ترکیب این مدل‌ها برای پیش‌بینی احساسات در تصاویر و متن مطلوب بوده است. نتایج این مطالعه می‌تواند به مدیران، تصمیم‌گیران، صاحبان و تولیدکنندگان محصولات و خدمات به منظور شناخت نظرات و بهینه کردن خدمات و محصولات خود کمک کند.

مراجع

1. Frazão X., Alexandre L.A. (2014) Weighted Convolutional Neural Network Ensemble. In: Bayro-Corrochano E., Hancock E. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2014. Lecture Notes in Computer Science, vol 8827. Springer, Cham. https://doi.org/10.1007/978-3-319-12568-8_82
2. Liu Y., Guo Y., S. Lew M. (2017) On the Exploration of Convolutional Fusion Networks for Visual Recognition. In: Amsaleg L., Guðmundsson G., Gurrin C., Jónsson B., Satoh S. (eds) MultiMedia Modeling. MMM 2017. Lecture Notes in Computer Science, vol 10132. Springer, Cham. https://doi.org/10.1007/978-3-319-51811-4_23
3. Devlin, J., Chang, M-W., Lee, K., Toutanova, K. (2019) Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics (ACL), pp. 4171–4186, https://doi.org/10.18653/v1/N19-1423.
4. Birjali, M., Kasri, M., Beni-Hssane, A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends, Knowledge-Based Systems, 226, 107134 (2021).
5. Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S-f., Pantic, M. A survey of multimodal sentiment analysis, Image and Vision Computing, 65, 3-14 (2017).
6. Ravi, K., Ravi, V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, Knowledge-Based Systems, 89, 14-46 (2015).
7. Hemmatian, F., Sohrabi, M.K. A survey on classification techniques for opinion mining and sentiment analysis. Artif Intell Rev 52, 1495–1545 (2019).
8. Schouten, K., Frasincar, F. Survey on Aspect-Level Sentiment Analysis, IEEE Transactions on Knowledge and Data Engineering, 28(3), 813-830 (2016).
9. Yue, L., Chen, W., Li, X. Zuo, W., Yin, M. A survey of sentiment analysis in social media. Knowl Inf Syst 60, 617–663 (2019).
10. Rao, Y., Lei, J., Wenyin, L. Li, Q., Chen, M. Building emotional dictionary for sentiment analysis of online news. World Wide Web 17, 723–742 (2014).
11. Liao, W., Zeng, B., Yin, X. Wei, P. An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. Appl Intell 51, 3522–3533 (2021).
12. Luo, Lx. Network text sentiment analysis method combining LDA text representation and GRU-CNN. Pers Ubiquit Comput 23, 405–412 (2019).
13. Wang, L., Niu J., Yu, S. SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis, IEEE Transactions on Knowledge and Data Engineering, 32(10), 2026-2039 (2020).
14. Sohrabi, M.K., Hemmatian, F. An efficient preprocessing method for supervised sentiment analysis by converting sentences to numerical vectors: a twitter case study. Multimed Tools Appl 78, 24863–24882 (2019).
15. Song, C., Wang, X-K. Cheng, P-F., Wang, J-Q., Li, L. SACPC: A framework based on probabilistic linguistic terms for short text sentiment analysis, Knowledge-Based Systems, 194, 105572 (2020).
16. Sailunaz, K., Alhajj, R. Emotion and sentiment analysis from Twitter text, Journal of Computational Science, 36, 101003 (2019).
17. Islam, M. R., Zibran, M. F. SentiStrength-SE: Exploiting domain specificity for improved sentiment analysis in software engineering text, Journal of Systems and Software, 145, 125-146 (2018).
18. Zhang, S., Wei, Z., Wang, Y., Liao, T. Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary, Future Generation Computer Systems, 81, 395-403 (2018).
19. Campos, V., Jou, B., Giro-i-Nieto, X. From pixels to sentiment: Fine-tuning CNNs for visual sentiment prediction. Image and Vision Computing, 65, 15-22 (2017).
20. Zhang, J., Chen, M., Sun, H., Li, D., Wang, Z. Object semantics sentiment correlation analysis enhanced image sentiment classification, Knowledge-Based Systems, 191, 105245 (2020).
21. Song, K., Yao, T., Ling, O., Mei, T. Boosting image sentiment analysis with visual attention, Neurocomputing, 312, 218-228 (2018).

22. Xue, L-Y., Mao, Q-R., Huang, X-H. Chen, J. NLWSNet: a weakly supervised network for visual sentiment analysis in mislabeled web images. Front Inform Technol Electron Eng 21, 1321–1333 (2020).
23. Cao, D., Ji, R., Lin, D. Li, S. Visual sentiment topic model based microblog image sentiment analysis. Multimed Tools Appl 75, 8955–8968 (2016).
24. Wu, L., Qi, M., Jian, M. Zhang, H. Visual Sentiment Analysis by Combining Global and Local Information. Neural Process Lett 51, 2063–2075 (2020).
25. Huang, F., Zhang, X., Zhao, Z., Xu, J., Li, Z. Image–text sentiment analysis via deep multimodal attentive fusion. Knowledge-Based Systems, 167, 26-37 (2019).
26. Majumder, M., Hazarika, D., Gelbukh, A., Cambria, E., Poria, S. Multimodal sentiment analysis using hierarchical fusion with context modeling, Knowledge-Based Systems, 161, 124-133 (2018).
27. Zhang, K., Zhu, Y., Zhang, W., Zhu, Y. Cross-modal image sentiment analysis via deep correlation of textual semantic, Knowledge-Based Systems, 216, 106803 (2021).
28. 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, Information Processing & Management, 56(6), 102097 (2019).
29. Zhou, J., Zhao, J., Huang, J. X., Hu, O. V., He, L. MASAD: A large-scale dataset for multimodal aspect-based sentiment analysis, Neurocomputing, 455, 47-58 (2021).
30. 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, Applied Soft Computing, 80, 387-399 (2019).
31. Do, H. H., Prasad P. W. C., Maag, A, Alsadoon, A., Deep learning for aspect-based sentiment analysis: a comparative review, Expert Systems with Applications, 118, 272-299 (2019).
32. Gao, J., et al., A survey on deep learning for multimodal data fusion. Neural Computation, 2020. 32(5): p. 829-864.
33. Ethayarajh, K., How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings. arXiv preprint arXiv:1909.00512 (2019).
34. Jafarian, H., Taghavi, A. H., Javaheri A., Rawassizadeh, R. Exploiting BERT to Improve Aspect-Based Sentiment Analysis Performance on Persian Language, 2021 7th International Conference on Web Research (ICWR), 2021, pp. 5-8, doi: 10.1109/ICWR51868.2021.9443131.
35. Ioffe, S. and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167, 2015.
36. Xie, S., Tu, Z. Holistically-nested edge detection, 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1395-1403, doi: 10.1109/ICCV.2015.164.
37. Yang, S., Ramanan, D. Multi-scale recognition with DAG-CNNs. 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1215-1223, doi: 10.1109/ICCV.2015.144.
38. Krizhevsky, A. (2009) Learning Multiple Layers of Features from Tiny Images. Technical Report TR-2009, University of Toronto, Toronto.
39. Sharma, C.a., et al., Task Report: Memotion Analysis 1.0 @SemEval 2020: The Visuo-Lingual Metaphor!