



انواع رمزگذارهای خودکار و کاربرد آنها در استخراج ویژگی و کاهش ابعاد تصاویر

فرزاد خدایاری

دانش آموخته کارشناسی ارشد هوش مصنوعی، دانشکده مهندسی کامپیوتر - دانشگاه تربیت دبیر شهید رجایی - تهران - ایران

احسان پازوکی

استادیار، دانشکده مهندسی کامپیوتر - دانشگاه تربیت دبیر شهید رجایی - تهران - ایران

چکیده

یک سری از چالش‌های برچسب‌گذاری تصویر، حجم بالای محاسبات در پردازش داده‌های تصویر و دقت روش‌های برچسب‌زدن است. با وجود روش‌های برچسب‌زنی تصویر با استفاده از شبکه‌های عمیق، هنوز کارایی این الگوریتم‌ها به حد مطلوب نرسیده است. و در کارهای بلادرنگ مشکل سرعت می‌تواند ناکارآمد باشد. کاهش ابعاد راهی برای کاهش پیچیدگی یک مدل و جلوگیری از برازش بیش از حد و رفع چالش مطرح شده است. برای کاهش ابعاد اولیه داده‌ها عمدتاً از روش‌های خطی استفاده می‌کنند. با این حال، از آنجایی که بسیاری از مشکلاتی که ما با آن مواجه هستیم سیستم‌های غیرخطی و متغیر با زمان هستند، تحقیقات فعلی در مورد روش‌های کاهش ویژگی غیرخطی نسبتاً بیشتر است.

رمزگذارهای خودکار نوعی شبکه عصبی هستند که یک نمایش ویژگی بهتر برای داده‌های ورودی با ابعاد بالا ایجاد، و همبستگی بین داده‌های ورودی را پیدا می‌کنند. رمزگذار خودکار به منظور کاهش ابعاد غیرخطی کارآمدتر از سایر روش‌های یادگیری چندگانه شناخته شده است. ما در این پژوهش، برای استخراج ویژگی از رمزگذارهای خودکار استفاده نمودیم و تصاویر مجموعه داده MIRFlickr را به معماری‌های مطرح شده که در Image Net کارایی بالایی داشته‌اند، داده شد اما مشخص شد هنوز هم این بردارهای ویژگی تصاویر با ابعاد بالا، برای الگوریتم‌های برچسب‌زنی خودکار تصاویر مناسب نیستند و پردازش آن‌ها باعث طولانی شدن زمان اجرای الگوریتم و همچنین کاهش دقت و کارایی الگوریتم خواهد بود. چون ما از روش مبتنی بر جستجو ۲PKNN برای برچسب‌گذاری خودکار تصاویر استفاده کردیم، بردارهای ویژگی با ابعاد بالا مشکل مزاحمت ابعاد دارد در نتیجه ما بردارهای ویژگی به دست آمده را به عنوان ورودی در اختیار رمزگذار خودکار پراکنده قرار دادیم که خروجی این روش، بردار ویژگی با ابعاد کم و مناسب برای الگوریتم مبتنی بر جستجو است. نتایج کاهش ابعاد در روش پیشنهادی استفاده از رمزگذار خودکار پراکنده معیار F^1 را نسبت به کارهای مطلوب قبلی به مقدار $19/0 \pm 0/3$ افزایش داد.

واژگان کلیدی: استخراج ویژگی، کاهش ابعاد، رمزگذار خودکار، مزاحمت ابعاد

۱- مقدمه

مجموعه داده‌هایی با ویژگی‌های زیاد، داده‌های با ابعاد بالا نامیده می‌شوند که اغلب اطلاعات اضافی زیادی، از جمله عوامل مرتبط یا تکراری در آن قرار دارد. کاهش ابعاد برای از بین بردن این تداخلات است. کاهش ابعاد ویژگی از پارامترهای موجود



برای تشکیل یک فضای ویژگی با ابعاد کم استفاده می‌کند و بر اثرات اطلاعات اضافی یا نامربوط غلبه می‌کند، بنابراین اطلاعات مؤثر موجود در ویژگی‌های اصلی را به ویژگی‌های کمتری ترسیم می‌کند (Jia et al, ۲۰۲۲). کاهش ابعاد راهی برای کاهش پیچیدگی یک مدل و جلوگیری از برازش بیش از حد است. دو دسته اصلی کاهش ابعاد وجود دارد: انتخاب ویژگی و استخراج ویژگی. داده‌های دارای ابعاد بالا برای الگوریتم‌های طبقه‌بندی به دلیل هزینه محاسباتی بالا و استفاده از حافظه مشکل‌ساز هستند (Khalid et al, ۲۰۱۴ ; Mishra and Sharma, ۲۰۲۱).

کاهش ابعاد راهی برای کاهش پیچیدگی یک مدل و جلوگیری از برازش بیش از حد و رفع چالش مطرح شده است. داده‌ها به طور فزاینده‌ای دارای ابعاد بالا مانند اطلاعات ژنومی، تصاویر، ویدئوها و متن هستند و کاهش ابعاد که برای تولید یک نمایندگی سطح بالا در نظر گرفته شده، ضروری است. مرحله پیش پردازش به این دلیل است که اگرچه یادگیری ماشینی مدل‌ها از نظر تئوری باید بتوانند روی هر عددی از ویژگی‌ها کار کنند، اما مجموعه داده‌های با ابعاد بالا همیشه باعث ایجاد یک مجموعه‌ای از مشکلات از جمله برازش بیش از حد، پیچیدگی محاسباتی بالا و مدل‌های بیش از حد پیچیده می‌شود که این مورد خود باعث یک موضوع شناخته شده به نام مزاحمت ابعاد می‌شود (Jia et al, ۲۰۲۲ ; Miyashiro, ۲۰۲۱). یکی دیگر دلیل از دلایل کاهش ابعاد این است که نمایش‌های سطح بالا می‌تواند به افراد کمک کند تا ساختار درونی داده‌ها را بهتر درک کنند.

با افزایش ابعاد داده‌ها، استخراج ویژگی اهمیت فزاینده‌ای پیدا می‌کند. با این حال، از آنجایی که بسیاری از مشکلاتی که ما با آن مواجه هستیم، سیستم‌های غیرخطی و متغیر با زمان هستند، تحقیقات فعلی در مورد روش‌های کاهش ویژگی غیرخطی نسبتاً بیشتر است (Jia et al, ۲۰۲۲). به عبارت دیگر فرایندی است که در آن با انجام عملیاتی بر روی داده‌ها، ویژگی‌های بارز و تعیین‌کننده آن مشخص می‌شود. روش‌های مبتنی بر استخراج ویژگی، یک فضای چندبعدی را به یک فضای با ابعاد کمتر نگاشت می‌دهند. این روش‌ها به دو دسته ی خطی و غیرخطی تقسیم می‌شوند. روش‌های خطی که ساده‌ترند و فهم آنها راحت‌تر است به دنبال یافتن یک زیرفضای تخت سراسری هستند. اما روش‌های غیرخطی که مشکل‌ترند و تحلیل آنها سخت‌تر است به دنبال یافتن یک زیرفضای تخت محلی می‌باشند. برای کاهش ابعاد اولیه داده‌ها عمدتاً از روش‌های خطی استفاده می‌کنند. با این حال، از آنجایی که بسیاری از مشکلاتی که ما با آن مواجه هستیم، سیستم‌های غیرخطی و متغیر با زمان هستند، تحقیقات فعلی در مورد روش‌های کاهش ویژگی غیرخطی نسبتاً بیشتر است (Mishra and Sharma, ۲۰۲۱). یک سری از چالش‌های برچسب‌گذاری تصویر، حجم بالای محاسبات در پردازش داده‌های تصویر و دقت روش‌های برچسب‌زدن است با وجود روش‌های برچسب‌زنی تصویر با استفاده از شبکه‌های عمیق، هنوز کارایی این الگوریتم‌ها به حد مطلوب نرسیده است و در کارهای بلادرنگ مشکل سرعت می‌تواند ناکارآمد باشد. در این مقاله برای رفع چالش‌های مربوط به برچسب‌گذاری تصویر، از روش‌های کاهش ابعاد مبتنی بر یادگیری عمیق با استفاده از رمزگذارهای خودکار استفاده نمودیم. ساختار مقاله به این صورت است که در قسمت دوم روش کاهش بعد عمیق شرح داده می‌شود و در ادامه ضمن بررسی انواع رمزگذارهای خودکار و مزایا و معایب آن‌ها در استخراج ویژگی و کاهش ابعاد، از روش کاهش ابعاد با رمزگذار خودکار پراکنده و رمزگذار خودکار پایه در مرحله استخراج ویژگی استفاده نمودیم سپس بردارهای ویژگی به دست آمده را در اختیار روش برچسب‌گذاری پیشنهادی قرار دادیم و در قسمت ششم نتایج آزمایشات بر روی مجموعه داده مربوطه، آمده است.

۲- کاهش ابعاد با یادگیری عمیق

کاهش ابعاد یک تکنیک مهم در یادگیری ماشینی و یادگیری عمیق است که تعداد ویژگی‌های یک مجموعه داده را کاهش می‌دهد و داده‌ها را قابل مدیریت‌تر و از نظر محاسباتی کارآمدتر می‌کند. کاهش ابعاد به ویژه در برخورد با داده‌های با ابعاد بالا مفید است. این داده‌های با ابعاد بالا می‌تواند منجر به پیچیدگی محاسباتی، تطبیق بیش از حد، و چالش‌های تجسمی شود، که در مجموع به عنوان مزاحمت ابعاد شناخته می‌شود. الگوریتم‌های یادگیری ماشینی و یادگیری عمیق به مقدار زیادی داده برای یادگیری تغییرپذیری، الگوها و نمایش‌ها نیاز دارند. اگر این داده‌ها شامل تعداد زیادی ویژگی باشد، این مورد می‌تواند به



مزامحت ابعاد منجر شود. مزامحت ابعادی که برای اولین بار توسط بلمن معرفی شد، توضیح می‌دهد که برای تخمین یک تابع دلخواه با دقت معین، تعداد ویژگی‌ها یا ابعاد مورد نیاز برای تخمین به‌طور تصاعدی رشد می‌کند. این مورد به ویژه در مورد داده‌های بزرگ که پراکندگی بیشتری را به همراه دارند، صادق است. پراکندگی در داده‌ها معمولاً به ویژگی‌هایی گفته می‌شود که دارای مقدار صفر هستند، این بدان معنا نیست که مقدار ازدست رفته است. اگر داده‌ها دارای ویژگی‌های پراکنده زیادی باشند، فضا و پیچیدگی محاسباتی افزایش می‌یابد. الیور کاس در سال ۲۰۰۲ نشان می‌دهد که مدل آموزش داده شده بر روی داده‌های پراکنده در مجموعه داده آزمایش ضعیف عمل می‌کند. به عبارت دیگر مدل در حین آموزش نویز را یاد می‌گیرد و نمی‌تواند به خوبی تعمیم دهند. وقتی داده‌ها پراکنده هستند، مشاهدات یا نمونه‌های موجود در مجموعه داده آموزشی به سختی خوشه‌بندی می‌شوند. زیرا داده‌های با ابعاد بالا باعث می‌شوند که هر مشاهدات در مجموعه داده با فاصله یکسان از یکدیگر ظاهر شوند. اگر داده‌ها معنی‌دار و غیر زائد باشند، مناطقی وجود خواهند داشت که نقاط داده مشابه با هم جمع شده و خوشه‌بندی می‌شوند، علاوه بر این، آنها باید از نظر آماری معنی‌دار باشند. از سوی دیگر داده‌های غیرپراکنده یا داده‌های متراکم داده‌هایی هستند که دارای ویژگی‌های غیرصفر هستند. جدا از اینکه حاوی ویژگی‌های غیرصفر هستند، آنها همچنین حاوی اطلاعاتی هستند که هم معنی‌دار و هم غیرزائد است. برای مقابله با مزامحت ابعاد، روش‌هایی مانند کاهش ابعاد استفاده می‌شود. تکنیک‌های کاهش ابعادی برای تبدیل ویژگی‌های پراکنده به ویژگی‌های متراکم بسیار مفید هستند. علاوه بر این، کاهش ابعاد نیز برای تمیزکردن داده‌ها و استخراج ویژگی استفاده می‌شود (Verleysen and François, ۲۰۰۵).

توسعه سریع فناوری اینترنت الزامات بالاتری را برای تجزیه و تحلیل هوشمند و پیش‌بینی داده‌های بزرگ ارائه می‌دهد. در دوره کلان داده، یک ساختار مدل یادگیری عمیق قدرتمندتر و پیچیده‌تر می‌تواند به‌طور مؤثرتر و دقیق‌تر ویژگی‌های داده‌های عظیم را ضبط کند و ماهیت داده‌های توصیف شده را بهتر بازگرداند، تا پیش‌بینی‌های دقیقی برای آینده انجام دهد. رویکرد یادگیری عمیق به دنبال یافتن ساختار داخلی داده‌ها و کشف رابطه واقعی بین متغیرها است. با این حال، در عمق شبکه یادگیری به‌دست آمده توسط یادگیری عمیق، آموزش نسبتاً دشوار است و این به چالش‌های توسعه آن تبدیل شده است. در سال ۲۰۰۶، هینتون خاطرنشان کرد که دشواری آموزش عمیق در شبکه با "اولیه سازی لایه به لایه" قابل حل است (Hinton et al, ۲۰۰۶; Hinton and Salakhutdinov, ۲۰۰۶). این امر باعث شد که یادگیری عمیق در زمینه یادگیری ماشین جوان سازی شود و باعث افزایش یادگیری عمیق در آکادمی و صنعت شده است. ویژگی‌های برتر آن در تشخیص گفتار و تحقیقات پزشکی منعکس شده است.

یادگیری عمیق در مقابل روش‌های یادگیری کم‌عمق مانند ماشین‌های بردار پشتیبان و حداکثر آنتروپی است. یادگیری کم‌عمق برای استخراج ویژگی‌های نمونه به تجربه‌مصلوحی پس از یادگیری متکی است. این یک ویژگی تک لایه بدون سلسله مراتب است. در مقایسه با یادگیری کم‌عمق، یادگیری عمیق در (۱) عمق ساختار مدل تأکید می‌شود و مدل یادگیری عمیق به‌طور کلی دارای بیش از پنج یا حتی ده لایه‌گره‌های لایه پنهان است. (۲) براهمیت یادگیری ویژگی تأکید می‌کند یعنی تغییر ویژگی‌های لایه داده نمونه اصلی توسط لایه را به یک فضای ویژگی جدید تغییر می‌دهد و پیش‌بینی یا طبقه‌بندی را ایجاد می‌کند که دشواری آن کمتر است (Yu et al, ۲۰۱۳). یادگیری عمیق می‌تواند به‌طور گسترده‌ای مورد استفاده قرار گیرد زیرا می‌تواند اطلاعات با ارزش‌تری را در داده‌های بزرگ با نیروی انسانی کمتر و مصرف زمان پیدا کند. روش‌های مشترک یادگیری عمیق از کاهش ابعاد ویژگی عمدتاً به روش مبتنی بر نظارت و مبتنی بر بدون نظارت و روش‌های مبتنی بر نیمه نظارت با توجه به تعداد نمونه‌های دارای برچسب تقسیم می‌شوند.

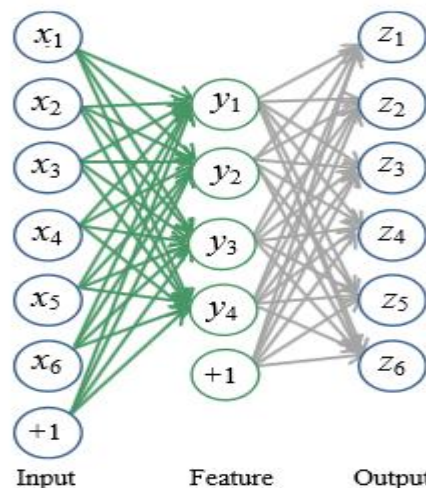
در این مقاله ما روش‌های یادگیری عمیق بدون نظارت برای کاهش ابعاد را بررسی خواهیم نمود. روش‌های یادگیری عمیق بدون نظارت برای کاهش ابعاد ویژگی، عمدتاً شامل رمزگذارهای خودکار عمیق، ماشین‌های محدود بولتزمن و شبکه‌های



باور عمیق هستند. با معرفی رمزگذارهای عمیق خودکار و شبکه های باور عمیق، در ادامه نحوه کاهش ابعاد ویژگی را با رمزگذارهای خودکار شرح می دهیم.

۳- رمزگذارهای خودکار

ایده رمزگذارهای خودکار (AE) برای اولین بار در دهه ۱۹۸۰ توسط هینتون و گروه PDP (Baldi, ۲۰۱۲) برای رسیدگی به مشکل پس انتشار بدون ناظر، با استفاده از داده های ورودی به عنوان ناظر، یعنی یادگیری بدون نظارت، معرفی شد. پس انتشار یک الگوریتم یادگیری نظارت شده است. در سال ۲۰۰۶، هینتون تلاش کرد تا از طریق DBN به این مسائل رسیدگی کند، جایی که، DBN در یک رویکرد حریصانه لایه-لایه آموزش داده می شود (Gogoi and Begum, ۲۰۱۷). رمزگذار خودکار یک شبکه عصبی سه لایه بدون نظارت است که شامل یک لایه ورودی، یک لایه پنهان و یک لایه خروجی است (که به آن لایه بازسازی نیز گفته می شود). ساختار معمولی AE در شکل ۱ نشان داده شده است. رمزگذار خودکار می تواند به تدریج بردارهای ویژگی خاص را به بردارهای ویژگی انتزاعی تبدیل کند، که به خوبی می تواند تبدیل غیر خطی را از فضای داده با ابعاد بالا به فضای داده با ابعاد پایین درک کند. فرآیند رمزگذار خودکار را می توان به دو مرحله رمزگذاری و رمزگشایی تقسیم کرد و این دو مرحله را می توان به صورت زیر تعریف نمود: فرآیند رمزگذاری از لایه ورودی به لایه پنهان (Binghao and Guodong, ۲۰۱۸).

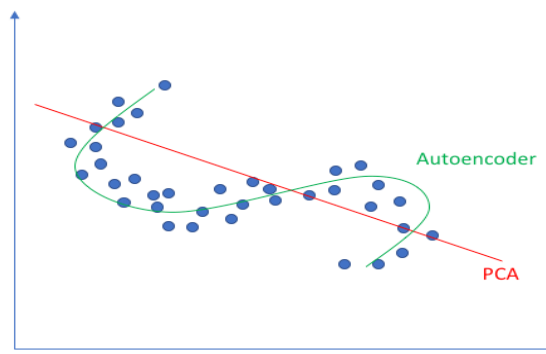


شکل ۱ - ساختار رمزگذار خودکار (Chen and Guo, ۲۰۲۳).

شبکه عصبی رمزگذار خودکار از روش های یادگیری بدون نظارت است و الگوریتم یادگیری رمزگذارهای خودکار برای کدگذاری کارآمد مجموعه داده به منظور کاهش ابعاد استفاده می شود در طول چند دهه گذشته، رمزگذارهای خودکار در لبه تکنولوژی در میان تحقیقات بر روی شبکه های عصبی مصنوعی بوده اند. البته اخیراً رمزگذارهای خودکار برای یادگیری مدل های تولیدی داده ها به کار گرفته شده اند. داده های ورودی ابتدا به یک نمایش انتزاعی تبدیل می شود سپس توسط تابع رمزگذار به فرمت اصلی تبدیل می گردد. به طور خاص تر، آموزش داده شده است تا ورودی را در برخی از نمایش ها رمزگذاری کند و بتوان ورودی را از آن نمایش بازسازی نمود. اساساً رمزگذار خودکار سعی می کند تابع هویت را در این فرآیند پیش بینی کند. یکی از مزیت های کلیدی رمزگذار خودکار این است که این مدل می تواند ویژگی های مفید را به طور مداوم در طول انتشار استخراج کند و اطلاعات بی فایده را فیلتر نماید. علاوه بر این، از آنجایی که بردار ورودی در فرآیند کدگذاری به یک نمایش با



ابعاد پایینتر تبدیل می‌شود، می‌توان کارایی فرآیند یادگیری را افزایش داد. رمزگذار برای نگاشت داده‌های دیده نشده^۱ به لایه میانی با ابعاد پایین استفاده می‌شود و یک بازنمایی خوب همیشه بر روی نقطه ای تمرکز می‌کند که هیچ اطلاعات مهمی در طول فشرده سازی از بین نرود. رمزگذارهای خودکار دقیقاً مانند تحلیل مؤلفه های اصلی (PCA) هستند که خود یک الگوریتم کاهش ابعاد است، اما تفاوت این است که PCA ماهیت خطی دارد درحالی که رمزگذارهای خودکار به دلیل معماری مبتنی بر شبکه عصبی، ماهیت غیرخطی دارند شکل ۲ (Nugroho et al, ۲۰۲۲). کاربردهای رمزگذار خودکار عبارتند از: حذف نویز تصاویر، تشخیص ناهنجاری در سری های زمانی، تشخیص نفوذ شبکه با تشخیص ناهنجاری، تولید سطوح و موسیقی بازی های ویدیویی، استخراج ویژگی، کاهش بعد: تجسم داده های با ابعاد بالا چالش برانگیز است t-SNE^۲ متداول ترین روش مورد استفاده است، اما با تعداد زیادی ابعاد (معمولاً بالای ۳۲) مشکل دارد. بنابراین رمزگذارهای خودکار به عنوان یک مرحله پیش پردازش برای کاهش ابعاد استفاده می‌شوند و این نمایش فشرده توسط t-SNE برای تجسم داده ها به صورت دو بعدی استفاده می‌گردد (Gogoi and Begum, ۲۰۱۷; Chai et al, ۲۰۱۹; Senefonte, ۲۰۲۲).



شکل ۲ - مقایسه رمزگذار خودکار و PCA (Nugroho et al, ۲۰۲۲).

۴- معماری رمزگذارهای خودکار

ساختار اصلی AE از یک لایه ورودی، یک لایه پنهان و یک لایه خروجی تشکیل شده است. در مرحله رمزگذاری، مفهوم رمزگذار خودکار یک بردار ورودی x را به بردار y با تابع رمزگذاری f_θ نگاشت می‌کند و سپس در مرحله رمزگشایی، بردار y را به بردار خروجی z نگاشت کرده که بردار ورودی را بازسازی می‌کند. تابع رمزگشایی g_θ می‌تواند وزن های W شبکه را با به حداقل رساندن خطای بازسازی $L(x, z)$ به عنوان یک تابع هزینه تنظیم کند که برای بهینه سازی پارامترهای شبکه استفاده می‌شود. بنابراین، تابع هدف AE را می‌توان به صورت زیر بیان نمود:

$$\min_{\theta} j_{AE}(\theta) = \min_{\theta} \sum_{i=1}^n L(x_i, z_i) = \min_{\theta} \sum_{i=1}^n L(x_i, g_{\theta}(f_{\theta}(x_i))) \quad (1)$$

که در آن x_i بعد i ام نمونه آموزشی، z_i بعد i ام داده های خروجی و n تعداد کل داده های آموزشی است. $L(x, z) = \|x - z\|^2$ نشان دهنده خطای بازسازی بین داده های ورودی و خروجی است. $y = f_{\theta}(x) = s(Wx + b)$ و $z = g_{\theta}(y) = s(W'y + b')$ به ترتیب توابع نگاشت رمزگذار و رمزگشا هستند. به ویژه s به عنوان یک تابع فعال سازی عمومی مانند تابع سیگموئید انتخاب می‌شود. توجه داشته باشید که بردار پارامتری که باید در AE تنظیم

^۱ _ unseen

^۲ - t-distributed Stochastic Neighbor Embedding



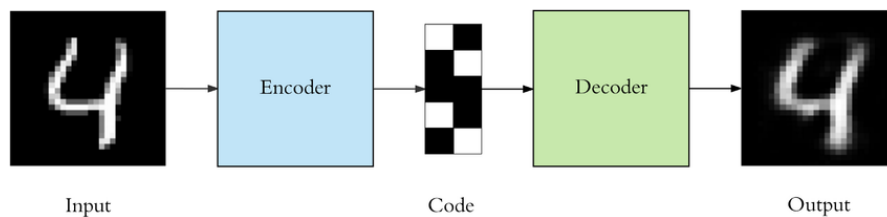
شود $\theta = \{w, b, w', b'\}$ است. که در آن b و b' بردارهای بایاس هستند و w و w' به ترتیب ماتریس‌های وزن رمزگذار و رمزگشا هستند. w' ترانهاد w است (Chai et al, ۲۰۱۹). شایان ذکر است که بعد لایه پنهان کوچکتر از لایه ورودی x است و لایه خروجی x' از لایه ورودی به لایه پنهان به عنوان یک فرآیند فشرده‌سازی و از لایه پنهان به لایه خروجی به عنوان یک فرآیند بازسازی در نظر گرفته می‌شود. هدف یک رمزگذار خودکار یافتن تابع $x' \approx x$ در حالی که لایه پنهان h کمتر از x' است. اگرچه شبکه عصبی با یک لایه پنهان می‌تواند انواع داده‌ها را بیان کند، اما دستیابی به توانایی تعمیم خوب هنوز دشوار است (Yuan et al, ۲۰۲۲). رمزگذار خودکار با استفاده از الگوریتم انتشار برگشتی اغلب براساس تابع هزینه میانگین مربعات خطا آموزش داده می‌شود، دلیل آن این است که رمزگذار خودکار معمولاً از طریق شبکه‌های عصبی استفاده می‌شود. و اگرایی Kullback-Leibler به اختصار (DKL) معیاری است که نشان می‌دهد چگونه یک توزیع احتمال با دیگری متفاوت است. (Myasar et al, ۲۰۲۱; Goodfellow et al, ۲۰۱۶).

۵- انواع رمزگذارهای خودکار

رمزگذارهای خودکار متعددی برحسب کاربرد معرفی شده اند که در ادامه چندین نمونه از این رمزگذارهای خودکار را معرفی می‌کنیم.

۵-۱- رمزگذار خودکار پایه

رمزگذار خودکار پایه از ۳ جزء تشکیل شده است (شکل ۳): رمزگذار، کد و رمزگشا. رمزگذار ورودی را فشرده کرده و کد را تولید می‌نماید، سپس رمزگشا تنها با استفاده از این کد، ورودی را بازسازی می‌کند. برای ساخت یک رمزگذار خودکار به سه چیز نیاز داریم: ۱- روش رمزگذاری، ۲- روش رمزگشایی، ۳- تابع هزینه برای مقایسه خروجی با هدف



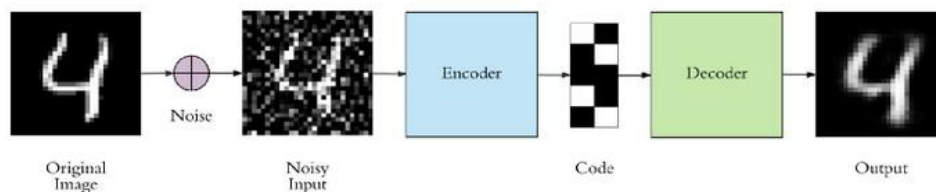
شکل ۳- کد در رمزگذار خودکار (Senefonte, ۲۰۲۲)

رمزگذارهای خودکار نوعی شبکه عصبی هستند که یک نمایش ویژگی بهتر برای داده‌های ورودی با ابعاد بالا ایجاد، و همبستگی بین داده‌های ورودی را پیدا می‌کنند. این یک شبکه عصبی پیشخور چند لایه است که ورودی را با پس انتشار نشان می‌دهد (Aslam et al, ۲۰۲۱). که در آن ورودی شبکه با خروجی یکسان است. ساختار شبکه را می‌توان به گونه‌ای طراحی کرد که مرکز شبکه کوچکتر از ورودی باشد. رمزگذار خودکار یک نمایش فشرده از ورودی را بدون نیاز به برچسب می‌آموزد. بنابراین، این یک روش یادگیری بدون نظارت است که می‌تواند برای مقادیر زیادی از داده‌های خام بدون فرآیندهای حاشیه نویسی غیرقابل حل اعمال شود. هینتون و همکاران در یک مطالعه اساسی روی رمزگذارهای خودکار (Hinton and Salakhutdinov, ۲۰۰۶) نشان داد که رمزگذارهای خودکار بازنمایی فشرده‌ای از تصاویر تولید می‌کنند که نمایش دقیق‌تری برای بازسازی تصاویر اصلی نسبت به تحلیل مؤلفه اصلی (PCA) است. PCA یک روش آماری است که اغلب برای کاهش ابعاد داده‌ها استفاده می‌شود (Aslam et al, ۲۰۲۱).



۲-۵- رمزگذار خودکار نویزی

کوچک نگه داشتن لایه کد، رمزگذار خودکار را مجبور می کند که نمایشی هوشمند از داده ها را بیاموزد. راه دیگری برای وادار کردن رمزگذار خودکار برای یادگیری ویژگی های مفید وجود دارد که اضافه کردن نویز تصادفی به ورودی های آن و بازیابی داده های اصلی بدون نویز است به این ترتیب رمزگذار خودکار نمی تواند به سادگی ورودی را در خروجی خود کپی کند زیرا ورودی همچنین دارای نویز تصادفی است. ما از آن می خواهیم که نویز را کم کند و داده های معنی دار زیربنایی را تولید نماید. این رمزگذار خودکار حذف نویز نامیده می شود (شکل ۴). ممکن است تصویر ورودی با نویز همراه باشد داده های نویز به ورودی رمزگذار خودکار تبدیل می شوند. رمزگذار خودکار اصلاً تصویر اصلی را نمی بیند. اما پس از آن انتظار داریم رمزگذار خودکار تصویر اصلی بدون نویز را بازسازی کند. تنها یک تفاوت کوچک بین اجرای رمزگذار خودکار حذف نویز و کدمعمولی وجود دارد. معماری به هیچ وجه تغییر نمی کند، فقط تابع تناسب فرق دارد. بقیه موارد دقیقاً یکسان است. ورودی رمزگذار خودکار تصویر نویز بوده و هدف مورد انتظار تصویر اصلی بدون نویز است (Chen and Guo, ۲۰۲۳).

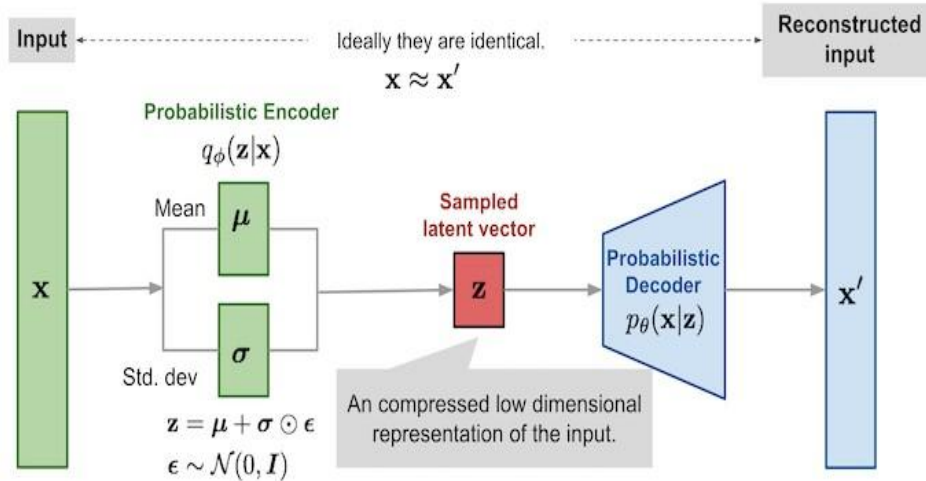


شکل ۴- رمزگذار خودکار نویزی (Tripathy and Tabasum, ۲۰۲۳)

۳-۵- رمزگذارهای خودکار متغیر

رمزگذارهای خودکار متغیر، رمزگذارهای خودکاری هستند که از روش نمونه گیری و تنظیم سازی Kullback-Leibler استفاده می کنند. رمزگذارهای خودکار متغیر قصد دارند فضای پنهان را هموارتر کنند، یعنی یک تغییر کوچک در X منجر به تغییر کوچکی در فضای پنهان Z و یک تغییر کوچک در Z منجر به تغییر کوچکی در X می شود. یک فضای نهفته باید صاف و دارای نقاط قابل قبول باشد تا موثرتر و دقیق تر واقع گردد. و این همان چیزی است که رمزگذار خودکار متغیر برای رسیدن به آن تلاش می کند. در VAE، رمزگذار نه تنها Z ، بلکه μ و σ را خروجی می دهد. پس از آن عملیات نمونه برداری Z را از بین این پارامترها انتخاب می کند و طبق معمول رمزگشا مانند قبل Z را می گیرد (شکل ۵). یک تکنیک نمونه برداری خوب بازسازی خوبی از نقاط داده و همچنین بازسازی خوبی از نقاط نزدیک به نقاط داده را فراهم می نماید. این فرآیند تضمین می کند که هر نقطه ای که نزدیک به مکان پنهانی است که در آن رمزگذاری کرده اید (ورودی X ، یعنی میانگین Z) می تواند به چیزی شبیه به X رمزگشایی شود. بنابراین فضای پنهان مجبور می شود به طور مداوم معنی دار باشد. هر دو نقطه نزدیک در فضای پنهان، تصاویر بسیار مشابه را رمزگشایی می کند. پیوستگی، همراه با ابعاد کم فضای پنهان، هر جهتی را در فضای پنهان مجبور می کند تا محور معنی داری از تغییرات داده ها را رمزگذاری نماید، و فضای نهفته را بسیار ساختارمند می کند. بنابراین برای دستکاری از طریق بردارهای مفهومی بسیار مناسب است.

در رمزگذار خودکار متغیر ما می خواهیم داده ها به صورت عادی در فضای Z به طور خاص، یک نرمال چند متغیره استاندارد $N(0, I)$ رمزگشا توزیع شوند. ما می توانیم مطمئن باشیم که همه چنین نقاطی با نقاط X معمولی بدون "حفره" مطابقت دارند. زیرا رمزگذار سخت کاری کند تا داده ها را فشرده کند، بنابراین فضا را هدر نمی دهد. پارامترهای یک VAE از طریق دو تابع هزینه آموزش داده می شوند: یک افت بازسازی که نمونه های رمزگشایی شده را مجبور می کند با ورودی های اولیه مطابقت داشته باشند، و یک فقدان منظم سازی که به یادگیری فضاهای پنهان خوب شکل گرفته و کاهش تطبیق بیش برآزش با داده های آموزشی کمک می کند. تابع هزینه منظم سازی از رمزگذار می خواهد که داده ها را در یک توزیع نرمال در فضای پنهان قرار دهد. و اگرایی کولبک-لایبلر $KL(p||q)$ اندازه گیری آماری عدم تشابه



شکل ۵ - ساختار رمزگذار خودکار متغیر (Crystal et al, ۲۰۱۹).

بین یک جفت توزیع p و q است. بنابراین، عددی است که وقتی p و q غیرمشابه باشند بزرگ است و وقتی شبیه باشند صفر است. تابع هزینه KL یک اصطلاح منظم در تابع هزینه VAE است. مثل همیشه، می‌توانیم با ضرب KL در یک اسکالر، منظم سازی را تنظیم کنیم. اگر خیلی قوی باشد مدل ما فرو می‌ریزد و اگر خیلی ضعیف باشد معادل رمزگذار خودکار کلاسیک خواهد بود (Carl, ۲۰۱۶; Gregor, et al, ۲۰۱۹; Crystal et al, ۲۰۱۹).

VAE یک مدل یادگیری عمیق است که ترکیبی از شبکه‌های عصبی و مفاهیم آماری مانند استنتاج بیزی است. این مدل از مدل‌های مولد در یادگیری ماشین است که به‌طور گسترده‌ای در حوزه‌های مختلف مورد استفاده قرار می‌گیرد. این مدل‌ها نه تنها برای بازسازی داده‌ها، بلکه برای تولید داده‌های جدید نیز کاربرد دارند. البته برای کاهش ابعاد و یادگیری نمایش‌های نهفته نیز استفاده می‌شود و می‌تواند داده‌های با ابعاد بالا (مانند تصاویر) را به فضای کم‌بعدی (فضای نهفته) تبدیل کند.

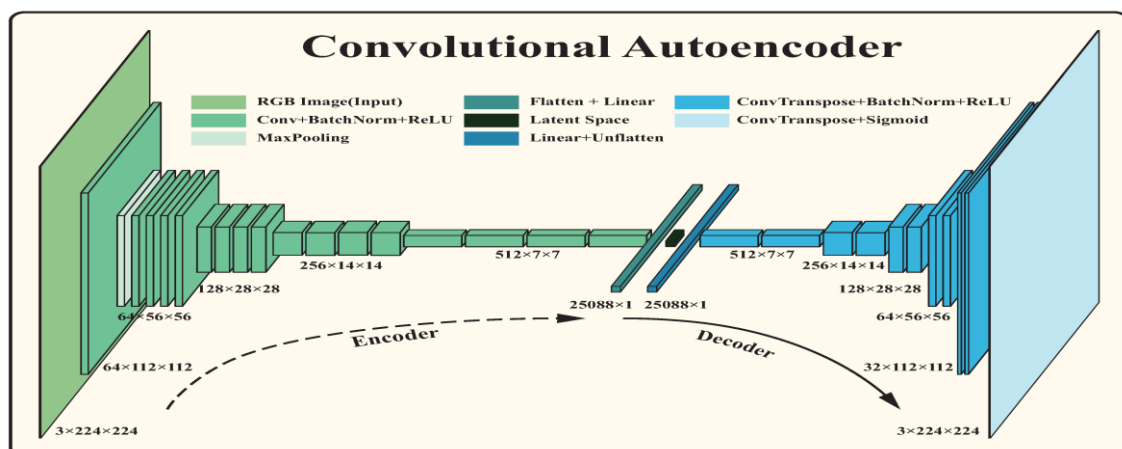
۴-۵- رمزگذار خودکار پشته‌ای

رمزگذار خودکار به منظور کاهش ابعاد غیرخطی (Tanvi et al, ۲۰۱۵) کارآمدتر از سایر یادگیری‌های چندگانه شناخته شده است، و همان‌طور که در شکل ۲-۱۵ نشان داده شده است، می‌توان آن را برای اجرای یادگیری عمیق روی هم قرار داد. اگرچه در (Riaz et al, ۲۰۱۸) آمده که در مقایسه با CNN، یک رمزگذار خودکار به اندازه کافی ویژگی‌های نماینده را از داده‌های با ابعاد بالا مانند تصاویر نمی‌آموزد، یادگیری ویژگی از طریق رمزگذار خودکار پشته‌ای به مقدار بسیار کمتری نیاز دارد. محاسبات آن را برای توصیف مناطق محلی در طول توسعه نمایش‌های تصویر BOVW جذاب می‌کند. (Olaode and Naghdy, ۲۰۱۹) برنامه رمزگذار خودکار پشته‌ای را به عنوان یک الگوریتم استخراج ویژگی تصویر محلی ارائه می‌کند. (Hinton and Salakhutdinov, ۲۰۰۶). در اینجا، انتقال از یادگیری بدون نظارت به تحت نظارت در حال انجام است. شبکه عصبی کامل تمام لایه‌های رمزگذار عمیق را به عنوان یک موجودیت واحد پردازش می‌کند به گونه‌ای که وزن‌ها در رمزگذار خودکار پشته‌ای بهبود می‌یابند. برای بهبود استحکام مدل و کاهش تأثیر تغییرات تصویر، محدودیت‌های متمایز اضافی (برچسب سازگار) را به رمزگذار خودکار استاندارد اضافه می‌کنند. تابع هزینه اقلیدسی استفاده از اصطلاح سازگاری برچسب را تضمین می‌کند که بردارهای ویژگی (نرون‌های لایه پنهان) به گونه‌ای بازیابی شوند که متمایز باشند. به همین دلیل، همه بردارهای ویژگی مربوط به برچسب کلاس مشابه شباهت بالایی دارند در حالی که تغییرپذیری یا تبعیض بین طبقاتی افزایش می‌یابد.

رمزگذار خودکار پشته ای یکی از روش‌های رایج در یادگیری عمیق است که برای کاهش بعد داده‌ها، استخراج ویژگی‌ها و همچنین برای بازسازی داده‌ها استفاده می‌شود. یکی از اصلی‌ترین کاربردهای رمزگذار خودکار پشته ای، کاهش بعد داده‌های با بعد بالا است (Hinton and Salakhutdinov, ۲۰۰۶). این تکنیک به خصوص در زمینه‌های پردازش تصویر و پردازش سیگنال بسیار مفید است. رمزگذارهای خودکار پشته ای قادر به استخراج ویژگی‌های مهم از داده‌ها هستند که می‌تواند به بهبود دقت مدل‌های پیش‌بینی کمک کند. همچنین برای پیش پردازش داده‌ها قبل از اعمال مدل‌های یادگیری ماشینی دیگر استفاده می‌شود تا کیفیت داده‌ها را بهبود بخشد. این مدل‌ها می‌توانند به بازسازی داده‌های ورودی با حذف نویز و افزایش کیفیت داده کمک کنند. در کاربردهای تشخیص ناهنجاری، رمزگذارهای خودکار پشته ای می‌توانند ناهنجاری‌ها را با بررسی اختلاف بین داده‌های بازسازی شده و داده‌های اصلی تشخیص دهند (Bengio, ۲۰۰۹).

۵-۵- رمزگذار خودکار پیچشی

رمزگذارهای خودکار کانولوشن یک نمایش تخصصی از رمزگذارهای خودکار هستند که برای مدیریت داده‌های با ابعاد بالا مانند تصاویر طراحی شده‌اند. مدل CA از یک رمزگذار برای فشرده سازی داده‌های ورودی در فضای پنهان کم بعدی و از رمزگشا برای بازسازی داده‌های اصلی با استفاده از سلسله مراتب فضایی تشکیل شده است. ساختار CA از یک رمزگذار، فضای پنهان و رمزگشا تشکیل شده است (Shivalila et al, ۲۰۲۵). پرکاربردترین شبکه‌های عصبی در الگوریتم‌های خوشه‌بندی رمزگذارهای خودکار پشته‌ای (SAE) هستند. رمزگذارهای خودکار پشته‌ای نیاز به پیش‌آموزش لایه‌ای دارد قبل از اینکه به صورت انتها به انتها تنظیم شود. وقتی لایه‌ها عمیق‌تر می‌شوند، روند پیش‌آموزش می‌تواند خسته‌کننده و زمانبر باشد. علاوه بر این، SAE با لایه‌های کاملاً متصل ساخته شده است که برای برخورد با تصاویر بی اثر هستند CAE برای یادگیری ویژگی‌های تصاویر مفید است و با حفظ ساختار محلی داده‌ها از تحریف فضای ویژگی جلوگیری می‌کند. رمزگذار خودکار پیچشی که می‌تواند به صورت سرتاسر آموزش داده شود، برای یادگیری ویژگی‌ها از تصاویر بدون برچسب طراحی شده است. CAE طراحی شده با ترکیب روابط فضایی بین پیکسل‌ها در تصاویر، نسبت به رمزگذارهای خودکار پشته‌ای برتری دارد. همانطور که از نام آن پیداست، CAE ترکیبی از رمزگذار خودکار و شبکه عصبی کانولوشن (CNN) است. رمزگذار خودکار یک شبکه عصبی مصنوعی است که از یک رمزگذار و یک شبکه عصبی رمزگشا تشکیل شده است. هنگامی که داده ورودی X به رمزگذار خودکار وارد می‌شود، در حالی که از چندین لایه کاملاً متصل در شبکه رمزگذار عبور می‌کند، به طور غیرخطی به یک بردار رمزگذاری شده Z تبدیل می‌گردد. سپس، از بردار رمزگذاری شده Z بازسازی X از طریق شبکه رمزگشا به دست می‌آید. به عنوان مثال، فرآیندهای رمزگذاری و رمزگشایی رمزگذار خودکار که با یک لایه پنهان هستند به صورت ریاضی به صورت زیر مدل شده است (Guo et al, ۲۰۱۷).



شکل ۶ - ساختار رمزگذار خودکار پیچشی (Zhang et al, ۲۰۲۴).



یکی از انواع شبکه‌های عصبی هستند که برای پردازش داده‌های پیچیده، به‌ویژه داده‌های تصویری، استفاده می‌شوند. این مدل‌ها ترکیبی از لایه‌های کانولوشنی و ساختار خودرمزگذار هستند و کاربردهای متنوعی دارند این مدل‌ها می‌توانند داده‌های تصویری را به نمایش‌های فشرده‌تر تبدیل کنند، که در کاهش حجم داده‌ها و افزایش سرعت پردازش مؤثر است. رمزگذار خودکار پیچشی یک رمزگذار خودکار است که در آن تبدیل غیرخطی توسط CNN به‌دست می‌آید. CNN دارای دو لایه متمایز از شبکه‌های عصبی کاملاً متصل معمولی لایه کانولوشنال و pooling است. رمزگذار خودکار معمولی معمولاً از دو لایه تشکیل شده است (شکل ۶). برای رمزگذاری $fw(\cdot)$ و رمزگشا $gu(\cdot)$ هدف آن یافتن یک کد برای هر کدام از نمونه ورودی با به حداقل رساندن میانگین مربعات خطا (MSE) است.

۶-۵- رمزگذار خودکار پراکنده

رمزگذارهای خودکار پراکنده دارای گره‌های مخفی بزرگتر از گره‌های ورودی هستند. آن‌ها می‌توانند ویژگی‌های مهم را از داده‌ها کشف کنند. یک رمزگذار خودکار پراکنده عمومی در جایی تجسم می‌شود که مبهم بودن یک گره با سطح فعال سازی مطابقت دارد. محدودیت پراکندگی بر روی لایه پنهان معرفی شده است. که برای جلوگیری از کپی لایه خروجی داده‌های ورودی است. پراکندگی ممکن است با عبارات اضافی در تابع هزینه در طول فرآیند آموزش، یا با مقایسه توزیع احتمال فعالسازی‌های واحد پنهان با مقدار دلخواه پایین، یا با صفر کردن دستی همه فعالسازی‌های واحد پنهان به غیر از قویترین واحدهای پنهان، به‌دست آید. رمزگذارهای خودکار پراکنده با اعمال محدودیت‌های پراکنده، الگوها را روی لایه‌های پنهان یاد می‌گیرند. تفاوت بین رمزگذار خودکار پراکنده و رمزگذارهای خودکار اساسی در تابع هزینه نهفته است.

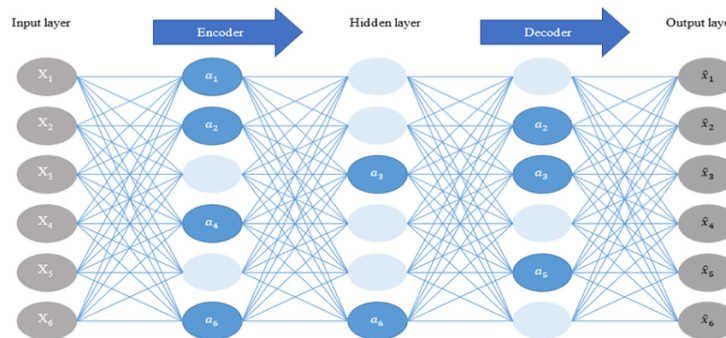
یکی از معایب رمزگذار خودکار محدود بودن تعداد واحدهای پنهان آن است. ما می‌توانیم رمزگذار خودکار را با استفاده از یک محدودیت پراکندگی منظم کنیم، به‌طوری که تنها بخشی از گره‌ها مقادیر غیر صفری داشته باشند که گره‌های فعال نامیده می‌شوند. به‌طور خاص، یک عبارت جریمه به تابع هزینه اضافه می‌کنیم به‌طوری که تنها بخشی از گره‌ها فعال می‌شوند (شکل ۷). این امر رمزگذار خودکار را مجبور می‌کند که هر ورودی را به صورت ترکیبی از تعداد کمی از گره‌ها نمایش دهد. این روش حتی اگر اندازه کد بزرگ باشد کار می‌کند، زیرا تنها زیر مجموعه کوچکی از گره‌ها در هر زمان فعال خواهند بود. رمزگذارهای خودکار پراکنده دارای جریمه پراکندگی هستند، مقداری نزدیک به صفر اما دقیقاً صفر نیست. جریمه پراکندگی علاوه بر خطای بازسازی بر روی لایه پنهان اعمال می‌شود. این کار مانع از بیش‌برازش بیش از حد نیز می‌شود.

رمزگذار خودکار پراکنده یک شبکه عصبی (NN) با چندین لایه رمزگذار خودکار است که معمولاً به عنوان یک روش DL برای کاهش ابعاد و یادگیری ویژگی استفاده می‌شود. رمزگذار خودکار پراکنده می‌تواند ویژگی‌های داده را استخراج کند به‌طور مؤثر، طبقه‌بندی را آسان‌تر می‌کند و ویژگی‌های داده‌های قوی‌تر را استخراج می‌کند (Zhang et al, ۲۰۱۸). این نوع شبکه یک روش متمایل به ویژگی بدون نظارت است که می‌تواند از طراحی ویژگی‌های فشرده و دستی جلوگیری کند. آزمایش‌ها در برنامه‌های مختلف مانند پردازش زبان طبیعی، بینایی رایانه و پردازش صدا دلگرم‌کننده هستند. هدف رمزگذار خودکار پراکنده این است که ورودی برابر با خروجی باشد. لایه پنهان در SAE را می‌توان به عنوان استخراج ویژگی لایه ورودی مشاهده کرد. اساساً رمزگذار خودکار پراکنده نوعی شبکه عصبی بدون نظارت است (Gogoi and Begum, ۲۰۱۷) و سعی می‌کند یک تابع $h_{w,b}(x) \approx x$ را یاد بگیرد. به عبارت دیگر، تلاش می‌کند تا یک تقریب از تابع هویت را یاد بگیرد تا x' را که مشابه x است، خروجی بگیرد. به نظر می‌رسد که تابع هویت یک کارکرد بی‌اهمیت بوده که در تلاش برای یادگیری است. اما با قرار دادن محدودیت‌هایی در شبکه، مانند محدود کردن تعداد واحدهای پنهان، می‌توان ساختار جالبی در مورد داده‌ها کشف کرد. به عنوان یک مثال عینی، فرض کنید ورودی‌های x مقادیر شدت پیکسل از یک تصویر $10 \times 10 \times 100$ (پیکسل) هستند، بنابراین $n = 100$ و $s_2 = 50$ واحد پنهان در لایه l_2 وجود دارد. توجه داشته باشید که $y \in \mathbb{R}^{100}$ نیز



داریم. از آنجایی که تنها ۵۰ واحد پنهان وجود دارد، شبکه مجبور است یک نمایش فشرده از ورودی را یاد بگیرد. یعنی تنها با توجه به بردار فعال سازی واحد پنهان $a^{(2)} \in \mathbb{R}^{50}$ ، باید سعی کند ورودی ۱۰۰ پیکسل را بازسازی کند. اگر ورودی کاملاً تصادفی بود مثلاً هر x_i از یک IID گاوسی مستقل از سایر ویژگی ها می آید، این عمل فشرده سازی بسیار مشکل خواهد بود. اما اگر ساختاری در داده ها وجود داشته باشد، برای مثال، اگر برخی از ویژگی های ورودی با هم مرتبط باشند، این الگوریتم قادر خواهد بود برخی از آن همبستگی ها را کشف کند (Ng, Sparse autoencoder, ۲۰۱۱).

در (Aslam et al, ۲۰۲۱) رمزگذار خودکار پراکنده با استفاده از الگوریتم یادگیری ویژگی بدون نظارت، ویژگی های ساختار یافته با ابعاد بالا نمایش داده های سرطانی یا غیرسرطانی را می آموزد. ورودی x با نمایش مربوطه در لایه ورودی رمزگذار به h تبدیل شده است. لایه پنهان h داده های ورودی را با ویژگی های جدید تجسم می کند. نمایش پنهان h از داده های ورودی جدید x' بازسازی می شود، این کار توسط رمزگشا در لایه خروجی انجام شده، حداقل اختلاف با آموزش رمزگذار خودکار بین ورودی x و مقدار بازسازی شده آن x' برای دستیابی به مقادیر پارامتر بهینه مشخص شده است. این اختلاف با استفاده از الگوریتم پس انتشار به دست آمده است. تابع هزینه SAE در رابطه (۲) نشان داده شده، که از سه اصطلاح تشکیل شده است.



شکل ۷ - ساختار رمزگذار خودکار پراکنده (Ullah et al, ۲۰۲۱)

$$J_T(W, b) = \frac{1}{n} \sum_{i=1}^n L(x(i)y(i)) + \sum_{j=1}^n KL(\rho \parallel \hat{\rho}_j) + \beta \|W\|_2^2 \quad (2)$$

اختلاف بین ورودی و نمایش بازسازی شده آن با جمله اول نشان داده می شود که مجموع میانگین مربعات خطا است. در ترم دوم رابطه (۲)، n تعداد لایه های پنهان را نشان می دهد و شاخص j شبکه را بر روی واحدهای پنهان جمع می کند. پارامتر α مقدار پراکندگی را نشان می دهد. به طور کلی، این مقدار تقریباً صفر است اما برابر با صفر نیست. فعال سازی هدف h با ρ و میانگین فعال سازی j امین واحد پنهان روی n داده آموزشی با $\hat{\rho}$ نشان داده می شود که با فرمول زیر محاسبه می شود (Aslam et al, ۲۰۲۱).

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n h_j(i) \quad (3)$$

شکل ۸ ساختار شبکه اصلی رمزگذار خودکار پراکنده سه لایه را به عنوان مثال نشان می دهد. لایه رمزگذاری l_1 داده های ورودی x را به لایه پنهان l_2 نگاشت می کند، که به عنوان یک نمایش ویژگی کم بعدی h از لایه رمزگذاری عمل می کند. لایه خروجی l_3 یک لایه ورودی معکوس است و هدف آن تبدیل h از l_2 به x است. برای کاهش شکاف بین x و x' ، SAE از ایده



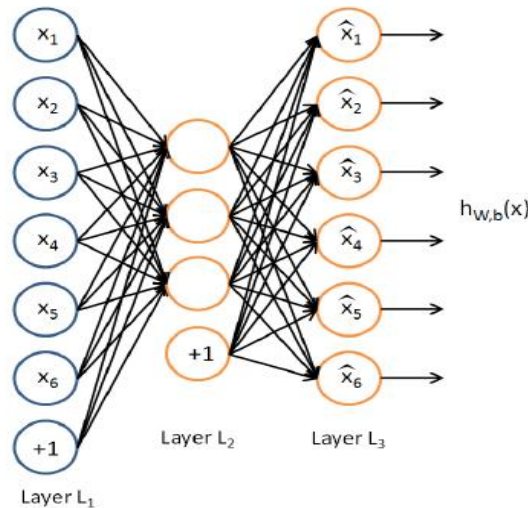
انتشار برگشتی برای بهینه سازی ماتریس وزن w و انحراف b از طریق تکرارهای متعدد استفاده می کند. در (Liu et al ۲۰۱۹)، تابع هزینه هسته آن را به عنوان یک نمونه واحد x تعریف می کند.

$$J(W, b; x) = \frac{1}{2} \|g_{W,b}(x)\|_2^2 \quad (4)$$

که در آن $g_{W,b}(x) = x'$ در واقع، $g_{W,b}(x)$ می تواند به عنوان یک تابع فعال سازی $f(W^T x)$ در نظر گرفته شود. $f(\cdot)$ یک نوع معمول از تابع سیگموئید می تواند باشد که در اینجا ما از تابع SELU استفاده می کنیم.

$$f(\alpha, x) = \lambda \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \quad (5)$$

با گسترش X به مجموعه آموزشی با n نمونه، تابع هزینه کل به صورت زیر تعریف می شود:



شکل ۸ - رمزگذار خودکار پراکنده با سه لایه

$$C(W, b) = \frac{1}{n} \sum_{i=1}^n J(W, b; x_i) + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{j=1}^{s_l+1} \sum_{i=1}^{s_l} (W_{ij}^{(l)})^2 \quad (6)$$

که در آن n_l تعداد لایه های رمزگذار خودکار را نشان می دهد l_1 به طور ثابت لایه رمزگذاری است، L_{n_l} لایه خروجی است، s_l به معنای تعداد عناصر در لایه خروجی است، و $W_{ij}^{(l)}$ به معنای ماتریس وزن است که بین l_1 با j واحد و با i واحد است. $C(W, b)$ معادله ای از w و b است و هدف به دست آوردن حداقل آن است. در مرحله اولیه آموزش شبکه، w و b باید مقداری اولیه شوند. به طور کلی، برای هر عنصر $W_{ij}^{(l)}$ و $b_i^{(l)}$ ، به طور تصادفی به یک عدد مثبت بسیار کوچک مقداردهی اولیه می شود (مثلاً $(0, \sigma^2)$ ، $\sigma < 0.01$). الگوریتم نزول گرادینت و نرخ یادگیری مناسب α برای کاهش تدریجی w و b تا زمان همگرایی استفاده می شود. با نزول گرادینت، فرمول تکراری به روز شده ماتریس وزن و بایاس به شرح زیر است:

$$W_{ij}^{(l)} \leftarrow W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} C(W, b), \quad (7)$$



$$b_{ij}^{(l)} \leftarrow b_{ij}^{(l)} - \alpha \frac{\partial}{\partial b_{ij}^{(l)}} C(W, b), \quad (8)$$

هر عنصر خروجی در لایه خروجی l_2 باید با $f(\cdot)$ فعال شود. اگر یک واحد نزدیک به ۱ باشد، فرض می‌کنیم که این یک نورون فعال است، و اگر واحد نزدیک به ۰ باشد، فرض می‌کنیم که این یک نورون غیرفعال است.

SAE از $a_j^{(L2)}(x)$ برای نشان دادن فعال شدن عنصر پنهان j در l_2 استفاده می‌کند، میانگین تابع فعال‌سازی تمام سلول‌های پنهان j در شبکه SAE به صورت زیر تعریف شده است:

$$\hat{\rho}_j = \frac{1}{d} \sum_{i=1}^d [a_j^{(L2)}(x^i)] s.t. \hat{\rho}_j = \rho, \quad (9)$$

جایی که ρ پارامتر پراکنده است، معمولاً مقدار زیادی نخواهد بود. برای اینکه نرخ فعال‌سازی لایه پنهان کم شود و فعال‌سازی لایه پنهان را محدود کند، SAE جریمه اضافی $\hat{\rho}_j$ اضافه می‌کند.

$$\sum_{j=1}^{s_2} KL(\rho \parallel \hat{\rho}_j) = \sum_{j=1}^{s_2} \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (10)$$

$$C_{sparse}(W, b) = C(W, b) + \beta \sum_{j=1}^{s_2} KL(\rho \parallel \hat{\rho}_j) \quad (11)$$

$KL(\cdot)$ و اگرایی Kullback-Leibler بین یک متغیر تصادفی برنولی با میانگین ρ و میانگین $\hat{\rho}_j$ است. s_2 تعداد واحدهای l_2 است. به طور خلاصه، تابع هزینه کلی است و اگرایی KL (KL-divergence) یک تابع استاندارد برای اندازه‌گیری تفاوت دو توزیع متفاوت است (Liu et al, ۲۰۱۹; Chung et al, ۲۰۲۴).

در (Liu et al, ۲۰۱۹) لی و همکاران از رمزگذار خودکار پراکنده بدون نظارت (SAE) برای کاهش ابعاد ویژگی‌های تصویر با ابعاد بالا استفاده کرده‌اند که هدف آن کاهش خطای بازسازی بین لایه شبکه عصبی ورودی و خروجی است. لایه پنهان تولید شده توسط تکرارها به عنوان ویژگی برای تحقق استخراج ویژگی از نمونه‌ها استفاده می‌شود تا حد امکان می‌توان از مزاحمت ابعاد ناشی از ویژگی‌های متعدد جلوگیری کرد. داده‌های با ابعاد پایین به طور قابل توجهی باعث بهبود کارایی درک تصویر اجتماعی در همان زمان می‌گردد و در مرحله استخراج ویژگی اجتناب از مزاحمت ابعاد و همبستگی برچسب‌ها هر دو در نظر گرفته شده، که با توجه به توانایی قدرتمند کاهش ابعاد، شبکه عصبی AE برای حل مشکل SIU پیشنهاد شده است.

در (Wang, ۲۰۱۶) بر توانایی رمزگذارهای خودکار در کاهش ابعاد تمرکز دارد. آزمایش‌ها نشان می‌دهند که نتایج رمزگذار خودکار واقعاً از سایر روش‌های کاهش ابعاد متفاوت است و در برخی موارد، رمزگذار خودکار نه تنها ابعاد را کاهش می‌دهد، بلکه می‌تواند ساختارهای تکراری را تشخیص دهد. نویسندگان مقاله معتقدند که این یک ویژگی خوب برای بسیاری از برنامه‌ها است و در موقعیت‌هایی با ساختارهای تکراری، رمزگذار خودکار مناسب‌تر است. همچنین در این مقاله به طور مقدماتی تأثیر تعداد گره‌های لایه پنهان را بر عملکرد رمزگذار خودکار بررسی می‌کند بنابه اطلاعات مقاله برای طبقه‌بندی بر روی مجموعه داده MNIST بدون شبکه، بهترین نتایج زمانی حاصل می‌شود که تعداد گره‌های لایه پنهان حول ابعاد ذاتی مجموعه داده تنظیم شود این مورد ممکن است رابطه احتمالی بین تعداد گره‌های لایه پنهان و ابعاد ذاتی مجموعه داده را نشان دهد.

در این پژوهش، ما ویژگی‌ها و کاربردهای انواع رمزگذارهای خودکار را مورد مطالعه قرار دادیم که به طور خلاصه نتایج مطالعات به این شرح است: رمزگذار پایه برای داده‌های ساده و غیرساختاری (مانند داده‌های جدولی یا بردارهای ویژگی)



عملکرد مناسبی دارد و پیاده‌سازی آن سریع بوده و برای کاهش ابعاد داده‌های با ابعاد متوسط مناسب است. اما برای داده‌های پیچیده (مانند تصاویر) عملکرد ضعیفی دارد و ممکن است به‌خوبی ویژگی‌های مهم را استخراج نکند. رمزگذار خودکار متغییر نیز برای داده‌هایی که نیاز به تولید نمونه‌های جدید دارند یا زمانی که توزیع داده‌ها مهم است مناسب ترند و می‌توان گفت از مدل‌های مولد بوده و برای داده‌های پیچیده (مانند تصاویر یا متن) مناسب است. فضای نهفته‌ای با توزیع مشخص (مانند گاوسی) ایجاد می‌کند اما پیچیدگی محاسباتی بالاتری دارد و ممکن است برای کاهش ابعاد ساده، بیش‌ازحد پیچیده باشد. رمزگذار خودکار پشته‌ای نیز برای داده‌های پیچیده و با ابعاد بسیار بالا مناسب بوده و از چندین لایه پنهان تشکیل شده است که می‌تواند ویژگی‌های سلسله‌مراتبی را یاد بگیرد در نتیجه برای کاهش ابعاد داده‌های با ساختار پیچیده (مانند تصاویر یا متن) مناسب است اما نیاز به تنظیم دقیق فرآیند یادگیری دارد و ممکن است زمان آموزش طولانی‌تری داشته باشد. رمزگذار خودکار پیچشی برای داده‌های ساختاری مانند تصاویر، ویدیوها یا سیگنال‌ها مناسب است چون از لایه‌های کانولوشنال استفاده می‌کند برای داده‌های شبکه‌ای (مانند تصاویر) بسیار مناسب است و می‌تواند ویژگی‌های محلی و فضایی را به‌خوبی یاد بگیرد. برای کاهش ابعاد تصاویر و داده‌های مشابه بسیار قدرتمند است در صورتی که برای داده‌های غیر ساختاری (مانند داده‌های جدولی) مناسب نیست و پیچیدگی محاسباتی بالاتری نسبت به رمزگذار خودکار ساده دارد. اما رمزگذار خودکار پراکنده برای داده‌هایی که نیاز به استخراج ویژگی‌های پراکنده دارند مناسب بوده و می‌تواند ویژگی‌های پراکنده و معنادار را استخراج نماید بنابراین برای کاهش ابعاد داده‌های با ابعاد بسیار بالا مناسب، اما از معایب آن این است که تنظیم پارامترهای پراکنده‌ی ممکن است دشوار بوده و پیچیدگی محاسباتی آن نیز بالا است. در ادامه ما برای بررسی عملکرد رمزگذارهای خودکار در کاهش ابعاد تصویر، بر روی دیتاست MIRFlickr آزمایشات متعددی انجام دادیم و در حالت‌های مختلف بدون استفاده از رمزگذار خودکار و با استفاده از معماری‌های VGGNet، ResNet و Densnet و با استفاده از رمزگذار پایه و رمزگذار خودکار پراکنده فاز استخراج ویژگی را انجام دادیم و نتایج حاصل از روش‌های کاهش بعد را در یک الگوریتم سنتی برچسب گذاری تصویر آزمایش نمودیم که نتایج به دست آمده را در قسمت بعد توضیح خواهیم داد.

۶- آزمایشات و نتایج

۶-۱- مجموعه داده

برای مقایسه روش‌های کاهش ابعاد در این کار، آزمایش‌های گسترده‌ای را روی مجموعه داده‌های تصویر اجتماعی MIRFlickr انجام دادیم. MIRFlickr شامل ۲۵۰۰۰ تصویر از فلیکر و ۱۳۸۶ کاربر است. ۳۸ نوع تگ مرتبط برای این تصاویر ارائه شده است

۶-۲- معیارهای ارزیابی

ما برای ارزیابی عملکرد روش برچسب زنی و صحت روش استخراج ویژگی خود از معیارهای رایج برچسب‌زنی تصاویر شامل دقت، فراخوانی و F_1 -Score و همچنین معیارهای MicroAUC و MacroAUC، که در (Liu et al, ۲۰۱۹) آمده است، استفاده می‌کنیم. معیار دقت عبارت است از نسبت تعداد تصاویری است که به‌درستی برچسب نسبت داده شده است به کل تصاویر پیش‌بینی شده برای آن برچسب. مطابق رابطه (۱۲) و یادآوری عبارت است از نسبت تعداد تصاویری که به‌درستی برچسب نسبت داده شده به تعداد تصاویری که دارای این برچسب هستند که در رابطه (۱۳) تعریف شده است. معیارهای دقت و یادآوری مکمل یکدیگر هستند و معیار F_1 میانگین وزنی دقت و یادآوری است که طبق رابطه (۱۴) به‌دست می‌آید و به ما کمک می‌کند تا نتایج کلی را بهتر ارزیابی کنیم.

$$Precision = \frac{TP}{TP + FP} \quad (12)$$



$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

TP: تعداد تصاویری که حاوی برچسب هستند و سامانه برچسب زنی به درستی به آن‌ها برچسب تخصیص داده است.

FP: تعداد تصاویری که دارای برچسب نیستند و سامانه برچسب زنی به اشتباه به آن برچسب اختصاص داده است.

FN: تعداد تصاویری که دارای برچسب هستند و سامانه برچسب زنی به آن‌ها برچسب تخصیص نداده است.

TP+FP: تعداد کل تصاویری است که سامانه برچسب تخصیص داده است.

برای بررسی اثربخشی روش خود، روشی را اتخاذ کردیم که AUC نامیده می‌شود. به عنوان یک شاخص ارزیابی، AUC نسبتاً قابل اعتماد است و به عنوان یک روش استاندارد برای مقایسه مدل چندبرچسبی در نظر گرفته می‌شود. به طور کلی، دو روش ارزیابی، MicroAUC و MacroAUC، برای ارزیابی عملکرد کلی بین چندین مفهوم و میانگین عملکرد همه مفاهیم استفاده می‌شود. در MicroAUC، بردارهای کلاس مختلف در یک بردار طولانی الحاق می‌شوند و سپس میانگین AUC محاسبه می‌شود. در MacroAUC ابتدا میانگین AUC هر کلاس محاسبه می‌شود و سپس میانگین مقدار AUC همه کلاس‌ها محاسبه می‌شود (Liu et al, 2019).

۶-۳- نتایج

برای مجموعه‌های داده MIRFlickr، از معماری‌های مختلف VGGNet و ResNet و Densnet که از قبل آموزش دیده شده را برای استخراج ویژگی تصویر اتخاذ می‌کنیم، که در نتیجه یک بردار ویژگی با ابعاد ۴۰۹۶ که نشان‌دهنده محتوای بصری هر تصویر است را به عنوان خروجی لایه آخر شبکه است دریافت می‌کنیم سپس بردار ویژگی با ابعاد بالا را به عنوان ورودی در اختیار رمزگذار خودکار پراکنده قرار می‌دهیم خروجی لایه پنهان رمزگذار خودکار پراکنده یک بردار با ابعاد ۱۰۲۴ از تصویر تشکیل می‌دهد که این کاهش ابعاد، بردار ویژگی کم بعدی از تصویر را در اختیار ما قرار می‌دهد.

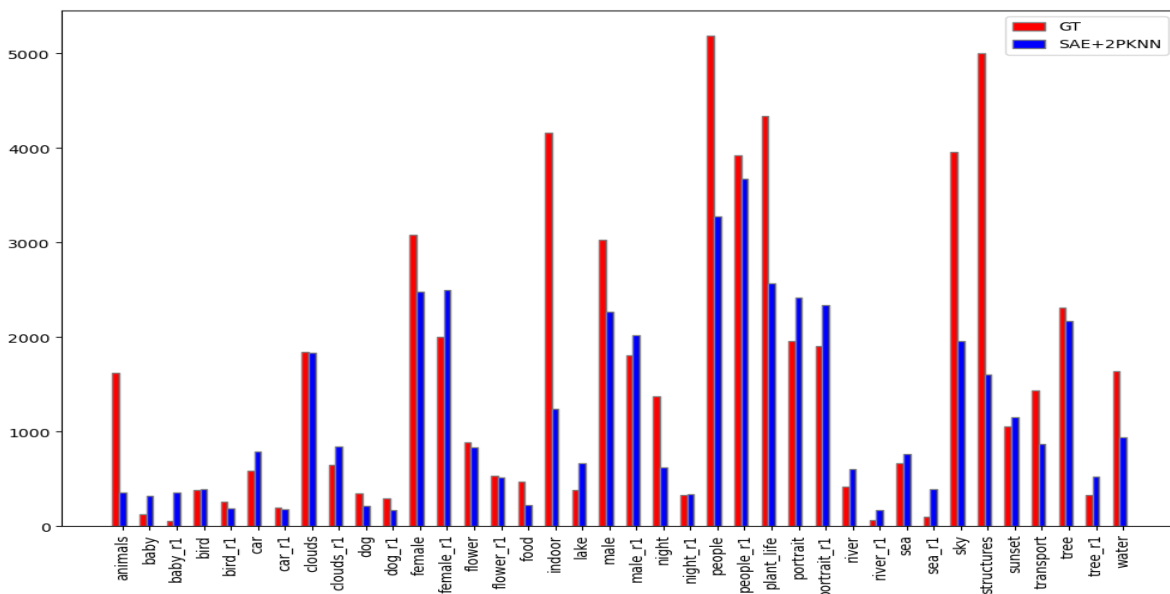
در این پژوهش، برای استخراج ویژگی از روش‌های یادگیری عمیق استفاده نمودیم و تصاویر مجموعه داده را به معماری‌های مطرح شده که در Image Net کارایی بالایی داشته‌اند، داده شد و ویژگی‌های با ابعاد بالا استخراج گردید. اما هنوز هم این بردارهای ویژگی تصاویر دارای ابعاد بالا، برای الگوریتم‌های برچسب زنی خودکار تصاویر مناسب نیستند و پردازش این بردارهای ویژگی باعث طولانی شدن زمان اجرای الگوریتم و همچنین کاهش دقت و کارایی الگوریتم خواهد بود. چون ما از روش مبتنی بر جستجو ۲PKNN برای برچسب گذاری خودکار تصاویر که یک الگوریتم سنتی اس، استفاده می‌کنیم، بردارهای ویژگی با ابعاد بالا هم مشکل مزاحمت ابعاد دارد و چالش مقیاس پذیری روش نیز وجود خواهد داشت. در نتیجه ما بردارهای ویژگی عمیق به دست آمده را به عنوان ورودی در اختیار رمزگذار خودکار پراکنده قرار می‌دهیم که بردار ویژگی خروجی این روش، بردار ویژگی با ابعاد کم و مناسب برای الگوریتم مبتنی بر جستجو خواهد بود در اینجا ما لایه های میانی رمزگذار و رمزگشا را با تعداد لایه های ۱۰۲۴، ۵۱۲، ۲۵۶ و ۲۰۴۸ آزمایش نمودیم و تابع فعال ساز را relu و selu قرار دادیم و خروجی حاصل از بردار ویژگی‌های به دست آمده نهایی با ابعاد ۵۱۲ و ۱۰۲۴ بعد را در مرحله بعد در اختیار روش ۲PKNN قرار دادیم، سپس نتیجه به دست آمده را ارزیابی نمودیم.

در ادامه معیارهای دقت بازخوانی و $F1$ به دست آمده حاصل از روش‌های مختلف مورد بررسی قرار گرفت که مشاهده شد کاهش ابعاد با روش رمزگذارهای خودکار پراکنده با ابعاد ۱۰۲۴، در مورد مجموعه داده مورد آزمایش مناسب‌تر به نظر می‌رسد. زیرا در کاهش ابعاد بیشتر از این مقدار، شاهد کاهش معنی دار معیارهای دقت، بازخوانی و $F1$ بودیم و مشخص شد که هر چند



با کاهش ابعاد بیشتر از این مقدار سرعت اجرای الگوریتم مبتنی بر جستجو ۲PKNN افزایش می‌یابد اما در معیارهای معرفی شده کاهش داشتیم و مقدار F^1 حاصل از کاهش ابعاد بیش از حد، حتی پایین‌تر از روش خطی PCA بود. بنابراین به نظر می‌رسد با کاهش ابعاد مناسب با استفاده از رمزگذار خودکار پراکنده، مشکل مزاحمت ابعاد برطرف می‌شود و زمان اجرای الگوریتم مبتنی بر جستجو به صورت معناداری کاهش می‌یابد.

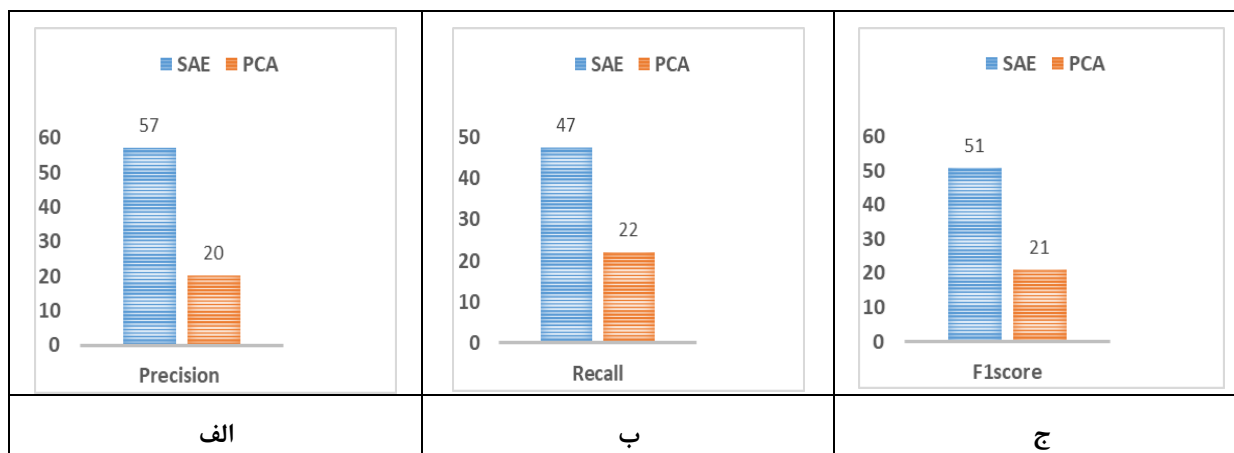
یکی از مشکلات الگوریتم های برچسب زنی خودکار تصاویر، مشکل عدم تعادل کلاس‌ها و برچسب‌گذاری ضعیف است (Cheng, ۲۰۱۸) با مشاهده شکل ۹ می‌توان نتیجه گرفت که بین مقادیر به دست آمده از روش پیشنهادی و تصاویر واقعی به غیر از چند کلاس برچسب تعادل نسبی برقرار است روش پیشنهادی در پیش‌بینی برچسب‌های کلاس structures, indoor و animals کارآمدی مطلوبی نداشته است. به عبارت دیگر مشکل عدم تعادل طبقاتی دارد یعنی



شکل ۹ - نمودار برچسب‌های پیش‌بینی شده با روش پیشنهادی

تعداد برچسب‌های پیش‌بینی شده تصاویر کمتر از تعداد برچسب‌های این کلاس از تصاویر می‌باشد و به نظر می‌رسد این مشکل مربوط به ویژگی‌های تصویر باشد که این روش استخراج ویژگی عمیق که یک روش استخراج ویژگی سراسری است در تشخیص کلاس animals کارایی مطلوب را نداشته است. یا در تشخیص برچسب‌های indoor و structures نیز به‌ماهیت تصویر مربوط است به‌علاوه اینکه در کلاس برچسب‌های female, river و portrait نیز برچسب‌های تخصیص داده شده روش پیشنهادی بیشتر از برچسب‌های واقعی این کلاس از تصاویر است. و این مورد به علت مشکل برچسب‌گذاری ضعیف رخ می‌دهد. البته به طور کلی این مقادیر با توجه به معیارهای ارزیابی الگوریتم خیلی قابل چشم‌گیر نبوده و می‌توان گفت که روش پیشنهادی به درستی مشکل عدم تعادل طبقاتی و برچسب‌گذاری ضعیف را رفع نموده است.

همانطور که در شکل ۱۰ مشاهده می‌شود روش استخراج ویژگی غیر خطی SAE در معیارهای دقت، یادآوری و F^1 به صورت معناداری نسبت روش خطی PCA عملکرد بهتری داشته است. البته در مورد این مجموعه داده نتیجه به این صورت مشاهده گردید. و در جدول ۱ نیز نتایج روش پیشنهادی با روش‌های قبلی مقایسه شده است.



شکل ۱۰- نمودار مقایسه معیارهای ارزیابی روش استخراج ویژگی نمودار قرمز مربوط به PCA و نمودار آبی مربوط SAE است.

جدول ۱- نتایج معیارهای ارزیابی روش برچسب‌زنی خودکار بر روی مجموعه داده MIRFLICKER با رمزگذارهای خودکار

MIRFLICKER Dataset				
ویژگی	روش	MicroAUC(%)	MacroAUC(%)	F1(%)
استاندارد	SAE+B*PKNN(Liu et al, ۲۰۱۹)	۷۱/۷±۰/۲	۷۰/۲±۰/۲	۳۴/۰±۰/۹
استاندارد	AE+B*PKNN	۷۳/۶±۰/۳	۷۱/۵±۰/۴	۳۷/۲±۰/۵
CNN(vgg۱۶)	SAE+B*PKNN	۸۹/۵±۰/۶	۸۹/۳±۰/۵	۵۰/۳±۱/۳
CNN(Resnet۱۰۱)	SAE+B*PKNN	۸۹/۶±۰/۶	۸۹/۱±۰/۹	۵۳/۰±۱/۲

۷- بحث و نتیجه‌گیری

استخراج ویژگی فرایند بسیار متداولی در انواع مختلف پردازش داده‌ها چون پردازش تصویر، پردازش صوت و غیره است. کاهش ابعاد که برای تولید یک نمایندگی سطح بالا در نظر گرفته شده، مهم و ضروری است. مرحله پیش پردازش به این دلیل است که اگرچه یادگیری ماشینی مدل‌ها از نظر تئوری باید بتوانند روی هر عددی از ویژگی‌ها کار کنند، اما مجموعه داده‌های با ابعاد بالا همیشه باعث ایجاد یک مجموعه‌ای از مشکلات از جمله برازش بیش از حد، پیچیدگی محاسباتی بالا می‌شوند و مدل‌های بیش از حد پیچیده نیز باعث یک موضوع شناخته شده مزاحمت ابعاد می‌شود. یکی دیگر از دلایل کاهش ابعاد این است که نمایش‌های سطح بالا می‌تواند به افراد کمک کند تا ساختار درونی داده‌ها را بهتر درک کنند.

هدف اصلی از رمزگذارهای خودکار، کدگذاری کارآمد داده است که این فرآیند تحت یادگیری بدون نظارت قرار می‌گیرد. رمزگذارهای خودکار به عنوان روش‌های استخراج‌کننده ویژگی و کاهش‌دهنده ابعاد مورد استفاده قرار می‌گیرند. رمزگذارهای خودکار چندین مزیت را برای برچسب‌گذاری تصویر و استخراج ویژگی ارائه می‌دهند که شامل استخراج خودکار ویژگی‌ها، کاهش ابعاد، بازسازی داده‌ها و مدیریت غیرخطی است و این مزایا آنها را به ابزاری قدرتمند برای کاربردهای مختلف مانند حذف نویز تصویر، فشرده‌سازی و بازیابی مبتنی بر محتوا تبدیل می‌کند. نتایج نشان می‌دهد استخراج ویژگی با استفاده از رمزگذارهای خودکار پراکنده که می‌تواند بردارهای ویژگی تصویر با ابعاد بالا را کاهش دهد و از مزاحمت ابعاد نیز جلوگیری نماید، باعث افزایش دقت و سرعت روش‌های برچسب‌زنی تصویر می‌شود.



منابع:

- Jia, Weikuan & Sun, Meili & Lian, Jian & Hou, Sujuan. (۲۰۲۲). Feature dimensionality reduction: a review. *Complex & Intelligent Systems*. ۸.
- Mishra, Deepti & Sharma, Saurabh. (۲۰۲۱). Performance Analysis of Dimensionality Reduction Techniques: A Comprehensive Review.
- S. Khalid, T. Khalil and S. Nasreen, (۲۰۱۴), "A survey of feature selection and feature extraction techniques in machine learning," *2014 Science and Information Conference*, London, UK, pp. ۳۷۲-۳۷۸,
- Miyashiro, K. Y., & Eberwine, J. H. (۲۰۲۱). Single cell transcriptomics and epigenomics methods provide high resolution genomics profiling of brain disorders. In *Epigenetics in Psychiatry* (pp. ۱۷۵-۱۹۶). Academic Press.
- Verleysen, M., François, D. (۲۰۰۵). The Curse of Dimensionality in Data Mining and Time Series Prediction. In: Cabestany, J., Prieto, A., Sandoval, F. (eds) *Computational Intelligence and Bioinspired Systems*. IWANN ۲۰۰۵. *Lecture Notes in Computer Science*, vol ۳۵۱۲. Springer, Berlin, Heidelberg.
- G. E. Hinton, R. R. Salakhutdinov, (۲۰۰۶), Reducing the Dimensionality of Data with Neural Networks. *Science* ۳۱۳, ۵۰۴-۵۰۷.
- Hinton, Geoffrey & Osindero, Simon & Teh, Yee-Whye. (۲۰۰۶). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*. ۱۸. ۱۵۲۷-۱۵۵۴.
- Yu, K. & Jia, L. & Chen, Y. & Xu, W.. (۲۰۱۳). Deep learning: yesterday, today, and tomorrow. ۵۰. ۱۷۹۹-۱۸۰۴.
- Baldi, P.. (۲۰۱۲). Autoencoders, Unsupervised Learning, and Deep Architectures. *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, in *Proceedings of Machine Learning Research*, ۲۷:۳۷-۴۹
- Gogoi, M., & Begum, S.A. (2017). *Image Classification Using Deep Autoencoders*. *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 1-5.
- Yan, Binghao & Han, Guodong. (۲۰۱۸). Effective Feature Extraction via Stacked Sparse Autoencoder to Improve Intrusion Detection System. *IEEE Access*. ۶. ۱-۱.
- Chen, S., & Guo, W. (۲۰۲۳). Auto-Encoders in Deep Learning-A Review with New Perspectives. *Mathematics*, 11(۸), ۱۷۷۷.
- Nugroho, Herminarto. (۲۰۲۰). Fully Convolutional Variational Autoencoder For Feature Extraction Of Fire Detection System. *Jurnal Ilmu Komputer dan Informasi*. ۱۳.
- Zhilei Yasi Wang, Wei Song, Huiling Wang, Fei Liu, (۲۰۱۹), A semi-supervised auto-encoder using label and sparse regularizations for classification, *Applied Soft Computing*, Volume ۷۷, Pages ۲۰۵-۲۱۷,
- Senefonte, Helen. (۲۰۲۲). Predicting mobility patterns based on profiles of social media users: tourists case study.
- Yuan, S., Zhang, H., Liu, Y. *et al.* Feature extraction-based intelligent algorithm framework with neural network for solving conditional nonlinear optimal perturbation. *Soft Comput* ۲۶, ۶۹۰۷-۶۹۲۴ (۲۰۲۲).
- Mundher, Myasar & Rahim, Mohd & Rehman, Amjad & Mehmood, Zahid & Saba, Tanzila & Naqvi, Rizwan. (۲۰۲۱). Automatic Image Annotation Based on Deep Learning Models: A Systematic Review and Future Challenges. *IEEE Access*. PP.
- Ian Goodfellow and Yoshua Bengio and Aaron Courville, (۲۰۱۶), *Deep Learning*, MIT Press.
- Aslam, Muhammad & Xue, Cuili & Chen, Yunsheng & Zhang, Amin & Liu, Manhua & Wang, Kan & Cui, Daxiang. (۲۰۲۱). Breath analysis based early gastric cancer classification from deep stacked sparse autoencoder neural network. *Scientific Reports*. ۱۱.
- Tripathy, S., Tabasum, M. (۲۰۲۳). Autoencoder: An Unsupervised Deep Learning Approach. In: Dutta, P., Chakrabarti, S., Bhattacharya, A., Dutta, S., Shahnaz, C. (eds) *Emerging Technologies in Data Mining and Information Security*. *Lecture Notes in Networks and Systems*, vol ۴۹۰. Springer
- Doersch, Carl. (۲۰۱۶). Tutorial on Variational Autoencoders. ۱۰, ۴۸۵۵/arXiv. ۱۶۰۶.۰۵۹۰۸.
- Karol Gregor, et al. "Temporal Difference Variational Auto-Encoder." *ICLR*, ۲۰۱۹.
- Wang, Crystal & Richardson, Yaateh & Sander, Ryan. (۲۰۱۹). Unsupervised Image Clustering and Topic Modeling for Accelerated Annotation.
- Patel, Tanvi & Kapadia, Mayank & Maisuria, Jemish. (۲۰۱۵). A Review on Content based Image Retrieval. *International Journal of Computer Applications*. ۱۳۲. ۲۲-۲۵. ۱۰.۵۱۲۰/ijca.۲۰۱۵۹.۷۵۸۴.
- Riaz, S., Arshad, A., & Jiao, L. (۲۰۱۸). Fuzzy Rough C-Mean Based Unsupervised CNN Clustering for Large-Scale Image Data. *Applied Sciences*, 8(۱۰), ۱۸۶۹.
- Olaode, Abass & Naghdy, Golshah. (۲۰۱۹). Local Image Feature Extraction using Stacked-Autoencoder in the Bag-of-Visual Word modelling of Images.
- Bengio, Y. (۲۰۰۹). Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, ۲(۱), ۱-۱۲۷.
- Shivalila Hangaragi, N. Neelima, Vivek Venugopal, Somnath Ganguly, Joyti Mudi, Joon-Ho Choi, (۲۰۲۵), CAE SynthImgGen: Revolutionizing cancer diagnosis with convolutional autoencoder-based synthetic image generation, *Alexandria Engineering Journal*, Volume ۱۱۵ (۲۰۲۵), Pages ۳۴۳-۳۵۴



- Guo, X., Liu, X., Zhu, E., Yin, J. (۲۰۱۷). Deep Clustering with Convolutional Autoencoders. , Neural Information Processing. ICONIP ۲۰۱۷. Lecture Notes in Computer Science(), vol ۱۰۶۳۵. Springer, Cham.
- Zhang, Y., Kong, L., Antwi-Afari, M. F., & Zhang, Q. (۲۰۲۴). An Integrated Method Using a Convolutional Autoencoder, Thresholding Techniques, and a Residual Network for Anomaly Detection on Heritage Roof Surfaces. *Buildings*, 14(۹), ۲۸۲۸.
- Ullah, Faizan & Salam, Abdu & Abrar, Muhammad & Ahmed, Masood & Ullah, Fasee & Khan, Atif & Alharbi, Abdullah & Alosaimi, Wael. (۲۰۲۱). Machine Health Surveillance System by Using Deep Learning Sparse Autoencoder. ۱۰,۲۱۲۰۲/rs.۲.rs-۹۱۱۳۸۵/v۱.
- Zhang, Changfan & Cheng, Xiang & Liu, Jianhua & He, Jing & Liu, Guangwei. (۲۰۱۸). Deep Sparse Autoencoder for Feature Extraction and Diagnosis of Locomotive Adhesion Status. *Journal of Control Science and Engineering*. ۲۰۱۸. ۱-۹.
- A. Ng, Sparse autoencoder, ۲۰۱۱. Cs۲۹۴a Lecture Notes.
- Chung, M., Archibald, R., Atzberger, P., & Solomon, J. M. (۲۰۲۴). Sparse L^1 -Autoencoders for Scientific Data Compression. *arXiv preprint arXiv:2405.14270*.
- Jianran Liu, Shiping Wang, Wenyuan Yang,(۲۰۱۹), Sparse autoencoder for social image understanding, *Neurocomputing*, Volume ۳۶۹, Pages ۱۲۲-۱۳۳.
- Yasi Wang, Hongxun Yao, Sicheng Zhao,(۲۰۱۶), Auto-encoder based dimensionality reduction, *Neurocomputing*, Volume ۱۸۴, Pages ۲۳۲-۲۴۲.



Types of Autoencoders and their application in feature extraction and image dimension reduction

Farzad Khodayari

Master's degree in Artificial Intelligence Shahid Rajaei Teacher Training, Computer Engineering Faculty Tehran Iran

Ehsan Pazoki

Assistant Professor Shahid Rajaei Teacher Training, Computer Engineering Faculty Tehran Iran

Abstract

A series of challenges in image tagging are the high computational volume in processing image data and the accuracy of image tagging methods. Despite image tagging methods using deep networks, the efficiency of these algorithms has not yet reached the desired level, and in real-time tasks, this speed problem can be inefficient. Dimensionality reduction is a way to reduce the complexity of a model and prevent overfitting. To reduce the initial dimensions of the data, linear methods are mainly used. However, since many of the problems we face are nonlinear and time-varying systems, current research on nonlinear feature reduction methods is relatively more. Autoencoders are a type of neural network that creates a better feature representation for high-dimensional input data and finds correlations between input data. Autoencoders are known to be more efficient in reducing nonlinear dimensions than other multiple learning methods. In this study, we used autoencoders to extract features and the MirFlicker dataset images were fed to the proposed architectures that have high performance in Image Net, but still these high-dimensional image feature vectors are not suitable for automatic image tagging algorithms and their processing will prolong the execution time of the algorithm and also reduce the accuracy and efficiency of the algorithm. Since we use the χ^2 PKNN prominence-based method for automatic image tagging, high-dimensional feature vectors have the problem of curse of dimensionality. As a result, we provided the obtained feature vectors as input to the sparse autoencoder, which will output a low-dimensional feature vector suitable for the prominence-based algorithm. The results of dimensionality reduction in the proposed method using the sparse autoencoder increased the F_1 criterion by 19.0 ± 0.3 compared to previous optimal works.

Keywords: Feature extraction, Dimensionality Reduction, Autoencoder, Curse of Dimensionality