

شرح‌نگاری خودکار تصویر با روش چرخش بلاک اصلی

حازم الرکابی، نسرین صوفی، هادی صدوقی یزدی و امیرحسین طاهری‌نیا

سطح پایین استفاده می‌شوند که از جمله آنها می‌توان به ویژگی‌های رنگ و بافت اشاره کرد. نتیجه جستجو به بهترین تناظر ممکن بین بردار ویژگی‌های استخراج‌شده و بردار ویژگی تصویر پرس‌وجو بستگی دارد. تنها مشکل روش‌های زیرمجموعه این دسته این است که کاربران علاقه‌ای به ویژگی‌های سطح پایین ندارند و نمی‌توانند تصاویر را بر اساس این ویژگی‌ها تفسیر کنند. آنها ترجیح می‌دهند با زبان‌های طبیعی به تفسیر تصاویر بپردازند. در دسته‌بندی و بازیابی تصاویر با استفاده از ویژگی‌های سطح پایین نمایش داده می‌شوند، زیرا تصویر یک آرایه بدون ساختار از پیکسل‌ها است. اولین گام از ادراک معنایی استخراج ویژگی‌های بصری مؤثر و کارا از همین پیکسل‌ها است. هدف اصلی روش‌های شرح‌نگاری تصاویر این است که به طور خودکار به هر تصویر آزمون چندین کلید واژه مرتبط نسبت دهند که این کلیدواژه‌ها بازتاب‌دهنده محتویات بصری آن تصویر هستند. یکی از رویکردهای بازیابی تصاویر مبتنی بر متن این است که در آن تصاویر به‌صورت دستی و به کمک انسان شرح‌نگاری می‌شوند و سپس این تصاویر مانند اسناد رسمی مورد بازیابی قرار می‌گیرند که البته به دلیل حجم گسترده تصاویر این کار غیر قابل انجام است و همچنین نظر افراد در شرح‌نگاری مؤثر خواهد بود. ایده اصلی شرح‌نگاری خودکار تصویر^۲ (AIA) روش‌هایی است که مدل‌های معنایی و مفهومی را با استفاده از مجموعه تصاویر نمونه بسیار زیاد به‌صورت خودکار یاد می‌گیرند و از این مدل‌های مفهومی برای برچسب‌زدن به تصاویر جدید و آزمون استفاده می‌کنند. هر تصویر دارای مفهومی است که این مفهوم از کنار هم قرار گرفتن ویژگی‌های تصویر به دست آمده و از آنجایی که اغلب اطلاعاتی که در شرح‌نگاری تصویر باید تحلیل گردند، غیر منفی هستند (مانند پیکسل‌های تصویر) و نیز به علت این که اطلاعات کاهش سطح یافته‌اند، نیاز بیشتری به دربرداشتن مقادیر غیر منفی دارند تا بتوانند با اطلاعات فیزیکی و واقعی تطابق داشته باشند. در این مقاله از روشی مرسوم به تجزیه نامنفی ماتریس^۳ (NMF) که تقریب ماتریس داده‌ها با مقادیر نامنفی است، استفاده شده و در روش پیشنهادی از روشی به نام چرخش بلاک اصلی^۴ [۱] برای حل NMF بهره گرفته شده است.

۱-۱- انگیزه

هدف اصلی این مقاله افزایش دقت و سرعت فرایند شرح‌نگاری خودکار تصاویر با به کارگیری الگوریتم NMF و استفاده از روش چرخش بلاک اصلی برای حل آن است. از جمله ویژگی‌های این سیستم، توانایی افزودن برخط کلاس جدیدی از داده‌ها به دانش خود است. علاوه بر این یادگیری دانش به صورت فشرده انجام می‌شود و توانایی آموزش بر اساس داده‌های دریافتی بدون نیاز به پردازش مجدد داده‌های قبلی، از جمله دیگر ویژگی‌های آن است. در این سیستم با ورود کلاسی از داده‌ها به

چکیده: سیستم‌های شرح‌نگاری خودکار تصاویر، وظیفه توصیف محتوای تصاویر به وسیله تخصیص برچسب به آنها را بر عهده دارند. هدف از انجام این تحقیق بهبود نتایج دقت و سرعت یک سیستم شرح‌نگار تصاویر است. اخیراً با توجه به رشد روزافزون تصاویر، فرایند شرح‌نگاری بر روی پایه‌های تصاویر به جای خودشان اجرا می‌گردد. یکی از این روش‌های جدید، پیاده‌سازی الگوریتم تجزیه نامنفی ماتریس (NMF) بر روی ویژگی‌های به دست آمده از تصاویر است. در روش پیشنهادی برای افزایش سرعت و کارایی بهتر سیستم شرح‌نگاری، برای اولین بار از روشی به نام چرخش بلاک اصلی برای حل NMF در شرح‌نگاری استفاده شده است. این روش با توانایی افزودن برخط کلاس جدیدی از داده‌ها به دانش خود و یادگیری دانش به صورت فشرده و علاوه بر آن، توانایی آموزش بر اساس داده‌های دریافتی بدون نیاز به پردازش مجدد توانسته از روش‌های پیشین ارائه‌شده برای حل NMF عملکرد بهتری را نشان دهد. در مرحله آموزش با روش چرخش بلاک اصلی ماتریس ضرایب و پایه تصاویر ورودی به دست می‌آیند. سپس در مرحله آزمون برای تصویر ورودی، توسط ویژگی‌های استخراج‌شده از تصویر و ضرایب به دست آمده از مرحله آموزش، ضریب تعلق تصویر آزمون به هر یک از کلاس‌های تصاویر آموزش محاسبه می‌گردد. سپس این ضریب در هنگام جستجو در میان تصاویر آموزش برای تخصیص برچسب به تصویر آزمون، دقت کار را افزایش می‌دهد. این جستجو توسط روش KNN بر روی پایه‌های تصاویر صورت می‌گیرد. برای آزمایش روش پیشنهادی از دو پایگاه داده Corel5K و داده‌های واقعی حیوانات (برگرفته از ۵۰۰ px) استفاده شده و نهایتاً با روش‌های موجود مقایسه شده که در پایگاه داده Corel5K به میزان دقت ۵۰/۲۰ و روی داده‌های واقعی به ۶۲/۸۹ رسیدیم که به‌طور قابل ملاحظه‌ای دقت افزایش یافته است.

کلیدواژه: تجزیه نامنفی ماتریس، چرخش بلاک اصلی، k- نزدیک‌ترین همسایه، شرح‌نگاری تصاویر.

۱- مقدمه

شرح‌نگاری خودکار تصاویر مسأله‌ای مهم در بسیاری از کاربردهای مختلف مانند بازیابی تصاویر بر اساس محتوا^۱ (CBIR) و یا ادراک تصاویر است. در واقع بازیابی تصاویر بر اساس محتوا زیرمجموعه‌ای از مبحث بازیابی تصاویر می‌باشد که تمرکز آن بر روی محتوای تصویر برای جستجوی موضوع یا هدفی خاص است. در CBIR ویژگی‌های محتویات

این مقاله در تاریخ ۸ اردیبهشت ماه ۱۳۹۷ دریافت و در تاریخ ۲۲ آبان ماه ۱۳۹۷ بازنگری شد.
حازم الرکابی، گروه کامپیوتر، دانشگاه فردوسی مشهد، مشهد، ایران، (email: hazim.al-rikabi@mail.um.ac.ir)
نسرین صوفی، گروه کامپیوتر، دانشگاه فردوسی مشهد، مشهد، ایران، (email: soufi.nasrin@mail.um.ac.ir)
هادی صدوقی یزدی (نویسنده مسئول)، گروه کامپیوتر، دانشگاه فردوسی مشهد، مشهد، ایران، (email: h-sadoghi@um.ac.ir)
امیرحسین طاهری‌نیا، گروه کامپیوتر، دانشگاه فردوسی مشهد، مشهد، ایران، (email: taheerinia@um.ac.ir)

2. Auto Image Annotation
3. Non-Negative Matrix Factorisation
4. Block Principal Pivoting

1. Content-Based Image Retrieval

روی محتوای تصویر انجام نمی‌گیرد بلکه متنی که همراه تصویر وجود دارد تحلیل و پردازش می‌شود. این متن ممکن است مستندات همراه تصویر یا محتوای صفحه وبی باشد که تصویر در آن وجود دارد.

۱-۲-۱-۲ روش‌های شرح‌نگاری خودکار مبتنی بر تصویر

یکی دیگر از روش‌های شرح‌نگاری خودکار تصاویر، روش مبتنی بر تصویر است. این روش نیازمند داشتن متن مرتبط با تصویر نیست. اساس کار در اینجا بر پردازش تصویر و استنتاج مفاهیم هر تصویر از خود تصویر است. هدف این است که با استفاده از ویژگی‌های سطح پایین مستخرج از تصویر بتوان ویژگی‌های سطح بالای آن را به دست آورد. منظور از ویژگی‌های سطح پایین تصویر ویژگی‌هایی مانند رنگ، بافت، شکل و غیره است در حالی که ویژگی‌های سطح بالا گویای مفاهیم و کلمات قابل برداشت از تصویر می‌باشد. یکی از چالش‌های مهم در این زمینه فاصله‌ای است که بین این دو دسته ویژگی وجود دارد و به شکاف معنایی معروف است، لذا به منظور شرح‌نگاری با استفاده از این روش باید شکاف معنایی را از بین برد. در واقع باید نگاهی بین ویژگی‌های سطح پایین و کلمات پیدا کرد و برای تولید این نگاشت از روش‌های یادگیری ماشین استفاده می‌شود. در واقع شرح‌نگاری تصاویر به نوعی به دسته‌بندی تصاویر تبدیل می‌شود که هر دسته بیانگر کلمه یا عبارتی در مورد تصویر است [۲]. فرایند دسته‌بندی و یادگیری نیازمند مجموعه‌ای از تصاویر مختلف حاشیه‌دار است. این مجموعه، مجموعه آموزش نامیده می‌شود. فرایند تولید سیستم شرح‌نگاری شامل انتخاب مجموعه آموزش، استخراج ویژگی‌ها از تصاویر و تولید نگاشت یا مدل بین ویژگی‌های سطح پایین و سطح بالای تصاویر با استفاده از الگوریتم‌های یادگیری ماشین است. بلوک دیاگرام کلی سیستم شرح‌نگاری در شکل ۱ آمده است.

۱-۲-۱-۲ انواع رویکردها در طراحی سیستم شرح‌نگاری تصویر

رویکردهای استفاده‌شده برای طراحی سیستم‌های شرح‌نگاری خودکار می‌تواند به سه گروه تقسیم‌بندی شوند. اولین گروه مدل‌های مولد هستند که تصاویر را به‌عنوان نمونه در نظر می‌گیرند که از یک توزیع آماری خاص نمونه‌برداری شده‌اند و پارامترهای توزیع مجموعه داده‌های آموزشی تخمین زده می‌شود. مدل‌های ترکیبی مثل GMM^1 [۳] و مدل‌های عناوین^۲ مانند تخصیص دیریکله پنهان^۳، تجزیه و تحلیل معنایی پنهان، تجزیه و تحلیل معنایی پنهان احتمالاتی [۴] و NMF دو زیرگروه مدل‌های مولد هستند. مطالعه برجسته این گروه روش $NMF-KNN$ [۵] است. این یک مدل مولد مبتنی بر جستجو است که یک مدل برتر برای هر تصویر را یاد می‌گیرد. ساخت یک مدل جداگانه برای هر تصویر آزمون، عملکرد را به طور قابل ملاحظه‌ای افزایش می‌دهد اما ساخت مدل نیاز به محاسبات زیادی دارد مخصوصاً وقتی تعداد زیادی تصویر آزمون داشته باشیم. گروه دوم مدل‌های جداساز نامیده می‌شوند که طبقه‌بندی جدا روی هر برچسب ایجاد می‌کند. هر تصویر به همه طبقه‌بندها اعمال می‌شود تا تعیین کند به کدام کلاس تعلق دارد. از جمله طبقه‌بندهایی که به این منظور استفاده شده‌اند می‌توان به SVM^4 ، یادگیری نمونه چندگانه^۵ و شبکه‌های عصبی^۶ اشاره کرد. یکی از بهترین کارهای این

توانایی شرح‌نگاری تصویر اضافه می‌شود و نوع عملکرد این سیستم‌های جستجوگر بر اساس نحوه عملکرد کاربران مهیا شده است. به طور مثال، کاربر با جستجوی موضوعی دلخواه تمام تصاویر شامل این برچسب را دریافت می‌کند که البته ممکن است برچسب‌های دیگری نیز داشته باشد. همچنین این امکان فراهم است که هزاران تصویر با برچسب به مجموعه داده اضافه گردد. علاوه بر آن با به کارگیری روش چرخش بلاک اصلی برای حل NMF ، موجب رفع محدودیت روش مبتنی بر مجموعه فعال شده و باعث افزایش سرعت حل مسأله می‌گردد.

در ادامه مقاله مروری بر کارهای صورت‌گرفته در حوزه مسئله شرح‌نگاری خواهیم داشت. در بخش سه به تشریح هر یک از اجزای سیستم شرح‌نگاری و معرفی ایده روش پیشنهادی در همان بخش می‌پردازیم. در بخش چهار نتایج آن را با سایر روش‌ها مقایسه و نهایتاً جمع‌بندی و نتیجه‌گیری خواهیم نمود.

۱-۲-۱-۲ مروری بر کارهای گذشته

این بخش برای درک بهتر سیستم شرح‌نگاری به توضیح و بررسی روش‌های شرح‌نگاری، انواع رویکردها در طراحی سیستم شرح‌نگاری تصاویر و بررسی سیستم‌های مبتنی بر NMF می‌پردازد.

۱-۲-۱-۱ روش‌های شرح‌نگاری تصاویر

به‌طور کلی شرح‌نگاری تصاویر به صورت دستی، خودکار و نیمه‌خودکار انجام می‌شود.

۱-۲-۱-۱ روش دستی

در این روش شرح‌نگاری تصاویر را انسان انجام می‌دهد. مهم‌ترین ویژگی این روش دقت آن است زیرا کلمات حاشیه توسط انسان تولید شده و لذا هوشمندانه و نزدیک به برداشتی است که از تصاویر استنتاج می‌شود. از آنجا که در بیشتر مواقع با مجموعه‌های بزرگی از تصاویر مواجه هستیم لذا توصیف و شرح‌نگاری هر تصویر به صورت دستی زمان‌بر و پرهزینه است. از طرفی کلمات به کار برده شده کاملاً تحت تأثیر شخصیت، روحیات، سطح علمی و دید فرد نسبت به مسایل مختلف است. این مسأله ممکن است منجر به تولید کلمات متفاوت برای تصاویر مشابه شود، تنها به این دلیل که توسط افراد متفاوت و یا یک فرد در موقعیت‌های متفاوت شرح‌نگاری شده باشند.

۱-۲-۱-۲ روش‌های خودکار/ نیمه‌خودکار

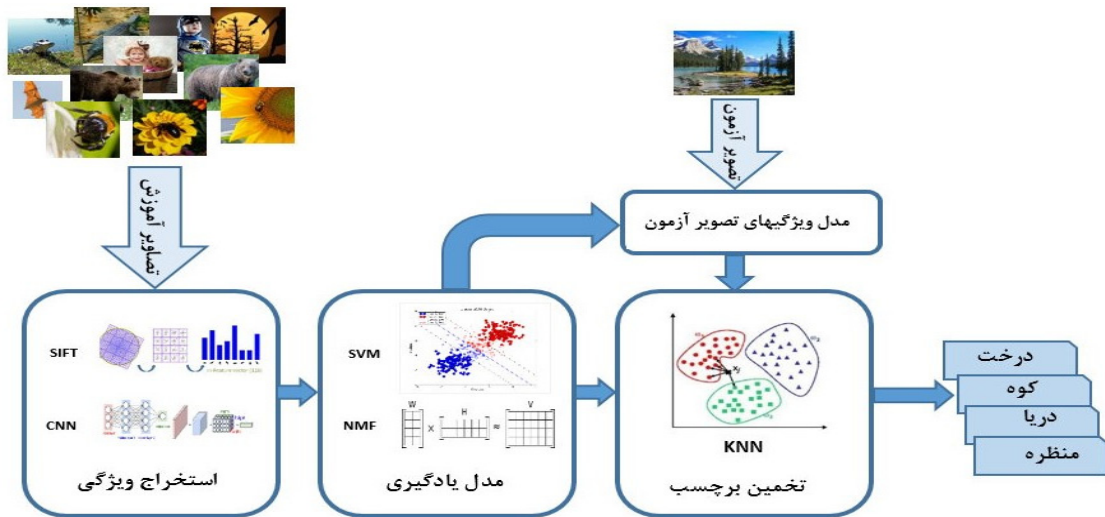
در روش خودکار سعی بر این است تا فرایند شرح‌نگاری تصاویر به طور کامل توسط ماشین انجام شود. دقت این گونه روش‌ها کمتر از دیگر روش‌هاست در حالی که در روش‌های نیمه‌خودکار کاربران در مورد حاشیه‌های تصویر تصمیم‌گیری می‌کنند و معمولاً رابط کاربری هوشمندی تولید می‌شود که اجرای عملیات به صورت دستی را کم می‌کند.

بعضی سیستم‌های خودکار یا نیمه‌خودکار قدرت یادگیری را از آنچه کاربران وارد می‌کنند نیز دارند تا در مورد تصاویر جدید حاشیه‌های دقیق‌تری تولید کنند. به طور کلی روش‌های خودکار شرح‌نگاری تصاویر به دو دسته کلی تقسیم می‌شوند: روش‌های مبتنی بر متن و روش‌های مبتنی بر تصویر.

۱-۲-۱-۲-۱ روش‌های شرح‌نگاری خودکار مبتنی بر متن

این روش‌ها تنها زمانی کاربرد دارند که در مورد هر تصویر اطلاعات متنی نیز وجود داشته باشد. در این حالت فرایند شرح‌نگاری شامل استخراج مفاهیم از متن مرتبط با تصویر خواهد بود و هیچ گونه پردازشی

1. Gaussian Mixture Models
2. Topic Models
3. Latent Dirichlet Allocation
4. Support Vector Machine
5. Multiple Instance Learning
6. Neural Network



شکل ۱: بلوک دیاگرام کلی سیستم‌های شرح‌نگاری خودکار تصاویر. ابتدا ویژگی‌های اولیه و متنوعی از تصاویر به نمایندگی از محتوای آنها استخراج می‌گردد. سپس رابطه بین این ویژگی‌ها و برچسب‌های درون مجموعه تصاویر آموزش، توسط یک سیستم یادگیر فراگرفته می‌شود. سرانجام در مرحله آزمون، برچسب‌هایی که محتوای تصویر مورد سؤال را توصیف می‌کنند، تخمین زده می‌شوند.

نمونه داده متناظر را تخمین می‌زند. در حقیقت فضای جدیدی می‌سازد و V نمایش داده در این فضا را نشان می‌دهد. با استفاده از چارچوب NMF در سیستم‌های شرح‌نگاری خودکار، عامل‌های پنهان اساسی از یک مجموعه داده استخراج می‌گردد، فضای پنهان مبتنی بر این عامل‌ها (بردارهای پایه U) ساخته می‌شود و نمایش تصاویر با استفاده از محاسبه عامل‌های موجود برای هر تصویر (ماتریس ضرایب V) به دست می‌آید.

یکی از توسعه‌های NMF، NMF چنددیده^۳ است که از ماتریس‌های داده چندگانه استفاده می‌کند. هر ستون از این ماتریس‌ها، داده یکسانی را با زاویه دید متفاوتی توصیف می‌کند. استراتژی دیگر، تجزیه یک دیدگاه و استفاده از ماتریس ضرایب آن برای پیدا کردن ماتریس پایه دیدگاه‌های دیگر است. در حقیقت در این روش‌ها فضای پنهان بر پایه اولین دیدگاه استخراج می‌شود و سایر دیدگاه‌ها نقش قابل توجهی در این پروسه ندارند. ثابت شده که تجزیه NMF منحصربه‌فرد نیست و ممکن است عامل‌های پنهان نامحدودی برای هر ماتریس پیدا شود. علاوه بر این، همه دیدگاه‌ها تصاویر یکسانی را نشان می‌دهند، بنابراین انتظار می‌رود که تقریباً مفاهیم پنهان در همه دیدگاه‌ها در نظر گرفته شود. آخرین مطالعات در مورد NMF چنددیده از این واقعیت‌ها استفاده کرده‌اند و سعی کرده‌اند به طور هم‌زمان همه دیدگاه‌ها را تجزیه کنند به طوری که یک ترم تعمیم‌کننده در تابع وجود دارد تا ماتریس ضرایب بین همه دیدگاه‌ها به اشتراک گذاشته شود. این تجزیه می‌تواند با قید نرم (ماتریس ضرایب باید شبیه باشد) یا سخت (ماتریس ضرایب باید یکسان باشد) انجام شود. در [۹] NMF مشترکی با قید سخت برای طبقه‌بندی چندکلاسه فعالیت انسان با یکی کردن دیدگاه‌های تصاویر مختلف ارائه شده است.

در [۱۰] NMF چنددیده با قید نرم برای مسأله خوشه‌بندی ارائه شده است که یک ساختار پنهان سازگار برای همه دیدگاه‌ها استخراج می‌کند. NMF-KNN از توسعه وزنی [۱۰] برای یک مدل مولد پرس‌وجو استفاده می‌کند به این صورت که یک مدل منحصر به فرد را برای هر تصویر آزمون یاد می‌گیرد که مبتنی بر نزدیک‌ترین همسایه‌هاست و یک قید اجماع برای ویژگی‌های متفاوت استخراج‌شده از آن تصویر آزمون اعمال

گروه KSVM-VT است. در [۶] یک روش ترکیبی طراحی شده که PLSA را به عنوان مرحله یادگیری مولد و مجموعه‌ای از طبقه‌بندهای SVM را برای طبقه‌بندی داده‌های چندبرچسب به عنوان مرحله یادگیری جداساز استفاده می‌کند. گروه سوم روش‌های مبتنی بر جست‌وجو است که ساده و قدرتمند هستند. این روش‌ها شباهت تصاویر آموزشی به تصویر آزمون را بر مبنای سنجش شباهت بصری جستجو می‌کند و برچسب‌های تصاویر آموزشی را به تصویر آزمون منتقل می‌کند. در حقیقت برای این گروه مرحله مدل‌سازی سیستم شرح‌نگاری خودکار، کوتاه است و تمرکز اصلی روی مرحله پیش‌بینی است که روش‌های یادگیری محلی از قبیل KNN به این گروه مربوط می‌شوند. استفاده از یادگیری متریک در کنار روش‌های مبتنی بر جست‌وجو، پیشرفت قابل توجهی داشته است. TagProb [۷] یکی از مطالعات مهم در این گروه است زیرا از مجموعه گسترده‌ای از ویژگی‌ها استفاده می‌کند و یادگیری متریک آن رویکردی مبتنی بر KNN است. این روش نتایج را با در نظر گرفتن برچسب‌های کمیاب بهبود می‌دهد. مجموعه ویژگی استفاده‌شده در TagProb در اغلب روش‌های سیستم شرح‌نگاری خودکار استفاده شده است.

۱-۲-۳ بررسی سیستم‌های مبتنی بر NMF

آقای لی و سونگ [۸] روش NMF را ارائه کردند که به طور گسترده در زمینه‌های مختلفی از قبیل جدایی منبع کور، طبقه‌بندی، تشخیص نفوذ^۴، شناسایی فعالیت‌ها، پردازش ویدئو، پردازش تصویر و بینایی کامپیوتر در طول دهه‌ها استفاده شده است. ماتریس نامنفی $X \in R^{n \times m}$ داده شده است. NMF با تقسیم داده‌ها به مجموعه‌ای از ماتریس‌های نامنفی $U \in R^{n \times k}$ (ماتریس پایه) و $V \in R^{m \times k}$ (ماتریس ضرایب یا نمایش)، یک فضای پنهان را آشکار می‌کند، بنابراین UV^T می‌تواند X را تخمین بزند. این امر با روش کمینه‌سازی تابع هدف زیر اتفاق می‌افتد

$$f = \min_{U, V} \|X - UV^T\|_F \quad \text{s.t. } U, V \geq 0 \quad (1)$$

اگر هر ستون X یک نمونه تصویر را نشان دهد، ستون‌های ماتریس U بردارهای پایه فضای جدید و سطرهای V ضرایب هر بردار پایه

1. Blind Source Separation
2. Intrusion Detection

این نمایش‌ها با استفاده از روش NMF تجزیه می‌گردد و فضاهای پنهان چندگانه پیدا شده که بردارهای پایه این فضاها یکسان نمایش داده می‌شود اما لزوماً محتوای یکسانی ندارند. این بدین معنی است که وقتی تصویری به این فضاها نگاشت داده شود نمایش آن در تمام فضاها شبیه هم است. در [۱۳] فرایند تجزیه به دو روش متفاوت طراحی شده است. یک روش MvNMF-SK نامیده می‌شود که تجزیه‌ای هم‌زمان با تعداد یکسانی از پایه، روی همه نمایش‌ها انجام می‌دهد و روش دیگر MvNMF-DK است که تجزیه‌ای جدا برای هر نمایش با تعداد متفاوتی از بردارهای پایه انجام می‌دهد.

۲- سیستم پیشنهادی

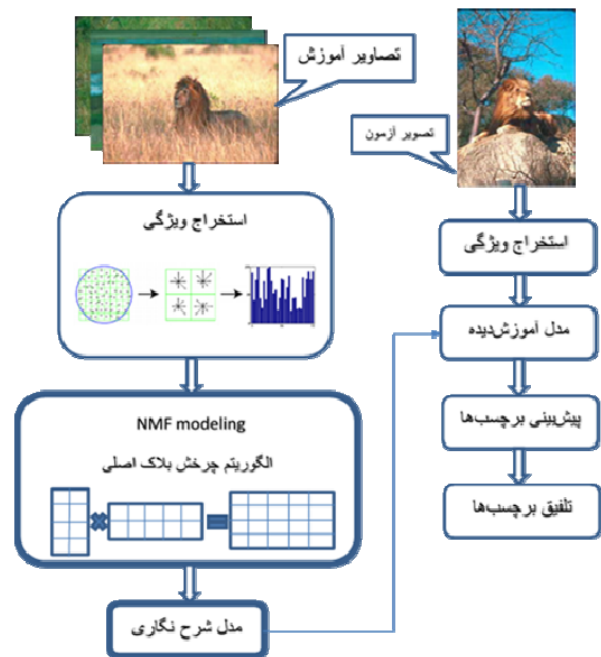
در شکل ۲ نحوه عملکرد سیستم پیشنهادی برای شرح‌نگاری تصاویر نشان داده شده است. بلوک‌های پررنگ شکل ۲، قسمت‌هایی را نشان می‌دهند که در این مقاله اصلاحات لازم در آنها صورت گرفته است. در این سیستم دو مرحله اساسی وجود دارد. در مرحله اول که مرحله آموزش نام دارد باید داده‌ها به عنوان ورودی به سیستم داده شوند. به دلیل حجم بالای تصاویر مورد استفاده که باعث کاهش سرعت پردازش می‌شوند، ویژگی‌های تصاویر استخراج می‌شوند. سپس وارد قسمت اصلی مرحله آموزش یعنی الگوریتم یادگیری ماشین می‌شویم. این قسمت ویژگی‌های تصاویر را به عنوان ورودی دریافت می‌کند. الگوریتمی که در این مرحله استفاده شده روش تجزیه نامنفی ماتریس (NMF) و الگوریتم مورد استفاده برای حل NMF روش چرخش بلاک اصلی در چارچوب ANLS^۵ است. در این فضا شباهت بین تصاویر با استفاده از ضرایب ماتریس NMF محاسبه می‌شود. لذا خروجی این الگوریتم، ماتریس ضرایب (H) و ماتریس پایه‌ها (W) است که در مرحله آزمون مورد استفاده قرار می‌گیرند. سپس وارد مرحله دوم یعنی مرحله آزمون می‌شویم. برای شروع مرحله آزمون یک تصویر به عنوان ورودی خوانده می‌شود و پس از استخراج ویژگی‌های آن، الگوریتم KNN^۶ اجرا می‌گردد تا نزدیک‌ترین کلاس‌ها به تصویر آزمون یافت شود. با پیدا شدن کلاس‌های مشابه، حال باید ویژگی‌های استخراج‌شده از تصویر و ماتریس پایه کلاس‌های مشابه به دست آمده از مرحله آموزش را به مسأله زیر بدهیم تا بردار ضرایب به دست آید

$$f = \min_{H_{test}} \|A_{test} - W \times H_{test}\|_F \quad (2)$$

و این مسأله به صورت زیر حل می‌شود

$$\frac{\partial f}{\partial H_{test}} = 0 \Rightarrow H_{test} = (inv(W^T \times W)) \times (W^T \times A_{test}) \quad (3)$$

منظور از inv معکوس یک ماتریس است. در گام بعدی برای به دست آوردن نزدیک‌ترین تصاویر به تصویر آزمون (منظور از A_{test} ارائه بصری تصویر آزمون است که بر مبنای ویژگی‌های استخراج‌شده از آن می‌باشد) باید بردار ضرایب (H_{test}) به دست آمده و ماتریس ضرایب آموزش داده شده را بار دیگر به الگوریتم KNN بدهیم. در این مرحله برچسب‌هایی برای تصویر آزمون پیش‌بینی می‌شود. در مرحله نهایی برچسب‌های پیش‌بینی شده به گونه مؤثری تلفیق داده می‌شوند و به این طریق سیستم



شکل ۲: روندنمای سیستم پیشنهادی.

می‌شود. در [۱۱] الگوریتم پیشنهادی FastTag به طور طبیعی می‌تواند شامل توصیف‌گرهای زیادی برای تصویر باشد و مشکلات پراکندگی برچسب را با روشی جدید حل کند. در این الگوریتم، داده‌های آموزشی خود را (تصاویر با برچسب‌های جزئی) به عنوان داده چنددیده بدون برچسب شرح می‌دهد و دو طبقه‌بند برای پیش‌بینی برچسب‌ها آموزش داده می‌شود. یکی مجموعه برچسب کامل (ناشناخته) را از برچسب‌هایی که در طول مرحله آموزش در دسترس است می‌سازد و دیگری نگاشتی را از ویژگی‌های تصویر به مجموعه برچسب ساخته‌شده یاد می‌گیرد. همچنین تابع ضرر پیوسته محدبی ارائه شده که این دو طبقه‌بند را با هم ترکیب کند. در [۱۲] الگوریتمی با دو مرحله ارائه شده است. در مرحله اول، مدلی برای ایجاد فضای جدید عوامل پنهان ساخته می‌شود. برای این منظور بعد از استخراج ویژگی، دو ماتریس داده ایجاد می‌شود (یکی از ویژگی‌های بصری و دیگری از اطلاعات برچسب‌ها). سپس الگوریتم NMF چندمنظوره^۱ با رویکرد مشترک بر روی این ماتریس‌ها اجرا می‌شود. بعد از آن فضای جدید عوامل پنهان با این ماتریس‌های پنهان ایجاد می‌گردد و تصاویر به این فضا نگاشت داده می‌شوند. در مرحله دوم یک روش مبتنی بر جستجو با استفاده از الگوریتم KNN وزن‌دار روی این فضای جدید اجرا می‌شود و برخی برچسب‌های تصاویر آزمون بر پایه نتایج این مرحله پیش‌بینی می‌شود. هدف اصلی در [۱۳] پیش‌بینی برخی برچسب‌های مربوط به تصاویر برای کمک به بازیابی معنایی^۲ است. بعد از استخراج ویژگی‌های بصری چندگانه، ماتریس داده (نمایش^۳) برای هر نوع ویژگی ساخته می‌شود به این صورت که بردار ویژگی هر تصویر مجموعه داده در یک ستون جدا گذاشته می‌شود. بنابراین ماتریس‌های داده چندگانه‌ای داریم که هر یک تصاویر را مبتنی بر ویژگی منحصر به فردی توصیف می‌کنند. علاوه بر این، ماتریس دیگری بر پایه برچسب‌های واقعی ساخته می‌شود که به آن نمایش برچسب^۴ گفته می‌شود. سپس همه

1. Multimodal
2. Semantic Retrieval
3. View
4. Tag View

5. Alternating Nonnegative Least Squares
6. K Nearest Neighbour

تخمین خوبی از (۴) با روش‌های نرم فرینوس^۴، واگرایی کولیک-لیبر^۵ و واگرایی برگمن^۶ انجام شده است. در اینجا نرم فرینوس مطرح می‌شود. عوامل W و H با حل مسأله بهینه‌سازی زیر پیدا می‌شوند

$$\min_{W \geq 0, H \geq 0} f(W, H) = \frac{1}{2} \|A - WH\|_F^2 \quad (۵)$$

مشکل (۵) این است که یک مسأله بهینه‌سازی نامحدب می‌باشد و این نشان می‌دهد که حل NMF ناممکن^۷ است و در عمل باید در جستجوی کمینه محلی باشیم. الگوریتم‌های زیادی برای حل (۵) تا کنون طراحی شده‌اند. اخیراً چندین الگوریتم مبتنی بر چارچوب حداقل مربعات نامنفی متناوب (ANLS) با کارایی خوبی ارائه شده است. الگوریتمی که هر ستون W و H را در هر تکرار به‌روزرسانی می‌کند و همچنین حداقل مربعات متناوب سلسله‌مراتبی^۸ (HALS) اخیراً ارائه شده است. الگوریتم سریع و جدیدی برای NMF با استفاده از روش چرخش بلاک اصلی در چارچوب ANLS ارائه شده است. الگوریتم‌های قبلی NMF با چارچوب ANLS از روش‌های مجموعه فعال^۹، projected gradient و projected quasi-Newton استفاده کرده‌اند. هدف استفاده از روش NMF برای سیستم‌های شرح‌نگاری تصویر، تعیین اثر عوامل پنهان اساسی (به‌عنوان مثال ماتریس پایه W) در نمایش محتوای تصاویر در کل مجموعه داده و سپس وقوع این عوامل (به‌عنوان مثال ماتریس ضرایب H) در هر تصویر است.

۲-۲-۲ چارچوب ANLS برای NMF

ابتدا برای حل مسأله (۴) روش ANLS را مرور می‌کنیم. در چارچوب ANLS متغیرها با روشی ساده به دو گروه، تقسیم و این دو گروه به نوبه خود به‌روزرسانی می‌شوند. این روش به صورت زیر خلاصه می‌گردد:

(۱) ابتدا $H \in R^{m \times k}$ با عناصر نامنفی مقداردهی اولیه می‌شود.

(۲) تکرار حل مسأله زیر تا زمانی که شرط توقف برقرار شود

$$\min_{W \geq 0} \|H^T W^T - A^T\|_F^2 \quad (۱-۶)$$

در اینجا H ثابت در نظر گرفته می‌شود

$$\min_{W \geq 0} \|WH - A\|_F^2 \quad (۲-۶)$$

و در اینجا W ثابت در نظر گرفته می‌شود.

(۳) ستون‌های W با نرم ۲ نرمال‌سازی می‌شوند و سپس ضرایب سطرهای H بر اساس بردارهای پایه W مقداردهی می‌گردد.

روش دیگر این است که ابتدا W و (۲-۶) مقداردهی شوند و سپس (۱-۶) تکرار گردد. نکته قابل توجه این است که هر زیرمسأله (۶) نمونه‌ای از مسأله حداقل مربعات نامنفی مقید^{۱۰} (NNLS) است. اگرچه مسأله (۵) نامحدب بود اما زیرمسأله‌های (۶) محدب هستند که می‌توان برای هر کدام راه‌حل‌های بهینه‌ای یافت.

برای این که همگرایی هر الگوریتم NMF مبتنی بر چارچوب ANLS را بررسی کنیم، پیدا کردن راه‌حل‌های بهینه (۶) در هر تکرار مهم است. بر

شرح‌نگاری تصویر پیاده‌سازی می‌شود. در ادامه به تشریح هر کدام از مراحل الگوریتم پیشنهادی می‌پردازیم.

۲-۱-۲ استخراج ویژگی‌های تصاویر

همان طور که گفته شد استخراج ویژگی‌های تصویر برای دو مجموعه داده آماده و واقعی با روش‌های SIFT^۱ و CNN انجام می‌شود. دلیل این انتخاب برای داده‌های آماده این است که باید شرایط مقایسه روش پیشنهادی با سایر روش‌ها یکسان باشد. در نتیجه از ویژگی‌های SIFT برای داده‌های آماده و از ویژگی‌های CNN که برای شرح‌نگاری مناسب‌ترند برای داده‌های واقعی استفاده گردیده است. در ادامه به توضیح مختصری از هر کدام می‌پردازیم.

SIFT ۱-۱-۲

SIFT به طور گسترده در بسیاری از برنامه‌های کاربردی مانند شناخت شیء، ردگیری ویدئو و غیره استفاده شده است. مراحل اصلی SIFT عبارتند از ساخت هرم گوسی تصاویر، اختلاف تصاویر هرم گوسی (DoG) و همچنین شناسایی نقاط کلیدی (دارای بیشترین مقدار) با همسایه‌های $3 \times 3 \times 3$ در دامنه DoG. هر نقطه کلیدی از موقعیتش در مقیاس خاص و جهت معین تشکیل شده است. تطبیق نقطه کلیدی بر اساس توصیف‌گری انجام می‌شود که شامل هیستوگرام با ۱۶ جهت (۸ بُعد برای هر کدام) است که این هیستوگرام از همسایه‌های 4×4 در اطراف نقطه کلیدی ایجاد می‌شود. بنابراین برای چنین توصیف‌گری یک بردار ویژگی ۱۲۸ تایی در نظر گرفته می‌شود [۱۴].

CNN ۲-۱-۲

شبکه‌های عصبی کانولوشن تا حد بسیار زیادی شبیه شبکه‌های عصبی مصنوعی می‌باشند. این نوع شبکه‌ها متشکل از نورون‌هایی با وزن‌ها و بایاس‌های قابل یادگیری (تنظیم) هستند. هر نورون تعدادی ورودی دریافت کرده و سپس حاصل ضرب وزن‌ها در ورودی‌ها را محاسبه می‌کند و در انتها با استفاده از یک تابع تبدیل (فعال‌سازی) غیر خطی نتیجه‌ای را ارائه می‌دهد. کل شبکه یک تابع امتیاز^۲ مشتق‌پذیر^۳ را ارائه می‌کند که در یک طرف آن پیکسل‌های خام تصویر ورودی و در طرف دیگر آن امتیازات مربوط به هر دسته قرار دارد.

۲-۲ کاهش بعد با تجزیه نامنفی ماتریس

همان طور که گفته شد برای حل مشکل شکاف معنایی در سیستم‌های شرح‌نگاری از روش NMF استفاده می‌شود. به این صورت که تصاویر را به فضای جدید برده و در این فضا شباهت بین تصاویر با استفاده از ضرایب ماتریس NMF محاسبه می‌شود. ایده اصلی، اتصال ویژگی‌های بصری و متنی به‌عنوان دو روش متفاوت برای پیدا کردن فضایی جدید با بعد پایین است که بتواند تصاویر را بهتر تفسیر کند.

NMF ۱-۲-۲

NMF به‌صورت زیر نوشته می‌شود. ماتریس ورودی A با بعد $m \times n$ داده شده که همه عناصر آن نامنفی است. هدف، پیدا کردن دو عامل $H \in R^{k \times n}$ و $W \in R^{m \times k}$ با عناصر نامنفی است

$$A \approx WH \quad (۴)$$

4. Frobenius Norm
5. Kullback-Leibler Divergence
6. Bregman Divergence
7. NP-Hard
8. Hierarchical Alternating Least Squares
9. Active Set
10. Nonnegativity-Constrained Least Squares

1. Scale-Invariant Feature Transform
2. Score Function
3. Differentiable

جواب مسأله $\min_{x \geq 0} \|Cx - b\|_1$ با جواب مسأله (۱۰) یکسان است و بنابراین شرایط KKT به صورت زیر است

$$y = C^T Cx - C^T b \quad (۱-۱۵)$$

$$x \geq 0 \quad (۲-۱۵)$$

$$y \geq 0 \quad (۳-۱۵)$$

$$x_i y_i = 0, \quad i = 1, \dots, q \quad (۴-۱۵)$$

اکنون مسأله آن است که چگونه عمل کنیم تا x و y یابیم که در این شرایط صدق کند. نکته‌ای که مطرح می‌شود این است که هر x و y که در شرایط (۱۵) صدق کند جواب مسأله است و هر جواب مسأله در این شرایط صدق می‌کند. الگوریتم چرخش بلاک اصلی بر اساس کنترل شرط $x_i y_i = 0, i = 1, \dots, q$ است.

مجموعه $\{1, \dots, q\}$ را به دو زیرمجموعه F و G تقسیم می‌کنیم به شکلی که $F \cap G = \emptyset$ و $F \cup G = \{1, \dots, q\}$ باشد. x_F, x_G, y_F و y_G را زیرمجموعه‌هایی از متغیرهای متناظر با اندیس‌هایشان در نظر می‌گیریم. در ابتدا $x_G = 0$ و $y_F = 0$ می‌باشد و در نتیجه شرط $x_i y_i = 0$ برقرار است. اکنون بر اساس (۱-۱۵)، x_F و y_G را به دست می‌آوریم

$$\begin{bmatrix} y_F \\ y_G \end{bmatrix} = C^T C \begin{bmatrix} x_F \\ x_G \end{bmatrix} - C^T b \quad (۱۶)$$

فرض می‌کنیم $M = C^T C$

$$\begin{bmatrix} y_F \\ y_G \end{bmatrix} = M \begin{bmatrix} x_F \\ x_G \end{bmatrix} - C^T b = \begin{bmatrix} M_{FF} & M_{FG} \\ M_{GF} & M_{GG} \end{bmatrix} \begin{bmatrix} x_F \\ x_G \end{bmatrix} -$$

$$C^T b = \begin{bmatrix} M_{FF} x_F + M_{FG} x_G \\ M_{GF} x_F + M_{GG} x_G \end{bmatrix} - C^T b = \quad (۱۷)$$

$$\begin{bmatrix} M_{FF} x_F \\ M_{GF} x_F \end{bmatrix} - C^T b$$

در نهایت رابطه زیر به دست می‌آید

$$y_G = M_{GF} x_F - (C^T b)_G = C_G^T C_F x_F - C_G^T b \quad (۱-۱۸)$$

و از خود رابطه اصلی می‌نویسیم

$$x_F = \arg \min_{x_F} \|C_F x_F - b\|_1 \quad (۲-۱۸)$$

که $x_G = 0$ و G اندیس‌هایی از x است که برای آن اندیس $x_i = 0$ می‌باشد. F و $y_F = 0$ اندیس‌هایی از y است که برای آن اندیس $y_i = 0$ می‌باشد.

می‌دانیم که از بین x_i و y_i حداقل یکی باید صفر باشد. چون $x_G = 0$ است اگر عضو جدیدی به آن اضافه شود، مثل این است که آن را برابر صفر گذاشته‌ایم. به همین ترتیب برای y_F هم برقرار است.

در ابتدا باید (۲-۱۸) را حل کرد و سپس جواب را در (۱-۱۸) قرار داد. مجموعه اندیس‌هایی که امکان‌پذیر نیستند را به صورت زیر

تعریف می‌کنیم

$$V = \{i \in F : x_i < 0\} \cup \{i \in G : y_i < 0\} \quad (۱۹)$$

سپس زیرمجموعه غیر تهی $\hat{V} \subset V$ را انتخاب کرده و F و G را به صورت زیر به‌روزرسانی می‌کنیم

اساس طراحی الگوریتم NMF مبتنی بر چرخش ANLS، روش خاصی برای حل (۶) باید ارائه می‌شود. الگوریتم قدیمی برای مسأله NNLS روش مجموعه فعال است. محدودیت بزرگ روش مجموعه فعال این است که فقط یک متغیر بین مجموعه‌های فعال در هر تکرار تغییر می‌کند و این وقتی ناشناخته‌ها بزرگ باشند، باعث کند شدن الگوریتم می‌شود. برای حل محدودیت روش مجموعه فعال، روش چرخش بلاک اصلی معرفی شده است.

ابتدا کمی به عقب‌تر برمی‌گردیم. در زیرمسأله‌های (۶)، حل هم‌زمان H و W کار بسیار مشکلی است و بنابراین بر اساس رویکرد BCD^۱ عمل می‌کنیم. در BCD یک گروه از متغیرها را ثابت نگه داشته و بر اساس یک متغیر بهینه‌سازی را انجام می‌دهیم

$$\min_x f(x_1, x_2, \dots, x_i, x_{i+1}, \dots, x_m) \quad (۷)$$

که تبدیل می‌شود به

$$\min_{x_i} f(x_1^k, x_2^k, \dots, x_i, x_{i+1}^{k-1}, \dots, x_m^{k-1}) \quad (۸)$$

$$\min_{x_i} f(x_1^k, x_2^k, \dots, x_i, x_{i+1}^k, x_{i+2}^{k-1}, \dots, x_m^{k-1}) \quad (۹)$$

یک حالت خاص BCD، رویکرد حل مسایل NMF است. در (۶) مشکل این است که یک W نداریم و هر W چندین W دیگر تولید می‌کند. در هنگام مقداردهی اولیه W ، در هر مرتبه نتایج متفاوتی ایجاد می‌شود، برای حل این مشکل ستون‌های W را برابر واحد قرار می‌دهیم و به عبارت دیگر آن را نرمال می‌کنیم (یعنی جمع عناصر هر ستون W برابر صفر باشد). راه‌حلی که برای این مسأله بیان می‌شود روش چرخش بلاک اصلی است که در ادامه توضیح خواهیم داد.

۲-۲-۳ روش چرخش بلاک اصلی

این روش همانند روش مبتنی بر مجموعه فعال است ولی در هر تکرار بیش از یک متغیر بین مجموعه فعال و غیر فعال جابه‌جا می‌شود. یعنی در هر ستون W و H در هر تکرار، اگر بیش از یک متغیر قابل جابه‌جایی بین دو مجموعه فعال و غیرفعال باشد این کار را انجام می‌دهد. این مسأله باعث کاهش زمان حل آن می‌شود.

ابتدا مسأله ساده‌تر زیر را در نظر می‌گیریم

$$\min_{x \geq 0} \|Cx - b\|_1 \quad (۱۰)$$

که در اینجا $x \in \mathbb{R}^{q \times 1}$ و $b \in \mathbb{R}^{p \times 1}$ ، $C \in \mathbb{R}^{p \times q}$ است. مسأله‌های (۶) قابل تبدیل به تعداد زیادی مسأله به شکل بالا هستند.

برای حل مسأله، $\min_{x \geq 0} \|Cx - b\|_1$ را در نظر می‌گیریم و تابع لاگرانژ را می‌نویسیم

$$\mathcal{L}(x, y) = \frac{1}{2} \|Cx - b\|_1 + y^T (-x) \quad (۱۱)$$

حال شرایط KKT را برای این مسأله می‌نویسیم

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial}{\partial x} \left(\frac{1}{2} \|Cx - b\|_1 + y^T (-x) \right) = 0 \quad (۱۲)$$

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{1}{2} (C^T Cx - C^T b) - y = 0 \quad (۱۳)$$

$$y = C^T Cx - C^T b \quad (۱۴)$$

ترکیب قانون تبادل کامل و قانون پشتیبان‌گیری روش چرخش بلاک اصلی برای مسأله NNLS با یک بردار سمت راست تک^۴ است که در الگوریتم ۱ خلاصه شده است. در این الگوریتم متغیر α به‌عنوان بافری است که تعداد دفعاتی که قانون تبادل کامل انجام می‌شود را می‌شمرد. وقتی $\alpha = 0$ باشد آن گاه قانون پشتیبان‌گیری اجرا می‌شود. حال اگر مسأله زیر را در نظر بگیریم

$$\min_{x \geq 0} \|CX - B\|_F^2 \quad (23)$$

که در اینجا $C \in \mathbb{R}^{p \times q}$ ، $B \in \mathbb{R}^{p \times r}$ و $X \in \mathbb{R}^{q \times r}$ است. الگوریتم ۲ را به سادگی برای هر بردار سمت راست $B = [b_1, \dots, b_r]$ اجرا می‌کنیم

$$\min_{x \geq 0} \left\| \begin{bmatrix} C_F & C_G \end{bmatrix} \begin{bmatrix} x_F \\ x_G \end{bmatrix} - b \right\|_F^2 \quad (24)$$

اگر $x_G = 0$ آن گاه

$$\min_{x_F \geq 0} \|C_F x_F - b\|_F^2 \quad (25)$$

برای حل ابتدا مشتق می‌گیریم

$$\frac{\partial}{\partial x_F} (\|C_F x_F - b\|_F^2) = 0 \quad (26)$$

$$\frac{\partial}{\partial x_F} (C_F^T x_F - b)^T (C_F^T x_F - b) = 0 \quad (27)$$

$$\frac{\partial}{\partial x_F} (x_F^T C_F^T C_F x_F - x_F^T C_F^T b - b^T C_F x_F + b^T b) = 0$$

$$C_F^T C_F x_F = C_F^T b \quad (28)$$

و (۱۵-۱) را بازنویسی می‌کنیم

$$y_G = C_G^T C_F x_F - C_G^T b \quad (29)$$

الگوریتم ۲، الگوریتم پیشرفته‌تری از الگوریتم ۱ را نشان می‌دهد. یکی از مزایایی که الگوریتم ۲ نسبت به الگوریتم ۱ دارد این است که با توجه به مشکلات NNLS ناشی از NMF، ماتریس C به طور معمول بسیار طولانی و باریک و محاسبه $C_F^T C_F$ ، $C_F^T b$ ، $C_G^T C_F$ و $C_G^T b$ بسیار هزینه‌بر است. به این منظور در الگوریتم ۲، $C^T b$ و $C^T C$ در ابتدا محاسبه و سپس آنها در تکرارهای بعدی دوباره استفاده می‌شوند. بنابراین چون اندازه ستون C کوچک است، حافظه مورد نیاز برای $C^T b$ و $C^T C$ کاهش می‌یابد و همچنین ماتریس X به طور معمول مسطح و وسیع است. در الگوریتم ۲ ایده مورد نظر این است که گروه‌هایی از ستون‌هایی که مجموعه یکسان، F و G و (۲۸) را حل می‌کند پیدا کنیم. با انجام این کار از انجام محاسبات تکراری جلوگیری می‌شود.

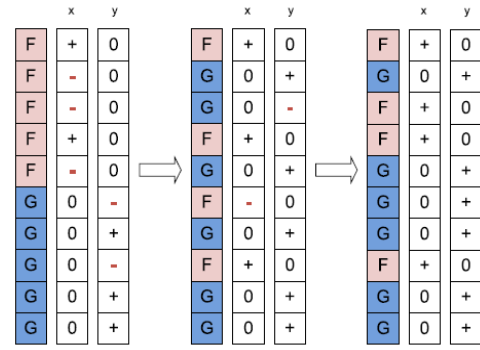
الگوریتم ۲ روش چرخش بلاک اصلی برای مسأله NNLS با بردارهای سمت راست چندگانه

ورودی: $B \in \mathbb{R}^{p \times r}$ و $C \in \mathbb{R}^{p \times q}$

خروجی: $X \in \mathbb{R}^{q \times r} = \arg \min_{x \geq 0} \|CX - B\|_F^2$

(۱) محاسبه $C^T C$ و $C^T B$

(۲) مقداردهی اولیه



شکل ۳: به روز رسانی F و G .

$$F = (F - \hat{V}) \cup (\hat{V} \cap G) \quad (20)$$

$$G = (G - \hat{V}) \cup (\hat{V} \cap F) \quad (21)$$

در شکل ۳ مثالی برای به‌روزرسانی F و G داریم.

x ها و y هایی که منفی هستند مقدارشان را صفر می‌کنیم و

منفی‌های x را به G و منفی‌های y را به F می‌دهیم.

نهایتاً الگوریتم چرخش بلاک اصلی به صورت زیر خلاصه می‌شود:

الگوریتم ۱) روش چرخش بلاک اصلی برای مسأله NNLS با یک بردار سمت راست تک

ورودی: $b \in \mathbb{R}^{p \times 1}$ و $C \in \mathbb{R}^{p \times q}$

خروجی: $x \in \mathbb{R}^{q \times 1} = \arg \min_{x \geq 0} \|Cx - b\|_F^2$

(۱) مقداردهی اولیه

$$F = \emptyset, G = \{1, \dots, q\}, x = 0, y = -C^T b, \alpha = 3, \beta = q + 1$$

(۲) محاسبه x_F و y_G طبق (۱۸)

(۳) تا زمانی که (x_F, y_G) نشدنی است ($x_F < 0, y_G < 0$) گام‌های زیر را تکرار کن:

(۴) محاسبه V طبق (۱۹)

(۵) اگر $|V| < \beta$ آن گاه $\hat{V} = V, \alpha = 3$

(۶) اگر $|V| \geq \beta$ و $\alpha \geq 1$ آن گاه $\hat{V} = V, \alpha = \alpha - 1$

(۷) اگر $|V| \geq \beta$ و $\alpha = 0$ آن گاه تنظیم \hat{V} با (۲۲)

(۸) به‌روزرسانی F و G با (۲۰) و (۲۱)

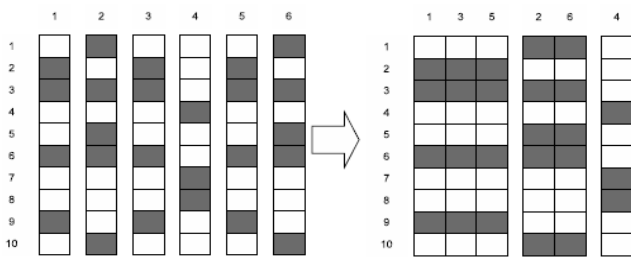
(۹) به‌روزرسانی x_F و y_G با (۱۸)

$|\hat{V}|$ نشان‌دهنده این است که چند متغیر در هر تکرار جابه‌جا می‌شود. اگر $|\hat{V}| > 1$ الگوریتم چرخش بلاک اصلی و اگر $|\hat{V}| = 1$ الگوریتم چرخش بلاک اصلی^۱ نامیده می‌شود. به منظور سرعت‌بخشیدن به جستجو معمولاً $\hat{V} = V$ را قرار می‌دهند که قانون تبادل کامل^۲ نام دارد. مشکل این روش این است که اگر همه V را در نظر بگیریم ممکن است داخل یک حلقه گیر کنیم، بنابراین به خاطر این که تکرار رخ ندهد از روشی دیگر به نام قانون پشتیبان‌گیری^۳ استفاده می‌کنیم که از فرمول زیر به‌دست می‌آید

$$\hat{V} = \{i : i = \max\{i \in V\}\} \quad (22)$$

پس به خاطر این که تکرار رخ ندهد اول سراغ آنها می‌رویم که اندیس‌های بزرگ‌تری دارند.

1. Single Principal Pivoting
2. Full Exchange Rule
3. Backup Rule



شکل ۴: گروه‌بندی بردارهای سمت راست.

که d_i فاصله نزدیک‌ترین همسایه i ام تا تصویر آزمون و eta تنظیم میزان پرت بودن داده‌ها است که در نهایت w_i وزن نزدیک‌ترین همسایه در نظر گرفته می‌شود.

۳- ارزیابی سیستم پیشنهادی

در این بخش به بررسی کارایی سیستم پیشنهادی بر اساس دو معیار ارزیابی آماری، دقت^۱، صحت^۲ و پارامتر ترکیبی F_1 می‌پردازیم. دقت عبارت است از نسبت مقادیر مثبت حقیقی^۳ (TP) به کل نمونه‌هایی که توسط مدل به صورت مثبت پیش‌بینی شده که به صورت زیر تعریف می‌شود (نسبت تعداد برچسب‌هایی که سیستم به درستی تشخیص داده است به برچسب‌های پیش‌بینی شده توسط سیستم)

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (32)$$

صحت عبارت است از نسبت مقادیر مثبت حقیقی (TP) به کل نمونه‌هایی که در پایگاه داده به صورت مثبت تعریف شده است. رابطه صحت به صورت زیر تعریف می‌شود (نسبت تعداد برچسب‌هایی که سیستم به درستی تشخیص داده است به برچسب‌های واقعی)

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (33)$$

که به جای این دو معیار می‌توان از یک معیار ترکیبی برای ارزیابی کارایی ارزیابی به نام F_1 استفاده نمود که دقت آشکارسازی و سطح عملکرد الگوریتم طراحی شده را نشان می‌دهد. این پارامتر مطابق رابطه زیر تعریف می‌شود

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (34)$$

در این روابط TP ^۴، FP ^۴ و FN ^۵ به ترتیب مثبت حقیقی، مثبت کاذب و منفی کاذب را نشان می‌دهند. سیستم مورد استفاده در کلیه آزمایش‌ها، Intel Core 2 Duo 2.0 GHz با ۴ گیگابایت حافظه است. روش پیشنهادی در محیط برنامه نویسی Matlab پیاده‌سازی شده و نسخه مورد استفاده برای پیاده‌سازی (R2016b) Matlab ۸٫۶ می‌باشد.

۳-۱ معرفی مجموعه داده‌های استفاده شده

برای ارزیابی سیستم پیشنهادی از دو مجموعه داده بهره برداریم که دسته اول مجموعه داده آماده و در دسترس CorelδK است. زیرا این

$$F_j = \emptyset, G_j = \{1, \dots, q\} \text{ for } j \in \{1, \dots, r\}, X = 0, y = -C^T B,$$

$$\alpha (\in \mathbb{R}^r) = \alpha, \beta (\in \mathbb{R}^r) = q + 1$$

(۳) محاسبه Y_{G_j} و X_{F_j} برای $j \in \{1, \dots, r\}$

(۴) طبق (۱۸) با استفاده از گروه‌بندی ستون‌ها

(۵) تا زمانی که (X_{F_j}, Y_{G_j}) نشدنی است $x_F < 0, y_G < 0$

گام‌های زیر را تکرار کن:

(۶) پیدا کردن ستون‌هایی که راه‌حل را نشدنی کند

$$I = \{j : (X_{F_j}, Y_{G_j}) \text{ is infeasible}\}$$

(۷) محاسبه V_j برای $j \in I$ طبق (۱۹)

(۸) برای همه $j \in I$ اگر $|V_j| < \beta_j$ آن گاه $\alpha_j = \alpha_j - 1, \hat{V}_j = V_j$

(۹) برای همه $j \in I$ اگر $|V_j| \geq \beta_j$ و $\alpha_j \geq 1$ آن گاه $\alpha_j = \alpha_j - 1, \hat{V}_j = V_j$

(۱۰) برای همه $j \in I$ اگر $|V_j| \geq \beta_j$ و $\alpha_j = 0$ آن گاه تنظیم \hat{V}_j

با (۲۲)

(۱۱) به‌روزرسانی F_j و G_j برای همه $j \in I$ با (۲۰) و (۲۱)

(۱۲) به‌روزرسانی X_{F_j} و Y_{G_j} برای $j \in I$ با (۱۸) با استفاده از گروه‌بندی ستون‌ها

در شکل ۴ مثالی را برای گروه‌بندی ستون‌ها می‌بینیم که $q = 10$ و $r = 6$ است. خانه‌های سیاه متغیرهای با اندیس F را نشان می‌دهد که با (۲۸) محاسبه می‌شود. با گروه‌بندی ستون‌هایی که مجموعه مشترک F دارند (مثلاً ستون‌های $\{1, 3, 5\}, \{2, 6\}, \{4\}$) می‌توانیم از محاسبات اضافی در حل (۲۸) جلوگیری کنیم. در عمل برای اجرای تکراری الگوریتم نیاز به یک معیار توقف است. برای NMF معیار می‌تواند این باشد که تابع هدف به کمینه محلی رسیده است یا نه.

۳-۲ طبقه‌بند KNN برای بازیابی و تلفیق برچسب‌ها

راه‌حل بصری برای تصمیم‌گیری در مورد چگونگی طبقه‌بندی یک داده آزمون بدون برچسب این است که به داده‌های آموزشی نزدیک به آن توجه کنیم. این روش به نام طبقه‌بندی k -نزدیک‌ترین همسایه شناخته می‌شود. در روش پیشنهادی برای بازیابی و تلفیق برچسب‌ها از یک روش دومرحله‌ای استفاده می‌کنیم که در هر مرحله از الگوریتم KNN استفاده می‌شود. در اولین مرحله ابتدا تصویر آزمون برای جستجوی نزدیک‌ترین کلاس‌ها در مجموعه داده آموزش به الگوریتم KNN داده می‌شود، سپس در انتهای این مرحله خوشه‌های نزدیک به تصویر آزمون را خواهیم داشت. باید توجه داشت که در مورد هر مجموعه داده این روش متفاوت است به این صورت که برای مجموعه داده CorelδK تصویر آزمون می‌تواند با تمام داده‌ها مقایسه شود ولی در داده‌های واقعی آن را با مراکز خوشه مقایسه می‌کنیم. در مرحله بعدی برای خوشه‌هایی که از مرحله قبل انتخاب شدند، نزدیک‌ترین تصاویر را پیدا می‌کنیم. این مرحله نیز با الگوریتم KNN انجام می‌شود. حال برچسب‌های تصاویر انتخاب شده را باید به روشی تلفیق کنیم. برداری در این مرحله برای هر خوشه به دست می‌آید که این بردار نشان‌دهنده وزن برچسب‌های تصاویر نزدیک به تصویر آزمون است و نهایتاً با بردارهای به دست آمده از سایر خوشه‌ها ترکیب شده و برچسب‌هایی که بیشترین وزن را دارند به عنوان برچسب‌های تصویر در نظر گرفته می‌شوند. وزن‌دهی بر اساس الگوریتم KNN انجام می‌گردد. همان طور که در بالا خلاصه‌ای از این الگوریتم گفته شد، فاصله‌های به دست آمده را به روش زیر وزن‌دهی می‌کنیم

$$w_i = \exp(-eta \times D_i) \quad (30)$$

$$D_i = d_i - \min(d) \quad (31)$$

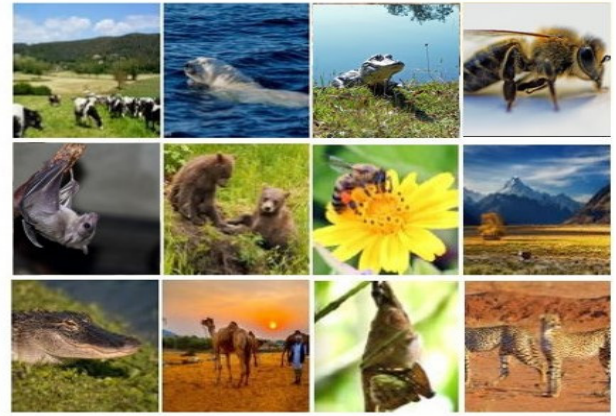
1. Precision

2. Recall

3. True Positive

4. False Positive

5. False Negative



شکل ۵: چند نمونه از تصاویر دادگان ۵۰۰ px و Corel5K به ترتیب در سمت راست و چپ.

الگوریتم زمانی حاصل می‌شود که بردارهای پایه برابر با ۳۰ باشند. همچنین با بالارفتن تعداد بردارهای پایه، زمان اجرای الگوریتم NMF که در فاز آموزش انجام می‌شود افزایش می‌یابد. در داده‌های واقعی با افزایش تعداد بردارهای پایه کارایی الگوریتم بهتر می‌شود و زمانی بیشترین کارایی را دارد که مقادیر Recall و Precision به هم نزدیک باشند. در اولین مرحله از اجرای الگوریتم KNN برای به دست آوردن خوشه‌های مشابه، در کل مجموعه داده به دنبال تصاویر نزدیک می‌گردیم تا نزدیک‌ترین خوشه به تصویر مورد نظر را بیابیم. به همین منظور تعداد همسایه‌ها به تعداد کل داده‌های آموزش است که در پایگاه داده Corel5K بین ۱ تا ۴۵۰۰ می‌تواند متغیر باشد. طبق آزمایشی که انجام دادیم بیشترین مقدار کارایی در تعداد همسایه ۳۰۰ است. داده‌های واقعی حیوانات به ۴ دسته طبقه‌بندی شده‌اند (تمساح، خفاش، خرس و زنبور عسل) و بنابراین تصویر آزمون را با داده‌های این کلاس‌ها مقایسه می‌کنیم. به این صورت که مراکز خوشه این کلاس‌ها را در نظر گرفته و با الگوریتم KNN نزدیک‌ترین خوشه را می‌یابیم. بنابراین بین ۱ تا ۴، تعداد همسایه‌هاست. در دومین مرحله از اجرای الگوریتم KNN در هر خوشه باید نزدیک‌ترین تصاویر را به تصویر آزمون پیدا کرد. چون تعداد داده‌های هر دسته در پایگاه داده Corel5K بین ۱ تا ۹۰ است، بنابراین بین ۱ تا ۹۰ متغیر KNN۲ را تغییر و این آزمایش را انجام دادیم. در مقدار ۲۰ بیشترین کارایی ملاحظه می‌شود. در داده‌های واقعی در هر دسته از این تصاویر تعداد ۱۵۰۰ تا ۳۰۰۰ تصویر وجود دارد و بنابراین تعداد همسایه‌ها را می‌توان بین این مقادیر قرار داد. با افزایش تعداد همسایه‌ها، کارایی الگوریتم بهتر شده است. پارامتر آخری که بررسی می‌کنیم تعداد برچسب‌ها است که در انتهای فاز آزمون مشخص می‌کنیم چه تعداد برچسب برای تصویر در نظر گرفته شود. در این آزمایش مشخص شده که با افزایش تعداد برچسب‌ها (با آزمایش روی هر دو پایگاه داده) کارایی الگوریتم در حال کاهش بوده و بهترین کارایی در تعداد ۵ برچسب است. نتایج این آزمایش‌ها در جداول ۱ و ۲ نشان داده شده است. همچنین در جداول ۳ و ۴ مثال‌هایی از شرح‌نگاری توسط روش پیشنهادی، برای تصاویر آزمون و تصاویر نویز پایگاه داده ۵۰۰px نشان داده شده‌است.

۲-۲-۳ مقایسه روش پیشنهادی با روش‌های موجود

در این بخش به مقایسه روش پیشنهادی با سایر روش‌های موجود در سال‌های اخیر می‌پردازیم. همان طور که در جدول ۵ مشاهده می‌شود الگوریتم پیشنهادی، کارایی به مراتب بهتری روی پایگاه داده Corel5K دارد. همچنین برای داده‌های واقعی نیز کارایی این روش آزمایش شد و به نتایج خوبی دست یافتیم که نتایج حاصل را در جدول ۶ ملاحظه می‌کنید.

پایگاه داده به عنوان یک استاندارد در بسیاری از سیستم‌های شرح‌نگاری تصاویر به کار می‌رود و همچنین برای ارزیابی بهتر سیستم پیشنهادی از مجموعه داده‌های واقعی برگرفته از سایت ۵۰۰ px استفاده نموده‌ایم. در شکل ۵، چند نمونه از تصاویر دادگان ۵۰۰px و Corel5K [۱۵] نشان داده شده است.

Corel5K: این مجموعه داده شامل ۵۰۰۰ عکس است که به صورت دستی با ۲۶۰ کلیدواژه متمایز برچسب‌گذاری شده‌اند. Corel5K به مجموعه‌های آموزش (۷۵٪) و آزمون (۲۵٪) تقسیم‌بندی می‌شود. مجموعه داده‌های آموزش و آزمون به ترتیب شامل ۴۵۰۰ و ۴۹۹ تصویر هستند. تصاویر به طور دستی با متوسط ۳/۴ برچسب از مجموعه ۲۶۰ برچسب شرح داده شده‌اند. همچنین داده‌ها یک برچسب گروهی نیز دارند که هر ۱۰۰ عکس دارای یک برچسب است، به عبارتی داده‌ها به ۵۰ دسته، طبقه‌بندی و ویژگی‌های این تصاویر با استفاده از روش SIFT استخراج شده‌اند.

داده‌های واقعی: داده‌های واقعی (برگرفته از سایت ۵۰۰ px) استفاده شده در این الگوریتم مشکلاتی را به همراه دارد. اول این که این داده‌ها دارای حجم بالایی هستند و دیگری این است که مانند داده‌های در دسترس و آماده، برچسب‌گذاری نشده‌اند و این کار را خودمان به طور دستی باید انجام دهیم. همچنین ویژگی‌های این تصاویر را با استفاده از روش CNN استخراج کرده‌ایم. این مجموعه شامل ۴ دسته عکس است (تمساح، خفاش، خرس و زنبور عسل) و بنابراین برچسب گروهی آنها شامل ۱ تا ۴ است. در زمان برچسب‌گذاری تصاویر آموزشی، دقت می‌کنیم که برچسب‌های تکراری حذف شوند. در هر دسته از تصاویر این پایگاه داده بین ۱۵۰۰ تا ۳۰۰۰ تصویر موجود است.

۲-۳ نتایج و آزمایش‌ها

برای نشان دادن کارایی الگوریتم پیشنهادی و بررسی رفتار آن در شرایط مختلف تعدادی آزمایش طراحی می‌نماییم. این آزمایش‌ها شامل بررسی تأثیر پارامترهای مختلف و ارزیابی کمی روش پیشنهادی در مقایسه با الگوریتم‌های موجود می‌باشد. در ادامه به تشریح هر یک از این آزمایش‌ها می‌پردازیم.

۱-۲-۳ تأثیر پارامترهای مختلف در روش پیشنهادی

الگوریتم پیشنهادی ۴ پارامتر مختلف دارد که در این آزمایش‌ها تأثیر مقدار این پارامترها را بیان می‌کنیم. همچنین این آزمایش‌ها را روی هر دو پایگاه داده بررسی خواهیم نمود. پارامتر اول تعداد بردارهای پایه الگوریتم NMF است (K-NMF). با انجام این آزمایش، بیشترین کارایی

جدول ۲: تأثیر پارامترهای مختلف بر روی بیشترین مقدار F1 و بیشترین مقدار صحت و دقت در پایگاه داده COREL5K.

	مقدار صحت	مقدار دقت	مقدار F1	تعداد بردارهای
	۵۰	۵۰	۵۰	تعداد پایه ۳۰
داده‌های	۵۰	دقت ۵۰	مقدار F1	تعداد تصاویر
Corel5K	۵۰	دقت ۵۰	مقدار F1	تعداد کلاس‌های
	۵۰	دقت ۵۰	مقدار F1	تعداد برچسب‌ها
	۵۰	دقت ۵۰	مقدار F1	۵

جدول ۱: تأثیر پارامترهای مختلف بر روی بیشترین مقدار F1 و بیشترین مقدار صحت و دقت در پایگاه داده PX ۵۰۰.

مقدار	مقدار	مقدار F1	تعداد تصاویر	مقدار
صحت ۶۵	دقت ۵۵	۶۰	همسایه ۱۵۰۰	داده‌های
مقدار	مقدار	مقدار F1	تعداد برچسب‌ها	واقعی
صحت ۶۵	دقت ۴۵	۵۴	۵	

جدول ۳: مثال‌هایی از شرح‌نگاری روش پیشنهادی و دستی برای تصاویر آزمون پایگاه داده PX ۵۰۰ (برای ۴ دسته از حیوانات: خرس، خفاش، زنبور عسل و تمساح). برچسب‌های برجسته‌انتهایی هستند که مطابقت دارند و برچسب‌های غیر برجسته مطابقت ندارند.



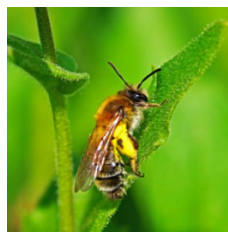
alligator, bee, flower, water, garden, insect, macro, nature, summer, yellow
summer, everglades, libelle, **alligator, white, crocodilian, light, lizard, natural, nature, reptile, studio, water**



animal, animals, bat, black, flying, nature, night, white, wild, wildlife
nature, night, vampire, animal, bat, **wildlife**

شرح‌نگاری سیستم

شرح‌نگاری انسان



bee, flower, flowers, garden, green, insect, macro, nature, summer, yellow
macro, nature, honey, **flowers, sigma, summer, bee**

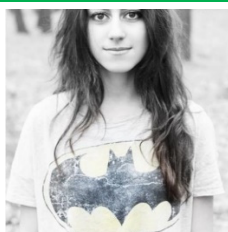


animal, animals, bear, brown, mammal, nature, wild, wildlife
alaska, animal, animals, bear, brown, fish, grizzly, **mammal, salmon, summer, water, white, wild, wildlife**

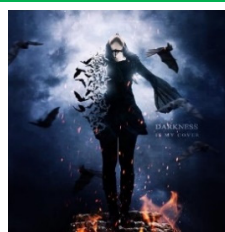
شرح‌نگاری سیستم

شرح‌نگاری انسان

جدول ۴: مثال‌هایی از شرح‌نگاری روش پیشنهادی و دستی برای تصاویر نویز از پایگاه داده PX ۵۰۰ برای دسته خفاش از حیوانات. برچسب‌های برجسته‌انتهایی هستند که مطابقت دارند و برچسب‌های غیر برجسته مطابقت ندارند.



animal, animals, baseball, bat, black, flying, girl, nature, white, wildlife
beautiful, beauty, **black, nature, black and white, girl, fashion, bat, batman, animal**



animal, bat, black, dark, girl, nature, night, white, wildlife
dark, model, creative, fear, cover, fire, darkness, **bat, photo-manipulation, animal, wildlife**

شرح‌نگاری سیستم

شرح‌نگاری انسان

می‌گردد. الگوریتم در مرحله آموزش با روش BPP-NMF ماتریس ضرایب و ماتریس پایه را به دست می‌آورد و با استفاده از این ماتریس‌ها جدول ۵: مقایسه روش پیشنهادی با سایر روش‌ها روی پایگاه داده COREL5K.

Method	Precision	Recall	F1
NMF-KNN [۵]	۳۸	۵۶	۴۵٫۲
۲PKNN+ML [۱۶]	۴۴	۴۶	۴۵

۴- بحث و نتیجه‌گیری

در این مقاله روشی برای شرح‌نگاری خودکار تصویر با استفاده از الگوریتم تجزیه نامنفی ماتریس (NMF) ارائه شد. در روش پیشنهادی برای مدل‌سازی از روشی به نام چرخش بلاک اصلی (BPP) برای حل NMF استفاده شده که این روش در دو مرحله آموزش و آزمون انجام

- [9] A. Eweawi, M. S. Cheema, and C. Bauckhage, "Discriminative joint non-negative matrix factorization for human action classification," in *Proc. German Conf. on Pattern Recognition, GPRC'13*, pp. 61-70, vol 8142. Springer, Berlin, 2013.
- [10] J. Liu, C. Wang, J. Gao, and J. Ha, "Multi-view clustering via joint nonnegative matrix factorization," in *Proc. of the SIAM Int. Conf. on Data Mining, SIAM*, pp. 252-260, Austin, TX, USA, 2-4 May 2013.
- [11] M. Chen, A. Zheng, and K. Weinberger, "Fast image tagging," in *Proc. 30th Int. Conf. on Machine Learning*, 9 pp., Atlanta, GA, USA, Jan. 2013.
- [12] R. Rad and M. Jamzad, "Automatic image annotation by a loosely joint non-negative matrix factorization," *IET Computer Vision*, vol. 9, no. 6, pp. 806-813, Dec. 2015.
- [13] R. Rad and M. Jamzad, "Image annotation using multi-view non-negative matrix factorization with different number of basis vectors," *J. of Visual Communication and Image Representation*, vol. 46, pp. 1-12, Jul. 2017.
- [14] M. T. Chao and Y. S. Chen, "Keyboard recognition from scale-invariant feature transform," in *Proc. IEEE Int. Conf. on Consumer Electronics-Taiwan, ICCE-TW'17*, pp. 205-206, Taipei, Taiwan, 12-14 Jun. 2017.
- [15] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth, "Object recognition as machine translation: learning a lexicon for a fixed image vocabulary," in *Proc. European Conf. on Computer Vision*, vol. 4, pp. 97-112, May 28-31 2002.
- [16] Y. Verma and C. Jawahar, "Image annotation using metric learning in semantic neighbourhoods," in *Proc. European Conf. on Computer Vision, ECCV'12*, vol. 3, pp. 836-849, Florence, Italy, 7-13 Oct. 2012.
- [17] S. Moran and V. Lavrenko, "A sparse kernel relevance model for automatic image annotation," *International J. of Multimedia Information Retrieval*, vol. 3, no. 4, pp. 209-229, Nov. 2014.
- [18] Y. Yang, W. Zhang, and Y. Xie, "Image automatic annotation via multi-view deep representation," *J. of Visual Communication and Image Representation*, vol. 53, pp. 368-377, Nov. 2015.
- [19] V. N. Murthy, E. F. Can, and R. Manmatha, "A hybrid model for automatic image annotation," in *Proc. of Int. Conf. on Multimedia Retrieval, ICMR'14*, pp. 369-376, Glasgow, United Kingdom, 1-4 Apr. 2014.
- [20] Z. Lu and Y. Peng, "Image annotation by semantic sparse recoding of visual content," in *Proc. of the 20th ACM Int. Conf. on Multimedia*, pp. 499-508, Nara, Japan, 29 Oct.-2-Nov. 2012.
- [21] Y. Xiang, X. Zhou, T. -S. Chua, and C. -W. Ngo, "A revisit of generative model for automatic image annotation using Markov random fields," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'09*, pp. 1153-1160, Miami, FL, USA, 20-25 Jun. 2009.
- [22] A. Makadia, V. Pavlovic, and S. Kumar, "A new baseline for image annotation," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'08*, vol. 3, pp. 316-329, Marseille, France, 12-18 Oct. 2008.

حازم الرکابی در سال ۱۳۸۵ مدرک کارشناسی علوم کامپیوتر خود را از دانشگاه مستنصریه در بغداد. و در سال ۱۳۹۶ مدرک کارشناسی ارشد مهندسی کامپیوتر از دانشگاه فردوسی در مشهد دریافت نمود. از سال ۱۳۸۷ در وزارت آموزش و پرورش در عراق در شهر (نجف اشرف) مشغول به فعالیت گردید و اینک نیز عضو هیأت علمی این وزارت می‌باشد. زمینه‌های علمی مورد علاقه نام‌برده متنوع بوده و شامل موضوعاتی مانند ایده‌های نو در معماری کامپیوتر، مدیریت شبکه‌های کامپیوتری، سیستم‌های توزیعی و رباتیک می‌باشد.

نسرین صوفی کارشناسی خود را در سال ۱۳۸۷ از دانشگاه پیام نور در رشته مهندسی کامپیوتر گرایش نرم افزار اخذ نمود و تحصیلات خود را در مقطع کارشناسی ارشد در رشته مهندسی کامپیوتر گرایش هوش مصنوعی از دانشگاه فردوسی مشهد در سال ۱۳۹۸ اخذ نمود. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: شبکه‌های عصبی عمیق و فیلتر کالمن.

هادی صدوقی یزدی هم‌اکنون استاد گروه مهندسی کامپیوتر در دانشگاه فردوسی مشهد می‌باشد. ایشان در سال ۱۳۷۳ مدرک کارشناسی مهندسی برق خود را از دانشگاه فردوسی مشهد و کارشناسی ارشد و دکتری مهندسی برق را به ترتیب در سال‌های ۱۳۷۵ و ۱۳۸۴ از دانشگاه تربیت مدرس دریافت نمود. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش سیگنال و شناسایی الگو.

SKL-CRM [۱۷]	۳۹	۴۶	۴۲
MVSAE [۱۸]	۳۷	۴۷	۴۲
SVM-DMBRM [۱۹]	۳۶	۴۸	۴۱
۲PKNN [۱۶]	۳۹	۴۰	۳۹٫۵
LJNMF [۱۲]	۳۵	۴۳	۳۹٫۱
FastTag [۱۱]	۳۲	۴۳	۳۶٫۷
TagProb [۷]	۳۳	۴۲	۳۶٫۹
SSRC [۲۰]	۳۱٫۵	۳۷	۳۴٫۱
MRFA [۲۱]	۳۱	۳۶	۳۳٫۳
JEC [۲۲]	۲۷	۳۲	۲۹٫۳
MvNMF-SK [۱۳]	۴۰٫۳	۴۵٫۸	۴۲٫۹
MvNMF-DK [۱۳]	۴۴	۴۷٫۵	۴۵٫۶
BPP-NMF	۴۹٫۸۸	۵۰٫۵۳	۵۰٫۲۰

جدول ۶: نتایج اجرای روش پیشنهادی روی داده‌های واقعی حیوانات.

Method	Precision	Recall	F1
BPP-NMF	۵۶٫۵۶	۷۰٫۸۲	۶۲٫۸۹

در مرحله آزمون، برای تصویر آزمون داده شده، برچسب‌ها را پیش‌بینی می‌کند. مرحله آزمون روش پیشنهادی شامل دو مرحله استفاده از الگوریتم k -نزدیک‌ترین همسایه (KNN) برای بازیابی و تلفیق برچسب‌ها است. ابتدا KNN برای به دست آوردن کلاس‌های مشابه می‌شود. برای ارزیابی و آزمایش روش پیشنهادی از دو پایگاه داده Corel5K و داده‌های واقعی حیوانات استفاده شد و نهایتاً برای پایگاه داده Corel5K که با روش‌های موجود در سال‌های اخیر مقایسه گردید، به میزان دقت ۵۰٫۲۰ رسیدیم که به طور قابل ملاحظه‌ای دقت الگوریتم افزایش یافته است و برای داده‌های واقعی نیز به دقت مناسب ۶۲٫۸۹ دست یافتیم.

مراجع

- [1] J. Kim and H. Park, "Fast nonnegative matrix factorization: an active-set-like method and comparisons," *SIAM J. on Scientific Computing*, vol. 33, no. 6, pp. 3261-3281, 2011.
- [2] C. F. Tsai and C. Hung, "Automatically annotating images with keywords: a review of image annotation systems," *Recent Patents on Computer Science*, vol. 1, no. 1, pp. 55-68, Jan. 2008.
- [3] J. Li and J. Z. Wang, "Real-time computerized annotation of pictures," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 985-1002, Jun. 2008.
- [4] D. Putthividhy, H. T. Attias, and S. S. Nagarajan, "Topic regression multi-modal latent Dirichlet allocation for image annotation," in *Proc IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'10*, pp. 3408-3415, San Francisco, CA, USA, 13-18 Jun. 2010.
- [5] M. M. Kalayeh, H. Idrees, and M. Shah, "NMF-KNN: image annotation using weighted multi-view non-negative matrix factorization," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'14*, pp. 184-191, Columbus, OH, USA, 23-28 Jun. 2014.
- [6] Z. Li, et al., "Learning semantic concepts from image database with hybrid generative/discriminative approach," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 9, pp. 2143-2152, Oct. 2013.
- [7] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Tagprop: discriminative metric learning in nearest neighbor models for image auto-annotation," in *Proc. IEEE 12th Int. Conf. on Computer Vision*, pp. 309-316, Kyoto, Japan, 29 Sept.-2 Oct. 2009.
- [8] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788-791, 21 Oct. 1999.

امیر حسین طاهری نیا کارشناسی خود را در رشته مهندسی کامپیوتر در سال ۱۳۸۳ از دانشگاه فردوسی مشهد اخذ نمود و کارشناسی ارشد و دکتری خود را در رشته مهندسی کامپیوتر به ترتیب در سال‌های ۱۳۸۵ و ۱۳۹۱ از دانشگاه صنعتی شریف اخذ نمود. وی از سال ۱۳۹۲ استادیار دانشکده فنی و مهندسی دانشگاه فردوسی مشهد است. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: امنیت چندرسانه‌ای، پنهان‌سازی اطلاعات، پردازش سیگنال چندرسانه‌ای.