مجله ماشین بینایی و پردازش تصویر
دوره دوم، شماره یک، بهار ۱۳۹۴

حاشیه‌نویسی خودکار تصاویر مبتنی بر خوشه‌بندی دوسطحی بصری و معنایی

سمانه بهرامی[۱] و محمد صنیعی آباده[۲]

چکیده

حاشیه‌نویسی خودکار تصاویر به ایجاد خودکار برچسب‌های متنی مطابق با محتوای بصری تصاویر دلالت دارد. اگرچه در دهه گذشته تحقیقات فراوانی در این زمینه انجام گرفته است اما وجود برچسب‌های متعدد و وجود شکاف معنایی میان این برچسب‌ها و ویژگی‌های سطح پایین بصری باعث کاهش دقت و کارایی این سامانه‌ها شده است. در این پژوهش یک روش حاشیه‌نویسی با استفاده از خوشه‌بندیدوسطحی بر مبنای ویژگی‌های کاهش یافته با الگوریتم وراثتی و نیز معانی پیشنهاد شده است.خوشه‌بندی باعث می‌شود تصاویر مشابه به هم از لحاظ بصری و نیز تصاویر مرتبط به هم از جهت معنایی در کنار هم قرار گرفته و حاشیه‌نویسی شوند. این کار علاوه بر تسریعحاشیه‌نویسی، کارایی قابل قبولی برای یک سامانه حاشیه‌نویسی نیز داشته است.برای ارزیابی روش، دو دادگان شناخته شده Corel5k و IAPR TC-12 انتخاب شده‌اند. نتایج به دست آمده روی این دو دادگان عملکرد قابل قبول روش پیشنهادی را در مقایسه با دیگر روش‌ها نشان می‌دهد.

کلید واژه‌ها

حاشیه‌نویسی خودکار تصاویر، خوشه‌بندیدوسطحی، الگوریتم وراثتی

۱ مقدمه

با رشد گسترده تکنولوژی‌های مرتبط با اینترنت، تصاویر و ویدئوها به سرعت در حال توسعه بر روی بستر اینترت هستند. چگونگی ساماندهی و مدیریت این اطلاعات حجیم، بحث داغی شده است که نیاز به راه حل‌های فوری دارد. بازیابی اطلاعات چندرسانه ای، گامی مؤثر جهت حل مشکل بیان شده و حاشیه-نویسی خودکار تصاویر گامی مهم و کلیدی در بازیابی اطلاعات چندرسانه‌ای می‌باشد.

حاشیه‌نویسی خودکار تصاویر فرایند انتساب خودکار معانی به تصاویر است. برای برچسب زنی به تصاویر بدون برچسب از ویژگی‌های تصاویر آموزش شامل ویژگی‌های سطح پایین و برچسب‌هایی که بیانگر معانی تصاویر هستند استفاده می‌شود. بنابراین تنها توجه به ویژگی‌های مهم بصری و تصاویر آموزشی شاخص می‌تواند به برچسب‌گذاری دقیق تصاویر کند.

مشکل شناخته‌شده‌ی شکاف معنایی، حاشیه‌نویسی تصاویر را به کار سختی تبدیل کرده است. چرا که استخراج معانی سطح بالا از تصاویر تنها با کمک ویژگی‌های سطح پایین چون رنگ و بافت مشکل است. یکی از چالش‌های حاشیه‌نویسی خودکار تصاویر، برچسب‌های ضعیف تصاویر مجموعه‌ی آموزش است. منظور از برچسب ضعیف این است که تناظر بین کلمات برچسب خورده به تصویر و نواحی تصویر مشخص نیست. برای هر تصویر تنها با کلماتی روبرو هستیم که به کل تصویر برچسب خورده‌اند و اینکه هر کلمه اشاره به کدام ناحیه از تصویر دارد نامعلوم است [۱]. از طرفی اعمال توصیفگرهای متعدد محلی و سراسری به تصاویر به دلیل اهمیت محتوای بصری تصاویر در ایجاد تمایز میان تصاویر و توجه به تمام ویژگی‌های بصری تصویری باعث ایجاد چالش ابعاد بالای ویژگی‌ها می‌شود.

اکثر تکنیک‌های خودکار حاشیه‌نویسی که تا کنون انجام گرفته‌اند مدلی را جهت یافتن ارتباط میان ویژگی‌های بصری و کلمات کلیدی پیشنهاد داده‌اند. در این پژوهش یک روش حاشیه-

این مقاله در مهرماه ۱۳۹۳ دریافت شد. در آبانماه بازنگری و در اردیبهشت‌ماه ۱۳۹۴ پذیرفته شد.

[۱] دانش آموخته کارشناسی ارشد دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، رایانامه: samaneh.bahrami@modares.ac.ir

[۲] استادیار،دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس
رایانامه: saniee@modares.ac.ir

نویسیجدید با استفاده از خوشه‌بندی دوسطحی بر مبنای ویژگی‌های کاهش یافته با استفاده از الگوریتم وراثتی و نیز معانی پیشنهاد شده است. نوآوری اصلی بیان شده در این پژوهش، انجام خوشه‌بندی به روشی جدید با در نظر گرفتن ارتباط میان ویژگی‌ها و معانی و سپس تفکیک تصاویر از هم با توجه به این ارتباط می‌باشد. خوشه‌بندی پیشنهاد شده، خوشه‌بندیدوسطحی بر مبنای هم محتوای بصری و هم‌معنا می‌باشد. ازآنجا که جهت حاشیه-نویسی رویکرد نزدیکترین همسایه به کار گرفته شده است، این نوع خوشه‌بندی باعث می‌شود که جهت یافتن همسایه‌های هر تصویر جدید تنها از تصاویر مرتبط با آن تصویر که در یک خوشه قرار گرفته‌اند استفاده شوند. علاوه بر این،خوشه‌بندی پیشنهادی به افزایش دقت و سرعت اجرای حاشیه‌نویسی کمک می‌کند.از طرفی از آنجا که چندین توصیفگر ویژگی محلی و سراسری جهت توصیف تصاویر استفاده شده‌اند، الگوریتم وراثتی جدیدی جهت انتخاب زیرمجموعه ویژگی‌های بهینه پیشنهادشده‌است که برخلاف سایر روش‌های انتخاب ویژگی موجود در حوزه حاشیه‌نویسی، قابلیت انتخاب ویژگی‌ها به صورتی کارآمد در حالت چندبرچسبی را دارد.

ادامه این مقاله به صورت زیر ساماندهی شده است. در فصل ۲ مروری بر پژوهش‌های پیشین در زمینه حاشیه‌نویسی تصاویر انجام می‌شود. روش پیشنهادی در فصل ۳ مورد بررسی قرار می‌گیرد. در فصل ۴ نتایج ارزیابی ارائه خواهند شد و بالاخره در فصل ۵ نتیجه‌گیری از پژوهش انجام شده بیان خواهد شد.

۲ پژوهش‌های پیشین

در دهه گذشته تحقیقات فراوانی در حوزه حاشیه‌نویسی تصاویر انجام گرفته است که می‌توان آن‌ها را به سه مدل گروه‌بندی کرد[۲]: مدل‌های احتمالاتی، مدل‌های مبتنی بر دسته‌بندی و مدل-های مبتنی بر نزدیکترین همسایه. اکثر مدل‌های احتمالاتی[۶-۳]احتمالات توأمی را روی محتوای تصویری و کلمات کلیدی جهت یافتن نگاشت میان ویژگی‌های تصویری و کلمات کلیدی تخمین می‌زنند. هدف این مدل‌ها، ساخت مدلی واحد برای تمامی کلمات کلیدی است که باعث ایجاد مدل‌سازی بهتری از نظر وابستگی‌های کلمات کلیدی به یکدیگر می‌باشند. مدل ترجمه[۱][۳] توسط دویگلو[۲] و همکاران ارائه شد. در این مدل تصاویر توسط الگوریتم قطعه‌بندی[۳] به نواحی تقسیم می‌شوند. سپس از هر ناحیه بردار ویژگی استخراج می‌شود. بردارهای ویژگی خوشه بندی[۴] می-شوند و به هر خوشه یک حباب[۵] گفته می‌شود. در مدل ترجمه، کلمات و حباب‌ها به عنوان دو زبان هم ارز در نظر گرفته می‌شوند و حباب‌ها به کلمات ترجمه می‌شوند. در این مدل احتمال شرطی یک کلمه به شرط یک حباب به یک مسئله‌ی بهینه سازی فرمول می‌شود. برای حاشیه نویسی تصویر جدید، ابتدا تصویر قطعه‌بندی می‌شود و از هر ناحیه بردار ویژگی استخراج می‌شود. برای هر ناحیه نزدیک‌ترین مرکز خوشه (حباب) انتخاب می‌شود و تصویر با حباب‌ها نمایه می‌شود. سپس با کمک نتایج بهینه سازیِ احتمالات شرطیِ یک کلمه به شرط یک حباب،احتمال هر حباب به یک کلمه ترجمه می‌شود. مدل CMRM[۶] توسط جون[۷] و همکاران ارائه شد [۴]. مدل CMRM هر تصویر از مجموعه آموزش را به صورت حباب‌های متناظر با نواحی تصویر و مجموعه کلمات برچسب خورده به تصویر نشان می‌دهد. احتمال توزیع توأم مشاهده‌ی کلمه کلیدی و حباب‌های هر تصویر، بر اساس نظریه‌ی بیز[۸] از روی تصاویر مجموعه‌ی آموزش تخمین زده می‌شود. برای حاشیه نویسی تصویر جدید، ابتدا تصویر قطعه بندی می‌شود و از هر ناحیه بردار ویژگی استخراج می‌شود. برای هر ناحیه نزدیک‌ترین مرکز خوشه (حباب) انتخاب می‌شود و تصویر با حباب‌ها نمایه می‌شود. سپس احتمال توأم هر کلمه و حباب‌ها محاسبه می‌شود و کلمات با بیشترین مقدار احتمال انتخاب شده و به تصویر برچسب زده می‌شود. مدل CRM[۹] توسط لاورنکو[۱۰] و همکاران ارائه شد [۵]. در این مدل فضای ویژگی‌ها به صورت پیوسته در نظر گرفته می‌شود. تصاویر ابتدا قطعه بندی شده و به نواحی تقسیم می‌شود و هر ناحیه با یک بردار ویژگی پیوسته نمایه می‌شود. در مدل CRM، بر خلاف مدل CMRM تصاویر با حباب‌ها نمایه نمی‌شوند، بنابراین نیازی به خوشه بندی بردارهای ویژگی نیست. تابع توزیع توأم بین نواحی تصویر (که یک فضای پیوسته است) و کلمات (که یک فضای گسسته است) تخمین زده می‌شود و از این توزیع برای حاشیه نویسی تصاویر جدید استفاده می‌شود. مدل MBRM[۱۱] که توسط فنگ[۱۲] و همکاران ارائه شده است [۶]، مشابه مدل CRM است با این تفاوت که تصاویر به جای قطعه بندی، به نواحی مستطیلی تقسیم می‌شوند و هر ناحیه‌ی مستطیلی با یک بردار ویژگی پیوسته نمایه می‌شود. همچنین بر خلاف مدل CRM که توزیع کلمات در یک تصویر را نمایی فرض می‌کند، در مدل MBRM توزیع کلمات در یک تصویر برنولی چندگانه در نظر گرفته می‌شود.

بعضی دیگر از مدل‌ها هر کلمه کلیدی را یک دسته جداگانه در نظر گرفته و بدین ترتیب مسئله حاشیه‌نویسی را به صورت چندین مسئله دسته‌بندی باینری تلقی می‌کنند. این مدل‌ها در گروه مدل-های مبتنی بر دسته‌بندی قرار می‌گیرند. اگرچه این مدل‌ها کارایی

[1] Translation model
[2] Duygulu
[3] Segmentation
[4] Clustering
[5] Blob

[6] Cross-Media Relevance Model
[7] Jeon
[8] Bayes Theorem
[9] Continuous Relevance Model
[10] Lavrenko
[11] Multiple Bernoulli Relevance Model
[12] Feng

بهتری نسبت به مدل‌های احتمالاتی دارند اما از آنجا که برای هر برچسب موجود در دادگان، یک دسته‌بند جداگانه آموزش می‌دهند، پیچیدگی و اتلاف زمان در این گروه از مدل‌ها بیشتر است. از طرف دیگر در حالتی که عدم‌توازن بین نمونه‌های مثبت و منفی هر برچسب در دادگان وجود داشته باشد، این مدل‌ها آسیب پذیر می‌باشند.مدل‌های احتمالاتی نسبت به مدل‌های مبتنی بر دسته‌بندی، مصرف داده کمتری دارند، ولیکن در شرایطی که تصاویر مشابه از نظر بصری، معانی متفاوتی را داشته باشند آسیب‌پذیر خواهند بود.در مدل SML[1] که توسط کارنیرو[2] و همکاران ارائه شده است [۷]، حاشیه نویسی خودکار تصاویر به یک مسئله‌ی دسته بندی[3] چند کلاسی فرمول شده است. در این مدل به تعداد کلمه‌ها، دسته بند دودویی یاد گرفته می‌شود. برای دسته بندی چند کلاسی از روش یک در برابر همه[4] استفاده می‌شود. در مرحله‌ی تست، کلمات باهم رقابت می‌کنند و کلمات با بیشترین رأی انتخاب می‌شود.مدل TMIML[5] یک چارچوب چند برچسبی چند نمونه‌ای[6] است[۸]. در این مدل تصاویر قطعه‌بندی[7] می‌شوند و هر ناحیه‌ی تصویر یک نمونه در نظر گرفته می‌شود. هر تصویر که شامل چندین نمونه متناظر با ناحیه‌های آن است، یک کیسه[8] در نظر گرفته می‌شود. برای هر کلمه‌ی w، کیسه‌ها برچسب مثبت و منفی می‌خورند. به این صورت که کیسه‌هایی که شامل کلمه‌ی w هستند برچسب مثبت و بقیه برچسب منفی می‌خورند. از میان کیسه‌های با برچسب مثبت، چند نمونه به عنوان نماینده‌ی کلمه‌ی w انتخاب شده و سپس تابع توزیع کلمه‌ی w از روی نمونه‌های انتخاب شده تخمین زده می‌شود.مدل HDGM[9][۹] مدلی سلسله مراتبی جهت حاشیه‌نویسی تصاویر می‌باشد. در این مدل ابتدا الگوریتم k-means جهت خوشه‌بندی تصاویر آموزشی بر اساس محتوای بصری آنها استفاده شده است، بدین معنی که هر تصویر آزمون به چندین خوشه منتسب می‌شود که اشاره به موضوع مشخصی دارند. سپس مدل SVM برای یادگیری تصاویر آموزشی داخل هر خوشه استفاده می‌شود و مدل دسته‌بندی ساخته می‌شود. هر تصویر آزمون با کمک مدل شاخته شده به خوشه‌ای منتسب می‌شود. الگوریتمی جهت گسترش معانی و تصاویر مرتبط‌استفاده می‌شود و سپس مجموعه تصاویر مرتبطی برای هر تصویر بدون برچسب به دست می‌آید. برچسب‌های این تصاویر مرتبط بر اساس مدل پیشنهاد شده وزنی را خواهند یافت و تصویر بدون برچسب توسط برچسب‌های با وزن بالاتر حاشیه‌نویسی می‌شود.

گروه سوم از مدل‌ها، یکی از قدیمی‌ترین، ساده‌ترین و در عین حال کارآمدترین مدل‌ها در دسته‌بندی یعنی مدل k نزدیک‌ترین همسایه می‌باشد. این مدل به خصوص در شرایطی که تعداد نمونه‌های آموزشی روزافزون است، بسیار کارآمد می‌باشد (از آنجا که در این رویکرد مدلی ساخته نمی‌شود، پس در شرایطی که تعداد تصاویر آموزش افزایش یابد برخلاف دو گروه قبلی نیازی به ساخت دوباره مدل ندارند). از تحقیقات بنیادی و پیشگامی که در زمینه حاشیه‌نویسی انجام گرفته‌اندو مبتنی بر رویکرد نزدیکترین همسایه می‌باشندمی‌توان به [۱۰] و [۱]اشاره کردTagProp.[۱۰]، یک مدل نزدیکترین همسایه وزن‌دار است که با درنظر گرفتن ترکیب وزن‌داری از حضور و عدم‌حضور کلمات کلیدی میان همسایه‌های هر تصویر آزمون، آن تصویر را برچسب‌گذاری می‌کند. مدل JEC[10] و Lasso[11] توسط ماکادیا[12] و همکاران ارائه شده است[1]. در این دو مدل از هر تصویر چندین بردار ویژگی استخراج می‌شود و سپس فاصله‌ی تصویر جدید با تصاویر آموزش بر اساس بردار ویژگی‌های استخراج شده محاسبه می‌شود. تفاوت مدل JEC و Lasso در روش محاسبه‌ی فاصله است. در مدل JEC فاصله براساس میانگین فاصله بردار ویژگی‌ها تعریف می‌شود (همه‌ی بردار ویژگی‌ها سهم یکسان در فاصله دارند) و در مدل Lasso برای محاسبه‌ی فاصله، به هر بردار ویژگی یک وزن تخصیص می‌دهد که با یک فرایند این وزن‌ها یاد گرفته می‌شود. عبدالله زاده حاشیه‌نویسی خودکار تصاویر را به عنوان یک مسئله دسته‌بندی چند برچسبی مدل کرده است و دو دسته‌بند چندبرچسبی بر مبنای قاعده نزدیکترین همسایه ارائه کرده است. دسته‌بندها بر اساس فاصله به همسایه یک وزن تخصیص می‌دهند و سپس بر اساس وزن و حاشیه‌های همسایه‌ها، تصویر جدید را حاشیه‌نویسی می‌کنند [۱۱].بهرامی و صنیعی، رویکردی معنایی با کمک الگوریتم وراثتی جهت برچسب زنی به تصاویر پیشنهاد داده‌اند. در مدل ارائه شده پس از دسته‌بندی تصاویر در فضاهای ویژگی گوناگون، به ازای هر تصویر بردارهای امتیازی وجود دارند که احتمال انتساب کلمات به تصویر را بیان می‌کنند. الگوریتم وراثتی جهت یافتن ارتباط میان فضاهای ویژگی و معانی پیشنهاد شده است. الگوریتم وراثتی ارتباط معنایی را با ارائه برداری وزن‌دار به ازای هر فضای ویژگی بیان می‌کند. بردار معنایی وزن‌دار برای هر فضا، در اصل بیانگر میزان قدرت تشخیص آن فضای ویژگی برای معانی مختلف است. در نهایت، این بردارهای معنایی، وزن‌های موجود در بردارهای امتیاز را تغییر می‌دهند به صورتی که کارایی حاشیه‌نویسی افزایش می‌یابد [۱۲].

اگرچه مرحله یادگیری نقش مهمی در حاشیه‌نویسی تصاویر دارد، ولیکن اینکه بر اساس چه ویژگی‌هایی یادگیری صورت می‌پذیرد نیز

[1] Supervised multiclass labeling
[2] Carneiro
[3] Classification
[4] One-versus-all
[5] Transductive Multi-Instance Multi-Label
[6] Multi-Instance Multi-Label
[7] Segmentation
[8] Bag
[9] Hierarchical discriminative and generative mode

[10] Joint Equal Contribution
[11] Penalized Logistic Regression
[12] Makadia

تأثیر فراوانی در کارایی نهایی دارد. استفاده از ویژگی‌های مناسب و بهینه یکی از بحث‌های اساسی در یادگیری ماشین می‌باشد و اغلب یکی از مهم‌ترین فاکتورها در کارایی سیستم دسته‌بندی است. رویکردهای انتخاب ویژگی زیادی در طول سال‌ها و در حوزه‌های مختلف پیشنهاد شده است. از کارهایی که در زمینه انتخاب ویژگی در حوزه حاشیه‌نویسی انجام گرفته‌اندمی‌توان به لی[1] و همکاران الگوریتم یادگیری Adaboost پویا با رویکرد انتخاب ویژگی مبتنی بر الگوریتم وراثتی جهت حاشیه‌نویسی تصاویر با استاندارد MPEG-7 پیشنهاد داده‌اند[۱۳]. در هر تکرار از یادگیری Adaboost، الگوریتم وراثتی جهت تولید پویای زیر مجموعه‌ای از ویژگی‌ها که بر اساس آنها دسته‌بندهای ضعیف ساخته می‌شوند، به کار گرفته می‌شود. البته لازم به ذکر است که در این مقاله تنها به ۲۰ دسته خاص از کلمات پرداخته شده است. ستیا[2] و همکاران نیز از Gaussian mixture model برای وزن دهی به ویژگی به صورت کارآمد استفاده کرده‌اند[۱۴].

هدف این پژوهش طراحی یک سامانه برچسب‌گذاری خودکار تصاویر مبتنی بر رویکرد نزدیک‌ترین همسایه می‌باشد. جهت افزایش کارایی و انتخاب مؤثر همسایه‌ها، رویکردی مبتنی بر خوشه‌بندی دوسطحی پیشنهاد شده است. از طرفی رویکرد انتخاب ویژگی‌ای متفاوت نسبت به سایر رویکردهای ارائه شده و مبتنی بر الگوریتم وراثتی در حالت چندبرچسبی پیشنهاد شده است.

۳ روش پیشنهادی

در این بخش از مقاله، روش پیشنهادی برای حاشیه‌نویسی تصاویر توضیح داده شده است. روش پیشنهادی به این صورت است: ابتدا ویژگی‌های بصری مناسب از تصاویر استخراج می‌شوند. بعضی از این ویژگی‌ها سراسری و بعضی محلی هستند تا از مزایای هر دو نوع ویژگی‌ها (سراسری و محلی) استفاده شود. بعد از استخراج ویژگی، ویژگی‌هایی از هر بردار ویژگی با استفاده از الگوریتم وراثتی انتخاب می‌شوند. سپس به خوشه‌بندی تصاویر در هر بردار ویژگی کاهش یافته پرداخته می‌شود. در هر بردار ویژگی تصاویر ابتدا از جهت مشابهت بصری از هم تفکیک می‌شوند و در خوشه‌هایی جای می‌گیرند، سپس تصاویر موجود در هر خوشه این بار از جهت مشابهت معنایی خوشه‌بندیمی‌شوند.

پس از خوشه‌بندیدوسطحی در هر بردار ویژگی، تصاویربدون برچسب به خوشه‌ها منتسب می‌شوند سپس هر تصویر جدید به کمک تصاویر هم خوشه‌اش دسته‌بندیمی‌شوند. خروجی هر دسته‌بند در هر فضای ویژگی برای هر تصویر بدون برچسب، یک بردار است که بردار امتیاز نام دارد. این بردار درجه عضویت هر کلمه برای تصویر را نشان می‌دهد. برای ترکیب نتایج دسته‌بندها در فضاهای ویژگی مختلف، بردارهای امتیاز با هم ترکیب می‌شوند و تصویر حاشیه‌نویسی‌می‌شود. شکل۱ شمای کلی از سامانه پیشنهادی را نشان می‌دهد. در ادامه بخش‌های مختلف سامانه پیشنهادی توضیح داده می‌شوند.

۳-۱ انتخاب ویژگی

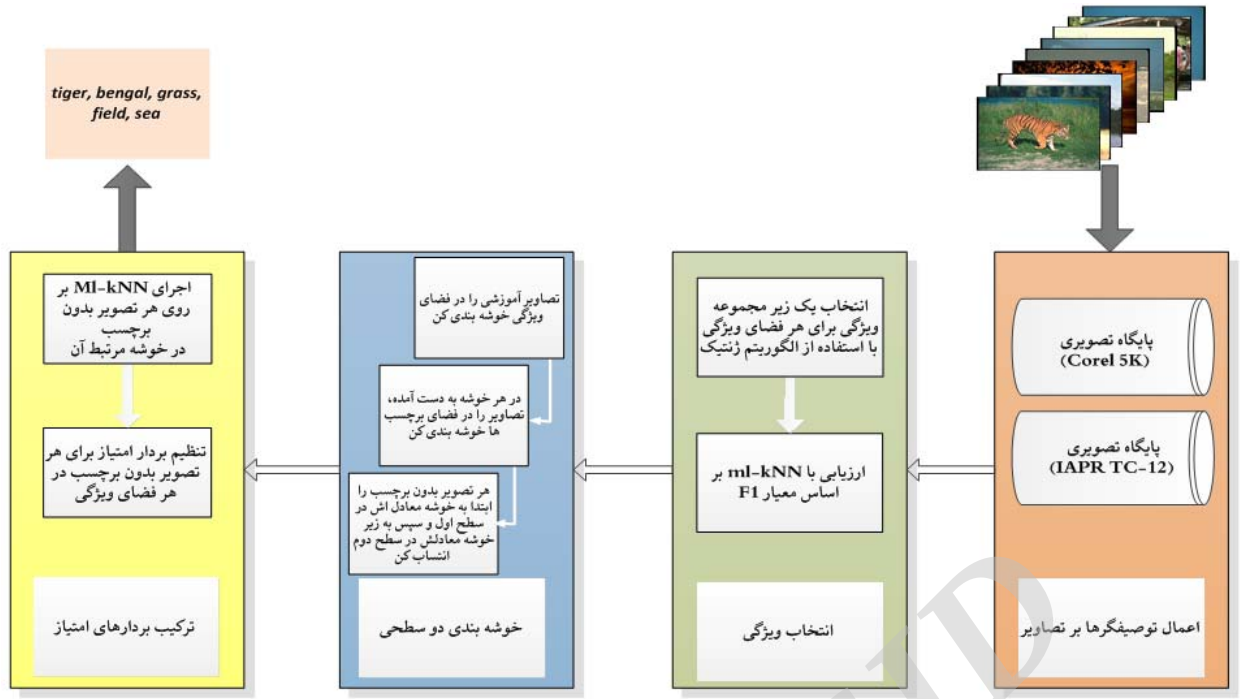
انتخاب ویژگی نقش مهمی در فرایند تحلیل داده دارد، چون اغلب، ویژگی‌های اضافه دقت و سرعت الگوریتم‌ها را کاهش می‌دهند. هدف از انتخاب ویژگی حذف ویژگی‌های اضافه و نامربوط است [۱۵]. با توجه به اینکه در این پژوهش از هر تصویرچندین بردار ویژگی استخراج شده است ممکن است بعضی ویژگی‌ها اضافی یا تکراری باشند و علاوه بر اینکه زمان اجرا را افزایش می‌دهند کارایی سامانه را نیز کاهش می‌دهند. در این پژوهش، جهت انتخاب ویژگی مبتنی بر مدل‌هایطبقه‌بند[3]، الگوریتم وراثتی پیشنهاد شده است. به طور کلی مدل‌های طبقه‌بند از یک دسته‌بند داخلی جهت دسته‌بندی زیر مجموعه ویژگی‌ها و تخمین کارایی آنها بر اساس معیارهایی چون F1 استفاده می‌کنند. از آنجا که در این پژوهش تصاویر در ۱۵ فضای ویژگی مختلف توصیف می‌شوند در نتیجه الگوریتم وراثتی نیز به صورت جداگانه بر روی هرفضا اجرا می‌شود و ابعاد هر فضا به صورت مستقل از فضاهای دیگر کاهش داده می‌شود. به همین دلیل مقدارپارامترهای الگوریتم وراثتی برای هر فضا متفاوت از فضای دیگر است. مقادیر مناسب پارامترها برای هر فضای ویژگی به صورت تجربی به دست می‌آید.

جزئیاتی از الگوریتم وراثتی جهت انتخاب ویژگی در زیر آمده است:

کد کردن کروموزم‌ها: به ازای هر ویژگی یک بیت در نظر گرفته می‌شود که حضور یا عدم حضور ویژگی را نمایش می‌دهد.

جمعیت اولیه: بسته به هر فضا چندین کروموزومبه صورت تصادفی تولید می‌شوند. برای تولید یک کروموزوم ابتدا تعداد یک‌ها به صورت تصادفی تولید شده و سپس یک‌ها به صورت تصادفی و یکنواخت در کروموزوم توزیع می‌شوند. برای مثال برای ویژگی شماره ۱۱ در دادگان Corel5k مقدار جمعیت اولیه ۴۰ انتخاب شده است.

[1] Li
[2] Setia
[3] Wrapper models



شکل۱ شمای کلی از سامانه پیشنهادی

تابع برازندگی: از معیار F1 برای سنجش برازندگی هر کروموزوم استفاده می‌شود. برای تعیین برازندگی هر کروموزوم ابتدا مجموعه ویژگی جدیدی تولید می‌شود که تنها شامل ویژگی‌هایی است که بیت معادل آنها در کروموزوم ۱ است، حال با توجه به این زیر مجموعه‌ویژگی جدید، مجموعه داده‌ای از تصاویر و ویژگی‌های انتخاب شده ایجاد می‌شود و دسته‌بندی چندبرچسبی روی تصاویر این مجموعه داده‌انجام می‌شود. برای دسته‌بندی چندبرچسبی، از الگوریتم چندبرچسبی kNNوزن‌دارتوضیح داده در بخش ۳-۳ استفاده می‌شود. به این صورت که برای هر تصویر بدون برچسبkنزدیک‌ترین همسایه‌اش از مجموعه آموزش انتخاب می‌شوند و سپس هرهمسایه براساس دوری و نزدیکی‌اش وزن می‌گیرد. برچسب‌های هر همسایه براساس وزن همان همسایه وزن‌دهی می‌شوند. درنهایت تنها چند برچسب با بالاترین وزن انتخاب می‌شوند و به تصویر برچسب می‌خورند. توضیحات بیشتر در مورد نحوه وزن‌دهی به برچسب‌های همسایه و دسته‌بندی در بخش ۳-۳ بیان شده است. حال که برچسب‌های تمام تصاویر آزمون معین شدند، باید کارایی حاشیه‌نویسی سنجیده شود. بدین منظور با کمک معیار F1بیان شده در بخش ۴-۱-۱ مقدار F1 نهایی محاسبه می‌شود. مقدار بدست آمده برای این معیار در اصل بیانگر کارایی حاشیه‌نویسی با کمک ویژگی‌های انتخاب شده است که همان برازندگی کروموزوم می‌باشند. شکل۲ نحوه سنجش برازندگی یک کروموزوم نمونه را مطابق توضیحات بیان شده نشان می‌دهد.

روش انتخاب: از روش انتخاب تورنمنت استفاده شده است. در این روش برای انتخاب یک والد، دو کروموزوم به طور تصادفی انتخاب شده و از بین این دو کروموزوم، کروموزومی که برازندگی بیشتری دارد به عنوان والد انتخاب می‌شود. با این روش کروموزوم‌های با برازندگی کم نیز شانس انتخاب خواهند داشت و از همگرایی زودرس و گیر افتادن در بهینه محلی جلوگیری میکند.

باز هم بسته به فضای ویژگی مورد نظر در هر تکرار تعدادی کروموزم جدید با عملگر تلفیق تولید شده و به همان تعداد کروموزوم‌ها با برازندگی کم حذف می‌شوند. رای مثال برای ویژگی شماره ۱۱ در دادگان Corel5k در هر نسل ۱۰ نمونه از بهترین فرزندان جایگزین ۱۰ کروموزوم با برازندگی کم می‌شوند.

تلفیق: از تلفیق یک نقطه‌ای با احتمالی که برای هر فضای ویژگی متفاوت است استفاده می‌شود. برای مثال برای ویژگی شماره ۱۱ در دادگان Corel5k احتمال تلفیق ۰,۵ انتخاب شده است.

جهش: درصدی از بیت‌های کروموزوم بسته به فضای ویژگی موردنظر تغییر می‌کند. البته نرخ جهش نسبتاً بالا انتخاب می‌شود تا پراکندگی کروموزوم‌ها بیشتر شده و الگوریتم در بهینه محلی گیر نیفتد. برای مثال برای ویژگی شماره ۱۱ در دادگان Corel5k احتمال جهش ۰,۱ انتخاب شده است.

شرط توقف: الگوریتم پس از ۲۰ تکرار که بهبودی در برازندگی بهترین کروموزوم حاصل نشود متوقف می‌شود.

تعداد اجراهای الگوریتم: الگوریتم وراثتی چندین بار اجرا شده و بهترین نتیجه به عنوان نتیجه نهایی انتخاب شده است و بدین ترتیب ابعاد بردار ویژگی کاهش داده می‌شود.

۳-۲ خوشه‌بندی دوسطحی

اکثر مدل‌های حاشیه‌نویسی کنونی از تمام تصاویر برای دسته‌بندی و حاشیه‌نویسی تصاویر جدید استفاده می‌کنند که باعث می‌شود تصاویر نامرتبط زیادی در حاشیه‌نویسی تصاویر بدون برچسب

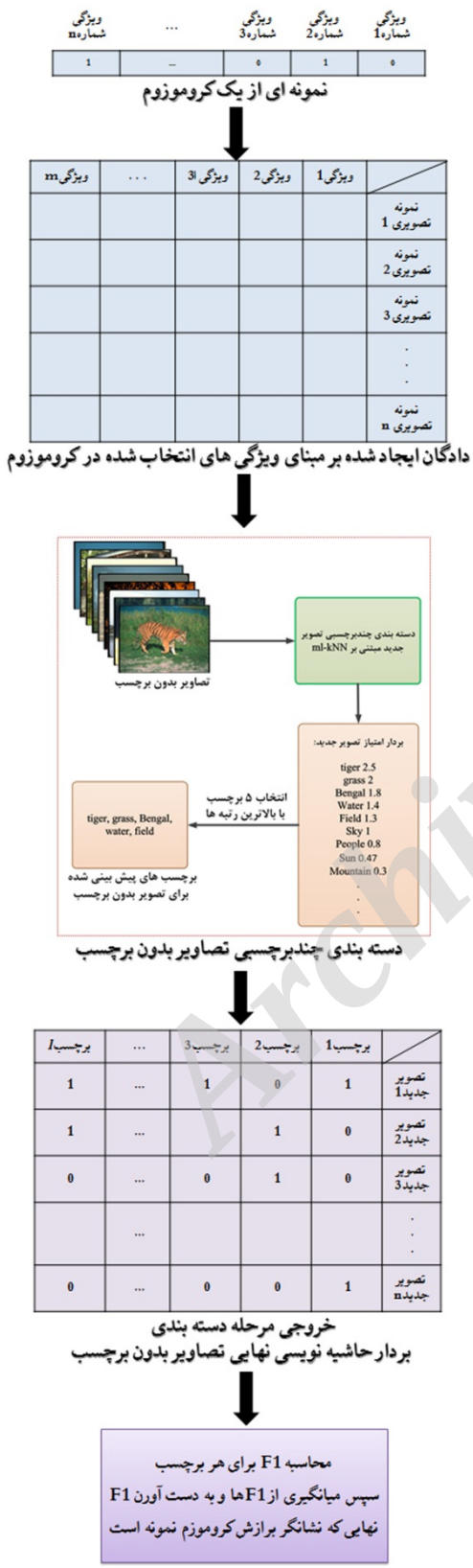
وارد شوند که علاوه بر کاهش کارایی حاشیه‌نویسی، زمان اجرا هم افزایش می‌یابد. در این سامانه جهت فایق آمدن بر این مسأله، تصاویر در دو سطح بر اساس ویژگی‌های بصری و نیز معانی(برچسب‌های تصاویر) با کمک الگوریتم خوشه‌بندی k-means از هم تفکیک می‌شوند.

در سطح اول تصاویر بر اساس ویژگی‌های بصری با الگوریتم k-means خوشه‌بندی می‌شوند. با این خوشه‌بندی تصاویری که از لحاظ بصری شبیه به هم هستند در یک خوشه قرار می‌گیرند که انجام این کار باعث افزایش دقت دسته‌بندی می‌شود. برای اینکه تصاویر مشابه از نظر معنایی هم به یکدیگر نزدیک شوند (چرا که تصاویر مشابه از جهت بصری لزوماً معنای یکسانی را به اشتراک نمی‌گذارند) در سطح دوم تصاویر داخل هر خوشه در فضای برچسب‌ها خوشه‌بندی می‌شوند. این خوشه‌بندی هم توسط الگوریتم k-means انجام می‌شود. منظور از فضای برچسب‌ها (معنایی)، فضایی است که در آن انتساب کلمات (معانی) به تصاویر مشخص شده است. فضای برچسب‌ها، فضایی دودویی است که در آن به ازای هر تصویر یک برداری دودویی به طول تعداد کلمات کلیدی دادگان تعریف شده است. اگر معنایی به تصویر برچسب خورده باشد، بیت متناظر با آن معنا در بردار یک و در غیر اینصورت صفر خواهد بود. پس از انجام این دو سطح خوشه‌بندی نیاز است تا خوشه مورد نظر هر تصویر آزمون معین شود تا دسته‌بندی هر تصویر بدون برچسب تنها با کمک تصاویر مرتبط با آن انجام شود. بنابراین تصاویر بدون برچسب باید به یکی از خوشه‌های معنایی سطح آخر منتسب شوند. برای انجام این کار ابتدا با محاسبه فاصله تصویر آزمون با مراکز خوشه‌های سطح اول، نزدیکترین خوشه یافته می‌شود و تصویر آزمون در سطح اول به نزدیکترین خوشه منتسب می‌شود، پس با این کار سایر خوشه‌های سطح اول و تمام زیرخوشه‌های آنها در سطح دوم نادیده گرفته می‌شوند. سپس مجدداً فاصله تصویر آزمون با مراکز خوشه‌های سطح دوم خوشه‌ای که به آن منتسب شده بود، محاسبه می‌شود و این بار تصویر آزمون به نزدیکترین خوشه معنایی منتسب می‌شود. شکل۳ نحوه این خوشه‌بندی‌ها را نشان می‌دهد.

از آنجا که تصاویر در چندین فضای ویژگی توصیف می‌شوند، خوشه‌بندی‌ها نیز به صورت جداگانه در هر فضا صورت می‌گیرد. بدین صورت که در نهایت ۱۵ نوع خوشه‌بندی متفاوت برای ۱۵ فضای ویژگی متفاوت به دست می‌آید و در مرحله دسته‌بندی نیز جهت دسته‌بندی هر فضا از خوشه‌بندی انجام گرفته برای همان فضا استفاده می‌شود یعنی جهت دسته‌بندی در یک فضای خاص، تصاویر آزمون تنها به کمک تصاویری که در خوشه موردنظر آنها در همان فضا وجود دارند دسته‌بندی می‌شوند.

۳-۳ دسته‌بندی چندبرچسبی

پس از اتمام مرحله قبلی، تصاویر مرتبط به هم داخل یک خوشه قرار گرفته‌اند. در این مرحله، نوعی از تکنیک‌های دسته‌بندی با نظارت، الگوریتم k نزدیک‌ترین همسایه (kNN) جهت دسته‌بندی



شکل۲ نحوه سنجش برازندگی یک کروموزوم

تصاویر جدید با کمک تصاویر مرتبط با آنها درخوشه مربوطه به کار گرفته شده است.

همان‌طور که گفته شد حاشیه‌نویسی خودکار تصاویر یک مسئله چند برچسبی است یعنی هر نمونه به چندین کلاس تعلق دارد. بنابراین دسته‌بند باید این قابلیت را داشته باشد که به هر نمونه چند برچسب بزند. برای این منظور از نسخه‌ای از دسته‌بند kNN به نام دسته‌بند kNN وزن‌دار چندبرچسبی استفاده شده است [۱۱].

الگوریتم موردنظر بدین صورت است که ابتدا فاصله بین نمونه جدید و تمام نمونه‌های آموزش در همان خوشه محاسبه می‌شود و سپس بر اساس رابطه۱وزن هر یک از k همسایه معین می‌شود. فرض کنید d معیار فاصله باشد، X بردار ویژگی تصویر جدید و X1,X2, ... , Xk همسایه‌های X هستند که بر اساس فاصله‌ی d(Xi,X) به صورتصعودی مرتب شده‌اند. d(Xi,X) را به اختصار با di نشان می‌دهند. همانگونه که در رابطه ۱ مشاهده می‌کنید وزن هر همسایه چون همسایه زام که با w_j نشان داده می‌شود، برابر خواهد بود با حاصل تفاضل فاصله دورترین همسایه تا تصویر $(d_k)X$ و فاصله خود همسایه j ام تا تصویر $(d_j)X$ بخش بر حاصل تفاضل فاصله دورترین همسایه تا تصویر $(d_k)X$ و فاصله نزدیکترین همسایه تا تصویرX (d_1). با توجه به این رابطه وزن نزدیکترین همسایه به تصویر X برابر با یک و وزن دورترین همسایه برابر با ۰ خواهد بود. به عبارت دیگر با توجه به رابطه ۱ وزن همسایه های تصویر X از دورترین همسایه تا نزدیکترین آنها در بازه [۰,۱] مقداردهی خواهند شد.

$$w_j = \begin{cases} \frac{d_k - d_j}{d_k - d_1} & if\, d_k \neq d_j \\ 1 & if\, d_k = d_j \end{cases} \qquad (۱)$$

برای حاشیه‌نویسی تصویر جدید ابتدا k همسایه‌ی I_k , ... ,I_1,I_2 که بر اساس فاصله به صورت صعودی مرتب شده‌اند، انتخاب شده و وزن‌های آنها W_k , ... ,W_1,W_2 محاسبه می‌شوند. سپس وزن هر همسایه در بردار حاشیه‌نویسی آن ضرب می‌شود(بردارحاشیه‌نویسی برداری دودویی به طول تعداد کلمات کلیدی دادگان است که در آن مقدار ۱ یعنی آن کلمه به تصویر برچسب خورده و مقدار ۰ یعنی آن کلمه به تصویر برچسب نخورده است) و بردارهای حاصل با هم جمع می‌شوند. بردار حاصل جمع، بردارامتیاز نامیده می‌شود که طول آن به اندازه تعداد کلمات کلیدی

شکل۳ نمایی از خوشه‌بندی دوسطحی و نحوه تعیین خوشه‌های معنایی برای تصاویر بدون برچسب

دادگان است و نشان می‌دهد که هر کلمه چه امتیازی برای آن تصویر آزمون کسب کرده است.از آنجا که تصاویر در چندین فضای ویژگی دسته‌بندی می‌شوند پس به ازای هر فضای ویژگی، یک بردار امتیاز بدست خواهد آمد. در نهایت جهت برچسب‌گذاری هر تصویر تمامی بردارهای امتیاز حاصل از تمام فضاهای ویژگی با هم جمع شده و بردار ویژگی نهایی را می‌سازند سپس چندین کلمه با بالاترین امتیازها از میان تمام کلمات موجود در بردار امتیاز نهایی انتخاب می‌شوند و به تصویر آزمون برچسب می‌خورند. شکل۴ نحوه ترکیب بردارهای امتیاز تصویر x را نشان می‌دهد. در این شکل، هر سطر، خروجی حاصل از اجرای دسته‌بندی در یک فضای ویژگی است که همان بردار امتیاز است. در این حالت تمام این بردارهای امتیاز با هم جمع می‌شوند و بردار امتیاز نهایی بدست می‌آید. پس از محاسبه‌ی بردار امتیاز، برای حاشیه‌نویسی تصویر جدید دو رویکرد وجود دارد: رویکرد اول این است که *1*کلمه با بیشترین امتیاز انتخاب شده و به تصویر برچسب زده شوند. به پارامتر *1* (تعداد کلماتی که به هر تصویر منتسب می‌شود) طول حاشیه‌نویسی[1] گفته می‌شود. رویکرد دوم این است که یک آستانه مشخص کرده و کلماتی که امتیاز بیشتر از آستانه دارند انتخاب شوند. مشکل رویکرد دوم چگونگی انتخاب آستانه‌ی مناسب است که به پارامترهای مسئله وابسته است. برای مثال اگر تعداد کلمات کلیدی دادگان ۱۰ برچسب باشد و ۵ توصیفگر ویژگی بر روی تصاویر اعمال شده باشند و نیز مقدار پارامتر طول حاشیه‌نویسی ۵ کلمه انتخاب شده باشد، همان‌طور که در شکل۴ مشاهده می‌کنید بعد از ترکیب این بردارهای امتیاز، ۵ کلمه با بیشترین امتیاز انتخاب شده و به تصویر برچسب می‌خورند. این کلمات با رنگ قرمز و زیر خط در بردار امتیاز نهایی نشان داده شده‌اند.

همان‌طور که گفته شد کاربرد حاشیه‌نویسی در بازیابی تصویر است. رتبه‌بندی تصاویر بر اساس مفهوم پرس‌وجو یک ویژگی ضروری برای سامانه‌های بازیابی تصویر است چون اکثر کاربران انتظار دارند که تصویر موردنظر خود را در همان ۱۰ یا ۲۰ تصویر اولی که بازیابی شده بیابند و بنابراین الگوریتم ما باید بتواند تصاویر مربوط به یک مفهوم (کلمه) را رتبه‌بندی کند و در زمان بازیابی تصاویر را بر اساس رتبه‌ی آنها مرتب کند. خروجی سامانه پیشنهادی پس از طی تمام مراحل برای هر تصویر جدید در هر فضای ویژگی یک بردار است که امتیاز هر کلمه را نشان می‌دهد. این بردارها با هم جمع می‌شوند و بردار امتیاز حاصل جمع به دست می‌آید. علاوه بر اینکه کلمات با بالاترین امتیازها از بردار حاصل جمع انتخاب شده و به تصویر برچسب می‌خورند، مقدار امتیاز هر کلمه برای تصویر ذخیره می‌شود تا از آن برای رتبه‌بندی تصاویر مربوط به یک کلمه استفاده شود. بنابراین الگوریتم پیشنهادی این قابلیت را دارد که تصاویر مربوط به هر مفهوم را رتبه‌بندی کند.

۴ ارزیابی و مقایسه

برای ارزیابی روش پیشنهادی و مقایسه با کارهای موجود، دو دادگان شناخته شده و استاندارد که در آنها تصاویر توسط انسان حاشیه‌نویسی شده‌اند، انتخاب شده است. دادگان Corel5k که در سال ۲۰۰۲ توسط دویگلو و همکاران [۳] و دادگان IAPR TC-12 که در سال ۲۰۰۶ توسط گروبینگر[2] و همکاران [۱۶] ارائه شده‌اند. دادگان Corel5k قدیمی‌تر بوده و کارهای زیادی برای ارزیابی روش خود از آن استفاده کرده‌اند. دادگان IAPR TC-12 شامل تصاویر مختلفی از ورزش و اکشن، انسان، حیوان، شهرها، مناظر و جنبه‌های دیگر زندگی امروزی می‌باشد. برای تصاویر هر دوی این دادگان، ۱۵ ویژگی بیان شده در جدول۱ استخراج شده‌اند. جدول۲ اطلاعات این دو دادگان را نشان می‌دهد. تصاویر آموزش و آزمون به صورت استاندارد توسط تولیدکنندگان دادگان تفکیک شده‌اند و در تمامی مقالات به همین صورت از مجموعه‌های آموزش و آزمون استفاده می‌شود و هدف حاشیه‌نویسی مجموعه تصاویر آزمون مشخصی می‌باشد. همان‌طور که در جدول۱ مشاهده می‌کنید برای نمونه در دادگان Corel5k برای بعضی از کلمات مجموعه آموزش تنها یک نمونه آموزشی وجود دارد ولی برای بعضی دیگر حتی تا ۱۰۰۴ تصویر آموزشی نمونه وجود دارد و این تنها پیچیدگی حاشیه‌نویسی را افزایش داده و باعث می‌شود با وجود کارهای فراوان انجام گرفته در این حوزه هنوز هم از کارایی قابل قبول دور باشند. شکل۵ چند تصویر به عنوان نمونه از دادگان Corel5k و IAPR TC-12 را نشان می‌دهند.

جدول۱ مشخصات دادگان Corel5k و IAPR TC-12

	Corel5k	IAPR TC-12
تعداد کل تصاویر	۴۹۹۹	۱۹۶۲۷
تعداد تصاویر مجموعه آموزش	۴۵۰۰	۱۷۶۵۵
تعداد تصاویر مجموعه آزمون	۴۹۹	۱۹۶۲
تعداد کل کلمات	۳۷۴	۲۹۱
کلمات ظاهر شده در مجموعه آموزش	۳۷۱	۲۹۱
کلمات ظاهر شده در مجموعه آزمون	۲۶۳	۲۹۱
تعداد برچسب‌های هر تصویر (کمترین-میانگین-بیشترین)	۱-۳,۵-۵	۱-۵,۷-۲۳
تعداد تصاویر آموزش برای هر کلمه (کمترین-میانگین-بیشترین)	۰ — ۴۲,۴ — ۱۰۰۴	۴۴ — ۳۴۷,۷ - ۴۹۹۹

[1] Annotation length

[2] Grubinger

	کلمه۱	کلمه۲	کلمه۳	کلمه۴	کلمه۵	کلمه۶	کلمه۷	کلمه۸	کلمه۹	کلمه۱۰
بردار امتیاز حاصل از دسته بندی در فضای ویژگی ۱	۰.۲	۰	۰.۴	۰	۰.۷	۰.۹	۰	۰.۵	۰.۷	۰.۲
بردار امتیاز حاصل از دسته بندی در فضای ویژگی ۲	۰	۰.۹	۰.۲	۰.۵	۰.۱	۰.۵	۰.۴	۰.۴	۰	۰.۷
بردار امتیاز حاصل از دسته بندی در فضای ویژگی ۳	۰.۸	۰	۰.۶	۰.۲	۰	۰.۵	۰.۴	۰	۰.۸	۰
بردار امتیاز حاصل از دسته بندی در فضای ویژگی ۴	۰.۷	۰.۱	۰.۴	۰.۸	۰.۵	۰	۰.۳	۰.۲	۰.۷	۰
بردار امتیاز حاصل دسته بندی در فضای ویژگی ۵	۰	۰.۴	۰	۰	۰.۷	۰.۸	۰.۱	۰.۹	۰.۲	۰.۲
بردار امتیاز نهایی حاصل تجمیع تمام بردار های امتیاز	۱.۷	۱.۴	۱.۶	۱.۵	۲	۲.۷	۱.۲	۲	۲.۴	۱.۱

شکل۴ مثالی از نحوه برچسب‌زنی نهایی به تصویر آزمون x

جدول۲ ویژگی‌های بصری استفاده شده از مرجع [۱۰]

شماره	نوع	روش توصیف	ابعاد بردار ویژگی
۱	محلی	هیستوگرام Hue با روش تشخیص نقاط مهم تور متراکم چندبعدی	۱۰۰
۲	محلی	هیستوگرام Hue با سه نوار افقی با روش تشخیص نقاط مهم تور متراکم چندبعدی	۳۰۰
۳	محلی	هیستوگرام SIFT با روش تشخیص نقاط مهم تور متراکم چندبعدی	۱۰۰۰
۴	محلی	هیستوگرام SIFT با سه نوار افقی با روش تشخیص نقاط مهم تور متراکم چندبعدی	۳۰۰۰
۵	سراسری	Gist	۵۱۲
۶	محلی	هیستوگرام Hue با روش تشخیص نقاط مهم هریس- لاپلاسی	۱۰۰
۷	محلی	هیستوگرام Hue با سه نوار افقی با روش تشخیص نقاط مهم هریس- لاپلاسی	۳۰۰
۸	محلی	هیستوگرام SIFT با روش تشخیص نقاط مهم هریس- لاپلاسی	۱۰۰۰
۹	محلی	هیستوگرام SIFT با سه نوار افقی با روش تشخیص نقاط مهم هریس- لاپلاسی	۳۰۰۰
۱۰	سراسری	هیستوگرام در فضای رنگ HSV	۴۰۹۶
۱۱	سراسری	هیستوگرام با سه نوار افقی در فضای رنگ HSV	۵۱۸۴
۱۲	سراسری	هیستوگرام در فضای رنگ LAB	۴۰۹۶
۱۳	سراسری	هیستوگرام با سه نوار افقی در فضای رنگ LAB	۵۱۸۴
۱۴	سراسری	هیستوگرام در فضای رنگ RGB	۴۰۹۶
۱۵	سراسری	هیستوگرام با سه نوار افقی در فضای رنگ RGB	۵۱۸۴

شکل۵ سمت راست چند نمونه از تصاویر دادگان Corel5k و سمت چپ چند نمونه از تصاویر دادگان IAPR TC-12

۴-۱ معیارهای ارزیابی

۴-۱-۱- ارزیابی حاشیه‌نویسی

جهت ارزیابی کیفیت حاشیه‌نویسی، دقت و فراخوان به ازای تک تک کلمات دادگان (v_i) محاسبه می‌شود.

$$Precision(v_i) = \frac{N_c}{N_s} \quad (۲)$$

$$Recall(v_i) = \frac{N_c}{N_r} \quad (۳)$$

$$F1(v_i) = 2.\frac{Precision(v_i)*Recall(v_i)}{Precision(v_i)+Recall(v_i)} \quad (۴)$$

که N_s، N_c و N_r به ترتیب، تعداد تصاویر نسبت داده شده در فاز آزمون، تعداد تصاویر صحیح نسبت داده شده در فاز آزمون و تعداد تصاویر نسبت داده شده در دادگان، به ازای هر کلمه v_iمی‌باشند.

پس از محاسبه نرخ‌های دقت و فراخوان برای تمام کلمات مجموعه آزمون، میانگین یکنواخت آنها محاسبه می‌شود. بنابراین کلماتی که به صورت مکرر در دادگان تکرار شده‌اند (که احتمالاً سامانه آنها را خوب یاد گرفته) و کلمات نادر همگی وزن یکسانی در محاسبه دقت و فراخوان سامانه دارند یعنی تفاوتی بین کلمات از نظر وزن و اهمیت وجود ندارد. دقت و فراخوانی کل به ترتیب با رابطه‌های ۵ و ۶ محاسبه می‌شوند.

معیارهای دقت و فراخوان هر کدام به تنهایی می‌توانند افزایش یابند. یعنی می‌توان فراخوان را در ازای کاهش دقت، افزایش داد و برعکس. بنابراین دقت و فراخوان هیچ کدام به تنهایی عملکرد سامانه را نشان نمی‌دهند. برای رسیدن به معیاری که عملکرد سامانه را بهتر نشان دهد معیار F1 (که بر اساس هر دو معیار دقت و فراخوان محاسبه می‌شود) استفاده می‌شود (رابطه‌ی ۷). به همین دلیل در مراحل مختلف این سامانه از معیار F1 به عنوان تابع برازندگی و معیار کارایی اصلی استفاده شده است.

$$Precision = \frac{\sum_{v_i \in vocabularyset} \text{Precision}(v_i)}{|vocabularyset|} \quad (۵)$$

$$Recall = \frac{\sum_{v_i \in vocabularyset} \text{Recall}(v_i)}{|vocabularyset|} \quad (۶)$$

$$F1 = 2.\frac{Precision*Recall}{Precision+Recall} \quad (۷)$$

علاوه بر سه معیار دقت، فراخوان و F1، تعداد کلماتی که فراخوان آنها بزرگتر از صفر است نیز در تحقیقات گزارش شده است. این معیار نشانگر این است که سامانه توانسته است چه تعداد کلمه یاد بگیرد. این معیار را به اختصار NZR[1] نشان داده می‌شود. به عبارت دیگر NZR نشان می‌دهد که سامانه برای چند پرس‌وجو (کلمه) می‌تواند تصاویر مرتبط را بازیابی کند. این معیار ارتباط مستقیمی با فراخوان دارد.

۴-۱-۲- ارزیابی رتبه‌بندی

اگرچه معیارهای دقت، فراخوان و F1 مهم‌ترین و رایج‌ترین پارامترها برای ارزیابی و مقایسه سامانه حاشیه‌نویسی تصاویر هستند، اما معیار مهم دیگری به نام میانگین دقت متوسط[2] (mAP) وجود دارد که کیفیت رتبه‌بندی تصاویر توسط سامانه را ارزیابی می‌کند. رتبه‌بندی تصاویر یک ویژگی مهم و ذاتی برای سامانه‌های بازیابی تصاویر است. میانگین دقت متوسط عملکرد بازیابی را ارزیابی می‌کند. به این صورت که اگر در تصاویر بازیابی شده، تصاویر مربوط در رتبه‌ای اول باشند، این معیار به یک نزدیک خواهد بود و هرچه تصاویر مربوط در رتبه‌های آخر قرار گیرند این معیار کمتر خواهد بود.

برای محاسبه میانگین دقت متوسط ابتدا دقت متوسط[3] (AP) برای پرس‌وجوی q محاسبه می‌شود. دقت متوسط همانند مرجع [۱۷] با رابطه‌ی ۸ محاسبه می‌شود که در آن i رتبه در دنباله بازیابی شده، n تعداد تصاویر بازیابی شده، P(i) دقت بازیابی در i تصویر اول (تعداد تصاویر مرتبط تا رتبه i ام تقسیم بر i) و rel(i)یک تابع است که اگر تصویر در رتبه i ام مرتبط با پرس‌وجو باشد، مقدار آن یک و در غیر اینصورت صفر خواهد بود.

$$AP\ (q) = \frac{\sum_{i=1}^{n}(P(i)*rel(i))}{|\{releventimages\}|} \quad (۸)$$

میانگین دقت متوسط هم با رابطه‌ی ۹ محاسبه می‌شود که در آن Nqتعداد پرس و جوها را نشان می‌دهد.

$$mAP = \frac{\sum_{q=1}^{Nq} AP(q)}{N_q} \quad (۹)$$

۴-۲ تنظیم پارامترها و ارزیابی مراحل

۴-۲-۱- انتخاب ویژگی

در این بخش کارایی مرحله انتخاب ویژگی به عنوان بخشی از سامانه پیشنهادی گزارش می‌شود. جدول۳ ابعاد کلی دادگان قبل و بعد از انتخاب ویژگی را نشان می‌دهد. برای دادگان Corel5k ابعاد باقی مانده بعد از انتخاب ویژگی ۲۱% از ابعاد اولیه را شامل می‌شود و برای دادگان IAPR TC-12 تنها ۸,۴۸% از ابعاد باقی مانده است. با حذف این ابعاد بالای ویژگی‌ها، علاوه بر اینکه زمان اجرا کاهش می‌یابد، کارایی سامانه افزایش می‌یابد.

برای نشان دادن کارایی انتخاب ویژگی، دو سناریو به کار گرفته شده است. در سناریوی اول تمام ویژگی‌های استخراج شده

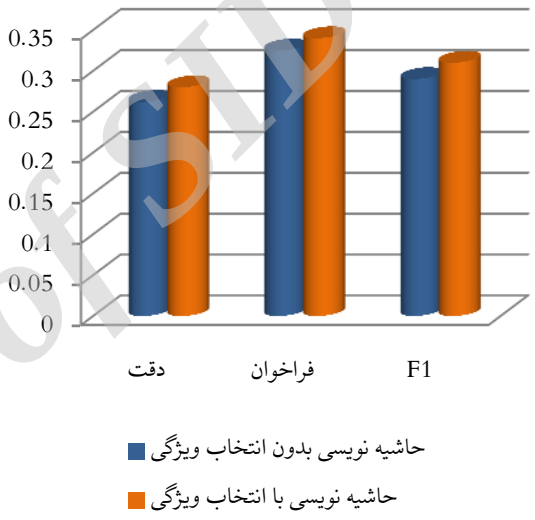
[1] Number of words with Non-Zero Recall

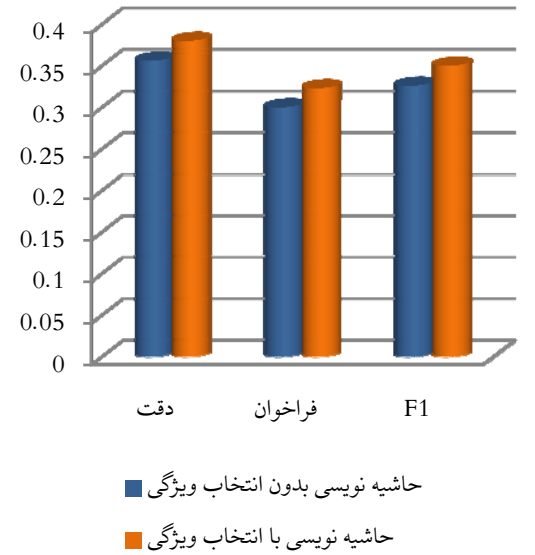
[2] Mean average precision
[3] Average Precision

استفاده شده است و در سناریوی دوم تنها از ویژگی‌هایی که توسط الگوریتم وراثتی در مرحله‌ی انتخاب ویژگی انتخاب شده‌اند، استفاده شده است. شکل۶ نتایج ارزیابی این دو سناریو را روی دو دادگان نشان می‌دهد. همان‌طور که مشاهده می‌کنید با انتخاب ویژگی در تمام معیارهای ارزیابی بهبود حاصل شده است.

جدول۳ تأثیر انتخاب ویژگی بر روی ابعاد دو دادگان

دادگان	مجموع ابعاد اولیه	ابعاد پس از کاهش	درصد کاهش
Corel5k	37152	7803	%۷۹
IAPR TC-12	37152	3152	%۹۱,۵۲





شکل۶ ارزیابی عملکرد انتخاب ویژگی روی هر دو دادگان، نمودار بالایی مربوط به دادگان Corel5k و نمودار پایینی مربوط به دادگان IAPR TC-12

۴-۲-۲- خوشه‌بندی دوسطحی

از آنجا که خوشه‌بندی پیشنهاد شده به صورت دوسطحیمی‌باشد، ارزیابی خوشه‌بندی باید پس از انجام هر دو سطح خوشه‌بندی محاسبه شود. در خوشه‌بندی پیشنهاد شده باید ابتدا خوشه‌بندی سطح اول را انجام داد و سپس خوشه‌های بدست آمده را در مجدداً در سطح دوم خوشه‌بندی کرد و سپس ارزیابی حاشیه‌نویسی را با خوشه‌های سطح دوم انجام داد. این نوع خوشه‌بندی باید به صورت تجربی و چندین مرتبه انجام شود و سپس بهترین خوشه‌بندی انتخاب شود.

برای نمونه جهت خوشه‌بندی تصاویر در فضای ویژگی ۷ ابتدا تصاویر بر اساس ویژگی‌ها به ۵ خوشه تفکیک می‌شوند و در مرحله دوم هر خوشه با توجه به اندازه و میزان تصاویری که در آن قرار دارند بر اساس فضای برچسب‌ها به چندخوشه تفکیک می‌شوند. جدول۴ خوشه‌بندی تصاویر بر اساس ویژگی شماره ۷ را به ۵ خوشه نشان می‌دهد. برای هر خوشه تعداد و درصد تصاویر آموزش موجود در آن و نیز تعداد تصاویر آزمونی که به خوشه منتسب می‌شوند بیان شده است. چنانچه خوشه شماره ۴ مجدداً بر اساس فضای معنا (برچسب‌ها) خوشه‌بندیشود ۳ خوشه به دست خواهد آمد. جدول۵ نحوه این خوشه‌بندی و تفکیک تصاویر آموزش و آزمایش موجود در خوشه ۴ از سطح اول را به ۳ خوشه نشان می‌دهد. بنابراین برای حاشیه‌نویسی تصاویر آزمون تنها درصد کمی (کمتر از ۱۰% به صورت میانگین) از کل تصاویر آموزش استفاده می‌شوند، که در دادگان با ابعاد بالاتر انجام این مرحله بسیار تأثیرگذارتر خواهد بود.با انجام آزمایشات متعدد برای تعیین تعداد خوشه‌ها برای هر دو سطح، این نتیجه حاصل شد که هر چه تعداد خوشه‌ها بیشتر باشد اگرچه مجموع مربعات خطا[1] برای هر خوشه و نیز مقدار مجموع کلمجموع مربعات خطای تمام خوشه‌ها کمتر می‌شود اما میزان F1 ای که با حاشیه‌نویسی تصاویر بر اساس آن خوشه‌ها به دست می‌آید کاهش می‌یابد و بالعکس؛ یعنی چنانچه تعداد خوشه‌های کمتری وجود داشته باشد، مجموع مربعات خطا افزایش می‌یابد اما کارایی حاشیه‌نویسی بهتر خواهد بود. دلیل کاهش کارایی سامانه با افزایش تعداد خوشه‌ها در این است که اگرچه با افزایش تعداد خوشه‌ها، خوشه‌های بهتری حاصل می‌شوند و اکثریت تصاویر درون خوشه‌ها مشابه ومرتبط هستند ولی احتمال انتساب اشتباه تصویر جدید به خوشه معادل‌اش با افزایش تعداد خوشه‌ها بیشتر خواهد شد. از آنجا که هدف افزایش کارایی حاشیه‌نویسی است تعداد خوشه‌ها به گونه‌ای انتخاب می‌شوند که میزان F1 حداکثر باشد. با توجه به اینکه ابعاد فضاهای ویژگی متفاوت هستند برای دادگان Corel5k تعداد خوشه‌ها در سطح اول از ۵الی ۷ و برای سطح دوم از ۳ الی۵ متغیر خواهند بود و برای دادگان IAPR TC-12 هم تعداد

[1] Sum of Squared Error

خوشه‌هابرای سطح اول و دوم به ترتیب از ۵الی ۱۰ و از ۴الی ۷ متغیر خواهند بود. شکل‌های ۷ و ۸ تاثیر تعداد خوشه‌ها را به ترتیب بر روی مجموع کلمجموع مربعات خطاهای خوشه‌ها و نیز F1 نهایی حاصل از اجرای حاشیه‌نویسی با خوشه‌بندی برای ویژگی شماره ۷ از دادگان Corel5k را نشان می‌دهند. همان‌طور که مشاهده می‌کنید با افزایش تعداد خوشه‌ها، مجموع مربعات خطا کاهش می‌یابد اما F1 کاهش می‌یابد. با توجه به نرخ F1 در شکل۸ تعداد ۵ خوشه برایخوشه‌بندی سطح اول در فضای ویژگی شماره ۷ انتخاب می‌شود.

جدول۴ تفکیک تصاویر پس از خوشه‌بندی بر اساس ویژگی شماره ۷ روی دادگان Corel5k

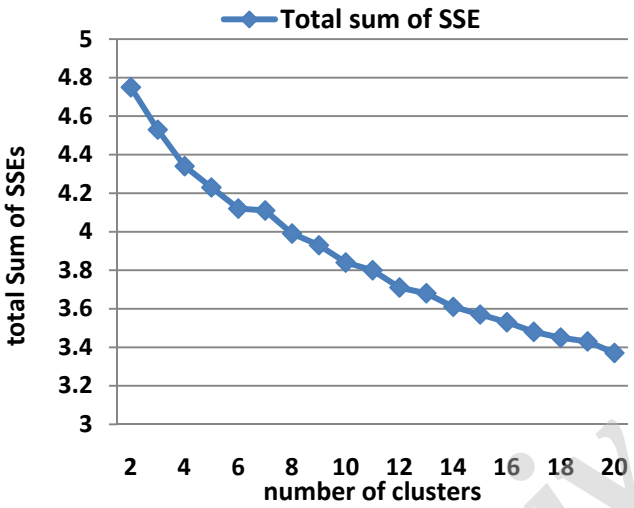
شماره خوشه	تعداد تصاویر آموزش در خوشه	تعداد تصاویر آزمایش در خوشه
۱	۱۹۹۸ (%۴۴,۴)	۲۰۶ (%۴۱,۳)
۲	۵۴۱ (%۱۲)	۶۰ (%۱۲)
۳	۳۱۹ (%۷,۱)	۴۳ (%۸,۶)
۴	۹۸۷ (%۲۱,۹)	۱۱۴ (%۲۲,۹)
۵	۶۵۵(%۱۴,۶)	۷۶ (%۱۵,۲)

جدول۵ تفکیک تصاویر خوشه ۴ ام از جدول۴ پس از خوشه‌بندی بر اساس فضای معنا دادگان Corel5k

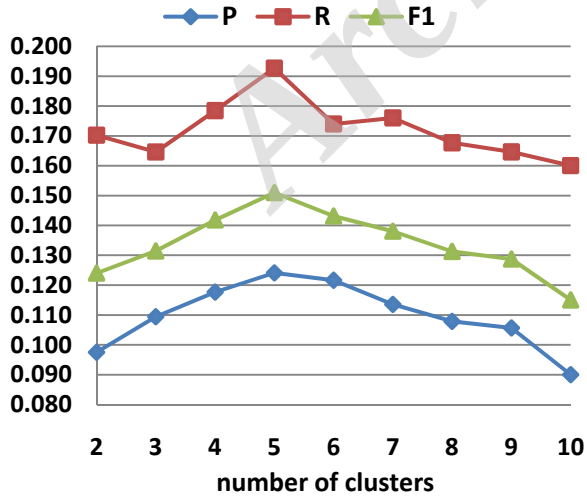
شماره خوشه	تعداد تصاویر آموزش در خوشه	تعداد تصاویر آزمایش در خوشه
۱	۴۱۰ (%۹,۱)	۴۶ (%۹,۳)
۲	۲۵۵ (%۵,۷)	۳۰ (%۶)
۳	۳۲۲ (%۷,۱)	۳۸ (%۷,۶)

برای نشان دادن کارایی خوشه‌بندی دوسطحی مانند روش انتخاب ویژگی ، دو سناریو به کار گرفته شده است. در سناریوی اول خوشه‌بندی صورت نمیگیرد و از تمام تصاویر برای حاشیه‌نویسی تصاویر آزمون انجام می‌شود و در سناریوی دوم خوشه‌بندی دوسطحی انجام می‌شود و جهت حاشیه‌نویسی تصاویر آزمون تنها از تصاویر مرتبط با آنها که در یک خوشه قرار گرفته‌اند، استفاده شده است. شکل۹ نتایج ارزیابی این دو سناریو را روی دو دادگان نشان می‌دهد.

همان‌طور که مشاهده می‌کنید با انجام خوشه‌بندی دوسطحی بیشترین بهبود در بخش فراخوان و به دنبال آن NZR است. علت بهبود فراخوان هم این است که توجه به برچسب‌های با گستردگی کم در دادگان با این خوشه‌بندی‌ها و مخصوصا خوشه‌بندی سطح دوم که بر اساس فضای معناست بیشتر شده است و احتمال اینکه این برچسب‌ها به تصاویر درست منتسب شوند بالاتر می‌رود. به عبارت دیگر خوشه‌بندی دوسطحی باعث می‌شود که تمرکز بر روی کلمات با پراکندگی کم در دادگان درون خوشه‌ها بیشتر شود در حالی که در حالت بدون خوشه‌بندی این کلمات تحت الشعاع کلمات مشهور دادگان قرار می‌گیرند. از طرفی از آنجا که ممکن است تصاویر آزمون در داخل خوشه‌ای که بهترین و مطلوبترین خوشه برای آنهاست قرار نگیرند و بنابراین با تصاویری حاشیه‌نویسی شوند که به آنها مرتبط نیست، در هر دو دادگان بهبود کمتری در نرخ دقت نسبت به فراخوان حاصل شده است. اما در نهایت از آنجا که معیار F1 بهبود داشته است این نتیجه حاصل می‌شود که انجام خوشه‌بندی در بهبود کارایی حاشیه‌نویسی تأثیر گذار است. البته همان‌طور که می‌دانید با خوشه‌بندی زمان اجرا نیز کاهش داده می‌شود چرا که برای حاشیه‌نویسی تصاویر آزمون تنها از تصاویر آموزش مرتبط با آنها و نه تمام تصاویر آموزش استفاده می‌شود بنابراین انجام خوشه‌بندی قابلیت مقیاس پذیری[1] سامانه را تقویت می‌کند.



شکل۷ تاثیر تعداد خوشه‌ها در مجموع کل مجموع مربعات خطاهای خوشه‌بندی برای ویژگی شماره ۷ از دادگان Corel5k



شکل۸ تاثیر تعداد خوشه‌ها در کارایی حاشیه‌نویسی برای ویژگی شماره ۷ از دادگان Corel5k

[1] Scalability

همان‌طور که نتایج دادگان نشان می‌دهد این بخش در افزایش کارایی حاشیه‌نویسی تأثیر گذار است و نیز باعث می‌شود در دادگان‌هایی با نمونه‌های تصویری بیشتر افزایش چشمگیری در سرعت حاشیه‌نویسی ایجاد شود. گرچه اشکال موجود در این گام این است که بعضی از تصاویر جدید در خوشه‌های نادرستی جای می‌گیرند که باعث می‌شود دقت حاشیه‌نویسی این تصاویر پایین بیاید.

شکل۱۰ شامل نمودارهایی است که تاثیر خوشه‌بندی دوسطحی را بر روی تمامی فضاهای ویژگی نشان می‌دهد. هر نمودار این شکل شامل دو میله به رنگ‌های آبی و نارنجی است که میله‌های نارنجی نرخ معیارها را با اجرای خوشه‌بندی و میله‌های آبی نرخ معیارها را بدون خوشه‌بندی نشان می‌دهد. همانگونه که مشاهده می‌کنید نمودارهای نرخ فراخوان و NZR در حالت خوشه‌بندی نسبت به حالت بدون خوشه‌بندی در اکثریت فضاهای ویژگی بهبود قابل ملاحظه‌ای داشته است و این بهبود در نرخ دقت کم‌تر است.

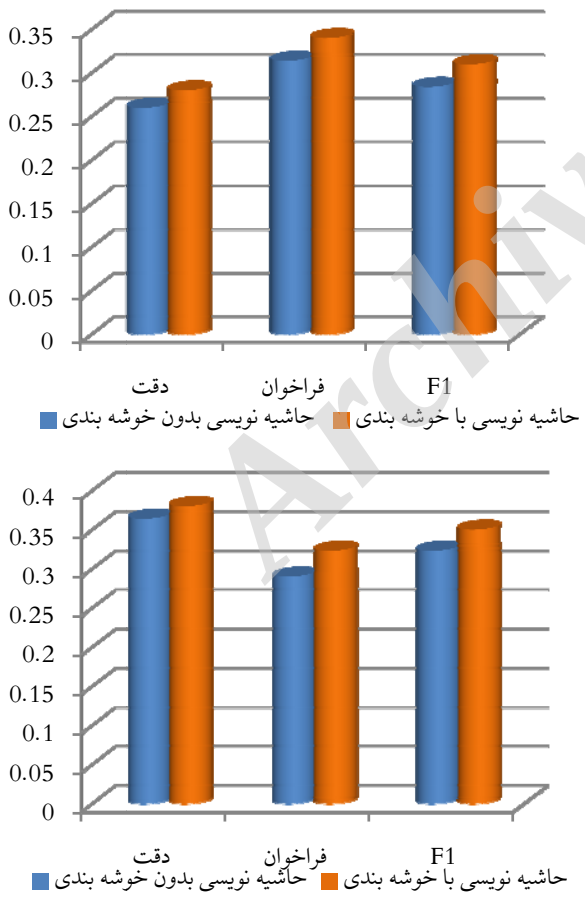
۴-۳ نتایج ارزیابی و مقایسه

✓ دادگان Corel5k

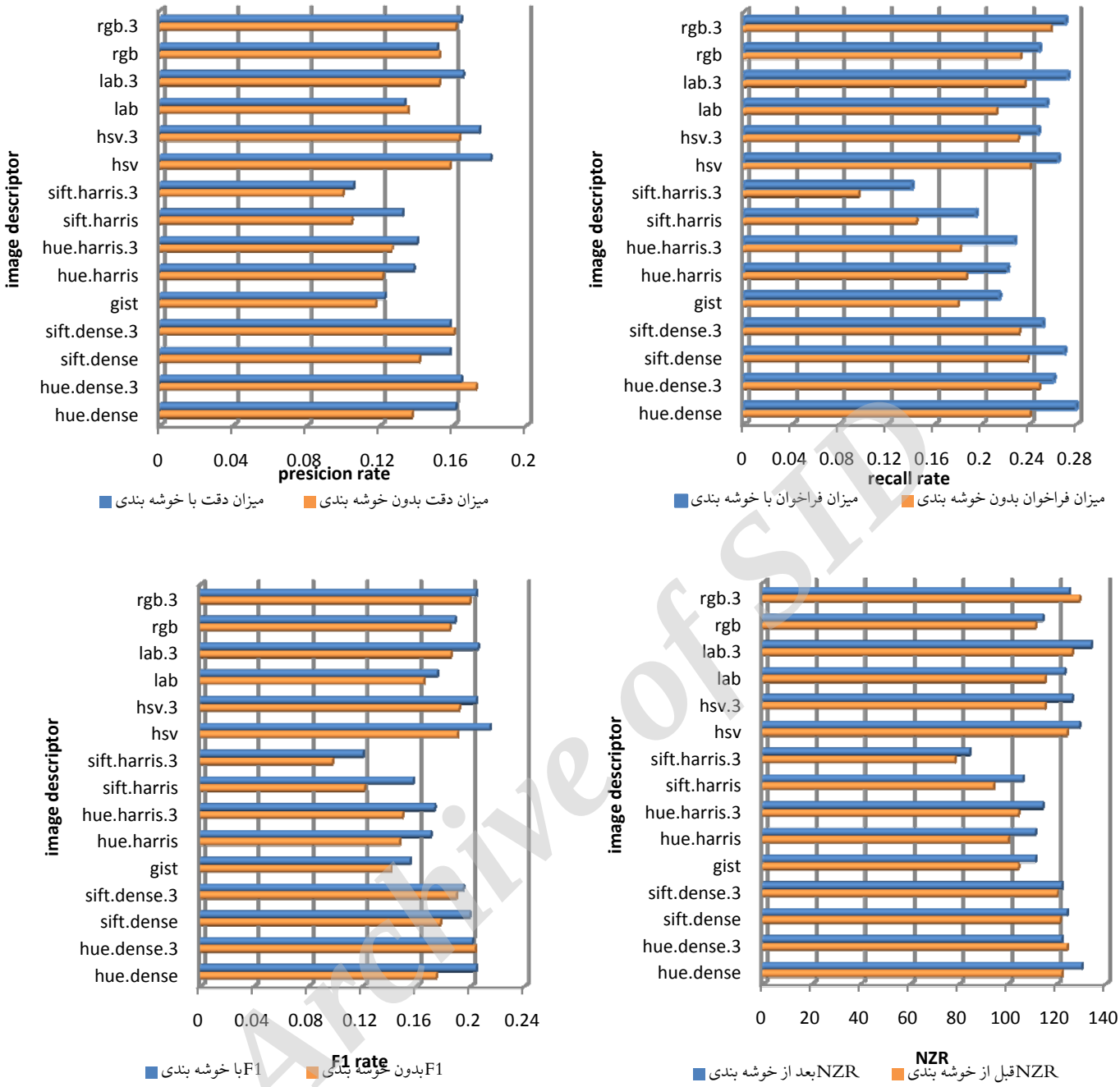
برای ارزیابی و مقایسه روش پیشنهادی با کارهای دیگر، چندین مدل انتخاب شده و نتایج ارزیابی بر اساس معیارهای دقت، فراخوان، F1 و تعداد کلمات با فراخوان غیر صفر (NZR) روی دادگان Corel5k در جدول۶ آمده است. در مقایسه عمکرد روش پیشنهادی با روش‌های حاشیه‌نویسی ارائه شده در مقالات [۱، ۲، ۸، ۹، ۱۱، ۱۲، ۱۷، ۱۸، ۱۹، ۲۰، ۲۱] صرفاً از نتایج گزارش‌شده در این مقالات استفاده شده است.

با توجه به اینکه نرخ دقت و فراخوان را می‌توان به تنهایی بالا برد (بالا بردن یکی در ازای کم شدن دیگری)، معیار F1، گزینه مناسب‌تری جهت مقایسه می‌باشد. اگرچه در روش پیشنهادی از نظر نرخ دقت بهبودی نسبت به بهترین روش بدست نیاورده است، اما از نظر معیار فراخوان عملکرد بهتری را ارائه می‌دهد که این بهبود در فراخوان به علت وجود بخش خوشه‌بندی دوسطحی می‌باشد. از آنجا که در مسئله حاشیه‌نویسی با پیچیدگی و دشواری بالای مسئله از توصیف تصاویر و ویژگی‌های سطح پایین بصری تا نحوه برچسب‌گذاری تصاویر دادگان و ماهیت چندبرچسبی و چندنمونه‌ای (اینکه هر برچسب متعلق به کدام ناحیه از تصویر است نامعلوم است) مواجه هستیم، لذا با بررسی سایر مقالات متوجه می‌شویم که برتری روش‌ها نسبت به یکدیگر تنها در بعضی از معیارها می‌باشد و بهبودهای گزارش شده در مقالات نسبت به هم اندک هستند. بنابراین از نظر معیار F1 بدست آمده، روش پیشنهادی عملکردی هم‌تراز با روش‌های موجود دارد که نشان دهنده‌ی عملکرد قابل قبول روش پیشنهادی در مقایسه با سایر روش‌ها است. روش پیشنهادی نسبت به بعضی مقالات چون[۱، ۲، ۸، ۹، ۱۱، ۱۷، ۱۸، ۱۹، ۲۱] از نظر معیار اصلی که همان F1 می‌باشد برتری دارد و هم‌تراز با مقالات [۱۲، ۲۰] می‌باشد.

شکل۱۱ چند نمونه از تصاویر حاشیه‌نویسی شده از دادگان Corel5k با سامانه ما را نشان می‌دهد. به هر تصویر ۷ برچسب زده شده است که برچسب‌های غلط با خط خوردگی مشخص شده‌اند. همانطور که درتصاویر مشاهده می‌کنید بعضی برچسب‌هایی که سامانه به تصاویر زده است اگرچه صحیح می‌باشند ولی به دلیل حاشیه‌نویسی ضعیف تصاویر در دادگان، این برچسب‌ها به تصویر نخورده است. به عنوان مثال برای تصویر سوم برچسب grass، برچسب صحیحی است چرا که این مفهوم در تصویر وجود دارند اما در دادگان این برچسب برای این تصویر بیان نشده است به همین دلیل انتساب این برچسب به تصویر گفته شده به عنوان برچسب غلط تلقی می‌شود و دقت سامانه را کاهش می‌دهد. همچنین در دادگان مشاهده می‌کنید که کلماتی چون کلمه Hawaii به تصویر برچسب خورده‌اند اما کلماتی از این دست، اسم مکان هستند و متاداده‌ی محتوا- مستقل محسوب می‌شوند که یادگیری این متاداده‌ها برای سامانه کامپیوتری ممکن نیست.



شکل۹ ارزیابی عملکرد خوشه‌بندی روی دادگان، نمودار بالایی مربوط به دادگان Corel5k و نمودار پایینی مربوط به دادگان IAPR TC-12 است.



شکل۱۰ تاثیر خوشه‌بندی بر روی معیارهای ارزیابی حاشیه‌نویسی برای هر توصیفگر تصویری در دادگان Corel5k

✓ دادگان IAPR TC-12

جدول۷ مقایسه روش پیشنهادی با روش‌های دیگر روی دادگان IAPR TC-12 را نشان می‌دهد که ردیف آخر مربوط به عملکرد روش پیشنهادی ما است. در مقایسه عمکرد روش پیشنهادی با روش‌های حاشیه‌نویسی ارائه شده در مقالات [۱، ۲، ۱۱، ۱۲، ۲۲، ۲۳، ۲۴] صرفاً از نتایج گزارش‌شده در این مقالات استفاده شده است. اگرچه از نظر نرخ دقت و فراخوان بهبودی نسبت به بهترین روش بدست نیامده است، اما از نظر معیار F1 بدست آمده، روش پیشنهادی عملکرد خوبی را نسبت به روش‌های موجود دارد که نشان دهنده‌ی عملکرد قابل قبول روش پیشنهادی در مقایسه با روش‌های دیگر است. سرعت بالای حاشیه‌نویسی تصاویر بدون برچسب و عدم پیچیدگی بالای روش پیشنهادی در مقایسه با سایر روش‌های با کارایی مشابه ولی با پیچیدگی بالا، روش پیشنهادی را به روشی مؤثرتر مبدل کرده است. شکل۱۲ چند نمونه از تصاویر حاشیه‌نویسی شده از مجموعه IAPR TC-12 با سامانه ما را نشان می‌دهد. به هر تصویر ۱۰ برچسب زده شده است که برچسب‌های غلط با خط خوردگی مشخص شده‌اند. در دادگان IAPR TC-12 اگرچه تعداد برچسب‌های تصاویر نسبت به تصاویر دادگان Corel5K بیشتر است اما مشکل حاشیه‌نویسی

ضعیف همچنین ادامه دارد و در بعضی موارد اگرچه سامانه برچسب‌های صحیحی به تصاویر زده است ولی این برچسب‌ها در دادگان ظاهر نشده‌اند.

از طرفی از آنجا که در هر دو دادگان تصاویر توسط کاربر انسانی با استانداردهای مختلفی حاشیه‌نویسی شده‌اند حالاتی وجود دارد که تصاویر مختلف توسط کلماتی متفاوت ولی با معانی مشابه برچسب گذاری شوند. برای مثال در دادگان IAPR TC-12 بعضی تصاویر با کلمه grass و بعضی دیگر با کلمه lawn حاشیه‌نویسی شده‌اند در حالیکه این دو برچسب کاملا هم‌معنی هستند و وجود این مسئله تنها پیچیدگی دادگان را افزایش و دقت سامانه را کاهش می‌دهد.

✓ ارزیابی رتبه‌بندی

همان‌طور که گفته شد رتبه‌بندی تصاویر مربوط به یک کلمه، یک خصوصیت حیاتی و لازم برای سامانه‌های بازیابی تصویر است. روش پیشنهادی ما تصاویر مربوط به یک کلمه را رتبه‌بندی می‌کند و تصاویر را بر اساس رتبه‌های آنها به کاربر نشان می‌دهد. برای ارزیابی کیفیت رتبه‌بندی، میانگین دقت متوسط برای روش پیشنهادی روی هر دو دادگان محاسبه شده و در جدول۸ آمده است. برای مقایسه کیفیت رتبه‌بندی روش، مدلهایی که قابلیت رتبه‌بندی دارند و میانگین دقت متوسط را گزارش کرده‌اند، انتخاب شده‌اند. نتایج روش پیشنهادی برای هر دو دادگان Corel5k وIAPR TC-12 آورده شده است.لازم به ذکر است که کیفیت رتبه‌بندی یک بار به ازای کلمات کلیدی موجود در بخش آزمون دادگان (تعداد کلمات کلیدی در بخش تست دادگان Corel5k برابر ۲۶۳ و برای IAPR TC-12 برابر ۲۹۱ است.) و یک بار به ازای کلمات با فراخوان غیر صفر هم (NZR) محاسبه می‌شود. شکل۱۳ نتایج بازیابی چند پرس‌وجو بر روی دادگان IAPR TC-12 را نشان می‌دهد. به ازای هر پرس‌وجو ۵ تصویر با بالاترین رتبه نشان داده شده است.

جدول۶ مقایسه عملکرد روش پیشنهادی با سایر روش‌ها روی دادگانCorel5k

نام مدل	P	R	F1	NZR
PLSA-FUSION-2010-[18]	۰,۱۹	۰,۲۲	۰,۲	۱۱۲
Transductive Multi-Instance Multi Label (TMIML)-2010-[8]	۰,۲۳	۰,۲۷	۰,۲۵	۱۳۰
Joint Equal Contribution (JEC)-2010-[1]	۰,۲۷	۰,۳۲	۰,۲۹	۱۳۹
Penalized Logistic Regression (Lasso)-2010-[1]	۰,۲٤	۰,۲۹	۰,۲٦	۱۲۷
Gussion Mixture Model (GMM)-2011-[19]	۰,۱۵	۰,۱۹	۰,۱۷	۹۳
Gaussian-multinomial PLSA (GM-PLSA)-2011-[17]	۰,۲۵	۰,۲٦	۰,۲۵	۱۲۵
Contextual Kernel and Spectral Methods (CKSM)-2011-[20]	۰,۲۹	۰,۳۵	۰,۳۱	۱٤۷
Hierarchically Discriminative and Generative Model (HDGM)-2012-[9]	۰,۲۹	۰,۳	۰,۳	۱٤٦
CVM-DP-2012-[21]	۰,۳۵	۰,۲۵	۰,۲۹	-
feature fusion and semantic similarity-2014-[2]	۰,۲۷	۰,۳۳	۰,۳	۱٤۱
DWML-kNN -2012-[11]	۰,۲۸	۰,۳۳	۰,۳	۱۳۳
IAGA -2014-[12]	۰,۳	۰,۳۲۷	۰,۳۱	۱۳۲
روش پیشنهادی	۰,۲۸	۰,۳٤	۰,۳۱	۱۳۷

تصویر				
حاشیه‌نویسی توسط انسان	People, restaurant, tables , tree	Bridge, arch, building, stone	Horses, foals, mare, tree	Jet, plane, sky, smoke
حاشیه‌نویسی توسط سامانه	Tree, people, ~~grass~~, tables, ~~cars~~, ~~palace~~, ~~building~~	~~Sky~~, ~~grass~~, bridge, ~~train~~, ~~field~~, stone, building	tree, ~~forest~~, ~~grass~~, mare, horses, ~~field~~, ~~flowers~~	~~Sky~~, ~~water~~, ~~clouds~~, ~~birds~~, jet, plane, smoke

شکل۱۱ کلمات پیش بینی شده توسط سامانه پیشنهادی در مقابل حاشیه‌نویسی توسط انسان برای چند تصویر نمونه در دادگان Corel5k

جدول۷ مقایسه عملکرد روش پیشنهادی با سایر روش‌ها روی دادگان IAPR TC-12

نام مدل	P	R	F1	NZR
Joint Equal Contribution (JEC)-2010-[1]	۰,۲۸	۰,۲۹	۰,۲۸	۲۵۰
Penalized Logistic Regression (Lasso)-2010-[1]	۰,۲۸	۰,۲۹	۰,۲۸	۲۴۶
Photo Annotation through Similar Images (PATSI)-2010-[22]	۰,۲۶	۰,۳۱	۰,۲۸	-
Image Annotation Using Group Sparsity (GS)-2012-[23]	۰,۳۲	۰,۲۹	۰,۳	۲۵۲
MLRank-2013-[24]	۰,۳۸	۰,۳۲	۰,۳۵	۲۵۹
-2014-[2]feature fusion and semantic similarity	۰,۲۹	۰,۲۹	۰,۲۹	۲۵۱
DWML-kNN -2012-[11]	۰,۴	۰,۳۱	۰,۳۵	۲۴۱
IAGA -2014-[12]	۰,۳۹۸	۰,۳	۰,۳۵	۲۴۴
روش پیشنهادی	۰,۳۸	۰,۳۲۳	۰,۳۵	۲۵۰

تصویر				
حاشیه‌نویسی توسط انسان	Balcony, car, church, front, house, people, side, street	Dog, grass, horse, landscape, mountain, people	Boy, cloud, desert, hair, landscape, mountain, rock, sky, stone, summit, tee-shirt, trouser	Bed, bedcover, lamp, night, painting, **room**, table, wall, window
حاشیه‌نویسی توسط سامانه	Building, front, house, palm, people, side, sky, square, street, tree	Group, hill, horse, man , mountain, people, rock, sky, tourist, woman	Desert, front, landscape, man, middle, mountain, people, rock, sky, woman	Bed, chair, curtain, man, room, table, wall, window, woman, wood

شکل۱۲ کلمات پیش بینی شده توسط سامانه پیشنهادی در مقابل حاشیه‌نویسی توسط انسان برای چند تصویر نمونه در دادگان IAPR

جدول۸ میانگین دقت متوسط روی دادگان Corel5k و IAPR TC-12

نام مدل	Corel5k dataset		IAPR TC-12 dataset	
	All 263 words	Words with recall > 0	All 291 words	Words with recall > 0
PLSA-FUSION-2010-[18]	0.26	0.3	-	-
JEC-2010-[1]	0.33	0.52	0.27	0.41
Lasso-2010-[1]	0.3	0.51	0.27	0.39
TMIML-2010-[8]	0.31	0.46	-	-
GM-PLSA-2011-[17]	0.32	0.37	-	-
ML-Rank-2013-[24]	-	-	0.42	-
DWML-kNN -2012-[11]	0.42	0.82	0.65	0.77
روش پیشنهادی	0.39	0.77	0.66	0.78

شکل۱۳ بازیابی معنایی نتایج روی دادگان IAPR TC-12. هر سطر ۵ نتیجه برتر معنایی مطابق با پرس‌وجوی معنایی در سمت چپ‌ترین ستون را نشان می‌دهد

۴-۴ پیچیدگی زمانی الگوریتم

در روش پیشنهاد شده زمانی که تصویر بدون برچسبی برای حاشیه‌نویسی وارد سامانه می‌شود تنها کافی است که ابتدا خوشه‌ی مرتبط با آن پیدا شود و سپس دسته‌بندی kNN بر روی آن تصویر با تصاویر درون خوشه انجام شود و در آخر هم بردارهای امتیاز به دست آمده از تمام فضاها با هم جمع زده شوند و چندین کلمه با بالاترین امتیاز به تصویر برچسب زده شوند.

برای حاشیه‌نویسی تصاویر بدون برچسب ابتدا بر اساس فاصله با مراکز خوشه‌های سطح اول، نزدیک‌ترین خوشه (C1) پیدا می‌شود و سپس، از بین مراکز خوشه‌های سطح دوم درون خوشه C1، مجددا نزدیک‌ترین خوشه (C2) را پیدا می‌شود و از بین تصاویر این خوشه (C2) نزدیکترین همسایه‌ها را انتخاب می‌شوند. اگر فرض کنید تعداد کل تصاویر n، تعداد خوشه‌های سطح اول m1، تعداد تصاویر هر خوشه سطح اول n/m1 ، تعداد خوشه‌های سطح دوم درون خوشه منتخب از سطح اول m2 و تعداد تصاویر هر خوشه سطح دوم درون خوشه منتخب از سطح اول برابر با n/m1*m2 باشد، پیچیدگی زمانی در این حالت از مرتبه O(m1+m2) + O(k*n/m1*m2) = O(k*n/m1*m2) خواهد بود (O(m1+m2) زمان پیدا کردن نزدیک‌ترین خوشه‌ها از سطح اول و دوم و O(k*n/m1*m2) زمان پیدا کردن k همسایه درون خوشه نهایی انتخاب شده در سطح دوم است). زمان حاشیه‌نویسی به تعداد بردارهای ویژگی‌ها، ابعاد آنها و تعداد تصاویر آموزشی رابطه مستقیم دارد که در این پژوهش با اعمال الگوریتم وراثتی برای کاهش ابعاد بردارهای ویژگی و نیز اعمال خوشه‌بندی برای کاهش تعداد تصاویر آموزشی این زمان به صورت قابل ملاحظه‌ای کاهش یافته است.

۵ نتیجه‌گیری

در این مقاله یک روش برای حاشیه‌نویسی خودکار تصاویر مبتنی بر خوشه‌بندی دوسطحی ارائه شد. از الگوریتم وراثتی جهت انتخاب زیر مجموعه ویژگی‌های بهینه در هر بردار ویژگی استفاده شده‌است که نتایج روی هر دو دادگان نشان می‌دهد علاوه بر اینکه عملکرد روش بهبود می‌یابد زمان حاشیه‌نویسی کاهش می‌یابد.

در هر بردار ویژگی، تصاویر در دو سطح، در سطح اول بر اساس ویژگی‌ها و در سطح دوم بر اساس برچسب‌هاخوشه‌بندی شدند. با این کار تصاویر مشابه به هم از لحاظ بصری و نیز تصاویر مرتبط به هم از لحاظ معنایی در یک خوشه جای گرفتند. با این کار در گام بعد، تصاویر بدون برچسب تنها با کمک تصاویری که با هم در یک خوشه مشابه و معنایی قرار گرفته‌اند‌دسته‌بندیمی‌شوند.

در پایان روش پیشنهادی روی دو دادگان شناخته شده پیاده سازی شد. نتایج روی این دو دادگان عملکرد قابل قبول روش پیشنهادی را در مقایسه با دیگر روش‌ها نشان می‌دهد.

مراجع

[1] A. Makadia, V. Pavlovic, and S. Kumar, “Baselines for image annotation,” *Int. J. Comput. Vis.*, vol. 90, no. 1, pp. 88–105, 2010.

[2] X. Zhang and C. Liu, “Image annotation based on feature fusion and semantic similarity,” *Neurocomputing*, 2014.

[3] P. Duygulu, K. Barnard, J. F. de Freitas, and D. A. Forsyth, “Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary,” in *Computer Vision—ECCV 2002*, Springer, pp. 97–112, 2002.

[4] J. Jeon, V. Lavrenko, and R. Manmatha, “Automatic image annotation and retrieval using cross-media relevance models,” in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 119–126, 2003.

[5] V. Lavrenko, R. Manmatha, and J. Jeon, “A model for learning the semantics of pictures,” in *Advances in neural information processing systems*, 2003.

[6] S. L. Feng, R. Manmatha, and V. Lavrenko, “Multiple bernoulli relevance models for image and video annotation,” in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, pp. II–1002, 2004.

[7] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos, “Supervised learning of semantic classes for image annotation and retrieval,” *Pattern Anal. Mach. Intell. IEEE Trans. On*, vol. 29, no. 3, pp. 394–410, 2007.

[8] S. Feng and D. Xu, “Transductive Multi-Instance Multi-Label learning algorithm with application to automatic image annotation,” *Expert Syst. Appl.*, vol. 37, no. 1, pp. 661–670, 2010.

[9] X. Ke, S. Li, and D. Cao, “A two-level model for automatic image annotation,” *Multimed. Tools Appl.*, vol. 61, no. 1, pp. 195–212, 2012.

[10] J. Verbeek, M. Guillaumin, T. Mensink, and C. Schmid, “Image annotation with tagprop on the MIRFLICKR set,” in *Proceedings of the international conference on Multimedia information retrieval*, pp. 537–546, 2010.

[11] عبدالله زاده مکی، داوود،**حاشیه نویسی خودکار تصاویر با ترکیب ویژگی‌های سراسری و ناحیه ای و استفاده از همبستگی کلمات**،پایان نامه کارشناسی ارشد مهندسی کامپیوتر، دانشگاه تربیت مدرس، دانشکده مهندسی برق و کامپیوتر، تهران، ۱۳۹۱، ص ۳۹ الی ۴۴.

[12] S. Bahrami, M. Saniee Abadeh, "Automatic image annotation using an evolutionary algorithm (IAGA)." In *Telecommunications (IST), 2014 7th International Symposium on*, pp. 320-325. IEEE, 2014.

[13] R. Li, J. Lu, Y. Zhang, and T. Zhao, “Dynamic Adaboost learning with feature selection based on parallel

genetic algorithm for image annotation," *Knowl.-Based Syst.*, vol. 23, no. 3, pp. 195–201, 2010.

[14] L. Setia and H. Burkhardt, "Feature selection for automatic image annotation," in *Pattern Recognition*, Springer, pp. 294–303, 2006.

[15] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1, pp. 273–324, 1997.

[16] M. Grubinger, P. Clough, H. Müller, and T. Deselaers, "The iapr tc-12 benchmark: A new evaluation resource for visual information systems," in *International Workshop OntoImage*, pp. 13–23, 2006.

[17] Z. Li, Z. Shi, X. Liu, and Z. Shi, "Modeling continuous visual features for semantic image annotation and retrieval," *Pattern Recognit. Lett.*, vol. 32, no. 3, pp. 516–523, 2011.

[18] Z. Li, Z. Shi, X. Liu, Z. Li, and Z. Shi, "Fusing semantic aspects for image annotation and retrieval," *J. Vis. Commun. Image Represent.*, vol. 21, no. 8, pp. 798–805, 2010.

[19] F. Shi, J. Wang, and Z. Wang, "Region-based supervised annotation for semantic image retrieval," *AEU-Int. J. Electron. Commun.*, vol. 65, no. 11, pp. 929–936, 2011.

[20] Z. Lu, H. H. Ip, and Y. Peng, "Contextual kernel and spectral methods for learning the semantics of images," *Image Process. IEEE Trans. On*, vol. 20, no. 6, pp. 1739–1750, 2011.

[21] M. Wang, F. Li, and M. Wang, "Collaborative visual modeling for automatic image annotation via sparse model coding," *Neurocomputing*, vol. 95, pp. 22–28, 2012.

[22] M. Stanek, B. Broda, and H. Kwasnicka, "Patsi—Photo annotation through finding similar images with multivariate Gaussian models," in *Computer Vision and Graphics*, Springer, pp. 284–291, 2010.

[23] S. Zhang, J. Huang, H. Li, and D. N. Metaxas, "Automatic image annotation and retrieval using group sparsity," *Syst. Man Cybern. Part B Cybern. IEEE Trans. On*, vol. 42, no. 3, pp. 838–849, 2012.

[24] Z. Li, J. Liu, C. Xu, and H. Lu, "MLRank: Multi-correlation Learning to Rank for image annotation," *Pattern Recognit.*, vol. 46, no. 10, pp. 2700–2710, 2013.

سمانه بهرامی مدرک کارشناسی ارشد خود را در رشته مهندسی کامپیوتر از دانشگاه تربیت مدرس در سال ۱۳۹۳ دریافت کرد. زمینه‌های پژوهشی مورد علاقه او، پردازش تصویر، الگوریتم‌های تکاملی و داده‌کاوی می باشد.

محمد صنیعی آباده مدرک کارشناسی خود را در سال ۱۳۸۰ در رشته مهندسی کامپیوتر (نرم افزار) از دانشگاه صنعتی اصفهان دریافت کرد. وی کارشناسی ارشد و دکتری خود را در گرایش هوش مصنوعی و روباتیک به ترتیب در دانشگاههای علم و صنعت ایران (۱۳۸۱) و صنعتی شریف (۱۳۸۶) اخذ نمود. دکتر صنیعی از سال ۱۳۸۸ عضو هیآت علمی دانشگاه تربیت مدرس می باشند. ایشان عضویت در بسیاری از مجامع معتبر علمی را در کارنامه خود دارند. برخی از این مجامع عبارتند از: بنیاد ملی نخبگان، کمیته داوران اولین و دومین دوره مسابقات جشنواره بین‌المللی روباتیک خوارزمی، کمیته علمی طراحی و تدوین برنامه جامع آموزشی رشته مهندسی فناوری اطلاعات در مرکز تحقیقات مخابرات ایران، کمیته علمی انجمن سیستمهای فازی ایران، کمیته اجرایی سمپوزیوم CNDS 2012 . ایشان برگزیده اولین نمایشگاه اختراعات و نوآوری‌های ایران در جشنواره ملی نوآوری و شکوفائی. دکتر صنیعی مولف دو کتاب، بیش از ۳۰ مقاله در مجلات معتبر و بالغ بر ۵۰ مقاله در کنفرانسهای شاخص خارجی و داخلی هستند. زمینه های تخصصی ایشان عبارتند از: داده‌کاوی در پزشکی، بیوانفورماتیک، بانک، بیمه، بورس، امنیت و غیره، متن‌کاوی، وب‌کاوی، نظرکاوی و تحلیل محتوایی متن، کاوش داده‌های عظیم، تصویرکاوی پزشکی، کاربرد روش‌های فرامکاشفه‌ای در کشف دانش و سیستم‌های فازی تکاملی.