علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۵، شماره ۲، ۱۳۹۶
صفحات ۱۳- ۲۶
مقاله عادی

نشریه علوم رایانش و فناوری اطلاعات

ارائه روشی متنی بر محتوا و جذابیت برای خلاصه‌سازی مجموعه تصاویر اجتماعی

زهرا ریاحی سامانی محسن ابراهیمی مقدم

دانشکده مهندسی و علوم کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران

چکیده

روش‌های خلاصه‌سازی تصویر امروزه به ابزاری مهم برای استفاده از مجموعه‌های تصویری بزرگ تبدیل شده‌اند. این روش‌ها مجموعه کوچکی از تصاویر را به‌عنوان نماینده از یک مجموعه بزرگ‌تر انتخاب می‌کنند. مطالعات جدید نشان داده‌اند که تصاویری که در روش‌های خلاصه‌سازی تصویر انتخاب می‌شوند باید از یک سو پوشش اطلاعاتی مناسبی از مجموعه اولیه فراهم کنند و از سوی دیگر معیارهایی مانند کیفیت و زیبایی تصاویر را در انتخاب تصاویر خلاصه مد نظر قرار دهند. با علم به این موارد، در این مقاله روشی خودکار برای خلاصه‌سازی مجموعه‌های تصویری با به کارگیری و مطالعه دو دسته ویژگی با عنوان ویژگی‌های محتوایی و ویژگی‌های جذابیت ارائه شده است. در روش پیشنهادی دو دسته ویژگی معرفی شده به‌طور خودکار و با استفاده از بستر شبکه‌های اجتماعی و هستان‌شناسی دامنه استخراج شده و تاثیرگذاری آنها در خلاصه‌سازی تصاویر با استفاده از بازخورد کاربران مدل می‌شود. خلاصه نهایی با ترکیب خلاصه ساخته شده به‌وسیله دو دسته ویژگی به‌دست آورده می‌شود. نتایج روش پیشنهادی با تصاویر خلاصه‌ای که توسط گروهی از کاربران بر روی مجموعه‌ای از تصاویر انتخابی از فلیکر انجام شده است مقایسه شده است. آزمایش‌ها نشان می‌دهد که روش پیشنهادی در مقایسه با روش‌های جاری خلاصه‌سازی تصویر توانسته است به بهبود سی و پنج درصدی پوشش اطلاعاتی در خلاصه کردن مجموعه تصویری به میزان یک درصد حجم اولیه و بهبود چهار درصدی پوشش اطلاعاتی در خلاصه کردن مجموعه تصویری به میزان پنج درصد حجم اولیه دست یابد. همچنین روش پیشنهادی توانسته است به بهبود سی و نه درصدی در ساختن خلاصه‌های شبیه خلاصه‌های دستی دست یابد.

کلمات کلیدی: تصاویر اجتماعی، خلاصه‌سازی تصویر، بازخورد کاربر.

۱- مقدمه

با ظهور دوربین‌های دیجیتال و تلفن‌های هوشمند اندازه مجموعه‌های تصویری رو به افزایش است. مطالعات نشان می‌دهد که روزانه ۵۰۰ میلیون تصویر در اینترنت بارگذاری می‌شود (فقط با در نظر گرفتن Flickr، Facebook، Instagram وSnapchat) [۱]. این مسئله موجب افزایش تقاضا برای روش‌های خلاصه‌سازی تصویر شده است که به کاربرها در مرور و پیدا کردن مجموعه‌های تصویری کمک می‌کنند.

تصاویر اجتماعی یا تصاویر بارگذاری شده در شبکه‌های اجتماعی و سایت‌های اشتراک تصویر معمولاً با داده‌های جانبی همراه هستند. به عنوان مثال اگر کاربری آنها را جذاب ببیند می‌تواند آنها را نشانه کند و یا بر روی آنها یادداشت بنویسد. این اطلاعات جانبی و همچنین اطلاعات جانبی دیگر که در دوربین‌های رقمی موجود هستند امروزه در سیستم‌های چندرسانه‌ای استفاده می‌شوند.

هدف یک سیستم خلاصه‌سازی مجموعه تصویر آن است که با دریافت یک مجموعه بزرگ از تصاویر بتواند مجموعه کوچک‌تری از تصاویر را از مجموعه اولیه به عنوان تصاویر نماینده یا تصاویر خلاصه انتخاب کند، به گونه‌ای که مجموعه انتخابی را بتوان جوهره و یا چکیده‌ای از مجموعه اولیه دانست [۲].

روش‌های اولیه در خلاصه‌سازی تصویر معیارهای پوشش[۱] و تکرار[۲] را دنبال می‌کردند، به این ترتیب که تصاویر مجموعه خلاصه باید تا حد ممکن شبیه مجموعه اولیه بوده و تا حد ممکن ناشبیه به همدیگر باشند [۳]. روش‌های جدیدتر این ایده را مطرح می‌کنند که معیارهای دیگری مانند کیفیت یا جذابیت هم برای انجام عمل خلاصه‌سازی مهم هستند. رادیناک و همکاران [۴] یک آزمایش بررسی جمعی انجام داده‌اند و در آن به تحلیل اینکه کاربر انسانی چگونه خلاصه‌سازی تصویر را انجام می‌دهد پرداخته‌اند. نتایج این مطالعه نشان می‌دهد که تصاویری که

به عنوان خلاصه انتخاب می‌شوند از نظر معنایی شبیه تصاویر دیگر هستند. از طرف دیگر، این نتایج نشان می‌دهد که معیارهای دیگری مانند کیفیت و زیبایی تصاویر نیز در خلاصه‌سازی مهم هستند [۴].

در این مقاله به معرفی و استخراج خودکار دو دسته ویژگی برای خلاصه‌سازی تصویر پرداخته‌ایم. دو دسته ویژگی مطرح شده شامل ویژگی‌های محتوایی و ویژگی‌های جذابیت هستند.

هدف ویژگی‌های محتوایی پیدا کردن تصاویری از مجموعه اولیه است که در انتها بتوانند پوشش اطلاعاتی مجموعه اولیه را در تعداد کمتر تصویر فراهم کنند. رویکردی که ما در این مقاله برای ویژگی‌های محتوایی به کار گرفته‌ایم رویکردی معنایی است و علاوه بر شباهت بصری تصاویر، مفاهیمی که در آنها وجود دارند نیز در نظر گرفته شده است. روش‌هایی که تا به امروز بر روی پوشش اطلاعاتی تمرکز داشتند به طور عمده بهینه‌سازی معیارهای تصویری یا عددی را دنبال می‌کردند [۵]. این روش‌ها معمولاً شباهت بصری را در نظر می‌گیرند. به تازگی روش‌هایی مطرح شده‌اند که به مدل کردن اطلاعات معنایی مجموعه تصویر پرداخته‌اند [۲]. در این مقاله بهبودی بر روی مدل معنایی [۲] ارائه شده است. روش ارائه شده در [۲] بر مبنای استفاده از هستان‌شناسی[۳] دامنه و استخراج ویژگی‌هایی تحت عنوان ویژگی‌های هستان‌شناسی است. جزئیات این ویژگی‌ها در قسمت ۲-۳ بحث خواهد شد. در این مقاله ویژگی‌های هستان‌شناسی معرفی شده در [۲] بهبود داده شده‌اند و همچنین دسته دومی از ویژگی‌ها تحت عنوان ویژگی‌های جذابیت برای خلاصه‌سازی تصویر معرفی و به کار گرفته شده‌اند.

ویژگی‌های جذابیت شامل کیفیت تصویر، احساس، اثر بخشی و زیبایی تصویر هستند. برای استخراج ویژگی‌های جذابیت، در این مقاله پیشنهاد شده است که از اطلاعات موجود بر روی تصاویر بارگذاری شده در شبکه‌های اجتماعی استفاده شود. شبکه‌های اجتماعی به کاربران خود اجازه می‌دهند که رفتارهای اجتماعی بر روی تصاویر انجام دهند. به عنوان مثال اگر یک تصویر برای یک کاربر جذاب باشد می‌تواند آن را پسند[۴] کند، بر روی آن یادداشت[۵] بنویسد و یا مجدداً به اشتراک[۶] بگذارد. مطالعات جدید نشان می‌دهد که این فراداده‌ها می‌توانند برای نشان دادن جذابیت تصویر به کار گرفته شوند [۶]. در این مقاله از این بستر برای پیدا کردن تصاویر جذاب استفاده شده است. جزئیات این ویژگی‌ها در قسمت ۱-۳ بحث خواهد شد.

همچنان که در بخش‌های آتی توضیح داده خواهد شد ویژگی‌های محتوایی و جذابیت دارای ابعاد بالا بوده و به محتوای تصاویر حساس هستند. ارائه یک روش یادگیری با حضور این بعد بالای ویژگی، نیازمند تعداد زیادی داده آموزشی است. بنابراین در این مقاله پیشنهاد شده است که خلاصه‌سازی را با استفاده از ویژگی‌های محتوا و جذابیت به طور جداگانه انجام داده و سپس خلاصه نهایی با ترکیب دو خلاصه ساخته شده به دست آورده شود. بدین ترتیب سه پیمانه برای خلاصه‌سازی تصویر تعریف شده است. پیمانه اول پیمانه محتوایی نام دارد و امتیازی برای میزان پوشش محتوایی هر تصویر - از مجموعه اولیه - استخراج می‌کند. پیمانه دوم پیمانه جذابیت نام دارد و برای هر تصویر امتیازی برای میزان جذابیت آن تصویر استخراج می‌کند. سومین و آخرین پیمانه، پیمانه تجمیع و خلاصه‌سازی است که تصاویر خلاصه را با بررسی اطلاعات هر دوی پیمانه‌های محتوایی و جذابیت تولید می‌کند.

مسئله بعدی انتخاب روش ترکیب نتایج پیمانه‌های محتوایی و جذابیت است. در این مقاله از روش جمع وزن‌دار استفاده شده است. انتخاب وزن‌ها با استفاده از یک آزمایش بازخورد از کاربر و در یک مرحله آموزش انجام شده است. در واقع وزن‌ها نشان‌دهنده این مسئله هستند که هر کدام از دو گروه ویژگی معرفی شده چه میزان در تولید خلاصه تاثیرگذار هستند. جزئیات آزمایش بازخورد کاربر در قسمت ۳-۳ توضیح داده شده است.

برای ارزیابی روش پیشنهادی یک مجموعه خلاصه انسانی برای مجموعه‌ای از تصاویر استخراجی از فلیکر در دامنه شهرها و مناطق ایجاد شده است. نتایج خلاصه‌سازی با نتایج کاربر انسانی مقایسه شده است. نتایج نشان می‌دهد که روش پیشنهادی نتایج بهتری در مقایسه با روش‌های دیگر در حوزه خلاصه‌سازی تصویر و همچنین در مقایسه با روش‌های تک معیاره دارد.

نوآوری‌های این مقاله را به صورت زیر می‌توان نام برد:

۱- یک روش خودکار برای خلاصه‌سازی تصاویر اجتماعی ارائه شده است که محتوا و ویژگی‌های جذابیت را هم‌زمان در نظر می‌گیرد.

۲- دو مجموعه قابل تفسیر از ویژگی برای خلاصه‌سازی پیشنهاد شده و میزان اثر بخشی آنها در خلاصه‌سازی تصویر با استفاده از بازخورد کاربران به دست آورده شده است.

۳- روش‌هایی خودکار برای استخراج ویژگی‌های معرفی شده به کار گرفته شده و روش‌های موجود در این حوزه بهبود داده شده‌اند.

ساختار مقاله در ادامه به صورت زیر است. در قسمت بعدی در مورد کارهای مرتبط در حوزه خلاصه‌سازی تصویر صحبت خواهد شد. سپس به معرفی روش پیشنهادی پرداخته و پس از آن نتایج روش پیشنهادی ارائه شده و با کارهای مرتبط مقایسه می‌شود.

۲- کارهای مرتبط

در قسمت قبلی مقاله به خطوط اصلی تحقیقات در حوزه خلاصه‌سازی مجموعه تصاویر اشاره کردیم. در این بخش به بررسی جزئی‌تر روش‌های مطرح در حوزه خلاصه‌سازی تصویر می‌پردازیم.

عمدتاً روش‌های مطرح در خلاصه‌سازی مجموعه تصاویر رویکرد بدون نظارت دارند. به تازگی روش‌هایی با رویکرد با نظارت نیز مطرح شده‌اند [۱]. رویکردهای با نظارت معمولاً مبتنی بر روش‌های یادگیری‌ای هستند که تلاش می‌کنند شیوه خلاصه‌سازی دستی که به وسیله انسان انجام می‌شود را یاد بگیرند. این گونه روش‌ها معمولاً در مرتبه تصاویر چند ده تایی قابل به‌کارگیری هستند، اما در مقابل، روش‌های بدون نظارت بر روی تعداد چند صد و یا چند هزار تصویر تمرکز دارند. از بعد دیگر مقایسه، می‌توان روش‌های خلاصه‌سازی تصویر را به دو دسته عمده تقسیم کرد. روش‌های خلاصه‌سازی مبتنی بر ویژگی‌های تصویری و روش‌های خلاصه‌سازی چند حالته[۷].

روش‌های خلاصه‌سازی تصویری، تنها از ویژگی‌های تصویری استفاده می‌کنند. سیستم‌های خلاصه‌سازی چند حالته، از اطلاعات جانبی تصاویر نیز برای انجام عمل خلاصه‌سازی استفاده می‌کنند. تصاویر موجود بر روی وب و یا تصاویر به اشتراک‌گذاری شده بر روی شبکه‌های اجتماعی دارای داده‌های جانبی هستند. به عنوان مثال، کاربران می‌توانند بر روی آنها داده‌های متنی یادداشت کنند و یا اگر کاربری تصویری را جذاب ببیند می‌تواند آن را به اشتراک بگذارد و یا می‌تواند تصویر را جزء تصاویر مورد علاقه‌اش قرار دهد. همچنین دوربین‌های دیجیتال، امروزه اطلاعاتی از زمان و مکان تصویر را ثبت می‌کنند. روش‌های متعددی در ادبیات موضوع مطرح شده است که از این اطلاعات جانبی استفاده می‌کنند [۴، ۷، ۸]. اکثر این روش‌ها در حضور داده‌های جانبی معمولاً موفق عمل می‌کنند اما این داده‌ها همیشه در دسترس نیستند. در ادامه در ابتدا به شرح روش‌های خلاصه‌سازی تصویری خواهیم پرداخت و پس از آن روش‌های خلاصه‌سازی چند حالته مطرح می‌شوند.

یکی از مطرح‌ترین روش‌های خلاصه‌سازی تصویری روش [۳] است. نویسندگان در این مقاله یک نسخه حریصانه از الگوریتم k-means را برای خلاصه‌سازی مجموعه تصاویر ارائه داده‌اند. الگوریتم پیشنهادی به طور تکراری

تصاویری را برای خلاصه انتخاب می‌کند که شباهت مجموعه انتخابی با مجموعه اولیه را بیشینه و شباهت میان مجموعه انتخابی را کمینه کند. روش پیشنهادی در [۵] هم بهینه‌سازی مشابهی را دنبال می‌کند. این روش بر این مبنا استوار است که تصاویر مجموعه اولیه را می‌توان با ترکیب خطی و تنک ویژگی‌های محلی تصاویر خلاصه بازسازی کرد. تابع هدف [۵] به صورت زیر تعریف می‌شود:

$$\min_{D,A} \sum_i \|x_i - D\alpha_i\|^2 + \lambda \sum_i \|\alpha_i\| + \beta \sum_{j \neq k} \mathrm{corr}(d_k, d_j) \qquad (۱)$$

که در آن مولفه اول اختلاف تصاویر اولیه با تصاویر بازسازی شده به وسیله ترکیب خطی تصاویر خلاصه است، مولفه دوم شرط تنک بودن را بر آورده می‌کند و مولفه سوم شباهت ویژگی‌های تصاویر خلاصه به یکدیگر را کمینه می‌کند. این روش از یک نسخه از الگوریتم شبیه‌سازی حرارت برای بهینه‌سازی تابع هدف استفاده می‌کند.

الگوریتم‌های خوشه‌بندی نیز برای خلاصه‌سازی مجموعه‌های تصویری به طور گسترده‌ای به کار گرفته شده‌اند. این روش‌ها مبتنی بر خوشه‌بندی مجموعه تصاویر و سپس انتخاب تصاویر نماینده از هر خوشه هستند. روش‌های خوشه‌بندی متفاوتی در ادبیات موضوعی برای این کار به کار گرفته شده‌اند: مانند الگوریتم انتشار وابستگی [۹] و خوشه‌بندی طیفی [۱۰]. روش‌های خلاصه‌سازی مبتنی بر خوشه‌بندی تصویر، ویژگی‌های مختلفی مانند ویژگی‌های محلی [۱۰]، سراسری [۹، ۱۱] و یا ترکیبی از هر دوی آنها [۱۲، ۱۳] از تصاویر استخراج می‌کنند.

روش‌های مرکزیت گراف نیز برای خلاصه‌سازی تصاویر استفاده شده‌اند [۱۴-۱۶]. یکی از مطرح‌ترین آن‌ها روش مطرح شده در [۱۴] است. در این روش از الگوریتم PageRank بر روی گراف مشابهت تصویر برای خلاصه‌سازی تصویر استفاده شده است. در گراف مشابهت نودها تصاویر هستند و شباهت بین آنها وزن یال گراف را مشخص می‌کند و الگوریتم PageRank مبتنی بر مرکزیت مقدار ویژه عمل می‌کند. مرکزیت مقدار ویژه مرکزیتی است که تعداد همسایه‌های یک راس و همچنین اهمیت همسایه‌های آن راس را به عنوان معیاری از اهمیت آن در نظر می‌گیرد و مبتنی بر این اصل است که راسی که به یک راس مهم نزدیک‌تر است اهمیت بالاتری در مقایسه با راس‌های دورتر دارد. این مرکزیت در مقابل مرکزیت درجه مطرح می‌شود که در آن اهمیت هر راس به عنوان تابعی از تعداد یال‌های آن در نظر گرفته می‌شود. اگر A_{ij} را به صورت شباهت بین دو راس i و j (معادله ۲) و e_i را به عنوان مرکزیت بردار ویژه‌ی مجموعه تصاویر تعریف شود e_i، به صورت معادله‌ای تعریف می‌شود که مرکزیت درجه همسایه‌های v را اندازه می‌گیرد [۱۷].

$$A_{i,j} = \mathrm{Sim}_{ij} \qquad (۲)$$

$$e_i = \frac{1}{\lambda} \sum_j A_{ij} e_j \qquad (۳)$$

در این رابطه، λ فاکتور نرمال‌سازی است و باعث می‌شود که جمع e_iها یک شود ($\sum_i e_i = 1$). با بازنویسی متغیرها به فرم برداری $E= (e_1, e_2, \ldots, e_n)^T$ معادله ۳ به صورت معادله ۴ بازنویسی می‌شود [۱۷]:

$$E = \frac{1}{\lambda} AE \qquad (۴)$$

$$AE = \lambda E \qquad (۵)$$

که در آن E به صورت بردار ویژگی ماتریس مجاورت A تعریف می‌شود. بنابراین مساله‌ی پیدا کردن تعدادی تصویر به عنوان نماینده از مجموعه اولیه تصاویر به مسئله پیدا کردن بردار ویژه ماتریس مجاورت تبدیل شده است [۱۷]. ریاحی و ابراهیمی مقدم در [۲] یک نسخه معنایی از گراف مشابهت را ساخته و از الگوریتم‌های مرکزیت گراف برای پیدا کردن تصاویر نماینده استفاده کرده‌اند. هم چنین در روش پیشنهادی در [۱۶] یک نسخه بهبود داده شده از الگوریتم PageRank پیشنهاد داده شده است. در این روش پس از استخراج هر کدام از تصاویر خلاصه، گراف اولیه براساس تفاوت با تصویر انتخابی وزن‌دهی مجدد می‌شود. به این ترتیب تصاویر خلاصه متنوع‌تر می‌شوند.

دسته دوم کارهای مرتبط، روش‌های خلاصه‌سازی تصویر چند حالته هستند. تصاویر اجتماعی و یا تصاویر بارگذاری شده بر روی وب دارای اطلاعات جانبی با محتوای متنی، زمانی، رفتاری یا مکانی هستند. روش‌های بسیاری در ادبیات موضوعی مطرح شده است که از یکی [۱۲] و یا ترکیبی [۴، ۷، ۱۸] از این اطلاعات جانبی برای خلاصه‌سازی تصویر استفاده می‌کنند. اکثر این روش‌ها در حضور این داده‌های جانبی موفق عمل می‌کنند.

اولین داده جانبی که در تحلیل‌ها استفاده شده است داده مکانی است. کندی و همکارن [۱۹] از داده‌های مکانی و متنی برای پیدا کردن تصاویر نماینده استفاده می‌کنند. به این صورت که ابتدا برچسب‌های متنی و اطلاعات مکانی را تحلیل کرده و برچسب‌های نماینده را به دست می‌آورند و سپس در مرحله بعد تصاویر مربوط به برچسب‌های نماینده انتخاب می‌شوند. جف و همکاران [۱۸] از ترکیبی از داده‌های متنی، مکانی، زمانی و اجتماعی و الگوریتم‌های خوشه‌بندی برای خلاصه‌سازی تصویر استفاده می‌کنند.

داده جانبی دیگری که به طور وسیعی به کاربرده شده است داده متنی است. پانگ و همکاران روشی را برای خلاصه‌سازی تصاویر استفاده کرده‌اند که در آن با روش‌های تحلیل متن همراه تصویر ابتدا عنوان‌های متنی نماینده را انتخاب کرده و سپس تصاویر مرتبط با آن عنوان‌ها را انتخاب می‌کنند [۲۰]. کامارگو و همکاران از یک روش فاکتورگیری نامنفی ماتریس برای پیدا کردن عنوان‌های نهفته در فضای متنی استفاده می‌کنند و سپس نتایج این فاکتورگیری را برای انجام فاکتورگیری نامنفی ماتریس در فضای تصویر به کار می‌برند [۸].

رفتارهای اجتماعی کاربران منبع دیگری از اطلاعات است که می‌توان از آن استفاده کرد. به عنوان مثال جیانگ و همکاران از ویژگی‌های تصاویری که روی شبکه‌های اجتماعی دارای یادداشت هستند برای خلاصه‌سازی مجموعه تصویر استفاده کرده‌اند [۲۱].

به تازگی روش‌های چند معیاره و با نظارت نیز معرفی شده‌اند. این روش‌ها مبتنی بر یادگیری خلاصه‌سازی از روی خلاصه‌سازی دستی انسان هستند. به عنوان مثال شیاتشک و همکاران [۱] روشی بانظارت برای خلاصه‌سازی تصاویر ارائه داده‌اند که بر مبنای بهینه‌سازی معیاری با نام V-rouge است که به صورت فاصله بین خلاصه دستی و اتوماتیک تعریف می‌شود.

رودیناک و همکاران [۴] یکی دیگر از روش‌های با نظارت و چند حالته را ارائه داده‌اند. روش پیشنهادی آن‌ها چند حالته است به این دلیل که از منابع اطلاعاتی متفاوتی مانند داده‌های متنی، مکانی و زمانی استفاده می‌کند و چند معیاره است به این دلیل که هم محتوا و هم ویژگی‌های زیبایی تصویر را در نظر می‌گیرد. آن‌ها از ماشین بردار پشتیبان برای رتبه‌بندی تصاویر استفاده کرده‌اند.

شن و همکاران [۷] روش دیگری ارائه داده‌اند که از مجموعه‌ای از داده‌های مکانی، زمانی و معنایی برای انجام خلاصه‌سازی استفاده می‌کند. آن‌ها از مدل مخلوط گاوسی برای انجام خوشه‌بندی استفاده می‌کنند و تصاویر را در تعدادی رویداد خوشه‌بندی می‌کنند. سپس از یک الگوریتم رتبه‌بندی برای رتبه‌بندی تصاویر استفاده می‌کنند.

جدول ۱ مقایسه‌ای بر روی روش‌های جدیدتر خلاصه‌سازی تصویر را نشان می‌دهد. این روش‌ها در نوع ورودی، میزان نظارت، رهیافت و روش خلاصه‌سازی متفاوت هستند در این مقاله روشی برای خلاصه‌سازی مجموعه تصویر ارائه می‌شود

که هم ارتباطات معنایی را برای خلاصه‌سازی در نظر می‌گیرد و هم پارامترهای کیفیت و جذابیت را لحاظ می‌کند.

۳- روش پیشنهادی

همان‌طور که در قسمت قبل عنوان شد، بیشتر روش‌های خلاصه‌سازی تصویر بر روی پوشش اطلاعاتی تمرکز داشته‌اند و روش‌های اندکی در ادبیات موضوعی وجود دارند که معیارهای دیگری مانند کیفیت و جذابیت را نیز در نظر گرفته‌اند.

در روش پیشنهادی از دو دسته ویژگی استفاده شده است. دسته اول ویژگی‌های محتوایی هستند که به صورت ویژگی‌های سطح بالا بر روی ویژگی‌های تصویری سطح پایین تصویر تعریف می‌شوند. هدف ویژگی‌های محتوایی این است که اطلاعات موجود در تصاویر را مدل کنند.

جدول ۱- مقایسه روش‌های خلاصه‌سازی تصویر

نام روش	حالت یا منبع اطلاعات	میزان نظارت	رهیافت
می [۲۲]	ویدئو	بدون نظارت	بهینه‌سازی معیار عددی
ناسی [۲۳]	ویدئو	بدون نظارت	آنالیز گراف
کامارو [۸]	تصویر+متن	بدون نظارت	بهینه‌سازی معیار عددی
شن [۷]	تصویر+مکان+زمان	بدون نظارت	خوشه‌بندی
ریاحی [۲]	تصویر	بدون نظارت	تحلیل گراف و روابط معنایی
فانگ [۲۴]	تصویر	با نظارت	یادگیری TOPIC ASPECTها
شیاتشک [۱]	تصویر	با نظارت	یادگیری توابع SUB-MODULAR
روش پیشنهادی	تصویر	بدون نظارت	چند معیاره

دسته دوم ویژگی‌ها، ویژگی‌های جذابیت هستند. این ویژگی‌ها به صورت ترکیبی از ویژگی‌های کیفیت، زیبایی و احساس بر روی ویژگی‌های سطح پایین تصویر تعریف می‌شوند.

روش پیشنهادی دارای یک گام[۸] آموزش است. در گام آموزش مدل‌های پیمانه محتوا و پیمانه جذابیت آموزش داده می‌شود و وزن‌دهی با استفاده از بازخورد کاربر انجام می‌شود. در گام آزمایش[۹] الگوریتم هر کدام از پیمانه‌ها بر روی تصاویر اعمال شده و نتایج آن‌ها با هم ترکیب می‌شود. گام آموزش در شکل ۱ نشان داده شده است.

ایده پیمانه محتوا بر مبنای مدل‌سازی معنایی اطلاعات است. هدف آن است که ویژگی‌های سطح بالای معنایی را از روی ویژگی‌های سطح پایین تصویری به دست آوریم. در این مقاله از یک بهبود انجام شده بر روی روش ارائه شده در [۲] استفاده کرده‌ایم. در این روش هستان‌شناسی دامنه به عنوان ورودی به سیستم داده می‌شود. سپس طبقه‌بندی طراحی می‌شود که تصاویر را به مفاهیم هستان‌شناسی دامنه طبقه‌بندی کند. هستان‌شناسی دامنه مفاهیم حائز اهمیت در یک دامنه و ارتباط بین آن‌ها را به سیستم خلاصه‌سازی تصویر معرفی می‌کند. فرض کنیم تصاویر اولیه مربوط به مکان‌های دیدنی یک شهر هستند. هستان‌شناسی دامنه، مفاهیم پراهمیت در یک دامنه مانند مکان‌های دیدنی مهم در یک شهر را به سامانه[۱۰] معرفی می‌کند. این مکان‌های دیدنی برای هر شهری متفاوت هستند به عنوان مثال مفاهیمی مانند fountain و amphitheater برای شهری مانند رم (مطابق با Colosseum و Trevi Fountain) و مفاهیمی مانند رودخانه و مسجد برای شهری مانند اصفهان در هستان‌شناسی دامنه وجود دارند. به بیان دیگر، هستان‌شناسی دامنه، دانش مورد نیاز درباره محتوا و زمینه تصاویر را به سیستم وارد می‌کند. در گام آموزش، طبقه‌بند با مفاهیم هستان‌شناسی دامنه آموزش داده می‌شود. در گام آزمایش از طبقه‌بند آموزش داده شده برای محاسبه ویژگی‌های هستان‌شناسی استفاده شده و سپس ویژگی‌های هستان‌شناسی برای محاسبه محتوای اطلاعاتی تصاویر استخراج می‌شود.

ایده پیمانه جذابیت پیدا کردن مدلی است که بتوان از روی آن ویژگی‌های سطح بالای جذابیت را از ویژگی‌های سطح پایین و تصویری به دست آورد. در این مقاله پیشنهاد شده است که از طبقه‌بندی استفاده شود که بتواند تصاویر جذاب را از غیرجذاب جدا کند. در گام آموزش، طبقه‌بند با تصاویر جذاب و غیرجذاب آموزش داده می‌شود. برای فراهم کردن داده آموزش برای طبقه‌بند، از رفتارهای کاربران در شبکه‌های اجتماعی استفاده شده است. به این ترتیب که دو مجموعه از تصاویر با برچسب‌های جذاب و غیرجذاب با استفاده از تصاویری که در شبکه‌های اجتماعی نشانه[۱۱] دارند جمع‌آوری می‌شود. رفتارهای کاربران نیز با توجه به محتوا و زمینه‌ی تصاویر متفاوت هستند. به عنوان مثال در شهری مانند رم دامنه‌ی تعداد نشانه‌ها بیشتر از شهری با تعداد کمتر بازدیدکننده است. بنابراین جذابیت باید برای محتواهای متفاوت به طور جداگانه‌ای مدل شود. با مدل کردن جذابیت در شبکه‌های اجتماعی، تصاویر جذاب و غیرجذاب به عنوان داده آموزشی به طبقه‌بند جذابیت داده می‌شود. از این طبقه‌بند در گام آزمایش برای ساختن امتیاز جذابیت استفاده می‌شود.

پس از گام آموزش نوبت به گام آزمایش می‌رسد. مراحل این گام در شکل ۲ نشان داده شده است. روش پیشنهادی شامل سه پیمانه است. دو پیمانه اول بر روی ویژگی‌های محتوایی و ویژگی‌های جذابیت کار می‌کنند. پیمانه اول برای هر تصویر امتیازی استخراج می‌کند که نشان‌دهنده میزانی از اطلاعات است که یک تصویر از کل مجموعه تصاویر را پوشش می‌دهد. این پیمانه را از این پس پیمانه محتوا می‌نامیم. پیمانه دوم بر روی ویژگی‌های جذابیت کار می‌کند و یک امتیاز به هر تصویر نسبت می‌دهد که نشان‌دهنده میزان جذابیت هر تصویر است. از این پس این پیمانه را پیمانه جذابیت می‌نامیم.

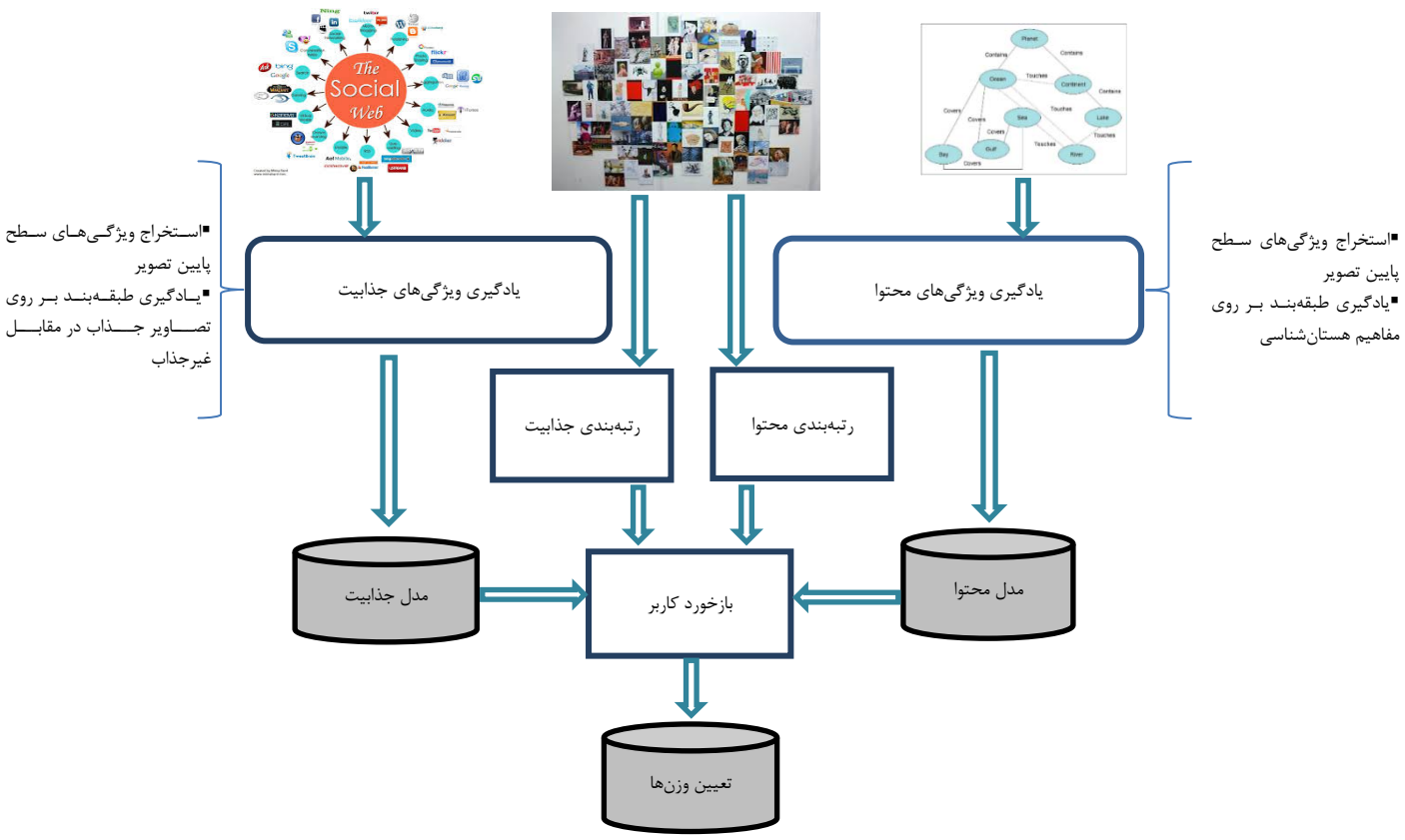
پیمانه سوم پیمانه تجمیع و خلاصه‌سازی است. هدف این پیمانه این است که خروجی دو پیمانه دیگر را ترکیب کرده و یک عدد نهایی و نشان‌دهنده میزان نماینده بودن تصویر استخراج کند. برای انجام تجمیع از روش جمع وزن‌دار استفاده کرده‌ایم. به این ترتیب هر تصویر یک امتیاز خلاصه‌سازی دریافت می‌کند که برابر جمع وزن‌دار امتیاز جذابیت و امتیاز محتوا است.

لازم به ذکر است که هستان‌شناسی دامنه و بستر شبکه‌های اجتماعی در گام آموزش استفاده می‌شوند و هیچ اطلاعات جانبی‌ای در گام آزمایش استفاده نمی‌شود. این مسئله موجب می‌شود روش پیشنهادی در شرایطی که اطلاعات جانبی در دسترس نیستند قابل به‌کارگیری باشد.

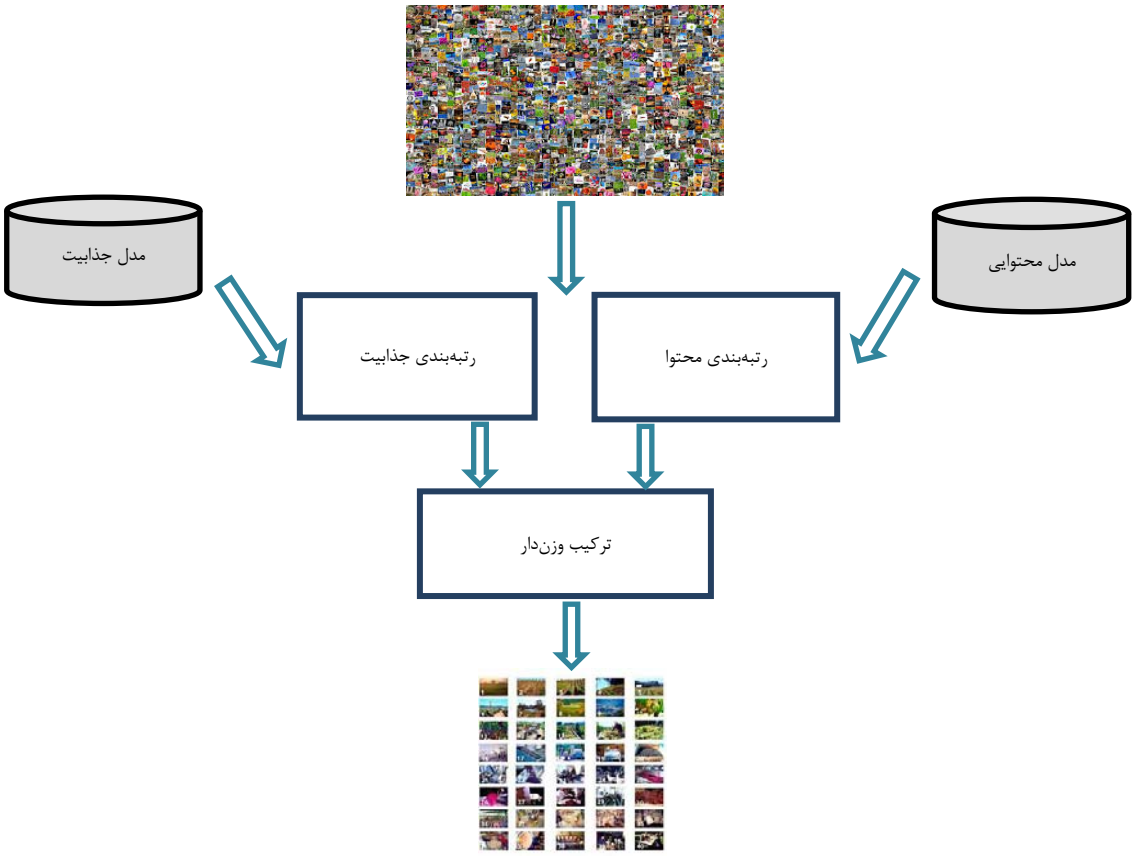
در ادامه جزئیات هر کدام از پیمانه‌ها توضیح داده شده است.

۳-۱- پیمانه جذابیت

هدف این پیمانه این است که به هر تصویر امتیازی نشان‌دهنده میزان جذابیت نسبت داده شود. برای این هدف در این قسمت طبقه‌بندی قرار داده شده است که بین تصاویر جذاب و غیرجذاب طبقه‌بندی انجام می‌دهد. این طبقه‌بند با استفاده از ویژگی‌های سطح پایین تصویر در مورد جذاب بودن یا عدم‌جذاب بودن یک تصویر تصمیم می‌گیرد. در گام آموزش، مجموعه‌ای از تصاویر جذاب و غیرجذاب به طبقه‌بند داده شده و طبقه‌بند آموزش داده شده در گام آزمایش برای تصمیم در مورد جذابیت تصاویر به کار گرفته می‌شود. در انتها، اطمینان تعلق هر تصویر به رده[۱۲] تصاویر جذاب به عنوان امتیاز جذابیت در نظر گرفته می‌شود.



شکل ۱- روش پیشنهادی گام آموزش



شکل ۲- روش پیشنهادی گام آزمایش

در اینجا اولین مسئله فراهم کردن تصاویر جذاب و غیرجذاب برای آموزش است. در روش پیشنهادی برای حل این موضوع از بستر شبکه‌های اجتماعی استفاده شده است. کاربران در شبکه‌های اجتماعی می‌توانند بر روی تصاویر فعالیت انجام دهند، به عنوان مثال وقتی یک تصویر برای آنها جذاب است کاربر می‌تواند آن را نشانه کند (مانند پسند[۱۳] در فیس‌بوک و یا علاقه[۱۴] در فلیکر). تحقیقات جدید نشان می‌دهد که این اطلاعات می‌توانند برای تعیین اینکه یک تصویر دارای جذابیت[۱۵] هست یا خیر مورد استفاده قرار بگیرند [۶]. هم‌چنین در برخی تحقیقات گزارش شده است که این ویژگی‌ها بعضاً از ویژگی‌های متنی و تصویری مهم‌تر هستند [۲۵]. پدرو و همکاران هم چنین گزارش می‌کنند که بین تعداد نشانه‌ها و بقیه رفتارهای اجتماعی مانند به اشتراک‌گذاری یا یادداشت همبستگی[۱۶] وجود دارد [۲۶]. بنابراین در این مقاله تصمیم گرفتیم از نشانه‌ها برای پیدا کردن تصاویر جذاب استفاده کنیم. به این ترتیب تصاویر به دو مجموعه جذاب و غیرجذاب براساس تعداد نشانه‌ها تقسیم می‌شوند.

برای انجام این کار یک مقدار آستانه برای تعداد نشانه‌ها تعریف می‌شود. به این ترتیب تصاویری که تعداد نشانه‌هایشان بالاتر از آستانه است به عنوان تصاویر جذاب و تصاویری که تعداد نشانه‌هایشان پایین‌تر از آستانه است به عنوان تصاویر غیرجذاب معرفی می‌شوند.

مسئله بعد مقدار آستانه برای نشانه‌ها می‌باشد. کردانی و ابراهیمی مقدم در مطالعه‌ای نشان داده‌اند که مقدار آستانه به محتوای تصاویر مورد نظر وابسته است [۶]. آن‌ها نشان می‌دهند که انتخاب یک مقدار آستانه سراسری امکان‌پذیر نبوده و مقدار آستانه باید با توجه به محتوای تصاویر انتخاب شود. هم‌چنین در این مطالعه آنها نشان داده‌اند که میانگین تعداد نشانه‌ها را می‌توان به عنوان آستانه برای تصمیم‌گیری در مورد جذابیت تصاویر استفاده کرد. بنابراین در این مقاله نیز ما مقدار آستانه را برای هر دامنه از تصاویر به طور جداگانه برابر با میانگین تعداد نشانه‌ها انتخاب کرده‌ایم.

با داشتن دو دسته از تصاویر جذاب و غیرجذاب از ویژگی‌های سطح پایین زیر برای آموزش جذابیت استفاده شده است. این ویژگی‌ها در سه دسته کیفیت ادراکی[۱۷]، حساسیت زیبایی شناختی[۱۸] و میزان تاثیر[۱۹] جای می‌گیرند و برای اولین بار در [۲۷] معرفی شده‌اند:

- کیفیت ادراکی ویژگی‌هایی از تصویر را در نظر می‌گیرد که کیفیت را تحت تاثیر قرار می‌دهند مانند روشنایی[۲۰]، درخشندگی[۲۱]، تباین[۲۲]، رنگ، وضوح و ماتی[۲۳]. روشنایی در دو بعد، درخشندگی[۲۴] و تباین[۲۵] تعریف می‌شود. درخشندگی به صورت میانگین شدت روشنایی[۲۶] پیکسل‌های تصویر سیاه- سفید و تباین، واریانس شدت روشنایی پیکسل‌های تصویر سیاه- سفید می‌باشند. نقشه برجسته قسمت‌های مهمتر یک تصویر را نشان می‌دهد، برای استخراج برخی ویژگی‌ها از تصویر ضرب شده در نقشه برجسته استفاده می‌شود. منطقه موضوعی[۲۷] تصویر به ناحیه‌ای گفته می‌شود که چگالی اشیاء موجود در تصویر در آن منطقه بیشتر است، بنابراین کلیات تصویر از آن قابل برداشت است. لذا این منطقه از اهمیت بالایی برخوردار است. به همین دلیل برخی ویژگی‌ها از منطقه موضوع تصویر استخراج می‌شوند.
- حساسیت زیبایی به این مسئله می‌پردازد که آیا یک تصویر از نظر زیبایی جذابیت دارد یا خیر. برای این دسته از ویژگی‌ها قانون سوم‌ها[۲۸]، سادگی و وزن بصری در نظر گرفته شده است. قانون سوم‌ها در عکاسی تاکید می‌کند که قرار گرفتن عناصر مهم تصویر در محل برخورد خطوط افقی و عمودی‌ای که تصویر را به سه قسمت تقسیم می‌کنند، بیش‌ترین توجه را جلب می‌کند. برای سادگی توزیع لبه‌ها در تصویر اولیه و تصویر ضرب در نقشه برجسته تصویر و همچنین مقادیر متمایز خلوص در تصویر اولیه و منطقه موضوع اصلی تصویر [۲۸] به دست آورده می‌شود و وزن بصری میزان مجزا بودن موضوع از پس زمینه را مشخص می‌کند.
- میزان تاثیر شامل وضوح و اثربخشی بوده و بیان‌کننده‌ی این است که چه عواطف مشخصی از تصویر احساس می‌شود و یا به عبارت دیگر تصویر چه احساساتی را برانگیخته می‌کند. در این مقاله از ویژگی‌های خط و رنگ استفاده شده است. مطالعات نشان می‌دهد، خطوط یک تصویر باعث ایجاد تاثیر است عاطفی می‌شوند [۲۹]. خطوط افقی با یک افق ثابت، آرامش و احساس امنیت را القا می‌کنند، خطوط عمودی واضح و مستقیم هستند و باعث ایجاد حس کرامت و ابدیت می‌شوند. از طرف دیگر، خطوط کج ناپایدار بوده و احساس پویایی را ایجاد می‌کنند. خطوط با جهت‌های بسیار مختلف بیان‌کننده‌ی بی‌نظمی و سردرگمی هستند. هر چه خط بلندتر، ضخیم‌تر و غالب‌تر باشد اثر روانی القا شده‌ی آن قوی‌تر می‌شود. برای پیدا کردن خطوط ایستا و پویا شیب خطوط در تصاویر با استفاده از تبدیل هاف به دست آورده شده و خطوط براساس زاویه آنها به کلاس ایستا و پویا نقسیم می‌شوند [۳۰].

برای ویژگی‌های رنگ سه دسته ویژگی زیر استفاده شده‌اند:

ویژگی اول احساسی: ویژگی احساسی اول القاکننده‌ی جفت کلمات معنادار سبک- سنگین[۲۹]، سرزنده- خسته‌کننده[۳۰]، پرطراوت- ویران[۳۱] و سرد- گرم[۳۲] است. بر پایه‌ی تئوری روانشناسی رنگ‌ها، رنگ‌های سایه روشن احساسات سبک، سرزنده، پرطراوت و گرم را منتشر می‌کنند. رنگ‌های سرد و گرم تأثیر متفاوتی روی احساس می‌گذارند، رنگ‌های گرم القاکننده‌ی عواطف شادی، سرزندگی و طراوت بوده در حالی‌که رنگ‌های سرد القاکننده‌ی عواطف ناراحتی، کسل‌کنندگی و افسردگی هستند.

ویژگی دوم احساسی: ویژگی احساسی دوم ارتباط قوی‌ای با جفت کلمات معنادار سخت- نرم[۳۳]، عالی- معتدل[۳۴] و پرجلوه- ظریف[۳۵] دارد. براساس تئوری روانشناسی رنگ‌ها، تصاویر با اشباع بالا القاکننده‌ی حس اعتدال و ملایمت و ظرافت می‌باشد، در حالی که رنگ‌های با اشباع پایین القاکننده‌ی احساسات عالی و پرجلوه است؛ به‌طور مشابه احساسات سخت- نرم ارتباط نزدیکی با تباین دارند [۳۰].

ویژگی سوم احساسی: ویژگی احساسی سوم ارتباط قوی‌ای با کلمات معنادار تخیل- واقعیت و قطعی- غیرقطعی دارد. براساس تئوری دریافت بصری، تباین و روشنایی بالا و لبه‌های تیز القاکننده‌ی واقعیت و قطعی بوده در حالی که تباین و روشنایی پایین و لبه‌های پهن‌تر القاکننده‌ی عواطف غیرقطعی و تخیل است [۳۰]. جزئیات این ویژگی‌ها در جدول ۲ نمایش داده شده است.

بردار ویژگی به دست آمده براساس این جدول، یک بردار ۴۹ بعدی است. سپس از تجزیه و تحلیل مولفه‌های اصلی[۳۶] بر روی بردار ویژگی ساخته شده برای کاهش بعد استفاده شده و از ماشین بردار پشتیبان با کرنل خطی برای طبقه‌بندی بین رده‌های جذاب و غیرجذاب استفاده شده است. با استفاده از ماشین بردار پشتیبان تصاویر مجموعه داده با برچسب‌های جذاب و یا غیرجذاب طبقه‌بندی می‌شوند.

برچسب تصاویر با بررسی $L_i \in [-1, 1]$ به وسیله طبقه‌بند پیش‌بینی می‌شود که در آن L_i فاصله نرمال شده خروجی ماشین بردار پشتیبان تا خط جداساز است. در این مقاله با استفاده از تابع سیگموید بر روی L_i میزان جذابیت تصویر به صورت زیر محاسبه شده است:

$$A_{x_i} = \frac{1}{1+e^{-L_i}} \qquad (۶)$$

که این مقدار به عنوان امتیاز جذابیت یک تصویر در نظر گرفته شده است.

جدول ۲- ویژگی‌های جذابیت تصویر

دسته ویژگی	نام ویژگی	تعداد بعد	توضیحات
کیفیت ادراکی	براقی[۳۷]	۲	میانگین (روشنایی) و واریانس (تباین) شدت تصاویر خاکستری [۲۶]
	رنگ	۷	میانگین و انحراف معیار اشباع[۳۸] و خلوص[۳۹] [۲۶، ۲۹]، تباین رنگ‌ها [۲۶] رنگارنگی[۴۰] (تفاوت رنگ تصویر در مقابل خاکستری) [۲۶] طبیعی بودن[۴۱] (ضریبی از حضور پوست، آسمان و چمن در تصویر که بعد از انتقال تصویر به فضای LUV به دست آورده می‌شود) ویژگی‌های رنگی [۳۱]
	وضوح و ماتی[۴۲]	۳	وضوح: میانگین و انحراف معیار لاپلاسین تصویر که با توجه به میانگین درخشندگی[۴۳] محلی نرمال شده است [۲۶]. تیری: فرکانس توزیع تبدیل FFT تصویر [۲۸]
	کیفیت موضوع	۱۲	برق، رنگ و تیزی تصویر ضرب شده در نقشه برجسته[۴۴] تصویر، و میزان تاری در منطقه موضوع تصویر
حساسیت زیبایی	قانون سوم‌ها	۱	قرار دادن موضوع تصویر در مکانی نزدیک نقطه قوی مرکزی [۳۲]
	سادگی	۴	توزیع لبه‌ها در تصویر اولیه و تصویر ضرب در نقشه برجسته تصویر، شمارش مقادیر متمایز خلوص در تصویر اولیه و منطقه موضوع اصلی تصویر [۲۸]
	وزن بصری	۱	میزان مجزا بودن موضوع اصلی از پس‌زمینه [۳۳]
میزان تاثیر	پویایی	۲	تعداد نسبی و طول خطوط ایستا و پویا [۳۴]
	احساس رنگ	۱۷	هیستوگرام طراحی شده برای بررسی تاثیر احساسی رنگ تصویر: میزان سرد یا گرم بودن رنگ تصویر، سخت یا نرم بودن رنگ تصویر، واقعی یا تخیلی بودن رنگ تصویر [۳۵]

۳-۲- پیمانه محتوا

برای مدل کردن اطلاعات محتوایی از هستان‌شناسی دامنه استفاده شده است. هستان‌شناسی دامنه حاوی مفاهیم مهم در یک دامنه و ارتباط بین آنها است. به بیان دیگر، هستان‌شناسی دامنه دانش عمومی در زمینه تصاویر را به سیستم می‌دهد. شکل ۳ یک مثال از هستان‌شناسی مربوط به تصاویر شهر پاریس را نشان می‌دهد. برای انجام خلاصه‌سازی محتوایی تصاویر، در ابتدا تصاویر مجموعه اولیه با استفاده از طبقه‌بند به مفاهیم مختلف هستان‌شناسی دامنه طبقه‌بندی شده و پس از آن ویژگی‌های هستان‌شناسی بر روی خروجی طبقه‌بند تعریف می‌شوند.

برای انجام طبقه‌بندی از یک روش طبقه‌بندی سلسله مراتبی استفاده شده است. این روش در [۳۶] پیشنهاد داده شده است. به این ترتیب که در هر نود در هستان‌شناسی یک طبقه‌بند قرار داده شده است. وظیفه هر طبقه‌بند آن است که بین رده‌هایی که زیر مفهوم مفهوم جاری هستند طبقه‌بندی انجام دهد. در هر نود در هستان‌شناسی، طبقه‌بندی بین زیرمفاهیم مفهوم جاری انجام شده و نودی که تصویر به آن طبقه‌بندی می‌شود برای کاوش مجدد انتخاب می‌شود. این کار به صورت تکراری انجام می‌شود تا به مفاهیم برگ برسیم. در انتها هر تصویر به دنباله‌ای از مفاهیم در هستان‌شناسی طبقه‌بندی شده و هر تصویر با سبد کلمات در فضای هستان‌شناسی نمایش داده می‌شود.

برای هر تصویر یک مجموعه از ویژگی‌های هستان‌شناسی به صورت زیر تعریف می‌شوند [۲].

$$F_i = \left(C_i^1, \dots, C_i^j, \dots, C_i^c\right) \quad (۷)$$

که در آن C^j مفهومی در فضای هستان‌شناسی است و C_i^j ویژگی هستان‌شناسی تصویر i با توجه به مفهوم j است که به صورت زیر محاسبه می‌شود و c تعداد مفاهیم است [۲].

$$C_i^j = \alpha_j \times Conf_i^j \quad (۸)$$

در این رابطه $Conf_i^j$ اطمینان به تعلق تصویر i به مفهوم j است. که با استفاده از تابع سیگموید بر روی خروجی طبقه‌بند به دست می‌آید. اگر L_i^j فاصله نرمال شده خروجی ماشین بردار پشتیبان تا خط جداساز برای تعلق تصویر i به کلاس متعلق به مفهوم j باشد در این مقاله با استفاده از تابع سیگموید بر روی L_i^j اطمینان به تعلق تصویر i به کلاس متناظر با مفهوم j مطابق معادله زیر به دست آورده شده است.

$$Conf_i^j = \frac{1}{1+e^{-L_i^j}} \quad (۹)$$

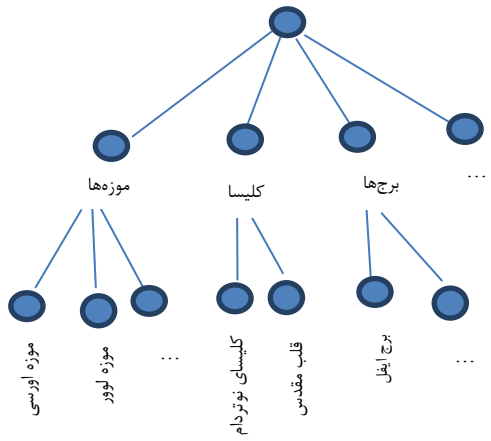
α_j به کمک رابطه ۹ محاسبه می‌شود و برابر است با جمع اطمینان به تعلق تصاویر به رده j تقسیم بر جمع اطمینان به تعلق تصاویر به همه رده‌ها [۲].

$$\alpha_j = \frac{N_j}{N_t} = \frac{\sum_{i=1}^{n} C_i^j}{\sum_{j=1}^{c}\sum_{i=1}^{n} C_i^j} \quad (۱۰)$$

دلیل استفاده از α_j این است که تعداد رخداد مفاهیم در محاسبه ویژگی‌های هستان‌شناسی در نظر گرفته شود. به این ترتیب مفاهیمی که بیشترین رخداد را دارند تاثیر بیشتری روی ویژگی‌های هستان‌شناسی دارند. تصاویر خلاصه باید نماینده مجموعه اولیه باشد پس به مفاهیم پررخداد شانس بیشتری برای انتخاب داده می‌شود [۲].

به‌عنوان مثال، بردار ویژگی تصاویر مرتبط با شهری مانند پاریس، شامل مفاهیمی که در هستان شناسی آن شهر هستند مانند (برج، کلیسا، موزه و ...) می‌باشد. برای هر تصویر میزان تعلق آن به مفاهیم مختلف ($Conf_i^j$) به‌دست آورده شده و با استفاده از آن، بردار ویژگی برای هر تصویر ساخته می‌شود. در کنار آن پارامتر α برای هر مفهوم محاسبه شده که برابر اطمینان به تعلق تصاویر به یک مفهوم تقسیم بر اطمینان به تعلق تصاویر به بقیه مفاهیم است. در انتها بردار ویژگی تصویر از ضرب $Conf_i^j$ در α_j به دست می‌آید (رابطه ۸).

در این مقاله از نودهای میانی طبقه‌بند برای ساختن ویژگی‌های هستان‌شناسی استفاده شده است.



شکل ۳- هستان‌شناسی دامنه مربوط به شهر پاریس

به عنوان معیار فاصله، از فاصله کسینوسی بین بردار ویژگی تصویر و بردار ویژگی مجموعه استفاده شده است.

$$\text{ContentScore}_{iC} = \frac{F_i . F_C}{|F_i||F_C|} \quad (۱۱)$$

که در آن F_i و F_C بردارهای ویژگی مجموعه و تصویر هستند. بردار ویژگی مجموعه از میانگین‌گیری بردار ویژگی تصاویر به دست می‌آید و نقطه نشانگر ضرب داخلی بین بردارهای ویژگی است. این معیار فاصله نشان می‌دهد که یک تصویر چقدر از اطلاعات کل مجموعه تصاویر را در فضای هستان‌شناسی پوشش می‌دهد. هر چه مقدار ContentScore در فضای هستان‌شناسی بزرگتر باشد نشان‌دهنده آن است که آن تصویر پوشش اطلاعاتی بالاتر و نمایندگی بالاتری دارد.

۳-۳- تجمیع و خلاصه‌سازی

هدف این پیمانه جمع‌بندی بین امتیازهای محتوایی و جذابیت و انجام خلاصه‌سازی است. ایده‌ی به کار گرفته شده در این مقاله، استفاده از جمع وزن‌دار امتیازهای جذابیت و معنایی است. ورودی این قسمت مجموعه‌ی تصاویر ورودی و امتیازهای جذابیت و معنایی و خروجی آن لیست رتبه‌بندی شده‌ی تصاویر براساس امتیاز کلی آنها است. به این ترتیب جمع وزن‌دار امتیازهای معنایی و جذابیت به دست می‌آید و به متغیر TotalScore مقدار دهی می‌شود.

$$\text{TotalScore}_i = \alpha \times \text{ContentScore}_i + \beta \times \text{AttractivenessScore}_i \quad (۱۲)$$

α و β وزن‌هایی هستند که به امتیازهای محتوایی و جذابیت نسبت داده می‌شود و تصاویر خلاصه براساس مقدار TotalScore انتخاب می‌شوند. برای تصمیم‌گیری در مورد آلفا و بتا یک آزمایش بازخورد از کاربران انجام شده است. جزئیات این آزمایش در ادامه مطرح شده است. نتایج حاکی از آن است که هر دوی ویژگی‌های محتوایی و جذابیت تقریبا به یک اندازه در خلاصه‌سازی مجموعه تصاویر مهم هستند.

آزمایش انتخاب آلفا و بتا بدین صورت طراحی شده است کهبه ازای ۱۲ شهر و هر شهر ۱۰۰ تصویر (در مجموع ۱۲۰۰تصویر) از میان تصاویر موجود برای آموزش انتخاب شدند. این تعداد تصویر با توجه به دو ملاحظه زیر انتخاب شده‌اند:

- در دسترس بودن کاربر انسانی برای ارزیابی
- مطابقت با روش‌های مشابه یادگیری در حوزه خلاصه‌سازی تصویر، روش‌های مشابه در حوزه یادگیری با نظارت نیز از حدود صد تصویر به ازای هر مفهوم استفاده می‌کنند. در نظر گرفتن بیشتر از صد تصویر در خلاصه‌سازی برای کاربر انسانی دشوار است [۱].

تصاویر به پیمانه‌های استخراج ویژگی‌های محتوایی و ویژگی‌های جذابیت داده شده و ده تصویر با امتیاز بالای محتوایی و جذابیت به دست آورده می‌شود. تصاویر بازیابی شده به کاربران نشان داده شد و از آنها خواسته شده که به مجموعه‌ی اولیه نگاه کرده و سپس یک نمره‌ی میزان نماینده بودن به تصویر بازیابی شده نسبت دهند. امتیاز میزان نماینده بودن با استفاده از جدول ۳ نسبت داده شده است.

جدول ۳- امتیازدهی به تصاویر براساس میزیان نماینده بودن

غیر نماینده	نمایندگی پایین	نمایندگی متوسط	نمایندگی بالا
۰	۱	۲	۳

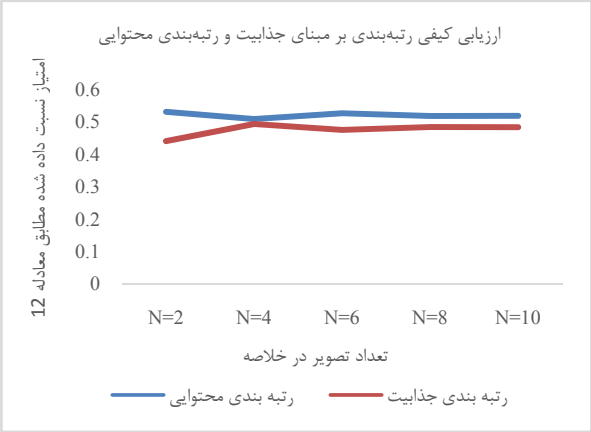
$$\text{Weight}_{set} = \frac{\sum_{j=1}^{U}(\text{grade}_j(1) + \sum_{i=2}^{M} + \frac{\text{grade}_j(i)}{\log_2 i})}{U*M} \quad (۱۳)$$

بعد از انجام ارزیابی توسط کاربران یک امتیاز کل برای میزان نماینده بودن مطابق معادله‌ی ۱۲ نسبت داده می‌شود. در معادله‌ی ۱۲ منظور از Set خلاصه محتوایی و یا خلاصه جذابیت است، U تعداد ارزیابی‌های موجود یا به بیان دیگر تعداد کاربرانی است که در آزمایش بازخورد کاربر حضور داشته‌اند، M تعداد کل تصاویر موجود در خلاصه است و $\text{grade}_j(i)$ امتیازی است که کاربر j به تصویر iام نسبت داده است. این روش نمره‌دهی از روش تخفیف افزایش تجمعی[۴۵] در بازیابی اطلاعات [۳۷] و با در نظر گرفتن این مسئله که روش‌های بازیابی قوی‌تر باید تصاویر را در مرتبه‌ی بالاتری بازیابی کنند به کار گرفته شده است. وزن نهایی نرمال شده به دست آمده از رابطه ۱۳ اهمیت هر مجموعه را نشان می‌دهد. این مقدار برای مقادیر متناظر آلفا و بتا در نظر گرفته می‌شود.

$$\alpha = \frac{\text{Weight}_{\text{ContentRankedSet}}}{\text{Weight}_{\text{AttractivenessRankedSet}} + \text{Weight}_{\text{ContentRankedSet}}}$$

$$\beta = \frac{\text{Weight}_{\text{AttractivenessRankedSet}}}{\text{Weight}_{\text{AttractivenessRankedSet}} + \text{Weight}_{\text{ContentRankedSet}}} \quad (۱۴)$$

شکل ۴ نتایج ارزیابی‌های کاربران را نشان می‌دهد. محور عمودی وزن‌های تعریف شده در معادله ۱۲ را نشان داده و محور افقی اندازه‌های متفاوت خلاصه را نشان می‌دهد. از شکل این نتیجه‌گیری حاصل می‌شود که هر دوی این ویژگی‌ها دارای اهمیت یکسانی هستند ولی ویژگی‌های محتوایی را می‌توان حائز اندکی اهمیت بیشتر دانست.



شکل ۴- ارزیابی خلاصه‌سازی محتوایی و جذابیت

۴- ارزیابی روش پیشنهادی

در این بخش به ارزیابی روش پیشنهادی خواهیم پرداخت. بدین ترتیب ابتدا در مورد پیش‌فرض‌های روش و مجموعه داده مورد استفاده شرح خواهیم داد. سپس به ارزیابی روش پیشنهادی خواهیم پرداخت. در قسمت اول نتایج روش پیشنهادی را با خلاصه‌های ساخته شده به وسیله‌ی انسان مقایسه کرده‌ایم و در بخش دوم پیمانه محتوا را با معیارهای کمی ارزیابی کرده‌ایم.

۴-۱- پیش‌فرض‌ها

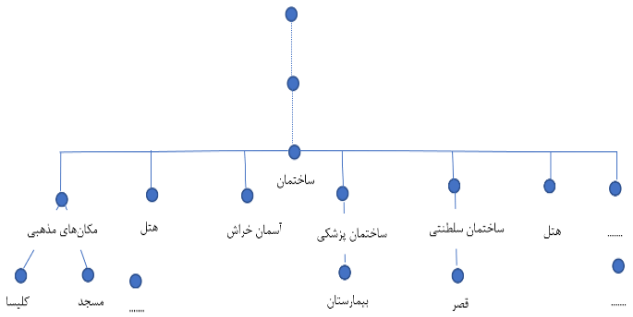
روش پیشنهادی تنها به خلاصه‌سازی می‌پردازد و برای از بین بردن داده‌های نویزی روشی پیشنهادن می‌کند. همچنین فرض بر این است که تصاویر حول موضوع – یا موضوعات مشخصی- هستند. بنابراین اگر داده‌ها نویزی باشند و یا تصاویر حول موضوع مشخصی نباشد روش به درستی عمل نخواهد کرد.

۴-۲- مجموعه داده

متاسفانه مجموعه داده استانداردی با تعداد تصویر زیاد برای خلاصه‌سازی مجموعه تصویر تا به حال ارائه نشده است، گرچه اخیرا تلاش‌هایی برای ایجاد مجموعه داده استاندارد خلاصه برای تعداد صدها تصویر انجام شده است [۱]. به عنوان مثال [۱] از ۱۴۰۰ تصویر مرتبط با ۱۴ مکان مختلف استفاده کرده است که به وسیله کاربر انسانی خلاصه‌سازی شده است، اما نویسندگان مجموعه داده خود را منتشر نکرده‌اند.

تصاویری که برای آزمایش روش پیشنهادی به کار گرفته شده است ۱۲۰۰۰ تصویر مربوط به ۱۲ شهر مختلف است که از فلیکر انتخاب شده‌اند. سپس از FlickrAPI برای استخراج و تحلیل آماری ۱۰۰۰۰ تصویر بالاتر مربوط به هر شهر و انتخاب آستانه برای مقدار جذابیت استفاده شده است. برای هر شهر ۵۰۰ تصویر جذاب و ۵۰۰ تصویر غیرجذاب بازیابی شده و با استفاده از آن طبقه‌بند جذابیت آموزش داده شده است.

هستان‌شناسی مربوط به هر شهر از روی DBpedia استخراج شده و سپس برای هر مفهوم، مفهوم منطبق با آن از وردنت بازیابی شده است. نودهای برگ در هستان‌شناسی استخراج شده و به طور بازگشتی پدر آن نودها انتخاب شده است. در حالتی که برای یک نود بیش از یک مسیر به ریشه وجود داشته، مسیر دارای بیشترین همپوشانی با مسیرهای دیگر انتخاب شده است. در انتها از ImagNet برای آموزش طبقه‌بند استفاده کرده‌ایم. جزئیات بیشتر در مورد روش استخراج هستان‌شناسی و نگاشت آن به وردنت در [۲] توضیح داده شده است. به عنوان نمونه یک زیرگراف استخراج شده از وردنت برای مفهوم ساختمان در شکل ۵ نشان داده شده است.



شکل ۵- نمونه‌ای از زیر گراف استخراجی از وردنت برای مفهوم ساختمان

۴-۳- مقایسه با خلاصه دستی

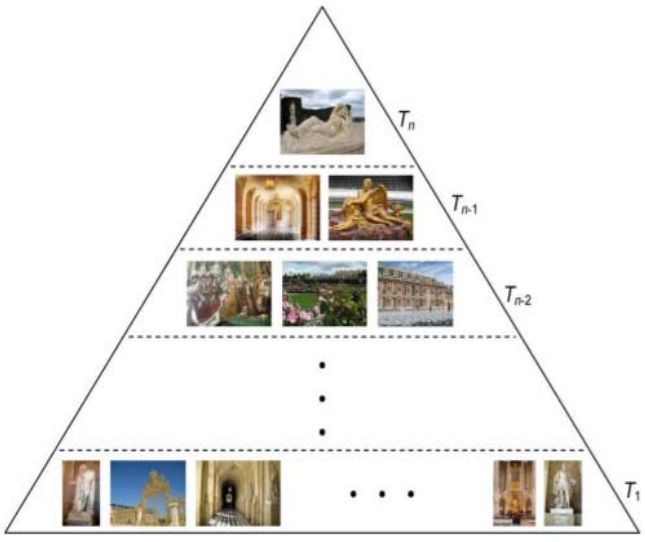
همان‌طور که ذکر شد در قسمت اول مقایسه، نتایج روش پیشنهادی را با خلاصه‌های ساخته شده به وسیله‌ی انسان مقایسه کرده‌ایم. برای انجام این کار ۱۲۰۰ تصویر مربوط به ۱۲ شهر مختلف انتخاب شده است (۱۰۰ تصویر برای هر شهر). تصاویر اولیه را به مجموعه‌ای از ۱۶ دانشجوی تحصیلات تکمیلی نمایش داده و از آنها خواستیم مجموعه‌ی کوچکی از تصاویر را به عنوان خلاصه انتخاب کنند. در انتها خلاصه‌ی تولید شده به وسیله‌ی روش پیشنهادی را با خلاصه‌های دستی تولید شده به وسیله‌ی کاربران مقایسه کرده‌ایم.

در این قسمت به روشی نیاز بود که بتواند نتایج خلاصه‌سازی را با مجموعه‌ای از خلاصه‌های انسانی مقایسه کند. رودیناک و همکاران [۴] نشان داده‌اند که معمولا توافق پایینی بین خلاصه‌های تولید شده به وسیله کاربران انسانی وجود دارد. تعداد تصاویر مشترک بین دو خلاصه انسانی میانگینی نزدیک به یک دارد. این مسئله انتخاب رهیافت‌هایی [۳۹]- که معیارهایی برای خلاصه‌های VERT [۳۸] و ROUGE مثل دستی و اتوماتیک در دامنه‌های متن و ویدیو هستند- را با سختی روبرو می‌کند.

بر همین اساس رودیناک و همکاران روشی هرمی برای مقایسه‌ی مجموعه‌ای از خلاصه‌های دستی و خلاصه‌ی اتوماتیک ارائه داده‌اند. در این روش تصاویر در یک هرم قرار داده می‌شوند- شکل ۵- خلاصه nام هرم هستند در n به این صورت که تصاویری که در لایه دستی حضور دارند. به این ترتیب تصاویر در لایه‌ی اول هرم تنها توسط یک کاربر انتخاب شده‌اند و هر چه در هرم بالاتر می‌رویم تصاویر توسط ام تصویر وجود n کاربران بیشتری انتخاب شده‌اند، تا اینکه در لایه دارد که توسط همه‌ی کاربران انتخاب شده است. رودیناک و همکاران تصویر در آن که حداکثر s با تعداد S عنوان می‌کنند که یک خلاصه مقدار ارزیابی شده را می‌گیرد باید به ترتیب زیر ساخته شده باشد [۴] و معیار d را برای ارزیابی خلاصه دستی ساخته شده به کار می‌برند. این معیار تعداد تصاویر مشترک بین خلاصه‌ی اتوماتیک و تصاویر موجود در هر لایه هرم را می‌شمارد و در عین حال به تصاویر لایه‌های بالای هرم امتیاز بیشتری نسبت می‌دهد [۴]:

$$d_{max} = \sum_{i=\theta+1}^{n} i \times |T_i| + \theta \times (N_R - \sum_{i=\theta+1}^{n} |T_i|) \quad (۱۵)$$

$$\theta = \max_i(\sum_{j=i}^{n} |T_i|) \geq N_R$$



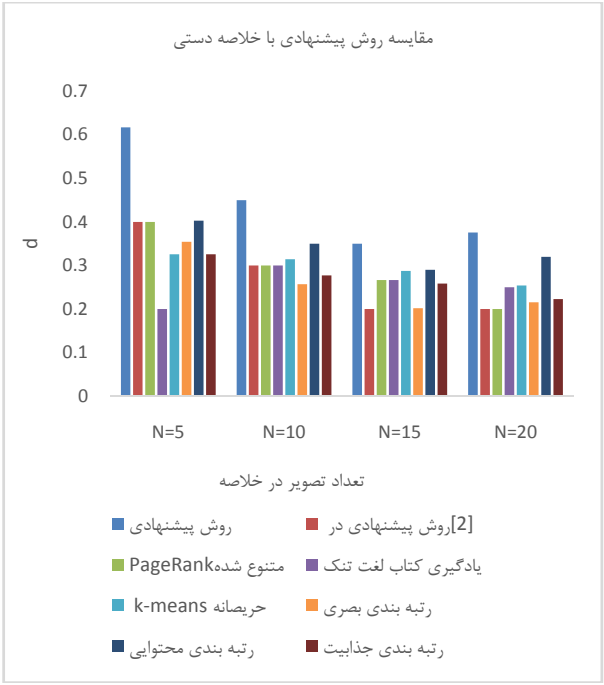
شکل ۶- روش هرم پیشنهادی در رودیناک و همکاران [۴]

$$d = \frac{1}{d_{max}} \times \sum_{i=1}^{n} i \times |T_i \cap R| \quad (۱۶)$$

ما نیز در این مقاله از این روش برای ارزیابی روش پیشنهادی استفاده کرده‌ایم. روش پیشنهادی در این مقاله با روش‌های زیر مقایسه شده است.

- k-means حریصانه: این روش در[۳] مطرح شده است. این روش یک تابع هدف را بهینه می‌کند که شباهت مجموعه‌ی انتخابی با مجموعه‌ی اولیه را بیشینه و شباهت مجموعه انتخابی با خودش را کمینه می‌کند.
- رتبه‌بندی بصری: این روش در [۱۴] به کار برده شده است. در این روش ازPageRank برروی ویژگی‌های تصویری برای انجام خلاصه‌سازی استفاده می‌شود.
- شبیه‌سازی حرارت با یادگیری تنک: تابع هدف پیشنهادی در این روش بر این مبنا استوار است که تصاویر مجموعه اولیه را می‌توان با ترکیب خطی و تنک مجموعه خلاصه بازسازی کرد. سپس از شبیه‌سازی حرارت برای بهینه‌سازی این تابع هدف استفاده شده است [۵].
- PageRank متنوع شده: این روش بهبودی را بر روی الگوریتم PageRank ارائه می‌کند. در این روش پس از استخراج هر کدام از تصاویر خلاصه، گراف اولیه براساس تفاوت با تصویر انتخابی وزن‌دهی مجدد می‌شود. به این ترتیب تصاویر خلاصه متنوع‌تر می‌شوند [۱۶].
- رتبه‌بندی جذابیت: در این روش تصاویر براساس امتیاز جذابیت رتبه‌بندی می‌شوند. به عبارت دیگر این روش، خروجی پیمانه جذابیت به تنهایی را در بر دارد.
- رتبه‌بندی محتوا: در این روش تصاویر براساس امتیاز محتوا رتبه‌بندی می‌شوند. به عبارت دیگر این روش، خروجی محتوا به تنهایی را در بر دارد.

شکل ۶ نتایج ارزیابی را نشان می‌دهد. محور عمودی پارامتر d است که با استفاده از معادله‌ی ۱۴ محاسبه شده است و محور افقی اندازه‌های مختلف خلاصه را نشان می‌دهد. همان‌طور که در شکل قابل ملاحظه است روش پیشنهادی نتیجه بهتری در مقایسه با روش‌های [۱۴]، [۳]، [۵] و [۱۶] که تنها معیارهایی عددی را بهینه می‌کردند داشته و هم‌چنین نتایج در مقایسه با استفاده از پیمانه‌های محتوا و جذابیت بهبود یافته است. بدین ترتیب روش پیشنهادی توانسته است به بهبودی و نه درصدی در ساختن خلاصه‌های شبیه خلاصه‌های دستی دست یابد.



شکل ۷- ارزیابی روش پیشنهادی، مقایسه با خلاصه دستی

شکل ۸ مجموعه‌ای از ۳۲ تصویر مربوط به شهر پاریس را نمایش می‌دهد و شکل ۹ نتایج خلاصه‌سازی با روش پیشنهادی را نشان می‌دهد. همان‌طور که از شکل مشخص است مجموعه‌ی اولیه شامل تصاویر تکراری زیادی است که طیف وسیعی از معیارهای جذابیت را پوشش می‌دهد و مجموعه‌ی انتخابی یک مجموعه ترکیبی است که هم پوشش اطلاعاتی و هم معیارهای جذابیت را در نظر گرفته است.

۴-۴- ارزیابی پیمانه محتوا

در این بخش به ارزیابی پیمانه محتوا می‌پردازیم. داده‌های تست برای این قسمت از ۱۲۰۰۰ تصویر انتخابی از فلیکر انتخاب شده‌اند. برای ارزیابی پیمانه محتوا از معیار معرفی شده در [۲] استفاده می‌کنیم. در روش پیشنهادی در [۲] از نودهای انتهایی در هستان‌شناسی برای ساخت ویژگی‌های هستان‌شناسی استفاده شده است، اما در این مقاله از همه نودهای میانی از ریشه تا برگ به عنوان ویژگی‌های تاثیرگذار در ساخت ویژگی‌های هستان‌شناسی استفاده شده است. معیار معرفی شده به صورت زیر است.

$$\gamma = \frac{\text{Coverage}}{\text{Redundancy}} \qquad (۱۷)$$

Coverage یا پوشش، این معیار شباهت مجموعه‌ی انتخابی را با مجموعه‌ی اولیه اندازه می‌گیرد و Redundancy یا تکرار، شباهت مجموعه‌ی اولیه را با خودش در نظر می‌گیرد. تصاویر نماینده‌ی انتخابی باید مشابه با مجموعه اولیه و غیرمشابه با همدیگر باشند و به عبارت دیگر پوشش بالا و تکرار پایین نشان‌دهنده‌ی سیستم خلاصه‌سازی بهتر است. در این مقاله از معیار γ برای ارزیابی روش خلاصه‌سازی استفاده شده است. هر چه معیار گاما بزرگتر باشد روش پیشنهادی کاراتر است.

محاسبه‌ی پوشش و تکرار به روش دستی دشوار است. بنابراین در [۲] پوشش و تکرار با شمارش تعداد انطباق‌های SIFT بین دو تصویر شمرده شده‌اند[۴۶]. شکل ۹ نتایج مقایسه روش پیشنهادی با روش به کار گرفته شده در [۲] را نمایش می‌دهد. نسبت γ را برای اندازه‌های مختلف خلاصه به دست آورده‌ایم. اندازه‌ی خلاصه بین یک درصد تا ۵ درصد اندازه مجموعه اولیه تغییر داده شده است. همان‌طور که از شکل پیداست روش پیشنهادی که از همه نودهای هستان‌شناسی برای ساخت ویژگی استفاده می‌کند در مقایسه با روش [۲] که تنها از نودهای برگ برای ساخت ویژگی‌های هستان‌شناسی استفاده می‌کنند نتایج بهتری در خروجی‌ها دارد. هم‌چنین در مقایسه با سایر روش‌ها که تنها از معیار عددی برای خلاصه‌سازی استفاده می‌کنند کاراتر عمل کرده است. بدین ترتیب روش پیشنهادی در مقایسه با روش‌های جاری خلاصه‌سازی تصویر توانسته است به بهبود سی و پنج درصدی پوشش اطلاعاتی در خلاصه کردن مجموعه تصویری به میزان یک درصد حجم اولیه و بهبود چهار درصدی پوشش اطلاعاتی در خلاصه کردن مجموعه تصویری به میزان پنج درصد حجم اولیه دست یابد.

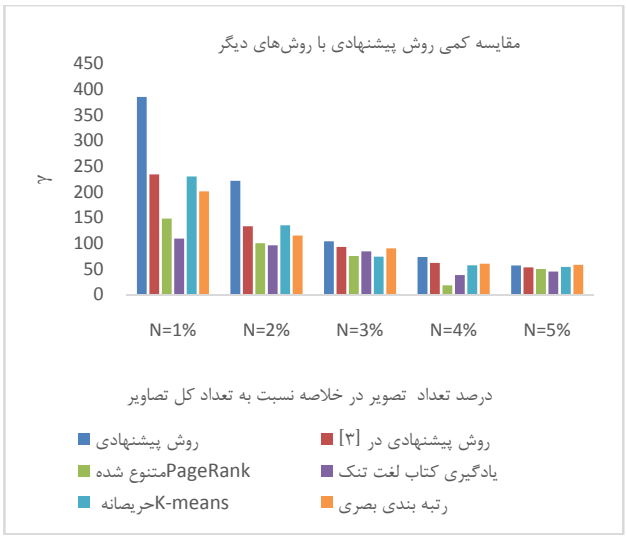
اکثر این روش‌ها به بهبود معیارهای عددی و تصویری پرداخته و از الگوریتم‌های حریصانه و یا مرکزیت گراف برای خلاصه‌سازی استفاده می‌کنند. روش پیشنهادی با استفاده از هستان‌شناسی دامنه، هم از معیارهای معنایی برای خلاصه‌سازی استفاده می‌کند و هم اطلاعاتی از فضای ویژگی تصاویر را با استفاده از هستان‌شناسی در اختیار دارد. به همین دلیل از روش‌های حریصانه و عددی بهتر عمل می‌کند.

شکل ۸- مجموعه‌ای از ۳۲ تصویر مربوط به شهر پاریس

شکل ۹- نتیجه خلاصه‌سازی تصاویر شکل ۹ با روش پیشنهادی

همان‌طور که مشاهده می‌کنیم در اندازه‌های کوچک‌تر خلاصه، روش پیشنهادی بهبود چشمگیرتری نسبت به بقیه روش‌ها دارد. از آنجایی که با افزایش تعداد تصاویر خلاصه، تصاویر بیشتری در خلاصه شرکت می‌کنند، شانس پوشش اطلاعاتی مجموعه اولیه افزایش می‌یابد. تا اینکه با انتخاب اندازه خلاصه برابر با اندازه مجموعه اولیه، معیار پیشنهادی برای همه روش‌ها یک مقدار به عنوان خروجی بر می‌گرداند- و این مقدار برابر یک است-. به این ترتیب روشی کاراتر عمل می‌کند که در تعداد پایین خلاصه بتواند پوشش اطلاعاتی مناسبی از مجموعه اولیه ارائه دهد.

همچنین همان‌طور که در بخش‌های قبل توصیف شد، در این مقاله از طبقه‌بند سلسله مراتبی استفاده شده و از همه نودهای هستان‌شناسی برای ساخت ویژگی استفاده شده است. بدین‌ترتیب و تعداد ویژگی‌ها در مقایسه با [۲] افزایش می‌یابند. این مساله منجر به افزایش حافظه مورد نیاز برای نگه‌داری ویژگی‌ها می‌شود. با این حال اگر طبقه‌بندی را صورت موازی در همه نودها انجام دهیم می‌توان با زمان اجرایی مشابه با روش [۲] ولی با مصرف حافظه بیشتر به جواب مطلوب رسید.



شکل ۱۰- ارزیابی پیمانه محتوا

جدول ۴- مقایسه زمان اجرای روش‌ها

روش	زمان اجرا برای صد تصویر بر حسب ثانیه
روش پیشنهادی	~۲۱
ریاحی و ابراهیمی مقدم [۲]	~۱۸
K-means حریصانه [۳]	~۹۵
رتبه‌بندی بصری [۱۴]	~۳۱
یادگیری افزونه و تکراری [۵]	~۲۱
PageRank متنوع شده [۱۶]	~۱۹

۴-۵- مقایسه زمان اجرا

در گام آخر به مقایسه زمان اجرای روش‌های مورد مقایسه پرداخته‌ایم. برای این کار روش‌های مورد بررسی، بر روی ششصد تصویر اجرا شده و زمان اجرای آنها را در جدول ۴ گزارش داده شده است. مجموعه داده مورد استفاده برای این آزمایش، ششصد تصویر برای شش شهر {پاریس، رم، بارسلونا، سن پطرزبورگ، هونگ کنگ، لندن} است که زمان اجرا بین شش مجموعه صد تایی مربوط به هر شهر، میانگین گرفته شده و به عدد بالاتر گرد شده است. بستر سخت‌افزاری یک دستگاه با واحد پردازش مرکزی Intel 2.5 GHz Core i-7 و ۸ گیگا بایت حافظه رم است. همان‌طور که در جدول می‌بینیم کمترین زمان اجرا مربوط به روش [۲] است و پس از آن روش‌های [۵]، [۱۶] و روش پیشنهادی با اختلاف کمتری قرار می‌گیرند. بیشترین زمان اجرا مربوط به K-means حریصانه [۳] است. زمان ذکر شده تنها برای آزمایش مدل‌ها می‌باشد. لازم به ذکر است که زمانی نیز برای آموزش مدل پیشنهادی لازم است که این زمان بیش از چند ده ساعت به طول می‌انجامد.

۵- نتیجه‌گیری

در این مقاله روشی چند معیاره و خودکار برای خلاصه‌سازی مجموعه‌ای از تصاویر اجتماعی مطرح شد. دو پیمانه‌ی محتوا و جذابیت در روش پیشنهادی معرفی شده است که هر کدام مجموعه‌ی متفاوتی از ویژگی‌ها را در نظر می‌گیرند. یک روش خودکار نیز برای استخراج ویژگی‌ها ارائه شد. خلاصه نهایی با ترکیب خروجی این دو پیمانه تولید می‌شود. نتایج آزمایش‌ها نشان می‌دهد روش پیشنهادی در مقایسه با روش‌های دیگر مطرح در حوزه خلاصه‌سازی تصویر که تک معیاره و عددی بودند به بهبود بالای سی درصد در ساخت خلاصه‌های شبیه خلاصه انسانی و بهینه‌سازی معیار عددی دست یافته است. در مقابل، روش پیشنهادی روشی مبتنی بر یک فار یادگیری بوده و میزان حافظه مصرفی این روش در مقابل برخی روش‌های دیگر بالاتر است.

این مقاله جهت‌های متفاوتی را برای کارهای آینده معرفی می‌کند. ترکیب نتایج پیمانه‌های جذابیت و محتوا به سادگی با استفاده از جمع وزن‌دار انجام شده است. روش‌های پیشرفته‌تری در ادبیات برای رتبه بندی چند معیاره وجود دارد. استفاده از این تکنیک‌های چند معیاره برای ترکیب نتایج دو پیمانه ایده‌ی جالبی برای دنبال کردن است.

مراجع

[1] S. Tschiatschek, R. K. Iyer, H. Wei, and J. A. Bilmes, "Learning Mixtures of Submodular Functions for Image Collection Summarization," in *Advances in Neural Information Processing Systems*, 2014, pp. 1413-1421.

[2] Z. R. Samani, and M. E. Moghaddam, "Aknowledge-based semantic approach for image collection summarization," *Multimedia Tools and Applications,* pp. 1-23, 2016.

[3] I. Simon, N. Snavely, and S. M. Seitz, "Scene summarization for online image collections," in Computer Vision, 2007. ICCV 2007, IEEE 11th International Conference on IEEE, 2007, pp. 1–8.

[4] S. Rudinac, M. Larson, and A. Hanjalic, "Learning crowdsourced user preferences for visual summarization of image collections," *Multimedia, IEEE Transactions on,* vol. 15, pp. 1231-1243, 2013.

[5] C. Yang, J. Shen, J. Peng, and J. Fan, "Image collection summarization via dictionary learning for sparse representation," *Pattern Recognition,* vol. 46, pp. 948–961, 2013.

[6] M. Kardaani, and M. E. Moghadam, "Attractive social image extraction based on users' social behaviors," in *2015 9th Iranian Conference on Machine Vision and Image Processing (MVIP)*, 2015, pp. 105-110.

[7] X. Shen, and X. Tian, "Multi-modal and multi-scale photo collection summarization," *Multimedia Tools and Applications,* pp. 1-15, 2015.

[8] J. E. Camargo, and F. A. González, "Multimodal latent topic analysis for image collection summarization," *Information Sciences,* vol. 328, pp. 270-287, 2016.

[9] Y. Jia, J. Wang, C. Zhang, and X.-S. Hua, "Finding image exemplars using fast sparse affinitypropagation," in *Proceedings of the 16th ACM international conference on Multimedia*, 2008, pp. 639-642.

[10] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg, "Mapping the world's photos," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 761-770.

[11] R. H. van Leuken, L. Garcia, X. Olivares, and R. van Zwol, "Visual diversification of image search results," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 341-350.

[12] J .a. G. Fan, Yuli, H. Luo, D. A. Keim, and Z. Li, "A novel approach to enable semantic and visual image summarization for exploratory image search," in *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, 2008, pp. 358-365

[13] Y. H. Yang, P. T. Wu, C. W. Lee, K. H. Lin, W. H. Hsu, and H. H. Chen, "ContextSeer: context search and recommendation at query time for shared consumer photos," in *Proceedings of the 16th ACM international conference on Multimedia*, 2008, pp. 199-208.

[14] Y. Jing, and S. Baluja, "Visualrank:Applying pagerank to large-scale image search," *Pattern Analysis and Machine*

[27] B. Geng, L. Yang, C. Xu, X.-S. Hua, and S. Li, "The role of attractiveness in web image search," in *Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 63-72.

[28] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, pp. 419-426.

[29] R. Datta, D. Joshi, J. Li, and J. Wang, "Studying aesthetics in photographic images using a computational approach," *Computer Vision–ECCV 2006,* pp. 288-301, 2006.

[30] J. Itten, "The Art of Color: The Subjective Experience and Objective Rationale of Color (trans. Ernest van Haagen)," ed: New York: Reinhold Publishing Corporation, 1973.

[31] K.-Q. Huang, Q. Wang, and Z.-Y. Wu, "Natural color image enhancement and evaluation algorithm based on human visual system," *Computer Vision and Image Understanding,* vol. 103, pp. 52-63, 2006.

[32] S. Bhattacharya, R. Sukthankar, and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 271-280.

[33] Y. Luo, and X. Tang, "Photo and video quality evaluation: Focusing on the subject," *Computer Vision–ECCV 2008,* pp. 386-399, 2008.

[34] J. Machajdik, and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 83-92.

[35] W. Wei-ning, Y. Ying-lin, and J. Sheng-ming, "Image retrieval by emotional semantics: A studyof emotional space and feature extraction," in *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*, 2006, pp. 3534-3539.

[36] Z. Abdollahpour, Z. R. Samani, and M. E. Moghaddam, "Image classification using ontology based improvedvisual words," in *2015 23rd Iranian Conference on Electrical Engineering*, 2015, pp. 694-698.

[37] K. Järvelin, and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Transactions on Information Systems (TOIS),* vol. 20, pp. 422-446, 2002.

[38] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Text summarization branches out: Proceedings of the ACL-04 workshop*, 2004.

[39] Y. Li, and B. Merialdo, "VERT: automatic evaluation of video summaries," in *Proceedings of the international conference on Multimedia*, 2010, pp. 851-854.

Intelligence, IEEE Transactions on, vol. 30, pp. 1877-1890, 2008.

[15] L. Yang, "Mining canonical views from internet image collections," Citeseer, 2011.

[16] J. Wang, L. Jia, and X.-S. Hua, "Interactive browsing via diversified visual summarization for image search results," *Multimedia systems,* vol. 17, pp. 379-391, 2011.

[17] D. Poole, *Linear algebra: A modern introduction*: Cengage Learning, 2014.

[18] A. Jaffe, M. Naaman, T. Tassa, and M. Davis, "Generating summaries and visualization for large collections of geo-referenced photographs," in *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, 2006, pp. 89-98.

[19] L. S. Kennedy, and M. Naaman, "Generating diverse and representative image search results for landmarks," in *Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 297-306.

[20] Y. Pang, Q. Hao, Y. Yuan, T. Hu, R .Cai, and L. Zhang, "Summarizing tourist destinations by mining user-generated travelogues and photos," *Computer Vision and Image Understanding,* vol. 115, pp. 352-363, 2011.

[21] J.-W. Jeong, H.-K. Hong, J.-U. Heu, I. Qasim, and D.-H. Lee, "Visual Summarization of the Social Image Collection Using Image Attractiveness Learned from Social Behaviors," *Multimedia and Expo (ICME), 2012 IEEE International Conference on,* pp. 538-543, 2012.

[22] S. Mei, G. Guan, Z. Wang, S. Wan, M. He, and D. D. Feng, "Video summarization via minimum sparse reconstruction," *Pattern Recognition,* vol. 48, pp. 522-533, 2015.

[23] S. U. Naci, U. Damnjanovic, B. Mansencal, J. Benois-Pineau, C. Kaes, M. Corvaglia, E. Rossi, and N. Aginako, "The COST292 experimental framework for rushes summarizationtask in TRECVID 2008," in *Proceedings of the 2nd ACM TRECVID video summarization workshop*, 2008, pp. 40-44.

[24] H. Fang, W. Lu, F. Wu, Y. Zhang, X. Shang, J. Shao, and Y. Zhuang, "Topic aspect-oriented summarization via group selection," Neurocomputing, vol. 149, pp. 1613–1619, 2015.

[25] R. van Zwol, A. Rae, and L. Garcia Pueyo, "Prediction of favourite photos using social, visual, and textual signals," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1015-1018.

[26] J. San Pedro, and S. Siersdorfer, "Ranking and classifying attractiveness of photos in folksonomies," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 771-780.

زهرا ریاحی سامانی مدرک کارشناسی خود را از دانشگاه شریف و مدرک کارشناسی‌ارشد و دکتری خود را از دانشگاه شهید بهشتی اخذ نموده و اکنون محقق دوره پسا دکتری در دانشگاه پنسیلوانیا است. تحقیقات مورد علاقه ایشان به کارگیری تکنیک‌های محاسباتی، یادگیری و تحلیلی در راستای استخراج اطلاعات عمیق و پیچیده از داده‌های تصویری است.
آدرس پست‌الکترونیکی ایشان عبارت است از:
z_rsamani@sbu.ac.ir

محسن ابراهیمی مقدم از سال ۱۳۸۵ به عنوان عضو هیات علمی دانشگاه شهید بهشتی فعالیت می‌کند. ایشان هر سه مقطع تحصیلی را در رشته مهندسی کامپیوتر در دانشگاه صنعتی شریف به پایان رسانده‌اند و هم‌اکنون به عنوان دانشیار دانشکده مهندسی و علوم کامپیوتر دانشگاه شهید بهشتی مشغول هستند. ایشان مسئول آزمایشگاه تحقیقات پردازش تصویر سیستم‌های توزیع شده هستند و تا کنون بالغ بر ۶۰ مقاله ژورنال و ده‌ها مقاله کنفرانس منتشر کرده‌اند. از علائق تحقیقاتی ایشان می‌توان به زمینه‌های پردازش تصویر و بینایی ماشین اشاره کرد.
آدرس پست‌الکترونیکی ایشان عبارت است از:
m_moghadam @sbu.ac.ir

اطلاعات بررسی مقاله:

تاریخ ارسال: ۱۳۹۶/۰۶/۰۶
تاریخ اصلاح: ۱۳۹۶/۰۷/۱۰
تاریخ قبول شدن: ۱۳۹۶/۱۰/۰۴
نویسنده مرتبط: دکتر محسن ابراهیمی مقدم، دانشکده مهندسی و علوم کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران.

[1]Coverage
[2]Redundancy
[3]Ontology
[4]Like
[5]Comment
[6]Share
[7]Multi Modal
[8]Phase
[9]Test
[10]System
[11]Bookmark
[12]Class
[13]Like
[14]Favorite
[15]Appeal
[16]Correlation
[17]Perceptual Quality
[18]Aesthetic Sensitivity
[19]Affective Tune
[20]Brightness
[21]Luminance
[22]Contrast
[23]Sharpness and Blur
[24]Brightness
[25]Contasrt
[26]Intensity
[27]Subject Region
[28]Rule of Thirds
[29]Light-Heavy
[30]Lively-Tedious
[31]Vibrant-Desolate
[32]Cool-Warm
[33]Hard-Soft
[34]Magnificent-Modest
[35]Showy-Elegant
[36]Principle Component Analysis
[37]Lightning
[38]Saturation
[39]Hue
[40]Colorfulness
[41]Naturalness
[42]Sharpness and Blur
[43]Luminance
[44]Saliency Map
[45]Discounted Cumulative Gain (DCG)
[46]Scale Invariant Feature Transform