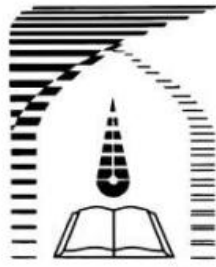


اللَّهُمَّ صَلِّ وَسَلِّمْ وَبَارِكْ عَلَى سَيِّدِنَا مُحَمَّدٍ



دانشگاه تربیت مدرس

دانشکده مهندسی برق و کامپیوتر

پایان نامه کارشناسی ارشد گرایش مهندسی نرم افزار

حاشیه نویسی معنایی تصاویر با استفاده از یک رویکرد تکاملی

سمانه بهرامی

استاد راهنما:

جناب آقای دکتر محمد صنیعی آباده

استاد مشاور:

جناب آقای دکتر نصرالله مقدم چرگری

آذرماه ۱۳۹۳



تاییدیه اعضای هیات داوران حاضر در جلسه دفاع از پایان نامه کارشناسی ارشد

خانم سمانه بهرامی پایان نامه ۶ واحدی خود را با عنوان حاشیه نویسی معنایی تصاویر با استفاده از یک رویکرد تکاملی در تاریخ ۱۳۹۳/۹/۲۵ ارائه کردند. اعضای هیات داوران نسخه نهایی این پایان نامه را از نظر فرم و محتوا تایید کرده، پذیرش آنرا برای اخذ درجه کارشناسی ارشد مهندسی کامپیوتر نرم افزار پیشنهاد می کنند.

عضو هیات داوران	نام و نام خانوادگی	رتبه علمی	امضا
استاد راهنما	دکتر محمد صنیعی آباده	استادیار	
استاد مشاور	دکتر نصراله مقدم چرکری	دانشیار	
استاد ناظر	دکتر فواد قادری	استادیار	
استاد ناظر	دکتر محمود فتحی	استاد	
مدیر گروه (یا نماینده گروه تخصصی)	دکتر فواد قادری	استادیار	

دستورالعمل حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی دانشگاه تربیت مدرس

مقدمه: با عنایت به سیاست‌های پژوهشی دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیات علمی، دانشجویان، دانش‌آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان‌نامه، رساله و طرحهای تحقیقاتی که با هماهنگی دانشگاه انجام شده است، موارد ذیل را رعایت نمایند:

ماده ۱- حقوق مادی و معنوی پایان‌نامه‌ها / رساله‌های مصوب دانشگاه متعلق به دانشگاه است و هرگونه بهره‌برداری از آن باید با ذکر نام دانشگاه و رعایت آیین‌نامه‌ها و دستورالعمل‌های مصوب دانشگاه باشد.

ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان‌نامه / رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و استاد راهنما مسئول مکاتبات مقاله باشد.

تبصره: در مقالاتی که پس از دانش‌آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان‌نامه / رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

ماده ۳- انتشار کتاب حاصل از نتایج پایان‌نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با مجوز کتبی صادره از طریق حوزه پژوهشی دانشگاه و بر اساس آئین‌نامه‌های مصوب انجام می‌شود.

ماده ۴- ثبت اختراع و تدوین دانش فنی و یا ارائه در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان‌نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق حوزه پژوهشی دانشگاه انجام گیرد.

ماده ۵- این دستورالعمل در ۵ ماده و یک تبصره در تاریخ ۱۳۸۴/۴/۲۵ در شورای پژوهشی دانشگاه به تصویب رسیده و از تاریخ تصویب لازم‌الاجرا است و هرگونه تخلف از مفاد این دستورالعمل، از طریق مراجع قانونی قابل پیگیری می‌شود.

نام و نام خانوادگی: سمانه بهرامی

امضاء



۹۳،۹،۲۹

آیین نامه چاپ پایان نامه (رساله) های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله) های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله) ی خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثار علمی» دانشگاه اطلاع دهد.

ماده ۲: در صفحه سوم کتاب (پس از برگ شناسنامه) عبارت ذیل را چاپ کند:
«کتاب حاضر، حاصل پایان نامه کارشناسی ارشد/ رساله دکتری نگارنده در رشته مهندسی کامپیوتر است که در سال ۱۳۹۳ در دانشکده مهندسی برق و کامپیوتر دانشگاه تربیت مدرس به راهنمایی جناب آقای دکتر محمد صنیعی آماده و مشاوره جناب آقای دکتر نصرالله مقدم چرکری از آن دفاع شده است.»

ماده ۳: به منظور جبران بخشی از هزینه های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به «دفتر نشر آثار علمی» دانشگاه اهدا کند. دانشگاه می تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.

ماده ۴: در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده رابه عنوان خسارت به دانشگاه تربیت مدرس، تأدیه کند.

ماده ۵: دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت بهای خسارت، دانشگاه می تواند خسارت مذکور را از طریق مراجع قضایی مطالبه و وصول کند؛ به علاوه به دانشگاه حق می دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶: اینجانب سمانه بهرامی دانشجوی رشته مهندسی کامپیوتر مقطع کارشناسی ارشد تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی: سمانه بهرامی

تاریخ و امضا: ۱۳۹۳/۹/۲۹



تقدیم به

قابل ستایش ترین انسان های زندگی من،

پدر و مادر عزیزم

به جبران گوشه ای از زحمات بی دریغشان

پاس نامه

پاس خدای را سزا است که غیر از او را نمی خوانم و اگر بخوانم هم پاسخی نمی شنوم،
پاس خدای را سزا است که به غیر از او دل نمی بندم و اگر بندم هم دلم را می شکنند و هم پشتم را خالی می کند،
خدایی که به او تکیه می کنم و او کرامی ام می دارد و دست نوازش بر سرم می کشد،
خدایی که از من بی نیاز است اما با من دوستی می کند و به من محبت می ورزد،
خدایی که با من بر دباری می کند کوی که هیچ کناهی نکرده ام، به گونه ای در من می نگرند که انکار هیچ خطایی از من سر نزده است، با من چنان رفتار می کند کوی بیج لغزشی نداشته ام.
به راستی چنین خدایی سایش بر انگیز است، به حقیقت دوست داشتنی و به وقع سجده کردنی است.
ضمن پاس و سایش به درگاه ایزدمنان که به من توانایی داد که با استعانت از او بتوانم این پژوهش را انجام دهم بر خود لازم می بینم که از دلگرمی و تشویق اساتید و دوستان که در نگارش این مجموعه مریاری نمودند قدر دانی کنم:
جناب آقای دکتر محمد صنیعی آبا، استاد راهنمای کرامی که در طول نگارش این مجموعه بارها بنیانی های عالمانه و بجایشان، سکاذا شایسته ای در هدایت این پایان نامه بوده اند.
جناب آقای دکتر نصرالله مقدم چرکری، استاد مشاور عزیز که با سه صدر و حسن خلق، مشاوره این تحقیق را پذیرفتند و در طول نگارش این مجموعه همواره از نظرات کارشناسانه شان بهره جستیم.
هم چنین از جناب آقای دکتر فواد قادری (ممتحن داخلی) و جناب آقای دکتر محمود فتحی (ممتحن خارجی) که زحمت قضاوت این پژوهش را بر عهده دارند کمال تشکر دارم.

چکیده

با رشد گسترده تکنولوژی‌های مرتبط با اینترنت، تصاویر و ویدئوها به سرعت در حال توسعه بر روی بستر اینترنت هستند. چگونگی ساماندهی و مدیریت این اطلاعات حجیم، بحث داغی شده است که نیاز به راه‌حل‌های فوری دارد. بازیابی اطلاعات چندرسانه‌ای، گامی مؤثر جهت حل مشکل بیان شده و حاشیه‌نویسی خودکار تصاویر گامی مهم و کلیدی در بازیابی اطلاعات چندرسانه‌ای می‌باشد. حاشیه‌نویسی خودکار تصاویر به ایجاد خودکار برچسب‌های متنی مطابق با محتوای بصری تصاویر دلالت دارد. در این پژوهش روشی بر مبنای روش‌های تکاملی برای این مسئله پیشنهاد شده است. الگوریتم ژنتیک هم به عنوانی گامی در پیش پردازش داده‌های تصویری و کمک به کاهش ابعاد بردارهای ویژگی در فضاهای گوناگون بصری و نیز در ترکیب نتایج حاشیه‌نویسی فضاهای ویژگی گوناگون پیشنهاد شده است.

در این پژوهش بر آنیم تا از هر دو نوع ویژگی‌های محلی و سراسری به منظور بهره بردن از مزایای هر دو نوع شیوه توصیف تصاویر استفاده کنیم. بنابراین چندین بردار ویژگی شامل ویژگی‌های محلی و ویژگی‌های سراسری از تصاویر استخراج شده‌اند. از آنجا که با چندین بردار ویژگی و در هر بردار با ابعاد بالایی از ویژگی‌ها مواجه هستیم، الگوریتم ژنتیک جهت انتخاب زیرمجموعه ویژگی‌های مؤثر به کار گرفته شده است. جهت دسته‌بندی تصاویر بدون برچسب به‌گونه‌ای که تنها تصاویر مرتبط به هر تصویر جدید در دسته‌بندی آن تصویر دخالت داده شوند، طی یک فرایند خوشه‌بندی دولایه بر مبنای محتوای بصری و هم‌چنین فضای معنایی، تصاویر مرتبط به هم در یک خوشه قرار می‌گیرند. الگوریتم ژنتیک جهت ترکیب معنایی و کارای نتایج فضاهای ویژگی مبنی بر این‌که هر فضای ویژگی در شناسایی بعضی برچسب‌های خاص سرآمد سایر فضاهای ویژگی می‌باشد و سپس ارائه نتیجه نهایی که همان برچسب‌های انتخاب شده برای تصویر بدون برچسب است پیشنهاد شده است.

جهت ارزیابی روش، دو دادگان شناخته شده‌ی Corel5k و IAPR TC-12 استفاده شده‌اند. روش پیشنهادی برای دادگان Corel5k سبب افزایش کارایی سامانه با در نظر گرفتن معیار F1 از ۰.۲۵ به ۰.۳۳۳ و برای دادگان IAPR TC-12 از ۰.۳ به ۰.۳۶۴ شده است. در ارزیابی عملکرد حاشیه‌نویسی، روش پیشنهادی در مقایسه با سایر روش‌ها روی دادگان Corel5k عملکردی معادل با بهترین روش و برای دادگان IAPR TC-12 بهبود داشته است. هم‌چنین روش پیشنهادی از نظر کیفیت رتبه‌بندی تصاویر (معیار میانگین دقت متوسط) در مقایسه با سایر روش‌ها روی دادگان Corel5k بهبودی نداشته است ولیکن در دادگان IAPR TC-12 نسبت به بهترین روش بهبود حاصل شده است.

کلمات کلیدی: بازیابی تصاویر، حاشیه‌نویسی خودکار تصاویر، حاشیه‌نویسی معنایی تصاویر، الگوریتم ژنتیک، دسته‌بندی چندبرچسبی، خوشه‌بندی دولایه

فهرست مطالب

صفحه	عنوان
ت.....	فهرست جدول‌ها
ث.....	فهرست شکل‌ها
۱.....	فصل ۱- کلیات
۲.....	۱-۱- تعریف مسئله
۲.....	۱-۱-۱- حاشیه‌نویسی متن-مبنای سنتی
۳.....	۲-۱-۱- بازیابی تصویر مبتنی بر محتوا
۵.....	۳-۱-۱- حاشیه‌نویسی تصاویر وب
۶.....	۴-۱-۱- حاشیه‌نویسی خودکار تصاویر
۸.....	۲-۱- چالش‌های حاشیه‌نویسی خودکار تصاویر
۸.....	۳-۱- کاربردها
۱۰.....	۴-۱- رویکرد پایان‌نامه و نوآوری‌ها
۱۱.....	۵-۱- ساختار پایان‌نامه
۱۲.....	فصل ۲- پیشینه‌ی پژوهش
۱۳.....	۱-۲- مقدمه
۱۳.....	۲-۲- ویژگی‌های بصری
۱۶.....	۱-۲-۲- یادگیری چند نمونه‌ای
۱۸.....	۲-۲-۲- تبدیل نمایش چند نمونه‌ای به نمایش تک نمونه‌ای
۲۰.....	۳-۲- الگوریتم‌های یادگیری
۲۰.....	۱-۳-۲- روش‌های مبتنی بر مدل بندی احتمالاتی
۲۳.....	۲-۳-۲- روش‌های مبتنی بر دسته‌بندی
۲۶.....	۴-۲- جمع بندی
۲۹.....	فصل ۳- روش پیشنهادی
۳۰.....	۱-۳- مقدمه
۳۱.....	۲-۳- ویژگی‌های بصری
۳۳.....	۳-۳- انتخاب ویژگی
۳۶.....	۴-۳- خوشه‌بندی دولایه بصری و معنایی
۳۷.....	۵-۳- دسته‌بندی چندبرچسبی
۳۹.....	۶-۳- ترکیب معنایی وزن‌دار

۴۶	قابلیت رتبه‌بندی	۷-۳
۴۷	جمع بندی	۸-۳
۴۸	ارزیابی نتایج	فصل ۴
۴۹	مقدمه	۱-۴
۴۹	دادگان‌های تصویری	۲-۴
۵۱	معیارهای ارزیابی	۳-۴
۵۱	ارزیابی حاشیه‌نویسی	۱-۳-۴
۵۲	ارزیابی رتبه‌بندی	۲-۳-۴
۵۳	تنظیم پارامترها و ارزیابی مراحل حاشیه‌نویسی	۴-۴
۵۳	پارامترهای الگوریتم ژنتیک	۱-۴-۴
۵۹	پارامترهای خوشه‌بندی	۲-۴-۴
۶۴	پارامترهای تعداد همسایه‌ها و طول حاشیه‌نویسی	۳-۴-۴
۶۹	نتایج ارزیابی و مقایسه	۵-۴
۷۴	پیچیدگی زمانی الگوریتم	۶-۴
۷۶	نتیجه‌گیری و کارهای آتی	فصل ۵
۷۷	نتیجه‌گیری	۱-۵
۷۹	کارهای آتی	۲-۵
۸۱	مراجع	
۸۴	واژه نامه انگلیسی به فارسی	

فهرست جدول‌ها

صفحه	عنوان
۲۷	جدول ۱-۲ : تعدادی از تحقیقات در حوزه حاشیه نویسی خودکار تصاویر
۳۲	جدول ۱-۳: ویژگی‌های بصری استفاده شده از [۲۳]
۵۰	جدول ۱-۴: مشخصات دادگان Corel5k و IAPR TC-12
۵۵	جدول ۲-۴: تأثیر انتخاب ویژگی بر روی ابعاد دادگان‌ها
۵۶	جدول ۳-۴: ارزیابی عملکرد انتخاب ویژگی روی هر دو دادگان
۵۸	جدول ۴-۴: ارزیابی عملکرد ترکیب معنایی وزن‌دار روی هر دو دادگان
	جدول ۵-۴: تفکیک تصاویر پس از خوشه‌بندی براساس فضای ویژگی شماره ۷ (توصیفگر hue.harris.3)
۵۹	روی دادگان Corel5k
	جدول ۶-۴: تفکیک تصاویر خوشه شماره ۴ از جدول ۴-۵ پس از خوشه‌بندی براساس فضای معنا روی دادگان Corel5k
۵۹	جدول ۷-۴: ارزیابی عملکرد خوشه‌بندی دولایه روی هر دو دادگان
۶۲	جدول ۸-۴: ارزیابی مرحله به مرحله روش پیشنهادی روی هر دو دادگان
۶۳	جدول ۹-۴: مقایسه عملکرد روش پیشنهادی با سایر روش‌ها روی دادگان Corel5k
۶۹	جدول ۱۰-۴: مقایسه عملکرد روش پیشنهادی با سایر روش‌ها روی دادگان IAPR TC-12
۷۱	جدول ۱۱-۴: میانگین دقت متوسط روی دادگان‌های Corel5k و IAPR TC-12
۷۳	

فهرست شکل‌ها

صفحه	عنوان
۳	شکل ۱-۱: مسئله سلیقه افراد در برچسب زدن تصاویر.....
۴	شکل ۲-۱: یک سامانه بازیابی تصویر مبتنی بر محتوا [۲].....
۵	شکل ۳-۱: متن‌های محیطی و کلمات کلیدی استخراج شده از یک تصویر وب [3].....
۹	شکل ۴-۱: برچسب‌های تولید شده برای تصویر سمت چپ توسط وب سایت ALIPR.com.....
۹	شکل ۵-۱: نتایج بازیابی برای پرس‌وجوی "zebra" از وب سایت ALIPR.com.....
۱۵	شکل ۱-۲: تناظر بین نوع ویژگی‌ها و مفاهیم. [۱۰].....
۱۶	شکل ۲-۲: تفاوت چهار چارچوب یادگیری [۴].....
۱۷	شکل ۳-۲: یک مثال از چارچوب چند نمونه‌ای چندبرچسبی در حاشیه‌نویسی تصاویر [۱۵].....
۱۸	شکل ۴-۲: تقسیم تصویر به بلوک‌های همپوشان [۲۶].....
۱۹	شکل ۵-۲: چارچوبی برای ترکیب نمایش چند نمونه‌ای و نمایش تک نمونه‌ای تصاویر [۹].....
۳۱	شکل ۱-۳: شمای کلی از سامانه پیشنهادی.....
۳۲	شکل ۲-۳: چگونگی ایجاد دو هیستوگرام برای هر تصویر.....
۳۵	شکل ۳-۳: نحوه سنجش برازندگی یک کروموزوم در انتخاب ویژگی.....
۳۷	شکل ۴-۳: نمایی از خوشه‌بندی دولایه و نحوه تعیین خوشه‌های معنایی برای تصاویر بدون برچسب ..
۳۹	شکل ۵-۳: سودکد الگوریتم DWML-kNN [۳۷].....
۴۲	شکل ۶-۳: بردارهای امتیاز حاصل از فضاها و ویژگی گوناگون برای تصویر جدید x و ترکیب جمعی ساده آنها برای حاشیه‌نویسی تصویر x
۴۳	شکل ۷-۳: کروموزومی نمونه در قالب ماتریسی که هر سطر آن یک بردار معنایی وزن‌دار برای یک فضای ویژگی می‌باشد.....
۴۴	شکل ۸-۳: ایجاد بردارهای امتیاز جدید حاصل از ضرب درایه‌های نظیر به نظیر بردار امتیاز یک فضای ویژگی برای تصویر جدید x از شکل ۶-۳ با بردار وزن‌دار معنایی آن فضا از کروموزوم شکل ۷-۳ و سپس ترکیب جمعی تمامی این برداری جدید برای حاشیه‌نویسی تصویر x
۴۶	شکل ۹-۳: نحوه حاشیه‌نویسی تصاویر بدون برچسب توسط سامانه پیشنهادی.....
۵۰	شکل ۱-۴: چند نمونه از تصاویر دادگان Core15k.....
۵۰	شکل ۲-۴: چند نمونه از تصاویر دادگان IAPR TC-12.....
۵۴	شکل ۳-۴: نمودارهای F1 حاصل به ازای مقادیر مختلف برای پارامترهای جمعیت اولیه، تولید نسل، احتمال تلفیق و احتمال جهش برای مرحله انتخاب ویژگی برای بردار ویژگی شماره ۱۱ (توصیفگر hsv.3) از دادگان Core15K.....

شکل ۴-۴: نمودارهای F1 حاصل به ازای مقادیر مختلف برای پارامترهای جمعیت اولیه، تولید نسل، احتمال تلفیق و احتمال جهش برای مرحله انتخاب ویژگی برای بردار ویژگی شماره ۱۱ (توصیفگر hsv.3) از دادگان IAPR TC-12 ۵۵

شکل ۵-۴: نمودارهای F1 حاصل به ازای مقادیر مختلف برای پارامترهای جمعیت اولیه، تولید نسل، احتمال تلفیق و احتمال جهش برای مرحله ترکیب معنایی وزن دار روی دادگان Corel5k ۵۷

شکل ۶-۴: نمودارهای F1 حاصل به ازای مقادیر مختلف برای پارامترهای جمعیت اولیه، تولید نسل، احتمال تلفیق و احتمال جهش برای مرحله ترکیب معنایی وزن دار روی دادگان IAPR TC-12 ۵۸

شکل ۷-۴: تاثیر تعداد خوشه‌ها در مجموع کل انحراف از معیارهای خوشه‌بندی برای فضای ویژگی شماره ۷ از دادگان Corel5k ۶۰

شکل ۸-۴: تاثیر تعداد خوشه‌ها در کارایی حاشیه‌نویسی برای فضای ویژگی شماره ۷ از دادگان Corel5k ۶۱

شکل ۹-۴: تاثیر کارایی مرحله به مرحله‌ی بخش‌های مختلف روش پیشنهادی روی دادگان Corel5k مطابق با جدول ۸-۴ ۶۳

شکل ۱۰-۴: تاثیر کارایی مرحله به مرحله‌ی بخش‌های مختلف روش پیشنهادی روی دادگان IAPR TC-12 مطابق با جدول ۸-۴ ۶۴

شکل ۱۱-۴: تاثیر پارامتر k در دسته‌بندی چندبرچسبی kNN بر عملکرد حاشیه‌نویسی برای فضای ویژگی شماره ۷ (توصیفگر hue.harris.3) در دادگان Corel5k ۶۵

شکل ۱۲-۴: تاثیر پارامتر k در دسته‌بندی چندبرچسبی kNN بر عملکرد حاشیه‌نویسی برای فضای ویژگی شماره ۱۱ (توصیفگر hsv.3) در دادگان Corel5k ۶۵

شکل ۱۳-۴: تاثیر پارامتر طول حاشیه‌نویسی بر معیارهای ارزیابی دقت، فراخوان و F1 روی دادگان Corel5k ۶۷

شکل ۱۴-۴: تاثیر پارامتر طول حاشیه‌نویسی بر معیارهای ارزیابی دقت، فراخوان و F1 روی دادگان IAPR TC-12 ۶۷

شکل ۱۵-۴: تاثیر پارامتر طول حاشیه‌نویسی بر معیار NZR روی دادگان Corel5k ۶۸

شکل ۱۶-۴: تاثیر پارامتر طول حاشیه‌نویسی بر معیار NZR روی دادگان IAPR TC-12 ۶۸

شکل ۱۷-۴: کلمات پیش‌بینی شده توسط سامانه پیشنهادی در دادگان Corel5k ۶۹

شکل ۱۸-۴: کلمات پیش‌بینی شده توسط سامانه پیشنهادی در دادگان IAPR TC-12 ۷۲

شکل ۱۹-۴: بازیابی معنایی نتایج روی دادگان IAPR TC-12 ۷۴

فصل ۱- کلیات

۱-۱- تعریف مسئله

در سال‌های اخیر با گسترش اینترنت و دسترس‌پذیری آسان وسایل ضبط تصاویر مانند دوربین رقی، دوربین موبایل و اسکنر، حجم تصاویر رقمی به شدت افزایش پیدا کرده است و به صورت انفجاری در حال رشد است. اکثر این تصاویر مخصوصاً مجموعه تصاویر شخصی بدون دسته‌بندی و بدون اطلاعات اضافی ذخیره می‌شوند. بنابراین نیاز به ابزاری کارا برای جستجو و بازیابی تصاویر بیش از پیش احساس می‌شود. در دو دهه‌ی اخیر تحقیقات زیادی در بخش بازیابی تصویر انجام شده است. رویکردهای کلی بازیابی تصویر به چهار دسته تقسیم می‌شوند [۱]: رویکرد اول حاشیه‌نویسی متن-مبنا^۱ سنتی، رویکرد دوم بازیابی تصویر مبتنی بر محتوا^۲، رویکرد سوم حاشیه‌نویسی تصاویر وب و رویکرد چهارم حاشیه‌نویسی خودکار تصاویر^۳.

۱-۱-۱ حاشیه‌نویسی متن-مبنای سنتی

این رویکرد در دهه ۱۹۷۰ مطرح شده است. در این رویکرد تصاویر به صورت دستی حاشیه‌نویسی می‌شوند یعنی انسان تصاویر را یکی یکی بررسی کرده و براساس محتوای آنها به تصاویر برچسب زده می‌شود. در این رویکرد پرس‌وجوی کاربر در قالب یک متن به سامانه داده می‌شود و سامانه با جستجو در پایگاه داده، تصاویر مربوطه را بازیابی می‌کند. دو مشکل اصلی برای این سامانه‌ها وجود دارد: مشکل نخست این که برچسب زدن تصاویر به صورت دستی توسط انسان زمان‌بر و هزینه‌بر است و در بعضی کاربردها مانند تصاویر ماهواره‌ای و تصاویر پزشکی برچسب زدن تقریباً ناممکن است. مشکل دوم درک مختلف افراد از یک تصویر واحد است، هر فرد بنا به سلیقه و درک خود برچسب‌های متفاوتی به یک

¹ Text based annotation

² Content based image retrieval (CBIR)

³ Automatic image annotation (AIA)

تصویر نسبت دهد. به عنوان مثال افراد مختلف برچسب‌هایی متفاوتی مانند شهر، خیابان، ماشین، درخت، آسمان، افراد، ساختمان و ... به شکل ۱-۱ نسبت می‌دهند.

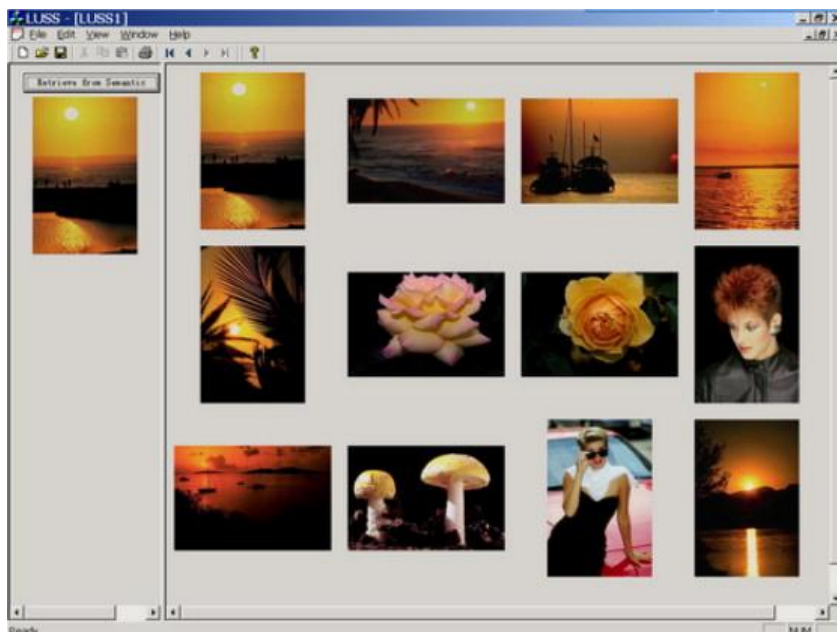


شکل ۱-۱: مسئله سلیقه افراد در برچسب زدن تصاویر

۱-۱-۲- بازیابی تصویر مبتنی بر محتوا

در رویکرد بازیابی تصویر مبتنی بر محتوا، تصاویر با محتوای بصری مانند رنگ و بافت بازیابی و شاخص‌گذاری می‌شوند. ساختار این سامانه‌ها به این صورت است که کاربر یک تصویر را به عنوان پرس‌وجو به سامانه ارائه می‌کند، سامانه ویژگی‌های بصری تصویر پرس‌وجو را استخراج کرده و با ویژگی‌های بصری تصاویر پایگاه داده مقایسه می‌کند و براساس تابع شباهت، فاصله‌ی تصویر پرس‌وجو و تصاویر پایگاه داده را محاسبه و شبیه‌ترین تصاویر را به کاربر نمایش می‌دهد. اصلی‌ترین مشکل سامانه‌های بازیابی تصویر مبتنی بر محتوا وجود شکاف معنایی^۱ بین ویژگی‌های سطح پایین و مفاهیم سطح بالا است. شکل ۱-۲ مثالی از یک سامانه بازیابی تصویر مبتنی بر محتوا را نشان می‌دهد که به ازای یک تصویر پرس‌وجو، چند تصویر مرتبط بازیابی شده‌اند.

^۱ Semantic gap



شکل ۱-۲: یک سامانه بازیابی تصویر مبتنی بر محتوا [۲]

برای کاهش شکاف معنایی در این سامانه‌ها از بازخورد ربط^۱ استفاده می‌شود. در بازخورد ربط کاربر تصاویر بازیابی شده را بررسی کرده و تصاویر مربوط و نامربوط را مشخص می‌کند. سامانه از روی این اطلاعات کاربر با الگوریتم‌های یادگیری ماشین سعی می‌کند مدل مفهومی که در ذهن کاربر است را یاد بگیرد و نتایج را بهبود بخشد.

اگرچه در سال‌های اخیر تحقیقات زیادی روی سامانه‌های بازیابی تصویر مبتنی بر محتوا تصاویر انجام شده است، اما این سامانه‌ها با مشکلاتی مواجه است که اصلی‌ترین آن، مشکل شکاف معنایی است. مشکل دیگر این است که به خاطر ابعاد بالای ویژگی‌های بصری، کارایی^۲ و مقیاس‌پذیری^۳ این سامانه‌ها کم است. هم‌چنین در این رویکرد کاربر باید یک تصویر به عنوان پرس‌وجو به سامانه ارائه کند که ممکن است تصویر کاملاً مناسب برای مفهومی که در ذهن دارد در دسترس نباشد.

در سامانه‌های بازیابی تصویر مبتنی بر محتوا، برای کاهش شکاف معنایی از بازخورد ربط استفاده می‌شود که کاربر باید تصاویر بازیابی شده را بررسی کرده و به آنها برچسب مثبت و منفی بزند که این کار از حوصله‌ی کاربران خارج است. مشکل دیگر این است که در بازخورد ربط یادگیری به صورت برخط انجام می‌شود و تمام تصاویر پایگاه داده دسته‌بندی و رتبه‌بندی می‌شوند که با افزایش اندازه‌ی پایگاه داده زمان یادگیری افزایش می‌یابد.

¹ Relevance feedback


² Efficiency

³ Scalability

۱-۱-۳- حاشیه‌نویسی تصاویر وب

تفاوت بازیابی تصاویر در حوزه‌ی وب با روش‌های قبل، داشتن اطلاعات اضافی در صفحه‌ی وب است که استخراج مفاهیم را آسان‌تر می‌کند. در صفحات وب معمولاً برای هر تصویر عنوان و یا توضیحاتی در مورد آن نیز موجود است. از طرف دیگر URL فایل تصویر اغلب دارای یک ساختار سلسله‌مراتبی واضح است و برخی از اطلاعات در مورد تصویر از جمله رده تصویر را دارد. علاوه بر این، سند HTML شامل برخی از اطلاعات مفید در عنوان تصویر، تگ ALT، متن توصیفی اطراف تصویر، لینک، و غیره است.

در این رویکرد عمدتاً از الگوریتم‌های متن کاوی^۱ برای استخراج کلمات کلیدی از اطلاعات اطراف تصویر استفاده می‌شود. شکل ۱-۳ یک مثال از کلمات کلیدی استخراج شده از متن‌های محیطی اطراف تصویر توسط سامانه پیشنهادی تحقیق [۳] را نشان می‌دهد. در این تحقیق دو نوع از متن‌های محیطی^۲ ET_{Normal} و ET_{Caption} استخراج شده است: ET_{Caption} شامل متونی هستند که توسط تگ‌های HTML، مانند آدرس تصویر و متن ALT تنظیم شده‌اند. از سوی دیگر، ET_{Normal} شامل متون عادی است که خارج از هر تگ HTML قرار دارند. ET_{Caption} به آسانی با استفاده از تگ‌های قابل استخراج است. اما استخراج ET_{Normal} مشکل است، چون تشخیص این که کدام قسمت از متن به تصویر مربوط است کار مشکلی است.

image	environmental texts	environmental keywords
	<p><i>ET-Normal:</i> "The Fleurieu Prize for Australian Landscape 2002." "by Joe Furlonger"</p> <p><i>ET-Caption:</i> "../images/2002winner_furlonger.jpg"</p>	Fleurieu Prize Australian Landscape Joe Furlonger images winner jpg

شکل ۱-۳: متن‌های محیطی و کلمات کلیدی استخراج شده از یک تصویر وب [3]

مشکل این رویکرد این است که حاشیه‌نویسی با استفاده از اطلاعات صفحه‌ی وب حاوی تصویر، مستقل از محتوای تصویر است، بنابراین چنانچه ارتباط بین متن اطراف تصویر و محتوای تصویر ضعیف باشد کلمات استخراج شده توصیف‌کننده‌ی محتوای تصویر نخواهند بود. هم‌چنین برای بعضی از تصاویر

¹ Text mining

² Environmental Text

ممکن است هیچ توضیح و متنی موجود نباشد که در این صورت نمی‌توان با این رویکرد آنها را حاشیه‌نویسی کرد.

۱-۱-۴- حاشیه‌نویسی خودکار تصاویر

حاشیه‌نویسی خودکار تصاویر، فرایند اضافه کردن متاداده^۱ در قالب عنوان یا کلمات کلیدی به تصاویر توسط سامانه کامپیوتری است. حاشیه‌نویسی تصاویر معمولاً برای دسترسی و جستجوی ساده و سریع تصاویر انجام می‌شود. حاشیه‌نویسی دستی تصاویر به دلیل سلیقه‌های مختلف افراد و همچنین هزینه بر و زمان بر بودن برای پایگاه داده‌های بزرگ عملی نیست. همچنین در بعضی تصاویر وب ممکن است ارتباط بین متن و تصویر ضعیف باشد و یا متنی موجود نباشد.

متاداده‌های که تصویر یا ویدیو با آن حاشیه‌نویسی می‌شوند به دو دسته تقسیم می‌شوند: محتوا-مستقل^۲ و محتوا-توصیف^۳. متاداده‌های محتوا-مستقل اطلاعاتی هستند که به محتوای تصویر مربوط هستند ولی مستقیماً آن را شرح نمی‌دهند، مانند نام عکاس، نام محل، تاریخ و متاداده‌های محتوا-توصیف اطلاعاتی هستند که مستقیماً به محتوای تصویر مربوط هستند و از روی ویژگی‌های سطح پایین و یا ارتباط معنایی بین مؤلفه‌های تصویر قابل استخراج هستند. استخراج اطلاعات محتوا-مستقل از محتوای تصویر تقریباً غیر ممکن است بنابراین تمرکز تحقیقات حاشیه‌نویسی تصاویر روی استخراج متاداده‌های محتوا-توصیف است.

ایده‌ی اصلی روش‌های حاشیه‌نویسی خودکار این است که از روی تعداد زیادی تصاویر حاشیه‌نویسی شده مدل مفاهیم معنایی یاد گرفته شود و از این مدل برای حاشیه‌نویسی تصاویر جدید استفاده کنند. برخلاف رویکرد بازیابی تصویر مبتنی بر محتوا، در این رویکرد پس از آن که مدل مفاهیم یاد گرفته شد، تصاویر به صورت آفلاین حاشیه‌نویسی می‌شوند و با یک فایل معکوس^۴ شاخص‌گذاری می‌شود. هنگام بازیابی تنها یک جستجوی ساده برای پیدا کردن کلمه‌ی موردنظر کاربر انجام می‌شود.

در این چارچوب پرس‌وجوها به صورت متنی هستند که مستقیماً با مفاهیم سطح بالا مرتبط هستند. این رویکرد مشکلات رویکرد بازیابی تصویر مبتنی بر محتوا و رویکرد حاشیه‌نویسی متن-مبنای سنتی را ندارد و کار با آن برای کاربران راحت است. همچنین روش‌های بازیابی با متن موجود به راحتی با این رویکرد وفق پذیر هستند.

^۱ Metadata

^۲ Content-independent

^۳ Content-descriptive

^۴ inverted file

ورودی سامانه حاشیه‌نویسی، تصویر و خروجی آن برچسب‌های متنی است که به تصویر ورودی زده می‌شود. بسته به این که به هر تصویر چند برچسب متنی منتسب می‌شود، روش‌های حاشیه‌نویسی به دو دسته‌ی حاشیه‌نویسی تک برچسبی^۱ و حاشیه‌نویسی چندبرچسبی^۲ تقسیم می‌شوند [۱].

✓ حاشیه‌نویسی تک برچسبی

در حاشیه‌نویسی تک برچسبی به هر تصویر تنها یک برچسب زده می‌شود. به عبارت دیگر در این روش‌ها چند رده^۳ (مفهوم) داریم که هر تصویر در یکی از آنها قرار می‌گیرد به همین دلیل به این روش‌ها، رده بندی تصویر^۴ نیز گفته می‌شود. برای این منظور ویژگی‌های سطح پایین تصاویر استخراج شده و به یک دسته‌بند^۵ داده می‌شود. دسته‌بند از روی دادگان آموزشی مدل مفاهیم را یاد می‌گیرد و تصاویر جدید را دسته‌بندی می‌کند. الگوریتم‌های دسته‌بندی شامل ماشین بردار پشتیبان^۶ (SVM)، شبکه عصبی مصنوعی^۷ (ANN) و درخت تصمیم^۸ (DT) به طور گسترده استفاده شده‌اند [۱].

✓ حاشیه‌نویسی چندبرچسبی

در حاشیه‌نویسی چندبرچسبی به هر تصویر چندین برچسب زده می‌شود. یک تصویر شامل اطلاعاتی زیادی است، ممکن است چندین شیء در یک تصویر وجود داشته باشند که توان آنها را با یک کلمه یا مفهوم نشان داد. بنابراین روش‌های حاشیه‌نویسی تک برچسبی که در آن به هر تصویر تنها یک برچسب زده می‌شود برای کاربردهایی مانند موتور جستجو مناسب نیستند.

برای حاشیه‌نویسی چندبرچسبی باید از روش‌های یادگیری چندبرچسبی استفاده کرد. تعریف رسمی یادگیری چندبرچسبی به این صورت است: هدف یادگیری تابع $f_{MLL}: X \rightarrow 2^Y$ از دادگان مفروض $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ است که X مجموعه‌ی نمونه‌ها، x_i یک نمونه $(x_i \in X)$ ، Y مجموعه‌ی برچسب‌ها و $y_i \subseteq Y$ برچسب‌های x_i را نشان می‌دهد [۴].

در اکثر مراجع به جای حاشیه‌نویسی تک برچسبی، از رده بندی تصویر و به جای حاشیه‌نویسی چندبرچسبی، از حاشیه‌نویسی استفاده می‌شود. در این پایان‌نامه نیز منظور از حاشیه‌نویسی، حاشیه‌نویسی چندبرچسبی است.

¹ Single labelling annotation

² Multi-labelling annotation

³ Category

⁴ Image categorization

⁵ Classifier

⁶ Support vector machine

⁷ Artificial neural network

⁸ Decision tree