دانشگاه الزهرا
دانشکده فنی و مهندسی

پایان نامه
جهت اخذ درجه کارشناسی ارشد
رشته مهندسی کامپیوتر گرایش هوش مصنوعی

عنوان

بازیابی تصاویر پزشکی مبتنی بر محتوا

استاد راهنما
دکتر رضا عزمی

دانشجو
ندا ناصری

آبان ماه سال ۱۳۹۱

بسم الله الرحمن الرحيم

دانشگاه الزهرا
دانشکده فنی و مهندسی

پایان نامه
جهت اخذ درجه کارشناسی ارشد
رشته مهندسی کامپیوتر گرایش هوش مصنوعی

عنوان

بازیابی تصاویر پزشکی مبتنی بر محتوا

استاد راهنما
دکتر رضا عزمی

استاد مشاور
دکتر رباب انبیایی

دانشجو
ندا ناصری

آبان ماه سال ۱۳۹۱

کلیه دستاوردهای این تحقیق متعلق به

دانشگاه الزهرا (س) است.

تقدیم به

همسر مهربانم

به پاس حمایت های همیشگی اش

و پدر و مادر عزیزم

به خاطر لطف و مهر بی پایانشان

با سپاس از

جناب آقای دکتر عزمی و سرکار خانم دکتر انبیایی

که با راهنمایی و حمایت های بی دریغ خود یاری ام نمودند

چکیده

رشد نمایی تولید تصاویر، نیاز به تولید ابزارهای نو برای مدیریت، بازیابی و نمایش دادن آنها را برای کاربردهای مختلف ایجاب می‌کند. متدهای جست وجوی مبتنی بر متن که به طور گسترده‌ای به کار می‌روند، به خوبی جواب‌گوی مقاصد بازیابی نیستند بلکه تحقق این خواسته، متدهای دسترسی‌ای را می‌طلبد که پرس‌وجوهایی فراتر از پرس‌وجوهای مبتنی بر متن ساده یا تقاضاهایی فراتر از تطابق دقیق فیلدهای پایگاه‌داده را فراهم کند. بازیابی تصویر مبتنی بر محتوا یک شاخه‌ی تحقیقی از بینایی ماشین است که در چند دهه‌ی اخیر بدین منظور توسعه یافته است.

در حوزه‌ی پزشکی تصاویر و به ویژه تصاویر دیجیتال در حجم رو به رشدی تولید می‌شوند و در تشخیص و درمان به کار می‌روند. مدیریت و دسترسی به این انبار تصویر بزرگ رفته‌رفته پیچیده‌تر می‌شود که پژوهش‌های خاص نیازمندیهای این حوزه را می‌طلبد.

در این پروژه ترکیبی از ضرایب موجک و ویژگی‌های LBP برای استخراج بردار ویژگی استفاده شده است. برای بهبود شباهت معنایی از بازخورد ارتباط کاربر استفاده شده است. برای بدست آوردن متد بازیابی بهینه ترکیبی از چند روش یادگیری ماشین به کار رفته است. قابل ذکر است که این روش‌ها به جای معیار فاصله‌ی ساده که در الگوریتم‌های کلاسیک به کار می‌رفت ارائه شده است. روش پیشنهادی بر روی تصاویر ماموگرافی پایگاه داده‌ی MIAS پیاده‌سازی شده است.

واژگان کلیدی: بازیابی تصاویر پزشکی مبتنی بر محتوا، بازخورد ارتباط، تبدیل موجک، کلاس‌بند SVM

فهرست مطالب

فصل اول: مقدمه ۱

۱-۱ مقدمه ۲

۱-۲ سیر تغییرات سیستم‌های بازیابی تصویر و ساختار سیستم ۳

۱-۳ سیستم‌های بازیابی تصویر مبتنی بر محتوا ۵

۱-۴ اصطلاحات مرتبط ۶

۱-۵ سازماندهی مطالب ۷

فصل دوم: بازیابی تصویر مبتنی بر محتوا ۹

۲_۱ مقدمه ۱۰

۲-۲ بلوک‌های سازنده‌ی یک سیستم CBIR ۱۰

۲-۳ استخراج ویژگی و نمایش آن ۱۱

۲-۳-۱ ویژگی‌های محلی و سراسری ۱۲

۲-۳-۲ رنگ ۱۳

۲-۳-۳ بافت ۱۳

۲-۳-۴ شکل ۱۴

۲-۳-۵ موقعیت مکانی ۱۵

۲-۴ قطعه بندی تصویر ۱۶

۲-۵ شاخص‌گذاری تصویر و کاهش بعد فضای ویژگی ۱۸

۲-۶ پرس‌وجو ۱۹
۲-۶-۱ سطوح معنایی جست‌وجو ۲۰
۲-۶-۲ انواع هدف در جست‌وجو ۲۱
۲-۶-۳ فرمول‌سازی (پردازش) پرس‌وجو ۲۲
۲-۷ اندازه‌ی شباهت ۲۴
۲-۸ مفهوم شکاف معنایی ۲۸
۲-۹ روش‌های کاهش شکاف معنایی ۲۹
۲-۹-۱ هستان‌شناسی شئ ۲۹
۲-۹-۲ یادگیری ماشین ۳۱
۲-۹-۳ بازخورد ارتباط (RF) ۳۷
فصل سوم: بازیابی تصاویر پزشکی ۴۲
۳-۱ مقدمه ۴۳
۳-۲ فیلدهای کاربرد پزشکی و فایده‌های بالینی ۴۴
۳-۳ استخراج ویژگی ۴۵
۳-۳-۱ متن ۴۵
۳-۳-۲ ویژگی‌های دیداری ۴۶
۳-۴ ارزیابی‌های سیستم ۵۰
۳-۵ پایگاه‌داده‌های مورد استفاده ۵۳
۳-۶ کارآیی ۵۴
۳-۷ چالش‌ها در CBIR ۵۵
۳-۷-۱ محتوا ۵۵
۳-۷-۲ ویژگی ۵۶
۳-۷-۳ کارآیی ۵۷
۳-۷-۴ قابلیت استفاده ۵۸
۳-۸ چالش‌ها و فرصت‌های CBIR در رادیولوژی ۶۱
فصل چهارم: سیستم پیشنهادی و ارزیابی ۶۵
۴-۱ مقدمه ۶۶
۴-۲ تبدیل موجک گسسته (DWT) ۶۷
۴-۳ الگوی دودویی محلی (LBP) ۷۰
۴-۳-۱ الگوی دودویی محلی یکنواخت ۷۲

۴-۳-۲ الگوهای دودویی محلی مستقل از چرخش..... ۷۳
۴-۴ دسته‌بندی توسط ماشین‌های بردار پشتیبان (SVM) ۷۴
۴-۵ روش Rocchio ۷۷
۴-۶ تست t ۷۸
۴-۷ الگوریتم پیشنهادی ۷۹
۴-۸ آزمایش ۸۱
۴-۹ ارزیابی ۸۲
فصل پنجم: جمع‌بندی و چشم‌انداز آینده ۸۵
جمع‌بندی ۸۶
چشم‌انداز آینده و پیشنهادات ۸۷
مراجع ۸۸

فهرست شکل‌ها و جداول

شکل ۱_۱ نمونه‌ای از سرآیند DICOM ۷

شکل ۲-۱ نمودار جریان فرآیند یک سیستم CBIR ۱۱

شکل۲-۲: نمودار سیستم CBIR با متد CLUE ۳۶

شکل ۲-۳: نمودار یک سیستم CBIR با بازخورد ارتباط ۳۸

جدول ۳-۱: خلاصه‌ی ویژگی‌های مورد استفاده‌ی سیستم‌های بازیابی پزشکی ۵۰

جدول ۳-۲: روش‌های برخورد با برخی چالش‌های سیستم‌های CBIR ۶۰

شکل ۴-۱: شمای کلی یک سیستم بازیابی تصویر ۶۶

شکل ۴-۲: ساختار درختی تبدیل موجک دوبعدی برای تصاویر ۶۸

شکل ۴-۳: نمونه‌ای از تجزیه‌ی موجک تصویر ۶۹

شکل ۴-۴: شمای تجزیه‌ی موجک چندسطحی ۷۰

شکل ۴-۵: محاسبه‌ی کد LBP برای یک پیکسل ۷۱

شکل ۴_۶ نحوه‌ی انتخاب الگوی دودویی محلی مستقل از چرخش ۷۳

شکل ۴_ ۷ جداسازی دوکلاس با ماشین بردار پشتیبان ۷۶

شکل ۴_۸ حرکت پرس‌وجو در فضای ویژگی در روش Rocchio ۷۸

شکل ۴-۹: شمای الگوریتم مورد استفاده در سیستم بازیابی ۸۰

شکل ۴_۱۰ برش ناحیه‌ی مطلوب از تصویر اولیه ۸۱

جدول ۴-۱: ارزیابی روش مورد استفاده بر اساس شماره‌ی تکرار ۸۳

شکل ۴-۱۱: دقت بازیابی طبق معیار precision بر اساس شماره تکرار الگوریتم‌ها ۸۳

شکل ۴-۱۲: منحنی ROC مربوط به نتایج بازیابی ۸۴

فصل اول:

مقدمه

۱-۱ مقدمه

ما در عصر تکنولوژی اطلاعات چندرسانه‌ای زندگی می‌کنیم. امروزه کاهش شدید هزینه‌ی ذخیره‌سازی، دسترسی گسترده به دستگاه‌های دیجیتال و وب جهانی، به رشد تولید محتوای چندرسانه‌ای و توزیع آن در سراسر جهان شتاب داده است. در میان تنوع محتواهای چندرسانه‌ای، شایع‌ترین و پرکاربردترین آنها تصاویر هستند. پیشرفت تکنولوژی این امکان را می‌دهد که تصاویری با استفاده از فضاپیماها حتی میلیون‌ها مایل دورتر از زمین تولید شوند. مراکز بیمارستانی و پزشکی همه روزه تعداد رو به رشدی از تصاویر دیجیتال از وجوه مختلف را برای تصمیم‌سازی بالینی و اهداف پژوهشی تولید می‌کنند. تکنولوژی‌های مورد نیاز مثل دوربین‌های دیجیتال، تلفن‌های همراه چندرسانه‌ای و کامپیوترهای شخصی، با قیمت مناسب برای استفاده‌ی عمومی در دسترس هستند. در نتیجه، تولید و ذخیره‌سازی تصاویر دیجیتال شخصی مثل عکس‌های تعطیلات و عکس‌های دوستان و اعضای خانواده ساده‌تر شده است. علاوه بر این مثال‌ها، زمینه‌های بسیار دیگری برای تولید تصاویر وجود دارند.

رشد نمایی تصاویر، نیاز به تولید ابزارهای نو برای مدیریت، بازیابی و نمایش دادن آنها را برای کاربردهای مختلف ایجاب می‌کند. این ابزارها در زمینه‌های مختلفی مثل سنجش از راه دور[1]، مد، پیشگیری از جرایم، نشر، پزشکی، معماری و غیره مورد نیاز کاربران است.

متدهای جست وجوی مبتنی بر متن که به طور گسترده‌ای به کار می‌روند، به خوبی جواب‌گوی مقاصد بازیابی نیستند بلکه تحقق این خواسته، متدهای دسترسی‌ای را می‌طلبد که پرس‌وجوهایی فراتر از پرس‌وجوهای مبتنی بر متن ساده یا تقاضاهایی فراتر از تطابق دقیق فیلدهای پایگاه‌داده را فراهم کند. بازیابی تصویر مبتنی بر محتوا یا CBIR[2] یک شاخه‌ی تحقیقی از بینایی ماشین است که در دو دهه‌ی اخیر توسعه یافته است. برنامه‌ها و ابزارهای زیادی در

[1] Remote Sensing
[2] Content Based Image Retrieval

این مدت برای فرمول‌بندی و اجرای پرس‌وجوهای مبتنی بر محتوای دیداری یا شنیداری و یا کمک به جست‌وجو در انبارهای بزرگ چندرسانه‌ای تولید شده‌اند اما هنوز هیچ راه‌حل کلی با توجه به تنوع زیاد پایگاه‌داده‌ها با انواع مختلف اسناد و ویژگی‌های مختلف بدست نیامده است. همچنین سؤالات بسیاری راجع به سرعت، توصیف‌کننده‌های معنایی یا تفسیر تصویر هدف بدون پاسخ مانده است.

در حوزه‌ی پزشکی تصاویر و به ویژه تصاویر دیجیتال در حجم رو به رشدی تولید می‌شوند و در تشخیص و درمان به کار می‌روند. تعداد تصاویر پزشکی تولید شده به شدت در حال افزایش است. به عنوان مثالی از حجم تصاویر تولیدی بخش پزشکی می‌توان به دپارتمان رادیولوژی بیمارستان جنیوا[1] اشاره کرد که تنها در سال ۲۰۰۲، در هر روز ۱۲۰۰۰ تصویر تولید شده است که این میزان رو به رشد می‌باشد. ویدیو و تصاویر تولیدشده در قلب‌شناسی و آندوسکوپی، منابع داده‌ی بسیار بزرگی هستند که در نظر است به صورت PACS[2] مجتمع شوند. مدیریت و دسترسی به این انبار تصویر بزرگ رفته‌رفته پیچیده‌تر می‌شود. بیشتر دسترسی به این سیستم‌ها مبتنی بر هویت بیمار، یا خصوصیات پژوهشی است که در استاندارد DICOM[3] نیز تعیین شده است.

۱-۲ سیر تغییرات سیستم‌های بازیابی تصویر و ساختار سیستم

نسل اول سیستم‌های بازیابی تصویر در اواخر دهه‌ی ۱۹۷۰ تولید شدند که بیشتر شبیه سیستم‌های مدیریت پایگاه‌داده‌ی سنتی یا سیستم‌های بازیابی متن بودند [12]. در سیستم‌های اولیه از تفسیر[4]های دستی به عنوان عبارات شاخص‌گذاری استفاده می‌شد که این تفسیرها هم

[1] Geneva
[2] Picture Archiving and Communication System
[3] Digital Imaging and Communications in Medicine
[4] annotation

محتوای تصویر و هم دیگر فراداده‌ها مثل نام فایل، قالب تصویر، تولیدکننده‌ی آن، تاریخ و غیره را بیان می‌کرد. متأسفانه انتساب دستی صفات متنی زمان‌بر و هزینه‌بردار است. وقتی پایگاه‌داده بسیار بزرگ و پویا باشد، مثل تصاویر موجود در وب، تفسیر دستی همه‌ی تصاویر تقریباً غیرممکن است. به علاوه، از آنجا که تفسیرهای تصاویر ممکن است بسته به زمینه‌ی بررسی آن متفاوت باشد، کلمات کلیدی مبتنی بر تفسیر هم طبیعتاً وابسته به موضوع است. مسئله‌ی دیگر این است که توصیف برخی خصوصیت‌های دیداری تصاویر (مثل بافت‌های خاص و شکل‌ها) با متن مشکل و گاهی غیرممکن است. آنچه که در این جا لازم است، به کار بردن توصیف محسوس‌تری از محتوای دیداری است که رابطه‌ی نزدیک‌تری با درک انسان داشته باشد [2, 3].

برای غلبه بر محدودیت‌های بازیابی تصویر مبتنی بر متن، در اوایل دهه‌ی ۱۹۹۰ سیستم‌های بازیابی تصویر مبتنی بر محتوا (CBIR) پدید آمدند. در CBIR، تصاویر به طور خودکار یا نیمه-خودکار، با استفاده از ویژگی‌هایی که مستقیماً از محتوای دیداری آنها بدست آمده‌اند، شاخص‌گذاری می‌شوند. این ویژگی‌های دیداری با تکنیک‌های پردازش تصویر بدست می‌آیند. این شاخص‌گذاری تصویر، با شاخص‌گذاری مبتنی بر کلیدواژه‌ی تفسیرهای تصویر تفاوت اساسی دارد چرا که در این روش، صفات مطلوب نواحی یا کل تصویر، توابع پیچیده‌ای در مقیاس پیوسته هستند. فرآیند متداول در CBIR به طور خلاصه به این شرح است [3, 13]:

- برای استخراج ویژگی‌های سطح پایین، از تکنیک‌های پردازش تصویر و بازشناسی الگو استفاده می‌شود.
- برای یک ویژگی، نمایشی به یک فرم برداری و یک مفهوم شباهت تعیین می‌شود و تصویر به صورت مجموعه‌ای از ویژگی‌ها بیان می‌شود.

- در نهایت بازیابی‌های تصویر بر مبنای محاسبه‌ی شباهت در فضای ویژگی انجام می‌شود و نتایج بر اساس مقدار شباهت محاسبه‌شده، مرتب می‌شوند.

۱-۳ سیستم‌های بازیابی تصویر مبتنی بر محتوا

هر چند جرقه‌ی سیستم‌های نخستین در آغاز دهه‌ی ۱۹۸۰ زده شد ولی اکثر مقالات به سیستم‌هایی مثل QBIC[1] [32] محصول IBM، به عنوان سرآغاز بازیابی تصویر مبتنی بر محتوا اشاره می‌کنند. قطعاً QBIC [1993] سرشناس‌ترین سیستم تجاری بازیابی تصویر است. دیگر سیستم تجاری برای بازیابی تصویر و ویدیو، Virage [1996] است که مشتریان سرشناسی مثل CNN دارد. اکثر سیستم‌های موجود، حاصل کار آکادمیک هستند. معروف‌ترین آنها Candid [1995]، Photobook [1996]، NeTra [1997] هستند که همه از ویژگی‌های بافت و رنگ ساده‌ای برای تشریح محتوای تصویر استفاده می‌کنند. استفاده از ویژگی‌های سطح بالاتر مثل بخش‌های قطعه‌بندی‌شده‌ی تصویر برای پرس‌وجو، با سیستم Blobworld [1998] معرفی شد. PicHunter [2000] یک مرورگر تصویر است که به کاربر کمک می‌کند تا یک تصویر را در پایگاه داده به صورت دقیق پیدا کند. این کار با نمایش تصاویر به کاربر به طوری که بهبود اطلاعات[2] حداکثر شود انجام می‌شود. برخی سیستم‌ها مثل Viper، WIPE یا Compass به صورت نسخه‌های تجربی روی وب موجود هستند. اکثر این سیستم‌ها معماری بسیار مشابهی برای پیمایش[3]، بایگانی و شاخص‌گذاری تصویر دارند که شامل ابزارهایی برای استخراج ویژگی‌های دیداری، ذخیره و بازیابی کارآی این ویژگی‌ها، اندازه‌گیری فاصله یا محاسبه‌ی شباهت و یک نوع واسط گرافیکی کاربر است.

[1] Query By Image Content
[2] Information Gain
[3] browsing

۱-۴ اصطلاحات مرتبط

PACS: در تصویربرداری پزشکی، برای فراهم آوردن ذخیره‌سازی به‌صرفه، بازیابی سریع تصویر، دسترسی به تصاویری که از وجوه مختلف گرفته شده‌اند و دسترسی هم‌زمان به سایت‌های مختلف، سیستم‌های بایگانی و انتقال تصویر توسعه یافتند. تصاویر و گزارش‌های الکترونیکی به صورت دیجیتال از طریق PACS فرستاده می‌شوند. بدین ترتیب نیاز به فایل‌های دستی، بازیابی یا حمل‌ونقل پوشه‌های فایل از بین می‌رود. قالب فراگیر برای ذخیره و انتقال تصاویر PACS، DICOM است. داده‌ی غیرتصویری مثل اسناد اسکن‌شده می‌توانند با استفاده از قالب‌های استاندارد تجاری مثل PDF در DICOM تعبیه شوند. یک PACS از چهار بخش اصلی تشکیل شده است: یک جنبه‌ی تصویربرداری مثل CT[۱] یا MRI[۲]، یک شبکه‌ی امن برای انتقال اطلاعات بیماران، ایستگاه‌های کاری برای تفسیر و بازبینی تصاویر و در نهایت یک بایگانی برای ذخیره و بازیابی تصاویر و گزارش‌ها. با ترکیب تکنولوژی وب، PACS این قابلیت را دارد که دسترسی به‌موقع و کارآیی به تصاویر، تفسیرها و داده‌های مربوط داشته باشد.

DICOM: استاندارد تصویربرداری دیجیتال و مخابره در پزشکی است که توسط انجمن تولیدکنندگان الکتریکی ملی (NEMA[۳]) به همراه کالج رادیولوژی امریکا (ACR[۴]) تولید شده است. این استاندارد بیشتر قالب‌های تصویری پزشکی را در بر می‌گیرد و ویژه‌ی انتقال اطلاعات و مخابره بین ماشین‌های تصویربرداری است. یک فایل DICOM از دو بخش تشکیل شده است. بخش اول، سرآیندی است که شامل اطلاعاتی از قبیل نام بیمار، نوع اسکن، ابعاد تصویر و غیره

[1] Computerised Tomography
[2] Magnetic Resonance Imaging
[3] National Electrical Manufacturers Association
[4] American College of Radiology

است. بخش دوم داده‌ی تصویری است که می‌تواند به فرم فشرده‌شده (bitmap) یا غیرفشرده (jpeg, gif, ...) باشد. نمونه‌ای از سرآیند DICOM در شکل ۱_۱ مشاهده می‌شود.

```
First 128 bytes: unused by DICOM format
Followed by the characters 'D','I','C','M'
This preamble is followed by extra information e.g.:

0002,0000,File Meta Elements Group Len: 132
0002,0001,File Meta Info Version: 256
0002,0010,Transfer Syntax UID: 1.2.840.10008.1.2.1.
0008,0000,Identifying Group Length: 152
0008,0060,Modality: MR
0008,0070,Manufacturer: MRIcro
0018,0000,Acquisition Group Length: 28
0018,0050,Slice Thickness: 2.00
0018,1020,Software Version: 46\64\37
0028,0000,Image Presentation Group Length: 148
0028,0002,Samples Per Pixel: 1
0028,0004,Photometric Interpretation: MONOCHROME2.
0028,0008,Number of Frames: 2
0028,0010,Rows: 109
0028,0011,Columns: 91
0028,0030,Pixel Spacing: 2.00\2.00
0028,0100,Bits Allocated: 8
0028,0101,Bits Stored: 8
0028,0102,High Bit: 7
0028,0103,Pixel Representation: 0
0028,1052,Rescale Intercept: 0.00
0028,1053,Rescale Slope: 0.00392157
7FE0,0000,Pixel Data Group Length: 19850
7FE0,0010,Pixel Data: 19838
```

شکل ۱_۱ نمونه‌ای از سرآیند DICOM

۱-۵ سازماندهی مطالب

در ادامه‌ی این پژوهش و در فصل دوم سیستم بازیابی تصویر مبتنی بر محتوا و ساختار آن معرفی می‌شود. سپس بخش‌های مختلف این سیستم در ادامه‌ی فصل تشریح می‌شوند. انواع ویژگی‌های به کار رفته، نحوه‌ی ارائه‌ی پرس‌وجو، جنبه‌ها و انگیزه‌های جست‌وجو، ملاک‌های تعیین شباهت و روش بازیابی، مفهوم شکاف معنایی و روش‌های برخورد با آن و دیگر مطالب مربوطه بررسی می‌شوند.

در فصل سوم به معرفی اهداف و کاربردهای این تکنولوژی در حوزه‌ی پزشکی می‌پردازیم. مسائل مربوط به این دامنه‌ی خاص از نیازهای این حوزه گرفته تا ویژگی‌های تصاویر پزشکی، رویکردهای ارائه‌شده و جنبه‌های قابل بررسی بیان می‌شود.

در فصل چهارم سیستم بازیابی پیشنهادی و روش‌های مورد استفاده تشریح گردیده است و در پایان این فصل به ارزیابی این سیستم پرداخته شده است.

در بخش پایانی، فهرستی از مراجع مورد استفاده و منابع مفید ارائه شده است.