

در هم سازی چند رسانه ای و شبکه بندی

بسیاری از شرکت های اینترنتی به طور مکرر، داده های چند رسانه ای ناهمگن را مدیریت می کنند. وب سایت های رسانه های اجتماعی از جمله فیسبوک، توئیتر و یوتیوب همراه با برنامه های موبایل نظیر اینستاگرام، اسنپ چت و وی چت، همگی با مسئله مشابهی روبرو بوده اند- ذخیره سازی کارآمد آن ها، نمایه بندی، جست و جو، تحلیل و درک داده های چندرسانه ای چگونه است؟ از این روی، هدف ما حل مسئله با مطالعه دو موضوع رایج در چند رسانه ای: هش کردن و شبکه بندی می باشد.

هش چند رسانه ای

ما به بررسی دو روش مختلف مرتبط با هش چند رسانه ای- هش مبتنی بر یادگیری سطحی و هش مبتنی بر یادگیری عمیق- پرداخته و فنون پیشرفته را برای ذخیره سازی چند رسانه ای، نمایه بندی و بازیابی می کنیم.

هش کردن (در هم سازی) با یادگیری سطحی

هشینگ توجه زیادی را از طرف محققان و متخصصان در زمینه کامپیوتر، یادگیری ماشینی، داده کاوی، بازیابی اطلاعات و سایر زمینه ها، جلب کرده است. طیف وسیعی از فنون درهم سازی و هشینگ برای کد گذاری اسناد، تصاویر، ویدئو و سایر انواع داده ها به مجموعه ای از کد های دو دویی توسعه یافته است، که موجب حفظ تشابهات در میان داده های اولیه می شود. با این کد های دو دویی، جست و جو های تشابه را می توان در دیتاست ها به دلیل کارایی بالای مقایسه زوجی با استفاده از فاصله همینگ انجام داد.

تلاش های اولیه در هشینگ، بر استفاده از تغییرات تصادفی در ساخت دستورات هش متمرکز بوده است. معرف ها و نمونه های شناخته شده نظیر هشینگ حداقل و هشینگ حساس به محل وجود داشته است. مینهش، تشابه جاکارد را برآورد کرده است در حالی که LSH، متریک های تشابه و فواصل مختلف را در نظر می گیرد نظیر فاصله ℓ_p برای $p \in (0, 2]$ ، تشابه کوسینوس و تشابه هسته ای. به دلیل هشینگ تصادف، بیت های بیشتری در هر جدول هش یا درهم سازی برای دست یابی به دقت بالا نیاز است. این معمولا موجب کاهش فراخوانی شده و از

این روی جداول درهم سازی برای دست یابی به صحت بالای نزدیک ترین همسایه بازیابی شده نیاز است. تعداد کل بیت های هش مورد استفاده در یک زمینه را می توان تا هزار ، تخمین زد.

در فراتر از طرح های هشینگ تصادفی مستقل از داده ها، یک روند اخیر در یادگیری ماشینی، توسعه فنون هشینگ وابسته به داده ها است تا یک مجموعه از کد های هش فشرده بر اساس مجموعه داده های آموزشی یاد گرفت. کد های دو دویی در این سناریو به دلیل سادگی و صحت محاسبه، محبوب بوده است. طرح هشینگ فشرده می تواند به جست و جوی نزدیک ترین همسایه پس از کد گذاری کامل مجموعه داده ها در کد های دو دویی کوتاه و ترکیب آن ها به یک جدول در هم سازی کمک کند. به علاوه، هشینگ فشرده برای ذخیره داده های بزرگ مقیاس اهمیت بسیار دارد. برای مثال، حفظ صد میلیون نمونه هر کدام با 100 بیت دو دویی، هزینه کم تر از 1.5 گیگابایت را دارد و به آسانی در حافظه قرار می گیرد.

برای ایجاد کد های هش فشرده موثر، روش های مختلف نظارت شده و نظارت نشده ارایه شده اند. روش هشینگ نظارت نشده پیشرفته، موسوم به هش گراف گسسته، از مفهوم گراف های تکیه کاه برای پوشش دادن ساختار همسایه در یک دیتاست گروهی استفاده کرده و سپس یک مدل هشینگ مبتنی بر گراف را در کل دیتاست فرموله می کند. این مدل منوط به روش های بهینه سازی گسسته جدید برای دست یابی به بیت های هش غیر همبسته است. روش DGH عملکرد بهتری از روش های هشینگ نظارت نشده نظیر کمی سازی تکراری، هشینگ طیفی و هشینگ کراف دارد و به این ترتیب قادر به پوشش دادن همسایه داده های خام در فضای کد گسسته خواهد بود

روش هشینگ نظارت شده پیشرفته، هشینگ گسسته نظارت شده، از اطلاعات برچسب نظارت شده استفاده کرده و هشینگ را از حیث طبقه بندی خطی، فرموله می کند که در آن کد های دو دویی یادگرفته شده برای طبقه بندی بهینه هستند. SDH از روش بهینه سازی مشترک استفاده می کند که به طور مشترک قادر به یادگیری دسته بند های خطی و دو دویی است. روش SDH عملکرد بهتری از روش های هشینگ نظارت شده قبلی دارد بسیاری از روش های هشینگ جالب دیگر نظیر هشینگ سند، ویدئو، داده های ساختار مند و بین رسانه ای وجود دارد. توجه کنید که همه روش ها بستگی به الگوریتم های یادگیری سطحی دارد. با این وجود، به دلیل سرعت بالای هشینگ مبتنی بر یادگیری سطحی، فنون هشینگ پیشرفته به طور گسترده در حافظه چند رسانه ای، نمایه

بندی، و بازیابی به خصوص در شرایط چند رسانه ای در ابزار های تلفن همراه استفاده می کنند. چندین شرکت نظیر اسنپ چت، پینترست، سنس تایم و فیس++ از فنون هشینگ مناسب برای مدیریت و جست و جو از طریق میلیون ها یا میلیارد ها تصویر استفاده می کنند

درهم سازی از طریق یادگیری عمیق

از سال 2006، یادگیری عمیق، موسوم به شبکه های عصبی عمیق، توجه زیادی را به خود جلب کرده است و از این روی تلاش های تحقیقاتی در طیف وسیعی از زمینه ها از جمله تشخیص صدا، چشم کامپیوتر، یادگیری ماشینی و متن کاوی انجام شده اند. هدف داده کاوی، یادگیری در مورد تماد ها و الگو های ویژگی قوی برای داده های پیچیده و نیز استفاده از یادگیری عمیق برای استفاده از کد های هش است و به صورت یک مدل دو دویی از داده ها استفاده می شود. از این روی، ما دو فنون هشینگ توسعه یافته مرتبط با یادگیری عمیق را معرفی می کنیم: شبکه عصبی کانولوشن و هشینگ شبکه عصبی عمیق.

فنون هشینگ قبلی با تکیه بر شبکه های عصبی عمیق، از بردار ویژگی های بصری استخراج شده از تصویر به عنوان ورودی استفاده می کند. کیفیت کد های هش تولید شده بستگی به کیفیت ویژگی های دستی دارد. برای حذف این مانع، رویکرد CNNHash، برای تلفیق یادگیری ویژگی تصویر و یادگیری کد هش به مدل یادگیری مشترک توسعه داده شده است.

این مدل متشکل از یک مرحله از کد های هش تقریبی یادگیری با توجه به اطلاعات نظارت شده و مرحله ای از آموزش شبکه عصبی کانولوشن عمیق است. مرحله دوم با استفاده از قدرت CNN قادر به یادگیری هم زمان ویژگی ها و مد های هش با استفاده از پیکسل تصویر خام است. CNN شامل سه لایه جمع اوری کانولوشن، یک لایه ارتباطی استاندارد و یک لایه خروجی با دستورات سافت مکس است. کد های هش نهایی توسط کمی سازی فعال سازی لایه خروجی تولید می شود.

اگرچه رویکرد CNNHash نیازمند یادگیری تقریبی کد های هش برای راهنمایی یادگیری مدل سازی تصویر و کد های هش می باشد، یک رویکرد جدید موسوم به DNNHash بوده است. از طریق DNNHash، مدل سازی تصویر و کد ها از اطلاعات نظارت شده برای آموزش CNN عمیق استفاده کرده و منجر به یک معماری عمیق برای هش نظارت شده می شود. خط لوله با معماری هشینگ عمیق شامل سه ساختار اصلی زیر است

- یک سه گانه از تصاویر که به سی ان ان تغذیه می شوند و بر روی آن میزان تلفات را می توان برای شناسایی اطلاعات نظارت شده طراحی کرد

- یک زیر شبکه مشترک با طیف وسیعی از لایه های کانولوشن برای تولید ویژگی های تصویر

- یک ماژول تقسیم و کد گذاری برای تقسیم ویژگی های تصویر در کانال های متعدد که هر یک از آن ها به یک بیت کد گذاری می شود

در ماژول تقسیم و کد گذاری، یک لایه مرتبط و یک لایه هش وجود دارد. در نهایت، یک کد هش از تصویر از طریق استانه سازی خروجی لایه هش ارایه می شود. DNNHash عملکرد بهتری از CNNHash دارد و رویکرد های هشینگ نظارت شده مبتنی بر یادگیری سطحی از حیث صحت جست و جوی تحقیقات باید در نظر گرفته شوند

با این حال برای هر دو DNNHash و CNNHash، توجه کنید که محققان هنوز به بررسی زمان برای تولید کد هش نپرداخته است. در سناریو های تحقیق دنیای واقعی، سرعت تولید هش بایستی بسیار سریع باشد. نگرانی های زیادی برای بررسی سرعت هش رویکرد های مبتنی بر شبکه عصبی به خصوص یادگیری ویژگی تصویر وجود دارد زیرا درهم سازی تصویر با یادگیری عمیق می تواند در مقایسه با رویکرد های مبتنی بر یادگیری سطحی، طولانی تر باشد.

شبکه بندی چند رسانه ای

در این جا ما به معرفی شبکه های اطلاعات چند رسانه ای اخیر می پردازیم. به عنوان مثال، ما یک کنفرانس چند رسانه ای را ارایه می کنیم که از هر دو اطلاعات متنی و بصری برای دست یابی به کنفرانس ها استفاده می کند

شبکه های اطلاعات چند رسانه های

پیشرفت های اخیر در زمینه فناوری وب به خصوص در سیستم های ذخیره سازی بزرگ مقیاس و سریع، امکان انتشار اخبار و را به طور عمقی داده است. با این حال، این پیشرفت ها موجب بروز مسائلی نظیر کاهش اطلاعات رسانه ای و محتوی خبری شده است در بسیاری از سناریو های نوظهور نظیر مواجهه با بلاای طبیعی، توییت ها و خبر ها به طور مکرر تکرار می شوند. انتشار این پیام ها و صفحات نا خوشایند و ناگوار است و از این روی خلاصه سازی خودکار توییت ها و اخبار لازم است و رتبه بندی اساسی ترین شیوه اطلاع رسانی به کاربران است

یک راه حل منفعل و رایج، تشویق کردن کاربران برای استفاده از کلمات کلیدی بیشتر در زمان تایپ کوئری است. با این حال، بدون دانش قبلی و محدودیت های کلمات، ایجاد یک فهرست رتبه بندی برای موضوعات جذب توجه از اهمیت برخوردار است. تغییرات اخیر در موتور جست و جو گوگل، مولفه جست و جوی تصویر را ترکیب کرده و از تحلیل محتوی ناهمگن استفاده کرده است. با این وجود، ارتباط بین تصاویر و کلمات کلیدی توسط کاربران تعیین می شود و لذا کیفیت جست و جو بسیار بهینه می شود.

راه حل های فعال تلاش می کنند تا اطلاعاتی را خلاصه کنند که بر مدالیته های تک داده ای متمرکز هستند. محققان یک روش رتبه بندی صفحه حساس به محتوی متن را برای استخراج کلمات کلیدی از توپتر به عنوان شیوه ای برای خلاصه سازی محتوی توپتر ارائه کرده اند. از چشم انداز جدید، تخصیص دیشلت پنهان برای تفسیر و حاشیه نویسی تصاویر استفاده شده است ولی این لزوماً از مدوالیته های داده های مختلف استفاده می کند محققان رویکرد های رتبه بندی توپیت را توسعه داده اند با این حال آن ها بر مدالیته داده ها متمرکز است. ارتباط بین تصاویر و کلمات کلیدی توسط کاربران تعیین می شود.

سایر راه حل ها برای تحلیل روابط و یا ارتباط میان نمونه ها شامل پیچ رنگ و ویژوال رنگ است. اولی به طور گسترده ای در شبکه های ناهمگن استفاده می شود و یژال رنگ که از پیچ رنگ استفاده می کند یک روش ارتباط مبتنی بر محتوی است با این حال محدود به شبکه های همگن است.

یک مدل MINet اخیراً برای ایجاد هستی شناسی سیستم رتبه بندی قوی ارائه شده است. و هدف آن تلفیق استنباط و ایجاد ارتباط میان اطلاعات چند مودالیته استخراج شده از داده های ناهمگن است. در فراتر از رویکرد های رتبه بندی، که برای شبکه های همگن طراحی شده است، بسیاری از محققان از یک سری رویکرد های نوین برای استفاده از ویژگی های رتبه بندی استفاده می کند و از این روی امکان استفاده از minET وجود دارد.

کنفرانس های چند رسانه ای

با این حال، این شبکه های اطلاعاتی که در آن ها گره نهایی می تواند یک رویداد باشد، از رویداد های افزونه رنج برده و یا کاهش کارایی ناشی از گره های مکرر می باشد زیرا داستان های مشابه توسط عوامل خبری رشد کرده است. به علاوه، برای تقویت اثر بر روی مخاطبانف رویداد ها و داستان ها، زمان های مختلف گزارش شده به خصوص تلویزیون و نیز پخش رادیویی است.

این ویژگی‌ها نیازمند روش‌های خودکار برای استفاده از اطلاعات خوشه‌ای و نیز حذف افزونگی هستند. یک روش جدید پیشنهادی قادر به مدیریت اطلاعات بصری و متنی نیست. از این روی این روش می‌تواند به طور کامل از اخبار تلویزیونی حاوی صوت و ویدئو باشد.

یک نقطه شروع خوب برای کنفرانس چند رسانه‌ای، پردازش کپشن‌های بسته است که با ویدئو همراه هستند. این CCها توسط یک سیستم تشخیص صدای خودکار تولید شده و از این روی امکان دست‌یابی به رونوشت‌ها توسط فونتیک ورودی وجود دارد. جدا از اخبار کتبی، یک پخش خبری اغلب بهدلیل برنامه TV ارایه می‌شود امکان طبقه‌بندی توصیفات CC در کوتاه مدت وجود دارد. این دو توصیف دارای یک هم‌پوشانی حداقل هستند اگرچه آن‌ها وابسته به یک دیگر نیز می‌باشند. برای مثال گزارش‌گران معمولاً از داستان‌های پیش‌زمینه استفاده کرده و قادرند تا رویداد‌های استخراج شده از CC را استفاده کنند.

چالش‌های دیگر حاصل از اشتباهات در CC و خطاهای ناشی از اپراتورهای انسانی و سیستم‌های ASR است. برای مثال، در دو پخش مشابه، که در آن‌ها پیلون‌های خاصی گزارش می‌شود، کپشن بسته مطلوب بوده است. دیگر موضوع جدید وجود نقص در CC است: اردن باید دو زندانی ISIS را به دلیل تلافی دست‌گیری خلبان اردنی اعدام کند. به این ترتیب امکان استفاده از یک رویکرد کنفرانسی برای خوشه‌بندی وجود دارد. رویداد‌های مرگ را می‌توان به اتفاقات مشابه طبقه‌بندی کرد. ویدئوها اغلب توصیف کوتاهی از محتوا ارایه کرده و هر دوی آن‌ها از چارچوب‌های ویدئویی برای نمایش دست‌گیری خلبان اردنی استفاده می‌کنند از این روی دو رویداد را می‌توان به یک صورت در نظر گرفت

در حقیقت، گزارش‌گران و کانال‌های تلویزیونی از محتوی ویدئوی مشابه برای توصیف یک داستان استفاده کرده‌اند و این در حالی است که آن‌ها از کلمات و عبارات‌های مختلف ارایه می‌کنند. به این ترتیب چالش‌ها در روش‌های حل اختلاف بستگی به اطلاعات استخراج شده با تشابه بصری دارد.

کارها و مطالعات مشابه به بررسی علائم و نشانه‌های ارتباطی با متن پرداخته‌اند. با این حال این روش‌ها بر ارتباط مفاهیم تصویر با مراکز موجود در متن مبتنی بوده‌اند و از این روی امکان تمایز یک نهاد از رویداد در اسناد وجود دارد، چون تعاریف مفاهیم نیازمند هر دو است. به علاوه مطالعات 21-23، به بهبود تشخیص محتوی بصری

با معرفی ویژگی های متنی پرداخته اند ضمن این که کار های اخیر مسیر های مخالفی با استفاده از اطلاعات بصری برای بهبود حل اختلاف اتخاذ کرده اند.

در آینده، ما از فنون بحث شده برای استفاده از روش های یادگیری کاربردی در زمینه های چند رسانه ای واقعی استفاده می کنیم. برای مثال، هدف ما توسعه شبکه های عصبی عمیق برای روش های درهم سازی مبتنی بر یادگیری با سرعت هش زمان واقعی است. هدف ما معرفی شبکه های حافظه انتها به انتها برای درک اطلاعات متنی و بصری و استفاده از روش های چند رسانه ای است.