



دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر
سمینار کارشناسی ارشد گرایش هوش مصنوعی

عنوان:
یادگیری بدون برد با شبکه‌های عمیق
Deep Zero-Shot Learning

نگارش:
سید محسن شجاعی
۹۳۲۰۷۹۷۹

استاد راهنما:
دکتر مهدیه سلیمانی

استاد ممتحن داخلی:
دکتر حمیدرضا ربیعی

چکیده: مسئله یادگیری از صفر^۱ به دنبال پیش‌بینی دسته‌هایی در زمان آزمون است که در زمان آموزش هیچ داده‌ای از آن دسته‌ها مشاهده نکرده است. شناسایی این دسته‌ها با یک سری اطلاعات جانبی صورت می‌گیرد. در یک مسئله دسته‌بندی تصاویر، یادگیری بدون برد به این صورت است که تعدادی تصویر به همراه برچسب و اطلاعات جانبی به الگوریتم داده می‌شود. در زمان آزمون یک سری اطلاعات جانبی مربوط به دسته‌های جدید و تصاویری بدون برچسب وجود دارد و هدف برچسب‌گذاری تصاویر با دسته‌های جدیدی است که اطلاعات جانبی آن‌ها داده شده. ویژگی‌های بصری و متون توصیف کننده یک دسته، مثال‌هایی از اطلاعات جانبی مورد استفاده در این نوع مسائل هستند. در این گزارش حالت‌های مختلف تعریف مسئله یادگیری از صفر معرفی می‌شود. سپس کارهای پیشین انجام شده مورد بررسی قرار می‌گیرد. در ادامه یک راه‌حل پیشنهادی با استفاده از تعریف فضای نهان بر اساس کلاس‌های دیده شده در زمان آموزش معرفی می‌کنیم. در پایان نتایج روش ارائه شده با نتایج روش‌های پایه مقایسه خواهد شد.

واژه‌های کلیدی: یادگیری از صفر، یادگیری بازنمایی، شبکه‌های عمیق

۱ مقدمه

در حوزه یادگیری ماشین مسئله استاندارد یادگیری با نظارت به صورت‌های مختلف توسعه یافته است. به کمک این روش‌ها، یادگیری ماشین از عهده‌ی کارهای بسیار چالش‌برانگیزتری برآمده است. بر خلاف پارادایم سنتی یادگیری با نظارت که فرض می‌کند داده‌های فراوانی از تمام دسته‌ها برای آموزش در اختیار قرار دارد، عموم این روش‌ها به دنبال کم کردن نیاز به داده‌های برچسب‌دار در زمان آموزش هستند. یادگیری نیمه‌نظارتی^۲ [۱] برای استفاده کردن از حجم زیاد داده‌های بدون برچسب موجود در جریان آموزش پیشنهاد شده است. یادگیری از تک نمونه^۳ [۲] سعی می‌کند یک دسته را تنها بوسیله یک نمونه‌ی برچسب‌دار از آن و البته با کمک نمونه‌های برچسب‌دار از سایر دسته‌ها شناسایی کند. انتقال یادگیری^۴ [۳] سعی می‌کند دانش به دست آمده از داده‌های یک دامنه یا برای انجام یک وظیفه را به داده‌های دامنه‌ی دیگر یا وظیفه‌ی دیگری روی داده‌ها منتقل کند. هیچ‌کدام از این روش‌ها یا به داده‌های برچسب‌دار را برای دسته‌هایی که مایل به تشخیص آن هستیم را به طور کامل از بین نمی‌برد. برای دستیابی به چنین هدفی، [۴] مسئله یادگیری از صفر را صورت‌بندی می‌کند. به این صورت که به دنبال یافتن یک دسته‌بند هستیم در حالی که داده‌های آموزش برای بعضی از کلاس‌ها هیچ نمونه‌ای در بر ندارد. در عوض یک توصیف از تمامی کلاس‌ها موجود است. این حالت به خصوص وقتی که تعداد دسته‌ها بسیار زیاد است رخ می‌دهد. برای مثال در بینایی ماشین تعداد دسته‌ها برابر انواع اشیای موجود در جهان است و جمع‌آوری داده‌های آموزش برای همه اگر غیر ممکن نباشد به هزینه و زمان زیادی احتیاج دارد. خصوصاً که همانطور که در [۵] نشان داده شده، تعداد نمونه‌های موجود برای هر دسته از قانون Zipf پیروی می‌کند. یک مثال دیگر رمزگشایی فعالیت ذهنی فرد است [۶]؛ یعنی تشخیص کلمه‌ای که فرد در مورد آن فکر یا صحبت می‌کند بر اساس تصویری که از فعالیت مغزی او تهیه شده است. طبیعتاً در این مسئله تهیه تصویر یا سیگنال فعالیت مغزی برای تمامی کلمات لغت‌نامه ممکن نیست. یک موقعیت دیگر که توصیف مسئله یادگیری از صفر بر آن منطبق است دسته‌بندی دسته‌های جدید است، مانند تشخیص مدل‌های جدید محصولات مثل خودروها که یک دسته‌ای است که در زمان آموزش اصولاً وجود نداشته است. یادگیری از صفر نیز مانند بسیاری از مسائل در یادگیری ماشین با توانایی‌های یادگیری در انسان ارتباط دارد و الهام از یادگیری انسان‌ها در شکل‌گیری‌اش بی‌تاثیر نبوده است. برای مثال انسان قادر است بعد از شنیدن توصیف «حیوانی مشابه اسب با راه‌راه‌های سیاه و سفید» یک گورخر را تشخیص دهد. یا تصویر یک اسکوتر را با توصیف «وسیله‌ای دو چرخ، یک کفی صاف برای ایستادن، یک میله صلیبی شکل با دو دستگیره» تطبیق خواهد داد.

در این نوشتار روی مسئله دسته‌بندی تصاویر از صفر تمرکز می‌کنیم؛ به این معنی که داده‌هایی که مایل به دسته‌بندی آن هستیم تصاویر هستند. در نتیجه در زمان آموزش تعدادی تصویر به همراه برچسب آن‌ها موجود است. برچسب‌هایی که در زمان آموزش وجود دارند را دسته‌های دیده شده یا دسته‌های آموزش می‌نامیم. همچنین یک نوع اطلاع جانبی هر یک از دسته‌های آموزش را وصف می‌کند؛ به این اطلاعات جانبی توصیف می‌گوییم. در زمان آزمون تصاویری ارائه می‌شود که به دسته‌هایی غیر از دسته‌های آموزش تعلق دارند. به این دسته‌ها با نام دسته‌های آزمون یا دسته‌های دیده‌نشده اشاره می‌کنیم. همچنین اطلاعات جانبی مربوط به این کلاس‌ها نیز در اختیار قرار می‌گیرد و بایست هر تصویر به یکی از توصیف‌ها نگاشته شود. در برخی روش‌ها فرض می‌شود توصیف دسته‌های آزمون هم در زمان آموزش قابل دسترسی است. توصیف‌ها ممکن است به صورت یک بردار از ویژگی‌های بصری [۷]، عبارات زبان طبیعی [۸، ۹، ۱۰] و یا دسته‌بندهای یادگرفته شده [۱۱] باشند. بردار ویژگی مرسوم‌ترین شکل توصیف کلاس است. ویژگی‌ها با توجه به نوع مسئله و گستردگی

دسته‌ها تعیین می‌شوند. اکثر ویژگی‌ها، ویژگی‌های بصری هستند مانند شکل (مانند گرد یا مستطیلی)، جنس (مانند چوبی یا فلزی) و عناصر موجود در تصویر (مانند چشم، مو، پدال و نوشته). برخی ویژگی‌ها هم ممکن است مستقیماً در تصویر قابل مشاهده نباشند برای مثال در یک مجموعه دادگان که دسته‌ها انواع حیوانات هستند [۱۲]، علاوه بر ویژگی‌های بصری، ویژگی‌هایی چون اهلی بودن، سریع بودن یا گوشت‌خوار بودن هم وجود دارد.

مباحث ادامه این گزارش به این صورت است: در بخش ۲ صورت‌های مختلفی مسئله یادگیری از صفر را با توجه به نوع اطلاعات جانبی مورد استفاده بیان کرده و روش‌های پیشین ارائه شده برای حل آن‌ها مرور می‌کنیم. در بخش ۳؟ یک روش پیشنهادی بیان می‌شود و نتایج عملی آن در بخش ۴؟ ارائه و روش‌های دیگر مقایسه می‌شود. بخش ۶ به کارهای آتی، جدول زمان‌بندی پژوهش و جمع‌بندی اختصاص دارد.

۲ کارهای پیشین

یک نحوه مدل‌سازی یادگیری از صفر، آن طور که در [۶] بیان شده، تبدیل آن به دو زیر مسئله است. مسئله اول یادگیری یک نگاشت از مجموعه تصاویر به یک فضای میانی که توصیف کلاس‌ها در آن قرار دارند و مسئله دوم یادگرفتن یک دسته‌بند که اعضای فضای میانی را به برچسب‌ها دسته‌بندی کند. در این نحوه مدل‌سازی فضای معنایی توصیف‌ها داده‌شده فرض می‌شود. این درحالی‌ست که بسیاری از اوقات توصیف‌ها به صورت خام قابل استفاده نیستند. برای مثال وقتی اطلاع جانبی از نوع متن است را در نظر بگیرید، فضای متون فضایی با بعد بسیار بالاست و لازم است که خود به یک فضای میانی نگاشته شود. یادگیری نگاشت از توصیف‌ها به فضای میانی ممکن است به صورت هم‌زمان و اشتراک بعضی پارامترها با سایر قسمت‌های مدل یادگرفته شود؛ در نتیجه لازم است یادگیری این نگاشت را هم جزو چارچوب ارائه شده در نظر بگیریم. در این نحوه مدل‌سازی یک چارچوب کلی برای بسیاری از روش‌های ارائه شده در یادگیری از صفر است. البته روش‌هایی هم وجود دارند که در این چارچوب ننگینند. در این بخش، با توجه به فراگیری این چارچوب ابتدا توصیف رسمی و معرفی نمادگذاری برای آن ارائه می‌شود. سپس روش‌های ذیل این چارچوب را مرور کرده و در پایان سایر روش‌ها را بیان می‌کنیم.

۱,۲ نمادگذاری

تصاویر را با $x \in \mathbb{R}^d$ نشان می‌دهیم که d ابعاد داده را نشان می‌دهد. توصیف‌ها را با $c \in \mathbb{R}^a$ نمایش می‌دهیم. a ابعاد توصیف‌هاست. مجموعه دسته‌های دیده‌شده را با \mathcal{S} و دسته‌های دیده‌نشده را با \mathcal{U} و مجموعه کل برچسب‌ها را با \mathcal{Y} نشان می‌دهیم که $\mathcal{Y} = \mathcal{U} \cup \mathcal{S}$. همچنین s و u به ترتیب تعداد هر کدام از دسته‌ها را نشان می‌دهد. c^y که $y \in \mathcal{U} \cup \mathcal{S}$ بردار توصیف کلاس y را نشان می‌دهد.

فرض می‌کنیم در زمان آموزش N_s تصویر از دسته‌های دیده شده به همراه برچسب و توصیف دسته‌ها موجود است. $X_s \in \mathbb{R}^{N_s \times d}$ مجموعه تصاویر، Y بردار برچسب‌ها و $C_s \in \mathbb{R}^{s \times a}$ مجموعه توصیف‌های دسته‌های آموزش است. X_u و C_u بطور مشابه برای دسته‌های آزمون تعریف می‌شوند.

فضای میانی را با \mathcal{M} و ضرب داخلی آن را با $\langle \cdot, \cdot \rangle$ نشان می‌دهیم. $\pi: \mathbb{R}^d \rightarrow \mathcal{M}$ و $\psi: \mathbb{R}^a \rightarrow \mathcal{M}$ نگاشت‌هایی از فضای تصاویر و توصیفات به این فضا هستند. یادگیری نگاشت‌های π و ψ ممکن است به صورت مستقل از هم انجام شود یا اینکه هم‌زمان یادگرفته شوند. در نهایت باید دسته‌بندی از \mathcal{M} به برچسب‌ها داشته باشیم: $\phi: \mathcal{M} \rightarrow \mathcal{Y}$. در خیلی از موارد دسته‌بندی را تنها روی دسته‌های آزمون در نظر می‌گیریم، یعنی برد ϕ تنها \mathcal{U} را شامل می‌شود نه تمام برچسب‌ها را. در ساده‌ترین حالت ϕ یک دسته‌بند نزدیک‌ترین همسایه در نظر گرفته می‌شود، یعنی برچسب نمونه آزمون x با رابطه ۱ پیش‌بینی خواهد شد:

$$y^* = \arg \max_{y \in \mathcal{U}} \langle \pi(x), \psi(c^y) \rangle \quad (1)$$

البته این انتخاب برای ϕ محدودیت‌های شناخته شده‌ای دارد. از جمله این که تمامی ابعاد از اهمیت یکسانی برخوردار هستند، درحالی‌که ممکن است بعضی ویژگی‌ها قابلیت جداسازی بهتری داشته باشند.

چارچوب فوق را می‌توان به روش‌های احتمالی هم تعمیم داد، به این صورت که π و ψ به صورت توزیع‌های احتمال شرطی تغییر پیدا می‌کنند. این تعمیم به صورت دقیق‌تر در بخش ۱۰.۲،۲ بررسی خواهد شد.

۱.۱،۲ کران خطا

تعریف و فروض یادگیری از صفر با حالت معمول دسته‌بندی متفاوت است. در نتیجه کران‌هایی که امکان‌پذیر بودن دسته‌بندی را با استفاده تعداد محدودی نمونه ضمانت می‌کنند را نمی‌توان در اینجا به کار بست. برای ارائه کران‌های خطای دسته‌بندی از صفر فرض‌های ساده‌کننده‌ای به مسئله اضافه شده است. برای این منظور فرض می‌شود که یادگیری نگاشت ψ مستقل از π انجام شده و رابطه بین توصیف‌ها و برچسب دسته‌ها رابطه‌ای یک به یک است. با این دو فرض می‌توان $\psi(c^y)$ را امضای دسته‌ی y نامید.

در [۶] با فرض دودویی بودن هر بعد از امضای دسته‌ها، کرانی بر اساس فاصله همینگ^۵ میان امضای دسته‌ی صحیح و مقدار پیش‌بینی شده ارائه می‌شود. در [۹] از نتایج مشابه در حوزه تطبیق دامنه برای کران‌دار کردن خطا استفاده می‌شود. در این جا کران ارائه شده بر اساس تفاوت توزیع‌های داده‌های آموزش و آزمون است. در آن نوشتار راهی برای تخمین تفاوت این دو توزیع در حالت کلی ارائه نمی‌شود. تنها به دو حالت حدی اشاره می‌شود که در صورت یکسان بودن توزیع‌ها، کران ارائه شده همان کران مشهور VC [۱۳] خواهد بود. هم‌چنین درحالی‌که امضای کلاس‌ها بر هم کاملاً عمود باشد کران برای احتمال خطا بزرگتر از یک شده و اطلاعاتی در بر ندارد.

۲،۲ پیش‌بینی ویژگی

همان‌طور که در بخش ۱ اشاره شد، بردار ویژگی مرسوم‌ترین نوع توصیف کلاس‌هاست. نخستین کارها روی یادگیری از صفر در بینایی ماشین [۷، ۱۲]، روش پیش‌بینی مستقیم ویژگی‌ها را پیشنهاد داده‌اند. در این حالت سعی می‌شود بردار ویژگی از روی تصویر ورودی بازسازی شود. آن‌گاه از میان دسته‌های دیده نشده، دسته‌ای که بردار ویژگی‌اش بیشترین شباهت را با بردار پیش‌بینی شده دارد به عنوان برچسب معرفی می‌شود. با ادبیات چارچوب معرفی شده این روش را این گونه توصیف می‌شود که فضای میانی M همان فضای بردار ویژگی در نظر گرفته شده است در نتیجه نگاشت ψ نگاشت همانی است و هدف تنها یادگرفتن نگاشت π است. اهمیت این روش‌ها از یک طرف بخاطر داده‌های بسیاری است که با فراداده‌ها^۶ و دنباله‌ها^۷ همراه شده‌اند که به صورت بردار ویژگی قابل مدل‌سازی هستند. دلیل دیگری برای اهمیت این روش‌ها این است که در مواردی هم که توصیف‌ها از نوع بردار ویژگی نیستند، ابتدا ψ به صورت مستقل یادگرفته می‌شود و بعد از آن با در نظر گرفتن $\psi(c^x)$ بعنوان بردار ویژگی کلاس‌ها مسئله به حالت مورد بحث این بخش تبدیل خواهد شد.

اگر ویژگی‌ها دودویی باشند. این مسئله را می‌توان نوعی دسته‌بندی چند برچسبی^۸ دانست که مدت زیادی است در حوزه یادگیری ماشین مورد مطالعه قرار گرفته است [۱۴]. البته دسته‌بندی چندبرچسبی با یادگیری از صفر از طریق پیش‌بینی ویژگی تفاوت‌هایی دارد. در اولی خروجی الگوریتم یک بردار ویژگی است و ترکیب‌های بسیار زیادی از مقادیر برای آن مجاز هستند، در دومی خروجی نهایتاً یک برچسب از کلاس‌های دیده نشده است و بردار ویژگی یک مقدار میانی برای رسیدن به این خروجی است. هم‌چنین همه ترکیب‌ها از ویژگی‌ها مجاز نیستند و تنها به تعداد دسته‌ها بردار ویژگی معتبر وجود دارد. در صورتی که ویژگی‌ها پیوسته باشند مسئله پیش‌بینی آن‌ها می‌تواند به صورت یک مسئله رگرسیون در نظر گرفته شود که برای در نظر گرفتن ارتباط ویژگی‌های مختلف باید با مدل‌های رگرسیون ساختاریافته [۱۵] حل شود. روش‌های معمول رگرسیون مانند فرایند گاوسی هر ویژگی را به صورت جداگانه یاد گرفته و ارتباط میان ابعاد در نظر گرفته نخواهد شد [۹]. مانند حالت دودویی مشکل دیگر تفاوت اساسی این مسئله با یادگیری از صفر است. در این مسئله به دنبال خطای کمتر در ویژگی‌های پیش‌بینی شده هستیم درحالی‌که در مسئله یادگیری از صفر این خطا اهمیتی ندارد و الگوریتم با دقت برچسب‌گذاری سنجیده می‌شود.

۱.۲.۲ روش‌های احتمالی

یکی از نخستین روش‌های پیش‌بینی ویژگی در [۱۲] ارائه شده است. فرض کنید در زمان آموزش نمونه‌های $\{(x_i, y_i)\}_{i=1}^{N_s}$ به همراه بردار ویژگی دسته‌های آموزش C_y در اختیار قرار گرفته است. در نسخه اول این روش که DAP نام دارد استفاده از داده‌های آزمون تنها به صورت یادگیری دسته‌بندی برای هر یک از ویژگی‌هاست. این یادگیری با فرض استقلال ابعاد ویژگی‌ها انجام می‌شود، یعنی $P(c|x) = \prod_{i=1}^d P(c_i|x)$. و هر یک از $P(c_i|x)$ با یک رگرسیون منطقی^۹ روی کل داده‌ها (مستقل از برجسب آن‌ها) تخمین زده می‌شود. همچنین احتمال پیشین وقوع هر یک ویژگی‌ها، $P(c_i)$ ، به صورت تجربی با توجه به تعداد وقوع تعیین می‌شود. در نهایت احتمال پسین هر کدام از برجسب‌های آزمون $u \in \mathcal{U}$ از این رابطه بدست می‌آید:

$$P(u|x) = \sum_c P(u|c)P(c|x) = \frac{P(u)}{P(c^u)} \prod_{i=1}^d P(c_i^u|x) \propto \prod_{i=1}^d \frac{P(c_i^u|x)}{P(c_i^u)} \quad (۲)$$

در نسخه دیگر این روش که IAP نام دارد تخمین $P(c_i|x)$ تغییر داده می‌شود به این صورت که ابتدا یک دسته‌بند چند دسته‌ای یعنی $P(y_k|x)$ روی داده‌ها یاد گرفته می‌شود و سپس رابطه ویژگی‌ها و برجسب‌ها به صورت قطعی مدل می‌شود. در نهایت خواهیم داشت:

$$P(c_i|x) = \sum_{k=1}^s P(y_k|x) \mathbb{I}(c_i = c_i^{y_k}) \quad (۳)$$

که $\mathbb{I}(p)$ وقتی که شرط p برقرار باشد برابر ۱ و در غیر این صورت صفر است.

علاوه بر این دو نسخه، این روش به حالت‌های دیگری هم توسعه داده شده است. برای مثال در [۱۶] وزن‌دهی متفاوت برای مدل‌سازی اهمیت هر کدام از ویژگی‌ها به مدل اضافه شده است. این روش دو کمبود مهم دارد، اول این که فرض استقلال میان ویژگی‌ها بسیار غیر واقعی است. برای مثال ویژگی‌های بصری خاک و صحرا وابستگی واضحی وجود دارد. مشکل دوم این است که یادگیری دسته‌بندی برای هر ویژگی بدون توجه به این است که از خروجی آن در دسته‌بندی دیگری استفاده می‌شود و معیار ارزیابی عمل‌کرد خطای دسته‌بند دوم است و خطای پیش‌بینی ویژگی‌ها به طور مستقیم اهمیت ندارد. نویسندگان [۱۷] برای حل این مشکل پیشنهاد می‌کنند فرض یک به یک بودن نداشت بین بردارهای ویژگی و برجسب‌ها را در نظر بگیریم و به عبارت دیگر دسته‌بند ϕ معرفی شده در بخش؟؟ یاد گرفته شود. پیش‌بینی ویژگی‌ها مانند مدل DAP با رگرسیون منطقی انجام می‌شود با این تفاوت که یادگیری پارامترهای آن‌ها و ϕ به صورت مشترک انجام می‌شود. ϕ یک نداشت خطی در نظر گرفته می‌شود: $\phi(\pi(x)) = R\pi(x)$. دو محدودیت روی مقادیر R اعمال می‌شود. یک محدودیت سطری و یک محدودیت ستونی. محدودیت سطری مانع از این می‌شود که فاصله همینگ سطرها از حدی کمتر بشود. دقت کنید که در یک دسته‌بند خطی به شکل بالا، هر سطر مانند نماینده از یک دسته است که میزان شباهت با آن میزان تعلق به آن دسته را می‌سنجد. در نتیجه این محدودیت تضمین می‌کند که بردار ویژگی نماینده هر دسته با دسته‌های دیگر متفاوت باشد. محدودیت ستونی یک مقدار حداکثری برای همبستگی میان ستون‌ها در نظر می‌گیرد تا به این صورت اطلاعات تکراری در ویژگی‌ها وجود نداشته باشد. نویسندگان این مقاله استدلال می‌کنند که با این دو محدودیت باعث حذف ویژگی‌های تکراری و ویژگی‌های غیر بصری (مانند بدبو بودن) خواهد شد.

نویسندگان [۱۸] برای در نظر گرفتن ارتباط بین ویژگی‌ها و ارتباط ویژگی‌ها با برجسب نهایی روش‌های مدل‌سازی مباحث^{۱۰} را از حوزه یادگیری در متن اقتباس می‌کنند. همچنین نویسندگان [۱۹] برای این کار یک چارچوب بر اساس مدل‌های گرافنی احتمال معرفی می‌کنند. در این چارچوب یک شبکه بیزی^{۱۱} برای مدل کردن این روابط در نظر گرفته می‌شود که ساختار آن با کمک روش‌های یادگیری ساختار^{۱۲} شناخته می‌شود.

۲.۲.۲ نگاشت‌های خطی

چند روش اخیر وجود دارد که علی‌رغم ساده بودن نتایج بهتری از روش‌های قبلی کسب کرده‌اند. در این روش‌ها نگاشت ψ همانی، دسته‌بند ψ دسته‌بند نزدیک‌ترین همسایه و نگاشت π خطی $\pi(x) = xW$ در نظر گرفته شده‌اند. اما معرفی توابع هزینه یا جمله‌های منظم‌سازی^{۱۳}

هوشمندانه‌تر باعث شده که نتایج بهتری به دست بیاورند. یکی از این روش‌ها که در [۲۰] معرفی شده، تابع هزینه‌ای ارائه می‌دهد که هم خطای دسته‌بندی، هم خطای پیش‌بینی ویژگی‌ها را در نظر می‌گیرد. این تابع هزینه چنین شکلی دارد:

$$L(W) = \frac{1}{N_s} \sum_{n=1}^{N_s} \lambda_{r_{\Delta}(x_n, y_n)} \sum_{y \in \mathcal{Y}} \max(\bullet, l(x_n, y_n, y)) \quad (۴)$$

$$l(x_n, y_n, y) = \mathbb{I}(y \neq y_n) + x_n W c_y - x_n W c_{y_n} \quad (۵)$$

که در آن $r_{\Delta}(x_n, y_n) = \sum_{y \in \mathcal{Y}} \mathbb{I}(l(x_n, y_n, y) > \bullet)$ و λ_k یک تابع نزولی از k است. این تابع، پیش‌بینی اشتباه ویژگی‌ها را این گونه جریمه می‌کند که به ازای برچسب نادرستی که رتبه بالاتری از برچسب صحیح در دسته‌بندی دریافت کرده، جریمه‌ای متناسب با امتیاز برچسب ناصحیح در نظر گرفته می‌شود. ضریب نزولی λ_k میزان جریمه را برای برچسب‌های غلط در رتبه‌های بالا بیشتر در نظر می‌گیرد.

یک روش دیگر که در [۲۱] ارائه شده، نگاشت‌های مشابهی را استفاده می‌کند. همچنین تابع هزینه آن شکل ساده نرم ۲ را دارد. مسئله‌ی بهینه‌سازی تعریف شده به این شکل است:

$$\underset{W \in \mathbb{R}^{d \times a}}{\text{minimize}} \|X_s W C_s\|_{Fro}^2 + \Omega(W) \quad (۶)$$

که $\Omega(W)$ یک جمله منظم‌سازی است که به این صورت تعریف می‌شود:

$$\Omega(W; X_s, C_s) = \gamma \|W C_s\|_{Fro}^2 + \lambda \|X W\|_{Fro}^2 + \beta \|W\|_{Fro}^2 \quad (۷)$$

که تابع هزینه فوق تنها دسته‌بندی اشتباه را جریمه می‌کند. مناسب نبودن تابع هزینه نرم ۲ برای خطای دسته‌بندی مسئله‌ای شناخته شده در یادگیری ماشین است و عمل‌کرد خوب این تابع در این روش شاید در نگاه اول عجیب بنظر برسد. اگر در جمله منظم‌سازی تعریف شده دقت کنیم این مسئله روشن‌تر خواهد شد. تابع هزینه نرم ۲ به این علت که حتی دسته‌بندی‌های صحیح را اگر مقداری غیر از مقدار تعیین شده (معمولاً یک) داشته باشند، به اندازه فاصله‌شان از این مقدار جریمه می‌کند. اما جمله منظم‌سازی تعریف شده اصولاً مانع بزرگ شدن مقدار پیش‌بینی شده خواهد شد. جمله اول در معادله (۴؟) را می‌توان اندازه بردار تصویر متوسط برای هر دسته دانست. جمله دوم مقدار بردار ویژگی پیش‌بینی شده برای هر دسته است و جمله سوم هم که یک جمله معمول است که پارامترهای نگاشت را کنترل می‌کند. در زمان آزمون برای نمونه x مقدار $x W C_u$ را محاسبه کرده و دسته‌ای که درایه‌ی متناظرش بیشترین مقدار را دارد به عنوان پیش‌بینی معرفی می‌کنیم. یک ویژگی این روش این است که با انتخاب $\beta = \gamma \lambda$ در معادله (۴؟) بهینه‌سازی معادله (۴؟) جواب بسته خواهد داشت؛ در نتیجه زمان اجرای این روش بسیار کمتر از سایر روش‌هایی است که مرور شد.

یک روش خطی دیگر که مستقیم از ویژگی‌ها استفاده نمی‌کند کاری است که در [۲۲] معرفی شده. این روش از تنها از نام هر کلاس به عنوان توصیف بهره می‌برد. با توجه به این در این روش نام‌ها مستقل از اطلاعات دیگر مسئله به بردارهایی نگاشته می‌شوند، بردارهای حاصل را می‌توان مانند بردار ویژگی در سایر مسائل به حساب آورد؛ در نتیجه این روش را ذیل عنوان پیش‌بینی ویژگی مرور می‌کنیم. این روش ابتدا برای بدست آوردن بردارهای مربوط به نام‌ها از مدل مشهور word2vec [۲۳] با پیش‌آموزش روی مقالات ویکی‌پدیای انگلیسی استفاده می‌کند. این روش هم‌چنین برای ویژگی‌های تصویر از شبکه عصبی برنده چالش ILSVRC 2012 استفاده می‌کند. * این روش نیز π را خطی و دسته‌بند ϕ را نزدیک‌ترین همسایه در نظر می‌گیرد. تابع هزینه مورد استفاده از این روش یک تابع هزینه‌ی رتبه‌بند است به این معنی که مانند [۲۰] به ازای برچسب‌هایی که امتیاز بیشتری نسبت به برچسب صحیح کسب کرده‌اند، جریمه در نظر می‌گیرد:

$$L((x_n, y_n); W) = \sum_{y \neq y_n} \max(\bullet, \text{magrin} - x_n W c_{y_n} + x_n W c_y) \quad (۸)$$

* استفاده از مقادیر نورون‌های لایه پگال اول شبکه‌های عصبی به عنوان ویژگی‌های بصری در بسیاری از روش‌های دیگر نیز صورت گرفته است؛ در نتیجه این قسمت جزوی از نگاشت π در نظر گرفته نمی‌شود، بلکه این مقادیر را به عنوان مجموعه تصاویر (X) تلقی می‌کنیم.

۳ روش ارائه شده

۴ نتایج پیاده‌سازی

۵ کارهای آتی

جدول ۱,۵: جدول زمان‌بندی

عنوان فعالیت	مدت زمان لازم	درصد پیشرفت	زمان اتمام
مطالعه روش‌های پیشین	۳ ماه	۱۰۰	شهریور ۹۲
پیاده‌سازی روش‌های پایه و بررسی تأثیر پارامترهای مختلف بر آن‌ها	۱ ماه	۱۰۰	مهر ۹۲
طرح ایده پیشنهادی	۲ ماه	۸۰	آذر ۹۲
پیاده‌سازی روش پیشنهادی، بررسی و مقایسه با سایر روش‌ها	۳ ماه	۶۰	اسفند ۹۲
نگارش مقاله	۱ ماه	۰	فروردین ۹۳
جمع‌بندی و نگارش پایان‌نامه	۲ ماه	۰	خرداد ۹۳

۶ جمع‌بندی

در این گزارش، سیستمی برای دسته‌بندی جویبار داده ارائه کردیم که در آن، برای کاهش هزینه مربوط به برچسب زدن داده‌ها، زیرمجموعه‌ای از داده‌ها را با استفاده از روش انتخاب فعال برای برچسب‌زنی انتخاب می‌کرد و برای برچسب‌زنی آن‌ها، به جای استفاده از افراد متخصص، از جمع‌سپاری استفاده می‌کند. در بخش؟؟، به بررسی ساختار این سیستم و اجزای آن پرداختیم و آن را به سه بخش اصلی سیستم انتخاب فعال، سیستم دسته‌بندی جویبار داده و بخش تجمیع نظرات افراد غیرمتخصص تقسیم کردیم و چالش‌های موجود در هر قسمت را بیان کردیم. در بخش؟؟، به بررسی روش‌های ارائه شده در هر یک از این حوزه‌ها پرداختیم و دلایل قوت و ضعف هر یک را مورد بررسی قرار دادیم. در قسمت ۳، با تمرکز بر روی بخش دسته‌بند نظارتی جویبار داده، چارچوبی کاملاً نظام‌مند و احتمالاتی برای دسته‌بندی جویبار داده ارائه کردیم که با مدل کردن مسئله تغییر مفهوم به عنوان یک مسئله خوشه‌بندی پویا و استفاده از مدل‌های ناپارامتری بیزی به یک روش کارا برای دسته‌بندی دست یافتیم. سپس با مقایسه روش‌های پایه و روش پیشنهادی، موفقیت آن را بر روی یک مجموعه داده واقعی نشان دادیم. نقاط قوت و ضعف برخی روش‌های ارائه شده در حوزه دسته‌بندی جویبار داده نیز در جدول ۲,۶، به اختصار شرح داده شده است.

مراجع

- [1] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-Supervised Learning*. Cambridge, MA: MIT Press, 2006.
- [2] E. G. Miller, *Learning from one example in machine vision by sharing probability densities*. Ph.D. thesis, MIT, 2002.
- [3] S. J. Pan and Q. Yang, "A survey on transfer learning," *Knowledge and Data Engineering, IEEE Transactions on*, vol.22, pp.1345–1359, 2010.
- [4] H. Larochelle, D. Erhan, and Y. Bengio, "Zero-data learning of new tasks," in *National Conference on Artificial Intelligence (AAAI)*, pp.646–651, 2008.
- [5] R. Salakhutdinov, A. Torralba, and J. Tenenbaum, "Learning to share visual appearance for multiclass object detection," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pp.1481–1488, 2011.

جدول ۲,۶: مقایسه روش‌های ارائه شده در حوزه دسته‌بندی جویبار داده

نام روش	سال ارائه	مزایا و معایب
وزن‌دهی به درست‌نمایی [۴]	۲۰۱۱	+ مدلی احتمالاتی و نظام‌مند - استفاده از تنها یک مدل - عدم قابلیت یادگیری مفاهیم تکرارشونده - سرعت پایین در بازیابی دقت پس از تغییر مفهوم
ترکیب دسته‌بندهای پایه [۴]	۲۰۱۱	+ غنی کردن فضای فرضیه با ترکیب مدل‌های ساده + پشتیبانی از مفاهیم تکرارشونده - قوانین به‌روزرسانی مکاشفه‌ای - تعداد زیاد دسته‌بندهای پایه - نداشتن مکانیسمی برای محدود کردن تعداد دسته‌بندها
خوشه‌بندی بردار ویژگی استخراج شده از دسته‌ها [۴]	۲۰۱۰	+ پشتیبانی از مفاهیم تکرارشونده - نداشتن مکانیسمی برای محدود کردن تعداد دسته‌بندها - حساسیت زیاد به پارامترها - یکسان فرض کردن مفهوم تمامی داده‌های یک دسته
روش گروهی مبتنی بر دقت [۴]	۲۰۱۳	+ پشتیبانی از مفاهیم تکرارشونده + روشی احتمالاتی برای انتخاب دسته‌بندی که باید به‌روز شود - روشی مکاشفه‌ای برای انتخاب مفهوم یک داده - یکسان فرض کردن مفهوم تمامی داده‌های یک دسته
روش دسته‌بندی مبتنی بر مدل مخلوط فرآیند دیریکله [۴]	۲۰۰۹	+ ارائه مدلی غیرخطی بر مبنای ترکیب دسته‌بندهای ساده خطی + تعیین تعداد دسته‌بندهای مورد نیاز بر حسب پیچیدگی مدل - عدم پشتیبانی از جویبار داده

- [6] M. Palatucci, G. Hinton, D. Pomerleau, and T. M. Mitchell, "Zero-shot learning with semantic output codes," in *Advances in Neural Information Processing Systems (NIPS)* 22, pp.1410–1418, 2009.
- [7] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing Objects by Their Attributes," in *Computer Vision and Pattern Recognition (CVPR)*, *IEEE Conference on*, pp.1778–1785, 2009.
- [8] R. Socher, M. Ganjoo, C. D. Manning, and A. Ng, "Zero-shot learning through cross-modal transfer," in *Advances in Neural Information Processing Systems (NIPS)* 26, pp.935–943, 2013.
- [9] M. Elhoseiny, B. Saleh, and A. Elgammal, "Write a classifier: Zero-shot learning using purely textual descriptions," in *Computer Vision (ICCV)*, *IEEE Conference on*, pp.2584–2591, 2013.
- [10] M. Norouzi, T. Mikolov, S. Bengio, Y. Singer, J. Shlens, A. Frome, G. Corrado, and J. Dean, "Zero-shot learning by convex combination of semantic embeddings," in *International Conference on Learning Representations (ICLR)*, 2014.
- [11] F. X. Yu, L. Cao, R. S. Feris, J. R. Smith, and S.-F. Chang, "Designing Category-Level Attributes for Discriminative Visual Recognition," in *Computer Vision and Pattern Recognition (CVPR)*, *IEEE Conference on*, pp.771–778, 2013.
- [12] C. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Computer Vision and Pattern Recognition (CVPR)*, *IEEE Conference on*, pp.951–958, 2009.
- [13] V. Vapnik. *Statistical learning theory*. Wiley New York, 1998.
- [14] G. Tsoumakas and Katakis, "Multi Label Classification: An Overview," *International Journal of Data Warehousing and Mining*, vol.3, no.3, pp.1–13, 2007.
- [15] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learnin*. New York: Springer, 2009.

- [16] M. Suzuki, H. Sato, S. Oyama, and M. Kurihara, "Transfer learning based on the observation probability of each attribute," in *Systems, Man and Cybernetics (SMC), IEEE International Conference on*, pp.3627–3631, 2014.
- [17] D. Mahajan, S. Sellamanickam, and V. Nair, "A joint learning framework for attribute models and object descriptions," in *Computer Vision (ICCV), IEEE International Conference on*, pp.1227–1234, 2011.
- [18] X. Yu and Y. Aloimonos, "Attribute-based transfer learning for object categorization with zero/one training example," in *Computer Vision (ECCV), European Conference on*, vol.6315, pp.127–140, 2010.
- [19] X. Wang and Q. Ji, "A unified probabilistic approach modeling relationships between attributes and objects," in *Computer Vision (ICCV), IEEE International Conference on*, pp.2120–2127, 2013.
- [20] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid, "Label-embedding for attribute-based classification," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pp.819–826, 2013.
- [21] B. Romera-Paredes and P. H. S. Torr, "An Embarrassingly Simple Approach to Zero-shot Learning," *Journal of Machine Learning Research*, vol.37, 2015.
- [22] A. Frome, G. S. Corrado, J. Shlens, S. Bengio, J. Dean, M. Ranzato, and T. Mikolov, "DeViSE: A Deep Visual-Semantic Embedding Model," in *Advances in Neural Information Processing Systems (NIPS) 26*, pp.2121–2129, 2013.
- [23] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems (NIPS) 26*, pp.3111–3119, 2013.

۷ واژه‌نامه

^۱ Zero-Shot Learning	^۵ Hamming	^{۱۰} Topic Modeling
^۲ Semi-supervised learning	^۶ Meta-data	^{۱۱} Bayesian Network
^۳ One-shot learning	^۷ Tag	^{۱۲} Structure Learning
^۴ Transfer Learning	^۸ Multilabel Classification	^{۱۳} Regularization Term
	^۹ Logistic Regression	