Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال اول، شماره دوم، پاییز و زمستان ۱۴۰۱، صفحات ۴۳ - ۲۹

تلفیق روش‌های یادگیری شات محدود جهت بهبود عملکرد طبقه‌بندی تصاویر با مجموعه داده‌های کم

علی بشیری [۱]، علی‌محمد لطیف [۲*]

[۱] دانشجوی کارشناسی ارشد دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران

[۲] دانشیاردانشکده مهندسی کامپیوتر – بخش هوش مصنوعی، دانشگاه یزد، یزد، ایران

مقاله پژوهشی

تاریخ دریافت:
۱۴۰۰/۱۱/۲

تاریخ پذیرش:
۱۴۰۱/۰۷/۱۰

کلیدواژه‌ها:
طبقه‌بندی، یادگیری شات محدود، فرایادگیری، یادگیری متریک

نویسنده‌مسئول:
alatif@yazd.ac.ir

چکیده

باوجود پیشرفت چشم‌گیر روش‌های هوش مصنوعی در سال‌های اخیر، همچنان نیاز به داده‌های فراوان برای یادگیری این روش‌ها مشاهده می‌شود. به‌منظور رفع این نیاز، الگوی جدید یادگیری ماشین به نام یادگیری شات محدود پیشنهادشده است. یکی از روش‌های مطرح در این حوزه رویکرد شبکه‌های نمونه اولیه است که درواقع ترکیبی از روش‌های یادگیری متریک و فرا یادگیری است. در این شبکه‌ها، طبقه‌بند سعی می‌کند تا با توجه به تنها تعداد کمی از نمونه‌های هر کلاس جدید، نسبت به این کلاس‌ها تعمیم یابد. در این پژوهش سعی بر آن شد تا شکل تغییریافته‌ای از شبکه‌های نمونه اولیه برای حل مساله طبقه‌بندی شات محدود پیشنهاد شود. در ابتدا به‌منظور بهبود عملکرد در شبکه‌های نمونه اولیه، به‌جای فاصله اقلیدسی از فاصله ماهالانوبیس برای اندازه‌گیری فاصله بین نمونه‌ها استفاده شد. این کار موجب بهبود عملکرد این شبکه‌ها در طبقه‌بندی تصاویر **omniglot** و **miniImageNet** شد به‌طوری‌که شبکه پیشنهادی توانست به ترتیب به دقت‌های ۹۹/۱٪ و ۶۸/۵٪ بر روی این دو مجموعه داده دست یابد. در بخش بعدی رویکردی کلی معرفی شد که می‌تواند معماری شبکه‌های عصبی کانولوشنی را با استفاده از الگوریتم ژنتیک به‌طور خودکار بهبود ببخشد. در این پژوهش از این رویکرد به‌طور خاص بر روی مجموعه داده‌های **omniglot** با معماری اولیه پیشنهادی در شبکه‌های نمونه اولیه استفاده‌شده است. درنهایت با استفاده از این رویکرد و جایگزینی معماری پیشنهادی آن با معماری اصلی شبکه نمونه اولیه دقت شبکه بهبودیافته و توانسته به‌دقت ۹۹/۵٪ دست یابد.

10.22034/ABMIR.2022.2812:

E-ISSN: 2821-2037 /© 2023. Published by Yazd University This is an open access article under the CC BY 4.0 License (https://creativecommons.org/licenses/by/4.0/).

۱- مقدمه

در سال‌های اخیر، به دلیل ظهور سیستم‌های محاسباتی قدرتمند، مجموعه داده‌های بزرگ [۱،۲]، مدل‌ها و الگوریتم‌های پیشرفته [۳]، هوش مصنوعی سرعت خود را افزایش داده تا شبیه انسان عمل کند و انسان را در بسیاری از زمینه‌ها شکست دهد.

روش‌های فعلی هوش مصنوعی باوجوداینکه در سال‌های اخیر پیشرفت چشمگیری در زمینه‌های مختلف داشته‌اند، اما نمی‌توانند خود را با چند نمونه آموزشی کم تعمیم و آموزش دهند. برنامه‌های موفقیت‌آمیز هوش مصنوعی متکی به یادگیری از داده‌های مقیاس بزرگ هستند. در مقابل، انسان‌ها با استفاده ازآنچه درگذشته آموخته‌اند قادر به یادگیری سریع کارهای جدید هستند. به‌عنوان‌مثال، کودکی که چگونگی جمع‌کردن را یاد گرفته می‌تواند به‌سرعت دانش خود را برای یادگیری ضرب با چند مثال انتقال دهد. مثال دیگر این است که با توجه به چند عکس از یک غریبه، کودک به‌راحتی می‌تواند همان شخص را از میان تعداد زیادی عکس شناسایی کند.

برای حل این مساله باید به این سؤال پاسخ داد که چگونه می‌توان برنامه‌های رایانه‌ای را ساخت که به‌صورت تجربی و خودکار بهبود می‌یابند [٤]. به‌منظور پاسخ‌گویی به این سؤال و یادگیری از تعداد محدودی از نمونه‌های آموزشی برچسب زده‌شده، الگوی جدید یادگیری ماشین به نام یادگیری شات محدود [۵] پیشنهادشده است.

یادگیری شات محدود مبتنی بر این مفهوم است که می‌توان الگوریتم‌های قابل‌اعتمادی برای انجام پیش‌بینی بر روی مجموعه داده‌های حداقلی ایجاد کرد. از انگیزه‌های استفاده ازاین‌روش می‌توان به موقعیت‌های زیر اشاره کرد:

- هرگاه کمبود داده‌های تحت نظارت وجود داشته باشد، مدل‌های یادگیری ماشین اغلب قادر به انجام نرمال‌سازی قابل‌اعتماد نیستند.
- هنگام کار با یک مجموعه داده عظیم، برچسب‌گذاری صحیح داده‌ها پرهزینه است.

یکی از شیوه‌های رایج برای بهبود عملکرد و کارایی روش‌های موجود در حوزه‌های مختلف، ترکیب روش‌های مختلف با یکدیگر است. در زمینه یادگیری شات محدود نیز می‌توان از این شیوه بهره برد و روش‌های مختلف را باهم ترکیب کرد. یکی از روش‌های مطرح در این حوزه رویکرد شبکه‌های نمونه اولیه[1] است که درواقع ترکیبی از روش‌های یادگیری متریک و فرایادگیری است. در این شبکه‌ها، طبقه‌بند سعی می‌کند تا با توجه به تنها تعداد کمی از نمونه‌های هر کلاس جدید، نسبت به این کلاس‌ها تعمیم یابد. در این پژوهش سعی بر آن شد تا شکل تغییریافته‌ای از شبکه‌های نمونه اولیه برای حل مساله طبقه‌بندی شات محدود پیشنهاد شود. در ابتدا به‌منظور بهبود عملکرد در شبکه‌های نمونه اولیه، به‌جای فاصله اقلیدسی از فاصله ماهالانوبیس برای اندازه‌گیری فاصله بین نمونه‌ها استفاده شد. در بخش بعدی رویکردی کلی معرفی شد که می‌تواند معماری شبکه‌های عصبی کانولوشنی را با استفاده از الگوریتم ژنتیک به‌طور خودکار بهبود ببخشد. با استفاده از این رویکرد و جایگزینی معماری پیشنهادی آن با معماری اصلی شبکه نمونه اولیه دقت شبکه به‌گونه‌ای مطلوب بهبودیافت.

۲- یادگیری شات محدود

ازآنجاکه یادگیری شات محدود یک زیرمجموعه در یادگیری ماشین است، قبل از تعریف آن بهتر است نگاهی به تعریف یادگیری ماشین در علوم کامپیوتری بیندازیم.

تعریف (۱) [٦]: طبق این تعریف اگر یک برنامه کامپیوتری بتواند با استفاده از کسب تجربه (E) در انجام برخی از وظایف که به آن محول شده است (T) بهبودی را در کارایی (p) سیستم ایجاد کند فرایند یادگیری صورت پذیرفته است.

به‌عنوان‌مثال یک کار طبقه‌بندی تصویر (T) را در نظر بگیرید. یک برنامه یادگیری ماشین می‌تواند دقت طبقه‌بندی (P) را از طریق تجربه (E) کسب‌شده توسط آموزش بر روی تعداد زیادی از تصاویر دارای برچسب بهبود بخشد. مثال دیگری که در این زمینه وجود دارد، برنامه رایانه‌ای آلفاگو[2] است که قهرمان انسانی را در اجرای بازی باستانی (T) گو[3] شکست داده است. این برنامه

[1] Prototypical Networks
[2] AlphaGo
[3] Go

رایانه‌ای با آموزش بر روی یک پایگاه داده (E) که از بازی حدود ۳۰ میلیون نفر از متخصصان انسانی و همچنین بازی مکرر با خود، میزان پیروزی (P) خود را در برابر حریفان بهبود می‌بخشد. خلاصه این مثال‌ها را می‌توان در جدول (۱) قابل‌مشاهده است.

تعریف (۲): یادگیری شات محدود نوعی از مسائل یادگیری ماشین است (که توسط T و P مشخص می‌شود)، و در آن فقط تعداد محدودی از نمونه‌های آموزشی برچسب‌گذاری شده برای هدف T وجود دارد.

جدول (۱): نمونه‌هایی از مسائل یادگیری ماشین بر اساس تعریف (۱)

وظیفه T	تجربه E	کارایی P
طبقه‌بندی تصاویر	تصاویر دارای برچسب در مقیاس بزرگ برای هر کلاس	دقت طبقه‌بندی
بازی باستانی گو	یک پایگاه داده حاوی حدود ۳۰ میلیون حرکت ثبت‌شده از متخصصان انسانی و سوابق بازی خود	نرخ برنده شدن

مسائل موجود در یادگیری شات محدود عمدتاً جز مسائل یادگیری ماشین تحت نظارت[1] هستند. به‌طور مثال، در مسئله طبقه‌بندی [۷]، طبقه بند یادگیری را تنها بر اساس تعداد کمی از نمونه‌های برچسب زده‌شده برای هر کلاس انجام می‌دهد.

طبقه‌بندی شات محدود یک طبقه‌بندی h را یاد می‌گیرد که برچسب yi را برای هر ورودی x پیش‌بینی می‌کند. در مسائل طبقه‌بندی شات محدود معمولاً یک مسئله طبقه‌بندی به‌صورت N-way-K-shot [۸] تعریف می‌شود، که در آن مجموعه آموزشی Dtrain شامل KN نمونه از کلاس‌های N است که در هر کلاس K نمونه وجود دارد.

۳- کارهای پیشین

اگرچه مرز روشنی بین روش‌های یادگیری شات محدود وجود ندارد اما به‌طورکلی این روش‌ها را می‌توان به چهار دسته تقسیم کرد:

- روش‌های یادگیری متریک
- روش‌های فرا یادگیری
- روش‌های داده افزایی
- روش‌های مبتنی بر معنا

۳-۱ روش‌های یادگیری متریک

در روش‌های یادگیری متریک هدف یادگیری نگاشتی از تصاویر به یک فضای تعبیه‌شده است که در آن تصاویر از دسته‌های یکسان به هم نزدیک‌تر بوده و از دیگر دسته‌ها فاصله داشته باشند به امید آن‌که فضای ایجادشده در مورد دسته‌های دیده نشده هم صادق باشد.

مقالات و پژوهش‌های گسترده‌ای در حوزه یادگیری متریک انجام‌شده است [۹،۱۰] که در اینجا برخی از کارهای مرتبط با روش پیشنهادی را به‌طور خلاصه بیان می‌کنیم. در اولین مطالعات صورت گرفته در این حوزه، گلدبرگر و همکاران در سال ۲۰۰۴ مقاله‌ای تحت عنوان آنالیز مؤلفه‌های همسایگی (NCA) [۱۱] ارائه دادند که در آن سعی بر آن دارند که فاصله ماهالانوبیس را برای به حداکثر رساندن دقت K – نزدیک‌ترین همسایه (KNN) در فضای تبدیل‌شده، یاد می‌گیرد.

در پژوهش آن‌ها یک روش جدید برای یادگیری اندازه‌گیری فاصله ماهالانوبیس برای استفاده در الگوریتم طبقه‌بندی KNN پیشنهادشده است. در این الگوریتم تلاش بر این است تا دقت حاصل از الگوریتم KNN را به‌طور تصادفی بر روی مجموعه آموزشی به حداکثر برساند. همچنین این روش می‌تواند یک تعبیه خطی با ابعاد کم از داده‌های دارای برچسب را یاد بگیرد که می‌تواند برای تجسم داده‌ها و طبقه‌بندی سریع مورداستفاده قرار گیرد. برخلاف سایر روش‌ها در آن زمان، مدل طبقه‌بندی آن‌ها غیر پارامتری است و هیچ فرضی در مورد شکل توزیع کلاس یا مرزهای بین آن‌ها وجود ندارد.

در سال ۲۰۱۷ مطالعه‌ای توسط جیک اسنل و همکاران [۱۲] به‌عمل‌آمده است که در آن شبکه‌های نمونه اولیه را برای حل مشکل طبقه‌بندی شات محدود پیشنهاد کردند. شبکه‌های نمونه اولیه یک فضای متریک را یاد می‌گیرند که در آن می‌توان طبقه‌بندی را با محاسبه فاصله هر نمونه از هر کلاس در بازنمایی اولیه انجام داد. تجزیه‌وتحلیل انجام‌شده توسط آن‌ها نشان می‌دهد که برخی از تصمیمات ساده در طراحی معماری شبکه می‌توانند پیشرفت چشمگیری نسبت به رویکردهای اخیر شامل انتخاب‌های پیچیده

[1] Supervised

معماری و روش‌های فرا یادگیری، داشته باشند. در این پژوهش سعی بر آن شده است تا بامطالعه دقیق این کار و پیاده‌سازی آن، بتوان بهبودی بر روی این رویکرد ایجاد کرد.

۳-۲ روش‌های فرا یادگیری

روش‌های فرا یادگیری به‌جای این‌که بر روی مجموعه‌ای از نمونه اشیاء آموزش داده شوند، بر روی مجموعه‌ای از وظایف شات محدود، که به آن‌ها اپیزود هم گفته می‌شود، باهدف آموختن یک استراتژی یادگیری که امکان سازگاری با وظایف جدید را فراهم می‌آورد، آموزش داده می‌شوند.

در مطالعه‌ای فین و همکاران [۱۳] یک الگوریتم برای فرا یادگیری پیشنهاد کردند که یک مدل آگنستیک هست، به این معنی که با هر مدلی که با گرادیان نزولی آموزش داده می‌شود و برای انواع مسائل طبقه‌بندی، رگرسیون و یادگیری تقویتی مناسب است، سازگار هست. در این روش، پارامترهای مدل به‌طور صریح آموزش داده می‌شوند به‌طوری‌که تعداد کمی از مراحل گرادیان با مقدار کمی از داده‌های آموزشی از یک وظیفه جدید، عملکرد نرمال‌سازی خوبی را در آن کار ایجاد می‌کند. درواقع، این روش مدل را آموزش می‌دهد تا به‌راحتی تنظیم شود.

بسیاری از رویکردهای فرا یادگیری برای یادگیری شات محدود، به یادگیرنده‌های پایه ساده مانند طبقه بند نزدیک‌ترین همسایه متکی هستند. بااین‌حال، حتی در روش‌های یادگیری شات محدود، پیش‌بینی کننده‌های خطی آموزش‌دیده شده با اختلاف می‌توانند عمومیت بهتری ارائه دهند.

لی و همکاران [۱٤] پیشنهاد کردند از این پیش‌بینی کننده‌ها خطی به‌عنوان یادگیرنده‌های پایه استفاده شود تا بازنمایی‌هایی را برای یادگیری شات محدود یاد بگیرد و نیز به‌خوبی مسئله میان اندازه ویژگی‌ها و کارایی را در میان طیف وسیعی از مجموعه داده‌های معیار نشان دهد. هدف آن‌ها یادگیری ویژگی‌های تعبیه‌شده‌ای است که به‌خوبی تحت یک طبقه‌بندی خطی برای دسته‌های جدید تعمیم می‌یابد. برای رسیدن به این هدف، از دو ویژگی طبقه بندهای خطی بهره استفاده‌شده است: مشتق‌گیری ضمنی از شرایط بهینه مسئله محدب و فرمول دوگان مسئله بهینه‌سازی. این رویکرد بانام MetaOptNتوانسته است به بهترین میزان عملکرد رویکرد بر روی مجموعه داده‌های miniImageNet، tieredImageNet، CIFAR-FSو FC100 رسد.

۳-۳ روش‌های داده افزایی

به‌موازات ایجاد روش‌های فرا یادگیری، روش‌های داده افزایی مطرح شدند. روش‌های داده افزایی نیز بسیار محبوب هستند. ایده این روش‌ها این است که با تولید و تقویت داده‌های جدید سعی در جبران مجموعه داده‌های محدود شود.

گوو و همکارانش [۱۵] در سال ۲۰۱۷ یک مدل مولد زبان برای تولید جملات ساختگی جدید پیشنهاد کردند که ابتدا یک جمله اولیه را از مجموعه آموزشی به‌عنوان نمونه برداشته و سپس آن را در یک جمله جدید ویرایش می‌کند. در مقایسه با مدل‌های مولد زبان سنتی که جملات را به‌صورت از چپ به راست یا با نمونه‌برداری ازجمله، از ابتدا ایجاد می‌کردند، مدل پیشنهادی آن‌ها ابهامات موجود در مدل‌سازی زبان را بهبود می‌بخشد و بر اساس ارزیابی انسانی، خروجی‌هایی باکیفیت بالاتر تولید می‌کند. علاوه بر این، این مدل یک بردار ویرایش شده نهفته ایجاد می‌کند که معانی قابل تفسیر مانند تشابه جملات و قیاس‌های سطح جمله را به تصویر می‌کشد.

۳-٤ روش‌های مبتنی بر معنا

در حال حاضر، روش‌های مبتنی بر معنا در حال افزایش هستند. از یادگیری صفر شات الهام گرفته‌شده‌اند که طبقه‌بندی فقط بر اساس توضیحات متنی یا ویژگی‌ها انجام می‌شود. درحالی‌که می‌توان از سؤالات معنایی اضافه هنگامی‌که نمونه‌های بصری کمیاب هستند، کمک گرفت. یادگیری از یک یا چند نمونه بصری یکی از قابلیت‌های مهم انسان از بدو تولد است، اما هنوز هم برای سیستم‌های مدرن هوش مصنوعی یک چالش مهم است.

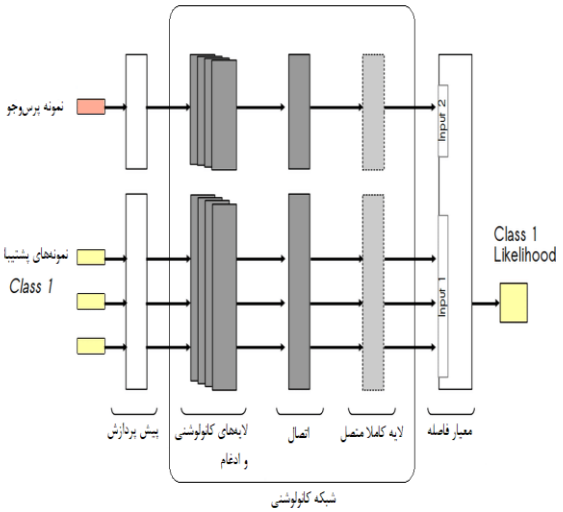
درحالی‌که پیشرفت‌های چشمگیری در یادگیری شات محدود از چند نمونه تصویر حاصل شده است، اما توجه کمتری به توضیحات کلامی که معمولاً هنگام ارائه یک شیء جدید به نوزادان ارائه می‌شود، شده است. شوارتز و همکاران [۱٦] در مقاله‌ای، بر نقش معنایی اضافی متمرکزشده‌اند که می‌تواند یادگیری بصری با تعداد کمی از تصاویر را به‌طور قابل‌توجهی تسهیل کند. با تکیه‌بر پیشرفت‌های اخیر در یادگیری شات محدود با اطلاعات معنایی

اضافی، آن‌ها نشان دادند که پیشرفت‌های بیشتر با استفاده از معناشناسی غنی‌تر و منابع معنایی چندگانه امکان‌پذیر است. با استفاده از این ایده‌ها، آن‌ها توانستند نتایج جدیدی را بر روی مجموعه داده معیار miniImageNet به جامعه علمی ارائه دهند.

٤- روش پیشنهادی

از آن جا که روش پیشنهادی مبتنی بر روش شبکه‌های نمونه‌های اولیه است ابتدا بهتر است نگاهی کلی به این روش و چگونگی کارکرد آن انداخته شود.

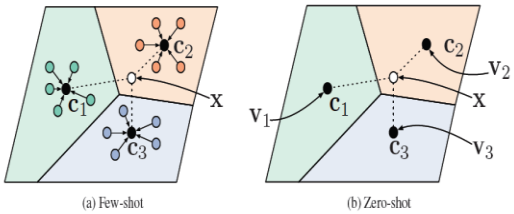
رویکردی که در شبکه‌های نمونه اولیه [۱۲] مطرح‌شده است مبتنی بر این ایده است که یک خوشه‌بندی وجود دارد که در آن نقاط در اطراف یک نمایش نمونه اولیه منفرد برای هر کلاس قرار می‌گیرند. به‌منظور انجام این کار، شبکه یک نگاشت غیرخطی ورودی به فضای نگاشتی که با استفاده از شبکه عصبی ایجادشده است را یاد می‌گیرد و یک نمونه اولیه کلاس را به‌عنوان میانگین مجموعه پشتیبانی آن در فضای نگاشت در نظر می‌گیرد. سپس طبقه‌بندی برای یک نقطه پرس‌وجوی جاسازی‌شده با پیدا کردن نمونه اولیه نزدیک‌ترین کلاس انجام می‌شود. نمودار جریان این شبکه در شکل (۱) آورده شده است. روش پیشنهادی این تحقیق بر دو بخش معیار فاصله و شبکه‌های کانولوشنی تمرکز دارد.



شکل (۱): نمودار جریان شبکه‌های نمونه اولیه

همان‌طور که در شکل (۲) سمت چپ مشاهده می‌کنید نمونه‌های اولیه شات محدود c_k به‌عنوان میانگین نمونه‌های پشتیبانی تعبیه‌شده برای هر کلاس محاسبه می‌شوند. در شکل سمت راست نیز نمونه‌های اولیه صفر شات c_k با استفاده از نگاشت کلاس فراداده v_k تولید می‌شوند. در هر دو حالت، نقاط پرس‌وجوی تعبیه‌شده از طریق یک تابع بیشینه هموار بر اساس فواصل موجود تا نمونه‌های اولیه کلاس مطابق فرمول زیر طبقه‌بندی می‌شوند [۱۲]:

$$p_\varphi(y=k|x) \propto \exp(-d(f_\varphi(x), c_k)) \qquad (۱)$$



شکل (۲): شبکه‌های نمونه اولیه در سناریوهای چند شات و بدون شات[۱۲]

در این پژوهش، شبکه‌های نمونه اولیه فقط برای مسئله یادگیری شات محدود فرمول‌بندی و تنظیم‌شده است. ارتباطات با استفاده از شبکه‌های تطبیقی در محیط یک شات رسم شده‌اند و تابع فاصله اساسی مورداستفاده در مدل تحلیل می‌شود. به‌طور خاص، می‌توان شبکه‌های نمونه اولیه را به مسئله خوشه‌بندی ارتباط داد [۱۷] تا استفاده از میانگین هر کلاس به‌عنوان نمونه‌های اولیه را هنگامی‌که فاصله‌ها با یک واگرایی برگمن محاسبه می‌شوند را توجیه کرد.

در طبقه‌بندی شات محدود یک مجموعه پشتیبانی کوچک از N نمونه برچسب‌گذاری شده وجود دارد که به‌صورت $\{(x_1, y_1), \ldots, (x_N, y_N)\}$ نمایش داده می‌شود و در آن‌ها $x_i \epsilon R^D$ بردار ویژگی D بعدی یک نمونه و $y_i \in \{1, \ldots, K\}$ برچسب مربوطه است. S_k نشان‌دهنده مجموعه‌ای از نمونه‌های برچسب‌گذاری شده باکلاس k است.

شبکه‌های نمونه اولیه درواقع یک بازنمایی M بعدی برای هر $c_k \in R^M$ یا نمونه اولیه از هر کلاس را از طریق تابع نگاشت $f_\varphi: R^D \rightarrow R^M$ با پارامترهای قابل یادگیری φ محاسبه می‌کند. هر نمونه اولیه یک بردار میانگین از نقاط پشتیبانی تعبیه‌شده دارد که متعلق به کلاس خود اوست و از طریق رابطه زیر محاسبه می‌شود [۱۲]:

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\varphi(x_i) \qquad (۲)$$

با توجه به تابع فاصله $d: R^M \times R^M \to [0, +\infty)$، شبکه‌های نمونه اولیه یک توزیع بر روی کلاس‌ها برای نقطه پرس‌وجو x ایجاد می‌کنند که این توزیع بر اساس یک تابع بهینه هموار[1] بر روی فاصله نسبت به نمونه‌های اولیه در فضای موجود عمل می‌کند. تابع این توزیع در فرمول ۳ آورده شده است [۱۲]. در این فرمول k تعداد کلاس‌های درست در فضای ایجادشده است.

$$p_\varphi(y = k|x) = \frac{\exp(-d(f_\varphi(x), c_k))}{\sum_{k'} \exp(-d(f_\varphi(x), c_{k'}))} \qquad (۳)$$

فرایند یادگیری با به حداقل رساندن منفی لگاریتم احتمال $J(\varphi) = -Logp_\varphi(y = k|x)$ از کلاس درست k به‌وسیله الگوریتم بهینه‌سازی گرادیان نزولی تصادفی پیش می‌رود. اپیزودهای آموزشی با انتخاب تصادفی زیرمجموعه‌ای از کلاس‌ها از مجموعه داده آموزشی و سپس انتخاب زیرمجموعه‌ای از نمونه‌ها در هر کلاس برای عمل به‌عنوان مجموعه پشتیبانی که به‌عنوان نقاط پرس‌وجو عمل می‌کنند، مورداستفاده قرار می‌گیرند.

٤-١ رویکرد پیشنهادی اول: جایگزینی معیار فاصله

اولین رویکردی که در این پژوهش مورد آزمایش قرارگرفته است استفاده از فاصله ماهالانوبیس به‌جای فاصله اقلیدسی است. همان‌طور که در نمودار جریان شبکه نمونه اولیه نشان داده شد پس‌ازاینکه نمونه‌های ورودی از طریق یک شبکه کانولوشنی به یک فضای تعبیه‌شده نگاشت شدند حالا نوبت به خوشه‌بندی آن‌ها با استفاده از یک طبقه بند می‌رسد. هر طبقه بند نیاز به یک معیار فاصله برای اندازه‌گیری فاصله بین نقاط پرس‌وجو تا نماینده هر کلاس که به آن نمونه اولیه گفته می‌شود دارد.

در نسخه اصلی شبکه‌های نمونه اولیه [۱۲] از معیار فاصله اقلیدسی برای انجام این کار استفاده‌شده است. بنا به دلایلی که در ادامه توضیح داده خواهد شد، فاصله ماهالانوبیس جایگزین فاصله اقلیدسی شد. دو مشکلی اساسی هنگام استفاده از فاصله اقلیدسی وجود دارد. مشکل اول ناشی از وجود رابطه یا همبستگی بین محورها است. اگر مقدارهای مربوط به محور افقی و عمودی با یکدیگر ارتباط یا همبستگی داشته باشند، فاصله اقلیدسی بزرگ‌تر از حالتی است که مستقل باشند. به‌این‌ترتیب به نظر می‌رسد که فاصله اقلیدسی وزن بیشتری به مشاهدات مرتبط باهم می‌دهد.

به‌این‌ترتیب اطلاع از مختصات یک نقطه می‌تواند در تعیین مختصات نقطه دیگر کمک کند و به‌این‌ترتیب آگاهی از مختصات یک نقطه می‌تواند برای جایگزین اطلاعات مربوط به نقطه دیگر باشد. در حالیکه در فاصله اقلیدسی درکی نسبت به این موضوع وجود ندارد و فاصله‌ها دوباره محاسبه شده و می‌تواند درمجموع، میانگین فاصله‌ها را تحت تأثیر قرار دهد. البته ازآنجایی‌که متغیرها یا مؤلفه‌های مشخص‌شده در مؤلفه‌های اصلی مستقل از یکدیگر ایجاد می‌شوند، معمولاً این مشکل در این حالت بروز نمی‌کند.

مشکل دوم و اساسی هنگام استفاده از فاصله اقلیدسی، محاسبه میانگین و انحراف استاندارد داده‌ها است. اگر واقعاً مجموعه داده شامل نقاط پرت باشد، میانگین و انحراف استاندارد را تحت تأثیر خود قرارداده و محاسبات را دچار مشکل می‌کنند. توجه داشته باشید که داده‌های پرت واقعاً متعلق به توزیع اصلی داده‌ها نیستند. مقدارهای پرت ممکن است براثر خطای اندازه‌گیری یا موارد خاص به وجود آمده باشند که نمی‌توانند گویای رفتار واقعی و دائمی پدیده تصادفی باشند. برای خروج از این بحران از فاصله ماهالانوبیس استفاده می‌کنیم.

از دیگر اعضای واگرایی برگمن فاصله ماهالانوبیس هست که سعی شد در این پژوهش به‌جای فاصله اقلیدسی از این معیار به‌عنوان معیار فاصله برای خوشه‌بندی ورودی‌ها در رویکرد شبکه‌های نمونه اولیه استفاده شود. این تابع فاصله توسط چاندار ماهالانوبیس، دانشمند و آمارشناسی هندی در سال ۱۹۳٦ میلادی معرفی شد.

شیوه محاسبه فاصله ماهالانوبیس برای برداری از نقطه‌های $x=(x_1,x_2,x_3,\dots,x_N)^T$ به‌صورت زیر است. در فرمول ٤ از نماد D_M برای نمایش فاصله نقطه X از توزیعی با بردار میانگین μ و ماتریس واریانس-کواریانس S^2 استفاده‌شده است [۱۲].

$$D_M(x) = \sqrt{(x-\mu)^T S^{-1}(x-\mu)} \qquad (٤)$$

از آن‌جا که در مسائل یادگیری شات محدود تعداد داده‌ها کم بوده و از طرفی فاصله ماهالانوبیس نیازمند معکوس ماتریس کواریانس داده‌ها است این نگرانی ایجاد می‌شود که ماتریس حاصل دچار

[1] Softmax

تکینگی[1] شود، اما در عمل سعی شد تا داده‌ها را در هر اپیزود به گونه‌ای تقسیم کرد که ماتریس کواریانس دچار این مشکل نشود. نکته دیگری که در این زمینه مطرح است این است که استفاده از فاصله ماهالانوبیس به دلیل وجود ماتریس کواریانس داده‌ها می‌تواند مقداردهی اولیه پارامترها را تسهیل نماید.

استفاده از فاصله ماهالانوبیس به‌جای فاصله اقلیدسی و ایجاد یکسری تغییرات دیگر در معماری و الگوریتم پیشنهادی شبکه‌های نمونه اولیه باعث شد تا در بعضی از آزمایش‌ها نتایج بهتری نسبت به روش پیشنهادی اسنل و همکاران اتخاذ شود که در فصل بعد نتایج حاصل از آن را ارائه داده و با نتایج بقیه روش‌ها مقایسه خواهیم کرد.

۴-۲ رویکرد پیشنهادی دوم: طراحی خودکار شبکه عصبی کانولوشنی

شبکه‌های عصبی کانولوشنی (CNN) در سال‌های اخیر موفقیت قابل‌توجهی در وظایف تشخیص تصویر به دست آورده‌اند و در برنامه‌های بینایی کامپیوتری مختلف اعمال می‌شوند [۱۸]. این شبکه‌ها به‌طور قابل‌توجهی دقت را در وظایف طبقه‌بندی تصویر بهبود بخشیدند. بااین‌حال، طراحی چنین شبکه‌هایی نیاز به دانش غنی از حوزه شبکه‌های عصبی کانولوشنی دارد. ازآنجاکه عملکرد CNN ها به‌شدت به داده‌های بررسی‌شده متکی است، محدودیت‌های عمده‌ای برای این روش طراحی وجود دارد.

با توجه به چالش‌های موجود در زمینه طراحی شبکه‌های عصبی کانولوشنی بهتر است رویکردی را ارائه داد تا این شبکه‌ها را به‌صورت خودکار طراحی نماید. در رویکرد پیشنهادی دوم، سعی بر آن شده است تا یک روش فرا ابتکاری کارآمد با استفاده از الگوریتم ژنتیک برای بهبود خودکار معماری یک شبکه‌های عصبی کانولوشنی معین برای رسیدگی ارائه شود. هدف از استفاده از این رویکرد در شبکه‌های نمونه اولیه بهبود عملکرد شبکه عصبی کانولوشنی تعبیه‌شده در این شبکه‌ها است.

برای دستیابی به این هدف، لازم است تا یک رمزگذار انعطاف‌پذیر ژن طراحی شود که حداکثر طول بلوک‌های ساختمانی در CNN را محدود نمی‌کند. با این طرح رمزگذاری ژن، انتظار می‌رود که معماری تکامل‌یافته به‌CNN ها برای دستیابی به عملکرد خوب در حل وظایف مختلف کمک کند. علاوه بر این، توسعه عملگرهای انتخاب، ترکیب و جهش مرتبط که می‌توانند با استراتژی‌های رمزگذاری ژن طراحی‌شده معماری‌ها کنار بیایند، الزامی است.

الگوریتم ۱ ساختار روش پیشنهادی را تشریح می‌کند. در الگوریتم پیشنهادی، با دادن معماری CNN ورودی، اندازه جمعیت و تعداد نسل‌ها، الگوریتم از طریق یک سری فرآیندهای تکاملی شروع به کار می‌کند و درنهایت یک معماری بهبودیافته برای CNN ورودی کشف می‌کند.

الگوریتم ۱: چارچوب الگوریتم پیشنهادی

Input: input CNN architecture, papulation size **P**, number of generation **G**.
Output: optimized input CNN architecture.
population = generate_population(num_pop)
For t ← 0 to G **do**
 For k ← 0 to population **do**
 Select two individual from P_t;
 Generate a new individual i_k using crossover;
 Randomly apply soft mutation to i_k;
 Evaluate the fitness of i_k;
 Evolve P_t with i_k;
 End for
 $P_{t+1} \leftarrow P_t$
End for
Select the best individual from p_t and decode it the corresponding CNN architecture.

در ابتدا، جمعیت اولیه با استفاده از تابع generate_population به‌طور تصادفی ساخته می‌شود. این تابع جمعیت اولیه را بر اساس یکسری مقادیر اولیه از پیش تعریف‌شده برای هر کروموزوم به‌طور تصادفی می‌سازد. سپس، تکامل برای تعداد ورودی نسل‌ها شروع می‌شود.

در هر مرحله از تکامل، یک فرزند جدید محاسبه می‌شود، بر روی مجموعه داده‌های داده‌شده ارزیابی می‌شود و با جمعیت تولیدشده در مرحله قبل مقایسه می‌شود. به‌طور خاص، هر فرد از فرزندان جدید، که معماری خاصی از CNN را رمزگذاری می‌کند، با انتخاب دو فرد از جمعیت، و سپس ترکیب آن‌ها با استفاده از عملگرهای ژنتیکی تولید می‌شود. پس‌ازآن، فرد جدید مورد ارزیابی

[1] Singularity

قرار می‌گیرد و با جمعیت مقایسه می‌شود. جمعیت افراد انتخابی باقی‌مانده به نسل بعدی از جمعیت فعلی تبدیل می‌شوند.

- **استراتژی رمزگذاری**

چهار بلوک مختلف (به‌عنوان‌مثال، لایه کانولوشن، لایه ادغام، لایه حذف و لایه کاملاً متصل) در معماری CNN وجود دارد. بنابراین، آن‌ها باید در یک فرد برای تکامل رمزگذاری شوند. ازآنجاکه هدف از انجام این رویکرد استفاده از آن به‌طور خاص در شبکه نمونه اولیه و بر روی مجموعه داده Omniglot است سعی بر آن شد تا از معماری پیشنهادی آن الگوریتم برای بهبود پارامترهای آن استفاده کرد.

- **ارزیابی شایستگی**

هدف از ارزیابی شایستگی ارائه یک معیار کمی از کیفیت یک فرد است. ازآنجایی‌که الگوریتم پیشنهادی به حل وظایف طبقه‌بندی تصویر می‌پردازد، خطای طبقه‌بندی بهترین استراتژی برای محاسبه تناسب یک فرد است. برای محاسبه خطای طبقه‌بندی، لازم است مجموعه داده اصلی را به یک مجموعه آموزشی D_{train} و یک مجموعه اعتبارسنجی D_{val} تقسیم کرده و سپس هر فرد را در D_{train} آموزش داده و سپس آن را روی D_{val} آزمایش کرد. درنهایت بهترین فرد، فردی است که کمترین مقدار تابع هزینه یا بیشترین مقدار دقت را بر روی مجموعه داده اعتبارسنجی داشته باشد.

- **عملگر انتخاب**

عملگر انتخاب تصمیم می‌گیرد که کدام افراد از جمعیت برای تولید فرزندان جدید در هر تکرار استفاده شوند. روش‌های مختلفی برای انجام این کار وجود دارد. در این رویکرد از روش چرخ رولت که یکی از معروف‌ترین آن‌ها است استفاده می‌کنیم. در این روش هر یک از کروموزوم‌ها بسته به میزان شایستگی احتمال انتخاب شدن دارند. به‌عبارت‌دیگر هر چه یک کروموزوم بهتر باشد احتمال انتخاب شدن آن برای تولید نسل بعدی بیشتر است و برعکس هر چه کروموزوم بدتر باشد، احتمال انتخاب شدن آن برای تولید نسل بعدی کمتر است. رابطه ٤ نحوه انتخاب فرد توسط چرخ رولت را نشان می‌دهد [18].

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_i} \qquad (٤)$$

در این رابطه f_i نشان‌دهنده میزان برازندگی ورودی i و p_i احتمال انتخاب ورودی i را نشان می‌دهد.

- **عملگر ترکیب**

عملگر ترکیب با ترکیب بخش‌هایی از دو فرد منتخب، فرد جدیدی تولید می‌کند. عملگر متقاطع الگوریتم پیشنهادی از متقاطع یک نقطه‌ای در الگوریتم‌های ژنتیک سنتی الهام گرفته‌شده است. در این عملیات ابتدا هر فرد والد به‌طور تصادفی به دو قسمت تقسیم می‌شود و سپس با گرفتن قسمت اول از یک والدین و قسمت دوم از دیگری یک فرد جدید به دست می‌آید.

- **عملگر جهش**

همان‌طور که اشاره شد عملگر جهش باعث ایجاد تنوع ژنتیکی در بین افراد می‌شود و تلاش می‌کند تا از همگرایی الگوریتم ژنتیک به حداقل محلی جلوگیری کند. درروش پیشنهادی برای پیاده‌سازی این عملگر از تغییر تعداد تکرارها در فرایند یادگیری مدل استفاده شد. به این صورت متغیر تصادفی تعریف شد که اگر مقدار آن از حد مشخصی کمتر باشد تعداد تکرارهای فرایند یادگیری به‌طور تصادفی بین ۱ تا ۱۰ گام افزایش می‌یابد. در فصل بعد به بررسی نحوه پیاده‌سازی این رویکرد و نتایج حاصل از آن پرداخته خواهد شد.

- **جمع‌بندی**

در این بخش به بررسی شبکه‌های نمونه اولیه که درواقع مدل اصلی مورداستفاده در این پژوهش است، پرداخته شد و به جزئیات پیاده‌سازی آن به تفکیک بخش‌های مختلف موردبررسی قرار گرفتند. همان‌طور که مشاهده شد این شبکه از دو بخش شبکه عصبی کانولوشنی و یک طبقه‌بند متریک به‌منظور انجام عملیات طبقه‌بندی تصاویر استفاده می‌کند.

در ادامه سعی شد تا با استفاده از دو رویکرد پیشنهادی، یعنی تغییر معیار فاصله و استفاده از الگوریتم ژنتیک در بهبود ساختار شبکه عصبی کانولوشنی شبکه نمونه اولیه، عملکرد این شبکه را بهبود ببخشیم.

در رویکرد اول با توجه به دلایل مطرح‌شده و توجه به این نکته که استفاده از فاصله ماهالانوبیس به دلیل وجود ماتریس کواریانس داده‌ها می‌تواند مقداردهی اولیه پارامترها را تسهیل نکند، معیار

اندازه‌گیری فاصله در طبقه بند شبکه تغییر داده شد. البته باید توجه داشت به دلیل استفاده از معکوس ماتریس کواریانس داده‌ها در فاصله ماهالانوبیس سربار محاسباتی زیادی به شبکه متحمل می‌شود.

در رویکرد پیشنهادی دوم، یعنی استفاده از الگوریتم ژنتیک در بهبود ساختار شبکه عصبی کانولوشنی، سعی شد تا با استفاده از الگوریتم ژنتیک پارامترهای مختلف شبکه عصبی پیشنهادی مانند تعداد لایه‌های کانولوشنی، اندازه فیلترها، توابع فعال‌سازی، الگوریتم بهینه‌سازی را به‌طور خودکار پیش‌بینی و در شبکه استفاده کرد. سپس هر شبکه‌ای که بیشترین دقت طبقه‌بندی را بر روی مجموعه داده‌ها داشت به‌عنوان شبکه بهینه انتخاب‌شده و در شبکه اصلی نمونه‌های اولیه استفاده خواهد شد. این رویکرد یک رویکرد کلی خواهد بود که برای سایر کاربردها نیز قابل‌استفاده است.

۵- ارزیابی روش پیشنهادی

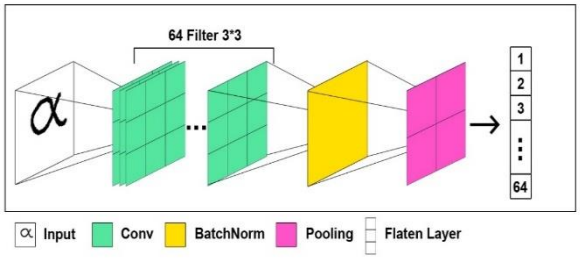
در این پژوهش ارزیابی‌ها بر روی دو مجموعه داده Omniglot و MiniImageNet نسخه ILSVRC2012 انجام‌شده است. خلاصه‌ای از مشخصات این مجموعه داده‌ها در جدول (۲) آورده شده است.

جدول (۲): مجموعه داده‌های تصویر برای محک زدن رویکردهای یادگیری شات محدود

مجموعه داده	توصیف	وضوح تصاویر
Omniglot	۱۶۲۳ کاراکتر دست‌نویس مختلف از ۵۰ زبان مختلف. ۲۰ نمونه برای هر یک از ۱۶۲۳ کاراکتر	۱۰۵×۱۰۵
MiniImageNet	زیرمجموعه Imagenet. تصاویر دنیای واقعی. ۳۸۴۰۰ تصویر آموزشی، ۱۲۰۰۰ تصویر تست و ۹۶۰۰ تصویر اعتبار سنجی	۵۰۰×h

رویکرد اول: تغییر معیار فاصله در الگوریتم شبکه نمونه اولیه

در این پژوهش روش اسنل و همکارانش دنبال شده است [12]. این کار با تغییر اندازه تصاویر در مقیاس خاکستری به 1×۲۸×۲۸ و افزایش کلاس کاراکترها به‌وسیله چرخش‌های متعدد ۹۰ درجه انجام شد. از ۱۲۰۰ کاراکتر به‌علاوه چرخش (درمجموع ۴۸۰۰ کلاس) برای آموزش و از مابقی تصاویر به همراه چرخش برای آزمایش استفاده شد. معماری که توسط اسنل و همکاران معرفی شد، از چهار بلوک متحرک تشکیل‌شده است. هر بلوک شامل ۶۴ فیلتر ۳ × ۳ کانولوشن، لایه نرمال‌سازی دسته‌ای، یک تابع غیرخطی ReLU و یک‌لایه ادغام 2× 2 هست. شمای کلی این بلوک‌های می‌توانید در شکل ۳ مشاهده کنید.



شکل (۳): شمای کلی بلوک‌های کانولوشنی

هنگامی‌که این شبکه با این معماری روی تصاویر Omniglot 28 × 28 اعمال می‌شود، منجر به ایجاد فضای خروجی ۶۴ بعدی خواهد شد. در این پیاده‌سازی از یک رمزگذار یکسان هم برای نقاط پشتیبانی و هم برای پرس‌وجو استفاده‌شده است. درنهایت مدل پیشنهادی از طریق الگوریتم بهینه‌سازی آدام[1] آموزش‌دیده شد. نرخ یادگیری اولیه ۳-۱۰ و تعداد قسمت‌های[2] اجرایی شبکه ۲۰۰ بار در نظر گرفته‌شده است. همچنین به‌منظور مقایسه عملکرد روش پیشنهادی با دیگر روش‌های موجود در حوزه یادگیری شات محدود که همگی تنها بر اساس دقت روش خود را ارزیابی کرده‌اند در این پژوهش نیز از معیار دقت برای این منظور استفاده شد.

در ادامه شبکه‌های نمونه اولیه با استفاده از فاصله ماهالانوبیس بجای فاصله اقلیدسی در سناریوهای ۱ شات و ۵ شات با اپیزودهای آموزشی شامل ۶۰ کلاس و ۵ نقطه پرس‌وجو در هر کلاس آموزش داده شد. طبق نتایج حاصل‌شده دریافت شد که مطابقت دادن عکس آموزشی با عکس آزمایشی و استفاده از کلاس‌های بیشتر در هر اپیزود آموزشی، مفید است.

[1] Adam

[2] Epoch

نتایج حاصل از رویکرد پیشنهادی با پژوهش‌های پایه مختلف، ازجمله آمار عصبی [23]، Meta - Learner LSTM [22]، MAML[13] و هر دو نسخه تنظیم دقیق و تنظیم‌نشده شبکه‌های تطبیقی [19] مقایسه شد. نتایج به‌دست‌آمده در جدول ۳ نشان داده‌شده است و همان‌طور که مشاهده می‌کنید با روش‌های مشابه موجود در این مجموعه داده قابل‌رقابت است.

جدول (۳): دقت طبقه‌بندی یادگیری شات محدود بر روی مجموعه داده Omniglot

مدل	معیار فاصله	دقت 5-way		دقت 20-way	
		۱ شات	۵ شات	۱ شات	۵ شات
آمار عصبی [۲۳]	کسینوسی	۹۸٪/۱	۹۹٪/۵	۹۳٪/۲	۹۸٪/۱
MAML [۱۳]	–	۹۸٪/۷	۹۹٪/۹	۹۵٪/۸	۹۸٪/۹
شبکه‌های تطبیقی [۱۹]	کسینوسی	۹۸٪/۱	۹۸٪/۹	۹۳٪/۸	۹۸٪/۵
شبکه‌های تطبیقی [۱۹]	–	۹۷٪/۹	۹۸٪/۷	۹۳٪/۵	۹۸٪/۷
شبکه‌های نمونه اولیه [۱۲]	اقلیدسی	۹۸٪/۸	۹۹٪/۷	۹۶٪/۰	۹۸٪/۹
روش پیشنهادی	ماهالانوبیس	۹۸٪/۹	۹۹٪/۶	۹۷٪/۱	۹۹٪/۱

مجموعه داده miniImageNet که در اصل توسط وینیلز و همکارانش [7] ارائه‌شده از مجموعه داده بزرگ‌تر ImageNet ILSCVR 12 استخراج‌شده است [24]. تقسیمات[1] مورداستفاده توسط وینیلز و همکارانش شامل ۶۰۰۰۰ تصویر رنگی با اندازه ۸۴×۸۴ هست که این تعداد داده در ۱۰۰ کلاس که هرکدام شامل ۶۰۰ نمونه تصویر است، تقسیم‌بندی شده‌اند. در آزمایش‌های انجام‌شده در این پژوهش، از تقسیمات معرفی‌شده توسط راوی و لاراشل [20] به‌منظور مقایسه مستقیم با الگوریتم‌های سطح بالا برای یادگیری شات محدود استفاده می‌شود.

تقسیمات آن‌ها از مجموعه متفاوتی از ۱۰۰ کلاس استفاده می‌کند که به ۶۴ کلاس آموزشی، ۱۶ کلاس اعتبارسنجی و ۲۰ کلاس آزمایشی تقسیم‌شده‌اند. در این پژوهش از روش آن‌ها با آموزش در ۶۴ کلاس آموزشی و با استفاده از ۱۶ کلاس اعتبارسنجی برای نظارت بر عملکرد تعمیم استفاده‌شده است.

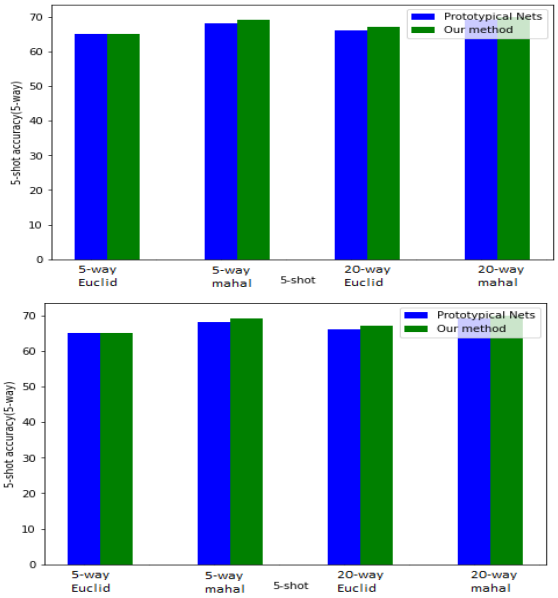
در این بخش نیز از همان معماری چهار بلوکی استفاده می‌شود که در آزمایش‌های omniglot بخش قبل بکار رفته است، اگرچه در اینجا به دلیل افزایش اندازه تصاویر، فضای خروجی ۱۶۰۰ بعدی ایجاد می‌شود. در این بخش از اپیزودهای ۳۰ کلاس برای طبقه‌بندی ۱ شات و از اپیزودهای ۲۰ کلاس برای طبقه‌بندی ۵ شات استفاده‌شده است.

در ادامه روش پیشنهادی با روش پایه‌ای که توسط راوی و لاروشل گزارش‌شده است مقایسه می‌شود [20]، که شامل یک رویکرد ساده از نزدیک‌ترین همسایه بر ویژگی‌هایی است که یک شبکه طبقه‌بندی در ۶۴ کلاس آموزشی آموخته است. روش‌های پایه دیگر که روش پیشنهادی با آن‌ها مقایسه شد عبارتنداز دو نوع شبکه‌های منطبق (معمولی و FCE) و Meta-Learner LSTM هستند که به‌خوبی تنظیم‌نشده‌اند. همان‌طور که در جدول ٤ مشاهده می‌شود، شبکه نمونه اولیه پیشنهادی با جایگزینی معیار فاصله توانسته است تا نسبت به شبکه‌های نمونه اولیه دقت طبقه‌بندی را هم در ۱ شات و هم در ۵ شات بهبود ببخشد.

جدول (٤): دقت‌های طبقه‌بندی شات محدود بر روی مجموعه داده miniImageNet.

مدل	معیار فاصله	دقت 5-way	
		۱ شات	۵ شات
فرا یادگیرنده LSTM [23]	–	٤٣٪/٤٤	٦٠٪/٦٠
MAML [13]	–	٤٨٪/٧٠	٦٣٪/١٥
شبکه‌های تطبیقی [19]	کسینوسی	٤٣٪/٤٠	٥١٪/٠٩
شبکه‌های تطبیقی FCE[19]	کسینوسی	٤٣٪/٥٦	٥٥٪/٣١
شبکه‌های نمونه اولیه [12]	اقلیدسی	٤٩٪/٧٠	٦٨٪/٢٠
روش پیشنهادی	ماهالانوبیس	٤٩٪/٦٩	٦٨٪/٥٠

[1] Split



شکل (٤): مقایسه تأثیر معیار فاصله و تعداد کلاس‌ها در هر قسمت آموزشی بر دقت طبقه‌بندی ۵ کلاس برای شبکه نمونه اولیه و شبکه‌های پیشنهادی ما بر روی مجموعه داده miniImageNet.

در ادامه سعی بر آن شد تجزیه‌وتحلیل بیشتری انجام شود تا تأثیر معیار فاصله و تعداد کلاس‌های آموزشی در هر قسمت بر عملکرد شبکه‌های نمونه اولیه و روش پیشنهادی تعیین شود.

در شکل (٤) فاصله اقلیدسی را در مقابل فاصله ماهالانوبیس و اپیزودهای آموزشی ۵ کلاس در مقابل ۲۰ کلاس در سناریوهای ۱ شات و ۵ شات با ۱۵ نقطه پرس‌وجو در هر کلاس در هر اپیزود مقایسه می‌شود. همان‌طور که در شکل مشخص است ۲۰ کلاس به‌دقت بیشتری نسبت به ۵ کلاس دست پیدا می‌کند و افزایش دشواری طبقه‌بندی ۲۰ کلاس به شبکه کمک می‌کند تا بهتر تعمیم یابد، زیرا مدل را مجبور می‌کند در فضای تعبیه‌شده تصمیمات دقیق‌تری اتخاذ کند. همچنین استفاده از فاصله ماهالانوبیس عملکرد را تا حد قابل‌قبولی در مقایسه بافاصله اقلیدسی بهبود می‌بخشد.

رویکرد دوم: طراحی خودکار معماری بلوک‌های کانولوشنی

در این رویکرد، ابتدا روش پیشنهادی به‌طور آزمایشی بر روی کار طبقه‌بندی تصویر با استفاده از مجموعه داده CIFAR-10 [25] آزمایش شد که در آن تعداد کلاس‌ها ۱۰ است. تعداد تصاویر آموزشی و اعتبارسنجی به ترتیب ۵۰۰۰۰ و ۱۰۰۰۰ است و اندازه هر تصویر ۳۲×۳۲ است. پس از اطمینان از صحت اجرایی بودن کار، روش پیشنهادی بر روی مجموعه داده اصلی شبکه یعنی مجموعه داده Omniglot بررسی شد. ازآنجاکه این مجموعه داده به‌منظور یادگیری و آزمایش بر روی روش‌های یادگیری شات محدود توسعه داده‌شده است لذا فرایند یادگیری آن نیز متفاوت از مجموعه داده‌های دیگر است.

در الگوریتم پیشنهادی، انتخاب پارامترها باکیفیت راه‌حلی که فرد می‌خواهد به دست آورد، منابع محاسباتی موجود و اندازه مجموعه داده در نظر گرفته‌شده تعیین می‌شود. مصالحه بین این مشخصات، زمان محاسباتی موردنیاز الگوریتم برای همگرایی به یک راه‌حل را مشخص می‌کند. به‌طور خاص، اگر منابع محاسباتی موجود فقط محدود باشد، زمان محاسباتی موردنیاز الگوریتم پیشنهادی ممکن است چالش‌برانگیز شود.

هر شبکه عصبی کانولوشنی پارامترهای متعددی دارد که می‌توان توسط الگوریتم ژنتیک مقدار آن‌ها را بهبود بخشید. اما در این پژوهش پارامترهای انتخابی که به‌عنوان کروموزوم‌های الگوریتم ژنتیک در نظر گرفته‌شده‌اند، به شرح زیر می‌باشند:

- تعداد فیلترهای هر بلوک
- اندازه فیلترها
- نرخ یادگیری
- گام حرکت
- تابع فعال‌سازی
- تعداد تکرارها در فرایند یادگیری

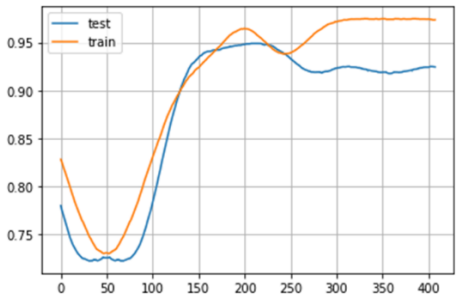
هر یک از این پارامترها به‌عنوان کروموزوم‌های الگوریتم ژنتیک در نظر گرفته می‌شود که در هر بار اجرا الگوریتم مدل را به یکی از مقادیر مختلف برای آن‌ها آموزش می‌دهد و دقت حاصل از آن را به‌عنوان میزان شایستگی آن مدل در نظر می‌گیرد. با این کار مدل با بهترین مقادیر برای پارامتر به‌عنوان مدل نهایی معرفی می‌شود.

الگوریتم پیشنهادی بر روی یک معماری پایه CNN برای مجموعه داده Omniglot آزمایش‌شده و CNN حاصل ارائه و تحلیل می‌شود. معماری CNN مورداستفاده برای تکامل شامل ٤ بلوک پی‌درپی است که در هر بلوک تعدادی لایه کانولوشنی، لایه

نرمال‌سازی، لایه ادغام و لایه کاملاً متصل است. هدف نهایی این روش پیدا کردن پارامترهای بهینه برای لایه‌های هر بلوک است.

نتایج حاصل از آزمایش نشان می‌دهد که مدل شبکه عصبی کانولوشنی پایه در صورت داشتن پارامترهایی به تعداد ۱۲۸ فیلتر با اندازه ۳*۳، گام حرکت ۱، نرخ یادگیری ۰/۰۰۳، تابع فعال‌سازی relu در ۴۰۰ تکرار توانست بهترین نتیجه را بر روی مجموعه داده Omniglot خواهد داشت که دقتی در حدود ۰/۹۳ خواهد داشت.

مدل CNN حاصل که توسط الگوریتم پیشنهادی تولیدشده است، پس از آموزش برای ۳۰۰ دوره، بر روی مجموعه داده‌ای با ۲۰ کلاس که در هر کلاس ۵ داده وجود دارد (20 way-5 shot)، دقتی در حدود ۹۳٪ بر روی مجموعه داده آزمایشی دارد که در شکل ۵ قابل‌مشاهده است.



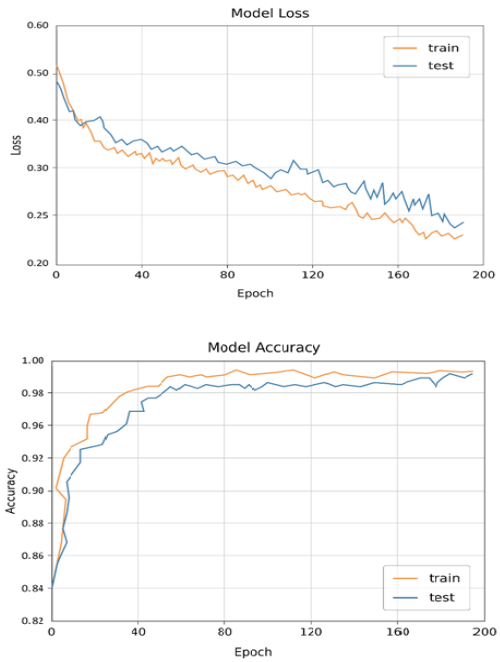
شکل (۵): دقت حاصل از مدل پیشنهادی بر روی مجموعه داده Omniglot

پس‌ازاینکه مدل بهینه توسط الگوریتم پیشنهاد شد حال بایستی این معماری پیشنهادی را با معماری اصلی در شبکه‌های نمونه اولیه جایگزین کرده و شبکه را مجدداً آموزش داده و نتایج را موردبررسی قرار دهیم. ازآنجاکه مدل پیشنهادی توسط الگوریتم فقط به‌صورت 20 way-5 shot موردبررسی قرارگرفته است لذا در شبکه‌های نمونه اولیه هم فقط به همین صورت موردبررسی قرار داده‌شده است و از روش 1 shot چشم‌پوشی شد.

جدول ۵: دقت طبقه‌بندی یادگیری شات محدود با استفاده از معماری پیشنهادی بر روی مجموعه داده Omniglot

مدل	معیار فاصله	دقت20way-5 shot
آمار عصبی [23]	کسینوسی	۹۸٪/۱
MAML [13]	–	۹۸٪/۹
شبکه‌های تطبیقی [19]	کسینوسی	۹۸٪/۵
شبکه‌های تطبیقی [19]	–	۹۸٪/۷
شبکه‌های نمونه اولیه [12]	اقلیدسی	۹۸٪/۹
روش پیشنهادی	**اقلیدسی**	**۹۹٪/۵**

شکل (٦): نمودارهای دقت و هزینه شبکه نمونه اولیه با معماری پیشنهادی بر روی مجموعه داده Omniglot



با جایگزینی مدل پیشنهادی با مدل اصلی دقت شبکه نمونه اولیه به ۹۹/۵٪ افزایش پیدا کرد. این نتیجه نشان می‌دهد که با تغییر معماری شبکه CNN الگوریتم شبکه نمونه اولیه می‌تواند تا حد قابل قبولی دقت خود را بر مجموعه داده Omniglot افزایش دهد. نتایج به‌دست‌آمده در جدول ۵ نشان داده‌شده است و همان‌طور که مشاهده می‌کنید با روش‌های مشابه موجود در این مجموعه داده قابل‌رقابت است.

استفاده از مدل شبکه عصبی کانولوشنی ایجادشده توسط الگوریتم پیشنهادی در شبکه‌های نمونه اولیه توانست به‌دقت ۹۹٪/۵ و زیان[1] ۲٪/۳ در 200 دوره آموزشی دست یابد. منحنی‌های دقت و زیان در شکل ٦ نشان داده‌شده است.

[1] Loss

٦- نتیجه‌گیری

به‌طورکلی هدف از این پژوهش ترکیب روش‌های یادگیری متریک و فرا یادگیری با یکدیگر و بررسی عملکرد این روش ترکیبی بر روی مجموعه داده‌های مختلف بود. برای این منظور رویکرد شبکه‌های نمونه اولیه به‌عنوان روش مبنا در نظر گرفته شد که درواقع یکی از پرکاربردترین روش‌ها برای انجام این کار است. سپس سعی شد هر دو بخش متریک و اپیزودیک (فرا یادگیری) این شبکه را به‌طور اصولی و علمی مورد بازیابی و تغییر قرار داده شود تا به نتایج مطلوب‌تری دست‌یافت.

در ادامه سعی شد تا با استفاده از دو رویکرد پیشنهادی، یعنی تغییر معیار فاصله و استفاده از الگوریتم ژنتیک در بهبود ساختار شبکه عصبی کانولوشنی شبکه نمونه اولیه، عملکرد این شبکه را بهبود ببخشیم.

در رابطه با رویکرد اول، نتایج به‌دست‌آمده نشان می‌دهد که این روش نتوانسته به‌طور چشمگیری دقت طبقه‌بندی شبکه را بهبود ببخشد. همچنین به دلیل استفاده از معکوس ماتریس کواریانس داده‌ها در فاصله ماهالانوبیس سربار محاسباتی زیادی به شبکه متحمل می‌شود که با توجه به بهبود نه‌چندان چشم‌گیر در عملکرد شبکه، لذا استفاده از آن چندان به‌صرفه نخواهد بود.

به‌طورکلی هدف از این پژوهش ترکیب روش‌های یادگیری متریک و فرا یادگیری با یکدیگر و بررسی عملکرد این روش ترکیبی بر روی مجموعه داده‌های مختلف بود. برای این منظور رویکرد شبکه‌های نمونه اولیه به‌عنوان روش مبنا در نظر گرفته شد که درواقع یکی از پرکاربردترین روش‌ها برای انجام این کار است. سپس سعی شد هر دو بخش متریک و اپیزودیک (فرا یادگیری) این شبکه را به‌طور اصولی و علمی مورد بازیابی و تغییر قرار داده شود تا به نتایج مطلوب‌تری دست‌یافت.

در ادامه سعی شد تا با استفاده از دو رویکرد پیشنهادی، یعنی تغییر معیار فاصله و استفاده از الگوریتم ژنتیک در بهبود ساختار شبکه عصبی کانولوشنی شبکه نمونه اولیه، عملکرد این شبکه را بهبود ببخشیم.

در رابطه با رویکرد اول، نتایج به‌دست‌آمده نشان می‌دهد که این روش نتوانسته به‌طور چشمگیری دقت طبقه‌بندی شبکه را بهبود ببخشد. همچنین به دلیل استفاده از معکوس ماتریس کواریانس داده‌ها در فاصله ماهالانوبیس سربار محاسباتی زیادی به شبکه متحمل می‌شود که با توجه به بهبود نه‌چندان چشم‌گیر در عملکرد شبکه، لذا استفاده از آن چندان به‌صرفه نخواهد بود.

References

[1] M. A. Turing, Computing machinery and intelligenc. Mind 59, 236, 433–433, 1950.
[2] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li and L. Fei-Fei. ImageNet, A large-scale hierarchical image database. Computer Vision and Pattern Recognition. 248–255, 2009.
[3] A.Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks, Neural Information Processing Systems, 1097–1105, 2012.
[4] M. Mohri, A. Rostamizadeh, and A. Talwalkar.Foundations of machine learning, MIT Press, 2018.
[5] S.Kadam, V.Vaidya, Review and Analysis of Zero, One and Few Shot Learning Approaches, Intelligent Systems Design and Applications, 100-112, 2018.
[6] S. Hochreiter and J. Schmidhuber, Long short-term memory, Neural Computation, 1735–1780, 1997.
[7] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, et al. Matching networks for one shot learning. InAdvances inNeural Information Processing Systems. 3630–3638, 2016.
[8] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. InInternational Conference on Machine Learning. 1126–1135, 2017.
[9] Aurélien Bellet, Amaury Habrard, and Marc Sebban. A survey on metric learning for feature vectors and structured data, 2013.
[10] Brian Kulis. Metric learning: A survey. Foundations and Trends in Machine Learning, 5(4):287–364, 2012.
[11] Jacob Goldberger, Geoffrey E. Hinton, Sam T. Roweis, and Ruslan Salakhutdinov. Neighbourhood components analysis. In Advances in Neural Information Processing Systems, pages 513–520, 2004.
[12] J. Snell, K. Swersky, and R. S. Zemel. Prototypical networks for few-shot learning. InAdvances in NeuralInformation Processing Systems. 4077–4087, 2017.
[13] Chelsea Finn, Pieter Abbeel, Sergey Levine; “Model-agnostic meta-learning for fast adaptation

of deep networks.", Proceedings of the 34th International Conference on Machine Learning - Volume 70, Pages 1126-1135, 2017.
[14] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, Stefano Soatto; "Meta-Learning with Differentiable Convex Optimization.", The IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[15] Kelvin Guu, Tatsunori B. Hashimoto, Yonatan Oren and Percy Liang; "Generating Sentences by Editing Prototypes.", Transactions of the Association for Computational Linguistics, Volume 6, p.437-450, 2018.
[16] Eli Schwartz, Leonid Karlinsky, Rogerio Feris, Raja Giryes, Alex M. Bronstein; "Baby steps towards few-shot learning with multiple semantics.", arXiv:1906.01905v1 [cs.CV], 2019.
[17] Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, and Joydeep Ghosh. Clustering with bregman divergences. Journal of Machine Learning Research, 6(Oct):1705–1749, 2005.
[18] D. Goldberg and J. Holland, Genetic algorithms and machine learning, Machine Learning, vol. 3, 95-99, 1988.
[19] Oriol Vinyals, Charles Blundell, Tim Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In Advances in Neural Information Processing Systems, pages 3630–3638, 2016.
[20] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. International Conference on Learning Representations, 2017.
[21] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In IEEE Computer Vision and Pattern Recognition, pages 1–9, 2015.
[22] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. International Conference on Machine Learning, 2017.
[23] Harrison Edwards and Amos Storkey. Towards a neural statistician. International Conference on Learning Representations, 2017.
[24] L. van d. Maaten and G. Hinton, Visualizing data using t-sne, Journal of Machine Learning Research, 2579–2605, 2008.
[25] A.Krizhevsky, Learning multiple layers of features from tiny images, Technical report, 2009.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 1, No. 2, Autumn & Winter 2022, pp. 29 - 43

Integration of few-shot learning methods to improve image classification performance with small data sets

Ali Bashiri, Ali Mohammad Latif*

[1]Computer Engineering Department, Yazd University, Yazd, Iran

Article Information

Original Research Paper

Received:
2022 January 22

Accepted:
2022 October 02

Keywords:
Classification, few-shot learning, meta-learning, metric learning

Corresponding Author*:
alatif@yazd.ac.ir

Abstract

Despite the significant advancement of artificial intelligence methods in recent years, there is still a need for a lot of data to learn these methods. To meet this need, a new machine learning model called Few-shot learning has been proposed. One of the methods in this field is the prototypical networks approach, which is actually a combination of metric learning and meta-learning methods. In these networks, the classifier tries to generalize to these classes according to only a small number of samples of each new class. In this study, an attempt was made to propose a modified form of the prototype networks to solve the finite shot classification problem. Initially, in order to improve the performance of the prototype networks, instead of the Euclidean distance, the Mahalanobis distance was used to measure the distance between the samples. This improved the performance of these networks in classifying omniglot and miniImageNet images so that the proposed network was able to achieve 99.1% and 68.5% accuracy on these two datasets, respectively. The next section introduces a general approach that can automatically improve the architecture of convolutional neural networks using a genetic algorithm. In this research, this approach has been used specifically on omniglot datasets with the proposed primary architecture in prototype networks. Finally, by using this approach and replacing the proposed architecture with the main architecture of the prototype network, the network accuracy was improved and was able to achieve 99.5% accuracy.

10.22034/ABMIR.2022.2812:

E-ISSN: 2821-2037 /© 2023. Published by Yazd University This is an open access article under the CC BY 4.0 License (https://creativecommons.org/licenses/by/4.0/).