

ماشین یادگیری مفرط عمیق: رویکرد ترکیبی یادگیری افزایشی برای طبقه‌بندی داده‌های جریانی

جواد حمیدزاده و منا مرادی

مناسب به نظر نمی‌رسند. منابع تولیدکننده داده‌های جریانی، توان پردازشی و حافظه محدودی دارند، از سوی دیگر امکان دسترسی نامحدود به داده‌ها وجود نداشته و داده‌های ذخیره‌شده در حافظه نمی‌توانند چندین بار مورد دسترسی قرار بگیرند. در نتیجه میزان حافظه مورد نیاز در این منابع باید تا حد امکان کاهش یافته و در عین حال کیفیت نتیجه طبقه‌بندی نیز حفظ شود.

ورود پیوسته داده‌ها: داده‌های جریانی یکی پس از دیگری و به ترتیب زمانی به سیستم وارد می‌شوند؛ بنابراین لازم است مدل یادگیرنده با ورود اطلاعات جدید، سریعاً آموزش دیده و به روز رسانی شود.

نویز: یکی از مشکلات رایج داده‌کاوی، وجود نویز در داده‌ها است که ممکن است به علت برچسب‌گذاری اشتباه داده‌ها و یا نویز در مقادیر ویژگی‌ها رخ دهد. در داده‌های جریانی، همراهی نویز با تغییر مفهوم، مسئله را دشوارتر می‌سازد زیرا ممکن است نویز به اشتباه، تغییر مفهوم تلقی شود.

رانش مفهوم: منظور از مدل یا مفهوم، قواعد و روابطی است که به کمک آنها می‌توان داده‌ها را به طبقه‌های مختلف منتسب کرد. در داده‌های جریانی، توزیع داده‌ها از قبل مشخص نبوده و ممکن است متأثر از عوامل بیرونی، پیوسته با زمان تغییر کند. در این صورت، مدلی که برای داده‌های گذشته معتبر بوده است با ورود داده‌های جدید، کارایی خود را از دست خواهد داد. در یک دسته‌بندی کلی، رانش مفهوم از نظر سرعت وقوع به پنج دسته تقسیم می‌شود:

رانش مفهوم تدریجی^۱: در این نوع رانش، یک تغییر هموار و تدریجی از مفهومی به مفهوم دیگر رخ می‌دهد. مفهوم میانی تولیدشده، یکی از مفاهیم آغازین یا پایانی است.
 رانش مفهوم ناگهانی^۲: در این نوع رانش، یک تغییر ناگهانی در محتوای کلاس رخ داده و کلاس دیگری ظاهر می‌شود.
 رانش مفهوم افزایشی^۳: در این نوع رانش، ضمن تغییر از مفهومی به مفهوم دیگر، چندین مفهوم میانی دیگر ظاهر می‌شوند. مفاهیم میانی تولیدشده، ترکیبی از مفاهیم آغازین و پایانی هستند.
 رانش مفهوم بازگشتی^۴: در این نوع رانش، پس از گذشت زمان، مفاهیم از قبل مشاهده شده، مجدداً ظاهر خواهند شد.
 رانش کوتاه^۵: در این نوع رانش، داده پرت با مفهوم موجود آمیخته

چکیده: داده‌های جریانی متشکل از داده‌هایی است که به ترتیب و با سرعت و حجم زیاد به سیستم وارد می‌شوند. توزیع این داده‌ها ناپایدار بوده و در طول زمان ممکن است تغییر کنند. با توجه به اهمیت این نوع داده‌ها در حوزه‌هایی مهم نظیر اینترنت اشیا، تسریع عملکرد و افزایش توان عملیاتی تحلیل داده‌های بزرگ جریانی به عنوان موضوعی مهم، مورد توجه محققین است. در روش پیشنهادی، از مفهوم یادگیری ترکیبی برخط در مدل بهبودیافته ماشین یادگیر مفرط به منظور طبقه‌بندی داده‌های جریانی استفاده شده است. به دلیل استفاده از رویکرد افزایشی، در هر لحظه تنها یک بلوک داده بدون نیاز به دسترسی به داده‌های پیشین یاد گرفته می‌شود. همچنین با بهره‌گیری از رویکرد آداپتو، وزن‌دهی به طبقه‌بندی‌کننده‌های پایه و تصمیم‌گیری در مورد حفظ و یا حذف آنها بر اساس کیفیت پیش‌بینی‌ها انجام می‌شود. مزیت دیگر روش پیشنهادی، بهره‌گیری از رویکرد مبتنی بر صحت طبقه‌بندی‌کننده جهت شناسایی رانش مفهوم است که منجر به تسهیل انطباق مدل و افزایش کارایی آن می‌شود. آزمایش‌ها بر روی مجموعه داده‌های استاندارد انجام گردید و روش پیشنهادی به طور میانگین با کسب ۰/۹۰٪ خاص بودن، ۰/۶۹٪ حساسیت و ۰/۸۷٪ صحت توانست اختلاف معناداری با دو روش رقیب داشته باشد.

کلیدواژه: داده‌های جریانی، رانش مفهوم، ماشین یادگیری مفرط، یادگیری افزایشی.

۱- مقدمه

کاوش در داده‌های جریانی^۱ در سال‌های اخیر توجه بسیاری از محققان را به خود جلب کرده است. داده‌های جریانی یک رشته از داده‌ها مانند $x_1, x_2, \dots, x_i, \dots$ هستند که در زمان‌های $T_1, T_2, \dots, T_i, \dots$ ($T_i < T_{i+1}$) به صورت پیوسته تولید می‌شوند و نیازمند پاسخ سریع می‌باشند. داده‌های جریانی دارای ویژگی‌هایی هستند که آنها را از داده‌های سنتی متمایز می‌کند [۱]:

حجم زیاد داده‌ها: داده‌های جریانی به صورت پیوسته و در حجم بالا تولید می‌شوند. بنابراین امکان ذخیره‌سازی تمام آنها وجود ندارد و در نتیجه، فرض در دسترس بودن همه داده‌ها در یک زمان صحیح نیست.
سرعت بالای ورود داده‌ها و زمان محدود: به دلیل سرعت بالای ورود داده‌ها، روش‌های سنتی با پیچیدگی محاسباتی بالا برای داده‌های جریانی

این مقاله در تاریخ ۲۲ بهمن ماه ۱۳۹۹ دریافت و در تاریخ ۱ آذر ماه ۱۴۰۰ بازنگری شد.

جواد حمیدزاده (نویسنده مسئول)، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه سجاد، مشهد، ایران، (email: j_hamidzadeh@sadjad.ac.ir)
 منا مرادی، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران، (email: mmoradi@semnan.ac.ir)

1. Streaming Data

2. Concept Drift
3. Gradual
4. Sudden
5. Incremental
6. Recurrent
7. Abrupt



شکل ۱: دسته‌بندی روش‌های طبقه‌بندی داده‌های جریانی.

گرادیان کاهش می‌ماند پس انتشار خطا است و همگرایی به کمینه^۴ محلی و نیاز به تنظیم پارامترهای متعدد از خصوصیات آنها است [۱۵].

به منظور غلبه بر این مشکلات، روش ماشین یادگیری مفرط^۵ (ELM) [۱۶] ارائه گردید. برخلاف الگوریتم‌های یادگیری مبتنی بر گرادیان کاهش^۶ که نیازمند تنظیم کل پارامترهای شبکه شامل وزن‌ها و بایاس‌های لایه‌های شبکه پیش‌خور با یک لایه پنهان در طول فرایند یادگیری تکراری هستند، ELM وزن‌های اتصالی از لایه ورودی به لایه پنهان و بایاس‌های نورون‌های مخفی را به طور تصادفی تولید می‌نماید. علاوه بر این، وزن‌های اتصالی از لایه پنهان به لایه خروجی به طور تحلیلی با معکوس کردن لایه پنهان و ضرب خروجی آموزشی و با استفاده از روش حداقل مربعات تعیین شده است و در نتیجه، قابلیت تعمیم خوب با هزینه محاسباتی بسیار پایین دارد و از آنجا که پس از تکمیل فرایند آموزش شبکه، به تنظیم پارامتر در شبکه نمی‌پردازد آموزش، سریع‌تر انجام می‌شود. علی‌رغم مزایای گفته‌شده، ELM ممکن است با کاستی‌هایی همراه باشد:

۱) به دلیل تعیین تصادفی پارامترهای نورون‌های لایه پنهان ممکن است بعضی از پارامترهای غیر بهینه تولید شوند که در عملکرد و ثبات شبکه تأثیر خواهند داشت. این مسئله در مدل‌هایی با تعداد زیادی از نورون‌های پنهان ممکن است منجر به پیچیدگی زیاد الگوریتم شود.

۲) از آنجا که این روش برای به حداقل رساندن خطای آموزش از کل داده‌های آموزشی استفاده می‌کند، ممکن است مشکل بیش‌برازش را به همراه داشته باشد.

۳) در کاربردهای برخی که نیازمند به روز رسانی مداوم مدل‌ها است، الگوریتم ELM پس از ورود داده‌های جدید، آنها را به داده‌های قبلی اضافه کرده و کل داده‌ها مجدداً آموزش داده خواهند شد که منجر به صرف هزینه و زمان محاسباتی زیادی خواهد گردید.

تا کنون روش‌های گوناگونی برای بهبود الگوریتم ELM ارائه شده‌اند [۱۷]. مؤلفان در [۱۸] و [۱۹] با افزایش و در [۲۰] و [۲۱] با کاهش تعداد نورون‌های لایه مخفی و با بهره‌گیری از یادگیری افزایشی، سعی در بهبود کارایی ELM داشتند. روش [۲۲] به منظور افزایش قابلیت تعمیم ELM معرفی گردید. در [۲۳] و [۲۴] نیز با استفاده از یادگیری افزایشی، طبقه‌بندی‌کننده‌های داده‌های جریانی دارای رانش مفهوم ارائه شده، از قابلیت تعمیم و سرعت قابل قبولی برخوردار هستند. در [۲۵] روشی با هدف بهینه‌سازی پارامترهای ELM معرفی گردید.

در این مقاله، به منظور طبقه‌بندی داده‌های جریانی در حضور رانش مفهوم، طبقه‌بندی‌کننده مبتنی بر یادگیری ترکیبی با استفاده از مدل

می‌شود که در این حالت، تغییر مفهوم موقتی بوده و داده‌های بعدی را تغییر نخواهد داد.

یکی از مسایل مهم در طراحی طبقه‌بندی‌کننده داده‌های جریانی آن است که نتیجه طبقه‌بندی باید به صورت مداوم به‌روز شده و هر زمان که کاربر مایل به مشاهده نتایج باشد، بتواند این کار را انجام دهد. همچنین الگوریتم طبقه‌بندی داده‌های جریانی، علاوه بر صحت بالا باید مقیاس‌پذیر بوده و به تکامل داده‌ها نیز توجه نماید. از دیدگاه دسترسی به داده، روش‌های طبقه‌بندی داده‌های جریانی به دو دسته تقسیم می‌شوند:

۱) روش‌های مبتنی بر بلوک: در این روش‌ها، داده‌های جریانی در بلوکی با اندازه مشخص وارد شده و طبقه‌بندی و شناسایی رانش مفهوم به ازای هر بلوک انجام می‌شود. عیب این روش‌ها، ناتوانی در تشخیص رانش مفهوم ناگهانی است [۲].

۲) روش‌های برخی: در این روش‌ها، پنجره‌ای که حاوی تعداد N داده اخیر است در نظر گرفته شده و با ورود هر داده، فرایند طبقه‌بندی و شناسایی رانش مفهوم انجام می‌شود. عیب این روش‌ها، ناتوانی در شناسایی رانش‌های تدریجی و بازگشتی است [۳] و [۴].

روش‌های احتمالاتی، مبتنی بر همسایگی، مبتنی بر ساختار درختی، مبتنی بر یادگیری ترکیبی و شبکه‌های عصبی، پنج رویکرد پرکاربرد در طبقه‌بندی داده‌های جریانی هستند (شکل ۱). روش‌های احتمالاتی معمولاً از الگوریتم بیز و روش‌های مبتنی بر همسایگی [۵] از الگوریتم kNN استفاده می‌نمایند. الگوریتم VFDT^۱ و نسخه‌های توسعه‌یافته آن از مشهورترین روش‌های طبقه‌بندی داده‌های جریانی هستند که از ساختار درختی بهره می‌برند [۶]. یادگیری ترکیبی نیز یکی از رویکردهای پرکاربرد جهت بهبود فرایند یادگیری این گونه از داده‌ها است که در آن، ساختار مدل یادگیری به گونه‌ای تنظیم می‌شود که قابلیت تعمیم زیادی دارد، به گونه‌ای که به خطای آزمون کمتری نسبت به مدل‌های ساخته‌شده با یک طبقه‌بندی‌کننده واحد دست یابند [۷] و [۸]. در شرایطی خاص، یادگیری ترکیبی بهتر از یادگیری بر اساس یک مدل واحد عمل می‌نماید [۹]. در [۱۰] تا [۱۳] روش‌های مبتنی بر الگوریتم آدا بوست^۲ [۱۴] برای طبقه‌بندی داده‌های جریانی ارائه گردید. از مهم‌ترین نقاط قوت این روش‌ها، وزن‌دهی به نمونه‌ها و اصلاح وزن نمونه‌های ورودی در هر مرحله و همچنین کاهش خطای طبقه‌بندی است؛ هرچند مهم‌ترین نقطه ضعف این روش‌ها، پیچیدگی زمانی و مشکل بودن تنظیم پارامترها است. همچنین شبکه‌های عصبی پیش‌خور^۳ با یک لایه پنهان با تعداد دلخواه پارامتر و توابع فعال‌سازی دلخواه به طور موفقیت‌آمیزی در طبقه‌بندی داده‌های جریانی کاربرد دارند. اگرچه این نوع از شبکه‌های عصبی از مشکل زمان زیاد فرایند آموزش رنج می‌برند زیرا الگوریتم یادگیری‌شان بر اساس روش‌های

4. Minimum

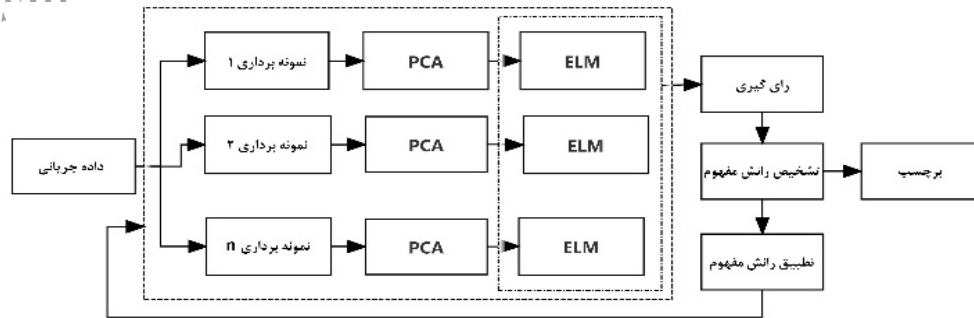
5. Extreme Machine Learning

6. Descent Gradient

1. Very Fast Decision Tree

2. AdaBoost

3. Feed Forward Neural Networks



شکل ۲: روندنمای DELM.

می‌پردازد. با در نظر گرفتن وجود تنها یک لایه مخفی در ELM، بنابراین فرایند استخراج ویژگی نیازمند توجه بیشتری است. بدین منظور، روش [۳۳] با استفاده از مفهوم کرنل و نگاشت به فضای با ابعاد بالاتر، سعی در بهبود استخراج ویژگی در ELM دارد.

بررسی کارهای انجام شده در حوزه طبقه‌بندی داده‌های جریانی مبتنی بر ماشین یادگیری مفرط نشان می‌دهد که در اغلب روش‌ها با ورود داده جدید، آموزش به صورت مجدد انجام شده و راهکار یادگیری تدریجی مغفول مانده یا در صورت استفاده از رویکرد یادگیری تدریجی، به چالش رانش مفهوم توجهی نشده و بنابراین با در نظر گرفتن این مسایل، روش پیشنهادی ارائه گردیده است. همچنین روش پیشنهادی با استفاده از یادگیری ترکیبی و حفظ دانش معتبر کسب شده از داده‌های پیشین می‌تواند پیش‌بینی‌های دقیق‌تری در صورت وقوع رانش مفهوم داشته باشد.

۳- ماشین یادگیری مفرط عمیق

روش پیشنهادی با نام DELM با هدف طبقه‌بندی داده‌های جریانی و بر اساس رویکرد آدابوست و با طبقه‌بندی‌کننده‌های پایه ELM تعریف می‌شود و برای درک بهتر، روندنمای آن در شکل ۲ نشان داده شده است. آدابوست یکی از مهم‌ترین رویکردهای یادگیری ترکیبی است که به منظور افزایش قابلیت تعمیم و دقت طبقه‌بندی‌کننده‌های پایه استفاده می‌شود. آدابوست نسبت به داده‌های نویزی و پرت حساس است، ولی نسبت به مشکل بیش‌برازش به بیشتر الگوریتم‌های یادگیری برتری دارد. در الگوریتم آدابوست در هر دور $t = 1, 2, \dots, T$ ، یک طبقه‌بندی‌کننده ضعیف، اضافه و در هر فراخوانی بر اساس اهمیت داده‌ها، وزن‌ها به‌روز می‌شوند. در هر تکرار، وزن داده‌های غلط طبقه‌بندی شده افزایش می‌یابند و وزن داده‌های درست طبقه‌بندی شده کاهش داده می‌شوند و بنابراین طبقه‌بندی‌کننده جدید بر داده‌هایی که سخت‌تر یاد گرفته می‌شوند، متمرکز خواهد بود [۳۴].

فرض کنید داده‌های جریانی، یک رشته از داده‌های عددی مانند $T_1, T_2, \dots, T_i, \dots$ هستند که در زمان‌های $(x_i \in \mathbb{R}^P)$ $x_1, x_2, \dots, x_i, \dots$ تولید می‌شوند. چنانچه $y_i \in \{0, 1\}$ برچسب داده x_i تولید شده در زمان T_i باشد، دوتایی (x_i, y_i) نشان‌دهنده یک داده برچسب‌دار در زمان T_i است. برای پیش‌بینی برچسب y_{i+1} مربوط به داده x_{i+1} ، روش پیشنهادی از مکانیسم پنجره لغزان با اندازه N استفاده می‌کند؛ بدین معنی که در زمان جاری T_i ، N_i داده برچسب‌دار اخیر در بلوک B_i قابل دسترس هستند و برای آموزش طبقه‌بندی‌کننده مورد استفاده قرار می‌گیرند. در گام بعد به منظور کاهش زمان پردازش، استخراج ویژگی به روش تحلیل مؤلفه‌های اصلی (PCA) انجام می‌شود

بهبود یافته ماشین یادگیری مفرط به نام یادگیری مفرط عمیق (DELIM) ارائه شده است. به دلیل بهره‌مندی از رویکرد یادگیری افزایشی، روش پیشنهادی به نمونه‌های موجود در بلوک داده‌های پیشین بی‌نیاز بوده و در عین حال می‌تواند کارایی‌اش را در صورت عدم اعتبار طبقه‌بندی‌کننده حفظ نماید. این روش با استفاده از داده‌های بلوک زمان جاری می‌تواند با تغییرات رخ داده در داده‌های جریانی تطبیق یافته و وجود داده‌های مرزی و نویزی، تأثیری بر عملکرد آن نخواهد داشت. همچنین روش پیشنهادی به دلیل تلفیق با روش آدابوست، توانسته است بر مشکل بیش‌برازش ELM غلبه نماید.

سازمان‌دهی مقاله به صورت زیر است: کارهای پیشین در بخش دوم آورده شده است. روش پیشنهادی و آزمایش‌های انجام شده در بخش‌های سوم و چهارم بیان گردیده‌اند و همچنین در بخش پنجم، نتیجه‌گیری و کارهای آینده ارائه شده‌اند.

۲- کارهای پیشین

از آنجا که روش پیشنهادی مبتنی بر ماشین یادگیری مفرط است، این بخش به مروری بر تعدادی از روش‌های ارائه شده در این حوزه می‌پردازد. در [۱۲] طبقه‌بندی‌کننده برخط داده‌های جریانی ایستا با استفاده از مفاهیم ترکیبی و معماری ELM ارائه گردید. در روش مذکور، ایجاد مدل بر اساس همبستگی میان طبقه‌بندی‌کننده‌ها و تغییرات خطای آزمایش است. وزن‌دهی اولیه توسط الگوریتم آدابوست و در نهایت بهبود آنها با استفاده از نظریه بازی‌ها انجام می‌شود. مؤلفان در [۲۶] طبقه‌بندی‌کننده برخط ترکیبی مبتنی بر معماری ELM برای مجموعه داده‌های بزرگ ارائه نمودند. روش [۲۷] از رویکرد یادگیری برخط در ELM برای طبقه‌بندی داده‌های جریانی دارای رانش مفهوم استفاده می‌کند. در این روش جهت حفظ قابلیت تعمیم به هنگام وقوع رانش مفهوم، به تعداد گره‌های لایه مخفی افزوده می‌شود. در [۲۸] رویکرد طبقه‌بندی داده‌های جریانی مبتنی بر ELM با کمک استخراج ویژگی بدون نظارت و یادگیری ترکیبی ارائه گردید. در روش [۲۹] به حل مسئله عدم توازن داده‌های جریانی مبتنی بر ELM پرداخته شده است. در [۳۰] طبقه‌بندی‌کننده مقاوم به نویز و رانش مفهوم داده‌های جریانی مبتنی بر ELM با استفاده از تخمین گر M و مکانیسم فراموشی متغیر معرفی گردید. در [۳۱] رویکرد ترکیبی برخط و وزن‌دهی مبتنی بر چگالی جهت تنظیم وزن داده‌های جریانی نامتوازن و بهبود ELM ارائه شده است.

در سال‌های اخیر با افزایش محبوبیت یادگیری عمیق از آن برای بهبود ELM بهره گرفته شده است. مرجع [۳۲] با استفاده از تخصیص اتوانکدر به هر لایه مخفی ELM به حل مسئله طبقه‌بندی داده‌های جریانی

Archive of SID

طبقه‌بندی‌کننده‌ای انتخاب می‌شود که توانایی بیشتری در تشخیص نمونه‌های غلط طبقه‌بندی‌شده را داشته باشد.

با ورود $k+1$ امین بلوک داده، به روز رسانی بلوک داده (رابطه (۹))، ماتریس خروجی لایه پنجم H_{k+1} (رابطه (۱۰)) و ماتریس هدف y_{k+1} (رابطه (۱۱))، به صورت زیر انجام می‌شود

$$B_{k+1} = \{x_i, y_i\}_{i=\sum_{j=1}^{k+1} N_j}^{\sum_{j=1}^{k+1} N_j} \quad (9)$$

$$H_{k+1} = \begin{bmatrix} g(w_1, b_1, x_{\sum_{j=1}^{k+1} N_j}) & \dots & g(w_L, b_L, x_{\sum_{j=1}^{k+1} N_j}) \\ \vdots & \ddots & \vdots \\ g(w_1, b_1, x_{\sum_{j=1}^{k+1} N_j}) & \dots & g(w_L, b_L, x_{\sum_{j=1}^{k+1} N_j}) \end{bmatrix}_{N_{k+1} \times L} \quad (10)$$

$$y_{k+1} = [y_{\sum_{j=1}^{k+1} N_j}^T \dots y_{\sum_{j=1}^{k+1} N_j}^T]_{N_{k+1} \times d}^T \quad (11)$$

که در آن N_j اندازه بلوک داده با برچسب زمانی j است. همچنین ماتریس وزن خروجی $\beta^{(k+1)}$ از (۱۲) و (۱۳) به دست می‌آید

$$\beta^{(k+1)} = \beta^{(k)} + K_{k+1}^{-1} H_{k+1}^T (T_{k+1} - H_{k+1} K_k^{-1}) \quad (12)$$

$$K_{k+1} = K_k + H_{k+1}^T H_{k+1} \quad (13)$$

از آنجا که ماتریس وزن خروجی β به محاسبات پیچیده وارون ماتریس نیازمند است، بنابراین از (۱۴) برای کاهش محاسبات استفاده می‌شود

$$K_{k+1}^{-1} = (K_k + H_{k+1}^T H_{k+1})^{-1} = K_k^{-1} - K_k^{-1} H_{k+1}^T (I + H_{k+1} K_k^{-1} H_{k+1}^T)^{-1} H_{k+1} K_k^{-1} \quad (14)$$

و در ادامه، رویکرد پیشنهادی در مواجهه با تغییرات داده‌های جریان ذکر می‌گردد.

DELIM برای به روز رسانی طبقه‌بندی‌کننده‌ها از مکانیسم یادگیری توالی افزایشی استفاده می‌کند. به عبارتی دیگر این روش، خود راهی برای مقابله با تغییر تدریجی مفهوم پیدا نموده و این اطمینان را می‌دهد که نیازی به آموزش مجدد و مستمر نخواهد داشت. رویکرد روش پیشنهادی در شناسایی تغییرات موجود در داده‌های جریان بدین صورت است: فرض می‌کنیم برای بلوک داده B_i در زمان i ام، میزان خطای طبقه‌بندی‌کننده p_i است که یک متغیر تصادفی بوده و از توزیع برنولی پیروی می‌کند. میزان پراکندگی خطا از (۱۵) محاسبه می‌شود

$$s_i = \sqrt{\frac{p_i(1-p_i)}{N_i}} \quad (15)$$

که N_i تعداد نمونه‌های موجود در بلوک است. با در نظر گرفتن مقادیر s_i و p_i ، تغییر داده‌های جریان به سه حالت پایدار، اختلال و تغییر مفهوم دسته‌بندی می‌شود.

حالت پایدار: اگر $p_i + s_i \leq p_{\min} + 2s_{\min}$ و $p_i < \varepsilon$ باشد، بدین معناست که میزان خطای طبقه‌بندی‌کننده‌ها در سطح پایینی است و مفهوم داده‌های جریان در جایی که ε آستانه باشد ثابت است. بنابراین طبقه‌بندی‌کننده‌ها برای طبقه‌بندی داده‌های جریان فعلی مناسب بوده و نیاز به تطبیق ندارند [۱۲].

حالت اختلال: اگر $p_i + s_i \geq p_{\min} + 2s_{\min}$ و $p_i < \varepsilon$ باشد، بدین معناست که میزان خطای طبقه‌بندی‌کننده‌ها همچنان در سطح پایینی

و بدین ترتیب، تأثیر ویژگی‌هایی که کمترین ارتباط را با طبقه مربوط دارند حذف می‌شود ($x_i \in \mathbb{R}^d$).

اکنون با در نظر گرفتن الگوریتم آداپوست با طبقه‌بندی‌کننده‌های پایه ELM و با فرض تعداد تکرار T ، مقداردهی اولیه وزن داده‌ها به صورت $D_i(i) = 1/N$ که $i = 1, \dots, N$ ، $t = 1, \dots, T$ بوده و همچنین با فرض در اختیار داشتن اولین بلوک داده $B_i = \{(x_i, y_i)\}_{i=1}^N$ با اندازه N ، $B_i = \{(x_i, y_i)\}_{i=1}^N$ وزن w و بایاس b ، از (۱) برای محاسبه ماتریس خروجی لایه پنجم H با تعداد L نورون و ماتریس هدف اولیه y استفاده می‌گردد

$$H_i = \begin{bmatrix} g(w_1, b_1, x_1) & \dots & g(w_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ g(w_1, b_1, x_N) & \dots & g(w_L, b_L, x_N) \end{bmatrix}_{N \times L} \quad (1)$$

$$y_i = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times d} \quad (2)$$

به همین ترتیب، وزن خروجی اولیه $\beta^{(1)}$ به صورت (۳) است

$$\beta^{(1)} = K_1^{-1} H_1^T y_i \quad (3)$$

که در آن $K_i = (1/c) + H_i^T H_i$ و $y_i = [t_1, t_2, \dots, t_N]^T$ است. برای طبقه‌بندی‌کننده‌های ضعیف H ، طبقه‌بندی‌کننده h_i به گونه‌ای انتخاب می‌شود که میزان خطا نسبت به توزیع، کمینه گردد

$$h_i = \arg \min_{h \in H} \sum_{i=1}^N D_i(i) I(y_i \neq h(x_i)) \quad (4)$$

در (۴)، $I(\cdot)$ یک تابع نشان‌گر است که در صورت صحت، مقدار آن برابر یک و در غیر این صورت برابر صفر می‌شود.

خطای h_i از (۵) محاسبه می‌گردد

$$\varepsilon_i = \sum_{i=1}^N D_i(i) I(y_i \neq h(x_i)) \quad (5)$$

اگر $|\varepsilon_i - 0.5| \leq \beta$ و یک آستانه از پیش تعیین شده باشد، نیازی به به روز رسانی وزن داده‌ها نیست. در غیر این صورت به روز رسانی با استفاده از (۶) صورت می‌گیرد

$$D_{t+1}(i) = \frac{D_t(i) \exp(\alpha I(y_i) \neq h_t(x_i))}{Z_t} \quad (6)$$

معمولاً $\alpha_i \in \mathbb{R}$ برای $\alpha_i = 0.5 \ln(1 - \varepsilon_i) / \varepsilon_i$ در نظر گرفته می‌شود و Z_t یک عامل نرمال‌سازی با مقدار $\sum_{i=1}^N D_i(i) \exp(-\alpha_i y_i h_t(x_i))$ است که موجب می‌گردد D_{t+1} یک توزیع احتمالاتی مجاز را نشان دهد.

خروجی نهایی طبقه‌بندی‌کننده‌ها از (۷) محاسبه می‌شود

$$H(x) = \text{sign} \sum_{t=1}^T \alpha_t h_t(x) \quad (7)$$

توجه داشته باشید که توزیع D_t به گونه‌ای به‌روز می‌شود که

$$-\alpha_t y_i h_t(x_i) \begin{cases} < 0 & y_i(i) = h_t(x_i) \\ > 0 & y_i(i) \neq h_t(x_i) \end{cases} \quad (8)$$

بدین ترتیب، به آن دسته از داده‌های x_i که طبقه‌بندی‌کننده h_t آنها را غلط طبقه‌بندی می‌کند وزن بیشتری تخصیص می‌یابد. در روش آداپوست، زمانی که طبقه‌بندی‌کننده‌ها بر اساس توزیع D_{t+1} سنجیده می‌شوند،

Archive of SID

جدول ۱: مجموعه داده‌ها.

مجموعه داده	تعداد نمونه‌ها	تعداد ویژگی‌ها	تعداد طبقه‌ها
Abalone	۴۱۷۷	۸	۳
Segment	۲۳۱۰	۲۳	۲
Credit Card	۳۰۰۰	۱۹	۷
Waveform	۵۰۰۰	۳۶	۶
Website Phishing	۲۴۵۶	۲۱	۳
Satellite Image	۶۴۳۵	۳۶	۶

همبستگی متیوز^۳ (MCC) (رابطه (۲۰)) که به صورت زیر تعریف می‌شوند

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (۱۶)$$

$$TPR = \frac{TP}{TP + FN} \quad (۱۷)$$

$$TNR = \frac{TN}{TN + FP} \quad (۱۸)$$

$$F_1 = \frac{2TP}{2TP + FP + FN} \quad (۱۹)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (۲۰)$$

ضریب همبستگی متیوز بیانگر وابستگی میان متغیر پیش‌بینی شده و مقادیر واقعی آن بوده و مقداری در بازه $[-1, +1]$ دارد. مقدار $+1$ نشان‌دهنده پیش‌بینی فاقد خطای طبقه‌بندی‌کننده، مقدار 0 بیان‌کننده پیش‌بینی تصادفی و مقدار -1 عدم تطابق کامل میان مقادیر پیش‌بینی شده و مقادیر واقعی است. در محاسبه معیارهای مذکور از نمادهای FN ، FP ، TN و TP استفاده شده که این نمادها عبارت هستند از:

FN : تعداد داده‌هایی که به اشتباه منفی تشخیص داده شده‌اند.

FP : تعداد داده‌هایی که به اشتباه مثبت تشخیص داده شده‌اند.

TN : تعداد داده‌هایی که به درستی منفی تشخیص داده شده‌اند.

TP : تعداد داده‌هایی که به درستی مثبت تشخیص داده شده‌اند.

علاوه بر این، کارایی روش پیشنهادی از نظر زمان اجرا سنجیده شده است. منظور از زمان اجرا، فاصله زمانی میان لحظه ورود داده به الگوریتم تا اعلام نتایج شبیه‌سازی توسط الگوریتم است.

۳-۴ نتایج

مقایسه کارایی روش DELM با روش CWEOS-ELM [۱۲] و E-OS-ELM [۲۶] توسط نرم‌افزار متلب ۲۰۱۶ روی سیستمی با پردازنده ۲/۳۰ گیگاهرتز و ۸ گیگابایت RAM و با استفاده از اعتبارسنجی متقابل به روش Leave-one-out انجام شده است. در DELM با استفاده از آزمون و خطا، تعداد طبقه‌بندی‌کننده‌ها $K = 10$ و حد آستانه $\varepsilon = 0.3$ تنظیم شده است. همچنین به مجموعه داده Abalone، ۲۵ نورون در لایه مخفی، به مجموعه داده‌های Segment و Credit هر کدام ۵۰ نورون، به Waveform تعداد ۵۰۰ نورون و به Website phishing و Sattelite image نیز هر کدام ۱۰۰ نورون تخصیص یافته است. تابع فعال‌سازی، سیگموئید انتخاب شده و نتایج آزمایش‌ها در جدول ۲ آمده‌اند.

Requirements: Data stream s , Block B_i , k and d, ε, K classifiers;

while $s \neq \text{null}$ **do**

Get B_i ;

Employ PCA

if $p_i + s_i \leq p_{\min} + \gamma s_{\min}$ and $p_i < \varepsilon$ **then**

The data stream is stable. The ensemble classifier remains unchanged.

if $p_i + s_i \geq p_{\min} + \gamma s_{\min}$ and $p_i < \varepsilon$ **then**

Incremental learning is done by updating the classifiers;

else if $p_i + s_i \geq p_{\min} + \gamma s_{\min}$ and $p_i \geq \varepsilon$ **then**

Concept drift has happened;

Ignore all classifiers and retrain them

End if

Return label

End while

شکل ۳: الگوریتم ۱- شناسایی تغییرات در روش پیشنهادی DELM.

است، اما عملکرد طبقه‌بندی‌کننده‌ها نوسان بزرگی دارد. در این حالت، طبقه‌بندی‌کننده‌ها هشدار می‌دهند و DELM از مکانیسم یادگیری توالی افزایشی برای به روز رسانی هر طبقه‌بندی‌کننده استفاده می‌کند.

حالت تغییر مفهوم: اگر $p_i + s_i \geq p_{\min} + \gamma s_{\min}$ یا $p_i \geq \varepsilon$ باشد، این نشان می‌دهد که تغییر داده‌های جریانی بسیار چشم‌گیر است یا عملکرد طبقه‌بندی‌کننده‌ها در سطح پایینی است. در این شرایط، مدل طبقه‌بندی برای جریان داده فعلی مناسب نبوده است و بنابراین همه طبقه‌بندی‌کننده‌ها، حذف و مجدداً فرایند آموزش انجام می‌گردد. چگونگی مواجهه با تغییرات موجود در داده‌های جریانی توسط روش پیشنهادی در شکل ۳ نشان داده شده است.

۴- آزمایش‌ها

اکنون برای بررسی کارایی روش پیشنهادی، از دو طبقه‌بندی‌کننده داده‌های جریانی مبتنی بر ماشین یادگیری مفرط استفاده شده است. در ادامه، مجموعه داده‌های مورد استفاده، معیار ارزیابی و نتایج حاصل از آزمایش‌ها تشریح گردیده است.

۴-۱ مجموعه داده‌ها

در جدول ۱ مجموعه داده‌های استخراج‌شده از مخزن UCI [۳۵] ذکر تعداد نمونه‌ها، ویژگی‌ها و طبقه‌های آنها مشاهده می‌شوند. مجموعه داده‌ها از کاربردهای مختلف نظیر داده‌های زیستی (Abalone)، داده‌های تصویری (Segment, Sattelite Image)، داده‌های تراکنش مالی (Credit Card)، داده‌های تشخیص نفوذ (Website Phishing) و فیزیک (Waveform) بوده و تنوع زیادی از نظر تعداد نمونه‌ها (۲۳۱۰ تا ۳۰۰۰۰)، تعداد ویژگی‌ها (۸ تا ۳۶) و تعداد طبقه‌ها (۲ تا ۷) دارند. مقادیر ویژگی‌های این مجموعه داده‌ها شامل اعداد حقیقی و صحیح و فاقد مقادیر گم‌شده هستند.

۴-۲ معیارهای ارزیابی

معیارهای مورد استفاده در ارزیابی روش پیشنهادی عبارت هستند از صحت (رابطه (۱۶))، حساسیت^۱ (رابطه (۱۷)) برای بررسی دقت الگوریتم در تشخیص نمونه‌های مثبت، خاص‌بودن^۲ (رابطه (۱۸)) برای بررسی دقت الگوریتم در تشخیص نمونه‌های منفی، F_1 (رابطه (۱۹)) و ضریب

1. Sensitivity

2. Specificity

3. Matthews Correlation Coefficient

معیار	میانگین	وجود اختلاف معنادار
MCC	۰٫۸۸	-
F1	۰٫۸۱	CWEOS-ELM
Specificity	۰٫۹۰	CWEOS-ELM, E-OS-ELM
Sensitivity	۰٫۶۹	CWEOS-ELM, E-OS-ELM
Accuracy	۰٫۸۷	CWEOS-ELM, E-OS-ELM
Time (s)	۶٫۸۰	-

علاوه بر این، همان‌طور که مشاهده می‌شود روش پیشنهادی از نظر زمان اجرا نیز تفاوت معناداری با دو روش رقیب ندارد. بدین معنا که یادگیری افزایشی استفاده‌شده در روش پیشنهادی توانسته است ضمن حفظ نتایج قابل قبول عملکردی، موجب کاهش محاسبات پیچیده و در نتیجه، کم‌شدن زمان اجرا شود.

۵- نتیجه‌گیری و کارهای آینده

در روش پیشنهادی، یک رویکرد ترکیبی یادگیری افزایشی مبتنی بر ماشین یادگیری مفراط برای طبقه‌بندی داده‌های جریان به نام DELM ارائه گردید که توانست در مقایسه با دو روش دیگر مبتنی بر یادگیری مفراط عمیق از نظر معیارهای خاص بودن، حساسیت و صحت برتری یابد. همچنین از نظر معیارهای F1، ضریب همبستگی متیوز و زمان اجرا تفاوت معناداری با روش‌های رقیب مشاهده نگردید. الگوریتم آدابوست به کار گرفته شده در روش پیشنهادی بر مشکل بیش‌برازش ELM غلبه می‌نماید. همچنین استفاده از رویکرد یادگیری افزایشی و به روز رسانی طبقه‌بندی‌کننده‌ها تنها در صورت رخداد رانش مفهوم، می‌تواند از آموزش مجدد طبقه‌بندی‌کننده‌ها به ازای هر تغییر در توزیع داده‌ها توسط ELM جلوگیری نماید. البته هنوز مشکل دیگر ELM یعنی تنظیم مقاداردهی تصادفی پارامترهای نورون‌های لایه پنهان پابرجا بوده و می‌تواند تأثیر منفی بر عملکرد بهینه روش پیشنهادی داشته باشد.

به عنوان کارهای آینده، ارائه راهکارهای مناسب برای حل مسایل مقاداردهی پارامترهای نورون‌های لایه مخفی و پیش‌بینی برچسب داده‌های جریانی در شرایطی که ویژگی‌های جدید به فضای ویژگی آن افزوده می‌شود در نظر گرفته شده‌اند.

مراجع

- [1] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: a review," *IEEE Trans. on Knowledge Data Engineering*, vol. 31, no. 12, pp. 2346-2363, Dec. 2018.
- [2] X. Zheng, P. Li, X. Hu, and K. Yu, "Semi-supervised classification on data streams with recurring concept drift and concept evolution," *Knowledge-Based Systems*, vol. 215, Article ID: 106749, Mar. 2021.
- [3] J. Ko and M. Comuzzi, "Keeping our rivers clean: information-theoretic online anomaly detection for streaming business process events," *Information Systems*, vol. 104, Article ID: 101894, Feb. 2022.
- [4] H. Tavasoli, B. J. Oommen, and A. Yazidi, "On utilizing weak estimators to achieve the online classification of data streams," *Engineering Applications of Artificial Intelligence*, vol. 86, no. C, pp. 11-31, Nov. 2019.
- [5] H. D. Dilectin and R. B. V. Mercy, "Classification and dynamic class detection of real time data for tsunami warning system," in *Proc. Int. Conf. on Recent Advances in Computing and Software Systems*, pp. 124-129, Chennai, India, 25-27 Apr. 2012.
- [6] G. Liu, H. Cheng, Z. Qin, Q. Liu, and C. Liu, "E-CVFDT: an improving CVFDT method for concept drift data stream," in *Proc. Int. Conf. on Communications, Circuits and Systems, ICCAS'13*, vol. 1, pp. 315-318, Chengdu, China, 15-17 Nov. 2013.

جدول ۲: مقایسه معیارهای ارزیابی به درصد (بهترین نتایج، پررنگ شده‌اند).

مجموعه داده	معیار	E-OS-ELM	CWEOS-ELM	DELM
Abalone	MCC	۰٫۶۴	۰٫۸۲	۰٫۶۲
	F1	۰٫۷۶	۰٫۷۸	۰٫۸۴
	Specificity	۰٫۴۸	۰٫۷۸	۰٫۷۸
	Sensitivity	۰٫۵۵	۰٫۵۴	۰٫۵۹
	Accuracy	۰٫۶۲	۰٫۶۳	۰٫۶۳
	Time (s)	۱٫۱۹	۱٫۲۲	۱٫۵۸
Segment	MCC	۰٫۷۸	۰٫۸۲	۰٫۹۸
	F1	۰٫۷۶	۰٫۸۷	۰٫۹۲
	Specificity	۰٫۷۶	۰٫۸۷	۰٫۹۹
	Sensitivity	۰٫۷۹	۰٫۸۲	۰٫۸۹
	Accuracy	۰٫۹۴	۰٫۹۳	۰٫۹۶
	Time (s)	۰٫۱۵	۰٫۱۹	۱٫۵۶
Credit Card	MCC	۰٫۷۵	۰٫۸۵	۰٫۹۵
	F1	۰٫۲۷	۰٫۱۴	۰٫۳۴
	Specificity	۰٫۶۹	۰٫۸۶	۰٫۸۹
	Sensitivity	۰٫۱۱	۰٫۱۴	۰٫۳۰
	Accuracy	۰٫۸۱	۰٫۸۲	۰٫۸۳
	Time (s)	۴٫۸۴	۵٫۵۲	۶٫۹۷
Waveform	MCC	۰٫۶۸	۰٫۸۱	۰٫۸۴
	F1	۰٫۷۳	۰٫۷۶	۰٫۹۴
	Specificity	۰٫۶۰	۰٫۸۳	۰٫۸۶
	Sensitivity	۰٫۶۷	۰٫۷۴	۰٫۸۱
	Accuracy	۰٫۸۸	۰٫۸۹	۰٫۹۰
	Time (s)	۰٫۵۶	۰٫۹۸	۱٫۹۹
Website Phishing	MCC	۰٫۷۹	۰٫۷۴	۰٫۹۴
	F1	۰٫۸۵	۰٫۸۵	۰٫۹۲
	Specificity	۰٫۷۳	۰٫۸۰	۰٫۹۳
	Sensitivity	۰٫۵۳	۰٫۶۵	۰٫۶۶
	Accuracy	۰٫۹۲	۰٫۹۳	۰٫۹۴
	Time (s)	۰٫۴۱	۰٫۶۰	۰٫۶۴
Satellite Image	MCC	۰٫۷۵	۰٫۸۵	۰٫۹۴
	F1	۰٫۹۴	۰٫۸۲	۰٫۹۰
	Specificity	۰٫۷۶	۰٫۸۸	۰٫۹۶
	Sensitivity	۰٫۷۹	۰٫۷۷	۰٫۸۹
	Accuracy	۰٫۹۶	۰٫۹۷	۰٫۹۷
	Time (s)	۱۹٫۱۶	۲۳٫۱۴	۲۸٫۰۳

همچنین برای نشان‌دادن معنی‌دار بودن آزمایش‌ها از آزمون آنالیز واریانس (Turkey's Honestly (HSD)) استفاده گردیده است. نتایج مندرج در جدول ۳ نشان می‌دهند که روش پیشنهادی از نظر معیار ضریب همبستگی متیوز با هیچ کدام از دو روش رقیب، اختلاف معناداری ندارد. از نظر خاص بودن، حساسیت و صحت با هر دو روش اختلاف معنادار داشته اما از نظر معیار F1 تنها با روش CWEOS-ELM اختلاف معنادار دارد. از آنجا که روش مذکور برای داده‌های جریانی ناپیوسته طراحی شده و در برابر رانش مفهوم راه حلی ارائه نداده است، بنابراین نتایج به دست آمده قابل توجیه می‌باشند.

- [24] W. Guo, T. Xu, K. Tang, J. Yu, and S. Chen, "Online sequential extreme learning machine with generalized regularization and adaptive forgetting factor for time-varying system prediction," *Mathematical Problems in Engineering*, vol. 2018, Article ID: 6195387, 31 May 2018.
- [25] J. Xie, et al., "GSPSO-LRF-ELM: grid search and particle swarm optimization-based local receptive field-enabled extreme learning machine for surface defects detection and classification on the magnetic tiles," *Discrete Dynamics in Nature and Society*, vol. 2020, Article ID: 4565769, 15 May 2020.
- [26] Y. Lan, Y. C. Soh, and G. B. Huang, "Ensemble of online sequential extreme learning machine," *Neurocomputing*, vol. 72, no. 13-15, pp. 3391-3395, Aug. 2009.
- [27] S. Xu and J. Wang, "Dynamic extreme learning machine for data stream classification," *Neurocomputing*, vol. 238, pp. 433-449, May 2017.
- [28] O. Aydogdu and M. Ekinci, "A new approach for data stream classification: unsupervised feature representational online sequential extreme learning machine," *Multimedia Tools and Applications*, vol. 79, no. 37, pp. 27205-27227, Oct. 2020.
- [29] W. Li-Wen, G. Wei, and Y. Yi-Cheng, "An online weighted sequential extreme learning machine for class imbalanced data streams," *J. of Physics: Conf. Series*, vol. 1994, no. 1, Article ID: 012008, 10 pp., Chongqing, China, 9-11 Jul. 2021.
- [30] W. Guo, "Robust adaptive online sequential extreme learning machine for predicting nonstationary data streams with outliers," *J. of Algorithms & Computational Technology*, vol. 13, Article ID: 1748302619895421, 18 Dec. 2019.
- [31] Y. Zhang, W. Liu, X. Ren, and Y. Ren, "Dual weighted extreme learning machine for imbalanced data stream classification," *J. of Intelligent & Fuzzy Systems*, vol. 33, no. 2, pp. 1143-1154, 2017.
- [32] B. Mirza, S. Kok, and F. Dong, "Multi-layer Online Sequential Extreme Learning Machine for Image Classification," pp. 39-49, 2016.
- [33] S. Ding, L. Guo, and Y. Hou, "Extreme learning machine with kernel model based on deep learning," *Neural Computing and Applications*, vol. 28, no. 8, pp. 1975-1984, Aug. 2017.
- [34] B. Krawczyk, L. Minku, J. Gama, J. Stefanowski, and M. Wozniak, "Ensemble learning for data stream analysis: a survey," *Information Fusion*, vol. 37, pp. 132-156, Sept. 2017.
- [35] UC Irvine Machine Learning Repository, <https://archive.ics.uci.edu/ml/index.php> (Accessed 03/13, 2020).
- [7] J. Guan, W. Guo, H. Chen, and O. Lou, "An ensemble of classifiers algorithm based on GA for handling concept-drifting data streams," in *Proc. 6th Int. Symp. on Parallel Architectures, Algorithms and Programming*, pp. 282-284, Beijing, China, 13-15 Jul. 2014.
- [8] M. A. M. Raja and S. Swamynathan, "Ensemble learning for network data stream classification using similarity and online genetic algorithm classifiers," in *Proc. Int. Conf. on Advances in Computing, Communications and Informatics, ICACCI'16*, pp. 1601-1607, Jaipur, India, 21-24 Sept. 2016.
- [9] Y. Lv, et al., "A classifier using online bagging ensemble method for big data stream learning," *Tsinghua Science Technology*, vol. 24, no. 4, pp. 379-388, Aug. 2019.
- [10] W. Chen, Q. Sun, J. Wang, J. J. Dong, and C. Xu, "A novel AdaBoost and CNN based for vehicle classification," *IEEE Access*, vol. 6, pp. 60445-60455, 2018.
- [11] H. Zhao, H. Yu, D. Li, T. Mao, and H. Zhu, "Vehicle accident risk prediction based on AdaBoost-SO in vanets," *IEEE Access*, vol. 7, pp. 14549-14557, 2019.
- [12] H. Yu, X. Sun, and J. Wang, "Ensemble OS-ELM based on combination weight for data stream classification," *Applied Intelligence*, vol. 49, no. 6, pp. 2382-2390, 15 Jun. 2019.
- [13] D. Vitorio, E. Souza, and A. L. I. Oliveira, "Using active learning sampling strategies for ensemble generation on opinion mining," in *Proc. 8th Brazilian Conf. on Intelligent Systems, BRACIS'19*, pp. 114-119, Salvador, Brazil, 15-18 Oct. 2019.
- [14] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. of the 13th Int. Conf. on Machine Learning*, pp. 148-156, Bari, Italy, 3-6 Jul. 1996.
- [15] B. L. S. da Silva, F. K. Inaba, E. O. T. Salles, and P. M. Ciarelli, "Outlier robust extreme machine learning for multi-target regression," *Expert Systems with Applications*, vol. 140, Article ID: 112877, Feb. 2020.
- [16] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, Dec. 2006.
- [17] S. Zhang, W. Tan, and Y. Li, "A survey of online sequential extreme learning machine," *Proc. of 5th Int. Conf. on Control, Decision and Information Technologies, CoDIT'18*, pp. 45-50, Thessaloniki, Greece, 10-13 Apr. 2018.
- [18] G. B. Huang, M. B. Li, L. Chen, and C. K. Siew, "Incremental extreme learning machine with fully complex hidden nodes," *Neurocomputing*, vol. 71, no. 4-6, pp. 576-583, Jan. 2008.
- [19] G. Feng, G. Huang, Q. Lin, and R. Gay, "Error minimized extreme learning machine with growth of hidden nodes and incremental learning," *IEEE Trans. on Neural Networks*, vol. 20, no. 8, pp. 1352-1357, Aug. 2009.
- [20] H. J. Rong, Y. S. Ong, A. H. Tan, and Z. Zhu, "A fast pruned-extreme learning machine for classification problem," *Neurocomput.*, vol. 72, no. 1-3, pp. 359-366, Dec. 2008.
- [21] Y. Miche, et al., "OP-ELM: optimally pruned extreme learning machine," *IEEE Trans. on Neural Networks*, vol. 21, no. 1, pp. 158-162, Jan. 2010.
- [22] N. Liu and H. Wang, "Ensemble based extreme learning machine," *IEEE Signal Processing Letters*, vol. 17, no. 8, pp. 754-757, Aug. 2010.
- [23] N. Liang, G. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Trans. on Neural Networks*, vol. 17, no. 6, pp. 1411-1423, Nov. 2006.

جواد حمیدزاده تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد مهندسی کامپیوتر به ترتیب در سال‌های ۱۳۷۴ و ۱۳۷۶ از دانشگاه صنعتی شریف و در مقطع دکتری مهندسی کامپیوتر در سال ۱۳۹۱ از دانشگاه فردوسی مشهد به پایان رسانده است و هم‌اکنون دانشیار دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه سجاد می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: یادگیری ماشین، شناسایی الگو، محاسبات نرم و کاربردهای آن، شبکه‌های کامپیوتری.

منا مرادی تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد مهندسی کامپیوتر به ترتیب در سال‌های ۱۳۸۷ و ۱۳۹۲ از دانشگاه آزاد اسلامی مشهد و دانشگاه آزاد اسلامی فردوس به پایان رسانده است و هم‌اکنون دانشجوی دکتری دانشکده مهندسی برق و کامپیوتر دانشگاه سمنان می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: یادگیری ماشین، شناسایی الگو، محاسبات نرم و یادگیری انتقالی.