

تحلیل مقایسه‌ای الگوریتم‌های کا نزدیک‌ترین همسایه، ژنتیک، ماشین بردار پشتیبان، درخت تصمیم و حافظه کوتاه مدت بلند در یادگیری ماشین

امیرمهدی حسینی و کوروش جمال‌پور

دانشگاه دامغان

چکیده

یادگیری ماشین یک فناوری پر رونق دوران جدید است که کامپیوترها را قادر می‌سازد به طور خودکار از داده‌های پیشین خوانده و تفسیر کنند. این فناوری از الگوریتم‌های متعدد برای ساخت مدل‌ها با طبیعت ریاضی استفاده می‌کند و سپس با استفاده از داده‌های گذشته و دانش، پیش‌بینی‌ها برای داده‌های جدید انجام می‌دهد. اخیراً، این فناوری برای شناسایی متن، تشخیص گفتارهای نفرت‌انگیز، سیستم‌های پیشنهادی، تشخیص چهره و موارد دیگر به کار گرفته شده است. در این مقاله، به طور مفصل بررسی شده‌اند تمامی جنبه‌های مربوط به پنج الگوریتم یادگیری ماشین به نام‌های K-Nearest Neighbor، Algorithm Genetic (KNN)، Machine Vector Support (GA)، Tree Decision (SVM)، (DT) و (Long Short Term Memory) (LSTM) که یک پیش نیاز برای ورود به حوزه یادگیری ماشین است. این مقاله نوری افکنده بر نتایج و استنتاج‌های جدید مرتبط با این الگوریتم‌ها از طریق تحقیق و بررسی مقالات اخیر که تحقیقات کمی و کیفی را در مسائل زمان واقعی، به ویژه تجزیه و تحلیل پیش‌بینی در زمینه‌های چند رشته‌ای انجام داده‌اند. این مقاله همچنین درباره منشأ وضعیتی این الگوریتم‌ها صحبت می‌کند که اگرچه در مقالات قبلی به ندرت مورد بحث قرار گرفته است، اما نکته‌ای برجسته برای علاقه‌مندان و آماتورهای ML است. برای توضیح و درک دقت، قدرت و قابلیت اعتماد الگوریتم‌ها، آن‌ها به طور جامع از نظر کیفی و کمی مورد بررسی و تحقیق قرار گرفته‌اند که در آن شبکه LSTM و الگوریتم SVM رفتار برتری نسبت به سایرین را نشان داده‌اند. این مقاله به تمامی سوالات مربوط که ممکن است در زمان مطالعه این الگوریتم‌ها پیش بیاید، از منشأ آن‌ها تا تعریف، روش‌های اجرا، برنامه‌های زمان واقعی همراه با شواهد نوآورانه کافی، بهره‌ها و معاملات اصلی پاسخ داده است؛ در نهایت یک مقایسه دقیق از عملکرد آن‌ها بر اساس اساس‌های کمی و کیفی ارائه شده است. برای خاتمه، این مقاله همچنین دامنه آینده الگوریتم‌های ML و هوش مصنوعی در زمان‌های آینده و نقش آن‌ها در اتوماسیون و توسعه جامع، نه تنها در جنبه‌های مرتبط با تکنولوژی بلکه همچنین در جنبه‌های انسانی، در نهایت پیروی شده از نتایج قابل اعتماد و مرتبط حاصل از این تحقیق جامع است.

۹

کلمات کلیدی

یادگیری ماشین، k نزدیک‌ترین همسایه، الگوریتم ژنتیک، ماشین بردار پشتیبان، درخت تصمیم، الگوریتم حافظه کوتاه-مدت بلند.

۱ مقدمه

کنند و تصمیم بگیرند، زیرا از مدل‌های ریاضی استفاده می‌کنند که توسط این الگوریتم‌های یادگیری ماشین با کمک داده‌های آموزشی (که مجموعه داده‌های نمونه موجود است) ساخته شده‌اند. برای بیان یک بیانیه مسئله پیچیده، اگر بخواهیم نوعی پیش‌بینی انجام دهیم، الزام نیست کد کامل مسئله را طراحی و نوشته شود، به جای آن فقط با ارائه مطالعات موجود به الگوریتم، می‌توان توسط رایانه یک مدل ریاضی یا منطقی ساخت تا نتیجه را پیش‌بینی کند. به طور

یادگیری ماشین، ترکیبی از مفاهیم آماری و دانش علمی رایانه هاست. این اصطلاح توسط آرتور ساموئل در سال ۱۹۵۹ ابداع شد و اکنون به عنوان زیر مجموعه‌ای از هوش مصنوعی شناخته می‌شود. الگوریتم‌های یادگیری ماشین امکان پردازش و طبقه‌بندی خودکار داده‌های جدید بر اساس اطلاعات قدیمی را برای پردازنده‌ها یا رایانه‌ها فراهم می‌کنند. بدون برنامه نویسی جامع، رایانه‌ها می‌توانند پیش‌بینی

کلی، یادگیری ماشین به سه دسته عمده تقسیم می‌شود: یادگیری بدون نظارت، یادگیری نظارت شده و یادگیری تقویتی.

- یادگیری نظارت شده: در این دسته، ماشین با داده‌های نمونه دارای برچسب ارائه شده برای آموزش، که بر اساس آن بعداً خروجی‌ها را پیش‌بینی می‌کند. نظریه یادگیری نوع نظارت شده بر کلمه 'نظارت' تمرکز دارد، جایی که هدف آن نقشه برداری داده‌های مرتبط با ورودی به داده‌های مرتبط با خروجی است. این روش بدون شک به مقدار قابل توجهی از کاربری انسانی برای ساخت مدل نیاز دارد، اما در نهایت منجر به اجرای سریع‌تر یک کار پیچیده می‌شود. یادگیری ماشین نظارت شده یک دسته گسترده از یادگیری ماشین است. این به طبقه‌بندی بیشتر به الگوریتم‌های رگرسیون و طبقه‌بندی تقسیم می‌شود.

- یادگیری بدون نظارت: یادگیری بدون نظارت امکان می‌دهد تا ماشین بدون هیچ نظارتی یاد بگیرد. در یادگیری بدون نظارت، یک مجموعه داده غیر جدا شده و بدون برچسب به ماشین ارائه شده و الگوریتم باید بر روی داده‌ها بدون هیچ نظارتی عمل کند. این نظریه به هدف دارد عناصر داده ورودی را که الگوهای مشابهی نشان می‌دهند دوباره گروه بندی کند. در این نظریه امکان پیش‌بینی هیچ نتایجی وجود ندارد و ماشین تلاش می‌کند تا بر اساس حجم عظیمی از داده‌ها درک‌های مهمی ارائه دهد. این دوباره به زیرشاخه‌های خوشه‌بندی و انجمن تقسیم می‌شود.

- یادگیری تقویتی: این تئوری به عنوان یک مکانیزم مبتنی بر بازخورد وجود دارد، جایی که فرد یادگیرنده برای هر حرکت صحیح پاداش می‌گیرد و برای عمل نادرست مجازات می‌شود. با این انگیزه‌ها، یادگیرندگان می‌توانند سیستم را تصحیح کرده و عملکرد آن را افزایش دهند. در این نوع یادگیری، فرد اساساً با محیط ترکیب می‌شود و سعی می‌کند بیشتر درباره آن کشف کند. همانطور که قبلاً اشاره شد، دو دسته زیر یادگیری نظارت شده وجود دارد: رگرسیون و طبقه‌بندی. الگوریتم‌های متعلق به زیر دسته رگرسیون مفید هستند زمانی که متغیر ورودی به نحوی با متغیر خروجی مرتبط است و الزام است متغیرهایی از طبیعت پیوسته مانند سهام یا برخی از روندهای جمعیتی پیش‌بینی شود. در حالی که الگوریتم‌های طبقه‌بندی مفید هستند زمانی که نتیجه از نوع زمینه‌ای است مانند "دایره یا مثلث، درست یا غلط، راست یا چپ، بله یا خیر" و غیره.

۲ K نزدیک ترین همسایه

کا-نزدیک‌ترین-همسایه یا K-NN که یکی از الگوریتم‌های حیاتی و موثر در تفکیک داده‌ها است، قادر است تا انتخاب اصلی برای پیاده‌سازی باشد، به‌ویژه زمانی که داده‌های موجود نسبتاً مبهم باشد.

این الگوریتم توسط اوولین فیکس و جوزف هاجز در سال ۱۹۵۱ برای بررسی جداکننده ارایه شد، زمانی که تصمیم‌گیری درباره چگونگی چگالی‌های احتمالاتی با استفاده از تخمین پارامتری نسبتاً چالش برانگیز بود. در سال ۱۹۶۷، چند ویژگی مرتبط با این الگوریتم محاسبه شد؛ به عنوان مثال هنگامی که 'k' برابر ۱ است و 'n' به بی‌نهایت نزدیک می‌شود، محدودیت خطا یا اشتباه طبقه‌بندی K-NN بالاتر از دو برابر نرخ خطای بیز است. پس از بررسی این ویژگی‌ها و خصوصیت‌های خاص، تحقیق و آزمایش از طریق دوره‌های طولانی برای شمارش روش‌های جدید ردیابی، بهبودهایی برای نرخ خطای بیز، روش‌هایی که فقط بر اساس فاصله اعتماد می‌کنند، روش‌های محاسبات نرم و رویکردهای دیگر انجام شد. الگوریتم K-NN در زیر نوعی از روش یادگیری نظارت شده قرار دارد و یکی از آسان‌ترین الگوریتم‌های استفاده شده در یادگیری ماشین است. اگرچه مناسب برای طبقه‌بندی و همزمان‌سازی هر دو استفاده می‌شود، اما اصولاً برای طبقه‌بندی اشیاء استفاده می‌شود. این الگوریتم بسیار کارآمد است و برای اختصاص هر مقدار گم‌شده و باز نمونه‌برداری داده‌ها استفاده می‌شود. برای مجموعه داده‌ها، این الگوریتم پیش‌بینی ارتباط بین داده‌های پنهان و داده‌های موجود را انجام می‌دهد و بر اساس آن پیش‌بینی، داده‌های جدید را به دسته‌بندی موجود نزدیکی بیشتر با آن مطابقت دارند. بنابراین، داده‌های تازه می‌توانند توسط الگوریتم K-NN به‌طور قطعی دسته‌بندی شوند. این الگوریتم نقطه‌ای شکل داده‌های جدید را بر اساس ترتیب داده‌های همسایه آن مرتب می‌کند. K-NN همچنین می‌تواند به عنوان الگوریتم یادگیری تنبلی معرفی شود، زیرا مجموعه داده ابتدایی تنها در ابتدا ذخیره می‌شود، اما فرآیند یادگیری مجموعه داده‌های آموزش تازه در صورت نیاز به طبقه‌بندی یا پیش‌بینی داده‌های جدید انجام نمی‌شود. همچنین این بی‌پارامتری طبیعی است، یعنی در K-NN هیچ روش یا شکل پیش‌تعیین‌شده‌ای برای رابطه بین ورودی و خروجی وجود ندارد. در شکل ۲، دو حالت وجود دارد، تومور خفیف یا بدخیم. یک نقطه داده جداگانه برای مشخص کردن بیشتر یا بدخیم انتخاب شده است. در این حالت، الگوریتم K-NN می‌تواند به آسانی به تحلیل‌گران در روند طبقه‌بندی نقطه جدید مختلف از مجموعه داده کمک کند، بر اساس شباهت یا شاخص مشابهت نقطه با هر دو مورد موجود. الگوریتم K-NN می‌تواند زمانی استفاده شود که مجموعه داده مورد نظر دارای برچسب و بدون نویز باشد.

۱-۲ عملکرد الگوریتم k نزدیک ترین همسایه

حرف 'K' موجود در K-NN به تعداد همسایه‌ها (داده‌هایی که نزدیک‌ترین به نقطه داده جدید هستند) اشاره دارد. تعیین یک مقدار مناسب برای K فرآیند اصلی این الگوریتم است. برای دقت بیشتر، حیاتی است که فرد مقدار صحیح K را انتخاب کند، و این فرآیند به تنظیم پارامتر معروف است. مقدار بسیار پایینی برای K مانند ۱ یا ۲ می‌تواند به نتایج نویزی منجر شود، در حالی که مقدار بسیار بالا در

۲-۲-۲ توضیحات و نتایج

جریان روشی که برای این تحقیق پیروی شده است، در شکل ۵ (a) نشان داده شده است. تقسیم مجموعه داده به دو بخش، نسبت‌های استفاده شده برای هر دو ۵۰:۵۰، ۶۶:۳۴ و ۷۵:۲۵ می‌باشد، جایی که توزیع از داده‌های کلاهداردی به داده‌های غیر-کلاهداردی است. تقسیم‌بندی را می‌توان در شکل ۵ (b) دید. علاوه بیشتر، جداول ۱، ۲، ۳ بیشتر درباره‌ی تقسیم‌بندی مجموعه داده توضیح می‌دهند (برای اطلاعات بیشتر به جدول ۴ مراجعه کنید). از جداول ۵، ۶، ۷ مشاهده می‌شود که الگوریتم رگرسیون لجستیک نتایج دقیق‌تر و قابل اعتمادتری نسبت به الگوریتم‌های نویو بیز و K-NN ارائه داده است. الگوریتم K-NN همانطور که از شکل‌های بالا مشخص است، بدترین عملکرد را از بین تمام الگوریتم‌ها نشان داده است، این به خاطر مجموعه داده آموزشی نمونه‌ای کوچک است، زیرا شباهت بالایی بین داده‌های کلاهداردی و غیر-کلاهداردی وجود دارد و بنابراین الگوریتم قادر به طبقه‌بندی به صورت کارآمد بین این دو دسته نبوده است.

۳-۲ مزایای الگوریتم KNN

الگوریتم K-NN یک الگوریتم آسان برای حل مسائل است. الگوریتم K-NN مقاوم و تحمل‌پذیر نسبت به نویز موجود در مجموعه داده استفاده شده برای آموزش می‌باشد. این الگوریتم سریع، آسان برای تفسیر و موثر است حتی اگر مجموعه داده به اندازه کافی بزرگ باشد.

۴-۲ معایب الگوریتم KNN

تصمیم‌گیری برای انتخاب مقدار مناسب برای K پیچیدگی است، زیرا گاهی نتایج را به شدت تغییر می‌دهد. زیرا نیاز است که فاصله نوع اقلیدسی بین هر نقطه داده‌ای متعلق به مجموعه داده استفاده شده برای آموزش محاسبه شود، منجر به هزینه بالای محاسبه شده می‌شود.

۳ الگوریتم ژنتیک

در دهه ۱۹۵۰، ریاضیدان انگلیسی به نام آلن تورینگ یک دستگاه معرفی کرد که قرار بود نظریه‌ها یا اصول تکاملی را شبیه‌سازد. شبیه‌سازی‌های وابسته به تکامل کامپیوتری در سال ۱۹۵۴ توسط نیلز آل باریسلی آغاز شد، که از دستگاه‌ها و کامپیوترهای موجود در دانشگاه پرینستون در مؤسسه مطالعات پیشرفته استفاده می‌کرد. با این حال، در میان مخاطبان به خوبی شناخته نشد. پس از آن، در سال ۱۹۵۷، متخصص ژنتیک کمی الکس فریزر اهل استرالیا، یک مجموعه مقالات مرتبط با شبیه‌سازی انتخاب مصنوعی ارگانیسم‌ها را کار کرد و منتشر کرد. پس از آن، شبیه‌سازی‌های کامپیوتری مرتبط با تکامل توسط بیولوژیست‌های مختلف در دهه ۱۹۶۰ به وجود آمدند و تکنیک‌ها در متون فریزر و بورنل و کرسی منتشر شد و تمامی جنبه‌های اصلی الگوریتم‌های ژنتیک پوشش داده شد. علاوه بر این، مجموعه‌ای از مقالات منتشر شده توسط هانس-یواخیم

برخی موارد ممکن است ابهام ایجاد کند، بسته به مجموعه داده مقدار ثابتی برای K وجود ندارد، با این حال، یکی از مقادیر استاندارد K که غالباً آن را به خود می‌گیرد عدد '۵' است، یعنی برای فرآیند اکثریت‌گیری، ۵ همسایه نزدیک‌تر به نقطه داده جدید در نظر گرفته می‌شوند. برای جلوگیری از اشتباهات و ابهامات میان دو کلاس مجموعه داده، به طور کلی، مقدار نادری از K مناسب می‌باشد. یک دیگر از محاسبه‌ی مبتنی بر فرمول برای K می‌تواند از این فرمول انجام شود: و n تعداد کل نقاط داده را نمایش می‌دهد. دنبال شده توسط آن، فاصله از نوع اقلیدسی نقاط موجود در مجموعه داده تا نقطه داده جدید محاسبه می‌شود. برای انجام این کار حیاتی است که مجموعه داده به شکل گرافیکی نمایش داده شود. فاصله اقلیدسی به شکلی که در شکل ۳ نشان داده شده است محاسبه می‌شود. پس از محاسبه‌ی ارزش‌های فواصل اقلیدسی تمام نقاط از نقطه داده جدید، باید دقت شود به کدام کلاس اکثر از همسایه‌های نزدیکشان تعلق دارند به عنوان مثال، در $K=5$ و سپس پس از محاسبه دقیق، آن کلاس را به نقطه داده تعلق داده شده برای طبقه‌بندی، الصاق کرد. مثل شکل ۴، می‌توان نتیجه گرفت که نقطه به کلاس A تعلق دارد، زیرا دارای ۳ همسایه نزدیک (اکثریت) از آن دسته است.

۲-۲ مقایسه‌ی الگوریتم‌های یادگیری ماشین رگرسیون لجستیک، بیز ساده و KNN برای تشخیص کلاهداردی کارت اعتباری - برنامه‌ی کاربردی اخیر

۱-۲-۲ زمینه‌ی کار اخیر

کارت‌های اعتباری به دلیل پیشرفت بی‌وقفه فناوری اینترنت امروزه به عنوان یک روش گسترده برای پرداخت‌ها پذیرفته شده‌اند. با این وصف، تقلب‌های بانکی امروزه نیز بیشتر شنیده می‌شوند نسبت به قبل، که به طور دائمی بر بخش‌های مختلفی از جامعه تأثیر گذاشته‌است، از افراد تا مؤسسات. با هر ویژگی امنیتی پیشرفته، فریب‌گران راه‌های جدیدی برای نزدیک شدن به قربانیان پیدا می‌کنند. یکی از نقاط ضعف موجود در اطلاعات کارت اعتباری، انحراف داده است که باعث پیش‌بینی ناکارآمد کلاهداردی‌های آتی می‌شود. این تحقیق انجام شده توسط فیاض ایتو و همکاران (۲۰۲۰) از سه تقارن پایگاه داده برای هدف مطالعه استفاده می‌کند و علاوه بر این، یک روش زیر نمونه‌برداری بر روی پایه تصادفی برای مجموعه داده‌های انحرافی انتخاب شده است. تحقیق آزمایشی انجام شده توسط فیاض ایتو و همکاران (۲۰۲۰) شامل سه الگوریتم، نزدیک‌ترین همسایه، نویو بیز و رگرسیون لجستیک است. معیارهای ارزیابی که توسط آن‌ها برای اندازه‌گیری مورد بررسی قرار گرفته‌اند شامل دقت، خصوصیت، حساسیت، اندازه‌گیری اف، مساحت زیر منحنی و دقت است. نتیجه نهایی تحقیق نشان داده‌است که رگرسیون لجستیک نتایج قابل اطمینان‌تری نسبت به دو الگوریتم دیگر استفاده شده را ارائه داده است.

برمرمن حاوی تنوع گسترده‌ای از راه حل‌ها برای مسائل مربوط به انتخاب، جهش و باز ترکیبی که به بهینه‌سازی وابسته است، بود. همچنین، جنبه‌های مربوط به الگوریتم‌های ژنتیک مدرن نیز توسط برمرمن در کار تحقیقی خود پوشش داده شد. تا دهه ۱۹۷۰، تکنیک تکامل مصنوعی تا آنجا که باید شناخته نشده بود تا زمانی که اینگو ریچنبرگ و هانسپل شوفل پژوهش‌هایشان را در دهه ۱۹۶۰ و ۱۹۷۰ ارائه کردند و ریچنبرگ و گروه‌شان به‌طور صحیح راحل‌هایی برای مواقع پیچیده مهندسی از طریق اصول ژنتیک و تکامل فراهم کرده بودند. روشی جایگزین برای مسائل تکاملی توسط لارنس جی. فوگل ارائه شد، اصولاً برای تولید هوش مصنوعی. در ابتدا، مفهوم برنامه‌ریزی تکاملی از ماشین‌های حالت متناهی برای پیش‌بینی محیطی استفاده می‌شد و تکنیک‌های انتخاب و تغییر برای طراحی پیش‌بینی استفاده می‌شدند. در نهایت، در اوایل دهه ۱۹۷۰، جان هالند بود که قادر بود الگوریتم‌های ژنتیک را از طریق کتابش تطبیق در سیستم‌های طبیعی و مصنوعی منتشر کند که جریان کار او با تحلیل سلولی خود، که به صورت شخصی توسط او و دانشجویانش انجام شد، شروع شد. مطالعات و تحقیقات مربوط به الگوریتم‌های ژنتیک اصولاً تئوری‌ای بودند تا اواسط دهه ۱۹۸۰ که در پیتمبورگ، پنسیلوانیا اولین کنفرانس بین‌المللی در مورد الگوریتم‌های ژنتیک برگزار شد. براساس مفاهیم بیولوژیکی مهم انتخاب طبیعی و وراثت، الگوریتم‌های ژنتیک بنیان‌های، الگوریتم‌های جستجو و بهینه‌سازی هستند. آن‌ها را می‌توان به عنوان یک دسته بازیافت شده از یک دامنه نسبتاً گسترده محاسبه به نام محاسبات تکاملی خواند. الگوریتم ژنتیک عمدتاً یک الگوریتم بهینه‌سازی مبتنی بر احتمال است. مشابه ژنتیک در زیست‌شناسی، در اینجا، چندین راحل که بدست آمده، جهش و باز ترکیبی را تجربه می‌کنند که در نهایت منجر به زاییدن جدید ترکیبی می‌شوند، تازه نوزادان، که توسط تکرار این فرآیند برای چند نسل بعد به دست می‌آیند. هر فرزند از میزان تناسبی تعیین شده برخوردار است که توسط ارزیابی تابع هدف آن اندازه‌گیری می‌شود و در پایان، افراد پرتراکم‌تر احتمال بیشتری برای تولید نسلی با تناسب بیشتر دارند. این تکنیک تضمین می‌کند که موجودیت‌های یا راحل‌های در نسل‌های پی‌رو به صورت صحیح‌تر درست می‌گردند تا نسل نهایی به دست آید. تولید فرزندان بر اساس اصل زیر انجام می‌شود. ۱. کاراکترها یا اجسام برای منابع تلاش می‌کنند و سپس تولید می‌شوند. ۲. کاراکترها با امتیاز تناسب بالاتر نسل می‌کنند تا فرزندان تولید کنند. ۳. بهترین ژن‌ها از کروموزوم‌های پدری به نسل‌های پی‌رو منتقل می‌شوند. ۴. بنابراین، با پیشرفت هر نسل جدید، آنها بهتر و مناسب‌تر برای محیط حاکم می‌شوند.

۱-۳ فضای جستجو

تمام جمعیت در منطقه خاصی که فضای جستجو نامیده می‌شود محدود شده است. هر موجود حاضر در اینجا دارای یک کلید یا راحل برای مسئله داده شده است. هر کروموزوم به عنوان یک بردار

با طول منظوری رمزگذاری می‌شود. پس از انتخاب و ایجاد نسل اولیه، الگوریتم ژنتیک منجر به تکامل گروه می‌شود، با استفاده از فرایندهای انتخاب، گذردهی و جهش.

۲-۳ انتخاب

در این فرایند، اساساً کروموزوم‌هایی که امتیاز تناسب بالاتری دارند، جستجو می‌شوند و اجازه داده می‌شود نسل‌های پی‌روی بهتر و رقابتی‌تر را تولید کنند تا ژن‌های بهتر و جذاب‌تری را منتقل کنند.

۳-۳ رمزگذاری

قالبی که یک کروموزوم به آن دست یافته‌است، دارای داده‌های مربوط به خروجی یا راه حلی است که نمایانگر آنها است. یکی از روش‌های متداول برای رمزگذاری استفاده از یک رشته دودویی است که در شکل ۶ نشان داده شده‌است. هر کروموزوم می‌تواند از این فرمت رمزگذاری شود. هر بیت حاضر در رشته، شامل بخشی از راحل خروجی است.

۴-۳ گذردهی

در گذردهی، دو کروموزوم پدری از طریق فرایند انتخاب، انتخاب شده و نقطه تصادفی برای گذردهی ژن‌ها مشخص می‌شود. بعد از انجام گذردهی، جدیدین نوزاد به وجود می‌آیند.

۵-۳ جهش

برای جلوگیری از همگرایی زودرس جمعیت، ژن‌های تصادفی وارد نوزادهای تازه تولید شده می‌شوند تا انواع موجود در جمعیت تشویق شوند.

۶-۳ تشخیص چهره بر اساس بهینه‌سازی الگوریتم ژنتیک

- برنامه کاربردی اخیر

۱-۶-۳ پس‌زمینه کار اخیر

موراد موسی و همکاران (۲۰۱۸)، بر روی روش تشخیص چهره بر اساس تحلیل مولفه اصلی معروف و روش تبدیل کوزین متمرکز انجام دادند. برای توسعه یک عملیات تشخیص چهره پایدار و قابل اطمینان، لازم است ابتدا به انتخاب ویژگی‌ها توجه کرد، که مسئول لغو سر و صداها غیر ضروری، داده‌های اضافی و ویژگی‌های متعدد دیگر نامربوط هستند. با این حال، توسعه الگوریتم ژنتیک، که الگوریتمی نسبتاً جدیدتر برای انتخاب ویژگی‌ها است، می‌تواند برای رفع این مسئله استفاده شود. برای استفاده از الگوریتم ژنتیک به منظور حل یک مسئله، لازم است راحل‌های موثر را در زنجیره‌های بیت قابل اندازه‌گیری کد کرد تا شامل کروموزوم‌های آمده از نقاط خاص شوند. هدف نهایی استفاده از اپراتورهای ژنتیک و توسعه تمییز معقولی میان کروموزوم‌ها است. موراد موسی و همکاران (۲۰۱۸)، یک سیستم تشخیص چهره را با استفاده از الگوریتم ژنتیک در کنار ترکیبی از تحلیل M تبدیل تبدیل کوزین متمرکز - تحلیل مولفه اصلی (DCT-PCA) طراحی کردند که برای کاهش بعد و انتخاب

ویژگی بر روی یک مجموعه تصاویر چهره انسان به کار برده شد. نتایج ارائه شده توسط مورد موسی و همکاران (۲۰۱۸)، کارایی این روش را نسبت به کارهای قبلی نشان می‌دهد.

۲-۶-۳ شرح و نتایج

مورد موسی و همکاران (۲۰۱۸)، از سه طرح استاندارد به نام موسسه علم و فناوری دانشگاه منچستر، (UMIST) آزمایشگاه تحقیقاتی اولیوتی (ORL) و بیل که در جدول ۸ زیر ارائه شده است، به همراه چهره‌های نماینده (شکل ۹) که برای تست استفاده شدند، استفاده کردند. پایگاه داده‌ها به طور تصادفی برای آموزش یا برای مجموعه‌های تست استفاده شدند و تمام ترتیبات مصلحی برای پژوهش استفاده شدند. نتایج میانگین و مشاهدات ارائه شده است. برای این پژوهش آزمایشی از نسخه R2015a MATLAB استفاده شد و ژن‌های قفل شده به ۳۰ گرفته شد. متغیرهای مختلف دیگری که برای پژوهش مورد نظر بررسی شده‌اند، در جدول ۹ نشان داده شده‌اند. نتایج ارائه شده توسط مورد موسی و همکاران (۲۰۱۸) در جدول ۱۰ ارائه شده‌اند و رویکرد تعقیب شده توسط آنها کمک کرده است تا این سیستم شناسایی چهره به نرخ تشخیص ۹۹٪ برسد. آشکار است که این روش نوین منجر به بهبود ۸٪ نسبت به کارهای قبلی شده است. بنابراین، این رویکرد مبتنی بر الگوریتم ژنتیک باعث موفقیت در بهبود کارایی و سرعت این سیستم تشخیص چهره شده و در انتخاب مناسب ضرایب مورد نیاز کمک کرده است.

۷-۳ مزایای الگوریتم ژنتیک

الگوریتم ژنتیک عملکرد بسیار پایداری در مقایسه با خروجی‌های محلی بیشینه یا کمینه فراهم می‌کند. آنها ارتقاء داده‌های بزرگ فضای حالت را فراهم می‌کنند. در مقایسه با سیستم‌های هوش مصنوعی سنتی، آنها نسبت به ورودی‌ها، ورودی‌های متغیر و سر و صدا ضعیف نمی‌شوند. الگوریتم‌های ژنتیک توابع متمایز و ناپایدار را بهبود می‌بخشند. این الگوریتم به داده‌ها یا اطلاعات تقلیدی نیاز ندارد. از نظر گسترده‌تر و بهینه‌تر، چند برابر با روش‌های ابتدایی است.

۸-۳ معایب الگوریتم ژنتیک

یکی از معایب محتمل الگوریتم ژنتیک این است که اغلب می‌تواند منجر به همگرایی زودرس جمعیت شود، به دلیل یکنواختی ژن‌ها. این امر هر گونه تحقیق مفیدی را باز می‌دارد. با اینکه این الگوریتم به اندازه زیادی به اطلاعات در مورد بیان مسئله نیاز ندارد، اما طراحی یک تابع هدف و دستیابی به عملیات چالشی است. اعمال الگوریتم ژنتیک زمان‌بر است.

۴ ماشین بردار پشتیبان

ماشین‌های بردار پشتیبان (SVM) توسط الکسی چرفوننکس و ولادیمیر ان. واپنیک در سال ۱۹۶۳ ایجاد شدند. از زمان

کشف ماشین‌های بردار پشتیبان، این تکنیک گسترده‌ترین استفاده را در مسائل جداسازی و طبقه‌بندی تصویر، هابیرمتن و متن پیدا کرده است. این الگوریتم‌ها بسیار پیشرفته هستند و می‌توانند برای تشخیص متن دست‌نوشته و همچنین برای مرتب‌سازی پروتئین‌ها در آزمایشگاه‌های زیست‌شناسی استفاده شوند. آنها در بسیاری از حوزه‌های دیگری مانند خودروهای خودران، چت‌بات‌ها، تشخیص چهره و غیره استفاده می‌شوند. از جمله یکی از الگوریتم‌های یادگیری نوع نظارت شده بسیار رایج، الگوریتم ماشین بردار پشتیبان برای مسائل رگرسیون و طبقه‌بندی طراحی شده است. الگوریتم SVM به سرخط مناسب تصمیم‌گیری یا مرز، معروف به هابیرپلاین، که فضای بعدی- n را به کلاس‌های گوناگون تقسیم می‌کند با هدف قرار دادن نقطه‌های مختلف در رده مناسب، هدفمند است. در الگوریتم SVM، نقاط بردار استوانه‌ای بسیاری که Support Vectors نام دارد، انتخاب می‌شوند که در ایجاد یک هابیرپلاین مناسب کمک می‌کند. الگوریتم SVM کاربردهای خود را در تشخیص چهره‌ها، طبقه‌بندی برخی از تصاویر، دسته‌بندی متون و غیره دارد. یک مثال در شکل ۱۱ نشان داده شده است که در آن دو دسته تضادی با استفاده از الگوریتم SVM طبقه‌بندی شده‌اند. به عنوان مثال، یک شخص با تصویر یک گربه غریب با برخی ویژگی‌های مشابه به سگ مواجه می‌شود. بنابراین، برای ایجاد یک مدل که به دقت تشخیص دهد آیا یک سگ است یا یک گربه، الگوریتم SVM بسیار مفید است.

۱-۴ انواع ماشین بردار پشتیبان

۱-۱-۴ نوع خطی ماشین‌های بردار پشتیبان

الگوریتم SVM نوع خطی مفید است در مواردی که داده‌ها باید به صورت خطی جدا شوند، که به این معنا است که مجموعه داده می‌تواند به دو کلاس تقسیم شود که توسط یک خط مستقیم جدا شوند.

۲-۱-۴ نوع غیرخطی ماشین‌های بردار پشتیبان

الگوریتم SVM نوع غیرخطی مفید است در مواردی که داده‌ها به صورت غیرخطی جدا شوند، که به این معنا است که مجموعه داده نمی‌تواند با استفاده از یک خط مستقیم به کلاس‌ها تقسیم شود.

۲-۴ هابیرپلاین و بردارهای پشتیبان در الگوریتم SVM

۱-۲-۴ هابیرپلاین

هابیرپلاین الگوریتم SVM به عنوان بهترین مرز تصمیم‌گیری از بین مرزهای تصمیم ممکن تعریف می‌شود که به روش دقیقی کلاس‌ها را در فضای بعدی- n دسته‌بندی می‌کند. ویژگی‌های مجموعه داده ابعاد هابیرپلاین را تعیین می‌کند، به این معنا که اگر مجموعه داده ۲ ویژگی داشته باشد، به این معناست که هابیرپلاین یک بعدی و اگر ۳ ویژگی داشته باشد، به این معناست که هابیرپلاین دو بعدی است. هابیرپلاینی که حاشیه‌های بیشینه دارد، به این معنا که فاصله بین دو نقطه داده بیشینه باشد، اولویت دارد.

۲-۲-۴ بردارهای پشتیبان

بردارهای پشتیبان نزدیکترین نقاط داده‌ای هستند که موقعیت هایپرپلین را تحت تأثیر قرار می‌دهند. به دلیل حمایت آنها از هایپرپلین، آنها به عنوان بردارهای پشتیبان شناخته می‌شوند.

۳-۴ عملکرد الگوریتم SVM

۱-۳-۴ SVM خطی

مدل کارایی الگوریتم SVM می‌تواند با استفاده از یک مثال توضیح داده شود. فرض کنید یک مجموعه داده دارای دو شیء مختلف (قرمز و زرد) و دو ویژگی، مانند X_1 و X_2 است. یک الگوریتم طبقه‌بندی‌ای که بتواند جفت مختصات X_1 و X_2 را به درستی در یکی از دو رنگ قرمز یا زرد جدا کند، مطلوب است. از آنجایی که این یک فضای ۲ بعدی است و دارای دو ویژگی است، بنابراین راحت‌تر است که این دو دسته را فقط با یک خط مستقیم تفکیک کرد، اما بسیاری از خطوط مستقیم امکان‌پذیر هستند. نقش این الگوریتم در اینجا معرفی می‌گردد که خط تصمیم مناسبترین را از بین تمام خط‌های یا مرزهای تصمیم انتخاب می‌کند؛ این مرز تصمیم بهترین مرز تصمیم نام دارد. الگوریتم SVM بردارهای پشتیبان یعنی نزدیکترین نقاط از مرز تصمیم در دو کلاس را مشخص می‌کند. الگوریتم SVM فاصله بین هایپرپلین و بردارها را بیشینه می‌کند تا یک راحل بهینه را تأمین کند.

۲-۳-۴ SVM غیرخطی

در نظر داشته باشید که داده‌ها به صورت غیرخطی ترتیب داده شده‌اند. در اینجا نمی‌توان به سادگی یک خط مستقیم رسم کرد. بنابراین، برای جدا کردن این نقاط داده، یک بعد دیگر نیاز است. برای داده‌های به صورت خطی، تنها دو بعد استفاده شده است یعنی X و Y اما برای داده‌های به صورت غیرخطی، یک بعد سوم اضافه می‌شود. از آنجایی که کار در یک فضای ۳ بعدی مشغول است، این در واقع یک صفحه است که موازی با محور Z است. با تبدیل آن به یک فضای ۲ بعدی $Z = 1$ ، یک دایره با شعاع ۱ واحد به دست می‌آید که در شکل ۱۳ (شکل ۱۲ را ببینید) نشان داده شده است.

۴-۴ استفاده از الگوریتم SVM برای تشخیص سرطان

پستان - کاربرد اخیر

۱-۴-۴ پیش‌زمینه کار اخیر

جنی ای. ام. سایدی-گیبونز و همکاران (۲۰۱۹) تحقیقات دقیقی در مورد الگوریتم‌های خاص یادگیری ماشین که می‌توانند برای پیش‌بینی سرطان، به ویژه سرطان پستان، استفاده شوند، انجام دادند. الگوریتم‌های یادگیری ماشین بسیار کارآمد هستند و می‌توانند در علوم پزشکی برای تشخیص زودرس یا پیشداوری بسیاری از بیماری‌های فاتال مفید باشند. در تحقیقات آزمایشی آنها، طرح‌های پیش‌بینی مبتنی بر الگوریتم‌های مختلف برای تشخیص سرطان بر اساس موادی که از جرم پستان استخراج شده، انجام شد. الگوریتم‌های استفاده شده

در کار تحقیقاتی آنها شامل شبکه‌های عصبی مصنوعی تک‌لایه، الگوریتم ماشین بردار پشتیبان با هسته تابع پایه گرد، و مدل تخطی عمومی (GLM) بود. تقریباً ۴۵۶ نمونه از جرم‌های پستان برای ارزیابی و ۲۲۷ نمونه برای اعتبارسنجی استفاده شدند. قبل از آزمایش الگوریتم‌ها و مدل‌ها در مجموعه اعتبارسنجی برای تشخیص بیماری، آنها با استفاده از نمونه‌های ارزیابی آموزش دیده شدند. به منظور مقایسه عملکردهای مدل‌های الگوریتمی موردنظر، معیارهای ارزیابی که توسط جنی ای. ام. سایدی-گیبونز و همکاران (۲۰۱۹) استفاده شد، حساسیت، خصوصیت و دقت بودند. پس از تحقیقات انجام شده توسط آنها، مشخص شد که الگوریتم SVM بیشینه مساحت زیر منحنی و دقت را نسبت به دو الگوریتم دیگر فراهم کرد.

۲-۴-۴ توضیحات و نتایج

مجموعه داده‌ای که برای تحقیقات آنها استفاده شد، مجموعه داده سرطان پستان ویسکانسین بود که از طریق مخزن یادگیری ماشین دانشگاه کالیفرنیا آبرون که به صورت رایگان در دسترس است، قابل دسترسی است. ویژگی‌های جرم‌های پستان که در این تحقیق به آن‌ها تمرکز داده شده است، هسته‌های سلولی هستند که تحت انجام سوزن-نمونه‌برداری دقیق (FNA) یک روش تشخیصی پزشکی استاندارد در زمینه انکولوژی تجربی شدند. یک شکل نمونه از جرم پستان در شکل ۱۴ نشان داده شده است. تعداد نمونه‌های استفاده شده برای این تحقیق در واقع تعداد نمونه‌هاست که به یک شناسه یکتا (ID) اختصاص یافته‌اند و به جز آن، ویژگی‌های مختلف دیگر متصل شده‌اند. ستون کلاس نمایش داده شده در شکل حاصل، تشخیص است که بیماری سرطانی بدخیم یا مہلک است که بستگی به نمونه‌برداری ضخیم سوزن (FNA) که سرطانی بودن آن یا خیر بود، دارد. همان‌طور که می‌توان در جدول ۱۱ دید، ۲۴۱ نمونه به خوب‌خویم و ۴۵۸ نمونه هم‌پیوندیم بوده‌اند. نمونه‌های هم‌پیوندیم دارای کلاس دو، در حالی که نمونه‌های خوب‌خیم دارای کلاس چهار هستند. در این تجربه نه ویژگی برای آزمون وجود دارد که در جدول ۱۱ نشان داده شده است، که هر ویژگی بر اساس یک مقیاس از ۱ تا ۱۰ ارزیابی شده است. هرچه مقدار به ۱۰ نزدیکتر باشد، ویژگی طبیعتاً بیشتری بدخیم است و هرچه مقدار به ۱ نزدیکتر باشد، ویژگی طبیعتاً بیشتری مہلک است. با وجود اینکه تمام الگوریتم‌ها یک سبک کاری بسیار متنوع دارند، آنها سطحی مناسب از دقت، حساسیت و خصوصیت را در عملکرد خود نشان داده‌اند و الگوریتم SVM بهترین عملکرد را با دقت ۹۶.۰ در جدول ۱۲ نشان داده شده است.

۵-۴ مزایای الگوریتم SVM

الگوریتم SVM بیشترین تطابق را در مواردی ارائه می‌دهد که تقسیم واضحی بین کلاس‌ها وجود دارد. SVM کارایی بیشتری در فضاهای با ابعاد بالاتر نشان می‌دهد. SVM نیز در مواقعی که تعداد فضاهای بعدی از مقدار نمونه‌های موجود بیشتر باشد به طور

مؤثر عمل می‌کند. SVM از نظر حافظه نسبت به سایر ویژگی‌ها کاربردی واقعی است.

۶-۴ معایب الگوریتم SVM

الگوریتم SVM قادر به کارایی مناسب برای مجموعه داده‌های بزرگ نمی‌باشد. SVM برای مواردی که مجموعه داده‌ها دارای حجم بزرگی از سر و صدا هستند، به عملکرد مطلوبی نمی‌رسد که در عمل خیلی اغلب اتفاق می‌افتد. SVM تحت عملکرد ضعیفی دارد در مواردی که مقدار عددی ویژگی‌های هر نقطه داده بیشتر از نمونه‌های داده‌ی آموزشی باشد. الگوریتم SVM برای طبقه‌بندی انجام شده توسط آن توجیه احتمالی ندارد.

۵ درخت تصمیم

الگوریتم درخت تصمیم (DT) که به دسته الگوریتم‌های یادگیری نظارت‌شده تعلق دارد، بیشتر برای حل مسائل طبقه‌بندی ترجیح داده می‌شود، اما به هر دو طریق می‌توان در حالت طبقه‌بندی و در حالت پیش‌بینی استفاده کرد. این الگوریتم شامل گره‌های داخلی که ساختار شاخه‌ها را نشان می‌دهند، مجموعه داده که نتیجه‌ی ارائه شده توسط الگوریتم را نشان می‌دهد، و هر گره برگ که یک نتیجه را نمایندگی می‌کند، می‌باشد. دو گره وجود دارد: اول گره تصمیم که برای تصمیم‌گیری استفاده می‌شود و شاخه‌های مختلفی دارد؛ و دوم گره برگ که خروجی گره‌های تصمیم است و دیگر شاخه‌هایی ندارد. این الگوریتم نام خود را به دلیل شباهتی که به دلیل یک درخت دارد بدست آورده است. گره ریشه نقطه شروعی است که به شاخه‌های مختلفی توسعه می‌یابد و سازه‌ای شبیه به درخت را شکل می‌دهند. درخت تصمیم به اختصار درخت را بر اساس پاسخ به سوال تقسیم می‌کند، به سوالی از نوع بله یا خیر.

۱-۵ طبقه‌بندی درخت‌های تصمیم

۱-۱-۵ درخت تصمیم دارای متغیر خوشه‌ای

درخت تصمیم دارای متغیر خوشه‌ای به عنوان هدف. مثال: - جمله مشکل با داشتن متغیر هدف به عنوان "آیا با پرتاب سکه، شیر ظاهر می‌شود یا خیر" (مشاهده شکل ۱۵).

۲-۱-۵ درخت تصمیم دارای متغیر ثابت

درخت تصمیم دارای متغیر ثابت به عنوان هدف. مثال: - آیا شخص می‌تواند یک وام را پس دهد یا نه. در صورتی که بانک‌ها اطلاعات درآمد را نداشته باشند، که یک متغیر مهم در این مورد است، آنگاه می‌توان یک درخت تصمیم برای پیش‌بینی درآمد ماهیانه یک فرد بر اساس عوامل مختلفی مانند دارایی‌ها، استاندارد زندگی، شغل و غیره برپا کرد. در اینجا مقادیری که پیش‌بینی می‌شوند برای متغیرهای پیوسته نوعی است.

۲-۵ اصطلاحات درخت تصمیم

گره ریشه: بخش ابتدایی از درخت تصمیم که از آن شروع به تقسیم کل داده می‌شود وارد مجموعه‌های مختلف ممکن می‌شود که همگن هستند.

گره برگ: گره نهایی به عقب که دیگر تقسیم درختی امکان‌پذیر نیست.

تقسیم: شامل فرایند تقسیم گره اصلی به زیر گره‌ها بر اساس محدودیت‌های ارائه شده می‌شود.

زیر درخت: تقسیم یک سلسله مراتب به یک زیر درخت یا شاخه منجر می‌شود.

قالب‌بندی: شامل حذف شاخه‌های بیش از حد از درخت تصمیم به منظور به دست آوردن نتایج بهینه است. در واقع اندازه درخت را بدون تاثیر بر دقت کاهش می‌دهد. این از دو نوع، قابلیت هزینه و قبول خطا تقسیم بندی است.

گره والدین و کودک: این گره پایه نامیده می‌شود که همچنین گره والدین نیز نامیده می‌شود، درحالی که گره‌های باقی‌مانده به‌سادگی گره‌های کودک نامیده می‌شوند.

۳-۵ اندازه‌گیری انتخاب ویژگی‌ها

اندازه‌گیری انتخاب ویژگی (ASM) شامل جمع‌آوری ویژگی بهینه مربوط به گره منبع و همچنین زیرگره‌ها است. دو عملکرد اصلی برای ASM عبارت‌اند از:

۱-۳-۵ بهره اطلاعات

بهره اطلاعات همانطور که از نام پیداست، میزان اطلاعاتی که توسط یک ویژگی در مورد کلاس ارائه می‌شود را محاسبه می‌کند. گره تقسیم می‌شود و درخت بر اساس ارزش‌های بهره اطلاعاتی ساخته می‌شود.

۲-۳-۵ شاخص Gini

شاخص Gini میزان خلوص یا اصالت را که در ایجاد یک الگوریتم درخت تصمیم استفاده می‌شود اندازه‌گیری می‌کند. ویژگی‌های کوچک شاخص Gini بهتر از ویژگی‌هایی با شاخص Gini بزرگتر توسط الگوریتم درخت تصمیم در هنگام تصمیم‌گیری ترجیح داده می‌شوند.

۴-۵ مراحل ساخت درخت تصمیم

گره ریشه، به نام X که شامل کل مجموعه داده است، به عنوان نقطه شروع درخت در نظر گرفته می‌شود. با استفاده از ASM، بهترین ویژگی مطابق از مجموعه داده را جستجو نمایید. X را به زیربخش‌هایی با مقادیر با کیفیت بهتر تقسیم می‌کند. فقط با استفاده از ویژگی ایده‌آل گره‌های درخت تصمیم را توسعه دهید. به تکرار گره‌های منحصر به فرد درخت تصمیم، با استفاده از زیرمجموعه‌های موجود مجموعه داده ایجاد شده در X ، توسعه دهید. این فرایند

را ادامه دهید تا به نقطه‌ای برسید که دیگر امکان نداشته باشید به زیربخش‌ها برسید. این گره نتیجه‌ی نهایی نهایی به عنوان گره برگ شناخته می‌شود.

۵-۵ پیش‌بینی بیماری کبد با استفاده از تکنیک‌های مختلف درخت تصمیم-برنامه نویسی اخیر

۱-۵-۵ زمینه کار اخیر

بیماری‌های مرتبط با کبد یکی از بیماری‌های مهلکی است که می‌تواند بر جان انسان‌ها تأثیر گذار باشد. کشف هر گونه تکنولوژی که بتواند این‌گونه بیماری‌ها را در مراحل ابتدایی پیش‌بینی کند، برای نجات جان انسان‌ها بسیار مفید است. نازمون نهار و همکاران (۲۰۱۸)، این تحقیق را در این زمینه انجام داده‌اند با اختلاف و مقایسه انواع مختلفی از الگوریتم‌های درخت تصمیم برای کمک به پیش‌بینی بیماری کبد در مراحل ابتدایی. الگوریتم‌های درخت تصمیم در بسیاری از زمینه‌ها به ویژه در حوزه علوم پزشکی مورد استفاده قرار می‌گیرند. نازمون نهار و همکاران (۲۰۱۸)، از مجموعه داده‌ای که شامل ویژگی‌هایی از قبیل بیل‌روبین مستقیم، بیل‌روبین کل، جنسیت، عامل سن، پروتئین‌های کلی و غیره استفاده کردند. تکنیک‌های درخت تصمیمی که در این تحقیق آزمایشی مورد آزمایش قرار گرفتند، شامل درخت مدل لجستیک (LMT)، جاوا (J۴۸)، درخت تصادفی، درخت تصمیم دهانه، جنگل تصادفی، درخت خطای کم کننده و درخت بازتراشی خطا (REPTree) بودند. مطالعه تجربی آنها نشان داد که تصمیم دهانه نتایج مطمئن و دقیق‌تری ارائه کرده است.

۲-۵-۵ توضیحات و نتایج

هدف اصلی این تحقیق کشف این است که آیا یک بیمار تحت تأثیر بیماری‌های مرتبط با کبد قرار دارد یا خیر، با استفاده از انواع مختلف الگوریتم‌های درخت تصمیم است. تکنیک‌های مختلف براساس معیارهای ارزیابی مختلفی نظیر دقت، خطا مطلق میانگین، آماره کاپا، زمان اجرا، دقت، بازخاست و غیره آزمایش و مقایسه شدند. نازمون نهار و همکاران (۲۰۱۸)، از ابزار استخراج داده قدرتمندی به نام ویکا استفاده کردند تا دقت انواع الگوریتم‌ها را با استفاده از آنها روی مجموعه داده‌های مختلف آزمایش نمایند.

۶-۵ مزایای الگوریتم درخت تصمیم

پیچیدگی بسیار پایین - این الگوریتم بسیار ساده قابل فهم است و نیازی به دانش ویژه مرتبط با آمار برای تفسیر آن ندارد. مفید برای کاوش داده - همچنین می‌تواند در مراحل کاوش داده مورد استفاده قرار گیرد زیرا الگوریتم درخت تصمیم یکی از سریعترین الگوریتم‌ها در ایجاد یا شناسایی ویژگی‌های جدید است. کمتر نیاز به تمیزکاری داده - به طور مقایسه‌ای نیاز کمتری به مراحل تمیزکاری داده دارد و تحت تأثیر مقادیر و داده‌های گم‌شده نیست. بدون محدودیت نوع داده - این قادر است به طور انعطاف‌پذیر با متغیرهای عددی و همچنین

متغیرهایی با طبیعت خوشه‌ای کنار بیاید. روش غیر پارامتریک - درخت تصمیم از یک روش غیر پارامتریک استفاده می‌کند، که به معنی نبود هیچ گونه فرضیه‌ای درباره توزیع فضایی است.

۷-۵ معایب الگوریتم درخت تصمیم

بیش‌آموزش - بیش‌آموزش یکی از مسائل عملی اصلی بر روی مدل درخت تصمیم است. با این حال، با تنظیم پرونده و محدودیت پارامترهای مدل، مشکلات بیش‌آموزش می‌تواند کاهش یابد. نامناسب برای متغیرهای پیوسته - درخت تصمیم برخی از اطلاعات ارزشمند را از دست می‌دهد در هنگام دسته‌بندی متغیرها در دسته‌های مختلف.

۶ الگوریتم حافظه کوتاه-مدت بلند (LSTM)

به دلیل پس‌انتشار با یادگیری مداوم واقعی یا زمانی، سیگنال‌های حاوی خطا که به سمت عقب در زمان حرکت می‌کنند، ممکن است ناپدید شوند یا بزرگ شوند؛ جابجایی‌های زمانی سیگنال حاوی خطا به طرز قابل توجهی به اندازه وزن‌ها بستگی دارد. در صورت بزرگ شدن، وزن‌ها به احتمال زیاد شروع به نوسان کردن می‌کنند و در صورت ناپدید شدن، یا زمان مصرف شده برای یادگیری اتصالات با تأخیرهای زمانی بلند از حد بیرون می‌رود، یا در بدترین حالت به دلایلی کار نمی‌کند. به عنوان درمان، الگوریتم حافظه کوتاه-مدت بلند، (LSTM) نوعی جدید از شبکه‌های عصبی مکرر در سال ۱۹۹۱ به وجود آمد که توسط Hochreiter Sepp و Schmidhuber Jurgen توسعه یافت تا سیستم‌های موجود را پیش‌رویی کند و مشکلات پس‌انتشار خطا مورد بحث فوق را برطرف کند. نسخه اصلی این الگوریتم حافظه کوتاه-مدت بلند فقط شامل سلول‌ها، دروازه‌های ورودی و خروجی بود. این الگوریتم قادر است تا در شکاف‌های زمانی بیش از گام‌ها پل روبرو بیاند، حتی زمانی که دنباله‌های استفاده شده برای ورود، غیرقابل فشرده‌سازی یا نویزی هستند، در حالی که از دست دادن توانایی شکاف زمانی کوتاه جلوگیری می‌کند. حافظه کوتاه-مدت بلند، طراحی شده توسط Hochreiter و Schmidhuber، یک نوع ویژه از شبکه عصبی مکرر (RNN) است که در برابر وابستگی‌های طولانی‌مدت به طور پیش‌فرض همراه با آن مجهز است. در الگوریتم LSTM، ورود یک گام کنونی خروجی گام قبلی است، و این امر با حل مشکلات وابستگی‌های طولانی‌مدت RNN که در آن RNN پیش‌بینی دقیقی از اطلاعات اخیر انجام می‌دهد اما قادر به پیش‌بینی داده‌های ذخیره شده در حافظه طولانی‌مدت نیست، بهبود یافته است. با افزایش طول شکاف، کارایی RNN کاهش می‌یابد. برخی از کاربردهای اصلی LSTM شامل توضیح تصاویر، تولید چتبات‌های خطنویسی برای پاسخ‌گویی به سوالات و موارد مختلف دیگر هستند.

۱-۶ ساختار LSTM

ساختار LSTM که شامل چهار شبکه عصبی و بلوک‌های حافظه مختلفی به نام سلول‌ها است، در زیر تصویب شده است. دروازه‌ها

تغییرات حافظه را بر داده‌های ذخیره شده در سلول‌ها انجام می‌دهند. دروازه‌ها سه نوع هستند.

۱-۱-۶ دروازه فراموشی

اطلاعاتی که دیگر نیازی به آن‌ها نیست، از سلول با استفاده از دروازه فراموش حذف می‌شوند. ورودی در یک زمان خاص، و خروجی سلول قبلی با استفاده از ماتریس‌های وزن‌دار ضرب می‌شوند و با اضافه شدن بایاس، به بیرون می‌روند. برای دریافت یک خروجی دودویی، نتیجه از آنالیز یک عملکرد فعال‌سازی عبور می‌کند. اگر خروجی '۱' باشد، اطلاعات در حالت سلول حفظ می‌شود و اگر خروجی '۰' باشد، پاک می‌شود.

۲-۱-۶ دروازه ورود

این دروازه وظیفه افزودن اطلاعات حیاتی به حالت سلول را انجام می‌دهد. اطلاعات از طریق یک تابع سیگموئید پردازش می‌شوند و مقادیری که باید نگه داشته شوند، تصفیه می‌شوند. مرحله بعد شامل ایجاد بردار با استفاده از تابع \tanh می‌شود که یک خروجی از ۱- تا ۱+ را تولید می‌کند که شامل تمام مقادیر ممکن از ۰ است. در نهایت، مقادیر بردار و نتایج تصفیه شده تابع سیگموئید با هم ضرب شده تا نتایج مفید مشتق شوند.

۳-۱-۶ دروازه خروجی

بر اساس داده‌های ذخیره شده در حالت سلول فعلی، خروجی اعلام می‌شود. در ابتدا، با استفاده از تابع \tanh یک بردار برای مقادیر سلولی ایجاد می‌شود. مرحله بعد شامل تنظیم اطلاعات با استفاده از تابع سیگموئید و تصفیه مقادیری که باید نگه داشته شوند است. در نهایت، حاصل ضرب مقادیر تنظیم شده و مقادیر برداری به عنوان خروجی ارسال می‌شود که به عنوان ورودی برای سلول بعدی عمل می‌کند.

۲-۶ عملکرد LSTM

مرحله اول نیاز به تصمیم‌گیری در مورد حذف اطلاعات غیرضروری از حالت سلولی دارد. این تصمیم‌گیری‌ها توسط "لایه دروازه فراموش" که یکی از لایه‌های سیگموئیدی است، حل می‌شوند. در زمان تصمیم‌گیری، x و h در نظر گرفته می‌شوند و نتایج برای تمام اعداد متعلق به سلول C می‌تواند هر عددی در بازه ۰ تا ۱ باشد. در صورتی که خروجی '۱' باشد، این نشان می‌دهد که اطلاعات باید ذخیره شوند، در حالی که '۰' نشان می‌دهد که اطلاعات باید دور انداخته شوند. پس از آن، حالا باید برنامه‌ریزی کرد که چه اطلاعاتی باید در سلول‌ها ذخیره شود. این فرآیند از دو بخش تشکیل شده است. ابتدا، لایه دروازه استفاده شده برای ورود، که همچنین یک لایه پوشش سیگموئیدیست، مقادیری را که باید به‌روز شوند حل می‌کند. ثانوی، یک بردار جدید بنام t توسط یک لایه تانچ، که برای اضافه شدن در این حالت استفاده می‌شود، تولید می‌شود. در پایان، حالا مهم است که برنامه‌ریزی شود که خروجی، که براساس حالت

سلول است، تصمیم گرفته شود، با این حال، خروجی یک خروجی تصفیه‌شده خواهد بود. بنابراین، ابتدا، یک لایه سیگموئیدی بخشی از حالت سلول را که باید به عنوان خروجی ارائه شود انتخاب می‌کند. پس از آن، حالت این سلول از طریق \tanh گذر داده می‌شود (برای محدود کردن نتایج از ۱- تا ۱) و سپس می‌توان آن را با ضرب آن به همراه نتیجه لایه دروازه سیگموئیدی، برای به‌دست آوردن خروجی دقیق مورد نظر، افزایش داد.

۳-۶ مدل LSTM-RNN برای پیش‌بینی نیاز به بار برق

- کاربرد اخیر

۱-۳-۶ زمینه کار اخیر

از زمان تکامل خانه‌های هوشمند، تخمین و پیش‌بینی نیاز به بار برق الکتریکی مسأله‌ای بسیار مهم است، اصلماً به دلیل اینکه شرکت‌ها و انجمن‌های مربوط به برق و الکتریسیته، می‌توانند برنامه‌ریزی و زمان‌بندی موثرتری برای بارها داشته باشند و میزان تولید اضافی انرژی را کاهش دهند. Bouktif Salah و همکاران (۲۰۱۸)، تحقیقات تجربی را در مورد استفاده از مدل الگوریتم LSTM برای پیش‌بینی بار برق با استفاده از انتخاب ویژگی و الگوریتم ژنتیک انجام دادند. آن‌ها هدف داشتند که یک مدل مبتنی بر LSTM بسازند تا مدل‌های پیش‌بینی برای برنامه‌ریزی و زمان‌بندی بار را طراحی کنند. بسیاری از الگوریتم‌های غیرخطی و خطی آموزش داده شدند تا مناسب‌ترین یکی به عنوان پایه انتخاب شود، با استفاده از پارامترهای متناسب و در آخر استفاده از الگوریتم ژنتیک برای تعیین تاخیر زمانی بهینه و لایه‌هایی که باید توسط شبکه LSTM استفاده شوند. داده‌های مصرف برق شهری فرانسه برای تحقیق و تحلیل استفاده شدند. آن‌ها از طریق تحقیقات تجربی خود اثبات کردند که مدل LSTM نتایج بسیار دقیق‌تری ارائه می‌دهد در مقایسه با مدل‌های یادگیری ماشین که با تنظیم پارامترهای فوق‌پارامتر بهینه شده است. نتایج آن‌ها نشان داد که شبکه LSTM با استفاده از ویژگی‌های زمانی محدود شده، تمام ویژگی‌ها و ویژگی‌های سری زمانی پیچیده را کسب کرده و با خطای متوسط ریشه مربع و خطای میانگین مطلق کوچک‌تری برای یک فضای شهری بزرگ در مورد پیش‌بینی و پیش‌بینی نشان می‌دهد.

۲-۳-۶ شرح و نتایج

Bouktif Salah و همکاران (۲۰۱۸)، یک مدل توسعه دادند که از سیستم پوشش‌دهنده و متنوع، تاخیر زمانی منطقی و لایه‌های برای مدل LSTM استفاده می‌کند، و در نهایت، الگوریتم ژنتیک آن‌ها را قادر به کنترل بیش‌اندازم‌گیری می‌کند و به‌دست آوردن پیش‌بینی دقیق‌تر و قابل اعتمادتر. آن‌ها مجموعه‌داده‌های بزرگی را برای یک فضای شهری که یک بازه زمانی حدود ۹ سال در تعریف ۳۰ دقیقه داشت، جمع‌آوری کردند استفاده کردند که با استفاده از آن، یک سیستم شامل LSTM-RNN را آموزش دادند تا میزان متوسط نیاز به بار برق الکتریکی را پیش‌بینی کنند. معیارهای

ارزیابی که برای تحلیل استفاده کردند شامل ضریب واریانس، خطای مربع میانگین ریشه، و خطای مطلق میانگین بودند. به عنوان بخشی از تحقیقات تجربی خود، سیستم طراحی شده LSTM-RNN با استاندارد یادگیری ماشین مقایسه شد، و مدل طراحی شده نتایج بهتری را در میان مدل‌های غیرخطی و خطی ارائه داد. نمودارها و نتایج مربوط به الگوریتم LSTM بدین صورت است: نمودار بالا نشان دهنده مصرف برق به مگاوات از ژانویه تا فوریه سال ۲۰۱۱ است. جداول ۱۵، ۱۶ عملکرد مدل‌های یادگیری ماشین از جمله مدل LSTM در مجموعه آزمایشی را نشان می‌دهد، که با دیدن که می‌توان گرفت که مدل LSTM در این مورد با دیگر مدل‌ها برتر است، خطای مربع میانگین ریشه (RMSE) و خطای مطلق میانگین (MAE) به طور مقایسه‌ای در مورد LSTM بسیار کوچکتر است. علاوه بر این، نمودار نشان‌دهنده در شکل ۲۰، تفاوت بین بار واقعی و پیش‌بینی شده توسط مدل LSTM را نشان می‌دهد. تحقیق انجام شده توسط Bouktif Salah و همکاران (۲۰۱۸) نشان می‌دهد که مدل LSTM نتایج بسیار دقیقی برای پیش‌بینی بار برق فراهم کرده است.

۴-۶ مزایای الگوریتم LSTM

توانایی پل‌رفتن از تاخیرهای زمانی بزرگ توسط گسترش خطای ثابت داخل سلول حافظه LSTM ها در مقابل کاهش گرادین‌ها مقاوم هستند. LSTM ها قادر به مدیریت وابستگی‌های موالی طولانی‌مدت هستند. LSTM ها نیازی به تنظیم دقیق پارامتر ندارند. LSTM ها حافظه‌ای تا زمان بیشتری را داشته و نشان می‌دهند. LSTM ها در مقیاس پیش‌بینی دقت بالایی از خود نشان می‌دهند.

۵-۶ معایب الگوریتم LSTM

LSTM ناتوان در حل مشکلات ناپدیدشدن گرادین به طور کامل هست زیرا سلول پیچیده‌تر شده است. LSTM ها زمان و منابع به مقدار زیادی نیاز دارند برای آموزش، به عبارت دیگر، نیاز به پهنای باند حافظه بسیار بالا دارند. بنابراین، در مورد سخت‌افزار ناکارآمد می‌باشند. با افزایش تقاضا برای استخراج داده، جستجویی برای مدل‌های دارای زمان ذخیره‌سازی بلندتر وجود دارد. شروع وزن بر LSTM ها تصادفی تاثیر می‌گذارد و باعث می‌شود آن‌ها شبیه یک شبکه عصبی خوراک به نظر برسند. مشکلات نصب که حتی با الگوریتم قطع شبکه برطرف نمی‌شوند.

۷ مقایسه کمی الگوریتم‌ها

۱-۷ مطالعه موردی برای مقایسه الگوریتم‌های K-NN ، SVM و درخت تصمیم

برای انجام تحلیل کمی، مقاله تحقیقی با عنوان "مطالعه مقایسه ای الگوریتم SVM KNN و درخت تصمیم برای پیش‌بینی عملکرد دانش‌آموزان" نوشته شده توسط Wiyono Slamet و همکاران در سال ۲۰۲۰، مورد بررسی قرار گرفته است، که در آن یک مجموعه

داده در زمان واقعی، شامل ۶ متغیر مختلف توسط آن‌ها برای انجام تحقیق جامع انتخاب شده بود. کار تحقیقی آن‌ها ادامه‌ای بود از کارهای مختلفی که در گذشته برای پیش‌بینی عملکرد دانش‌آموزان بر اساس چندین الگوریتم یادگیری ماشین منتشر شده بود. پلتفرم استفاده شده برای تحلیل آن‌ها، studio R بود، و آن‌ها نتایج قابل اعتماد و معتبری ارائه دادند پس از انجام دقیق فرآیندهای مختلف مانند جمع‌آوری داده‌ها، پیش‌پردازش داده‌ها، ساخت مدل‌های قوی پس از آموزش، اعتبارسنجی و آزمایش مدل‌ها بر اساس الگوریتم‌های مختلف، و در نهایت مقایسه و ارزیابی آن‌ها، به نحو کمی. شش متغیر موجود در مجموعه داده، شامل معدل نمره، معدل نمره ارشد، شهر محل زندگی، رشته تحصیلی، نوع مدرسه، حرفه والدین، و عملکرد دانش‌آموزان به صورت جدول ۱۷ نمایش داده شده است. آن‌ها به استفاده از پیش‌پردازش داده پرداختند به منظور حذف باگ‌های خاص مرتبط با مقادیر نقطه داده‌ای گم شده یا ویژگی‌های مختلف، به عنوان مثال، به منظور ایجاد مدل‌های قوی‌تر و قابل اعتماد. یکی از روش‌های متداول برای انجام پیش‌پردازش، تقسیم داده موجود به داده‌های آموزش و آزمایش است، که از آن اولی برای آموزش و ساخت مدل استفاده می‌شود، در حالی که دومی برای آزمایش مدل از حیث دقت در پیش‌بینی نیاز است. Wiyono Slamet و همکاران از یک رویکرد مشابه برای محاسبه و اجرای این تحقیق استفاده کردند، و در نهایت سه مدل را بر اساس پارامترهای عملکرد مختلف برای پیش‌بینی عملکرد دانش‌آموزان مقایسه و ارزیابی کردند. همانطور که در بالا بحث شد، قبل از پردازش داده، آن‌ها را به داده‌های آموزش و آزمایش به نسبت ۷۵:۲۵ تقسیم کردند، که در آن تصمیم گرفته شد ۱۱۴۸ نمونه با ۶ پیش‌بین کننده و ۲ کلاس استفاده شود، با پشتیبانی از اعتبارسنجی ۱۰ تا ۱۰ برابر از ۱۰ fold تکرار شده است. مدل ایجاد شده توسط آن‌ها در جدول ۱۸ نمایش داده شده است، پس از آن مدل با داده‌های آزمایش آزمایش شده، و ماتریس های گیجگاه نتیجه‌آور برای الگوریتم‌های SVM KNN و درخت تصمیم به ترتیب در جداول ۱۹، ۲۰ و ۲۱ نمایش داده شده است. قبل از آزمایش این مدل‌ها در جدول ۱۸، دیده شد که الگوریتم KNN عملکرد بهتری دارد زمانی که $k=3$ ، و دقت ۵۹.۹۴٪ جدول ۲۳ مقایسه کمی خلاصه نتایج این مطالعه موردی بر اساس مقاله تحقیقی نوشته شده توسط Wiyono Slamet و همکاران است که شامل سه الگوریتم SVM KNN و درخت تصمیم؛ برای پیش‌بینی عملکرد دانش‌آموزان، می‌باشد.

۲-۷ مطالعه موردی برای مقایسه الگوریتم ژنتیک و الگوریتم LSTM

برای مقایسه و تجزیه و تحلیل کمی الگوریتم‌های دوم باقی‌مانده، یعنی الگوریتم ژنتیک و الگوریتم حافظه کوتاه‌مدت و بلند مدت (LSTM)، مقاله تحقیقی با عنوان "روشی بر اساس CNN GA و LSTM برای پیش‌بینی جریان گردشگران روزانه در اماکن دیدنی" نوشته شده

توسط Lu Wenxing و همکاران، منتشر شده در سال ۲۰۲۰، مورد بررسی قرار گرفته است و مطالعه موردی از این مقاله که به مقایسه الگوریتم‌ها کمک می‌کند، در زیر آورده شده است. این مقاله نوشته شده توسط Lu Wenxing و همکاران، هدف ایجاد یک مدل برای پیش‌بینی جریان گردشگران را داشت تا امکان جذاب و دیدنی معمولی به طور صحیح حفظ و اداره شود. با این حال، از آنجایی که هیچ مدلی نمی‌تواند به تنهایی پیش‌بینی دقیقی را انجام دهد به دلیل داده‌های بسیار متغیر، نویسندگان این مقاله روی یک مدل کار کردند که از شبکه‌های عصبی پیچشی (CNN) همراه با یک الگوریتم یادگیری عمیق یعنی الگوریتم حافظه کوتاه‌مدت و بلندمدت و در نهایت بهینه‌سازی شده توسط الگوریتم ژنتیک، برای پیش‌بینی گروه روزانه یک مکان به نام Huangshan در چین استفاده کردند. به عنوان بخشی از اجرای تحقیقاتشان، آن‌ها در ابتدا نقشه‌های ویژگی پیوسته را از انواع مختلف داده‌ها مانند داده‌های هواشناسی، جستجوی شبکه و غیره تشکیل دادند. در ادامه، استخراج بردار توسط شبکه پیچشی رخ داد، و پس از استخراج موفق، بردارهای مشتق‌شده به شبکه LSTM برای پیش‌بینی داده‌های سری زمانی غذا داده شد. به طور سنتی، مجموعه داده در حال استفاده قبل از مرحله پیش‌بینی به پیش‌پردازش و نرمال‌سازی می‌گذرد. مدل طراحی شده از نظر عملکردش به صورت کمی مقایسه شده است، بدون بهینه‌سازی با الگوریتم ژنتیک و با بهینه‌سازی با الگوریتم ژنتیک با استفاده از پارامترهای عملکرد مشترک، مانند خطای مطلق میانگین (MAE)، خطای مطلق درصدی میانگین (MAPE)، شاخص توافق (IA) و ضریب همبستگی شخصی. بعد از مقایسه منصفانه بین LSTM-CNN، LSTM-GA-LSTM-CNN و مدل GA انجام شده توسط نویسندگان، مدل شامل الگوریتم ژنتیک (GA-LSTM-CNN) حدود ۲۲.۸ جدول ۲۵ مقایسه عملکرد مدل‌های مختلفی را نشان می‌دهد که برای پیش‌بینی جریان گردشگران بر روزانه در شهری به نام Huangshan در چین استفاده شده است. همانطور که از جدول مشخص است، اگر الگوریتم‌ها برای عملکرد فردی مدنظر قرار گیرند، LSTM حدود ۵ جدول ۲۶ نتایجی که برای ضریب همبستگی پیرسون (r) جرد مقدار کوچکی الگوریتم GA نسبت به رشته LSTM ارائه داده شده است. جدول ۲۷ نتایجی که برای شاخص توافق (IA) جرد عملکرد LSTM به طور روشن از عملکرد الگوریتم ژنتیک فراتر رفته است. بنابراین، براساس نتایج و محاسبه سه پارامتر عملکرد، MAPE r و IA، این مطمئن شدن نتیجه می‌دهد که LSTM در عملکرد برتر به الگوریتم ژنتیک برای چنین پیش‌بینی‌های تحلیلی، است.

۸ دامنه‌ی آینده

به دلیل ویژگی‌های انقلابی یادگیری ماشین، دامنه‌ی آن روزبه روز گسترش می‌یابد. صنعت خودرو یک مثال است که نمایانگر نوآوری‌های عالی به کمک یادگیری ماشین می‌باشد. برندهای

معروف اتومبیل‌ها، مانند تسلا، تویوتا، مرسدس بنز، گوگل، نپسان و غیره مقادیر بزرگی پول در این حوزه سرمایه‌گذاری کرده‌اند تا برنامه‌های نوآورانه‌ای با استفاده از یادگیری ماشین و سایر هوش مصنوعی‌ها تدارک ببینند. خودروی خودران معروفی که توسط تسلا به وجود آمده، با استفاده از سنسورهای اینترنت اشیا (IoT)، یادگیری ماشین، دوربین‌های با وضوح بالا و غیره ساخته شده است که تنها نیاز به ورود انسان برای برنامه‌ریزی مقصد مورد نظر به سیستم دارد و بقیه کار توسط ماشین انجام می‌شود، یعنی انتخاب مسیر مناسب، خالی از ترافیک و تضمین رساندن مسافر به مقصدش به صورت ایمن. رباتیک یک حوزه دیگر است که به طور مداوم در میان دانشمندان، محققان و حتی مردم عادی مورد بحث قرار دارد. یادگیری ماشین و هوش مصنوعی امکان ابداعاتی همچون ربات قابل برنامه‌ریزی اولین بار در سال ۱۹۵۴ با نام Unimate و سپس ایجاد اولین ربات هوش مصنوعی به نام Sophia را امکان‌پذیر کرده است. در این حوزه دامنه‌ی روشنی برای تحقیقات وجود دارد و انتظار می‌رود آینده ربات‌های ایجاد شده با استفاده از یادگیری ماشین و هوش مصنوعی و فناوری‌های انقلابی دیگر که قادر به انجام وظایف مشابه انسان‌ها در همه حوزه‌ها شامل پزشکی باشند. یادگیری ماشین هنوز هم باید فراتر از حدود بررسی شود، و یکی از حوزه‌هایی که به شدت کمک می‌کند در بررسی یادگیری ماشین، محاسبه‌ی کوانتوم است. این محلولیت وقوع مکانیکی شبیه به سوپرپوزیشن و پیچیدگی کوانتوم را تشکیل می‌دهد (برای جزئیات بند ۲۸ را ببینید). به بررسی جزئیات نوآورانه‌ی کاربردهای پنج الگوریتم یادگیری ماشین در مقاله پرداخته شده است. در زیر، جدولی که شامل برنامه‌های نوآورانه، نویسندگان مقالات مربوطه و یافته‌های آن‌ها برای خلاصه‌سازی توضیحات فوق، آورده شده است (برای جزئیات بند ۲۹ را ببینید).

۹ نتیجه‌گیری

این مقاله یک مطالعه مقایسه‌ای از الگوریتم‌های یادگیری ماشین، KNN ژنتیک SVM درخت تصمیم و LSTM به همراه برخی از کاربردهای نوآورانه‌ی اخیرشان را ارائه می‌دهد که در زمینه تحقیقات آینده دارای دامنه‌ی عظیمی می‌باشد. الگوریتم‌ها و مفاهیم مربوطه به جزئیات خیلی زیادی توضیح داده شده‌اند، از آغاز تا مهمترین کاربردهای جدیدشان. این مقاله نوری بر بسیاری از جنبه‌های حیاتی افکار پرتاب می‌کند، مانند زمانی و در چه شرایطی الگوریتم‌ها بر هایجسته نمودند و چگونه در سناریوی امروزی مفید بودند برای کاربردهای پیش‌بینی واقعی زمان و کاربردهای دیگر. بینش‌های متعلق به روش‌های پیاده‌سازی این الگوریتم‌ها نیز به طور جزئی مورد بحث قرار گرفته است و نتایج و عملکرد آن‌ها در کارهای تحقیقی اصیل و نو آور بحث شده است. مقایسه‌ی دقیقی از انواع الگوریتم‌های یادگیری ماشین براساس مبنای کیفی و کمی صورت گرفته است، و همچنین به صورت جدولی خلاصه

شده است. پس از انجام یک بررسی جامع و تحقیق در این حوزه، ما قادر بودیم به نتایج مهمی برسیم که شبکه LSTM و الگوریتم SVM از بهترین نتایج ارائه داده‌اند زمانی که به پیش‌بینی تحلیلی در برنامه‌های واقعی زمانی مرتبط با حوزه‌های چندرشته‌ای مانند پزشکی، تقلب‌های بانکی، شناسایی چهره، پیش‌بینی عملکرد دانش‌آموز، پیش‌بینی مصرف برق و غیره می‌پردازیم. شبکه LSTM شبکه یادگیری عمیق با بازخورد است و دارای این مزیت است که اطلاعات مورد نیاز را حفظ می‌کند، که این امر به آن امکان می‌دهد تا نتایج بسیار دقیقی را ارائه بدهد در زمینه پیش‌بینی. در نهایت، دامنه‌ی آینده تاکید می‌کند که تقاضای مورد انتظار و محبوبیت یادگیری ماشین و هوش مصنوعی در آینده، که انتظار می‌رود با انسان‌ها را در زمینه‌های مختلف پشتیبانی نماید یا کاملاً آن‌ها را جایگزین کند و تمام روند اتوماتیک کردن را در مقیاس و سرعت بزرگتر با کمک تحقیقات پیشرفته و دقیقتری فراهم آورد.