مجله علمی مطالعات پژوهشی در مهندسی برق آینده، ۱۴۰۳، دوره ۱، شماره ۱، ص ص ۱-۱۲

مجله علمی مطالعات پژوهشی
در مهندسی برق آینده
www.journalhi.com/pow/

چالش ها و رویکرد های جدید در کاربرد های اینترنت اشیا، پردازش تصویری و ماشین لرنینگ

علی کوهستانی [۱]

۱- استادیار گروه برق-مخابرات دانشگاه صنعتی قم

اطلاعات مقاله

مقاله پژوهشی کامل
دریافت: ۲۷ اسفند ۱۴۰۱
پذیرش: ۱۰ اردیبهشت ۱۴۰۲
ارائه در سایت: ۱۲ خرداد ۱۴۰۲

کلید واژگان:
اینترنت اشیا
پردازش تصویری
ماشین لرنینگ

چکیده

در عصر کنونی انقلاب صنعتی چهارم، دنیای دیجیتال دارای داده های زیادی مانند داده های اینترنت اشیا، داده های امنیت سایبری، داده های تلفن همراه، داده های تجاری، داده های رسانه های اجتماعی، داده های بهداشتی و غیره است. برای تجزیه و تحلیل هوشمندانه این داده ها و توسعه برنامه های هوشمند و خودکار مربوطه، دانش هوش مصنوعی، به ویژه یادگیری ماشینی کلید اصلی است. انواع مختلفی از الگوریتم های یادگیری ماشین مانند یادگیری نظارت شده، بدون نظارت، نیمه نظارت و تقویت در منطقه وجود دارد. یادگیری عمیق، که بخشی از یک خانواده گسترده تر از روش های یادگیری ماشین است، می تواند داده ها را در مقیاس بزرگ بصورت هوشمند تجزیه و تحلیل کند. گره های شرکت کننده در شبکه های اینترنت اشیا معمولاً محدود به منابع هستند، که باعث می شود اهداف حملات سایبری را جلب کنند. مشخصات منحصر به فرد گره های اینترنت اشیا باعث می شود که راه حل های موجود برای در برگرفتن کل طیف امنیتی شبکه های اینترنت اشیا کافی نباشد. تکنیک های یادگیری ماشین و یادگیری عمیق، که می توانند هوش جاسازی شده را در دستگاه ها و شبکه های اینترنت اشیا فراهم کنند، می توانند برای مقابله با مشکلات مختلف امنیتی استفاده شوند. در این مقاله، ما به طور سیستماتیک الزامات امنیتی، بردارهای حمله و راه حل های امنیتی فعلی برای شبکه های اینترنت اشیا را مرور می کنیم. سپس ما خلاءهای موجود در این راه حل های امنیتی را که خواستار رویکردهای ML هستند، روشن می کنیم. ما همچنین در مورد راه حل های موجود MLبرای پرداختن به مشکلات امنیتی مختلف در شبکه های اینترنت اشیا به طور مفصل بحث می کنیم.

Challenges and new approaches in the applications of Internet of Things, image processing and machine learning

Ali Kuhestani[1]

1- Assistant Professor of Electricity-Communications Department, Qom University of Technology.

Article Information

Original Research Paper
Received 18 March 2023
Accepted 02 October 2023
Available Online 04 October 2023

Keywords:
Internet of Things
Image processing
Machine learning

Abstract

In the current era of the fourth industrial revolution, the digital world has a lot of data, such as IoT data, cyber security data, mobile phone data, business data, social media data, health data, etc. To intelligently analyze this data and develop the corresponding intelligent and automated programs, the knowledge of artificial intelligence, especially machine learning, is the key. There are different types of machine learning algorithms such as supervised, unsupervised, semi-supervised and reinforcement learning in the area. Deep learning, which is part of a broader family of machine learning methods, can intelligently analyze large-scale data. Nodes participating in IoT networks are usually resource-constrained, which makes them attractive targets for cyber attacks. The unique characteristics of IoT nodes make existing solutions insufficient to cover the entire security spectrum of IoT networks. Machine learning and deep learning techniques, which can provide embedded intelligence in IoT devices and networks, can be used to tackle various security problems. In this article, we systematically review the security requirements, attack vectors, and current security solutions for IoT networks. We then highlight the gaps in these security solutions that call for ML approaches. We also discuss in detail the existing ML solutions to address various security problems in IoT networks.

۱- مفاهیم اولیه

ما در عصر داده ها زندگی می کنیم، جایی که همه چیز در اطراف ما به یک منبع داده متصل است، و همه چیز در زندگی ما به صورت دیجیتالی ضبط می شود [۱، ۱۳]. به عنوان مثال، دنیای الکترونیکی فعلی دارای انواع مختلفی از داده ها است، مانند داده های اینترنت اشیا، داده های امنیت سایبری، داده های شهر هوشمند، داده های تجاری، داده های تلفن های هوشمند، داده های رسانه های اجتماعی، داده های سلامت، داده های COVID-19، و خیلی بیشتر. داده ها می توانند ساختار یافته، نیمه ساختاریافته یا غیر ساختاری باشند، و به طور خلاصه در بخش مورد بحث قرار می گیرند."انواع داده های واقعی و تکنیک های یادگیری ماشین"، که روز به روز در حال افزایش است.

استخراج بینش از این داده ها می تواند برای ایجاد برنامه های مختلف هوشمند در دامنه های مربوطه استفاده شود. برای ساخت یک سیستم امنیت سایبری خودکار و هوشمند مبتنی بر داده، می توان از داده های مربوط به امنیت سایبری استفاده کرد. برای ساخت برنامه های تلفن همراه هوشمند آگاه از زمینه، می توان از داده های مربوطه مربوط به تلفن همراه استفاده کرد که بلافاصله و به روشی هوشمندانه از اطلاعات لازم برای استخراج بینش یا دانش مفید از اطلاعات، که برنامه های واقعی در آنها مبتنی است.

۱.۱ ماشین لرنینگ

معمولاً سیستم هایی را فراهم می کند که توانایی یادگیری و افزایش تجربه را به طور خودکار و بدون برنامه ریزی خاص داشته باشند و به طور کلی به عنوان محبوب ترین فن آوری های اخیر در چهارمین انقلاب صنعتی شناخته می شود [۱۳، ۱۴]. معمولاً اتوماسیون مداوم ساختهای متداول و روشهای صنعتی، از جمله پردازش داده های اکتشافی، با استفاده از فن آوری های هوشمند جدید مانند اتوماسیون یادگیری ماشین است. به طور کلی، اثربخشی و کارایی یک راه حل یادگیری ماشین به ماهیت و مشخصات داده ها و عملکرد الگوریتم های یادگیری بستگی دارد. در زمینه الگوریتم های یادگیری ماشین، تجزیه و تحلیل طبقه بندی، رگرسیون، خوشه بندی داده ها، مهندسی ویژگی ها و کاهش ابعاد، یادگیری قاعده انجمن، یا تکنیک های یادگیری تقویت برای ایجاد موثر سیستم های مبتنی بر داده وجود دارد [۲، ۱۶]. علاوه بر این، یادگیری عمیق از شبکه عصبی مصنوعی نشات می گیرد که می تواند برای تجزیه و تحلیل هوشمندانه داده ها مورد استفاده قرار گیرد، که به عنوان بخشی از خانواده گسترده تر رویکردهای یادگیری ماشین شناخته می شود. دلیل آن این است که هدف الگوریتم های یادگیری مختلف متفاوت است، حتی نتیجه الگوریتم های یادگیری مختلف در یک دسته مشابه بسته به ویژگی های داده ممکن است متفاوت باشد. با توجه به اهمیت و پتانسیل "یادگیری ماشین" برای تجزیه و تحلیل داده های ذکر شده در بالا، در این مقاله، ما یک دیدگاه جامع در مورد انواع مختلف الگوریتم های یادگیری ماشین ارائه می دهیم که می تواند برای افزایش هوش و توانایی های یک برنامه استفاده شود .

بنابراین، سهم اصلی این مطالعه توضیح اصول و پتانسیل تکنیک های مختلف یادگیری ماشین و کاربرد آنها در زمینه های مختلف کاربرد در دنیای واقعی است که قبلاً ذکر شد. مشارکت های اساسی این مقاله به شرح زیر است: - برای تعیین دامنه مطالعه ما با در نظر گرفتن ماهیت و ویژگی های انواع مختلف داده های دنیای واقعی و توانایی های فنون مختلف یادگیری.

برای ارائه یک دیدگاه جامع در مورد الگوریتم های یادگیری ماشین که می تواند برای افزایش هوش و قابلیت های یک برنامه داده محور استفاده شود.

ما به طور خلاصه الگوریتم های مختلف یادگیری ماشین را در بخش بعدی بحث می کنیم و توضیح می دهیم و به دنبال آن مناطق مختلف برنامه های کاربردی واقعی مبتنی بر الگوریتم های یادگیری ماشین مورد بحث و جمع بندی قرار می گیرند.

۱.۲ اینترنت اشیا

به عنوان شبکه ای به هم پیوسته و توزیع شده از سیستم های جاسازی شده در ارتباط است که از طریق فن آوری های ارتباطی سیمی یا بی سیم ارتباط برقرار می کند. موارد موجود در اینترنت اشیا به اشیا مربوط به زندگی روزمره ما اعم از لامپ های هوشمند خانگی مانند لامپ هوشمند، آداپتور هوشمند، کنتور هوشمند، یخچال هوشمند، فر هوشمند، کولر گازی، سنسور دما، ردیاب دود، دوربین IP گرفته تا پیچیده تر اشاره دارد. دستگاه هایی مانند دستگاه های شناسایی فرکانس رادیویی، ردیاب های ضربان قلب، شتاب سنج ها، سنسورها در پارکینگ و طیف وسیعی از حسگرهای دیگر در اتومبیل و غیره مقیاس عظیم شبکه های اینترنت اشیا چالش های جدیدی مانند مدیریت این دستگاه ها، مقدار زیاد داده، ذخیره سازی، ارتباطات، محاسبات و امنیت و حریم خصوصی را به همراه دارد. ناهمگنی داده های تولید شده توسط اینترنت اشیا جبهه دیگری برای سازوکارهای پردازش داده های فعلی ایجاد می کند.

۱.۳ماشین لرنینگ

می تواند به ماشین ها و دستگاه های هوشمند کمک کند تا دانش مفید را از داده های تولید شده توسط دستگاه یا انسان استنباط کنند. همچنین می تواند به عنوان توانایی یک دستگاه هوشمند در تغییر یا خودکار کردن وضعیت یا رفتار براساس دانش که به عنوان بخشی اساسی برای یک راه حل اینترنت اشیا در نظر گرفته می شود، تعریف شود. از تکنیک های ML در کارهایی مانند طبقه بندی، رگرسیون و تخمین چگالی استفاده شده است. انواع برنامه ها مانند بینایی رایانه ای، تشخیص تقلب، انفورماتیک زیستی، شناسایی بدافزار، احراز هویت و تشخیص گفتار از الگوریتم ها و تکنیک های ML استفاده می کنند.

به روشی مشابه، ML می تواند برای ارائه خدمات هوشمند در اینترنت اشیا استفاده شود. در ادامه، ما در مورد نظرسنجی های موجود بحث می کنیم که قبلاً در ادبیات مربوط به جنبه های مختلف امنیت در شبکه های اینترنت اشیا از طریق تکنیک های ML منتشر شده است.

نظرسنجی های موجود IoT دارای ادبیات غنی است و تا به امروز، نظرسنجی های بسیاری منتشر شده است که جنبه های مختلف امنیت اینترنت اشیا را پوشش می دهد. از نظر دانش موجود، بیشتر نظر سنجی ها در ادبیات بر روی تکنیک های ML مورد استفاده در اینترنت اشیا نیست. علاوه بر این، نظرسنجی های موجود یا مخصوص برنامه هستند یا طیف کاملی از امنیت و حریم خصوصی در شبکه های اینترنت اشیا را شامل نمی شوند. ما موضوعاتی را که در این نظرسنجی ها و موارد مربوطه را در نظرسنجی خود بیان می کنیم، بیان می کنیم.

در جدول ۱، ما نظرسنجی های مربوط به نقش ML و DL در امنیت و شبکه های اینترنت اشیا را پوشش می دهیم.

در شبکه های اینترنت اشیا ادبیات غنی در مورد تکنیک های مبتنی بر ML و DL موجود است، اما ما فقط به جنبه های امنیتی شبکه های اینترنت اشیا و نقش ML و DL در پرداختن به چالش های امنیتی در شبکه های اینترنت اشیا تمرکز می کنیم.

به طور خاص، مقاله ذکر شده بر روی روش های ML و DL برای بحث در مورد کاربرد آنها در امنیت در لایه های مختلف تمرکز دارد، در حالی که ما مشکلات امنیتی موجود را در حوزه های مختلف عملکردی بررسی می کنیم و در مورد راه حل های ML و DL برای مسائل امنیتی مانند احراز هویت، مجوز بحث می کنیم. - بدون در نظر گرفتن لایه ها، DDoS، بدافزار و غیره.

برای پر کردن شکاف ها، ما یک بررسی جامع از تکنیک های ML و DL مورد استفاده در امنیت اینترنت اشیا انجام می دهیم.

برای تمرکز بیشتر بر روی جنبه کاربردی اینترنت اشیا، ما عمیق تر به راه حل های امنیتی مبتنی بر ML و DL در اینترنت اشیا می پردازیم. برای این منظور، ما همچنین در مورد چالش های تحقیقاتی موجود و مسیرهای آینده برای تحقیقات بیشتر در مورد ML و DL برای شبکه های اینترنت اشیا بحث می شود. هدف ما از بین بردن شکاف بین الزامات امنیتی اینترنت اشیا و توانایی های ML و DL است که به شما کمک می کند تا چالش های امنیتی فعلی شبکه های اینترنت اشیا را برطرف کنید.

مشارکتهای اصلی این مقاله را می توان به شرح زیر خلاصه کرد:

۱) ما در حال حاضر یک بررسی منظم و جامع از نقش ماشین و مکانیسم های یادگیری عمیق در اینترنت اشیا ارائه می دهیم.

۲) ما نتایج پیشرفته ML و DL در شبکه های اینترنت اشیا را با تمرکز بر امنیت و حریم خصوصی شبکه های اینترنت اشیا شرح می دهیم.

در واقع، ما به طور دقیق، شرایط امنیتی، سطح حمله در اینترنت اشیا را بررسی می کنیم و سپس در مورد راه حل های مبتنی بر ML و DL برای کاهش حملات امنیتی در شبکه های اینترنت اشیا بحث می کنیم.

۲- تحقیقات پیشین

معمولاً در دسترس بودن داده ها به عنوان کلیدی برای ساخت یک مدل یادگیری ماشین یا سیستم های دنیای واقعی مبتنی بر داده در نظر گرفته می شود [۱۳، ۱۴]. داده ها می توانند اشکال مختلفی داشته باشند، مانند ساختار یافته، نیمه ساختاری یا غیر ساختاری [۲، ۷]. علاوه بر این، "فراداده" نوع دیگری است که به طور معمول داده های مربوط به داده ها را نشان می دهد. در طرح های کاملاً مشخص، مانند پایگاه داده های نسبی، داده های ساختاریافته معمولاً در قالب جدول ذخیره می شوند.

ساختار ناپذیر: از طرف دیگر، هیچ قالب یا سازمانی از پیش تعریف شده برای داده های بدون ساختار وجود ندارد، این امر ضبط، پردازش و تجزیه و تحلیل را که بیشتر شامل متن و مواد چندرسانه ای است، بسیار دشوارتر می کند.

داده های حسگر، ایمیل ها، ورودی های وبلاگ، ویکی ها و اسناد پردازش کلمه، فایل های PDF، فایل های صوتی، فیلم ها، تصاویر، ارائه ها، صفحات وب و بسیاری از انواع دیگر اسناد تجاری را می توان به عنوان داده های بدون ساختار در نظر گرفت.

نیمه ساختار یافته: داده های نیمه ساختاری مانند داده های ساختاری که در بالا ذکر شد در یک پایگاه داده رابطه ای ذخیره نمی شوند، اما دارای ویژگی های سازمانی خاصی هستند که تجزیه و تحلیل را آسان می کنند.

فراداده: این فرم عادی داده نیست، بلکه "داده در مورد داده" است.

تفاوت اصلی بین "داده" و "فراداده" این است که داده ها صرفاً ماده ای هستند که می توانند چیزی را نسبت به خصوصیات داده های سازمان طبقه بندی، اندازه گیری یا حتی مستند کنند.

ناهمگنی: در یک شبکه اینترنت اشیا، تعداد زیادی از دستگاه های مختلف با قابلیت ها، ویژگی ها و پروتکل های ارتباطی مختلف با یکدیگر ارتباط برقرار می کنند.

چنین ناهمگنی از یک سو ارتباط بین پلتفرمی بین دستگاه های مختلف را امکان پذیر می کند، اما از سوی دیگر چالش های جدیدی را به شبکه اینترنت اشیا وارد می کند. برخی از این چالش ها شامل طراحی شبکه و معماری ذخیره سازی برای دستگاه های هوشمند، پروتکل های کارآمد ارتباط داده ها، شناسایی فعال و محافظت از اینترنت اشیا از حملات مخرب، استاندارد سازی فن آوری ها و دستگاه ها و رابط های برنامه و غیره است.

اتصال متقابل: انتظار می رود دستگاه های اینترنت اشیا to به زیرساخت های اطلاعاتی و ارتباطی جهانی متصل شوند و از هرجای هر زمان و هر زمان قابل دسترسی هستند. ارتباط در مجاورت: یکی دیگر از ویژگیهای بارز اینترنت اشیا the ارتباط در مجاورت بدون درگیر کردن مقامات مرکزی مانند ایستگاههای پایه است. ارتباطات دستگاه به دستگاه از ویژگی های ارتباط نقطه به نقطه مانند ارتباط کوتاه برد اختصاصی و فناوری های مشابه بهره می برد. معماری اینترنت سنتی بیشتر به سمت ارتباطات شبکه محور متمایل است در حالیکه جداسازی شبکه ها و خدمات ارتباطات دستگاه محور و محتوا محور را امکان پذیر می سازد که باعث تمیز کردن سرویس IoT می شود. [۴۳،۴۴].

اینترنت فعلی زمین بازی جذاب حملات امنیتی است، از هک های ساده گرفته تا نقض امنیت هماهنگ شده در سطح شرکت ها که بر صنایع مختلف مانند بهداشت و درمان و تجارت تأثیر منفی گذاشته است. تا به امروز، مسائل امنیتی و حریم خصوصی از دیدگاه های مختلف از جمله امنیت ارتباطات، امنیت داده ها، حریم خصوصی، امنیت معماری، مدیریت هویت، تجزیه و تحلیل بدافزار و غیره به طور گسترده در حوزه اینترنت اشیا مورد تحقیق قرار گرفته است.

بر اساس این طبقه بندی، شباهت های اساسی بین مسائل امنیتی در حوزه سنتی IT و اینترنت اشیا وجود دارد. با این حال، نگرانی اصلی اینترنت اشیا محدودیت های منابع است که مانع تصویب راه حل های پیچیده امنیتی موجود در شبکه های اینترنت اشیا می شود. راه حل های امنیت و حریم خصوصی در اینترنت اشیا نیاز به طراحی لایه ای و الگوریتم های بهینه سازی شده دارد.

به عنوان مثال به دلیل محدودیت های محاسباتی، دستگاه های اینترنت اشیا ممکن است برای کنار آمدن با امنیت و حریم خصوصی به نژادهای جدید رمزنگاری بهینه شده و سایر الگوریتم ها نیاز داشته باشند. یک رویکرد کلان امنیت و حریم خصوصی نسبت به اینترنت اشیا از راه حل های امنیتی موجود و همچنین توسعه مکانیزم های هوشمند، قوی، تکاملی و مقیاس پذیر جدید برای رسیدگی به چالش های امنیتی در اینترنت اشیا نامزد می شود [۲۰].

جدول ۱، ما نظرسنجی های مربوط به نقش ML و DL در امنیت و شبکه های اینترنت اشیا

Year	Paper	Topic(s) of the survey	Related sections in our paper	Enhancements in our paper
2017	[21]	IoT authentication and access control	Sec. III	Detailed security and privacy solutions through ML and DL in IoT
2017	[۲۱]	Security in the edge layer of IoT	Sec. III	Coverage of entire IoT from security, privacy standpoint
2018	[۲۳]	Security of IoT framework architectures	Sec. III	In-depth coverage of security and privacy in generic IoT with focus on state-of-the-art ML and DL techniques
2018	[۲۴]	IoT-enabled attacks on different sectors and assess different attacks in critical infrastructure	Sec. III	In-depth coverage of security issues, threats, attacks, and solutions in generic IoT
2018	[۲۵]	Security threats in IoT	Sec. III	Enhanced threats, attacks, and solutions in IoT
2019	[۲۶]	Data security in IoT and data lifecycle	N/A	Covering in-depth security issues and their ML- and DL-based solutions in IoT
2018	[۲۷]	Blockchain and SDN solutions for IoT	N/A	Generic IoT and in-depth review of the ML- and DL-based solutions in multiple domains
2018	[۳۳]	Resource scheduling techniques in IoT	N/A	Detailed investigation of security in IoT and state-of-the-art techniques based on ML and DL
2017	[۲۸]	Threats and vulnerabilities in IoT applications, architecture and possible attacks	Sec. III	Enhanced threat landscape, requirements, attacks, and their respective solutions in generic IoT security spectrum
2017	[۲۹]	IDS in IoT, detection methods, placement and validation strategies	Sec. III	Detailed coverage of security and privacy issues and state-of-the-art based on ML and DL
2019	[۳۴]	Security of IoT applications in different domains	Sec. III	Coverage of generic IoT applications with solutions, independent of particular domains
2019	[۳۰]	Current development in IoT security, challenges, simulators, and tools	Sec. III and V	In-depth and more detailed survey of the security requirements, threats, attacks, and solutions in IoT
2016	[۳۵]	Secure routing protocols in IoT	N/A	Focus on the applications security and privacy
2018	[۳۱]	Security challenges in IoT and sensor networks	Sec. III	Detailed security challenges, attacks, and solutions in IoT
2017	[۳۶]	Trust models for service management in IoT	N/A	Coverage of different aspects of security with solutions based on ML and DL
2017	[۳۷]	Communication standards in IoT	N/A	ML- and DL-based generic security solutions in IoT
2017	[۲۲]	System architecture, and security, privacy in edge-/fog-based IoT	Sec. III	Enhanced coverage of state-of-the-art ML- and DL-based security and privacy in generic IoT
2017	[۳۸]	Health-care communications standards in IoT	N/A	Focus on generic IoT and in-depth coverage of the security solutions
2018	[۳۹]	Architecture, scheduling, network technologies, and power management of IoT operating systems	N/A	Focus on the current solutions for generic IoT applications and future research directions
2018	[۴۰]	Open issues and challenges in IoT and enabling technologies	Sec. VI	Enhanced state-of-the-art and research challenges in ML-driven IoT security
2018	[۴۱]	IoT data analytics through DL	N/A	In-depth review of ML- and DL-based security solutions in IoT
2019	[۴۲]	data fusion in IoT applications through ML	N/A	In-depth coverage of ML and DL in different aspects security in IoT

جدول ۲، انواع مختلفی از تکنیک های یادگیری ماشین

Learning type	Model building	Examples
Supervised	Algorithms or models learn from labeled data (task-driven approach)	Classification, regression
Unsupervised	Algorithms or models learn from unlabeled data (Data-Driven Approach)	Clustering, associations, dimensionality reduction
Semi-supervised	Models are built using combined data (labeled + unlabeled)	Classification, clustering
Reinforcement	Models are based on reward or penalty (environment-driven approach)	Classification, control

برای تجزیه و تحلیل این داده ها در یک حوزه مشکل خاص، و استخراج بینش یا دانش مفید از داده ها برای ساخت برنامه های هوشمند در دنیای واقعی، انواع مختلفی از تکنیک های یادگیری ماشین را می توان با توجه به توانایی یادگیری آنها استفاده کرد، که مورد بحث قرار گرفته است در ادامه الگوریتم های یادگیری ماشین به طور عمده به چهار دسته تقسیم می شوند: یادگیری تحت نظارت، یادگیری بدون نظارت، یادگیری نیمه نظارت شده و یادگیری تقویت، همانطور که در شکل نشان داده شده است. نظارت شده: یادگیری نظارت شده معمولاً وظیفه یادگیری ماشین است تا یاد بگیرد تابعی که ورودی را به یک خروجی براساس نمونه جفت ورودی-خروجی ترسیم کند. پیش بینی برچسب کلاس یا احساسات یک قطعه از متن، مانند یک توییت یا یک بررسی محصول، به عنوان مثال، طبقه بندی متن، نمونه ای از یادگیری نظارت شده است. نیمه نظارت شده: یادگیری نیمه نظارت شده را می توان ترکیبی از روشهای نظارت شده و بدون نظارت فوق الذکر تعریف کرد، زیرا این روش هم بر روی داده های دارای برچسب و هم بدون برچسب عمل می کند [۲، ۱۴]. بنابراین، این بین یادگیری "بدون نظارت" و یادگیری "با نظارت" قرار می گیرد. برخی از زمینه های کاربردی که از یادگیری نیمه نظارت استفاده می شود شامل ترجمه ماشینی، کشف تقلب، برچسب گذاری داده ها و طبقه بندی متن است. در جدول ۲، انواع مختلفی از تکنیک های یادگیری ماشین را با مثال جمع بندی می کنیم.

استقرار فراگیر تعداد زیادی دستگاه باعث افزایش سطح حمله در سیستم اینترنت اشیا می شود. از آنجا که دستگاه های اینترنت اشیا- محدود به منابع هستند، استفاده از مکانیزم های پیچیده امنیتی در برابر حملات معروف امکان پذیر نیست.

در حملات فیزیکی، مهاجمان به دستگاه ها دسترسی مستقیم دارند و جنبه های مختلف دستگاه ها را دستکاری می کنند.

برای دستیابی به دستگاههای فیزیکی، مهندسی اجتماعی یکی از برجسته ترین روشهایی است که مهاجمان به دستگاهها دسترسی پیدا می کنند و حمله واقعی را انجام می دهند که از آسیب فیزیکی به دستگاه گرفته تا شنود، کانالهای جانبی و سایر حملات مربوطه را شامل می شود.

علیرغم این واقعیت که از فناوری های مختلفی در لایه فیزیکی اینترنت اشیا استفاده می شود، ماهیت حملات فیزیکی بیشتر شبیه و نیازمند رویکردهای مهندسی اجتماعی است.

برای شروع حملات فیزیکی، مهاجمان باید در مجاورت دستگاهها / سخت افزار با اهداف مختلف مانند از بین بردن فیزیکی سخت افزار، محدود کردن طول عمر آن، به خطر انداختن مکانیسم ارتباطی، دستکاری در منبع انرژی و غیره باشند. تزریق گره مخرب به شبکه همچنین می تواند باعث حمله انسان در وسط شود که به آن مهاجم امکان می دهد امتیازات را افزایش داده و حملات دیگری را نیز انجام دهد.همانطور که گفته شد، مهاجمان همچنین از روشهای مختلف مهندسی اجتماعی برای دسترسی فیزیکی به سخت افزار / دستگاهها برای اهداف مختلف مانند حملاتی که قبلاً ذکر کردیم، استفاده می کنند.حملات مهندسی اجتماعی مربوط به محیط فیزیکی شبکه ها است و کاهش آنها دشوار است. با این حال، آگاهی بهتر و مکانیسم های کنترل دقیق دسترسی به کاهش چنین حملاتی کمک می کند [۴۶و۴۷].

۳- خلاصه وظایف و الگوریتم های یادگیری ماشین

در این بخش، ما الگوریتم های مختلف یادگیری ماشین را مورد بحث قرار می دهیم که شامل تجزیه و تحلیل طبقه بندی، تحلیل رگرسیون، خوشه بندی داده ها، یادگیری قاعده ارتباط، مهندسی ویژگی برای کاهش ابعاد و همچنین روش های یادگیری عمیق است.

تجزیه و تحلیل طبقه بندی طبقه بندی به عنوان یک روش یادگیری نظارت شده در یادگیری ماشین در نظر گرفته می شود، که به یک مسئله مدل سازی پیش بینی نیز اشاره دارد، جایی که برچسب کلاس برای یک مثال داده شده پیش بینی شده است.

در ادامه، ما مشکلات معمول طبقه بندی را خلاصه می کنیم.

طبقه بندی چند کلاسه: به طور سنتی، این به آن دسته از وظایف طبقه بندی گفته می شود که بیش از دو برچسب کلاس دارند.

طبقه بندی چند کلاسه برخلاف وظایف طبقه بندی باینری، اصل نتایج عادی و غیر عادی را ندارد.طبقه بندی چند برچسب: در یادگیری ماشینی، طبقه بندی چند برچسب در مواردی که یک مثال با چندین کلاس یا برچسب در ارتباط است، یک نکته مهم است.بنابراین، این یک تعمیم از طبقه بندی چند کلاسه است، جایی که کلاسهای درگیر در این مسئله از نظر سلسله مراتبی ساخته شده اند و هر مثال ممکن است همزمان در بیش از یک کلاس در هر سطح سلسله مراتبی، به عنوان مثال، طبقه بندی متن چند سطحی باشد.الگوریتم های طبقه بندی بسیاری در ادبیات ماشین آلات و علوم داده داده شده است [۲، ۹،۱۶]. در ادامه، ما متداول ترین و رایج ترین روش هایی را که به طور گسترده در مناطق مختلف کاربرد استفاده می شود، جمع بندی می کنیم.

۳.۱ حملات فیزیکی

در حملات فیزیکی، مهاجمان به دستگاه ها دسترسی مستقیم دارند و جنبه های مختلف دستگاه ها را دستکاری می کنند.

برای دستیابی به دستگاههای فیزیکی، مهندسی اجتماعی یکی از برجسته ترین روشهایی است که مهاجمان به دستگاهها دسترسی پیدا می کنند و حمله واقعی را انجام می دهند که از آسیب فیزیکی به دستگاه گرفته تا شنود، کانالهای جانبی و سایر حملات مربوطه را شامل می شود.

علیرغم این واقعیت که از فناوری های مختلفی در لایه فیزیکی اینترنت اشیا استفاده می شود، ماهیت حملات فیزیکی بیشتر شبیه و نیازمند رویکردهای مهندسی اجتماعی است.

برای شروع حملات فیزیکی، مهاجمان باید در مجاورت دستگاهها / سخت افزار با اهداف مختلف مانند از بین بردن فیزیکی سخت افزار، محدود کردن طول عمر آن، به خطر انداختن مکانیسم ارتباطی، دستکاری در منبع انرژی و غیره باشند. تزریق گره مخرب به شبکه همچنین می تواند باعث حمله انسان در وسط شود که به آن مهاجم امکان می دهد امتیازات را افزایش داده و حملات دیگری را نیز انجام دهد.همانطور که گفته شد، مهاجمان همچنین از روشهای مختلف مهندسی اجتماعی برای دسترسی فیزیکی به سخت افزار / دستگاهها برای اهداف مختلف مانند حملاتی که قبلاً ذکر کردیم، استفاده می کنند.حملات مهندسی اجتماعی مربوط به محیط فیزیکی شبکه ها است و کاهش آنها دشوار است. با این حال، آگاهی بهتر و مکانیسم های کنترل دقیق دسترسی به کاهش چنین حملاتی کمک می کند. مهاجمان می توانند با تنظیمات این گره ها، آنها را همیشه بیدار نگه

دارند تا باتری تخلیه شود. شایان ذکر است که حتی برای تشخیص نفوذ در لایه های فوقانی، به عنوان مثال حمله مسیریاب، می توان بردارهای حمله زیادی را در برد. در لایه پیوند، الزامات امنیتی مختلف مانند محرمانه بودن، اصالت داده و یکپارچگی، امنیت معنایی و امنیت در برابر حملات مختلف مانند حملات پخش مجدد و کنترل دسترسی توسط IEEE 802.15.4 پشتیبانی می شود.

۳.۲ تجزیه و تحلیل Linear Discriminant

تجزیه و تحلیل Linear Discriminant، تجزیه کننده مرز تصمیم گیری خطی است که با تطبیق تراکم های شرطی کلاس بر داده ها و اعمال قانون بیز ایجاد می شود [۳، ۹]. این روش همچنین به عنوان تعمیم متمایز کننده خطی فیشر شناخته می شود، که مجموعه داده داده شده را در فضای بعدی قرار می دهد، یعنی کاهش ابعادی که پیچیدگی مدل را به حداقل می رساند یا هزینه های محاسباتی مدل حاصل را کاهش می دهد. LDA از نزدیک با ANOVA و تجزیه و تحلیل رگرسیون، که به دنبال بیان یک متغیر وابسته به عنوان ترکیبی خطی از سایر ویژگی ها یا اندازه گیری ها است، ارتباط دارد. برخی معیار های رایج در تجزیه تحلیل مدل ها می تواند به شرح زیر باشد.

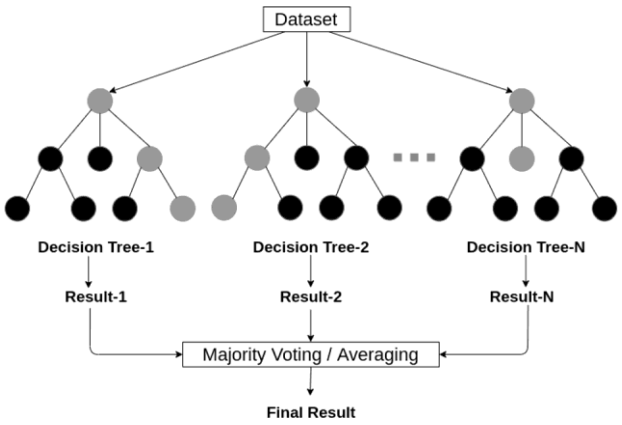
رگرسیون لجستیک: مدل آماری رایج مبتنی بر احتمال دیگر که برای حل مسائل طبقه بندی در یادگیری ماشین استفاده می شود، رگرسیون لجستیک است. رگرسیون لجستیک به طور معمول از یک تابع لجستیک برای تخمین احتمالات استفاده می کند، که در معادله معادل تابع سیگموئید تعریف شده ریاضی نیز گفته می شود.

K- نزدیکترین همسایگان: K-Nearest Neighbours یک "یادگیری مبتنی بر نمونه" یا یادگیری غیرتولیدی است که به عنوان الگوریتم "یادگیری تنبل" نیز شناخته می شود. بر اساس معیارهای تشابه، نقاط داده جدید را طبقه بندی می کند.

ماشین بردار پشتیبان: در یادگیری ماشین، یکی دیگر از تکنیک های رایج که می تواند برای طبقه بندی، رگرسیون یا سایر کارها استفاده شود، ماشین بردار پشتیبانی است. به طور مستقیم، هواپیمای فوق العاده، که بیشترین فاصله را از نزدیکترین نقاط داده آموزش در هر کلاس دارد، به یک جدایی شدید دست می یابد زیرا، به طور کلی، هرچه حاشیه بیشتر باشد، خطای تعمیم طبقه بندی کننده کاهش می یابد.

درخت تصمیم: درخت تصمیم یک روش یادگیری نظارت شده غیر پارامتری شناخته شده است. روش های یادگیری DT هم برای طبقه بندی و هم برای کارهای رگرسیون استفاده می شود. نمونه ها با بررسی ویژگی تعریف شده توسط آن گره، از گره ریشه درخت شروع می شوند و سپس به شاخه درخت مربوط به مقدار صفت منتقل می شوند.

جنگل تصادفی: طبقه بندی جنگل تصادفی به عنوان یک تکنیک طبقه بندی گروهی شناخته شده است که در زمینه یادگیری ماشین و علم داده در مناطق مختلف کاربرد استفاده می شود. این روش از مجموعه موازی استفاده می کند که متناسب با چندین طبقه بندی کننده درخت تصمیم به طور موازی است، همانطور که در شکل ۱ نشان داده شده است. در زیر نمونه های مختلف مجموعه داده و از رای گیری اکثریت یا میانگین ها برای نتیجه یا نتیجه نهایی استفاده می کند. بنابراین مشکل بیش از حد مناسب را به حداقل می رساند و دقت و کنترل پیش بینی را افزایش می دهد. مدل یادگیری RF با چندین درخت تصمیم معمولاً دقیق تر از یک مدل مبتنی بر درخت تصمیم واحد است. برای ساختن یک سری درخت تصمیم با تنوع کنترل شده، این ترکیب ترکیبی از بوت استرپ و انتخاب ویژگی تصادفی است. هم با مشکلات طبقه بندی و هم رگرسیون سازگار است و برای مقادیر مقوله ای و پیوسته به خوبی متناسب است.



شکل ۱ چندین طبقه بندی کننده درخت تصمیم به طور موازی

افزایش گرادیان شدید:

افزایش گرادیان، مانند جنگل های تصادفی در بالا، یک الگوریتم یادگیری گروه است که مدل نهایی را بر اساس مجموعه ای از مدل های منفرد، به طور معمول درختان تصمیم، تولید می کند. از شیب برای به حداقل رساندن عملکرد از دست دادن استفاده می شود، مشابه نحوه استفاده شبکه های عصبی از نزول شیب برای بهینه سازی وزن ها. افزایش گرادیان شدید نوعی تقویت گرادیان است که هنگام تعیین بهترین مدل تقریب های دقیق تری را در نظر می گیرد. این شیب های مرتبه دوم عملکرد ضرر را محاسبه می کند تا از دست دادن و تنظیم پیشرفته را به حداقل برساند، که باعث برازش بیش از حد می شود و تعمیم و عملکرد مدل را بهبود می بخشد. نزول شیب تصادفی: نزول شیب تصادفی یک روش تکرار شونده برای بهینه سازی عملکرد هدف با ویژگی های صافی مناسب است، جایی که کلمه "تصادفی" به احتمال تصادفی اشاره دارد. گرادیان شیب تابعی است که درجه تغییر یک متغیر را در پاسخ به تغییرات متغیر دیگر محاسبه می کند. بگذارید، میزان یادگیری است و Ji هزینه هزینه آموزش به عنوان مثال، سپس معادله است. نشان دهنده روش به روزرسانی وزن نزولی شیب تصادفی در تکرار j است.

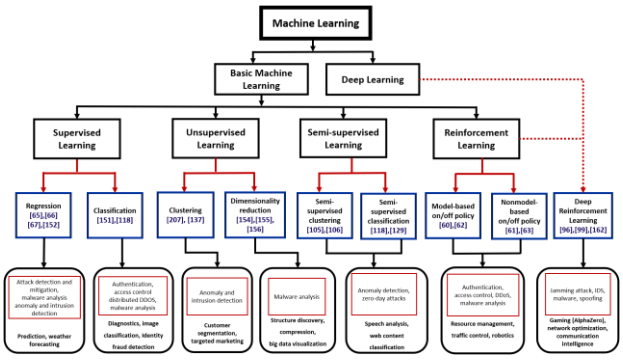
طبقه بندی مبتنی بر قانون: از اصطلاح طبقه بندی مبتنی بر قاعده می توان برای هر طرح طبقه بندی استفاده کرد که از قوانین IF-THEN برای پیش بینی کلاس استفاده می کند. چندین الگوریتم طبقه بندی مانند Zero-R، One-R، درختان تصمیم [۱۰، ۱۱]، DTNB، زبان آموز Ripple Down Rule، هرس تکراری تکراری برای تولید خطای کاهش با توانایی تولید قانون وجود دارد. درخت تصمیم یکی از متداول ترین الگوریتم های طبقه بندی مبتنی بر قاعده در میان این تکنیک ها است زیرا دارای چندین مزیت است، مانند تفسیر آسان تر. توانایی مدیریت داده های با ابعاد بالا ؛ سادگی و سرعت دقت خوب و قابلیت تولید قوانینی برای طبقه بندی روشن و قابل درک انسان. قوانین مبتنی بر درخت تصمیم گیری همچنین دقت قابل توجهی را در یک مدل پیش بینی برای موارد آزمایش غیب فراهم می کند. از آنجا که قوانین به راحتی قابل تفسیر هستند، این طبقه بندی های مبتنی بر قاعده اغلب برای تولید مدلهای توصیفی استفاده می شوند که می توانند سیستمی شامل موجودیتها و روابط آنها را توصیف کنند.

۴- یادگیری عمیق

مکانیزم Deep Learning یک روش یادگیری ماشین است که از ANN گرفته شده است. شبکه عصبی از سلولهای عصبی متصل شده از طریق اتصالات وزنی تشکیل شده است. برای دستیابی به مجموعه خروجی مطلوب، روش یادگیری نظارت شده یا نظارت نشده با شبکه مرتبط است. یادگیری با استفاده از داده های دارای برچسب و بدون برچسب از تکنیک های یادگیری تحت نظارت یا بدون نظارت انجام می شود، به ترتیب با تنظیم تکرار وزن در هر جفت نورون دنبال می شود. (شکل ۲)

مکانیزم Deep Learning به دلیل محاسبه توزیع شده و، یادگیری و تجزیه و تحلیل مقدار زیادی از داده های بدون برچسب، طبقه بندی نشده و بدون نظارت شناخته شده است. این یک مدل سلسله مراتبی از یادگیری و نمایش ویژگی های ایجاد شده توسط فرآیند لایه لایه در مغز انسان را توسعه می دهد. مدل های DL با ارائه مدل سازی طبقه بندی بهتر و تولید نمونه داده های بهتر، به برنامه های مختلف ML مانند تشخیص گفتار، بینایی رایانه و NLP کمک می کنند.

CNN و RNN پرکاربردترین معماری یادگیری عمیق هستند.



شکل ۲ مکانیزم Deep Learning

۴.۱ یادگیری تقویت عمیق:

DL یکی از انواع تکنیک های ML است که برای تقریب، طبقه بندی و پیش بینی عملکرد استفاده می شود در حالی که RL نوع دیگری از تکنیک های ML است که برای تصمیم گیری استفاده می شود که در آن یک عامل نرم افزار با تعامل با یک محیط بیش از موارد مختلف، در مورد اقدامات بهینه یاد می گیرد. ایالت ها.

با ترکیب DL و RL، نمایندگان می توانند به تنهایی یاد بگیرند و سیاست خوبی برای دستیابی به حداکثر پاداش های بلند مدت ارائه دهند. در این روش، RL از DL برای یافتن بهترین سیاست کمک می گیرد و DL تقریب مقادیر عمل را برای یافتن کیفیت یک عمل در یک حالت معین انجام می دهد. DL قادر به یادگیری از الگوهای پیچیده است اما مستعد طبقه بندی نادرست است.

در این شرایط، RL دارای یک توانایی قدرتمند برای یادگیری خودکار از محیط و بدون هیچ گونه ساخت ویژگی است و به DL در طبقه بندی کارآمد کمک می کند. DQN یکی از نمونه های چنین ادغامی است که در آن یادگیری Q با یک NN عمیق ترکیب می شود. هنگامی که داده های با ابعاد بالا در ورودی ها ارائه می شود، نماینده DQN سیاست ها را با استفاده از RL می آموزد، در حالی که از CNN عمیق برای تقریب عملکرد و مقدار استفاده می شود. به طور رسمی، ادغام هر دو به عنوان DRL شناخته می شود که اساساً ترکیبی از RL و DL است. DL ویژگی هایی را از داده های آموزش استخراج می کند و در RL، یک عامل اقدامات متقابل با یک محیط را انجام می دهد و سعی می کند پاداش تجمعی را به حداکثر برساند.

تکنیک های یادگیری ماشین مورد استفاده در امنیت اینترنت اشیا در ادامه، ما در مورد الگوریتم های مختلف ML با تمرکز بر مشکلات اساسی امنیت و حریم خصوصی در شبکه های اینترنت اشیا بحث می کنیم، همانطور که در جدول IV نشان داده شده است. دقیق تر، احراز هویت، شناسایی و کاهش حمله، حملات توزیع نشده خدمات، تشخیص ناهنجاری و نفوذ و تجزیه و تحلیل بدافزار را در نظر می گیریم.

الگوریتم های یادگیری تحت نظارت با داده های دارای برچسب کار می کنند و در شبکه های اینترنت اشیا برای سنجش طیف، برآورد کانال، فیلتر کردن تطبیقی ، مشکلات امنیتی و محلی سازی استفاده می شوند. SVM، Naive Bayes، Random Forest، درخت تصمیم، الگوریتم های طبقه بندی گسترده ای هستند که به طور گسترده مورد استفاده قرار می گیرند.

این الگوریتم ها با عنوان "Instance-based" نیز شناخته می شوند که با جستجوی مشابه ترین داده های آموزشی، برای هر مشاهده جدید پیش بینی می کنند. خانواده الگوریتم های یادگیری بدون نظارت با داده های بدون برچسب سروکار دارند و از داده های ورودی به روشی ابتکاری استفاده می کنند. الگوریتم تحت نظارت ML مانند SVM، DT و Naive Bayes نیز در امنیت اینترنت اشیا استفاده می شود. با در نظر گرفتن الگوریتم های ML بدون نظارت، خوشه بندی K-means و سلسله مراتبی دو الگوریتم خوشه بندی محبوب هستند.

خوشه بندی K-mean بیشترین محبوبیت را دارد زیرا یک الگوریتم ساده و انعطاف پذیر است که خوشه ها را براساس فاصله هندسی بین نقاط داده تشکیل می دهد. به دلیل محدودیت های منابع، دستگاه های اینترنت اشیا may ممکن است قادر به میزبانی یا اجرای الگوریتم های محاسباتی پیچیده برای هر نوع کار مانند ارتباطات، تجزیه و تحلیل و پیش بینی نباشند.

الگوریتم های اصلی DRL برای امنیت و حفظ حریم خصوصی مورد استفاده قرار می گیرند: گرادیان عمیق سیاست تعیین گرایانه، DQN مداوم، تکرار تجربه اولویت بندی شده، یادگیری N- مرحله ای ناهمزمان، SARSA عمیق و شبکه دوئل DQN ..

محدودیت های کاربردی یادگیری ماشین در شبکه های اینترنت اشیا بسیاری از تکنیک های سنتی ML ذاتاً کارآمد و مقیاس پذیر نیستند تا بتوانند داده های اینترنت اشیا را مدیریت کنند و بنابراین به تغییرات قابل توجهی نیاز دارند. ۲) تجزیه و تحلیل داده های ناهمگن: داده های بی سیم می توانند از منابع مختلفی از جمله سیستم های اطلاعاتی شبکه ای و دستگاه های سنجش و ارتباط تولید شوند. داده های تولید شده در شبکه های اینترنت اشیا دارای ماهیت متنوعی با انواع مختلف، قالب ها و معناشناسی هستند، بنابراین ناهمگنی نحوی و معنایی را به نمایش می گذارند. ناهمگنی نحوی به تنوع در انواع داده ها، قالب های پرونده، طرح های رمزگذاری و مدل های داده اشاره دارد. برنامه اینترنت اشیا می تواند ترکیبی از داده های ساخت یافته، نیمه ساختار یافته یا غیر ساختاری را ایجاد کند. اگر داده های برچسب گذاری شده داشته باشیم، می توان از الگوریتم های طبقه بندی استفاده کرد و اگر داده ها بدون برچسب هستند، می توان از الگوریتم خوشه بندی برای گروه بندی و تجمیع داده های موجود استفاده کرد. بحث قبلی، به طور مساوی، برای توابع مربوط به امنیت در اینترنت اشیا

where که در آن داده های زمان واقعی برای بردارهای احتمالی حمله مانند نفوذ پردازش می شوند، قابل اجرا است.

یک روش DL مبتنی بر RNN برای تکنیک تجزیه و تحلیل بدافزار در IoT پیشنهاد کرده است. نویسندگان برنامه های مبتنی بر Advanced RISC Machines را در اینترنت اشیا در نظر گرفتند. نویسندگان مدل های خود را با مجموعه داده های مختلف بدافزار موجود آموزش می دهند و سپس چارچوب خود را با بدافزار جدید آزمایش می کنند.

دقیق تر، تکنیک DL مبتنی بر RNN به ٪۹۸ دقت در شناسایی بدافزار جدید در داخل برنامه IoT دست یافت. نویسندگان با استفاده از روش های یادگیری فضای ویژه و شبکه های کانولوشن عمیق، طبقه بندی بدافزار را در برنامه های IoT سازگار با ARM انجام دادند. در تجزیه و تحلیل خود، نویسندگان از روش Class-Wise Information Gain برای انتخاب ویژگی ها که در آن نمونه های نرم افزارهای مخرب و بدافزار برای آموزش انتخاب شده بودند، استفاده کردند.

نویسندگان گزارش کردند که طبقه بندی های استفاده شده به دقت ٪۹۹.۶۸ در شناسایی بدافزار با دقت ٪۹۸.۵۹ و فراخوان ٪۹۸.۳۷ دست پیدا می کنند. چارچوب شناسایی براساس ANN است و با هر دو نرم افزار خوش خیم و بدافزار در سیستم عامل اندروید آزمایش شده است. نتایج آزمایشی با خانواده صحیح بدافزار به میزان قابل توجهی از ٪۹۶ تا ٪۹۸ بدافزار اندروید دست یافت. مکانیسم شناسایی بدافزار DDoS مبتنی بر تشخیص تصویر در شبکه های اینترنت اشیا را پیشنهاد داد. در این راه حل، نویسندگان ابتدا دو خانواده مهم بدافزار، یعنی Mirai و Linux را جمع آوری و طبقه بندی می کنند. پس از آن CNN در مقیاس کوچک برای طبقه بندی تصاویر در نرم افزارهای مخرب و بدافزار اعمال می شود.

نتایج تجربی گزارش شده توسط نویسندگان نشان می دهد که روش CNN-based به ٪۹۴ دقت در طبقه بندی بدافزارها و بدافزارهای DDoS دست می یابد در حالی که به ٪۸۱.۸ دقت در شناسایی دو خانواده بدافزار فوق الذکر دست می یابد. از رمزگذارهای خودکار عمیق برای شناسایی حملات Botnet در اینترنت اشیا استفاده می کند. در این راه حل ها، نویسندگان رفتار شبکه را استخراج کرده و سپس با استفاده از رمزگذارهای خودکار عمیق، رفتار ناهنجار شبکه را جدا می کنند. از شبکه های Deep Q با تکنیک یادگیری Q برای پرداختن به مسائل امنیتی مانند احراز هویت، کنترل دسترسی و تجزیه و تحلیل بدافزار در برنامه های مراقبت های بهداشتی شبکه های اینترنت اشیا استفاده می شود.

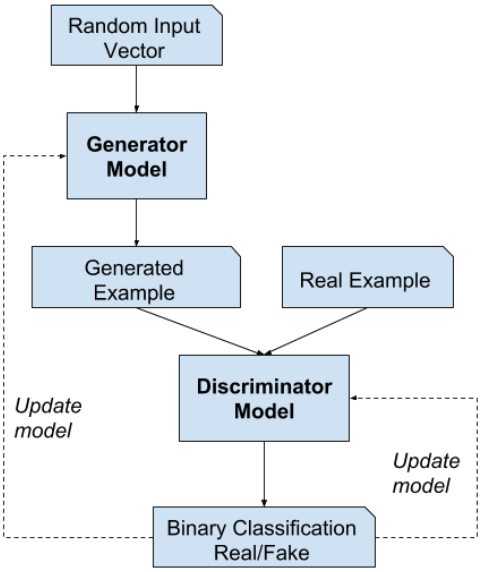
نویسندگان از روش یادگیری Q برای تجزیه و تحلیل داده های بیماران از طریق شبکه های لایه لایه Deep Q برای تأیید اعتبار و تجزیه و تحلیل بدافزار استفاده کردند. نویسندگان گزارش دادند که شبکه های Deep Q در مقایسه با MLP و آموزش مقادیر برداری، انرژی کمتری مصرف می کنند.[۴۸،۴۹،۵۰،۵۱].

۴.۲ امنیت اینترنت اشیا-مبتنی بر شبکه خصمانه تولیدی

GAN ها با استفاده از آنها در یک زمینه خصمانه انقلابی در روش های ML ایجاد کرده اند. در زمینه امنیت اینترنت اشیا، GAN ها نتایج قابل توجهی را برای حل مشکلات امنیتی مختلف در شبکه های اینترنت اشیا ارائه داده اند. همانطور که قبلاً ذکر شد، تشخیص نفوذ برای شبکه های اینترنت اشیا بسیار مهم است و ثابت شده است که GAN ها در شناسایی نفوذ در اینترنت اشیا موثر هستند. بر خلاف GAN های سنتی، نویسندگان GAN های توزیع شده را برای جلوگیری از هزینه های ارتباطی GAN متمرکز توزیع کردند. علاوه بر این، در صورت GAN متمرکز، دسترسی به تمام نقاط داده از دستگاه های اینترنت اشیا باید داده شود. در حالی که در GAN توزیع شده، داده ها توسط دستگاه های اینترنت اشیا به اشتراک گذاشته نمی شوند، بلکه وزن ها در میان گره های اینترنت اشیا به اشتراک گذاشته می شوند.

نویسندگان گزارش دادند که GAN های توزیع شده می توانند حدود ٪۲۰ دقت تشخیص بالاتر، ٪۲۵ دقت بالاتر و ٪۶۰ مثبت کاذب پایین تر در مقایسه با GAN متمرکز داشته باشند. به همین ترتیب، از طریق اینتراتور و همکاران استدلال می شود که GAN ها لزوماً نمونه های معتبر جدیدی تولید نمی کنند و بنابراین، ممکن است از قدرت در تشخیص ناهنجاری برخوردار نباشند. نویسندگان یک GAN چند تبعیض را پیشنهاد کردند که از دو تبعیض استفاده می کند که یکی شبکه متراکمی برای اطمینان از کیفیت نمونه ها است و دیگری رمزگذار خودکار برای تشخیص نفوذ است. همچنین در نتایج گزارش شده است که خود رمزگذار GAN عملکرد خود رمزگذار پایه را بهتر می کند. در این کار، نویسندگان از طبقه بندی کننده برای پیش بینی صحیح نمونه ها استفاده می کنند و سپس یک پیکربندی خصمانه دیگر، پیش بینی را که یک مدل معمولی GAN است، سخت تر می کند. نویسندگان گزارش دادند که آنها بیش از ٪۸۰ دقت به دست آوردند که کمتر از GAN های توزیع شده است.

جدا از تشخیص نفوذ، GAN ها همچنین برای تجزیه و تحلیل بدافزار و شناسایی حمله (DDoS) استفاده شده اند.



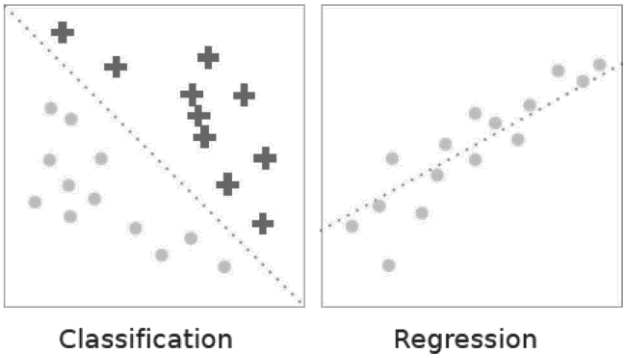
شکل ۳. معماری کلی شبکه Generate Adversarial (GAN)

۵- رگرسیون خطی و پردازش تصویری

تجزیه و تحلیل رگرسیون شامل چندین روش یادگیری ماشین است که امکان پیش بینی متغیر نتیجه مداوم را بر اساس مقدار یک یا چند متغیر پیش بینی کننده فراهم می کند. مهمترین تمایز بین طبقه بندی و رگرسیون این است که طبقه بندی برچسبهای کلاس متمایز را پیش بینی می کند، در حالی که رگرسیون پیش بینی یک مقدار مداوم را تسهیل می کند. شکل ۶ نمونه ای از چگونگی تقسیم بندی را با مدلهای رگرسیون متفاوت نشان می دهد. مدل های رگرسیون اکنون در زمینه های مختلف از جمله پیش بینی یا

پیش بینی مالی، برآورد هزینه، تجزیه و تحلیل روند، بازاریابی، برآورد سری های زمانی، مدل سازی پاسخ دارو و بسیاری موارد دیگر به طور گسترده مورد استفاده قرار می گیرند. برخی از انواع آشنای الگوریتم های رگرسیون رگرسیون خطی، چند جمله ای، رگرسیون لاسو و ریج و غیره هستند. رگرسیون خطی رابطه ای را بین متغیر وابسته و یک یا چند متغیر مستقل با استفاده از بهترین خط مستقیم مناسب ایجاد می کند.

$$y = a + bx + e$$
$$y = a + b1x1 + b2x2 + \cdots + bnxn + e,$$



شکل ۴ طبقه بندی در مقابل رگرسیون در طبقه بندی خط نقطه ای نشان دهنده یک مرز خطی است که دو کلاس را از هم جدا می کند. در بازگشت، خط نقطه ای رابطه خطی بین دو متغیر را مدل می کند.

جایی که a رهگیری است، b شیب خط است و e اصطلاح خطا است از این معادله می توان برای پیش بینی استفاده کرد مقدار متغیر هدف بر اساس متغیر پیش بینی شده (ها). رگرسیون خطی چندگانه یک توسعه از رگرسیون خطی ساده است که به دو یا چند متغیر پیش بینی کننده اجازه می دهد تا یک متغیر پاسخ، y را به عنوان یک تابع خطی [۲] تعریف شده در معادله، مدل کنند. در حالی که رگرسیون خطی ساده فقط ۱ متغیر مستقل دارد که در معادله تعریف شده است.

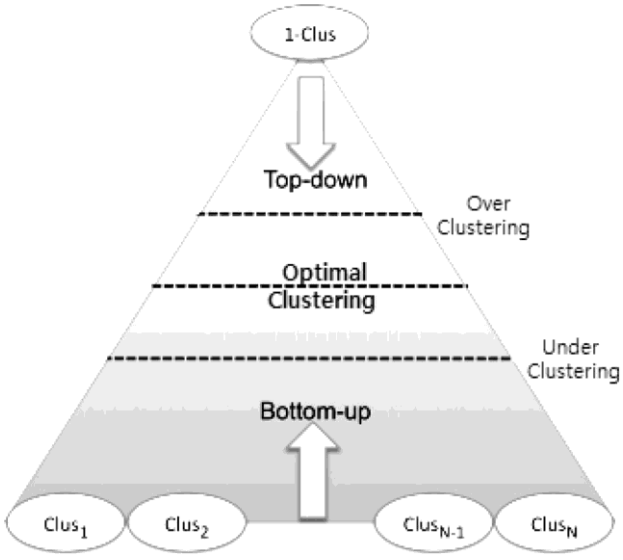
رگرسیون چند جمله ای: رگرسیون چند جمله ای شکلی از تحلیل رگرسیون است که در آن رابطه بین متغیر مستقل x و متغیر وابسته y خطی نیست، بلکه درجه چند جمله ای n n در x است [۹]. معادله رگرسیون چند جمله ای نیز از رگرسیون خطی (رگرسیون چند جمله ای درجه) حاصل شده است ۱) معادله، که به شرح زیر تعریف شده است:

$$y = b$$
$$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots + b_n x^n + e.$$

در اینجا y خروجی پیش بینی شده / هدف است، b0، b1، bn ... ضرایب رگرسیون هستند، x یک متغیر مستقل / ورودی است. به عبارت ساده، می توان گفت که اگر داده ها به صورت خطی توزیع نشوند، در عوض درجه نهم از چند جمله ای است، بنابراین برای به دست آوردن خروجی مطلوب، از رگرسیون چند جمله ای استفاده می کنیم. رگرسیون LASSO و ridge: رگرسیون LASSO و Ridge به دلیل توانایی جلوگیری از جایگزینی بیش از حد و کاهش پیچیدگی مدل، به عنوان تکنیک های قدرتمندی شناخته می شوند که معمولاً برای ساخت مدل های یادگیری در حضور تعداد زیادی ویژگی استفاده می شوند. . مدل رگرسیون LASSO از تکنیک تنظیم L1 استفاده می کند که از جمع شدگی استفاده می کند، که "مقدار مطلق اندازه ضرایب" را مجازات می کند. از طرف دیگر، رگرسیون خط الراس از قاعده گذاری L2 استفاده می کند، که "اندازه مربعات ضرایب" است. بنابراین، رگرسیون خط الراس وزنه ها را مجبور به کوچک بودن می کند اما هرگز مقدار ضریب را روی صفر قرار نمی دهد و یک راه حل غیر پراکنده انجام می دهد. به طور کلی، رگرسیون LASSO برای بدست آوردن زیرمجموعه ای از پیش بینی ها با از بین بردن ویژگی های کم اهمیت بسیار مفید است و رگرسیون خط الراس برای زمانی مفید است که مجموعه داده دارای "چند خطی بودن" باشد که به پیش بینی کننده های مرتبط با سایر پیش بینی ها اشاره دارد. کاهش یافته توسط:٪ شخصیت ها:

آنالیز خوشه ای

تجزیه و تحلیل خوشه، همچنین به عنوان خوشه بندی شناخته می شود، یک روش یادگیری ماشین بدون نظارت برای شناسایی و گروه بندی نقاط داده مرتبط در مجموعه داده های بزرگ بدون نگرانی برای نتیجه خاص است. این مجموعه ای از اشیا را به گونه ای گروه بندی می کند که اشیا in در همان رده، خوشه نامیده می شوند، از جهتی شباهت بیشتری به یکدیگر از اشیا in موجود در گروههای دیگر [۲]. این روش اغلب به عنوان یک روش تجزیه و تحلیل داده ها برای کشف روندها یا الگوهای جالب در داده ها، به عنوان مثال، گروه های مصرف کننده بر اساس رفتار آنها استفاده می شود. در طیف گسترده ای از حوزه های کاربردی، مانند امنیت سایبری، تجارت الکترونیکی، پردازش داده های تلفن همراه، تجزیه و تحلیل سلامت، مدل سازی کاربر و تجزیه و تحلیل رفتاری، می توان از خوشه بندی استفاده کرد. در ادامه، انواع مختلف روشهای خوشه بندی را به طور خلاصه مورد بحث و بررسی قرار می دهیم.



شکل ۵ مدل سازی کاربر و تجزیه و تحلیل رفتاری

روش های تقسیم بندی: بر اساس ویژگی ها و شباهت های موجود در داده ها، این روش خوشه بندی داده ها را به چندین گروه یا خوشه دسته بندی می کند. متداول ترین الگوریتم های خوشه بندی مبتنی بر روش های پارتیشن بندی K-means، K-Mediods، CLARA و غیره است.

روش های مبتنی بر تراکم: برای شناسایی گروه ها یا خوشه های مجزا، از این مفهوم استفاده می شود که یک خوشه در فضای داده یک منطقه مجاور از تراکم نقطه بالا است که توسط مناطق مجاور با تراکم نقطه کم از

سایر خوشه ها جدا شده است. روش های مبتنی بر تراکم معمولاً با خوشه هایی از داده های مشابه و چگالی و ابعاد بالا دست و پنجه نرم می کنند.

روش های مبتنی بر سلسله مراتب: خوشه بندی سلسله مراتبی به طور معمول به دنبال ایجاد سلسله مراتبی از خوشه ها، به عنوان مثال، ساختار درخت است. استراتژی های خوشه بندی سلسله مراتبی به طور کلی به دو نوع تقسیم می شوند: رویکرد تجمع گرا - یک روش "پایین به بالا" که در آن هر مشاهده از خوشه خود شروع می شود و جفت خوشه ها به صورت یکجا ترکیب می شوند، سلسله مراتب را بالا می برند و "از بالا به پایین" تقسیم می شوند. رویکردی که در آن همه مشاهدات در یک خوشه آغاز می شود و تقسیمات بصورت بازگشتی انجام می شود، همانطور که در شکل ۷ نشان داده شده است، به سمت پایین سلسله مراتب حرکت می کند.

روش های مبتنی بر شبکه: برای مقابله با مجموعه داده های گسترده، خوشه بندی مبتنی بر شبکه به ویژه مناسب است. برای بدست آوردن خوشه ها، اصل ابتدا خلاصه کردن مجموعه داده با نمایش شبکه و سپس ترکیب سلول های شبکه است.

خوشه بندی K-means: خوشه بندی K-mean یک الگوریتم سریع، قوی و ساده است که وقتی مجموعه داده ها به خوبی از یکدیگر جدا شوند، نتایج قابل اتکایی را ارائه می دهد. از آنجا که مقادیر افراطی می توانند به راحتی بر روی یک میانگین تأثیر بگذارند، الگوریتم خوشه بندی K-means نسبت به نقاط دور حساس است. خوشه بندی با تغییر میانگین: خوشه بندی با تغییر متوسط یک روش خوشه بندی غیرپارامتری است که نیازی به دانش قبلی در مورد تعداد خوشه ها یا محدودیت های شکل خوشه ندارد. خوشه بندی تغییر میانگین هدف این است که "Blobs" را در یک توزیع یا تراکم صاف از نمونه ها کشف کند. در موارد با ابعاد بالا، که تعداد خوشه ها به طور ناگهانی تغییر کند، الگوریتم تغییر میانگین به خوبی کار نمی کند.

روش DBSCAN : ایده اصلی DBSCAN این است که اگر یک نقطه به یک خوشه نزدیک شود، اگر به نقاط زیادی از آن خوشه نزدیک باشد. DBSCAN، برخلاف k-means، به مشخصات پیشینی تعداد خوشه ها در داده ها نیازی ندارد و می تواند خوشه های دلخواه شکل پیدا کند.

خوشه بندی سلسله مراتبی تجمعی: متداول ترین روش خوشه بندی سلسله مراتبی که برای گروه بندی اشیا in در خوشه ها بر اساس شباهت آنها استفاده می شود، خوشه بندی تجمعی است. مزیت اصلی خوشه بندی سلسله مراتبی تجمعی نسبت به k-means این است که سلسله مراتب ساختار درختی حاصل از خوشه بندی جمع کننده اطلاعاتی تر از مجموعه غیر ساختاری خوشه های مسطح است که توسط k-means برگردانده می شود، که می تواند به تصمیم گیری بهتر در مربوط کمک کند. حوزه های کاربرد.

کاهش ابعاد و یادگیری ویژگی

تمایز اصلی بین انتخاب و استخراج ویژگی ها این است که "انتخاب ویژگی" زیرمجموعه ای از ویژگی های اصلی را نگه می دارد، در حالی که "استخراج ویژگی" ویژگی های کاملا جدیدی ایجاد می کند.

انتخاب ویژگی ها: انتخاب ویژگی ها، همچنین به عنوان انتخاب متغیرها یا ویژگی ها در داده ها شناخته می شود، فرآیند انتخاب زیر مجموعه ای از ویژگی های منحصر به فرد برای استفاده در یادگیری ماشین ساخت و مدل علم داده است.

آزمون مجذور کای، آزمون تحلیل واریانس، ضریب همبستگی پیرسون، حذف خصوصیات عودی، برخی از تکنیک های معروف هستند که می توانند برای انتخاب ویژگی مورد استفاده قرار گیرند.

استخراج ویژگی: در یک مدل یا سیستم مبتنی بر یادگیری ماشین، تکنیک های استخراج ویژگی معمولاً درک بهتری از داده ها، راهی برای بهبود دقت پیش بینی و کاهش هزینه محاسباتی یا زمان آموزش را فراهم می کنند. هدف "استخراج ویژگی" [۶، ۱۲] کاهش تعداد ویژگی ها در یک مجموعه داده با تولید ویژگی های جدید از ویژگی های موجود و سپس کنار گذاشتن ویژگی های اصلی است. سپس اکثر اطلاعات موجود در مجموعه اصلی ویژگی ها را می توان با استفاده از این مجموعه کاهش یافته جدید جمع کرد. تجزیه و تحلیل مولفه های اصلی اغلب به عنوان یک روش کاهش بعد برای استخراج فضای بعدی با ایجاد اجزای نام تجاری جدید از ویژگی های موجود در یک مجموعه داده استفاده می شود.

در طیف گسترده ای از حوزه های کاربردی، مانند امنیت سایبری، تجارت الکترونیکی، پردازش داده های تلفن همراه، تجزیه و تحلیل سلامت، مدل سازی کاربر و تجزیه و تحلیل رفتاری، می توان از خوشه بندی استفاده کرد. در ادامه، انواع مختلف روشهای خوشه بندی را به طور خلاصه مورد بحث و بررسی قرار می دهیم.

روش های تقسیم بندی: براساس ویژگی ها و شباهت های موجود در داده ها، این رویکرد خوشه بندی، داده ها را به چندین گروه یا خوشه طبقه بندی می کند. دانشمندان یا تحلیل گران به طور معمول تعداد خوشه ها را بسته به ماهیت برنامه های هدف، به صورت پویا یا ایستا تعیین می کنند تا برای روش های خوشه بندی تولید کنند. متداول ترین الگوریتم های خوشه بندی مبتنی بر روش های پارتیشن بندی K-means، K-Mediods، CLARA و غیره است.

الگوریتم های خوشه بندی معمولی بر اساس تراکم DBSCAN، OPTICS و غیره هستند. روش های مبتنی بر سلسله مراتب: خوشه بندی سلسله مراتبی به طور معمول به دنبال ایجاد سلسله مراتبی از خوشه ها، به عنوان مثال، ساختار درخت است.

روش های مبتنی بر شبکه: برای مقابله با مجموعه داده های گسترده، خوشه بندی مبتنی بر شبکه به ویژه مناسب است.

برای به دست آوردن خوشه ها، اصل ابتدا خلاصه کردن مجموعه داده با نمایش شبکه و سپس ترکیب سلول های شبکه است. آیا الگوریتم های استاندارد خوشه بندی مبتنی بر شبکه هستند.

روش های مبتنی بر مدل: الگوریتم های خوشه بندی مبتنی بر مدل به طور عمده دو نوع دارند: یکی که از یادگیری آماری استفاده می کند و دیگری بر اساس روشی برای یادگیری شبکه عصبی.

روش های مبتنی بر محدودیت: خوشه بندی مبتنی بر محدودیت یک رویکرد نیمه نظارت شده برای خوشه بندی داده است که از محدودیت هایی برای ترکیب دانش دامنه استفاده می کند.

۶- چالش های تحقیق در آینده

DL - یک اندازه برای همه مناسب نیست: تکنیک های DL بسیار خاص برنامه هستند که در آن یک مدل آموزش داده شده برای حل یک مسئله ممکن است نتواند برای مسئله دیگری در حوزه مشابه عملکرد خوبی داشته باشد.

این مدل ها معمولاً باید با داده های مربوطه آموزش ببینند تا برای سایر مشکلات مشابه مورد استفاده قرار گیرند.

این ممکن است برای برخی از شبکه های ثابت مشکلی ایجاد نکند. با این حال، برای برنامه های IoT در زمان واقعی، استفاده از چنین مدل هایی دشوار است.

شبکه های عصبی جعبه های سیاه هستند: شبکه های عصبی عمیق مانند Blackbox عمل می کنند، زیرا ما نمی دانیم که چگونه هر مدل DL با دستکاری داده های ورودی با استفاده از نورون ها در لایه های به هم پیوسته پیچیده، به نتیجه می رسد.

در این راستا، ML و DL نیز مستعد این تأثیر هستند در صورتی که با ایجاد تغییر جزئی در داده های ورودی به سیستم یادگیری، تغییرات عظیمی در خروجی که مدل آموخته شده است، ایجاد می شود.

این پدیده تکنیک های ML و DL به کار رفته در اینترنت اشیا را در معرض حملات امنیتی قرار می دهد، جایی که مهاجمان به عمد داده های ورودی را تغییر می دهند تا سیستم ناپایدار شود.

با توجه به مقدار زیاد داده های تولید شده توسط دستگاه های اینترنت اشیا، پیاده سازی تکنیک های DL در دستگاه های لبه ای دشوار خواهد بود.

پایداری مدلهای DL در مواردی که اطلاعات تازه موجود بر روی مدل آموزش دیده تأثیر بگذارد نیز مهم است.

الزامات بیش از حد مناسب و پارامترهای بیش از حد: آموزش خارج از خط از ثبت داده های ثابت و یادگیری از نمونه های محدود روی سیستم واقعی، اعتبار تصمیم گیری در مدل های DL را بسیار تحت تأثیر قرار می دهد.

در واقع، کارآیی یک مدل ML براساس توانایی آن در عملکرد خوب در یک مجموعه داده جدید ارزیابی می شود و نه از طریق عملکرد آن بر روی داده های آموزشی که به آن داده می شود.

به دلیل تفاوت در توزیع آموزش و آزمون مجموعه داده ها، طبقه بندی کننده های ML معمولاً هنگام استفاده در برنامه های دنیای واقعی از کار می افتند.

به طور معمول مدل ML بر روی یک مجموعه داده آموزش خاص آموزش می بیند و نمونه های آموزش را به خاطر می سپارد، اما تعمیم آن را برای داده های جدید و شرایط جدید نمی آموزد. در نتیجه، خطاهایی در مجموعه داده های دیده نشده و در حین مجموعه داده های آموزش، به ویژه در مدل های پیچیده با پارامترهای بسیار زیاد در مقایسه با تعداد مشاهدات، رخ می دهد.

این موارد به صورت انتخابی یا تصادفی انتخاب می شوند و می توانند با حتی تغییر جزئی در این پارامترها، تغییر زیادی در عملکرد مدل ها ایجاد کنند. یادگیری تحت نظارت به دلیل مجموعه داده های ثابت پایدار در نظر گرفته می شود در حالی که RL و DRL اصلاً پایدار نیستند.

چالش های مربوط به داده های اینترنت اشیا برای تکنیک های ML و DL مبتنی بر داده، چالش های مربوط به در دسترس نبودن مجموعه داده های مناسب و کافی وجود دارد که در زیر بحث شده است. این داده ها می توانند حاوی اطلاعات شخصی و حیاتی باشند که نه تنها کاربران، بلکه رفتار و سبک زندگی آنها را نیز شناسایی می کند. داده های تولید شده توسط BAN و سایر برنامه های مربوط به مراقبت های بهداشتی ممکن است حریم خصوصی کاربر را به خطر بیندازد و داده های حاصل از خانه هوشمند منجر به افشای سبک زندگی شخصی و همچنین رفتار شود.

تا به امروز، بسیاری از تکنیک های ناشناس سازی استفاده شده است که اطلاعات را قبل از استفاده برای تجزیه و تحلیل، ناشناس می کند. با این حال، تحقیقات همچنین نشان داده است که می توان با تزریق داده های نادرست، تکنیک های گمنام سازی را هک کرد و مدل های آموزشی را به خطر انداخت. جمع آوری داده ها با حفظ حریم خصوصی و ناشناس ماندن می تواند چالش برانگیز باشد.

۱) سوالاتی از قبیل نحوه استفاده از الگوریتم های ML و DL بر روی چنین داده هایی و میزان حفظ حریم خصوصی توسط الگوریتم های ML و DL باید به آنها پاسخ داده شود. تولید داده های مصنوعی برای آموزش و آزمایش مدل های DL می تواند از نظر محاسباتی بسیار گران باشد.

۲) عدم تعادل داده ها: برای یک سیستم اینترنت اشیاء، هنگامی که حملات به ندرت اتفاق می افتند، مجموعه داده های جمع آوری شده برای ML یا DL بسیار متعادل نیستند. این داده های عدم تعادل می تواند به طور قابل توجهی بر عملکرد طبقه بندی کننده حمله یا روش های IDS تأثیر بگذارد.

۳) همجوشی داده ها: برای ساخت مدل های ML و DL باید تلفیق داده ها از دستگاه های مختلف اینترنت اشیا و عناصر شبکه انجام شود.

۴) چالش های قانونگذاری برای ML در امنیت اینترنت اشیا The هجوم خدمات و برنامه های کاربردی اینترنت اشیا dom در حوزه های مختلف بحث قانونگذاری را در میان جامعه تحقیقاتی و صنعت برانگیخته است.

برخی از حوزه های اینترنت اشیا هنوز با سیاست های کارآمد و قابل قبول قانونی درگیر هستند. فناوری اتومبیل مستقل از IoT در سفارشی سازی تجربه کاربر بهره مند می شود. با این حال، هیچ قانون روشنی برای تجاری سازی فناوری اتومبیل خودمختار و همچنین استفاده از داده های تولید شده توسط این فناوری ها برای آموزش و تجزیه و تحلیل در دسترس نیست. بیمه چالش دیگری برای چنین فناوری است که در آن سخت است که تصمیم بگیریم چه کسی را بیمه کنیم.

این چالش ها به همان اندازه بر فناوری های پشتیبانی شده توسط اتومبیل خودمختار از جمله اینترنت اشیا affect تأثیر می گذارد. علاوه بر این، داده های تولید شده توسط این فناوری ها برای مکانیسم ML برای یادگیری و مدل سازی رفتارهای مختلف مورد نیاز خواهد بود.

اجرای مقررات عمومی حفاظت از داده ها و مقررات مختلف در مورد واردات و صادرات الگوریتم های رمزنگاری نیز امنیت IoT را با چالش های مهمی روبرو می کند.

علاوه بر این، قوانین مختلف در مورد کاربردهای مختلف اینترنت اشیا مانند خانه هوشمند، سلامت الکترونیکی هوشمند و غیره تحت قوانین مختلف در کشورهای مختلف قرار دارد، یک راه حل امنیتی ممکن است برای مناطق مختلف کارساز نباشد.

در طیف گسترده ای از حوزه های کاربردی، مانند امنیت سایبری، تجارت الکترونیکی، پردازش داده های تلفن همراه، تجزیه و تحلیل سلامت، مدل سازی کاربر و تجزیه و تحلیل رفتاری، می توان از خوشه بندی استفاده کرد. در ادامه، انواع مختلف روشهای خوشه بندی را به طور خلاصه مورد بحث و بررسی قرار می دهیم.

روش های تقسیم بندی: براساس ویژگی ها و شباهت های موجود در داده ها، این رویکرد خوشه بندی، داده ها را به چندین گروه یا خوشه طبقه بندی می کند. دانشمندان یا تحلیل گران به طور معمول تعداد خوشه ها را بسته به ماهیت برنامه های هدف، به صورت پویا یا ایستا تعیین می کنند تا برای روش های خوشه بندی تولید کنند. متداول ترین الگوریتم های خوشه بندی مبتنی بر روش های پارتیشن بندی K-means، K-Mediods، CLARA

و غیره است. الگوریتم های خوشه بندی معمولی بر اساس تراکم DBSCAN، OPTICS و غیره هستند.

روش های مبتنی بر سلسله مراتب: خوشه بندی سلسله مراتبی به طور معمول به دنبال ایجاد سلسله مراتبی از خوشه ها، به عنوان مثال، ساختار درخت است.

روش های مبتنی بر شبکه: برای مقابله با مجموعه داده های گسترده، خوشه بندی مبتنی بر شبکه به ویژه مناسب است.

برای به دست آوردن خوشه ها، اصل ابتدا خلاصه کردن مجموعه داده با نمایش شبکه و سپس ترکیب سلول های شبکه است.

آیا الگوریتم های استاندارد خوشه بندی مبتنی بر شبکه هستند.

روش های مبتنی بر مدل: الگوریتم های خوشه بندی مبتنی بر مدل به طور عمده دو نوع دارند: یکی که از یادگیری آماری استفاده می کند و دیگری بر اساس روشی برای یادگیری شبکه عصبی.

روش های مبتنی بر محدودیت: خوشه بندی مبتنی بر محدودیت یک رویکرد نیمه نظارت شده برای خوشه بندی داده است که از محدودیت هایی برای ترکیب دانش دامنه استفاده می کند.

۷- شبکه عصبی مصنوعی و یادگیری عمیق

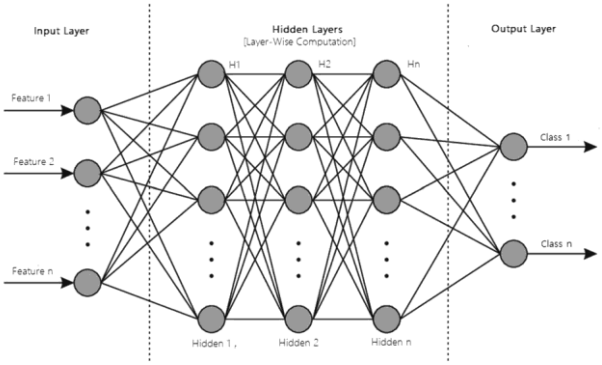
یادگیری عمیق بخشی از خانواده گسترده ای از رویکردهای یادگیری ماشین مبتنی بر شبکه های عصبی مصنوعی با یادگیری بازنمایی است. یادگیری عمیق با ترکیب چندین لایه پردازشی، مانند لایه های ورودی، پنهان و خروجی، برای یادگیری از داده ها، یک معماری محاسباتی را فراهم می کند. مزیت اصلی یادگیری عمیق نسبت به روشهای سنتی یادگیری ماشین عملکرد بهتر آن در موارد مختلف، به ویژه یادگیری از مجموعه داده های بزرگ است [۱۴، ۱۷]. در ادامه، ما انواع مختلفی از روش های یادگیری عمیق را که می تواند برای ساخت مدلهای داده محور موثر برای اهداف مختلف استفاده شود، مورد بحث قرار می دهیم.

MLP: معماری پایه یادگیری عمیق، که به آن شبکه عصبی مصنوعی feed-forward نیز گفته می شود، perceptron چند لایه نامیده می شود. یک MLP معمولی یک شبکه کاملاً متصل است که از یک لایه ورودی، یک یا چند لایه مخفی و یک لایه خروجی تشکیل شده است، همانطور که در شکل نشان داده شده است. هر گره در یک لایه با وزن مشخص به هر گره در لایه زیر متصل می شود.

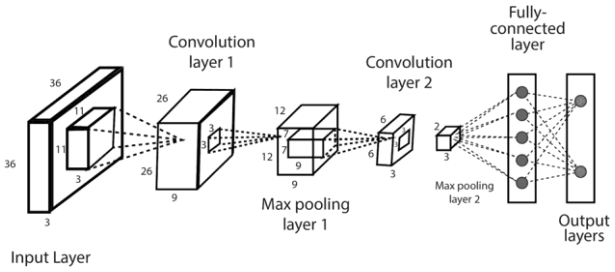
CNN یا ConvNet: شبکه عصبی کانولوشن طراحی ANN استاندارد را تشکیل می دهد، متشکل از لایه های کانولوشن، لایه های جمع کردن و همچنین لایه های کاملاً متصل.

LSTM-RNN: حافظه کوتاه مدت طولانی یک معماری مصنوعی شبکه عصبی است که در زمینه یادگیری عمیق استفاده می شود. شبکه های LSTM برای تجزیه و تحلیل و یادگیری داده های متوالی، مانند طبقه بندی، پردازش و پیش بینی داده ها بر اساس داده های سری زمانی، که آنها را از سایر شبکه های متمایز متمایز می کند، مناسب هستند. علاوه بر این متداول ترین روشهای یادگیری عمیق که در بالا بحث شد، چندین رویکرد یادگیری عمیق دیگر برای اهداف مختلف در منطقه وجود دارد.

رمزگذار خودکار: یکی دیگر از فنون یادگیری است که به طور گسترده ای برای کاهش ابعاد و همچنین استخراج ویژگی در کارهای یادگیری بدون نظارت استفاده می شود. یک شبکه اعتقادی عمیق معمولاً از شبکه های ساده و بدون نظارت مانند ماشین های محدود بولتزمن یا رمزگذاران خودکار و یک شبکه عصبی تولید مجدد تشکیل شده است. یک شبکه تبلیغاتی مولد نوعی شبکه برای یادگیری عمیق است که می تواند داده هایی با ویژگی های نزدیک به ورودی داده واقعی تولید کند. یادگیری انتقال در حال حاضر بسیار رایج است زیرا می تواند شبکه های عصبی عمیق را با داده های نسبتاً کم، که معمولاً استفاده مجدد از یک مسئله جدید با یک مدل از قبل آموزش دیده است، آموزش دهد.



شکل ۷ ساختاری از یک مدل شبکه عصبی مصنوعی با چندین لایه پردازش



شکل ۸ شامل چندین لایه کانولوشن و استخر نمونه ای از شبکه عصبی کانولوشن (CNN یا vNet) - Con

فناوری یادگیری ماشینی می تواند به مشاغل کمک کند تا تاریخ خرید مشتریان خود را تجزیه و تحلیل کنند و براساس خرید و ترجیحات خود، پیشنهادهای محصول سفارشی را برای خرید بعدی خود ارائه دهند.

با استفاده از مدل سازی پیش بینی بر اساس تکنیک های یادگیری ماشین، بسیاری از خرده فروشان آنلاین، مانند آمازون، می توانند موجودی را بهتر مدیریت کنند، از موقعیت های خارج از انبار جلوگیری کنند و تدارکات و انبار را بهینه کنند.

تکنیک های یادگیری ماشینی شرکت ها را قادر می سازد بسته ها و محتوایی متناسب با نیازهای مشتریان خود بسازند و به آنها امکان می دهد ضمن جذب مشتریان جدید، مشتریان موجود را نیز حفظ کنند.

NLP و تجزیه و تحلیل احساسات: پردازش زبان طبیعی شامل خواندن و درک زبان گفتاری یا نوشتاری از طریق رایانه است [۸، ۱۳]. بنابراین، NLP به رایانه کمک می کند تا متن را بخواند، سخنرانی را بشنود، آن را تفسیر کند، احساسات را تحلیل کند و تصمیم بگیرد که چه جنبه هایی قابل توجه است، جایی که می توان از تکنیک های یادگیری ماشین استفاده کرد.

به طور کلی، تجزیه و تحلیل احساس به عنوان یک کار یادگیری ماشین در نظر گرفته می شود که متن را برای قطبیت، مانند "مثبت"، "منفی" یا "خنثی" همراه با احساسات شدیدتر مانند بسیار شاد، شاد، غمگین، بسیار غمگین، عصبانی، دارای تجزیه و تحلیل می کند علاقه، یا علاقه مند نیست و غیره

تشخیص تصویر، گفتار و الگو: بازشناسی تصویر نمونه ای شناخته شده و گسترده از یادگیری ماشین در دنیای واقعی است که می تواند یک شی object را به عنوان یک تصویر دیجیتالی شناسایی کند.

چندین روش یادگیری ماشین مانند طبقه بندی، انتخاب ویژگی، خوشه بندی یا روش برچسب گذاری توالی در منطقه استفاده می شود. [۴، ۵]. یادگیری ماشین را می توان در مراحل مختلف کشاورزی پایدار مانند مرحله قبل از تولید - برای پیش بینی عملکرد محصول، خصوصیات خاک، نیاز آبیاری و غیره استفاده کرد.

محیط توسعه برنامه تلفن همراه با قدرت هوش مصنوعی، به ویژه تکنیک های یادگیری ماشین از طریق قابلیت یادگیری آنها از داده های متنی، بسیار تغییر کرده است [۱۳، ۱۸]. بنابراین، توسعه دهندگان برنامه های تلفن همراه می توانند به یادگیری ماشین برای ایجاد برنامه های هوشمند اعتماد کنند که بتواند رفتار، پشتیبانی و سرگرمی انسانها را درک کند.

برای ساخت سیستم های مختلف شخصی مبتنی بر داده مبتنی بر آگاهی از زمینه، مانند مدیریت وقفه هوشمند، توصیه هوشمند تلفن همراه، جستجوی هوشمند آگاه از زمینه، تصمیم گیری که به طور هوشمندانه به کاربران نهایی تلفن های همراه در یک محیط رایانه ای فراگیر کمک می کند، تکنیک های یادگیری ماشین مناسب.

برای پیش بینی وقایع آینده در زمینه های مختلف، می توان از روش های طبقه بندی استفاده کرد [۱۵،۱۹]. بنابراین، تکنیک های مختلف یادگیری در فرقه بحث شده است.

"وظایف و الگوریتم های یادگیری ماشین" می تواند به ایجاد برنامه های سازگار و هوشمند متناسب با تنظیمات کاربران تلفن همراه کمک کند.

نتیجه گیری

امنیت و حریم خصوصی اینترنت اشیا از اهمیت فوق العاده ای برخوردار است و نقشی محوری در تجاری سازی فناوری اینترنت اشیا دارد. راه حل های امنیتی و حریم خصوصی سنتی از تعدادی از مسائل مربوط به ماهیت پویای شبکه های اینترنت اشیا رنج می برند. مجموعه داده های مورد نیاز برای الگوریتم های ML و DL هنوز کمیاب هستند، که باعث می شود معیار کارآیی راه حل های امنیتی مبتنی بر ML و DL کار دشواری باشد. ما درباره چالش های امنیت و حریم خصوصی در اینترنت اشیا، بردارهای حمله و الزامات امنیتی بحث کرده ایم. ما تکنیک های مختلف ML و DL و کاربردهای آنها را برای امنیت اینترنت اشیا شرح داده ایم.

سپس ما در مورد راه حل های امنیتی موجود بحث کرده و چالش های باز و مسیرهای تحقیقاتی آینده را شرح داده ایم. برای کاهش برخی از کاستی های رویکردهای یادگیری ماشین در امنیت اینترنت اشیا، مبانی نظری DL و DRL باید تقویت شود تا عملکرد مدل های DL و DRL بر اساس پارامترهایی مانند پیچیدگی محاسبات، یادگیری کمی شود. کارایی، و همچنین استراتژی های تنظیم پارامتر.

مراجع

1. Cao L. Data science: a comprehensive overview. ACM Comput Surv (CSUR). 2017;50(3):43.
2. Han J, Pei J, Kamber M. Data mining: concepts and techniques. Amsterdam: Elsevier; 2011.
3. John GH, Langley P. Estimating continuous distributions in bayesian classifiers. In: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc. 1995; 338–345
4. Kamble SS, Gunasekaran A, Gawankar SA. Sustainable industry framework: a systematic literature review identifying the current trends and future perspectives. Process Saf Environ Protect. 2018;117:408–25.
5. Kamble SS, Gunasekaran A, Gawankar SA. Achieving sus- tainable performance in a data-driven agriculture supply chain: a review for research and applications. Int J Prod Econ. 2020;219:179–94.
6. Liu H, Motoda H. Feature extraction, construction and selection: A data mining perspective, vol. 453. Springer Science & Business Media; 1998.
7. McCallum A. Information extraction: distilling structured data from unstructured text. Queue. 2005;3(9):48–57.
8. Otter DW, Medina JR , Kalita JK. A survey of the usages of deep learning for natural language processing. IEEE Trans Neu- ral Netw Learn Syst. 2020.
9. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30.
10. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1:81–106.
11. Quinlan JR. C4.5: programs for machine learning. Mach Learn. 1993.
12. Sarker IH, Alqahtani H, Alsolami F, Khan A, Abushark YB, Siddiqui MK. Context pre-modeling: an empirical analysis for classification based user-centric context-aware predictive modeling. J Big Data. 2020;7(1):1–23.
13. Sarker IH, Hoque MM, MdK Uddin, Tawfeeq A. Mobile data science and intelligent apps: concepts, ai-based modeling and research directions. Mob Netw Appl, pages 1–19, 2020.
14. Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A. Cybersecurity data science: an overview from machine learning perspective. J Big Data. 2020;7(1):1–29.
15. Sarker IH, Watters P, Kayes ASM. Effectiveness analy- sis of machine learning classification models for predicting personalized context-aware smartphone usage. J Big Data. 2019;6(1):1–28.
16. Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2005.
17. Xin Y, Kong L, Liu Z, Chen Y, Li Y, Zhu H, Gao M, Hou H, Wang C. Machine learning and deep learning methods for cyber- security. IEEE Access. 2018;6:35365–81.
18. Zhu H, Cao H, Chen E, Xiong H, Tian J. Exploiting enriched contextual information for mobile app classification. In: Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012; pages 1617–1621
19. Zulkernain S, Madiraju P, Ahamed SI. A context aware interruption management system for mobile devices. In: Mobile Wireless Middleware, Operating Systems, and Applications. Springer. 2010; pages 221–234
20. A. Mosenia and N. K. Jha, "A comprehensive study of security of internet-of-things," *IEEE Transactions on Emerging Topics in Com- puting*, vol. 5, pp. 586–602, Oct 2017.
21. J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet of Things Journal*, vol. 4, pp. 1125–1142, Oct 2017.
22. M. Ammar, G. Russello, and B. Crispo, "Internet of things: A survey on the security of iot frameworks," *Journal of Information Security and Applications*, vol. 38, pp. 8 – 27, 2018.
23. I. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz, and J. Lopez, "A survey of iot-enabled cyberattacks: Assessing attack paths to critical infrastructures and services," *IEEE Communications Surveys Tutorials*, vol. 20, pp. 3453–3495, Fourthquarter 2018.
24. F. Restuccia, S. DOro, and T. Melodia, "Securing the internet of things in the age of machine learning and software-defined networking," *IEEE Internet of Things Journal*, vol. 5, pp. 4829–4842, Dec 2018.
25. J. Hou, L. Qu, and W. Shi, "A survey on internet of things security from data perspectives," *Computer Networks*, vol. 148, pp. 295 – 306, 2019.
26. D. E. Kouicem, A. Bouabdallah, and H. Lakhlef, "Internet of things security: A top-down survey," *Computer Networks*, vol. 141, pp. 199 – 221, 2018.
27. F. A. Alaba, M. Othman, I. A. T. Hashem, and F. Alotaibi, "Internet of things security: A survey," *Journal of Network and Computer Applications*, vol. 88, pp. 10 – 28, 2017.
28. B. B. Zarpelo, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in internet of things," *Journal of Network and Computer Applications*, vol. 84, pp. 25 – 37, 2017.
29. M. binti Mohamad Noor and W. H. Hassan, "Current research on in- ternet of things (iot) security: A survey," *Computer Networks*, vol. 148, pp. 283 – 294, 2019.
30. K. Sha, W. Wei, T. A. Yang, Z. Wang, and W. Shi, "On security challenges and open issues in internet of things," *Future Generation Computer Systems*, vol. 83, pp. 326 – 337, 2018.

31. J. Granjal, E. Monteiro, and J. S. Silva, "Security for the internet of things: A survey of existing protocols and open research issues," *IEEE Communications Surveys Tutorials*, vol. 17, pp. 1294–1312, thirdquarter 2015.
32. Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in internet-of-things," *IEEE Internet of Things Journal*, vol. 4, pp. 1250–1258, Oct 2017.
33. A. Chowdhury and S. A. Raut, "A survey study on internet of things re- source management," *Journal of Network and Computer Applications*, vol. 120, pp. 42 – 60, 2018.
34. P. Asghari, A. M. Rahmani, and H. H. S. Javadi, "Internet of things applications: A systematic review," *Computer Networks*, vol. 148, pp. 241 – 261, 2019.
35. D. Airehrour, J. Gutierrez, and S. K. Ray, "Secure routing for internet of things: A survey," *Journal of Network and Computer Applications*, vol. 66, pp. 198 – 213, 2016.
36. J. Guo, I.-R. Chen, and J. J. Tsai, "A survey of trust computation mod- els for service management in internet of things systems," *Computer Communications*, vol. 97, pp. 1 – 14, 2017.
37. V. Gazis, "A survey of standards for machine-to-machine and the internet of things," *IEEE Communications Surveys Tutorials*, vol. 19, pp. 482–511, Firstquarter 2017.
38. S. B. Baker, W. Xiang, and I. Atkinson, "Internet of things for smart healthcare: Technologies, challenges, and opportunities," *IEEE Access*, vol. 5, pp. 26521–26544, 2017.
39. F. Javed, M. K. Afzal, M. Sharif, and B. Kim, "Internet of things (iot) operating systems support, networking technologies, applications, and challenges: A comparative review," *IEEE Communications Surveys Tutorials*, vol. 20, pp. 2062–2100, thirdquarter 2018.
40. A. olakovi and M. Hadiali, "Internet of things (iot): A review of enabling technologies, challenges, and open research issues," *Computer Networks*, vol. 144, pp. 17 – 39, 2018.
41. M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for iot big data and streaming analytics: A survey," *IEEE Com- munications Surveys Tutorials*, vol. 20, pp. 2923–2960, Fourthquarter 2018.
42. W. Ding, X. Jing, Z. Yan, and L. T. Yang, "A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion," *Information Fusion*, vol. 51, pp. 129 – 144, 2019.
43. D. B. J. Sen, "Internet of Things - Applications and Challenges in Technology and Standardization," *IEEE Transactions in Wireless Personal Communication*, May 2011.
44. U. S. Shanthamallu, A. Spanias, C. Tepedelenlioglu, and M. Stanley, "A Brief Survey of Machine Learning Methods and their Sensor and IoT Applications ," *IEEE Conference on Information, Intelligence, Systems and Applications*, March 2018.
45. E. Fernandes, A. Rahmati, K. Eykholt, and A. Prakash, "Internet of things security research: A rehash of old ideas or new intellectual challenges?," *IEEE Security Privacy*, vol. 15, no. 4, pp. 79–84, 2017.
46. J. Chen, S. Li, H. Yu, Y. Zhang, D. Raychaudhuri, R. Ravindran, H. Gao, L. Dong, G. Wang, and H. Liu, "Exploiting icn for real- izing service-oriented communication in iot," *IEEE Communications Magazine*, vol. 54, pp. 24–30, December 2016.
47. Y. Liu, Y. Kuang, Y. Xiao, and G. Xu, "Sdn-based data transfer security for internet of things," *IEEE Internet of Things Journal*, vol. 5, pp. 257– 268, Feb 2018.
48. E. B. Karbab, M. Debbabi, A. Derhab, and D. Mouheb, "Maldozer: Automatic framework for android malware detection using deep learn- ing," *Digital Investigation*, vol. 24, pp. S48 – S59, 2018.
49. J. Su, D. V. Vargas, S. Prasad, D. Sgandurra, Y. Feng, and K. Sakurai, "Lightweight classification of iot malware based on image recognition," *CoRR*, vol. abs/1802.03714, 2018.
50. Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breit- enbacher, and Y. Elovici, "N-baiotnetwork-based detection of iot botnet attacks using deep autoencoders," *IEEE Pervasive Computing*, vol. 17, pp. 12–22, Jul 2018.
51. P. Mohamed Shakeel, S. Baskar, V. R. Sarma Dhulipala, S. Mishra, and M. M. Jaber, "Maintaining security and privacy in health care system using learning based deep-q-networks," *Journal of Medical Systems*, vol. 42, p. 186, Aug 2018.