Годишник Телекомуникации 2021, том 8, с. 53 - 64

Yearbook Telecommunications 2021, vol. 8, p. 53 - 64

eISSN 2534-854X https://telecommunications.nbu.bg/bg/godishnik-telekomunikacii

DOI: https://doi.org/10.33919/YTelecomm.21.8.6

ПРИЛОЖЕНИЕ НА ПОДХОДИ НА ИЗКУСТВЕН ИНТЕЛЕКТ И МАШИННО ОБУЧЕНИЕ В КИБЕРСИГУРНОСТТА

Васил Кърдев, Росен Пасарелски

APPLICATION OF APPROACHES TO ARTIFICIAL INTELLIGENCE AND MACHINE TRAINING IN CYBER SECURITY

Vasil Kadrev, Rosen Pasarelski

Резюме: Интерес представлява как машинното обучение се използва в киберсигурността както за защитни, така и за престъпни дейности, вкл. и за кибератаки, насочени към модели на машинно обучение. Цел на разработката, е да се покаже как машинното обучение може да се приложи в контекста на сигурността както от гледна точка на защита, така и на атака, както и потенциалните заплахи, насочени към моделите на машинно обучение.

Получени са резултати, използвайки аналитична описателна методология въз основа на литературни източници, като е разгледано използването на техники за изкуствен интелект и машинно обучение за повишаване на киберсигурността, както и най-важните области на приложение, които подобряват киберсигурността. От друга страна са разгледани особеностите и приложенията на изкуствения интелект и машинното обучение в услуга на киберпрестъпниците.

По отношение на приносната част, е направен задълбочен и обширен анализ на особеностите и приложенията на изкуствения интелект и машинното обучение, от две гледни точки - киберсигурност и киберпрестъпления.

Ключови думи: киберсигурност, киберпрестъпления, изкуствен интелект, машинно обучение

Abstract: Of interest is how machine learning is used in cybersecurity for both security and criminal activities, incl. and for cyberattacks targeting machine-learning models. The aim of the development is to show how machine learning, can be applied in the context of security in terms of both protection and attack, as well as the potential threats posed to machine learning models. Results were obtained using analytical descriptive methodology based on literature sources, examining the use of artificial intelligence techniques and machine learning to increase cybersecurity, as well as the most important areas of application that improve cybersecurity. On the other hand, the features and applications of artificial intelligence and machine learning in the service of cybercriminals are considered. Regarding the contribution part, an in-depth and extensive analysis of the features and applications of artificial intelligence and machine learning from two points of view (cybersecurity and cybercrime) has been made.

Keywords: cybersecurity, cybercrime, artificial intelligence, machine learning

1. ВЪВЕДЕНИЕ

При изкуствения интелект AI (Artificial Intelligence) се описват концепциите, които, с помощта на интелигентни компютърни програми, позволяват функционирането на компютрите (интелигентни агенти, интелигентни машини) да изглежда разумно, чрез способности за анализ на околната среда и предприемане на действия, които увеличават възможностите за постигане на определени цели. Общи за повечето подобласти на изследванията на изкуствения интелект, са задачи като възможности за анализ, обучение, планиране, общуване, възприемане, както и способността за управление и манипулиране на обекти.

Изкуственият интелект се дефинира като „способността на дадена система да интерпретира правилно външните данни, да се учи от тях и да използва тези знания за

постигане на специфични цели и задачи чрез гъвкава адаптация“. Теорията на изкуствения интелект се основава на хипотезата, че основно човешко качество като интелигентността, може да бъде толкова точно описано, че да бъде симулирано от машина. Все още не е определено теоретично кои условия покриват изискването за интелигентност (съществуват обаче, множество хипотези) [1-4].

За реализация на „интелигентни“ системи има два подхода:

- символно (семиотично, низходящо) на основата на моделиране на мисловни процеси;
- невро-кибернетично (невронно, възходящо) на основата на моделиране на мозъчни структури, неврони.

Понастоящем методите на изкуствения интелект се използват за разрешаване на множество специфични задачи в различни области, вкл. при Интернет на нещата IoT (Internet of Things) [5]. Що се отнася до намаляване на грешките в оперативните задачи и намиране на аномалии, изкуственият интелект изпреварва човешките способности и компетентност, като играе важна роля при оценка на грешките [6].

Изкуственият интелект, като решение за киберсигурност, може да помогне за защита на организациите от интернет заплахи, да идентифицира различни типове зловреден софтуер, да осигури практически стандартите за сигурност и да помогне за създаването на по-добри стратегии за превенция и възстановяване. Представлява интерес, въз основа на особеностите на концепцията за изкуствен интелект, да се идентифицират най-важните области на изкуствения интелект, които могат да се използват в киберсигурността и да се изясни тяхната конкретна роля (особено приложението на машинно обучение (самообучение), извличане на данни, дълбоко обучение и експертни системи) за поддържане на киберсигурността в организациите [7-12].

Същевременно, всички тези изброени ползи от прилагането на методите на изкуствения интелект, представляват интерес и за киберпрестъпниците, както в непосредствената им дейност (например: неоторизиран достъп, автоматизиран зловреден софтуер, усъвършенствани фишинг атаки), така и при дейности, свързани с компрометиране на продуктите за машинно обучение (например: увреждане на данните за обучение, промяна на модела за машинно обучение, избягване на откриване чрез модели на машинно обучение).

2. СПЕЦИФИКА НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ В КИБЕРСИГУРНОСТТА

Основният въпрос е какви подходи и приложения с използване на изкуствен интелект могат да се прилагат в областта на киберсигурността. В тази връзка е необходимо да се изясни концепцията за изкуствения интелект и основните негови области, които пряко се отнасят до киберсигурността.

2.1. Постановка на проблематиката

В тази връзка, съществуват дискусии още през 50-те и 60-те години на миналия век за това дали машината може да изпълнява цялата работа, която хората могат да извършват в ежедневието си. Ако машината има само частични способности за решаване на проблеми и разсъждения, но не е в състояние да изпълни пълните човешки когнитивни способности, се говори за "слаб AI". При концепцията за "силен AI" се включват задачите, които хората традиционно изпълняват, при това се прилага широк спектър от фонови знания и съществува някаква степен на самоуправление и самоосъзнаване.

Изкуственият интелект е област от компютърните науки, в която се създават системи, които могат да функционират интелигентно и независимо, с използване на механизми за вземане на решения, наподобяващи функционалността на мозъка. С AI, машината се

Годишник Телекомуникации 2021, том 8, с. 53 - 64

Yearbook Telecommunications 2021, vol. 8, p. 53 - 64

eISSN 2534-854X https://telecommunications.nbu.bg/bg/godishnik-telekomunikacii

DOI: https://doi.org/10.33919/YTelecomm.21.8.6

обучава от опита чрез обработка на големи количества данни и разпознаване на модела в тях. Например, функционалностите при разпознаването на лица, както и самоуправляващото се превозно средство, се основават на машинно обучение и обработка на естествен език, които са подмножество на AI.

AI включва много свързани области и технологии, като машинно обучение, дълбоко обучение, невронна мрежа, обработка на естествен език и др. (имащи пряко отношение към киберсигурността):

- Машинното обучение прилага множество технологии, които позволяват на компютрите "да мислят" чрез математически алгоритми въз основа на събраните данни и конкретни инструкции и правила. Вместо да се програмира компютъра стъпково, такъв подход предоставя насоки, които позволяват обучение, без формални програмни инструкции. Примери могат да бъдат разпознаване на глас, лица, проследяване на необичайно поведение и др.
- Последващото развитие на машинното обучение води до т. нар. "дълбоко обучение", което представлява компютърен модел, който изпълнява задачи за категоризация/ класификация, директно от изображения, текст или звук. Този модел, от своя страна използва невронна мрежа с множество слоеве, при което точността на разпознаване е по-голяма.
- Невронните мрежи може да се използват, например, като система (модел) за разпознаване (като разпознаване на лица и почерк), и които в крайна сметка, с използване на дълбоко обучение, способстват за внедряване на машинно обучение.
- Обработка на естествен език, при което машините анализират езика и речта (естествения говор) чрез приложения за разпознаване на реч и др.

Огромният потенциал на технологиите с изкуствен интелект може да се използва за подобряване на киберсигурността, тъй като с нарастващия напредък в областта на информационните технологии, престъпниците използват киберпространството, за да извършват различни киберпрестъпления, вкл. и с използване на същите тези технологии.

Както AI, така и киберсигурността са широки области. Съвместно прилагани, използваните в тях технологии, може да се използват, както за намаляване на рисковете в организациите, така и за откриване на киберзаплахи (измами и др.). Например, използването на технологии с изкуствен интелект, улеснява откриването и реакцията на заплахи (злонамерен софтуер), като се използват предишни данни за кибератаки за да се определи най-добрия начин на действие.

Организациите внедряват AI, което позволява множество решения, свързани със сигурността (например по отношение на информация за сигурността), а управлението на събития помага на анализаторите на сигурността да се подобри откриването на всякакви заплахи в корпоративната мрежа. Използването на AI в много случаи може да бъде по-ефективно при откриването на зловреден софтуер.

2.2. Категории на подходите за машинно обучение (самообучение)

Алгоритмите за ML може да се групират в няколко категории в зависимост от използвания подход за обучение (най-подходящият алгоритъм за ML се избира най-вече относно колко е интензивна изчислително и колко бърза е дадена техника и в зависимост от типа на приложението) - табл. 1.

Таблица 1. Категории на подходите за обучение (самообучение)

Самообучение с учител (Supervised Learning)
При алгоритмите за самообучение с учител се построява математически модел на съвкупност от данни (обучаващи данни, вкл. набор от обучаващи примери), за който се дефинира вход и желан изход. Има две фази в самообучението с учител - фаза на обучение и фаза на тестване. Наборите от данни, използвани за фазата на обучение, имат известни етикети (labels). При алгоритмите се научава връзката между входните стойности и етикетите, и се правят опити да се предвидят изходните стойности на тестовите данни. Използват се, например, алгоритми за класификация (за разпознаване на цифри и реч, диагностика, за откриване на измами с идентичност), като "метод на опорните вектори" (Support Vector Machines), "най-близък съсед" (Nearest Neighbor), "случайна гора" (Random Forest), също "линейна регресия" (Linear Regression) и др.
Самообучение без учител (Unsupervised Learning)
Алгоритмите за самообучение без учител приемат съвкупност от данни, съдържаща само входящи стойности, и намират дадена структура в данните, например групи или клъстери. Тези алгоритми се обучават от тестови данни, които не са надписани, класифицирани или категоризирани, като няма налични етикети. Алгоритмите в тази категория идентифицират модели (шаблони) за тестване на данни, както и за групиране на данни, или за предсказване на бъдещи стойности. Самообучението без учител се занимава с проблеми, свързани с намаляване на размерите, използвани при визуализация на големи данни (big data), извличане на характеристики или откриване на скрити структури. Освен това, се използва за групиране на проблеми, сегментиране на клиенти и целенасочен маркетинг.
Смесено самообучение (Semi-supervised Learning)
Това е комбинация от предишните две категории, като се използват както маркирани (етикетирани) данни, така и немаркирани данни. Работи най-вече като самообучение без учител, с възможните ползи от използването на етикетирани данни.
Самообучение с утвърждение RL (Reinforcement Learning)
При този подход за обучение, алгоритмите се опитват да предвидят резултат за даден проблем, въз основа на набор от параметри за настройка, след което изчисленият изходен резултат се преобразува като входен параметър и се изчислява нов изход, докато се намери оптималният изходен резултат. Този подход се използва от изкуствените невронни мрежи ANN (Artificial Neural Networks) и дълбокото обучение (Deep Learning).

Тъй като заплахите срещу киберсигурността непрекъснато се променят и развиват, е необходима автоматизирана и незабавна реакция. Следователно методите на машинно обучение, особено дълбокото обучение, което не изисква непременно предишно обучение или разчитане на предходни класификации, предоставени от експерти, може би са особено важни като приложение на подходите на AI към киберсигурността.

3. ПРИЛОЖЕНИЯ НА МАШИННОТО ОБУЧЕНИЕ В КИБЕРСИГУРНОСТТА

3.1. Особености на приложенията

При нарастващите заплахи към киберсигурността, проучванията се фокусират върху машинното обучение и неговия богат набор от инструменти и техники за идентифициране, както и за реагиране на сложни кибератаки. Машинното обучение би могло да се използва в различни области на киберсигурността, като предоставя подходи за анализ с цел откриване и реагиране на атаки. Също така, може да подобри процесите за повишаване на сигурността, като автоматизира рутинните задачи и улесни анализаторите на сигурността да работят бързо с полуавтоматизирани задачи.

От друга страна, в практиката на киберсигурността често възникват проблеми свързани с обработката на данни, подходящи за машинно обучение. Причината за това е наличието на огромно количество големи и небалансирани набори от данни, недостигът на време,

Годишник Телекомуникации 2021, том 8, с. 53 - 64

Yearbook Telecommunications 2021, vol. 8, p. 53 - 64

eISSN 2534-854X https://telecommunications.nbu.bg/bg/godishnik-telekomunikacii

DOI: https://doi.org/10.33919/YTelecomm.21.8.6

необходимо за извършване на ръчна категоризация, и уникални характеристики на полетата, като категоризация на семантиката, които увеличават разликата между техническия опит и статистическото моделиране. За решаването на тези проблеми се изисква използването на приложения въз основа на дълбоко обучение, като метод на машинното обучение.

Друга важна област, това е методологията за извличане на данни, която има за цел да получи ценна информация и да намери значими скрити модели и тенденции от огромен брой данни в бази данни, които при традиционните статистически подходи не могат да се открият. Това е обширна област, която включва използването на: машинно обучение, бази данни, статистика, експертни системи, високопроизводителни изчисления, невронни мрежи и др. Извличането на данни се извършва по различни начини (например клъстериране, класификация, регресионни модели, анализ на връзки, обобщение и др.).

Експертните системи са един от ефективните инструменти на AI, също и в областта на киберсигурността, и представляват софтуерни пакети, чрез които се получават отговори на запитвания от клиенти, или при необходимост, се предоставя друг софтуерен пакет. Тези системи представляват експертни знания, които се съхраняват в определена област на приложение. Също така те включват механизъм за анализ/ разсъждения за ориентиране към отговорите в светлината на предоставената информация/ въпроса от клиента и друга допълнителна информация.

Може да се разгледат някои популярни приложения на машинното обучение в киберсигурността както следва: откриване и класификация на заплахите, машинно обучение при оценка на риска в мрежата, автоматизиране на рутинните задачи за сигурност и оптимизиране на анализа.

3.2. Откриване и класификация на заплахите

Алгоритмите за машинно обучение могат да бъдат внедрени в приложения за идентифициране и реагиране на кибератаки, преди те да са се разпространили. Това обикновено се постига с помощта на модел, разработен чрез анализ на големи данни за събитията относно сигурността и идентифициране на използвания модел на злонамерените дейности. В резултат, подобни дейности се откриват и обработват автоматично. Наборът от данни за обучение на такива модели обикновено включва предишни идентифицирани и записани показатели IOC (Indicators of Compromise), които след това се използват за изграждане на модели, чрез които може да се наблюдават и идентифицират заплахите и да се реагира на тях в реално време. Също така, с наличието на набори от данни на IOC, може да се използват алгоритми за класификация от машинното обучение, за да се идентифицира различното поведение на зловредните програми в тях и съответно да ги класифицираме. Съществуват изследвания въз основа на поведенческо-базирани рамки за анализ, които използват техники за клъстериране и класификация на машинното обучение, за анализ на поведението на множество зловредни програми.

С използване на така обучените модели става възможно автоматизиране на процеса на откриване и класифициране на нов зловреден софтуер, чрез което анализаторите на сигурността могат бързо да идентифицират и класифицират нов тип заплаха и да отговорят адекватно, като използват решения, базирани на данни. Например, използвайки исторически набор от данни, съдържащ подробни събития от атаки на рансъмуера WannaCry, моделът може с машинно обучение да се научи да идентифицира подобни

атаки, което прави възможно автоматизирането на процеса на идентификация и реагиране на подобни атаки.
Техники за машинно обучение са използвани също и за класификация на IP трафика, водещо до автоматизиране на процеса при системи за откриване на проникване, които също така могат да се използват за идентифициране на поведенчески модели при DDOS атаки. Съществуват изследвания (относно единични, хибридни и ансамблови класификации), които са фокусирани върху анализирането на множество решения при системи за откриване на проникване с използване на машинно обучение.

3.3. Автоматизиране на рутинните задачи за сигурност и оптимизиране на анализа

Машинното обучение може да се използва за автоматизиране на повтарящи се задачи, извършвани от анализаторите на сигурността.
Това може да стане чрез:

- анализиране на записи/отчети за минали действия, предприети от анализатори на сигурността за успешно идентифициране и реагиране на определени атаки;
- използване на тези знания за изграждане на модел, който може да идентифицира подобни атаки и да реагира адекватно без човешка намеса.

Въпреки че е трудно да се автоматизира пълният процес на защита и да се замени човекът-анализатор на сигурността, има определени аспекти на анализа, които може да се автоматизират чрез машинно обучение, вкл. откриване на злонамерен софтуер, анализ на мрежови журнали, както и оценки на уязвимостта, като анализ на мрежовия риск. Чрез включване на машинното обучение в дейностите за поддържане на сигурността, човек и машина могат да обединят усилията си и да постигнат неща, със скорост и качество, които иначе биха били невъзможни.
Все повече задачи е възможно да се автоматизират благодарение на развитието на изкуствения интелект, и определени задачи, които в момента се изпълняват от хора, ще бъдат поети от машини. По-съществено е наличието на синергичен ефект, т.е. комбинацията от изкуствен интелект и човешки интелект дава далеч по-добри резултати, отколкото всеки ще даде сам по себе си.
Поради тази причина, понастоящем се наблюдава възход на компаниите, използващи изкуствен интелект, с фокус не само върху създаването на продукти за автоматизиране на задачи с изкуствен интелект, но и върху създаването на продукти, които подобряват и допълват производителността на човека-анализатор (посредством обобщаване и анализ на огромни обеми данни).
В допълнение (с оглед на подобряване на дейностите на анализаторите на сигурността), с цел създаване на приложения, които генерират правила за класифициране на мрежови връзки, са проведени проучвания за използването на алгоритми за машинно обучение, като генетични алгоритми и инструменти за вземане на решения.
Други подходи използват прилагане на когнитивна архитектура за създаване на автоматизирана система за вземане на решения за киберзащита, с умения на експертно ниво, базирани на това как хората разсъждават и учат. Обикновено анализаторите на киберсигурността трябва да отделят време за реакция на множество събития, които понякога включват фалшиви положителни резултати, които в повечето случаи се оказват загуба на време. Проведени са проучвания, за да се покаже, че алгоритмите на класификаторите за машинно обучение могат да бъдат обучени от данни за предупредителни сигнали, за да се идентифицират и разграничават фалшивите положителни резултати от истинските положителни. По такъв начин е възможно да се създаде автоматизирана система, която ще предупреждава анализатора само за сценарии, които включват истински положителни резултати.

Годишник Телекомуникации 2021, том 8, с. 53 - 64

Yearbook Telecommunications 2021, vol. 8, p. 53 - 64

eISSN 2534-854X https://telecommunications.nbu.bg/bg/godishnik-telekomunikacii

DOI: https://doi.org/10.33919/YTelecomm.21.8.6

3.4. Машинно обучение при оценка на риска в мрежата

Използването на машинно обучение за получаването на количествени мерки при присвояване на оценки на риска в различни сегменти от мрежата, позволява на организациите да приоритизират своите ресурси за киберсигурност, по отношение на различните оценки на риска. Машинното обучение може да се използва за автоматизиране на този процес чрез анализ на исторически набори от данни за кибератаки и определяне кои области в мрежите са били най-вече подложени на определени видове атаки. Използването на машинно обучение дава предимства, тъй като получените резултати не само ще се основават на познанията за домейните в мрежите, но най-важното е, че резултатите ще се управляват от данни. В такъв случай ще може да се определи количествено вероятността и степента на въздействие на атака по отношение на дадена област на мрежата, което може да помогне на организациите да намалят вероятността да бъдат жертви на атаки до допустимия риск.

Проведени са проучвания върху използването на алгоритми за машинно обучение (като K-Nearest Neighbor, Support Vector Machines и алгоритми Random Forest) за анализ и групиране на мрежови активи въз основа на тяхната свързаност.

Други проучвания са фокусирани върху това как устройствата на IoT, свързани към малки и средни предприятия (МСП) SMEs (Small and Medium Sized Enterprises), могат да се използват за осъществяване на "обедни атаки" (lunch attacks)[1] срещу МСП. Разработени са системи, използващи машинно обучение, които използват принципа на взаимното потвърждаване, за да анализират огромни обеми предупредителни сигнали (alerts) в мрежите на организацията, за да определят оценките на риска, като вземат предвид свързаността и асоциациите между различни мрежови единици.

4. ПРИЛОЖЕНИЯ НА МАШИННО ОБУЧЕНИЕ ЗА КИБЕРПРЕСТЪПЛЕНИЯ

Точно както машинното обучение е обещаващ инструмент за справяне с нарастващите киберзаплахи, както беше показано, то е и инструмент, който може да се използва от злонамерени нападатели. Например, има проучвания, които показват възможността киберпрестъпниците да използват машинно обучение, за да създадат интелигентен злонамерен софтуер, който може да надхитри настоящите интелигентни защитни системи. Областта на машинното обучение напредва с бързи темпове, предлагайки обещаващи решения за киберзащита, но също дава възможност на киберпрестъпниците да го използват при извършване на по-сложни мащабни атаки, насочени срещу модели на машинно обучение. Както анализаторите на сигурността, така и киберпрестъпниците, активно търсят иновативни техники/ технологии за изкуствен интелект, които да добавят към арсенала си от кибероръжия. Например, както специалистът по киберзащита активно анализира данни, за да разбере по-добре моделите на нападателите, самите нападатели също могат да откраднат данни за потребителите и да ги анализират, за да създадат по-добре своите атаки.

Пример за такива приложения е нелегален достъп и анализ на имейли на целеви потребители с цел използването му за създаване на по-добри фишинг атаки.

[1] The term "lunchtime attack" refers to the idea that a user's computer, with the ability to decrypt, is available to an attacker while the user is out to lunch. Терминът "атака по време на обяд" се отнася до идеята, че компютърът на потребителя, с възможност за дешифриране, е на разположение на нападателя, докато потребителят е на обяд.

Някои популярни категории техники за атака, базирани на машинно обучение, включват: - методи за неоторизиран достъп; - автоматизиран зловреден софтуер; - усъвършенствани фишинг атаки.

4.1. Неоторизиран достъп.

Машинното обучение може да се използва за получаване на неразрешен достъп до системи, например като тези, включващи CAPTCHA, т.е. програма или система, предназначени да разграничат човешкия от машинния вход, обикновено като начин за осуетяване на нежелана поща и автоматично извличане на данни от уебсайтове.

По същество CAPTCHA[2] е тест за сигурност (използван в информатиката), за който се смята, че може да бъде издържан само от човек. Компютър генерира прост въпрос, чийто отговор е очевиден за човек, но не и за друг компютър. Типичен CAPTCHA тест е показването на разкривени букви, които потребителят трябва да въведе. CAPTCHA понякога е определян като обратен тест на Тюринг, защото е проверка, при която потребител трябва да докаже на компютър, че е човек, за разлика от стандартния тест на Тюринг, при който машина иска да докаже на човек, че е човек.

Една област, която е силно повлияна от машинното обучение, е тази на машинното зрение, при което една машина е обучена да идентифицира обекти. Това е технология, използвана в самоуправляващите се автомобили, където колите разчитат на машинно обучение, за да идентифицират и избягват препятствия. Тъй като машините могат да идентифицират обекти в изображенията, те могат да бъдат обучени да заобикалят системата, базирана на CAPTCHA, която разчита, като условие за упълномощаване на потребител, да идентифицира обектите в изображение. Също така, алгоритмите за машинно обучение, като невронни мрежи, които се опитват да имитират човешкия мозък, могат да бъдат обучени да ускоряват и автоматизират техниките за социално инженерство, като например отгатване на потребителски пароли, като обучават модела с големи масиви от данни, съдържащи данни за хакнати преди това потребителска информация, включително техните потребителски имена и пароли и всякакъв вид информация, която може да се използва за подобряване на процеса на отгатване.

Проведени са множество изследвания за това как машинното обучение може да се използва за неоторизиран достъп до системите. Примерите включват PassGAN, които могат да генерират много достоверни предположения за пароли, като използват т.нар. генеративни състезателни мрежи GAN (Generative Adversarial Networks) и реални изтичания на пароли, за да научат разпространението на реалните пароли. Някои проучвания, с цел успешен неоторизиран достъп, се фокусират върху използването на машинно обучение за генериране на пароли при атаки с груба сила (brute force attacks) в реално време, които се базират на тестване на различни варианти на пароли.

Някои изследвания се фокусират върху използването на дълбоко обучение за заобикаляне на CAPTCHA без човешка намеса. Други изследвания се фокусират върху използването на машинно обучение за клониране на човешки глас, за което съществуват и приложения.

4.2. Автоматизиран зловреден софтуер.

Обикновено създаването на зловреден софтуер включва писане на злонамерена програма, която в повечето основни случаи може да бъде идентифицирана от програмите за сигурност, които имат записи на зловредните програми. Има обаче случаи, при които машинното обучение се използва за генериране на зловреден код, който програмите за

[2] Терминът „CAPTCHA“ е въведен през 2000 от Люис фон Ан, Манюел Блум, Никълъс Дж. Хопър (от Университета „Карнеги Мелън“) и Джон Ленгфърд (IBM) и е съкращение от „Completely Automated Public Turing test to tell Computers and Humans Apart“ (напълно автоматизиран публичен тест на Тюринг за разграничаване на компютри от хора). Настоящата официална версия на CAPTCHA е reCAPTCHA.

Годишник Телекомуникации 2021, том 8, с. 53 - 64

Yearbook Telecommunications 2021, vol. 8, p. 53 - 64

eISSN 2534-854X https://telecommunications.nbu.bg/bg/godishnik-telekomunikacii

DOI: https://doi.org/10.33919/YTelecomm.21.8.6

сигурност не могат да открият, вкл. системи, базирани на машинно обучение. Друг пример включва DeepLocker, злонамерен софтуер, функциониращ с AI, разработен от изследователи на IBM, който е в състояние да използва разпознаване на лица, разпознаване на глас и геолокация, за да идентифицира целта си преди да започне атаката. Има множество изследвания за използването на машинно обучение за генериране на компютърен код в сценарии, при които AI може да напише кода без човешка намеса (например проучване, проведено в Microsoft).

4.3. Усъвършенствани фишинг атаки.

Машинното обучение може да се използва за извършване на усъвършенствани фишинг атаки (spear phishing attacks), например чрез незаконно събиране на истински данни от имейли на целеви лица и подаване на данните към модел на машинно обучение, който след това може да се учи от данните, да извлече контекст от тях и да генерира имейли, които изглеждат истински, като тези, от които се е научил. След това резултатът може да бъде включен в автоматизиран процес, като по този начин се ускорява ефективността и скоростта, с които киберпрестъпниците могат да предприемат целенасочени фишинг атаки.

При някои фишинг атаки се използва социалното инженерство, за да се получи незаконно информация за целевите потребители. Социалното инженерство е популярен вид техника за атака, която използва измама, за да манипулира хората, за да получи личната им информация. Има проучвания върху използването на машинно обучение за извършване на сложни атаки с използване на социалното инженерство. Примерите включват проучвания с използване на невронна мрежа[3] (мрежа със специализирана памет LSTM (Long Short-Term Memory) и управляем рекурентен блок GRU (Gated Recurrent Units)), които се обучават от дадена публикация в социалните медии, извлечена в съответствие с дадено целево време, с цел манипулиране на потребителите, при кликване върху измамни URL адреси. Подобни подходи могат да се използват за имейл базираните фишинг атаки.

5. СИГУРНОСТ НА ПРОДУКТИТЕ ЗА МАШИННО ОБУЧЕНИЕ

От началото на компютърната революция, киберпрестъпниците винаги са търсили начини за използване на уязвимости на софтуера и извършване на злонамерени дейности. С експлозивния растеж на технологията за изкуствен интелект, киберпрестъпниците започват да търсят начини за използване на уязвимости в тази област. Атаките върху системите за машинно обучение обикновено се обсъждат в контекста на състезателното машинно обучение (Adversarial Machine Learning), което се занимава с прилагането на техники за машинно обучение към задачи, свързани със сигурността, като биометрично разпознаване, филтриране на нежелана поща, проникване в мрежата и откриване на злонамерен софтуер.

Атаките срещу алгоритмите за машинно обучение могат да бъдат категоризирани в три области:

- атаки, насочени към промяна на набори от данни за обучение и въвеждане на уязвимости в крайния модел;
- атаки, насочени към увеличаване на процента грешки на крайния модел;
- атаки, целящи да направят възможно, определен набор от записи да бъдат класифицирани или интерпретирани от модела по желание на нападателя.

[3] https://en.wikipedia.org/wiki/Recurrent_neural_network

Моделът за машинно обучение се изгражда чрез подаване на данни в компютърен алгоритъм, който след това обучава модела от данните, а по-късно може да се използват обучените модели, за да се предскажат или класифицират неизвестни данни. Крайният продукт на модела за машинно обучение може да бъде, например, просто уравнение, което след това се транслира като компютърен код, който при получаване на съответен вход, генерира изход под формата на класификация или прогноза.
От краткото описание на приложението на моделите за машинно обучение, може да се види, че киберпрестъпниците могат да неутрализират продукта за сигурност, с използване на машинно обучение, като променят данните за обучение или променят крайните параметри на модела.

5.1. Увреждане на данните за обучение

От практиките за машинно обучение е добре известно, че успехът на проектите, свързани с машинно обучение зависи до голяма степен от качеството на данните. Това означава, че ако моделът се тренира с увредени данни, това ще доведе до увредени резултати, независимо от това колко моделът е усъвършенстван. Възможно е киберпрестъпникът да получи достъп до обучителния набор данни на модела на машинно обучение и да промени данните преди началото на обучението, без знанието на инженерите по машинно обучение, т.е. данните вече ще бъдат фалшифицирани, което ще доведе до моделиране, базирано на грешни данни. Следователно, окончателният модел вече няма да бъде надежден, тъй като е бил обучен с фалшифицирани данни и няма значение колко е добър процесът на моделиране, прогнозите или класификациите, със сигурност няма да бъдат подходящи. Също така, друг сценарий може да бъде в ситуация, в която даден модел е проектиран да се обучава всеки път, когато получава нови записи. В този случай, киберпрестъпникът може да захрани модела с фалшифицирани данни, който ще се обучи от тях и резултатът от моделирането ще е отрицателен. Проведени са множество проучвания за анализиране и защита срещу атаки за увреждане на данни.

5.2. Промяна на модела за машинно обучение

Ако киберпрестъпникът получи неправомерно достъп до модел на машинно обучение и промени параметрите му, това ще доведе до дирижирани резултати. Например, ако алгоритъмът на обучен и внедрен модел за машинно обучение, може да бъде представен математически като y = 1 + 2x (където x е входният параметър, а y е резултатът от модела), тогава, ако киберпрестъпникът промени уравнението като y = 1 - 2x, може да се види, че това очевидно ще доведе до грешни прогнози и съответно до погрешни (катастрофални) решения.

5.3. Избягване на откриване чрез модели на машинно обучение

Този случай се отнася за атаки, които имат и за цел да избегнат разкриване. Това може да се случи в ситуации, когато нападателят променя данните, използвани при тестване, с цел избягване да бъде класифициран като заплаха по време на стандартните системни операции. Като пример в проучванията, за да се покаже как може да се правят такива атаки, са използвани биометричните системи.

От разгледаните особености се вижда, че при проектите за машинно обучение е жизненоважно сигурността да се взема напълно сериозно. Необходимо е да се вземат подходящи мерки за наблюдение и защита на моделите за машинно обучение и техните набори от данни.

Годишник Телекомуникации 2021, том 8, с. 53 - 64

Yearbook Telecommunications 2021, vol. 8, p. 53 - 64

eISSN 2534-854X https://telecommunications.nbu.bg/bg/godishnik-telekomunikacii

DOI: https://doi.org/10.33919/YTelecomm.21.8.6

6. ЗАКЛЮЧЕНИЕ И ИЗВОДИ

Въз основа на резултатите, получени с помощта на аналитична и описателна методология, за основните характерни случаи на използване на машинно обучение, дълбоко обучение и методи за извличане на данни, може да се обобщят възможностите за целите на киберсигурността, например, за откриване на проникване, на злонамерен софтуер и на спам.

Резултатите също така показват, че ефективността на използването на методи за машинно обучение за целите на киберсигурността, е възможно да се ограничава от различни слабости, което изисква постоянно преоценяване и внимателно коригиране на параметри, което, от своя страна, е трудно за автоматизация.

Освен това, главно когато една и съща функционалност се прилага за идентифициране на разнообразни заплахи, ефективността на определянето е неприемливо ниска. Това може да бъде преодоляно чрез използване на различни дневници (machine-based workbooks) за идентифициране на конкретни специфични заплахи.

Също така, тъй като машинното обучение е на сравнително ранен етап, не може да се стигне до категорично заключение относно неговата ефикасност за целите на киберсигурността. В перспектива може да се очаква развитие, особено с отчитане на т. нар. състезателно обучение, но ролята на дълбокото обучение понастоящем се очертава като водещо средство на машинно обучение, което може да допринесе за повишаване на киберсигурността. Например, системите за киберсигурност, базирани на алгоритми за дълбоко обучение, имат съществени предимства, като намаляване на обема на ръчно идентифициране на модели за подозрително поведение, и в тази връзка способността за подобряване на ефективността.

За откриване на злонамерен софтуер се използват различни стратегии и алгоритми за обработка при извличането на данни от огромен набор от информация, които са с различна ефективност. Всяка от стратегиите за извличане на данни има различна насоченост, като откриване на аномалии, откриване на злоупотреба и хибридно откриване. Алгоритмите за извличане на данни могат да се използват при всяка стратегия, но все пак имат силни и слаби страни. Например, използваните алгоритми, при откриване на злонамерен софтуер, са: обучение чрез дърво за вземане на решения (Decision Tree Learning), класификатор на Бейс (Native Bayes Classifier NB), K-най-близък съсед (K-Nearest Neighbour) и метод на опорните вектори (Support Vector Machine). Някои от тези алгоритми имат критични ограничения относно: сложност, нарастващи изисквания за памет и/или за висока изчислителна мощност.

Понастоящем алгоритмите за извличане на данни могат да откриват злонамерен софтуер и да го класифицират, но същевременно технологиите за разработване на зловреден софтуер се развиват непрекъснато. В тази връзка е изключително важно да се разработват нови алгоритми за извличане на данни, които да бъдат бързи и мащабируеми за откриване и класифициране на зловреден софтуер.

В заключение, киберсигурността е критична и жизненоважна за защита на данните, информацията и системите, а много области и приложения на изкуствения интелект могат да допринесат за повишаване на киберсигурността, като машинно обучение, дълбоко обучение, алгоритми за извличане на данни и експертни системи.

Ясно е, че машинното обучение е мощен инструмент, който може да се използва за автоматизиране на сложни кибердейности за защита и за атака. Следователно, тъй като киберпрестъпниците също се възползват от машинното обучение в техния арсенал от кибероръжия, се очаква да изпитаме по-сложни и мащабни атаки, използващи AI.

Следователно, е от жизненоважно значение специалистите по сигурността, както и по машинно обучение да бъдат в крак с последните постижения в машинното обучение, вкл. състезателно машинно обучение, за да използват потенциалните приложения за сигурност, свързани с AI.

Направеният обзор може да послужи като основа за бъдещи изследвания, които могат да се фокусират върху анализирането на съществуващите решения за сигурност и различните предизвикателства, при използването на машинното обучение за разработване и внедряване на мащабируеми системи за киберсигурност в различни сфери.

ЛИТЕРАТУРНИ ИЗТОЧНИЦИ (REFERENCES):

1. REGE, Manjeet and Raymond Blanch K. MBAH. Machine Learning for Cyber Defense and Attack. In: *Data Analytics 2018: The Seventh International Conference on Data Analytics* [online]. 2018, pp. 73-78 [viewed 6 December 2021]. ISBN 978-1-61208-681-1. Ghent University. Available from: https://biblio.ugent.be/publication/8603211
2. KABBAS, Azzah, Atheer ALHARTHI and Asmaa MUNSHI. Artificial Intelligence Applications in Cybersecurity. *International Journal of Computer Science and Network Security* [online]. 2020, vol. 20(2), pp. 120-124 [viewed 6 December 2021]. Available from: http://search.ijcsns.org/07_book/html/202002/202002016.html
3. TRUONG, Thanh Cong, Quoc Bao DIEP and Ivan ZELINKA. Artificial Intelligence in the Cyber Domain: Offense and Defense. *Symmetry* [online]. 2020, vol. 12(3), p. 410 [viewed 6 December 2021]. ISSN 2073-8994. MDPI. Available from: https://www.mdpi.com/2073-8994/12/3/410
4. *Artificial intelligence and cybersecurity: Opportunities and Challenges. Technical workshop summary report* [online]. National Science and Technology Council, 2020 [viewed 6 December 2021]. Available from: https://www.nitrd.gov/pubs/AI-CS-Tech-Summary-2020.pdf
5. СИМЕОНОВА, Ц. *Развитие на перспективните технологии в „Интернет на свързаните неща" IoT (Internet of Things)*. София: Асеневци, 2021. ISBN 978-619-7586-25-1.
6. СИМЕОНОВА, Ц. Особености на влиянието на човешкия фактор като източник на уязвимости върху информационната сигурност. *Сборник доклади от университетска научна конференция 28-29 май 2020 г., В. Търново.* Велико Търново: ИК на НВУ „Васил Левски", 2020, с. 2020-2030. ISSN 2367-7481.
7. СИМЕОНОВА, Ц. Методи за анализ и оценка на риска в областта на киберсигурността при системи SCADA. *Сборник доклади от университетска научна конференция 28-29 май 2020 г., В. Търново*. Велико Търново: ИК на НВУ „Васил Левски, 2020, с. 1998-2008. ISSN 2367-7481.
8. ПЕТРОВ, Г. *Развитие на Интернет и отворените системи.* Част 1. София: Авангард Прима, 2017. ISBN 978-619-160-834-8.
9. СИМЕОНОВА, Ц. и В. КЪДРЕВ. Развитие на изкуствения интелект и неговото приложение в телекомуникационните мрежи и услуги. *Сборник доклади от Научна конференция с международно участие на НВУ „В. Левски", В. Търново, 27-28.05.2021*. Велико Търново: ИК на НВУ „Васил Левски, 2021, с. 2376-2386. ISSN 2367-7481.
10. IVANOVA, Yoana. Assessment of the probability of cyberattacks on transport management systems. Publication of Union of Scientists in Bulgaria. *International Journal on Information Technologies and Security*. 2018, vol. 10(4), pp. 99-106. ISSN 1313-8251.
11. IVANOVA, Yoana. Modelling the Impact of Cyber Attacks on the Traffic Control Centre of an Urban Automobile Transport System by Means of Enhanced Cybersecurity. *MATEC Web of Conferences BulTrans-2017* [online]. 2017, vol. 133 [viewed 6 December 2021]. ISSN 2261-236X. Available from: https://doi.org/10.1051/matecconf/201713307001
12. СИМЕОНОВА, Ц. Прилагане на дървовидни методи за интегриране на оценката на сигурността и безопасността при управлението на риска. *Сборник доклади от университетска научна конференция 28-29 май 2020 г.* Велико Търново: ИК на НВУ „Васил Левски", 2020, стр. 2009-2019. ISSN 2367-7481.

Информация за авторите:
доц. д-р Васил Къдрев, НБУ департамент „Телекомуникации", vkadrev@nbu.bg
доц. д-р Росен Пасарелски, НБУ департамент „Телекомуникации", rpasarelski@nbu.bg

Contacts:
Assoc. Prof. Vasil Kadrev, PhD, New Bulgarian University, Department Telecommunications, vkadrev@nbu.bg
Assoc. Prof. Rosen Pasarelski, PhD, New Bulgarian University, Department Telecommunications, rpasarelski@nbu.bg

Дата на постъпване на ръкописа (Date of receipt of the manuscript): 10.06.2021
Дата на приемане за публикуване (Date of adoption for publication): 30.09.2021