Universidade Estadual de Londrina

ISRAEL FAUSTINO BOTELHO JUNIOR

TRANSFERÊNCIA DE CONHECIMENTO PARA DETECÇÃO DE ATAQUES EM INTERNET DAS COISAS

LONDRINA

2024

ISRAEL FAUSTINO BOTELHO JUNIOR

TRANSFERÊNCIA DE CONHECIMENTO PARA DETECÇÃO DE ATAQUES EM INTERNET DAS COISAS

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciência da Computação da Universidade Estadual de Londrina para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof(a). Dr(a). Bruno Bogaz Zarpelão

LONDRINA

2024

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Sobrenome, Nome.
Título do Trabalho : Subtitulo do Trabalho / Nome Sobrenome. - Londrina, 2017.
100 f. : il.

Orientador: Nome do Orientador Sobrenome do Orientador.
Coorientador: Nome Coorientador Sobrenome Coorientador.
Dissertação (Mestrado em Ciência da Computação) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Ciência da Computação, 2017.
Inclui bibliografia.

1. Assunto 1 - Tese. 2. Assunto 2 - Tese. 3. Assunto 3 - Tese. 4. Assunto 4 - Tese. I. Sobrenome do Orientador, Nome do Orientador. II. Sobrenome Coorientador, Nome Coorientador. III. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação. IV. Título.

ISRAEL FAUSTINO BOTELHO JUNIOR

TRANSFERÊNCIA DE CONHECIMENTO PARA DETECÇÃO DE ATAQUES EM INTERNET DAS COISAS

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciência da Computação da Universidade Estadual de Londrina para obtenção do título de Bacharel em Ciência da Computação.

BANCA EXAMINADORA

Orientador: Prof(a). Dr(a). Bruno Bogaz Zarpelão
Universidade Estadual de Londrina

Prof. Dr. Segundo Membro da Banca
Universidade/Instituição do Segundo Membro da Banca – Sigla instituição

---

Prof. Dr. Terceiro Membro da Banca
Universidade/Instituição do Terceiro Membro da Banca – Sigla instituição

---

Prof. Ms. Quarto Membro da Banca
Universidade/Instituição do Quarto Membro da Banca – Sigla instituição

Londrina, de 2024.

Este trabalho é dedicado às crianças adultas
que, quando pequenas, sonharam em se
tornar cientistas.

AGRADECIMENTOS

Os agradecimentos principais são direcionados à Gerald Weber, Miguel Frasson, Leslie H. Watter, Bruno Parente Lima, Flávio de Vasconcellos Corrêa, Otavio Real Salvador, Renato Machnievscz[1] e todos aqueles que contribuíram para que a produção de trabalhos acadêmicos conforme as normas ABNT com LaTeX fosse possível.

Agradecimentos especiais são direcionados ao Centro de Pesquisa em Arquitetura da Informação[2] da Universidade de Brasília (CPAI), ao grupo de usuários *latex-br*[3] e aos novos voluntários do grupo *abnTeX2*[4] que contribuíram e que ainda contribuirão para a evolução do abnTeX2.

[1] Os nomes dos integrantes do primeiro projeto abnTeX foram extraídos de <http://codigolivre.org.br/projects/abntex/>

[2] <http://www.cpai.unb.br/>

[3] <http://groups.google.com/group/latex-br>

[4] <http://groups.google.com/group/abntex2> e <http://abntex2.googlecode.com/>

"Não vos amoldeis às estruturas deste mundo, mas transformai-vos pela renovação da mente, a fim de distinguir qual é a vontade de Deus: o que é bom, o que Lhe é agradável, o que é perfeito.
(Bíblia Sagrada, Romanos 12, 2))

BOTELHO, I. F. B.. **Transferência de conhecimento para detecção de ataques em Internet das Coisas**. 2024. 31f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual de Londrina, Londrina, 2024.

RESUMO

Com a crescente disponibilidade de circuitos integrados baratos e o acesso à Internet remoto, o número de dispositivos IoT (Internet das Coisas) aumentou significativamente nos últimos anos. No entanto, esse crescimento também trouxe um aumento nas ameaças à segurança desses dispositivos, tornando até mesmo objetos cotidianos, como lâmpadas e geladeiras, alvos de ataques. A detecção desses ataques é necessária para garantir a integridade desses dispositivos, um dos recursos que vem sendo explorado para essa função é o aprendizado de máquina, que traz diversas técnicas e algoritmos capazes de automatizar o processo de detecção de ataques. Além disso, existe a dificuldade de treinar modelos de treinar modelos com exemplos de situações de ataque, já que essas amostras não são tão fáceis de encontrar ou produzir. Dessa forma, reaproveitar modelos já treinados seria uma forma de superar este obstáculo. Nesse contexto, este trabalho tem como objetivo explorar técnicas de transferência de conhecimento para reaproveitar modelos de aprendizado de máquina de dispositivos semelhantes, reduzindo o custo e o esforço necessários para treinar novos modelos. Essa abordagem promete melhorar a segurança e a eficácia da detecção de ataques em dispositivos IoT.

Palavras-chave: Transferência de conhecimento, Internet das coisas, Aprendizado de máquina, Aprendizado profundo.

BOTELHO, I. F. B.. **Transfer learning to detect attacks on Internet of Things devices**. 2024. 31p. Final Project (Bachelor of Science in Computer Science) – State University of Londrina, Londrina, 2024.

ABSTRACT

The growing availability of affordable integrated circuit and remote internet access has led to a substantial increase in the number of IoT (Internet of Things) devices in recent years. However, this growth has also witnessed a surge in security threats targeting these devices, making even everyday objects like lamps and refrigerators vulnerable to attacks. Detecting these attacks is necessary to ensure the integrity of these devices, and one of the resources being explored for this function is machine learning, which offers various techniques and algorithms capable of automating the attack detection process. Additionally, there is the challenge of training models with examples of attack situations, as such samples are not easy to find or produce. Thus, reusing already trained models would be a way to overcome this obstacle. In this context, this work aims to explore knowledge transfer techniques to reuse machine learning models from similar devices, reducing the cost and effort needed to train new models. This approach promises to improve the security and effectiveness of attack detection in IoT devices.

Keywords: Transfer learning, Internet of Things, Machine learning, Deep learning.

LISTA DE ILUSTRAÇÕES

Figura 1 – Arquitetura de três camadas de IoT. Fonte: Wu et al.[1] 16
Figura 2 – Representação do modelo de transferência de conhecimento. Fonte: Ribani e Marengoni[2] 21

LISTA DE TABELAS

Tabela 1 – Uso de diferentes abordagens nas diferentes configurações. Fonte: Pan e Yang[3] 22

LISTA DE ABREVIATURAS E SIGLAS

IoT	*Internet of Things*
IDS	*Intrusion Detection System*
IA	Inteligência Artificial
TL	*Transfer Learning*

SUMÁRIO

1 INTRODUÇÃO 13

2 FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA E ESTADO DA ARTE 15
2.1 Internet das Coisas 15
2.2 Vulnerabilidades e Ataques em IoT 16
2.3 Detecção de Ataques 18
2.4 Aprendizado de Máquina 19
2.5 Transferência de Conhecimento 20
2.6 Trabalhos relacionados 22

3 CONCLUSÃO 25

REFERÊNCIAS 26

APÊNDICES 30

ANEXOS 31

1 INTRODUÇÃO

Sistemas de Internet das Coisas (IoT - *Internet of Things*) estão amplamente presentes nos dias de hoje, em 2023 segundo a Amazon[1], houveram mais de 17 milhões de dispositivos conectados à Alexa, uma assistente de voz. Esses sistemas descrevem uma rede de objetos conectados a sensores, programas ou outras tecnologias, com o objetivo de trocar informações com outros sistemas, geralmente por meio da Internet [4]. O uso de IoT varia desde eletrodomésticos, como lâmpadas e geladeiras, até veículos aéreos não tripulados e sensores em barragens. Devido à diversidade desses sistemas e a sua alta versatilidade, houve um aumento significativo na sua fabricação e uso nos últimos anos, atingindo 14,3 bilhões de dispositivos em 2023 [5].

No entanto, apesar de sua ampla utilização, a segurança desses dispositivos é precária, tornando-os suscetíveis a ataques. Diversas vulnerabilidades são comumente encontradas, como senhas fracas, *backdoors*, falha de autenticação, entre outras [6]. Um reflexo disso é o aumento expressivo de *malwares* voltados para IoT, que alcançou 112,3 milhões de instâncias de *malware* em 2022, o que representa um aumento de 66% em relação a 2021 [7].

Normalmente, pacotes transmitidos e recebidos por dispositivos conectados à rede apresentam um padrão de comportamento. Por esse motivo, a partir do momento em que ocorre um ataque, é possível diferenciar sua atividade das atividades normais [8]. Uma forma comum de detectar essas atividades é por meio do uso de algoritmos de aprendizado de máquina, que conseguem analisar e descrever comportamentos com base em padrões [9].

Contudo, os sistemas IoT são geralmente compostos por dispositivos heterogêneos, ou seja, com diferentes configurações e fabricantes, o que resulta na falta de padronização entre eles [9]. Além disso, modelos de aprendizado de máquina são treinados para aplicações específicas, o que requer muito tempo computacional e um grande conjunto de dados [10, 11]. Portanto, ao alterar o ambiente no qual o sistema se encontra, a acurácia das detecções de ataques pode diminuir e até mesmo exigir um novo treinamento [11].

Uma forma de reduzir a necessidade de novos treinamentos é através da transferência de conhecimento, um método de aprendizado de máquina. O método utiliza informações de uma tarefa, como um modelo de classificação, e aplica em outra tarefa, permitindo o acúmulo de conhecimentos de diferentes modelos previamente treinados [11, 12]. Como resultado da aplicação desse modelo, pode-se aproveitar a similaridade entre os dispositivos IoT para reduzir o tempo e custo necessários para o treinamento de modelos em

[1] aboutamazon.com.br/noticias/dispositivos/numeros-mostram-o-quanto-alexa-conquistou-os-coracoes-dos-brasileiros

novos ambientes [10].

Diante disso, o principal objetivo deste trabalho é explorar e analisar diferentes abordagens de aprendizado de máquina com transferência de conhecimento, a fim de tornar o treinamento de modelos para sistemas IoT mais eficiente e preciso. Para alcançar esse propósito, inicialmente será realizado um levantamento bibliográfico para investigar a literatura existente sobre o tema. Em seguida, serão identificadas e selecionadas as abordagens de aprendizado de máquina e transferência de conhecimento mais adequadas para a problemática da segurança de sistemas IoT. Após isso, será feita a seleção e identificação de conjuntos de dados capaz de fornecer diversas situações de testes, e por fim, serão realizados testes capazes de medir a acurácia com ou sem a transferência de conhecimento.

2 FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA E ESTADO DA ARTE

2.1 Internet das Coisas

Internet das Coisas, também conhecida como *Internet of Things* (IoT), é um termo criado em 1999 por Kevin Ashton, que descreve um sistema no qual objetos do mundo real podem ser conectados à Internet por meio de sensores [13]. Atualmente, devido ao advento de circuitos integrados de computadores baratos, o termo se tornou popular e abrange diversos objetos do dia a dia que tiveram suas funções integradas à Internet [13, 14].

Um sistema IoT normalmente funciona coletando e trocando dados em tempo real, utilizando diversos objetos ao nosso redor, como sensores e dispositivos inteligentes. Com a popularização do conceito de casa inteligente, muitos dispositivos cotidianos passaram a ser integrados à Internet, resultando em um aumento na disponibilidade de dados. Alguns desses objetos que podemos encontrar facilmente em residências incluem lâmpadas inteligentes, geladeiras inteligentes, assistentes de voz, entre outros.

A arquitetura mais bem aceita pode ser visualizada na Figura 1 e é dividida em três camadas: percepção, rede e aplicação.

- **Camada de percepção:** A camada de percepção atua como a interface inicial entre o mundo físico e o mundo digital da IoT. Sua responsabilidade é identificar objetos e coletar informações por meio de sensores e dispositivos [1, 15].
- **Camada de rede:** A camada de rede é responsável por transmitir as informações obtidas pela camada física. Para isso, utiliza protocolos de rede, como TCP/IP [1, 15].
- **Camada de aplicação:** A camada de aplicação representa a interface final da IoT com os usuários e as aplicações específicas. Ela se concentra em atender às demandas das indústrias e das necessidades sociais, possibilitando a aplicação inteligente da IoT em diversos cenários [1, 15].

Além disso, há diversos modelos de comunicação para dispositivos, que estabelecem regras e estruturas para que eles possam interagir entre si e com outras aplicações [13]. Esses modelos de comunicação são classificados em: comunicação dispositivo-para-dispositivo (*Device-to-Device Communication*), comunicação dispositivo-para-*Gateway* (*Device-to-Gateway Communication*) e comunicação dispositivo-para-internet (*Device-to-Internet Communication*).

Figura 1 – Arquitetura de três camadas de IoT. Fonte: Wu et al.[1]

O modelo de comunicação dispositivo-para-dispositivo ocorre quando um dispositivo se comunica diretamente com outro, sem a necessidade de utilizar um serviço intermediário. Diferentes métodos podem ser empregados para possibilitar essa comunicação, como Bluetooth e Wi-Fi. Por outro lado, o modelo de comunicação dispositivo-para-*Gateway* ocorre quando se utiliza um serviço auxiliar ou um *Gateway*, um dispositivo capaz de conectar duas redes, para permitir que um dispositivo IoT se comunique com a internet ou faça uso de serviços em nuvem. Um exemplo de comunicação dispositivo-para-*Gateway* ocorre quando diversas lâmpadas inteligentes são conectadas a um roteador, permitindo controla-las por um telefone. Por fim, o modelo de comunicação dispositivo-para-internet ocorre quando um dispositivo consegue acessar a internet ou um serviço em nuvem diretamente [13, 16].

2.2 Vulnerabilidades e Ataques em IoT

Ataques no âmbito da cibersegurança podem ser definidos como uma tentativa de contornar dos serviços de segurança e violar as politicas de segurança de um sistema, tendo como objetivo afetar sua operação ou fazer usos de suas informações [17]. Essas tentativas muitas vezes envolvem técnicas de engenharia social, ou o uso de *malware* e a exploração de vulnerabilidades.

Em sistemas IoT, vulnerabilidades podem ser encontradas devido à falta de planejamento voltado para a segurança, muitas vezes causada pelo barateamento do custo de produção e pela falta de atualizações que garantam uma maior integridade da sua segurança [18]. De acordo com [19] existem 9 classes de vulnerabilidades:

1. **Captação insuficiente de energia:** Dispositivos IoT têm energia disponível limi-

tada e não necessariamente possuem meios de recuperá-la automaticamente, possibilitando que um invasor gaste essa energia com ataques, deixando o dispositivo inoperante.

2. **Autenticação inadequada:** Devido às restrições que afetam seu poder computacional, dispositivos IoT frequentemente empregam um mecanismo de autenticação fraco. Esse mecanismo pode ser explorado por invasores para violar a integridade dos dados e adicionar nós maliciosos na rede.

3. **Encriptação imprópria:** Com as restrições de recursos dos dispositivos, a eficácia e eficiência de algoritmos de encriptação tendem a cair, possibilitando brechas para que invasores os contornem e tenham acesso a informações sensíveis [19].

4. **Portas abertas desnecessárias:** Diversos dispositivos IoT deixam portas abertas sem necessidade enquanto executam serviços vulneráveis. Essa prática permite que um invasor por meio de ataque acesse a porta e conecte-se ao dispositivo.

5. **Controle de acesso insuficiente:** O uso de senhas pouco complexas, a ampla utilização de senhas padrão e a concessão desnecessária de permissões a certos usuários dos dispositivos IoT possibilitam que um invasor as descubra e obtenha acesso não autorizado ao dispositivo.

6. **Capacidade imprópria de gerenciamento de versões:** Idealmente, os dispositivos IoT deveriam manter seu *firmware* e programas atualizados para minimizar vetores de ataques, forma de invasores entrar no sistema, e melhorar seu funcionamento. No entanto, a maioria dos fabricantes não fornece atualizações de segurança e, quando o fazem, muitas vezes falta garantia de integridade, o que permite que sejam modificadas maliciosamente.

7. **Práticas incorretas de programação:** Muitos *firmwares* desenvolvidos para dispositivos IoT são programados com falhas de segurança conhecidas. Essas práticas inadequadas de programação permitem que invasores explorem facilmente as falhas, ganhando acesso às informações e até mesmo ao dispositivo.

8. **Mecanismos insuficientes de auditoria:** Muitos dispositivos IoT falham em oferecer um procedimento de registro das atividades, o que muitas vezes permite que atividades maliciosas ocorram de forma silenciosa.

Portanto, de acordo com [20], há quatro formas de atacar explorando vulnerabilidades nesses dispositivos:

- **Ataques físicos**: São mais concentrados na parte física dos dispositivos e requer estar fisicamente próximo a eles. De modo geral, esses ataques são voltados em

comprometer a integridade física do equipamento, interferir nos protocolos de comunicação como no *RF Interference on RFIDs* ou modificando códigos e programas como no *Malicious Code Injection*.

- **Ataques de rede**: Estes são mais focados em interferir nos protocolos de rede aplicados e precisa que o invasor tenha acesso à rede dos dispositivos. Esses ataques podem ler e interceptar mensagens com o objetivo de obter informações sensíveis como o *Traffic Analysis Attack* e o *Man in the Middle Attack* ou derrubar um serviço ou deixá-lo inoperante como o *Denial of Service*.

- **Ataques de programas**: São realizados por programas maliciosos, tais como vírus, *spyware* e cavalo de Troia. Por serem realizados de dentro do dispositivo, inicialmente é necessário infectá-lo. Os ataques dessa categoria podem variar consideravelmente, dependendo do programa utilizado. Por exemplo o Mirai transforma o dispositivo infectado em um *bot* que obedece aos comandos do invasor para realizar ataques de negação de serviço distribuídos (DDoS) [21].

- **Ataques de criptografia**: Estes ataques consistem em descobrir a chave privada utilizada pelo aparelho em suas mensagens. Diversas abordagens de criptoanálise podem ser implementadas e dependem muito das falhas em cada método de criptografia. Por exemplo um ataque do tipo *Man in the Middle* possibilita que o atacante intercepte a troca de chaves entre dispositivos e conseguir a chave.

2.3 Detecção de Ataques

A detecção de ataques em dispositivos de IoT é uma tarefa necessária para garantir a segurança desses sistemas altamente conectados. Existem diversas ameaças que podem visar os dispositivos IoT, e para cada tipo de ataque, é necessário encontrar uma solução adequada. No entanto, a simples combinação de todas essas soluções, quando implementadas, pode impactar negativamente o desempenho geral dos dispositivos IoT, o que se torna um dos problemas de segurança nesse contexto [22].

Para lidar com esse desafio, foram propostas diferentes técnicas e ferramentas que se concentram em estudar o comportamento do sistema. Uma dessas ferramentas é o sistema de detecção de intrusão (IDS - *Intrusion Detection System*), cujo propósito principal é detectar qualquer atividade suspeita que ocorra na rede alvo [23]. Com essa finalidade, sensores são posicionados em locais estratégicos para a captura de tráfego de rede, cabeçalhos de pacotes, requisições de serviço, mudanças de arquivos e chamadas de sistema [24].

No IDS, é possível a detecção com base em padrões de ataques previamente conhecidos (detecção por assinatura). Nesse método, o sistema identifica um ataque comparando

os traços da atividade na rede com os traços de ataques pré-instalados no banco de dados do IDS. Esse método permite a identificação eficaz de ataques conhecidos, mas caso seja um novo tipo de ataque, o IDS não é capaz de identificá-lo [23, 25].

Também é possível a detecção com base em comportamentos fora do padrão no tráfego de rede (detecção por anomalia). Nesse método, o sistema identifica qualquer anomalia na rede, analisando seu comportamento e verificando se a atividade foge do padrão normal. Esse método permite a detecção de vários tipos de ataques sem a necessidade de conhecimento prévio, no entanto, pode apresentar várias análises imprecisas, como falsos positivos [23, 25].

Além disso, é possível a detecção com base em regras estabelecidas pelo sistema, como a detecção baseada em especificação. Esse método consiste em identificar ataques quando sua atividade não está em conformidade com as especificações criadas. Esse método permite distinguir comandos inesperados de ataques, mas caso o ataque respeite as regras, ele passa despercebido [23, 25].

Outra técnica amplamente utilizada para a detecção de ataques em IoT é o uso de aprendizado de máquina. Nesse caso, os dados coletados dos dispositivos IoT são usados para treinar modelos de aprendizado de máquina. Esses modelos são então empregados para observar e detectar possíveis ataques na rede IoT [22].

2.4 Aprendizado de Máquina

O aprendizado de máquina é uma subárea da inteligência artificial (IA) e um ramo em ascensão dos algoritmos computacionais. Ele foi projetado para imitar a inteligência humana e aprender com o ambiente ao seu redor [26].

Neste campo, utilizam-se algoritmos e modelos estatísticos para aprender a executar tarefas sem depender de instruções explícitas [27]. Esses algoritmos são construídos a partir de modelos gerais com parâmetros ajustáveis que, ao receberem diferentes valores, realizam diversos cálculos para otimizar os critérios de desempenho [28].

O processo de aprendizagem é dividido em duas etapas principais: treinamento e inferência. Durante a fase de treinamento, o processo de aprendizagem ocorre de forma repetitiva e incremental. Os algoritmos processam exemplos um após o outro, ajustando gradualmente os parâmetros do modelo para aprimorar o desempenho. Na fase de predição, os exemplos são processados com base nos parâmetros calculados, resultando em saídas rotuladas [28].

Além disso, o aprendizado de máquina é categorizado, com base em informações passadas durante o treinamento, em três categorias principais: supervisionado, não supervisionado e semi-supervisionado.

- **Aprendizado de máquina supervisionado:** No aprendizado de máquina supervisionado, utiliza-se um conjunto de dados pré-definidos para estimar um valor desconhecido. O treinamento envolve uma série de valores em pares ordenados (entrada, saída), onde a saída é rotulada com o resultado esperado [29]. Isso resulta em valores numéricos contínuos (regressão) ou valores discretos que preveem rótulos (classificação) [30, 31].

- **Aprendizado de máquina não-supervisionado:** No aprendizado de máquina não supervisionado, não são usados rótulos. Durante o treinamento, apenas as informações de entrada são processadas, sem qualquer saída associada a um resultado. No entanto, é possível construir uma estrutura formal para esse modelo com base no estabelecimento de relações entre as entradas [32]. Isso é comumente usado em tarefas como agrupamento, regras de associação e redução de dimensionalidade [31, 32, 33].

- **Aprendizado de máquina semi-supervisionado:** O aprendizado de máquina semi-supervisionado, por sua vez, é uma combinação das abordagens supervisionada e não supervisionada, utilizando tanto dados rotulados quanto não rotulados [26, 31].

2.5 Transferência de Conhecimento

Humanos têm intrinsecamente a habilidade de usar o que aprenderam em uma tarefa para aprender algo similar, utilizando por exemplo o conhecimento de andar de bicicleta para aprender a dirigir uma moto. Inspirado por esse mecanismo, a transferência de conhecimento (TL - *Transfer Learning*) tem como objetivo utilizar conhecimentos adquiridos por um treinamento de aprendizado de máquina em outro, visando melhorar seu desempenho e precisão ou reduzir o número necessário de rótulos no modelo alvo [34].

A transferência de conhecimento pode ser descrita a partir de dois conceitos centrais:

- **Domínio**: Um domínio $\mathcal{D}$ é composto por um espaço de características $\mathcal{X}$ e uma distribuição marginal $P(X)$, onde $\mathcal{X}$ é um conjunto contendo todas as possíveis características [3, 34]. Por exemplo, em um cenário de análise do tráfego de rede, cada métrica possível é considerada uma característica e faz parte do conjunto $\mathcal{X}$.

- **Tarefa**: Uma tarefa $\mathcal{T}$ é composta pelo conjunto de todos os possíveis rótulos $\mathcal{Y}$ e um modelo ou função preditiva $f(\cdot)$ que prediz um rótulo com base no domínio onde é aplicada [3, 34]. A tarefa é aprendida com base no domínio dado e prediz um rótulo para cada valor x dele. Alguns exemplos clássicos de tarefas são: classificação, regressão e agrupamento.

Formalmente, a transferência de conhecimento é descrita como: dado um domínio de origem $\mathcal{D}_S$ e sua tarefa $\mathcal{T}_S$ e um domínio alvo $\mathcal{D}_T$ e sua tarefa $\mathcal{T}_T$, a transferência de conhecimento é o processo de ajudar a melhorar a função $f_T(\cdot)$ na tarefa alvo com o conhecimento adquirido em $\mathcal{D}_S$ e $\mathcal{T}_S$, onde $\mathcal{D}_S \neq \mathcal{D}_T$ ou $\mathcal{T}_S \neq \mathcal{T}_T$ [2]. Isso significa que, com o conhecimento adquirido após aprender a tarefa $\mathcal{T}_S$, podemos melhorar a função de predição da tarefa $\mathcal{T}_T$, processo ilustrado na Figura 2.

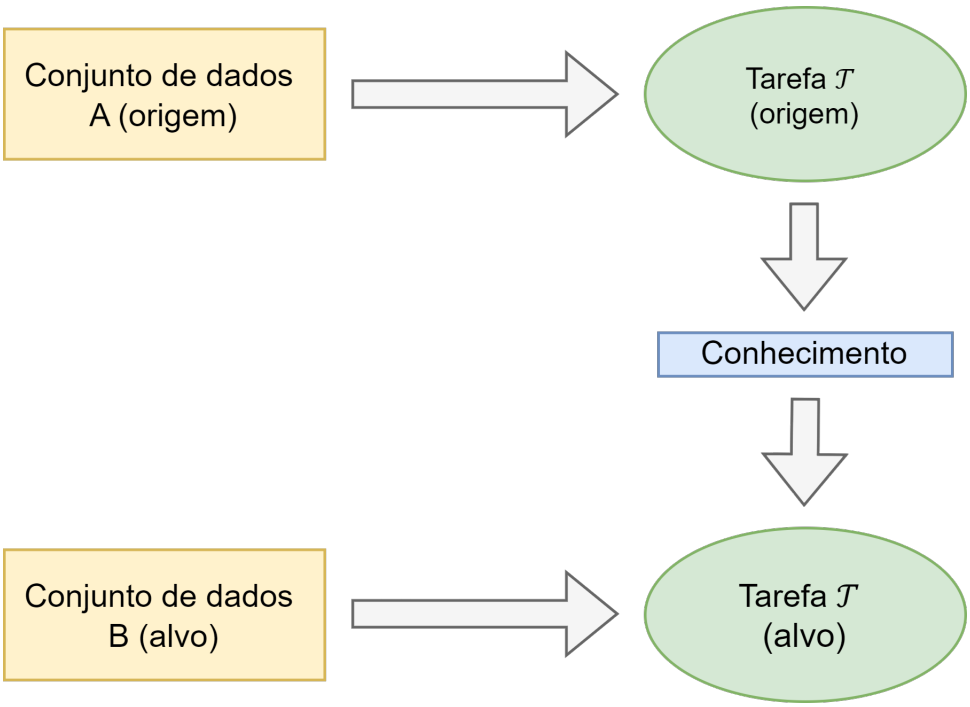

Figura 2 – Representação do modelo de transferência de conhecimento. Fonte: Ribani e Marengoni[2]

No trabalho realizado por [3] a transferência de conhecimento é categorizada em três configurações baseadas na diferença entre os domínios e tarefas de origem e alvo, sendo elas:

- **Transferência de conhecimento indutiva**: Diz respeito ao cenário em que $\mathcal{T}_S$ e $\mathcal{T}_T$ são diferentes e os domínios podem ou não ser iguais. Diante disso, são necessários alguns dados rotulados no domínio alvo para induzir um modelo objetivo preditivo $f_T(\cdot)$. Além disso, nessa configuração, dependendo da disponibilidade de rótulos no domínio alvo, pode-se categorizar em dois casos: um similar ao aprendizado multitarefa, que tem como objetivo aprender duas ou mais tarefas simultaneamente e outro ao aprendizado auto-didata, que visa utilizar conceitos de aprendizados não-supervisionados para aprimorar a tarefa supervisionada dada[35].
- **Transferência de conhecimento transdutiva**: Ocorre quando $\mathcal{T}_S$ e $\mathcal{T}_T$ são iguais, mas os domínios são diferentes. Nesse caso, o domínio alvo não possui dados rotulados, enquanto muitos dados rotulados existem no domínio de origem. Os dois

casos possíveis são quando os espaços de características são diferentes ou quando as distribuições são diferentes.

- **Transferência de conhecimento não-supervisionada**: Acontece quando $\mathcal{T}_S$ e $\mathcal{T}_T$ são diferentes, mas diferente da indutiva essa configuração tem como foco resolver problemas não-supervisionados.

Com base nas configurações descritas acima, [3] também descreve quatro tipos de abordagens podem ser aplicadas aos problemas de transferência de conhecimento:

- **Transferência de instância**: Essa abordagem consiste em reutilizar parte dos dados do domínio de origem no domínio alvo. Também descrita como rebalancear, esse método visa aprimorar a tarefa alvo inserindo dados da origem.
- **Transferência de representação de características**: Essa abordagem consiste em encontrar uma boa representação das características visando minimizar as divergências entre os domínios. Uma vez encontrada, a representação pode ser enviada para a tarefa alvo.
- **Transferência de parâmetros**: Essa abordagem parte da ideia que modelos individuais para tarefas semelhantes compartilham alguns parâmetros comuns ou distribuições anteriores de hiper-parâmetros dos modelos. Por exemplo pesos da função de perda ou termos específicos de uma determinada tarefa.
- **Transferência de conhecimento relacional**: Essa abordagem lida com problemas de domínios relacionais, que diferente das outras abordagens não assume que os dados são independentes e identicamente distribuídos. Nesse método é transferido as relações dos dados do domínio de origem para o domínio alvo.

É possível ver onde cada abordagem pode ser aplicada na Tabela 1.

Abordagem	TL indutiva	TL transdutiva	TL não-supervisionada
Transferência de instância	x	x	
Transferência de representação de características	x	x	x
Transferência de parâmetros	x		
Transferência de conhecimento relacional	x		

Tabela 1 – Uso de diferentes abordagens nas diferentes configurações. Fonte: Pan e Yang[3]

2.6 Trabalhos relacionados

As ideias por trás da transferência de conhecimento foram propostas pela primeira vez em 1995 em uma oficina do NIPS-95[1] chamada "*Learning to Learn*". Nos últimos anos,

[1] Conteúdo disponível em: socrates.acadiau.ca/courses/comp/dsilver/NIPS95_LTL/

a pesquisa em detecção de ataques em IoT começou a considerar o uso de transferência de conhecimento devido à baixa capacidade computacional e à escassez de dados para treinar modelos complexos para esses dispositivos.

No trabalho de Khoa et al. [36], a transferência de conhecimento é utilizada para contornar a indisponibilidade de rótulos. No modelo apresentado, são propostas duas redes que operam aprendizado federado de forma independente, onde a primeira possui rótulos abundantes e a segunda não possui. A principal contribuição desse trabalho vem do fato de que as redes não precisam ter as mesmas estruturas de dados, permitindo uma aplicação mais realista. Algo similar ocorre no trabalho de Mehedi et al.[37], que propõem um modelo de rede neural residual baseado em transferência de conhecimento profunda (P-ResNet). Este modelo demonstra uma capacidade de adaptação eficaz em cenários com pouca informação rotulada disponível.

Além disso, outro problema que os modelos de aprendizado de máquina podem enfrentar é o desbalanceamento dos dados. No estudo conduzido por Chen et al. [38], é proposto um modelo de adaptação de domínio adversário aprimorada por informações (IADA), baseado no treinamento de redes neurais com domínio adversário (DANN). Este modelo busca mitigar os desafios apresentados pelo desbalanceamento dos dados ao melhorar a capacidade de adaptação entre diferentes domínios por meio de técnicas avançadas de aprendizado adversário.

No trabalho de Chen et al. [39], é explorado o uso de processamento de linguagem natural em conjunto com transferência de conhecimento profunda para criar um Sistema de Detecção de Intrusões (IDS) capaz de se adaptar mais eficientemente a uma rede heterogênea. Nele o uso de incorporação de palavra (WE - *Word Embedding*) contorna as diferenças nas características alinhando os domínios.

Outro cenário que explora muito bem os conceitos de transferência de conhecimento é o seu uso para acelerar e tornar mais eficientes os treinamentos em dispositivos com restrições de recursos. No trabalho realizado por Yilmaz et al. [40], são abordados os dispositivos LLN (Low Power and Lossy Networks), que possuem diversas limitações. Neste estudo, a programação genética foi utilizada para desenvolver algoritmos de detecção de intrusão mais adaptáveis, transferindo conhecimentos para novos aparelhos e para novos tipos de ataques. Isso resultou em uma abordagem mais rápida e eficiente, com maior precisão em relação aos modelos tradicionais.

No trabalho de Zhang et al. [41], a transferência de conhecimento é adotada na abordagem de transferência de instância, juntamente com o aprendizado federado, com o objetivo de criar um modelo leve para detecção de intrusão em que os dados não são independentes e identicamente distribuídos. Nesse modelo, é utilizado como base de treinamento um conjunto de dados público, gerando um modelo inicial que é distribuído aos clientes da rede.

Ademais, uma forma de tornar o treinamento mais eficiente é reutilizar modelos antigos para que os novos aprendizados não partam do zero. Tal ideia é explorada no trabalho de Santos et al. [42], onde são utilizadas técnicas de aprendizado por reforço em conjunto com aprendizado de máquina para gerar modelos de detecção de intrusão mais robustos e eficientemente atualizáveis.

Por fim, também são explorados outros usos para a transferência de conhecimento na literatura, como no trabalho de Wang et al. [43], que investiga diversas formas de aumentar a eficiência da sexta geração de comunicação sem fio (6G), e no trabalho de Otoum et al. [44], que estabelece uma analogia entre aprendizado federado e aprendizado dividido, visando comparar sua precisão.

3 CONCLUSÃO

REFERÊNCIAS

[1] WU, M. et al. Research on the architecture of Internet of Things. In: *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*. [S.l.: s.n.], 2010. v. 5, p. V5–484–V5–487. ISSN: 2154-7505.

[2] RIBANI, R.; MARENGONI, M. A Survey of Transfer Learning for Convolutional Neural Networks. In: *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*. [S.l.: s.n.], 2019. p. 47–57. ISSN: 2474-0705.

[3] PAN, S. J.; YANG, Q. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 10, p. 1345–1359, out. 2010. ISSN 1558-2191. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[4] O que é IoT? (Internet das Coisas) | Oracle Brasil. Disponível em: <https://www.oracle.com/br/internet-of-things/what-is-iot/>. Acessado em: 8 de junho de 2023.

[5] STATE of IoT 2023: Number of connected IoT devices growing 16 percent to 16.7 billion globally. Disponível em: <https://iot-analytics.com/number-connected-iot-devices/>. Acessado em: 8 de junho de 2023.

[6] WANG, A. et al. An inside look at iot malware. In: CHEN, F.; LUO, Y. (Ed.). *Industrial IoT Technologies and Applications*. Cham: Springer International Publishing, 2017. p. 176–186. ISBN 978-3-319-60753-5.

[7] SONICWALL. *SonicWall Cyber Threat Report*. [S.l.], 2023. Disponível em: <https://www.sonicwall.com/medialibrary/en/white-paper/2023-cyber-threat-report.pdf>.

[8] GHORBANI, A. A.; LU, W.; TAVALLAEE, M. *Network intrusion detection and prevention: concepts and techniques*. [S.l.]: Springer Science & Business Media, 2009. v. 47.

[9] GUL, M. J.; SYED, M. K.-u.-R. R. Network attack detection in iot using artificial intelligence. In: *2023 International Multi-disciplinary Conference in Emerging Research Trends (IMCERT)*. [S.l.: s.n.], 2023. I, p. 1–6.

[10] LIU, X. et al. Toward deep transfer learning in industrial internet of things. *IEEE Internet of Things Journal*, v. 8, n. 15, p. 12163–12175, 2021.

[11] VU, L. et al. Deep transfer learning for iot attack detection. *IEEE Access*, v. 8, p. 107335–107344, 2020.

[12] OLIVAS, E. S. et al. *Handbook of research on machine learning applications and trends: Algorithms, methods, and techniques: Algorithms, methods, and techniques*. [S.l.]: IGI global, 2009.

[13] ROSE, K.; ELDRIDGE, S.; CHAPIN, L. The Internet of Things: An Overview.

[14] O que é a Internet das Coisas (IoT). Disponível em: <https://aws.amazon.com/pt/what-is/iot/>. Acessado em: 9 de setembro de 2023.

[15] JING, Q. et al. Security of the Internet of Things: perspectives and challenges. *Wireless Networks*, v. 20, n. 8, p. 2481–2501, nov. 2014. ISSN 1572-8196. Disponível em: <https://doi.org/10.1007/s11276-014-0761-7>.

[16] NOUR, B. et al. A survey of Internet of Things communication using ICN: A use case perspective. *Computer Communications*, v. 142-143, p. 95–123, jun. 2019. ISSN 0140-3664. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0140366418309228>.

[17] STALLINGS, W. *Cryptography and network security: principles and practice*. 4th ed. ed. Upper Saddle River, N.J: Pearson/Prentice Hall, 2006. OCLC: ocm63126393. ISBN 978-0-13-187316-2.

[18] SIVARAMAN, V. et al. Smart IoT Devices in the Home: Security and Privacy Implications. *IEEE Technology and Society Magazine*, v. 37, n. 2, p. 71–79, jun. 2018. ISSN 1937-416X. Conference Name: IEEE Technology and Society Magazine. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8371556>.

[19] NESHENKO, N. et al. Demystifying IoT Security: An Exhaustive Survey on IoT Vulnerabilities and a First Empirical Look on Internet-Scale IoT Exploitations. *IEEE Communications Surveys & Tutorials*, v. 21, n. 3, p. 2702–2733, 2019. ISSN 1553-877X. Conference Name: IEEE Communications Surveys & Tutorials. Disponível em: <https://ieeexplore.ieee.org/document/8688434?denied=>.

[20] DEOGIRIKAR, J.; VIDHATE, A. Security attacks in IoT: A survey. In: *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. [s.n.], 2017. p. 32–37. Disponível em: <https://ieeexplore.ieee.org/document/8058363>.

[21] MARGOLIS, J. et al. An In-Depth Analysis of the Mirai Botnet. In: *2017 International Conference on Software Security and Assurance (ICSSA)*. [s.n.], 2017. p. 6–12. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8392610>.

[22] AL-TALEB, N.; SAQIB, N. A. Attacks Detection and Prevention Systems for IoT Networks: A Survey. In: *2020 International Conference on Computing and Information Technology (ICCIT-1441)*. [S.l.: s.n.], 2020. p. 1–5.

[23] AL-HADHRAMI, Y.; HUSSAIN, F. K. DDoS attacks in IoT networks: a comprehensive systematic literature review. *World Wide Web*, v. 24, n. 3, p. 971–1001, maio 2021. ISSN 1573-1413. Disponível em: <https://doi.org/10.1007/s11280-020-00855-2>.

[24] ZARPELãO, B. B. et al. A survey of intrusion detection in Internet of Things. *Journal of Network and Computer Applications*, v. 84, p. 25–37, abr. 2017. ISSN 1084-8045. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1084804517300802>.

[25] LIAO, H.-J. et al. Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, v. 36, n. 1, p. 16–24, jan. 2013. ISSN

1084-8045. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1084804512001944>.

[26] NAQA, I. E.; MURPHY, M. J. *What is machine learning?* [S.l.]: Springer, 2015.

[27] SARKER, I. H. et al. Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, v. 7, n. 1, p. 41, jul. 2020. ISSN 2196-1115. Disponível em: <https://doi.org/10.1186/s40537-020-00318-5>.

[28] ALPAYDIN, E. *Machine Learning, revised and updated edition*. [S.l.]: MIT Press, 2021. Google-Books-ID: 2nQJEAAAQBAJ. ISBN 978-0-262-36535-2.

[29] GEETHA, T.; SENDHILKUMAR, S. *Machine Learning: Concepts, Techniques and Applications*. [S.l.]: CRC Press, 2023.

[30] MICHIE, D.; SPIEGELHALTER, D. J.; TAYLOR, C. C. Machine learning, neural and statistical classification. 1994.

[31] AYODELE, T. O. Machine learning overview. *New Advances in Machine Learning*, IntechOpen, v. 2, p. 9–18, 2010.

[32] GHAHRAMANI, Z. Unsupervised learning. In: _____. *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. p. 72–112. ISBN 978-3-540-28650-9. Disponível em: <https://doi.org/10.1007/978-3-540-28650-9_5>.

[33] RAYMER, M. et al. Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, v. 4, n. 2, p. 164–171, 2000.

[34] WEISS, K.; KHOSHGOFTAAR, T. M.; WANG, D. A survey of transfer learning. *Journal of Big Data*, v. 3, n. 1, p. 9, maio 2016. ISSN 2196-1115. Disponível em: <https://doi.org/10.1186/s40537-016-0043-6>.

[35] RAINA, R. et al. Self-taught learning: transfer learning from unlabeled data. In: *Proceedings of the 24th international conference on Machine learning*. Corvalis Oregon USA: ACM, 2007. p. 759–766. ISBN 978-1-59593-793-3. Disponível em: <https://dl.acm.org/doi/10.1145/1273496.1273592>.

[36] KHOA, T. V. et al. Deep Transfer Learning: A Novel Collaborative Learning Model for Cyberattack Detection Systems in IoT Networks. *IEEE Internet of Things Journal*, v. 10, n. 10, p. 8578–8589, maio 2023. ISSN 2327-4662. Conference Name: IEEE Internet of Things Journal. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9868083>.

[37] MEHEDI, S. T. et al. Dependable Intrusion Detection System for IoT: A Deep Transfer Learning Based Approach. *IEEE Transactions on Industrial Informatics*, v. 19, n. 1, p. 1006–1017, jan. 2023. ISSN 1941-0050. Conference Name: IEEE Transactions on Industrial Informatics. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9749858>.

[38] CHEN, Y. et al. Cross-Domain Industrial Intrusion Detection Deep Model Trained With Imbalanced Data. *IEEE Internet of Things Journal*, v. 10, n. 1, p. 584–596, jan. 2023. ISSN 2327-4662. Conference Name: IEEE Internet of Things Journal. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9868087>.

[39] CHEN, D.; ZHANG, F.; ZHANG, X. Heterogeneous IoT Intrusion Detection Based on Fusion Word Embedding Deep Transfer Learning. *IEEE Transactions on Industrial Informatics*, v. 19, n. 8, p. 9183–9193, ago. 2023. ISSN 1941-0050. Conference Name: IEEE Transactions on Industrial Informatics. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9976274>.

[40] YıLMAZ, S.; AYDOGAN, E.; SEN, S. A Transfer Learning Approach for Securing Resource-Constrained IoT Devices. *IEEE Transactions on Information Forensics and Security*, v. 16, p. 4405–4418, 2021. ISSN 1556-6021. Conference Name: IEEE Transactions on Information Forensics and Security. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9478846>.

[41] ZHANG, J. et al. Federated Learning for Distributed IIoT Intrusion Detection Using Transfer Approaches. *IEEE Transactions on Industrial Informatics*, v. 19, n. 7, p. 8159–8169, jul. 2023. ISSN 1941-0050. Conference Name: IEEE Transactions on Industrial Informatics. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9927327>.

[42] SANTOS, R. R. d. et al. Reinforcement Learning for Intrusion Detection: More Model Longness and Fewer Updates. *IEEE Transactions on Network and Service Management*, v. 20, n. 2, p. 2040–2055, jun. 2023. ISSN 1932-4537. Conference Name: IEEE Transactions on Network and Service Management. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9893186>.

[43] WANG, M. et al. Transfer Learning Promotes 6G Wireless Communications: Recent Advances and Future Challenges. *IEEE Transactions on Reliability*, v. 70, n. 2, p. 790–807, jun. 2021. ISSN 1558-1721. Conference Name: IEEE Transactions on Reliability. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9388790>.

[44] OTOUM, S.; GUIZANI, N.; MOUFTAH, H. On the Feasibility of Split Learning, Transfer Learning and Federated Learning for Preserving Security in ITS Systems. *IEEE Transactions on Intelligent Transportation Systems*, v. 24, n. 7, p. 7462–7470, jul. 2023. ISSN 1558-0016. Conference Name: IEEE Transactions on Intelligent Transportation Systems. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9756883>.

Apêndices

Anexos