IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 13 July 2023, accepted 14 August 2023, date of publication 18 August 2023, date of current version 23 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3306452

RESEARCH ARTICLE

AI-Enabled Trust in Distributed Networks

ZHIQI LI[1,2], WEIDONG FANG[1,2,3], (Member, IEEE), CHUNSHENG ZHU[4], (Member, IEEE), ZHIWEI GAO[5], AND WUXIONG ZHANG[1,2], (Member, IEEE)

[1]Science and Technology on Microsystem Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 201899, China
[2]School of Electronic, Electrical, and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[3]Shanghai Research and Development Center for Micro-Nano Electronics, Shanghai 201210, China
[4]College of Big Data and Internet, Shenzhen Technology University, Shenzhen, Guangdong 518118, China
[5]Ceprei Certification Body, Fifth Electronics Research Institute of Ministry of Industry and Information Technology, Guangzhou, Guangdong 510610, China

Corresponding author: Weidong Fang (weidong.fang@mail.sim.ac.cn)

This work was supported in part by the Shanghai Natural Science Foundation under Grant 21ZR1461700, in part by the Shanghai Science and Technology Innovation Action Plan under Grant 21DZ1200600, in part by the National Key Research and Development Program under Grant 2020YFB2104202, and in part by the Special Projects for Key Research and Development Tasks in the Autonomous Region under Grant 2022B01009.

ABSTRACT Cybersecurity, as a crucial aspect of the information society, requires significant attention. Fortunately, the concept of trust, originating from the field of sociology, has been under extensive research in order to enhance cybersecurity by evaluating the trustworthiness of nodes with artificial intelligence (AI) techniques in distributed networks (DNs). However, the scalability issues faced by AI-enabled trust hinder its integration with the DNs. Currently, there is a lack of a comprehensive review article that explores the current state of AI-enabled trust development applications. This paper aims to address this gap by providing a review of the state-of-the-art AI-enabled trust in DNs. This review focuses on the concept of trust and how it can be facilitated through AI, particularly utilizing machine learning and deep learning methods. Additionally, the paper provides a comprehensive comparison and analysis of three key domains in the field of AI-enabled trust: trust management (TM), intrusion detection system (IDS), and recommender systems (RS). Some open problems and challenges that currently exist in the field are manifested, and some suggestions for future work are presented.

INDEX TERMS Artificial intelligence, machine learning, trust, distributed networks, cybersecurity.

I. INTRODUCTION

Distributed network is a prevalent form of network topology where nodes are spread across different locations, and there is no central node governing the network. This decentralized structure entails multiple terminals and offers several advantages, including enhanced stability, rapid processing speed, and flexibility. As a result of these benefits, distributed networks have found extensive applications over the past decades. In recent years, different types of networks have been extensively deployed across various domains of life. However, the rapid development of the information society has brought forth numerous security challenges. Consequently, hackers and malicious competitors have launched a significant number of network intrusions and attacks with the intention of disrupting targeted networks. This growing concern regarding the security of distributed networks can be observed through the search trend analysis of ‘‘Distributed Networks + Cyber Security’’ on Google from January 2010 to July 2023, as depicted in Figure 1 (source: Google Trends[1]).

Over the past decade, there has been a consistent growth in the search trend for cyber security in distributed networks, indicating a significant increase in concern for this topic. The reason behind this heightened interest is the potentially immeasurable or irreversible consequences of network crashes caused by malicious attacks. Thankfully, researchers have dedicated substantial efforts to addressing these cyber security issues. Numerous information security and privacy protection schemes have been proposed to mitigate these challenges. Examples include the use of encrypted transmissions and the configuration of firewalls or virtual private

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

[1] https://trends.google.com/trends/explore?date=2010-01-01%202023-07-05&q=Distributed%20Networks%20%2B%20Cyber%20Security

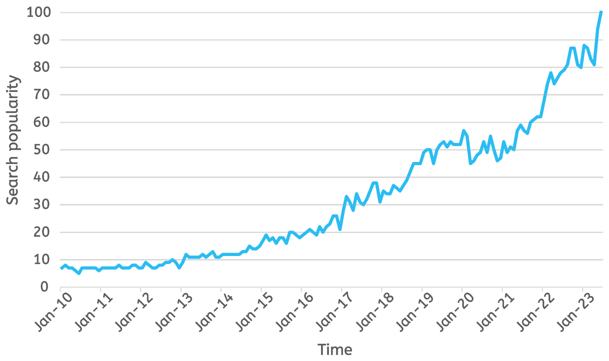

FIGURE 1. **The search trend of "distributed networks + cyber security" on google.**

networks (VPNs) in network interactions. These traditional schemes have proven effective in countering external attacks.

Despite the effectiveness of traditional security schemes against external attacks, there is still a need for further improvement, particularly when it comes to addressing internal attacks within DNs. To enhance network security in such scenarios, an increasing number of researchers have turned to the concept of "trust" for decision-making and intrusion prevention. The concept of trust originates from sociology and refers to the interpersonal belief that individuals will uphold their promises to each other, resulting in corresponding benefits [1]. Trust has garnered substantial attention across various fields, particularly in open applications based on networks such as peer-to-peer (P2P), ad hoc, web services, cloud computing, and the Internet of Things (IoT). The growing interest in trust indicates its intriguing and significant role in enhancing network security.

To enhance the accuracy and efficiency of trust evaluation in models, researchers employ various methods such as fuzzy algorithms or game theory. While game theory and other pure algorithms offer valuable insights, they can be complex and may not be easily programmed to address data-driven problems. In the fields of cybersecurity and communications, AI technologies have emerged as highly valuable [2], [3]. In the context of trust evaluation, AI can play a crucial role by analyzing historical data and improving the accuracy and efficiency of the evaluation process.

The three primary applications of trust are trust management (TM), intrusion detection system (IDS), and recommender system (RS). Recently, there has been a significant increase in the number of survey papers focusing on the applications of AI-enabled trust in cybersecurity [4], [5]. For example, many researchers have investigated existing trust management schemes and summarized trust evaluation methods used in the Internet of Things [6], [7]. Furthermore, J. Wang et al. conducted an excellent review on machine learning-based trust evaluation, systematically surveying the applications of machine learning (ML) in trust evaluation [8]. However, they did not analyze the machine learning-based trust evaluation in IDS and RS.

Therefore, this paper aims to provide readers with a comprehensive review paper that analyzes trust in the three main application domains in DNs: TM, IDS, and RS. We take the origin and nature of trust as a starting point, and systematically describe the general situation of trust and the advantages of AI-enabled trust. The paper explores the latest state-of-the-art works proposed within the past five years, compares them, in order to more accurately analyze current state of development of the field. Then, the advantages and disadvantages of these works are analyzed. The paper also highlights open challenges and suggests several insightful future work in the field.

The contributions of our work could be summarized as follows:

1) This paper categorizes and summarizes the advantages and disadvantages of commonly used machine learning and deep learning methods in AI-enabled trust in DNs.
2) This paper provides an overview of the state-of-the-art research on AI-enabled trust in DNs, focusing on three aspects: trust management, intrusion detection system, and recommender system.
3) Based on the analysis of related work proposed in the past five years, this paper concludes the existing open problems and challenges, and further proposes some suggestions for future work.

The remaining sections of the paper are organized as follows. In Section II, we provide a comprehensive review of the concept of trust in DNs and compare several commonly used machine learning and deep learning methods for trust evaluation. Section III focuses on the application of AI-enabled trust in three main domains: trust management, intrusion detection system, and recommender system. Recent related works and the utilized techniques are analyzed and compared, highlighting their advantages and disadvantages. Section IV identifies and discusses the existing open problems and challenges. Section V presents some perspectives for future work. Finally, in Section VI, the conclusion is drawn by summarizing the key findings and presenting the contributions made throughout the paper.

II. TRUST AND AI

Trust plays a vital role in strengthening interpersonal relationships and serves as the foundation for establishing social order. Over time, the study of trust has become an interdisciplinary field that encompasses various domains. It has been integrated into fields such as business management, economics, engineering, and computer science [9]. In DNs, the evaluation of trust values for network nodes is crucial for establishing trust relationships among these nodes.

Trust as a subjective and fuzzy parameter holds significant relevance in the interaction process between objects. Evaluating trust in these interactions serves as a foundation for making informed decisions to enhance the overall security of the interaction. Currently, AI is being employed to enhance the accuracy and efficiency of trust evaluation. Consequently, AI-enabled trust has emerged as a prominent research focus, garnering widespread attention within the field.

A. TRUST

In the following, we analyze the nature of trust in DNs by exploring three aspects: definition, properties, and evaluation. We provide a comprehensive definition of trust in the context of the network environment. Then, we discuss some properties of trust in DNs, examining its inherent characteristics. Lastly, we delve into the evaluation of trust in DNs, exploring various approaches used to assess trustworthiness.

1) DEFINITION

Trust is a subjective and abstract concept deeply rooted in sociology and psychology. the definition of trust can vary significantly due to differences in the field of study, the specific objects and subjects being considered, and the contextual factors involved [10]. Therefore, there is no unanimous consensus on a single, widely accepted definition of trust. Instead, researchers have the flexibility to define trust based on the specific scenario they are studying and identify the factors that influence it.

In 1993, P. Denning provided a definition of trust as an evaluation of the ability of a person, organization, or object to perform by given requirements on a behavioral domain [11]. Building on it, F. Azzedin further defined the trust in networks as follows: trust refers to the ability of an entity to change over time, along with its corresponding behavior [12].

Trust can be understood as a relationship that exists between nodes within a network. It can be characterized as the subjective probability or possibility of one node exhibiting the desired behavior as perceived by another node [13]. When the actions and behaviors of node B align with the expectations of node A, it can be said that node A trusts node B. In the context of node interactions, trust can be described as follows: Node B may be considered trustworthy by node A when node A believes that node B will strictly adhere to the expected and required behavior.

2) PROPERTIES

Trust in networks involves assumptions, expectations, and behaviors, making it a concept that relates to both subjective beliefs and objective reality. Drawing upon the definitions of trust in sociology and psychology, we can summarize the properties of trust in networks as follows:

- *Dynamicity*: Trust exhibits a dynamic and changeable nature influenced by both subjective and objective factors. The level of trust between parties tends to increase as the number of successful interactions grows. Conversely, trust diminishes when interactions result in failures or negative outcomes.
- *Subjectivity*: Trust is not solely determined by the historical behavior of the trustee, rather, it is also influenced by the subjective judgments of different trustors. These judgments can be influenced by various factors, including changes in the trustor's status or circumstances.
- *Hard to get, easy to lose*: When an interaction fails, the decrease in trust is typically greater than the increase in trust resulting from a successful interaction.
- *Unequal*: Due to the subjective nature of trust, the degree of trust between two entities may not be equal. It can vary depending on individual perceptions, experiences, and specific interactions.
- *Partial transferability*: Trust relationships are often transferable, meaning that if node A trusts node B and node B trusts node C, it does not necessarily imply that node A trusts node C. The transferability of trust is valid only under specific conditions and cannot be assumed in all cases.
- *Time-decaying*: The reliability of a trust value diminishes over time. When evaluating trust, the weight assigned to a trust value assessed further back in time should be lower compared to more recent trust values. By assigning a higher weight to recent evaluations, a more accurate representation of the current state of trust can be achieved.

Furthermore, there are other properties of trust that have not been explicitly listed. Researchers could delve deeper into the nature of trust within social interpersonal relationships and develop definitions that align more closely with real-life expectations.

3) EVALUATION

The fact that trust relationships cannot be fully automated means that quantitative measurement of trust is so difficult. However, trust can be measured as the level of authentication and access permissions. Trust is indeed quantifiable, and a trust degree serves as a quantitative representation of the trustworthiness of a trustee as perceived by the trustor. It reflects the trustor's assessment of the trustee's honesty, reliability, and the judgment of their future behavior. Historical interaction experiences can be utilized to obtain a trust degree, which can also be referred to as a trust value or trust rating.

Trust degrees can be represented as a binary variable in some cases. However, trust relationships are not limited to a binary distinction between trust and distrust. To provide a more detailed understanding of trust, researchers often quantify it using continuous trust values. Typically, trust values range between 0 and 1 [14], [15]. A trust value of 1 indicates complete trustworthiness of a node, while a value of 0 suggests complete untrustworthiness. Additionally, trust values can also be represented using discrete trust levels, providing further granularity in expressing the varying degrees of trust. These different representations of trust values allow for a more nuanced understanding and measurement of trust in different contexts.

Trust degree can be assessed by considering both the direct trust level (DTD) and indirect trust level (ITD). DTD represents the level of trust established through direct interactions and experiences between two entities. On the other hand,

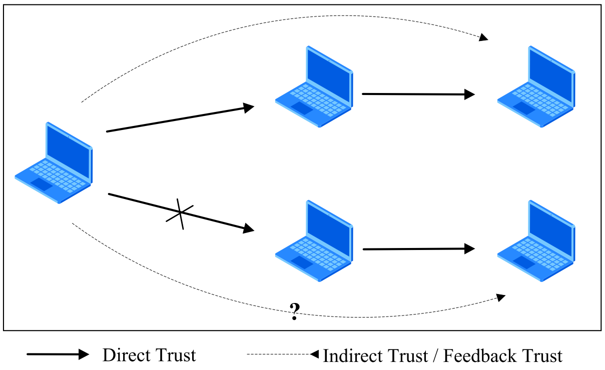

FIGURE 2. Direct trust & indirect trust.

ITD, also referred to as recommendation trust degree, feedback trust degree, or reputation, reflects the degree of trust between entities based on indirect recommendations from a third-party intermediary. However, the reliability of ITD can be challenging to guarantee due to the instability of third-party entities and the presence of potentially malicious intermediaries. These factors may introduce uncertainties into the trust evaluation process. There is a schematic diagram of direct trust and feedback trust evaluation in Figure 2.

The overall trust degree, also referred to as comprehensive trust degree or global trust degree, represents a comprehensive evaluation of the trustee's trustworthiness. It is derived by the trustor through a combination of DTD and ITD. Depending on the specific context and attributes of the monitoring events, various approaches can be employed to effectively represent and assess the trust degree of nodes, thus facilitating trust-related decision-making processes.

B. TRUST BEING POWERED BY AI

In distributed networks, the utilization of AI techniques can enhance the efficiency and accuracy of evaluating trust between nodes. In the subsequent discussion, we introduce several commonly employed AI techniques for trust evaluation.

1) MACHINE LEARNING

Machine learning is a subfield of artificial intelligence that leverages existing data to make predictions or responses to unfamiliar data in the future. It involves the use of computers to mimic the process of human learning through observation, enabling the development of systems that enhance their performance by leveraging historical data and experience [16].

ML methods can be categorized into three main types: supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning algorithms are provided with labeled output data, which is utilized to train the models and achieve the desired outcomes based on these labels. The process of supervised learning typically involves training and testing stages [17]. On the other hand, unsupervised learning aims to extract valuable insights from input datasets that do not have predefined class labels [18].

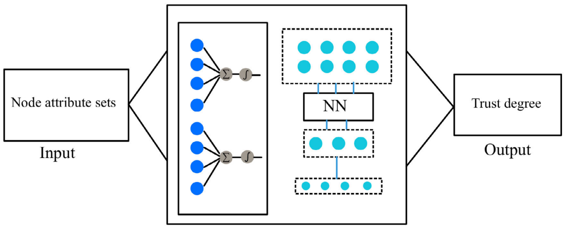

FIGURE 3. Structure of supervised classification models for trust evaluation.

In supervised learning, the choice of classification method plays a crucial role. The general structure of a supervised classification model as illustrated in Figure 3.

Machine learning and related technologies are rapidly evolving. The powerful capabilities are also applicable in trust evaluation. Trust evaluation heavily relies on historical interaction data, and the essence of machine learning is to make predictions based on historical data. By incorporating machine learning techniques into trust evaluation, the trust values of systems and target entities can be dynamically updated. This infusion of machine learning makes the trust evaluation model more dynamic and accurate.

Some commonly used ML methods are briefly introduced and their advantages and disadvantages are summarized in Table 1. We present the types of these methods, where "S" denotes supervised learning methods, "U" represents unsupervised learning methods and "Semi" represents semi-supervised learning methods.

Incorporating machine learning methods in calculating the trust level enhances the adaptability of the model. This enables the model to be responsive to new interactions, thereby improving its dynamic adaptability. Additionally, the utilization of a penalty mechanism effectively mitigates the impact of false feedback, particularly from deceitful nodes, including collusive deception nodes.

2) DEEP LEARNING & NEURAL NETWORK

Deep learning (DL) plays a crucial role in feature extraction and perception. Features can be learned from data without human-designed feature extractors in deep learning. Deep learning effectively combines multiple layers of representation learning methods, enabling the extraction of valuable information from the data for classification and prediction [19].

Deep learning technologies perform operations using multiple consecutive layers. Many layers are interconnected, and each layer receives the output of the previous layer as input. Starting from the original data, the representation of each layer becomes into a higher-level representation, thus the intricate structure is discovered in the high-dimensional data [20]. In the following, we introduce several deep learning technologies commonly used in trust evaluation.

(a) Convolution Neural Network (CNN)

In a CNN, multiple layers consisting of two-dimensional planes and multiple neurons are employed. Unlike traditional

TABLE 1. Comparison of machine learning methods.

Methods	Type	Description	Avantages	Disavantages
Decision Tree (DT)	S	DT constructs a model by serializing the problem division on data features.	DT exhibits fast training speed and good scalability.	DT is prone to overfitting and may become complex when dealing with datasets that contain a large number of features.
Naive Bayes (NB)	S	NB is based on the assumption of Bayesian theorem and the independence assumption between features.	NB possesses faster training speed and excellent scalability.	The disregard for the correlation between features may lead to suboptimal performance, especially in cases where feature correlation is high.
k-Nearest Neighbors (KNN)	S	KNN classifies or regresses based on the distance measurement between samples.	KNN is simple and easy to understand, suitable for multi-class problems, and capable of online learning.	KNN has a high computational complexity, resulting in lower classification efficiency for large-scale datasets.
Logistic Regression (LoR)	S	LoR is commonly used for binary classification problems. It is based on the linear regression model and applies the sigmoid function for classification.	LoR has high computational efficiency, interpretable models, and can provide probability estimates between classes.	As a linear model, its modeling capability for complex non-linear relationships is limited.
Support Vector Machine (SVM)	S	SVM performs classification by finding the maximum margin hyperplane in the feature space.	SVM performs well in handling high-dimensional and low-feature datasets, and can be applied to non-linear problems through kernel techniques.	SVM may incur significant computational costs when applied to large-scale datasets or datasets with high levels of noise.
K-means Clustering (K-means)	U	K-means groups similar data points together and aims to maximize the similarity within clusters while minimizing the similarity between clusters.	K-means is intuitive and easy to comprehend, and performs well when dealing with large-scale datasets. It does not require pre-specifying the number of clusters, and can automatically form clustering results at different levels.	Determining an appropriate value for k is not easy, and the clustering results are highly sensitive to the initial centroids and the selection of outliers.
Hierarchical Clustering	U	The process can be represented using a dendrogram. Each branch of the dendrogram represents a merging or splitting step, while the leaf nodes represent individual data points or final clusters.	It does not require the prior specification of the number of clusters to be formed. The clustering tree can be visualized as a dendrogram, allowing for an intuitive observation of the relationships between data points.	When dealing with large datasets, this algorithm exhibits high computational complexity and lacks robustness in handling outliers and noisy data.
Density-Based Spatial Clustering of Applications with Noise (DBSCAN)	U	DBSCAN classifies data points into core points, boundary points, and noise points based on the density of data points, determining the shape and number of clusters.	DBSCAN does not require pre-specifying the number of clusters and can automatically discover clusters of arbitrary shapes. It is robust against noisy data points.	It may have poorer performance in clustering high-dimensional data and clusters with varying densities.
Gaussian Mixture Model (GMM)	U	GMM treats data as a mixture of multiple Gaussian distributions, each with its own mean and covariance matrix.	GMM can fit various shapes of data distributions and provide probability estimates for each data point belonging to each component.	GMM has high computational complexity, and the results are greatly influenced by the choice of initial values.
Graph-based Semi-Supervised Learning (GSSL)	Semi	GSSL utilizes a graph to represent the relationships between data samples and leverages the information from unlabeled data to enhance the performance of classification or regression.	GSSL can effectively utilize information from unlabeled data. It represents data through a graph structure and handles local structures and similarities in the data better. It is suitable for large-scale datasets.	The process of graph propagation suffers from error accumulation, which may result in inaccuracies in the propagated results.

neural networks (NNs) where layers are fully connected, CNNs exhibit sparse connectivity, with each neuron being connected to only a limited portion of the preceding layer. This characteristic of CNNs enables them to effectively handle dense connections between deep neural network (DNN) layers, facilitating the classification of high-dimensional data within the input layer [21].

A typical CNN comprises several key components: convolutional layers, pooling layers, fully connected layers, as well as input and output parts. The dimensionality and depth of each layer can be tailored to suit the specific requirements of the CNN being constructed. The CNN's architecture, characterized by its interconnectedness, offers advantages such as parameter reduction and improved training speed. This connectivity also leads to reduced evaluation time, thereby enhancing overall system efficiency.

(b) Deep Belief Network (DBN)

The deep belief network is a neural network architecture composed of multiple layers of random variables. It is built upon the restricted Boltzmann machine (RBM), which serves as its fundamental building block. DBN employs two main training methods: unsupervised pre-training and supervised fine-tuning. By sequentially stacking pre-trained RBMs and learning the probability distribution of the data layer by layer, DBN extracts features with multiple probabilities.

One notable characteristic of DBN is the ability to pass the patterns learned at the top layer back to the input layer using conditional probabilities. This facilitates global fine-tuning through the use of the Backpropagation (BP) algorithm. DBN exhibits a high level of flexibility, allowing for easy extension. It strives to preserve the original features' characteristics while reducing their dimensionality. DBN finds common application in trust degree evaluation, as it can effectively reduce the evaluation error rate and provide a better representation of trust degrees. Vitalkar et al. employed a DBN to design a vehicle self-assembling network intrusion detection mechanism [22]. The results demonstrated that the DBN algorithm achieved higher accuracy in network intrusion detection compared to other machine learning methods.

(c) Stacked Auto Encoders (SAE)

SAE is a deep automatic coding model that consists of multiple autoencoders serving as its basic structural units. Each autoencoder comprises an encoding layer and a decoding layer, and it accomplishes signal input through iterative encoding and decoding processes. The training procedure of SAE involves the utilization of the greedy layer-wise unsupervised pretraining algorithm, which plays a crucial role in DNN preprocessing.

In the context of trust evaluation and intrusion detection, the utilization of SAE can enhance accuracy performance. Rao et al. proposed a two-stage hybrid intrusion detection scheme [23], wherein unsupervised SAE with smooth L1 regularization was employed in the first stage to promote autoencoder sparsity. In the second stage, a DNN was utilized for attack prediction and classification. The proposed model outperformed traditional models in terms of overall detection rate and false positive rate.

In addition, there are several other deep learning methods commonly used in trust evaluation in distributed networks. Lin et al. proposed a trust evaluation model based on a long short-term memory (LSTM) network using the particle swarm optimization (PSO) [24]. The model has the advantages of both algorithms. The vanishing and exploding gradient phenomena of traditional recurrent neural network (RNN) are avoided. It can find the globally optimal initial weights and thresholds to provide a more accurate trust evaluation. We briefly introduce some commonly used deep learning methods and summarize their advantages and disadvantages in Table 2.

C. ADVANTAGES OF AI-ENABLED TRUST

Many proposals have been put forward to enhance network security, and the utilization of trust has emerged as a prominent trend. For instance, trust evaluation systems employing game theory or fuzzy theory have gained attention. AI techniques have entered a booming stage of development, which can be applied in many fields, including cybersecurity and communication. In comparison to AI techniques, trust evaluation methods based on game theory or traditional algorithms tend to be more intricate and lack the ability to effectively address data-driven problems through programming.

As the external environment changes, trust needs to be adjusted automatically. The trustworthiness of target entities is an important factor that affects the dynamic adjustment of the system. Trust evaluation systems empowered by AI have the capability to perceive changes in the requirements of target entities and make necessary adaptations, such as updating the trust value assigned to entities. Moreover, the incorporation of a feedback correction function enhances the evaluation process, rendering it more dynamic, precise, and unbiased.

Additionally, trust is inherently influenced by subjective human judgments. Therefore, when evaluating trust, it becomes imperative to account for irrational behavior. Traditional algorithms typically assume the rationality of all participants, which is not always accurate. Thus, gaining an understanding of human behavioral patterns becomes crucial. Leveraging machine learning techniques enables the modeling of node behaviors based on past behavioral data to enhance the accuracy.

AI-enabled trust is commonly applied in the following three domains: TM, IDS, and RS. Within the field of trust management, the integration of AI technologies allows systems to effectively identify and utilize potential features for more accurate trust value calculations. This enables the systems to adapt to events and changes within a dynamically evolving environment. Furthermore, AI technologies empower IDSs to swiftly adapt to the rapidly changing network topology. This adaptability grants them significant advantages in detecting new types of intrusions and potential ones. Through the utilization of AI-enabled trust, recommender systems are capable of delving deeper into mining user and item features. In scenarios involving vast or rapidly changing data, an RS built upon AI-enabled trust possesses numerous advantages, such as higher recommendation accuracy and faster processing speed.

III. AI-ENABLED TRUST IN DISTRIBUTED NETWORKS

AI-enabled trust is commonly applied in three main domains: trust management, intrusion detection system, and recommender system. In this section, we analyze these applications of AI-based trust in DNs and conduct a comparative analysis of some related work.

A. TRUST MANAGEMENT

In 1996, M. Blaze introduced the concept of ''trust management'' as a solution to the security challenges faced by network services on the Internet. Blaze's work also encompassed the introduction of trust management mechanisms within distributed systems. The TM model serves as a framework

TABLE 2. Comparison of deep learning methods.

Methods	Description	Avantages	Disavantages
CNN	CNN consists of convolutional layers, pooling layers, and fully connected layers, featuring shared weights and local connections.	CNN can automatically learn spatial hierarchical features. Parameter sharing reduces the number of parameters, while local connections capture local relationships.	CNN exhibits limited generalization ability to large-scale transformations.
RNN	RNN processes and propagates past information through recurrent connections.	RNN has a memory function and can handle variable-length input and output sequences. It stores and utilizes past information.	RNN struggles to capture long-term dependencies, often leading to issues like vanishing or exploding gradients, and it has high training complexity.
LSTM	LSTM captures long-term dependencies and is suitable for handling sequential data.	LSTM addresses the issues of gradient vanishing and exploding in RNN, enabling it to capture long-term dependencies effectively.	Compared to regular RNNs, LSTM has a larger number of parameters, resulting in longer training and inference times.
Generative Adversarial Network (GAN)	GAN consists of a generator and a discriminator, engaging in adversarial learning during the training process.	GAN can generate realistic synthetic samples and has the ability to generate new samples.	GAN training process is unstable and prone to mode collapse and mode collapse problems, making it challenging to evaluate the quality of the model.
DBN	DBN constructs a deep structure by stacking multiple RBMs and utilizes a greedy layer-wise training approach for pre-training and fine-tuning.	DBN has strong feature representation learning capabilities, automatically learning data features and demonstrating better generalization ability.	DBN requires training multiple RBMs layer by layer, which can demand significant computational resources and time. Its performance is affected by multiple hyperparameters, making tuning difficult.
Deep Q-Learning (DQL)	DQL combines DNN with the Q-learning algorithm to solve Markov Decision Processes (MDPs) with large state spaces.	DQL uses deep neural networks to approximate the Q-value function, allowing it to handle large and continuous state spaces. DQL can learn nonlinear function approximation and therefore can learn complex decision strategies.	DQL relies on a large number of samples for training and is highly sensitive to sample quality. Additionally, DQL may encounter convergence difficulties and comes with high computational costs.
Deep Autoencoders (DAE)	Autoencoder is composed of an encoder and a decoder, used for unsupervised learning and feature extraction.	DAE can perform unsupervised learning and feature extraction on data, exhibiting certain robustness.	DAE may introduce information loss during the encoding and decoding of input data, potentially limiting the learned representations.

for assessing trust and establishing trust relationships among nodes/entities. TM primarily emphasizes the metrics of trust within these relationships and employs mathematical models for trust evaluation. Trust management models can be categorized based on their architectural structure, namely centralized, semi-distributed, and fully distributed. In contrast to the centralized model, a distributed trust management model operates without a central server. This paper specifically focuses on the distributed trust management model.

Trust management can typically be categorized into two main classifications: credential-based (or policy-based) static trust management (CSTM) and behavior-based dynamic trust management (BDTM), as depicted in Figure 4. In BDTM, the trustworthiness of the target entity is dynamically assessed based on its behavior history and current behavior. Behavior-based trust encompasses both direct trust and indirect trust.

In recent years, numerous machine learning and deep learning-based TM models have been introduced. Lin et al. proposed a TM model that incorporates machine learning techniques and social relationships to evaluate trust [25]. This approach involves assessing trust through the examination of node behavior and employing a data training model. In a similar vein, Ma et al. presented a behavioral model that utilizes the LSTM neural network for predicting future node behavior [26]. In the following, we further provide a detailed description of CSTM and BDTM.

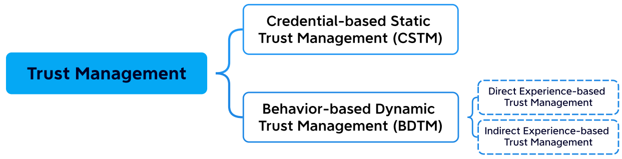
Trust Management
Credential-based Static Trust Management (CSTM)
Behavior-based Dynamic Trust Management (BDTM)
Direct Experience-based Trust Management
Indirect Experience-based Trust Management

FIGURE 4. Classification of trust management.

1) CREDENTIAL-BASED STATIC TRUST MANAGEMENT (CSTM)

The field of trust management has its roots in identity authentication and authorization. M. Blaze coined the term ‘‘trust management’’ to refer to the process of establishing security policies, obtaining security credentials, and evaluating whether the collection of security credentials meets the corresponding security policy. In the context of CSTM,

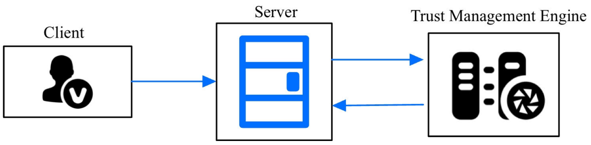

FIGURE 5. Framework of CSTM.

it grants entities access to resources based on their demonstrated trustworthiness. In the case of a single-origin scenario, employing a distributed static trust mechanism in the form of trust queries proves to be an effective solution for addressing security trust issues.

Trust relationships are typically assessed by employing digital credentials or credential chains. Once a system has established trust in the identity of an entity or confirmed its membership in a trusted organization, the utilization of certificates alone is deemed satisfactory for accomplishing authentication authorization and establishing a trust relationship.

Figure 5 presents the fundamental structure of the CSTM model. The trust management engine is the core of the whole trust management framework. It can return the permission judgment result of approval or rejection based on the input request, trust credentials, and security policy.

CSTM employs a program to validate trust relationships, necessitating the implementation of intricate security strategies by developers for conducting trust evaluation. However, it appears that CSTM is ill-suited for managing dynamically evolving trust relationships during runtime. In addition, CSTM mainly focuses on analyzing identity and authorization information. Once an information relationship is established, it generally fails to consider the influence of entity behavior on the existing trust relationship.

2) BEHAVIOR-BASED DYNAMIC TRUST MANAGEMENT (BDTM)

In 1994, Marsh conducted a pioneering study on feature-based trust management techniques from various sociological and behavioral perspectives [27]. In contrast to CSTM, the BDTM model emphasizes a comprehensive consideration of multiple factors, particularly the behavioral context, which significantly influences the establishment and management of trust relationships. Additionally, the BDTM model highlights the importance of dynamically collecting relevant subjective and objective factors that undergo changes over time. This allows for the immediate evaluation and management of entity trustworthiness. Consequently, the viability of entities can be dynamically updated and evolved within the framework of the BDTM model.

The subjectivity and uncertainty of trust are considered in BDTM. It gathers evidence related to the object and assesses the level of trust using various calculation models. Direct experience and indirect experience, known as witness information, are the most commonly utilized information in computing trust.

Compared with traditional TM, BDTM exhibits the following distinctive characteristic. Firstly, BDTM needs to gather trust-related information and translate it into various quantitative inputs that directly impact the trust relationship. Secondly, BDTM necessitates the continuous monitoring and adjustment of the trust relationship during the trust management process. This is achieved through the examination of multiple attributes associated with the trust relationship. Consequently, the managed trust network experiences heightened complexity and uncertainty. Last but not least, BDTM employs distributed trust evaluation and decision-making mechanisms to address the coordination challenges prevalent in trust management among diverse entities. By encompassing these characteristics, BDTM offers a more comprehensive and flexible approach to trust management, facilitating effective management of the complex dynamics within trust networks.

3) RELATED WORK ON AI-BASED TM

In Section II, it is evident that traditional artificial intelligence algorithms such as DNN are often referred to as ‘‘black boxes’’ due to their lack of interpretability in making predictions or decisions. Explainable Artificial Intelligence (XAI) has been proposed as a solution to this problem. Mahbooba et al. explore the concept of XAI to enhance trust management by investigating decision tree models in the IDS domain, due to the DT algorithm offers offer higher interpretability [28]. Therefore, combining decision trees with XAI can provide more intuitive explanations and understanding of the model's decisions. Their work achieved 100% accuracy, precision, and recall on the KDD benchmark dataset. It enhances trust management by enabling human experts to comprehend the underlying data evidence and causal reasoning.

Jyothi and Patil presented a trust mechanism for detecting selfish nodes in vehicular ad hoc networks (VANETs) based on the DBN-based red fox optimization (RFO) algorithm [29]. This authentication scheme simultaneously satisfied the security and privacy objectives in VANET environments. RFO is an optimization algorithm inspired by the behavior of red foxes in their prey searching process. The proposed method outperforms other existing methods, such as KNN and ANN, in terms of computational cost, accuracy, precision, recall, and communication overhead. However, it should be noted that this model can only evaluate and manage trust within VANETs.

Reinforcement learning (RL) is an AI technique that utilizes machine agents to solve problems by training robust machine learning systems through a combination of dynamic programming and supervised learning [30]. RL incorporates a mechanism that strikes a balance between exploration and exploitation, allowing it to explore new actions and strategies while leveraging existing knowledge and experience to make more accurate evaluations. Mayadunna and Rupasinghe introduced a trust framework based on reinforcement learning to calculate the trust values of user nodes in social networks [31]. In their study, they selected specific features of the social network as training features and utilized the presence of edges between nodes as label information.

Although RL has shown effectiveness in evaluating trust, it faces challenges in dealing with state spaces of high dimensionality. Deep Q-Learning (DQL) is an advanced form of reinforcement learning that combines DNNs with Q-learning algorithms to optimize the Q-value function, facilitating the selection of optimal actions [32]. DQL has demonstrated remarkable capabilities in handling high-dimensional states and action spaces, as well as improved learning efficiency and stability. He et al. introduced a model that utilizes deep Q-learning to make optimal decisions for automated network resource allocation [33]. By relying on automated decision-making instead of manually crafted or explicit control rules, the proposed model exhibits enhanced adaptability to dynamic changes in network conditions.

Deep Q-network (DQN) is based on the deep Q-Learning algorithm and incorporates DNNs to handle high-dimensional state spaces. In DQN, action selection is performed by using the currently estimated value function to find the action with the highest estimated value. However, such estimates can suffer from high estimation bias, leading to unstable and inefficient training. To address this issue, the concept of a target network is introduced in Double DQN. The target network is a separate network from the estimation network (online network) and is used for computing the target values. Several trust models based on Double DQN have been proposed, demonstrating acceptable accuracy and errors [34], [35], at the cost of significant computational resources and time.

The effectiveness of trust reasoning, which relies on trust propagation, is influenced by various factors, including the length of the path and the chosen aggregation strategy. Ghavipour and Meybodi proposed a novel aggregation strategy and a heuristic algorithm called DLATrust, which leverages distributed learning automata (DLA) and builds upon standard collaborative filtering techniques [36]. DLA is an approach rooted in RL principles, where a collective of interacting autonomous agents collaborate to solve problems. By harnessing distributed algorithms, DLA offers the advantages of flexibility and scalability. DLATrust aims to identify reliable paths between two users and infer trust values by employing the proposed aggregation strategy. In their study, the trust network was considered as a static graph. However, trust weights can dynamically change over time. To address this concern, the researchers introduced a dynamic trust propagation algorithm named DyTrust, which facilitates the inference of trust between indirectly connected users [37].

While significant research efforts have been devoted to establishing trust networks among users, limited attention has been given to analyzing their characteristics. Chen et al. proposed a trust evaluation framework based on machine learning in their study [38]. The researchers classified user features into four groups, including profile-based features, behavior-based features, feedback-based features, and link-based features. They then developed a lightweight feature selection method to assess the effectiveness of each feature and identify the optimal combination of features from users' online records. This approach provides a more comprehensive understanding of trust relationships and holds the potential to enhance the accuracy of trust evaluations.

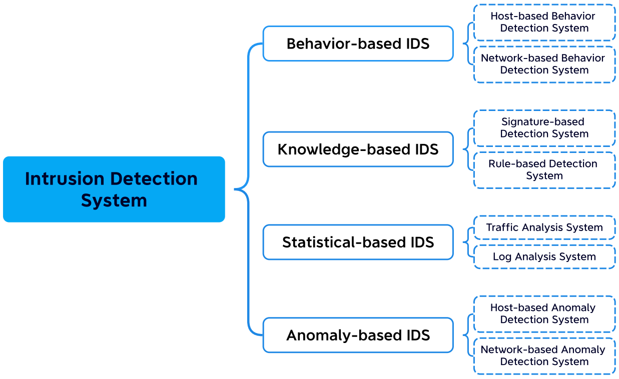
Intrusion Detection System
Behavior-based IDS
Host-based Behavior Detection System
Network-based Behavior Detection System
Knowledge-based IDS
Signature-based Detection System
Rule-based Detection System
Statistical-based IDS
Traffic Analysis System
Log Analysis System
Anomaly-based IDS
Host-based Anomaly Detection System
Network-based Anomaly Detection System

FIGURE 6. Classification of intrusion detection system.

Table 3 provides a comprehensive overview of recently proposed TM works utilizing AI techniques. We classify these works into two categories: credential-based and behavior-based. Furthermore, within the behavior-based category, we further distinguish between direct experience-based (DE) and indirect experience-based (IE) methods. Additionally, we present the employed techniques and experimental results regarding accuracy. Finally, we conduct an analysis of the advantages and disadvantages of the related work.

In recent years, the global outbreak of the COVID-19 pandemic has highlighted the importance of public health on a global scale. In their study, Fang et al. proposed a trust management scheme that utilizes dynamic aging weights within the framework of the Internet of Medical Things (IoMT) [44]. The scheme focuses on collaborative behaviors between two nodes, employing a higher aging weight to limit the rapid growth of trust value among regular nodes. In the event of non-cooperative behaviors, a lower aging weight is employed to reduce the potential risks associated with compromised nodes.

B. TRUST-BASED INTRUSION DETECTION SYSTEM

Intrusion detection is an essential mechanism employed to identify unauthorized access to networks through the analysis of network traffic in order to uncover covert malicious activities [45]. Malicious behaviors can be distinguished from normal network behaviors by using an intrusion detection scheme. In Figure 6, we depict the commonly used classifications of IDS.

The utilization of AI-enabled trust in the design of an IDS enhances both the accuracy and efficiency of detection. This approach effectively accommodates intricate network structures and enables the detection of a broader range of attack methods, leading to improved monitoring accuracy

TABLE 3. Comparison of related work on trust management.

Scheme	Technique	Credential-based	DE	IE	Results	Advantages	Disadvantages
DFTE [25]	RL	×	✓	✓	FAR=4.8% MDR=2.55%	DFTE achieves high trust evaluation accuracy under different granularity requirements and performs well in terms of participation rate and data reliability.	DFTE requires a considerable amount of trial and error and iterative optimization to find the optimal strategy.
MWHZ [26]	LSTM	×	✓	✓	Accuracy=98.3% MSE=0.05 R^2=0.88	The model exhibits good accuracy, low error, and short training time.	The process of calculating trust values is based on a specific edge computing architecture of the Internet of Things, lacking flexibility.
MTSS [28]	DT	×	×	✓	Accuracy=100% Precision=100% Recall=100%	Decision trees show improved accuracy compared to other methods, with low computational costs.	The model is prone to overfitting.
NP [29]	DBN	×	✓	✓	Accuracy=94% Precision=90% Recall=90%	The model demonstrates good accuracy performance.	The developed method is limited to the trust prediction process in VANETs only.
MR [31]	RL	×	✓	×	-	It can accurately evaluate trust.	It has a weak ability to handle high-dimensional state spaces.
HLYH [33]	DQL	×	×	✓	-	It can automatically optimize decisions for network resource allocation based on variable network conditions.	High energy consumption.
DDQN-Trust [34]	Double DQN	×	✓	✓	Accuracy=94.0%	The model outperforms approaches based on DQN and random scheduling in terms of accuracy performance.	The scheduling process incurs high overhead and has a large time complexity.
DDQN-D2D [35]	Double DQN	×	✓	✓	-	It enhances offloading probability and reduces average latency consumption to some extent.	The training process consumes significant computational resources and time.
DLATrust [36]	DLA	×	×	✓	MAE=0.115 Precision=96.32% Recall=97.38%	It can effectively identify reliable trust paths and predict trust with higher accuracy.	It cannot effectively predict dynamic trust.
DyTrust [37]	DLA	×	×	✓	MAE=0.115 Precision=95.43% Recall=99.43%	It can accurately infer dynamic trust.	High energy consumption and time cost.
CYLY [38]	ML	×	✓	✓	Accuracy=96%	It improves the accuracy of trust evaluation by analyzing the overlap between positive and negative feature distributions.	It has a weak context-awareness ability for different scenarios and time.
LM [39]	SVM	✓	×	×	Accuracy=96.6%	The model has good accuracy performance.	It incurs high computational overhead and requires appropriate parameter settings.
SMS [40]	KNN	×	✓	×	Accuracy=100% Precision=100% Recall=100%	The model exhibits good accuracy performance.	The dataset used for model training has a limited variety of attributes.
NeuralWalk [41]	NN	×	✓	✓	F1-score=91.6%	It can accurately predict unknown trust relationships in an inductive manner.	High time cost.
iSim [42]	Hybrid ML	×	✓	×	MAE=0.13	It has low time complexity and strong context awareness ability.	Poor stability, with significant performance differences when tested on different datasets.
Medley [43]	Time encoding	×	×	✓	Accuracy=73.3%	It can capture continuous time features and assign weights using an attention mechanism.	High energy consumption and time cost.

and efficiency. Moreover, the deployment of an IDS in the network results in reduced risks of network paralysis and privacy breaches caused by cyber-attacks, ultimately enhancing network security.

1) DATASET ON IDS WITH AI-ENABLED TRUST

In the following, we present commonly employed datasets in the design of IDSs. One such dataset is the KDD CUP99 dataset, which has gained extensive usage since 1999 and currently holds the status of being the most widely utilized intrusion detection dataset. Over the years, it has served as a benchmark dataset for numerous research projects in the field of intrusion detection, earning it the name of the KDD benchmark dataset. Nevertheless, it is worth noting that due to its imbalanced distribution of data types, the results obtained from this dataset tend to exhibit a bias towards the more frequently occurring data.

Later, in order to address the limitations of the KDD CUP99 dataset, a new standard dataset called NSL-KDD was introduced. It has a balanced data distribution with no duplicates and thus does not favor more frequent data. However, NSL-KDD lacks a representative sample of typical attacks. As a result, another dataset called UNSW-NB15 was proposed by the Australian Centre for Cyber Security (ACCS) and was closer to the data from real-world network. In addition, more recent datasets such as CICIDS2017 and CICIDS2018 have also been developed. These datasets offer researchers additional resources for studying intrusion detection.

However, commonly used datasets often exhibit a significant class imbalance, with a much larger number of normal instances compared to anomalous instances. This inherent imbalance can introduce bias in models, causing them to overfit to normal data and disregard anomalous samples during the training. Furthermore, due to the imbalanced dataset, an IDS may exhibit high detection rates for certain specific types of attacks, while not performing adequately in detecting all types of attacks. Therefore, generating a rich and high-quality dataset is crucial.

The typical steps involved in generating a diverse and excellent dataset are as follows. Firstly, it is necessary to determine the desired data types and features for the dataset, such as network traffic data, log files, and system events. Next, real intrusion detection datasets like KDD CUP99 and NSL-KDD are collected. In cases where existing data is insufficient or lacks diversity, data augmentation and synthesis methods can be employed to generate more diverse data, thereby expanding and diversifying the dataset and improving the model's generalization ability. Subsequently, appropriate features are selected and extracted based on the requirements and issues of the intrusion detection system. Additionally, the dataset needs to be cleaned and preprocessed to ensure its quality. Lastly, continuous updates and improvements should be made to the dataset to enable the trained model to detect the latest types of intrusions.

2) RELATED WORK ON IDS WITH AI-ENABLED TRUST

In recent years, numerous trust-based IDS with AI technologies have been proposed. By leveraging ML and DL algorithms, an IDS has the capability to continuously learn and optimize, thereby enhancing its accuracy and efficiency. Additionally, through trust mechanisms, the IDS can identify entities with a high level of trust and label them as trustworthy, thereby reducing the likelihood of false positives. Moreover, trust-based IDS can better defend against internal attacks, in addition to external attacks. For entities with a low level of trust, trust-based IDS can employ stricter measures to ensure the legitimacy of their actions.

The sea lion optimization (SLO) algorithm is a heuristic optimization algorithm based on the behavior of sea lions in the natural world. It simulates the optimization strategies observed in sea lions' behaviors such as hunting, reproduction, and territory competition. Kagade and Jayagopalan proposed a novel IDS in wireless sensor networks (WSNs) using a deep learning model [46]. They introduced a new approach called self-improved sea lion optimization (SI-SLnO) model. According to this strategy, a multidimensional two-layer hierarchical trust model is employed to evaluate the trust of cluster heads and nodes, considering content trust, honesty trust, and interaction trust. Initially, the optimal cluster heads are selected within the sensor nodes, prioritizing those with high energy levels. Lastly, intrusion detection based on deep learning is performed using an optimized NN, with training accomplished through the proposed SI-SLnO algorithm's optimal weight adjustment process. Experimental results demonstrate the accuracy of this intrusion detection model. However, NN-based model requires abundant and high-quality datasets for training.

Federated learning is a distributed machine learning approach aimed at training models using locally available data distributed across multiple devices or data centers while ensuring user privacy protection. By decentralizing the model training process onto local devices, federated learning avoids the need for transmitting sensitive data in a centralized training setting. Several works have been proposed based on federated learning [47], [48], [49]. By leveraging the diversity of these datasets, federated learning enables the training of more comprehensive and robust IDSs. Although federated learning reduces the amount of data transferred, network communication is still required during the model parameter transmission, which can introduce delays and communication overhead.

Lingam et al. proposed an algorithm to tackle the challenges posed by social botnets [50]. Their algorithm focuses on detecting social botnets by introducing a trust model that incorporates two parameters: direct trust and indirect trust. These parameters are utilized to identify reliable paths within online social networks (OSNs). Direct trust is calculated using Bayesian theory, while indirect trust is determined using Dempster-Shafer (D-S) theory. In a subsequent study conducted two years later, the researchers constructed a weighted signed Twitter network graph [51]. This graph assigns weighted edges based on behavioral similarity and trust values between online social accounts. Additionally, two algorithms were developed by the researchers: social botnet community detection (SBCD) and deep autoencoder-based SBCD (DA-SBCD). The objective of these algorithms is to accurately detect social bots in OSNs.

As illustrated in Table 4, we provide an overview of the related work, including the techniques, datasets utilized, and their experimental results. In addition, we include the advantages and disadvantages of each work in Table 4.

In Table 4, compared to IDSs based on traditional classification methods, neural network-based IDSs generally exhibit higher accuracy. However, this improvement in performance comes at the expense of longer training times and increased energy consumption.

C. TRUST-BASED RECOMMENDER SYSTEM

A recommender system serves the purpose of not only providing information services, but also establishing connections between target entities (typically users) and recommended

TABLE 4. Comparison of related work on intrusion detection system.

Scheme	Technique	Dataset	Accuracy (%)	Precision (%)	Recall (%)	Advantages	Disadvantages
SLnO-NN [46]	NN	-	95.0	92.0	93.0	Accuracy is high and computation time is short.	High quality and diversity train data is required.
FL-IDS [47]	Federated Learning (FL)	Aposemat IoT-23	92.5	90.7	86.3	FL-based models and ensemble learning-based models show comparable accuracy performance.	There are significant communication overheads and high computational costs.
DTF [48]	FL	Cooja Dataet	96.0	96.0	97.5	High accuracy with lightweight design.	There are considerable communication overheads and high computational costs.
FELIDS [49]	FL	CSE-CIC-IDS2018	94.2	98.0	99.0	High accuracy.	There are substantial communication overheads and high computational costs.
LRS [50]	Bayes, D-S theory	The Fake Project	-	84.5	-	It leverages Bayesian theory for direct trust determination and D-S theory for indirect trust determination, combining the strengths of both approaches.	The training dataset exhibits low accuracy and lacks diversity.
DA-SBCD [51]	DAE	The Fake Project	-	90.9	87.7	By integrating deep autoencoder models with trust and similarity values, it achieves more accurate botnet detection.	The model fails to consider the temporal and spatial characteristics of social robots, resulting in low temporal correlation.
SMO-DBN [52]	DBN	KDD CUP99	90.0	90.0	92.0	The proposed method demonstrates superior accuracy and a low false alarm rate.	When the number of hidden neurons is small, the false alarm rate exceeds 10%.
CPMA [53]	SVM	-	96.0	-	-	The proposed detection framework outperforms others in terms of detection performance and accurately identifies the attack pattern of malicious nodes.	As the number of relay nodes is limited, the accuracy rate tends to decrease. However, as the number of relay nodes increases, the system size also progressively expands.

entities (usually items or products). The recommendation process involves information filtering, wherein RSs aim to predict the level of interest that an entity may have in a recommended entity. The implementation of a recommender system involves employing various methods to effectively connect target entities with recommended entities and achieve the desired recommendation outcome.

1) TRUST IN RS

In general, the recommendation process of a RS follows a series of steps. Initially, entity models are constructed, and the target entity database is generated by extracting target entity information from preferences and historical data. Meanwhile, the features of the recommended entities are extracted to build recommended entity models, which form the recommended entity database. Subsequently, a recommendation algorithm is employed to generate a recommendation list or assign a final rating to the recommended entities. Finally, the resulting information is provided to the target entity.

Leveraging trust relationships among entities has the potential to enhance the reliability of recommendations. Numerous trust-based recommendation schemes have been proposed, which explore the explicit or implicit trust dynamics existing between entities. They make recommendations to targeted target entities based on trust relationships.

The implementation of recommendation algorithms typically relies on a dedicated database that encompasses historical behavior or rating data pertaining to target entities. Thus, during the initial phase, accurately recommending the appropriate products to these target entities becomes challenging [54]. When a new target entity or recommended entity enters the system, essential data required to establish a trust relationship is absent. As a result, the recommender system faces difficulties in gauging entity preferences and executing recommendation mechanisms, leading to what is commonly known as the cold start problem. If a new target entity does not receive appropriate advice promptly upon entering the system, its experience of using the system will be compromised, which may even cause the target entity to relinquish the use of the system [55].

2) RELATED WORK ON RS WITH AI-ENABLED TRUST

Based on AI technologies, recommender systems can predict users' preferences and behavior patterns more accurately. As a result, in recent years, many recommender systems based on AI-based trust have been proposed. Root mean square error (RMSE) is one of the commonly used metrics to evaluate the accuracy of recommender systems. A smaller RMSE value indicates a higher accuracy, as it reflects the closeness between the predicted results and the actual user interests. Compared to other metrics like mean absolute error, RMSE is more sensitive and penalizes larger error values, thus better reflecting the overall performance of recommender systems.

In the evaluation of recommender systems, federated learning can help collect and integrate user behavior data distributed across different devices, leading to a more accurate understanding of user interests and behavior patterns, thereby providing personalized recommendation services. Wahab et al. proposed a federated learning-based approach to address the item cold-start problem in recommender systems [56]. The proposed model demonstrates low RMSE and runtime. However, designing recommender systems based on federated learning presents certain challenges, such as device heterogeneity and high communication costs.

Furthermore, there have been related works based on deep learning and neural networks [57], [58], [59], [60]. Experimental results have shown that deep learning and neural networks can capture latent features and user behavior patterns hidden in the data, enabling more accurate prediction of user interests and preferences. However, they require significant computational resources and data for training and lack interpretability.

Existing RSs often struggle to handle large datasets, leading to a decrease in their performance. In this study, the authors propose a hyper-tuned RBM and develop a reliable recommender system based on its design [61]. They establish a mathematical model-based objective function to measure users' recommendation/trust scores. The results demonstrate that the model not only achieves acceptable accuracy but also handles large datasets as inputs effectively.

As illustrated in Table 5, we provide an overview of recent related work in the field of AI-based RS with trust. In addition, we have also indicated whether the related work considers mitigating the cold start problem with "Y" for yes and "N" for no. Besides, we present the datasets utilized in each study and evaluate their performance in terms of accuracy, measured by the RMSE metric, along with an analysis of their advantages and disadvantages.

In Table 5, it can be observed that some recommender systems have not taken the cold start problem into consideration. Additionally, it is worth noting that MovieLens serves as a commonly utilized dataset, and recommender systems leveraging AI-enabled trust typically exhibit low RMSE values.

Similarly, the approach of combining another scheme to optimize the performance of one scheme is also applicable in an RS. O'Donovan and Smyth conducted a comparative study involving two distinct trust evaluation models [68]. These models were integrated into the standard collaborative filtering algorithm, employing a trust-based weighting method and a trust-based filtering method, respectively. The experimental results showed that the integration of the trust management model led to an improvement in recommendation accuracy.

The selection of a dataset is a crucial aspect in the design of an RS based on AI-enabled trust. Commonly utilized datasets in the domain of movie recommender systems include the MovieLens dataset, Netflix dataset, and FilmTrust. These three datasets encompass a wealth of attributes and labels pertaining to movies. Moreover, the MovieLens dataset is accessible in sizes of 10k, 100k, and 1M, offering varying degrees of data granularity. In the realm of social recommender systems, the Epinions and Ciao datasets enjoy popularity. These datasets are frequently employed for training and evaluating models centered around social recommendations [69].

IV. OPEN PROBLEMS & CHALLENGES

Despite significant advances in AI-enabled trust technologies for DNs, numerous challenges still remain. In the following analysis, we present some current open problems and challenges in terms of the application domains.

A. TRUST MANAGEMENT

After being optimized using AI techniques, a trust management model demonstrates improved accuracy and efficiency in evaluating trust between nodes. Nevertheless, numerous challenges and problems still require mitigation and resolution, which can be primarily observed in the following aspects.

Firstly, the energy consumption of these devices increases as a result of integrating machine learning algorithms for trust evaluation into the devices of DNs [70]. While the utilization of ML algorithms featuring intricate logic or multiple layers may enhance performance, the task of conserving node energy poses inherent challenges that are hard to circumvent. To address this challenge, researchers can consider distributing computational tasks across multiple devices or servers for processing. By utilizing distributed computing, the workload on individual devices can be alleviated, thereby reducing energy consumption.

Secondly, the accuracy of trust evaluation is typically low during the initial stage. Trust evaluation relies on historical data, but the initial data is often sparse. So, the initial trust of nodes is difficult to evaluate accurately in the system's early stages [71]. However, the issue of data sparsity can be addressed through algorithmic advancements. To mitigate this problem, Rahim et al. proposed a novel trust-based scheme called TrustASVD++, which builds upon the ASVD++ algorithm [72]. This approach leverages trust data associated with target entities and utilizes matrix decomposition techniques to effectively alleviate the problem of data sparsity.

Thirdly, some trust management models exhibit limited applicability due to the diverse nature of network structures. There is no universally applicable trust management scheme, and it is necessary to propose appropriate protocols for each specific trust management scheme [73]. The forms of trust management may vary across different network topologies or application scenarios. Introducing adaptive mechanisms allow for the dynamic adjustment of parameter and strategy in the trust management model based on different network structures and environmental conditions.

Last but not least, cross-origin trust is a challenging aspect to achieve. Presently, most TM models primarily focus on evaluating trust within a single origin, neglecting the crucial aspect of cross-origin evaluation. Consequently, these models lack the capability to facilitate cross-origin interaction and evaluate trust between nodes from different origins. Researchers can address the issue of cross-domain trust

TABLE 5. Comparison of related work on recommender system.

Scheme	Cold Start	Technique	Dataset	RMSE	Advantages	Disavantages
WRBC [56]	Y	FL, DQL	Epinions	1.50	The accuracy of cold-start recommendations has been improved, with lower RMSE, MAE, and runtime.	Training costs are high and depend on the quality of the dataset.
TTDNN [57]	Y	DNN	Last.fm	-	Addressing the sparsity issue in data, reducing the computational complexity of the recommender system.	Accuracy performance is not high.
RSLCNet [58]	N	LSTM, DNN	MovieTweetings	1.73	Both the accuracy and effectiveness in terms of MAE and RMSE have been improved.	The model requires significant computational resources for training and inference and is highly sensitive to the order of input data.
DLCRS [59]	N	DL	MovieLens 1M	0.90	Accuracy and effectiveness have been significantly improved.	Training time is long.
CMJ [60]	Y	DNN	-	0.86	The model exhibits low loss rate and high accuracy.	Training time is long and depends on the quality of the dataset.
JGRT [61]	N	Hyper-tuned RBM	MovieLens 100K, MovieLens 1M, Film Trust	0.53 0.43 0.25	The loss error and RMSE are low, indicating good accuracy.	The computational complexity slightly increases.
JGT [62]	Y	Cluster Analysis	MovieLens 1M	0.50	The loss error and RMSE are low, demonstrating good accuracy.	There are numerous fake anomalies and low transparency.
KB [63]	Y	DAE	Epinions, Ciao	0.89 0.82	Compared to other methods, there is a 10-20% improvement in RMSE.	The dataset used for model training has few attribute types.
DNN-MF+SSO [64]	N	DNN Social Spider Optimization (SSO)	Yahoo!, TripAdvisor	0.69 0.76	Compared to other benchmark works, the MAE and RMSE are lower.	It requires a large amount of data for training, and high quality and diversity of data are necessary.
SSS [65]	N	ML	HVBP	0.20	In comparison to CNN and SVM-based approaches, we achieved an average loss reduction of approximately 34% and 3%, respectively.	Training time is long and depends on the quality of the dataset.
TECSRS [66]	N	Linear Regression, ANN, RF	QWS	0.14 0.29 0.10	The RF-based recommender system achieves the lowest MSE, indicating high accuracy.	RF-based recommender systems are prone to overfitting and perform poorly on high-dimensional sparse data.
RSA [67]	N	NB	An online dataset	-	In addition to providing recommendations, we also assure users that the recommendations come from trustworthy sources, effectively protecting information and privacy.	The model may be affected by the "curse of dimensionality".

by utilizing a proxy server. The proxy server is positioned between the client and the target server, allowing it to receive client requests and forward them to the target server.

B. INTRUSION DETECTION SYSTEM

IDS is one of the most common and effective tools for defending against cyber-attacks. Almost all of the related work on IDS shown in Table 4 demonstrate a remarkable accuracy rate of 90%. Moreover, certain machine learning models have even achieved a detection accuracy of 99% for specific attack types. The introduction of machine learning has significantly improved the accuracy of the IDS. However, this field also faces numerous challenges.

Firstly, some IDSs may experience low detection speed, which becomes a significant challenge in scenarios with high network data traffic. The real-time nature of IDS operation necessitates rapid detection, and if the detection speed is inadequate to match the pace of attacks, the efficiency of intrusion detection is considerably compromised. The utilization of parallel computing can enhance the processing speed of intrusion detection systems. Performing appropriate preprocessing on input data can reduce computational complexity and workload. For instance, data normalization, filtering, or sampling can be applied to decrease data volume or eliminate noise.

Secondly, IDS often exhibit limitations in their independent learning capabilities. Furthermore, the number of attack types is growing rapidly, underscoring the importance of timely updates to the detection database in IDS. Failure to do so can result in the system falling behind the evolving attack techniques and failing to identify emerging intrusion attempts. Consider utilizing more sophisticated and flexible models, such as deep learning models, to better capture complex attack patterns. Regularly update the models, monitor the latest attack trends, and incorporate new training data for model updates.

Thirdly, intrusion detection datasets are commonly characterized by a significant class imbalance, where the number of normal data instances far exceeds that of anomalous data instances. This inherent imbalance poses a challenge during training as models tend to exhibit a bias towards overfitting to normal data while neglecting anomalous samples.

Furthermore, while IDS may achieve high detection rates for certain types of attacks, they may not perform as effectively in detecting all attack types. Therefore, it is crucial to create a comprehensive and well-balanced dataset.

Lastly, the deployment of an IDS necessitates stringent requirements on network architecture. Nowadays, there exist numerous network configurations and compatibility issues may arise during IDS deployment, particularly when network topology changes or protocol switches occur. These compatibility issues can potentially result in subpar detection outcomes. Cross-domain training and transfer learning can be employed to alleviate this issue. Cross-domain training involves training the model on datasets from multiple environments, enabling the model to be more robust to network structure and topology variations across different environments. Transfer learning, on the other hand, involves training the model on a source domain and then applying the trained model to a target domain to improve its performance in the target domain.

C. RECOMMENDER SYSTEM

Recommendation algorithms are primarily driven by big data analytics. Therefore, it is difficult to accurately assess trust values and thus to recommend the most appropriate recommended entity to the target entity. There are several main issues and problems with the RS as currently designed using AI-enabled trust evaluation methods.

Firstly, the cold start problem poses a significant challenge. As RS heavily rely on historical data to generate recommendations, accurately providing recommendations during the initial start-up phase becomes more challenging. Upon analyzing Table 5, it becomes evident that certain studies have overlooked the cold start problem, necessitating researchers to focus on further optimization in future investigations. By combining various recommendation algorithms, such as content-based recommendation, collaborative filtering, and popularity-based recommendation, it is possible to alleviate the cold-start problem and reduce its impact.

Secondly, recommender systems are prone to malicious attacks. Hackers or vicious competitors often engage in malicious activities, such as inflating their own or their peers' trust values by providing artificially high scores. Simultaneously, they deliberately assign low scores to their competitors in order to diminish the trustworthiness of other legitimate entities. These actions result in disparities between the recommended information and the actual reality, thereby impacting the evaluation of trust ratings and interfering with the selection of target entities. By leveraging reinforcement learning techniques, RSs can continuously adjust strategies to adapt to the modifying behaviors of attackers, thereby enhancing the robustness of the system.

Thirdly, the scarcity of data can result in significant errors. The provision of accurate recommendations relies on a substantial volume of reliable data. When the available data is too limited, the accuracy of system recommendations tends to be compromised. To address this issue, Shambour et al. introduced a novel fusion-based multi-criteria collaborative filtering model, aimed at enhancing the effectiveness and personalization of hotel recommendations [74]. It was observed that as the data sparsity increased, the mean absolute error (MAE) of this model also increased, while the coverage decreased.

Last but not least, RSs utilizing AI-enabled trust face ethical and privacy challenges. RSs require the collection and analysis of users' personal data, and personal data breaches or unauthorized use of user information can potentially infringe upon the privacy of users. To mitigate this issue, techniques such as data anonymization and de-identification can be employed to protect user privacy. Additionally, differential privacy techniques can be utilized to add noise to the recommendation results, further safeguarding user privacy. Furthermore, recommender systems may produce unfair outcomes due to algorithmic biases or discrimination against certain users or groups. To address this issue, designers need to conduct algorithmic audits and tests to identify and rectify potential biases. In addition, the utilization of federated learning also offers the advantage of preserving user privacy, as the raw data does not need to leave the user's device, and only model updates are shared.

AI has its unavoidable weaknesses, such as the problems of datasets. The most significant problems with the current datasets are the staleness of data and the imbalance between normal data and attack data. Moreover, some datasets have poor correlation and an imbalance between the train sets and test sets. Though the problem of the dataset may be mitigated by optimizing the dataset, it still is a key challenge.

V. FUTURE WORK

Certainly, proposed models and systems with AI-enabled trust have showcased significant research contributions within the realms of cybersecurity and privacy protection. However, there are still several challenges and issues exist, necessitating further resolutions and alleviation. Optimistically, it rather proves the research potential in the field. In future work, researchers can consider the following suggested items.

A. A TRUST MANAGEMENT SYSTEM BASED ON ENSEMBLE LEARNING WITH DNN

In ensemble learning (EL), the integration of multiple machine learning methods is utilized to optimize the solution. However, EL suffers from the drawback of prolonged training time and high computational costs. Therefore, a potential approach to address this issue prior to employing EL for trust evaluation is the design of a DNN that can discern suitable candidate machine learning methods. Neural networks exhibit greater adaptability to the environment in comparison to traditional algorithms. By pre-selecting methods, subsequent training time and computational costs in the ensemble learning process can be mitigated.

Regarding implementation, the first step is to build a DNN model capable of accepting input data and pertinent features. The model learns to predict the most effective machine learning methods for a given problem. Inputs to the model may encompass problem descriptions, dataset features, and relevant domain knowledge. Outputs can manifest as scores or probabilities assigned to each machine learning method, indicating their suitability. To train the DNN, a dataset encompassing diverse machine learning methods and corresponding performance metrics must be prepared. This dataset can be manually annotated by experts or derived from existing machine learning problems and performance evaluations. Subsequently, the DNN is trained using this dataset to acquire an understanding of the correlations between different machine learning methods and problems.

The design of a DNN for selecting appropriate machine learning methods effectively addresses the challenges associated with lengthy training time and high computational costs in ensemble learning. This approach leverages the adaptability of neural networks to the environment and their comprehension of problem complexity. Nevertheless, it necessitates overcoming challenges pertaining to dataset construction and training, as well as validating and adjusting the selected machine learning methods.

Besides, the utilization of EL can be considered for the selection of cluster heads in DNs, aiming to enhance node energy conservation and extend the network's overall lifespan [75]. Furthermore, researchers have the opportunity to explore additional parameters that can provide a more accurate evaluation of trust. For instance, Siddiqui et al. proposed a trust evaluation model that achieves high performance by incorporating three parameters: the similarity rate (SMR) of content and services between two vehicles, the familiarity rate (FMR) of the trustor towards the trustee, and the packet delivery ratio (PDR) [76]. Simulation results demonstrated that the classification accuracy obtained by averaging these parameters surpassed that of classifications based on single parameters.

B. AN INTRUSION DETECTION SYSTEM WITH A HYPERPARAMETER AUTO OPTIMIZER

By studying proposed related works on AI-based IDS, disparities in performance have been observed among various IDSs. These differences arise from variances in the adoption of machine learning classifiers and training datasets when dealing with different attack types.

To tackle this issue, a potential solution involves implementing a hyperparameter automatic optimizer for mitigation. This approach entails the integration of multiple trained detection subsystems into the IDS, enabling an automatic optimization process. Initially, conventional detection subsystems identify a sufficient number of intrusion events, allowing the IDS to analyze the most prevalent intrusion types in the present environment. Subsequently, the hyperparameter automatic optimizer selects the most appropriate detection subsystem and hyperparameters to adapt to the prevailing intrusion environment. This method offers the advantage of combining multiple independent intrusion detection subsystems, thereby enhancing detection capabilities for various attack types in comparison to a single IDS. Additionally, it improves adaptability in dynamic environments. Moreover, through the integration of multiple detection subsystems, the IDS can gather information from diverse perspectives and feature sets, leading to enhanced detection accuracy and coverage.

However, the implementation of this approach necessitates extensive training and experimentation to construct and optimize multiple detection subsystems, while determining the optimal configuration for hyperparameters. Furthermore, it is crucial to ensure coordination and consistency among different detection subsystems to avoid overfitting or conflicting situations. Thus, developing an efficient hyperparameter automatic optimizer is a complex task that requires comprehensive consideration of various factors, alongside rational experiment design and algorithm development.

In conclusion, an IDS that incorporates a hyperparameter automatic optimizer effectively addresses the performance disparities observed among IDS trained on different machine learning classifiers or datasets, when confronted with various attack types. Its advantages lie in offering broader attack detection capabilities and adaptability to dynamic environments. Nevertheless, it also presents challenges related to algorithm development and experiment design that need to be overcome.

C. AN XAI-BASED HYBRID RECOMMENDER SYSTEM USING THE KNOWLEDGE GRAPH TECHNOLOGY

Explainability refers to the capacity of a model to allow human understanding of its decision-making process, even by individuals without expertise in machine learning. XAI enhances user trust and system acceptance by providing an understanding of the rationale behind specific predictions or decisions made by AI systems. An recommender system built upon XAI present a novel approach that surpasses traditional one. This is primarily attributed to the enhanced ability of XAI technology to comprehend the relationships and associations among items, resulting in more accurate and personalized recommendations.

To address the cold-start problem arising from data sparsity, various methods have been proposed. One such method is the utilization of Biased Tensor Factorization (BTF), which has demonstrated high effectiveness in mitigating data sparsity issues [69]. By incorporating biases between users and items, BTF can fill in missing values in the data, thus improving the accuracy of recommendations. Furthermore, researchers can explore the integration of knowledge graphs to alleviate the cold-start problem. Knowledge graphs employ a graph structure to represent relationships between entities, including target entity-recommended entities, target entity-target entities, and recommended entity-recommended entities. Leveraging these relationships, recommender systems can gain better insights into users' interests and preferences, leading to more precise recommendations. The use of knowledge graphs enhances initial trust and the accuracy of recommendations.

Moreover, hybrid recommendation methods serve as an effective strategy for addressing the cold-start problem. These methods combine multiple recommendation algorithms and techniques, surpassing the limitations of individual approaches. By integrating different recommendation methods, hybrid approaches can leverage the strengths of each, providing more comprehensive and personalized recommendations.

VI. CONCLUSION

In this paper, we have conducted a review of the research field of AI-enabled trust and highlighted its immense potential. By introducing the concept of trust, models can make decisions more comprehensively and securely. Furthermore, leveraging AI techniques to extract latent features allows for more accurate evaluation of trust in nodes or entities. AI-enabled trust evaluation with the ability to perceive changes in target entity demands and dynamically adjust various aspects, including updating entity trust values.

We have analyzed some commonly used ML and DL algorithms in trust evaluation, and compared their advantages and disadvantages. Additionally, we have analyzed and compared some related works on AI-enabled trust, which have been categorized based on the three widely adopted applications of AI-enabled trust: TM, IDS, and RS. We find that although many works based on AI-enabled trust have achieved respectable accuracy, striking a balance between accuracy, training cost, and system complexity often proves challenging.

Subsequently, we have summarized the current open problems and challenges in the field of AI-enabled trust and proposed corresponding solutions. Finally, we provide three suggestions for future work. EL is a promising method in the field of trust evaluation, and the XAI can enhance user trust and system acceptance by providing an understanding of the decisions made by AI systems. The field of AI-enabled trust still faces numerous issues and challenges that should receive focused attention in subsequent research endeavors.

REFERENCES

[1] N. V. Oza, T. Hall, A. Rainer, and S. Grey, "Trust in software outsourcing relationships: An empirical investigation of Indian software companies," *Inf. Softw. Technol.*, vol. 48, no. 5, pp. 345–354, May 2006.

[2] Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, "Communication-efficient edge AI: Algorithms and systems," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2167–2191, 4th Quart., 2020.

[3] F. Kamoun, F. Iqbal, M. A. Esseghir, and T. Baker, "AI and machine learning: A mixed blessing for cybersecurity," in *Proc. Int. Symp. Netw., Comput. Commun. (ISNCC)*, Oct. 2020, pp. 1–7.

[4] M. Elsayed and M. Erol-Kantarci, "AI-enabled future wireless networks: Challenges, opportunities, and open issues," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 70–77, Sep. 2019.

[5] I. H. Sarker, M. H. Furhad, and R. Nowrozy, "AI-driven cybersecurity: An overview, security intelligence modeling and research directions," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–18, Mar. 2021.

[6] Z. Yan, P. Zhang, and A. V. Vasilakos, "A survey on trust management for Internet of Things," *J. Netw. Comput. Appl.*, vol. 42, pp. 120–134, Jun. 2014.

[7] J. Guo, I.-R. Chen, and J. J. P. Tsai, "A survey of trust computation models for service management in Internet of Things systems," *Comput. Commun.*, vol. 97, pp. 1–14, Jan. 2017.

[8] J. Wang, X. Jing, Z. Yan, Y. Fu, W. Pedrycz, and L. T. Yang, "A survey on trust evaluation based on machine learning," *ACM Comput. Surveys*, vol. 53, no. 5, pp. 1–36, Sep. 2020.

[9] Z.-P. Fan, W.-L. Suo, B. Feng, and Y. Liu, "Trust estimation in a virtual team: A decision support method," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10240–10251, Aug. 2011.

[10] A. Jøsang and D. McAnally, "Multiplication and comultiplication of beliefs," *Int. J. Approx. Reasoning*, vol. 38, no. 1, pp. 19–51, Jan. 2005.

[11] P. J. Denning, "A world lit by flame," *ACM Commun.*, vol. 36, no. 12, pp. 170–171, Dec. 1993.

[12] F. Azzedin and M. Maheswaran, "Integrating trust into grid resource management systems," in *Proc. Int. Conf. Parallel Process.*, Aug. 2002, pp. 47–54.

[13] A. R. Sfar, E. Natalizio, Y. Challal, and Z. Chtourou, "A roadmap for security challenges in the Internet of Things," *Digit. Commun. Netw.*, vol. 4, no. 2, pp. 118–137, Apr. 2018.

[14] A. I. A. Ahmed, S. H. Ab Hamid, A. Gani, S. Khan, and M. K. Khan, "Trust and reputation for Internet of Things: Fundamentals, taxonomy, and open research challenges," *J. Netw. Comput. Appl.*, vol. 145, Nov. 2019, Art. no. 102409.

[15] W. Fang, N. Cui, W. Chen, W. Zhang, and Y. Chen, "A trust-based security system for data collection in smart city," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4131–4140, Jun. 2021.

[16] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H. T. Lin, *Learning From Data*. New York, NY, USA: AML Book, 2012.

[17] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: Classification and comparison," *Int. J. Comput. Trends Technol.*, vol. 48, no. 3, pp. 128–138, Jun. 2017.

[18] M. E. Celebi and K. Aydin, *Unsupervised Learning Algorithms*. Berlin, Germany: Springer, 2016.

[19] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[21] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1449–1457.

[22] R. S. Vitalkar, S. S. Thorat, and D. V. Rojatkar, "Intrusion detection for vehicular ad hoc network based on deep belief network," *Comput. Netw. Inventive Commun. Technol.*, vol. 75, pp. 853–865, Sep. 2022.

[23] K. N. Rao, K. V. Rao, and P. Reddy, "A hybrid intrusion detection system based on sparse autoencoder and deep neural network," *Comput. Commun.*, vol. 180, pp. 77–88, Dec. 2021.

[24] Z. Lin, H. Yanwen, X. Jie, F. Xiong, L. Qiaomin, and W. Ruchuan, "Trust evaluation model based on PSO and LSTM for huge information environments," *Chin. J. Electron.*, vol. 30, no. 1, pp. 92–101, Jan. 2021.

[25] H. Lin, S. Garg, J. Hu, X. Wang, M. J. Piran, and M. S. Hossain, "Data fusion and transfer learning empowered granular trust evaluation for Internet of Things," *Inf. Fusion*, vol. 78, pp. 149–157, Feb. 2022.

[26] W. Ma, X. Wang, M. Hu, and Q. Zhou, "Machine learning empowered trust evaluation method for IoT devices," *IEEE Access*, vol. 9, pp. 65066–65077, 2021.

[27] S. P. Marsh, "Formalising trust as a computational concept," M.S. thesis, Sch. Natural Sci., Univ. Stirling, Stirling, U.K., 1994.

[28] B. Mahbooba, M. Timilsina, R. Sahal, and M. Serrano, "Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model," *Complexity*, vol. 2021, pp. 1–11, Jan. 2021.

[29] N. Jyothi and R. Patil, "An optimized deep learning-based trust mechanism in VANET for selfish node detection," *Int. J. Pervasive Comput. Commun.*, vol. 18, no. 3, pp. 304–318, Dec. 2021.

[30] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, no. 1, pp. 237–285, Jan. 1996.

[31] H. Mayadunna and L. Rupasinghe, "A trust evaluation model for online social networks," in *Proc. Nat. Inf. Technol. Conf. (NITC)*, Oct. 2018, pp. 1–6.

[32] J. Fan, Z. Wang, Y. Xie, and Z. Yang, "A theoretical analysis of deep Q-learning," in *Proc. 2nd Conf. Learn. Dyn. Control*, Jun. 2020, pp. 486–489.

[33] Y. He, C. Liang, F. R. Yu, and Z. Han, "Trust-based social networks with computing, caching and communications: A deep reinforcement learning approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 66–79, Jan. 2020.

[34] G. Rjoub, O. A. Wahab, J. Bentahar, and A. Bataineh, "Trust-driven reinforcement selection strategy for federated learning on IoT devices," *Computing*, pp. 1–23, Apr. 2022, doi: 10.1007/s00607-022-01078-1.

[35] Y. Bai, D. Wang, G. Huang, and B. Song, "A deep reinforcement learning-based social-aware cooperative caching scheme in D2D communication networks," *IEEE Internet Things J.*, vol. 10, no. 11, pp. 9634–9645, Jun. 2023, doi: 10.1109/JIOT.2023.3234705.

[36] M. Ghavipour and M. R. Meybodi, "Trust propagation algorithm based on learning automata for inferring local trust in online social networks," *Knowl.-Based Syst.*, vol. 143, pp. 307–316, Mar. 2018.

[37] M. Ghavipour and M. R. Meybodi, "A dynamic algorithm for stochastic trust propagation in online social networks: Learning automata approach," *Comput. Commun.*, vol. 123, pp. 11–23, Jun. 2018.

[38] X. Chen, Y. Yuan, L. Lu, and J. Yang, "A multidimensional trust evaluation framework for online social networks based on machine learning," *IEEE Access*, vol. 7, pp. 175499–175513, 2019.

[39] J. Lopez and S. Maag, "Towards a generic trust management framework using a machine-learning-based trust model," in *Proc. IEEE Trustcom/BigDataSE/ISPA*, Helsinki, Finland, Aug. 2015, pp. 1343–1348.

[40] S. A. Siddiqui, A. Mahmood, Q. Z. Sheng, H. Suzuki, and W. Ni, "Towards a machine learning driven trust management heuristic for the Internet of Vehicles," *Sensors*, vol. 23, no. 4, p. 2325, Feb. 2023.

[41] G. Liu, C. Li, and Q. Yang, "NeuralWalk: Trust assessment in online social networks with neural networks," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Apr. 2019, pp. 1999–2007.

[42] X. Gao, W. Xu, M. Liao, and G. Chen, "Trust prediction for online social networks with integrated time-aware similarity," *ACM Trans. Knowl. Discovery From Data*, vol. 15, no. 6, pp. 1–30, May 2021.

[43] W. Lin and B. Li, "Medley: Predicting social trust in time-varying online social networks," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, May 2021, pp. 1–10.

[44] W. Fang, C. Zhu, T. M. Ma, W. Zhang, B. Li, L. Yi, F. Xu, T. Zhang, and B. Wang, "Dynamic aging weight scheme for trust model in Internet of Medical Things," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2021, pp. 3366–3369.

[45] B. Mukherjee, L. T. Heberlein, and K. N. Levitt, "Network intrusion detection," *IEEE Netw.*, vol. 8, no. 3, pp. 26–41, May 1994.

[46] R. B. Kagade and S. Jayagopalan, "Optimization assisted deep learning based intrusion detection system in wireless sensor network with two-tier trust evaluation," *Int. J. Netw. Manage.*, vol. 32, no. 4, Feb. 2022, Art. no. e2106.

[47] W. Huang, T. Tiropanis, and G. Konstantinidis, "Federated learning-based IoT intrusion detection on non-IID data," in *Proc. Global IoT Summit (GIoTS), Internet Things*, Dublin, Ireland, Jun. 2022, pp. 326–337.

[48] R. Alghamdi and M. Bellaiche, "A cascaded federated deep learning based framework for detecting wormhole attacks in IoT networks," *Comput. Secur.*, vol. 125, Feb. 2023, Art. no. 103014.

[49] O. Friha, M. A. Ferrag, L. Shu, L. Maglaras, K.-K.-R. Choo, and M. Nafaa, "FELIDS: Federated learning-based intrusion detection system for agricultural Internet of Things," *J. Parallel Distrib. Comput.*, vol. 165, pp. 17–31, Jul. 2022.

[50] G. Lingam, R. R. Rout, and D. V. L. N. Somayajulu, "Detection of social botnet using a trust model based on spam content in Twitter network," in *Proc. IEEE 13th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2018, pp. 280–285.

[51] G. Lingam, R. R. Rout, D. Somayajulu, and S. K. Das, "Social botnet community detection: A novel approach based on behavioral similarity in Twitter network using deep learning," in *Proc. 15th ACM Asia Conf. Comput. Commun. Secur.*, Oct. 2020, pp. 708–718.

[52] H. N. Bhor and M. Kalla, "TRUST-based features for detecting the intruders in the Internet of Things network using deep learning," *Comput. Intell.*, vol. 38, no. 2, pp. 438–462, Apr. 2022.

[53] L. Liu, X. Xu, Y. Liu, Z. Ma, and J. Peng, "A detection framework against CPMA attack based on trust evaluation and machine learning in IoT network," *IEEE Internet Things J.*, vol. 8, no. 20, pp. 15249–15258, Oct. 2021.

[54] K. R, P. Kumar, and B. Bhasker, "DNNRec: A novel deep learning based hybrid recommender system," *Expert Syst. Appl.*, vol. 144, Apr. 2020, Art. no. 113054.

[55] A. L. V. Pereira and E. R. Hruschka, "Simultaneous co-clustering and learning to address the cold start problem in recommender systems," *Knowl.-Based Syst.*, vol. 82, pp. 11–19, Jul. 2015.

[56] O. A. Wahab, G. Rjoub, J. Bentahar, and R. Cohen, "Federated against the cold: A trust-based federated learning approach to counter the cold start problem in recommendation systems," *Inf. Sci.*, vol. 601, pp. 189–206, Jul. 2022.

[57] S. Ahmadian, M. Ahmadian, and M. Jalili, "A deep learning based trust- and tag-aware recommender system," *Neurocomputing*, vol. 488, pp. 557–571, Jun. 2022.

[58] H. Daneshvar and R. Ravanmehr, "A social hybrid recommendation system using LSTM and CNN," *Concurrency Comput., Pract. Exper.*, vol. 34, no. 18, Apr. 2022, Art. no. e7015.

[59] M. F. Aljunid and M. Dh, "An efficient deep learning approach for collaborative filtering recommender system," *Proc. Comput. Sci.*, vol. 171, pp. 829–836, Jan. 2020.

[60] S. S. Choudhury, S. N. Mohanty, and A. K. Jagadev, "Multimodal trust based recommender system with machine learning approaches for movie recommendation," *Int. J. Inf. Technol.*, vol. 13, no. 2, pp. 475–482, Jan. 2021.

[61] G. K. Jha, M. Gaur, P. Ranjan, and H. K. Thakur, "A trustworthy model of recommender system using hyper-tuned restricted Boltzmann machine," *Multimedia Tools Appl.*, vol. 82, pp. 8261–8285, Jul. 2022.

[62] G. K. Jha, M. Gaur, and H. K. Thakur, "A trust-worthy approach to recommend movies for communities," *Multimedia Tools Appl.*, vol. 81, no. 14, pp. 19655–19682, Jan. 2022.

[63] M. Kherad and A. J. Bidgoly, "Recommendation system using a deep learning and graph analysis approach," *Comput. Intell.*, vol. 38, no. 5, pp. 1859–1883, Oct. 2022.

[64] B. B. Sinha and R. Dhanalakshmi, "DNN-MF: Deep neural network matrix factorization approach for filtering information in multi-criteria recommender systems," *Neural Comput. Appl.*, vol. 34, no. 13, pp. 10807–10821, Feb. 2022.

[65] H. Singh, M. B. Singh, R. Sharma, J. Gat, A. K. Agrawal, and A. Pratap, "Optimized doctor recommendation system using supervised machine learning," in *Proc. Int. Conf. Distrib. Compt. Netw. (ICDCN)*, Kharagpur, India, Jan. 2023, pp. 360–365.

[66] T. Kulkarni, P. De, S. Lawande, V. Sinha, S. Deshpande, and S. Kelkar, "Trust evaluation and cloud service recommendation system based on machine learning techniques," in *Proc. Int. Conf. Comput., Commun. Green Eng. (CCGE)*, Sep. 2021, pp. 1–6.

[67] K. Rrmoku, B. Selimi, and L. Ahmedi, "Application of trust in recommender systems—Utilizing naive Bayes classifier," *Computation*, vol. 10, no. 1, p. 6, Jan. 2022.

[68] J. O'Donovan and B. Smyth, "Trust in recommender systems," in *Proc. Int. Conf. Intell. Target entity Interfaces*, 2005, pp. 167–174.

[69] J. Zhao, W. Wang, Z. Zhang, Q. Sun, H. Huo, L. Qu, and S. Zheng, "TrustTF: A tensor factorization model using user trust and implicit feedback for context-aware recommender systems," *Knowl.-Based Syst.*, vol. 209, Dec. 2020, Art. no. 106434.

[70] W. Fang, C. Zhu, F. R. Yu, K. Wang, and W. Zhang, "Towards energy-efficient and secure data transmission in AI-enabled software defined industrial networks," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 4265–4274, Jun. 2022.

[71] K. Siau and W. Wang, "Building trust in artificial intelligence, machine learning, and robotics," *Cutter Bus. Tech. J.*, vol. 31, no. 2, pp. 47–53, Mar. 2018.

[72] A. Rahim, M. Y. Durrani, S. Gillani, Z. Ali, N. U. Hasan, and M. Kim, "An efficient recommender system algorithm using trust data," *J. Supercomput.*, vol. 78, no. 3, pp. 3184–3204, Feb. 2022.

[73] E. Ahvar and M. Fathy, "BEAR: A balanced energy-aware routing protocol for wireless sensor networks," *Wireless Sensor Netw.*, vol. 2, no. 10, pp. 793–800, 2010.

[74] Q. Y. Shambour, A. A. Abu-Shareha, and M. M. Abualhaj, "A hotel recommender system based on multi-criteria collaborative filtering," *Inf. Technol. Control*, vol. 51, no. 2, pp. 390–402, Jun. 2022.

[75] W. Fang, W. Zhang, W. Yang, Z. Li, W. Gao, and Y. Yang, "Trust management-based and energy efficient hierarchical routing protocol in wireless sensor networks," *Digit. Commun. Netw.*, vol. 7, no. 4, pp. 470–478, Nov. 2021.

[76] S. A. Siddiqui, A. Mahmood, W. E. Zhang, and Q. Z. Sheng, ''Machine learning based trust model for misbehaviour detection in Internet-of-Vehicles,'' in *Proc. Int. Conf. Neural Inf. Proc. (ICONIP)*, Sydney, NSW, Australia, 2019, pp. 512–520.

ZHIQI LI received the B.E. degree in electronic and information engineering from Nanjing Normal University, Nanjing, China, in 2022. He is currently pursuing the M.E. degree with the Shanghai Institute of Microsystem and Information Technology (SIMIT), Chinese Academy of Sciences, Shanghai, China. His research interests include machine learning and trust model.

WEIDONG FANG (Member, IEEE) received the B.E. degree in industrial electrical automation from Shandong University, Jinan, China, in 1993, the M.E. degree in communication and electronic systems from the China University of Mining & Technology, Beijing, in 1998, and the Ph.D. degree in electromagnetic fields and microwave techniques from Shanghai University, Shanghai, China, in 2016. He is currently an Associate Professor with the Shanghai Institute of Microsystem and Information Technology (SIMIT), Chinese Academy of Sciences, Shanghai, China. His current research interests are information security and energy efficiency in the IoT/WSN/VANET, including trust model, secure network coding, and secure routing protocol.

CHUNSHENG ZHU (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from The University of British Columbia, Canada, in 2016. He is an Associate Professor with the College of Big Data and Internet, Shenzhen Technology University, China. He has authored more than 100 publications. His research interests mainly include the Internet of Things, wireless sensor networks, cloud computing, big data, social networks, and security.

ZHIWEI GAO received the Ph.D. degree in control theory and control engineering from Tongji Unviersity, China. He is a Senior Engineer with Ceprei Certification Body. He has authored more than ten publications published by refereed international journals and conferences. His research interests mainly include cloud computing and the Internet of Things (IOT).

WUXIONG ZHANG (Member, IEEE) received the B.E. degree in information security from Shanghai Jiao Tong University, Shanghai, China, in 2008, and the Ph.D. degree in communication and information systems from the Shanghai Institute of Microsystem and Information Technology (SIMIT), Chinese Academy of Sciences, Shanghai, in 2013. He is currently a Professor with SIMIT. His research interests include beyond third-generation mobile communication systems and vehicular networks.