Wang, Jingwen; Jing, Xuyang; Yan, Zheng; Fu, Yulong; Pedrycz, Witold; Yang, Laurence T.

A Survey on Trust Evaluation Based on Machine Learning

Published in:
ACM Computing Surveys

DOI:
10.1145/3408292

Published: 28/09/2020

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Wang, J., Jing, X., Yan, Z., Fu, Y., Pedrycz, W., & Yang, L. T. (2020). A Survey on Trust Evaluation Based on Machine Learning. *ACM Computing Surveys*, *53*(5), Article 107. https://doi.org/10.1145/3408292

A Survey on Trust Evaluation Based on Machine Learning

JINGWEN WANG and XUYANG JING, Xidian University
ZHENG YAN, Xidian University and Aalto University
YULONG FU, Xidian University
WITOLD PEDRYCZ, University of Alberta
LAURENCE T. YANG, St. Francis Xavier University

Trust evaluation is the process of quantifying trust with attributes that influence trust. It faces a number of severe issues such as lack of essential evaluation data, demand of big data process, request of simple trust relationship expression, and expectation of automation. In order to overcome these problems and intelligently and automatically evaluate trust, machine learning has been applied into trust evaluation. Researchers have proposed many methods to use machine learning for trust evaluation. However, the literature still lacks a comprehensive literature review on this topic. In this article, we perform a thorough survey on trust evaluation based on machine learning. First, we cover essential prerequisites of trust evaluation and machine learning. Then, we justify a number of requirements that a sound trust evaluation method should satisfy, and propose them as evaluation criteria to assess the performance of trust evaluation methods. Furthermore, we systematically organize existing methods according to application scenarios and provide a comprehensive literature review on trust evaluation from the perspective of machine learning's function in trust evaluation and evaluation granularity. Finally, according to the completed review and evaluation, we explore some open research problems and suggest the directions that are worth our research effort in the future.

CCS Concepts: • **Computing methodologies** → *Machine learning*; *Learning paradigms*;

Additional Key Words and Phrases: Trust evaluation, machine learning, performance metrics, evaluation requirements

The work is supported in part by the National Natural Science Foundation of China under Grants 61672410 and 61802293, the National Postdoctoral Program for Innovative Talents under grant BX20180238, the Project funded by China Postdoctoral Science Foundation under grant 2018M633461, the Academy of Finland under Grants 308087, 314203 and 335262, the open grant of the Tactical Data Link Lab of the 20th Research Institute of China Electronics Technology Group Corporation, P.R. China under grant CLDL-20182119, the Shaanxi Innovation Team project under grant 2018TD-007, and the 111 project under grant B16037.

Authors' addresses: J. Wang, X. Jing, and Y. Fu, State Kay Lab of ISN, School of Cyber Engineering, Xidian University, 266 Xinglong Section of Xifeng Road, Xi'an, Shaanxi 710126, China; emails: {wjwen0914, xuyangjing91}@163.com, ylfu@xidian.edu.cn; Z. Yan, State Key Lab of ISN, School of Cyber Engineering, Xidian University, No. 2 South Taibai Road, Xi'an, Shaanxi 710071, China and the Department of Communications and Networking, Aalto University, Konemiehentie 2, P.O.Box 15400, Espoo 02150, Finland; email: zyan@xidian.edu.cn; W. Pedrycz, the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, T6R 2V4, Canada; email: wpedrycz@ualberta.ca; L. T. Yang, the Department of Computer Science, St. Francis Xavier University, Antigonish, NS, B2G 2W5, Canada; email: ltyang@gmail.com.

0360-0300/2020/09-ART107 $15.00
https://doi.org/10.1145/3408292

ACM Reference format:
Jingwen Wang, Xuyang Jing, Zheng Yan, Yulong Fu, Witold Pedrycz, and Laurence T. Yang. 2020. A Survey on Trust Evaluation Based on Machine Learning. *ACM Comput. Surv.* 53, 5, Article 107 (September 2020), 37 pages.
https://doi.org/10.1145/3408292

1 INTRODUCTION

Trust evaluation is the process of quantifying trust using attributes that affect trust. It has been widely used in different fields to facilitate decision making [Jiang et al. 2016; Chen et al. 2014]. In a sensor network, trust assessment based on malicious node detection can help ensure the security of the network [Feng et al. 2011; He et al. 2012; Ye et al. 2017; Zhang et al. 2018]. In a social network, trust evaluation helps users build social relationships, reduce the risk of social activities and improve the quality of social networking [Hao et al. 2014; Jiang et al. 2014; Niu et al. 2017]. In e-commerce, researchers use trust evaluation to help them choose trading services [Wang and Lin 2008; Zhang et al. 2014]. Trust evaluation is also used to ensure secure interactions between automated agents and to promote the success of automation in a multi-agent system [Schmidt et al. 2007]. In Peer-to-Peer (P2P) networking, trust evaluation helps identifying interactive objects, ensuring resource sharing with friendly peers, and combating malicious peers [Li et al. 2011]. In service provision, trust evaluation helps service requesters select an appropriate service from a large number of candidates [Li et al. 2014; Liu and Jia 2015; Liu et al. 2014; He et al. 2018]. Therefore, it becomes evident that research direction in trust evaluation is worth pursuing.

There are some existing surveys on trust evaluation. Jøsang et al. [2007] summarized the trust and reputation system used in online transactions, including some trust evaluation methods. In the context of Internet of Things (IoT), Yan et al. [2014b] investigated existing trust management schemes and summarized the trust evaluation methods used there. Guo et al. [2017] analyzed the trust evaluation methods for managing services in IoT. Wahab et al. [2015] surveyed trust evaluation methods for web services. Pinyol and Sabater-Mir [2013] presented a review of trust evaluation methods for open multi-agent systems. Zhu and Yan [2016] surveyed the methods of trust evaluation in e-commerce. Bansal and Kohli [2019] summarized trust evaluation methods for website. Ahmed et al. [2019] gave a summary of trust assessment schemes in cross-cloud alliance. Although the above literature summarizes the methods of trust evaluation for various scenarios, there is a lack of a review on trust evaluation methods based on machine learning.

Most traditional trust evaluation methods are based on the experience of direct and indirect interaction between a trustor and a trustee. However, when there is no interaction experience between the trustor and the trustee, traditional trust evaluation methods are not so applicable. At the same time, there are cases where the data used for trust evaluation are incomplete and the evaluation process ignores other valuable data, which greatly impact the accuracy of trust evaluation. Meanwhile, most traditional trust evaluation methods determine trust by aggregating trust factors through weighting and other relevant calculations. However, it is difficult to determine the weights, thus hard to ensure evaluation accuracy. In response to these problems, many researchers suggested using machine learning to make trust evaluation intelligent and accurate.

Meanwhile, the data used to evaluate trust is becoming big inevitably due to the fast growth of online services, social networking and mobile communications. In recent years, big data has gradually become a research hotspot in both academic and industrial circles. In brief, it comes with visible features such as Velocity, Volume, Value, Variety, and Veracity (5V). Due to the complex, high-dimensional, and variable characteristics of big data, it is difficult to conduct trust evaluation in a real, messy, and complex big data environment. Big data often accompanies with a large-scale distributed computing scenario, e.g., social networking, pervasive computing, P2P network-

ing and grid computing. Due to large amount of data and their complex structure, traditional trust evaluation methods become ineffective to calculate trust values that can be analyzed based on techniques of big data. As an important and commonly used technology for dealing with big data, machine learning can systematically and effectively process data and improve computational efficiency. Therefore, many researchers have proposed using machine learning to evaluate trust.

Machine learning aims to generate a model from data with computers, namely learning through machines. As early as the late 1950s, people have been studying it since the advent of Artificial Intelligence (AI) [Martensa 1959], and it is the core of AI. It can build models based on data and use models to simulate human intelligence activities. At present, there have been a lot of machine learning algorithms proposed. According to their characteristics, these algorithms realize several categories of learning: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning [Jing et al. 2018].

Researchers have proposed many methods by using machine learning for trust evaluation. However, the literature still lacks a comprehensive literature review on this topic. We found that most existing related literature reviews focus on traditional trust evaluation methods. Although Liu et al. [2018] summarized existing methods of using machine learning for trust prediction, they only focused on the methods in online social networking. However, the application of trust evaluation is not limited to social networks, but also many other scenarios, such as multi-agent systems, service environments and ad-hoc networks. In this article, we offer a comprehensive summary and analysis on trust evaluation methods based on machine learning in a number of typical application domains. Nowadays, with the popularity of artificial intelligence and the ubiquity of big data, it becomes necessary to summarize the existing trust evaluation methods that adopt machine learning, understand the pros and cons of these methods, specify research challenges and determine new directions of future research efforts, hoping to improve existing methods and propose efficient trust evaluation methods with the ability of dealing with big data.

This article conducts a comprehensive review of the state-of-the-art of trust evaluation based on machine learning in order to reach the above goals. The main contributions of this article can be summarized as follows:

(i) We analyze the requirements that should be satisfied for evaluating trust with machine learning. Employing these requirements as criteria, we assess the performance of trust evaluation methods that adopt machine learning.
(ii) We review the existing methods by classifying them based on application scenarios and further the machine learning's function in trust evaluation and evaluation granularity, and further utilize the proposed criteria to evaluate the advantages and shortcomings of each method.
(iii) According to the results of the completed review and analysis, we identify open issues and further indicate directions of future research.

We organize the remainder of this article as follows: Section 2 gives a brief introduction on the fundamentals of both trust evaluation and machine learning, followed by the requirements that should be satisfied by the trust evaluation methods. According to the proposed requirements, we offer a taxonomy of existing trust evaluation methods, identify a number of categories, and review them in Section 3. On the basis of the literature review, we explore unsolved research issues and point out promising future research directions in the field of machine learning-based trust evaluation in Section 4. Finally, we conclude this survey in the last section.

2 PREREQUISITES

In this section, we briefly introduce the characteristics of trust and some traditional trust evaluation methods, as well as some basic knowledge of machine learning. We also discuss the advantages of using machine learning to evaluate trust. Based on the discussion, we put forward the requirements that a qualified trust evaluation method should meet, which can serve as the uniform criteria used to evaluate all existing methods.

2.1 Trust Evaluation

2.1.1 The Characteristics of Trust. Trust is a multidisciplinary concept that is not easy to define. Trust is fundamentally a belief or estimate. It is an entity's subjective expectations of the future behavior of other entities [Mui et al. 2002]. Different scholars have different definitions of trust.

Although there are different definitions of trust in different areas, trust exhibits the following consistent characteristics [Yan 2013]: subjectivity, dynamic, context-awareness, incomplete transitivity, time decay, asymmetry, and measurability.

2.1.2 Traditional Trust Evaluation Methods. Trust evaluation is the process of quantifying trust with attributes that influence trust. The methods of trust evaluation show diversity in different application scenarios.

Traditional trust evaluation methods mainly use trust-related attributes to assess trust [Yan 2010, 2013]. The evaluation models include but are not limited to: Bayesian inference [Wang and Wu 2014], (weighted) average models [Yan 2013], subjective logic [Jøsang et al. 2007], Dempster-Shafer theory [Wu et al. 2015], fuzzy logic [Alnasser and Sun 2017], fuzzy cognitive map [Yan and Prehofer 2011], information entropy model [Dai et al. 2008], game theory [Mehdi et al. 2017], and cloud models [Zhang et al. 2018]. Table 1 shows the advantages and disadvantages of a number of traditional trust evaluation methods. Although these methods can measure trust, they also put forward high formal mathematical requirements for evaluation. Sometimes, it is hard to apply the models into practice. The selection of trust-related attributes and the formulation of rules greatly affect the accuracy of trust evaluation. Existing work mainly applies static data analysis (such as exploratory factor analysis and confirmatory factor analysis [Yan et al. 2012, 2013b]) and literature analysis for justification. It lacks intelligence and dynamic support on trust impact factor selection and rule formulation.

Many existing methods apply linear aggregation to aggregate different trust-impact attributes during trust evaluation in such application domains as multi-agent systems, social networks [Yan et al. 2013a, 2014a], service environments, and ad-hoc networks.

Through the analysis of the traditional trust evaluation methods in the above application scenarios, we find that when prior knowledge is lacking, the traditional trust evaluation methods are no longer applicable. At the same time, using linear combination to aggregate trust features such as experience, knowledge, recommendations and so on to express trust seems rough without theoretical and practical support. Thus, trust evaluation accuracy could be suspicious.

2.2 Benefits of Machine Learning for Trust Evaluaton

From the above introduction to traditional trust evaluation methods, it can be seen that there are some irreplaceable advantages to use machine learning for trust evaluation.

First, trust evaluation based on machine learning can overcome the "cold start" and "zero knowledge" problems of traditional methods. Traditional trust evaluation methods use direct historical interaction information and indirect recommendation information to calculate trust values. But when the trustee is a newcomer, this information does not exist, which leads to traditional methods become ineffective. In this case, the trust evaluation based on machine learning can establish a trust model by using other available trust-related feature data to perform trust evaluation. Mean-

Table 1. Comparison of Traditional Trust Evaluation Methods

Trust Evaluation Method	Advantages	Disadvantages
Bayesian Inference	Using Bayesian inference to calculate probability to represent trust value, the method is simple.	Confusing the subjectivity and uncertainty of trust with randomness of probability.
Dempster-Shafer Theory	Do not need to know the *a priori* probability, have a good representation of uncertainty.	The data need to be independent, the theory of synthetic rules is weak.
Fuzzy Logic	A good solution of the fuzzy problem of trust.	The establishment of rules and the determination of membership functions could be difficult.
Subjective Logic	Introduce subjective factor and facilitate trust reasoning and comprehensive calculation.	The construction of operators is difficult.
Entropy-based model	Information entropy is used to portray the uncertainty of trust relationship.	Insufficient consideration of factors that affect trust.
Game Theory	Theory of game theory is broad, while paying attention to the degree of completeness of information, which makes the results close to reality.	It is difficult to formulate the rules of game.
Cloud-based	It fully integrates ambiguity and randomness and describes the uncertainty of trust.	Trust combination and transfer calculation often utilize cloud theory. It is still difficult to use it to calculate trust value based on trust evidence.

while, many traditional trust evaluation methods represent trust by using a linear combination of direct and indirect trust values, where the weights used for combination are hard to decide in many practical cases. This type of method impacts the accuracy of trust evaluation, which hopefully can be improved with the methods of machine learning.

Second, in the case of processing big data to evaluate trust, using machine learning can ensure accurate results. Big data provides sufficient sources of data for trust evaluation, making it more accurate [Huang and Chen 2019]. In large-scale networking scenarios such as social networks, due to enormous data and complex data structure, the use of traditional trust evaluation methods becomes complicated and difficult, which often leads to inaccurate evaluation results. However, machine learning as a primary means of processing big data has its own specific advantages in dealing with big data. Therefore, in dealing with big data, compared with other trust evaluation methods, machine learning is obviously more appropriate for trust evaluation [Han et al. 2019].

Third, from the perspective of artificial intelligence, machine learning shows great convenience for trust evaluation. Machine learning, which is the main technology of artificial intelligence, also examines how to use computers to simulate human behaviors. It uses existing data (or experiences) to get a model for later computation. This process fits well with human thinking patterns. Trust evaluation is a subjective behavior of human beings. Thus, it is appropriate to use machine learning to evaluate trust.

Fourth, the process of trust evaluation using machine learning is intuitive and easy to understand. The basic process of using machine learning to solve the problem of trust evaluation is certain. Its general process can be roughly divided into three steps, namely data preprocessing, model selection, and final model determination. The details are as follows: First, we need to transform the raw data into useful features through data preprocessing, which is the process of extracting

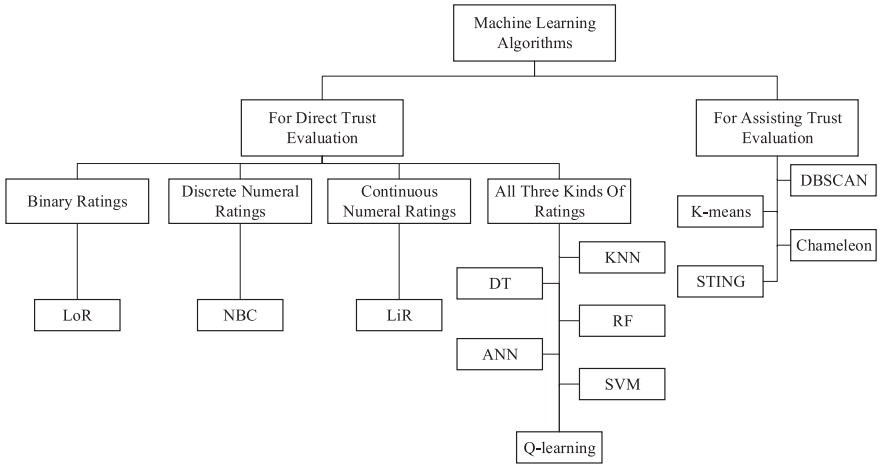

Fig. 1. Classification of machine learning algorithms based on their functions in trust evaluation and evaluation granularity.

meaningful features from raw data with missing values, repeated values, noise, and high dimensions by means of data cleaning, data fusion, feature selection, and other methods. Then, we need to choose the most appropriate learning algorithm from the many possible algorithms of machine learning to build a model for trust evaluation. Finally, because there are often parameters in the algorithm that need to be set, the performance of the model is often significantly different due to different parameter configurations. Therefore, after selecting the algorithm, the parameters should be adjusted and the parameters with the best performance should be selected to determine the final model. The process of trust evaluation using machine learning simulates the process of human decision-making, that is, making decisions based on existing experiences. This process is simple and effective to implement and easy to understand.

2.3 Machine Learning

Stefik [1985] defined machine learning as: "a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E." Machine learning relates to pattern identification, statistical learning, data mining, speech recognition, computer vision, natural language processing, and intrusion detection. Due to the needs of big data processing and requests of artificial intelligence in many fields, machine learning becomes more and more important. Machine learning has been widely applied into analyzing big data. Its main target is to generate models from data with computers by applying a learning algorithm. Thus, later on, when facing a new situation, the model will help in corresponding judgment. Nowadays, there are a lot of learning algorithms. By analyzing the machine learning algorithms used for trust evaluation, we classify them into two main categories according to their functions played in trust evaluation. One is for direct trust evaluation, and the other is for assisting trust evaluation (e.g., for data processing) before another method is applied to evaluate trust, as shown in Figure 1. For the first type of algorithms, we can further divide them into a number of sub-classes based on trust evaluation granularity: binary ratings, discrete numeral ratings, and continuous numeral ratings, as well as the algorithms that can support all above three ratings. In what follows, we briefly introduce these algorithms and summarize their advantages and disadvantages in Table 2.

Table 2. Comparison of Machine Learning Methods

Algorithms	Description	Advantages	Disadvantages
LiR	The main task of determining a regression model is accessing the parameters of the linear function, which can be obtained by minimizing a loss (cost) function that is used to measure a fitting degree.	Simple calculation and implementation. Support continuous numerical trust value in trust evaluation.	Cannot fit non-linear data.
LoR	As a classification algorithm, the output variables of LoR model should be qualitative variables. So, the output variables of the LoR model are qualified by setting a threshold.	Simple to implement, small amount of calculation, fast speed and low storage. Support binary ratings in trust evaluation.	Not applicable to large feature space, easy to underfit.
KNN	First, compare each feature in the new data with the corresponding feature in a labeled training dataset. Then, in the labeled dataset, extract the classification labels of the k most similar data to the new data. Finally, the category of the new data is the one that is the most frequent occurrence in a number of k category labels.	Model is easy to understand. Support binary ratings, continuous numerical and discrete numerical trust value in trust evaluation.	High dependence on sample balance.
NBC	Use Bayes theorem to classify by calculating the probability of different categories under specific conditions.	Have a solid mathematical foundation, insensitive to missing data. The algorithm is simple. Support binary ratings and discrete numerical trust value in trust evaluation.	Cannot apply to feature-related datasets.
DT	Two kinds of nodes exist in the DT. One is the non- leaf node that corresponds to a feature or attribute and has two or more child nodes. The other is the leaf node that corresponds to a category. Each edge in the DT represents a decision condition.	Can be visualized. The decision-making process is intuitive and clear. Support binary ratings, continuous numerical and discrete numerical trust value in trust evaluation.	Easy to overfit, poor generalization performance.
RF	RF essentially consists of multiple decision trees, each of which is slightly different from others.	No overfitting, can be used in high- dimensional datasets. Support binary ratings, continuous numerical and discrete numerical trust value in trust evaluation.	Low performance for small datasets and low-dimensional datasets.
SVM	SVM aims to find the maximum-margin hyperplane, which means that the interval between the hyperplane and the nearest data point on each side is the largest.	Good for high- and low-dimensional data processing, strong generalization ability. Support binary ratings, continuous numerical and discrete numerical trust value in trust evaluation.	Determining a kernel function based on actual problems is difficult.
ANN	It is composed of a series of simple units that are arranged in layers and densely connected to each other. Input variables are processed by these units to get output variables.	High accuracy, strong robustness, and fault tolerance. Support binary ratings, continuous numerical and discrete numerical trust value in trust evaluation.	Many parameters, complex model and low interpretability.
K-Means	It is a partitioning method based on distance to organize objects into multiple mutually exclusive groups or clusters.	Simple and easy to understand and implement, low time complexity.	Need to set an initial value, very sensitive to noise and outliers.

(Continued)

Table 2. Continued

Algorithms	Description	Advantages	Disadvantages
DBSCAN	The basic idea of DBSCAN is to start from a certain core point. Then the method continuously expands to a region with a high density of data, so as to obtain a maximal region containing the core point and boundary points [Ester et al. 1996]. Any two points in the region are density-connected.	Can identify noise point and find spatial clustering of any shape, high clustering speed.	When density is uneven, clustering quality is poor.
Chameleon	Chameleon is a graph partitioning algorithm. It is used to gain some relatively small sub-clusters from the *k*-nearest neighbor graph in order to minimize edge cutting.	It has a strong ability to find high-quality clusters of any shape.	The choice of the value of K is a problem.
STING	It is a kind of grid-based clustering technique [Wang et al. 1997]. On the basis of grid multiresolution, the space is divided into matrix units corresponding to different resolutions.	Grid structure facilitates parallel processing and incremental updates. It is efficient.	The bottom-level granu- larity has a great impact on cluster quality. Too thick or too fine could cause some problems.
Q-learning	Its main idea is that the agent perceives information from the environment, selects appropriate behaviors according to its own state, changes its state and obtains corresponding rewards or punishments from it, thereby correcting strategy accordingly.	It is adaptive and can improve strategy based on environmental feedback. Support binary ratings, continuous numerical and discrete numerical trust value in trust evaluation.	When the number of state-action pairs is large, convergence is slow and learning is inefficient.

2.3.1 Machine Learning Algorithms for Direct Trust Evaluation. In trust evaluation schemes, these machine learning algorithms are used to directly calculate trust values or judge whether the evaluated object is trustworthy based on trust-related attributes. In this type of machine learning algorithms, based on learning granularity of machine learning (i.e., the granularity of the result data of trust evaluation), we divide the existing algorithms into binary ratings machine learning, discrete numeral ratings machine learning, continuous numeral ratings machine learning, and machine learning for all three types. The following is our brief introduction to these algorithms.

2.3.1.1 Algorithms for Binary Ratings. Using this type of machine learning algorithms for trust evaluation, evaluation results are divided into two classes: trusted and untrusted. Generally, the binary classification algorithms in machine learning can achieve this goal. The following lists these algorithms:

Logistic Regression (LoR). LoR is one type of classification algorithms. The establishment process of LoR model is roughly the same as that of LiR. Their difference is that the range of output variables in LiR is controlled to be within [0, 1] by applying the Sigmoid function [Tolles and Meurer 2016]. Similar to LiR, the modeling process of LoR is simple and fast. But it is not applicable for processing non-linear data. In addition, LoR is a binary classification algorithm, not applicable to multi-classification problems.

2.3.1.2 Algorithms for Discrete Numeral Ratings. Based on this type of algorithms for trust evaluation, the final evaluations results are expressed by discrete numbers. For example, López and Maag used the numbers 0, −1, and 1 to represent trust evaluation results, where 0 indicates "neutrally trusted", 1 indicates "trustworthy", and −1 indicates "untrustworthy" [López and Maag 2015].

Naive Bayes Classification (NBC). NBC is a common classification algorithm supported by Bayesian theory [Russell and Norvig 2009]. NBC based on classical mathematics theory has a

solid mathematical foundation. It is simple with stable classification efficiency. However, it is not sensitive to missing data. The algorithm assumes that the features are independent with each other, but this assumption is often not true in practical applications. When there are a large number of features or the correlation between features is high, the classification efficiency is not good.

2.3.1.3 Algorithms for Continuous Numeral Ratings. The use of this type of algorithms in trust evaluation can eventually obtain continuous numerical trust values. For example, Hauke et al. [2013] used decimals between 0 and 1 to indicate the degree of trust. The closer to 0 the value is, the lower the trust degree, and vice-versa.

Linear Regression (LiR). LiR is the most basic algorithm in machine learning. As the name implies, it represents a linear relationship between input and output. That is, the model is a linear function that maps input variables to output variables [Freedman 2005]. LiR is a regression model based on linear variables, usually the modeling process is simple and fast. However, it is not suitable for non-linear data.

2.3.1.4 Algorithms for All Three Types of Ratings. This type of algorithms is suitable for obtaining binary ratings, discrete numeral ratings, and continuous numeral ratings.

K-Nearest Neighbor (KNN). There are three steps to classify new data with KNN [Altman 1992]. A KNN algorithm is highly dependent on sample balance regarding each category of samples. When samples are extremely unbalanced, the classification results will be biased.

Decision Tree (DT). DT organizes rules in a tree to classify samples [Quinlan 1987]. The generation process of a decision tree is very intuitive, easy to understand and explain. However, when the learning of training data is too thorough, it is easy to cause overfitting, which leads to poor model generalization.

Random Forest (RF). RF essentially consists of multiple decision trees, each of which is slightly different from others [Ho 1998]. RF is good at dealing with high-dimensional data, feature loss data, and unbalanced data. But it may be less effective in processing low-dimensional data.

Support Vector Machine (SVM). As a kind of classification algorithm, SVM aims to find the maximum-margin hyperplane, which means that the interval between the hyperplane and the nearest data point on each side is the largest [Cortes and Vapnik 1995]. It can simplify the difficulty of solving high-dimensional space problems. Therefore, it can process high-dimensional and low-dimensional data. But determining the kernel function based on actual problems is a difficult task.

Artificial Neural Networks (ANN). ANN simulates a biological learning system, that is, an extremely complex network of interconnected neurons [Dreyfus 1990]. Generally, ANN has high accuracy and strong parallel distributed processing ability. However, it requires a large number of parameters, such as initial values of weights, thresholds, and the settings of network topology. The learning process cannot be observed and the output results are difficult to explain, which affect the credibility and acceptability of the results.

Q-Learning. Q-learning is a common algorithm in reinforcement learning. Its main idea is that the agent perceives information from the environment, selects appropriate behaviors according to its own state, changes its state, and obtains corresponding rewards or punishments from it, thereby correcting strategy accordingly [Watkins and Dayan 1992]. Its strategy is updated in real time. But modeling realized with this algorithm is complicated and difficult to understand.

2.3.2 Machine Learning Algorithms for Assisting Trust Evaluation. This type of machine learning algorithms plays an auxiliary role in trust evaluation tasks. They can be utilized to process data

for trust evaluation. For example, in the absence of trust attribute information, machine learning is used to cluster similar users according to user attributes, and then for those users lacking trust information, trust evaluation is performed with the trust data of the user that does not lack trust information [Zhang et al. 2016].

K-Means Algorithm. K-means is a partitioning method based on distance [Lloyd 1982] to organize objects into multiple mutually exclusive groups or clusters. However, it is hard to determine the value of K and is also sensitive to noise and outliers.

DBSCAN Algorithm. The full name of DBSCAN is "Density-Based Spatial Clustering of Applications with Noise." It is a density-based clustering algorithm. DBSCAN can discover arbitrarily shaped clusters in a noisy spatial data set by dividing sufficiently high-density areas. The basic idea of DBSCAN is to start from a certain core point. Then the method continuously expands to a region with a high density of data, so as to obtain a maximal region containing the core point and boundary points [Ester et al. 1996]. Any two points in the region are density-connected. This clustering algorithm is efficient. It can handle noise points and derive arbitrary shapes of spatial clusters in an effective way. However, when the data density is not uniform and the distance between clusters is far apart, the quality of clustering is poor.

Chameleon. Chameleon is a kind of hierarchical clustering algorithm. It uses knowledge of *k*-nearest neighbors and dynamic modeling [Karypis et al. 1999]. Chameleon does not rely on a static and user-provided model. It can automatically adapt to the internal features of the merged cluster. This merging process facilitates the discovery of high-quality clusters of arbitrary shapes. However, it uses a KNN graph, so the selection of the value of K is a problem.

STING. It is a kind of grid-based clustering technique [Wang et al. 1997]. The clustering quality of STING is greatly influenced by the bottom-level granularity of the grid structure. If the granularity is very small, the processing cost will increase significantly. If the granularity is too coarse, the quality of the cluster analysis will be reduced. Therefore, granularity determination is important.

2.4 Requirements and Criteria

In this section, we propose a set of requirements that can serve as the criteria for evaluating trust evaluation methods based on machine learning.

Effectiveness. The most basic and important requirement on a trust evaluation method is that it can accurately provide the trust value of a trustee. Therefore, a qualified trust evaluation method must ensure accurate evaluation results to prove its effectiveness. Effectiveness can be represented by a number of indicators, such as recall, precision, accuracy, and F-score. Table 3 shows the formulas used to compute the above indicators.

Appropriate Data and Algorithm. To study trust evaluation with machine learning, we should consider its two important parts. One is the data for training the model, the other is the algorithm for building the model. Choosing the right data and algorithms helps achieve a highly accurate evaluation. Therefore, we should explore whether a method considers the impacts of training data selection and algorithm selection on the trust evaluation.

Robustness. Attacks may occur during the process of trust evaluation, for instance, conflict behavior attack, on-off attack, collision attack, Sybil whitewashing attack and bad-mouthing attack (refer to Table 4 for details) [Yan et al. 2012, 2013a, 2014b]. These attacks have an impact on the outcome of the trust evaluation, allowing us to get a result that is contrary to the fact. Trust

Table 3. Calculation Formulas of Effectiveness Indicators

Indicators	Formulas
Accuracy	$accuracy = \frac{TP+TN}{TP+FN+FP+TN}$
Precision	$precision = \frac{TP}{TP+FP}$
Recall	$recall = \frac{TP}{TP+FN}$
F-score	$F-score = \frac{P \times R}{2 \times (P+R)}$
True Positive (TP) represents the quantity of outcomes that predict a positive class as a positive class. False Positive (FP) represents the quantity of outcomes that predict a negative class as a positive class. True Negative (TN) represents the quantity of outcomes that predict a negative class as a negative class. False Negative (FN) represents the quantity of outcomes that predict a positive class as a negative class. P: Precision; R: Recall.	

Table 4. Attacks on Trust Evaluation

Attack	Description
Bad-mouthing attack	This attack usually exists in a trust management or reputation system that considers recommendations. It defames the trustworthiness of good parties or inflate the trustworthiness of malicious parties by providing dishonest recommendations, feedback, or votes.
On-off attack	When the trust values of malicious nodes performing on-off attacks are significantly reduced, attackers can perform good behaviors to increase their trust values over a period of time. And when their trust values reach a certain level, they begin to execute malicious behaviors.
Conflict behavior attack	By carrying out different behaviors in different time or domains, attackers can cause conflicts with normal users, thus impairing the recommendation trust of good users.
Collusion attack	Several attackers have reached an agreement to form a conspiracy group and control the trust evaluation result of a certain target by submitting false feedback in an organized manner.
Whitewashing attack	Attackers with very low trust discard their existing identity and, as newcomers, hide their bad history and reenter a system.

evaluation methods should consider preventing these attacks so that the trust evaluation methods are not affected by these attacks. We use robustness to express this quality attribute.

Privacy Protection. The data used for trust evaluation will inevitably contain some private information that users do not want to disclose. In the process of conducting trust evaluation, we should give full protection to ensure that private data should not be leaked, as data owners expect. Such a trust evaluation method can be well accepted and recognized since it is a responsible method. Therefore, privacy protection of users and trust evidence should be considered while conducting trust evaluation.

Context-Awareness. It is a basic characteristic of trust. Trust evaluation methods should support context-awareness. That is, when the application scenario, context or environment changes, the evaluation scheme can sense it and adaptively adjust itself to dynamically fit into a new situation.

Subjectivity. It is also a basic characteristic of trust. The trust evaluation should be able to portray the subjectivity of trust. With this way, the expression of trust can be close to reality.

Trust evidence used for trust evaluation should reflect the subjective view of the trustor. The trust evaluation result should be personalized based on the judgement of a particular individual.

Computational Overhead. It is an indicator stating how efficient a given method is. The lower the computational overhead, the faster the algorithm runs, indicating its high efficiency. The efficiency of a trust evaluation method can be reflected by the computational overhead of its applied machine learning algorithm.

3 MACHINE LEARNING FOR TRUST EVALUATION

In this section, we first classify the trust evaluation methods into six categories based on their application scenarios, and then group the existing work based on the function played by the machine learning in trust evaluation, i.e., machine learning for direct trust evaluation and machine learning for assisting trust evaluation. For the first type that uses machine learning to evaluate trust directly, we further classify existing methods into three sub-classes based on trust evaluation granularity: models with binary ratings, models with discrete numeral ratings, and models with continuous numeral ratings. Next, we evaluate various trust evaluation methods by utilizing the criteria proposed in the previous section. Finally, we analyze and discuss the existing solutions based on the results of our literature review. Table 5 presents a review summary.

3.1 Trust Evaluation Based on Machine Learning in Social Networks

In this section, we review existing trust evaluation methods based on machine learning in social networks.

3.1.1 Machine Learning for Direct Trust Evaluation.

3.1.1.1 Models with Binary Ratings. Liu et al. [2008] presented a method to use user behavior information to establish a trust network. They considered the features obtained from the attributes of individual users and the attributes of the interactions between the users. Then, for training a model, they utilized a supervised learning approach to gain a model, which was used to derive the users' trust relationship. They divided the trust factors into user factors and interaction factors, made a more detailed division between the two types of factors, and selected algorithms such as DT, NB, LoR, and SVM to perform experiments on the data collected from Epinion. Experimental results show that using NB or SVM classifier with interactive features performs better than using only user features. The features they selected for trust prediction are of statistical nature. Thus, this method does not capture subjectivity. Context awareness is not supported, either. The protection of the privacy of trust evidence and resistance against possible attacks were not explored. Computational overhead was not evaluated in this work, either.

Zolfaghar and Aghaie [2011] proposed a trust prediction method, which is time-aware, to constantly update a trust network. They predicted the current moment of trust connection based on the previous moment of trust network connection relationship. In terms of feature selection, they selected the features obtained through the calculation corresponding to the five factors of knowledge, relationship, reputation, similarity, and personality. They used MLP algorithm to predict trust. Both static and dynamic approaches were applied. The difference is that the static method uses only one snapshot of the trust network and dynamic one uses a series of snapshots before the time of evaluation. Then, they conducted experiments based on Epinion data and found that the dynamic approach is more effective and supports context awareness. However, privacy protection and malicious attack resistance were not considered. Its F-score is 0.8 and its performance is moderate. This method is applicable for social web applications. Notably, the trust factors used in the method are not subjective. Computational overhead was not mentioned in this work.

Table 5. Summary and Comparison of Trust Evaluation Methods Using Machine Learning (a)

Role of ML	Type of ML	Ref.	Sc	Al	SF	TR	Ef	PP	Rb	Ca	Su	CO	Remarks
Social Networks													
DTE	SL	Liu et al. [2008]	OC	NBC or SVM	User factors and interaction factors	N.A.	Precision: 0.72	N	N	N	Y	N.A.	Using both user features and interaction features for trust prediction is proved better than other methods.
DTE	SL	Zolfaghar and Aghaie [2011]	SWA	MLP	Structure attributes and contextual information	{0, 1}	Predicti on: 0.80	N	N	Y	N	N.A.	Propose a time-based trust prediction and trust network update method.
DTE	SL	Zolfaghar and Aghaie [2012]	SN	DT, SVM, LoR, BN	Trust-inducing factors (knowledge, reputation, relationship, similarity and personality factors)	{trust ed, untr usted}	Accuracy: 0.9377, ROC: 0.97	N	N	N	N	N.A.	Propose a framework of trust inducing factors in social networks and use these factors to predict trust.
DTE	SL	Khadangi and Bagheri [2013]	Facebo ok	MLP, KNN, SVM	Interaction and profile features	Binary class	Accuracy: 0.83	N	N	N	Y	N.A.	Using interaction and profile information to predict the trust of users in Facebook.
DTE	SL	Zhao and Pan [2014]	SN	SVM	9 features about users and their relationships	{1,−1}	P: 0.83, R: 0.97, F: 0.89	N	N	N	N	N.A.	Propose a framework to evaluate trust and set up trust network.
DTE	SL	Wang et al. [2016]	SN	Multila yer-NN	Inducing factors	Binary class	Accuracy: 0.914	N	N	N	N	N.A.	A trust prediction method using inducing factors based on DS theory and NN.
DTE	SL	Wang [2017]	SN	LoR	Traditional trust value and auxiliary information	{0, 1}	Accuracy: 0.90	N	N	N	N	N.A.	Conbine with traditional trust evaluation algorithms to improve accuracy and stability.
DTE	Semi	Papaoiko nomou et al. [2015]	OSN	SNN	User ratings	{positi ve, nega tive}	Accuracy: 0.71-0.96	N	N	N	Y	N.A.	Have a good performance on trust inference.
DTE	SL	Chen et al. [2019b]	SN	BN	User profile, behavior and interaction data	{trust, distrus t}	Accuracy: 0.9516, Recall: 0.9832	N	N	N	Y	N.A.	Use BN to make trust decisions without building trust relationships.

Table 5. Summary and Comparison of Trust Evaluation Methods Using Machine Learning (b)

DTE	SL	Chen et al. [2019a]	SN	LR, ANN	User features	{trusted, untrusted}	Accuracy: 0.996	N	N	N	N	N.A.	Use user features and machine learning to evaluation user trust in social networks.
DTE	RL	Kim and Song [2011]	SN	Q-learning	User relationship and direct trust values	value	N.A.	N	N	Y	N	$O(n+m)$	Use short-term direct relationships to predict long-term indirect relationships and have a good performance.
ATE	UL	Chen et al. [2016]	MSN	Clustering	Contact behavioral and user attributes	tuple	Prediction, Recall and F-score: > 0.8	N	N	N	N	N.A.	Suitable for large-scale MSNs and contain trust aggregation and transfer methods.
ATE	UL	Zhang et al. [2016]	SN	K-means or fuzzy c-means	User rating and relationship	value	N.A.	N	N	N	N	N.A.	Mitigate the sparseness of explicit trust graph and improve the ability to predict trust.
Multi-Agent Systems													
DTE	SL	Liu et al. [2013]	MAS	LDA/DT	Local knowledge (past interactions)	N.A.	FR (false rate): < 0.1; < 0.19	N	Y	N	N	N.A.	Propose a trust architecture for large-scale distributed systems.
DTE	SL	Nguyen and Bai [2018]	MAS	BN	Contextual data (environmental factors, targets and temporal factors)	{0,1}	Accuracy: <=1.0	N	N	N	N	N.A.	Dynamically evaluate the trust of the agent group using Bayesian network.
DTE	RL	Aref and Tran [2015], Aref and Tran [2017]	MAS	Q-learning	N.A.	Four categories	N.A.	N	Y	Y	N.A.	N.A.	Conbine fuzzy logic and Q-learning and enhance accuracy.
DTE	SL	Zhou et al. [2015]	MAS	Deep belief network and RBM	N.A.	N.A.	N.A.	N	Y	N	N.A.	$O(n)$	Can defend against context-correlated attacks and be accurate.
Service Environments													
DTE	SL	Ma et al. [2009]	Epinion	SVM	19 Users & their interactions	{trust, non-trust}	F-score: 0.67-0.81, 0.73-0.85	N	N	N	Y	N.A.	Choose unusual features and have a good performance with sparse trust relationship.

Table 5. Summary and Comparison of Trust Evaluation Methods Using Machine Learning (c)

DTE	SL	Mohanty et al. [2010]	Service-oriented environment	BPNN, BNN, J48, SVM, TreeNet	Quality of service attributes	{trusted, untrusted}	Accuracy: 0.9972	N	N	N	N	N.A.	Use multiple quality of service attributes and machine learning algorithms to evaluate trust of web services.
DTE	SL	Korovaiko and Thomo [2013]	OC	RT or SVM	From user-similarity and rater-reviewer interactions	{trust, distrust}	Prediction: 0.8 (RT), 0.73 (SVM)	N	N	N	N	N.A.	Choose features that make trust prediction performance improved.
DTE	SL	Olteanu et al. [2013] Wawer et al. [2014]	Web pages	Classification	Contents, social and (GI) features	Binary class	Accuracy: 0.75	N	N	N	N	N.A.	Predict trust for webpages.
DTE	SL	Mao et al. [2017]	SeOE	POS-NN	QoS attributes	{1,−1}	Precision: 0.8828	N	N	N	N	N.A.	Use neural network to evaluate trust, and PSO is used to optimize neural network initial settings.
DTE	Semi	Jayasinghe et al. [2019]	IoT Services	k-means and SVM	Knowledge information (relationship, spatial, credibility and temporal)	{0,1}	Recall: 1.0 TP: 0.9813	N	N	N	N	N.A.	Combine clustering and classification methods for trust evaluation of services in the IoT.
DTE	SL	Wu [2010]	C2C E-Commerce	FNN	Technical, merchant and customer factors	Four categories	N.A.	N	N	N	Y	N.A.	Reflect the ambiguity and uncertainty in trust evalutation.
DTE	SL	Yahyaoui and Zhioua [2013], Yahyaoui and Al-Mutairi [2016]	SeOE	HMM/PSA	Derived by calculating entropy/8 self-defined attributes	5 or 11 categories	F: > 0.75 /F: > 0.82	N	N	N	N	$O(n^2)$	Use their own defined trust model for trust evaluation on services.
DTE	SL	Mashinchi et al. [2011]	SeOE	Fuzzy LiR	QoS attributes	Fuzzy value	Accuracy: 0.8	N	N	N	N	N.A.	Add fuzzification concept and facilitate the portrayal of the uncertainty of trust.
DTE	SL	Mao and Lin [2016]	SeOE	PSO-NN	QoS attributes	value	Percision: 0.87	N	N	N	N	N.A.	A well-performing method by using PSO-NN.
Ad Hoc Networks													
DTE	SL	Imana et al. [2010]	MANET	RBF-NN	Ten novel attributes	[0,1]	Accuracy: 0.9869	N	N	Y	N	N.A.	A high accuracy method for predicting reputation values by using RBF-NN.

Table 5. Summary and Comparison of Trust Evaluation Methods Using Machine Learning (d)

DTE	SL	Trofimova et al. [2017]	Ad hoc networks	MLP	PDRs	{trusted, distrusted}	Accuracy: 0.98	N	N	N	N	N.A.	The method of generating training data is given.
DTE	Semi	Han et al. [2019]	UASN	K-means and SVM	Communication trust, packet trust, and energy trust	{0,1}	Accuracy: 0.97	N	N	N	N	N.A.	Use unsupervised learning and supervised learning to assess trust in UASN.
DTE	SL	El-Sayed et al. [2020]	VN	DT	Vehicle interaction information	{trusted, untrusted, uncertain}	Precision: 0.9234, Recall: 0.9164	N	Y	Y	N	N.A.	Propose an entity-centric trust evaluation scheme for the nodes in vehicle networks.
Other Scenarios													
DTE	SL	Yuan et al. [2006]	PeCE	NBC	*A priori* knowledge and recommendation information	{0,1}	N.A.	N	N	Y	Y	N.A.	Propose a dynamic trust decision model that can be used in pervasive environments.
DTE	SL	Huang and Chen [2019]	Crowdsourcing	RF, DT, BPNN, SVM	Factors from initial reputation dimension, evaluation dimension, transaction dimension and punishment dimension	{1,−1}	Accuracy: 0.928	N	N	N	N	N.A.	Estabilsh a multi-dimensional reputation indicator system and propose a reputation evaluation method using machine learning.
DTE	SL	D'Angelo et al. [2017]	PeCE	NBC	Extracted by Apriori algorithm	{1,0,−1}	Accuracy: 0.92	N	Y	N	N	N.A.	The trust evaluation method can identify three basic attacks.
DTE	SL	Huang et al. [2005]	P2P	BP-NN	Transaction result sequences	value	N.A.	N	N	Y	N	$O(n)$	Use neural network in a P2P environment to derive the local trust value.
DTE	SL	Song [2005]	Online Scenarios	ANN	Trust opinions	value	Accuracy: 0.938	N	N	Y	Y	N.A.	Use ANN to make trust decisions based on heterogeneous recommendations.
				HMM	The expertise and trustworthiness of recommendations	value	Accuracy: >= 0.9	N	N	N	N	N.A.	Use HMM to evaluate the recommender's reputation based on the expertise and trustworthiness of recommendations.
				ANN	Local trust evaluations	value	Accuracy: 0.944	N	N	N	Y	N.A.	Use ANN to evaluate global trust in large distributed systems.

Table 5. Summary and Comparison of Trust Evaluation Methods Using Machine Learning (e)

ATE	UL	Ni and Luo [2008]	GC	Hierar chical cluste ring	Local security policy and reputation	value	N.A.	N	N	N	N	N.A.	Suitable for large-scale and dynamic-grid computing.
Common Frameworks													
DTE	SL	López and Maag [2015]	All	SVM	N.A.	{1,0,-1}	Accuracy: 0.96	N	N	N	N.A.	N.A.	Propose a universal trust management framework for trust assessment.
DTE	SL	Hauke et al. [2013]	N.A.	RF	N.A.	value	Accuracy: 0.59	N	N	N	N.A.	N.A.	The method can map estimator output to a belief logic representation.

Y: Yes; N: No; N.A.: Not Available; ML: Machine Learning; Ref: Reference; Sc: Scene; Al: Algorithm; SF: Selected Features; TR: Trust Representation; Ef: Effectiveness; PP: Privacy Protection; Rb: Robustness; Ca: Context-awareness; Su: Subjectivity; CO: Computational Overhead; RL: Reinforcement Learning; Semi: Semi-Supervised Learning; SL: Supervised Learning; UL: Unsupervised Learning; (O/M)SN: (Online/Mobile) Social Network; GC: Grid Computing; MANET: Mobile Ad Hoc Networks; SeOE: Services-oriented Environment; PeCE: Pervasive Computing Environment; SWA: Social Web Applications; OC: Online Communities; UASN: Underwater Acoustic Sensor Networks; IoT: Internet of Things; FLR: Fuzzy Linear Regression; FNN: Fuzzy Neural Network; LSOS: Large-Scale Open System; MAS: Multi-Agent Systems; DT: Decision Tree; BN: Bayesian Net-work; PSO: Particle Swarm Optimization; VN: Vehicle Network; HMM: Hidden Markov Model; DTE: Direct Trust Evaluation; ATE: Assisting Trust Evaluation.

Zolfaghar and Aghaie [2012] proposed a framework using trust-inducing factors to address the sparseness of trust networks, and mapped qualitative factors into measurable features for trust prediction. The entire prediction process is divided into five steps. First, the trust assessment problem is analyzed and mapped into a data mining problem. Second, the initial data is collected. The data is divided into structural data and contextual data. Third, the data is preprocessed. The obtained data are reputation factor, knowledge factor, similarity factor, propensity factor, and relationship factor. Fourth, machine learning is used for modeling. Fifth, the learned model is evaluated. The real data from Epinions were used in the experiment, and the comparison is made by decision tree, SVM, logistic regression, Bayesian network, and neural network, respectively. The experimental results show that the prediction results using decision tree are the best, with an accuracy of 0.9377 and a ROC of 0.97. The applied factors in the trust evaluation are not subjective data. The method does not consider privacy protection and defense against attacks. It is not a context-aware method. Nor does it investigate the computational overhead.

Khadangi and Bagheri [2013] proposed a method for evaluating the trust relationship between users in Facebook. They used the Facebook application to collect data. Then, on the basis of the Pearson correlation coefficient between features and trust, they selected features from user interaction information and profile information. After that, they chose KNN, SVM, and MLP to train a model and predict trust using the features they selected. They compared experimental results and found that all of them are effective and MLP achieves the highest accuracy. The selected features are obtained from interactive information, which is subjective. In the process of data collection, there is no protection on the users' sensitive and private data, and no resistance on any possible malicious user attacks. Context-awareness is not supported in the method. Computational overhead was not discussed in this work.

Zhao and Pan [2014] presented a framework of trust assessment by applying machine learning in the context of social networks, which consists of five parts, namely, collecting data, extract-

ing features, training model, predicting trust, and building trust networks. It uses the data collected from social networks with the value of trust to extract features and establish a model to perform trust prediction. Then, a trust network can be established based on trust relationships between users. They select nine features about users and their relationship. They chose SVM to train model and divided trust relationship into trust and distrust, respectively represented by 1 and −1. Weibo's[1] data were used in experiments. After studying the experimental results, we found that the framework has a good ability to predict trust. Selected features are based on statistics. They do not contain subjective information. The algorithm can be changed, but not context-aware. Meanwhile, the framework does not consider privacy protection although having the ability to resist some attacks. In addition, there is no consideration on computational overhead.

Wang et al. [2016] proposed a method of trust prediction by applying multilayer neural network and Dempster-Shafer theory. In terms of feature selection, they divided the induction factors of distrust and trust into three categories (they are homophily, social status, and emotion tendency) and defined their specific representations. Prediction architecture is comprised by an input layer, a fusing layer, and a decision layer. It can be divided into five units: input unit, evidence processing unit, mass combining unit, fusing unit, and decision unit. They mapped the input set represented by the inducing factors to evidence prototypes, built a mass function based on it, and then used multilayer NN to fuse local factors and global factors, respectively, for trust evaluation. Experiments were performed using Epinion data. From the experimental results, we can see that the accuracy of this method is 0.914. The features selected by the program are mainly for social networks and can be used in social-network-related applications. Since the features applied are not subjective, thus this method cannot support the subjective of trust. The method is neither adaptive nor context-aware. It does not protect privacy and cannot defend against attacks. Computational overhead was not considered, either.

Wang [2017] proposed a method for calculating trust values in social network scenarios by applying machine learning. They used the trust values calculated by a traditional method and some additional information as training features, and represented the users' social relationship with a graphical structure. Nodes represent users and edges represent the users' relationships. After obtaining a labeled training set, the method uses a logistic regression method to model. The main task of determining the logistic regression model is also to solve function parameters, which can be obtained by minimizing a loss (cost) function that is used to measure fitting degree. The data of Tencent[2] Weibo were used in experiments. The experimental results showed that the method using machine learning has higher accuracy by comparing it with other traditional trust evaluation methods. But this method does not take into account the subjectivity of trust and context awareness. Meanwhile, the privacy protection of data was not considered. Nor can it defend against possible attacks. Also, there is no analysis on the computational overhead of this method.

Papaoikonomou et al. [2015] presented a method for predicting the trust level between two users by using user ratings on items with a semi-supervised learning way. They applied a signed social graph to represent a social network. The sign on the edge of the graph is composed of positive sign and negative sign, which indicate that the relationship between users is trustworthy or untrustworthy. The problem of trust prediction is converted to the problem of edge sign prediction and the solution is consist of two steps. The first step is the user encoding stage. They utilized the Restricted Boltzmann Machine to generate a binary low-dimensional code for each user by using the user's rating information. The second step is the sign prediction phase. They used autoencoder networks to classify the binary codes obtained in the previous step. Autoencoder conducts

[1] weibo.com.
[2] t.qq.com.

unsupervised learning of all the data obtained. After that, a supervised neural network performs a classification task. The experiment, which used datasets containing 130K users from Epinions, sets up different autoencoder architectures and labeled datasets of different data sizes by using accuracy measurement as an evaluation indicator. The results of experiment indicate that this method performs well and the accuracy value can be up to 0.9592. User rating information used in this method is subjective. But the method is not context-aware and does not consider privacy protection and defense against attacks. They did not analyze computational overhead.

Chen et al. [2019b] proposed a trust model for social networks. The model uses Bayesian networks to make decision on trust. It selects user-related data such as user profile information, behavior information, and interaction data as user features, and uses Bayesian networks to classify users into trust and distrust. They obtained datasets from Facebook and Twitter to build models, and selected different features from different datasets. From the Facebook dataset, they selected 11 user features, and selected 12 user features from the Twitter dataset. The accuracy of the model tested by the Facebook dataset is 0.8717 with a recall rate 0.8421. The accuracy of the model tested by the Twitter dataset is 0.9516 with a recall rate 0.9832. The user features used in this model contain user preference information, which are subjective. However, the model is not context-aware. It cannot defend against attacks on machine learning and protect privacy. There is no discussion on computational cost.

Chen et al. [2019a] presented a trust evaluation framework for online social networks. This framework is based on machine learning and applies user features to make a trust decision. They divided the user features into four categories and designed a feature selection method to select optimal features. After that, they used machine learning to build a model in order to judge whether a user is trustworthy or not. They used the user data in Twitter to test the trust model. Eight machine learning methods were tested. Results showed that the performance of trust evaluation using logistic regression and neural network was the best, with an accuracy rate up to 0.996. The feature data adopted are user behavior, relationship and other related data, which are not subjective. The framework does not support context awareness. Meanwhile, it does not consider privacy protection of user data and cannot resist common attacks. The computational overhead of the framework was not analyzed.

3.1.1.2 Models with Continuous Numeral Ratings. Kim and Song [2011] gave a trust evaluation method in social networks based on trust propagation and reinforcement learning, which can use short-range direct relationships to predict long-distance indirect relationships. There are two steps to build the model. In the first step, a social network consisting of users and their direct trust values represented by a graph is preprocessed. The second step is to use the Q-learning algorithm to update the two objective functions according to immediate feedback. Q-learning is based on a Markov decision-making process. And an optimal strategy is obtained by iterating and updating a value function that represents the accumulated rewards based on specified states and actions. Then, they derive the trust value based on an expected trust path strength. This method's computational complexity is $O(n + m)$. Experimental results indicate that the best method of trust evaluation is to use the full path for trust estimation in combination with max-min and weighted aggregation. This model can be dynamically adjusted to support context awareness. But it does not reflect the subjective of trust. Meanwhile, it does not consider privacy protection and cannot defend against attacks.

3.1.2 Machine Learning for Assiting Trust Evaluation. Chen et al. [2016] used a clustering algorithm to evaluate trust for Mobile Social Networks (MSNs). They represented the social network as a graph structure and calculated trust in Implicit Social Behavioral Graph (ISBG). First, they proposed a clustering algorithm for community discovery in MSNs to find ISBG. After that, they

set the contacts rank according to the group affiliations of different users. Finally, they calculated the trust values in the group according to contact behaviors and user attributes, and put forward the method of trust aggregation and transfer calculation. In the experimental part, they performed simulations to refine precision, recall, and F-Score to make accuracy metrics most suitable. Experimental results show that this method performs well and can be used to calculate, aggregate and transfer trust values, which are represented as tuples. But this method does not reflect the subjective of trust and lacks support on context-awareness. Meanwhile, it does not provide privacy protection and cannot resist against attacks. There is no discussion on computational overhead.

Zhang et al. [2016] offered a method of trust prediction on the basis of co-cluster to mitigate the sparseness of an explicit trust graph and improve its ability to predict trust. They graphically represented social networks with users and items and utilized the users' rating on items and the relationship between the users to predict trust. They mapped users and items to a shared potential space and used *k*-means or fuzzy *c*-means to find subgroups. Then, they calculated the implicit and explicit similarities of subgroups separately. In each subgroup, the similarities are aggregated. And the predicted results from different subgroups are merged to get a final prediction result. Experimental results show that this method is not pretty good, but it is better than some other trust prediction methods. It is extensible because it works on different datasets. However, this method does not support the subjective of trust and context-aware trust evaluation. Privacy protection and protection against attacks were not considered. Also, the authors did not evaluate computational overhead.

3.2 Trust Evaluation Based on Machine Learning in Multi-Agent Systems

In this part, we review existing trust evaluation methods based on machine learning in multi-agent systems.

3.2.1 Machine Learning for Direct Trust Evaluation.

3.2.1.1 Models with Binary Ratings. Liu et al. [2013] presented a trust architecture using machine learning in a large distributed system, that is composed of a knowledge collector, a trust calculation engine, and a storage component. In the trust calculation engine, they proposed a method for trust assessment using machine learning algorithms. They classified historical transactions into two categories: successful and unsuccessful. Then, linear discriminant analysis or DT was used to train a model by using feature vectors and category labels as training data for evaluating potential transactions. They experimented with real-world data collected from the Internet auction sites Allegro and eBay. The experimental results show the good performance and robustness of the proposed method. However, this work does not consider privacy protection. The features used to train the model are objective information. Meanwhile, context awareness was not considered. There are too many factors that need to be determined for trust evaluation in a specific scenario. In this work, computational overhead was not mentioned.

Nguyen and Bai [2018] presented a scheme for dynamically evaluating the trust of agent group using Bayesian networks. The scheme uses contextual data to evaluate trust. Contextual data includes three kinds of factors: some environment factors, targets (such as relationship, feedback, etc.) and temporal factors. Then, mutual information was used for determining features and trustworthiness was decided by applying a Bayesian network. The authors developed a system with 560 agents to perform testing. In the best case, the accuracy can reach 1.0, indicating that the method is effective. Features used are not subjective. And this method is not context-aware. Mean-

while, it does not support privacy protection and attack defense. The calculation overhead was not specified.

3.2.1.2 Models with Discrete Numeral Ratings. Aref and Tran [2015] presented a method to evaluate trust, which is used in multi-agent systems. They used Q-learning to evaluate trust to enhance the response of a trust model to dynamic changes in the multi-agent systems. Then they [Aref and Tran 2017] proposed another method based on the previous trust evaluation method, which combines fuzzy logic and the Q-learning algorithm. It is conducive to embody the ambiguity and uncertainty of trust. First, the direct trust value is computed using a Q-learning algorithm, and then the direct trust value, the indirect trust value and the average time delay are fuzzified. After that, they used the inference rules in a fuzzy logic system to get a final result. They conducted simulation experiments in different scenarios. From the experimental results, we learn that this method can enhance accuracy and resist attacks. Meanwhile, it supports context-awareness. However, it was not clear whether the method can support subjectivity. And there are no measures for data privacy protection. Neither of the above works discussed computational overhead.

3.2.1.3 Models with Continuous Numeral Ratings. Zhou et al. [2015] designed a Context-Awareness Stereotypical Trust deep learning framework (CAST) for inferring *a priori* trust value in an evidence-sparse context by learning seven context-awareness stereotypes in evidence-dense contexts. They trained an inference model for each of seven stereotypes by using a deep model and employed evidences with trust values for model training. Then they used the value of the smallest output from the seven inference models as the required *a priori* trust value to defend against context-correlation attacks. They conducted simulation-based experiments and used Root-Mean-Square Deviation (RSMD) to measure accuracy. The experimental results indicate that the method is effectivity with high accuracy. It can resist context-correlation attacks and its applicability is sound. They used time complexity to represent the computational overhead of the method, which is related to the amount of layer-wise connections. However, the framework does not describe the selected specific features. It does not support context awareness, privacy protection, and subjectivity since all of them were not considered.

3.3 Trust Evaluation Based on Machine Learning in Service Environments

In this part, we survey trust evaluation methods based on machine learning in service-oriented environments.

3.3.1 Machine Learning for Direct Trust Evaluation.

3.3.1.1 Models with Binary Ratings. Ma et al. [2009] proposed two methods, Personalized Classification Method (PCM) and Cluster-based Classification Method (CCM), to solve the problem of trust prediction when the user's trust relationship is sparse. They treat the trust prediction task as a classification task and use machine learning to complete it. They chose two datasets from Epinions Datasets in order to evaluate and compare the methods. In PCM, they used user pairs presented by 19 user interaction features and trust or distrust labels as a training set. Then, they chose SVM to train classifier to classify user pairs. In CCM, they adopted a divided hierarchical cluster method to cluster users based on the density of connected neighborhood in a trust web. After that, they also used user pair to train classifiers like PCM. The difference is that CCM trains a classifier for each user cluster. In experiments, they used F-Measure to compare the performance of PCM, CCM, a Global Classification Method (GCM), and a trust prediction method based on trust propagation called MoleTrust. The results showed that the selected features are valid. The performance of CCM is between PCM and GCM, which is related to the number of user clusters. And the performance of trust prediction methods based on classification is better than MoleTrust. We can find that PCM and CCM can apply into such an application scenario where online users can com-

ment. Most of the features selected in the method are ratings and reviews, which are subjective. But when the trust evidence is determined, the trained classifier will not change any more, so the method does not support context-awareness. In this article, the authors did not consider privacy protection, attack defense, and computational overhead.

Mohanty et al. [2010] proposed using a series of quality of service attributes and multiple machine learning algorithms to evaluate whether the web service is trustworthy. The attributes used include price, availability, reputation, throughput, reliability, security, response time, accuracy, latency, integrity, regulatory, accessibility, and robustness. The evaluation process of the model is divided into three steps: feature selection, classification model training, and rule generation. This method was evaluated by using classification algorithms such as Back Propagation Neural Network (BPNN), Probabilistic Neural Network (PNN), Group Method of Data Handling (GMDH), Classification and Regression Trees (CART), ID3 decision tree (J48), TreeNet, and SVM, working on a dataset gained from Web Service Crawler Engine (WSCE). Testing results show that a highest accuracy rate can be reached by using J48 or TreeNet, which is 0.9972. The most important attributes were found as throughput, response, reliability, documentation and capacity to succeed. The method is suitable for trust evaluation on web services, but it cannot support context awareness and subjectivity. It does not have the ability to protect privacy and prevent attacks. Meanwhile, the authors did not provide computational overhead.

Korovaiko and Thomo [2013] proposed a method in online communities for predicting users trust relationships. They used a classification algorithm to solve the trust prediction problem. Because if there are a lot of similarities between users, it will be easy for them to trust each other. In addition, if a user makes some high-quality comments on multiple different products, the user will be treated as trustworthy. So, they extracted five features from user-similarity interactions and extracted three features from rater reviewer interactions. Then, they quantified the eight features and used the Kolmogorov-Smirnov test to sort them. They selected SVM and RF as classifiers. A comparative experiment was conducted on the Epinion online community to verify whether the method using selected features is more effective than existing methods. The results indicated that the performance of trust prediction using the features in this method is 5–20% higher than that of the previous ones. The method achieves a good performance, but the used features are not subjective. Meanwhile, context awareness and computational overhead were not discussed. There is no privacy protection on data and no consideration on attack resistance.

Olteanu et al. [2013] presented a scheme for predicting the trust of a webpage. They selected 22 features from content features and social features. Then, they selected regression and two-category methods for learning model and prediction. They experimented with a dataset of 1000 URLs with trust levels established by Microsoft and chose supervised learning algorithms such as SVM, extremely randomized trees (ERT) for regression or SVM, DT, ERT, and NBC for classification. Its accuracy is around 0.75 and it has a good performance. On the basis of this method, Wawer et al. [2014] proposed adding General Inquirer features into feature selection, which can improve predictive performance. The features used in these two methods are not subjective and context-aware. Both of them do not consider privacy protection and attack resistance.

Mao et al. [2017] used neural network to determine the non-linear relationship between the quality of service attributes and the trustworthiness of a service in order to evaluate the trust of cloud services. Meanwhile, the method uses particle swarm optimization to optimize the initial settings of neural network and improves the evaluation accuracy. The method consists of three stages. In the preparation stage, the main task is to analyze the problem, determine the structure of the neural network, and the input and output of a prediction model. Then, in optimization stage, PSO is used to optimize the initial settings of the neural network, and an optimal model is determined by continuously testing the model with training data. Finally, in prediction phase, the

model obtained by training is used to predict the trust degree of cloud services. Experiments were conducted by using the public dataset QWS to compare different algorithms, such as Bayesian network, CART, J48, SVM, FLRA and BPNN. The experimental results show that the method can reach prediction precision as high as 0.8828. The attributes used for trust prediction are not subjective. The method is suitable for service-oriented scenarios without context awareness. Meanwhile, it cannot protect privacy and does not have the ability to defend against attacks. The computational overhead was not presented in this work.

Jayasinghe et al. [2019] used machine learning for trust assessment of IoT services. The reason for using machine learning is that in evaluation process, the influence of trust attributes on the trust value is expressed by weights, but determining an appropriate weight is a complex task. Therefore, intelligent methods are needed to determine import attributes and trust boundary. This method uses some knowledge information, such as relationship, spatial, credibility, and temporal as trust attributes. Due to the scarcity of labeled data, unsupervised learning was applied to classify data, and then the data is used for supervised learning to obtain an evaluation model. The unsupervised learning and supervised learning methods are *k*-means and SVM. A simulation experiment was performed by using the dataset from CRAWDAD. The experimental results show that the recall rate is 1.0 and the true positive rate is 0.9813. The trust attributes are not subjective. The method is not context-aware. Privacy protection and attack defense were not considered in this work. Computational overhead was not discussed.

3.3.1.2 Models with Discrete Numeral Ratings. Wu [2010] proposed a trust evaluation model for C2C online transactions. Because of many ambiguities and uncertainties in the evaluation process, it uses fuzzy neural networks for modeling. Two parts make up the process, first the input data is blurred, and then the fuzzy data is applied for a neural network formation. The method uses fuzzy sets to represent the ambiguity aspect. There are nine factors used in the method and they are divided into three categories such as technical factors, merchant factors, and customer factors. And some factors are subjective such as consumer psychology. But the accuracy of the method cannot be confirmed. Meanwhile, it did not consider privacy protection of data and resistance to possible attacks. The method does not support context-awareness. The authors did not evaluate computational overhead.

Yahyaoui and Zhioua [2013] proposed a method to evaluate the trust sequence of observed web services based on self-defined trust patterns by using Hidden Markov Model (HMM). They first defined five categories of trust patterns. Then, they selected attributes according to entropy values to represent trust observations and trained HMM based on trust patterns. Finally, they used the trained HMM to match the observed trust sequence with given trust patterns to determine the trust category of the observed web service. Experiments were conducted with the real-world dataset Quality Web Services. Precision was applied as a performance measure. The results show that the precision of the method is about 0.8. The complexity of associated computation is $O(n^2)$. However, HMM suffers from uncertainty in probability distributions. They also presented future work, such as extending the trust patterns. Later on, Yahyaoui and Al-Mutairi [2016] proposed an improved method. They defined 11 trust patterns classes. In terms of pattern matching, the HMM was abandoned and a rule-based Prefix-Suffix Algorithm was proposed. They also defined eight attributes to describe the trust sequence and presented merge rules. Compared with the original method [Yahyaoui and Zhioua 2013], the performance of the improved method is better. Its accuracy is about 0.95. However, both of the above methods do not consider the threat of potential

attacks. Meanwhile, the data selected for trust evaluation are not subjective, and the methods do not support context awareness.

3.3.1.3 Models with Continuous Numeral Ratings. Mashinchi et al. [2011] proposed a trust evaluation method for web services. It was designed on the basis of fuzzy linear regression. They selected quality of service attributes as features. Then, they used a training dataset to establish a fuzzy linear regression model to express the functional relationship between the QoS values and a delivered service trust value. The input is numeric data. The output is fuzzy data in the model. To make the output more intuitive, they defuzzied the output data and classified the data into defined categories. They used fetching data from the Web to conduct experiments. The results showed that compared with SVM, DT, and LiR, the method is more accurate. The method is suitable for being applied into Service-Oriented Computing (SOC). It is also applicable to other scenarios since QoS attributes are not fixed and can be adaptively changed in this method. Since the features used to train the model do not contain subjectively conscious information, the method is not subjective. And there is no privacy protection for sensitive data, no consideration on possible attacks and no support on context-awareness. Computational overhead was not discussed.

Mao and Lin [2016] presented a method for evaluating the trust of network services using neural network. They selected Quality of Service (QoS) attributes as trust features. The training set is composed of QoS values and trust values. Then, they determined the appropriate initialization parameters for the NN by using a Particle Swam Optimization (PSO) algorithm. After that, they selected some common strategies such as back propagation to further train NN to obtain the model for predicting service trust values. They experimented on a public QoS dataset QWS and found that the method performs well compared to some other methods. QoS attributes are targeted to a service-based environment, thus they are not subjective. Meanwhile, context awareness was not considered. It did not consider privacy protection of data and resistance to possible attacks. There is no analysis of computational overhead.

3.4 Trust Evaluation Based on Machine Learning in Ad-hoc Networks

We survey the literature on trust evaluation using machine learning in ad-hoc networks in this subsection.

3.4.1 Machine Learning for Direct Trust Evaluation.

3.4.1.1 Models with Binary Ratings. Imana et al. [2010] offered a method, which uses Radial Basis Function Neural Networks (RBS-NN), to predict the trust of the nodes in Mobile Ad Hoc Networks (MANETs). They chose ten attributes to model the nodes. Training datasets consist of ten attributes representing the nodes and the reputation values of the nodes. They used RBS-NN to learn training datasets to derive the mapping of node attributes and node reputation values. After that, it is regarded as a predictor for trust evaluation of the nodes with unknown values. To predict the credit value of the current state of a node, they used the data from multiple time points in the previous period of time to train the model to obtain its credit value. Simulation based experiments showed that the method performs well and its accuracy is about 0.98. Notably, the ten attributes used in the method are not subjective. The assessment results are influenced by time, so the program has very good dynamic adaptability and support context awareness. But it does not consider the protection of sensitive data and the defense against possible attacks. Also, they did not evaluate computational overhead.

Trofimova et al. [2017] presented a method for trust assessment by using a neural network in ad hoc networks. They determined whether a node is trustworthy based on Packet Delivery Ratio (PDR) of the node and set a threshold to distinguish them. After comparing multiple neural network models, they chose Multilayer Perceptron (MLP) and back propagation algorithms for trust

evaluation. Under specific network arrangement, they randomly assigned PDRs to intermediate nodes and obtain training data through some calculations. The simulation results indicate that the scheme is effective. However, it does not provide a specific implementation method for training data. Meanwhile, once there is a change in the network, the topology used to generate the training data should be reworked. Therefore, the dynamic adaptability of this method is poor. That is, context awareness was not considered. The training data are not subjective, thus cannot reflect subjectivity of trust in assessment. Privacy protection for data was not taken into account in this work. This method does not consider resistance to possible attacks in trust evaluation. Computational overhead was not analyzed.

Han et al. [2019] proposed to use machine learning to assess the trust of sensor nodes in underwater acoustic sensor networks. In this scenario, there are sparse distribution of underwater nodes, weak communication signals, high delay, narrow loan, and other characteristics, which make trust evaluation difficult. Thus, machine learning is selected to overcome these problems and evaluate trust. The proposed method contains two phases and uses communication trust, packet trust, and energy trust as features. The first phase is an unsupervised learning phase. After obtaining the feature data, the data are divided into two clusters by using a *k*-means algorithm. The purpose is to label the data. The process of *k*-means is roughly as follows. First, determine the total of categories, i.e., the value of K that is 2 therein. Then, select two data points randomly from the dataset as initial centroids. After that, calculate the similarity between each data point and each centroid and put the data points into the cluster where the most similar centroid is located. Then, re-determine the centroids and repeat the above steps, the process iterates until the centroids no longer change. The second phase is a supervised learning phase, which uses the labeled data obtained in the first phase and selects the SVM algorithm to train the model for trust evaluation. Simulation-based experiments were performed. The experimental results show that the accuracy of the method can reach 0.97. However, the method cannot protect privacy or resist attacks. It does not consider computational overhead and is not context-aware. The characteristics used for trust evaluation are not subjective.

3.4.1.2 Models with discrete numeral ratings. El-Sayed et al. [2020] proposed an entity-centric trust model for nodes in a vehicle network. The model uses direct experiences and recommendations to calculate trust values based on vehicle interaction time, distance and other indicators. The trust value and a DT model are used to build decision rules, and a decision is divided into three types according to the trust value: trustworthy, untrustworthy, and uncertain. For some trust values with errors, the scheme chooses an ANN model to adaptively adjust. The model was tested by applying it into a vehicle network that contains 5 road slide units and 30 vehicle nodes with a transmission range of 300 meters. Experimental results show that the model has good performance and robustness, its precision rate can reach 0.9234 with a recall rate 0.9164. The attributes used to calculate the trust value are not subjective. DT and ANN were used in trust modeling and evaluation. When context changes in real time, the model can be adjusted adaptively. Therefore, this model supports context awareness. However, privacy protection was not considered.

3.5 Trust Evaluation Based on Machine Learning in Other Scenarios

In this part, we review the literature on trust evaluation based on machine learning in other scenarios, such as pervasive computing, grid computing, and peer-to-peer networking.

3.5.1 Machine Learning for Direct Trust Evaluation.

3.5.1.1 Models with Binary Ratings. Yuan et al. [2006] proposed a model for dynamic trust decisions in a pervasive environment. The factors used in the trust model are prior probability, trust level, past interaction history, time influence, and peer recommendation. An NBC algorithm was

selected for trust decision and used twice. When using NBC for the first decision, only prior knowledge of service provider was used. If a decision result cannot be obtained, then recommendation information is added to derive a final decision. Subjectivity of this method is well supported. Simulation experiments demonstrated its dynamic decision-making ability, but did not verify its effectiveness. This model can dynamically evaluate trust with context-awareness. However, data privacy protection and defense against possible attacks were not considered. Computational overhead was not evaluated, either.

Huang and Chen [2019] proposed a model for evaluating the reputation of crowdsourcing participants using a random forest based on linear discriminant analysis. The scheme is divided into five steps. First, a multi-dimensional reputation index system is established, and data is collected and preprocessed. The second step is to reduce the data dimension, and the third step is to standardize the data. The fourth step is to select a subset of features and use machine learning for modeling. The fifth step is to verify the validity of the model. By using the data from the zbj.com platform, multiple experiments were conducted, where different dimensionality reduction methods and different machine learning algorithms were combined for reputation evaluation. The final results show that using the combination of linear discriminant analysis and random forest makes accuracy up to 0.928. The data indicators used for reputation evaluation come from four dimensions: initial reputation dimension, penalty dimension, evaluation dimension, and transaction dimension. None of them are subjective. The method is not context-aware. Defense attacks and privacy protections were not considered, either. There is no explanation of computational overhead in this work.

3.5.1.2 Models with Discrete numeral ratings. D'Angelo et al. [2017] proposed an *a priori* algorithm and an NBC-based trust model in pervasive computing. First, they represented the interactions between entities with the tuples of the nine properties they defined. According to the trust score that is one of the nine properties, the interactions are divided into three categories: trustworthy, dubious, and untrustworthy. After that, they used the *a priori* algorithm to extract features. Finally, they used the NBC algorithm to determine the trust category for the data to be evaluated. Simulation experimental results indicate that this method can identify counting-based, time-based, and context-based attacks with high accuracy. It is robust to some extent. But it does not consider the protection of data obtained from other devices. The selected attributes are not subjective. This method does not support context awareness. Computational overhead was not discussed.

3.5.1.3 Models with Continuous Numeral Ratings. Huang et al. [2005] proposed a method for calculating a peer's local trust value using Back Propagation Neural Network (BP-NN) in Peer-to-Peer (P2P) networking. They processed the transaction result sequences of peers to get constrictive transaction result sequences and used them with their global trust values as training data to decide the quantity of input, hidden layers, and nodes of each layer, and some other parameters of the neural network model. After processing the obtained transaction result sequence to get a constrictive transaction result sequence, they put it as an input into the trained neural network model to derive the corresponding local trust value. On the basis of JXTA (Juxtapose, an open source P2P protocol), they utilized a P2P data backup system for experiments, but the results did not show that the method is effective. They used time complexity to represent computational overhead, which is $O(n)$. The data used for training and testing are not subjective but time-sensitive, so the method has good dynamic adaptability and context awareness. However, the method does not consider the protection of sensitive data and the prevention of possible attacks.

Song [2005] proposed three trust evaluation models. All of them are suitable for online scenarios. The first model makes trust decisions on unknown parties based on heterogeneous recommendations. The recommendation trust network model was trained by ANN by using the trust opinions

of qualified recommenders and requesters of information providers. A back-propagation algorithm was used to train the model. Experiments were performed by using simulation results of movie file sharing in a P2P network with 50 agents. The experimental result shows that the accuracy of the model can reach 0.938. The model is adaptable and can deal with changes in trust behaviors. The model uses trust opinions as features. It is subjective. However, it does not take into account attack defense and privacy preservation, nor does it analyze computational overhead. The second model uses HMM to evaluate a recommender's reputation based on the expertise and trustworthiness of recommendations. The performance of this model was evaluated through simulation experiments. The results show that the model has good performance and its accuracy is at least 0.9. However, it is not context-aware, subjective, nor can it resist attacks or protect privacy. Meanwhile, its computational cost was not analyzed. The third model was proposed to manage the global reputation of users in a distributed system. It uses the back-propagation algorithm to evaluate local agent trust and builds neural networks to obtain global reputation based on local trust. The performance of the model was evaluated by building a simulation experiment with 10 users, 4 reputation communities, and 2,000 transactions. The result shows that its accuracy can reach up to 0.944. This model is context-aware. But it cannot resist attacks and protect privacy. It is not subjective, either. Notably, its computational overhead was not evaluated.

3.5.2 Machine Learning for Assisting Trust Evaluation. Ni and Luo [2008] used hierarchical clustering algorithm to evaluate trust of grid entities in grid computing virtual organization. They cluster grid entities into four categories by using local security policy and reputation as the characteristics of the grid entities, which are respectively very trustworthy, trustworthy, untrustworthy, and absolutely untrustworthy. When trust evaluation is performed on grid entities in different virtual organizations, they obtained trust value by considering the security level of trust relationship. This method is aimed at the trust assessment of virtual organizations in grid computing. Subjectivity and context-awareness were not supported. Privacy protection and attack resistance were not considered in the assessment of trust, either. The authors did not analyze computational overhead.

3.6 Common Trust Evaluation Frameworks Based on Machine Learning

In this section, we review some common trust evaluation framework based on supervised machine learning.

3.6.1 Machine Learning for Direct Trust Evaluation.

3.6.1.1 Models with Binary Ratings. López and Maag [2015] put forward a universal trust management framework and proposed a corresponding trust assessment scheme. They had unified mathematical definitions of trust features, context, trustor, trustee, and so on. They used a machine learning method to convert the trust evaluation problem into a multi-classification problem. The trust relationship is divided into untrustworthy, neutrally trusted, and trustworthy, and is respectively represented as −1, 0, 1. Considering that the data may be linearly inseparable, they chose an SVM algorithm with radial basis function kernel and presented a new algorithm for determining the optimal parameters in SVM. They utilized trust data containing two trust features to conduct a simulation experiment and their achieved accuracy is up to 0.96. The selected features were not described in the framework, so it is not possible to determine whether they are subjective. At the same time, it uses supervised learning to model. When the context changes in real time, it cannot adjust adaptively. So, this scheme does not support context awareness. Moreover, privacy protec-

tion was not considered, no defense against attacks, nor investigation on computational overhead was conducted.

3.6.1.2 Models with Continuous Numeral Ratings. Hauke et al. [2013] proposed the requirements of trust assessment using supervised learning methods. First, in training phase, trust assessment requested a training dataset represented by features and labels. During evaluation phase, they used the training dataset and selected RF, DT, and KNN algorithms to perform trust value calculation. After that, they derived a discrete trust value. However, trust represents a subjective probability, so they convert the obtained value into a probability form to better represent the trust relationship. In addition, they also proposed ways to map the evaluation results to a Certain Logic opinion space. They used the data in a hotel booking website to conduct experiments. The results showed that the method is effective in prediction. But some details were not disclosed, such as which features were selected. There is no comment on privacy protection of sensitive data and on how to deal with possible attacks. But the universality of this method is broad. Context-awareness and computational overhead were not examined.

3.7 Discussion

In Table 4, we compare the existing methods of trust evaluation using machine learning based on the following criteria:

—Privacy protection:
-Yes: The work considers or takes measures to protect the collected trust evidence.
-No: The work does not consider or take measures to protect the collected trust evidence.
—Robustness:
-Yes: The method can withstand harsh environments, i.e., it can work normally in the presence of some of the previously discussed attacks.
-No: The method cannot withstand harsh environments, i.e., it cannot work normally in the presence of some of the attacks mentioned above.
—Context-awareness:
-Yes: The method can perceive the change of context or environment and adjust dynamically and adaptively.
-No: The method cannot perceive the change of context or environment and cannot adjust dynamically and adaptively.
—Subjectivity:
-Yes: There are references to the subjectivity of trust, or the use of some subjective information or data as evidence of trust in evaluation.
-No: There are no references to the subjectivity of trust, and no use of some subjective information or data as evidence of trust in evaluation.
—Computational Overhead: The complexity of trust evaluation computation is referred as its computational overhead.

From the above reviews, we summarize the use of machine learning for trust evaluation in different scenarios. In multi-agent systems and social networks, there are four purposes of using machine learning for trust evaluation. The first is the use of other available data to evaluate trust when historical transaction information or historical evidence is scarce and unavailable. The second is to improve the accuracy of evaluation by combining machine learning and other computing models. The third is for handling complex data relationships. The fourth is to achieve accurate evaluation in a big data situation. In service-oriented systems, trust evaluation using machine learning was mainly applied to perform intelligent calculations to improve accuracy in the case of complex

data relationships. In ad-hoc networks, trust evaluation using machine learning mainly use the attributes specific to ad hoc networks to perform trust evaluation in the absence of interactive information in order to ensure accuracy.

Meanwhile, we find that the process of trust evaluation using supervised learning is roughly the same. Existing works treated the trust evaluation as a classification or regression problem and used a labeled dataset to train the model for trust assessment. However, existing works' application scenarios, selected features and the specific algorithms are normally different. For unsupervised learning, we find that because there is no available data label, it cannot directly treat trust evaluation as a classification problem like supervised learning. Unsupervised learning can only cluster similar data according to the attributes or features of data. On this basis, the method based on unsupervised learning selects other methods for the calculation or prediction of trust values. Although using clustering algorithms cannot directly perform trust evaluation, the clustering algorithms can aggregate data with similar attributes, reducing the amount of data of subsequent analysis, making the quality of the data used for trust evaluation higher. In the trust evaluation methods based on semi-supervised learning, when training the model, there is less requirement for the amount of data with labels than the methods based on supervised learning, that is, large amounts of unlabeled data and partially labeled data can be used for model training. In this way, we can use semi-supervised learning for trust evaluation when we know that there are few labeled data. For existing methods using reinforcement learning, they all use Q-learning algorithms. Because reinforcement learning dynamically adjusts the model by interacting with the environment, these methods support context awareness.

In Table 4, we notice that the existing methods of trust evaluation using machine learning mostly use supervised learning, while the other three types of machine learning algorithms are less used. The reason for this phenomenon is that supervised learning is simple to understand and widely used. Meanwhile, there are more studies on supervised learning than the other three. At the same time, the application scenarios of these methods are large-scale distributed networks such as social networks and service-oriented environments. The features selected in the same scenario are similar, e.g., in Ma et al. [2009], Zhao and Pan [2014], Yahyaoui and Zhioua [2013], and Korovaiko and Thomo [2013], the features related to the relationship between users were selected. Notably, most methods simply use a few categories to represent the results of trust evaluation, which is not conducive to reflect the fuzziness and uncertainty of trust. We can find that most methods are effective. In particular, privacy protections on trust evidence are not explored in all methods. Only a few of the existing methods can resist attacks that occur in trust evaluation [D'Angelo et al. 2017; Zhou et al. 2015; Liu et al. 2013; Aref and Tran 2015, 2017]. Among the existing methods, we roughly divide those with contextual awareness into two categories. One is to use supervised learning. The selected training data is related to time series [Huang et al. 2005; Imana et al. 2010; Zolfaghar and Aghaie 2011; Yuan et al. 2006]. The other is to use reinforcement learning, which itself continually optimizes strategies based on real-time feedback from performing operations [Kim and Song 2011; Aref and Tran 2015, 2017]. Also, there are few schemes that reflect the subjectivity of trust [Ma et al. 2009; Khadangi and Bagheri 2013; Yuan et al. 2006; Wu 2010; Papaoiko nomou et al. 2015]. Methods that consider computational overhead are also rare. Most works do not consider this quality attribute, which was only considered in Zhou et al. [2015], Yahyaoui and Zhioua [2013], Yahyaoui and Al-Mutairi [2016], Huang et al. [2005], and Kim and Song [2011].

4 OPEN PROBLEMS AND FUTURE RESEARCH DIRECTIONS

4.1 Open Problems

Based on the above review and the comparative analysis of the literature, we identify some open issues in the field of trust evaluation based on machine learning.

First, a generic solution for trust evaluation based on machine learning is still missing. Machine learning is a powerful artificial intelligence method that continuously improves its performance by reorganizing existing knowledge structures [Han et al. 2019]. The trust evaluation methods based on machine learning show that, in the case of sparse data, complex data relationships, and big data, using machine learning can effectively perform trust evaluation. However, most of the trust evaluation methods based on machine learning are applied into social network scenarios, while still limited in other scenarios. Based on the above review, the literature still lacks a generic method of machine-learning-based trust evaluation that can be applied into various fields. Research is highly expected to investigate the possibility of achieving a common solution.

Second, most of the trust evaluation results obtained through machine learning are coarser-grained. Trust is a subjective concept with uncertainty and should be represented by continuous value [Khadangi and Bagheri 2013]. From the above review, we found that most of the existing methods use techniques of supervised learning for trust evaluation. They treat the trust evaluation problem as a classification problem. Although this kind of methods is simple, effective, and fast, its ultimate representation of trust is divided into two or a limited number of main categories such as trusted and distrusted. Such a coarser-grained representation cannot sufficiently reflect the subjectivity and uncertainty of trust.

Third, how to determine features, algorithms, and algorithm integration in trust evaluation based on machine learning needs to be investigated. We found that the choice of features and algorithms has a significant impact on trust evaluation performance. As studied in Liu et al. [2008], selecting different features or different algorithms has an impact on evaluation performance. Compared with Wawer et al. [2014] and Olteanu et al. [2013], adding some features can improve the performance of evaluation, but at the same time, it may bring a certain amount of computational overhead. Therefore, when using machine learning to conduct trust evaluation, which features to select, which algorithm to choose and how to combine them to make a method effective and efficient and perform best are still worth studying.

Fourth, the method of data labeling is not well explained or explored. When using the supervised learning algorithms for trust evaluation, in addition to selecting appropriate features and algorithms, the acquisition of labeled datasets is also a critical step. Existing trust evaluation methods using supervised learning do not specify how to obtain labeled data for model training. But the quality of the training dataset directly affects the performance of the model. Trust varies in different scenarios. Any training datasets are not universal. In the existing literature, although there is a certain number of researches on the selection of features and machine learning algorithms, few studies explore how to acquire training dataset and how to label data.

Fifth, existing methods using machine learning for trust evaluation generally do not consider privacy protection on the data used for evaluation and seldom concern the robustness of evaluation. The data used for trust evaluation generally contains the private information of related entities. Therefore, in a trust evaluation method, data privacy protection is inevitably essential and should be considered. By reviewing the existing literature, we can see that none of existing methods of trust evaluation based on machine learning consider this issue. At the same time, the possible attacks on trust evaluation that we mentioned in Section 2 should also be considered. But most of the existing works do not prove whether their methods can withstand these attacks. Therefore, problems with regard to security exist in the current works.

Finally, most of the existing methods do not address or ignore computational overhead, thus do not pay special attention to evaluation efficiency. For an algorithmic approach, the analysis of computational and storage overhead is important and should not be overlooked. It is an important item in performance evaluation metrics to judge the merit of a method. Thus, trust evaluation based on machine learning should also consider the cost and complexity of computation and make trade-off with other quality attributes, like precision, accuracy, and fine-grainedness.

4.2 Future Research Directions

Based on our review and the discussion on open research issues, we notice that this research topic is still in its infancy. Many important open issues attract attention and deserve special efforts. Herein, we point out some potential future research directions about trust evaluation based on machine learning.

First, fine-grained trust evaluation based on machine learning with subjectivity and dynamic support should be explored. Under the premise of ensuring effectiveness, we need to study whether traditional trust evaluation methods can be combined with machine-learning-based methods, making the results of trust evaluation granular and relevant to reality. In most of the methods reviewed in this article, the representation of trust does not reflect its uncertainty and subjectivity, as well as time dependency. We believe that later research should make efforts to make machine-learning-based trust evaluation effectively reflect the subjectivity of trust. Meanwhile, trust values change over time [Korovaiko and Thomo 2013]. Time-dependent features should also be considered in learning. And at the same time, the representation of trust should be fine-grained, thus can really reflect the nature of trust. In addition, combining machine learning with other trust-computing methods could be a good way to improve evaluation accuracy by overcoming the shortcomings of each.

Second, automatic learning of feature selection and algorithm selection should be investigated. Current work focuses on specific scenarios by studying which features and algorithms can be selected to make the trust evaluation results accurate. For the features and algorithms used in trust evaluation, the existing literature is case-specific, lacks a common solution. In the subsequent studies, we should investigate whether it is possible to automatically find a suitable combination of features and algorithms based on previous experiences, so as to make the trust evaluation more effective and generic, i.e., applying machine learning into feature selection and algorithm selection.

Third, how to obtain trustworthy labeled data when using supervised learning for trust evaluation is an important issue that is worth studying. In the methods of trust evaluation using supervised learning, the features used for trust evaluation varies in different scenarios with quite different training datasets. For different application scenarios, how to acquire a suitable training set could be the most difficult task in research. Because the training dataset needs to be labeled, how to mark the labels in a trustworthy way is worth studying since it highly impacts the quality of the final trust evaluation results. Subsequent research should put efforts on this aspect. How to refer to the datasets with labels in another application domain and let them benefit a working application domain is also an interesting topic worth studying.

Fourth, the security of trust evaluation is worth special efforts. In our survey, the issue of data privacy protection has not been touched. At the same time, there are only few methods that consider resistance to some possible attacks. Therefore, subsequent research can start with the security and privacy protection of trust evaluation based on machine learning, so that its security and privacy can be guaranteed. The usage of some emerging technologies, such as federal learning and differential privacy to protect the privacy of sensitive data in trust evaluation can be considered.

Fifth, the research to develop a framework for trust evaluation as a service that can be applied into multiple scenarios is highly expected. Future research should explore the general applicability

of the trust evaluation based on machine learning by taking into account computational overhead, communication cost and other practical quality attributes and explore its practicability of deployment in a real world. In this case, how to automatically collect trust evidence, update and optimize the models used for trust evaluation should be considered. Meanwhile, how to refer to the models built up in different domains and integrate them to form a common model is an interesting research topic.

Sixth, we highly suggest attempting different machine learning methods to evaluate trust. From the survey, we note that there are many researchers who explored using supervised learning for trust evaluation. Unsupervised learning has no dependence on labeled data and reinforcement learning has good dynamic adaptability. Future research can make more efforts on applying reinforcement learning, unsupervised learning and semi-supervised learning for trust evaluation since they have their own advantages to support the nature of trust and easy process of trust evaluation.

Last but not the least, it is highly suggested integrating machine learning methods with other emerging technologies to evaluate trust, such as knowledge fragment fusion [Zheng et al. 2019], correlation computation [Yan et al. 2017], fusion pattern learning [He et al. 2018], relation discovery [Chen et al. 2014], and so on. Zheng et al. [2019] provided a significant way of constructing a fragmented knowledge graph. Zheng's method named "knowledge forest" can efficiently represent and compute knowledge, which can be adopted for trust evaluation. Discovering relations of human beings from mass data has also been effectively achieved with the method presented by Chen et al. [2014]. It would be an interesting research topic to perform trust evaluation by intelligently fusing fragmented knowledge about trust, learning fusion pattern and extracting trust relation from mass data with machine learning. Thus, we can solve "zero knowledge" and "cold start" issues in an effective manner and achieve highly accurate trust evaluation based on widely available information in the cyber world.

5 CONCLUSIONS

This article gives a thorough survey on existing trust evaluation methods based on machine learning. We first introduced the basic knowledge and characteristics of trust and machine learning and analyzed the benefits of using machine learning for trust evaluation. Meanwhile, we summarized traditional trust evaluation methods and machine learning algorithms, and compared their advantages and disadvantages. Then, we discussed the requirements that a good trust evaluation method should satisfy in order to figure out the criteria to justify the quality of a trust evaluation method based on machine learning. In particular, we divided the existing trust evaluation methods into a number of categories according to their application scenarios, the functions of machine learning algorithms in trust evaluation and evaluation granularity. By employing the proposed evaluation criteria, we performed a thorough review and comment of each method's advantages and drawbacks. According to the completed review, we found that this research topic is still in the course of initial development. There is a list of critical open issues that should be solved. Finally, we proposed future research directions to attract special efforts and investigation.

REFERENCES

Usama Ahmed, Imran Raza, and Syed Asad Hussain. 2019. Trust evaluation in cross-cloud federation: Survey and requirement analysis. *ACM Comput. Surv.* 52, 1, Article 19 (Feb. 2019), 37 pages. DOI: https://doi.org/10.1145/3292499

Aljawharah Alnasser and Hongjian Sun. 2017. A fuzzy logic trust model for secure routing in smart grid networks. *IEEE Access* 5 (2017), 17896–17903. DOI: https://doi.org/10.1109/ACCESS.2017.2740219

N. S. Altman. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46, 3 (1992), 175–185. DOI: https://doi.org/10.2307/2685209

Abdullah Aref and Thomas Tran. 2017. A hybrid trust model using reinforcement learning and fuzzy logic. *Computational Intelligence* 34, 2 (2017), 515–541. DOI: https://doi.org/10.1111/coin.12155

Abdullah M. Aref and Thomas T. Tran. 2015. A decentralized trustworthiness estimation model for open, multiagent systems (DTMAS). *Journal of Trust Management* 2, 1 (March 2015), 3. DOI: https://doi.org/10.1186/s40493-015-0014-4

Himani Bansal and Shruti Kohli. 2019. Trust evaluation of websites: A comprehensive study. *International Journal of Advanced Intelligence Paradigms* 13, 1-2 (2019), 101–112. DOI: https://doi.org/10.1504/IJAIP.2019.099946

Shuhong Chen, Guojun Wang, and Guojun Jia. 2016. Cluster-group based trusted computing for mobile social networks using implicit social behavioral graph. *Future Generation Computer Systems* 55 (2016), 391–400. DOI: https://doi.org/10.1016/j.future.2014.06.005

Xu Chen, Yuyu Yuan, Lilei Lu, and Jincui Yang. 2019b. A multidimensional trust evaluation framework for online social networks based on machine learning. *IEEE Access* 7 (2019), 175499–175513.

Xu Chen, Yuyu Yuan, and Mehmet A. Orgun. 2019a. Using Bayesian networks with hidden variables for identifying trustworthy users in social networks. *Journal of Information Science* (July 2, 2019), 1–16. DOI: https://doi.org/10.1177/0165551519857590

Yanping Chen, Qinghua Zheng, and Wei Zhang. 2014. Omni-word feature and soft constraint for Chinese relation extraction. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, MD, 572–581. DOI: https://doi.org/10.3115/v1/P14-1054

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20 (1995), 273–297. DOI: https://doi.org/10.1007/BF00994018

Hongjun Dai, Zhiping Jia, and Xiaona Dong. 2008. An entropy-based trust modeling and evaluation for wireless sensor networks. In *Proceedings of the 2008 International Conference on Embedded Software and Systems*. 27–34. DOI: https://doi.org/10.1109/ICESS.2008.31

Gianni D'Angelo, Salvatore Rampone, and Francesco Palmieri. 2017. Developing a trust model for pervasive computing based on a *priori* association rules learning and Bayesian classification. *Soft Computing* 21, 21 (Nov. 2017), 6297–6315. DOI: https://doi.org/10.1007/s00500-016-2183-1

Stuart E. Dreyfus. 1990. Artificial neural networks, back propagation, and the kelley-bryson gradient procedure. *Journal of Guidance, Control, and Dynamics* 13, 5 (1990), 926–928. DOI: https://doi.org/10.2514/3.25422

H. El-Sayed, Alexander H. Ignatious, P. Kulkarni, and S. Bouktif. 2020. Machine learning based trust management framework for vehicular networks. *Vehicular Communications* 25 (2020), 100256. DOI: https://doi.org/10.1016/j.vehcom.2020.100256

Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD'96)*. AAAI Press, 226–231.

Renjian Feng, Xiaofeng Xu, Xiang Zhou, and Jiangwen Wan. 2011. A trust evaluation algorithm for wireless sensor networks based on node behaviors and D-S evidence theory. *Sensors* 11, 2 (2011), 1345–1360. DOI: https://doi.org/10.3390/s110201345

David Freedman. 2005. *Statistical Models: Theory and Practice*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139165495

Jia Guo, Ing-Ray Chen, and Jeffrey J. P. Tsai. 2017. A survey of trust computation models for service management in Internet of Things systems. *Computer Communications* 97 (2017), 1–14. DOI: https://doi.org/10.1016/j.comcom.2016.10.012

Guangjie Han, Yu He, Jinfang Jiang, Ning Wang, Mohsen Guizani, and James Adu Ansere. 2019. A synergetic trust model based on SVM in underwater acoustic sensor networks. *IEEE Transactions on Vehicular Technology* 68, 11 (Nov. 2019), 11239–11247. DOI: https://doi.org/10.1109/TVT.2019.2939179

Fei Hao, Geyong Min, Man Lin, Changqing Luo, and Laurence T. Yang. 2014. MobiFuzzyTrust: An efficient fuzzy trust inference mechanism in mobile social networks. *IEEE Transactions on Parallel and Distributed Systems* 25, 11 (Nov 2014), 2944–2955. DOI: https://doi.org/10.1109/TPDS.2013.309

Sascha Hauke, Sebastian Biedermann, Max Mühlhäuser, and Dominik Heider. 2013. On the application of supervised machine learning to trustworthiness assessment. In *Proceedings of the 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*. 525–534. DOI: https://doi.org/10.1109/TrustCom.2013.5

Daojing He, Chun Chen, Sammy Chan, Jiajun Bu, and Athanasios V. Vasilakos. 2012. A distributed trust evaluation model and its application scenarios for medical sensor networks. *IEEE Transactions on Information Technology in Biomedicine* 16, 6 (Nov. 2012), 1164–1175. DOI: https://doi.org/10.1109/TITB.2012.2199996

Huan He, Qinghua Zheng, and Bo Dong. 2018. VUSphere: Visual analysis of video utilization in online distance education. In *Proceedings of the 2018 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 25–35. DOI: https://doi.org/10.1109/VAST.2018.8802383

Tin Kam Ho. 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 8 (Aug. 1998), 832–844. DOI: https://doi.org/10.1109/34.709601

Baohua Huang, Heping Hu, and Zhengding Lu. 2005. Identifying local trust value with neural network in P2P environment. In *Proceedings of the 2005 1st IEEE and IFIP International Conference in Central Asia on the Internet*. 1503–1505. DOI: https://doi.org/10.1109/CANET.2005.1598194

Yanrong Huang and Min Chen. 2019. Improve reputation evaluation of crowdsourcing participants using multidimensional index and machine learning techniques. *IEEE Access* 7 (2019), 118055–118067. DOI: https://doi.org/10.1109/ACCESS.2019.2933147

Eyosias Y. Imana, Fredric M. Ham, William Allen, and Richard Ford. 2010. Proactive reputation-based defense for MANETs using radial basis function neural networks. In *Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN)*. 1–6. DOI: https://doi.org/10.1109/IJCNN.2010.5596945

Upul Jayasinghe, Gyu Myoung Lee, Tai-Won Um, and Qi Shi. 2019. Machine learning based trust computational model for IoT services. *IEEE Transactions on Sustainable Computing* 4, 1 (Jan. 2019), 39–52. DOI: https://doi.org/10.1109/TSUSC.2018.2839623

Wenjun Jiang, Guojun Wang, and Jie Wu. 2014. Generating trusted graphs for trust evaluation in online social networks. *Future Generation Computer Systems* 31 (Feb. 2014), 48–58. DOI: https://doi.org/10.1016/j.future.2012.06.010

Wenjun Jiang, Jie Wu, Feng Li, Guojun Wang, and Huanyang Zheng. 2016. Trust evaluation in online social networks using generalized network flow. *IEEE Trans. Comput.* 65, 3 (Mar. 2016), 952–963. DOI: https://doi.org/10.1109/TC.2015.2435785

Xuyang Jing, Zheng Yan, and Witold Pedrycz. 2018. Security data collection and data analytics in the Internet: A survey. *IEEE Communications Surveys Tutorials* 21, 1 (2018), 586–618. DOI: https://doi.org/10.1109/COMST.2018.2863942

Audun Jøsang, Roslan Ismail, and Colin Boyd. 2007. A survey of trust and reputation systems for online service provision. *Decision Support Systems* 43, 2 (Mar. 2007), 618–644. DOI: https://doi.org/10.1016/j.dss.2005.05.019

G. Karypis, Eui-Hong Han, and V. Kumar. 1999. Chameleon: Hierarchical clustering using dynamic modeling. *Computer* 32, 8 (Aug. 1999), 68–75. DOI: https://doi.org/10.1109/2.781637

Ehsan Khadangi and Alireza Bagheri. 2013. Comparing MLP, SVM and KNN for predicting trust between users in Facebook. In *Proceedings of ICCKE 2013*. 466–470. DOI: https://doi.org/10.1109/ICCKE.2013.6682864

Young Ae Kim and Hee Seok Song. 2011. Strategies for predicting local trust based on trust propagation in social networks. *Knowledge-Based Systems* 24, 8 (Dec. 2011), 1360–1371. DOI: https://doi.org/10.1016/j.knosys.2011.06.009

Nikolay Korovaiko and Alex Thomo. 2013. Trust prediction from user-item ratings. *Social Network Analysis and Mining* 3, 3 (Sept. 2013), 749–759. DOI: https://doi.org/10.1007/s13278-013-0122-z

Bixin Li, Liao Li, Hareton Leung, and Rui Song. 2014. PHAT: A preference and honesty aware trust model for web services. *IEEE Transactions on Network and Service Management* 11, 3 (Sept. 2014), 363–375. DOI: https://doi.org/10.1109/TNSM.2014.2325771

Xiaoyong Li, Feng Zhou, and Xudong Yang. 2011. A multi-dimensional trust evaluation model for large-scale P2P computing. *Journal of Parallel Distributed Computing* 71, 6 (June 2011), 837–847. DOI: https://doi.org/10.1016/j.jpdc.2011.01.007

Fengming Liu, Li Wang, Lei Gao, Haixia Li, Haifeng Zhao, and Sok Khim Men. 2014. A web service trust evaluation model based on small-world networks. *Knowledge-Based Systems* 57 (2014), 161–167. DOI: https://doi.org/10.1016/j.knosys.2013.12.015

Haifeng Liu, Ee-Peng Lim, Hady W. Lauw, Minh-Tam Le, Aixin Sun, Jaideep Srivastava, and Young Ae Kim. 2008. Predicting trusts among users of online communities: An epinions case study. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC'08)*. ACM, New York, 310–319. DOI: https://doi.org/10.1145/1386790.1386838

Lianggui Liu and Huiling Jia. 2015. Trust evaluation via large-scale complex service-oriented online social networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45, 11 (Nov. 2015), 1402–1412. DOI: https://doi.org/10.1109/TSMC.2015.2406858

Shushu Liu, Lifang Zhang, and Zheng Yan. 2018. Predict pairwise trust based on machine learning in online social networks: A survey. *IEEE Access* 6 (2018), 51297–51318. DOI: https://doi.org/10.1109/ACCESS.2018.2869699

Xin Liu, Gilles Tredan, and Anwitaman Datta. 2013. A generic trusted framework for large-scale open systems using machine learning. *Computational Intelligence* 30, 4 (2013), 700–721. DOI: https://doi.org/10.1111/coin.12022 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/coin.12022

S. Lloyd. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28, 2 (Mar. 1982), 129–137. DOI: https://doi.org/10.1109/TIT.1982.1056489

Jorge López and Stephane Maag. 2015. Towards a generic trust management framework using a machine-learning-based trust model. In *Proceedings of the 2015 IEEE Trustcom/BigDataSE/ISPA - Volume 01 (TRUSTCOM'15)*. IEEE Computer Society, Washington, DC, 1343–1348. DOI: https://doi.org/10.1109/Trustcom.2015.528

Nan Ma, Ee-Peng Lim, Viet-An Nguyen, Aixin Sun, and Haifeng Liu. 2009. Trust relationship prediction using online product review data. In *Proceedings of the 1st ACM International Workshop on Complex Networks Meet Information & Knowledge Management (CNIKM'09)*. ACM, New York, NY, USA, 47–54. DOI: https://doi.org/10.1145/1651274.1651284

Chengying Mao and Rongru Lin. 2016. QoS trust rate prediction for web services using pso-based neural network. In *2016 International Conference on Advanced Cloud and Big Data (CBD)*. 68–74. DOI: https://doi.org/10.1109/CBD.2016.022

Chengying Mao, Rongru Lin, Changfu Xu, and Qiang He. 2017. Towards a trust prediction framework for cloud services based on PSO-driven neural network. *IEEE Access* 5 (2017), 2187–2199. DOI: https://doi.org/10.1109/ACCESS.2017.2654378

Henrik H. Martensa. 1959. Two notes on machine "Learning". *Information and Control* 2, 4 (1959), 364–379. DOI: https://doi.org/10.1016/S0019-9958(59)80014-0

M. Hadi Mashinchi, Lei Li, Mehmet A. Orgun, and Yan Wang. 2011. The prediction of trust rating based on the quality of services using fuzzy linear regression. In *Proceedings of the 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*. 1953–1959. DOI: https://doi.org/10.1109/FUZZY.2011.6007745

Muhammad Mohsin Mehdi, Imran Raza, and Syed Asad Hussain. 2017. A game theory based trust model for vehicular ad hoc networks (VANETs). *Computer Networks* 121, C (July 2017), 152–172. DOI: https://doi.org/10.1016/j.comnet.2017.04.024

Ramakanta Mohanty, V. Ravi, and M. R. Patra. 2010. Web-services classification using intelligent techniques. *Expert Systems with Applications* 37, 7 (2010), 5484–5490. DOI: https://doi.org/10.1016/j.eswa.2010.02.063

Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. 2002. Notions of reputation in multi-agents systems: A review. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems*. 280–287. DOI: https://doi.org/10.1145/544741.544807

Tung Doan Nguyen and Quan Bai. 2018. A dynamic Bayesian network approach for agent group trust evaluation. *Computers in Human Behavior* 89 (2018), 237–245. DOI: https://doi.org/10.1016/j.chb.2018.07.028

Xudong Ni and Junzhou Luo. 2008. A clustering analysis based trust model in grid environment supporting virtual organizations. In *Proceedings of the 22nd International Conference on Advanced Information Networking and Applications - Workshops (AINA workshops 2008)*. 100–105. DOI: https://doi.org/10.1109/WAINA.2008.162

Danmei Niu, Lanlan Rui, Haoqiu Huang, and Xuesong Qiu. 2017. A service recovery method based on trust evaluation in mobile social network. *Multimedia Tools and Applications* 76, 3 (Feb. 2017), 3255–3277. DOI: https://doi.org/10.1007/s11042-016-3963-4

Alexandra Olteanu, Stanislav Peshterliev, Xin Liu, and Karl Aberer. 2013. Web credibility: Features exploration and credibility prediction. In *Proceedings of the 35th European Conference on Advances in Information Retrieval (ECIR'13)*. Springer-Verlag, Berlin, 557–568. DOI: https://doi.org/10.1007/978-3-642-36973-5_47

Athanasios Papaoikonomou, Magdalini Kardara, and Theodora Varvarigou. 2015. Trust inference in online social networks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 (ASONAM'15)*. ACM, New York, 600–604. DOI: https://doi.org/10.1145/2808797.2809418

Isaac Pinyol and Jordi Sabater-Mir. 2013. Computational trust and reputation models for open multi-agent systems: A review. *Artificial Intelligence Review* 40, 1 (June 2013), 1–25. DOI: https://doi.org/10.1007/s10462-011-9277-z

J. R. Quinlan. 1987. Simplifying decision trees. *International Journal of Man-Machine Studies* 27, 3 (1987), 221–234. DOI: https://doi.org/10.1016/S0020-7373(87)80053-6

Stuart Russell and Peter Norvig. 2009. *Artificial Intelligence: A Modern Approach* (3rd ed.). Prentice-Hall.

Stefan Schmidt, Robert Steele, Tharam S. Dillon, and Elizabeth Chang. 2007. Fuzzy trust evaluation and credibility development in multi-agent systems. *Applied Soft Computing* 7, 2 (March 2007), 492–505. DOI: https://doi.org/10.1016/j.asoc.2006.11.002

Weihua Song. 2005. *Evaluating Online Trust Using Machine Learning Methods*. Ph.D. Dissertation, Louisiana Technical University, Ruston, LA, United States. AAI3170187.

Mark J. Stefik. 1985. Machine learning: An artificial intelligence approach: R. S. Michalski, J. G. Carbonell and T. M. Mitchell, (Tioga, Palo Alto, CA); 572 pages. *Artificial Intelligence* 25, 2 (1985), 236–238. DOI: https://doi.org/10.1016/0004-3702(85)90005-0

Juliana Tolles and William J. Meurer. 2016. Logistic regression: Relating patient characteristics to outcomes. *JAMA* 316, 5 (08 2016), 533–534. DOI: https://doi.org/10.1001/jama.2016.7653

Yelena Trofimova, Alexandru Mihnea Moucha, and Pavel Tvrdik. 2017. Application of neural networks for decision making and evaluation of trust in ad-hoc networks. In *Proceedings of the 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. 371–377. DOI: https://doi.org/10.1109/IWCMC.2017.7986315

Omar Abdel Wahab, Jamal Bentahar, Hadi Otrok, and Azzam Mourad. 2015. A survey on trust and reputation models for web services. *Decision Support Systems* 74, C (June 2015), 121–134. DOI: https://doi.org/10.1016/j.dss.2015.04.009

Guanghao Wang and Yue Wu. 2014. BIBRM: A Bayesian inference based road message trust model in vehicular ad hoc networks. In *Proceedings of the 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*. 481–486. DOI: https://doi.org/10.1109/TrustCom.2014.137

Wei Wang, Jiong Yang, and Richard R. Muntz. 1997. STING: A statistical information grid approach to spatial data mining. In *Proceedings of the 23rd International Conference on Very Large Data Bases (VLDB'97)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, 186–195.

Xin Wang, Ying Wang, and Hongbin Sun. 2016. Exploring the combination of Dempster-Shafer theory and neural network for predicting trust and distrust. *Computational Intelligence Neuroscience* 2016, Article 23 (Jan. 2016), 1 page. DOI: https://doi.org/10.1155/2016/5403105

Yuji Wang. 2017. The trust value calculating for social network based on machine learning. In *Proceedings of the 2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Vol. 2. 133–136. DOI: https://doi.org/10.1109/IHMSC.2017.145

Yan Wang and Kwei-Jay Lin. 2008. Reputation-oriented trustworthy computing in e-commerce environments. *IEEE Internet Computing* 12, 4 (July 2008), 55–59. DOI: https://doi.org/10.1109/MIC.2008.84

Christopher J. C. H. Watkins and Peter Dayan. 1992. Q-learning. *Machine Learning* (1992), 279–292. DOI: https://doi.org/10.1007/BF00992698

Aleksander Wawer, Radoslaw Nielek, and Adam Wierzbicki. 2014. Predicting webpage credibility using linguistic features. In *Proceedings of the 23rd International Conference on World Wide Web (WWW'14 Companion)*. ACM, New York, 1135–1140. DOI: https://doi.org/10.1145/2567948.2579000

Yuping Wu. 2010. Research of trust degree evaluation for C2C E-commerce based on fuzzy neural network. In *Proceedings of the 2010 2nd International Conference on E-business and Information System Security*. 1–4. DOI: https://doi.org/10.1109/EBISS.2010.5473508

Yue Wu, Fanchao Meng, Guanghao Wang, and Yi Ping. 2015. A Dempster-Shafer theory based traffic information trust model in vehicular ad hoc networks. In *Proceedings of the 2015 International Conference on Cyber Security of Smart Cities, Industrial Control System and Communications (SSIC)*. 1–7. DOI: https://doi.org/10.1109/SSIC.2015.7245329

Hamdi Yahyaoui and Aisha Al-Mutairi. 2016. A feature-based trust sequence classification algorithm. *Information Sciences* 328 (2016), 455–484. DOI: https://doi.org/10.1016/j.ins.2015.08.008

Hamdi Yahyaoui and Sami Zhioua. 2013. Bootstrapping trust of Web services based on trust patterns and hidden Markov models. *Knowledge and Information Systems* 37, 2 (Nov. 2013), 389–416. DOI: https://doi.org/10.1007/s10115-012-0554-1

Zheng Yan. 2010. *Trust Modeling and Management in Digital Environments: From Social Concept to System Development*. IGI Global, Hershey, PA. DOI: https://doi.org/10.4018/978-1-61520-682-7

Zheng Yan. 2013. *Trust Management in Mobile Environments: Autonomic and Usable Models* (1st ed.). IGI Global, Hershey, PA.

Zheng Yan, Yu Chen, and Yue Shen. 2013a. A practical reputation system for pervasive social chatting. *Journal of Computer System Sciences* 79, 5 (Aug. 2013), 556–572. DOI: https://doi.org/10.1016/j.jcss.2012.11.003

Zheng Yan, Yu Chen, and Yue Shen. 2014a. PerContRep: A practical reputation system for pervasive content services. *The Journal of Supercomputing* 70, 3 (Dec. 2014), 1051–1074. DOI: https://doi.org/10.1007/s11227-014-1116-y

Zheng Yan, Yan Dong, Valtteri Niemi, and Guoliang Yu. 2013b. Exploring trust of mobile applications based on user behaviors: An empirical study. *Journal of Applied Social Psychology* 43, 3 (2013), 638–659. DOI: https://doi.org/10.1111/j.1559-1816.2013.01044.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1559-1816.2013.01044.x

Zheng Yan, Xuyang Jing, and Witold Pedrycz. 2017. Fusing and mining opinions for reputation generation. *Information Fusion* 36 (2017), 172–184. DOI: https://doi.org/10.1016/j.inffus.2016.11.011

Zheng Yan and Christian Prehofer. 2011. Autonomic trust management for a component-based software system. *IEEE Transactions on Dependable and Secure Computing* 8, 6 (Nov 2011), 810–823. DOI: https://doi.org/10.1109/TDSC.2010.47

Zheng Yan, Peng Zhang, and Robert H. Deng. 2012. TruBeRepec: A trust-behavior-based reputation and recommender system for mobile applications. *Personal and Ubiquitous Computing* 16, 5 (June 2012), 485–506. DOI: https://doi.org/10.1007/s00779-011-0420-2

Zheng Yan, Peng Zhang, and Athanasios V. Vasilakos. 2014b. A survey on trust management for internet of things. *Journal of Network and Computer Applications* 42 (2014), 120–134. DOI: https://doi.org/10.1016/j.jnca.2014.01.014

Zhengwang Ye, Tao Wen, Zhenyu Liu, Xiaoying Song, and Chongguo Fu. 2017. An efficient dynamic trust evaluation model for wireless sensor networks. *Journal of Sensors* 2017 (Oct. 2017), 1–16. DOI: https://doi.org/10.1155/2017/7864671

Weiwei Yuan, Donghai Guan, Sungyoung Lee, and Youngkoo Lee. 2006. A dynamic trust model based on naive bayes classifier for ubiquitous environments. In *Proceedings of the 2nd International Conference on High Performance Computing and Communications (HPCC'06)*. Springer-Verlag, Berlin, 562–571. DOI: https://doi.org/10.1007/11847366_58

Tong Zhang, Lisha Yan, and Yuan Yang. 2018. Trust evaluation method for clustered wireless sensor networks based on cloud model. *Wireless Networks* 24, 3 (01 Apr 2018), 777–797. DOI: https://doi.org/10.1007/s11276-016-1368-y

Weiyu Zhang, Bin Wu, and Yang Liu. 2016. Cluster-level trust prediction based on multi-modal social networks. *Neurocomputing* 210 (2016), 206–216. DOI: https://doi.org/10.1016/j.neucom.2016.01.108

Xiuzhen Zhang, Lishan Cui, and Yan Wang. 2014. CommTrust: Computing multi-dimensional trust by mining e-commerce feedback comments. *IEEE Transactions on Knowledge and Data Engineering* 26, 7 (July 2014), 1631–1643. DOI: https://doi.org/10.1109/TKDE.2013.177

Kang Zhao and Li Pan. 2014. A machine learning based trust evaluation framework for online social networks. In *Proceedings of the 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications (TRUSTCOM'14)*. IEEE Computer Society, Washington, DC, 69–74. DOI: https://doi.org/10.1109/TrustCom.2014.13

Qinghua Zheng, Jun Liu, Hongwei Zeng, Zhaotong Guo, Bei Wu, and Bifan Wei. 2019. Knowledge forest: A novel model to organize knowledge fragments. *Science China (Information Sciences)* (2019).

Peng Zhou, Xiaojing Gu, Jie Zhang, and Minrui Fei. 2015. *A priori* trust inference with context-aware stereotypical deep learning. *Knowledge-Based Systems* 88, C (Nov. 2015), 97–106. DOI: https://doi.org/10.1016/j.knosys.2015.08.003

Yuquan Zhu and Zheng Yan. 2016. A survey on trust evaluation in e-commerce. In *Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications (MobiMedia'16)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, Belgium, 130–139.

Kiyana Zolfaghar and Abdollah Aghaie. 2011. Evolution of trust networks in social web applications using supervised learning. *Procedia Computer Science* 3 (2011), 833–839. DOI: https://doi.org/10.1016/j.procs.2010.12.137

Kiyana Zolfaghar and Abdollah Aghaie. 2012. A syntactical approach for interpersonal trust prediction in social web applications: Combining contextual and structural data. *Knowledge-Based Systems* 26 (2012), 93–102. DOI: https://doi.org/10.1016/j.knosys.2010.10.007

Received December 2018; revised June 2020; accepted June 2020