Received July 21, 2018, accepted August 24, 2018, date of publication September 10, 2018, date of current version October 8, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2869699

Predict Pairwise Trust Based on Machine Learning in Online Social Networks: A Survey

SHUSHU LIU[1], LIFANG ZHANG[1], AND ZHENG YAN[1,2], (Senior Member, IEEE)

[1]Department of Communication and Networking, Aalto University, 02150 Espoo, Finland
[2]State Key Laboratory of Integrated Services Networks, School of Cyber Engineering, Xidian University, Xi'an 710071, China

Corresponding author: Zheng Yan (zyan@xidian.edu.cn)

This work was supported in part by the Academy of Finland under Grant 308087, in part by the NSFC under Grant 61672410, Grant 61802293, and Grant U1536202, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Program 2016ZDJC-06, in part by the National Key Research and Development Program of China under Grant 2016YFB0800704, and in part by the 111 Project under Grant B08038 and Grant B16037.

ABSTRACT Trust plays a crucial role in online social networks where users do not communicate or interact with each other in a direct face-to-face manner. Although many researchers have already conducted comprehensive studies on trust computing like trust evaluation, pairwise trust prediction is still relatively under explored especially with machine learning methods which can overcome the disadvantages of both linear predication and trust propagation. This survey aims to fill this gap and first provides an overview of state-of-the-art researches in pairwise trust prediction using machine learning techniques, especially in the context of social networking. Specifically, we present a workflow of trust prediction using machine learning and summarize current available trust-related datasets, classifiers and different metrics used to evaluate a trained classifier. Also, we review, compare, and contrast the literature for the purpose of identifying open issues and directing future research.

INDEX TERMS Trust prediction, machine learning, social networks, trust evaluation.

I. INTRODUCTION

Online Social Networks (OSNs) such as Facebook and Twitter have become a main media for people to socialize and for governments and enterprises to deliver their services. Moreover, OSNs have gone beyond socialization and information dissemination and have entered into nearly every part of our daily life such as commercial, enterprise and entertainment. As a result, the ubiquitous OSNs have become an important infrastructure of our society. However, online social communities often have a huge amount of data generated by users such as discussions, reviews and ratings. Such tremendous data might be irrelevant, unreliable or untrustworthy but filtering them out manually is impossible. In such situation, trust can help users to dissect relevant and reliable information, and thus address the information overload and credibility problems. In addition, users do not necessarily know each other in such an uncertain and semi-virtual environment as OSNs. Trust can enable users to receive high quality recommendations and opinions even from those with whom they are unacquainted. Moreover, trust is often employed to fight against spams and attacks [1] and thus improving the security of OSNs. Sherchan *et al.* [2] has identified trust as a critical factor to the social capital of an online community.

The growing popularity of OSNs combined with the pivotal role of trust to them has spurred increasing amount of research on constructing trust in OSNs. Ruan and Durresi [4] surveyed trust inference and potential local trust related attacks in online social communities. Jøsang *et al.* [9] analyzed researches on trust and reputation systems for online service provision. Although there are several surveys about trust computation, they rarely investigated pairwise trust prediction that describes a relationship between two users of OSNs, especially with machine learning based methods, which have shown to be effective in inferring latent trust relations and outperform other conventional methods in terms of accuracy [16], [19].

The application of machine learning to trust prediction in OSNs has many advantages. Commonly, trust models are based on a linear combination of input features such as knowledge, experience, and reputation, and the weight of this linear combination determines the importance of each feature [68]. However, the concept of trust is much more complex and subjective, for instance, two low value features can contribute to a high trust when combined together other than treated separately. A simple combination of features is not enough to formulate trust correctly, while machine learning based methods have high prediction accuracy for both linear and non-linear modeling. Besides, the best selection of weights is still an open problem that needs to be solved. Apart from linear combination, many works tried to model trust prediction into trust propagation that propagates trust values through a web of trust. The effectiveness of this approach highly depends on the connectivity of the known web of trust and can be quite poor when the connectivity is very sparse which is often the case in OSNs. In contrast, machine learning based methods are flexible and have high generality in integrating trust inducing factors especially when some information is unavailable or highly sparse.

Since the literature still lacks a thorough survey on this topic despite its rapid development, this survey aims to fill this gap and provides an overview of state-of-the-art researches in pairwise trust prediction using machine learning techniques. Specifically, we define trust and discuss its characteristics from multiple perspectives. We also summarize a workflow of trust prediction using machine learning and analyze current available trust-related datasets, classifiers and different metrics used to evaluate a trained classifier. In addition, we review existing studies and enumerate their strengths and limitations. Finally, we discuss open research problems and propose a number of potential future research directions in this research field. This survey is addressed to researchers and practitioners interested in applying machine learning to predict pairwise trust in online social networks. In general, the main contributions of this paper are summarized as follows:

- To the best of our knowledge, this is the first work synthesizing the pairwise trust prediction in OSNs with focus on machine learning methods. We summarize the pipeline of machine learning based trust prediction and analyze recent ten years' work accordingly.
- We provide a comprehensive summary and comparison of available datasets regarding OSNs under uniform standards.
- We summarize open research issues and challenges in machine learning based trust predictions regarding OSNs based on in-depth literature study and analysis. We also propose a number of future research directions that worth further investigation.

The rest of the paper is organized as follows. Section II reviews some related work. Section III presents trust definition, trust properties, trust inducing factors and uniform criteria of trust prediction in the context of OSNs. In Section IV, we comprehensively summarize and compare available datasets regarding OSNs. Section V walks through the pipeline of predicting trust with machine learning. Section VI discusses existing trust prediction systems and enumerates their strengths and limitations. In Section VII, we discuss open research issues and propose future research directions in this research field. Finally, a conclusion is given in the last section.

II. RELATED WORKS

As OSNs are the main social media and advertisement platform nowadays, more and more related researches are conducted on it especially in the field of trust computing which is requisite for online communication. Here, we are going to present several survey papers in the research of trust.

Trust related surveys have been applied in many domains, such as E-commerce, Mobile Ad Hoc Network (MANET), Internet of Things (IoT) and so on. Zhu and Yan [62] analyzed trust evaluation with referring to five main E-commerce scenarios: Business-to-Customer (B2C), Business-to-Business (B2B), Peer-to-Peer (P2P), Customer-to-Customer (C2C), Government-to-Customer (G2C). Xu and Yan [63] discussed trust evaluation in MANET which is a multi-hop temporary and autonomic network comprised of a set of mobile nodes. They reviewed papers in recent ten years and compared them under a uniform criteria. Yan *et al.* [64] did a comprehensive literature review on trust management in IoT with focus on topics like trust evaluation, trust framework, data perception trust, identity trust and privacy preservation and etc. They advised that researches in IoT should be oriented and driven by practical needs and demands like power-efficient technologies and lightweight trust management. Abdelghani *et al.* [66] discussed trust management in a new paradigm, social internet of things which is a combination of IoT and social networks.

In the field of OSNs, references [2] and [4] are the two most related work. Sherchan *et al.* [2] presented a survey addressing three aspects of social trust: trust information collection, trust evaluation and trust dissemination. While Ruan and Durresi [4] paid more attention to trust modelling, trust inference and attacks. For the attack analysis, they focused on local trust related attacks like naive attack, traitor attack, whitewashing attack, collusion attack instead of global trust related attack [67].

For the other related works, Cho *et al.* [10] provided an efficient and relevant background knowledge on how to model trust in a given domain. They mainly focused on the properties and formulation of trust factors in different contexts. Chomphoosang *et al.* [65] did a research on watermark techniques that can be used in social networks to protect the content of information. Pranata *et al.* [61] conducted a usability and effectiveness analysis on trust rating systems through anonymous online survey. Compared with binary rating systems, 5-star rating systems and notation-based rating systems are preferred in terms of usability.

Although there are several surveys about trust computation, they rarely investigated pairwise trust prediction, especially with machine learning based methods, which have shown to be effective in inferring latent trust relations and outperform other conventional methods in terms of accuracy. Many conventional methods just considered a weighted function like sum of trust inducing factors [68]. Therefore, in this paper we provide a comprehensive survey on machine learning based pairwise trust prediction in OSNs.

III. BACKGROUND AND CRITERIA OF PAIRWISE TRUST PREDICTION IN OSNs

A. PAIRWISE TRUST DEFINITION AND PROPERTIES

Pairwise trust in OSNs can be described as the edge of a directed graph $G = (V, E)$ where V represents a set of unique users and $E \subset V \times V$ denotes a set of edges (also known as relationships) between users. In our problem, each edge has a label to indicate the trust between users. We may have the label information for some edges, which is encoded as E^L, and for the other edges encoded as E^U, we do not have the label information, where $E = E^L \cup E^U$. Our general objective is to predict the trust level of unlabled edges in E^U based on the available information in the OSNs. For the sake of simplicity, pairwise trust is also denoted as trust in the rest part of this paper.

Further, we can summarize the properties of pairwise trust as follows:

- Subjective. Trust evaluation is highly impacted by the personal preferences, biases and dispositions of a trustor. Thus, subjectivity has been recognized as the inherent nature of trust. The subjective nature of trust hints that trust evaluation should be personalized. However, the subjective property of trust also means that objective trust assessment may not be possible since evidence may be uncertain, incomplete, and conflicting in reality.
- Contextual. Trust computing depends on specific contexts. More specifically, a user who is trusted to be a good reviewer in products related to home and garden might not be trusted to be a good recommender for restaurants, as shown in real world dataset Epinions [11].
- Dynamic. Normally, trust needs to be reassessed when new evidences have arrived or the social contexts (e.g. task, risk) have been changed. The outcome of prior trust decision also plays a role in updating trust. Specifically, the positive outcome of previous trust decision increases trust and vice versa. Moreover, trust decays over time. Many trust models assign more weights on recent evidences.
- Asymmetric. A trusts B does not mean B also trusts A. Asymmetry also means that A and B have unequal degree of trust toward each other. However, the asymmetric gap tends to be mitigated by interactions.
- Propagative. If A trusts B, and B trusts C with whom A has no connection, the trust of A towards C can be deduced from the degree of trust A to B and B to C. Such propagation can be used to create a trust chain.
- Fragile. Trust is highly sensitive to negative interactions. Specifically, negative evidence destroys trust much easier than positive evidence establishes trust. Trust building is much more difficult than trust destroying.

B. PAIRWISE TRUST MEASURING SCALES

Modeling trust as a computational concept cannot avoid specifying its scale. Different trust scaling methods have been proposed in the literature. However, there is no standard to scale trust yet. We present different trust scaling methods here and discuss their pros and cons in the context of OSNs. We classify trust scaling into three types: namely, binary, discrete/nominal values and continuous values.

a) **Binary.** Modeling trust as a binary concept is simple and efficient when it comes to decide if an entity should be trusted or not. Moreover, nearly all social networks that enable their users to specify trust relationship treat trust as binary. For example, in Epinions, users can either trust or distrust another user. Facebook users can either be friends or not. As a result, modeling trust as a binary concept has the advantage of having real life datasets for experiments. Another advantage is that machine learning in classifying problems is relatively mature. However, binary scale is obviously very restrictive and thus loses resolution. It cannot express the degree of trust or distrust. In addition, asking users to provide a binary trust might be challenging to them. This is because users might neither trust nor distrust another user due to the lack of evidence.

b) **Discrete/Nominal value.** Nominal value expresses different levels of trust in words. The nominal words can be complete distrust, moderate distrust, moderate trust, and complete trust. These nominal values can also be represented with discrete numbers. We thus classify them into one group herein. Unlike binary value, discrete or nominal trust value enables a user to explain how much he or she trusts another. However, no trust-related dataset contains more than two levels of trust.

c) **Continuous value.** Continuous value is more expressive and can describe the extent to which one trusts another. Marsh scales trust in $[-1, 1]$ where 0 indicates the state of being completely ignorant or uncertain [14]. Continuous value is very powerful and flexible in the sense that it can express both binary and discrete or nominal trust. The continuous trust can be converted to binary trust by comparing its value with a certain threshold set by users. Such a threshold can be tuned to satisfy different requirements of various applications and trustors. Specifically, important tasks and stricter users can have higher threshold. The discrete or nominal version of continuous trust can be obtained by specifying a range for each of the nominal value. However, no trust-related dataset contains continuous trust value. Therefore, to model trust as continuous value, one must conduct a pricey survey to obtain such continuous trust.

C. CRITERIA OF PAIRWISE TRUST PREDICTION

To clarify the underlying issues of current work and identify open research issues for future study, we propose a number of criteria for assessing the machine learning based pairwise trust prediction models based on the properties of pairwise trust as mentioned above. To provide trustworthy OSNs trust prediction models, researches should achieve the following goals:

- *Trustworthiness:* the trust prediction should be robust to overcome various potential attacks on it, such as self-promoting attacks, bad-mouthing attacks, whitewashing attacks, discriminatory attacks, denial of service, orchestrated and etc.
- *Adaptability:* the trust relationships are normally dynamically changed due to the leaving and joining of different users. Besides, the outcome of prior trust prediction also plays a role in updating trust. Moreover, trust decays over time.
- *Usability:* trust prediction in online social network should consider the subjective opinion of participants (both trustor and trustee) and be helpful with regard to the interactions.
- *Privacy:* user privacy should be concerned when user data is collected by a central party since the private information like hobby, religion or health status can be induced from user's behavior data.
- *Accuracy:* the accuracy of trust prediction should be ensured without any doubt.
- *Efficiency:* the algorithms for trust prediction should be efficient in order to dealing with large-scale datasets and dynamically manage trust predictions in OSNs.
- *Uniformity:* It is preferred to offer a uniform model considering users subjective factors with trusted credibility for trust prediction.
- *Comprehension:* The trust prediction should concern various trust inducing factors in a comprehensive way. This is essential for achieving accurate trust prediction and recommendation.
- *Generality:* Trust prediction for various systems and service can be commonly or widely used in different application scenarios, which is a preferred objective for trust prediction.

IV. DATASETS

In machine learning, dataset represents experience, which is one of the most essential components of machine learning. Therefore, in order to train an effective trust prediction model for OSNs, the first essential step is to obtain a high quality dataset. Among the trust-related datasets used in papers we have reviewed, except a simulated one [16], the rest of them are from real life applications. Ideally, the relevant official companies offer their trust-related datasets and allow them to be downloaded free of charge. Such official datasets have the advantage of being comprehensive. Specifically, they can include sensitive data such as distrust information, which is inaccessible to the public and is thus uncrawlable. The companies usually anonymize their datasets in order to protect the privacy of their users. However, this kind of datasets is limited. The only publicly available official dataset is the Extended Epinions dataset. Moreover, the official datasets confine us to construct trust inducing features based only on the information provided by them. As a result, some researchers choose to crawl data by themselves.

Crawled datasets have the advantage of being flexible in the sense that we have freedom to choose what data to crawl. We classify the crawled datasets into two types depending on whether the crawled OSN supports user specified trust or not. For OSNs with publicly available trust relationship such as Epinions, trust can be crawled with other information such as reviews, ratings, and profiles. For OSNs like Facebook and Twitter that do not enable users to specify trust, however, crawling trust is impossible. Instead, to collect a dataset for this kind of OSNs, publicly accessible data such as posts, follows, shares, and profiles must be crawled first and then ask a trust evaluator to assign a trust relationship for a given pair of users according to their crawled data which will be presented to the trust evaluator. It is preferred that the trust evaluator in the survey is exactly the trustor as in [19], [28]. This is because trustor related features such as trustor leniency could not be modeled from his data anymore if the trust evaluator and trustor is not the same person. However, collecting such trust values via survey is expensive because a system need to be designed carefully and scientifically to gather trust from a reasonable large number of participants. However, current datasets collected this way [19] only asked a small number of participants to evaluate trust. In addition, their collecting systems are quite primitively since they ignore many factors a good survey should have. We refer to the datasets with crawled trust as fully crawled dataset and the datasets with surveyed trust as surveyed dataset. None of the datasets contains privacy sensitive data such as block list and the private profile information. Since none of surveyed dataset is publicly available, we focus on some of typical fully crawled datasets in the subsequent discussion.

Most of the fully crawled datasets are from review websites. This is because review sites are the most common kind of social networks that support user specified trust. The most typical ones of them are general product review sites (e.g. Epinions, Ciao) and movie review sites (e.g. Filmtrust, Flixster). In these websites, users can publish textual reviews along with relevant ratings about products they have purchased or used or movies they have watched. Other users can rate the reviews in terms of helpfulness. Moreover, these websites allow their users to maintain a trust list where they can add those whose reviews are consistently valuable to them. In addition to the review sites, some researches have also employed other websites that support certain kind of social relationships to study trust. The most typical ones are Flixster, Wikipedia vote network and Slashdot.

The main reason for our dataset research is because we notice that the statistics about some datasets in the literature are not consistent even though they came from the same source. A good example is the statistics about Film Trust-related dataset in [17] and [18]. In addition, the units used in the statistics are inconsistent. Some papers use file size and others the number of records. We use the number of records to summarize the publicly available datasets in the sense that it is more consistent with machine learning terminology.

TABLE 1. Extended Epinions dataset.

File	# of records	Record information
User trust	841,372	{user id, other id, trusted value (1,-1), date} (Trust: 85.3%)
Article (user reviews)	1,198,115	{content id, author id, subject id}
Article rating	3,024,664	{content id, rater id, rating score, status, date, last modified, object type, verticle id}
Rating score: 1: Not Helpful 2: Somewhat Helpful 3: Helpful 4: Very Helpful 5: Most Helpful		

Table 7 shows the comparison of different datasets in OSNs in the literature.

A. EXTENDED EPINION DATASET

Epinions datasets have been used extensively to predict trust since they provide both web of trust and user behavioral data. As a result, many different Epinions datasets exist and are publicly available. The most widely used one is the extended Epinions dataset[1] provided directly by Epinions staff. The dataset contains both trust and distrust relations. Specifically, the dataset consists of three files as shown in Table 1. It is worth to note that the rating information in the dataset is about reviews written by another user instead of about products. The great advantage of this dataset is that it includes the inaccessible distrust information, which also explains its wide usage.

B. EPINIONS DATASETS

Both Tang *et al.* [11][2] and Massa and Avesani [21][3] published their crawled datasets from Epinions. Both of their datasets contain product ratings and trust relationships. However, the datasets from Tang *et al.* contain more information. Specifically, Tang *et al.* published four different versions of dataset they crawled from Epinions. One is raw dataset without any preprocessing and others differ from each other depending on whether information about time or category is included or not. Table 2 presents the most comprehensive version that includes both category and time information. It is worth to note that only relative time is available since the time stamps was obtained by splitting time points into 11 parts. However, none of these two datasets contains distrust information.

C. CIAO DATASETS

Two Ciao trust-related datasets are publicly available so far. One was crawled by Tang *et al.* [11]. Another one called CiaoDVD was crawled by Guo *et al.* [22] for the entire DVD category.[4] Tang *et al.* also crawled the textual content of each review, which, however, is not published. As shown in the table, this dataset contains item ratings and binary trust relationship. The rating contains both the ratings given by users to products in specific categories and the global helpfulness of this rating perceived by other users. The personal view of the rating helpfulness, which describes user interactions in terms of review rating, is, however, not included in this dataset. On the other hand, CiaoDVD contains review rating that describes the personal view of the review helpfulness. However, not all users of CiaoDVD with trust relationship provide movie or review ratings. Specifically, only 1438 of 4658 different users with explicit trust have issued trust relationship, 4455 of them have rated review, and 2740 of them have rated movies. Both Ciao dataset and CiaoDVD include the creating time of a product rating. Such time information is desirable since it enables the modeling of the dynamic property of trust features. However, none of them contains the established time of a trust relationship, which is needed when modeling the dynamic property of trust. In addition, none of the datasets includes information about when a review helpfulness rating is given. Moreover, both of them have only trust relationship but no information about distrust.

TABLE 2. Crawled datasets from Epinions and Ciao.

Dataset	File	# of records	Record information
Epinions_ Tang [11]	Product rating	922,267	{userid, productid, categoryid, rating, helpfulness, time stamps}
	Trust	300,548	{UserId1, UserId2, timestamps(integers)}
Epinions_ Massa [21]	Product rating	664,824	{user_id, item_id, rating_value}
	Trust	487,182	{Trustor id, trustee id, trust value (all 1)}
Ciao [11]	Product rating	284,086	{userid, productid, categoryid, rating, helpfulness, time stamps}
	Trust	57,544	{UserId1, UserId2}
CiaoDVD [22]	Movie rating	72,665	{userID, movieID, genreID, reviewID, movieRating, date}
	Review-ratings	1,625,480	{userId, reviewId, reviewRating (helpfulness)}
	Trust	40,133	{trustorId, trusteeId, trustRating (all 1)}

D. FLIXSTER

Flixster is a social movie site where users can rate movies out of five stars, write a textual review on a movie page, discover new movies and meet people with similar movie taste. Table 3 shows the Flixster friendship network dataset crawled by Javier Parra [23].[5] The friendship is interpreted as trust relationship in [13]. Since the dataset contains only friendship and thus no distrust information is available.

[1] http://www.trustlet.org/extended_epinions.html
[2] https://www.cse.msu.edu/ tangjili/trust.html
[3] http://www.trustlet.org/downloaded_epinions.html
[4] https://www.librec.net/datasets.html
[5] http://socialcomputing.asu.edu/datasets/Flixster

TABLE 3. Datasets of Flixter and FilmTrust.

Dataset	File	# of records	Record information
Flixster [23]	nodes	2,523,386	All the node ids used in the dataset
	Social relationship	9,197,337	{UserId1, UserId2} (Friendship)
FilmTrust [18][6]	Movie rating	35,497	{userId, movieId, movieRating [0.5,4.0]}
	Trust	1,853	{trustorId, trusteeId, trustRating (all 1)}

[6] https://www.librec.net/datasets.html

This dataset is restricted in the sense that no information about item ratings or review ratings is included. As a result, no user interactions can be extracted from this dataset. It is worth to note that the dataset used in [12] is a different dataset, which is not publicly available anymore.

E. FILMTRUST DATASET

FilmTrust is a trust-based movie sharing and rating website. In FilmTrust, users can add friends and thus create his own social network. In addition, they are required to indicate how much they trust each one of their added friends in terms of movie rating in the form of assigning them a trust rating. This explicitly specified trust or an inferred trust is then used to personalize a predictive movie rating for each user by weighting the ratings from other users according to how much the user trusts these raters [17]. Table 3 shows the dataset crawled by Guo *et al.* [18] from the entire FilmTrust website. As shown in the table, the dataset is quite small, including only 1508 different users. This might result from the fact that FilmTrust was developed for academic research instead of for real life application, thus it does not attract large enough number of users. The trust rating value in the dataset is always one. Therefore, only trust relationship is included and distrust information is unavailable. We notice that the information of this dataset presented in different papers seems to be contradictory. For example, in [18], the scale of movie ratings is shown to be [1, 5] while in [17] the scale is [0.5, 4.0]. In addition, the number of trust relationship is claimed to be 2853 [17], while in the public dataset it is 1853. The number of ratings in [17] is 70,998, which is different from the real dataset. The number of ratings in [18] is in the unit of file size, which cannot be compared with that of [17]. Thus, we analyze the publicly available dataset and present herein the information consistent with it.

TABLE 4. Datasets of wikipedia vote network.

Dataset[7]	# of votes	Record information	Positive percentage
Wiki-RfA [26]	198,275	{voter name, votee name, vote, result, year, date, text}	72.8%
Wiki-Elec [24], [25]	114,040	{result, election closed time, votee (id and name), nominator (id and name), [vote, voter id, time, voter name]}	73.6%
Vote: (1:support, 0:neutral, -1:oppose)			
Result: (promote 1, not promote 0)			
Note: our statistic is obtained from the publicly available dataset, which differs from that of [25], which discards neutral votes.			
[7] https://snap.stanford.edu/data/wiki-Vote.html			

TABLE 5. Datasets of slashdot.

Dataset	# of records	Relationship	Positive percentage
Soc-sign-Slashdot090221	549,202	Binary (1, -1)	77.4%
Soc-sign-Slashdot090216	545,671	Binary (1, -1)	77.39%
Soc-sign-Slashdot081106	516,575	Binary (1, -1)	76.73%
Record Information: {FromNodeId, ToNodeId, Sign (1, -1)}			

F. WIKIPEDIA VOTE NETWORK (DUMP)

Wikipedia is a free online encyclopedia collectively written and edited by volunteers all over the world. An editor of Wikipedia who wants to become an administrator must issue a request for adminship (RfA) either by himself or another community member. Whether the editor gets promotion or not is decided by how much votes he can receive from any other Wikipedia users. Other users can either support, oppose or be neutral about the promotion by voting. Table 4 summarizes the dataset dumped and parsed by West *et al.* [26] and Leskovec *et al.* [24], [25] from Wikipedia page edit history. Wiki-Elec includes voting history from January 2008 while wiki-RfA covers all voting data since the adoption of the RfA feature in 2003 through May 2013. In addition, wiki-RfA contains textual comments accompanied with each of the votes. The vote has been employed to indicate a trust or distrust relationship between two users. The neutral votes, however, are often discarded [20], [26].

G. SLASHDOT

Slashdot is a news website about current primary technologies. The users of Slashdot can tag those whose comments they like as friends and those whose comments they dislike as foes thanks to the Slashdot Zoo introduced in 2002. The friends and foes relationship have often been used to represent the trust and distrust [25], [26]. Table 5 presents datasets about tag relationships in Slashdot crawled at different times, which are indicated in their file names. Specifically, Soc-sign-Slashdot090221 indicates that it was crawled on February 21 2009. However, the datasets include no other information than the trust links.

H. BITCOIN WEB OF TRUST

Bitcoin web of trust describes the trust relationship among people who trade on Bitcoin platforms such as Bitcoin OTC or Bitcoin Alpha. Members of the Bitcoin platforms can rate others from −10 (total distrust) to 10 (total trust) in steps

TABLE 6. Datasets of Bitcoin OTC and Bitcoin alpha.

Dataset[8]	# of records	Record information	Relationship	Positive percentage
Bitcoin OTC	35,592	{rater, ratee, rating, time}	Discrete Trust [-10, 10]	89%
Bitcoin Alpha	24,186	{rater, ratee, rating, time}	Discrete Trust [-10, 10]	93%

[8] https://snap.stanford.edu/data/

of one. The Bitcoin web of trust is employed to prevent transactions with fraudulent and risky users, which happen more often on such platforms as Bitcoin OTC or Bitcoin Alpha whose members are anonymous. Table 6 shows the information about the two publicly available datasets about Bitcoin web of trust. As shown in the table, the datasets include only the trust relationships between two users. None of them contains data about user interactions. The time is measured as the elapsed seconds since Epoch.

I. ANALYSIS AND DISCUSSION

The contents of a dataset decide what features can be extracted from them, thus affecting the algorithms that can be applied to them. A dataset without any time information, for example, cannot model the dynamic property of trust. Thus, we summarize their characteristics here in Table 7.

First, most of publicly available datasets are crawled by researchers. As a result, most of them do not include distrust relationship due to privacy issue. However, they contain some desirable information about user interactions such as product rating. Some of them also include rating times and category information that are not included even in official datasets (e.g. Epinions_Tang, Ciao, CiaoDVD).

Second, most of datasets with available distrust relationship, however, have no data about user interactions (e.g. Wikipedia, Slashdot, Bitcoin). Extended Epinions dataset is the only one that contains both distrust and data about user interactions. Moreover, it is the only dataset that includes the information about when the trust or distrust relationship is created. However, none of the datasets contains user information even not simple demographic info such as age, gender, and occupation.

Third, the size of a dataset influences heavily the performance of a machine learner. As Domingos has pointed out "a dumb algorithm with lots and lots of data beats a clever one with modest amounts of it" [28]. In terms of trust-related dataset, except the extreme small Filmtrust dataset, most of them are in the order of millions. Datasets related to Bitcoin and Ciao are in the order of tens of thousands. The largest one is Flixster, which is in the order of billions. However, Flixster dataset provides no information about user interactions. From the perspective of rating data, Extended Epinions dataset has the largest number of review rating, which is much larger than the other datasets that contain review rating information.

Forth, all the datasets mentioned above that provide both trust and distrust relationships consist of overwhelmingly more trust than distrust. As shown in the above tables, the percentage of positive trust is above 70%. The most widely used dataset extended Epinions contains as high as 85% percent of trust. This is called class imbalance problem. The problem is inherent with trust-related datasets resulting from the fact that users are generally reluctant to specify distrust relations. Special care should be taken for trust prediction in that it is often modeled as a supervised classification problem that is sensitive to the class imbalanced distribution.

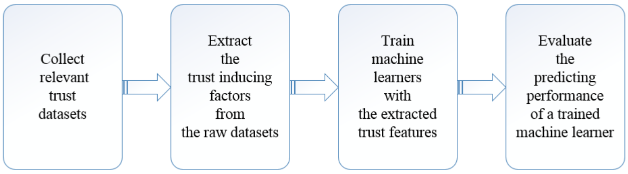
Collect relevant trust datasets
Extract the trust inducing factors from the raw datasets
Train machine learners with the extracted trust features
Evaluate the predicting performance of a trained machine learner

FIGURE 1. The pipeline of trust prediction models.

V. PIPELINE OF TRUST PREDICTION WITH MACHINE LEARNING

In this section, we outline the pipeline of trust prediction models according to the definition of machine learning. Machine learning is broadly defined by Mitchell 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" [15]. According to this definition, the trust prediction task can be framed as:

- Training experience E: a dataset containing the behavior history and the trust relationship of given user pairs
- Task T: predict the trust relationship for a pair of users given their behavior history
- Performance measure P: percentage of trust relationship correctly predicted

With this framework, we summarize the pipeline for predicting trust with machine learning technique in Fig. 1. As shown in the figure, the first step is to obtain or collect a dataset as we summarized in Section IV. The second step is to analyze and preprocess the dataset in order to quantify qualitative trust inducing factors. With quantitative trust inducing factors, next is to train machine learning models and evaluate their effectiveness according to the performance measure P. The trained models with satisfied performance can then be integrated to real life applications. Next, we discuss each of them in more detail.

A. FEATURE ENGINEERING

Feature engineering makes use of domain knowledge to create useful features from a raw dataset. It is needed since raw dataset is usually not amenable to learning. All the datasets mentioned above, for example, contains no direct information about trust inducing factors. Therefore, we need to construct a new feature set that can represent the underlying problem better in order to improve the predictive performance of a machine learning algorithm. Feature engineering is essential

TABLE 7. Summary of different datasets in OSNs.

Dataset	# of pairwise trust rating	interaction data	trust /distrust	time	trust rating	user profile	textual comment	category	percentage
Epinions_Extended	841,372	Y	both	Y	(-1,1)	N	N	N	85.3%
Epinions_Tang [11]	300,548	Y	trust	Y	1	N	N	Y	-
Epinions_Massa [21]	487,182	Y	trust	N	1	N	N	N	-
Ciao [11]	57,544	Y	trust	Y	1	N	N	Y	-
CiaoDVD [22]	40,133	Y	trust	Y	1	N	N	Y	-
Flixster [23]	9,197,337	N	trust	N	1	N	N	N	-
FilmTrust [18]	1,853	Y	trust	N	1	N	N	N	-
Wiki-RfA [26]	198,275	N	both	Y	(-1,0,1)	N	Y	Y	72.8%
Wiki-Elec [24], [25]	114,04	N	both	Y	(0,1)	N	Y	N	73.6%
Soc-sign-Slashdot090221	549,202	N	both	N	(-1, 1)	N	N	N	77.4%
Soc-sign-Slashdot090216	545,671	N	both	N	(-1, 1)	N	N	N	77.39%
Soc-sign-Slashdot081106	516,575	N	both	N	(-1, 1)	N	N	N	76.73%
Bitcoin OTC	35,592	N	both	Y	[-10,10]	N	N	N	89%
Bitcoin Alpha	24,186	N	both	Y	[-10,10]	N	N	N	93%
Y: included N: not included -: not mentioned or considered									

to a machine learning project. It is the key element in deciding whether a machine learning project can succeed or not. This is because a good feature set with many independent features that each correlate well with the result makes the learning process much easier [28]. Moreover, well engineered features can help to avoid over fitting and reduce training time and cost since they allow the use of less complex algorithms that are faster to run and easier to maintain. However, feature extraction is challenging in the sense that it requires domain-specific knowledge. Therefore, feature extraction typically consumes most of the effort in a machine learning task.[9] This is also the case for extracting features to predict a trust relationship. Most of current efforts in the literature have focus on selecting a good enough set of features. The literature often borrows knowledge from social and psychological science to construct features for trust prediction. The knowledge includes, for example, the homophily theory, the structural balance theory (a friend's friend is a friend) and social status theory. These theories are often examined on trust-related datasets first before training a model. The features can be created either from the contextual interaction data between two users or from the structural trust network underlying raw datasets. Features derived from the interaction history represents direct knowledge of a user towards his candidate trustee. However, they might be unavailable for two users who has not involved any kind of interaction yet. Features obtained from the trust network can be the number of incoming trust or distrust, common neighbors and so on. They have the advantage of being able to be applied to various applications without modification. However, they rely heavily on the sparse trust specified explicitly by users.

In the context of OSNs, trust is more closely related to the social trust derived from the discipline of sociology and psychology. "Social trust between two individuals is often studied by examining interaction history, similarity in preferences, background, demographics, reputation or recommendation from third parties, different life experiences and so forth" [10]. Here, we classify the trust features according to the view of both trustor and trustee with considering different context as the following five aspects, namely, trustor's objective factors, trustor's subjective factors, trustee's objective factors, trustee's subjective factors and context [60]. Table 8 shows the detail of trust features.

- *Trustor's objective factors.* It refers to the criteria or policies specified by the trustor for a trust decision such as the trustor's standards, regulations, laws and etc. Normally, it can be specified and formalized into a paradigm.
- *Trustor's subjective factors.* Different from objective factors, trustor's subjective factors rely more on trustor's backgrounds, experience, and personalities. It includes confidence, disposition (also known as propensity which describes how easy a trustor tends to trust others), trustor's dependence and etc.
- *Trustee's objective factors.* It includes reputation, ability, integrity, dependability and so on. Particularly, reputation which is a global measure of each system entity built upon observed evidence is the most often used one.

[9] http://blog.kaggle.com/2013/04/10/qa-with-xavier-conort/

TABLE 8. Trust inducing factors.

Trustor's objective factors	goal; trustor's standards; laws.
Trustor's subjective factors	propensity; Confidence; (subjective) expectations or expectancy; subjective probability; willingness; belief; attitude; feeling; intention; faith; hope; trustor's dependence and reliance.
Trustee's objective factors	reputation; ability; integrity; dependability; predictability; timeliness; behaviors; strength; privacy preservation.
Trustee's subjective factors	honesty; benevolence; goodness; propensity.
Context	situations entailing risk; structural; risk; domain of action; environment (time, place, involved persons); purpose of trust.

- *Trustee's subjective factors.* Honesty, benevolence, goodness and propensity are the main subjective factors of trustee. The propensity of a trustee describes his/her tendency to be trusted. Many studies model it with the average of all the ratings the trustee has received in OSNs.
- *Context.* It describes when, where and why the trust relationship is applied. Notably, the inducing factors could be different and paid different attention by a trustor in different situations and contexts. In this paper, we focus on the context of OSNs where trust predictions are implied mainly for social activities.

It is worth to note that the number of trust features can be numerous. Moreover, these features might be independent, dependent, or conflicting with each other. Thus, it is desirable to examine the correlation among the constructed features before training a model. It is also advisable to consider the availability of features in a dataset. This is because using sparse features would result in a very small training dataset which may result in a low prediction accuracy.

B. MACHINE LEARNING MODELS

In this subsection, we analyze the machine learning models most widely applied in predicting trust. At the same time, it is remarkable that machine learning models cannot achieve the best performance by just combining trust features and standard classifiers without considering optimization techniques. Hence, we also present optimization techniques in terms of imbalance problems, new data integration and over-fitting.

1) MACHINE LEARNING ALGORITHMS

Machine learning algorithms are designed to guide the program to learn from experience. As we have mentioned before, trust prediction has often been modeled as a classification problem in literature. Classification is also the most mature and widely used machine learning algorithm. Thus, we will concentrate on the commonly used classification algorithms in the subsequent discussion.

A classifier maps an input vabriable into a class. Generally, the input variable is represented by a vector of discrete and/or continuous values while the class is typically labeled by a single discrete value [28]. Table 9 lists and compares the most common classification algorithms applied to predict trust in OSNs. These algorithms represent the best-known and most widely used classifiers. Therefore, an overwhelming number of tutorials concerning them exist in the Internet. They have also been extensively studied in academy, thus we will not investigate them in detail.

TABLE 9. Comparison of typical classifiers.

Classifier	Pros	Cons
DT	good at handling discrete or categorical features; no parameters to tune	hard to incorporate new instances; bad performance on imbalanced datasets
K-NN	no parameters to tune	require large storage space; high requirement on feature selection
NB	converge quickly; high accuracy even with a relatively small dataset; robust to missing values; easy to incorporate new instances	only work on discrete features
SVM	its complexity is independent of the number of applied features	require large training datasets
NN	perform better in non-linear classification	feature sensitive; time consuming compared with other models
LR	incorporate new data easily; interpretability; robust	cannot deal with continuous outcomes
Ensemble	can reduce the generalization error; robust, accurate and precise predictions	introduce lots of extra computation

a: DECISION TREE (DT)

Decision trees sort training instances based on their feature values and has no parameters to tune [31]. They are thus good at handling discrete or categorical features. However, incorporating new instances into a built tree needs to rebuild it. In addition, decision tree is often pruned in order to improve generalization and avoid over-fitting. However, pruning can eliminate the leaves belonging to the minority class, thus having detrimental effect for imbalanced datasets [32]. Combining the pruning with sampling can help to mitigate such negative effect.

b: K-NEAREST NEIGHBORS (K-NN)

K-NN is learned by simply storing the training instances. As a result, it requires large storage space. When a dataset is of huge size, which is the case in predicting trust for real life applications, the storage requirement would be too high to be practical. In addition, k-NN is sensitive to irrelevant features [31]. In other words, the presence of an irrelevant feature might lead to false classification. Because of this, k-NN puts high requirement on feature selection.

c: NAIVE BAYES (NB)

NB is reputed to converge quickly and can accomplish high accuracy with a relatively small dataset. In addition, NB has advantage of being robust to missing values and being easily used as incremental learners. However, NB assumes independence among input features, which is difficult to guarantee. It is also inappropriate for datasets to contain too large number of features. Moreover, NB works on discrete features, requiring continuous features to be discretized in most cases [31].

d: SUPPORT VECTOR MACHINE (SVM)

SVM is a non-probabilistic binary classifier that depends on a small number of selected support vectors. Thus, its complexity is independent of the number of applied features. In addition, SVM can accomplish non-linear classification with kernel trick, which transforms inputs into a high-dimensional space. SVM works on continuous features while discrete features pose challenges for it. Generally, SVM requires large training sets to reach its maximum prediction accuracy [31].

e: NEURAL NETWORK (NN)

NN models include Multilayer Perceptron (MLP) and Radical Basis Function (RBF). Their performance is comparable to that of decision trees but seldom better and their training generally takes longer time than decision trees. However, NN models perform better than decision trees in tasks that needs diagonal partitioning or in non-linear classifying. Selected features can affect the training efficiency of NN models adversely. This is because irrelevant features can make NN training very inefficient or even impractical [31].

f: LOGISTIC REGRESSION (LR)

Logistic regression is a probabilistic classifier that allows input features to correlate with each other. Unlike decision trees, logistic regression incorporate new data easily.

g: ENSEMBLE

Ensemble algorithms construct multiple different classifiers separately and the classifying decisions of these classifiers on a new instance are then combined to predict a class for the new instance. This way, ensemble modeling can reduce generalization error and generate more robust, accurate, and precise predictions than its constituent individual members do. As a result, ensemble methods have won many machine learning contests, a prominent example of which is the Netflix Grand Prize [33]. A good example of ensemble approach is random forest that combines multiple decision trees. However, ensemble introduces lots of extra computation.

2) OPTIMIZATION TECHNIQUES

In order to guarantee a high performance of trust prediction task, we propose the following problems and solutions need to be considered when training a machine learning model.

a: IMBALANCE

Trust-related datasets are inherently imbalanced. All the datasets mentioned in section IV that provide both trust and distrust relationships consist of overwhelmingly more trust than distrust. The percentage of positive trust is all above 70%. The most widely used dataset extended Epinions contains as high as 85% percent of trust. This is called class imbalance problem. The problem is inherent with trust-related datasets resulting from the fact that users are generally reluctant to specify distrust relations. Special care should be taken for trust prediction in that it is often modeled as a supervised classification problem that is sensitive to the class imbalanced distribution. The sensitivity arises because most of the classifiers aim to minimize misclassification rate (or increase accuracy) in the model training phase. In this way, all the misclassification error is assumed to have the same cost, which is far from the case in real world applications. In our case, for example, the cost of misclassifying a distrust relationship as trust is more serious than misclassifying a trust as distrust. More specifically, the class imbalance problem decreases the sensitivity of a classifier to the minority class, resulting a poor precision and recall for the minority class. Therefore, we must apply certain data balancing techniques to solve the imbalance problem before the model training process. We summarize herein the most adopted ones of them.

- *Re-sample the original datasets.* This technique creates a balanced dataset by over-sampling the minority class or by under-sampling the majority class. The under-sampling strategy can build a better classifier than the over-sampling method. Sampling can also be performed in a stratified manner, which divides examples into different mutually exclusive strata and then samples equal number of examples from every strata.
- *SMOTE (Synthetic Minority Over-sampling Technique).* SMOTE combines the under-sampling approach with special over-sampling technique that generates synthetic minority samples instead of replicating the minority class [29]. The combination is shown to have better classifying performance (in terms of ROC) than the mere under-sampling approach. In addition, SMOTE can improve the prediction performance of minority class. However, it cannot increase the overall accuracy.
- *Cost sensitive technique.* This approach assigns different costs to different types of misclassification errors [30]. The cost can depend either on the class or on the examples, resulting to class-dependent and example-dependent cost sensitive classifications respectively. Cost sensitive technique can be used to assign higher cost to the error of the minority class and thus force the learning process to model the minority more accurately.

Many cost sensitive variants of the above standard classifiers have been proposed in the literature. Some examples of them are cost-sensitive C4.5 [34], cost-sensitive NB [35], cost-sensitive SVM [36], cost-sensitive MLP [37], cost-sensitive logistic regression [38] and cost-sensitive ensemble of decision trees [39].

Generally, the over-sampling approach increases learning time while under-sampling reduces available training data. Moreover, the sampling methods including SMOTE distorts example distribution, which may seriously affect some classification algorithms. In addition, they are only applicable to the binary classification problems or particular multi-class problems with special cost matrix [30]. Cost sensitive technique can overcome these shortcomings of the sampling approaches.

b: NEW DATA INTEGRATION

As for the dynamic property of training data, an ideal model should be able to be updated to incorporate new data. Two mechanisms exist to accomplish this. One is to re-train the model periodically with bagging algorithms. Another one is to apply incremental learning algorithms that can adapt the parameters of a learning model with new data instances without losing its current knowledge [40]. Model re-training is easy to comprehend and it enables us to test different models easily with different data combinations. The re-training can be triggered manually or automatically as proposed in [41]. However, re-training has to buffer old data, which might become impossible as new data accumulates overtime. On the other hand, incremental learning does not need to store old data. It is also much faster. In addition, incremental learning is ideal for applications where data must be discarded after some time in order to protect privacy. Many incremental versions of above standard classifiers have been proposed. Incremental SVM [42] and NN [43] are some of the examples.

To sum up, the cost-sensitive and incremental version neural networks or SVM is more appropriate for our trust prediction task theoretically.

c: OVER-FITTING

Testing a trained model with the same dataset used in training process is methodological wrong. Since the model learned with such a method would predict poorly on unseen data even though it performs perfectly on training data. This is called over-fitting. However, a good classifier should be able to generalize beyond the training examples. In machine learning, "it is generation that counts" [28]. Therefore, a trained model must be tested on unseen data. To accomplish this, holding out a portion of available data for later testing is a common practice. However, holding out data reduces the amount of available data that can be used in training phase. A solution to combat this problem is called cross validation. Cross validation randomly splits a dataset into n mutually exclusive subsets of approximately equal size, then trains a classifier n times iteratively, and each time uses n-1 subsets for training while the remaining one for testing. The average on the testing results from each of the iteration is then used to describe the classifier performance. Cross validation is particularly useful for tasks with only small number of training examples. A special case of cross validation is the leave-one-out technique that leaves out only one sample for testing and uses all the remaining samples to construct a training dataset. Such a technique can make full use of available data by maximizing the training data most. However, it introduces higher variance and requires more computation resources. Therefore, 5 or 10-fold cross validation is often used.

TABLE 10. The confusion matrix of trust prediction and formulas for common performance metrics.

	Actual positive (trust)	Actual negative (distrust)	Accuracy = $\frac{(TP+TN)}{(TP+FP+FN+TN)}$
			Recall = True Positive Rate (TPR)
			$= \frac{TP}{(TP+FN)}$
			Precision = $\frac{TP}{(TP+FP)}$
			F-measure = $\frac{2Precision*Recall}{(Precision+Recall)}$
			False Positive Rate (FPR)= $\frac{FP}{(TN+FP)}$

Predicted positive (trust)	TP	FP	
Predicted negative (distrust)	FN	TN	
Confusion matrix			Formulas concerning metrics under consideration

C. PERFORMANCE EVALUATION

In this subsection, we discuss the most commonly used performance metrics used in binary classification in the context of trust prediction. But before we get into the detail of performance metrics, we have a quick view with confusion matrix since the definitions of typical performance metrics such as accuracy and recall are derived from it.

1) CONFUSION MATRIX

Table 10 presents the most commonly used confusion matrix in trust prediction. The table also contains the formulas defining the measures under consideration in the subsequent discussion. As shown in the table, a trust binary classifier can produce four types of outcomes, namely, true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Positives and negatives herein represent trust and distrust respectively. Next, we present their interpretations and impacts in real life applications.

- TP. The number of examples that are correctly classified as trust. Correctly classifying as many trust classes as possible can help to improve the quality of trust based services such as trust based recommendation.
- FP. The number of distrusted instances that are incorrectly labeled as trust. Misplacing distrust as trust can lead to catastrophic consequences for some applications. For example, when distrust is misclassified as trust in Epinions, the reviews that are supposed to be blocked for a user are displayed to him. The unwanted reviews

TABLE 11. The confusion matrix of two example classifiers.

	Actual trust	Actual distrust		Actual trust	Actual distrust
Predicted trust	780	180		620	20
Predicted distrust	20	120		180	180
	800	200		800	200
Classifier 1 (accuracy: 0.8, precision: 0.81, recall: 0.98)				Classifier 2 (accuracy: 0.8, precision: 0.97, recall: 0.78)	

might irritate the user so much that he quits the use of Epinions. Such misclassified trust would incur much more serious consequence in applications that involved money. This is because users might be misled to trade with fraudulent users that are misclassified as trust, and thus suffer financial loss.

- TN. The number of examples with distrust relationship that are correctly predicted as distrust. Correctly predicting distrust relationship is crucial for some applications. For example, in Epinions, blocking distrusted users can free users from irritation of low quality or spam reviews. In the case of bitcoin, correctly classifying distrust can help to protect users from trading with others they distrust, preventing users from potential financial loss.
- FN. The number of trusted items that are misclassified as distrust. Misclassifying trust to distrust might result in poor services in the sense that it rules out the useful trust information. In addition, users might be deprived of opportunities to interact or cooperate with others they actually trust.

Based on the above discussion, we argue that FP and TN are more important than TP and FN in trust prediction for real life applications. However, working out a set of formulas to describe their relative gains of correct prediction and cost of incorrect prediction is a daunting task that involves multiple disciplines.

2) PERFORMANCE METRICS

Next, we discuss the most commonly used performance metrics in classification problems in more detail, which includes accuracy, precision, F-measure, ROC, and PR.

a: ACCURACY

Accuracy represents the percentage of correctly classified instances. As shown in the formula, accuracy assumes that TP is as important as TN, which is not the case in trust prediction. For instance, Table 11 shows the confusion matrix of two trust classifiers with the same 80% accuracy. However, classifier 2 clearly predicts distrust much better than classifier 1 and thus it is preferred since the capacity to classify distrust correctly is more valued in the domain of trust prediction. Therefore, accuracy alone cannot differentiate these two classifiers and thus is not suitable for trust prediction.

b: PRECISION/RECALL

Unlike accuracy, precision and recall can differentiate the two classifiers in table 11. For example, both precision and recall can reflect the strength of above two classifiers in predicting trust correctly, displaying their great advantage over accuracy. However, neither precision nor recall can evaluate the capability of above two classifiers in predicting distrust correctly. For instance, the high precision and recall of classifier 1 cannot reflect its poor capacity in recognizing a distrust relationship. Note that this problem can be addressed by utilizing inverse recall and precision where distrust is interpreted as positives and trust as negatives. In this case, however, the precision and recall fail to judge how well a classifier predicts trust. In conclusion, precision and recall can only reflect a classifier's capacity in predicting one particular class. If precision and recall have to be used, we argue that it is better to use them on distrust prediction. It is worth to note that recall and precision often have inverse relationship. Specifically, increasing recall is often achieved at the expense of decreasing precision. Therefore, they are often used together.

c: F-MEASURE

F-measure combines precision and recall into a single metric. F-measure can be used to obtain a trade-off between precision and recall. However, like precision and recall, F-measure cannot reflect how well a classifier handles negative cases [44].

d: RECEIVER OPERATOR CHARACTER (ROC)

ROC curve plots FPR on x-axis and TPR/recall on y-axis. Therefore, it can evaluate a classifier's ability to predict both positive and negative class, thus solving the problems encountered by precision and recall. According to the definition, the point (0, 1) on the ROC graph represents a perfect classifier that can label all the positive instances correctly while make no mistake in classifying negative instances. Classifiers sitting close to the point have high recall while with low FPR and thus are preferred. In addition, the diagonal line on the ROC space represents a randomly guessing classifier, which can conveniently be used as a baseline. Fig. 2 shows a ROC curve depicting the above two classifiers. As shown in the figure, classifier 2 is better than classifier 1 since it classifiers both trust and distrust quite well while classifier 1 performed poorly in predicting distrust. In this sense, ROC can evaluate a classifier better than precision or recall. Generally, using a single scalar quantity to represent and compare the performance of different models is more convenient and thus is desired. In this case, the area under the curve (AUC) is used instead. Moreover, Tom Fawcett recommend to average the ROC curves of multiple test sets when comparing classifiers in order to take into account variance [45].

Note that, a discrete classifier that outputs only a class label is presented as a single point in ROC curve. On the other hand, a probability/score classifier that generates

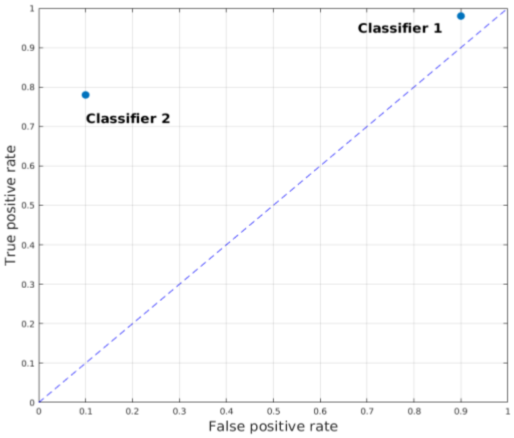

FIGURE 2. The ROC curve of classifier1 and classifier 2.

TABLE 12. The confusion matrix of two classifiers for balanced and imbalanced datasets.

Balanced dataset 1				Imbalanced dataset 2	
	Actual trust	Actual distrust		Actual trust	Actual distrust
Predicted trust	2000	800		2000	100
Predicted distrust	400	1600		400	200
	2400	2400		2400	300
Classifier 3 (precision: 0.71, recall: 0.83, FPR:0.33)				Classifier 4 (precision: 0.95, recall: 0.83, FPR: 0.33)	

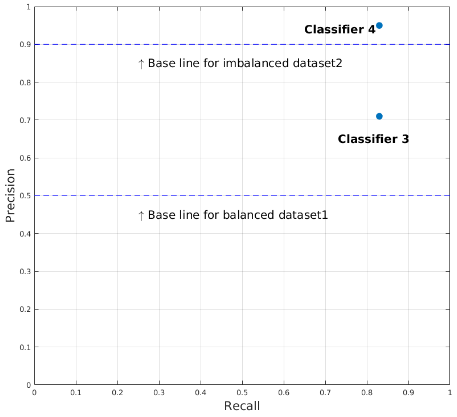

FIGURE 3. The PR curve of classifier 3 and classifier 4.

a probability or score to represent the degree of an example belonging to a class classifies the example by comparing the probability or score with a predefined threshold. As a result, one threshold yields a classifier. Such a classifier with varying thresholds can be depicted as a curve on ROC space [45]. Therefore, ROC has another advantage of being able to visualize the performances of a probability/ scoring classifiers with different thresholds. ROC can thus be also used to select an optimal threshold for a probability/scoring classifier. Due to these attractive properties, ROC curve is the most popular technique to evaluate a binary classifier. However, ROC graph has its own shortcomings. First, ROC cannot be used directly to compare different families of classifiers. This is because the scores of different classifying models might have different scales. Using a common threshold to compare them without proper calibration is thus problematic. Second, discrete classifiers must be converted to scoring versions to generate a full ROC curve. Third, ROC is unsuitable for applications with different error costs (e.g. our trust prediction) unless certain transformation is performed on it. This is because standard ROC curve assumes equivalent cost to different types of errors [45]. In addition, ROC curve is insensitive to class skews and changes in class distribution. In such scenarios, ROC curves could be misleading when it is applied to evaluate and compare different classifiers [46]. For example, ROC curve cannot differentiate classifier 3 and 4 shown in Table 12 despite the fact that classifier 4 with higher precision is more desired.

e: PRECISION-RECALL (PR) CURVE

PR curve plots recall on x-axis and precision on y-axis. Even though it shares many common characteristics with ROC curve, PR curve is different in terms of the following perspectives. First, the optimal classifiers sit on upper right hand corner in the PR space. Second, unlike ROC with fixed base line, the base line of PR curve changes with the ratio of positive instance to negative instance (as shown in Fig. 3). Most importantly, PR curve is perceived superior to ROC curve for tasks with imbalanced datasets since PR plots precision that can capture the poor performance of a classifier for imbalanced datasets. As a result, PR curve has been proposed to be an alternative to ROC curve for tasks with strongly imbalanced datasets [46]. For example, PR curve can present the superior of classifier 4 to classifier 3 which ROC fails to do. Therefore, PR curve is more appropriate for trust prediction. However, a transformation should be performed on it in order to handle different error costs. Moreover, averaging the PR curves across different test sets might provide more solid comparison.

To sum up, an evaluation process must be executed after training a model in order to evaluate the performance of the trained model. In addition, evaluation is needed when comparing several different models to select an optimal one. A poor performance metric that ignores the distribution of classes in a dataset or that neglects some domain-specific aspects of a classifier might give undue credit for some models while fail to give other models enough recognition. In such situations, selecting a proper evaluation metric is of great importance.

VI. LITERATURE REVIEW

In this section, we review the literatures in recent ten years and pay more attention to the works published in recent five years. To retrieve the reviewed articles, we search in all

TABLE 13. Summary of different trust prediction models.

Paper	Features	Classifier	Dataset	Training size	Class distribution	Cross validation	Performance metric
[48]	User factors and interaction factors	SVM_RBF, NB, DT, LR	Epinions	40,462	Imbalanced	5-fold	Precision, recall, F1 of trust
[49]	User factors, WR	SVM_LK	EpinionsVideo, EpinionsTrustlet	-	Balanced	5-fold	F1 of trust
[50]	Product rating, user profile, TN	SVM_LK	@cosme, Extended Epinions	2000	Balanced	-	Precision, recall, F1 of trust
[51]	Trustee reputation	RBF, MLP, SVM_LK	Extended Epinions	210,999	SMOTE	2, 5, 10, 15, 20 fold	Precision, recall of both trust and distrust
[52]	WR	SVM	Extended Epinions	2000	Balanced	5-fold	F1 of trust
[53]	WR, TN	Multilayer NN	Epinions		Balanced	-	Accuracy
[12]	WR, TN, user profile	C5.0 DT, SVM, LR, BN, NN	Epinions_Massa		Balanced	10-fold	Precision, recall, F1 of trust, accuracy, AUC of ROC
[54]	User behavior	Unsupervised	Epinions		Balanced	-	Accuracy
[55]	WR, TN	LR	Extended Epinions	57,626	Balanced	5-fold	Precision, recall, F1
[13]	WR, TN	L2-regularized LR	Extended Epinions	744 + 424	Imbalanced	-	-
[56]	Rating, TN	SVM_RBF	Own collected data	624	-	-	Precision, recall, F1 of trust, accuracy
[16]	-	SVM_RBF	Simulated	-	-	5-fold	Accuracy
[57]	WR	SVM, RF	Epinions	400,000 + 1,600,000 & 2000	Imbalanced & BalancedY	10-fold	Precision, recall, F1 and ROC area
[3]	WR, TN	Ensemble method	?	6000	Balanced	10-fold	Accuracy and ROC curve
[58]	WR, TN	C4.5 DT	Extended Epinions	2000	Balanced & imbalanced	100-fold	Accuracy and F1 measure
[20]	TN	SVM, JRip and J48 DT, RBFN, AdaBoost, NB, MLP	Extended Epinions, Wikipedia	75,760 & 547,694	Balanced & Imbalanced	10-fold	Accuracy, recall, precision of both trust and distrust, F1, ROC area

WR: Review write-rate interaction
TN: Trust Network
LR: Logistic Regression
SVM_LK: SVM with linear kernel
SVM_RBF: SVM with RBF kernel
DT: Decision Tree
NN: Neural Network
BN: Bayesian Network
?: Not sure whether it is Epinions_Massa or Extended Epinions.

authoritative databases, such as IEEE Explorer, ACM library, Springer library, and Science Direct, by using keywords: trust or distrust or trust evaluation or trust prediction or trust relationship or trust mining or trust model or signed ties and social network and machine learning. We classify our review into four categories: classification, regression, ensemble models and others. The summary of different trust prediction models is presented in Table 13.

A. CLASSIFICATION

In the domain of OSNs, trust prediction is usually mapped into a typical classification problem which can be further divided into binary classification and multi-class classification. Normally, unknown relationships are assigned into a trust or distrust which is known as binary classification. While in multi-class classification scenario, it is mapped into fine-grained relationships, namely, trust, neutral trust, distrust and so on.

1) BINARY CLASSIFICATION

Most of the current works model trust prediction as a binary classification problem.

Liu *et al.* [47] conducted the first study that applies machine learning to classify pairwise trust. They proposed to apply supervised learning to predict trust between two users according to the features derived from both their individual actions and their previous interactions which thus guaranteed the adaptability and usability of trust prediction. They also considered inducing factors in a comprehension way. The authors first systematically identified the roles of a user as review writer, review rater and review commenter and the interaction types among different users as write-rate, rate-rate, write-write, write-comment, and comment-comment. In their training, they derived 576 user features and 821 interaction features and applied them in training. The experiment was conducted on dataset crawled from the reviews of videos and DVDs in Epinions. They applied 5-fold cross validation and each fold has about 25% positive instances while 75% negative instances. They evaluated different trained classifiers with precision, recall, F value and 25% precision. The experimented classifiers included decision trees, NB, logistic regression, SVM with linear kernel and SVM with RBF kernel. Considering the number of inducing factors and selected classifiers, they achieved a medium efficiency level. The prediction accuracy was also attained which were above 97% in all cases, sometimes quite close to 100% and they found the better performance of SVM with RBF kernel and NB. As the inherent feature of machine learning, uniformity and generality were preserved. Unfortunately, trustworthiness and privacy were not discussed in this paper.

Ma *et al.* [48] conducted another line of similar study which thus inherited uniformity, comprehension and generality. They also derived features from individual user actions and interactions but only focused on write-rate interacting type. The usability and adaptability were satisfied then. However, due to the lack of distrust in the dataset, they ignored distrust relationship and only focused on classifying trust from non-trust. Instead of training a single classifier with all available data, they trained cluster-specific or personalized classifiers using a subset of data. More specifically, they first grouped similar users into clusters and trained a classifier for each of the clusters. In their personalized classifiers, they trained a classifier for each individual user using data involving the user. However, these two types of classifier introduced high training overhead and the training data might be too small for passive users with little or no interactions at all. Therefore, the authors had to conduct their experiments on a subset of datasets by excluding users without many interactions. In their training, they used the 19 most trust-predictive features obtained from [47] to train SVM classifiers. They trained classifiers for both highly active and non-active users, each of which has 62,851 and 22,163 trust records. F1 was used to evaluate different classifiers. Their experiments showed good prediction accuracy on both cluster-specific and personalized classifiers trained with all available data. This study also used very small data to train their classifiers. It also ignored class imbalanced problem. Moreover, it cannot predict trust for user pairs without write-rate interactions.

The study of Matsuo and Yamamoto [49] is also among the first attempts to predict trust with machine learning. They achieved usability assuming that two users who trust each other have similar product rating behaviors, thus trust and rating exert bidirectional effects. They interpreted similarity between two users in terms of their profiles, product ratings and trust relations. Specifically, they extracted 79 similarity features and used them to train a SVM classifier with linear kernel. The training dataset is composed of randomly selected 1000 pairs of users with trust relations and another 1000 user pairs without any relation representing distrust. They found that product rating features and trust features contributed more to the prediction performance than user profile features. However, most of the 79 extracted features were redundant. For example, the similarity of product rating behaviors between two users were calculated by three different measures, namely matching coefficient, cosine similarity and Jaccard coefficient. Each of the measures represents one feature even though all of them describe the same thing. The redundant features would introduce extra training cost and cause a low efficiency in real life. Moreover, this study ignored the user review interactions upon which the trust relationship is built. The observed prediction accuracy was 80% in average. For the other criteria, the proposed method inherited comprehension, uniformity and generality, but ignored adaptability, trustworthiness and privacy problems.

Graña *et al.* [50] utilized trustee reputation to train classifiers in order to combat trust unpredictability caused by the lack of interaction history. The proposed model fulfilled uniformity, comprehension and generality. Specifically, the reputation feature is represented as a vector composed of trust values about trustee expressed by users who the trustor trusts. The dimension of reputation vectors is fixed by discarding users whose trustor has small number of trusted users or down-sampling users whose trustor has more than specified number of trusted users. Their reputation feature database constructed this way includes 210,999 instances with 90.02% of them belonging to the trust class and thus is highly imbalanced. The authors then trained RBF, MLP and SVM with linear kernel using the constructed reparation database. The trained classifiers were evaluated in terms of the recall and precision of both trust and distrust class. The experiments varied the number of folds in cross validation to imitate system growth. Small number of folds resembled big growth in the future. In this way, the authors explored the resiliency of different classifiers to future system growth. Their experiment showed the high resiliency of SVM. Moreover, their experiments showed the poor recall of distrust class for imbalanced database. Increasing feature vector size was shown to be able to reduce such bias towards the prevalent class. In addition, the author also investigated the influence of SMOTE balancing technique. They concluded that SMOTE could reduce the gap between precisions of trust and distrust class. Moreover, SMOTE can improve the recall of distrust without affecting that of trust class. To summarize, even though this study is the first to study the resiliency of a classifier to future growth and the influence of balancing techniques, the ignorance of some other important trust features such as user review interactions induced the absence of adaptability and usability.

Wang *et al.* [52] combined Dempster-Shafter theory and multilayer neural network to predict both trust and distrust. However, such combination obtained a high prediction accuracy at the cost of efficiency. The study borrowed the concept from sociology and psychology to divide trust and distrust inducing factors into three types, namely, homophily, status theory, and emotion tendency. The study then mapped the quantified inducing factors to an evidence prototype composed of quantitative intervals in order to improve reliability and reduce complexity. The evidences were further processed with a small multilayer neural network (evidence processing unit) in order to deal with potential dependence and conflict among evidences. The processed evidences were then fed into the proposed trust and distrust predicting framework that combined Dempster-Shafter theory and neural network. Before experiments, the study removed user pairs without the proposed features. As a result, only 40% of the labled data (157,838 records) was used for training. Moreover, the study constructed a balanced dataset and employed prediction accuracy as performance metric. The experiments showed that emotion tendency predicted distrust better than predicting trust. The study further applied logistic regression to analyze the feature effectiveness in predicting trust and distrust. In addition, the study conducted several more experiments to compare its proposed scheme with SVM and C5.0 decision tree and other methods. The experiments proved the superiority of the proposed framework to other methods. To sum up, the study has the advantage of using relative large and balanced dataset in terms of accuracy. Moreover, it predicts the important distrust. It is also the only study to apply emotion tendency to predict trust and distrust. However, this study was still short of adaptability and usability without considering prior trust prediction and interaction history when predicting the trust relationship of user pairs.

The study in [12] first identified five main trust inducing factors as knowledge, reputation, relationship, similarity and personality. The five qualitative factors were then quantified and normalized based on the data from user interactions and trust network which complied with usability. But adaptability was understudied without discussion on trust decaying and other dynamically changed factors. The study used Epinions_Massa dataset. However, it sampled only 1000 users containing 54,162 labeled records for experiments in order to overcome limited computing resources. Moreover, the study applied both under-sampling and oversampling method to construct a balanced dataset. The experiments examined multiple classification models such as SVM with RBF kernel, logistic regression, Bayesian network, neural network, decision tree which guaranteed uniformity, comprehension and generality. These models were evaluated in terms of different performance metrics. The observed evaluation presented a high accuracy on the C5.0 decision tree and neural network. In addition, the analysis concluded that knowledge, relationship and similarity factors predict better than the other factors. The results also observed that structural features were better than contextual features in predicting trust. To summarize, the study has the advantage of evaluating models from multiple perspectives. However, it ignored adaptability and sacrificed the efficiency for accuracy with sampling strategy.

Zhao and Pan [55] proposed a trust evaluation framework that based trust on six features of a trustee and three features between a candidate trustor and trustee. The study is the first one to incorporate user profile information into trust prediction. The extraction of features met the requirement of usability but they failed to acquire adaptability since the absence of dynamically changed factors. The framework first collected its training dataset by conducting a survey to ask eight students to assign a trust or distrust label for their Weibo followers. The collected dataset contained 624 user pairs. The nine features were quantified from the dataset. The study trained a SVM with RBF kernel using 516 records which showed a relatively high computation cost compared with other models. Moreover, since the limitation of training data, the accuracy of the trained SVM classifier was relatively low in terms of precision, recall, F1 and accuracy. To sum up, the paper used very small training dataset and did not analyze the ratio between trust and distrust in the dataset. The dataset was unrepresentative since the survey was conducted among only a small number of participants. The study also ignored the performance of distrust prediction.

Korovaiko and Thomo [56] proposed to predict trust by applying user similarity and the interactions between review writers and review raters which, same to the above paper, achieved usability while failed to fulfill adaptability. The study first analyzed the Epinions dataset to point out that rater-reviewer interactions are sparse while similarity is prevalent. To alleviate the problem of sparseness, the paper investigated the similarity in much broader perspectives. Specifically, it measured similarity in terms of overall ratings, high ratings and low ratings, categories, review rating and review writing. After that, the authors proposed a Personalized Trust Prediction model to infer pairwise trust based on the opinions of the closest trustees of a candidate trustor towards his candidate trustee. The model weighted the opinions according to the rater-writer and similarity features of the candidate trustor and his trustees. Moreover, the authors proposed to employ the Kolmogorov-Smirnov test for ranking features based on their trust discriminatory capability which also induced more computation cost. Their ranking results observed that features derived from low ratings are more discriminative compared to those from high ratings. The study constructed two training sets from the original Epinions dataset. The first one was imbalanced and contained 400, 000 trust statements and 1,600, 000 lack-of-trust statements. The second one included 1000 trust and 1000 lack-of-trust. The study then trained Random Forestry (RF) classifiers with 30 J48 decision trees on both datasets. It also trained SVM classifiers but only on the second small dataset. Moreover, the training was conducted on three different feature sets, namely, all 22 features, the top 7 features ranked by Kolmogorov-Smirnov test and 8 features from [51]. The results showed a positive prediction accuracy and the better performance of RF than that of SVM in terms of precision, recall, F1 score and ROC area. To sum up, the study provided many new perspectives to construct trust features and observed a high prediction accuracy. They also held the uniformity, comprehension and generality.

Borzymek *et al.* aimed to investigate the efficiency of a classifier trained with review based attributes in predicting pairwise trust and distrust in social networks and the potential improvement brought by incorporating them into classifiers trained with more universal graph based attributes [57].

Since they only adopted simple factors like review based and graph based attributes other than the time related or subjective opinion of participants, they achieved little usability and adaptability. For experiments, the study created three groups of training sets from the Extended Epinions dataset. Each of the training sets contained 2000 records with different proportion of trust and distrust or requirement of historical interaction records. The authors trained three C4.5 decision trees on each training set with review-based attributes, graph attributes and the combined attributes. The results observed improvement of combined attributes in prediction accuracy, F measure of trust and distrust. Moreover, the paper investigated the prediction of important distrust. However, the training sets were very small. In addition, the authors only examined the decision trees.

2) MULTI-CLASS CLASSIFICATION

López and Maag [16] proposed a generic framework to enable machine learning trust models to collect trust information and exchange trust evaluations. The trust management framework employed the RESTful web-service architecture in order to entitle a wide range of devices. The proposed framework is noteworthy for two main reasons. First, the framework enabled a trustor to express at what context to evaluate his candidate trustees since the study assumed that trust features vary from trustor to trustor and from context to context. Second, the framework allowed trustors to decide how to interact with their trustees. The proposed method in this paper can resist whitewashing attacks by setting a forgetting factor which is a function that factors newer interactions and places more importance on newer interactions, and respectively lower importance in older interactions. But it is still vulnerable to other attacks. Moreover, the framework incorporated the training data collecting process, which most of the trust prediction models have ignored. Specifically, the framework obtained its labeled training datasets by collecting trust features first and then annotating them by the system administers or developers. It included both adaptability and usability related factors. The framework also included a trust evaluating engine, which modeled trust prediction as a multi-class problem. Specifically, the study chose to use SVM classifiers with medium efficiency. To summarize, the study has proposed the first multi-class trust prediction model using machine learning. Moreover, the paper provided a framework to communicate trust features and trust values within an application. However, the study conducted experiments with simulated data containing only two trust features in a single context. Moreover, the study provided no statistics about the simulated dataset. In addition, the authors labeled the simulated features themselves, which is not scientific. This is because trust is a subjective concept and thus is more credible to be annotated by the candidate trustors, which can be achieved by adding a survey model to the framework. All of these makes it difficult to validate the prediction accuracy as high as 96.61% of the proposed model claimed in the paper. Inherently, it fulfilled the requirement of uniformity, comprehension, and generality. However, the privacy problem was not included.

B. REGRESSION

Fang *et al.* [13] proposed a trust and distrust framework to predict continuous trust and distrust. The framework took into consideration both interpersonal and impersonal factors of trust and distrust. The interpersonal factors contained four features of a trustee: benevolence, competence, integrity and predictability. All of them were modeled from user rating data. On the other hand, the impersonal factors were modeled from trust network and included the trust in-degree, distrust in-degree, trust out-degree, distrust out-degree of a trustee. Since they only considered the trustee factor, they provided partially usability. The study applied the quantified eight features to train two logistic regression models to predict trust and distrust in a comprehensive way, respectively. It is worth to note that the framework used the expected probability that a user trusts another to represent the predicted trust. Therefore, the predicted trust and distrust is in the form of continuous values. The study used three datasets, namely, Extended Epinions, FilmTrust and Flixter. From the Extended Epinions dataset, two subsets were sampled. One of the sampled subset contained 744 trust and 424 distrust. Another one included 3443 trust and 1398 distrust. The experiments trained L2-regularized logistic regression on the two sampled subsets. However, the FilmTrust and Flixter dataset were only used for validation due to their lack of distrust information. The experimenting results showed a high efficiency and prediction accuracy. The study then applied the predicted trust and distrust to refine trust network by removing unreliable trust or distrust links which further improved the adaptability of the model. In addition, the study applied the newly predicted trust to improve recommendation systems. To sum up, the study is the first one to investigate continuous trust and distrust. However, it incorporated only trust features about trustees and ignored those of trustors. It also assumed that trust features correlate linearly with trust and distrust, which might not meet the criteria of uniformity and generality. In addition, the study assumed that trust features are independent with each other and thus is incapable to capture the possible dependency among the trust features. Moreover, the training datasets are very small and thus are not representative.

Another work proposed by Ma *et al.* [54] also adopted a regularized logistic regression model to predict trust for recommendation systems. Similarly, this model achieved high efficiency and comprehension. They absorbed three trust factors including personal, interpersonal and impersonal factors and the personal factors which referred to trustor bias/propensity and trustee bias guaranteed usability. However, they ignored adaptability features. These features were then used to train a regularized logistic regression model. The training used a subset of data down-sampled from the Extended Epinions dataset due to limited computing resources. The down-sampled dataset contained 306,773 trust records and 28,813 distrust records. The authors further constructed a balanced dataset containing 28,813 trust and distrust to combat the imbalance problem. The experiments were conducted on both imbalanced and balanced datasets. The results showed that the trustor and trustee bias and similarity correlated with trust most closely and positively. However, the impersonal factors had little influence on trust prediction. The experiments also observed that the model trained with all three factors achieved the best prediction performance in terms of precision, recall and F1. Additionally, the authors applied the predicted trust to the trust based recommendation systems, and confirmed that predicted trust could work as well as the explicit trust specified by users.

C. ENSEMBLE MODELS

Zolfaghar and Aghaie [3] developed a framework to predict both trust and distrust with machine learning. The study employed four trust-inducing factors including knowledge, reputation, similarity and personality. The authors then derived eight quantitative features for these four factors from both trust network and contextual information. As for the dataset, the authors claimed to have used Epinions_Massa [18]. However, the statistics about the dataset listed in this paper were inconsistent with Epinions_Massa. Despite pre-processing, the trust statement would not be more than that of Epinions_Massa. In addition, Epinions_Massa contained only trust statements while this study claimed to include also distrust statements, which is very confusing. We observe that the statistics are more consistent with that of Extended Epinions dataset. However, without confirmation from authors, we cannot assert this observation. The study applied an ensemble approach that employed a voting mechanism to aggregate 5 different families of classifiers, namely, Radial Basis Function Neural network (RBFN), J48 decision tree, Naive Bayes, logistic regression and SVM. The authors created a balanced training set from the original dataset with 3000 trust pairs and 3000 distrust pairs. Although The experiments observed higher computation cost, the ensemble method outperformed each of its constituent classifiers in prediction capacity. They also ranked the eight features according to their prediction capacity. Results showed higher ranks of structural features than that of contextual features.

D. OTHERS

Nguyen *et al.* [51] applied trust antecedent framework to model trust prediction for online rating systems. Trust antecedent framework identified four factors to predict trust between two users which are the ability, benevolence and integrity of a trustee and the trust propensity of a trustor. Usability is obtained by the selected factors. However, adaptability related factors were neglected. The framework assumed that a user is trusted if s/he is thought by another user to have the ability to deliver expected outcome, to want to do good with the trustor (benevolence), and to adhere to a set of good moral principles (integrity). The trust propensity describes how easy a trustor to trust others in general. The study quantified each of the four qualitative factors into measurable features from review write-rate interacting data. More specifically, the ability of a trustee is quantified by the average rating received from his candidate trustor and the number of his review rated by his candidate trustor. The trustee benevolence is measured by his leniency, which is quantified by the relative difference between the review ratings of his candidate trustor and the actual quality of the reviews. The trustee integrity is modeled with his global trustworthiness measured by the number of users that trust him. In addition, the trust propensity of a trustor is measured by the global leniency s/he shows to his or her trustees and the number of user s/he trusts. With these derived features, the study combined them differently, resulting into eight different models. The eight models were evaluated with F1 score on a dataset containing 1000 trust and 1000 distrust pairs randomly selected from the whole Extended Epinions dataset. In addition, the authors also combined the most important features identified by [47] and their eight features to train a SVM classifier. The training used 5-fold cross validation. However, the result showed a rather limited prediction accuracy which according to author can be explained by the poor capturing ability of trust propensity. For the other criteria, this paper still held the uniformity, comprehension and generality.

Yang *et al.* [53] proposed to use unsupervised or semi-supervised algorithm to predict signed social ties such as trust or distrust for the unsigned social relationships such as acquaintance. Such newly available signed social ties can capture richer information than traditional social relationships since acquainted users do not necessarily trust each other. Predicted signed social ties can add a new meaningful dimension to the acquaintance relationship, enabling service providers to improve their service. The authors first empirically validated the correlation between user behaviors and the signs of social relationship. Confirmed by the validation, the authors designed unsupervised behavior relation interplay (BRI) model and examined it on a balanced dataset down-sampled from the Epinions dataset. However, the exact number of records of the balanced dataset was not provided in their paper. The results showed that the BRI model achieved prediction accuracy higher than 70%. In addition, the authors designed the semi-supervised version of BRI, which made use of a small amount of data with labeled social ties. The experiments observed that the prediction accuracy increased as more labeled data was used in training. Moreover, the authors further extended their BRI model to encode structural balance theory and social status theory. The experiments observed that social status theory improved BRI more than the structural balance theory. In particular, BRI with social status can achieve accuracy comparable to the supervised leave-one-out logistic regression method. The study also demonstrated that the predicted signed social ties could predict user behaviors better than Homophily. To sum up, the unsupervised and semi-supervised methods have the advantage of uniformity and generality applicable to a wide range of OSNs. Still, trustworthiness and adaptability were not concluded.

The study in [20] proposed to predict trust based on reputation in the context of social networks. The study first extracted reputation vectors of fixed size from the Extended Epinions dataset and Wikipedia dataset. However, the extracted vectors were limited in adaptability and usability since the limitation of data information. Besides, in order to obtain reputation vectors of fixed size, instances with less witness must be discarded while instances with more witness than the specified size must be down-sampled. However, discarding instances wastes the valuable available data. Thus, as the main contribution of this paper, the authors proposed to use the conditional probabilistic of the reputation sets with variable size to obtain a fixed size feature set. The two training feature datasets extracted from the Extended Epinions and Wikipedia datasets contained 75,760 and 547,694 instances, respectively. Both of them are imbalanced. The Wikipedia feature set included 84.45% of trust and 15.55% of distrust while the Epinions feature set contained 89% of trust and 11% of distrust. For the experiments, the study has examined eight different machine learning algorithms such as SVM and decision trees. Thus, it inherited with uniformity, comprehension and generality. The results observed poor recall and precision for distrust class. Therefore, the authors employed the SMOTE technique to balance the feature sets, which was shown to have improved the precision and recall of distrust. Moreover, the experiments showed that the probabilistic reputation feature exhibited strong discriminatory power. The authors have also examined the resiliency of the trained classifiers to the continuous growing size of the social network. To sum up, the study has provided new methods to construct reputation features. Moreover, the paper has reported the performance of distrust prediction. However, extracting reputation feature from the sparse trust graph might be problematic. It is preferred to combine such feature with other features derived from richer information such as user interactions.

VII. DISCUSSIONS AND FUTURE DIRECTIONS

Section VI demonstrates that current studies have investigated many aspects of trust prediction using machine learning. They show that applying machine learning technique to predict trust is effective and promising. However, the studies leave some problems unaddressed. We summarize them herein and propose some potential directions for solving them.

A. OPEN ISSUES AND CHALLENGES

In order to find current problems, then possible to propose new research directions, we use the proposed criteria to analyze and measure existing work as shown in Table 14. The detail of criteria is listed in section III-C. In the table, Y represents that the property is satisfied. N represents the method cannot satisfy the property. Y(partially) represents that the trust prediction method can satisfy the property partially. Blank represents the property was not mentioned or considered in the article. H, M, L represents high, medium and low, respectively. Based on the literature review and evaluation according to the proposed criteria, we find a number of open issues in machine learning based trust prediction in OSNs.

First, we can conclude that only one reviewed paper [16] discussed the robustness of trust prediction to overcome potential attacks. The proposed method in this paper can resist whitewashing attacks by setting a forgetting factor which is a function that factors newer interactions and places more importance on newer interactions, and respectively lower importance in older interactions. But it is still vulnerable to other attacks.

Second, adaptability is seldom discussed except [16], [47], [48]. Modeling the important dynamic property of trust would be important since trust relationships are inherently dynamically changed due to the leaving and joining of different users, trust decay over time or the prior trust experience in updating trust. However, no information about the trust building and destroying time is publicly available. Therefore, this also poses a big challenge for modeling trust dynamicity with crawled datasets.

Third, no effort has concentrated on privacy preservation in the existing studies in this field based on our literature review. User privacy should be concerned when their private data like relationships, rating, review opinions are collected by a central party since the private information like hobby, religion or health status can be induced from user's behavior data.

Forth, most of the current works model trust prediction as a binary classification problem. Until now, only one of the studies treats trust prediction as a multi-class problem [16] and only two studies consider trust as a continuous value [13], [54]. However, we have argued that measuring trust with a multi-class metric or a continuous value is more fine-grained and thus more expressive. Therefore, it would be beneficial to exanimate the regression algorithms to predict trust with a continuous value.

Fifth, no solution is holistic by satisfying all criteria. As illustrated in Table 14, none of them can satisfy all criteria and provide an ideal trust prediction method.

B. FUTURE RESEARCH DIRECTIONS

All above open issues motivate future research. We further suggest a number of promising research directions about machine learning based trust prediction in OSNs as below.

1) **Robustness trust prediction.** Potential attacks and malicious behaviors should be identified in trust prediction. The pairwise trsut prediction models should be robust to resist various potential attacks when the participants are socializing with each other in OSNs. In addition, the trust prediction models should also have the ability to identify the type of attack or malicious behaviors since it is very helpful to optimize the model.

TABLE 14. Comparasion of different trust prediction models.

Paper	Trustworthiness	Adaptability	Usability	Privacy	Accuracy	Efficiency	Uniformity	Comprehension	Generality
[48]		Y	Y		H	M	Y	Y	Y
[49]		Y	Y		H	L	Y	Y	Y
[50]		N	Y		M	M	Y	Y	Y
[51]		N	N		L	L	Y	Y	Y
[52]		N	Y		L	M	Y	Y	Y
[53]		N	N		H	L	Y	Y	Y
[12]		N	Y		H	M	Y	Y	Y
[54]		N	Y		M	M	Y	Y	Y
[55]		N	Y		H	H	Y	Y	Y
[13]		N	Y(partially)		H	H	N	Y	N
[56]		N	Y		L	M	Y	Y	Y
[16]	Y	Y	Y		H	M	Y	Y	Y
[57]		N	Y		H	M	Y	Y	Y
[3]		N	Y		H	L	Y	Y	Y
[58]		N	N		L	M	Y	Y	Y
[20]		N	N		L	M	Y	Y	Y

2) **Context-awareness.** Context-awareness should be concerned in trust prediction with adaptability. The trust relationship is inherently dynamic and social networks normally hold such such properties as mobility and complexity, contributing to the frequent changes of trust. Thus, trust prediction should be context-aware. Therefore, the trust prediction mechanism in OSNs should be aware of the background and requirements of a social task in order to take appropriate action. For example, users of Facebook put an emphasize on interests of trustee while users of Epinion may more interest in the knowledge reputation of trustee. At the same time, factors like times and previous experience should also be considered when building a secure and reliable trust prediction model.

3) **Privacy preservation.** Privacy should be preserved and enhanced. Nowadays, people pay more attention to their personal privacy especially in OSNs, because the uncertainty of the other side, which could be a stranger in many online application scenarios. Normally, cryptography and data fusion are two main approaches to achieve the privacy of machine learning based prediction, but at the same time, they can result in the loss of efficiency or accuracy, which need to be trade-off.

4) **Fine-grained prediction.** Fine-grained trust prediction should be provided. The fine-grained definition of trust provides different trust levels to help the trustor to accurately decide how to interact with the trustee. Compared with the trust or distrust selection, multilevel trust gives a trustor more signals to make a decision. Based on the fact that a user usually cannot make a distinct decision about a stranger when lack of background information, the requirement on fine-grained trust prediction is highly expected in practice.

5) **Comprehension.** A holistic trust prediction model is expected. The trust prediction model should fully meet all or most of the criteria. Various kinds of potential attacks should be reliably detected and identified. The prediction is convinced only when it is context aware and adapatable to factors such as time, subjective policies and previous experiences. For predicting trust, user privacy should be protected when collecting user data. In addition, a fine-grained trust prediction is preferred. In a word, a holistic model should be the ultimate pursuit in future investigations.

VIII. CONCLUSION

In this paper, we reviewed and analyzed trust prediction with machine learning techniques in OSNs from many perspectives. Specifically, we summarized and discussed the pipeline of trust predicting models. We analyzed the feature extracting process and recognized its crucial role in predicting trust. Moreover, we explored different trust-related datasets and reviewed their contents and statistics, figuring out their characteristic properties of being imbalanced. The exploration on the datasets has also fixed some inconsistencies about them in different papers. In addition, we discussed different families of machine learners and presented their cost-sensitive and incremental variants that can help fighting against class imbalance and continuous system growth problems. The important performance metrics were also discussed before we walked through existing trust prediction models and enumerated their pros and cons. Machine learning has gained high attention in recent years. The first attempts to apply machine learning to predict trust have been proved effective and promising. However, we found a number of open research issues based on our survey and tried to propose a number of future research directions in this field. With growing interests in trust prediction with machine learning, we believe there will be much progress in this area.

REFERENCES

[1] A. Mohaisen, N. Hopper, and Y. Kim, "Keep your friends close: Incorporating trust into social network-based sybil defenses," in *Proc. INFOCOM*, vol. 11, Apr. 2011, pp. 336–340.

[2] W. Sherchan, S. Nepal, and C. Paris, "A survey of trust in social networks," *ACM Comput. Surv.*, vol. 45, no. 4, Aug. 2013, Art. no. 47.

[3] K. Zolfaghar and A. Aghaie, "Mining trust and distrust relationships in social Web applications," in *Proc. Intell. Comput. Commun. Process. (ICCP)*, Aug. 2010, pp. 73–80.

[4] Y. Ruan and A. Durresi, "A survey of trust management systems for online social communities—Trust modeling, trust inference and attacks," *Knowl.-Based Syst.*, vol. 106, pp. 150–163, Aug. 2016.

[5] Z. Yan, X. Jing, and W. Pedrycz, "Fusing and mining opinions for reputation generation," *Inf. Fusion*, vol. 36, pp. 172–184, Jul. 2017, doi: 10.1016/j.inffus.2016.11.011.

[6] Z. Yan, Y. Chen, and Y. Shen, "PerContRep: A practical reputation system for pervasive content services," *J. Supercomput.*, vol. 70, no. 3, pp. 1051–1074, 2014.

[7] Z. Yan, Y. Chen, and Y. Shen, "A practical reputation system for pervasive social chatting," *J. Comput. Syst. Sci.*, vol. 79, no. 5, pp. 556–572, 2013, doi: 10.1016/j.jcss.2012.11.003.

[8] Z. Yan, P. Zhang, and R. H. Deng, "TruBeRepec: A trust-behavior-based reputation and recommender system for mobile applications," *J. Pers. Ubiquitous Comput.*, vol. 16, no. 5, pp. 485–506, Jun. 2012.

[9] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decis. Support Syst.*, vol. 43, no. 2, pp. 618–644, Mar. 2007.

[10] J.-H. Cho, K. Chan, and S. Adali, "A survey on trust modeling," *ACM Comput. Surv.*, vol. 48, no. 2, Nov. 2015, Art. no. 28.

[11] J. Tang, H. Gao, and H. Liu, "mTrust: Discerning multi-faceted trust in a connected world," in *Proc. 5th ACM Int. Conf. Web Search Data Mining*, Feb. 2012, pp. 93–102.

[12] K. Zolfaghar and A. Aghaie, "A syntactical approach for interpersonal trust prediction in social web applications: Combining contextual and structural data," *Knowl.-Based Syst.*, vol. 26, pp. 93–102, Feb. 2012.

[13] H. Fang, G. Guo, and J. Zhang, "Multi-faceted trust and distrust prediction for recommender systems," *Decis. Support Syst.*, vol. 71, pp. 37–47, Mar. 2015.

[14] S. P. Marsh, "Formalising trust as a computational concept," Ph.D. dissertation, Dept. Math. Comput. Sci., Stirling Univ., Stirling, Scotland, 1994.

[15] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[16] J. López, and S. Maag, "Towards a generic trust management framework using a machine-learning-based trust model," in *Proc. Trustcom/BigDataSE/ISPA*, vol. 1, Aug. 2015, pp. 1343–1348.

[17] J. Golbeck, "Trust and nuanced profile similarity in online social networks," *ACM Trans. Web*, vol. 3, no. 4, 2009, Art. no. 12.

[18] G. Guo, J. Zhang, and N. Yorke-Smith, "A novel Bayesian similarity measure for recommender systems," in *Proc. IJCAI*, Aug. 2013, pp. 2619–2625.

[19] E. Khadangi and A. Bagheri, "Comparing MLP, SVM and KNN for predicting trust between users in Facebook," in *Proc. Comput. Knowl. Eng. (ICCKE)*, Oct./Nov. 2013, pp. 466–470.

[20] J. D. Nuñez-Gonzalez, M. Graña, and B. Apolloni, "Reputation features for trust prediction in social networks," *Neurocomputing*, vol. 166, pp. 1–7, Oct. 2015.

[21] P. Massa and P. Avesani, "Trust-aware recommender systems," in *Proc. ACM Conf. Rec. Syst.*, Dec. 2007, pp. 17–24.

[22] G. Guo, J. Zhang, D. Thalmann, and N. Yorke-Smith, "ETAF: An extended trust antecedents framework for trust prediction," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2014, pp. 540–547.

[23] R. Zafarani and H. Liu. (2009). *Social Computing Data Repository at ASU*. [Online]. Available: http://socialcomputing.asu.edu

[24] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proc. 19th Int. Conf. World Wide Web*, Apr. 2010, pp. 641–650.

[25] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2010, pp. 1361–1370.

[26] R. West, H. S. Paskov, J. Leskovec, and C. Potts. (Sep. 2014). "Exploiting social network structure for person-to-person sentiment analysis." [Online]. Available: https://arxiv.org/abs/1409.2450

[27] S. Kumar, F. Spezzano, V. S. Subrahmanian, and C. Faloutsos, "Edge weight prediction in weighted signed networks," in *Proc. Data Mining (ICDM)*, Dec. 2016, pp. 221–230.

[28] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[29] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[30] P. Domingos, "MetaCost: A general method for making classifiers cost-sensitive," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 1999, pp. 155–164.

[31] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, pp. 3–24, Jun. 2007.

[32] N. V. Chawla, "C4. 5 and imbalanced data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure," in *Proc. ICML*, vol. 3, Aug. 2003, p. 66.

[33] Y. Koren, "The BellKor solution to the Netflix grand prize," *Netflix Prize Document.*, vol. 81, pp. 1–10, Aug. 2009.

[34] K. M. Ting, "An instance-weighting method to induce cost-sensitive trees," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 3, pp. 659–665, May 2002.

[35] A. Ibáñez, C. Bielza, and P. Larrañaga, "Cost-sensitive selective Naïve Bayes classifiers for predicting the increase of the *h*-index for scientific journals," *Neurocomputing*, vol. 135, pp. 42–52, Jul. 2014.

[36] P. Cao, D. Zhao, and O. Zaiane, "An optimized cost-sensitive SVM for imbalanced data learning," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, Apr. 2013, pp. 280–292.

[37] C. L. Castro and A. P. Braga, "Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 888–899, Jun. 2013.

[38] A. C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive logistic regression for credit scoring," in *Proc. Mach. Learn. Appl. (ICMLA)*, Dec. 2014, pp. 263–269.

[39] B. Krawczyk, M. Woźniak, and G. Schaefer, "Cost-sensitive decision tree ensembles for effective imbalanced classification," *Appl. Soft Comput.*, vol. 14, pp. 554–562, Jan. 2014.

[40] A. Gepperth and B. Hammer, "Incremental learning algorithms and applications," in *Proc. Eur. Symp. Artif. Neural Netw. (ESANN)*, 2016, pp. 357–368.

[41] A. Bifet, G. Holmes, B. Pfahringer, and R. Gavalda, "Improving adaptive bagging methods for evolving data streams," in *Proc. Adv. Mach. Learn.*, 2009, pp. 23–37.

[42] G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," in *Proc. Adv. Neural Inf. Syst.*, 2001, pp. 409–415.

[43] R. Polikar, L. Upda, S. S. Upda, and V. Honavar, "Learn++: An incremental learning algorithm for supervised neural networks," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 31, no. 4, pp. 497–508, Nov. 2001.

[44] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.

[45] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers," *Mach. Learn.*, vol. 31, no. 1, pp. 1–38, 2004.

[46] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, p. e0118432, Mar. 2015.

[47] H. Liu *et al.*, "Predicting trusts among users of online communities: An Epinions case study," in *Proc. 9th ACM Conf. Electron. Commerce*, Jul. 2008, pp. 310–319.

[48] N. Ma, E.-P. Lim, V.-A. Nguyen, A. Sun, and H. Liu, "Trust relationship prediction using online product review data," in *Proc. 1st ACM Int. Workshop Complex Netw. Meet Inf. Knowl. Manage.*, Nov. 2009, pp. 47–54.

[49] Y. Matsuo and H. Yamamoto, "Community gravity: Measuring bidirectional effects by trust and rating on online social networks," in *Proc. 18th Int. Conf. World Wide Web*, Apr. 2009, pp. 751–760.

[50] M. Graña, J. D. Nuñez-Gonzalez, L. Ozaeta, and A. Kamińska-Chuchmała, "Experiments of trust prediction in social networks by artificial neural networks," *Cybern. Syst.*, vol. 46, nos. 1–2, pp. 19–34, Feb. 2015.

[51] V.-A. Nguyen, E.-P. Lim, J. Jiang, and A. Sun, "To trust or not to trust? Predicting online trusts using trust antecedent framework," in *Proc. 9th IEEE Int. Conf. Data Mining*, Dec. 2009, pp. 896–901.

[52] X. Wang, Y. Wang, and H. Sun, "Exploring the combination of Dempster-Shafer theory and neural network for predicting trust and distrust," *Comput. Intell. Neurosci.*, vol. 2016, Jan. 2016, Art. no. 23.

[53] S. H. Yang, A. J. Smola, B. Long, H. Zha, and Y. Chang, "Friend or frenemy?: Predicting signed ties in social networks," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2012, pp. 555–564.

[54] X. Ma, H. Lu, and Z. Gan, "Implicit trust and distrust prediction for recommender systems," in *Proc. Int. Conf. Web Inf. Syst. Eng.*, Nov. 2015, pp. 185–199.

[55] K. Zhao and L. Pan, "A machine learning based trust evaluation framework for online social networks," in *Proc. IEEE 13th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Sep. 2014, pp. 69–74.

[56] N. Korovaiko and A. Thomo, "Trust prediction from user-item ratings," *Social Netw. Anal. Mining*, vol. 3, no. 3, pp. 749–759, 2013.

[57] P. Borzymek and M. Sydow, "Trust and distrust prediction in social network with combined graphical and review-based attributes," in *Proc. KES Int. Symp. Agent Multi-Agent Syst., Technol. Appl.* Berlin, Germany: Springer, Jun. 2010, pp. 122–131.

[58] S. Shehnepoor, M. Salehi, R. Farahbakhsh, and N. Crespi, "NetSpam: A network-based spam detection framework for reviews in online social media," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1585–1595, Jul. 2017.

[59] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proc. 49th Annu. Meeting Assoc. for Comp. Linguistics, Hum. Lang. Technol.*, vol. 1, Jun. 2011, pp. 142–150.

[60] Z. Yan and S. Holtmanns, "Trust modeling and management: From social trust to digital trust," in *Proc. Comput. Secur., Privacy Politics, Current Issues, Challenges Solutions*, 2008, pp. 290–323.

[61] I. Pranata, G. Skinner, and R. Athauda, "A survey on the usability and effectiveness of Web-based trust rating systems," in *Proc. IEEE/ACIS 12th Int. Conf. Comput. Inf. Sci. (ICIS)*, Jun. 2013, pp. 455–460.

[62] Y. Zhu and Z. Yan, "A survey on trust evaluation in e-commerce," in *Proc. 9th EAI Int. Conf. Mobile Multimedia Commun.*, 2016, pp. 130–139.

[63] G. Xu and Z. Yan, "A survey on trust evaluation in mobile ad hoc networks," in *Proc. 9th EAI Int. Conf. Mobile Multimedia Commun.*, 2016, pp. 140–148.

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

[65] P. Chomphoosang, P. Zhang, A. Durresi, and L. Barolli, "Survey of trust based communications in social networks," in *Proc. 14th Int. Conf. Netw.-Based Inf. Syst. (NBiS)*, Sep. 2011, pp. 663–666.

[66] W. Abdelghani, C. A. Zayani, I. Amous, and F. Sèdes, "Trust management in social Internet of Things: A survey," in *Proc. Conf. e-Bus., e-Services e-Soc.* Cham, Switzerland: Springer, 2016, pp. 430–441.

[67] K. Hoffman, D. Zage, and C. Nita-Rotaru, "A survey of attack and defense techniques for reputation systems," *ACM Comput. Surv.*, vol. 42, no. 1, 2009, Art. no. 1.

[68] J. Golbeck, "Introduction to computing with social trust," in *Computing With Social Trust*. London, U.K.: Springer, 2009, pp. 1–5.

SHUSHU LIU received the B.Sc. and M.Sc. degrees in computer science from Soochow University, Suzhou, China, in 2014 and 2017, respectively. She is currently pursuing the Ph.D. degree with the Department of Communication and Networking, Aalto University, Espoo, Finland. Her research interests are in social network, machine learning, and data security and privacy.

LIFANG ZHANG received the B.Sc. degree in electrical engineering from Beijing Forestry University, Beijing, China, in 2012, and the M.Sc. degree from the Department of Communication and Networking, Aalto University, Espoo, Finland, in 2016. Her research interests are in network security and data privacy.

ZHENG YAN (M'06–SM'14) received the B.Eng. degree in electrical engineering and the M.Eng. degree in computer science and engineering from Xi'an Jiaotong University, Xi'an, China, in 1994 and 1997, respectively, the second M.Eng. degree in information security from the National University of Singapore, Singapore, in 2000, and the Licentiate of Science and the Doctor of Science in Technology in electrical engineering from the Helsinki University of Technology, Helsinki, Finland, in 2005 and 2007, respectively. She is currently a Professor at Xidian University, Xi'an, China, and a Visiting Professor and a Finnish Academy Research Fellow at Aalto University, Espoo, Finland. Her research interests are in trust, security and privacy, security-related data analytics. She served as a Steering Committee Co-Chair for the IEEE Blockchain Conference and a general/program chair for over 20 international conferences. She serves as an Associate Editor of *Information Sciences*, *Information Fusion*, the IEEE Internet of Things Journal, the IEEE Access Journal, JNCA, and *Security and Communication Networks*.