Financial Defaulter Detection on Online Credit Payment via Multi-view Attributed Heterogeneous Information Network

Qiwei Zhong
Alibaba Group
Hangzhou, China
yunwei.zqw@alibaba-inc.com

Yang Liu*
Institute of Computing Technology, Chinese Academy of Sciences
Beijing, China
liuyang17z@ict.ac.cn

Xiang Ao†*
Institute of Computing Technology, Chinese Academy of Sciences
Beijing, China
aoxiang@ict.ac.cn

Binbin Hu
Ant Financial Services Group
Hangzhou, China
bin.hbb@antfin.com

Jinghua Feng
Alibaba Group
Hangzhou, China
jinghua.fengjh@alibaba-inc.com

Jiayu Tang†
Alibaba Group
Hangzhou, China
jiayu.tangjy@alibaba-inc.com

Qing He*
Institute of Computing Technology, Chinese Academy of Sciences
Beijing, China
heqing@ict.ac.cn

ABSTRACT

Default user detection plays one of the backbones in credit risk forecasting and management. It aims at, given a set of corresponding features, e.g., patterns extracted from trading behaviors, predicting the polarity indicating whether a user will fail to make required payments in the future. Recent efforts attempted to incorporate attributed heterogeneous information network (AHIN) for extracting complex interactive features of users and achieved remarkable success on discovering specific default users such as fraud, cash-out users, etc. In this paper, we consider default users, a more general concept in credit risk, and propose a multi-view attributed heterogeneous information network based approach coined MAHINDER to remedy the special challenges. First, multiple views of user behaviors are adopted to learn personal profile due to the endogenous aspect of financial default. Second, local behavioral patterns are specifically modeled since financial default is adversarial and accumulated. With the real datasets contained 1.38 million users on Alibaba platform, we investigate the effectiveness of MAHINDER, and the experimental results exhibit the proposed approach is able to improve AUC over 2.8% and Recall@Precision=0.1 over 13.1% compared with the state-of-the-art methods. Meanwhile, MAHINDER has as good interpretability as tree-based methods like GBDT, which buoys the deployment in online platforms.

*Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS). Also at University of Chinese Academy of Sciences, China.
†Corresponding author.

WWW '20, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7023-3/20/04.
https://doi.org/10.1145/3366423.3380159

KEYWORDS

Financial defaulter detection; Multi-view attributed heterogeneous information network; Meta-path encoder

ACM Reference Format:
Qiwei Zhong, Yang Liu, Xiang Ao, Binbin Hu, Jinghua Feng, Jiayu Tang, and Qing He. 2020. Financial Defaulter Detection on Online Credit Payment via Multi-view Attributed Heterogeneous Information Network. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3366423.3380159

1 INTRODUCTION

With the globalization of the digital economy, internet financial institutions are incessant to gather. WeChat Pay, as one of the biggest mobile payment platforms, announced it has more than 1 billion daily active personal users in February 2019. Its Chinese competitor Alipay also claimed to have surpassed the 1 billion mark this year[1]. These platforms are meanwhile serving millions of enterprise users. Such a huge number of active users facilitates the emergence of online credit payment services produced by these platforms. That is, users are able to sign an electronic contract with a platform and choose to pay by their credit pledges. The credit pledge will be cleared after the user repays the money within the time stipulated in the contract.

Analogous to traditional credit card services in commercial banks, default user detection [12, 13, 19, 22–24, 32, 35] is at the heart of risk forecasting and management for online credit payment services. The purpose of default user detection is to predict whether a user[2] will fail to make required payments in the future. Hence this problem is generally formulated as a binary classification problem, and a classifier such as tree-based model or neural

[1] https://www.merchantsavvy.co.uk/mobile-payment-stats-trends/
[2] In this paper, a user could either refer to a personal user or an enterprise user. We will use the term "user" to denote both of them unless otherwise specified hereafter.

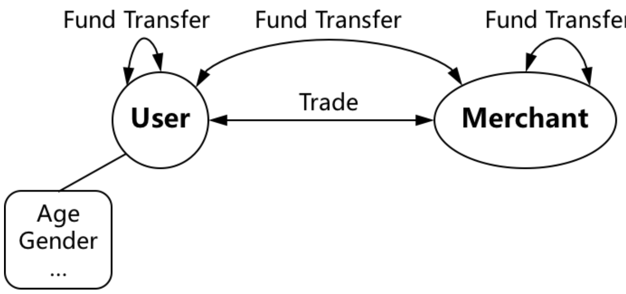


Figure 1: A toy example of AHIN schema.

network is trained by feeding user-related features through cumbersome feature engineering. Conventional approaches may extract user-related features from various aspects, e.g., user profiles [13], personal credit history [35] , trading transactions [2, 3, 24], and social activities [19], etc.

Recently, a few research efforts exploiting the interactive relations between users have been shown effective on detecting specific anomalies (e.g., fraud, cash-out, malicious users) over financial behaviors [13, 19, 32]. In these works, transactions are modeled in an attributed heterogeneous information network (AHIN) where users, merchants, and etc. are treated as nodes in the network. For example, Figure 1 illustrates a toy example of AHIN schema in which users and merchants are denoted by different types of nodes. These nodes have various interactions, such as trading and fund transfer, which are represented as links in the network. Meanwhile, the nodes may have additional attributes for users, e.g., age, gender, etc. Then the default user detection can be formulated as a node classification problem on AHIN. Under such setting, besides the traditional statistical features of a user (a node in the AHIN), more complex features could be extracted through its neighbors matched by pre-defined meta-paths (a relation sequence connecting two nodes) [13] or aggregation of node neighbors [32].

Despite the remarkable successes of existing researches on specific abnormal user detection, the more general financial default user detection on online credit payment service is still underexplored. The specialized challenges for this problem are summarized as follows.

- *Endogeny.* Financial default is not completely driven by objective factors. Users could be subjectively reluctant to afford when they raise the debt, which derives to the endogenous factors for the defaulters. However, capturing these endogenous aspects of default users is challenging and a more accurate user profile modeling approach is in urgent demand.
- *Adversary.* Some illegal economic activities such as cash-out are highly related to financial default [13]. The suspicious criminals may deliberately construct complex behaviors, e.g., transferring money among a lot of users and trying to make the fund transfer path as long as possible, to avoid regulation and gloss up the evidences. Hence, how to perceive fine-grained behavioral patterns is becoming an intractable problem.
- *Accumulation.* Meanwhile, not all the defaulters are malicious. For example, some enterprise defaulters may result from the vicious circle impacted by upstream or downstream enterprises. In another word, financial default has its accumulated and transitive features on local structures, which gives a natural motivation to exploit interactive relations among multiple users.

To remedy these challenges, we propose a novel approach coined MAHINDER (**M**ulti-view **A**ttributed **H**eterogeneous **I**nformation **N**etwork based financial **DE**fault use**R** detection) in this paper.

First, for better user profile modeling to the prediction task, multiple user relationships are simultaneously utilized, and they constitute a *multi-view attributed heterogeneous information network* (MAHIN for short). Figure 2 exhibits an example of the MAHIN in our considered scenarios. In this figure, besides containing different types of nodes, the MAHIN covers different kinds of user relationships, i.e., *social view* denoting social connections, *fund view* describing trading and fund transfer behaviors, and *device view* presenting user-device login operations. Meanwhile, both nodes and links in MAHIN may have attributes such as "age" and "asset" for node "Tom" and "time" and "amount" for link "trade". We define meta-paths on these views to extract interactive features of users. Different meta-paths might have distinct weights to decide personalized characteristics. Hence an attention mechanism is adopted to de-emphasize irrelevant views during training, further improving user profile modeling.

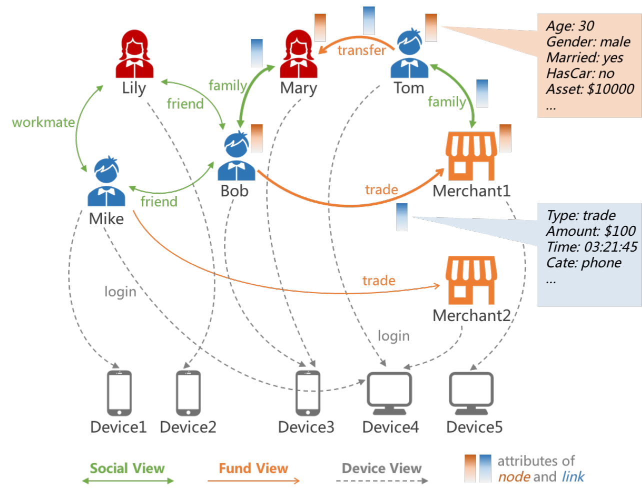


Figure 2: An MAHIN example.

Second, to mine fine-grained behavioral patterns, MAHINDER carefully takes the local structures of assigned meta-paths into account and devises a *meta-path based path encoder* to encode attributes on both nodes and links. This component integrates two LSTM variants to model node and link sequences consisting of their respective attributes, and the hidden representations are distilled by attribute attentions on both node and link levels. Through this way, fine-grained behavioral patterns on both levels could be simultaneously extracted. Finally, all the features are aggregated together and a softmax classifier is trained to identify defaulters.

The contributions of this work are summarized as follows.

- We are the very first attempt to consider financial default user detection over online credit payment service, which is crucial for risk forecasting and management in internet financial institutions.
- We formulate the problem as a node classification problem over multi-view attributed heterogeneous information network and propose a novel model MAHINDER to solve it. This method employs meta-path over different views to thoroughly model user profile and devises a meta-path based path encoder to capture local structural patterns on both nodes and links. Attention mechanisms are adopted on node, link and meta-path levels to automatically learn the importance of attributes and data views.
- Experiments on a real-world dataset from Alibaba demonstrates the effectiveness of MAHINDER. It can not only outperform the compared state-of-the-art approaches but also have as good interpretability as tree-based methods.

The remainder of the this paper is organized as follows. Section 2 surveys the related researches in the literature. Section 3 introduces definitions and the problem statement of this paper. Section 4 details the proposed MAHINDER approach. Section 5 illustrates the experiments and Section 6 concludes the paper.

2 RELATED WORK

In this section, we will review the related studies from two aspects, namely financial defaulter detection and heterogeneous information network.

2.1 Financial Defaulter Detection

Default user detection is at the heart of credit risk forecasting and management for traditional credit card services in commercial banks and emerging online credit payment services. Extensive literature from economic and data mining fields have been studied on specific types of default, such as fraud [12, 22–24, 32], cash-out [13, 18, 39], and malicious [19], etc.

Conventional methods regard each user as an individual and extract statistical features from different aspects like user profiles and historical behaviors, and detect specific default by supervised learning approaches like logistic regression, support vector machines, tree-based methods or neural networks [4, 22, 23, 35, 39]. For instance, Bhattacharyya et al. [4] analyzed various aspects of user attributes elaborately such as categorical attributes, transaction types, etc., and derived 16 important attributes (e.g., average amount over 3 months, total number of transactions with the same merchant during last month, etc.) to evaluate logistic regression, support vector machines and random forests for fraud detection. Zhang et al. [39] collected more than 5000 original features from four different aspects, namely the seller features, the buyer features, the transaction features and the historical features, and used distributed deep forest model on an extra-large scale dataset with more than 100 millions of training samples to automatically detect cash-out users.

Users may have rich interactions with each other in financial scenarios. A few of recent works start to utilize the graph-based methods to model complex interactive patterns of users [13, 18, 19, 32]. For example, Li et al. [18] investigated the detection of fraudulent cash-out by building a semi-supervised learning algorithm that automatically tunes the prior and parameters in Markov Random Field while inferring labels for every node in the graph. Liu et al. [19] proposed a graph neural network to learn discriminative embeddings from heterogeneous account-device graphs based on two fundamental weaknesses of attackers (i.e., device aggregation and activity aggregation) for malicious account detection. For the heterogeneous graph consisting of various types of nodes, they proposed an attention mechanism to learn the importance of different types of nodes, while using the sum operator for modeling the aggregation patterns of nodes in each type. Hu et al. [13] proposed a meta-path based graph embedding method for user cash-out prediction via employing meta-path based neighbors to exploit user attributes and structural information. While in this work, we consider a more general scenario for defaulter detection in online credit payment service. Multiple views of user interactive behaviors are modeled as a multi-view attributed heterogeneous information network. A meta-path based path encoder is devised to learn both node and link attributes as user preference representations, and attention mechanisms are adopted on both attribute and view levels, which makes our work distinct to the existing ones.

2.2 Heterogeneous Information Network

In the past couple of years, many works have been devoted to mining heterogeneous information network [26] for different applications such as classification [14, 40], clustering [29], relation inference [36, 38], recommendation [25], and anomaly detection [19]. Li et al. [17] firstly considered attributes of nodes and extended HIN to attributed heterogeneous information network (AHIN) to enrich node information, which further emerges subsequent researches over AHIN. For example, Li et al. [17] proposed a dynamic AHIN embedding framework to support downstream machine learning tasks. Hu et al. [13] proposed to detect cash-out users as a node classification problem in AHIN. While in this work, we aim to detect default users based on multi-view attributed heterogeneous information network, which has not been explored before.

Meta-path, a relation sequence connecting two nodes, is widely explored in mining HIN. To name some, Sun et al. [28] first proposed the concept of meta-path in HIN and PathSim was used to evaluate the similarity of same-typed nodes based on symmetric paths. After that, many research works [11, 34] extended PathSim by incorporating richer information such as transitive similarity, temporal dynamics and supportive attributes. Meng et al. [20] studied how to discover meta-paths automatically and investigated how to generate meta-paths that can best explain the relationship between these node pairs.

Meta-paths have also been used for node embeddings in HIN. It aims at learning a low-dimensional and dense representation for nodes in HIN. Metapath2vec [6] formalized meta-path based random walks to construct the heterogeneous neighborhood of a node and then leveraged a heterogeneous skip-gram model to perform node embeddings. HIN2Vec [8] carried out multiple prediction training tasks jointly based on a target set of relationships to learn latent vectors of nodes and meta-paths in the HIN.

With the advancement of deep learning, graph neural network [10, 16, 30] has gained a lot of attention, and many methods [33, 37] based on graph neural network for HIN have also been developed. For instance, HetGNN [37] is a heterogeneous graph neural network model and is able to capture both structure and content heterogeneity. HAN [33] proposed both node-level and semantic-level attentions to learn the importance of nodes and meta-paths.

3 PROBLEM STATEMENT

In this section, we present the problem statement for financial defaulter detection. Prior to that, we overview several concepts which are helpful for problem statement.

Definition 1. View-specific Network. Given a user relationship v as a view, a view-specific network is denoted as $\mathcal{G}^v = \{\mathcal{V}^v, \mathcal{E}^v, X_{\mathcal{V}^v}, X_{\mathcal{E}^v}\}$. Here, $\mathcal{V}^v = \mathcal{U} \cup \mathcal{S}^v$ where $\mathcal{U}$ is the users set we need to classify, $\mathcal{S}^v$ and $\mathcal{E}^v$ denote view-specific nodes and links, respectively. $X_{\mathcal{V}^v} \in \mathbb{R}^{|\mathcal{V}^v|\times d_1}$ and $X_{\mathcal{E}^v} \in \mathbb{R}^{|\mathcal{E}^v|\times d_2}$ are respective attribute matrix for nodes and links in $\mathcal{G}^v$, where d_1, d_2 indicate the dimensions of attributes, respectively.

For example, as shown in Figure 2, if $\mathcal{U} = \{Bob, Tom\}$ and considering the social-specific network, $\mathcal{S}^{social}$ should be the social-related user set of $\mathcal{U}$, i.e., $\mathcal{S}^{social} = \{Mike, Lily, Mary, Merchant1\}$, and $\mathcal{E}^{social}$ is their relations set, i.e., $\mathcal{E}^{social} = \{Bob \xleftrightarrow{family} Mary, Bob \xleftrightarrow{friend} Mike, \ldots\}$. The attributes for the node Tom, e.g. "Age", "Gender" and etc, are stored in $X_{\mathcal{V}^{social}}$.

With the concept of view-specific network, we further extend it to multi-view attributed heterogeneous information network to integrate richer information about the relations of nodes and links.

Definition 2. Multi-view Attributed Heterogeneous Information Network. A multi-view attributed heterogeneous information network (MAHIN), denoted as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, X_{\mathcal{V}}, X_{\mathcal{E}}\}$, consists of an m-views node set $\mathcal{V} = \mathcal{U} \cup \mathcal{S}^{v_1} \cup \cdots \cup \mathcal{S}^{v_m}$ with an attribute information matrix $X_{\mathcal{V}} \in \mathbb{R}^{|\mathcal{V}|\times d_1}$, and the corresponding link set $\mathcal{E} = \mathcal{E}^{v_1} \cup \cdots \cup \mathcal{E}^{v_m}$ with an attribute information matrix $X_{\mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|\times d_2}$. An MAHIN is also associated with a node type mapping function $\phi : \mathcal{V} \to \mathcal{A}$ and a link type mapping function $\psi : \mathcal{E} \to \mathcal{R}$. $\mathcal{A}$ and $\mathcal{R}$ denote the sets of predefined node types and link types, where $|\mathcal{A}| + |\mathcal{R}| > 2$.

For example, Figure 2 exhibits an example MAHIN with 3 views, namely *social*, *device*, and *fund* views with corresponding attributes over nodes and links. It is worth noting that the concept of AHIN is a special case of our definition, which only considers individual view and the attribute information of nodes.

In MAHIN, the concept of meta-path [27] is utilized to model the interaction between two nodes. That is, two nodes can be connected via various different semantic paths, which are called meta-paths.

Definition 3. Meta-path. A meta-path $\mathcal{P}$ is denoted as a path in the form of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} A_{l+1}$, which describes a composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_l$ between nodes A_1 and A_{l+1}, where $\circ$ denotes the composition operator on relations. We further denote the node sequence of $\mathcal{P}$, i.e. $A_1 \to A_2 \to \cdots \to A_{l+1}$, as the node sub-path, which is denoted as $\mathcal{P}^o$, and the link sequence of $\mathcal{P}$, i.e. $R_1 \to R_2 \to \cdots \to R_l$, as the link sub-path, denoted as $\mathcal{P}^l$.

For example, if meta-path $\mathcal{P}$ is $User \xrightarrow{fund} User \xrightarrow{social} User$, a path matched based on it could be $Bob \xrightarrow{trade} Merchant1 \xrightarrow{family} Tom$ as the MAHIN shown in Figure 2. The node sub-path is then *Bob* $\to$ *Merchant1* $\to$ *Tom* and the corresponding link sub-path is *trade* $\to$ *family*.

Finally we present the problem statement of financial defaulter detection with the concepts mentioned above.

Problem Statement of Financial Defaulter Detection. Given an MAHIN $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, X_{\mathcal{V}}, X_{\mathcal{E}}\}$ consisting of m specific views, our purpose is to detect defaulters from the target user set $\mathcal{U} \subset \mathcal{V}$. We assign a label $y_u \in \{0, 1\}$ on each user $u \in \mathcal{U}$ to indicate whether he/she is a defaulter or not. Hence, given the MAHIN $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, X_{\mathcal{V}}, X_{\mathcal{E}}\}$ and the training set $\mathcal{D} = \{(u, y_u)\}$, our financial defaulter detection problem is to predict the default probability p_v of user v in the testing set.

4 THE MAHINDER MODEL

In this section, we present the proposed MAHINDER model. Firstly, meta-paths across different views are collected to describe multiple aspects of structural information in MAHIN. Then, we devise a novel neural component to capture interactions among nodes and links by encoding their corresponding attributes. Furthermore, considering that different nodes, links and meta-paths may have different importance, a series of attention mechanisms are designed to model user preferences towards them. The overall architecture of the proposed model is shown as Figure 3.

4.1 Meta-path on MAHIN

In this subsection, we introduce the meta-paths used in our model. To better motivate these meta-paths, we first demonstrate some statistical analysis on the three different views (i.e., social, device, and fund view) of our experimental dataset in this paper. The results are shown as Figure 4. For each view, we count the number of neighbors who are defaulters (called default neighbor), and divide all users into three distinct groups with respect to the number of their default neighbors (shown #Default Neighbors in Figure 4). Figure 4 (a) demonstrates the probability of defaulters in each group[3]. The lifting percentages of probability of default in different user groups against users without any default neighbor (i.e., #Default Neighbors=0) are presented in Figure 4 (b). Furthermore, for social view, the probability of defaulters for different relations when #Default Neighbors>0 is illustrated as Figure 4 (c). From these three figures, we have the following observations.

(1) Macroscopically, different views have different impacts on users, yielding different probability of defaulters and lifting percentages within the same group as shown in Figure 4 (a) and (b). It inspires us that different views may have different importance on users, and cross-view meta-paths might be more effective for extracting informative features for users since they integrate multi-view of data simultaneously.

(2) Microscopically, as shown in Figure 4 (c), we observe that different types of relations also have different impacts. It implies that links may play a quite important role but used to be ignored inadvertently. Combined with the adversarial scenario, it urges us

[3]Note that the probability of default is computed within each group, which derives that three bars in the same view do not necessarily add up to 100%.

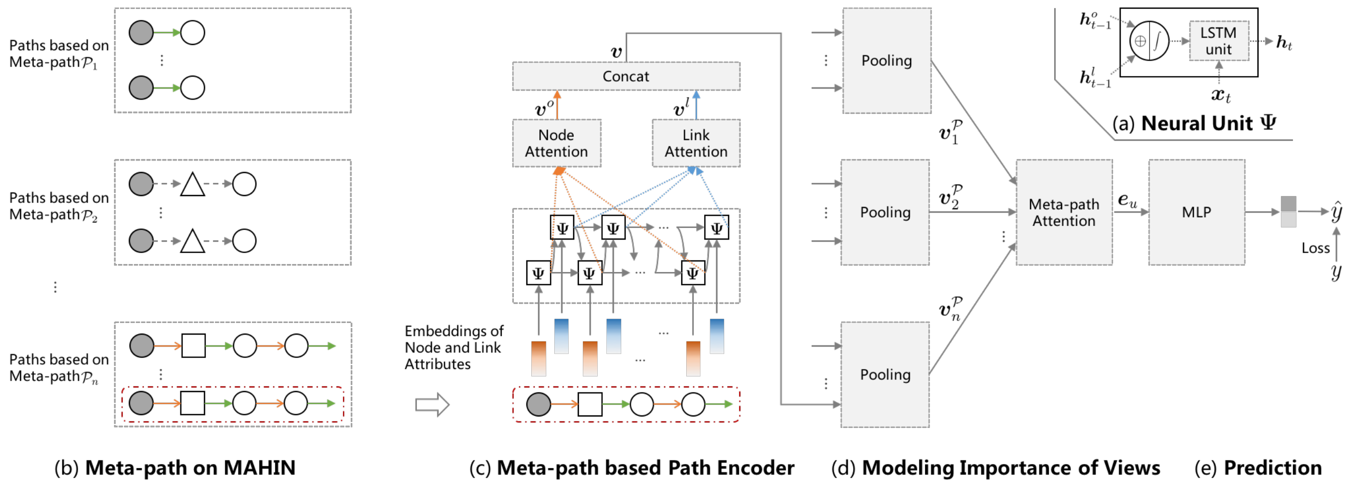


Figure 3: The architecture of the proposed MAHINDER.

to elaborately explore the modeling of links and collect deeper/high-order paths of nodes and links rather than only the first-orders.

Inspired by the observations above, we collect paths of nodes and links for a user based on pre-defined meta-paths which are discriminative on detecting financial defaulters, as follows.

(1) intra-view meta-paths:

- UsU: $User \xrightarrow{social} User$
- UdU: $User \xrightarrow{device} User$
- UfU: $User \xrightarrow{fund} User$
- UsUsU: $User \xrightarrow{social} User \xrightarrow{social} User$
- UfUfU: $User \xrightarrow{fund} User \xrightarrow{fund} User$

(2) cross-view meta-paths:

- UdUsU: $User \xrightarrow{device} User \xrightarrow{social} User$
- UfUsU: $User \xrightarrow{fund} User \xrightarrow{social} User$
- UfUsUfU: $User \xrightarrow{fund} User \xrightarrow{social} User \xrightarrow{fund} User$

These meta-paths are devised from the perspective of financial default detection and can reflect some typical default phenomenon. For example, fraud users may register many accounts via social-related ID cards and login them on limited devices, which combines the social and device relationships and derives the meta-path “UdUsU”. Cash-out users may transfer money among a lot of nodes and try to make the fund path as deep as possible, which derives meta-paths like “UfUfU” and “UfUsUfU”, etc. Among the devised meta-paths, the first five meta-paths model the respective views, and the rests are cross-view paths that integrate different relations among distinct views.

Path instances on MAHIN are sampled according the defined meta-paths as the user features. All the paths start from the target users we need to classify. In another word, the first node of a sampled path is the target user. For each meta-path $\mathcal{P}$, a frequency threshold $\theta^{\mathcal{P}}$ is specified to restrict the quantity of sampled paths for one user, which is set to 100 in our implementation.

4.2 Meta-path based Path Encoder

Unlike existing methods [13, 33] that incorporate meta-path aggregated neighbors as node features, our MAHINDER elaborately models both the node sub-path and the link sub-path by a LSTM-based path encoder. The encoder reflects both the independence and dependence of the two aspects in a simple yet effective manner.

The encoder models dual sequences namely the node sub-path and the link sub-path. A single neural unit in our path encoder is denoted as $\boldsymbol{\Psi}$, which is shown as Figure 3 (a). Two hidden vectors $\boldsymbol{h}^o_{t-1} \in \mathbb{R}^{d_1}$, $\boldsymbol{h}^l_{t-1} \in \mathbb{R}^{d_2}$ from node sequence and link sequence, and an input vector $\boldsymbol{x}_t$ (the attributed representations of a node or a link, which are jointly learnt with the model) are fed into $\boldsymbol{\Psi}$ at each time step t. $\oplus$ shown in Figure 3 (a) is a concatenation which is defined as

$$\boldsymbol{h}'_{t-1} = [\alpha \boldsymbol{h}^o_{t-1}; (1-\alpha)\boldsymbol{h}^l_{t-1}] \tag{1}$$

where α is a coefficient, which can be manually set up or automatically trained, to control the importance of the two hidden inputs, and $\boldsymbol{h}'_{t-1} \in \mathbb{R}^{d_1+d_2}$. The module $\int$ in our neural unit shown in Figure 3 (a) is defined as

$$\boldsymbol{h}_{t-1} = \sigma(\boldsymbol{W}_\phi \boldsymbol{h}'_{t-1} + \boldsymbol{b}_\phi) \tag{2}$$

where σ is the element-wise sigmoid activation function, $\boldsymbol{W}_\phi$ is the weight parameter, $\boldsymbol{b}_\phi$ is the bias term. $\boldsymbol{h}_{t-1}$ and $\boldsymbol{x}_t$ are fed into vanilla LSTM unit to generate hidden vector $\boldsymbol{h}_t$.

Intuitively, not all the nodes/links of a path have the same importance. We thus adopt attention mechanisms to capture user preferences towards nodes/links. Specifically, we define the node attention mechanism to merge $\boldsymbol{h}^o_1, \boldsymbol{h}^o_2, \cdots, \boldsymbol{h}^o_L$ into a distilled representation as follows, where L is the path length.

$$\boldsymbol{z}^o_i = \tanh(\boldsymbol{W}_a \boldsymbol{h}^o_i) \tag{3}$$

$$a^o_i = \frac{\exp(\boldsymbol{q}_a^{\mathrm{T}} \boldsymbol{z}^o_i)}{\sum_{j=1}^{L} \exp(\boldsymbol{q}_a^{\mathrm{T}} \boldsymbol{z}^o_j)} \tag{4}$$

$$\boldsymbol{v}^o = \sum_{i=1}^{L} a^o_i \boldsymbol{h}^o_i \tag{5}$$

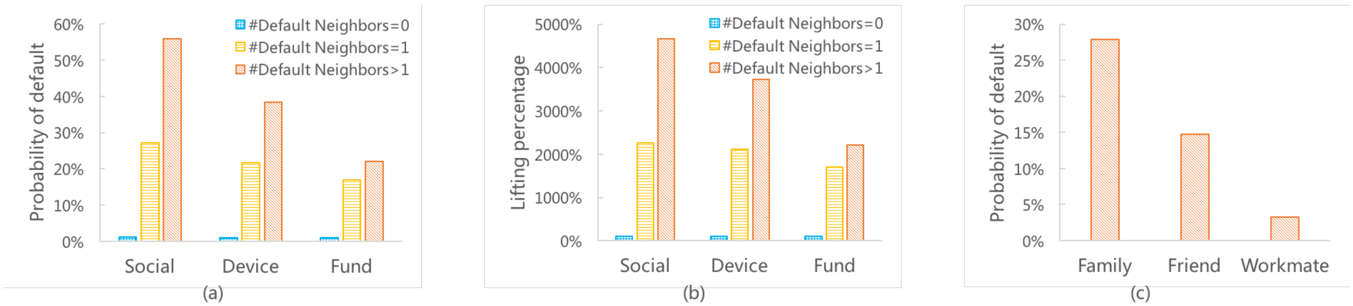


Figure 4: The statistical data analysis on the defaulters on MAHIN of the experimental dataset in this paper. (a) and (b): The probability and the lift of defaulters according to the number of defaulters in their direct neighbors over different views. (c): The probability of defaulters under different link types within a same view.

where W_a is a trainable weight matrix, q_a is a random initialized vector and a^o is the attention weights over nodes. v^o is the distilled representation of the node sub-path.

The distilled representation of the link sub-path v^l can be obtained in a very similar way, and we thus omit the details. The final output by a path encoder is the concatenated vector of v^o and v^l, denoted as v, which is continuously utilized by the subsequent computations.

4.3 Modeling Importance of Views

Recall that we sample at most $\theta^{\mathcal{P}}$ for each pre-defined meta-path $\mathcal{P}$. Hence we can obtain $\theta^{\mathcal{P}}$ distilled representations of sampled paths. We aggregate them by an average pooling and derive an overall representation for meta-path $\mathcal{P}$, denoted as $v^{\mathcal{P}}$. Such representation encodes specific features of users from corresponding meta-paths.

Different kind of users (such as fraud user, cash-out user, etc.) are likely to have diverse preferences over meta-paths from multiple views. We thus devise a meta-path attention mechanism to capture user preferences towards meta-paths.

In more detail, we define the preliminary attention mechanism to merge all representations of meta-paths $v_1^{\mathcal{P}}, v_2^{\mathcal{P}}, \cdots, v_n^{\mathcal{P}}$ into a distilled representation as follows.

$$z_i' = \tanh(W_a' v_i^{\mathcal{P}}) \tag{6}$$

$$a_i' = \frac{\exp({q_a'}^{\mathrm{T}} {z_i}')}{\sum_{j=1}^{n} \exp({q_a'}^{\mathrm{T}} {z_j}')} \tag{7}$$

$$e_u = \sum_{i=1}^{n} a_i' v_i^{\mathcal{P}} \tag{8}$$

where ${W_a}'$ is a trainable weight matrix, q_a' is a random initialized vector and a' denote the attention weights over different meta-paths. The output vector e_u is the final user representation generated by MAHINDER, and it will be used by the classifier for prediction.

4.4 Model Training

After modeling the complex interactions above, the obtained final representation e_u is fed into multiple fully connected neural networks and a regression layer with a sigmoid unit, as follows:

$$z_u = ReLU(W_L \cdots ReLU(W_1 e_u + b_1) + b_L) \tag{9}$$

$$p_u = \sigma(w_p^T z_u + b_p) \tag{10}$$

where, W_* and b_* respectively denote the weight matrix and the bias vector for each layer, $ReLU(\cdot)$ is the element-wise rectified linear unit function [21], and w_p and b_p are the weight vector and the bias, respectively. Here p_u is the predicted default probability of user u.

Finally, our model is trained with cross entropy loss with regularization. The loss function is defined as

$$\mathcal{L}(\Theta) = \sum_{\langle u, y_u \rangle \in \mathcal{D}} (y_u \log(p_u) + (1 - y_u) \log(1 - p_u)) + \lambda \|\Theta\|_2^2 \tag{11}$$

where y_u is the ground truth, Θ is the parameter set of the proposed model, and λ is the regularizer parameter.

4.5 Discussion

In this subsection, we discuss the advantages of MAHINDER compared with the recent alternatives and discuss the scalability of our method.

The proposed MAHINDER is a flexible framework to leverage rich attribute and structural information over multi-view heterogeneous information network. Unlike existing methods [13, 33] which only consider meta-path based neighbors as the user context, our approach adopts meta-path based path encoder to model interactive behaviors among the target user to its first-order and high-order neighbors. As a result, not only neighbors' profile but also the relationships to the target user are simultaneously considered in our method. Meanwhile, our adopted meta-paths may either be within the same view or cross different views since defaulters may have various complex reasons failing to pay the requirements. However, most of existing works on specific default detection merely take meta-path within the same view even though the data simultaneously contains multiple views.

Besides, compared to traditional network embedding methods [5, 31] which represent nodes via simple adjacency matrix, our method is also more suitable for large-scale networks, benefited from reducing the dimension of original input from $O(|\mathcal{V}|)$ to $O(\mathcal{D})$, where $|\mathcal{V}|$ is the total number of nodes in network and $\mathcal{D}$ is the feature dimension after discretization for each user ($\mathcal{D} \ll |\mathcal{V}|$). Additionally, for a cold-start user which never appears in training set, the proposed model can efficiently learn the representation through his/her meta-path based paths in networks when he/she has some links in the network.

5 EXPERIMENTS

In this section, we investigate the effectiveness of the proposed model. We conduct extensive experiments on a large-scale real-world dataset. First, we verify the prediction accuracy on detecting defaulters from the dataset. Next, we perform ablation test to demonstrate the effectiveness of every component in our model. Then we look closer to the data and demonstrate the interpretability of our method by case studies.

5.1 Dataset

We collect a real-world dataset from an online credit payment service provided by Alibaba Group. It contains 1.38 million users (ranging from 2019/01/01 to 2019/03/31) for training and 0.51 million users (ranging from 2019/05/01 to 2019/05/31) for testing, chronologically. In the testing set, there are 20.58% unseen target users. According to the general definition in financial area, we define the positive samples as users who default within one month and the negative samples as users who do not default within one month. It is noteworthy that the time interval between training and testing set should not be less than one month since we need the data in the next month when defining the label. The positive rate is around 0.5% in the dataset. The data statistical information is exhibited in Table 1.

Table 1: The statistical information of dataset.

Dataset	#Positive	#Negative	#Total	#Positive Rate
Training	6,950	1,374,355	1,381,305	0.503%
Testing	2,522	511,116	513,638	0.491%

Based on the dataset, we construct a multi-view attributed heterogeneous information network under the premise of complying with security and privacy policies. Three views are adopted here, namely social view, fund view and device view, as shown in the MAHIN example in Figure 2. The MAHIN altogether contains 14.98 million nodes (four different types) and 168.86 million links (six different types composed of the relationships from the three views). For attributes of nodes and links, we collect 100 attributes for each user (node), such as user profile, credit information, purchasing behaviors, and asset information, etc. For each relation (link), we construct 45 attributes such as link type (e.g., family, trade, login, etc.), first/last related time, amount, interaction frequency, etc. Table 2 exhibits the corresponding statistical information.

5.2 Compared Methods

We compare with several state-of-the-art and representative methods including tree-based methods and (A)HIN embedding methods to verify the effectiveness of our proposed method. Two kinds of strategies are applied to HIN embedding based baselines for fair comparisons. In the first strategy, the interactive features of a target user is obtained by the meta-path defined in [13], which is to extract meta-path based neighbors by first-order proximity. The second strategy is to extract interactive features of a target user following the meta-paths defined in our paper. Hence high-order proximity of users are considered in this strategy. However, since the baselines could not model the attributes on links, we still use the meta-path based neighbors to aggregate features of a target user's neighbors.

- **GBDT** [7]: a scalable tree-based model for feature learning and classification task, and widely used in various areas.
- **DeepForest** [39, 42]: a deep model based on decision trees holding three characteristics, i.e., layer-by-layer processing, in-model feature transformation and sufficient model complexity. We feed only attribute features of a target user into the tree-based methods, i.e., GBDT and DeepForest.
- **HAN** [33]: a semi-supervised graph neural network which employs node-level attention and semantic-level attention simultaneously. We use **HAN** to denote the variant implementation of the first strategy mentioned above, and $\textbf{HAN}_{S2}$ denotes the second.
- **HACUD** [13]: a heterogeneous network method which employs meta-path based neighbors to exploit structural information, feature and path attentions to learn the importance of attributes and meta-paths. We use **HACUD** to denote its original model presented in their paper, and $\textbf{HACUD}_{S2}$ for its variant that adopts the second strategy mentioned above.
- **MAHINDER**: our proposed method. We also derive six variants of MAHINDER to comprehensively compare and analyze the performances of its each component. They are,
- $\textbf{MAHINDER}_{\backslash S}$: removing social view and its corresponding meta-paths.
- $\textbf{MAHINDER}_{\backslash D}$: removing device view and its corresponding meta-paths.
- $\textbf{MAHINDER}_{\backslash F}$: removing fund view and its corresponding meta-paths.
- $\textbf{MAHINDER}_{\backslash L}$: removing link information and its corresponding attention module, and only modeling the node sub-paths.
- $\textbf{MAHINDER}_{\backslash EnAtt}$: removing node and link attention mechanisms in path encoder.
- $\textbf{MAHINDER}_{\backslash MpAtt}$: removing attention mechanism modeling importance of views.

5.3 Implementation Details

We generate paths of nodes and links based on MaxCompute[4]. It allows us to sample path length up to twenty and longer. We implement the proposed model based on TensorFlow [1]. We randomly initialize the model parameters with a xavier initializer [9] and choose Adam [15] as the optimizer. During the training, the positive examples are upsampled to keep the positive rate at around 10% in our dataset. Moreover, we set the batch size to 512, the learning rate to 0.01 and set the regularizer parameter $\lambda = 0.01$ to prevent overfitting. For the other comparison methods, we optimize their parameters according to their papers. Moreover, for all compared methods, we implement them on parameter server based distributed learning systems [41] for scaling up to large-scale datasets. Five-run-average values on testing dataset are reported.

[4]https://www.alibabacloud.com/product/maxcompute

Table 2: The statistical information of our MAHIN in the experiment.

Dataset	Type		Attribute		Total
	Number	Examples	Number	Examples	
Node	4	User/ Merchant/ Phone/ Computer	100	NodeType/ [*User Profiles*]: Age/Gender/Married/IsVIP/... [*Credit Information*]: CreditScore/IsInBlacklist/... [*Purchase Behaviors*]: PurchaseAmountAYear/... [*Asset Information*]: Asset/HasCar/HasFactory/...	14,984,670
Link	6	Family/Friend/Workmate/ Trade/Transfer/ Login	45	LinkType/ [*Social*]: FirstRelatedTime/... [*Fund*]: TradeCategory/TransferAmount/... [*Device*]: LoginTime/StayMinute/...	168,864,052

Table 3: Performances of different methods on the dataset. The subscripts indicate the increasing value compared to GBDT.

Metric	GBDT	DeepForest	HAN	HACUD	HAN_{S2}	$HACUD_{S2}$	MAHINDER
AUC	$0.891_{/0.000}$	$0.914_{/0.023}$	$0.920_{/0.029}$	$0.925_{/0.034}$	$0.927_{/0.036}$	$0.930_{/0.039}$	$\mathbf{0.953_{/0.062}}$
$R@P_{0.1}$	$0.403_{/0.000}$	$0.411_{/0.008}$	$0.424_{/0.021}$	$0.433_{/0.030}$	$0.456_{/0.053}$	$0.490_{/0.087}$	$\mathbf{0.564_{/0.161}}$

5.4 Metrics

We use two widely adopted metrics to measure the performance of financial defaulter detection, namely **AUC** and $\mathbf{R@P_N}$.

The first metric **AUC** is the area under the ROC Curve and is defined as:

$$AUC = \frac{\sum_{u \in \mathcal{U}^+} rank_u - \frac{|\mathcal{U}^+| \times (|\mathcal{U}^+|+1)}{2}}{|\mathcal{U}^+| \times |\mathcal{U}^-|} \tag{12}$$

Here, $\mathcal{U}^+$ and $\mathcal{U}^-$ denotes the positive and negative set in the testing set, respectively. And $rank_u$ indicates the rank of user u via the score of prediction.

The second metric $\mathbf{R@P_N}$ means the **R**ecall when the **P**recision equals **N**. In our dataset, we set **N**=10%. Since the positive rate in credit payment services is low in general (about 0.5% in our dataset), this metric that lifts 20 times in our dataset (10% vs 0.5%) indicates the ability to detect top-ranked positive samples and balance the impact on the real-world business system.

The higher **AUC** and $\mathbf{R@P_N}$ indicate the higher performance of the approaches.

5.5 Main Results

Table 3 demonstrates the main results of all compared methods on the dataset. The major findings from the experimental results can be summarized as follows:

(1) We can clearly observe that our model MAHINDER outperforms all the baselines by a large margin. Its AUC, with reported value of 0.953, is at least 4% higher than the tree-based methods, and $R@P_{0.1}$ gets 15% higher at the same time. Furthermore, MAHINDER is more advanced than the state-of-the-art heterogeneous network methods, i.e., HAN and HACUD, with about 2.8% increased AUC and 13.1% increased $R@P_{0.1}$, respectively. That is, the usage of MAHIN and the further exploring on the local behavioral patterns make it more superior to the competitors. Besides, the obvious improvement of $R@P_{0.1}$ indicates that the model can detect more top-ranked default users under the same precision. This is critical to the real-world system when leveraging the business effect and interception rate.

(2) For baselines, DeepForest gets better performances than GBDT among the tree-based methods. It achieves better AUC and $R@P_{0.1}$ via deeper modeling the user attributes. Moreover, it can be further seen that HIN embedding methods, e.g., HAN and HACUD, are more effective than tree-based methods due to taking advantage of structural information. AUC is increased by more than 1% and $R@P_{0.1}$ is improved by 4.5%. In addition, we observe that the performances of HAN and HACUD could be further improved by feeding our proposed multi-views information. Both HAN_{S2} and $HACUD_{S2}$ outperform their original models. These observations demonstrate that it is effective to model attributes and structural information based on multiple views simultaneously.

(3) For the unseen target users in the testing set, AUC and $R@P_{0.1}$ are 0.949 and 0.562, respectively, which are close to the overall performance. It demonstrates that our model can efficiently learn newcomers' representation.

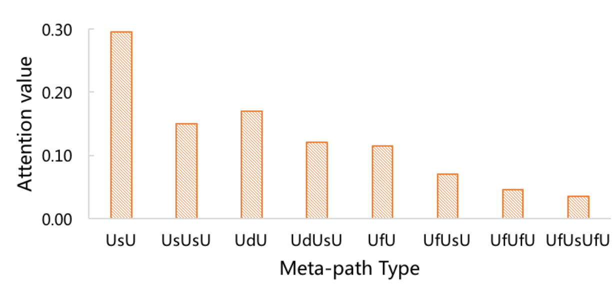


Figure 5: The average attention values of meta-paths corresponding to multiple views.

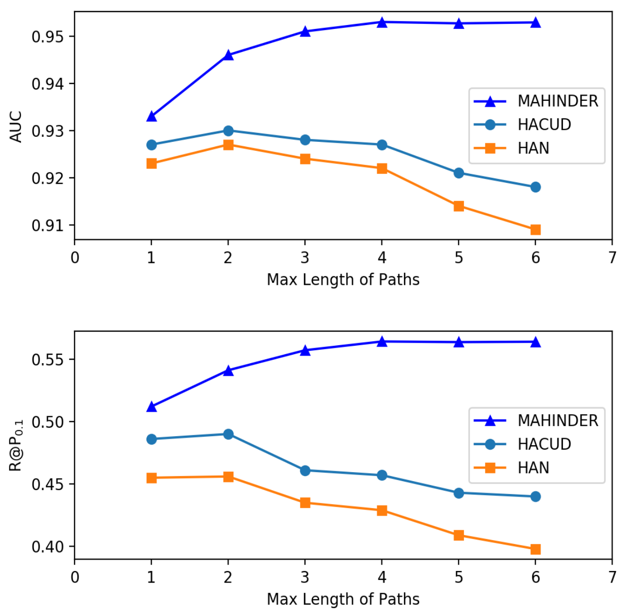


Figure 6: The performances of MAHINDER, HACUD and HAN on different max length of paths.

5.6 The Effects of Meta-paths

Next, we investigate the effects of different meta-paths. First, we report the attentive weights of meta-paths with all views in meta-path attention layer shown as Figure 5. From the figure we can see that the contribution of first-order nodes and links (i.e., UsU, UdU, UfU) is of great importance for defaulter detection, and *social view* > *device view* > *fund view* on the dataset. Besides, it is worth noting that the long meta-paths also contribute a lot of the weights, such as UsUsU, UdUsU, and UfUsU, etc. In other words, the proposed MAHINDER could benefit from these complex interactive information of attributes and structures for prediction.

To further evaluate the effects of long meta-paths, we vary the max length of paths and rerun our MAHNINER and two strong baselines, namely HACUD and HAN, for comparison. Figure 6 presents the results of AUC and R@P$_{0.1}$ of the three compared methods when varying the max length of sampled paths. We can observe that the performance of our model improves obviously with the increase of max length, and the trend of the curves become flatter when max length $\geq$ 5. However, the two metrics of the other two methods begin to decrease when the max length of path is longer than 2. We analyze the reasons as follows. Since both HACUD and HAN simply adopt meta-path based neighbors to aggregate the target user's features, it may bring noise when the neighbor is too far away. Hence both HACUD and HAN achieve best performance when the max length of path equals 2. However, since our MAHINDER is equipped with meta-path based path encoder, the sequential patterns among the target node and the neighbor nodes are learned, which results in a monotonically increasing. It gives us an encouraging result that long paths intuitively might have positive effects especially under the adversarial features of the malicious defaults.

5.7 Ablation Test

Furthermore, we perform the ablation test for our MAHINDER, and the results are shown in Table 4.

5.7.1 The effects of views. Firstly, we demonstrate the effectiveness of different views by removing the corresponding view information (i.e., removing the corresponding meta-paths) respectively. Compared the second to the fourth columns with the last column in Table 4, we can clearly see that both AUC and R@P$_{0.1}$ get worse by removing any view-specific information. It is the worst by removing social view, which means social view has more significant impact on detecting defaulter in our dataset. The results reflect the importance of macroscopically modeling multiple views as well, since every view has a positive contribution for our task.

5.7.2 The effects of fine-grained behavior modeling. Next, to further verify the importance of local behaviors modeling and attention mechanisms applied on path encoder and meta-paths, we take a comparison with our approach and its three variants, as shown in Table 4. The first variant MAHINDER$_{\backslash L}$ removes the links and their associated operations. We could clearly observe that MAHINDER$_{\backslash L}$ performs worse than our full model, which illustrates the links are useful. Meanwhile, its decreasing values (v.s. MAHINDER) reflect that link patterns play a significant role in default user detection. However, MAHINDER$_{\backslash L}$ still gets a performance of 0.936 on AUC and 0.525 on R@P$_{0.1}$ and clearly outperforms HAN and HACUD, which indicates the effectiveness of the path encoder and multi-view meta-paths.

The second variant is to remove the attention mechanism in path encoder module, namely MAHINDER$_{\backslash EnAtt}$. The results show that it performs worse (AUC: 0.945, R@P$_{0.1}$: 0.543) than MAHINDER, implying that different nodes/links of a path have different importance for defaulter prediction. Furthermore, MAHINDER$_{\backslash MpAtt}$, the third variant that removes attention mechanism when modeling importance of views, also becomes worse. It proves that different views depict users' preferences from different perspectives. More visual analysis will be detailed in next subsection. Besides, even though removing some important information or components, the variants are still better than most of the baselines. Again, our full system MAHINDER achieves the best performance by integrating the multiple views and modeling the local behavioral patterns elaborately. It demonstrates that sufficiently modeling these information could better reflect users' features on default user prediction.

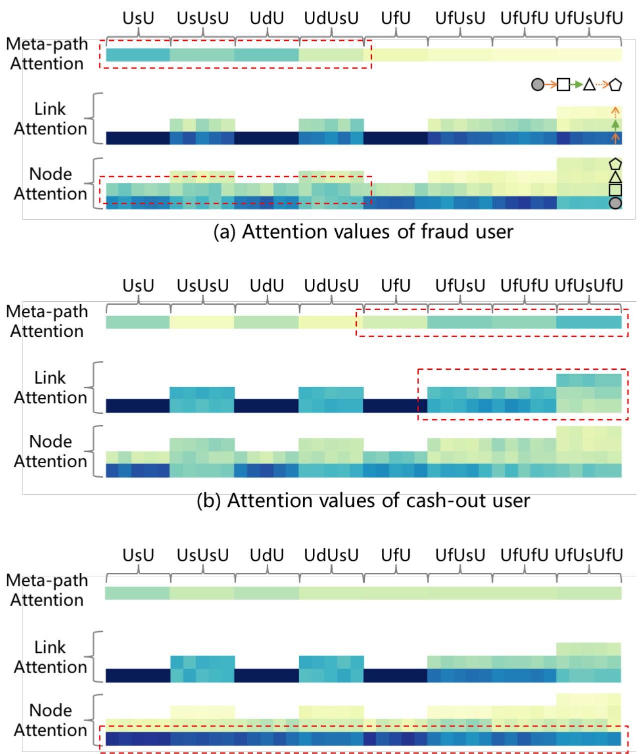
5.8 Visualization

We next look closer to the data and visualize the average attention values of three specific defaulters manually confirmed on the testing set (i.e., fraud users, cash-out users, and unintentional defaulters), as shown in Figure 7. For a clear visualization, we only plot 5 paths of a meta-path, resulting 40 of the horizontal axis length[5]. The vertical axis from the bottom to the top represents the direction of a path. A path example is shown in Figure 7 (a). For each path, the darker the color, the higher the attention weight.

[5]Note that the number of all the meta-paths used in this paper is eight.

Table 4: Performances of ablation test on the proposed MAHINDER method. The subscripts indicate the decreasing value compared to MAHINDER.

Metric	MAHINDER$_{\backslash S}$	MAHINDER$_{\backslash D}$	MAHINDER$_{\backslash F}$	MAHINDER$_{\backslash L}$	MAHINDER$_{\backslash EnAtt}$	MAHINDER$_{\backslash MpAtt}$	MAHINDER
AUC	0.929$_{/-0.024}$	0.934$_{/-0.019}$	0.938$_{/-0.015}$	0.936$_{/-0.017}$	0.945$_{/-0.008}$	0.942$_{/-0.011}$	**0.953$_{/0.000}$**
R@P$_{0.1}$	0.487$_{/-0.077}$	0.510$_{/-0.054}$	0.521$_{/-0.043}$	0.525$_{/-0.039}$	0.543$_{/-0.021}$	0.536$_{/-0.028}$	**0.564$_{/0.000}$**



(a) Attention values of fraud user

(b) Attention values of cash-out user

(c) Attention values of unintentional defaulter

Figure 7: Attention values of three specific types of defaulters, namely fraud users (a), cash-out users (b) and unintentional defaulters (c). For every meta-path, the weights of five path instances are visualized. For each path instance, the corresponding weights on node and link attentions are exhibited from the bottom to the top along with the vertical direction. For example, the attention weight on the target user is shown on the bottom row of "Node Attention" in each sub-figure. For every path visualization, the darker color indicates the higher attention weight.

The results show that different specific defaulters have different preferences. The fraud users have higher attention values on social and device views (e.g., UsU, UdU, etc.) and mainly pay close attention to first-order neighbors (e.g., device aggregation and activity aggregation mentioned in Section 2), while the cash-out users have higher attention values on fund and social views (e.g., UfUsU, UfUsUfU, etc.) and pay more attention to high-order links. We conjecture the possible reason for these phenomena is they try to make the fund transfer path deeper to avoid regulation and gloss up the evidences. Besides, the unintentional users exhibit totally different patterns that they have higher attention value on themselves (default due to their own reasons) and more uniform distribution of attention values on all views.

From the visualizations, we recognize that our MAHINDER has good interpretability on results. Recall that defaulter detection systems are usually deployed for financial-related applications. The interpretability of MAHINDER might buoy the deployment on online platform since it may facilitate the confidence of financial decision makers.

6 CONCLUSION AND FUTURE WORK

In this paper, we investigated the study of default user detection on online payment service platforms. By elaborately analyzing the characteristics of financial defaulters on online payment service platforms, we propose a HIN based model MAHINDER to solve the problem. First, we consider multiple relationships of users on online payment service platforms and construct a multi-view attributed heterogeneous information network for better presenting user profiles. Then, both intra-view and cross-view meta-paths are utilized to capture users' related interactive features. A meta-path based path encoder equipped with attention mechanisms is proposed to model the sampled paths in a fine-grained manner. Another attention layer over different meta-paths is used to screen unrelated views. Experiments on large-scale datasets demonstrate that MAHINDER is effective in detecting default user on the platforms and the interpretability of the model is helpful in inferring the default reasons.

Most existing meta-path based approaches for financial defaulter detection related researches are two-stage methods. That is, first extracting features based on pre-defined meta-paths and then designing node classification models by feeding these extracted features. In future, seeking an end-to-end trained financial defaulter detection model could be an interesting problem. Meanwhile, because of the complex reasons and characteristics for financial default, e.g., endogeny, adversary and accumulation, etc., designing interpretable models for this problem is also an interesting direction.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable comments and suggestions. The research work is supported by the National Key Research and Development Program of China under Grant No. 2017YFB1002104, the National Natural Science Foundation of China under Grant No. 61976204, U1811461, 91846113, the Project of Youth Innovation Promotion Association CAS. This work is also supported by Alibaba Group through Alibaba Innovative Research Program.

REFERENCES

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A System for Large-scale Machine Learning. In *OSDI*. 265–283.

[2] Xiang Ao, Ping Luo, Jin Wang, Fuzhen Zhuang, and Qing He. 2018. Mining precise-positioning episode rules from event sequences. *IEEE Transactions on Knowledge and Data Engineering* 30, 3 (2018), 530–543.

[3] Xiang Ao, Haoran Shi, Jin Wang, Luo Zuo, Hongwei Li, and Qing He. 2019. Large-scale frequent episode mining from complex event sequences with hierarchies. *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, 4 (2019), 1–26.

[4] Siddhartha Bhattacharyya, Sanjeev Jha, Kurian Tharakunnel, and J. Christopher Westland. 2011. Data Mining for Credit Card Fraud: A Comparative Study. *Decision Support Systems* 50, 3 (2011), 602–613.

[5] Hanjun Dai, Bo Dai, and Le Song. 2016. Discriminative Embeddings of Latent Variable Models for Structured Data. In *ICML*. 2702–2711.

[6] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. Metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *KDD*. 135–144.

[7] Jerome H Friedman. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* (2001), 1189–1232.

[8] Taoyang Fu, WangChien Lee, and Zhen Lei. 2017. Hin2vec: Explore Meta-paths in Heterogeneous Information Networks for Representation Learning. In *CIKM*.

[9] Xavier Glorot and Yoshua Bengio. 2010. Understanding the Difficulty of Training Deep Feedforward Neural Networks. In *AISTATS*. 249–256.

[10] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *NIPS*. 1024–1034.

[11] Jiazhen He, James Bailey, and Rui Zhang. 2014. Exploiting Transitive Similarity and Temporal Dynamics for Similarity Search in Heterogeneous Information Networks. In *DASFAA*. 141–155.

[12] Bethany Hoogs, Thomas Kiehl, Christina Lacomb, and Deniz Senturk. 2007. A Genetic Algorithm Approach to Detecting Temporal Patterns Indicative of Financial Statement Fraud. *Intelligent Systems in Accounting, Finance & Management: International Journal* 15, 1-2 (2007), 41–56.

[13] Binbin Hu, Zhiqiang Zhang, Chuan Shi, Jun Zhou, Xiaolong Li, and Yuan Qi. 2019. Cash-out User Detection based on Attributed Heterogeneous Information Network with a Hierarchical Attention Mechanism. In *AAAI*. 946–953.

[14] Ming Ji, Jiawei Han, and Marina Danilevsky. 2011. Ranking-based Classification of Heterogeneous Information Networks. In *KDD*. 1298–1306.

[15] Diederik P Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

[16] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

[17] Jundong Li, Harsh Dani, Xia Hu, Jiliang Tang, Yi Chang, and Huan Liu. 2017. Attributed Network Embedding for Learning in A Dynamic Environment. In *CIKM*. 387–396.

[18] Yuan Li, Yiheng Sun, and Noshir Contractor. 2017. Graph Mining Assisted Semi-supervised Learning for Fraudulent Cash-Out Detection. In *ASONAM*. 546–553.

[19] Ziqi Liu, Chaochao Chen, Xinxing Yang, Jun Zhou, Xiaolong Li, and Le Song. 2018. Heterogeneous Graph Neural Networks for Malicious Account Detection. In *CIKM*. 2077–2085.

[20] Changping Meng, Reynold Cheng, Silviu Maniu, Pierre Senellart, and Wangda Zhang. 2015. Discovering Meta-paths in Large Heterogeneous Information Networks. In *WWW*. 754–764.

[21] Vinod Nair and Geoffrey E. Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In *ICML*. 807–814.

[22] Carsten AW Paasch. 2008. Credit Card Fraud Detection Using Artificial Neural Networks Tuned by Genetic Algorithms. *Ph.D. Thesis. Hong Kong University of Science and Technology* (2008).

[23] Jon T. S. Quah and M. Sriganesh. 2008. Real-time Credit Card Fraud Detection Using Computational Intelligence. *Expert Systems With Applications* 35, 4 (2008), 1721–1732.

[24] Pediredla Ravisankar, Vadlamani Ravi, G Raghava Rao, and Indranil Bose. 2011. Detection of Financial Statement Fraud and Feature Selection Using Data Mining Techniques. *Decision Support Systems* 50, 2 (2011), 491–500.

[25] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and S Yu Philip. 2018. Heterogeneous Information Network Embedding for Recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2018), 357–370.

[26] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and S Yu Philip. 2016. A Survey of Heterogeneous Information Network Analysis. *IEEE Transactions on Knowledge and Data Engineering* (2016), 17–37.

[27] Yizhou Sun and Jiawei Han. 2012. Mining Heterogeneous Information Networks: Principles and Methodologies. *Synthesis Lectures on Data Mining and Knowledge Discovery* 3, 2 (2012), 1–159.

[28] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. Pathsim: Meta Path-based Top-k Similarity Search in Heterogeneous Information Networks. In *VLDB*. 992–1003.

[29] Yizhou Sun, Brandon Norick, Jiawei Han, Xifeng Yan, Philip S Yu, and Xiao Yu. 2013. Pathselclus: Integrating Meta-path Selection with User-guided Object Clustering in Heterogeneous Information Networks. *ACM Transactions on Knowledge Discovery from Data* 7, 3 (2013), 11.

[30] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph Attention Networks. In *ICLR*.

[31] Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural Deep Network Embedding. In *KDD*. 1225–1234.

[32] Daixin Wang, Jianbin Lin, Peng Cui, Quanhui Jia, Zhen Wang, Yanming Fang, Quan Yu, Jun Zhou, Shuang Yang, and Yuan Qi. 2019. A Semi-supervised Graph Attentive Network for Financial Fraud Detection. In *ICDM*.

[33] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous Graph Attention Network. In *WWW*. 2022–2032.

[34] Kun Yao, Hoi Fong Mak, et al. 2014. PathSimExt: Revisiting PathSim in Heterogeneous Information Networks. In *WAIM*.

[35] Dianmin Yue, Xiaodan Wu, Yunfeng Wang, Yue Li, and Chao-Hsien Chu. 2007. A Review of Data Mining-based Financial Fraud Detection Research. In *WiCOM*. 5519–5522.

[36] Chuxu Zhang, Chao Huang, Lu Yu, Xiangliang Zhang, and Nitesh V Chawla. 2018. Camel: Content-aware and Meta-path Augmented Metric Learning for Author Identification. In *WWW*. 709–718.

[37] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V. Chawla. 2019. Heterogeneous Graph Neural Network. In *KDD*. 793–803.

[38] Chuxu Zhang, Lu Yu, Xiangliang Zhang, and Nitesh V Chawla. 2018. Task-guided and Semantic-aware Ranking for Academic Author-paper Correlation Inference. In *IJCAI*. 3641–3647.

[39] YaLin Zhang, Xiaolong Li, Yuan Qi, ZhiHua Zhou, Jun Zhou, Wenhao Zheng, Ji Feng, Longfei Li, Ziqi Liu, Ming Li, and et al. 2019. Distributed Deep Forest and its Application to Automatic Detection of Cash-Out Fraud. *ACM Transactions on Intelligent Systems and Technology* 10, 5 (2019).

[40] Yizhou Zhang, Yun Xiong, Xiangnan Kong, Shanshan Li, Jinhong Mi, and Yangyong Zhu. 2018. Deep Collective Classification in Heterogeneous Information Networks. In *WWW*. 399–408.

[41] Jun Zhou, Xiaolong Li, Peilin Zhao, Chaochao Chen, Longfei Li, Xinxing Yang, Qing Cui, Jin Yu, Xu Chen, Yi Ding, and Yuan Alan Qi. 2017. KunPeng: Parameter Server Based Distributed Learning Systems and Its Applications in Alibaba and Ant Financial. In *KDD*. 1693–1702.

[42] Zhihua Zhou and Ji Feng. 2017. Deep Forest: Towards An Alternative to Deep Neural Networks. In *IJCAI*. 3533–3539.