Understanding the Impact of Startups’ Features on Investor Recommendation Task via Weighted Heterogeneous Information Network

Sen Wu, Ruojia Chen, Guiying Wei, Xiaonan Gao, and Lifang Huo

School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China

Correspondence should be addressed to Xiaonan Gao; gaoxiaonan0001@163.com

Received 24 December 2020; Revised 23 March 2021; Accepted 12 April 2021; Published 22 April 2021

Academic Editor: Huajiao Li

Investor recommendation is a critical and challenging task for startups, which can assist startups in locating suitable investors and enhancing the possibility of obtaining investment. While some efforts have been made for investor recommendation, few of them explore the impact of startups’ features, including partners, rounds, and fields, to investor recommendation performance. Along this line, in this paper, with the help of the heterogeneous information network, we propose a FEatures’ COntribution Measurement approach of startups on investor recommendation, named FECOM. Specifically, we construct the venture capital heterogeneous information network at first. Then, we define six venture capital metapaths to represent the features of startups that we focus on. In this way, we can measure the contribution of startups’ features on the investor recommendation task by validating the recommendation performance based on different metapaths. Finally, we extract four practical rules to assist in further investment tasks by using our proposed FECOM approach.

1. Introduction

Venture capital (VC) is an important source of funding for startups [1]. It plays a key role in startups’ sustainable growth and performance [2] and the innovation in the economy [3]. Investment companies provide financial support for startups. Besides that, they provide startups with necessary experience and knowledge for their development. However, there is information asymmetry between startups and investment companies, and it is challenging for startups to find suitable investors, especially for the startups that have no previous investors before [4]. Finance is an integral part of the startups’ process, yet obtaining it is challenging [5]. For this reason, an investor-filtering system for startups can be extremely beneficial [4]. Therefore, achieving information filtering and then recommending suitable investors for startups who can meet their personalized financial needs by using recommendation techniques have become a key problem to be solved in the domain of VC.

In the domain of VC, the extensive recommendation studies mainly focus on helping VC companies to find suitable startups [1, 6–8], while effective methods that support startups in finding investors are relatively rare. Xu et al. [4] achieved the investor recommendation by constructing a tripartite network representation containing virtual links. Antonio et al. [9] pointed out that finding a suitable investor who is interested in the startups’ fields usually requires long-term research, which is especially difficult for new startups. Furthermore, recommending suitable investors who are interested in the startups’ features is also challenging due to their inexperience. However, as for the new startups which have not been invested before, the resources they can utilize is limited. They have no previous investors to be mined, and the features of startups are the limited resources available for finding suitable investors for them. Along this line, we try to employ the startups’ features to study the investor recommendation problem, by measuring the contribution of startups’ features on the investor recommendation performance. The features we care about in this paper include partners, rounds, and field. An example of startups is shown in Table 1, the startup Ziroom and Bitmain have the same round, the startup Vxiaohe is in
Angel Investment, and the former two startups have the same investor, while the investor of the third startup is different from them. Therefore, we consider that the feature of round may affect the decision of investors. And, there may be various impacts on investors by different startups’ features. To this end, how to measure the contribution of startups’ features on the investor recommendation performance is a vital task for startups. In this paper, we aim to address startups’ features contribution measurement problem.

For exploring the impact of startups’ features on the investor recommendation task, there are two challenges that need to be solved. (1) How to represent the features related to the investor recommendation task? (2) How to recommend a investor to a startup based on the focused features? Notably, the VC network is a typical heterogeneous information network (HIN), which contains multiple types of nodes and multiple edges between nodes. There is rich structural information and semantic information in the VC network, which provides us the opportunity to address the two challenges based on HIN.

HIN has been successfully applied in recommender systems [10]. Therefore, in this paper, we propose a features’ contribution measurement approach of startups on investor recommendation named FECOM, based on HIN. The purpose of this paper is to explore the impact of startups’ features on the investor recommendation task and extract the rules to assist in further investment tasks. Specifically, we integrate various types of VC data into the form of HIN at first, by which the venture capital heterogeneous information network (VC-HIN) can be constructed. Then, the venture capital metapaths (VC-metapaths) are defined to represent the features of startups that we focus on. Based on this, we propose the FECOM approach, which can conduct the investor recommendation task by utilizing different VC-metapaths. In this way, we can measure the contribution of startups’ features on the investor recommendation task according to different VC-metapaths, and extract beneficial rules to assist in further investment tasks, by analysing the recommendation performance.

The main contributions of this paper can be summarized as follows:

(1) We define the venture capital heterogeneous information network (VC-HIN), by which the structural information contained in the VC network can be expressed.

(2) We formulate the venture capital metapaths (VC-metapaths) to represent the features related to the investor recommendation task, and the semantic information contained in the VC network can be indicated.

(3) We propose a features’ contribution measurement approach of startups on investor recommendation named FECOM. On the basis of it, we can analyse the impact of startups’ features on investor recommendation by validating the recommendation performance based on different metapaths.

(4) We extract four practical rules by exploiting the proposed FECOM approach, which are beneficial for further investment tasks.

The paper is structured as follows. In Section 2, we introduce the related works. In Section 3, we propose two definitions about the proposed FECOM, which are the basis of FECOM. In Section 4, we provide the framework and the details of the proposed FECOM. In Section 5, we explore the impact of startups’ features on investor recommendation and extract four practical rules to assist in further investment tasks through experiments. In Section 6, we conclude the paper.

### 2. Related Work

In this section, we briefly review related works on the VC, VC recommendations, and HIN.

In recent years, related research studies in the domain of VC have continued to grow. Chircop et al. [11] studied the influence of religious belief on VC decision-making. Andrieu and Groh [12] established a startups’ investor selection model and found that startups would select investors based on the factors such as the quality of the investor support. Tian et al. [13] confirmed the nonlinear relationship between the geographic distance between the investors and the startups and the startups’ performance through the research. Cheng and Tang [14] conducted the research on the impact of investment partner selection strategy, the industry of the startup, and geographic uncertainty on VC. Wu et al. [15] studied the influence of the network structure on VC alliance’s successful exit from emerging markets. Zhang et al. [16] used multilayer network analysis to simulate the spread of risks in the VC market when external shocks affect investment companies or startups. Luo [17] put forward the countermeasures for the company’s financial risk investment by predicting the stock premium. Salamzadeh and Kesim [5] tried to conceptualize the phenomenon “startup” and recognize the challenges they might face. Antonio et al. [9] examined the association between the presence of venture capital and the growth of startups.

However, as for the VC recommendation, there are few research studies focusing on that. Stone et al. [1] demonstrated the effectiveness of the collaborative filtering recommendation technique in the new application domain of finance and improved the recommendation quality of
investment opportunities. Zhao et al. [6] proposed a portfolio optimization framework to solve the problem of information filtering in VC, which can predict the new investment of investment companies. Wu et al. [7] applied a personalized recommendation technique to the domain of VC recommendation to complete the recommendation of VC projects. Xu et al. [4] combine with prob diffusion algorithm, develop the investor recommendation by constructing a tripartite network representation containing virtual links. The existing recommendation research studies in the domain of VC mainly concentrate on the recommendation of VC startups and seldom study the recommendation of investors based on the needs of startups. However, due to the information asymmetry between investors and startups, the financing process is usually challenging, especially for the startups that have not yet been invested before [4]. As the features of startups without investment are the limited resources available for them to find suitable investors, we try to exploit the features and measure the contribution of startups’ features on the investor recommendation performance and then extract beneficial rules to assist in the further investment recommendation problem by analysing the recommendation performance.

As an emerging direction, HIN has attracted the attention of many scholars. HIN contains rich structural and semantic information [18]. The recommendation based on HIN can naturally simulate the rich information contained in the network and then realize personalized recommendation. Node similarity measurement, as a key part of information extraction from HIN-based recommendation, has been widely studied. Haveliwala [19] proposed the personalized PageRank algorithm (PRANK recommendation), which measures the similarity between nodes by randomly walking between nodes; Jeh and Widom [20] proposed the SimRank algorithm, which calculates the similarity between nodes through the similarity of their neighbors. However, both PRANK and SimRank ignored the semantic information contained in the HIN. In addition to those algorithms, there are many HIN-based recommendation techniques considering the metapath-based similarity [21]. Sun et al. [18] proposed the PathSim algorithm based on symmetric metapaths to achieve similarity measurement between nodes of the same type in the network; Lao and Cohen [22] proposed the PCRW algorithm to measure the similarity between different types of nodes in the network through a random walk model with path constraints; Shi et al. [23] proposed the HeteSim algorithm based on the arbitrary metapath, which can measure the similarity between any type of node in the network. The above three algorithms all use metapath-based similarity to measure the node similarity of network semantic information, but they do not consider the network structure information. All the above similarity measurement techniques have some limitations because they fail to consider the structure or semantic information of HIN at the same time. Lee et al. [24] proposed the PathRank algorithm to calculate the similarity between nodes by using the one-way random walk model. It realizes the measurement of nodes’ similarity on the basis of fully extracting the structure and semantic information of HIN. Compared with the above node similarity measurement methods, the PathRank algorithm can utilize the rich information contained in the VC network. Furthermore, we will calculate the scores of investors to be recommended on the basis of the PathRank algorithm in the following sections by considering the rich structural information and semantic information in the VC network.

Among the above existing literature studies, there is less attention from scholars that has been devoted to the evaluation of the impact of startups’ features on the investor recommendation task. Motivated by these observations, we propose a features’ contribution measurement approach of startups on investor recommendation named FECOM. According to the proposed FECOM, the practical rules could be obtained to support the investment task by evaluating the recommendation performance based on different metapaths.

3. Preliminaries

In this section, we propose two definitions about the proposed FECOM approach, namely, venture capital heterogeneous information network (VC-HIN) and venture capital metapaths (VC-metapaths) in the VC network, which are the basis of the model. Specifically, VC-metapaths can provide various semantic information for the investor recommendation task.

Definition 1. Venture capital heterogeneous information network (VC-HIN).

Venture capital heterogeneous information network (VC-HIN) is defined as a weighted heterogeneous information network constructed based on venture capital data, including different types of nodes and weighted edges. Specifically, VC-HIN is indicated by $G = (V, E)$, where $V = \{C, P, R, F, I\}$ is the node set containing five kinds of node, namely, $C$-company, $P$-company partner, $R$-investment round, $F$-field, and $I$-investor, and $E$ refers to the edge set including 14 kinds of edge types between different nodes, which are directed and reveal the relations between corresponding nodes. As for the weights of edges in VC-HIN, we define different weight calculation methods for different kinds of edges. In terms of the edge between $I$ and $C$, its weight is the total number of times of the investor who used to invest in the startup. For the edge between $I$ and $P$, the weight is the number of times of the investor invested in the companies built by $P$. Similarly, as to the edge between $I$ and $R$, its weight is the number of times of the investor invested in companies in round $R$. And, the weight of the edge between $I$ and $F$ is the number of times of the investor invested in companies in field $F$. Particularly, the weights of edges between $C$ and $P$, $P$, $R$, and $F$ are defined as 1.

An example of VC-HIN is shown in Figure 1. There are two company nodes $C_1$ and $C_2$, two investor nodes $I_1$ and $I_2$, two company partner nodes $P_1$ and $P_2$, two round nodes $R_1$ and $R_2$, and one field node $F_1$. The lines represent the edges between nodes, and thicker lines indicate the edges with larger weights. For example, the line between $I_2$ and $F_1$ is thicker than that between $I_1$ and $F_1$, which indicates that the
weight of the edge linking \( I_2 \) and \( F_1 \) is larger than the edge linking \( I_1 \) and \( F_1 \). It proves that the investor \( I_2 \) invests more in companies in the field \( F \) than \( I_1 \). Similarly, the weight of the edge linking \( I_1 \) and \( C_1 \) is larger than the edge linking \( I_2 \) and \( C_1 \), revealing that the company \( C_1 \) is invested by the investor \( I_1 \) more than \( I_2 \). Furthermore, the edges between different nodes are directed and reveal the relations between corresponding nodes, and the edges have two kinds of directions “+” and “−”. For example, the edge \( C_1 \rightarrow I \) implies that the company \( C_1 \) is invested by the investor \( I \). And, the edge \( I_1 \rightarrow I \) implies that the investor \( I_1 \) invests in the company \( C_1 \).

**Definition 2.** Venture capital metaphats (VC-metapaths).

The venture capital metaphats (VC-metapaths) in VC-HIN are defined specifically to the investor recommendation task studied in this paper. We first formulate the venture capital heterogeneous information network schema TG = (A, R) shown in Figure 2 with the node type mapping function \( \phi_A: V \rightarrow A \) and the edge type mapping function \( \phi_E: E \rightarrow R \), where A represents the node type set and R represents the edge type set. Based on TG, we define VC-metapaths, which can represent the compound relation between different nodes in VC-HIN. A VC-metapath can be represented as \( g_k = v_1 \rightarrow \phi^V v_2 \ldots v_i \rightarrow \phi^V v_i + 1 \), where each element in set \( \{v_1, v_2, \ldots, v_i\} \) represents a node in the network, and the element in set \( \{e_1, e_2, \ldots, e_l\} \) represents an edge in VC-HIN. We are supposed to recommend suitable investors for startups; therefore, the first node in the VC-metapath is fixed as a company node, and the last node must be an investor node. As shown in Figure 2, there are five kinds of node types in the set A and 14 kinds of edge types including directions in the set R.

As we focus on the contribution of the target startups’ features on the investor recommendation task, the commonly used features include startups’ partners, rounds, and fields. As shown in Table 2, we formulate six VC-metapaths based on these three kinds of startups’ features. The longer the metapath, the fuzzier the corresponding semantic information and the lower the reliability of the similarity measure [18]. Therefore, we define that there will be no more than four nodes in VC-metapaths. The six VC-metapaths cannot only exploit the semantic information behind the different types of nodes and edges in VC-HIN but also simulate several classical recommendation methods, such as collaborative filtering recommendation and content-based filtering. Table 2 lists the six VC-metapaths, their semantic analysis conclusions, and similar methods. For example, the VC-metapath CPI can mine the semantic information behind startups’ partners and the edge linking \( P \) and \( I \) contained in the metapath. This metapath recommends investors who have invested in the partners of target companies, which is similar to the paradigm of content-based recommendation.

Figure 3 shows the six VC-metapaths, where the edges contained in the metapath are represented by solid lines. CPI and CPCF are shown in Figure 3(a), where CPI consists of the edge linking \( C \) and \( P \) and the edge linking \( P \) and \( I \), while CPCF is composed of the edge linking \( C \) and \( P \) and the edge linking \( C \) and \( I \). For the metapath CRI and CRCF in Figure 3(b), the edge linking \( C \) and \( R \) and the edge linking \( R \) and \( I \) make up CRI, and CRCF is comprised of the edge linking \( C \) and \( R \) and the edge linking \( C \) and \( I \). Similarly, CPI and CIFC are shown in Figure 3(c), where CPI consists of the edge linking \( C \) and \( F \) and the edge linking \( F \) and \( I \), while CIFC is composed of the edge linking \( C \) and \( F \) and the edge linking \( C \) and \( I \).

Based on the basic concepts mentioned above, the proposed FECOM can consider the structure information contained in VC-HIN and the semantic information contained in multiple VC-metapaths at the same time, which will fully mine the rich information contained in the VC network and is able to explore the impact of startups’ features to the investor recommendation task. The details about FECOM will be introduced in the next section.

4. **FECOM Approach**

In this section, we first provide the overall framework of the proposed FECOM approach and then illustrate its details.

4.1. **Overview of FECOM.** In this section, we introduce the framework of FECOM.

The framework of the proposed FECOM is shown in Figure 4, consisting of two main components:

1. **Constructing Layer.** We construct the VC-HIN and specify the VC-metapaths of VC-HIN in this layer
Table 2: VC-metapaths of VC-HIN

<table>
<thead>
<tr>
<th>VC-metapaths</th>
<th>Semantic analysis</th>
<th>Similar method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPCI</td>
<td>Recommend investors who have invested in the companies with the same partners as target startups</td>
<td>Collaborative filtering recommendation</td>
</tr>
<tr>
<td>CPI</td>
<td>Recommend investors who have invested in the partners of target startups</td>
<td>Content-based recommendation</td>
</tr>
<tr>
<td>CRCI</td>
<td>Recommend investors who have invested in the companies with the same rounds as target startups</td>
<td>Collaborative filtering recommendation</td>
</tr>
<tr>
<td>CRI</td>
<td>Recommend investors who have invested in the rounds of target startups</td>
<td>Content-based recommendation</td>
</tr>
<tr>
<td>CFCI</td>
<td>Recommend investors who have invested in the companies with the same fields as target startups</td>
<td>Collaborative filtering recommendation</td>
</tr>
<tr>
<td>CFI</td>
<td>Recommend investors who have invested in the fields of target startups</td>
<td>Content-based recommendation</td>
</tr>
</tbody>
</table>

Figure 3: Examples of VC-metapaths. (a) The metapath CPI and CPCI. (b) The metapath CRI and CRCI. (c) The metapath CFI and CFCI.

Figure 4: Framework of the proposed FECOM where the left part shows the constructing layer and the right part presents the recommending layer.
Complexity

firstly. The VC-HIN is built according to the VC data collected in advance, which is similar to Figure 1; however, it is bigger and more complex than Figure 1. The VC-metapaths are designed according to the information to be mined; since the target startups in this paper have no investors before, the partners, rounds, and fields of them can be exploited to realize the investor recommendation. As shown in Table 2 and Figure 3, the VC-metapaths in this paper have been designed. Herein, we give two kinds of transition matrices, the VC-HIN transition matrix $T_1$ and the VC-metapath transition matrix $T_{gk}$, $k \in \{1, 2, \ldots, 6\}$, for each VC-metapath $g_k$, $k \in \{1, 2, \ldots, 6\}$. Since there are six metapaths in Table 2, $k$ is an integer less than or equal to 6. The VC-HIN transition matrix $T_1$ is extracted from VC-HIN, which can express the structural information of VC-HIN. Similarly, the VC-metapath transition matrix $T_{gk}$ is calculated based on the VC-metapath, indicating the semantic information contained in the metapath.

(2) Recommending Layer. This layer requires the VC-HIN transition matrix $T_1$ and the VC-metapath transition matrix $T_{gk}$ as input and conducts the investor recommendation. The recommending layer contains two personalized investor recommendation strategies, which are based on the single metapath and the mixed metapath. The investor recommendation of the mixed metapath is derived from recommendation results of different single metapaths. As shown in the recommending layer, the VC-HIN transition matrix $T_1$ and the VC-metapaths’ transition matrices $T_{g1}$, $T_{g2}$, and $T_{g3}$ are obtained in the constructing layer. Next, we utilize the VC-HIN transition matrix $T_1$ and the VC-metapaths’ transition matrices containing $T_{g1}$, $T_{g2}$, and $T_{g3}$ in Figure 4, respectively, to obtain the optimal investor recommendation sets $S_1$, $S_2$, and $S_3$ based on the single metapaths $g_1$, $g_2$, and $g_3$. The VC-metapaths’ transition matrices $T_{g1}$ and $T_{g2}$ with better recommendation effectiveness will be selected to generate the mixed metapath transition matrix $T_{gm}$ of the mixed metapath $g_m$ after comparing the recommendation performance on each single metapath. In the end, the optimal investor recommendation set $S_m$ can be obtained based on the VC-HIN transition matrix $T_1$ and the mixed metapath transition matrix $T_{gm}$. By evaluating the performance of the recommendation set, we can analyse the impact of the corresponding VC-metapath on the investor recommendation task.

Next, based on the framework of FECOM mentioned above, the technical details about the constructing layer and recommending layer are interpreted, respectively.

4.2 Constructing Layer. In the constructing layer, we are supposed to construct the VC-HIN and design the VC-metapaths based on the VC data collected before firstly. Next, we compute the VC-HIN transition matrix $T_1$ and the VC-metapath transition matrix $T_{gk}$ based on the adjacency matrix $P$ and the edge adjacency matrix $P_{Rn}$, $n \in \{1, 2, \ldots, 14\}$, respectively. Since there are 14 edge types in $T = (A, R)$, $n$ is an integer less than or equal to 14.

For the adjacency matrix $P$, it is represented as a $|V| \times |V|$ matrix, where $|V|$ is the total numbers of all the nodes in VC-HIN. And, the element $P(i, j)$ of $P$ represents the weight of the edge between the node $v_i$ and $v_j$.

Based on the adjacency matrix $P$, the VC-HIN transition matrix $T_1$ can be computed according to equation (1), which is $\frac{4}{3}|V| \times |V|$ matrix:

\[
T_1(i, j) = \begin{cases} 
\frac{w_{ij}}{\sum_j P(i, j)} & \text{if } e_i, \text{ linking } v_i \text{ and } v_j \in E \land i \leq |V| \land j \leq |V|, \\
0, & \text{otherwise},
\end{cases}
\]

(1)

where $T_1(i, j)$ is the element of the $i$th row and the $j$th column of $T_1$, representing the probability of the transition between the node $v_i$ and the node $v_j$.

As for the edge adjacency matrix $P_{Rn}$, it is also a $|V| \times |V|$ matrix, where $R_n$ is one of the edge types in the edge type set $R$ and $|V|$ is the total numbers of all the nodes in VC-HIN. The element $P_{Rn}(i, j)$ of $P_{Rn}$ reveals the weight of the edge linking the node $v_i$ and $v_j$ whose edge type belongs to the edge type $R_n$.

For the metapath $g_k = v_1 \longrightarrow^{e_1} v_2 \ldots v_i \longrightarrow^{e_i} v_i + 1$ mentioned in Section 3, its transition matrix $T_{gk}$ can be calculated after normalizing the edge adjacency matrix $P_{Rn}$. The normalized $P_{Rn}$ will be represented as $N_{Rn}$, which is defined as follows:

\[
N_{Rn}(i, j) = \begin{cases} 
\frac{w_{ij}}{\sum_j P_{Rn}(i, j)} & \text{if } e_i, \text{ linking } v_i \text{ and } v_j \in R \land i \leq |V| \land j \leq |V| \land n \leq |R|, \\
0, & \text{otherwise},
\end{cases}
\]

(2)

The VC-metapath transition matrix $T_{gk}$ can also be denoted as a $|V| \times |V|$ matrix, which can be represented as follows:
The basic idea is that the transition matrix of the metapath \( g_k \) can be obtained by multiplying the normalized edge adjacency matrices of all the edges contained in the metapath \( g_k \).

In this section, on the basis of the adjacency matrix \( P \) and the edge adjacency matrix \( P_{\text{normalized}} \) mentioned above, we introduce the calculation process of the VC-HIN transition matrix \( T_{\Sigma} \) and the VC-metapath transition matrix \( T_{g_k} \). Hereafter, we interpret the recommending layer in detail.

4.3. Recommending Layer. For exploring the contribution of different startups’ features on investor recommendation task \( T_{g_k} \), we utilize different types of metapaths to carry out the investor recommendation, respectively, in this layer. We will explain how to obtain the investor recommendation set for target startups under the single metapath and the mixed metapath in this section.

4.3.1. Investor Recommendation Based on the Single Metapath. The single metapaths are the VC-metapaths defined based on the venture capital heterogeneous information network schema, which can represent the compound relation between different nodes in VC-HIN. In this paper, the single metapath refers to the six VC-metapaths shown in Table 2 and Figure 3. The investor recommendation based on the single metapath mainly relies on the VC-metapaths mentioned before. We can obtain the investor recommendation sets according to the partners, rounds, and fields of startups. The investor recommendation based on the single metapath such as CPI, CRI, and CFI can simulate the hybrid recommendation. It can realize the investor recommendation by mixing startups’ various features including the partner, round, and field information. For example, the mixed metapath based on CPI and CRI can consider the partners and rounds of the startups at the same time.

The generation process of the mixed metapath is shown in Figure 5. Firstly, we can obtain the VC-HIN transition matrix \( T_{\Sigma} \) and the VC-metapaths’ transition matrices \( T_{g_1} \), \( T_{g_2} \), and \( T_{g_3} \) for the metapath \( g_1 \), \( g_2 \), and \( g_3 \) from the constructing layer. Then, the investor recommendation sets based on different single metapaths can be selected to generate the mixed metapath. Meanwhile, several single metapaths with better recommendation effectiveness will be selected to generate the mixed metapath.

We realize the investor recommendation based on the single metapaths in this section. Along this line, the investor recommendation based on the mixed metapath will be provided.

4.3.2. Investor Recommendation Based on the Mixed Metapath. The mixed metapath is generated by the single metapaths mentioned above. In this paper, the mixed metapath is built by combining several single metapaths with good recommendation performance, where the number of metapaths to generate the mixed metapath is not fixed. We choose single metapaths based on their performance. The investor recommendation based on the mixed metapath can simulate the hybrid recommendation. It can realize the investor recommendation by mixing startups’ various features including the partner, round, and field information. For example, the mixed metapath based on CPI and CRI can consider the partners and rounds of the startups at the same time.

As shown in the descending order:

\[
\overline{r} = u_{\text{trans}} \times \overline{r} \times T_{\Sigma} + u_{\text{path}} \times \overline{r} \times \overline{T_{g_k}} + u_{\text{rec}} \times \overline{r}.
\]  

(4)

As shown in equation (4), \( \overline{r} \) represents the score set of all the investors in the descending order. We first initialize and then continuously optimize \( \overline{r} \) by equation (4). When \( \overline{r} \) no longer changes, it will be the optimal score set. \( u_{\text{trans}} \), \( u_{\text{path}} \), and \( u_{\text{rec}} \) are all the probability weight parameters, \( \overline{r} \) is represented as a \( 1 \times |V| \) vector, and the element \( \overline{r}(i) \) of \( \overline{r} \) is 1 when its corresponding node \( v_i \) is the target startup.

Next, as there are several probability weight parameters in equation (4), we will obtain different investor recommendation sets by setting different values for the parameters. After comparing the performance of each single metapath under different parameter configurations, we can acquire the optimal performance of each single metapath. Meanwhile, the optimal investor recommendation sets can be produced. Finally, several single metapaths with better recommendation effectiveness will be selected to generate the mixed metapath \( g_m \).

As shown in equation (5), \( T_{g_m} \) represents the mixed metapath transition matrix of the mixed metapath \( g_m \), which is a \( |V| \times |V| \) matrix just like the VC-metapath transition matrix \( T_{g_k} \). \( T_{g_k} \) represents the VC-metapath
transition matrix of the metapath $g_k$, and $w_k$ is the metapath weight of the metapath $g_k$.

After getting the mixed metapath transition matrix $T_{g_m}$, we will introduce the investor recommendation based on the mixed metapath. Firstly, we can obtain the score set of the investors by inputting the VC-HIN transition matrix $T_{g}$ and the mixed metapath transition matrix $T_{g_m}$ into equation (4). Then, the optimal investor recommendation set of the mixed metapath can be collected after comparing the performance of the mixed metapath under different parameters.

In brief, after comparing the optimal investor recommendation sets based on different single metapaths and the mixed metapath by using evaluation metrics, the optimal investor recommendation set can be obtained. Along this line, we can know the contribution of each VC-metapath on the investor recommendation task; based on this, the impact of startups’ features will be explored.

5. Experiments

In this section, we first evaluate the recommendation performance of FECOM on different single metapaths and the mixed metapath through real VC data. Along this line, we can analyse the effectiveness of different metapaths to the investor recommendation task.

5.1. Dataset. The experimental dataset is collected from http://www.itjuzi.com. We delete the investment events with missing fields, rounds, and partners and remove the investors who have invested less than three times and their corresponding investment events, so as to avoid the cold-start problem for investors, which leaves us with 13573 investment events. As shown in Table 3, there are 6816 companies, 1328 investors, 4862 partners, 15 rounds, and 19 fields in the experimental dataset.

We divide the dataset into the training set and the testing set in the chronological order. The training set takes the first 80% of the original data, and the testing set takes the rest part. Besides, we aim to recommend investors to new startups that have no previous investors before; to this end, there are only 1602 investment events in the testing set. Since some startups appear in multiple investment events, there are actually 995 startups involved in the testing set.

5.2. Evaluation Metrics. We choose two evaluation metrics, the ranking score (RS) [25] and the AUC (area under receiver operator curve) [26], to evaluate the effectiveness of our proposed FECOM in the investor recommendation task.

5.2.1. Ranking Score (RS). The ranking score (RS) is an important indicator to measure the relative ranking of the investors in the recommendation set. For the target startup, the relevant investor refers to the investor whose investment record exists in the testing set, and the irrelevant investor refers to the investor who has not invested in the target startup in the training set. RS can measure the relative
rank of relevant investors and measure the degree of 
uniformity between the recommendation set and the actual 
investor set. The smaller the RS, the more accurate the 
relative ranking of the investors in the recommendation set.
On the contrary, the higher the RS, the less accurate the 
ranking. Since there are few actual investors for the target 
startup, the RS is a better evaluation metric than Precision 
and Recall [4].

As shown in equation (6), \( L_u \) represents the number of 
investors to be ranked for the target startup and \( l_{ua} \) rep- 
resents the ranking of the investors to be ranked in the 
recommemded list. The mean ranking score can be obtained 
by averaging the ranking scores of all the startups:

\[
RS_{ua} = \frac{l_{ua}}{L_u}.
\]

5.2.2. AUC (Area under Receiver Operator Curve). The AUC 
can be approximated as the probability that the utility of the 
relevant investor is higher than the utility of the irrelevant 
investor. It can measure the ability of the recommendation 
technique to distinguish relevant investors from irrelevant 
investors. For the target startup, the higher the AUC, the 
stronger the ability of the recommended technique to dis-

5.3. Recommendation Performance on VC-Metapaths. We 
evaluate the investor recommendation based on the single 
metapath and the mixed metapath according to the evalua-
tion metrics mentioned above in this section.

5.3.1. Recommendation Performance on Single Metapaths. 
For locating the optimal recommendation result on each 
metapath, we set multiple probability parameters \( \omega_{\text{trans}} \), 
\( \omega_{\text{path}} \), and \( \omega_{\text{re}} \) where \( \omega_{\text{trans}} \) ranges from 0 to 0.4, \( \omega_{\text{path}} \) ranges from 0.6 to 1, and \( \omega_{\text{re}} \) ranges from 0 to 0.2, and the step size 
of them are all 0.05. Specifically, when \( \omega_{\text{trans}} \) and \( \omega_{\text{re}} \) are 0, 
the recommendation effectiveness is poor, which is not 
reported in this paper. The values of RS and AUC based on 
each single metapath are shown in Tables 4 and 5. The values 
that need to be paid attention are in bold.

As shown in Table 4, among all the metapaths, when 
\( \omega_{\text{trans}} \) is 0.2, \( \omega_{\text{path}} \) is 0.6, and \( \omega_{\text{re}} \) is 0.2, the recommendation 
based on the metapaths CPCI and CPI performs better. The 
best RS of CPCI is 0.211807, while the best RS of CPI is 
0.213887. Furthermore, the investor recommendation based 
on the metapath CPCI performs better than the metapath 
CPI. Besides, when \( \omega_{\text{trans}} \) is 0.35, \( \omega_{\text{path}} \) is 0.6, and \( \omega_{\text{re}} \) is 0.05, 
and the recommendation based on the metapaths CRCI, 
CRI, CFCI, and CFI performs best. Except for this, the 
recommendation based on the metapaths CFCI and CFI 
performs better than the metapaths CRCI and CRI.

As for AUC results shown in Table 5, when \( \omega_{\text{trans}} \) is 0.15, 
\( \omega_{\text{path}} \) is 0.7, and \( \omega_{\text{re}} \) is 0.15, the investor recommendation 
based on the metapath CPCI performs best. Furthermore, 
when \( \omega_{\text{trans}} \) is 0.15, \( \omega_{\text{path}} \) is 0.7, and \( \omega_{\text{re}} \) is 0.15, the investor 
recommendation based on the metapath CPI performs best. 
Among all the metapaths, the investor recommendation 
based on the two metapaths performs better than the other 
metapaths. At the same time, when \( \omega_{\text{trans}} \) is 0.2, \( \omega_{\text{path}} \) is 0.6, 
and \( \omega_{\text{re}} \) is 0.2, the performance of the investor recom-

5.3.2. Recommendation Performance on the Mixed Metapath. 
After optimizing the investor recommendation based on 
different single metapaths mentioned above, we plot their 
performance in Figures 6 and 7. The investor recommen-
dation based on the metapath CPCI and CPI performs better 
than those based on other four metapaths apparently. 
Therefore, we choose the two better single metapaths CPCI 
and CPI. Next, we will select the optimal parameter con-
figuretion and VC-metapaths’ CPCI and CPI to generate the 
mixed metapath and conduct the investor recommendation. 
The metapath weight parameter of the single metapath CPCI 
is set to \( w_1 \), the metapath weight parameter of CPI is set to 
\( w_2 \), and \( w_1 + w_2 = 1 \). Furthermore, in the experiment, the 
step size of the two parameters is 0.1.

The experiment results of the investor recommendation 
on the mixed metapath are shown in Figures 8 and 9. It 
shows that the RS and AUC of the investor recommendation 
based on the mixed metapath vary with the metapath weight 
parameters. The RS gradually decreases as \( w_1 \) increases. 
Meanwhile, the performance of the investor recommenda-


CPCI is better, compared with the results based on the mixed metapath.

5.4. Analysis of VC-Metapaths. After assessing the performance of the investor recommendation based on the single metapath and the mixed metapath, we can now analyse the impact of each defined single metapath on the investor recommendation task and summarize four rules to assist in further investment tasks.

1. The partners of startups contain much more precise semantic information for investor recommendation than the fields and rounds of startups.

As shown in Tables 4 and 5, among all the VC-metapaths, the investor recommendation based on
the metapaths CPCI and CPI performs better than the other metapaths. It indicates that the metapaths CPCI and CPI can imply rich semantic information. Correspondingly, the partners involved in the two metapaths contain more precise semantic information than the fields and rounds of startups. Therefore, the partners of startups could be followed when recommending investors for startups. For example, as for the startup Ziroom shown in Table 1, if we recommend investors for it based on its partners, rounds, or fields, its partners Lin Xiong and Guowei Li could contain more useful semantic information than its field Real Estate and its round A, and we could obtain good recommendation performance by mining its partners.

(2) The fields of startups include richer related information about investor recommendation than the rounds.

Tables 4 and 5 show that the recommendation based on the metapaths CFCI and CFI performs better than the metapaths CRCI and CRI, which illustrates that the metapaths CFCI and CFI include more plentiful information than the metapaths CRCI and CRI. Along this line, the fields of startups contain richer information than the rounds. Thereby, we could focus on their fields in the investor recommendation task. Also, for the startup Ziroom shown in Table 1, if we recommend investors on the basis of its rounds or fields, its field Real Estate could contribute more to the recommendation task than its round A, and we could get good recommendation effectiveness by utilizing its field.

(3) The companies with the same partners as the target startup could contribute much more information to the investor recommendation task, than its partners.

As shown in Table 4, the investor recommendation based on the VC-metapath CPCI is more effective than the investor recommendation based on the metapath CPI. The results indicate that it will make the relative ranking of relevant investors in the corresponding investor recommendation set more accurate when we mine the companies with the same partners as the target startup and recommend investors for the target startup. However, when we focus on the partners of startups and recommend investors for the target startup who have invested in the partners of the target startup, the accuracy of the relative ranking of relevant investors will be reduced.
Therefore, the companies with the same partners as
the target startup could contribute much more in-
formation to the investor recommendation task,
than its partners. For example, if we recommend
investors for the startup Ziroom shown in Table 1
based on its partners, Lin Xiong and Guowei Li, we
could first focus on the companies with partners Lin
Xiong and Guowei Li and then recommend the
investors who have invested in those companies for
Ziroom.

(4) It will lead to worse performance, when the investor
recommendation task is conducted by simulta-
neously utilizing the companies with the same
partners as the target startup and the partners of the
target startup.

The experimental results shown in Tables 4 and 5
illustrate that the partners of startups are more
important than the other features, such as fields and
rounds of startups. As shown in Figures 8 and 9,
when the metapath weight parameter \( w_1 \) is 1 and \( w_2 \)
is 0, the RS and AUC of the investor recom-
menation based on the mixed metapath both perform
best. As the metapath weight parameter \( w_1 \) is 1 and
\( w_2 \) is 0, it indicates that the performance of investor
recommendation based on the single metapath CPCI
is better, compared with the result based on the
mixed metapath. It shows that the mixed metapath
generated by the metapaths CPCI and CPI will make
the semantic information contained in the metapath
fuzzier, which will lead to worse recommendation
performance. For example, if we recommend in-
vestors for the startup Ziroom shown in Table 1
based on its partners Lin Xiong and Guowei Li, we
define the investors who have invested in the
companies with the same partners as the first kind of
investor and the investors who are interested in the
partners as the second kind of investor, and we could
only recommend investors for Ziroom based on the
first kind of investor. Furthermore, it will make the
recommendation performance worse if we recom-
 mend investors for Ziroom on the basis of the two
kinds of investors.

6. Conclusions

In this paper, we investigate how to mine practical rules to
assist in further investment tasks, by measuring the impact
of startups’ features on the investor recommendation task.
Along this line, we propose a features’ contribution mea-
surement approach of startups on investor recommenda-
tion, named FECOM. Specifically, we first construct the
VC-HIN and design six VC-metapaths based on three
startups’ features, including partners, rounds, and fields.
Then, the two recommendation strategies contained in
FECOM could be accomplished by exploiting different
types of VC-metapaths. To this end, we could analyse the
impact of startups’ features on investor recommendation
by validating the recommendation performance and
summarize the practical rules beneficial for further in-
vestment tasks. Finally, four practical rules about the
startups’ features for the investment task are extracted
based on a real-world dataset, which can be employed to
guide in further investor recommendation.

However, if there is a large amount of VC data to be
processed, the proposed FECOM will be time consuming.
In the future, we will put efforts into reducing the time cost of
FECOM. Besides, we will study the optimal number of single
metapaths to generate the mixed metapath.

Data Availability

The data used to support the findings of this study are
available from the corresponding author via gao-
lian0001@163.com upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest
regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural
Science Foundation of China (no. 71971025).

References

n recommendation for venture finance,” in Proceedings of the
22nd ACM International Conference on Information &
Knowledge Management, pp. 1865–1868, ACM, San Francisco,
CA, USA, December 2013.
capital investment in startups’ sustainable growth and per-
formance: focusing on absorptive capacity and venture cap-
italists’ reputation,” Sustainability, vol. 12, no. 8, Article ID
innovation: what we know and how much we still need to
learn,” Journal of Economic Perspectives, vol. 34, no. 3,
investors for new startups by integrating network diffusion
and investors' domain preference,” Information Sciences,
and challenges,” in Proceedings of the 4th International
Conference on Employment, Education and Entrepreneurship
(EEE), Belgrade, Serbia, October 2015.
capital investment recommendation,” in Proceedings of the
9th ACM Conference on Recommender Systems, pp. 75–82,
Vienna, Austria, September 2015.
algorithm based on heterogeneous information network,”
International Journal of Computers Communications &
based on meta path ranking in heterogeneous information
network,” in Proceedings of the 7th IEEE International
Complexity


