AlphaVC: A Reinforcement Learning-based Venture Capital Investment Strategy

Hao Zhong
ESCP Business School
hzhong@escp.eu

Zixuan Yuan
HKUST (Guangzhou)
zixuanyuan@hkust-gz.edu.cn

Denghui Zhang
Stevens Institute of Technology
dzhang42@stevens.edu

Yi Jiang
ESCP Business School
yjiang@escp.eu

Shengming Zhang
Rutgers University
shengming.zhang@rutgers.edu

Hui Xiong
HKUST (Guangzhou)
xionghui@ust.hk

Abstract

Venture capital investments play a powerful role in fueling the emergence and growth of early-stage startups. However, only a small fraction of venture-backed startups can survive and exit successfully. Prior data-driven prediction based or recommendation based solutions are incapable of providing effective and actionable strategies on proper investment timing and amounts for startups across different investment rounds. In this paper, we develop a novel reinforcement learning-based method, AlphaVC, to facilitate venture capitalists' decision-making. Our policy-based reinforcement learning agents can dynamically identify the best candidates and sequentially place the optimal investment amounts at proper rounds to maximize financial returns for a given portfolio. We retrieve company demographics and investment activity data from Crunchbase. Our methodology demonstrates its efficacy and superiority in both ranking and portfolio-based performance metrics in comparison with various state-of-the-art baseline methods. Through sensitivity and ablation analyses, our research highlights the significance of factoring in the distal outcome and acknowledging the learning effect when making decisions at different time points. Additionally, we observe that AlphaVC concentrates on a select number of high-potential companies, but distributes investments evenly across various stages of the investment process.

Keywords: venture capital, reinforcement learning, portfolio optimization

1. Introduction

Venture capital (VC) firms play a crucial role in the entrepreneurial finance market, providing considerable impetus to the growth and development of startups. In recent years, there has been a significant increase in global venture capital investments, with investment amounts surpassing $670 billion in 2021 (KPMG, 2021). Despite the strides made by the venture capital industry, it continues to experience a high rate of investment failure. Evidence shows that only 25% of venture-backed startups (those that have received at least one round of venture capital investments) are able to survive and grow (Gage, 2012). This high investment failure ratio creates an empirical puzzle that attracts scholars' attention to the topic of VC decision-making.

Indeed, the literature on VC decision-making has been growing exponentially since 2009. The literature that investigates how VCs make decisions can be organized along three major traditions: (1) the trait-based view (e.g., Dimov et al., 2007; Malmström et al., 2017), (2) the cognition/behavior-based view (e.g., Baum & Silverman, 2004; Maxwell et al., 2011), and (3) the knowledge-based view (e.g., Carpentier & Suret, 2015; Zacharakis & Meyer, 2000). Each tradition provides factors that explain VCs' decision-making strategies. Although existing studies offer extremely rich insights into how VCs make decisions, our current understanding of VC decision-making faces three challenges. First, very few scholars investigate the overall decision ***quality***, i.e. whether invested ventures will receive the next round of investment and eventually go to IPO. Most scholars focus on understanding whether VCs have invested or not or whether VCs have the willingness to invest. Second, existing research treats each VC decision in isolation. This means that whatever VCs decide to do in one situation at time t does not influence their decision-making in another situation at time $t + 1$. We have little knowledge of the ***learning*** effect of an earlier investment decision on a later decision. Finally, as a consequence of not considering the distal outcome or the learning effect between actions, most research focuses on *demystifying* VC's decision-making process instead of *prescribing* good solutions that help VCs make better decisions.

In this paper, our primary aim is to empirically

URI: https://hdl.handle.net/10125/106906
978-0-9981331-7-1

examine the optimal way of investing. Specifically, we ask the following research question: *how could investors determine the timing and amount of investment in target startups?* To capture the dynamic nature of the investing process, we propose to tackle this problem from a reinforcement learning (RL) perspective. The main objective of our RL model, *AlphaVC*, is to instruct the agents (acting as investors) on how to distribute specific portions of funds among various startups across multiple rounds of investment. Extensive experimentation has been performed to assess the efficacy of our proposed AlphaVC model on datasets associated with two industry sectors, namely *Financial Services* and *Information Technology*. The experimental results show that AlphaVC outperforms other baseline models concerning both ranking and portfolio-based performance metrics. Our sensitivity analysis also shows that without considering the overall decision quality, i.e., whether the invested ventures will receive the next round of investment and reach IPO, and the agents' past decisions, AlphaVC would experience a sharp decline in her decisions to an optimal investment portfolio. Moreover, through a sensitivity analysis on investment strategy parameters and through comparing AlphaVC with human agents, we find that AlphaVC focuses its attention on a few high-potential companies but disperses the investment amount in a balanced manner across different stages in the investment process.

2. Theoretical background

2.1. Literature of VC Decision Making

The literature around VC decision-making can be organized along three major traditions: (1) the trait-based view, (2) the cognition/behavior-based view, and (3) the knowledge-based view.

In the first tradition (the trait-based view), scholars primarily examine the influence of entrepreneurial characteristics, such as gender and passion, on investment decisions (e.g., Malmström et al., 2017). Scholars also investigate how the traits of VC investors, such as their consistency, finance capacity, portfolio, reputation, and confidence level (e.g., Drover et al., 2017; Petty & Gruber, 2011; Zacharakis & Meyer, 2000) influence their investment decisions. In this approach, good decision-making is assumed to belong to a particular type of VC. The point of this tradition has been to enumerate a set of characteristics that could describe good VCs as an entity. Those VCs with good reputations, good financial capability, and who are confident are more likely to make good decisions. Underlying this tradition is the assumption that only the VCs with the right traits could make good decisions.

The second strand of research, the cognitive/behavior-based view, shifted the focus from who the right VCs are to what the VCs do. In this tradition, research focuses on looking at cognitive antecedents of VC decision-making processes. For example, the research investigates the link between investment decisions and the similarities between venture capitalists and members of a venture team (e.g., Forlani & Mullins, 2000). Other factors such as perceived uncertainty (e.g., Forlani & Mullins, 2000), escalation of commitment (Devigne et al., 2016), and cognitive biases on investment decisions (e.g., Baum & Silverman, 2004; Maxwell et al., 2011) are also considered factors that affect VC decision-making quality. In general, the cognitive/behavior-based view proposes a variety of decision-making criteria that influence investors' willingness to invest.

Finally, in the stream of knowledge-based view, researchers focus on studying how VC's accumulated knowledge influences VC investment decisions. Specifically, scholars aim to understand which types of past experiences in startup teams can lead to greater success and interest from VCs and how VCs' experiences in specific industries or with certain types of companies may inform their investment decisions. For example, factors such as past financial expertise (e.g., Baum & Silverman, 2004; Dimov & De Clercq, 2006), past VC experience (e.g., Shepherd & Zacharakis, 2002; Shepherd et al., 2003; Wesley II et al., 2022), past industrial experience (e.g., Franke et al., 2008; Wesley II et al., 2022), social network (e.g., Wang, 2016) are all examined to understand how and why VCs make decisions to select and to invest in entrepreneurs.

2.2. Challenges of Understanding VC Decision Making

Although existing studies offer extremely rich insights into how VCs make decisions, our current understanding of VC decision-making faces three challenges. First, the empirical focus of the VC decision-making process is incomplete (McMullen & Dimov, 2013). Most existing papers focus on proximate outcomes, measured by whether VCs make the decisions to invest (e.g., Dimov & De Clercq, 2006; Franke et al., 2008; Wang, 2016) or whether VCs have the willingness to invest (e.g., Drover et al., 2017; Wesley II et al., 2022). Very few scholars investigate distal outcomes of VCs' decisions, such as the overall decision ***quality***, i.e. whether the invested ventures will receive next round of investment and eventually go to IPO. Without knowing the overall quality of VC decisions, we cannot advise how VCs could make investment decisions in practice to enhance new venture survival and success.

Second, the lack of empirical focus on studying the quality of VC decisions may be due to the persistence of early conceptualization that a VC investment decision is a single-moment event, which happens in isolation from the overall decision-making process. Indeed, existing research treats each VC decision in isolation. This means whatever VCs decide to do in one situation at time t does not influence their decision-making in another situation at time $t + 1$. Under this assumption, we could not tell, for example, how a VC could sequentially ***learn*** to make better decisions of investing based on their past actions. For example, in the behavior/cognition research strand, research typically focuses on specific moments when decisions were made. We have little knowledge about whether a VC's decision to invest in a venture at a nascent stage has any influence on the VC's decision to invest when the venture is at a later stage. Even for longitudinal studies that intend to capture the dynamic nature of the investing process (McMullen & Dimov, 2013), scholars rarely assume any learning effect among decisions. In fact, most scholars map out decision points without talking about how earlier decisions influence latter ones (e.g., Carpentier & Suret, 2015; Dimov & De Clercq, 2006; Petty & Gruber, 2011; Wang, 2016).

Third, the focus on approximate outcomes may be a result of empirical limitations that have inhibited scholars' ability to "see" over a longer span of time. That is, most research does not take objective retrospective historical data into account; rather, scholars rely on survey, experimental, and longitudinal data generated through VC's own perceptions (Gompers et al., 2009; Nanda et al., 2020). The existing research designs do not deepen our ability to recognize important patterns of possible investment actions and their consequences. Retrospective data, on the other hand, can provide us with more fine-grained levels of analysis, such as temporal dynamics of decisions effects over distal outcomes. To date, research has clearly highlighted the importance of considering VC decision-making process. Evidence shows that there are differences in the apparent relevance of criteria at different stages of the VC decision-making process. For example, Petty and Gruber (2011) find that the main reasons for rejecting a business proposal in the early stages of the fund lifecycle are not the same as the main reasons for rejection later on in the life of the venture fund. Because VC decision-making is a complex and long-term task that often involves multiple rounds of investment spanning over several years (e.g., Dimov & De Clercq, 2006; Petty & Gruber, 2011; Wang, 2016), it becomes necessary for scholars to consider the process aspect of VC decision-making. The consequence of not considering process is that most research focuses on *demystifying* VC's decision-making process instead of *prescribing* good solutions that help VC make decisions. Therefore, to capture the dynamic nature of the investing process, we propose a new paradigm of methods that could advance our current understanding of VC decision-making.

2.3. Addressing Theoretical and Empirical Incompleteness

To address the theoretical and empirical incompleteness, we propose to study VC decision-making in two steps. First, we use real, enriched VC-investing-startup data that includes IPO labels to do research. In addition to collecting information on whether a VC invested in a company, we also gather data on the subsequent development of the invested company, such as whether it progresses to the next funding round and eventually achieves a successful IPO. Such data mining approaches have become popular and have been used to predict business success, such as company strategy (Yankov, 2012), collective intelligence (Dellermann et al., 2021), ventures' success rate (Bonaventura et al., 2020; Zhang et al., 2021). By conducting analyses on this more comprehensive data, we aim to gain a new understanding of how to assist VCs in making *qualified* investment decisions.

Second, building on the rich empirical data, we introduce the Markov Decision Process (MDP) as a mathematical framework for modeling decision-making. MDP falls under reinforcement learning, which is defined as actors learning what to do - how to map situations to actions - so as to maximize a numeral reward signal (Sutton & Barto, 2018). Scholars have used reinforcement learning in risk management (Buehler et al., 2019), optimizing portfolio (Moody & Saffell, 2001), and asset allocation (Almahdi & Yang, 2017), among others. In our case, we use MDP as a reinforcement learning technique and look at how VC investors take different actions to maximize reward under different situations.

2.4. Technical Methods for VC Investment

There are numerous studies aiming to seek signals of business success using machine learning and data mining approaches. Yankov (2012) studied 42 success prediction models and major success factors, identifying industry structure, company strategy, and interaction of strategy with structure as a successful model pattern. Hadley et al. (2018) instead included a people-centric network for analyzing the startup's success. Zhang et al. (2021) modeled VC firms, people, and startups into a heterogeneous business information network and used a scalable heterogeneous graph Markov NN to predict if early-stage startups could receive a series-A round.

Additional research that is highly pertinent to our study includes portfolio optimization through the application of reinforcement learning techniques. This line of research can be traced back to (Moody & Saffell, 2001) and has drawn more attention due to the heat of deep neural network-boosted reinforcement learning. Besides, Almahdi and Yang (2017) adopts recurrent reinforcement learning (RRL) for asset allocation and variable weight portfolio allocation. Their proposed coherent risk-adjusted objective function yields better return performance than other objective functions, i.e. the Sharpe ratio and the Sterling ratio.

3. Research Problem and Data

Supposed that we consider N investment rounds $\{t_1, t_2, ..., t_N\}$ and a predetermined set of M startups $\{c_1, c_2, ..., c_M\}$ for investment. Our objective is to allocate a proportionate amount of funding to a selected group of companies across various rounds, with the aim of maximizing the financial return within the portfolio. The output of the model is an allocation matrix, denoted as $\mathbf{A} \in \mathbb{R}^{N \times M}$, where each element $a_{ij} \in [0, 1]$ indicates the proportion of the total funding assigned to company c_j in round t_i, subject to the constraint of $\sum_{ij} a_{ij} = 1$.

Our data is sourced from *Crunchbase*, a widely acclaimed dataset for its comprehensive coverage of entrepreneurial activities and venture capital investments. This platform provides access to a vast network of market intelligence and analytics spanning a broad range of industries and geographies. We utilize Crunchbase to gather a substantial amount of information about startups, investors, and investment activities for our study. We collected data on companies founded between 2010 and 2018, as well as investment activities during the same period. Our data covers a wide range of industries, including financial services, information technology, healthcare, and consumer goods. These companies are located in various countries, including but not limited to the United States, China, and the United Kingdom. To obtain a more comprehensive characterization of the startups, we undertake feature engineering with our sample data. Table 1 presents the features focusing on four perspectives of characterizing the startups, including investments, investors, location, and industry (Sharchilev et al., 2018; Xu et al., 2022).

4. Research Methodology

4.1. Startup Screening

To facilitate subsequent training of the RL model, it is imperative that we undertake a screening process to eliminate startups that are unlikely to succeed in the follow-on round. This will enable us to focus the efforts and resources of our RL model on startups that have a greater potential for success and a more promising future. By incorporating a binary label indicating the likelihood of a company receiving funding in the subsequent round, we can train our XGBoost model effectively using the prepared features (see Table 1) and make accurate predictions about the funding prospects of the startups. Compared to optimizing a sequential VC portfolio, predicting follow-on investment rounds is a simpler task that could potentially yield higher accuracy. Using this approach and integrating the predictions generated by the XGBoost model into our framework (as illustrated in Figure 1), the RL agent can better assign funding ratios to those unmasked companies and reduce the frequency of errors.

4.2. RL-based Investment Strategy Learning

4.2.1. States. In our setting, the *agent* refers to the venture capital investor. The state of our VC agent includes relevant investment details of the VC environment being observed, as well as historical decision data of the agent. Assuming there are a total of T investment rounds, for each round $t \in \{1, 2, ..., T\}$, the VC agent's state is denoted as $\mathbf{S}^t \in \mathbb{R}^{M \times 2d}$. At each investment round t, we establish the agent's state $\mathbf{S}^t$, which consists of two parts: the *factual state* $\mathbf{S}^t_f \in \mathbb{R}^{M \times d}$ and the *decision state* $\mathbf{S}^t_d \in \mathbb{R}^{M \times d}$. The agent states are formed by concatenating both states:

$$\mathbf{S}^t = \mathbf{S}^t_f \oplus \mathbf{S}^t_d, \tag{1}$$

where $\oplus$ is the concatenation operator. The factual state $\mathbf{S}^t_f$ typically includes the financial features of the company and industry. It is calculated based on the factual statistical features of companies in a given pool. The decision state $\mathbf{S}^t_d$ contains information about the agent's historical investment decisions. It is calculated using the hidden states of the Long Short-term Memory Networks (LSTMs) (Hochreiter & Schmidhuber, 1997) in the agent's policy network (see Section 4.2.2).

Factual states. The factual state $\mathbf{S}^t_f$ contains factual information collected from all M startups in the pool. Specifically, to form the initial representation of $\mathbf{S}^t_f$ during each investment round t, we use the $K = 14$ features as summarized in Table 1, namely $\{\mathbf{f}^t_1, \mathbf{f}^t_2, ..., \mathbf{f}^t_K\}$, that can be gathered and computed for each company at every investment round. The initial representation is denoted as $\widetilde{\mathbf{S}}^t_f \in \mathbb{R}^{M \times K}$ and is computed by concatenating the features together, i.e., $\mathbf{f}^t_1 \oplus \mathbf{f}^t_2 \oplus, ..., \oplus \mathbf{f}^t_K$. Subsequently, a fully-connected layer is applied to the initial fourteen-dimensional feature vector, which projects the vector into a hidden space consisting of d dimensions, resulting in the formation of $\mathbf{S}^t_f \in \mathbb{R}^{M \times d}$:

$$\mathbf{S}^t_f = \sigma(\mathbf{W}_1 * \widetilde{\mathbf{S}}^t_f + \mathbf{b}_1), \tag{2}$$

Table 1: Summary of Our Extracted Features

Category	Feature	Remarks
Investments	INVEST_COUNT	Total number of investments it has received
	INVEST_FREQ	Average months per investment since its establishment
	LAST_INVEST_AMOUNT	Amount of the last investment
	LOG_LAST_INVEST_AMOUNT	Logorithmed amount of the last investment
Investor	AVG_INVESTORS_PER_ROUND	Average number of investors per investment round
	AVG_INVEST_COUNT_BY_INVESTOR	Average number of the past investments per its investors
	REPEATED_INVESTOR_COUNT	Number of investors who has invested this company more than once
Location	IPO_RATIO_BY_CITY	Ratio of companies IPOed in its city
	IPO_RATIO_BY_CITY_1Y	Ratio of companies IPOed in its city in recent 1 year
	IPO_RATIO_BY_CITY_3Y	Ratio of companies IPOed in its city in recent 3 year
Industry	AVG_IPO_RATIO_BY_INDUSTRY_1Y	Average ratio of IPO exits among its industry terms in recent 1 year
	AVG_ACQ_RATIO_BY_INDUSTRY_1Y	Average ratio of companies which got acquired among its industry terms in recent 1 year
	AVG_IPO_RATIO_BY_INDUSTRY_3Y	Average ratio of IPO exits among its industry terms in recent 3 year
	AVG_ACQ_RATIO_BY_INDUSTRY_3Y	Average ratio of companies which got acquired among its industry terms in recent 3 year

where $\mathbf{W}_1 \in \mathbb{R}^{d\times K}$ and $\mathbf{b}_1 \in \mathbb{R}^{M\times d}$ denote the parameters of the fully-connected layer, and σ denotes the common sigmoid function for non-linear activation.

Decision states. Unlike factual states, which only contain factual information about the environment, *decision states* serve the purpose of encapsulating the past decision-making behavior of the RL agent. In our agent's policy network, decision states are derived from the hidden states generated in the previous time step (denoted as $\mathbf{h}_{t-1} \in \mathbb{R}^{M\times d_h}$) using the LSTM modules. The decision states are calculated as follows:

$$\mathbf{S}_d^t = \sigma(\mathbf{W}_2 * \mathbf{h}_{t-1} + \mathbf{b}_2), \tag{3}$$

where $\mathbf{W}_2 \in \mathbb{R}^{d\times d_h}$ and $\mathbf{b}_2 \in \mathbb{R}^{M\times d}$ are the parameters of the second fully-connected layer, which projects the hidden states to a d-dimensional space. LSTMs are ideal for this problem as they can capture both current and past information about an agent's hidden states. Further information regarding the LSTM networks can be found in Section 4.2.2.

Actions. In our context, we assume that the RL agent will take a sequence of T actions corresponding to T investment rounds. Suppose that we have M startups in the pool, each of these actions is equivalent to a specific row $\mathbf{A}^t$ in the funding allocation matrix $\mathbf{A} \in \mathbb{R}^{T\times M}$. However, each $\mathbf{A}^t$ $(t \in \{1, 2, ..., T\})$ is independent of the others, we inadvertently sacrifice the preservation of sequential patterns across different rounds.

To mitigate this issue, we have established a two-part framework for the agent's actions. These are defined as the *stage-wise action* denoted by $\mathbf{a}_s \in \mathbb{R}^{T\times 1}$ and the *company-wise action*, which is represented by $\mathbf{A}_c \in \mathbb{R}^{T\times M}$. We can then obtain the final funding allocation matrix $\mathbf{A}$ by combining these two actions. A stage-wise action involves investors determining the allocation of funding at each round based on the total available funds. A company-wise action involves determining the percentage of funding that each individual company receives based on the prior stage-wise action.

Stage-wise action. For each round t, the scalar $a_s^t \in \mathbf{a}_s$ represents the total funding investment ratio at the current round. The stage-wise action adheres to the constraint $\sum_{t=1}^{T} a_s^t = 1$ to ensure that all funding is invested by the last round. Our empirical results suggest

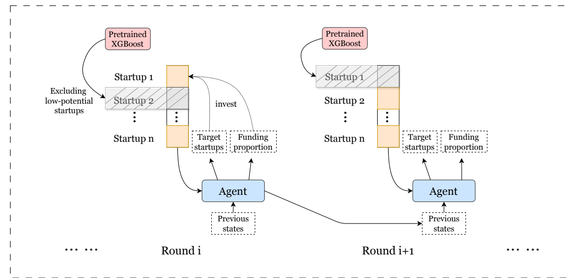

Figure 1: Our RL-based Model Framework

restricting $0 \le a_s^t \le 0.5$ to prevent potential bias in the loss function during model training. We found that this bias could occur when assigning disproportionate weight to an early investment stage.

Company-wise action. Each investment round t involves a company-wise action $\mathbf{a}_c^t \in \mathbf{A}_c$, which is a vector with M dimensions, where M represents the number of companies in the pool. Specifically, the value of each element $\mathbf{a}_c^t[i]$ ranges from 0 to 0.5 to indicate the proportion of funding invested in the i-th company for the current round, where $i \in \{1, 2, ..., M\}$. It is imperative that the sum of all elements in $\mathbf{a}_c^t$, $\sum_{i=1}^{M} \mathbf{a}_c^t[i]$, equals 1 to conform to our constraint.

The final allocation matrix $\mathbf{A}$ can be computed through a series of steps that involve the orthogonal actions described above. Specifically, each row $\mathbf{A}^t$ can be derived by taking the product of the column vector $\mathbf{a}_c^t$ and the corresponding scalar a_s^t. By repeating this process for all rounds, we can generate the entire allocation matrix $\mathbf{A}$ that meets the required constraints.

4.2.2. Policy network. We present our framework in Figure 1 to provide a more comprehensive overview. Our framework involves assessing the state of all companies in the investment pool and passing this information to the agent's policy network in every round of investment. Before investment decisions are made at each round, we first utilize an XGBoost model to identify and exclude companies that are likely to fail in the subsequent rounds (as discussed in Section 4.1). The agent's responsibility is then to make projections for the best funding allocation matrix ($\mathbf{A}$) that designates the most suitable companies to invest in and determine the corresponding ratio of funding. Its final goal is to achieve the highest possible cumulative reward.

LSTM network. Within our policy network, it is imperative to properly capture the sequential patterns in the agent states. Following its original definition (Hochreiter & Schmidhuber, 1997), given a sequence of input data $\mathbf{x}_t \in \mathbb{R}^n$, a memory cell $\mathbf{c}_t \in \mathbb{R}^d$ and a hidden state $\mathbf{h}_t \in \mathbb{R}^d$, we denote the LSTM unit as:

$$(\mathbf{h}_t, \mathbf{c}_t) = \mathbf{LSTM}(\mathbf{h}_{t-1}, \mathbf{c}_{t-1}, \mathbf{x}_t, \theta), \tag{4}$$

where θ represents all the model parameters. In our policy network, the input at each time step for the LSTM network is the agent state at the current round: $\mathbf{x}_t = \mathbf{S}^t$.

Multi-task prediction layer. In addition to capturing sequential patterns in the agent states, it is also imperative to develop a component that predicts the next stage-wise and company-wise actions. We thus design the following multi-task prediction layer based on the current LSTM hidden state $\mathbf{h}_t$ and RL agent state $\mathbf{S}^t$:

$$F_1(\mathbf{A}_c^t | \mathbf{S}^t) = \sigma(\mathbf{S}^t \, \mathrm{ReLU}(\mathbf{W}_3 \mathbf{h}_t)), \tag{5}$$

$$F_2(\mathbf{a}_s^t | \mathbf{S}^t) = \sigma(\mathbf{W}_5 \, (\mathbf{S}^t \oplus \mathrm{ReLU}(\mathbf{W}_4 \mathbf{h}_t))), \tag{6}$$

where $\mathbf{W}_3 \in \mathbb{R}^{2d \times d_h}$, $\mathbf{W}_4 \in \mathbb{R}^{2d \times d_h}$, $\mathbf{W}_5 \in \mathbb{R}^{2d}$, are the matrices of learnable weights. $ReLU(\cdot)$ and $\sigma(\cdot)$ denote the ReLU activation function and softmax layer, respectively. The function $F_1(\cdot)$ is responsible for predicting the company-wise actions, while $F_2(\cdot)$ is tasked with predicting the actions at the stage level.

4.2.3. Reward. In the context of optimizing a VC portfolio, the reward should align with the potential financial returns to investors while portfolio companies continue to grow and expand. Although the most reliable option for the reward function is to utilize the realized investment return, it is challenging to obtain the actual investment return due to the non-disclosure of startup valuations before or after investment rounds. Consequently, an alternative option is to use significant milestone events, such as acquisitions or public listings, as proxies for successful exits to approximate the investment return. However, these events are sparse, and our experiments reveal that they have limited capacity to improve the RL model effectively.

To this end, we design the reward to capture startups' ability to secure *subsequent investments* to determine their sustained pattern of growth and success over time. We define the reward of our RL model as follows:

$$\mathcal{L}_{reward} = \mathbb{E}_{s, A_i^t} \sim \pi(\theta; \cdot) \sum_{t=t_i+1}^{N} \gamma^{t-(t_i+1)} H(\mathbf{A}_i^t), \tag{7}$$

where $\pi(\theta; \cdot)$ denotes the previously defined policy network, t_i indicates the current investment round of startup i, and γ is the discounted rate. In particular, $H(\mathbf{A}_i^t)$ is an indicator of whether the startup secures a follow-on investment:

$$H(\mathbf{A}_i^t) = \begin{cases} 1, & \text{receiving a follow-on investment} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Note that "receiving a follow-on investment" indicates that the startup has successfully obtained additional funding for a subsequent round of investment.

To promote diversity and prevent the RL agent from concentrating on a restricted set of startups, a regularization term is added into the loss function. This term uses the KL-divergence $\mathrm{KLD}(\cdot, \cdot)$ to evaluate the similarity between two investment decision matrices $\mathbf{A}^t$ and $\hat{\mathbf{A}}^t$ on a round-by-round basis:

$$\mathcal{L}_{div} = \frac{1}{T} \sum_{t=1}^{T} \mathrm{KLD}(\mathbf{A}^{\mathrm{t}}, \hat{\mathbf{A}}^{\mathrm{t}}), \tag{9}$$

where $\hat{\mathbf{A}}^t$ denotes the estimated $\mathbf{A}^t$. Finally, the training objective is to minimize the loss:

$$\mathcal{L} = -(\mathcal{L}_{reward} - \beta \mathcal{L}_{div}), \tag{10}$$

where β is a hyperparameter that determines the significance of diversity regularization.

5. Strategy Evaluation

5.1. Experimental Setup

Data Summary. Our experimental study specifically targets the **Financial Services ($\mathcal{FS}$)** and **Information Technology ($\mathcal{IT}$)** sectors. We have provided the data summaries for these two industries in Table 2. The datasets comprise approximately 10,000 investment transactions between venture capital firms and startups over a span of 20 years (2000-2021). The $\mathcal{FS}$ dataset consists of around 2.6K startups and 1.9K investors, whereas the $\mathcal{IT}$ dataset has a slightly higher volume (3.7K startups and 2.4K investors). Also, we observe that the $\mathcal{FS}$ dataset has slightly more investment rounds per startup (4.01) compared to the $\mathcal{IT}$ dataset (2.99). Additionally, the average amount of funds raised per startup in the $\mathcal{FS}$ dataset (USD 16.5M) is significantly greater than that in the $\mathcal{IT}$ dataset (USD 7.1M). Further information regarding the datasets is in Table 2.

Baselines. We have conducted a comparative analysis using two existing lines of research. The first group includes various ML models designed for one-step-ahead forecasting, i.e., predicting the follow-on investment round. This category includes the following baselines: Logistic Regression (Kleinbaum et al., 2002), LSTM (Hochreiter & Schmidhuber, 1997), GRU (Chung et al., 2014), and XGBoost (Chen & Guestrin, 2016). The second set comprises RL-driven methodologies for making sequential decisions, such as DQN (Mnih et al., 2015), Dueling DQN (D-DQN) (Wang et al., 2016), PPO (Schulman et al., 2017), and DDPG (Lillicrap et al., 2019).

	Financial Services ($\mathcal{FS}$)				Information Technology ($\mathcal{IT}$)			
# Total Investment Records	10,482				11,006			
# Total Companies	2,614				3,681			
# Total Investors	1,887				2,353			
	Mean	**Std**	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**
# Invested companies per investor	1.57	1.73	1	42	1.42	1.27	1	31
# Investments received per company	4.01	4.75	1	43	2.99	3.17	1	31
# Investment rounds received per company	1.36	0.72	1	6	1.27	0.61	1	5
\$ Money raised per company	16.5M	67.2M	1,000	1,435.9M	7.1M	22.1M	524.29	350.6M
\$ Money raised per company per year	11.5M	42.1M	1,000	1,104.6M	5.4M	16.0M	524.29	350M

Table 2: Summary of Our Data Sample

	Financial Services ($\mathcal{FS}$)				Information Technology ($\mathcal{IT}$)			
	TR	RC	Hits@10	RR	TR	RC	Hits@10	RR
LR	0.086	1.560	0.321	0.549	0.042	1.153	0.183	0.474
GRU	0.040	1.226	0.167	0.696	0.060	1.333	0.400	0.757
LSTM	0.077	1.375	0.500	0.959	0.078	1.396	0.308	0.945
XGBoost	0.129	1.095	**0.964**	0.921	0.137	1.188	**0.992**	1.043
DQN	0.208	1.218	0.364	0.208	0.176	1.546	0.273	0.176
D-DQN	**0.404**	1.091	0.636	0.404	**0.349**	1.407	0.455	0.349
PPO	0.208	1.250	0.333	0.208	0.208	1.250	0.333	0.208
DDPG	0.150	1.273	0.273	0.150	0.208	1.273	0.364	0.208
AlphaVC	0.311	**1.667**	0.852	**1.202**	0.316	**1.672**	0.857	**1.507**

Notes. We highlight the best approach in bold and the runner-up underlined.

Table 3: Overall Performance

Evaluation metrics. (1) Total rewards (TR) computes the total rewards received by the agent, given by TR $= \sum_{i=1}^{M_s} \sum_{t=1}^{T} \gamma^{t-(t_i+1)} H(\mathbf{A}_i^t)$, where M_s is the number of selected companies.

(2) Average investment round count (RC) calculates the average number of investment rounds, given by RC $= \frac{1}{M_s} \sum_{i=1}^{M_s} T_i$, where T_i stands for the number of investment rounds considered for company i.

(3) Hits@k represents the percentage of companies receiving follow-on investments that are also ranked within top-k positions: $Hits@k = \frac{P_{u,q} \cap R_{u,q}(k)}{k}$, where P_c is a set of selected companies, and $R_c(k)$ records the top-k matched invested companies. Here $k = 10$.

(4) Risk-adjusted Return (RR) indicates the expected differential return per unit of systematic risk: $RR = \mathbb{E}_r/\sigma_r$, where $\mathbb{E}_r$ is the average return and σ_r is the weighted average of the probabilities of XGBoost-based positive predictions at each round.

5.2. Main Results

We present the performance results of both the baseline models and our proposed AlphaVC in Table 3. There are several interesting findings. First, our model has demonstrated outstanding performance in discovering high-potential startups, as evidenced by achieving either the highest or runner-up performance on both datasets across all metrics. This finding reaffirms that our model excels in overall performance, proving its reliability in effectively identifying high-potential startups. We have also noted that our model exhibits a considerable lead over the runner-up with respect to RC and RR performance metrics. This observation implies that our model has a propensity to take into account more investment rounds (higher RC), resulting in an increase in portfolio return over time (higher RR). Second, regarding the evaluation metrics of Hits@10, it was observed that XGBoost exhibits the best performance while our model ranks second. Although XGBoost is highly effective in identifying startups that are likely to receive follow-on investments, it has exhibited limited potential in generating enduring investment returns, as reflected by its inferior performance on the RR metric compared to our model. Third, our AlphaVC has demonstrated consistently strong performance across multiple industries, which attests to the resilience of our portfolio-based returns optimization strategy. This stability is a testament to the efficacy of our framework in producing dependable and consistent results.

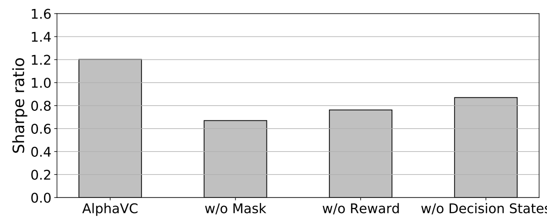

Figure 2: Ablation Analysis

5.3. Ablation Analysis

In our study, we utilize the ablation analysis to investigate the effect of 1) the XGBoost-based masking module, 2) the reward function, or 3) the decision states on the performance of our AlphaVC. We present the findings of our ablation analysis for the $\mathcal{FS}$ dataset are presented in Figure 2. Similar results are also achieved with $\mathcal{IT}$, affirming the consistency of our conclusion.

Our investigation begins with the unique design of the masking module based on XGBoost. The primary goal of this module is to eliminate low-potential companies, thus reducing the search space for the agent. Lacking this module could potentially impede the agent's ability to effectively identify the most promising prospects, given a larger pool. The outcomes obtained from the figure's comparison between AlphaVC and "w/o Mask" led us to affirm that the exclusion of the XGBoost-based masking module results in a substantial

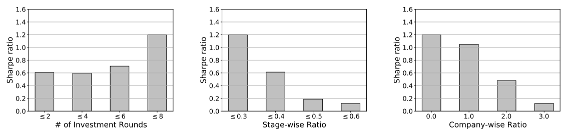

Figure 3: Varying Investment Strategy Parameters

decrease in our model's performance. The results provide compelling evidence that the module is an indispensable component of our model.

The reward function is designed to direct the agent towards making coherent investment decisions, thereby enhancing their investment returns. Without it, the agent would not expect distal investment returns from the portfolio companies, similar to VCs who fail to take into account future investment returns. Our model's performance significantly declines when we compare it to ”w/o Reward” in Figure 2, highlighting the crucial role of the reward function in our model. Ignoring the future returns can lead to poor investment decisions, i.e., the overall ***quality*** of the investment decisions.

The decision states in our model encapsulate the investment decisions made by the agent in the past, and removing them would result in the agent disregarding past actions when making future investment decisions. This approach is disadvantageous for the agent since previous investment experience could provide valuable insights into future investment scenarios. We find that our model's performance experiences a significant decrease when we compare it to “w/o Decision States” in Figure 2. This finding underscores the importance of past investment decisions and implies that an investor could sequentially ***learn*** to make better investment decisions based on their past actions.

5.4. Sensitivity Analysis on Investment Strategy Parameters

The results presented in Figure 3 evaluate the impact of three crucial parameters of the investment strategy on the model performance: the number of investment rounds, the stage-wise investment ratios, and the investment diversity coefficient. First, we vary the maximum number of investment rounds from 2 to 8 to examine its effect on the model's performance. Note that a greater number of investment rounds grants the agent more flexibility in making investment decisions concerning stages. The result shows a noticeable upward trend in the model's performance as the parameter's upper bound increases. This illustrates that our agent is capable of making intelligent decisions when confronted with more adaptable investment scenarios. Second, we find that a smaller upper bound for stage-wise ratios is linked to higher future rewards. This result supports the notion that a diversified investment strategy across stages leads to better outcomes, as discussed in our first finding above. As such, we suggest that investors diversify their investments across different stages to maximize their returns. Third, the company-wise diversity coefficient has a significant impact on performance. Specifically, we find that smaller coefficients lead to a higher level of performance. This finding suggests that a viable strategy for achieving optimal results may involve limiting an agent's focus to a few companies with high potential rather than spreading investments across a larger number of mediocre companies. By doing so, agents can potentially achieve superior results and enhance the overall performance of their investment portfolios.

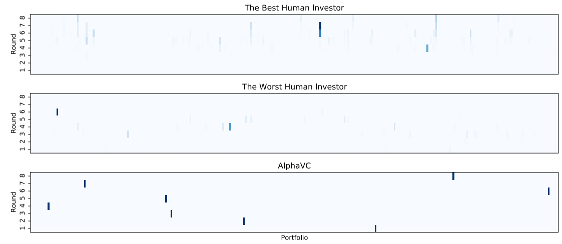

Figure 4: Human Investors vs. AlphaVC

5.5. Investment Strategies of AlphaVC vs. Human Investors

We perform a comparative analysis of the investment strategies developed by AlphaVC and two human investors: the best and the worst performers in the $\mathcal{FS}$ sector. Figure 4 displays heatmaps depicting their investment strategies, with each heatmap showing the investment rounds (up to eight) on the vertical axis and portfolio companies on the horizontal axis. The investment activities on the portfolio companies are represented by tiny vertical sticks, where a darker shade indicates a higher proportion of funding. Then we discovered several interesting findings. Notably, the top-performing human investor appears to be risk-averse and devotes a higher proportion of their investments towards later-stage companies. The worst performer appears to exercise greater caution, investing in fewer companies in general. We also discover that both human investors made significant investments in a small pool of companies, evidenced by the lack of dark sticks in their heatmaps. On the other hand, our AlphaVC agent tends to distribute money across an adequate number of companies that span various stages – from early to late stages. This investment approach emphasizes balance and diversification, with the objective of achieving superior future returns.

6. Conclusion

In this paper, we developed a novel reinforcement learning-based model, *AlphaVC*, to aid venture capitalists in making intelligent investment decisions.

To demonstrate the effectiveness of our method, we sourced VC investment data from Crunchbase and conducted an assessment of our model within *Financial Services* and *Information Technology* industry sectors. Our method demonstrated clear superiority over various baseline methods, as shown by the experimental results, which evaluated both ranking and portfolio-based performance metrics. Through the ablation analyses, we show the importance of considering the distal outcome and of assuming a learning effect in between decisions made at different points in time. Compared to other available solutions, our proposed approach offers a significant advantage of intelligently determining the optimal investment timing and amount to maximize financial returns across the entire portfolio.

We are also aware of several limitations of our work. First, VC investment strategies may depend on the financial performance of portfolio companies, which, unfortunately, is unavailable in our current dataset. As part of our future plans, we intend to explore alternative data sources that provide financial performance information for portfolio companies to improve our RL model. Second, our current evaluation metrics primarily focus on assessing which startups to invest in, such as TR and Hits@k. Another question pertains to the extent to which the model effectively allocates funds in proportion. In the future, it is worth considering additional metrics in assessing the allocation of investment in an optimized portfolio.

References

Almahdi, S., & Yang, S. Y. (2017). An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. *Expert Systems with Applications*, *87*, 267–279.

Baum, J. A., & Silverman, B. S. (2004). Picking winners or building them? alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups. *Journal of Business Venturing*, *19*(3), 411–436.

Bonaventura, M., Ciotti, V., Panzarasa, P., Liverani, S., Lacasa, L., & Latora, V. (2020). Predicting success in the worldwide start-up network. *Scientific Reports*, *10*(1), 1–6.

Buehler, H., Gonon, L., Teichmann, J., & Wood, B. (2019). Deep hedging. *Quantitative Finance*, *19*(8), 1271–1291.

Carpentier, C., & Suret, J.-M. (2015). Angel group members' decision process and rejection criteria: A longitudinal analysis. *Journal of Business Venturing*, *30*(6), 808–821.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling.

Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., & Leimeister, J. M. (2021). Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method.

Devigne, D., Manigart, S., & Wright, M. (2016). Escalation of commitment in venture capital decision making: Differentiating between domestic and international investors. *Journal of Business Venturing*, *31*(3), 253–271.

Dimov, D., & De Clercq, D. (2006). Venture capital investment strategy and portfolio failure rate: A longitudinal study. *Entrepreneurship Theory and Practice*, *30*(2), 207–223.

Dimov, D., Shepherd, D. A., & Sutcliffe, K. M. (2007). Requisite expertise, firm reputation, and status in venture capital investment allocation decisions. *Journal of Business Venturing*, *22*(4), 481–502.

Drover, W., Busenitz, L., Matusik, S., Townsend, D., Anglin, A., & Dushnitsky, G. (2017). A review and road map of entrepreneurial equity financing research: Venture capital, corporate venture capital, angel investment, crowdfunding, and accelerators. *Journal of Management*, *43*(6), 1820–1853.

Forlani, D., & Mullins, J. W. (2000). Perceived risks and choices in entrepreneurs' new venture decisions. *Journal of Business Venturing*, *15*(4), 305–322.

Franke, N., Gruber, M., Harhoff, D., & Henkel, J. (2008). Venture capitalists' evaluations of start–up teams: Trade–offs, knock–out criteria, and the impact of vc experience. *Entrepreneurship Theory and Practice*, *32*(3), 459–483.

Gage, D. (2012). *The venture capital secret: 3 out of 4 start-ups fail*. Retrieved April 27, 2022, from https://online.wsj.com/article/SB10000872396390443720204578004980476429190.html

Gompers, P., Kovner, A., & Lerner, J. (2009). Specialization and success: Evidence from venture capital. *Journal of Economics & Management Strategy*, *18*(3), 817–844.

Hadley, B., Gloor, P. A., Woerner, S. L., & Zhou, Y. (2018). Analyzing vc influence on startup success: A people-centric network theory approach. In *Collaborative innovation networks* (pp. 3–14). Springer.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., & Klein, M. (2002). *Logistic regression*. Springer.

KPMG. (2021). *Kpmg venture pulse q4 2021* (tech. rep.). KPMG.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2019). Continuous control with deep reinforcement learning.

Malmström, M., Johansson, J., & Wincent, J. (2017). Gender stereotypes and venture support decisions: How governmental venture capitalists socially construct entrepreneurs' potential. *Entrepreneurship Theory and Practice*, *41*(5), 833–860.

Maxwell, A. L., Jeffrey, S. A., & Lévesque, M. (2011). Business angel early stage decision making. *Journal of Business Venturing*, *26*(2), 212–225.

McMullen, J. S., & Dimov, D. (2013). Time and the entrepreneurial journey: The problems and promise of studying entrepreneurship as a process. *Journal of Management Studies*, *50*(8), 1481–1512.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *nature*, *518*(7540), 529–533.

Moody, J., & Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, *12*(4), 875–889.

Nanda, R., Samila, S., & Sorenson, O. (2020). The persistent effect of initial success: Evidence from venture capital. *Journal of Financial Economics*, *137*(1), 231–248.

Petty, J. S., & Gruber, M. (2011). "in pursuit of the real deal": A longitudinal study of vc decision making. *Journal of Business Venturing*, *26*(2), 172–188.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms.

Sharchilev, B., Roizner, M., Rumyantsev, A., Ozornin, D., Serdyukov, P., & de Rijke, M. (2018). Web-based startup success prediction. *Proceedings of the ACM International Conference on Information and Knowledge Management*, 2283–2291.

Shepherd, D. A., & Zacharakis, A. (2002). Venture capitalists' expertise: A call for research into decision aids and cognitive feedback. *Journal of Business Venturing*, *17*(1), 1–20.

Shepherd, D. A., Zacharakis, A., & Baron, R. A. (2003). Vcs' decision processes: Evidence suggesting more experience may not always be better. *Journal of Business Venturing*, *18*(3), 381–401.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second). The MIT Press.

Wang, Y. (2016). Bringing the stages back in: Social network ties and start-up firms' access to venture capital in china. *Strategic Entrepreneurship Journal*, *10*(3), 300–317.

Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., & Freitas, N. (2016). Dueling network architectures for deep reinforcement learning. *International Conference on Machine Learning*, 1995–2003.

Wesley II, C. L., Kong, D. T., Lubojacky, C. J., Saxton, M. K., & Saxton, T. (2022). Will the startup succeed in your eyes? venture evaluation of resource providers during entrepreneurs' informational signaling. *Journal of Business Venturing*, *37*(5), 106229.

Xu, R. R., Chen, H., & Zhao, J. L. (2022). Sociolink: Leveraging relational information in knowledge graphs for startup recommendations. *Journal of Management Information Systems*, *Forthcoming*.

Yankov, B. (2012). Overview of success prediction models for new ventures. *International Conference Automatics and Informatics*, *12*, 13–16.

Zacharakis, A. L., & Meyer, G. D. (2000). The potential of actuarial decision models: Can they improve the venture capital investment decision? *Journal of Business Venturing*, *15*(4), 323–346.

Zhang, S., Zhong, H., Yuan, Z., & Xiong, H. (2021). Scalable heterogeneous graph neural networks for predicting high-potential early-stage startups. *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2202–2211.