Leveraging Search Technology to Augment Investment Practices by Venture Capitalists
Tiffany Jaya, Pavan Kurapati, Nishant Velagapudi
W201 Week 8 - Group Report 1

Domain
Venture Capital

Venture Capitalists raise money from a pool of capital known as limited partners, which include accredited individuals and large financial institutions. Their goal is to get a return of at least 3-5 times committed capital by purchasing shares of start-up companies and selling them for significant gains when these companies go public. Typically, VC operates under the 2 and 20 structure whereby they will charge limited partners 2% annually for operating expenses and 20% profit for selling the company's shares. So it is in their best interest to assist entrepreneurs for success to gain as much of the 20% profit as possible.

The majority of investments made by a venture capitalist fail. This is unavoidable: as the success rate of start-up companies is prohibitively small. However, the advent of data science can give VC firms more information to decide who they choose to invest in.

Key Stakeholders
- *Entrepreneurs :* They have the idea and need funding
- *Investors :* Need high returns
- *Investment Bankers :* Companies to sell
- *Venture Capitalists :* Make market for the above three

To tie our stakeholders into the introduction to the domain: entrepreneurs own start-ups and work to publicize/grow them. Investors are interested in growing their capital, while investment bankers want to see their investments go public and sell. Venture capitalists take money from investors and investment bankers and identify promising start-ups. They invest in their choices and then work to ensure that these start-ups succeed.

Core Business Problems
- ***How do we pick right Industry and Technology?***

The Harvard Business Review argues that "the myth is that venture capitalists invest in good people and good ideas. The reality is that they invest in good industries."[1] Deciding what is a good industry is challenging: the timing can be volatile and missing the mark is extremely expensive. HBR gives the example of CD-roms: "in 1983, more than 40 venture-funded companies and more than 80 others existed. By late 1984, the industry market value had plunged from $5.4 billion to $1.4 billion. Today only five major players remain."[1]

- ***How does region impact VC funding?***

VCs suffer from a locality bias - they are more likely to invest in companies that are geographically near to them. However, while Asia's total venture capital investment has remained steady, the VC activity in the Americas and Europe fell significantly ($82 -> $72 billion

and $18 to $16 billion respectively). The locality bias is highly logical - local businesses are easier to familiarize, involve less risk of the unknown (either in terms of legislature or market pitfalls).

- ● **At what stage should an investment be made?**

Part of deciding what to invest in is deciding what "stage" of a company is safe to invest in. Investing later in the process means that the companies are more stable and shares will be more expensive. At the same time, investing earlier in the life cycle of a company is riskier. 80% of the money invested by VCs goes into companies in their "adolescent phase." However, it is possible that risk averse behavior leads to greater investment in slightly older start-ups. Companies in their true startup phase have less of a track record and thus attract less attention. On the other hand, there is significant investor caution for late stage deals - potentially because companies requiring late stage investment have some hidden problems. Median deal sizes have increased for startup stage companies from 2015 to 2016, while median deal sizes for adolescent companies has held steady.

- ● **How does crowdfunding impact Venture Capitalists business model?**

Some businesses have gone into the practice of crowdfunding instead of looking for true Venture Capital backing. Crowdfunding poses a threat to eat into VC opportunity, as entrepreneurs will be able to promise products instead of company equity to receive infusions of cash. This threat will hopefully push venture capitalists to be willing to adapt new technologies to gain an advantage. VCs may be unused to the idea of concept competition - as they have solely owned the domain for quite some time. An example is home robots or small technology (wearables, etc.) - Jibo raised more than $1 million in crowd funding, eliminating the need and downside for entrepreneurs of venture capitalism.4

Decision To Be Made

Venture capitalists need to determine what companies are worth investing in and at what point should they be invested in. Korver highlights how difficult this seemingly simple decision is, and states that "it takes seven years and $30 million to train a venture capitalist."2 These decisions are further complicated by the currently low availability of data: "the average venture capitalist only makes a few decisions a year, and it takes years to see if these decisions result in good or bad outcomes."

Fallacies & Biases

Part of the reason why this occurs is because limited partners fall prey to the social norm bias. They accept the risk of investing with VCs and raise huge funds even though VCs have failed to return their capital. They don't held VCs accountable which is part of the reason why the problem persists.

But out of all the participants in the domain, venture capitalists who rely more on an intuition based approach are captive to the common fallacies we encountered on the daily basis. For the longest time, VCs tend to find their own field immune to the change. This special pleading

fallacy is partly a result of a historical but successful paradigm established by the fact that VCs used to be a relatively small group of individuals who have access to the best deals and therefore have better success than they do see now. But with increase funds and the rise of new approaches (such as Y Combinator, Kickstarter, or super angels), there are the same number of great startup ideas but with more competition vying for the best deals.

In addition, the human mind is not great at piecing together facts. We commonly weave a thread of story through the facts so that we can explain to ourselves what isn't there and fool ourselves into thinking that we can predict the future. VCs are encountering the same narrative fallacy. They want to overreach correlation between two unlike things. It's the same as a person who wants to fly and creates a large feathered wing because he observed things that flew have feathers. Similarly, VCs often times over-extrapolate and think if it worked for Uber it should work for Uber for X.

Since the goal of VCs is to get a high return on their invested capital, they have a tendency to view the return on startup as bimodal outcome: either it is a total bust or a unicorn like Google. But startup returns are not bimodal; they exhibit more of a power law curve. Sure, there will be 15-20 unicorns which will break into the $100 million revenue, but most companies will be somewhere along the middle, not a total bust but not a unicorn either, and it is plausible to build these companies into the $100 million in revenue, but it will take years to do so (approximately 6-12 years). The adherence to the bimodal returns fallacy causes two bad behaviors. One, VCs feel the need to do every deal so that you can catch one big winner. Second, companies that are anointed as unicorns are valued higher than public market is willing to pay for these companies.

To make matter worse, venture capitalism have long feedback cycles. A typical partner in VC firm holds only one or two deals a year. In a span of 10 years, VC might only invest in 15 different companies, so they will not know if their decision in investing a certain company result in a good or bad outcome. It becomes easy to confuse luck with skill. For this reason, it is easy for VCs to use any information to confirm their way of thinking; they can be subject to a confirmation bias or a group mentality by validating with other VCs their investment in a coffee session. The result is a biased judgment or a group of VCs going after the same market at the same time. It's bad news all around.

Proposal
Data Collection
We see two problems to address: the first is industry identification and the second is determining the best time to invest to maximize potential profits and minimize risk. We propose a composite approach to aggregating a dataset that will be uniquely comprehensive for the purpose of Venture Capitalism. We will aggregate stock market data, web hits, and IPO dates for various companies. We can collect stock market data through a variety of APIs available online. We can aggregate web hits and text blurbs about various start-ups by using a web-crawler (though this will need to be an ongoing process and must be restricted to an explicit list of start-up names). We can also retrieve a list of IPO dates and offering prices from an online source (https://www.iposcoop.com/, for example).

Analysis
<u>Web Crawling</u>

        We can use the number of times that a start-up company is mentioned as a proxy metric for public "awareness" about the company/industry. We can calculate sentiment scores in each blurb of text that mentions the company name as a rough measure of how positively/negatively the company is viewed. We can then plot number of hits/sentiment over time versus the IPO date and offering price itself. We can retain the source of the text blurb and use it to calculate a "weight" (e.g. a mention on Wired means a lot more than a mention on a smaller blog). Stock prices will give us a proxy metric for how industries as a whole are performing as well.

        Our goal is to answer two large questions: where should VCs invest (industry or company specific), and when should they invest (relative to start-up lifecycle). Web crawling is a novel data source to address these problems and we believe we can use these data to identify "hot" or trending industries and view historical data to identify patterns in investment time (within company life-cycle) and ultimate result.

Implementation

        We can implement this work using cloud resources: either AWS or Azure is ideal for web crawling since the dedicated resources are elastic. We can then perform our text analysis using pre-trained sentiment models and existing NLP libraries. There are significant challenges in performing these analyses: we have high velocity and high volume data. We have computational complexity in just counting name hits - let alone performing sentiment analysis on each snippet. We also need to enable transparency into our data - we do not want to fall into the trap of presenting insights in a silo, thus excluding subject matter experts who could either challenge or augment our findings.

Impact

        We expect our analysis to give VC's a unique advantage by getting a read on how popular a company is. Further, VC's taking an active role in publicizing a company can track how their efforts are impacting a company's visibility. Charting historical sentiment/hits data versus IPO date and offering price will also give our VC's insight into when investment should be made. We think that this data will be a powerful complement to existing risk profiles/opportunity evaluations.

Potential Obstacles

        We think that our approach could be doubted, as traditional venture capitalists who rely more on an intuition based approach may need to see hard evidence that our data can correctly identify companies that will perform well. Additionally, VC firms will be very hesitant to share data/methodologies with one another, so our data needs must remain agnostic from any particular venture capitalist for our approach to be successful. VC's might face confirmation bias - they will be most likely to adapt our methods so long as they confirm the conclusions they come to and reject our methods if they disagree with the conclusions. Our approach takes on two of the three V's of big data - volume and velocity. Web crawling is a resource intensive

practice that involves significant technical challenge. Parsing crawled text, assigning sentiment, and weighting the resulting counts and sentiment through source scoring will add another degree of challenge.

References

1. https://hbr.org/1998/11/how-venture-capital-works
2. https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2017/01/venture-pulse-q4-2016-report.pdf
3. https://www.kauffmanfellows.org/journal_posts/applying-decision-analysis-to-venture-investing/
4. https://techcrunch.com/2012/07/29/the-paradox-of-vc-seed-investing/
5. https://www.fastcompany.com/1826876/googles-creative-destruction
6. https://mattermark.com/venture-capitalists-cognitive-bias-and-why-past-isnt-always-prelude/
7. http://www.acuitasinvestments.com/insights/Kauffman_VC_paper.pdf
8. https://medium.com/@ArtofJina/the-vc-industrys-identity-crisis-e83639d2ff60
9. https://www.kauffmanfellows.org/journal_posts/applying-decision-analysis-to-venture-investing/
10. https://www.mckinsey.com/industries/high-tech/our-insights/a-machine-learning-approach-to-venture-capital
11.