Factors that influence startup investment

Revised Research Proposal

W.M.R. Shelton

Component: Research Proposal

Supervisors: Professor Melinda Hodkiewicz, Dr Tim French

Degree: BPhil(Hons) (24 point project)

University: The University of Western Australia

Background

Technological advances have made launching startups more accessible than ever before. Customers can be accessed easily through the Internet and launching a startup can be done from a bedroom. However, startups remain competitive and risky endeavours. Startups can be unprofitable for years so entrepreneurs look for incubators, accelerators, angel investors and venture capital firms to support them through this developmental period. Aside from funding, investors hold experience and networks that can accelerate startup growth. Investors act as scouts, able to identify the potential of new startups, and as coaches, able to help startups realise that potential [1].

Startups must convince investors to support them throughout their development, but this process can be burdensome and time-consuming. Investors find it difficult to evaluate startup potential for investment because metrics of performance often do not exist or are uncertain [2]. Popularity of online databases like AngelList and CrunchBase, which offer information on startups, investments and investors, is evidence of a desire for better methods of assessing startup potential. By 2014, over 1,200 investment organisations (including 624 venture capital firms) were members of CrunchBase's Venture Program, mining CrunchBase's startup data to help inform their investment decisions [3].

Investment comes with trade-offs for startups. The majority of venture capital-backed startups end in bankruptcy [4]. Investors are protected from these losses because the minority of their investments that are successful have outsized returns: 85% of venture capital returns come from 10% of investments [4]. Investors seek to optimise the risk-reward trade-off by pressing startups to grow rapidly, frequently raise funding rounds and make quick, centralised decisions [5]. The rapid growth demanded by investors is generally incompatible with public company structures, due to reporting and compliance requirements [6]. Accordingly, we see venture capital-backed startups delaying Initial Public Offerings (IPO). Time taken to IPO has doubled in the past 20 years [7].

Aim

Startups remaining privately-held for longer shifts value creation to the private markets. Microsoft's market capitalisation grew 500-fold following its IPO in 1986, but for Facebook to grow to the same extent since its IPO in 2012 its capitalisation would exceed the total global equity market. Investment in late-stage startups is approaching all-time highs as public market investors enter the private markets [7]. Given this situation, it is important to understand how factors that influence investment change through a startup's development. A clear gap in the academic literature exists in this area. Studies of the determinants of startup investment have common weaknesses. This study will address these weaknesses in three ways:

Larger Sample Size Previous studies are restricted in sample size. Most studies have samples of fewer than 500 startups [8, 9], or between 500 and 2,000 startups [10, 11, 12, 13, 14], and only a few have large scale samples (more than 100,000 startups) [15, 16]. Sample size is more critical to model development than the sophistication of machine learning algorithms or feature selection [17]. Startups databases (e.g. CrunchBase) and social networks (e.g. Twitter) offer larger data sets than those previously studied. We expect data collected from these sources will lead to the discovery of additional features and higher accuracy in startup investment prediction.

Developmental Focus Prior work focuses on early-stage investment in startups, primarily equity crowdfunding [18, 8, 16, 19] and angel investing [14]. The functions and objectives of startups change through their development [20]. For example, early stages of funding are characterised by uncertainty about technical validity and market fit [21]. In this setting, patents are a strong signal to investors, but may become less so in later rounds. Generally, we expect signals that attract investment in startups will change over time.

Rich Features Prior work focuses on basic company features (e.g. the headquarters' location, the age of the company) for startup investment predictive models [18, 9]. Semantic text features (e.g. patents, media) [10, 19] and social network features (e.g. co-investment networks) [13, 16, 11] may also predict startup investment. We expect a model that includes semantic text and social network features alongside basic company features could lead to better startup investment prediction.

We will develop software that collects and processes information about startups to predict their likelihood of raising investment at different stages of their

development. This study has potential for scholarly, policy and firm-specific implications. Our scholarly contribution is a conceptual framework for startup investment, based on work by Ahlers et al. [8]. Our conceptual framework posits that startup investment is a product of two factors: startup potential and investment confidence. We will test this framework with respect to startup development using cross-sectional and longitudinal analyses. We aim to contribute to the understanding of the determinants of startup investment, with a focus on how they change over time. Ultimately, we hope that we can encourage greater investment in startups.

Method

Data Collection   We will develop an automated data collection system that will provide a platform on top of which we can easily perform our analyses. Our primary data sources are CrunchBase, AngelList, Twitter and PatentsView. We will start with a focus on CrunchBase and then develop systems to match the other sources. CrunchBase data can be accessed in multiple ways. The simplest format are comma separated files (CSV) that hold data about each relation in their database (e.g. funding rounds, investors). A current CSV dump of the database can be requested at any time from CrunchBase. We have also retrieved several older CSV dumps that can be compared with current data for longitudinal studies. CSV dumps provide a subset of the attributes in the CrunchBase data set. To get the full data set requires access through their application programming interface (API). We will develop a crawler that can continually traverse the API iteratively to effectively mirror the CrunchBase data set locally for further analyses. Our master database is likely to be a Sqlite server. We are also investigating distributed solutions, including using Spark.

Machine Learning Analyses   We will manipulate and combine the data collected from our data sources into a labelled data set appropriate for the application of supervised machine learning algorithms. Primary labels will be whether a startup receives funding at each funding round. We may also investigate measures of startup performance (e.g. survival time, exit). We will compare and evaluate machine learning algorithms to find which algorithms suits this task best. We have collected six historical CSV dumps from CrunchBase spanning the period from October 2013 to the present. We will match companies across these data sets to test the robustness of our model across time and to see whether the gradient of change in different features can provide greater accuracy to our model than the static features.

Timeline

Please see below (Table 1) for a schematic of the proposed methodology.

S:W	Date	Task
2:03	Fri 19 August	Draft proposal due
2:14	Wed 09 November	Revised proposal due
2:14	Wed 09 November	Literature review due
2:16	Fri 25 November	Data collected
2:21	Fri 30 December	Completed main experiments
1:08	Fri 28 April	Draft dissertation due
1:10	Fri 12 May	Seminar title and abstract due
1:13	Mon 29 May	Final dissertation due
1:13	Fri 02 June	Poster due
1:13	29 May - 02 June	Seminar
1:17	Mon 26 June	Corrected dissertation due

Table 1: Proposed timeline

Software and Hardware Requirements

This project will be developed primarily in Python using scikit-learn, a free open-source machine learning library [22]. Sqlite may be used to prepare datasets for processing. The system will be hosted on a public compute cloud, likely Amazon Web Services. Free academic licenses for CrunchBase and AngelList have been requested.

Bibliography

[1] BAUM, J. A., AND SILVERMAN, B. S. 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 (2004), pp. 411–436.

[2] SHANE, S., AND CABLE, D. Network ties, reputation, and the financing of new ventures. *Management Science* 48, 3 (2002), pp. 364–381.

[3] PATIL, A. *CrunchBase's Venture Program Members Are Making Startup Data Better Than Ever.* `https : / / info . crunchbase . com / 2015 / 01 / crunchbases-venture-program-members-are-making-startup-data-better-than-ever/`. Online; accessed 18 05 2015. Jan. 2015.

[4] SAHLMAN, W. *Risk and reward in venture capital.* 2010.

[5] FRIED, J. M., AND GANOR, M. Agency costs of venture capitalist control in startups. *New York University Law Review* 81 (2006), p. 967.

[6] WIES, S., AND MOORMAN, C. Going public: how stock market listing changes firm innovation behavior. *Journal of Marketing Research* 52, 5 (2015), pp. 694–709.

[7] NATIONAL VENTURE CAPITAL ASSOCIATION *2016 National Venture Capital Association Yearbook.* `http : / / www . nvca . org / ?ddownload = 2963`. Online; accessed 06 Nov 2016. Mar. 2016.

[8] AHLERS, G. K., ET AL. Signaling in equity crowdfunding. *Entrepreneurship Theory and Practice* 39, 4 (2015), pp. 955–980.

[9] GIMMON, E., AND LEVIE, J. Founder's human capital, external investment, and the survival of new high-technology ventures. *Research Policy* 39, 9 (2010), pp. 1214–1226.

[10] HOENEN, S., ET AL. The diminishing signaling value of patents between early rounds of venture capital financing. *Research Policy* 43, 6 (2014), pp. 956–989.

[11] YU, Y., AND PEROTTI, V. Startup Tribes: Social Network Ties that Support Success in New Firms. In: *Proceedings of 21st Americas Conference on Information Systems.* 2015.

[12] AN, J., JUNG, W., AND KIM, H.-W. A Green Flag over Mobile Industry Start-Ups: Human Capital and Past Investors as Investment Signals. In: *PACIS 2015 Proceedings*. AIS Electronic Library, 2015, p. 67.

[13] WERTH, J. C., AND BOEERT, P. Co-investment networks of business angels and the performance of their start-up investments. *International Journal of Entrepreneurial Venturing* 5, 3 (2013), pp. 240–256.

[14] CROCE, A., GUERINI, M., AND UGHETTO, E. Angel Financing and the Performance of High-Tech Start-Ups. *Journal of Small Business Management* (2016). ISSN: 1540-627X. DOI: 10.1111/jsbm.12250.

[15] SHAN, Z., CAO, H., AND LIN, Q. Capital Crunch: Predicting Investments in Tech Companies. Unpublished thesis. Stanford University. Available at http://www.zifeishan.org/files/capital-crunch.pdf. 2014.

[16] CHENG, M., ET AL. Collection, exploration and analysis of crowdfunding social networks. In: *Proceedings of the Third International Workshop on Exploratory Search in Databases and the Web*. ACM. 2016, pp. 25–30.

[17] CARUANA, R., KARAMPATZIAKIS, N., AND YESSENALINA, A. An empirical evaluation of supervised learning in high dimensions. In: *Proceedings of the 25th International Conference on Machine learning*. ACM. 2008, pp. 96–103.

[18] BECKWITH, J. Predicting Success in Equity Crowdfunding. Unpublished thesis. Joseph Wharton Research Scholars. Available at http://repository.upenn.edu/joseph_wharton_scholars/25. 2016.

[19] YUAN, H., LAU, R. Y., AND XU, W. The determinants of crowdfunding success: A semantic text analytics approach. *Decision Support Systems* 91 (2016), pp. 67–76.

[20] MCMULLEN, J. S., AND DIMOV, D. Time and the entrepreneurial journey: The problems and promise of studying entrepreneurship as a process. *Journal of Management Studies* 50, 8 (2013), pp. 1481–1512.

[21] HSU, D. H., AND ZIEDONIS, R. H. Patents As Quality Signals For Entrepreneurial Ventures. In: *Academy of Management Proceedings*. Vol. 2008. 1. Academy of Management. 2008, pp. 1–6.

[22] PEDREGOSA, F. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* (2011).