

生存時間解析モデルを用いたベンチャー企業の成長予測

Predicting Growth of Startups using Survival Analysis

株田 達矢 *¹ 森 純一郎 *¹

Tatsuya Kabuta

Junichiro Mori

*¹ 東京大学大学院工学系研究科技術経営戦略学専攻

Graduate School of Engineering, The University of Tokyo

Identifying the factors which affects growth of ventures is important for both investors and entrepreneurs. In this paper, we apply the Survival Analysis, a statistical method for the duration of probabilistic events, to model the movement of ventures' financing stages. In addition, we use Big Data available on web such as investment history of investors and text data of news articles to show some factors such as the network features and appearance on news articles affect on the venture growth.

1. 序論

新しい技術や新しいビジネスモデルを中核に、急速な成長を目指す新興企業がベンチャー企業である。経済の成熟化あるいは停滞が進む中、これらベンチャー企業は重要な存在であり、ベンチャー企業の成長要因の分析への関心が高まってきた。

まず古典的な手法として、その企業単体の属性、すなわち事業領域、ビジネスモデル、資金、人材、特許等を説明変数とし企業の成長をモデル化するアプローチが生まれた [Gompers 08]。また、企業を取り巻く資本や人材のネットワークに注目した研究がなされた [Shane 02]。しかしながら、これらの研究の多くは、成長の結果 (Exit するか否か等) のみに着目し、成長のスピードを捉えたものは少ない。現実には、あるベンチャーが成功するか否かだけでなく、どの程度の速さで成長するのか、リターンを得られるのはいつなのかといった視点が、投資家や起業家にとって重要である。また、分析に用いるデータについて、昨今のインターネットの普及および World Wide Web の拡大により、我々が利用可能な情報の量も飛躍的に増大した。例えば、多くの VC (ベンチャーキャピタル) はその投資先を公開している。他にも、今日ではベンチャー企業が多くのニュースメディアに取り上げられ、また我々はそれを Web 上で容易に読むことが可能である。これらはいずれも、投資家や記者などの人の判断が投資や取材という形で顕在化したものであるため、企業の潜在能力や成長性を反映している可能性が高い。

そこで本研究では、Web で入手可能なこれらの企業情報、投資家の投資履歴とニュース記事のテキストデータを用いて、ベンチャー企業の成長スピードを分析した。具体的には、CrunchBase (<https://www.crunchbase.com>) の企業データおよび TechCrunch (<http://techcrunch.com>) のニュース記事情報を用い、企業の投資ステージの推移を生存時間解析の手法で分析することで、投資活動とニュース記事情報がその後のベンチャーの成長と関連していることを示した。このように生存時間解析を用いたベンチャー企業の成長分析としては、企業の地域や業種情報を用いたもの [Giot 07] などがなされているが、本研究のように、ニュース記事に代表されるソーシャルな情報や有力 VC との関係に着目したものはなされていない。

本研究で検証する仮説は以下の二つである。

仮説 1

ニュースメディアで取り上げられる企業はそうでない企業に比べ、その後成長する可能性が高く、また成長スピードも早い。

仮説 2

有力な投資家の投資を受けた企業はそうでない企業に比べ、その後成長する可能性が高く、また成長スピードも早い。

結論としては、これらの仮説は正しいことが実験の結果わかった。すなわち、有力な投資家から投資を受けた企業、およびニュースメディアで取り上げられた企業はそうでない企業に比べ、次の投資ステージに進む可能性が高く、またその期間は平均的には短くなることが明らかになった。

2. 手法

2.1 ベンチャー企業の成長

ベンチャー企業はその成長の過程で、VC や個人投資家から幾たびかの資金調達を行う。このようなベンチャー企業への投資はその段階に応じて、次のようなカテゴリーに分けることができる。

シード投資

事業を始めるための立ち上げ資金。

この時点ではビジネスはまだアイデア段階であることが多い。調達した資金は主にユーザ調査や製品のプロトタイプ製作に利用される。

シリーズ A

外部からの初の本格的な資金調達。通常 VC や CVC はこの段階で参画する。

さらなる技術開発や、優秀な人材の獲得に用いられる。

シリーズ B

シリーズ A 以上の大型調達となる。通常この段階ではプロダクトが出来上がっており、収益予測も立っていることが多い。調達資金は主にユーザの獲得や事業の収益化に用いられる。

シリーズ C+

シリーズ B 以降も資金が必要な場合、シリーズ C, D,

連絡先: 株田達矢, 東京大学大学院工学系研究科, 東京都文京区本郷 7-3-1, 03-5841-1161, t.kabuta@gmail.com

E...というように追加投資が行われる．資金はさらなる事業拡大のために用いられる．

また，ある程度成長した起業は Exit (IPO もしくは売却売却)へと至る．本研究では，投資シリーズが次に進むことあるいは Exit に至ることをもってベンチャー企業の「成長」と定義する．

2.2 データセット

本研究では，ウェブで利用できる大規模な情報ソースである，CrunchBase が提供する企業データおよび TechCrunch のニュース記事に着目した．

CrunchBase は TechCrunch が運営するスタートアップ関連のデータベースであり，2015 年 8 月末時点で，IT，バイオ，エネルギー等テック領域における 26 万 2660 社の企業と 1 万 3628 の金融機関，37 万 9285 人の個人のデータが登録されている．

TechCrunch は 2005 年にサービスを開始した，スタートアップ領域に特化した Web メディアである．主に IT，バイオ，エネルギー分野のスタートアップ企業の紹介や業界の重要なニュースを発信しており，累計記事数は 2015 年 11 月時点で 13 万 3727，月間ページビュー数は 2600 万に達する．

本研究では，次のステージの投資あるいは Exit がなされるまでの期間を分析し，ベンチャー企業の成長速度に影響する要因を探る．いま，ベンチャー企業の資金調達を一定の確率で発生する事象と捉えたと，その間隔はある確率分布に従う．生存時間解析はそのような問題を扱うためのモデルである．

2.3 生存時間解析

一定の確率で生起する事象 (イベント) の発生間隔を扱う問題は，医学や生物学，工学といったさまざまな分野で現れる．生存時間解析はそのような問題に対し威力を発揮する．生存時間解析はその名の通り，まずは医学の領域で，患者の治療開始時点から死亡までの期間を解析するために発展した．すなわち，二つの患者群に試薬と偽薬を投与し，それぞれの集団間で生存時間に有意な差があるか，またそれはどの程度かといった問題を通じて，薬の効能を明らかにすることが初めの応用例である．また，データの打ち切りにも対応出来るのが特徴である．「死亡」「生存」といった言葉にとられることなく，広くイベント一般が発生するまでの時間間隔を分析する際にも，生存時間解析は有効である．

いま，ベンチャー企業の資金調達や Exit をある確率で発生するイベントと捉えたと，その発生間隔は生存時間解析を用いて分析することができる [Hochberg 07]．また，観測期間内に次の投資ステージに達しなかった企業については打ち切りデータとみなすことができる．そのような点でも，ベンチャーの資金調達間隔を生存時間解析でモデル化することは理にかなっている．

時刻 t でイベントが発生する確率の確率分布が $f(t)$ で与えられると仮定する．このとき，次で定義される $S(t)$ を生存関数と呼ぶ．

$$S(t) = Pr(T > t) = \int_0^t f(x)dx \quad (1)$$

生存関数は，イベントが時刻 t までに発生しない，すなわち生存時間が t 以上である確率を表す．本研究ではこれを，ベンチャー企業が時刻 t までに次の投資ステージに進まず，現在のラウンドに留まっている確率とみなす．

また，次のように定義される $h(t)$ をハザード関数と呼ぶ．

$$h(t) = \lim_{\Delta t \rightarrow +0} \frac{Pr(t < T < t + \Delta t)}{\delta t} = \frac{f(t)}{S(t)} \quad (2)$$

ハザード関数は，イベントが t まで発生しなかった場合に，次の Δt に発生する確率，すなわち時刻 t の瞬間発生率を表す．生存時間と共変量の関係を調べる場合，ハザード関数として次の形を仮定することが多い．これを Cox 比例ハザードモデル [Cox 72] という．

$$h(t) = h_0(t) \exp\left(\sum_{i=1}^k \beta_i z_i\right) \quad (3)$$

ここで z_i を共変量ベクトル， β_i をパラメータである．右辺前半部は時間 $h_0(t)$ のみに依存する部分で，ハザード関数のベースラインを表す．一方後半部は各データにより異なり，共変量がハザード関数に及ぼす影響を表す．本研究では，ニュース記事における言及や VC との関係によって企業の成長スピード (資金調達間隔) がどのように変化するかに関心がある．従ってセミパラメトリックな Cox モデルを用いればよく， $h_0(t)$ について具体的な形を考える必要はない．

2.4 Cox モデルを用いたベンチャー企業の成長モデル

2.3 節で説明した生存時間解析の枠組みを用いて，ベンチャー企業の成長を予測するモデルを構築する．例として，シードステージにある企業がシリーズ A に進むまでの期間をモデル化する問題を考える．

いま， T_s を観測開始時点， T_e を観測終了時点とする． T_s 時点でシードステージにあった企業を集め，その集合を C とする．次に， T_s から現在までの投資履歴を調べ，もし企業 c_i が次の投資を受けていたらその時点を T_1 とし， $DURATION(i) = T_1 - T_s$ を企業 c_i のシードシリーズの生存時間とする．なお，複数回の投資が同一ラウンドにカテゴライズされている場合，最も古いものを投資日とする．追加投資が見つからなかった場合，そのデータは打ち切りとし， $DURATION(i) = T_e - T_s$ を企業 c_i のシードステージの生存時間とする．また，打ち切られた企業は $CENS(i) = 0$ ，そうでない企業は $CENS(i) = 1$ とする．ここで，この場合の $DURATION(i)$ がシード投資からシリーズ A までの期間そのものではないことに注意したい．観測開始時点において最初の投資日からの日数が企業によって異なっている．従来研究の多くでは，観測開始時点企業ごとに定め，純粋に投資の発生間隔のみを分析対象としている．しかし本研究では，景気の変動等による影響を除去するため，また利用した情報の取得可能日を明示的に示すため，あらかじめ定められた同一期間内の企業群を対象に分析を行っている．代わりに，観測開始時点での投資日からの経過日数は共変量としてモデルに組み込む．

このようにして得られた $DURATION$ と様々な共変量との関係が，本研究の関心事項である．

まず，仮説 1 を検証するために，次のような共変量を考える．

ARTICLE_FLAG

開始観察時点以前に記事で言及されたか否か

次に，仮説 2 を検証するために，次の共変量を考える．

WITH_TOPVCs_PR

「有力 VC」から投資を受けているか否か

「有力 VC」の選出法としては様々なものが考えられるが、ここでは「有力 VC」として「共投資ネットワーク上における PageRank 上位 20 社」という定義を採用した。

共投資ネットワークとは、投資家間の共投資関係すなわち「共通の企業を投資先として持つ」という関係をグラフ化したものである。いまあるベンチャー起業 X に投資家 A,B,C が投資しているとすると、このとき、後から参画した企業から先駆者へと有向枝を張ることで構成されるのが、共投資ネットワークである。

PageRank とは、ネットワーク上のノードに対して定義される値で、より多くの枝を集めるノードほど、また有力なノードから枝を集めるほど、高いスコアがつくように設計されている。その性質から、ウェブページの重要度の指標として用いられている。

共投資ネットワークでは、後から参画したプレイヤーが先駆者へと枝を張るため、より早く投資を実行するほど、また投資した企業がのちに多くの投資家を惹きつけるほど、多くの枝を集める。すなわち、「将来有望なベンチャー企業を早い段階で発見し、育てた VC」が多くの枝を集める。そのため、共投資ネットワーク上の PageRank を見ることによって、投資家の「目利き力」を評価できると考えられる。このような考えから、今回は VC の影響度を PageRank を用いて順位付けする手法を採用した。

以上、二つの仮説に対応する二つの共変量の有意性を検定することが、本実験の主な目的である。しかし、ベンチャー企業の成長は他の様々な要因によっても規定されている。そこで本研究では関連研究を踏まえ、次の共変量もモデルに組み込んだ。

AGE 企業設立から観測開始時点までの経過日数

DAYS_SINCE_INVESTED 対象ステージの投資を受けた日から観測開始時点までの経過日数

MILESTONES 観測開始時点から直前にマイルストーンがあるか否か(2 値)

IN_BA シリコンバレーを含む米国ベイエリアにあるか否か(2 値)

IN_US 米国(ベイエリア除く)にあるか否か(2 値)

BIO バイオテクノロジー系企業

ENVIRONMENT 環境系企業

HARDWARE ハードウェア系企業

ENTERPRISE エンタープライズ系 IT 企業

COMMERCE E コマース系企業

AD 広告系企業

GAME ゲーム企業

FINANCE 金融系企業

EDUCATION 教育系企業

IN_BA 米国ベイエリア

IN_US ベイエリア以外ん米国

IN_OTHERS 米国以外の地域

LOG_TOTAL_FUNDING 観測開始時点までの累計資金調達額の対数。

これらはいずれも企業の基本的な属性である。また、ネットワーク的属性として、投資ネットワークに着目し、それに基づき計算されるスコアを共変量として用いた。ここで、投資-被投資ネットワークとは、投資家と企業の投資する/される関係を表すネットワークのことである。今回は次の 3 つの基本的なスコアを用いた。

DEGREE_CENTRALITY 次数中心性。

CLOSENESS_CENTRALITY 近接中心性

BETWEENNESS_CENTRALITY 媒介中心性

投資-被投資ネットワークおよび前述の共投資ネットワークは観測開始時点以前のすべての CrunchBase 上の投資履歴をもとに作成した。また、各指標については観測開始時点のネットワーク構造に基づいて計算し、その後のネットワーク自体の変化は考慮に入れていない。

3. 実験

2.4 節で導入したモデルを用いて、ベンチャー企業の成長スピード(シード~シリーズ A~シリーズ B~シリーズ C および各ステージ~Exit)の分析を行った。

3.1 実験方法

企業データについては前述の CrunchBase から、2015 年 11 月に取得したデータを用いた。次に TechCrunch の記事データについては、同年 12 月に取得したデータを使用した。

分析は、2013 年 1 月を開始時点とし、そこから遡って 2 年以内、すなわち 2011 年 1 月から 2012 年 12 月までに対象投資ステージに到達した企業が、その後の 2013 年 1 月 2015 年 8 月に次のステージに進むまでを対象とした。また、各ステージから Exit までの期間の分析に当たっては、次の変数をモデルに加えた。

LAST_ROUND

最後の投資ステージの種類。シードシリーズ A シリーズ B それ以降の 4 値。

3.2 実験の結果

図 1 にシードステージ~シリーズ間の、図 2 に各ステージから売却までの分析結果を記す。ここで表の coef は共変量の影響の大きさを、 $\exp(\text{coef})$ はそこから計算される相対ハザードを、 $\text{se}(\text{coef})$ は係数の標準誤差を、 z および $\Pr(>|z|)$ はそれぞれ z 値、 p 値である。右の星の数は p 値の大きさを示しており、0~0.1% が星 3 つ、0.1~1% が 2 つ、1~5% が 1 つ、5~10% がピリオドに対応する。

4. 考察

4.1 記事属性とベンチャー企業の成長速度

シード~シリーズ A~シリーズ B~シリーズ C すべての段階において、TechCrunch で一度でも社名が出た企業はそうでない企業に比べて、次のステージに行く率が高く、またその期間も短いことがわかる。影響の大きさとしては、地域属性や事業領域属性よりも大きい。また、Exit についても同様の効果が見られる。

図 1: Cox モデル：シードステージからシリーズ A まで

共変量	coef	exp(coef)	se(coef)	z	Pr(> z)
AGE	-0.510	0.600	0.040	-12.48	<2e-16 ***
DAYS_SINCE_INVESTED	-0.448	0.638	0.158	-2.84	4.51E-03 **
IN_BA	0.114	1.121	0.111	1.03	3.03E-01
IN_US	0.021	1.021	0.101	0.21	0.833615
BIO	-1.046	0.351	0.254	-4.108	4.00E-05 ***
ENVIRONMENT	-0.457	0.632	0.284	-1.608	0.107928
HARDWARE	-0.047	0.953	0.203	-0.232	0.81629
ENTERPRISE	0.320	1.377	0.089	3.56	3.71E-04 ***
COMMERCE	-0.34	0.706	0.306	-1.135	0.256172
AD	0.118	1.126	0.16	0.74	0.4589
GAME	0.021	1.021	0.156	0.137	0.891
FINANCE	0.193	1.214	0.155	1.25	0.211
EDUCATION	0.371	1.450	0.167	2.227	0.025 *
CLOSENESS_CENTRALITY	-0.018	0.982	0.048	-0.37	0.708
DEGREE_CENTRALITY	0.026	1.026	0.035	0.748	0.454
BETWEENNESS_CENTRALITY	-0.086	0.917	0.0492	-1.755	7.93E-02 .
WITH_TOPVCs_PR	0.316	1.372	0.113	2.799	0.0051 **
LOG_TOTAL_FUNDING	1.437	4.211	0.057	24.83	2e-16 ***
MILESTONES	-0.046	0.954	0.134	-0.343	0.731
ARTICLE_FLAG	0.332	1.394	0.0921	3.612	0.00030 ***

Concordance= 0.847
Rsquare= 0.203

図 2: Cox モデル：各投資ステージから Exit (売却) まで

共変量	coef	exp(coef)	se(coef)	z	Pr(> z)
AGE	0.137522	1.1474269	0.0512879	2.681	0.007332 **
DAYS_SINCE_INVESTED	0.3681708	1.4450888	0.1406589	2.617	0.008858 **
LAST_ROUND_SEED	-0.2998682	0.7409159	0.1465248	-2.047	0.040704 *
LAST_ROUND_A	-0.0325501	0.967974	0.1356838	-0.24	0.81041
LAST_ROUND_B	0.0555649	1.0571377	0.136151	0.408	0.683191
IN_BA	0.4808114	1.6173862	0.1135813	4.233	2.30E-05 ***
IN_US	0.4207926	1.5231683	0.1016985	4.138	3.51E-05 ***
BIO	0.1655362	1.1800257	0.1665874	0.994	0.320374
ENVIRONMENT	-0.6042466	0.546486	0.3086747	-1.958	0.050283 .
HARDWARE	-0.2145169	0.8069312	0.2029778	-1.057	0.290581
ENTERPRISE	0.3202982	1.3775386	0.0861874	3.716	0.000202 ***
COMMERCE	-0.8824005	0.4137884	0.1113904	-2.145	0.031959 *
AD	0.2539216	1.2890708	0.1245498	2.039	0.041478 *
GAME	0.0848834	1.0885902	0.1381346	0.614	0.538886
FINANCE	-0.0004816	0.9995185	0.1771946	-0.003	0.997831
EDUCATION	-0.3295919	0.7192172	0.234119	-1.408	0.159191
CLOSENESS_CENTRALITY	0.5666272	1.7623131	0.0786136	7.208	5.69E-13 ***
DEGREE_CENTRALITY	0.1102067	1.1165089	0.039735	2.774	0.005545 ***
BETWEENNESS_CENTRALITY	-0.0422309	0.9586484	0.052671	-0.802	0.422676
WITH_TOPVCs_PR	0.1382654	1.1482803	0.100581	1.375	0.169235
LOG_TOTAL_FUNDING	-0.1343708	0.8742658	0.0633343	-2.122	0.03387 *
MILESTONES	0.2127313	1.2370522	0.0983396	2.163	0.030523 *
ARTICLE_FLAG	0.6891383	1.9919983	0.0900381	7.654	1.95E-14 ***

Concordance= 0.61 (se = 0.012)
Rsquare= 0.072 (max possible= 0.99)

4.2 ネットワーク属性とベンチャー企業の成長速度

有力な VC とのつながりの有無はより直接的に成長に寄与していることがわかった。本研究では有力 VC の選出法として、共投資ネットワークにおける PageRank を利用した方法を採用したが、このようにして選ばれた VC からの投資がベンチャーの成長を早めることから、PageRank による VC の評価法自体の妥当性を示すことができたと考えられる。

売却による Exit については、近接中心性および次数中心性が高い企業は買収されやすいことが明らかになった。これらのスコアが高い企業は、Google や Facebook 等主要企業とのネットワーク上における距離も近いと言える。これらはいずれもベンチャー企業の買収に積極的な企業であり、彼らとの距離の近いほど買収の交渉が起りやすく、結果として Exit のしやすさにつながると考えられる。

4.3 その他属性

その他の共変量にも、成長に影響を与えるものがあつた。

設立から長い時間が経過したベンチャーは次の投資も起りにくいことがわかった。一方で売却および IPO による Exit については、古い企業ほど起りやすいことがわかった。Exit できるほどの規模になるにはある程度の時間がかかると考えられることから、これは自然な結果であると言える。

事業領域については、どの段階においてもバイオ系ベンチャーは成長が遅いことがわかった。これは、バイオ系事業は研究開発等に時間がかかるためと考えられる。また、シード～シリーズ A の初期においては、エンタープライズ系 IT 企業の成長が早いことがわかった。売却による Exit については、エンタープライズ系 IT および広告系が起りやすく、E コマースは起りにくいという結果となった。ただし、VC の投資領域にはトレンドがあるため、分析期間が異なれば違う結果が出る可能性もある。

地域情報に関しては、意外なことに、企業の所在地と成長の速度には有意な関係は見られなかった。一方で、売却による Exit の起りやすさに関しては、ベイエリアおよび他の米国地域は他国と比べ高い（ハザード比にして 1.52 1.63 倍）という結果が得られた。

財務属性については、累計の資金調達数が多い企業ほど次の投資も起りやすいことがわかった。より多くの資金を集めているところほど投資家にとって魅力的な事業を手がけていると言え、成長も早いと考えられる。

5. 結論

本研究では、生存時間解析の枠組みを用いて、ベンチャー企業の成長をモデル化した。具体的には、ベンチャーの投資ラウンドの推移（シード～シリーズ A～シリーズ B～シリーズ C～…～Exit）を成長の指標とし、その発生間隔を分析することで、ベンチャーの成長確度および成長スピードに影響を与える要因を特定した。結果として、ニュース記事に代表されるようなソーシャルな情報や、投資ネットワーク情報が、企業の成長の予兆として利用できることを明らかにした。また、事業領域や設立年数等によって成長スピードに有意な差があることもわかった。

本研究では各共変量による相対ハザードの分析に力点を置いたが、今後の発展として、基準ハザード関数自体を具体的に構成することで、ベンチャー企業の絶対的な成長確率を与えることが考えられる。これは VC や事業会社の投資戦略を考える上で有用であろう。

参考文献

- [Cox 72] Cox, D. R.: Regression models and life-tables, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 187–220 (1972)
- [Giot 07] Giot, P. and Schwiendbacher, A.: IPOs, trade sales and liquidations: Modelling venture capital exits using survival analysis, *Journal of Banking & Finance*, Vol. 31, No. 3, pp. 679–702 (2007)
- [Gompers 08] Gompers, P., Kovner, A., Lerner, J., and Scharfstein, D.: Venture capital investment cycles: The impact of public markets, *Journal of Financial Economics*, Vol. 87, No. 1, pp. 1–23 (2008)
- [Hochberg 07] Hochberg, Y. V., Ljungqvist, A., and Lu, Y.: Whom you know matters: Venture capital networks and investment performance, *The Journal of Finance*, Vol. 62, No. 1, pp. 251–301 (2007)
- [Shane 02] Shane, S. and Cable, D.: Network ties, reputation, and the financing of new ventures, *Management Science*, Vol. 48, No. 3, pp. 364–381 (2002)