

人材の転職履歴情報を素性とした ベンチャー企業評価

Startup Evaluation with employee
turnover info as Social Capital

上野山 勝也*¹ 大澤 昇平*¹ 松尾 豊*¹
Katsuya Uenoyama Shohei Osawa Yutaka Matsuo

*¹ 東京大学大学院
The University of Tokyo

This paper proposes a new method to evaluate startups using employee turnover history. We apply the scheme to Information Technology Industry and show the example. Also we use Crunchbase which has huge informations of talents turnover history. We define a series of math to predict the Exit of IPO and do forecast the success with SVM. Also we do feature generation with employee turnover network and select features which maximize precision and recall in the SVM discriminator. As a result, we got a good result with with employee turnover history and validate the method will work well.

1. はじめに

経済の成熟に伴いベンチャー企業への期待が高まっている。一方で、ベンチャー企業の事業活動は不確実性が高く、上場企業と異なり公開情報が少ないため、企業価値をなにかしらの方法で早期に評価出来れば、資金や人材の流入もより活発かつ効率的に行われるであろう。

これまでマイクロ経済学に於ける産業構造分析や企業動向分析の領域及びファイナンスの金融工学の領域で、ベンチャー企業評価 [Davilaa 03] や成功の予兆発見 [Chang 04] は行われてきた。

これまではデータを取得できるサンプル数が少なかったため、因子間の相関や因果の分析に留まる研究が多く、ベンチャー企業評価を機械学習の手法を用いた研究は少なかった。

一方で、最近 Web 上にベンチャー企業のデータが数多く公開されて始めている。本稿は、これらの Web 上に増加するデータを機械学習の手法を適用するに十分なサンプル数であると考へ、ベンチャー企業の企業価値評価を行おうという試みである。

本論文では、Web 上に増加するベンチャー企業の情報を用い機械学習の手法を活用し Exit を予測する 2 クラス分類器をベンチャー企業評価モデルとして構築する。Exit が予測される企業が企業価値が高いと考へる。対象企業への転職ネットワーク構造に着目し素性生成を行い、分類精度を最も向上させる素性群を調べる事でベンチャー企業評価モデルの評価を行う。

2. ベンチャー企業の既存の評価手法

古典的なベンチャー企業評価のアプローチは recourse based view [He 09] と呼ばれ、企業の成功は如何に社内資源を最適に配置するかにあるという立場をとる。分析対象の企業のどの属性情報が成功に影響を与えるかを解明しようという試みである。

次に、社会のネットワーク化に伴い、network resource combinations (以下 NRC) [Zeng 10] というアプローチが台頭してくる。対象企業の活動を単一ノードとして捉えるのではなく、その顧客、投資家、取引先といった外部の actor (行為者) をノードとし、ノード間が持つ資金流入、人の転職や取引関係といった関係情報をリンクとし、総体としてのネットワークとして捉

```

企業情報:Google,
foundedyear:1998年9月7日
address1: city:Mountain View, state code:CA,
従業員数:28000,
business category: search,
totalmoneyofraised:$25.1M,
foundedround, roundcode:angel, raisedamount:100000,
人材情報: nameA
所属企業 1: Research In Motio Title: Technical Director
過去所属企業 2: Appear Networks Systems, Title: CTO
過去所属企業 3: Ubitexx, Title: VP Development, Title: CTO
過去所属企業 4: Medhand International AB, Title: CTO

```

図 1: CrunchBase 掲載企業情報の例

え KSF を解明するというアプローチである。ノード間を資金、人材、技術、知識といった有形、無形の資産が流通することで価値が創造され企業が成功する [He 09] という立場である。現在も、ベンチャー企業を取り巻くネットワーク構造が成功にどう寄与するかを解明するために多くの研究が行われている。人的な資産がベンチャー企業の成功に利いていることが確認され、人的な資産という意味である **ソーシャルキャピタル** というキーワードで語られ、研究は加速している [He 09][Yli-Renko 02]。本研究も NRC の立場でソーシャルキャピタルに着目する。特に転職ネットワークが持つ素性を元に機械学習により Exit の予測を行う事でベンチャー企業の評価を行う。

ソーシャルキャピタルを分析するのに統計的に優れたサンプル数をベンチャー企業群から集めるのは実質的に不可能であったが、Web 上の情報を活用する事で本稿はそれを実現する

3. CrunchBase

CrunchBase (<http://techcrunch.com/tag/crunchbase/>) は企業及び投資家や従業員といった人材情報を掲載しており、2012年7月時点で、11万5422件の企業情報とそこに在籍する13万1179人分の人材情報が掲載されている。企業情報は多岐にわたる。例を図1に示す。また人材情報は氏名、役職、過去の在籍企業名等が掲載されている。本稿ではその中でも2万5406件の人材の転職履歴情報に着目する。

転職履歴情報の全体感を把握するために、企業をノードとし人材の転職フローをリンクとする方向性を持つ人材の転職ネットワーク図を図2に描画した。リンクは転職人数が2名以上

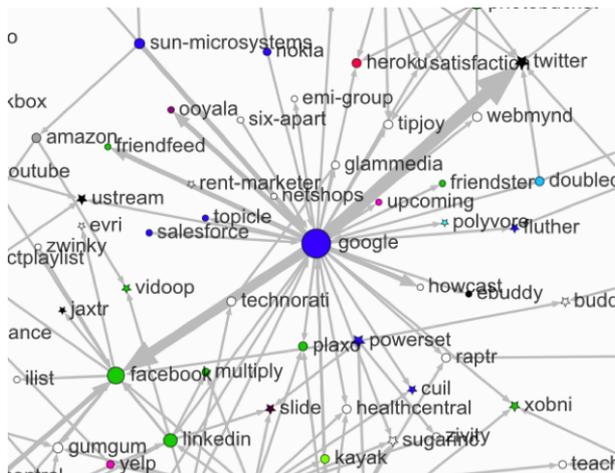


図 2: Google 周辺の転職人材の移動

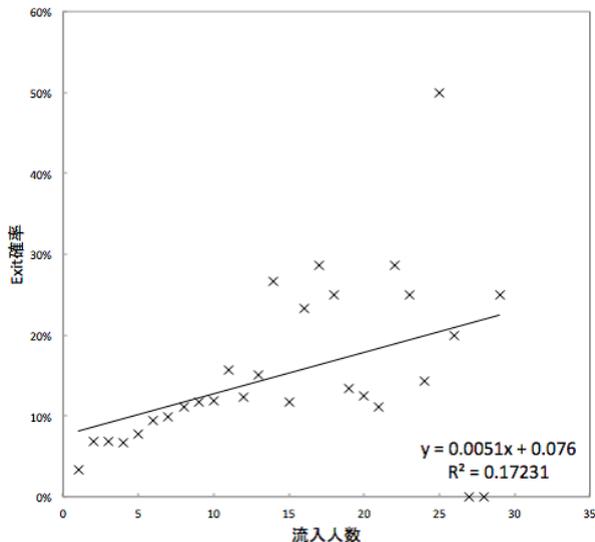


図 3: 人材の流入と Exit 確率の分布

のリンクを描画した。各ノードは企業を表し、リンクの太さは移動している人材の多さを表している。Yahoo や Google といった一流企業から小さなベンチャー企業に人材が流入しているのがわかる。一流企業を離れ、人材が流入するベンチャー企業はそうでないベンチャー企業より人材に撮っては魅力的で成功しそうな企業である可能性がある。

次に、人材の流入数と Exit 確率の分布を図 3 に示す。R 値が 0.17 で人材の流入数と Exit 確率は正に相関している事がわかる。つまり人材の流入がベンチャー企業の Exit 確率を高めていることがわかる。一方でこれだけで人材流入数が KSF であると結論できる事は出来ず、より体系的に素性生成を行い KSF を明らかにしていく必要がある。例えば、人材の流入経路や過去に所属していた企業の特徴が利いている可能性もある。次章に提案する手法によりソーシャルキャピタルに関する素性をより体系的に選択し、Exit に与える影響を分析する。

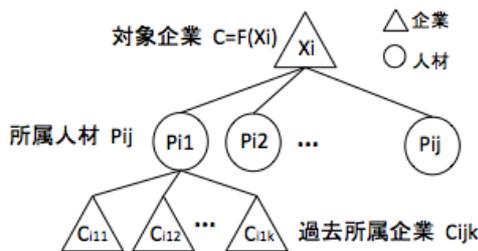


図 4: ソーシャルキャピタル係数の定量化モデル (n, m 追記)

4. 提案手法

ベンチャー企業が Exit する／しないを識別する 2 値分類器をベンチャー企業価値評価のために構築する。Exit と予測される企業が企業価値が高いと考える。分類アルゴリズムは、機械学習の分野で 2 値分類に標準的に用いられる SVM(サポートベクターマシーン)を用いる。各ノードに対し、転職ネットワーク構造及び資金流入ネットワーク構造を用いて素性生成 (Feature Generation) を行い、Exit の予測精度を高める素性選択 (Feature Selection) を実施することで KSF を特定する。より高精度に Exit する／しないを識別できる素性群は、ベンチャー企業の設立から成功に至る複雑な因果連鎖の中で Exit に影響を与えている素性群であると考えられる。特に、本稿では人材転職ネットワークに着目しこのソーシャルキャピタルに関連する因子が成功の予測に影響を与えるという仮説の元にソーシャルキャピタルの定量化モデルを構築し、Exit の予測を試みる。

4.1 転職ネットワーク構造を用いた素性生成

転職ネットワーク構造に基づく素性を元にノードである企業の Exit する／しないの予測を実施した。

$$C_i \in C \tag{1}$$

$$\phi(C_i) = \phi_{financial}(C_i) * \phi_{social}(C_i) \tag{2}$$

まず転職ネットワークを構成するベンチャー企業集合 C を考える。ベンチャー企業集合 C に属するベンチャー企業 C_i に対し、KSF であると想定される素性群 $\phi(C_i)$ を定義する。素性群 $\phi(C_i)$ は調達資金に関する素性群 $\phi_{financial}(C_i)$ とソーシャルキャピタルに関する素性群 $\phi_{social}(C_i)$ の積で表される。素性選択の複数シナリオにおける分類精度を測定し分類精度を最大化させる素性群を選択することでベンチャー企業の KSF を特定する。仮説として、ソーシャルキャピタルに関する因子 $\phi_{social}(C_i)$ を素性に組み込むことで予測精度が向上するのではないかと考えた。その仮説を検証するために、複数のシナリオで素性選択を行い、人材の流入に関する因子を素性に組み込む事で予測の精度がどのように変わるかを確認した。ベンチャー企業の KSF の候補として素性にはソーシャルキャピタルに関する素性群 $\phi_{social}(C_i)$ と金融資本に関する素性群 $\phi_{financial}(C_i)$ を定義した。詳細は表 1 に示す。

$\phi_{financial}(C_i)$ として 1. 累計資金調達金額, 2. 資金調達回数, 3. 企業の各成長段階における資金調達有無 の 3 つを定義

表 1: ベンチャー企業の KSF 候補

$\phi_{financial}(C_i)$: 金融資本に関する因子	
ϕ_{f1}	: 累計資金調達金額
ϕ_{f2}	: 各資金調達ラウンドでの資金調達有無
ϕ_{f3}	: 累計資金調達回数
$\phi_{social}(C_i)$: ソーシャルキャピタルに関する因子	
ϕ_{s1}	: 過去所属組織の従業員数
ϕ_{s2}	: 過去所属組織の Exit 有無
ϕ_{s3}	: 過去所属組織の累計資金調達金額
ϕ_{s4}	: 過去所属組織の人材輩出数

した. $\phi_{social}(C_i)$ として 1. 過去在籍企業の従業員数, 2. 過去在籍企業が Exit したか, 3. 過去在籍企業の累計資金調達額, 4. 過去在籍企業の人材輩出数の 4 つを定義した.

次に $\phi_{social}(C_i)$ の定量化モデルを図 4 に示す. 2 つのステップにより, $\phi_{social}(C_i)$ を求める. 第一に, $\phi_{social}(C_i)$ は現在対象企業 C_i に所属する各人材 P_{ij} に定義されるソーシャルキャピタル係数 $\phi_{social}(C_{ijk})$ の和である. 第二に, 各人材 P_{ij} のソーシャルキャピタル係数は j 人の人材 P_{ij} が過去に所属していた k 個の企業 C_{ijk} に振られるソーシャルキャピタル係数 $\phi_{social}(C_{ijk})$ の和である. つまり, 対象企業 C_i の $\phi_{social}(C_i)$ を式 3 として定式化される.

$$P_{ij} \in P_i \quad (3)$$

$$\phi_{social}(C_i) = \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \phi_{social}(C_{ijk}) \quad (4)$$

4.2 素性選択によるモデル評価

前項により定義された素性群 $\phi(C_i)$ に対し, 素性選択を行う事でベンチャー企業の KSF を特定する. 素性選択は大別すると, フィルタリング法 [He 09] とラッパー法 [He 09] の二つが存在する. フィルタリング法とは学習の前に事前に素性の特性に基づき素性群の選別を行う手法である. 一方で, ラッパー法とは特徴量の部分集合を使って実際に学習し, 精度を最大にする素性群の部分集合を特定する. 本稿はラッパー法を用いる. 前項により定義した, ベンチャー企業 C_i の KSF であると想定できる素性群 $\phi(C_i)$ に対し, Exit する/しないを高精度で識別する分類器を形成する素性群のセット φ を KSF として特定する. 入力クラスとなる素性群のセットは全パターンで実験を行い, 最も高精度を得る素性群 $\phi(C_i)$ を特定する.

$$X_i \in \phi(C_i), C_i \in \{-1, 1\} \quad (5)$$

$$X_i \in \varphi(C_i), \varphi \subset \phi, C_i \in \{-1, 1\} \quad (6)$$

次に基底関数として式 (7) を用いた.

$$x, x^2, \sqrt{x}, \log x \quad (7)$$

5. ベンチャー企業の評価結果

対象データとして, CrunchBase 上から取得したデータの内, 設立年度が記載されておりかつ在籍する人材が 1 名以上記載されている企業 4512 社を対象とした. その内 Exit している企業数は 363 社である. 対象データをランダムに学習データ 2256 サンプル (うち, 正例 181 社) とテストデータ 2256 サンプル (うち, 正例 182 社) に分割し, 学習及びテストを行っ

た. SVM のパラメータとしてはグリッドサーチにより特定した $C = 30, 000$ 及び $\gamma = 0.1$ を用いた.

図 5: 各シナリオにおける予測精度と適合率

	ϕ_{f1}	ϕ_{f2}	ϕ_{f3}	ϕ_{s1}	ϕ_{s2}	ϕ_{s3}	ϕ_{s4}	$x, x^2, \sqrt{x}, \log x$		
								Pre.	Rec.	F
RP								8.1	8.1	8.1
FC	*							100	0.7	1.4
		*						14.3	0.6	1.2
			*					17.6	1.8	3.3
	*	*	*					NaN	0.0	NaN
	*	*	*	*				100	0.7	1.4
	*	*	*					12.5	0.6	1.1
	*	*	*	*				16.1	2.7	4.7
FC+SC	*	*	*	*				28.6	3.6	6.3
	*	*	*	*	*			23.8	3.0	5.4
	*	*	*	*	*	*		22.6	4.1	7.0
	*	*	*	*	*	*	*	29.4	3.0	5.4
	*	*	*	*	*	*	*	32.1	5.2	9.0
	*	*	*	*	*	*	*	22.2	3.6	6.1
	*	*	*	*	*	*	*	26.5	5.1	8.6
	*	*	*	*	*	*	*	28.6	3.6	6.3
	*	*	*	*	*	*	*	30.4	4.0	7.0
	*	*	*	*	*	*	*	27.6	4.5	7.7
	*	*	*	*	*	*	*	18.8	3.5	5.9
	*	*	*	*	*	*	*	25.9	4.1	7.1
	*	*	*	*	*	*	*	29.0	5.2	8.8
	*	*	*	*	*	*	*	18.2	3.6	5.9
	*	*	*	*	*	*	*	21.6	4.5	7.5

組成選択 (Feature Selection) においてどの素性を素性群に含めるかで 2 の 7 乗引く 1 の 127 通りのシナリオが存在する. まず, $\phi_{financial}(C_i)$ のみを含めて Exit の予測を行った際の結果を図 5 のシナリオ FC に示す. 次に, $\phi_{social}(C_i)$ も含めて Exit の予測を行った際の結果をシナリオ FC+SC に示す.

第一に, $\phi_{social}(C_i)$ を含めないシナリオよりも含めたシナリオの方が大幅に精度及び適合率が良いことが確認できる. F-Score でみると FC では 1.1 から 4.7 であるが, FC+SC であれば 5.4 から 9.0 で約 2 倍から 4 倍の精度になっている事が分かる. つまり, $\phi_{social}(C_i)$ を含めて SVM 分類器を作成することにより精度の高い Exit 予測が可能になる事が証明された. これが本稿の最も大きな主張であり, これまで余り議論される事なかった人材に関する転職履歴情報を用いてベンチャー企業の評価を実施する事で, より高い精度と適合率でベンチャー企業の Exit が予測できる事が証明された.

第二に, どのような $\phi_{social}(C_i)$ を含めると, より予測精度が高まるのかであるが, 本実験の結果によると ϕ_{s1} と ϕ_{s3} を含めた場合が最も精度高く Precision が 32.1 であり Recall が 5.2 であり F-score が 9.0 であることがわかる. また ϕ_{s1} から ϕ_{s4} の 4 つの各素性が含まれているシナリオの F-Score の平均値は, ϕ_{s1} , ϕ_{s2} , ϕ_{s3} , ϕ_{s4} の順番に 7.4, 6.8, 7.3, 7.0 であり, ϕ_{s1} , ϕ_{s3} , ϕ_{s4} , ϕ_{s2} の順番に予測精度の向上に寄与している事が言える. これらの 4 つの素性は各々独立ではないが, 平均値で見ると従業員数が多い企業からの転職者が多い企業, 次に過去の資金調達金額が多い企業からの転職者が多い企業が高い確率で Exit していることがわかる. 一方で, 過去所属組織の人材輩出数や Exit の有無はそれらに比べると Exit 予測の精度に寄与しない事がわかる. これらにより大企業や資金調達に成功しているにも関わらず, 転職してきている人材が多い企業は高い確率で Exit しており, 優秀な人材を惹き付ける何かしらの理由があり, それが Exit に寄与していると解釈する事が出来る.

以上により, 転職履歴情報を用いる事でよりよいベンチャー企業評価が可能となる事が示された.

参考文献

- [Davilaa 03] Antonio Davilaa, George Foster, Mahendra Gupta.: Venture capital financing and the growth of startup firms, *Journal of Business Venturing* Vol 18 689-708 (2003)
- [Chang 04] Sea Jin Chang.: Venture capital financing, strategic alliances, and the initial public offerings of Internet startups, *Journal of Business Venturing* Vol 19 721-741 (2004)
- [He 09] Jiang He, M. Hosein Fallah.: Is inventor network structure a predictor of cluster evolution?, *Technological Forecasting and Social Change*, Vol. 76, pp 91-106 (2009)
- [Yli-Renko 02] H Yli-Renko.: Social capital, knowledge, and the international growth of technology-based new firms *International Business Review*, Vol 11 pp 279-304 (2002)
- [Zeng 10] S.X. Zeng, X.M.Xie, C.M.Tam.: Relationship between cooperation networks and innovation performance of SMEs, *Technovation* Vol 30 pp. 181-194 (2010)