

# 人材の転職履歴情報を素性とした ベンチャー企業のExit予測

上野山 勝也<sup>1,a)</sup> 大澤 昇平<sup>1,b)</sup> 松尾 豊<sup>2,c)</sup>

受付日 2013年10月9日, 採録日 2014年7月11日

**概要:** 経済の成熟にともないベンチャー企業への期待が高まっている。ベンチャー企業の上場または事業売却（以降 Exit）を高い精度で予測できれば、資金や人材はより適正なベンチャー企業に移動するため価値は高い。一方で、ベンチャー企業の Exit を予測するモデルは、データ取得の制約からベンチャー企業の「社内資源」特に資金調達に関する素性を元にしたものが多かった。一方で本稿は「社外資源」である創業メンバーや従業員が持つ過去の人間関係に関わる資産が Exit に寄与しているという仮説に基づき Exit を予測する手法を提案する。Web 上に構築されている Cruchbase という人材データベースを活用することで、これまでデータ取得が困難であった転職履歴情報を活用することで Exit を予測する手法を提案する。2 万社に対し人材の転職履歴情報を用いて Exit の予測を行うことで、用いない手法より 10 ポイント高い精度で Exit を予測できることを確認した。

**キーワード:** SVM, 素性選択, ソーシャルキャピタル, ベンチャー企業, ウェブマイニング

## Startup Evaluation with employee turnover info as Social Capital (version 2014/06/29)

KATSUYA UENOYAMA<sup>1,a)</sup> SHOHEI OSAWA<sup>1,b)</sup> YUTAKA MATSUO<sup>2,c)</sup>

Received: October 9, 2013, Accepted: July 11, 2014

**Abstract:** 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.

**Keywords:** support vector machine, feature generation, social capital, startups, web mining

### 1. はじめに

経済の成熟にともないベンチャー企業への期待が高まっている。IT 業界を中心に創業コストが下がり、米国では現

在年間 70 万社以上のベンチャー企業が創業され、年々その数は増加している。大手企業は、ベンチャー企業が持つ技術や人材を買収により内製化することで成長している。IT 業界は分裂と合体を繰り返し産業が渾然一体となって成長している。米国の IT 企業に於ける人材流動性は非常に高く、その流動性が経済成長を支えているともいえる。

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、成長の予測が困難である。成功要因を発見しそれを他のベンチャー企業に活用することで成功確率を高められれば、またより早い段

<sup>1</sup> 東京大学大学院工学系研究科  
School of Engineering, The University of Tokyo, Bunkyo,  
Tokyo 135-8656, Japan

<sup>2</sup> スタンフォード大学  
Stanford University, Stanford, CA 94305, United States

a) uenoyama@weblab.t.u-tokyo.ac.jp

b) ohsawa@weblab.t.u-tokyo.ac.jp

c) matsuo@weblab.t.u-tokyo.ac.jp

階でベンチャー企業の成長を予測できれば、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おける産業構造分析や企業動向分析の領域およびファイナンスの金融工学の領域で、ベンチャー企業の成功要因の分析 [1] や Exit 予測という試み [2] は行われてきた。ベンチャー企業の成功要因となる因子は Key Success Factor [6] (以下 **KSF**) と呼ばれ、ベンチャー企業が成功に至るまでの複雑な因果連鎖の中に再現性ある KSF という因子が存在するという立場で、既存研究の多くはその解明に努めてきた。因子間の相関や因果を統計的な手法を用いて分析し KSF の存在を証明するというアプローチであった。因子設定の対象は企業内に留まらず、社外の投資家や顧客との関係性 [4] といった社外の因子を対象を広げている。

一方で、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、ベンチャー企業の事業売却もしくは IPO (以降 Exit と呼ぶ) を機械学習の手法を用いて統計的に予測しようという研究は少なかった。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。Crunchbase<sup>\*1</sup>や Angel List<sup>\*2</sup>など、ベンチャー企業のデータベースが構築され、サンプル数は数万を超え急速に増加している。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え、Exit の予測、成功要因の特定を行おうという試みである。

本稿では、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 クラス分類器を構築する。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ことでモデルの最適化を行うと同時に、ベンチャー企業の KSF の特定を行う。

構成としては、2 章でベンチャー企業の旧来の分析手法について外観し、3 章で人材データベースである CrunchBase の説明を行う。4 章で今回提案する転職ネットワークを活用した Exit 予測のための分類器を定義する。5 章で分類精度に基づき分類器の素性選択を行うことでベンチャー企業の Exit 予測および KSF の特定を行う。6 章で本稿の手法の研究的意義を再考する。

## 2. 関連研究

### 2.1 ベンチャー企業の評価指標

企業の評価に関する研究は、社内資源のみを用いたものと、社内資源に加え社外資源を用いたものに二分される。本研究では、古典的アプローチである前者を resource based view、後者を network resource combinations (以下

NRC) アプローチと呼んで区別する。本研究は 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予測を行う。以下では、それぞれの特徴および課題について述べる。

古典的なベンチャー企業評価のアプローチは resource based view [5] 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。Sambasivan ら [7] は起業家の個人スキルが Venture Paformance に与える影響をマレーシアのベンチャー企業 1,275 社にアンケートし検証した。また Ardichvili [9] は Personal Quality は企業パフォーマンスに影響を与えず、Motivation や戦略が Performance に影響を与えることを証明した。このように、ベンチャー企業の属性情報に着目し、成功という結果に対する因果関係を分析するというアプローチであった。このようなアプローチの持つ課題点として、同種の社内資源を持つ場合に、人材流入や取引先などの社外の要因の区別ができないといった課題点があった。

次に、社会のネットワーク化にともない、NRC [8] 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とらえるのではなく、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、ノード間を持つ資金流入、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とらえ 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。Gassel ら [10] は bio-tech 業界を、1 つの企業下のネットワークではなく network 資産が重要と考え、ネットワーク資産を評価し、ネットワーク資産がベンチャー企業の KSF であることを示した。他の研究 [3], [11] でも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。このアプローチの課題点は、データの収集である。古典的アプローチに加えて、社外にまで調査対象を広げるといのは、データの収集コストが数倍以上に膨れ上がってしまう。そのため、NRC のケーパビリティの高さに比べて、実務上は古典的アプローチを取らざるを得なかった。

前述のとおり、これまではソーシャルキャピタルを分析するのに統計的に有意な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。ベンチャー企業は世界中に点在し情報はあまり公開されず研究は限定的であった。しかし、近年 Web 上に多くのベンチャー企業のデータが公開され始めている、本研究ではベンチャー企業

\*1 <http://citeseer.ist.psu.edu/>

\*2 <http://scholar.google.com>

に勤務する被雇用者の過去の転職履歴に着目した方法論を示す。本研究はデータが限定的であったベンチャー企業分析に対し近年増加する Web 上の情報を活用し定量的な分析を行った新たな研究である。

### 2.2 ウェブ上に増加するベンチャー企業情報

Web 上で公開されるデータの激増している。CrunchBase はインターネット業界を中心とした企業と人材の基礎情報が公開されており。人材の転職履歴情報も掲載されている。Angel List は多くのエンジェルやベンチャー企業で働く人々の情報が公開され、どのエンジェルがどのベンチャー企業に着目しているかという情報も増えている。Crunchbase には現在 20 万件のベンチャー情報が蓄積し、Angel List には 2 万 6 千件の人材情報が蓄積されており、これらは日々増加の一途をたどっている。

こうしたデータを活用した研究も行われつつあり、Werthら [13] は CrunchBase に蓄積する共同出資関係に着目し、ベンチャーキャピタリスト間の共同出資関係がベンチャー企業の成功に与える影響を分析し、Alexy ら [14] はベンチャーキャピタリストが持つ人間関係を成功確率の関係を分析した。一方で、特に被雇用者が持つソーシャルキャピタルがベンチャー企業の成功に与える影響に着目した研究は筆者らの知る限り行われておらず、本研究は被雇用者の転職履歴に KSF があるのではないかと仮説の元、論旨を展開する。

## 3. CrunchBase

### 3.1 ベンチャー企業データベース

CrunchBase (<http://techcrunch.com/tag/crunchbase/>) とはアメリカ TechCrunch 社が構築したベンチャー企業のデータベースである。取材先の選定のために構築されたが企業情報を各企業が編集でき、現在では PR や採用のために各社が情報を自ら掲載している。企業の実績や人材の経歴を PR するための情報基盤として主に米国で使われている。図 1 は企業および人材に関する掲載情報の一例である。本稿で使用した項目を中心に記載している。

```

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
total money ofraised:$25.1M
founded round, round: angel, raised amount:100000,
round: seriesA, raised amount:200000,
:
    
```

図 1 CrunchBase に掲載される企業情報  
Fig. 1 Company information on CrunchBase.

### 3.2 蓄積データ

CrunchBase は企業およ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 5,422 件の企業情報とそこに在籍する 13 万 1,179 人分の人材情報が掲載されている。企業情報は企業名、企業設立年数、資金調達金額、タイミング、会社概要など多岐にわたる。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、Twitter アカウント、ホームページや Blog の URL などが掲載されている。本稿ではその中でも 2 万 5,406 件の人材の転職履歴情報に着目する (図 2)。

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 4 に描画した。リンクは転職人数が 2 名以上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。図 3 は図 4 の拡大図であるが、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分かる。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と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 5 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ことが分かる。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分かる。また人材移動の測定が Web のニュースなどに公表されないシリコンバレーという土地に粘着した情報が存在しそれを検出できるセンサである可能性が示唆されている。一方でこれだけで人材流入数が KSF であると結論できることはできず、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。たとえば、人材の流入経路や過去に所属していた企業の特性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。

今回は CrunchBase の情報を用いているが、転職履歴情報を別の方法で取得できるのであれば別のデータを用い

```

nameA: Tom
現所属企業 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
過去所属企業 4: Medhand International AB, Title: CTO
nameB: Mike
現所属企業 1: Grockit Title: VP of Marketing
過去所属企業 2: Jangl, Title: VP Marketing
過去所属企業 3: Outspark, Title: VP of Marketing
nameC: Sarah
現所属企業 1: Google.inc Title: Software Engineer
:
    
```

図 2 CrunchBase に掲載される転職履歴情報  
Fig. 2 Turnover histories on CrunchBase.



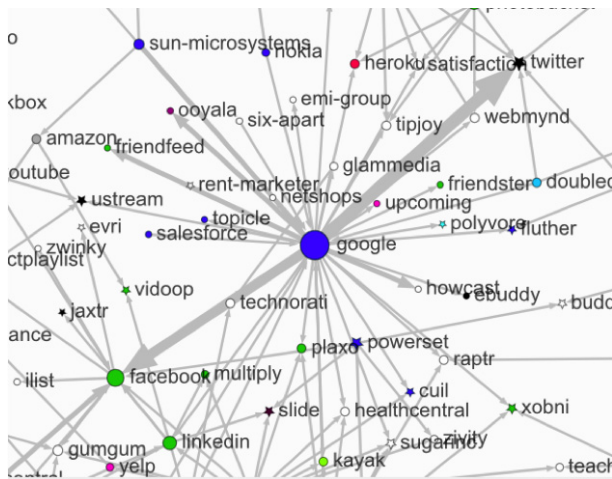


図 3 Google 周辺の転職人材の移動  
Fig. 3 Turnover network around Google.

でも実施できる汎用的な手法を提案する、データの取得可能性がウェブにより広がったため、本稿は Crunchbase のデータを用い手法の有効性を主張するが、他のデータでも適用可能な汎用的手法の提案を意図している。

#### 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 値分類器を作成する。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ン) を用いる。各ノードに対し、転職ネットワーク構造およ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

ソーシャルキャピタルおよび資金流入のネットワーク構造を用い素性生成を行う。転職ネットワークおよび、資金流入ネットワークに関する素性からノードとなる企業を Exit する/しないに分類するタスクはリンクマイニング<sup>\*3</sup>の分野における「リンクに基づく分類 (Link-based classification)」であるといえる。ネットワーク構造を持つデータに対し学習や予測を行う研究は多く行われており、本研究も該当分野の研究の潮流に則る。Backstorm ら [15] は、ノードの周

<sup>\*3</sup> LinkKDD と呼ばれるワークショップが 2003 年から開催されており、また ACM SIGKDD の会誌である Explorations でも Link Mining の特集が組まれている。

辺のネットワーク構造を用い新たな属性を生成することでリンクに基づくノード分類の有効性を示した。松尾ら [17] は、リンクに基づく分類のための素性生成に、体系的にネットワーク構造を用いる方法論を提案している。本稿でも同様に、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情報を元にソーシャルキャピタルを測定する意味に言及する。ソーシャルキャピタルとは対象人材に帰属する知識や人間関係といった無形資産を指し、本稿でも、分類対象となる企業に転職を通じ流れ込む人材が、ソーシャルキャピタルを保有していると考えられる。つまり対象企業のソーシャルキャピタルは、対象企業に転職してきた人材集合が保有するソーシャルキャピタルの和として定量化される。次に対象人材がどの程度のソーシャルキャピタルを保有しているかをどう定量化するかという論点が残る。本稿では転職情報に着目し、過去の転職履歴でどのような経験や知識を積み人間関係を育ててきたかを定量化することで対象人材が保有するソーシャルキャピタルを定量化する。一方で転職履歴情報からソーシャルキャピタルを定量化するには、ソーシャルキャピタルをどのような式で定量化するか、過去所属企業どうしを独立事象ととらえるのか非独立事象ととらえるのか等議論の余地が多くある。本稿では、過去所属企業どうしを独立ととらえ、ソーシャルキャピタルをどのような式で定量化するかに議論を絞る。つまり対象人材が持つソーシャルキャピタルは過去所属企業群から得るソーシャルキャピタルの和であると考えその定量化の式を特定する。複数の式を素性群に組み込み、分類精度を最大化する素性選択をすることで、最適な定量化の式を特定する。

次に以上のソーシャルキャピタルの定量化の議論を定式化する。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こと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

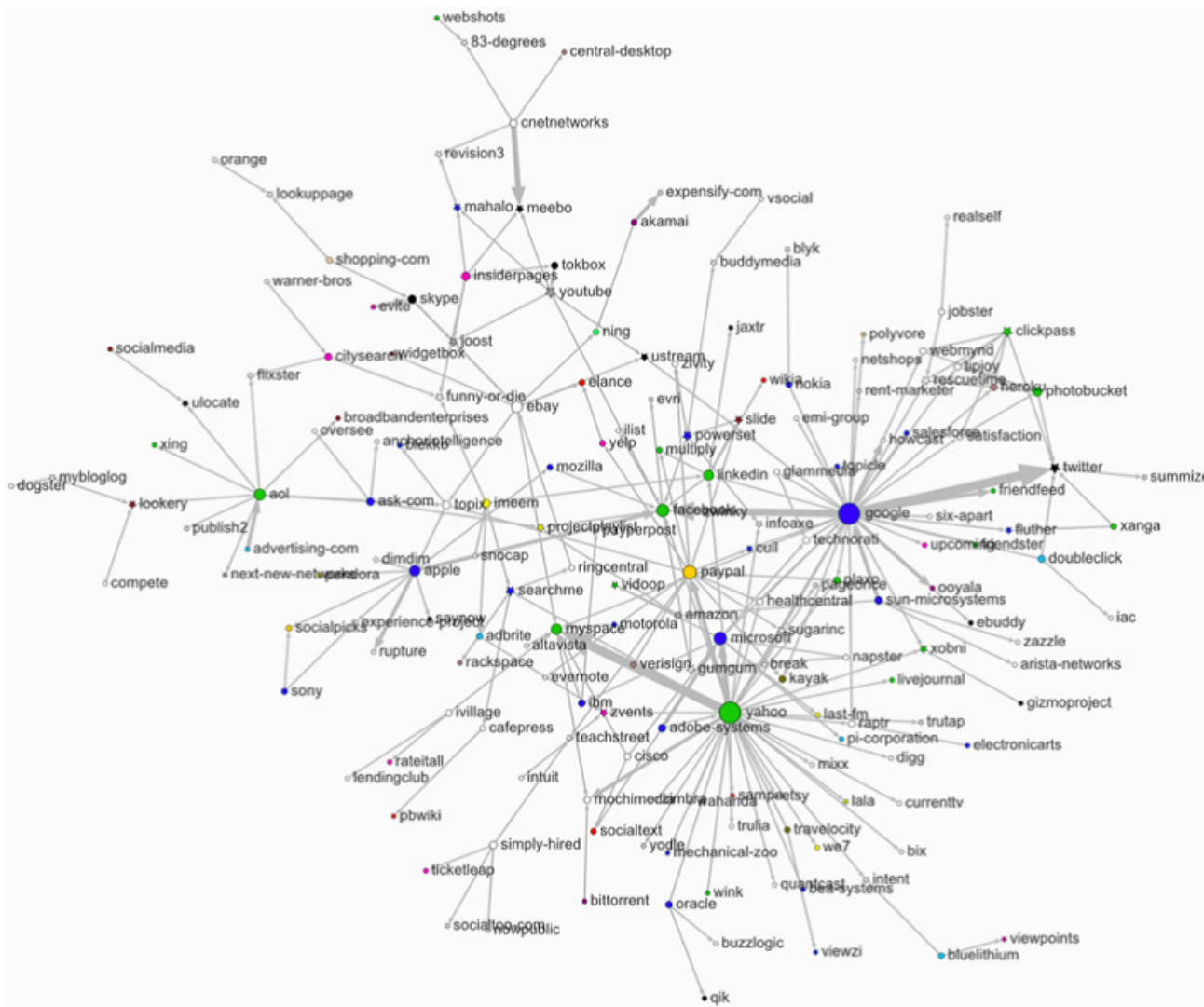


図 4 CrunchBase により作成したベンチャー企業間に於ける人材の転職ネットワーク  
 Fig. 4 Turnover network generated by information on CrunchBase.

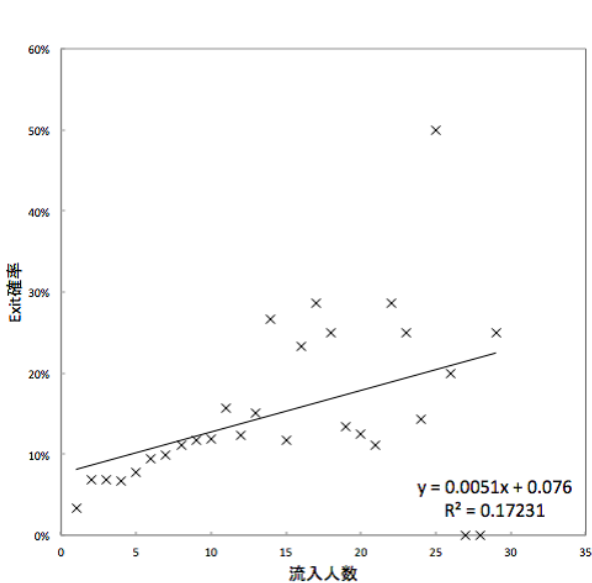


図 5 人材の流入と Exit 確率の分布  
 Fig. 5 Number of turnovers vs. Exit probability.

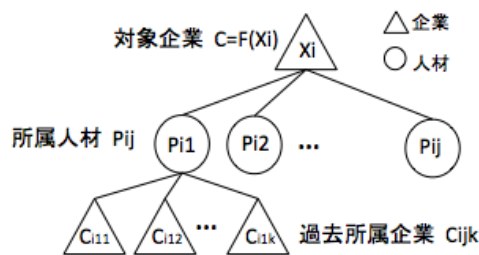


図 6 ソーシャルキャピタル係数の定量化モデル (n, m 追記)  
 Fig. 6 Framework for social capital quantification.

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。  
 $\phi_{financial}(C_i)$  に資金調達タイミングの素性を採用することで企業の成長ステージごとの分析, 考察が可能とな

表 1 ベンチャー企業の KSF 候補  
Table 1 Key success factors.

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

るが、今回はデータサンプル数の制約から、企業の成長ステージごとの分析は対象としない。

次に  $\phi_{social}(C_i)$  の定量化モデルを図 1 に示す。2つのステップにより、 $\phi_{social}(C_i)$  を求める。第 1 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 2 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 KSF の特定

前節により定義された素性群  $\phi(C_i)$  に対し、素性選択を行うこと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、フィルタリング法とラッパー法 [16] の 2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。前節により定義した、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  $\varphi$  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。

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

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

## 4.3 SVM 分類パラメーターの最適化

### 4.3.1 モデルパラメーターの最適化

次に分類精度を向上させるため、SVM 分類器の最適化を行う。カーネル関数は標準的な RBF カーネルを用いた。RBF カーネルを式 (4) に示す。

$$k(\mathbf{x}, \mathbf{y}) = \exp(-\gamma|\mathbf{x} - \mathbf{y}|^2) \quad (7)$$

ベンチャー企業の 2 値分類はたとえば企業が多くの資金調達をしたからといって必ずしも Exit するわけでない。つまり実態としてベンチャー企業のカテゴリは境界平面で完全に分離できるわけではなく、一定程度の重なりがあると考えられる。したがって、境界に一定程度の重なりを許すソフトマージン SVM を用いた。また分類器の精度を向上させるため、ソフトマージン SVM が持つパラメータ  $C$  と RBF カーネルが持つパラメータ  $\gamma$  の最適な値をグリッドサーチで特定した。

### 4.3.2 基底関数の最適化

次に基底関数の最適化を考える。人材や資金という因子が Exit に対し、線形に影響力を増すという基底関数も想定できるが、チームや人材といったものは非線形に影響が増していきどこかの閾値を超えたタイミングで Exit に対する影響が増加するという基底関数も考えられる。したがって実際には下記 4 パターンの基底関数を定義し、どの基底関数が精度高い予測につながるかを評価することで、SVM の最適化を計った。

$$x, x^2, \sqrt{x}, \log x \quad (8)$$

## 5. 評価

### 5.1 対象データ

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 4,512 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 2,256 サンプル（うち、正例 181 社）とテストデータ 2,256 サンプル（うち、正例 182 社）に分割し、学習およびテストを行った。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  $C = 3 \times 10^4$  および  $\gamma = 0.1$  を用いた。

### 5.2 評価指標

評価指標には、適合率 (precision)  $P$ 、再現率 (recall)  $R$ 、 $F$  値 (F-value) の 3 つを用いる。テストデータ中に含まれる Exit した企業集合  $E$  と、本手法が予測した企業集合  $\tilde{E}$  に対して、それぞれは次式によって定義される。

$$P = \frac{|E \cap \tilde{E}|}{|\tilde{E}|}, R = \frac{|E \cap \tilde{E}|}{|E|}, F = \frac{PR}{P + R} \quad (9)$$

### 5.3 ベンチャー企業の Exit 予測結果

表 1 に示す通り、候補となる素性は 7 種類存在するため、素性選択 (Feature Selection) においてどの素性を素性群に含めるかで  $2^7 - 1 = 127$  通りのシナリオが存在する。このうち、素性群  $\phi_{financial}(C_i)$  のみから素性を選択



表 2 各シナリオにおける予測精度と適合率  
Table 2 Precision and recall by KSFs.

	$\phi_{f1}$	$\phi_{f2}$	$\phi_{f3}$	$\phi_{s1}$	$\phi_{s2}$	$\phi_{s3}$	$\phi_{s4}$	$x$	$x^2$			$\sqrt{x}$			$\log x$			$x, x^2, \sqrt{x}, \log x$				
									Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.	F	Pre.	Rec.
RP								8.1	8.1	8.1	8.1	8.1	8.1	8.1	8.1	8.1	8.1	8.1	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	-	-	-	-	-	-	-	-	-	-	-	
	*	*	*				<b>16.1</b>	<b>2.7</b>	<b>4.7</b>	-	-	-	-	-	-	-	-	-	-	-	-	
FC+SC	*	*	*	*				28.6	3.6	6.3	16.7	1.7	3.1	25.0	1.8	3.3	21.7	3.0	5.3	27.5	6.5	10.5
	*	*	*		*			23.8	3.0	5.4	15.8	1.8	3.2	20.0	1.8	3.3	28.6	3.4	6.1	24.1	4.1	7.1
	*	*	*	*	*			22.6	4.1	7.0	15.8	1.8	3.2	15.0	1.8	3.2	20.0	2.4	4.3	24.6	8.0	12.1
	*	*	*			*		29.4	3.0	5.4	24.0	3.5	6.1	25.0	2.4	4.3	26.1	3.5	6.2	19.1	5.3	8.3
	*	*	*	*		*		<b>32.1</b>	<b>5.2</b>	<b>9.0</b>	21.4	3.4	5.9	<b>23.8</b>	<b>3.0</b>	<b>5.3</b>	27.6	4.5	7.8	23.2	7.5	11.3
	*	*	*		*	*		22.2	3.6	6.1	13.6	1.8	3.1	18.2	2.4	4.3	15.8	1.8	3.2	19.7	7.5	10.8
	*	*	*	*	*	*		26.5	5.1	8.6	16.0	2.4	4.1	15.4	2.3	4.0	24.0	3.4	6.0	19.7	7.3	10.7
	*	*	*	*		*		28.6	3.6	6.3	16.7	1.7	3.1	25.0	1.8	3.3	21.7	3.0	5.3	27.5	6.5	10.5
	*	*	*			*		30.4	4.0	7.0	<b>33.3</b>	<b>4.1</b>	<b>7.3</b>	25.0	2.4	4.4	22.7	3.0	5.2	21.4	5.3	8.5
	*	*	*		*	*		27.6	4.5	7.7	20.7	3.6	6.1	14.3	1.7	3.1	22.7	3.0	5.2	19.2	8.0	11.2
	*	*	*	*	*	*		18.8	3.5	5.9	13.3	2.4	4.0	15.4	2.4	4.1	24.0	3.5	6.1	18.4	8.2	11.3
	*	*	*		*	*		25.9	4.1	7.1	21.9	4.1	7.0	23.5	2.3	4.2	<b>25.8</b>	<b>4.8</b>	<b>8.1</b>	18.3	6.3	9.3
	*	*	*	*		*	*	29.0	5.2	8.8	19.4	4.1	6.8	22.7	3.0	5.2	17.6	3.6	5.9	19.1	7.6	10.9
	*	*	*		*	*	*	18.2	3.6	5.9	18.4	4.2	6.8	14.3	2.4	4.1	21.2	3.9	6.6	18.3	8.4	11.5
*	*	*	*	*	*	*	21.6	4.5	7.5	16.3	4.1	6.6	14.3	2.2	3.9	17.1	3.6	5.9	<b>24.7</b>	<b>11.0</b>	<b>15.2</b>	

するシナリオを FC (FinanCial) によって表記し、FC におけるすべての素性に加え、素性群  $\phi_{social}(C_i)$  の素性を加えるシナリオを FC+SC (SoCial) により表記する。また、比較対象として、一様分布からランダムに予測を行う RP (Random Predictor) もシナリオに加える。

それぞれのシナリオに対する実験結果を表 2 に示す。まず、 $\phi_{social}(C_i)$  を含まないシナリオよりも含めたシナリオの方が大幅に精度およ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ことが分かる。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 予測が可能になることが証明された。

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ことの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ことで、より高い精度と適合率でベンチャー企業の Exit が予測できることが証明された。

次に、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を分析する、本実験の結果によると  $\phi_{s1}$  と  $\phi_{s3}$  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 であることが分かる。また  $\phi_{s1}$  から  $\phi_{s4}$  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、 $\phi_{s1}$ ,  $\phi_{s2}$ ,  $\phi_{s3}$ ,  $\phi_{s4}$  の順番に 7.4, 6.8, 7.3, 7.0 であり、 $\phi_{s1}$ ,  $\phi_{s3}$ ,  $\phi_{s4}$ ,  $\phi_{s2}$  の順番に予測精度の向上に寄与していることがいえる。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 していることが分かる。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ことが分かる。これらにより大企業や資金調達に成功しているにもかかわらず、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ことができる。

最後に、どの基底関数がより Exit の予測精度を高めるかであるが、基底関数が  $x$ ,  $x^2$ ,  $\sqrt{x}$ ,  $\log x$  の順番に F-score の最高値が 9.0, 7.3, 5.3, 8.1 となっており、 $x$ ,  $\log x$ ,  $x^2$ ,  $\sqrt{x}$  の順番に Exit の予測精度が高いことが分かる。一方で、基底関数を  $x$ ,  $x^2$ ,  $\sqrt{x}$ ,  $\log x$  すべて組み合わせ Exit 予測すると、F-score が 15.2 となり最も高くなる。これにより基底関数を最適化することでさらに Exit の予測精度が高まっていくことが分かった。 $\phi_{s1}$  から  $\phi_{s4}$  の各素性に異なる基底関数を適応し最適解を求める余地もあり、その点は今後改善していきたい。

## 6. 考察

本章では、本手法の実務的な有用性と改善点について考察する。第1に、情報が少ないベンチャー企業を過去の資金調達に関する情報だけでなく、従業員の過去の転職履歴情報をデータとして用いることで、より高い精度でExitが予測できるという意味で有効であると考えられる。特に本手法はベンチャーキャピタルが投資先のポートフォリオを組むにあたって、ベンチャー企業が将来的にExitするかどうかを見抜けるといった点で有用であると考えられる。

第2に、本手法は、より創業間もない成長ステージにいる企業に対し有用であると考えられる。ある程度成長した企業の場合、資金の調達状況や従業員数、設立年度などの素性をExitしそうかどうかは判断材料にできるが、創業間もない企業の場合、将来的にExitするかどうかを判断する材料は少ないためである。

最後に、本手法において今後改善できる点に関し考察する。本手法は  $\phi_{financial}(C_i)$  にシリーズ A やシリーズ B<sup>\*4</sup> といった資金調達タイミングの素性を組み込むことで、企業の成長ステージごとの分析が可能となる。本稿ではデータサンプル数の制約から資金調達タイミングを素性として組み込まなかったが、今後 Web 上のデータがさらに増加すると考えられるため、本手法に企業の成長ステージを考慮することで本手法を発展させることが可能となると考えられる。

## 7. おわりに

本稿では、ベンチャー企業の転職履歴情報を活用することで、資金調達額や従業員数など、社内の資源のみを素性に用いた手法に比べて高精度にベンチャー企業のExitが予測できる手法を示した。具体的には機械学習におけるSVM2値分類器を構築し分類精度を最大化する素性選択を行うことでより高精度のExit予測を実現した。実際にIT業界のベンチャー企業を例にとり、ソーシャルキャピタルに関連する素性を用いより高精度にベンチャー企業のExitが予測できることを証明した。

機械学習の観点から述べると、リンクマイニングの領域のLinked Based Classificationをソーシャルキャピタルの1つである転職履歴ネットワーク上の素性を活用することで行った。リンクマイニングにおけるノード分類という問題をノードを企業とし、リンクをソーシャルキャピタルネットワークを対象に適応しベンチャー企業のExit予測に対して有用な知見が得られることが確認できた。

今後の研究課題としてはモデルの最適化およびデータ精度の向上があげられる。これらはこの種の問題に付き物の論点であるが、下記に具体的に述べる。

\*4 成長ステージに応じた資金調達の段階の呼称。初期段階から順にシリーズ A, B, C と呼ぶ。

本研究では、転職履歴情報に基づき素性生成をすることでベンチャー企業のExit予測する手法を構築した。素性の候補は4章にて選択し、予測精度により素性選択を行ったが、素性選択をより体系的に行えばより良いモデルを構築できる可能性は残っている。また素性ごとに適応する基底関数を変更することでモデルの予測精度も改善できる余地もある。より体系的なモデルの最適化は今後の研究課題である。

本研究では、オープンソース型のデータベースであるCrunchBaseを用いているが、CrunchBaseに記載されている情報がすべて正しいかどうかは現段階では判断できない。今後、Web上に公開されるデータが増加することでデータの信頼性を分析する素地が整ってくる。この点は今後の改善点の1つである。また、本研究では、IT業界のベンチャー企業を取り上げたが、原理的には他の業界にも応用が可能である。今後は、IT業界に限らず、他の業界のベンチャー企業のExitの予測に本手法を適用していきたいと考えている。

ベンチャー企業のExitを予測することで、より資金および人材の流入先が最適化されることは企業や個人にとって有用である。今後は、精度を向上させること、他の分野にも適用し有効性を検証することなどが課題である。

本研究では、転職履歴情報に基づき素性生成をすることでベンチャー企業のExit予測する手法を構築した。素性の候補は4章にて選択し、予測精度により素性選択を行ったが、素性選択をより体系的に行えばより良いモデルを構築できる可能性は残っている。また素性ごとに適応する基底関数を変更することでモデルの予測精度も改善できる余地もある。より体系的なモデルの最適化は今後の研究課題である。

## 参考文献

- [1] Davila, A., Fostera, G. and Gupta, M.: Venture capital financing and the growth of startup firms, *Journal of Business Venturing*, Vol.18, pp.689–708 (2003).
- [2] Chang, S.J.: Venture capital financing, strategic alliances, and the initial public offerings of Internet startups, *Journal of Business Venturing*, Vol.19, pp.721–741 (2004).
- [3] He, J. and Fallah, M.H.: Is inventor network structure a predictor of cluster evolution?, *Technological Forecasting and Social Change*, Vol.76, pp.91–106 (2009).
- [4] Blomstermo, A., Erikssona, K., Lindstrandb, A. and Sharma, D.D.: The perceived usefulness of network experiential knowledge in the internationalizing firm, *Journal of International Management*, Vol.10, pp.355–373 (2004).
- [5] Barney, J.B.: Strategic Factor Markets, Expectations, Luck and Business Strategy, *Management Science*, pp.1231–1241 (1986).
- [6] Mollera, K. and Anttilaa, M.: Marketing capability—A key success factor in small business?, *Journal of Marketing Management*, Vol.3, No.2, pp.185–203 (1987).



- [7] Sambasivan, M., Abdul, M. and Yusop, Y.: Impact of personal qualities and management skills of entrepreneurs on venture performance in Malaysia: Opportunity recognition skills as a mediating factor, *Technovation*, Vol.29, No.11, pp.798-805 (2009).
- [8] Zeng, S.X., Xie, X.M. and Tam, C.M.: Relationship between cooperation networks and innovation performance of SMEs, *Technovation*, Vol.30, pp.181-194 (2010).
- [9] Ardichvili, A.: Learning and knowledge sharing in virtual communities of practice: Motivators, barriers and enablers, *Advances in Developing Human Resources*, Vol.10, pp.541-554 (2008).
- [10] Gassel, K. and Pascha, W.: Milking partners or symbiotic know-how enhancement? International versus national alliances in Japan's biotech industry, *International Business Review*, Vol.9, pp.625-640 (2000).
- [11] Yli-Renko, H.: Social capital, knowledge, and the international growth of technology-based new firms, *International Business Review*, Vol.11, pp.279-304 (2002).
- [12] de Carvalho, A.G., Calomiris, C.W. and de Matos, J.A.: Venture capital as human resource management, *Journal of Economics and Business*, Vol.60, pp.223-255 (2008).
- [13] Werth, J.C. and Boert, P.: Co-investment networks of business angels and the performance of their start-up investments, *International Journal of Entrepreneurial Venturing*, Vol.5, No.3, pp.240-256 (2013).
- [14] Alexy, O.T., Block, J.H., Sandner, P. and Wal, A.L.J.T.: Social capital of venture capitalists and start-up funding, *Small Business Economics*, Vol.39, No.4, pp.835-885 (2012).
- [15] Backstorm, L., Huttenlocher, D., Lan, X. and Kleinberg, J.: Group formation in large social networks: Membership-Growth, and Evolution, *Proc. SIGKDD'06* (2006).
- [16] Das, S.: Filters, Wrappers and a Boosting-based Hybrid for Feature Selection, *Proc. 18th International Conference on Machine Learning*, pp.74-81 (2001).
- [17] 唐門 準, 松尾 豊, 石塚 満: リンクに基づく分類のためのネットワーク構造を用いた属性生成, 情報処理学会論文誌, No.49, No.6, pp.2212-2223 (2008).



上野山 勝也 (正会員)

2013年東京大学工学部技術経営戦略学博士課程修了。博士(工学)東京大学グローバル消費インテリジェンス寄付講座特任助教。機械学習, 人工知能, 特に高次 Web マイニングの企業活動への応用領域に興味がある。人工

知能学会会員。



大澤 昇平 (正会員)

2010年筑波大学第三学群情報学類卒業。2012年筑波大学大学院システム情報工学科コンピュータサイエンス専攻修了。同年より, 東京大学大学院工学系研究科技術経営戦略学専攻博士後期課程入学。専門は, ソーシャルメ

ディア分析, 人工知能。



松尾 豊 (正会員)

1997年東京大学工学部電子情報工学科卒業。2002年同大学院博士課程修了。博士(工学)。同年より, 産業技術総合研究所研究員。2005年10月よりスタンフォード大学客員研究員。2007年10月より, 東京大学大学院工学系

研究科総合研究機構/知の構造化センター/技術経営戦略学専攻准教授。人工知能学会編集委員長。専門は, Web マイニング, 人工知能, ビッグデータ分析。