

取引ネットワーク情報を用いた企業の成長分析と企業間の関係抽出

Corporate growth analysis and corporate relationship extraction using transaction network information

小林 真輝人^{*1} 浅谷 公威^{*2} 佐々木 一^{*3} 森 純一郎^{*2,3} 坂田 一郎^{*2,3}
 Kobayashi Makito Asatani Kimitaka Sasaki Hajime Mori Junichiro Sakata Ichiro

^{*1} 東京大学工学部システム創成学科 Program for Social Innovation, Faculty of Engineering, The University of Tokyo
^{*2} 東京大学工学系研究科技術経営戦略学専攻 Department of Technology Management for Innovation, Graduate School of Engineering, the University of Tokyo

^{*3} 東京大学政策ビジョン研究センター
 Policy Alternatives Research Institute, the University of Tokyo

Many factors that affect the growth of companies have been found, but existing studies are analyzing them from individual factors. Also, research focusing on inter-company relationships that are thought to have a major impact on growth and analyzing growth is scarce. In this research, we extracted the competitors from the trading network, and predicted the growth potential of the enterprise in a complex manner using that information and various resource information of the company. As a result, we confirmed the increase in prediction accuracy and identified important factors for growth for each industry.

1. 序論

本章では、本研究の背景と、位置づけ、目的と貢献について述べる。

1.1 背景

成熟期を迎えた我が国の経済社会において、各企業にとっての成長はより困難を伴う目標となっている。企業経営者に加え、投資家や政策立案者にとってどのような企業や業種に成長が見込めるかといった見通しに関わる情報は一層重要な材料となっている。

これまで企業の成長予測に用いられてきた指標は、売上高、利益率、資本金を含む企業内リソースに関する情報が一般的であった。近年では、企業同士のネットワーク的な位置付けが重要であるという視点に注目が集まってきている[Takayasu2007]。本研究では、環境条件として、取引によって形成されるコミュニティ、企業の競合関係に注目する。

取引ネットワークとは、企業をノードとして、取引関係にある企業間にエッジを生成することで作成される社会ネットワークのことを指す。取引関係を元に行っていることから、信頼関係に裏付けられたネットワークであると考えられ、ものやサービスだけでなく、情報や暗黙知を含めた知識共有の経路になっていると理解されている。

競合関係とは、同一もしくは代替可能な製品を扱い、市場を共有し顧客を奪い合う企業間の関係(潜在的な競合を含む)である。競合企業との競争の過程で、企業は製品を差別化し業務を効率化し新しい知識を求め、一方で競争力のない企業は市場を奪われること等により淘汰される。企業間の競合関係は、企業の成長性に大きく影響し、また競合企業群内の企業は製品特性、市場変動等の点で外部の企業と比較して特性が近いと考えられる。

1.2 本研究の位置づけ

企業の成長は複数の要因が複雑に影響しあい達成されるものであると考えられるが、既存の研究[Navaretti 2014]では個別

の因子から成長を分析しており、成長因子の複合的な分析は十分に行われていない。

競合関係は、同業の企業やサプライチェーンの集合であるクラスターや近隣企業とは異なり、企業の実環境条件に注目したものであり、このような企業間の関係性をネットワークの視点から抽出する研究は乏しい。さらに、競合関係が企業活動に大きな影響を与えることが分かっているが[Cheung 2012]、競合企業の情報と関連付けて企業の成長性を定量的に分析する研究は少ない。

そこで、本研究では取引データと企業情報から競合関係と企業のリソース情報を抽出し、それらを特徴量として機械学習を利用することで、競合関係の情報を含む複数の因子を複合的に分析し、企業の成長性を予測した。

1.3 目的と貢献

本研究の目的は、取引ネットワークから競合関係を抽出し、当該情報とその他のリソース情報とを共に機械学習を用いて複合的に分析し、企業の成長性を予測する手法を提案することである。

競合関係の情報は、取引先推薦、決算予測にも応用可能である。また、ケーススタディや特徴量の追加、より多くの競合関係の特定によって、さらに成長性予測の精度を高めることができれば、投資や経営の大きな助けとなる。本研究はこれを実現するための基礎となる手法を提案し、一つのケーススタディを提示した。

2. 提案手法

本章では、競合関係の定義と企業の成長予測に用いた手法について述べる。本研究では、競合関係を定義して、取引データから抽出した。続いて、抽出した競合関係を含む特徴量を用いて企業の成長性を予測した。

2.1 競合関係の定義

本研究独自の視点である、競合関係の定義について述べる。本研究では、以下の3つの条件を満たす企業群が競合関係にあると定義した。

条件 1. 業種が同じ

条件 2. 取引先の業種のコサイン類似度が $1/\sqrt{2}$ 以上

条件 3. 直接取引関係がない

条件 1 に関して、同一の製品または代替可能な製品・サービスを生産・提供し、市場を共有する企業のみが競合関係であると考えた。

条件 2 に関して、小分類が同じ企業でも大きく種類の異なるサプライチェーンに組み込まれている企業同士は、競合関係になりえないと考えた。コサイン類似度は、取引企業群の小分類の分布を表したベクトルを正規化し内積をとることで計算した。コサイン類似度の閾値 N については、3 つの値($N = \sqrt{3}/2, 1/\sqrt{2}, 1/2$) について競合企業数の分布を比較した(表 2.1)。 N の値を小さくしすぎると、小分類が同じ企業を集めた集合との差異が小さくなり、競合企業を特定する意味が失われる。一方で、 N の値を大きくしすぎると、十分な数の競合企業群が見つからず、検証が困難になる。また、競争企業数のばらつきが大きいことは極端な構造を持つ競合企業群の影響が大きくなるため、好ましくない。このような理由から、競合企業数が極端に大きくなる直前であり、競合企業数のばらつきが最も小さくなった $1/\sqrt{2}$ に決定した。

条件 3 は、取引関係があるということは、それぞれの企業がサプライチェーンの上流と下流に位置するということであり、そのような企業同士が競合関係にあることは考えにくいためである。

この定義に基づいて、取引ネットワークから競合関係を抽出することで、手動ではラベル付けが困難な 2010 年度の愛知に関する取引企業の 7.57%に当たる大量の企業の競合関係を見つけることができた。

表 2.1. コサイン類似度の閾値による競合関係の変化

コサイン類似度	競合保有企業割合	競合企業数平均	競合企業数標準偏差	最大競合企業数
$\sqrt{3}/2$	5.58%	17.52	33.50	149
$1/\sqrt{2}$	7.57%	14.04	29.54	149
$1/2$	20.8%	33.50	64.54	628

2.2 成長性予測

企業の成長性予測には機械学習の教師あり学習による分類を用いた。2010 年度、2011 年度のデータを用いて 2012 年度の成長企業を予測した。

特徴量は、企業情報やネットワーク特徴量、連携能力の指標である連携指数、クラスター情報、企業の取引構成、競合関係に関する情報などからなり、全 69 項目である。

モデルにはサポートベクターマシンとランダムフォレストを検討した結果、より高い精度が得られたランダムフォレストを用いた。

企業成長に大きな影響を与えると考えられる要因の一つに、時勢の業種全体の成長に対する影響がある。本研究の目的は、成長性の大きい企業の特徴を分析することであり、このような時勢の影響は予測のノイズとなる可能性がある。そのため、中小企業のデータを用いた実験では、特徴量とラベルの売上高に、その年度の同一業種の企業の売上高の平均との差を用いて予測を行った。

3. 使用データ

本章では、本研究で使用したデータの概要と、フィルタリング、成長の指標について述べる。

3.1 データ概要

本研究では、帝国データバンク株式会社によって提供されている 2010 年から 2014 年までの国内企業を対象とした取引データを用いた。ケーススタディとして、愛知県の企業に関する取引をサンプリングした。愛知県は日本の工業の中心である中京工業地帯の中核をなし、十分な企業数が存在する。また、トヨタを中心

とするサプライチェーンが存在することにより、密な取引ネットワーク構造が期待できる。愛知県に関する 322469 取引、159001 社のデータを使用して取引ネットワークを作成した。成長性の予測には 5 年間データに欠落がない企業 120116 社のみを使用した。

3.2 フィルタリング

データには多種多様な企業が含まれており、このようなデータを対象にしては予測精度に限界があると考えられる。そこで、フィルタリングを行い均一なデータセットを得た。企業の規模(従業員数)、競合関係が発見されたこと、業種(IT、自動車関連製造業、化学工業)、を条件とした(表 3.1)。

日本においては、企業数の観点では大企業の占める割合は極めて小さいが、一方、売上高や従業員数においては大きな割合を占める。大企業は中小企業とは大きく異なる性質を示すため、大企業と中小企業が入り混じったデータの予測は困難になると考えられる。そのため、従業員数をもとに中小企業 111064 社を選別し、そのうちの 10000 社の企業をランダムサンプリングしたデータセットを作成した。

また、企業の業種により産業構造や付加価値を生産する過程が異なるため、成長に強く影響を与える因子は異なると考えられる。そこで、特定の業種(IT、自動車関連製造業、化学工業)であることを条件に企業をフィルタリングし業種による差異を検証した。IT、自動車関連製造業、化学工業を選択したのは、参入障壁や産業構造が大きく異なるため、それぞれが異なる成長性に影響する因子を持つと考えたからである。

本研究で独自の定義に基づいて抽出した競合関係が成長性に与える影響を調べるために、競合関係が見つかった企業のみからなるデータセットを使用して、予測を行った。取引情報の欠落のない 120116 社データの 8.63%に当たる 10368 社の企業に関する競合関係を抽出した。

表 3.1. 各データセットの企業数

データセット	中小企業	競合保有企業	IT	自動車関連製造業	化学工業
企業数	111064	10368	1431	1474	1098

3.3 成長企業のラベル

本研究では売上高を用いて成長企業を定義した。従業員数や店舗数は経営者の判断で決定され、利益は企業の置かれた短期的な状況で変化すると考えたためである。

成長企業を翌年の売上高変化率がデータセット内で上位 $N\%$ である企業と定義して $N = 10, 20, 30, 40, 50$ それぞれについて検討した。どのデータセットも売上高変化率の平均は 1 程度であり、標準偏差は業種のデータセットで小さくなった(表 3.2)。

表 3.2. 各データセットの売上高変化率の平均と標準偏差

	中小企業	競合保有企業	IT	自動車関連製造業	化学工業
平均	1.085	1.059	1.062	1.041	1.016
標準偏差	1.925	1.093	0.351	0.233	0.225

4. 実験

実験は python によるプログラムの作成・実行によって行った。実験は競合関係の抽出・分析と企業の成長性予測からなる。

4.1 競合関係の抽出・分析

2010年度の愛知県に関する取引データを使用して、競合関係を定義に基づき抽出し以下の3つの分析を行った。

- 競合企業群の売上高、従業員数の分析
- 乗り換え取引の分析
- 競合企業群の売上高変化率の分析

分析では、比較対象として、小分類が同じ企業を見つかった競合関係の数だけランダムに選択した企業の集合であるダミー企業群を作成した。

本研究で使用したデータでは企業名が確認できないため、抽出した競合関係の妥当性を直接検討することができない。そこで、規模の大きく異なる企業同士はその競争力が大きく異なるため競合関係になりにくい(仮定1)、競合企業は同一または代替可能な製品を扱うため、競合企業間で取引の乗り換えが発生しやすい(仮定2)、という2つの仮定の下、抽出した競合関係の妥当性を検証した。

まず、仮定1を検証した。競合企業群がどのような性質を持つ集合なのか調べるため、競合企業群が見つかった企業の売上高と従業員数を企業全体と比較した。次に、企業と競合関係にある企業の規模の違いを調べるため、競合関係が見つかった企業間の2010年度の売上高比率と従業員数比率をダミー企業群と比較した。

次に、仮定2を検証するため、企業との取引を中止し小分類が同じ別の企業との取引を開始する事象を乗り換え取引と定義して、乗り換え取引の発生頻度を調べた。2010年を基準として、取引が停止した企業が停止した年に企業群内の企業との2010年と比較した新しい取引の有無によって乗り換え取引を判断し、取引を停止した年度に乗り換えが起こったとした。乗り換え取引を停止した年に開始した取引に限定しなかったのは、企業が現在の取引を中止する前に乗り換え先の企業との取引を開始する移行期間を設ける可能性があるためと判断したためである。

最後に、企業の成長予測を行うにあたり、企業の競合関係と企業の成長性の関係性を検証した。乗り換え取引が競合関係にある企業間で発生しやすいことにより、競合企業群全体の売上高は一定に保たれるという仮説をたて、これを検証するため、2010年度から2011年度の競合企業群全体の売上高変化率とダミー企業群全体の売上高変化率を比較した。

4.2 企業の成長性予測

取引ネットワークと企業情報から抽出した企業のリソース情報から企業の成長性を予測するモデルを作成し、以下の2つの分析を行った。

- 競合関係の成長性に与える影響分析
- 業種ごとの精度比較と成長因子分析

精度の評価にはF値を用いた。

競合関係の企業の成長性への影響を詳しく調べるため、競合保有企業のデータセットと中小企業のデータセットを用いて予測を行い、精度を比較した。さらに、特徴量の重要性を比較した。

業種ごとの成長性の違いを調べるため、IT、自動車関連製造業、化学工業のデータセットについて予測を行い、精度を比較した。さらに、業種ごとの成長に影響を与える因子を検証するために、特徴量の重要性を比較した。

5. 結果

本章では、競合関係の分析と企業の成長性予測の結果について述べる。

5.1 競合関係の分析結果

競合保有企業の売上高、従業員数の分析の結果は表5.1、表5.2のようになった。競合保有企業の方が全企業より、売上高、従業員数ともに平均と標準偏差が小さくなった。また、競合企業群の方がダミー企業群より、売上高比率、従業員数比率ともに平均と標準偏差が小さくなった。

表5.1. 競合保有企業と企業全体の売上高、従業員数比較

	売上高平均 (標準偏差)	従業員数平均 (標準偏差)
競合保有企業	2157 (19329)	38.35 (215.7)
全企業	9504 (6191076)	74.48 (851.6)

表5.2. 競合企業群とダミー企業群の売上高比率、従業員数比率比較

	売上高比率平均 (標準偏差)	従業員数比率平均 (標準偏差)
競合企業群	7.65 (76.65)	3.75 (21.12)
ダミー企業群	32.25 (1145)	10.60 (145.7)

企業から競合企業群の乗り換え取引の分析の結果、347件の乗り換え取引を発見した。ダミー企業群では10回の試行で平均222.3件、標準偏差は9.56の乗り換え取引が発見できた。4年間に中止された取引は13500件あったため、中止された取引の2.57%で、競合企業群の企業への乗り換えが確認できた。また、同一業種に対する乗り換え取引は6817件あったため、同一業種に対する乗り換え取引全体の5.09%が競合企業間で行われたことが分かった。

競合企業群全体の売上高変化率の分析の結果、売上高変化率は競合企業群においては、平均1.018、標準偏差0.255であったのに対し、ダミー企業群においては、平均1.04、標準偏差1.044であった。競合企業群はダミー企業群に比べて売上高変化率のばらつきが大幅に小さくなった。

5.2 企業の成長性予測結果

競合関係の成長性に与える影響分析の結果は表5.3のようになった。競合関係企業だけを対象とすることで、中小企業を対象とした場合よりも、全てのNにおいて予測精度が向上した。重要性が大きい特徴量を比較した結果は表5.4のようになった。競合保有企業の成長性には競合関係に関する特徴量の影響が大きかった。

表5.3. 競合保有企業と中小企業の予測精度比較

N	10	20	30	40	50
競合	0.87	0.75	0.66	0.60	0.60
中小	0.86	0.74	0.63	0.57	0.55

表5.4. 競合保有企業と中小企業の負うような特徴量の比較

	競合	中小
1	利益	売上高
2	利益変化率	操業年数
3	競合企業売上高合計×競合企業合計売上高変化率×競合企業合計売上高に対する自社売上高比率	利益
4	売上高	一人当たり売上高
5	競合企業平均操業年数	近接中心性

業種ごとの精度比較と成長因子分析の結果は表 5.5 のようになった。業種によるフィルタリングでは、中小企業と比較して予測の精度は向上しなかった。重要性が大きい特徴量を比較した結果は表 5.6 のようになり、業種ごとに重要な特徴量は異なった。

表 5.5. 中小企業と業種ごとの予測精度比較

N	10	20	30	40	50
IT	0.84	0.69	0.68	0.58	0.56
自動車	0.87	0.72	0.64	0.61	0.52
化学	0.84	0.74	0.61	0.58	0.56

表 5.6. 業種ごとの重要な特徴量の比較

	IT	自動車	化学
1	売上高変化率	従業員数変化率	一人当たり売上高
2	一人当たり売上高	クラスター合計売上高	売上高変化率
3	従業員数変化率	近接中心性変化率	従業員数
4	操業年数	クラスター合計従業員数	近接中心性
5	利益変化率	発注企業平均媒介中心性	クラスター平均従業員数

6. 考察

本章では、競合関係の妥当性と企業の成長性について考察する。

6.1 競合関係の妥当性

競合関係の見つかった企業の規模が、ランダムに同じ業種の企業を選出した場合に比べて、近しくなった。この結果は、極端に規模の違う企業間では、競争が起きにくく競合関係が発生しにくいことと適合している。極端に規模の大きい企業に競合関係が見つからなかった。一般に大企業は取引数が多く、その業種も多様になる傾向があり、取引企業の業種のコサイン類似度が大きくなる企業が見つかりにくくなるためだと考えられる。

企業が同一製品の取引について企業 A から企業 B に取引企業を変更するとき、企業 A と企業 B の間には競合関係があるといえる。乗り換え取引に関する分析では、競合企業群内での乗り換え取引数がダミー企業群内の平均値の 1.5 倍となった。これは競合関係の妥当性を裏付ける結果である。同一業種に対する乗り換え取引がどの程度、同一または代替可能な製品に関して行われた取引であるかは本データからは明らかにできないが、5.09%という割合は見つかった競合関係が企業が潜在的に保持する競合関係の極一部である可能性を示している。

以上の考察から、本研究で抽出した競合関係は、中小企業を対象としたものであり、市場を共有する企業の中では部分的である可能性はあるものの、競合関係に対する知見と矛盾しておらず、比較的妥当性があるといえる。本研究の抽出手法では、企業の主業のみを考慮しており、実際には複数の業種にまたがって活動する企業が数多く存在する。このような企業のそれぞれの市場の競合関係をどのように分析するかは今後の課題である。

競合企業群内の企業の売上高の合計はそれらの企業が共有する市場の規模を表しており、その値が一定である傾向があるということは、市場の規模の急激な変化が起こりにくいことを意味している。競合関係は、地理的情報や直接の取引関係とは異なる視点で、企業の環境条件から企業を紐づける試みであり、解釈のしやすさとして優れ、クラスターリング、決算予測、取引先推薦に応用可能である。

6.2 企業の成長性

競合関係を保有する企業を条件にフィルタリングすることで予測精度が向上したことには、競合関係に関する特徴量が、企業成長を説明したこと、競合関係を保有する企業をサンプリングすることで予測のノイズとなる企業データが減少したことが寄与していると考えられる。競合企業群全体の情報は市場の実態を表し、変化が緩やかであるため予測に有用であると思われる。また、競合関係はコントロール可能な指標であり、企業経営に反映させることができる。

一般的には、業種別に予測を行うことで同一業種内の特徴量の差異を詳細に判別しやすくなり精度が上昇すると考えられている。しかし、本研究では業種別にみると特徴量の重要性に大きな違いがあったにもかかわらず、予測精度が向上しなかった。これは、業種によって際立った特徴量があり、かつその業種の企業が中小企業全体に対して成長性が高い場合、業種の際立った特徴量を用いて分類することで精度が向上すること、業種別のデータセットの企業数が少なかったことが影響していると思われる。

また重要な特徴量から IT は新興企業の動きと優秀な人材の確保が、自動車関連製造業は早くからサプライチェーンの中心に組み込まれることが、化学工業では設備投資を行い、業務を効率化して競争力を高めることが重要であると考えられる。

地域差に対するケーススタディの積み重ねや、さらなる競合関係の特定、特徴量の追加などによって高い精度で予測が達成できれば効率的な投資やポイントを押さえた企業運営が可能になる。

7. 結論

本研究では、取引ネットワークから競合関係を抽出し、この情報と多様なリソース情報を、機械学習を用いて複合的に分析することで、企業の成長性を予測する手法を提案した。本研究で抽出した競合関係は、乗り換え取引が発生しやすく、全体での売上高が一定になる傾向があるなどの性質を示し、予測可能性、解釈のしやすさに優れ、クラスターリングをはじめとする様々な研究に応用可能である。ケーススタディ、特徴量の追加、競合関係の特定などにより、さらに精度が高い予測が可能になれば、投資や経営の大きな助けとなる。

謝辞

本研究にあたって、指導教員である坂田一郎先生をはじめ、多くの方々にご助力いただきました。特に、貴重なデータを提供してくださった株式会社帝国データバンクのご協力に、感謝いたします。

参考文献

- [Takayasu 2007] Takayasu, M., Sameshima, S., Ohnishi, T., Ikeda, Y., Takayasu, H., & Watanabe, K.: Massive Economics Data Analysis by Econophysics Methods-The case of companies' network structure., Annual Report of the Earth Simulator Center April, the Earth Simulator Center, 2007.
- [Navaretti 2014] Navaretti, G. B., Castellani, D., & Pieri, F.: Age and firm growth: evidence from three European countries., Small Business Economics, Springer, 2014.
- [Cheung 2012] 3) Cheung, M., & Zhuang, J.: Regulation games between government and competing companies: Oil spills and other disasters., Decision Analysis, Informs, 2012.