

取引により形成されるコミュニティが 新興企業の成長に与える影響の分析

布施 智章

概要：革新的な技術やビジネスモデルで急激な成長を遂げ、新たな市場・産業・雇用を創出する新興企業に対し、成長要因を特定し、成長を予測する研究は起業家、投資家双方から注目されている。本研究では、企業間の取引関係をネットワーク構造として捉え、取引が密な集団を企業コミュニティとして定義し、それらが新興企業の成長にどのような影響を与えているか、ランダムフォレストを用いた学習器で分析を行った。その結果、新興企業は所属するコミュニティ、特にコミュニティ内での相対的な規模の大きさが企業成長に影響を与えていることを示した。また、学習器を用いて一定の精度で新興企業の成長予測が可能であることを検証した。

1. 序論

1.1 研究背景

近年、革新的な技術やビジネスモデルを基に急激な成長を遂げているベンチャー企業を始めとした新興企業に注目が集まっている。現に、世界の時価総額ランキングTOP10に入っているGoogle社やFacebook社などは、創業からまだ20年も経っていないような新興企業である。新興企業は、安定志向の大手企業ならば避けるようなビジネス上のリスクを犯すことができ、それまでにはない新たな技術やビジネスモデルをもって革新をもたらし、新たな市場・産業、そして雇用を生み出す可能性を秘めている [1]。

こうした背景から、新興企業の成長要因分析 [2] に対する関心が集まってきており、「成長を遂げる新興企業と、そうでない新興企業にはどのような違いがあるか」という問いに対し、多くの研究がされてきた。従来の経営学的手法では、企業の事業領域、ビジネスモデル、資金、人材、特許などの企業属性といった企業の内側にある情報を説明変数としたものが多く、例えば企業の財務指標に着目したものが [3]。しかし、これらの質的情報を大量に収集することは困難である。

一方、企業同士のネットワークに着目した研究の1つに、企業間の取引関係をネットワーク構造として捉え、ネットワーク構造やコミュニティの分析を行うことで新たな産

業やイノベーションの創出を目指す研究もされている [4]。企業がどのような企業と取引関係にあるのかという情報は経営の意思決定が反映されているのである。こうした研究は、ネットワークに代表されるような企業の外側の環境情報も、イノベーション活動の成否や企業の成長に影響を与えることを示している。

そこで本研究では、企業間の取引ネットワークに着目し、ネットワークの構造やコミュニティが新興企業の成長にどのような影響を与えているかを分析した。具体的には、企業間取引ネットワークにおいて、取引が密な集団を企業コミュニティとして定義し、ネットワーク特徴量や所属するコミュニティの情報などを説明変数として新興企業の成長結果を学習し、従来手法で用いられている企業属性といった企業の内側にある質的情報を使用せずに、企業の基本情報とネットワークという環境情報を用いて、新興企業の成長予測が可能であるか検証を行った。

本研究の研究課題は、以下の2つを検証することである。

- 企業間の取引関係をネットワークとしてトポロジカルに捉えて分析を行い、新興企業の成長要因を特定する
- 従来手法で用いられている企業属性の情報を使用せずに、新興企業の成長予測を行う

結論としては、新興企業の成長は所属する企業コミュニティによって最も多くの影響を受け、特に所属するコミュニティ内での相対的な規模（売上高・従業員数）が大きいほど成長を遂げていることを推定した。また、それらの情

^{†1} 現在、東京大学工学系研究科技術経営戦略学専攻

報を元に新興企業の成長予測を行い、一定の精度での予測が可能であることが明らかになった。

2. 手法

2.1 データセット

本研究では、共同研究先でもある株式会社帝国データバンクにより提供された、企業間取引データを利用して分析を行った。このデータは、取引の受注・発注各々の企業をノード、取引関係の有無をエッジとすることで、企業間取引ネットワークを形成することができる。さらに発注→受注という金流を方向とする、有向ネットワークとも考えることができる。

2.2 コミュニティの検出手法

複雑ネットワークのコミュニティ検出方法の一つとして、Newman の提唱するモジュラリティ [5] という検出結果の指標を使用するものがある。モジュラリティは頂点集合 V の分割 $C = \{V_1, V_2, \dots, V_k\}$ を引数とする関数であり、コミュニティ内のネットワークが密であり、コミュニティ間のネットワークが疎であるほど高い値を出す指標である。ネットワークを隣接行列として表現した際の (i, j) 成分を A_{ij} 、総リンク数を m 、頂点 i を通るエッジの本数を k_i 、頂点 i が属するコミュニティを c_i 、 $\delta(c_i, c_j)$ を頂点 i, j が同じコミュニティに属するときは 1、それ以外の場合に 0 となるクロネッカーのデルタとしたとき、モジュラリティ $Q(C)$ は次式で表される。

$$Q(C) = \frac{1}{2m} \sum_{i \in V} \sum_{j \in V} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j) \quad (1)$$

コミュニティ検出は多くの場合モジュラリティ最大化と同等である。今回はモジュラリティを用いたコミュニティ検出手法の一つにであり、Blondel らによって提案された Louvain 法 [6] を用いた。Louvain 法は以下のアルゴリズムにより計算され、高速かつ高いモジュラリティを導出できる。

2.3 企業の成長予測手法の概略

本研究では、新興企業を設立から 10 年以内の企業（つまり、設立年が 2006 年 1 月以降と記載のあるもの）と定義する。次に、提案手法の概略を図 1 に示す。まず、企業間取引データを用いて、目的変数となる正例・負例のデータセットを作成し、それと並行して様々な説明変数を作成する。最後に、学習器による予測を行うことで、新興企業の成長予測を行う。

2.4 学習器の説明変数

学習器に用いた説明変数は、以下の通り企業情報、ネットワーク特徴量、コミュニティ指標の 3 つに分類できる。

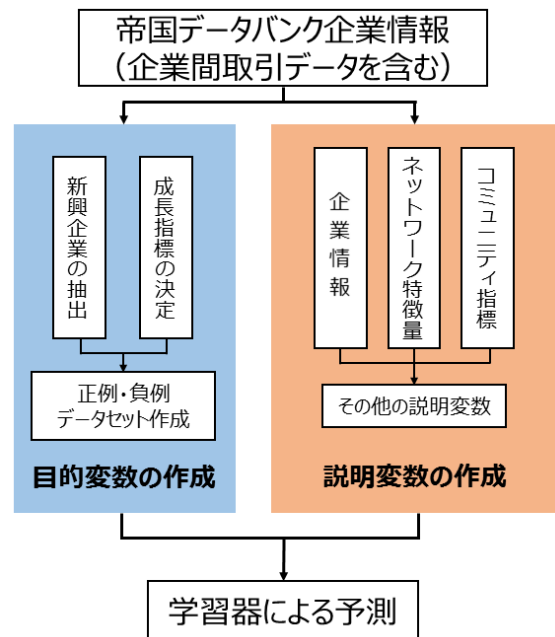


図 1 提案手法の概略

なお以下の説明変数は、とくに断りがない限り 2014 年度の単年度のデータを用いている。

2.4.1 企業情報

従業員数 2014 年度での企業の従業員数

業種 A~M 産業分類区分の大分類が、A であれば 1、それ以外を 0 ととる変数

2.4.2 ネットワーク特徴量

入次数 当該企業へ発注を行っている企業数。2014、2012 年度の 2 年間の指標を利用

出次数 当該企業から受注を受けている企業数。2014、2012 年度の 2 年間の指標を利用

近接中心性 他の頂点との距離の平均値の逆数

媒介中心性 ある頂点が他の 2 点間の最短路上にあるかどうかを示す指標

固有ベクトル中心性 ネットワークの隣接行列の最大固有ベクトル

ページランク 有向ネットワーク上のランダム・ウォークを利用した指標。2014、2012 年度の 2 年間の指標を利用

z 値 コミュニティ内でのハブ度を表す指標

P 値 コミュニティ間のコネクタ度を表す指標

次数の変化 2014 年度の入(出)次数 - 2012 年度の入(出)次数

ページランク変化率 2014 年度のページランク ÷ 2012 年度のページランク

2.4.3 コミュニティ指標

企業数 コミュニティ内の企業数

総売上高 コミュニティ内の企業の売上高の総和

平均売上高 コミュニティ内の企業の売上高の平均値

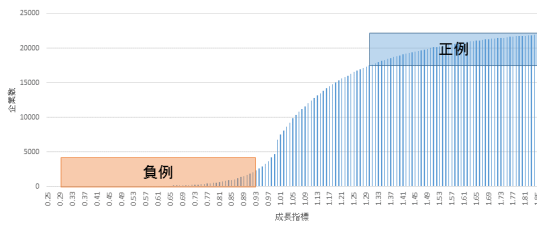


図 2 成長指標と企業数の累積度数分布

総従業員数 コミュニティ内の企業の従業員数の総和
平均従業員数 コミュニティ内の企業の従業員数の平均値
平均設立年 コミュニティ内の企業の設立年の平均値
新興企業比率 コミュニティ内の企業のうち新興企業の占める比率
成長企業比率 コミュニティ内の企業のうち成長企業（売上高変化率が1以上）の占める比率
相対売上高 企業の売上高 ÷ 平均売上高
相対従業員数 企業の従業員数 ÷ 平均従業員数

2.5 学習器の目的変数

成長指標をもとに、2.3 で定義した新興企業群を正例・負例に分けたデータセットを作成する。つまり、正例であれば1、負例であれば0を返すような2値データを目的変数として利用する。本研究では1年間での売上高成長率を複数年（2～4年）で平均したものを新興企業の成長指標として定義する。成長指標を横軸に、企業数の累積度数分布を表したものを図2に示す。

今回の成長予測では、図の通りに正例・負例を設定した。正例データは年平均30%以上の成長率を遂げる新興企業であり、企業数は6954社（新興企業群のうち、成長指標が上位約28%）となった。一方、負例データは年平均-7%の成長率となる新興企業であり、企業数は2020社（新興企業群のうち、成長指標が下位約8.3%）となった。

2.6 ランダムフォレスト

本研究では、成長予測を行う学習器として、ランダムフォレストを用いた。ランダムフォレストは、Breimanによって提案された識別・回帰を行う機械学習アルゴリズムの一つであり、複数の決定木を弱識別機として用い、それらの結果を統合する集団学習である [7]。

2.6.1 学習器の性能評価

学習器の性能評価方法として、本研究ではROC曲線のAUCを用いた。ROC(Receiver Operating Characteristic)曲線とは、受信者動作特性曲線とも呼ばれ、陽性率を縦軸、偽陽性率を横軸にとり、閾値を変化させながら2つの値が変化する様子をプロットしたものである。ROC曲線の下側の面積を求めるAUCという指標はアルゴリズムの識別性能評価としてよく利用されている [8]。完全にランダム

表 1 主要コミュニティの指標

コミュニティ番号	多い業種とその割合	新興企業の比率
0	製造業：51.7%	0.045
1	製造業：42.4%	0.052
2	サービス業：62.1%	0.152
7	建設業：49.7%	0.082
8	卸・小売業・飲食店：63.3%	0.080
10	建設業：57.8%	0.061
13	卸・小売業・飲食店：54.9%	0.078
20	サービス業：59.1%	0.188

な識別であればAUCは0.5となり、1に近づくほど識別性能は高いと評価できる。

2.6.2 説明変数の重要度

ランダムフォレストでは説明変数の重要度を計算することができるが、中でも今回はランダムフォレストで回帰を行った際によく利用されるIncMSEを用いた。IncMSEは、ある説明変数をランダムにシャッフルした場合にどれほどエラーが増えるかという指標である。つまり、重要である説明変数であるほど、値をシャッフルした場合にエラーが増大するという考えに基づいている。一般的にはエラーには平均二乗誤差 (Mean Square Error)

$$MSE(c) \equiv \frac{1}{n} \sum_{i=1}^n (x_i - c)^2 \quad (2)$$

が使われる。

3. 分析結果

3.1 コミュニティの検出結果

Louvain法により、507611社ある企業を2773個のコミュニティに分割することができた。中でもデータセットに含まれる新興企業数が多い上位8つのコミュニティの指標を表1にまとめた。表から分かる通り、各々のコミュニティは特定の産業分類区分が多いコミュニティとなっていることがわかる。

3.2 学習器の性能評価

学習器の性能評価として、ROC曲線を図3に示す。ROC曲線のAUCは0.7393となっており、精度は非常に高いとまでは言えないが、新興企業の成長をある程度予測可能であると評価できる。また、学習器の訓練は複数回行っても正答率とAUCの値は安定しており、実験の再現性も確保できている。

3.3 説明変数の重要度

説明変数の重要度を示すIncMSEの計算結果を図4に示す。説明変数を企業情報・ネットワーク特徴量・コミュニティ指標の3つの分類に分けて見ると、企業情報が最も重要度は低い説明変数群であるとわかる。次に、説明変数の重要度を細かく見ていくと、相対売上高や相対従業員数と

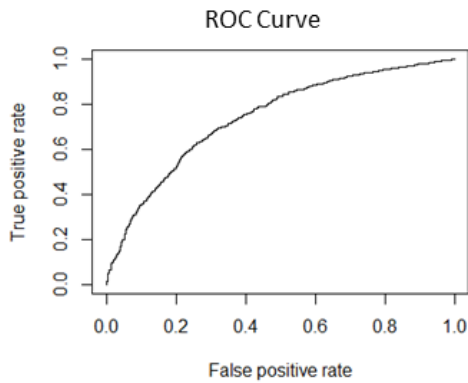


図 3 学習器の性能評価：ROC 曲線

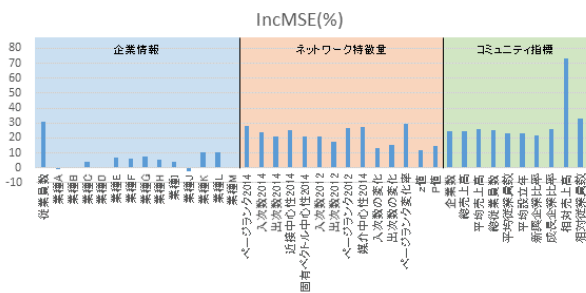


図 4 説明変数の重要度：IncMSE

いった指標が最も重要度が高いことがわかる。また、企業情報の中では従業員数、ネットワーク特徴量の中ではページランクや各種の中心性といった指標の重要度が高い。一方で、企業情報の中でも産業分類情報や、ネットワーク特徴量のうち入次数・出次数の変化、z 値 P 値といった指標はそこまで重要度が高くないため、企業の成長にさほど影響を与えないと考えられる。また、この値だけでは、説明変数の値が正に影響を与えているか、負に影響を与えているかの判断できない。そのため、重要な説明変数を正例と負例ごとに平均値を算出すると、おおよそこれらの説明変数が正に影響を与えており、値が大きいほど新興企業が成長していることがわかった。

以上の議論では、説明変数の中でも企業情報・ネットワーク特徴量・コミュニティ指標のどの説明変数群が最も重要度が高いかの判断はしかねていた。そこで、各々の分類の説明変数を学習に用いなかった場合、学習器の性能がどのように変化することで、説明変数群の重要度を議論する。各々の分類の説明変数を学習に用いない場合の AUC を算出し、比較を行ったものを表 2 に示す。表の通り、コミュニティ指標を用いなかった場合、AUC が 0.7393 から 0.6864 に低下し、最も学習器の性能が低下することがわかった。

3.4 コミュニティによる説明変数の重要度の差異

3.1 では、分割されたコミュニティの産業分類区分に偏りがあることを説明した。この時、コミュニティ毎に説明変数の重要度に差異はあるだろうか。そこで表 1 の 8 つの

表 2 説明変数の選択による AUC の比較

学習に用いた説明変数	ROC 曲線の AUC
全て	0.7393
企業情報以外の 2 分類	0.7307
ネットワーク特徴量以外の 2 分類	0.7287
コミュニティ指標以外の 2 分類	0.6864

コミュニティについて、正例と負例の説明変数の値に統計的な有意差があるといえるかどうか、検定を行う。今回は、標本の正規性も等分散性も仮定しないノンパラメトリックな検定として、Brunner-Munzel 検定 [9] を用いて正例・負例の 2 標本のデータにおいて p 値を計算した。

計算の結果、従業員数・相対売上高・相対従業員数はどのコミュニティも正例と負例データで有意差があることがわかった。また、サービス業が多くを占め新興企業の比率も高い 2 つのコミュニティ (2,20) については、相対売上高や相対従業員数以外にも、媒介中心性・近接中心性・ページランク (2014 年度) などの説明変数も正例・負例データで有意差があることが確認できた。

4. 考察

4.1 企業間取引ネットワークによる新興企業の成長予測

3.2 で議論した通り、企業間取引ネットワークにおけるネットワーク特徴量、及び所属するコミュニティの情報を説明変数とすることで、新興企業が成長を遂げるかどうかを一定の精度で予測することができた。つまり、従来手法で用いられる企業属性情報を説明変数に用いなくても、企業間の取引をネットワークとして捉え、トポロジカルに分析を行うことで、新興企業の成長を予測できることを示した。ネットワーク上で重要なポジショニングをとり、適切な企業コミュニティに所属している新興企業ほど、取引を通じて取得した情報を元に経営の意思決定を行い、自身の成長につなげているということが言える。しかし、あくまで予測できることは新興企業が短期的に成長を遂げるのが可能かどうかの情報だけであった。企業の成長の度合いやスピードの予測を行う場合は、企業間取引ネットワークの構造を時系列的に分析することが必要であると考えられる。

4.2 企業コミュニティが新興企業の成長に与える影響

企業間取引ネットワークの中でも、所属するコミュニティがどのようなものであるかということが新興企業の成長に大きな影響を与えていることがわかった。その中でも特に、コミュニティ内での相対的な規模 (売上高・従業員数) が大きい新興企業ほど成長を遂げている。つまり、コミュニティ内である程度の規模まで企業を成長させ、創業からのスタートダッシュを成功させることが、今後の成長を最も左右すると思われる。また、相対従業員数も同様

に、創業時期にある程度の規模まで事業を拡大し、顧客を開拓することができるほどの従業員を抱えていることが、今後の企業成長につながると考えられる。

コミュニティという観点では、 z 値や P 値は新興企業の成長にさほど影響を与えない変数であることがわかっている。つまり、同一コミュニティ内で多くの企業と取引関係を結ぶことや、できるだけ多くのコミュニティにまたがるように取引関係を結ぶことは、企業の成長という点においてさほど重要ではないと考えられる。

4.3 取引ネットワーク上でのポジションが与える影響

入次数や出次数、及びその変化は新興企業の成長に影響を与えなかった。つまり、単純な取引数の大小や増減では新興企業は成長しない。その他のネットワーク特徴量やその変化率が重要度の上位にきていることからわかるように、取引する企業が変化することで取引ネットワーク上でのポジションがどうなるかということが新興企業の成長に大きな影響を与えると考えられる。また、ページランクや中心性といった指標を大きくするためには、ネットワーク上でどのようなポジションの企業と取引関係を持つということがかなり重要になってくる。従って、ネットワーク上で重要なポジションをとる企業と取引を結ぶほうが、自身のネットワーク上での重要度も上がりやすくなり、企業を成長させることができると考えられる。

4.4 所属コミュニティによる差異

3.4で議論した通り、どのコミュニティにおいても相対売上高・相対従業員数が正例・負例データで統計的に有意な差があることが確認できた。さらに、サービス業が多くを占め、新興企業の比率が高い2つのコミュニティにおいては、媒介中心性・近接中心性・ページランク(2014年度)といったネットワーク特徴量の指標も統計的な有意差が確認できた。つまり、特にサービス業を主事業としている新興企業では、取引を通して取得できる情報の価値が高いのではないかと考えられる。従って、サービス業を主事業としている新興企業と多くの取引関係にある新興企業は、ネットワーク上で重要なポジショニングをとる様々な企業と取引関係を構築し、自身のネットワーク特徴量を大きくすることが、企業成長へつなげていくために重要であると考えられる。

一方で、サービス業以外が多くを占めるコミュニティでは、ネットワーク特徴量の重要度はそこまで高くない。それには、以下のような理由が考えられるだろう。

- 製造業・建設業・卸小売業飲食店といった分類の企業の多くは、企業と取引関係を持つことでサプライチェーンを築いており、取引ネットワーク上でのポジションよりも、取引の内容そのものが企業成長に影響を及ぼしている

- 製造業・建設業のような新興企業の比率が低い業種は、本来の意味での新興企業ではなく、既存企業の子会社や関連会社のような設立の新しい企業が、分析対象となる新興企業の定義に当てはまっているという可能性も考えられる。そうした企業は取引先の企業よりも、親会社の業績や業務内容が成長要因につながりやすいため、ネットワーク特徴量といった指標の重要度は低くなりやすいと考えられる

5. 結論

本研究では、企業間の取引ネットワークに着目し、ネットワークの構造やコミュニティが新興企業の成長にどのような影響を与えているかを分析した。具体的には、企業間取引ネットワークを用いて取引が密な集団を企業コミュニティとして定義し、ネットワーク特徴量や所属するコミュニティの情報などを説明変数とし、新興企業の成長結果を目的変数とすることで学習を行った。結果として、殆どの新興企業では所属するコミュニティが成長に大きな影響を与えており、さらに一部のコミュニティに所属する新興企業は取引ネットワーク上で重要なポジションにあるほど成長しているということがわかった。また、従来手法で用いられている企業属性情報を使用せずに学習を行ったとしても、一定の精度で新興企業の成長予測を行うことができた。これを以て、本研究の課題の検証を行うことができた。特に、コミュニティ内で相対的に企業の規模が大きく、創業から一定の規模まで拡大できた新興企業ほど、成長を遂げているということがわかった。

さらに本研究は、新興企業の産業分類や所属するコミュニティによって、影響を与える説明変数が異なることも示すことができた。これは、起業家やベンチャーキャピタルにとって、企業にとって有用な取引関係の構築の仕方を示唆しており、今後の企業成長のために経営の意思決定要因の1つとすることができると考えられる。

一方で、本研究の課題にも言及しておこう。本研究で用いた企業間取引ネットワークは全ての取引を同質なものとみなして構造の分析を行っているが、取引は中身によって企業に及ぼす影響がかなり異なってくるので、取引の内容を定量化し、ネットワークのエッジに重み付けをすることが望ましい。しかし実際のビジネスにおいては、取引の持つ価値やその内容は相手企業によって大きく異なっており、取引の詳細は当事者同士でしか知り得ない秘匿性の高い情報を多く含んでいる。

以上を踏まえ、今後の研究の方向性として、次のようなものが考えられる。

1つは、取引を行う企業間の関係や規模、事業内容といった属性情報を正しく捉え、取引の価値を推定することである。ビジネス上で秘匿性の高い情報を正しく推定することができれば、非常に経済的価値の高い研究となるだろう。

そのためには、今回の分析手法に加え、取引ネットワーク上の距離が近い企業の情報に加え、既存手法で用いられている企業属性情報を含めた様々な情報を統合的に組み合わせる必要がある。さらに、既存の企業に対してヒアリングを重ね、多くの実例を集積することも重要であると考えられる。

次に考えられるのは、取引ネットワークの時系列的な変化の分析である。本研究はネットワークの構造と成長との時系列的な因果関係を検証するに至っていない。また、目的変数を企業の成長速度や度合いとすることにより、より大きな知見を得ることができる。さらに、ベンチャー企業の時系列的な分析に用いられているような、投資を含めた財務情報などの定量的な指標や、ネットニュースなどの話題性といった定性的な指標を加えることで、より詳細な情報をもとに、企業の経営判断を助力することにつながるだろう。

参考文献

- [1] S.Wennekers, R. Thurik. 1999. Linking entrepreneurship and economic growth. *Small business economics*. 13(1), 27-56.
- [2] S.J.Chang. 2013. Venture capital financing, strategic alliances, and the initial public offerings of Internet startups. *Journal of Business Venturing*. 19(5), 721-741.
- [3] Bertoni F., Colombo M. G., Grilli, L. 2013. Venture capital investor type and the growth mode of new technology-based firms. *Small Business Economics*. 40(3), 527-552.
- [4] Yoshiyuki Takeda, Yuya Kajikawa, Ichiro Sakata, Katsumori Matsushima. 2008. An analysis of geographical agglomeration and modularized industrial networks in a regional cluster: A case study at Yamagata prefecture in Japan. *Science Direct, Technovation*. 28(8), 531-539.
- [5] M.E.J. Newman, M.Girvan. 2004. Finding and evaluating community structure in networks. *Physical Review E*. 69(026113).
- [6] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008(October 2008).
- [7] Leo Breiman. 2001. Random Forests. *Machine Learning*. 45(1), 5-32.
- [8] Hanley, J. A. McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 143(1), 29-36.
- [9] Brunner E., Munzel U. 2000. The nonparametric Behrens-Fisher problem: asymptotic theory and a small-sample approximation. *Biometrical journal*. 42(1), 17-25.