

令和 5 年 5 月 22 日現在

機関番号：11301

研究種目：挑戦的研究（萌芽）

研究期間：2020～2022

課題番号：20K20799

研究課題名（和文）自然言語処理を用いたアントレプレナーシップの測定と教育プログラム改善への展開

研究課題名（英文）Measurement of entrepreneurship using natural language processing and application to the improvement of education program

研究代表者

伊藤 彰則（Ito, Akinori）

東北大学・工学研究科・教授

研究者番号：70232428

交付決定額（研究期間全体）：（直接経費） 4,600,000円

研究成果の概要（和文）：アントレプレナー教育の学習者が書いたテキストから個人的アントレプレナーシップ志向（IE0）を推定することが本研究の目的であった。期間中、関連する実際の授業の受講者からテキストを集めたが、それ自体で識別モデルを作成するほどの量を集めることができなかった。そこで、パーソナリティを学習するための既存のデータを用いて、上記のデータに関するパーソナリティ分析を行い、アントレプレナーシップと無関係なテキストとの比較を行うことでIE0自動推定のための手がかりを探索した。その結果、MBTIパーソナリティのうち、J-P（分析的～感性的）の軸に違いが表れることがわかった。

研究成果の学術的意義や社会的意義

起業家教育はさまざまな教育機関で行われているが、その成果をどのように計測するかは十分に研究されていない。特に重要なのは起業をする意欲をいかにして測るかである。これを表すのが個人的アントレプレナーシップ志向（IE0）であり、アンケート等でこれを測る方法は確立されている。しかし、教育成果の計測という意味ではアンケートを使うことは適切でなく、教育活動の中で得られた成果を使ってIE0を測ることが望ましい。そこで今回は、教育内容に関する感想文をもとに、機械的にIE0を測ることを試みた。結果として直接IE0を測ることはできなかったが、パーソナリティの中にIE0と関連する軸があることが明らかになった。

研究成果の概要（英文）：The purpose of this project is to develop a model that estimates the individual entrepreneurial orientation (IE0) from the text written by students of an entrepreneurship-related class. We gathered text data from students of the related classes, but the amount of the text is insufficient to train a statistical model. Therefore, we exploited data for training personality trait and developed a personality estimation model. Then we analyzed the entrepreneurship-related text and the irrelevant text using the developed model. The analysis result suggested that the entrepreneurship is related to the J-P dimension of the MBTI personality trait, which represents judgment-perception.

研究分野：音声言語処理

キーワード：アントレプレナー教育 アンケート 個人的アントレプレナー志向 自然言語処理

## 1. 研究開始当初の背景

日本におけるアントレプレナー教育(起業家教育)は初等中等教育から大学までの幅広い教育機関で実践が始まっている。教育内容的には、座学と起業を疑似体験する演習、ビジネスプランコンテストなどを組み合わせることが多い[中島 2019]。アントレプレナー教育は、知識や技能の習得そのものだけが目的ではなく、実際に起業する意欲を涵養することも重要な目的であるため、その評価は通常の科目のようにペーパーテストでは測れず、評価自体に困難を伴う[樋口 2017]。

近年の教育分野では、先験的な理論に基づく教授法だけでなく、エビデンスに基づく教育が重要視されてきており、それはアントレプレナー教育も例外ではない。アイデアの評価やアントレプレナーシップの評価は点数をつけるペーパーテストでは測れないため、アンケートによる自己評価[青木 2014]や、ルーブリックを用いて学生同士が評価し合うピアチェック[松前 2016]が行われている。これらの評価方法は一定の信頼性があるとされているが、これらの評価に加え、授業の成果物を用いた客観的な教育効果測定が望まれていた。

一方、従来のアントレプレナー教育の研究における定量的エビデンスとしてはアンケートが主流であり(例えば[Ruskobaara&Pihkala2015])、その結果を因子分析や重回帰分析によって処理することが多い。これは教育プログラムや学習者の心理的傾向を分析するための手段としては妥当であるが、実際の教育現場で教育プログラムの一環として行う方法としては限界がある。

より効果的な教育を行うためには適切な評価が不可欠であり、そのためにはアントレプレナー教育の効果、特にアントレプレナーシップを客観的かつ定量的に測定する手法の開発が不可欠である。

アントレプレナーシップを定量化する試みとして、起業家志向(Entrepreneurial Orientation, EO)がある。教育の観点からは個人的起業家志向(Individual EO, IE0)が重要である[Bolton&Lane2018]。Bolton と Lane は、個人的起業家志向を定量化する方法として、アンケートによって起業家志向を「リスク志向」「イノベーション志向」「積極性」の3つの尺度で評価する方法を提案した。一方、学生に対する教育の成果を計測するという目的からは、アンケートではなく、調査レポートなどの課題の結果から IE0 が推定できることが望ましい。

## 2. 研究の目的

本研究では、アントレプレナー教育プログラムを受講した大学生が書いた文章から、本人の IE0 を自動推定する技術を開発する。また、この技術により定量化した結果を教育プログラムにフィードバックし、教育プログラムを改善していくプロセスを確立する。

そのため、本学で実施するアントレプレナー教育プログラムを主なフィールドとしてデータ収集を行う。教育プログラム内での受講生からのアウトプット(レポート等のテキスト)を収集すると同時に、同プログラム内において何度かアンケートを実施し、各教育段階における受講生の IE0 を測定する。

次に、収集されたデータを利用し、テキストから IE0 を推定するモデルを開発する。テキストから筆者の感性情報を推定する手法はセンチメント分析(Text Sentiment Analysis)と呼ばれ、特に Twitter などの SNS からの評判推定や感情推定などは近年急速に発展している[Yadollahi2017]。本研究ではこれを応用して、テキストから IE0 を推定する。近年のセンチメント分析では、数千万~数十億文のデータから深層学習を使って推定を行うのがトレンドであ

るが、本研究でのデータ量はこれよりも4桁以上小さいので、Bag of Word や決定木など、少量の学習データでも安定したモデルが得られる技術を中心に検討を行う。

### 3. 研究の方法

プロジェクトの最終目標は、文章から学生の起業家精神（典型的にはIE0）を推定することである。この目標を達成するために、まず東北大学のアントレプレナー関連講義の参加者を対象にボランティアを募り、講義について書いた感想文を収集する。対象とした講義は次の4つである。

- アントレプレナー入門（全学）
- 医療機器ビジネス学（医工学研究科）
- 情報技術経営論（情報科学研究科）
- 経営学入門（経済学部）

これらの参加者から十分な数のデータが集まれば、センチメント分析のためのモデルを応用してIE0を推定する。一方、データが十分集まらなかった場合には、アントレプレナーシップに関連した公共データを使ってモデルを作り、アントレプレナーシップ推定への応用を検討する。

### 4. 研究成果

上記の講義受講者からボランティアを募ってデータを収集したところ、14名から応募があった。これらのデータはモデルを推定するには少なすぎるため、上記の計画に従い、公共データからモデルを作成してアントレプレナーシップ推定に応用することにした。このときのデータおよびモデルとして、性格（パーソナリティ）推定のものを応用することにした。

文章から性格を推定するモデルはいくつか存在する。本モデルは、MehtaらによるBERTベースの性格推定モデルを参考に開発されたものである。性格特性として、4次元（E-I、N-S、F-T、J-P）を持つMBTIを使用した。まず、テキストがBERTモデルに入力され、次にCLSの出力が1つの隠れ層を持つ全結合層に供給される。出力は4つの2値分類結果であり、損失は次元ごとに計算され、学習によって総損失を最小化する。

分析対象データは、2022年4月から7月にかけて東北大学で開催された起業関連授業に参加した14名の実験参加者によるテキストである。「授業を受けた動機、授業を終えた後の感想などを記述する」文章と、「授業を終えた後の自分について、起業するために持っておくべきものなど気づいたことを記述する」文章の2つを記述してもらった。テキストは日本語で記述した。開発したモデルは英語しか受け付けられないため、DeepLを用いて英語に翻訳した。機械翻訳による分類精度の低下などの影響が考えられるが、その影響は（あるとしても）比較的小さいことを確認している。

比較のために、他に2つのテキストを収集した。ひとつは、Webから収集した大学の授業の感想で、起業とは関係ないものである。起業に関連するテキストと同じように分割した。もう1つは、Twitterから収集した、大学の授業とは無関係のツイートである。

性格分類モデルを学習するためのテキストコーパスは、KaggleのMyers-Briggs Personality Type Datasetを使用した。このデータセットは、Personality Caféというウェブサイトから1人あたり50件の投稿があり、8675個のMBTIラベルがある。

これらのデータを学習に使用する際、分析用のデータが機械翻訳されたものであることを考慮して、データを増強した。学習データは英語だったので、DeepLを使ってそれを日本語に翻訳し、さらにAzure Cognitive Serviceを使って英語に翻訳し直した。英語から日本語への翻訳と日本語から英語への翻訳で異なる翻訳器を使用することで、同じ文に翻訳し直されることを防いだ。再翻訳されたデータを元のデータとマージすることで、学習データの量を2倍に増やし

た。また、MBTI の 16 の特性が同じサンプル数になるように、少ないカテゴリの文章をアップサンプリングした。モデルの学習時には、データの 80%と 20%をそれぞれ学習データと検証データとして使用した。

まず、性格推定モデルの学習を行った。その結果、N-S 識別が最も精度が高く、他の次元も 95%以上の精度を示している。200 エポック後の最終的な精度は平均で 96.0%であった。

次に、このモデルを用いて、収集したデータの性格特性を測定した。その結果、データソースによらず、ほとんどのデータが ISF(内向性、感覚的、感情的)に分類されることがわかる。また、Kaggle のデータでは 4 つの特徴の精度が 95%以上であったことから、3 つの特徴の判別はデータソース (Personality Café) 特有の表現に依存することが示唆された。

一方、J-P(判断-知覚)次元では 3 つのデータソースの値が変化しているようであった。その有意性を確認するために、性格値とデータソースを因子とする二元配置 ANOVA を実施した。その結果、特性値の効果 ( $F(3,1094)=1176.97, p<2 \times 10^{-16}$ )、データソースの効果 ( $F(4,1094)=4.28, p=0.014$ ) とともに 5%水準で有意であった。

この解析結果は、テキストデータのほとんどが類似の形質 (ISFx) に分類されたため、疑問が残る。ここで、J-P の特性の違いについて考えてみよう。分析結果が書き手の特性の違いを正しく反映しているとすれば、起業関連のテキストは他のデータと比較して「Perception」特性である傾向がある。Kerr のレビューでは、起業家の特性は、ビッグファイブ特性の「開放性」と「良心性」の両方が高い傾向にあるとされている。一方、我々の分析は MBTI の特性に基づいている。Furnham によれば、J-P 次元は開放性と負の相関があり、良心的と正の相関がある。この結果は、起業家関連の文章が「知覚的」である(「判断的」でない)傾向があることを示唆しているので、起業家は開放性が高い傾向があるが、良心的であることとは一致しないという Kerr の分析に一致する。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 深澤佑樹, 伊藤彰則, 能勢隆
2. 発表標題 言語からの性格推定におけるモデルとデータ量の影響に関する研究
3. 学会等名 音響工学研究会
4. 発表年 2022年

1. 発表者名 Akinori Ito, Kotaro Takeda, Shuiti Ishida
2. 発表標題 Personality Analysis of Entrepreneurial Text for Entrepreneurship Education
3. 学会等名 International Conference on Natural Language Processing (国際学会)
4. 発表年 2023年

1. 発表者名 Yuki Fukazawa, Akinori Ito, Takashi Nose
2. 発表標題 Effect of Data Size and Machine Translation on the Accuracy of Automatic Personality Classification
3. 学会等名 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	石田 修一  (Ishida Shuichi)  (00326539)	東北大学・工学研究科・教授    (11301)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	武田 浩太郎  (Takeda Kotaro)  (80727214)	東北大学・工学研究科・講師     (11301)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関