

# 社内研修の評価及び人材育成の効率化 を目的とした社内SNSの分析

芳賀 あかり<sup>1,a)</sup> 富田 陽介<sup>2,b)</sup> 山下 晃弘<sup>1,c)</sup> 松林 勝志<sup>1,d)</sup>

**概要：**多くの企業では様々な社内研修が行われているものの、十分な効果検証を行っている企業は少ない。一方でSNSを社内コミュニケーションに活用する企業も増えており、そこに蓄積されたデータを分析すれば、研修の効果検証やキャリア教育に役立てられる可能性がある。本研究では社内SNSを開発する企業との共同研究として、SNSに蓄積されたデータを自然言語処理技術によって分析する。分析結果から、個人に合わせた研修の提案や、一人ひとりの特徴を可視化し、研修の効率化を実現することが研究の目的である。

**キーワード：**社内研修、自然言語処理、OJT、Off-JT

## Analysis of internal social media for in-house job training aimed at improving the efficiency of human-resource development

### 1. はじめに

この研究は、社内研修をターゲットとした研究である。日本には新卒一括採用制の伝統があり、新卒採用は、採用後の企業による積極的な能力開発を伴ってきた。この方式は同時に初期職業訓練を企業が行うことを当然とした[1]。このような背景から現在ほとんどの企業は自社で研修の仕組みを持っており、社員を会社で育てていくという考えが強い。そのため多くの企業が社内研修に力を入れており、研修の成果は会社の成長に直結している。

ところが多くの企業では、研修生は研修で学んだ内容の10%程度しか現場で実践できていないことがわかっている[2]。これは研修の効果測定に基づくフィードバックや改善が十分に行われていないことが背景にあり、実務に結びつかないような研修を実施している場合も多い。しかし、社員一人ひとりの研修の効果を人手で把握するのは多くの時間と労力が必要となり現実的ではない。このため、研修の効果測定の効率化は企業にとって重要な経営課題となっ

ている。本研究ではこの効果測定の指針として以下の2つを取り上げる。

(1) 研修生のスキルやモチベーションなどの状態がどう変化したか

(2) 研修内容は実務に結びついているか

このような研修の効果を定量化し自動的に評価する方法は確立されておらず、有効な仕組みを提案する研究も見当たらない。

一方で、企業におけるSNSの利用は、2005年以降相次いでおり、たとえば、NTT東日本の「Sati」やNTTデータの「Nexti」があげられる[3][4]。このように近年多くの企業が社内でのコミュニケーションにSNSを利用しており、その中で社内研修にもSNSを利用する企業も増えている。社内研修に利用されるSNSには様々な情報が蓄積されており、例えば研修の課題や試験、上司と研修生の会話の内容などがある。本研究では、これらを利用して上述の2つの指針の評価を目指す。

本研究の大きな目的は、社内SNSのデータを自然言語処理技術によって分析し研修の効果測定を効率化することである。その第一段階として、本研究では指針(1)に着目し、研修生の学習行動状態の推定と比較を目指す。社内研修に関連する先行研究との違いは、対象とするデータセットが含む情報である。評価自体を目的としたアンケートではなく

<sup>1</sup> 東京工業高等専門学校、情報工学科、制御情報研究室

<sup>2</sup> 株式会社ウーシア、取締役

a) s15141@tokyo.kosen-ac.jp

b) tomida@ousia.me

c) yamashita@tokyo-ct.ac.jp

d) matsu@tokyo-ct.ac.jp

く、研修の課題や研修生と上司の会話などを含んだデータを分析すれば、研修そのものの評価ができると考えられる。

## 2. システムとデータセット

本研究では、株式会社ウーシアが開発・展開している「研修フォローアップアプリケーション Core」[5]によって得られたデータを利用する。Core は社内研修用の SNS で、研修の報告や研修生と上司の発言などを管理している。このシステムを利用する研修生はそれぞれ複数の研修プログラムを受講している。各会社がそれぞれいくつかの研修プログラムを持っており、研修生は一定期間ごとに研修プログラムを受講する。研修生はグループに属し、同じグループの研修生は同じ研修プログラムを受講する。1つの研修プログラムが終了した後、適時研修内容の実践状況を報告するための質問フォームが作られる。このフォームには、研修プログラムに関連する質問が複数あり、質問への回答には上司や他の研修生がコメントすることができる。質問内容はグループのトレーナーが設定する。質問と回答の例を表 1 に示す。

データセットには上記の研修プログラムや、質問フォーム、質問フォームの質問、回答、コメントなどが含まれている。データセットの概要を表 2 に示す。本論文では、質問フォームの回答を分析する。以降、質問フォームへの回答を「回答」と呼ぶ。

## 3. アプローチ

### 3.1 基本設計

1 章の指針(1)を評価する方法として、本研究では図 1

表 1 質問と回答の例

コース	マネージャー研修
質問	自身・部下の成功に向けて取り組んでいることはなんですか？
回答（受講者 A）	計画的に行動できるように心がけています。また、部下とともに定期的な振り返りの時間を持ち、課題を共有し、対策の検討をおこなっています。
回答（受講者 B）	担当者ごとに異なる認識を持っていたり、異なる立場であったりするため、合意をとっていくことは難しいと感じました。話し方や立ちふるまいなどを、試行錯誤しながら取り組んでいきたいと思います。

表 2 質問と回答の例

会社数	19 社
グループ数	70 グループ
ユーザ数	1263 名
フォーム数	331 フォーム
回答数	12538 件
回答の単語数	5~609 単語

のような評価モデルを提案する。

図 1 は研修生が研修プログラム A, B, C を順番に受けている様子を表しており、研修生の状態は研修プログラムごとに変化している。このように研修プログラム受講後に、研修生の状態に変化があると仮定する。これを仮説 a とする。このとき、それぞれの状態を推定できれば時系列で比較することができる。本研究では、研修プログラム受講後の研修生の状態は、研修生が研修プログラム受講後に記述する回答から推定できるのではないかと考えた。これを仮説 b とする。そこで、質問フォームへの回答に注目し、研修生の状態を推定する。

はじめに、データセットの特徴を知るために潜在的トピック分析を行なった。その結果に基づき研修生の状態について議論し、研修生の状態モデルを決定した。次に、決定したモデルに基づいて手作業でラベル付きデータを作成した。最後にラベル付きデータを用いて研修生の状態の推定を行った。

### 3.2 アルゴリズムと実装方法

それぞれのアルゴリズムと実装方法について説明する。

#### 3.2.1 潜在的トピック分析

潜在的トピック分析には LDA(Latent Dirichlet Allocation) を用いる。LDA はトピックモデルの 1 つで、文書の確率的生成モデルである。学習段階では文章から固定数のトピックを抽出し、各トピックは複数の単語と単語に対する重みが付与される [6]。

本研究で扱うデータセットは、ラベル付きデータが存在せずどのような特徴があるかわからないため、教師なしアルゴリズムであり潜在的なトピックを抽出できるトピックモデルを選んだ。その中でも回答が複数のトピックを含む可能性があるため、1つの文書に含まれるトピックの割合を推定できる LDA を利用した。

#### 3.2.2 ラベル付きデータの作成

ラベル付きデータの作成では、3.2.1 節で定義したモデルに基づいて手作業で回答にラベルを付ける。ラベルは、Core を運営する(株)ウーシアと協力して LDA による分析結果を考察し「反省」「理解」「決意」「該当なし」「わからない」の 5 つに決定した。本研究では、振り返りや感じ

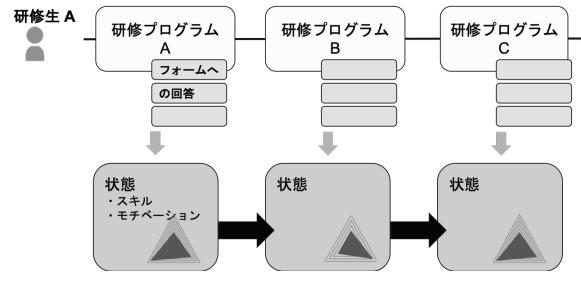


図 1 評価モデル

たことを述べている場合は「反省」、反省の深堀りや学んだことである場合は「理解」、次に何をするかや目標についてである場合は「決意」と定義した。これは4.2節で述べるカテゴリと対応している。また、3つのラベルのどれにも該当しない場合は「該当なし」、「該当なし」か判断がつかない場合は「わからない」というラベルをつける。さらに、1つの回答に複数の状態が含まれている可能性があるため、回答を読点で区切り文単位でラベル付けを行う。

データにラベルを付ける人のことをアノテータと言う。アノテータは上記で説明したラベルについて理解を統一する必要があるため、不特定多数をアノテータとすることは難しい。また、ラベル付けには社内研修についてのある程度の専門知識が必要となる。これらの理由から、Coreを開発、運営している株式会社ウーシアの方々にラベル付けを委託した。ラベル付けの際には、社内でラベル定義の理解を統一する時間を設けて頂いた。表3に、アノテータ数とデータ数を示す。本研究ではアノテータ数が限られており、ラベル付けの作業は個人への負担が大きい。そのため、doccano[7]というラベル付けツールを利用した。doccanoは人手でのラベル付けを容易にするツールであり、単純な操作かつ、少ないステップでラベル付けができる。ラベル付け画面は図2のようになっている。アノテータはショートカット又はクリックでラベルを選ぶ。1つのラベルに決定できない場合も考えられるため、「該当なし」と「わからない」以外のラベルは複数選択可とする。ラベル付け終了後、複数のラベルが付いているデータは、後述の分類問題を簡単にするためにラベルを1つに決定する。ラベルの決定方法は、ラベルのばらつきを調査し、その結果によって調整する。

### 3.2.3 研修生の学習行動状態の分類

次に作成したラベル付きデータを用いて分類する。分類で用いる、TF-IDF, Doc2Vec, PCA, t-SNE, SVMについて説明する。

表3 アノテータとデータの数

アノテータ数	5名
データ数	6079件
ラベル付けされたデータ数	3130件
データの平均単語数	8~66文字



図2 ラベル付け画面

TF-IDF(Term frequency-inverse document frequency)はコーパス内の文書にとって各単語がどれほどその文章の特徴を表現しているかを示す数値である。主に情報検索やテキストマイニングに使用され、単語の出現頻度(TF)と逆文書頻度(IDF)の積で求められる。TF-IDF値は単語の出現頻度に比例するが、「ある」「私」などの特定の頻出単語は重要ではない。IDFはこのような頻出用語の重みを下げ、希少用語の重みを増やすための数値である[8]。

Doc2Vec(Document to vector)は、文書をベクトル化するアルゴリズムで、ニューラルアーキテクチャを使用して前後に出現する単語から単語ベクトルを学習するWord2Vecの拡張として提案された。Doc2Vecは、Word2Vecと同様に前後に出現する文書から文書ベクトルを学習する[9]。

PCA(Principal component analysis)は、次元圧縮のアルゴリズムの1つであり。変数の相関構造を推定し、関連の強い変数をまとめることで高次元のベクトルを次元圧縮する[10]。

t-SNEは、次元圧縮のアルゴリズムの1つであり、高次元データの視覚化に優れている。各データポイントを一对比較により2次元または3次元に圧縮することで高次元データを視覚化する。生成される視覚化は、ほとんどすべてのデータセットで他の手法によって生成される視覚化よりも大幅に優れていることがわかっている[11]。

SVM(Support vector machine)は、与えられたデータを超空間上で複数の集合へと分離する際、マージンを最大にすることによってマージン最大化によって分離超平面を求め、データの境を学習するモデルである。ここでマージンとは直感的にはデータ店から分離超平面までの距離である[12]。

本研究ではこれらのアルゴリズムを用いて次の方法で分類を行う。まず、ラベル付きデータをそれぞれベクトル化する。次に、ベクトルを次元圧縮したものをプロットし、可視化する。可視化したグラフを目で確認し、分類できるかどうかを判断する。分類ができれば、グラフ内のラベル境界線を学習し、分類機を作成する。最後に、テストデータを分類機で分類し、精度を評価する。以下に、それぞれの実装方法を述べる。

ラベル付きデータのベクトル化は2つの方法で行う。1つ目はDoc2Vecを用いて300次元にベクトル化する方法である。Doc2Vecは前後から文章の特徴量を学習するため文の流れを考慮できる可能性がある。実装にはGensimのライブラリ[13]を用いる。2つ目は、n-gramにスコアを付ける方法である。ラベル付きデータを目視で確認したところ言い回しなどの共通点があった。このため連続した表現の考慮が可能で単純な考え方であるn-gramを採用した。n-gramを用いたベクトル化は次の手順で行う。まず、出現するn-gram全てにこの後説明する方法でスコアをつける。次に文のスコアを文が含むn-gramのスコアを用いて

求める。このとき、文のスコアは3次元で表され、これをその文のベクトルとする。このベクトル化では、次元が3となり既にグラフに表せる次元数であるため次元圧縮は行わない。

スコアの計算方法は、各ラベルへの出現回数とTF-IDFスコアの2パターンである。各ラベルへの出現回数は以下の手順で求める。まず、ラベル付きデータに出現する全てのn-gramについて、「反省」のラベルがついた文に出現したとき「反省」に1点加える、というように、n-gramの出現回数を各ラベルで求める。ここで、1つのn-gramに、「反省」、「理解」、「決意」の3軸の出現回数スコアが求められる。次に、文に含まれるn-gramの出現回数スコアの和を軸ごとに求め、その文のスコアとする。

各データのTF-IDFスコアは以下の手順で求める。まず、ラベル付きデータに出現する全てのn-gramのTF-IDFスコアを計算した。TF-IDFスコアはラベルごとに計算されるため、1つのn-gramに、「反省」、「理解」、「決意」の3軸のスコアが求まる。次に、文に含まれるn-gramのスコアの和を軸ごとに求め、その文のスコアとした。

nの値は、n=1, 2, 3の3つで実験する。上記のスコアとnの値を変化させた全ての組み合わせの場合でn-gramを比較し、最も良いものを採用する。比較ではスコアが高いn-gram上位5件を抽出し、ラベルごとに差が出るか、専門用語が入らないかなどの観点で評価を行う。

ベクトルの次元圧縮とグラフへのプロットはPCAとt-SNEを用いた場合の2パターンで実験を行なった。どちらの場合も、二次元及び三次元に圧縮しグラフにプロットする。

このグラフを目視で確認したとき、ラベルごとに分布が偏りプロットした空間上で複数のグループに別れれば、高い精度で分類ができる可能性がある。そのため、まずデータをプロットし目視で結果を確認した。プロットの際はラベルごとに色分けをし、各ラベルの分布を確認できるようにした。実装にはscikit-learnのライブラリ[14]を用いる。

ラベルの境界線の学習はSVMで行う。ラベル付きデータのラベルを正解とし、「反省」「理解」「決意」「該当なし」の4のグループに分かれるよう境界線を学習させる。このとき、「わからない」のラベルがついたデータはノイズ除去のため使用しない。最後に、テストデータを用意し学習させたモデルで分類したときの正解率を評価する。

### 3.2.4 教師なしアルゴリズムを用いた特徴語の可視化

さらに、潜在的トピック分析とは別に、各回答にどのような特徴があるのか、教師なしアルゴリズムで分析する。まずユーザを一人選び、TF-IDFスコアを用いて研修プログラムごとに回答の特徴語を調査する。TF-IDFスコアの計算に用いるコーパスは選んだユーザの全回答とし、文書は1つの研修プログラムでの回答全体とする。次に、TF-IDFスコアの高い上位5単語を特徴語とし、研修プログラムご

トピック	上位5単語				
	確認	人間関係	講演会	活用	訪問
1	計画	実践	踏まえる	難しい	面談
2	駆使	質問	スライド	用意	情報検索
3	把握	顧客	効果	確認	スライド

とに比較する。ユーザを変えて同じ手順で結果を比較し、傾向を調査する。

### 3.2.5 前処理

これらのアルゴリズムを使用するにあたり、データの前処理を行う。はじめに文章から記号と英字を除外し、形態素解析エンジンMeCabを用いて分かち書きする。MeCabは条件付き確率場に基づく高い解析精度を持ち、現存の日本語の形態素解析エンジンChaSenやKAKASIと比べて高速に動作する[15]。辞書はmecab-ipadic-neologd[16]を使用した。このとき、潜在的トピック分析、特徴語の抽出では、わかりやすい特徴量を抽出するため名詞、動詞、形容詞以外の形態素を除外する。ラベル付きデータを用いた分類では、語尾や言い回しも含めるため、特定の形態素の除去は行わない。

## 4 実験

潜在的トピック分析、研修生の状態モデルの決定、ラベル付きデータを用いた状態の推定の3段階で実験・調査を行なった。

### 4.1 潜在的トピック分析

まず潜在的なトピックの調査を行なった。本研究では3.1節に示すように研修生のフォームへの回答を分析するが、一回の回答では文章が短くトピックが適切に別れないため、同じ研修生の複数の回答を結合したものに入力とした。LDAにより抽出されたトピックを表4に示す。

### 4.2 研修生の状態モデルの決定

4.1節の結果から、回答がどのような内容を含むのかCoreを開発する企業の方々と考察し、以下の3つのカテゴリに分けられるという仮説を立てた。これを仮説cとする。

- 反省 : 振り返り、感じたこと
- 理解 : 反省の深堀、学んだこと
- 決意 : 次何をするか、目標

これらは、研修生がどのように研修に取り組んでいるか、という学習行動状態を表すと言える。この3つの状態を基に、本研究では研修生の状態モデルを提案する。これを図3に示す。

### 4.3 研修生の状態の推定

決定したモデルに基づいて研修生の状態を推定した。

#### 4.3.1 ラベル付きデータの作成

提案したモデルに基づき、株式会社ウーシアの方5名に協力して頂きラベル付きデータの作成を行なった。この際、1つの回答に複数の状態が含まれている可能性があるため、回答を読点で区切り、文単位でラベル付けを行った。また、ラベル付けの基準が各データで変わらないようにするため、ラベル付けするデータの順番はアノテータごとにランダムに決定した。

本研究では「反省」「理解」「決意」がそれぞれ300件以上集まるまでラベル付けを行なった。この作業は後述のラベル付きデータを用いた分類と同時に進め、その結果を見ながら、ラベル付けするデータ数を決定した。各ラベルの分布を図4に示す。

図4より、「理解」と「該当なし」のラベルがついたデータが多く、「反省」と二倍近く差がついた。この結果についてアノテータと議論したところ、「該当なし」とラベル付けされたものは、研修内容などの単なる出来事を書いているものが多いことがわかった。また、これは研修内容を回答させる質問が多数存在することが原因であるとわかった。

#### 4.3.2 ラベルのばらつきの調査

本研究では後述の分類問題を簡単にするために、各データのラベルを1つに決める。そのため、ラベルのばらつきを調査しラベルの決定方法を議論した。このとき、「わからない」ラベルが付いたものは除外した。調査の結果、「わからない」を除いた2923件の内、付与されたラベルがすべて同じものは259件、異なるラベルが1つずつ付けられたものは272件、一人がラベル付けしたものは2259件であった。そこで、異なるラベルが付いたデータをランダムに数件選び、ラベルを確認した。その結果、「理解」と「反省」や、「反省」と「該当なし」が同じデータに付けられる傾向があることがわかった。また、類似しにくいと考えられる「反省」と「決意」が付けられている場合もあり、ラベルのばらつきが大きいことがわかった。

また、一人のみがラベル付けしたデータは信頼性が低く、複数人が同じラベルをついているものが信頼性が高い。そこで、複数人がラベルを付けたデータの内、多数決でラベルが決まるデータ数を調査した。その結果、該当

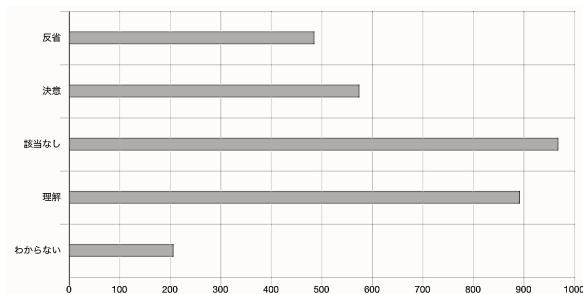


図4 ラベル付け結果の分布

するデータは391件であった。これは全データの1.3割程度となってしまう。しかし、アノテータの中の1名はこのような分析に対して専門性を有しており、このアノテータのラベルは高い信頼性を有しているため、採用することとした。このアノテータを高信頼アノテータと呼ぶこととする。以下にラベルの決定方法をまとめる。

複数人がラベルを付けたデータの内、多数決でラベルが決まる場合はそのラベルをデータのラベルとする。さらに、ラベルが同票の場合または高信頼アノテータのみがラベル付けしている場合、高信頼アノテータのラベルをデータのラベルとする。この場合、使用可能なデータは720件となり、全データの2.4割となるため、この方法を採用した。

#### 4.3.3 ラベル付きデータを用いた分類

この実験では、まず、ラベル付きデータをそれぞれベクトル化し、空間上にプロットする。さらに、SVMを用いてラベルの境界線を求め、分類を行う。データのベクトル化は2種類の方法で行なった。1つ目はDoc2Vecによるベクトル化、2つ目はn-gramのスコアによるベクトル化である。

1つ目のDoc2Vecを用いたベクトル化では、2次元への次元圧縮にPCAとt-SNEを用いた場合の2パターンで実験を行なった。全データをPCAで次元圧縮を行いプロットしたものを図5に、t-SNEで次元圧縮を行いプロットしたもの図6に示す。

PCAは各ラベルのデータが全体に散らばっていて分布に偏りは見られない。可視化に特化しているt-SNEでも、外れているデータはあるが全体的に偏りが少ない。このため、Doc2Vecでベクトル化した場合、ラベルを区切る境界線を引くことはそのままの状態では困難である。

2つ目のn-gramを用いたベクトル化について述べる。まず、3.2節で述べた方法でn及びn-gramへのスコアの種類の決定をし、次に各データに3つのスコアを付与した。事前調査としてn=1,2,3の場合の、スコアの上位5件を比較した結果、bi-gramのTF-IDFスコアが最も重複がなく専門用語が入らない結果となった。これはTF-IDFスコアでは一般的な表現を排除できるためであると考えられる。また、uni-gramは連続的な表現を考慮できないためbi-gramの方が良い結果になったと考えられる。tri-gramでは重複

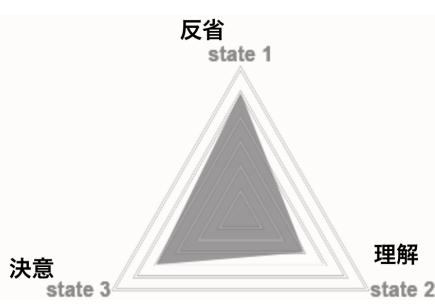


図3 研修生の状態モデル

が多く、これはスコアの差が小さくなるためであると考えられる。したがって、 $n=2$ 、スコアはTF-IDFスコアを用いることに決定した。このときの、各ラベルのbi-gramとスコアを表5に示す。

次に、決定したn及びスコアを用いて文を3次元にベクトル化した。各文をプロットした結果を図7に示す。この結果から該当なし以外は綺麗に別れることがわかった、

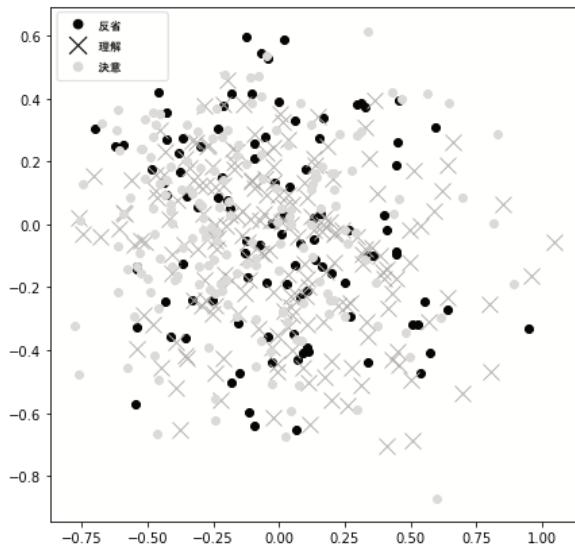


図5 PCAで次元圧縮した場合

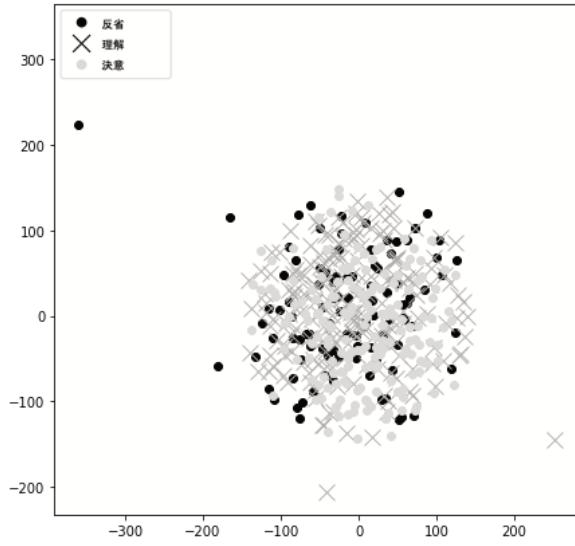


図6 t-SNEで次元圧縮した場合

表5 各ラベルのbi-gramとスコア

反省		理解		決意	
bi-gram	スコア	bi-gram	スコア	bi-gram	スコア
反省し	0.00327	が重要	0.00246	いきます	0.00362
が課題	0.00233	重要だ	0.00207	でいき	0.00307
課題です	0.00205	学びまし	0.00188	たいです	0.00230
と反省	0.00205	を学び	0.00132	行きたい	0.00167
不足し	0.00164	実感し	0.00129	付けて	0.00167

最後に、SVMで分類を行った。ただし学習データとテストデータはランダムに選びその比は8:2とした。その結果0.944の正解率が得られた。分類機が予測したラベルを図8に示す。分類結果は図7と対応しており図7が正解ラベルとなっている。この結果から、4.2節で提案した3軸の学習行動状態で分類ができたと言える。また、該当なしも含めて高い精度が得られたため該当なしの除去にも利用できる可能性がある。

#### 4.4 教師なしアルゴリズムを用いた特徴の可視化

この実験ではTF-IDFを用いて研修プログラムごとに特徴語の可視化をした。本研究ではデータセットから数人のユーザを選び結果を比較したが、ここではその内の一人の

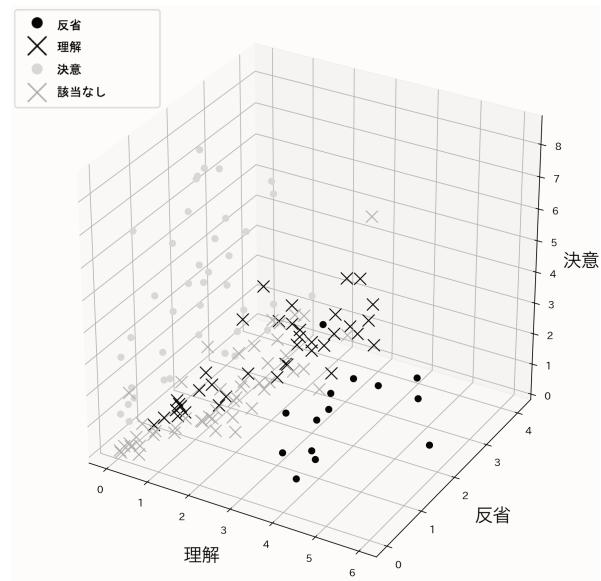


図7 各文のbi-gramのTF-IDFスコア

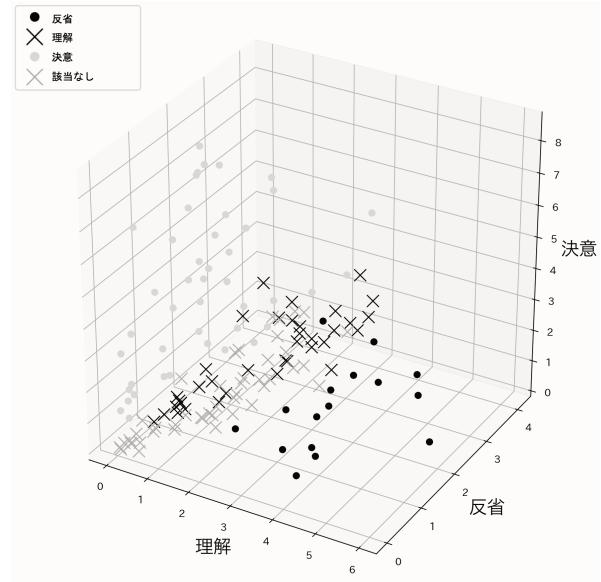


図8 予測ラベル

結果を紹介する。このユーザをユーザ A と呼ぶ。ユーザ A は研修プログラム A から F までを順番に受講していた。研修プログラムごとの特徴語を図 9 に示す。

太字で示した単語はユーザの学習行動状態を表すと考えられる。「学ぶ」という単語からは、ユーザが知識を得たいという状態だと考えられる。また、「自身」、「関連性」、「分解」、「具体的」は章で説明した学習行動状態モデルの「理解」を意味すると考えられる。さらに、「活用」、「検討」は学習行動状態モデルの「決意」に該当すると考えられる。

その他の単語は、「マーケット」や「事業戦略」など、研修の内容や専門用語と考えられるものが多い。何を示すかわからない単語も数個存在した。

## 5. 考察

本研究の目的は 1 節の指針 (1) に着目し、研修生の学習行動状態を推定することであった。推定のために、本研究では当初以下の 3 つの仮説を立てた。仮説 a : 研修プログラムで学習行動状態が変化すること仮説 b : 回答から学習行動状態が推定できること仮説 c : 学習行動状態は 3 つあることこの章では、これらの仮説と実験結果を照らし合わせ、仮説の真偽や期待できること考察する。

### 5.1 仮説の検証

ラベル付けの結果、提案した 3 次元の学習行動状態に該当しない場合が多く存在することがわかった。さらに 4.3.3 節で回答を分類したところ、提案した 3 軸で、「反省」「理解」「決意」に加え該当なしを含めた 4 つに高い精度で分類できた。この結果から仮説 b の実現可能性を示すことができ、仮説 c については提案した 3 軸の学習行動状態が存在するが、状態に関係ない回答もあることがわかった。この学習行動状態に関係のない回答は本研究の分類方法で除去できる可能性がある。また、アノテータとのディスカッションから、該当なしがつけられたラベルは单なる事実という場合が多かったことがわかった。これにより、潜在トピック分析で得られた専門用語や研修内容などが入ったトピックは、該当なしに当たるのではないかと考えられる。4.3.3 説の分類と 4.4 節の特徴語分析では、研修プログラムごとに実験結果が変化し仮説 a が正しい可能性があることがわかった。これは時系列での比較が可能になるということ

であり、研修プログラムの効果の自動的評価ができるようになることが期待できる。また、研修生によって実験結果に傾向があり、研修生同士、過去の研修生との比較ができる可能性があるとわかった。

### 5.2 学習行動状態の分類による可能性

今回の研究で、学習行動状態を分類できることがわかった。同じ方法で様々な軸を定量化できる可能性があり、本研究で実施した方法を用いて他の特徴が見つかることも期待できる。また、学習行動状態の比較ができたとしても、それが良い変化かどうか判断するのは難しい。しかし、データがさらに蓄積されれば優秀な生徒の過去の変化を学習することで、研修生の変化が良いものかどうかの評価ができる可能性がある。これは指針 (1) の評価を解決する。

本研究の結果を実システムに適用する際、考えられる実装内容は 2 つある。1 つ目は研修生自身、上司、研修担当者に研修生の学習行動状態をフィードバックするシステムである。研修生自身へのフィードバックは、研修生が自身を客観的に見ることができ自分が見えていないところと見えているところがわかるようになることが期待できる。また、どのような社員と自身が似ているかがわかることで、今後のキャリアプラン構築の手助けになる可能性がある。上司へのフィードバックでは、部下の特徴、傾向の可視化の実装を考えている。これにより部下のスキルやモチベーションを把握できるようになれば、実務での指導がより効率的になることが期待できる。また、人材配置にも役立てる可能性がある。研修担当へのフィードバックは、研修プログラムの影響が定量的にわかるため研修プログラムの分析に役立つと考えられる。以上のように、研修のフィードバックができるようになることは、3 つの立場において研修の大きな効果向上につながる可能性がある。

2 つ目は上司への研修内容の自動的な報告である。社内研修は、会社全体でサポートする環境を作ることが大切である。そのためには周りの社員から理解を得る必要があり。そのためには何をしているかや研修の現状を報告することが不可欠である。このシステムを利用することで研修状況を自動的に簡潔に伝えられれば、研修に理解のある環境を構築し易くなると期待できる。指針 (2) の評価については、実務で使っている他の SNS 「slack」 やコメントの分析を考えている。実務内容や、何に対して課題感を持っているかを定量的に分析できれば、実務につながる研修を提案できる可能性がある。

研修についての研究は手法が確立されておらず、データセット構築手法なども含め多くの実験が必要になる。そのためディスカッションしながら時間をかけて進めることになるが、研修を様々な面で効率化できる可能性が大いにある。

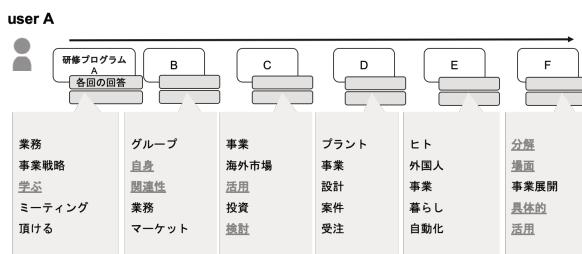


図 9 ユーザ A の研修プログラムごとの特徴語

## 6. おわりに

本研究では、社内研修の自動的な評価の仕組みを実際のシステムに適用することが最終目標であり、その第一段階として質問フォームへの回答に注目し、研修生の学習行動状態の推定と比較を行った。まず潜在的なトピックを調査し、「研修生の状態モデル」を提案した。次に、提案したモデルに基づいてラベル付きデータを作成し分類を行ったところ、bi-gram 単位で TF-IDF スコアをつけ SVM で学習することで 0.964 の正解率が得られた。今後の課題は、どの状態にも該当しないデータの定義を改善すること、別の軸でラベル付けをし分類すること、実際のシステムに実装し有用性の評価することである。

謝辞 株式会社ウーシアの皆様には実験方針についてご討論頂き、さらにデータセット構築にご協力頂いたことに深謝する。

## 参考文献

- [1] 小杉 礼子: 新卒一括採用の功罪と対応 : 移行の再組織化, 産業教育学研究 (2014).
- [2] 中原 淳, 島村 公俊, 鈴木 英智佳, 関根 雅泰: 研修開発入門 「研修転移」の理論と実践, ダイヤモンド社 (2018)
- [3] 小豆川裕子: 株式会社 NTT データ経営研究所, 企業内 SNS の利用状況と効果をめぐって-NTT データのケースを中心に入手先 <<http://warp.da.ndl.go.jp/info:ndljp/pid/11175166/www.keieiken.co.jp/monthly/2010/1008-08/>> (参照 2020-02-20).
- [4] 加藤 菜美絵, 諏訪 博彦, 太田 敏澄: 企業内 SNS 導入に関する利用者調査, 情報処理学会論文誌, Vol.55, No.1, 221-229(2014).
- [5] OUSIA: changing the ecology of learning 入手先 <<https://ousia.me/>> (参照 2020-02-20).
- [6] D. Ramage, D. Hall, R. Nallapati, C. D. Manning: *Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora* EMNLP 248–256(2009).
- [7] doccano: Text annotation for Human 入手先 <<https://doccano.herokuapp.com/>> (参照 2020-02-20).
- [8] J. E. Ramos, Using TF-IDF to Determine Word Relevance in Document Queries, In Proceedings of the first instructional conference on machine learning, January 2003.
- [9] JH Lau, T Baldwin: *An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation*, Proceedings of the 1st Workshop on Representation Learning for NLP, pp. 78–86 (2016).
- [10] S Wold, K Esbensen, P Geladi: Principal Component Analysis, Elsevier Science Publishers B.V, 37— 52 (1987)
- [11] L Maaten, G Hinton, Visualizing Data using t-SNE, Journal of Machine Learning Research 9, 2579–2605, Nov 2008.
- [12] 平 博順, 春野 雅彦: *Support Vector Machine* によるテキスト分類における属性選択, 情報処理学会論文誌, (2000)
- [13] Gensim: Doc2Vec paragraph embeddings 入手先 <<https://radimrehurek.com/gensim/models/doc2vec.html>> (参照 2019-02-12).
- [14] scikit-learn, scikit-learn Machine Learning in Python 入手先 <<https://scikit-learn.org/stable/index.html>> (参照 2019-02-12).
- [15] Github. mecab: MeCab (和布蕉) とは, 入手先 <<https://taku910.github.io/mecab/>> (参照 2020-02-06).
- [16] Github: Neologism dictionary for MeCab 入手先 <<https://github.com/neologd/mecab-ipadic-neologd>> (参照 2019-08-30).