情報処理学会論文誌　教育とコンピュータ　Vol.7　No.3　12–23 (Oct. 2021)

ショートペーパー

社内研修の評価およ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[2,a)]　富田 陽介[3,b)]　山下 晃弘[1,c)]　松林 勝志[1,d)]

受付日 2020年7月7日，再受付日 2020年11月30日，
採録日 2021年6月19日

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，1人1人の特徴を可視化し，研修の効率化を実現することが研究の大きな目的である．その第1段階として，本研究では研修生の「学習行動状態」を定義し統計的な分析に基づく定量化を提案した．また，実データに基づいて時系列での可視化を行い，学習行動状態が変化する様子について考察した．その結果，研修によって学習行動状態が変化する過程や，個人ごとにその変化の過程が異なることが明らかとなり，今後，社内研修自体の評価や人事評価，キャリアプラン支援などに活用できる可能性を示した．

キーワード：教育の設計・測定・評価，テキスト処理，自然言語，学習過程，社内研修，OJT，Off-JT

Analysis of Internal Social Media for In-house Job Training Aimed at Improving the Efficiency of Human-resource Development

Akari Haga[2,a)]　Yosuke Tomida[3,b)]　Akihiro Yamashita[1,c)]　Katsushi Matsubayashi[1,d)]

Received: July 7, 2020, Revised: November 30, 2020,
Accepted: June 19, 2021

Abstract: Most Japanese companies have their own in-house training systems. However, only a few companies have systems to fully verify the training effectiveness. Furthermore, an increasing number of companies are using social media for both internal communications and in-house training discussions. Therefore, by analyzing this accumulated data in social media, it is possible to verify the effects of in-house training and career education. This research is a joint study with a company that develops internal social-media system for in-house training. We analyze the accumulated data in a practical system by using natural language processing. This research aims to visualize both training proposals tailored to each individual and the characteristics of each trainee from the analysis results. It also aims to improve the effectiveness and the efficiency of in-house training systems. In this paper, as the first step, we defined the "state of learning behavior" of trainees and proposed quantification the state. And we visualized in time series based on real data and considered how state of learning behavior changes. As a result, we found that the state of learning behavior changed by training. And the process of change is different for each individual. We showed the possibility of using it for evaluation of in-house training, personnel evaluation and career plan support.

Keywords: design, measurement, and evaluation of education, job training, text processing, natural language processing, learning processes, OJT, Off-JT

© 2021 Information Processing Society of Japan

1. はじめに

本研究は，社内研修の効率化および研修効果の向上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分かっている [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，実務に結び付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 1 人 1 人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，人が何かを学び実践するまでの心理状態や行動のことを学習行動状態と定義し，研修の効果測定の指針として以下の 2 つを取り上げる．

(1) 研修生の学習行動状態がどう変化したか
(2) 研修内容は実務に結び付いているか

自然言語処理分野においては，SNS 上の投稿から投稿者の感情を推定する研究が広く行われている．足立らは SNS のデータから感情語の相関を解析することで，ネット上のやりとりを通じて人々の間にどのように情報や感情が伝搬するかという問題を研究を行っている [3]．高橋らは Twitter の投稿に対するコメントの感情分析を行い炎上を検出する研究を行っている [4]．

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 [5], [6]．このように近年多くの企業が社内でのコミニュケーションに 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，たと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．

しかし，社内研修に関する投稿から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．そこで本研究では，一般の SNS 分析で用いられる自然言語処理の手法を社内研修専用の SNS にも応用し，「社内研修に関する投稿」の分析によって，上述の 2 つの指針の評価を目指す．社内で利用される SNS に蓄積されたテキストデータに対して自然言語処理に基づく分析を行い，研修生の学習行動状態の変化や，研修内容の実務への回答を結び付けて評価を目指す研究は少ない．本研究の目的は，社内 SNS の実データを分析することで社内研修と学習者の状態を分析評価する新たな手法を確立することである．まずデータをトピック分析することによりどのような軸があるのかを選定し，TF-IDF（Term frequency-inverse document frequency）を素性とした SVM（Support vector machine）でデータ内のそれぞれの文がどの軸に関連する記述であるかを分類することで，単なる感情ではなく学習行動状態を可視化する．

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，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」[7] 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．

[1] 東京工業高等専門学校情報工学科
Department of Computer Science, National Institute of Technology, Tokyo College, Hachioji, Tokyo 193–0997, Japan
[2] 電気通信大学 I 類メディア情報学プログラム
School of Informatics and Engineering, The University of Electro-Communications, Chofu, Tokyo 182–8585, Japan
[3] 株式会社ウーシア
Chief Technology Officer at Ousia, Inc., Nagoya, Aichi 464–0804, Japan
[a)] h1810750@edu.cc.uec.ac.jp
[b)] tomida@ousia.me
[c)] yamashita@tokyo-ct.ac.jp
[d)] matsu@tokyo-ct.ac.jp

© 2021 Information Processing Society of Japan

表 1　質問と回答の例
Table 1　Example of questions and answers.

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　データセット概要
Table 2　Dataset structure.

会社数	19 社
グループ数	70 グループ
ユーザ数	1,263 名
フォーム数	331 フォーム
回答数	12,538 件
回答の単語数	5〜609 単語

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** 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潜在的トピック分析を行った．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い，時系列で比較を行った．

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 [8]．

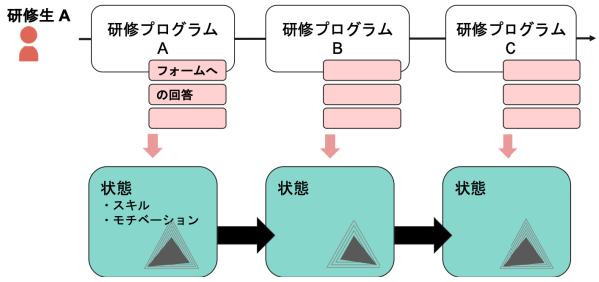


図 1　評価モデル
Fig. 1　evaluation model.

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分からないため，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 項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 を運営する株式会社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」の 4 つに決定した．本研究では，振り返りや感じ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」とした．

ラベル付けを小規模で試験的に実施したところ，「反省」と「該当なし」が類似しており，「該当なし」かどうかも判断がつかない場合があることが分かった．そのため「わからない」というラベルを追加し，「反省」，「理解」，「決意」，「該当なし」，「分からない」の 5 つのラベルでラベル付けを実施することに決定した．

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，回答を句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いう．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

© 2021 Information Processing Society of Japan

図 2　ラベル付け画面
Fig. 2　annnotation interface.

表 3　アノテータとデータの数
Table 3　Annotation settings.

アノテータ数	5 名
ラベル付けされたデータ数	3,130 件
データの平均単語数	8〜66 文字

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 [9] 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また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「分からない」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（Document to vector），PCA（Principal component analysis），t-SNE，SVM について説明する．

TF-IDF 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 [10]．

Doc2Vec 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，文書 ID と文書に出現する複数の単語を入力とし，単語を連結して次の単語を予測する [11]．

PCA 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 [12]．

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対比較により 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分かっている [13]．

SVM 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点から分離超平面までの距離である [14]．

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，高次元のデータに対して SVM によって分類器を作成し，分類の可否について検証を行う．また同時に，ベクトルを次元圧縮したものをグラフとしてプロットし，可視化する．可視化したグラフを目視で確認し，分類の状況について確認を行う．分類の可否の状況をふまえて，多次元空間上で決定されたラベル境界線を使用した分類器を作成する．最後に，構築した分類器によってテストデータを分類し，精度を評価する．以下に，それぞれの実装方法

© 2021 Information Processing Society of Japan

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 [15] を用いる．2 つ目は，形態素 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い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 すべてにこの後説明する方法でスコアをつける．次に文全体のスコアを，その文が含む 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 となりすで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すべての n-gram について，「反省」のラベルが付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すべての 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 種類，つまり形態素の uni-gram，bi-gram，tri-gram の 3 種類で実験する．上記のスコアと n の値を変化させたすべての組合せの場合で 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った．どちらの場合も，2 次元および 3 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分かれれば，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 [16] を用いる．

ラベルの境界線の学習は RBF カーネル（Radial basis function kernel）を用いた 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 つ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，「分からない」のラベルが付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．

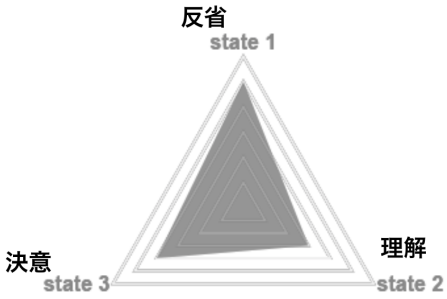


図 3　研修生の状態モデル
Fig. 3　A model of trainee's state.

3.3　学習行動状態の時系列比較

3.2.3 項の学習行動状態分類モデルで利用した TF-IDF スコアを用いて研修生の回答のスコアを計算し，時系列で比較を行う．学習行動状態は 3 軸のスコアで推定されるため，スコアを**図 3** のようなレーダチャートに表すことにする．研修生はランダムにサンプリングし，株式会社ウーシアの方々とディスカッションしながら研修プログラムごとに研修生がどのような変化をしたかを考察する．

3.3.1　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 1 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．

3.3.2　前処理

これらのトピック分析，TF-IDF，SVM 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 [17]．辞書は mecab-ipadic-neologd [18] 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，分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，語尾やいい回しも含めるため，特定の形態素の除去は行わない．

© 2021 Information Processing Society of Japan

表 4　潜在的トピック
Table 4　Latent topic.

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

表 5　文の具体例と学習行動状態
Table 5　Answers examples and learning behavior states.

文の具体例	学習行動状態
メンバーの意見を傾聴し時間をかけて話し合うことはできていません.	反省
メンバーの積極性を向上させる施策を打つ必要があるのかなと思います.	反省
しっかりとキーワードをお伝えすることで少し信用を頂けたのかな考えています.	理解
自分は合意形成を行っていく点においては長けていると考えている.	理解
自分のスタイルが確立できるよう日々精進したいと思います.	決意
メンバーに安心感を与えられるようなマネジメントを心がけたい.	決意
これからはプラン B に焦点が当たるのではないかという意見がありました.	該当なし
部門メンバー全員で 2 回目の打ち合わせを開催いたしました.	該当なし

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，1 回の回答では文章が短くトピックが適切に分かれないため，同じ研修生の複数の回答を結合したものを入力とした．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

表 5 に回答に含まれる文の具体例とそのカテゴリを示す．表 5 の 1 つ目の文は自身の反省を書いているためカテゴリは「反省」に分類できる．一方で 7 つ目の文はメンバからの意見であり自分の意見ではない．また最後の文も反省ではなく単なる事実である．**表 6** に質問と回答の例を示す．このように他人の意見や単なる事実を書いている文は質問内容に影響される場合が多いため 3 つのカテゴリには該当しないと判断することとした．

表 6　質問が回答に影響している例
Table 6　Affect of the question.

質問 1	アップデートした機能を上司やメンバーに何回くらい共有できましたか？
質問 1 に対する回答 1	共有は 3 回行いました.
質問 1 に対する回答 2	機能は 2 週に 1 回程度は更新しており，上司には定期的な面談で必ず共有しておりますので，回数としては 4 回です.
質問 2	シナリオを職場共有して得られた意見
質問 2 に対する回答 1	このシナリオが発現する可能性は低いとの意見があった.
質問 2 に対する回答 2	これからはプラン B に焦点が当たるのではないかという意見がありました.

また，これらのカテゴリを世界的な教育の評価法とされているカークパトリックの 4 段階評価法 [19] と照らし合わせるといくつか類似している点がある．以下にカークパトリックの 4 段階評価法を示す．

- Reaction（反応）：研修に対する満足度の評価
- Learning（学習）：学習到達度の評価
- Behavior（行動）：行動変容の評価
- Results（業績）：学習者や職場の業績向上度合いの評価

仮説 c は，研修中の被験者の回答を自然言語処理で分析した結果導かれた仮説であり，実データの分析から得られたものである．この仮説 c が，結果としてカークパトリックの 4 段階評価法と類似性を持つ点は興味深い．カークパトリックの 4 段階評価法を用いた研修効果の測定は，アンケートやヒアリング，試験などで明示的な測定を行うことを前提とするが，提案手法は研修によって蓄積されたテキストデータを自然言語処理によって分析し，研修効果の測定を達成しようとする試みである．本手法はより簡易かつリアルタイムに効果測定を可能にする手法として，カークパトリックの 4 段階評価法による効果測定を補完できる可能性があると考えている．

一方で，カークパトリックの 4 段階評価法の中には「業績」のレベルも定義されているが，本研究で使用したデータセットはあくまで研修期間中に得られたデータであり，そこからは研修後の業績の情報は得られないため，提案する学習行動状態のモデルには対応する状態が含まれなかったものと考えられる．

© 2021 Information Processing Society of Japan

以上の理由から，回答に含まれる内容は「反省」,「理解」,「決意」の 3 つのカテゴリに分けられるとした．

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いえる．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 名に協力していただきラベル付きデータの作成を行った．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，回答を句点で区切り，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」のラベルが付いたデータが多く,「反省」と 2 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分かった．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分かった．

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,「分からない」ラベルが付いたものは除外し,「反省」,「理解」,「決意」,「該当なし」のラベルが付いたデータのみを用いてラベルを決定することとした．調査の結果,「分からない」を除いた 2,923 件のうち，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，1 人がラベル付けしたものは 2,259 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分かった．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分かった．

また，1 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付けているものの方が信頼性は高い．そこで，複数人がラベルを付けたデータのうち，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

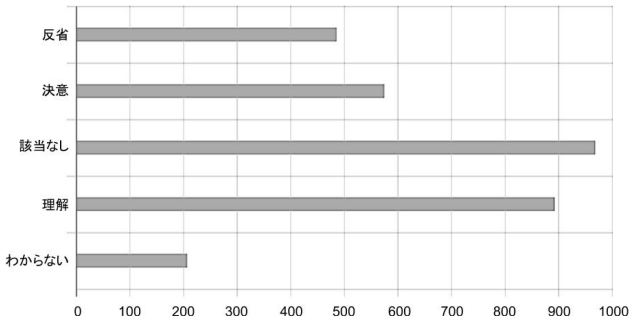


図 4　ラベル付け結果の分布
Fig. 4　Label distribution.

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．そのうえ，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．以降このアノテータを高信頼アノテータと呼ぶことにする．高信頼アノテータのラベルと多数決で決まるラベルの一致率を調べたところ，93.3%のラベルが一致した．以上の理由から以下のラベルの決定方法を採用する．

複数人がラベルを付けたデータのうち，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を**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 および 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

© 2021 Information Processing Society of Japan

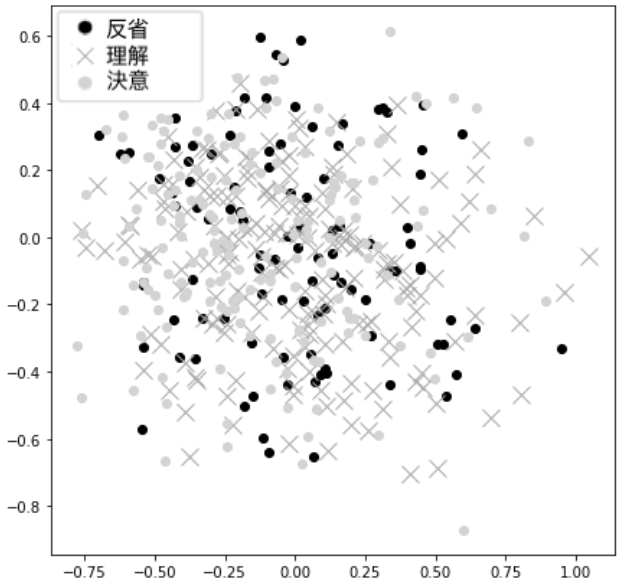


図 5　PCA で次元圧縮した場合
Fig. 5　Dimensional compression using PCA.

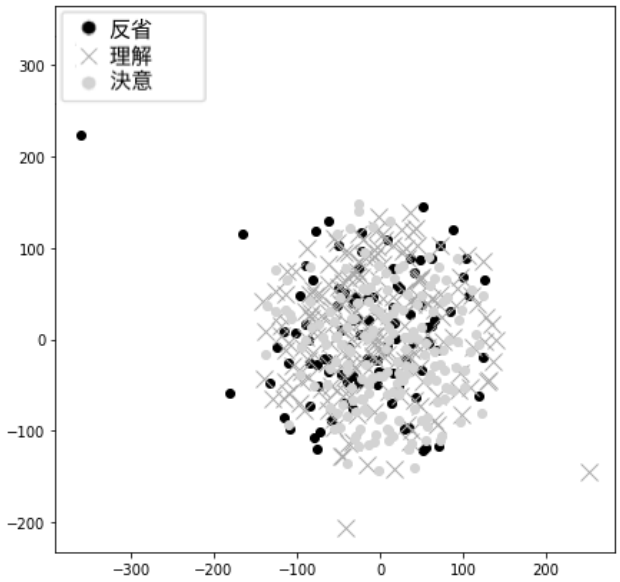


図 6　t-SNE で次元圧縮した場合
Fig. 6　Dimensional compression using t-SNE.

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 7** に示す．

次に，決定した n およ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分かれることが分かった，最後に，RBF カーネルを用いた 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

表 7　各ラベルの bi-gram とスコア
Table 7　bi-gram scores.

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

8 : 2 とした．その結果 0.944 の正解率が得られた．分類器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いえる．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　学習行動状態の時系列比較

この実験では 4.3.3 項の分類モデルで利用したスコアの計算法を用いて研修生の学習行動状態を研修プログラムごとに推定し，比較を行った．まず，研修プログラム終了時に研修生が記述した回答をプログラムごとに結合し，分類モデルで用いた TF-IDF スコアの計算法で回答のスコアを計算する．スコアは「反省」,「理解」,「決意」の 3 軸で計算されるため，3 つのスコアをレーダチャートに表し時系列で比較を行った．ここでは一部の実験結果を抜粋して説明する．**図 9** に研修生の学習行動状態の変化を示す．図 9 より，研修生 A のレーダチャートの変化を見てみると，最初の 2 回は理解に大きく偏っており最後に決意が大きい三角形になっている．このことから前半の研修で得られた理解を反省につなげることができ，最後に決意につながっていると考えられる．一方で，B は三角形が小さいものがあるなど，研修の成果がなかなか結び付いていないように見える．

実際のプログラムを確認したところ，プログラム A，B では自社戦略，プロセスの理解をし，プログラム C では理解から自身の役割を考え，プログラム D で将来の自社の戦略モデルを考えるという内容であり，レーダチャートと対応しているといえる．しかし，レーダチャートがどちらも決意に傾いたプログラム D においては「～すべきか」という質問がされており，質問内容が回答へ与える影響が大きいことが考えられる．

この結果から，時系列での新しい比較の方法ができることが分かった．しかしながら，学習行動状態自体の変化なのか，質問内容の影響によるものなのかということを考える必要がある．

4.5　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

© 2021 Information Processing Society of Japan

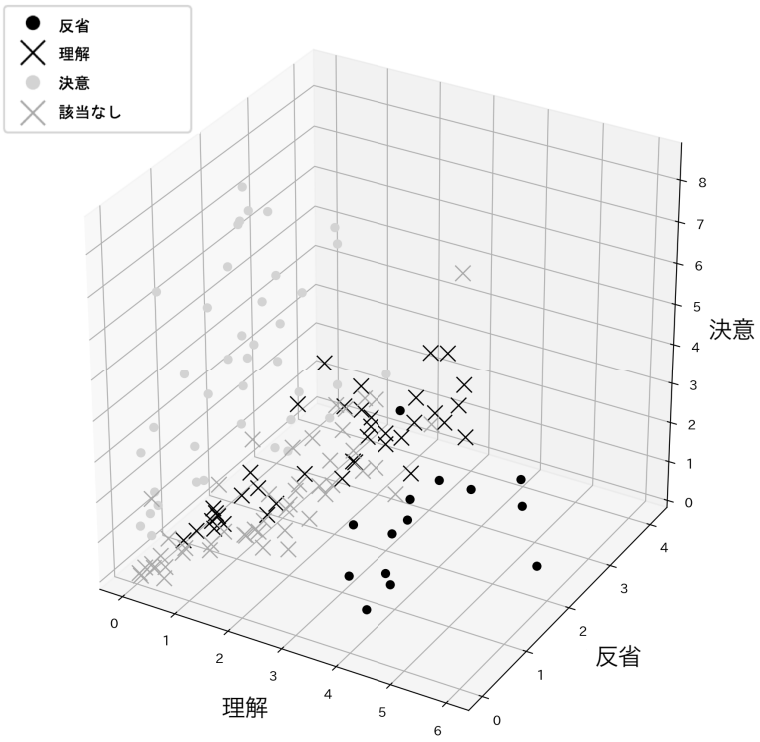


図 7　各文の bi-gram の TF-IDF スコア
Fig. 7　bi-gram TF-IDF scores.

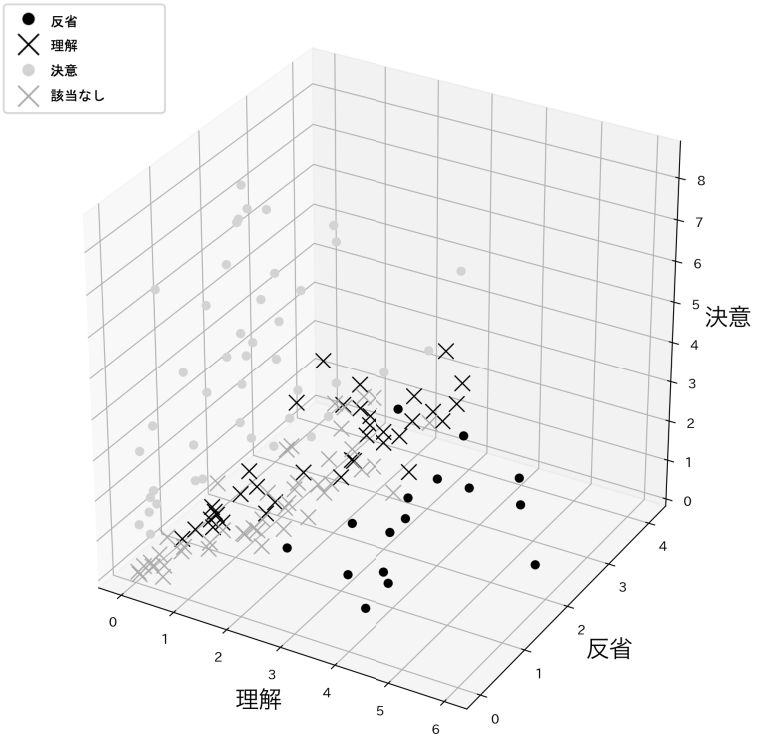


図 8　予測ラベル
Fig. 8　Predicted labels.

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，ここではそのうちの 1 人の結果を紹介する．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 **10** 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

© 2021 Information Processing Society of Japan

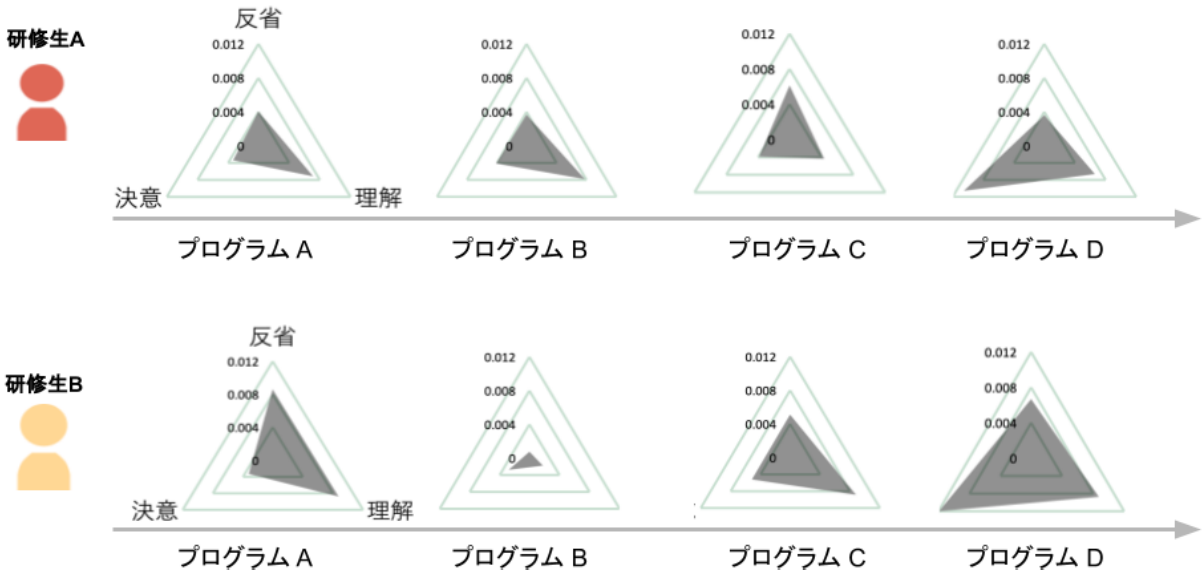


図 9　研修生の学習行動状態の変化

Fig. 9　Changes in the learning behavior of trainees.

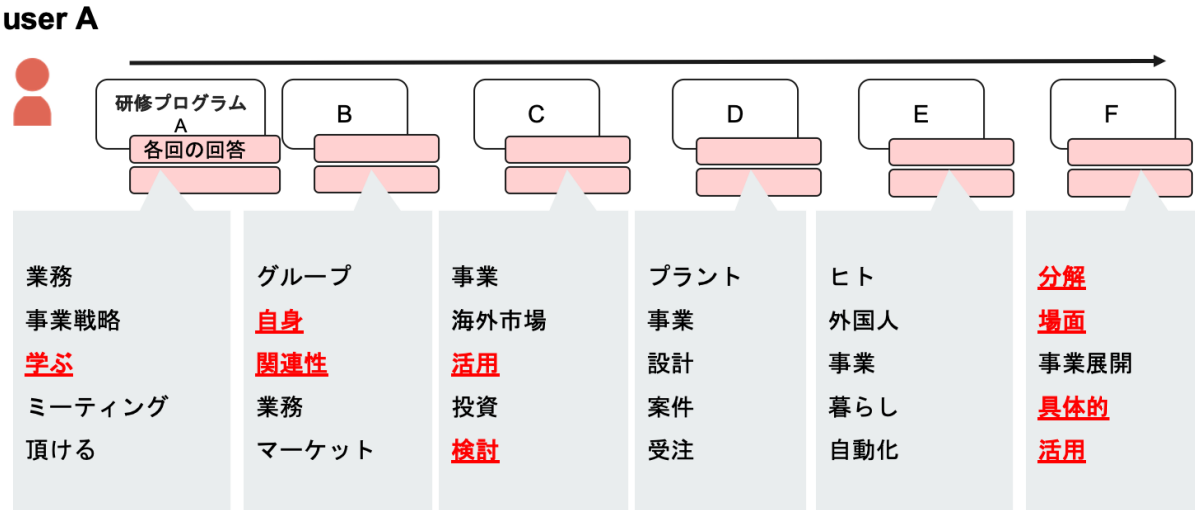


図 10　ユーザ A の研修プログラムごとの特徴語

Fig. 10　Feature words of UserA in each program.

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．何を示すか分からない単語も数個存在した．

5.　考察

本研究の目的は 1 章の指針 (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分かった．さらに 4.3.3 項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，状態に関係しない回答もあることが分かった．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，該当なしが付けられたラベルは単なる事実という場合が多かったことが分かった．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，該当なしにあたるのではないかと考えられる．4.3.3 項の分類と 4.4 節の時系列での比較，4.5 節

© 2021 Information Processing Society of Japan

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分かった．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分かった．

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分かった．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．

本研究では「文章の意味」の分析は行っておらず，あくまで形態素レベルでの統計的な処理である．そのためテキストの中に否定文や文脈に依存した表現が存在する場合，そのテキストが表現する内容を適切に分析できない懸念がある．たとえば「理解できたとは思いません」という文は「理解できた」という形態素 bi-gmram が「理解」のスコアを高くし，「理解」が強い文であると判定されてしまう．このようなことを防ぐために，文脈を考慮する手法が今後必要である．

また，文の書き方がスコアに影響を及ぼす問題が考えられる．たとえば，研修生は文章を書く際に文の自然さを確保するため意図的に同じいい回しを避ける場合があり，このときスコアが小さくなる可能性がある．そのため，より適切にスコアを算出するようスコアの量子化などを行うことを考えている．さらに研修生が高いスコアを出す書き込みを狙って文を書いてしまうという問題も考えられ，企業の方と話し合いながら対策を考えていく必要がある．

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分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分かることで，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られる可能性がある．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分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やす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．

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 1 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 0.944 の正解率が得られた．今後の課題は，別の軸でラベル付けをし分類すること，学習行動状態の推定結果と企業の人事担当の方の評価を照らし合わせ評価すること，実際のシステムに実装し有用性を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いただき，さらにデータセット構築にご協力いただ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] 足立　悠，戸田幹人：SNS における感情表現とその相互作用関係の抽出，情報処理学会研究報告，Vol.2015-NL-220, No.11 (2015).
[4] 高橋直樹，檜垣泰彦：Twitter における感情分析を用いた炎上の検出と分析，電子情報通信学会技術報告，LOIS2016-86 (2017).
[5] 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-

© 2021 Information Processing Society of Japan

08/〉（参照 2020-02-20）.
[6] 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, pp.221–229 (2014).
[7] OUSIA: changing the ecology of learning, available from 〈https://ousia.me/〉 (accessed 2020-02-20).
[8] Ramage, D., Hall, D., Nallapati, R. and Manning, C.D.: Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora, *EMNLP*, pp.248–256 (2009).
[9] doccano: Text annotation for Human, available from 〈https://doccano.herokuapp.com/〉 (accessed 2020-02-20).
[10] Ramos, J.E.: Using TF-IDF to Determine Word Relevance in Document Queries, *Proc. 1st Instructional Conference on Machine Learning* (Jan. 2003).
[11] Lau, J.H. and Baldwin, T.: An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation, *Proc. 1st Workshop on Representation Learning for NLP*, pp.78–86 (2016).
[12] Wold, S., Esbensen, K. and Geladi, P.: *Principal Component Analysis*, pp.37–52, Elsevier Science Publishers B.V. (1987)
[13] Maaten, L. and Hinton, G.: Visualizing Data using t-SNE, *Journal of Machine Learning Research*, Vol.9, pp.2579–2605 (2008).
[14] 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).
[15] Gensim: Doc2Vec paragraph embeddings, available from 〈https://radimrehurek.com/gensim/models/doc2vec.html〉 (accessed 2019-02-12).
[16] scikit-learn, scikit-learn Machine Learning in Python, available from 〈https://scikit-learn.org/stable/index.html〉 (accessed 2019-02-12).
[17] Github. mecab：MeCab（和布蕪）とは，入手先 〈https://taku910.github.io/mecab/〉（参照 2020-02-06）.
[18] Github: Neologism dictionary for MeCab, available from 〈https://github.com/neologd/mecab-ipadic-neologd〉 (accessed 2019-08-30).
[19] Kirkpatrick Partners: The Kirkpatrick Model, available from 〈https://www.kirkpatrickpartners.com/OurPhilosophy/The-Kirkpatrick-Model〉（参照 2021-06-12）.

推薦文

本論文は社内研修において，社内 SNS への投稿内容に対して自然言語処理技術を適用することによって研修生の学習行動状態をモデル化し，その状態を推定する手法を提案している．この手法を実データを用いて評価し，状態分類の適合率が 90%を超えるという結果を得ており，研修の効率化や効果の向上に役立つ研究となることが期待される．研修生の心理状態を正しく評価できているかを検証するには，SNS 投稿の分析に留まらず，インタビューなどのさらなる評価が必要であるが，本論文は第 1 著者がジュニア会員であったときに投稿されたものであり，今後，研究を発展させることで，大きな成果を得ることが期待できる．

（情報処理学会理事・論文誌「教育とコンピュータ」アドバイザー　西田 知博）

芳賀 あかり

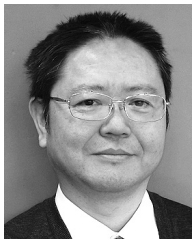
1999 年生．2020 年東京工業高等専門学校情報工学科卒業．電気通信大学 3 年在学．自然言語処理に関する研究に興味を持つ．

富田 陽介

1979 年生．2000 年沼津工業高等専門学校制御情報工学科卒業．2009 年 Banana Systems 株式会社取締役．2013 年コーチ・ユナイテッド株式会社技術責任者．2016 年株式会社ビルディット代表取締役（現任）．2017 年株式会社ウーシア取締役 CTO（現任）．「研修フォローアップアプリケーション Core」ほか，教育・育成にかかわるソフトウェアサービスの開発・運用に従事．

山下 晃弘　（正会員）

1983 年生．2008 年北海道大学大学院情報科学研究科修士課程修了．2010 年同博士課程修了．博士（情報科学）．2011 年株式会社調和技研代表取締役．2013 年東京工業高等専門学校情報工学科助教．2017 年同准教授．組み込みシステム，機械学習，ウェブマイニングに関する研究に従事．情報処理学会，人工知能学会，日本工学教育協会各会員．



松林 勝志　（正会員）

1965 年生．1989 年山梨大学大学院修士課程修了．2005 年同博士課程修了．博士（工学）．1991 年東京高専助手．1999 年 Scotland Dundee University 在外研究員．2007 年東京高専教授．組み込みシステムに関する研究に従事．2013 年関東工学教育協会業績賞．2013 年日本工学協会工学教育賞・文部科学大臣賞．情報処理学会，日本工学教育協会各会員，日本工学教育協会特別教育士（工学・技術）．

© 2021 Information Processing Society of Japan