

Title	自然言語処理による360度評価のフィードバックテキストの情報抽出と評価方法の提案
Sub Title	
Author	横山, 彰士(Yokoyama, Akito) 林, 高樹(Hayashi, Takaki)
Publisher	慶應義塾大学大学院経営管理研究科
Publication year	2020
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2020年度経営学 第3759号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40003001-00002020-3759

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

慶應義塾大学大学院経営管理研究科修士課程

学位論文 (2020 年度)

論文題名

自然言語処理による 360 度評価のフィードバックテキストの情報抽出と評価方法の提案

指導教員	林 高樹
副指導教員	山尾 佐智子
副指導教員	林 洋一郎
副指導教員	岩本 隆

氏 名	横山 彰士
-----	-------

論文要旨

所属ゼミ	林高樹研究会	氏名	横山 彰士
(論文題名) 自然言語処理による 360 度評価のフィードバックテキストの情報抽出と評価方法の提案			
(内容の要旨) 【目的】 本研究は、360 度評価におけるテキスト情報の予測力の評価方法の提案を目的とする。本研究を日本企業の HR 領域における課題、及び非構造化データの活用における課題に対するアプローチとして、HR 領域のテキスト情報（非構造化データ）の応用の提案の一つとしたい。 【研究方法】 本研究の目的達成のための小目的は、①テキストから抽出した情報単体での予測力の評価方法の提案、②数値評価データに現れないテキスト情報の予測力の評価方法の提案、③数値評価データとテキスト情報を同時に使用した予測性能が高い手法の提案である。使用データは、株式会社リンクアンドモチベーション モチベーションエンジニアリング研究所から提供を受けた 360 度評価に関する各種データ（評価結果のスコア、各設問の数値評価データ、フィードバックテキスト）である。 【新規性】 本研究では、人事評価のテキスト情報に対して、ディープラーニング技術をベースとした最新の自然言語処理技術である BERT を採用している点。加えて、人事評価において構造化データと非構造化データを同時に使用した予測モデルを提案している点で新規性があると考えられる。 【結果】 Permutation 検定の結果、自然言語処理により抽出したテキスト情報は、360 度評価結果のスコアに対して有意に予測力を持っていると結論付けられた。追加的情報価値の観点においても、数値評価データのみを使用した回帰モデルの予測性能 (RMSE=4.77, $R^2=0.51$) と、テキスト情報を加えた回帰モデルの予測性能 (RMSE=3.75, $R^2=0.75$) の比較から、有意 ($p<0.01$) に予測性能が向上していることが分かった。非線形モデル（機械学習）を用いた結果は、勾配ブースティング回帰の予測性能 (RMSE=2.78) のみが線形回帰モデルの予測性能 (RMSE=4.07) を越える結果となった。 【考察】 今回の研究で得られた結果から、HR 領域におけるテキスト情報の有効利用の可能性を明らかにし、テキスト解析による情報抽出とその情報価値の評価に至る方法論を提示することができた。 特定企業のデータセットであるため、直ちに一般化されるものではないが、テキスト情報の評価方法や利活用の方法は、一般性を有しており、今後、他企業や他業界内の同種の人事評価データに対しても適用可能なものである。			

目次

1	序論	4
1.1	日本型 HRM における諸問題	4
1.1.1	ICT の活用における課題	4
1.1.2	新任管理職の能力開発における課題	5
1.2	非構造化データの増加と活用	6
1.3	研究目的	7
1.4	先行研究サーベイ	8
2	研究方法	10
2.1	使用データ	10
2.1.1	数値評価データの概要	10
2.1.2	テキストデータの概要	14
2.2	手法	15
2.2.1	テキスト解析手法	15
2.2.2	統計手法	16
2.2.3	機械学習手法	16
3	結果	18
3.1	テキスト解析	18
3.2	テキスト情報の予測力の評価	20
3.3	線形モデルを用いたテキスト情報の追加的情報価値の評価	23
3.4	非線形モデルを用いた予測性能の評価	27
4	考察	29
5	本研究の課題と限界	31
6	参考文献	34
	付録	36
	付録 1 主成分の解釈について	36
	因子分析	36
	主成分のラベリング	39
	付録 2 Model B モデル詳細	42
	付録 3 ソースコード	44

1 序論

1.1 日本型 HRM における諸問題

1.1.1 ICT の活用における課題

「新卒一括採用」「年功序列」「終身雇用」が特徴の日本型人材マネジメントは、第二次世界大戦後の高度経済成長期から日本企業における組織運営や人材マネジメントの基盤となっている。これらの制度は現在でも大企業を中心に一般的に普及している。このことは経済産業省(2019)が発表している「企業の戦略的人事機能の強化に関する調査」の中でも報告されている。日本企業の成長を支えた日本型 HRM(Human Resource Management：人的資源管理)は、2000 年前後を境にして、見直し、改革といった活動が活発化している。

その背景には、ビジネス環境の激変、ICT(Information and Communication Technology：情報通信技術)の発達・進展による業界構造の変化、グローバル化、マクロ経済の不況、労働市場の変化といったさまざまな動きに対し、従来の日本型 HRM の手法では対応できないことが挙げられる。

前述の経済産業省(2019)の調査では、HR(Human Resources：人的資源、人材、人事)領域におけるテクノロジー及びビッグデータの活用について、今後目指すべき方向性が示されている。ビッグデータを活用した人材マネジメント施策立案・実行・モニタリングを実現するためには、データ解析結果に基づく意思決定に対して心理的抵抗を低減していくためのマネジメント経験、成功体験を積み上げていくようなロードマップを描くことが重要であると述べている。しかし、ICT を活用した効果的な人材マネジメントを実現するためには、人事部門の ICT リテラシー、問題解決能力を向上する必要があることが示唆されており、この課題が HR 領域における ICT 活用を推進する上での課題となっている。

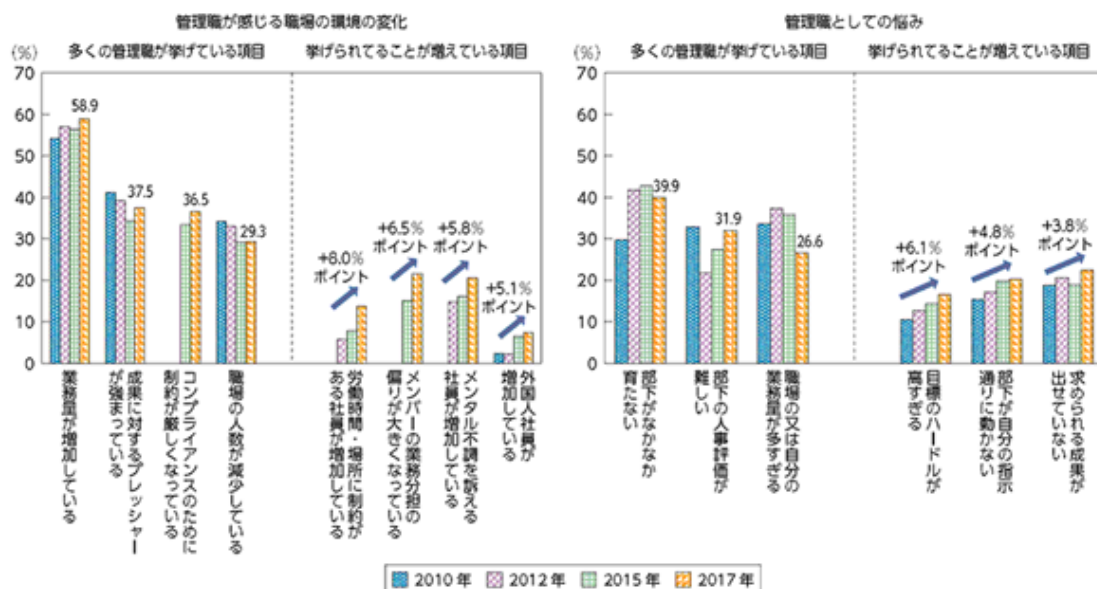


【図 1-1】HR テクノロジーの全体像
(出所：経済産業省(2019)より筆者作成)

1.1.2 新任管理職の能力開発における課題

藤江ら（2014）は、現在の日本企業では、組織のフラット化や外部環境への対応における人員構成の変化にともない、チームマネジメントや部下の指導といった昇格前のマネジャーやリーダーの代理経験が減少する傾向があると述べている。そのため、マネジメント能力が十分に向上する前に昇格してしまうケースや、昇格後も管理職がプレイヤー実務を多く抱えざるを得なくなり、管理職に求められる戦略立案や部下育成がおろそかになるケースなどが増えている。結果として、管理職としての必要な知識・能力が向上しづらい環境になっている。

厚生労働省（2019）は、「平成 30 年度版 労働経済の分析——働き方の多様化に応じた人材育成の在り方について——」の中で、「管理職が感じる職場の環境の変化や管理職としての悩みについて」というトピックにおいて、下図 1-2 の通りの調査結果を報告している。職場環境の変化としては「業務量が増加している」が最も多く挙げられている。本調査が始まった 2010 年から過去最高水準となっている。管理職の悩みとしては、「部下がなかなか育たない」「部下の人事評価が難しい」「職場の又は自分の業務量が多すぎる」を多く挙がっている。



【図 1-2】 管理職が感じる職場の環境の変化や管理職としての悩みについて
 （出所：厚生労働省 平成 30 年度版 労働経済の分析）

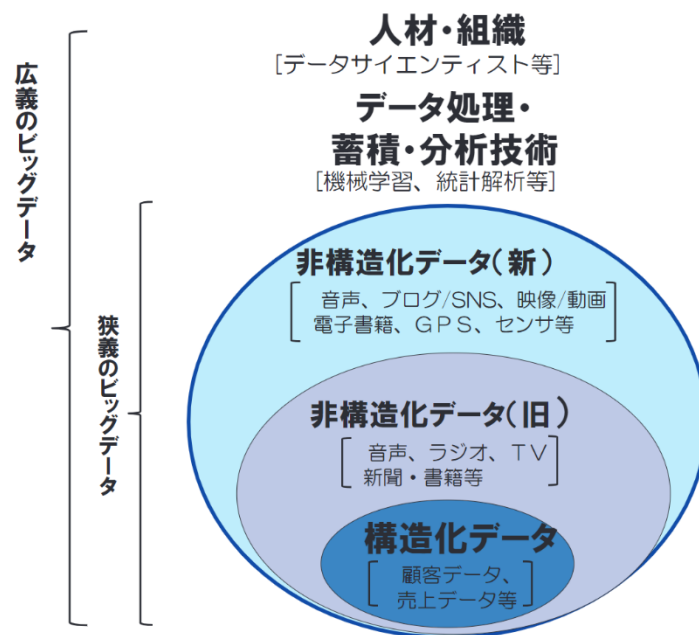
筆者の調査した限りでは、新任管理職の能力開発の課題を取り上げた先行研究は海外には存在しない。このことについては、企業側、労働者側のキャリア開発への考え方が日本企業と欧米企業では異なることが原因と考えられる。山尾（2020）は、東京とカリフォルニア州で人事責任者へのヒアリング調査を行っており、その中で労働市場の特徴と人材育成投資への影響を指摘している。流動性の高い労働市場においては、キャリア入社割合が非常に高く、その場合スキルや経験が多様であるため、企業側はカスタマイズされたプログラムを提供する必要があり、コーチング、上長との 1 on 1 によるキャリア相談、キャリア開発のためのアサインメントへの配属といっ

た、従業員個人に合わせた機会提供を可能にする人事制度設計に注力していると述べている。企業側は従業員のスキルに合わせた配属と成長の機会を提供するという姿勢であるため、先述した日本企業の新任管理職における課題のように、従業員のスキルにマッチしない昇格や、統一的な方法でマネジメント能力向上を図るような思考にはならないと考えられる。

マネジャーにとって必要なリーダーシップ・コンピテンシーの強化については、Zenger ら(2011)が、現役のリーダー約3万人の360度評価データ25万件を分析しており、昇進を見送られたマネジャーがリーダーシップの改善などの事例を挙げている。また同論文では、リーダーからマネジャーへの昇格における重要なリーダーシップ・コンピテンシーの強化には、360度評価の有用であることも示唆している。

1.2 非構造化データの増加と活用

総務省(2013)は「平成25年版 情報通信白書」の中で、ビッグデータ概念として下図1-3の通り、整理している。



【図 1-3】 ビッグデータ概念
(出所：総務省(2013)より引用)

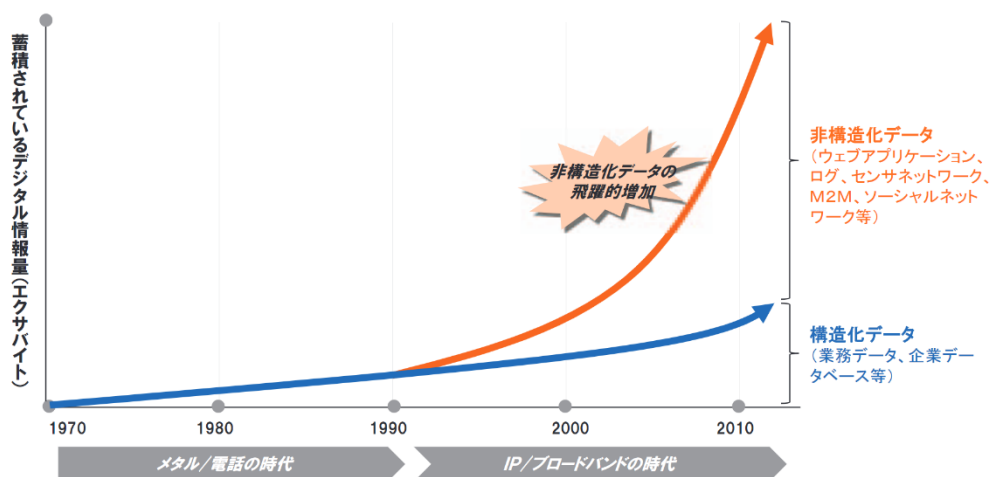
構造化データは、POS データや顧客データ、人事評価スコアなど、列と行からなるフォーマットで管理が可能であり、データベースやシステム間インターフェイスとして活用可能なデータのことを指す。非構造化データは、自由記述形式のテキスト、音声や画像といった構造定義されていないデータを指し、構造化データと取り扱い方法や管理方法も異なる。その特徴から専門的なツールや手法を使用しなければ理解できない、もしくは分析できないため、企業の多くは大量の非構造化データを持ちながらも、その価値や重要性を見出すことができず、限定的な利活用の方法に留まっているか、利用目的がないまま保管している。

HR 領域における非構造化データの活用は、従業員へのテキスト形式のアンケート・業務日報のテキスト分析による離職防止やリテンション施策や、ハイパーフォーマーの従業員の顧客とのコミュニケーションを分析することで行動特性のパターン化に活用している。HR 領域においても非構造化データであるテキストを分析することが有用である一方で、経済産業省(2019)で挙げた課題から限定的な利用に留まっている状況にある。

そして、近年、非構造化データは飛躍的な増加が見られ、今後もその傾向はさらに強くなると予想されている(図1-4)。理由としては、電子メールの浸透、Office やアドビ社の製品の普及など、コミュニケーションツールとしてのデータ活用が挙げられる。また、e-文書法などの法規制により、紙媒体から電子媒体への転換が推進されたことも、増加の理由として挙げられる。

加えて、新型コロナウイルスの感染拡大による2020年4月の緊急事態宣言以降、テレワークの導入が急激に進んだ。これまで口頭で指示やフィードバックされていた内容は、ビデオ通話やメール、チャットといったコミュニケーションツールを介在した形で伝達されるようになった。コミュニケーションツールの変化について、筆者の調べた限り定量的な大規模調査はまだ確認できないが、テレワーク特有のハラスメントのリスクとして「リモートハラスメント(リモハラ)」「テレワークハラスメント(テレハラ)」などという言葉が使われ始めている(児玉(2020))。

上司から部下への指示やフィードバックを含めた対面でのコミュニケーションが、コミュニケーションツールを通して、テキスト・音声・映像といった非構造化データに置き換えられるため、今後さらなるデータ増加が見込まれる。企業としてはそれらを埋没させることなく、前述のリスクへの対処やHR 領域での利活用を進めていくべきであると考えられる。



【図 1-4】構造化データと非構造化データの増加スピードの違い

(出所：総務省(2013)より引用)

1.3 研究目的

本研究は、360 度評価におけるテキスト情報の予測力の評価方法の提案を目的とする。本研究を日本企業の HR 領域における課題、及び非構造化データの活用における課題解決に対するアプローチとして、HR 領域のテキスト情報(非構造化データ)の利活用の方法論の一つとして提案したい。

360 度評価は実務で最も多く採用されている人事評価方法の一つである。この評価方法の利点は、同僚から広く判断材料を集めることによって、納得感のある公正な評価ができる点であり、360 度評価結果のスコアは Zenger ら (2011) が指摘するように、リーダーシップの開発において重要な指標となる。企業の人材開発部門は、従業員の 360 度評価結果のスコアを定点観測の指標として管理・分析し、スコアの高い従業員の行動特性を明らかにすることや、低いまま留まっている従業員を発見し早期に対応することが求められる。

しかし、360 度評価の問題点として、評価結果に対して評価者の性格特性や自己防衛による過大評価または過小評価といったバイアスがかかってしまう点が挙げられる (Bernardin ら (2015)、Zenger ら (2011))。この問題の解決案として、兵藤 (2015) は、360 度評価において被評価者に対して評価者が強みと弱みについてコメントしたテキストを使用した人事評価方法を提案している。テキスト情報には評価者が被評価者に対して抱いている感情や課題意識が含まれており、数値評価の情報には表れない情報を抽出することができることを示唆している。一方で、単語ベースでの極性判断を行うテキストマイニング手法の制度面での限界も示唆されている。

上記より、本研究の目的の達成のため、360 度評価における非構造化データである「フィードバックのテキスト情報」の「360 度評価結果のスコア」に対する予測力を評価する。より具体的には、360 度評価を構成する構造化データである数値評価データと、「フィードバックのテキスト情報」を、「360 度評価結果のスコア」の予測に同時使用した場合における予測性能の向上の有無を検証し、「フィードバックのテキスト情報」の情報価値を評価する。

1.4 先行研究サーベイ

人事評価におけるテキストマイニングを用いた研究事例及び、フィードバックのテキスト情報を使った研究事例について調査を行った。また、人事評価において予測モデルの構築に構造化データと非構造化データを同時に使用した研究については、筆者の調べた限り存在しなかった。ここでは、企業業績の数値情報と CSR レポートのテキスト情報を同時使用した予測モデルの構築の研究事例を紹介する。

・テキストマイニングを用いた 360 度評価における研究事例

兵藤 (2015) は、人材評価の方法として従来より採用されている数値評価をベースとした 360 度評価ではなく、テキストデータのテキストマイニングによる 360 度評価を新たに提案している。評価者が被評価者の強みと弱みについてコメントしたテキストを扱っており、従来の 360 度評価における問題点であるバイアスの発生への対処として、コメントした内容は、評価者と被評価者が複数回やりとりしてお互いが納得するバイアスの生じない評価結果を導く工夫がされている。テキストマイニングの方法としては、単語の極性辞書を利用してコメントごとに得点を付与し、総得点を人事評価としている。テキスト情報による人事評価と従来のアンケートデータによる人事評価を比較した結果、双方に弱い相関があったものの、完全に一致するということはなく、テキストマイニングによる人事評価は「職場従業員間の評価関係」を表す方法として有効であるこ

とを示唆している。課題として、単語ベースでの極性分析には精度向上の限界があること、加えて、単語の意味解釈のために手作業で辞書に単語を追加しており、2015年時点での自然言語処理の技術では、精度向上及び改善は困難かつ複雑な作業を求められていた。

この点については、現在の最先端の自然言語処理技術で解析できれば、単語だけでなく文脈の意味を捉えることができ、テキストの持つ情報の損失を抑えながら、360度評価結果のスコアに対して説明力を持つ情報を抽出できると考える。

・フィードバックのテキスト情報の研究事例

Abe1ら(2020)は、上司からのフィードバックのテキスト情報が従業員の行動・モチベーションに与える影響についての実証実験を行っている。インドの1,800人のギグエコノミーワーカーを雇用し、架空のマネジャーの性別を用意した上で、労働者のパフォーマンスへのフィードバックをランダムに配信した。フィードバックには、ポジティブフィードバック、ネガティブフィードバックが用意されており、ワーカーのパフォーマンスに関わらず、それらのフィードバックがランダムに配信された。実験結果として、ネガティブフィードバックは、ワーカーの態度を中程度に悪化させる一方で、義務的・自発的な仕事の両方で、仕事への努力の提供を増加させることがわかった。対照的に、ポジティブフィードバックは、態度にも努力の提供にも影響を与えなかった。この結果から、フィードバックそのものの性質として、フィードバックはその受容後の行動に影響を与え、フィードバックの内容によって受容した人間の行動の性質が変化させる。また、この実験は、対面でのフィードバックではなく配信という形を取っているため、見えないマネジャーからのネガティブフィードバックによる態度の悪化とパフォーマンスの向上は直感に相違ない。対面で業務を行う現場では、上司と部下の関係性の構築とポジティブとネガティブを含んだフィードバックを行うことで、部下の態度の悪化を抑止しつつ、パフォーマンス向上に向けた建設的なフィードバックが行うことができると考える。

上記のような、ポジティブ・ネガティブといったフィードバックやテキストの極性に焦点を当てた先行研究は存在するが、上司からのフィードバックの質的な面の検証を行っている先行研究は筆者の調べた限り存在しなかった。質的な面での検証は、上司の性格や伝え方(テキストであれば書き方)、使用する語彙や文章量といったテキスト情報特有の制約が多くあり、分析が困難である一方、数値データや極性の情報には含まれないニュアンスや語彙の使用の傾向などが表れる。近年、自然言語処理技術の向上により大量のデータに対して文脈を考慮した上での定量化が可能となり、フィードバックテキストから人材マネジメントに寄与する新たな情報を抽出できる可能性がある。

・構造化データと非構造化データを予測に同時利用した研究事例

Leeら(2020)は、企業の経営成績と企業の社会的責任(CSR)の関係を考慮した上で、構造化データである財務情報と非構造化データであるCSRレポートのテキストを同時に使用することで、業績予測のための意思決定フレームワークを提案している。CSRレポートの多面的側面に対して、単語出現の分布を元にテキストのトピックを特定する潜在的ディリクレ配分法(LDA)を使って、各企業のCSRの方針を特定し、CSR方針と財務情報との関係性を明らかにした上で、CSR関連情報

を含んだ線形回帰モデルと含んでいないモデルに分けて業績予測精度を比較した。予測精度はわずかな改善であったが、これまで業績予測に対して数値情報と同時に考慮されていなかった CSR レポートのテキストを分析可能な形に定量化し、ハイブリッドなアプローチとして提案されたモデルは、企業の業績予測の有望な代替手段であることを示唆している。本事例は、非構造化データであるテキスト情報を、構造化データと併用した形で予測に用いることで、テキスト情報に含まれる予測力を検証できる可能性を示唆している。

先行研究レビューより、本研究は以下の2点で新規性があると考えられる。

一つ目は、HR 領域、特に 360 度評価のテキスト情報に対して、極性辞書や LDA といった単語ベースのテキストマイニング手法ではなく、文脈まで捉えることができる最新の自然言語処理技術を採用することである。

二つ目は、360 度評価の数値評価データ（構造化データ）とフィードバックのテキスト情報（非構造化データ）を同時に使用した 360 度評価結果のスコアの予測モデルを検討することで、テキスト情報の予測力の評価、及び人材開発の現場で十分に活用されていないテキスト情報の応用方法を提案することである。

2 研究方法

研究目的の達成のため、以下の小目的を設定する。

- ① テキストから抽出した情報単体での予測力の評価方法の提案
- ② 数値評価データに現れないテキスト情報の予測力の評価方法の提案
- ③ 数値評価データとテキスト情報を同時に使用した予測性能が高い手法の提案

2.1 使用データ

本研究に使用するデータは株式会社リンクアンドモチベーション（以下、LMI）から提供を受けたデータである。分析対象は、2018 年 9 月～2020 年 5 月に 360 度評価及びフィードバックを実施した従業員のうち、研修前後の 360 度評価結果のスコアが確認可能な従業員、1040 名とした。

なお、以下では、当該データの仕様や内容に沿って、研究方法を具体的に述べていくが、方法自体は一般的なものであり、他のデータセットにも適用が可能である。

2.1.1 数値評価データの概要

- ・ 360 度評価

LMI の 360 度評価では、情報提供、情報収集、判断行動、支援行動の 4 つの機能に沿って設問を構成されている。

【表 2-1】 360 度評価の設問の分類と内容

情報提供	戦略情報の提供	自部署とそのメンバーをとりまく環境（業界動向、顧客のニーズ、競合の動向、会社方針等）に関する情報を把握し、メンバーに情報を提供する。
情報提供	役割情報の提供	メンバーが業務を遂行するにあたって必要となる情報（自部署の使命・目標等）を提供すると共に、役割や責任範囲を明確にし、具体策を示す。
情報収集	役割情報の収集	自部署内の連携状況、他部署との連携状況を把握すると共に、メンバー個々人の持ち味や要望をつかむ。
情報収集	進捗状況の収集	自部署の業務の進捗状況を把握し、個々人の成果を確認すると共に、トラブルの発生や個々人の課題を把握する。
判断行動	基準提示	メンバーに求める行動や考え方のポイント、業務に関する評価基準を明確に提示する。
	評価行動	提示した基準に従って納得感のある評価を行うとともに、メンバーには公平に接する。
支援行動	人材開発と業務支援	部下にノウハウの伝授、チャレンジする機会の提供を行い、部下のコンディションを気にかかけ、壁にぶつかった際には適切にサポートする。
	支援行動と意欲喚起	業務の背景や意義の伝達、部下へ適切な権限委譲を行い、またオープンに部下の意見を募り、優れた意見や企画を取り入れる。

（出所：株式会社リンクアンドモチベーション HP¹）

機能ごとに 10 項目の設問（計 40 項目）が用意されており、各設問を 5 段階（1～5）で評価する。評価者は同僚である上司・部下である。評価者は被評価者への期待度と満足度について、それぞれ同一の設問に対して評価を行うため、合計で 80 項目の設問に回答する。

満足度は、現状のスキルや姿勢についての評価であり、各設問に該当する課題意識が大きいほど最大評価 5 から減点されていき、伸ばしてほしい長所や克服してほしい短所を強調させるようなスコアの付け方がされる。期待度は、役割上期待される項目については値が高くつけられ、役割上必要ない項目や、あえて低い値をつけることで高い値を目立たせて強調するようなスコアの付け方がされる。期待度と満足度は同時に測定されているため、双方には関連性がない。設問に対する数値評価は、被評価者に公開される情報であり、360 度評価におけるバイアスが生じる可能性がある。

360 度評価結果のスコアは、期待度と満足度から LMI 独自のモデルにより算出されており、偏差値に基づきスコアが付けられる。本研究では、将来の 360 度評価結果のスコアの予測のため、予測対象となる 360 度評価結果のスコアと、予測対象の 2 四半期前の 360 度評価結果のスコアを使用する。

2 種類の 360 度評価結果のスコア、上司からの期待度 40 項目、上司からの満足度 40 項目、部下からの期待度 40 項目の平均値、部下からの満足度 40 項目の平均値の統計量について、表 2-2 に記載する。なお、各設問の内容については、LMI の提供サービスの一部であるため、非公開となる。

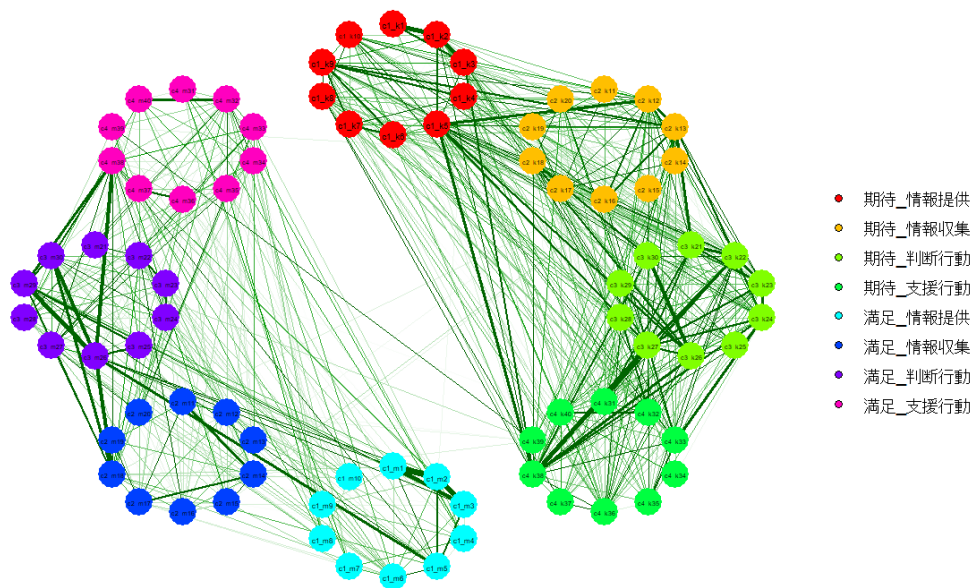
¹ https://solution.lmi.ne.jp/hr_development/c/middle_management/lincoln_feedback

【表 2-2】各数値の要約統計量

項目名	機能	平均値	標準偏差	中央値	最小値	最大値
予測対象の360度評価スコア	-	60.20	6.68	61.25	29.9	73.2
2 四半期前の360度評価スコア	-	60.47	6.76	61.3	27.6	73.7
部下からの期待度平均	-	4.31	0.57	4.4	0	5
部下からの満足度平均	-	4.04	0.59	4.135	0	4.88
上司からの期待度項目1	情報提供	4.30	0.64	4	2	5
上司からの期待度項目2	情報提供	4.37	0.59	4	2	5
上司からの期待度項目3	情報提供	4.51	0.55	5	2	5
上司からの期待度項目4	情報提供	4.49	0.59	5	2.5	5
上司からの期待度項目5	情報提供	4.57	0.53	5	3	5
上司からの期待度項目6	情報提供	4.52	0.57	5	2	5
上司からの期待度項目7	情報提供	4.46	0.58	4.5	3	5
上司からの期待度項目8	情報提供	4.46	0.55	4.5	3	5
上司からの期待度項目9	情報提供	4.48	0.58	4.6	2.5	5
上司からの期待度項目10	情報提供	4.42	0.60	4.5	2	5
上司からの期待度項目11	情報収集	4.23	0.62	4	2	5
上司からの期待度項目12	情報収集	4.60	0.51	5	3	5
上司からの期待度項目13	情報収集	4.55	0.57	5	2	5
上司からの期待度項目14	情報収集	4.33	0.66	4	2	5
上司からの期待度項目15	情報収集	4.21	0.68	4	2	5
上司からの期待度項目16	情報収集	4.49	0.55	4.5	2.5	5
上司からの期待度項目17	情報収集	4.56	0.54	5	3	5
上司からの期待度項目18	情報収集	4.55	0.53	5	3	5
上司からの期待度項目19	情報収集	4.60	0.53	5	3	5
上司からの期待度項目20	情報収集	4.44	0.55	4.5	3	5
上司からの期待度項目21	判断行動	4.53	0.57	5	2.5	5
上司からの期待度項目22	判断行動	4.68	0.48	5	3	5
上司からの期待度項目23	判断行動	4.60	0.51	5	3	5
上司からの期待度項目24	判断行動	4.71	0.47	5	3	5
上司からの期待度項目25	判断行動	4.47	0.58	4.5	2.5	5
上司からの期待度項目26	判断行動	4.60	0.53	5	3	5
上司からの期待度項目27	判断行動	4.68	0.49	5	3	5
上司からの期待度項目28	判断行動	4.54	0.54	5	3	5
上司からの期待度項目29	判断行動	4.45	0.59	4.5	2	5
上司からの期待度項目30	判断行動	4.67	0.50	5	3	5
上司からの期待度項目31	支援行動	4.59	0.54	5	3	5
上司からの期待度項目32	支援行動	4.55	0.55	5	2	5
上司からの期待度項目33	支援行動	4.67	0.49	5	3	5
上司からの期待度項目34	支援行動	4.50	0.60	5	2	5
上司からの期待度項目35	支援行動	4.69	0.48	5	3	5
上司からの期待度項目36	支援行動	4.70	0.46	5	3	5
上司からの期待度項目37	支援行動	4.47	0.55	4.5	2	5
上司からの期待度項目38	支援行動	4.64	0.51	5	3	5
上司からの期待度項目39	支援行動	4.11	0.84	4	1	5
上司からの期待度項目40	支援行動	4.65	0.51	5	3	5

項目名	機能	平均値	標準偏差	中央値	最小値	最大値
上司からの満足度項目1	情報提供	4.09	0.72	4	2	5
上司からの満足度項目2	情報提供	4.25	0.70	4	2	5
上司からの満足度項目3	情報提供	4.37	0.70	4.5	2	5
上司からの満足度項目4	情報提供	3.96	0.74	4	2	5
上司からの満足度項目5	情報提供	4.15	0.76	4	2	5
上司からの満足度項目6	情報提供	4.31	0.70	4	2	5
上司からの満足度項目7	情報提供	4.12	0.70	4	2	5
上司からの満足度項目8	情報提供	4.39	0.68	4.5	1	5
上司からの満足度項目9	情報提供	4.14	0.65	4	2	5
上司からの満足度項目10	情報提供	4.02	0.75	4	1	5
上司からの満足度項目11	情報収集	3.90	0.71	4	2	5
上司からの満足度項目12	情報収集	4.07	0.66	4	2	5
上司からの満足度項目13	情報収集	4.27	0.64	4	2	5
上司からの満足度項目14	情報収集	3.91	0.70	4	2	5
上司からの満足度項目15	情報収集	4.12	0.66	4	2	5
上司からの満足度項目16	情報収集	4.16	0.62	4	2	5
上司からの満足度項目17	情報収集	4.06	0.68	4	2	5
上司からの満足度項目18	情報収集	4.23	0.69	4	0	5
上司からの満足度項目19	情報収集	4.01	0.71	4	2	5
上司からの満足度項目20	情報収集	4.45	0.62	4.5	1	5
上司からの満足度項目21	判断行動	4.20	0.68	4	2	5
上司からの満足度項目22	判断行動	4.42	0.61	4.5	2	5
上司からの満足度項目23	判断行動	4.21	0.64	4	2	5
上司からの満足度項目24	判断行動	3.97	0.72	4	1	5
上司からの満足度項目25	判断行動	4.28	0.71	4	2	5
上司からの満足度項目26	判断行動	4.18	0.75	4	1	5
上司からの満足度項目27	判断行動	4.59	0.59	5	2	5
上司からの満足度項目28	判断行動	4.20	0.73	4	1	5
上司からの満足度項目29	判断行動	4.14	0.74	4	1	5
上司からの満足度項目30	判断行動	4.21	0.74	4	2	5
上司からの満足度項目31	支援行動	4.17	0.75	4	1	5
上司からの満足度項目32	支援行動	4.27	0.66	4	2	5
上司からの満足度項目33	支援行動	4.54	0.59	5	2	5
上司からの満足度項目34	支援行動	4.19	0.72	4	1	5
上司からの満足度項目35	支援行動	4.26	0.76	4	1	5
上司からの満足度項目36	支援行動	4.30	0.70	4	2	5
上司からの満足度項目37	支援行動	4.37	0.66	4.5	2	5
上司からの満足度項目38	支援行動	4.38	0.69	4.5	1	5
上司からの満足度項目39	支援行動	4.16	0.71	4	2	5
上司からの満足度項目40	支援行動	4.28	0.68	4	2	5

図 2-2 において、期待度 40 項目、満足度 40 項目の各項目の相関をネットワーク図で可視化した。期待度・満足度間には相関がないもしくは弱いことが分かる。また、4つの機能を越えた相関が見られたが、現場業務においては複合的な能力が求められることに起因すると考えられる。



【図 2-2】 上司からの期待度・満足度の相関のネットワーク図

2.1.2 テキストデータの概要

LMI の提供サービスである研修事業において、研修に参加する従業員の上司から研修講師に向けて、以下の3つの質問についてテキストで回答し、事前に提出する。

- ・ 役割：あなたの部下が現在担っているミッション・役割は何ですか？
- ・ 課題：本人に克服してもらいたいところや伸ばしてもらいたいところはどこですか？
- ・ 期待：あなたの部下には今後どのように成長してもらいたいですか？

研修講師は、この回答をもとに研修に参加する従業員への指導方法を考え、個人に合わせた研修プログラムを実施する。

以下は、コメントの例である。各質問への回答テキストから筆者が作成した。

- ・ 役割：あなたの部下が現在担っているミッション・役割は何ですか？
 - (例1) 中小企業向けのコンサルティング事業の営業リーダー。2名のメンバーをもちながらマネジメントタスクをこなす。
 - (例2) 納品支援の部隊のリーダー若手メンバーや新人の成長支援。パートナー社員のマネジメントを担当。
- ・ 課題：その中で本人に克服してもらいたいところや伸ばしてもらいたいところはどこですか？
 - (例3) 強みはコミュニケーション力である。相手を盛り上げるコミュニケーション力・協調性があり、初対面でも仲良くなることができる。課題としては、外部情報の収集である。新技術を探求し、習得してほしい。コミュニティ形成の新たな技術を自ら策定し、実践してほしい。

・期待：あなたの部下には今後どのように成長してもらいたいですか？

(例4) 世界一のコミュニティナビゲーターとなる。イベントに登壇すると、素晴らしいナビゲーションで場を盛り上げられる。

(例5) あやふやなまま受け入れ行動に移すことができる試行力は強み。調和関係は自信をもってもらってよい。一つの事象に捉われすぎてしまい、解決するまでの曖昧力が課題。そのスキルを伸ばして成長して欲しい。

本研究では、先行研究の事例から、テキスト情報にある長所と短所に関する記載、またはポジティブフィードバック、ネガティブフィードバックが従業員の行動や評価に与える影響を考慮するため、「課題」に関する回答テキストを使用する。この回答テキストは被評価者である従業員には非公開であり、360度評価の問題点であるバイアスを排除した情報として扱うことができる。

2.2 手法

本研究で使用する各種手法について紹介する。

2.2.1 テキスト解析手法

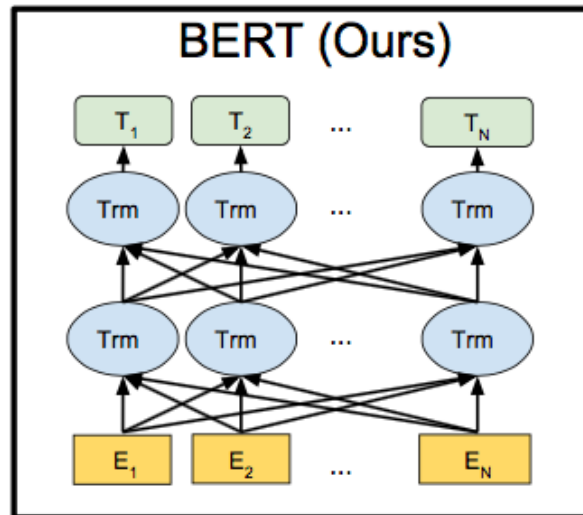
テキストの特徴を数値表現に変換するため、自然言語処理モデル「BERT」を使用する。BERTは、2018年10月にGoogleのDevlinら(2018)により発表された自然言語処理モデルである。

BERTの特徴として「文脈を読むことが可能になった」ことを挙げられる。BERTにはTransformerというアーキテクチャ(構造)が組み込まれており、MLM(Masked Language Modeling: 複数箇所が空白になっている文章の単語予測)とNSP(Next Sentence Prediction: 2文が連続した文か判定)を学習させることで、単語レベル・文章レベルで双方向から学習でき、「文脈を読むこと」が実現した(図2-2)。

もう一つの特徴は、ファインチューニングが可能であることが挙げられる。教師あり学習のタスクである感情分析のような2値分類や、トピック分類といったタスクに対して、BERTの出力部分を変えるだけでさまざまなタスクに応用可能である点である。

本研究では、東北大学 乾・鈴木研究室が公開しているWikipedia日本語ページを事前学習済みのBERTのモデル²を使用し、事前学習済みモデルが抽出する768次元の特徴量を分析に使用する。

² <https://github.com/cl-tohoku/bert-japanese>



【図 2-3】 BERT モデル概念図
(出所：Devlin ら(2018)より引用)

2.2.2 統計手法

- ・線形回帰

ある変数の値（目的変数）を、別の変数（説明変数）の値に基づいて予測するために使用される。目的変数の値を最適に予測する 1 つ以上の説明変数を使用して、線形方程式の係数が推定される。本研究の線形回帰モデルにおける係数の推定には、最小二乗法を使用する。

- ・Permutation 検定（並び替え検定）

ある指標を定め、標本データからその指標を計算し、その後標本データのサンプルを非復元抽出によってランダムに並び替え、その指標を繰り返し計算し、その指標の値の分布（帰無分布）を得る。標本データから計算された指標値が、その帰無分布においてどのくらい極端な値か（分布のテイル確率）を評価することで、P 値（経験 P 値, empirical p-value）を得る。

機械学習分野では、分類器の予測性能の評価に対して Permutation 検定を応用する試みがなされている (Ojala ら(2010), Golland ら(2003), Molinaro ら(2005))

- ・Cross Validation (交差検証)

汎化性能を評価する統計的な手法である。本研究では、10-fold Cross Validation を扱う。データセットを 10 個のサブセットに分割して、そのうち、8 個のサブセットで学習し、残りの 2 個のサブセットで検証するという作業をすべての組み合わせに対して行う検証方法である。データの分割方法に依存しない、解析結果の妥当性の検証・確認を行う手法である。

2.2.3 機械学習手法

将来の 360 度評価結果のスコアの予測モデルとして、数値評価データとテキスト情報を同時に

使用する非線形回帰モデルを構築するにあたり、以下の機械学習手法を用いる。

- ・多層パーセプトロン（ニューラルネットワーク）

多層パーセプトロンは、順伝播型ニューラルネットワークの一つであり、少なくとも入力層、隠れ層、出力層の3つのノードの層からなります。入力層のノード以外のノードは非線形活性化関数を使用するニューロンであり、誤差逆伝播法（バックプロパゲーション）と呼ばれる教師あり学習手法を利用することで、線形分離が不可能なデータを識別できる。

- ・ランダムフォレスト回帰

ランダムフォレストは、複数の弱学習器を組み合わせるアンサンブル学習の手法の一つ。決定木を複数作った上で、結果を多数決・平均化することによって、決定木における過学習を平準化する手法である。

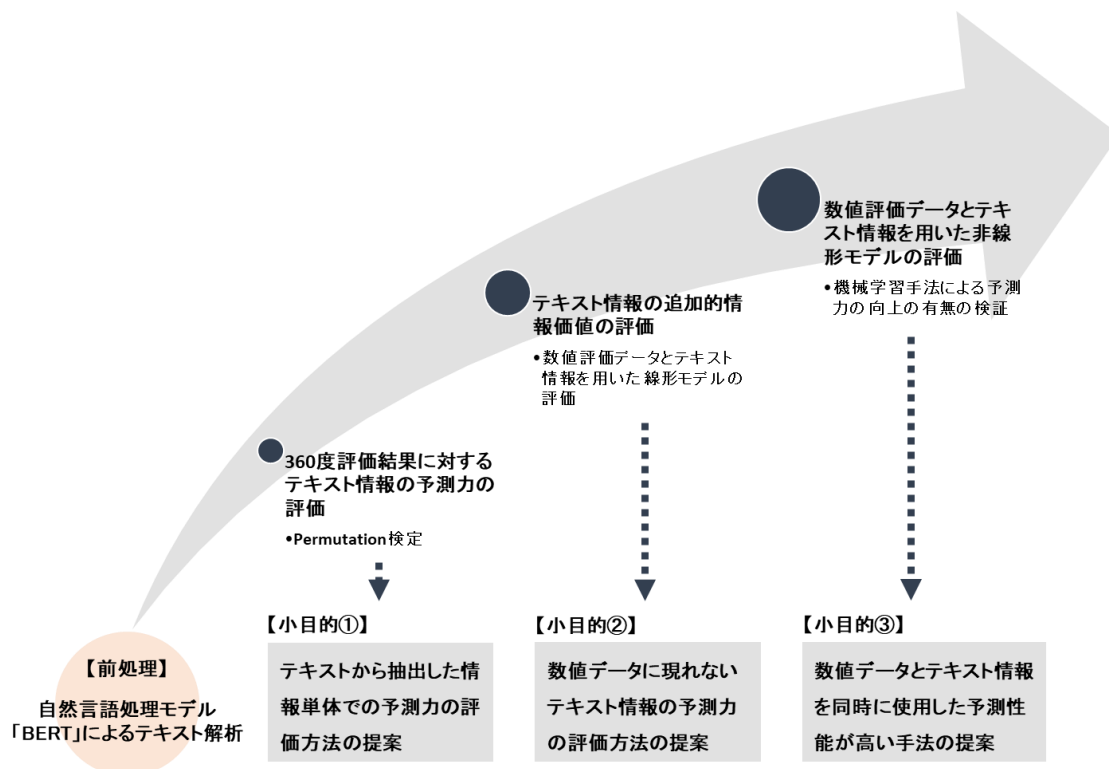
- ・勾配ブースティング回帰

勾配ブースティングは、アンサンブル学習の手法の一つであり、逐次的に弱学習器を構築していく手法である。弱学習器を1つずつ順番に構築していき、新しい弱学習器を構築する際に、それまでの弱学習器の結果を利用して損失関数を最小化していく。

3 結果

本研究にて提案する 360 度評価のフィードバックテキスト情報の評価の方法論について、以下にプロセス、ツールを挙げる。

- ① 前処理として、フィードバックのテキスト情報を最新の自然言語処理モデル BERT を用いて解析し、各テキストから特徴量を抽出する。高次元かつ個々の特徴量の意味解釈が困難であるため、主成分分析により次元圧縮した主成分得点も用意する。
- ② 360 度評価結果のスコアに対する①のテキスト情報の予測力を評価する。Permutation 検定と交差検証を用いて予測力を検証する。
- ③ 線形モデルにおけるテキスト情報による追加的情報価値を評価する。数値評価データのみを説明変数に含む線形回帰モデルと、テキスト情報の特徴量を加えた線形回帰モデルの予測性能を比較する。もしテキスト情報の持つ予測力が数値評価データに包含されている場合は有意に差が出ないことが予想される。
- ④ ③の結果において、線形性の制約がテキスト情報の持つ価値をうまく引き出せていない可能性もあるため、③の結果をベンチマークとしつつ、非線形構造を有する複数の機械学習手法の予測性能を評価する。



【図 3-1】 提案方法論の概要

3.1 テキスト解析

自然言語処理モデル BERT を使用して、フィードバックのテキスト情報から分散表現ベクトル (768 次元の特徴量) を抽出する。768 次元の特徴量は、個々の特徴量の意味解釈が困難であるた

め、主成分分析により次元圧縮した主成分得点も計算し、以降の分析に使用する。

768次元という高次元のデータセットの主成分分析については、青嶋ら(2013)は、「高次元データは豊富な情報を有するものの、それが巨大なノイズに埋もれ見つけ難いことである、これらの理由から、通常の変量解析法では高次元データの推測に精度を保証することができず、間違っ了解析結果に導くことさえある。」と述べており、高次元の特徴量を持つデータセットに対して従来の主成分分析を用いることによる問題点を指摘している。本研究では、テキスト情報にある長所と短所に関する記載、ポジティブフィードバック、ネガティブフィードバックが混在する「課題」に関する回答テキストを使用しているため、文書内で取り上げられるトピックや使用される単語は多岐にわたり、また文書量も千差万別で、ポジティブ/ネガティブ表現も混在している。そのため、768次元のデータについて、スパース性や外れ値の存在を前提とした分析が必要である。本研究では、従来の主成分分析ではなく、Crouxaら(2007)により提案された、スパース性への対応及び外れ値にロバスト性を持つ主成分分析手法のアルゴリズムを適用した。

Rパッケージ「pcaPP」のPCAGrid³関数及びsPCAGrid関数を用いる。PCAGridは、平面内のグリッド検索アルゴリズムを使用して、外れ値にロバスト性を持つ主成分を計算した。また、sPCAGridは、上記に加え、スパースネス制約を適用した主成分を計算した。

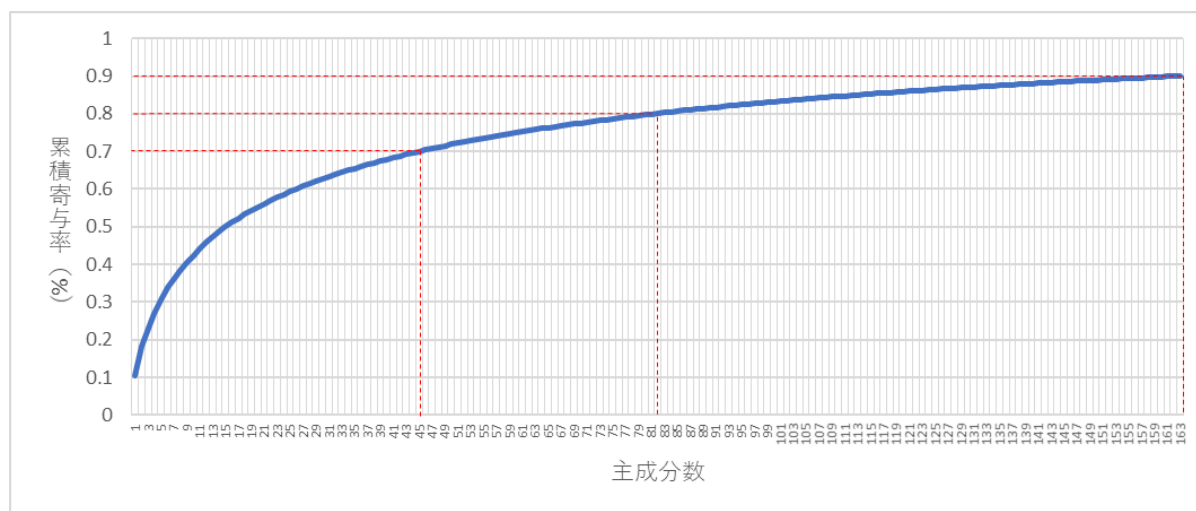
表3-1のとおり、主成分分析の結果、第15主成分までで、累積寄与率が50%となり、第45主成分で70%、第82主成分で80%、第163主成分で90%となった。

【表3-1】主成分分析の抽出結果

	標準偏差	寄与率	累積寄与率
Comp1	1.723	10.49%	10.49%
Comp2	1.478	7.71%	18.20%
Comp3	1.171	4.85%	23.05%
Comp4	1.049	3.89%	26.94%
Comp5	1.012	3.62%	30.56%
Comp6	0.953	3.21%	33.76%
Comp7	0.830	2.44%	36.20%
Comp8	0.790	2.20%	38.40%
Comp9	0.749	1.98%	40.38%
Comp10	0.745	1.96%	42.34%
Comp11	0.707	1.77%	44.11%
Comp12	0.682	1.64%	45.75%
Comp13	0.660	1.54%	47.29%
Comp14	0.643	1.46%	48.75%
Comp15	0.597	1.26%	50.01%
⋮	⋮	⋮	⋮
Comp45	0.335	0.39%	70.02%
⋮	⋮	⋮	⋮
Comp82	0.235	0.19%	80.10%
⋮	⋮	⋮	⋮
Comp163	0.149	0.07%	90.03%

³ <https://www.rdocumentation.org/packages/pcaPP/versions/1.9-73/topics/PCAGrid>

【表 3-2】 累積寄与率のグラフ



3.2 テキスト情報の予測力の評価

将来の 360 度評価結果のスコアに対して、テキスト情報が予測力を持っているかを調べる。テキスト情報が予測力を持たない場合、 i 番目のデータの 360 度評価結果のスコアに対して、 i 番目のテキスト情報による予測結果と、別の (j 番目の) テキスト情報による予測結果には、有意な差は生まれないと考えられる。テキスト情報が予測力を持っているかを判断するため、Permutation 検定を行う。以下、検定の手順について記載し、概要については図 3-1 に図示した。

1040 件のデータセットを訓練用データと検証用データに分割して予測性能を算出するが、予測性能がデータの分割の仕方に依存する可能性があるため、K-fold Cross Validation を用いる。K-fold Cross Validation では、データセットを K 分割して、「8 割を訓練用データセット」「2 割を検証用データセット」とする。下記の①～④の作業について、全てのデータが必ず一度は訓練用データセット、検証用データセットとして利用されるように使用データを変えながら K 回繰り返す。Cross Validation によって得られた K 個の p 値に対し、フィッシャーの統合確率検定を用いた統合 P 値を算出することで、データの分割の仕方に依存しない予測性能を検証することができる。統合 P 値の計算には、scipy ライブラリ⁴を使うことにする。

- ① 将来の 360 度評価結果のスコアを被説明変数として、テキスト情報を説明変数とする線形回帰モデルを構築し、得られた RMSE を RMSE(0) とする。線形回帰モデルの構築においては、Stepwise 法による変数選択を行う。
- ② 次に、Permutation 検定（並び替え検定）を行う。テキスト情報を非復元抽出で並び替える、つまり、360 度評価結果のスコアの並び順を固定したまま、テキスト情報（説明変数）を並び替えることで、説明変数と被説明変数がランダムな組み合わせとなる。①と同様に線形回帰モ

4

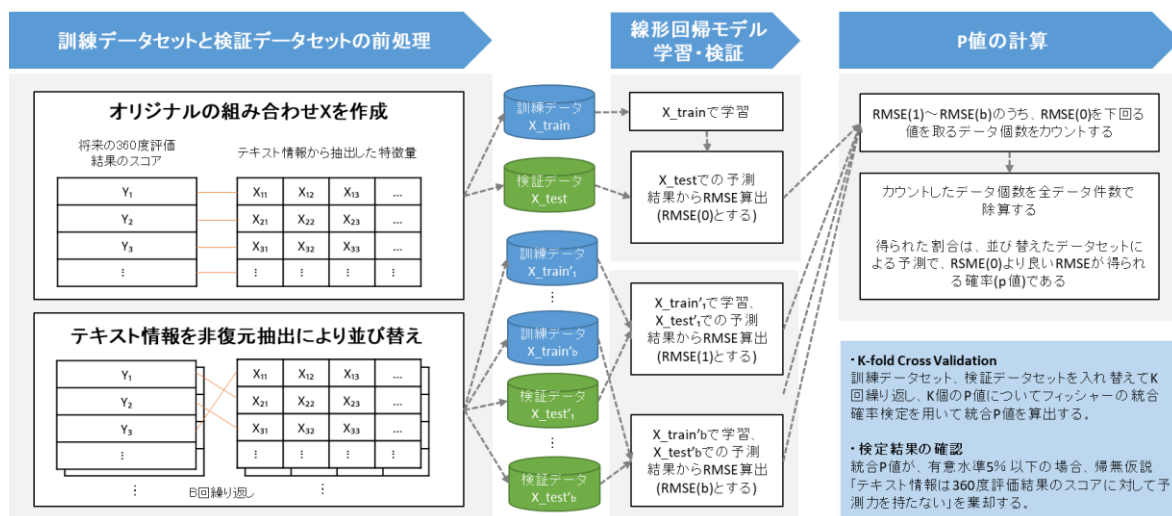
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.combine_pvalues.html#scipy.stats.combine_pvalues

デルを構築する。

- ③ ②について、組み合わせを変えながら B 回繰り返し、得られた RMSE を RMSE(1)～RMSE(B) とする。
- ④ ③で得られた B 個の RMSE のうち、RMSE(0)以下の数値を取った個数 r を数えて P 値を計算する： $P=(r+1)/(B+1)$ 。

帰無仮説を「テキスト情報は 360 度評価に対して予測力を持たない」とする。RMSE(0)が RMSE(1)～RMSE(B)によって作られる分布の中で極端に小さければ、p 値が小さくなり、帰無仮説は棄却されることになる。

以下では、 $B=1000$ 、 $K=10$ と設定した。また、有意水準として、 $\alpha=0.05$ と設定した。

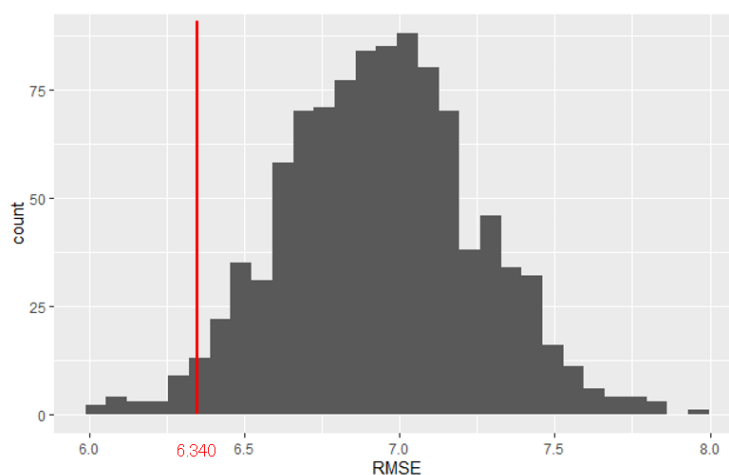


【図 3-2】 Permutation 検定の概要

Cross Validation 及び Permutation 検定を行った結果は、下表の通りである。Cross Validation のデータ分割の際、ランダムシードを固定しているため、各モデルは各回で同一のデータセットに対して学習・検証を行っている。

【表 3-3】 Permutation 検定結果 (B=1000)

特徴量	主成分得点 (50%)		主成分得点 (70%)		主成分得点 (80%)		主成分得点 (90%)			
	RMSE(0)	p値	RMSE(0)	p値	RMSE(0)	p値	RMSE(0)	p値		
1回目	6.340	0.016	6.481	0.528	6.651	0.613	6.780	0.491	6.819	0.054
2回目	5.455	0.005	6.260	0.021	6.399	0.092	6.196	0.006	6.572	0.044
3回目	6.672	0.212	6.511	0.030	6.478	0.041	6.457	0.019	6.760	0.033
4回目	6.001	0.051	7.730	0.876	7.450	0.016	7.623	0.014	7.883	0.068
5回目	6.540	0.213	7.394	0.009	7.203	0.001	7.131	0.001	7.246	0.001
6回目	5.881	0.002	6.235	0.173	6.445	0.515	6.495	0.285	6.883	0.276
7回目	6.538	0.382	6.745	0.251	6.984	0.641	7.256	0.717	8.306	0.924
8回目	4.984	0.004	6.005	0.418	5.965	0.099	6.485	0.699	6.532	0.109
9回目	5.567	0.091	6.884	0.345	6.897	0.106	6.634	0.003	7.092	0.052
10回目	4.916	0.010	7.191	0.591	7.184	0.169	7.353	0.280	7.871	0.557
統合P値	5.E-08		0.010		6.E-05		8.E-07		2.E-05	



【図 3-3】 特徴量を使った回帰モデルの 1 回目の RMSE (1) ~RMSE (1000) のヒストグラム例

テキスト情報から抽出した 768 次元の特徴量を使用した線形回帰モデルは、Stepwise 法による変数選択を行った特徴量を説明変数としている。なお、変数選択を行わない場合について、特徴量がデータ件数 ($n=1040$) を下回るため、最小二乗推定は可能であるが、768 次元の特徴量をそのまま説明変数として扱ったモデルの RMSE が著しく悪かった。そのため、Stepwise 法による変数選択を行っている。

また、主成分得点を使用した線形回帰モデルは、累積寄与率 50%までの主成分の主成分得点、70%までの主成分の主成分得点、80%までの主成分の主成分得点、90%までの主成分の主成分得点を説明変数としており、Stepwise 法による変数選択は行っていない。

図 3-2 に示した例について、RMSE(1)～RMSE(1000)の分布において、RMSE(0)=6.340 を下回るデータが非常に少ないことが分かる。

検定の結果、特徴量を用いた線形回帰モデル、累積寄与率 70%、80%、90%までの主成分得点を用いた線形回帰モデルにおいて、統合 p 値が有意水準 ($\alpha=0.05$) を下回るため、帰無仮説の「テキスト情報は 360 度評価に対して予測力を持たない」を棄却し、対立仮説の「テキスト情報が 360 度評価に対して予測力を持っている」を採択する。一方、累積寄与率 50%及び 70%までの主成分得点を用いた線形回帰モデルの p 値は有意水準 ($\alpha=0.05$) を上回るため、帰無仮説は棄却されず、予測力を持っているとは言い難い。

P 値の本来の使い方ではないが、主成分得点については累積寄与率 80%までの主成分を用いたモデルの統合 P 値が、累積寄与率 90%のモデルの統合 P 値を下回るため、モデルに組み込む主成分数が多ければ良いとは言えず、最適な数があることが予想される。これは、768 次元の特徴量をそのまま説明変数として扱ったモデルの性能が悪いことからわかる。

3.3 線形モデルを用いたテキスト情報の追加的情報価値の評価

将来の 360 度評価結果のスコアに対して、360 度評価を構成する数値評価データ（予測対象の 360 度評価から 2 四半期前の 360 度評価結果のスコア、同じく 2 四半期前の上司からの期待度・満足度計 80 項目、部下からの期待度平均・満足度平均）のみを説明変数に含む線形回帰モデルと、数値評価データに加えてテキスト情報を説明変数に含む線形回帰モデルの予測性能を比較し、テキスト情報を説明変数に追加することによる予測力を評価する。テキスト情報が持つ予測力が数値評価データに含まれている場合、双方の線形モデルの予測性能には有意な差が見られないはずである。

以下の手順で行う。

- ① 将来の 360 度評価結果のスコアを被説明変数として、数値評価データ（2 四半期前の 360 度評価結果のスコア、上司からの期待度 40 項目、上司からの満足度 40 項目、部下からの期待度平均値、部下からの満足度平均値）を説明変数とする線形回帰モデル (Model A) を構築する。線形回帰モデルの構築においては、Stepwise 法による変数選択を行う。
- ② ①の説明変数に、テキスト情報の特徴量を加えた線形回帰モデル (Model B)、テキスト情報の累積寄与率 50%までの主成分得点を加えた線形回帰モデル (Model C-1)、70%までの主成分得点を加えた線形回帰モデル (Model C-2)、80%までの主成分得点を加えた線形回帰モデル (Model

C-3)、90%までの主成分得点を加えた線形回帰モデル (Model C-4) を構築する。Model B においては、モデル構築に際し、Model A の説明変数を固定して、テキスト情報の説明変数についてのみ Stepwise 法による変数選択を行う。主成分得点を加えた Model C-1~C-4 については、Stepwise 法による変数選択を行わない。

- ③ 双方の線形回帰モデルの性能を表す RMSE、決定係数 R^2 、自由度調整済み決定係数 R^2 を算出する。また、分散分析により予測結果に有意な差があるか調べる。
- ④ 3.2 と同様の手順で「テキスト情報が 360 度評価結果のスコアに対して追加的情報価値を持たない」を帰無仮説として、Permutation 検定を行う。

表 3-4 に、Model A の説明変数、偏回帰係数、標準誤差、t 統計量の値、t 検定したときの有意確率を記載する。

【表 3-4】 Model A の説明変数、偏回帰係数ほか

変数名		Estimate	Std. Error	t value	Pr(> t)
(Intercept)	切片	17.38	3.24	5.37	0.00 ***
BeforeScore	2四半期前の360度評価スコア	0.71	0.02	34.64	< 2e-16 ***
MemEmean	部下の期待度平均値	0.56	0.23	2.39	0.02 *
E7	期待度_情報提供	0.45	0.30	1.48	0.14
E9	期待度_情報提供	-0.65	0.33	-2.00	0.05 *
E11	期待度_情報収集	-0.62	0.28	-2.18	0.03 *
E12	期待度_情報収集	-0.60	0.36	-1.65	0.10 .
E14	期待度_情報収集	-0.41	0.27	-1.52	0.13
E16	期待度_情報収集	0.97	0.33	2.93	0.00 **
E17	期待度_情報収集	0.81	0.32	2.51	0.01 *
E27	期待度_判断行動	-0.86	0.37	-2.36	0.02 *
E29	期待度_判断行動	-0.62	0.27	-2.27	0.02 *
E33	期待度_支援行動	0.59	0.37	1.60	0.11
E39	期待度_支援行動	0.47	0.20	2.39	0.02 *
S4	満足度_情報提供	0.48	0.21	2.28	0.02 *
S8	満足度_情報提供	-0.40	0.24	-1.64	0.10
S17	満足度_情報収集	-0.46	0.23	-2.01	0.04 *
S29	満足度_判断行動	0.55	0.23	2.41	0.02 *
S35	満足度_支援行動	0.53	0.26	2.01	0.04 *
S38	満足度_支援行動	-0.56	0.24	-2.31	0.02 *
S40	満足度_支援行動	0.36	0.23	1.52	0.13

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

続いて、モデルの予測性能を比較する。3.2と同様にデータの分割の仕方への依存を回避するため、10-fold Cross Validationを行い、RMSE、決定係数 R^2 、自由度調整済み決定係数 R^2 の平均値を算出した。なお、決定係数 R^2 、自由度調整済み決定係数 R^2 は内挿データ（訓練データ）に対する値であり、RMSE は外挿データ（検証データ）に対する値である。また、「テキスト情報が360度評価結果のスコアに対して追加的情報価値を持たない」を帰無仮説として Permutation 検定を行った。

主成分得点を使用したモデルについては、Stepwise 法による変数選択を行っていないため、分散分析により追加的に加えた主成分得点の変数の値が有意に回帰の誤差分散を減少させるか否かを確認した。

【表 3-5】モデルの性能比較 (RMSE、 R^2 、自由度調整済み R^2)

モデル	RMSE	R^2	Adj R^2
Model A	4.77	0.51	0.50
Model B	3.75	0.75	0.69
Model C-1	4.76	0.51	0.50
Model C-2	4.76	0.53	0.50
Model C-3	4.80	0.54	0.50
Model C-4	4.83	0.58	0.49

【表 3-6】分散分析結果 (Model A と Model C-1~C-4)

モデル	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Model A	1019	23204				
Model C-1	1004	22750	15	453.52	1.2986	0.1959
Model C-2	974	22156	30	593.9	0.8502	0.6982
Model C-3	937	21620	37	536.22	0.6224	0.9628
Model C-4	856	19930	81	1689.49	0.8958	0.7301

表 3-5 から、テキスト情報の特徴量を説明変数に加えた Model B の予測性能 (RMSE3.75、決定係数 $R^2=0.75$ 、自由度調整済み $R^2=0.69$) は、Model A の予測性能 (RMSE=4.77、決定係数 $R^2=0.51$ 、自由度調整済み $R^2=0.50$) より高いことが分かる。テキスト情報の主成分得点を説明変数に加えた Model B-2 の予測性能 (RMSE4.76) についても、Model A の予測性能 (RMSE=4.77) よりわずかな改善が見られているが、表 3-6 より、主成分得点を使用したすべてのモデルで p 値が 0.05 以上となっており、Model A との予測性能に有意な差異がないことが分かった。

【表 3-7】 Permutation 検定結果 (B=1000)

	特徴量		主成分得点 (50%)		主成分得点 (70%)		主成分得点 (80%)		主成分得点 (90%)	
	RMSE(0)	p値	RMSE(0)	p値	RMSE(0)	p値	RMSE(0)	p値	RMSE(0)	p値
1回目	5.155	0.793	4.688	0.075	4.611	0.007	4.674	0.025	5.019	0.146
2回目	3.458	0.002	4.431	0.962	4.530	0.907	4.671	0.870	4.755	0.567
3回目	3.646	0.150	5.251	0.132	5.241	0.045	5.505	0.546	5.718	0.402
4回目	4.097	0.393	4.871	0.361	4.942	0.402	5.152	0.706	5.506	0.718
5回目	3.667	0.005	4.602	0.138	4.661	0.122	4.715	0.155	4.970	0.094
6回目	4.412	0.463	4.308	0.442	4.473	0.785	4.631	0.819	5.248	0.943
7回目	4.656	0.307	5.377	0.963	5.426	0.932	5.581	0.955	6.065	0.927
8回目	3.435	0.091	4.672	0.412	4.821	0.714	5.031	0.858	5.479	0.795
9回目	4.027	0.118	4.786	0.113	4.906	0.496	5.061	0.627	5.440	0.700
10回目	3.147	0.014	5.522	0.166	5.657	0.569	5.688	0.405	5.935	0.350
統合P値	9.E-05		0.143		0.160		0.663		0.732	

Permutation 検定は、数値評価データの説明変数の組み合わせは固定し、テキスト情報のデータの説明変数のみ組み合わせをランダムに入れ替えて行った。表 3-7 から、テキストの特徴量を用いた統合 p 値のみが有意水準 ($\alpha=0.05$) を下回り、帰無仮説を棄却し、対立仮説を採択することとなった。主成分得点については、すべてのモデルで統合 p 値が有意水準 ($\alpha=0.05$) を上回り、帰無仮説を棄却できないという結論となった。

3.4 非線形モデルを用いた予測性能の評価

ここまで線形回帰モデルを用いて、テキスト情報の予測力の評価を行ってきた。線形回帰モデルでは、線形性の制約ゆえ、テキスト情報の価値を有効活用できていない可能性がある。そこで本節では、非線形モデルを含む機械学習手法を用いたモデルを構築することで、非線形モデルへの変更によって予測性能が向上するか検証する。

機械学習手法として、多層パーセプトロン回帰、ランダムフォレスト回帰、勾配ブースティング回帰を使用する。手法の使用に伴い、データセットに適したハイパーパラメータの探索を行う。ハイパーパラメータは機械学習のアルゴリズムの挙動を設定するパラメータを指す。機械学習手法ごとに異なる値のハイパーパラメータがある。グリッドサーチで全てのパラメータの組み合わせを試し、Cross Validation で RMSE を最小化する値を探索する。

・多層パーセプトロン回帰（ニューラルネットワーク）

パラメータ	説明
hidden_layer_sizes	隠れ層の層の数と、ニューロンの数を指定する。
solver	活性化関数を指定する。
activation	最適化手法を指定する。

・ランダムフォレスト回帰

パラメータ	説明
ctiterion	分割基準を指定する。
max_depth	木の深さの最大値を指定する。
n_estimators	木の生成数を指定する。
max_features	ノードを分割するときに考慮すべき特徴量のランダムなサブセットのサイズを指定する。

・勾配ブースティング回帰

パラメータ	説明
learning_rate	学習率を指定する。
max_depth	木の深さの最大値を指定する。
min_samples_leaf	木の生成数を指定する。
max_features	ノードを分割するときに考慮すべき特徴量のランダムなサブセットのサイズを指定する。

Model B の説明変数を機械学習手法のインプットとし、手法を変えたことにより予測性能が向上するかを検証する。非線形モデルを前提としているため、モデル性能については RSME によって比較する。以下、手法による RMSE の差異と、グリッドサーチにより調整したハイパーパラメータを記載する。ハイパーパラメータは、テキスト情報追加のモデルのものである。

【表 3-8】手法による RMSE の差異比較

モデル	数値評価データ	テキスト情報追加
	RMSE	RMSE
線形回帰モデル Model A	4.73	-
線形回帰モデル Model B	-	3.61
多層パーセプトロン回帰モデル	4.91	4.33
ランダムフォレスト回帰モデル	5.01	4.55
勾配ブースティング回帰モデル	3.57	2.81

・多層パーセプトロン回帰（ニューラルネットワーク）

パラメータ	値	説明
hidden_layer_sizes	100, 100	中間層 2 つ（各中間層のニューロン数が 100 個、100 個）の 4 層のニューラルネットワーク
solver	lbfgs	準ニュートン法の BFGS の一種の最適化手法
activation	relu	中間層の活性化関数に Relu 関数を設定

・ランダムフォレスト回帰

パラメータ	値	説明
criterion	mae	分割基準に MAE(平均絶対誤差)を設定
max_depth	4	木の深さの最大値を 4 に設定
n_estimators	10000	木の生成数を 10000 本に設定
max_features	auto	特徴量のサブセットのサイズを n_estimators と同じ 100 に設定

・勾配ブースティング回帰

パラメータ	値	説明
learning_rate	0.1	学習率を 0.1 に設定
max_depth	5	木の深さの最大値を 5 に設定
n_estimators	10000	木の生成数を 10000 本に設定
max_features	log2	特徴量のサブセットのサイズを $\log_2(n_estimators)$ に設定

表 3-8 から、数値評価データのみを説明変数に用いた場合、及びテキスト情報を追加した場合においても、勾配ブースティング回帰を用いたモデルで RMSE の改善が見られた。一方で、多層パーセプトロン回帰、ランダムフォレスト回帰では RMSE の改善が見られなかった。

なお、線形回帰モデルの表 3-5 の RMSE と、表 3-8 の RMSE が異なる理由は、非線形回帰モデルとの予測性能を比較する上で、線形回帰モデルと非線形回帰モデルで使用するデータセットの条件が変わらないようにしているため、表 3-5 で線形モデル間の予測性能の比較に使用したデータセットとは異なっているためである。

4 考察

本研究は、360 度評価のテキスト情報の応用の検討に向けて、大きく分けて二つの実証分析を行った。一つ目は、テキスト情報が将来の 360 度評価結果のスコアに対して予測力を持っているかについての実証分析である。二つ目は、将来の 360 度評価結果のスコアの予測に非線形モデルを用いることで線形モデルと比較して予測力が向上するかの実証分析である。

一つ目について、Permutation 検定、及び線形回帰モデルを使った予測性能の比較により、BERT

によって抽出したテキスト情報が、将来の 360 度評価結果のスコアに対して予測力を持っていることが分かった。一方、テキスト情報の特徴量の意味解釈を目的とした主成分分析によって抽出した主成分得点については、単体では多くの主成分を用いる場合には予測力を持っているが、追加的情報価値の評価においては、線形モデルの枠組み内ではあるが予測力を持たないことが分かった。768 次元で文書の特徴を表している特徴量に対して、次元圧縮を行うことで、本来持っている情報価値をそぎ落としてしまったと考えられる。この結果については、対象データのみ限定された結果、または BERT の仕様に依存する可能性、採用した事前学習済みモデルの仕様に依存する可能性があるが、テキスト情報の有用性及びその評価方法について一つの示唆を与えることができた。

二つ目について、機械学習手法である勾配ブースティング回帰を用いたモデルにおいて、RMSE の改善が見られた。一方、多層パーセプトロン回帰、ランダムフォレスト回帰については、線形回帰モデルより予測性能が低いという結果であった。勾配ブースティング回帰については予測性能が向上する一方で、ハイパーパラメータに影響されやすいため、チューニングを注意深く行う必要があり、かつモデルの解釈が困難である。非線形回帰モデルとの比較について検証した上で、解釈可能なモデルを選択する場合については、線形回帰モデルを採用することも問題ないとする。

本研究では、360 度評価結果のスコアの予測における上司からのフィードバックテキストの追加的情報価値の評価において、モデルの RMSE の改善はわずかであったが、上司のフィードバックが部下に与える影響を定量的に評価することができた。部下育成の課題は前々から挙げられていたが、フィードバックの内容を定量的に評価しようという取り組みはまだ少ないと感じられる。今後、コロナ禍における上司・部下間の 1on1 マネジメントの重要性の高まりや、労働市場の流動性の高まりに対応するために、企業の人事部門としては上司による部下の育成やマネジメントに関する手法をより積極的に分析する必要がある。管理職研修などにおいても、フィードバックの内容についてその効果を定量的に分析した上で、より具体的なフィードバックの手法や指針を示すべきであるとする。

また、今まで HR 領域の ICT 活用のボトルネックであった人事部門の ICT リテラシー、問題解決能力を向上については、本研究で扱ったオープンソースの自然言語処理モデルやライブラリといった公開されたツールを使用することで、一定の専門性を保持していれば比較的容易にデータを分析・評価可能な形に成形することができる。本研究は、HR 領域の非構造化データであるテキスト情報の応用方法を提案することを目的とし、具体的なアプローチとして 360 度評価のフィードバックテキストを扱ったが、HR 領域においては 360 度評価に限らず、眠っている人事データを活用してマーケティング分野のように仮説と検証を繰り返していけるような体制を作っていくべきであるとする。本研究の方法論が、HR 領域における ICT 活用の一つの示唆となることを期待する。

5 本研究の課題と限界

本研究の主な課題と限界について、以下の三点が挙げられる。

・現在の自然言語処理技術の限界について

近年、自然言語処理技術が急速に進化している原因として、深層学習（ディープラーニング）を用いた精度の向上が挙げられる。同時に、深層学習の特徴でもある出力結果の解釈困難性について、BERT でも同様の課題が挙げられている。BERT モデルの学習過程やその内容、文脈の解釈の方法などの仕組みの理解については発展途上であり、本研究での課題となっている BERT が出力する特徴量がどのような判断基準によって出力されているのか、という点を解き明かすことは非常に困難である。本研究では、特徴量の解釈を目的として主成分分析による次元圧縮を行ったが、抽出した第 1～第 15 主成分得点について、数値評価データとの相関、単語の共起ネットワークの形成、文面の目視による確認を行い、解釈を試みた（付録 1）。主成分にラベリングを行ったが、最終的には恣意的な判断が介在してしまうこと、実務においては大量の文章に目を通して手動で解釈することは現実的でないことが分かった。自然言語処理技術は業務の現場で発展を続けてきたため、どうしても現場では性能に注目されてしまうが、今後は業務への応用の幅を広げるためにも、深層学習をベースにした自然言語処理技術の研究が進み、学習済みモデルに対する解釈可能性が高まることに期待したい。

・フィードバックテキストの構成の限界について

テキスト情報は非常にノイズレベルが高いが、定量データには表れない言葉のニュアンスなどが表れる点で、利活用の価値がある。しかし、本研究で扱ったテキストは、データ分析を前提としておらず、フォーマットの自由度が非常に高いものであった。例えば、長所・短所がバランスよく記載されているものや、どちらかがないもの、どちらもないもの、といったような記載レベルの違いが顕著であった。より純度の高い情報を抽出可能とするには、長所・短所の記載のフォームを分けるなど、記載の自由度を残しながらも分析の前処理を簡素化する仕組みを一つ一つ入れていく必要がある。このような観点は、データ分析をすることで初めて気づく点であり、今後企業はこのような小さな改善を繰り返しながらテキスト情報の活用を目指すべきである。テキスト解析を前提としたデータ取得のフォーマットの改善提案としては以下の三つを挙げる。

- 記述対象のテーマを細分化し文章のラベル付けを容易にする

本研究のケースでは、「役割」「課題」「期待」といったようなテーマに対して、自由に記述するという形式であった。本論文の分析に使用した「課題」のコメントについても、強みや弱みの混在、強みの記述においても言及されているトピックは多様性を持っていた。自然言語処理において文章のラベリングは、解析の容易性及び拡張性を持たせる上で、非常に重要である。そのため、コメントの書き手に対して、可能な限りトピックを選択させる仕組みが効果的である。例えば「課題」の場合、まず「強み」「弱み」のどちららについて記述するかについて選択を行い、その後「判断行動」「支援行動」といった 360 度評価の評価項目に紐づくトピックを選択し、最終的に選択したトピックについてコメ

ントを行う。書き手の作業量は増えるが、文章へのラベリングには必要な作業と言える。

- 箇条書き、体言止め、句点の制約を設ける

本研究で使用したコメントにおいては、1つの段落として記述されているもの、箇条書きによるもの、体言止めで統一されているものなど、記述方法が統一されていなかった。最新の自然言語処理では、文脈の把握、前後の文章同士の繋がりまで含めて、単語や文章の特徴を解析する。箇条書きや体言止めは、人間の読みやすい形式ではあるが、機械が読む場合には適切ではないと言える。例えば、今回のような「強み」「弱み」の記載が混在している場合には、「・リーダーシップ」とだけ記載された内容を「リーダーシップが発揮できている」のか、「リーダーシップが発揮できていない」「リーダーシップを発揮してほしい」という意味なのか、文面から捉えることができない。

また、前後の文章の関係性を見る上で、句点の間を一つの文章として捉えるため、「・リーダーシップ・顧客への対応」といった箇条書きの文章を一つの文章として認識してしまう。「リーダーシップ。顧客への対応。」というように句点を記載することで、前半部分と後半部分が別の意味を持った文章として正しく認識される。

- 社内でのみ使用している固有名詞の利用の制約を設ける

自然言語処理のモデルでは事前に学習した単語や文章、辞書などをもとにテキスト解析を行うため、人物名・組織名・会社名・商品（サービス）名といった単語については解析が困難となるケースがある。しかしながら、本研究で使用したデータの場合は、コメントの本来の目的は分析ではなく、研修講師に従業員の抱える課題を伝えることであった。社内でのコミュニケーションの効率化という点では、具体性の持つ固有名詞は非常に有用である。このような例も含め、目的に合わせたフォーマットの検討と選択が必要となる。

・データ制約における課題について

本研究で使用したデータは特定企業の特定データであり、論文内で検討した各分析手法は、当該データの仕様に合わせて適用された。もとより、今回の研究で得られた実証分析結果は直ちに一般化されるものではない。しかしながら、本論文のテキスト情報の評価方法や利活用の方法は、一般性を有しており、今後、他企業や他業界内の同種の人事評価データに対しても適用可能なものである。もちろん、本手法を適用するためには当該データの仕様に合わせる形で、適宜カスタマイズする必要がある。

謝辞

最後に、研究にご協力いただいた方々へのお礼の言葉をもって、当論文を締めくりたいと思います。まず、本研究についてさまざまなアドバイスを頂き、研究室での実りある1年を送らせていただきました、林高樹教授に感謝したいと思います。次に、本研究のきっかけを作っていた副査の岩本先生、データをご提供いただいた株式会社リンクアンドモチベーションの大島様、林幸弘様にも感謝いたします。そして、副査の山尾先生、林洋一郎先生には、よりよい研究になるよう度重なるアドバイスをいただき、最後まで寄り添っていただきました。林高樹研究室のメンバー（金野さん・楊さん・吉田さん・渡邊さん）は非常に心地よく、雰囲気の良い研究室であったため、コロナ禍において対面の機会が少なくなってしまっていたことが非常に残念です。

本研究を始めとして、大学院での充実した2年目を送ることができたのは、皆様のおかげだと感じています。本当にありがとうございました。

6 参考文献

- Abel, M., & Buchman, D. (2020). The Effect of Manager Gender and Performance Feedback: Experimental Evidence from India, IZA Institute of Labor Economics, No. 13871.
- Bernardin, H. J., Thomason, S., Buckley, M. R., Kane, J. S. (2015). Rater rating-level bias and Accuracy in performance appraisals: the impact of rater personality, performance management competence, and rater accountability, *Human Resource Management*, 55(2).
- Crouxa, C., Filzmoser, P., & Oliveira, M. R. (2007). Algorithms for Projection-Pursuit robust principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 87, 218-225.
- Devlin, J., Chang, M., Lee, K., Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Google AI Language.
- Golland, P., Fischl, B. (2003). Permutation tests for classification: Towards statistical significance in image-based studies. *International Conference on Information Processing and Medical Imaging*, pages 330-341.
- Langer, N., Slaughter, S. A., & Mukhopadhyay, T. (2014). Project managers' practical intelligence and project performance in software offshore outsourcing: A field study, *Information Systems Research*, 25, 364-384.
- Lee, M., Huang, Y. L. (2020). Corporate Social Responsibility and Corporate Performance: A Hybrid Text Mining Algorithm, *Sustainability* 2020, 12(8), 3075.
- Molinaro, A. M., Simon, R., & Pfeiffer, R. M. (2003). Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, 21(15):3301-3307.
- Ojala, M., Garriga, G. C. (2010). Permutation Tests for Studying Classifier Performance. *Journal of Machine Learning Research* 11 (2010) 1833-1863.
- Zenger, J. H., Folkman, J. R., & Edinger, S. K. (2011). Making yourself indispensable. *Harvard Business Review*, 89, 84-90.
- 青嶋誠, 矢田和善 (2013). 高次元データの統計的方法論, 日本統計学会誌第 43 巻, 123-150.
- 経済産業省 (2019). 平成 30 年度産業経済研究委託事業「企業の戦略的人事機能の強化に関する調査」, https://www.meti.go.jp/shingikai/economy/jinzai_management/pdf/20190329_04.pdf (参照日 2020-12-25).
- 厚生労働省 (2019). 平成 30 年度版 労働経済の分析—働き方の多様化に応じた人材育成の在り方について—, <https://www.mhlw.go.jp/wp/hakusyo/roudou/18/dl/18-1.pdf> (参照日 2020-12-25).
- 児玉小百合 (2020). 「テレワークに潜む「リモハラ」の危険」『日本経済新聞』, 2020-5-31 (参照日 2020-12-25).
- 兵藤誠 (2015). 「テキストマイニングを使った人材の 360 度評価」に関する考察, 慶應義塾大学大学院経営管理研究科, 修士学位論文, 2014 年度経営学, 第 2990 号.
- 藤江嘉彦, 宮崎陽司, 山岸健太郎 (2014). 管理職への役割転換における学習・実践を促進するモデル, 経営行動科学学会第 17 回大会発表論文集.

山尾佐智子(2020). 企業の人材マネジメントの実態～日米ヒアリング調査より～, 公益財団法人
日本生産性本部生産性総合研究センター 生産性レポート Vol.17「日本企業の人材育成投資の
実態と今後の方向性～人材育成に関する日米企業ヒアリング調査およびアンケート調査報告
～」, 2020年12月, 4-26.

付録

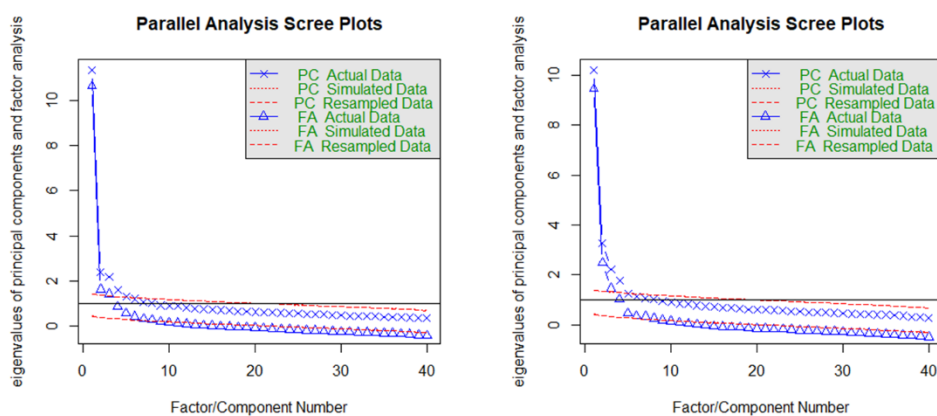
付録1 主成分の解釈について

各主成分がテキスト情報のどのような特徴を表しているのか解釈を試みた。

因子分析

主成分と数値評価データの相関の確認を行うにあたり、満足度・期待度の各40項目に対してあらかじめ探索的因子分析を行うことで、解釈可能な変数因子の抽出を試みる。因子ごとに計算した因子得点と主成分の値の相関を確認し、各主成分が表す特徴の理解を図る。

因子数を決めるにあたり、代表的なアプローチ方法として①固有値1以上の基準、②スクリーテスト、③MAP、④平行分析、⑤最尤解のカイ2乗検定がある。本分析において因子数の決定を行うために平行分析とMAPを用いる。平行分析の結果、固有値1以上の因子について、期待度では因子数は3つ、満足度については因子数4つという結果が出た。



【図】 平行分析（右：期待度、左：満足度）

次に、Rパッケージ「psych」のVSS関数を用いて、MAP（最小平均偏相関）による推奨因子数を確認する。期待度については、最小の因子数が4因子であったが、BICの値からは6因子が推奨された。できるだけ因子負荷量が低い項目を落とさず、解釈可能な因子を検討するため、まずは因子数を6つに設定して分析を進める。

```

> print(MAPml)

Very simple Structure
Call: vss(x = object, fm = "ml")
VSS complexity 1 achieves a maximum of 0.81 with 1 factors
VSS complexity 2 achieves a maximum of 0.85 with 2 factors

The Velicer MAP achieves a minimum of 0.01 with 4 factors
BIC achieves a minimum of -2357.6 with 6 factors
Sample size adjusted BIC achieves a minimum of -717.85 with 8 factors

Statistics by number of factors
  vss1 vss2  map dof chisq  prob sqresid fit RMSEA  BIC SABIC complex eChisq SRMR eCRMS eBIC
1 0.81 0.00 0.0099 740 4979 0.0e+00 30 0.81 0.074 -162 2189 1.0 7978 0.070 0.072 2837
2 0.50 0.85 0.0085 701 3823 0.0e+00 24 0.85 0.065 -1047 1180 1.5 5235 0.057 0.060 366
3 0.40 0.78 0.0068 663 2745 1.7e-250 20 0.87 0.055 -1860 245 1.9 2870 0.042 0.046 -1736
4 0.38 0.75 0.0065 626 2118 5.4e-161 18 0.89 0.048 -2231 -243 2.0 1913 0.034 0.038 -2436
5 0.31 0.65 0.0069 590 1770 3.4e-118 16 0.90 0.044 -2328 -454 2.3 1461 0.030 0.035 -2638
6 0.32 0.63 0.0074 555 1498 1.1e-87 15 0.90 0.040 -2358 -595 2.4 1133 0.026 0.031 -2723
7 0.29 0.60 0.0082 521 1310 1.4e-69 14 0.91 0.038 -2309 -654 2.6 944 0.024 0.030 -2676
8 0.26 0.50 0.0089 488 1122 6.4e-52 14 0.91 0.035 -2268 -718 2.9 748 0.021 0.027 -2642

```

【図】MAPによる推奨因数（期待度）

満足度については、最小の因子数が4因子であったが、BICの値からは7因子が推奨された。こちらについても、まず因子数は7つに設定して進めることとする。

```

> print(MAPml)

Very simple Structure
Call: vss(x = object, fm = "ml")
VSS complexity 1 achieves a maximum of 0.74 with 1 factors
VSS complexity 2 achieves a maximum of 0.82 with 2 factors

The Velicer MAP achieves a minimum of 0.01 with 4 factors
BIC achieves a minimum of -2265.56 with 7 factors
Sample size adjusted BIC achieves a minimum of -668.49 with 8 factors

Statistics by number of factors
  vss1 vss2  map dof chisq  prob sqresid fit RMSEA  BIC SABIC complex eChisq SRMR eCRMS eBIC
1 0.74 0.00 0.0143 740 6405 0.0e+00 36 0.74 0.086 1265 3615 1.0 12575 0.088 0.090 7434
2 0.56 0.82 0.0094 701 4208 0.0e+00 25 0.82 0.069 -662 1564 1.4 5783 0.060 0.063 913
3 0.47 0.78 0.0076 663 3056 1.4e-302 21 0.85 0.059 -1550 556 1.7 3262 0.045 0.049 -1344
4 0.42 0.75 0.0067 626 2214 6.6e-176 18 0.87 0.049 -2135 -146 1.8 1930 0.034 0.039 -2418
5 0.39 0.68 0.0072 590 1835 1.2e-127 17 0.88 0.045 -2264 -390 2.1 1538 0.031 0.035 -2561
6 0.38 0.65 0.0076 555 1594 4.1e-101 15 0.89 0.042 -2261 -499 2.2 1231 0.028 0.033 -2624
7 0.35 0.61 0.0082 521 1354 2.4e-75 15 0.90 0.039 -2266 -611 2.4 984 0.025 0.030 -2635
8 0.35 0.60 0.0089 488 1172 4.1e-58 14 0.90 0.037 -2218 -668 2.6 808 0.022 0.028 -2582

```

【図】MAPによる推奨因数（満足度）

続いて、Rパッケージ「psych」のfa関数を用いて、因子抽出を行う。因子抽出には最尤法を採用、因子回転にはプロマックス法を採用した。

下表のとおり、期待度は6因子を抽出し、因子負荷量0.3以下の項目が確認されなかったため、項目を落とさずに因子を抽出することができた。なお、各項目の質問内容が属する、4つのカテゴリーを記載している。各因子は項目の質問内容とカテゴリーを確認の上、因子にラベリングして解釈可能な形とした。

【表】期待度の因子負荷量とラベル

項目	カテゴリー	能力	行動		姿勢		
		自己・他者情報の共有	業務情報の共有	業務情報の収集	組織貢献への姿勢	成果・品質への姿勢	他メンバーの状況の把握
		ML1	ML2	ML3	ML4	ML5	ML6
e33	支援行動	0.681	-0.114	0.097	-0.027	-0.025	-0.034
e6	情報提供	0.622	0.063	-0.006	-0.019	-0.111	0.154
e10	情報提供	0.591	-0.079	-0.053	-0.014	0.136	0.159
e18	情報収集	0.573	-0.123	-0.118	-0.198	0.514	0.202
e7	情報提供	0.57	0.104	0.115	0.077	-0.273	0.194
e9	情報提供	0.532	0.19	0.236	0.018	-0.216	0.09
e20	情報収集	0.525	0.01	-0.153	0.093	0.104	0.032
e4	情報提供	0.442	0.141	0.113	0.103	-0.127	0.332
e8	情報提供	0.391	0.08	-0.172	0.312	-0.029	-0.057
e35	支援行動	0.37	-0.092	0.048	0.21	0.057	-0.094
e16	情報収集	0.363	-0.042	0.349	-0.016	0.063	0.09
e2	情報提供	0.092	0.698	-0.025	-0.046	0.149	-0.03
e1	情報提供	-0.098	0.696	-0.001	-0.09	0.133	-0.113
e3	情報提供	0.107	0.546	-0.033	-0.011	0.115	-0.145
e11	情報収集	0.017	-0.037	0.733	-0.141	-0.053	0.052
e19	情報収集	0.041	-0.048	0.695	-0.068	-0.076	-0.178
e14	情報収集	-0.337	0.074	0.543	0.235	0.065	0.144
e17	情報収集	0.007	0.039	0.51	-0.06	0.12	0.091
e12	情報収集	0.265	0.032	0.466	-0.148	0.139	0.049
e13	情報収集	0.238	0.097	0.418	0.132	-0.062	0.012
e21	判断行動	-0.043	0.012	0.391	0.344	0.016	-0.11
e23	判断行動	0.001	-0.128	0.369	0.18	0.338	-0.033
e22	判断行動	0.096	-0.018	0.328	0.239	0.135	-0.144
e40	支援行動	0.063	-0.065	-0.126	0.747	0.084	0.043
e32	支援行動	0.174	-0.163	-0.044	0.585	0.025	0.025
e37	支援行動	0.023	0.101	-0.117	0.577	0.025	0.011
e28	判断行動	0.101	0.01	0.024	0.44	0.034	0.061
e39	支援行動	-0.337	-0.029	0.162	0.422	0.13	0.194
e34	支援行動	0.002	-0.095	0.26	0.398	-0.122	0.09
e31	支援行動	0.068	0.012	0.234	0.391	-0.06	-0.068
e27	判断行動	0.323	0.019	0.055	0.343	0.052	-0.278
e36	支援行動	0.218	0.065	-0.108	0.338	0.137	0.085
e38	支援行動	0.198	0.071	-0.022	0.149	0.396	-0.172
e26	判断行動	0.034	0.069	-0.029	0.012	0.688	-0.01
e29	判断行動	-0.143	0.119	0.054	-0.035	0.665	0.124
e30	判断行動	0.005	0.048	0.013	0.068	0.588	0.05
e25	判断行動	-0.123	0.055	-0.048	0.132	0.565	0.066
e5	情報提供	0.234	0.142	0.133	-0.009	0.354	-0.027
e15	情報収集	0.332	-0.145	0.006	0.049	0.132	0.59

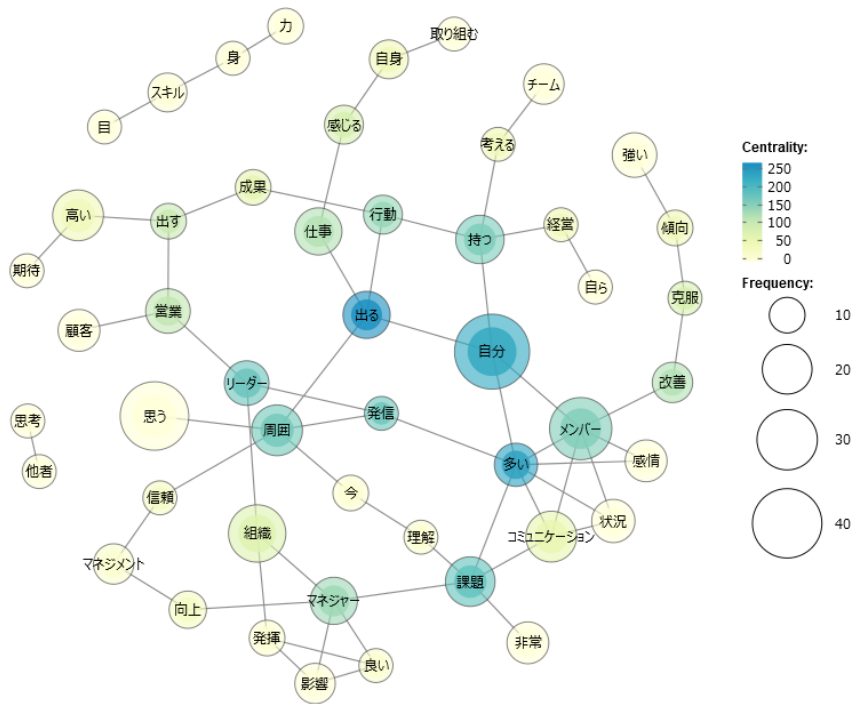
満足度は、1 回目に 2 項目が因子負荷量 0.3 以下となり、これらの項目を取り除いて、再度による推奨因子数を確認し、BIC の値から 6 因子が推奨され、2 回目は 6 因子で抽出した。2 回目については 3 項目が因子負荷量 0.3 以下となり、これらの項目を取り除いて、3 回目の因子抽出をおこなった。BIC の値から 6 因子が推奨された。結果、3 回目については、下表のとおり、因子負荷量 0.3 以下の項目はなく、項目内容とカテゴリーから因子にラベリングを行った。

【表】満足度の因子負荷量とラベル

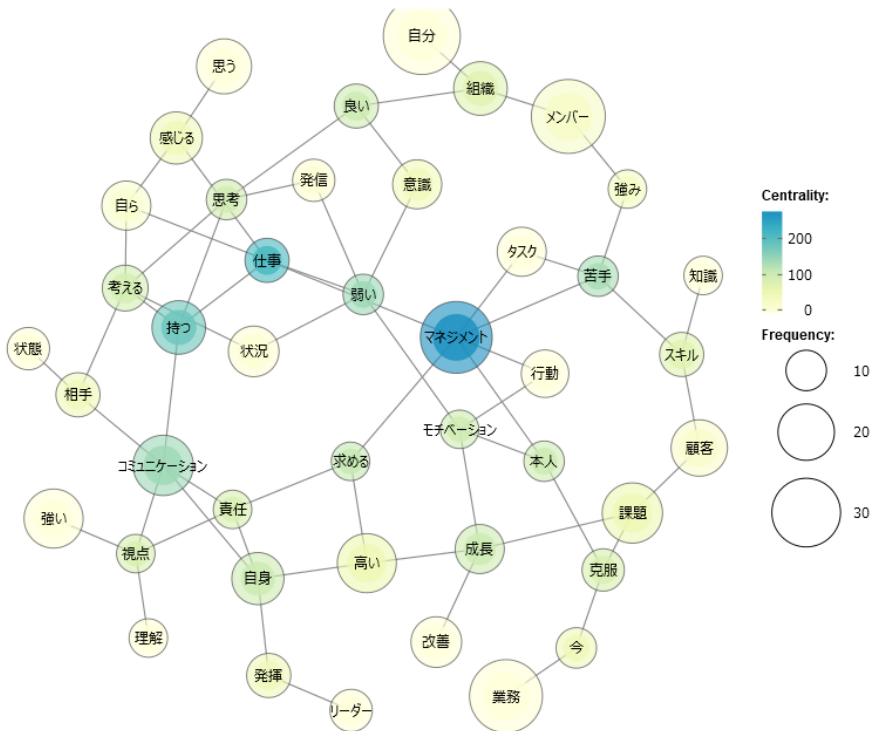
項目	カテゴリー	能力	行動			姿勢	
		業務タスク管理能力	業務情報の共有	業務情報の収集	自己表現・発信力	成果・品質への姿勢	組織貢献への姿勢
		ML1	ML2	ML3	ML4	ML5	ML6
s26	判断行動	0.804	0.068	-0.06	-0.012	0.031	-0.118
s29	判断行動	0.706	0.023	-0.027	-0.035	0.066	-0.119
s38	支援行動	0.654	-0.122	-0.096	-0.015	0.112	0.2
s18	情報収集	0.55	0.034	0.243	-0.074	0.077	-0.084
s25	判断行動	0.536	0.081	0.073	0.029	-0.19	0.064
s30	判断行動	0.513	-0.091	0.247	-0.231	0.279	0.051
s5	情報提供	0.477	0.177	-0.237	0.202	0.058	-0.011
s2	情報提供	-0.013	0.804	0.028	-0.007	0.062	0.094
s1	情報提供	0.009	0.784	0.191	-0.123	0.04	0.083
s3	情報提供	0.253	0.534	-0.122	0.077	-0.024	0.028
s11	情報収集	-0.174	0.06	0.78	-0.046	0.073	-0.056
s14	情報収集	-0.062	0.093	0.691	-0.001	0.127	-0.085
s15	情報収集	0.039	0.018	0.498	0.062	-0.056	0.122
s12	情報収集	0.011	0.024	0.456	0.049	0.23	-0.08
s17	情報収集	0.123	0.026	0.453	0.17	0.082	-0.111
s16	情報収集	0.039	-0.021	0.432	0.065	0.016	0.091
s39	支援行動	0.089	-0.141	0.39	0.069	-0.104	0.186
s4	情報提供	-0.02	0.154	0.059	0.624	-0.083	-0.033
s31	支援行動	-0.051	-0.179	0.048	0.595	0.066	0.032
s34	支援行動	0.027	-0.213	-0.011	0.525	0.105	0.241
s7	情報提供	-0.01	0.001	0.114	0.519	-0.008	0.003
s9	情報提供	-0.029	0.217	0.16	0.438	0.091	-0.134
s13	情報収集	-0.024	0.04	0.341	0.38	-0.079	0.046
s6	情報提供	0.228	0.108	-0.131	0.344	0.156	0.004
s23	判断行動	-0.033	0.092	0.222	-0.033	0.46	0.088
s24	判断行動	0.306	-0.043	0.009	0.082	0.435	-0.138
s22	判断行動	0.031	0.001	0.169	0.13	0.379	0.039
s27	判断行動	0.249	0.006	0.006	-0.102	0.332	0.281
s8	情報提供	-0.062	-0.025	-0.003	-0.024	-0.054	0.622
s40	支援行動	-0.09	-0.113	0.051	0.066	0.105	0.614
s20	情報収集	0.089	0.127	0.077	-0.258	0.019	0.601
s37	支援行動	-0.078	0.184	-0.01	0.105	-0.174	0.565
s35	支援行動	-0.089	0.179	-0.243	0.071	0.243	0.564
s32	支援行動	-0.089	-0.049	0.103	0.106	0.175	0.545
s28	判断行動	-0.044	0.032	0.035	0.004	0.275	0.497
s36	支援行動	0.218	0.02	-0.028	0.106	-0.188	0.475

主成分のラベリング

各主成分が表現する特徴を理解した上で、各主成分にラベリングを行う。各主成分得点の上位10%と下位10%を抽出した上で、共起ネットワークによるテキスト群全体の傾向を可視化、スコアデータ（抽出した因子）との単回帰による相関の確認を行い、その上で文脈の解釈のために実際の文章を目視で確認する。例として第一主成分の共起ネットワークと、各因子との相関を記載する。



【図】 第一主成分の共起ネットワーク（上位 10%）



【図】 第一主成分の共起ネットワーク（下位 10%）

	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分	第6主成分	第7主成分	第8主成分	第9主成分	第10主成分
期待度_自己・他者情報の共有	-0.026	0.001	0.063	-0.081	0.028	-0.188	-0.084	0.006	0.140	-0.061
期待度_業務情報の収集	-0.072	-0.012	0.166	-0.149	0.071	-0.129	-0.069	0.104	0.111	0.019
期待度_組織貢献への姿勢	-0.152	0.037	0.138	-0.103	0.087	-0.102	-0.052	0.121	0.094	-0.009
期待度_成果・品質への姿勢	-0.103	0.064	0.075	-0.092	0.011	-0.086	-0.050	0.061	0.019	-0.106
期待度_業務情報の共有	-0.087	-0.011	-0.012	-0.041	-0.052	-0.122	-0.005	-0.005	0.020	-0.067
期待度_他メンバーの状況の把握	0.036	0.019	0.017	-0.028	0.015	-0.108	-0.100	0.032	0.007	-0.042
満足度_業務タスク管理能力	0.066	0.040	0.004	-0.010	0.007	-0.094	-0.006	0.005	-0.005	0.035
満足度_業務情報の共有	0.049	0.018	-0.049	-0.003	0.017	-0.092	-0.052	-0.015	0.076	-0.010
満足度_業務情報の収集	-0.045	0.056	0.129	-0.052	-0.003	-0.028	-0.039	0.057	0.023	0.014
満足度_自己表現・発信力	-0.062	-0.009	0.055	-0.051	0.032	-0.109	0.019	0.013	0.076	-0.040
満足度_成果・品質への姿勢	0.000	0.071	0.050	0.025	0.042	-0.014	-0.047	-0.037	0.017	0.017
満足度_組織貢献への姿勢	0.063	-0.012	0.095	-0.013	0.063	-0.119	-0.007	0.020	0.062	-0.039

【図】 期待度・満足度の因子と各主成分の相関行列

いくつかの主成分得点では、因子と主成分の弱い相関が確認できたが、ほとんどのテキスト群は同一のトピックであっても、ポジティブ要素とネガティブ要素を含んでおり、因子得点の高低との関係性が見えにくくなっている。

確認の結果から、主成分得点が表現する特徴について解釈を試みた。第11主成分以降は、傾向の判別が難しいため、ラベルの付与は以下の第10主成分までとした。ラベリングの結果については、下図に記載する。

主成分	ラベル	概要
第1主成分	ポジティブ/ネガティブ	値が高いと、ポジティブな記述が多い傾向
第2主成分	セルフコントロール	値が高いと、感情やセルフコントロールに関する記述が多い傾向
第3主成分	マネジメント	値が高いと、業務マネジメントに関する記述が多い傾向
第4主成分	顧客・メンバーの信頼獲得	値が低いと、顧客やメンバーの信頼構築に関する記述が多い傾向
第5主成分	主体性・自主性	値が高いと、主体性や自主性、積極性（チャレンジ精神）に関する記述が多い傾向
第6主成分	リーダー・マネジャーの動き	値が低いと、リーダーやマネジャーの動きや働き方・考え方に関する記述が多い傾向
第7主成分	顧客の理解	値が高いと、顧客の理解に関する記述が多い傾向
第8主成分	コミュニケーション・交渉	値が低いと、顧客・メンバーとのコミュニケーションや交渉に関する記述が多い傾向
第9主成分	いまある強みの向上	値が低いと、伸ばしたい強みに関する記述が多い傾向
第10主成分	タスク管理	値が低いと、納期やタスク管理に関する記述が多い傾向

【図】 主成分のラベリングの結果

Model B-2

【表】 Model B-2 の説明変数、偏回帰係数ほか

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	18.91	2.98	6.35	0.00	***
BeforeScore	0.68	0.02	28.65	< 2e-16	***
MemEmean	0.51	0.27	1.88	0.06	.
E7	0.43	0.30	1.43	0.15	
E9	-0.67	0.33	-2.05	0.04	*
E11	-0.64	0.28	-2.27	0.02	*
E12	-0.57	0.36	-1.56	0.12	
E14	-0.43	0.27	-1.60	0.11	
E16	0.94	0.33	2.85	0.00	**
E17	0.81	0.32	2.52	0.01	*
E27	-0.93	0.37	-2.53	0.01	*
E29	-0.60	0.27	-2.20	0.03	*
E33	0.61	0.37	1.65	0.10	.
E39	0.50	0.20	2.54	0.01	*
S4	0.52	0.21	2.46	0.01	*
S8	-0.40	0.24	-1.66	0.10	.
S29	-0.45	0.23	-1.95	0.05	.
S35	0.55	0.23	2.44	0.02	*
S38	0.53	0.26	2.03	0.04	*
S40	-0.55	0.24	-2.29	0.02	*
S17	0.36	0.23	1.54	0.12	
Comp2	-1.62	0.85	-1.90	0.06	.
Comp1	-1.33	0.77	-1.73	0.08	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

付録3 ソースコード

Permutation 検定、グリッドサーチによるハイパーパラメータの探索、機械学習手法の検証の Python のソースコードを記載する。

• Permutation_test.py

```
import numpy as np
import pandas as pd
from statistics import mean
from math import sqrt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import KFold

df = pd.read_csv('C:¥¥Users¥¥data.csv')

reg = LinearRegression()

rmse0_list = []
mean_list = []
p_list = []

kf = KFold(n_splits = 10, shuffle = True, random_state = 1234)

for train_index, test_index in kf.split(df):
    train = df.iloc[train_index]
    test = df.iloc[test_index]
    Y_train = train['AfterScore']
    X_train = train.drop(columns = ['AfterScore'])
    Y_test = test['AfterScore']
    X_test = test.drop(columns = ['AfterScore'])

    model = reg.fit(X_train, Y_train)
    rmse0 = sqrt(mean_squared_error(Y_test,model.predict(X_test)))

    print(rmse0)
    rmse0_list.append(rmse0)
rmse_list = []
```

```

for num in range(1000):
    X_train = X_train.sample(frac=1)
    X_test = X_test.sample(frac=1)

    model = reg.fit(X_train, Y_train)
    rmse = sqrt(mean_squared_error(Y_test,model.predict(X_test)))
    rmse_list.append(rmse)

target = [i for i in rmse_list if i < rmse0]
p = len(target)/len(X_train)
print(p)
p_list.append(p)

mean_list.append(mean(rmse_list))
print(mean_list)

print(rmse0_list)
print(p_list)

mean = mean(p_list)
print(mean)

```

- ・グリッドサーチによるハイパーパラメータの探索

Gridsearch.py

```
import numpy as np
import pandas as pd
from math import sqrt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import make_pipeline
from sklearn.ensemble import RandomForestRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import GridSearchCV

df = pd.read_csv('C:¥¥Users¥¥data.csv')

y = df['AfterScore']
X = df.drop(columns = ['AfterScore'])

(X_train, X_test, y_train, y_test) = train_test_split(X, y, test_size=0.2, random_state=1234)

estimator_ml = MLPRegressor()
estimator_rf = RandomForestRegressor()
estimator_gb = GradientBoostingRegressor()

param_grid_ml = {'hidden_layer_sizes':[(100,), (100, 10), (100, 100, 10), (100, 100, 100, 10)],
                  'solver':['lbfgs', "sgd", "adam"],
                  'activation':['relu', 'logistic', 'tanh']}

param_grid_rf = {'criterion':['mse', 'friedman_mse', 'mae'],
                  'max_depth': [2,3,4,5,6,7],
                  'n_estimators': [1,10,50,100]}

param_grid_gb = {'criterion':['mse', 'friedman_mse', 'mae'],
                  'learning_rate':[0.001,0.01,0.1,0.2],
                  'max_depth':[3,4,5,6,7],
```

```

        'loss':['ls', 'lad', 'huber', 'quantile'],
        'n_estimators': [25,50,75,100,125]}

gs_ml = GridSearchCV(estimator=estimator_ml,
                    param_grid=param_grid_ml,
                    scoring='rmse',
                    n_jobs=-1)

gs_rf = GridSearchCV(estimator=estimator_rf,
                    param_grid=param_grid_rf,
                    scoring='rmse',
                    n_jobs=-1)

gs_gb = GridSearchCV(estimator=estimator_gb,
                    param_grid=param_grid_gb,
                    scoring='rmse',
                    n_jobs=-1)

gs_ml = gs_ml.fit(X_train, y_train)
gs_rf = gs_rf.fit(X_train, y_train)
gs_gb = gs_gb.fit(X_train, y_train)

rmse_ml = sqrt(mean_squared_error(y_test,gs_ml.predict(X_test)))
print("RMSE = ", rmse_ml)
print('ベストパラメータ:',gs_ml.best_params_)

rmse_rf = sqrt(mean_squared_error(y_test,gs_rf.predict(X_test)))
print("RMSE = ", rmse_rf)
print('ベストパラメータ:',gs_rf.best_params_)

rmse_gb = sqrt(mean_squared_error(y_test,gs_gb.predict(X_test)))
print("RMSE = ", rmse_gb)
print('ベストパラメータ:',gs_gb.best_params_)

```


- ・機械学習手法の検証

Machine_learning.py

```
import numpy as np
import pandas as pd
from statistics import mean
from math import sqrt
from sklearn.linear_model import LinearRegression
from sklearn.neural_network import MLPRegressor
from sklearn.ensemble import GradientBoostingRegressor, StackingRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import KFold

df = pd.read_csv('C:¥¥Users¥¥data.csv')

kf = KFold(n_splits = 10, shuffle = True, random_state = 1234)
for train_index, test_index in kf.split(df):
    train = df.iloc[train_index]
    test = df.iloc[test_index]
    Y_train = train['AfterScore']
    X_train = train.drop(columns = ['AfterScore'])
    Y_test = test['AfterScore']
    X_test = test.drop(columns = ['AfterScore'])

    lm_rmse_list = []
    ml_rmse_list = []
    rf_rmse_list = []
    gb_rmse_list = []

    reg1 = LinearRegression()
    reg2 = MLPRegressor(activation='relu',hidden_layer_sizes=(100,100), solver
='lbfgs')
    reg3 = RandomForestRegressor(max_features='auto',criterion = 'friedman_mse
',max_depth=3,n_estimators=10000)
    reg4 = GradientBoostingRegressor(max_features='log2',learning_rate=0.1,los
s='ls',max_depth=5,n_estimators=10000)

    model1 = reg1.fit(X_train, Y_train)
```

```
model2 = reg2.fit(X_train, Y_train)
model3 = reg3.fit(X_train, Y_train)
model4 = reg4.fit(X_train, Y_train)

rmse1 = sqrt(mean_squared_error(Y_test,model1.predict(X_test)))
rmse2 = sqrt(mean_squared_error(Y_test,model2.predict(X_test)))
rmse3 = sqrt(mean_squared_error(Y_test,model3.predict(X_test)))
rmse4 = sqrt(mean_squared_error(Y_test,model4.predict(X_test)))

lm_rmse_list.append(rmse1)
ml_rmse_list.append(rmse2)
rf_rmse_list.append(rmse3)
gb_rmse_list.append(rmse4)

print("model1 RMSE = ", mean(lm_rmse_list))
print("model2 RMSE = ", mean(ml_rmse_list))
print("model3 RMSE = ", mean(rf_rmse_list))
print("model4 RMSE = ", mean(gb_rmse_list))
```