内閣府経済社会総合研究所委託業務

AI 技術の導入が雇用環境へ及ぼす影響の評価手法に関する調査研究報告書（2023年度）

受託者：慶應義塾大学 SFC 研究所

1. はじめに

1.1 背景

本業務の目的は Michael Webb による論文「The Impact of Artificial Intelligence on the Labor Market」(2019) [1] に基づき、AI 技術が国内雇用環境へ及ぼす影響に関して、暴露度スコアに関する試算を行うことである。この目的を達成するため、上記論文を参考にした自然言語解析プログラムを作成し、それを用いて職業データベース JobTag (version 3.01 (2023 年 3 月公開) [2] ならびに特許データ提供サービス J-PlatPat [3]、JP-NET [4]、特許庁からダウンロードサービス [5] 経由で取得した全特許（2023 年 10 月取得時）を用いた暴露度スコアを計算してきた。

Webb 論文における暴露度スコア計算法は、特許データベースと職業データベースを入力とする。特許データベースからはタイトルを抽出して動詞-名詞ペアを取得する。一方、職業データベースからは職業のタスクを取得する。特許タイトルと職業タスクの重なり度合いを、両者から抽出した動詞名詞ペアにより計算する。

Webb 論文では特許データベースとして Google Patents Public Data (IFI CLAIMS Patent Services 提供) を用いており、職業データベースとして O*NET (US Department of Labor ならびにその後継である Dictionary of Occupational Titles (DOT) を用いている。

Webb 法の日本語版を作成するには、データとアルゴリズムに関する検討が必要である。なぜなら Webb 法で用いられた特許データならびに職業データは日本の特許や日本の職業ではなく、また、Webb 法で用いられた名詞・動詞ペアは英語では有効だが、日本語では有効だとは限らないからである。

1.2 アプローチ

1.2.1 職業データ

本研究では職業データとして、JobTag (日本語版 O*NET) を用いた。JobTag の正式名称は職業情報提供サイトであり、日本版 O*NET とも呼ばれる。JobTag のウェブサイトによれば、これは「ジョブ」（職業、仕事）、「タスク」（仕事の内容を細かく分解したもの、作業）、「スキル」（仕事をするのに必要な技術・技能）等の観点から職業情報を分かりやすく説明したものである。JobTag の目的は、求職者等の就職活動や企業の採用活動等を支援とのことである。

我々が用いた JobTag (version 3.1, 2023 年 3 月公開) には 484 件の職業が登録されている。その一例には「メガネ販売」がある。この中には「どんな仕事？」欄で仕事の概要が説明されており、「タスク（職業に含まれる細かな仕事）」欄で様々なタスク内容が説明されている。また「実施率」なるパラメタが各タスクに対して与えられている。「実施率」が取りうる値の範囲は 0%から 100%である

1.2.2 特許データ

本研究ではまず J-PlatPat, JP-NET を用いた特許取得を行った。データ検索に際しては、検索キーワードを設定する必要がある。そのキーワードに該当する特許データを得ることができる。ただし、毎回の取得可能データ件数に限りがある。J-PlatPat は 500 件程度、JP-NET は 10,000 件程度である。総計では J-PlatPat から 5,660 件、JP-NET から 190,256 件の特許を取得した。これらでは特許件数が十分ではなかったため、特許庁に依頼して全ての特許データを取得し、利用可能にした。この件数は 9,421,030 件だった。

1.2.3 アルゴリズム

入力データとして JobTag ならびに各種特許データを用いることを前節で述べた。それらのデータを用いて、職業データと特許データの重なり度合を計算するために、本研究では 2 つのアルゴリズムを構築した。
第一のアルゴリズムは名詞だけを用いる手法であり、これを NounPhraseExtractor と記載する。これは名詞の出現頻度を計数した後、頻度ベクトルを構成し、ベクトル間のコサイン類似度を計算する。
第二のアルゴリズムは「名詞」を「サ変動詞 or 動詞」する（例：ご飯を食べる、清掃を実施）ならびに「名詞」の「サ変名詞」（例：実験の評価、モデルの学習）となるペアの抽出である。これを MorePhraseExtractor と呼ぶ。NounPhraseExtractor は網羅性が高いが品質が低く、MorePhraseExtractor はその逆となる。そこで MorePhraseExtractor の網羅性を高めるために、我々は更に Word2Vec 法 [6] の利用を検討し、また、特許データ件数の増加を実施した。

2. 実施概要

2.1 期間、スケジュール、内容

(1) 実施期間

令和 5 年 8 月 15 日から令和 6 年 3 月 22 日

(2) 実施項目と実施スケジュール

本業務では 4 つのタームを設計し、各タームでは下記内容を実施する予定だった。一つのタームは 7 週だった。

1. ターム 1：環境構築、データのダウンロード、データ整形プログラムの作成、名詞動詞ペア抽出プログラムの作成と評価
2. ターム 2：暴露度スコアの試算、名詞動詞ペア抽出プログラムの改善
3. ターム 3：名詞動詞ペア抽出プログラムの改善
4. ターム 4：同上

(3) 実施内容
本件業務は具体的には次のように実施した。

1. ターム 1：環境構築、データのダウンロード、データ整形プログラムの作成、名詞動詞ペア抽出プログラムの作成と評価(8/22-10/12)
    1.1. 8/22-29：
        1.1.1. J-PlatPat, JP-NET、ダウンロード開始
    1.2. 8/29-9/5：
        1.2.1. ストップワード除去
        1.2.2. 人工知能技術により増減した職業の考察
    1.3. 9/5-12：
        1.3.1. NounPhraseExtractor の作成
        1.3.2. 9 月 8 日：第一回研究会実施
    1.4. 9/12-19
        1.4.1. JP-NET 調査
        1.4.2. 特許データ(J-PlatPat, JP-NET) のデータ理解、フォーマット理解
    1.5. 9/19-26
        1.5.1. 特許データ(J-PlatPat, JP-NET) のデータ理解、フォーマット理解を継続。
        1.5.2. WordNet の調査

1.5.3. 特許庁全件データ（以下、JPO データと表記）利用を検討
1.5.4. MorePhraseExtractor の設計を開始
1.6. 9/26-10/5
1.6.1. MorePhraseExtractor の実装を継続
1.6.2. Webb スコアの妥当性について考察
1.6.3. 日本語版 O*NET の正解データについて検討
1.7. 10/5-10/12
1.7.1. 論文調査
1.7.2. 従来論文のスコアを検討
2. ターム 2：暴露度スコアの試算、名詞動詞ペア抽出プログラムの改善(10/12-11/30)
2.1. 10/12-10/19
2.1.1. MorePhraseExtractor の実装を継続
2.1.2. Word2Vec 法の設計
2.2. 10/19-10/26
2.2.1. MorePhraseExtractor と Word2Vec の結合作業
2.2.2. 論文調査を継続
2.3. 10/26-11/2
2.3.1. MorePhraseExtractor と Word2Vec の結合によるゼロスコア問題の検討
2.3.2. LDA パイプライン完成
2.4. 11/2-11/9
2.4.1. Webb スコアの妥当性を検討
2.5. 11/9-11/16
2.5.1. 回帰スコアの検討
2.6. 11/16-11/23
2.6.1. Word2Vec における類義語計算の高速化に関して検討
2.6.2. Embedding と行列計算による高速化を検討
2.6.3. MorePhraseExtractor のリファクタリングを実施
2.7. 11/23-11/30
2.7.1. 暴露度スコアの試算完了
2.7.2. JPO データからのデータ取得方法を模索
3. ターム 3：名詞動詞ペア抽出プログラムの改善（11/30-1/18）
3.1. 11/30-12/7
3.1.1. 特許データ件数の増減とゼロスコア数の関係を調査
3.1.2. MorePhraseExtractor, NounPhraseExtractor, RAKE による結果の分析
3.2. 12/7-12/14
3.2.1. JPO データの HDD からマシンへの移動完了
3.2.2. 12/7：第二回研究会の事前説明

2.2 研究会開催履歴

表 1. 研究会開催履歴

	日時	議事
第一回研究会	2023 年 9 月 8 日	1. 暴露度スコアの計算方法を理解 2. 特許のデータと日本版 O*NET のデータを取得 3. 日本版暴露度スコアの算出方法を調査
第二回研究会	2023 年 12 月 11 日	1. 暴露度スコア計算の報告 2. 暴露度スコアの評価と分析 3. 課題と今後の進め方
第三回研究会	2024 年 2 月 5 日	1. 特許検索エンジン構築 2. 暴露度スコアのアップデート 3. Webb 論文と本研究成果の比較

2.3 有識者

- 新谷 元継（東京大学大学院経済学研究科・経済学部　教授）
- 山本 勲（慶應義塾大学商学部　教授）
- 宮崎 崇史（LINE ヤフー株式会社　リサーチ・エンジニア）
- 小山田 昌文（NEC データサイエンス研究所 主席研究員）

3. 使用したデータセット

本プロジェクトに参考にする [1] では、職業に関するデータを O*NET [7] から取得している。O*NET とは米国労働省が公開している職業情報データベースおよび職業情報サイトであり、米国職業分類に含まれる約 900 職種について、具体的な能力、必要な知識、向いている興味や価値観等を共通尺度上で数値化したデータを提供している。

本プロジェクトでは、日本国内における人工知能による暴露度スコアの計算を目的としているため、以下に示す日本版 O*NET (JobTag) [2] を用いた。

3.1 日本版 O*NET (JobTag)

3.1.1 日本版 O*NET (JobTag) とは

厚生労働省によって職業情報提供サイト JobTag（日本版 O*NET）が日本版 O*NET として開発された。2024 年 3 月 20 日現在、合計 549 職種に関する職業情報が提供されている。日本版 O*NET は、様々な職業について、職業の性質を**「タスク」（仕事の内容を細かく分解したもの、作業）、「スキル」（仕事をするのに必要な技術・技能）等**の観点から職業を解説し、求職者の就職活動や企業を採用活動の支援を目的としている。

以下に日本版 O*NET のトップページのスクリーンショットを示す。このように、求職者がいくつかの質問に答えることで適職を知ることができたり、キーワードやあらかじめ定められたテーマを選択することで該当する職業を調べたりすることができる。

図 1. JobTag（日本版 O*NET）のトップページ

3.1.2 職業ページ

本プロジェクトで職業に関する情報を抽出対象である職業ページについて述べる。日本版 O*NET のトップページから検索を進めていくと最終的には職業ページ（スクリーンショットを以下に示す）に辿り着く。このように、日本版 O*NET では各職業について、以下の情報を提供している。

- 「どんな職業か？」
  - どのような職業であるか一段落程度の自由記述テキスト
- 「タスク（職業に含まれる細かな仕事）」：
  - タスク記述：どのような作業を行うのか一文程度の自由記述テキスト。
  - 実施率：アンケートに基づいて集計した当該職業に対する重要性を示した数字である。これは 0 から 100%の間の数値で表現される。

本プロジェクトでは、これらの自由記述テキストから職業に関わる情報を抽出し、暴露度スコアの計算を行う。[1]に倣い、**タスク記述**と**実施率**を使用した。ただし、（オリジナルである）O*NET に比べて、タスク記述が短いため、代替または追加案として「どんな職業か？」を用いた結果も報告する。

なお、これらの情報はエクセルまたは CSV フォーマットでダウンロード可能であり、本プロジェクトでは提供されているデータを用いた。

図 2.　職業ページ：「どんな仕事？」の一例。ここでは豆腐製造、豆腐職人に関する「どんな仕事？」が記載されており、その仕事内容が記述されている。JobTag が提供する職業には、この項目は存在する。

図 3. 職業ページ：タスク記述と実施率の一例。ここでは豆腐製造、豆腐職人に関する「タスク内容」と「実施率」が記載されている。タスク内容は、この職業で行うタスクの内容を表す。実施率とは、その職業に就職している人の中で、そのタスクを実施している人の割合を表す。

3.1.3 タスク記述が欠損している職業

JobTag (version 3.1, 2023 年 3 月公開) には 484 職業が存在する。そののうち、37 職業に関してはタスク記述が欠損している。タスク記述がなければ、本業務で利用することができない。そのため、これらの職業は分析対象から除外した。その結果、**447 職業**を対象にすることとなった。これは[1]で用いた O*NET が提供する 964 職業に比べて少ない。これらの職業を表 2 に示す。

表 2. JobTag においてタスク記述が欠損している職業

タスク記述が欠損している職業	
1	国際公務員
2	産業用ロボット開発技術者
3	産業用ロボットの設置・設定
4	産業用ロボットの保守・メンテナンス
5	太陽光発電の企画・調査
6	太陽光発電の設計・施工
7	太陽光発電のメンテナンス
8	ネット通販の企画開発
9	植物工場の栽培管理
10	自動車組立
11	生産用機械組立
12	自動車技術者
13	精密機器技術者
14	電気技術者
15	電気通信技術者
16	医療用画像機器組立
17	織布工/織機オペレーター

18	木材製造
19	紡績機械オペレーター
20	食品技術者
21	貴金属装身具製作
22	医薬品製造
23	タイヤ製造
24	化粧品製造
25	石油精製オペレーター
26	ゲームクリエーター
27	アニメーター
28	ブックデザイナー
29	接客担当（ホテル・旅館）
30	営業（IT）
31	港湾荷役作業員
32	ドローンパイロット
33	セキュリティエキスパート（脆弱性診断）
34	NPO 法人職員（企画・運営）
35	データエンジニア
36	独立系ファイナンシャル・アドバイザー（IFA）
37	タンクローリー乗務員

3.2 特許文献データの収集

3.2.1 J-PlatPat

J-PlatPat [3] は、独立行政法人工業所有権情報・研修館が提供する産業財産権情報検索サービスである。キーワード、検索範囲文献、公知日を指定して検索を行い、500 件以下の場合、開発の名称と要約を含む CSV をダウンロードできる。表 3 の 13 個のキーワードに基づいて、5,660 件の特許情報を取得した。

表 3. J-PlatPat からの特許収集で用いたキーワード

キーワード
ウェブマイニング
オントロジー
ディープラーニング
データマイニング
ビッグデータ
モンテカルロ法
画像認識
機械学習
自然言語処理
人工神経回路網
人工知能
探索木

J-PlatPat では、要約データを含む特許情報を取得する場合、一度に取得できる件数は 500 件に限定されている。取得件数が多すぎるとエラーになってしまい、結果が一切得られない。そのため、J-PlatPat を用いて大規模なデータを収集することは困難である。今回は JogTag に対応するなるべく多数の特許を取得する必要があった。そのため、J-PlatPat を用いて、大規模な特許収集をすることは不可能であるとの結論を、研究実施中に得た。

3.2.2 JP-NET

JP-NET [4] とは、日本パテントサービス株式会社が提供する特許情報検索サービスである。キーワード、検索範囲文献、公開日を指定して検索を行うことができる。検索結果が 10,000 件以下の場合に限り、開発の名称と要約を含む CSV をダウンロード可能である。J-PlatPat 同様にダウンロード数に関する制限があるものの、JP-NET の制限は、J-PlatPat よりも緩く、本研究の目的によりふさわしいものだった。ただし、JP-NET は一回のダウンロードに 10 分程度の時間を必要とするため、多数回の試行には膨大な労力が必要だった。

特許を取得するために、表 4 の 5 個のキーワードに基づいて、特許データのダウンロードを、長時間を要しつつ、試行錯誤した。その結果、190,256 件の特許情報を取得することができた。上記の通り、JP-NET での一回の取得可能件数は 10,000 件であり、J-PlatPat よりも本業務の目的に適切ではあったものの、20 万件程度しか特許データを取得できなかった。

本研究提案の申請時には、J-PlatPat と JP-NET を用いて特許データを収集することを考えていた。そして、これらのデータを用いて、第二回研究会までは分析を行ってきた。しかしながら、特許データ件数の多寡により、暴露度スコアの品質に違いが生じる可能性を感じ、件数増加のための処理が必要となっていた。

表 4. JP-NET からの特許収集で用いたキーワード

キーワード
画像認識
機械学習
自動運転
質疑応答
人工知能

3.2.3 特許庁からの全件データ取得

J-PlatPat にしても JP-NET にしても、特許データを提供するものではあるが、その大元のデータは特許庁により管理されているものである。そして特許庁は、これらの特許データを、一括して提供している。このデータを特許庁データ（JPO データ）と記載する。

JPO データは、特許庁の一括ダウンロードサービスによって提供されるデータである。 特許庁の一括ダウンロードサービス [5] を使用することで、1993 年から 2023 年までの公開特許情報を取得可能だった。このデータサイズが大きすぎたために、ネットワーク経由でのダウンロードは不可能であり、それゆえに HDD を特許庁へ送付し、全ての特許データを当該 HDD へ複製していただき、それを返送していただく、という手順を取った。データ取得依頼を行ったのは、2023 年 10 月頃である。公表特許公報は 7 ヶ月後に公開されることを鑑みると、我々が取得したデータは 2023 年 3 月ごろまでに出願されたものだと考えられるだろう。

次、HDD のデータをコンピュータへと複製し、その内容を確認したところ、データは PDF 形式となっていた。この PDF に対して、テキストを抽出するプログラム PDF2TXT を適用しようとした。ところが、この適用には失敗した。その理由は、PDF に格納されていたデータが、いわゆる PDF フォーマットではなく、画像データを PDF に無理やり変換したものだったからである。

そこで PDF データに対して OCR 処理を施すことにより、テキストを抽出することとした。ただし抽出されたテキストには OCR により不適切なものも含まれてしまったため、このエラー処理も実施した。これらのデータには PDF フォーマットのみならず ISO フォーマットなども含まれていたが、それらもテキストへと変更をした。これらのデータを、一旦、ファイルシステムのディレクトリに保存をした。

3.2.4 特許検索システムの構築

特許データがテキスト形式になっていようとも、それが検索可能になっていなければ、J-PlatPat, JP-NET で実施したような、キーワード検索が不能である。JPO データは特許データから構成されるものの、その中には AI に深い関連のあるもののみならず、あらゆる特許が含まれているからである。この問題を解決するために、検索エンジンと呼ばれるソフトウェアに JPO データを登録することとした。検索エンジンとしては ElasticSearch [8] を選定した。ElasticSearch は、大量のデータ処理を目的とした検索エンジンであり、高速なキーワード検索や時間軸検索を可能にする。これにより、発明の名称と要約に対する任意のキーワードを検索可能にした。

様々な問題があったものの、最終的に JPO データを ElasticSearch へ格納し、検索可能にすることができた。特許はものによっては長大であり、それら全てを本研究では用いる訳ではない。本研究の目的は、あくまでも JobTag と全ての特許データの関連性を機械的に計算することであるから、特許文献の長短により、特許の特徴が変動して表現されることは不適切である。また、OCR 処理には膨大な時間を要するために、全ての特許情報を OCR 処理していては、本業務が契約期間内に終わら

ないことが確実であることが判明していた。この問題を解決するために、特許データの 1 ページ目だけを OCR 処理することとした。特許データの 1 ページ目には「要約」という欄が存在する。この情報を用いて特許情報の特徴を表現することとした。特許検索システム構築の流れを下図に示す。

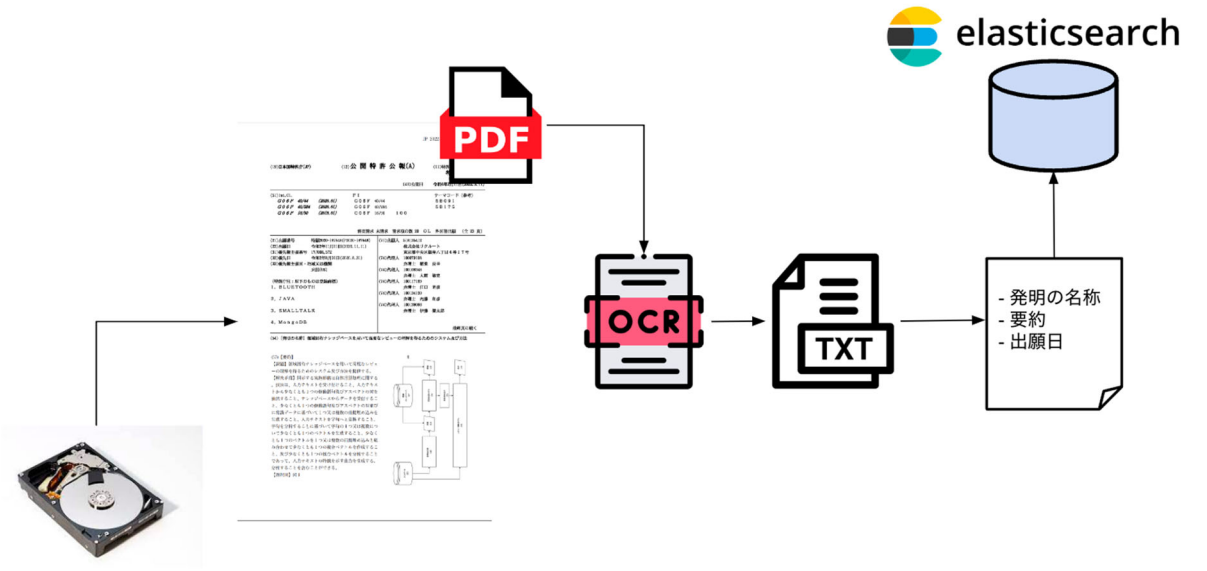

図 4.　OCR を用いた特許庁データからの構造化データ抽出と検索インデクス構築

ElasticSearch を用いて集計したところ、公開特許情報に含まれる公開特許と公表特許の総件数は 9,421,030 件だった。既存の特許検索サービスとの比較を表 5 に示す。我々が自作した特許検索システムは、従来のサービスよりも、遥かに容易に特許を検索・取得可能である。

表 5.　特許取得方法の比較

特許取得方法	一度に取得可能な件数	一回の応答時間	分析に利用可能な特許件数
J-PlatPat	500 程度	数分	5,660
JP-NET	10,000 程度	5~10 分	190,256
独自特許 DB (ElasticSearch)	無制限	数秒	9,421,030

この独自特許検索システムに対して、表 6 に示すキーワードを用いて特許データを収集した。このキーワードを選定した理由は、特許庁が公開している「AI 関連発明の出願状況調査 報告書」[9] の「別添 2（AI コアキーワード」に記載されていたからである。時期については 2014 年以降とした。理由は同報告書で、2014 年以降に AI に関する特許が急増していると述べられていたからである。

表 6. 独自特許データベースからの特許収集で用いたキーワード

キーワード			
機械学習	マシンラーニング	学習アルゴリズム	教師あり学習
教師なし学習	半教師あり学習	ニューラルネット	パーセプトロン
アクティブラーニング	サポートベクタマシン	バックプロパゲーション	ファインチューニング
過剰適合	シグモイド関数	深層学習	誤差逆伝播
オートエンコーダ	自己符号化	ボルツマンマシン	潜在表現
次元削減	強化学習	Q 学習	long short term
長短期記憶	敵対生成ネット	表現学習	転移学習
アンサンブル学習	遺伝的アルゴリズム	deep reinforcement	ネオコグニトロン
群知能	スワームインテリ	SVM	コネクショニスト
ランダムフォレスト	決定木	ベイジアンネット	勾配ブースティング
XGboost	AdaBoost	ロジティック回帰	勾配降下法
潜在セマンティクス	潜在ディリクレ	隠れマルコフ	特徴選択
ワードエンベッド	確率的アプローチ	ファジィ理論	混合ガウス
トピックモデル	チャットボット	ロボアドバイザー	オントロジー
エキスパートシステム	マルチエージェントシステム	帰納論理プログラミング	リカレントニューラル
知識ベース	人工知能	artificial intelligence	畳み込みニューラル
トランスフォーマ	machine learning	学習モデル	neural network
教師ありトレーニング	教師なしトレーニング	半教師ありトレーニング	レインフォースメントラーニング
コネクショニズム	過剰学習	活性化関数	ディープラーニング
deep Q	Q ラーニング	long short-term	長・短期記憶
generative adversarial network	サポートベクターマシン	ロジスティックリグレッション	潜在的セマンティクス

能動学習	スウォームインテリ	ディシジョントリー	ベイズモデル
gradient boost	XGBoost	Adaboost	ファジー理論
マルコフ確率	潜在的ディリクレ	確率論的アプローチ	トピックモデリング
ワードエンベッディング	リインフォースメントラーニング	ワードエンベディング	確率論的アルゴリズム
帰納プログラミング	計算知能	畳込ニューラル	再帰型ニューラル
教師あり訓練	教師なし訓練	半教師あり訓練	過適合
deep learning	deep belief	Q learning	ディシジョンツリー
ベイズモデリング	アダブースト	潜在意味	latent dirichlet
マルコフモデル	確率的テクニック	ファジィ論理	トピック分析
conventional neural	再帰的ニューラル	ベイズ推定	潜在的意味
マルコフモデリング	単語分散表現	確率論的テクニック	ファジー論理
トピックラベル	recurrent neural	潜在概念	マルコフネット
単語埋め込み	確率的手法	ファジィロジック	トピック抽出
潜在的概念	マルコフ過程	単語埋めこみ	確率論的手法
ファジーロジック	latent semantic	単語埋込	確率的方法
ファジィ制御	確率論的方法	ファジー制御	確率アルゴリズム
確率的アルゴリズム			

4.日本版暴露度スコアの算出

4.1 開発手法

Webb の手法に基づいて暴露度スコアを計算するには、特許データと職業データから名詞・動詞ペアを抽出する必要がある（e.g. automatic calculation）。Webb の論文で使用されたデータの言語は英語であり、単純なアルゴリズムで名詞・動詞ペアを抽出することが可能である。しかし、日本語の文章から名詞・動詞ペアと同等のフレーズを抽出するのは容易ではない。

4.1.1 係り受け解析とルールベースに基づく手法

日本語における係り受け解析と日本語版 Webb 手法 (JaWebbNaive)

[1] の日本語化において、名詞・動詞ペア抽出の基本アルゴリズムの日本語化は自明ではないため、日本語における係り受け解析について解説したのち、具体例と実装を用いた例を交えて名詞・動詞ペア抽出のアプローチとして妥当であるか述べる。

[1] では、使用した実装・ライブラリに関する記載はないものの、dependency parsing algorithm（係り受け解析アルゴリズム） (Honnibal and Johnson 2015) [10] を用いて名詞・動詞ペアを抽出したという記述がある。係り受け解析とは、自然言語処理における構文解析 (syntactic parsing) の一種であり、単語間の依存関係を解析し、構造化する解析手法の総称である。ただし、**そもそも言語が異なる上に、単語間の依存関係の種類や単語の分割方法について複数の異なる定義が存在するため**、完全に同様の係り受け解析を行うことは困難である。

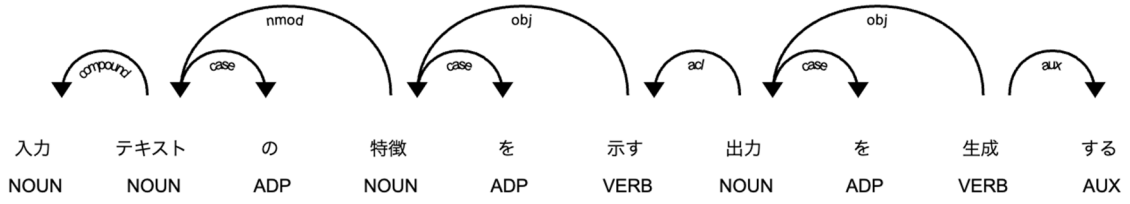

図 5. 係り受け解析の例

そこで本プロジェクトでは、Universal Dependency （UD）[11] と呼ばれる言語横断的に統一的な構造解析の仕組みを用いた言語解析器である GiNZA[12] を使用した。

具体的にはこの処理には以下の３つの処理が行われている。

1. 単語分割：入力文を単語に分割する。形態素解析とも呼ばれる。
2. 品詞タグ付与：ステップ１で分割された単語に対して品詞を推定し、付与する。この例では名詞（NOUN）、動詞（VERB）という品詞が付与されている。
3. 係り受け解析：単語間の依存関係を依存タグと共に付与する。この例では、依存関係は矢印で表され、たとえば obj（目的語）、case（助詞）といった依存タグが付与されている。

名詞（NOUN）と動詞（VERB）に注目をすると上記の例では、直接の依存関係が存在するペアは

- 示す (VERB) → 特徴 (NOUN): (obj)
- 生成 (VERB) → 出力 (NOUN): (obj)

の二つであり、それぞれ（特徴, 示す）、（出力, 生成）という名詞・動詞ペアの候補として抽出することができる。そのため、obj (object) 依存タグ

- 動詞 → 名詞: (obj)

というルールを対象に名詞・動詞ペアを抽出すればよいことがわかる。

実際に、発明の名称を係り受け解析した結果を示す。この例では「履歴及び時系列の共同分析に基づく異常の特性評価のためのシステム及び方法」という発明の名称に対して同様の係り受け解析を行った。

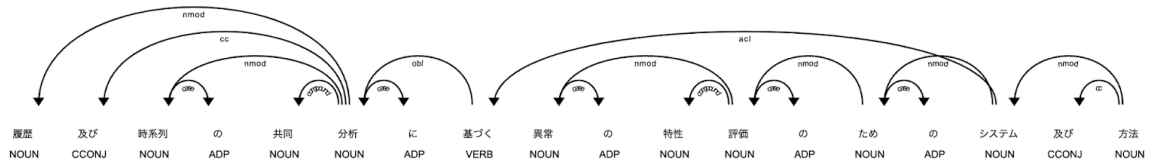

図 6. 複雑な係り受け解析の例

この例では、動詞と名詞が直接依存関係にあるのは

- 基づく→分析: (obl)

のみであり、obj 依存タグを伴う依存関係にある単語ペアは存在しない。なお、obl (oblique) 依存タグ は、当該単語に対する形容関係を表し（「分析」に「基づく」）、いわゆる動詞・名詞ペアの関係として抽出するには不適切である。

このように、日本語係り受け解析を用いて名詞・動詞ペア抽出を試みると、抽出の網羅性が低くなる恐れがある。これは後述するように日本語文法上は名詞が動詞化するサ変動詞のようなものが数多く出現するためである。そのため、**英語とは異なり、日本語において係り受け解析の愚直な利用では [1] で得られたような品質の名詞・動詞ペアの抽出は難しいと考えられる**。詳しくは提案アルゴリズムの解説および実験結果で述べる。

この係り受け解析によって特許データおよびタスク記述から名詞・動詞ペアを抽出し、[1]で開発された相対頻度に基づく手法に基づいて暴露度スコアを計算する。本稿ではこの手法を **JaWebbNaive** と呼ぶ。

4.1.2 MorePhraseExtractor の提案：品詞タグ抽出とルールベースに基づく手法

JaWebbNaive で抽出できる名詞・動詞ペアは限られている。なぜなら、英語の動詞に相当する日本語は「測る」のような動詞のほかに、「評価」などのサ変名詞があるからである。例えば、JaWebbNaive では「実験の評価を行う」という文から（評価, 行う）を抽出することができるが、（実験, 評価）は抽出できない。また、（評価, 行う）よりも（実験, 評価）が持つ情報の方が職業と特許の性質を反映していることや、英語の動詞に相当する日本語の単語は多くの場合、サ変名詞であることからわかるように、サ変名詞を含む名詞・動詞ペアを抽出できないことは正確な暴露度スコアを計算する上での大きな問題である。

以上の問題を解決するため、**MorePhraseExtractor** を開発した。
MorePhraseExtractor は「名詞」を「サ変動詞 or 動詞」する（例：ご飯を食べる、清掃を実施）ならびに「名詞」の「サ変名詞」（例：実験の評価、モデルの学習）となるペアを抽出することが可能な手法である。JaWebbNaive と違う点は、名詞と動詞の依存関係の他に、サ変動詞またはサ変名詞を含んだ依存関係の抽出が可能なことである。
これを実現するために、MorePhraseExtractor は係り受け解析による依存関係の特定だけでなく、品詞タグや直前の単語のチェックなどを行う。
また、既存の言語解析器では不適当な単語や記号を抽出する可能性があったので、名詞・動詞ペア抽出ルール以外にもルールを設定し、そのような問題を回避するよう設計されている。

MorePhraseExtractor により特許データと職業データから名詞・動詞ペアを抽出した後、[1]の式に従い暴露度スコアを計算する。

$$rf_c^t = \frac{f_c^t}{\sum_{c \in C^t} f_c^t}. \quad (1)$$

$$Exposure_{i,t} = \frac{\sum_{k \in K_i} \left[w_{k,i} \cdot \sum_{c \in S_k} rf_c^t \right]}{\sum_{k \in K_i} \left[w_{k,i} \cdot |\{c : c \in S_k\}| \right]}. \quad (2)$$

式（1）はあるキーワードの特許データ中の相対頻度を表す。式（2）の分子は職業データから抽出したキーワードの相対頻度の和とタスク実施率の積、分母は職業データから抽出したフレーズの個数とタスク実施率の積を表す。
なお、日本版 O*NET において、ある職業の全タスク実施率の和が 1 を超える場合があるので、暴露度スコア計算時にはタスク実施率に正規化を施した。

4.1.3 その他のキーワード抽出手法

MorePhraseExtractor はなるべく[1]で抽出されたようなフレーズと似た性質のフレーズを抽出する手法だが、日本語の文章では名詞が持つ意味が重要であることが多いという我々の知見から、NounExtractor と NounPhraseExtractor をさらに開発した。どちらも品詞タグの情報を用いることで単語の中から名詞を抽出する手法であるが、NounExtractor が単語の最小単位を抽出するのに対し（e.g. 報告書を書く -> “報告”, “書”を抽出）、NounPhraseExtractor は複合語の名詞を抽出するような違いがある（e.g. 報告書を書く -> “報告書”を抽出）。
キーワードを抽出した後は MorePhraseExtractor で抽出した際と同じように[1]の式に従い暴露度スコアを計算する。

4.1.4 キーワード抽出に基づく手法

RAKE を用いた手法 (RAKE-Extractor)

名詞・動詞ペア以外の代替手法として、各文から重要な意味を持つキーワードを抽出する手法の利用が考えられる。重要度の計算方法には様々な観点があり、ドメイ

ンによって大きく異なるが、汎用的かつ高速なキーワード抽出アルゴリズムとして Rapid Automatic Keyword Extraction (RAKE) [13] が知られている。

RAKE は意味を持たない機能語（日本語では、「て、に、を、は」といった助詞など）をストップワードとして除去した後に、単語をフレーズとしてまとめ上げ、フレーズ同士が同じ文に出現した回数を重みとする共起グラフを構築する。その共起グラフを用いて、ノードの次数（より多くのフレーズと共起しているか）を出現回数で正規化したものをスコアとして用いる。

本プロジェクトでは日本語実装である rake_ja [14] を用いた。

4.1.5 トピックモデルを用いた手法

[1]の特許データ中と職業データ中のキーワードの相対頻度から暴露度スコアを計算する手法の代替として、文章データのトピックを判断するトピックモデルが考えられる。トピックモデルの代表格として知られる LDA [15] では、文章中の単語の共起性に着目し、文章に潜む複数のトピックの分布をモデル化する。

本プロジェクトにおいては、NounExtractor を用いて、特許データと職業データから抽出した名詞に対して、トピック数 150 とする LDA を適用した。その結果として得られた、特許データ群の潜在トピック分布と各職業データの潜在トピック分布の類似度を、各職業の暴露度スコアとして算出した。

4.2 提案手法と Webb 手法の比較

提案手法と Webb 手法で最も違うのはキーワードの抽出アルゴリズムである。前述の通り、Webb 手法では dependency parsing algorithm のみで英文から目的の名詞・動詞ペアを取得しているが、提案手法は日本語の文章からキーワードを抽出するため、係り受け解析の他に品詞タグのチェックや様々なルールを組み合わせた抽出手法となっている。

また、Webb 手法では WordNet を使用することで英単語の類義語を考慮しながら暴露度スコアを計算するが、提案手法では類義語の判定を行っていない。これは、後述の通り日本語版の WordNet が意味の大きく異なる二つの単語を類義語として判定する可能性があるからである。

以上の違いがあるが、暴露度スコアの計算式は Webb 手法のものを再現している。また、職業データは Webb 手法で O*NET を使用するので、提案手法でもそれに対応する日本版 O*NET を用いる。特許データに関しては、特許庁より入手したものを使用する。これらの違いを表 7 に示す。

表 7. 提案手法と Webb 法の比較

	提案手法	**Webb 手法 [1]**
計算式	式（2）	式（2）
抽出	係り受け+品詞タグ+ルールベース	Dependency parsing algorithm
類義語	判定なし	判定あり（WordNet）
職業データ	日本版 O*NET（549 職種）	O*NET（約 900 職種）
特許データ	特許庁データ	Google Patents Public Data（US）
特許データ件数	約 942 万件	不明（論文に詳細記載なし）
特許データ検索条件	変更可能	不明（論文に詳細記載なし）

4.2.1 類義語を用いた表記揺れへの対応

抽出された似た意味を持つ動詞・名詞ペアが表記揺れによって全く異なる動詞・名詞ペアとして扱われてしまう可能性がある。

そのような表記揺れに対応するため、[1] では WordNet を用いて同じ意味を持つ類義語をまとめた。WordNet とは synset と呼ばれる同じ概念の語をクラスタにまとめ、synset 同士の関係を階層的に記述した概念辞書である。Synset を同じ意味を持つ類義語クラスタと捉えることで、WordNet を類義語辞書として用いることができる。

[1] では以下のように、動詞・名詞ペアを構成する各単語を conceptutal category (synset を指しているものと思われる) に対応づけると述べている。

> This allows me to assign each of the nouns occurring in my verb-noun pairs to a single conceptual category, for a given conceptual level. I use “aggregated verb-noun pair” to refer to a pair consisting of a verb and a noun conceptual category. For the conceptual level that includes “person”, for example, “recognize economist” would be part of the aggregated verb-noun pair “recognize person”.

しかしながら、ひとつの単語は複数の概念を持つことがあり、WordNet においてはある単語が複数の異なる synset に存在する可能性がある。そのため、[1] が述べる概念への対応づけは自明ではない。

WordNet は英語を対象言語としているため、日本語版である日本語 WordNet [16]を用いて予備実験を行った。具体的には、提案手法である MorePhraseExtractor を分析用に用いた JP-NET 全データに対して適用した抽出結果から、名詞、動詞それぞれについて同じ synset に含まれる単語を類義語として抽出した。

抽出フレーズ頻度上位である「計画」について日本語 WordNet を用いて類義語を取得した結果、「計画」は 5 つの異なる synset に登録されており、合計 55 個もの異なる類義語が取得された。取得された類義語を以下に示す。

['考える', '策略', '計', 'プランニング', '企図', '劃策', '狙う', '画策', '図る', '方略', '企劃', '腹積り', '策する', 'スキーム', '謀る', '企画', '目論見', '按ずる', '案ずる', 'プラン', '計画すること', 'ストラテジー', '立案', '工作', '策す', '目論む', '腹積もり', '籌策', '設計', '企てる', '企み', '巧む', '計図', '計略', '計策', '案', '後図', '構想', '企む', '作戦計画', '意図', '企て', '心積もり', '目論', '策', 'デザイン', '志す', '仕組む', '戦略', '企らむ', '企', 'プログラム', '目論み', '謀', '予定', 'ディザイン']

これを見てわかるように、「設計」、「工作」、「デザイン」、「プログラム」といった産業的観点において意味が大きく異なる単語が同じ概念として用いられていることがわかる。これらをまとめてしまうことにより、意味の粒度が粗くなり、暴露度スコア計算の質が低下するおそれがある。そのため、本プロジェクトでは（日

本語）WordNet を用いた概念への対応づけによる表記揺れの対応を行わないことにした。また、代替手法として Word2Vec を用いた表記揺れの対応を実装し、予備実験を行ったところ有効性を確認したが、計算量が多く、後に述べるとおり本プロジェクトではデータ数を増やすことで表記揺れの影響を小さくすることを目指した。

5. 日本版暴露度スコア計算システム

ここでは、開発した日本版暴露度スコア計算システムの構成と内容について述べる。本システムは表 8 に示す環境で開発を行った。

表 8. 開発環境

項目	内容
使用プログラミング言語（バージョン番号）	Python 3.10.12
使用ライブラリ（バージョン番号）	gensim==4.3.2 hydra-core==1.3.2 japanize-matplotlib==1.1.3 matplotlib==3.6.0 nltk==3.6.5 numpy==1.24.4 omegaconf==2.3.0 pandas==1.5.3 rake-ja==0.0.1 scikit-learn==1.3.0 scipy==1.8.0 seaborn==0.11.2 sentence-transformers==2.2.2 spacy==3.6.0 sphinx-rtd-theme==1.3.0 Sphinx==7.1.2 statsmodels==0.14.0 elasticsearch==8.12.0
開発行数	約 3,500 行
その他使用したツール	ElasticSearch（検索エンジン構築のため）

5.1 システム構成

システム構成図を以下に示す。当システムは以下の計算機能を持つ。

- タイトル・要約抽出
- 職業・タスク定義文抽出
- 暴露度スコア計算
- 暴露度スコア評価

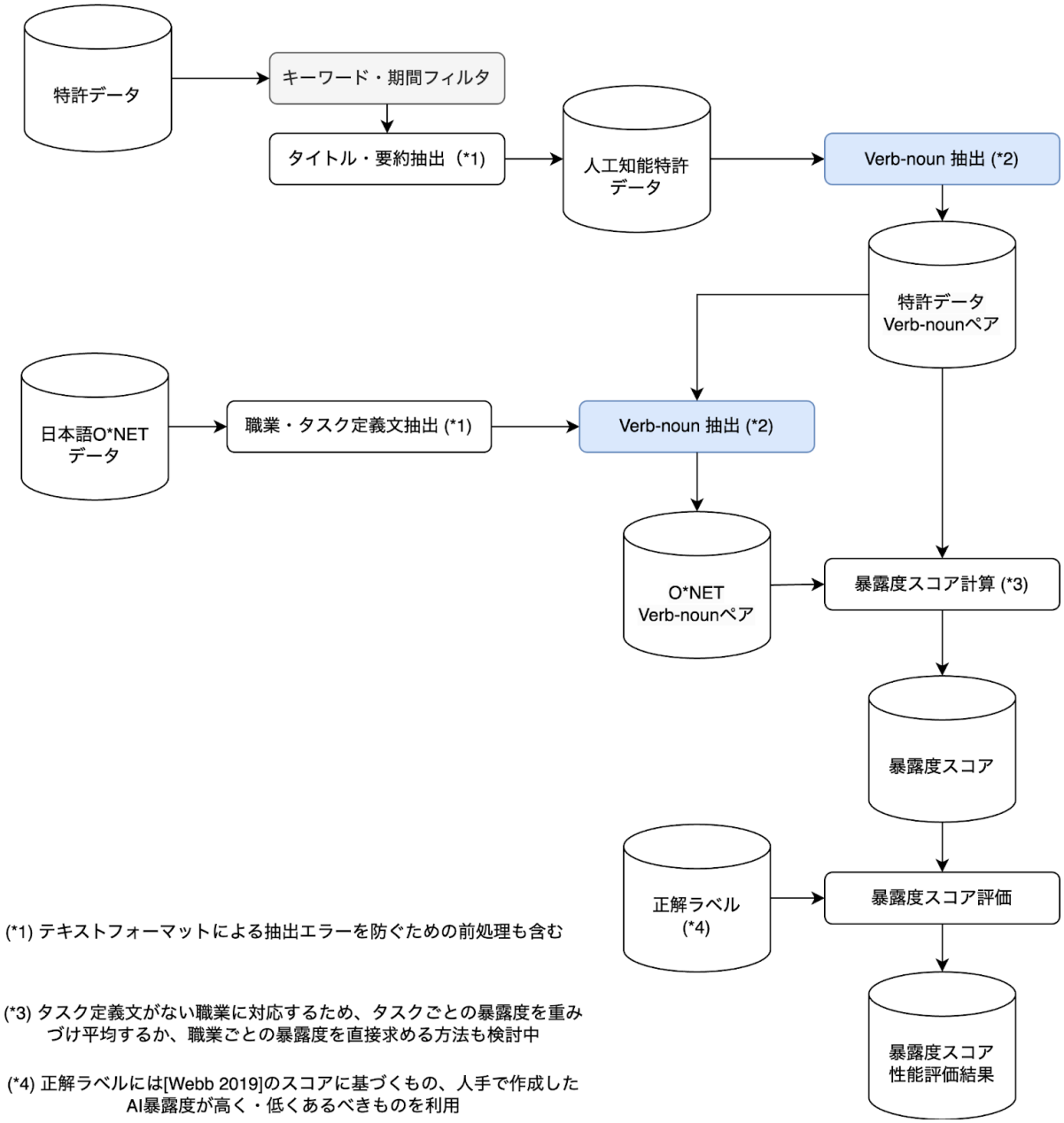

図 7. システム構成図

5.2 ソースコードと利用方法

本システムは Python コードで実装した。上述のとおり、本プロジェクトでは複数の手法を様々な設定で用いたため、これらの設定を全て設定ファイルとして記述し、コマンドを一回実行するだけで暴露度スコア及び各種分析結果のファイルが自動生成されるパイプラインを実装した。これにより、以下のような要件を満たすことができる。

- 特許データを変えた際に暴露度スコアの比較分析
- 手法の設定（例：閾値）を変更して暴露度スコアの再計算

- 新しい手法の実装とパイプラインへの追加

以下、具体的な設定ファイルの例を示す。システム図に示したとおり、本システムは暴露度スコア計算と、算出スコア評価の２つのステップから構成されるため、それぞれのステップに対する設定ファイル（YAML 形式）を用意する。あとは以下の２つのコマンドを実行するだけで、本報告書に記載されている結果を再現することができる。

```
$ python scripts/create_makefile.py
$ make
```

1 つ目のコマンドでは Makefile と呼ばれる、ファイル間の依存関係と生成方法を記載した make コマンド向けの設計書が生成され、make コマンドを用いることで各種計算、分析、評価に必要な Python プログラムが実行される。

```yaml
onet_loader:
  _target_: esri_labor_ai.ONETLoader
  desc_filepath: "data/ja-onet/IPD_DL_description_3_00_readable.csv"
  task_filepath: "data/ja-onet/IPD_DL_numeric_3_01_readable.csv"
patent_loader:
  _target_: esri_labor_ai.JPlatPatLoader
  filepath:
    - "data/patents/jplatpat/探索木__all__all.csv"
    - "data/patents/jplatpat/画像認識__2021__all.csv"
    - "data/patents/jplatpat/画像認識__2022__all.csv"
    - "data/patents/jplatpat/人工知能__2023__all.csv"
    - "data/patents/jplatpat/画像認識__2023__all.csv"
    - "data/patents/jplatpat/機械学習__all__A61.csv"
    - "data/patents/jplatpat/機械学習__all__G01.csv"
    - "data/patents/jplatpat/機械学習__all__G05.csv"
    - "data/patents/jplatpat/機械学習__all__G16.csv"
    - "data/patents/jplatpat/機械学習__all__H04.csv"
    - "data/patents/jplatpat/自然言語処理__2020__all.csv"
    - "data/patents/jplatpat/自然言語処理__2021__all.csv"
    - "data/patents/jplatpat/自然言語処理__2022__all.csv"
    - "data/patents/jplatpat/自然言語処理__2023__all.csv"
    - "data/patents/jplatpat/オントロジー__all__all.csv"
    - "data/patents/jplatpat/モンテカルロ法__all__all.csv"
    - "data/patents/jplatpat/人工神経回路網__all__all.csv"
    - "data/patents/jplatpat/ビッグデータ__all__all.csv"
    - "data/patents/jplatpat/データマイニング__all__A61.csv"
    - "data/patents/jplatpat/データマイニング__all__G06.csv"
    - "data/patents/jplatpat/データマイニング__all__G16.csv"
    - "data/patents/jplatpat/データマイニング__all__H04.csv"
    - "data/patents/jplatpat/ウェブマイニング__all__all.csv"
    - "data/patents/jplatpat/ディープラーニング__2020__all.csv"
    - "data/patents/jplatpat/ディープラーニング__2021__all.csv"
    - "data/patents/jplatpat/ディープラーニング__2022__all.csv"
    - "data/patents/jplatpat/ディープラーニング__2023__all.csv"
pipeline:
  _target_: esri_labor_ai.SimpleExtractionPipeline
  extractor:
    _target_: esri_labor_ai.MorePhraseExtractor
output:
  base_dir: outputs
  ai_exposure_score_filename: "ai_exposure_score.csv"
  extract_data_filename: "extract_data.csv"
```

図 8. 暴露度スコア計算設定ファイルの例

```yaml
evaluators:
  affected_occupation:
    _target_: esri_labor_ai.AffectedOccupationRankingEvaluator
    onet_filepath: "data/ja-onet/IPD_DL_description_3_00_readable.csv"
    golden_filepath: "data/labels/affected_nonaffected_occupations_20230911-v0.json"
  webb_ranking_topk:
    _target_: esri_labor_ai.WebbRankingEvaluator
    k: 20
    similarity_threshold: 0.7
  mapped_aiscore_threshold-None:
    _target_: esri_labor_ai.MappedAIScoreCorrelationEvaluator
    similarity_threshold: null
  mapped_aiscore_threshold-0.7:
    _target_: esri_labor_ai.MappedAIScoreCorrelationEvaluator
    similarity_threshold: 0.7
  zero_count:
    _target_: esri_labor_ai.ZeroCountEvaluator
analyzers:
  comparative_table:
    _target_: esri_labor_ai.ComparativeTableAnalyzer
  regression_analysis__komatsu_mugiyama_2021:
    _target_: esri_labor_ai.RegressionAnalysisAnalyzer
    preset_name: "komatsu_mugiyama_2021"
  regression_analysis__v1:
    _target_: esri_labor_ai.RegressionAnalysisAnalyzer
    features:
      - "規則的（ルーチンやスケジュールが決まっている）"
      - "芸術"
      - "他者の反応の理解"
      - "交渉"
      - "説得"
    feature_labels:
      - "規則的"
      - "芸術"
      - "理解"
      - "交渉"
      - "説得"
  regression_analysis__condition:
    _target_: esri_labor_ai.RegressionAnalysisAnalyzer
    features:
      - "社会的認知・地位"
      - "良好な対人関係"
      - "労働条件（雇用や報酬の安定性）"
      - "労働安全衛生"
      - "組織的な支援体制"
      - "専門性"
      - "奉仕・社会貢献"
      - "私生活との両立"
  regression_analysis__character:
    _target_: esri_labor_ai.RegressionAnalysisAnalyzer
    features:
      - "現実的"
      - "研究的"
      - "芸術的"
      - "社会的"
      - "企業的"
      - "慣習的"
      - "達成感"
      - "自律性"
  regression_analysis__education:
    _target_: esri_labor_ai.RegressionAnalysisAnalyzer
    features:
      - "高卒未満"
      - "高卒"
      - "専門学校卒"
      - "短大卒"
      - "高専卒"
      - "大卒"
      - "修士課程卒（修士と同等の専門職学位を含む）"
      - "博士課程卒"
    feature_labels:
      - "高卒未満"
      - "高卒"
      - "専門学校卒"
      - "短大卒"
      - "高専卒"
      - "大卒"
      - "修士課程卒"
      - "博士課程卒"
  ai_score_scatter_plot:
    _target_: esri_labor_ai.AIScoreScatterPlotAnalyzer
  score_distribution_analysis:
    _target_: esri_labor_ai.ScoreDistributionAnalyzer
output:
  eval_score_filename: "eval_score.csv"
```

図 9. 暴露度スコア評価設定ファイルの例

プログラム出力結果の例を以下に示す。このように 1 設定ファイルにつき、設定ファイルと同様の名称のディレクトリが作成され、そのディレクトリ以下に暴露度スコア（ai_exposure_score.csv）をはじめ、暴露度スコア評価結果および分析結果が eval ディレクトリ以下に保存される。

```
outputs/jpnet-all-001__morephrase-webb-001
├── ai_exposure_score.csv
├── calculate_score.log
├── eval
│   ├── ai_score_scatter_plot
│   │   ├── ai-score-scatter__felten_aioe.png
│   │   └── ai-score-scatter__webb_v1.png
│   ├── comparative_table
│   │   ├── felten_aioe__bottomk.csv
│   │   ├── felten_aioe__topk.csv
│   │   ├── webb_v1__bottomk.csv
│   │   ├── webb_v1__topk.csv
│   │   ├── webb_v2__bottomk.csv
│   │   └── webb_v2__topk.csv
│   ├── eval_score.csv
│   ├── evaluate_score.log
│   ├── regression_analysis__character
│   │   ├── regression_analysis.pdf
│   │   └── regression_analysis.png
│   ├── regression_analysis__condition
│   │   ├── regression_analysis.pdf
│   │   └── regression_analysis.png
│   ├── regression_analysis__education
│   │   ├── regression_analysis.pdf
│   │   └── regression_analysis.png
│   ├── regression_analysis__komatsu_mugiyama_2021
│   │   ├── regression_analysis.pdf
│   │   └── regression_analysis.png
│   ├── regression_analysis__v1
│   │   ├── regression_analysis.pdf
│   │   └── regression_analysis.png
│   └── score_distribution_analysis
│       ├── score_dist_density.png
│       ├── score_dist_hist.png
│       └── score_dist_stats.csv
├── extract_data.csv
├── task_weight_phrase_count.csv
└── task_weight_phrase_count_with_patents.csv
```

図 10. 本システムの出力結果の一例

6.算出スコアの評価と分析

6.1 定量評価を用いた予備実験

6.1.1 人手ラベルを用いた評価

人工知能技術の影響を受ける・受けない職業を人手で設定し、算出暴露度スコアがどれほど正解職業を正しくランキングしているかによって評価を行った。

- 影響を受ける職業の例
  - タクシー配車オペレーター、損害保険事務、通信販売受付事務、データ入力
- 影響を受けない職業の例
  - 引越作業員、すし職人、そば・うどん調理人

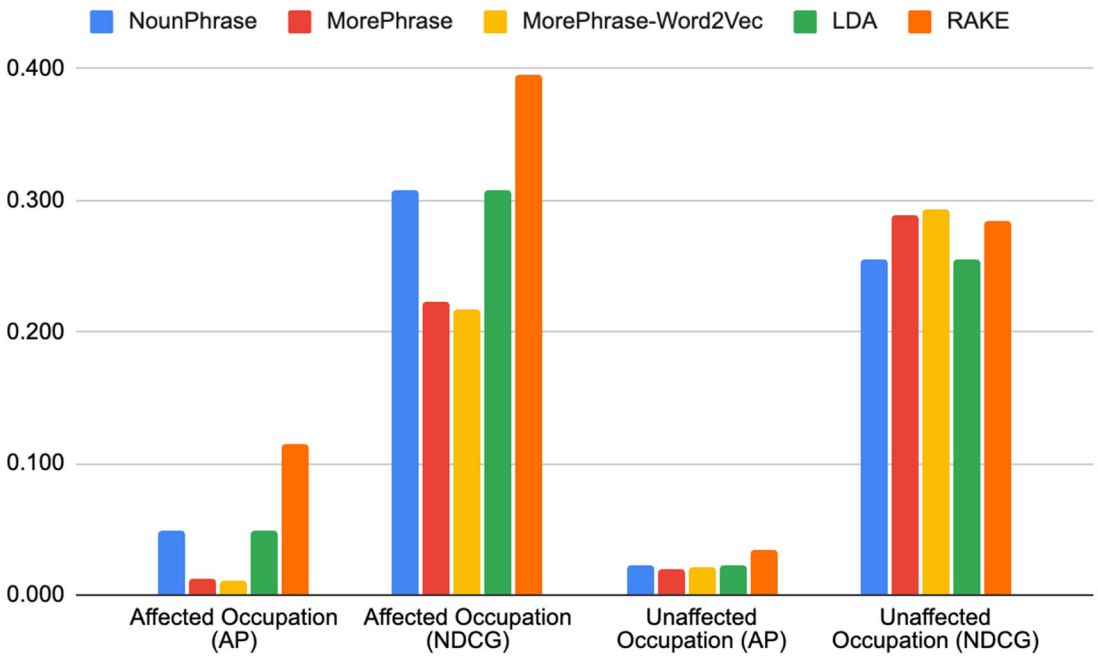

図 11. 人手ラベルを用いた評価

6.1.2 既存研究のデータを用いた評価

既存研究のラベルを正解データとみなして評価を行った。具体的には[1] [18] を利用した。これらの研究では英語と米国を対象言語・対象国としているため、そのままでは日本版 O*NET に適用できない。そこで職業名の一致を取るために、多言語埋め込みモデルを用いて職業名を一致させた。多言語埋め込みモデルは言語横断的にテキストを固定次元の埋め込みに写像するモデルであり、これを用いて英語、日本語で記述された職業名を埋め込み事件に写像、コサイン類似度が最も高いものを対応する職業として選択した。

評価指標としては以下の指標を用いた。

- １）既存スコアと算出スコアにおける上位・下位職業の一致率
- ２）相関係数（Pearson、Kendall、Spearman）

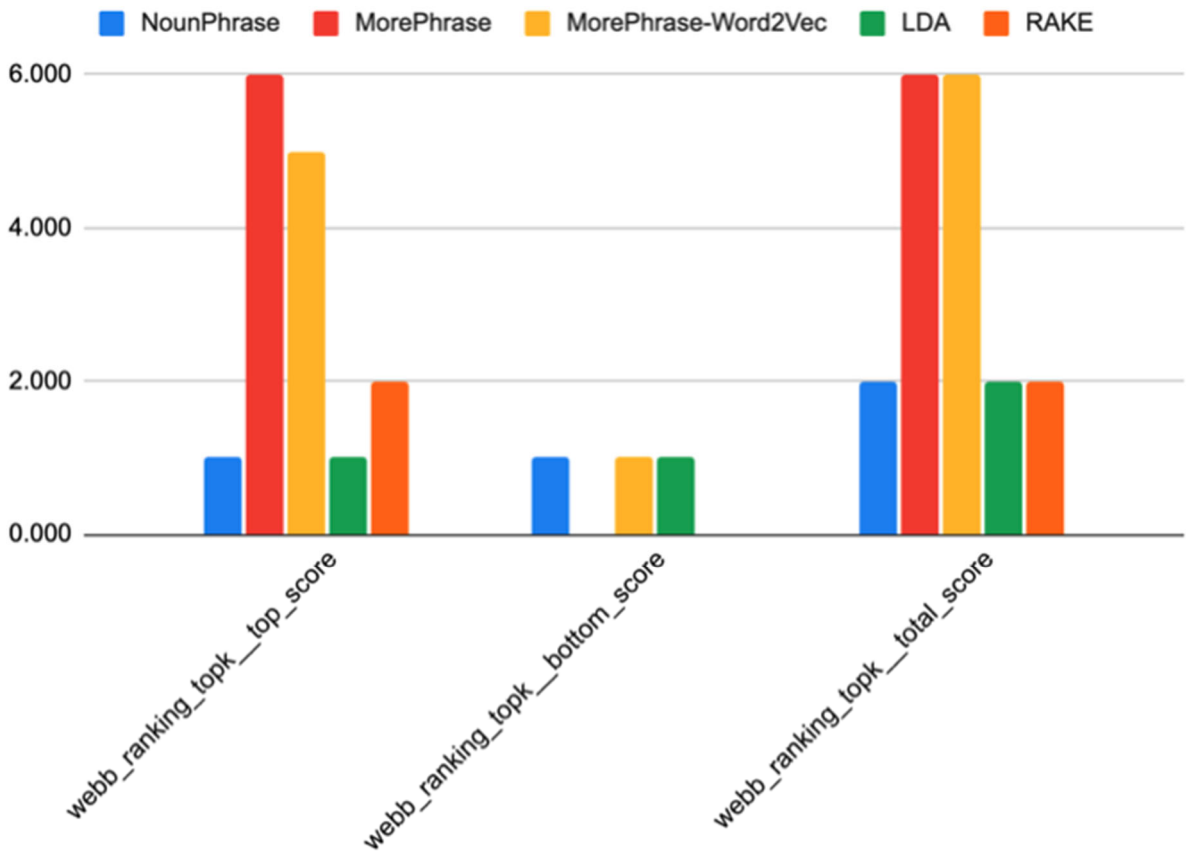

図 12.　既存研究のデータを用いた評価（上位・下位の一致率）

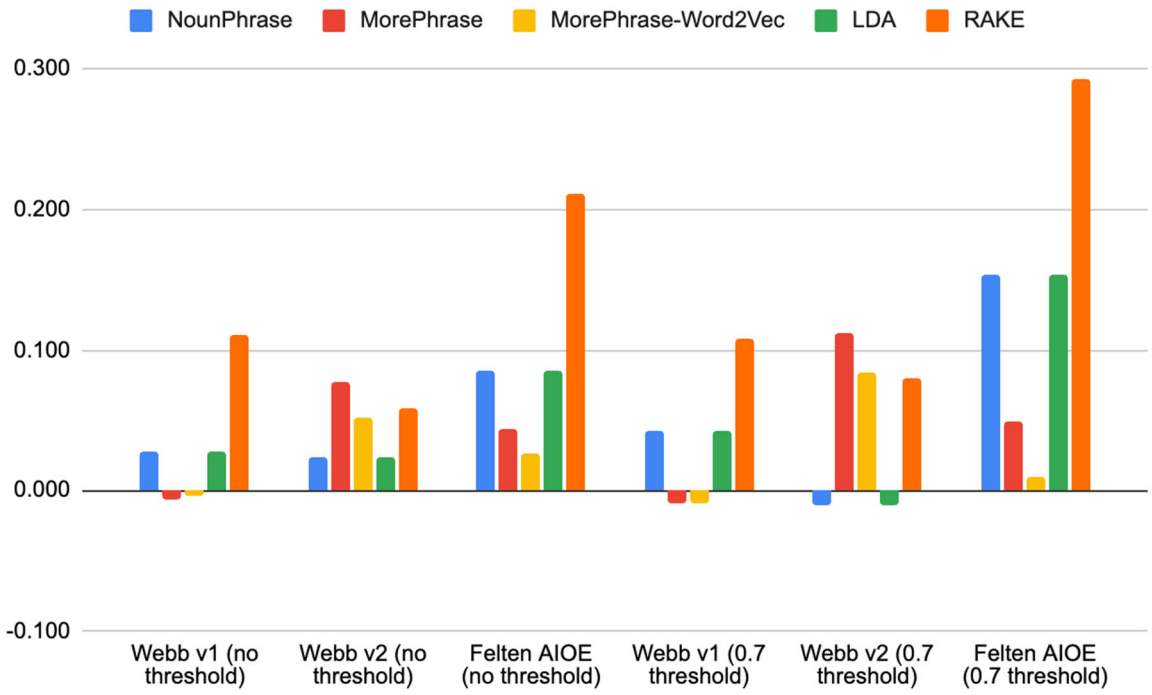

図 13. 既存研究のデータを用いた評価（相関係数）

6.1.3 分析に用いる特許文献データ量と算出スコア評価値の比較

[1] では Google Patent Data を使用したとだけ述べており、具体的な特許件数についての記述はない。そこで我々は暴露度スコア計算に用いる特許文献の件数を変化させることで、使用データ数と暴露度スコアの質の関係について基礎分析を行った。以下に J-PlatPat 全データと一部データを用いて算出された暴露度スコアの比較を示す。上述の[1] [18] との相関係数について、全データを用いた暴露度スコアの方が高い値を示した。これにより、**より多くの（人工知能技術に関わりのある）特許文献を利用することが重要である**ことが示唆された。この結果を受けて、本プロジェクトでは計算量が許す限り、より多くの特許文献データを用いる方針とした。

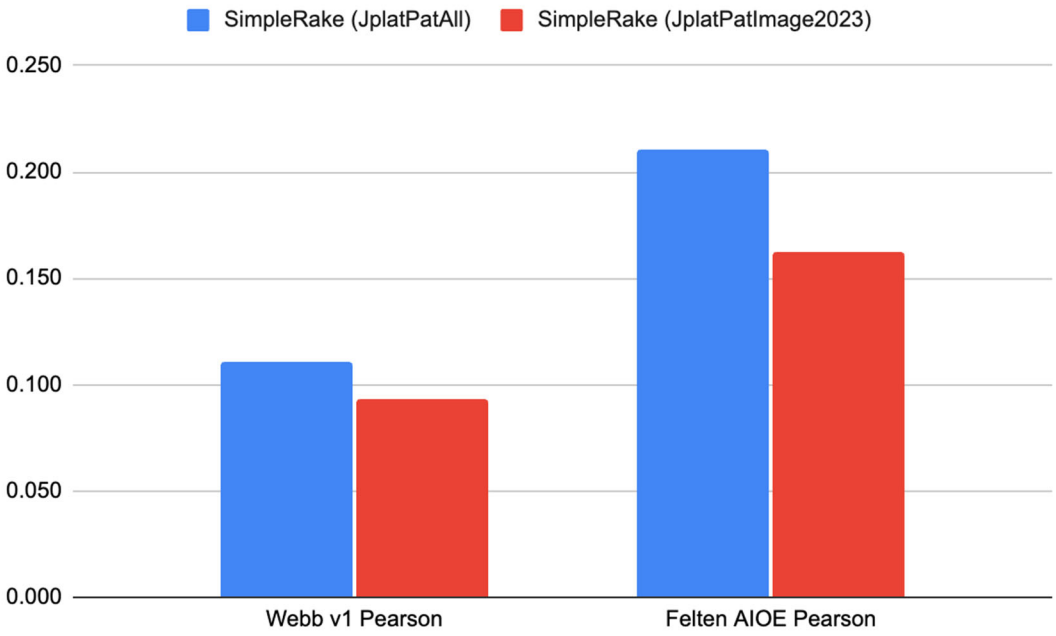

図 14. 特許文献データ量と算出スコア評価値の比較

6.1.4 定量評価により得られた知見

上記の定量評価を通じて以下の知見が得られた。

- 異なる評価指標は異なる傾向を捉えている。
  - 正確性 (precision) と網羅性 (coverage/recall) の二側面どちらに注力した指標であるか。
- より多くの特許文献データを使うことが重要
  - [1] では利用データのサイズ、フィルタ条件に関しては考察していない
  - 特許庁データを検索エンジンに格納することで柔軟な設計が可能

ただし、精度の高い定量評価を行うためには、品質の高いラベルデータが必要であり、本プロジェクトでは以下に述べる既存研究で用いられた分析手法、定性・比較分析を中心に算出暴露度スコアの評価を行うこととした。

6.2 Webb 手法の単純な日本語化の課題

さきほど係り受け解析の単純な適用では名詞・動詞ペアが適切に抽出されないおそれがあることを述べた。そこで、実際の特許データを用いて分析を行った。具体的には、JP-NET 全データを用いて以下の比較を行った。

- JaWebbNaive（係り受け解析を用いた日本版 Webb 手法）
- MorePhraseExtractor
- NounExtractor

また、職業タスク記述の抽出対象の妥当性を検証するため、以下の 2 つの比較を行った。

- 職業のタスク記述および実施率
- 「どんな職業？」（各職業につき記述はひとつなので実施率=1 として計算）

そのため、全 6 種類の組み合わせについて実験を行った。

6.2.1 結果: タスク記述を用いた場合

表 9. タスク記述を用いた場合の結果

JaWebbNaive	MorePhraseExtractor	NounExtractor
管理__する: 59	報告書__作成: 32	等: 424
手続き__する: 27	計画__立てる: 21	者: 418
計画__立てる: 24	相談__受ける: 20	ため: 309
手配__する: 23	書類__作成: 19	商品: 205
作業__する: 23	書__作成: 14	管理: 201
調整__する: 23	依頼__受ける: 13	書: 195
準備__する: 21	資料__作成: 13	客: 187
作業__行う: 20	指示__出す: 11	情報: 186
確認__する: 20	契約書__作成: 11	(: 185
清掃__する: 19	状況__把握: 10	作業: 170

6.2.2 結果: 「どんな職業？」を用いた場合

表 10. 「どんな職業？」を用いた場合の結果

JaWebbNaive	MorePhraseExtractor	NounExtractor
作業__行う: 74	依頼__受ける: 22	(: 999
管理__行う: 42	役割__果たす: 19	こと: 789
業務__行う: 33	計画__立てる: 18	情報: 776
指導__行う: 28	書類__作成: 17	者: 388
調整__行う: 22	状況__把握: 16	）: 358
開発__行う: 20	注意__払う: 16	場合: 208
設計__行う: 19	相談__受ける: 15	装置: 156
依頼__受け: 19	顧客__要望: 13	ため: 116
清掃__行う: 19	業務__担当: 12	データ: 100
販売__行う: 19	注文__受ける: 11	第: 72

6.3 暴露度スコア算出結果の分析

6.3.1 スコア分布

表 11. 算出暴露度スコアの統計値

統計値	スコア
count	447
mean	0.0702
std	0.0999
min	0
25%	0.0152
50%	0.0389
75%	0.0808
max	1.0000

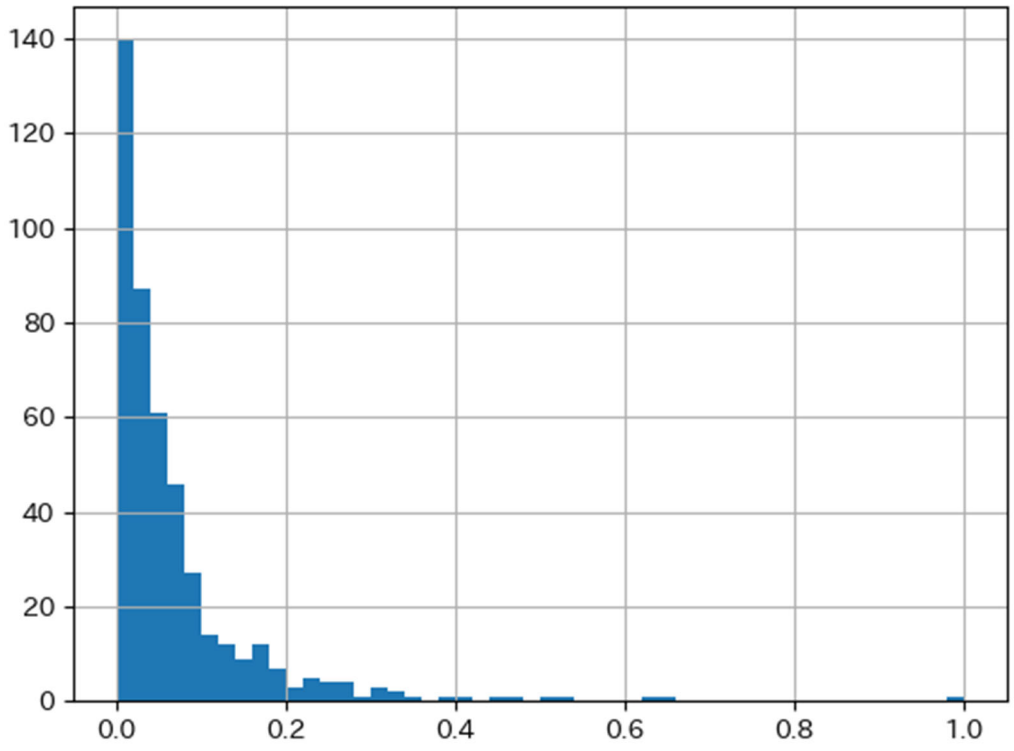

図 15.　暴露度スコアの分布

6.3.2 回帰分析を用いた暴露度スコアの妥当性検証

日本版 O*NET ではサーベイに基づいて各職業の性質（例：現実的、研究的、企業的）および従事者の学歴（例：高卒、大卒、修士課程卒）が数値化されている。[1] [17] における回帰分析を参考にして、これらの数値を**説明変数**、算出暴露度スコアを**被説明変数**とみなして回帰分析を行い、各職業の性質と人工知能による影響の強さの関係を分析した。本分析では、以下の属性を説明変数として用いた。

- （職業性質）現実的、研究的、芸術的、社会的、企業的、慣習的、達成感、自律性
- （学歴）高卒未満、高卒、専門学校卒、短大卒、高専卒、大卒、修士課程卒、博士課程卒
- （定型・非定型）非定型分析、非定型相互、定型認識、定型手仕事、否定型手仕事

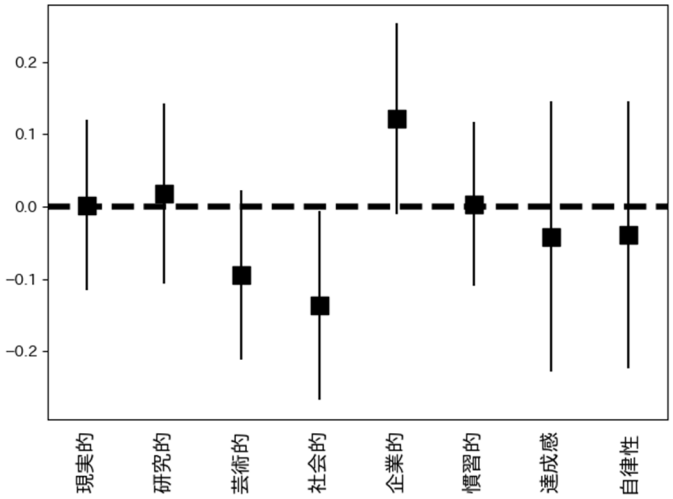

図 16. 職業の性質と暴露度スコアの関係

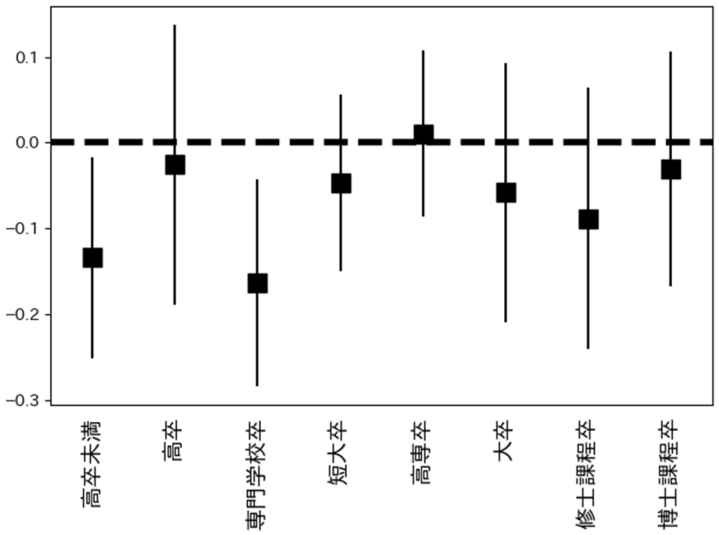

図 17. 学歴と暴露度スコアの関係

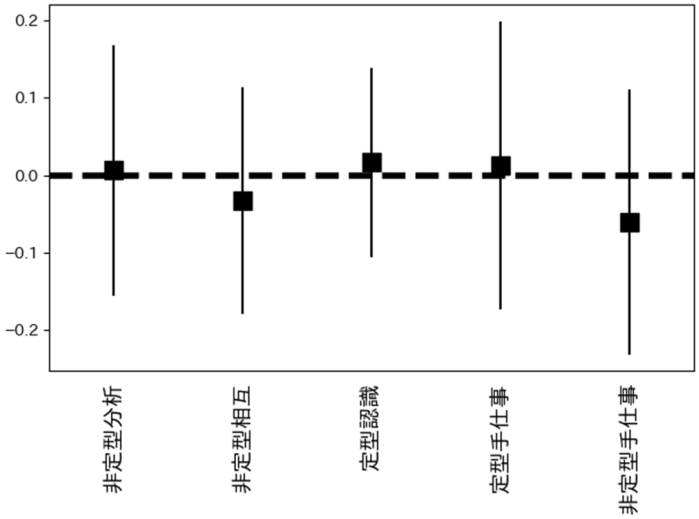

図 18. タスクの定型・非定型と暴露度スコアの関係

6.3.3 既存研究との比較による算出暴露度スコアの妥当性検証

予備実験では相関係数によって評価を行ったが、ここでは算出暴露度スコアを職業カテゴリにまとめることで定性的な分析を行う。具体的には以下の手順で**人手により日本版 O*NET の 511 職業を職業カテゴリに対応づけた**。O*NET の職業は職業大分類 23 カテゴリへの対応情報が米国労働省労働統計局によって定義されているので、職業カテゴリごとに暴露度スコアの平均を比較することで、算出暴露度スコアと[1][18]の結果の比較が可能となる。

以下に職業大分類 23 カテゴリを示す。

11-0000 Management Occupations
13-0000 Business and Financial Operations Occupations
15-0000 Computer and Mathematical Occupations
17-0000 Architecture and Engineering Occupations
19-0000 Life, Physical, and Social Science Occupations
21-0000 Community and Social Service Occupations
23-0000 Legal Occupations
25-0000 Educational Instruction and Library Occupations
27-0000 Arts, Design, Entertainment, Sports, and Media Occupations
29-0000 Healthcare Practitioners and Technical Occupations
31-0000 Healthcare Support Occupations
33-0000 Protective Service Occupations
35-0000 Food Preparation and Serving Related Occupations
37-0000 Building and Grounds Cleaning and Maintenance Occupations
39-0000 Personal Care and Service Occupations
41-0000 Sales and Related Occupations
43-0000 Office and Administrative Support Occupations
45-0000 Farming, Fishing, and Forestry Occupations
47-0000 Construction and Extraction Occupations
49-0000 Installation, Maintenance, and Repair Occupations
51-0000 Production Occupations
53-0000 Transportation and Material Moving Occupations
55-0000 Military Specific Occupations

図 19. 職業大分類 23 カテゴリの一覧（List of SOC Occupations）

図 20 に算出暴露度スコア平均の散布図を示す。[1] に対しては相関係数 0.298 (p-value = 0.179) と統計的優位性は得られなかったものの、正の相関の傾向が読み取れる。一方で [18] に対しては 0.541 (p-value < 0.01) と有意な相関が得られた。

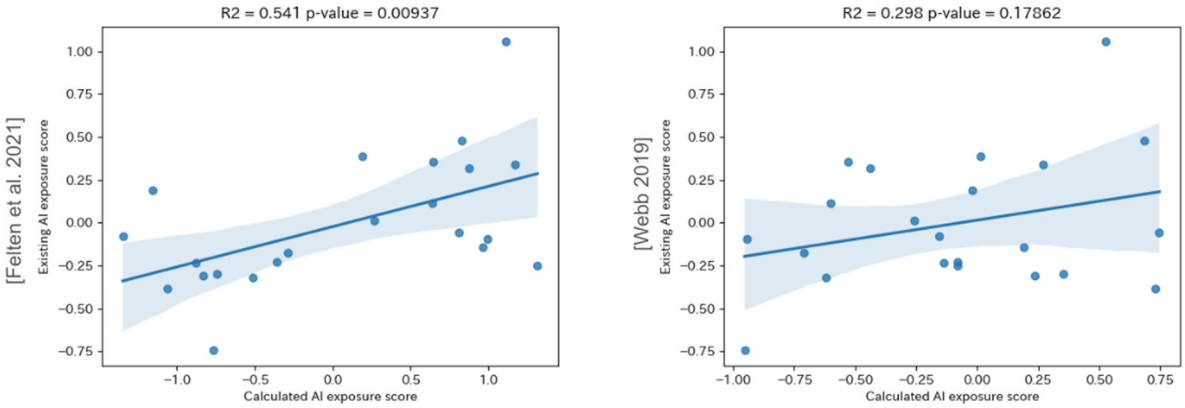

図 20. 職業大分類 23 カテゴリごとの算出暴露度スコアと既存暴露度スコアの比較（左: [18] 右: [1]）

6.4 定性評価

6.4.1 Webb の結果

ここでは定性評価を行う。Webb 論文における暴露度スコアの top-20, bottom-20 を下記に示す。Webb 論文では本研究とは異なる特許データと職業データを用いている。

表 12. Webb 法の結果（AI の影響を受けやすい 20 件、受けにくい 20 件)

Webb 論文の結果（[1]より抜粋）	
Top-20 (AI の影響を受けやすい 20 件)	Bottom-20 (AI の影響を受けにくい 20 件)
Clinical laboratory technologies and technicians	Animal caretakers, except for farm
Power plant operators	Mail carriers for postal service
Pest control occupations	Funeral directors
Purchasing agents and buyers of farm products	Art/entertainment performers and related occs
Dispatchers	Food preparation workers
Optometrists	Subject instructors, college
Construction inspectors	Shoemaking machine operators
Physicists and astronomists	Sales counter clerks
Chemical engineers	Clergy
Supervisors, forestry and logging workers	Hotel clerks
Locomotive operators: engineers and firemen	Sales demonstrators, promoters, and models
Marine engineers and naval architects	Barbers
Elevator installers and repairers	Correspondence and order clerks
Atmospheric and space scientists	Payroll and timekeeping clerks
Metallurgical and materials engineers	Postal clerks, excluding mail carriers
Knitters, loopers, and toppers textile operatives	Sales workers
Water and sewage treatment plant operators	Housekeepers, maids, butlers, and cleaners
Typesetters and compositors	File clerks
Production checkers, graders, and sorters in manufacturing	Bartenders
Textile cutting and dyeing machine operators	Locksmiths and safe repairers

6.4.2 本研究の結果と考察

提案手法 MorePhraseExtractor を、JobTag ならびに全特許（9,421,030 件）から表 6 に示したキーワードを用いて抽出した AI に関する特許（1,389,111 件）に適用し、暴露度スコアを計算した。ここではその下位 20 件、上位 20 件について示し、考察を行う。なお、全結果については 6.4.4 節に掲載する。

表 13. 提案手法の結果（AI の影響を受けにくい 20 件)

本研究の結果	
順位	Bottom-20（AI の影響をうけにくい）
1	きもの着付指導員
2	すし職人
3	中華料理調理人
4	酪農従事者
5	日本料理調理人（板前）
6	IT コンサルタント
7	解体工
8	新聞配達員
9	バーテンダー
10	会社経営者
11	そば・うどん調理人
12	映像編集者
13	コールセンターオペレーター
14	新聞記者
15	ジュエリーデザイナー
16	メイクアップアーティスト
17	ラーメン調理人
18	保温工事
19	ソムリエ
20	葬祭ディレクター

6.4.2.1 Bottom-20

AI による影響が少ないと考えられる、bottom-20 の結果を表 13 に示す。なお、暴露度スコアが 0 だった 4 件の職業（検察事務官、イラストレーター、麻薬取締官、通訳ガイド）は除外している。

この結果に対して考察を行う。これらの中には職人的な職業が下記のように存在する。この中で 1〜5 は調理に関する職業である。

1. すし職人
2. 中華料理調理人
3. 日本料理調理人（板前）
4. そば・うどん調理人
5. ラーメン調理人
6. ジュエリーデザイナー
7. メイクアップアーティスト
8. 映像編集者

また、これらの中には身体を多く用いる職業が下記のように存在する。調理に関する職業（上記 1〜5）も身体を用いるものだが、それらを除いたものだけ示す。

1. 酪農従事者
2. 解体工
3. 新聞配達員
4. 保温工事

上記とは別に、これらの中には、コミュニケーション能力が必要な職業が下記のように存在する。なお、本業務で用いた特許データは、前述の通り、2022 年 4 月までに出願されたものだけである。そのため、対話などに関して優れた能力を有する AI 技術は、特許としてはあまり存在していない可能性がある。

1. 着物着付け指導員
2. バーテンダー
3. コールセンターオペレーター
4. ソムリエ

最後に、判断が求められる職業が bottom-20 には下記のように含まれる。

1. IT コンサルタント
2. 新聞記者
3. 会社経営者
4. 葬祭ディレクター

以上より、調理を行う職業、職人的職業、身体を用いる職業、ヒトとコミュニケーションを行う職業、そして判断を行う職業については、AI による影響が少ないという結果が得られたと考えられる。

なお、暴露度スコアが 0 だった 4 件の職業（検察事務官、イラストレーター、麻薬取締官、通訳ガイド）のうち、検察事務官、麻薬取締官、イラストレーターについては、職人的素養があると考えられ、検察事務官、麻薬取締官、通訳ガイドについては、コミュニケーションに関する素養が必要だと考えられる。

6.4.2.2 Top-20

AI による影響が大きいと考えられる、top-20 の結果を表 14 に示す。

表 14. 提案手法の結果（AI の影響を受けやすい 20 件)

本研究の結果	
順位	Top-20（AI の影響を受けやすい）
1	データ入力
2	めっき工
3	自転車販売
4	地方公務員（行政事務）
5	CG 制作
6	生産・品質管理技術者
7	旅行会社カウンター係
8	観光バスガイド
9	化粧品訪問販売
10	ネット通販の運営
11	船員
12	林業作業
13	検査工（工業製品）
14	工場労務作業員
15	データサイエンティスト
16	シューフィッター
17	外務公務員（外交官）
18	食品営業（食品メーカー）
19	キッティング作業員 （PC セットアップ作業員）
20	産業廃棄物処理技術者

この結果に対して考察を行う。なお、本研究ではあくまでも影響を受けるタスクが存在するかどうかという観点から暴露度スコアが計算されていることに注意される必要がある。まず、コンピュータを用いてルーティンワークを行う下記の職種は影響を受けやすいと考えられる。

1. データ入力
2. CG 製作
3. データサイエンティスト
4. キッティング作業員（PC セットアップ作業員）

他の職業については、陽にコンピュータを用いるとは書かれていないものの、ルーティンワークに関することがタスク内容に記述されている。ルーティンワークは自動化されやすい作業である。そのため、ある程度影響を受けているように考えられる。他方、「林業作業」「外務公務員（外交官）」のように、影響が低いと考えられる物も存在する。

6.4.2.3 Webb 法の課題

ここに、動詞・名詞ペアだけを抽出している Webb 法の限界があると考えられる。本研究の結果についても、Webb の結果についても、低暴露度スコア、すなわち AI に影響を受けにくい職業については、直観的に解釈が容易である。ところが、AI に影響を受けやすい職業については、頷けるものもあるが、そうではない物も存在する。

この限界、あるいは精度に関する課題は、Webb 法のみならず、JaWebbNaive 法、ならびに、MorePhraseExtractor 法でも同様に存在する。精度を向上させるならば、職業のタスク記述から一部を抽出するよりも、その全体の意味を把握し、それを以って、特許データ記述との突き合わせを行う方が好ましいと考えられる。なぜなら記述から一部のみを抽出することで、重要な情報が失われる可能性があるからである。

Webb 法の優れた点は、暴露度スコア計算を、専門家が有する知識により形成された仮説に基づくのではなく、ヒトの能力では咀嚼困難な程の莫大なデータを機械に自動処理させるデータ駆動に基づく点である。このアプローチには発展の余地がある可能性が、本研究により示唆された。

6.4.3 タスクごとの名詞・動詞の寄与率の分析

[1] では、抽出された名詞・動詞ペアが最終的な暴露度スコアにどのように寄与しているのかを定性分析するために以下のような分析を行っていた（Table 1）。

Table 1: **Tasks and exposure scores for precision agriculture technicians.**

Task	Weight in occupation	Extracted pairs	AI exposure score x100
Use geospatial technology to develop soil sampling grids or identify sampling sites for testing characteristics such as nitrogen, phosphorus, or potassium content, ph, or micronutrients.	0.050	(develop, grid)	0.050
		(identify, site)	0.234
		(test, characteristic)	0.084
Document and maintain records of precision agriculture information.	0.049	(maintain, record)	0.000
Analyze geospatial data to determine agricultural implications of factors such as soil quality, terrain, field productivity, fertilizers, or weather conditions.	0.048	(analyze, datum)	0.469
		(determine, implication)	0.837
Apply precision agriculture information to specifically reduce the negative environmental impacts of farming practices.	0.048	(apply, information)	0.000
		(reduce, impact)	0.151
Install, calibrate, or maintain sensors, mechanical controls, GPS-based vehicle guidance systems, or computer settings.	0.045	(maintain, sensor)	0.000
Identify areas in need of pesticide treatment by analyzing geospatial data to determine insect movement and damage patterns.	0.038	(identify, area)	0.234
		(analyze, datum)	0.469
		(determine, movement)	0.502

Notes: Table displays six of the twenty-two tasks recorded for precision agriculture technicians in the O*NET database. For each task, the weight is an average of the frequency, importance, and relevance of that task to the occupation, as specified in O*NET, with weights scaled to sum to one. The verb-noun pairs in the third column are extracted from the task text by a dependency parsing algorithm. The AI exposure score for an extracted pair is equal to the relative frequency of similar pairs in the titles of AI patents. The score multiplied by 100 is thus a percentage; for example, pairs similar to "determine implications" represent 0.84% of pairs extracted from AI patents.

図 21.　[1] における Table 1 の抜粋

本プロジェクトでも、同様の分析を行うための分析ロジックを実装し、各職業・各タスクについて、使用されたアルゴリズムで抽出された名詞・動詞ペアのスコアを抽出し、分析を行った。表 15 では、暴露度スコアによって人工知能の影響を最も受けると判断された職業「**データ入力**」についてタスク記述と抽出された動詞・名詞ペア及びその相対頻度を示す。このようにそれぞれのタスクに対して、どのような名詞・動詞ペアがどのように暴露度スコアに寄与しているのかを可視化できる。

表 15. タスク記述と暴露度スコアに寄与した抽出フレーズの例

職業	タスク記述	実施率（正規化後）	抽出フレーズ	相対頻度（x1000)
データ入力	新しく担当する案件の内容や納期、分担を確認する。	0.0731	(分担, 確認)	0.0000
	キーボードを操作し、文字や数値などのデータをコンピュータに入力する。	0.3866	(データ, 入力)	0.8999
			(キーボード, 操作)	0.0000
	データ入力のための専用端末を操作する。	0.1607	(端末, 操作)	0.0262
	ＯＣＲ（光学式文字読取装置）でデータを読み取り、結果を確認する。	0.0293	(データ, 読む)	0.2240
		0.0293	(結果, 確認)	0.0260
	データを要求された通りの形式にとりまとめ、指定の場所に保存する。	0.1677	(データ, 要求)	0.0291

6.4.4 日本版暴露度スコア計算結果

447 件の職業に対する暴露度スコアの計算結果を表 16 に示す。暴露度スコアが高いほど、その職業は AI による影響が大きいと考えられる。

表 16. 日本版暴露度スコア

職業名	暴露度スコア
データ入力	0.000295398
めっき工	0.000193991
自転車販売	0.000186359
地方公務員（行政事務）	0.000153762
CG 制作	0.000153123
生産・品質管理技術者	0.000137081
旅行会社カウンター係	0.000133081

観光バスガイド	0.000123676
化粧品訪問販売	0.000115136
ネット通販の運営	0.00010543
船員	9.78E-05
林業作業	9.50E-05
検査工（工業製品）	9.39E-05
工場労務作業員	9.36E-05
データサイエンティスト	9.14E-05
シューフィッター	8.78E-05
外務公務員（外交官）	8.18E-05
食品営業（食品メーカー）	8.11E-05
キッティング作業員（PC セットアップ作業員）	8.05E-05
産業廃棄物処理技術者	7.91E-05
キャリアカウンセラー/キャリアコンサルタント	7.67E-05
マーケティング・リサーチャー	7.58E-05
遊園地スタッフ	7.53E-05
バックヤード作業員（スーパー食品部門）	7.29E-05
鉄骨工	7.07E-05
光学機器組立	6.92E-05
薬剤師	6.75E-05
鉄道運転計画・運行管理	6.58E-05

日本語教師	6.54E-05
非破壊検査技術者	6.47E-05
証券アナリスト	6.26E-05
営業課長	6.01E-05
診療放射線技師	5.84E-05
運用・管理（IT）	5.70E-05
建築施工管理技術者	5.63E-05
社会保険労務士	5.60E-05
銀行・信用金庫渉外担当	5.43E-05
経理課長	5.34E-05
航空管制官	5.32E-05
清涼飲料ルートセールス	5.29E-05
舞台照明スタッフ	5.26E-05
キャディ	5.22E-05
学校事務	5.18E-05
フォークリフト運転作業員	5.06E-05
人事課長	5.05E-05
Web マーケティング（ネット広告・販売促進）	5.02E-05
航空機開発エンジニア（ジェットエンジン）	5.02E-05
フロント（ホテル・旅館）	4.96E-05
ピアノ調律師	4.80E-05

ハンバーガーショップ店長	4.75E-05
中小企業診断士	4.73E-05
原子力技術者	4.68E-05
製版オペレーター、DTP オペレーター	4.61E-05
ビデオレンタル店店員	4.58E-05
製品包装作業員	4.57E-05
医療ソーシャルワーカー	4.39E-05
施設管理者（介護施設）	4.39E-05
介護タクシー運転手	4.38E-05
商品企画開発（チェーンストア）	4.36E-05
医薬情報担当者（MR）	4.26E-05
金属プレス工	4.11E-05
はり師・きゅう師	4.04E-05
バイオテクノロジー研究者	4.00E-05
一般事務	3.95E-05
臨床開発モニター	3.94E-05
総務課長	3.92E-05
広報コンサルタント	3.90E-05
非鉄金属製錬技術者	3.83E-05
理学療法士（PT）	3.68E-05
雑誌記者	3.68E-05

インダストリアルデザイナー	3.65E-05
AI エンジニア	3.62E-05
冷凍加工食品製造	3.50E-05
ビル施設管理	3.50E-05
医学研究者	3.46E-05
ペストコントロール従事者（害虫等防除・駆除従事者）	3.46E-05
ディスパッチャー（航空機運航管理者）	3.45E-05
印刷オペレーター	3.39E-05
広告営業	3.36E-05
計器組立	3.35E-05
ハウス野菜栽培者	3.19E-05
IR 広報担当	3.18E-05
診療情報管理士	3.12E-05
企業法務担当	3.08E-05
放送記者	3.07E-05
エレベーター据付	3.02E-05
柔道整復師	2.95E-05
鍛造工/鍛造設備オペレーター	2.94E-05
貿易事務	2.93E-05
ピッキング作業員	2.88E-05
建設・土木作業員	2.84E-05

農業技術者	2.81E-05
発電所運転管理	2.80E-05
鉄筋工	2.76E-05
バイオテクノロジー技術者	2.76E-05
サッシ取付	2.73E-05
臨床工学技士	2.71E-05
老人福祉施設生活相談員	2.67E-05
測量士	2.63E-05
中学校教員	2.63E-05
法務技官（心理）（矯正心理専門職）	2.63E-05
通信販売受付事務	2.61E-05
ファイナンシャル・プランナー	2.55E-05
建具製造	2.55E-05
介護事務	2.54E-05
ディーラー	2.53E-05
証券外務員	2.53E-05
CAD オペレーター	2.51E-05
経営コンサルタント	2.45E-05
弁理士	2.40E-05
プロジェクトマネージャ（IT）	2.40E-05
路線バス運転手	2.39E-05

入国警備官	2.38E-05
染色工/染色設備オペレーター	2.36E-05
鉄道線路管理	2.35E-05
せり人	2.34E-05
海上保安官	2.30E-05
コンプライアンス推進担当	2.28E-05
看護助手	2.26E-05
野菜つけ物製造	2.25E-05
商社営業	2.21E-05
調剤薬局事務	2.21E-05
送電線工事	2.16E-05
救急救命士	2.16E-05
メガネ販売	2.15E-05
パラリーガル（弁護士補助職）	2.14E-05
録音エンジニア	2.11E-05
機械設計技術者	2.10E-05
代理店営業（保険会社）	2.10E-05
引越作業員	2.08E-05
特許審査官	2.08E-05
営業事務	2.07E-05
保健師	2.06E-05

児童相談所相談員	2.05E-05
タイル工	2.03E-05
細胞検査士	2.03E-05
分析化学技術者	2.03E-05
リフレクソロジスト	2.03E-05
ボイラーオペレーター	2.00E-05
プラスチック成形	1.98E-05
障害者福祉施設指導専門員（生活支援員、就労支援員等）	1.98E-05
広告デザイナー	1.98E-05
福祉事務所ケースワーカー	1.95E-05
看護師	1.95E-05
パイロット	1.95E-05
さく井工/ボーリング工	1.91E-05
トラック運転手	1.91E-05
生産・工程管理事務	1.89E-05
気象予報士	1.88E-05
総務事務	1.88E-05
医療機器開発技術者	1.88E-05
福祉用具専門相談員	1.86E-05
観光バス運転手	1.86E-05
産婦人科医	1.81E-05

国際協力専門家	1.81E-05
型枠大工	1.80E-05
マンション管理員	1.80E-05
道路パトロール隊員	1.79E-05
客室乗務員	1.78E-05
ガソリンスタンド・スタッフ	1.76E-05
ホテル・旅館支配人	1.75E-05
土木施工管理技術者	1.73E-05
銀行等窓口事務	1.71E-05
化学製品製造オペレーター	1.69E-05
Web デザイナー	1.68E-05
専門学校教員	1.67E-05
人事コンサルタント	1.67E-05
ブロック積み	1.65E-05
防水工	1.65E-05
航海士	1.65E-05
ビル清掃	1.63E-05
外科医	1.63E-05
電子機器技術者	1.62E-05
航空整備士	1.59E-05
自動車営業	1.58E-05

不動産鑑定士	1.57E-05
福祉ソーシャルワーカー	1.57E-05
作業療法士（OT）	1.56E-05
林業技術者	1.55E-05
学童保育指導員	1.54E-05
通訳者	1.53E-05
印刷営業	1.51E-05
テレビ・ラジオ放送技術者	1.51E-05
家電修理	1.51E-05
高分子化学技術者	1.51E-05
小児科医	1.51E-05
半導体製造	1.48E-05
内部監査人	1.48E-05
児童指導員	1.46E-05
ごみ収集作業員	1.44E-05
ガラス食器製造	1.42E-05
鉄道車掌	1.42E-05
内科医	1.42E-05
精神科医	1.42E-05
大工	1.41E-05
犬訓練士	1.41E-05

歯科衛生士	1.39E-05
電子機器組立	1.39E-05
セキュリティエキスパート（情報セキュリティ監査）	1.38E-05
フラワーショップ店員	1.37E-05
助産師	1.36E-05
システムエンジニア（業務用システム）	1.36E-05
マンション管理フロント	1.36E-05
果樹栽培者	1.36E-05
配管工	1.32E-05
倉庫作業員	1.32E-05
国家公務員（行政事務）	1.31E-05
ファインセラミックス製造技術者	1.30E-05
ファンドマネージャー	1.29E-05
フランチャイズチェーン・スーパーバイザー	1.28E-05
建築板金	1.24E-05
建築設計技術者	1.23E-05
住宅・不動産営業	1.23E-05
図書館司書	1.23E-05
電気工事士	1.22E-05
しょうゆ製造	1.20E-05
ビール製造	1.20E-05

報道カメラマン	1.20E-05
石工	1.19E-05
言語聴覚士	1.18E-05
金型工	1.18E-05
海上自衛官	1.16E-05
消防官	1.16E-05
溶接工	1.15E-05
弁護士	1.14E-05
NC 工作機械オペレーター	1.14E-05
靴製造	1.14E-05
宇宙開発技術者	1.14E-05
土木設計技術者	1.13E-05
テクニカルライター	1.13E-05
ハウスクリーニング	1.13E-05
経理事務	1.12E-05
雑誌編集者	1.11E-05
水産ねり製品製造	1.10E-05
薬学研究者	1.09E-05
入国審査官	1.05E-05
建築塗装工	1.04E-05
内装工	1.03E-05

レンタカー店舗スタッフ	1.02E-05
治験コーディネーター	1.00E-05
広報・PR 担当	9.96E-06
リサイクルショップ店員	9.95E-06
施設介護員	9.78E-06
こん包作業員	9.73E-06
刑務官	9.67E-06
陸上自衛官	9.65E-06
セキュリティエキスパート（オペレーション）	9.61E-06
プラント設計技術者	9.58E-06
臨床検査技師	9.50E-06
水族館飼育員	9.41E-06
警察官（都道府県警察）	9.16E-06
労働基準監督官	9.02E-06
動物看護	9.00E-06
電車運転士	8.92E-06
M&A マネージャー、M&A コンサルタント/M&A アドバイザー	8.63E-06
公認会計士	8.60E-06
税務事務官	8.45E-06
清酒製造	8.42E-06
半導体技術者	8.40E-06

損害保険事務	8.39E-06
宅配便配達員	8.39E-06
Web ディレクター	8.31E-06
情報工学研究者	8.24E-06
人事事務	8.21E-06
特別支援学校教員、特別支援学級教員	8.19E-06
みそ製造	8.03E-06
飲食チェーン店店員	7.88E-06
法務教官	7.85E-06
歯科医師	7.82E-06
施設警備員	7.80E-06
広告ディレクター	7.79E-06
紙器製造	7.71E-06
鋳造工/鋳造設備オペレーター	7.55E-06
ヘルプデスク（IT）	7.55E-06
視能訓練士	7.49E-06
ファッションデザイナー	7.42E-06
ツアーコンダクター	7.27E-06
パタンナー	7.24E-06
潜水士	7.19E-06
プログラマー	7.19E-06

税理士	7.14E-06
スーパー店員	7.09E-06
左官	7.08E-06
土地家屋調査士	6.91E-06
クリーニング師	6.89E-06
自動車教習指導員	6.87E-06
保育士	6.84E-06
ハム・ソーセージ・ベーコン製造	6.82E-06
とび	6.72E-06
企画・調査担当	6.65E-06
学習塾教師	6.64E-06
花き栽培者	6.63E-06
豆腐製造、豆腐職人	6.57E-06
学芸員	6.51E-06
コピーライター	6.47E-06
西洋料理調理人（コック）	6.40E-06
カウンセラー（医療福祉分野）	6.29E-06
義肢装具士	6.27E-06
アウトドアインストラクター	6.24E-06
ネイリスト	6.22E-06
駅務員	6.16E-06

司法書士	6.07E-06
製本オペレーター	6.01E-06
通関士	5.99E-06
商業カメラマン	5.98E-06
アナウンサー	5.94E-06
職業訓練指導員	5.91E-06
高等学校教員	5.84E-06
畜産技術者	5.81E-06
自動運転開発エンジニア（自動車）	5.65E-06
受付事務	5.62E-06
水産技術者	5.43E-06
ソフトウェア開発（パッケージソフト）	5.34E-06
ソフトウェア開発（スマホアプリ）	5.32E-06
駅構内売店店員	5.30E-06
惣菜製造	5.29E-06
デジタルビジネスイノベーター	5.28E-06
鉄道車両清掃	5.26E-06
稲作農業者	5.24E-06
航空自衛官	5.23E-06
英会話教師	5.23E-06
積卸作業員	5.22E-06

沿岸漁業従事者	5.21E-06
行政書士	5.18E-06
起業、創業	5.17E-06
土木・建築工学研究者	5.15E-06
動画制作	4.98E-06
放送ディレクター	4.90E-06
スポーツインストラクター	4.88E-06
建設機械オペレーター	4.79E-06
ベビーシッター	4.77E-06
システムエンジニア（組込み、IoT）	4.68E-06
携帯電話販売	4.63E-06
知的財産サーチャー	4.53E-06
コンビニエンスストア店員	4.50E-06
システムエンジニア（Web サイト開発）	4.45E-06
物流設備管理・保全	4.44E-06
スーパー店長	4.42E-06
アートディレクター	4.41E-06
ミシン縫製	4.38E-06
銀行支店長	4.33E-06
タクシー運転手	4.30E-06
社会教育主事	4.30E-06

美容師	4.27E-06
タクシー配車オペレーター	4.15E-06
化粧品販売/美容部員	4.09E-06
ワイン製造	4.00E-06
パン製造、パン職人	3.97E-06
スーパーレジ係	3.95E-06
給食調理員	3.87E-06
動物園飼育員	3.72E-06
医療事務	3.68E-06
家庭裁判所調査官	3.58E-06
インテリアデザイナー	3.56E-06
洋菓子製造、パティシエ	3.49E-06
秘書	3.46E-06
裁判官	3.38E-06
保険営業（生命保険、損害保険）	3.33E-06
デパート店員	3.31E-06
ホールスタッフ（レストラン）	3.29E-06
陶磁器製造	3.26E-06
乳製品製造	3.17E-06
自動車整備士	3.17E-06
ダンプカー運転手	3.15E-06

産業廃棄物収集運搬作業員	3.15E-06
フラワーデザイナー	3.10E-06
家政婦（夫）	2.98E-06
CD ショップ店員	2.88E-06
ブライダルコーディネーター	2.84E-06
造園工	2.81E-06
家具製造	2.80E-06
OA 機器営業	2.77E-06
トリマー	2.74E-06
調理補助	2.72E-06
空港グランドスタッフ	2.67E-06
調教師	2.64E-06
小学校教員	2.61E-06
医薬品販売/登録販売者	2.54E-06
図書編集者	2.47E-06
エステティシャン	2.45E-06
栄養士	2.43E-06
訪問介護員/ホームヘルパー	2.41E-06
電器店店員	2.33E-06
看板制作	2.24E-06
獣医師	2.23E-06

陶磁器技術者	2.22E-06
スポーツ用品販売	2.19E-06
テレビカメラマン	2.14E-06
検察官	2.09E-06
知的財産コーディネーター	1.99E-06
理容師	1.98E-06
かばん・袋物製造	1.91E-06
幼稚園教員	1.72E-06
スクールカウンセラー	1.64E-06
和菓子製造、和菓子職人	1.62E-06
トレーラートラック運転手	1.62E-06
あんまマッサージ指圧師	1.61E-06
客室清掃・整備担当（ホテル・旅館）	1.61E-06
船舶機関士	1.57E-06
かん詰・びん詰・レトルト食品製造	1.48E-06
駐車場管理	1.43E-06
書店員	1.39E-06
厩舎スタッフ	1.39E-06
舞台美術スタッフ	1.35E-06
介護支援専門員/ケアマネジャー	1.25E-06
手話通訳者	1.25E-06

グラフィックデザイナー	1.17E-06
システムエンジニア（基盤システム）	1.16E-06
インテリアコーディネーター	1.12E-06
ルート配送ドライバー	1.12E-06
翻訳者	1.08E-06
アロマセラピスト	1.08E-06
ホームセンター店員	1.03E-06
ペットショップ店員	1.03E-06
検針員	1.03E-06
音楽教室講師	1.02E-06
ベーカリーショップ店員	1.01E-06
衣料品販売	1.00E-06
送迎バス等運転手	9.95E-07
歯科技工士	9.25E-07
雑踏・交通誘導警備員	9.01E-07
フードデリバリー （料理配達員）	8.82E-07
カフェ店員	8.49E-07
葬祭ディレクター	8.03E-07
ソムリエ	7.99E-07
保温工事	7.79E-07
ラーメン調理人	7.04E-07

メイクアップアーティスト	6.88E-07
ジュエリーデザイナー	6.69E-07
新聞記者	6.10E-07
コールセンターオペレーター	5.39E-07
映像編集者	5.36E-07
そば・うどん調理人	4.92E-07
会社経営者	4.86E-07
バーテンダー	4.72E-07
新聞配達員	3.80E-07
解体工	3.07E-07
IT コンサルタント	2.51E-07
日本料理調理人（板前）	2.27E-07
酪農従事者	2.19E-07
中華料理調理人	1.53E-07
すし職人	1.17E-07
きもの着付指導員	8.18E-08
通訳ガイド	0
麻薬取締官	0
イラストレーター	0
検察事務官	0

7. まとめ

本業務の目的は“The Impact of Artificial Intelligence on the Labor Market,” Michael Webb, (2019) に基づき、AI 技術が国内雇用環境へ及ぼす影響に関して、暴露度スコアに関する試算を行うことだった。

この調査のため、上記論文を参考にした自然言語解析プログラム MorePhraseExtractor を作成し、それを用いて職業データベース JobTag (version 3.01 (2023 年 3 月公開) ならびに特許データを用いて分析を行った。特許データ利用のために、独自のデータベースを構築し、900 万件を超える特許を indexing した。

JobTag と AI に関する 180 万件程度の特許を対象に MorePhraseExtractor を用いて関係性を分析した結果、低暴露度スコアの職業として、調理を行う職業、職人的職業、身体を用いる職業、ヒトとコミュニケーションを行う職業、そして判断を行う物が出力された。一方、高暴露度の職業として、比較的頭脳的なルーティンワークを行う物が出力された。

今後の課題は、利用する語句の増加による、精度の向上である。

参考文献

[1] Webb, Michael, The Impact of Artificial Intelligence on the Labor Market (November 6, 2019). Available at SSRN: https://ssrn.com/abstract=3482150 or http://dx.doi.org/10.2139/ssrn.3482150

[2] 職業情報提供サイト（日本版 O*NET）JobTag. https://shigoto.mhlw.go.jp/User/

[3] 特許情報プラットフォーム J-PlatPat. https://www.j-platpat.inpit.go.jp/

[4] 特許情報検索サービス JP-NET, 日本パテントデータサービス株式会社. https://www.jpds.co.jp/jp-net/jp-net.html

[5] 特許情報の一括ダウンロードサービスについて, 特許庁, https://www.jpo.go.jp/system/laws/sesaku/data/download.html

[6] Mikolov, Tomas; Sutskever, Ilya; Chen, Kai; Corrado, Greg S.; Dean, Jeff (2013). Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems. Bibcode:2013arXiv1310.4546M.

[7] O*NET. https://www.onetonline.org/

[8] ElasticSearch. https://www.elastic.co/jp/

[9] AI 関連発明の出願状況調査報告書, 特許庁審査第四部審査調査室, 2023 年 10 月. https://www.jpo.go.jp/system/patent/gaiyo/sesaku/ai/document/ai_shutsugan_chosa/hokoku.pdf

[10] Matthew Honnibal and Mark Johnson. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1373–1378, Lisbon, Portugal. Association for Computational Linguistics. https://aclanthology.org/D15-1162/

[11] Universal Dependencies. https://universaldependencies.org/

[12] GiNZA - Japanese NLP Library. https://megagonlabs.github.io/ginza/

[13] Rose, Stuart & Engel, Dave & Cramer, Nick & Cowley, Wendy. (2010). Automatic Keyword Extraction from Individual Documents. 10.1002/9780470689646.ch1. https://www.researchgate.net/publication/227988510_Automatic_Keyword_Extraction_from_Individual_Documents

[14] Rapid Automatic Keyword Extraction algorithm for Japanese. https://github.com/kanjirz50/rake-ja

[15] Blei, David & Ng, Andrew & Jordan, Michael. (2001). Latent Dirichlet Allocation. The Journal of Machine Learning Research. 3. 601-608. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

[16] 日本語 WordNet. https://bond-lab.github.io/wnja/

[17] Carl Benedikt Frey, Michael A. Osborne, The future of employment: How susceptible are jobs to computerisation?, Technological Forecasting and Social Change, Volume 114, 2017, Pages 254-280, ISSN 0040-1625, https://doi.org/10.1016/j.techfore.2016.08.019.

[18] Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. Strategic Management Journal, 42(12), 2195–2217. https://doi.org/10.1002/smj.3286

[19] Nils Reimers, Iryna Gurevych: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. EMNLP/IJCNLP (1) 2019: 3980-3990

[20] sentence-transformers/stsb-xlm-r-multilingual, https://huggingface.co/sentence-transformers/stsb-xlm-r-multilingual

付録

O*NET と日本版 O*NET (JobTag) の職業の対応づけ

日本版 O*NET における職業は O*NET とは独立して定義されており、対応づけがなされていない。そのため、[1] と同じような傾向を得ているのかを判断することが困難である。そこで本プロジェクトでは、多言語埋め込みモデルと呼ばれる、言語横断的にテキスト情報を同じ埋め込み時間に写像する技術を用いて職業の対応づけを行った。具体的には以下の実装とモデルを利用した。

- 手法: SentenceBERT [19]
- モデル: sentence-transformers/stsb-xlm-r-multilingual [20]

このモデルを利用することで英語・日本語言語問わずテキストで表現された職種（たとえば、家電修理、Home Appliance Repairers）を 768 次元の埋め込み空間に写像できる。そしてコサイン類似度が最も高い（異なる言語で記述された）職種を対応づけた。

結果を表 17 に示す。

表 17. O*NET と JobTag の対応付け

jaonet_occ	mapped_occ	similarity
家電修理	Home Appliance Repairers	0.8080365
プラント設計技術者	Manufacturing Engineers	0.71363544
医療用画像機器組立	Medical and Clinical Laboratory Technicians	0.8254237
織布工/織機オペレーター	Fabric and Apparel Patternmakers	0.683335
染色工/染色設備オペレーター	Tapers	0.6265575
ミシン縫製	Sewing Machine Operators	0.8751166
木材製造	Model Makers, Wood	0.89094096
合板製造	Tapers	0.77064645
家具製造	Furniture Finishers	0.8647002
紡織設備管理・保全	Fabric and Apparel Patternmakers	0.7743417

紙器製造	Paperhangers	0.891086
紡績機械オペレーター	Textile Cutting Machine Setters, Operators, and Tenders	0.8530166
建具製造	Architectural Drafters	0.8122555
食品技術者	Food Science Technicians	0.9480306
靴製造	Shoe Machine Operators and Tenders	0.7111251
かばん・袋物製造	Tapers	0.5125893
漆器製造	Painting, Coating, and Decorating Workers	0.76922905
貴金属装身具製作	Precious Metal Workers	0.75492823
医薬品製造	Pharmacists	0.8133412
生産・品質管理技術者	Manufacturing Production Technicians	0.85918355
タイヤ製造	Tire Builders	0.8163295
化粧品製造	Shampooers	0.6972207
石油精製オペレーター	Derrick Operators, Oil and Gas	0.845958
化学製品製造オペレーター	Chemical Technicians	0.9100596
原子力技術者	Nuclear Engineers	0.95397824
発電所運転管理	Power Plant Operators	0.88943434
分析化学技術者	Chemical Technicians	0.915329
陶磁器技術者	Potters, Manufacturing	0.786487
ファインセラミックス製造技術者	Mechatronics Engineers	0.75845253
石工	Stonemasons	0.8969589
花火師	Fire Investigators	0.7487862

高分子化学技術者	Chemists	0.83333874
バイオテクノロジー技術者	Bioinformatics Technicians	0.9107669
宇宙開発技術者	Aerospace Engineers	0.8403341
航空機開発エンジニア（ジェットエンジン）	Aerospace Engineers	0.87744653
システムエンジニア（業務用システム）	Computer Systems Engineers/Architects	0.8279343
プログラマー	Computer Programmers	0.84957373
システムエンジニア（Web サイト開発）	Software Developers, Systems Software	0.8557565
システムエンジニア（組込み、IoT）	Electronics Engineering Technicians	0.74025214
ソフトウェア開発（パッケージソフト）	Software Developers, Systems Software	0.8314333
ソフトウェア開発（スマホアプリ）	Software Developers, Applications	0.7883528
システムエンジニア（基盤システム）	Microsystems Engineers	0.81763965
運用・管理（IT）	Management Analysts	0.7619641
ヘルプデスク（IT）	Helpers--Roofers	0.6941259
セキュリティエキスパート（オペレーション）	Security Management Specialists	0.9251884
プロジェクトマネージャ（IT）	Technical Directors/Managers	0.7835518
データサイエンティスト	Computer and Information Research Scientists	0.75568646

デジタルビジネスイノベーター	Online Merchants	0.62941974
AI エンジニア	Computer Systems Engineers/Architects	0.73981494
Web デザイナー	Web Developers	0.8488827
Web ディレクター	Web Administrators	0.8587145
動画制作	Film and Video Editors	0.7863227
CG 制作	Tapers	0.54589105
ゲームクリエーター	Video Game Designers	0.75357664
アートディレクター	Art Directors	0.82586753
広告デザイナー	Advertising and Promotions Managers	0.7767864
広告ディレクター	Advertising and Promotions Managers	0.8509049
グラフィックデザイナー	Graphic Designers	0.94536495
コピーライター	Tapers	0.7747121
ディスプレイデザイナー	Actuaries	0.66404366
インテリアデザイナー	Interior Designers	0.8538337
インテリアコーディネーター	Audio and Video Equipment Technicians	0.7529808
カラーコーディネーター	Audio and Video Equipment Technicians	0.68503046
ファッションデザイナー	Fashion Designers	0.9887327
パタンナー	Tapers	0.73707557
イラストレーター	Graphic Designers	0.7234589
アニメーター	Multimedia Artists and Animators	0.7261357
看板制作	Segmental Pavers	0.7401829

テクニカルイラストレーター	Technical Writers	0.8366934
スタイリスト	Hairdressers, Hairstylists, and Cosmetologists	0.7442702
ブックデザイナー	Graphic Designers	0.6302994
テキスタイルデザイナー	Fashion Designers	0.8046219
フラワーデザイナー	Floral Designers	0.96525156
ジュエリーデザイナー	Fashion Designers	0.60512197
フードコーディネーター	Cooks, Restaurant	0.7516927
舞台美術スタッフ	Makeup Artists, Theatrical and Performance	0.6613509
舞台照明スタッフ	Fire Inspectors	0.63845265
インダストリアルデザイナー	Electromechanical Equipment Assemblers	0.7422556
商業カメラマン	Photographers	0.6064398
テレビカメラマン	Camera Operators, Television, Video, and Motion Picture	0.76432014
報道カメラマン	Reporters and Correspondents	0.8403102
テクニカルライター	Technical Writers	0.84824854
製版オペレーター、DTP オペレーター	Dredge Operators	0.66717046
印刷オペレーター	Printing Press Operators	0.8631213
製本オペレーター	Tapers	0.6973822
IT コンサルタント	Information Technology Project Managers	0.7854566
広報コンサルタント	Assessors	0.70240563

人事コンサルタント	Assessors	0.81193686
知的財産コーディネーター	Cytogenetic Technologists	0.581714
知的財産サーチャー	Real Estate Brokers	0.6323348
土木・建築工学研究者	Architectural Drafters	0.8485152
情報工学研究者	Information Technology Project Managers	0.86449844
医学研究者	Medical and Clinical Laboratory Technicians	0.8386251
科学捜査研究所鑑定技術職員	Tapers	0.5544194
薬学研究者	Pharmacists	0.79997885
バイオテクノロジー研究者	Bioinformatics Scientists	0.9055724
エコノミスト	Economists	0.9524352
特別支援学校教員、特別支援学級教員	Special Education Teachers, Middle School	0.8840527
学習塾教師	Teacher Assistants	0.9230224
日本語教師	Teacher Assistants	0.7604748
英会話教師	English Language and Literature Teachers, Postsecondary	0.87723464
職業訓練指導員	Training and Development Managers	0.8222765
社会教育主事	Education Administrators, Preschool and Childcare Center/Program	0.73916936
救急救命士	Emergency Medical Technicians and Paramedics	0.7955282
外科医	Surgeons	0.9406006

小児科医	Pediatricians, General	0.7783011
内科医	Hospitalists	0.76776814
精神科医	Psychiatrists	0.9202298
産婦人科医	Tapers	0.71383226
治験コーディネーター	Clinical Research Coordinators	0.6901745
医療ソーシャルワーカー	Healthcare Social Workers	0.9433347
福祉ソーシャルワーカー	Social and Human Service Assistants	0.87197685
施設管理者（介護施設）	Administrative Services Managers	0.81620145
カウンセラー（医療福祉分野）	Rehabilitation Counselors	0.88738203
スクールカウンセラー	Educational, Guidance, School, and Vocational Counselors	0.85619646
ベビーシッター	Childcare Workers	0.76192
フロント（ホテル・旅館）	Interior Designers	0.64403427
客室清掃・整備担当（ホテル・旅館）	Maids and Housekeeping Cleaners	0.79711646
接客担当（ホテル・旅館）	Concierges	0.74682236
ホールスタッフ（レストラン）	Concierges	0.7464917
飲食チェーン店店員	Cooks, Restaurant	0.7017307
調香師	Tapers	0.7353754
アロマセラピスト	Orthoptists	0.75577015
リフレクソロジスト	Tapers	0.6346307
葬祭ディレクター	Funeral Service Managers	0.884135

きもの着付指導員	Tapers	0.66845644
速記者、音声反訳者	Editors	0.55672425
パラリーガル（弁護士補助職）	Tapers	0.6932105
秘書	Executive Secretaries and Executive Administrative Assistants	0.7479203
受付事務	Tapers	0.7241908
一般事務	Public Relations Specialists	0.67018926
データ入力	Data Entry Keyers	0.73428
経理事務	Assessors	0.77000666
営業事務	Administrative Services Managers	0.7785929
人事事務	Assessors	0.78441596
総務事務	Executive Secretaries and Executive Administrative Assistants	0.6370772
企画・調査担当	Program Directors	0.73692226
調剤薬局事務	Pharmacists	0.87869275
介護事務	Home Health Aides	0.70816916
生産・工程管理事務	Manufacturing Engineers	0.8252953
銀行等窓口事務	Financial Examiners	0.7615157
貿易事務	Sales Agents, Securities and Commodities	0.65353835
損害保険事務	Insurance Underwriters	0.8421894
通信販売受付事務	Administrative Services Managers	0.6993358
学校事務	Education Administrators, Elementary and Secondary School	0.8508426

医療事務	Tapers	0.71225256
広報・PR 担当	Tapers	0.6661872
IR 広報担当	Actuaries	0.6202391
企業法務担当	Accountants	0.7074558
コンプライアンス推進担当	Actuaries	0.65080255
医薬品販売/登録販売者	Pharmacy Aides	0.7361367
リサイクルショップ店員	Recycling and Reclamation Workers	0.88732266
携帯電話販売	Telephone Operators	0.56241655
CD ショップ店員	Retail Salespersons	0.5700676
営業（IT）	Actuaries	0.74327415
保険営業（生命保険、損害保険）	Insurance Underwriters	0.8330109
銀行・信用金庫渉外担当	Credit Authorizers	0.7848614
ディーラー	Gaming Dealers	0.7198385
マーケティング・リサーチャー	Market Research Analysts and Marketing Specialists	0.8690115
証券アナリスト	Financial Analysts	0.80044913
商品企画開発（チェーンストア）	Supply Chain Managers	0.7727711
OA 機器営業	Tapers	0.7078592
証券外務員	Financial Analysts	0.77091014
化粧品販売/美容部員	Hairdressers, Hairstylists, and Cosmetologists	0.74052054
化粧品訪問販売	Makeup Artists, Theatrical and Performance	0.6890554
清涼飲料ルートセールス	Bartenders	0.5988177
自転車販売	Bicycle Repairers	0.696752

アクチュアリー	Outdoor Power Equipment and Other Small Engine Mechanics	0.5571971
内部監査人	Auditors	0.8972819
ファンドマネージャー	Investment Fund Managers	0.9122416
M&A マネージャー、M&A コンサルタント/M&A アドバイザー	Management Analysts	0.791944
代理店営業（保険会社）	Insurance Underwriters	0.8987427
マンション管理員	Assessors	0.73649037
マンション管理フロント	Concierges	0.6847977
雑踏・交通誘導警備員	Traffic Technicians	0.8113055
ボイラーオペレーター	Pipe Fitters and Steamfitters	0.73911715
トラック運転手	Heavy and Tractor-Trailer Truck Drivers	0.65991926
トレーラートラック運転手	Heavy and Tractor-Trailer Truck Drivers	0.87275153
ダンプカー運転手	Taxi Drivers and Chauffeurs	0.5708338
送迎バス等運転手	Bus Drivers, Transit and Intercity	0.84595764
介護タクシー運転手	Taxi Drivers and Chauffeurs	0.7527054
ルート配送ドライバー	Pile-Driver Operators	0.70396876
宅配便配達員	Postmasters and Mail Superintendents	0.69976896
新聞配達員	Broadcast News Analysts	0.7096252
倉庫作業員	Industrial Production Managers	0.7559446

ピッキング作業員	Grinding and Polishing Workers, Hand	0.77094686
ハウスクリーニング	Maids and Housekeeping Cleaners	0.802937
ペストコントロール従事者（害虫等防除・駆除従事者）	Chemical Equipment Operators and Tenders	0.7598383
製品包装作業員	Packaging and Filling Machine Operators and Tenders	0.80409217
工場労務作業員	Manufacturing Production Technicians	0.8646332
バックヤード作業員（スーパー食品部門）	Food Batchmakers	0.70651627
調理補助	Food Cooking Machine Operators and Tenders	0.7348328
給食調理員	Cooks, Restaurant	0.85195553
ごみ収集作業員	Recycling and Reclamation Workers	0.83942974
産業廃棄物処理技術者	Chemical Plant and System Operators	0.7779892
産業廃棄物収集運搬作業員	Recycling and Reclamation Workers	0.79040515
積卸作業員	Welders, Cutters, and Welder Fitters	0.7454427
こん包作業員	Helpers--Roofers	0.71122974
港湾荷役作業員	Marine Engineers	0.66704184
ブリーダー	Tapers	0.76809835
自然保護官（レンジャー）	Park Naturalists	0.71675897
酪農従事者	Agricultural Technicians	0.7224847
水産養殖従事者	Aquacultural Managers	0.77675176
稲作農業者	Farmworkers and Laborers, Crop	0.8853669

ハウス野菜栽培者	Farmworkers and Laborers, Crop	0.7342558
果樹栽培者	Tree Trimmers and Pruners	0.7945987
花き栽培者	Tree Trimmers and Pruners	0.80019194
畜産技術者	Agricultural Technicians	0.82751656
沿岸漁業従事者	Fishers and Related Fishing Workers	0.85947704
動物看護	Animal Trainers	0.82869077
ドローンパイロット	Commercial Pilots	0.69656074
国会議員	Tapers	0.6351422
国際協力専門家	Human Resources Specialists	0.54105073
会社経営者	Lodging Managers	0.80223507
玩具（おもちゃ）製作	Model Makers, Metal and Plastic	0.61602837
法務技官（心理）（矯正心理専門職）	Administrative Law Judges, Adjudicators, and Hearing Officers	0.7443931
入国審査官	Tapers	0.72940916
検察事務官	Police Detectives	0.6391331
労働基準監督官	Assessors	0.7424065
セキュリティエキスパート（脆弱性診断）	Security Management Specialists	0.9228275
特許審査官	Licensing Examiners and Inspectors	0.6087332
NPO 法人職員（企画・運営）	Human Resources Managers	0.69781065
データエンジニア	Computer and Information Systems Managers	0.7683994
独立系ファイナンシャル・アドバイザー（IFA）	Investment Fund Managers	0.6482097

タンクローリー乗務員	Podiatrists	0.6886788
検査工（工業製品）	Industrial Engineers	0.83834016
食品営業（食品メーカー）	Food Service Managers	0.8586242
自動運転開発エンジニア（自動車）	Automotive Engineers	0.8737799
医療機器開発技術者	Medical Appliance Technicians	0.8830438
臨床開発モニター	Clinical Research Coordinators	0.80282235
キッティング作業員（PC セットアップ作業員）	Computer Operators	0.77806556
フードデリバリー （料理配達員）	Food Service Managers	0.80273676
セキュリティエキスパート（デジタルフォレンジック）	Security Management Specialists	0.7499757
セキュリティエキスパート（情報セキュリティ監査）	Security Management Specialists	0.8899118
風力発電のメンテナンス	Wind Energy Engineers	0.8092741
総務課長	Executive Secretaries and Executive Administrative Assistants	0.66144514
人事課長	Assessors	0.7147963
経理課長	Tapers	0.7577602
営業課長	Administrative Services Managers	0.75446635