

Web データに基づく質問応答システムを用いた情報処理技術者試験応募法人の推定 Estimation of Corporations Applying for the Information Technology Engineer Examination Using a QA System Based on Web Data

奥村 明俊[†] 市瀬 規善[†] 久寿居 大[†] 石川 開[†] 鳥澤 健太郎[‡] 大竹 清敬[‡]
Akitoshi Okumura Noriyoshi Ichinose Dai Kusui Kai Ishikawa Kentaro Torisawa Kiyonori Ohtake

1. はじめに

我が国のデジタル社会のビジョンである「デジタルの活用により、一人ひとりのニーズに合ったサービスを選ぶことができ、多様な幸せが実現できる社会」[1]の実現に向け、IPA は、社会の仕組みをデジタルに対応させていくためのアーキテクチャ設計・データ標準化[2]を推進しながら、デジタルトランスフォーメーション (DX) の促進、サイバー・フィジカル一体化時代におけるサイバーセキュリティの確保やデジタル人材育成等に取り組んでいる。デジタル人材の育成と活用には、スキル・レベルの可視化に向けた環境整備が必要である。IPA では、情報技術者試験のオンライン化や通年試験化により受験者の利便性向上を図って応募者の拡大を図っている[3]。応募者拡大には、広報活動が必要であり、SNS や Web サイトを活用した情報発信、パンフレットの配布、説明会の開催などを行っている。従来、情報系の学校や IT ベンダー企業をターゲットとして広報活動を行ってきたが、近年、応募者の属性が多様化している。特に、IT に関する基本的リテラシーの資格試験である IT パスポート試験は、文科系の学生やユーザ企業の社員など従来と異なる所属の応募者が増加している。広報活動の効果的なターゲットングのため、応募者の属性や応募者の多い法人を推定する必要がある。特に、パンフレット配布や説明会開催は、リソースとコストに限りがあるので、精度の高いターゲットングが求められる。パンフレット配布や説明会開催は、法人を対象に実施しているので、潜在的応募者が多い法人を見込み顧客として正確に推定することがターゲットングの課題である。

本論文では、Web データに基づく質問応答システムを用いて情報処理技術者試験に応募する法人を推定する手法を提案し、提案手法を実装した WISDOM-DX について述べる。2 節では、潜在顧客を推定する手法と情報処理技術者試験のターゲットングの課題について述べる。3 節では、WISDOM-DX による応募法人の推定手法について説明する。4 節では、東京証券取引所上場企業 3640 社について、WISDOM-DX およびベースラインとして Google 検索を適用して推定した結果を報告する。5 節では、WISDOM-DX とベースラインの推定結果について考察し今後の課題について述べる。

2. 潜在顧客の推定

一般に、法人や個人などの潜在顧客を推定するため、以下の手法が用いられている：

1) 顧客への直接調査

対象者全般に広くアンケート調査やインタビューなどを実施し、潜在的顧客の意識や行動を直接的に把握する。例えば、どのような商品やサービスに関心があるか、実際にどのような商品を購入したことがあるか、サービスを利用したことがあるかなどを調査する。大規模かつ精緻に実施す

る場合、時間とコストを要するのでタイムリーに実施することは容易ではないが、精度の高い推定が可能となる。

2) 顧客の実績データの活用

既存顧客など過去実績のある顧客のデータを分析し、共通的な属性を抽出することで潜在顧客を推定する。顧客データから同一ユーザを発見する User Identity Linkage (UIL) として研究され、広告配信や顧客分析などに用いられている。顧客の実績データを教師データとして、学習技術を用いて顧客の行動履歴からユーザ同定を行うことが可能である[4]。

3) 顧客の発信情報の活用

法人や個人が Web などソーシャルメディア上に発信した情報を分析することで、どのような取組みや活動を行っているかを把握し、関連する商品やサービスの潜在顧客となるかを推定する。例えば、DX や女性活躍など企業経営に関する活動を推進している企業を Web 情報から推定することが可能である[5][6]。このような活動を推進している企業は、DX や女性活躍推進に関するセミナーやイベントの潜在顧客となる。

実際には、これら 3 つの手法を組み合わせて用いることが多い。情報処理技術者試験の場合、パンフレット配布や説明会開催といった広報活動のターゲットングのために、法人に対して大規模に直接リサーチを行うことはコストの観点から難しい。現状、応募者の実績データとして所属情報が十分に蓄積されていないので、顧客の実績データを活用することは現実的ではない。そこで、まず顧客の発信情報を活用して、法人のターゲットングを行う。具体的には、Web 情報を活用して情報処理技術者試験に関心の高い法人を推定するシステムを開発し、その結果をもとにターゲットングして重点的な広報活動を行う。

3. WISDOM-DX による応募法人の推定

3.1 WISDOM-DX の目的

従来、広報活動の担当者は、重点的に注力すべき法人をターゲットングするため、各法人について Web 検索を行い情報処理技術者試験に対する関心度合いを推定していた。担当者は、検索結果を見ながら検索を繰り返し時間と手間をかけてターゲットングしていた。そのため、ターゲットングは属人的な作業となりがちで、定観測的な評価や組織的活動として展開することが困難であった。我々は、企業の経営活動を評価するために開発した WISDOM-DX[5][6]を活用して、情報処理技術者試験に応募する可能性の高い法人を自動的に推定するシステムを実現する。法人の経営者は、広報、IR、プロモーションなどを通じて、経営方針や戦略、具体的な活動を Web 上に発信している。また、優れた活動はメディアによって取り上げられることも多い。DX のような活動テーマ名と法人名の AND 条件による Google 検索結果数は、企業の経営活動を評価する一つ

[†] 独立行政法人 情報処理推進機構 (IPA)

Information-technology Promotion Agency, Japan

[‡] 国立研究開発法人 情報通信研究機構 (NICT)

National Institute of Information and Communication

の目安になる。検索結果の多い企業ほど積極的に取り組んでいる傾向があり、Web 情報が評価のための重要な情報源となる[5]。もちろん、単純な検索には検索誤りが存在し、検索精度の改善が必要である。WISDOM-DX は、評価テーマに関して 5W1H (when, who, where, what, why, how) の 6 種類の質問文を生成し、Web データに基づく質問応答システムを用いて企業活動を評価する。6 種類の質問応答から回答集合を生成してスコアリングする。その際、予め用意した優良企業事例を学習データとして 6 種類のスコアを積み上げして統合的なスコアを算出して企業の取組みを評価する。企業の DX 活動に関して、東京証券取引所上場企業 3371 社を WISDOM-DX によってランキングして、人手による 3 種類の評価において高評価の企業に着目しどのように分布するかを検証した結果、66.7% から 79.2% が、WISDOM-DX 上位 10% と一致した[5]。また、女性活躍推進企業に関して、東京証券取引所上場企業 3380 社を WISDOM-DX によってランキングし、人手による 3 種類の評価において高評価の企業に着目しどのように分布するかを検証した結果、68.1% から 82.3% が、WISDOM-DX 上位 10% と一致した[6]。WISDOM-DX による企業活動の評価は、上位 10% において複数のテーマで 67% 以上人手による評価と一致している。

今回、IT パスポート試験に関心の高い法人を推定するために WISDOM-DX 上に必要なテーブルや辞書を実装した。IT パスポート試験は、IT やセキュリティに関する基礎知識を問う情報処理技術者試験である。WISDOM-DX によって東京証券取引所上場企業 3640 社の IT パスポート試験への関心の高さを推定し、関心が高いと思われる上位 5% (182 社) が広報活動の注力対象となり得るかを検証する。

3.2 WISDOM-DX の処理概要

図 1 の WISDOM-DX のシステム構成に従って、IT パスポート試験に関心の高い法人を推定する処理を説明する。

5W1H 質問生成は、入力された組織リストの各組織に対し、ドメイン辞書のキーワードと質問展開テーブルを用いて、組織の情報教育活動の 5W1H に関する 6 つの質問タイプの質問文 (クエリ) を生成する。組織リストの例を図 2 に示す。図 2 の例は、質問展開テーブルの質問タイプ 1 から質問タイプ 6 までのテンプレートの <sub> のスロットに「組織 A」、「組織 B」を入力することを示している。質問展開テーブルの例を図 3 に示す。図 3 の例では、質問タイプ 1 のテンプレートとして「<sub>はどのように<obj>を<pred>か」が記述され、以降、質問タイプ 2 から質問タイプ 6 までの質問テンプレートが記述されている。ドメイン辞書の例を図 4 に示す。図 4 の例では、質問タイプ 1 から質問タイプ 6 のテンプレートの <obj> のスロットに「IT やセキュリティの基礎知識」を、<pred> のスロットには「学んでいる」を入力することを示している。5W1H 質問生成は、組織リストとドメイン辞書の記述から質問展開テーブルのスロットを埋め、質問タイプ 1 として「組織 A はどのように IT やセキュリティの基礎知識を学んでいるか」というクエリを生成する。同様の手順により、質問タイプ 2 として「組織 A はどこで IT やセキュリティの基礎知識を学んでいるか」、質問タイプ 3 として「組織 A は誰が IT やセキュリティの基礎知識を学んでいるか」、質問タイプ 4 として「組織 A は IT やセキュリティの基礎知識で何を学んでいる

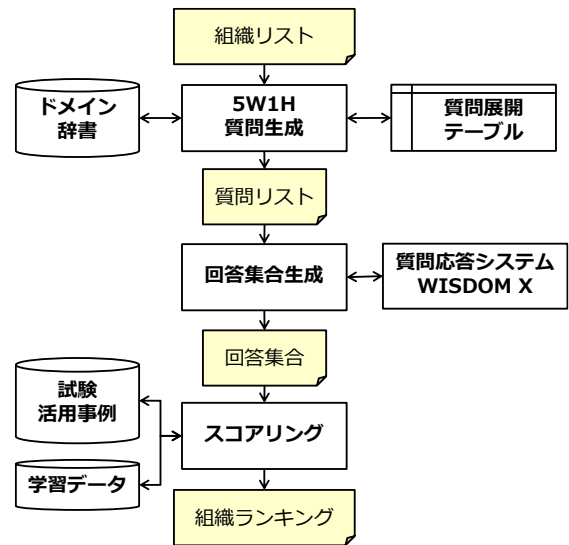


図 1 WISDOM-DX システム構成

質問タイプ	企業リスト
1,2,3,4,5,6	<sub> ← 組織A
1,2,3,4,5,6	<sub> ← 組織B

図 2 組織リストの例

質問タイプ	質問展開テーブル
1	<sub>はどのように<obj>を<pred>か
2	<sub>はどこで<obj>を<pred>か
3	<sub>は誰が<obj>を<pred>か
4	<sub>は<obj>で何を<pred>か
5	<sub>はなぜ<obj>を<pred>か
6	<sub>はいつ<obj>を<pred>か

図 3 質問展開テーブルの例

質問タイプ	ドメイン辞書
1,2,3,4,5,6	<obj> ← IT やセキュリティの基礎知識
1,2,3,4,5,6	<pred> ← 学んでいる

図 4 ドメイン辞書の例

質問タイプ	回答パッケージ	尤度	回答の URL
1	新入社員研修でITの知識・技術の基礎を身につけ、その後日常のビジネス活動と能力開発のための研修を通じて、ITのプロとしてのスキルを研鑽する。	0.769	https://www.inews.jp/2021/06/...html
1	入社後1ヶ月の技術研修では、未経験者をシステムエンジニアに育て上げてきました。未経験者を対象とした研修カリキュラムが組まれているため、IT知識が無い人でも研修期間中に基礎を丁寧に教えます。	0.646	https://it.impress.co.jp/articles/xxx

図 5 質問タイプ 1 の回答集合例

か」、質問タイプ5として「組織AはなぜITやセキュリティの基礎知識を学んでいるか」、質問タイプ6として「組織AはいつITやセキュリティの基礎知識を学んでいるか」をクエリとする質問リストを生成する。

回答集合生成は、質問リストのクエリを質問応答システム WISDOM X[7][8]に入力し、Webデータ 91.8億ページからクエリに対する回答パッセージ、回答パッセージの確からしさを示す尤度、回答の元となったWeb情報のURLの3つ組を回答トリプレットとして回答集合を生成する。回答集合は、質問タイプごとに生成されるので1社につき6セットとなる。質問タイプ1のクエリに対する回答集合の例を図5に示す。

通常、スコアリングは、回答集合の「回答量」、「類似度」、「尤度」に基づいてスコアを計算する。回答量は、回答パッセージ数であり図5の例では2となる。尤度は、回答パッセージの確からしさを示す値であり、図5の例ではそれぞれ0.769と0.646である。「類似度」は、回答パッセージと試験の活用事例とのテキスト間の類似度である。試験活用事例は、過去の実績として公開された取り組みの紹介記事からテキストデータを抽出したものである。類似度は、テキスト間の共通の語彙、語彙の重要性、尤度を組み合わせた8つのスコア関数で算出される。8つのスコア関数を6つの質問タイプに対する回答集合に適用し、マルチクエリスコアアンサンブル法 (MQSE) [5]によって統合的なスコアを算出する。学習データは、質問タイプごとに算出されたスコアを統合するパラメータの学習 (重み付け) に用いられるもので、ITパスポート試験に応募した実績のある組織と実績のなかった組織の二値データである。今回、ITパスポート試験の活用事例が存在しないので、類似度はすべて均等に1とする。また、学習データとして利用できるほど実績データが存在しないので重みはすべて均等に1を割り付けることとする。その結果、6つの質問タイプの回答集合の尤度スコアを均等に加算したものが各組織の統合スコアとして算出されることになる。スコアリングは、最終的に入力リストに記載された組織の統合スコアの大きい順にランキングして出力する。

4. 評価実験

4.1 目的と方法

実験の目的は、東京証券取引所上場企業3640社を対象としてITパスポート試験の応募者増加につながる新たな法人を見込み顧客として見つけることである。現実的に広報活動可能な200社以内に絞り込んで、重点的に広報活動を行う予定である。そのため、WISDOM-DXを用いて法人のITやセキュリティの基礎教育に関する取り組みを評価し、評価スコアの高い法人をITパスポート試験に対する関心が高い見込み顧客とする。ITパスポート試験は、ITやセキュリティに関する基礎知識を問う試験なので、3節で述べた「どのようにITやセキュリティの基礎知識を学んでいるか」など6種類の質問文を用いる。WISDOM-DXの有効性を比較検証するために、法人名、ITもしくはセキュリティ、基礎知識のキーワードによるGoogle検索をベースライン手法として実行する。具体的には、「法人名 AND (IT OR セキュリティ) AND 基礎知識」を検索条件としてGoogle Custom Searchを実行し、検索件数の大きい順に3640社をランキングする。ベースラインは、従来の広報担当者が行

っていた作業を簡単にモデル化したものである。評価の目安として、ITパスポート試験のバウチャーを購入した実績のある45社を用いる。バウチャーは、ITパスポート試験の受験手数料支払いに使用できる電子的な前売りチケットであり、法人が団体として申し込むものである。ITパスポート試験のバウチャーによる応募者は24,590名であり、全応募者の約9.7%である。応募者の所属する法人の約10%がバウチャーを購入しており、バウチャー購入は法人がITの基礎教育に関心が高いことを示す目安になる。そこで、バウチャー購入企業に着目し、WISDOM-DXとベースラインによる3640社のランキング中にどのように分布するかを比較検証する。今回の目的は、バウチャー購入企業を識別することではなく、バウチャー購入企業の分布から見込み顧客を見つけることである。ランキング上位の企業は、バウチャー購入実績はなくてもITパスポート試験に関心が高いと思われるので見込み顧客とする。WISDOM-DXとベースラインがランキングした結果において、バウチャー購入企業がどのように分布するかを検証し、それぞれのランキングの上位5%の182社、7.5%の273社、10%の364社の中にバウチャー購入企業が含まれる割合を評価し、上位5%の企業に関して内容を分析する。

4.2 評価結果

バウチャー購入企業が、WISDOM-DXとベースラインによるランキング上位5% (182社)、7.5% (273社)、10% (364社)の中にそれぞれ含まれる割合を図6に示す。WISDOM-DXの上位5%、7.5%、10%に含まれる割合は、それぞれ、20% (9/45)、26.7% (12/45)、35.6% (16/45)であった。また、ベースラインの上位5%、7.5%、10%に含まれる割合は、それぞれ、13.3% (6/45)、15.6% (7/45)、20%

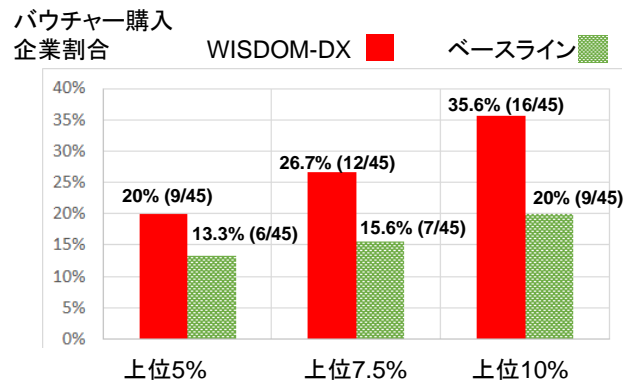


図6 バウチャー購入企業の割合



図7 WISDOM-DXにおける分布

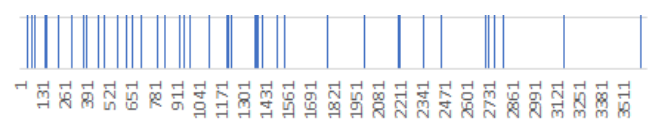


図8 ベースラインにおける分布

(9/45) であった。バウチャー購入企業が、WISDOM-DX とベースラインによる 3640 社のランキングの中でどのように分布するかを図 7 と図 8 にそれぞれ示す。横軸はランキング順位、縦線はその順位に存在する企業を示している。図 7 において、WISDOM-DX によって検索されずスコアが算出されなかったバウチャー購入企業 27 社は、最下位の 3640 位にプロットされた。図 8 において、ベースラインによるスコアが最下位であるバウチャー購入企業は 3601 位にプロットされた。バウチャー購入企業は、WISDOM-DX では上位と下位に分かれ、ベースラインでは全般に分布する結果となった。

5. 考察

東京証券取引所上場企業 3640 社に WISDOM-DX とベースラインを適用した結果、バウチャー購入企業が含まれる割合は、それぞれの上位 5%, 7.5%, 10% の企業に対して図 6 に示すように 6.7% から 15.6% の優位性があった。また、バウチャー購入企業がランキング上位に含まれる分布から WISDOM-DX はベースラインよりも見込み顧客を上位に絞り込んでいる。しかし、Web 検索では、検索誤りによって対象とした企業と異なる企業が検索されることがある。そこで、WISDOM-DX とベースラインのそれぞれの上位 5% の 182 社の検索結果の内容について分析した。WISDOM-DX とベースラインの共通の上位 5% 企業 (グループ A) は 44 社、WISDOM-DX のみ上位 5% の企業 (グループ B) は 138 社、ベースラインのみ上位 5% の企業 (グループ C) は 138 社であり、検索誤りの割合 (誤り率) は表 1 に示す結果であった。

	グループ A	グループ B	グループ C
企業数	44 社	138 社	138 社
誤り率	2.3% (1/44)	6.5% (9/138)	13.8% (19/138)

表 1 WISDOM-DX とベースラインの上位 182 社の誤り率

各グループの検索誤りの主な要因は以下の通りである：

1) グループ A

検索誤りは 1 社、誤り率は 2.3% であった。この企業は転職幹旋会社であり、Web ページには IT やセキュリティの基礎教育を行っている他の企業について掲載されていた。また、この転職幹旋会社が提供するコンテンツを用いた IT 基礎教育についても記載されていたため、この企業自身が主体的に IT の基礎知識の学習に取り組んでいると解釈され回答メッセージに高い尤度を与えられた。

2) グループ B

検索誤りは 9 社、誤り率は 6.5% であった。企業名が短い場合、企業名を部分文字列として含む異なる企業の Web ページが検索され回答メッセージが生成されることがあった。例えば、株式会社 A が対象企業の場合、株式会社 AB が、Y 株式会社対象企業の場合、XY 株式会社対象として検索され回答メッセージが生成された。

3) グループ C

検索誤りは 19 社、誤り率は 13.8% であった。グループ B と同様に、短い企業名を部分文字列として含む異なる企業が検索結果に含まれていた。また、他社の事例を紹介しているメディア関連会社、セミナー運営会社、グループ A と同種の転職幹旋会社が含まれていた。

WISDOM-DX による上位 5% 企業 (グループ A とグループ B の和集合) とベースラインによる上位 5% 企業 (グループ A とグループ C の和集合) の検索誤り率は、それぞれ、5.5% (10/182) と 11% (20/182) であり、WISDOM-DX の見込み顧客の検索誤り率はベースラインの半分であった。今後、WISDOM-DX の検索誤りに対処して精度向上を図る。また、WISDOM-DX とベースラインの上位 5% 企業に IT パスポート試験に関するアンケート調査を実施して評価と情報収集を行う。さらに、試験活用事例や学習データを充実して、WISDOM-DX の結果をもとに広報活動を実施するスキームを確立し、IT パスポート試験以外の情報処理技術者試験の広報活動にも活用する。

6. おわりに

IT パスポート試験の広報活動として重点的に注力すべき法人をターゲットにするため、Web データに基づく質問応答システムを用いた WISDOM-DX を開発した。東京証券取引所上場企業 3640 社に WISDOM-DX を適用した結果、バウチャー購入企業が含まれる割合は、Google 検索によるベースラインと比べて、上位 5%, 7.5%, 10% の企業に対して 6.7% から 15.6% の優位性があった。また、WISDOM-DX による上位 5% 企業とベースラインによる上位 5% 企業の検索誤り率は、それぞれ 5.5% と 11% であり、WISDOM-DX の見込み顧客の検索誤り率はベースラインの半分であった。今後、アンケート調査を実施して試験活用事例や学習データを充実し、WISDOM-DX の結果をもとに広報活動を実施するスキームを確立し、IT パスポート試験以外の情報処理技術者試験の広報活動にも活用する。

謝辞

株式会社日本システムアプリケーションの赤石沢元博氏、岡部雅世氏、NICT の木俣豊氏、内元清貴氏、IPA の齊藤裕氏、下田忠義氏、本多康弘氏ほか関係者のご支援に感謝の意を表します。

参考文献

- [1] デジタル庁, “デジタル社会の実現に向けた重点計画”, 2022. [<https://www.digital.go.jp/policies/priority-policy-program/>]
- [2] 奥村明俊, “Society 5.0 の実現に向けた挑戦者へのエール”, 情報処理 Vol.62, No.6, pp. 284-287 (2021).
- [3] 本多康弘奥村明俊, “情報処理技術者試験における実施方式の改革 - 「新たな日常」を踏まえた試験の実現に向けて - ” 情報処理 Vol.64, No.4, pp. 166-173. (2023)
- [4] 草野元紀, 小山田昌史: 行動パターンを基にした異なるドメインに対するユーザ同定技術, 第 13 回データ工学と情報マネジメントに関するフォーラム, 日本データベース学会, (2021)
- [5] 奥村明俊, 市瀬規善, 久寿居 大, 石川 開, 鳥澤 健太郎, 大竹 清敬, “Web データに基づく企業のデジタルトランスフォーメーション活動評価システム WISDOM-DX の検証”, 第 15 回データ工学と情報マネジメントに関するフォーラム, 日本データベース学会, (2023)
- [6] 奥村 明俊, 市瀬規善, 久寿居 大, 石川 開, 鳥澤 健太郎, 大竹 清敬, “Web データに基づく企業経営活動評価システム WISDOM-DX の検証” 2023 年度人工知能学会全国大会 (第 37 回) (2023)
- [7] 国立研究開発法人情報通信研究機構 (NICT), “WISDOM X (ウィズダム エックス) とは?”, <https://www.wisdom-nict.jp/#top>
- [8] Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Ryu Iida, Masahiro Tanaka, and Julien Kloetzer “A semi-supervised learning approach to why-question answering”, In Proceedings of AAAI-16, pp. 3022-3029, (2016)