

WEB データに基づく質問応答システムを用いた 企業のデジタルトランスフォーメーション活動の分析と評価

久寿居 大^{†1} 石川 開^{†1} 奥村 明俊^{†1}
鳥澤 健太郎^{†2} 大竹 清敬^{†2}

概要:

デジタルトランスフォーメーション (DX) の促進は、我が国にとって喫緊の課題である。DX 促進と関連事業拡大のためには、企業における DX の取組の調査・分析・評価と状況把握が不可欠である。一般的に、このような活動は、アンケートやヒアリングなどの調査に基づく専門家の分析・評価によって行われる。その結果、手間や費用などコストがかかり、タイムリーかつ迅速な現状把握が困難である。また、記述式回答に関する評価は、専門家によって属人性があり一貫性にかけるという課題がある。本論文では、専門家による分析プロセスを評価軸ごとに質問応答の形式でモデル化し、アンケート調査を実施せずとも WEB データに基づく質問応答システムを用いて企業の DX 活動を自動的に分析・評価可能なシステム、WISDOM-DX を提案する。WISDOM-DX は、企業の DX 活動に関して、質問応答結果をもとにスコアリングし DX の進展度合いを評価する。システムが WEB データに基づき企業の DX 活動を自動的に分析・評価するので、属人性を排除した結果を迅速に得ることができる。本システムを DX 銘柄 2021 にアンケート回答した企業 464 社に適用し、正解企業 48 社を識別する実験を行ったところ、AUPR が 0.350 であり、システムがランキングした上位 48 件の正解率は 33.3% であった。不正解の原因について分析した結果、企業名の分割による単純な検索エラーや、業種や自己資本利益率などの制限によるものが多く、対応可能であることがわかった。企業名の分割と業種制限に対応することにより 11 社について改善でき 25 社 (52.1%) を正解とできる見通しを得た。約 5 割の精度で DX 先進企業を自動的に推定可能であり、従来できなかった客観的な観点からの大規模な分析と評価の自動化が可能になる。

キーワード: デジタルトランスフォーメーション, DX, テキストマイニング, 経営戦略分析, 質問応答, 小論文自動採点

Analyzing Enterprise Activities for Digital Transformation using QA System based on WEB information

DAI KUSUI^{†1} KAI ISHIKAWA^{†1} AKITOSHI OKUMURA^{†1}
KENTARO TORISAWA^{†2} KIYONORI OHTAKE^{†2}

Abstract:

Promoting digital transformation (DX) is one of the most urgent issues for Japanese society. In order to enhance the DX in organizations, it is essential to determine their status by surveying, analyzing, and evaluating their current DX initiatives. This determination is generally performed by experts using questionnaires, interviews, and other surveys. This requires much time and expense, making it difficult to obtain the current status in a timely and prompt manner. In addition, there is a problem of inconsistency because the interpretation of descriptive answers may differ among the experts. This paper proposes WISDOM-DX, a system that automatically analyzes and evaluates an organization's DX initiatives instead of questionnaires using a question-and-answer system filled from Web information. By modeling the analysis process of experts in the form of questions and answers for each evaluation axis, WISDOM-DX evaluates the progress status of organizations' DX initiatives by scoring them based on the results of the question-and-answer system. The system automatically analyzes and evaluates a company's DX activities based on Web data, which makes it possible to obtain consistent results in a timely and efficient manner. The system was evaluated with 464 companies that replied to the DX Brand 2021 questionnaires by comparing the results that DX experts evaluated using the questionnaires. The AUPR was 0.350. True positive rate was 33.3% about 48 correct companies. Our error analysis proved that 11 false positives were correctable because they were caused by segmentation errors of company names and lack of knowledge on award criteria. The true positive rate is 52.1% when all the 11 false positives are corrected. It is possible to automatically evaluate the DX initiatives with an accuracy of higher than 50% in a larger scale from the objective view-points that was not conventionally achieved.

Keywords: Digital Transformation, DX, Text Mining, Business Strategy Analysis, Question and Answering, Automated Essay Scoring

1. はじめに

我が国は、Society5.0の実現に向けてデータ利活用とデジタル・ガバメントの観点から社会全体のデジタル化に取り組んでいる[1,2]。このような取組みに対して、行政機関は人的リソースや財源を最大限有効活用して、効果的な施策

のエビデンスに基づいた立案と迅速な評価が求められる[3]。今般、COVID-19により、社会経済活動全般のデジタルトランスフォーメーション(DX)が、我が国にとって喫緊の課題であることが鮮明になった。DXは、2004年にストルターマン教授が「ITの浸透が、人々の生活をあらゆる面でより良い方向に変化させる」概念として提唱した。我が

^{†1} 独立行政法人 情報処理推進機構 (IPA)
Information-technology Promotion Agency, Japan

^{†2} 国立研究開発法人 情報通信研究機構 (NICT)
National Institute of Information and Communications Technology

国では、2018年にDX推進のガイドラインが発表され、DXは「企業がビジネス環境の激しい変化に対応し、データとデジタル技術を活用して、顧客や社会のニーズを基に、製品やサービス、ビジネスモデルを変革するとともに、業務そのものや、組織、プロセス、企業文化・風土を変革し、競争上の優位性を確立すること」と定義されている[4]。企業は、COVID-19の影響により事業継続の危機にさらされ、テレワークなど社内のITインフラや就業に関するルールを迅速かつ柔軟に変更し環境変化に対応できた企業と、対応できなかった企業の差が拡大している。DXの成否が日本の将来を左右すると言っても過言ではない。日本のDX進展のためには、企業など組織のDXの状況をタイムリーかつ適切に把握し、有効事例を広く共有する必要がある。そのため、各企業のDXに対する取り組みを調査・分析・評価する活動が行われている。評価活動は、アンケートやヒアリングなどの調査に基づいて専門家が分析して行うことが多い[5]。大規模調査では、アンケート回答者と分析の専門家の手間や費用などコスト低減が課題となる。まず、回答者にとって回答そのものが手間であり、詳細な調査のため項目や記述内容が増えると負担感が増大し回答率が低下する。また、回答者が自由に記述した内容の分析（質的分析）は、専門家にとっても負担となる。そのため大規模調査をタイムリーに実施するのは容易ではない。質的分析は、数値化されたデータに対する量的分析とは異なり、分析手法が標準化ないし規格化されていると言え難く[7]、解釈が恣意的であり解釈に至る過程が不明確になることもある[8]。専門家一人で把握できるデータ量に限界があり、人手による質的分析には見落としや主観による偏りの問題が指摘されている[8]。このような属人性を排除した評価の一貫性担保も重要な課題である。本論文では、これらの課題を解決して大規模な調査をタイムリーに実施するため、専門家による分析のプロセスを質問応答の形式でモデル化し、アンケート調査を行うことなくWEBデータに基づく質問応答システムを用いて企業のDX活動を評価するWISDOM-DXシステムを提案する。WEBデータに基づいて企業を評価することにより属人性を排除した迅速な評価と、膨大な企業の中から上位に入るような、他社の見本となる先進的なDXへの取り組みを行っている企業を自動推定可能にすることが目的である。2章では、関連調査研究として、DXに関する調査、アンケート分析、解答の自動採点と質問応答について概説する。3章では、DX調査事例としてDX銘柄2021の調査や選定プロセスについて述べる。4章では、WISDOM-DXの内容とシステム構成について説明する。5章では、提案システムをDX銘柄2021に適用し実験した結果について報告する。6章では、実験結果について考察し今後の課題について述べる。

2. 関連調査と研究

2.1 DXに関する調査

近年、DXに関して進展状況の分析や助成事業の採択審査のため、民間調査会社[9,10,11]、業界団体[12,13,14]、地方自治体[15,16]、行政機関[17,18,19]など多数の組織によって調査が実施されている。調査の対象は、民間企業、自治体や官公庁など公的機関、各種団体など様々であるが、アンケートやヒアリングの結果や提案書などをもとに有識者など専門家が分析・評価することが一般的である。アンケートの回答は、選択式と自由記述式が用いられる。選択式はあらかじめ作成された選択肢を回答者が選び、自由記述式は回答者が自分の言葉で表現する。選択式は大量のデータの確保や回答者の分類など量的分析が容易であり、自由記述式は回答者の主張や意図などの質的分析に適する。経済産業省が東京証券取引所と選定したDX銘柄の場合、国内企業を対象に選択式と自由記述式のアンケート調査を行い、評価委員が優れた企業を選定している[17,18]。DX銘柄に関しては3章で詳細に述べる。

2.2 自然言語処理によるアンケート分析

企業の経営戦略活動分析として、アンケートの自由記述やインタビューから得られたテキストデータを計量的に分析するテキストマイニングが広く活用されている[5]。テキストマイニングとは、定性的な特徴をもつテキストを定量的に分析する技術である[20]。テキストデータを統計的に解析し、その中から重要な語句や単語間の関係を明らかにする。テキストマイニングという言葉は1990年代半ばから使われるようになり[21]、さらにテキストデータに対し計量的手法で解析を行うためのテキストマイニングソフトウェアパッケージの開発も進んだ。テキストマイニングを行えば、大量の言語データを量的に整理・分析し、言葉や言葉同士の関係を掘り出し可視化できる。それによりデータの全体像を把握し、探究の方向を得ることや、解釈の根拠を数学的に示すことも可能になる[7]。2000年代から、多くの分野で活用されるようになった[20]、経営戦略分析として、後藤らは、企業の戦略事例にアナロジーという概念を導入し、中小企業の経営者が、自社の状況と最も類似する事例を抽出する方法を示した[22]。企業戦略立案のための手法であるSWOT分析の視点を取り入れたアナロジー評価指標を提案し、自社のビジネス環境に適合する戦略事例を効率的に発見するモデルを提案した。日経ビジネスの記事から202社の事例を対象とし、大分類として戦略、強み、弱み、機会、脅威の5分類から小項目であるアナロジー評価指標185項目を抽出した[23]。このアナロジー評価指標を用いて、類似性の高い戦略事例を効率的に検索することを実証した。DXの分析には、どのような視点から評価指標を抽出するかを明らかにしなければならないが、経営分析のアプローチとして有益な示唆を与えている。

2.3 自然言語処理による解答採点と質問応答

記述式問題の解答を自動採点する研究が、AES (Automated Essay Scoring) として行われている[24]。AESは、計算機を使ってエッセイを採点するタスクの総称である。TOEFL テストの採点にも用いられる e-rater [25]や日本語における小論文評価システム Jess [26] があげられる。内容語の Bag of Words を素性とする SVM を用いる手法[27]や、ニューラルネットワーク(NN)を用いて文章分類問題を解く研究が行われている。NN のモデルとしては、文の木を基に再帰的に文のベクトルを構成する再帰型ニューラルネットワーク(RNN) [28] や、局所的な情報を畳み込むことでフレーズ単位での判断を下せる畳み込みニューラルネットワーク(CNN) [29] などがある。これらのモデルは映画レビューの極性判定や、質問文の分類などのタスクにおいて用いられ、ベースライン手法である SVM 相当以上の精度を達成した。寺田らは、記述式問題の回答が全体として正解か不正解を自動採点する手法を提案した[30]。高校学習範囲の試験問題の解答データに正解と不正解のラベルを付け、CNNにより、単語よりも大きな単位での有用な素性の抽出をして解答の分類を行った。数百人規模の解答データを用いて実験したところ、90%の精度を達成した。DX 活動のような正解が自明ではない記述の評価手法としての効果は不明であるが、採点の指針を見出せば、参考となる技術である。

質問に対する回答の生成は、質問応答システムとして研究されている[31]。2011年に質問応答システム(ワトソン)が米国のクイズ番組でチャンピオンに圧勝したことが大きな話題となった。ワトソンは、問題の内容を分析して、事前に収集された大量のテキスト情報から問題の解答候補とその根拠・確信度を計算し、高い確信度の候補を解答とした[32]。近年、自治体や企業の問い合わせ対応などで、AIチャットボットの実用化が進んでいる[33]。ワトソンもチャットボットも回答には、情報源となる百科事典やシナリオ、FAQなどデータを必要とする。質問応答能力は、準備できるデータの質や量に依存しデータをいかに獲得・更新するかが課題となる。情報源として WEB データやウイキペディアを活用する手法も提案されている[34,35,36]。大規模データの活用は、オープンな質問応答に活用され国際的なコンテストで技術開発が進められている[37]。すべての質問に完全に回答できるものではなく、実用として活用するためにはタスク設計が重要な課題となる。

3. DX 銘柄 2021 の選定

DX 銘柄とは、東京証券取引所の上場企業の中から、企業価値の向上につながる DX を推進するための仕組みを社内に構築し、優れたデジタル活用の実績が表れていると、経済産業省と東京証券取引所とに共同で選定された企業である。DX 銘柄 2021 の選定にあたって、DX 調査事務局が、2020年11月から、東京証券取引所の国内上場会社約 3,700

社に対して、「経営ビジョン・ビジネスモデル」「戦略」「戦略実現のための組織・制度等」「戦略実現のためのデジタル技術の活用・情報システム」「成果と重要な成果指標の共有」「ガバナンス」の6項目に関するアンケート調査を実施した。アンケート調査は選択式と自由記述式の2種類である。銘柄選定は、一次評価と二次評価の2つのプロセスで行われた。一次評価として、DX 銘柄 2021 調査に回答した企業 464 社からのアンケートの選択式回答項目と各企業の3年平均 ROE (自己資本利益率) によるスコアリングが行われる。次に、評価者9名から構成される評価委員会においてアンケートの記述式回答などから取り組みについて評価され、DX 銘柄 2021 (グランプリを含む 28 社) と DX 注目企業 2021 (20 社) などが選定され7か月後に発表された[17]。民間調査会社や業界団体においても調査開始から発表まで数か月かかることが多く[14]、調査に要するコストは小さくない。DX 銘柄 2021 と同様の調査を 2015 年から 2020 年まで実施した時のアンケート回答率は 6%~15%と高くない [18]。回答する企業の負担も大きく、回答および調査・分析のコストを低減しつつ、一貫性のある評価の大規模かつタイムリーな実現が今後の課題である。DX 銘柄に選定された企業は、単に優れた情報システムの導入、データの利活用にとどまらず、デジタル技術を前提としたビジネスモデルと経営の変革にチャレンジし続けていると評価された企業である。選定企業の優良な取組が他の企業における DX の参考となることが期待されている。

4. 質問応答システムを用いた DX 分析・評価

4.1 WISDOM-DX

本論文では、調査・分析・評価のコスト低減と属人性を排除した一貫性のある評価の課題を解決するために、アンケート調査を行うことなく WEB データに基づく質問応答システムを用いて企業の DX 活動を自動的に分析・評価するシステム、WISDOM-DX を提案する。

専門家による評価は、ドメインごとの評価軸にそった質問応答と考えられる。例えば、DX 銘柄 2021 においては、3章で述べた6項目について記述式回答を評価する。そこで、専門家による分析プロセスを評価軸ごとに質問応答の形式でモデル化し、WEB データに基づく質問応答システムを用いて企業の DX 活動を自動的に分析・評価する。WISDOM-DX のシステム構成を図1に示す。WISDOM-DX は、まず入力された企業リストの各企業に対してドメイン評価モデルのキーワードと質問展開テーブルから複数の質問文を生成する。次に、質問応答システム WISDOM X を活用して、WEB データを基に質問文に対する応答結果を出力する。応答結果から DX 摘要を生成し、評価参照データベースと照合して類似度を計算しスコアリングする。企業から自由記述回答が得られる場合は、DX 摘要に自由記述回答を合わせてスコアリングする。最後に、企業ごとのス

コアをランキングして出力する。このように、企業の DX に関して、WEB データに基づいた自動分析・評価により、属人性を排除した結果を迅速に得られる。

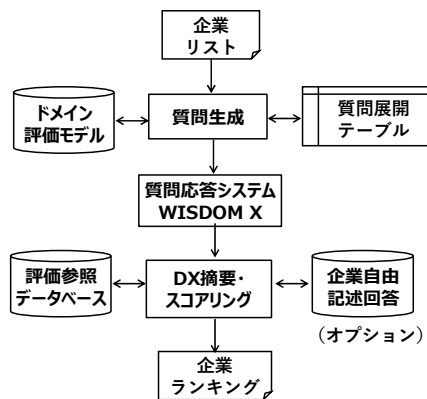


図 1 WISDOM-DX システム構成

Figure 1 WISDOM-DX System Configuration

4.2 質問応答システム

質問応答システムとして WISDOM X を用いる[36]. WISDOM X は、WEB60 億ページから抽出した情報を基にして、ファクトイド型 (例えば、「地球温暖化は何をもたらす?」)、なぜ型 (例えば、「なぜ地球温暖化が起きる?」)[38]、どうやって型 (例えば、「どうやって地球温暖化を防ぐ?」) どうなる型 (例えば、「地球温暖化が進むとどうなる?」)[39]、定義型 (例えば、「地球温暖化とは何?」) といったタイプの質問に回答できる。WISDOM X は 2015 年より公開されていたが、2021 年 3 月に 350GB のテキストで事前学習した BERT や、BERT と敵対的学習と呼ばれる深層学習技術[40,41,42]を組み合わせた独自改良版が導入され、精度が向上し、回答できる質問のバリエーションも増大した。WISDOM X の構築には、ミドルウェアとして RaSC [43]を開発して種々の言語処理モジュールを多数の計算機上で並列に効率良く運用している。従来の検索エンジンは、入力されたキーワードに関連する文書の URL を提供するだけであるが、WISDOM X は、質問に対して名詞句や動詞句、文など必要な範囲でピンポイントな回答を出力する。なお、WISDOM X は、一般にその出力結果を WISDOM X 以外のシステム開発やデータベース構築に利用することを許諾していない。WISDOM-DX は、NICT の了解のもと WISDOM X を活用している。

4.3 質問生成

最終的なスコアリングの精度を高めるためには、各企業の DX に関する情報だけを精度よく収集することが重要である。そのため WISDOM X の様々なタイプの質問に柔軟に回答できるという特徴を活用する[36]。今回は低コストで大規模に DX に優れた企業を調査するという目的に合わせて、調査・分析のコスト低減を優先して、質問のタイプは一般によく使われる 5W1H をベースとし、ドメイン評価

モデルを用いた質問文の生成は行わなかった。精度を重視し、企業から自由記述回答が得られる場合には、自由記述回答の分析により、ドメインに特徴的な語を抽出し、それらをドメイン評価モデルに加えた質問文を生成することでより幅広く情報収集が可能になる。

5W1H をベースとした質問文は以下の 6 タイプである。

- 1) どうやって:「X 社はどうやって DX を達成したか?」
- 2) どこで:「X 社はどこで DX を推進しているか?」
- 3) 誰が:「X 社は誰が DX を推進しているか?」
- 4) 何を:「X 社は DX で何を達成できたか?」
- 5) なぜ:「X 社はなぜ DX を達成できたか?」
- 6) いつ:「X 社はいつから DX を推進しているか?」

DX 銘柄 2021 の選定では、3 章で述べたように 6 項目に関してアンケート調査が実施された。アンケートの 6 項目の情報は、以下の 4 タイプの質問によって主に収集される。

- 1) どうやって:「経営ビジョン・ビジネスモデル」「戦略」「戦略実現のためのデジタル技術の活用・情報システム」の 3 項目
- 2) どこで:「戦略実現のための組織・制度等」の 1 項目
- 3) 誰が:「戦略実現のための組織・制度等」「ガバナンス」の 2 項目
- 4) 何を:「成果と重要な成果指標の共有」の 1 項目

さらに、残りのタイプ (「なぜ」、「いつ」) の質問文を付け加えることによりアンケート項目に含まれないような各社の DX に関する情報も収集されることが期待される。

4.4 DX 摘要とスコアリング

1) 評価参照データベース

評価参照データベースは、企業の DX に関する情報、例えば、企業の名称、業種、銘柄コードといった属性情報に加えて、開示された外部機関からの受賞歴や評価情報を保持する。また、企業の基本情報として有価証券報告書の内容なども記載される。これらの情報を用いて、業種や従業員数など DX 以外の属性によってスコアリング結果を分類し、ランキングをコントロール可能となる。本論文の実験では、評価参照データベースとして、2020 年 DX 銘柄[18]や 2016 年から 2019 年までの攻めの IT 銘柄や注目企業[18]として公開された取り組みを DX 優良企業の事例として参照する。これらの取り組みは、インターネット検索サイトのグーグルでも検索でき、WISDOM X においても検索されるサイトとして登録する。

2) DX 摘要・スコアリング

DX 摘要は、各企業の DX に関する質問に対して WISDOM X の出力結果をまとめたものである。評価参照データベースと DX 摘要を用いて、以下に述べる複数のスコア関数で各企業のスコアを計算し、提案するマルチクエリアンサンブル手法によって各企業の総合スコアを計算する。総合スコアによる企業ランキングが最終的に出力される。

3) スコア関数

企業の DX に対する重要な取り組みや成果は、報道、IR、プロモーションなどを通じて WEB 上に露出するので、企業の DX への取り組みの水準は WISDOM X から得られる DX 摘要の質や量から推定可能と考えられる。そこで、我々は、DX 摘要に含まれる記事数を量的特徴量、過去の DX 銘柄や注目企業の活動紹介記事との類似度を質的特徴量として DX 摘要をスコアリングするアプローチを提案する。以降、DX 摘要の具体的なスコア関数を検討する。DX 摘要の量的特徴量を用いたスコアの定式化としては、まず単純に、次式のように記事数をスコアとする方法が考えられる。

$$Score_{cnt}(\mathbf{D}_t, q) = \sum_{\{d_t | \mathbf{D}_t\}} 1$$

ここで、 \mathbf{D}_t は DX 摘要、 q は WISDOM X に入力された質問文、 d_t は DX 摘要 \mathbf{D}_t に含まれる各記事を表す。また、WISDOM X は、各記事 d_t の質問文 q に対する回答らしさ、すなわち信頼度を推定するため、この信頼度を考慮した次式のような定式化も可能である。

$$Score_{cnt_conf}(\mathbf{D}_t, q) = \sum_{\{d_t | \mathbf{D}_t\}} conf(d_t, q)$$

ここで、 $conf(d_t, q)$ は WISDOM X から得られる、各記事 d_t のクエリ q に対する信頼度を表す。次に、質的特徴量を用いたスコアの定式化として、評価参照データベース中の過去の DX 銘柄や注目企業の活動紹介と DX 摘要の間の語彙的類似度に基づく次式のような定式化が可能である。

$$Score_{sim}(\mathbf{D}_t, \{w_h\}, q) = \sum_{\{w_t | \mathbf{D}_t\}} sim(w_t, \{w_h\})$$

ただし、 $sim(w_t, \{w_h\})$ は、DX 摘要 \mathbf{D}_t に含まれる単語 w_t と評価参照データベース中の文書に含まれる単語集合 $\{w_h\}$ の間の類似度をあらわす。本研究では品詞フィルタリングを適用し、固有名詞を除く名詞、動詞、形容詞のみを単語に用いる。語彙的類似度には、文脈非依存の単語埋め込みベクトルを用いた次式を用いる。

$$sim(w_t, \{w_h\}) = \max_{\{w_h\}} \frac{\mathbf{v}(w_t) \cdot \mathbf{v}(w_h)}{\|\mathbf{v}(w_t)\| \|\mathbf{v}(w_h)\|}$$

ここで、 $\mathbf{v}(w_t)$ および $\mathbf{v}(w_h)$ はそれぞれ、単語 w_t および w_h の単語埋め込みベクトルを表す。この類似度に基づくスコア $Score_{sim}$ も、前述の信頼度との組み合わせが可能である。具体的には次式のように定式化できる。

$$Score_{sim_conf}(\mathbf{D}_t, \{w_h\}, q) = \sum_{\{d_t | \mathbf{D}_t\}} \sum_{\{w_h\}} sim(w_t, \{w_h\}) \cdot conf(d_t, q)$$

さらに、品詞フィルタリングに加え、単語の重要度を考慮する。情報検索で一般的な $idf(w)$ を用いれば、次式のように定式化できる。

$$Score_{sim_idf}(\mathbf{D}_t, \{w_h\}, q) = \sum_{\{w_t | \mathbf{D}_t\}} sim(w_t, \{w_h\}) \cdot idf(w_t)$$

単語頻度 $tf(w)$ も併用すれば、次式の定式化となる。

$$Score_{sim_tf_idf}(\mathbf{D}_t, \{w_h\}, q) = \sum_{\{w_t | \mathbf{D}_t\}} sim(w_t, \{w_h\}) \cdot tf(w_t) \cdot idf(w_t)$$

単語の重要度を考慮した二つのスコア $Score_{sim_idf}$ および $Score_{sim_tf_idf}$ は、さらに信頼度との組み合わせが可能である。 $Score_{sim_idf}$ への信頼度の考慮は次のように定式化できる。

$$Score_{sim_idf_conf}(\mathbf{D}_t, \{w_h\}, q) = \sum_{\{d_t | \mathbf{D}_t\}} \sum_{\{w_t | d_t\}} sim(w_t, \{w_h\}) \cdot idf(w_t) \cdot conf(d_t, q)$$

同様に、 $Score_{sim_tf_idf}$ の場合は次式のようになる。

$$Score_{sim_tf_idf_conf}(\mathbf{D}_t, \{w_h\}, q) = \sum_{\{d_t | \mathbf{D}_t\}} \sum_{\{w_h\}} sim(w_h, \{w_t\}) \cdot tf(w_h) \cdot idf(w_h) \cdot conf(d_t, q)$$

以上、計 8 種類のスコア関数を DX 摘要のスコアリングに用いる。

4) マルチクエリアンサンブル

提案手法は、4.2 節で述べたように、6 種類の質問文を生成する。その結果、6 種類の DX 摘要に対してそれぞれスコアが計算される。本論文では、これらのスコアを以下に述べるアンサンブル法で 1 つに統合する方法を提案する。

Step 1: 6 種類の質問文 $\{q\}$ から得られた各 DX 摘要 \mathbf{D}_t に対して、前節のスコア関数を適用し、スコアを求める。

Step 2: Step 1 で求められた DX 摘要のスコアを、質問文の種類が同じグループに分けてソートすることにより、質問文の種類ごとの各企業の企業集合における順位 $rank(\mathbf{D}_t, q)$ を求める。

Step 3: 質問文の種類ごとに求めた企業の順位から総合スコア $Score_{ensemble}$ を次式で求める。

$$Score_{ensemble}(\mathbf{D}_t, \hat{c}_q) = \sum_{\{q\}} \hat{c}_q \cdot rank(\mathbf{D}_t, q)^{-1}$$

ここで、 $\{q\}$ は 6 種の質問文集合、 $rank(\mathbf{D}_t, q)$ はある質問文 q に対する企業の順位、 \hat{c}_q は結合的数を表す。結合定数は、ランキングの評価尺度を目的関数として、これを最大化する値を用いる。例えば AUPR を評価尺度に用いる場合、以下の式で最適化を行う。

$$\hat{c}_q = \operatorname{argmax}_{c_q} AUPR(Score_{ensemble}(\mathbf{D}_t, c_q), y_{true})$$

ここで、 y_{true} は、二値分類の正解データを表す。本実験では過去年度の DX 銘柄や注目企業を正例、それ以外の銘柄を負例とする教師データを用いる。

5) 企業自由記述回答

入力された企業のアンケートの自由記述回答が利用できる場合、スコアリングに活用する。今回は、企業名を入力としてスコアリングするので活用しない。

5. 実験

5.1 方法

WISDOM-DXは膨大な企業の中から上位に入るDX先進企業を自動推定可能にすることが目的である。DX 銘柄2021のエントリ企業464社の中からDX 銘柄2021(28社)とDX 注目企業2021(20社)の合計48社を上位約10%の正解企業として提案システムによって識別可能かを評価する。ただし、DX 銘柄2021は、業種ごとに原則2社ずつ選定、ROEが高い企業を加点方式により評価、などの条件が加わっている。WISDOM-DXではDX 銘柄2021を100%抽出することが目的ではなく、あくまでWISDOM-DXの性能評価のために利用する。今回の実験では、WISDOM-DXに企業の業種情報やROEの情報は与えておらず、WEBから収集可能な自然文のみでの評価である。ベースラインの識別方法として、インターネットで利用可能な検索サイトとしてGoogleを用いる。企業名とデジタルトランスフォーメーションというキーワードで検索して検索件数の大きい順にランキングする。提案システムとして、4.4に記載したマルチクエリアンサンブルの手法でランキングする。

5.2 結果

性能を直観的に把握するために、ベースラインと提案システムによる上位48件に正解企業48社が含まれる件数を調べた結果、ベースラインが11件(22.9%)で、提案システムが16件(33.3%)であった。評価尺度はAUPR(Area Under the Precision-Recall Curve)を用いる。AUPRは、Precision-Recall曲線下の面積を計算したものである。一般的に、分類システムの適合率と再現率はトレードオフの関係にあるのでPrecision-Recall曲線は右下がりの曲線になるが、AUPRスコアが高ければ高いほどシステムの予測精度が良いことを示す。Googleの検索件数でランキングしたベースラインではAUPRは0.202、提案手法のマルチクエリアンサンブルでは、スコア関数 $sim_tf_idf_conf$ に対するAUPRが最も高く、0.350であった。提案手法の

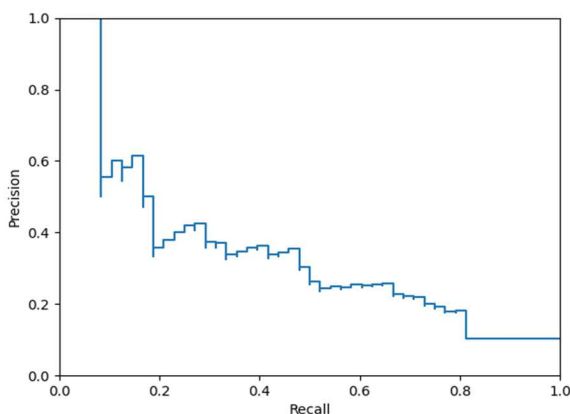


図2 提案手法のPrecision-Recall曲線

Figure 2 Precision-Recall Curve of the proposed method

Precision-Recall 曲線を図2に示す。

6. 考察

6.1 上位48社中の不正解企業

1) 企業名が分割されて検索

提案システムによる上位48社には正解ではない企業(False Positive)が32社含まれている。そのうち6社は、企業名が一単語としてではなく分割されて検索された結果、不適当なDX摘要が生成されたことによる。例えば、企業名が「あいうえお」のように平仮名連続と「銀行」のような普通名詞の組合せや「システム」「データ」といった普通名詞のみで構成される場合、一単語の企業名として検索されない結果がDX摘要として生成される。一単語の企業名が含まれないDX摘要は、排除することで対応する。

2) 業種による制限

DX 銘柄2021では、業種ごとに原則2社に限定されて選定される[17]。提案システムによる上位48社中には情報・通信業の業種において正解である2社が含まれているが、正解ではない企業5社も含まれている。これらの5社は業種による制限によって選定されなかったと思われる。

3) 一次評価による選定

DX 銘柄2021の選定においては、各企業の選択式項目の回答および3年平均ROE(自己資本利益率)によるスコアによって一次評価が行われる[17]。提案システムによるFalse Positive 32社から上述した1,2)の企業13社を除く19社は、一次評価によって選定されなかった可能性がある。一次評価結果について関連機関に確認したところ、12社が一次評価によって選定されないことが分かった。

False Positive 32社の不正解理由を表1に示す。上述した1,2,3)以外の理由によるその他として9社が存在する。これら9社については、WEB情報からだけでは識別困難なのか、他に原因によるものか詳細な分析が必要である。

表1 False positive 32社の不正解理由

Table 1 Error reasons of false positive 32 companies

不正解の理由	件数
1) 企業名が分割されて検索	6
2) 業種制限	5
3) 一次評価による選定	12
4) その他	9

6.2 上位48社にランクされなかった正解企業

提案システムによる上位48社よりも下位にランクされた正解企業(False Negative)は32社であった。そのうち、9社については、WISDOM Xの応答数がゼロであり、8社については応答数が10件未満であった。これらの企業については、DX摘要がほとんど生成されず、上位にランキングされなかった。これらの企業は、Google検索では存在するものであり、WISDOM Xのクロール対象として追加す

ることにより改善が図れる。他の 15 社については、WISDOM X の応答数は少量ではあるが存在する。つまり、DX 摘要は少量ながらも生成されたが、スコアリングによって上位にランキングされなかった企業である。応答数を増やすためには、一般的に、企業名を同義語や短縮語に置き換えて質問生成を追加することが有効である。今回、汎用性を重視してドメイン評価モデルを活用しなかったが、ドメイン評価モデルを用いた同義語や短縮語、質問文中の単語の置き換えによって質問生成の拡大により DX 摘要を充実させることも有効と思われる。DX のドメイン評価モデルは、例えば、ビジョン、ビジネスモデル、戦略、経営計画、組織、制度、デジタル技術、情報システム、生産性、経営トップといったキーワードである。企業の自由記述回答を活用することにより、ドメイン評価モデルをさらに充実することが可能となる。

6.3 スコアリング手法の改善

6.1 節で述べた、企業名が分割されて検索、業種制限については、スコアリング手法の改善によって対応可能である。まず、企業名が分割されて検索された WISDOM X の応答については、検索結果中に一つの連続した単語として存在しない場合、DX 摘要から削除するか重みを減らすことでランキングを下げられる。また、事業制限については、企業の業種コードをパラメータとして、業種ごとにランキングして業種制限に対応することが考えられる。一次評価による選定については、ROE は公開されていて参照可能であり ROE を反映したスコアリングは可能である。

6.4 企業 DX 活動の自動分析・評価

提案システムによるランキング上位 48 社の中には、16 社 (33.3%) の正解が含まれ、6.3 節で述べた改善を行うことで、上位 48 社中の正解ではない企業 (False Positive) 32 社中の 24 社について改善可能性がある。上位 48 社よりも下位にランクされた正解企業 (False Negative) 32 社中の 17 社についても改善できる可能性がある。これらの改善策のうち WISDOM X の応答数が少なかった企業を WISDOM X のクローリング対象として追加し企業名の分割と業種制限に対応することにより 11 社について改善でき 25 社 (52.1%) を正解とできる見込みがある。また、本システムにより上位 48 社にランキングされなかった企業であっても高くランキングされた企業は、一定以上の DX 推進活動を行っていると考えられる。東京証券取引所の国内上場会社約 3700 社を対象にした場合でも、アンケート未提出企業も含めて、従来の評価に準ずる形で約 5 割の精度で WEB データに基づいて DX 先進企業 (トップ 10% の企業) を推定可能であり、従来はできなかった客観的な観点からの大規模な分析と評価の自動化が可能になる。

6.5 今後の課題

評価参照データベースとの類似度からスコアリングした場合、参照データベースに登録されていないような独自

性の高いユニークな取り組みは、高く評価されない可能性がある。そのような事例は、DX 銘柄 2021 においては、ほとんど見当たらないが、専門家による評価が必要となる事例に備える必要がある。専門家によって何らかの形で高く評価された事例は、タイムリーに評価参照データベースに登録して対応できるようなスキームを構築する予定である。

7. おわりに

専門家による分析プロセスを評価軸ごとに質問応答の形式でモデル化し、WEB データに基づく質問応答システムを用いて企業の DX 活動を自動的に分析・評価するシステム WISDOM-DX を開発した。本システムを DX 銘柄 2021 にアンケート回答した企業 464 社に適用し、正解企業 48 社を識別する実験を行ったところ、AUPR が 0.350 であり、システムがランキングした上位 48 件の正解率は 33.3% であった。不正解の原因について分析した結果、企業名の分割による単純な検索エラーや、業種や ROE などの選定における制限によるものが多く、対応可能であることがわかった。企業名の分割と業種制限に対応することにより 11 社について改善でき 25 社 (52.1%) を正解とできる見込みである。東京証券取引所の国内上場会社約 3700 社を対象にした場合でも、アンケート未提出企業も含めて、従来の評価に準ずる形で約 5 割の精度で DX 先進企業 (トップ 10% の企業) を自動的に推定可能であり、従来はできなかった客観的な観点からの大規模な分析と評価の自動化が可能になる。本システムにより DX についてアンケート調査を実施することなく WEB データに基づいて大規模な調査・分析が実現可能であることを示した。

謝辞 経済産業省の平井裕秀氏と田辺雄史氏、NICT の内元清貴氏と田仲正弘氏、IPA の市瀬規善氏、山田宏幸氏、野村治彦氏、中沢潔氏、境真良氏、柳原聡子氏、光安沙輝子氏、井上佳春氏、松村文子氏など関係者の皆様のご支援に厚く御礼申し上げます。最後に、南山大学青山幹雄教授の日本の DX に対する計り知れないご貢献に感謝し敬意を表するとともに深い哀悼の意を表します。

参考文献

- [1] 奥村明俊:デジタルアーキテクチャデザイン特集 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. (May 15, 2021)
- [2] 首相官邸:デジタル社会の実現に向けた改革の基本方針, (Dec.25,2020)
<https://www.kantei.go.jp/jp/singi/it2/dgov/201225/siryou1.pdf>
- [3] 小林庸平:日本におけるエビデンスに基づく政策形成(EBPM)の現状と課題”, 日本評価研究 Vol. 20, No.2, pp.33-48, (July 2020)
- [4] 経済産業省:DX 推進ガイドライン. (Dec 12, 2018)
<https://www.meti.go.jp/press/2018/12/20181212004/20181212004.html>
- [5] 町田佳世子:質的研究におけるテキストマイニング活用の利

- 点と留意点, SCU Journal of Design & Nursing Vol. 13, No. 1, pp.47-53 (2019)
- [6] 佐藤郁哉: 質的データ分析の基本原則と QDA ソフトウェアの可能性. 日本労働研究雑誌 665 : pp.81-96, (Dec. 2015)
- [7] 岡部大祐: 計量的テキスト分析の基礎. 田崎克也編, コミュニケーション研究のデータ解析. ナカニシヤ出版, 京都, pp.189-201, (2015)
- [8] 今井多樹子, 高瀬美由紀, 佐藤健一: 質的データにおけるテキストマイニングを併用した混合分析法の有用性. 日本看護研究学会雑誌 41(4), pp.685-700, (2018)
- [9] 日経 BP 総合研究所: DX サーベイ 2 With コロナ時代の実態と課題分析, (Nov. 25, 2020)
<https://info.nikkeibp.co.jp/nxt/campaign/b/279660/>
- [10] 矢野経済研究所: 2020 デジタルトランスフォーメーション(DX)市場の現状と展望. (Jul. 30, 2020)
https://www.yano.co.jp/press-release/show/press_id/2487
- [11] IDC Japan 株式会社 国内企業のデジタルトランスフォーメーション動向調査結果を発表 (De.7 2020)
<https://www.idc.com/getdoc.jsp?containerId=prJPJ47071820>
- [12] 一般社団法人日本情報システム・ユーザー協会: 企業 IT 動向調査 2021. (Apr.28, 2021)
https://juas.or.jp/library/research_rpt/it_trend/
- [13] 一般社団法人情報サービス産業協会: 社会のデジタルトランスフォーメーション(DX)推進に貢献する情報サービス企業のあり方. (May 30, 2019)
<https://www.jisa.or.jp/publication/tabid/272/pdid/30-J007/Default.aspx>
- [14] 一般社団法人日本 CTO 協会:DX 動向調査レポート 2021 年度版. (Apr.12,2021) <https://cto-a.org/news/2021/04/12/4956/>
- [15] 東京都: DX 推進実証実験プロジェクト(第 1 期) (Mar.19, 2021)
<https://www.metro.tokyo.lg.jp/tosei/hodohappyo/press/2021/03/19/05.html>
- [16] 神奈川県: 神奈川県 DX プロジェクト推進事業, (May 17, 2021) <https://www.pref.kanagawa.jp/docs/sr4/dx-project.html>
- [17] 経済産業省:デジタルトランスフォーメーション銘柄 (DX 銘柄) 2021. (Jun.7, 2021)
https://www.meti.go.jp/policy/it_policy/investment/keiei_meigara/dx-report2021.pdf
- [18] 経済産業省: DX 銘柄/攻めの IT 経営銘柄. (June 11, 2021)
https://www.meti.go.jp/policy/it_policy/investment/keiei_meigara/keiei_meigara.html
- [19] 総務省: 令和 2 年度版情報通信白書 日本企業のデジタル・トランスフォーメーション推進に向けて
https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/html/n_d133440.html
- [20] 小木しのぶ: テキストマイニングの技術と動向, 計算機統計学 28(1), pp.31-40, (2018)
- [21] 金明哲: コーパスとテキストマイニング. 石田基広, 金明哲編著, コーパスとテキストマイニング. 共立出版, 東京, pp.1-14, (2012)
- [22] 後藤正幸,原田繁幸,田邊亘:戦略事例の構造化による戦略アナロジー評価モデルの構築. 経営情報学会 2007 年秋季全国研究発表大会. (2007)
- [23] 週刊日経ビジネス 2000 年 12 月 11 日~2003 年 9 月 1 日号
- [24] Zixuan Ke, Vincent Ng:Automated Essay Scoring: A Survey of the State of the Art, Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence Survey track. Pages 6300-6308. <https://doi.org/10.24963/ijcai.2019/879>
- [25] Yigal Attali and Jill Burstein. Automated essay scoring with erater v.2. Journal of Technology, Learning, and Assessment, (2006).
- [26] 石岡恒憲, 亀田雅之. コンピュータによる小論文の自動採点システム Jess の試作. 計算機統計学, Vol. 16, No. 1, pp. 3-19, (Dec 2003)
- [27] 中島功滋. 機械学習を利用した短答式記述答案の自動識別. 日本教育工学会第 26 回全国大会, pp.639-640. 日本教育工学会第 26 回全国大会講演論文集, (2010).
- [28] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng and Christopher Potts: Recursive deep models for semantic compositionality over a sentiment tree-bank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1631-1642, (2013).
- [29] Kim Yoon. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746-1751, (2014).
- [30] 寺田凜太郎, 久保顕大, 柴田知秀, 黒橋禎夫, 大久保智哉: ニューラルネットワークを用いた記述式問題の自動採点, 言語処理学会第 24 回年次大会発表論文集, pp.370-373. (2016)
- [31] 奥村学監修: 質問応答システム自然言語処理シリーズ 2, コロナ社. (Aug. 2009)
- [32] 金山博, 武田浩一: Watson:クイズ番組に挑戦する質問応答システム, 情報処理, Vol.52, No.7, pp. 840-849, (Jun. 15, 2011).
- [33] IT トレンド:2021 年版チャットボット比較 27 選! AI 搭載・シナリオ型別に紹介, (Jun.2021)
<https://it-trend.jp/chatbot/article/639-0003#chapter-3>
- [34] 三原英理, 藤井敦, 石川徹也. World Wide Web を用いたヘルプデスク指向の質問応答システム, 第 4 回情報科学技術フォーラム講演論文集, pp. 163-166, (Aug. 2005)
- [35] 相濱佑介, 土屋誠司, 渡部広一: Wikipedia を情報源とした質問応答システムの検討, 情報科学技術フォーラム講演論文集 18th, pp.165-166. (Aug.20 2019)
- [36] 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>
- [37] Sewon Min, et al: NeurIPS 2020 EfficientQA Competition: Systems, Analyses and Lessons Learned, (Jan. 2021)
<https://colinraffel.com/publications/arxiv2021neurips.pdf>
- [38] 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).
- [39] Chikara Hashimoto, Kentaro Torisawa, Julien Kloetzer, Motoki Sano, István Varga, Jong-Hoon Oh, and Yutaka Kidawara: Toward future scenario generation: Extracting event causality exploiting semantic relation, context, and association features. In Proceedings of ACL 2014, pp. 987-997, (2014).
- [40] John-Hoon Oh, Kazuma Kadowaki, Julien Kloetzer, Ryu Iida and Kentaro Torisawa: Open domain why-question answering with adversarial learning to encode answer texts. In Proceedings of ACL 2019, pp. 4227- 4237, (2019).
- [41] Kazuma Kadowaki, Ryu Iida, Kentaro Torisawa, Jong-Hoon Oh, and Julien Kloetzer: Event causality recognition exploiting multiple annotators' judgments and background knowledge. In Proceedings of EMNLP 2019, pp. 5820-5826, (2008).
- [42] John-Hoon Oh, Ryu Iida, Julien Kloetzer, and Kentaro Torisawa: BERTAC: Enhancing transformer-based language models with adversarially pretrained convolutional neural networks. In Proceedings of ACL-IJCNLP 2021, pp. 2103- 2115, (2021).
- [43] Masahiro Tanaka, Kenjiro Taura, and Kentaro Torisawa.: Automatic resource management for program orchestration in large-scale data analysis. In Proceedings of IPDPS 2017, pp.1088-1097. (Jun.30 2017)