

令和 5 年 5 月 23 日現在

機関番号：10101

研究種目：基盤研究(C)（一般）

研究期間：2017～2022

課題番号：17K06250

研究課題名（和文）抽象的概念構造表現を介した人-システム循環によるデータモデリング

研究課題名（英文）Data Modeling with Human-System Circulation via Abstract Conceptual Structure Representation

研究代表者

湧田 雄基（Wakuda, Yuki）

北海道大学・大学院教育推進機構・特任准教授

研究者番号：00377847

交付決定額（研究期間全体）：（直接経費） 3,600,000円

研究成果の概要（和文）：本研究では、データ分析により得た結果を抽象化した概念構造に対応付け、意味関係を極力自動的に解釈可能とすることで、文章表現による分析結果の提示の実現を目指した。研究に用いたデータは、数値や文字列が混在するテキストデータを対象とした。それらのデータについて、変数間の意味的な関係性を識別し、これを外部知見データとして分析に導入し、分析結果の向上を試行する点が本研究の特徴である。インフラ維持管理データの語彙関係を示す概念構造（オントロジー）のプロトタイプを制作し、これを用いた分析結果の説明文の生成を試行した。さらに、具体的な分野として、社会インフラの維持管理において、データ分析並びに結果解釈を試行した。

研究成果の学術的意義や社会的意義

本研究の成果は、ビッグデータからモデリングを行った結果を抽象化した概念構造に紐づけ、文章表現を行う事である。特に、インフラメンテナンス分野を対象として、データ分析結果を実務者への説明を日本語の文章表現によって行うことで、データサイエンスを広く様々な人に活用してもらえよう努めた点で、社会的意義がある。このとき、データ分析により構築した「モデル」による出力結果と、この結果に対する人による解釈とを対比することを実現し、これにより、AIや機械学習の実務を通じた運用の実現を目指したものである。

研究成果の概要（英文）：In this study, we aimed to map the results obtained from data analysis to an abstracted conceptual structure, to interpret semantic relations as automatically as possible, and to present the analysis results in a written form. The data used in the study were text data containing a mixture of numerical values and character strings. The unique feature of this study is that it attempts to improve the analysis results by identifying semantic relationships among variables in tabular data and introducing them into the analysis as external knowledge data. In this study, a prototype of a conceptual structure (ontology) showing lexical relationships among infrastructure maintenance data was produced, and an attempt was made to generate explanatory text for the analysis results using this ontology. In addition, data analysis was conducted in the specific field of social infrastructure maintenance management.

研究分野：データサイエンス

キーワード：データサイエンス 機械学習 オントロジー AI DX

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

申請者は、システム(ロボット)による自動的な状況(Context)認識手法について研究を行ってきた。その手法として、映像によって観測したシーンを様々なセンサ等の情報から、「何が(主体)」、「何を(相互作用対象)」、「どうしている(行為)」といった抽象概念構造でシーンを解釈・表現する研究を行った。この概念構造は、Cognitive Ontology (コグニティブオントロジ)と呼んでおり、これまでの研究で、映像中の静的シーンによる Context の解釈/表現/記述の提案を行った。

本研究提案では、以上の考え方をインフラ維持管理におけるデータ分析の解釈性向上という課題に適用し、より実用に近い研究課題も含めた研究テーマを設計し研究を行うものである。近年、インフラ維持管理の対象構造物の老朽化や維持管理人材減少、コスト制約への対応が喫緊の課題とされている。これらの課題に対する解決策として ICT/IoT/AI など情報技術の貢献への期待は大きい。申請者の従前の研究を通じ、データ分析の実用化のための解くべき技術課題を以下3つに絞った。また、これらの課題解決に対し、システム(分析側)と人との間を繋ぐ新たな技術体系の考案とプロトタイプシステムを用いた実務志向の評価を含む研究が必須であると考えた。

【インフラ分野でのデータ分析実用化に向けた技術課題】

- a. データクレンジング(大規模な量的/質的データ中のデータエラーを発見し修正する)
- b. 人の知見(暗黙知)を分析に導入する
- c. 分析結果に対する根拠/解釈を説明する(解釈性)

以上のうち、本研究で取り組む b および c は、人工知能を真に実用化するために解決が必要な課題である。これらの課題解決によって、「機械が得意な処理」と「人でしか為せない意思決定などの役割」との分担を明確にしながら連携し、業務の循環によりデータサイエンスの性能を向上する。

2. 研究の目的

本研究では、統計や機械学習を活用し大規模なデータから潜在的な知見を抽出するデータ分析(データモデリング)において、特に実用/応用に近づけるためには、分析内容や結果と応用展開との間のギャップを試行錯誤で詰めていくことが必要である。この段階に於いて、応用に関する知見(ドメイン知識)を分析に導入する手法や、分析結果を人に対して根拠を含めて説明する手法については、大いに研究の余地がある。

本研究実施の目的は、データ分析の取り組みにおける、機械と人との連携による「分析性能の向上」と「分析結果の解釈性向上」である。具体的な分野として「土木/社会インフラの維持管理(以降、インフラ維持管理)」を対象とし、インフラ維持管理に関するデータ(設計/施工/補修/点検/運用/環境)を分析し、維持管理業務の高度化を目指す。

以上の目的実現のため、以下2つの技術目標を設定する。

人の暗黙知を分析に導入する手法の類型化と評価(人の知見を分析に導入)

データ分析結果を解釈するためのオントロジ設計と実装(分析結果の根拠/説明の手法)

上の技術目標の達成により、人とシステムとの間のコミュニケーションを通じたデータサイエンスサイクル(循環)によるデータモデリング手法構築とその効果の評価を実施する。

3. 研究の方法

研究提案書に記載の実施計画に基づき、研究計画を次の3つの研究課題に分割し、推進した。

- 「研究課題1 抽象的概念を用いたデータモデリング」
- 「研究課題2 抽象的概念を用いた分析結果の表現技術」
- 「研究課題3 研究成果の評価」。

上の研究課題推進により、人の知見/暗黙知を分析に導入する手法の類型化とその効果の評価を行う。人が有する知見(暗黙的と位置付けられる工学的/現場経験的な知見)を分析に導入するための方策として、データ調製や分析手法等、段階毎に類型化するとともにその効果を評価する。

また、データモデルの解釈と文章化のための記述方法の類型化と評価を行う。維持管理データを対象に、統計および機械学習を用いたデータモデリングを適用する。その上で、対象データのデータモデル表現に必要な要素(Entity/Relation)を列挙し、類型化する。これにより、データモデル解釈と文章化のためのオントロジを設計する。

4. 研究成果

研究期間前半において、分析対象データの準備および、統計量算出の前処理を実施した。変数に含まれる値の出現頻度を分析することで、頻出値および値の共起の解析を行った。その際、対象データの値の分布を用いて、変数認識を行う為のロジック・辞書・処理スクリプトの構築を行った。これらにより、分析対象データテーブルを読み込めば、変数が有する値や値の分布を解析する事で、対象変数の意味/概念を推定し、変数と意味との対応付けが可能となった。

この際の変数認識では、値の形式(数量,文字列,日付,緯度経度)などの判別について、値の形式により行い、値の頻出値の辞書参照や、値の占有率調査により、変数の認識を試行した。認識した変数に対して、目的変数/説明変数の区別を行い、線形判別分析および決定木分析を用いたデータモデリングを行った。

また、研究期間後半において、対象とするインフラ維持管理データの語彙関係を示す概念構造(オントロジ)のプロトタイプを制作し、これを用いた分析結果の説明文の生成を試行した。

さらに、データ分析事例を増やし、ロジスティック回帰、ランダムフォレスト等のバギングや、勾配ブースティング等のブースティング手法を試行し、これらの分析結果の解釈に必要なオントロジの拡充を行った。特に、複数の異なる手法による分析結果の性能評価を実施するとともに、パラメータ選定における特徴の差異についての研究を行った。

機械学習におけるモデルの考察については、Feature Importance 及び SHAP 値を参照した手法について、複数手法における評価を行った。さらに、具体的な分野として、社会インフラの維持管理において、除雪業務等のデータを題材としたデータ分析を行い、分析の実施に加え、計画時に目指した維持管理業務の高度化に向けた課題について、機械可読なデータの集約や実務者のスキルアップについて課題として考察した。以上の結果のインフラ維持管理業務での実用を想定し、データ分析と可視化に関するプロトタイプシステムの検討を進めた。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 0件）

| | |
|------------------------------------------------------------------|-----------------------|
| 1. 著者名 小山元希, 湧田雄基, 吉田啓佑, 石川幸宏, 榎谷祐輝, 田中大介, 小西真治 | 4. 巻 3 |
| 2. 論文標題 XGBoost のハイパーパラメータに関する数学的考察及びインフラ維持管理への応用 | 5. 発行年 2022年 |
| 3. 雑誌名 AI・データサイエンス論文集, 第3巻, 第J2号 | 6. 最初と最後の頁 826-847 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.11532/jsceiii.3.J2_826 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |
| 1. 著者名 湧田雄基, 山下明美, 吉田啓佑, 龍田斉, 関和彦, 有井賢次, 熊谷兼太郎, 中畑和之, and 長沼諭 | 4. 巻 2 |
| 2. 論文標題 インフラ維持管理業務での機械学習活用に向けたモデリング試行と結果の解釈に関する一考察 | 5. 発行年 2021年 |
| 3. 雑誌名 AI・データサイエンス論文集 | 6. 最初と最後の頁 437-446 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.11532/jsceiii.2.J2_437 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |
| 1. 著者名 湧田雄基, 猪村元, 石川雄章 | 4. 巻 1 |
| 2. 論文標題 インフラマネジメントにおけるデータサイエンスとその迅速化を支援するデータ利活用基盤 | 5. 発行年 2020年 |
| 3. 雑誌名 AI・データサイエンス論文集, Vol. 1, No. J1 | 6. 最初と最後の頁 25-34 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.11532/jsceiii.1.J1_25 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |
| 1. 著者名 湧田雄基, 磯部正幸, 石川雄章 | 4. 巻 70 |
| 2. 論文標題 インフラ維持管理におけるデータサイエンスの実践 | 5. 発行年 2018年 |
| 3. 雑誌名 建設機械施工Vol.70 No.9 9月号 | 6. 最初と最後の頁 18-24 |
| 掲載論文のDOI (デジタルオブジェクト識別子) なし | 査読の有無 無 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

〔学会発表〕 計8件（うち招待講演 2件 / うち国際学会 1件）

| |
|----------------------------------------------------------|
| 1. 発表者名 湧田雄基, 行木孝夫, 小森大地, 吉田啓佑 |
| 2. 発表標題 北海道大学DX教育連携部門の取組 ~ 除雪データを活用した分析試行及びDX教育への展開 ~ |
| 3. 学会等名 第38回寒地技術シンポジウム, 第特1-1巻, 2022. |
| 4. 発表年 2022年 |

| |
|---------------------------------------------------------------|
| 1. 発表者名 小山元希, 湧田雄基, 阿部真育, 石川幸宏, 榎谷祐輝, 田中大介, 山口秀明 |
| 2. 発表標題 検査データを用いた地下鉄トンネル打音点検業務の効率化 |
| 3. 学会等名 映像情報メディア学会技術報告, Vol. 45, No. 4, pp. 201-202, 2021. |
| 4. 発表年 2021年 |

| |
|---------------------------------------------------------------|
| 1. 発表者名 長谷川蒼, 湧田雄基, 鈴木弘武 |
| 2. 発表標題 高速道路橋 RC 床版の健全度判定におけるデータ分析活用の検討 |
| 3. 学会等名 映像情報メディア学会技術報告, Vol. 45, No. 4, pp. 199-200, 2021. |
| 4. 発表年 2021年 |

| |
|------------------------------------------------------------------------------------------------------------|
| 1. 発表者名 Wakuda Yuki, Isobe Masayuki, Ishikawa Yusho |
| 2. 発表標題 Data Science Support System for Infrastructure Maintenance on Cloud |
| 3. 学会等名 17th International Conference on Computing in Civil and Building Engineering ICCCBE 2018 (国際学会) |
| 4. 発表年 2018年 |

| |
|---------------------------------------|
| 1. 発表者名 湧田雄基, 石川雄章, 磯部正幸 |
| 2. 発表標題 インフラ維持管理データサイエンスの体系化と試行支援 |
| 3. 学会等名 平成30年度 土木学会全国大会第73回年次学術講演会 |
| 4. 発表年 2018年 |

| |
|-----------------------------------------------------------|
| 1. 発表者名 湧田雄基 |
| 2. 発表標題 インフラ維持管理におけるデータサイエンスの活用 |
| 3. 学会等名 平成30年度 土木学会研究討論会「膨大な点検データと性能評価の齟齬をどうする！」(招待講演) |
| 4. 発表年 2018年 |

| |
|---------------------------------------------------------------|
| 1. 発表者名 湧田雄基, 安達慎一, 三浦孝智, 榎谷祐輝, 小川大貴, 小西真治, 篠原秀明, and 石川雄章 |
| 2. 発表標題 重回帰分析に基づく地下鉄トンネル上床のはく離・はく落発生要注意区間の推定 |
| 3. 学会等名 平成29年度土木学会全国大会 第72回年次学術講演会, 2017.9.12 |
| 4. 発表年 2017年 |

| |
|----------------------------------------------------|
| 1. 発表者名 湧田雄基 |
| 2. 発表標題 インフラ維持管理データサイエンスの体系化と試行支援 |
| 3. 学会等名 数理・データサイエンスの人材育成と社会応用, 2018.3.16 (招待講演) |
| 4. 発表年 2018年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

| | 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|--|---------------------------|-----------------------|----|
|--|---------------------------|-----------------------|----|

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---------|---------|
|---------|---------|