

令和 5 年 6 月 13 日現在

機関番号：12102

研究種目：基盤研究(B) (一般)

研究期間：2018～2022

課題番号：18H03244

研究課題名(和文)数量データに基づくWeb情報の信頼性検証と高信頼情報の生成

研究課題名(英文)Credibility Validation of Web Information and Generation of Credible Information based on Quantitative Data

研究代表者

加藤 誠 (Makoto P., Kato)

筑波大学・図書館情報メディア系・准教授

研究者番号：00646911

交付決定額(研究期間全体)：(直接経費) 13,100,000円

研究成果の概要(和文)：本研究課題では、Webに記述されている情報に対してその根拠となる数量データを対応づけることによってWeb情報の信頼性の担保を、また、数量データとWeb情報の対応関係から信頼性の高い情報の自動生成を実現する。この目的のために本研究課題では、以下の3つの課題に取り組んだ：1. Web上の数量データの「意味」を理解し、その意味に基づいて欠損値や誤抽出などを補完する。2. 数量データとテキストなどで記述された情報の対からその対応関係を発見し、数量データと情報の対応付けを行う。3. 学習した数量データと情報の対応関係から、意思決定において有用な情報を数量データから自動生成する。

研究成果の学術的意義や社会的意義

本研究課題では、Web上の数量データから情報を自動的に導出する方法を確立し、Web情報の信頼性の担保と信頼性の高い情報の自動生成を実現することにより、人々が信頼性の高い情報によって意思決定が行える社会を実現することを目的としている。また、現在不足しているデータサイエンティストの代わりに数量データから価値ある情報を生み出すことで、より多くの人々がビッグデータ社会の恩恵を受けることを狙っている。学術的には、これまであまり関連付けられていなかった数量データとテキストという2つの異なるモダルの対応関係を発見しようとする研究である。

研究成果の概要(英文)：This research project aims to make the reliability of Web information verifiable by mapping quantitative data to information on the Web, and to automatically generate highly reliable information based on the relationship between quantitative data and Web information. To this end, this research project addressed the following three issues: 1. Understanding the "meaning" of quantitative data on the Web for completing missing values and incorrectly extracted data; 2. Discovering the correspondence between quantitative data and information described in text; 3. Based on the learned correspondence between quantitative data and information, automatically generating information useful for decision making from the quantitative data.

研究分野：情報検索

キーワード：情報検索 データマイニング 数量データ 信頼性

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

ソーシャルネットワークサービスやユーザ生成型コンテンツの発展・普及により、多くの非専門家が情報を発信することが可能となり、Web 情報の信頼性がより深刻な社会問題となってきた。特に、医療情報に関する誤情報の氾濫は喫緊の課題であり、検索エンジン企業による分析(White, 2013)によれば、医療系質問に対する 50%以上の Web 検索結果は誤りを含んでいると報告されている。また、日本においても医療情報キュレーションサイトの信頼性がメディアに大きく取り上げられたばかりである。誤った情報に基づいて何らかの意思決定を行った場合には、自身にとって最適な選択を行えないだけでなく、深刻な被害を受ける可能性もある。医療分野に限らず、情報の信頼性の問題は人々の購買活動・人生設計(就学・就職等)や企業での経営判断、政治参画においても大きな影響を与えている問題である。

ここで、データと情報という 2 つの概念を整理することで、本研究課題における信頼性をより正確に定義する。これらの概念に対しては様々な定義があるが、本研究課題では Rowley と Hartley による定義(Rowley and Hartley, 2006)に則り、データを「目的を伴った処理や構造化がされていない、離散的かつ客観的な事実や観測値」、情報を「疑問文(何、誰、どこ、いつ、など)に答えるために処理され構造化されたデータ」とであると定義する。例えば、未加工の映像や写真、センサー機器などから得られた測定値、アンケートなどで得られた集計値、実験などで観測された値などがデータであり、Web ページに記述された文章や知識ベースなどに含まれる宣言的な知識表現が情報に該当する。これに基づいて、本研究課題では信頼性が高い情報を「多くのデータから導出された情報」と定義する。この定義を用いた場合、ある情報に対してそれを導出したデータが明示されていれば、情報の信頼性を評価することが可能である。しかしながら、一般にこの過程が明らかである場合は稀である。そのため、人々が Web 上の情報の信頼性を評価することは一般には困難であると考えられる。

2. 研究の目的

本研究課題の目的は、Web 上の数量データから情報を自動的に導出する方法を確立し、Web 情報の信頼性の担保と信頼性の高い情報の自動生成を実現することにより、人々が信頼性の高い情報によって意思決定が行える社会を実現することである。

3. 研究の方法

本研究課題では下記の課題の達成を目的とする。

(1) 数量データの収集とその集約に基づいた数量データの意味づけ

Web 上の数量データの多くは表形式によって表現されているものの、フォーマットの違いや表現の曖昧性、ラベルの曖昧性などの要因によって、各数値が何を意味しているのか理解することは容易ではない。そのため、まず数量データを大規模に収集し、それらのうち、どの数量データが同一であるのか、包含関係にあるのか、また、ある数量データが別の数量データに対して特定の演算を行った結果であるのかを推定し、数量データに特化したオントロジを自動的に構築する。各数量データに対してオントロジ中の要素を対応付けることが「数量データの意味づけ」である。これに加えて、欠損値や誤抽出などを数量データの関係性(同一性、包含、集約関係等)を用いて補完を行う方法を提案する。

(2) 数量データと情報の対応付けに基づく情報の信頼性評価

Web 上に記述されている情報を導出するような数量データを特定することを目的とする。特に、Web ページに記述されている宣言的知識を検証対象の情報とする。たとえば、「 県 市の人口は 人である」や「 県 市は治安が悪い」といった情報が例として挙げられる。これらの宣言的知識と数量データを対応付け、もし数量データが信頼できるのであって両者に齟齬があるのであれば宣言的知識の信頼性は低く、また、合致しているのであれば信頼性は高いと判断することができる。本課題における技術的な問題は、どのようにしてテキストと数量データを対応付けるかである。

(3) 数量データからの高信頼情報の生成

数量データと情報を対応付けることができるのであれば、数量データから情報を生成することも可能であると考えられる。例えば、複数の数量データに基づいて、それらの数量データから読み取れる内容を言語化することができれば、それは「多くのデータから導出された情報」であるため、信頼性の高い情報を生成できていると言える。また、テキストに限らず、数量データから何らかの意思決定に必要な情報を取り出して容易に理解可能な形式にすることで、信頼性の高い情報による意思決定が可能になると考えられる。

4. 研究成果

(1) 数量データの収集とその集約に基づいた数量データの意味づけ

表からの量的データ属性間の関係抽出

本研究では、表中の数値を表す属性の中から同一属性の対、および、上位下位関係にある属性の対を抽出する問題に対して取り組んだ（藤岡・加藤・吉川，2020）。

我々は量的データ属性であっても、同一であるような属性を抽出する方法として「異なる表中の2つの属性において、同一のエンティティを表すタブルの多くが同じ数値を含むのであればそれらの属性は同一である」という考えに基づいた方法を提案する。

量的データ属性間の上位下位関係の抽出方法としては、ある属性のすべての下位属性の値をタブルごとに合計すれば、各タブルにおけるその属性の値と等しくなるはずである、という考えを利用する。例えば、「穀物消費量」と各種穀物の消費量が記載された表において、タブルごとに米や小麦、トウモロコシの消費量などを合計し、それが同じタブルの「穀物消費量」の値と等しくなれば、「米消費量」や「小麦消費量」などは「穀物消費量」の下位属性であると判断する。

提案手法の有効性を示すために、Webからクローラされ数値を含むと判断された、103,706個の表を利用して実験を行った。同一属性抽出の問題については、比較手法としてカラム名の類似度を用いた方法と、単に値が一致するタブル数の割合を用いた方法を採用し比較を行った。実験結果から、提案手法は適合率および再現率の面から優れた結果を示し、また、従来の属性名の類似性に基づく手法とはまったく異なる属性対が得られることが明らかになった。上位下位関係にある属性対の抽出についても実験を行い、比較手法よりも高い適合率と再現率を提案手法によって実現できることを示した。

グラフニューラルネットワークを用いた表形式データの見出し認識

本研究では、表形式データにおける見出しとその階層関係の認識（見出し認識）を行う方法を提案した（笹治・加藤，2022）。図1に見出しと階層関係の例を示す。例えば、「2人以上の世帯」と「総数」、「世帯主」は見出しであり、「2人以上の世帯」と「総数」などは階層関係にある。

見出し階層の認識タスクの既存研究では、複雑な構造をもつ統計表を画像化し、表の局所的な特徴に基づいて階層関係を判定している。

しかし、局所的な特徴による認識では、複数の罫線デザインで記述した見出しや、罫線そのものを使用せずに、図形を挿入した形で見出し階層を表現した場合に、分類器が誤認識してしまう問題がある。そこで、本研究では、スプレッドシート内の統計表を構成するセルの視覚的な情報（文字列や値、罫線など）をノード特徴とするグラフデータに変換し、グラフニューラルネットワーク（GNN）を適用することで、スプレッドシートの大域的な特徴を考慮して統計表における見出しとその階層関係の認識（見出し認識）を行う方法を提案している。

また、見出し認識の教師データを用意するためには大きな労力が必要となるため、教師なし表現学習を行うことによって認識タスクに有効なセル表現を学習し、少数の教師データであっても効果的な学習が可能となる方法を提案している。表現学習手法としては、教師なし表現学習手法のDeep Graph Infomax（DGI）を採用し認識タスクに有効な特徴表現を獲得する。

実験では、政府統計ポータルサイトe-Statで公開されているスプレッドシートを使用し、GNNの教師あり学習手法が従来の機械学習による手法よりも見出し階層の認識で高い精度を達成できることを示した。また、DGIの埋め込みを用いた分類器の精度はGNNの教師あり学習手法よりも低くなった。見出しセルの識別ではノード初期特徴ベクトルを用いた分類器の精度が最も高くなることを示した。このことから、GNNの教師あり学習手法は見出し階層の認識に効果的であることが明らかになった。

	総数	2人以上の世帯 ▲			
		総数 ■	世帯主 ■	世帯主の配偶者 ■	その他 ■
正規の職員・従業員	3534	2922	1522	511	
非正規の職員・従業員 ▲	2045	1741	428	820	
パート・アルバイト ■	1407	1230	204	675	
パート ●	988	887	136	615	
アルバイト ●	421	343	69	60	

■ 見出しセル ▲ 見出しの階層関係
■ 数値セル ● 見出しの階層関係

図1. スプレッドシートにおける見出しとその階層関係の例

(2) 数量データと情報の対応付けに基づく情報の信頼性評価

統計データ検索と統計データ特定

本研究では、Webページに記述されている宣言的知識がどの統計データに基づいて書かれているかを特定する方法について、評価方法、および、実現方法についての提案を行った（中野・加藤，2021）（中野・加藤，2023）。

本研究で扱った問題は、宣言的知識中に数値が含まれている場合を対象としており、その数値

がどの統計データから得られたのかを特定する問題と、数値が統計データ内のどの箇所から得られているかを特定する問題の2つに分割することが可能である。以下では、2つの問題設定、すなわち、統計データ検索問題、統計データ特定問題ごとに説明を行う。

統計データ検索問題は、宣言的知識をクエリ、統計データ集合をコレクションとした、アドホック検索の問題として定式化を行った。さらに、クエリに含まれる文書と検索対象文書の統計データの双方のフィールドを考慮した検索モデルを提案した。本論文が取り組む統計データ検索問題においては、検索対象文書である統計データがメタデータや表などの複数のフィールドを持つ。情報検索においては、このように検索対象文書が複数のフィールドを持つ場合の検索はこれまで研究されており、BM25を拡張したBM25Fなどの検索モデルが研究されてきた。一方で、本論文の統計データ検索問題においては、クエリに含まれる文書もタイトルなどの付随する情報から複数のフィールドを持つという特徴がある。そこで本研究では、検索対象文書の統計データのフィールドだけではなくクエリのフィールドまで考慮した検索モデルとして、BM25Fを拡張したBM25FFを提案した。統計データ検索問題に対する提案手法の有効性を評価するために、データセットの構築と実験も行った。データセットの構築においては、クエリとしてWikipediaの記事を用い、被引用統計データとして政府統計データのポータルサイトであるe-Statの統計データを用いた。実験の結果、提案手法はクエリと統計データのうちの片方のフィールドのみを考慮する手法やいずれのフィールドも考慮しない手法と比較して、約2.2~3.8倍の性能を発揮することが判明した(表1のMRRを提案手法であるBM25FFと他の手法を比較)。加えてフィールドごとの分析を行った結果、エンティティに関連するクエリのフィールドが統計データの検索に重要である可能性が示された。

表1. 統計データ検索問題に対する性能の比較

	MRR		v.s. BM25 W/T/L	Hit@10		Hit@20		Hit@100
BM25	0.094		-/-/-	0.129		0.266		0.422
BM25F	0.139	(+48.0%)	219/111/113	0.318	(+147.4%)	0.345	(+29.7%)	0.442 (+4.8%)
QF-BM25	0.080	(-14.8%)	243/ 65/135	0.237	(+84.0%)	0.266	(-0.1%)	0.637 (+50.9%)
BM25FF	0.305[†]	(+225.0%)	311/ 41/ 91	0.395[†]	(+206.9%)	0.444[†]	(+66.8%)	0.731[†] (+73.3%)

統計データ特定問題は、宣言的知識と統計データが与えられたときに、統計データ内から宣言的知識の根拠となったセルを特定する問題である。我々は本問題に対して、ニューラル言語モデルを用いた新たな手法を検討した。統計データ特定問題に類似する問題として表質問応答問題が存在しており、表質問応答の多くの手法においては質問文と表全体をニューラル言語モデルに入力し、質問に対する答えを抽出・生成する。しかしながら、既存の表質問応答で想定されている表の行数や列数は統計データと比較すると小さい。そのため入力長の制限があるニューラル言語モデルにおいて、既存の表質問応答の手法をそのまま用いることは難しい。そこで本研究では表全体ではなく、行ごと・列ごとの情報を利用する表質問応答手法をベースとしたセル特定手法を提案した。この手法においてはまずテキストと表を入力として、行と列のそれぞれについてテキストに適合する確率を計算する。その後、テキストに対する行の適合確率と列の適合確率の積を取ることでテキストとセルの適合確率を計算する。さらにこの手法に対して統計データに対するセル特定問題に特有の問題を考慮しいくつかの改良を行った。

統計指標とテキストの対応付け

本研究課題では、「治安が悪い」といったような、数量データとの対応関係が明らかではないようなテキスト表現に対して、どのような統計指標が根拠となりうるかという問題について取り組んだ(笹治・加藤・小山田, 2023)。

根拠となりうる統計指標を発見する手法として、「含意認識を用いた手法」と、「相互情報量を用いた手法」の2つを提案した。前者の手法は、統計指標の相対的な量(大きいや小さいなど)に関する前提文をもとに、テキスト表現が含意される確率を求め、統計指標を含意確率によって順位付けする。例えば、「ある市は犯罪件数が多い」から「ある市は治安が悪い」かを含意認識すると「犯罪件数」という指標の含意確率が求められる。

後者の手法は、統計指標の値を入手することであるテキスト表現の予測しにくさがどの程度減るかを調べ、不確かさの減少度でランキングを作成する。例えば、「高齢者数」に関する統計量を観測することで、「高齢化が進んでいる」というテキストの不確かさはある程度減少すると考えられ、この不確かさをトピックモデルによって算出する。

本タスクに関する実験を行うため、市区町村を対象とし、対象となる統計指標(783種類)のデータセットを構築した。構築したデータセットに基づき、2つの評価指標(P@10とnDCG@10)でベースライン手法と提案手法の比較を行った。その結果、含意認識を用いた提案手法が両方の評価指標で最も高い性能を示した。

(3) 数量データからの高信頼情報の生成

数量データからの説明文生成

表形式データからその説明文を生成する方法について研究を行った(Xu, Shinden, and Kato, 2021). 提案手法では, データを特定のルールに従って単純な文に変換し, T5 や GPT-2 などの事前学習済み言語モデルに入力することによってデータの説明文生成を行った. 事前学習済み言語モデルへはデータから作られた文に加えて, 数量データを含む文書中の他の文のうち, 説明文生成に有用だと期待される文を入力として与えた.

数量データの自動可視化

情報を「疑問文に答えるために処理され構造化されたデータ」であると定義した場合には, ある目的を持って可視化されたデータは, データに基づいて生成された高信頼情報であると考えられる. 本研究では, 可視化の目的を表す「視覚化意図」に基づいて数量データを自動的に視覚化する方法について提案を行った(Maruta and Kato, 2022). ここで, 自動的な可視化とは, 視覚化意図とデータに基づいて, 適切な可視化種類(折れ線グラフ, 円グラフ, 棒グラフなど)および可視化列の予測という2つのタスクから構成される.

可視化種類の予測では双方向アテンションモデルという, 2つのデータを入力し, 片方のデータからもう片方のデータの重要な部分を予測する手法を適用した. 双方向アテンションモデルが可視化意図の中の可視化種類の予測に効果的な単語を数量データから特定し, 数量データの中の可視化種類の予測に効果的な列を可視化意図から特定することで, より効果的に可視化種類の予測を行う.

可視化列の予測では事前学習済みモデル BERT を用いて, 可視化意図と数量データの各列の対応関係を推定する. 具体的には, 可視化意図と数量データの各列のヘッダを BERT に入力し, 各ヘッダと可視化意図の類似度を推定することによって, 可視化列の予測を行う.

可視化種類の予測では, 双方向アテンションを用いた手法がより単純なベースライン手法よりも最も高い予測精度を示した. また, 数量データのみを用いた手法と可視化意図のみを用いた手法を比較すると, 可視化意図のみを用いた手法の方が高い予測精度を示した. さらに, 可視化意図と数量データを組み合わせた手法が最も高い予測精度を示したことから, 可視化意図は可視化種類の予測に有用であることが示された.

可視化列の予測では, BERT を用いた手法がベースライン手法よりも高い性能を示した. また, 提案した手法の中で, Pairwise-Column BERT という, 列のペアを BERT に入力し Pairwise 損失を用いて学習を行った手法が最も高い予測精度を示した.

< 引用文献 >

- (White, 2013) R. White. "Beliefs and Biases in Web Search" ACM SIGIR, pp. 3-12, 2013.
(Rowley and Hartley, 2006) J. Rowley, R. Hartley. Organizing Knowledge: An Introduction to Managing Access to Information. Ashgate Publishing, 2006.
(藤岡・加藤・吉川, 2020) 藤岡 周平, 加藤 誠, 吉川 正俊: 表からの量的データ属性間の関係抽出. 情報処理学会論文誌データベース (TOD) 13, pp. 10-21, 2020.
(笹治・加藤, 2022) 笹治 拓矢, 加藤 誠: グラフニューラルネットワークを用いた表形式データの見出し認識. 電子情報通信学会論文誌 D J105-D, No. 5, pp. 360-371, 2022.
(中野・加藤, 2021) 中野 優, 加藤 誠: クエリと文書のフィールドを考慮した被引用統計データの検索. 情報処理学会論文誌 データベース (TOD) 14, pp. 49-60, 2021.
(中野・加藤, 2023) 中野 優, 加藤 誠: 被引用統計データのセル特定手法の検討. 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 (DEIM 2023), 2023.
(笹治・加藤・小山田, 2023) 笹治 拓矢, 加藤 誠, 小山田 昌史: 含意認識モデルとニューラルトピックモデルを用いた統計指標検索. 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 (DEIM 2023), 2023.
(Xu, Shinden, and Kato, 2021) Junjie H. Xu, Kohei Shinden, Makoto P. Kato: Table Caption Generation in Scholarly Documents Leveraging Pre-trained Language Models. Proc. of the 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE 2021), pp.963-966, 2021.
(Maruta and Kato, 2022) Atsuki Maruta, Makoto P. Kato: Intent-Aware Data Visualization Recommendation. Data Science and Engineering 7, pp. 301-315, 2022.

5. 主な発表論文等

〔雑誌論文〕 計14件（うち査読付論文 14件 / うち国際共著 1件 / うちオープンアクセス 0件）

1. 著者名 Makoto P. Kato, Hiroaki Ohshima, Ying-Hsang Liu, Hsin-Liang Chen	4. 巻 -
2. 論文標題 A Test Collection for Ad-hoc Dataset Retrieval	5. 発行年 2021年
3. 雑誌名 Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval	6. 最初と最後の頁 2450-2456
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3404835.3463261	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 該当する
1. 著者名 Junjie H. Xu, Kohei Shinden, Makoto P. Kato	4. 巻 -
2. 論文標題 Table Caption Generation in Scholarly Documents Leveraging Pre-trained Language Models	5. 発行年 2021年
3. 雑誌名 Proceedings of the 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE 2021)	6. 最初と最後の頁 963-966
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/GCCE53005.2021.9621890	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 Atsuki Maruta, Makoto P. Kato	4. 巻 -
2. 論文標題 Intent-aware Visualization Recommendation for Tabular Data	5. 発行年 2021年
3. 雑誌名 Proceedings of the 22nd International Conference on Web Information Systems Engineering (WISE 2021)	6. 最初と最後の頁 252-266
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-91560-5_18	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 中野 優, 加藤 誠	4. 巻 14
2. 論文標題 クエリと文書のフィールドを考慮した被引用統計データの検索	5. 発行年 2021年
3. 雑誌名 情報処理学会論文誌データベース (TOD)	6. 最初と最後の頁 49-60
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 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 Wiradee Imrattana-trai, 加藤 誠, 吉川 正俊	4. 巻 18-J
2. 論文標題 ゼロショット学習によるテキストからのエンティティプロパティ同定	5. 発行年 2020年
3. 雑誌名 日本データベース学会和文論文誌	6. 最初と最後の頁 -
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 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 Jun-Li Lu, Makoto P. Kato, Takehiro Yamamoto, Katsumi Tanaka	4. 巻 13(2)
2. 論文標題 Searching for Microblogs Referring to Events by Deep Dynamic Query Strategies	5. 発行年 2020年
3. 雑誌名 情報処理学会論文誌データベース (TOD)	6. 最初と最後の頁 1882-7799
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 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 藤岡 周平, 加藤 誠, 吉川 正俊	4. 巻 13(3)
2. 論文標題 表からの量的データ属性間の関係抽出	5. 発行年 2020年
3. 雑誌名 情報処理学会論文誌データベース (TOD)	6. 最初と最後の頁 10 ~ 21
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 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 IMRATTANATRAI Wiradee, KATO Makoto P., TANAKA Katsumi, YOSHIKAWA Masatoshi	4. 巻 E101.D
2. 論文標題 Entity Ranking for Queries with Modifiers Based on Knowledge Bases and Web Search Results	5. 発行年 2018年
3. 雑誌名 IEICE Transactions on Information and Systems	6. 最初と最後の頁 2279 ~ 2290
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transinf.2017EDP7372	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 笹治 拓矢、加藤 誠	4. 巻 J105-D
2. 論文標題 グラフニューラルネットワークを用いた表形式データの見出し認識	5. 発行年 2022年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 360 ~ 371
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2021DET0003	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 中野 優, 加藤 誠	4. 巻 1
2. 論文標題 被引用統計データのセル特定データセットの構築	5. 発行年 2022年
3. 雑誌名 論文誌「データドリブンスタディーズ」	6. 最初と最後の頁 -
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 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 Maruta Atsuki, Kato Makoto P.	4. 巻 7
2. 論文標題 Intent-Aware Data Visualization Recommendation	5. 発行年 2022年
3. 雑誌名 Data Science and Engineering	6. 最初と最後の頁 301 ~ 315
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41019-022-00191-7	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 Yamamoto Yusuke, Yamamoto Takehiro	4. 巻 -
2. 論文標題 Personalization Finder: A Search Interface for Identifying and Self-controlling Web Search Personalization	5. 発行年 2020年
3. 雑誌名 JCDL '20: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020	6. 最初と最後の頁 37 ~ 46
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3383583.3398519	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 Imrattanatrai Wiradee, Kato Makoto P., Yoshikawa Masatoshi	4. 巻 -
2. 論文標題 Identifying Entity Properties from Text with Zero-shot Learning	5. 発行年 2019年
3. 雑誌名 Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval	6. 最初と最後の頁 195 ~ 204
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3331184.3331220	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 Kato Makoto P., Imrattanatrai Wiradee, Yamamoto Takehiro, Ohshima Hiroaki, Tanaka Katsumi	4. 巻 -
2. 論文標題 Context-Guided Learning to Rank Entities	5. 発行年 2020年
3. 雑誌名 Proceedings of the 42nd European Conference on IR Research	6. 最初と最後の頁 83 ~ 96
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-45439-5_6	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 -

〔学会発表〕 計26件 (うち招待講演 0件 / うち国際学会 2件)

1. 発表者名 和久井 拓斗, 加藤 誠
2. 発表標題 統計データ収集のためのフォーカストクローラ
3. 学会等名 第14回データ工学と情報マネジメントに関するフォーラム (DEIM 2022)
4. 発表年 2022年

1. 発表者名 中野 優, 加藤 誠
2. 発表標題 被引用統計データのセル特定データセットの構築
3. 学会等名 第14回データ工学と情報マネジメントに関するフォーラム (DEIM 2022)
4. 発表年 2022年

1. 発表者名 Makoto P. Kato, Hiroaki Ohshima, Ying-Hsang Liu, Hsin-Liang Chen
2. 発表標題 Overview of the NTCIR-15 Data Search Task
3. 学会等名 Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-15) (国際学会)
4. 発表年 2020年

1. 発表者名 中野 優、加藤 誠
2. 発表標題 誤引用検証のための被引用統計データの検索
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 Junjie H. Xu, Wiradee Imrattana-trai, Makoto P. Kato
2. 発表標題 Toward Description Generation for Tables in Scientific Articles
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 丸田 敦貴、加藤 誠
2. 発表標題 視覚化意図を考慮したデータの効果的な視覚化方法の推定
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 笹治 拓矢、加藤 誠
2. 発表標題 グラフニューラルネットワークを用いたスプレッドシートの見出し認識
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 Ryota Mibayashi, Huulong Pham, Naoaki Matsumoto, Takehiro Yamamoto and Hiroaki Ohshima
2. 発表標題 Uhai at the NTCIR-15 Data Search Task
3. 学会等名 Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-15) (国際学会)
4. 発表年 2020年

1. 発表者名 勝又千夏、山本祐輔
2. 発表標題 社会的不利益と後悔の回避のためのソーシャルメディアデータの忘却ニーズに関する調査
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 永野里佳奈、山本祐輔
2. 発表標題 芸術作品に興味を促すキャプションの自動生成
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 清水勇祐、山本祐輔
2. 発表標題 ウェブ検索における便益とプライバシーリスクを考えさせる情報提示手法
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 梅田浩郎、山本祐輔
2. 発表標題 ユーモアセンスを向上させる文章作成支援インターフェース
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 若月祐樹、山本祐輔
2. 発表標題 どこでもプレスト:実世界情報をトリガとするアイデア
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 奥瀬雄哉、山本祐輔
2. 発表標題 分野横断の架け橋となる情報推薦
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 鈴木雅貴、山本祐輔
2. 発表標題 確証バイアスとウェブ検索行動の関係分析
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 庵谷拓輝、山本祐輔
2. 発表標題 評価値に翻弄されないレビュー情報探索を促す検索結果スニペット
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 齊藤史明、山本祐輔
2. 発表標題 トピック理解のためのより貪欲な情報探索を促進する問いかけ文の提示
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 崔 洙瑚、橋口 友哉、木村 壘、大島 裕明
2. 発表標題 ウェブ閲覧URL列へのTransformer Encoderの適用による広告閲覧ユーザの属性推定
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 顧 席銘、窪内 将隆、山本 岳洋、大島 裕明
2. 発表標題 教師なしドメイン適応を用いた粒子形状評価
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 川原 敬史、橋口 友哉、湯本 高行、大島 裕明
2. 発表標題 事故概要説明文の文脈理解による傷病の程度の推定
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 吉村 拓真、ファム フーロン、和田 真弥、大島 裕明
2. 発表標題 空間的特徴量とパーソナリティ分析を用いたストレス推定
3. 学会等名 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 (DEIM 2021)
4. 発表年 2021年

1. 発表者名 Wiradee Imrattanatrai, 加藤 誠, 吉川 正俊
2. 発表標題 ゼロショット学習によるテキストからのエンティティプロパティ同定
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2019年

1. 発表者名 白髪 宙海, 村本 直樹, 高橋 克郎, 大島 裕明
2. 発表標題 飲食店レビューにおける再訪問ユーザの行動の分析
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2019年

1. 発表者名 白髪 宙海, 村本 直樹, 大島 裕明
2. 発表標題 食べログの再訪問レビューの発見
3. 学会等名 第11回Webとデータベースに関するフォーラム
4. 発表年 2018年

1. 発表者名 笹治 拓矢, 加藤 誠, 小山田 昌史
2. 発表標題 合意認識モデルとニューラルトピックモデルを用いた統計指標検索
3. 学会等名 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 (DEIM 2023)
4. 発表年 2023年

1. 発表者名 中野 優, 加藤 誠
2. 発表標題 被引用統計データのセル特定手法の検討
3. 学会等名 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 (DEIM 2023)
4. 発表年 2023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
研究分担者	山本 祐輔 (Yamamoto Yusuke) (50625431)	名古屋市立大学・データサイエンス学部・准教授 (23903)	
研究分担者	大島 裕明 (Ohshima Hiroaki) (90452317)	兵庫県立大学・情報科学研究科・准教授 (24506)	
研究分担者	吉川 正俊 (Yoshikawa Masatoshi) (30182736)	京都大学・情報学研究科・教授 (14301)	
研究分担者	山本 岳洋 (Yamamoto Takehiro) (70717636)	兵庫県立大学・情報科学研究科・准教授 (24506)	

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
---------	---------