

令和元年5月30日現在

機関番号：22604

研究種目：基盤研究(B) (一般)

研究期間：2015～2018

課題番号：15H02780

研究課題名(和文) 動向情報エコシステム実現のための情報アクセス・活用支援

研究課題名(英文) Support of Information Access and Utilization for Trend Information Ecosystem

研究代表者

高間 康史 (Takama, Yasufumi)

首都大学東京・システムデザイン研究科・教授

研究者番号：20313364

交付決定額(研究期間全体)：(直接経費) 12,500,000円

研究成果の概要(和文)：本研究では多種多様な動向情報へのアクセス・活用を支援する基盤技術について研究し、情報の活用が既存情報へ付加価値を与えると同時に、新しい情報の創出につながる動向情報エコシステムの実現を目的とする。(1)動向情報のアクセス・活用を支援する技術、(2)ベストプラクティス共有の枠組み、(3)開発したシステムによる実証実験に取り組み、(1)時系列データを対象とした検索エンジンの機能拡張、テキストストリームデータを対象としたモニタリング支援システムの開発、(2)クエリ推薦や知見共有の枠組みの研究などを行った。(3)に関して、クエリ推薦によるベストプラクティス共有等に関して有効性が認められる結果が得られた。

研究成果の学術的意義や社会的意義

ICTの急速な進化によりもたらされる、多種多様かつ膨大なデータの生成・流通・蓄積を促進し、資源として活用する事で新たな価値の創造や社会的課題の解決につなげる取り組みが活発化している。本研究は個々のユーザーによる動向情報へのアクセスおよび活用を支援する技術・システムに関するものであり、現状進められているデータ流通・公開のインフラ整備と組み合わせることで、データの利活用がデータの流通・公開を促進し、さらなる価値の創造を持続的に生み出す動向情報エコシステムの実現に将来的に貢献することが期待できる。

研究成果の概要(英文)：This research project aims to study an ecosystem for trend information that leads to new information creation through the sharing of the best practice of information access and utilization. Sub-goals of the project include: (1) technologies for supporting access and utilization of trend information, (2) framework for sharing the best practice, and (3) user experiments using the developed systems. As the main contributions of the project, (1) we developed an advanced version of the search engine designed for retrieving time series data, and monitoring support systems for text stream data. As for (2), we developed the framework for best-practice sharing, such as query recommendation function and the sharing of findings from trend information. Through the user experiments using the developed system, the effectiveness of sharing the best practice of information access/utilization is confirmed.

研究分野：ウェブインテリジェンス

キーワード：ウェブインテリジェンス 動向情報

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

ICTの急速な進化によりもたらされる、多種多様かつ膨大なデータの生成・流通・蓄積を促進し、資源として活用する事で新たな価値の創造や社会的課題の解決につなげる取り組みが活発化している。本研究では、データの利活用がデータの流通・公開を促進し、さらなる価値の創造を持続的に生み出す構図を情報エコシステムとして捉える。現在では、オープンデータやデータ・マーケットプレイスなどの枠組みや、Twitter やオンラインニュースなどの多様なサービスにより、データの流通・公開のインフラは整備されているが、十分活用されているとは言いがたい現状にある。今後情報エコシステムとして機能させていくためには、個々のユーザにおける情報アクセス・活用を促進する必要があると考える。そのためには、データの大規模化だけでなく、多様性やストリーム性といったデータの質の変化を考慮した新たなアクセス・活用支援手段の確立が必要と考える。

データの多様性に関しては、異なるデータリソースを組み合わせる事で新たな価値を生み出す事が期待されているが、分散して存在するリソースの中から新たな価値を生み出す組合せを発見する事は現状困難である。また、オンラインニュースなどストリーム性のあるテキストデータへのアクセス支援として、オンラインニュースのキュレーションサービスが現在人気を集めているが、欲しい情報が必要な時に存在するとは限らないためデータを確認すべきタイミングを見極める事が重要という、ストリームデータの特性は十分考慮されていない。

2. 研究の目的

本研究では流行や時系列データの時間的変動といった、多種多様な動向情報へのアクセス・活用を支援する基盤技術について研究し、情報の活用が既存情報へ付加価値を与えると同時に、新しい情報の創出につながる動向情報エコシステムの実現を目的とする。挑戦的萌芽研究で開発した動向情報検索システム(コンテキスト検索エンジン)などを活用して、情報活用のベストプラクティスを共有可能な知識情報基盤を確立する。

3. 研究の方法

情報エコシステムを実現するためには、情報アクセスや活用を個人的かつ単発的な行為に留めるのではなく、ベストプラクティスを蓄積、共有する事で、既存情報に付加価値を与えると共に、新たなデータの創出・公開を促進する仕組みが必要と考える。例えば、リソースの有効な組合せは、リソースの価値を高める付加価値と見なす事ができる。これをベストプラクティスとして共有できれば、既存リソースの活用促進につながる事が期待できる。本研究では、データの多様性、ストリーム性を動向情報という統一的な観点から捉える事で、多数の利用者による多様な情報アクセス・活用行為間のベストプラクティス共有を実現する。以上より、研究課題は以下の3点に大別される。

(1) 動向情報のアクセス・活用を支援する技術を確認する。前述のコンテキスト検索エンジンの改良やデータベースの拡充を行う他、オンラインニュースやBBQ(電子掲示板)、Twitter等のテキストストリームデータを対象としたモニタリング支援システムを開発する。

(2) 情報活用を個人の単発的な行為にとどめるのではなく、アクセスや活用に関するベストプラクティスを共有する事で、既存情報(リソース)に付加価値を与えると共に、新たな情報の創出につながる枠組みを実現する。

(3) 上記(1)(2)で開発したシステムを実際に運用し、ベストプラクティスの共有、およびそれによるデータの活用・公開が促進される効果について検証する。

4. 研究成果

(1) 動向情報のアクセス・活用支援技術

コンテキスト検索エンジンは、時系列データを対象とした検索エンジンであり、Webから収集した各種時系列データに対し、値がピークをとる時期や急下降した時期など、特徴的変動が観測された時期を検索することなどが可能である。図1に、開発したコンテキスト検索エンジンのスクリーンショットを示す。本研究では、コンテキスト検索エンジンに対し以下の拡張を行った。



・検索可能な時系列データの拡張：
特に、Wikipedia 閲覧データが今後

図1 コンテキスト検索エンジンのスクリーンショット

の実験に有益な内容を含むことを発見し、データ収集、前処理などを行いデータベースに追加を行った。これにより、136万件以上の多様なアイテムを検索可能とした。

・検索結果ランキング機能の導入：ランク学習を利用する点は一般的な Web 検索エンジンと同様であるが、コンテキスト検索エンジンに特有のユーザ意図を分析し、意図ごとに学習を行う方法を提案し、実装した。また、学習に利用する素性についても、周期性や変化の急峻さ、長期的トレンド、各特徴的変動の発生頻度など、時系列データに特有のものを検討し、システムに取り入れた。

・検索結果画面の改良：一般的な Web 検索エンジンでは、クエリとして指定した単語周辺の本文をスニペットとして検索結果画面に含めることで、実際にページを閲覧するかどうかを判断する手掛かりを提供している。同様の役割を果たすものとして、小型の可視化である Sparkline に着目し、これを検索結果画面に導入した。

・論理演算機能の導入：各クエリに対する検索結果間で論理演算を適用する機能を導入した。これにより、例えば 2016 年 1 月、2017 年 1 月、2018 年 1 月にそれぞれピークを持つアイテムを検索し、それらの AND をとることで、1 月に周期的にピークを持つアイテムの発見が可能となる。

モニタリング支援システムに関しては、オンラインニュースなどのテキストストリームデータを対象として研究を進めたが、機能的な面では確認タイミングに着目した支援技術が本研究の特徴として挙げられる。テキストストリームデータはユーザにとって有益な情報を含んでおり、定期的に関連するユーザは多数存在するが、随時到着する情報を常時確認することは不可能である。一般には、休憩時間などにこれらストリームデータを確認するなど、メインタスクの合間に確認作業が行われる。見方を変えれば、モニタリングを行うためにメインタスクが中断されることになるため、作業割込みによる知的生産性の低下が懸念される。本研究では、メインタスクの途中でテキストストリームの確認を行う状況を想定し、「ユーザが重要視する目安に基づいてテキストストリームを確認できるタイミング」を適切な確認タイミングと定義する。適切な確認タイミングをユーザ自身が判断するための手掛かりとなる指標を提示することによりモニタリング支援を行う。

適切な確認タイミングに影響を与える要因として、貯まっている（確認前の）テキストストリーム量、その重要度、メインタスクの状況の 3 点を考慮する。これらに関する手がかりとして、蓄積データ量ゲージ、重要データリスト、作業度グラフの 3 種類の可視化結果をユーザに提示する。また、明度によりユーザの注意を惹く度合いを調整する。

20 代の工学系大学生、大学院生 12 名に試作したシステムを 5 時間利用してもらい、評価実験を行った結果、以下の結果が得られた。

・確認時の平均データ量について、データ量を重要視した実験協力者とそれ以外の実験協力者では有意差が見られた。

・確認時に重要記事が複数あった割合について、重要データの配信を重視した実験協力者とそれ以外の実験協力者で有意差がみられた。

これらの結果より、実験協力者ごとに異なる確認タイミングに応じて、提案システムは適切な手がかりを提示できていることが確認された。

(2) ベストプラクティス共有の枠組み

コンテキスト検索エンジンにおいて、その検索履歴から有益なリソースの組合せを発見可能なクエリを推薦する機能を、ベストプラクティス共有機能として導入した。具体的には、ユーザが連続して実行したクエリ対 (A, B) に着目し、B についての検索結果に対するインタラクションの方が増加した場合にベストプラクティス候補と判断し、推薦に利用する。推薦クエリの決定方法は、クエリ構造に基づく推薦、検索ユーザの嗜好に基づく推薦の 2 種類を提案した。前者では、推薦の際には、入力されたクエリと A が構造的に類似する場合に B を推薦する。後者では、クエリログからユーザの好むアイテムを推定し、ユーザが入力したクエリと A の検索結果が類似し、かつ B の検索結果上位にユーザの好むアイテムが多く含まれる場合に B を推薦する。

また、オープンにできないデータ（ローカルデータ）も検索可能となるようにシステムを拡張し、検索結果に基づき発見した知見のみをデータジャケットの枠組みを用いて共有可能な仕組みも導入した。

(1) で述べたモニタリング支援システムにおいては、ユーザごとに確認タイミングの決定基準が質（データ量あるいは重要記事など）・量（どの程度蓄積されたときに確認するかなど）とも異なることから、ユーザごとに過去の確認状況を分析し、過去に確認したときと同様の状況になったときにユーザの注意を明度変化により引き付ける機能を導入した。具体的には、前回確認時からの経過時間が当該ユーザの平均確認間隔を越えているとき、データ蓄積量が当該ユーザが確認した際の平均値を越えているときなどに、ユーザの注意を引き付けるように設定した。(1) で述べた実験において、提案システムによる明度変化が確認タイミングに影響を与えたとみなせる実験協力者の存在を確認した。

(3) ベストプラクティス共有に関する実証実験

上記で述べたコンテキスト検索エンジンを用いて、データ分析に関する演習を複数回行い、ベストプラクティス共有に関する実証実験を行った。以下に、実施した実証実験の一部についてその概要を示す。

(3-1) クエリ推薦機能に関する実証実験

ベストプラクティス候補を収集する予備実験、クエリ推薦を行う本実験の2回に分けて実施した。予備実験では、工学系大学生・大学院生15名に協力してもらい、コンテキスト検索エンジンを用いて3か所の空欄を含む年表を作成してもらい、他の人に空欄にあてはまるアイテムを回答してもらうゲーム形式の実験を行った。クイズの正解者、正解された問題作成者の両方に得点が入るようにし、簡単すぎず、かつ誰も答えられないようなものではない、適度な難易度の問題を作成した場合に高得点が得られるようにした。これにより、問題作成時に多様な検索を試行する動機を与えることで、多様なベストプラクティス候補を収集することを目的とした。

本実験では、工学系大学生・大学院生10名に協力してもらい、コンテキスト検索エンジンを用いて予備実験で作成したゲームに回答してもらった。この際、(2)で説明した手法によりクエリの推薦を行った。実験の結果、回答してもらった合計30問のうち、クエリ構造の類似度に基づく推薦は16問で推薦したクエリが利用され、ユーザの好みに基づく推薦は6問で利用され、クエリ推薦によるベストプラクティスの共有効果が示された。

(3-2) 効率的な検索行動の共有に関する実証実験

本実験では、コンテキスト検索エンジンを利用した検索行動ログを分析し、クエリ生成に関する不十分な理解や検索結果の効率的な探索方法などについて明らかにした結果から、効率的な検索行動に関するガイドラインを作成した。作成したガイドラインの一部を以下に示す。

- ・「同時期に同じ特徴的変動を示したアイテムの検索」は多数のアイテムがヒットするが多いので、関心のある期間を指定して検索する方がよい
- ・特徴的変動のうち、MAXとMINをそれぞれ指定した検索結果のAND演算は無意味
- ・検索結果を詳しく確認するのは、検索結果をある程度絞り込んでからにする
- ・出来事についての関連アイテムを検索する場合は、ニュース記事などでその出来事の発生時期を確認してから検索を行う

工学系大学生・大学院生11名に協力してもらい、指定した新聞記事を読んでもらい、それに関連するアイテムを発見してもらう実験、および自身の関心あるアイテムと何らかの関連があるアイテムを見つけてもらう実験などを行った。ガイドラインを読まずに一度実験してもらった後、ガイドラインを読んでから実験を再度行ってもらった。両実験で検索行動に変化がみられるか分析したところ、以下の違いが観測された。

- ・1クエリで検索を終える場合の割合が、47.9%から33.3%に低下
- ・複雑な検索クエリの増加：アイテムと期間を組み合わせたクエリの割合が38%から62%に増加
- ・検索結果下位まで確認する傾向の増加
- ・利用する特徴的変動の変化：MAXの多用からPEAKの多用へ。(PEAKはアイテムにつき複数回発生しうるため、関連アイテムの発見などに有効であることが想定される。)

以上より、検索行動に関するベストプラクティスを共有する効果が認められる結果が得られた。

5. 主な発表論文等

[雑誌論文](計17件)

1. S. Song, S. Yamada, Ambient Lights Influence Perception and Decision-Making, *Frontiers in Psychology*, Vol. 9, 2019, 10.3389/fpsyg.2018.02685 (査読あり)
2. P. Cheng, K. Kobayashi, T. Hashimoto, K. Saito, Y. Takama, Development of Property Search System for Renovation, *知能と情報*, Vol. 30, pp. 815-822, 2018, 10.3156/jsoft.30.6_815 (査読あり)
3. M. Endo, M. Hirota, S. Ohno, H. Ishikawa, Best-time Estimation Method using Information Interpolation for Sightseeing Spots, *International Journal of Informatics Society*, Vol. 10, pp. 97-105, 2018 (査読あり)
4. M. Okabe, S. Yamada, Clustering Using Boosted Constrained k-Means Algorithm, *Frontiers in Robotics and AI*, Vol. 5, p. 18, 2018, 10.3389/frobt.2018.00018 (査読あり)
5. H. Ishikawa, R. Chbeir, An integrated data model for management and mining of social big data: Concepts and Applications, *International Journal on Advances in Networks and Services*, Vol. 10, pp. 70-81, 2017 (査読あり)

6. S. Kori, K. Yamaguchi, Y. Zhu, S. Takiguchi, Y. Kato, H. Ishikawa, Y. Takama, Application of Search Engine Focusing on Trend-related Queries to Market of Data, Intelligent Decision Technologies, Vol. 10, pp. 263-272, 2016, 10.3233/IDT-150250 (査読あり)
7. K. Okamoto, R. Yamanishi, M. Matsushita, Sound-effects Exploratory Retrieval System Based on Various Aspects, IEEJ Transactions on Electronics, Information and Systems, Vol. 136, pp. 1712-1720, 2016, 10.1541/ieejieiss.136.1712 (査読あり)
8. Y. Takama, M. Okumura, Interactive Visualization System for Monitoring Support Targeting Multiple BBS Threads, International Journal on Intelligent Decision Technologies, Vol. 9, pp. 391-403, 2015, 10.3233/IDT-140232 (査読あり)

[学会発表] (計 40 件)

1. 岡部, アンサンブル学習に基づく半教師ありクラスタリング, 電気学会スマートシステムと制御技術シンポジウム 2019, 2019.
2. S. Fujii, M. Hirota, D. Kato, T. Araki, M. Endo, H. Ishikawa, Analysis of the Difference of Movement Trajectory by Residents and Tourists using Geotagged Tweet, GEOProcessing 2019, 2019.
3. 高間, 吉田, 確認タイミングに着目したテキストストリームモニタリング支援システムの提案, 第 12 回 W12 研究会, 2018.
4. R. Takami, Y. Takama, Proposal of visual analytics interface for time series data employing trajectory manipulation, PacificVis2018, 2018.
5. Y. Shirai, S. Hattori, Y. Takama, Analyzing Resident's Lifestyle from Household Electricity Consumption Data, ISCIIA&ITCA2018, 2018.
6. 岡久, 高間, ローカルデータ・公開データを統合利用可能なコンテキスト検索エンジンの提案, JSAI2018, 2018.
7. 高見, 高間, 軌跡の直接操作に基づく時系列データの視覚的分析支援インタフェースの設計と評価, JSAI2018, 2018.
8. 岩崎, 松下, 予想材料に基づく為替の動向情報分析を目的としたニュース記事分類手法, 第 20 回インタラクティブ情報アクセスと可視化マイニング研究会予稿集, 2018.
9. Y. Takama, K. Yamaguchi, L. Chen, H. Ishikawa, Introduction of Sparkline as Snippet to Context Search Engine Result Page, TAAI2017, 2017.
10. 佐藤, 高間, コンテキスト検索エンジンを利用した時系列データマイニングの提案, JSAI2017, 2017.
11. 盛山, 坂井, 松下, 動向情報可視化システムを用いたユーザの行動モデルの分析に関する研究, JSAI2017, 2017.
12. H. Ishikawa, R. Chbeir, A Data Model for Integrating Data Management and Data Mining in Social Big Data, The Ninth International Conferences on Advances in Multimedia, 2017.
13. S. Song, S. Yamada, Expressing Emotions through Color, Sound, and Vibration with an Appearance-Constrained Social Robot, HRI2017, 2017.
14. 手塚, 山口, 高間, コンテキスト検索エンジンを対象としたランキング機能の提案, JSAI2016, 2016.
15. 吉田, 高間, ストリームデータモニタリングにおける確認タイミングの判断支援インタフェースの提案, JSAI2016, 2016.
16. M. Yasuo, N. Shirozu, M. Matsushita, Assessing Spatiotemporal Context of User's Daily Behavior to Facilitate Subtle Deviation, 5th Asian Conference on Information Systems, 2016.
17. 内藤, 古田, 松下, 時系列データの探索的分析を支援する可視化システム: 記事と時系列データのアラインメント方式の提案, 第 13 回 SIG-AM, 2016.
18. 盛山, 松下, 高間, コンテキスト検索による情報探索行為を支援するための可視化インタフェースの提案, JSAI2015, 2015.
19. 山口, 諸, 桑折, 高間, コンテキスト検索エンジンのインタフェース向上に関する検討, JSAI2015, 2015.
20. 高間, Y. Zhu, 桑折, 山口, 瀧口, 動向に関する問いに答えるコンテキスト検索エンジンの開発, 第 10 回 SIG-AM, 2015.
21. 手塚, 山口, Y. Zhu, 桑折, 高間, コンテキスト検索エンジンへのランキング機能の導入に関する検討, 第 11 回 SIG-AM, 2015.
22. S. Kori, Y. Zhu, K. Yamaguchi, S. Takiguchi, Y. Takama, Analysis of User's Behaviour Based on Search Intentions for Information Retrieval Using Search Engines, TAAI2015, 2015.
23. Y. Moriyama, M. Matsushita, Y. Takama, Visual User Interface to Supporting Information Seeking Behavior in Context Searching, TAAI2015, 2015.

〔図書〕(計 2件)

1. 石川(監著), 横山, 廣田(著), フルスタック JavaScript と Python 機械学習ライブラリで実践するソーシャルビッグデータ-基本概念・技術から収集・分析・可視化まで, コロナ社, 192 ページ, 2019.
2. 高間, 情報可視化 データ解析・活用のためのしくみと考え方, 森北出版, 128 ページ, 2017.

〔産業財産権〕

出願状況(計 0件)

取得状況(計 0件)

〔その他〕

ホームページ等

<http://krectmt3.sd.tmu.ac.jp/cse.html>

6. 研究組織

(1)研究分担者

研究分担者氏名: 山田 誠二

ローマ字氏名: (YAMADA, seiji)

所属研究機関名: 国立情報学研究所

部局名: コンテンツ科学研究系

職名: 教授

研究者番号(8桁): 50220380

研究分担者氏名: 岡部 正幸

ローマ字氏名: (OKABE, masayuki)

所属研究機関名: 県立広島大学

部局名: 経営情報学部

職名: 講師

研究者番号(8桁): 50362330

研究分担者氏名: 松下 光範

ローマ字氏名: (MATSUSHITA, mitsunori)

所属研究機関名: 関西大学

部局名: 総合情報学部

職名: 教授

研究者番号(8桁): 50396123

研究分担者氏名: 石川 博

ローマ字氏名: (ISHIKAWA, hiroshi)

所属研究機関名: 首都大学東京

部局名: システムデザイン学部

職名: 教授

研究者番号(8桁): 60326014

(2)研究協力者

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。