

令和元年6月21日現在

機関番号：24201

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00307

研究課題名(和文) データ分析過程の共有による知識創発支援

研究課題名(英文) Knowledge Emergence Support by Sharing Data Analysis Process

研究代表者

砂山 渡 (Sunayama, Wataru)

滋賀県立大学・工学部・教授

研究者番号：40314398

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)：1) テキストマイニングツールTETDMに操作ログの保存と再現機能を実装した。2) ログデータ共有環境として、保存されたデータ分析環境の操作ログを、他のユーザと共有するため、任意にアップロード、ダウンロードすることが可能なサーバの構築を行った。3) データ分析プロセスになぞらえたTETDM上で実施するチュートリアル課題、ならびにツール選択支援機能を用意した。4) 収集したログデータからスキル獲得を支援する操作履歴の推薦を行い、スキル獲得に繋がる効果があるかを実験的に検証した。

研究成果の学術的意義や社会的意義

データ分析の需要や必要性が高まっている反面、データ分析として具体的にどのような手順で分析を行えば良いのかわからない人が多く、そのためには手本を見せる必要があった。提案システムによって、他人がどのように分析操作を行っているかを確認することができるようになり、その操作をまねて、自ら分析方法を学ぶことができるようになった。また情報推薦分野では、例えばユーザの購買行動に類似するユーザが購入した商品を推薦することが多いが、本システムでは逆に類似しない、ユーザがまだ獲得していないスキルに関わる操作履歴を推薦するシステムを提案している点が新しい。

研究成果の概要(英文)：1) Functions for saving and loading operation history had implemented to a text mining tool TETDM. 2) A server had constructed to share uploaded operation histories. 3) Tutorials and support functions for selecting mining tools had prepared for learning data analysis process. 4) According to experimental results, the proposed methods were effective to mining skill acquisition.

研究分野：知能情報学

キーワード：知識創発 テキストマイニング データサイエンス スキル獲得 操作履歴

1. 研究開始当初の背景

ビッグデータの活用に注目が集まる中、マイナンバー制度が施行されるなど、データを収集する環境が整えられるとともに、データマイニングと呼ばれるデータ処理技術の開発が盛んに行われている。しかし現状では、必ずしもビッグデータが有効に活用されておらず、その原因として、人間がデータを活用する際の環境が整えられていないこと、ならびにデータを処理してその分析を行える人材が少ないことが挙げられる。

2. 研究の目的

本研究では、多くの人々が共通のツールを利用してデータ分析を行う際の、データ分析過程を共有できる枠組みにより、お互いにデータ分析のためのノウハウを共有でき、データ活用に向けた知識創発を支援できる環境を構築する。

知識創発のプロセスは、多様な分析の結果を集める「発散」フェーズと、集められた結果をまとめる「収束」フェーズからなり、本研究ではこの発散フェーズにおいて、ユーザ個人が得られる分析結果の多様性を増すことを目指す。すなわち、ユーザがデータ分析を行う際に、独力で得られる分析結果と比較した際に、他のユーザのデータ分析時の操作を見て、その操作から得た発想に基づいて操作と分析を行うことで、多様な分析結果を集められるようにする。これにより大局的で客観的な知見を見いだす知識創発の支援を行う。

3. 研究の方法

データ分析過程の共有による知識創発支援の枠組みに必要な構成要素を実装する。本枠組みは、データ分析過程として、データ分析環境(TETDM)上のユーザの操作ログを、ユーザが利用(保存または再現)するための環境と、操作ログの集合を管理して共有(操作ログを任意にアップロード、ダウンロードすることが可能)するための環境からなる。

また、提案する枠組みの中で、操作ログを用いて、効果的に知識創発を支援する方法の検討と実装を行う。また、データの意味解釈を目的としたデータ分析作業において、知識創発の支援効果を実験により検証する。

4. 研究成果

1) 操作ログの保存と再現機能を実装した。操作ログの保存として、データ分析環境をユーザが起動してから終了するまでの操作をログとして保存する機能を実装した。あるマイニングツールがセットされてから他のツールがセットされるまでの、ひとつの分析操作を表す操作ログとして、マウスやキーボード操作を対象として、各操作情報を記録する機能を実装した。

操作ログの再現として、ひとつの操作ログを入力として与え、データ分析環境上で、操作を再現する機能を実装した。

2) ログデータ共有環境として、保存されたデータ分析環境の操作ログを、他のユーザと共有するため、任意にアップロード、ダウンロードすることが可能なサーバの構築を行った。

操作ログのアップロードとして、ユーザがログデータ共有のためのサーバにアクセスした際に、操作ログをアップロードするためのサーバ上の機能を実装した。また、ユーザがログデータ共有のためのサーバにアクセスした際に、アップロードされた操作ログの一覧を表示し、その中から参照したい操作ログを選択してダウンロードできる機能を実装した。

ログデータ共有環境の構築においては、より多くのユーザに継続的に利用してもらうための機能を実装した。具体的には、サーバへのアクセス回数や、操作履歴のアップロード回数に応じて、サーバ上で利用可能な機能を解放する、またユーザを褒める機能を実装した。実験の結果、これらの機能が、ユーザの利用を促す効果があることが確認された。

3) データ分析プロセスになぞらえたTETDM上で実施するチュートリアル課題、ならびにツール選択支援機能を用意した。すなわち、テキスト分析における重要な単語として、テキスト全体やテキストを条件で絞り込んだ部分テキストの高頻度の単語に着目させる課題を用意した。またこれらの単語が出力された原因を探ってもらう課題を用意した。これらの課題をTETDM上に実装し、課題ごとの操作ログを収集できる環境を構築した。

有効な操作履歴を推薦する際に、ユーザ自身がある程度適切なツールを初めに選択しないと、データベース中のどの履歴でも良い履歴として推薦されてしまうことがわかった。そこで、分析の目的に応じて妥当性のあるツールを選択しやすくするインタフェースを構築した。すなわち、ユーザが分析したい単位(単語、文書、文)と分析内容(分析単位が単語の時は、重要語の抽出や共起単語の抽出など)を選択すると、自動的にツールがセットされる。

4) 収集したログデータからスキル獲得を支援する操作履歴の推薦を行い、スキル獲得に繋がる効果があるかを実験的に検証した。まず推薦の方法として、ユーザの操作履歴とデータベース中にある操作ログ集合とを比較し、ユーザが行っていない操作を含む操作ログ、ならびに分析に必要な操作を積極的に行っている操作ログに高い評価を与え、推薦することとした。この推薦の結果、ユーザは推薦された操作ログに含まれるユーザにとって新しい操作を真似する傾向があったこと、および操作ログの参照によって新たな考察結果が得られる傾向があったことを確認した。

5. 主な発表論文等

〔雑誌論文〕(計 6 件)

[1]砂山渡・石田純太・川本佳代・西原陽子：類推による説明スキルの獲得支援システム，情報処理学会論文誌，査読有，Vol.59, No.10, pp.1922 -- 1931, (2018).

[2]Wataru Sunayama, Tomoya Matsumoto, Yuji Hatanaka, and Kazunori Ogohara: Data Mining Framework for Treating both Numerical and Text Data, International Journal of Service and Knowledge Management, 査読有, Vol.2, No.1, pp.1 -- 18, (2018).

[3]Wataru Sunayama, Yuki Shibata, and Yoko Nishihara: Topic Recommendation Method related to a Present Topic for Continuing a Conversation, Information Engineering Express, 査読有, Vol.3, No.1, pp.19 -- 28, (2017).

[4]砂山渡・高橋麻祐・川本佳代：ストーリー提供機能とライフ機能を用いた学習意欲の向上と維持の枠組み，日本知能情報ファジィ学会誌，査読有，Vol.29, No.3, pp.501 -- 509, (2017).

[5]砂山渡・竹岡駿・西村和則：テキストマイニングのための統合環境 TETDM の利用意欲向上のためのゲームモードの開発，日本知能情報ファジィ学会誌，査読有，Vol.29, No.2, pp.558 -- 566, (2017).

[6]砂山渡・高間康史・徳永秀和・串間宗夫・西村和則・松下光範・北村侑也：統合環境 TETDM を用いた社会実践，人工知能学会論文誌，査読有，Vol.32, No.1, NFC-A, pp.1 -- 12, (2017).

〔学会発表〕(計 19 件)

[1]中江剛士, 砂山渡, 畑中裕司, 小郷原一智:新しい分析操作を含む操作履歴の提示によるテキストマイニングスキルの伝達支援,第21回人工知能学会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資料, pp.96 -- 101, (2019).

[2]Ryuichi Omi, Yoko Nishihara, and Ryosuke Yamanishi: Extraction of Paraphrases using Time Series Deep Learning Method, International MultiConference of Engineers and Computer Scientists, (2019)

[3]中川拓郎・砂山渡・畑中裕司・小郷原一智:着目点の明示によるデータ分析支援,第18回人工知能学会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資料, pp.34 -- 39, (2018).

[4]中江剛士・砂山渡・畑中裕司・小郷原一智:有効な操作履歴の評価によるテキストマイニングスキルの伝達支援,第31回人工知能学会全国大会, 2M2-0S-34a-5, (2017).

[5]松本友哉・砂山渡・畑中裕司・小郷原一智:データマイニングとテキストマイニングの連携によるデータ分析支援,第15回人工知能学会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資料, pp.14 -- 19, (2017).

他 14 件

〔図書〕(計 0 件)

〔産業財産権〕
出願状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
出願年:
国内外の別:

取得状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
取得年:
国内外の別:

〔その他〕

<https://tetdm.jp>

6 . 研究組織

(1)研究分担者

研究分担者氏名:西原陽子

ローマ字氏名:(NISHIHARA, Yoko)

所属研究機関名:立命館大学

部局名:情報理工学部

職名:准教授

研究者番号(8桁):70512101

(2)研究協力者

研究協力者氏名:

ローマ字氏名：

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。