

学術研究実績報告書

申請書との変更点およびその理由(内容、日程、実施場所、参加者等で変更があれば記入)

コロナ禍による制約のために研究期間を1年延長した

研究実績概要

研究代表者(申請者氏名・所属機関・職名):石川 博・東京都立大学システムデザイン研究科・教授
 共同研究者(氏名・所属機関・職名):遠藤雅樹・職業能力開発総合大学校・情報通信ユニット・准教授
 研究課題名:ビッグデータ時代に求められる統合的仮説生成方法論の体系化
 研究期間:2019年 4月 1日 ~ 2021年 3月 31日

概要:(1,000字以内で記述)

(A)ソーシャルビッグデータの統合仮説生成の方法論とその基盤となる理論・モデルの基本検討.

(A1)統合的データモデルの提案: 一般にデータモデルは基本となるデータ構造とそれに対する操作からなる. ソーシャルビッグデータ応用はデータ構造として, 集合の集まりを基本とするデータマイニングと集合を基本とするデータ管理とからなるハイブリッドなエコシステムである. そこで両者を統一的に表現するデータ構造として集合族(集合の集まり)を導入した. 次に集合族を入出力とし, データマイニング操作と集合族固有操作(縮退形であるデータ集合操作を含む)からなる基本操作を提案した. 本研究ではソーシャルビッグデータ応用における仮説生成手順全体の記述のためにハイブリッドなエコシステムを纏めて記述・説明でき, プログラミング言語より抽象度の高いデータモデルを提案した.

(A2)仮説生成方法論の構築: このデータモデルアプローチに従って, 複数データソースを統合的に利用して仮説を生み出す方法論として以下のような一般化差分法を提案した. 本研究では複数の分野(数学, 経済, 科学, 医学, 画像処理, 機械学習等)で独立に発展してきた差分方式の体系化を目指し, IT による具体化を行う. 提案する差分法はどのような空間の差分を対象にするかで以下のように分類できる. それぞれに具体的応用を付する.

- (ケース1)時系列データの異なる時点におけるデータの差分
 - ・ 事象の変化(違い)を判断する.
- (ケース2)同一の空間データの異なる時点における差分
 - ・ 動体や変異の検出をする.
- (ケース3)実空間ではなく概念空間における概念データを対象とした差分
 - ・ 概念のコアな意味を見つける.
 - ・ 未知と既知の概念間の差分を表現し, 表現された差分で未知の概念の理解を助ける.
- (ケース4)仮説間における差分
 - ・ 複数の仮説間の差分から新しい仮説を得る.

(B)先進的ユースケースの収集及びそれを対象にした実現可能性の評価実験の実施.

JAXA, 自動車部品メーカーなどと連携して, ソーシャルビッグデータへの統合的仮説生成方式を実践することで有効性を確認した.

(A) (B)の研究成果は以下のような国際論文誌と国際会議の論文にまとめられた.

- Hiroshi Ishikawa, Yukio Yamamoto, Social Big Data: Concepts and Theory, Springer Nature Trans. Large Scale Data Knowl. Centered Syst. 47: 51-79 (2021)
- Hiroshi Ishikawa, Yasushi Miyata, Social Big Data: Case Studies, Springer Nature Trans. Large Scale Data Knowl. Centered Syst. 47: 80-111 (2021)
- Yukio Yamamoto, Hiroshi Ishikawa, Data Management in Japanese Planetary Explorations for Big Data Era. ACM WIMS 2020: 88-90
- Munenori Takahashi, Masaki Endo, Shigeyoshi Ohno, Masaharu Hirota, and Hiroshi Ishikawa, Automatic detection of tourist spots and best-time estimation using social network services, IWIN 2020