

平成 30 年 5 月 11 日現在

機関番号：55501

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K03639

研究課題名(和文) 中小企業のためのビッグデータ可視化システムの研究開発

研究課題名(英文) Research and development of big data visualization system for SMEs

研究代表者

内田 保雄 (Uchida, Yasuo)

宇部工業高等専門学校・経営情報学科・教授

研究者番号：70321487

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：中小企業において高い専門性を持たないユーザでもビッグデータの分析・活用を容易に行え、操作性に優れたビッグデータ可視化システムのプロトタイプを開発した。ベースシステムとして、Cytoscapeというオープンソースのグラフデータ統合/解析/可視化プラットフォームを用いた。操作の容易性および利便性を重視して設計し、ユーザとCytoscapeをつなぐ独自のGUIインターフェースを構築した。そして、開発したプロトタイプを中小企業コンサルタントに試用してもらい中小企業への適合性の評価を得た。以上により、中小企業ユーザにおいても手軽に自社の保有する大量のデータを可視化して分析を行うことを可能にした。

研究成果の概要(英文)：Even users who do not have high expertise in small and medium enterprises can easily analyze and utilize big data and developed prototype of big data visualization system with excellent operability. As the base system, we used Cytoscape, an open source graph data integration / analysis / visualization platform. Designed with emphasis on ease of operation and convenience, we built a unique GUI interface that connects users and Cytoscape. And we got evaluation of conformity to small and medium enterprises by having trial of small prototype consultant for developed prototype. Through the above, it has made it possible for users of small and medium enterprises to easily visualize and analyze large amounts of data owned by their own company.

研究分野：情報システム

キーワード：ビッグデータ 中小企業 可視化

1. 研究開始当初の背景

ビッグデータの定義にはいろいろあるが、著名なものとしてはマッキンゼーの論文が挙げられる。この論文では、おもに量的側面からとらえた定義となっている。そして、7つのキーとなる洞察をおこなっている。また、シンボリックな定義としては META グループ(現ガートナー社)の研究報告で使われている「3V」モデルがある。すなわち、ボリューム(volume、データ量)、速度(velocity、入出力データの速度)、バラエティ(variety、データタイプとデータ源の範囲)の視点から定義されている。

しかしながら、最近ではさまざまな事例や応用の展開とともにその概念や定義も変貌しつつある(図1)。そのため、欧米はもちろん我が国においても、急展開で学術的研究から実務的応用までの広範囲にしのぎを削る一大潮流となっている。

前述のガートナー社は、すでに昨年のリサーチ分析レポートにおいて、「2013年の展望: ビッグデータと情報インフラストラクチャ」と題して分析をおこなっている。そのおもな所見は次のとおりである。

- ビッグデータを活用する企業はそうでない同業他社に比べ、既に優れた経済的な成果を得ている。
- ビッグデータのユースケースをサポートするスキルを獲得できるかどうか、既に問題になっている。この状況は短期的には改善されない。
- 適切な戦略や組織モデルの必要性がより重要になっている。
- 30%強が、論理データウェアハウスの実装の進め方を理解するための支援を求めている。

一方我が国では、総務省の「平成24年版情報通信白書」において、「特集 ICT が導く震災復興・日本再生の道筋」の中で、ビッグデータの活用が謳われている。しかしながら、日本情報システム・ユーザー協会による「企業IT動向調査報告書2014」によれば、「1割に満たないビッグデータ活用企業」と分析されており、活用は大企業を中心に進むとみられている。

このような状況下であって、中小企業においても勝ち残りのためにビッグデータの活用が課題となってきた。そして、中小企業向けの研究開発や技術支援が喫緊の課題となっている。しかしながら、中小企業におけるビッグデータ活用のための方策や今後の方向性に関する学術的な研究は、ほとんどおこなわれていないのが現状である。

2. 研究の目的

われわれは、研究課題「中小企業の情報基盤を強化するクラウドシステムの研究開発」〔基盤研究(C)〕(2012年度~2014年度)に取り組んできており、中小企業の情報基盤を構築・強化するための情報システム・モデ

ルとして、中小企業に適合的なクラウドシステムのモデルを提案し、プロトタイプを開発するとともに利活用時における問題点についても考究してきた。そこで、次なる研究ステージとして、本研究ではそこで得られた知見を生かして、中小企業に適合的なビッグデータ可視化システムのフレームワークを検討・設計するとともに、中小企業における戦略性のある情報基盤のありかたや運用・管理時の問題点を実証的に解明することを目指すものである。

3. 研究の方法

研究の中心となるビッグデータの可視化については、さまざまな手法が考えられるが、本研究では4次元の可視化を追求する(図2)。「4次元」の考え方については、我々が別の理論的研究において扱ってきた多次元の入力のオートマトンの理論を応用する。従来の可視化手法の多くは、2次元あるいは3次元の可視化であるが、本研究ではさらに1次元を加えて4次元とする。4次元目の軸はアニメーションとして実現する。このことにより、従来の静的な可視化では見えなかった現象の分析も可能になると期待できる。なお、ユーザーインターフェースはシンプルにして使いやすいツールとしての開発を目指すものである。

本研究は、大きく4つのフェーズに分けられる。まず、先行研究の調査として、文献研究をおこなう。ここでは、大企業の場合と中小企業の場合との差異の分析に重点を置きながら考察する。つぎに、先進事例の実地調査による事例研究をおこなう。ここでは、実際の運用・管理における問題点や課題を中心に調査する。つづいて、システムの設計・開発に取り組む。ここでは、中小企業に適合的なシステムのモデルを考案し、プロトタイプとして実装する。さらに、最終フェーズとして、実証システムによる運用と評価をおこない、知見としてまとめる。

まず、ビッグデータの企業への適用および中小企業における情報基盤に関するこれまでの研究蓄積、および最新の動向を体系的に整理する。この作業は、文献研究をおこなう、先進的な取り組みをおこなっている中小企業の事例を調査する、という方法により進める。については、ビッグデータおよび中小企業における情報基盤に関する主要な学術文献のサーベイとレビューをおこなう。とくにビッグデータの利用を強力に推進している欧米の文献のサーベイは不可欠である。そして、欧米での先進的な事例が日本に応用可能かどうかについて検討する。また

については、ビッグデータの導入を積極的に推進している中小企業あるいは中小企業の情報化を支援している公共機関を実際に視察し、事例研究をおこなう。そこでは、さまざまなビッグデータ可視化システム構築の事例から、その効果や問題点を考察する。

これらの調査研究は研究代表者を中心としておこなうが、研究分担者もその一翼を担い、情報収集やデータ分析作業をおこなう。なお、研究が計画どおりに進まないときは、中間成果発表を圧縮して情報収集と分析に注力することで対応する。

続いて、前年度の研究調査およびその分析と考察にもとづいて、ビッグデータ可視化システムの企画・設計の作業に入る。このフェーズでは、研究代表者と研究分担者でシステムの仕様について分析・検討を重ねることにより、仕様を確定していく。なお、研究が計画どおりに進まないときは、中間成果発表を最小限に圧縮して研究補助者のマンパワーを増やすことで対応する。

さらに、開発したシステムの実証実験をおこない、また評価もおこなう。

本システムは、オープンソースの基盤技術をベースとして構築する。具体的には、大規模データの分散処理を支える Java ソフトウェアフレームワークであるオープンソースの Apache Hadoop を導入する。そして、ワークロード分散システム (Hadoop MapReduce)、キーバリューストア (HBase)、分散ファイルシステム (Hadoop Distributed File System) により構成する。次に、オープンソースのグラフ可視化プラットフォームである Gephi をカスタマイズして設計・構築する。これにより、中小企業においても操作や分析が容易なシステムを実現することが可能となる。

また、特に重点をおく研究ミッションとしては、「ビッグデータの活用の在り方について」でも指摘されている、ビッグデータの活用に関する研究開発の動向の日米比較において、研究開発投資額の圧倒的な違いと活用者の裾野の広がりをかんがみると、我が国における中小企業での活用にかかる現状の問題点分析と今後の普及支援のための研究開発が挙げられる。

4. 研究成果

本研究の作業は大きく分けて 1. Cytoscape の制御のためのプログラム構築, 2. GUI インターフェースの構築に分かれる。ユーザの行う操作をより容易にするため、構築する GUI のウィジェットはファイル読み込み指定とグラフ表示指定の極めてシンプルなものとする。本研究ではユーザが使用するデータを自社のデータベースに蓄積されたデータと想定してシステムを構築する。

上記研究作業について、それぞれ構築・実験を行った。

[実験 1]

実験 1 では研究作業の 1. Cytoscape の制御のためのプログラム構築について、ふるさと納税受入件数・受入額に関するデータを用いて実験を行う。

[実験データ]

A) 都道府県隣接リスト(2010 年 海上隣接を含む)

データ数: 47 都道府県

B) ふるさと納税に関する現況調査結果 (正)平成 28 年 6 月 14 日 18 時 30 分掲載 各自治体のふるさと納税受入額及び受入件数 (平成 20 年度~平成 27 年)

データ数: 1788 市区町村

C) 平成 22 年 都道府県・市区町村別統計表(一覧表)

データ数: 1969 都道府県市区町村

[データ加工]

Cytoscape への入力には以下の 2 種類のデータが存在する。

・グラフデータ: ノード同士のつながりを表す

・テーブルデータ: 各ノードの属性を表す

本実験では実験データのうち, A をグラフデータ, B と C をテーブルデータとした。グラフデータの作成に当たっては Python ライブラリの NetworkX を用いる。テーブルデータに関しては 1 列目に主キーを持つてくる必要がある。

[スタイル]

Cytoscape による可視化の際のスタイルを以下のように設定した。なお, label Font Size と Size に関しては受入件数と受入額の場合でそれぞれ実験を行った。

Full Color

全国, 都道府県 : #666666

区 : #99FF99

政令市 (特別区部を含む) : #66CCFF

市 : #FF9999

町, 村 : #FFFF99

Shape

全国, 都道府県 : V

他 : Ellipse

実験結果

実験結果は以下の通りである。プログラミングによる可視化が来ていることがわかる (図 1、図 2)。

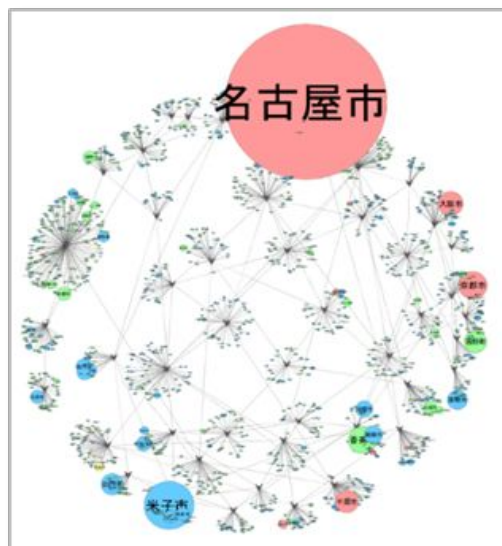


図 1 受入件数に関する可視化結果

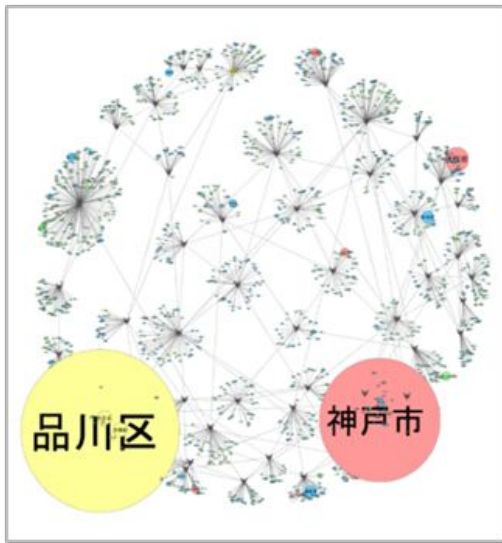


図2 受入額に関する可視化結果

〔実験2〕

実験2では研究作業の2.GUI インターフェースの構築について、実験を行う。操作は“ファイルを開く”と“上書き保存”、“名前を付けて保存”の非常にシンプルなものとする。ファイルを開くと、データを読み込み Cytoscape にデータが送られ、可視化結果が返ってきて画像として表示されるという仕組みとなる。

〔アクション〕

・開く

ファイルダイアログを開く
 選択したデータファイルを読み込む
 可視化結果が表示される

・保存

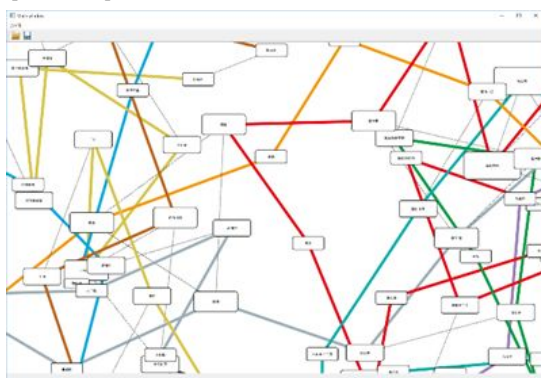
- 上書き保存

保存ファイルと同じ名前のファイルがあれば上書き保存、なければ名前を付けて保存する

- 名前を付けて保存

ファイルダイアログを開き、名前を付けて保存する

〔完成図〕



5. 主な発表論文等

〔雑誌論文〕(計4件)

Yasuo Uchida, Yui Okamoto, Seigo Matsuno & Makoto Sakamoto, A STUDY OF

VISUALIZATION SYSTEM FOR SMES IN JAPAN, Proceedings of The 3rd CHREST International Conference, pp.1-6, 2017

Yasuo Uchida, Miao Xinyun, Seigo Matsuno, Yasushi Iha and Makoto Sakamoto, Preliminary Study on a System for Visualization of Big Data in SMEs, Proceedings of Engineering and Technology Innovation, vol.7, pp.20-24, 2017

Yasuo Uchida, Miao Xinyun, Seigo Matsuno, Yasushi Iha and Makoto Sakamoto, Technical Review of a System for Visualization of Big Data for SMEs, 2016 5th EEM International Conference on Public Administration & Management (EEM-PAM 2016), Advances in Education Research, Volume 91, pp.35-39, December 20-21, 2016

Yasuo Uchida, Seigo Matsuno, Yasushi Iha, Makoto Sakamoto, Comparison of tools for visualization of big data in SMEs, Journal of Scientific Research and Development, Volume 3, Special Issue 1, pp.97-103, 2016

〔学会発表〕(計1件)

Yasuo Uchida, Seigo Matsuno, Yasushi Iha, Makoto Sakamoto, "Comparison of tools for visualization of big data in SMEs", The 2016 3rd International Postgraduate Conference on Business Administration and Management (IPCBAAM 2016), 2016

6. 研究組織

(1)研究代表者

内田 保雄 (UCHIDA, Yasuo)
 宇部工業高等専門学校・経営情報学科・教授

研究者番号：70321487

(2)研究分担者

松野 成悟 (MATSUNO, Seigo)
 宇部工業高等専門学校・経営情報学科・教授

研究者番号：30290795

(3) 研究分担者

坂本 真人 (SAKAMOTO, Makoto)
 宮崎大学・工学部・准教授

研究者番号：50196101