

ビッグデータ分析と活用による地域活性化教育

Experimental Study on Big Data Analytics for Regional business Development

河合博子^{*1}, 竹上 健^{*1}, 河合岳夫^{*2}, 根本勇^{*3}
 Hiroko KAWAI^{*1}, Takeshi TAKEGAMI^{*1}, Takeo KAWAI^{*2}, Isamu NEMOTO^{*3}

^{*1}高崎商科大学

^{*1}Takasaki University of Commerce

世田谷区役所

^{*2}City of Setagaya

フェーズプランニング

^{*3}PhasePlanning inc

Email: h-kawai@uv.tuc.ac.jp

あらまし：現在、ビッグデータ分析と活用を通じて大学の教員と学生および企業や行政が連携して農産物の販路拡大、観光地集客、高齢者の健康増進・医療連携、中心市街地の活性化課題等の改善案を創出する教育展開を検討している。本稿では、ビッグデータ分析と活用の基本的な学習手順と教育環境システム的环境について述べる。

キーワード：ビッグデータ活用分析，地域活性化課題，Excel，Hadoop，NoSQL

1. はじめに

ビッグデータ分析活用は、その情報処理基盤やクラウドの利用環境の普及、ベンダーやSIによる大学の講義支援(2012のEMCジャパンと立教大学、2013のNECと会津大学の例等)によって、その大学における実践的な授業展開が可能となりつつある。

そのようなビッグデータ分析活用で解を見出した課題の例として、農産物の販路拡大、高齢者の健康増進・医療連携、中心市街地の活性化、観光地集客などが挙げられる。

大学が地域の企業や行政とともに近未来を真剣に考え、既存の価値観や枠組みを超えて想定外の領域でビジネスイノベーションを創造することが求められている課題である。

本稿では、商学部の情報系学生を対象とするビッグデータの分析活用の基本的な学習手順と現在準備中のICT教育システム環境について述べる。

2. ビッグデータで得られる知見や効果

ビッグデータとは、議論は百出で明確な定義は定まっていないが、概して「事業に役立つ知見を導出するため」のデータであり、典型的なデータベースシステムでは格納・検索・共有・分析・可視化が困難なデータで、非構造なものを含む、多様、容量が膨大、高解像度、生成処理と処理速度の特徴を持つデータのことをいう。

ビッグデータ分析の知見には、既存の事例から、(1)現状や実績の把握、(2)変化の検出、(3)原因の仮説、(4)未来の予測の4つに集約できる。

(1)現状や実績の把握とは、人口分布や交通状況、過去5年の株取引内容、現在の電気使用量などの情報から「何が起きたか」を明らかにするものである。

(2)変化の検出とは、つぶやき数の上昇や下降傾向、Webサイトのアクセス速度の変化、脈拍や血圧の変

表1 ビッグデータ活用の効果

ビッグデータ活用の効果
・顧客満足度の向上(顧客嗜好分析・マーケティング・商品リコメンド・商品開発)
・リソース利用の最適化(交通・電気等利用状況把握等)
・異常の早期検出(ログや計測値等の情報分析し、システムや人体の異常の早期発見)
・リアルタイム情報提供(逐次発生する情報分析と提供)
・セキュリティ向上(防犯カメラ等の動画や画像分析・犯罪予防)

表2 ビッグデータ目的設定からシナリオ検証

例)観光集客
1)目的設定
・あるべき姿(観光客を増やしたい)
・現状(観光客が増えていない)
・改善ポイント(集客の良いコースを企画・リコメンドの改善・Webページの改善)
目的・観光コースをより正確に把握し、集客に良いコースを企画する。
・顧客とコースの相関を分析し、顧客の嗜好を導出する。
2)活用シナリオの立案
目的を実現するための活用シナリオを検討。
・利用する分析手法を検討するために各手法の種類と特徴を理解する。
3)活用シナリオの検証
・トライアルを実施し、効果や課題を洗い出す。
・生成・転送・格納・整理・分析の各段階で発生する問題点を列挙する。
・実行の場合はシステムと業務内容を検討する。
・非実現の場合は、シナリオの見直しを勧告する。
4)ビッグデータ活用が実現可能かどうかの検証
システムの導入(Hadoop,NoSQL,データ分析ツール,アプライアンス,CEPなどビッグデータに関連した技術の選択)
①番組み型(Hadoop(HDFS),NoSQL,RDBを利用)
②リアルタイム型(CEPを利用)

化の検出等「何が変わったか」を可視化するものである。(3)原因の仮説とは、今まで気づかなかった原因と結果の関係について「なぜ起きたのか」を精度の高い知見を導出するものである。(4)未来の予測では、実績値をもとに、今後起きうる現象を予測する。例えば、つぶやき、取引状況、経済状況から株価を予測するなどがこれにあたる。未来予測ではより高度なデータ分析が必要なため、データ分析の専門家や高度な分析ツールの利用も必要となる。

このビッグデータ活用の効果については、表1ビッグデータ活用の効果に示すように顧客満足度の向

上など5点に集約することができる。

3. ビッグデータ分析活用の学習プロセス

ビッグデータの分析活用には、情報処理基盤や解析手法についての知見が必要であるが、それ以前に、適用する業界や事業を知り、何よりもデータ解析をする目的を把握した上で仮説を立て、データから導いた最適化プロセスを実務に落とせるようにしなければならない。

表2は、ビッグデータの分析の目的設定からシナリオ検証までのプロセスを観光地集客の例で示している。検証で用いられるツールはシステム環境により選択が必要となる。表3では、データの生成から分析までの学習すべき技術を示している。表4には、各分析手法とその特徴を示している。表のクラスター分析の前(検定)まではいわゆる EXCEL の関数などを用いた統計分析である。ビッグデータの分析ではクラスター分析以降の高度な分析手法が必要となる。

4. ビッグデータシステム環境

表5に、使用するビッグデータ活用を支える基本技術を示す。今回の教育環境は、NoSQLにはHbase, MongoDBを使い、クラウドデータベーススペースで利用する。また図1に、ビッグデータ処理と基盤技術の関係を示す。

NoSQL のシステム構築には、Hbase, MongoDB等を研究者アカウントのあるデータセンターのサーバの仮想マシン上に数台を用意する。学生はPC 端末から UltraVNC ソフトウェアを使いデータセンターのデータベースにリモート接続し、それらをクラウド環境で使用する。

5. まとめ

本稿では、大学・行政・企業が連携して地域活性化の課題解決に取り組むためのビッグデータ分析と活用手順、およびそのための教育環境システムについて述べた。そのシステムでは、(1)地域のビジネス課題をビッグデータの分析の活用の対象として取り組むために目的を明確にして再構成すること、(2)ビッグデータ分析、統計モデルの作成のために EXCEL の統計分析を学習し、適切な分析レポートをビジュアル化すること、(3)情報処理基盤とハードやソフトの長所と短所を理解することなどに重点を置くことなどが要求される。なお、この教育環境システムは、工学系でない学生が情報基盤やビッグデータの分析手順を理解しやすいようにできるだけ簡易な設計にしていきたいと考えている。

参考文献

- (1) 河合博子: “ビッグデータの功罪—ビッグデータが創造する豊かさと危うさ” 同志社商学第64号巻6号 太田進一教授古稀祝賀記念号,同志社商学会, pp.262-276 (2013)
- (2) 河合博子: “ビッグデータの功罪～イノベーションとセキュリティリスク～” 高崎商科大学叢書第7号商学

表3 データの生成から分析

5) データの生成(方法の選択)
既存データ利用が新規データの生成 新規の場合必要なデータを生成する仕組みを検討する(入力ICカード・スマホアプリによる自動収集・センサーによる自動生成・アプリによるログビデオ等の動画等)
6) データの転送(方法の選択)
(固定PC・移動端末・センサー)
7) データの蓄積(方法の選択)
リレーショナルデータベース(非構造化・スケールアウトに非適応) 多次元データベース(MDB)集計データが定まっているとき NoSQLデータベース(非構造化データ・スケールアウト向き・複雑な処理が実行できない)
Hadoop(プログラムの作成)
8) データ整理(分析できる形式にデータを整理)
データクレンジング(コード変換・表記変換・不正データの発見と修正・名寄せ) ・マスターデータ管理(MDM) ・非構造化データから構造化データへの変換(テキストマイニング・画像認識・音声認識)
9) データの分析
・データ検索, レポートツール ・データ検索, レポートツールOLAPツール ・データマイニングツール

表4 各分析手法とその特徴

種類	分析手法	表示・調査の特徴
定形レポート	ランキング	属性の並べ替え, 順位表示
	Zチャート	値・累積値・移動合計の表示 短期的・長期的なデータの傾向
	ABC分析	値の累積値データランク分け
非定形レポート	レーダチャート	データの大きさ, 中心からの距離
	ドリリング	データ間の階層情報データ特徴
	スライス & ダイス	集計項目/表示条件切替 データ特徴
アラート	分散分析	2つのデータ集合の違い 誤差/必然分析
	検定	2つのデータ集合の平均値/分散の違い
データ分類	クラスター分析	似たデータのグルーピング
	クラス分類	データが所属するカテゴリ予測
原因分析	相関分析	2項目間の関係
	アソシエーション分析	複数データの組合せの相関ルール 頻出品度の高いアイテムを抽出
予測	回帰分析	実績値から予測式を作成, 値予測
	時系列分析	傾向・循環性・季節変動 不規則変動の特徴・将来予測

表5 システム構成とビッグデータの基盤技術

10) システム構成
・システム構成 ・Webアプリケーション型 ・クラウド型
11) ビッグデータを支える技術
分散処理 NoSQL CEP クラウドコンピューティング インメモリ技術 アプライアンス

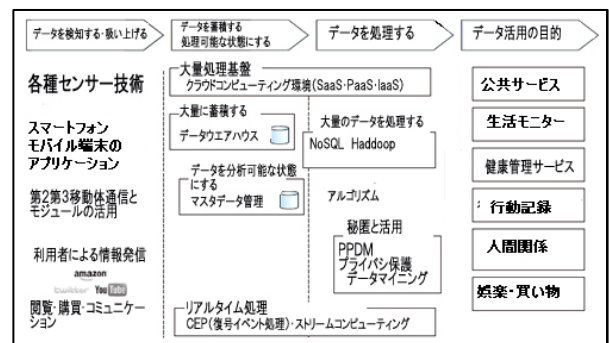


図1 ビッグデータ処理と基盤技術の関係
(出典: 河合 2012p103)

- 研究II, pp.101-112 (2012)
- (3) 竹上健: “表計算ソフトのVBAマクロの学部授業での教育実践” 高崎商科大学叢書第7号商学研究II, pp.122-133 (2012)