

[2014年度シンポジウム講演]

ビッグデータの現状と情報社会への影響 ～情報システム学の意義を踏まえて～

井上 健太郎

この記事は、第7回情報システム学会シンポジウム（2014.5.17）におけるシンポジウム講演の口述内容をまとめたものです。

1. はじめに

ビッグデータ関係の講演をするのは、実はもう 5～6 回目です。ユーザーの活用事例を主に報道する「情報ストラテジー」にいた経験が長いので、私はビッグデータに関してはユーザー寄りの視点で見えています。ですので、時にはベンダーの動きに対して少し斜に構える発言があつて、「おや？」と思うところがあるかもしれません。ユーザー側の視点でマネジメントにどう影響が及ぶのか、あるいはなぜマネジメントがうまくかみ合わないのかという視点で見ているので、そういう人間が話しているのだと思っただけだと幸いです。

まず会場の空気を知りたいので、皆さんに質問させていただきます。

1.1 「ビッグデータ」という言葉

「ビッグデータ」というのは、あなたにとってどんな言葉でしょうか。選択肢を四つ提示させていただきました。

「わくわくする」言葉だと思われる方はどの

Kentaro Inoue

日経 BP 社

コンピュータ・ネットワーク局イノベーション

ICT 研究所 兼 総合コンテンツ事業部

第7回情報システム学会シンポジウム

[シンポジウム講演] 2014年05月17日受付

© 情報システム学会

ぐらいいらっしゃいますか。3人ぐらいですね。

では、「IT 業界の仕掛けた販促用語」であり、ちょっと私は引いて見えていますという方はどのぐらいいらっしゃいますか。20人ぐらいでしょうか。

「技術進化の象徴」であり、あくまで技術的な話題として捉えているという方はどのぐらいいらっしゃいますか。結構多いですね。先ほどと同じぐらいです。

「その他」という方はいらっしゃいますか。ちなみに、どんなイメージでしょうか。

(フロア 1) パズワード。

(井上) 2番目に近い感じですね。大体2番目と3番目が半々ぐらいでした。

私は、ビッグデータブームは Hadoop などの OSS ツールの進化と、ネットサービス会社の隆盛を背景として生まれてきた活用用語であり、全く空虚な言葉ではないと思います。技術的な裏付けがあるのです。

ただ、ネット企業特有の業務を通じて発展してきたデータ活用プロセスが未消化のまま、一般企業に展開されている感じがあるのです。その未消化さが企業によっては何か響かなかつたり、空回りしている感じだったり、関係ないという感じだったりするのだと思います。

まず、そういうことがビッグデータに関するゆがみとしてあるということです。

2. 基礎知識

2.1 定義

昨年、私が JISA の情報サービス産業白書でビッグデータのビジネスの動向について書いたとき、ビッグデータの定義をこのような表にまとめました。

| | |
|------------|--|
| IDC ジャパン | 少なくとも以下の一つを満たすデータ <ul style="list-style-type: none"> ・100T バイト以上の規模で収集されるデータ ・ハイスピードストリーミングによって受けるデータ ・毎年 60%以上の成長率で生成されるデータ |
| 矢野経済研究所 | ①非常に大容量であるため処理や分析に特別な技術を要するデータ、②データベース、文書、音声、映像などを含む多様な種類のデータ、③極めて高頻度で発生し超高速に処理されるデータ、の三要素のいずれかを含むデータを総称する |
| ガートナー ジャパン | 高度な洞察、意思決定、プロセス自動化のために、コスト効果が高く革新的な形態の情報処理が求められる、大量、高速、多様性の特徴を1つ以上備える情報資産 |
| 野村総合研究所 | 特になし（調査を担当する研究員の担当分野に応じて異なる） |

IDC Japan では、大きいデータ、またはハイスピードストリーミング、成長率の高いデータのいずれかという定義をしています。矢野経済研究所では、大容量、または音声・映像等を含む多様なもの、または高頻度で発生し高速で処理されるデータという定義をしています。ガートナー ジャパンでも似たような定義をしています。野村総合研究所が、意外なことに調査員によって定義が違うということです。このぐらいビッグデータに関する定義は実は混沌としているということをお示しするためにまとめました。

「三つの V」(Velocity, Volume, Variety)

というのを聞かれたことがあると思います。高速性と大容量さと多様性です。ただ、実際にはビッグデータというタイトルがついた特集記事でもこれら三つの V とは別な要素を持った事例が強調されることもあります。データそのものがユニークだとか、分析に工夫があつて高度であるとか、分析結果の業務への落とし込み方に知恵があるとか、成果が明快であるといった要素です。スモールデータよりビッグデータの方が直ちに業績貢献度が高いと我々報道側も考えているわけではないので、ビッグデータ事例と、いわばスモールデータ事例が混在した特集記事も作られているというのが実情でしょう。

2.2 ブームの特徴は温度差

最近、トーマス・H・ダベンポートさんという有名人が『データ・アナリティクス 3.0』(邦訳版は日経 BP 社) という本を出しました。その本の中で、『ハーバード・ビジネス・レビュー』の読者 1000 人にアンケートしたところ、「活用している」が 28%、「自社にビッグデータに対する戦略がある」が 23%、「自社内への影響を検討済み」が 6%、「自社にビジネスへの役立て方を理解している」が 3.5% だったと書いてあります。ダベンポートさんが今のアメリカではデータ活用が非常に進んでいるかということ、意外とそうではないと正直に書いてくださっているのです。特に「自社内への影響を検討済み」がアメリカでもたった 6% しかありません。

ガートナー ジャパンが昨年 5 月ごろ発表したデータ (<http://www.gartner.co.jp/press/html/pr20130513-01.html>) は、2012 年 11 月時点のアンケートに基づいており、「よく知っている」というのは多少増えているのですが、「非常に関心がある」というのは全体の 2.5% しかなくて、「多少関心がある」を入れても実は結局 2 割もない。「全く関心がない」人が 2 割、バズワードという「IT 業界のはやり言葉として冷静に見ている」人が 6 割もいたということです。

2.3 特徴 1

ビッグデータは大ブームであるかのように言

われていますが、同時代的な盛り上がり方をしているようでしていない部分がありまして、雑誌のアンケート結果を、回答者の属性別に見ていると、けっこう「一般社員」「課長係長」「部長」といった属性ごとに、時期によって別々な動きをしています。ビッグデータという言葉は恐らくアメリカと同時代的に日本に輸入されたので、アーリーアダプターとレイトマジョリティの温度差が、時期によって比較的顕著に出たのではないかと思います。これもビッグデータの一つの特徴だと思います。

こうした温度差は、所属企業の会社の規模によっても生じています。例えば日本情報システム・ユーザー協会 (JUAS) から最近出た、「企業 IT 動向調査 2014」の冊子を見ますと、売上高 1 兆円以上の規模の会社では、母数は 38 件ですが、約 7 割は前向きに捉えているとあり、それ以下の規模の企業ではがくんとこの割合が下がります。

2.4 特徴 2

こうした温度差が背景にあるのだと思いますが、「IT 部門＝抵抗勢力説」というのが、最近やたら強化されてしまった感があります。

JUAS の調査結果によると、ビッグデータに対する捉え方を問うた設問において、「前向きに捉える」IT 部門が約 2 割です。「しばらくは静観」が約 5 割、「懐疑的に捉えている」が 7%、「分からない」が 2 割ということで、そんなに前向きではない。ただ、同じ調査で IT 部門に自社の経営者の意識を尋ねたところ、「わからない」という答えが最多ではあるのですが、経営者は自分たちよりもさらに前向きではないと考えている IT 部門も多いようです。

これも「日経コンピュータ」の記事です (<http://itpro.nikkeibp.co.jp/article/COLUMN/20140221/538286/>)。IT 部門の自己評価では、抵抗勢力だと自覚しているのは約 4% しかいないのですが、利用部門にアンケートを取ると「うちの IT 部門は抵抗勢力だ」という人が 14.8% いて、10 ポイントもギャップがあるのです。

最近、ベンダー等に私が取材しているときも

「ユーザーの IT 部門はビッグデータ活用推進の抵抗勢力」というコメントが出るのが非常に増えている印象があります。ちなみにこうした見方は、ビッグデータ周辺だけではなく、モバイル端末の活用推進の話題においても時々出ます。

2.5 特徴 3

もっとも、ビッグデータの活用推進に慎重な人が多いのにもそれなりの理由があります。おそらく主な要因は、具体的な活用情報、事例情報が乏しいことでしょう。

JUAS の調査で「ビッグデータ活用における課題」を問うた結果でも、「導入する目的の明確化」が最多であり、活用目的がよく分からないというユーザーが多いのです。活用すべき理由をユーザーに納得させる事例情報が全然足りないということだと思います。

先ほど、ネット企業のデータ活用ぶりを見て旗を振っても、一般企業がなかなか咀嚼できない、ということをちらっと申し上げました。JUAS の調査を見る限り、一般企業ではビッグデータとは何かについて、「顧客データ」「取引データ」という回答が圧倒的に多く、「Web アクセスのログ」は意外と少ないのです。ネット企業がソーシャルゲームなどの新サービスを立ち上げてログデータをどんどん取得するのはまったく違ったスタート地点から、一般企業では「ビッグデータ活用」の検討が始まっていることがわかります。

その一方で、経営者が業績貢献を語ってくれる事例を、IT 業界はずっと探し続けています。

その一つはコマツでした。多分皆さんご存じだと思いますが、コマツには KOMTRAX というシステムがありまして、ショベルカー等に GPS やエンジンの稼働状態を知らせるセンサーが付いており、無線でコマツに吸い上げています。その稼働状況を吸い上げて、メンテナンスサービスの効率化、あるいは販売戦略に生かしているということです。KOMTRAX のおかげで中国の景気減速を未然に検知して対処できたという成果報告も、確か去年の決算短信に

は入っていました。

ところが今年1月に日経新聞に「曇った千里眼」という記事が出ました。インドネシアの景気に関しては、KOMTRAXの神通力が通用しなかったという内容です。その記事によると、KOMTRAXだけでは市場の先行きを読むにはまだまだデータ不足で、インドネシアの為替レートが想定外に変化したので、見通しが狂ったということのようです。ビッグデータ活用の成功事例はなかなか長続きしていない状況です。

その一方でソフトバンクモバイルも昨年からはビッグデータ活用を力強く語り始めました。スマホアプリの利用履歴を分析して、どこで電波が混雑しているか、アンテナが不足しているかをビッグデータで分析して、効率的にアンテナを設置できているので、顧客サービスの満足度向上が比較的効率的にできますということを、孫正義社長は言っています。

このように経営者が力強くビッグデータ活用を推してくれる事例がもっといろいろな業種から出てくるといいのですが、実際は少ないので、IT業界も啓蒙に悩んでいるという状況だと思います。

3. Google Trend で見る流行度合い

ネット上に「Googleトレンド」というツールがあります。例えばGoogleトレンドのトップページで「情報システム学会」を検索すると、折れ線グラフが出てきます。Googleサーチでこのキーワードが検索された回数を時系列で表示しています。

こういうものを使って、私は流行をよく見るのです。

3.1 「ビッグデータ」関連 IT ツールの販促をめぐる地殻変動

ビッグデータと、その関連用語を幾つか検索してみました。「ビッグデータ」はアルファベットの「bigdata」と、片仮名の「ビッグデータ」と合わせた結果です。それから、「データサイエンティスト」「Hadoop」。Hadoopというのは、ビッグデータの処理でよく使われる分散バッチ

処理ソフトです。それから、よく最近話題になる「R言語」、フリーウェアのデータ分析言語です。

「bigdata + ビッグデータ」は、大体2011年の初めぐらいから徐々に盛り上がり、2012年で一定のレベルに達して、2013年にまた一段盛り上がったのですが、今のところ2013年夏以降は伸び悩んでいます。

「Hadoop」は昨年あたりから下り坂です。

「データサイエンティスト」は、昨年の夏がピークで、これも直近は下り坂となっています。

「R言語」は、あまり流行と関係なく一定数の関心があるようです。

アメリカでの推移はこうではないのです。アメリカに限定して調べると、「Bigdata」と「Hadoop」はひたすら右肩上がり、勢いは全く鈍っていません。

イギリスも調べてみましたが、大体同じような感じ。「Hadoop」「Bigdata」「Data Scientist」いずれも右肩上がりです。ドイツは、私はドイツ語が分からないので調べていません。

このような感じで、どうも日本だけが最近では盛り上がり切っていない。特に「Hadoop」への関心が下がってしまっているのは、日本のIT業界の一つの問題点を暗示していると思います。最近、IDCジャパンが、OSSを積極的に活用している会社の方が売上高は増加傾向だという調査結果を出していました (<http://www.idcjapan.co.jp/Press/Current/20140108Apr.html>)。ネット企業が数台規模のクラスタリングでデータを蓄積し、活用法を見いだしながら数百~数千ノードのビッグデータ情報基盤を構築するうえで、HadoopのようなOSSは非常に大きな役割を果たして来ています。一般企業がHadoopに関心がないとすれば、このこともビッグデータブームの足をひそかに引っ張っている可能性があるかと私は考察します。

先ほど、「IT部門=抵抗勢力説」について触れましたが、一部のITベンダーは既に、マーケティング政策を変え始めました。IT部門を介さずに利用部門に直接リーチするために、Sierではなく、別な業態の会社を販促パートナーとして重視する動きが出ています。

例えば、日本マイクロソフトがエクセルで「power BI」というツールを出して、「ビッグデータを民主化する」というキャッチフレーズで販促を始めているのですが、この販促のパートナーとなったのは、ソフトバンクグループのマーケティング支援会社である Agoop、ソーシャルメディア分析サービスのジー・サーチ、モバイル系のユーザーの行動を分析する仕組みを持つビデオリサーチインタラクティブなどです。今のところ提携相手はこうしたマーケティング支援のベンチャーで、SIer は名前が挙がっていません。MS の関係者に意図を確認したところ、マーケティングの強化に関心の高い利用部門に直接リーチしたいので、こうした新たなパートナーを開拓しているとのことでした。

マイクロソフトと BI ツールで競合するウィングアークも、最近ではブレインパッドなどのマーケティング系の会社と協力して営業したりセミナーを開催したりしています。

ほかにもここ 2~3 年で、電通や博報堂と大手 IT ベンダーが協業するなどの動きもあり、販促パートナーとしての SIer の存在感がビッグデータの分野では、やや小さくなっているように感じられます。

このような、いわば「IT 部門飛ばし」や「SIer 外し」がアメリカではどうなのかと思い、昨日まで調べていたのですが、やはり多少はありそうです。最近出た「ハーバード・ビジネス・レビュー」5 月号の「アナリティクス競争元年」という特集の中に、「データの品質管理は IT 部門に任せるな」という論文が載っていました。要は IT 部門にデータ品質の問題を担当させても、うまくいったケースはないということです。データ品質というのは、作成した瞬間に決まるものであって、データ作成における業務プロセスを改めさせる権限が IT 部門にない以上は、任せても仕方がないということを言っています。

最近、CIO(最高情報責任者)よりも CMO(チーフ・マーケティング・オフィサー=最高マーケティング責任者)の影響力の増加を予想する報道も見かけるようになったのも、IT 部門の存在感が少々薄くなっていることの表れかもしれません。

ただし、IT 部門を飛ばして、経営企画や利用部門に直接リーチする営業戦略がそううまくいくのかどうかは、正直言って疑問もあります。ある大手ユーザーのマーケティング関係者に IT ベンダーの営業の評判を最近聞いてみたのですが、評価は芳しくありませんでした「売りたいものが最初から決まっていて、それを売ることにしか考えていない」という意見で、なかなか会話が噛み合っていないような印象でした。

システムの要件を取り決めるやり取りも、IT 部門ではなく CMO が相手となるとますます大変になるはずですが。

ですので、ビッグデータブームの裏で、「誰に売り込むか」「誰と売り込むか」を再考する動きが IT ベンダーの間で出ていることは間違いのないのですが、やはり IT 部門にもっと影響力を発揮してもらうのが IT ベンダーにとって望ましいという揺り戻しが来る可能性もあり得るように思います。今後の方向性はまだ必ずしも明確ではありません。

3.2 行き過ぎた「データサイエンティスト待望論」への疑問

データサイエンティストの人材像に関する議論も決着はまだ見えきってはいません。いろいろな講演や報道を踏まえると、現状では下記のようなことがデータサイエンティストの人材像について言われています。

- ・業務改善の仮説を立てて、事業部門の現場に提案できる
- ・データの収集・加工ができる
- ・統計手法を駆使できる
- ・分析結果を、現場が分かるように説明し、勘と度胸のマネジメントを科学的に否定し、新たな発見で信頼を得る

大阪ガスでデータ分析組織を率いている河本薫さんという有名な方がいらっしゃいます。河本さんは、データを業務改善に生かすうえで、データ分析のスキルが高度なものかどうかはそれほど重要だとは考えていないようです。ある

講演では「分析スキルについては、ツールの操作が分かればいいと思っています。統計手法の高度さに凝るよりも、データを集められるだけ集めて多く使う方が、いい結果が出ることが多いです」と言っていました。

これがまた重要なのですが、「現場の勘と度胸を頭から否定すると大体けんかになるので、否定してはいけない」と彼は言っています。このぐらい現場には密着してやらなければいけないということで、非常に苦労人だということがよく分かるコメントです。

このほかデータサイエンティストの苦労話として他の企業の人からも

- ・分析結果を出しても、「そんなの分析しなくてもおれたちは分かっていた」と利用部門から反応されることが結構多い
- ・利用部門に回帰ロジスティックやツリー分析など説明しても拒否反応を起こされる。むしろ、やみくもに高度な手法を適用するよりも分かりやすい手法で分析して説明することを心掛けている
- ・段取りと話法が大事。早い段階でどういう指標を良くしたいのかについて合意を得る、一方的に結論を提示するのではなく複数の選択肢を提示して相手に選ばせながら進める、等

といった話も聞きました。

すなわち、企業で待望されているデータサイエンティストの人材像は、単なる統計屋さんではないということは確かです。

ただ、こうしたビジネススキルをあれこれと身につけたデータサイエンティストなくしてビッグデータ活用は進まない、というのはちょっと理想論に走り過ぎている気がします。現場には本当にデータで説得されたいという気持ちがあるのか。そもそも業務改善・改革マインドが現場にあるのか。そういったことまで含めて冷静に議論すべきだと思うのです。経営層や利用部門の側の意識改革とセットで、むしろ組織デザインの問題として捉える方が現実的なのではないでしょうか。

これは私の仮説ですが多くの企業では、経営層や IT 部門、利用部門のそれぞれで、ビッグ

データに向き合う姿勢が以下のようにバラバラのように見えます。

経営者の多くは、「経営状態の可視化」は望んでいるのですが、データの活用イメージまでは考えずに、「IT 部門が活用方法を提案してくるべきだ」と活用方法の考案を丸投げしている話がよく聞こえてきます。

一方、日本の IT 部門の立場で考えると、どれだけ価値があるのかもわからないデータの有効な活用法を、Hadoop のような OSS を使いながら、発見してくれ、といわれるのは、これまでの基幹系システム構築などとはまるで違った業務です。しかも、一般企業でいうビッグデータは顧客データだと JUAS の調査結果で出ているわけですから、コンプライアンスや情報セキュリティの問題も関わってきます。しかしこうした IT 部門の戸惑いが、利用部門や経営者に理解されているとは思えません。

利用部門の意識の壁としては、「自分たちの業務ノウハウをデータで形式知化されて素直に喜べるのか」という問題があります。あるコンサルティングファームの BI コンサルタントによれば、データで得られた知見に基づいて A/B テストをして成果を見極めることさえ、きちんとやろうとしない企業が珍しくないそうです。

企業内のデータサイエンティストは、使えそうなデータを探して、データの欠損を埋めたり異常値を修正したりコードや単位を他のデータと揃えたりといった整備作業だけで、かなり時間を取られるようです。その一方で、現場観察から仮説を立てること、分析結果で人を動かすことが、得意な人は少数派でしょう。

大阪ガスの河本さんの講演にあった話ですが、データサイエンティストには四つの壁があるということでした。

- ① データの壁。ビジネス課題の解決に必要なデータが完全にそろえることはまずありません。
- ② 分析の壁。人材が確保できるかも知りますが、有効な分析結果を短期間で出せるかということも結構大変だといえます。
- ③ KKD の壁。勘と経験と度胸の壁を乗り越

えられるかは、分析結果が出てみないと実は分かりません。

- ④ 費用対効果の壁。どう評価を受けるか、どう経営層に納得してもらうかということです。

ネットサービス企業の業務は、もともとデータ活用に向いています。顧客の行動データをウェブで簡単に取ってこられますし、データ活用して改善する対象は多くの場合、ページデザイン程度です。A/Bテストも簡単にできます。

しかし一般企業では、営業部隊のアクションを変えとか、コールセンターのアウトバウンドサービスを変えなきゃだといった提案をデータサイエンティストがして、関係者を巻き込んでいくのは大変だと思います。

こういったデータに基づいた改善活動が多くの企業に広まっていくには、卓抜したデータサイエンティスト人材の出現を待つだけではなく、組織デザインの在り方ももっと議論が盛り上がってよいような気がします。

4. おわりに

本日の話を総括すると、ユーザーの業種や組織の立場によって温度差が非常にあります。しかし、いろいろなところで壁にぶつかってビッグデータの活用推進活動は停滞する可能性もあります。

特に、ユーザー企業における組織デザインの議論が置き去りにされているので、ぜひ情報システム学会の方にその辺を頑張って啓蒙していただきたいと私は思っています。

最後に一つ余談ですが、最近、機械学習や人工知能が技術分野で非常に盛り上がっています。ただし「過適合・過学習」の問題が、あまり世間的に話題になっていないのは、私はちょっと怖いと思います。

ビッグデータの何がノイズで何が本質的なシグナルかを見極めないで機械学習をしても、将来予測は当たらないので、気を付けないといけません。この過適合・過学習の怖さについては手前みそで恐縮ですが、日経 BP から『シグナ

ル&ノイズ 天才データアナリストの予測学』という本が出ていて詳しく書いてあります。非常に面白いので、ぜひご一読いただきたいです。

あまり将来を指し示す結論はなかったかもしれませんが、私なりの問題提起を今日はさせていただきます。ご清聴ありがとうございました（拍手）。