

Over the AI ― AIの向こう側に(11):

おうちにやってくる人工知能 ～ 国家や大企業によるAI技術独占時代の終焉

<http://eetimes.jp/ee/articles/1705/31/news023.html>

今回のテーマは「おうちでAI」です。といっても、これは「AIを自宅に実装すること」ではなく、「週末自宅データ分析およびシミュレーション」に特化したお話になります。さらに、そうなると避けては通れない「ビッグデータ」についても考えてみたいと思います。そして、本文をお読みいただく前に皆さんにも少し考えていただきたいのです。「ビッグデータって、いったいどこにあるのだと思いますか？」

2017年05月31日 11時30分 更新

[江端智一, EE Times Japan]



今、ちまたをにぎわせているAI(人工知能)。しかしAIは、特に新しい話題ではなく、何十年も前から隆盛と衰退を繰り返してきたテーマなのです。にもかかわらず、その実態は曖昧なまま……。本連載では、AIの栄枯盛衰を見てきた著者が、AIについてたっぷりと検証していきます。果たして”AIの彼方(かなた)”には、中堅主任研究員が夢見るような”知能”があるのでしょうか――。[⇒連載バックナンバー](#)

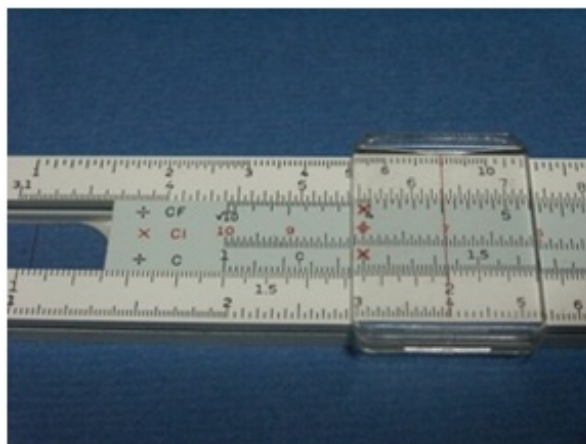
潤沢な計算リソースの使い道が「初音ミク」？

スタジオジブリの映画「風立ちぬ」の中で、飛行機的设计をするために、エンジニアたちが、「計算尺」を使っているシーンが登場します。

計算尺とは、稼働式の2つの定規を組み合わせて、目盛りを合わすことで、計算を行うものです。

私は高校に通っていた頃、選択教科の「計算尺クラブ」に入っていました。

―― なんて、私、そんなクラブに所属していたんだろう？



出典: <https://ja.wikipedia.org/wiki/計算尺>

ちょっと調べてみたのですが、その当時、電卓はもう、それほど高価なものではなく、普通に使われていました。私が大学に入学する頃には、BASIC言語が使える関数電卓があったくらいです——が、まあ、それはさておき。

計算尺も、ソロバンと同様に、練習問題集(ドリル)というのがありましてですね、それが、なんというか、「誤差付き」なんです。確か、解答欄の記載はこんな感じだったかと思います。

“ $4.67 \sim 4.84 \times 10^8$ ”

この範囲の解答であれば、この計算尺による計算結果は「正解」になったということです*1)。

*1) 計算尺ではケタ数の計算はできなかったの、ケタ計算だけは別紙で計算しておく、などの手間が必要でした。

このような誤差を認容する計算手段(計算尺)で、航空力学の最先端技術の塊である戦闘機(いわゆる、“零(ゼロ)戦”など)が設計されていたことは、本当に驚くべきことです。

ちなみに、あの東京タワーも、内藤多仲博士率いるチームが3カ月かけて、計算尺で構造の力学の計算をしていたそうです。



さて、コンピュータの影も形もなかった時代を経て、国家でも企業でない私たち個人に、潤沢な計算処理手が、PCやスマホという形で、掌(てのひら)に落ちてきました。

以前にも申し上げましたが——例えば、ボーカロイドの映像や音声を作り出す計算リソースの消費量は、当時、私が在学していた大学の大型計算機センターの全能力をはるかに超えています。

しかし、その潤沢な計算リソースの使い道が「YouTube」や「スーパーマリオ」や「初音ミク*2」……???

*2) 関連記事: [初音ミクを生んだ“革命的”技術を徹底解剖!ミクミクダンス、音声、作曲…\(Business Journal\)](#)

—— なんか、違う。

そもそも、私たちは、たった40年くらいの間に人類が誕生して以来、経験したこともないような脅威の計算リソースを手に入れたというのに、なぜ、私たちは全然「幸せ」になっていないのだろう?

コンピュータの性能向上は私たちを苦しめている!?

以前、私は、1980年代のスーパーコンピュータと、5000円で昨年(=2016年)購入したボードコンピュータの性能比較をして、計算能力だけでも3千倍～60万倍も向上していることを示しました([該当記事](#))。

コンピュータの性能の比較(再掲)

いつの時代も、人工知能研究は、「コンピュータの性能」を言い訳にしてきたが…(例:「もし、100倍あれば…」)

マシン名称		値段	パフォーマンス	メインメモリ
VAX11/780 (1980年頃) (ミニスパコン)		1億2千万円	500KIPS	5Mバイト
Raspberry Pi 3B(2016年) (格安ボードコンピュータ)		5000円	1581MIPS	1Gバイト
比率		0.000042倍	3162倍	200倍

もう、言い訳できる状況にない

単純な計算であれば、私たちの1日8時間の労働時間は10秒くらい(8時間÷3000)になっても良さそうなのですが ———— そこまで言わないまでもにしても ———— 1日4時間労働くらいの日常は実現してほしかった。

むしろ、コンピュータによる計算能力の向上は、私たちを苦しめる方向のみに働いている、としか思えません。

どうせ「苦しめる」のなら、AIが政府を乗っ取り、外交を進め、戦争を勝手に始め、人類を滅亡させてしまう、くらいのことをやってくれても良さそうなものですが ———— 現在のAI技術では、この境地には遠く及びません([参考記事](#))。

計算尺の時代に生きた人は、コンピュータによって管理された世界によって、人類を幸せにするユートピア(あるいは人類を不幸にするディストピア)がくると信じていたと思うのです。

だから ————

もしも、私が、タイムマシンで、計算尺しかなかった時代にいつて、『潤沢なコンピュータリソースの未来の成果物は、これです!』といて、「初音ミク」を紹介したら、

—— 絶対に叱られる



と思う。

おうちでAI——DIY AI

こんにちは、江端智一です。

今回は、ちょっと趣向を変えまして、「おうちでAI —— Do It Yourself AI (DIY-AI)」という観点からお話したいと思います。

今や、パソコンの計算リソースは、20世紀最後のスーパーコンピュータ(スパコン)の性能を上回っており、少なくとも第2次AIブームのAI技術であれば、軽々と自宅のパソコンで実現できるレベルになっています。

今回は、冒頭のような、ざっくりした比較ではなく、過去のデータや法則に基づき、できるかぎり正確に計算してみました。

計算に際しては、第1次AIブームが終わりを迎えた時であって、パソコン用のCPUとしてインテル(Intel)が開発したワンチップのマイクロプロセッサ「4004」が登場した時期でもある“1970年”を基準に計算リソースの比較を試みました。

まず、パソコンが搭載するメモリの容量を調べてみました。

ムーアの法則

半導体の集積率は18カ月で2倍になる」という半導体業界の経験則

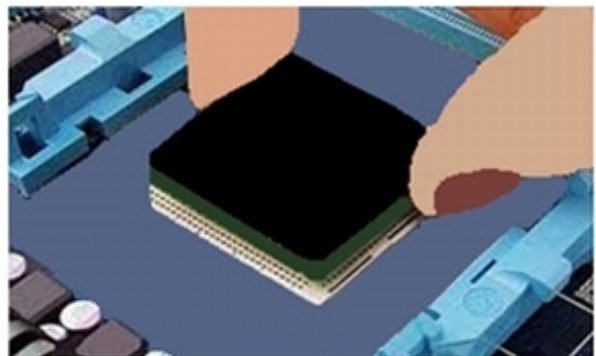
$$\text{パソコンのメモリ搭載量の比率(ザックリ)} = 2^{\frac{12}{18}(X\text{年} - 1970\text{年})} \quad (1970\text{年を基準})$$

エクセルで計算してみた

年(西暦)	1970年	1979年	1984年	2001年	2010年	2017年
比率	1	64	645	166万	1億	27億
登場したパソコン等	(登場前)	PC-8001	Dynabook J-3100	Windows XP	iPad	
						

単純にムーアの法則にのっとると、メインメモリはざっくり27億倍程度になっていることが分かります。

さらに、パソコンの頭脳であるCPU(中央処理装置)性能の推移も調べてみました。



CPUの性能

MIPS値の単純な比較

年(西暦)	1970 年	1979 年	1984 年	2001 年	2010 年	2017 年
比率	1	10	43	4万	38万	174万

参考文献 <https://ja.wikipedia.org/wiki/MIPS>

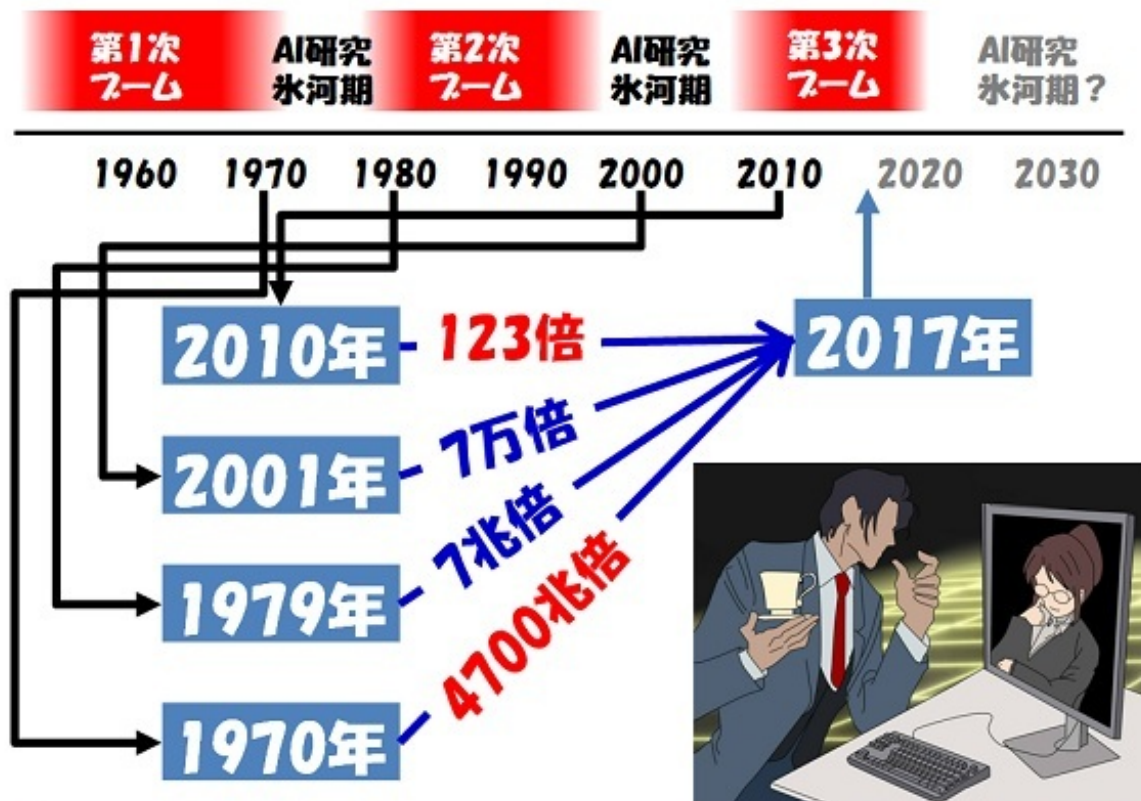
こちら、100万倍を軽く上回る、すごい高速化が図られています。メモリやCPUには「冬の時代」はないのかと、正直ねたましい気持ちになります。

4700兆倍、賢くなっても、「ナンシー」が現れる兆候すらない

さて、上記の2つのデータを用いて、以下のような計算をしてみました。

コンピュータはどれだけ頭が良くなった？

「頭の回転速度が**2倍**になって、記憶力が**2倍**になったら、
「**4倍(2×2)頭が良くなる**」という単位」を導入したら
こうなった



**4700兆倍も頭が良くなっても、
未だに「ナンシー」の片鱗すら現れない**

第1次AIブームの時から、パソコンは頭が4700兆倍も良くなっていて、ほんの7年前からですら、100倍以上も賢くなっています。それなのに、AIの世界においては、ナンシー*4)が出現する兆候すらありません。

*4) 参考記事: [我々が求めるAIとは、碁を打ち、猫の写真を探すものではない](#)

そんなわけで、最近の私は、現在のコンピュータアーキテクチャを踏襲したままで、「人間の代わりとなる人工知能」を作るのは、もうダメなんじゃないかな、と思っているわけです。

もっとも、「人間の代わりになる」というぜいたくを言わないのであれば、日進月歩で増え続けるコンピュータの計算リソースに対応しながら、AI技術も発展を続けてきたと思います。

膨大な計算リソースがAIの発展を妨げている？

しかし、最近、私は、この膨大な計算リソースが、逆にAI技術の研究開発を妨げ始めているのではないかと考えるようになってきました。

そのきっかけとなったのは、自宅で日本人1億2700万人分のオブジェクトをパソコンの中に作って、ここ10年から1000年オーダーの日本の人口の変化などを調べていた時です*5)。

*5) 参考記事: <http://www.kobore.net/diary/?date=20140822>

1億2700万個の構造体のオブジェクトの、39世代目の、7500万番目くらい(番号を特定できない)のポイントが、ヌルポイントを指し示して、セグメンテーションフォールトするというバグで、シミュレーションを続けることができなくなったのです。

そして、このデバッグは、最悪でした。

コアダンプも取れなかったので(原因不明)、バグの発生箇所を特定するために、1億2700万人分のシミュレーションのデータを、パソコンのディスプレイで表示しなければならなかったのですが、——とにかくオブジェクト数や計算量が膨大で、私がパソコンの前を離れ、風呂に入ってしまったら、まだ1世代の計算が終了していない……という有様でした。

AI技術についても同じことがいえます。

AI技術は、膨大な計算を必要とするものが多く、そのアルゴリズムは(原理はともあれ)実装は恐ろしく面倒です。上記のように、メモリリークなどが理由で、プログラムがクラッシュしてくれるなら、まだ幸運なケースです。大抵の場合は、間違った計算をし続けた挙句、正常終了してしまいます。

この場合、期待通りの結果が得られたのか、得られなかったのかを判断する手段がなく、しかも、得られなかった場合、それがAI技術の限界によるものか、あるいはバグによるものかを、判定する方法がないからです。

これが、世間がこれほどまでに「AI」で騒いでいるのに、現場(特にものづくりの現場)で、一向にAIの導入が進んでいかない理由の1つです。

この問題の本質は、

AI技術が膨大な計算リソースを必要としながら、その膨大な計算リソースによって、そのAI技術の処理内容が把握できなくなる

という、巨大なジレンマにあります。

20世紀のスパコンを超えてしまった現代のパソコン

昔のコンピュータ(パソコンを含む)を使ったAI技術の開発は——今考えると、実に長閑(のどか)なものでした(遠い目)。

ソースコードに printf() を埋め込んで、あるいは、GDB (Gun Debugger) で、1行ずつステップを追いかけて、バグを見つけていきましたし、デバッグの最中に実装したアルゴリズムの間違いに気が付くこともできたものです。

第1次AIブームの主演は、スーパーコンピュータ(スパコン)であり、そして、スパコンは、白衣を着た何十人もの研究員によって、24時間体制で、管理運用され続ける、祭壇に祭り上げられる神様のような存在でした(江端の勝手なイメージですが)。

しかし、今やパソコンは、20世紀のスーパーコンピュータ(スパコン)の性能を軽く追い越しています。パソコンは「個人で管理し、制御することができる」という意味において、「パーソナル」な「コンピュータ」だったわけですが、今や、私たち、プロのITエンジニアですら、パソコンの処理の中身を、追跡して理解するのが難しくなっています。

かつてパソコンのソフトウェアは自分で作るものでした(ソフトウェアがなかったから)。信じられないかもしれませんが、私が小学生の頃には、プログラミング教育なんかなくても、自分でゲームをプログラミングしていた子どもがいたのです(確か、「パソコン少年」と言われていました)。

今や、パソコンは、その所有者が、自力で新しい価値(新しいソフトウェアなど)を作り出す装置としては、その役目を終った ―― かのようにも見えます。

しかし、決してそんなことはないと思うのです。

例えば、たった一人でパソコンを使って、キャラクターを創作し、楽曲を作り、そのキャラクターに歌を歌わせて、踊らせる若者がいる ―― これは、紛れもない、パソコンによるパーソナルな新しい価値の創出だと思うのです。

ならば、私も、諦めるべきではないと思うのです。

私は、第2次AIブームの挫折組の一人で、予算のないひもじさで、AI研究員から脱落したヘタレであり、そして、私は、私以外の多くの若手研究員が数年以内にこの第3次AIブームの挫折組として脱落していく残酷な未来も見えています(参考記事)。

しかし、今の私は、「週末自宅研究員」という概念にたどりついています。

加えて、今や、大学や会社ではなく、自宅に20年前のスパコンレベルのコンピュータが鎮座しているのです。これで、今回も諦めてしまったら、私は、本当に、言い訳のできない真のヘタレ研究員に成り下がってしまう、と思うのです。



出典: http://biz-journal.jp/2013/01/post_1323_2.html

おうちでできるAIは？

とはいえ、私の能力でできることは知れている、という自覚はあります。例えば、私が「おうちでAI —— Do It Yourself AI (DIY-AI)」のノリで、「ナンシー」を作り出せる可能性は、絶望的に低いと思います。しかし、それでも「週末自宅研究員」と「週末自宅ライター」を続けていくという覚悟だけはできております。

私の言う「おうちでAI」の狙いは、自分（だけ）でAI技術を開発、適用して、自分のため（だけ）にAI技術を使うことにあります。

そこで、今回、そのような「おうちでAI」という観点から、どのようなアプリケーションやサービスが考えられるか、ざっくり考えてみました。

おうちでAI “DIY-AI”

実用的な、Do It YourselfのAIアプリは何か？

項目	内容	実現性
(1) 自宅/実家 見守りシステム	老親や留守番の子ども(や猫)を見守る	○ ある程度のIT技術が必要
(2) 介護/育児 ロボット	介護や育児の支援をしてくれるロボット	△ コストと安全性が今後の課題
(3) 恋人BOT	自分を元気づけ、慰めてくれる自動応答エンジン	○ 簡単。だが、逆に人生が虚しくなるかも
(4) 渋滞/事故 予測	実家帰省時の渋滞予測、鉄道の人身事故遅延を予測して、快適な生活を手に入れる	× 行政が生データ出すことはほとんどない。予測するより、ニュースを見た方が早い
(5) ホログラム 3D嫁/夫	自宅内で3Dで投影される伴侶	× 技術開発、設備投資と運用資金で破産しかねない
(6) 老人ホーム 4.0(*)	自分自身を介護する、全自動自宅の構築	○ ある程度のIT技術が必要

こうやって考えてみると、あんまり「ない」

(*) 関連記事: [「老人ホーム 4.0」がやって来る](#)

こうやって見てみると、自宅にこそ必要なAI技術というのは、家事、育児、介護に偏りがちにな

りますし、単なる情報処理だけでは足りずに、どうしても駆動機械が必要（例えば、宅内エレベータ、自動運転の車椅子など）となることも問題となります。

となると、施工技術、配電技術、そしてIT技術なども必要となり、エンジニアでさえ難しい話になると思います。

ですので、今回の「おうちでAI」については、「自宅への実装」についてはいったん忘れることとして、週末自宅研究員の「週末自宅データ分析」とか、「週末自宅シミュレーション」に特化して、話を進めることにします。

ビッグデータ解析

ここから、後半になります。

今回は、AI技術では言うまでもなく、AI技術と無関係な分野であっても、避けては通れない、ビッグデータ解析についてお話したいと思います。

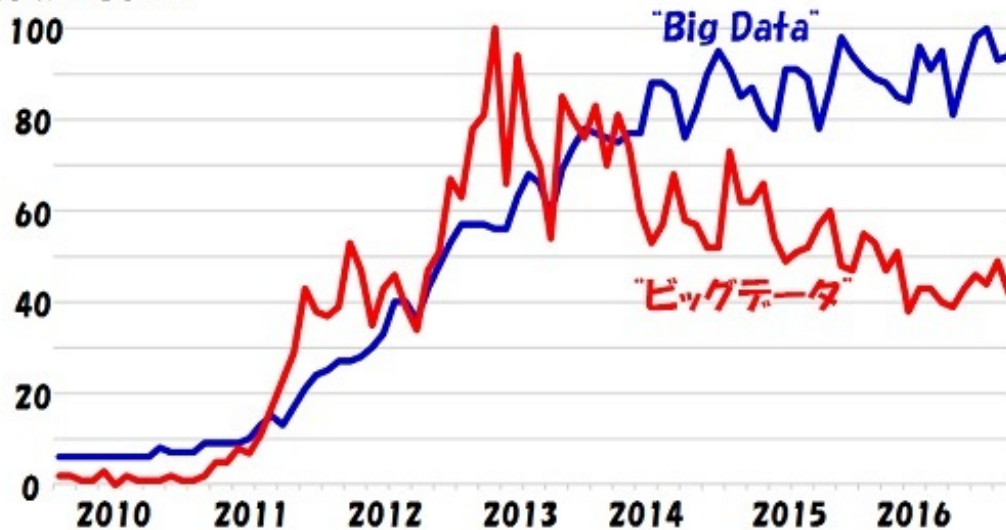
といっても、ビッグデータの解説記事など、検索すればヤマほど出てきますので、今回は、「週末自宅データ分析」という観点に集中して、論じたいと思います。

2～3年前、新聞の第1面にヤマのように登場していた「ビッグデータ」ですが、今や「人工知能」という用語に置き換わった感があります。これが、現在どのような状況にあるのか（「ユビキタス」とか「ロングテール」とか「Web 2.0」とかのように消えてしまったのか）を、調べてみました。

「ビッグデータ」はバズワードだったのか？

Googleトレンドによる“記事数の比率”の推移

[記事数の比率(%)]



国内は減少傾向だが、世界全体では上昇傾向

このように、日本ではピークを越えた感じがありますが、「ビッグデータ」という言葉が消えているとは言えません。また世界的に見れば、“Big Data”の記事が、いまだに増加傾向にあることが分かります。私は、これは「人工知能」技術が「ビッグデータ」と抱き合わせになっているからではないか、と考えています。

ところで、私はかつて、会社の業務で、数千万行のデータ処理を命じられ、その中から「それがどんなものかは知らんが、売り上げに貢献する傾向(トレンド)を見つけ出せ」という、当時のビッグデータ幻想に狂っていた、典型的なエライ人のムチャな命令で ―― この世の地獄を味わった一人です。

『ゴミ箱の中にはゴミしか入っておらず、そのゴミ箱が100万個、1億個に増えても、そこに、間違ってもダイヤモンドが入っているわけがない』 ―― この当たり前の事実すら分からなくさせてしまう「技術バズワード」が、私は、いつだって大っ嫌いです。

そして、私が今回調べた限り、ビッグデータに関する広告や記事の中には、その地獄の中で私が体験した3つの事項(以下に列挙)に対する記載がありませんでした。

「ビッグデータ」で語られないこと

なぜとか？ なぜと語らないのか？

問題点	内容
沢山のデータが「ある」だけ	各拠点に、各拠点毎のフォーマットで記載された(下手すると)手書きのデータがあるだけ
永遠に定まらないデータのフォーマット	(1)データの内容がデコボコ。(2)記載方法がバラバラ、(3)データの解釈もバラバラ、(4)フォーマットを作っている最中にさえ、項目の追加要求が発生する
だいたい、エクセルで足る	国家、行政、巨大企業を除けば、ほとんどの会社はビッグデータなど持っていない(日本の中小企業数は99.7%)

ビッグデータの作成時間→半年
ビッグデータの解析時間→数十秒
なんて話はザラ

はっきりいって、この事実を、悪意で意図的に隠蔽(いんぺい)しているとは思えないほどです。

ダイヤのようなトレンドデータは出てこない

ビッグデータの取り扱いの大半は、小学生の計算ドリルや漢字の書き取り以上に単調で、気が遠くなるほど膨大な作業時間がかかり、そして、多くの場合、その犠牲に見合った成果「ダイヤモンドのようなトレンドデータ」は出てきません。

もしあなたが「ビッグデータ解析で業績が2倍アップした」とかいう記事を目にしたら、ぜひ、私に紹介してください。私が徹底的に検証します。大抵の場合は、ビッグデータ解析うんぬん以前に、自社の財務や製品計画の定量化(数値化)も行わずに、会社経営を続けてきた、マヌケな会社の話であることが多いです。

「ビッグデータ」はどこにあるのか

さて、ここからは、本題の「趣味の週末自宅データ分析」に立ち戻り、検証を続けていきたいと思います。

前述の数千万行のデータの解析をやっていた時、私は世間が「ビッグデータ」とやかましく騒いでいる一方で、「ビッグデータ」が世の中のどこにもないことに気がついていました。

企業が業務データをオープンデータとして公開するわけがないのは、当たり前としても、当時「ビッグデータ」を推進していた政府自身が、データを公開していないのです。

私は当時、「GPS電波の偏差状況を計測することで、地震の予知が可能となるかもしれない」という論文を読んで、GPSを管轄している行政庁のWebサイトにアクセスして、自宅に、GPS電波状況の生データをリアルタイムで取得して、解析するシステムを作ろうと考えていました。

しかし、そのデータを入手するのに、面倒な手続が必要になる上に、もらえるデータが「3カ月前のデータ」。揚げ句「CD-ROMで配送、手数料あり」と記載されているのを見て――キレました。そんな、3カ月もの間、冷蔵庫の外に放置しつづけた生肉……もとい、生データに、何の価値があるのか、と憤慨していました。

で、今回、あらためて調べてみたのですが、やっぱり、生のデータを公開しているような行政機関は見つけられませんでした。

「ビッグデータ」はどこにあるのか(1)

何の権限もない個人が使える公的データ

入手方法	名称	概要
インターネット等からアクセス可能	e-Stat 政府統計の総合窓口	とても役に立つのだけど、解析後のデータ (ビッグデータ)ではない
	j-STAT MAP 地図による小地域分析	解析結果を表示するGISのGUIであるが、 ビッグデータではない
	特許情報プラットフォーム (J-PlatPat)	特許出願の内容の(ほぼ)全部が読めるが、 ビッグデータではない
	市町村が提供する、災害リスクマップ	リスクは表示されるが、その根拠となる 数値データ が開示されている訳ではない
情報公開制度に基づく、官公庁への開示請求	エクセル等で、必要なデータを作成して、送付してくれる(手数料のみ)	元の生データを丸ごと開示という請求はできない →使用目的を明示して請求する必要あり

「個人」向けに公開されている生ビッグデータはない

もっとも上記のe-Statなどは、コラム執筆に際してさんさんお世話になってきましたし、鉄道人身事故のデータを開示してくれた国土交通省さんにも感謝しています。ですが、それでも「生データが欲しい」という気持ちに変わりはありません。

例えば、『国勢調査の全データを、生データで開示してくれたら』—— と思うことは、度々あります。生データさえあれば、私の考える珍奇な仮説(例えば、25歳以内で2回離婚した男性はその後、何度も結婚を繰り返す、など)を読み取れる(かもしれない?)からです。生のビッグデータを使って、私の独自の観点の仮説検証をしてみたいという誘惑は大きいのです。

もっとも、国勢調査の全データが、生データで全公開されたら、仮に名前が伏せられていたと

しても、本人にたどりつくのはそれほど難しくないので、プライバシーの問題から、許可されないのも当然だとは思いますが。

個人として使えるビッグデータとは？

しかし、私たちが個人として使えるビッグデータが、全くないかと言えば、そうでもありません。例えば、次のようなビッグデータは、使い放題です。

「ビッグデータ」はどこにあるのか(2)

何の権限もない個人でも絶対に使えるビッグデータ

入手方法	内容	メリット/デメリット
自分の電子メールのログ / 自分のブログ	自分専用の世界最強のビッグデータ	過去の事例が完全に記録されているので、ものすごく助かる / 嫌なことを思い出し、落ち込んだり、はらわたが煮えくり返ったりする
他人のブログ / SNS / Twitter	日本中(世界中)の人が、気楽に本音を晒(さら)しまくっているビッグデータ	世間の本音が聞ける / 自分の意見をさらし続けている人は、ぶっちゃけ「不注意な人」だから、データの偏りに留意する必要がある

最低でも、テキストマイニングツールは必要

自分の電子メールのログというのは結構、侮れないのです。あれはビッグデータというよりは、自分専用のデータベースといってもよいです。

例えば、1年前にやった(父の)確定申告の手続き方法、購入するのに身分証明書が必要な劇薬系除草剤の名前とか、そんなものを思い出すのに、とても便利です。仕事でお世話になった人の名前を思い出せない時には、仕事の内容で適当に検索すれば、その人の名前をすぐに見つけ出すことができます。

自分のブログは、コラムを執筆する時のネタ元として、なくてはならないものですし、特にパソコンの設定方法や、試しに作ってみたプログラムなどが、どれほど後々に重宝するかは、言うまでもありません。

また、ネットに転がっている、他の人のブログやTwitterのログがなければ「ダイエット」や「飛び込み自殺」の記事は書けなかっただろうと思います([参考記事一覧](#))。これらは、十分に「趣味の週末自宅データ分析」に欠かせない、私にとっての貴重なビッグデータです。

しかし、個人的に言えば、これらのデータは、まだまだ「ビッグ」なデータとは言えません。私にとって、「ビッグデータ」とは ―― それを「ビッグデータ」と呼んでよいのかはさておき ―― 私が自作したシミュレーションプログラムが吐き出す、膨大な計算結果です。

「ビッグデータ」はどこにあるのか(3)

シミュレーションプログラムが吐き出す膨大なデータ

入手方法	内容	課題/狙い
1億2700万人のシミュレータの結果	(仮想の)日本国民1億2700万人の仮想日本人の振舞いを、調べる	(1)データが膨大すぎて、ぶっちゃけシミュレーションの結果が全然分かん
人口3億人の合衆国の大統領選挙のシミュレーション結果	(仮想の)合衆国の国民に100万回ほど、選挙を繰り返してもらう	(2)データ解析用のコードを実装するのが、思いっきり面倒くさい → ビッグデータ解析ツールを使って、ラクをしたい

「自分で作ったシミュレータが、何を出力しているのか全然分かん」の対応が必要

データの行数が億の単位を超えると、本当に「何がなんだか分からない」という状態になります。そんな膨大なデータを視認するのは、人智を超えた所業です(今、Windowsのコマンドプロンプトの画面(80行=19cmで換算)で測って見たら、1億行のデータは237kmになることが分かりました(東京～浜松間くらい)。これを100年分やると(以下省略)。

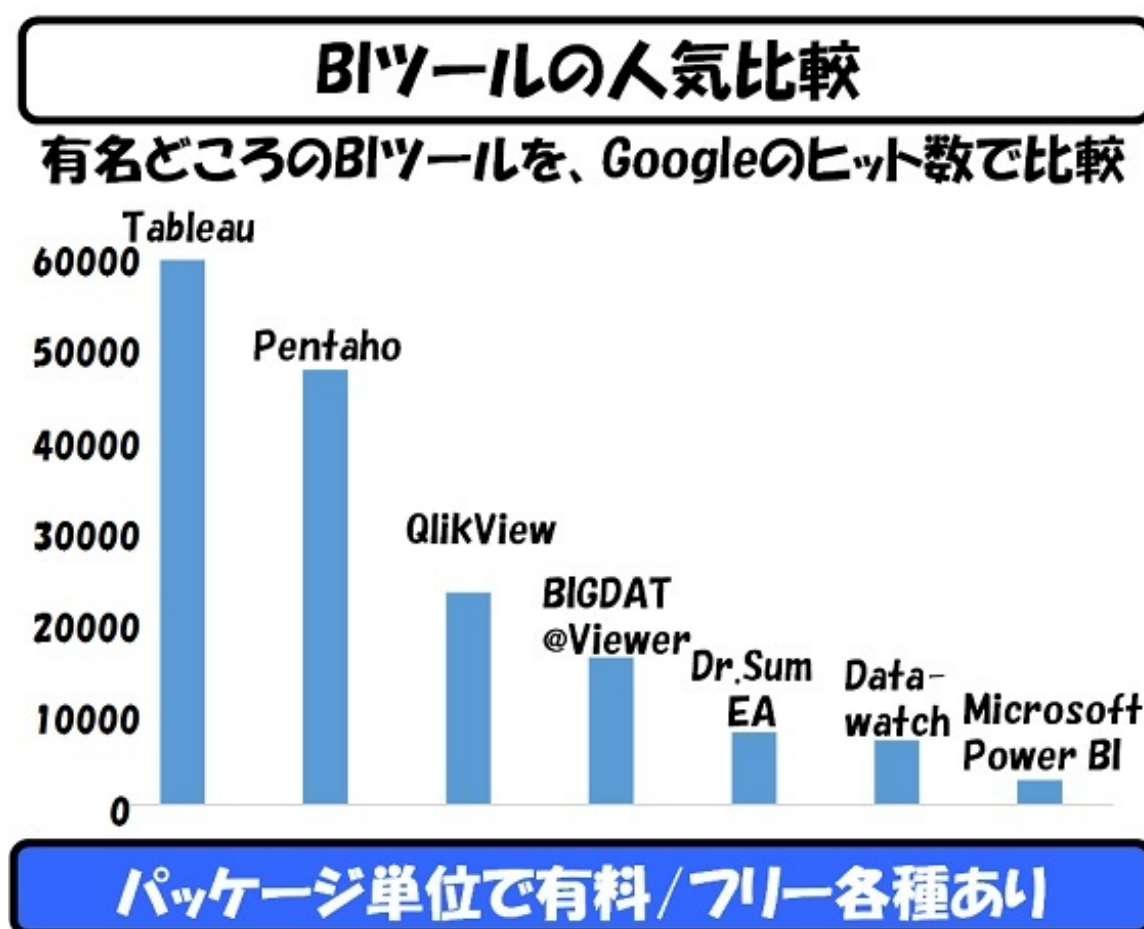
[米国大統領選挙のシミュレーション](#)では、14時間ほど要したのですが、その結果を見た後で、解析用のプログラムコードを追加するなどという作業をやっていたら、1日1回しかシミュレータを回せないことになり、何のためにシミュレーションやっているのか、分からなくなります。

このように、シミュレーションにおいても、『コンピュータの計算能力向上のおかげで、やれることがヤマほど増えて、そのためにできないことがヤマほど増える』という、典型的な負のループに陥るわけです。

自作シミュレーション結果を解析する手法

ではここからは、週末自宅研究員として、自分のメールログやネット上のTwitterの情報、そして、自作シミュレーションの結果を解析するために、どんな手段や方法を用いればよいのかの検討に入りたいと思います。

まず、Google検索でざっくり有名どころのBI(ビジネスインテリジェンス)ツールを調べてみました。



BIとは、企業に蓄積された大量のデータを収集して分析するためのツールです。さまざまな機能がありますが、一言で言えば、ビッグデータを見える化(グラフ化)するものと言い切ってもよいかと思います。

今回、これらのBIツールのデモビデオをYoutubeでレビューしました。感想は以下の通りです。

おうちでビッグデータ解析(1)

YouTubeでBIツールのデモをざっくり見た所感

項目	内容
(江端にとっての)メリット	(1)行き届いた多様なデータ表示機能 (2)マウスだけで操作可能 (3)充実している附属機能(データ予測等)
(江端にとっての)デメリット	(1)個人で使うには 高い (ものがある) (2)個人で使うには、 いらん機能が多すぎる (ものがある) (3) サーバ とか クライアント とか ストレージ とか、 パッケージ の追加とか、 システム構築 とか、 ライセンス管理 とかを考えると「 ウンザリ 」する その他、「江端家に インメモリDB を購入する金があるか!」など

大げさすぎる

個人で使うには、「ちょっと、いろいろ無理かなあー」という感じでした。

最も無理に感じたのは「値段」の問題です。「週末自宅研究員」には、解析ツールに投資するお金はないのです。

実際、今でも、さまざまな会社に参加して、頼み込んで、各種のプラットフォームやツールやデバイスをお借りして、解析を続けているくらいです(『連載コラムで使いたい』と申し上げると、多くの場合、快く貸していただけるようです)。

ですから、フリーのツールを構築し、使いこなすように、日々努力を続けることは、「週末自宅研究員」にとっては、必須(というか必死)のミッションであるのです。

おうちでビッグデータ解析(2)

「金がない」週末ライターという観点から、フリーのツールを検討

名前	内容
R	(1)コマンド一発の統計解析が可能 (2)プログラミング言語(R言語)もある (3)しかし、「簡単な計算を、高速に」、というコンセプトではない(みたい)
Hadoop	(1)とにかく、取り扱えるデータの規模がめちゃくちゃにでかい(テラとかペタデータに特化したバッチ処理システム) (2)複数のDBを束ねて取り扱える(分散コンピューティング環境)

「パソコン1台だけ」で、「ちょいちょい」
という感じではない

私は今、データ解析やシミュレーションを手軽に行えるツールの1つとして「MATLAB / Simulink」を使っているのですが、今後は「R」の使用も視野に入れて、勉強していかなければと思っています。しかし、さすがに自宅で「Hadoop」を構築することはなさそうです。自宅で分散データベース構築するのは、「いくらなんでも、ないわー」と思いますので。

やはり王道は「Excel」だ

というわけで、今さらながらの結論ですが、やっぱり王道はExcelですね。最近のExcelはすごくなって、100万行くらいのデータを軽く扱えます(以前は6万行程度が上限で、実際のところ1万行以上のデータでは、よく固まって(フリーズして)いました)。

しかし、100万と1億の間の2ケタの壁は、どうにも高く、全てをExcelで済ませるというのは無理です。そこで以前、1億人2700万人のシミュレーションに取り組んだ時に、「Aktblitz III」という解析ツールを使わせてもらっていました。

おうちでビッグデータ解析(3)

数億行のデータ解析を1台のパソコンでやっつける

要件	内容	評価
エクセル	今や、100万行 (x1万6千列)の データを扱える	江端が必要とするデータ 解析のほとんどは、これで 十分
Aktblitz III	1億人分のオフ ジェクトを扱うシ ミュレータでは、こ れが最後の頼い だった	1億(最大20億行)のデー タ解析を数秒で実施 #ただ、最初のデータの ローディングに10分以上 かかるのが難

**Aktblitzを試した時に、PCがスパコンになっ
たと感じた(本当)**

Aktblitz IIIを使った時に、「ついに、パソコンで1億行を超えるデータ解析を数秒で得られる時代になったのか」と、隔世の感がありました。

実際に、Aktblitz IIIを開発された方にインタビューさせていただき、「アルゴリズムの工夫で、ここまでの高速化を実現した」というお話を聞いた時には、泣きそうになるほど感動したのを覚えています*6)。

*6)もし、あなたが、私を泣かせたいなら、「ビジネスモデル」や「社会トレンドの未来予測」なんぞの話でなく、「技術」の話を持ってきてください。

今回、「パソコンで億行」という観点で、ざっくり調べてみたのですが、例えばPower BIファミリーの中の「Power Pivot」なども、十分に速いらしいです([参考記事](#))。

いずれにしても、「ビッグデータ解析 = 大型計算機センター」というパラダイムが、「ビッグデータ = おうちのパソコン」に移行しつつあると言えそうです。

もっとも、今後、データ量が、「億」から「兆」、そして「京」になっていく可能性もありますから、断定的に言うのは差し控えたいと思います。

そういえば、『パソコンに、32メガ(単位に注意)もの巨大なメモリを搭載する必要が、一体ど

ここにある!』と私を叱責した、かつての上司 ―― 今、すっかりエライ人になっていますが ―― 今度、公の場(懇親会など)でこの話を振ってみようと思っています(私、けっこう執念深いんです)。

ITエンジニアですら、中身を把握するのが難しい

それでは、今回のコラムの内容をまとめてみたいと思います。

【1】今回の前半は、記憶容量(メモリ)の増加量と、CPUの処理速度の歴史を振り返って、パソコンの性能が、過去数年間でも100倍以上、パソコン創成期から比較すると4800兆倍という驚異的な進化を遂げてきたことを示しました。

これに対するAI技術の進歩を鑑みた時、「現在のコンピュータアーキテクチャの元での『人間の代わりとなる人工知能』を作るのは無理なのではないか」という私の所感を述べました。

【2】このようなコンピュータの驚異的な性能向上によって、AI技術の発展が促されているという側面はあるものの、逆に、その性能向上によって、開発者ですらAI技術の処理内容が分からなくなるという、大きなジレンマが発生することを、江端のプログラム開発現場の具体例で説明しました。

【3】私たちは、これまで先人が手に入れることができなかった潤沢な計算リソースを、すでに自宅に持っていることを示し、来たる「AI冬の時代」に対しても、国家や企業の思惑を超えて、週末や休日の自宅研究員として、研究を継続できることに言及しました。

【4】「おうちでAI」という新しいコンセプトを提示し、現時点においては自宅で実現可能なAI技術のサービスをいくつか紹介しました。しかし、その多くが、稼働装置を伴うものになるため、当面の間は、「週末自宅データ分析」とか、「週末自宅シミュレーション」に特化して検討することとして、この方法について解説しました。

【5】「ビッグデータ」という言葉が氾濫していた数年前に、「ビッグデータ」のうたい文句を真に受けた幹部によって、地獄のデータ解析作業を強いられた、気の毒な研究員のエピソードを紹介しました。また「ビッグデータ」を叫んでいた当事者(政府など)が、ろくすっぽ、まともな「ビッグデータ」を開示しておらず、今も開示していないことを明らかにしました。

【6】週末自宅研究員として、ビッグデータ解析に必要となるツールなどについて、江端家の事例を使って具体的に説明しました。

以上です。

□

計算尺で戦闘機や東京タワーを設計していた時、エンジニアたちは、自分たちが何を作っているか、そしてそれがどのような特性を有し、どのように適用、発展させるべきかを、正確に理解していたと思います。

しかし、ここ数十年のコンピュータの爆発的な性能向上によって、私たちエンジニア（特にITエンジニア）は、私たち自身が何を作っているのかを、（辛うじて）理解していても、その作っているものの内容を、把握することすら難しくなっているように思えます。

特に、AI技術に関しては —— 特にニューラルネットワーク技術などでは —— 「なんか知らんけど、上手く動いている」の世界を地で行っています。これは、理解可能な原理や仕組みを確実に積み上げて、1つの製品やサービスを完成させる、という現代工学の基本原則に反しているかのようにすら思えます。

しかし、その一方、

—— このパソコンの中では、一体、何が作られて、何が行われているんだろう

と、コンピュータやAI技術のブラックボックス的な怖さに怯えつつも、出てきた答えに驚愕（きょうがく）するあの瞬間は、なんともいえずにワクワクするものなのです。

⇒「Over the AI —— AIの向こう側に」[⇒連載バックナンバー](#)



Profile

江端 智一（えばた ともいち）

日本の大手総合電機メーカーの主任研究員。1991年に入社。「サンマとサバ」を2種類のセンサーだけで判別するという電子レンジの食品自動判別アルゴリズムの発明を皮切りに、エンジン制御からネットワーク監視、無線ネットワーク、屋内GPS、鉄道システムまで幅広い分野の研究開発に携わる。

意外な視点から繰り出される特許発明には定評が高く、特許権に関して強いこだわりを持つ。特に熾烈（しれつ）を極めた海外特許庁との戦いにおいて、審査官を交代させるまで戦い抜いて特許査定を奪取した話は、今なお伝説として「本人」が語り継いでいる。共同研究のために赴任した米国での2年間の生活では、会話の1割の単語だけを拾って残りの9割を推測し、相手の言っている内容を理解しないで会話を強行するという希少な能力を獲得し、凱旋帰国。

私生活においては、辛辣（しんらつ）な切り口で語られるエッセイをWebサイト「[こぼれネット](#)」で発表し続け、カルト的なファンから圧倒的な支持を得ている。また週末には、LANを敷設するために自宅の庭に穴を掘り、侵入検知センサーを設置し、24時間体制のホームセキュリティシステムを構築することを趣味としている。このシステムは現在も拡張を続けており、その完成形態は「本人」も知らない。

本連載の内容は、個人の意見および見解であり、所属する組織を代表したものではありません。

関連記事



[外交する人工知能 ～ 理想的な国境を、超空間の中に作る](#)

今回取り上げる人工知能技術は、「サポートベクターマシン(SVM)」です。サポートベクターマシンがどんな技術なのかは、国境問題を使って考えると分かりやすくなります。そこで、「江端がお隣の半島に亡命した場合、“北”と“南”のどちらの国民になるのか」という想定の下、サポートベクターマシンを解説してみます。



[「英語に愛されないエンジニア」のための新行動論とは、“愛”である](#)

今回は、最終回とは別に「総括編」をお届けします。「英語に愛されないエンジニア」と「英語に愛されるエンジニア」。どちらがよかったかと問われれば、私は間違いなく後者を選ぶでしょう。「英語に愛されないことで、別の幸せがあった」などと締めくくるつもりは毛頭ありません。ただし、本連載で展開してきた新行動論のバックボーンにあったのは、常に“愛”でした。



[もはや我慢の限界だ! 追い詰められる開発部門](#)

コストの削減と開発期間の短縮は、程度の差はあれ、どの企業にとっても共通の課題になっている。経営陣と顧客との間で“板挟み”になり、苦しむ開発エンジニアたち……。本連載は、ある1人の中堅エンジニアが、構造改革の波に飲まれ“諦めムード”が漂う自社をどうにかしようと立ち上がり、半年間にわたって改革に挑む物語である。



[カップ1つ作るのに、ご主人様とメイドは4000回会話する](#)

産業用ネットワーク「EtherCAT」の世界を「ご主人様」と「メイド」で説明し続けてきた本連載。今回からは最終章と題し、EtherCATを開発したベッコフとEtherCAT Technology Groupの方々へのインタビューの様子を紹介していきます。今回は、EtherCATを使って感じた不満、疑問を遠慮なく、ぶつけてみました。



[「スカイネット」が現実には？ AIの未来像は“自我”がキーワード](#)

10年前は、「最も革新的なAIは、自己学習すること」と言われていた。実際、自己学習するAIの開発は進んでいて、実用化が可能なレベルまで達しているものもある。“自己学習”がもう一歩進めば、AIに“自我”が芽生える可能性も否定できない。映画「ターミネーター」に登場する「スカイネット」のような、自我を持つ人工知能がAIの未来像になるのだろうか。



[「キャリアは自分で切り開く」に踊らされない](#)

長く仕事をする上で、キャリアデザインというのはとても重要です。キャリアデザインには、社内異動の他、より高みを目指すための、あるいはキャリアチェンジのための転職も含まれるでしょう。しかし、転職雑誌やサイトで目立つ「キャリアは自分で切り開け!」という言葉に踊らされてはいけません。そもそもキャリアデザインというのは、それほど簡単なものではないからです。まずは、筆者が提示する「3つの質問」について答えを考えてみてくだ

さい。

Copyright© 2017 ITmedia, Inc. All Rights Reserved.

