

面白い本。読みながら、ネット検索して学びを深めていける。オーパーツ、オーグメント・デバイスなどの言葉も知ったし、クッキーにもサードパーティクッキーというのがあって個人の行動監視ができることを知った。

仮に 30 年前の人々がスマートフォンを見たら、**オーパーツ**（その時代、その場所に存在することが不自然なもの。ヘッジスの水晶髑髏やコスタリカの石球が有名）であることは間違いない。ダグラス・エンゲルバート（マンマシンインタフェースの先駆的な研究者。マウスを開発）が提唱し、「**夢想**」とも許された**オーグメント・デバイス**（人間の能力を拡張する装置）そのものである。世の中の大多数の人が、いきなりドラえもんを手中にしたに等しい。

こうした性質があるため、コンピュータに関わる技術者はデータが失われることをとても嫌う。冒頭に記したように、恐怖しているといってもよい。これは**コンピュータの運用に関わった技術者であれば、生涯逃れることのできない悪夢**である。顧客の機密情報を消してしまったら、お詫びのしようもないのだ。

遠隔地バックアップ、世代管理、フルバックアップ、**差分バックアップ**、**増分バックアップ**……、初級者向けのテキストに出てくる用語だけを並べても、バックアップ（予備として使う複製）に関連したものがいくつも登場する。新人研修などでも、「重要なデータは最低 3 つは複製を作りましょう」などと言われることがあるし、「せっかく複製を作っても、オリジナルと同じ場所に置いては火災などで同時に失われることがあるので、遠隔地にも配置しましょう」とアドバイスされる。

・・・

そこで毎日のバックアップの各々を残しておくのである。これが**世代管理**というもので、毎日（とは限らない、毎週でも毎時間でも、その企業の事情に応じて取得する）全範囲のコピーを保存するフルバックアップ、基準になる日（たとえば月曜日）を決めて全範囲をコピーし、火～日曜日については基準日から変更のあった部分だけを保存する**差分バックアップ**、基準日に全範囲をコピーし、以降は当日分の変更点だけを保存する**増分バックアップ**などの方法が用意されている。

また、近年進む**デジタルフォレンジック**（デジタル鑑識）に対応するために、世代管理されたデータを数年分保存する企業すら増えている。デジタルフォレンジックとは、情報システムに関する記録や成果物を、高い精度で残しておく技術や運用方法の総称で、何らかのトラブルが生じたときに法的な証拠として使うことを想定している。そのため、取得しておくデータは機器操作記録、ある時点でのメモリ上の記憶内容など広範にわたり、かつ取得しておいたデータが第三者によって**破壊・改竄**されていないことも証明する必要がある。

どんなところにも抜け道はある。たとえば、スイッチにはミラーポートと呼ばれる管理目的のケーブル差し込み口がある。見た目は通常の差し込み口と同じだが、管理目的であるがゆえの特権が与えられており、その1つにスイッチを経由していくパケットをすべて見るができる権利がある。管理者が本当に管理目的で使うのであればよい。しかし、第三者がそこにパソコンを接続することができれば、スイッチを経由していくパケットはすべて見る事ができてしまうのだ。

インターネットの特徴に「ベストエフォート」がある。パケットを届けるために最大限の努力はするが、届かなくても保証はしないという考え方である。乱暴なようにも聞こえるが、このおかげでインターネットにおける通信コストは劇的に下がった。そして、「最大限の努力」を担っているのがルータである。パケットを中継するルータは、周囲のルータと経路情報をやり取りして、どの道が最善か、故障等で使えなくなっている道がないか、常にチェックしている。一番いい道を選ぶためである。また、道が混んでいたり、故障したりしていた場合、別の道を探したり、パケットを自身に保持しておくこともする。送信可能になったときに、送り出すためだ（保持していられないくらいパケットが溜まると、廃棄することもある）。

それではクラウドはどうだろうか。自宅のパソコンにかわって、計算能力の生成を担うのはデータセンタと呼ばれるサーバの集合体である。ここでいうサーバとは、パソコンより処理能力が大きく、各種の高可用性対策（壊れない、直しやすい）を施したマシンのことで、コンピュータであるという点においてパソコンと本質的な違いはない。

とはいうものの、パノプティコンは監視社会の議論を始めるためのベースラインを構築するものであることは現代でも論をまたない。ここで少しだけおさらいしておこう。

・・・

パノプティコンは、この思想のもとに設計された刑務所である【写真④】。

・・・

パノプティコンは美しい円柱状の建物で、円柱の表面に受刑者の部屋が配置されている。円柱のなかはがらんどろになっていて、中央に監視塔がある。こうすることで、看守は極めて少ない人数で多くの受刑者を監視することができる。



セッション管理の技術はいくつもあるが、現在最もよく利用されているのはクッキーである。クッキーはクライアント/サーバシステムで稼働し、サーバ（サービスを提供する側のコンピュータ、この場合はホームページを見せるウェブサーバ）が作成して、クライアント（サービスを享受する側のコンピュータ、この場合はホームページを見せてもらうウェブクライアント/ブラウザ）に送信し、クライアントの記憶装置に保存する。

クッキーは単純なテキストファイルなので、興味がある方は中身を見ることが出来る。クッキーをどこに保存しているかは、ブラウザの設定画面で確認が可能だ。たとえば、インターネットエクスプローラであれば、メニューバーの「ツール」→「インターネットオプション」→全般タブの「閲覧の履歴」の「設定」→インターネット一時ファイルタブの「ファイルの表示」とクリックしていけば、保存場所を見ることが出来る 【図 8】。

「インターネット一時ファイル」とあるように、最近見たホームページや画像のキャッシュ（再度見るときに、高速で再現できるよう、パソコン内に一時保存しておくこと）と一緒に保存されているためやや紛らわしいが、cookie と名付けられたテキストファイルを見つけることができるはずである。一般的なテキストエディタ（ウィンドウズであれば「メモ帳」）で開けば、中を確認できる。

人間が読める形で情報、たとえばサイトの URL が善かれているかもしれないし、意味不明な数値が書かれているかもしれない。クッキーは用途に制限がないため、サーバがちょっとしたメモ帳代わりに使うこともある。しかし、多くの場合は数字の羅列に意味があって、その数値で利用者やパソコンを特定したり、セッションを特定したりする。



【図8】 「ファイルの表示」をクリックすることで、コンピュータ内に保存されているクッキーを見ることが出来る。

実際に、クッキーが使われることで、いちいちパスワードを入力しなくても恩恵にあずかっている人も多いはずである。

・・・

多くのウイルス対策ソフトは、**ウイルスチェックの際にクッキーを消す**。利便性だけを考えれば迷惑この上ない。そもそも、ふだん見ているサイトのクッキーをなぜわざわざ消す必要があるのだろうか。

もちろん、行動の追跡ができるからである。クッキーは基本的には、そのサイトからしかアクセスすることができない。たとえば、新潮社のウェブサーバが作って、ホームページを見に来た人のパソコンに保存したクッキーは、**新潮社のウェブサーバだけが読み出すことができる**。これが基本で、**ファーストパーティクッキー**という。

最初はこれでうまくいっていたが、ウェブ上の広告などが当たり前になってくると、不都合が起こってきた。A社のホームページにB社の広告が載っているようなケースである。ファーストパーティクッキーの考え方では、A社のウェブサーバがクッキーを作って利用者のパソコンに保存する。すると、B社はクッキーを使うことができない。

B社はA社のホームページに間借りさせてもらっている状態とはいえ、クッキーは使いたいわけである。クッキーを使うことで広告をいつ見てもらえたか、広告を見た後どのページに移動したかなどの情報を蓄積できるからだ。

この問題を解決するために、現在見ているウェブサーバ以外のサーバが発行したクッキーも使うことができるようにした。これが**サードパーティクッキー**である。

サードパーティクッキーはバナー広告などでよく使われる。それ自体は悪いことではない。利用者がどんな広告を見て、どんな行動を取っているかを知れば、もっとよい広告を打って利用者の利便性を向上させることも可能だからだ。

だが、行きすぎると、行動の捕捉につながる。たくさんの広告をさまざまなサイトに出稿している広告代理店であれば、利用者の行動履歴を統合して緻密に分析することができる。利用者はサイトを利用するときに、そのサイト上で何をしていたのかが記録に残るくらいは認識しているだろう。だが、サイトを移動すれば、行動記録は途切れると考えるのがふつうである。企業が違うのだから、順当な判断である。実際、A社のサイトでケーキを買って、でダイエットタブレットを買っても、その履歴は結びつかない。

しかし、サードパーティクッキーはこれを結びつけてしまう。広告代理店C社が、A社とB社に広告を出稿して、クッキーを発行していた場合、**サイトをまたがった行動を記録することが可能**である。冒頭で、クッキーは一種の監視システムだと書いた理由の1つである。

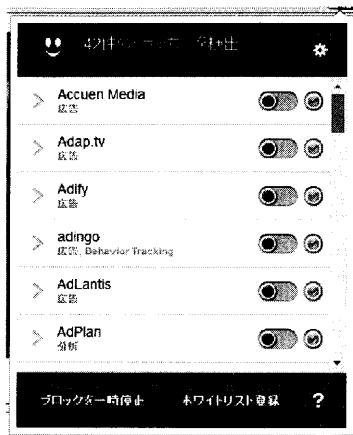
利用者にしてみれば、まったく認識していない企業（あるホームページを見に行った際に、そのページに掲載されているバナー広告を出稿している企業をすみずみまでチェックする人は稀）にウェブ上の活動記録を作られ、その情報を**自分のパソコンの片隅に記録されてしまう**のである。

「クッキーでウェブ上の行動を追跡されているかもしれない」と言われても、自分の問題として実感を伴って受け止められる人のほうが少数派だろうが、一度「**Ghostery**」というプラグインを自分のブラウザに組み込んで試してみるとよい。あるウェブページを訪問した際に、どれだけのクッキーによって行動を追跡されているかがわかる。

【図9】は、あるアダルトサイトにアクセスした際、クッキーをはじめとする追跡技術がどのくらい使われたかを示した一覧である。ここにリストされたサイトに情報が送られ、図中に記された「広告」「分析」「**Behavior Tracking**」などに使われている。

追跡されていることがわかりにくい、対処しにくいという点において、サードパーティクッキーは問題を孕んでいると言えるだろう。

だからウイルス対策ソフトも、クッキーを危険であると検出して消してしまうのである。対策ソフトによっては、クッキーをスパイウェアとして報告するものもある。



【図9】あるアダルトサイトでは、42のサービスに情報を送信していた。追跡行為であるとして、ウイルス対策ソフトがクッキーを消去することもある。

ここで犯罪予測の話題が登場するのは意外だろうか。し

かし、同じ予測には違いない。その人が未来に購入する商品がわかるなら、未来に犯す犯罪も、未来にかかる病気も予測することができる。かくのごとく、**現在のデータマイニングシステムの精度は高いのだ。**

罪を犯しそうな特性や行動、環境などは現時点でも概ねわかっている。その精度を高めて、入力するデータ量も増大すれば（監視社会化により、至る所にセンサーや監視カメラがある。クレジットカードの購買履歴だって分析に使える）、**ある人物がいつ罪を犯すのか予測できるようになる日が来る。そんなに遠い未来の話ではない。**現時点ではさすがに予測に基づいた逮捕は行われていない。しかし、テロを未然に防止するためのプロファイリングは、それに近いことをしているし、アメリカでは性犯罪をおこした人にGPS端末を取り付ける州があり、韓国、フランスでも同様である。いずれも、性犯罪は再犯率が高いことを受けての措置である。

私の子供も自閉症スペクトラム障害である。だから、彼女の思考はトレースできると思う。乏しい医学知識を技術屋風に解釈すると、知的障害が中央処理装置に不具合を抱えているのに対して、自閉症スペクトラム障害は入出力装置に不具合を抱えているイメージである。入力と出力でトラブルが生じがちなので、突然走り出すなどの突拍子もない行動（出力）が起こることがある。自閉傾向を持つ子は総じて迷子になりやすい。

JR 東日本は車両に追加センサーを載せ、収集

する情報を拡大した。それを使ったサービスの皮切りとなったのが、**山手線トレインネット**であ

る。

乗客はスマートフォンにアプリをインストールすることで、山手線トレインネットを利用することができる。現時点で使うことができるサービスは運行状況、車両ごとの混雑状況と室温、乗降駅情報などである。それだけ聞くと、従来からデジタルサイネージなどで提供されていたと感じるかもしれないが、このサービスは利用者がどの車両に乗っているかまで識別して、それに見合った情報を提供してくれる。1号車に乗っているのであれば、1号車から最短の出口案内を表示する。

山手線トレインネットでは、この位置識別に超音波を使う（NTT ドコモが開発した、Air stamp という音波を使ったチェックインシステムを採用している）。場所を特定する技術としてはGPSがあるが、車両を正確に特定するにはGPSでは誤差が大きい。トンネルなどで電波が届かないこともある。そこで、人間には聞こえない周波数の音波を流し、スマートフォンのマイクに拾わせるのである。車両ごとに異なる音源を鳴らしておけば、マイクが拾った音を分析することで、どの車両に乗っているかが正確に把握できる。施設への入退館などを管理することができる技術である。GPSより精密に人の動きを監視する技術の浸透が進んでいることを示す事例である。

~~~~~

山手線トレインネットは、スマートフォン用アプリである「JR 東日本アプリ」に実装されている機能の一つで、山手線の運行情報、電車の走行位置や各電車の混雑度を車両単位で見ることができるものである。

<http://www.jreast-app.jp/>

~~~~~

ヤフーの議席予測はその後も継続されていて、2013年の参院選でも予測を行った。その予測の様子や、関連モデル、投影モデルの簡単な解説は今でもアーカイブを読むことができる（<http://event.yahoo.co.jp/bigdata/senkyo201307/>）。

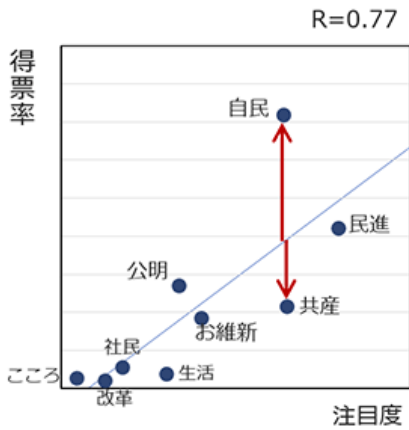
この選挙では121議席が争われたが、ヤフーの議席予測精度はさらに向上していて、投影モデルが111議席の予測を的中させている。人間であれば、票読み屋として一生喰っていけるレベルである。恐ろしいのは、ヤフーのビッグデータ分析は政治的な文脈を何もっていないことだ。どの党がどんな政治主張を持っているのか、どんな歴史的経緯があるのか、民意との乖離はどうかといったデータをまったく持たない。議席予測を行うアナリストが駆使するこうしたツールを用いず、単に検索数だけで議席の予測をしているのである。1980年代にエキスパートシステムが流行したことがあった。人間の専門家の思考と作業手順をトレースして同じ知見を導き出すシステムである。ヤフーの選挙分析はエキスパートシステムですらなく、人間がこの問題を考えるときとまったく異なるアプローチで正解に近い値を導き出している。

~~~~~

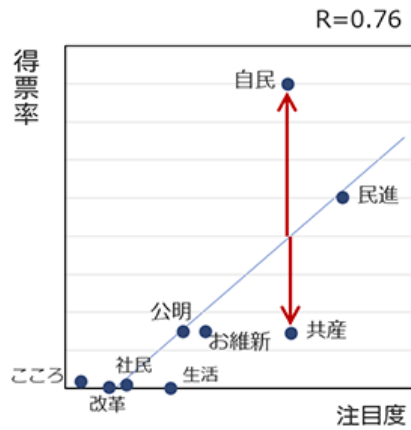
<https://about.yahoo.co.jp/info/bigdata/election/2016/03/>

本来は2013年参院選のグラフのようにすべてのプロットが直線上か誤差のほとんどない近い位置に並ぶという結果になるはずでしたが、比例区・選挙区ともに自民党と共産党が大きく乖離しました。

比例区



選挙区



~~~~~

まさかそんなことが、という意外な相関が発見されることもある。図書館の利用率は学業や生活全般に与える影響の大きい項目だが、1年生の秋に頻繁に図書館を使い始めた学生は何故か成績が向上する傾向がある。学食の利用頻度が高い学生は、成績もよい。

【図 13】は、ある企業のコミュニケーションサービスの利用規約である。利用者はサービスの利用に先立って、この利用規約に同意しなければならない。規約がきちんと明示されているのはよいことだが、このつらつらと並べられた長文をはたして利用者全員が読むのだろうか？ まず期待できないだろう。そして、注目したいのは枠で囲まれた改訂前の「3.」の記述である。「当社は、すべての会員記述情報を無償で複製その他あらゆる方法により利用し、また、第三者に利用させることができるものとします」とある。これは、**個人情報を守る規約では決してない**。この企業は個人情報を守る気はなく、利用者の同意を取り付けることによって、あらゆるサービスに利用しようと目論んでいる。

...

守ろうとしているのは、個人情報ではなく企業と担当者の立場である。

実はこの事例は、2013年にJR 東日本と日立製作所の間で行われた実際の取引である。**JR 東日本が Suica の乗降履歴をビッグデータの活用のために、日立製作所に販売しようとしたのだ**。suica の乗降履歴はマーケティング資料として非常に有益で、これを販売しない手はないとしたものだが、どのような匿名化を行うのか、どの範囲のデータを売るのがはっきりしなかったため、利用者が反発した。その後、クレームが続く中で、「駅の乗降履歴、利用額、日時、性別、年齢」が提供されるデータで、氏名、電話番号、住所、生年月日、物販履歴は対象外であることが示されたが、とにかく何らかの匿名化処理を施せばいいだろうとの考えが社会的な信用も

厚い一部上場企業にも浸透していることを示す事例であったと言える。

私たちは、「匿名化」にもっと注意を払わなければならない。実名（特定）と匿名の間に、仮名（識別）があると考えておくべきだ。本来の匿名からは、実名に至ることはできないはずだが、仮名であれば実名から仮名を作ったルールや対応表を入手することができれば、実名を知ることができる。実名に至るまでの情報の入手難易度や、必要な手数によって強い仮名と弱い仮名があることも指摘できる。こうした事実があるのに、多くのケースで仮名を含めて「匿名」と表現され、匿名であるから情報を提供しても大丈夫という説明がまかり通るのは、安全性の是非を別にしても、アンフェアである。

ブログの記事で言及される情報から自宅の絞り込みが行われ、有望と思われる地点にはスネークと呼ばれる協力者が実際に赴き、写真との照合を行った。自宅前と思われる箇所から掲示板に書き込みをすると、その家の窓が開き、スネークの様子を窺う行動が見られたことから、自宅が特定された。

グーグルは、どちらかといえばプライバシーを公にするサービスを好む企業だが、**使われないサービスは容赦なく廃棄**することでも有名である。鳴り物入りで公開されたグーグルデスクトップや、グーグルノートブックなども利用者数が期待した水準でないとわかると即座に提供を取り止めた。