

パーソナル情報の信託業務における価値創造 ～情報の価値と非公開情報の補完の視点から～

辰巳 憲一*

1 はじめに

プライバシーを保護しながらパーソナル情報を信託会社などの組織で一元管理する、という社会経済的システムに関心が広がっている。情報を提供する個人には、それなりに報酬を得ながら、膨大になる個人情報を手元で管理できる、メリットももたらされる。

該当する組織が行う信託業務には、取り扱う商品が微妙であるため、慎重に、付加価値を創造しなければならないという高度なミッションが課される。本稿では、パーソナル情報信託業務における付加価値創造に関わる様々な問題点を指摘し、いくつか解決策を展開する。個人や組織（企業を含む）には、パーソナル情報の非公開を許しながら、プライバシーや企業機密を侵さずに、何らかの形でパーソナル情報を広く販売する、という観点がハイライトである。

本稿は、プライバシー情報を保護しながら信託で一元管理する、個人情報信託を経済分析した辰巳 [2011b] の続編である。前研究と同様に、本稿でも法学的なアプローチはとらない。なお、本稿では最近の慣例に従い、前稿で用いた個人情報という言葉ではなく、パーソナル情報という表現を用いた。

非公開パーソナル情報を広く受け入れてもらうために、データ・情報の価値の所在を分析した後、その補間・補完の分析に向かう。本研究は、また、情報セキュリティ問題とも強く関係している。情報セキュリティについては、辰巳 [2011a]、辰巳 [2011b]、辰巳 [2011c]、辰巳 [2012a]、辰巳 [2012b]、などを参照。さらには Tatsumi and Goto [2010]、辰巳・後藤 [2010] と Goto and Tatsumi [2012] はリアル・オプション理論を用いた情報セキュリティ投資行動の理論的分析である。

*) 学習院大学経済学部教授。On Enhancing Value Added in Trust Activity of Privacy Information ~An Analysis of Substitution~。内容などの連絡先：〒171-8588豊島区目白1-5-1 学習院大学経済学部、TEL(DI) : 03-5992-4382, Fax : 03-5992-1007, E-mail: Kenichi.Tatsumi © gakushuin.ac.jp (ご送信される場合◎は@に置き換えてご利用ください。)

本稿の前半の情報の価値とベンダー行動については、様々な意見をいただいた内外の金融証券データ・情報分野のベンダー関係者の方々、化学分析情報提供会社の方々、に感謝したい。ならびに後半のパーソナル情報の補完については、様々な研究会での報告や質問に対する返答などから有益なヒントを提供いただいた。講師の各氏に、感謝したい。

2 情報価値とその創造

情報の価値と何か、そしてその創造にまつわる様々な問題を、まず展開しておこう。

2-1 データ・情報

2-1-1 データ・情報の一般的な分類

データ・情報と一括りせず、データと情報を分ける場合、データは数字、情報は（場合によってデータを含み）文字・記事や画像・絵、を指すケースが多いように見受けられる。情報にデータを含んでいると解釈する場合、データは情報の元になる、と表現される¹⁾。情報にデータ（数字）が含まれる代表的な例は経済・金融証券・企業ニュースである。いわゆるベンダーが提供するものは、ふつう、データと情報に分けられる。その結果ベンダーはそれらの内容によって2つに分けられる。会社数からみると、数字のデータだけを有料で提供したり配信する会社の方が断然多い。

最近注目される情報は、個人の行動履歴、趣味ならびに個人間の繋がり、などであり、必ずしも数字（データ）ばかりではない²⁾。しかも、体系化されておらず、非構造化データと呼ばれる。

データ・情報は入手の難易度、作成コスト、などの入手コストによっても分けられる。このような分類をする意義は、付けられる価格は当然それによって変わってくるからである。探偵など、人手によって手数をかけて手に入れた情報は、様々な形になるが、高コストである。

「情報」というものは、人の頭の中にある「記憶」、仕事をすることによって人の身についた「ノウハウ」、記録媒体に保存された形で存在する「電子データ」など、様々な形のものが存在する。これらのうち、前2者である「記憶」と「ノウハウ」の情報は、直接売買できない。人との付き合いや従業員の雇用を通じてのみ入手できる。たとえ元従業員が元の勤め先から持ち出したとしても企業は取り戻すことは絶対できない。

データの作成と一口で言っても、紙に書かれていたり、口頭で述べられた数字をそのままデジタル化するという単純なレベルから、それらから意味のある形で数字を抜き出し（そのためには専門知識が必要になる）、デジタル化するという一段高いレベルまである。作られたデータや獲られた情報をそのままコピーすれば、多くの場合違法ではあるが、極めて低コストで利用可能になる。

データ・情報の分類方法や分類基準は、後述のとおり、これら以外にも、いくつもある。

-
- 1) 「データが情報となるには、目的のために体系化され、具体的な仕事に向けられ、意思決定に使われなければならない（ドラッカー [1995]）」。データと情報の言葉の使い方は本稿とは少し違うが、本稿が意図する内容はドラッカーの意図する内容と同じである。ドラッカーの定義に基づくと、データを利用可能な情報にいかにか転化するかが求められる。情報力とは情報を入手する力ではなく解釈して利用する力である。データや情報を、特に情報を、さらに昇華した使い方がある。「組織の機能や仕事に必要な情報を最も有効に扱うことのできる規模が、組織にとって最適の規模である」。ちなみに、「組織が大きくなれば、中身の大部分は外部環境から遠ざかる。そのため生命の維持に不可欠な栄養素を供給すべき内部機関が複雑になる。こうして規模が複雑さを招く」（いずれもドラッカー [2001]）。
 - 2) シュナイター [2010] は、ソーシャルネットワークで、ユーザーが提供しているデータには6種類がある、と指摘する。本文以下にほぼそのまま引用している。

2-1-2 情報の価値について

情報の価値という言葉でデータや情報の価値を表すものとしよう。

(1) 情報の価値の源

情報の価値は、まず、その希少性が決め手になる。他にない、珍しい、競合する他社（者）は持っていない、などの観点から希少であるかどうかで価値が決まる。顧客や企業の機密情報を合法的に（少しでも多く）獲ることが出来れば、確実に将来を予測することが出来、様々な経済活動から（高い）価値を得ることができる、からである。

もちろん希少性だけで情報の価値が決まるわけではない。ニュース速報のように速報性が問われる情報もあれば、逆に歴史的資料のように信憑性・正確性や現状保存性が重要となる情報もある。

情報の価値は、受け手によって変化する。全く同じ情報であっても、受け手の環境や関心によって必要とされる希少性、速報性や信憑性は変わってくるからである。

(2) 情報価値の構築

収集だけでなく、蓄積それ自体、蓄積方法と「組み合わせ」が情報の価値を生み出す。さまざまな種類の情報を収集し、場合によって誤差を修正し、雑音を排除（クリーニング）して、蓄積し、それらを組み合わせると高い価値のある情報に体系化するという作業がなされて、情報が価値を持つようになる。また、例えば蓄積方法としては、決められた場所に、決められた長さの範囲内で、決められた項目を蓄積していく方法がある。このようにして初めて、それが情報として活用できるようになる。

様々な場所に散逸していると、そもそも情報として体をなさない。そして、このように統一されていないと、再利用できない、様々な主体の間で情報の共有ができない、管理が難しい、などの問題が生じる。また、関係者が情報をそれぞれ勝手に更新したり、その内容がどう変更されたのかを把握できない、ということも往々にして起こる。それゆえ、このように統一化、体系化されていなければ、再利用、情報共有、管理ができない、ということである。

(3) 利用と所有

ベンダーと利用契約しデータを使うことになって、それをダウンロードして誰もが改めて気付き事柄がある。今や、そのデータは自分のファイル・ディスク中に存在し、必要な時にはいつでもエクセルからアクセスし、エクセル上で処理し、終わればまた自分のファイルに戻す。もはや自分のデータになっている、のである。

データ・情報が、常時アクセスできコントロール可能な場所にある限り、利用と所有の両者の間に違いは大きくない。これがデータ・情報の特徴で、利用と所有を区別できない場合がある、のである。データ・情報の所有・保有することとそれを購入することとの違いについては、機能面の境界が不明で曖昧なところがある。それゆえ、データ・情報の使用料・利用料と購入代金との違いについても、会計上の違いは定められていても、実態は曖昧である。

クラウドであっても、これらのアクセス度とコントロール度の条件がある限り、両者の間に違いはない。しかしながら、海外所在のデータセンターが政変でアクセスを止められたり、何らかの要因で新政府によって差し押さえられる可能性が例え0.01%でもあれば、この条件は満たされない。それゆえ、海外へ流れたデータ・情報の場合、利用も所有も出来なくなる可能性がある、のである。

(4) 対応のコスト

ふつうのデータ・情報では金額のしれた再購入費用で済むところ、特殊なデータ・情報についてはそれをはるかに超える費用が利用や保有・購入に際してかかる場合がある。それらは、唯一無二でしかも復元不能のデータ・情報や個人情報データに関するもので、それらを破損、漏洩など、盗難されるなどした場合である。

例えば、プライバシーを侵すなどの事件事故が起こった時は、取り返しの付かないことであり、その処理費用は膨大になる。しかも、事後的には該当の金額を補償のために支払うことになるが、従来、事前にこれらに見合う予算を立てることは大変難しかった。将来的には、この費用も保険料のような形で、計上することになる。また、その金額はデータ・情報の価値のなかに組み込まれるべきである。

2-1-3 データ・情報の売買取引

(1) データ・情報の供給者

データ・情報仲介業者いわゆるベンダーを通じて取引されるのは、通常、公開可能なデータ・情報であって公開されているもの、あるいは、元々は公開できないが、官庁が収集しそれらを何らかの形で集計して公開するデータ・情報である。

ユーザーが利用したいデータ・情報によって、データ・情報ベンダー（提供者）が異なる場合が多い。それゆえ、ユーザーは複数のベンダーと契約して提供を受けなくてはならないことが多くなっている。

ほぼ同じデータ・情報であっても複数のソースから提供される場合がある。例えば、取引所の価格等のデータであれば、顧客が直接回線を接続してデータ・情報の保有者（例えば取引所）から直接取得する場合、と（取引所と契約した者である）データ・情報ベンダー等から情報を間接的に取得する場合の2つのケースがある。このような二次的なデータ・情報ベンダーが存続しえる1つの理由は、それらベンダーが取引所のデータ（価格と取引量のデータ）だけでなく、他のデータ・情報（例えば、経済データ）も同時に提供し、ユーザーの利便性が高まっているからである。同時に提供されるものが、ソフトを動かせるプラットフォーム、文章情報、というケースもある。

(2) データ・情報の提供媒体

ユーザーがデータ・情報を取得する手段には、回線などを直接接続してオンラインで取得する（手段としてはインターネットや電子メール、機器としてはPC 端末やモバイル端末）以外に、紙媒体、CD-ROM、USB 等可搬記録媒体など様々にある。手段が異なれば、料金も異なる。

電子出版では、紙や印刷のコスト、物流費（低廉であるが通信費はかかる）がかからない。それゆえ、値下がりの傾向にある。しかしながら、冊子体の雑誌が電子ジャーナルになることで、購読価格が上昇した例がある。これは、初期投資が価格に上乘せされたからである。

(3) 売買・提供されるデータ・情報の一般的質

データ・情報にはその利用に関する制限や権利関係の問題³⁾がある。例えば、あるデータ・情報をベンダーから直接取得する場合と第三者から二次取得する場合とではいくつかの観点か

3) 取得したデータ・情報をそのまま、または加工したものも含め、自社内以外で利用することは、契約に基づく場合以外、禁じられる（二次利用の禁止）。また、取得した情報は、自社内以外の第三者へ開示または漏洩することは、契約に基づく場合以外、禁じられる（秘密の保持）。

ら情報の質が異なる。例えば取引所データについては、リアルタイム情報が提供されるケース（即時性）もあるが、東証、NYSE 及び AMEX などの価格データは実際の取引から最低20分遅れで表示され、その他のデータでも15分遅れで表示される。

マスコミ等を通じて一般向けに公開されるデータ・情報であっても、契約したユーザー（情報会員と呼ばれる）が取得できる情報では、データ・情報の量（網羅性）および精度に差がある。例えば粒度においては、値の平均値化や中位値化がされている、あるいは端数の処理の仕方が違う、など正確性が違うのである。

網羅性や正確性で他に優れた情報ソースが存在する場合があるのも事実であるが、データ・情報ベンダーから購入するデータは一般にまず自社システムへの接続性（そのように設定されるので）に優れている。そして第二に即時性の利点がある。

ちなみに、データベースがいくつかあるといっても、このようにデータの質が異なるので、価格を一律に比較することは困難になる。

2-1-4 データ・情報ベンダーの経済学的特性

（1）商品の性質と産業構造

利用したいデータ・情報によって、データ・情報を提供するベンダーや当該情報を取得する手段が異なるため、料金も異なる。また、ベンダーごとに提供するデータ・情報の精度が異なり、サービスやその価格にばらつきがある。そのため、このようなベンダー・サービスに関しては取引される商品・サービスが規格化されていない、と言える。そして、商品・サービスを相互に比較することは極めて困難になっている。

データ・エクスポートが可能となっている場合、ベンダーは他から購入したデータに自身のデータを組み合わせて更に新しい価値あるデータを作ることも可能となる。このような加工販売がデータの価値を高める1つの要素となる。

それゆえ、商品・サービスの境界が定義できないのだから、市場に範囲は厳密には存在せず、それらの価格である「市場価格」が形成されていない、と認識されることさえある。市場は広めに捉えた方が無難であり、価格競争が十分できない産業である、ことは事実であると筆者は考えている。

（2）コスト構造と産業構造

作られたデータは、そのままコピーしたり、他のデータと組み合わせて更に新しいデータを作ることもできる。この場合当初に作成されたデータのコストはとてども低くなる。これを、経済学的には、固定費用的要素が強いと表現する。このように、固定費用比率が高い、特殊な財・サービスを提供する、このデータ・情報提供産業は独特の産業構造と企業行動を生じさせることになった。

シャピロ&バリアン [1999a] は、このような業種の姿を次のように描いている。「情報財の最大のメリットは、複製が容易でコストがかからないため、規模の経済が効きやすいことである。しかし、情報財は固定費が非常に高く、競合他社が同一市場に多数参入した場合の価格競争に耐えきれなくなってしまう。」ここで、情報財とは、デジタル化された製品、ウェブ上のサービスなど、を指す。

規模の経済が強い一般の産業では寡占化は必至である。しかしながら、情報関連産業では、違っており、多様な構造をしている。

データ・情報は、物理的な財と違って、ベンダーが売却しても（消去などをしなければ）手

元に残しておく（手元に残っている）ので、ベンダーは何度でも売る、ということができる。これを多重利用効果と呼んでおこう。当初固定費用は膨大でも、既述のように、規模が大きくなれば平均固定費用を無視できる程小さくできる。それだけでなく、追加生産費用は極めて低いのである。この多重利用効果は利点であり、同時に（汎用品は）容易に悪用される欠点を内包している。

（3）公共性

データ・情報ベンダーには、一人あるいは一社でも顧客が利用したい希望を持ち続けるのであれば、少なくとも契約期間が切れるまで、場合によっては顧客が望むならばそれを超えてデータ・情報を供給し（少なくとも質問に答え）ていかなければならないという義務と責任が生じる。メーカーが製造を停止した機器のメンテを行わなければならないのと同様である。データ作成企業としての社会的使命といってもよい。サービス提供の継続性という一種の公共性なのである。この公共性をまったく無視すれば、企業の存続は望めないから、ある程度従うしかない。経済的要因があつて、提供できない、拒否せざるをえない事態が生じた場合にはベンダーは顧客との交渉が必要になる。

2-2 データ・情報提供業の課題

（1）ユーザーとベンダーの関係～利益相反

価格や取引データを提供するベンダー自身あるいはその親会社・関連会社が、提供するデータ・情報の源である価格付けや取引仲介業務を直接行なっている場合、利益相反が起こる可能性がある。公表する情報が取引者の利益を害する可能性があるからである。

そのため、このような場合、ベンダーは約定（均衡）価格や総取引量は公表するが、個別ケースの詳細は公表しない。気配と取引量を公表する場合は取引者の名前を公表しない、などの配慮⁴⁾がされる。これらの匿名化が、どれ程の効果があるかどうか（秘密はバレていないか）、他にどのような利益相反回避策があるか、などを調査した研究はないようである。

ちなみに、この観点から見ると、格付け会社については、根本的な利害の衝突を孕んでビジネスしている特殊な業種である。

（2）バージョン戦略

シャピロ&バリアン [1999a] は、前節での引用文に続けて、このような業種で個別企業はどう生き残る（べき）かを次のように描いている。「唯一の戦略は、顧客価値にリンクさせて、価格設定を多様化させていくことだ。付加価値の違いによって製品を差別化する、バージョン戦略が鍵となる。」

バージョンング (versioning)⁵⁾とは、提供する商品やサービスを、提供する時期、機能や品質、性能などが異なる複数のバージョンに分けて提示し、ユーザーに選択してもらおう売買手法のことである。

4) 例えば、M&A 仲介会社と M&A を含むデータの提供を業務とする会社が、同一グループに所属する場合は、「M&A は、当事者企業が、血の滲むような、熟慮と努力を重ねた結果成功に至るものであり、仲介企業は、一切表にでることなく、黒子に徹すべきものである」との方針で、グループ企業が仲介して成功に至った M&A デールについては、データ提供会社はどこが仲介したと言う情報は、一切公表をしないようにしている。

5) バージョニング戦略を展開する文献については、他に、吉川 [2002] などがある。

バージョンングで重要なのは、いかにして顧客により上位の高価メニューを選択してもらうかにある。それゆえ、コストが多少余分にかかっても、あえて戦略的に廉価メニューを用意して、高価メニューを相対的によく見せる方法も採られる。

（３）パーソナル情報の価値計算

特段有名でもない、普通の人の住所・氏名・生年月日・携帯電話番号などの個人情報、どれほど価値のあるものなのか。またそれが流出・漏洩することのマイナスはどれほどの大きいものなのだろうか。クレジットカード番号情報の価値は理解できる。また、流出元に関する情報がもう少しあれば、個人資産の額を予想できるようになり、流出・漏洩情報に価値が出てくる。

1つ1つは価値がない些細な、何げない、パーソナル情報も、集合体としては価値を持つ場合が多々あることが現在では認識され、広く分析されている。ビッグデータと呼ばれている対象が、ほとんどがパーソナル情報であることは周知なことである。

現代のビジネス社会では、企業が何らかの原因で個人情報、機微な情報を漏洩してしまった場合の金銭的な損害賠償金額を適正に計算する方法が求められている。それを漏洩した企業が単独で計算することは益々困難になると筆者は予想している。アウトソーシングが普通になるのではなかろうか。

（４）カスタマイズされたデータ・情報とコンサル

一般化された、汎用のデータ・情報ではなく、ユーザー企業毎にカスタマイズされた、換言すればパーソナライズされたデータ・情報が今後様々な局面で望まれるようになる、と筆者は予想している。例えば、企業は、抽象的な消費者の情報ではなく、「我が店舗に足を運ばれるお客様の情報が欲しい」のである。ユーザー企業はどのようなデータ・情報を望むのか、カスタマイズの次元は、名前、性、年齢、体格や所得、資産規模という伝統的なものだけでなく、気質、好み、将来に係わる変数、なども入れてデータ・情報の販売を行わねばならないであろう。データ・情報産業は、その結果、コンサル業と融合するようになる、と筆者は予想している。

3 パーソナル情報の活用と補完

ユーザー企業・組織に対して、必要ならば素あるいは元のデータを提供し、目的外の利用はないかを確認したり、漏洩防止策を確立することはデータ・情報提供者にとって重要な業務である。しかしながら、これらの業務だけでは、ビジネスにはならない（多くの場合赤字である）、ことをこれまで詳しく説明してきた。筆者は、パーソナル情報提供者にとっても同様である、と言いたい。次に、さらに具体的な付加価値創造業務の中身に立ち入っていくことにしよう。

3-1 パーソナル情報の匿名化

3-1-1 パーソナル情報の活用

（１）パーソナル情報活用の二律背反

交通・通勤渋滞回避や都市計画立案、医療技術進歩、といった現代の重大課題の解決に導く、パーソナル情報の活用を広げることは、結果として個々の個人さらに広くは社会に役立つ、のは事実である。あるいは、ビジネス・チャンスを生み、経済社会を活性化し、ひいては日本産

業の国際競争力を増強できると期待できる。

他方、パーソナル情報が広く公表されるようになると、多くの人々がマスコミなどの情報操作に踊らされることはなくなるが、プライバシー、人権が侵される恐れとモラルのない企業や政権による情報管理が強まることが懸念される。それゆえ、これらパーソナル情報の活用には二律背反関係がある。この二律背反は今や広く知られている。

(2) 匿名化

この背反する問題について技術的な側面から解決を探るのが、個人を特定できないように一定の条件で識別または分類するデータの匿名化技術である。

匿名化は、例えば年齢が19歳の人を10～20歳というような範囲のデータとしてまとめることで個人を特定しにくくするものである。企業であれば、例えば所属業種をフィルムメーカー、さらには化学産業、またさらには製造業というようにより広い範囲に所属させる。どの程度範囲を広めるか、その程度に応じて、k-匿名化などという記号におけるkが使われる。

匿名化技術としては、トップダウン、ボトムアップ、近似解、ヒューリスティクス、など多くの(効率的)アルゴリズムや新しい概念が提案されている。展望や文献などは、南 [2013]などを参照。

3-1-2 匿名化の課題と限界

匿名化にはいくつかの課題と限界がある。本稿との係わりで、いくつか紹介しておこう。

(1) 匿名化とアクセス制御

匿名化は、誰(パーソナル情報の販売先)にどの程度(粒度、頻度など)情報を開示するかを規定するアクセス制御に合わせて、行われなければならないだろう。アクセス度が高い権利者には、匿名化の程度が低いデータを提供できる。その結果、アクセス権の設定は重要な作業になる。

その極端な事例は、本人である。パーソナル情報保有者本人は、完全アクセス可能で、非匿名データを知る権利、その修正(もし誤っているならば)を(該当の委員会へ)申請する権利を持っているべきである。そもそも、民主主義社会において基本的な事柄は、パーソナル情報は本人の同意なしに、そのまま第三者に提供してはならないという点であろう。

(2) 匿名化データ相互の関係

匿名化されたデータであっても、外部知識と組み合わせれば、個人の特特定が可能になる場合がある。住所データや性別、あるいはその他行動履歴を利用すれば個人を特定できてしまう。さらに、同様に、複数の匿名化データを組み合わせれば、個人を特定できる場合もある。「個人が特定できる情報と特定できない情報との境界は、もはやない」という認識は確立されてきたということである。

暗号解読度という視点と同様に、復元可能性という視点が重要視されて、復元可能性のより低い技術の発展がなされなければならないということである。

(3) 匿名化と情報損失

匿名化という技術は、パーソナル情報を様々に利活用することを前提に、研究されている。そのため、情報損失という観点が取られ、その程度が議論される。

得られている知見には、準識別情報(詳しくは後述)の内容が似た同士を結合するような匿名化をすれば、情報損失が小さくなる。外れ値がある場合、その影響から、情報損失が大きく、匿名化後のデータの有用性が大幅に減少する、など幾つか重要な点がある。

他方で、現状の匿名化処理はデータの有用性を十分生かしていない、という意見も根強い。

3-2 非公開パーソナル情報の補完

様々あるアンケート調査において、ある共通の特徴が観察されている。例えば、年収に関する質問では、ある水準以下の年収の人はその項目に答えたくないと同欄のまま提出する傾向がある、ことが知られている。未婚か既婚かを聞く質問では、特定の年齢層の人が答えてくれなかったりする。誰でも公開したくない情報があるものなのである。

3-2-1 匿名化、情報公開拒否と補完

匿名化は、筆者の見るところ、どうも議論の出発点として、パーソナル情報あるいはプライバシー情報の強制的な公開が前提になっているように見受けられる。しかる後、その情報をどう「ぼかす」かが科学的な課題となっているのではないかと思う。

筆者の考えでは、パーソナル情報あるいはプライバシー情報は、その保有者が公開するか非公開にするかどうか、自身の判断で、選択できる制度作りの方が望ましい、と思う。現在、匿名化技術の発展に研究努力が重ねられている。それはそれで無駄ではないが、進むべき方向は違っているように筆者は思えてならない。

公開するかどうかを情報保有者自身が選択できる非公開選択制度（非公開選択制度ではなく「非」を取った公開選択制度と呼んでもよいかもしれないが、「非公開にできる」のどという点を強調するため、敢えて非公開選択制度という名称に拘った。）は、制度的に大変好ましい。その理由は、まず第一に、自由主義に適した制度であること。第二に、公開することを強制すれば人々に嘘をつかせたり、極秘行動を生む、あるいは（プライバシーを守るために）わざわざ回避行動をとらざるえないことになってしまう、大変住みにくい社会を作り出す恐れがある、からである。

(1) 情報収集拒否さえも可能に

現在、公開を拒否するだけでなく、収集段階から拒否することを認める動きにある。収集お断りは公開お断りの前段階と解釈して当然である。

米国連邦取引委員会（FTC）は米国時間2010年12月2日、プライバシー保護対策の新たな提案に盛り込んだ個人情報管理手段「Do Not Track」について説明した。これはオンライン行動の追跡を消費者が拒否できるようにする手段で、複数のWebブラウザに通用する仕組みであるという。

FTCは、ユーザーに適したコンテンツやサービスを効率的に提供できるという点で行動ターゲティング広告の価値を認めているが、その手法は透明性が不十分で、プライバシー侵害の懸念があると指摘する。そのため、消費者が行動ターゲティング広告の配信を拒否したり、特定データのみの収集を許可したりできるシンプルかつ使いやすい手段が必要だと、米国議会に対して説明した（以上、2つのパラグラフは報道記事から引用）。

日本でも、周知のように、既に、いくつかのサイトで同様な仕組みが取り入れられている。

(2) 非公開選択制度の確立

どんな事柄をどんなデバイスを使って、どう発言するのが基準となって、誰がプライバシー保護を得て、誰がプライバシー保護を得ないのかが決まるとすれば、民主主義国家の社会システムとしては不備である。声の大きい人だけがプライバシーを守られることは避けるべきである。それゆえ、明文化されたプライバシー情報非公開選択制度を確立することが重要になる。

当然例外はある。特に、公共の場に向けられた監視カメラで撮られた情報は公開するか非公開にするかの選択の対象になるべきではないだろう。むやみに無意味な公開に対して個人は抗議すべきだが、犯罪や社会道徳に反する事例は、非公開の選択対象になるべきではない。

自分のプライバシー情報に対する価値に対して、個人は正しい認識や感覚を持っていない事実がよく観測され指摘され報告される。そうだからといって、デタラメな制度作りとプライシング（料金設定）をするべきではない。むしろ、そのような人々を啓蒙しながら、パーソナル情報信託業務は事業拡大に取り組むべきである。

情報の利活用とプライバシーを両立させるためには、非公開選択制度しかない。日本でも、その方向に進むのではないかと、筆者は予想している。経済産業省は、2013年5月企業が顧客の個人情報をも2次利用するための指針をまとめ、購入履歴や性別などのうち消費者が同意した情報に限り、他の企業への販売などを認める、と報道されている。消費者が選択するというプライバシーに配慮した指針を策定して、企業の個人情報ビッグデータ利活用を後押しし、顧客データの分析を通じた市場開拓を促すのである。

しかしながら、消費者庁は個人情報保護を巡る規制緩和には慎重であるため、先行きはまだ不確かである、とも見込まれている。（いずれも「個人情報、企業利用に指針 開示、消費者が選択」2013年5月10日付日経新聞記事参照。）この点に関して、筆者が先にも強調したポイントを繰り返しておきたい。つまり、公開選択制度という名称では、規制緩和と解釈されてしまうのではないかと思う。公開選択制度ではなく、非公開選択制度という名称を使えば、プライバシーは選択すれば非公開にできる制度という点が強調できるものと思う。

（3）非公開の意味

情報通信のファイバー網からデータを傍受していた事実が暴露され発覚した事件が米国で起こり、「サイバー空間にプライバシーはない」と、恐怖を持って、理解されるようになっていく。

この事件の要点は、何か特定のターゲットを定めて盗聴するよりも、一切切っすべてを記録（監視）することの方が、技術とデバイスの飛躍的な進歩により、簡単になってしまった、結果にすぎない。監視（盗聴）され記録されたが閲覧されず分析されずに眠っているデータは無限にあることになる（報道記事参照）。

どうせ眠るだけだとしても我々は安心してはおれない。万一の（自分は犯罪者になってしまう）場合、当局はそんなに苦勞をしなくても、技術進歩により、短時間で検索でき、情報は取り出せる、ようになるかもしれないからである。

このような時代に、自身の情報を非公開することに実質的な意味があるのか、と疑問に思う人がいても当然である。しかしながら、「私はこの情報を非公開にした」と「その情報を誰か知っている人・組織がいる」とは捉え方は違う。これら2つの命題は、筆者は、矛盾しないと思っている。特に、知っているのが国である場合、国を信頼しておれば、広く拡散しないだろうと思ってしまう。国の機関は知っているが、他の人は知らないという状態を「非公開」と考える人は多いのではないだろうか。そもそも、国・政権は我々が選挙で選んだのだから過半の人は信頼していると見て良いだろう。本稿もそういう考え方を取りたい。

（4）事業創造

非公開情報が多くなりすぎると、データとして有効に活用できなくなると恐れをもつ人がいるかもしれない。そこで、情報の価値を損なわず非公開情報を「公開情報のように活用させる」技が望まれる。しかも、プライバシー情報を保護しながら、行われなければならない。

一般に、事業を興隆させるためには、販売する商品を増やす、あるいは・さらには、販売する市場を増やす、ことが必須になる。この観点からは、非公開のデータを含んだままでは、データは売れない。不完全データを、適切に補間・補完して、データを売る必要がある。

さらに、データは、生データ、素データ、のままでは誰も買ってくれない。データの利用方法は、カスタマイズされる方向に向かっている。カスタム・データは、技術的に可能な範囲で、データ利用者が望む粒度と頻度で提供されなければならない。

3-2-2 情報非公開の選択

3-2-2-1 公開拒否の統計学的分析

(1) 非公開率

先の例を続ければ、年収を問われて答えられない人も、自分が興味のある、自慢できる項目には答えるかもしれない。財務諸表が公表されない小企業も、業況判断、金融機関の融資態度など、答えてくれる項目はある。非公開率は、変数によって、大きく異なるということである。従来非公開であったが、事態が変化し、公開されるようになったという場合もあろう。

非公開率とは情報を非公開にすることを選択した人・組織の数が全体に占める比率である。欠損値発生メカニズムを明らかにして欠損値の補間に役立てようとするのと同じように、非公開にされる選択の理由を解明することは非公開値の補間・補完にとって重要になってくる。

(2) 非公開原因の大分類

欠損値発生メカニズムの研究（例えば Little and Rubin [2002] 参照。彼らはその他の分類や統計分析技法の解説も行っている。新しいところでは他に Ton, Jeroen and Sander [2010] などの研究もある。）を参考に分類すると、非公開は大きく分けて3つの要因で起こる。

1つ目は、SCAR (secrecy completely at random) と呼べる、非公開が完全にランダムに決定されているケース。

2つ目は、SAR (secrecy at random) と呼べる、非公開するかどうかは他の変数が誘引となって決断される（未婚か既婚か聞く質問では特定の年齢層が答えてくれない）ケース。

3つ目は、NSAR (not secrecy at random) と呼べる、非公開するかどうかは非公開した当該変数の値に依存する（年収の例では、非公開が年収自体の水準の値に依存する）というもので、無視できないケースであり、欠損値研究では取り扱いが非常に難しいとされている。

これらは、概念を整理し要因を大別し統計分析に役立てようとしただけなので、さらに深い非公開の分析が望まれる。

3-2-2-2 公開拒否の経済分析

(1) 公開拒否の経済学

公開を拒否する理由を社会科学的に分析できないだろうか。組織の利益や社会的価値の評価、などによって非公開が決意される、と考えられる。犯罪行為であることが明瞭である場合、社会的モラルに反する行為である場合、競争相手を利する場合、もそうである。しかしながら前二者は、普通、経済分析からは省かれる。

情報の経済学によって分析できる分野も存在するように思われる。自己以外が当該情報を保有するか保有しないかで、個人・組織の最適化の結果が異なれば、それらを比較し、効用・利益水準の高い方を選ぶことによって、公開の可否を選択することが可能になる。これが分析方法のエッセンスである。

(2) 公開拒否理由と個人・組織の属性の間の独立性

公開拒否の理由が、個人・組織の属性と独立であるか否かが、分析上まず重要になる。まったく独立であることは希であるかもしれない。一般には、個人・組織の特性、諸変数、時期などに依存するものと考えられる。

公開拒否の理由が個人・組織の属性と独立でないケースとしては、いくつか考えられる。例えば小規模企業は、財務諸表を持続的に作成できず決算数字に欠損が発生する傾向がみられる、などが知られている。この欠損は非公開という形でも公表される場合もある。その理由は、財務に詳しい人材の不在等などと言いつけられているが、融資や投資に不利にならないように、非公開にされるという理由もあるのではないかと予想される。

機微情報については、個人が劣等感を持っている事柄を非公開にする選択をしてしまうことは多いにある。しかしながら、人によって劣等感を感じるレベルは異なる。

欠損値の最適な補間方法は時期や変数によって、異なるべきである、という結論が得られている（後掲の引用文献参照）ようである。それは、翻って、時期や変数によって、公開拒否の理由が異なるようになるからである。時期に応じて公開拒否になる例として、就活時や結婚前には、1つ位他言できない事柄があるものであることから想像できよう。

その他の特性として考えられるのは、公開拒否要因は相関する、という点である。飲酒する人は煙草も吸う、年収の低い人は未婚である、などいくつか具体例が存在する。

3-2-3 補間あるいは補完とデータ・情報の分類

3-2-3-1 補間あるいは補完

(1) 補間あるいは補完とは

分析者が分析するデータには、様々な理由で、データ数値が一部分存在しないということが頻繁に起こる。この欠損値を合理的な方法で補って、(統計)分析するために、完全なデータにすることを補間 (interpolation) あるいは補完 (substitution) という。

データには、周知のように、クロスセクション・データと時系列データがある。両者の性質が異なるので、クロスセクション・データと時系列データの補間は違って当然である。

(2) クロスセクション・データの補間

人間の顔についての断片情報であれば、顔の構造分析で蓄積された知見が補間を助けるように、経済データであれば、実証された経済モデルが補間を助ける。時間軸でのデータ補間には、動学モデルが必要になるが、経済学分野ではそれが著しく発達しているとは言えない。ところが、クロスセクション的には多くの経済モデルがあり、多くが検証されている。

辰巳・松葉 [2008] では、時系列データにおける補間方法の分析と考察を行った。宇都宮・園田 [2001] は、様々なクロスセクション補間方法を展望し、特定の応用例に関する補間方法の優劣⁶⁾についても論じている。さらに、現行の官庁統計で取られている補間方法を説明して

6) 宇都宮・園田 [2001] は、短観 (全国企業短期経済観測調査) の代表的な計数項目である設備投資、売上高、経常損益について、(1) 現行の方法 (平均値補完, mean imputation), (2) 回答のある直近調査回の値の代入 (横置き補完と呼ばれる), (3) 前回調査回の値に層内の回答者の前回調査比 (3月調査時における計画の場合は、初回調査であることから前年比) 伸び率を乗じた値の代入 (伸び率補完) の3つを選択肢として比較検討したところ、現行の平均値補完よりも、その他の手法の方が精度は高いとのシミュレーション結果が得られた。また、横置き補完と伸び率補完を比較すると、総じてみれば、横置き補完の精度が高いことを示唆していると判断された。

いる。平川・鳩貝 [2012] はその研究を引き継ぎ、さらに、欠測項目以外のデータを用いて補完する手法⁷⁾や欠損率の活用を検討する。

なお、幾つかの基本的な補間法はデータ処理のパッケージ・ソフト（エクセルなど）に組み込まれており、容易に利用できるようになっている。

3-2-3-2 補間あるいは補完するデータの種類

個票ではなく、それらを適切にある程度集計したデータ・情報は、小集計、セルあるいは層と呼ばれる。本稿では、以降、層という名称で統一しよう。公開拒否の理由だけでなく、個票か層であるかによって、広くはデータの種類によって、補間の方法は異なってしかるべきである。

(1) パーソナルなデータ・情報の種類

パーソナル情報に含まれる諸属性は識別情報、準識別情報、その他情報の3つのカテゴリーに分類される。この分類法は公的な使い方になりつつあり、具体的な項目例はよく引用される政府答申・報告書（例えば『パーソナルデータの利用・流通に関する研究会報告書』http://www.soumu.go.jp/main_content/000231357.pdf）から転載した図表1に掲げた。静的準識別情報の中に入れられる、生年月日、性別（最近は第三と第四の性もあるが）は、時間と共に変化するという性質は存在しない。

データのとる値に注目する分類もある。男女、陰性陽性（病気の）、などのような2値データ、あるいは職業のようなカテゴリーの多値データ、などもある。頻度分布でみた特性では、最近、外れ値あるいはファットテイルの多いデータ、に関心が寄せられている。

(2) データの粒度

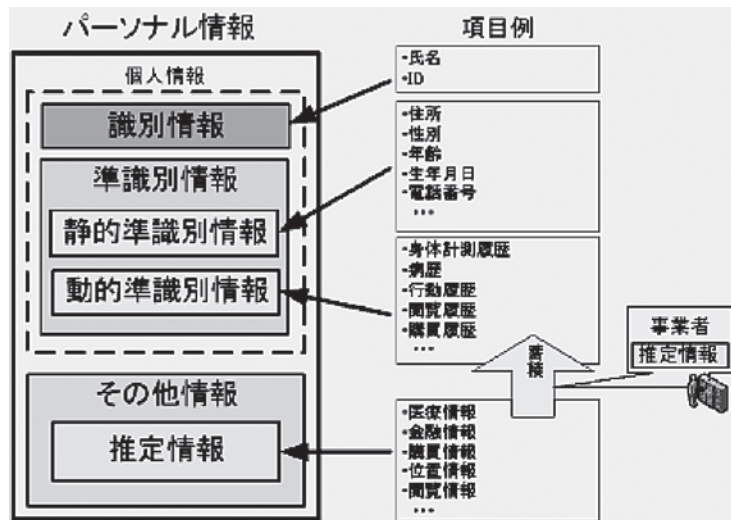
個票と層（セルあるいは小集計）の区別は重要である。層の例としては、資産規模別、資本金別、所得階層別あるいは税額別、年齢（創設後年数）別、などの分類がある。これらは大小があり、順番を付けて並べることができる。そして一般に、組織や人は小から大へ時間とともに動いていく。粒度がある、その他の分類例としては、先に一例を挙げた業種分類以外に、地域別、などがある。

層では、複数のデータが層内に既に存在していることが前提になっており、1つあるいはいくつかが欠損しても、層内の他のデータから何らかの代替値、推定値が得られることが多い。このことが大きな特徴になっている。

ここで1点だけ説明しておきたい。粒度を粗くしていくという手順は、匿名化を進めることに似ている。しかしながら、匿名化の理論は現時点ではまだ抽象論であり、具体的な議論に立ち至れば、両者の議論は類似してこよう。研究は相互に補完できるのではないかとみられる。

7) 欠測項目以外のデータを用いて補完する手法として平川・鳩貝 [2012] は、さらに、項目間比率補完を検討する。この手法は、欠測項目と他の項目との比率（項目間比率）を、欠測項目と対応する他の調査項目に乗じることで、欠測値データを算出する手法である。

平川・鳩貝 [2012] は、宇都宮・園田 [2001] で示唆されていた通り、経済変動が大きい時期では伸び率補完と横置き補完の精度の差異は大きい、こと。そして、時期や変数によって、最適な補完方法は異なる、という結論を得ている。



図表1 パーソナル情報分類

(3) コントロールの可否によるデータ・情報の分類

データ・情報の発信元、製造元、収集元がコントロール可能かどうかによって分けることもできる。コントロール可能とは、知られたいくなければ、自身が慎重でありさえすればよい、という意味である。あるいは、取り消しができるという意味でもある。

コントロールできるデータ・情報としては、

(a) 自主的に公開するデータ・情報、

がある。合理的な人間を前提にすればこれらの情報は公開することが決定されたものである。

コントロールできないデータ・情報としては、次の5つがある。

(b) 公開が必然的に要求される（住民票などの登記、入会などの登録、税制）データ・情報。これは、書き入れなければ、法律に反する、入会できない、から人は公開するに過ぎない。入会することが必然でなく入会を当初からやめれば、公開しなくてよい。

(c) 自身の医療、行動履歴のデータ・情報、
 ここには、病歴、購買履歴、金融証券取引履歴、なども入る。

(d) コントロールできない場所（例えば他者のブログ）に書き込み公開したデータ・情報。

(e) 他者が述べる（自身に対する意見などの）評価データ・情報、

この情報も、真偽にかかわらず、意思に反して一部あるいは全部が広く世間に流布・拡散してしまう。

(f) 自身の社会的関係のデータ・情報、

どれ位自分の意思でどれ位深くお付き合いしているかは第三者に分からない、ところがある。第三者からは、仲良しとみられていても、本人達は相互にそう考えていない、場合も多いのではないかと思う。それが、仲良しとみられてしまうという不本意がコントロールできないという意味である。

図表2には、以上の分類を表にした。シュナイアー [2010] は、ソーシャルネットワークを

事例に、ユーザーが提供しているデータを6種類に分けている。その分類と対応できるように、説明文をほぼそのまま、しかしながら順番は変更して、引用している。

本稿で分析対象になるのは、(a) から (c) のデータ・情報である。(d) から (f) については、間違っていることに気付いたら修正するよう説得するしかない。あるいは、非公開にしてもらうようお願いするしかない。

図表2 パーソナル情報の分類

		本文での分類	ソーシャルネットワークでユーザーが提供しているデータの分類（シュナイアー [2010] 参照）
コントロール困難	公開拒否可能	(a) 自主的に公開するデータ・情報	公開データ (disclosed data)：ブログや写真、メッセージ、コメントなど、自分のページに掲載するデータのこと。
		(b) 公開が必然的に要求される（住民票などの登記、入会などの登録、税制）データ・情報	サービス・データ (service data)：これはサイトに登録する際に入力するデータで、本名、年齢、クレジットカード番号などがあるだろう。
		(c) 自身の医療、行動履歴のデータ・情報	行動データ (behavioral data)：自分が何を、誰とするとした習慣をそのサイトが集めるもの。オンラインゲームをすとか、何かのトピックについて書くことが多いとか、どんなニュース記事にアクセスするかといったこと。
	公開拒否不可能	(d) コントロールできない場所（例えば他者のブログ）に書き込み公開したデータ・情報	委託データ (entrusted data)：他人のページに投稿するデータ。上の公開データと同じようなものだが、違いは自分でコントロールができないことだ。コントロール権は相手にある。
		(e) 他者が述べる（自身に対する）意見や評価のデータ・情報	付随データ (incidental data)：他人が自分について掲載、投稿するデータ。ブログや写真など内容は上と同じだが、これについても自分にコントロール権はなく、そもそもデータを生み出したのも自分ではない。
		(f) 自身の社会的関係のデータ・情報	派生データ (derived data)：他のデータから由来して自分を特定するデータ。たとえば、あなたの友達の80%がゲイならば、あなた自身もゲイである可能性が強いとみなされるといったようなこと。

(4) 補間・補完にあたっての課題

以上の考察に基づいて、非公開情報の補完にどのような基本的な基準・課題が存在するか、次節以降順次みていくことにしよう。その前に、いくつか付随的な説明を行っておきたい。

データが「非公開である」という事実から、隠された情報の推論が可能な場合が存在する。特に、2値変数が該当する場合は深刻である。それゆえ、非公開のまま、データを外部に提供するのは問題なのである。

ノイズをデータに追加して、ノイズ+データを提供することは好ましくない。つまり、これらは何ら意味の無い偽のデータなのである。データ利用者にとって、こういう手法は役立たない。

非公開決定に影響する他要因をできるだけコントロールするのが望ましい。具体的に述べれば、公開拒否理由と個人属性を独立にしてしまう補間・補完方法が好ましい。

3-2-4 非公開データの補間・補完

3-2-4-1 非公開データの補間・補完方法の検討

データの補間・補完方法には、大きく分けて次の3つがある。平均値補完 (mean imputation) 法、近傍値代入法、そして代替値代入法 (substitution method) である。順に説明していこう。

(1) 平均値補完法

平均値補完法は、層 (セルあるいは小集計) 内の非公開データの補間に限られるが、層内の非公開値に代わって、公開されたデータから計算された層内平均値を代入する。データの特性によっては、平均値に代えて、中央値が用いられるべきであろう。

個票の補間に応用する場合には、事後的に層化を行い、しかる後に計算した層内平均値を代入する事後層化調整 (post-stratification adjustment) 法がある。実際の欠損値補完では、この方法が適用されるケースも多い。

しかしながら、いくつか懸念される点が存在する。まず統計処理上の問題として、平均値補完法では、層内のデータの分散は過小になる、点がある。

当該平均値が非公開値の不偏推定値である場合、プライシーが侵されることになるのであろうか。代入値が不偏推定値である事実が公開されない限り、プライシーが侵されないことはない、と思われる。

(2) 近傍値代入法

近傍値代入法は、個票でも層でも共通な技法であり、公開者のなかから非公開者の属性に最も近い者を抽出し、その該当数値 (だけ) を非公開データに代入する方法である。

属性の最も近い者という観点については、属性は多次元であるため、いろいろ多様な近傍概念を用いることができる。目的やデータの特性に応じて、多数のなかから選べるメリットがある。

層内標本分散に関しては、その過小評価が、平均値補完法に比べて、是正される。しかしながら、真の値に比べれば、依然として過小評価である。

公開拒否の理由が個人の属性と独立でない時、この代入方法では、確実に偏りのある結果が得られてしまう。さらに、関連する極めて深刻な問題が存在する。平均値は、抽象的な概念であり、特定の人間を意識することは少ない。しかしながら、近傍値とは似た人のことなので数字には人間臭さが残ってくる恐れがある。非公開を決意した人は、該当値について、思わぬ人の思わぬ値が自分の非公開値に代わって代入されていることを知ることにできれば、それが気に入らないということも起こりえる。

(3) 代替値代入法

代替値代入法は、該当する変数以外のデータを利用して回帰分析等の適切な方法で推定値を求め、これを代入する方法である。

なお、他の調査データではなく、当該データの過去の実現値 (もし存在すれば) を用いた時系列分析することも考えられる。過去のデータを代入する手法はヒストリカル・データ代入法 (historical data substitution method) とも呼ばれる。

代替値代入法を使えば、個票レベルでデータが再現される。しかしながら、結果は計測技法の特性、実施方法の適否の判断に大きく依存する。

欠損値補完では、具体的に次のような方法も取られる。本研究で用いている用語で記述して

非公開値補完に応用できるように表現すると、最近時の公開値が存在していることを前提に、データのある直近公開時の実現値を代入（横置き補完と呼ばれる）、あるいは前回公開の値に層内公開者の前回比伸び率平均値を乗じた値を代入（伸び率平均値補完と呼ばれる）、する方法である。

非公開項目以外のデータを用いて補完する手法としては、さらに、項目間比率補完がある。この手法は、非公開項目と他の項目との比率（項目間比率と呼ぶ）を、非公開項目と対応する他の調査項目に乗じることで、非公開値データを導出する手法である。この比率に経済学的意味があれば、解釈可能になり大変理解しやすいものとなる。

3-2-4-2 非公開データの補完に適用

これらの方法を、非公開データの補完に適用するとしたら、どのような問題があるか、どれが適当か、考えてみよう。都合により逆順で進める。

外のデータが獲られない、あるいは自身の過去のデータも非公開にされており一切存在しない、ということは普通に起こるので、代替値代入法は実際上適用できないケースが非常に多いと予想される。伸び率平均値補完、特に項目間比率補完は、経済学的に納得が獲やすくても、データ入手の制約が大きく影響し、採用を妨げる要因になる。

近傍値代入法は、上で述べた要因により非公開者の了解が得られないことも起こり得るため、適用できない可能性がある。「貴方様の非公開データは、次の方法で補完いたします」と事前に本人の了解をえることが必要になる。あるいは、本人には公開しないことを事前に了解をえることが必要になる。

平均値補完法は、欠損値の補間によく使われており、非公開値の補間にもっとも可能性が高いと考えられる。さらに、公開拒否の理由を個人・組織の属性と独立にできる可能性がある。つまり個人・組織の属性から公開拒否理由を推測出来なくできる大きな利点がある。しかしながら、層のなかに、複数の非公開データがある場合、そのすべてに同じ平均値は代入できない。考えられる問題は、この点である。しかしながら、この場合は、個々の非公開値毎に、事後層化し、都度平均値を計算していく方法が代案として考えられる。

3-2-4-3 補完方法に関するその他の問題

カテゴリ・データや2値データ、あるいは外れ値や異常値に関しては、どうすれば良いだろうか。まずは、補完方法に関する補充説明から始めよう。

(1) 多重代入法

前小節で説明したのは、単一代入法と呼ばれる技法であるが、その他の補間補完法として多重代入（multiple imputation）法がある。この方法は、1つの非公開値に複数の値を代入するという点が特徴である。このことで変数の確率分布の要素を具体化できる、それゆえプライバシーはさらにボヤけるメリットがあるが、様々な局面で計算量が多くなり過ぎるという欠点がある。そのため、欠損値補完では余り利用されていない。パーソナル情報提供業にどう適用するか、の検討は将来の仕事にしたい。

(2) 2値データやカテゴリ・データ

2値データには1と0の数値（例えば、「はい・いいえ」のアンケートでは「はい」に1、「いいえ」に0）を割り当て、層内平均値を計算して、平均値が0.5以上なら「はい」を代入するとよい。これは、層内では「はい」が「いいえ」より多いから、非公開値には「はい」と入れておく、措置に相当する。より高次のカテゴリ・データも同様に適用できる。

(3) 外れ値

データセットの中の他のデータから著しく乖離したデータ（つまり外れ値）は母集団を代表していない。外れ値は、別途、分析の当初から存在の有無を確認し、どういう対処を行うかを決定しておかなければならない。単なる集計ミス、処理のプロセスで発生した、などというケースも多い。検出方法、対応法などの先行研究を展望もしている石川・遠藤・白鳥 [2010] が詳しい。

外れ値や異常値が引き起こす問題に対処するには、平均値に代えて中央値を用いるのが有効であることが知られている。非公開値補完についても、同様である。

4 暗号化状態処理技術について

最後に、プライバシー保護技術の新しい動向との関連性に触れておかねばならない。それは、暗号化状態処理と呼ばれる技術に関してである。

4-1 暗号化状態処理技術の概略

議論を進めるための便宜上の都合で仮に、プライバシーを認めず、すべてのパーソナルなデータ・情報を報告させる、と想定してみよう。しかる後、データを暗号化し、検索も計算もすべて暗号化データで処理できれば、結果としてプライバシーは守れる、ことになる。

これらが、暗号化状態処理技術が理想とする目標で、暗号で検索し、暗号で計算し、悪意のあるデータ管理者から（も）パーソナル情報を守るのが狙いとなっている。この世界では、パーソナル情報を保有する個人自身の外では暗号だけが流通することになる。

背景は次のとおりである。暗号は、現在主として、データを移送する際や保管する際に利用されているに過ぎない。移送過程については盗聴されたり漏れることがないシステムや高度な暗号で対処する方向に進みつつある。しかしながら、そうであっても、保管中や保管が終了してデータを利用する際には、一時的あるいは永久に復号する（平文に直す）必要があり、その際不正アクセスされたり、サーバ管理者などが（復号鍵を）不正利用する、恐れが懸念されている。

そこで、暗号のまま検索したり、一旦暗号化された文章を復号することなく別の鍵で復号する（代理人再暗号化と呼ばれる）技術、さらには暗号化されたデータ同士の演算を行う技術が開発されている。これらが、総称して暗号化状態処理技術と呼ばれるもので、その展望については清藤 [2014] を参照。

この方式のもう1つの特徴は、データ利用希望者にはデータそのものではなく、分析希望を聞いてデータセンターが暗号計算し、復号化した分析結果だけが提供される、点である。そして分析結果だけからは、データ利用希望者が元データを推測できないようになっている。データ自体が欲しい利用者には、断るしかない。

4-2 暗号化状態処理技術の限界

(1) 暗号化状態処理技術の将来

暗号化状態処理技術があれば、前節で展開した補完技術は要らないということではないことを2つの観点から説明しよう。まず第一に、暗号化状態処理技術の将来に関してである。

現行の暗号化状態処理技術は、特定の暗号技術（準同型性と呼ばれるタイプ）の下でのみ、稼働可能になっている。ところが、この暗号方式のセキュリティの水準は高くない。よりセキュリティの水準の高い暗号方式で暗号化状態処理技術が近い将来開発されるかどうかは全く分からない。それゆえ、有用な暗号化状態処理技術が実用化される保証はないことになる。

暗号化されたデータの統計解析は、本来、暗号検索と加算・乗算からなる暗号計算から構成される高度なものである。そして、ビッグデータの分析や高度で複雑な計量経済学的分析を行うのは不可能である。簡単なものでも、計算速度などの観点から満足いくものは、まだ開発されていない。この点だけは時間が解決してくれるかもしれない。

さらに、根本的で深刻な問題が横たわっている。暗号文と平文との間で変換する際に漏洩や搾取がよく行われるということであれば、いつもすべて暗号にしていまえば、というのが基本的な考え方である。しかしながら、すべてが暗号ならば、すべて平文であるのと変わらなくなるのではないだろうか。この暗号がこの物・人・事象などを指すという事実はいつかすべて判明してしまう、からである。

（２）暗号化状態処理技術と補完技術

第二に、暗号化状態処理技術と補完技術の関係に関してである。分析結果を手に入れるだけで満足できる利用者がいるとすれば、それは分析意図と分析結果が公になっても構わない者に限られる。それは、例えば大規模企業グループがデータセンターを保有し、利用者が同じグループに属す企業である場合であろう。その数は極めて限られる。

元データも欲しがると、あるいは元データだけを欲しがると利用者はかなり存在することが予想される。購入したデータに、自身のデータをそのまま加えたり、独自性を増す（例えば、新しい理論を開発することによって新測度を考案）ように加工して、新たに最終ユーザーに売り出す業者も、本稿前半部分で既述のように、存在する。非公開情報の信託業務は、事業として、このような要望を無視するべきではないだろう。データを提供する際には、非公開データの補完技術が必須になる。補完した後、希望するならば、暗号化すればよい。

それゆえ、暗号化状態処理技術と非公開データの補完技術は相容れないものではない。両者は、異なった目的を持った技術であり、互いに補い合えるのである。

5 まとめ

インターネットは、善意や信頼関係に基づく性善説で考案され、人間の悪意や犯罪者を想定して設計されていない。当初からセキュリティやプライバシーを想定して作られたものではない。それゆえ、既に大きな市場になっている情報セキュリティだけでなく、プライバシーに関しても起業家にとっては大きなビジネス・チャンスになるのではないかと思われる。IT企業が個人生活に過度に侵入していると感じる人は多いとする調査⁸⁾も出ている。

8) ロイターとイブソスが2014年3月11日—26日に米国で4781人から回答をえたアンケート調査では、グーグル、フェイスブック、アマゾン・ドットコム、アップル、マイクロソフト、ツイッターが個人生活の多くの領域に過度に進出していると懸念する回答は51%に達した。

また、IT企業による個人情報の取り扱いや情報管理の安全性については、約3分の2の回答者が「不安を感じる」と答えた。位置情報や支出パターン、個人的な情報交換に関するデータを収集する能力について、最も懸念されたのはグーグルとフェイスブックだった。

本研究で提案した非公開選択制度に残された課題はある。例えば、個人が、自身の意思で非公開にしたパーソナル情報を、(場合によって意図せず)自身で公開してしまう事態も起こりえる。この事態への対策、補償問題への対応、など考えておかねばならない。

非公開の情報が流布しないような多層で堅固なシステム作り、非公開に確かになっているかを該当の本人が確認できる仕組み作り、も求められる。これらの点は、情報セキュリティ問題であり、筆者は既に一部は研究している。

データ・情報は文字通り無数にある。事例に基づいたケース・スタディを積み重ねる必要があることを否定する者はいないだろう。その過程で、様々な問題が浮かび上がるかもしれない。「いいね！」も個人情報公開であるというケンブリッジ大学の研究もある⁹⁾。欠損値の補完法もまだまだ発展するかもしれない。非公開選択の経済分析も、様々な分野で具体的に展開されるかもしれない。パーソナル情報に関する非公開データの補完方法はまだまだ発展の見込みがあるのである。

参考文献

- ドラッカー、P. F. (Peter F. Drucker) 著, 上田惇生, 林 正, 佐々木実智男, 田代正美訳 [1995] 『未来への決断』, ダイヤモンド社, 1995年9月。
- ドラッカー、P. F. (Peter F. Drucker) 著, 上田惇生編訳 [2001] 『マネジメント [エッセンシャル版]』, ダイヤモンド社, 2001年12月。
- Goto, M. and Tatsumi, K., [2012] “The Theory of Optimal Investment in Information Security and Adjustment Costs: An Impulse Control Approach,” pp.73-96 in *Recent Advances in Financial Engineering 2011*, Proceedings of International Workshop, edited by Takahasi, A., Muromachi, Y. and Nakaoka, H., World Scientific Publishing, 2012.
- 平川貴大・鳩井淳一郎 [2012] 「ビジネスサーベイにおける欠測値補完の検討—全国企業短期経済観測調査(短観)のケース—」日本銀行ワーキングペーパーNo.12-J-8, 2012年8月。
- 石川篤史・遠藤俊介・白鳥哲哉 [2010] 「ビジネスサーベイにおける外れ値対応—全国企業短期経済観測調査(短観)のケース—」日本銀行ワーキングペーパー10-J-12, 2010年7月29日。
- Little, R. J. A. and Rubin, D. B., [2002], *Statistical Analysis with Missing Data*, Wiley, September, 2002.
- 南 和宏 [2013] 「プライバシー保護データパブリッシング」『情報処理(情報処理学会)』, 54(9), pp.938-946.
- Musil, S., [2013], 「Facebookの「いいね！」で個人的属性が分かる?—ケンブリッジ大が危険性を指摘」

IT企業が無人機を開発することに否定的だという回答は42%だったが、これらの企業による家電分野への進出に否定的との回答は13%にとどまった。(2014年4月4日ロイター)

- 9) 「いいね！」も個人情報公開である。ケンブリッジ大学の研究者らによると、Facebook上で5万8000人のユーザーのLike「いいね！」を調査することによって、ユーザーのIQや性別、性的指向、政治的および宗教的信念、さらには薬物使用までの、個人的属性を80%以上の精度で特定することができたという。Musil [2013] 参照。

Likeの対象は、写真や友達の近況アップデート、さらにスポーツやミュージシャン、書籍のページなどに対する支持である。それを、ユーザーが提供した人口統計学的プロフィールやそのほかの精神測定テストと比較したところ、性的指向を88%、民族性を95%、政治的傾向を85%の確率で正しく予測できたという。Likeからは、ウェブ検索クエリやウェブ閲覧履歴、クレジットカード利用明細と同様に、非常に多くの属性をうかがい知ることができる。それゆえ、その危険性を認識することが重要であると指摘された。

- CNET News, 2013年3月13日。
- シュナイアー, B., [2010] 「フェイスブックの爆発的増殖力が突きつける難題！「我々はプライバシーの制御を失う覚悟はあるのか」」『ダイヤモンド・オンライン』, 2010年11月22日。
- シャピロ, カール, & バリアン, ハル・R., [1999a] 『顧客価値をとらえるバージョニング戦略～付加価値の連鎖により収益機会を拡大させる』, ダイヤモンド ハーバード・ビジネス・ライブラリー, 1999年5月。
- シャピロ, カール, & バリアン, ハル・R. 著, 千本俤生監訳, 宮本喜一訳, [1999b] 『「ネットワーク経済」の法則』, IDG ジャパン, 1999年。
- 清藤武暢 [2014] 「暗号化状態処理技術を巡る最新動向」第15回情報セキュリティ・シンポジウム, 日本銀行金融研究所, 2014年3月5日。
- 辰巳憲一 [2011a] 「金融・経済活動における情報などの分割, バックアップと情報セキュリティ～金融セキュリティの経済学入門（Ⅰ）～」『学習院大学経済論集』, 2011年1月, pp.301-321。
- 辰巳憲一 [2011b] 「個人情報信託の経済分析～プライバシー情報を保護しながら信託で一元管理する～」『学習院大学経済論集』, 2011年7月, pp.83-109。
- 辰巳憲一 [2011c] 「情報セキュリティ問題とその進化」『学習院大学経済経営研究所年報』, 第25巻, 2011年12月, pp.15-26。
- 辰巳憲一 [2012a] 「情報セキュリティの階層型統合管理に関する経済分析（Ⅰ）」『学習院大学経済論集』, 2012年4月, pp.3-21。
- 辰巳憲一 [2012b] 「情報セキュリティの階層型統合管理に関する経済分析（Ⅱ）」『学習院大学経済論集』, 2012年7月, pp.99-116。
- Tatsumi, K. and Goto, M., [2010] “Optimal Timing of Information Security Investment,” in Moore, T., Pym, D. and Ioannidis, C. (eds.), *Economics of Information Security and Privacy*, Springer, 2010. (Chapter 11)
- 辰巳憲一・後藤 允 [2010] 「情報セキュリティとその投資の分析～研究報告書～」『学習院大学計算機センター年報』2010年12月, pp.49-62。
http://ci.nii.ac.jp/els/110008148617.pdf?id=ART0009662982&type=pdf&lang=jp&host=cinii&order_no=&ppv_type=0&lang_sw=&no=1303796391&cp=
- 辰巳憲一・松葉育雄 [2008] 「時系列データにおける補間方法の分析と考察」『学習院大学経済経営研究所年報』, 第22巻, 2008年12月, pp.35-43。
- Ton, W., Jeroen, P. and Sander, S., [2010], *Handbook of Statistical Data Editing and Imputation*, Wiley, 2010.
- 宇都宮浄人・園田桂子 [2001] 「「全国企業短期経済観測調査」における欠測値補完の検討」日本銀行ワーキングペーパーシリーズ01-11, 2001年8月29日。
- 吉川尚宏 [2002] 「プライシング戦略」, 野村総研, www.nri.co.jp/opinion/m_review/2002/pdf/nmr02-3.pdf, 2002年3月。