

ビッグデータの時代へ向けた研究の展望

田中 幹夫*

The Prospects of the Study toward the Era of Big Data

Mikio TANAKA

In late years, collection, analysis and use of huge data called “Big Data” have begun and brought results to marketing, demand prediction and creation of new business. In the field of the transportation plan, the administration and the maintenance of the railroad, it is considered to utilize the big data. And it is hoped that it leads to the improvement of the railroad administration and that of railroad users’ convenience. In this report, the prospects of the use of the big data in the railroad and those of the study on a large quantity of data are discussed.

キーワード：情報処理，輸送計画，需要予測，旅客流動，設備保守，信頼性

1. はじめに

近年、ビッグデータという言葉が良く使われる。従来
の方法では取扱困難な巨大規模データを指し、その分析
活用による業務改善や新ビジネス創出が期待されてい
る。ビッグデータとは文字通り巨大情報の意味で明確な
定義はないが、従来手法（データベース，処理アルゴ
リズム）やハードウェア（記憶装置）能力では処理が難
しい、数10ペタバイト以上を指す場合が多い。現在の相
当な大規模システムでも扱うデータの上限は数100テラ
程度と言われ、それ以上の領域を扱う、いわばビッグデ
ータ独特の手法の研究開発も進んでいる。サイズの大き
いだけでなく、データを世界中に分散配置し、並行的に
処理や検索を行う高度機能も開発されている。高速処理
可能なスーパーコンピュータの発展の寄与も大きいであ
ろう。ビッグデータという言葉自体はIT分野でよくあ
るブームの1つという見方もできるが、情報化によって
生まれた資産である巨大データの分析活用という流れは
確固たるものである。近年、「パラダイムシフト：第4
のパラダイムへ」¹⁾ という事が提唱されている（表1）。
ここで示唆されている、今後の「データ科学の時代」を

反映したものがビッグデータの時代と言えよう。なお本
稿では、前記のようなデータサイズに拘らず広く大量
データの分析・活用の展望を述べる。

2. ビッグデータ活用の背景

ビッグデータが注目を浴び、活用され、効果を上げて
いる背景として以下のような状況が挙げられる。

- ・ 情報システム・通信ネットワークが普及拡大し、それ
を通して膨大な情報資産が蓄積されている。
- ・ 人間やモノの様々な活動（動き）に関する大容量の情
報が記録・蓄積可能となった。
- ・ ハードウェア（処理装置，記憶装置，通信ネットワー
ク，センサ）の機能向上，大容量化が実現した。
- ・ データ分析処理技術が発展している。

これらを背景に、情報システムの対象を、色々な場
面で人についてもモノについても「マス（集団）から個へ」
転換する流れが生じている。従来、人を対象とした需要
予測・マーケティングでは、マスの特性を重視する、モ
ノが対象の場合は、対象種の平均像を想定して制御や監
視を行う方法が一般的であった。近年では視る対象が精
緻化され、個人の特長・嗜好を意識した対応（マーケテ
ィング等）を試みる、モノに対しては、対象個々の特性に
応じた最適な監視や制御を行なう、という方法が広がっ
ている。これにより社会全体として、効率向上，サービ
ス向上，省資源等が図られようとしている。

この状況で必要になるのが、人にしてもモノにしても、
個々の特性に関する情報である。従来、この種の情報を
正確・大規模に捉えるのは困難で「マス」の把握に留まっ
ていたが、現在は、インターネット利用や各種機器から
得られる情報（例えば携帯電話位置情報，自動改札情報

表1 パラダイムシフト¹⁾

	パラダイム	概要
1	経験科学	観察・観測で得た情報を基に経験的処理
2	理論科学	経験を分析し背後の論理・法則を探求
3	計算科学	解析解が得られぬ対象をコンピュータ能力で数値シミュレーション的に解く
4	データ科学	膨大なデータ、遍在・強力な計算能力の存在を前提にした科学研究

* 信号・情報技術研究部 主管研究員

特集：信号通信・運輸

など)を通した膨大な情報が、いわゆるビッグデータとして得られる時代になってきた。

監視や制御の電子化・コンピュータ化の普及も重要な背景である。鉄道の例として転てつ機の保守を考える。密着度合、鎖錠状態(ロック)、モーター電流値等が監視値として共通的に設定・マニュアル化され、それに基づき人手主体の保守が行われてきた。個別の特性、状況等が明らかになれば、設備個々に適合した、きめ細かい効果的・効率的保守が可能になる。しかし保守が全く人手で行われる時代、ある駅の30台の転てつ機の個別特性が明らかになったとしても、1台毎に調整方法や検査周期を変える、という方法は利点が少なく現実的でない。人的運用からも「この駅の転てつ機は、ある時期・方法で一斉に」という計画にならざるを得ない。監視や保守の電子化・遠隔化等が進めば、過去蓄積データを活用した、個々の機器特性に応じた個別措置も可能になる。

3. 鉄道におけるビッグデータ研究の展望

鉄道の様々な分野で、ビッグデータを活用・適用した研究が考えられる。ここでは幾つかの例を挙げてみたい。

3.1 利用者行動のモデル化と輸送計画の改善

鉄道の長所は「大量定型輸送」で発揮され、輸送力絶対量が不足していた産業革命期に大発展を遂げた。今後の社会を考えた場合、人口や必要輸送力の減少は必須である。鉄道は本来の長所を維持しつつ、自動車等の個別輸送機関と比べて大きな弱点である「個別ニーズへの対応」「ニーズ変動に対する柔軟性」を備えたシステムへの転換が望まれよう。

今迄、これを意識した「デマンド輸送方式」²⁾(利用者ニーズに応じた動的柔軟な輸送計画)(図1)、「曖昧予約方式」³⁾(柔軟予約による列車への動的旅客配分)等の研究・シミュレーションも試みられてきたが実現への道のりは遠い。「個別ニーズ対応」と「輸送効率性」の両立という難しい課題に対応するためには、個々の移動ニーズを正確精緻に捉え、予測し、うまく集約して大量定型

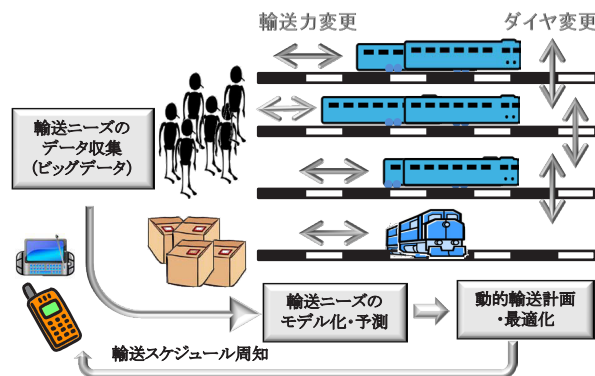


図1 動的(デマンド指向)輸送計画の概念

輸送に近づける、といったアプローチが求められる。このために利用者動向の精緻な分析・予測が必要である。

利用者動向の予測にはマイクロ経済学の効用理論に基づく行動選択モデルが一般に使われる。これは主体(購入者、利用者)が客体(購入商品、交通機関)を選択する際、主体、客体、状況、環境等に係る各種属性値から算出される効用値(主体からみた価値の高さ)によって各選択肢の確率分布が得られるというものである。例えば利用者が、どの交通機関・列車を選ぶかは主体属性(年齢等)、客体属性(料金、時間等)、環境属性(移動目的、天候等)等で各選択肢の効用値、選択確率が計算され、これにより旅客行動予測が可能になる。しかし上記パラメータ組合せは膨大で、必要な信頼性でのパラメータ同定は至難である。相当量のビッグデータが得られた場合、各種属性組合せに対する実効的なデータサイズが得られ、各属性値を行動モデル式に入れて個々の行動予測を積上げる事で精緻な予測に繋がる事が期待される。

上記は定常時における行動予測、輸送計画であるが、近年発生したような大災害による輸送停滞・停止時や、事故や設備故障による列車乱れ時など異常時の対応方法も重要である。交通機関が停止した際の利用者の行動選択(移動取止め/再開待ち/迂回等)の分析がビッグデータを基としてできれば、異常時対応の支援、帰宅困難者対策、利便性向上に繋がっていくであろう。

3.2 システムバランスの変化検知

我々を取り巻く状況は様々な要素から構成され、仮にN個の変数を持つシステムと考えれば、各時点の姿はN次元空間上の一点に位置づけられ、それが時間経過と共に動いていくというイメージになる。

通信トラフィック、コンピュータアクセス、人間の移動・位置等の状況も、ある膨大な次元の下でのシステム状態と考えられる。この状態の変化を検知する研究は最近盛んになっている。例えば通信トラフィックの量や内容から、サイバー攻撃やテロの危険性を確率的に求める、インターネット利用状況や検索キーワード変化から、景気や購買行動の変化を検知予測する等である。鉄道での一つの試みは旅客流動の変化検知である。これは長期的な輸送計画・設備計画策定や、需要開拓・マーケティング等に有用である。しかし、大局的变化を如何に検知するかは難しい課題である。一つの手法は、自動改札機データを分析し、各駅/各OD(発着)利用者等の多くの変数の相互関係を分析し、それを重回帰でモデル化するので概念を図2に示す⁴⁾。定常時には回帰と現実との差は、ある範囲内に収まるがそれを超える場合には何らかの状態変化が考えられる。その確率、定量的差分を推定するものである。利用者動向に影響する説明変数には膨大な種類が考えられるが、ここでは、他駅/ODの変数

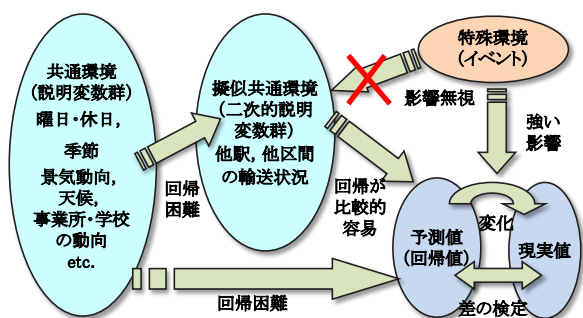


図2 イベント効果推定・検定の考え方

を二次的説明変数として使用することで重回帰の簡略化を図っている。その手法を用いてイベント効果による旅客増を定量的に評価した例を図3に示す。

同様の考え方は設備監視にも適用できる。軌道回路は列車検知、信号制御のベースとなる重要設備で、その特性監視は、輸送の安全性・安定性確保に重要である。しかし、この特性は様々な環境や設置条件に依存して常時変化し、健全性判定は困難である。過去データを基に各軌道回路の特性値間の関係を分析して常時の姿を求め、それからの外れの大きさと健全度を推定しようという方法が考えられる。このため、データに基づき、軌道回路

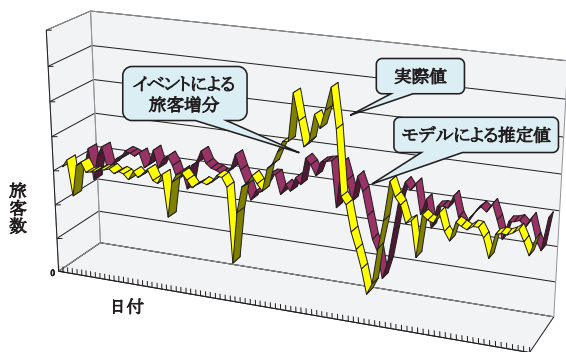


図3 イベントの検知と定量化

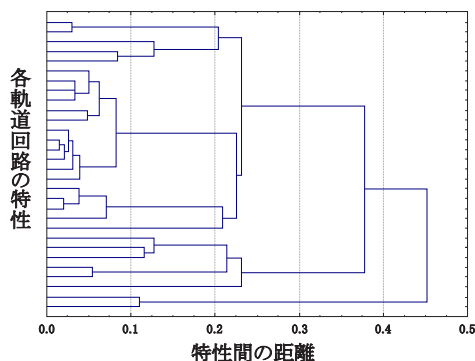


図4 軌道回路のクラスター分析（デンドログラム）

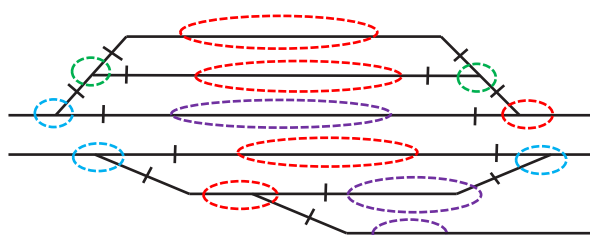


図5 軌道回路特性の分類（同色が相関大）

特性の相関分析やグルーピングを行う。図4は、ある区域の軌道回路特性のクラスター分析を行った例（デン

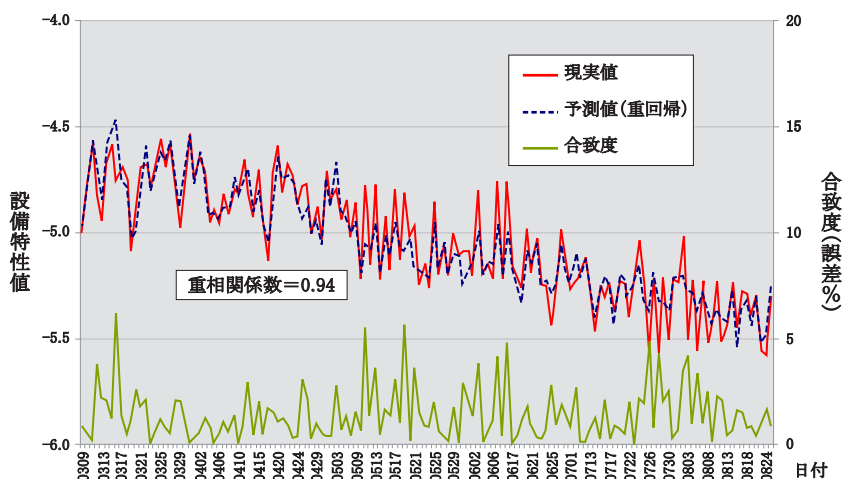


図6 設備特性の時系列変化の現実値と予測（重回帰）値

ドログラム)、図5はクラスター分析や相関分析により、類似特性の軌道回路を見出した例である。

図6は特性変化の検知例で、他設備特性値や環境変数を説明変数とする重回帰値と現実値とを併せて示したものである。両者の相関は高く、合致度チェックによる健全度変化の早期検知手法の研究を進めている。

3.3 設備の特性分析

鉄道輸送の安全性・安定性は多種多様な設備に支えられ、その健全性は保守検査や監視によって確認されている。従来の基本的考え方では、設備種類毎に保守方法(検査部位、方法、閾値等)や検査時期(周期)等が一律に定められる。現実の設備状態変化は、設備の使用・設置環境、経年、過去履歴等々に依存して様々であり、特定基準で判断できるものではない。現場保守担当者も、この状況は熟知しているものの、データが大規模効率的に収集・分析される状況ではないため、これに関する知識の共有、体系化、技術継承は進んでいない。

近年は監視装置が進展しデータが大規模・自動的に収集蓄積される例が増えている。これらデータを利用した分析の試みを示す(ここの結果はサンプルに基づく例で、一般的結論に至るものではない事に留意されたい)。

まず、電車線設備で多く使用される曲引装置について

特集：信号通信・運輸

表2 電車線設備の設置環境と処置率（曲引）

設置環境	全検査	処置要	処置不要	処置率
一般	131895	2767	129128	0.021
橋梁	5493	348	5145	0.063

表3 信号設備の使用環境と処置率（転てつ機）

列車方向	全検査	処置要	処置不要	処置率
背向	18366	943	17423	0.051
対向	52723	2397	50326	0.045

の分析例を挙げる。数万回の検査において何らかの処置が必要となった割合（処置率）を設置環境（一般/橋梁）別に示したものを表2に示す。この結果から、統計解析手法（分割表における χ^2 乗検定）により「処置率は環境種別に依存するか」の検定を行う。ここでは、有意水準0.05のもとで「差が有意」、即ち「橋梁区間ではトラブル発生確率が高い」という事実が導かれた。

次に転てつ機に関する結果を表3に示す。前記同様の統計的検定の結果、背向の転てつ機の方が、対向のものよりトラブル発生率が高いという結果になった。なお、各種属性に関して同様な分析を行った結果、以下のものであった。

- ・ 処置率の差が統計的有意となった属性
 - 対向/背向、形式（NS/TS、弾性/関節）、先端/ノーズ
 - ・ 処置率の差が統計的有意とならなかった属性
 - 片/両開き、右/左開き、本線/側線、番数、経年
- なお、これら結果が実際の保守担当者の直観・知見と合致するかは興味深い。曲引の例では、ほぼ合致する（一般に橋梁上は風雨に暴露され苛酷な状況）。転てつ機では、対向の方が（車輪がポイント先端部に突入してくる関係から）トラブルが多いのでは、という意見がやや多く、今回の結果は、それとは逆のものであった。

また、設備の寿命・交換時期の検討のために、設備の設置環境別に生存時間分析を行った例を図7に示す⁴⁾。これは過去の検査結果、トラブル発生状況のデータを基に信頼度、生存率（健全性）の変化を分析し、環境毎に取替時期の最適化を図ろうというものである。

上記のような検討を広範囲大規模、体系的に進め、いわゆるビッグデータを構築して分析活用できれば、個々の設備特性に応じた、より適切・効率的な保守方法（例えば属性によって方法や周期を変える）を見出せる可能性がある。また、熟練保守担当者の定性的な知見経験の中にある貴重な知識が定量化され（或いは否定され）、新たな事実の発見や技術継承の一助にもなるであろう。

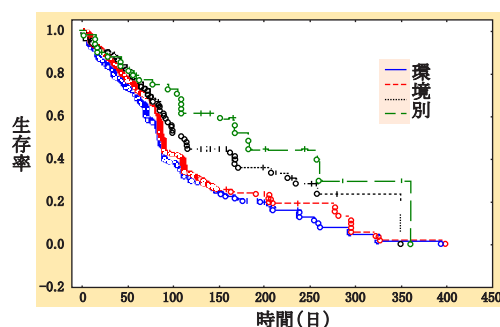


図7 設備の生存時間分析

4. おわりに

鉄道の各分野で共通データベースを備え、ビッグデータ活用を図るイメージを図8に示す。鉄道業務は複雑・専門的で分野毎縦割になりがちであるが貴重なビッグデータは各分野で共有し有効活用を図るべきであろう。

本稿で述べた内容は技術的側面であり、他に多くの課題がある。プライバシー保護の面から、個人情報利用はどの程度許容されるか、利便性の面から「運行ダイヤがニーズにより動的に変わる」形態が歓迎されるか、設備個々の特性に応じた保守は、現場の実務に馴染むか、法的保守基準との整合性を保てるか、等である。今後、各方面から御意見も頂戴し、将来に向けたビッグデータ活用の研究を考えていきたい。

文 献

- 1) Microsoft Research : The Fourth Paradigm: Data-Intensive Scientific Discovery, <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>.
- 2) 富井, 他: デマンド輸送実現のための課題, 鉄道総研報告, Vol.17, No.12, pp1-6, 2003
- 3) 田中, 他: 新しい列車予約方式の研究, 鉄道総研報告, Vol.11, No.8, pp37-42, 1997
- 4) 田中, 他: データマイニング手法の鉄道への適用の研究, 鉄道総研報告, Vol.14, No.7, pp7-12, 2000

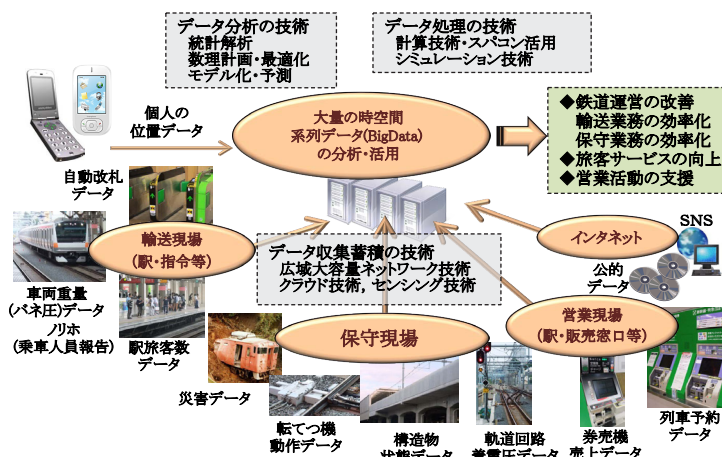


図8 鉄道におけるビッグデータ利用の総合的イメージ