

- 鉄道一般
- 車両
- 軌道
- 構造物
- 防災
- 電力
- 信号通信
情報
- 材料
- 環境
- 人間科学
- 浮上式鉄道

ビッグデータの時代と 鉄道輸送

近年は「ビッグデータ」と言われる巨大なデータの収集・活用が始まり、マーケティング、需要予測、新たなビジネスの創出などに成果をあげています。鉄道の輸送計画や運営・保守においても、大量データの収集分析と利用が考えられ、業務運営の改善や利便性向上などにつながる事が期待されます。本稿では、鉄道におけるビッグデータ利用や、大量データを対象とした研究の展望を述べたいと思います。



田中 幹夫
Mikio Tanaka
信号・情報技術研究部
主管研究員
【専門分野】情報工学,
電気工学

はじめに

インターネットを使っている方・・・『この商品を検索した人は、こんな商品も見ています』『この商品を買った人は、こんな商品も買っています』といったメッセージを目にしませんか？

20XX年、鉄道会社の輸送指令室・・・『こう動いた旅客は、次にこんな行動に移ります』『M分後にN駅ホームの旅客数はP人に達し注意が必要です』といったメッセージが指令員に提供されているかも知れません。

「ビッグデータ」という言葉を良く耳にするようになりました。従来の方法では取り扱いが困難な巨大規模のデータを指しており、その分析活用により現行業務の改善や新ビジネスの創出

が期待されています。公的にも米国オバマ大統領の施政方針演説¹⁾でビッグデータの重要性が言及され、関連研究への大規模予算措置が発表されました。これでさらにブームに拍車がかかったようです。

「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ありませんが、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)の能力では処理が難しい、数10ペタバイト(☞参照)以上を指す場合が多いようです。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、それを超えた領域を扱う、いわばビッグデータ独特の手法の研究開発も進んでいます。サイズの大きいだ

☞ ペタバイト

データの大きさを示す単位。1バイトは1文字分に相当します。ペタバイトは 10^{15} バイト。なお、1ペタ= 10^3 テラ= 10^6 ギガ= 10^9 メガ= 10^{12} キロ。

☞ データ科学の時代

近年、「パラダイムシフト：第4のパラダイムへ」³⁾という事が提唱されています。これは科学の流れを、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・法則を探索、③計算科学：解析解が得られぬ対象をコンピューター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、と捉え、今後は、④の時代に入っていく、というものです。

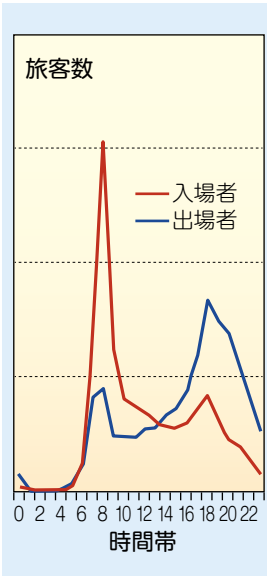


図1 駅の入/出場者数

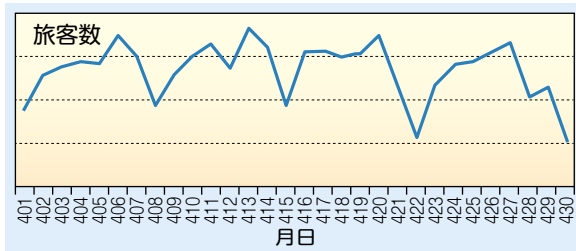


図2 都市間の旅客数

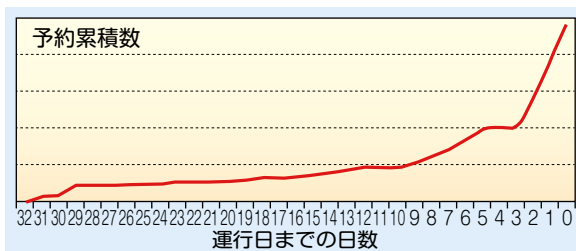


図3 列車の予約累積数

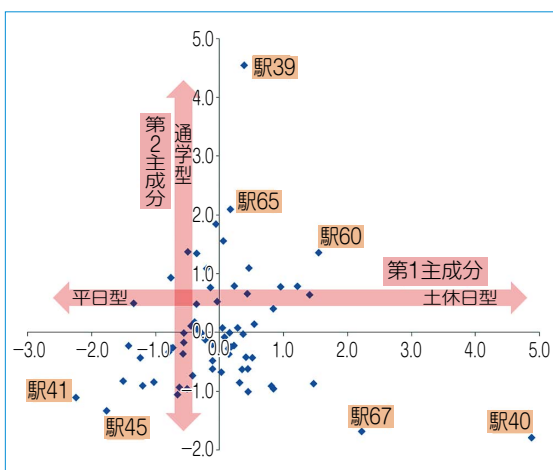


図4 駅の主成分分析(自動改札データ使用)

量が不足していた輸送力の提供という役割から大発展しました。大量定型輸送に適し、個別輸送と比較して格段に高い輸送力と効率が利点です。

日本の人口は減少期に入り、30年後は最盛期の80%と予想され、輸送量増加は見込めません。都市圏の利用者の多くは通勤通学者で、その減少割合はさらに大きくなるでしょう。今後は大量

定型輸送の長所の維持に加え、個別多様ニーズにも対応して利便性を高め、かつ輸送効率性を保つ柔軟な鉄道への転換が必要でしょう。また、例えば輸送量80%に対してコスト80%、さらにそれ以下にできる運営が求められます。

「個別ニーズ」への対応を犠牲に「大量定型輸送」に徹して長所を發揮してきた鉄道にとって、これは実に難題ですが、これに答えるため、情報技術、ビッグデータの活用は、1つの手掛りとなるのではないのでしょうか。

情報技術・ビッグデータ利用の試み

今まで「デマンド輸送方式」(利用者ニーズに応じた動的柔軟な輸送計画)、

「曖昧予約方式」(柔軟な予約による、列車への動的旅客配分)などの研究やシミュレーションは試みられてきました。しかし課題は大きく実現には至っていません。個別ニーズ対応と輸送効率性の両立のためには、個々の移動ニーズを正確精緻に捉え、予測し、それをうまく集約して大量定型輸送に近づける、といったアプローチが必要です。どんな技術、手法が必要でしょうか？

①利用者動向に係るデータ収集

今まで大都市交通センサス調査や乗務員によるノリホ作成(乗客数把握)などが行われてきましたが、正確性、規模、頻度が不十分です。今後求められるのは精緻で大規模なものです。自動改札・券売機データ、車両重量(重量から乗客数推定)、駅での画像処理、利用者携帯端末の位置情報などの利用が考えられます。データ例を図示します。図1は大都市ターミナル駅の1日の時間別入/出場者数、図2は1ヶ月間の2都市間発着(OD)旅客数、図3は長距離列車に対する予約の1ヶ月前からの累積数、図4は利用者情報(券種)からの駅特性分類(主成分分析)です。今までこれらデータに基づく旅客流動の研究は行われてきましたが、まだまだデータの精密さや収集体制が不十分で、個々の旅客特性と比較対照した精緻な分析やモデル作成は不可能でした。目標である「旅客ニーズ予測」には程遠い状況と言えます。今後、データ収集と解析が広がるビッグデータ時代には、格段に正確で分解能が高いデータを用いて、より適切で正確なモデルの仮定、現実に近い予測が期待されます。

②利用者ニーズのモデル化

利用者動向の予測にはミ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ます。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

けでなく、巨大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な機能²⁾も開発されています。大量データの高速度処理が可能なスーパーコンピュータの寄与も大でしょう。

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ますが、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、今後の「データ科学の時代」³⁾(註参照)を反映したものと言えます。なお本稿では、データサイズに拘らず広く大量データの分析・活用の展望を述べます。

鉄道の特徴と今後求められるもの

鉄道は19世紀の産業革命期、絶対

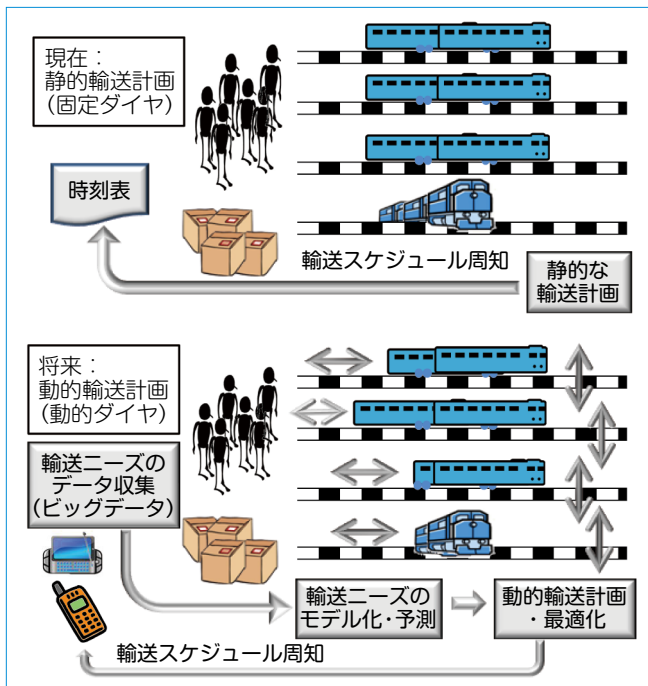


図5 静的輸送計画と動的輸送計画 (イメージ)

況、環境など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す。例えば利用者が、どの交通機関・列車を選ぶかは主体属性 (年齢など)、客体属性 (料金、時間など)、環境属性 (移動目的、天候など) などで各選択肢の効用値、選択確率が計算されます。これにより旅客行動予測が可能になるわけです。しかし上記パラメータの組み合わせは膨大で、必要な信頼性でのパラメータ同定は至難です。パラメータの組み合わせによっては欠測が生じる場合も多々あります。相当量のビッグデータが得られた場合、各種属性の組み合わ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み上げることで精緻な予測につながることを期待されます。

ちなみにデータ量はどの程度になるでしょう？凡その概算ですが、属性20種類、各特性3択、必要ケース数400としますと、標本サイズは $3^{20} \times 400 \approx 1$ テラとなります。データの大きさとしては数100テラバイト、さらに時系列で特性変化を追うためにはペタバイト領域となるでしょう。まさにビッグ

データ特有の規模となっていきます。

そこで、冒頭例のようなものも想像される訳です。オンラインショッピングなどで普及していますが、これは、購買・利用者の検索や実行動に係る大量データを蓄積・分析し、そこから予想動作の確率分布を見出すもので「協調フィルタリング」などの手法が使われます。

③動的な輸送計画と最適化

実務面からみると、動的な輸送計画作成というのは複雑で難易度が高く、バスのような「今日は客が多いから1台増やす」といった技は鉄道では至難です。線路容量、駅設備、車両運用、乗務員運用等々の複雑な要素を全て勘案して作成するのが輸送計画や列車ダイヤであり、人手による動的計画作成は現実的ではありません。オペレーションズリサーチの技術が基礎となり、前記のような輸送上の各種制約下で、ある目標値 (利便性や提供コストなど) を最大かつ最適にする手法が必要とされます。

④暗黙的パターン予約

受動的データ把握だけでは利用者動向の予測は困難と考えられます。それでは、どんな方法が考えられるでしょう？一案として全利用者に「(利

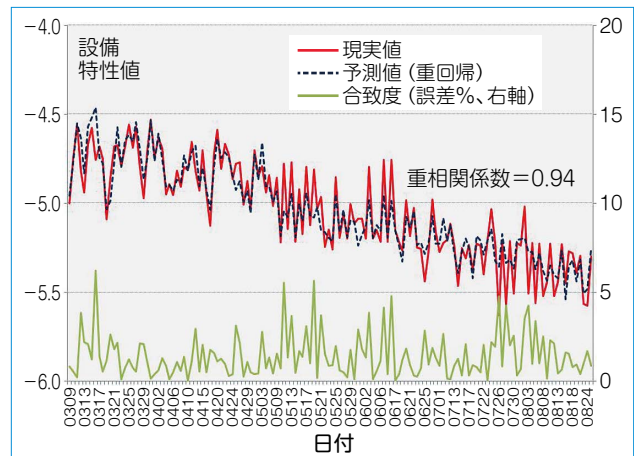


図6 ある保安装置特性値の時系列変化の現実値と予測 (重回帰) 値

用内容を事前に) 予約して頂く」方法も考えられます。多くの鉄道利用者の動きは「定型」「往復」パターンに沿ったものです。そのパターンを提供側はビッグデータを通じた受動的把握に加え、利用者から積極的に明示してもらう、ということです。この予約は利用者行動を拘束しない「緩い」「パターンの」予約ですが、この情報も加えることで予測精度は高まると考えられます。しかし、「予約」は煩わしいでしょう。予約パターンから外れない利用をした場合は料金面でインセンティブを与える方法も考えられます。現状の定期券も利用パターンを予約 (登録) していると解釈でき、定期割引運賃はインセンティブと言えるでしょう。

「全利用者が予約」とは突拍子ない考えと思われるかも知れませんが、飛行機利用者は事前にその個人名・全行程を提供する訳です。その予約情報で航空会社は必要により便や使用機材を変更し輸送効率を上げることが可能です。また利用サービス面では予約情報が有るゆえに、乗継客のために接続便出発を少々遅らして待たせる、といった対応も可能な訳です。この考え方を鉄道にも緩い形で適用できないか、ということになります。極端な例を挙げれば、本来停車する駅でも降車/乗車客がいなければ、駅通過が可能なのです (実際、バスでは行われていますね)。

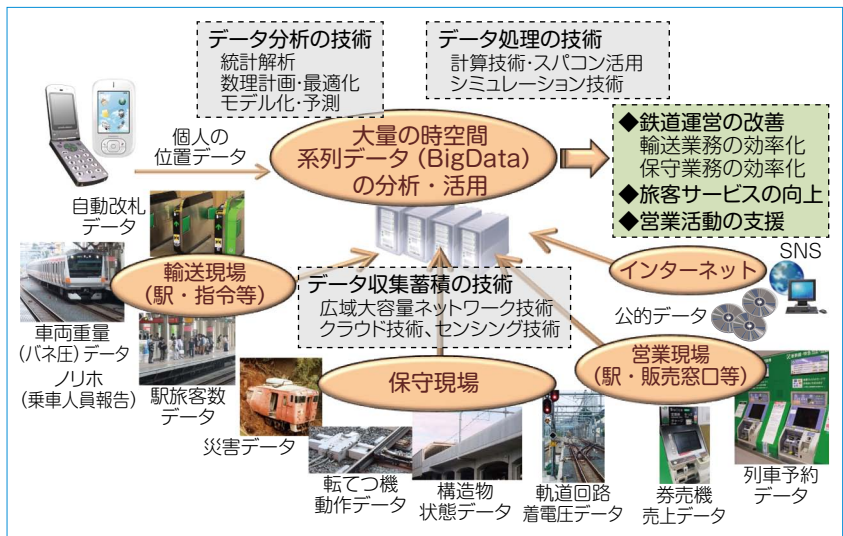


図7 鉄道におけるビッグデータ利用の総合的イメージ

⑤動的スケジュールの周知

現行では提供側がスケジュール(ダイヤ)を「時刻表」という形で公開します。動的計画となれば時刻表がダイナミックに変化する訳で、それを利用者が簡便に得られる必要があります。携帯電話、スマートフォンなどの普及でメディア的基盤は確立しつつあると言えるでしょう。

以上のイメージを図5に示します。

異常時の対応

近年の大災害時に課題になったのが輸送機関の異常時対応です。ビッグデータを通して利用者移動をリアルタイムで大量に収集・追跡できれば、任意の時点・箇所の旅客数、移動ニーズ発生量が予想できます。すなわち、災害などでの輸送停止や停滞時に、どんな事象が発生するか予測可能で、帰宅難民への対応、運転回復・再開はどのような方法が良いか、などの判断の支援につながります。

保守や安全のために

以上、輸送計画でのビッグデータ利用を述べましたが、鉄道のあらゆる分野で活用可能でしょう。例えば設備保守面でも基本的考えは同様です。設備履歴・特性・環境・検査結果などのデータを大規模に収集し、各属性別に正確精緻な

分析ができれば、個々設備の特性変化のモデル化や変化予測の可能性がります。より適切で効率的な保守時期の設定や寿命予測、設備トラブル発生防止に役立てることが可能でしょう。

図6は、その試みの一例で、ある保安設備の健全度変化の現実値と、環境や他設備データによる予測(重回帰)値とを合わせて示したものです。両者にはかなりの合致が見られ、合致度チェックによる設備特性変化の早期検知手法の研究が進められています。

安全についても同様です。安全や事故の管理のために「ハインリッヒの法則」があります。1:29:300という数字を耳にした人は多いでしょう。これは見方を変えれば、数件のデータを見ている限りはリスクの全体像が見えないが、300件集めれば、重故障の1や軽故障の29が推測できる、とも解釈可能です。時空間的に多くの事象を集めて分析し、極端領域での確率分布を求めて、安全の向上に資する可能性があります。

統計解析への指摘に「過去情報を用いる以上、実例が無ければ無力では」というものがあります。大規模地震や災害の予測は典型例でしょう。実例に依存せず極端事象の確率を求める研究も有ります⁴⁾。正規分布では確率分布は中央が山で両裾で急減しますが、裾

の様態に注目し発生データ範囲を超えた領域の確率予測を試みるものです。今後こういった研究もビッグデータ活用拡大に合わせ進展が期待されます。

おわりに

鉄道のさまざまな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7に示します。鉄道業務は複雑かつ専門的で分野毎の縦割りになりがちですが、大規模ビッグデータは全分野で共有し有効活用を図るべきでしょう。

本稿で述べた内容は技術的な側面であり、他の多くの課題があります。個人情報利用はどの程度許容されるか、プライバシー侵害をどう避けるか、という問題があり、鉄道事業者がビッグブラザー⁵⁾になるのか、という批判が出るかも知れません。利便性の面では「時刻表がニーズにより動的に変わる」形態が歓迎されるか、という論点もあるでしょう。利用者が望むのは「自由」でなく、例えば時刻表に「拘束」される形態なのだ、という少々心理学・哲学的議論もあるかも知れません⁶⁾。今後、各方面からご意見も頂戴し、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と思っています。RRR

文献

- 1) WhiteHouse : BigData is a BigDeal, <http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>
- 2) Hadoop.apache.org : Hadoop Releases, <http://hadoop.apache.org/releases.html>
- 3) Microsoft Research : The Fourth Paradigm : Data-Intensive Scientific Discovery, <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>
- 4) E.J.Gumbel : Statistics of Extremes, 1963
- 5) G.Orwell : 1984, 1949
- 6) E.S.Fromm : Escape From Freedom, 1941