

No.4

ビッグデータ収集・解析 技術の動向

「鉄道トレンドウォッチング」第4回では、ここ数年、耳にする機会が多い「ビッグデータ」に着目し、技術の現状と今後の活用について動向を報告します。

■ ビッグデータの対象領域

近年、情報通信技術であるICT (Information and Communication Technology) 技術の活用が広く進められています。ビッグデータ技術については、ICT技術の中に含まれる一つの重要な技術と考えられます。

また、ビッグデータに関係するものとして、IoT (Internet of Things: モノのインターネット) 技術、人工知能技術などがあげられます。IoTによって得られた大量のデータは、ビッグデータ収集・解析技術によって、推定や予測といったことに活用されます。この収集・解析技術には、データの取

集、蓄積、分析、見える化、といった技術も含まれています。

■ ビッグデータの現在に至るまで

「ビッグデータ」という用語は、2010年代に入って使われるようになりました。この用語が出現する前には、「データマイニング」という用語が用いられていました。どちらも基本的には、大量のデータから統計分析などにより有用な情報を見つけ出すことを指しています。ビッグデータのほうが、データ量が大きいというイメージで用いられていると考えられます。

データマイニングという用語が用いられていた時には、お店の売り上げデータから「ママに頼まれてスーパーで紙おむつを買ったパパは、ついでに缶ビールを買う」といったことが予測できました。ビッグデータ技術が盛ん

になるにつれて、こういった新たな知見をさらに得たいという考え方が出てきています。

ICT技術が急速に成長してきたことにより、現在では蓄積できるデータの量、データ伝送速度、データ処理能力などが高い水準に到達しました。また、近年データの情報源となるセンサー技術、SNS (ソーシャルネットワーク) 技術の利用が盛んになっています。このような背景から、ビッグデータを活用し、今まで得られなかった情報を取得することがさらに期待されています。

■ ビッグデータを扱うための技術

ビッグデータを扱うために、従来から用いられるデータ処理技術として、最適化、統計学、機械学習、テキストマイニングなどがあります。特にテキ

ストマイニング技術については、SNSの普及によって、大量のテキストデータを扱うニーズが出てきています。たとえば、SNS上で多く使用されている単語を調べることで、世の中でのようなことが注目されているかといったトレンドの把握ができます。

上記のような従来技術に加えて、NoSQLデータベースがあります。従来のデータベ

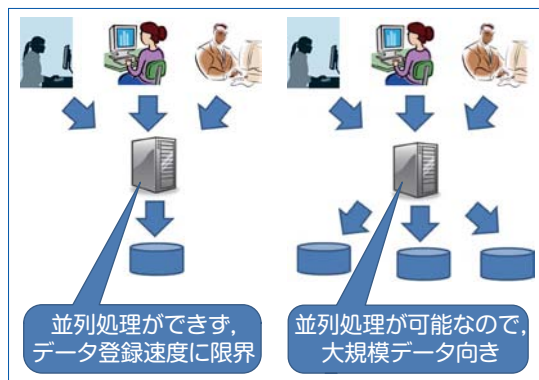


図1 RDBでの処理(左)とNoSQLでの並列処理(右)

表1 ビッグデータの活用例

対象	分析・予測内容
自動車	渋滞状況 災害時の通行可能場所
印刷機	故障時期
構造物 (トンネルや橋)	災害による被害規模 老朽化時期
百貨店	店内における顧客の流動
気象	長期予報から、季節商品 が売れる時期の需要
人の流れ	携帯電話の基地局情報を用いた 時間帯・エリアの状況
鉄道	旅客流動 線路や架線の劣化

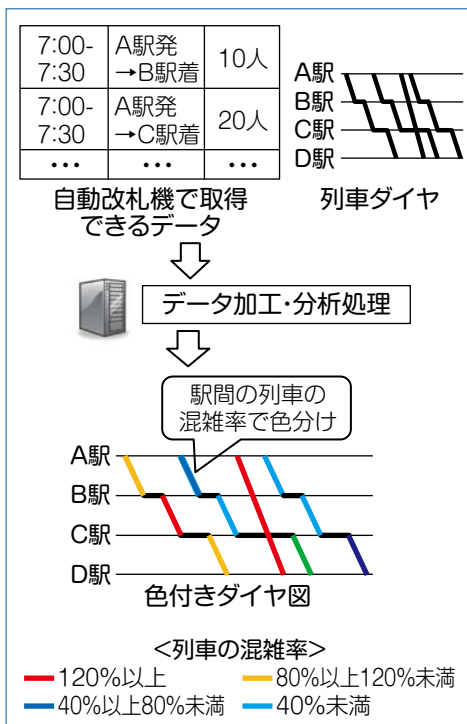


図2 列車の混雑率の推定

スでは、その定義や操作でSQL（構造化問い合わせ言語）というISO規格を用いたRDB（リレーショナルデータベース）が多く用いられてきましたが、複雑なデータ構造を持つので、大量のデータ処理には限界がありました。NoSQLは、ビッグデータ処理用に高速かつ同時並行処理を行いやすくするため、キーと値の対で定義するという、SQLにとらわれない（Not only SQL = NoSQL）シンプルなデータ構造をとることができます。このNoSQLデータベースは、facebookやmixiなどのSNSで使用されるようになってきました（図1）。

■ 近年のビッグデータの活用例

ビッグデータはさまざまな分野で活用されています。表1にビッグデータが活用されているおもしろいものをあげます。予測されるものとして、商品や食品の需要、構造物などの劣化、自動車や人の滞留状況といったものがあります。

■ ビッグデータの課題

ビッグデータ技術では、さまざまな



図3 営業列車での軌道検測

手段で大量のデータを活用できませんが、データを扱う上で直面する、いくつかの課題があります。

まずは、大量のデータを集めていても、どのように扱えばよいかわからないという課題があります。ビッグデータという魅力的な言葉にひかれ、データを分析して、何かを見つけ出したいという要望があります。しかし、大量にデータ

を集めても、具体的に何を^い見出せばよいかかわからず、データの収集・分析を断念することがあります。

また、ビッグデータに取り組みたいと思っても、思うように大量のデータを集められないといった課題があります。鉄道では、設備の故障など障害のデータが大量に集まることは少なく、分析するのに十分な量のデータを得られないことが多いと言われています。データの分析結果から事故を未然に防ぐことができれば理想的ですが、過剰に警報がでると、場合によってはその警報を見過ごす危険性もあります。

■ 鉄道への適用事例

ここでは、ビッグデータの鉄道への適用事例を2つあげます。

まず、改札機データを利用することによる旅客流動推定があげられます。旅客の改札機通過データを利用して、列車に乗車する人数の推定が可能となり、ダイヤ改正時の参考データとして活用されています（図2）。

また、営業列車にセンサーを付けて、列車自身、およびレールや架線など周

辺設備の状態のセンシングが行われています。これらのデータは従来得られなかったデータであり、大量かつ継続的に蓄積することによって、異常の予兆を早期に検知することが可能になります（図3）。

■ 今後の鉄道への活用

現在、鉄道設備は、定められた周期ごとに検査を行うことで管理されています。たとえば、橋りょうやトンネルなどの構造物の目視および打音による検査は、2年に1回、また、在来線の車両の交番検査（分解はしないが車両基地などで詳細に行う検査）は3か月に1回行われています。これらの検査周期は、国の基準などに基づいて設定されています。今後はデータを大量に収集することで、適切な検査周期を求めることが可能になると考えられます。

また、リアルタイムにデータを取得することにより、線路や架線などは、検査周期ごとの管理から、設備の状態に応じた管理を目指した取り組みが行われています。改札機データやインターネットのデータは、現在はいったん蓄積したものを後日分析して、旅客需要の予測に用いています。これらのデータをリアルタイムに収集・解析することが可能になれば、お客さまの需要に応じた列車の設定が可能になります。

今後も、ビッグデータを活用することにより、鉄道の設備検査が適切に行われ、お客さまが便利に列車に乗ることが期待されます。

（渡辺義大／企画室 戦略調査）