



SEMINAR REPORT

M2Mデータを活用するビッグデータ処理技術と今後の方向性



日本電気株式会社
ナレッジ研究所
エグゼクティブエキスパート
福島 俊一 氏

はじめに

M2Mによって実世界のビッグデータがたくさん集まってきます。まず、集まったビッグデータを分析することでどのような価値が生み出されるか、についてお話ししたいと思います。次に、どうやってそのような価値を生み出すか、ビッグデータというデータの山を、ビジネスや問題解決に活用できる宝の山に変える技術についてお話しします。これは機械学習の技術が中心になります。

そして、より高い価値を生み出すための技術チャレンジは何か、つまり、データから何か分析結果を導き出したとき、顧客がより大きな対価を支払ってくれるような結果を生み出すためには、技術的に何を追求していかなければいけないかについてお話しします。この追求すべきポイントは三つあると思っています。これらのポイントをしっかり伸ばしていくことで、ビッグデータがより役に立つものになり、ビジネスも拡大していくと考えています。

その後で、具体的に実世界のビッグデータを分析する最新の事例についてご紹介します。ひとつはカメラの映像データを対象とした映像解析技術、もうひとつは主にセンサデータ等を対象としたデータマイニング技術です。

実世界ビッグデータから生み出される価値

最初に、実世界ビッグデータから生み出される価値について

お話ししたいと思います。

M2Mの世界は、図1に示すように色々な機器がネットワークに繋がっていて、様々な実世界のデータがリアルタイムに流入してくる世界です。データの分析や活用という側面から眺めると、センサから温度や位置情報などの様々な種類のデータが入ってきて、それはある意味で実世界を様々な観点から捉えたものだと言えます。今どういう状況にあるのが、より正確に把握できるようになり、それに基づいた先の予測も、より正確にできるようになっていきます。これがM2Mから得られるデータの意義ではないかと思っています。そして、そのようなデータを分析することで、ユーザの抱えている問題あるいは社会が抱えている問題を解決していけるようになります。

では、実世界のデータが色々集まってきて、それを活用した事例、社会の問題を解決するような事例をいくつかご紹介しします。当社が関わっている案件を中心に、広いマクロな視点から順にミクロな視点まで、簡単にご紹介します。

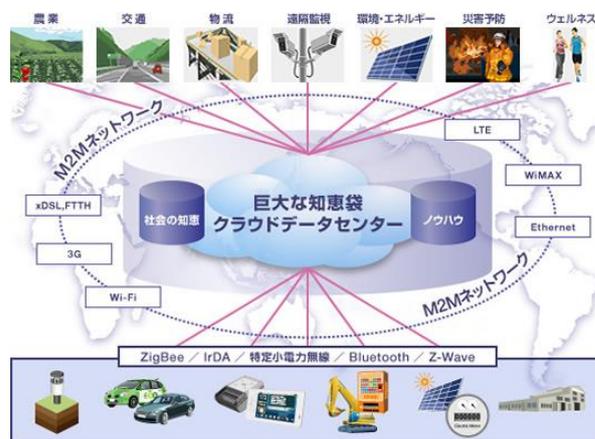


図1 M2Mの世界

(http://jpn.nec.com/solution/m2m/whats_connexive.html から引用)

■ 海底からのビッグデータ

海底には、地震や津波に関わるデータを取得するためのセンサが埋め込まれたケーブルが敷設されています。それによって海底の状況を知ることができ、地震や津波をリアルタイムに検

知して、より迅速な対処ができるようになります。

[参考] [http://jpn.nec.com/profile/mitatv/discover/01/
http://jpn.nec.com/techrep/journal/g09/n04/pdf/090413.pdf](http://jpn.nec.com/profile/mitatv/discover/01/http://jpn.nec.com/techrep/journal/g09/n04/pdf/090413.pdf)

■ 宇宙からのビッグデータ (図2)

宇宙には、大気中の湿度を測る衛星「しずく」や、温室効果ガスなどを測る衛星「いぶき」が打ち上げられています。それらの衛星から地表の状態や大気の状態をセンシングして、農業に影響する環境変化をいち早く知って予め対処する、地球温暖化対策するなど、宇宙から得られるデータの活用が考えられます。

[参考] <http://jpn.nec.com/ad/cosmos/index.html>
<http://jpn.nec.com/techrep/journal/g11/n01/pdf/110119.pdf>

■ 人工衛星により、地上の様々な状態がセンシング可能に

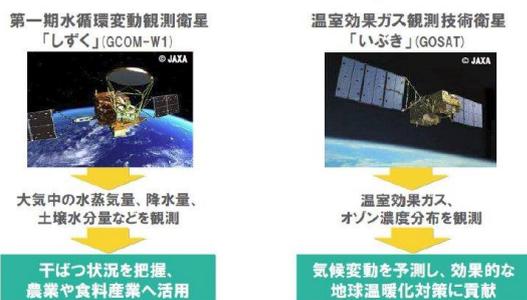


図2 宇宙から：人工衛星から地球観測

(<http://jpn.nec.com/bigdata/relate/ict-conference2013.html> から引用)

■ 車の動きからのビッグデータ

海底や宇宙というやや遠い世界から入りましたが、身近な例では車が挙げられます。車にGPSが搭載されて、どこを走行しているか位置情報をリアルタイムに追跡でき、車の移動速度もリアルタイムに知ることができます。あまり車が移動していないようなら、渋滞が発生していると予想され、それに応じて信号機をうまく制御するなどの活用が考えられます。さらには位置情報だけでなく、ワイパーの動きを見ることで、どこで雨が降っているかも分かります。この様に車から得られる情報を分析することで、道路の状況や場合によっては天候の情報も分かってきます。

[参考] <http://jpn.nec.com/techrep/journal/g09/n03/pdf/090325.pdf>
<http://jpn.nec.com/techrep/journal/g11/n04/pdf/110420.pdf>

■ 人の動きからのビッグデータ (図3)

センサではなくカメラに着目してみると、最近では防犯カメラが屋内外あちらこちらに設置されています。映像データは犯罪防止にも使われることが多いですが、店の中の映像データであれば、店内の改善に活用することも考えられます。

図3のように、店内に複数の監視カメラがあって、オーバーラップするように映像が撮られているならば、店内の人の動きを追跡することが可能です。背景画像との差分から何か動くものがあることがわかり、異なる角度から撮られた結果を対応付けて三角測量の原理を使うことで距離も分かり、姿形や顔らし

きものの位置から人が動いていると判断できるという仕掛けです。

[参考] <http://www.nec.co.jp/solution/tvsol/feature.html>
<http://jpn.nec.com/techrep/journal/g11/n03/pdf/110303.pdf>

■ 店舗内のカメラ映像から、購入しなかった来店者の行動を分析 POSデータ(購入した来店者)と組み合わせて、品揃えや陳列を改善

顔画像、動線、立ち寄り時間の分析から、これまでわからなかった「買う前」行動を発見

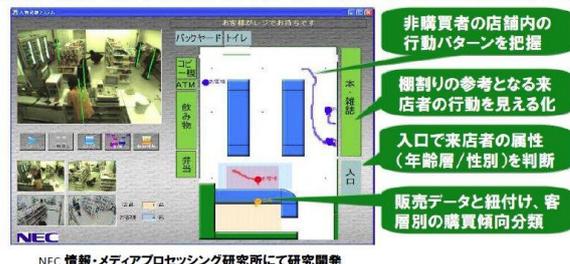


図3 人の動きから：動線分析マーケティング

従来、コンビニやスーパーなど小売業におけるPOSデータ分析では、商品を購入した顧客について、例えば、20代の女性がこの様なものを買った、40代の男性がこの様なものを買ったという情報をレジで記録することで、今後の品揃え等に活用してきました。しかし、何か欲しいものがあって店に入ってきて、ある棚のところへ見に行ったが無かったので、そのまま何も買わずに出て行ってしまった様なケースは、POSデータには反映されません。人の動きを解析すると、このような売り逃しのケース、買わなかった人に関する分析も可能になります。つまり、人の動きの分析がマーケティングのために新たな価値を生んだということになります。

■ お金の動きからのビッグデータ

以上の例とは異なり目には見えないものですが、実世界の重要な動きとして、金の動きというものがあります。銀行や金融機関のどこから金が入ってきて、どこへ出ていったという情報を繋ぎ合わせて分析すると、金融機関の間にはどのような関係があって、例えば、どこかの金融機関が破綻すると、その範囲がどこまで及ぶかといったことが事前に判断できるようになります。

以上、マクロなところからマイクロなところまで駆け足で眺めてきましたが、実世界のデータが色々な形で収集できるようになってきており、それを分析することで、ユーザが抱えている問題や社会の問題が解決できるようになってきていることをご理解いただけたものと思います。

従来はビッグデータというと、グーグル、アマゾン、あるいはツイッターのようなネット上のサービスに集まるデータを分析する、北米的なビッグデータの世界のことでした。しかし、それとは違って、センサやカメラのデータから実世界の活動が色々収集できるようになってきて、これを分析することで、ネット上のビッグデータの分析とはまた違った様々な価値が生み出されるようになってきました。ここまでご紹介してきた事例はそういった意味合いで、話を受け止めていただければと思います。

データの山を宝の山に変える技術

データから価値を生み出すということ、もう少し技術的な観点から掘り下げてみましょう。

図4では氷山に例えましたが、以前は見えるのが氷山の一角でしかなかったのが、センサやカメラが携帯・自動車など色々なものに取り付けられるようになり、今まで見えていなかったところまで見えるようになってきました。ただし、生のデータがばらばらと見えてきても、それが必ずしも価値を持つというわけではありません。見えてきたデータをうまく解析することで役立つデータに変える、意味を持った宝に変えるというのが今回のお話の中心であり、いまビッグデータが注目されている理由になります。

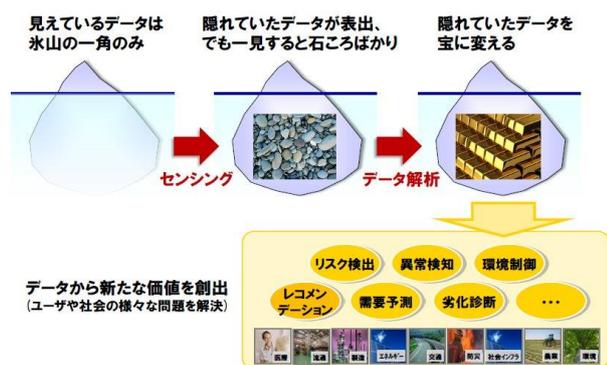


図4 ビッグデータの時代：データの山を宝の山へ

ビッグデータの処理技術としては、大量のデータを如何にして高速に処理するかという問題がありますが、この問題に関しては、ストリーム処理、並列処理、Hadoopなどの色々な方法が開発されて、対処できるようになってきました。ここで重要になってくるのは、単に大量のデータを高速に処理できるだけでなく、データからどのような価値が導き出せるのか、データの解釈結果をどれだけ有効に活用できるかという面です。それを突き詰める技術に関して最新のトレンドをお話したいと思います。

宝の山に変わったということは、ユーザーや社会の様々な問題を解決できるようなものに変えられたということです。例えば、リスクを検知できる、異常を検知できる、危険を事前に予知して対処できるようになるとか、あるいは、モノが今後どのくらい売れるのか、エネルギーが今後どの程度使われるのかなど需要を予測して、それに合わせて適切に制御できるようになるということです。

データの山を宝の山に変える技術は機械学習です。まず、観測されるデータの山を過去まで遡って眺めることで、そこに隠れている規則性を見つけることが機械学習のポイントとなります。その規則性に基づいて、新たに入力される観測データを解釈し、意思決定やアクションに結び付けることで問題解決を図ります。

もう少し具体的なイメージを示しましょう。図5の中央下のグラフは、電力データ、温度データ、商品の売れ行きなど、時間に合わせて観測値が変化するデータです。青線の部分が過去

のデータで、その変化の仕方に規則性を見出せたとき、時間軸方向に伸ばせば、将来どの様に変化するかを予想できます。また、この規則性から外れたデータが出てきたら、何か異常が起きているかもしれないと判断できます。この図は二次元で表したので、いとも簡単、見れば明らかですが、実際にはとても複雑な多次元データなので、これを人間の直感で見つけるのは甚だ難しく、コンピュータの力で規則性を見出そうというのが機械学習技術です。



図5 機械学習技術：データの山を宝の山に変える技術

また、図5の右下のグラフは、いまお話しした時系列データとは別のタイプの例になります。色々な属性を持ったデータをどこかで線引きして、こちらは正解、こちらは不正解、あるいは、これは何かかに該当している、該当していないなど、判別や認識を行うケースです。このようなケースについても、過去のデータから線引きに関する規則性を見出して、その規則性に基づいて、新たに入ってくるデータの判定が可能になります。

要するに、過去のデータを見て、そこに潜んでいる規則性を見つけることで、それは色々な条件値だったり、将来予測だったり、物の判定だったりしますが、意思決定・アクションに結びつけることができるというのが機械学習というものです。

機械学習技術によって、データから新たな価値を導き出すことはできますが、往々にして人間が一瞥して分かる、機械が大変な計算をして導き出した結果でも人間やユーザーや社会から見るとさほどのことはないということもあります。本当にお金を払ってもらえるような結果を出せる技術に持っていかないと意味がないわけです。

そのためのポイントは三つあると思っています。一つめは、人間には見えないものまで見える化することです。もう一つは、人間が捉えることが難しいような規則性や関係を導き出せることです。「そうだったのか」ということでお金を払ってもらえるような技術になります。三つめは、人間の能力を超えた規模にまで拡大することです。とても人間では扱えないような規模でも人間と同等あるいは人間以上の働きをしてくれるならば、お金を払う気になってもらえるということ です。

以上は抽象的なレベルの話でしたから、本当に人間には見えないものまで見える化するというのはどのようなことなのか、映像解析の例でご紹介します。さらに、人間が捉えることが困難な規則性・関係を見つけ出すこととは如何なることなのか、データマイニング技術の例でご紹介します。

より高い価値を生み出す映像解析技術

防犯カメラが色々な所に設置されていますが、それを使って特に都市や施設の安全を見守るために、不審者や不審な行動を見つけるアプリケーションが実用化されつつあります。例えば、何か事件が発生して、その事件に遭遇した人が「犯人はこんな感じの人でした」と告げると、それに該当しそうな人を過去や現在のカメラ映像データからピックアップする、あるいは、指名手配犯をブラックリストとして予め登録しておいて、それに該当しそうな人が出てきたら、見つかったことを知らせる。実社会の大量の映像データが集まってくると、それを解析することで、防犯の目的でとても役に立つアプリケーションを作ることができ、価値があるということになります。

映像解析の中でも顔の認証は既に普及してきています。空港やレジャー施設などでドアを入るときに、この人だったら入ってもよいといったローカルなゲート認証のレベルで実用化されているのはよくご存知だと思います。

それに対して、広い範囲でたくさん人が集まっている状況で、指名手配の人を見つける様なケース、いわば実世界映像のビッグデータへと広げていくと、ゲートでの顔認証に比べて非常に難しくなってきます。図6に、特に難しくなるポイントを三つほど挙げました。すなわち、顔の向きが正面ではない、顔の位置が動いて追跡が必要、監視カメラで遠くから撮っているので画像が低解像、という三つの問題が、認識が非常に難しくなる要因になります。このような問題への取り組みがなされてきた結果、顔の向きや環境が様々に変化してもしっかり人を区別できるとか、低解像な画像でもきちんと人を認識できるとか、映像監視の技術は大きく進化しており、映像のビッグデータ解析は、こんなことまでできるようになったのかと感じていただけるような相当高いレベルまで達しています。

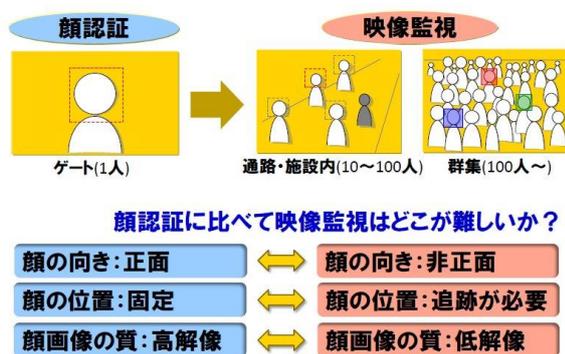


図6 映像ビッグデータ解析：顔認証から映像監視へ

このような映像解析の出発点は顔認証ですので、それをおさらいした上で、そこから技術がいかに進歩したかをお話します。顔認証というのは、図7に示すように、予め色々な顔が登録してあって、そこに新たに顔が入ってくると、この中の誰かを判断する技術です。一見すると、マッチングするだけの簡単な技術のように思われるかもしれませんが。

顔を写した画像・映像にもとに、それが誰であるかを判断する技術
(事前にDBに登録してある人の中の誰であるかを見つける)

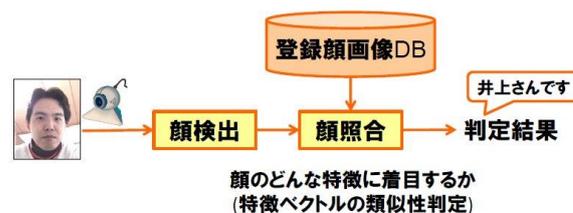


図7 顔認証とは？

しかし、ゲートできちんと正面を向いて顔を認識するのではなく、もっと色々な環境で認識する必要があることを考えると、顔について非常に多くのバリエーションが生じてきます。図8に示すように、顔の向きが異なる、眼鏡を掛けている／掛けていない、笑っている、照明の条件が変わる、年を重ねると顔つきが変わる、などの様々な変動が生じます。実世界の色々なシーンで顔認証しようとする、環境や対象の変動があっても同じ人物だと判定できることが、当然のことながら要求されます。これに対応するため単純に判定条件を緩めてしまうと、別人を同一人物と判定してしまう問題が生じるので、そのさじ加減が難しいのです。

具体的な事例をデモビデオで見ていただいた通り、眼鏡を掛けないで登録していた人が、眼鏡を掛けて入ってきても、さらには笑って入ってきても、同じ人だと自信を持って判定できています。



世界最高権威ベンチマークテストNIST MBE@2010でダントツ1位！

図8 顔認証の難しさは？

また、経年変化によって過去に登録した顔と今の顔が違っていても、それでも機械は正しく本人だと判断できています。3年前、5年前、16年前、23年前の顔でも本人だと判定しました。23年前の写真となると、人間でも二つの写真が同一人物だと判定できるかわからないと思いますが、機械の方は自信をもって二人が同一人物だと判定しました。

年を取っていたり、顔の向きが変わっていたり、眼鏡を掛けたりと、この様な変化に対してどこに着目すればその人かどうかを区別できるのか、特徴をうまく捕まえて判定しています。人間には本人かどうか分からなくなってしまう様なケースでも、機械の方が却って的確に特徴を判断して、同一人物かどうかを判定しています。ある種、人間の能力を超え始めていると感じています。

次なる進化は、防犯カメラ・監視カメラの画像からの顔認識です。

防犯カメラの広範囲映像からの顔認識のデモビデオを見ていただきましたが、防犯カメラの画像には同時に多くの人が映っています。あらかじめブラックリストに登録された人物が現れたら検出するというアプリケーションでは、カメラ映像中から顔の位置をリアルタイムに検出して、ブラックリスト中の人物と一致しないかを判別します。実際は同時に複数の人を、これは誰だ、これは誰だと区別して、その顔を追従しながら判定しているのです。このようなケースを想定すると、先ほど述べたように、表情が変わったり、顔の角度が変わったりする様な変動に追従して、顔の認識をしなければいけないことを分かっていただけだと思います。

このようなケースでもう一つ重要な問題は、遠くから撮った画像であるため低解像度だということです。誰の顔であるか識別するのに十分な解像度を確保することが必要になります。そのための技術は超解像と呼ばれている技術で、もとの画像からより精細な画像を自動的に作り出す画像処理技術です。防犯カメラで遠方から撮ったために入力画像がぼやけていてそのままでは顔認識ができなさそうなケースであっても、超解像技術を使うと認識できるような品質に上げることが可能になりつづあります。

顔に加えてナンバープレートの例をお見せしましたが、図9の例のように、かろうじて文字が読める程度のもや人の眼でもまったく判読不能なものが、それが超解像技術を使うと、この様に読めるようになります。特にこの例はびっくりされることが多いのですが、最先端はこれぐらいの技術レベルに達しつつあります。

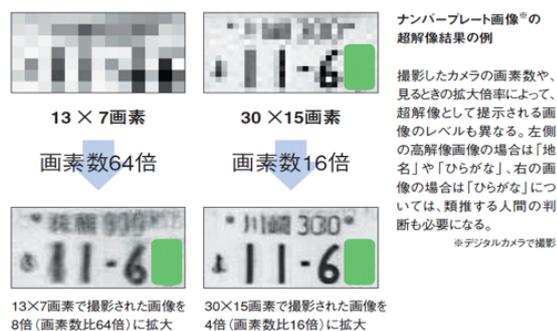


図9 超解像技術：低解像カメラ画像を高解像度化

(<http://jpn.nec.com/rd/innovation/feature/201203/index.html> から引用)

図10は超解像技術の種明かしです。このとき対象物を顔やナンバープレートに限定しているのが一つのポイントになります。限定しているので、これに関する学習を事前におけます。どうやって学習しているかというと、高解像なナンバープレート画像から低解像な画像を作っておいて、それを細かき断片に分けて、高解像な画像と低解像な画像のペアを予め辞書として登録しておきます。実際に低解像な画像が入ってきたら、この辞書を登録時と逆に適用して、低解像な画像断片は元々この様な高解像な画像断片だったと逆変換をすることで、高解像なナンバープレート画像が作れます。

実際のアプリケーションでは、ナンバープレートや顔のよう

に対象を絞るという想定にあまり無理がないケースは結構あるので、学習型超解像技術の実用化は進んできています。

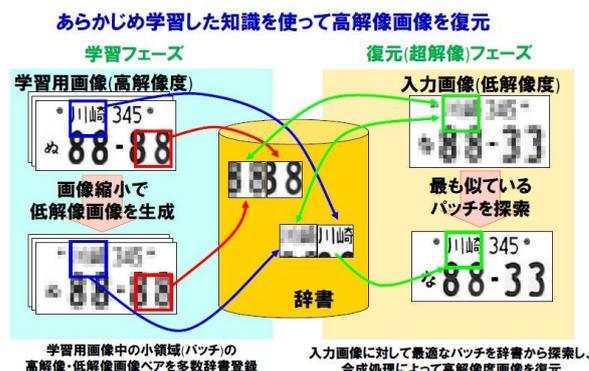


図10 学習型超解像方式

(<http://jpn.nec.com/rd/innovation/feature/201203/index.html> から引用)

[参考]

<http://jpn.nec.com/rd/innovation/feature/201203/index.html>

<http://jpn.nec.com/techrep/journal/g12/n02/pdf/120217.pdf>

以上は、実世界のビッグデータから生み出される価値の一つとして、これから益々大きくなっていくのであろう監視カメラや防犯カメラの映像のデータを使ってできること、特に人の追跡や人の判定はここまでできているということをお話ししました。

超解像を含めていろいろな具体例を見ていただいておりますが、人間にはとても見えなかったものまで機械では見えてしまう、人間には同一人物だと分らなかつたものが機械の特徴判定では分かってしまう。要するに、人間には見えないものまで見える化することで、人間にはできないことまで実現してくれてすごい、お金を払ってもいいと思ってもらえるようになります。

これが特定の小さい環境でしか実現できないのでは、さほど役に立たないかもしれませんが、人間の能力を超えた規模まで拡大させることでいっそう役に立つようになります。監視映像を人が張りついて見ていけば、何かおかしいことが起きれば気づきますが、実際問題として、人が24時間べったり張りついて見ているのは非常に耐え難いことです。また、カメラが何千個もあったとしたら、とても人間では追い切れません。あちこちから集めてきた大量の映像データを機械で分析することで、問題が起きているケースを見落とさず素早く察知できます。要するに、人間には見えないものまで見える化し、それを人間の能力を超えた規模までスケールさせることが、価値を大きく高めるためのポイントだということです。

より高い価値を生み出すデータマイニング技術

ここからは最先端のデータマイニング技術に話が変わります。こちらは多様なセンサデータから集まる大量の観測データを扱うものになります。

原子力発電所、プラント、トンネル、橋梁などの社会インフラを含む大規模物理システムは事故が起きると大きな問題にな

りますが、老朽化も進んでおり、早めの対策が求められています。事故を起こす前に予兆・リスクを事前に検知するために、対象システムに多数のセンサを取り付けてデータをまとめておくことが行われています。あるセンサデータの値がおかしくなったことで問題発生が直ちに分かるようなケースでは、特段データマイニングするほどのこともなく、あるデータが閾値を超えたら異常と判断すればよいことです。しかし、多種多様なセンサデータが複雑な関係をもって時間変化していると、どこで異常が起きているのか、一目見ただけでは分かりませんし、単純な閾値判定では済みません。

そのため従来はかなりの専門知識が必要でした。例えば、ここに取り付けた温度計がこれぐらいの値になって、ここのバルブがこれぐらいの圧力になると、この様な事態が発生している、これはかなり危険な状態ではないか、というような知識が必要です。個々のセンサデータの意味やデータ間の因果関係をよく分かっている専門家でない、何か異常が起きているのか判断できないと考えられていました。しかも、このような専門家はあまり多くないことも問題でした。

これに対して、インバリエント解析技術（SIAT：System Invariant Analysis Technology）という新たに開発した技術を使うことで、専門家に頼らなくても、専門知識がなくても、異常が起きていることを発見し、障害の未然防止・事前対処が可能になりました。しかも、実際に実験してみた例では、専門家よりも早く見つけたというケースも時折ありました。

従来専門家はセンサの意味をきちんと理解した上で、異常かどうかを判定するという考え方が、インバリエント解析技術は、特段センサの意味を知らなくても、網羅的にデータ間の関係を分析するデータマイニングによるアプローチをとったものです。図11はある発電所の例で、センサは3,000個～5,000個あり、これらのデータ値は刻々と変わっていきます。あるセンサと別のセンサの1対1の関係をシラミ潰しに見ていくと、このセンサのデータが上がっていくと別なセンサのデータが連動して上がっていくという、理由は分からないが関係が深いペアが少しずつ分かってきます。これを関係の不変性（Invariant）と呼びます。非常に関係が深いセンサ同士が分かった結果、あるときに正常な関係が崩れると、例えば先ほどの例では、片方のデータが上がっているのに片方のデータが上がらなくなったということが起きます、異常が起きているのではないかとということになります。

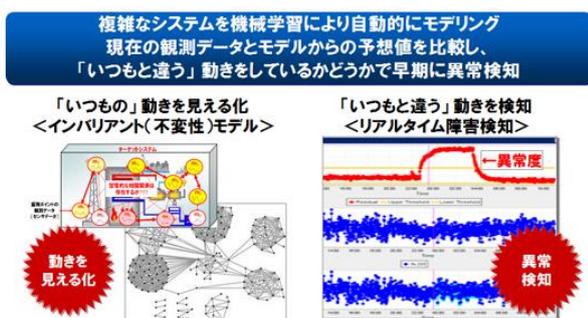


図11 インバリエント解析技術（SIAT）

(<http://jpn.nec.com/bigdata/analyze/index.html> から引用)

従来は1個1個のセンサのデータの意味を考えて異常かどうか

かを判断していたのに対して、機械学習による網羅的な相関分析によって、その関係の崩れを見つけるという、専門知識を全く使わないアプローチも可能になったということです。ときには専門家が見落としていたようなケースまで異常を見つけることができ、人間の判断能力を超えた複雑なケースについても何らかの規則性を見つけ出せるというレベルで、異常を事前に察知する異常検知・障害予防という新たな価値が提供できるようになります。

この様な考え方は、今は原子力発電所で実際に適用して有効性を確かめた段階です。トンネル、橋、航空機、水道管などの社会インフラにて、色々な問題を事前に見つけ出すのに展開していける技術ではないかと期待しています。

[参考] <http://jpn.nec.com/bigdata/analyze/index.html>

<http://jpn.nec.com/websam/invariantanalyzer/>

<http://jpn.nec.com/profile/mitatv/discover/16/index.html>

ここからは、もう一つ別の最先端のデータマイニング技術を紹介します。センサで集めた実世界のデータは非常に複雑で、データマイニングでルールを見つける従来の考え方ではうまくいかないことが分かってきました。その分かり易い例として、電力需要予測を取り上げます。

図12において、緑線が実際に測定されたデータです。図では1週間分しか見えないように見えますが、実際には2ヶ月分のデータを学習させました。その学習期間の実測データに対して回帰分析を使って関数を当てはめてみたのが青線です。右側の予測期間の青線は、この関数に基づいて今後はこのような値になるだろうと予測した結果を意味します。

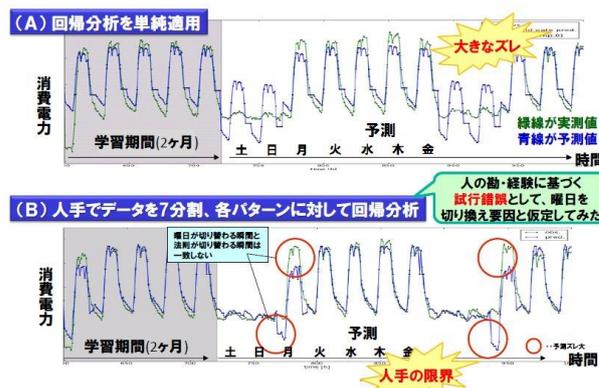


図12 電力需要予測の例：従来はどうしていたか？

しかし、実際の変化は緑線で、(A)では予測値である青線に対して緑線が全く合っていないところがあります。これは何かというと、単純に回帰分析だけでは精度を出せないということです。土日と平日では傾向が違うのが見た目にも明らかで、単純に一つのルールではカバーできない状況になっていることが分かります。

こういう状況に至ったとき、従来はどの様にしていたかという、あるパターンで幾つかのルールが切り替わっていると考えて、例えば曜日で分けてみます。曜日ごとにデータを分割して回帰分析を行い、それらを繋ぎ合わせたものが(B)です。人が勘と経験を働かせた試行錯誤によって(B)は、多少精度がよくなっていますが、やはりうまく予測できていません。

それではどうするのかというと、従来のやり方では試行錯誤が延々と続きます。曜日で分けたのではうまくいかないのなら、曜日を更に時間で細かく場合分けしてみる。電力だから天気の影響しているのではないかと、天気別にも分けてみる。どの様な場合分けをするとルールの変更に対応して良いフィッティングとなるのか試行錯誤を繰り返すしかなかった。というのが一筋縄ではいかない実世界データの分析の実状でした。

今回、新しい技術として異種混合学習というのを開発しました。この技術は場合分けと場合毎の最適ルールの発見を同時に自動的に行ってしまいう画期的な技術です。

この異種混合学習技術を用いた結果が図13の上側のグラフです。この結果によると、人が曜日別7つに場合分けしたり、それを時間や天気で分けたりと、あれこれ細かな場合分けを試行錯誤したのが、この技術を使うと、いちばんうまくフィッティングする場合分けとルールが自動で見つけられ、しかも、それが三つのルールで十分だったということです。人が一生懸命頑張っても誤差が5.8%ぐらいあったのが、この新技術が自動で見つけた結果は誤差2.7%ということで、人間が試行錯誤するよりも、人間の能力を超えた高い精度が出せたということです。

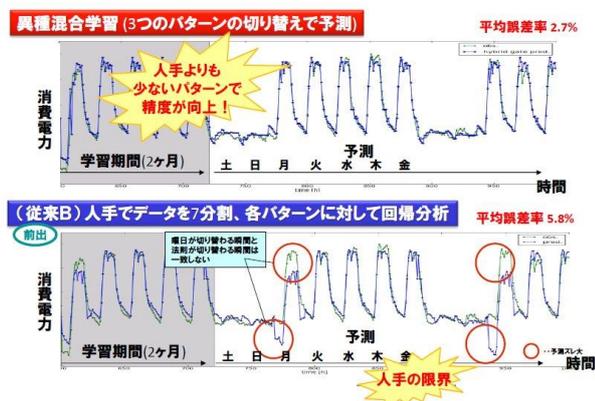


図13 電力需要予測の例：異種混合学習技術を適用すると

ここで、また価値の大きさという視点に戻ってみると、誤差が5.8%から2.7%になったぐらいで、大きな対価を払ってもらえるだろうかということになります。

色々人の話を聞いてみると、精度は人間並みもしくは人間よりも少し高い程度で十分だけれど、大規模にスケールできることが高い価値をもつケースが多いことがわかってきました。異種混合学習技術は、精度が高いというだけでなく、複雑なデータでも人間が介在しないで自動で規則性を見つけ出せるというのが特長です。複雑なデータから精度良く自動で規則性を見つけ出せるということは、色々な条件に対して同時にたくさんまとめて調べられることになります。

例えば電力需要の予測では、発電所のところでトータルどれだけの需要があるかを求める問題ならば、人間が一生懸命考えて答えを出せるならそれでも足りるかもしれません。しかし、各家庭や各オフィスでこの程度は電力を使うから、各地域でこの程度は電力を使う、この地域は電力供給を少し抑え目にした方がよいなどと、場所によってきめ細かに条件を変えるには、各家庭や各オフィスの需要を予測しなくてはならない。そうすると、この家庭ではこれぐらい使う、あちらのオフィスではこれぐらい使う、といった予測を毎日もしくは毎時やらなければ

いけなくなって、何千個、何万個という異なる条件で予測の計算をするのは、人間にはとても無理な相談です。

別な需要予測の例として、コンビニやスーパーなどでの生鮮食品の仕入れ問題があります。賞味期限のあるおにぎりなどに関して、発注量が多すぎると売れ残って廃棄しなければいけない羽目になり、少なめに発注するともっと買ってもらえる人がいたのに品切れになってしまったというようなことが起きるので、できるだけ適切な需要予測をして、発注したいわけです。

従来は、店長や経験深くて勘のよい発注担当者が、よくよく考えて発注数を決めるというやり方だったかもしれません。人間がすることなので、幾つかの売れ筋商品については適切な発注量を決められたかもしれませんが、それ以外の商品は手が回らず適当に発注するしかないこととなります。

異種混合学習技術では、自動化により、色々な条件のケースを一気に予測することが可能になりました。店舗によっては商品の売れ方は異なるので、全店舗×全商品という何千、何万もの組み合わせに対して、それぞれ明日はどの程度売れるか、個別のきめ細かな予測ができるようになります。精度が少し上がることも、自動で複雑なケースの予測ができることで、様々なケースを一気に自動処理できることが非常に大きな価値になることがわかってきました。

[参考] <http://jpn.nec.com/bigdata/analyze/index.html>

<http://jpn.nec.com/techrep/journal/g12/n02/pdf/120219.pdf>

M2Mを介して実世界の色々なデータが集まってきて、その集まったデータからルールを導いて予測や条件を見つけ出すとすると、実際は単純な一つのルールが当てはめられるようなものではなく、複雑なケースをうまく場合分けして、それぞれについてルールを見つけ出すことを自動化することが必要になってきます。ここまでできるようになってくると、実世界からM2Mで集まってくる複雑な振る舞いをするデータを活用できる場面が広がっていくと思います。

以上、人間に捉えるのが困難な規則性・関係も見つけ出している事例、規則性を見つけ出すことだけではなく、人間ではできない多くの組み合わせのケースについて一気に予測してしまう、つまり、人間の能力を超えた規模までスケールさせた事例をご紹介します、そのようなポイントを追求することで非常に高い価値が生まれることをお話ししました。

おわりに

M2Mが流通させる実世界のビッグデータから色々な価値が生み出されます。集まったデータが実世界のより正確な状況把握や、より正確な将来予測を可能にし、ユーザや社会の課題を解決するような価値を生むということです。実世界のビッグデータは、実世界の状況を分析・予測できる源として価値が大きくなってきているということです。

二つの分野でその具体的な例をお話ししました。映像解析によって安全を見守るといった価値が提供され、データマイニングによって需要予測や異常検知という価値が提供される事例を、NECの最新技術とともにご紹介しました。

その中でも触れたように、ユーザや社会に提供する価値を一層大きくするために、技術開発の方向性としては、①人間には

見えないものまで見える化する、②人間に捉えるのが困難な規則性・関係まで見つけ出す、さらには、③それらを人間の能力を超えた規模まで拡大させる、という三点が重要で、その方向で価値を高めていくのが、今後進むべき方向ではないかと考えています。

本講演録は、平成25年12月13日に開催されたSCAT主催「第91回テレコム技術情報セミナー」のテーマ「M2Mデータを活用するビッグデータ処理技術と今後の方向性」の講演要旨です。

*掲載の記事・写真・イラストなど、すべてのコンテンツの無断複写・転載・公衆送信等を禁じます。