

■ ビッグデータ実例（TBSラジオ「日本全国8時です」2012. 10. 18）

8月の最後に、いま情報社会で話題になっているビッグデータ技術について御紹介しましたが、そこでは技術的な内容が中心でした。

そこで今日は、すでに実際の社会に浸透している事例によって、その技術の威力を御紹介したいと思います。

阪神タイガースフアンの森本さんにとって今年は残念な年で、チームはクライマックスシリーズにも残ることができず、入場者数の対前年比が12球団でもっとも落ち込んだ球団になってしまいました。

一方、日本ハムは監督経験もなく、失礼ながら7年間の現役時代には際立った成績も残さず、引退してからは解説者として仕事をし、20年以上現場から遠離っていた栗山英樹監督がチームを優勝に導きました。

この差の原因を暗示する映画があります。

アメリカでは昨年公開され、日本では2日後にWOWOWで放送予定の「マネーボール」というアメリカ映画です。

これはアメリカンリーグ西地区に所属する「オークランド・アスレチックス」のゼネラルマネージャーであるビリー・ビーンが、選手の年俸総額が全30球団で下から3番目で、ニューヨーク・ヤンキースの3分の1という貧乏球団であったオークランド・アスレチックスを「セイバーメトリックス」と言われる分析理論を駆使して、全球団で最高勝率のチームに変貌させる物語で、2003年に「マネーボール：軌跡のチームをつくった男」というベストセラーになったノンフィクション小説を映画にしたものです。

映画ではブラッド・ピットが主演するビリー・ビーンが野球選手としての経験はないが「セイバーメトリックス」を駆使するイエール大学卒業の若い理論家の分析に基づいて選手を補強し、それほど資金を使わないで20連勝をはじめとする優秀な成績を残すという物語です。

例えば、選手の評価は打率が重要な指標ですが、打率の良い選手は年俸が高額なので、四球も含めた出塁率で評価して安い選手を獲得するとか、犠牲バントなどの犠打はアウトカウントを増やすだけであるから、いつも強打させるなど、従来理論を否定して、統計的に作戦を進めるのが「セイバーメトリックス」です。

そこで日本ハムと阪神に戻ると、日本ハムはセイバーメトリックスを基にしたBOS（ベースボール・オペレーション・システム）を駆使しているのに対し、阪神は、実は最初にBOSを開発した球団ですが、現場の反対で導入を断念した経緯があります。

これだけが原因ではありませんが、日本ハムと阪神の今年の差は、ここに一つの重要な要因があると思います。

しかし、セイバーメトリックスはビッグデータとしてはデータ量も少ない初歩で、さらに進んだシステムが次々と登場しています。

現在、グーグルで情報を検索すると、英語の項目などでは「翻訳する」という欄があり、そこをクリックすると64カ国語に翻訳することが可能です。

日本語に翻訳することはもちろん、アラビア語、タイ語、ハイチ語、ラトビア語など、我々には馴染みの少ない言語にも翻訳され、それぞれの言語の文字で表示されます。

私も大学で自動翻訳を多少研究したことがありますが、これまでは、それぞれの言葉の単語の辞書と構文の例を用意し、主語と述語と目的語を判定して置き換えるという方法が一般的でした。

しかしグーグルではまったく違う方法で翻訳しています。

これまで人間が翻訳した膨大な文章から何億もの事例を集めておき、翻訳対象となる文章に近い事例を見つけ出して翻訳する「統計的機械翻訳」という処理をしているのです。

しばらく以前まで、翻訳された日本語は意味不明で、単語を調べる手間が省けるとい程度でしたが、最近では十分に意味が理解できるほどに進歩しています。

そろそろインフルエンザの流行の時期が近付いてきましたが、外国に行くときなどは、その国でどの程度流行しているかが気になります。

そこで「グーグル・フルトレズ」というサイトを開きますと、流行の程度を5段階に色分けした世界地図が現れます。

どのようにして調べているかという、インフルエンザが流行してくると、それぞれの国の人々がインターネットでインフルエンザの予防法とか薬とかを調べるようになります。

そこで、グーグルが各国のインフルエンザ関係のサイトへのアクセス数を測定し、その多さによって流行の程度を推定しているのです。

アメリカのカリフォルニア州のサンタクルーズ市では予算の制約から警官の数が増やせない、市域全体を150m角の升目に区分けし、過去8年間に、それぞれの区画で、どのような犯罪が何曜日のどの時間に発生したかの統計を作成し、犯罪が発生しそうな時間になると、巡回中の警官に指示を出し、その区画に向かわせます。

そうすると、警官が待機している場所で自動車の窓ガラスを割って物を盗むというような犯罪が発生し、待ってましたと御用にすることが増えてきたそうです。

日本でも事例は増えており、ウェザーレポートのウェブサイトを開くと、日本列島に天候の状態や地震の発生を示す○印がびっしり表示された地図が現れます。

これは約30万人の会員が、スマートフォン経由で自分のいる位置の天気の状態を

送ってきた情報を表示しているのです。

気象庁の自動計測装置であるアメダスは全国に1300カ所しかありませんが、ウェザーレポートは少なくとも数万とか数千という個所から情報が送られてきますから、より細かな予報ができることとなります。

さらにトピックスという項目があり、各地の紅葉の状態や初冠雪の状態が会員から写真で送られてきますから、いま現在の状況を画像で見ることできます。

野村総合研究所が昨年調べた日本、アメリカ、中国の企業のビッグデータへの取り組みを調べた結果がありますが、10テラバイト以上のデータを解析して使用している企業の比率は、日本28%、アメリカ56%、中国37%と日本は出遅れています。

もちろん残念なことですが、逆に見れば、日本企業にはチャンスがあるとも考えることもできます。

ぜひ企業関係のリスナーの皆様はビッグデータ時代に対応する戦略を検討されたいと思います。