

駒場リサーチ  
キャンパス公開講演

## 「サイバーフィジカルサービスと明日への IT」

喜連川 優 (生産技術研究所 教授)

### ◎ 3.11 震災時における IT

#### サイバーフィジカルサービスと 明日への IT

＝3.11震災時におけるITメディア利用の解析を含めて＝

きつれがわ  
喜連川優

東大生研 教授  
戦略情報融合国際研究センター長  
東大地球観測データ統合連携研究機構長  
国立情報学研究所 客員教授  
文部科学省 科学官

この生産技術研究所というのは大変温かい研究所でして、他の組織と違いまして、全く互いの研究の足を引っ張らないという、すごくいい風土のところ、私、20年以上ここにいらさせていただいて大変ありがたいんですけども、ときどき冷たいこともありまして、きょうも、大体こういうのは司会の方がおられて紹介して下さるんですけども、「勝手にやりなさい」とのことで、自分で自分を紹介して、自分で話をして帰っていくという、極めてエコの講演会になっております。本日、お越しいただきましてありがとうございます。

きょうの講演のタイトルなんですけれども、「サイバーフィジカルサービスと明日への IT」としてみました。2番目のフレーズは最近つけざるを得ないような状況かなと思っております。本題に入ります前に、「学術総懺悔」と言われているところもありますが、3.11 という危機的な状況に学がどう貢献できるのかという、その真価が問われているのが最近ではないかと思っています。

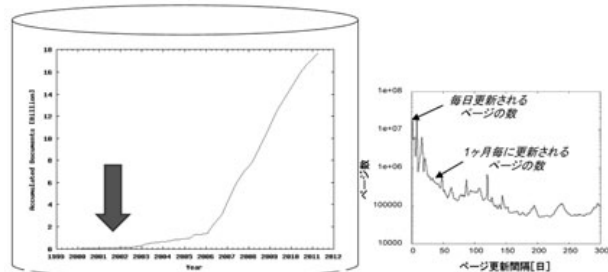
その中で、私どもが行っています研究の一つに関しまして、必ずしも直球ではございませんけれども、どんなふうにお役に立て得るかをまずご紹介させていただいて、それから本題に入らせていただきたいと思います。

#### ◎日本人ウェブアーカイブの構築

実は私どもでは10年以上にわたりまして、今、200億ページ弱くらいまでウェブのページを収集してまいりました。こんなにたくさんページを集めている機関は、ワールドワイドでも、インターネットアーカイブという組織があるの

### 日本語ウェブアーカイブの構築

- 12年間にわたり180億ページ(2011年4月)規模の日本語ウェブページを集積し、継続期間および規模においてアジア圏最大級のウェブアーカイブを構築
- 各URLの更新頻度に応じた可変周期収集



ですが、その次くらいで、アジア域では最大規模とされているのですが、この図を見ていただくとわかりますように、99年くらいから研究を開始しました。

このころは、回線の帯域が大変高価な時期でございまして、ウェブのクローリングをやっていると、大学の基盤センターから電話がかかってくる、「回線の8割も9割も使って、一体何をしていますのですか？」というような御小言を頂戴しながら、コソコソやっていたわけです。

当時、どうしてそういうことを考えたかといいますと、ここに書いてありますように、ITのメディアというのが今後非常に重要になってくるだろうと。つまりコンピュータというのは、最初、弾道計算をする、計算をする、情報の処理をするというところから入りまして、それが次の段階では情報管理する、つまり企業システムとして、大活躍してきました。そしてこの20世紀の終わりから21世紀にかけてコンピュータのメディアとしての存在感が非常に強くなってきたわけです。

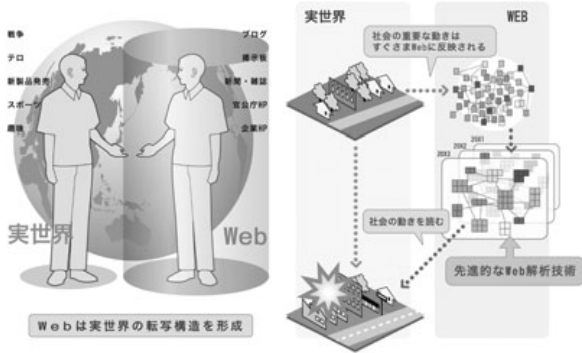
#### ◎実社会の射影としてのウェブ

この実世界の中で起こりますいろいろな事象は、新製品を発表したり、スポーツでだれが勝ったか、あるいは不信任案が出たりというような、重要な事象というのは本当に瞬時にウェブの世界、サイバー世界の中に反映されるようになります。そうしますと、これらの2つの実世界とサイバー世界は双対の関係にありまして、サイバー世界は実

実社会の射影としてのウェブ  
(ウェブは社会のセンサ)



Socio Sense



ヤシマ作戦

社会のセンサーと見なすことができます。いったん、サイバー世界に入ってしまうと、ここではコンピュータが処理対象として、何でもやることのできる空間ですので、先端的な手法を駆使しますと、社会の動きが見えるかもしれない、そんなことを「Socio Sense」という名前をつけてずっとやってきたわけでございます。

世の中にはいろいろな難しい問題があり、もちろんiPS細胞も難しいですし、ビッグバンも難しいです。社会がどう動いているのかというのがパッと把握することも難しい問題の1つと言えましょう。そういう課題にチャレンジしてきたという次第でございます。

◎ Socio Sense システム

Socio-Senseシステム



さて、ここで、3.11をITメディアからどう感じとれるか、あるいは3.11時にITメディアがどう動いたかをまずご紹介させていただきたいと思ひます。

◎ヤシマ作戦

この Socio Sense というシステムを共同で一緒に作ってきておられます豊田准教授からデモをご紹介させていただきます。

豊田准教授

ご紹介いただきました豊田と申します。

ウェブのメディアといっても、この数年で大きく変わってきていまして、一昔前はブログといたしまして、その日にあったことや読んだ本や、政治に対する意見とかをまとめてみんなに見てもらおうというようなのがはやってたんですけど、最近、Twitter というリアルタイムで友人同士でメッセージを共有するというようなサービスが出て来始めて、お互いのウェブの中でもメディアが複数あって、その「役割」というのがどうも変わってきているらしい、というのが見えてきております。

これは、喜連川・豊田研で自然言語処理を、ポスドクをやっていたら吉永君という方が作ったツールなんですけれども、ブログで何が話されているか、Twitter で何が話されているかというのを比較して、テキストの係り受けを細かく見られるというようなツールです。

上半分がブログで何が言われているか、下半分がTwitter で何が言われているか、というようなものになっていまして、ここでは“行う”という、一般的なキーワードですけれども、それで検索をかけたらどんなものが出てくるかというのが出ています。

各日に、男性が青で女性が赤なんですけど、男性が何件、女性が何件くらいしゃべっていたか、その下には係り受けの細かい情報が入っておりまして、これは「を」と書いていますが、これは「何々を行う」というふうに書いている「何々」の部分を一覧アップされているわけです。

地震の直後の3月11日には「募金活動を行う」とか「支援を行う」とか「ライブを行う」とか、そういうようなことがたくさん発言されていることがわかります。ブログもTwitter も「募金活動」が多く出てきているのがわかるんですけど、立ち上がりがいまいち違って、よくよく見ますと、3月11日、12日——12日にはTwitter のほうでは「募金活動」がブワッと上がってきているんですけど、ブログのほうでは送電とか支援とか、ちょっとニュースから引っ張ってきたようなキーワードのほうがまだ上位

にいる。

3月13日、14日と、だんだんとブログのほうでも「募金」というのが出てきてはいるんですけども、Twitterの立ち上がりが非常に早い、ブログのほうは若干遅れ気味で来るというような感じの現象があります。

さらに「募金活動」というところを細かく見てあげますと、これは「行う」とか「始める」とか、そういうのがいっぱい出てくるんですけども、Twitterのほうに12日から13、14とずっと上がって行って、だんだん収束、数が減っていくというようなパターンなのに対して、ブログのほうでは立ち上がりがジワジワ来て、ずっと同じペースをキープしていくというような感じの流れになってきています。

そういうわけで、基本的にはブログのほうに持続期間が長いようで、Twitterのほうは、急激な盛り上がりはあるんですけども、その後、だんだん減っていく、というようなメディアの違いがこの図で表れています。

さらにTwitterで、リアルタイムでいろいろな情報が共有されるということなんですけれども、それをさらに細かく見ていくとどうなるかというようなツールも現在作っております。

これは、Twitterの上で、リアルタイムでいろいろなことが言われているんですけども、それはだれに対するコメントだったり、メンションだったり、リツイートだったりするかというようなことがありますので、それで線を引っ張れるんですが、それを時系列で追いかけるというようなもので、一種、タイムマシンのようになっています。

時間を通して、どのようにコメントが動いていったかというのがわかるようにしています。地震が起きたのが3月11日の14時50分くらいだったと思うんですけども、このあたりから時間を進めていきますと、最初にポツッと現れた人が何を言っているかといいますと、「阪神大震災の経験者として、風呂にお水をためろ」とか「ガスの元栓閉めろ」、一般的なアドバイスみたいなものが最初に出てきていて、それがたくさんの人にコピーされているんですね。これは真ん中にいる人が最初の発言をした人です。周りの人は、「この発言は有用だから、私の友達にも見てほしい」といって、これをコピーしてまわっている人たちなんです。みんなが全員同じような文章を言っていることがわかる。もとの発言をコピーして、友達に「これは見ておけよ」というようなことを言っていることを表しています。

こういう塊がドッドッと最初のうち現れているわけなんですけど、これが地震が起きてから30分くらいのところなんです。ここからさらに時間を進めていきますと、どんどん発言が増えて行って、しばらくしてくると、一般的なアドバイスというよりは、具体的に「どこに逃げろ」とか、そういうような情報が現れてきます。例えば、ここを見ますと、「駅に集まらずに付近の避難所へ向かってください」と全国の避難場所一覧というようなものがここにありますよと

いうような情報が現れて、それもけっこういろいろな人にコピーされている、というようなことが起きていることがわかります。

さらに時間を進めていくと、こういう感じで、これは地震が起きてから3時間後くらいですけども、ものすごい勢いで有用だと思われる情報がたくさんの人にもものすごい勢いで共有されていく、というのがこのメディアの特徴であろうというふうに思われます。

例えば、この辺なんかには「東京都23区の避難所をまとめました」というような塊があって、ここを細かく見て行ってあげますと、すごい勢いでこのコメントをサポートする人が増えていっているのがわかると思います。この人は何をやった人かというと、東京都23区の避難所をまとめてGoogleマップに張り付けて公開したという人で、隣にいる人がこの話題を広げてくれた人なんです。

これ、時間をちょっと巻き戻しますけれども、この人が最初に発言して、これだけ広まった後で、有名人が突然ポツッと現れて、この人が何を言っているかといいますと、「これはすばらしい試みだから、みんなも情報提供してあげてください」というようなことを呼びかけているわけです。この人はTwitterの中では大変有名な人で、この人の発言は10万人以上が見ているような有名人で、その人が発言した途端にドッとこの話題が広まっていくという様子もわかります。

地震の他の情報もいろいろ見てみますと、こういうふうには有名人が「有用な情報」というのをみんなに伝えてあげて、その活動を支援してあげるというような状況が多く見られています。こういう状況が地震などによく見られているというのが興味深いところだと思います。

普通の、あまり有名でない人が始めた試みを有名人がサポートして行って話題が拡散していく、というような話もあるんですけども、一方では有名人が直接主体となって話題を広げていくような話も多々ありまして、例えば、募金活動なんていうので検索をかけてあげると、これは、3月12日あたりからですけども、丸の大きさがその人がどのくらい有名人か、何人くらいその人のメッセージを読んでいるかを示していますが、初期のころにかなりでかい人が真ん中において、Tポイントの公式のTwitterアカウントなんですけれども、Tポイントを被災地へ募金できますよというようなことを言っていて、それがものすごくたくさんの人に支持されているというような状況がわかります。その後、孫正義さんが出てきたり、有名人が次から次へと参加して行って、その人たちを中心に話題が広がっていくというようなパターンも見られます。

このように、話題によって情報が拡散していくパターンがいろいろ違うということがわかってきています。有名人が主体になって広げる話題や草の根で広がっていったものが有名人によってサポートされて広がっていく。どうい

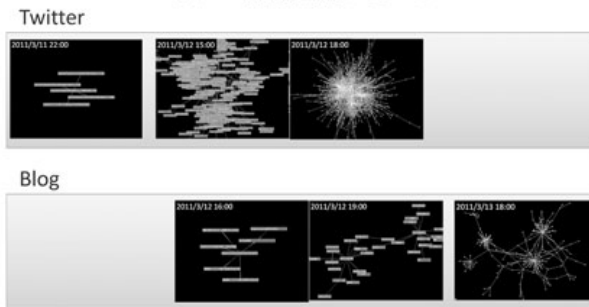
ふうにより有用な情報がネット上に広がっていくかというのを分析していく、というのが今後いろいろなことが起きたときに、何がウェブ上で起きるかというのを解析して上で重要になっていくと同時に、一方で、デマ情報みたいな不要な情報や問題がある情報が拡散するというような状況も多く見られていますので、こういう情報の拡散というのを分析していくことがこれから重要になっていくのではないかとこのように感じています。以上です。

### 喜連川

以上が3.11以降で、私どものクローラーで随分頑張っ、ツイッター情報を集めながら震災直後の状況などを解析した一例でございますけれども、ここのプロジェクターの解像度が足りませんものですから、これほどの大量の情報をお見せすることがなかなか不得手な状況になっています。もしお時間がありましたら、私どものE棟の5階の研究室に来ていただけますと4m×4mくらいの巨大なハイレゾリューションのディスプレイがございますので、それを見ていただきますともっと迫力あるデモをご覧頂けます。

### ◎ヤシマ作戦

#### ツイッタとブログの比較 ヤシマ作戦を例にとつて



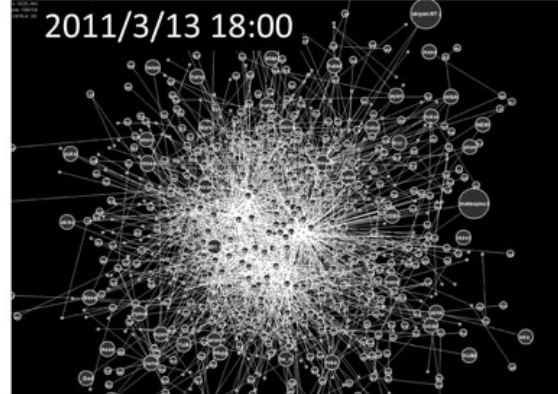
今、私どものほうからご紹介させていただいたように、募金活動もありますが、「ヤシマ作戦」も9.11で生まれた有名なソーシャルな動きの1つです。節電を呼びかける活動でアニメからその名前をとったものです。先ほど豊田准教授のほうからご紹介いただきましたようにブログで見ますと、翌日くらいから、しかしながらTwitterで見ますと、当日から、翌日になりますともうこんなに広がっていくところがございます、メディアの比較を行いますと、Twitterとブログというものの特性が非常に違うということがわかってまいります。リアルタイム性という面で見ますと、はるかにTwitterのパワーが大きくなってきていることが判りますが、それが最近の人気の的とも考えられます。

### ◎ Diffusion pattern

先ほどのデモでもご紹介しましたことをもう一度整理しますと、「Diffusion」という用語を英語では使いますが、情報がどう拡散していくのかという点に関しまして、とりわけ3.11時に緊急情報がどう広がっていくかについては誰

## ヤシマ作戦に見る 草の根的な拡散パターン

Majority twitterers have around 1000 followers

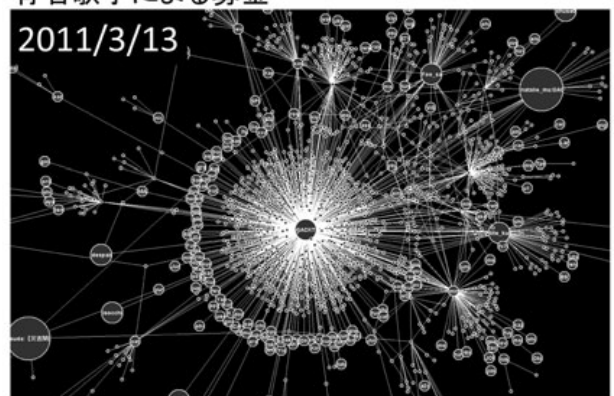


もよく判っていなかったわけですが、それに挑戦をしてみまして、この生研公開で始めにご紹介させていただいた次第です。先ほども説明させていただきましたが、最初は多くはグラスルート。そして、それがああるインフルエンサーによって広まるというようなパターンが多くなっております。ヤシマ作戦の場合も、最初、3月13日の6時前くらいまでは、大体フォロワーが1,000くらい、これは私自身でこのくらいの数で、普通の人という感じです。ところが、募金の例のように、ものすごく有名な人がいて、そこからフツと広がる場合もあります。一体、情報がどういうふうには伝播していつているのかというようなことを、3.11の様な危機的な状況時には、政府はモニタリングする術を持っておく必要があるのではないかと考えております。

### ◎有名歌手による募金

## インフルエンサーによる拡散パターン

有名歌手による募金



3.11時の情報拡散についてお話ししたわけですが、一般の企業におきましても、自社の新しい製品についてどんなふうにはレピュテーションが広がっているのか、あるいは何らかの障害があったときに、それがどのようにユーザに受け止められているのかというようなことも、全てリアルタイムに見れるようなシステムが作れる、そんな

時代になってきたということです。

◎ Nowism

## Nowism

ウェブ → ブログ → ツイッター

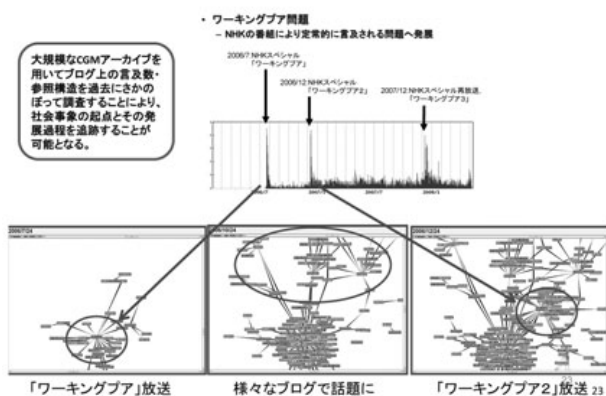
リアルタイム・メディア化へ

この様な潮流を Nowism という言い方をされる方もおられますが、ウェブからブログ、そして Twitter へと、いわゆる超リアルタイムメディアという方向に世の中がシフトしているということです。Twitter は、サービス開始当初、米国では地震がテレビで報道されるよりも Twitter のほうが数分早い、その数分のギャップが大きな価値と言っていたわけですが、最新のソリューションでは分単位のリアルタイムつぶやきストリーム解析が登上してきました。時代はリアルタイム化してきているということです。

しかしながら、だからといってどっちがいいとか、悪いかリアルタイムでなくてはいけないということではなく、多様なメディアを駆使する、そういう時代に入ってきておまして、申し上げたいことはそれを解析する基盤というものが重要だということです。

◎ワーキングプア

### ワーキングプア (ブログが社会のセンサ)



3.11 から離れますと例えば、我々は、ワーキングプアという問題をずっと解析してきました。先ほど、ブログの場合は比較的ライフタイムが長いという話をしましたが、これはものすごく長く続いております、何年にもわたって議論が続いている。ということは、逆にいいますと、こういう問題はやはり表層上のキーワードではなくて、国民に

とって非常に深刻な問題だということを感じ取ることが出来ます。我々はこういう多様なメディアを解析することによって時代を感じとれるようになってきつつあります。

◎最先端の開発支援プログラム

### 最先端研究開発支援プログラム

超巨大データベース時代に向けた  
最高速データベースエンジンの開発と  
当該エンジンを核とする  
戦略的社会サービスの実証・評価

東京大学 生産技術研究所  
戦略情報融合国際研究センター  
喜連川 優

さて、ここで3.11に関連する話題はおしまいにしまして、講演のタイトルを「明日へのIT」としてみましたが、私どもが現在進めております最先端研究開発支援プログラム、いわゆる FIRST というものをご紹介させていただきたいと思えます。

これは自民党が政権をとっておりました時代に、この一番端にありますように、研究費90億円を、30人ということでスタートしたのですが、なぜか民主党政権に途中で変わりまして、総額2,700億円の予定であったのが、政権交代で1,000億円に減額されました。それでもかなり大きな予算をいただきながら、私どものほうでは超巨大データベースへの挑戦というテーマで研究をさせていただいております。

もう少し正確に申し上げますと、「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証評価」という長いタイトルとなっております。

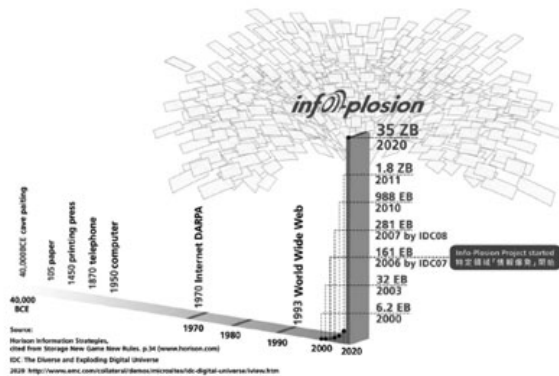
実は、科研費など、研究予算するときは、タイトルは最大30文字という字数制限があるんですが、このときは内閣が急いでおられたのか、制限を入れるのを忘れられたのか、文字数は自由で、たぶん採択された研究テーマの中でタイトルが一番長いのが我々の研究プロジェクトになっていると思えます。

ざっくり申し上げますと、データベースエンジンをつくるという研究と、それを使って新しいサービスを生み出す研究の2つから構成されております。

◎ IT にとって何が最大の変革的か

さて、たびたび色々ところで申し上げているのですが、21世紀の最大のITから見たときの変革点というのは、情報が爆発するように増えているという情報爆発現象

## 情報爆発: 21世紀の最も顕著な現象 ITにとっての最大の変革点の一つ



象と言えます。100年を経た後に振り返ったとして、これが非常に大きな変化の一つとしてあげられることはまず間違いないだろうと考えております。

### ◎ウェブ空間

先ほどお見せしましたように、我々のウェブアーカイブもこの様に急増大してきているわけですが、多くの方がITメディアが情報爆発を支えているというふうに一見お感じになられるわけです。

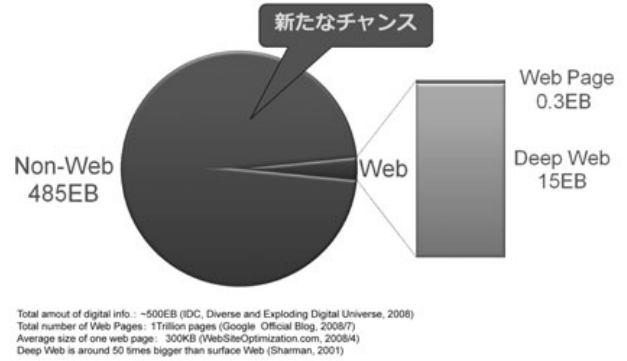
しかしながら、実はそうでもない、つまり情報空間全体を見たときに、ウェブがどれだけのものかというのをちょっと考えてみますと、このスライドの様にGoogleが2008年に、「僕たち、1兆ページくらい見つけましたよ」という発表をされておまして、同じ年にディープウェブの研究とか、ウェブページの平均サイズとか、いろいろな研究成果をいろいろなところが出ておます。それを合わせますと、いわゆるクロウラーが取ってくるサーフィスウェブの空間というのは、この赤い線、1本くらいではないんですね。この緑のところはディープウェブと言いますが、ダイナミックにウェブページが生成される部分ですが、問い合わせられた瞬間にそれが出てくるだけでありまして、たまたまクロウラーがそれを見なかったら、取れない情報で、これはかくれた領域になります。

そうしますと、この全体の情報空間に比べますと、Googleさんが「世の中のすべての情報をオーガナイズするんだ」とおっしゃられているんですけども、実のところ、そんなに大したことはないということになるわけです。

### ◎人が書く量はたいしたことはない。モノがしゃべる時代に!

じゃ、それで?ということになりますと、むしろ残っている領域の情報がまだまだプロミッシングであると、それは一言で言うと、こういうことになります。つまり、Twitterをいくらしゃべったところで、たかがつぶやき、つまり、140キャラクターなんていうものを1日に大体7,000万くらいツイート、最近では1億ツイートくらいだと言わ

## サイバー空間



Total amount of digital info: ~500EB (IDC, Diverse and Exploding Digital Universe, 2008)  
Total number of Web Pages: 1 Trillion pages (Google Official Blog, 2005/7)  
Average size of one web page: 300KB (WebSiteOptimization.com, 2008/4)  
Deep Web is around 50 times bigger than surface Web (Shaman, 2001)

れておりますけれども、人が書ける量なんて大したことがないんですね。それに比べますと、モノがしゃべる、これが重要になってきているということでございます。これを別の言葉でいいますと、いわゆる IOT, Internet of Things というような名前でも呼ばれる時代になってきました。

その一例として見ますと、例えばアメリカのスマートグリッドは1年間に1,000ペタバイト、つまり1エクサバイトのデータを生み出すとの予想が報告されています。これはフェイスブックも、こんなデータはさすがに出せないと思いますので、いわゆるモノがしゃべる時代というのが次の情報爆発の根源になってくるだろうということでございます。

### ◎超巨大データベースの登場

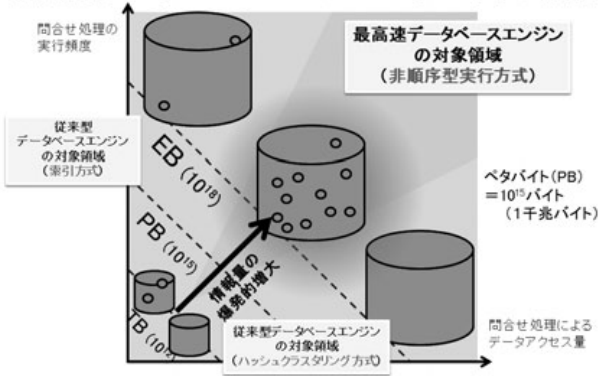
今のスマートグリッド例は、予想値ですが、現実世界でもデータベースの超巨大化は、進んでおまして、例えばウォールマートは、ちょっと古くなりましたが、2006年ごろに0.6ペタバイトくらいと言っておりました。さらに2008年に100社だったのが600社に拡大。そうすると、6倍するだけで、大体4ペタバイトくらい、今ですともうちょっと10ペタバイトくらいにはいっているのではないのでしょうか。

PGEというのは、スマートグリッドの先端的開発を進めているところでございますけれども、彼らも2009年くらいには3ペタバイトくらいで、近々4ペタくらいということをやっていますので、やはりこれも数ペタから10ペタくらい握っているような会社が出て来そうです。しかも両方とも共通することは人の書いたメンションではなくて、モノがしゃべっている、そういうデータだということがポイントです。その解析空間が重要な時代に入ってくるということです。

### ◎超巨大データベース時代に向けた超高速データベースエンジン

ここからちょっとテクニカルな話になるのですが、データベースを見ましたときに、原則、2つのアングルから見ることができます。それはどういうことかとい

### 超巨大データベース時代に向けた 最高速データベースエンジンのターゲット領域



ますと、非常に小容量、横軸は1回にアクセスするデータの量なんですけれども、例えば銀行で自分の口座の預金を見る場合、ほんのちょっと情報を見るわけです。そういうものはデータをとってくる量は少ないんですけれども、アクセスする頻度はべらぼうに高い、こういう領域がござい

ます。もう一つは、いわゆる解析をする、ビジネス的にはBI、Business Intelligence と呼ばれているようなものなんですけれども、たくさんのデータを処理する、その代わりに、それは時間がかかりますので1日にやれる頻度はすごく少ない、あるいは何日もかかったりします。この2領域がデータベースの特徴的な分類になります。

前者はこういうふうに大きなディスクがあったとき、ポツポツとしかアクセスしませんので、通常はインデックスといたしまして、索引を作るという方式で勝負をかけます。こちら側は、ディスクのデータ、ほぼ全体丸ごと見るということとなりまして、定番になっておりまして、我々自身がずっと長年やってきた方式なんですけれども、ハッシュクラスタリングの方式というものをここに適用することになります。

実は、大ざっぱに言いますと、この方式か、この方式か、どっちかをとるということです。つまり、たくさんとるならこっちをやる、ちょっとだけならこっちをするというような感じだと思っただけならばと思います。そのときに、データのボリュームがすごく少ないと、どっちをやっても大して変わらないんですね。ところが、データの量がだんだん大きくなっていきますと、どちらとも言い切れないような状況がクローズアップされてきます。つまり、点々がポツポツといっぱい出てくるような状況になってきて、ここになりますとパフォーマンスがずっと落ちてしまう。ただ、1エクサバイトというような、先ほどのスマートグリッドのような世界になりますと、そもそもこんなたくさんのデータを全部読むなんていうことはできません。かといって、ちょっとだけ読むということもない。つまり、

ある程度の量を読むというアプリケーションがこの超巨大なデータベース時代の一番のコアな領域になるのだろうと考えられ、その為のエンジンを開発しようと私どもは考えたわけでございます。

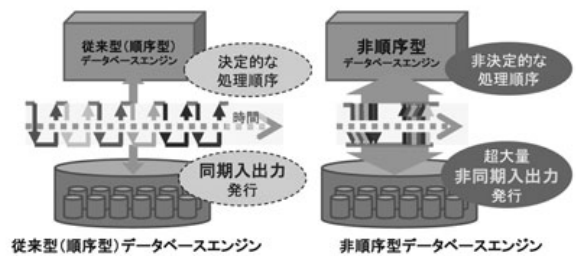
そんなことくらい、他の人だって考えるだろうというご質問もたぶんおありになられるかもしれないんですけども、データベースのベンダーはどちらかという上位方向をねらってまいりました。商用のデータベースシステムといますのは、XMLのインターフェースを入れたり、あるいは、ERP的なSAPとか、ピープルソフトとの連動というような、どちらかというハイヤーレベルのファンクションにどんどんシフトしていきまして、その根幹となるエンジンのところというのはそんなに変えられないというか、そこを変える元気が出ない。この部分は非常にコンプレキシの高いソフトウェアになっておりますので、よっぽどのがないとエンジンをやるというところには動かなかったのが過去の歴史だと我々は理解しております。

私共は、大学なんだから、やるなら根っこをやろうと。上位層も重要ですけども、ときどきは原点に戻って考えることが必要だろうということで、ここにございますような、非順序のエンジンというものを考えるに至ったわけでございます。

#### ◎非順序型データベースエンジン

### 非順序型データベースエンジン

- ・ 非順序型実行原理に基づく最高速データベースエンジン
- ・ 特徴
  - 超大量非同期入出力発行
  - ストレージ駆動型非順序型(Out-of-Order)実行
  - 実行時動的入出力スケジューリング



それは、簡単に言いますと、通常のデータベース、通常のソフトウェアというのは、同期 I/O といまして、「データをちょうだい」とストレージに言いますと、ストレージは「はい」とあげます。それをもらってから次のリクエストを出し、またその答えが返ってから、また次のリクエストを出すという、こういうやりとりをしまして、順序は赤、黄、緑、青というふうに決定的に、つまりこの順番は確実に決まった順序で実行されるというのが通常のデータベースの作り方なわけです。

そこに対して、私どもが考えましたのは、非同期の I/O を使おうというものです。この赤、黄、緑、青というリク

エストを返事が返って来る前に、欲しいものをどんどん言ってしまうわけです。返事が返ってくるものも、返事を出した順番とは全然違う順番で返ってきます。これはディスクに依存しますので、毎回毎回、順番が変わります。そういう意味で、決定的なシーケンスにはならない、ノンデータミニスティックな処理になるということでございます。

そういうややこしいことをやるんですけれども、やるとそれなりのご利益がございまして、決定的なものといえますのは、行って帰って、これはソフトウェアのレイヤーのスタックですけれども、データベースのレイヤーからOSのカーネルを経て実際のディスクドライブまでいきますと、そのレイヤー間をギョコンバタンしながらソフトウェアが動いていく。

我々の場合は、スレッドというものを山のように作りまして、I/Oのリクエストを出せるというような、そんなことを考えたわけでございます。

### ◎ Demo (ストレージの様子)

ここでちょっと直感的にどんなイメージなの、ということをご紹介したいと思います。最近のディスクストレージというのは、あまりピカピカしなくなったんですけれども、昔のディスクはけっこうピカピカ光りまして、こういうふうアクセスされると点がポツポツとつくような、何か頑張っている感じがして、という努力感を感じさせるような、そういうシステムです。

ここをご覧いただくとわかるんですが、これはアドレス空間なんですけれども、アクセスしたパターンのところ、アクセスするとそこに黒い点がポツと打つというのをしています。ノタノタノタノタと、こういうふう処理をしていっている様子がわかると思います。

ちなみに、これは5台のドライブで、ここにサーバーがあるという構成です。

これは、実は延々かかってしまい、長くかかりますので途中でやめます。これからが非順序実行(アウトオーダーエグゼキューション)、ここがバツバツとさっきよりもはるかに勢いよく光っているのがわかります。

[音声オン]

すごい音がしているのがわかるかと思うんですが、これでもう終わっちゃいました。先程のものに比べますととても速いことをご体感頂けたかと存じます。

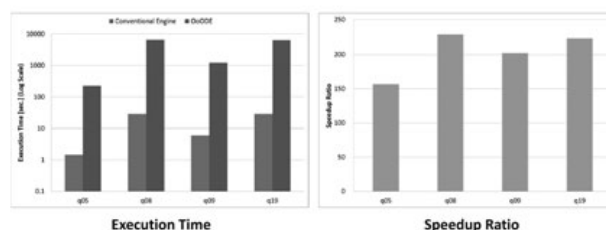
これは実は学会で論文賞を受賞させていただいたんですが、けっこう優秀な学生といいますが、先生がおられまして、発表した際に「これ、素晴らしいと思いますが、ディスクがどんどんつぶれるんじゃないですか」という質問を受けましてちょっと困った次第ですが、結局、今のI/Oに比べますと、多くの非同期I/Oを出すというようなことをやりますと、大きく加速させることができるということをご紹介させて頂いた次第です。

地震のためになかなか機器の導入が難しく、しかも新しい建屋の建築も難しかったんですけれども、生研の多くの方々のお力添えを頂戴しまして、この様なサーバーとストレージを導入出来ました。電気利用の制限があり、多くは今スイッチオフで、1つだけとりあえず動かせるという、そんなかわいそうな状況なんですけれども、ここで実行したときの性能をまとめたものがこれです。

### ◎日立HiRDBをベースにエンジンを開発

#### 日立HiRDBをベースにエンジンを開発 業界標準ベンチマークTPC-Hによる評価

- Scale factor 10,000 (Logical size = 10TB)
- Several queries tested in Conventional Engine and OoODE



左の図の縦軸が1秒、10秒、100秒、1,000秒と対数軸となっております。右図はこれを比で割ったものでございまして、青いのが通常のエンジンで、あれ、これ、逆さまですね、失礼しました。

これが普通のエンジンの性能で、これは我々のエンジンの性能です。ここに書いてありますように、10TBのスタンダードのベンチマークをデフォルトで使っております。

通常型のエンジンといえますのは、日立さんがお持ちになられているHiRDBというソフトウェアがあるんですけれども、そのデータベースの通常モードで動かすというのがこの実行時間でございまして、それを非順序型のエンジンにリプレースするとこれくらい速くなります。つまりハードウェア環境は全く変えないで固定にしておきまして、ソフトウェアだけを変える。通常ですと、ハードウェアをたくさん並べていって性能を上げるといってはありますが、ここではハードウェアをフィックスする。ソフトウェアだけを変えて、大体100倍以上の性能が出ているというところをご覧いただけるかと思えます。

ただ、先ほどこんな図を出させていただきましたように、ここここはそもそも、例えば1個か2個しか読む、あるいは全部読むというのは、これはもう最適解になっていますので、性能は上がりません。真ん中のところが上がるということでございまして、それを示しているのがこの図でございます。

先ほどの斜めのところを横にしたようなものなんですけれども、選択率がちょっとのレコードしか出さないものに



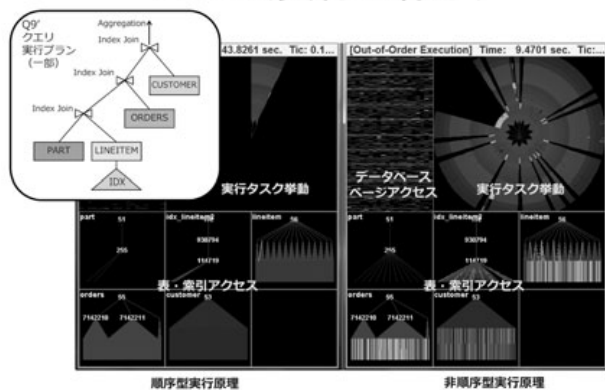
対しては数倍なんですけれども、もうちょっと越えてきますと、何十倍くらいに上がってきまして、100 倍、200 倍くらいまでのピークからずっとまた、全部読むものに対してはゲインが出ないというような、これくらいの領域で、高いパフォーマンスを示すということがだんだんわかってきているわけです。

実行の状況は非常に複雑で、最初、この様にきれいに整列して動いているんですけども、1 秒もすぎますと中はもうグチャグチャになっております。この中でおもしろいのは、非常に長い時間、I/O がかかっているようなものもありまして、これはグローバルなスケジューリングを行っているということに帰因しています。フェアネスなんていうことは一切気にしません。

マシンがどう動いているかという、もうちょっと内部の細かいところをご紹介しますと思います。

◎ Demo (実行の様子)

Demo(実行の様子)



今度は音は出ませんが、左側が通常のデータベースのエンジンです。ここからチョコチョコと動いているのがおわかりになると思いますし、これがいわゆるインデックスの構造なんですけれども、ピカピカと光っているところが、そこをアクセスしている状況を示しています。

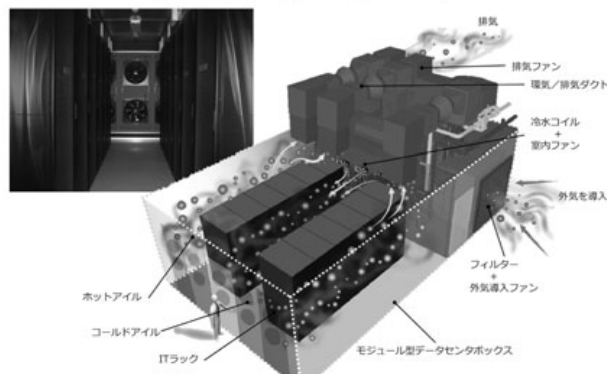
右は、アウトオブオーダーののですが、これを実行しますと、こういうふうには本来はこう回っていくところが、好き勝手に伸ばせるところをバンバン伸ばしているというのがおわかりいただけると思います。

それから、インデックスも、左がポツポツなのが右はパワーとアクセスしているというのがおわかりいただけると思うんですけども、レゾリューションの関係で、あまりきいに見えないんですが、こちらは延々かかって、まだここまでしかいかないことが判ります。やれるところから好き勝手にやって、処理がザッと終わっていくというのがおわかりいただけるかと存じます。こんなデータベースのエンジンを作っているということでございます。

◎外気冷却を用いた省エネ型データセンサ

このように、私共の新しいエンジンは、従来に比べまして、

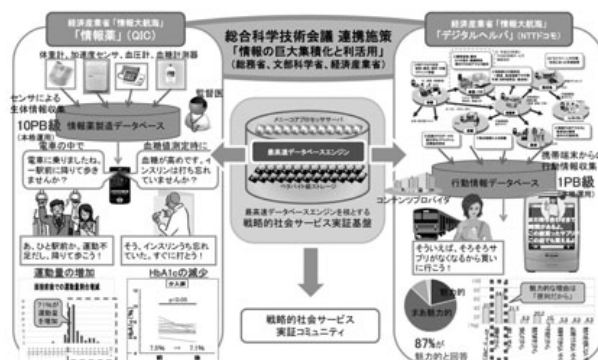
外気冷却を用いた  
省エネ型データセンサ



大きなデータに対してとても強力なものですが、エネルギーを消費しますので、それを入れている器も外気型のいわゆるコンテナっぽいものを導入して現在システム全体を開発しているところでございます。

◎巨大データを活用したサイバーフィジカルサービス

巨大データを活用した  
サイバーフィジカルサービス



タイトルの後半分の「該エンジンを使ったサービス」ですが、我々は、その当時は経済産業省の情報大航海プロジェクトを推進していたものですから、「情報薬」というものや「デジタルヘルパー」という、両方とも日本全体で実施しようとする軽くペタ級のデータを生み出すものですから、こんなものを当初イメージ致しておりました。これはアメリカ的に言いますと、サイバーフィジカルサービス、と呼べるかと思えます。

現在、後半のサイバーフィジカルサービスを実証する部分に関しましては、本日もおいでいただいております国立情報学研究所の坂内所長にサイバーフィジカルセンターをお作り頂きまして、この最先端プロジェクト全体の研究支援統括も国立情報学研究所安達先生にお世話になっている次第でございます。

◎サイバーフィジカルシステムって何だ？

さて、サイバーフィジカルって何なの、というところな

### ITの進化

- サイバー(ネット)世界の創出・発展



- 実世界とサーバ世界の融合  
(サイバーフィジカル)

のですが、フィジカルというのがついていると、つまり、先ほど「モノがしゃべる時代」と言ったわけですが、フィジカルというのは、“モノ”を指しているといっても言いかもしれません。

ITの進化という面で申しますと、いわゆるサイバー空間、アマゾンやネットフリックスというような世界から、だんだんより実世界を意識し、サイバー世界と実世界をよりタイトに結び付ける世界へと今ITが進化していると、流れを読むことが出来ます。4月の始めにサイバーフィジカルの国際会議がシカゴで開催されました。その起源は実はもっと古く、2006年ごろにアメリカのNSFが使い始めたようです。CPSの定義自身は、こんなふうになっております。system featuring tight combination over and coordination between the system's computational and physical element という言い方をしているということです。

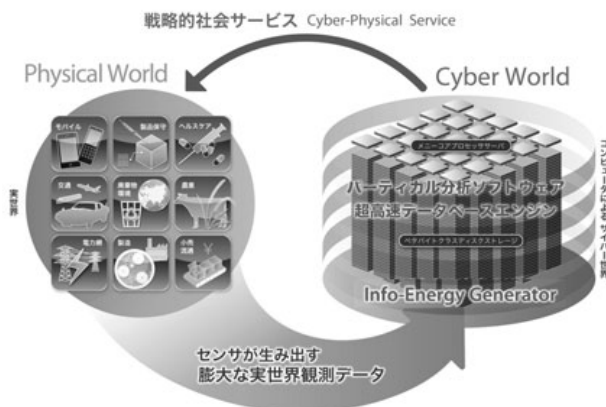
要するに、サイバーなパワー空間と物理空間をよりタイトにくっつける。別の言い方をしますと、従来はエンベデッドシステムというものだったわけですが、エンベッドがよりコネクタされる、そんな世界とも見ることが出来ます。NSFは、2年間、既にコールをかけておりますし、アメリカのNITRDの資料ですが、ITのトッププライオリティに指定するほどのプロジェクトになっております。先ほどから繰り返しておりますように、タイターインテグレーションというのが肝でございます。

対象領域はいろいろございまして、医療からファイナンスからトランスポーションからロジスティクス、スマートグリッド、農業、ビルディングなど、ありとあらゆるところに適用可能が模索されております。ちょっとキュートなところとございまして、メキシコと接しているアメリカですと無人のボーダーコントロールをどうするかとか、あるいはオイルを流すためのパイプラインはむき出しですので、穴を開けてプツと取っていく盗賊が山にいます。それをどうやってプロテクトするか、というようなクールなアプリケーションが沢山考えられています。

#### ◎基本アーキテクチャー

そういうものを全体構造としてみますと、この図のよう

### 基本アーキテクチャー



に考えられます。つまり、先ほど出ていましたような、多様なアプリケーションに於いて実世界がオブザバブルになる。つまり、センサから膨大なデータを吐き出し、それをクラウド上に持ってきて、先ほどご紹介しましたような高速のデータベースエンジンで必要なデータを超高速にアクセスする。その上にパーティカルなアナリティックスエンジンを入れまして、そこからサイバーフィジカルなサービス、つまり実世界にフィードバックをする、こういうループが描けるのではないかと。これが下名のCPSに対するViewです。どちらかというと上側の細い矢印のことを今までサイバーフィジカルサービスと呼んでいたわけ

#### ◎別の視点

ですけれども、裏を返しますと、下の太い矢印という観点から見ますと、まさに私共が推進して参りました「情報爆発」特定研究プロジェクトで言うところの「情報爆発からの価値創出」ということとございまして、情報爆発という名前を入れますと今やヒット件数が2,000万件くらいにもなるような時代になってくれたんですけれども、今年の

情報爆発現象の捉え方

infoplusion 55



震災日にたまたま開催していた最終シンポジウムが中断されるという、忘れることのできないプロジェクトになりました。いわゆる情報のオーバーロードという側面ではなく、大量情報をいかに利活用するか、という観点でこの情報爆発という文部科学省のプロジェクトを進めてまいりましたし、情報大航海という経済産業省のプロジェクトもお世話致しましたが、それもいわゆるサイバー空間上のウェブ情報ではない非ウェブの情報をどう活用するかという観点で推進して参りました。「情報爆発からの価値創出」を唱えてから6年くらいすると、一応、メディアの方もだんだん、ちょっと、これ、シャンパンから金貨が出てくるという、これで勘弁してくださいと日経BPさんに言われたのですが、読売さんも、エコノミストもというところで、その他色々な所で記事に取り上げて頂きました。

◎表現は違っても方向感と同じ

したがって、別にサイバーフィジカルという考え方自体は新しいものではなくて、2006年に生まれました後、2009年からファンディングがなされております。情報爆発は2004年にファンディングをアプライしまして、2005年から始動しまして、経産省の情報大航海は2006年にア

表現は違っても方向感と同じ

- 文科省 情報爆発 2005(2004) – 2010
- 米NSF Cyber Physical System 2009(2006)-
- 経済省 情報大航海 2007(2006) - 2009
- IBM Smarter Planet 09?
- HP Central Nervous System for the Earth 10
- IOT, M2M
- Big Data
- 喜連川最先端研究開発支援プログラム, FIRST(2010-2013)

プライしまして2007年からです。IBMのSmarter Planet,あるいはIOT, M2M, いろいろな表現がありますが、原則、ほとんど同じことを言おうしていると理解していいのではないかと思います。

◎BIG DATA

# Big Data

そして、最近出てきているのがこの何ともシンプルな言葉の“Big Data”でございまして、マッキンゼーが分厚い、1 cm 5 mm くらいのレポートが5月に出しました。Big Dataが次のイノベーションのフロンティアであるというように言っておりまして、まさに我々が情報爆発で言ったのと同じだなという気がするんですが、“torrent”, 激流だと言っています。50億の携帯電話が世の中で使われ、フェイスブックのようなソーシャルコミュニティが1カ月に300億のコンテンツを入れるようになって、それだけ情報量が増えている中で、IT spendingは5%のインクリースしかなく、40%のコンテンツインクリースがある。

◎Exhibit

これをどうやっていくかがこれからの勝負です、という指摘をしています。その根幹が先ほど言いましたように、「モノがしゃべる」という、つまりinternet of thingsからのデータのエクスポージョンが一番大きなポイントという言い方をしています。

しかも、それはいろんな領域で非常にポテンシャルベネフィットが大きく、これからの時代の次のITのドライバーはこの領域だろうというふうに、マッキンゼーは解析して

経済省 情報大航海プロジェクトの背景と目的

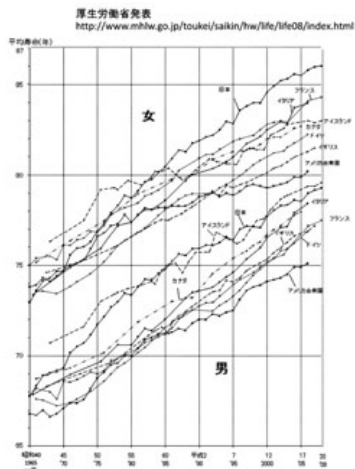


いるわけでございます。

◎人の健康

**史上最も長い  
寿命を生きる日本人**

- この25年間は日本が「世界一長寿国」
- 中国人は全体で73.0歳(2005)
- 韓国は、男75.7、女82.4歳(2006)



私どももそういう意味でいろいろな領域があり、最先端のプロジェクトの中では、多様な挑戦をしているのですが、とりあえずきょうは「人の健康」というものをちょっとご紹介させていただきたいと思います。

◎史上最も長い寿命, 生きる日本人

日本は、これが女性の平均寿命で、これが男性の平均寿命で、圧倒的に平均寿命が長く、25年間、日本が最長寿国になっています。

◎韓国に次ぐ世界で最低レベルの出生率

一方で、韓国が追い抜いていただきまして、世界第2位のバースレートを低くしている。ということはどういうことかといいますと、要するに高齢人口が大きく増えてきている。

◎年齢とともにね上がる医療費

一方で、医療費という観点で見ますとこんなふうになっていまして、お年をとられると医療費が上がるというのはいたしかたない話のわけですが、このカーブをいかに右にドリフトさせるかというのが国の健全な医療経済を実現するかということに対応致します。そのために、いわゆるメタボ健診が導入されたとも考えられます。

◎特定健診制度

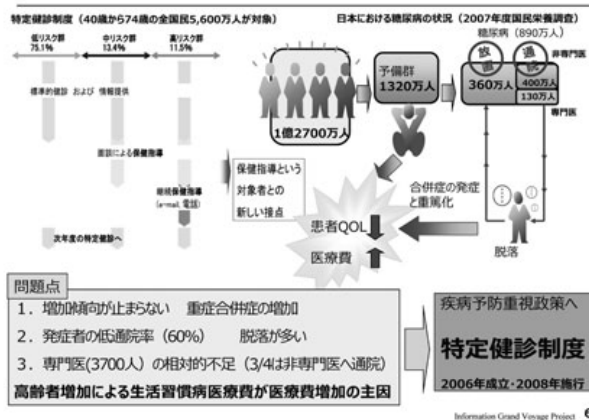
現在、特定健診の対象者5,600万人ですが、メタボになっている人は2,000万人弱くらいおられまして、こんな多くの人々は、IT無しに診れるわけがありません。又一番典型的なのは糖尿病ですけれども、糖尿病の専門医は100万人程度しかおられずITでなんとか重篤にならない様に生活の支援ができないか。

◎情報薬

そこで、情報大航海プロジェクトでいろいろ、あるいは情報爆発プロジェクトでも基礎研究をやってきたのですが、固いタブレットの薬ではなくて、情報を薬にしようという考えで九州大学の中島先生と東大の須藤先生にと



「次世代解析技術を活用した携帯情報端末による健康管理」



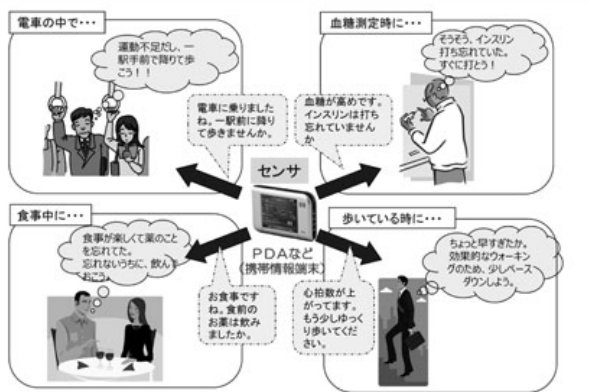
情報薬：適切なタイミングで提供される情報は人を健康にし、薬となり得る

言葉 ⇒ 運動量 ⇒ 「情報薬」 by IT

「情報薬」コンセプトは、札幌医科大学・辰巳治之教授が提唱されました。

り組んで頂きました。薬をのむのを忘れてるんじゃないの、という普通の薬摂取の注意に加えて、あんまり運動しないからエレベーターではなくて階段で上がりましょう、という様な情報を提供することでそれが薬と同じような効果を生み出すという考えです。昔はこの例も言っていたんですけど、最近、震災でエレベーターがほとんど止まっ

情報薬 (センサーによる情報爆発の活用)



Information Grand Voyage Project 71

ておりますので、言わなくてもそうなっていて、少々複雑な状況です。要するに加速度センサを使って運動量を測定し、それに基づいてリコメンを行うという手法でございます。電車も1駅前に降りて歩きましょう。とタイムリーにリコメンがユーザにプッシュされます。

◎情報薬による運動量の改善

そうしますと、たいいてい、みんなオビディエントな人ばかりですので、小生の様な大学の先生はへそまがりです、ときどき反対に運動量が減る人もいるんですが、「動け」というと大体の人が動いてくれる。そんなに被験者を選んだわけでもないんですけども、そういうことがわかりましたし、実際に定量的にデータがありますと、保健師さんとの会話に必要な時間が大きく軽減されるという意味でも経済効果が大きいことも判りました。

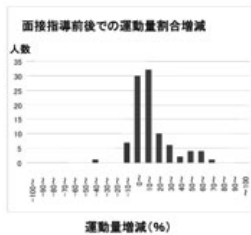
◎機械の学習 識別

情報爆発・情報大航海プロジェクトによる一步をより発展させ、最先端研究では行動認識の高度化を進めております。これは口で言っているのは簡単ですが、実はとても難しい。つまり、3軸の加速度センサがあるわけですけども、このセンサで一体何をしているんだ、ということをと

実証事業の成果

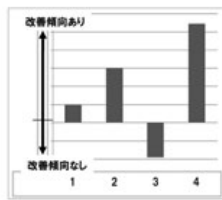
「100人実験」の成果 (1) (保健指導の効果測定)

モニター約100名中、90%で運動量が増加した(平均+9.6%増加)。“歩行”や“立つ”などの行動が増加し、“エレベータ”などの行動が減少していた。



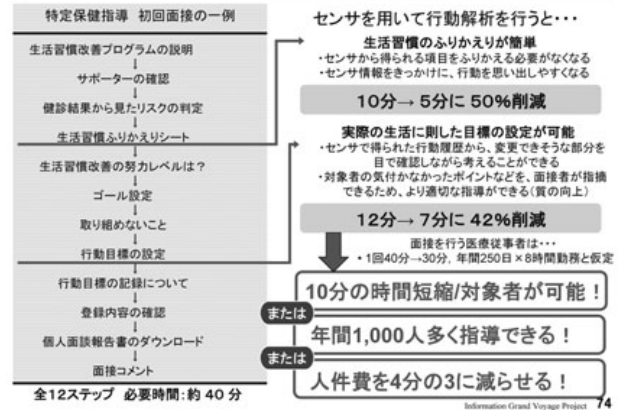
健康への意識によりクラスターリングをすると運動強度の変化に差が見られた

- 1.健康に留意しているグループ
- 2.健康に不安を抱えているグループ
- 3.健康に自信があるグループ
- 4.健康に漠然とした不安があるグループ



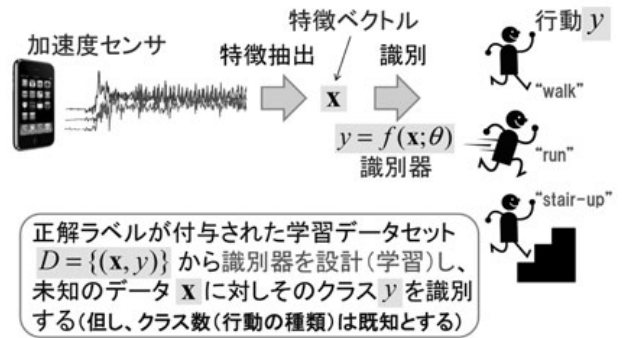
Information Grand Voyage Project 73

保健指導の大幅な効率化



Information Grand Voyage Project 74

機械学習に基づく自動行動識別



適応学習の必要性

加速度データは体格や運動能力等で個人差がある

学習データ(既存のセンサデータ)で学習した識別器はテストデータ(新規のセンサデータ)に対して、十分な識別精度が保証されない(未解決問題)

転移学習(共変量シフト適応)に基づき、適応的な行動識別器学習法を提案し、実データを用いた2クラス識別実験で有効性を検証

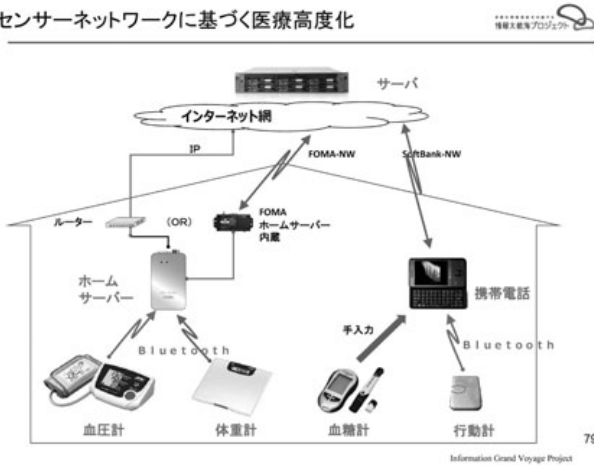
定することはそんなに簡単ではありません。

要するに、音声認識と全く同じことになりまして、不特定話者、つまり体重が大きい、小さい、背が高い、背が低い、いろいろな動きのある中で、これが運動量なんだということを学習するスキームというのはまだまだ大変なところでございまして、今、それを一生懸命、研究開発をしようとしているところでございます。詳細は省略させていただきますが、トランスファーラーニングという手法で幾つ

かの実験データを行いながら、ある特定の人で勉強したものをうまく他の人に適用するというをやりますと、かなり学習効率が上がる、精度が上がるということもわかってきているということでございます。

◎センサーネットによる医療の高度化

センサーネットワークに基づく医療高度化



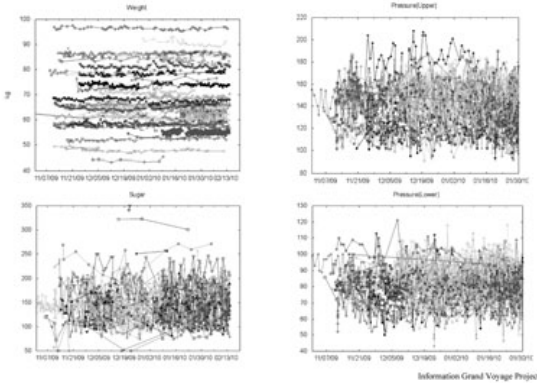
79

医療に関しましては、大航海プロジェクトでは50人ずつの介入群と非介入群で実証実験をしたわけですが、加速度センサだけではなくて血圧計とか、血糖とか、もっとシリアスなセンサを沢山つけた実験を行いました。

◎ケーススタディ

膨大なセンサー値

各測定機器で取得できた情報は以下のとおり



81

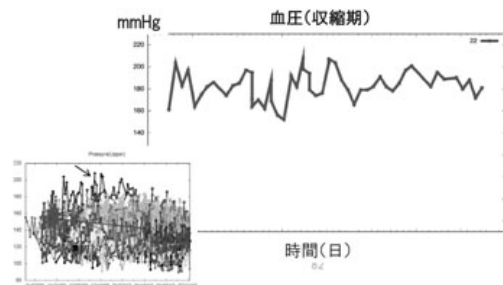
そうすると、これが一番興味深い例だったのですけれども、普通、病院に行って血圧を測りますと、緊張してプッと血圧が上がるんですけども、ある方は、普段は血圧が高いにもかかわらず、カッコいいふりをしようと思って、血圧をプッと下げる方がおられるんですね。こういうのを「逆白衣」と言うのだそうでして、そんな人を実際に発見出来ました。その人は、即入院ということになったそうです。

病気になられている方の常時モニタリングというのは非

実証事業の成果—ケーススタディ—

血圧のモニターにより病態が判明した症例

・通院時よりも日常の血圧が著しく高くなる「逆白衣高血圧症」の検出例。即入院治療となった



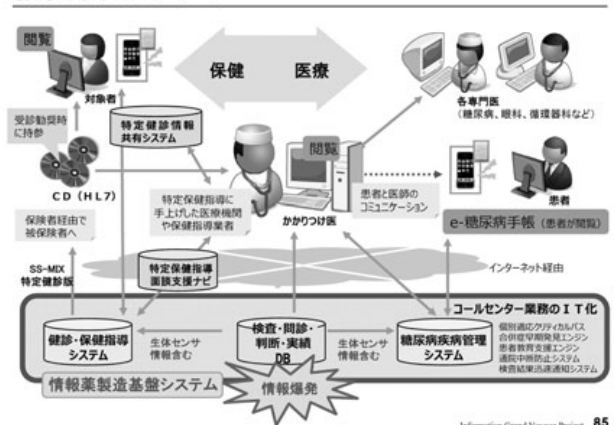
常に価値が高いということ、これもセンサによる膨大な情報爆発が肝で、次世代の斬新なITのソリューションというものがCPSから出てくるのではないかと考えております。

◎クラウド上にたまればたまるほど知の高度化が加速

クラウド上に

貯まれば貯まる程 知の高度化

情報薬製造基盤



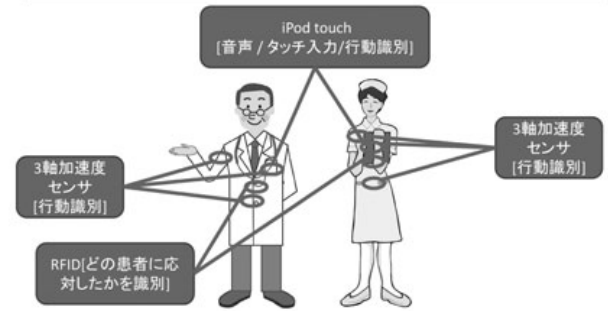
Information Grand Voyage Project 85

クラウド上にそういう知がどんどんたまることによりまして、より高度化する。つまり、今言いましたような知見がたくさん蓄積されることによって、その精度がますます上がってくる方向になると期待されます。時間がなくなりましたので、この後の例はちょっと省略させていただきたいと思っておりますけれども、最後に申し上げたいことは、情報

患者用のセンサ装着イメージ  
(フル装備、実際は疾病により選択)



医療従事者用のセンサ装着イメージ(2011年度)



- 2011年度は急性期病棟の医療行為を識別
- 2012, 2013年度は在宅医療・介護行為を識別

爆発をいかに料理するか、という時代にまさに入ってきたということでございましてしかも、その市場規模は極めて大きくマッキンゼーは30兆とも計算しております。この

様な Big Data, あるいは情報爆発に対して私共の取組についてご紹介をさせて頂きました。ご清聴, ありがとうございます。(拍手)