

IBM 東京基礎研 森本所長, ジンガジャパン 松原社長 インタビュー by 近山, 喜連川: 「ビッグデータが何をイネーブルするか? その本質を議論しよう。」

日本アイ・ビー・エム株式会社東京基礎研究所所長の森本典繁氏とジンガジャパン株式会社・代表取締役社長 CEO である松原健二氏の両名をお迎えし、本号のゲストエディタである喜連川優教授 (東京大学生産技術研究所) と近山隆教授 (本誌編集委員, 東京大学工学系研究科) を加えた4名が、スマーター・プラネット, ソーシャルゲーム, データエコシステムの観点からビッグデータについて刺激的議論を繰り広げます。ぜひお楽しみください。

喜連川: 本日はお時間を頂戴し有難うございます。ビッグデータというキーワードは、IT 分野において最も注目を受けている言葉の一つとなり、業界だけではなくアカデミアでも、多くの国際会議でもキーンと講演でもビッグデータが取り上げられることが多くなっています。今日は、ビッグデータの表層流ではなく、深層流について議論をできればと考えています。

Smarter Planet とビッグデータの接点

喜連川: IBM は世界中の都市でスマーター・○○というプロジェクトをやられて、スマート化とビッグデータは呼び方は異なりますが、ビッグデータが重要な役割を果たしており、非常に頑張っておられると思います。現状どこまで来ているのか、次の方向感について、まず森本さんから紹介いただけますと幸いです。

森本: IBM が Smarter Planet を提唱したのは2008年11月ですが、そもそもプラネットと言っていた時の最終ゴールは、岩野和生さん (注: 当時 IBM, 現三菱商事) が説明されていたように森羅万象のあらゆる情報を見て扱える、というものでした。重要なのは、そのデータを集めて何をするかということです。基本的には、データを集めてそれを処理して何らかのアクションに役立てるという PDCA のループが完成すると、それが1つのスマーター・○○ということになります。



喜連川優氏



森本典繁氏

例えばスマーター・ウォーターという適用分野 (ドメイン) では、アイルランドの港湾の管理データの見える化や、ニューヨーク郊外にあるハドソン川の河川の管理などに取り組んでいます。また、ストックホルム、シンガポール、オーストラリアでは交通シミュレーションや予測・制御を行うスマーター・トラフィックが行われています。IBM の伝統的な領域では、スマーター・コマースやファイナンスなどが挙げられます。しかし、情報を集めて交通渋滞が解消するというだけでは、シティーにおける人間生活の1つの側面にすぎません。今後は、これらのドメインが横とつながっていくと思われます。ここにビッグデータとの接点が出てくるのです。1つ1つのドメインの中では、限られた分野で限られた情報ですが、実際には2つや3つのドメイン、さらにはもっと広げてデータを重ねていかなければなりません。現在は大量のデータが取得されていて、いろんなデータがタダでアクセスできる状態になりつつあります。むしろデータ量の方が大きくて、それをプロセッシングする比率の方が小さい状態になるという逆転現象が起きています。

喜連川: はい、データがプロセッシングよりも主役になるというメッセージは判り易いステートメントだと思います。

森本: はい。一方、データをただ単に組み合わせて計算していたらすぐに計算量が爆発するので、どのデータを組み合わせるのか十分に検討する必要があります。これがビッグデータの今の一番のリスクではないかと思っています。この先 IBM がすることは、シナリオベースの分析、解析、具体的な目的を抽出することだと思っています。

複数のドメインが融合した1つの例はリオデジャネイロのケースです。リオの市長が素晴らしいリーダーシップをお持ちで、全31の政府機関が集まったシティー・コマンド・センターというのを作りました。そこには各政

府機関の代表者がいて、各端末が政府機関の全情報システムにつながり、データがリアルタイムに入ってきます。代表者はそこで即決するのです。例えば、ここに消防車を出す、AとBはここで一緒に行動せよ、警察と消防はここへ急行してCをせよ、などを全てそこで決められるのです。官僚的な組織では、各機関や省庁にお伺いを立てたり、無駄に長い承認プロセスを経ている間に、機を逸してしまいます。

喜連川：なるほど。

森本：そういう場があるという状態で、個別のドメインをつなげるというのを今実際にやりつつあります。気象データが地理データとつながって、どの辺のエリアが問題になりそうかということ予想することができ、交通局による交通制御につなげることができます。もしそのエリアに病院があるとしたら患者らを退避させなくては行けないので、消防車、そして救急車が向かう。このようにして初めてスマートな世界が実現されます。

まだ技術的な問題もあります。そもそもシティーのガバナンスが効いている点が成功に大きく寄与しているかもしれませんが、リオのケースを横につながるドメインの1つのモデルとしてやっていこうと考えています。

喜連川：なるほど、要素はできてきたので、それらをつなげる取り組みのフェーズに入ってきたと。そこはシナリオベースというか、目的やニーズに基づいてより大きなシステムをデザインする。まさにそこがビッグデータの力の見せ場ということですね。

このアプローチを日本に適用する時に、リオだとできたけれど日本ではこの辺が難しそうとかあるんでしょうか。

森本：日本での難しさは、やはりデータのオーナーシップとかプライバシーでしょう。今、「パーソナルデータの利用・流通に関する研究会」というのが総務省にあって、政府系や公的なデータを公開してなるべく使ってもらおうという取り組みをしています。これは大変重要な課題です。

喜連川：経済省でも dataMETI を推進しようとしており、ビッグデータ同様、オバマエフェクトで大きく動こうとしておりますね。小生もデータのエコシステムを回すためには、産業の委縮効果が出ないように、オーナーシップの問題を整理する必要があると常々主張しております。プライバシーは当然かと思いますが、日本は現行検索エンジンが著作権上 2010 年 1 月 1 日にホワイトになったわけですが、それまではグレイでした。グレイではダメなんです。産業が育たない。遅すぎて、大きく後塵を拝しました。二度と同じ事態にならないことが必須です。

森本：やはりデータのオーナーシップやプライバシーに関しては、ルールを明確に決めていかないとはいけません。せっかくここまでは使ってみたけど、ここからは許されないとか、実はその使い方は抵触していますと言われる。あるいは個人情報保護法もあって、それらがデータ利活用に二の足を踏ませてしまうことがあるという部分は否めないですね。

喜連川：それは日本だけの問題ではなく、世界中が同じ問題をはらんでいると思います。リオの場合、何かうまくいった秘訣みたいなものはあるんですか。

森本：あるとすれば、その 1カ所にデータを全部集中して、そこに権限を集中することができたという部分でしょうか。

喜連川：それは政府のものだからというオーナーシップによりデータを集めることができた、そういうシステムを作ったということは、ある種、拍手すべきことだと思います。

オーナーシップはリオ特有の状況があったという事実は理解できましたが、プライバシーの懸念は、多分もう世界中で共通の課題で、リオだから解けるというような問題でもないという気がします。

森本：そうだと思います。世界中で今まさにその議論をやっています。どういうところまで使っているのか悪いのか、こういう使い方ならいいとか悪いとか。そういうルールを明確に決めていくというのは重要です。日本の場合はちょっと難しい部分がありますね。世界といっても広いですから（笑）。

喜連川：日本は、まだそれほど IT が十分利活用されているとはいづら部分もある気もしますが、それでも全世界の標準から比べるとかなり高いレベルにあります。今後は、各国難しさが違いますが、いずれにしてもポイントは、先ほどおっしゃっていただいたドメイン横断（クロスドメイン）というのがキーワードになるという点はまさに同感です。

ソーシャルゲームにおけるビッグデータの活用

喜連川：Smarter Planet というのは、米国 NSF 的な言い方ではサイバーフィジカルシステムと呼ばれますが、物理世界をサイバー空間（あるいはクラウド空間）で集約してソリューションを産み出していく方法論におけるビッグデータ化と見なせます。

初期の Amazon ではデジタルメディアだけが物販されていましたが、これは 100%サイバー空間に閉じたサービスでした。ところが次第に非電子的な物を売るようにな



近山隆氏



松原健二氏

り、物流が重要になり、最近ではキバ・システムズ（注：Kiva Systems, 米国の物流センタ用ロボット・メーカー）を買収するなどバックヤードの重要性とその改革が目を見はります。いまだにサイバーだけで済んでいる世界というのは、ある意味で幸せにも見えます。しかし、そこで新たな大きな変革が起きているのがソーシャルゲームと言えましょう。

ここ1~2年のとにかく一番大きな動きは、完璧にモバイルシフトしたということです。そのモバイルシフトの中で、どこでも、いつでも、誰でもやれるようなソーシャルゲームが実現された。これが大変大きな起爆剤となり、この波に乗ったプレイヤーが大きな成功を収めています。膨大なデータの解析が肝という今までにないようなゲーム感をもたらしているのではないかと思いますので、その辺の動きについてご紹介いただけますでしょうか？

松原：私は、もともと家庭用ゲーム機の仕事をやっていて、その後オンラインゲーム、ソーシャルゲームという仕事をするようになりました。従来のゲームは作った後にユーザに提供して、そのフィードバックを別のチャンネルから得ていました。小説、映画、音楽など他のコンテンツも同じく、アンケート、売上げ、メディアからの批評などのチャンネルに依存していました。ところが、ゲームも他のコンテンツも、この10年位の間に、オンライン化によって大きく変わってきて、モバイルになってさらにシフトしてきたと思います。つまり、作り手が作った後から手を離れるという形から、サービス化してきて、作り手とユーザが非常に近い関係になってきました。

オンライン化によりユーザのあらゆる行動が手に入るようになったのはゲームならではの特徴だと思います。音楽は受動的なコンテンツなので、いくらオンライン化しても、作り手がユーザに音楽を渡した後、ユーザがプレイヤーのスイッチを入れてその音楽をどんな風に聞いているかというのはなかなか分かりません。ところがゲームというのはインタラクティブなので、そこが全て分かってくるのです。

ゲームに関するデータをどのように分析するかという

のは幾つか段階があったと思っています。初期は、何人プレーしたかとかの大雑把なデータの把握から始まりました。その時は売り切り型が支配的でした。それがソーシャルゲームになって、ゲームを初めに無料で提供して要所要所でお金を払って頂くというモデルになってきました。そうすると当然ながら、より売上げを増やすために、どの点はうまくいっているか／うまくいっていないかというのを、かなり事細かく分析するようになりました。いくらオンラインゲームと言っても、売り切り型だとそこまで細かく分析しても売上増にはなかなか直接に結びつかないのですが、現在のソーシャルゲームでは、たった今分析した結果に基づいて1時間後に手を打てば、そこでボンと売上げが伸びるという世界になっています。そういう点でデータの分析がものすごく重要になってきました。

大量にデータを集めて有効に分析をすることが重要ですので、ジンガ社 (Zynga) では、ゲームのいたるところにデータ採取ポイントを埋め込んでデータ分析をしています。新しく企画したゲームアイデアがうまく動いているかどうかをリアルタイムに把握して、ダメだったら直ちにすぐパラメータを変えたりしています。今まで家庭用ゲーム機はパッケージで販売していたので、リアルタイムに修正するということはおよそできませんでした。

大きなデータと言うと、現在ジンガ社全体でおおよそ月間3億人のお客様に遊んで頂いています。個々のゲームに関しても月間数千万人というレベルで遊んで頂いています。世界規模でサービスをしているので、もちろん男女差だとか年齢差だとかそういう分析もしていますが、地域ごとの特性とかさまざまな分析もしていて、この数年でさまざまな分析結果が蓄積されてきたと感じています。

喜連川：ゲームは、リアルタイムにユーザのニーズをセンスできる最先端のアプリケーションという気がします。ユーザが毎日ITに費やす時間がありますが、その有限の時間の中をいかに多く取るかが今のIT業界の最大の目的関数になっているのは周知の事実ですが、全体の約1/5をSNS(ソーシャル・ネットワーク・サービス)が占めて、残りの1/5をゲームが占めて、さらに残りの1/5を電子メールが占め、残りは個別なサイト群が占めます。つまり全体のほぼ2/5はSNSとゲームが占めている。サイバースペースのサービスの次の最も大きなプレイヤーはSNSとゲームです。検索というのはもはやメジャーではない世界になりつつあります。ところが、SNSは実はサービスではないのでビジネスモデル的には厳しくて苦戦しています。その点、ゲームは極めてクリアです。

ここからが質問です。ゲームにはゴールドファーミング（注：ゲームの通貨やアイテムを現実の通貨に換金する RMT（注：リアル・マネー・トレーディング）やキャラクタの育成代行などの商売）というダークサイドの側面がありますが、そのダークサイドを、先ほどのデータ分析で解き明かして健全な世界に持っていけないのでしょうか。つまり、今でも 1/3 以上は煩惱と言われるインターネットの世界を健全にしないと、子供達はますますすさんでいきますし、イジメが起きます。こういう世界でゲームの商売をやっている方々は、単に売上げを伸ばすだけでは社会責任を果たせないと思います。ゲーム業界にとっての未来観とビッグデータ観に関してお話をして頂けますか。

松原：特に日本市場では、規制や社会的な通念など、プレーする側も提供する側もあまり準備できずに急激に進んでしまった感があります。今の日本で一番特異的なゲームの要素は「ガチャ」でしょう。これは、簡単に言えば「くじ」の 1 つなんですけど今やソーシャルゲーム業界の中心的ビジネスモデルになっています。売上げを伸ばそうという側からすると、もっとガチャを引かせる方向に行くので、これを突き詰めると、健全性に問題が出てきます。一方でガチャというのは、昔から物理的なおもちゃ販売の機械としてあったわけです。こっちはそれほどの不健全さを伴わないで何十年間も存在していたのに、なぜバーチャルなガチャだけ不健全さを大きく騒がれているかがまだ明らかになっていません。

福袋のように中身が何だか分からないけど引くという行為には、何百年以上の歴史があると思います。ソーシャルゲームというサイバーの中で福袋を提供したら、ビジネス的に大成功してしまったというのが現状で、まだそれに対して健全さをきっちり担保するということには達していない。ではどうするか。青少年には購入できる限度額を設ける、もっとガチャで得られるアイテムの確率を分かりやすく表示するなどの改善を進めていますけど、まだまだ社会に広くコンセンサスを得られるレベルには至っていません。非常に難しい問題だと思いません。

個人的に言えば、ゲームの面白さの本質はガチャではないと思います。一方、収益を挙げるにはガチャに替わるような効率的な仕組みができていません。それはゲームの作り手の問題だと思っています。ガチャも選択肢の 1 つとしてゲームの楽しさを伝える仕組みを作りあげて、健全性と収益性を実現することが、次の市場の勝者につながると思います。この解が出てくるのは、もう少し先になるでしょう。

ビッグデータとの関連で言えば、今はゲームのデータの分析をして何が面白い／面白くないというところを見つけるのにとどまっていますが、次のステップは、ユーザがどういう個性・特性を持っているのか、各ユーザのプロファイリング・モデル化をして、じゃあこのユーザの特性に照らして、こういうゲームを提供しようというようなレベルだと思っています。ではこれが、インターネット上の他のアプリのサービスとしてすでに提供されているパーソナライゼーションとどう違うのか、それは、ゲームというのは最もインタラクティブなアプリの 1 つなので、ユーザの個性が最もうまく把握できる可能性があるという点です。まだ残念ながら、業界全体では手が付いていないところですが、ゲームのユーザのインタラクティブ性やチャットの分析などにより、データを一段抽象化して個人情報を取り除いた形でモデル化できれば、それを使って横展開——先ほどドメインが 1 つから横につながって展開するというお話がありましたけど、そういうことにつながるかなと思うのです。

ゲームにおけるビッグデータは、インタラクティブ性を伴うという特徴によって最もユーザの個性を含んでいると思います（雑音も沢山含まれています）。ゲームこそユーザが本当に楽しんでやっているものなので、そこにユーザの本質が現れるのではないかと。それをうまく引き出せれば、ユーザにとってもサービス提供側にとっても有益なユーザモデルの構築ができるし、そのモデルを利用したユーザとゲームのマッチングができると思っています。しかし、今はビッグデータを分析できていないし使いこなせていない。理論的な裏づけや実績もまだありません。

喜連川：究極的なゴールとして、個にフィットした楽しみを提供するという方向を目指すわけですね。

松原：はい、簡単に言えば、接待マージャンができる。接待マージャンは、皆さんご存じのように、強過ぎてもだめだし、弱過ぎてもだめだし、相手を気持ちよく勝たせてあげるというものです。

喜連川：しかしいずれにしても、ガチャを買うこととゲームを真に楽しむことは違うんだというロジックは、少々きれいごとのように聞こえます。つまり、多くの人々は情報検索の価値を十分認識している。だけどその価値に誰が金を払うのかというところに対しては誰も同意しなかった。だからしょうがないので、やや直球ではない広告がマネタイズ原資として入っている。それと同じようなことが、ガチャがゲームの中に入っているということとして起きている。人々は必ず本来の目的に対してお金を支払えるのかということに関しては、今の IT 観で

はやや難しい気もするんです。

先ほど松原さんは太古の昔から福袋があるとおっしゃいました。今、実は奈良で正倉院展というのをやっているんですけど、そこに双六の盤と玉のようなものが展示されていました。その説明に何と書かれていたかというところ、双六は最初はゲームだったのだそうです。ところが今と同じようにその双六で RMT が起こってお金を賭けるようになったらしいのです。それで政府は双六を禁止したと解説されていました。これが古来からの人間の煩惱というものではないでしょうか。そういうことを踏まえても、ガチャではない純粋なゲームの世界があると展望しているのかどうかということをお伺いしていいでしょうか？（笑）。

松原：例えば、韓国では政府がガチャを厳しく規制しています。オンラインゲームが若者の間で大流行し、そうすると起きてくる様々な問題への対応の一環として、規制されました。アメリカには規制がありません。アメリカでは実はガチャも福袋も好まれていないようですが、それは、アメリカでは中身が分からないものにお金を払いたくないという習慣が理由だと思っています。アジアは、日本を含めて福袋的なものが好まれていると思います。日本は今年、コンプガチャという、ガチャの部分的な機能について規制されましたが、その対応後もソーシャルゲーム市場は伸びています。さっき申しましたように仕組みが同じ実物のガチャが存在しているので、サイバーのガチャだけを禁止するのは公平性を欠くことになるでしょう。

喜連川：すると、アメリカのビジネスモデルは何なのですか。

松原：アメリカは、時間を節約するためにお金をかけるというすごくシンプルなモデルなのです。

ゲームには本や音楽と同じように中毒性があるので、早く進行させたいという欲求が生まれます。例えばジンガ社の CityVille という都市を育成するゲームでは、プレイヤーが自分の街に高層マンションを建てようとする、何人かの友達の手伝いが必要になります。そこで友達に「手伝って！」と声をかけ、友達が集まれば先に進めますが、集められるまでは一旦プレーが止まります。友達が多勢いるので3時間も待てば友達が手伝いを OK してくれるだろうと思う人はお金を払わない。すぐにでも続きをやりたいと思う人は、実際の友達に呼びかけないで、何十円程度のお金を払って CityVille の仮想の人を雇うことができます。

喜連川：その友人は現実の人間ですか。それとも人工知能のエージェントみたいなもので、プレイヤーのように見

せかけたソフトウェアという感じですか。

松原：CityVille では Facebook の友達につながっています。私がプレーしていて建物を建てようとなった時、3人友達の手伝いが必要だと言われます。では近山さん、喜連川さん、森本さんをお願いしようと思ったら私はポチ、ポチ、ポチと手伝いを依頼するボタンをクリックして3時間待ちます。もし今すぐ先にプレイを進めたいければ、何十円かを払って CityVille の仮想の人を、プレイヤーに見立てたソフトウェアですが、雇います。

森本：みなしでもう誰かが来たと考えるということですね。

喜連川：それはもうガチャと大して変わらないのではないですか。

松原：作り手がガチャと大きく違うと思っているのは、ランダム性がない点です。今の話だと何十円か払えば確実に私は予想している成果を得られます。

喜連川：確実性。確率的要素が入っているか否かが大きな違いだということですが、プレーしている側からすれば、やはり射幸心を煽られているという観点では、同じような気がするんですよ。

松原：それは映画でも音楽でも同じで、全てのコンテンツは中毒性に訴えないと商売できません。ある程度ユーザーの気持ちを煽るという姿勢はコンテンツの本質だと思われれます。だからそれを明確な形で示すのが欧米モデルで、ランダム性を入れているのがアジアで受け入れられるガチャということではないかと。

喜連川：明快なご説明有難うございます。

松原：コンテンツの本質の1つに没入感 (immersion) があるとされていて、その没入感を演出するために、惹きこむ、言いかえれば中毒性を出します。

森本：勉強中毒にさせると皆が喜ぶ（笑）。

喜連川：そうなんです。私は本当はそこに持っていきたいと思っていました。ゲームに没入感を持たせて、そこでストレス発散するというのも私は人生の中で重要なことだと思えます。しかし、ある一定レベル以上没入させてしまうと問題があるような時、もっと別のまっとうな (sane) 部分にもっていく方向性はないのでしょうか。先の勉強中毒のように。

松原：そうですね…

喜連川：例えば Linden Lab が Second Life を開発・運営した時にやった事業として、引きこもりの子供達を Second Life の世界で随分とトレーニングして、実社会で仕事ができるようにしたという実績を聞きました。全ての Second Life の事業がきれいではもちろんないでしょうけど、そういうものも CSR の1つとして明確に打ち出して

いました。ユーザから金を搾り取るためだけのゲームじゃない、もう少しほのぼのとした世界で「やっぱりビッグデータだよ」という例があると、情報処理学会らしいなと（笑）。

松原：いいですね、そういうのがあれば。

喜連川：無ければ無いでもいいですけどね。

松原：シリアスゲームと呼ばれるジャンルがあるのですが、エンタテインメントゲームとはまだ開きがあります。ゲームに関わらずエンタテインメントの目的はユーザを惹きこむ面白さにあり、それが最優先のプライオリティです。一方、シリアスゲームでは優先課題が異なります。例えばマナー教育とします。すると、身に付けるべきマナー全部をカバーすることが最優先の必須項目になってきたりしてしまう。そしてその途端に、エンタテインメントの面白さがどんどん失われてくるというのが、これまでのシリアスゲームですね。

研究者の方々が皆おっしゃるのは、ゲームはすなわち遊びであり、なかなか研究対象としにくいということです。ゲームを研究することはアウェーであるというネガティブな要素。一方では、研究成果などを表現する場においては、ゲームの判りやすさ楽しさを利用するというゲームのポジティブな要素。これら両面があります。後者の方はゲームの面白さの本質、本当に惹きこまれるものに関わっていて、作り手がどうやって作ろうかと考えている遊びのことで。シリアスゲームの課題は、ゲームの表面的な機能を利用するだけでなく、本質的な面白さを生み出す要素をバランス良く取り入れることだという感じがします。

喜連川：要するにゲームはもっと奥が深いものだ。学会としてそういうところをもう少し深掘りしていくのは、逆に言うと、有意義な要求かも知れないということですね。

松原：東京大学名誉教授の養老孟司先生はゲーム好きで、東京ゲームショーでゲーム大賞の選考委員長を永年務めておられます。以前、養老先生は賞授与式のスピーチの中で、ご自身が中毒になりたがり屋みたいなのところがあって、本も中毒だし、もちろん研究も中毒だしゲームも中毒だという趣旨のお話をされていました。

喜連川：ああ、なるほど。

松原：本も勉強もゲームも同じくらい中毒性をお持ちだということで、他のユーザに対しても同じレベルにできればいいのではないかと。

喜連川：世の中に役に立つ中毒はいいわけですが、ゲームは何の役にも立たないわけですね。

松原：それではそもそも、コンテンツや娯楽を全否定し

てしまいますね（笑）。

喜連川：養老先生がおっしゃっているように、中毒だから皆一緒と言われてしまうと、それはちょっと違うかなと、今思ったんですけどね。

松原：そういう点では、ゲームの目的はエンタテインメントとやはりリフレッシュです。ゲームに時間とお金を費やして、そこで得られた爽快感を生活の中に役立てていただければという感じです。

喜連川：私は足もみマッサージでリフレッシュするのが一番です。要するにビッグデータによってリフレッシュのパーソナライズが達成できて、この人はゲームでリフレッシュするのが一番いいだろうと分かると（笑）。

ビッグデータの2次利用とデータエコシステム

近山：今日は遅れて来てすみませんでした。今のゲームではまだまだユーザのプロファイリングがされていないということでしたが、例えば楽天の今回の招待論文には、複数サービス間でユーザプロフィールを共有することが重要だと書かれています。

購買サイトでは、ユーザがこういうものを欲しがっているらしいというプロフィールが比較的直接に出てきますよね。それが今後、センスしたデータがやたら多種多様になってきて、ゲームなど他の切り口から見ただころもプロフィールに加えられていきます。例えば、ここでそれぞれの紋章が刻まれた2つのドアがあった時に、こちらのドアから入りたがる人というのはこの本を買いたがるのか（笑）。現実はその単純ではないと思いますけど、そういった関連性はゲームだけでなく至るところで出てきそうです。

例えば楽天のように、一企業の中で閉じてやっているとこもあると思いますが、そうやって取得した情報をどうやって広く利用していくか。それは各企業それぞれの利益追求のためでもあるんですけど、結局のところユーザの快適性の向上に行き着くわけですね。これを実現するには、至るところにあるデータを上手に共有して使っていくことが今後大事になってきそうに思います。

松原：どういうユーザモデルを作ればいいのかという点では、恐らく広告等の分野の方が進んでいるかも知れません。心理学でもマイヤーズ・ブリッグズタイプ指標 (Myers-Briggs Type Indicator) など、個性の分類にはかなり歴史があります。

楽天は、物販あり金融ありトラベルありブログありゲームもありで、かなりグローバルにサービスをされています。

るので、閉じていながら多様なデータを得ることができると思います。さらに航空会社のマイレージなどがあると良いのですが、1つのサービスやプロバイダだけでは難しいので、恐らく都市レベルあるいは国家レベルで対応しないと実現できないかなと思います。

近山：そうですね。ひょっとすると逆に、そういった情報を個々の企業が独占することを禁止するというような制度が必要になってくるのかもしれない。

松原：私もその可能性があるかなと思います。そのためには個人情報のお話をクリアしなければいけなくて、面白い方向性かなと思います。

森本：2点コメントさせてください。まず、個人のプロファイリングとか遊びとか買った物とか、楽天のデータはコンテンツだと考えられます。これに対し、別の要因として、例えば天気とかその日はサッカーの試合をやっているとか別のイベントをやっているとか、環境情報がある。これらコンテンツとコンテキストの多分両方を、もっと組み合わせると、いろんな最終的な判断とかプロファイリングにつながるように思います。ここでのポイントは、コンテンツのデータは質が良くて数は比較的少ない。対してコンテキストは幾らでも広げようと思えば広げられる。だってそれこそ、違う国で大統領選があったという話って本当に関係があるのか/ないのかということまで行ってしまう。コンテキストには様々なものが入ってくるので処理が大変になります。

もう1つはデータ共有です。共有するのを強制するというのは最終的には難しいと思います。しかし、インセンティブがあれば、お互いデータを出し合うんですね。例えば銀行どうしでは、ブラックリストに相当するような危険情報などを共有しています。あるいは国どうしでも、犯罪者情報など本来すごく機密性が高いけれどもお互いのメリットになるようなものは積極的に共有しています。ヨーロッパの銀行間には ORX (Operational Riskdata eXchange) という仕組みがあって、本来なら自分の会社の独占的所有物 (proprietary) であっても、特定の相互メリットのためにリスク情報をお互い ORX へ拠出して皆でそれを共有しているのです。だから今回の「パーソナルデータの利用・流通に関する研究会」でも、何らかのインセンティブや共有するメリットをうまく共通化すべきだと思います。さっき喜連川先生がおっしゃっていた「その価値に誰が金を払うのか」というのは難しい問題です。ちょっとやそつとのデータならタダみたいなものですが、これが Peta バイト、Zeta バイトになると置いておくだけでもお金がかかりますので。

喜連川：法律で強制的にデータを提出させるというのは

もう論理的にあり得ない話だと思います。

松原：私のイメージは、個人情報を読み上げるといよりは、モデル化されたデータをリンクさせることでした。

喜連川：モデル化して個人にひも付かないビッグデータもすでに山のようにあって。そこで成功している例も幾らでもあります。それらをドメイン横断、あるいは相互参照 (co-cite) でくっつけることによって、ものすごく大きい価値が生まれる例は沢山あります。そういう領域が面白くて、私はそれをどう実現するかというメカニズムデザインがやはり一番難しいと思っています。

先日、内閣府の委員会でプレゼンをしてきたんですが、私の最後のスライドにはデータエコシステムと書いておきました。つまり、データを提供した者に対してどう利益再配分をするのかという、そのマーケットをいかに作るかということが究極であって、これさえできれば全部が動くと考えています。逆に言うと、それを実現するための規則やガイドラインというものをどう作っていくのか。これは自然科学の例を使って説明しました。

例えば、中国の山奥へ行ってデータを取ってきた人というのは、データをもらうために白酒 (パイチュウ) を3本ぐらい飲まなければいけなかった。だから肝臓を壊してこのデータを取ってきた。そのデータをなぜ無関係の第三者に無料であげなくてはいけないのかという話がある。そうすると、そのデータを使って書かれた論文に、その人の名前が必ず出る。つまり今の学術評価というのは、その最終知に対して評価をするというエコシステムが回っているんですね。

ところが今はデータのレイヤでエコシステムが回っていないんです。だから、いろんな実験をするたびにデータのレイヤでエコシステムをデザインし直さないと健全にいかないんです。データを取ってくれた人の労力、収集・整理した努力などに対して評価する見取り図を作らなくてはならない。

森本：その場合の唯一のインセンティブは、データを供出したら他の人のデータも見られるようになるということだけど、他の人のデータを見るモチベーションが無い人は供出するモチベーションもないということになるんですね。そこが難しい。

喜連川：例えば環境と健康というのは非常に密接なドメインで、データどうしのクロスも山のようにあるんです。だからデータのエコシステムを作った時、自分の所しか役に立たないデータと、いろんな所で役に立つデータと、どちらが多いかという、クロスで利用価値のあるデータの方が遥かに多いというのが私の予想です。

森本：もともとそのために集めましたというデータの取

得コストが一次利用で正当化されてしまっているならば、他の人が別の用途でデータを二次利用、三次利用するのは、ある意味、丸儲けになります。あるいは、何か理由があれば使っているよとなるかも知れませんが。

喜連川：いわゆるデータの来歴 (provenance) と言うんですけれども、例えば三次利用した時にそれが目的外使用かどうか電子透かしによって判定できて、目的外使用の場合でも、大元との契約がうまく成立し、利益がフィードバックされるようなフレームワークを作らなくてはいけないと思います。要するにそもそもデータの価値をもっと認めないといけない。

森本：そうですね。

喜連川：これは我々情報分野の人間にとって非常に大きなチャレンジで、データがアセットになるように是非そういう分野を今後開拓していく必要があります。私がデータはアセットであると言い出してからもう25年以上経っているんですけど、その間誰もアセットと言ってくれなかったのですが、今まさにそういう時代じゃないかなと。

森本：今の議論のポイントは、例えばお天気データを取得して天気予報をやって、エコシステムが1つ回ってしまっていると。その同じお天気データを使って他の人が別のビジネスをして儲けるかも知れないとなった時に、その利益がこの大元へ…

喜連川：戻るようにしないとイケないわけです。そういうことは今、すごく実際に起こっています。例えば、環境がらみですが、自宅に太陽光パネルを置こうかという話になった時に、雲はどれだけ来るのか、太陽がどれだけ自分の家の上を照らすのかということを知りたくて、まさにそこで天気予報データが活用されます。そこは天気予報じゃなくて、エネルギー予測で、別の利用形態となります。利益再配分がさらに進むようなことがたくさん起きていて、それが廻るようにしないとイケない。

森本：はい、起きています。

近山：その時、何が価値のあるものになるかというところが大変難しい。それと、元のデータから抽出したデータなのかどうかを識別することも難しい。

例えばある歌手の歌マネがあったとします。もちろん曲には著作権がついていますし、その歌自身にも著作権がついていますが、歌手の歌い方には多分著作権はついていません。歌を1回抽象化してしまうと著作権が外れてしまう。1回アブストラクションしたところで著作権以外のいろんな権利も外れてしまうでしょう。ビッグデータはまさにその抽象化のところが鍵です。どういう観点から抽象化したかということが新しいチャンスを生むわ

けですが、その時に元のデータは何だったのかがほとんど分からなくなっている。抽象化して得られたデータまでうまく透かしが入れられるかどうかという話だと思うんです。

喜連川：今、解けているかどうかという話ではなくて、そのデータが商品になった時に、データのトレーサビリティが効くかどうかというお話をされているんだと思います。それが解きやすい問題かどうかという観点からすると、ある場合では、まさに解きにくい問題になっているかも知れない。だからこそ、我々情報技術者の次のビッグデータのチャレンジなんです。

近山：そうなんですけど、この問題は解けるんだろうか。

まあ、多分部分的には解けて、うまく解ける範囲内では話はいまいくのでしょ。

喜連川：解けない場合は、原則として罰則規定を入れるしかないですね。個人情報保護法も解けない場合がいっぱいあるんですけども、

近山：まあ、そうですね。

森本：今もう普及しているデータとか無料で出ているデータはいいとして、将来本当に価値の高い目的でもってそのデータが欲しいとなった時、そのデータが最初はそんな価値があるとは思われてなかった、でもそのデータは本当は価値があることに気づくと。すると、俺のデータで3割も売り上げを伸ばしたでしょうみたいなことが起きる。ただ、特許の使用を追及するほどの厳しさを要請するかどうかは難しいですけど。

喜連川：そのようなことは今のビジネスでも起きていますよね。最初の契約で自分のデータはこの目的に対しては幾らで売るという具合に。予想以上にこんなに儲かるとは思わなかったけど、そうなっちゃったらそれはそれでいいわけですね。

森本：あと例えば、弁当屋がある人のデータを自由に使って3割売り上げを伸ばしたのはいいとして、では、ある人が今度から儲けの一部を自分によこせと弁当屋に言ったら、弁当屋は、じゃあデータが間違っていた時は損を払えと言うことにもなりますよね。だからそれは自己責任とのバランスがとれるようになっているわけで。

喜連川：それはさっき言った天気予報データの場合だと、これだけ晴れるぞと言っておいて実際は雲ばかりだったら太陽光パネルは全然何の役にも立たない。おいお前、どうしてくれると。だから、そういうケースで生じる不足額も扱うような契約が必要で、つまり、最終物ではないデータのレイヤーで我々はシステムをきっちりと考えていく必要がある。

森本：そうすると、ビッグデータのもう一つの課題とし

て、最初からデータの信頼性というサービスレベルを規定する、この天気予報は 8 割しか当たらないんですとか 6 割ですと言っておいて、そのサービスレベルで契約して頂く。通常のビジネスで今おっしゃったデータ売りをすると、そういう形になりますね。

喜連川：今はすぐ初期段階なので、皆さん違和感を抱いたり、こんなことでもいいのかということになると思うんですけど、今後 10 年ぐらいの長いスパンで見ると、そういうものが当たり前になってきて、逆に当たり前になるからこそ、そういうデザインをしっかり研究することが重要になると考えています。

ポストビッグデータ

松原：ビッグデータ活用の究極として、私は検索エンジンが要らなくなるだろうと思っています。検索エンジンって、明示的 (explicit) に自分が問合せに行かないと答えが戻ってこない。ところがビッグデータを活用して、知らない間に自分に関係することが暗黙的 (implicit) に判断できるようになってくれれば、検索エンジンで自分が探さなければいけない、どこそこに何か求めるということをする必要がなくなる。

近山：知りたいなという思う前に、答えがそこにあるという (笑)。

喜連川：そこまで言うとは何ていうかサイエンス・フィクションになってしまうから。

松原：そこまでは言い過ぎかもしれませんが、今明示的に検索エンジンに入力している情報から、もっとふだんのアプリを使っている中で得られる情報に移ってくるんじゃないのかなと。

喜連川：その議論はいわゆるインテントマイニング (intent mining) といって、もう過去 10 年ぐらい延々やってきた研究課題なんです。ところがユビキタスもアンビエントも同じで、もうお節介は勘弁してくれと (笑)。その達成レベルと我々が要求するレベルとは相当ギャップがある。私は Watson にはものすごく拍手喝采するんですけど、今松原さんがおっしゃっているレベルは、Watson のレベルよりもっと遥か上なので。だから、それをあまり今、やれそうだという感じは、私なんかは全然していません (笑)。

松原：分かりました。

喜連川：先日、ガートナーの IT イベントでのメッセージに、フォーチュン 500 の企業でビッグデータの試みは 85% は失敗するというのがありました。ビッグデータは必ずしもバラ色でないものだとした時、我々は次に何をすべきかというようなヒントを、ちょっとここで議論

したいと思います。ポスト・ビッグデータは何でしょうか。この議題はちょっと難し過ぎるかも知れないんですけど、ビッグデータがまあまあそこそこ色々利用可能になってきた時に、我々が次の切り札として取り組むべき重要な技術課題はどこだろうというもの。

森本：今うちで議論をしているのが、ビッグデータはデータがたくさんあって、いろんな種類があって、それから流れていて、すぐ消えるものもあり、そして曖昧だと。それを Volume, Velocity, Variety, Veracity の頭文字をとって 4V (フォービー) と呼んでいます。でもそれはデータのスタイルだけしか表現していない。全部のデータを無目的に処理するわけではないので、そのデータを実際に使うかとか、どう使うかとか、そのシナリオとかまで踏み込まないといけない。それは多分ビッグデータの次というよりは、ビッグデータをどう使うかという…

喜連川：それはもう十分ビッグデータの範疇ですよ (笑)。今はプレディクティブ (predictive) だと言われていて、次はプレスクリプティブ (prescriptive) だと言われています。また、今まさにおっしゃられた目的/目的外というのは、いわゆるダークビッグデータのことです。いろいろそういうパスワードがあるわけですが、はてさて未来は?

松原：ではビッグデータの前というのはどんなイメージだったのでしょうか。何からビッグデータに来て、次は何かというと、A から B、B から C となるので何となくちょっと想像が付きやすい。

喜連川：アメリカ OSTP 的に言うと、インターネットとスパコンが A で、B としての次がビッグデータだと見なされています。つまり、このインパクト感でビッグデータをアメリカは取り上げています。インターネットには、色々ネガティブな側面が多いですよ。いい所が半分、悪い所も半分あるような感じ。でもとにかく世界を変えた。これだけはみんな認めるんですよ。スパコンも地下核実験の代替という意味ではやはりネガティブな側面も大きい。しかし、このコンピューショナル・サムシングというのはやっぱり世界を根本的に変えた。計算が変えた、ネットが変えたと同時に、データが変える時代になった。これがビッグデータだという言い方になっています。そうすると計算とデータとつながる線 (ネット) があり、その次まだあるのか、もう終わりなのか。

森本：その流れで言うなら、IBM 的にはコグニティブと言っています。

喜連川：そう、コグニティブですね。

松原：何か表現系のような気がちょっとします。計算能力が高くなったおかげでデータ処理は速くなった。インターネットのおかげで情報のディストリビューションは

安価になり速くなった。でも見せ方というか、コグニティブにつながるかもしれませんが、人間の認知を助けるようなところには行ってないと思います。

森本：人間の感覚にもっと寄ってくるということですね。

松原：いくら計算が速くなっても、結局、最後の表現が悪かったら分からないで終わってしまうと思っているので。

喜連川：ただ、人間に近くなるというのは、もう昔から結構もう繰り返し繰り返し言われてきたことではありません。これだけ計算能力があるんだから、本来の計算の部分はもう十分過ぎるぐらいパワーがあるので、残りは全部ユーザインタフェースに使うよみたいな議論というのは延々されてきました。難しいところかも知れないんですが、その辺でビッグデータの次の話があれば。

松原：ちょっと直感で申し訳ないんですけど、機械から人に何か届けるのはこの10〜20年間全然解決していないんじゃないかなと。一方では、そこに届けるまでの計算速度とか量とかは、ものすごい速さで性能向上したような気がします。言ってみれば、ラスト30センチでいいんですけど、この世界って何か進歩してないなと感じます。そこは、翻訳かもしれませんし、何か単なるプレゼン能力かもしれないし、ビジュアルライズかもしれませんけど。

森本：いまだにハイパーリンクのリストが出てくるだけですからね。

松原：そうそう。相変わらず、そういう点では、最後の人間に認知してもらうところのプロセスというのは同じかなという。そこまでの手段は、あらゆることでスピードが速くなっているのという気がします。

喜連川：この前、内閣府の会議の場ですが、要するに我々電気系はマックスウェルの方程式で飯を食えている、土木建築屋さんはニュートン力学とかボイル・シャルルの法則がある。要するにある種、非常に基礎的な、かなり予測可能な定式化ができている領域なんですよ。これが生物というものに一步踏み込むともう全然違う。人間ではなくとも、今、植物工場とかで農業をやっていますが、予測が殆ど当たらないんですよ。つまり、対象の複雑さ (complexity) の空間がもうべらぼうに大きい。だから Watson で医療応用と言われても、実はやれることというのは極めて限られているとしか我々は言わざるを得ないのも事実ではないかと感じます。そうすると、人間にとってというのを言うのは簡単なんだけど、もうちょっと身の丈に合ったステップを踏んだほうがいいのではないかなと私は思います。やはり人は、想像を絶する複雑さがあるような気がするんだな。

森本：そうですね。単位電力あたりの処理能力という観点で見ると、人間とコンピュータの間には、まだ10万倍ほどの開きがありますからね。

喜連川：そうでしょうね。だからリーズナブルな目標を立てながら研究する。それが物理学とか天文学とかは上手なんです。コンピュータ・サイエンスの場合、連続的な中間目標が設定しづらいかも知れない。人が目標というのは言いやすいんだけど、そこはある種、究極なんです。その究極の手前を誰がデザインするかで勝負が決まるわけです。——という議論をしていると際限がないので、ここで止めましょうか (笑)。

近山：というか、今、コンピュータはアウトプットはアウトプットだと思っているでしょう。そこを、こちらの表情を見ながら表示してほしいよねというのがあります。つまり、講義していたって、学生の顔を見ながら「あつ、こいつ分かってない」と思いながら言い方を変えたりするでしょう。コンピュータはそれをやってくれないけど、「これ何？」という顔をしたら、そういう説明をして欲しいんですね (笑)。そんなの分かってるよという顔をしていたら、もうそこは省略したりとかね。

喜連川：こんなの分かってるよというような顔の学生がいるんだけど実は何も分かってない (笑)。

ポストビッグデータはなかなかクリアな道が見えたとはいえませんが、ビッグデータに関しましては、多様な論点を全く異なる応用を推進されておられるお二人から、新鮮な切り口で議論ができ、読者の方に従来とはやや異なるデジタルプラクティスならではの新しい視座をご提供できたのではないかと感じております。森本さん、松原さん、近山先生、本日は本当にありがとうございました。