

ビッグデータの未来

Valuenex コンサルティング株式会社 代表取締役社長 中村達生

目次

1. 人間の知覚限界を超える
2. 俯瞰分析
3. GOOGLE と APPLE の戦略を読む
4. データサイエンティスト
5. 空白地帯がビジネスチャンス領域
6. GOOGLE と HONDA のコラボレーション
7. 予測分析
8. アルゴリズム産業
9. 可視化の効果
10. 質疑応答

1. 人間の知覚限界を超える

本格的なビッグデータ活用時代をむかえ、本稿では大量の情報をいかにして把握するか、それから得られる情報でどのように未来を予測するかについて考察する。

2020年に、世の中の情報量は45ゼタバイト（テラバイトの10億倍）になる。更に2045年には人間が機械でやることにキャッチアップできなくなる「シンギュラリティ」が起きると言われている。

最初にこういう問いかけをしたい。これだけ情報の洪水に対処できるのは Panoramic View Analysis すなわち俯瞰解析である。いろんな情報を正確に、1枚の絵に落とし込めれば、そこから、いろいろな情報が得られる。「読む」のではなくて、「見る」にすれば、実は時間を削減できる。気付かなかったことが見えてくるのではないか。

諸説あるが、人間の頭の容量は4テラバイトとか12テラバイトと言われている。多分半分以上は生活のために使っているし、全く使われていない領域があるので、自由に使える領域は多分ハードディスク1個分もないのではないだろうか。

この非常に少ない頭の容量は、当然2045年になっても増えない。しかし、データは増える。皆さんは業務中にいろいろなドキュメント読むが、読んでいると多分間に合わなくなる。情報を把握する方法を今から変えないと、多分5年後、10年後には誰もが情報の洪水に流される。これを何らかの方法で、可視化する、あるいは何か別の方法で簡易化できないだろうかと考えている。

2. 俯瞰分析

書類の積み方には個人個人のクセがある。几帳面な人はきちんと分けし、分類化されて出てくる。分類化というのは非常に効率的なのだが、イノベーションは生まない。要するに昔の体系で分類している。既存の分類だから、そうした人に限って、新しいドキュメントが来るとどこに分類したらいいが分からなくなり困ることになる。面白いやり方の一つが、あの野口悠紀雄さんの超整理学で、新しいものを左から入れていって、右から見ていけば右が全部古いものになる見方をする。そうすると常に右から整理していけば良いというやり方である。しかしそれは一次元である。そこで、二次元的に表して、何らかの方法で可視化できれば、分布の特徴で、何が起きているかって読めればよいのではと考えた。

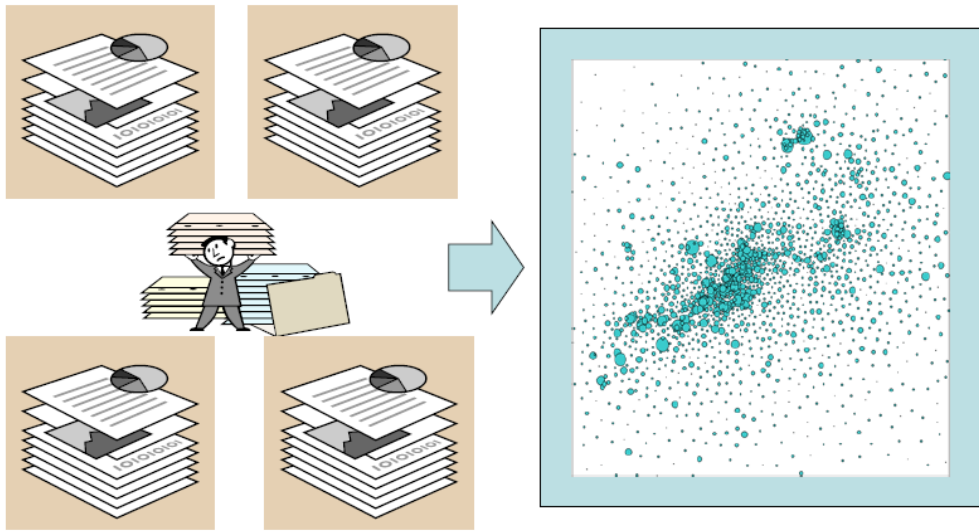


図1 大量の書類の処理「読む vs 見る」

図1右のドットの一つは、一つのドキュメントに対応している。このドットの距離っていうのは類似性の強さを現している。どのくらい内容が近いかということでドット距離が測られている。密集している領域には、たくさんの同じような文書があるのだということが分かる。この空白のあたりは、あまり文書がないということが分かる。それから離れた文書があるのだなということも分かる。

図2はヘルスケアという言葉を含む米国特許を単純に集めて可視化した図である。2万2763件の特許が該当する。ヘルスケアと書いているものを、明日までとにかく把握しろと言われたらこれに対応しなくてはならない。

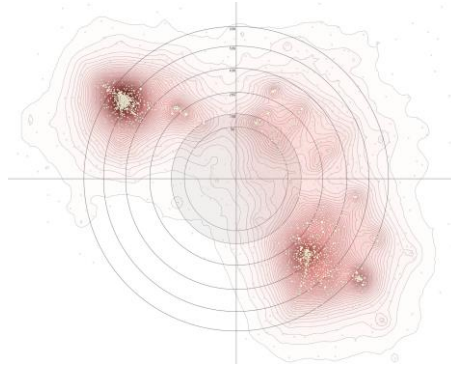


図 2 22763 件のヘムスケア技術特許の俯瞰解析

この俯瞰解析図、われわれのツールを使って得たものだが、大きく分けると二つの領域がある。一方が IT 関係で、デバイスなどである。最近よくある電子デバイスを使って、体調を測るというものである。もう一方は創薬を中心とした技術領域で化学関係のものである。

細かい等高線があるが、中間地帯、実はここに空白地帯があり、この空白地帯が脳科学の領域である。想像を巡らせていただきたいが、化学的な知識と IT 関係の知識を取り合わせたものが脳科学なので、非常に理解しやすい。現状ではこの領域に、いろいろな会社が向かっている。

しかもここに、キラッと光る技術がちらほらある。こういうときに、例えば自分たちがデバイスを作る会社だとすると、どうしたらいいのかという戦略を、この俯瞰図をベースに立てることになる。また、反対側の創薬分野には、人の臨床データと技術を持っているので、脳科学の人たちと組んで、真ん中を目指そうというやり方もある。両方持っている会社もあるのだが、残念ながら事業部が分離、独立しているので気が付いていなかった。その場合、この俯瞰図を見せて、実はあなた方が一番アドバンテージあると言うと、理解する。

この俯瞰図は時系列で見ていくと、どの会社がどこへ動こうとしていのか分かる。同時に、スタートアップカンパニーが結構存在しているのも分る。このコアの所は多分もう確立された技術だが、その終点では技術が急激に萌芽・成長してそれがスタートアップだったら早い段階で買収したほうが良いという戦略が提案できる。これがアメリカ型モデルである。

こういう分析をしていくときに、いわゆる SWOT 分析をする。強みは何で、弱みは何で、脅威が何で、チャンスが何だと。その中でも萌芽領域とかエマージング領域、それからホワイトスペース領域を見つけるというのが極めて重要である。

こういうビジュアライゼーション、可視化というのは、今たまたま特許の例をご紹介したが、ビッグデータ、つまりニュースデータ、ウェブ上の SNS データ、レビューデータなど、何にでも使える。

これをデータ別のレイヤーに重ね合わせていき、技術シーズ情報としての特許論文から一番マーケットに近い社会を表すような情報としての SNS データ、そして製品情報、これをうまくつなぐことによって、このマーケットに一番、この動きが一番近い商品は何かが見えてくる。また仮に商品がなければど

ういう技術が今後売れそうなのか、あるいは商品も技術もなければ、その空白に何を作るべきなのかというのが見えてくる。

日本では、よくビッグデータの活用とか、データがたくさんあるから、何かできないかと言うが、これはシーズ型、下からのボトムアップ型アプローチである。しかしマーケットが見えていないので、発散して、どこにも至らないで終わる。

しかしながら、Amazon の例もそうであるが、彼らは上から来ている、マーケットから来ている。マーケットから来て、何が必要で、コストが掛かるから無人で配達したい、そのための制御が必要だ。情報として何と何が必要だと、これを結合するためにどうしたらいいか、クラウドが必要だ。このように市場からアプローチしているのでごく自然につながる。

「もうビッグデータではない」という話がある。あえてスモールデータという人もいるが、私は、アプロプリエイトデータ、要するに適切データとっている。必要な情報の種類と量は人によって違う。だから、ビッグデータとは何かというときには適切な量にして分析をする必要がある。

何が適切かっていうと目的によるが、精度がどのくらい必要なかっていうことが重要である。先ほどレーダー図で可視化したけど、他のものと決定的に違うのはやはり精度である。ドットの位置が多少ずれただけで全然違う結果になってしまう。それがいかに精度良く配置されるかというのが非常に重要になってくる。

3. Google と Apple の戦略を読む

図 3 は Apple と Google の特許を一つの絵にしたものである。Apple と Google の特許には二つの種類がある。一つは、Apple と Google が出願人として出したもの、もう一つは Apple と Google が権利者として保有しているものである。両方が混在している。だから、ここからいろいろなことが分かる。例えば、Apple と Google が特許を出してないが、後から権利を購入して現在持っている、しかも最近急激に増やした特許は、その人たちが非常に興味を持っている特許であると理解できる。かつその特許は死滅しておらず、活かしているとなるともうそれは絶対事業としてやると判断できる。そして今、商品として現れてなければ、近い将来出るだろうということがほぼ予測できる。

これを自動で計算して、最近権利化した中でかつ急激に伸びている特許を選択すると、我々のツールではサマリーが出てくる。そうしたらサーモスタット（感温スイッチ）が出た。結局同社は、サーモスタットの会社を買収した。リモートでサーモスタットを制御するのは、彼らの中ではかなり重要である。周辺を今、研究し始めているが、彼らのコア技術は何も変わってない。

Google のコアというのは、サーチエンジンの技術と AdSense の技術である。サーチエンジンはいかに検索するのか、類似性をどう測るかという話、もう一つの AdSense はどうやって広告出すかである。しかしそれ以外のところは全部買収している。

その中でも特にサーモスタットの技術はこのあたりにあり、等高線で密度が濃く出るのが、そこが

急激に出てきている。これは、実はポートフォリオ的に見るとかなりリスクのある取り方だと思っている。要するに Google の得意としている領域と随分かけ離れていてシナジーがみられない。だから、次に考えられるのは、この間をこう買ってくるだろうということが推測できる。この周辺を捉えて、こちらにつないでくるだろうと。

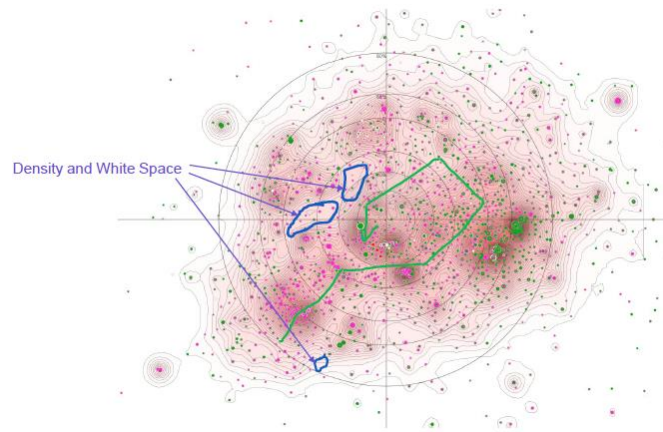


図 3 Google 特許の俯瞰解析図の一例

もう一つの特徴として、空白地帯が見える。幾つかの空白領域を分析してみると、一つはノイズリダクションの技術であった。ノイズリダクション、これは BOSE やソニーが非常に強い技術であるが、Apple というオーディオ機器もあるようなイメージがあるし、Google も Android でスマートホンをやっているが、実はこの部分が空白なのである。

バーチャルリアリティというときに、われわれはどうしても 3D と画像のほうに注目しがちだが、人間の知覚というのは 6 割画像で、次が音である。そうすると、音を持ってないことは結構弱みであると推測できる。画像が現実と同じぐらいのものが出たときに次にそのリアリティを高めるのは音である。しかし、彼らはノイズリダクションの技術を持ってない。ここに着目して、この技術を持っている会社を売り込めば、これは強い。こういう戦略が、実俯瞰解析図から、いとも簡単に見える。

4. データサイエンティスト

ツールではこのように出るのだが、これをどう読むかというのは、やはり少々経験が要る。そこで、シナリオライティングやビジョナリーが必要になる。いわゆるデータサイエンティストのスキルが必要である。データサイエンティストには三つの要件がある。一つは IT スキル、二つ目は統計的スキル、三つ目、これがビジネスセンスである。ビジネスが見えるということである。

アメリカでは、データサイエンティストたちが今 Google、Apple や SAMSUNG などに囲い込まれおり、なんと去年の年収が 1 人当たり 5000 万円という高額になっている。そこで、われわれは去年シリコンバレーに会社を設立し、データサイエンティスト要らずのツールを作ろうとしている。

ちょっとしたトレーニングすれば、データサイエンティストがいなくても、ビジネスができるという

売り込みをしたら、あれよ、あれよといろいろな人から引き合いが来ている状況である。

5. 空白地帯がビジネスチャンス領域

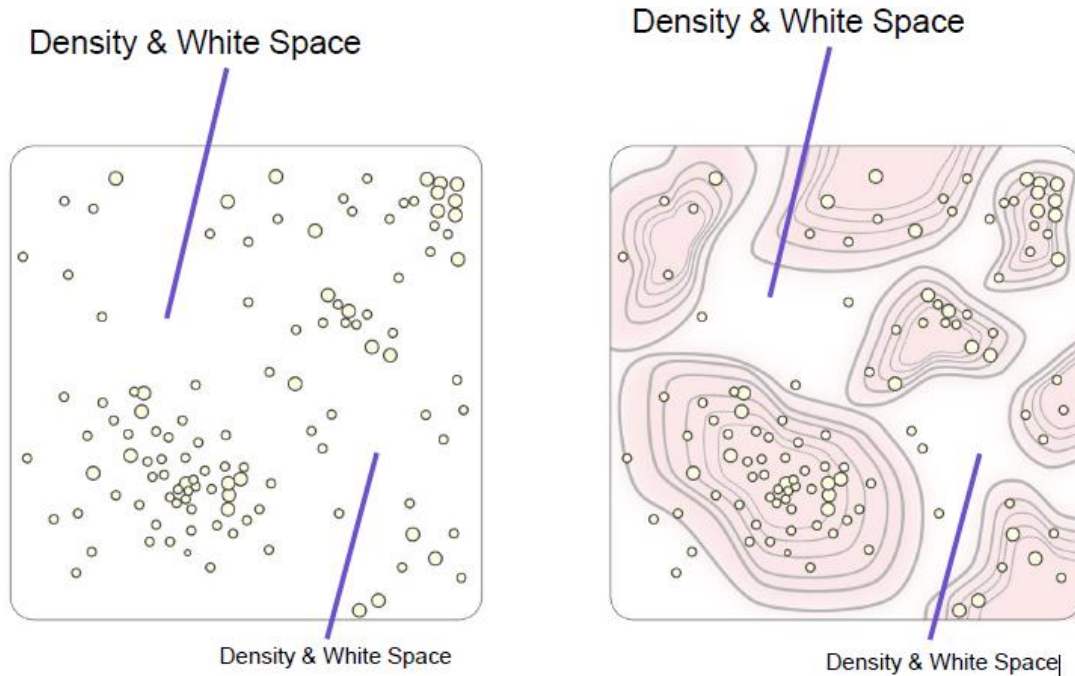


図 4 俯瞰図の中にある空白地帯と高密度地帯

図 4 左だけ見ると、何やらドットがちりばめられていて、よく分からないが、密度に合わせて等高線を書いていくと空白が見えてくる。この空白が非常に意味を持つ。例えば、研究開発で特許をこのように分析してこの空白が出てきたときに、三つの考え方があがる。

一つは、まだ誰も気が付いてないニッチな領域で本当にそれが特許として出てない。二つ目は、もう実は実用化されていて技術もあるのだけれど、秘匿していて、特許を出してない場合。例えば製造プロセスに使う技術とか、こうしたものは特許としては出にくい。三番目が永久機関など、あったらいいが物理現象的に不可能なもの、そうしたのも特許は出ない。

二つ目が以外と分かりやすい。ぽっかり穴が開く。なぜかというところコアの技術は模倣を嫌い秘匿し特許出願しない。しかしコア特許を出願しないときは周辺特許を出すからである。だからエッジが効いた穴になる。自動車の塗装技術などの特許はそのままでは出ないが、周辺の刷毛の技術や、植毛の技術など、ありとあらゆるものが出てきて、間接的にわかってくる。

6. Google と Honda のコラボレーション

図 5 は 2 年前にシリコンバレーのマウンテンビューという所にある Google 本社の知財責任者に見せたものである。

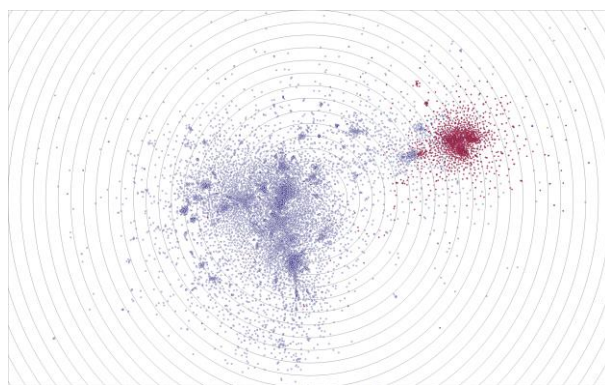


図 5 Google とホンダの保有特許の俯瞰解析図

Google とホンダの話である。アメリカに出ている Google の特許とホンダモーターの特許を同時に、一つの絵にプロットすると、これだけ技術領域が違う。当然のことながら一方は自動車メーカー、一方は検索エンジンの技術である。しかし、よく見ていくと、非常に近い技術がイメージセンシングの技術である。2年前であるから、ようやく Google のドライバーレスカー、自動運転走行車の技術が出てきたころである。こちらは、自動運転という概念で、イメージセンシングテクノロジーを捉えている。それに対して、ホンダモーターは、衝突安全という概念でこの技術を捉えている。でも中身は同じ。こういう状況のときに、Google がとるべき戦略は、コンピートして行くのか、それともコラボレーションして行くのかを投げかけに、マウンテンビューへ行って話をした。他のことはいろいろ話すのに、なぜかこの話だけは一切ノーコメントであった。資料も不要と言われた。そしたらなんと6カ月後に、彼らはコラボレーションをした。

7. 予測分析

ここからはプリディクティブ・アナリティクスの話である。先ほど、データサイエンティストの話をしたが、データ統計、計算能力、仮設定能力、この仮設定能力が、ビジョナリーであり、とにかく大事である。日本で、大企業の中で仮設定能力は、上司が喜ぶかどうかという仮設定である。そうではなく、もう少し未来から、もう少し世界からこの仮設定を設定しようという能力をいかに持つかが重要である。これは、ウェブサイトのリコメンデーションである。Amazon で何かを買うと、こんな物を買った人はこんな物も買いましたっていう、リコメンデーションがあるが、これはプリディクティブ・アナリティクスではない。自分は気が付いてなくて、システムが知っているだけであり、他の人が買った情報を流しているだけである。しかし日本で今、まだこれがある意味プリディクティブ・アナリティクスだと捉えられている節がある。

例えば先ごろ IBM のファミリーマートのシステム担当をされている方が言っていたのは、売れ筋の商品のランキングを出して、そのランキングで上位に上がっていた物で売れていなかった場合、それをお勧めしましょうということだそうである。上位の一位が唐揚げで、二位がポテトだそうである。ポテトを買った人に、「唐揚げもいかがですか」と勧めたらどうでしょうと。

これはなんの考えもない誤りである。油物が好きな人には2タイプある。すなわち唐揚げ派とポテト派がいる。ポテト好きな人に唐揚げを売ったら、それは油ものでメタボを進行させてしまう。

ここには、全く仮説が設定されていない。モデリングができてない。システム上で顧客の顔を見ず、現場を見ないでただ単にモデルだけでやっていたらとんでもないことになる一例である。

シリコンバレーのアパートでは電気料金が毎日スマートメーターで計測されている。昨日までどのくらい使ったかは、すぐ出る。ここからさまざまなサービス出てくる。このまま使ったら、一般的な広さの家よりも多く消費するなどの警告や、逆にまだ余裕があるからと新たなサービスを勧めるなどである。

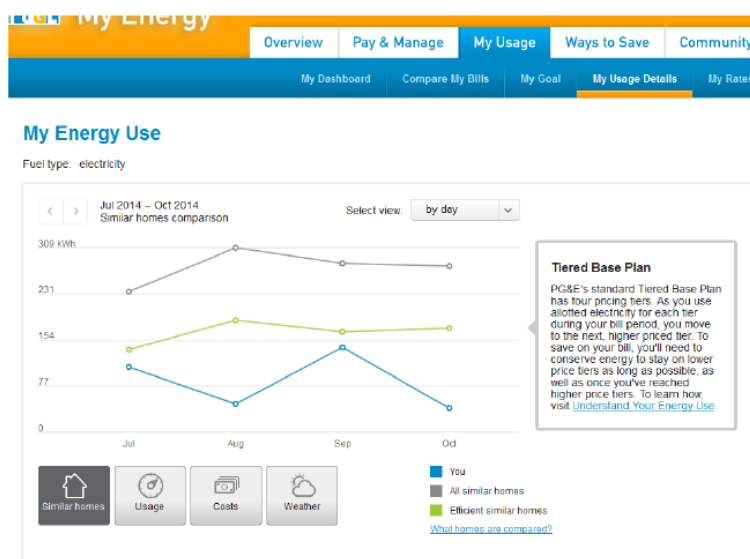


図 6 My Energy Use の電気家計簿の例

これは本当のプリディクティブ・アナリティクスではないかも知れないが、しっかりとモデリングができています。東京電力が最近、電気家計簿を作ったが1カ月単位、そのくらい遅れている。

8. アルゴリズム産業

ビッグデータがこれから非常に伸びる時に、日本として何をしていくのかという問いに、一つの答えになるかもしれないと思っているのは、アルゴリズム産業という考え方である。これは私の造語であり、そのような産業は、今は存在していない。

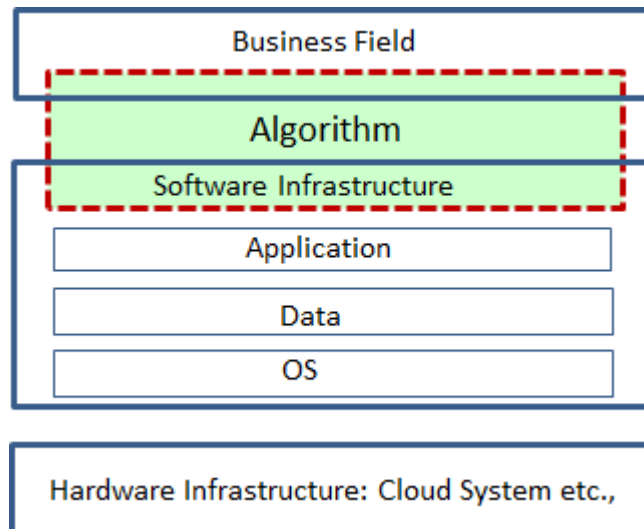


図 7 ビッグデータにおけるアルゴリズムの位置づけ

日本でノーベル物理・化学賞が多いのは、ハードウェアが劣るので、手書き計算で戦える分野であるからという人もいます。そういう意味で、このアルゴリズムに特化するのには得策ではないだろうか。例えばハードウェアやOS、アプリケーションは全部アメリカに取られてしまった。ハードウェアはもうクラウドになっている。一般の人は、もうどんなハードウェアが使われているかあまり気にしなくてもよくなっている。しかし、ソフトウェアから上の部分というのは、どんなビジネスに対してITでどうサービスをするのか、自由に発想できる。それをモデル化するというところで新しい産業をつくっていくべきではないか、アルゴリズム産業はそこをやろうというものである。

クラウドというのは非常に便利で、実はわれわれの仕組みはAmazonのクラウドサービスを使っているが、仮に、Googleのクラウドサービスがもっと良くなったとか、他でクラウドサービスに良いものがあれば、いつでも乗り換えられる。しかも下のソフトウェアのアーキテクチャーが多少変わっても、アルゴリズムは書き替え可能である。ロジックさえしっかり持っていればよい。

ビジネスとソフトウェアをつなぐ部分、身近なところでどんなビジネスができるかというのがこの図8である。これを縦に見ると、データベース、ハードウェア、ソフトウェアがあり、そしてビジネスがある。一方、横に見ると技術、製品、マーケット、ファイナンスとあるが、みなばらばらに専門家が事業をしている。例えばこのテクノロジーというと、知財ビジネスというものがある。でも、実は、こうマーケットとつなぐことによって、そのマーケットに必要な技術、あるいは、マーケットに必要な特許というのをうまくコンビネーションしていけば足りないものがなんであるか、どういう技術を今度作らなければならないかということが全部見えてくる。

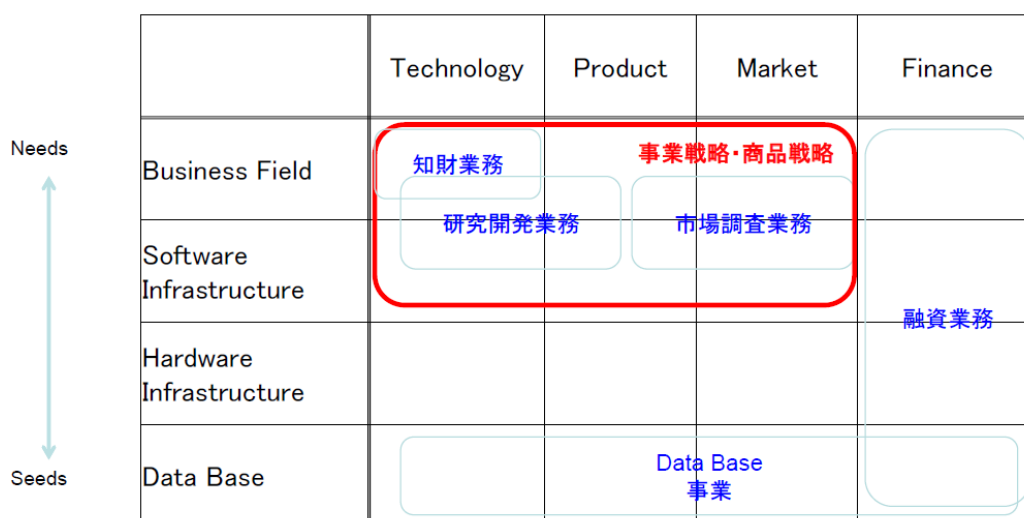


図 8 価値創造に向けたビジネスとソフトの戦略マップ

9. 可視化の効果

たくさんの情報を可視化すれば何ができるかと言えば、時間の節約、コスト削減、そして再現性の向上がある。誰がやっても同じ結果が得られる。専門知識が不要というのものもある。これまでであれば、ある技術については、その専門家に聞かないと分からなかった。あるいは、この部分は誰々に聞かないと、これは特許かどうか分からなかった。でもこれからいろいろな技術が融合する、新しいジャンルができると、もう専門知識だけに頼っていたのでは分からない。だからこういう新しい方法で、見ていく必要がある。

この図 9 がわれわれのツールの一画面であるが、レーダー図があり、いろいろなグラフがあり、この上にサマリーが出るようになっている。

アメリカのデザイン会社と改良検討をしているところである。日本のソフトウェアベンダーが作るとこのような感じになるが、世界で戦っていくには、やはりユーザーが目に触れる所が世界標準にならなければいけない。少し費用は掛かるが、世界標準のインターフェースを取り入れて、当然技術は戦えるもので、かつオリジナリティー、それからフィーチャー、日本らしさが入っていることが非常に重要であると思っている。今そのコンセプトづくり作りをやっている。

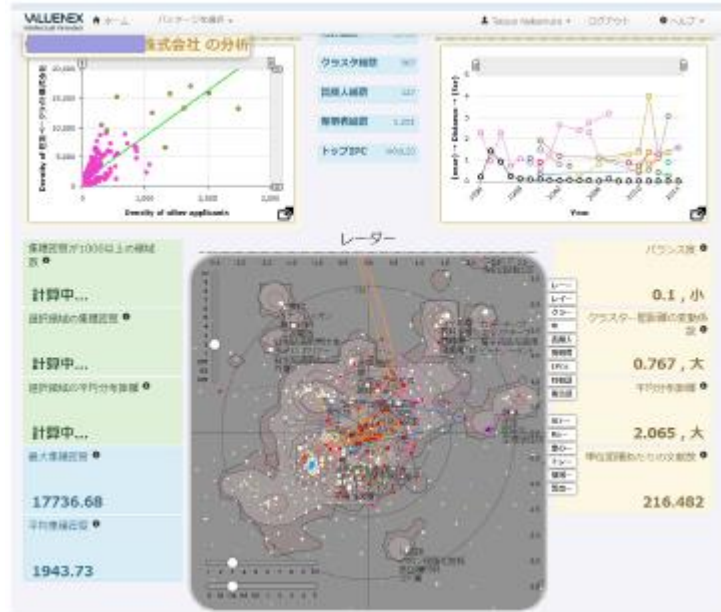


図 9 Valuenex が提供する解析ツール画面

講演終了

10. 質疑応答

質問者 1: ありがとうございます。ビッグデータの応用の可能性というのを一つ聞きたいのですが、ユーザーの潜在需要を見つけるということに印象を受けたのですが、このテクニカルは？

中村: ビッグデータの潜在需要探す方法ってということですか？

一足飛びに新しいものというよりも、今のものをリブレースできることからしていくと、結構あると思います。情報がうまく流れていないので、そこを連携したら非常に効率化するし、ビッグデータとして扱えるだろうと思います。例えば、レセプトデータとかですね、医療関係は結構、規制もあるので情報が壁で囲われちゃってそれが可能かどうかは別としてですが。そういう情報が閉じているところ、を、オープンにすると、新しいビジネスってかなり需要として出てくるのではないですか。

3年前 2012年の夏に、世界銀行がワシントンでオープンパブリックデータコンファレンスを開いた。国とか官公庁の持っているデータはどんどんオープンにして誰でも使えるようにしよう。そこでどんなアプリケーションを作るかっていうのは、民間に任そうと、こういう発想なのです。世界 33カ国の国と地域が参加している会議ですが、日本からはなかった。非常にショック受けました。

質問者 2: 25年前に、事例ベース推論というのをやっていたんですけど、AIをやっていたのです。その頃から思うと夢のような話になってきていますが、あくまでもデータの密度っていいですか、当時はモデルを入れるっていいですか、概念と概念の、そういうことによって、うまくデータの少なさを補っていたような気がしますけれども、さっきおっしゃっていたようにこういう解析したレーザーチャートからう

まく、分かりやすくするような解釈は何か、モデル、行動、人間思考モデル、あるいは、さっきイノベーション、技術、ファイナンスなど、あるでしょうか。

中村：そうですね、あの、これはあくまでも入り口なのですね。分析の入り口なのです。こう、地図があって、その地図をどう読み解くかっていうのは、その、目的に対してモデル化をしてかなきゃいけません。これは特許、技術なので、大抵の場合、使うのはSWOT分析です。強みが誰、どこで、弱みがどこ、あるいはコンペティター誰で、その人たちは何をやろうとしているのか。で、こういう情報がどうやったら簡単に出てくるかと、読まないでも、なるべく、なるべく、よ、さぼってですね、なるべく簡単に出すにはどうしたらいいか、というのが、こう、モデル化されて、入っていきます。

司会者：今日お話していただいたその、最後のところ。イノベーション政策と。やっぱりビッグデータとですね。こうも日本のその、イノベーション政策の中で、どううまく取り入れていくかということで頭悩ませていたのだという話をされましたね。中村講師のお話聞かれてどんな接点がありそうですか？

質問者3：おっしゃるとおり、出口から考えなくてはいけないっていうの、まさにおっしゃるとおりで、そこにアルゴリズム産業、非常に、納得感が高い話だなと思って聞いていました。その上で、じゃあ、医療のカルテを、もうどんどん電子化して、地域を越えて共有しましよみたいな話は、念頭にはあるんですが、それ以外でこのビッグデータという世界に対して、行政がどうアプローチできるのかなっていうとどうでしょうか。

中村(講師)：そこはもう、データ解放だなと僕は思っています。

司会者：人材育成ってことをね、いろんな人が言っていますが、それは、どうなのですか？

質問者3：人材、人材育成って結構その、何でもワードになっていまして。仮説設定力みたいなお話がありましたけど、どう、何をどうして、どういう能力を持つ人が本当に、社会に求められているのかっていう、もう一步具体化しないとジャストの人材育成っていうワードはよく使われているんですけど、実際何かやろうとすると、具体性に欠ける、という言葉なのです。単純にビッグデータを扱える人を育てましようって言っても全然、中身がない話になっていて、一步進んだ、どのような具体的なワードがあるかと思いました。

中村(講師)：多分、ビッグデータとは、早晚消えるというのは、当たり前になるから消えるのです。やめるからなくなるのではなく。当たり前になるから消えるのです。で、そのときには環境化すると思うのです。もう、Amazonが自動で、こう、何か届けてくれるような、環境になるかもしれません。私はそのときに、人間は、何をすべきか次のように考えています。推論とか、推論力とか、そういう機械になかなかできないことをする。人間がどう推論したかっていうのを脳科学的に取り出して、モデル化して埋めていく。

(了)

※尚、本稿は2015年4月4日開催のビジネスモデル実践研究会第16回研究会での講演内容を取り纏めたものです。