

ビッグデータを活用する確率モデリング技術

—社会実装の取り組みと課題—

本村 陽一[†]

(受付 2017 年 11 月 17 日；改訂 2018 年 3 月 15 日；採択 4 月 11 日)

要 旨

ビッグデータを活用した機械学習により現在、人工知能技術の実用化が劇的に進んでおり、それによる産業構造変革や Society5.0 と呼ばれるスマート社会の実現も期待されている。本稿では実社会のビッグデータとして利用者の ID が付いたサービス利用履歴データ (ID=POS データや ID 付きアンケート, ID 付き操作履歴など) から確率潜在意味解析 (PLSA), ベイジアンネットワークを用いて確率モデルを構築し, 利用者の行動や嗜好性を予測する確率モデリング技術について概説する。またそれによりサービスの価値や生産性向上の実現に寄与する人工知能技術としての応用例や社会実装を進めるための取り組みについても紹介する。

キーワード：サービス工学, 人工知能技術, 確率モデリング, ベイジアンネットワーク, 確率的潜在意味解析, ビッグデータ。

1. はじめに

物質的価値だけでなく生活の質 (QoL) や経験価値が重視される時代になり, 生活やサービス (コト) を直接の対象にして価値を生み出すことが期待されている。それにともない従来の技術開発の延長ではない, 新たなコトづくりのための技術開発が必要になっている。技術競争の激化が進み機能や品質の均質化が起これると, どの技術や製品を選んでも大差がないコモデティ化 (汎用化) が進む。製品は独自の付加価値を失うために製品選択の基準が価格にしかない状態となるために価格競争にさらされる。こうしたことから, これまで技術力を高めることで競争力を誇っていた日本の製造業が苦戦するようになってきたと言われている。製品の機能や技術自体では差別化することが難しくなると, 製品が使われる背景であるニーズや市場を知り, 市場に投入するタイミングをはかり, 製品の魅力を高める外装デザインや使い勝手など機能以外の付加価値も合わせて考えることに競争力の源泉がシフトする。技術が優位であっても, それが価値として十分に伝わらなければその技術は使われない。使われないので技術開発の優先順位が下がり, せっかく優れた技術であるのに開発が継続できず最悪の場合には技術が不当に安く放出されたり, 朽ち果てることすらある。競争力を高めるために全く新しいジャンルの製品開発によりイノベーションを起こすという考え方もあるが, 利用者のニーズやマーケットの動向やトレンド, 投入するタイミングを考えることなしに全く新しい製品の価値を理解してもらうことはさらに難しくなる。これまでに成功した新製品では技術それ自体の良さだけでなく, 利用者側に潜在的なニーズと適切なタイミング, 普及のための条件などが整っているところに投

[†] 産業技術総合研究所 人工知能研究センター：〒135-0064 東京都江東区青海 2-4-7

入されていたことも多い。つまり、利用者の側、ユーザーサイドの状況にももっと目を向けることが成功の確率を高めることにつながる。イノベーションとは技術革新のみではなく、革新的技術が受容されることを通じた社会革新であり、製造業がもっと利用者側の状況や潜在的ニーズを考慮して利用者にとって価値の高い製品を持続的に開発できるような構造変革を目指すことが重要になっている。

一方、国民総生産(GDP)や労働人口の7割近くがサービス産業に関連している。サービス産業の実態は幅広いが、顧客と直接接する、顧客接点の多いサービスでは、いわゆる「おもてなし」として日本のサービス全般に質が高いと言われている。しかし、一般にそのサービス価値が高いと言っても、それは定量的な指標で比較できるものになってはいない。顧客満足度という概念はあっても、それが定量化されていないためにサービスの質を正当な価格に反映することができず、わかる人にはわかるが大きな利益を生むことはできないということになる。品質のよいサービスを提供している従業員がその結果に対する貢献度を示すことが難しく、待遇の改善や仕事へのモチベーションを上げることが難しい問題がある。また品質の高いよいサービスを提供する技術があくまで個々の従業員の経験により支えられ再現性が低いために、大きな波及効果を上げることは難しい。これまでの製造業の成長を支えてきた産業技術が高品質な製品を大量に生産、提供するバリューチェーンとして再現性の高い仕組みを構築することで成長してきたように、サービス業でもレバレッジ(乗数効果)の大きな効果や生産性を上げるための仕組みづくりが重要である。サービスは提供されるものであると同時に、そのサービスを受け取る利用者側の反応でその品質が評価されるものである。つまりサービスでは本質的に利用者の側、ユーザーサイドの反応に目を向けることがとても重要になる。

こうした背景のもと、製品や製品の機能も含めた価値の提供であるサービスと、さらにその結果生じる利用者側に起こる反応も含めた価値共創の仕組みを明らかにし、それを工学的な枠組みとして確立するサービス工学の試みが行われてきた(吉川, 2008)。2008年経済産業省の研究拠点整備事業が開始されると同時に、産業技術総合研究所においてサービス工学研究センターが設立され(本村 他, 2008)、2012年にはサービス学会も誕生した。モノと違って、サービスは複数の人の相互作用(コト)として伝搬するので、モノの機能評価と違い、人の心理や行動、状況に基づく相互作用や経験を計算する方法が必要である。この価値共創の主体が人であることが、明示的な表現体系がすでに確立されてきたモノを対象に発展してきた工学の体系だけでは取り扱うことができない本質的な問題である。つまり、社会システム、つまり人の集団活動をいかに工学的、体系的に取り扱うことができるか、がサービス工学の大きな課題であり、そのためにはサービスのシステム観が重要になる(図1)。

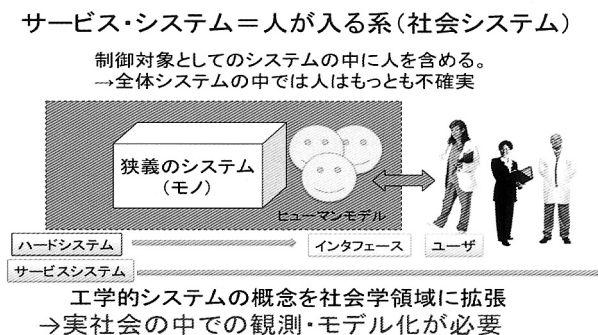


図1. 社会システムも含めたサービスのシステム観。

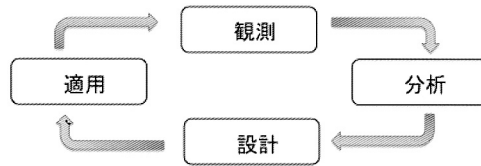


図 2. サービスの最適設計ループ。

サービスの特性はサービス提供と消費が同時に行われ、その品質はサービス利用者や状況に依存し、保存できないというもので、これらは同時性・異質性・消滅性というように整理されている。簡単に言えば、モノと違って、サービスは複数の人の相互作用(コト)として伝搬しており、人の心理や行動、状況に基づく相互作用や経験を計算する方法が必要というわけである。これは製品それ自体の価値ではなく、利用する時の価値についても成り立つ性質でもある。この価値の主体が人である、人起点であることが、これまでのモノを対象に発展してきた工学の体系だけでは取り扱うことができない本質的な問題である。サービスは無形であるが、サービス現場における人の相互作用の結果である行動履歴を大規模データとして客観的に観測することはできる。サービスの現場で起こる人の行動の結果を客観的に観測し、得られたデータを分析して得られる計算論的なモデルに基づいて、あるべきサービスを設計し、それを現場に適用するという「最適設計ループ」によって、価値を連続的に改良するアプローチを実践している(本村 他, 2008); (産業技術総合研究所, 2014) (図 2)。

このアプローチを実現するために、サービス中に生成されるデータを使って工学的に扱うことのできる計算モデルを活用してサービスを設計する方法がある。計算機の中で表し、実際に変数を計算したりすることで、あたかも実在するシステムを近似できるようなものを計算モデルという。この計算モデルによって、サービスシステムをとりあえず近似して説明できれば、その動きを模倣(シミュレート)することができる。すると、その動き方をいろいろと変えてみることで、より良い改善を考える。サービスの結果として起こる現象を説明する背後の関係を全部書き下すことは難しい。また、全ての人について成立する決定的なモデルを天下一的に見つけることも難しい。そこで、その日の行動を(例えば天気や曜日、利用者のライフスタイルなどから)確率的に予測し、その予測精度が実用上十分であるような不確実性を考慮した計算モデルを実際のデータから構築することで、サービスの現象を実用十分な範囲で近似できるモデルとして活用する。データからこうした計算モデルを構築する技術は機械学習(Machine Learning)と呼ばれる人工知能技術である。これらの技術はここ十数年の間、インターネットのデータについて使われ大きな成功を収め、今や日常の中でなくてはならないものになりつつある amazon.com や Google などの実際の IT サービスを支える基盤技術であり、さらに様々なモノがネットで接続される IoT(Internet of Things)時代においても注目されている技術である。現実の空間でのサービスの場合にはインターネットと違って機械学習やデータマイニングのために大量のデータを効率よく集める仕組みがまだ十分整っていない。しかし、電子マネーにより買い物をしたり、メンバーズカードに共通ポイントを貯めたりする利用者が増えており、さらに、今後行動履歴などの大量のデータが簡単に収集できるようになることが予想され、実社会物理空間(フィジカル)の活動と計算モデルの空間(サイバー)の活動が統合したサイバーフィジカルシステムと呼ばれるものとして社会活動をとらえる時期が来ている。サービス現場で得られるビッグデータを活用して実際の店舗やレストラン、病院などで起きている実際のサービス現象の計算モデルを作り、それを利用した支援技術を現場に提供できれば、実際のサービスの改善に役立つ。同様に製品の使われ方や利用した時の利用者が感じる価値を定量的に評価す

るモデルをデータから構築して、設計や製品の制御に活用すれば、それは製品開発や製品の制御にも役立つ仕組みでもある。本稿ではサービス現場で生成されるビッグデータとして利用者のIDをひもづけた購買履歴データ(ID-POSデータ)や共通ポイントデータ、行動履歴データなどを活用して確率的モデルを構築する事例を紹介し、これが人工知能技術として社会実装されると、産業構造変革にどのような役割を果たすかを議論する。

2. ビッグデータを活用する人工知能技術：確率モデリング

人の行動は毎回同じとは限らず決まった通りに動くものではない。したがって、不確実性を考慮したモデルが必須になる。また実生活場面の現象を説明するモデル化においては記述量・計算量の点から、扱う対象自体を完全に記述することは無理であるので、現象を確率的・統計的なものとして扱うことにする。人の行動が起こる確率を考えて、その行動が起こる条件として典型的な(相互情報量が高い)状況を見つけると、条件付確率 $P(\text{行動} | \text{状況})$ という形で不確実性を含めて表すことができる。さらに人のタイプごとにとる行動が異なる場合には、さらにこれを条件部に加えて $P(\text{行動} | \text{状況}, \text{人のタイプ})$ とすればよい。この人のタイプは利用者の「異質性」とも呼ばれる。この人の異質性はサービスにおける基本的な特性であり、これをいかに取り扱うかはサービス工学における重要な課題である。この条件付確率の条件部に入る変数を加えていくことで、来店行動やある商品を買う購買行動の確率を精度良く予測できれば、あるタイプの顧客がお店に来る人数やその顧客が買いそうな商品の数も推定できるので、適切な人員配置や商品の準備をすることで人員不足や品切れを防ぐことができる。つまりサービスの最適化がはかれる。また、日常業務の中でその日の状況や顧客のデータをさらに大量に持続的に集め、確率モデルも新たなデータによって更新することができれば、予測精度をさらに高めることができる。これを実現するために条件付確率のモデルをデータから構築する技術がベイジアンネットワークである(本村・岩崎, 2006)。

利用者の異質性については、行動が似ている人を集めてセグメントを作ることが行動の予測を行う場合には適している。実サービス中に集積されている購買行動のデータであれば、行動は購入した製品のIDとして見分けられる。会員カードなどの顧客IDと製品IDが記録されたID-POSデータに対して確率的潜在意味解析(PLSA)を活用して利用者の異質性を潜在クラスとして発見する事例(本村 他, 2012, 石垣 他, 2010; 2011a, Ishigaki et al., 2010, 石垣 他, 2011b)がある。大規模な数年分のID-POSデータを使って、数千人の顧客を比較的少数のセグメントに分類し、セグメントごとに商品選択確率や来店行動などの予測の精度が向上することなどが確認されている。個人のID付データの場合、個人の行動やプライバシーの保護が問題になるのに対して、適切なセグメントにより異質性を表す方法は情報量を失わずにプライバシーも保護できる(山下・本村, 2014)。こうした適切なセグメントを探索するアルゴリズムとして確率的潜在意味解析が応用できる。

3. 確率的潜在意味解析

確率的潜在意味解析(Probabilistic Latent Semantic Analysis: PLSA)は、顧客毎の商品の買いかた、商品から見ると買われ方が似ているものを同一クラスに併合する操作を繰り返すことによってID付きPOSデータから顧客と商品を同時分類しカテゴリ生成を行う方法である。具体的にはまず、顧客ごとに購入した商品の数を集計した共起行列を作成する。また、顧客と商品が所属する潜在クラスの数を決める。その上で、顧客 x_i 、商品 y_j 、潜在クラス z_k の関係を次式、

$$(3.1) \quad P(x_i, y_j) = \sum_k P(x_i|z_k)P(y_j|z_k)P(z_k)$$

としてモデル化し、次式の数尤度

$$(3.2) \quad L = \sum_i \sum_j n(i, j) \log P(x_i, y_j)$$

を最大化する方法が確率的潜在意味解析である。

本来は顧客ごとに購入した商品の数をベクトルとして考えて、類似した顧客をセグメント化したいが、商品の種類が数千点以上の規模である場合、それらの商品全てを購入することはほとんどない。したがってたいの顧客はそのうちほとんどの商品の購買数がゼロとなり、顧客毎の商品の数をベクトルにした類似度計算がうまくいかない。そこで、次元を圧縮するために潜在クラスを導入して、適当な初期値で割り当てた商品分類を顧客の特徴ベクトルとして顧客を分類する。次に今度は商品の類似性を求めるために、先に得られた顧客分類を特徴ベクトルとして使って、新たに商品分類を更新する。この操作を繰り返すことで対数尤度は増大し、やがて収束する。最適なカテゴリの数は事前に決めることは難しいため、情報量規準(AIC)に基づいて最適なクラスタ数を探索する。このアルゴリズムによって、数千から数万の膨大な数の顧客と商品を比較的少数の潜在クラスへ分類する確率ベクトル $P(x|z)$, $P(y|z)$ が得られる。

ここで得られたセグメントは商品選択を対象に情報量が大きくなるように顧客を分割する場合の解である。本質的にどのように分類することが望ましいかどうかはそのセグメントをどのような目的に用いるかに依存する。商品の購入を促すためのマーケティング施策として、提供する商品や商品情報を選ぶ商品選択に対する情報量が大きいセグメントに顧客を分類することが最適である。一方、ある一日や場所ごとに実施する施策を最適化したいのであれば、商品の代わりに、時間や場所を対象にしたPLSAを実行して得られた各セグメントに対して最適な施策を決定すべきであろう。このようにしてPLSAの結果得られるセグメントは顧客セグメントというだけでなく、商品や時間や場所のセグメントとして活用することもできる。このPLSAのセグメントがどういう意味を持つのかを説明するために、ベイジアンネットワークを用いてセグメント間の構造をモデル化する(本村, 2016b)。これまでサービス現場での実証研究(吉田・本村, 2013)や応用、企業による実用化も多数行われている。

4. ベイジアンネットワーク

日常生活中における現象をモデル化するためには人の行動や心理などからくる様々な不確実性に対処することが必要である。そのため確率モデルを使って対象をモデル化することで、知りたい変数の確率分布を推定し、おこりえる各状態の確率(確信度)を評価する枠組みが有効である。複数の確率変数の間の定性的な依存関係をグラフ構造によって表し、個々の変数の間の定量的な関係を条件付確率表で表したモデル、ベイジアンネットワークがある(本村・岩崎, 2006)。

ベイジアンネットワークの中の一つの子ノードに注目した依存関係、つまり一つの目的変数(従属変数: Y)と、それに対する説明変数(独立変数: X)の間の依存関係について着目するとX-Y空間を条件付確率表にしたがって量子化し、個々の確率値を割り当てたものになっており、これにより任意の非線形性、非正規性を表現できる。また複数の親ノードによる交互作用を表せること、つまり複数の親を持つ場合、非線形を持つ交互作用も表すことができることが現実の社会で起こる多様な事象をモデル化するために有用な特長になっている。日常生活場面では個人差、状況依存性などを反映する必要がある、この点で交互作用や非線形性、非正規性

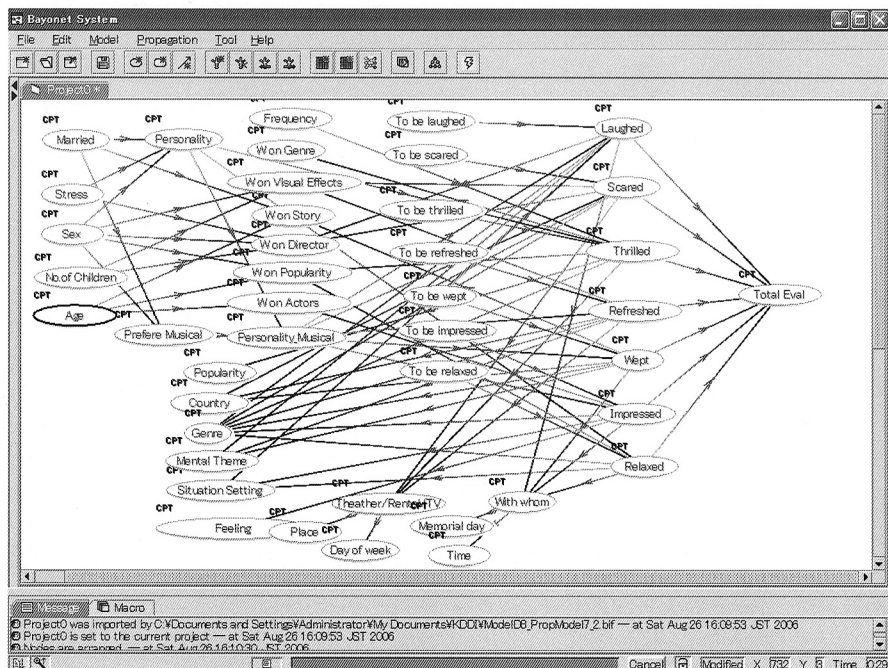


図 3. 映画推薦のためのベイジアンネット。

を含むモデル化が果たす役割が非常に大きい。また構築したベイジアンネットワークの一部の変数に観測値を代入して確率伝搬と呼ばれる計算を実行することで、それ以外の変数の確率分布が高速に求められるアルゴリズムが存在することも、様々な情報サービスに応用する上で重要な特長である。

ベイジアンネットワークと、その上で確率推論を実行することで、利用者の選択行動などを推定することができる。そこでユーザーが選択する可能性の高いコンテンツを推定することで情報推薦に応用できる。とくに新たに追加されたコンテンツにも対応できるようにコンテンツ属性を変数として用い、さらにユーザ属性や状況を表す変数もベイジアンネットワークのノードとしてモデルに組み込むことで、状況やユーザの傾向に応じた推薦が可能になる。筆者らとKDDI研究所のグループによる、携帯電話サービスのためにベイジアンネットワークを用いた映画コンテンツを推薦する事例(本村・岩崎, 2006)では約 1600 名の被験者に対して映画コンテンツを提示するアンケート調査により収集したユーザ属性、コンテンツ属性、コンテンツ評価履歴からベイジアンネットワークモデルを構築した(図 3)。

このモデルを用いて状況とユーザの嗜好性に応じて映画を推薦する携帯情報システムのプロトタイプを開発した。システムはデータベースから登録済みのユーザ属性情報と状況情報を使って確率推論を実行し、その結果選択される確率が高いと判断されたコンテンツを上位から推薦する。こうしてユーザの状況と嗜好性に応じて映画を推薦するサービスを実際に運用することで、持続的にデータの収集とモデルの改善が可能になる。こうして得られる各種の大規模データからベイジアンネットを構築すれば、人間の生活行動を推定する再利用可能な計算モデルとして様々な情報サービスにも活用できる(図 4)。

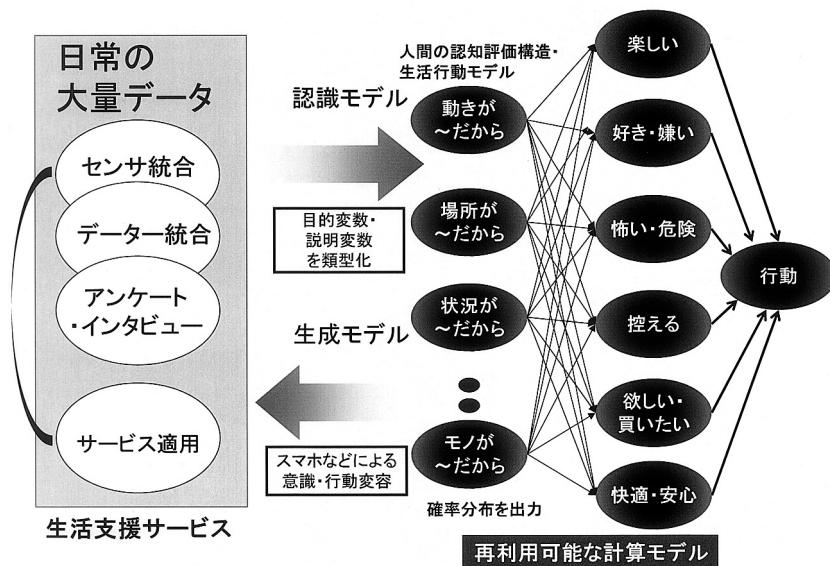


図 4. ビッグデータから構築される生活行動モデル.

5. 確率潜在意味構造モデリングと利用者のモデル化手法としての応用

確率的の潜在意味解析とベイジアンネットを組み合わせることで、ID-POS データや共通ポイントカードの使用履歴データなどのサービス現場で大量に集積されているビッグデータから顧客の異質性を表し購買行動や嗜好性、アンケート回答の推定を行うことができる確率モデル、利用者モデルが構築できる(図 5)。

この利用者モデルを使って、顧客それぞれに対して対応を個別に最適化することや、ある時間やエリア、ある商品に対して主たる利用者を推定して利用者の集団の特性を推定することでサービスを最適化する方法などが実現されはじめています。前者は会員カードやスマートホンなどと連携したレコメンド(情報推薦)やナビゲーション、後者は利用者セグメントごとの施策やサービスの最適化という形で実行されることになる(図 6)。またこれは従来、マーケティングやマネジメントとして実行されている業務のインテリジェント化としても位置付けることもでき、実際の産業応用としてはサービス分野の IT 化、あるいは人工知能技術の普及を進めるものとしても考えられる。今後、「サービスのシステム化」「社会基盤技術のインテリジェント

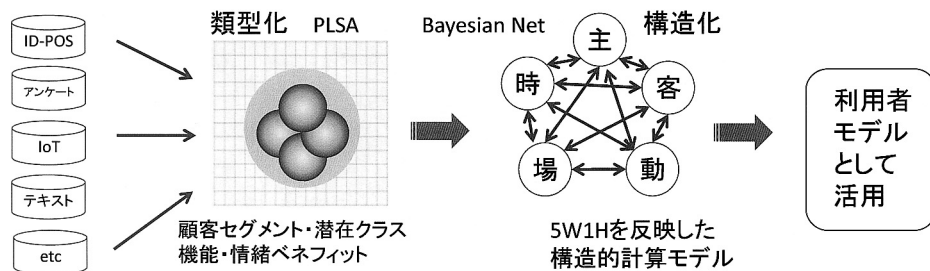


図 5. 確率的潜在意味解析(PLSA)とベイジアンネット.

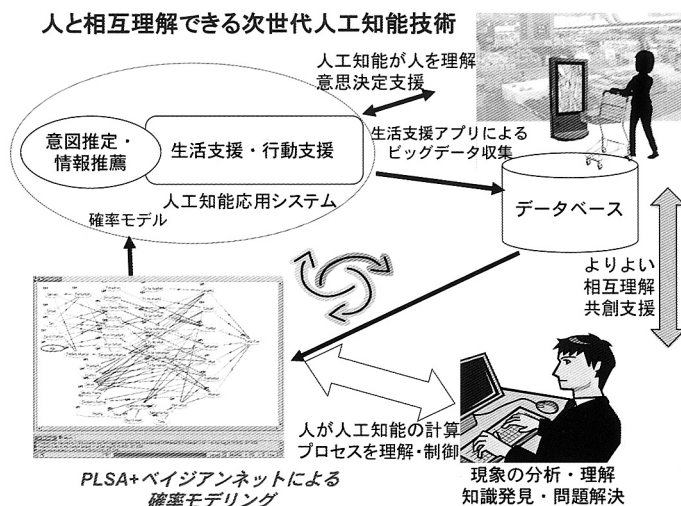


図 6. 人と相互理解できる AI による生活支援技術.

化」というものを考えた時、ビッグデータを計算モデル化して利用者の行動を確率的に予測することのできる人工知能技術をさらに多様なデータと組みあわせることによって、付加価値が高く、効率よく実行できる新たなサービスが生まれることになるだろう。

6. 産業構造変革のための人工知能技術としての役割

本稿で述べた確率的潜在意味解析(PLSA)とベイジアンネットワークを用いて確率モデルを自動的に構築する技術を情報サービスにも適用することが可能なソフトウェア(PLASMA)として実装した。このソフトウェアを使ったマーケティングサービスや、データを複数の企業が共有、連携して活用する新規事業などの立ち上げも開始され、こうした活動を支援するために後で述べる「人工知能技術コンソーシアム」が産総研内に設置され、本稿で述べた人工知能技術を活用することで、飛躍的な生産性の向上と、既存の価値創出の拡大を第一段階とし、さらにデータと計算モデルを共有することで、有機的に連携できる企業ネットワークを第二段階とした産業構造変革の実現に向けて活動している。

機械学習に基づく人工知能技術を社会の中で活用するためには、持続的なデータ収集を行う仕組みが不可欠であり、AI技術の研究開発はウォーターフォール型よりは、スパイラル型の研究推進、つまり生活やサービスの現場でビッグデータの収集と技術活用を併行して行う必要がある。そのため、産業応用プロジェクトの中では社会的なニーズの高い問題設定と、それに関与する多くのステークホルダーとの連携が不可欠になる。インターネットの発展とともに成長した巨大IT産業が、自身のサービスやビジネスを展開しながら最新の人工知能技術の研究開発を進めているのと同様、次世代人工知能技術の研究開発においても、現実的な場面における社会実装と技術検証、つまりユーザにとっての有用性や安全性、信頼性を初期の段階で示しながら性能を高度化するという方法論(growth hack)が有効であると考えられる(本村, 2016a)。そのため、単に技術を研究開発するだけでなく、実際にどのようなデータから機械学習が実行され、社会の中でどんなサービスを実現するかを研究開発と合わせて考える工夫も重要である(本村, 2015)。

Internet of Things (IoT) が爆発的に進み、実空間における様々な現象がビッグデータとして記

録され、それらが計算機空間でモデル化、シミュレーション可能になる Cyber Physical System (CPS) の概念がある。IoT デバイスの普及とそこから生成されるビッグデータを活用することで、社会の現象を計算モデルとして構築することで新たな現象が計算可能になる。本稿で述べた確率的潜在意味解析 (PLSA) とベイジアンネットを統合した手法はビッグデータを扱う統計手法であると同時に、これが社会実装されると人工知能技術として次のような役割を果たす。実空間の活動が IoT デバイスにより時間、空間情報とともにデータ化され、これが PLSA により情報量の高い潜在クラスに確率的に分類、すなわち確率ベクトルにコード化される。簡単のために所属確率が高いベクトルを 1 に、それ以外を 0 のように二値化されると考えると、実空間の現象の統計的な共起関係が PLSA の尤度極大化により、できるだけ情報量が高くなるようにデジタル化したことに相当する。さらにこのデジタル化された現象間の関係がベイジアンネットによって構造化されることで、時間や場所、人の特性、行動、対象物などがいわゆる「いつ」「どこで」「誰が」「何を」「どうした」という形で関係性や相互作用を確率的に表現できるように計算モデル化される。この計算モデルの入出力が実社会のサービスとして実行され、新たな実空間の現象を生じさせるならば、これは、計算モデルの空間(サイバー)上で、実空間(フィジカル)の現象が密に連携したサイバーフィジカルシステムの一つの実装方法となっている。

サイバー空間で最適化された計算結果をスマホのアプリやサービスを通じて人々に提供し、意思決定や行動を支援することで、良い現象の発生確率を上げ、事故などの良くない現象の発生確率を下げるという意味での物理世界の制御、マネジメントが可能になる。こうした IoT デバイスと AI 技術により構成される Cyber Physical System (CPS) を活用することで、産業構造変革を進め、生産性向上や付加価値の向上に寄与することが期待され、実際に 2018 年度から多数の政府系プロジェクトや企業活動が開始している。人工知能技術を活用し実社会の産業構造変革に貢献するためには、現実の社会構造や生活と乖離することなく、人々にとって扱いやすい形でその技術とサービスが提供され、制度や文化の進化とも歩調を合わせて社会実装を円滑に進めていく必要がある。

7. おわりに

機械学習やビッグデータを活用する人工知能技術により、サイバーフィジカルシステムを構築することで複雑な社会問題の解決も期待されている。ただし、計算モデルが高度で複雑なものになるにつれ、学習のために必要なデータ量が増大する。表層的に観測可能なセンサデータなどは比較的容易に取得できるが、人間行動の内部的状態は心理的なものであるため、被験者を用いたアンケート調査も必須になりコストが大きい。またデータを取得する上で、プライバシーの問題や、単に研究目的のためには協力が得られにくいという現実的な問題もある。またたとえ外部的な要因で観測容易な事象だとしても、実際に使う場面において、状況依存性の高い説明変数を網羅的に収集するためには、データを観測する環境が日常的な利用環境とできるだけ合致するように統制しておく必要がある。

そこで、こうした問題に対して実サービスと調査・研究を一体化すべきであるとする「サービスとしての調査・研究 (Research as a service)」という概念が提唱されている (本村, 2009)。調査・モデル化の段階とそのモデルを用いた応用を切り離すことなく、情報サービスを社会の中で実行しながら、そこで得られる観測や評価アンケート、利用者のフィードバック (心理的調査) の結果を網羅的に収集する。これは古くはサイバネティクス、また信頼性工学ではデミングサイクルとして知られる PDCA (Plan, Do, Check, Action) サイクルを、実問題を通じて回し続けることで、モデルを常に修正していくというものである。

不確実性に対する本質的な解決のためには対象を実データによりモデル化し、そのモデルを

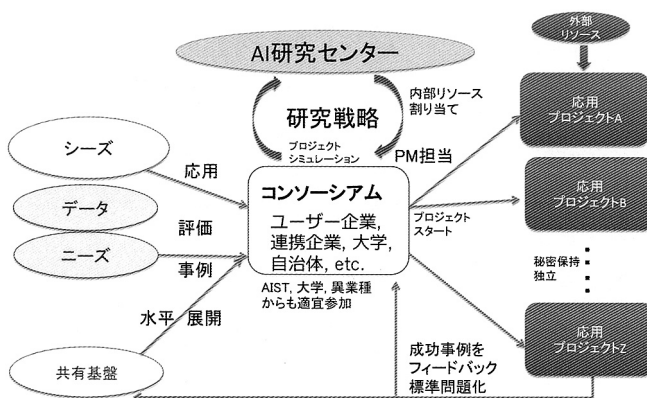


図 7. 産総研人工知能技術コンソーシアム。

用いて制御しながらさらにデータを収集する，というサイクルを永続的に続けるアプローチが重要になる．大規模データと人工知能技術を活用した情報システムを導入した実サービスの開発と応用を通じて，多様な生活者の特性を計算モデル化し，有用な知識モジュールとして社会全体で活用できるビッグデータ活用技術の仕組みを確立することが期待できるようになっている．こうした技術を産業基盤として社会の実サービスの中で大規模データの収集とモデルの構築，活用を持続することでさらに多くの情報サービスの実現が容易になる．こうしたビッグデータと人工知能技術により産業生産性を向上し，生活の質も向上することを目指し異業種連携を加速するための場づくりとして，産総研内に「人工知能技術コンソーシアム」が設立され，2018年度180社以上の企業が10を越えるワーキンググループ，関西，東海，九州の地域支部により活動し，ユーザ参加型の社会実装，技術評価を進めている(図7)。

共同で行動を行うコミュニティは実践コミュニティ(Community of practice)と呼ばれる．この実践コミュニティの形成や運用を人工知能技術によって支援し，誰もが自由に入出入りして参加したりできるような「場」ができることによって，柔軟にサービスシステムが共創的に価値を創出できるようになることも期待できる．こうした価値共創を持続的に再現，発展できる仕組みによって社会や暮らしを今よりもよいものにするという方向性を強く打ち出すことが，人工知能技術により新たな産業や社会構造を変革するための社会実装のために有効であると考えられる．

参 考 文 献

- 石垣司, 竹中毅, 本村陽一 (2010). 確率的潜在意味解析を用いた大規模 ID-POS と顧客アンケートの統合利用による顧客-商品の同時カテゴリ分類, 電子情報通信学会技術研究報告. NC, ニューロコンピューティング, **109**(461), 425-430.
- Ishigaki, T., Takenaka, T. and Motomura, Y. (2010). Category mining by heterogeneous data fusion using PdLSI model in a retail service, *Proceeding on IEEE International Conference on Data Mining (ICDM)*, 857-862.
- 石垣司, 竹中毅, 本村陽一 (2011a). 日常購買行動に関する大規模データの融合による顧客行動予測システム: 実サービス支援のためのカテゴリマイニング技術, 人工知能学会論文誌, **26**(6), 670-681.
- 石垣司, 竹中毅, 本村陽一 (2011b). 百貨店 ID 付き POS データからのカテゴリ別状況依存的変数間関係

- の自動抽出法, オペレーションズ・リサーチ, **56**(2), 77-83.
- 本村陽一 (2009). 大規模データからの日常生活行動予測モデリング, シンセシオロジー, **2**(1), 1-11.
- 本村陽一 (2015). サービス工学におけるビッグデータ活用技術, 日本機械学会誌, **118**(1163), 628-631.
- 本村陽一 (2016a). 次世代人工知能技術, 情報処理学会誌, **57**(5), 466-469.
- 本村陽一 (2016b). 第9章『ベイジアンネットワークと確率的潜在意味解析による確率的行動モデリング』, 『確率的グラフィカルモデル』(鈴木讓 他), 共立出版, 東京.
- 本村陽一, 岩崎弘利 (2006). 『ベイジアンネットワーク技術～ユーザ・顧客のモデル化と不確実性推論』, 東京電機大学出版局, 東京.
- 本村陽一, 西田佳史, 持丸正明, 赤松幹之, 内藤耕, 橋田浩一 (2008). サービスイノベーションのための大規模データの観測・モデリング・サービス設計・適用のループ, 人工知能学会誌, **23**(6), 736-742.
- 本村陽一, 竹中毅, 石垣司 (2012). 『サービス工学の技術～ビッグデータの活用と実践～』, 東京電機大学出版局, 東京.
- 産業技術総合研究所 (2014). 『社会の中で社会のためのサービス工学～モノ・コト・ヒトづくりのための研究最前線』, カナリア書房, 東京.
- 山下真一郎, 本村陽一 (2014). 確率的潜在意味解析を用いた集団匿名化法における来店店舗予測精度の評価, 人工知能学会全国大会 1L2-OS-17a-3.
- 吉川弘之 (2008). サービス工学序説—サービスを理論的に取り扱うための枠組み—, シンセシオロジー, **1**(2), 111-122.
- 吉田真, 本村陽一 (2013). ベイジアンネットワークによるセグメント説明モデルと映画推薦への応用, 第三回 Web インテリジェンスとインタラクション研究会.

Probabilistic Modeling Technology Using Big Data: Activity for Social Implementation

Yoichi Motomura

Artificial Intelligence Research Center, National Institute of Advance Industrial Science and Technology

Currently, the practical application of artificial intelligence is being dramatically advanced by machine learning using big data. These efforts are also expected to help realize industrial structural reform and the smart society (“Society 5.0”). In this paper, we introduce probabilistic modeling using probabilistic latent semantic analysis and Bayesian networks. To realize the value of service and improvement in productivity, the user’s behavior and preference are predicted by probabilistic models constructed from service history data (ID-POS data, questionnaire with ID, operation history with ID). Examples of real applications and efforts at social implementation are also discussed.