

AI 技術によるビッグデータ活用の課題

Issues of Big Data Utilization using Artificial Intelligence Technology

本村 陽一*¹

Yoichi Motomura

*¹ 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology

Issues of Big data utilization using AI technology are discussed in this paper. We have several difficulties to utilize Big data in the real society. In order to overcome such difficulties, novel artificial intelligence technology should be developed and it will be implemented into the real society by service providers. This approach will be a kind of an action research and new framework to proceed AI research is also required.

1. はじめに

いわゆるビッグデータはインターネット上のデータにとどまらず、企業内に集積される業務データや、各種のセンサにより記録される人の行動記録(ライフログ)や ID 付のカードや端末の利用履歴(パーソナルデータ)、電子マネーや共通ポイントカード履歴、交通系 IC カード利用履歴など多岐に渡る種類のデータとして現実に利用可能なものになっている。しかし個人の ID のついたデータはパーソナルデータとしてプライバシーへの配慮が必要であり、そのデータの取り扱いのリスクによって、データを活用するメリットがある程度理解されながらも、現実には活用が進んでいない場面も多く存在する。また大量に蓄積されたデータを有効に活用する技術が十分でないことから、活用の効果はデータを可視化することとどまり、その結果を解釈、実行する人のスキルに依存することから、データ活用人材の不足や育成の問題がボトルネックとなる事態も顕在化している。本発表では、こうした実社会のビッグデータ活用の課題について整理し、この課題を解決するために必要となる AI 技術への展望を述べる。

2. 実社会のビッグデータ活用

生活空間での現象を大量のデータ、いわゆるビッグデータとして観測し、集積できるようになった。機械学習や確率モデルなどの人工知能技術によって、このビッグデータを活用した計算モデルや支援システムを構築することで、生活空間内の現象やそのプロセス、ダイナミクスを直接の対象として課題を解決する工学的な手法を提供することが期待できる。

例えば産総研サービス工学研究センターでは生活中に生成される大量、大規模なデータ、いわゆるビッグデータを収集し、そのデータを通じて、生活空間中の現象がどのように生じるかを確率的な構造モデルとして構築し、そのモデルを活用して実際の現象を改善するための方法を推定し、その推定結果にもとづいて現場の行動を変化させる実証的研究が複数の具体的事例を通じて行ってきた[1]。そこではビッグデータを活用するための共通の枠組みとして、現象のモデル化手法が確立されてきている。生活空間中の現象を対象化して、計算可能なモデルとすることで、生活現場の現象が形式化、操作可能なものとなる。この生活中に起こる現象のモデル化による、生活現象の最適化は、これまでパターン認識やクイズのような問題の解決を可能にして

きた人工知能にとつての次の挑戦課題かも知れない。これまでの人工知能技術、機械学習においては問題となるドメインの大局的な知識やインターネット上の静的な(定常)データを対象にすることが多かったが、生活現場の詳細なビッグデータが活用できるようになれば、ある特定の生活地域や時間に局在する知識(ローカルナレッジ)を対象とすることができるようになり、こうしたローカルナレッジの活用はこれまでの人工知能技術が対象にすることが難しかった生活現場の支援技術を開発可能にすると期待できる。

3. 実社会ビッグデータ活用の課題

経済産業省の調査事業[2,3]などにおいて、地域活性化やサービス産業の生産性向上のためのビッグデータ活用の検討が進められてきた。実際の地域のニーズや企業間の協力にもとづいていくつかのシナリオや事例が具体的に検討されている。そこで明らかになってきたビッグデータ活用の課題としては次のようなものがある。

3.1 複数機関の間でのデータ活用の必要性と困難性

各事業者のサービス現場で収集される購買履歴や消費者の行動履歴などは社会の生活実態を忠実に表す情報である。現在はこうしたデータの活用は企業内のみで活用されていることが多く、またデータを活用できる人材不足などのため、その活用範囲は限定的なものにとどまっている。今後さらにデータの活用を進めることで地域の課題解決、例えば食の実態から地域の健康管理に発展させて地域における医療費の削減や行政コストの削減、効果向上などに資することも期待されているが、そのためには行政を含む複数機関がそれぞれ持つデータを効果的に活用することが求められる。しかし、現在はパーソナルデータを組織外に出すことが難しく、また複数機関が実ビジネスでは競合する場面もあることから、データを共有し、新たな目的に活用し、二次利用することには大きな障壁がある。

複数機関の間でのデータ活用の困難を打破するためには、複数機関や、利用者である市民にとっても共通のベネフィット、社会共有価値の創出が不可欠である。この社会共有価値が大きければ、それを目的にするために妥当な範囲の中で、コストやリスクの許容レベルが設定できる。逆に明確な社会共有価値の創出が目的として設定できなければ、ビッグデータ活用を協働で進めることは、機関のコスト負担や利用者のプライバシーリスクを払拭することができないことから非常に困難である。

連絡先: 本村陽一, 産業技術総合研究所, 東京都青海2-3-26, 03-3599-8355, y.motomura@ais.tog.jp

3.2 データ共有プラットフォームの構築

大きな社会共有価値を創出するためであれば、データの価値を共有するためのプラットフォームにより、複数機関が協働してデータの活用を進めることが期待できる。そのようなデータ共有プラットフォームに必要なことは、複数機関からのデータの収集と安全な管理、データを活用できるステークホルダーへの情報提供である。ここで問題となるのが、各機関が収集したデータがパーソナルデータである時で、その場合には、その取扱いが個人情報保護法にもとづいて適切に管理されなければならない。現時点ではパーソナルデータそのものではなく、k-匿名化と集計を行いパーソナルデータではない二次形式のデータや計算モデルに変換してから情報共有する方法が有望であると考えられる。計算モデルとして活用する技術は、データ活用の自動化を進め、データ活用人材の不足を解決できる可能性がある。元のデータを変換して情報損失をできるだけ抑えるには、情報量(尤度)に考慮して k-匿名化、集計を行う機能や方法を実装したデータ共有プラットフォームの構築が今後重要になる。

3.3 データ共有プラットフォームの運用

データ共有プラットフォームの技術だけではなく、それをどのように実社会の中で運用するべきか、についてもあらかじめ検討し、社会実装の方法を明らかにする必要がある。パーソナルデータを k 匿名化、集計した二次データについてはデータを生成した本人の同意は不要となるため、当初想定していなかった目的に対して新たな同意をとることなく、データが活用できる可能性が生まれる。しかし、複数機関が協働し、社会共有価値創出を目的とした地域や社会全体の知としてのデータ共有プラットフォームをどのように運用していくかについては、合意や意思決定の仕方、新たな仕組みが必要となると思われる。

4. 人工知能技術としての課題

これまでビッグデータを活用した事例としては、インターネットの自動翻訳や自動スマートホンで使われる音声認識や車の運転を支援する画像認識など、インターネットやセンサから各種のデータを集め、そのデータを用いて、高度な処理を行うコンピュータプログラム(人工知能)を学習させる機械学習と呼ばれる方法が圧倒的な成功を収めている。そこに共通するのは、観測されたデータを特定の正解パターンに分類し、認識するモデルを構築し活用する点である。しかし、この認識モデルを構築する方法では、現象を識別に決まったパターンに分類することはできても、現象を望ましい方向へ変化させることができない。そこで必要となるのが、「制御モデル」である。制御モデルというのは、まず現象を再現できるだけの構成要素と因果構造からなる生成モデルを構築し、さらにその生成モデルの中で後から我々が自由に変化させることのできる説明変数(制御変数)と、その制御変数によって影響を受けて変化させる目的変数(被制御変数)を区別したものである。生活中的現象をモデル化するためには必然的に人の行動がモデルの中に入り込む。しかし、社会の中の人の行動は必然的に不確実性があり、また人の行動をどのように計算論的にモデル化するかについてはまだ方法が十分に確立されているとは言えない。生活機能分類(ICF)にもとづく確率的なモデル化手法などは一般的な生活行動のモデル化のための一つの有望なアプローチであると考えられる[4]。

行動や環境、状況に関するデータの中から、ある行動 b が発現した背景となる説明変数としては個人要因 p 、や環境要因 e の中から影響が強いものを機械学習によって自動的に探索して、 $P(b|p, e)$ として行動を条件付き確率としてモデル化できる。状

況を表す変数群 e の間に時間や場所、目的などを含めた多段の因果的な構造をとることでコンテキスト依存性が表現できる。このようにして、人間の生活行動に関する因果構造「このような状況の時にはこの行動をとる」という生活機能の関係を多段の条件付き確率によって構造化し、その行動が何に起因して起こりやすいのかを理解し、その行動が起こる確率を予測することができる。このように表される条件付き確率分布は実世界で発生する現状($As\ is$)のモデルとなる。さらにモデルの中に制御変数も加えることで、望ましい確率が最大になる場合はどうすれば良いか、という理想的な状態($To\ Be$)を推定するような方法を明らかにすることも考えられる。こうした実社会の現象を人の行動(生活機能)を中心としてモデル化、それを活用して理想的な状態を作り出す支援技術を開発することは、ビッグデータを活用する人工知能技術の重要な課題となるだろう。またこうした支援技術を様々な機関が実社会で提供できることは生活を中心とした新たな産業やサービスの創出にもつながると期待できる。

5. おわりに

実社会の中で生活に関連する継続的に大量のデータが収集できれば、生活機能を確率的にモデル化する技術によって行動が起こる確率を推定したり、望ましい状態が起こりやすくするような制御をする機能を実現できる。しかし、そのためには集まるデータを活用することで生活者や社会にとって十分に価値の高い社会共有価値が得られる必要がある。つまり持続的にデータを収集し続けるためには意義の高いデータの活用方法を考え、社会的にも受容される技術やサービスを持続的に提供できる仕組みを動かすことが重要である。これは研究というよりはむしろ、成果の実用化や実践のようなものとして行われることになる。つまり本稿で述べたような実社会のデータを活用する人工知能技術の研究は純粋な基礎研究や理想的な実験環境での観測や解釈にとどまってしまうことが難しく、研究者自らが社会や生活の中に入り込んで、望ましい変化を作り出すことを同時に行うことが必要になる。こうした研究方法はアクション・リサーチと呼ばれる。この時、生活者と研究者の双方の立場で日常生活の中で制御モデルを表現し、意味を共有する変数としての共通言語が必要である。また制御モデルを使って現象を再現したり改善するためには、研究者も社会問題の解決のための活動に関与し、行政や企業などの機関とも協働することも必要ことが多い。こうした社会問題の解決にも踏み込んで研究者が継続して活動するためには新たな研究の進め方や社会的な仕組みも必要になってくるだろう。人工知能研究のあり方についての検討と実践(社会実験)は今後の重要な課題である。

参考文献

- [1] 産業技術総合研究所: 社会の中で社会のためのサービス工学, カナリア書房, 2014.
- [2] 経済産業省, 平成 25 年度我が国経済社会の情報化・サービス化に係る基盤整備(地域ビジネス振興支援に資するデータプラットフォーム構築とベンチマーキング手法開発に関する調査研究事業) 成果報告書, http://www.meti.go.jp/meti_lib/report/2014fy/E004195.pdf, 2014.
- [3] 経済産業省: 平成 26 年度 我が国経済社会の情報化・サービス化に係る基盤整備「パーソナルデータの地域での活用に関する調査研究事業」報告書, 2015.
- [4] 本村陽一: データに基づく生活機能構造の理解と分析—大規模データ活用による日常へのアプローチ—, 情報処理, vol.54, no.8, pp.787-790., 2013.