人類と人工知能　10 月 3 日個人課題

学籍番号：2322003
氏名：阿部　佑哉

ビッグデータとは、人間では全体を把握することが困難な巨大なデータ群のことを指す言葉だ。例として挙げられるのはソーシャルメディアデータやマルチメディアデータ、ウェブサイトデータなど様々なものがあり、活用範囲が幅広いため、あらゆる分野・業界でビッグデータの利活用が期待されている。そんなビッグデータだが、どうであればビッグデータとするといった明確な定義は存在しない。そのためビッグデータのエッセンスとして、一般的には Volume(量)、Variety(多様性)、Velocity(速度あるいは頻度)の「3 つの V」を高いレベルで備えていることが特徴とされている。また近年では、これに Veracity(正確性)と Value(価値)を加えた「5 つの V」をビッグデータの特徴とするとも言われている。

そんなビッグデータだが、問題点が 3 つある。1 つ目は、保守管理と運用の負荷が大きいことだ。人間には扱いきれないほどの膨大な量のデータは、保存するためのストレージの必要になるほか、データの選定はクレンジング(前処理)の負荷も増大する。当然、それらの作業をおざなりにすると分析作業の効率や分析精度が落ちてしまう。そのため、保守管理と運用の負荷の増大は問題点と言える。

2 つ目は、セキュリティ対策についてだ。ビッグデータの中にはパーソナルデータも含まれるため、万が一にも情報漏洩が起こらないようにセキュリティ対策が求められる。EU では GDPR(EU 一般データ保護規則)が施行され、個人データ保護の取り組みが進められた。特に WEB サイト上で取得する Cookie によるブラウジング情報の取得・利用については、EU をはじめ世界的に注視されている。最新の法規制やルールに関する情報をキャッチアップして、適宜対応していく必要がある。

3 つ目は、ハイスキル人材の不足である。ビッグデータを適切に運用するためには、データ活用に長じた高度なスキルを持った人材(データサイエンティスト・データアナリスト)の登用が有効だ。また技術的知見に加えて、ビジネスに対する深い洞察力を有している人材が望ましいとされている。しかし、「データ」と「ビジネス」の両面を高いレベルで満たす人材は特に不足しているため、人々は未だビッグデータを活用しきれない場面も多い。

ビッグデータと対照的に扱われるものとして、スモールデータがある。スモールデータとは、アクセスが容易で取り扱いしやすい形式の、有意義な洞察が可能なデータとされている。ビッグデータの特例のレコードだけを取り出したようなものもスモールデータとして取り扱われるので、元来のデータサイズではなく、利用する時点のデータサイズが基準となる。

ビッグデータと比較してのスモールデータの特徴として、取り扱いが簡単な点が挙げられる。ビッグデータは前述の通り膨大な量のデータを選定し前処理を行うという負荷があるが、スモールデータは扱うデータ量が少ないためその負荷が軽く、ある程度のデータ分析の知識や経験があれば簡単に扱うことができる。

現在このビッグデータ時代に、逆にスモールデータが注目されている。その 1 番の理由は、「ビッグデータの限界」についてだ。これまでの風潮としてビッグデータを活用することで、ありとあらゆる事象を観測したり、推測したりすることができるだろうと考えられてきたが、実際には、その予測や推測が必ずしも正しいものだとは言えないということがわかってきた。つまりビッグデータの限界とは、ビッグデータ活用は万能ではないということである。

ビッグデータとスモールデータとは場面によって使い分けるものではなく、補完関係である。ビッグデータにはビッグデータの、スモールデータにはスモールデータのそれぞれの特徴があり、どちらか一方が優れていてもう一方はそうではないというわけではない。これらが相互に補完し合う関係性で成り立っており、どちらのデータも重要なのだ。

発表がよかったグループは、g17 だと考える。理由としては、データサイエンスと既存の科学の違いを説明する際に、領域という分かりにくい要素を図で表現することによって直感的に伝わりやすくなっており、また例も挙げていることで聞き手ごとの解釈の差異を減らせているのではないかと感じたからだ。

スライドという視覚的に相手に伝えるツールを最大限に活かせていると感じたし、実際私がこの発表をスライドなしで聞いていたら何を言っているのか全く分からなかったことだろう。また、発表者の影山くんの態度もよかったと私は思った。ただのウェブサイトなどの文言を真似しただけではなく、自分で咀嚼して理解した上でわかりやすく伝えようとしているのが伝わってきたし、中西先生の質問に対しても毅然とした態度で自分の言葉で返答していたため、しっかり自分なりに理解しているんだと思った。