



大矢野 潤

◆ データサイエンス演習 ◆

近年、ビッグデータを利用したデータサイエンスは刮目すべき成果をあげている。ここで、統計処理、データマイニング、人工知能などは決して一過性のバズワードではなく、先人が築き上げてきた技術が、コンピュータの低価格化、インターネット上のIoT、SNSなどの隆盛とともに開花し始めているのである。これらの分野において歴史的に語られている「Garbage In, Garbage Out」という言葉にも注目したい。精度の高いシステムに誤りを多く含んだデータを入力すれば、正確に誤った結果を出力するのは当然である。

本演習では手に入れたデータを、現在のさまざまな技術を用いて可視化し、それが本当にゴミなのか、再利用可能な情報を含んでいるのかなどを確認する。さらに、データから有用な情報を抽出し利活用する技術と理論の構築を試みる。