

情報処理応用B 第8回

ビッグデータとIoT

藤田 一寿

1. ガイダンス，コンピュータの歴史
2. 情報の表現（文字，音声，画像）
3. ハードウェアと情報産業
4. ソフトウェアの種類・ライセンス・特許
5. インターネットとビジネス
6. コンピュータネットワークの仕組み
7. 情報セキュリティ
8. IoTとビッグデータ

9. 人工知能 -人工知能とは-
- 10.人工知能 -人工知能の実現方法-
11. 人工知能 -人工知能の技術-
12. 人工知能 -人工知能の研究教育-
13. 人工知能 -プロンプトエンジニアリング-
14. 人工知能 -次世代人工知能-
15. 人工知能 -人工知能のリスク-

今後人工知能の知識が必要となるので，大幅に人工知能の内容を追加しています。

ビッグデータとIoT

ビッグデータ

ビッグデータとは

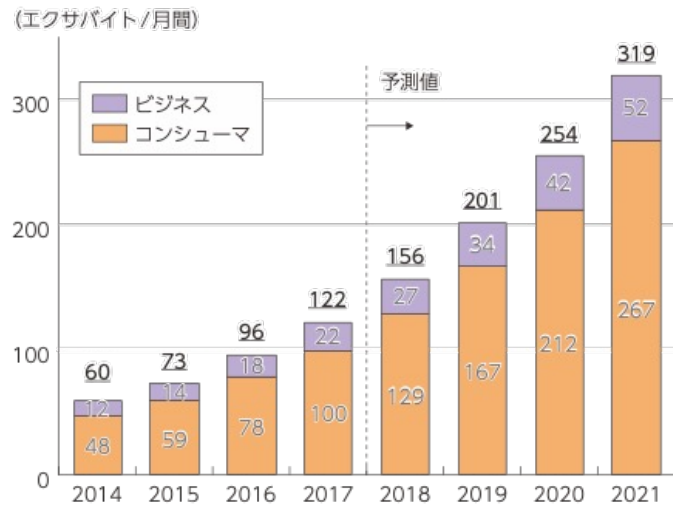
- インターネットでは、様々なデータが湯水の如く湧き出し、山のよう
に保存されている。
- 常時生成され続ける多種多様な大量のデータをビッグデータと
呼ぶ。
- ビッグデータの特徴
 - 3V(volume, velocity, variety)
 - 大量(volumeが大きい)
 - 生成量が多い(velocityが大きい)
 - 多種多様である(varietyに富む)
 - 多次元データである。

データの次元とは

定期試験のデータが有る。数学の点しかない場合、学籍番号と数学の2つの数値しかないので2次元データとなる。

数学、物理、化学、英語、現国の点があるなら学籍番号を含めると6次元データとなる。

世界のトラフィックの推移及び予測



(出典) CiscoVNI

ビッグデータデータはどこからくる？

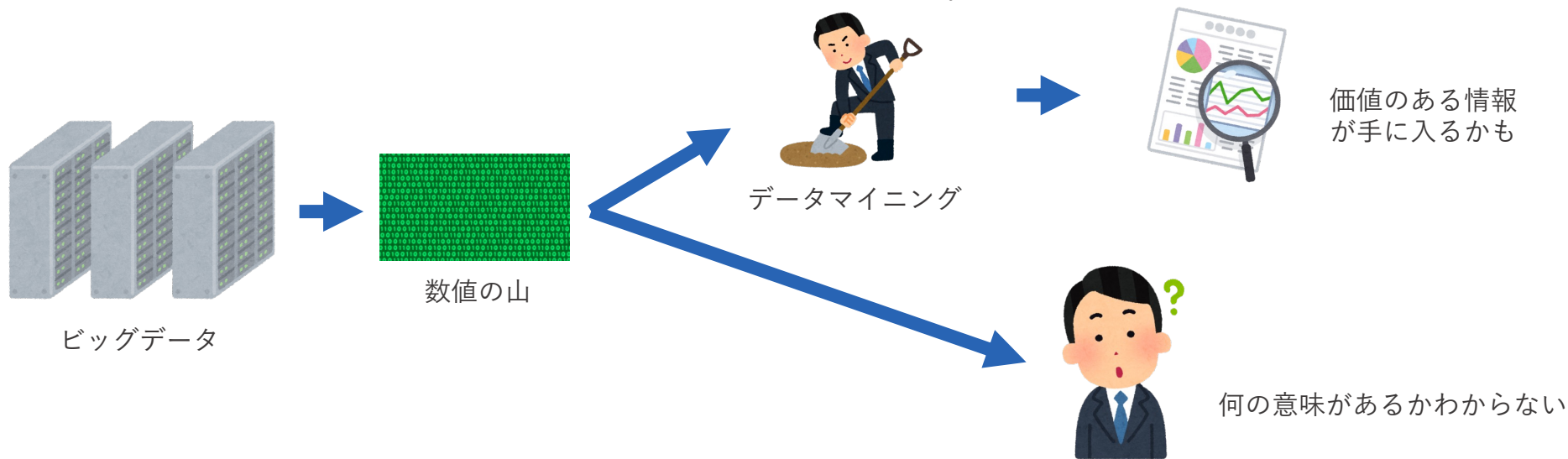


■ ビッグデータだけでは価値を生まない

- 大量のデータ（ビッグデータ）が蓄積されるが、そのままでは意味がない。
- ビッグデータからなにか価値のある事柄を抽出して、初めて意味がある。
 - データマイニングが必要

データマイニングとは

- 大量のデータから何か意味のある情報を取り出す。
 - データ+マイニング（採掘）
- データを解析することで予想外の関係が見出される事がある
- データマイニングの伝説
 - おむつを買った人はビールを買う傾向がある。

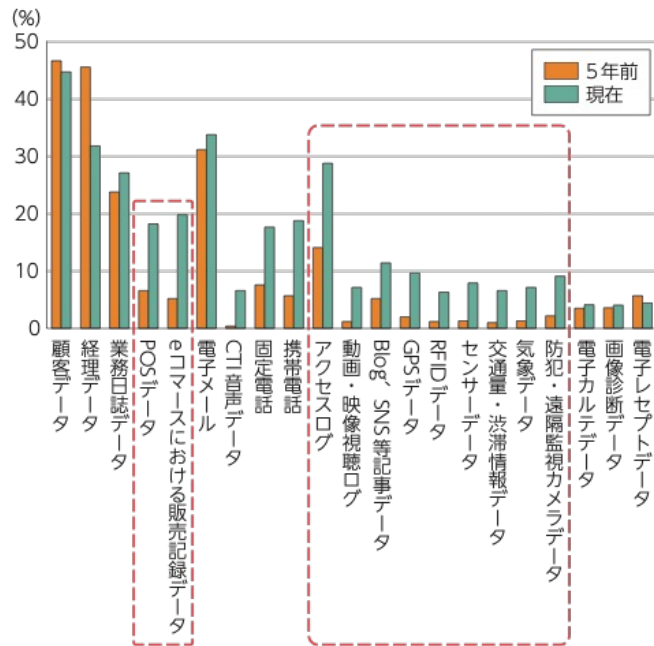


■ 応用例

- 大阪ガス
 - 過去の修理歴とコールセンターの依頼内容から交換部品の自動抽出
- Tsutaya
 - 購買歴を分析し、ソフトの仕入に活かす
- 日本食品
 - クーポンの効果的な発行
- ダイードロリンコ
 - 視線の動きを解析し、視線が行きやすい場所に主力製品を置くことで売上
- スシロー
 - 寿司皿にICタグを取り付け、寿司の情報を管理する。ICタグから得られた情報から需要を予測し、レーンに流すネタや量をコントロールする
- その他、需要予測、価格の決定など様々な場面で活用される。

分析に活用しているデータ

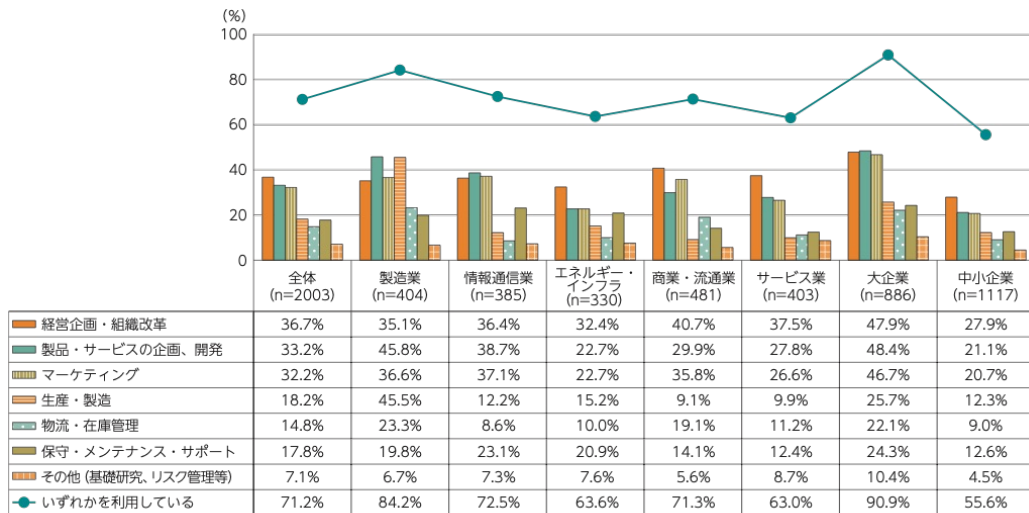
分析に活用しているデータ



業務データ 販売記録 顧客等とのコミュニケーション 自動取得 MtoM 医療

(出典) 総務省 (2020) 「デジタルデータの経済的価値の計測と活用の現状に関する調査研究」

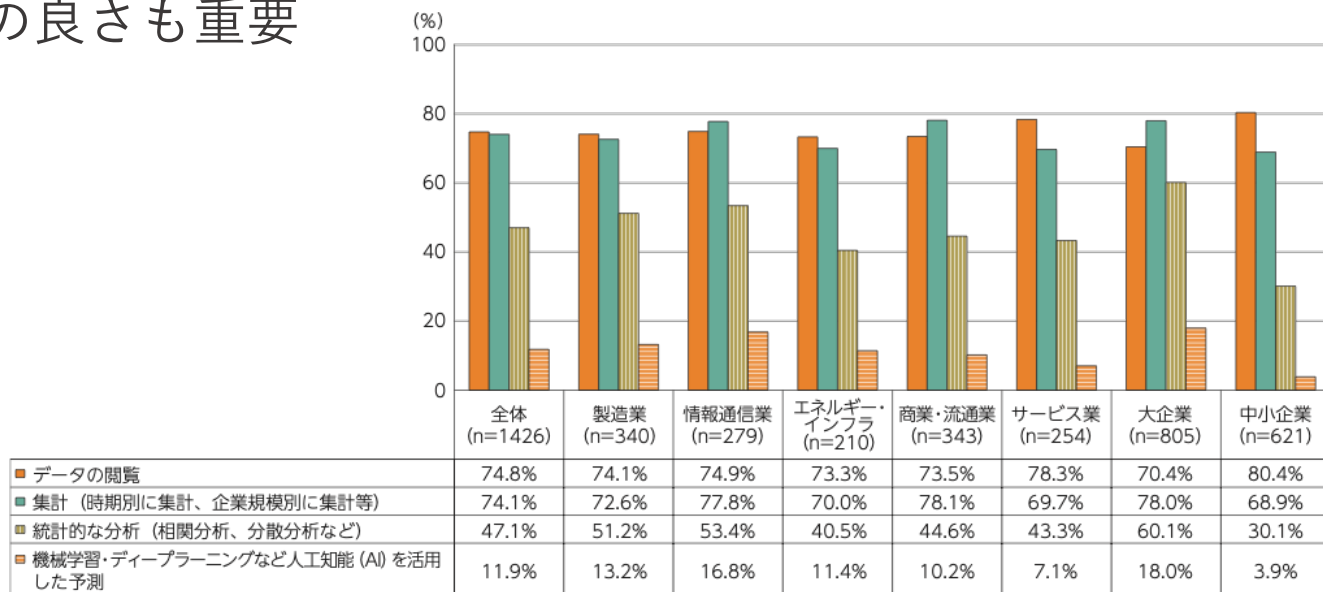
分析を活用している業務領域



(出典) 総務省 (2020) 「デジタルデータの経済的価値の計測と活用の現状に関する調査研究」

データの分析手法

- データの分析手法が重要ではなく、データ分析をして利益につなげることが重要
- 最新の手法、複雑な手法が必ずしも成果を上げるとは限らない
- データの質の良さも重要



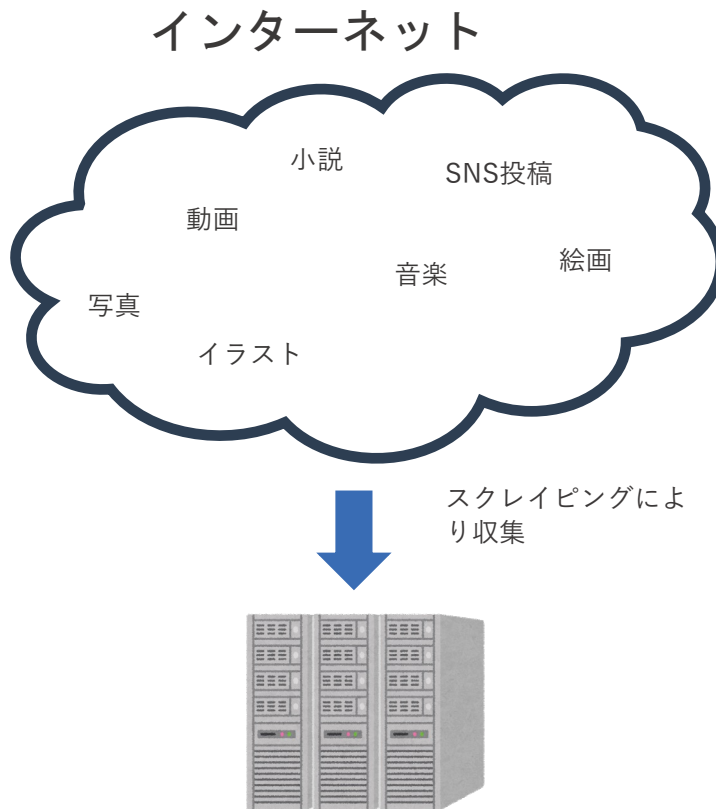
（出典）総務省（2020）「デジタルデータの経済的価値の計測と活用の現状に関する調査研究」

■ オープンデータ

- データが一切の制限無しで、全ての人が望むように利用・再掲載できるような形で入手すべきであるという考え方、またはそのデータ（wikipediaより改）
- オープンデータ基本指針におけるオープンデータの定義
 - 営利目的、非営利目的を問わず2次利用可能なルールが適用されたもの
 - 機械判読に適したもの
 - PDFで公開されているものはコンピュータで読み取りしにくく、せっかくデータを公開しても活用できない。
 - 無償で利用できるもの
- 例
 - 鯖江市が公開している路線や時刻表、現在走っている場所のデータを用い、バスの場所を調べるアプリを開発
 - 全国の自治体が公開している避難所情報を用い、避難所へのルート案内するアプリを開発

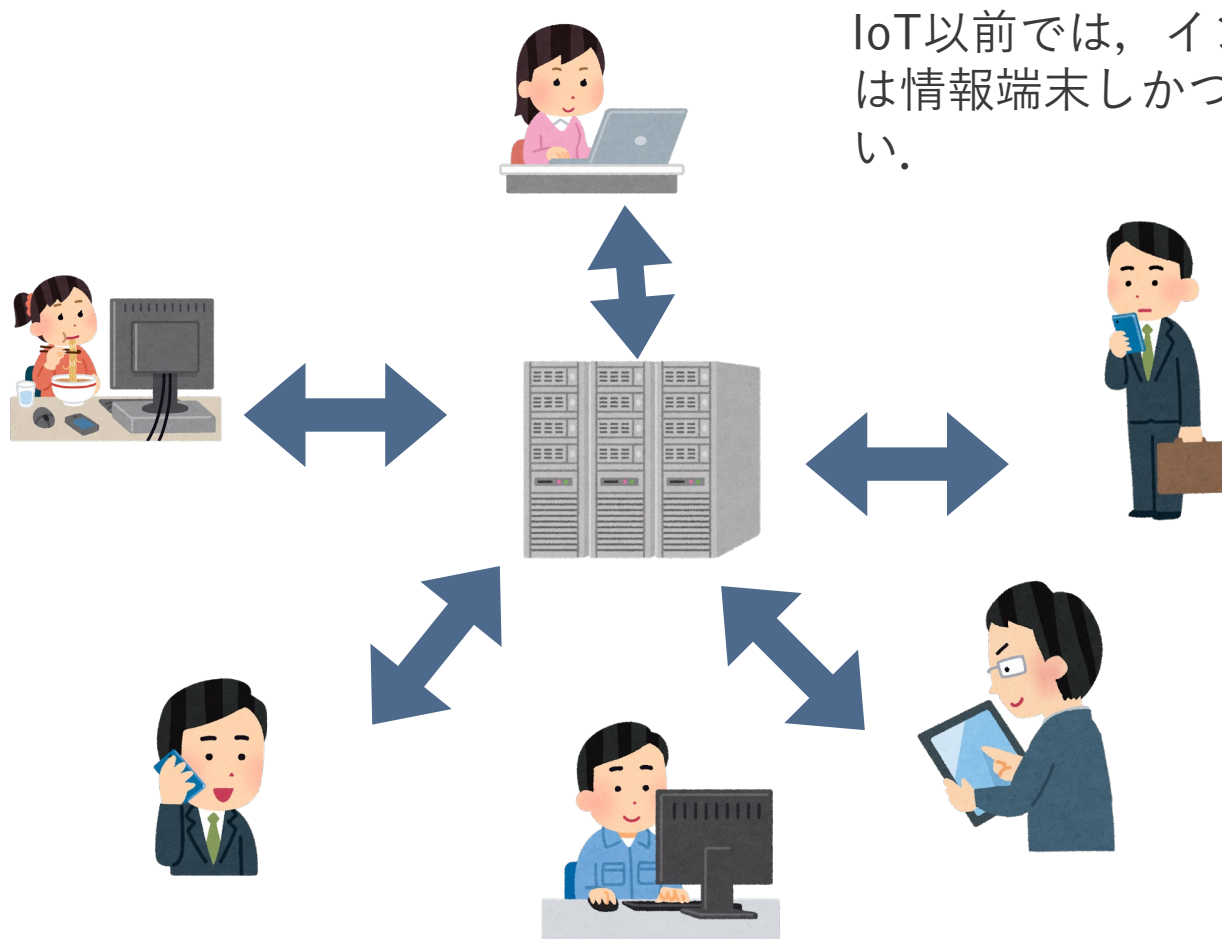
ビッグデータはすでに手元にある

- インターネット上にはすでに大量のデータが存在している。
- それを収集して活用する事もできる。
- ウェブサイトからデータを取り込む技術をスクレイピングという。
- ウェブサイトから定期的にデータを収集しデータベース化するソフトをクローラと言う。
 - 2022年話題になったイラスト生成人工知能は、イラスト投稿サイトからデータをスクレイピングした画像を使って学習したとされている。



IoT

インターネットと情報端末

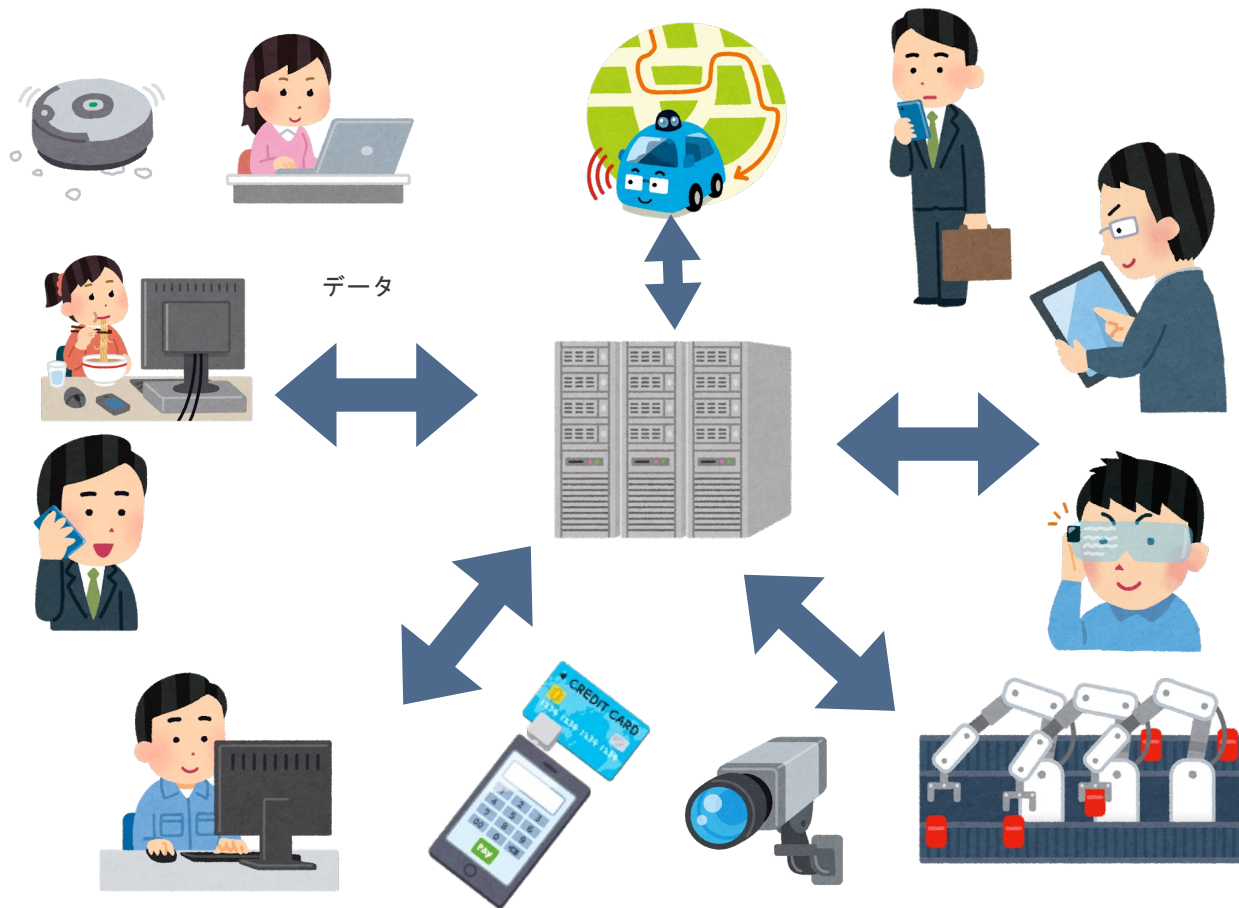


IoT以前では、インターネットには情報端末しかつながっていない。

■ IoTの時代

- IoT (Internet of Things)とはモノのインターネットとも呼ばれ、あらゆるモノがインターネットにつながる社会やその様子を表す.
- 現在でも、パソコンやスマートフォンだけではなく、テレビ、スマートスピーカー、車、監視カメラ、電子レンジなどもインターネットに繋がっている場合がある.
- IoT機器とは、インターネットに繋がるあらゆるモノのことで、特にパソコンやスマートフォンのような以前からインターネットにつながっていたIT機器以外を指すことが多い.

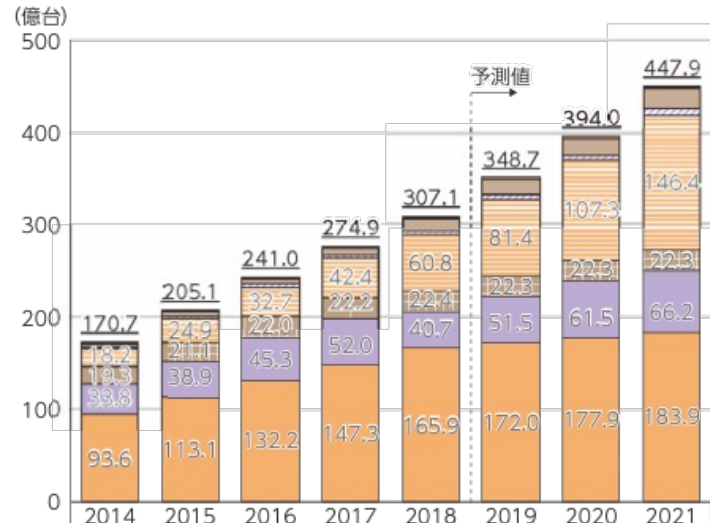
インターネットと機器



- IoT社会は、あらゆるものがインターネットにつながる社会
- つまり、あらゆるものがインターネットにつながると、あらゆるものからデータが送られてくる
- 莫大で様々なデータが貯まる → **ビッグデータ**
- ビッグデータから何か知識を得たい → **データマイニング**
- ビッグデータから何か判断したい → **人工知能**

IoTの現状と予測

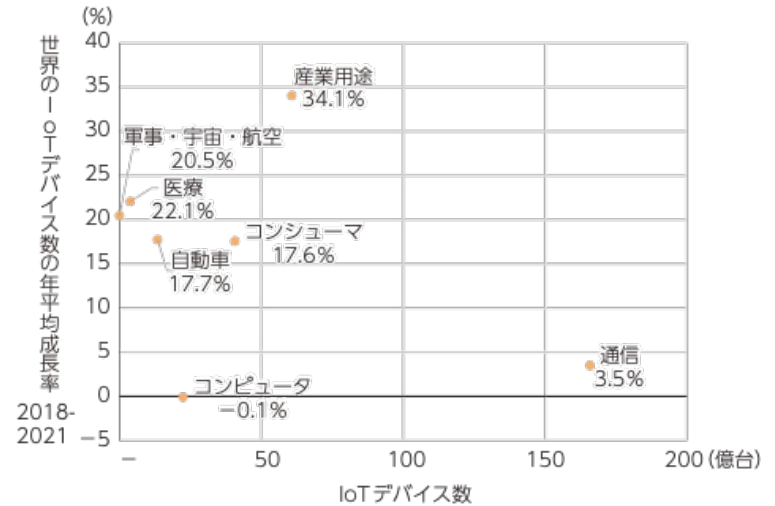
世界のIoTデバイス数の推移および予測



	2014	2015	2016	2017	2018	2019	2020	2021
通信	93.6	113.1	132.2	147.3	165.9	172.0	177.9	183.9
コンシューマ	33.8	38.9	45.3	52.0	40.7	51.5	61.5	66.2
産業用途	18.2	24.9	32.7	42.4	60.8	81.4	107.3	146.4
コンピュータ	19.3	21.1	22.0	22.2	22.4	22.3	22.3	22.3
医療	19.3	21.1	22.0	22.2	22.4	22.3	22.3	22.3
自動車	0.03	0.04	0.04	0.05	0.03	0.04	0.05	0
軍事・宇宙・航空	0.03	0.04	0.04	0.05	0.03	0.04	0.05	0

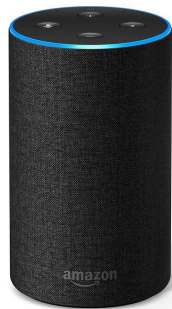
(出典) IHS Technology

分野・産業別のIoTデバイス数および成長率の予測



スマートスピーカ

- 対話型の音声操作に対応したAIアシスタント機能を持つスピーカー。内蔵されているマイクで音声を認識し、情報の検索や連携家電の操作を行う。（wikipediaより）

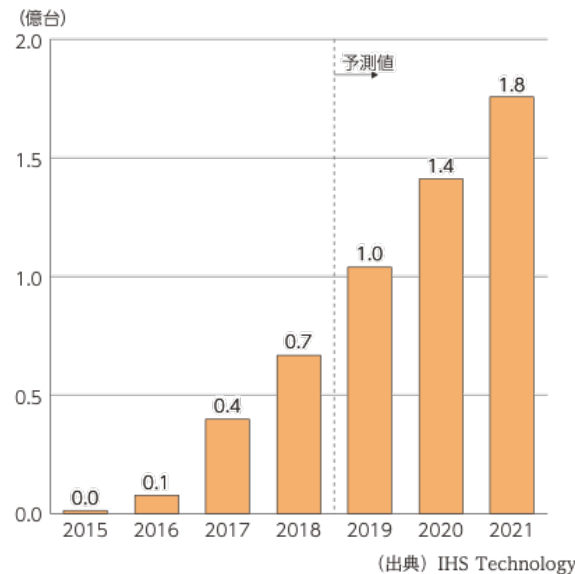


Amazon Echo



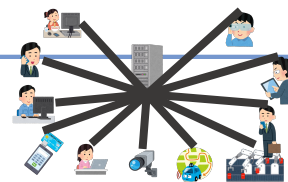
Google Home

世界のスマートスピーカ市場規模の推移と予測

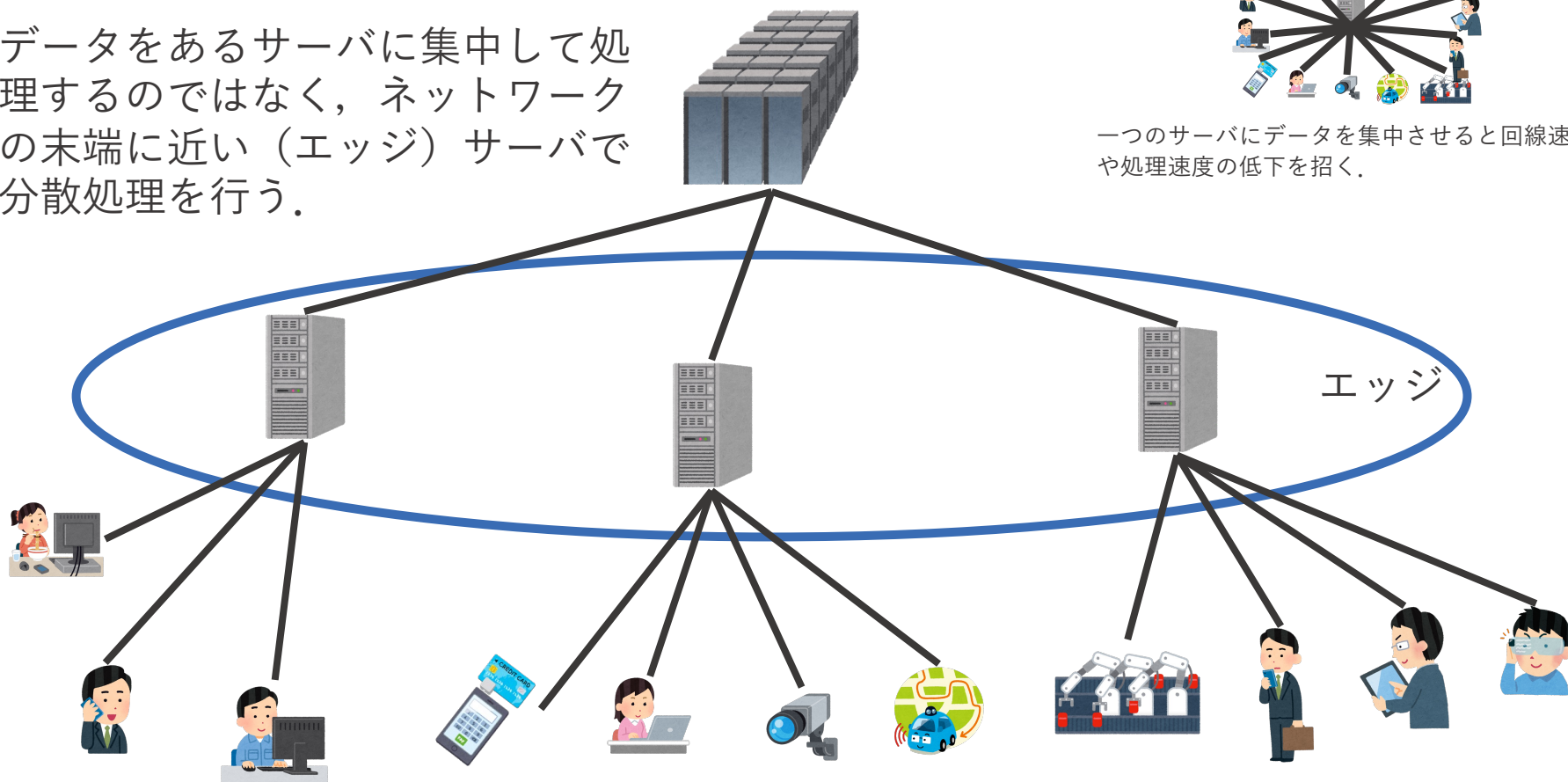


■ エッジコンピューティング

データをあるサーバに集中して処理するのではなく、ネットワークの末端に近い（エッジ）サーバで分散処理を行う。



一つのサーバにデータを集中させると回線速度や処理速度の低下を招く。

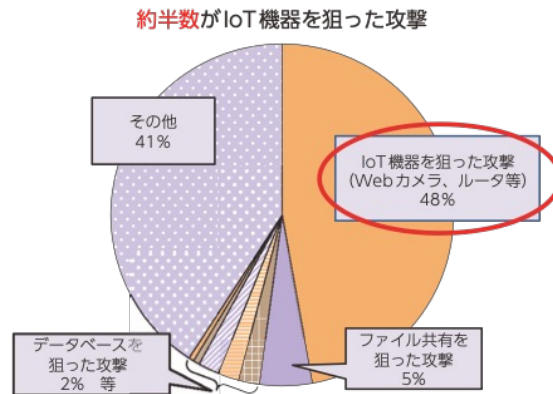


たくさんのIoT機器から大量のデータが送られる

IoT機器とセキュリティ

- IoT機器は、見た目がパソコンやスマートフォン (IT機器) に見えないため、一般ユーザだけではなくメーカーもセキュリティに対する意識が低くなる。
- IoT機器はIT機器と同等の機能を有するため (LinuxなどのOSが入っている)、IT機器と同等のセキュリティ対策をする必要がある。

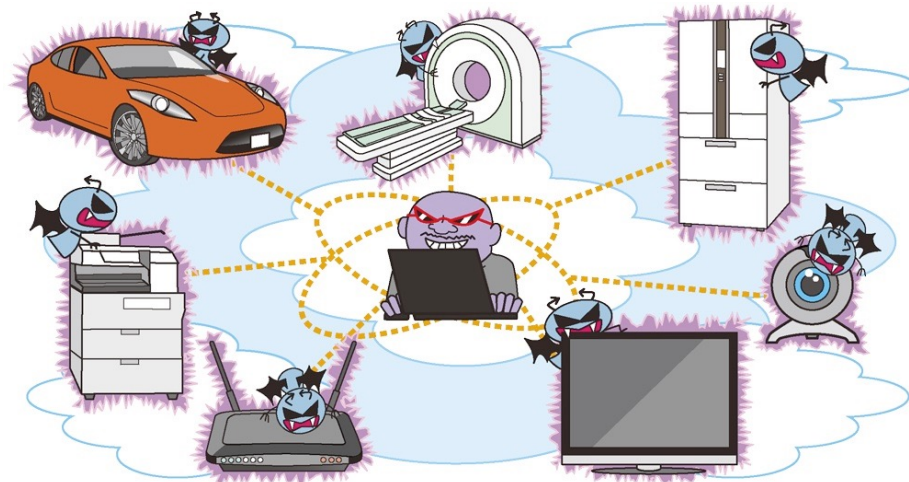
宛先ポート番号別の年間観測パケット数割合



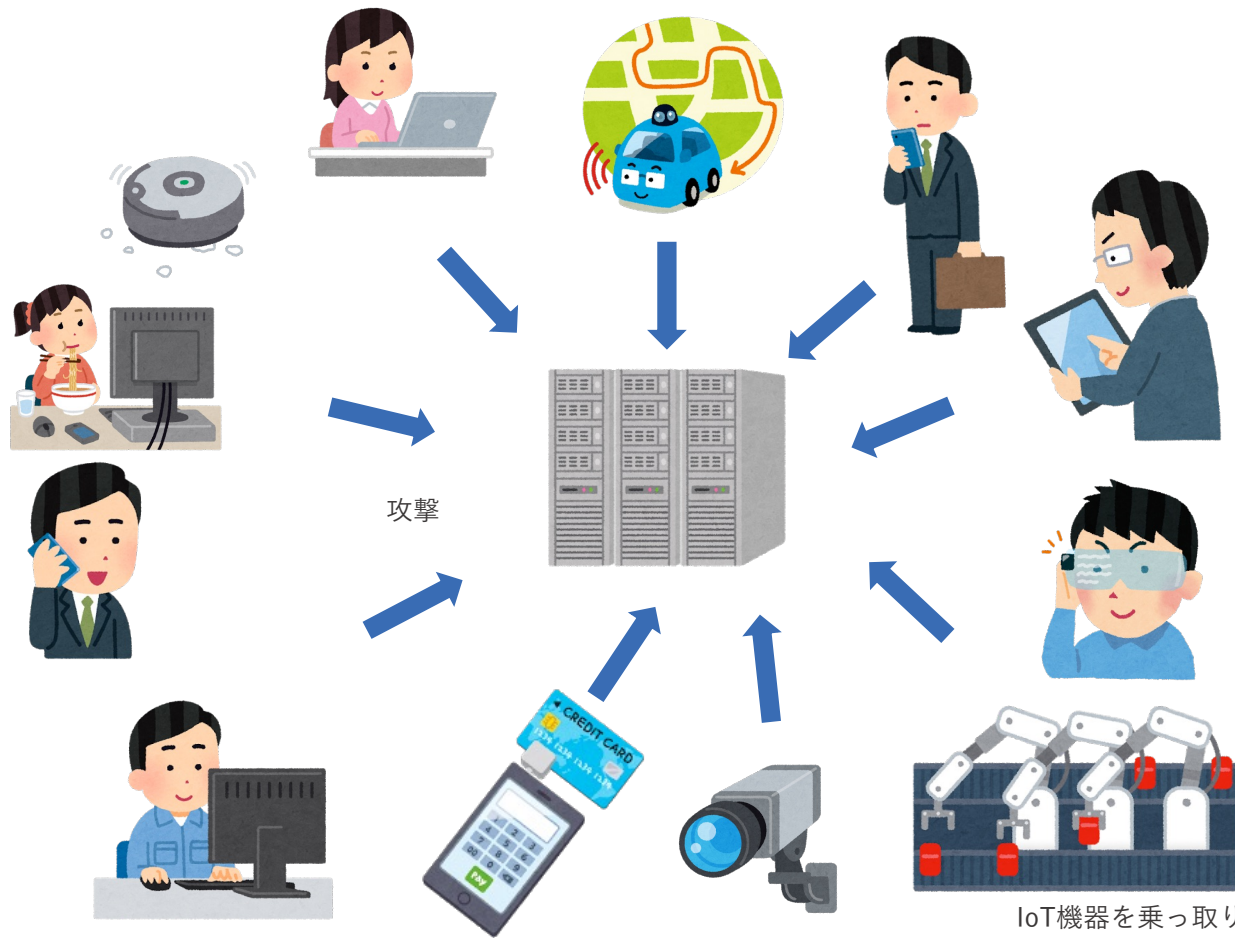
- *1 NICTERで観測されたパケットのうち、サービスの種類 (ポート番号) ごとに割合の多い上位から30位までを分析したもの。
- *2 IoT機器を狙った攻撃は多様化しており、ポート番号だけでは分類しにくいものなど、「その他」に含まれているものもある。

IoTのセキュリティ問題

- IoT機器にマルウェアを感染させ、マルウェアをばらまく.
- IoT機器マルウェアを感染させ、DDoS攻撃を行う.
- IoT機器から機密情報を窃取する.
- IoT機器を遠隔制御する.



IoTからの攻撃



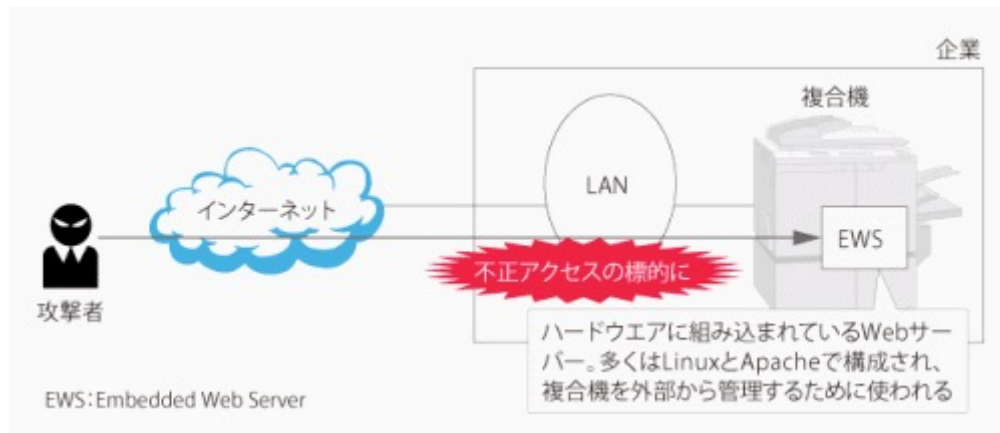
IoT機器を乗っ取り攻撃に参加させる

■ 監視カメラの事例

- 監視カメラはインターネットから動画の視聴，カメラの向きの制御などが行えるものがある。
- 監視カメラの画像の流出
 - 監視カメラの中には，画像がインターネットに公開される設定になっているものがある．世界中の監視カメラ画像を見られるサイトも存在する。
- 監視カメラの乗っ取り
 - 2018年上尾市の河川監視カメラに不正アクセスがあり，「I'm hacked. bye2」と監視カメラの画像に表示された．さらに，パスワードが変更され，制御不可能になった
(<https://www.sankei.com/region/news/180428/rgn1804280045-n1.htm>,
<https://scan.netsecurity.ne.jp/article/2018/05/01/40886.html>)

■ プリンタと複合機の事例

- プリンタや複合機はインターネットから印刷，スキャンされた資料の閲覧，設定の変更などが行えるものがある。
- 2013年複数の大学等で，複合機で読み取った情報がインターネット上で閲覧できる状態となっていたことが判明した。



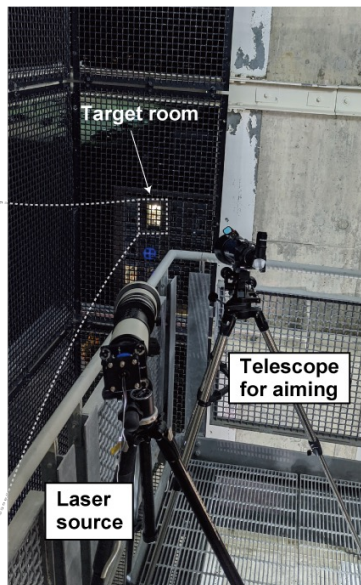
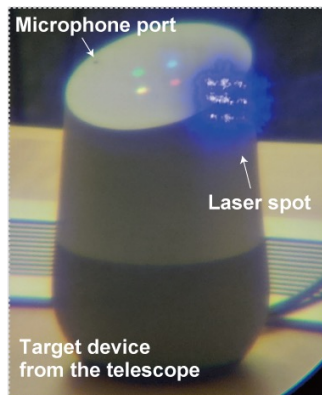
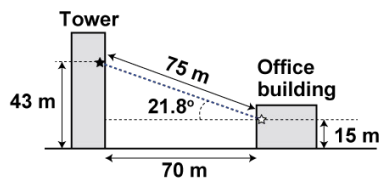
■ テスラモデルSの遠隔操作

- テスラ車の通信機能に存在した脆弱な仕様について車載情報端末用の車内ネットワークに侵入
- 車載情報端末のWebブラウザーに存在した脆弱性を攻撃して、任意のコードを実行可能に
- 車載情報端末のLinuxカーネルに存在した脆弱性を攻撃して、ルート権限を取得
- 情報端末用の車内ネットワークと、制御系ネットワーク（CAN）とをつなぐ「コントローラー」を攻撃して、コントローラーのファームウェアを書き換え
- コントローラーから電子制御ユニット（ECU）に偽のコマンドを送り自動車を遠隔操作

(<https://www.youtube.com/watch?v=c1XyhReNcHY>)

Light commands

- スマートスピーカーのマイクが音だけではなく光にも応答する特性を利用した攻撃。
- 音声信号をのせたレーザーをマイクに当てることで、スマートスピーカーを遠隔から操ることができる。



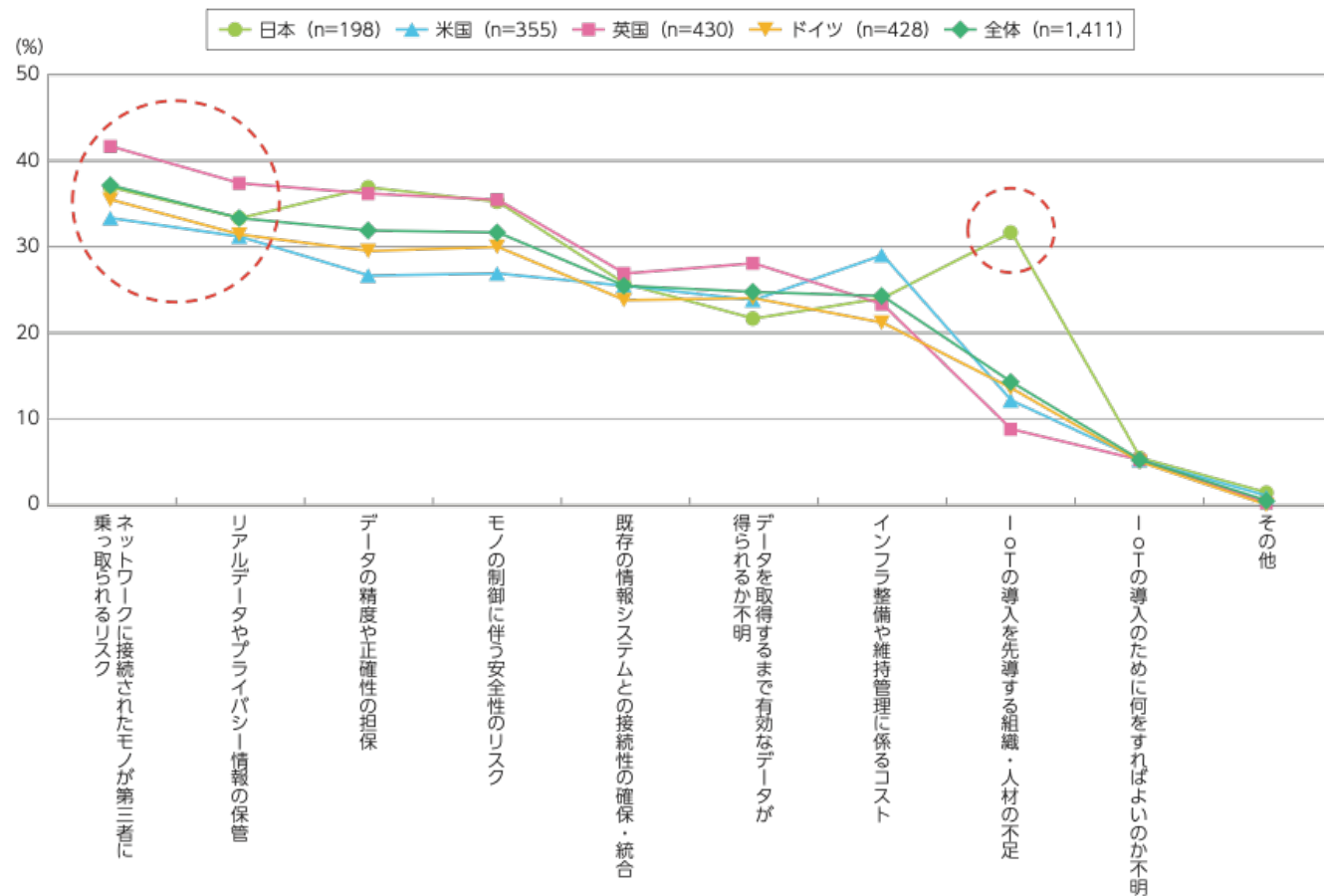


LIGHT COMMANDS

■ IoT機器のセキュリティ対策

- IoT機器は一見すると機能が限定され安全に見えるが、中身はパソコンと同じであると考えて、パソコンやネットワーク機器と同様のセキュリティ対策を行う。
- 利用者は少なくとも次の対策はする必要がある。
 - ソフトウェアを最新のものにする。
 - 初期設定を使わず、よりセキュリティの高い設定にする。
 - ID、パスワードだけではなく、ネットワークの設定も。

IoTの課題



(出典) 総務省「ICTによるイノベーションと新たなエコノミー形成に関する調査研究」(平成30年)

演習

■ 演習

- ビッグデータを企業が活用している事例はどれか。(基本情報平成29年秋期)
1. カスタマセンタへの問合せに対し、登録済みの顧客情報から連絡先を抽出する。
 2. 最重要な取引先が公表している財務諸表から、売上利益率を計算する。
 3. 社内研修の対象者リスト作成で、人事情報から入社10年目の社員を抽出する。
 4. 多種多様なソーシャルメディアの大量な書込みを分析し、商品の改善を行う。

■ 演習

- ビッグデータを企業が活用している事例はどれか。(基本情報平成29年秋期)
1. カスタマセンタへの問合せに対し、登録済みの顧客情報から連絡先を抽出する。
 2. 最重要な取引先が公表している財務諸表から、売上利益率を計算する。
 3. 社内研修の対象者リスト作成で、人事情報から入社10年目の社員を抽出する。
 4. 多種多様なソーシャルメディアの大量な書込みを分析し、商品の改善を行う。

■ 演習

- データマイニングの説明として、適切なものはどれか。(基本情報平成19年秋期)
 1. 大量のデータを高速に検索するための並行的アクセス手法
 2. 大量のデータを統計的，数学的手法で分析し，法則や因果関係を見つけ出す技術
 3. 販売実績や製造実績などの時系列データを大量に蓄積するデータベースの保存手法
 4. ユーザの利用目的に合わせて，部門別のデータベースを作成する技術

- データマイニングの説明として、適切なものはどれか。(基本情報平成19年秋期)
 1. 大量のデータを高速に検索するための並行的アクセス手法
 2. 大量のデータを統計的、数学的手法で分析し、法則や因果関係を見つけ出す技術
 3. 販売実績や製造実績などの時系列データを大量に蓄積するデータベースの保存手法
 4. ユーザの利用目的に合わせて、部門別のデータベースを作成する技術

■ 演習

- IoT(Internet of Things)を説明したものはどれか。(基本情報技術者試験平成28年春期)
1. インターネットとの接続を前提として設計されているデータセンタのことであり、サーバ運用に支障を来さないように、通信回線の品質管理、サーバのメンテナンス、空調設備、瞬断や停電に対応した電源対策などが施されている。
 2. インターネットを通して行う電子商取引の一つの形態であり、出品者がWebサイト上に、商品の名称、写真、最低価格などの情報を掲載し、期限内に最高額を提示した入札者が商品を落札する、代表的な CtoC 取引である。
 3. 広告主のWebサイトへのリンクを設定した画像を広告媒体となるWebサイトに掲載するバナー広告や、広告主のWebサイトの宣伝をメールマガジンに掲載するメール広告など、インターネットを使った広告のことである。
 4. コンピュータなどの情報通信機器だけでなく様々なものに通信機能をもたせ、インターネットに接続することによって自動認識や遠隔計測を可能にし、大量のデータを収集・分析して高度な判断サービスや自動制御を実現することである。

■ 演習

- IoT(Internet of Things)を説明したものはどれか。(基本情報技術者試験平成28年春期)
1. インターネットとの接続を前提として設計されているデータセンタのことであり、サーバ運用に支障を来さないように、通信回線の品質管理、サーバのメンテナンス、空調設備、瞬断や停電に対応した電源対策などが施されている。
 2. インターネットを通して行う電子商取引の一つの形態であり、出品者がWebサイト上に、商品の名称、写真、最低価格などの情報を掲載し、期限内に最高額を提示した入札者が商品を落札する、代表的な CtoC 取引である。
 3. 広告主のWebサイトへのリンクを設定した画像を広告媒体となるWebサイトに掲載するバナー広告や、広告主のWebサイトの宣伝をメールマガジンに掲載するメール広告など、インターネットを使った広告のことである。
 4. コンピュータなどの情報通信機器だけでなく様々なものに通信機能をもたせ、インターネットに接続することによって自動認識や遠隔計測を可能にし、大量のデータを収集・分析して高度な判断サービスや自動制御を実現することである。

- IoT(Internet of Things)の実用例として、適切でないものはどれか。(基本情報平成30年春期)
1. インターネットにおけるセキュリティの問題を回避する目的で、サーバに接続せず、単独でファイルの管理、演算処理、印刷処理などの作業を行うコンピュータ
 2. 大型の機械などにセンサと通信機能を内蔵して、稼働状況、故障箇所、交換が必要な部品などを、製造元がインターネットを介してリアルタイムに把握できるシステム
 3. 検針員に代わって、電力会社と通信して電力使用量を送信する電力メータ
 4. 自動車同士及び自動車と路側機が通信することによって、自動車の位置情報をリアルタイムに収集して、渋滞情報を配信するシステム

- IoT(Internet of Things)の実用例として、適切でないものはどれか。(基本情報平成30年春期)

1. インターネットにおけるセキュリティの問題を回避する目的で、サーバに接続せず、単独でファイルの管理、演算処理、印刷処理などの作業を行うコンピュータ
2. 大型の機械などにセンサと通信機能を内蔵して、稼働状況、故障箇所、交換が必要な部品などを、製造元がインターネットを介してリアルタイムに把握できるシステム
3. 検針員に代わって、電力会社と通信して電力使用量を送信する電力メータ
4. 自動車同士及び自動車と路側機が通信することによって、自動車の位置情報をリアルタイムに収集して、渋滞情報を配信するシステム