KOREA CAPITAL MARKET INSTITUTE
자본시장연구원

Data and Financial Innovation

October 2, 2019

Min Kyeong Kwon

Contents

1 Data Concepts
2 Data Applications
3 Data and Innovation
4 Implications

Data Concepts

- What does the dictionary define **data**?
  - Factual information used as a basis for making a decision or developing a theory;
  - Facts or information acquired by observation, experiment or investigation;
  - Information in the form of text, numbers, sound, and images that can be processed by a computer

- Widely used meaning of data
  - Things necessary to create information
    - Data → Information → Knowledge
      - E.g., measuring global temperatures for the past 100 years (data) → Global warming is underway (information)
    - Evaluated as 'the oil of the 21st century', 'a source of a competitive advantage that surpasses the traditional factors of production, such as capital and labor' and so on.

  - Information stored in some format that can be processed by the computer
    - Changes in the way data is collected or gathered and analyzed
      (1) Storing data by hand (manual data storage) → manual data analysis
      (2) Entering data manually into the computer (manual data entry) → data analysis using the computer
      (3) Accumulating data automatically in the computer, machines, sensors or other devices → computer-assisted data analysis

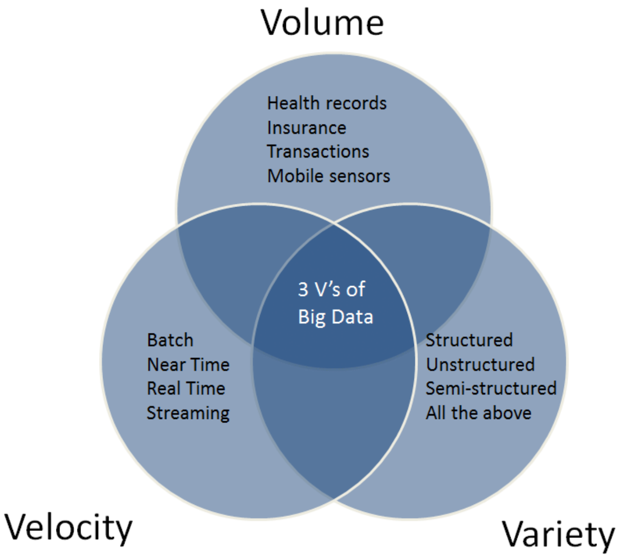
Data Concepts

Related keywords

Big data

- An extremely large dataset that cannot be stored or processed by traditional methods
- 3Vs (volume, velocity, and variety) are three defining properties of big data.
- Big data's growth started with the explosive growth of data driven by advances in sensors, IoT devices, smartphones, wireless networks, logs, cameras, microphones, RFID, and social networking platforms.
- Algorithms such as machine learning algorithms are also rapidly developed to analyze big data.

3 V's of Big Data


Volume
Health records
Insurance
Transactions
Mobile sensors
3 V's of
Big Data
Batch
Near Time
Real Time
Streaming
Structured
Unstructured
Semi-structured
All the above
Velocity
Variety

Source: Austin and Kusumoto (2016) (the diagram that describes the 3V definition of big data presented in Laney (2001))

Data Concepts

- Big data platform: Data collection → storage → processing → analysis → description and visualization
  - Big data employ differentiated methods including distributed data storage and processing.
  - Real time processing has been growing gradually.
  - Traditional data analytics focuses on uncovering causal relationships whereas big data analytics primarily examines data to discover correlations.

Key functions of big data platform components

Data collection	Data storage	Data processing	Data analytics	Description
Unstructured data collection Structured data collection ETL Web Robot EAI, ESB, FTP, etc. Open API	Raw data NoSQL Memory Search engine Data security	Batch processing Real time processing (CEP)	Text analytics Machine learning Statistics Data mining SNS analysis Predictive analytics (algorithms)	Visualization

Source: Ahn, Chunmo (2017)

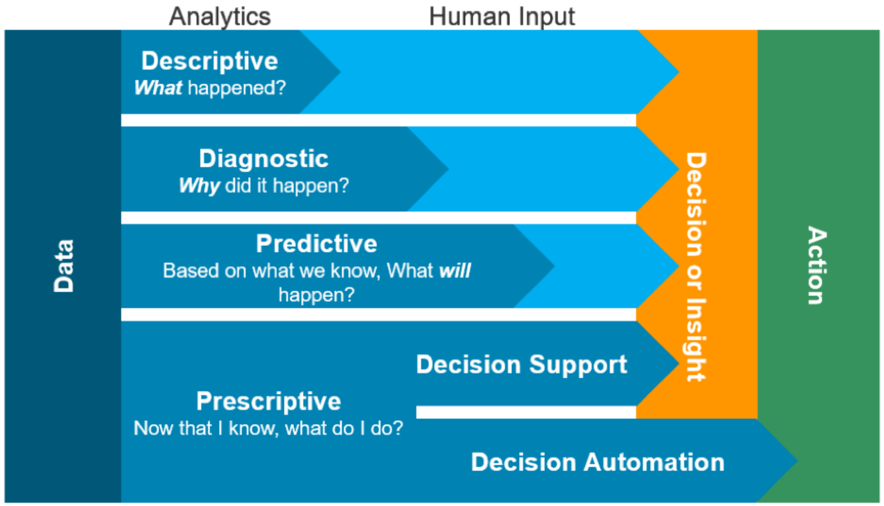
Cloud computing

- Computing services, such as data storage and computing power, delivered on demand by external providers
- Core infrastructure for big data applications

Data Concepts

Data applications

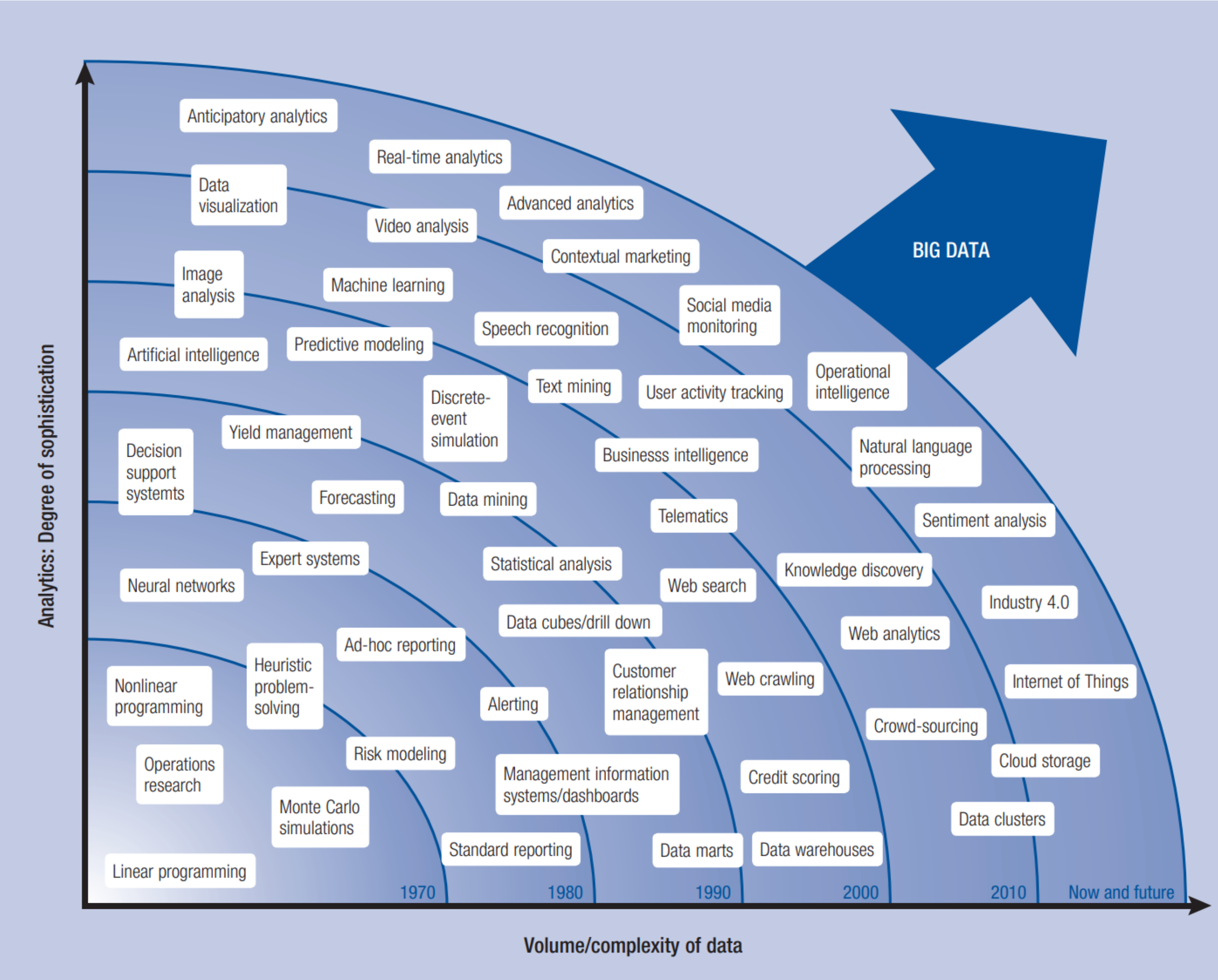
- Data are used to identify and solve problems.
- Increasing likelihood of data usage at each stage
  - (descriptive) What happened?
  - (diagnostic) Why did it happen?
  - (predictive) What will happen?
  - (prescriptive) What do I do?
- Change in the basis for decision-making from intuition to data.


Analytics
Human Input
Data
Descriptive
What happened?
Diagnostic
Why did it happen?
Predictive
Based on what we know, What *will* happen?
Prescriptive
Now that I know, what do I do?
Decision Support
Decision Automation
Decision or Insight
Action

Source: Gartner(2015)

Data Concepts

Evolution of data-based analytics

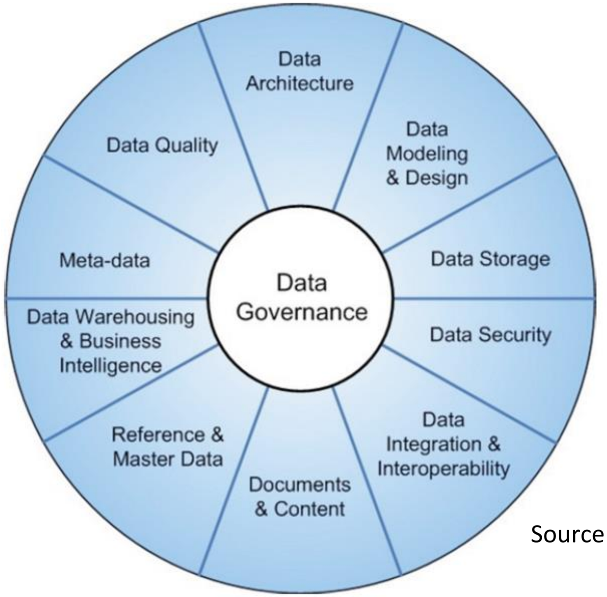

Anticipatory analytics
Real-time analytics
Data visualization
Advanced analytics
Video analysis
Contextual marketing
BIG DATA
Image analysis
Machine learning
Social media monitoring
Speech recognition
Predictive modeling
Artificial intelligence
Operational intelligence
Text mining
User activity tracking
Discrete-event simulation
Yield management
Decision support systemts
Businesss intelligence
Natural language processing
Forecasting
Data mining
Telematics
Sentiment analysis
Expert systems
Statistical analysis
Knowledge discovery
Neural networks
Web search
Industry 4.0
Data cubes/drill down
Ad-hoc reporting
Web analytics
Heuristic problem-solving
Customer relationship management
Nonlinear programming
Web crawling
Alerting
Internet of Things
Crowd-sourcing
Risk modeling
Operations research
Management information systems/dashboards
Credit scoring
Cloud storage
Monte Carlo simulations
Data clusters
Standard reporting
Data marts
Data warehouses
Linear programming
1970
1980
1990
2000
2010
Now and future
Analytics: Degree of sophistication
Volume/complexity of data

Source: Booz & Company, 2014, Big data maturity: An action plan for policymakers and executives

Data Concepts

Data management

- It is aimed at creating value by acquiring high quality data and actively using them.


Data Governance
Data Architecture
Data Modeling & Design
Data Storage
Data Security
Data Integration & Interoperability
Documents & Content
Reference & Master Data
Data Warehousing & Business Intelligence
Meta-data
Data Quality

Source: Cho, Wanseob (2017)

- Methods of data sharing
  - (1) Data exist and are stored separately.
  - (2) Data are managed in departmental silos.
  - (3) Enterprise-wise data sharing
  - (4) External data consolidation

Five stages of a company's analytical competitiveness

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Poor data quality; occasional generation of data analysis reports by individual business units	Generation of data analysis reports on a regular basis by individual business units	Partial business intelligence (BI) approach; lack of enterprise-wide data integration; data silos	Enterprise-wide data integration; data quality control; generation of reports with integrated enterprise data	Use of external data; advanced statistical analytics and optimization

Source: Kim, Okgi (2018)

Contents

1 Data Concepts

2 Data Applications

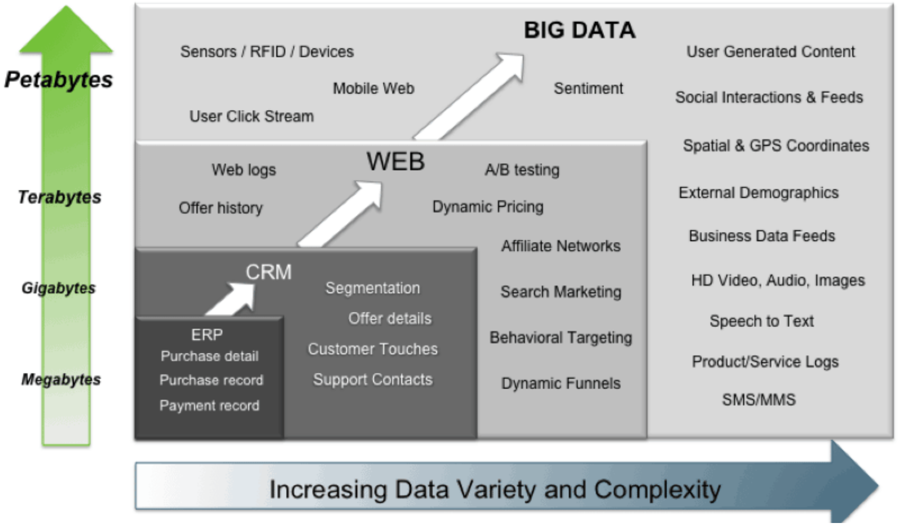
3 Data and Innovation

4 Implications

Data Applications

- Data applications in non-financial firms and public institutions
  - Types of data
    1. (1) Enterprise Resource Planning (ERP)
    2. (2) Customer Relationship Management (CRM)
    3. (3) Weblogs
    4. (4) Data generated by sensors, RFID, mobile web, user click streams, SNS and other various platforms

Size and types of data


Petabytes
Terabytes
Gigabytes
Megabytes
BIG DATA
Sensors / RFID / Devices
Mobile Web
User Click Stream
Sentiment
User Generated Content
Social Interactions & Feeds
Spatial & GPS Coordinates
External Demographics
Business Data Feeds
HD Video, Audio, Images
Speech to Text
Product/Service Logs
SMS/MMS
WEB
Web logs
Offer history
A/B testing
Dynamic Pricing
Affiliate Networks
Search Marketing
Behavioral Targeting
Dynamic Funnels
CRM
Segmentation
Offer details
Customer Touches
Support Contacts
ERP
Purchase detail
Purchase record
Payment record
Increasing Data Variety and Complexity

Source: Teradata

Data Applications

- Purposes of using data
  - (General) analysis of business performance using performance measures such as revenue and profit
  - (Manufacturing) backlog management, inventory and quality control, and maintenance and repairs
    - GE generates 75% of its revenue from providing maintenance and repairs based on data from sensors on its products.
    - Siemens has increased its productions volume eight-fold by analyzing data from facilities and re-adjusting its production lines.
    - Volvo increased efficiency in vehicle maintenance using data from sensors installed in Volvo cars.
  - (Sales and marketing) analysis of sales patterns and customers, and marketing
    - Customer acquisition and retention, and up-sell and cross-sell
    - A company's call center classifies customer propensities based on a history of previous customer interactions, and provides customized responses to customers. Customer feedback is used to develop a next-generation statistic model of customer behavior.
    - Product development which reflects the results of weblogs and social platform analysis
  - (Public sectors) healthcare, communications, welfare, transportation, environment, and crime/fraud detection
    - Crime predictions (predicting times and places that have high probability for crime occurrence), offender profiling, etc.
    - Traffic forecasting, toll road revenue estimation, and real time traffic information service
    - Preemptive response to the spread of bird flu or avian influenza

Data Applications

- Data applications in financial services firms
  - Driving factors behind big data usage (Gutierrez, 2014)
    - Channels for the delivery of financial products or services go online.
      - Exponential growth in the amount of customer data which are not accumulated offline
      - Significantly increased frequency of transactions in financial products due to the ease of executing financial transactions online
    - New sources of data generated from new platforms, such as social media data
      - Reasoning with data about a group to which an individual belongs to or relationships between individuals
    - Tighter risk exposure rules and reporting requirements for financial services firms

  - Types of data
    - Large amounts of data are generated by financial market infrastructures (systems) and market participants, just as data are accumulated by sensors installed in factory machines.
    - Market data
      - Market order and transaction data, press releases, disclosures, analyst reports, news reports, and social media data
      - Data collected by satellites and sensors, and external database
    - Customer data
      - Data on the use of channels by customers, including ATM, call center, online, and branch
      - Data on transactions in financial instruments, such as mortgage loans, and credit cards

Data Applications

- Purposes of using data
  - (Production) To understand market risk and sentiments, analyze credit risk, and make pricing decisions
    - To measure counterparty credit risk
    - To identify risk factors in complex financial instruments such as mortgages, and make pricing decisions
  - (Sales and marketing) To categorize customers according to their consumption behavior and risk profiles, provide services tailored to customers, and acquire customer touchpoints.
    - Capturing information about a client's engagement party through SNS → Recommending a loan product targeting a newly-wed couple
    - A woman in her 30s working at a company located in Yeoido → Recommending the most popular financial product in the same group
  - (Operations) To reduce operational costs, comply with tougher financial regulation and reporting requirements, and manage financial liquidity, strengthen internal controls, and detect abnormal trading activity and events.
    - Detecting an attempt made in Busan to withdraw money from a customer's account into which money has been deposited via ATM in Seoul ten minutes ago → Signs of abnormality
    - Korea Exchange (KRX) considers the establishment of a market oversight system based on machine learning.

Contents

1 Data Concepts
2 Data Applications
3 Data and Innovation
4 Implications

Data and Innovation

Open data

- Data that's available to everyone to access, use and share (Open Data Institute's website)

- Government data should be open and made available by anyone without restrictions.
  - G8 leaders signed the Open Data Charter at the G8 Summit in 2013.
    - Open data by default
    - Quality and quantity
    - Usable by all
    - Releasing data for improved governance
    - Releasing data for innovation
  - The data must be technically open as well as legally open (see Open Knowledge Foundation).
  - There are exemptions including information and data that could have national security, public good, or privacy implications.

- Machine-readable data
  - Data in a format that can be read, modified, converted or extracted by software

Data and Innovation

- Open access to government data
  - United States (US)
    - Adopted open data as part of Transparency and Open Government policies in 2009.
    - The federal government runs its open government data portal (data.gov), with proactive participation by state governments.
    - Private sector companies like Zillow and WestLaw generate revenue by leveraging open data.*
      * See Park, Kyeonghyun, et al. (2017)

  - United Kingdom (UK)
    - Embarked on its open data initiative in 2010 with David Cameron's inauguration as British prime minster.
    - Running its open government data portal (data.gov.uk).
    - Making efforts to facilitate the use and release of government data through Open Data Strategy (2014).
    - Private sector companies such as Open Corporates and Spend Network use government data to produce revenue.*
      * See National IT Industry Promotion Agency (2014).

Data and Innovation

- South Korea
  - Article 3.1 of the Act on Promotion of the Provision and Use of Public Data: “Every public institution shall endeavor to enable anyone to readily use public data and shall take measures necessary to promote universal access to the use thereof.”
  - The government’s strategy to revitalize the data industry (June 2018)
    - Selecting data whose demand from the private sector is high as national core data and providing open access to them early.
    - Providing greater access to private sector data of public good nature: Open access to big data on communications, portal search, news reports, distribution, and financial transactions that are in high need for industrial purposes.
  - Launched a public data portal (data.go.kr), and Seoul Open Data Portal.

Pilot projects in key sectors (proposed)

Sector	Institution Name	Core Data	Use Cases
Medicine & Healthcare	Health Insurance Review & Assessment Service/National Health Insurance Service	Patient personal information, medical department, disease name, healthcare benefits, medication, etc.	Patient-tailored diagnosis and treatment services, development of precision medical solutions , etc.
	Private healthcare facilities	Patient records, medical imaging data, prescription, etc.	
Transportation	Korea Transport Institute/Korea Transportation Safety Authority	Road conditions, road facilities management, traffic volume, location of an accident, damage, etc.	Analytics services to reduce traffic jams and find the causes of car accidents, etc.
	Tmoney	Hourly and regional mobility, the number of passengers getting on/off	
Finance	Bank of Korea/Korea Credit Information Services	Economy/finance statistics, retail/corporate loans, tax delinquency, defaults, bankruptcy, etc.	Development of customized financial services, insurance fraud analysis, development of delinquency prediction models, etc.
	Banks/insurance companies/credit card companies	Account info, loans, product sales/purchase, internet banking usage, customer complaints, credit card merchants, etc.	
Telecoms/media	Telecom operators/IPTV operators	Subscriber/location info, floating population, traffic per service/purchase details, etc.	Blocking the spread of infection, commercial area analysis, content recommendation service, ad strategy development, etc.
	Korea Press Foundation/Korea Broadcast Advertising Corp. (KOBACO)	Employees, current status of the advertising market, subscription patterns, digital contents, sales, etc.	

Source: Government, Extracted partly from the table in the Strategy to Revitalize the Data Industry (June 2018)

Data and Innovation

- Core public data (National Information Society Agency)
    - (Culture/tourism) book-lending data in national libraries; 3D raw scan data on national treasures; tourism and commercial areas information
    - (Transportation/logistics) parking lots; road signs and traffic information; electric vehicle charging locations
    - (Environment/climate) weather information service
    - (Medicine/healthcare) hospitals and pharmacies; ER facilities; patient datasets
    - (Industry/employment) electronic disclosures; information on the National Pension scheme
    - (Food/Health) ingredients of cosmetic products; foods and nutrients; feed composition
    - (Education) private educational institutions; public education data; school events calendar and lunch menu calendar
    - (Land) building register; land use register; real estate sale prices; public prices of property; demographic distribution
    - (Agriculture/fisheries) final auction prices at the wholesale market
    - (Welfare) disability organizations and facilities; volunteer work and training information

- Most viewed data on the Seoul Open Data Portal
    - Registered population by administrative district category, "dong/gu"
    - Subway map/bus map/information on the number of passengers getting on/off at each station
    - Real time subway arrival information/subway location data
    - Real time air quality/fine dust alerts
    - Information on floating population and businesses

Data and Innovation

Trading data between companies

- Vibrant trading of private sector data through a data exchange or through individual contracts in the US and China.
    - In those countries, the use of personal information is relatively easy.
    - Acxiom has data on over 10,000 attributes of 2.5 billion consumers around the world.
        - Financial services firms and retail firms combine open data with their internal data, and use them for marketing purposes, e.g., micro targeting.

- Trading personal or private sector data is relatively limited in South Korea because of stringent privacy regulation.
    - Popular data in the Data Store (Korea Data Agency or K-Data)
        - Dining code > opening hours, menus and prices of restaurants across the nation
        - BC Card > consumption data by region or by industry
        - KB Card > sales data per customer group according to customer profiles
        - Mobile T-Money > information on mass transit services
        - SK Telecom > data on weekday floating population in Seoul

Data and Innovation

MyData

Background

- Low personal data usage by individuals who are data subjects
- Non-competitive market environment due to information asymmetry
- Incomplete data transmission methods with respect to current account information service

Adoption of policies

- Granted individuals the right to data portability, allowing them to receive their personal data from (financial) institutions and transmit the data to third parties.
- Allowed the data subject to exercise his or her right to personal information in the sectors of healthcare, finance, telecommunications, etc.
  - Financial institutions are required to develop APIs and provide API access.

Pilot projects in major sectors (proposed)

Sector	Project Description	Participants
Medicine/Healthcare] Health management	Allow users to download health checkup results using smartphone health apps, manage health management information in an integrated manner, including the number of steps and heart rate-> real time health management	Hospitals, and mobile phone manufacturers (including five largest hospitals)
Finance] Asset management	Allow users to receive account transaction data and credit card purchase data using open API to get a consolidated picture of financial assets, and receive recommendations for financial products tailored to them -> stable financial planning and investment	Fintech firms, banks, and credit card companies (targeting 1 million customers)
Telecommunications] Service package recommendation	Allow telecom operators to download information on the amount of voice and data usage by their subscribers and recommend customized service packages to them -> help households save their communications spending.	KAIT, and telecom operators (targeting 200,000 customers)

Source: Government, Strategy to Revitalize the Data Industry (June 2018)

Data and Innovation

- EU
  - Adopted the right to data portability.
    - This right allows individuals to request that a data controller transmits the personal data they have provided to the controller in a structured, commonly used, machine-readable and interoperable formant directly to another.
  - Allowed third parties to have read and write access to data through API.
    - Read access enables a third party to look at customer data on account balance, transactions, etc. accumulated in a financial institution.
    - Write access enables a third party on behalf of a customer to send payment instructions to a financial institution.
  - Included account information service (AIS) and payment initiation service (PIS) in the scope of payment services subject to regulation.
- UK
  - Requiring financial institutions to provide APIs for financial product information in addition to read and write access, as in the EU.
  - Boosting the use of a payment accounts comparison service.
- Australia
  - Requiring financial institutions to grant read access and open access to financial product information.
  - Widened the scope of financial products subject to open access.
- US
  - Allowed third parties on behalf of customers to access the personal data of the customers through authoritative interpretation.
  - Encouraging third parties to use API access, not screen scrapping.

Data and Innovation

- South Korea
  - MyData Initiative (July 2018)
    - Included read access.
    - Adopted the right to data portability.
    - Requiring standard API to be used for data transmission between a financial institution and a third party.
    - Introduced a business for personal credit information management.
      - AIS, information account service, data analytics/consulting, investment advisory/discretionary investment service, financial product advice, etc.
    - Proposed the scope of financial products subject to open access
      - (Banks, cooperative banks, savings banks, and insurers) deposit account/credit card transaction data; loan/insurance policy information
      - (Securities companies) information on deposits and withdrawals of investor deposit accounts/CMA, and aggregate amount invested in financial instruments (stocks, investment funds, ELS, etc.)
      - (Telecoms) telecoms billing and payment information

  - Open Banking Initiative (February 2019)
    - Included write access.

Data and Innovation

Data consolidation and use cases

- Provision of aggregate data on real estate listings (e.g., Zillow)
  - Collecting and aggregating public real estate data on the actual sale price of homes, their size and structure, liens, maintenance fees, etc.
  - Providing prospective buyers with aggregate data by merging with internal data on for-sale listings, etc.
- Optimization of night bus routes
  - Designed bus routes using data on the locations and billing address of night time phone calls.
  - Leveraged data on taxi pickup/drop-off locations during night time.
- Prediction of the spread of bird flu
  - Consolidated farm data and vehicle movement data.
- Development of comprehensive air-quality index (CAI)
  - Collection of air quality data from various sources → dimensionality reduction → indexation
- Defeating diseases (Patients Like Me)
  - Social networking site where patients connect and share information with others suffering the same disease(s)
  - Accumulating structured data on diseases and using them in studies to defeat diseases.
  - 0.6 million subscribers with over 2,800 diseases, and 43 million disease data
  - Selling data that patients have uploaded anonymously to pharmaceutical companies and others.

Data and Innovation

Innovation in the financial services industry

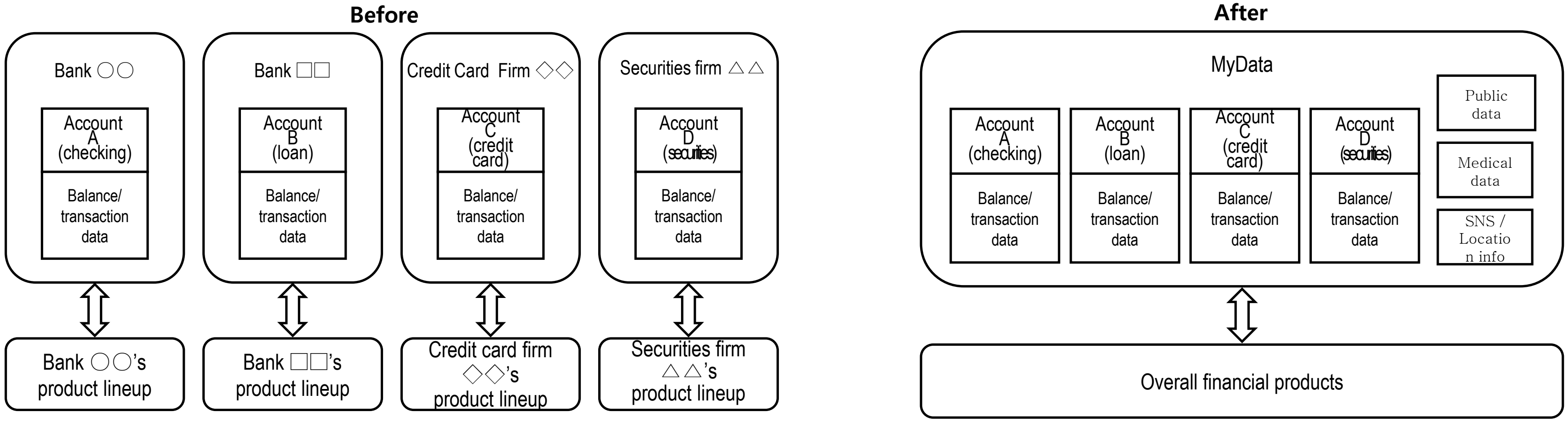
- Creation of financial instruments
  - Machine learning is applied to asset pricing (Gu et al., 2018).
  - New sources of data are collected through image recognition, natural language processing, SNS platforms, etc. and are put to use.
  - Competitive attempts are made to predict asset prices using data and platforms from hedge funds.
  - Capital distribution efficiency and financial stability in the market are expected to improve through sophisticated risk measurement.

- Improvement in credit scoring methods
  - Problems with existing methods
    - Data used in existing quantitative analysis are limited to financial transactions, utilities billing and payment, delinquency records, etc.
    - Low credit scores have been assigned to borrowers who lack the aforementioned records regardless of their actual repayment ability.
  - Assessing and checking the actual repayment ability using big data
    - Machine learning is used to analyze big data (thousands of attributes) in addition to dozens of traditional quantitative indicators.
    - The sentiment, behavior and social relationships of borrowers captured from SNS messages, emails, text messages, etc. are reflected in credit scoring.
    - A great deal of clients with low credit scores have received loans after demonstrating their adequate repayment ability.

Data and Innovation

Assisting consumers in make rational decisions.

- MyData help consumers allocate their financial assets efficiently.
  - Decrease in revolving credit, delinquency or overdraft.
- Offering financial consulting tailored to customers by identifying their consumption patterns, financial position, risk appetite, etc.
  - Customer data is used to find financial instruments whose terms and conditions are a best fit for customers.
  - Recommending low-cost alternatives to the product or service that a customer is now holding or using, with the same benefits the customer could receive.
  - Offering popular products among a group of consumers in the same age or similar income levels



Before

Bank ○○
Account A (checking)
Balance/ transaction data
Bank ○○'s product lineup

Bank □□
Account B (loan)
Balance/ transaction data
Bank □□'s product lineup

Credit Card Firm ◇◇
Account C (credit card)
Balance/ transaction data
Credit card firm ◇◇'s product lineup

Securities firm △△
Account D (securities)
Balance/ transaction data
Securities firm △△'s product lineup

After

MyData
Account A (checking) — Balance/ transaction data
Account B (loan) — Balance/ transaction data
Account C (credit card) — Balance/ transaction data
Account D (securities) — Balance/ transaction data
Public data
Medical data
SNS / Location info
Overall financial products

Data and Innovation

- Enhancing customer experience.
  - Increasing real time services
  - Providing life cycle or location-based services by consolidating data from telecoms, healthcare, and SNS.
    - Data consolidation between the financial services industry and other industries would add large value to the society.
  - Delivering seamless customer experience in alignment with other industries.
    - Financial services have large impacts on customer experience in distribution, airline, hotel industries.

- Data are being used by financial regulators.
  - Data are being leveraged by financial regulators to set priorities among urgent policy tasks.
    - The US Consumer Financial Protection Bureau (CFPB) in collaboration with SAS analyzes customer complaints data, identifies common issues, and sets priorities in addressing problems.

Contents

1 Data Concepts

2 Data Applications

3 Data and Innovation

4 Implications

Implications

- Financial services firms' response*
  - Building the enabling environment for data-based decision-making.
  - Simplifying and automating systems, and reducing operational costs through the adoption of software-as-a-service (SaaS).
  - Providing anytime/anywhere access using Cloud and API.
  - Strengthening data collection and analytical capabilities to understand and identify customer needs.
  - Paying attention to cyber security.
  - Acquiring related workforce and technologies.

* See Six priorities for 2020, PwC (2016)

Implications

Notes

- South Korea's privacy regulations are rigorous and stringent.
  - Tricky de-identification procedures, possible re-identification through consolidation of other data, availability of legal exemptions, etc.

- The infringement of personal rights occurs due to the indiscriminate use of personal information.
  - A surge in unwanted DM, voice phishing, etc.

- If too much weight is imposed on data only to find correlations without considering causal relationships, this would lead to misjudgments and cause side effects.
  - The US police stopped black or Hispanic drivers, predicting the high probability that a black or Hispanic male between the ages of 20 and 27 driving a used car possesses or uses prohibited drugs (Ahn Chunmo, 2017).

Implications

- Big data can be used to tackle social issues and problems.
  - Big data would provide a clue on how to solve social issues in healthcare, transportation, and others.
  - Examples of issues in the financial services sector
    - Conflicts of interest between financial institutions and customers arising from the sale of financial products
    - Retirement plan members' apathy and neglect towards retirement pension assets reaching about KRW 200 trillion
    - Lower investment returns of individual investors resulting from their irrational trading behavior
    - Poor returns from publicly offered funds
  - Data can be used to distill a problem to its essence and address it.