chapter2-augmented-perception

J. Adrián Sánchez

8/5/2017

The Parable of Google Flu: Traps in Big Data Analysis

- Big data has shown the value of these data, we are far from a place where they can supplant more traditional methods or theories.
- **"Big data hubris"** is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.
- The core challenge is that **most big data** that have received popular attention are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis.
- **Warning** that the big data were **overfitting** the small number of cases.
- All empirical research stands on a **foundation of measurement**. Is the instrumentation actually **capturing the theoretical construct of interest?** Is **measurement stable and comparable** across cases and over time? Are measurement **errors systematic?**
- Oddly, GFT bakes in an assumption that relative search volume for certain terms is **statically related to external events,** but search behavior is not just exogenously determined, it is also **endogenously cultivated by the service provider.**
- **Red team dynamics** occur when research subjects (in this case Web searchers) attempt to **manipulate the data-generating process to meet their own goals,** such as economic or political gain.
- **Transparency and Replicability.** Even if one had access to all of Google's data, it would be impossible to replicate the analyses of the original paper from the information provided regarding the analysis.
- Researchers need a better understanding of how these changes occur over time. Scientists need to replicate findings using these data sources across time and using other data sources to ensure that they are observing robust patterns and not evanescent trends.
- **Science is a cumulative endeavor,** and to stand on the shoulders of giants requires that scientists be able to continually assess work on which they are building.
- However, traditional **"small data"** often offer information that is not contained (or containable) in big data, and the very factors that have enabled big data are enabling more traditional data collection.
- **Use Big Data to Understand the Unknown**
- **Study the Algorithm**

The Digitization of Just About Everything

- **Sensors** -> The software's genius is to turn all the smartphones running it into 'sensors' that **upload constantly to the company's servers** their location and speed **information.**
- Call a **network effect** —a situation where the value of a resource for each of its users increases with each additional user.
- **The second machine age**

1. **Exponential improvement in computer gear**

2. **Digitization** - > this phenomenon as **"encoding information as a stream of bits."** Digitization, in other words, is the work of **turning all kinds of information** and media—text, sounds, photos, video, data **from instruments and sensors,** and so on **into the ones and zeroes** that are the native language of computers and their kin. -> **CONSEQUENCES**-> New ways of acquiring **knowledge** and **higher rates of innovation.**

Unique economic properties of digital information: such information is **non-rival**, and it has close to **zero marginal cost of reproduction.**

A. We humans like to **share and interact** -> **User-Generated Content**

B. Data Explosion -> **digitization yields truly Big Data.** ->

Are all of these exa- and zettabytes of digital data actually useful? It increases understanding by making huge amounts of data readily accessible, and data are the lifeblood of science.

3. **The style of innovation**

5 Ways Product Design Needs to Evolve for the Internet of Things

- What does **product design** mean when products are **no longer just physical goods?**
- When **machines are both connected and intelligent**, they become a **hybrid of products and services**.
- How their products will obtain the data they need to operate and how the **data their products create will be used for other product-service hybrids.**
- They must make **using the data services** embedded in smart products a **satisfying and even gratifying experience.**
- designing for usability, simplicity, quality, innovation and manufacturability all remain valid.
- Design for usability includes design for **"upgradability."**
- Design for simplicity includes design for **saving time**
- Design for quality includes **design for communication.** -> how the data created by their products will be consumed in other systems?
- Design for product innovation includes **design for discovery**. -> They can treat hybrids as R&D laboratories, capturing how buyers use services and testing features to uncover which ones customers find most valuable.
- Design for manufacturability includes design for **creating insight.** -> to create products that can make and carry out better decisions.

Making Advanced Analytics Work for You

- Powerful new business models that derive from an ability to exploit data.
- To capitalize on them, organizations would have to **make complex process changes** and **build employees' skills.**
- How companies and frontline **managers actually made decisions**, and **new demands for data management added complexity to operations.**

1. Companies must be able to identify, combine, and manage multiple sources of data.
2. They need the capability to build advanced analytics models for predicting and optimizing outcomes.
3. Management must possess the muscle to transform the organization so that the data and models actually yield better decisions.

A clear strategy for how to use data and analytics to compete, and **deployment of the right technology architecture** and **capabilities**.

- Sufficient time and energy in aligning managers across the organization in support of the mission.

The ability to see what was previously invisible improves operations, customer experiences, and strategy. Get **creative about the potential of external and new sources of data.** What decisions could we make if we had all the information we need?

Existing IT architecture may prevent the integration of siloed information, and managing unstructured data often remains beyond traditional IT capabilities. Many legacy systems were built to deliver data in batches, so they can't furnish continuous flows of information for real-time decisions. Connecting the most important data for use in analytics, followed by a cleanup operation to synchronize and merge overlapping data and then to work around missing information.

- Analytics models that allow managers to predict and optimize outcomes. With identifying the business opportunity and determining how the model can improve performance. **A pure data-mining approach** often leads to an **endless search** for what the data really say. What's the least complex model that would improve our performance?
- Create **simple visual interface** that highlighted projected workforce needs and necessary actions. Ultimately, that approach of using a simple tool to deliver complex analytics substantially improved workforce planning and reduced the need for new hires and overtime.
- Most organizations will need to **upgrade their analytical skills and literacy.** Adjusting culture and mind-sets typically requires a multifaceted approach that includes training, role modeling by leaders, and incentives and metrics to reinforce behavior.

Executives should concentrate on targeted efforts to **source data, build models, and transform the organizational culture.**

GE and the Industrial Internet

Sell outcomes-based services based on analytics and software? GE's **Industrial Internet initiativ**e proposed an **open, global network that connected machines, data, and people**, and provided **data synthesis and analysis** allowing for **real-time and predictive solutions** to optimize the complex operations of GE's varied customer base, including **predicting maintenance and repair needs and informing performance and operational decisions.**

There are new commercial- and consumption-based business models emerging. On the commercial side, there's the Industrial Internet as a service, which could include a range of things including remote monitoring and diagnostics, information services, platform-as-a-service (PAAS), or data management.

- new software headquarters;
- the launch of a common technology platform across GE's diverse industrial businesses;
- a thorough assessment across the organization of GE's software development expertise and readiness evaluation of its sales talent capable of supporting this new direction;
- and new and expanded partnerships with companies

New business model - deep familiarity and expertise with the specific customer and its sector - customize how it partnered with the customer and sold the software-enabled - allow significant access to internal operational data and some kind of value/revenue/profit-sharing arrangement - Diagnosing customer needs, establishing value, understanding consumption models, and applying provocation-based selling are very important.

- We have new **non-traditional 'competitors'** starting to approach our long-standing customers. IBM mostly, but SAP and big data start-ups are telling our customers they can **provide these analytics and services**.

Proposition - Software as a service represents a whole new beast for GE." GE had to identify and develop new opportunities, source and hire developer and sales expertise and talent, build the offerings, and price and coordinate sales of each offering through each division's sales mechanisms.

- Pursuing a comprehensive services business model that offered a range of **ancillary services** to its industrial products, **tying the customer** more closely to GE.
- Invest in the technology needed to **gather, process, and use the information generated by its machines**.

Build an Industrial Internet: GE Software

- Someone with a background in innovative software development; someone who was service oriented versus product development; and someone able to manage a start-up–like environment in a very large, complex company- -> **Developer talent** had to be addressed.
- Put a single, common platform in place, then be sure everyone was able to build on it.
- Connecting machines and sharing data raised questions of **security and compliance,** especially as regulations varied widely across the U.S., Europe, and the rest of the world. GE's **experiences with customers' data could lead to efficiencies and optimization**; they could also offer opportunities to **learn more about a customer's pain point or an industry's inefficiency.** But, the **intellectual property** of those discoveries belonged to whom?

Taking an Industrial Internet to Market

- Build sales capabilities to match the initiative's offerings, providing presales resources, an enablement engine, and postsales support.

What to engage?

- The easiest and quickest **beneficiaries have been GE businesses with minimal existing software-based services in their portfolio and, as a result, had fewer software development organizations with a vested interest in legacy technology.**

Building out the Ecosystem

- We partner with competitors.
- Start-ups also provided opportunities: venture investing aimed at accelerating early research in businesses that we'd suffocate if we brought [them] into GE or in businesses we don't necessarily want to own.