From big data to useful knowledge: challenges and opportunities in smart factories

Grzegorz J. Nalepa and Edyta Brzychczy

AGH University of Science and Technology

CHIST-ERA/IBDSF, Kraków, Poland, June 22, 2017

home.agh.edu.pl/gjn, geist.re, bigdata.agh.edu.pl
Outline

1. Smart Factories
2. Learning
3. Semantics
4. Context
5. Processes and Marketing
6. Outlook
Photo from Shutterstock
Factories
Learning
Semantics
Context
Process
Outro

1st
Mechanization, water power, steam power

2nd
Mass production, assembly line, electricity

3rd
Computer and automation

4th
Cyber Physical Systems

Photo from Wikipedia
Photo by ERIC LALMAND/AFP/Getty Images
Topics for this talk

1. learning
2. semantics
3. context
4. processes
Outline

1. Smart Factories
2. Learning
3. Semantics
4. Context
5. Processes and Marketing
6. Outlook
Big Data may be Huge Problem

Big Data

- we can record anything
- unlimited storage
- quite fast retrieval
- get it anywhere – cloud
- we can know more! (?)
- we can do more! (?)

Challenges

- *data* storage is not a solution
- distribution is nice but…
- goodbye ACID, welcome CAP
- no relationships, no *knowledge*
- statistics is not enough!
- to understand, we need to know what (how) questions to ask!
Machine Learning (and data mining)

Benefits
- automated training
- concept learning
- object classification
- pattern recognition
- (new!) knowledge discovery
- massive parallelism (use of)
- incremental models (storage!)

Methods
- learning from examples
- unsupervised learning
- clusterization
- deep learning
- learning from data streams
Outline

1. Smart Factories
2. Learning
3. Semantics
4. Context
5. Processes and Marketing
6. Outlook
50B CONNECTED THINGS BY 2020
### Internet of Things! (or World Wide Wide Mess?)

#### IoT
- *everything* gets connected
- accessible from *anywhere*
- sensor networks
- M2M communication
- „smart” X (house, car, city)
- *panopticon*-like monitoring

#### Challenges
- huge data! maybe our *storage* is not unlimited after all?
- *protocols* (opportunistic)
- energy efficiency
- can machines *understand* each other?
- can they communicate with *our concepts*?
Metadata, Ontologies, Semantics

**Methods**
- symbolic knowledge *representation*
- data vs. *knowledge*
- numbers vs. *concepts*
- automated *reasoning*
- inference with *rules*
- *Semantic* (social) Web

**Benefits**
- metadata – data interpretation
- processing with human concepts
- *shared* conceptualizations
- (semantically) Linked Open Data
- semantic *interoperability*
- *query*, not search

http://geovation.uk/linked-data-opportunity-for-developers

http://business.data.gov.uk/companies
Outline

1. Smart Factories
2. Learning
3. Semantics
4. Context
5. Processes and Marketing
6. Outlook
Data in dynamic environments is often uncertain and ambiguous.

Uncertainty sources:

- **Epistemic**
  - Lack of knowledge
  - Semantic ambiguity

- **Aleatoric**
  - Machine imprecision
  - Inherent randomness
Context-Aware Systems

context = any information that can be used to characterize the situation of an entity

**Hardware**
- mobile, wearable
- versatile sensors
- environment monitoring
- new interfaces
- ambient intelligence

**Software**
- *adaptation* of functionality
- learning of user *habits*
- reasoning with concepts
- *recommendation* (relevance)
- decision support
- *mediation* of context
- continuous *assistance*
Outline

1. Smart Factories
2. Learning
3. Semantics
4. Context
5. Processes and Marketing
6. Outlook
“Action is the foundational key to all success.”

- Pablo Picasso

http://www.flickr.com/photos/19387816@N00/196455538/
WHY?
Process (How?)

- What?
- Who?
- When?
- Where?
BPM Cycle

- Identification of processes
- Process modelling
- Process analysis
- Process redesign
- Implementation
- Monitoring and process control

© G.J. Nalepa + E. Brzychczy (AGH)
Process mining vs. mining process

Source: Korski. J, FAMUR S.A.
Process monitoring

Source: Korski. J, FAMUR S.A.
How about real process mining?
Process mining

http://prompt.processmining.it/static-images/pmini.jpeg
Smart factories need deep process analysis in process-oriented way
Gartner technology hype cycle (2016)

Expectations
- Connected Home
- Blockchain
- Smart Robots
- Micro Data Centers
- Gesture Control Devices
- IoT Platform
- Commercial UAVs (Drones)
- Affective Computing
- Smart Data Discovery
- Virtual Personal Assistants
- Brain-Computer Interface
- Conversational User Interfaces
- Volumetric Displays
- Smart Workspace
- Personal Analytics
- Quantum Computing
- Data Broker PaaS (dbPaaS)
- Neuromorphic Hardware
- Context Brokering
- General-Purpose Machine Intelligence
- 4D Printing
- Smart Dust

Innovation Trigger
- Cognitive Expert Advisors
- Machine Learning
- Software-Defined Security
- Autonomous Vehicles
- Nanotube Electronics
- Software-Defined Anything (SDx)

Peak of Inflated Expectations
- Natural-Language Question Answering
- Enterprise Taxonomy and Ontology Management

Trough of Disillusionment

Augmented Reality

Slope of Enlightenment

Virtual Reality

Plateau of Productivity

Years to mainstream adoption:
- less than 2 years
- 2 to 5 years
- 5 to 10 years
- more than 10 years
- obsolete
- before plateau

Source: Gartner (July 2016)
Take away messages

1. Machine learning (incremental) needed for BD
2. Semantics needed for interoperability in IoT
3. Context and uncertainty handling for reasoning
4. Process mining and management for analytics

Thank you for your attention!
Do you have any questions?

home.agh.edu.pl/gjn, geist.re, bigdata.agh.edu.pl