Formalizing expert knowledge through machine learning

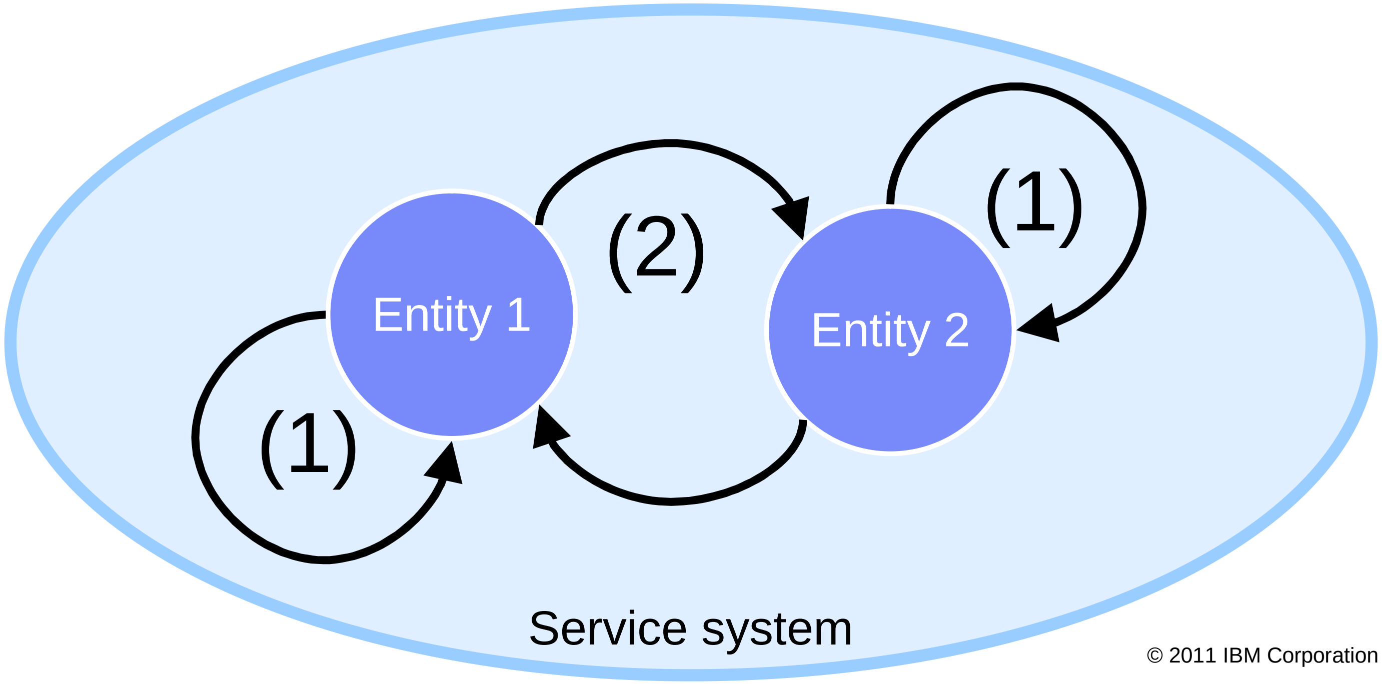
Tsuyoshi Idé
IBM Research - Tokyo

Contents

- □**Introduction**
  - –Why this is an important task
  - –How traditional approaches failed
- □**Case study: condition-based maintenance in the rail industry**
- □**Summary**

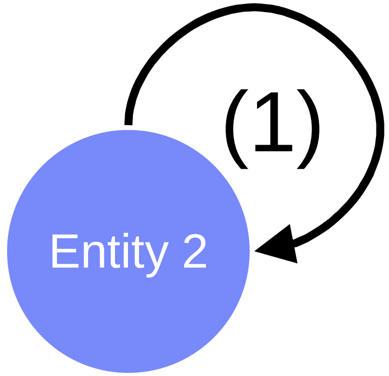
Value co-creation is the key concept of service systems

- **Two basic steps in value co-creation:**
- **(1) transformation**
  - Transform information by adding a value typically based on expert knowledge
- **(2) transfer**
  - Transfer the information to trigger further transformation



Formalizing expert knowledge is a key task in service science

- **The transformation step is a major source of vale creation**
- **This is usually triggered by certain expert knowledge**



- **Interesting question:**
  - –What experts are doing in this step?

> - **What kind of language is appropriate for knowledge representation?**
> - **How can we construct useful rules from experience?**

Contents

□**Introduction**

- Why this is an important task
- How traditional approaches failed

□**Case study: condition-based maintenance in the rail industry**

□**Summary**

Expert systems: failure and success
The failure of MYCIN

- **MYCIN (1970s)**
  - –Medical expert system developed in Stanford
  - –Used a large repository of IF-THEN rules
  - –Seemingly good results in academic benchmarks

  - –Never been used in practice

- **Issue: knowledge acquisition bottleneck**
  - –"Who prepares the complete knowledge base?"

Expert systems: failure and success
The victory of DeepQA

- **DeepQA (2011)**
  - Beat human quiz champions
  - Capable of handling open-domain questions
    - i.e. handles an infinite number of queries

- **New technologies?**
  - Rely on digitalized encyclopedia data like Wikipedia
  - Used machine learning to make a decision

Lessens learned from the history of expert system

□ **(Hopefully automatically) capturing expert knowledge is essential in practice**

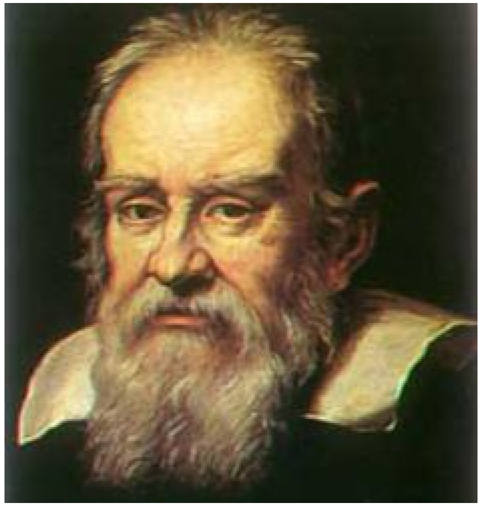
□ **The extracted decision rules must satisfy at least three criteria:**

- **Generalizeability**
  - Must be capable of handling unseen situations
- **Learnability**
  - Must capture the decision patterns automatically
- **Actionability**
  - Must provide insights understandable to humans

(For ref.) Mathematics is a powerful language. Galileo's comment may apply to service systems

- **"Philosophy is written in this grand book, the universe...."**
- **"It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures;...."**

"The Assayer", Galileo Galilei, 1623.



- **This may be true in service systems!**

Contents

- **Introduction**
  - Why this is an important task
  - How traditional approaches failed
- **Case study: condition-based maintenance in the rail industry**
- **Summary**

Wheel axle anomaly detection is a major topic in condition-based monitoring

This work

TSI
(Technical Specifications for Interoperability)

Axle bearings
(axle, gear, etc.)

Wheels/Tracks
(impact load, profile, hot/cold wheels, etc.)

- Temperature
- Sound/acoustic
- Strain
- Optical
- Acceleration
- Image
- ….

* TSI: European specification for high-speed trains

Monitoring temperature is a common approach, but anomaly detection is known to be very hard

- **Temperatures are very much dependent on external disturbances**
  - Weather: rainfall, wind, sunlight
  - Train speed, braking, …
  - Equipment configuration

- **Example: two wheel axles in different cars**
  - Dependencies on Car # and rainfall are clear

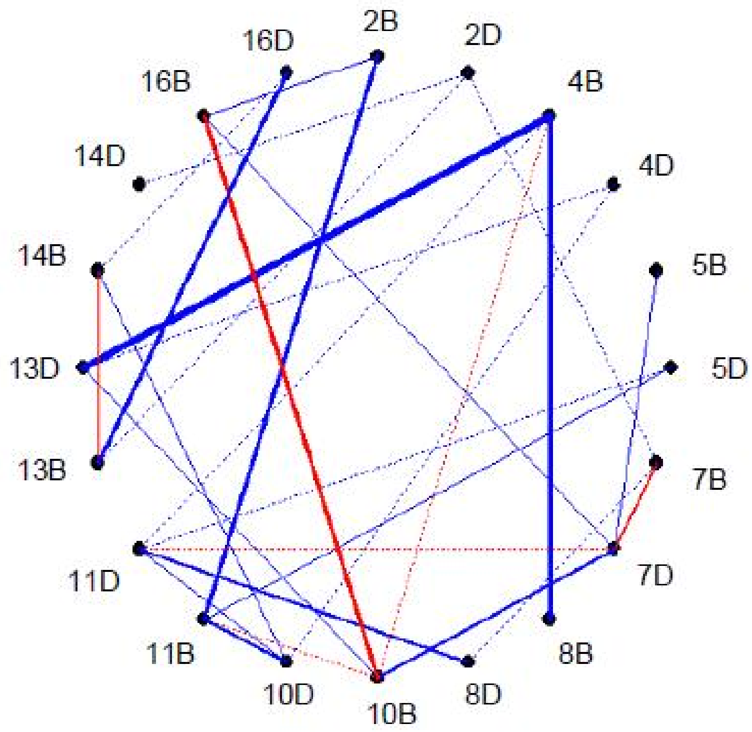
Technical challenges in hot box detection

- **How to eliminate the effect of climate**
  - Rainfall, wind, …
- **How to handle temperature differences in car positions**
  - Different cars may give different temperatures
- **How to handle temperature differences in axle positions**
  - Even in the same car, different axles may give different temperatures

Basic idea: relative comparison among dependent axles

- **Step 1: Discover the dependency between axles**
  - Dependency is automatically identified using a machine learning technique

- **Step 2: Perform relative comparison with dependent axles**
  - "Comparison" is mathematically performed in a probabilistic fashion



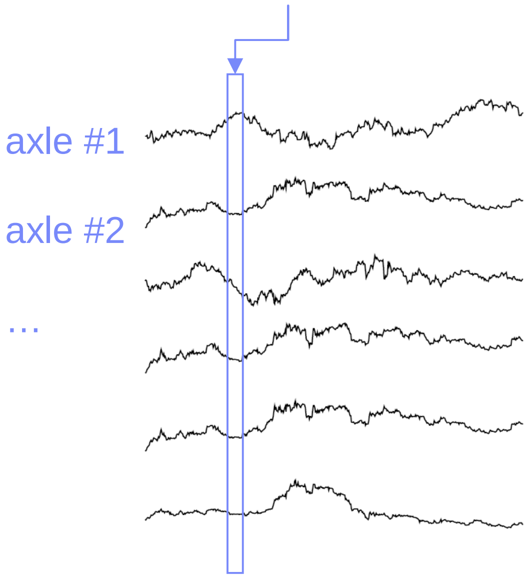
Note: Apart from mathematical expression, this approach shares the original idea of relative comparison with expert engineers.

Step1: Dependency discovery for anomaly detection

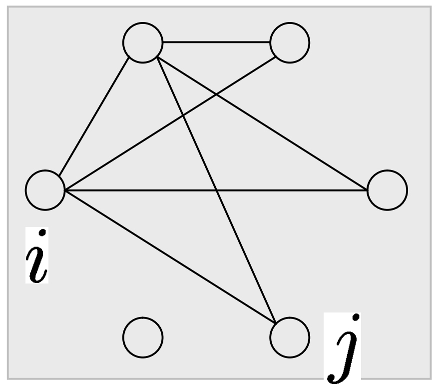
* IBM Anomaly Analyzer for Correlational Data

ANACONDA

- **The ANACONDA algorithm uses a *sparse structure learning* technique, which automatically finds a hidden dependency between variables**
  - Dependencies are identified based only on the previous recordings
  - Detailed knowledge of the system is not used

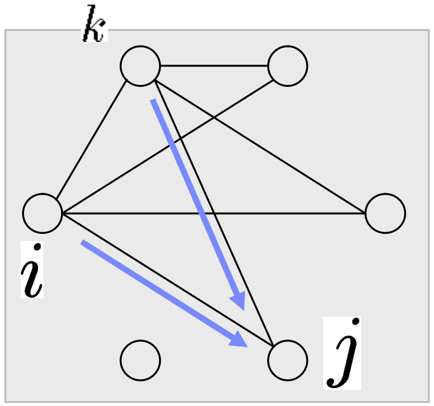


Dependency between variables



“Axle *i* is dependent on *j* ”

Step 2: How much does a temperature deviate from its expected value, given dependent variables?



Example:

- j-th axle is dependent on axles i and k
- The j-th temp. should be predicted by *i* and *k*
- Negative conditional log likelihood - $p(x_j \mid x_i, x_k)$ gives a measure of how much x_j deviates from its expectation

(anomaly score of the j^{th} variable) = - log $p(x_j \mid x_i, x_k)$

Conditional probability density function of the graphical Gaussian model

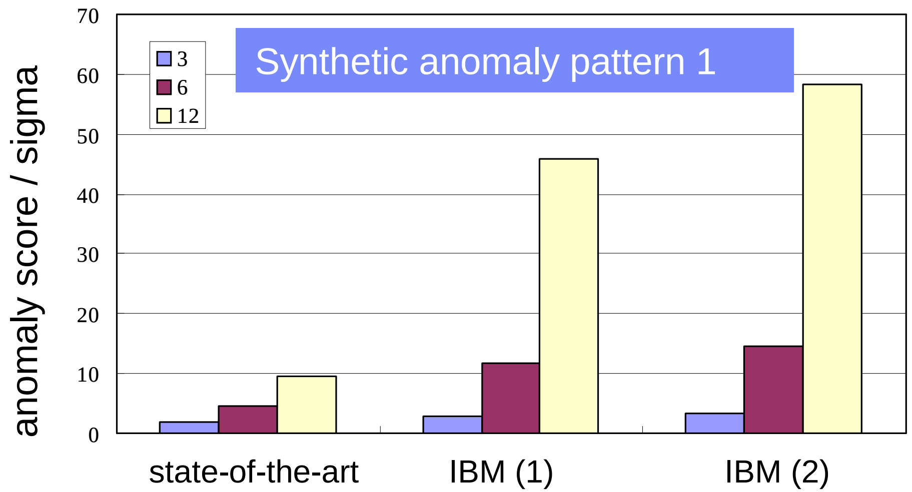
Wheel axle and gear box temperature data

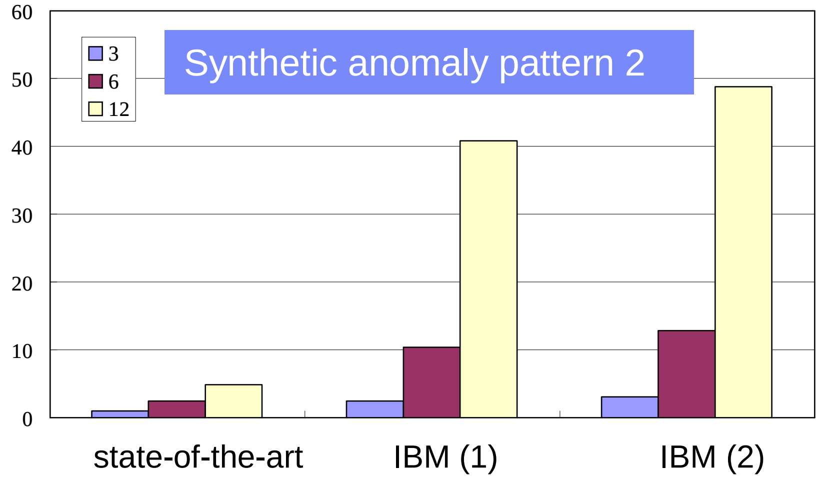
- **We are given a set of about 100 dimensional temperature vectors**
  - Typically measured using a wayside hot box detector
  - Each temperature vector is a recording at a single detector

Result: Our method showed much better detectability of known anomalies

- **Compared with a state-of-the-art method**
  - It is based on hard-coded expert knowledge
- **Performance measure: higher is better**

  (mean anomaly score of anomalous samples) / (std. dev. of anomaly scores of normal samples)
- **Results with synthetic as well as real anomalies clearly shows better performance of our method**
  - About one order of magnitude better





Contents

Summary

- **Formalizing expert knowledge is a key problem in service science**
- **The use of mathematics is a natural way for knowledge representation**
- **Machine learning is a systematic method for rule discovery**
- **As an example, we construct a rule for anomaly detection to encode expert knowledge in the rail industry**

Thank you!

Example: outsourced maintenance of high-speed trains

- **Entity 1: train operator**
  - Provides technical information
  - Receive a guarantee of safety
- **Entity 2: maintenance company**
  - Has expert engineers perform inspection
  - Trivially observed quantities are transformed based on expert knowledge

Mathematical and probabilistic representation is important for generalizability

□**General knowledge representation: IF-THEN rule**

– IF (predicate) THEN (consequent) ELSE (alternative) END IF

□**Our claim:**

- **Natural language is not a good starting point to represent the predicates and consequents**
- **One "natural" representation of a rule looks like:**

$$y = f(x|\mathcal{D})$$

Decision variable

observables

previous data

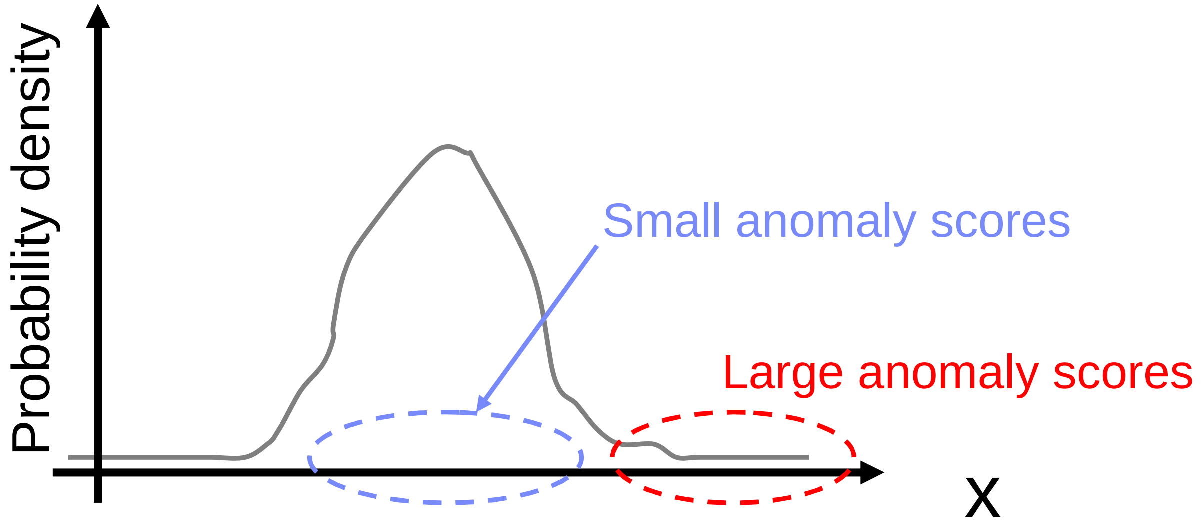
Example: anomaly detection of wheel axles

- **Decision variable y: anomaly score of each axle**
  - Representing how much anomalous an axle is
- **Observables x: temperatures of wheel axle boxes**
- **Data *D*:**
  - A set of previous measurements on the temperature under normal and abnormal conditions

Probabilistic approach is useful to build the rule

- **General strategy to build the rule $f(x|\mathcal{D})$ is to use probability distributions of the data *D***

- **Example: anomaly detection**
  - The anomaly score $f(x|\mathcal{D})$ can be defined based on the probability density of x given the data D



Machine learning give a systematic way to constructing mathematical rules

- **(Statistical) machine learning is based on probabilistic distribution**
  - e.g. $p(x|\mathcal{D})$ in the anomaly detection example

- **Machine learning is data-driven**
  - Decision functions are defined using the probability functions, which is identified in a data-driven fashion

•Machine learning, which is data-driven in nature, is a useful framework for rule discovery

Service Science Research Forum (May 10, 2012) Program agenda

- **14:00-14:05 Opening and Welcome**
  - Spohrer and Sawatani, Book editor
- **Session Chair Dr. Uchihira**
- **14:05-14:20 TBD**
  - Arai, Shibaura Institute of Technology
- **14:20-14:35 Community based participatory service engineering: case studies and technologies**
  - Motomura, National Institute of Advanced Industrial Science and Technology (AIST)
- **14:35-14:50 Methodology of Workshop-Based Innovative System Design based on Systems engineering and design thinking**
  - Yasui, Keio University
- **14:50-15:05 Human Behavior Observation for Service Science**
  - Matsunami, Osaka Gas CO. Ltd.
- **15:05-15:20 Value Co-Creation Process and Value Orchestration Platform**
  - Kijima, Tokyo Institute of Technology
- **15:20-15:35 Service design in tourism: Encouraging a cooperative relationship between professional design and non-professional design**
  - Hara, The University of Tokyo
- **15:35-16:00 Break**
- **Session Chair Dr. Hara**
- **16:00-16:15 Temporal-Spatial Communication for Nursing and Caregiving**
  - Uchihira, Toshiba Corp.
- **16:15-16:30 Formalizing expert knowledge through machine learning**
  - Ide, IBM Japan, Ltd.
- **16:30-17:10 Overall and Latest update**
  - Jim Spohrer, Book Editor
- **17:10-17:30 Discussion/Closing**