Markus Stocker

SFERC International Collaboration • August 23, 2012 • Kuopio, Finland

Acquisition and representation of knowledge about sensor observations

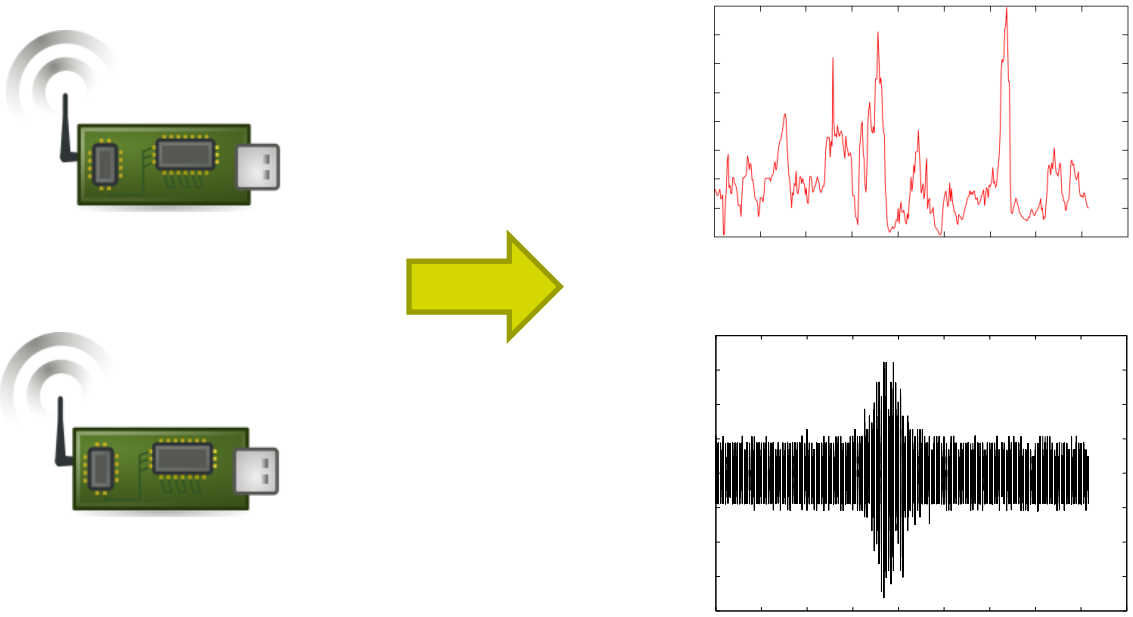
UNIVERSITY OF
EASTERN FINLAND

Overview

- For environmental monitoring, discuss
  - How to let computers represent knowledge acquired from sensor data
  - Knowledge about observations made by a sensor network
- We look into
  - Acquisition of sensor data in environmental monitoring
  - Processing of sensor data into data patterns
  - Classification of patterns and acquisition of knowledge
  - Representation of knowledge and automated inference
- Discussed for two use cases
  - Classification of vehicles observed in vibration sensor data
  - Representation of algal bloom observations in Lake Taihu

Introduction
Environmental monitoring

Environmental monitoring
Sensor data

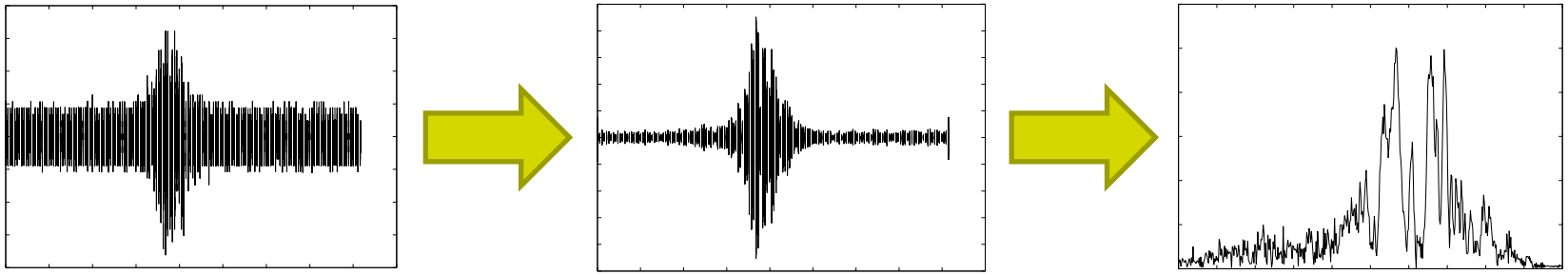


Sensor data
Tabular representation

Time	Dim 1	…	Dim N
2012-08-22 10:00:00	0.00512546	…	…
2012-08-22 10:00:01	0.00635674	…	…
2012-08-22 10:00:02	0.03467845	…	…
2012-08-22 10:00:03	0.13453564	…	…
2012-08-22 10:00:04	0.00004566	…	…
2012-08-22 10:00:05	0.00034563	…	…
2012-08-22 10:00:06	0.23045677	…	…
2012-08-22 10:00:07	0.34506677	…	…
2012-08-22 10:00:08	0.03402594	…	…
…	…	…	…

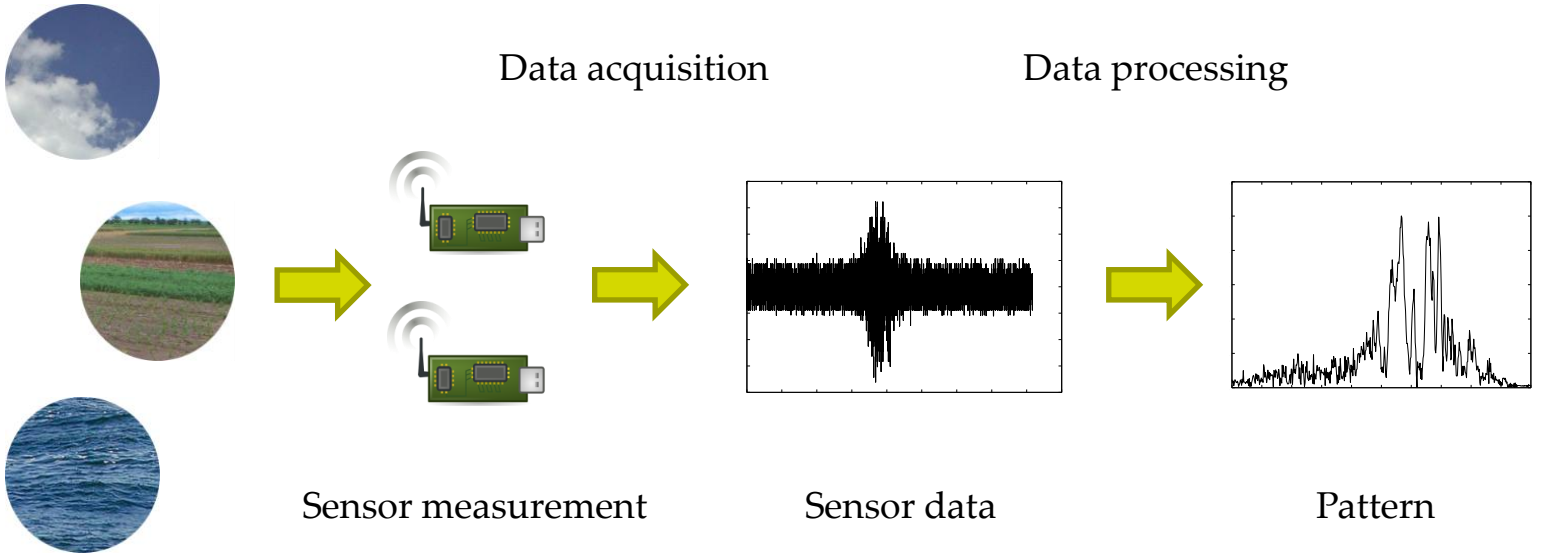
Environmental monitoring
Data processing

Bandpass filter



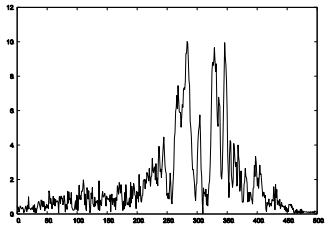
Fourier transform

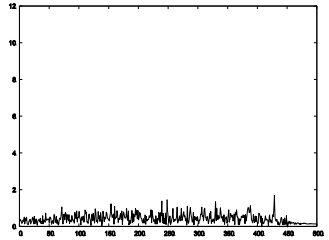
Environmental monitoring Recap

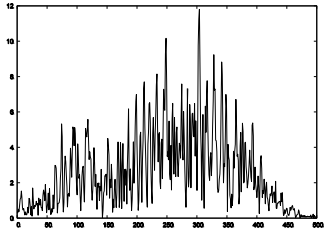


Properties of objects or events of the real world

Introduction
Great, and what next?



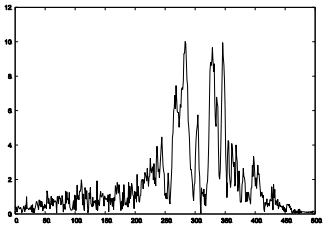


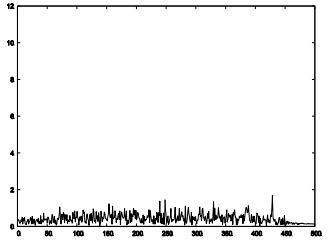


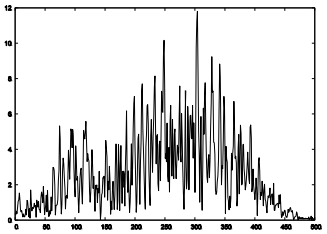
Problems and challenges

- What do those patterns actually mean
- How to deal with high-volume data
- How to continuously process data
- How to efficiently process data
- How to store and retrieve such data

Knowledge acquisition
Machine learning







Machine learning classification task (supervised)

Light vehicle

No vehicle

Heavy vehicle

Supervised learning
A very brief intro

Time	Dim 1	…	Dim N	Label	Pred.
2012-08-30 10:00:00	0.00512546	…	…	Lv	
2012-08-30 10:00:01	0.00635674	…	…	Lv	
2012-08-30 10:00:02	0.03467845	…	…	Hv	
2012-08-30 10:00:03	0.13453564	…	…	Lv	
2012-08-30 10:00:04	0.00004566	…	…	Hv	
2012-08-30 10:00:05	0.00034563	…	…	Hv	
2012-08-30 10:00:06	0.23045677	…	…	Lv	Lv
2012-08-30 10:00:07	0.34506677	…	…	Hv	Hv
2012-08-30 10:00:08	0.03402594	…	…	Lv	Hv
…	…	…	…	…	…

Training

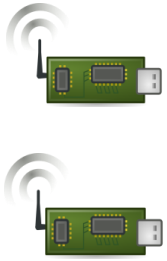
Test/Validation

Knowledge acquisition Recap

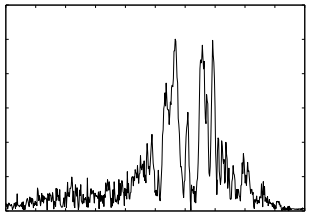
Properties of objects or events of the real world

Data processing

Machine learning



Sensor measurement



Pattern

Light vehicle

Observation

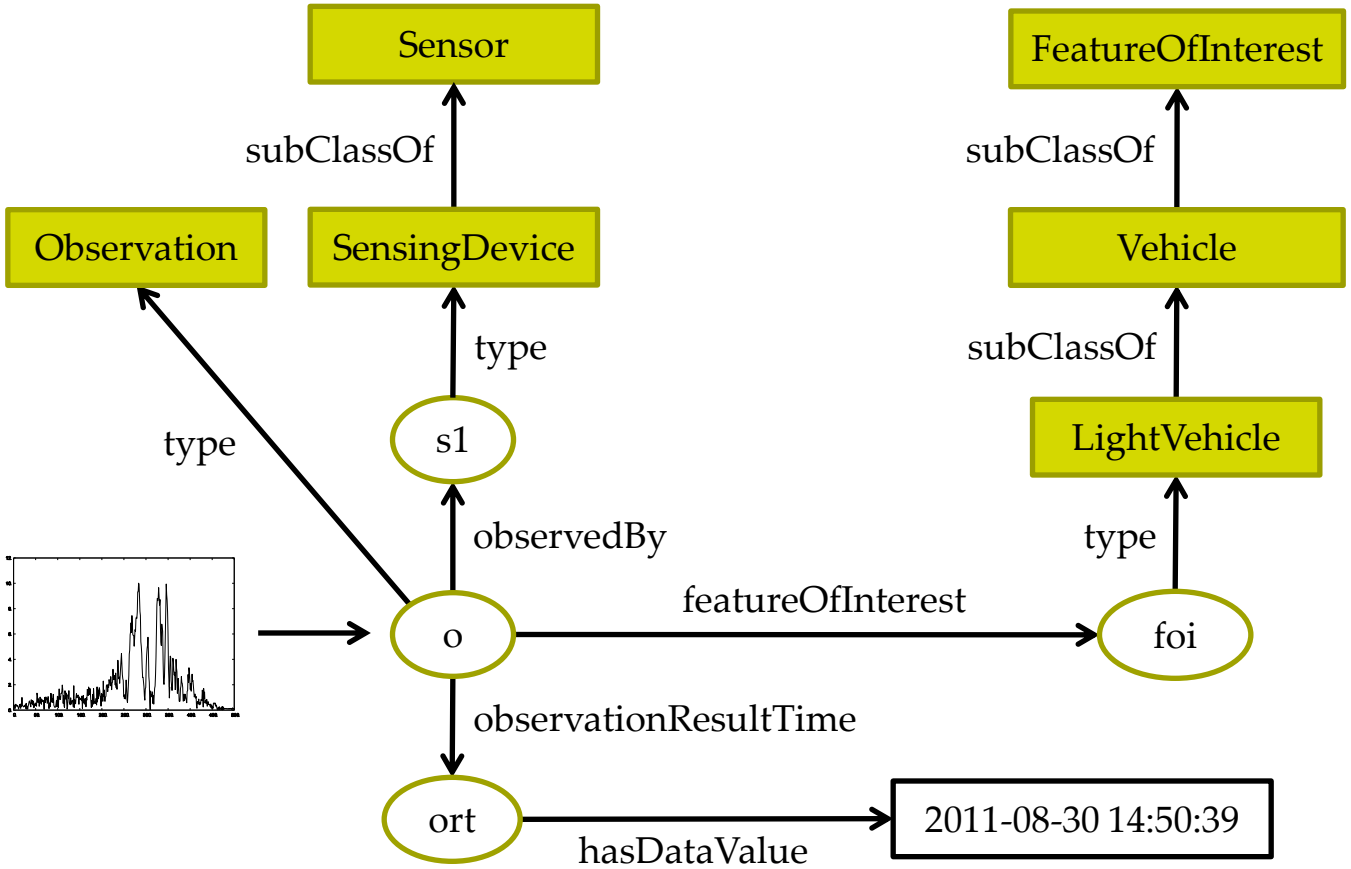
Knowledge acquisition
Great, and what next?

- We now know how to acquire abstract knowledge about observations
- It is a convenient method because
  - Computer does it automatically
  - Computer can do it continuously
  - Often computationally not too expensive
- Though there are downsides
  - Supervised learning is expensive
  - Domain specificity
- Now that the computer tells us “there is a light vehicle”
  - What can we do with this piece of information?
  - We let the computer represent it

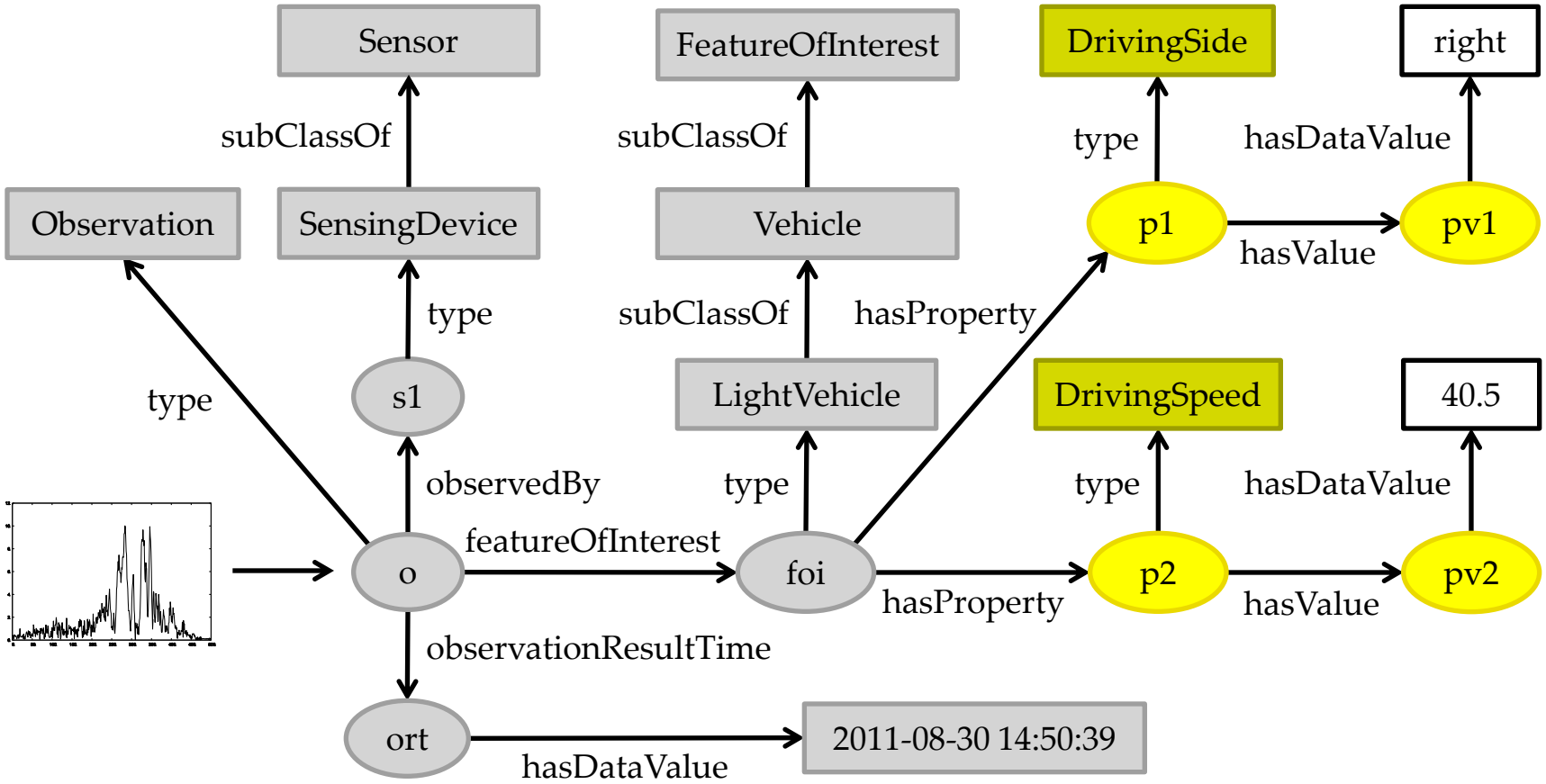
Knowledge representation
A very brief intro

- Looking at a street and seeing a vehicle we can describe it
  - For instance, “it is a heavy vehicle”
  - The vehicle is “on the right side of the road”
  - And we may know the time
  - And we might guess the speed
- A computer can do something similar
  - It already “knows” that
    - There is a vehicle observed by sensor S
    - And that it is a heavy vehicle
    - And what time it is
  - Given two sensors, it can
    - Infer the vehicle velocity

Knowledge representation
Observation



Knowledge representation
Automated inference

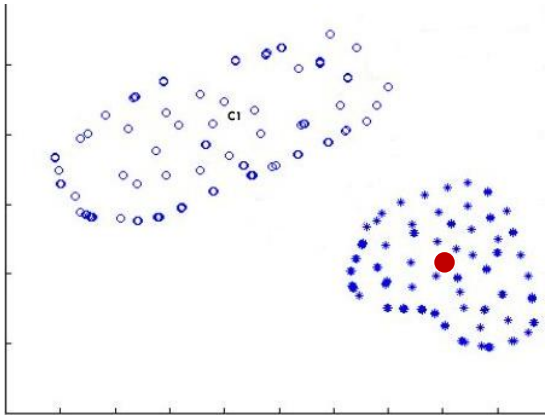


Putting it all together
Algal bloom observations in Lake Taihu

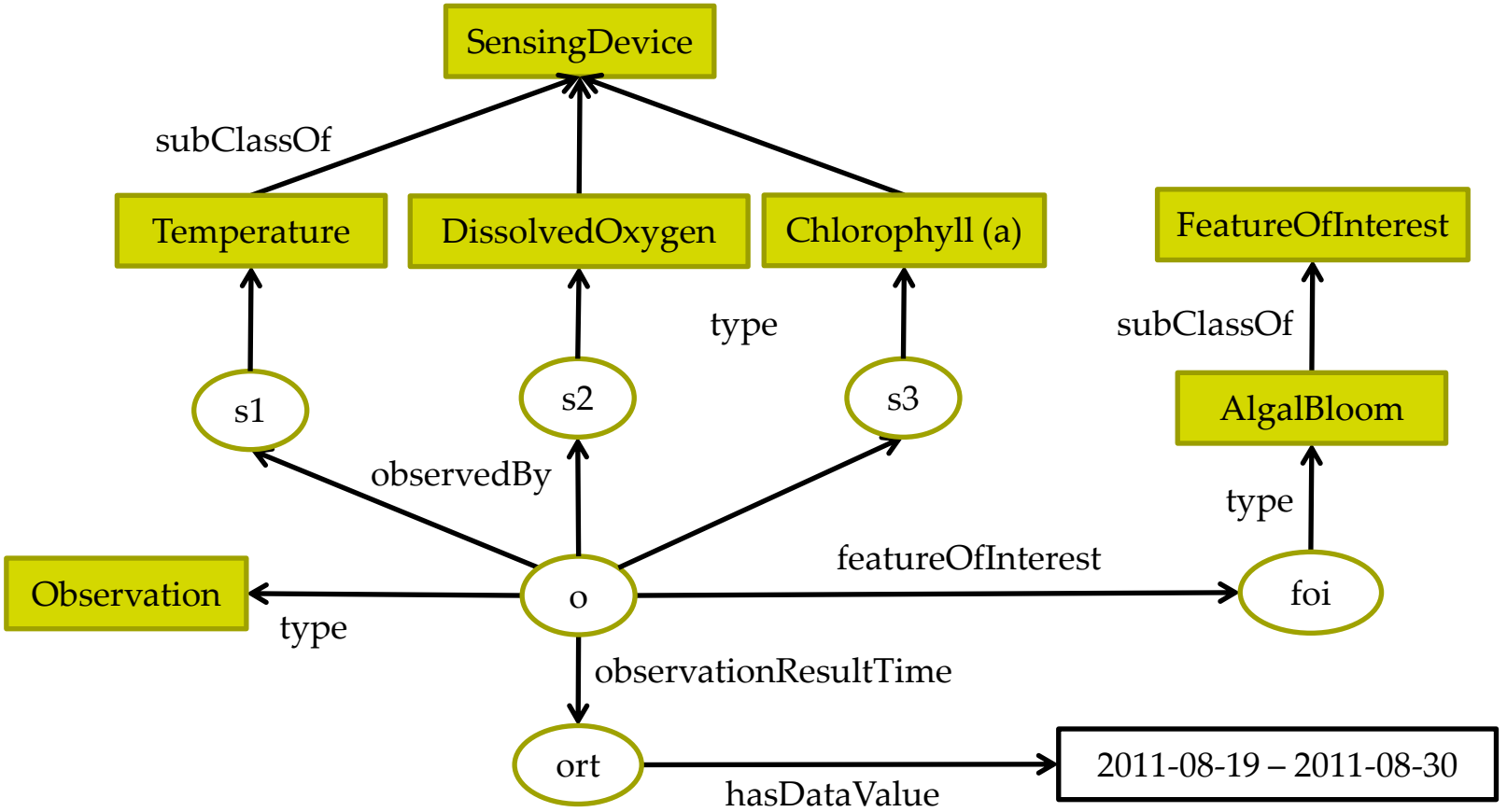
- Given
  - Sensor monitoring data
    - EMB01 station located in the Meilian bay of Lake Taihu, China
  - Sensor data for
    - Water temperature
    - Dissolved oxygen
    - Chlorophyll (a)
- Aim
  - Represent knowledge about observations for algal blooms

Algal bloom observations in Lake Taihu Knowledge acquisition

- We have a 3-dimensional dataset
  - Water temperature, dissolved oxygen, chlorophyll (a)
- As well as measurement values over 3 years
- Use clustering to learn a prototype algal bloom observation
  - From past algal bloom observations
- For the sake of argument, assume
  - Lower cluster is for algal bloom observations
  - Hence red dot is the prototype observation
- Classify future observations
  - According to similarity with prototype



Algal bloom observations in Lake Taihu Knowledge representation



Conclusions

- Discussed one type of environmental monitoring
  - Automated using sensor networks
- Looked at the data produced by such networks, time series
- Computational methods to process such data
  - To acquire interesting patterns
- Computational methods to classify patterns
  - To acquire conceptual knowledge
- Computational methods to represent knowledge
  - Pros of automated inference, abstraction, query, integration, sharing
- Discussed two use cases