On-line Fusion of Functional Knowledge Within Distributed Sensor Networks

Dominik Fisch, Bernhard Sick

Intelligent Embedded Systems Group
University of Kassel
www.ies-research.de

Final Colloquium of the DFG Priority Program 1183
"Organic Computing"
September 15./16. 2011
Nuremberg

Intelligent
Embedded Systems

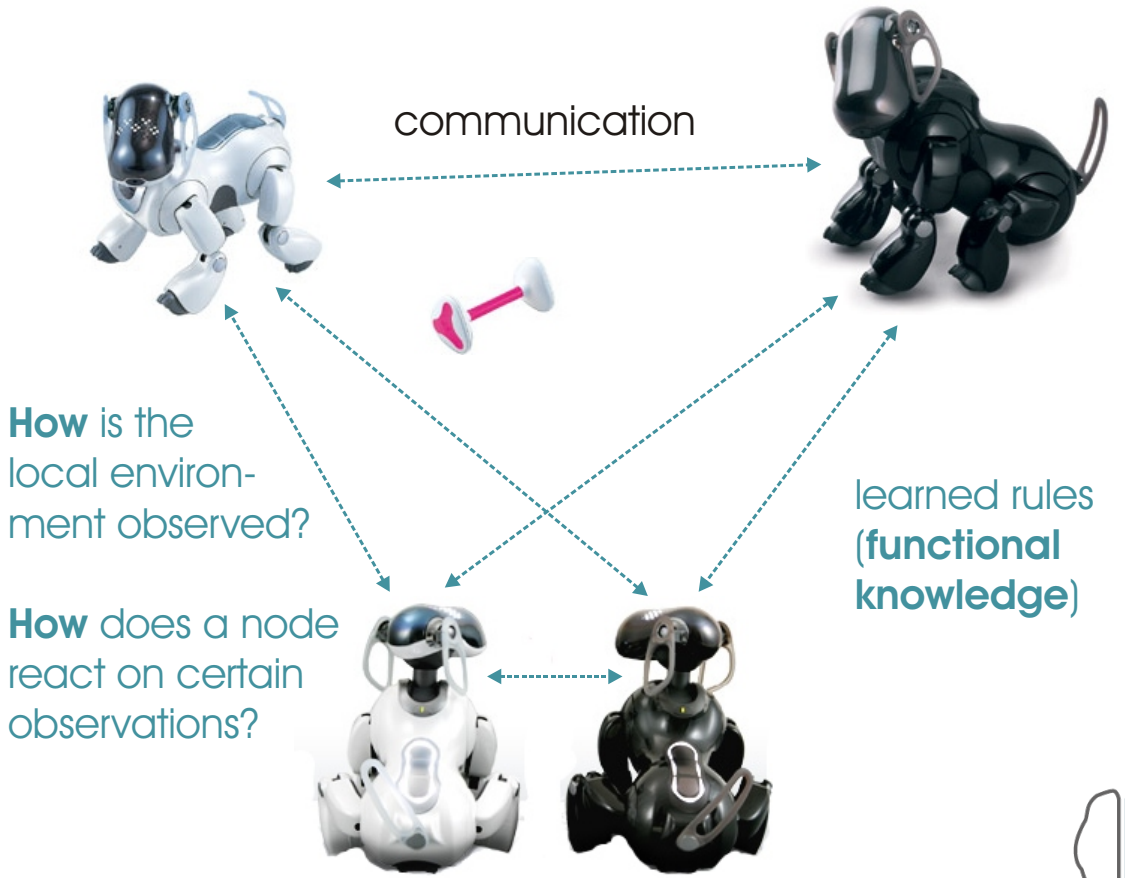
Outline of the Presentation

1. The beautiful idea
2. The ambitious plan
3. The hard reality
4. The real application
5. The final outcome

Intelligent Embedded Systems

The beautiful idea – 1

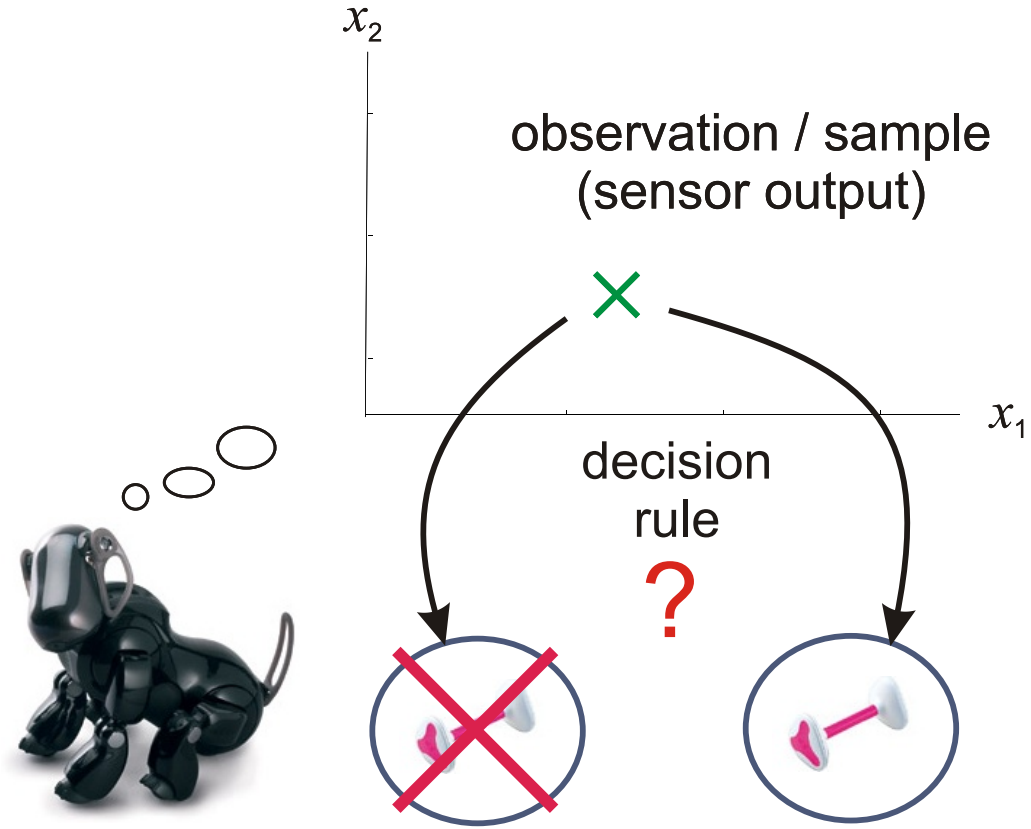
Collaboration of *organic agents* (i.e., intelligent systems such as teams of robots, smart sensor networks, or software agents) by exchanging learned rules instead of (or in addition to) observed samples.



Intelligent Embedded Systems

The beautiful idea – 2

Focus on classification problems with various applications, e.g., distributed intrusion detection in computer networks or strategy coordination in robotics.



Intelligent Embedded Systems

The beautiful idea – 3

Challenge:

Classification rules should not be “crisp” but consider the “uncertainty” of knowledge

In technical applications, *uncertainty* is caused by

- measurement errors
- transmission errors
- outliers
- missing values
- ...

Intelligent Embedded Systems

The ambitious plan – 1

Key research issues:

1. How can knowledge be represented or, in other words, which classifier paradigms can be used and how can these be trained from sample data (either off-line or on-line)?
2. How can a need to acquire new knowledge (novelty) or a possibility to discard outdated knowledge (obsoleteness) be detected?
3. How can knowledge (in form of rules) be extracted from one classifier and how can it be integrated in another classifier (e.g., fused with already existing knowledge)?
4. How can various knowledge properties be assessed numerically and how can a knowledge exchange process be improved by using this kind of meta-knowledge?

Intelligent Embedded Systems

The ambitious plan – 2

First attempt to approach a solution:

Use of radial basis function neural networks that – if trained appropriately – can be decomposed in a set of rules that are (from a functional viewpoint) similar to fuzzy rules.

Problem:

- The knowledge contained in these rules turned out to be quite "subjective", i.e., it is not potentially useful for other organic agents.

A lot more of basic research had to be done ...

Intelligent Embedded Systems

The hard reality – 1

Knowledge representation and offline acquisition:

- Classification rules are represented with a new kind of probabilistic classifier based on hybrid mixture models.
- The mixture models are hybrid as they combine different kinds of distributions (e.g., multinomial or Gaussian) for different dimensions of the input space.
- Parameters are found in a variational Bayesian approach, i.e., based on second-order probability theory (distributions are defined over the parameters of the classifier).
- Components of the mixture model describe processes that are assumed to "produce" the observed data, i.e., they model knowledge in an objective way.
- Classifiers can be trained partly unsupervised.

Intelligent Embedded Systems

The hard reality – 2

Online knowledge acquisition in dynamic environments:

- Techniques for novelty detection and obsoleteness detection are based on a very fast penalty/reward scheme derived from probabilistic considerations.
- The classifier can be adapted quickly as it can be trained incrementally.
- The detection techniques can also be applied to emergence detection and measurement, anomaly detection, or online clustering problems.

Intelligent Embedded Systems

The hard reality – 3

Knowledge extraction and fusion:

- Rules that are similar to fuzzy rules can be extracted from that classifier. Components of the mixture model are rule premises that are gradually assigned to classes.
- Similarity of rules can be measured by means of divergence measures from probability theory.
- Rules can be fused by combining their corresponding second-order distributions of parameters.
- Rules can be integrated into a classifier by adapting parameters such as mixture coefficients or rule conclusions.

Intelligent Embedded Systems

The hard reality – 4

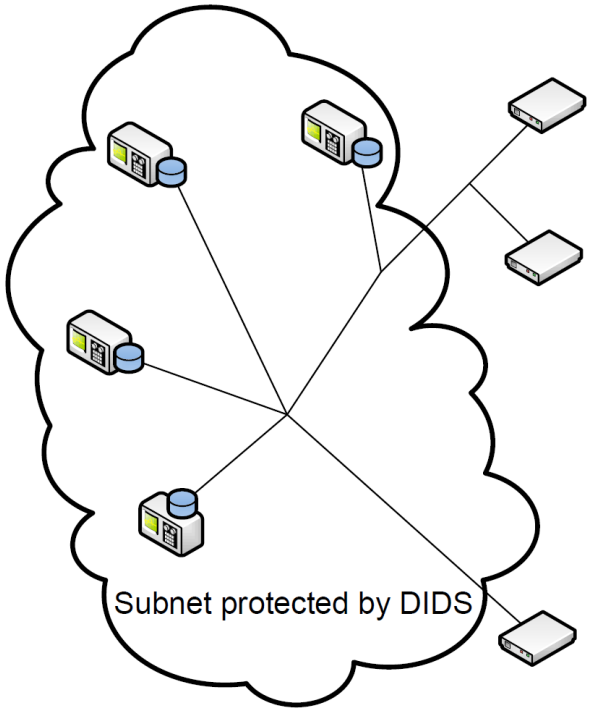
Knowledge assessment:

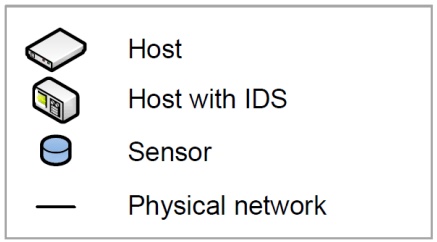
- Measures are defined that assess various properties of rules numerically, e.g., informativeness, importance, uniqueness, representativity, or comprehensibility.
- These measures are used to assess the "usefulness" of rules before they are integrated into a rule base by organic agents.
- The measures can also be applied to other data mining problems.

Intelligent Embedded Systems

The real application – 1

Distributed Intrusion Detection System (DIDS):



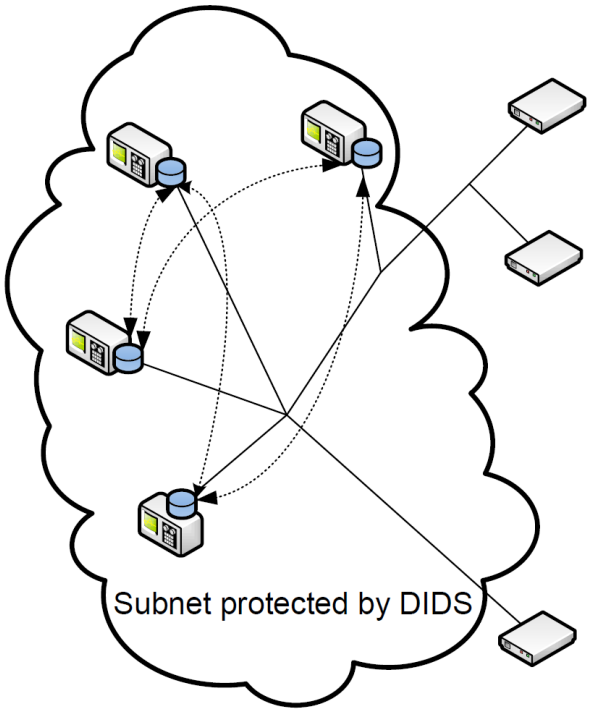


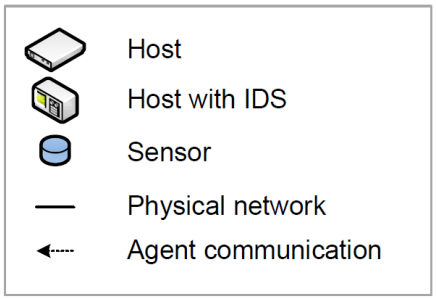
- IDS protect computer systems
  - Objective: Scan data for intrusions and alert administrator
  - Scanned data: Network traffic, log files, firewall messages, ...
- Advantages of DIDS
  - Enhanced scalability
  - No single point of failure

Intelligent Embedded Systems

The real application – 2

DIDS with OC techniques – knowledge exchange:





- Structurally similar Intrusion Detection Agents (IDA)
- Each IDA is **situation-aware** and **self-adaptive**:
  - performs its own detection task locally
  - is able to detect the need for new knowledge (i.e., new attack types)
  - is able to handle this situation, i.e., learn new classification rules
- IDA exchange learned rules
  - enables **pro-active** behavior
  - rule integration is controlled by an assessment of rules

Intelligent Embedded Systems

The real application – 3

Demonstration with parts of the KDD '99 network intrusion data set:

- Three collaborating IDA
- Initial classifiers trained with 2 000 background traffic records
- Locally learned rules are broadcasted
- Received rules are placed in a cache for further usefulness evaluation

Attack schedule:

Agent 1		Agent 2		Agent 3	
Time	Traffic	Time	Traffic	Time	Traffic
0 – 10 000	Normal	0 – 30 000	Normal	0 – 15 000	Normal
10 001 – 22 000	Portsweep	35 001 – 47 000	Portsweep	15 001 – 19 803	Back
22 001 – 32 000	Normal	47 001 – 50 000	Normal	19 804 – 50 000	Normal
32 001 – 36 000	Back				
36 001 – 50 000	Normal				

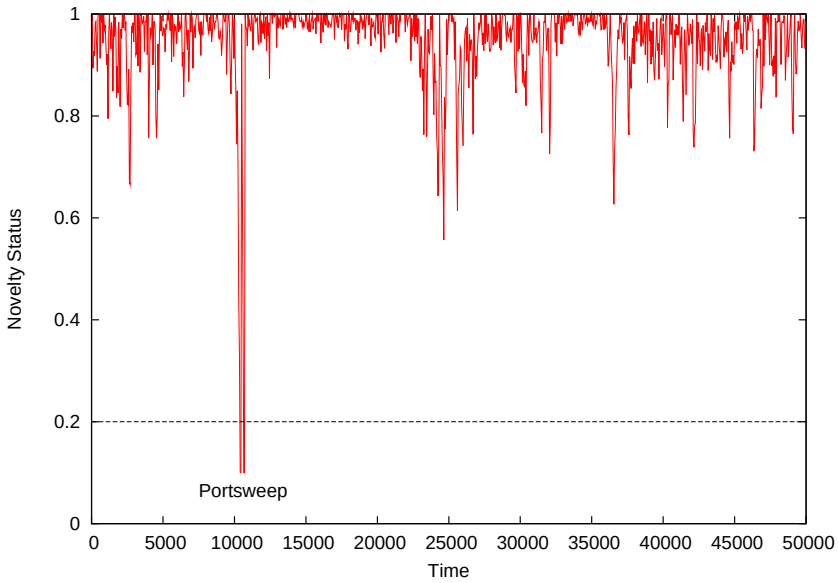
Intelligent Embedded Systems

The real application – 4

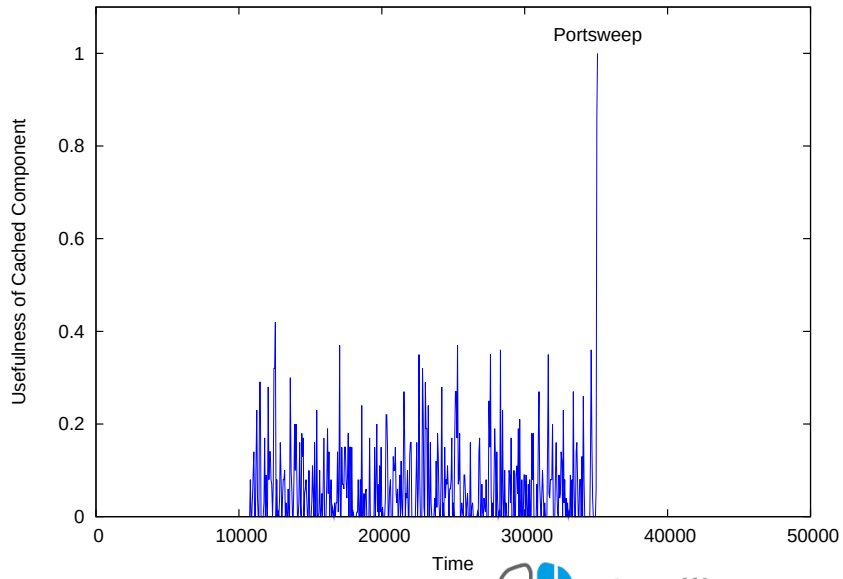
Knowledge exchange in action – example: Portsweep attack

- Agent 1: Start of Portsweep attack @ 10 000
- Agent 2: Start of Portsweep attack @ 35 000

Agent 1: Novelty measure



Agent 2: Usefulness measure



Intelligent Embedded Systems

The real application – 5

Comparison: No collaboration vs. collaboration

Agent 1

No collaboration:

True \ Pred.	Normal	Back	Portsweep
Normal	44 497 (96.7%)	123 (12.3%)	179 (6.0%)
Back	0 (0.0%)	877 (87.7%)	0 (0.0%)
Portsweep	1 503 (3.3%)	0 (0.0%)	2 821 (94.0%)

Collaboration:

True \ Pred.	Normal	Back	Portsweep
Normal	44 513 (96.8%)	33 (3.3%)	179 (6.0%)
Back	0 (0.0%)	967 (96.7%)	0 (0.0%)
Portsweep	1 487 (3.2%)	0 (0.0%)	2 821 (94.0%)

Intelligent Embedded Systems

The real application – 5

Comparison: No collaboration vs. collaboration

Agent 2

No collaboration:

True \ Pred.	Normal	Back	Portsweep
Normal	45 974 (97.8%)	–	482 (16.0%)
Back	–	–	–
Portsweep	1 026 (2.2%)	–	2 518 (84.0%)

Collaboration:

True \ Pred.	Normal	Back	Portsweep
Normal	46 274 (98.5%)	–	19 (0.6%)
Back	–	–	–
Portsweep	726 (1.5%)	–	2 981 (99.4%)

Intelligent Embedded Systems

The real application – 5

Comparison: No collaboration vs. collaboration

Agent 3

No collaboration:

True \ Pred.	Normal	Back	Portsweep
Normal	48 797 (100%)	153 (12.7%)	–
Back	0 (0.0%)	1 050 (87.3%)	–
Portsweep	–	–	–

Collaboration:

True \ Pred.	Normal	Back	Portsweep
Normal	48 797 (100%)	153 (12.7%)	–
Back	0 (0.0%)	1 050 (87.3%)	–
Portsweep	–	–	–

Intelligent Embedded Systems

The real application – 6

Collaboration of agents yields:

- **Improved classification performance**
- **Reduced number of human expert invocations**

The final outcome – 1

Publications:

- *Journals:*
  ACM TAAS 2011, IEEE Tr KDE 2011, IEEE Tr DSC 2011, INS 2010
- *Conferences and Workshops:*
  ICAART 2011, BICC 2010, SASO 2010, IJCNN 2009, IA 2009, SMCia/08, AHS-2008, FOCI 2007, CI-ALife 2007, ARCS 2007, ATC-06, SMCals/06, CIIW'05
- *Book Chapters:*
  Organic Computing – A Paradigm Shift for Complex Systems (ch 1.3 and ch. 3.3), Organic Computing (Ch. 4)
- *Best Paper Awards:*
  SASO 2010, SMCia/08, AHS-2008

Intelligent Embedded Systems

The final outcome – 2

We are very grateful to the DFG for the support of our project and to our colleagues in the priority program for excellent cooperation and very fruitful discussions!

Thank you! Any Questions?

Intelligent Embedded Systems