Towards Systematic Benchmarking of Activity Recognition Algorithms

Timo Sztyler, Christian Meilicke, Heiner Stuckenschmidt

CoMoRea´18

UNIVERSITÄT MANNHEIM

Content

1. Motivation
2. Benchmarking Ontology
3. Evaluation Framework
4. Running a Benchmark
5. Discussion & Next Steps

MOTIVATION

Motivation

What is Activity Recognition?

- Actions, Physical Activities, Activities of Daily Living, …

  ➡ No real agreement on the set of activities

- Sensors: Motion, Physiological, Proximity, Environmental, …
- People: Elderly, Children, Athletes, Handicapped, …

There are a lot of datasets which sounds good, but …

Idea

We propose a benchmarking framework, aiming to …

- … establish standards for datasets
- … directly compare AR algorithms
- … motivate researcher to join the competition

Our work is motivated by OAEI

Goal

We want to support researchers by …

- … providing a benchmark infrastructure
- … helping to identify relevant problems
- … easily compare different algorithms

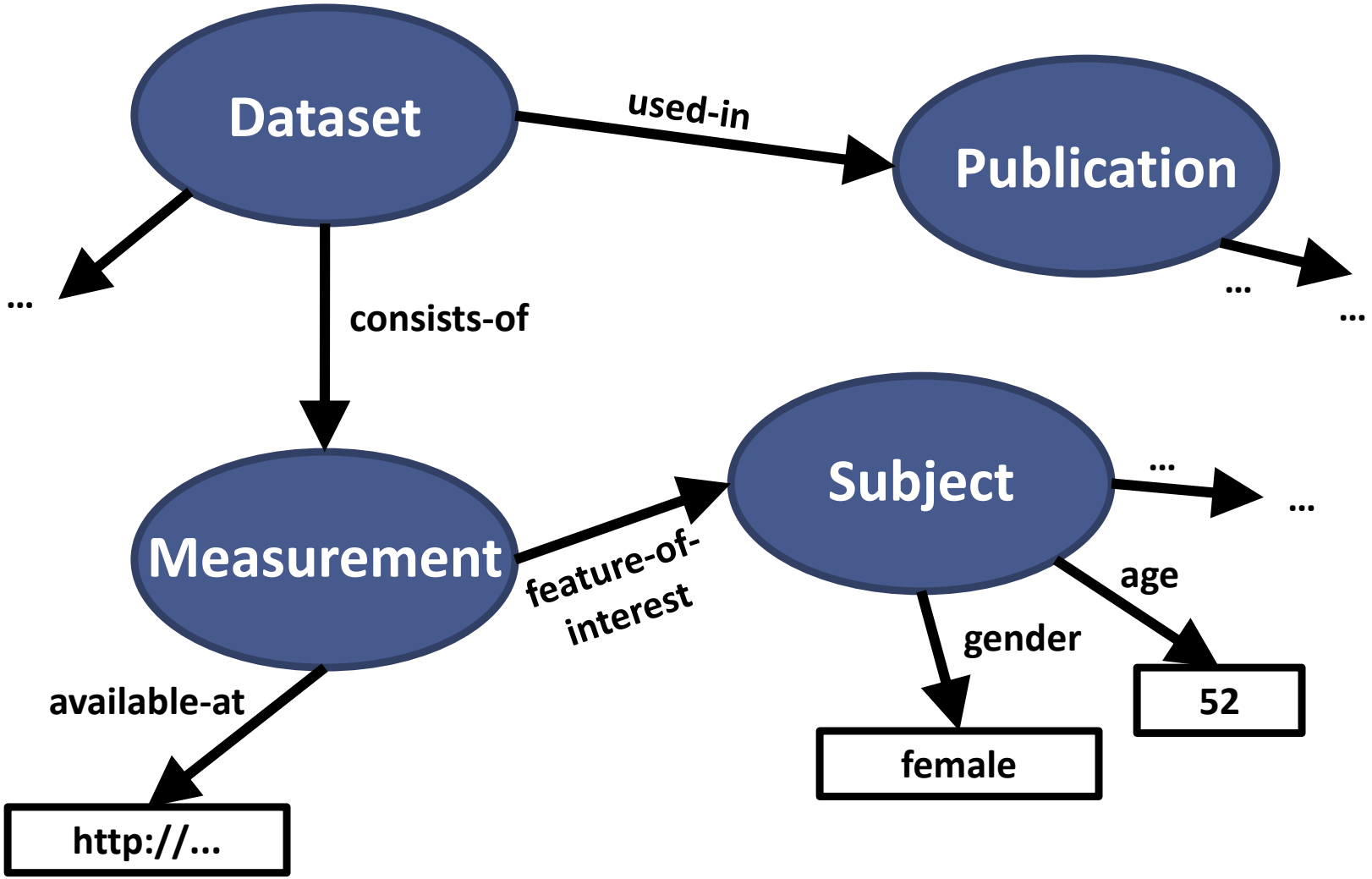
The framework should help to

- … find suitable datasets (method and scenario)
- … execute algorithms automatically for simple comparability

BENCHMARKING ONTOLOGY

Benchmarking Ontology

We built a web ontology by reusing existing models …



Measurements

Links a sensor to the phenomenon it measures …

- Activities and subjects are the phenomenon
  - We embed an activity hierarchy (“Ontology design pattern”)
  - Referred activities are gold standard annotations
- We adopted the concept of sensor types and its location

→ critical to identify suitable datasets

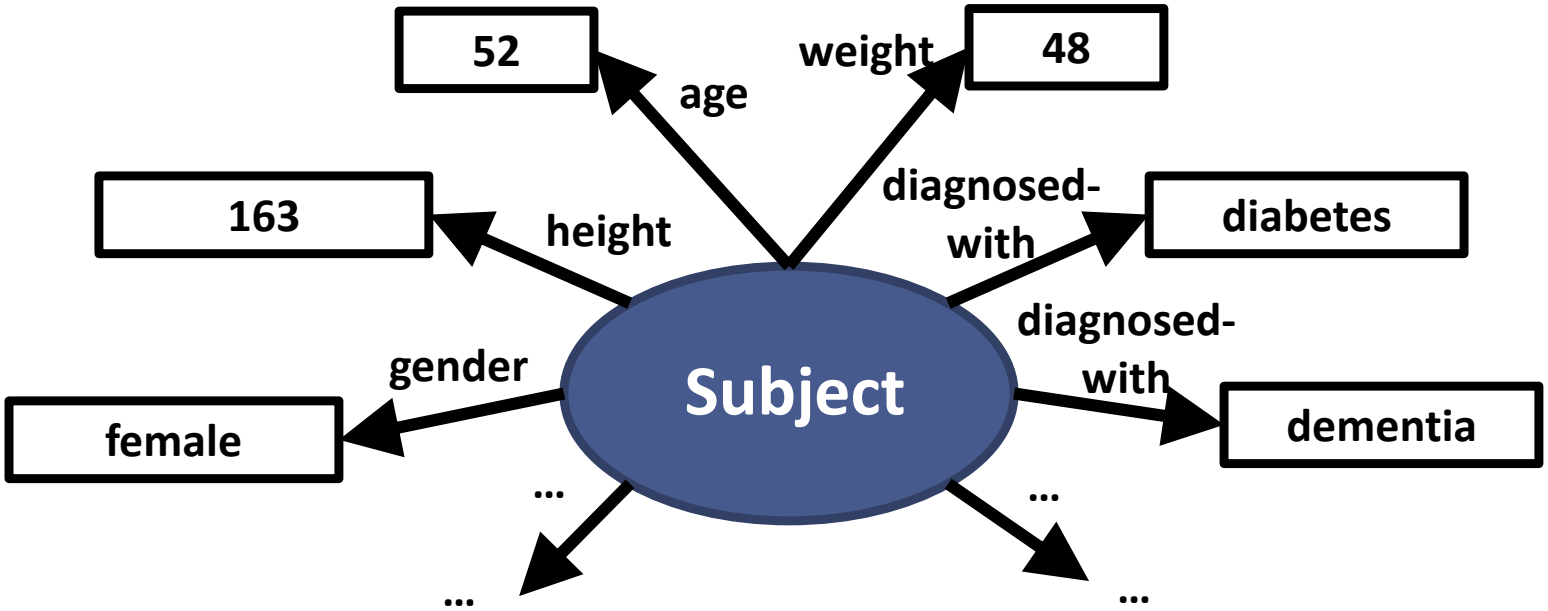
“time series” vs. “sequence of images”

Semantic Sensor Network Ontology

Subjects

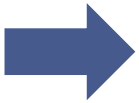
Information about the monitored subjects allow to …

- … benchmark specific algorithms (e.g. fall detection for elderly)
- … study approaches of transferring models
- … create synthetic datasets of certain subjects



Publications

So far, we focused on identifying datasets…



…but miss support for comparing results

The publication concept links to academic articles that use a particular dataset

- We do not plan to automate this process
  - The effort is not worth the effort
  - Maybe enhance the model by a result concept

Bibliographic Ontology BIBO

EVALUATION FRAMEWORK

Evaluation Framework

Activity Recognition Tool

- tool that is designed to solve the test suite

Evaluation Client

- piece of middle-aware (adapter)

Test suite

- fixed experimental setting and specific dataset
- can be simply accessed by a URL (standard web server)

Evaluation Client

We aim to minimize the effort …

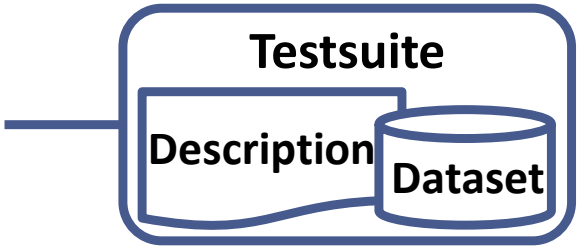
```
public void learnFrom(TrainingExample example);

public String classify(Example example);
```

➡ examples are predefined in a test suite

- Example: time interval, measurements and a label
- Independent of activity and sensor types
- Supports evaluation metrics (e.g. Recall)

Test Suite



Refers to a dataset and a description …

- relevant measures are not explicitly specified

➡ We want to keep it clear and simple

Dataset (CSV)

```
id,attr_time,attr_x,attr_y,attr_z
1,1438189989519,-0.006108,0.03481,-0.004581
2,1438189989520,-0.008246,0.030543,-0.002443
3,1438189989521,-0.009468,0.035430,-0.005192
4,1438189989522,-0.016187,0.00580,-0.004581
...
```

There are two constraints …

1) Dataset has to be a CSV file
2) consecutive sequence of points in time

Description (XML)

Ground Truth

```
<activitiyrecognition-suite name="example-name" description="some description">
  <trainingdata>
    <example targetclass="climbingdown">
      <sensordata type="Accelerometer" pos="forearm" source="p1/data/acc_climbingdown_forearm.csv"
        start="1435997166000" end="1435997167000" time-col="2"/>
      <sensordata type="Gyroscope" pos="forearm" source="p1/data/gyr_climbingdown_forearm.csv"
        start="1435997166000" end="1435997167000" time-col="2"/>
      <sensordata type="Magnetometer" pos="forearm" source="p1/data/mag_climbingdown_forearm.csv"
        start="1435997166000" end="1435997167000" time-col="2"/>
    </example>
    ... list further training examples here ...
  </trainingdata>
  <testdata> ... list test examples here ... </testdata>
</activitiyrecognition-suite>
```

Attributes

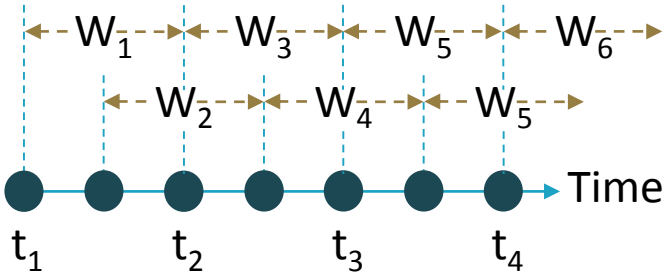
RUNNING A BENCHMARK

Running a Benchmark

We have generated seven test suites …

- We used a HAR dataset of a previous work
- Activities: running, standing, walking, …
- Seven on-body positions were recorded
- Accelerometer, Gyroscope, Magnetometer

We wrapped an existing activity recognition tool



Static windows, Overlapping

	Features
Time	Correlation coefficient (Pearson), entropy (Shannon), mean, mean absolute deviation, interquartile range (type R-5), kurtosis, median, standard deviation, variance
Frequency	Energy (Fourier, Parseval), entropy (Fourier, Shannon), DC mean (Fourier)

Results

Test suites and evaluation client were created independently …

- adaptation effort is feasible
- test suites can be used to get useful insights

Class	C	F	H	S	T	U	V
running	0.990	0.983	0.993	0.993	0.993	0.997	0.997
standing	0.984	0.965	0.971	0.997	0.991	0.965	0.975
jumping	0.952	1.000	0.976	1.000	0.976	1.000	1.000
walking	0.958	0.910	0.696	0.715	0.794	0.860	0.777
climbing up	0.952	0.803	0.873	0.887	0.839	0.902	0.966
climbing down	0.992	0.913	0.770	0.758	0.753	0.891	0.795
sitting	0.994	0.981	0.973	0.988	0.991	0.979	0.985
lying	0.990	0.987	0.990	0.987	0.981	0.990	0.990
avg.	0.977	0.943	0.905	0.916	0.915	0.948	0.936

DISCUSSION & NEXT STEPS

Discussion

Our proposal will only have an impact when …

- … it is used by a critical mass

We submitted this work to hear what other people think about this approach.

- Is it feasible and desirable?
- Do you already made experience with a similar idea?

➡ We support any attempt to use the proposed technology

Next Steps

We have to improve our central repository …

- … as WebProtégé can only grant editing rights

➡ We need an application that supports….

  - … fine grained rights (ownership)
  - … better search interfaces/algorithms
  - … more appropriated views and editing capabilities

- Contact dataset authors for discussions and collaborations

Thank you for your attention :)

CoMoRea´18