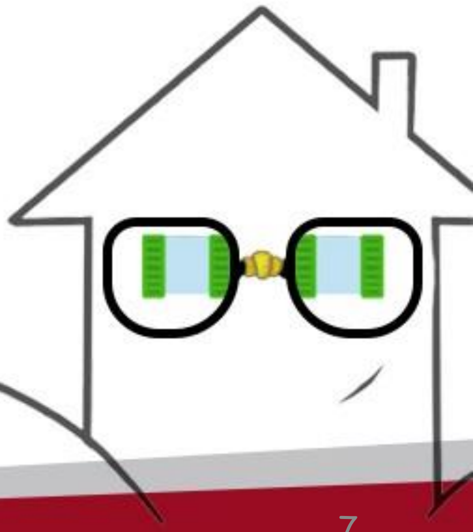
SAPIENZA
UNIVERSITÀ DI ROMA

Ambient Intelligence (Part 2)

PhD Course on Smart Environments: Technologies, state of the art and research challenges

Francesco Leotta
leotta@diag.uniroma1.it



Pattern Mining

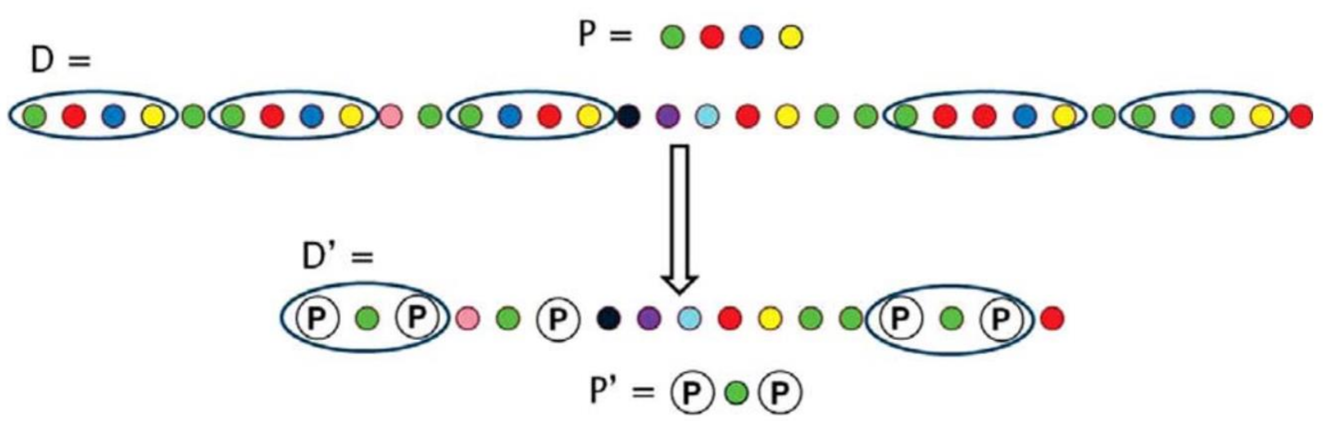
- Approaches related to pattern analysis in data mining take as input an event log and extract patterns of events obtained by windowing
- e.g., CASAS project employs a pattern mining technique to discover human activity patterns
  Allows to discover discontinuous patterns and variations
  - **Unsupervised learning**
  - **Used for labeling unlabeled activities…**
  - **…then HMM, CRF, SVM can be used**

Cook, Krishnan, Rashidi. Activity discovery and activity recognition: A new partnership. IEEE Trans. Cybern. 2013
Rashidi, Cook. COM: A method for mining and monitoring human activity patterns in home-based health monitoring systems. ACM Trans. Intell. Syst. Technol. (TIST) 2013

CASAS Pattern Mining Approach

- Based on a compression mechanism
  - A sequence pattern (P) is identified and used to compress the data set
  - A new best pattern (pattern P) is found in the next iteration of the algorithm

$$Compression = \frac{DL(D)}{DL(P) + DL(D|P)}$$



Emerging Patterns - EPs

- Are patterns of events that strongly characterize an activity or habit
- Given two databases of transactions T1 and T2
  - A transaction is a combination of variable assignments where all the variables are assigned
  - T1 is a set of transactions valid for a specific activity
  - T2 is a set of transactions from contrasting classes
- An itemset is an EP if its support in T1 wrt its support in T2 is high
  - The ratio is referred to as *growth rate*
- Usually supervised learning

Apriori Algorithm and AmI

- Seminal algorithm in data mining
- Initially introduced for smart supermarket
- Goal: find the frequent itemsets, i.e., products that are usually bought together
- Can be used in AmI in order to find:
  - Sensors usually triggering together
  - Sensors that when in specific state should trigger actions

Apriori Algorithm and AmI

- *Item* $v_i = d_k^{v_i}$
  - An assignment of a single variable to a single value out of its domain
- *Itemset* $C = \{v_i = d_k^{v_i} \mid \forall i \neq j : v_i \neq v_j\}$
  - A combination of *items* s.t. no variable is seen more than once
  - Constraints are returned by the algorithm in a form of *itemsets*
  - Itemsets can be turned if needed into rules A➔B == (not A or B)
- *Transaction*
  - An *itemset* that assigns values to all variables
- In smart spaces transactions are equivalent to situation
- Dataset is a sequence of transactions

Apriori Algorithm and AmI

- Given an itemset C, its support in T is defined as

$$Supp(C) = \frac{T^C}{W}$$

- T^C is the number of times the itemset C is in T, W is the total number of transactions

- An itemset is frequent if its support is above a minimum threshold value

Event Condition Action - ECA Rules

- Initially a specification-based method
- Can be compared to decision making in AI agents
  - Reflex agents with state take as input the current state of the world and a set of Condition-Action rules to choose the action to be performed
- An Event Condition Action - ECA rule basically has the form “ON event IF condition THEN action”, where conditions can take into account time
  - E.g.,

```
ON occurs ( Shower , Off , t0 )
IF context (BathroomHumidityLevel (>, 75))
THEN do (On, BathroomFan , t) when t = t0 + 10s
```

- **Unsupervised learning**
- **Mainly useful for decision making and analysis**

ECA Apriori Based Algorithm 1

- Define three categories of sensors
  - Type O sensors installed in objects, providing direct information about the actions of the users
  - Type C sensors providing information about the environment (e.g., temperature, day of the week)
  - Type M sensors providing information about the position of the user inside the house (e.g., in the bedroom)
- Events of the ECA rule always come from sets O and M
- Conditions expressed in terms of the values provided by Type C sensors
- The action part contains only Type O sensors (called mainSeT)
- Algorithm:
  1. Discover, for each sensor in the mainSeT, the set associatedSeT of O and M sensors potentially related to it as triggering events
     A. The method employed is APriori for association rules; association rules $X \Rightarrow Y$ are limited to those where cardinality of both X and Y is unitary and Y only contains events contained in mainSeT.
     B. This step requires a window size value to be specified in order to create transactions
  2. Discover the temporal relationships between the events in associatedSeT and those in mainSeT
  3. The conditions for the ECA rules are mined with a RIPPER learner

Aztiria, Augusto, Izaguirre, Cook. Learning Accurate Temporal Relations from User Actions in Intelligent Environments. 3rd Symposium of Ubiquitous Computing and Ambient Intelligence 2008.

ECA Apriori Based Algorithm 2 (1)

- Environment state can be represented as a set of variables with finite domain
- A ruleset describes domain of the smart environment and its expected behavior
  - Dependency between variables: $\neg([Shades = down] \wedge [Window = open])$
  - User preference: $(Room.Presence > 0) \Rightarrow (Room.Lamp = on))$
- Predicate logic over finite domains

Degeler, Lazovik, **Leotta**, Mecella. "Itemset-based mining of constraints for enacting smart environments". *In Proc. of ACOMORE Workshop, UBICOMP, 2014*

ECA Apriori Based Algorithm 2 (2)

- A rule can be represented either as **allowed** or as **forbidden** assignments
- Ex: $Jack.location = Room1 \Rightarrow Room1.Lamp = on$
  - Allowed: $(\neg Room1, on), (Room1, on), (\neg Room1, off)$
  - Forbidden: $(Room1, off)$
- The corresponding constraint:
  $$\neg(Jack.location = Room1 \wedge Room1.Lamp \neq on)$$
- Environments usually are more **permissive** than **restrictive**
- Rules are mined in the form of environmental constraints, i.e., we aim at finding **restricted combinations of values**
- To find a restriction we aim at detecting an **abnormal drop in frequency** of a combination of values w.r.t. its subsets
- Rules can be mined by using a variation of the **Apriori** algorithm

Degeler, Lazovik, **Leotta**, Mecella. "Itemset-based mining of constraints for enacting smart environments". *In Proc. of ACOMORE Workshop, UBICOMP, 2014*

SAPIENZA
UNIVERSITÀ DI ROMA

ECA Apriori Based Algorithm 2 (3)

- Dataset: sensor log $T = \{< S_1, w_1 > \ldots < S_{|T|}, w_{|T|} >\}$
  - Regular time intervals
    - consecutive sensor readings may represent the same situation
    - *w* is constant
  - Irregular time intervals
    - consecutive sensor readings contain changes
    - *w* depends on the amount of time the corresponding situation remained stable

- For an itemset *C* and a dataset *T*
  - T^C is the set of couples in a dataset *T* that contain the itemset *C*
  - $w(T^C)$ is the total weight of the itemset *C* in a dataset *T*

Degeler, Lazovik, **Leotta**, Mecella. "Itemset-based mining of constraints for enacting smart environments". *In Proc. of ACOMORE Workshop, UBICOMP, 2014*

ECA Apriori Based Algorithm 2 (4)

- A variation of the Apriori definitions to take into account weigths
- Given an itemset *C*, its support in *T* is defined as

$$Supp(C) = \frac{w(T^C)}{W}$$

- *W* is the total sum of weights over the entire dataset

$$W = \sum_{i=1}^{|T|} w_i$$

- An itemset is *frequent* if its support is above a minimum threshold value
- The algorithm seeks for a drop in frequency (support) of a set of variables with the addition of an item $v_i = d_j^{v_i}$

Degeler, Lazovik, **Leotta**, Mecella. "Itemset-based mining of constraints for enacting smart environments". *In Proc. of ACOMORE Workshop, UBICOMP, 2014*

Current Limitations/Opportunities (1)

- Several methods require an extensive labeling
  - **Feasible in the lab but more difficult in real settings**
  - **Human in the loop approach must be followed (e.g., NEST Thermostat)**
  - **Merge Specification and Learning based approaches**
  - If not labeling, at least segmentation
- ***In several cases, the recognition is performed, but it is coarse-grained***
  - To help the user throughout the pipeline (cf. step 5 in slide 21) is very challenging
  - Hierarchical models can be helpful

Labeling: *full annotation* of activity/habit instances

Segmentation: *surrogate* in which *only separation* is performed

Possible Segmentation Strategies

- **Explicit segmentation**. The stream is divided into chunks usually following some kind of classifier previously instructed over a training data set
- **Time-based windowing**. This approach divides the entire sequence into equal size time intervals
- **Sensor Event-based windowing**. This last approach splits the entire sequence into bins containing an equal number of sensor events

Krishnan, Cook. Activity recognition on streaming sensor data. Pervasive Mob. Comput. 2014

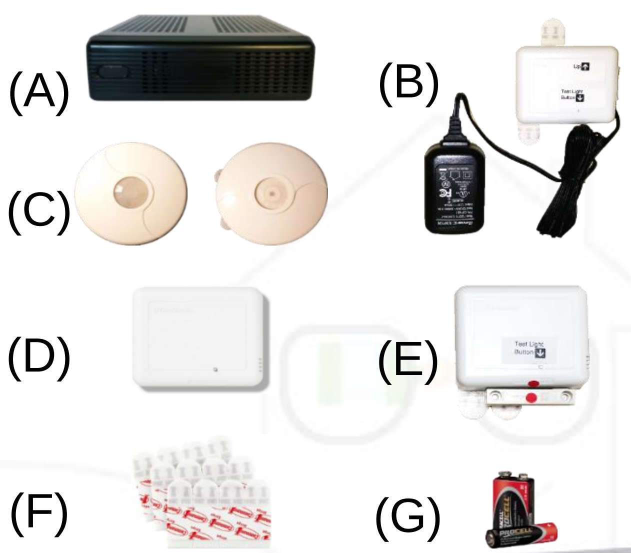
Current Limitations/Opportunities (2)

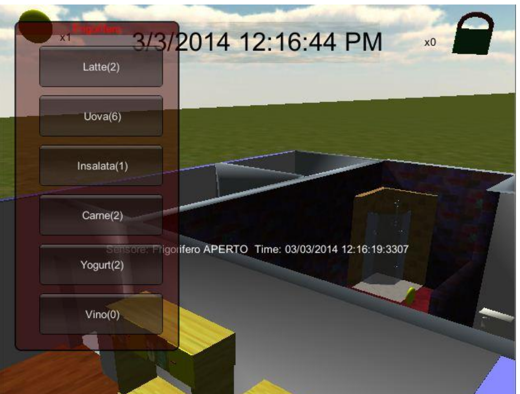
- The problem of multiple users is usually addressed only if labeling the dataset is possible
  - Not all devices/sensors can be equipped with tags
  - Tags are usually considered invasive by users
  - **Few papers so far addressing non-invasive sensors and/or without prelabeling**
- **The problem of knowledge update is usually not addressed**
- Visual analysis of human habits and activity is usually difficult
  - It may help to design better smart spaces

Rashidi, Cook. Keeping the resident in the loop: Adapting the smart home to the user. IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum. 2009

What about Datasets?

- Freely available datasets from the community for evaluation purposes
  - Limited availability: e.g., WSU CASAS installations
  - In some cases labeled with activities beginning and end markers
  - **No standard formats!!!**
  - **No guarantees about the quality of labeling**
- Smart Home in a Box
  - Freely distributed by WSU to research groups





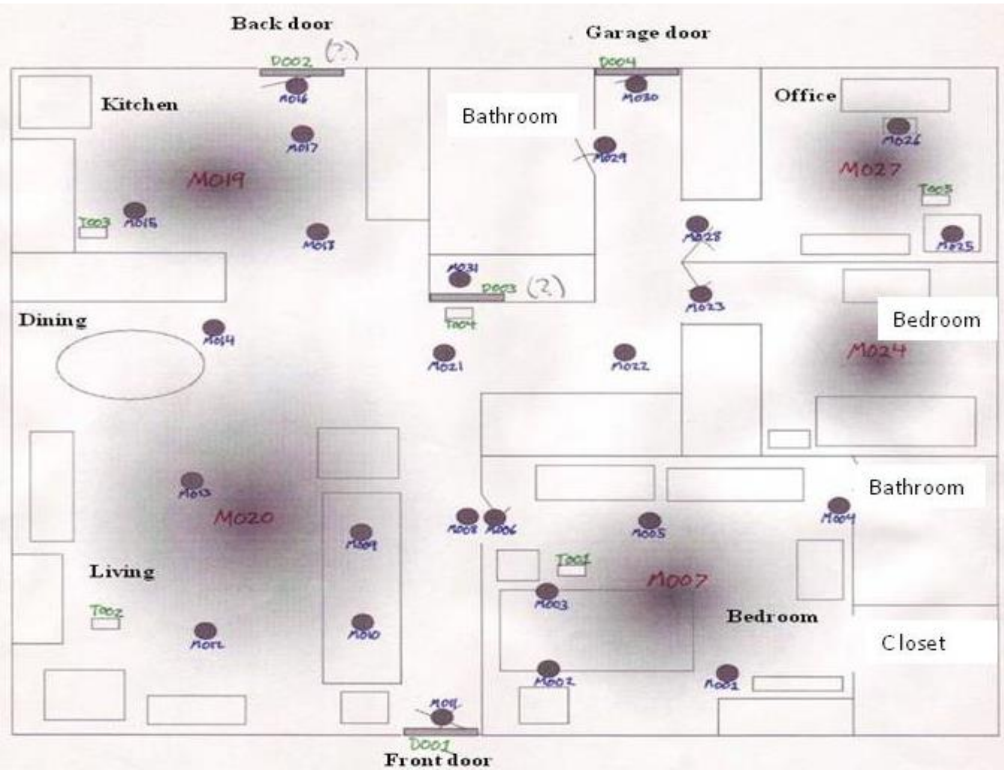
- Simulation
- Crowdsourcing
  - Ensuring truthfulness through clustering

Cucari, **Leotta**, Mecella, Vassos. "Collecting human habit datasets for smart spaces through gamification and crowdsourcing." <u>*In Proc. of GALA 2015*</u> (pp. 208-217)

CASAS

- http://casas.wsu.edu/datasets/
- Large collections of labeled and unlabeled data
  - In labeled data sensor measurements are associated with the activity the user was performing
- Single and multiple user data
- Mainly consistent of presence sensors installed with a grid layout on the ceiling
- Also data gathered through the Smart Home in a Box Program have been made available

CASAS




```
2010-11-04 05:40:46.310862 M003 OFF
2010-11-04 05:40:51.303739 M004 ON Bed_to_Toilet begin
2010-11-04 05:40:52.342105 M005 OFF
2010-11-04 05:40:57.176409 M007 OFF
2010-11-04 05:40:57.941486 M004 OFF
2010-11-04 05:43:24.021475 M004 ON
2010-11-04 05:43:26.273181 M004 OFF
2010-11-04 05:43:26.345503 M007 ON
2010-11-04 05:43:26.793102 M004 ON
2010-11-04 05:43:27.195347 M007 OFF
2010-11-04 05:43:27.787437 M007 ON
2010-11-04 05:43:29.711796 M005 ON
2010-11-04 05:43:30.279021 M004 OFF Bed_to_Toilet end
```

MXXX movement TXXX temperature DXXX door

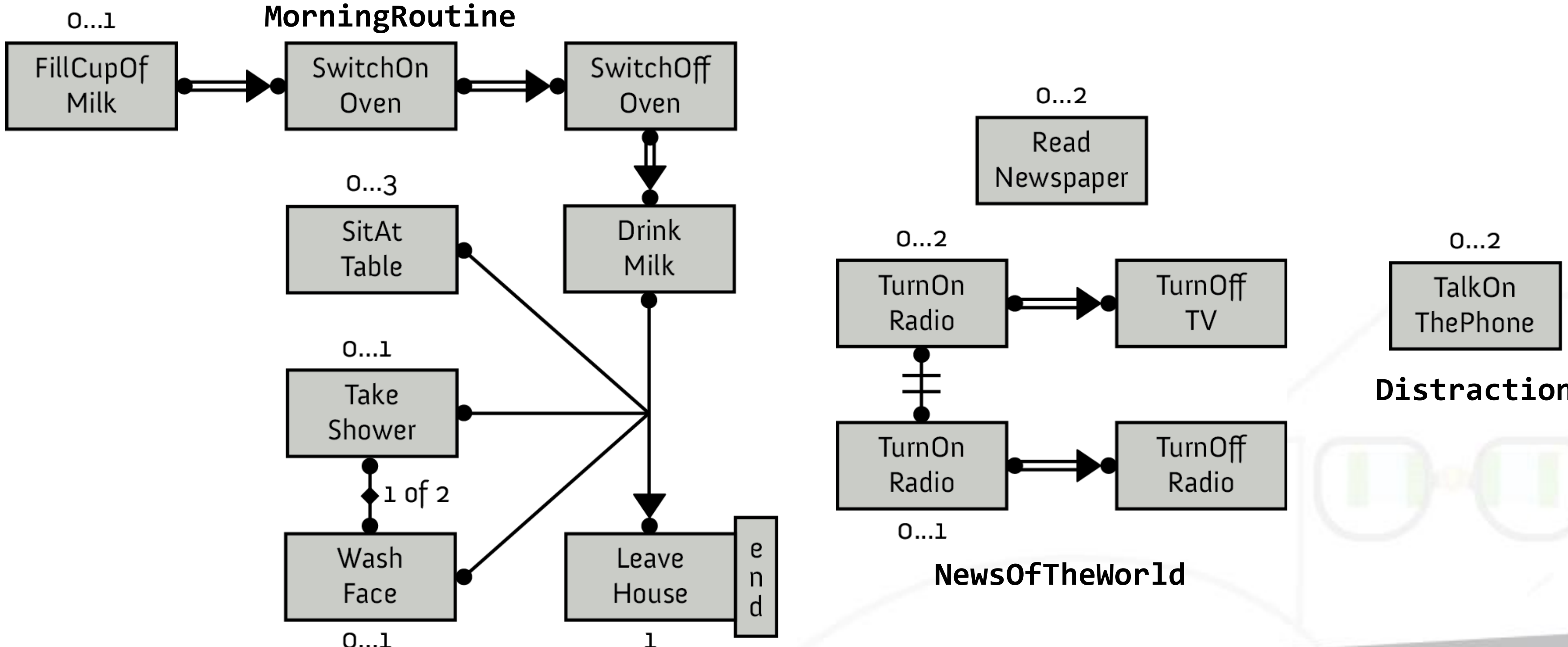
TRACEBASE

- https://github.com/areinhardt/tracebase
- A collection of power consumption traces
- Traces collected from individual electrical appliances
  - One sample per second
  - Average in the last 1s and 8s

```
14/01/2012 10:48:47; 151; 156
14/01/2012 10:48:48; 147; 151
14/01/2012 10:48:49; 147; 151
14/01/2012 10:48:50; 145; 149
14/01/2012 10:48:51; 145; 147
14/01/2012 10:48:52; 145; 147
14/01/2012 10:48:53; 143; 147
14/01/2012 10:48:54; 143; 145
14/01/2012 10:48:55; 143; 143
```

Simulated Data Through Planning

- A set of models defined in DECLARE
  - an LTL based business process modeling language



Caruso, Ilban, Leotta, Mecella, Vassos. Synthesizing daily life logs through gaming and simulation. Proceedings of the AwareCast 2013

Simulated Data Through Planning

- Iteratively select a random number of habit and distraction models
- Combine selected models by putting in logical AND their semantics
- Select a random compliant execution trace consisting of tasks called **h-actions**
- The sequence of all generated random executions is called habit trace
- **h-actions** are essentially high-level tasks to be executed in a generic environment → impossible to directly bind sensors
- **Hint: associate to each h-action a goal to be realized according to a lower-level underlying action theory**
- Employ a planner that produces a sequence of low-level atomic p-actions for each h-action in the trace, generating a more detailed planner log

Caruso, Ilban, Leotta, Mecella, Vassos. Synthesizing daily life logs through gaming and simulation. Proceedings of the AwareCast 2013

Simulated Data Through Planning

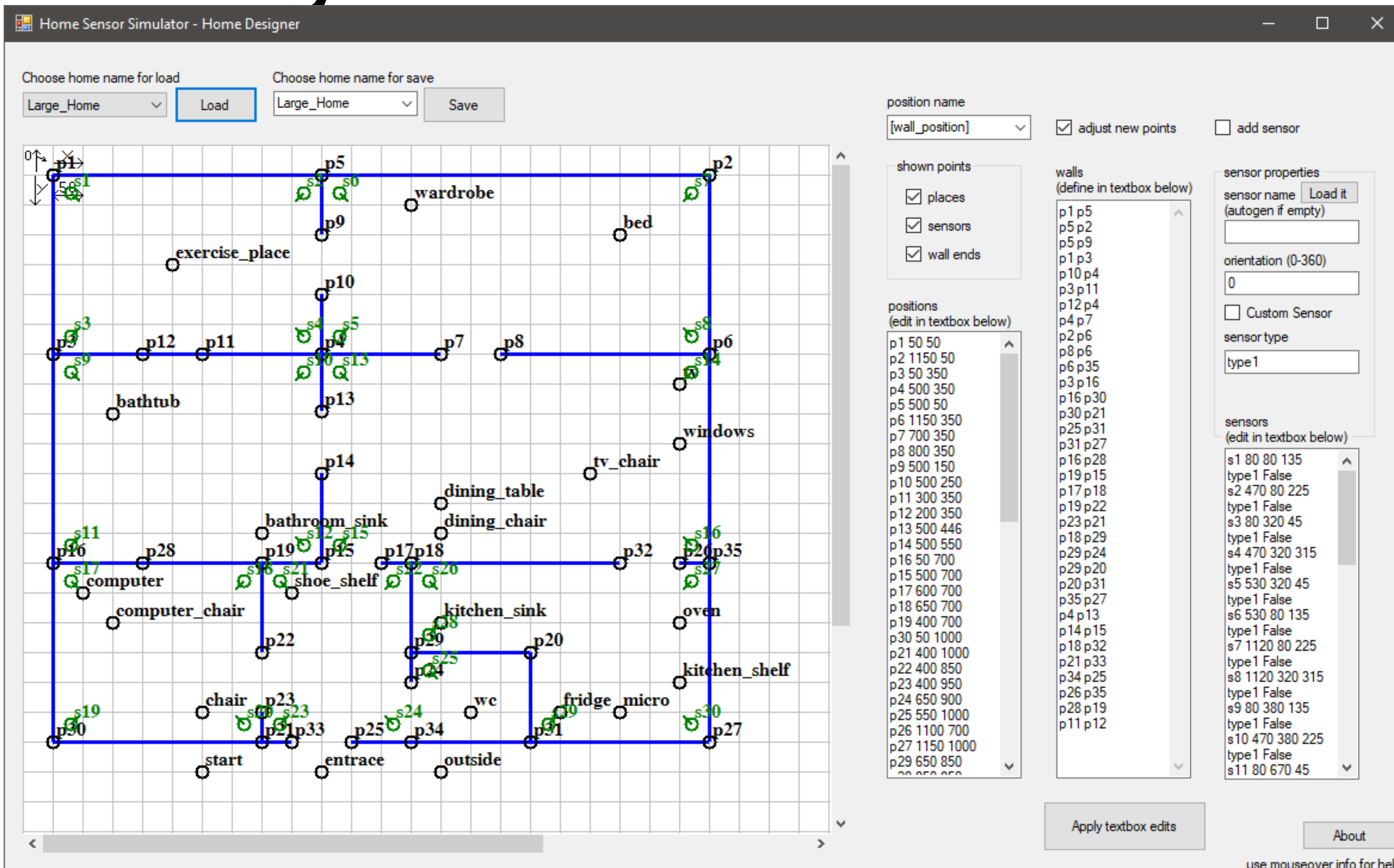
HABIT TRACE	PLANNER LOG
...	...
FillCupOfMilk	GOAL deviceState cupOfMilk filled (1) (moveToDevice livingroom newspaper kitchenDoor) (2) (changeDeviceState livingroom kitchenDoor closed open) (3) (moveToRoom livingroom kitchenDoor kitchen kitchenDoor) (4) (moveToDevice kitchen null cupOfMilk) (5) (changeDeviceState kitchen cupOfMilk empty filled)
TurnOnRadio	GOAL deviceState radio on (1) (moveToRoom kitchen cupOfMilk livingroom kitchenDoor) (2) (moveToDevice livingroom null radio) (3) (changeDeviceState livingroom radio off on)
TalkOnThePhone	GOAL usedDevice livingroom phone (1) (moveToDevice livingroom radio phone) (2) (useDevice livingroom phone)
...	...

Caruso, Ilban, Leotta, Mecella, Vassos. Synthesizing daily life logs through gaming and simulation. Proceedings of the AwareCast 2013

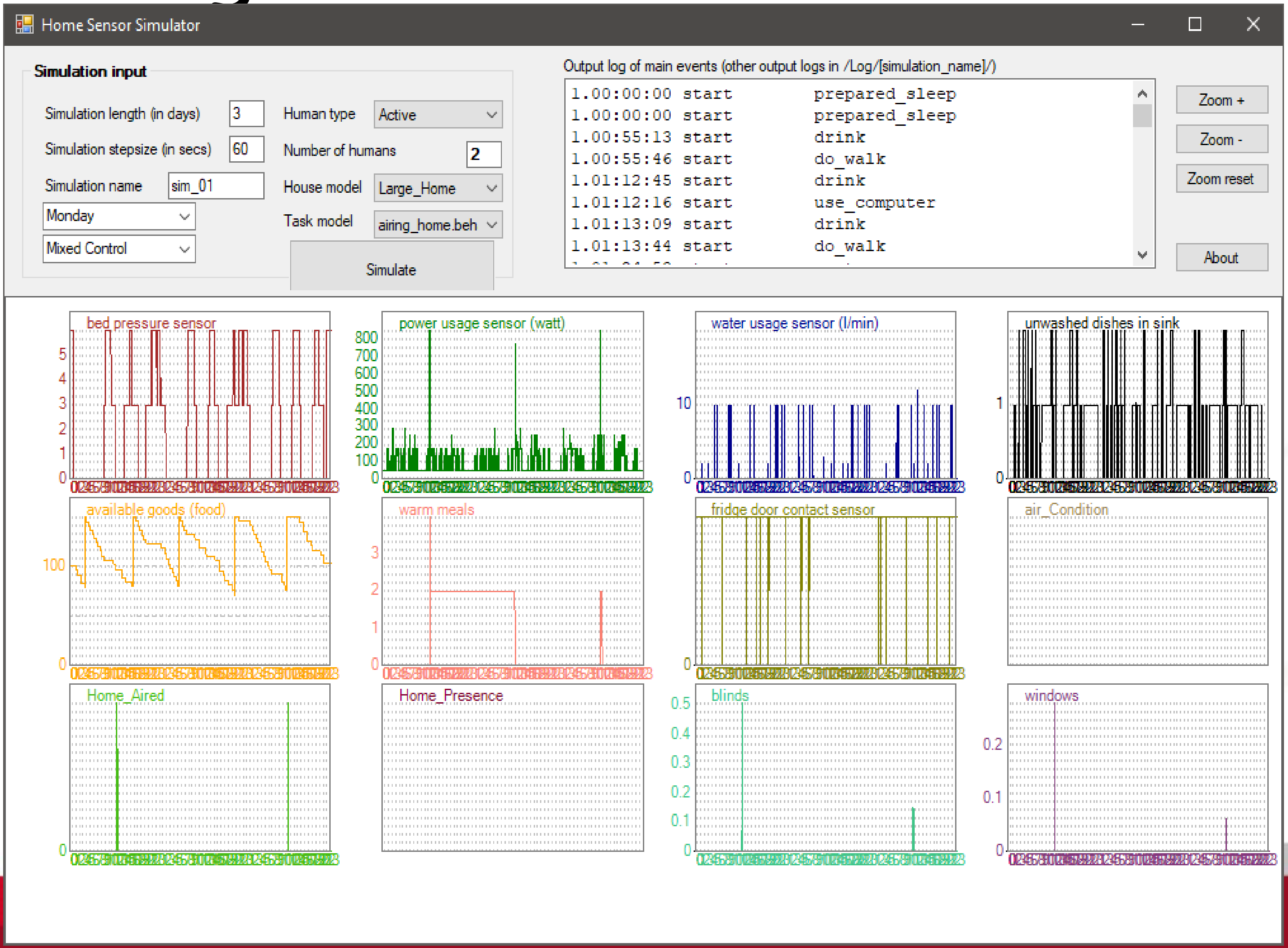
Simulated Data Through Physical Modeling

- Previous approach (planning) does not easily allow to model physical properties → low realism
- Human behavior driven by needs (e.g., thirsty)
- When certain conditions are verified, flows of actions are executed on the environment
- Actions change the state of the system and the state of the users
- Support for multiple users
- Sensor can be easily added and modeled through mathematical expressions

Simulated Data Through Physical Modeling



Simulated Data Through Physical Modeling



Final Project Ideas

- Given a freely available dataset, propose a strategy to:
  - Segment activities for single persons
    - E.g., inactivity based, based on some external knowledge
  - Segment multiple users
  - Automatically define habits based on clusters of activities
- From energy consumption data derive patterns of employment of devices
- Comparison of activity discovery algorithms
- Comparison of activity forecast algorithms
- Evaluation of simulated data
- Automatically define rule definition models (e.g., IFTTT) from learned models (E.g., ECA rules)
- …