Population-level Behavior Analysis on Smart Environment Sensor Data

Beiyu Lin

Ph.D. Candidate

WASHINGTON STATE UNIVERSITY

Research Goal

Research Goal

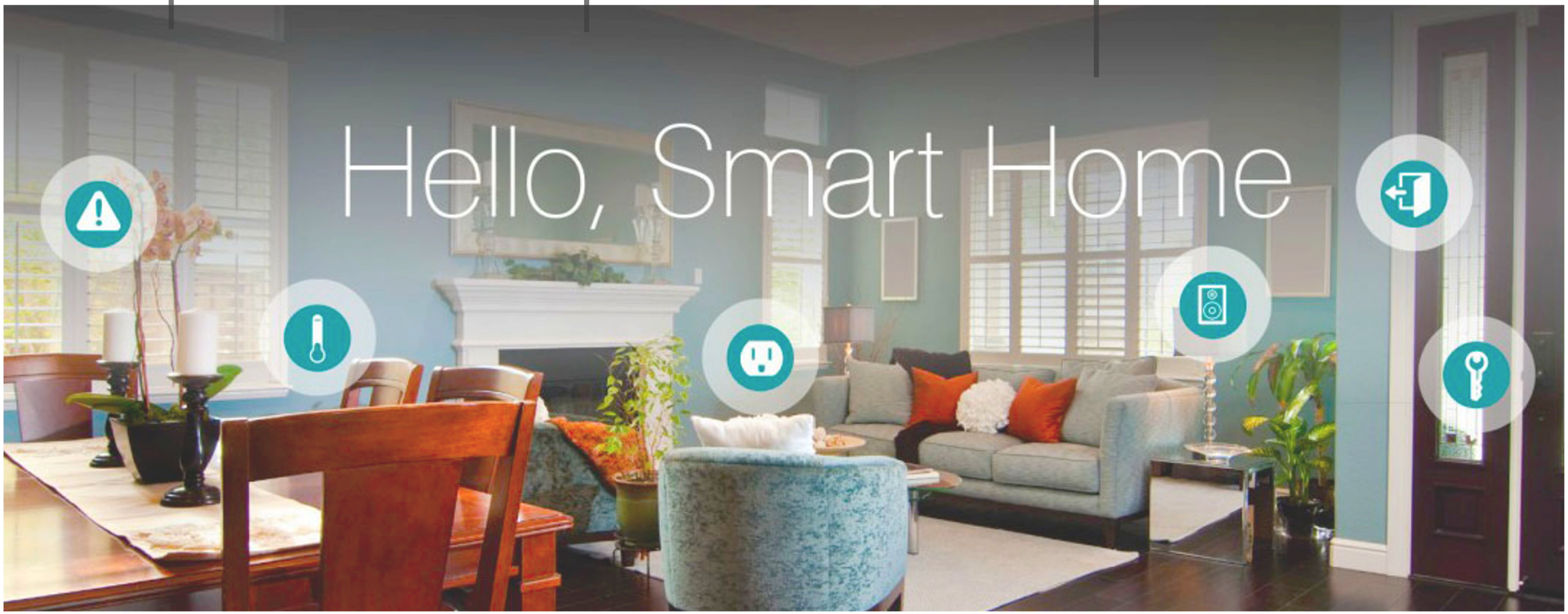
- Formally model population-level human behavior from ambient sensor data
- Use these models to compare population subgroups

Dissertation steps

Mathematical Model of Indoor Behavior

Markov Model of Indoor Behavior

Inverse Reinforcement Learning Model of Indoor behavior



Contributions

1. **Indoor behavior can be modeled by a Pareto distribution**
    - Innovation: heavy-tailed distribution (***Pareto distribution***)
    - Impact: the Pareto distribution and its properties, the 80/20 rule, can be used to model various human behavior patterns

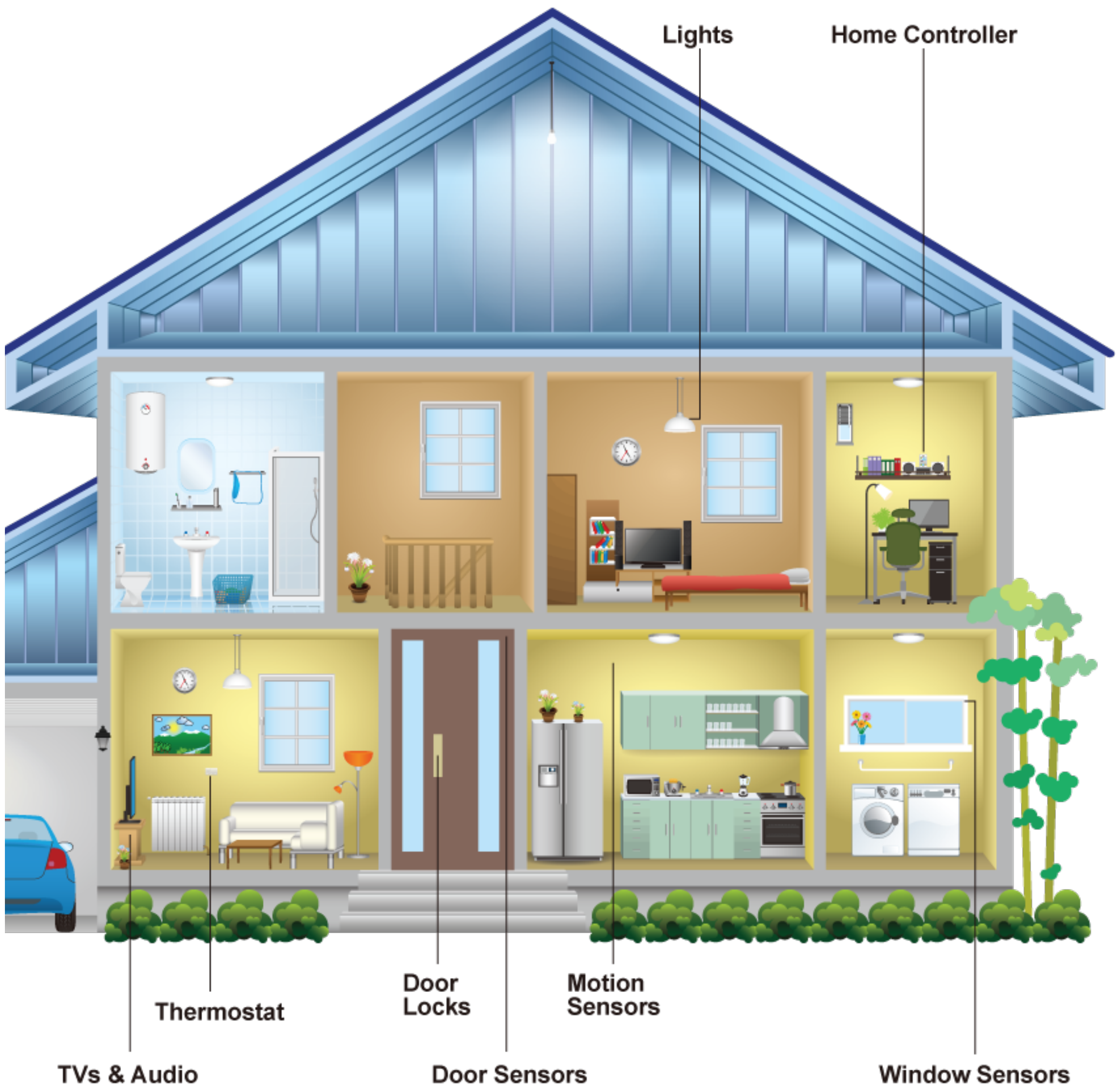
2. **Proposed a new measure for model selection**
    - Innovation: ***a new measure*** to balance the tradeoff between under/over-estimating the true Markov order
    - Impact: understand the complexity of sensor-observed human indoor behavior

3. **Proposed inverse reinforcement learning to learn indoor behavior**
    - Innovation: ***1st study*** of IRL to model indoor behavior
    - Impact: extract quantitative and qualitative measures of human behavior

4. **Distinguished population subgroups based on ambient sensor data**
    - Innovation: ***population-level behavior analysis*** *via ambient sensor data*
    - Impact: compare behavior for healthy older adults with behavior of older adults with chronic health conditions

Background: sensor data

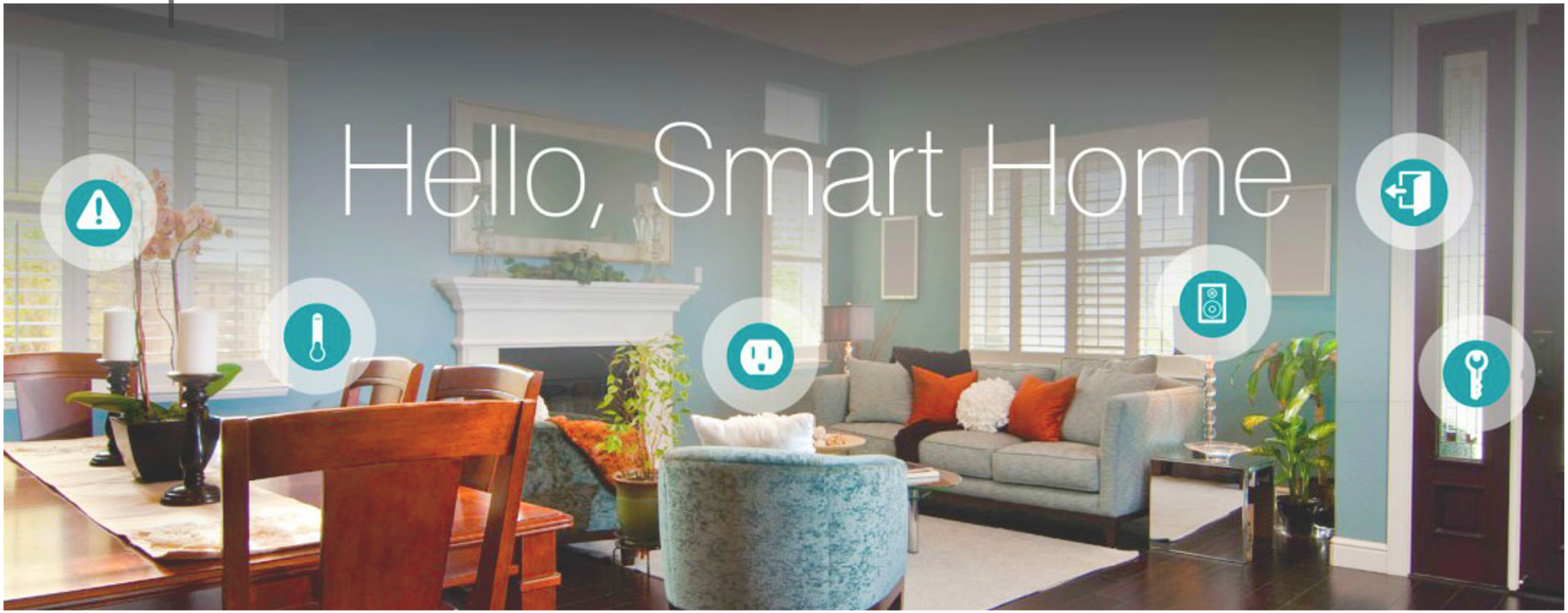
Sensor types: infrared motion(narrow/wide-area), ambient light, magnetic, and temperature sensors.



2011-06-13	21:48:43	Bathroom	ON	Personal_Hygiene
2011-06-13	21:48:44	Bathroom	OFF	Personal_Hygiene
2011-06-13	22:47:02	Bedroom	ON	Personal_Hygiene
2011-06-13	22:47:04	Bedroom	OFF	Sleep
2011-06-13	22:47:06	Bedroom	ON	Sleep
....				
...				
2011-06-14	10:11:24	Kitchen	ON	Wash_Dishes
2011-06-14	10:11:25	Kitchen	OFF	Wash_Dishes
2011-06-14	10:11:40	Kitchen	ON	Cook
2011-06-14	10:11:41	Kitchen	OFF	Wash_Dishes

Step 1

Mathematical Model of Indoor Behavior



Mathematical Model of Indoor Behavior - Hypothesis

We hypothesize that:

We can mathematically model human behavior at an individual and population level based on ambient sensor data.

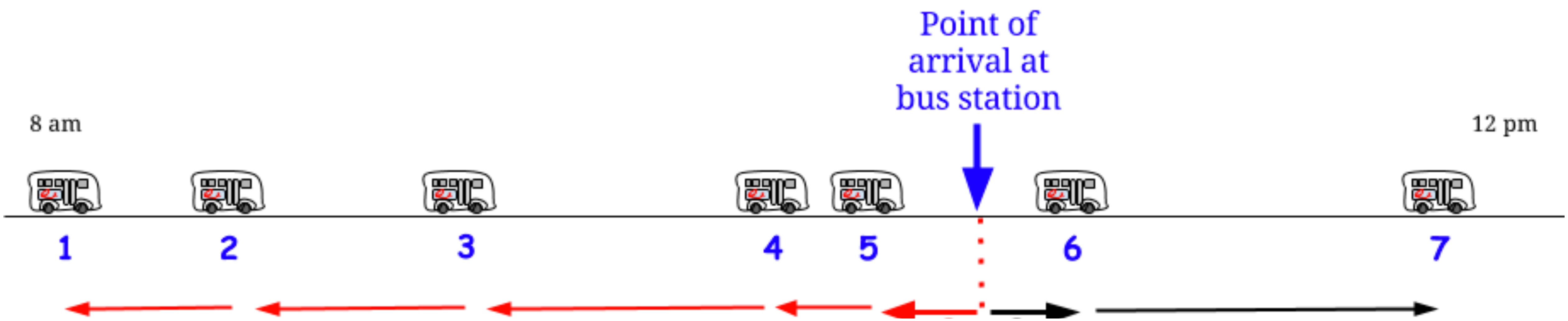
We first model temporal behavior features

We assume temporal behavior features can be:

activity durations, inter-arrival times of two successive activities, etc.

Mathematical Model of Indoor Behavior - Background

Inter-arrival times:



the time interval between two successive occurrences of activities

Mathematical Model of Indoor Behavior - Datasets

	% in each subgroup in total 99 smart homes
Number of residents	
Single-resident	46
Two-resident	18
Three or more-resident	4
Not reported	32
Age	
Young (age $\leq$ 35)	14
Middle aged (35 $<$ age $\leq$ 64)	9
Senior (age $>$ 64)	65
Not reported	12
Education level	
High school	10
Bachelor	19
Master	20
Doctorate	15
Not reported	36

	% in each subgroup in total 99 smart homes
Health status	
Healthy	57
Targeted health ailments	23
Cognitive decline	8
MCI	4
Dementia	3
Parkinson's Disease	1
Mobility limitations	4
Lung problems	2
Atrial fibrillation	2
Macular degeneration	1
Other health conditions	6
Not reported	20

Mathematical Model of Indoor Behavior - Algorithm

Algorithm 1 Population-based activity modeling

input : X sequences of sensor events with associated automatically labelled activity

A set of selected activities to analyze

output : the top 15 fitted distributions

C = ChangePointDetection(X) (identify set of change points),

S = SegmentData(C) (segment activities with start and end times),

E = ExtremeValue(S) (perform extreme value theory to filter out outliers),

D = SeparateData(E, A) (separate data into one set per analyzed activity)

for *each activity* a $\in$ A **do**

d = (D, a) (data for selected activity)

I = InterArrivalTimes(d) (calculate inter-arrival times of activities),

H = Histogram(I) (fit 82 distributions to the histogram of the inter-arrival times),

$T_{15,a}$ = SumSquareError(H) (select the top 15 distributions with sum of square errors)

end

return T_{15} *(return list of all top 15 distributions)*

Mathematical Model of Indoor Behavior - Methodology

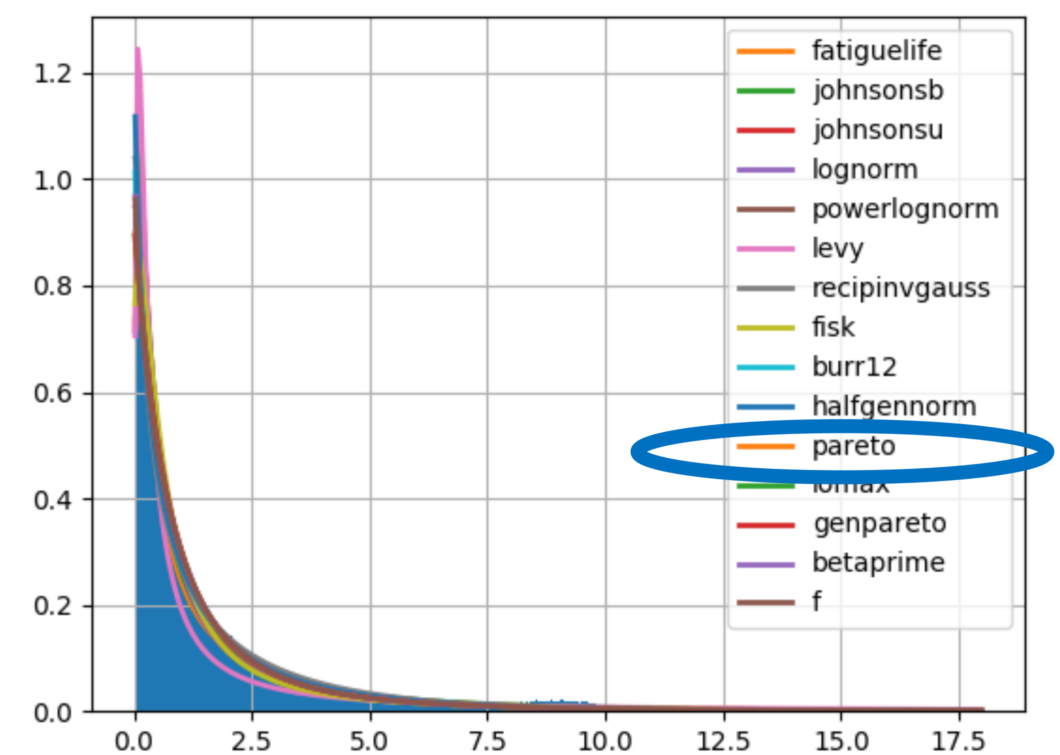
A sequence of smart home sensor data with labelled activities

Change Point	Date/Time	Activity Label
0	2012-08-24 16:06:06	Other_Activity
0	2012-08-24 16:06:07	Other_Activity
1	2012-08-24 16:21:24	Other_Activity
0	2012-08-24 16:12:26	Relax
0	2012-08-24 16:13:34	Relax
0	2012-08-24 16:13:35	Relax
0	2012-08-24 16:34:52	Relax
0	2012-08-24 16:34:53	Relax
0	2012-08-24 16:35:05	Relax
0	2012-08-24 16:35:06	Relax
0	2012-08-24 16:35:06	Relax
0	2012-08-24 16:35:07	Relax
0	2012-08-24 16:35:10	Relax
0	2012-08-24 16:35:10	Relax
0	2012-08-24 16:35:13	Relax
0	2012-08-24 16:35:14	Relax
1	2012-08-24 16:35:17	Relax
0	2012-08-24 16:35:17	Relax

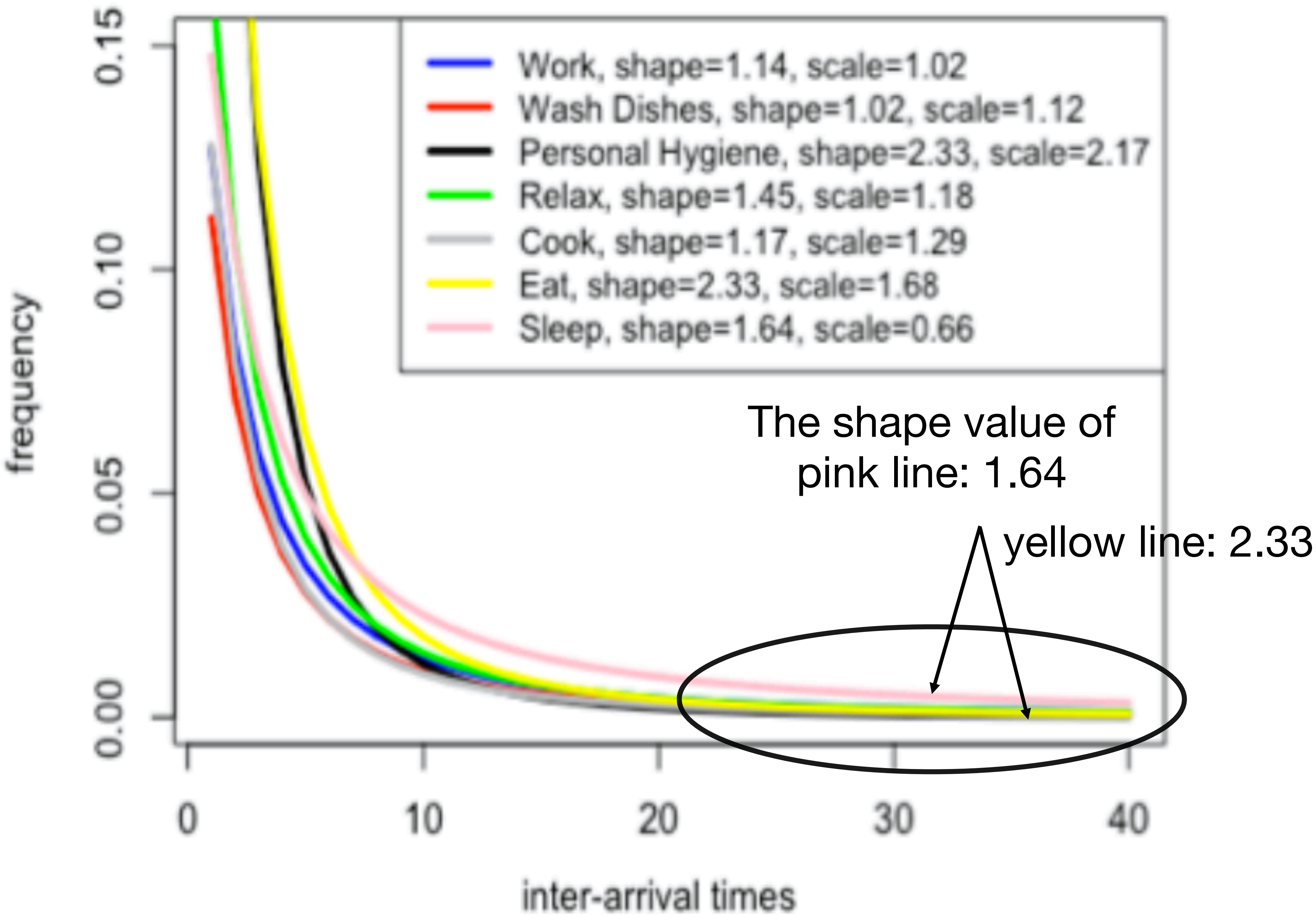
Segment Data

1	2012-08-24 16:21:24	Other_Activity
0	2012-08-24 16:35:14	Relax

Data Clean



Mathematical Model of Indoor Behavior – Results



Mathematical Model of Indoor Behavior - Experiments

Three experiments:

1. **Population-wide analysis**

    99 homes selected subset of activities

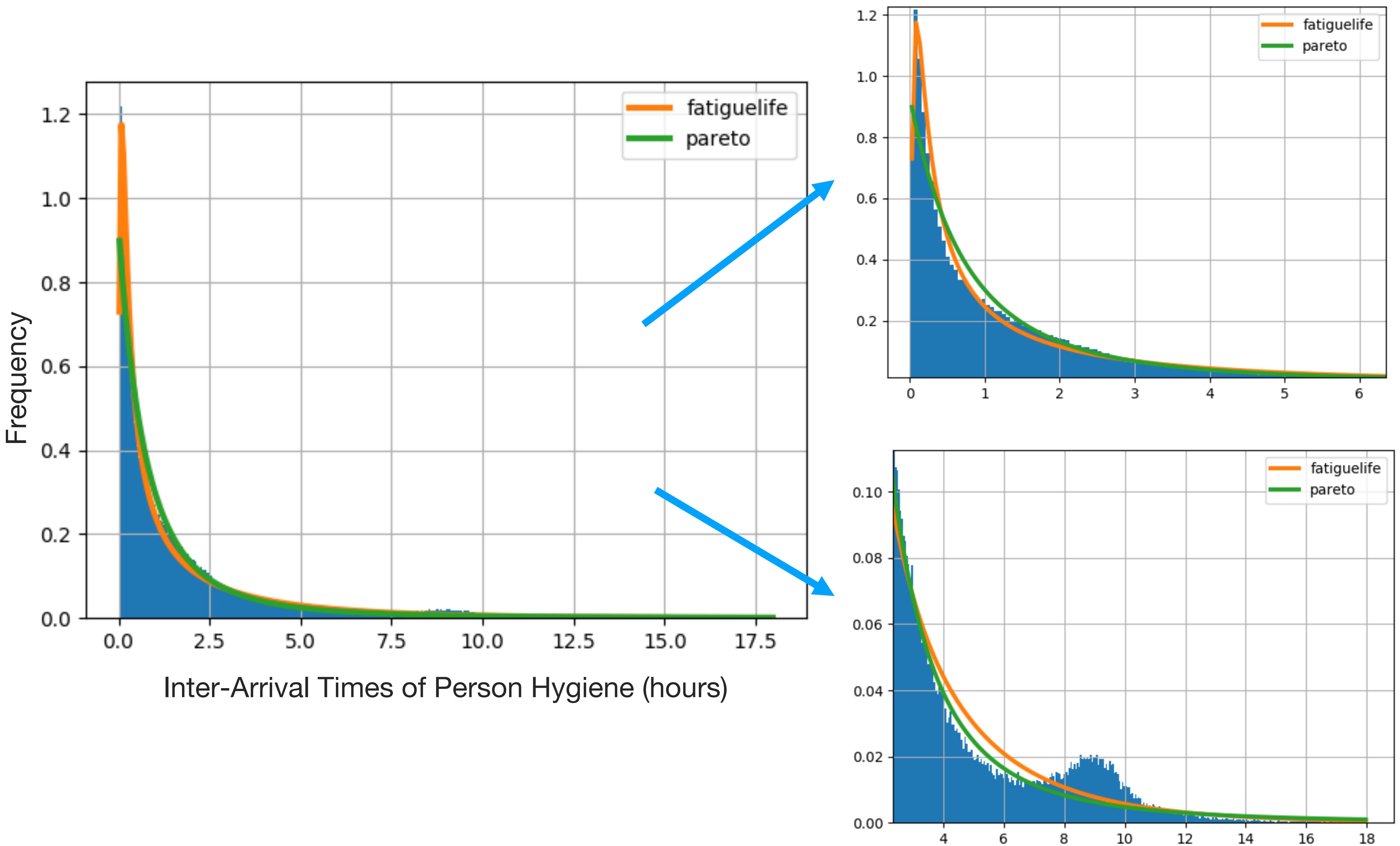
2. **Group-wide analysis on homes split into two groups**

    Healthy (57 homes) and Health complaints (23 homes)

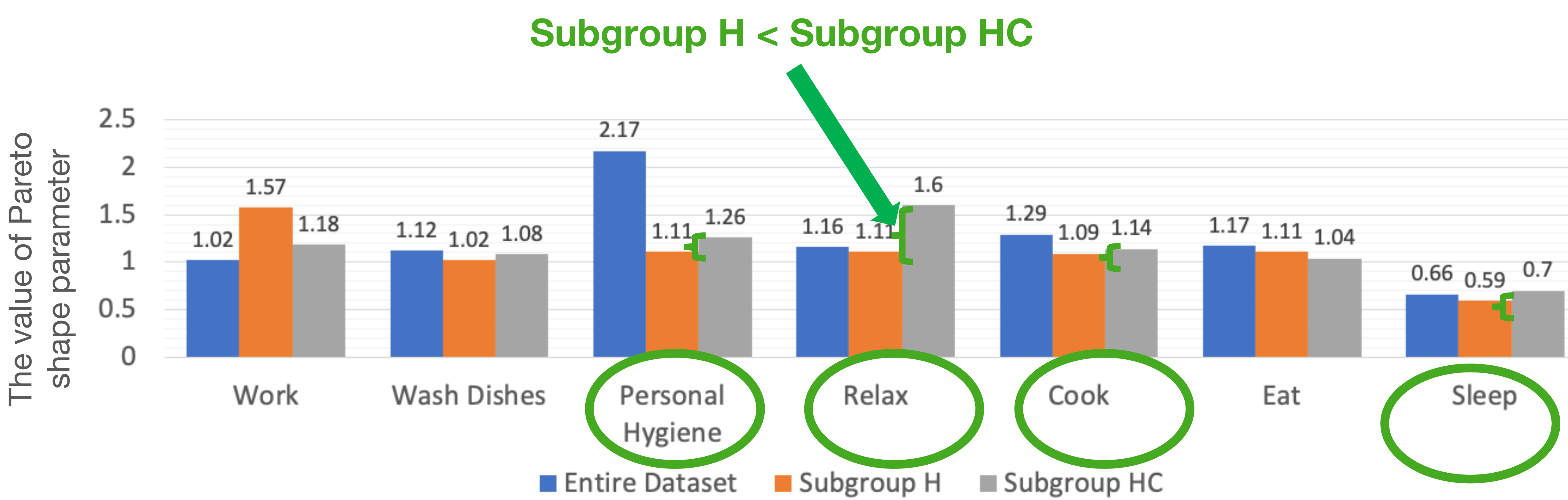
3. **Individual-based analysis on a single home with a more complete set of activities**

    An individual whose health condition is under watch

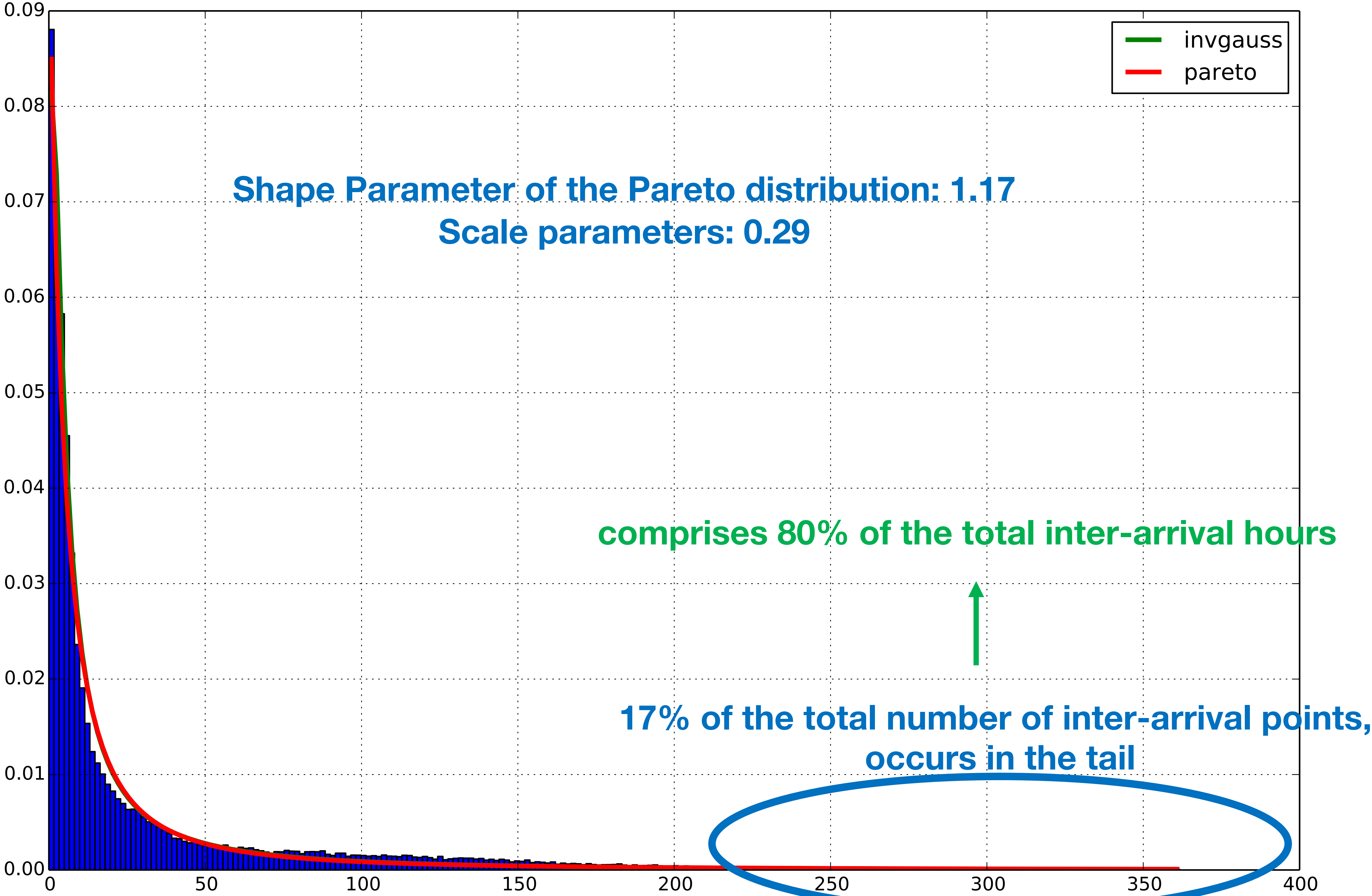
Results - Population-Wide Analysis



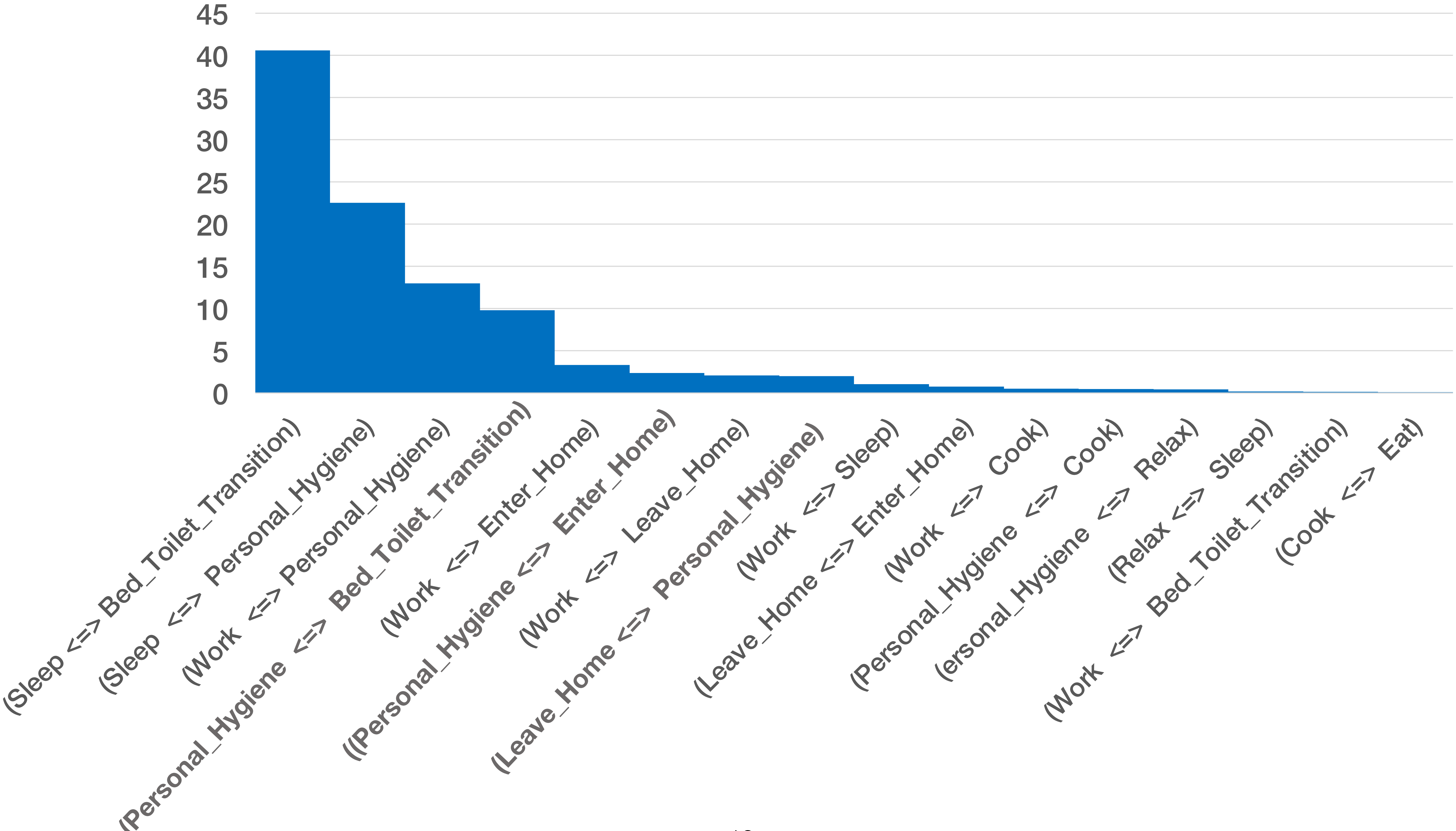
Results - Group-Wide Analysis



Results - Single-Person Analysis



Results - Single-Person Analysis



Mathematical Model of Indoor Behavior – Conclusions

1. **Indoor behavior can be formally modeled by many standard distributions**
2. **Pareto distribution fits many activity inter-arrival times**
3. **Thus we can use 80/20 rule to describe properties of the observed behavior**

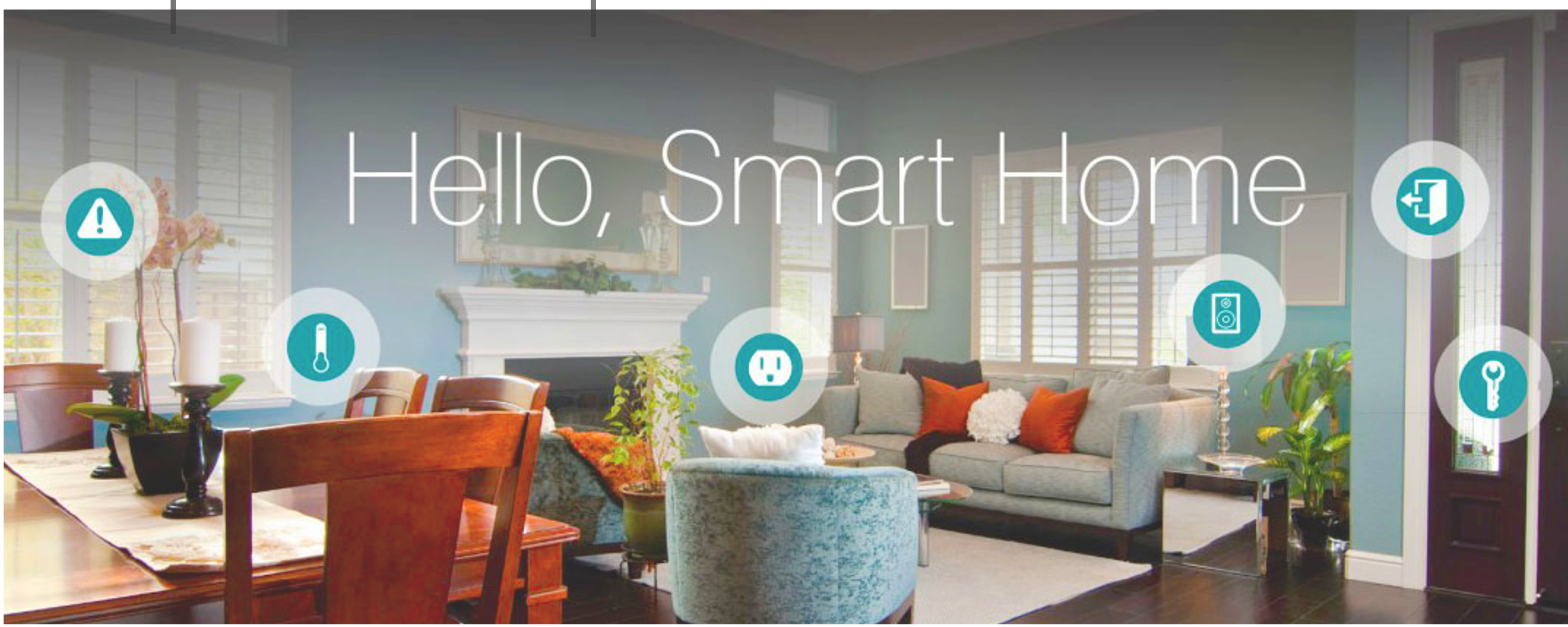
Contributions

1. **Indoor behavior can be modeled by a Pareto distribution**
   - Innovation: heavy-tailed distribution (***Pareto distribution***)
   - Impact: the Pareto distribution and its properties, the 80/20 rule, can be used to model various human behavior patterns

Step 2

Mathematical Model of Indoor Behavior

Markov Model of Indoor Behavior



Markov Model of Indoor Behavior - Hypothesis

1. Examine the question of whether a person's indoor activities are Markovian.

2. Identify the Markov order that best fits behavior-driven sensor data.

3. Select the orders of Markov models that best capture resident activities.

4. Perform this process on different population subgroups.
   (health status on formal models of human behavior)

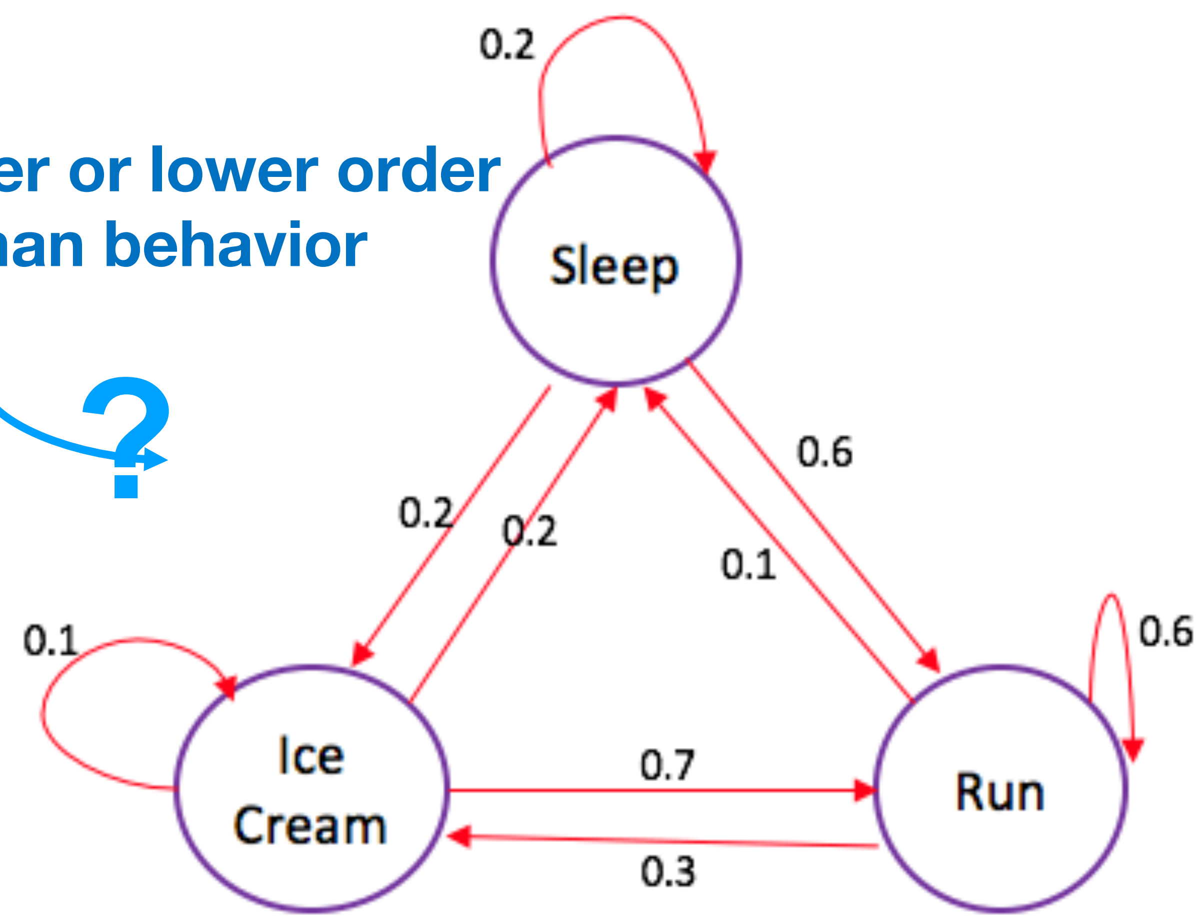
Markov Model - Definition

1st order Markov chain:

Examples: Weather Prediction; Gambling; Driver's Driving Patterns (Markov Decision Process);

higher-order or lower order of human behavior

?



Markov Model of Indoor Behavior – Markovian

1. Inter-arrival times of indoor activities can be described by Pareto distribution

2. A Pareto distribution is a special case of Lévy flight

3. A Lévy flight is a Markovian stochastic process

$\downarrow$

indoor resident behavior captured by sensors is a Markov process

Markov Model of Indoor Behavior – Order Selection

Current measures

$$\mathrm{AIC} = -2\log(\mathcal{L}(\hat{\theta}|\mathrm{data})) + 2\mathrm{K}$$

Good if N/K < 40

$$\cancel{\mathrm{AIC_c} = AIC + \frac{\mathrm{K}(\mathrm{K}+1)}{\mathrm{N}-\mathrm{K}-1}}$$

$$\cancel{\mathrm{CAIC} = -2\log(\mathcal{L}(\hat{\theta}|\mathrm{data})) + \mathrm{K}(\log(N)+1)}$$

$$\mathrm{BIC} = -2\log(\mathcal{L}(\hat{\theta}|\mathrm{data})) + \mathrm{K}\log(\mathrm{N})$$

$$\mathrm{HQIC} = -2\log(\mathcal{L}(\hat{\theta}|\mathrm{data})) + 2\mathrm{K}\log(\log(\mathrm{N}))$$

$\hat{\theta}$: maximum likelihood estimate

$\mathcal{L}(\hat{\theta}|\mathrm{data})$: likelihood of the model

K : # of model parameters

N : sample size

Markov Model of Indoor Behavior – Order Selection

Current measures

under-estimate

N	AIC			BIC			HQIC		
	<	=	>	<	=	>	<	=	>
500	97.98	2.02	0.00	100.00	0.00	0.00	100.00	0.00	0.00
1000	0.10	99.90	0.00	100.00	0.00	0.00	57.36	42.64	0.00
1500	0.01	99.99	0.00	81.07	18.93	0.00	6.13	93.87	0.00
2000	0.00	100.00	0.00	25.21	74.79	0.00	0.56	99.44	0.00
2500	0.00	100.00	0.00	4.39	95.61	0.00	0.05	99.95	0.00
3000	0.00	100.00	0.00	0.40	99.60	0.00	0.00	100.00	0.00
3500	0.00	100.00	0.00	0.13	99.87	0.00	0.00	100.00	0.00
4000	0.00	100.00	0.00	0.01	99.99	0.00	0.00	100.00	0.00
4500	0.00	[illegible]	[illegible]	[illegible]	100.00	0.00	0.00	100.00	0.00
5000	0.00	100.00	[illegible]	0.00	100.00	0.00	0.00	100.00	0.00
5500	0.00	97.89	2.11	0.00	100.00	0.00	0.00	100.00	0.00
6000	0.00	100.00	0.00	0.00	100.00	0.00	0.00	100.00	0.00
6500	0.00	100.00	0.00	0.00	100.00	0.00	0.00	100.00	0.00

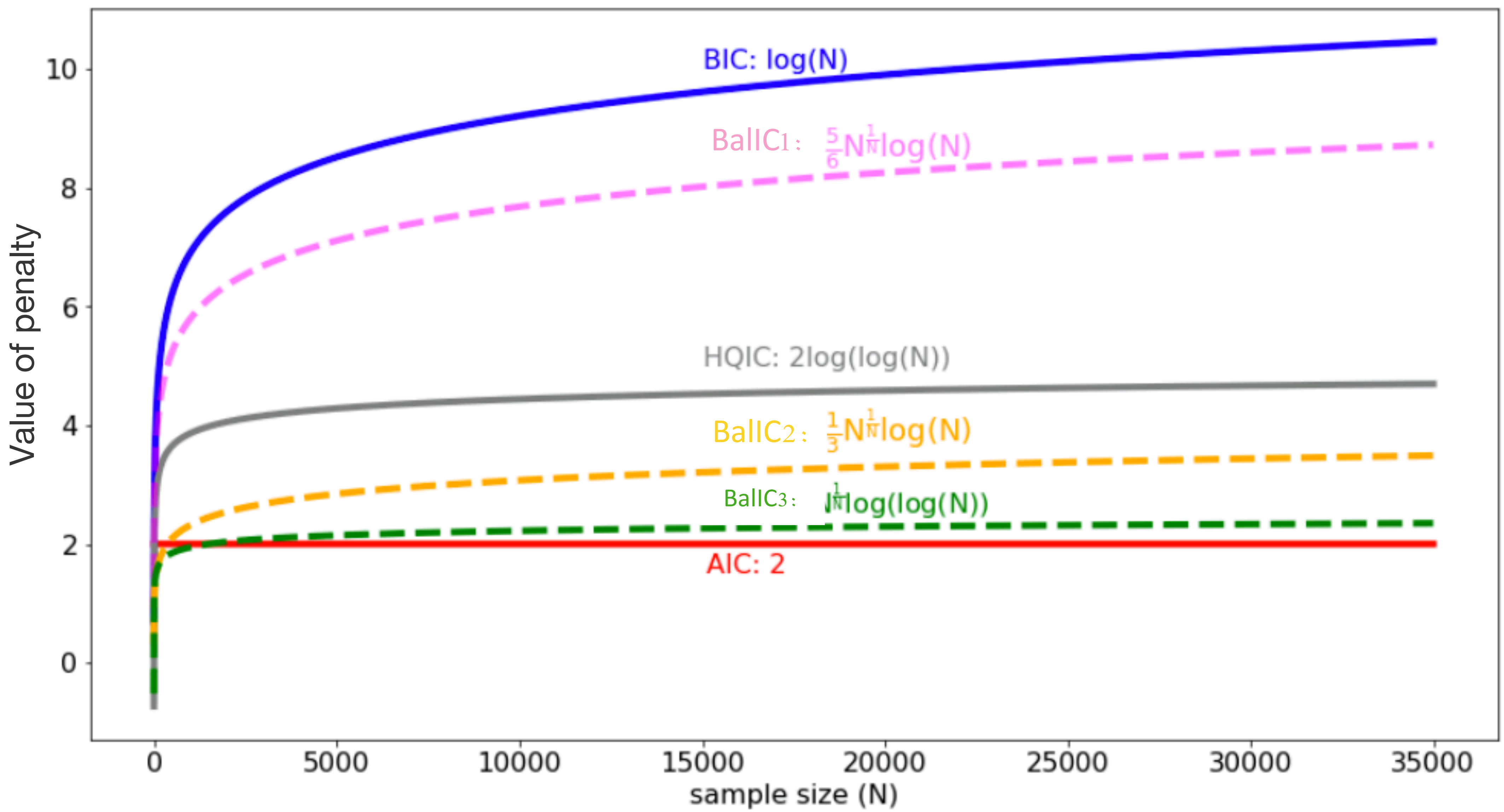
over-estimate

New Measure: BalIC (Balanced Information Criteria)

$$\mathrm{BalIC_1} = -2\log(\mathcal{L}(\hat{\theta}|\mathrm{data})) + \frac{5}{6}\mathrm{K}N^{\frac{1}{N}}\log(\mathrm{N})$$

$$\mathrm{BalIC_2} = -2\log(\mathcal{L}(\hat{\theta}|\mathrm{data})) + \frac{1}{3}\mathrm{K}N^{\frac{1}{N}}\log(\mathrm{N})$$

$$\mathrm{BalIC_3} = -2\log(\mathcal{L}(\hat{\theta}|\mathrm{data})) + \mathrm{K}N^{\frac{1}{N}}\log(\log(\mathrm{N}))$$



Performance of Current and New Measures

N	AIC			BIC		HQIC		BalIC_1		BalIC_2		BalIC_3
	$<$	$=$	$>$	$<$	$=$	$<$	$=$	$<$	$=$	$<$	$=$	$=$
500	0.10	99.90	0.00	100.00	0.00	55.16	44.84	99.60	0.40	24.22	75.78	100.00
1000	0.01	99.99	0.00	83.35	16.65	5.84	94.16	45.46	54.54	2.29	97.71	100.00
1500	0.00	100.00	0.00	23.51	76.49	0.53	99.47	6.23	93.77	0.21	99.79	100.00
2000	0.00	100.00	0.00	4.32	95.68	0.05	99.95	0.75	99.25	0.02	99.98	100.00
2500	0.00	100.00	0.00	0.67	99.33	0.00	100.00	0.07	99.93	0.00	100.00	100.00
3000	0.00	100.00	0.00	0.06	99.94	0.00	100.00	0.01	99.99	0.00	100.00	100.00
3500	0.00	99.93	0.07	0.01	99.99	0.00	100.00	0.00	100.00	0.00	100.00	100.00
4000	0.00	100.00	[illegible]	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	100.00
4500	0.00	[illegible]	[illegible]	[illegible]	100.00	[illegible]	[illegible]	0.00	100.00	0.00	100.00	[illegible]
5000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	100.00	[illegible]	[illegible]	[illegible]
5500	0.00	100.00	0.00	0.00	100.00	[illegible]	[illegible]	0.00	100.00	0.00	[illegible]	[illegible]
6000	0.00	100.00	0.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	100.00
6500	0.00	100.00	0.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	100.00

over-estimate

under-estimate

good

Markov Model of Indoor Behavior – Smart Home Data

2011-06-13 10:23:15.231817 Kitchen OFF Wash_Dishes
2011-06-13 10:23:16.156653 Kitchen ON Wash_Dishes
2011-06-13 10:23:17.277543 Kitchen OFF Wash_Dishes
2011-06-13 10:23:17.804728 Kitchen ON Wash_Dishes
2011-06-13 10:23:18.459437 Kitchen ON Cook
2011-06-13 10:23:19.612122 Kitchen OFF Cook
2011-06-13 10:23:19.739553 Kitchen OFF Cook
2011-06-13 10:23:20.253069 Kitchen ON Cook

Groups	The Selected Order Based on Different Measures			
	AIC	BIC	HQIC	BalIC_3
entire dataset	4	3	4	4
young	3	3	3	3
middle aged	3	2	3	3
middle aged double	3	2	3	3
middle aged single	3	2	2	3
middle aged healthy	3	2	3	3
middle aged health complaints	2	1	2	2
middle aged and senior	3	2	2	3
seniors	4	3	3	4
senior single	3	3	3	3
senior double	3	3	3	3
senior single healthy	3	3	3	3
senior single health complaints	3	3	3	3

Markov Model of Indoor Behavior – Conclusions

- Our new measure, $BalIC_3$, outperforms the others
    - dimension consistency
    - no under-estimating problem based on the experiments' results
- Human behavior can be expressed with a Markov chain (it is Markovian)
- Markov chain orders (and thus the observed human behavior complexity) depend on the nature of the sensor data and its level of detail

Contributions

1. **Indoor behavior can be modeled by a Pareto distribution**
   - Innovation: heavy-tailed distribution (***Pareto distribution***)
   - Impact: the Pareto distribution and its properties, the 80/20 rule, can be used to model various human behavior patterns

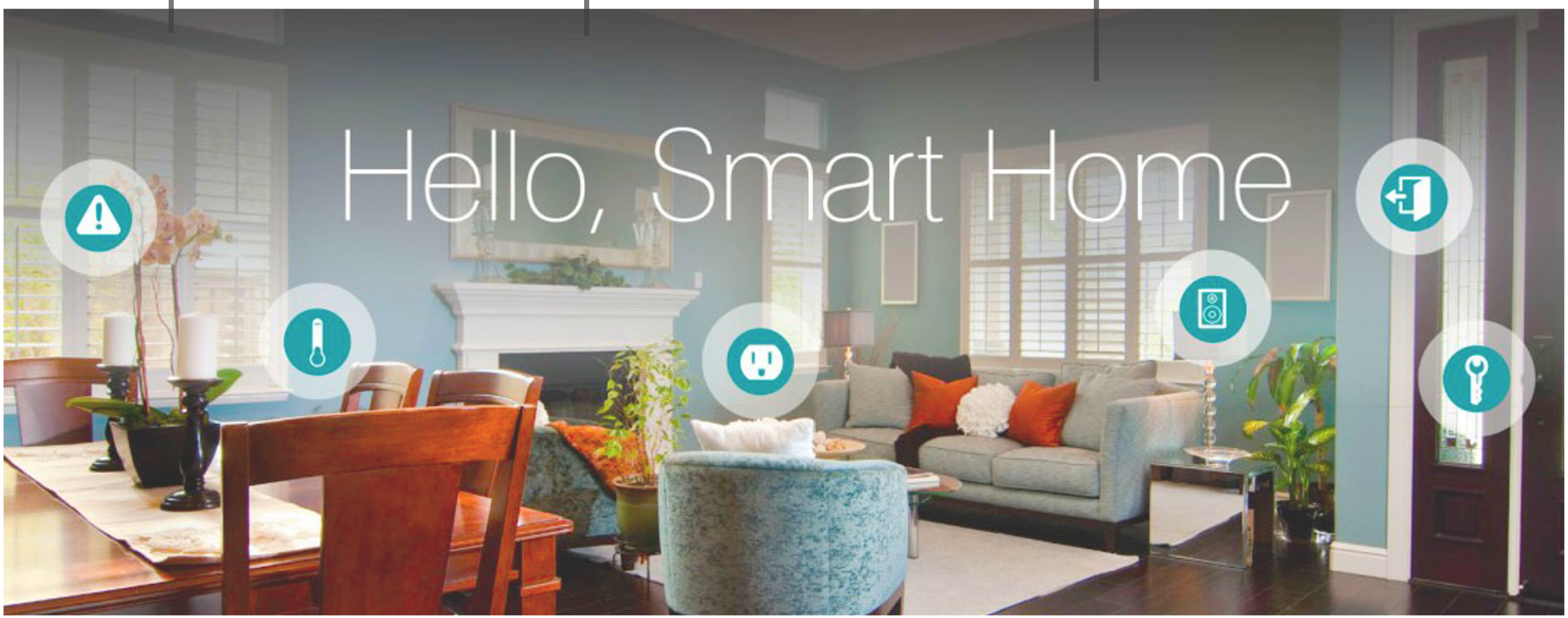
2. **Proposed a new measure for model selection**
   - Innovation: ***a new measure*** to balance the tradeoff between under/over-estimating the true Markov order
   - Impact: understand the complexity of sensor-observed human indoor behavior

Step 3

Mathematical Model of Indoor Behavior

Markov Model of Indoor Behavior

Inverse Reinforcement Learning Model of Indoor behavior



Inverse Reinforcement Learning (IRL) of Indoor Behavior - Hypothesis

1. Model navigation patterns as a Markov decision process

2. Design the spatio-temporal features of indoor navigations

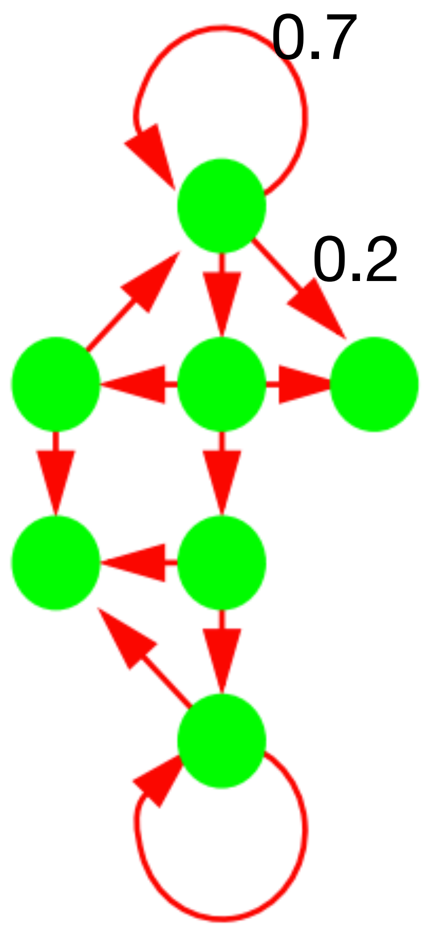
3. Learn a behavior strategy that is consistent with the observed movement patterns

4. Analyze behavior patterns and differences on different subgroups (eight actual smart homes)

Inverse Reinforcement Learning (IRL) of Indoor Behavior - Definitions

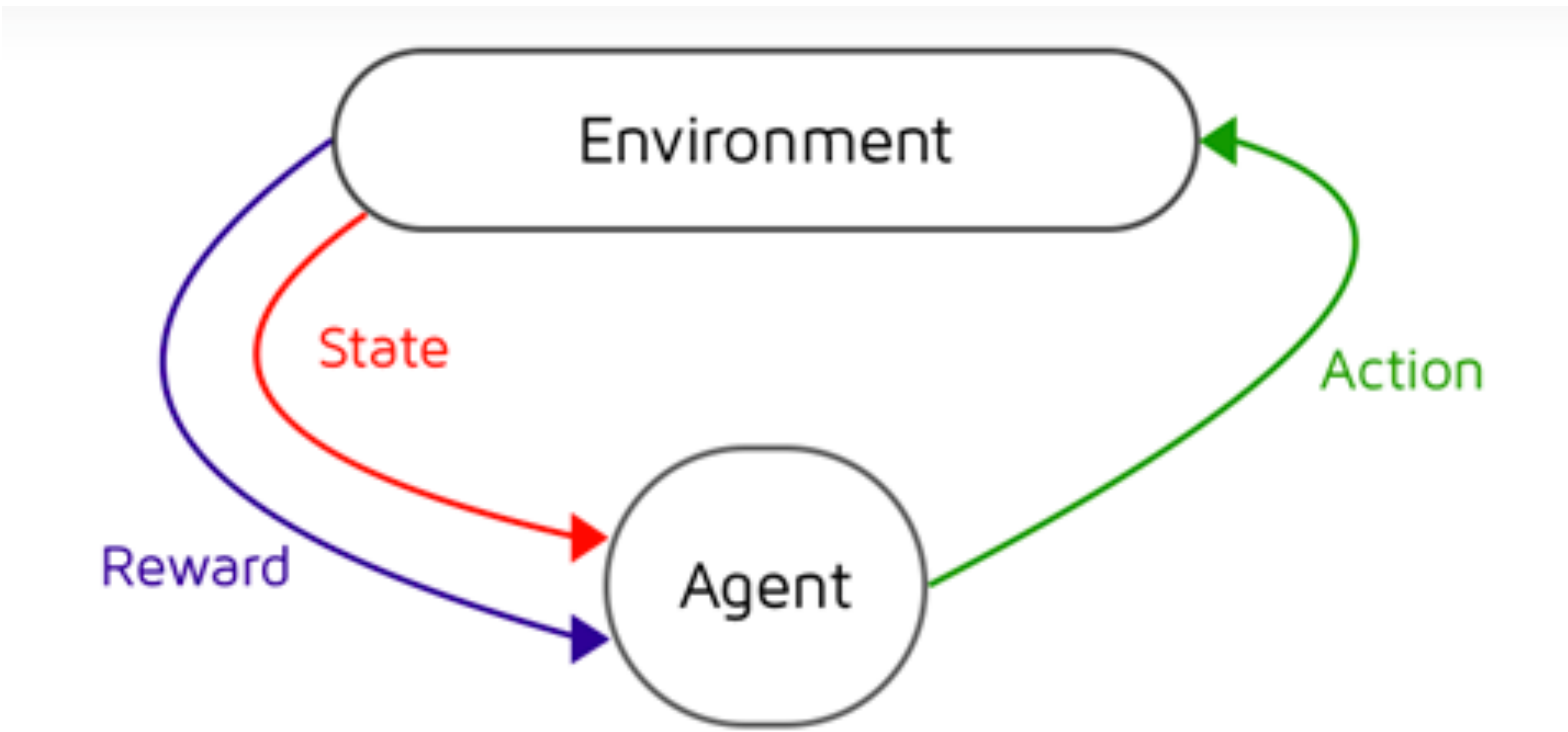
A *Markov Decision Process* (MDP) model contains:

- A set of possible states *S*
- A set of possible actions *A*
- A real valued reward function *R(s, a)*
- A transition matrix *P*

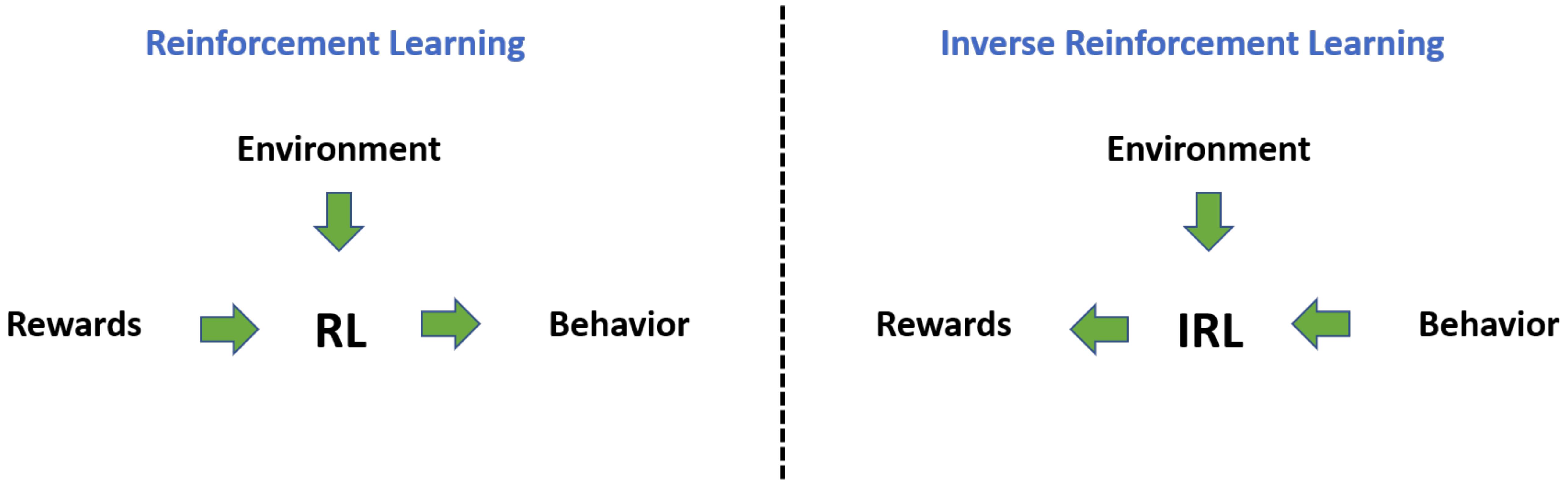


Inverse Reinforcement Learning (IRL) of Indoor Behavior - RL

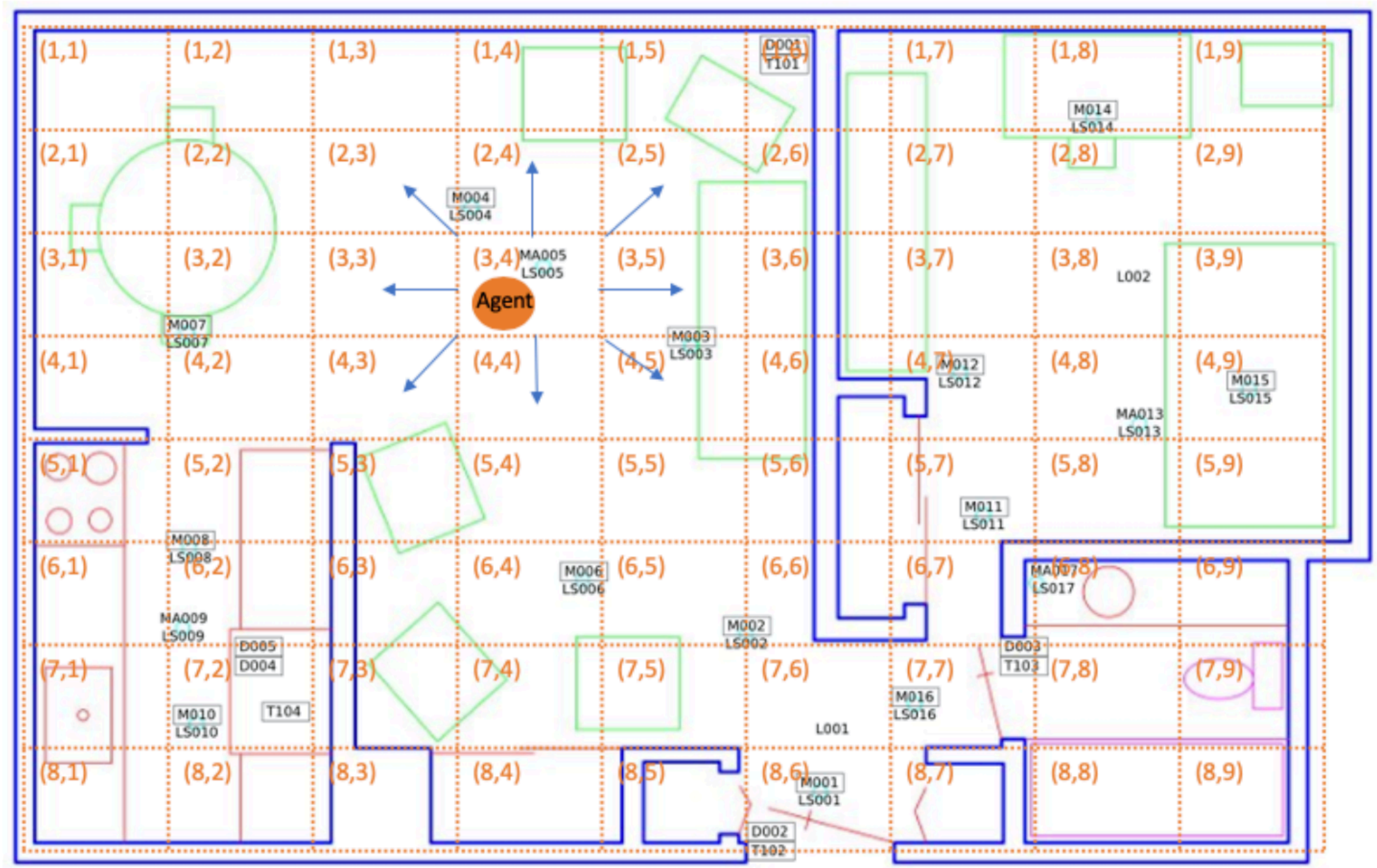
Reinforcement Learning (RL):



Inverse Reinforcement Learning (IRL) of Indoor Behavior – RL v.s. IRL



Inverse Reinforcement Learning (IRL) of Indoor Behavior – Navigation



d_toilet	d_bathroom_sink	d_livingroom_chair	d_kitchen_sink
d_bedroom	d_kitchen	d_livingroom	d_hallway
d_stove	d_office_chair	o_toilet	o_livingroom_chair
o_kitchen_sink	o_office_chair		

Inverse Reinforcement Learning (IRL) of Indoor Behavior – Algorithm

Algorithm : Resident Relative Entropy IRL (RRE-IRL)

input : a set of residents' trajectories $\mathcal{T}$

output : the reward function (the preference vector) θ

initialize: the preference vector θ

while $(\frac{\hat{\partial}}{\partial \theta_i} g(\theta) > \epsilon_i)$ **do**

calculate $\frac{\hat{\partial}}{\partial \theta_i} g(\theta) = \hat{\mu_i} - \frac{\sum_{\tau \in \mathcal{T}_N^\pi} \frac{A_{tran}}{\pi(\tau)} \exp(\theta \cdot \phi^\tau) \phi_i^\tau}{\sum_{\tau \in \mathcal{T}_N^\pi} \frac{A_{tran}}{\pi(\tau)} \exp(\theta \cdot \phi^\tau)} - \alpha_i \cdot \epsilon_i,$

update $\theta_i \longleftarrow \theta_i + \alpha_i \cdot \frac{\hat{\partial}}{\partial \theta_i} g(\theta)$

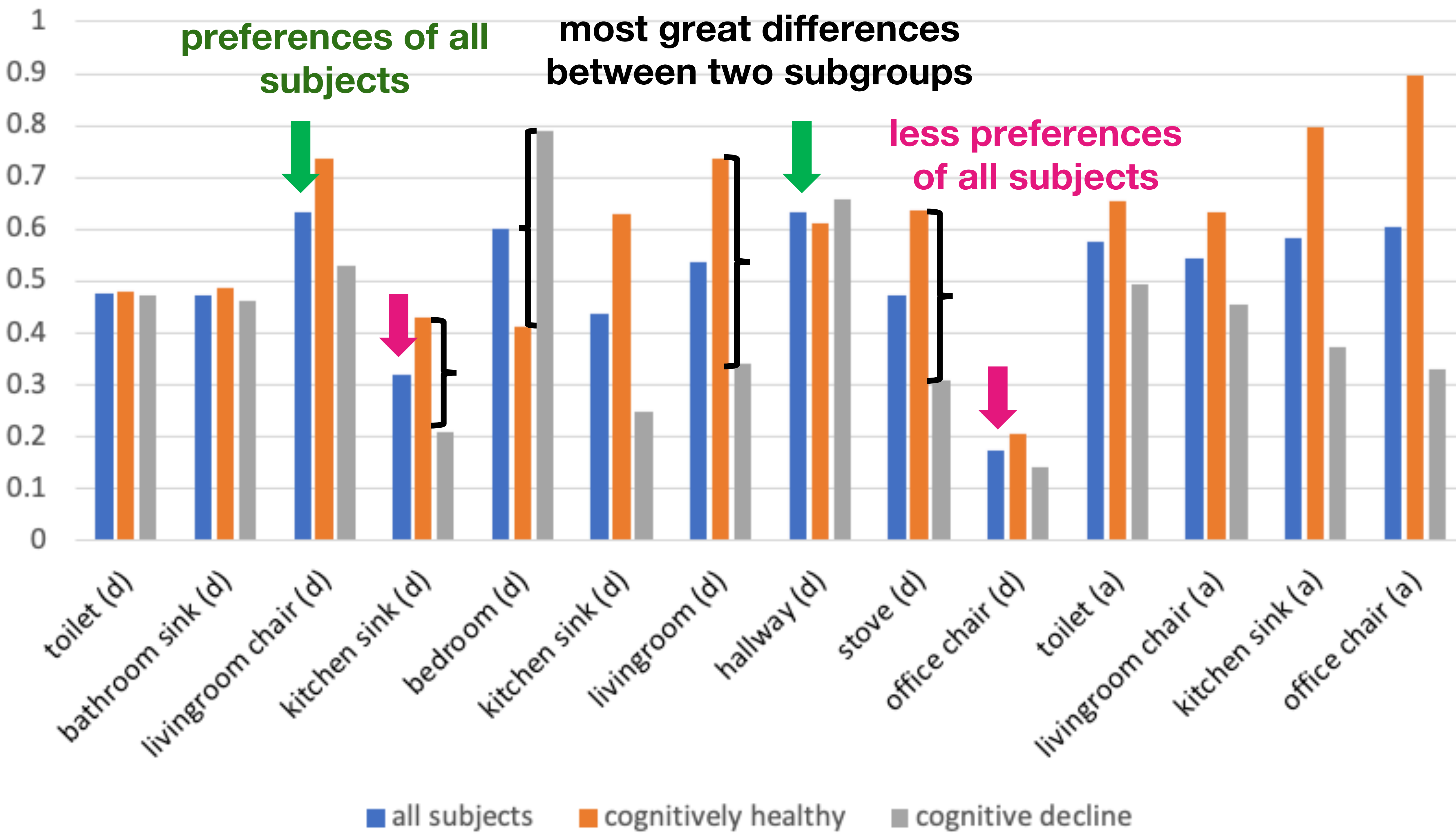
end

return θ

Inverse Reinforcement Learning (IRL) of Indoor Behavior – Datasets

Group	ID	Health Diagnosis	#Sensors	Duration of Data Collection	Number of Month-long Samples
Cognitive decline	Home 1	Mild Cognitive Impairment (MCI)	21 downward-facing motion (motion); 2 motion area (ma)	843 days	26
	Home 2	MCI	19 motion; 2 ma	223 days	7
	Home 3	MCI	26 motion; 0 ma	682 days	22
	Home 4	MCI, Early dementia	11 motion; 2 ma	149 days	5
Cognitively healthy	Home 5	Healthy	13 motion; 1 temperature	1788 days	56
	Home 6	Healthy	13 motion	1591 days	49
	Home 7	Healthy	18 motion; 2 ma	379 days	12
	Home 8	Healthy	10 motion; 1 ma	969 days	31

Inverse Reinforcement Learning (IRL) of Indoor Behavior – Results



Inverse Reinforcement Learning (IRL) of Indoor Behavior – Conclusions

1. Indoor navigation can be described as a Markov decision process

2. Spatial and temporal features can be fused into one model

3. Indoor behavior strategy can be learned via IRL

4. Cognitively declined residents can be distinguished from healthy population

Contributions

1. **Indoor behavior can be modeled by a Pareto distribution**
   - Innovation: heavy-tailed distribution (***Pareto distribution***)
   - Impact: the Pareto distribution and its properties, the 80/20 rule, can be used to model various human behavior patterns

2. **Proposed a new measure for model selection**
   - Innovation: ***a new measure*** to balance the tradeoff between under/over-estimating the true Markov order
   - Impact: understand the complexity of sensor-observed human indoor behavior

3. **Proposed inverse reinforcement learning to learn indoor behavior**
   - Innovation: ***1st study*** of IRL to model indoor behavior
   - Impact: extract quantitative and qualitative measures of human behavior

4. **Distinguished population subgroups based on ambient sensor data**
   - Innovation: ***population-level behavior analysis*** *via ambient sensor data*
   - Impact: compare behavior for healthy older adults with behavior of older adults with chronic health conditions

Future Work

Extend current work of smart home

- Develop models include sufficient information of *behavior fluctuations* (e.g., IRL, adversarial IRL, and generative adversarial imitation learning)
- Investigate temporal and spatial properties in *transfer learning scenarios*.

Mining rare behavior

- Propose a new framework for rare behavior detection
- Combine ideas from both adversarial training and classification

Model ICU patient’s behavior

- Patients’ mobility patterns
- ICU instruments’ data (time series data)

Publications

Journal Articles

Lin, B., Cook, D., Schmitter-Edgecombe, M., 2019. Using Continuous Sensor Data to Formalize a Model of In-Home Activity Patterns. Journal of Ambient Intelligence and Smart Environments.

Huangfu, Y., Lima, N.M., O'Keeffe, P.T., Kirk, W.M., Lamb, B.K., Pressley, S.N., Lin, B., Cook, D.J., Walden, V.P. and Jobson, B.T., 2019. Diel variation of formaldehyde levels and other VOCs in homes driven by temperature dependent infiltration and emission rates. Building and Environment, 159, p.106153.

Ghods, A., Caffrey, K., Lin, B., Fraga, K., Fritz, R., Schmitter-Edgecombe, M., Hundhausen, C. and Cook, D.J., 2018. Iterative Design of Visual Analytics for a Clinician-in-the-loop Smart Home. IEEE Journal of Biomedical and Health Informatics.

Kirk, W.M., Fuchs, M., Huangfu, Y., Lima, N., O'Keeffe, P., Lin, B., Jobson, T., Pressley, S., Walden, V., Cook, D. and Lamb, B.K., 2018. Indoor air quality and wildfire smoke impacts in the Pacific Northwest. Science and Technology for the Built Environment, 24(2), pp.149-159.

Lin, B., Huangfu, Y., Lima, N., Jobson, B., Kirk, M., O’Keeffe, P., Pressley, S.N., Walden, V., Lamb, B. and Cook, D.J., 2017. Analyzing the relationship between human behavior and indoor air quality. Journal of Sensor and Actuator Networks, 6(3), p.13.

Conference Proceedings

A. Musser, Lin, B., Cook, D., B. Jobson, M. Kirk, N. Lima, P. O'Keeffe, S. Pressley, V. Walden, Y. Huangfu, and B. Lamb. 2018. Indoor air toxic gases levels in a net-zero energy house under multiple ventilation system settings, the 15th Conference of the International Society of Indoor Air Quality and Climate.

A. Musser, Lin, B., Cook, D., B. Jobson, M. Kirk, N. Lima, P. O'Keeffe, S. Pressley, V. Walden, Y. Huangfu, and B. Lamb, 2018. The major role of temperature on indoor concentrations of air toxic VOCs in 9 houses based on in-situ high time resolution measurements, the 15th Conference of the International Society of Indoor Air Quality and Climate.

A. Musser, Lin, B., Cook, D., B. Jobson, M. Kirk, N. Lima, P. O'Keeffe, S. Pressley, V. Walden, Y. Huangfu, and B. Lamb, 2018. Simulations of indoor air quality based on future climate conditions. Conference of the International Society of Indoor Air Quality and Climate.

Posters

Lin, B., 2019. “Constructing an Ecologically-Valid Formal Markovian Model of Human Activity Patterns”, General Poster Session at Grace Hopper Celebration, Orlando, FL.

Lin, B., 2019. “Population Level Behavior Analysis and Its Applications in Healthcare”, Ph.D. Forum at SIAM International Conference on Data Mining Doctoral Forum, Alberta, Canada.

Lin, B., 2018 .“Identifying and modeling the patterns of human activity routines”, Research Exposition at Graduate and Professional Student Association, Pullman, WA

Thank you!