Machine Learning in Mobile and Sensor Networks

Qiang Yang, 杨强

http://www.cs.ust.hk/~qyang

Hong Kong University of Science and Technology

Joint Work with Dr. Rong Pan and students: Jie Yin, Xiaoyong Chai, Jeff Pan and Dou Shen Kangheng Wu

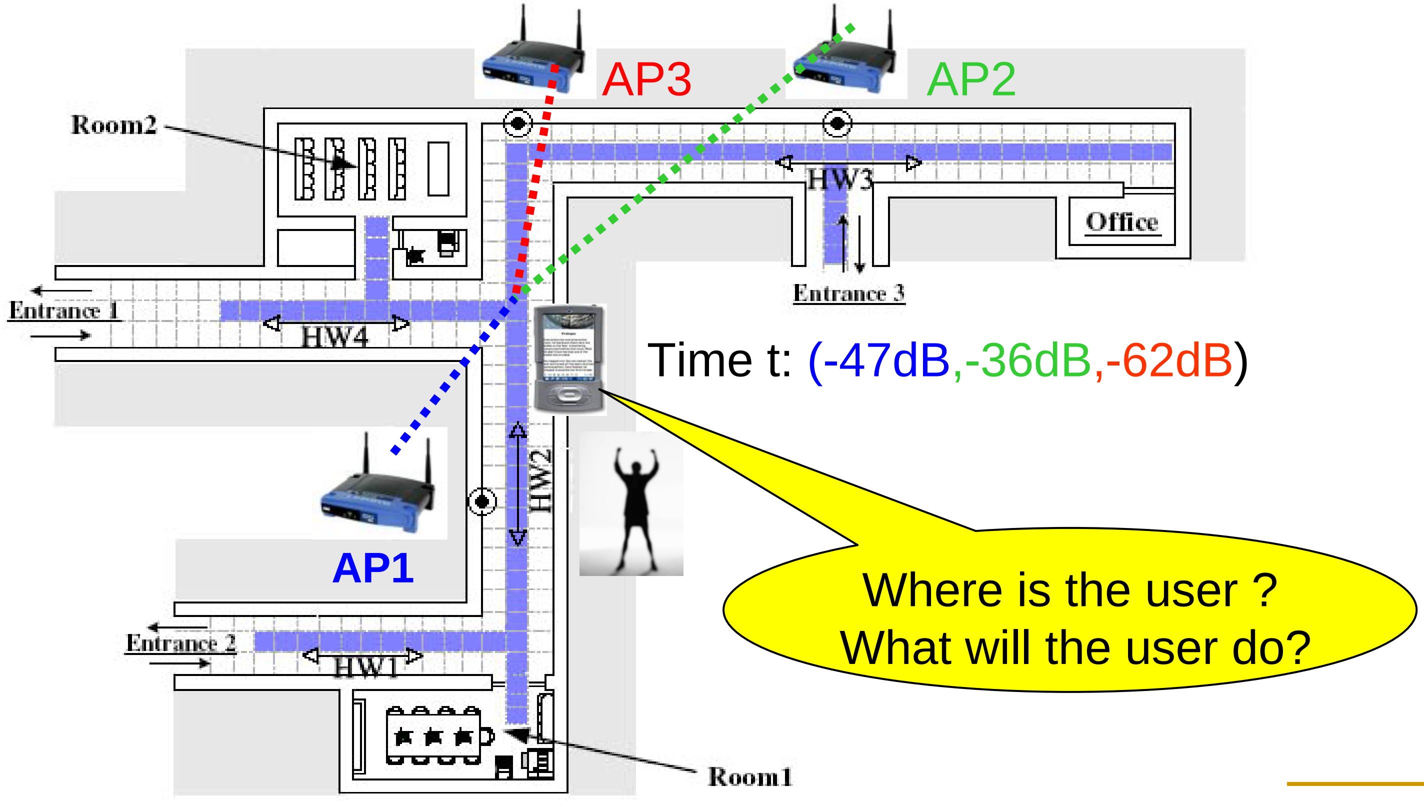
Context-Aware Computing: A Solution

- A central theme in context-aware computing is to build *predictive models of human behavior*
  - Where is the user? (location estimation)
  - What is her ultimate goal? (activity recognition)

(courtesy CMU)

Problem Domain: Wireless Enviornment

- A user with a mobile device walks in an indoor wireless environment (802.11b WLAN)



What are the learning problems?

- Learning to classify low level sequences
  - Location, actions and goals
  - Semi-supervised Learning
- Learning to segment sequences into discrete activities
- New Machine Learning Problem: Data Migration
- Semi-supervised Learning:
  - To reduce calibration efforts
- Training data migration problem
- Distributed Learning problem

Machine Learning in Pervasive Computing: A Video Demonstration

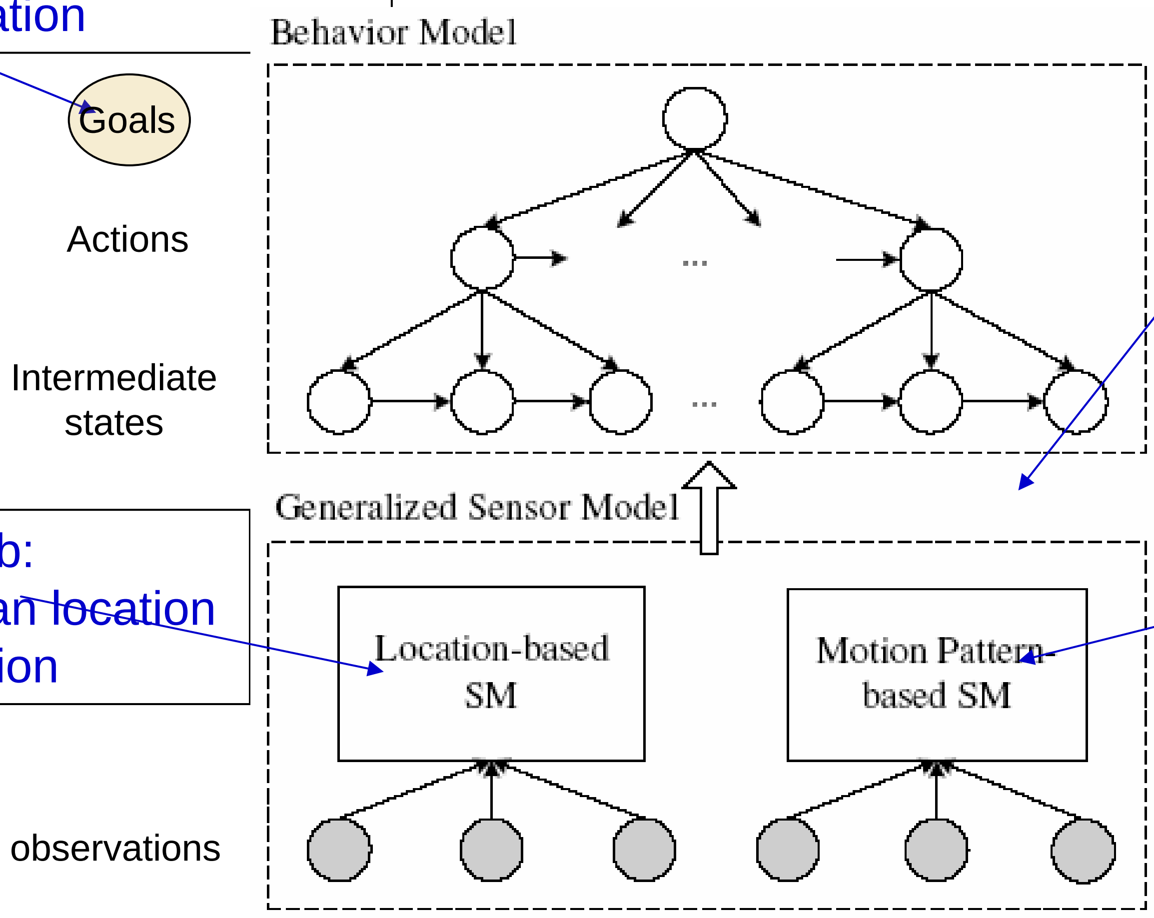
- http://www.cs.ust.hk/~qyang/sensor2.wmv

Probabilistic Goal Recognition: Architecture

ML Prob: semi-supervised classification

ML Problem: Distributed Learning

ML Prob: Model Migratoin



ML Prob: Bayesian location estimation

ML Prob: segmentation

ML Problem: Bayesian Learning on Location

- Two phases: offline Training and online Localization
- **Offline phase** – collect samples to build a mapping function l=f(s) from signal space S to location space L where s ∈ S and l ∈ L

Loc.	Time	(AP1,AP2,AP3)
(1,0)	1s	(-60,-50,-40) dB
(2,0)	2s	(-62,-48,-35) dB
…..	….	(… , … , …)dB
(9,5)	9s	(-50,-35,-42) dB

Training…

l=f(s) where
s ∈ S and l ∈ L

- **Online phase** – given a new signal s , estimate the most likely location l from l=f(s)
  - s* = (-60,-49,-36)dB , compute f(s)=l as estimated location

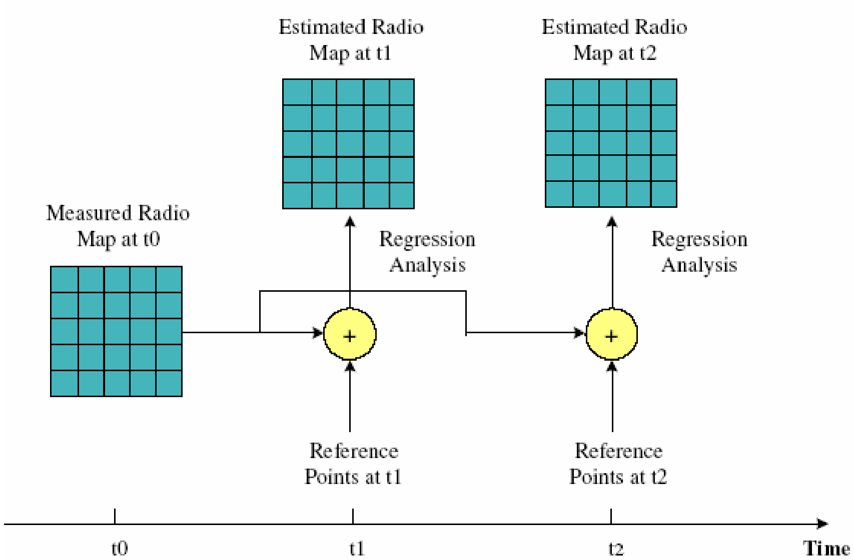
Related Work in Building the Sensor Model

- Microsoft Research's RADAR [Bahl and Padmanabhn, 2000]
  - K-Nearest-Neighbor Method
  - Offline - for each location, compute the mean signal
  - Online – estimate location with KNN and triangulation
- Maryland's Horus [Moustafa Youssef et al. ,2003], Rice U. Robotics-based System [Ladd et al., 2002, 2004]
  - Maximum Likelihood Estimation
  - Offline - for each location, train the **Radio Map** of each AP at each location
  - Online - apply Bayes' rule and (user dynamics) for estimation

- Major issues
  - Radio map changes with time
    - How to adapt for other times?
    - Yin et al. [IEEE Percom 2005]
  - Reduce human calibration by user unlabelled traces
    - Semi-supervised Learning
    - Chai et al. [IEEE Percom 2005]

ML Problem: Data and Model Migration

- Key idea: collect radio map once, and then adapt the radio map using reference points and a regression analysis



- During initial time period *t0*:
  - At each location *i*, we learn a predictive function f_{ij} for the *jth* AP, based on the reference points

$$S_{est}(i, j, t_0) = f_{ij}(\overrightarrow{r(t_0)})$$

- During the online phase (time period *t*):

$$S_{est}(i, j, t) = f_{ij}(\overrightarrow{r(t)})$$

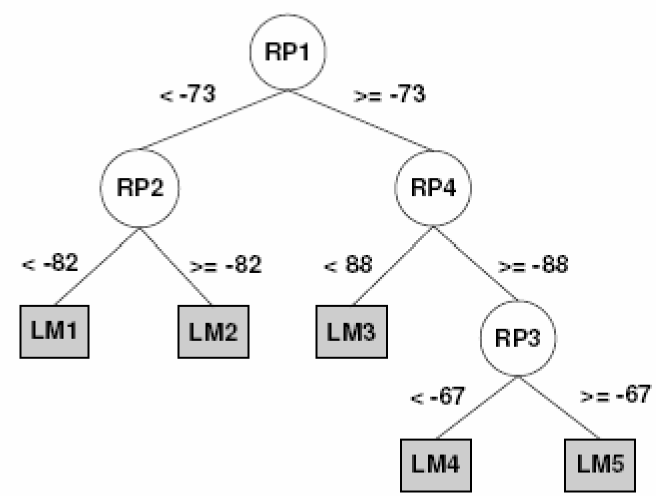
$$D_i(t) = \sqrt{\sum_{j=1}^{p} (s_j(t) - ss_j(t))^2}$$

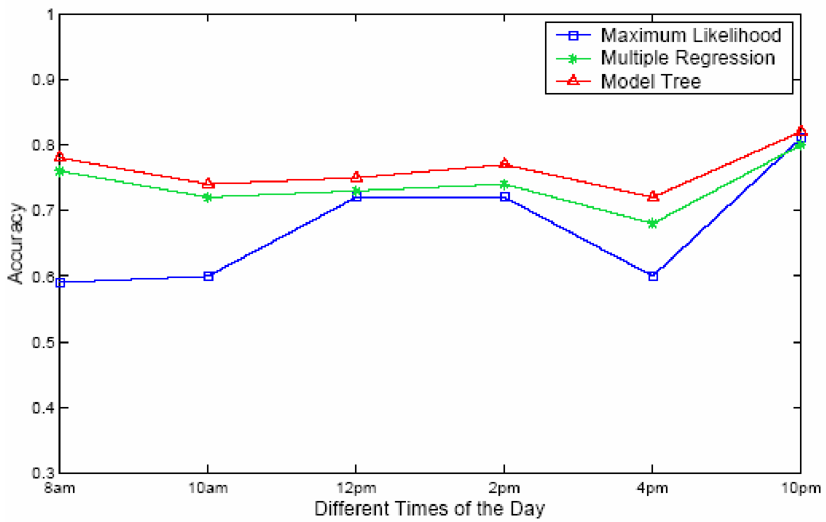
Critical Issue: learn predictive function *fij*

- -- mapping between the signal-strength values received by the mobile client and the reference points.
- Two algorithms via regression analysis
  - A multiple-regression based algorithm (Linear Model LM)

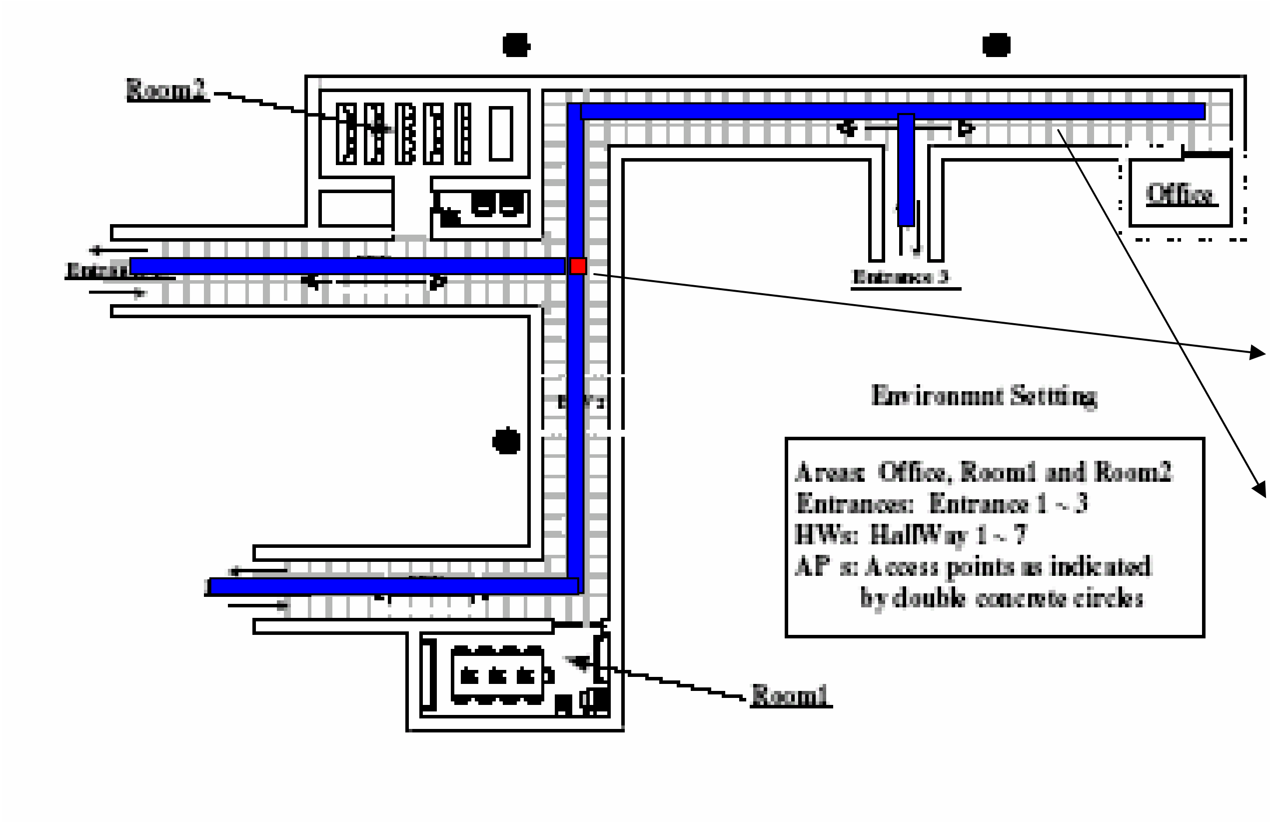
$$s = a_0 + a_1 * r_1 + a_2 * r_2 + \ldots \varepsilon$$

  - A model-tree based algorithm (see result at 1.5m)





ML Problem: Semi-Supervised Learning



Sampling time (N_s)

Number of locations sampled (N_l)

Total amount of calibration effort: $N_s \times N_l$

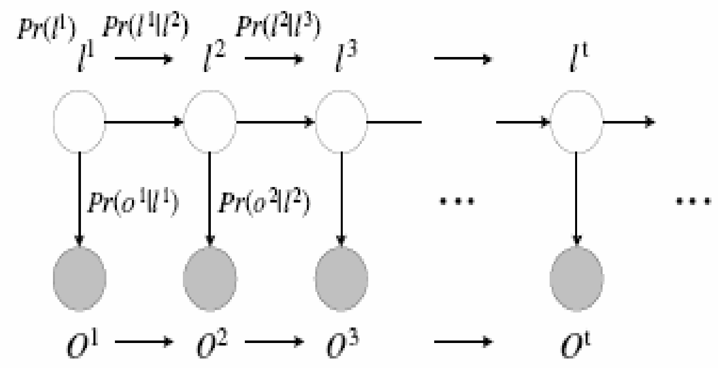
Semi-supervised Learning Framework: Using Unlabeled Traces to Improve the Radio Map

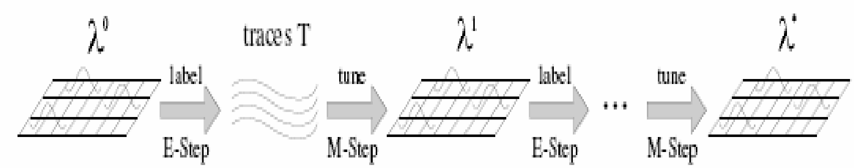
- What is a user trace
  - – A sequence of signal strength measurements recorded when a user holding a wireless device navigates in the environment

Trace #	Observation Sequences			
	t_1	t_2	...	t_k
1	$(AP_1$:-57) $(AP_2$:-33) $(AP_3$:-51)	$(AP_1$:-56) $(AP_2$:-30) $(AP_3$:-62)	$(AP_1$:-55) $(AP_2$:-36) $(AP_3$:-56)	$(AP_1$:-52) $(AP_2$:-62) $(AP_3$:-47)
2	$(AP_1$:-62) $(AP_2$:-57) $(AP_3$:-55)	$(AP_1$:-39) $(AP_2$:-41) $(AP_3$:-32)	$(AP_1$:-46) $(AP_2$:-45) $(AP_3$:-43)	$(AP_1$:-41) $(AP_2$:-43) $(AP_3$:-27)

Modeling User Traces Using Hidden Markov Model (HMM)

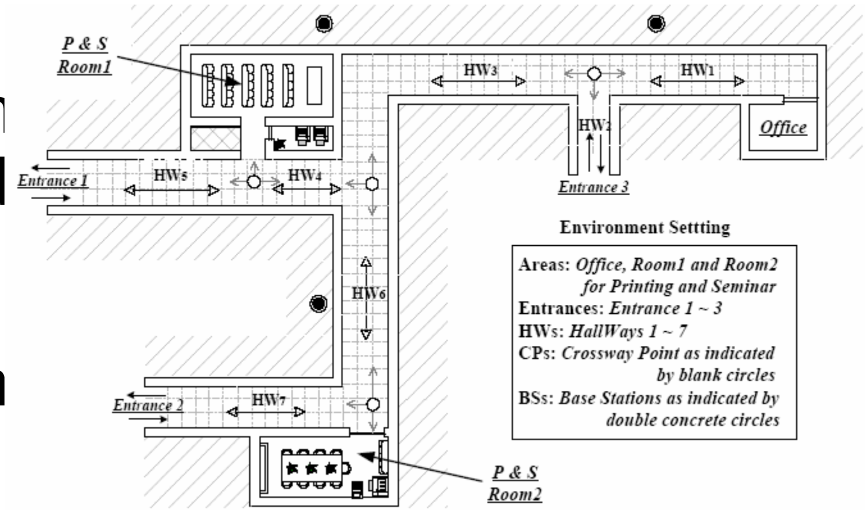
- An HMM is a quintuple <L, O, λ, A, π>
  - L: location-state space – $\{l_1, l_2, \ldots\ l_n\}$
  - O: observation space – $\{o_1, o_2, \ldots\ o_m\}$
  - λ: radio map – $\{ \Pr(o_j \mid l_i) \}$
  - A: location-state transition – $\{ \Pr(l_j \mid l_i) \}$
  - π: initial state distribution – $\{ \Pr(l_i) \}$
- HMM model parameter θ = (λ, A, π)





Experimental Setting

- The environment is modeled a space of 99 locations, each representing a 1.5-meter grid cell.
- Sensor readings contain sign strength measurements from base stations.
- Sensor model construction: 100 signal samples at each location.



99 locations

Reducing the calibration effort: result

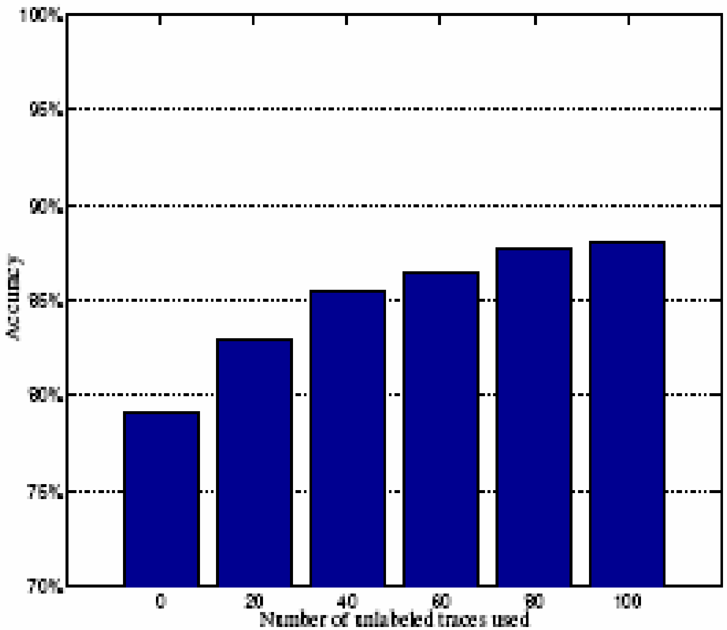


Fig. 16. Improvement achieved through using an increasing number of traces ($N_s = 20, N_l = 99$)

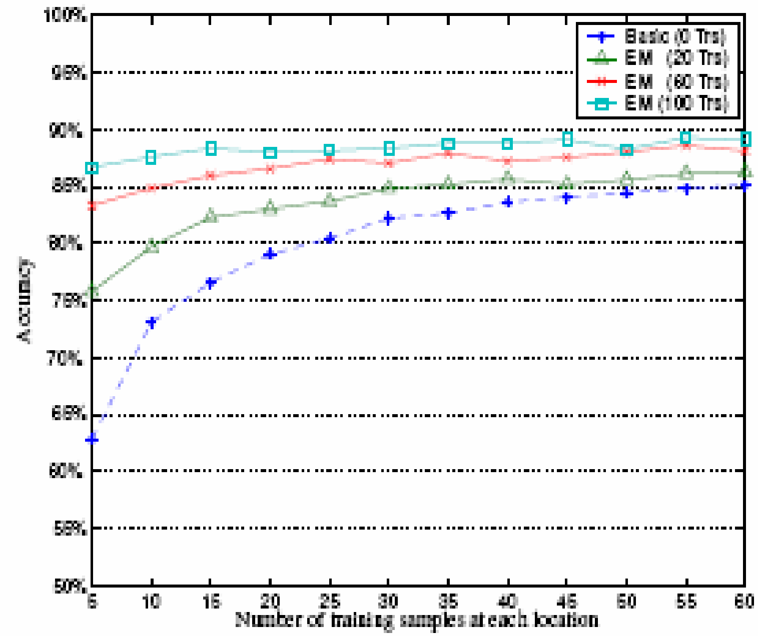
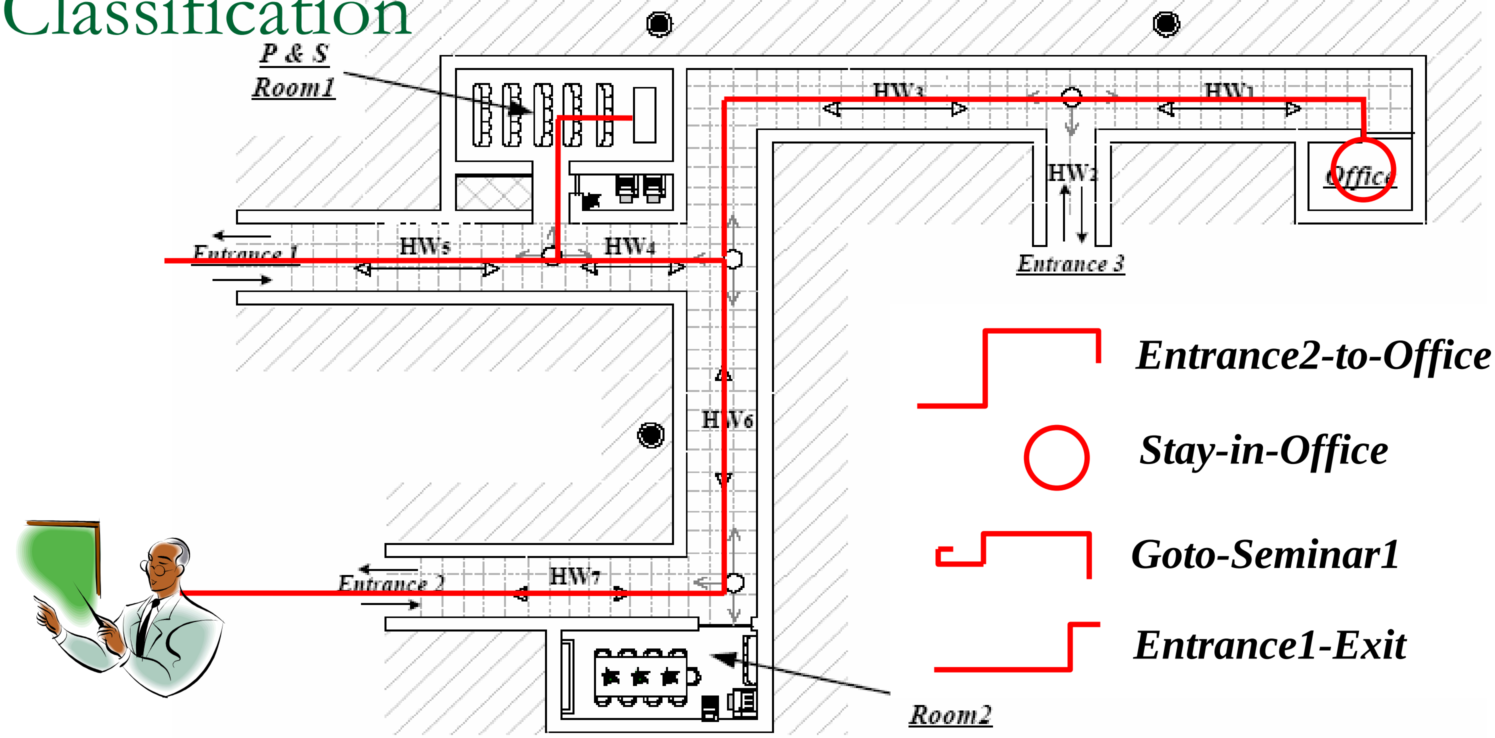


Fig. 17. Effect of using a varying number of traces to reduce the sampling time

IEEE PERCOM 2005 Article, Chai et al.

ML Prob: Multi-Dimensional Time-Series Classification



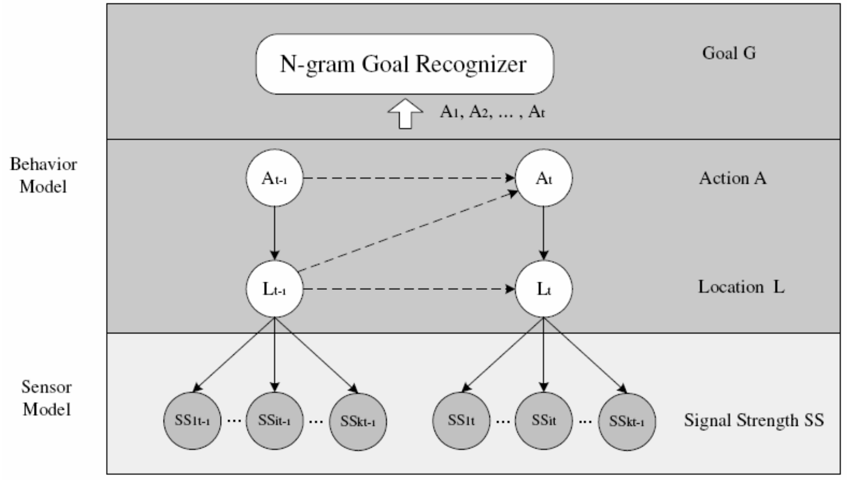
- Objective:
  - Infer what he is doing → actions are not directly observable — ***Sensor-based***
  - recognize his ultimate goal → more than one goal is achieved — ***Multiple-goal***

In: Proceedings of the AAAI ’2004, ‘2005

How to recognize a user's goals?

- Problem: how to ensure that goal recognition framework is robust?
- Previous Work:
  - HMM and DBN based: restricted to high-level inferences [Albrecht et al. 98] [Han & Veloso 00]
  - Sensor-based DBN: monolithic architecture but inflexible

    [Nguyen et al.03] [Bui 03] [Liao et al.04]
- Our Method:
  - A two-level recognition architecture

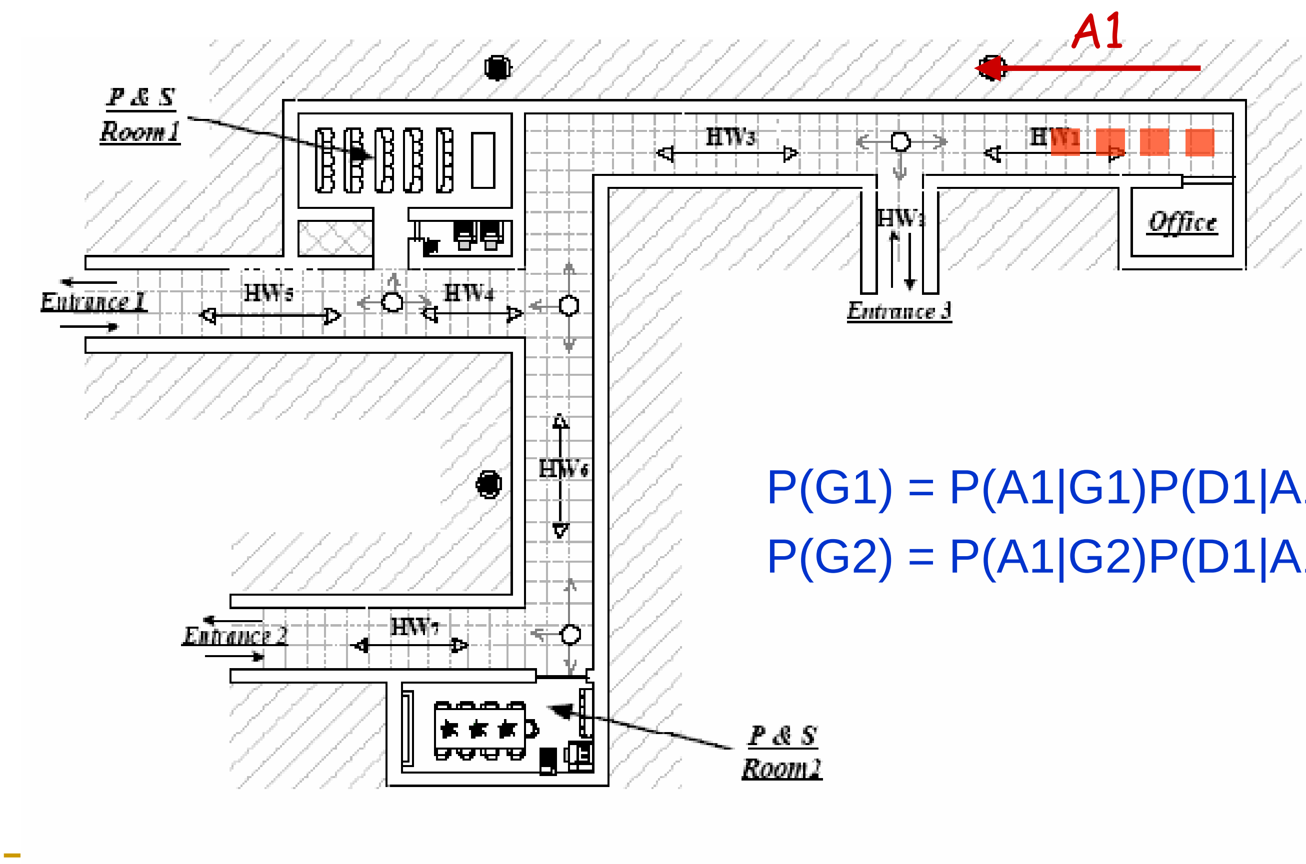


$$\begin{aligned} G^* &= \arg\max P(G|A_1, A_2, \ldots, A_t) \\ &= \arg\max P(G|A_{1:t}). \end{aligned}$$

$$\begin{aligned} G^* &= \arg\max \frac{P(A_{1:t}|G)P(G)}{P(A_{1:t})} \\ &= \arg\max P(A_{1:t}|G)P(G) \end{aligned}$$

An Example

G1: Go-to-Print-in-Room1   G2: Go-to-Seminar-in-Room2



Signal vector:

<58, 60, 45>

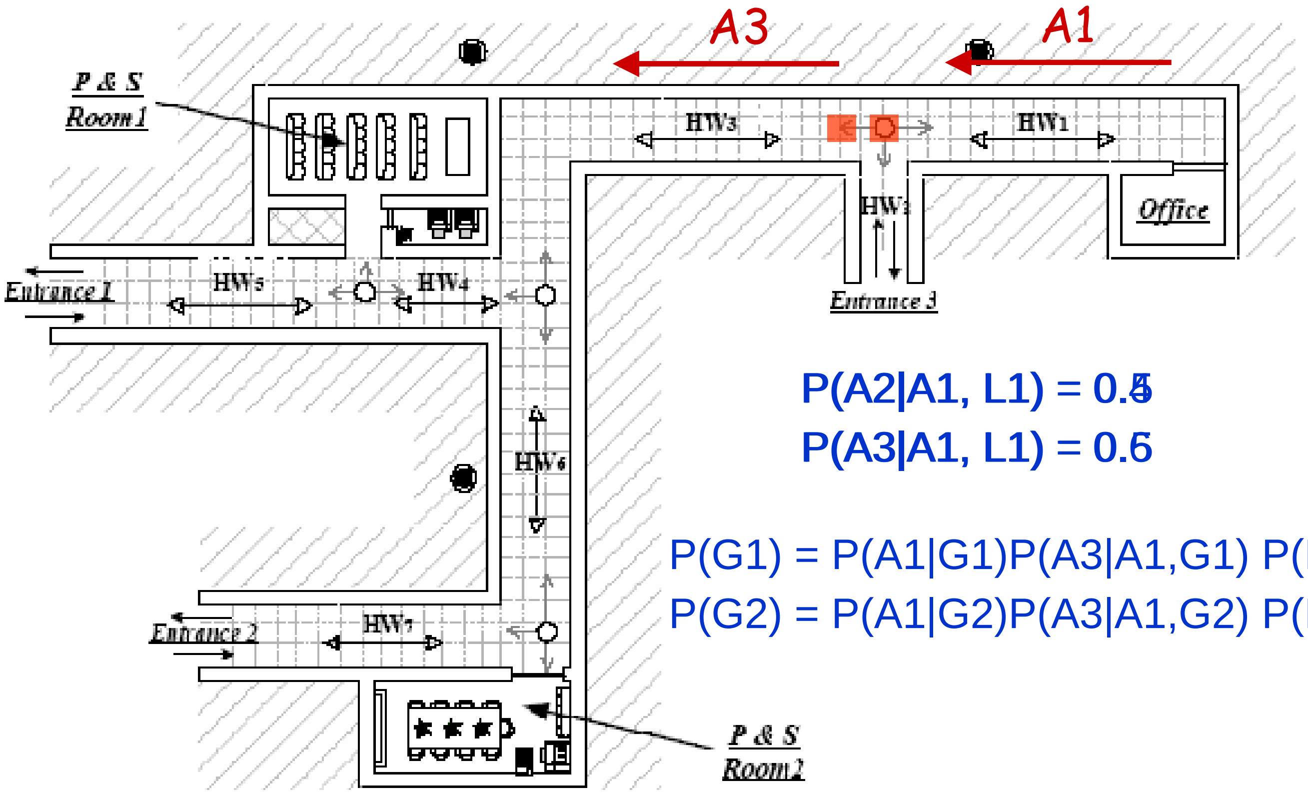
<56, 59, 48>

P(G1) = P(A1|G1)P(D1|A1) = 0.5

P(G2) = P(A1|G2)P(D1|A1) = 0.5

An Example

G1: Go-to-Print-in-Room1   G2: Go-to-Seminar-in-Room2



P(A2|A1, L1) = 0.4

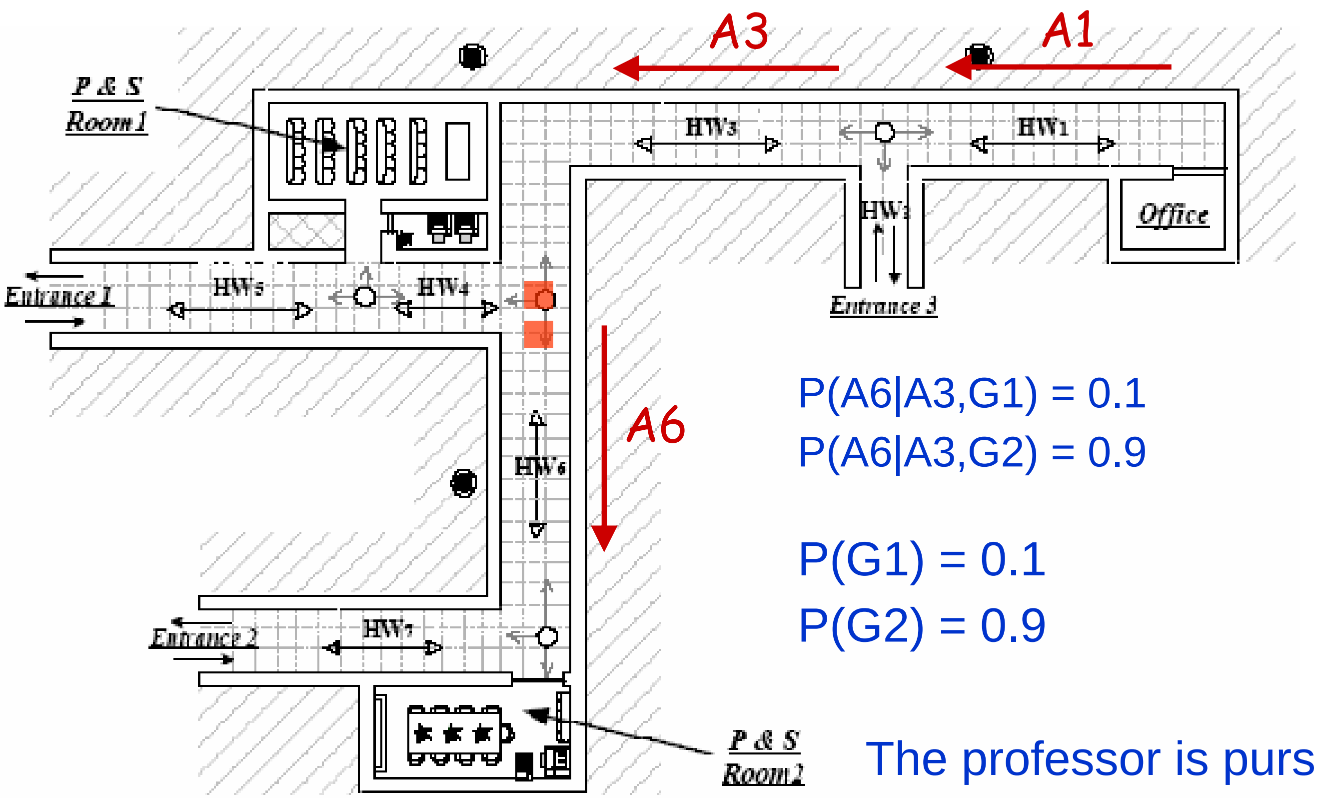
P(A3|A1, L1) = 0.6

P(G1) = P(A1|G1)P(A3|A1,G1) P(D1|A1)P(D3|A3) = 0.5

P(G2) = P(A1|G2)P(A3|A1,G2) P(D1|A1)P(D3|A3) = 0.5

An Example

G1: Go-to-Print-in-Room1   G2: Go-to-Seminar-in-Room2



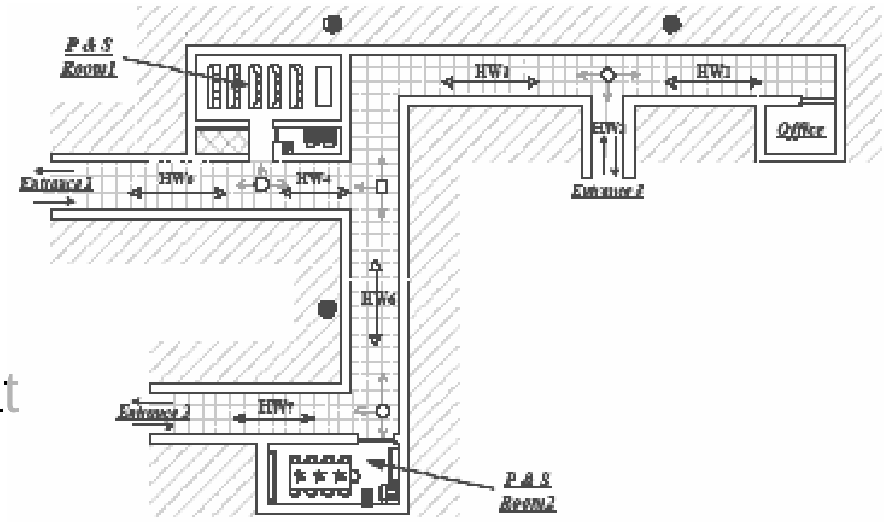
P(A6|A3,G1) = 0.1
P(A6|A3,G2) = 0.9

P(G1) = 0.1
P(G2) = 0.9

The professor is pursuing G2.

Environment and Data Set

- 99 locations (a 1.5-meter grid cell)
- 8 out of 25 base stations
- Data for sensor model:
  - 100 samples were collected at each location
- Evaluation Data: about 600 traces(19 goals)
  - 3-fold cross-validation

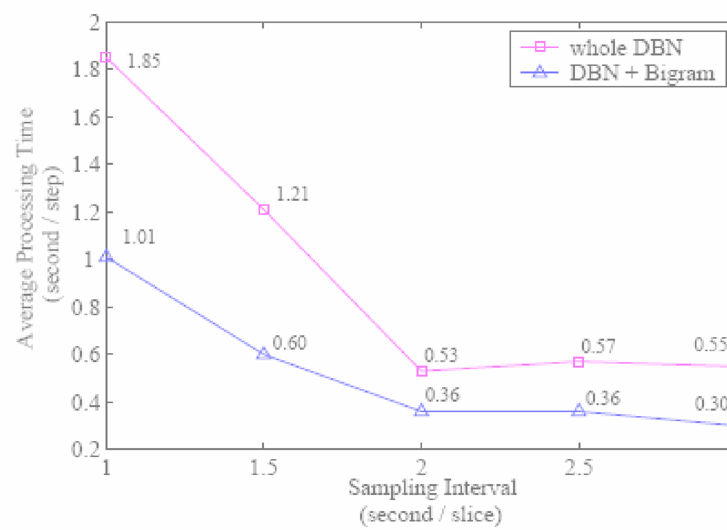


99 locations
10 actions
19 goals

Evaluation Criteria

Sampling Interval	1s	1.5s	2s	2.5 s	3s
Whole DBN	89.5 %	87.1 %	84. 2%	75. 4%	71. 9%
DBN + Bigram	90.5 %	83.2 %	82. 1%	74. 7%	72. 6%

- Efficiency
  - The average processing time for each observation in the on-line recognition.
- Accuracy
  - The number of correct recognition divided by the total number of recognition.
- Convergence rate
  - The average number of observations, after which the recognition converges to the correct answer, over the average number of observations for those traces which converge.

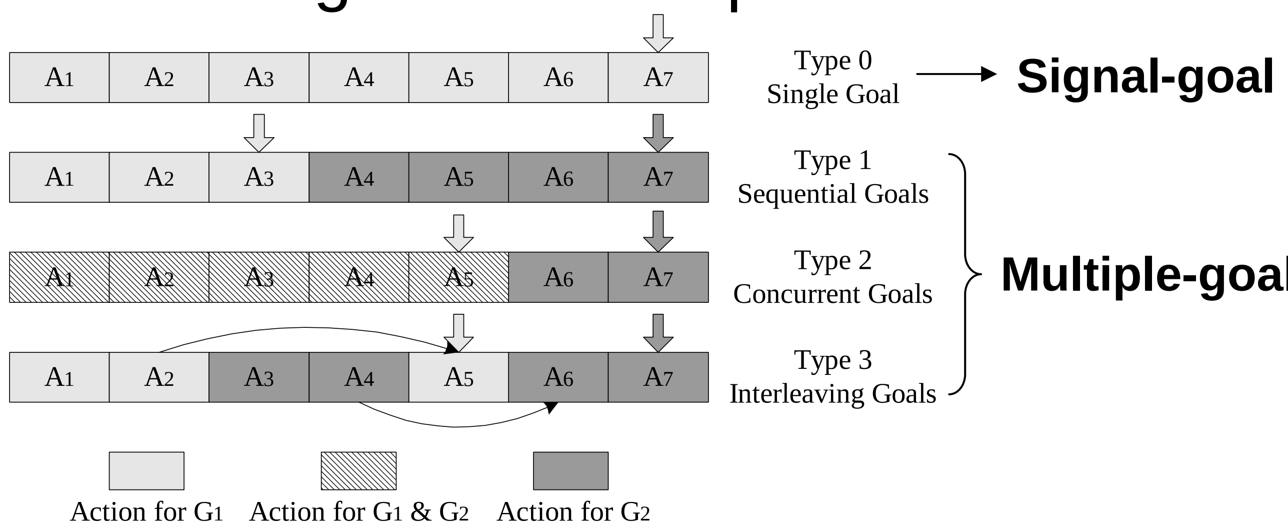


Sensor-Based Multiple-Goal Recognition==Time-series Classification

- Recognition based on sensory readings

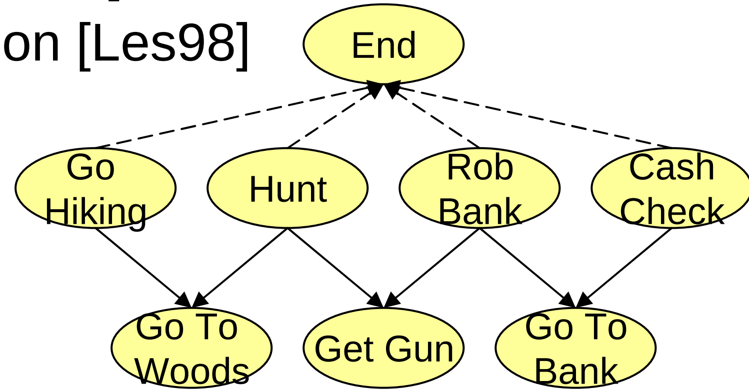
Trace #	Observation Sequences				Goal
	t_1	t_2	...	t_k	
1	$(AP_1$:-57) $(AP_2$:-33) $(AP_3$:-51)	$(AP_1$:-56) $(AP_2$:-30) $(AP_3$:-62)	$(AP_1$:-55) $(AP_2$:-36) $(AP_3$:-56)	$(AP_1$:-52) $(AP_2$:-62) $(AP_3$:-47)	G_1
2	$(AP_1$:-62) $(AP_2$:-57) $(AP_3$:-55)	$(AP_1$:-39) $(AP_2$:-41) $(AP_3$:-32)	$(AP_1$:-46) $(AP_2$:-45) $(AP_3$:-43)	$(AP_1$:-41) $(AP_2$:-43) $(AP_3$:-27)	G_2 G_3

- Multiple-goal in a single action sequence



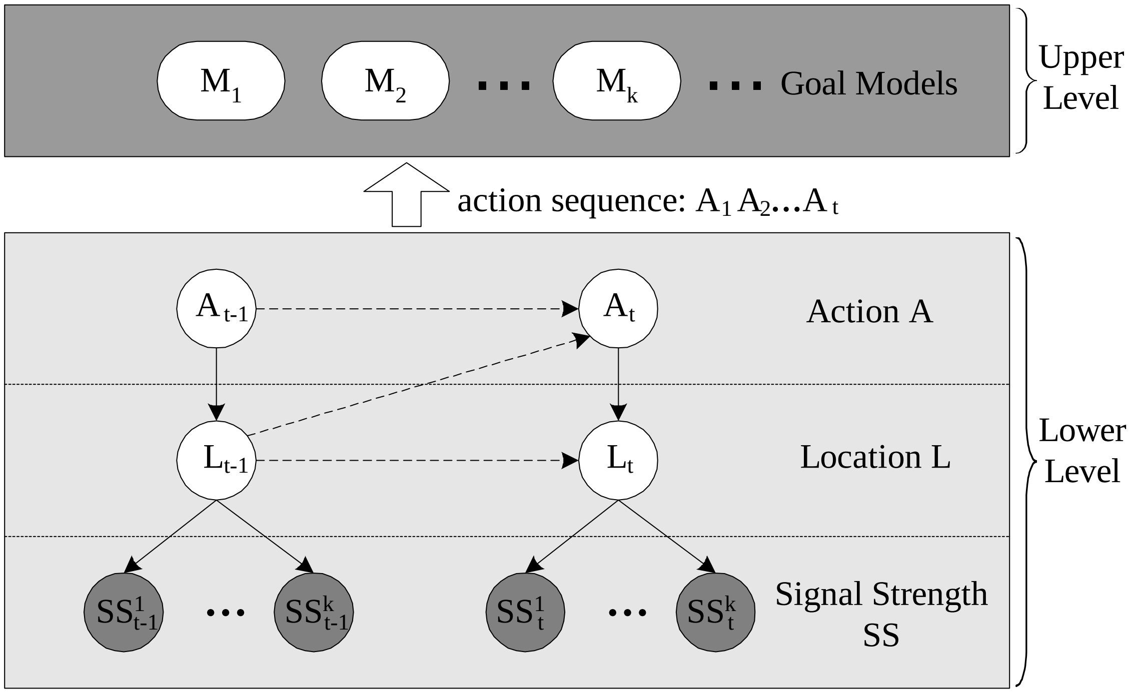
Plan Recognition and Activity Recognition

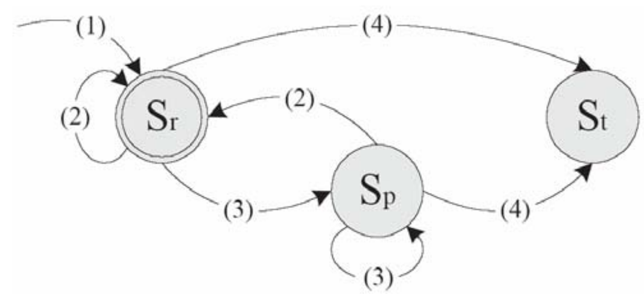
- Two categories of approaches:
  - Consistency approaches
    - Formal theory of plan recognition [Kau87]
    - Scalable and adaptive goal recognition [Les98]
  - Probabilistic approaches
    - Hidden Markov models
    - Bayesian Net and dynamic BN



- Limitations
  - Logic-based assume actions are given, and cannot deal with uncertain signals
  - Bayesian approaches must have a winning goal; but there may be several concurrent goals

Framework of Sensor-Based Multiple-Goal Recognition





(1) Instantiate; (2) Evolve; (3) Suspend; (4) Terminate

Two-level multiple-goal recognition framework

Model Instantiation and evolution

- A default-goal model M_0 is instantiated when a goal model is created at time t
  - M_0 is added into the model set $\mathbf{M}$
  - $L_t(M_0) = \pi_0 Q_0(A_t)$
- A goal model M_k is instantiated
  - whenever $\pi_k Q_k(A_t) \geq \pi_0 Q_0(A_t)$
  - $Acc(M_k) = A_t$ and M_k is added into $\mathbf{M}$
  - $L_t(M_k) = \pi_k Q_k(A_t)$

Experimental Setting

- Actions and goals:

AID	Name	AID	Name
A_1	*Walk-in-HW1*	A_2	*Walk-in-HW2*
A_3	*Walk-in-HW3*	A_4	*Walk-in-HW4*
A_5	*Walk-in-HW5*	A_6	*Walk-in-HW6*
A_7	*Walk-in-HW7*	A_8	*Print_R1*
A_9	*Seminar_R1*	A_{10}	*Print_R2*
A_{11}	*Seminar_R2*	—	—

GID	Name	GID	Name
G_1	*"Print-in-Room2"*	G_2	*"Seminar-in-Room2"*
G_3	*"Print-in-Room1"*	G_4	*"Seminar-in-Room1"*
G_5	*"Go-to-Office"*	G_6	*"Exit-through-Entrance3"*
G_7	*"Exit-through-Entrance1"*	G_8	*"Exit-through-Entrance2"*

- ***eight*** goals, 850 single-goal traces
- Multiple-goal traces are synthesized:
  - Segments of single-goal traces are pieced together to generate connective traces containing multiple goals.

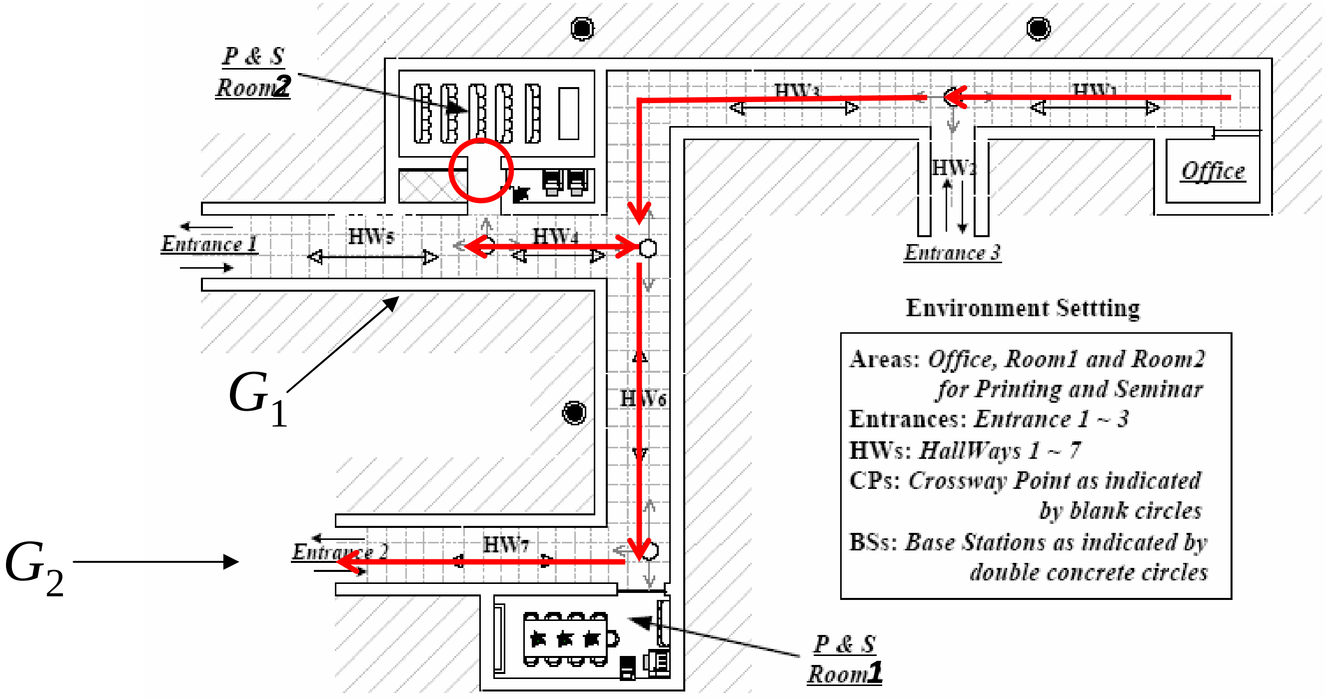
Comparison Targets & Evaluation Criteria

- Three algorithms:
  - *MG-Recognizer* [Cha05]
  - *SG-Recognizer* [Yin04]
  - *BHMM-Recognizer* [Han99]
- Three criteria:
  - Recognition accuracy
  - Inference efficiency
    - Measured in terms of the number of models instantiated
  - Scalability
    - *w.r.t.* the number of goals modeled
    - *w.r.t.* the number of goals contained in a single trace

An Example

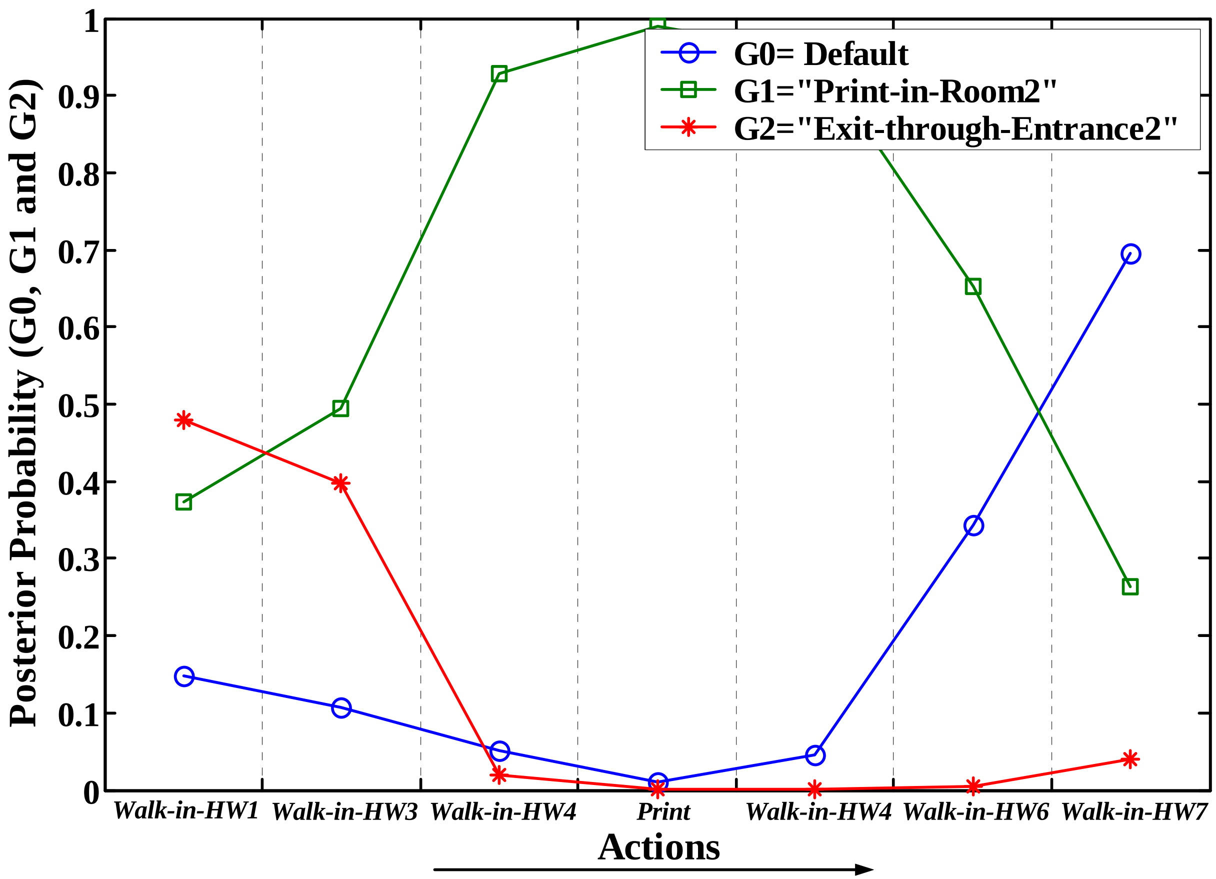
- Two goals are achieved in a single trace:

  G_1 = "*Print-in-Room2*" and G_2 = "*Exit-through-Entrance2*"



Recognition Accuracy

- *SG-Recognizer*



Accuracy and Efficiency

- Recognition accuracy:

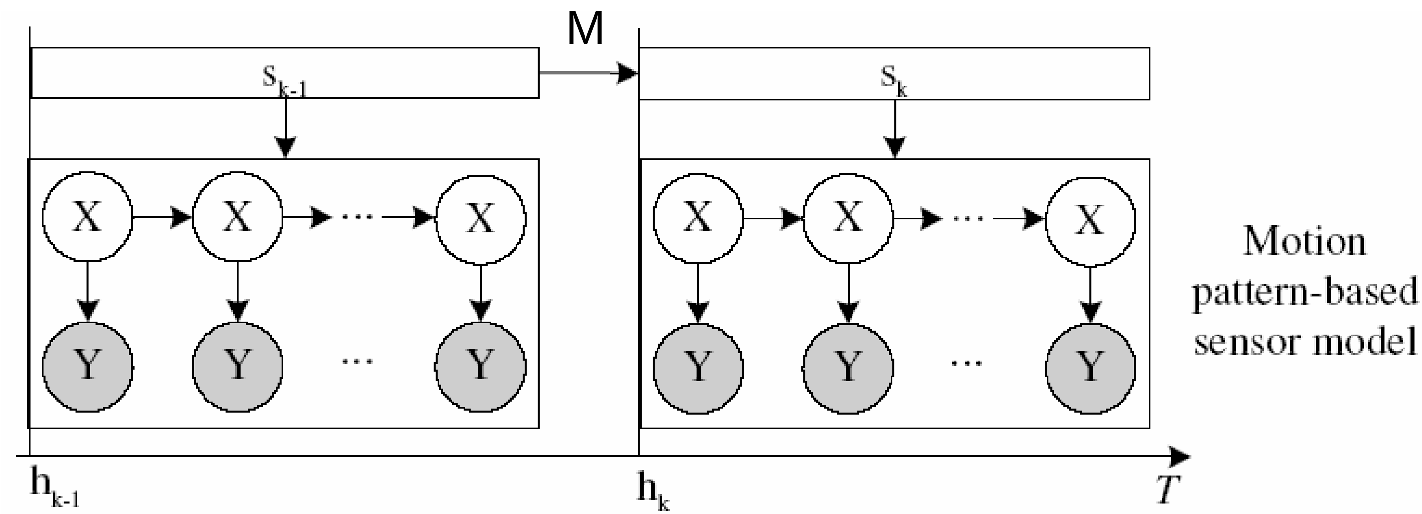
Recognizer	*SG-Recognizer*	*BHMM-Recognizer*	*MG-Recognizer*
Single-Goal	97.8%	95.5%	94.6%
Multiple-Goal	24.5%	79.1%	91.4%

- Inference Efficiency:

Recognizer	*SG-Recognizer*	*BHMM-Recognizer*	*MG-Recognizer*
Single-Goal	9	20.7	6.5 + 3.7
Multiple-Goal	9	28.7	6.6 + 4.8

ML Prob: Segmentation and Feature Selection in Multi-dimensional Time-series

- Learning a Probabilistic Segmentation Model [AAAI 2005]
- We partition an observation sequence Y into Ns segments



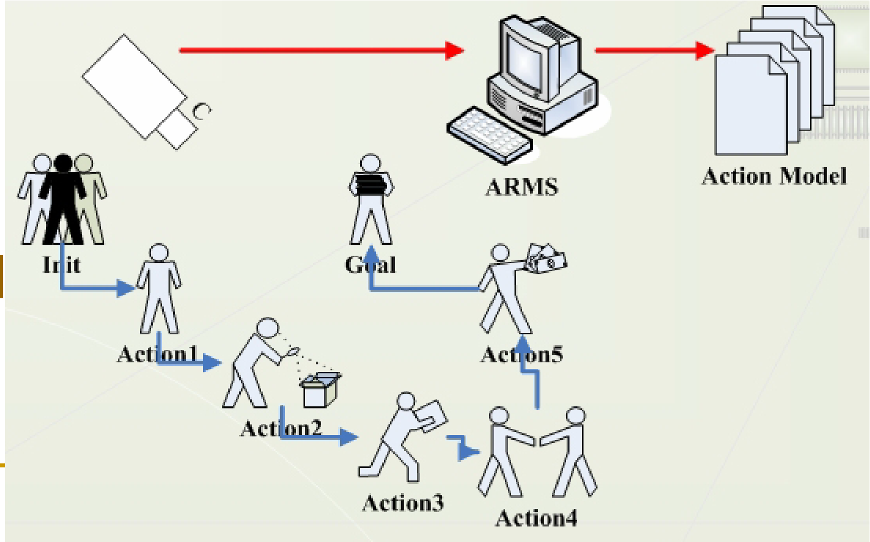
  - Segment labels $S = \{s_1, s_2, \ldots, s_{N_s}\}$ and segmentation points $H = \{h_1, h_2, \ldots, h_{N_s}\}$

ML Prob: Statistical Relational Learning: Action Model Learning

- Input: observed plans
  - $init_1, a_{11}, a_{12}, a_{13}, \ldots, a_{1n}, goal_1$
  - $init_2, a_{21}, a_{22}, a_{23}, \ldots, a_{2m}, goal_2$
  - …
- Output: action models; e.g.
  load (x - hoist y - crate z - truck p - place)
  - pre: (at x p), (at z p), (lifting x y)
  - del: (lifting x y)
  - add: (at y p), (in y z),
    - (available x), (clear y)
- Main Issue:
  - Automatically guess an initial action model
    Then allow humans to edit these models

Key contribution:

- can learn action models even when no intermediate state observations are available



Distributed Learning In Sensor Networks

- *Distributed Regression:an Efficient Framework for Modeling Sensor Network Data,* Carlos Guestrin et al, IPSN 2004
- Key Insight:
  - Robustness of learned model is key in sensor networks
    - Nodes may be added to the network or fail
    - Communication is unreliable, and link qualities change over time
  - Distributed machine learning:
    - Pieces of models are learned at different sites
    - A central model is integrated together.

Conclusions and Future Work

- Sensor and Wireless Networks provides grounds for new Machine Learning Research
  - Semi-supervised Classification on time series
  - Data and Model Migration
  - Multi-dimensional Segmentation
  - Distributed Learning of robust models

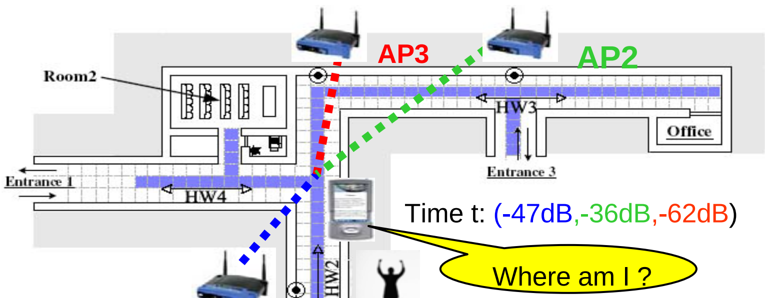
Our Other Work in 2005 (HKUST)

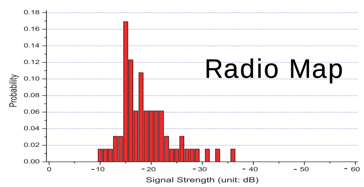
- **See http://www.cs.ust.hk/~qyang**
- **Research in ubiquitous computing**
  - [IJCAI05p] J.F. Pan, J.T. Kwok, Q. Yang and Y.Q. Chen. "Accurate and Low-cost Location Estimation Using Kernels" **IJCAI 2005**, Edinburgh, UK, Aug 2005
  - [AAAI05y] J. Yin, D. Shen, Q. Yang and Z.N. Li "Activity Recognition through Goal-Based Segmentation". **AAAI 2005**, Pittsburg, PA USA, July 2005
  - [AAAI05c] X.Y. Chai and Q. Yang, "Multiple-Goal Recognition From Low-level Signals" **AAAI 2005**, Pittsburg, PA USA, July 2005
- **Research in Web mining**
  - KDD-CUP 2005: Champion on all three awards (see http://webproject1.cs.ust.hk/q2c/)
  - J.T Sun, D. Shen, H.J. Zeng, Q. Yang, Y.C Lu and Z. Chen "Web Page Summarization Using Clickthrough Data". **ACM SIGIR** 2005. Brazil. August 2005.
  - Z. Chen, G.R. Xue, Y. Yu,, Q. Yang, "Exploiting the Hierarchical Structure for Web Link Analysis" **ACM SIGIR** 2005. Brazil. August 2005.
  - W.S. Xi, Z. Chen, G.R. Xue, Y. Yu, H.J. Zeng, Q. Yang, and C.X. Lin "Scalable Collaborative Filtering Using Cluster-based Smoothing" **ACM SIGIR** 2005. Brazil. August 2005.
- **Research in case-based reasoning & Machine Learning**
  - [AAAI05p] R. Pan, Q. Yang, J.F. Pan and L. Li. Competence Driven Case-Base Mining. *Proceedings of the **AAAI 2005***, Pittsburg, PA USA, July 2005

Accurate and Low-cost Indoor Location Estimation Using Kernels

J.F. Pan, J.T. Kwok, Q. Yang and Y.Q. Chen, ***IJCAI 2005***

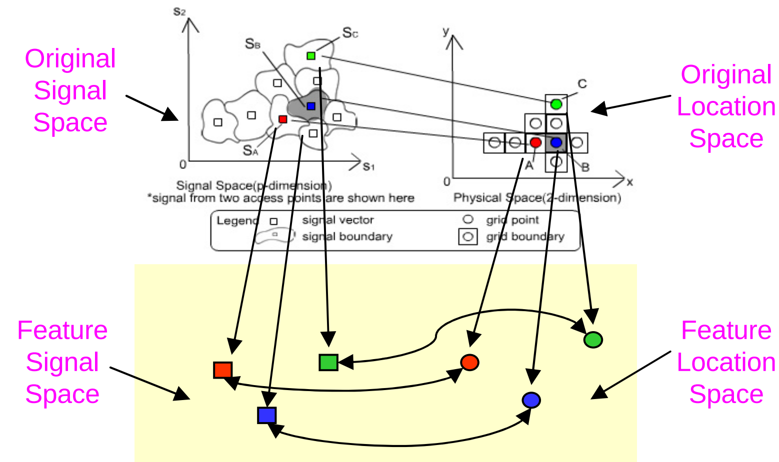
Problem A user with a mobile device walks in an **indoor** wireless environment (Covered by WiFi signal)





Motivation

- Similar signals may not necessarily be nearby locations, or vice versa
- Maximize correlation between signal and location under feature transformation



Methodology **Kernel Canonical Correlation Analysis**

- Proposed by [D.R Hardoon et al. 2004]
- Two non-linear Canonical Vectors **Wx** & **Wy**

$$\mathbf{Wx = X}\ \alpha \quad \mathbf{Wy = Y}\ \beta$$

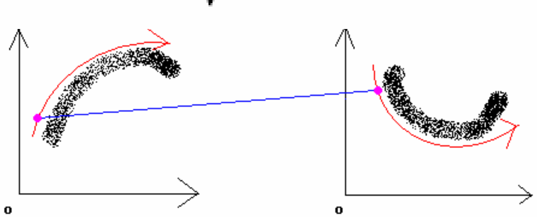
- K is the kernel

$$\Phi: x \rightarrow \Phi(x)$$

$$K(x, z) = \langle \Phi(x), \Phi(z) \rangle$$

- Maximize the correlation of projections

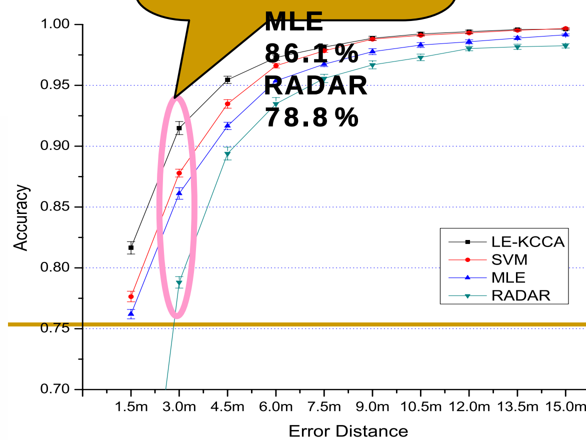
$$\rho = \max_{\alpha,\beta} \frac{\alpha' K_x K_y \beta}{\sqrt{\alpha' K_x^2 \alpha \cdot \beta' K_y^2 \beta}}.$$



Experiment Results

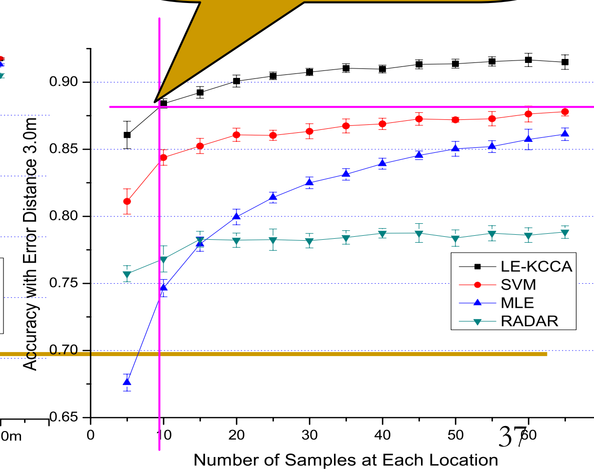
Accuracy (Error in 3.0m)

LE-KCCA 91.6%
SVM 87.8%
MLE 86.1%
RADAR 78.8%



Reduce Calibration Effort

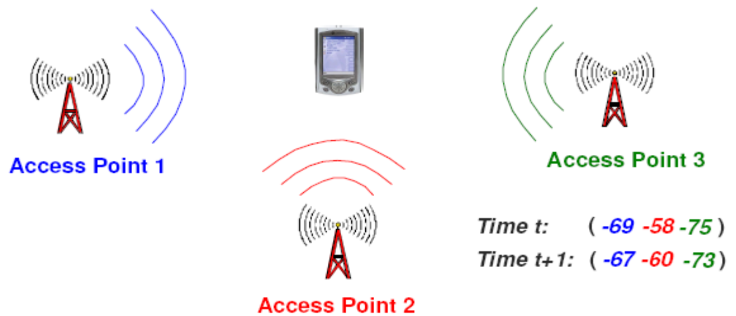
Outperform the others using 10-15 samples from each location



Activity Recognition through Goal-Based Segmentation

Jie Yin, Dou Shen, Qiang Yang and Ze-Nian Li, AAAI 2005

- Application Domain: Wireless Environment



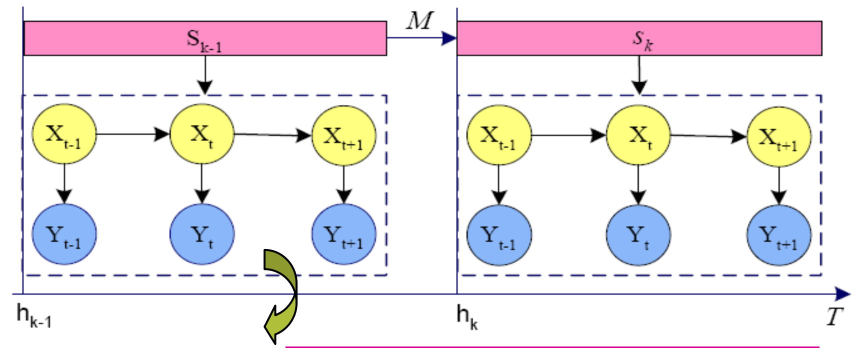
- Trace Database on Signal-Strength Readings

Trace ID	Signal-strength Sequences				Goal ID
	t_1	t_2	...	t_n	
1	(AP1:-80) (AP2:-78) (AP3:-62)	(AP1:-81) (AP2:-77) (AP3:-64)	(AP1:-64) (AP2:-69) (AP3:-71)	(AP1:-68) (AP2:-84) (AP3:-81)	G_1
2	(AP1:-66) (AP2:-85) (AP3:-84)	(AP1:-63) (AP2:-81) (AP3:-78)	(AP1:-69) (AP2:-64) (AP3:-59)	(AP1:-78) (AP2:-78) (AP3:-61)	G_2

- Goal-Based Segmentation algorithm
  - High-level goals can be recognized from low-level signal segments
  - Each segment define a motion pattern

Sequence Segmentation Model
Activity Recognition Model

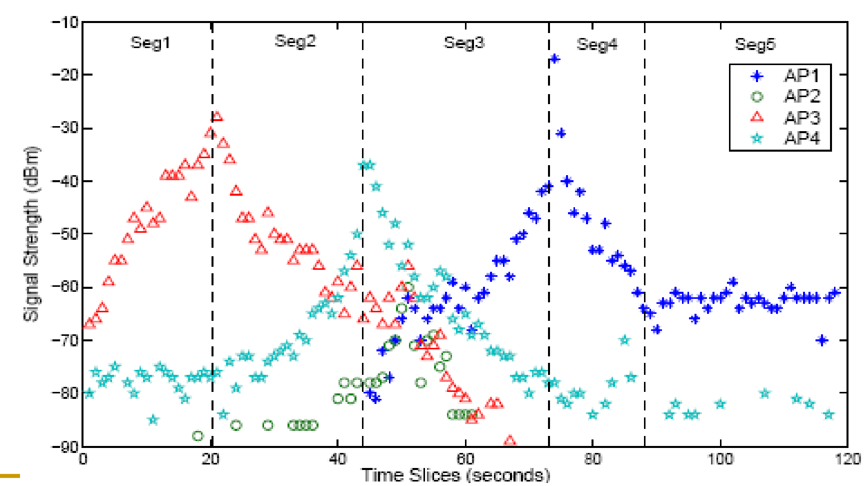
- Probabilistic Segmentation Model



Motion Pattern:

Dynamics Model: $X_{t+1} = \Lambda_t X_t + W_t$
Observation Model: $Y_t = C_t X_t + B_t$

- Illustration on Sensory data

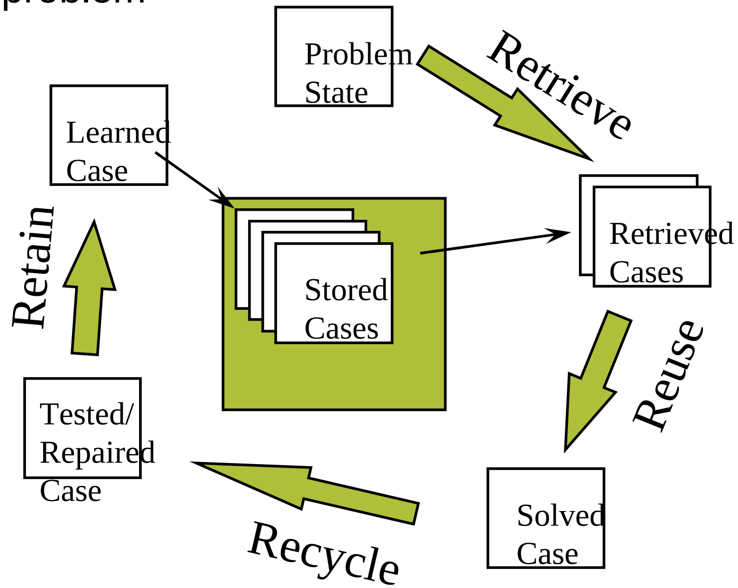


Reduce human effort in calibration for activity recognition!

Competence Driven Case-Base Mining

R. Pan, Q. Yang, J.F. Pan, L. Li , AAAI 2005

- Case-Based Reasoning: Using previous cases to solve a current problem



•Step 1: learn the distribution of the original sample – using KFDA

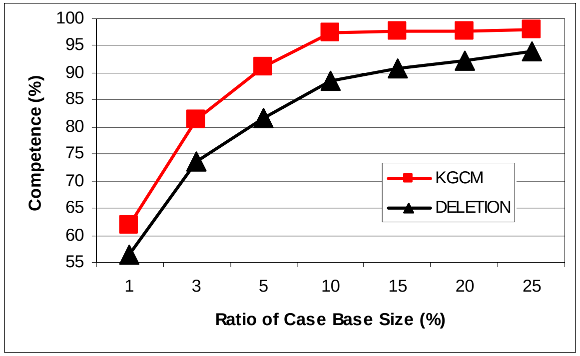
•Step 2: Mining cases by considering the distribution and diversity

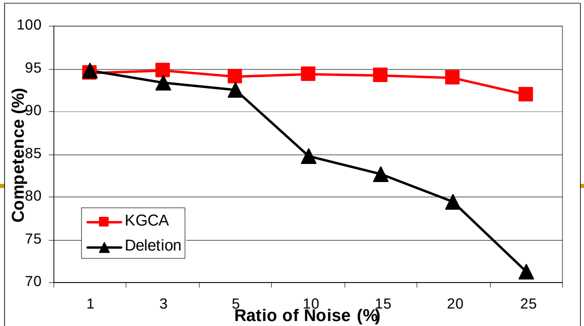
Problem: How to automatically obtain a quality case base from the raw data?

What is the quality metric?

	Customer Database			
	Income	Married	Cars	Approved?
Sammy	50K	n	1	?
Beatrice	50K	y	1	Yes
Dylan	80K	n	2	Yes
Mathew	30K	n	1	No
Larry	40K	n	0	No
Basil	80K	n	1	Yes

Applicants



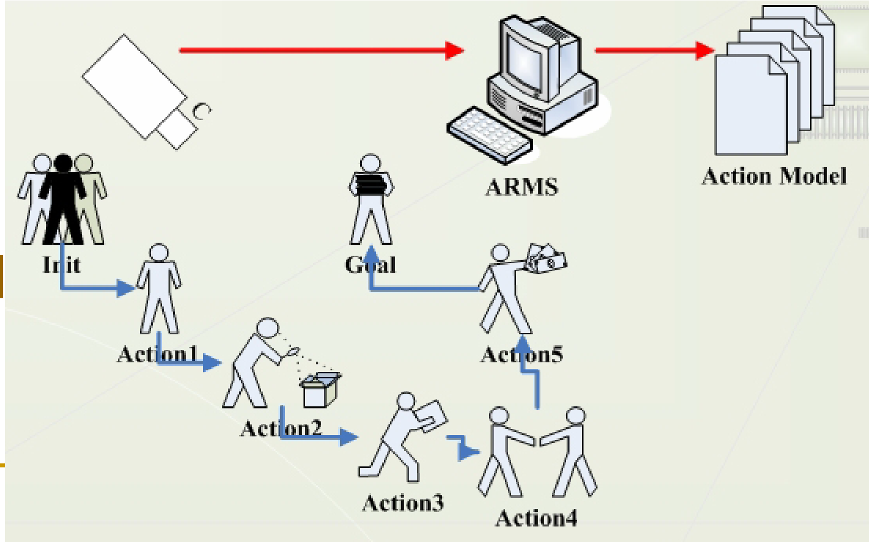


ARMS: Learning Action Models

ICAPS 2005

Winner: the first ICAPS/KE competition 2005, CA, USA

- Input: observed plans
  - $init_1, a_{11}, a_{12}, a_{13}, \ldots, a_{1n}, goal_1$
  - $init_2, a_{21}, a_{22}, a_{23}, \ldots, a_{2m}, goal_2$
  - …
- Output: action models; e.g.
  load (x - hoist y - crate z - truck p - place)
  - pre: (at x p), (at z p), (lifting x y)
  - del: (lifting x y)
  - add: (at y p), (in y z),
    - (available x), (clear y)
- Main Issue:
  - Automatically guess an initial action model
    Then allow humans to edit these models

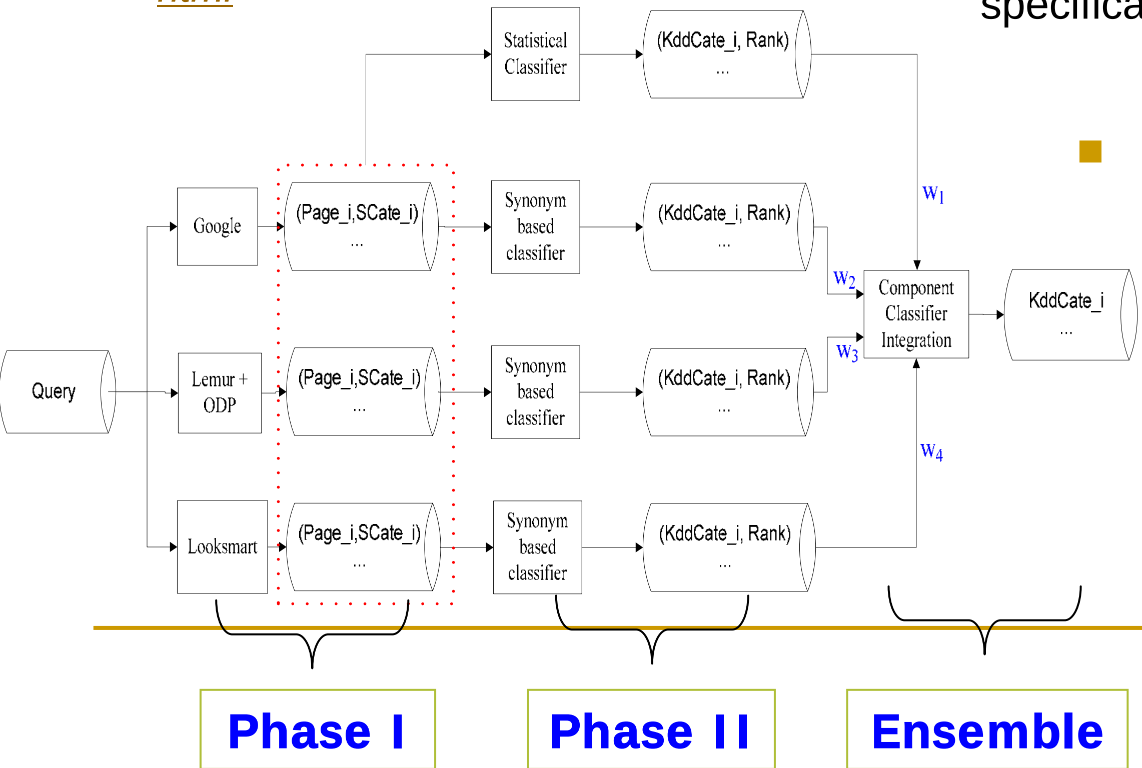


ACM KDD-CUP 2005 -- Winner

HKUST Team: *D. Shen, R. Pan, J.T. Sun, J.F. Pan, K.H. Wu, J. Yin and Professor Q. Yang*

- **Task**
  - Categorize 800,000 queries into 67 predefined categories;
- **Limitation**
  - "There is no restriction on what data you can/can't use to build your models." *From http://kdd05.lac.uic.edu/kddcup.html*

- **Key Characteristics**:
  - No training data
  - Meaning of Queries: ambiguous
    - A query usually contains too few words;
    - Queries often have more than one meaning.
  - Semantics of Categories: uncertain
    - Only the names of Categories, no more specification;



- HKUST won *all three awards* for KDDCup 2005:
  - **Query Categorization Precision Award,**
  - **Query Categorization Performance Award**
  - **Query Categorization Creativity Award**

KDDCUP Winners Aug 2005

Thank You!

- http://www.cs.ust.hk/~qyang