JANUARY 18-19, BERKELEY, CA
PACIFIC EARTHQUAKE ENGINEERING RESEARCH CENTER
2018 Annual Meeting
UC BERKELEY • CALTECH • OSU • STANFORD • UC DAVIS • UC IRVINE • UC LOS ANGELES • UC SAN DIEGO • UNR • USC • U WASHINGTON

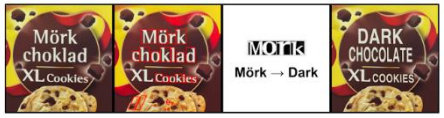
TBSI 清华-伯克利深圳学院 Tsinghua-Berkeley Shenzhen Institute

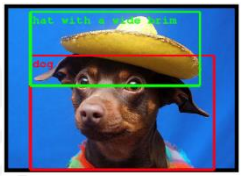
Learning with Confidence: Theory and Practice of Information Geometric Learning from High-dim Sensory Data

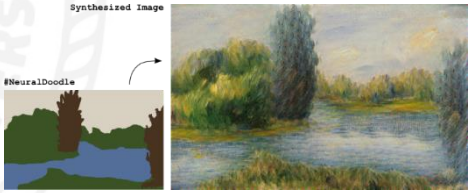
Professor Lin Zhang
Department of Electronic Engineering, Tsinghua University
Co-director, Tsinghua-Berkeley Shenzhen Institute

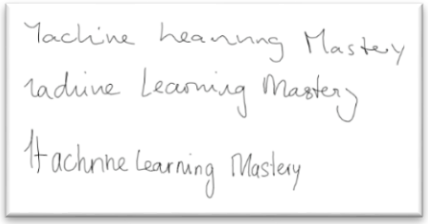
Joint work with Profs. Shao-Lun Huang (TBSI),
Lizhong Zheng (MIT), Pei Zhang (CMU)

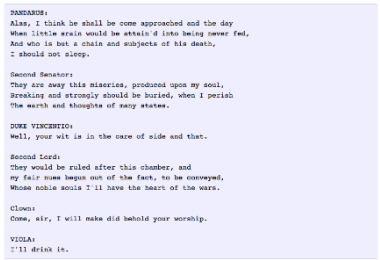
Who is NOT working on Machine Learning?





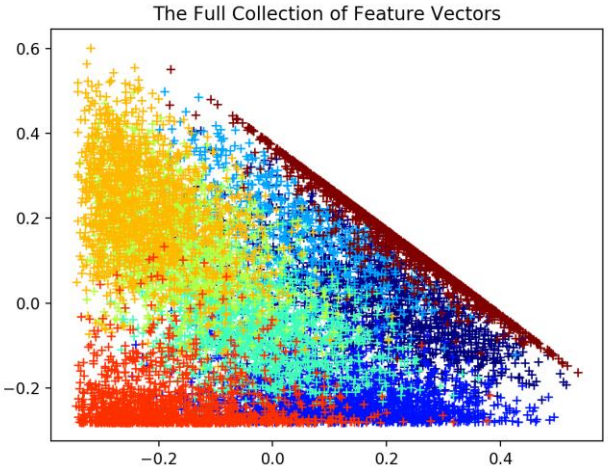


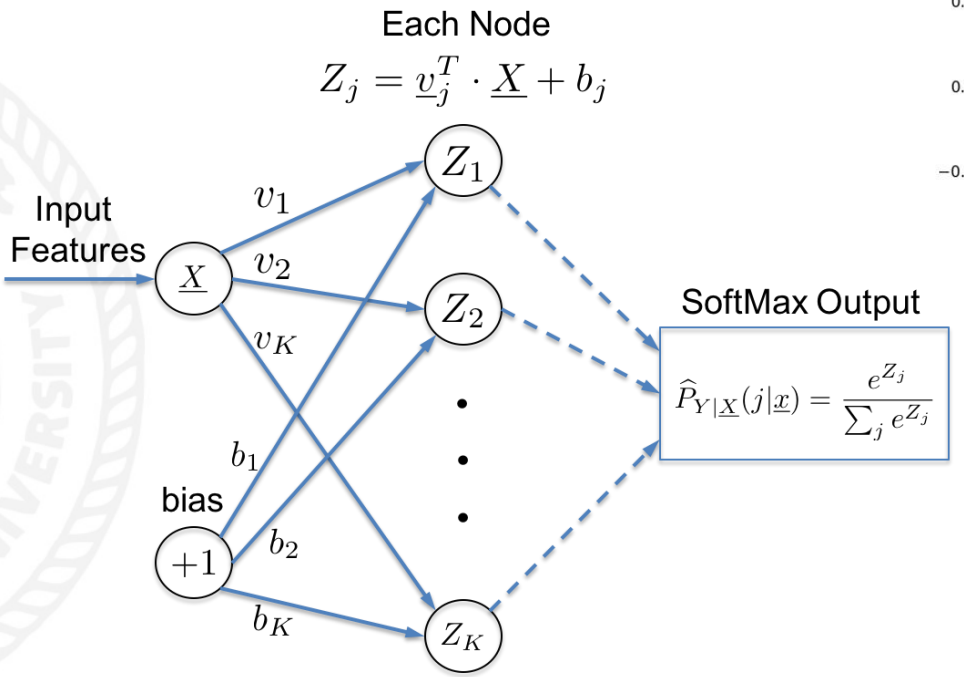




A Simple Example of Using Neural Networks

A Classification Problem





Standard Neural Network Packages

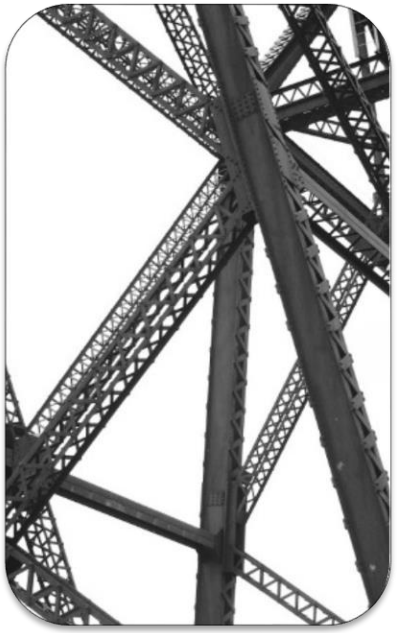
```
model=Sequential()
model.add(Dense(Cy, activation='softmax', input_dim=Dx))

sgd = SGD(lr=0.01, decay=1e-6, momentum=0.9, nesterov=True)
model.compile(loss='categorical_crossentropy',  optimizer=sgd,  metrics=['accuracy'])
model.fit(X,Labels, verbose=0, epochs=200, batch_size=200)
```

- Fitting: decide weights by stochastic gradient descent.
- It works, but how?

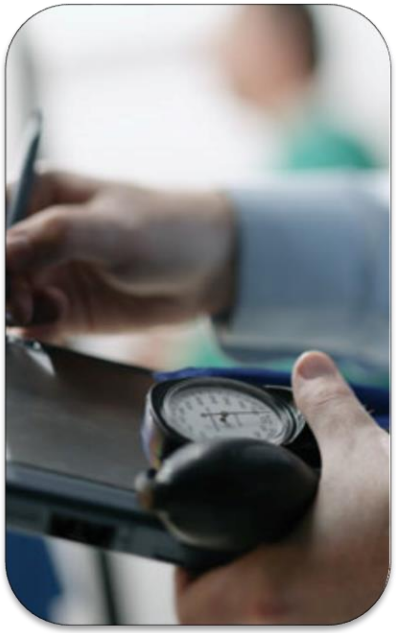

“What I cannot create, I do not understand. ”
—— Richard Feynman

We need confidence while learning

Fraud Detection



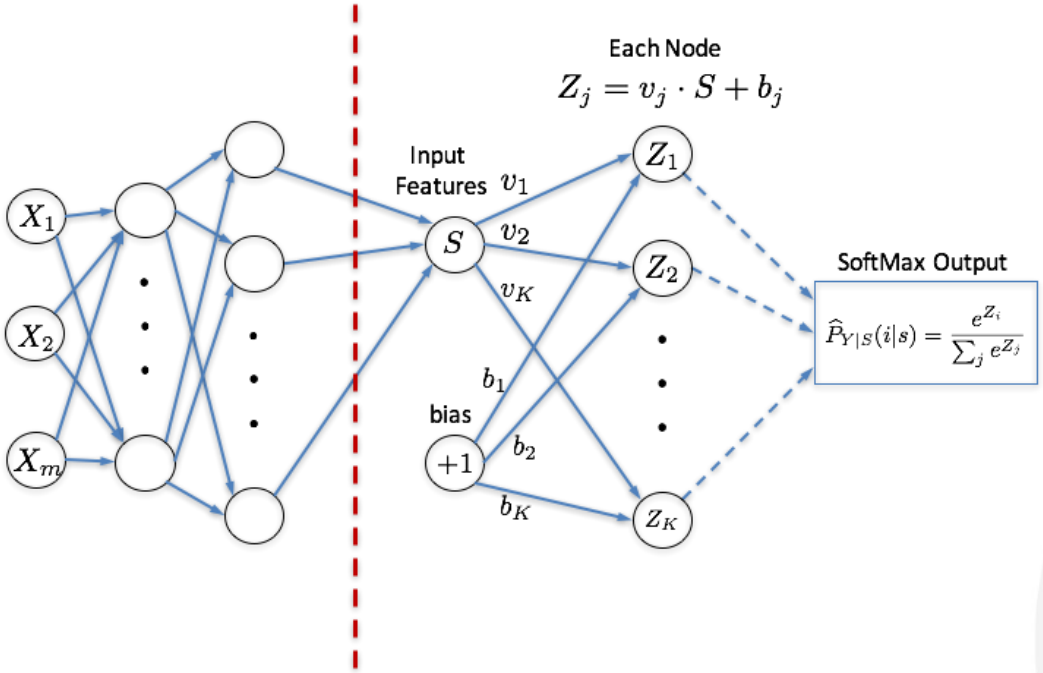
Structural Engineering



Medical Applications

Business Data Analytics

What About Deep Networks



What is it doing before the last layer?

Our Wish List for a Good Learning Theory

- More General:
  - Many different types of data.
  - Different time scales and qualities.
  - Not clear what we are looking for.

- More Flexible: General Purpose Processing and Information Market
  - Labels, Experts, and Fake News
  - Sensitive Information

- More Guarantees:
  - How Good Is Your Data?
  - Does It Solve My Problem?

Good Theory: Information Theory

- Information Theory: Claud Shannon
  - How much information do you obtain from an observation?
  - Measured in units of "bits"
  - Universal interface for compression and transmission
- To make it simple: the more surprised, the more information
- Kullback-Leibler divergence $D(P||Q)$: the distance between two distributions
- Metrics with Operational Meaning
  - Limit of compression
  - Channel capacity

"Frequently the messages have meaning; ... These semantic aspects of communication are irrelevant to the engineering problem."

— *C. E. Shannon*

Generalization of Information Theory: from Bit to Vector

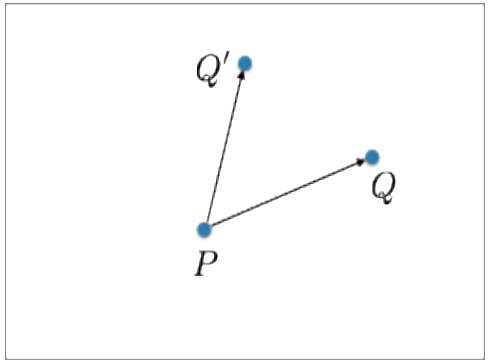
- Hope for new metrics:
  - Captures semantics: what is a partial piece of information, and what is it about?
  - Computable: not just from models but directly from data;
  - Backward Compatible;
  - Operational Meaning: related to inference performance;
  - Don't take away too much!

Distribution Space	$Q \leftrightarrow [Q(x) - P_0(x), x \in \mathcal{X}]$
Information Vector Space	$\phi \leftrightarrow \left[\frac{Q(x) - P_0(x)}{\sqrt{P_0(x)}}, x \in \mathcal{X}\right]$
Functional Space	$\text{LLR} \leftrightarrow \left[\log \frac{Q(x)}{P_0(x)} \approx \frac{Q(x) - P_0(x)}{P_0(x)}, x \in \mathcal{X}\right]$

- Backward compatibility

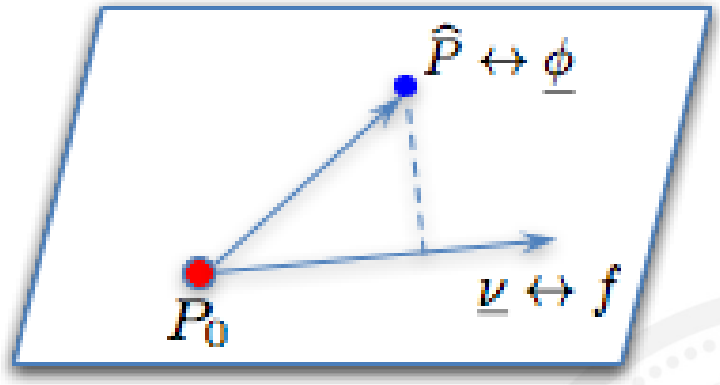
$$D(P||Q) \approx \|\underline{\phi}^P - \underline{\phi}^Q\|^2$$

- Most importantly, information now have directions



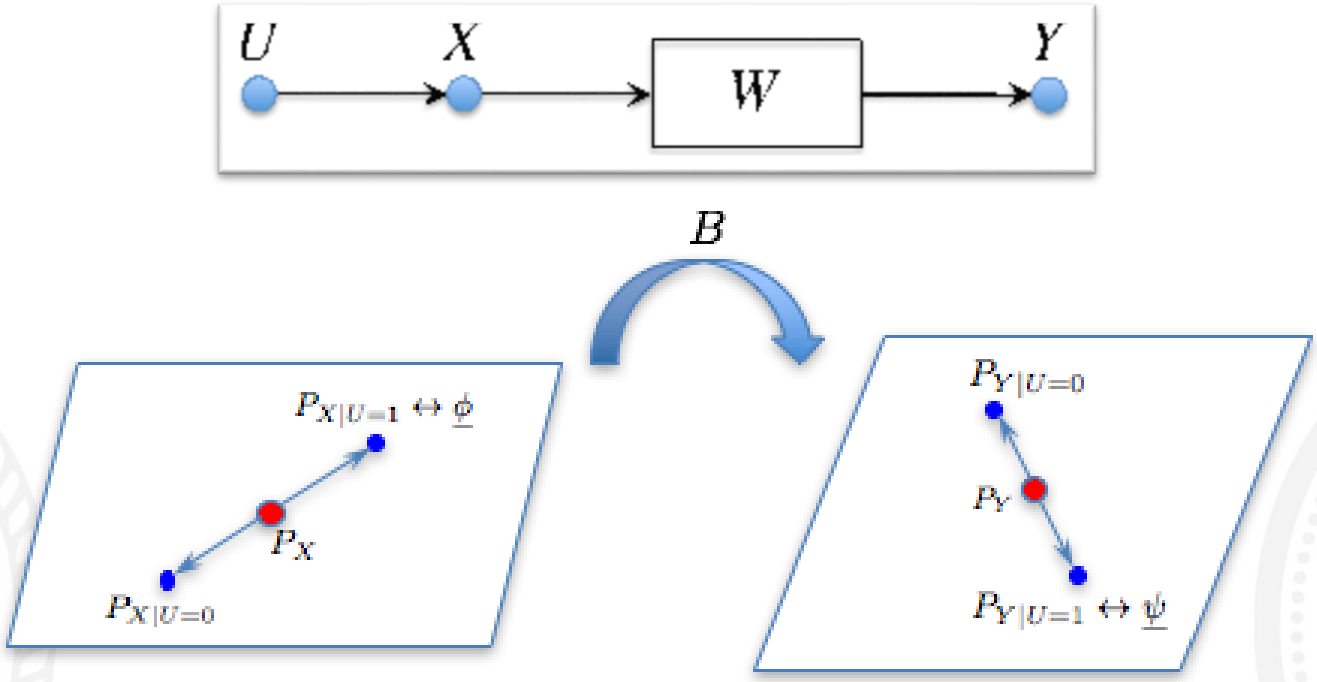
Distribution Space

Let's Draw a Picture



- Total information = length of $\underline{\phi}$ displacement vector
- Each score $f(\cdot)$ we evaluate corresponds to a particular direction in functional space;
- Partial Information, Score = $\mathbb{E}_{\widehat{P}}[f(X)] = \langle \phi, \nu \rangle$ projection

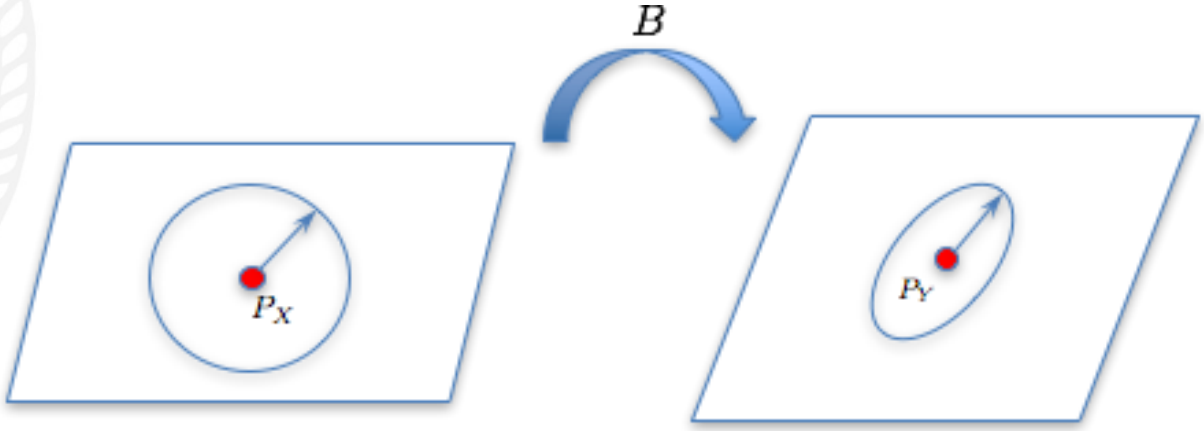
An easy problem: Detection Problem



- Some feature U of X that we want to detect from
  - noisy observations Y
  - Example like from behavior decide user profile

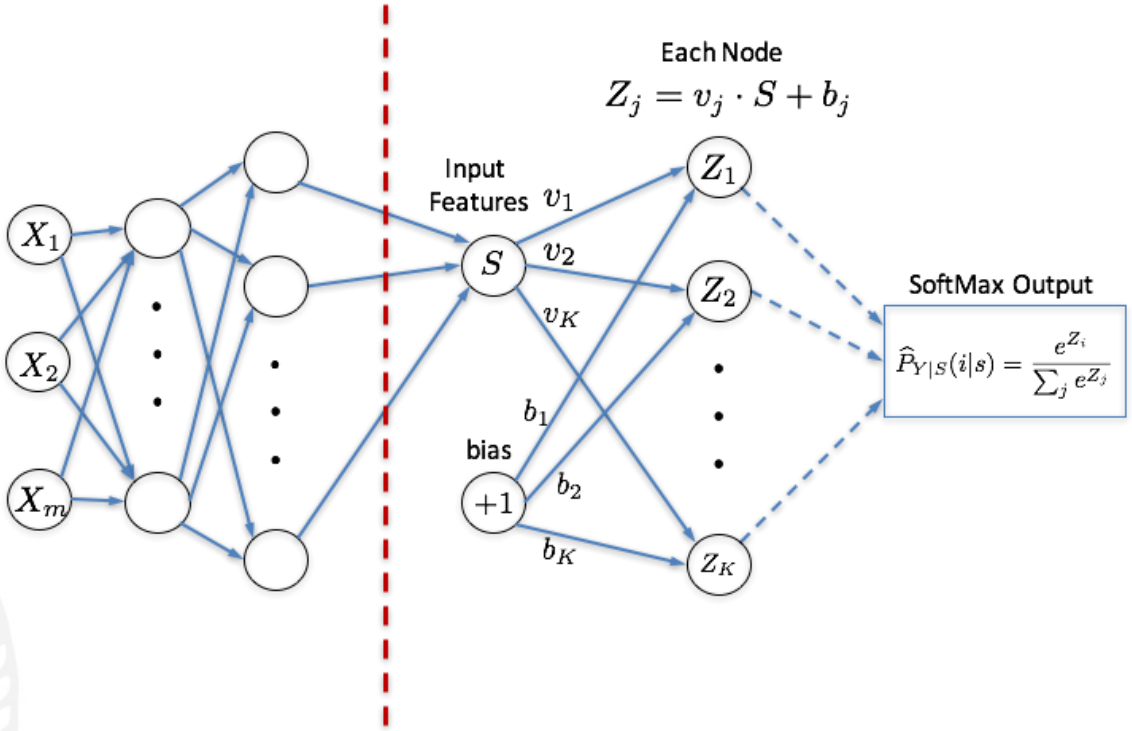
A Harder Problem:
What if We Don't Know What To Detect?

- Need processed data to be used for multiple purposes;
- Need to reduce dimensionality before learning the models;
- Cannot name what attribute we wish to make inference (e.g. recommendation, community detection)
- Good news: the picture still holds, and we can find the set of features with best performance on average



$$\max_g \quad \mathbb{E}_{U-X}[D(P_{g(Y)|U=1}||P_{g(Y)|U=0})]$$

Put Theory to Work: Neural Networks



- The goal of Neural Networks is the same: to pick the “useful” features of high dimensional observations.
- Supervision: want to specify the dependence between the inputs and the labels
- Forward/Backprop = Alternating conditional expectation, with constraints

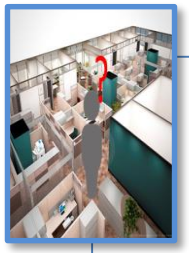
What Do We Gain Conceptually?

- Every weight in every layer computes conditional expectations;
- Two way selection of informative features, where are they?
- Other learning algorithms viewed the same way: PCA, CCA, Compressed Sensing…

Where Do We Go from Here?

- Generic information vs. task-specific information: where do we put the prior, costs, and other constraints?
- Supervised vs. unsupervised: common information between more than two random variables – multi-terminal neural networks?
- Separation vs. no separation: transfer learning/multi-task learning, data sharing.

NOW, SOME EXAMPLES



PLoc: Powerline indoor occupancy sensing

MetroEYE： Measuring Fine-grained Metro Interchange Time

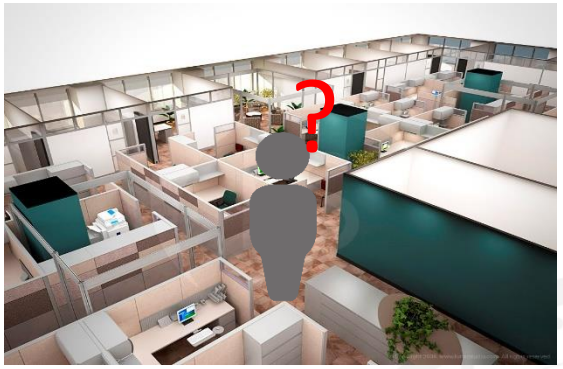
Occupancy detection via Footstep induced building vibration

电力线室内定位 | PLoc: Occupancy Sensing via Power Line

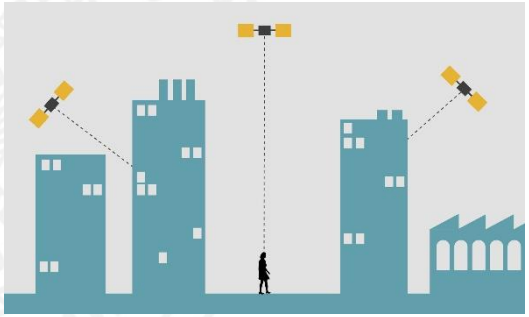
Indoor Localization and Occupancy Sensing

Indoor localization information is essential in many pervasive applications in commercial buildings.

- Estimate the user walking patterns.
- Measure occupancies of room for energy saving.
- Optimize space utilization.



Global infrastructure based: GPS, etc.

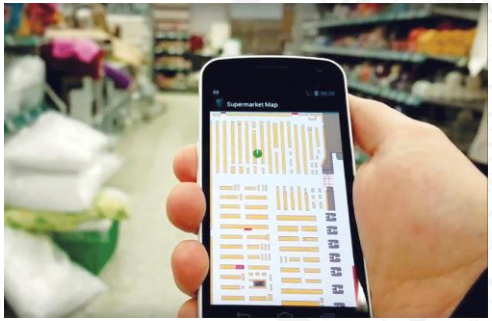


- Bad performance due to blockage of satellite signals.

Local infrastructure based: Camera, microphone, PIR sensors, UWB radar, etc.

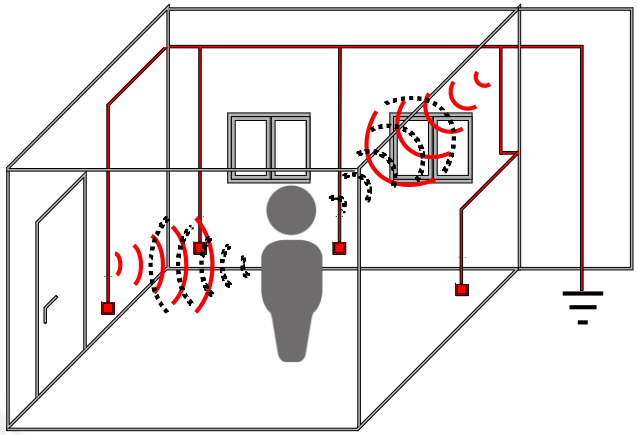
- Large deployment and maintenance costs.

Wearable: Smartphones, etc.

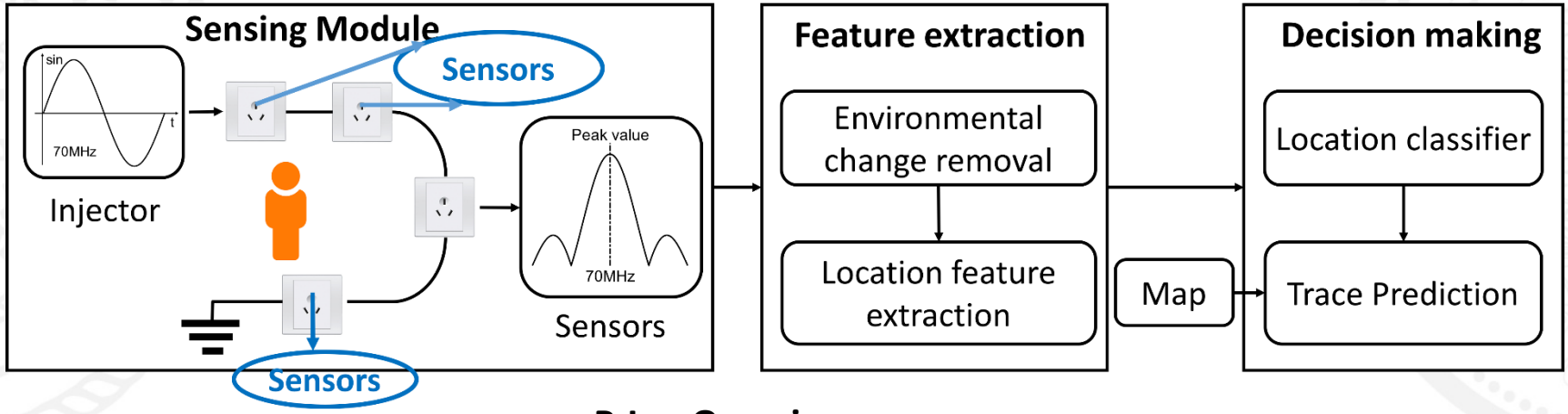


- The hardware is not carried all the time.

Our method: P-Loc Powerline-Localization

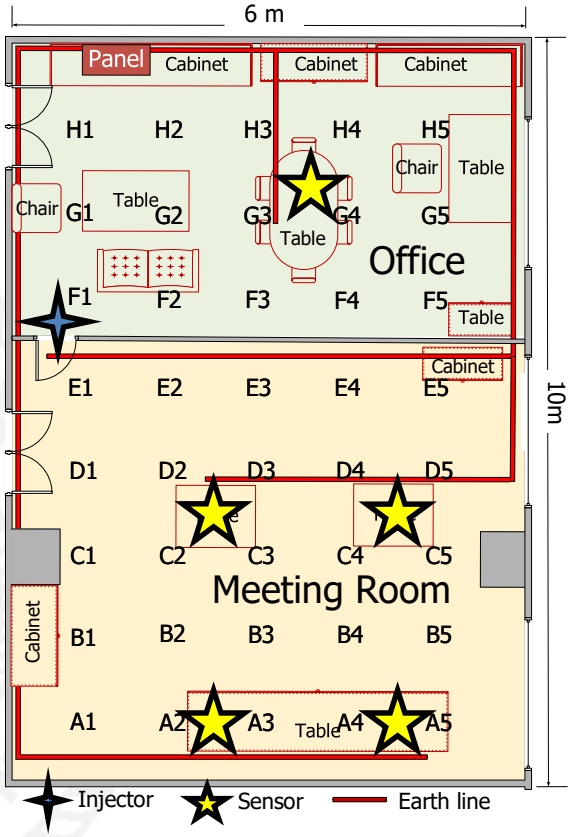


- Human body: **conductor**.
- Powerline: viewed as an **antenna**.
- Human body's location <- Signal changes captured by the powerline.

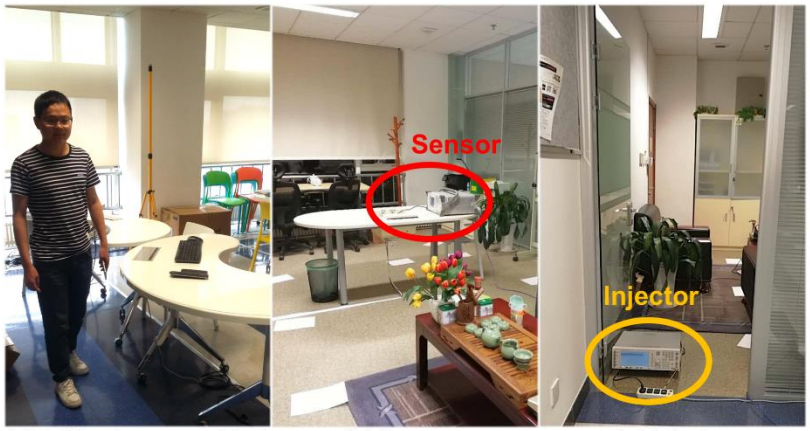


P-Loc Overview

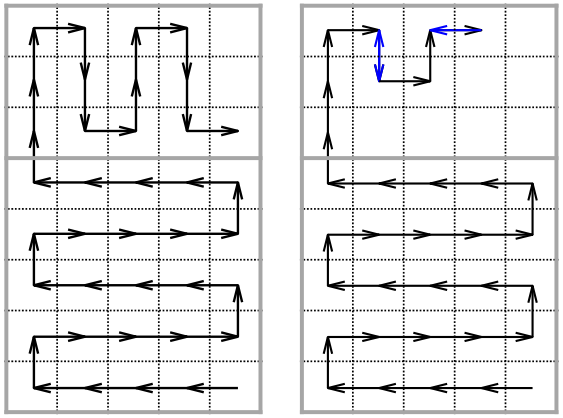
Performance



Experiment Setup



(a) Walking human subject (b) Sensors (c) Injector



IPSN 2017 Best Poster

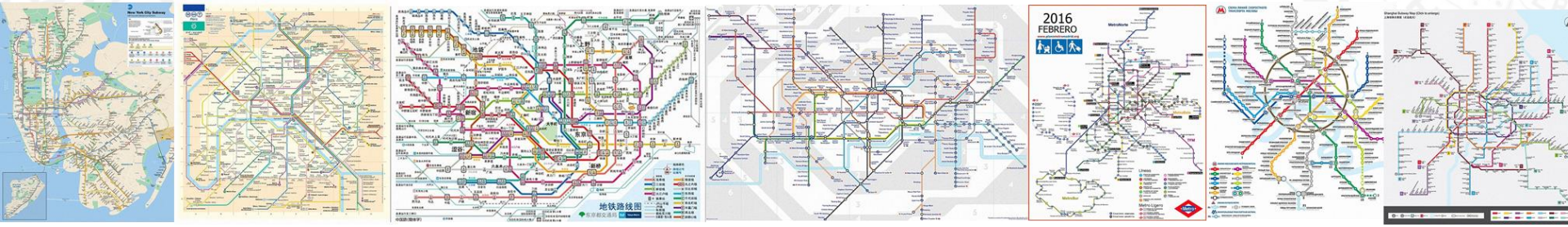
93 % tracking accuracy

MetroEYE：Measuring Fine-grained Metro Interchange Time via Smartphones

TBSI 清华-伯克利深圳学院 Tsinghua-Berkeley Shenzhen Institute

Motivation

- Underground metro has been a major solution to urban traffic problem
  - 55 Countries, 140 Cities with Underground Metro Systems
  - Daily passengers: Shanghai 11.3 M, Beijing 12.7 M, London 4.8 M
- Metro Networks are very Complex
  - Beijing: 53 Interchanging stations, including 3 3-line interchanging stations.
  - Interchanging time is highly variable and a major impacting factor of QoS
  - One minute interchanging time saving for each passenger result in 24 yr time saving every day
- Understanding Interchanging time is important
  - Better planning of station layout: platform, stairs, elevators, etc.
  - Tracking the congestion level of passenger.
  - Assist optimizing the timetable of metros.



MetroEYE: Approach

First tier model:

Magnetic Field

GSM RSSI

Acceleratio

Cost-Sensitive
Naive Bayes Model

In-Metro Interchange

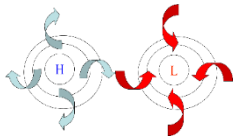
Passenger state inference

Second tier model

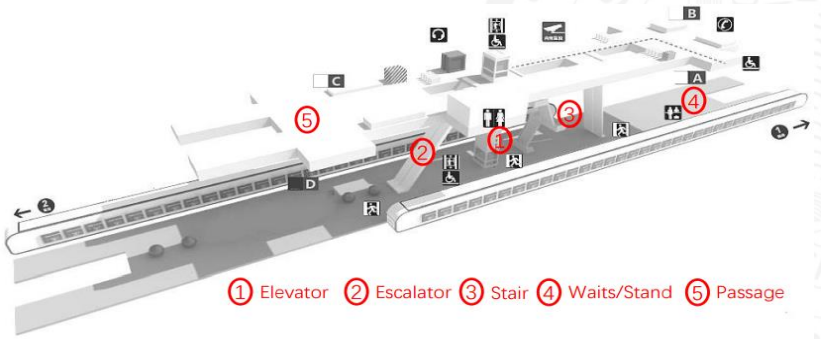
Interchange

Acceleration

Pressure



CRF



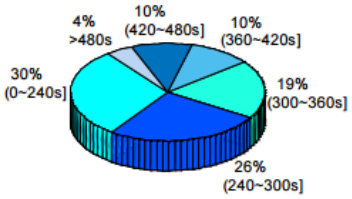
Fine-grained interchange case inference

MetroEYE: Results

Mobiqui-tous 2017 Best Paper

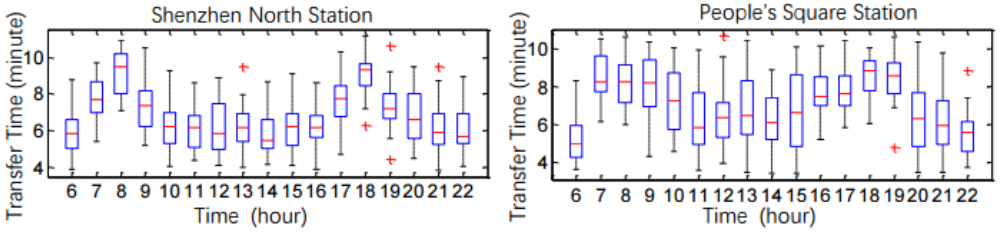
Distribution of Interchange Time Method

Distribution of interchange time

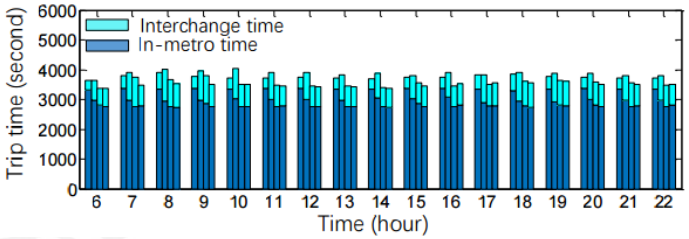


Stations with interchange time over 360s

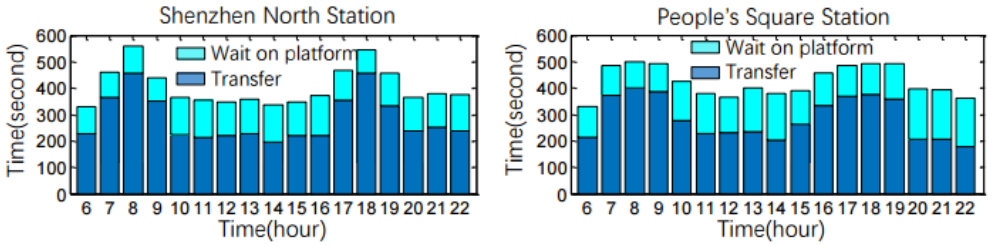
Interchange time variance of day



Interchange Time and *In-Metro* Time



Analysis of Waiting Time for Metro



Person Identification via Steps-induced Vibration

Person identification in smart building enables

- Elderly/Child monitoring
- Enhanced security
- Energy usage profiling

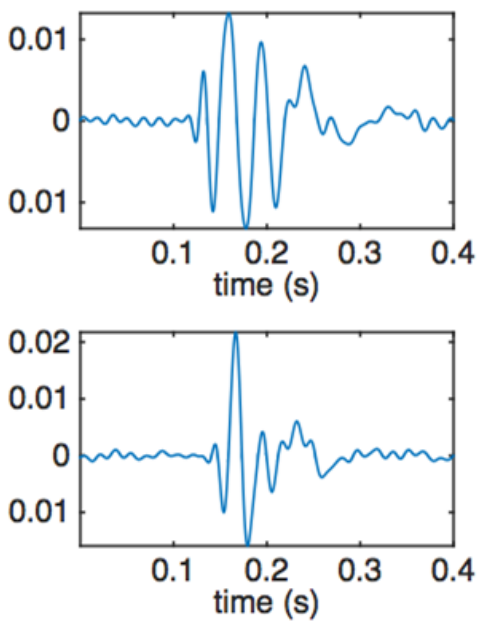
Human footsteps induced floor vibration

- People's gaits are unique (for identification)
- Unique gait induces unique floor vibration
- Floor vibration sensing is sparse, passive and constraint-less

Person Identification through Floor Vibration

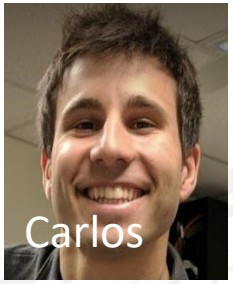
1. Footstep induces vibration

2. Sensing



3. Vibration signal feature extraction

features



4. Classification/ Identification

Person Identification through Floor Vibration

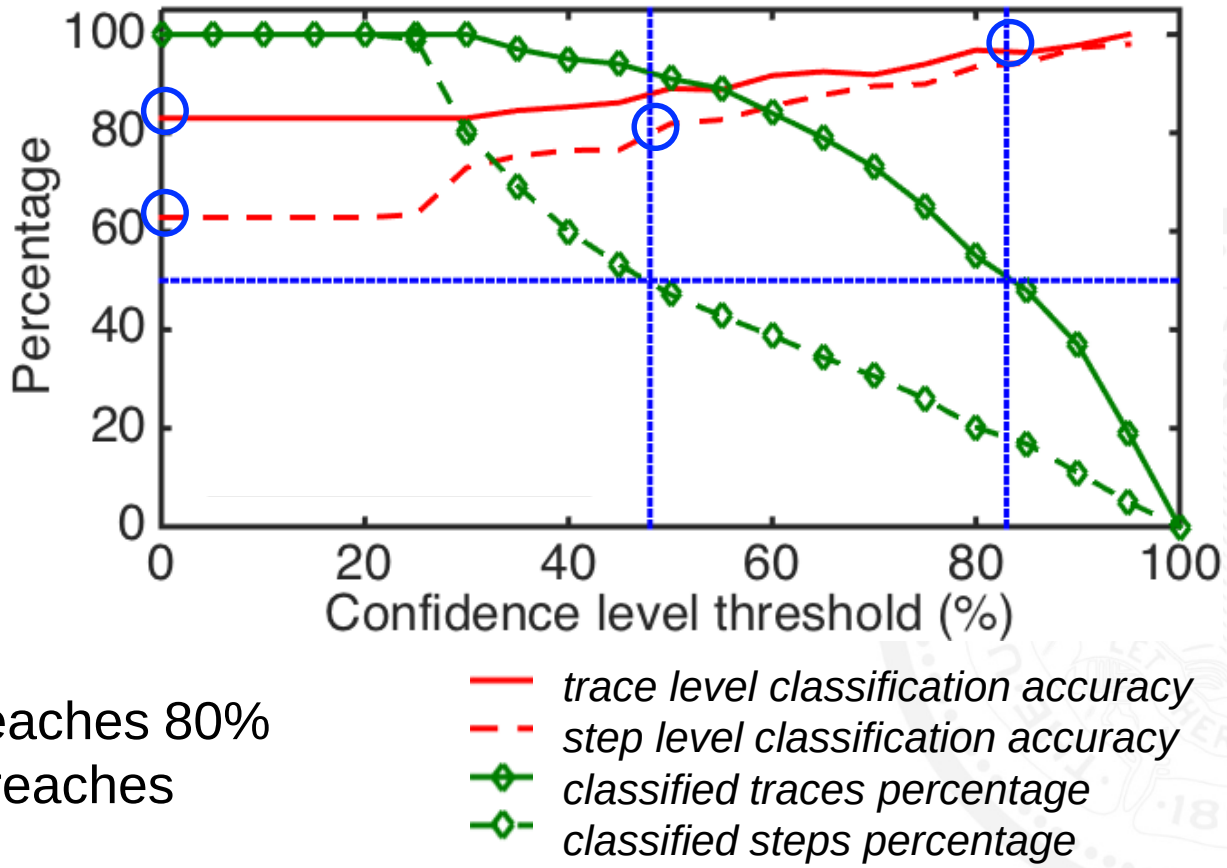
Identify 5 people

Non-thresholding

- Step level classification reaches over 60% accuracy
- Trace level classification accuracy improves over 20%

Thresholding (50% classifiable)

- Step level classification reaches 80%
- Trace level classification reaches 96.5%



About Tsinghua-Berkeley Shenzhen Institute (TBSI.edu.cn)

- Established in 2015, by UC Regent, Tsinghua University with supports from Shenzhen government
- Faculty: 20 Berkeley Professors, 30 Tsinghua Professors, 22 TBSI full-time professors
- Degree program: Ph.D. and Dual master degree, now with 200 students

TBSI is

- a US-style university in China;
- a research hotel;
- a hub for scholars from different disciplines;
- a portal to access China's global problem.