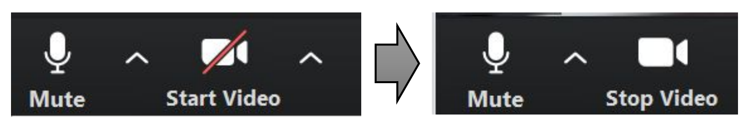
EE5104 Data Science

- Today
  - Sensor Data Science : Process Overview

- Turn your video on 

- Check out lecture materials posted in the syllabus

Up-to-date Syllabus: http://tiny.cc/y3wouz

An Overview of Sensor Data Analysis Process (Mobile vs. Fixed Sensing Cases)

Young Tae Noh

KENTECH

Sensor Data Gathering & Processing

- Getting sensor data
  - From which sensors? (e.g., motion sensors, current sensors)
  - From where? Phone (wearable) vs. factory (stationary)
  - How? (e.g., wireless or wired, hierarchical?)
- Processing sensor data
  - Why? For what? (e.g., activity recognition or fault detection)
  - How (procedure)
    - Sensor data processing pipeline: collect □ segment □ extract □ classify
    - Sensor fusion – leveraging multiple sensors for better classification

Overview

- Mobile Sensing with Smartphones
  - A Survey of Mobile Phone Sensing, IEEE Com Mag, 2010
- Sensor Data Processing Pipeline
  - A tutorial on human activity recognition using body-worn inertial sensors, 2014
- Industrial Applications: Machine Condition Monitoring and Fault Diagnosis
  - Design and deployment of industrial sensor networks: experiences from a semiconductor plant and the north sea, ACM SenSys 2005
  - Novel Industrial Wireless Sensor Networks for Machine Condition Monitoring and Fault Diagnosis, Liqun Hou and Neil W. Bergmann, IEEE Transactions on Instrumentation and Measurement, 2012

Mobile Sensing with Smartphones

A Survey of Mobile Phone Sensing, IEEE Com Mag, 2010

Contents

- Applications
- Eco-system Players
- Scale of Mobile Sensing
- Sensing Paradigm
- Mobile Sensing Architecture
    - Sense
    - Learn
    - Inform, Share, Persuasion
- Privacy Issues

Galaxy S20 Sensors

Accelerometer

Magnetometer (Compass)

Gyroscope

Ambient Light

Proximity

Camera

Voice

Pressure (Barometer)

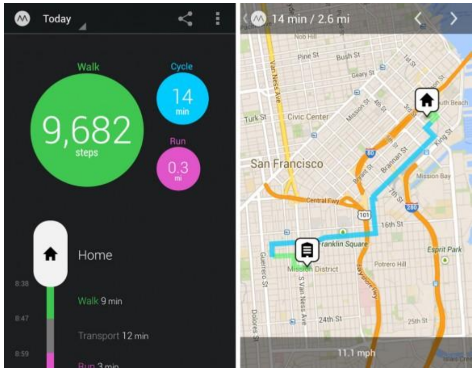
NFC

Heart Rate

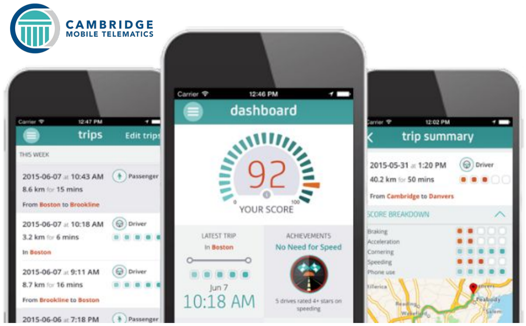
Fingerprint scanner

Applications

- Health and Well Being
  - Promoting personal fitness (UbiFit Garden, Move, Google Fit)
- Transportation
  - Traffic conditions (MIT VTrack, Nokia/Berkeley Mobile Millennium)
  - Driving behaviors (MIT DriveWell)
- Social Networking
  - Sensing presence (Dartmouth CenceMe)
- Environmental Monitoring
  - Measuring pollution (UCLA PIER)



Move

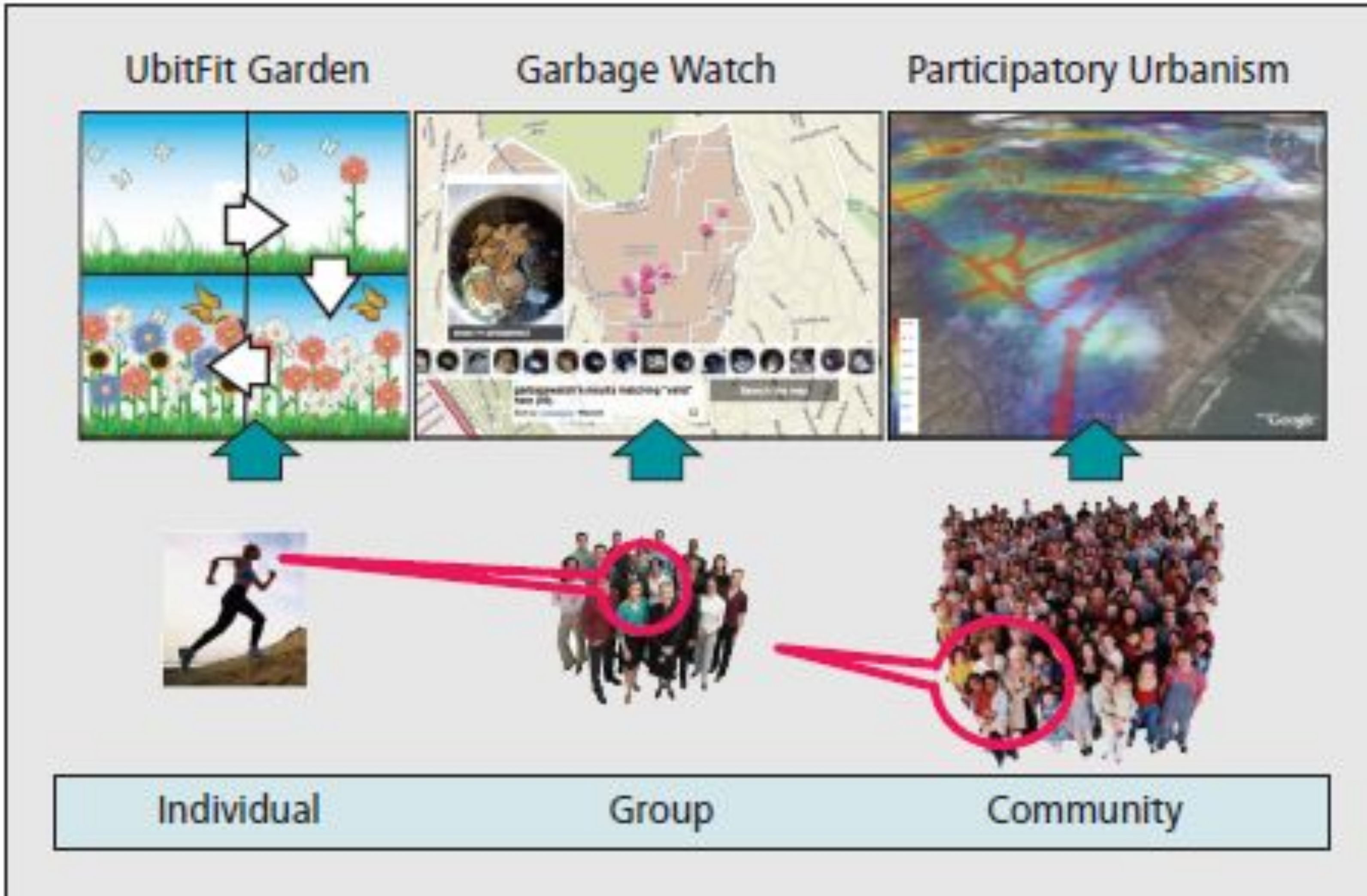


Drive Well

Eco-system Players

- Multiple vendors
  - Apple AppStore
  - Google Play (Android Market)
  - Microsoft Mobile Marketplace
- Developers
  - Startups
  - Academia
  - Small Research laboratories
  - Individuals
- Critical mass of users

Scale of Mobile Sensing



Sensing Paradigm

- Participatory: active sensor data collection by users
  - Example: managing garbage cans by **<u>taking photos</u>**
  - Advantages: supports complex operations
  - Challenges:
    - Quality of data is dependent on participants

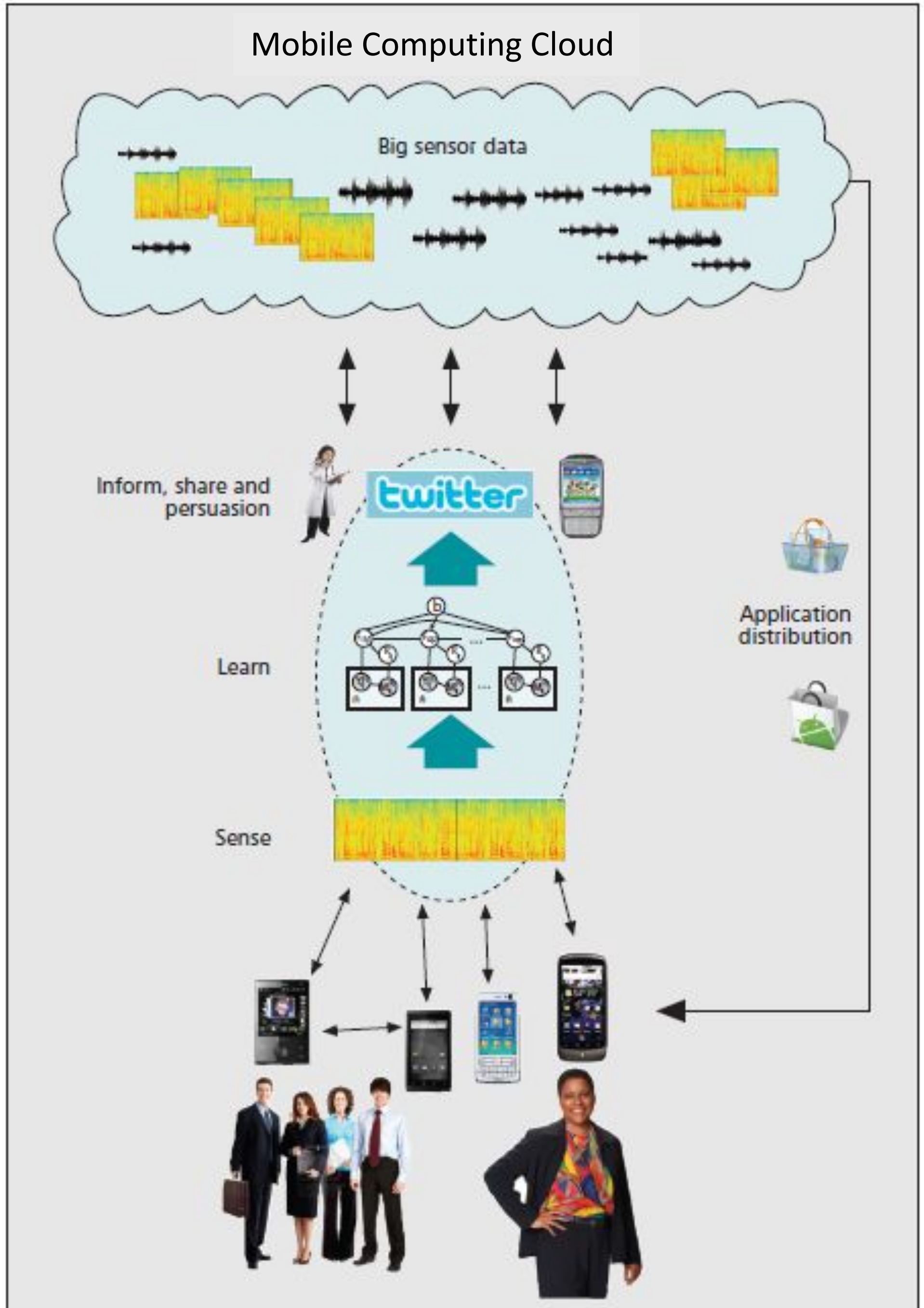
- Opportunistic: automated sensor data collection
  - Example: collecting GPS location traces from users' phone
  - Advantages: lowers burden placed on the user
  - Challenges:
    - Technically hard to build – people underutilized
    - Phone context problem (dynamic environments)

Mobile Sensing Architecture

INFORM, SHARE AND PERSUASION

LEARN

SENSE

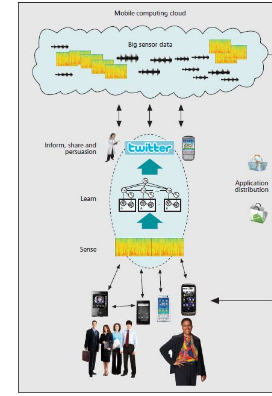


Sense

INFORM, SHARE, PERSUASION

LEARN

SENSE



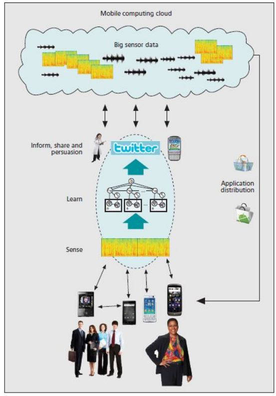
- Programmability
  - Managing smartphone sensors with system APIs
  - Challenges: fine-grained control of sensors, portability (OS & sensor variation)
- Continuous sensing
  - Resource demanding (e.g., computation, battery)
  - Energy efficient algorithms; trade-off between accuracy and energy consumption
- Phone context
  - Dynamic environments affect sensor data quality
  - Some solutions:
    - Admission controls for removing noisy data
    - Collaborative multi-phone inference (i.e., using multiple sensors)
- Time consuming
  - Most labor intensive work in sensor data science
  - Sensor data + label collection

Learn

INFORM, SHARE, PERSUASION

LEARN

SENSE

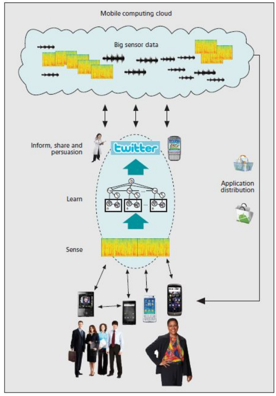


- Integrating sensor data
  - Data mining and statistical analysis
- Learning algorithms
  - Supervised: data are hand-labeled (e.g., cooking, driving)
  - Semi-supervised: some of the data are labeled
  - Unsupervised: none of the data are labeled
- Example: human behavior and context modeling
  - Activity classification
  - Mobility pattern analysis (place logging)
  - Noise mapping in urban environments

INFORM, SHARE, PERSUASION

LEARN

SENSE



Learn: Scaling Models

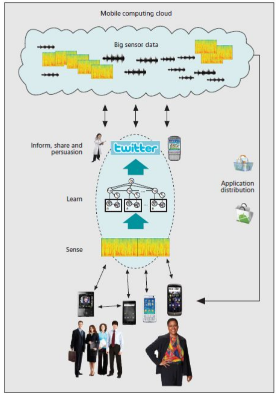
- Scaling model to everyday uses
  - Dynamic environments; personal differences
  - Large scale deployment (e.g., millions of people)
- Models must be adaptive and incorporate people into the process
- If possible, exploit wisdom of crowd (or crowdsourcing) to improve data classification and solutions
- Challenges:
  - Lack of common machine learning toolkits for smartphones
  - Lack of large-scale public data sets
  - Lack of public repository for sharing datasets, code, and tools

Inform, Share, Persuasion

INFORM, SHARE, PERSUASION

LEARN

SENSE



- Sharing
  - Data visualization, community awareness, and social networks
- Personalized services
  - Profile user preferences, recommendations, persuasion
- **Persuasive technology** – systems that provide tailored feedback with the goal of changing user's behavior
  - Motivation to change human behavior (e.g., healthcare, environmental awareness)
  - Methods: self-reflection, goal setting, social competitions
  - Interdisciplinary research combining behavioral and social psychology with computer science

Privacy Issues

- Respecting the privacy of the user is the most fundamental responsibility of a mobile sensing system
- Reconstruction type attacks
  - Reverse engineering collected data to obtain invasive information
- Second-hand smoke problem
  - How can the privacy of third parties be effectively protected when other people wearing sensors are nearby?
  - How can mismatched privacy policies be managed when two different people are close enough to each other for their sensors to collect information?

Privacy Issues

- Understanding of privacy issues of novel mobile and wearable technologies is required
- Furthermore, stronger techniques for protecting people's privacy are needed

- Current solutions
  - Cryptography
  - Privacy-preserving data mining
  - Processing data locally versus cloud services
  - Group sensing applications is based on user membership and/or trust relationships

Summary

- Applications
  - Health & Well-being, transportation, SNS, environmental monitoring,
- Eco-system Players
  - Vendors, developers, users
- Scale of Mobile Sensing
  - Individual, Group, Community
- Sensing Paradigm
  - Participatory vs. Opportunistic
- Mobile Sensing Architecture
  - Sense
  - Learn
  - Inform, Share, Persuasion
- Privacy Issues

Case Study of Mobile Sensing: *Human Activity Recognition using Smartphones*

A tutorial on human activity recognition using body-worn inertial sensors, Andreas Bulling, Ulf Blanke, Bernt Schiele, ACM Computing Surveys (CSUR) Volume 46 Issue 3, January 2014

About Activities

- High-level activities
  - Giving a lecture, having a breakfast, playing soccer…
- Low-level activities
  - Lying on a bed, standing still, running, walking…



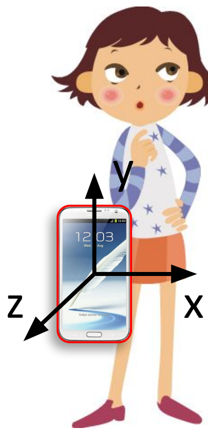



Case Study

- Consider three target activities to recognize
  - Running
  - Standing still
  - Lying on a bed

- How can we recognize these activities using your smartphone’s motion sensors?

Activity Recognition Process



Motion sensors

accelerometer compass gyroscope

↓ Data acquisition and pre-processing

Sensor Data

↓ Segmentation

Data Segment

↓ Feature extraction

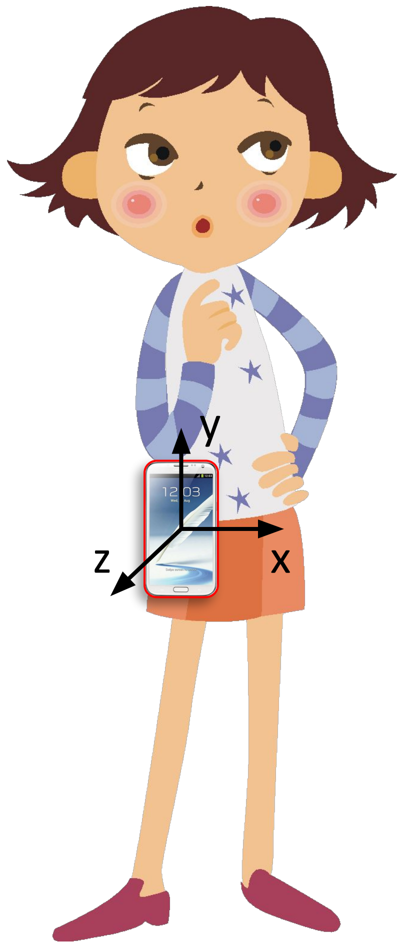
Features

↓ Model building & Classification (Inference)

Activity

Activity Recognition Process

Phone shaken?	Phone orientation?	Current Activity
No	Upright	Standing still
Yes	Upright	Running
No	Lying down	Lying on a bed
Yes	Lying down	Nothing (=Null)



Average value of accelerometer y-axis sensor signals for the last 2 seconds (if avg. Y-axis >= alpha, it is upright; otherwise, lying)

Variance of accelerometer sensor signal for the last 2 seconds (if variance >= beta, it is shaken)

Activity Recognition Process

Phone shaken?	Phone orientation	Activity
No	Upright	Standing still
Yes	Upright	Running
No	Lying down	Lying on a bed
Yes	Lying down	…?

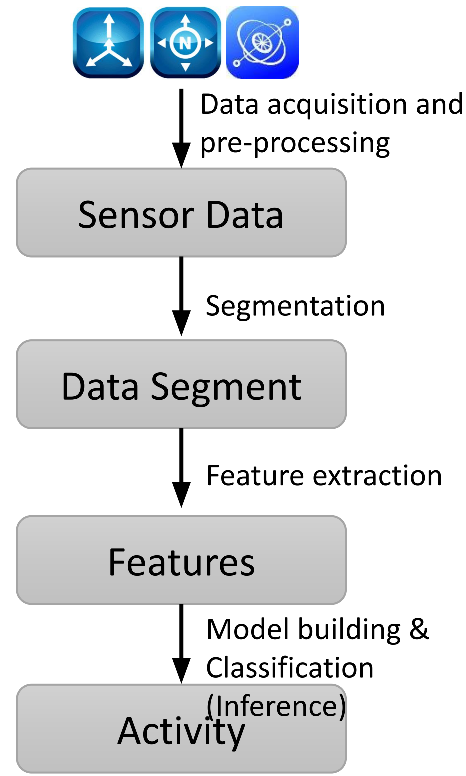
Classification

Average value of accelerometer y-axis sensor signals for the last 2 seconds

Feature extraction

Segmenting (=windowing)

Variance of accelerometer sensor signal for the last 2 seconds



Data Acquisition & Pre-processing

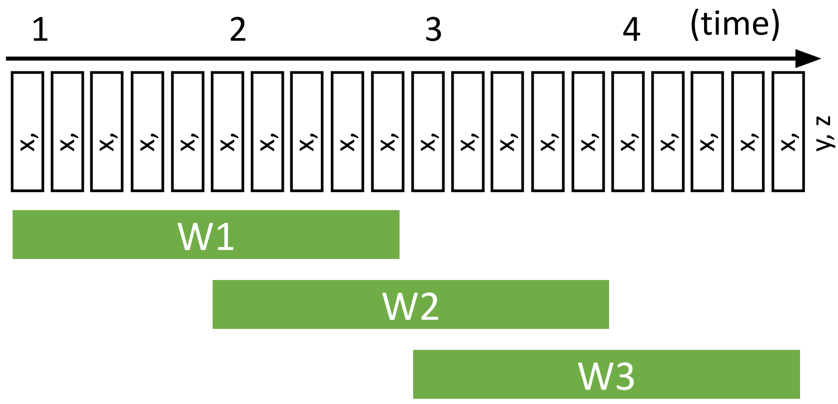
- Collecting a stream of sensor data (e.g., using Android's sensor manager interface)
- Since most sensors provide data on some regular basis, we also need to know **sampling rate** (will learn more about this during DSP sessions)
- An accelerometer, for example, may provide a stream of tuples of real numbers representing the acceleration in x, y and z-direction with 5 Hz
- Cf) Android's sensing rate configuration
  - Predefined rates: SENSOR_DELAY_NORMAL, UI, GAME, FASTEST
    - Or, the desired delay between events in microseconds
  - Actual rate is device-dependent; e.g., Nexus 5 (Normal/UI: 15 Hz; Game: 50 Hz; Fastest: 200 Hz)
  - Note that your smart devices will not guarantee such rates, and actual rate is dependent on its operating conditions (e.g., workload)

Data Segmentation

- For feature extraction, we need to "identify" those data segments that are likely to contain information about activities (known as "activity detection" or "spotting)
  - **Sliding window**: using a window (=frame) of samples, and simply slide that window with fixed overlapping (e.g., 50%)
  - Energy based: different activities have different activity "intensities" (or energy) (e.g., rest vs. others) – moving avg. can be used for automatic segmentation
- In our example, to recognize basic physical activities we collect the data of 2 seconds from the accelerometer.
- This corresponds to 10 readings of the acceleration data (if sampling rate is fixed to 5Hz)

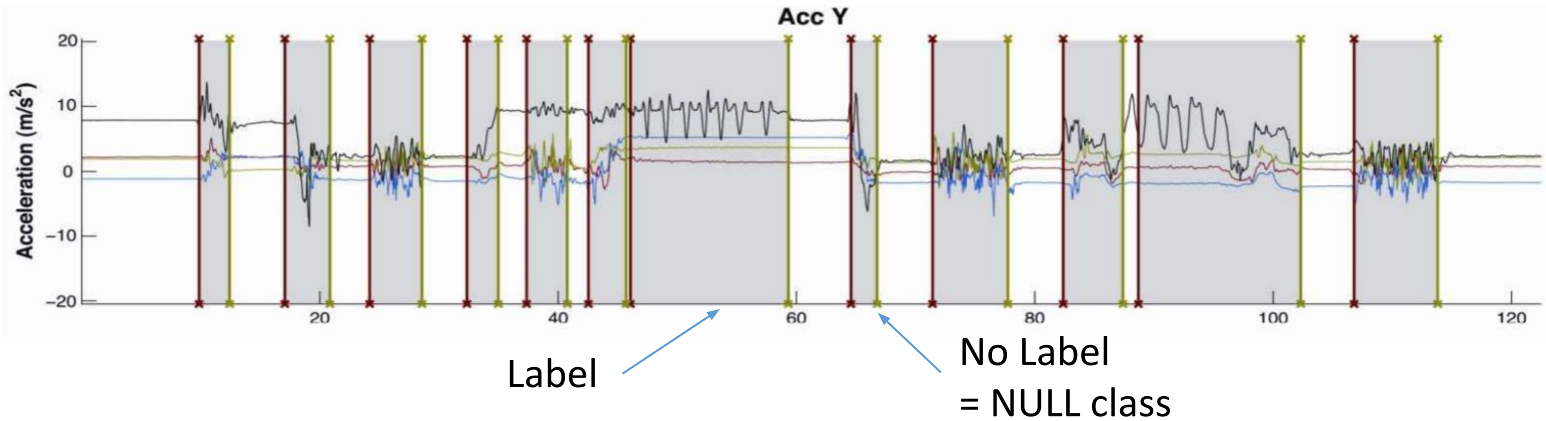
Example

Segmenting the sensor data using a window of 2 seconds with an overlap of 1 second



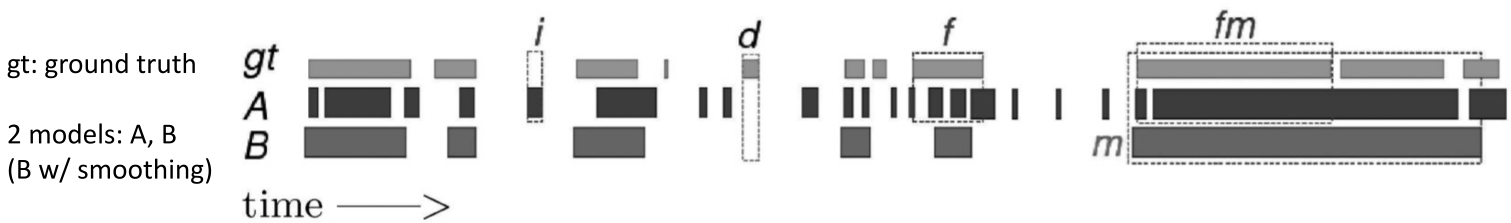
Practical Issues

- Continuous sensing – not every data has labels
  - Sliding window may result in wrongfully classifying "NULL" class to some other classes



https://doi.org/10.4108/ICST.BODYNETS2009.6036

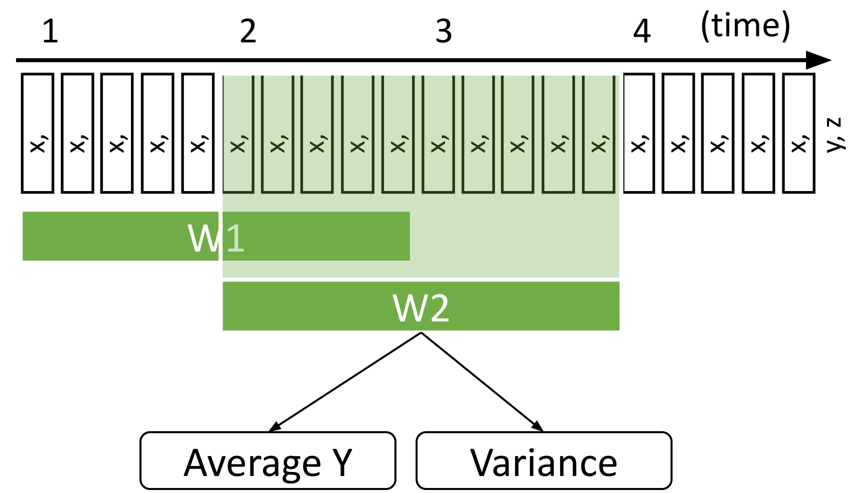
- Errors: insertion, deletion, fragmentation, merging



https://dl.acm.org/doi/pdf/10.1145/1889681.1889687

Feature Extraction

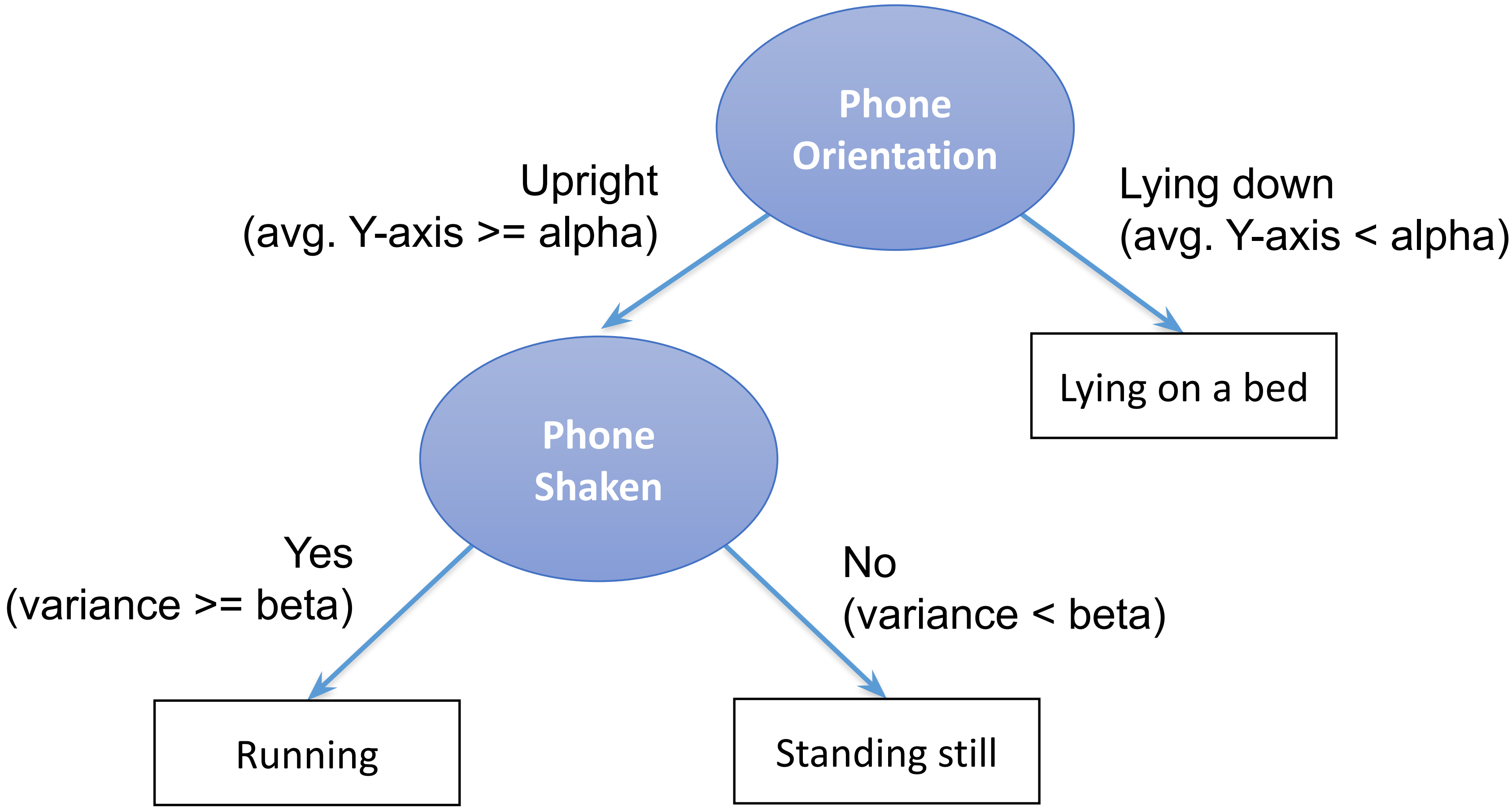
- How to extract features?
  - Signal-based features – mean, variance, kurtosis
  - Body model features – exploiting prior knowledge about human kinematics
  - Event-based features – if there are any events (e.g., a sequence of eye movements – saccades, fixations, or blinks)
  - Multilevel features – duration, frequency, co-occurrence, clustered data/labels
- In our examples, we extract the following features:
  - Average value of accelerometer's y-axis signals
  - Variance of accelerometer sensor signals



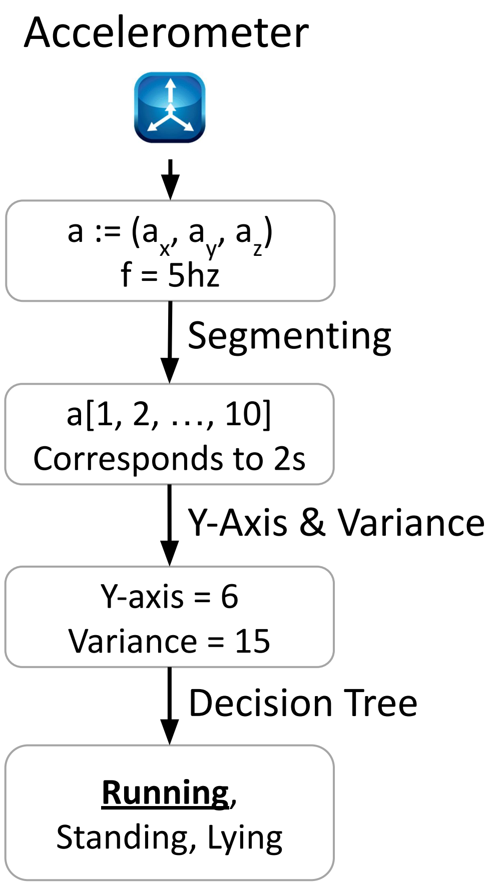
Classification

- After extracting important features from the raw data, we use a classifier to determine the current activity
- Many different classification algorithms exist, and depending on the application domain, there may be one algorithm that shows the best performance
- Depending on the algorithm, the result is either a crisp decision (e.g., decision tree), or a probability distribution over activities (e.g., Naïve Bayes)
- Learning algorithms: supervised vs. unsupervised (based on whether training dataset is used or not)

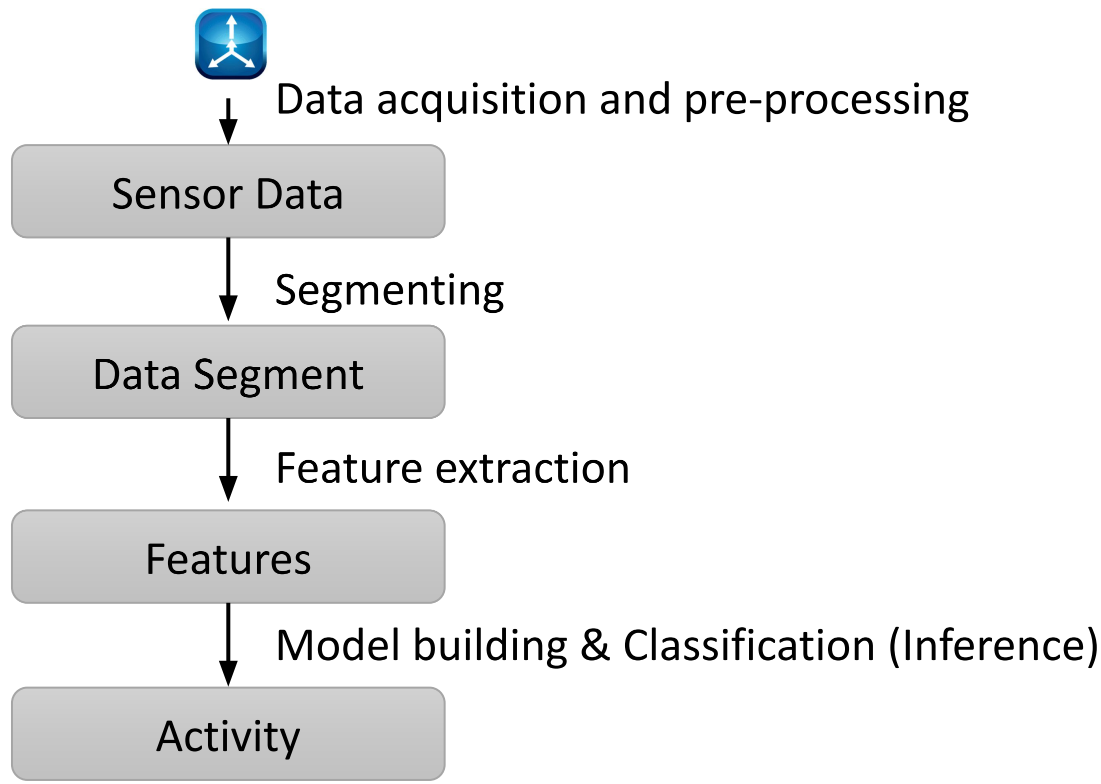
Classification Example: Decision Tree



Summary



Accelerometer
$a := (a_x, a_y, a_z)$
f = 5hz
Segmenting
a[1, 2, …, 10]
Corresponds to 2s
Y-Axis & Variance
Y-axis = 6
Variance = 15
Decision Tree
Running,
Standing, Lying



Data acquisition and pre-processing
Sensor Data
Segmenting
Data Segment
Feature extraction
Features
Model building & Classification (Inference)
Activity

Application Research Papers

- RecoFit: Using a Wearable Sensor to Find, Recognize, and Count Repetitive Exercises, ACM CHI 2014
- iSleep: Unobtrusive Sleep Quality Monitoring using Smartphones, Sensys 2013
- Dog's life: Wearable Activity Recognition for Dogs, Ubicomp 2013
- Automatic Assessment of Problem Behavior in Individuals with Developmental Disabilities, Ubicomp 2012
- Using Mobile Phones to Determine Transportation Modes, ACM TOSN 2010

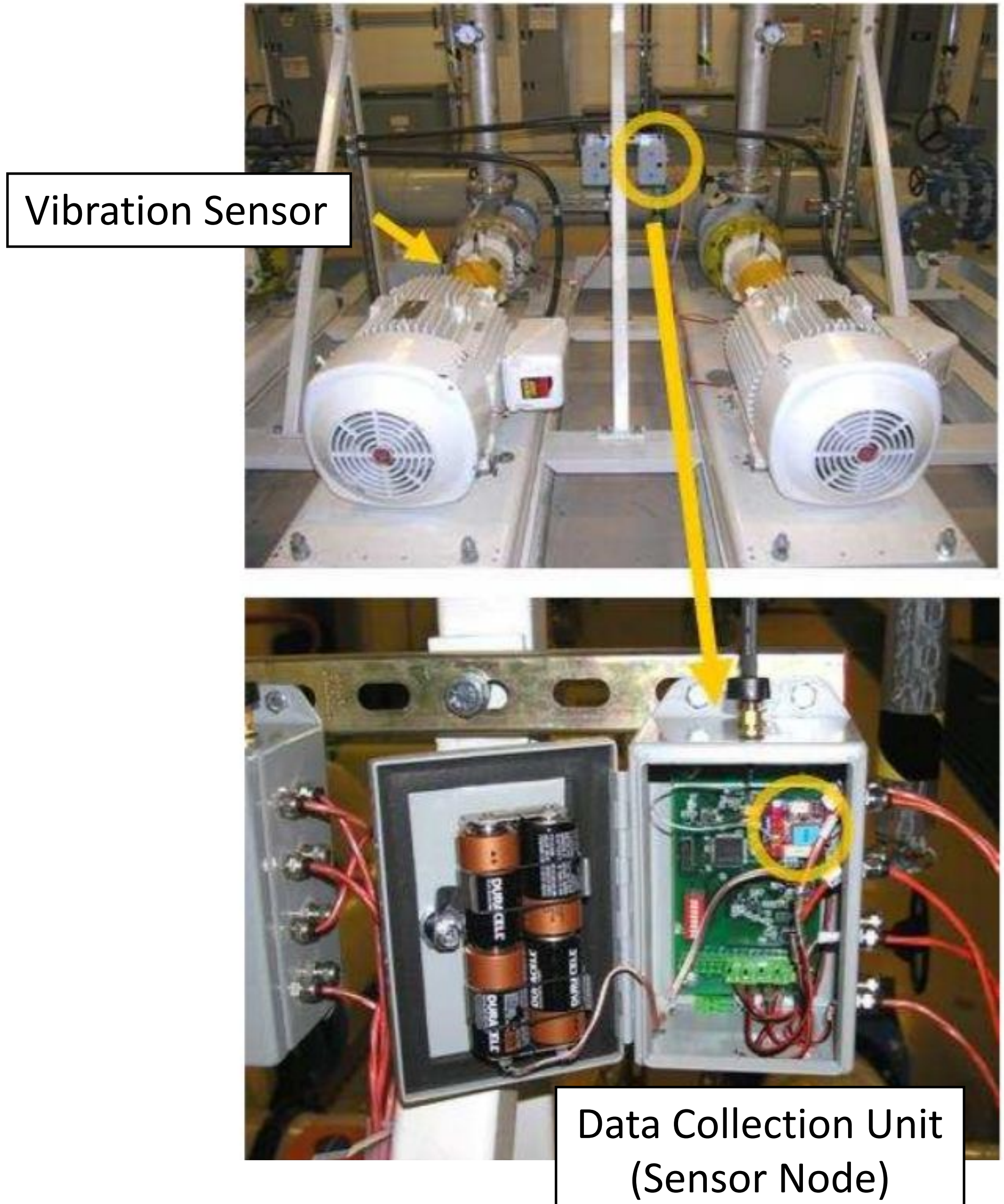
Industrial Applications: Machine Condition Monitoring and Fault Diagnosis

***Design and deployment of industrial sensor networks: experiences from a semiconductor plant and the north sea**, ACM SenSys 2005*

Predictive Analysis for Machine Condition Monitoring using Sensors

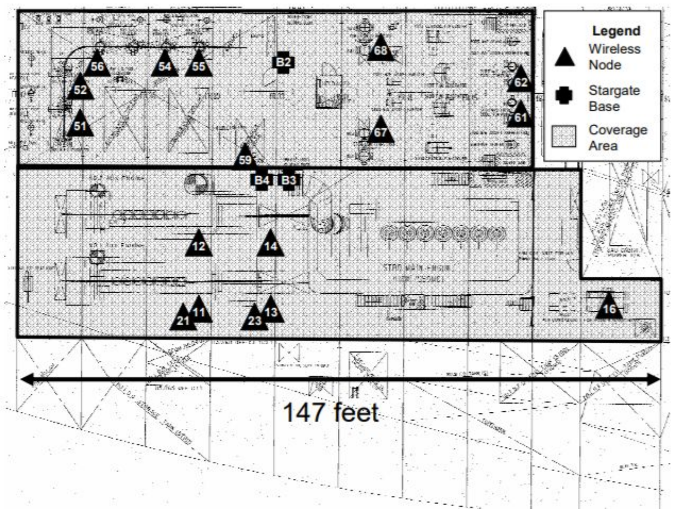
- Vibration analysis: checking amplitude/frequency patterns of machine vibration
- Oil analysis: checking particles, viscosity, acidity
- Temperature analysis w/ thermal camera or temperature sensors (e.g., thermocouples): finding abnormal heat sources by comparing to baseline data for temperature changes
- Ultrasonic detection: checking ultrasonic frequencies to detect faults (e.g., corrosion, wear patterns)

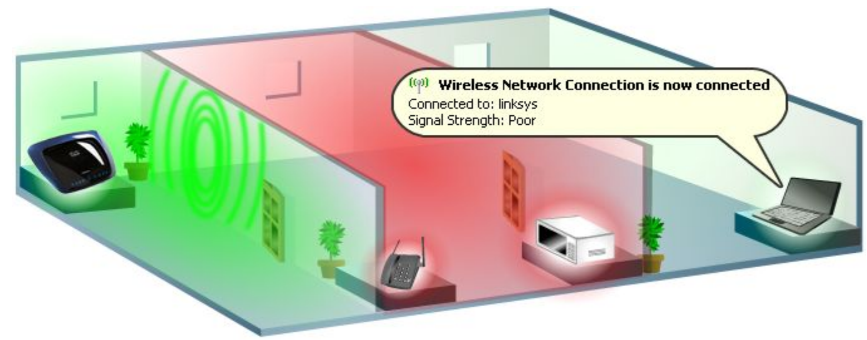
Sensors



How to Configure Sensor Nodes? Wireless vs. Wired?

How to Configure Sensor Nodes? Wireless vs. Wired? 1) RF Coverage/Bandwidth/Interference





RF Coverage: How are sensors wirelessly **covered**? How much **bandwidth** can they deliver?

- Conduct a **site-survey** to identify shadows caused by obstructions in the environment
- Help to decide whether to add resources, such as relay nodes or additional gateways, to ensure coverage

RF Interference: How is the **quality** of wireless connections?

- Are there any interference sources?
- What's the network Quality of Service (QoS) (e.g., bandwidth, delay) under RF interference?

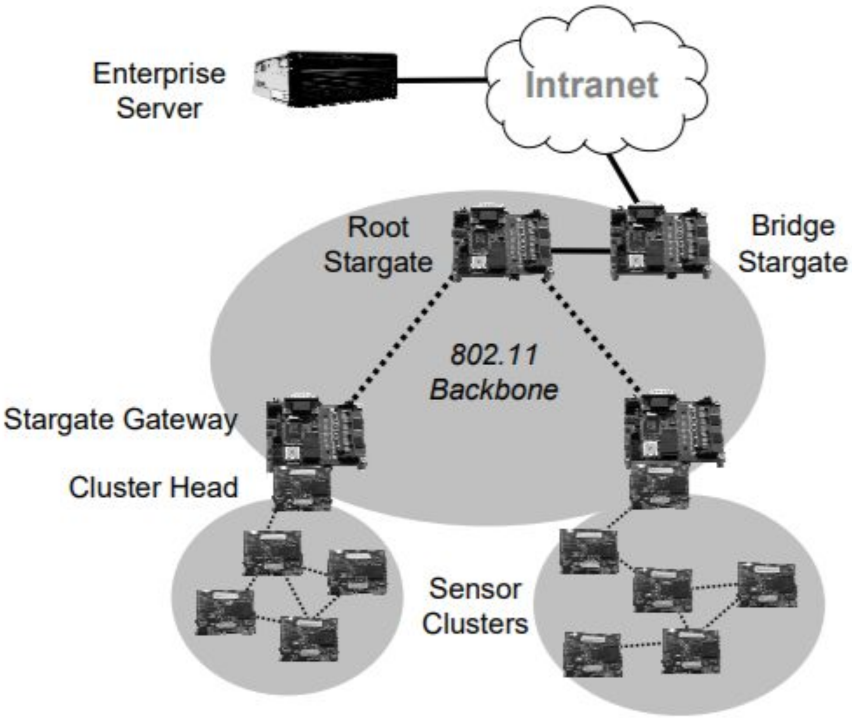
How to Configure Sensor Nodes? Wireless vs. Wired? 2) Cost

	Manual Collection	Online System	Wireless Data / Wired Power
# Wired APs	0	450	35
# Wireless APs	0	0	875
# Analyzers	8	1	1
Hardware Costs			
Sensors (installed)	$1,260,000	$1,260,000	$1,260,000
Wired APs	$0	$2,250,000	$17,500
Wireless APs	$0	$0	$262,500
Analyzers	$144,000	$18,000	$18,000
Installation Costs			
Wired APs	$0	$3,375,000	$262,500
Wireless APs	$0	$0	$1,726,974
Labor (Collection Costs)	$168,000	$3,360	$3,360
Total Costs	**$1,572,000**	**$6,906,360**	**$3,550,834**
Total Costs w/o Sensors	**$312,000**	**$5,646,360**	**$2,290,834**

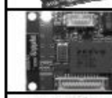
Industry Requirements

- Fault tolerance and reliability
- Long-lived battery powered operations
- Maintainable
- Seamless integration into existing applications
- Security

Deployment Experience

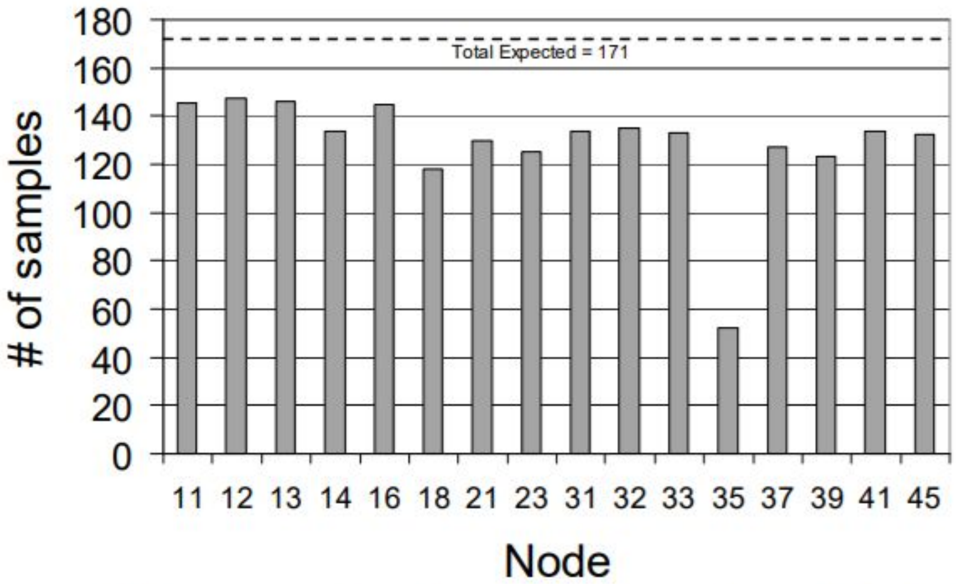


Network Architecture

Platform	Description
	Mica2 Sensor Node: Atmel AtMega128L, Chipcon 900 Mhz radio, Battery powered.
	Intel Mote Sensor Node: ARM Core, Zeevo Bluetooth radio, Battery Powered
	Stargate Gateway Node: Intel XScale® processor (PXA255), 802.11b radio, serially-connected Mica2/Intel Mote, wall powered.

- 1 Server
- 4 Stargates
- 26 sensor nodes
- 150 accelerometers

Deployment Experience



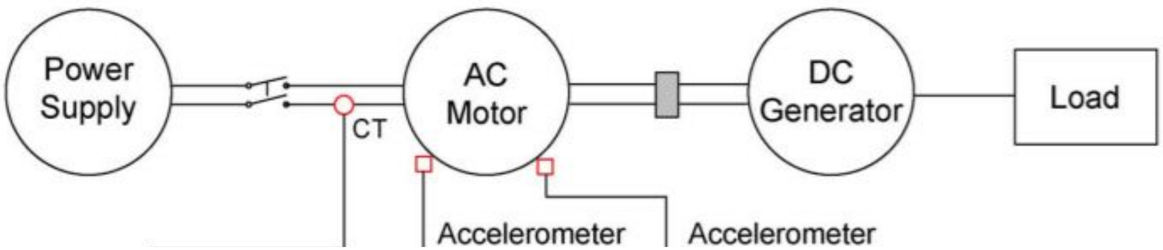
Histogram of total number of vibration samples received/node from the starboard deployment of a 19 week period

- Majority of the nodes successfully delivered results at least **80%** of the time
- **Failures** were highly correlated within a particular sensor network cluster

Industrial Applications: Machine Condition Monitoring and Fault Diagnosis

Novel Industrial Wireless Sensor Networks for Machine Condition Monitoring and Fault Diagnosis, *Liqun Hou and Neil W. Bergmann, IEEE Transactions on Instrumentation and Measurement, 2012*

AC Motor Failure Monitoring

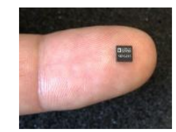


Current Values (Amperes)

Acceleration Values (m/s^2)

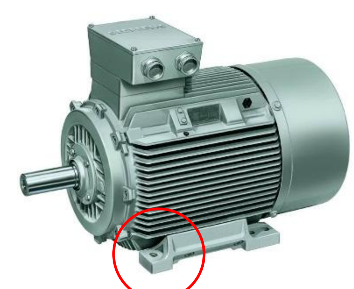
Current transformer (CT) sensor

Industrial vibration sensor



MEMS sensor

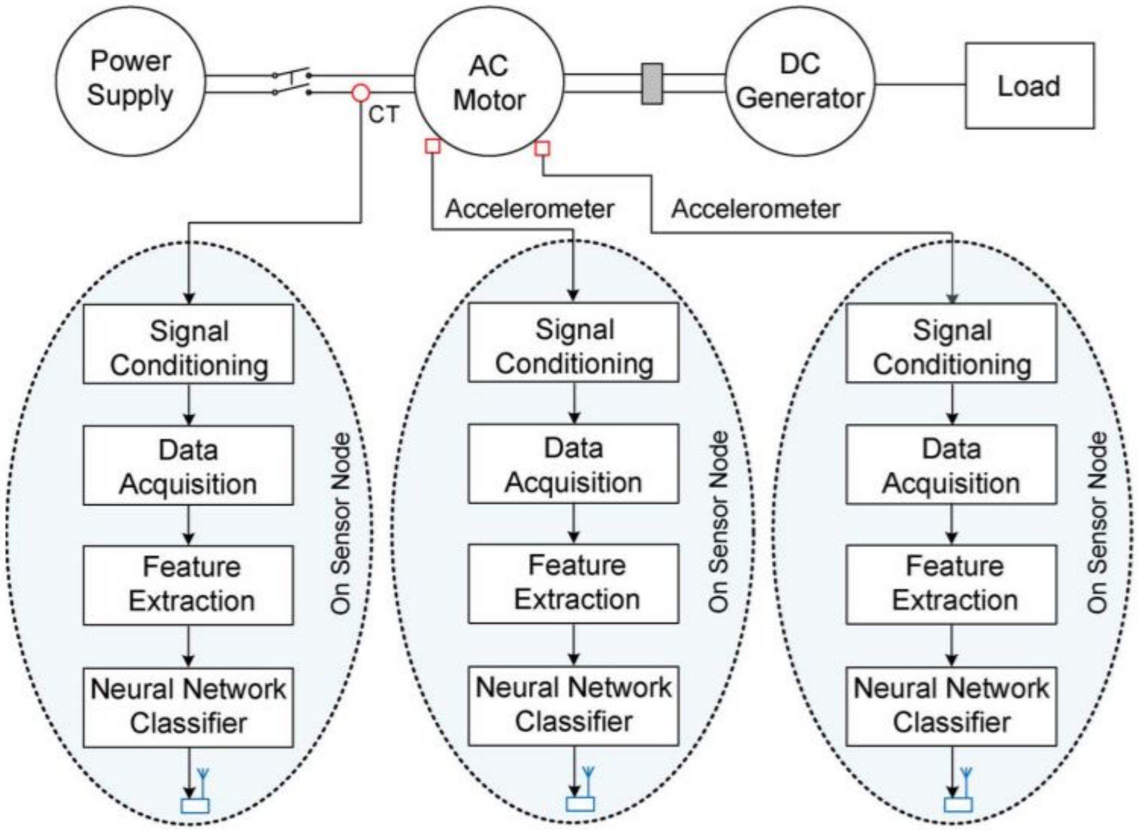
- (H_OC) Healthy motor without resistor load on the DC generator side (the dc generator is open circuit)
- (H_R15) Healthy motor with 15-Ω resistor load on the DC generator
- (LF_R15) Fault 1: motor subjected to loose feet with 15-Ω load
- (18g I_R15) Fault 2: 18-g imbalance with 15-Ω load



Loose feet

Cf. CT sensors: Current transformers (CTs) are sensors that measure alternating current (AC)

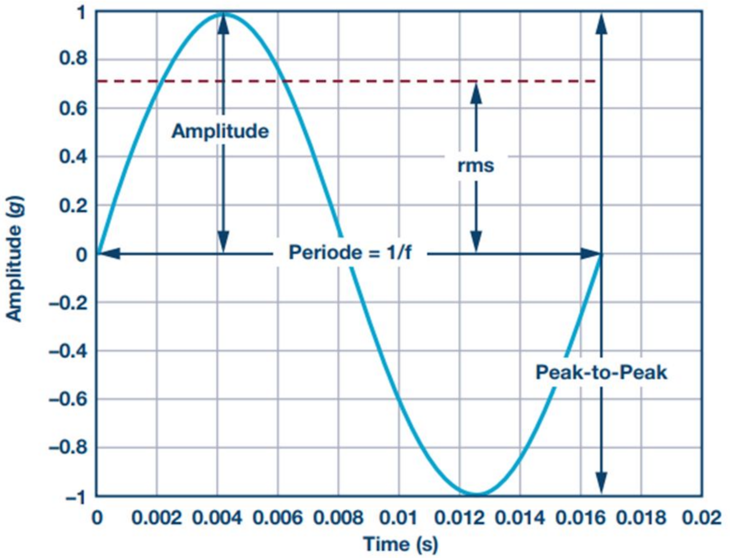
AC Motor Failure Monitoring



Stator current (SC) **Vibration/Head (VH)** **Vibration/End (VE)**

Cf. CT sensors: Current transformers (CTs) are sensors that measure alternating current (AC)

Feature Extraction



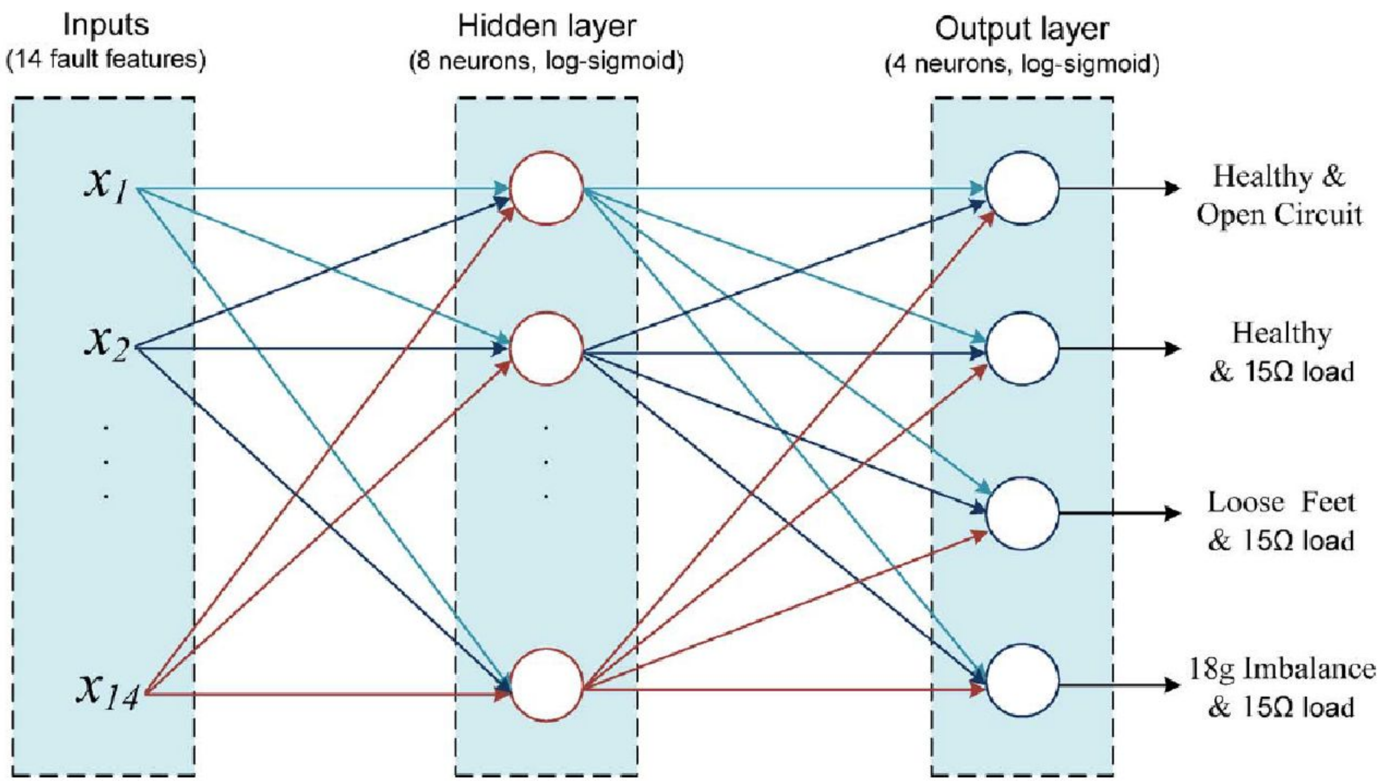
		Time domain	Frequency Domain					
Vibrations	→	P-P, Variance	(2,3) f_b	(5,6) f_b	(7-9) f_b	(10,11) f_b	(13,14) f_b	(16,17) f_b*
			(19,20) f_b	(21,22) f_b	(23-25) f_b	(26-67) f_b	(68-93) f_b	>93 f_b
Current	→	P-P, Variance						

* This frequency component is about 100 Hz, twice of the line frequency. f_b is the resolution of FFT, $f_b = 6$ Hz.

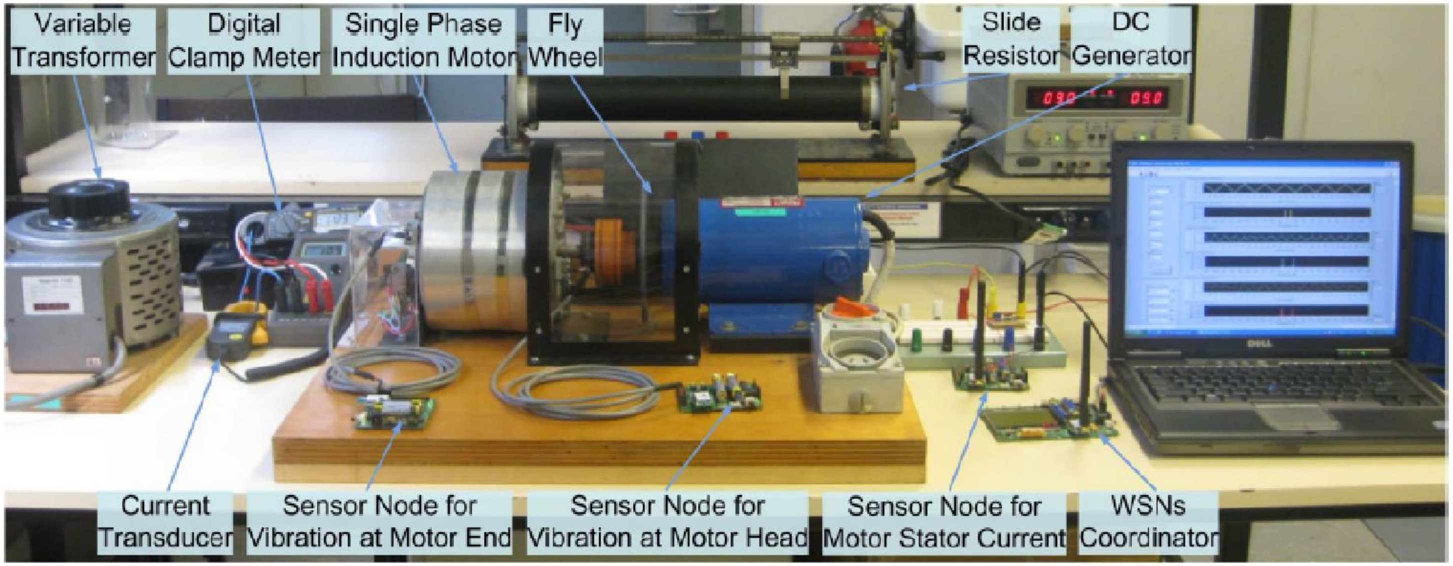
* Acceleration signature peak-to-peak (P-P) amplitude and variance values from a window of 512 samples

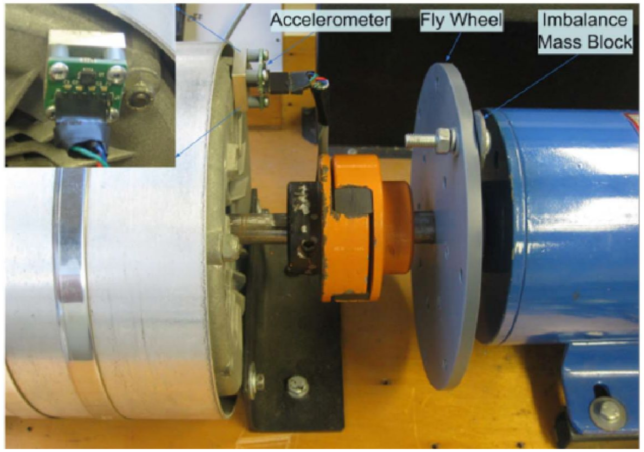
Cf. CT sensors: Current transformers (CTs) are sensors that measure alternating current (AC)

Model Training



Data Collection Environment





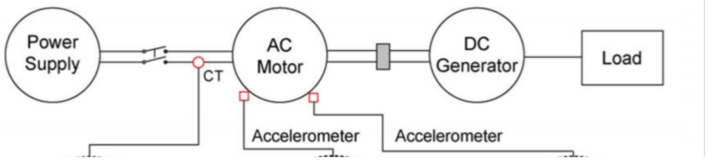
- Motor rotational speed is about 960 r/min (= 16 r/s)
- JN5139 sensor board:
  - IEEE 802.15.4 and ZigBee protocols
- Dataset:
  - Training data: 15 measurements for each condition (a total of 60 measurements)
  - Testing data: 6 measurements for each condition (a total of 24 measurements)

Performance Comparison

Test Pattern	Classifier	Classification Results			
		H_OC	H_R15	LF_R15	18g I_R15
H_OC*	VH	**0.7705**	0.2175	0.0112	0.0005
	VE	**0.8826**	0.0677	0.0285	0.0210
	SC	**0.9936**	0.0004	0.0004	0.0053
H_R15	VH	0.4897	**0.5094**	0.0003	0.0003
	VE	0.1721	**0.7713**	0.0011	0.0553
	SC	0.0005	**0.9256**	0.0005	0.0732
LF_R15	VH	0.0004	0.0250	**0.9724**	0.0019
	VE	0.0950	0.1033	**0.7856**	0.0159
	SC	0.0004	0.0032	**0.9957**	0.0004
18g I_R15	VH	0.0004	0.0582	0.0004	**0.9407**
	VE	0.0390	0.1455	0.0030	**0.8123**
	SC	0.0005	0.9190	0.0056	**0.0748**

The worst diagnosis results for each motor operating condition are shaded

H_OC *Healthy motor without resistor load (open circuit)*
H_R15 *Healthy motor with 15-Ω resistor load on the DC generato*
LF_R15 *Fault 1: motor subjected to loose feet with 15-Ω load*
18g I_R15 *Fault 2: 18-g imbalance with 15-Ω load*



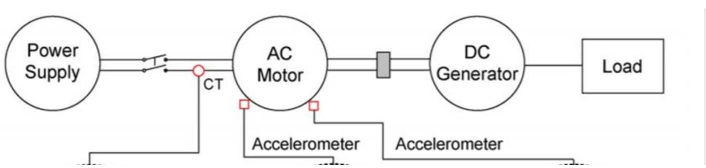
Stator current (SC) **Vibration/Head (VH)** **Vibration/End (VE)**

Performance Comparison

Test Pattern	Classifier	Classification Results			
		H_OC	H_R15	LF_R15	18g I_R15
H_OC*	VH&VE	**0.9776**	0.0211	0.0004	0.0000
	VH&VE&SC	**0.9989**	0.0000	0.0000	0.0000
H_R15	VH&VE	0.1671	**0.8318**	0.0000	0.0000
	VH&VE&SC	0.0001	**0.9982**	0.0000	0.0000
LF_R15	VH&VE	0.0001	0.0033	**0.9960**	0.0000
	VH&VE&SC	~~0.0000~~	~~0.0000~~	**~~0.9990~~**	~~0.0000~~
18g I_R15	VH&VE	0.0000	0.0109	0.0000	**0.9883**
	VH&VE&SC	~~0.0000~~	~~0.1456~~	~~0.0000~~	**~~0.8401~~**

The best diagnosis results with conditional fusion are shaded

H_OC *Healthy motor without resistor load (open circuit)*
H_R15 *Healthy motor with 15-Ω resistor load on the DC generato*
LF_R15 *Fault 1: motor subjected to loose feet with 15-Ω load*
18g I_R15 *Fault 2: 18-g imbalance with 15-Ω load*



Stator current (SC) **Vibration/Head (VH)** **Vibration/End (VE)**

Summary

- Getting sensor data
  - From which sensors? (e.g., motion sensors, current sensors)
  - From where? Phone (wearable) vs. factory (stationary)
  - How? (e.g., wireless or wired, hierarchical?)

- Processing sensor data
  - Why? For what? (e.g., activity recognition or fault detection)
  - How (procedure)
    - Sensor data processing pipeline: collect □ segment □ extract □ classify
    - Sensor fusion – leveraging multiple sensors for better classification

Summary

- Mobile Sensing with Smartphones
  - Scale of mobile sensing (individual, group, community)
  - Sensing paradigm: participatory vs. opportunistic
  - Sense => Learn => Inform, Share, and Persuade
- Sensor Data Processing Pipeline
  - Key steps: Data Collection => Segmentation => Feature Extraction => Model Building => Evaluation
- Industrial Applications: Machine Condition Monitoring and Fault Diagnosis
  - Manual vs. Online (wired vs. wireless) : cost & efficiency matters
  - Wireless sensor networks: lower costs (no wires), but be careful about RF inferences, data loss, and battery issues
  - Multiple sensors can be used for diagnosing machine faults