Smart Phone-Based Sensor Mining

A tutorial

These slides available from http://storm.cis.fordham.edu/~gweiss/presentations.html

Gary M. Weiss
Fordham University
gweiss@cis.fordham.edu

What is a Smart Phone?

- What is a smart phone and what does it do? What devices can it replace?
  - Play along and for now forget the topic of this talk
  - A smart phone is:
    - A mobile wireless communication device (a "phone")
    - A network computer: Web access, email, and computing
    - A music device (MP3 player) and a gaming device
    - A camera & video recorder
    - A calendar, address book, memo pad– a PDA
    - Also a very <u>diverse</u> sensor array

Can You Guess the Sensors?

- What sensors are found on smart phones?
  - Audio sensor (microphone)
  - Image sensor (camera, video recorder)
  - Tri-Axial Accelerometer
  - Location sensor (GPS, cell tower, WiFi)
  - Proximity sensor (infrared); Light sensor
  - Magnetic compass; Temperature sensor
  - Virtual/calculated sensors:
    - Proximity (via light), gravity, orientation, gyroscope

The Advent of Smart Phones

- Smart phone growth is extremely strong
  - In 4th quarter of 2010 exceeded PC sales first time[1]
- Smart phones becoming ubiquitous
  - We carry them everywhere we go
- Smart phones are becoming more powerful
  - Faster, more memory, and more sensors!
- Other devices behave similarly (have sensors)
  - Portable game & MP3 players (Gameboy, iPod Touch), tablet computers (iPad, Xoom)

Data & Sensor Mining

- Data mining: application of computational methods to extract knowledge from <u>data</u>
  - Most data mining involves inferring predictive models, often for classification
- Sensor mining: application of computational methods to extract knowledge from <u>sensor data</u>
- Smart phone sensor mining: …
- This tutorial does not focus on mining *methods*
  - Since the methods are not new but smart phone sensor mining is new

THE RIGHT TIME FOR SMART PHONE SENSOR DATA MINING

"The number of diverse and powerful sensors on smart phones, combined with their mobility and ubiquity, combined again with their increasing computational power, makes this the right time for work on Smart Phone-Based Data Mining"

– Gary Weiss

Goals for this Tutorial

- Provide basic introduction to the area
  - Taxonomy of the work that has been done
  - Highlight some of the many applications
- Encourage/motivate/promote R&D
  - Creative applications waiting to be discovered!
- Identify challenges and opportunities
  - Highlight relevant engineering issues

Who Might be Interested in This?

- This tutorial will not be overly technical and should be of interest to a wide audience
  - Those interested in expanding use of data mining
  - Those interested in expanding use of sensors
  - Those interested in mobile communications and ubiquitous computing
  - Those interested in interesting software apps and impacting the world (and perhaps getting rich)

A Little Bit About Myself …

- Previous research focused on fundamental issues related to data mining (class imbalance)
  - While important, not so interesting to undergrads and little immediate impact
- Two years ago started what is now WISDM
  - Android based with papers on activity recognition, and hard and soft biometrics, design & architecture
  - In process of deploying working apps
  - Project has ability to make impact on large population of users

Tutorial Overview

- Relatively quick overview:
  - Tour of main application areas
  - Research challenges and engineering issues
- More detailed examination
  - Some common themes & issues
  - Survey of key application areas
  - Architecture and design Issues
- Finishing Touches
  - Relevant workshops, conferences, & journals

Overview of Application Areas

- Who is the user?
  - Biometric identification & identifying traits
- What is the user doing?
  - Activity recognition
- Where and When is the user?
  - Location and spatial based data mining applications
  - Temporal based data mining applications
- Who, What, Where, When, and Why?
  - Social networking & context sensitive applications

Overview of Architecture & Design

- Mobile platforms:
  - which platform to use & tradeoffs
- Resource constraints
  - Battery, CPU, RAM, bandwidth, …
    - Moore's law implies battery biggest future concern
- Security and privacy
- Architecture
  - How much on client vs. server

Survey of Application Areas

But First: Common Themes & Issues

Method for Collecting Training Data

- Training data is needed to build predictive models for activity recognition etc.
- For some applications labeled training data requires no extra effort (e.g., hard biometrics)
  - The label is the identity and if we know the owner of the phone then labels are easy
- For many applications labels are not free
  - Researcher can control the training phase
  - But for popular apps we need easy self-training
    - One study has users label activities[2] & another location types[21]

Two Types of Predictive Models

- Universal Model vs. Personal Model
  - Universal model: built on one set of users and then applied to everyone else
    - No requirement on new user– no run-time training
  - Personal model: acquire training data for user & then generate model
    - Places data collection requirement on user, but may sometimes by easily automated
  - Personal models almost always do significantly better, even using much less training data[15,16]

Feature Extraction

- Sensor data is time-series data
- Common data mining prediction algorithms expect "examples" and not time-series
  - Typical method moves a sliding window across data to extract higher level features
    - Average acceleration per axis, distribution of acceleration values, speed from GPS data, etc.
    - WISDM uses a 10 second window for activity recognition[15]
    - Other study uses ~7s window with 50% overlap[4]
  - Alternative is to use time series prediction methods directly, but few applications do this

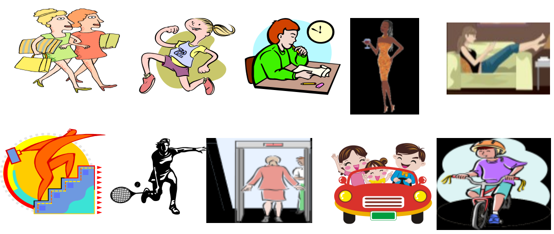
Crowdsourcing

- Crowdsourcing is the outsourcing of a task to a large group or community of people
  - Examples: ESP Game (Google Image Labeler), Amazon Mechanical Turk
- By collecting phone sensor data from many users can create useful apps
  - In “The Dark Knight” Batman relies on a distributed sensor network to track The Joker
  - Google Navigator & many location-based apps

Non-intrusive Interaction

- Ubiquitous sensor mining applications often require non-intrusive interaction with user
  - Apps may provide useful but non-essential information and cannot be distracting
  - PeopleTones[17] system detects and notifies you when a buddy is near using vibrotactile cues.
    - Semantically meaningful auditory cues are most useful
    - PeopleTones has special software to convert auditory cues into vibrations.
  - CenceMe[21] allows user to bind a gesture to action or state (e.g., a circle means “going to lunch”.

ACTIVITY RECOGNITION



Why is Activity Recognition Useful?

- Context-sensitive applications
  - Handle phone calls differently depending on context
  - Play music to suit your activity
  - Fuse with other info (GPS) for better results
    - Can confirm you are on subway vs. traveling in a car[19]
  - Untold new & innovative apps to make phones smarter
- Tracking & Health applications
  - Track overall activity levels and generate fitness profiles
  - Detect dangerous situations (falling); care of elderly[5]
- Social applications
  - Link users with similar behaviors (joggers, hunters)

Activity Recognition w/o Smart Phones

- Dedicated accelerometers placed on a variety of body parts[2,13,14,25]
- A single accelerometer but custom hardware
  - Pedometers (limited function); FitBit[8]
- Multi-sensor solutions
  - eWatch[19]: accelometer + light sensor, multiple locs.
  - Smartbuckle: accelerometer + image sensor on belt
- Use Phone but not a central component
  - Motionbands[10] multi-sensor/location transmits data to smart phone for storage

Location on Body of Smart Phone

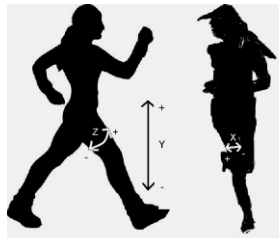
- The location of the smart phone will impact activity recognition
  - WISDM study currently assumes phone in pocket[15]
  - CenceMe study showed pocket and belt clip yield similar results[21]
  - Phone in pocket book & elsewhere needs study
- Phone orientation can have impact
  - WISDM study indicates may not be a problem
  - Can correct for orientation using orientation info

Smart Phone Accelerometer

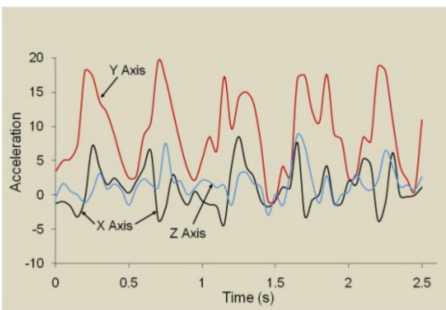
- Measures acceleration along 3 spatial axes
- Detects/measures gravity
  - Orientation impacts g values
- Measurement range typically -2g to +2g
  - Okay for most activities but falling yields higher values
  - Range & sensitivity may be adjustable
- Sampling rates ~20-50 Hz
  - Study found 20Hz required for activity recognition[4]
  - WISDM project found could not reliably sample beyond 20Hz (50ms) and this might limit activity recognition[18]

Accelerometer Data for Six Activites

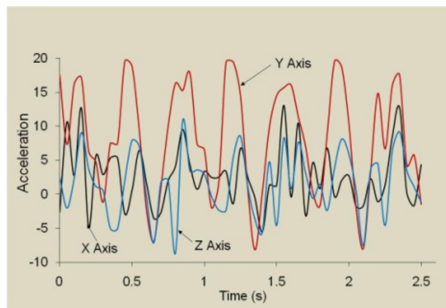
- Accelerometer data from Android phone[15]
  - Walking
  - Jogging
  - Climbing Stairs
  - Lying Down
  - Sitting
  - Standing



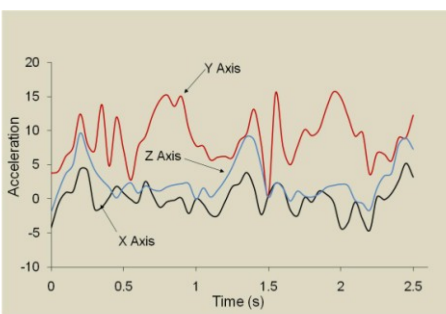
Accelerometer Data for "Walking"



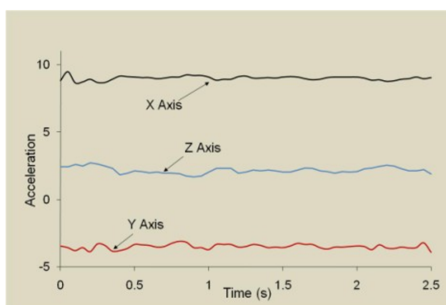
Accelerometer Data for "Jogging"



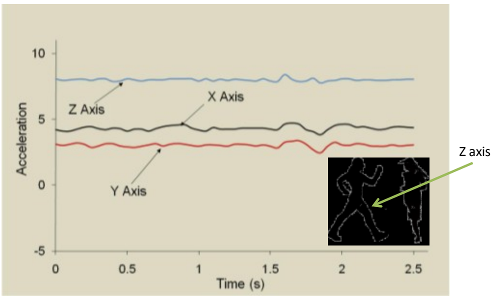
Accelerometer Data for "Up Stairs"



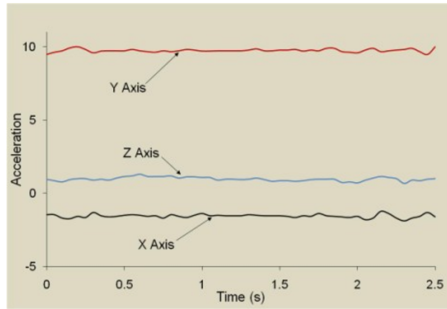
Accelerometer Data for "Lying Down"



Accelerometer Data for "Sitting"



Accelerometer Data for "Standing"



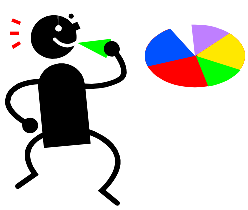
Fall Detection

- Mainly focused on helping the elderly
  - Aging populations will yield great future challenges
- Mostly camera & accelerometer based
  - May also use acoustic or pressure sensors
  - GE QuietCare: camera-based system (nursing homes)
- Accelerometer-based approach [11,24,27]
  - Sensor at waist generally best
  - Threshold-based mechanism[3] (between 2.5g and 3.5g)
  - Elderly don't accelerate quickly so fall detection easier
  - Most data from simulated falls

Determining Transportation Modes

- Nokia n95 system[23] uses GPS & Accelerometer
  - GIS info may be missing or mode may be ambiguous
  - Modes: stationary, walking, running, biking, motorized
  - Precision & recall both equal 91.3% using a decision tree and 93.6% when using DT combined with HMM
  - Using generalized classifier drops accuracy only 1.1%
  - To save power shuts off GPS when inside
    - Triggers GPS based on change in primary cell phone tower
    - GPS lock takes a while so even trying it occasionally saps power
- Alternatives:
  - use GPS & GIS info[22] or only accelerometer

ACTIVITY RECOGNITION RESULTS



Non-Phone Based System

Activity	Accuracy	Activity	Accuracy
Walking	89.71	Walking carrying items	82.10
Sitting & Relaxing	94.78	Working on Computer	97.49
Standing Still	95.67	Eating or Drinking	88.67
Watching TV	77.29	Reading	91.79
Running	87.68	Bicycling	96.29
Stretching	41.42	Strength-training	82.51
Scrubbing	81.09	Vacuuming	96.41
Folding Laundry	95.14	Lying Down & Relaxing	94.96
Brushing Teeth	85.27	Climbing Stairs	85.61
Riding Elevator	43.58	Riding Escalator	70.56

Non-Phone Based System (cont)

Classifier	Personalized Model	Universal Model
Decision Table	36.32	46.75
Instance-Based	69.21	82.70
C4.5	71.58	84.26
Naïve Bayes	34.94	52.35

Universal models perform best. The increase in the amount of data more than compensates for the fact that people move differently. This does not appear to be the case for phone based systems with measurements on one body location.

WISDM Activity Recognition[15]

- Smart-phone based (Android)
- Six activities: walking, jogging, stairs, sitting, standing, lying down (more to come)
- Labeled data collected from over 50 users
- Data transformed via 10-second windows
  - Accelerometer data sampled (x,y,z) every 50m
  - Features (per axis):
    - average, SD, ave diff from mean, ave resultant accel, binned distribution, time between peaks

WISDM Activity Recognition[15]

- The 43 features used to build a classifier
  - WEKA data mining suite used, multiple techniques
  - Personal, universal, hybrid models built
    - Universal models built using leave-one-out validation
- Architecture (for now) uses "dumb" client
- Basis of soon to be released actitracker service
  - Provides web based view of activities over time

WISDM Results

- WISDM results[15] are presented using:
  - Confusion matrices and accuracy
- Results are shown for various things
  - Personal, universal, and hybrid models
  - Most results aggregated over all users but a few per user to show how performance varies by user
  - Results for 6 activities (ones shown in the plots)

WISDM Universal Model- IB3 Matrix

72.4% Accuracy		Predicted Class					
		Walking	Jogging	Stairs	Sitting	Standing	Lying Down
Actual Class	Walking	2209	46	789	2	4	0
	Jogging	45	1656	148	1	0	0
	Stairs	412	54	869	3	1	0
	Sitting	10	0	47	553	30	241
	Standing	8	0	57	6	448	3
	Lying Down	5	1	7	301	13	131

WISDM Personal Model- IB3 Matrix

98.4% accuracy		Predicted Class					
		Walking	Jogging	Stairs	Sitting	Standing	Lying Down
Actual Class	Walking	3033	1	24	0	0	0
	Jogging	4	1788	4	0	0	0
	Stairs	42	4	1292	1	0	0
	Sitting	0	0	4	870	2	6
	Standing	5	0	11	1	509	0
	Lying Down	4	0	8	7	0	442

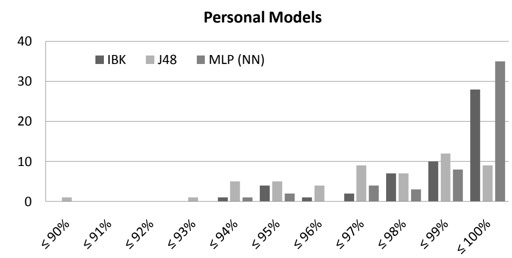
WISDM Hybrid Model- IB3 Matrix

97.1% Accuracy		Predicted Class					
		Walking	Jogging	Stairs	Sitting	Standing	Lying Down
Actual Class	Walking	3028	2	32	2	2	0
	Jogging	5	1803	5	1	0	0
	Stairs	86	13	1288	3	0	0
	Sitting	4	1	6	903	2	24
	Standing	2	0	14	1	520	3
	Lying Down	3	2	5	22	0	421

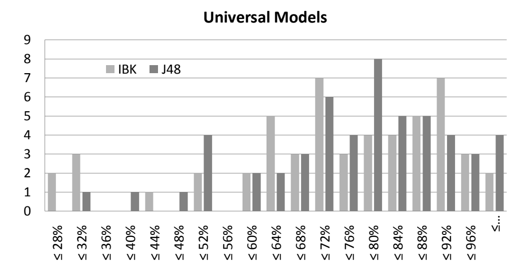
WISDM Accuracy Results

	% of Records Correctly Classified						
	Personal			Universal			Straw Man
	IB3	J48	NN	IB3	J48	NN	
Walking	99.2	97.5	99.1	72.4	77.3	60.6	37.7
Jogging	99.6	98.9	99.9	89.5	89.7	89.9	22.8
Stairs	96.5	91.7	98.0	64.9	56.7	67.6	16.5
Sitting	98.6	97.6	97.7	62.8	78.0	67.6	10.9
Standing	96.8	96.4	97.3	85.8	92.0	93.6	6.4
Lying Down	95.9	95.0	96.9	28.6	26.2	60.7	5.7
Overall	98.4	96.6	98.7	72.4	74.9	71.2	37.7

WISDM Per-User Performance



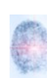
WISDM Per-User Performance



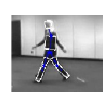
CenceMe Results[21]

	Sitting	Standing	Walking	Running
Sitting	0.682	0.282	0.364	0.000
Standing	0.210	0.784	0.006	0.000
Walking	0.003	0.046	0.944	0.008
Running	0.008	0.070	0.177	0.745

BIOMETRIC IDENTIFICATION








Biometrics

- Biometrics concerns unique identification based on physical or behavioral traits
  - Hard biometrics involves traits that are sufficient to uniquely identify a person
    - Fingerprints, DNA, iris, etc.
  - Soft biometric traits are not sufficiently distinctive, but may help
    - Physical traits: Sex, age, height, weight, etc.
    - Behavioral traits: gait, clothes, travel patterns, etc.

Biometrics for Everyone

- Equipment getting smaller, cheaper
- Biometrics needs sensors and processing
  - Laptops have sensors and processing
    - Face recognition now an option
- Smart phones also have sensors & processing!
  - Camera might be relevant, but so is accelerometer
- Substantial work on gait based biometrics
  - Much of it is vision based since can be used widely
    - Airports, etc.

Gait-Based Biometrics

- Numerous accelerometer-based systems that use dedicated and/or multiple sensors
  - See related work section of *Cell Phone-Based Biometric Identification*[16] for details
- Two smart phone-based biometric systems
  - Possible uses
    - Phone security (e.g., to automatically unlock phone)[9]
    - Automatic device customization[16]
    - To better track people for shared devices
    - Perhaps for secondary level of physical security

Using Time Delay Embeddings

- System from McGill university[9]
  - Provides alternative way of extracting features
  - Used methods from nonlinear time series analysis
  - Uses fewer than a dozen features
  - Runs entirely on Android HTC G1 phone
  - Collected 12-120 seconds of data from 25 people
  - Results: 100% accuracy!
  - Video clip from Discovery channel[7]
    - Shows that can quickly identify a user and use it to unlock phone

WISDM Biometric System

- Same setup as WISDM activity recognition
  - Same data collection, feature extraction, WEKA, …
- Used for identification and authentication
  - Identification means predicting identity from pool of all users (36 in this study)
  - Authentication is a binary class prediction
- Evaluate single and mixed activities
  - Evaluate using 10 sec. and several min. of test data
    - Longer sample classify with “Most Frequent Prediction”

WISDM Biometrics Data Set

	Walk	Jog	Up	Down	Aggregate (Total)
Sum	2081	1625	632	528	4866
%	42.8	33.4	13.0	10.8	100

WISDM Biometric Prediction Results

	Aggregate	Walk	Jog	Up	Down	Aggregate (Oracle)
J48	72.2	84.0	83.0	65.8	61.0	76.1
Neural Net	69.5	90.9	92.2	63.3	54.5	78.6
Straw Man	4.3	4.2	5.0	6.5	4.7	4.3

	Aggregate	Walk	Jog	Up	Down	Aggregate (Oracle)
J48	36/36	36/36	31/32	31/31	28/31	36/36
Neural Net	36/36	36/36	32/32	28.5/31	25/31	36/36

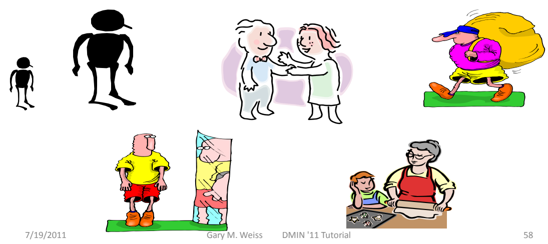
WISDM Biometric Authentication Results

- Authentication results:
  - Positive authentication of a user
    - 10 second sample: ~85%
    - Most frequent class over 5-10 min: 100%
  - Negative Authentication of a user (an imposter)
    - 10 second sample: ~96%
    - Most frequent class over 5-10 min: 100%

Biometric Identification Summary

- Can do remarkably well with short amounts of accelerometer data (10s – 2 min)
- Results may not be good enough for rigorous applications but sufficient for many
  - Automatic customization
  - First level security
    - The system described in the Discovery channel clip unlocked the phone using biometrics to avoid entering a password, which also could be used

TRAIT IDENTIFICATION



Soft Biometrics and User Traits

- Soft biometrics traits are not distinctive enough for identification unless combined with other traits
  - Sex, height, weight, …
- But do we have better uses for these "soft" traits than for identification?
  - As data miners, of course we do!
  - We want to know everything we possibly can about a person. Somehow we will exploit this.
    - We could use weight to improve calories burned

Expanding the Definition of Trait

- Normally think about traits as being:
  - Unchanging: race, skin color, eye color, etc.
  - Slow changing: Height, weight, etc.
- But want to know everything about a person:
  - What they wear, how they feel, if they are tired, etc.
- I have not seen this goal stated in context of mobile sensor data mining
  - It is the focus of *Identifying user traits by mining smart phone accelerometer data*[26]

Related Work

- Very little explicit work on this topic
  - Some work related to biometrics but incidental
    - Work on gait recognition mentions factors that influence recognition, like weight of footwear & sex
- Other communities work in related areas
  - Ergonomics & kinesiology study factors that impact gait
    - Texture of footwear, type of shoe, sex, age, heel height
    - Interaction between gait speed, obesity, and race

WISDM Trait Identification[26]

- Data collected from ~70 people
  - Accelerometer and survey data
  - Survey data includes anything we could think of that might somehow be predictable
    - Sex, height, weight, age, race, handedness, disability
    - Type of area grew up in {rural, suburban, urban}
    - Shoe size, footwear type, size of heels, type of clothing
    - # hours academic work , # hours exercise
  - Too few subjects investigate all factors
    - Many were not predictable (maybe with more data)

WISDM Trait Identification Results

Accuracy 71.2%	Male	Female
Male	31	7
Female	12	16

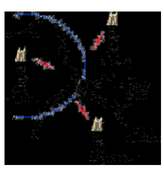
Accuracy 83.3%	Short	Tall
Short	15	5
Tall	2	20

Accuracy 78.9%	Light	Heavy
Light	13	7
Heavy	2	17

Trait Identification Summary

- A wide open area for data mining research
  - A marketers dream
- Clear privacy issues
- Room for creativity & insight for finding traits
- Probably many interesting commercial and research applications
  - Imagine diagnosing back problems via your mobile phone via gait analysis …

LOCATION-BASED APPLICATIONS



Significant Locations

- Significant locations are important locations
  - Usually defined based on frequency with which one person or a population visits a location
- Extract locations where people stay and then cluster them to merge similar points
  - Stay points: points a user has spent more than *ThresTime* in within *ThresDistance* of the point[12]
  - Interesting locations: locations that include stay points from many (>*ThresCount*) people

Significant Location System[12]

- Data collected from 165 users over 2 years
  - 62 users contains 3.5M GPS points
  - *ThresTime* = 20 min and *ThresDistance* = 0.2 KM
    - Allows us to ignore most cases where sitting in traffic

User	# GPS Points	# Stay Points	# Interesting Locations Visited
User 1	910,147	469	9
User 2	860,635	181	8
User 3	753,678	134	13
User 4	188,480	82	4
User 5	89,145	8	1

Significant Location System[12]

- Table below holds top most interesting places
  - Results show that subjects are highly educated
  - Can characterize and group people by the interesting places that they visit

Latitude	Longitude	Frequency	Interesting Locations
40.00	116.327	309	Main Building, Tshingua Univ.
39.976	116.331	122	China Sigma Center, Microsoft China R&D
40.01	116.315	74	Da Yi Tea Culture Center, Tea House
39.975	116.331	58	Cuigong Hotel
39.985	116.32	36	Loongson Technology Service Center

Significant Locations: Assoc. Rules

- Locations visited in a day can represent itemset
  - Mary: {Supermarket, Park, Post Office, School}
  - John: {Supermarket, Park, School, McDonald's}
- Rule: {Supermarket, Park} → {School}

Improving Transportation

- Use location data from many users (crowdsource)
  - Avoid congested roads: Google Navigator
  - Manage traffic dispersion
  - Mine historical data to predict traffic patterns
- Augment road maps with lane information
  - determine lane boundaries
  - deviation of a car → save a life
  - dynamic lane closures: short of cars in a lane → accident or roadwork

Role of Location in Social Networks

- Build Social Communities based on location
  - Proximity
  - Time
  - Frequency
- Google Latitude
  - "See where friends are and what they are up to"
- Facebook "Check-Ins"
  - "Check-In" to a certain location using a cell phone, created by a Facebook user, tag friends
  - See who else is in this location

A Mining Safety Application[1]

- Heavy equipment in mining is dangerous
  - Collisions, open pits, bad visibility
  - Tend to move fast when moving between areas
  - Existing systems use GPS for collision avoidance
    - So lots of GPS data
  - Goal is to use GPS data to improve mine safety
    - Risk assessment & operator guidance
    - Beyond immediate collision warnings
      - Collision avoidance may not be effective if context ignored

A Mining Safety Application (cont.)

- Situational awareness– context matters
- Dependent on location within mine & activity
  - Example: at main excavation site being loaded with copper ore
  - Don't alarm when a vehicle loads or unloads another
- Helps to have knowledge of significant places
  - Care about places where vehicle interactions differ
    - Haulage roads, intersections, loading bays, parking lots
    - Here length of stay not used to determine significant place
    - Once determine type of places can link/fuse on map

A Mining Safety Application (cont.)

- Speed is critical & significant places classified as high or low speed
  - High speed: haulage roads and (high interaction) intersections
  - Low speed: dumping, parking, etc. where vehicles tend to bunch up
- Crowdsourcing since data from all vehicles
  - Know type of vehicle and speeds
    - so have good idea where loading, hauling etc occurs
    - Can identify normal mining functions
    - Can identify normal characteristics (speed, closeness, etc.)

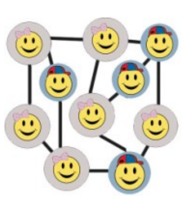
Integration with Other Info/Apps

- Learn more about locations using other info
  - Activity impacts location
    - walk/jog in park
    - drive on roads
    - sleep in hotel/house
  - Demographics impacts location
    - High schools have lots of teenagers
      - May know age from some phone apps
- All of this works in other direction too
  - Location impacts activity, tells us something about those at the site

Some Location-Based Apps

- iMapMy* where * = {Run, Walk, Ride, Hike}
  - tracks route, distance, pace, & more in real-time
  - Share the details of your fitness activities with friends & family, via email, Facebook, or Twitter
  - This data can be mined for exercise-related info
- WHERE helps you discover & share favorite places
  - Recommendation engine <u>learns</u> your preferences and recommends great places
    - Create lists of your favorite places and share with friends

SOCIAL NETWORKING APPLICATIONS



CenceMe Application

- Sensing meets mobile sensor networks[21]
- Classifiers:
  - <u>Audio classifier</u> uses microphone to determine if human voice is present (based on frequency)
  - <u>Conversation classifier</u> uses this info to identify a conversation (human voice must exceed threshold)
    - > 85% accuracy in noisy indoor environments
  - <u>Activity classifier</u> (DT) uses accelerometer and determines sitting, standing, walking, running

CenceMe Application

- Social context classifier derived from multiple sources
  - Neighborhood info: CenceMe buddies around?
  - Social status: uses conversation & activity classifier
    - Can tell if talking to buddies at a restaurant, alone, or at a party
    - Partying and dancing are social status states that use activity and sound volume (volume used to identify parties)
- Mobility mode detector uses GPS to determine if in a vehicle or not (standing, walking, running)
- Location classifier uses GIS info and (shared) user created bindings to map to a icon and location type

CenceMe Application

- Summarize info by using social stereotypes or behavior patterns, calculated daily and viewable
  - Nerdy: based on being alone, lots of time in libraries, and few conversations
  - Party Animal: frequency & duration of parties, level of social interaction
  - Cultured: frequency & duration of visits to museums, theatre
  - Healthy: physically active (walking, jogging, cycling)
  - Greeny: low environmental impact (walk not drive)

CenceMe Application

- Based on user study of 22 people over 3 weeks the things people liked the most:
  - Location information
  - Activity & conversation information
  - Social context
  - Random images
    - When your phone is open the phone takes & posts pics
    - People like it because it forms a daily diary
      - "Oh yeah … that chair … I was in classroom 112 at 2PM"

CenceMe Application

- One survey comment was:
  - "CenceMe made me realize I'm lazier than I thought and encouraged me to exercise a bit more"

CenceMe Results

	Sitting	Standing	Walking	Running
Sitting	0.682	0.282	0.364	0.000
Standing	0.210	0.784	0.006	0.000
Walking	0.003	0.046	0.944	0.008
Running	0.008	0.070	0.177	0.745

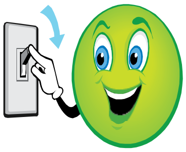
	Conversation	Non-Conversation
Conversation	0.838	0.162
Non-Conversation	0.368*	0.632

Architecture & Design Issues

Resource Issues, Platform Considerations, Client vs. Server Responsibilities, Security & Privacy

RESOURCE ISSUES

Power, RAM & CPU



Sensors are not a Priority

- Example of sensors not being a priority
  - The Android OS tries to preserve battery life
  - Screen hibernation is one key to saving power
    - But screen hibernation puts sensors to sleep![18]
      - Continuous monitoring of sensors was either not considered or viewed as secondary
      - Developers debate whether this is a feature or a bug
      - Work around: CPU "Wake Lock" which prevents hibernation; we compensate by turning screen off
      - We don't think this is the ideal solution (CPU still in normal mode)

Power Consumption

- GPS and GSM localization take lots of power
  - Turn off GPS when not needed/when inside[23]
    - uses cell towers not GPS to determine when go outside
  - Sample at lower rate if acceptable to application
    - But because GPS lock takes time and energy, small reductions in high sampling rates not helpful
      - CenceMe says Nokia takes 120s for lock & active 30s more
    - PeopleTones[17] buddy notification checks every 90 sec.
    - Use adaptive sampling rate (e.g., PeopleTones increases rate when buddy is transitioning from near to far).

Power Consumption

- Uploading data can take significant power
  - Upload via cellular network takes even more if cell phone tower is far away
  - WiFi takes less so if not time-sensitive, send when WiFi available
- Sleep cycles may improve battery life for various applications
  - CenceMe noted little benefit for sleep cycle <10s but longer sleep cycles really hurt the application[21]

Power Consumption Nokia n95[23]

Activity	Power (Watts)
Phone Idle	0.054
Accelerometer Sampling (32 Hz)	0.111
GPS Assisted Lock	0.718
GPS Lock	0.407
GPS Sampling (1 Hz)	0.380
Music Player	0.447
Video Player (Screen on)	0.747
Active Call	0.603
Gaming (Screen On)	1.173
Generating Features & Executing Classifier	0.003
App to Determine Transport Mode	0.425

CenceMe Power Consumption

Activity	Power (Watts)
No CenceMe & Idle	0.08
CenceMe & no user interaction	0.90
Conversation & Social Setting Classifier (rest idle)	0.80
Activity Classifier (rest idle)	0.16

- Results for Nokia N95
- Running full CenceMe suite: 6.22 ± 0.59 hours
  - Not ideal, needs further power optimization

Power Consumption for WISDM[18]

Activity	Power (Watts)
Android	0.001
Sensor Collector	0.043
Lit up Screen	0.525

- Battery Test on HTC EVO with GPS off
- Sensor Collector is WISDM App to collect and store sensor data, but does not apply predictive models to it.
- Sensor collector has minimal impact on battery life, thus it is feasible to continuously collect sensor data.
  - When device on idle, SensorCollector takes 6.6% of power

Memory & CPU Usage Nokia n95[23]

Activity	CPU %	RAM (MB)
Phone Idle	2.18	28.91
Active Call	2.31	30.00
Music Player	30.86	30.26
Video Player	14.63	32.58
Game Playing	97.34	37.52
App to Determine Transport Mode	6.91	29.64

Memory & CPU Usage Nokia n95[21]

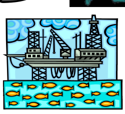
Activity	CPU %	RAM (MB)
Phone Idle	2	34.08
Accel. & Activity Classification	33	34.18
Audio Sampling & Classification	60	34.59
Activity, Audio, & Bluetooth	60	36.10
CenceMe	60	36.90

Resource Issues Summary

- In almost all cases power is much more of a limiting resource than CPU or RAM
- Typical sensor mining apps might drain the battery in 6 or 7 hours
  - This is not really acceptable for apps that are designed to run continuously.
  - We need to work hard to only use power when needed (adaptively)
  - May not be a good solution at this time

MOBILE PLATFORM CONSIDERATIONS





Mobile Platform Considerations

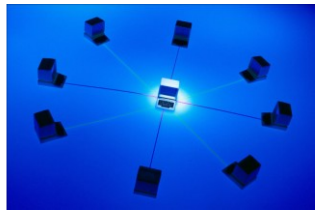
Criterion	Apple iOS	Android	Windows Phone 7
Language	Objective C	Java	Visual Basic
Language Popularity	Low (Difficult)	High	Low
Multiprocessing	No	Yes	Yes
Developer Tools:			
Free	No	Yes	Yes
Documentation	Limited	Extensive	Emerging
Open Source	No	Yes	No
App Approval	Strict Oversight	None	Some Oversight
Market Share	13.80%	14.50%	< 6%
Hardware Venders	Apple	Many	Many

Mobile Operating System Comparison[18]

WISDM Project Experiences

- Adopted Android because easy to program, easy to deploy, free, open, & multi-vendor
- Android was changing quickly when started
  - Big differences between versions
- Many vendors → lots of compatibility testing
  - Found bugs in some versions but not others
- Would Apple let us post our app? Not sure. Android little oversight.
- WEKA data mining suite written in Java

CLIENT VS. SERVER RESPONSIBILITIES



Division of Client and Server Tasks

- Division of labor has tradeoffs
  - More processing on client (phone) means:
    - Application/platform more scalable
    - Increased privacy
    - Bigger drain on power, CPU, & RAM, but not bandwidth
  - More processing on server means:
    - Data captured for future research and other uses
    - Can exploit data not otherwise available (crowdsourcing)

Division of Client and Server Tasks

Client Type:	1/Dumb	2	3	4	5	6/Smart
Data Collection	●	●	●	●	●	●
Data Transformation		●	●	●	●	●
Classification			●	●	●	●
Model Generation				●		●
Data Storage					●	●
Data Reporting					●	●

Division for CenceMe Application[21]

- Backend servers generate higher level "facts" based on phone classifications ("primitives")
  - Audio classifier runs on phone to detect presence of human voice but server executes conversation classifier
  - Higher level facts include social context (meeting, partying, dancing), significant places, & crowdsourcing
- Features generated from raw data on the phone
- Activity classifier trained off line on server but universal model exported to phone (small DT)

SECURITY & PRIVACY

Security and Privacy

- Security policies vary widely
  - Some mobile OS's have strict security policies
    - Symbian requires properly signed keys to remove restrictions on using certain APIs
  - Android has few restrictions
    - My WISDM project has had no problem tapping into sensors and transmitting results
    - Android does notify the user of services that are used
      - SYSTEM PERMISSIONS FOR WISDM SensorCollector
        - » ACCESS_COARSE_LOCATION, ACCESS_FINE_LOCATION
        - » INTERNET, WAKE_LOCK, WRITE_EXTERNAL_STORAGE

Security and Privacy

- Applications that access sensor data can easily spy on you (they do by design)
  - Location data is probably most sensitive
  - A few bad apps could damage the field
  - Note below from http://www.androidspysoftware.com

Security and Privacy

- Even legitimate applications have to be concerned with privacy & security
  - For example, WISDM will encrypt data in transit, include secure accounts with passwords, etc.
  - Need to ensure than any aggregated info is made public only if cannot be traced to individual
- As research study WISDM needs to be careful
  - Do we want others to know where we are 24x7, when we are active, asleep, etc?

Security and Privacy

- What to do?
  - Make it clear what you are monitoring and storing
  - Provide application level control for the user
  - For example, allow the users to turn on/off monitoring of specific sensors and show which ones are on
  - Of course if they use an option to upload the information to Facebook then little privacy!
- Since legitimate and illegitimate apps function alike, no easy way to distinguish them
  - Could try to use only certified apps, but quite limiting

Security & Privacy: iPhone Controversy

- **Why is my iPhone logging my location?**

  The iPhone is not logging your location. Rather, it's maintaining a database of Wi-Fi hotspots and cell towers around your current location, some of which may be located more than one hundred miles away from your iPhone, to help your iPhone rapidly and accurately calculate its location when requested. Calculating a phone's location using just satellite data can take up to several minutes. iPhone can reduce this time to just a few seconds by using Wi-Fi hotspot and cell tower data to quickly find GPS satellites, and even triangulate its location using just Wi-Fi hotspot and cell tower data when GPS is not available (such as indoors or in basements). These calculations are performed live on the iPhone using a crowd-sourced database of Wi-Fi hotspot and cell tower data that is generated by tens of millions of iPhones sending the geo-tagged locations of nearby Wi-Fi hotspots and cell towers in an anonymous and encrypted form to Apple.

Security & Privacy: iPhone Controversy

- **People have identified up to a year's worth of location data being stored on the iPhone. Why does my iPhone need so much data in order to assist it in finding my location today?**

  This data is not the iPhone's location data—it is a subset (cache) of the crowd-sourced Wi-Fi hotspot and cell tower database … to assist the iPhone in rapidly and accurately calculating location. <u>The reason the iPhone stores so much data is a bug we uncovered and plan to fix shortly</u>. We don't think the iPhone needs to store more than seven days of this data.

- **When I turn off Location Services, why does my iPhone sometimes continue updating its Wi-Fi and cell tower data from Apple's crowd-sourced database?**

  It shouldn't. This is a bug, which we plan to fix shortly.

Relevant Resources

- Conferences & Workshops (partial list)
  - International Workshop on Knowledge Discovery from Sensor Data (SensorKDD-11)
  - International Workshop on Mobile Sensor Networks (MSN-11)
  - International Joint Conference on Biometrics (IJCB-11)
  - ACM Conference on Embedded Networked Sensor Systems (SenSys 2011)
  - International PhoneSense Workshop on Sensing Apps. on Mobile Phones

Journals

- International Journal of Wireless Sensor Networks
- International Symposium on Wearable Computers
- International Conference on Pervasive Computing
- Relevant AI and Data Mining Journals

My Contact Information

- Gary Weiss
  - Fordham University, Bronx NY 10458
  - gweiss@cis.fordham.edu
  - http://storm.cis.fordham.edu/~gweiss/
- WISDM Information
  - http://www.cis.fordham.edu/wisdm/
    - WISDM papers available: click “About” then “Publications”
  - Sensorcollector eventually available for collecting sensor data (sensorcollector.com)
  - Actitracker will shortly allow you to log in and track your activities via our Android app (actitracker.com)

Special Thanks To …

- WISDM research group
  - Current Members
    - Anthony Alcaro, Alex Armero, Shaun Gallagher, Andrew Grosner, Margo Flynn, Jeff Lockhart, Paul McHugh, Luigi Paterno, Tony Pulickal, Greg Rivas, Priscilla Twum, Bethany Wolff, Jack Xue
  - Key Former Members
    - Jennifer Kwapisz, Sam Moore, Shane Skowron, Alvan Wong

References

1. Agamennoni, G., Nieto, J., and Nebot, E. 2009. Mining GPS data for extracting significant places, *Proceedings of the 2009 IEEE international conference on Robotics and Automation*.
2. Bao, L. and Intille, S.. 2004. Activity recognition from user-annotated acceleration data, *Lecture Notes Computer Science*, vol. 3001, pp. 1-17.
3. Bourke, A.K., O'Brien, J.V., and Lyons, G.M. 2007. Evaluation of threshold-based tri-axial accelerometer fall detection algorithm, *Gait & Posture* 26(2): 194-99.
4. Bouten, C.V., Koekkoek, K.T., Verduin, M., Kodde, R., and Janssen, J.D. 1997. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity, *IEEE Transactions on Bio-Medical Engineering*, 44(3):136-147.
5. Brezmes, T., Rersa, M., Gorricho, J-L, and Cotrina, J. 2010. Surveillance with Alert Management System using Conventional Cell Phones, *Proceedings of the 5th International Multi-Conference on Computing in the Global Information Technology*, 121-125.
6. Cho, Y., Nam, Y., Choi, Y-J, and Cho, W-D. 2008,.Smart-Buckle: human activity recognition using a 3-axis accelerometer and a wearable camera, *HealthNet*.
7. Discovery channel video about a Smart phone-based biometric system for securing smart phones (based on the research in X16). The relevant portion is about 2/3 thru the video clip which contains two segments. Url: http://watch.discoverychannel.ca/#clip370449

References

8. FitBit. http://www.fitbit.com
9. Frank, J., Mannor, S., and Precup, D. 2010. Activity and gait recognition with time-delay embeddings, *Proceedings of the 24th AAAI Conference on Artificial Intelligence*.
10. Gyorbiro, N., Fabian, A., and Homanyi, G. 2008. An activity recognition system for mobile phones, *Mobile Networks and Applications*, 14 (1), 82-91.
11. Ketabdar, H., and Polzehl., T. 2009. Fall and emergency detection with mobile phones, *Assets '09 Proc. of the 11th International ACM SIGACCESS Conference on Computers and Accessibility* ACM, 241-42.
12. Khetarpaul, S., Chaujan, R., Gupta, S.K., Subramaniam, L.V., and Nambiar, U. 2011. Mining GPS data to determine interesting locations, *Proceedings of the 8th International Workshop on Information Integration on the Web*.
13. Krishnan, N., Colbry, D., Juillard, C., and Panchanathan, S. 2008. Real time human activity recognition using tri-Axial accelerometers, In *Sensors, Signals and Information Processing Workshop*.
14. Krishnan, N., and Panchanathan, S. 2008. Analysis of low resolution accelerometer data for continuous human activity recognition, in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 3337-3340.
15. Kwapisz, J.R., Weiss, G.M., and Moore, S.A. 2010. Activity recognition using cell phone accelerometers, *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data*, 10-18.

References

16. Kwapisz, J.R.,Weiss, G.M., and Moore, S.A. 2010. Cell phone-based biometric identification, *Proceedings of the IEEE Fourth International Conference on Biometrics: Theory, Applications and Systems*.
17. Li, K.A., Sohn, T.Y., Huang, S, and Griswold, W.G. 2008. PeopleTones: A System for the detection and notification of buddy proximity on mobile phones, Proceedings of the 6th International Conference on Mobile Systems.
18. Lockhart, J.W., Weiss, G.M., Xue, J.C., Gallagher, S.T., Grosner, A.B., and Pulickal, T.T. 2011. Design considerations for the WISDM smart phone-based sensor mining architecture, In *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*, San Diego, CA.
19. Maurer, U., Smailagic, A., Siewiorek, D., and Deisher, M. 2006. Activity recognition and monitoring using multiple sensors on different body positions, In *IEEE Proceedings on the International Workshop on Wearable and Implantable Sensor Networks*, 30(5).
20. Menn, J. February 8, 2011. Smartphone shipments surpass PCs. Retrieved from http://www.ft.com/cms/s/2/d96e3bd8-33ca-11e0-b1ed-00144feabdc0.html#axzz1L2wKclC7
21. Miluzzo, E., Lane, N.D., Fodor, K, Peterson, R., Lu, H., Musolesi, M., Eisenman, S.B., Zheng, X., and Campbell, A.T. 2008. Sensing meets mobile social networks: the design, implementation and evlauation of the CenceMe application, Proceedings of the 6th ACM on Embedded Network Sensor Systems, 337-350.

References

22. Patterson, D., Liao, L., Fox, D, and Kautz, H. 2003. Inferring high-level behavior from low-level sensors. *Lecture Notes in Computer Science*, Springer-Verlag, 73-89.
23. Reddy, S. Mun, M. Burke, J. Estrin, D, Hansen, M. and Srivastava, M. 2010. Using mobile phones to determine transportation modes. *ACM Transaction on Sensor Networks*, 6(2).
24. Sposaro, F., and Tyson, G. 2009. iFall: An android application for fall monitoring and response, *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.
25. Tapia, E.M., Intille, S. et al. 2007. Real-Time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor, In *Proc. of the 2007 11th IEEE International Symposium on Wearable Computers*.
26. Weiss, G.M., and Lockhart, J.W. 2011. Identifying user traits by mining smart phone accelerometer data, *Proceedings of the 5th International Workshop on Knowledge Discovery from Sensor Data.*
27. Zhang, T., Wang, J., Liu, P., and Hou, J. 2006. Fall detection by embedding an accelerometer in cellphone and using KFD algorithm, *International Journal of Computer Science and Network Security*, 6(10): 277-284.