Mobile Sensing IV
Activity Recognition

Spring 2015
Petteri Nurmi
Learning Objectives

• What activity recognition is? Why is it interesting?
• What are the main research challenges in (mobile) activity recognition?
• What types of (mobile) activity recognition applications there are?
• How to determine which activities to recognize?
• What is an activity recognizer? How to construct one?
• What aspects should be taken into consideration when evaluating activity recognizers?
What is Activity Recognition?

- Activity recognition is the process of recognizing activities, which can be defined as:
  
  * Activity is defined as human behavior in terms of what is being done when, during a specified period of time.

- One of the most popular subfields of mobile sensing
- Also widely investigated in wearable and “pervasive” sensing (i.e., infrastructure-based) research

Bit of history

• First approaches were based on computer vision
  • Detecting user actions, sign language, etc. from continuous video streams
• First mobile solutions were based on wearable sensors (sensor boxes or worn to clothing)
  • Medical monitoring, assisted living, industrial sector applications (e.g., manufacturing)
• Mobile phones for recent development
  • Mainly motion and transportation routines
• Consumer electronics
  • Microsoft Kinect, Nintendo Wii
  • Philips DirectLife, Nike+
Accelerometer

- Device that measures *proper* acceleration ("g-force")
  - Relative to free fall → device on the surface of table measures acceleration equal to gravity *upwards*
  - Units can be either m/s² or g-force (g)
- Current devices incorporate *triaxial* accelerometers
  - In research literature many examples of 1 and 2 dimensional accelerometers
- Measurements influenced by gravity
  - Needs to be removed if goal is to estimate extent of motion; discussed during later lecture
Gyroscope

- Device that measures rotational motion
  - Smartphones contain (triaxial) MEMS gyroscopes that return angular velocity, i.e., speed of rotation
  - Unit can be rad/s, degrees/s, or revolutions/s (RPS)
- Often integrated on the same chip as accelerometer
  - Cost of using accelerometer + gyroscope smaller than the sum of their individual costs
- Several error sources
  - Constant bias, average output when no rotation applied
  - White noise
  - Flickering noise, temperature noise, calibration noise
Sensors for Pervasive Sensing

- Pervasive sensing refers to situations where sensing is “seamlessly” embedded into the environment
  - Mobile sensing special case of pervasive sensing
- Wide variety of sensors can be used
  - Cameras (e.g., commercial spaces)
  - Power usage sensors, switches (light), proximity / infrared sensors
- Not covered on the course
  - But techniques taught on the course can be used for these sensors as well
Research Challenges

• *Intraclass* variability
  • Same activity performed differently by different individuals

• *Interclass* similarity
  • Classes fundamentally different, but exhibit similar movements (e.g., drinking coffee vs. drinking water)

• *NULL* class
  • Typically only few parts of data stream relevant ➔ irrelevant classes can interfere with detection

• *Class imbalance*
  • Some activities (sitting, walking) occur frequently, whereas others occur very infrequently (e.g., taking a sip)
Research Challenges

• Cross-device generalization
  • Sensor and measurement characteristics contain significant variation across devices
  • Varying sensor placement (and device orientation)
    • trouser pocket, hand, jacket pocket, backpack, handbag
• Sampling rate variations
  • Interdevice variations: different devices have different sampling rates for sensors
  • Intradevice variations: same device may adapt sampling rates during the execution of application
• Activities can be concurrent and interleaving
  • Multiple activities occurring at the same time
Research Challenges: Device Orientation

• Devices can be in different orientations, extracted features should be insensitive to device orientation

• Rotation invariance
  • Function whose value does not change when arbitrary rotation is applied on the measurements

• Potential solution:
  • Apply rotation invariant norm on the signals
    – L2 norm, i.e., Euclidean distance $\sum_i \sqrt{s_i^2}$
  • Extract features from the normed signal

• Alternative is to estimate device orientation and transform signals to global reference system
Mobile phones can have different placements

- Urban mobility: trouser pockets, jacket pockets, bags (handbag or back bag), and the user’s hand
- Social situations: bag, pocket, on a surface
- Home situations: lying somewhere
  - Particularly at home and workplaces, devices not necessarily within arms length

Optimally would need *placement insensitive* features

- In practice next to impossible design
- Instead, collect data from dominant placements, and assess performance against varying the placement
Research Challenges: Activity Spotting

• Activity spotting refers to identifying when a relevant activity takes place
  • Requires *segmentation* of the sensor signals
    – Detecting different “levels” in the data
  • Segmentation performed by identifying changes, or change events in the data
    – Active/inactive detection
    – Detecting specific movements, e.g., rotation of wrist, etc.
  • Once data segmented, compares data within each segment against activity categories
    – If sufficient match, assume relevant activity takes place
Activity spotting can also help to validate ground truth labels.
Application Categories for AR: Sport and Behaviour Tracking

- Fitness trackers popular example of AR
  - Recognizing overall level of physical activity
  - Recognizing type of activity
- Has been predominantly carried out using wearable sensors, but increasingly on smartphones
- In fitness tracking, target is to estimate metabolic consumption accurately
- Alternatively more generic movement tracking, estimating where people travel from/to and how
- Mainly considers accelerometers and physiological sensors, in particular heart rate
  - Some use for barometers (altitude)
  - Location sensors also used often, GPS, WiFi, GSM
Application Categories for AR: Care Monitoring

- Active research topic, particularly for elder care and people with cognitive disabilities (e.g., dementia)
  - Automatically assess the care needs of people
  - Monitor regularity of behaviour and alert care personnel if irregularities observed (e.g., forgot medication)
- Most of the work contains a mix of infrastructure sensors and wearable sensors
  - “Smart homes”: microphones & cameras (seldom for privacy reasons), electric switches, etc.
  - Mobile devices: mainly accelerometer and microphone
Application Categories for AR: Gesture Recognition

- Refers to the detection of hand or body motions that are intend to communicate information
  - Command gestures: e.g., pinch zooming, multifinger scrolling etc.
- Overall very large topic area, in terms of mobile sensing mainly used for following:
  - Motion control gestures (Wii or mobile phone), e.g., for home electronics
  - Airwriting
  - Authentication
- Main challenge is recreating gesture trajectory reliably from sensor information
Bao & Intille (2004): Activity Recognition from User-Annotated Acceleration Data

- One of the first (and most cited) AR papers
  - 5 biaxial accelerometers attached to different locations
    - upper left arm, right wrist, right hip, upper left leg, lower right leg
  - 20 different activities
- Features
  - 6.7s windows, 50% overlap (512 samples 76.25Hz)
  - DC, energy, mean, frequency-domain entropy, correlation
- Classifiers:
  - Nearest Neighbour, C4.5 (decision tree), Naive Bayes
  - Best results with C4.5: 71.6% in user-specific and 84.3% in leave one-user-out testing

- Relies on an external wearable sensing device that connected via Bluetooth
  - Accelerometer, digital compass, and barometer
  - Attached to the hip
- Supports 8 activities
  - Sitting, standing, walking, walking up/down stairs, riding elevator up/down, brushing teeth
- Originally extract 651 features, boosting-based selection, top 50 used for final classifier
- Classification:
  - Boosted decision stumps: around 80% overall accuracy
  - Boosted decision stumps + HMM: around 85%
Hemminki, Nurmi, Tarkoma (2013): Accelerometer-Based Transportation Mode Detection on Smartphones

- Distinguishing 8 different transportation modalities using smartphone accelerometers: stationary, walk, standing, car, bus, train, metro, tram
- Preprocessing: Gravity elimination (lecture V), transformation to horizontal and vertical planes
- Features: statistical (11), time domain (4), frequency domain (12), and a set of so-called peak and segment features
- Classifiers: Hierarchical classification framework, boosting-based decision tree + Hidden Markov Model
- Results: Over 80% recognition accuracy, train and metro lowest (65%) mainly because they are confused with each other
Which activities should I select?

- In many application domains, existing standardized categorizations can be used as basis
  - Health care: activities of daily living and instrumented activities of daily living
  - Sports: physical activity categorizations, e.g., compendium of physical activities
  - Behaviour monitoring: time use surveys
  - Goal is to cover “most important” ones, i.e., so that the monitoring captures the activities most of the time
Activities of Daily Living

- Activities of daily living (ADL)
  - Basic tasks of everyday life: eating, washing, moving
- Instrumental activities of daily living (IADL)
  - More complex tasks of everyday life that require interactions with some instrument: managing finances, preparing meals, shopping, managing medications
- Particularly the recognition of ADLs has received much attention in activity recognition literature
- Important application for assisted living, including people with disabilities or diseases and elder care
  - Assistance needs related to capability of carrying out ADLs (and related tasks) independently
Activities of Daily Living

• WHO activity categorization
  • Intended for providing a standardized categorization for identifying and classification disabilities

• 9 primary ADL categories
  1. Learning and applying domain knowledge
  2. General tasks and demands
  3. Communication
  4. Mobility
  5. Self care
  6. Domestic care
  7. Interpersonal interactions and relationships
  8. Major life areas
  9. Community, social, and civic life
Compendium of Physical Activities

- Defines 21 high level categories of physical activities, and several subcategories
- For each category gives a so-called MET rate (metabolic intensity), which can be used to estimate overall level of physical activity for a user

<table>
<thead>
<tr>
<th>Category</th>
<th>Activity Type</th>
<th>Subcategory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycling</td>
<td>Conditioning Exercise</td>
<td>Dancing</td>
</tr>
<tr>
<td>Fishing &amp; Hunting</td>
<td>Home Activities</td>
<td>Home Repair</td>
</tr>
<tr>
<td>Inactivity</td>
<td>Lawn &amp; Garden</td>
<td>Miscellaneous</td>
</tr>
<tr>
<td>Music Playing</td>
<td>Occupation</td>
<td>Running</td>
</tr>
<tr>
<td>Self care</td>
<td>Sexual Activity</td>
<td>Sports</td>
</tr>
<tr>
<td>Transportation</td>
<td>Walking</td>
<td>Water Activities</td>
</tr>
<tr>
<td>Winter Activities</td>
<td>Religious Activities</td>
<td>Volunteer Activities</td>
</tr>
</tbody>
</table>
Time Use Surveys

- National statistical institutes carry out so-called time use surveys that measure population time use
  - Diary-based survey, usually few days per participant
  - Measures activity type & time spent on activity
    - 5 primary categories, and several subcategories
- Example from Finnish statistical institute
  - Employment (working & commuting)
  - Personal care and needs (cleaning, sleeping, washing)
  - Studying
  - Leisure activities (11 categories)
  - Miscellaneous / Non-classified

Activity Recognizer

• A system that implements the sensing pipeline for the purpose of recognizing some activity (-ties)
  • Thus, a chain containing a set of preprocessing, feature extraction, and inference techniques

• Some differences between the design and deployment phase
  • Design: Investigate large sets of features and classifiers to find best possible
  • Deployment: Include ONLY the “best” features and classifier(s) as given by evaluation

• But how to detect which ones are best?
Feature and Sensor Selection

- **Feature selection** refers to the process of selecting a subset of relevant features
  - Basic idea to estimate how related a feature is with the “target” variable (i.e., class that is being sensed)
  - E.g., pointwise mutual information and correlation
- **Things to consider**
  - As mentioned, activity recognition data imbalanced
    - Many simple feature selection techniques poor under class imbalance
  - Features come potentially from different sensors, including a second feature from a sensor cheaper than adding a new sensor all together
Minimum Redundancy and Maximum Relevance

- Basic idea:
  - Minimize redundancy $V$
    - $I(i,j)$ information gain between features $i$ and $j$
  - Maximize relevance $W$
    - $h$ is target class/category
- For continuous variables:
  - F-statistic used for relevance
  - Correlation used for redundancy
- Different ways to combine redundancy and relevance into a single measure
  - MID: Mutual Information difference
  - FCD: F-test correlation difference

\[
\text{mRMR} = \max_S \left[ \frac{1}{|S|} \sum_{f_i \in S} I(f_i; c) - \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j) \right].
\]
Boosting and AdaBoost

- Machine learning technique that combines several weak classifiers into a single strong one
- AdaBoost
  - One of the most popular boosting techniques
  - Each sample assigned a weight that determines its importance
  - Over a series of rounds T, learns (weak) classifiers that minimize classification error on the weighted sample
  - After each round, re-evaluates weights of samples and trains the next classifier
- Note: typically each classifier uses 1-2 features ➔ automatic support for feature selection
AdaBoost cont.

1. Initialization:
   • Initialize weight of each sample \( w_j = 1 / n \)
   • Initialize weight of each weak learner \( \alpha_i = 1 / T \)

2. For \( i = 1 \ldots T \) (T number of weak learners)
   • \( h_i \) = base learner learned using weights \( w_j \)
   • Test learner \( i \) on all data
   • Calculate error of learner on weighted data: \( \varepsilon_i = \sum w_j E_j \)
     - \( E_j = 1 \) if the sample misclassified, otherwise \( E_j = 0 \)
   • \( \alpha_i = 0.5 \times \log((1 - \varepsilon_i) / \varepsilon_i) \)
   • Loss = \( \exp(-\alpha_i y_j h_i(x_j)) \)
   • Set new weights \( w_j = w_j \times \text{Loss} / \sum w_j \)
Evaluating Activity Recognizers: Best Practices

• Thus far we have discussed metrics for evaluation
  • Classification: precision, accuracy, recall, F1-score
  • Event-based: fragmentation, overfill, underfill
  • Energy: power consumption or impact on battery life
• But how to design experiments where to calculate?
  • In practice should test against as many research challenges as possible
  • User-dependent and user independent testing
  • Testing against different placements, window sizes, classification algorithms, etc.
Evaluating Activity Recognizers: Best Practices

• Labelled training data should be collected so that all different aspects can be evaluated
  • Several placements at the same time
    – Several devices / placement to get out hardware variations
  • Several sensor modalities at a time
  • User variability requires collecting data from several different types of users
  • Temporal variability: several datasets at different times
  • Controlled vs. real-world: placements require control, everyday usage unconstrained ➔ need test both

• Variety of other factors
  • E.g., in Finland, weather always a significant factor
  • Radio signal environments, public transportation etc.
Evaluating Activity Recognizers: Testing & Cross-Validation

- In practice, testing is carried out using a procedure called *cross-validation*
  - Data split into k-folds, k-1 folds used for training a model and one fold used for testing
  - Process repeated so that each fold is once serving as the test set, results obtained as averages over the folds
  - *Stratified cross-validation*: folds determined so that the distribution of data in them reflects original distribution
- Why is this awesome?
  - Person-dependent testing: k-fold CV on one user’s data
  - Person-independent testing: leave-one-user-out CV
  - Placement testing: leave-one-placement-out CV
Example: Accelerometer-based transportation mode detection

- Two scenario based data collections
  - Scenarios follow a predefined route, different each time to avoid overfitting to characteristics of route
  - Scenario I: 9 participants
  - Scenario II: 7 participants
  - 6 devices / participants, 2 phones in 3 placements
- Supplemented with everyday data
  - Collected from Helsinki, Germany, Japan
  - Provides insights into generalization performance
Example: Accelerometer-based transportation mode detection

<table>
<thead>
<tr>
<th>TMode</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationary</td>
<td>95.1 (2.0)</td>
<td>72.4 (2.1)</td>
</tr>
<tr>
<td>Walk</td>
<td>92.7 (2.4)</td>
<td>92.4 (2.0)</td>
</tr>
<tr>
<td>Bus</td>
<td>84.8 (3.6)</td>
<td>79.6 (0.5)</td>
</tr>
<tr>
<td>Train</td>
<td>74.7 (4.0)</td>
<td>80.8 (1.9)</td>
</tr>
<tr>
<td>Metro</td>
<td>69.6 (3.8)</td>
<td>81.3 (1.7)</td>
</tr>
<tr>
<td>Tram</td>
<td>88.4 (2.5)</td>
<td>87.9 (1.6)</td>
</tr>
<tr>
<td>Mean</td>
<td>84.2 (3.0)</td>
<td>82.4 (1.6)</td>
</tr>
</tbody>
</table>

Evaluation:

1. Overall accuracy
   - K-fold cross-validation for user-dependent
   - Leave-one-user out for user-independent

2. Influence of device placement
   - Leave-one-placement out cross-validation

3. Generalization
   - Train with scenario, test with everyday data

4. Robustness
   - Event-based metrics
Advanced Topics: Activity Discovery

• Refers to the identification of activities from sensor signal streams
  • Identifying how many classes there are in the data
  • Effectively a clustering problem
    • Extract features from data and cluster them. Use a heuristic for selecting optimal number of clusters
    • Signal “factorization”: transform signals and perform clustering on the transformed signals
      – Clustering the coefficients of the factorized signals
  • Can be combined with active learning
    • Identify most informative points given the clustering and prompt for labels from the user
Advanced Topics: Community Similarity Networks

- Looks at how classifier training can utilize data collected from other users
- Community similarity network (CSN)
  - Graph where vertices are users, and edges are *similarities* between them
  - Egocentric graphs, i.e., each user has separate CSN
  - Similarity looks at how similar a person is to another person, in terms of patterns, not in terms of raw sensors
- Similarity-aware boosting
  - Similar idea as in Adaboost, but weights of data samples determined using similarities
Emerging topic, looks at how the feature extraction phase can be removed by transforming the signals.

- Signal represented as superposition of basis vectors
- Activation of basis vector can be used as feature instead of explicit feature extraction phase
- Examples of techniques: PCA, Sparse Coding
Summary

- Activity recognition: detection of human actions
  - One of the most important categories for sensing apps
- Several research challenges:
  - Data variability, particularly across users, placements, sensing equipment, and over time, class imbalance, “NULL” class, and annotation of signals
  - Selection of activities driven by categorizations done by statistical institutes / governmental organizations
  - Activity recognizer = implementation of the sensing pipeline for recognizing activities
    - Need for feature, sensor, and classifier selection
    - Complex evaluation procedures, need to isolate and test against all main sources of variability
References


• Junker, H.; Amft, O.; Lukowicz, P. & Tröster, G., Gesture spotting with body-word inertial sensors to detect user activities, Pattern Recognition, 2008, 41, 2010-2024

• Lester, J.; Choudhury, T. & Borriello, G., A Practical Approach to Recognizing Physical Activities, Proceedings of the 4th International Conference on Pervasive Computing (PERVASIVE), Springer - Verlag, 2006, 3968, 1 - 16


