Network Community Behavior to Infer Human Activities

Lane, Nicholas D.
Pengyu, Li
Zhou, Lin
Zhao, Feng

Mohammad Alodadi
IS809
• Activity recognition
  – Unreliable in noisy environment
  – Relies on the user’s personal information
  – Isolated individual – models solely on the sensor data collected at personal scale

• Socially connected individuals have correlated activity patterns
  – Two people with strong social connection
    • Share similar behaviour
    • Food selection, transport, sleep pattern

• Community behaviour
• Opportunity to improve the recognition of behaviour/activity
  – Improve the inference
• Capture the collective behaviour of the people in community
• Network community behaviour framework
  – Considers personal (acc, mic)
  – and community scale information
  – Improve accuracy
Activity Recognition

• Conventional approach
  – At personal scale - Individual Sensing
  – Gather sensor data and classify
  – Assumes people operate in social vacuum/isolated

• Networked Community Behavior (NCB)
  – Community Sensing
  – Personal as well as community scale data
  – Identify communities with co-related behaviour
  – Improve accuracy by using hints
Behaviour Modelling

Should We Be Ignoring the Behavior of the Wider User Population?

Sleep  Eating  Places  Mood
Community Behaviour in Daily Life

- Community behaviour
  - Correlation among behavioural patterns are strongest
    - With strong social tie
  - Sleep pattern, transportation habit
Community behaviour

• Key idea: Within community, people respond similarly
  – Given similar context and situation
  – Adopt the behaviour of others
  – Both ways
• How group of (similar) people commute
  – Similar weather, traffic condition
• People exchange information
  – Change their previous behaviour
  – Homophily
Explore dataset

• Mobile sensing dataset
• Captures
  – sleep duration (4 duration categories)
  – and transportation mode (car, bus, subway, walk)
• Tests strongly connected people tend to have more correlated pattern
• Social links----participant co-location and trajectory similarity
• Compute correlation – Jaccard coefficient
Personal Sensing Linked with Community Behaviour

Car, bus, subway, walk
21 days, 27 participant

51 participant, 90 days

Exploit the additional community scale information to recognize activities
Specially when individual sensor data is noisy
Observations from socially connected individual can provide signal

Bottom 20%

Figure 1: Pairs of people with strong social links have correlated behavior in diverse activities, such as, transportation and sleep.

Co-location and trajectory similarity
Networked Community Behaviour Framework

- Personal Sensing as well as community-scale sensing
- Robust
- Two key phases
  - Hierarchical Network – connects personal and community
  - Networked Community Behaviour Learning
    - Trains classifiers and Performs collective inference
Figure 2: Networked Community Behavior Framework
Hierarchical network

Community scale
- Time-series Location Estimates
- Social Network Mining
- Community Discovery
- Community Feature Extraction

Personal scale
- Activity-related Sensor Data
- Framing and Feature Extraction
- Activity Node Representation

Common trajectories and locations

Figure 2: Networked Community Behavior Framework
Collective inference

**Community scale**
- Time-series Location Estimates
- Activity-related Sensor Data
- Social Network Mining
- Community Discovery
- Community Feature Extraction

**Personal scale**
- Activity-related Sensor Data
- Framing and Feature Extraction
- Activity Node Representation

Two types of classifier
Make weak decisions

**Figure 2:** Networked Community Behavior Framework
Hierarchical NCB Network

Top layer: Community scale relationship

Bottom layer: Personal Sensing Layer

- Node represents activities + feature (attribute)
- Links two layers
- Attribute: sensor data + ground truth + soft decision

Figure 3: Hierarchical NCB Network
Hierarchical NCB Network

• Personal Sensing Layer=>conventional approach
  - Bottom Layer
  - Sensor data framed and features are extracted
    - Acc, audio, gps
  - Activity Node Representation
    - Each activity – one node
    - How to represent activities?
    - Sleep/mood
    - Frequent events (walk), count

Figure 3: Hierarchical NCB Network
Hierarchical NCB Network

- Community Behaviour Layer
- Three stage process
  - Mining Social Network Graph
    - Identify social ties
  - Discover Communities
    - Newman Girvan Algorithm
    - Edge-betweenness centrality
  - Community Feature Extraction

<table>
<thead>
<tr>
<th>Category</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity &amp; Duration</td>
<td>{NumColoc, NumLoc} × {All, Evening, Weekend} NumHours, NumWeekdays, BoundingBoxArea</td>
</tr>
<tr>
<td>Location Diversity</td>
<td>{Avg, Med, Var, Min, Max} × {Entropy, Freq} {Avg, Med, Var, Min, Max} × UserCount</td>
</tr>
<tr>
<td>Mobility Regularity</td>
<td>SchEntropy × {LH, LD, LHD} SchSize × {LH, LD, LHD}</td>
</tr>
<tr>
<td>Specificity</td>
<td>{Min, Avg, Max} × TFIDF, PerObvTogether</td>
</tr>
<tr>
<td>Structural Properties</td>
<td>NumMutualNeighbors {Neighborhood, Location} × Overlap</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centrality</td>
<td>Betweenness, Eigenvector, Closeness, Degree [1]</td>
</tr>
<tr>
<td>Other</td>
<td>Average Network Clustering Coefficient [10]</td>
</tr>
</tbody>
</table>

Table 2: Community Features.
Mining social network

• Social ties
  – Co-location, user trajectories
  – GPS, Wifi data

• What is co-location?
  – 30 meters-10 mints

• Compute weight
Community Identification Algorithms

- Hierarchical
- Girvan-Newman
- Radicchi et al.
- Spectral Bisection
- Clique percolation
- Louvain

How to measure the quality of a community partition?
Modularity

- **Modularity** measures the group interactions compared with the expected random connections in the group.

- In a network with m edges, for two nodes with degree $d_i$ and $d_j$, expected random connections between them are $d_i d_j / 2m$.

- The interaction utility in a group:
  $$\sum_{i \in C, j \in C} A_{ij} - d_i d_j / 2m$$

- Modularity:
  $$\frac{1}{2m} \sum_{C} \sum_{i \in C, j \in C} A_{ij} - d_i d_j / 2m$$

Expected Number of edges between 6 and 9 is $5*3/(2*17) = 15/34$
Community Detection Algorithm

• Divisive Method: Newman-Girvan
  – Calculate the edge betweenness for all edges in the network.
  – Remove the edge with the highest betweenness.
  – Recalculate betweennesses for all edges affected by the removal.
  – Repeat until no edges remain.
Community Detection Algorithm

- Divisive Method: Newman-Girvan
  - *Edge betweenness*
    - $\text{betweenness}(e_{ij}) = \text{number of times} e_{ij} \text{ appears in all shortest paths}$
    - High betweenness edges are more "central"
In Execution

0 cuts

100 cuts

120 cuts

500 cuts
Community Detection

Illustration of community detection using dendrogram

Modularity maximization

Modularity

0.3

0.41

0.35
Networked Community Behaviour Learning

Two step process
• S-1: NCB trains two classifiers using personal and community scale information
  - Expose soft decisions
• S-2 Collective inference using Relaxation-Labeling
  - Uses network weight and classifier output => final inference

• Personal Sensing Classifier
  - Data observed from individual users (personal sensing layer)
  - Classifier trained with a different dataset – sensor data features are labelled by hand with ground truth
  - Classification based on the (unlabelled) data gathered from personal sensing Layer
  - Applied to new activity node – produces a vector of soft decision
    - Confidence value
Networked Community Behaviour Learning

- Community Behaviour Classifier
  - Makes soft inference for activity nodes
- Classification is done based on aggregated feature
  - Operates on combination
    - User node attribute
    - Soft decision vectors from the activity nodes of the network neighbourhood

**Figure 4: Community Behavior Classifier**
Network neighbourhood

- Dictates which activity and user nodes will generate feature for Community Behaviour Classifier
- Determined by the user node
  - Who generates the data
- Neighbourhood: All adjacent nodes + active nodes

Figure 4: Community Behavior Classifier
Networked Community Behaviour Learning

- Community Behaviour Classifier
  - Classification is done based on aggregated feature
  - Network neighbourhood
  - Aggregated neighbourhood feature vector
    - Community features (structural metrics)
    - Soft-decision vectors

\[ aV_j = \sum_{i \in \mathcal{N}} V_i i c_i e^{-k t_{i,j}} w_{p,q} c_f k \]

- Classification-user node=>activity classes of user’s activity nodes
  - Based on strong social ties
  - Belong to communities that have correlated behaviour between members for this activity
NCB Framework

Community scale  Personal scale

- Time-series Location Estimates
- Activity-related Sensor Data

- Social Network Mining
- Framing and Feature Extraction
- Community Discovery
- Activity Node Representation

- Community Behavior Collective Inference
  - Personal Sensing Classifier
  - Community Behavior Classifier

- Hierarchical Network
  - User Nodes
  - Activity Nodes

**Figure 2:** Networked Community Behavior Framework
Community Behaviour Collective Inference

- Relaxation-Labeling

**Algorithm 1: Relaxation-Labeling under NCB**

1. For $\forall$ unlabeled $a_j \in$ Hierarchical NCB Network
2. $a_j . \text{per}V \leftarrow$ Personal Sensing Classifier ($a_j$)
3. End

4. While $ca_{cv} \geq ca_{th}$ and iterations below threshold $\leq it_{th}$
5. For $\forall$ unlabeled $a_j \in$ Hierarchical NCB Network
6. $\text{temp}V \leftarrow$ Community Behavior Classifier ($a_j$)
7. $a_j . \text{com}V \leftarrow \beta \text{temp}V + (1 - \beta)a_j . \text{per}V$
8. End
9. End

10. $a_j . \text{inference}V \leftarrow a_j . \text{com}V$
Evaluation

• Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Activity Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transportation</td>
<td>{ bus, subway, car, walk, stationary}</td>
</tr>
<tr>
<td>Sleep</td>
<td>{ 9+ hours, 9 to 6 hours, 6 to 3 hours, &lt;3 hours }</td>
</tr>
<tr>
<td>Diet</td>
<td>{ fast food, cafeteria &amp; food court western restaurant, ethnic restaurant}</td>
</tr>
</tbody>
</table>
| Mood      | { PA – above median pleasure/activeness  
pA – below median pleasure/activeness  
PA – above median pleasure/activeness  
pA – below median pleasure/activeness } |

Table 3: Dataset Activity Classes.

• Benchmark – classifier
  – Single-classifier
Datasets. Table 3 lists all activity classes contained in the four datasets used in our evaluation. Our first dataset, *Transportation* comprises 51 people who collect GPS traces of their transportation mode behavior for three months. Data is collected not only with smartphones but using other personally worn GPS-enabled devices (e.g., PDAs, personal navigation devices). Users self-report ground-truth transportation modes. Next, *Mood* contains 25 people all of whom provide phone usage data (e.g., SMS, email, phone calls, application usage, web browsing, and location). Participants provide usage data for two months; in addition they complete a brief mood survey instrument (based on the Circumplex model [29]) multiple times per day. Under the Circumplex model, user moods are represented as a pair of values corresponding to two dimensions of mood (pleasure and activity). As noted in Table 3, we use four discrete class of mood adopting the methodology of [21]. Similarly, *Sleep* contains participant provided sleep duration surveys for 27 people over 21 days. Subjects carry Nexus One smartphones which collect microphone, accelerometer and again phone usage data (e.g., time spent recharging). We source *Transportation*, *Mood* and *Sleep* datasets externally, from the authors of [34],[21] and [19] respectively. The remaining dataset *Diet* we collect ourselves. We provide 20 people with smartphones for 1 month, each phone collects GPS or WiFi data (enabling location estimates). Participants work for the same company and agree to either complete a food diary on their phone or take photos of their meals. Meals that occur at home are ignored. We determine coarse ground-truth of daily meals by coding meals from the food diary or photos into 4 meal
Evaluation

• Classifier performance comparison

Both uses same underlying feature, training data, classifier design. Gain is due to crowdsourcing.
Evaluation

• Classifier performance comparison
  • CDF of per person average classification accuracy
  • X-axis=>fraction of users having accuracy < p
  • Curves shifted bottom right
    • Large fraction of population
      Experiences high accuracy

Figure 6: CDF of per-person classification accuracy for all datasets. NCB provides more uniform classifier accuracy for the entire user population than the benchmark classifiers.
Evaluation

- Performance Gains
- Based on correlative pattern

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Social Network Link Strength</th>
<th>NCB Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Transportation</td>
<td>0.20</td>
<td>0.51</td>
</tr>
<tr>
<td>Mood</td>
<td>0.30</td>
<td>0.51</td>
</tr>
<tr>
<td>Sleep</td>
<td>0.48</td>
<td>0.71</td>
</tr>
<tr>
<td>Diet</td>
<td>0.08</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 4: Population correlation level present in all datasets.

Figure 7: NCB increases classification accuracy for individuals with higher levels of correlated behavior within the social network.
Evaluation

• Performance Gains

NCB performs better in the noisy env.

Figure 8: When single-user classifier struggles, NCB provides its largest performance gains. Here we show representative relationships between the performance gain in using NCB compared with the accuracy of the benchmark classifiers.
Critique

• Positive points:
  – Relying on community behavior in order to enhance activity recognition was innovative.
  – The model can overcome outliers.

• Negative points:
  – Social ties are different depending on the application.
  – Temporal data are not used in the classifiers.
  – Should have involved social theories in social ties definition such as social correlation, and social balance theory.