A Heterogeneous Clustering Approach For Human Activity Recognition

Sabin Kafle
Overview

• Problem statement and Motivation
• Preliminaries
  • Gait events and EMG signal
  • Dynamic Linear Model (DLM)
  • Hierarchical Normal Model (HNM)
  • Generalized Dirichlet Distribution
• Our Approach
• Experiments and results
• Conclusion and Future Work
Motivation

• Availability of sensors in everyday life has enabled capture of contextual information from Human Behavior in real-time.

• This has lead to the significant research interest in Human Activity Recognition.

• Despite this interest, two significant aspect of such analysis have been overlooked
  • Performance variation across different sensors
  • Availability of labeled data in large scale
Motivation

• Sensor data is also used in development of prosthetic hands and limbs with automated control
  • Neuromuscular signals captured by EMG sensors are used to understand the intent of user
  • The recognition of intent enables such prosthetic devices to provide automated control in certain direction; e.g. lifting of limb while walking up the stairs in case of prosthetic limbs.
• While significant progress has been made for intent recognition, unsupervised learning of such intent has been distinctly lacking.
Example of an HAR system using sensors

Problem Statement

• The known heterogeneity in the data source limits the applicability of current approaches to the data set.

• The unreliability of label sensors along with the general lack of prior knowledge about the distinct number of classes in most sensor data leads us to base our work on nonparametric approach.

• We also need an approach which is able to incorporate many data preprocessing steps applied to supervised learning.

• We develop, implement and test a new Bayesian Semi-parametric approach for clustering heterogeneous HAR data.
  • We focus on a specific sensor (EMG) and specific activity (terrain detection) for application of our approach.
EMG data sample
Gait Cycle

• Gait cycle starts from toe-off phase and ends with toe-off of the same leg.
• Two states are significant in a gait cycle, Toe Off (TO) and Heel Strike (HS)
Muscles significant for Terrain Detection
Features extracted from EMG signals

• Time-domain features are usually extracted from EMG signals due to their less computational cost and efficient classification ability. Some of the popular features used for gait event detection from EMG signals include (Gupta et al. 2014):
  • Mean
  • Variance
  • Mean and variance trend
  • Windowed mean and variance difference
  • Maximum Amplitude
  • Root mean square

• Each of these features are extracted over a certain time-range or certain gait phase stages, thus reducing the dimension of time series itself and making online computation feasible.
Our Approach

• Our approach is based on combination of Hierarchical Normal model, Sampling Space model and Beta two-parameter truncated process with Generalized Dirichlet Distribution as random probability measure.
  • Break input signal into latent variables using Sampling space model
  • Use hierarchical normal model to introduce hierarchies present in the data
    • Perform Clustering based on subset of parameters from sampling space model
    • Use Beta two-parameter truncated process for clustering parameters
• We use a Gibbs Sampler with Metropolis Hastings sub-chain to obtain posterior samples
Sampling Model (Nieto-Barajas et al. (2014))

• We use a sampling model based on DLMs to represent time series.
  \[ y_i = Z \alpha_i + X \beta_i + \theta_i + \epsilon_i \]

• Z and X are design matrices of dimension T x p and T x d respectively

• p x 1 dimensional \( \alpha_i \), d x 1 dimensional \( \beta_i \) and T x 1 dimensional \( \theta_i \) are vectors representing the parameter of the model.

• We use \( \theta_i \) and \( \beta_i \) for clustering the time series, while \( \alpha_i \) is used for improving the fit of the model.

• The usage of selective parameters for clustering enables us to use only the relevant component of time series for clustering.
Hierarchical Normal Model

• We use Hierarchical Normal Models to introduce heterogeneity in our work.

• HMNs are popularly used in human and biological sciences for data that has natural hierarchy.

• Each level represents a level in hierarchy (sensor, person etc.)
Nonparametric Hierarchical Models

- We cast the clustering problems in terms of random variables.
- For a Gibbs sampler, we iteratively draw values from $\pi(p, Z|K, X), \pi(K|p, Z, X)$

\[
\begin{align*}
(X_i|Z, K) &\sim \pi(X_i|Z_{K_i}), \\
(K_i|p) &\sim \sum_{k=1}^{N} p_k \delta_k (.), \\
(p, Z) &\sim \pi(p)\pi(Z)
\end{align*}
\]

\[K = (K_1, ..., K_n), Y_i = Z_{K_i}\]
• For each cluster, the hierarchy is represented as:
Feature Inclusion

• We extend the Sampling Space model to include multiple time-series as set of independent processes
  • This is required to incorporate multiple features of EMG signal into our approach

• The Generalized Dirichlet Distribution parameter \( p \) and \( K \) which is responsible for sampling the cluster membership can be updated in two possible ways:
  • Update \( K \)s with each feature series being considered independent of one another
  • Maintain multiple \( p \)’s and \( K \)’s and use majority voting from feature series to determine cluster membership
Data

• We have a large collection of EMG sensor data collected while walking on different terrains [walking on Level Ground (LG), Ramp Ascent (RA), Ramp Descent (RD), Stair Ascent (SA) and Stair Descent (SD)]
  • Sensors are placed on different limb muscles to measure EMG signals specific to that muscle
  • Each person is subject to the same set of experimental procedure regardless of difference in gait (locomotion) length.
• Earlier research has suggested that EMG recordings are specific to the muscle and person the sensor is placed on. (Reaz et al. 2006)
• Given the EMG recordings, we want to group data collected on same (similar) terrain into a single cluster.
Experiments

- We have EMG signal dataset of
  - 7 EMG sensors applied to different muscles during gait events
  - 12 normal subject
  - Total of 9450 records of gait cycles (on hierarchy of sensor and person)

- We consider a hierarchical model with specification from Sensor to Person data.

- To make each time series of equal length
  - only 0.5s of data both before and after HC and TOs are considered (considered to be most significant phase for gait event detection)

- Each feature series is of length 23 to make the experiments computationally viable and increase relevance of features.

- We also check the viability of ensemble approach and combination approach for feature series

- Because of unbalanced data size for different class, we use **subsampling** to test our approach

- We use heterogeneity score to compare the cluster quality.

- We let the sampler burn-in for 1000 iterations and then sample a cluster every 200 iterations
Results

• For 5 clusters with hierarchy and a single (mean) feature
  • Percentage of data of each class in different clusters

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level Ground</td>
<td>5.6</td>
<td>76.3</td>
<td>9.17</td>
<td>5.5</td>
<td>3.43</td>
</tr>
<tr>
<td>Ramp Ascent</td>
<td>4.5</td>
<td>80.1</td>
<td>6.7</td>
<td>5.2</td>
<td>3.3</td>
</tr>
<tr>
<td>Ramp Descent</td>
<td>6.0</td>
<td>76.8</td>
<td>9.1</td>
<td>5.2</td>
<td>2.9</td>
</tr>
<tr>
<td>Stair Ascent</td>
<td>5.4</td>
<td>77.7</td>
<td>8.39</td>
<td>5.1</td>
<td>3.2</td>
</tr>
<tr>
<td>Stair Descent</td>
<td>5.8</td>
<td>74.5</td>
<td>10.8</td>
<td>5.6</td>
<td>3.2</td>
</tr>
</tbody>
</table>

CONCLUSION
EMG signals by themselves are not very good for classification
Results

• For 5 clusters with hierarchy and all (13) features
  • Percentage of data of each class in different clusters
  • The majority voting is not considered

<table>
<thead>
<tr>
<th>Clusters generated by Gibbs Sampler</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level Ground</td>
<td>38.0</td>
<td>25.2</td>
<td>16.8</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Ramp Ascent</td>
<td>23.8</td>
<td>43.0</td>
<td>15.6</td>
<td>10.0</td>
<td>7.6</td>
</tr>
<tr>
<td>Ramp Descent</td>
<td>24.8</td>
<td>16.8</td>
<td>12.8</td>
<td>7.3</td>
<td>38.3</td>
</tr>
<tr>
<td>Stair Ascent</td>
<td>24.8</td>
<td>15.0</td>
<td>39.5</td>
<td>12.2</td>
<td>8.5</td>
</tr>
<tr>
<td>Stair Descent</td>
<td>11.4</td>
<td>15.6</td>
<td>26.8</td>
<td>36.9</td>
<td>9.3</td>
</tr>
</tbody>
</table>

CONCLUSION
Feature series do help with classification
## Results (Experiment 1)

<table>
<thead>
<tr>
<th>Data Label</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
<th>C12</th>
<th>C13</th>
<th>C14</th>
<th>C15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level Ground</td>
<td>0</td>
<td>33.69</td>
<td>22.8</td>
<td>15.54</td>
<td>10.77</td>
<td>6.9</td>
<td>3.93</td>
<td>2.38</td>
<td>2.2</td>
<td>1.13</td>
<td>0.6</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramp Ascent</td>
<td>0</td>
<td>35.05</td>
<td>23.41</td>
<td>15.61</td>
<td>9.26</td>
<td>6.88</td>
<td>3.31</td>
<td>2.91</td>
<td>1.85</td>
<td>0.93</td>
<td>0.79</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramp Descent</td>
<td>0</td>
<td>35.01</td>
<td>24.36</td>
<td>14.4</td>
<td>11.12</td>
<td>6.67</td>
<td>4.22</td>
<td>0.94</td>
<td>1.64</td>
<td>1.05</td>
<td>0.47</td>
<td>0.12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stair Ascent</td>
<td>0</td>
<td>23.02</td>
<td>16.67</td>
<td>10.2</td>
<td>35.49</td>
<td>6.58</td>
<td>3.51</td>
<td>1.81</td>
<td>1.7</td>
<td>0.79</td>
<td>0.11</td>
<td>0.11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stair Descent</td>
<td>0</td>
<td>23.57</td>
<td>16.9</td>
<td>11.19</td>
<td>32.86</td>
<td>6.67</td>
<td>3.33</td>
<td>2.14</td>
<td>1.67</td>
<td>0.71</td>
<td>0.48</td>
<td>0.36</td>
<td>0</td>
<td>0.12</td>
<td>0</td>
</tr>
</tbody>
</table>

For clusters being chosen based on what is considered **adequate** by model:
- Columns are clusters obtained from model
- Configuration:
  - Hierarchy: Yes
  - Majority Voting: Yes
## Results (Experiment 2)

<table>
<thead>
<tr>
<th>Data Label</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
<th>C12</th>
<th>C13</th>
<th>C14</th>
<th>C15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level Ground</td>
<td>0</td>
<td>50</td>
<td>31.58</td>
<td>7.89</td>
<td>5.26</td>
<td>5.26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramp Ascent</td>
<td>0</td>
<td>70</td>
<td>10</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramp Descent</td>
<td>0</td>
<td>50</td>
<td>33.33</td>
<td>0</td>
<td>16.67</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stair Ascent</td>
<td>0</td>
<td>70.83</td>
<td>16.67</td>
<td>8.33</td>
<td>4.17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stair Descent</td>
<td>0</td>
<td>73.33</td>
<td>23.33</td>
<td>3.33</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

For clusters being chosen based on what is considered **adequate** by model
- Columns are clusters obtained from model
- Configuration:
  - Hierarchy: NO
  - Majority Voting: Yes
### Results (Experiment 3)

<table>
<thead>
<tr>
<th>Data Label</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
<th>C12</th>
<th>C13</th>
<th>C14</th>
<th>C15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level Ground</td>
<td>0</td>
<td>36.07</td>
<td>22.02</td>
<td>14.94</td>
<td>9.05</td>
<td>5.48</td>
<td>4.17</td>
<td>3.1</td>
<td>2.14</td>
<td>1.61</td>
<td>0.83</td>
<td>0.36</td>
<td>0.18</td>
<td>0.06</td>
<td>0</td>
</tr>
<tr>
<td>Ramp Ascent</td>
<td>0</td>
<td>37.7</td>
<td>21.43</td>
<td>13.49</td>
<td>10.05</td>
<td>5.95</td>
<td>3.84</td>
<td>3.44</td>
<td>1.46</td>
<td>1.98</td>
<td>0.4</td>
<td>0.26</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramp Descent</td>
<td>0</td>
<td>36.53</td>
<td>23.19</td>
<td>14.4</td>
<td>9.37</td>
<td>6.79</td>
<td>4.22</td>
<td>2.69</td>
<td>1.76</td>
<td>0.7</td>
<td>0</td>
<td>0.23</td>
<td>0</td>
<td>0</td>
<td>0.12</td>
</tr>
<tr>
<td>Stair Ascent</td>
<td>0</td>
<td>24.38</td>
<td>36.51</td>
<td>13.61</td>
<td>8.84</td>
<td>6.01</td>
<td>4.08</td>
<td>3.63</td>
<td>1.02</td>
<td>0.79</td>
<td>0.34</td>
<td>0.45</td>
<td>0.34</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stair Descent</td>
<td>0</td>
<td>19.4</td>
<td>37.98</td>
<td>16.31</td>
<td>9.76</td>
<td>6.79</td>
<td>4.17</td>
<td>2.62</td>
<td>0.95</td>
<td>1.07</td>
<td>0.71</td>
<td>0</td>
<td>0.24</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

For clusters being chosen based on what is considered **adequate** by model
- Columns are clusters obtained from model
- Configuration:
  - Hierarchy: YES
  - Majority Voting: NO
# Results (Experiment 4)

<table>
<thead>
<tr>
<th>Data Label</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
<th>C12</th>
<th>C13</th>
<th>C14</th>
<th>C15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level Ground</td>
<td>0</td>
<td>65.79</td>
<td>26.32</td>
<td>5.26</td>
<td>0</td>
<td>2.63</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramp Ascent</td>
<td>0</td>
<td>60</td>
<td>30</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramp Descent</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stair Ascent</td>
<td>0</td>
<td>66.67</td>
<td>20.83</td>
<td>12.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stair Descent</td>
<td>0</td>
<td>63.33</td>
<td>20</td>
<td>3.33</td>
<td>10</td>
<td>3.33</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

For clusters being chosen based on what is considered **adequate** by model
- Columns are clusters obtained from model
- Configuration:
  - Hierarchy: NO
  - Majority Voting: NO
Results (Comparison with popular approaches)

- **Our Approach**: 39.1%
- **Complexity Invariant Distance (CID)**: 30.7%
- **Correlation Based Dissimilarity (COR)**: 29.3%
- **Dynamic Time Warping (DTW) measure**: 30.7%
- **Euclidean Distance (EUCL)**: 36%
- **Periodogram-based distances (PER)**: 25.3%
- **Bayesian nonparametric time series clustering**: 26%
Evaluation

• Usually there is one large cluster followed by smaller clusters (due to rich gets richer phenomenon of Dirichlet process) especially for EMG signals without feature extraction

• Introducing hierarchy improves the performance of the model.
  • We believe this is due to the heterogeneity inducing components being represented in the hierarchical parameters rather than the clustering parameters
Future Works and Conclusion

• A semi-parametric approach for clustering Human Activities with natural hierarchy is explored.
  • Having hierarchies helps in getting better clusters, generally
• Future work includes applying the approach to new datasets with non-linear state-space models.
Questions??


References