A Heterogeneous Clustering Approach for Human Activity Recognition

Sabin Kafle(✉) and Dejing Dou

Department of Computer and Information Science,
University of Oregon, Eugene, USA
{skafle,dou}@cs.uoregon.edu

Abstract. Human Activity Recognition (HAR) has a growing research interest due to the widespread presence of motion sensors on user's personal devices. The performance of HAR system deployed on large-scale is often significantly lower than reported due to the sensor-, device-, and person-specific heterogeneities. In this work, we develop a new approach for clustering such heterogeneous data, represented as a time series, which incorporates different level of heterogeneities in the data within the model. Our method is to represent the heterogeneities as a hierarchy where each level in the hierarchy overcomes a specific heterogeneity (e.g., a sensor-specific heterogeneity). Experimental evaluation on Electromyography (EMG) sensor dataset with heterogeneities shows that our method performs favourably compared to other time series clustering approaches.

Keywords: Time series · Heterogeneous clustering · Bayesian semiparametrics · Human Activity Recognition

1 Introduction

The widespread availability of sensors in everyday lives enables us to capture contextual information from underlying human behavior in real-time. This has led to the significant research focus on Human Activity Recognition (HAR) using sensor data [15]. Sensor data is used to determine the specific activity performed by the user at that instant, using either statistical or machine-learning approach. Despite a significant interest on HAR research, real-world performance variations across different sensors have been overlooked [15].

A significant research problem based on use of sensor networks is development of sensor-based automatic prosthetic limbs. The sensor network (usually an EMG sensor network) is used for detecting the intention of the user of the prosthetic limbs in order to provide a better control mechanism to the prosthetic limbs. The sensor network provides data related to the neural intent of the user, which is then interpreted by the prosthetic limb control mechanism to enable certain degree of freedom to the limb motion. For example, the control system is able to recognize whether the user is walking along a level ground or climbing up the

S. Madria and T. Hara (Eds.): DaWaK 2016, LNCS 9829, pp. 68–81, 2016.
DOI: 10.1007/978-3-319-43946-4_5

stairs based on the neural impulse of the user (inferred from sensor data using statistical and machine learning models), which then triggers an intent specific freedom on the prosthetic limbs; e.g., automated rising of the prosthetic limb when the user is climbing up the stairs. While significant progress has been made in the development of prosthetic limbs with such control mechanisms [5], most of the work focus on having a prosthetic limb trained to a specific user only. There is a distinct lack of research in unsupervised learning of user intent from such sensor data.

We focus on developing an unsupervised approach to recognize the user intent based on the sensor data. We treat the sensor data as a time series which is the most natural interpretation of such data. While time series clustering is a significant research area with many different approaches proposed, most of them are inapplicable to our current problem. Time series clustering usually cluster the data obtained from same or similar data source, which is not true for our case. Moreover, most of the approaches require the number of clusters (or activity) in the data to be predetermined which is not always feasible in sensor data. Another challenge lies in the interpretation of the sensor data itself. The sensor data comprises of additive noises and have been found to be inefficient in representing the user intent as raw data themselves [14]. Time and frequency domain features are extracted from sensor data which are then used in machine-learning models for intent interpretation.

In this work, we address the challenges of performing unsupervised learning approach on sensor datasets. We first introduce the heterogeneities in the dataset as a hierarchy with each level in the hierarchy representing a specific heterogeneity. Next, we perform clustering using Bayesian semiparametric approach to mitigate the problem of pre-specifying the number of clusters in the dataset. Our approach learns the number of clusters (or activities) present in the dataset as a parameter of the model, which is capped by some large number that is considered to be an upper limit on the possible number of clusters. Finally, we also develop a feature series clustering approach where we obtain features from the sensor dataset, which is then used to cluster the input data. For evaluation of our approach, we use an EMG sensor dataset collected while the subject performs a walking motion into different terrains. Our dataset consists of eight EMG sensors placed on different limb muscles of a single person during the data collection phase. We find that having a hierarchy to eliminate heterogeneity in the data helps in obtaining better clustering performance. Our method outperforms other approaches which treat the sensor data as a time series in unsupervised learning.

The paper is organized as follows. In Sect. 2, we review previous work on sensor data usage for activity recognition followed by brief description of time series clustering algorithms. We also introduce Hierarchical Normal Model, which is our approach for eliminating heterogeneities in sensor data. We describe our approach in Sect. 3 which includes description of our method and different parameters within the approach. We present experimental results in Sect. 4 and conclude in Sect. 5.

2 Background and Related Work

2.1 Activity Recognition Using EMG Sensors

EMG (Electromyographic) sensors measure electrical current generated in skeletal muscles during its contraction representing neuromuscular activity. The contraction of skeletal muscle is initiated by impulses in the neuron to the muscle and is usually under voluntary control, which are captured by surface EMG sensors. Such signals are significant for detection of gait events of individual.

A *gait* is defined as someone's manner of ambulation or locomotion, involving the total body [2]. The two main phases of gait cycle are the stance phase and the swing phase. A complete gait cycle comprises of - "Heel Strike" (HS), "Flat Foot" (FF), "Mid Stance" (MS), "Heel Off" (HO), "Toe Off" (TO) and "Mid Swing" (MS). Two phases of gait cycle have been found to be most effective in recognizing locomotion mode. One is Heel Strike (also called initial contact), a short period which begins the moment the foot touches the ground while the other is toe-off (also called pre-swing phase), a period when the toe begins to take stance. Activity recognition mechanism based on *gait cycles* involve extracting features from these two phases before classification. Each gait is classified as belonging to a particular activity; e.g., walking up the stairs or walking down the ramp.

Most of the work in EMG signal based terrain identification (also called locomotion mode identification or gait event detection) is based on using classification algorithms, which depend on having labelled training data. The earlier work for EMG signal analysis is based upon wavelet analysis [8] and auto-regressive models [1]. It was demonstrated by [13] that there is a difference in EMG signal envelope among level-ground walking and descending and ascending a ramp, with conclusion that EMG signals from hip-muscles could be used to classify the locomotion modes. The more recent approaches are based on using the features extracted from EMG signals for training a machine learning classification model.

The EMG signals by themselves are random signals with zero mean, but have significance during stages where the muscle contraction is maximum [4]. The features extracted from EMG signals are crucial for getting proper classification accuracy during prediction. The features extracted during the *150* ms phase before and after the "Heel Strike" and "Toe Off" is found to be most accurate for terrain identification [5]. The time domain features which are significant for gait event detection [3] are - *Mean, Variance, Mean Trend, Variance Trend, Windowed Mean Difference, Windowed Variance Difference and Auto-regressive coefficients.*

2.2 Time Series Clustering

Time series clustering is one of the most fundamental and complex task in data mining research. Time series clustering algorithms are usually applied by either converting the popular static clustering approaches to handle time series or by modifying time series to make static clustering methods applicable [9].

One of the most popular approach for static data clustering is k-means or k-mediods, which generate spherical-shaped cluster with a distance measure being considered for deciding cluster membership. Another popular approach for clustering is hierarchical clustering which generate clusters in agglomerative manner (assign each data as an individual cluster and proceed with merging to generate ideal cluster) or divisive manner (partition the data based on some metrics). This approach requires some cluster quality check metrics to determine the best cluster partition. Density-based clustering approach grows a cluster as long as the density of the "neighbourhood" exceeds some threshold value. Model-based clustering approach assumes a model for each cluster and attempts to best fit the data to the assumed model.

The most used approach for clustering time series data is based on computing the similarity measure between different time series and then using the similarity measure to obtain either a spherical cluster partition using *k-means* algorithm or a non-spherical cluster partition using *fuzzy k-means*. Another approach is to extract features from time series and then use those features to perform clustering, either by using a multinomial distribution (when the number of clusters is known *apriori*) or a Dirichlet Process (when the number of clusters is not known apriori). The detail survey and comparison of different time series approach can be found on [11].

2.3 Hierarchical Normal Model

Many different kinds of data, including observational data collected in human and biological sciences, have a hierarchical structure. Sensor signals such as EMG have a natural hierarchy where the measurement of each person is grouped under an individual person and each type of sensor is grouped under that particular sensor. This natural hierarchical tendency of data requires multi-level analysis, which can be incorporated using Hierarchical Normal Models (HNM). HNMs were first studied in the context of biological and human sciences where family, race, geographical location introduces a natural hierarchy in the data [10].

A hierarchy of normal distribution is considered in hierarchical normal model. The top-most level of hierarchy includes a prior for mean and variance of the model (joint prior or distinct prior). A mean value is sampled from the prior, which is then used to sample different means for Level 1 of hierarchy, with the variance obtained from variance prior. For each sub-hierarchy in Level 2, the mean is sampled from each parent sample separately. The variance at each level can be either estimated from the data of that group and kept fixed or obtained from Gibbs sampler step for variance.

An example representation of Hierarchical Normal Model (HNM) is given in Fig. 1. In the figure, the hierarchy moves from left to right. On each level of hierarchy, different components incorporate differences present in that level of hierarchy. We start with a base distribution and add heterogeneities as we move from left to right at each level. The existence of such hierarchies is the result of differentiation in all kind of activities (e.g., different gait events for different person and differing sensor metrics for different muscle activation).

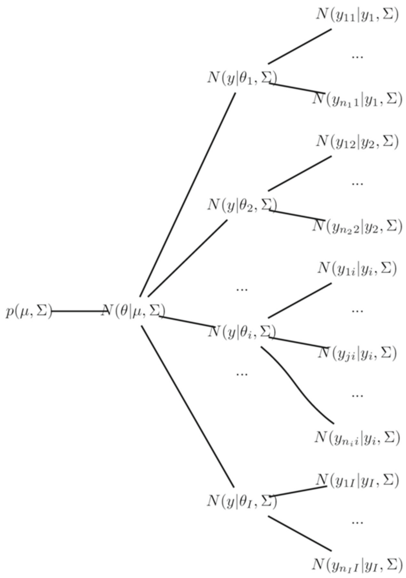

Fig. 1. A Hierarchical Normal Model (HNM).

3 Our Approach

In this section, we describe our model for clustering EMG sensor data. In our approach, we address three key challenges in clustering sensor data:

- Sensor data, especially EMG sensor data, by itself cannot be used for classification or clustering purposes, since it is a noisy time series with *zero mean*. We need to filter out noise from the sensor data before it can be used for activity recognition purposes.
- Heterogeneity is a key challenge and bottleneck for clustering EMG sensor data. Person and Sensor based heterogeneity is a significant impediment to clustering such sensor data.
- Usage of EMG sensor data for activity recognition is heavily dependent on features extraction. We need to incorporate the features from EMG signals for performing clustering, since they have been found to be more useful for activity recognition. Our experimental results also show that using features extracted from EMG sensor data gives much better clustering performance than using raw signal from EMG sensor as input.

We aim to address the above mentioned challenges with our approach. We explain our approach on three subsections each related to the above-mentioned aspects of sensor data analysis. We also borrow some ideas from statistical analysis of time series data and use them heavily in our approach.

3.1 Sensor Data Representation

We represent the sensor data as a time series of $T \times 1$ dimension, also called a vector y_i where i subscript is used to represent the i^{th} time series in the sensor dataset of size N.

We break-down a single EMG sensor data into three distinct latent variables. We use the sampling model [12] to represent sensor data.

$$\boldsymbol{y_i} = \boldsymbol{Z\alpha_i} + \boldsymbol{X\beta_i} + \boldsymbol{\theta_i} + \boldsymbol{\epsilon_i}, i = 1, 2, ..., n \tag{1}$$

where, ϵ_i is $T \times 1$ dimensional random noise. The other three parameters given in the Eq. 1 represent three different latent variables:

- α_i is $p \times 1$ dimensional vector representing the non-clustering components of the sensor data. It is used to enhance the fit of the data to the cluster core to which that data belongs to. For example, mean of the sensor data cannot be significant aspect for clustering such data but can be represented by α_i to make the sensor data fit to its cluster better.
- β_i is the $d \times 1$ dimensional vector representing the clustering but non-autoregressive components of the sensor data. It is used to cluster the sensor data based on several components of time series including trends. d represents the number of components of time series that are considered for clustering. A polynomial trend would require the value of d to be 3.
- θ_i is the $T \times 1$ dimensional vector representing the Auto Regressive $AR(1)$ components of the sensor data assuming stationarity in the time series. We do not assume the sensor data to be stationary but only consider some components represented by θ_i as an auto-regressive component.

Our approach does not require explicit specification of different aspects of sensor data that are considered for clustering or not but they are learned during the training phase automatically. The matrices Z and X are design matrices of dimensions $T \times p$ and $T \times d$ respectively.

We provide Bayesian treatment to our approach. This is done by assuming that sensor data and latent variables are generated by multivariate Normal distributions. We, then represent a single sensor data as a function of multivariate Normal Distribution given below:

$$\begin{aligned} \boldsymbol{f(y_i) \propto N_T(Z\alpha_i + X\beta_i + \theta_i, \sigma^2_{\epsilon_i} I)} \\ \alpha_i \sim N(0, \Sigma_\alpha) \\ \beta_i \sim N(\overline{\beta_{s,r,k}}, \Sigma_{\beta,s,r,k}) \\ \theta_i \sim N(\overline{\theta_{s,r,k}}, \Sigma_{\theta,s,r,k}) \end{aligned} \tag{2}$$

where parameters with s, r, k as subscript represent the specific values obtained after incorporating the heterogeneity into the non-heterogeneous clustering parameters.

For simplicity, we assume covariance matrices to be diagonal matrices for Multivariate Normal Distribution and each diagonal elements obtained as a sample from Inverse-Gamma (IGa) prior. This completes the specification of our approach to separate noisy components from sensor data.

3.2 Incorporation of Heterogeneities in Sensor Data

We incorporate heterogeneities in EMG sensor data by using a Hierarchical Normal Model (HNM) where each level of hierarchy represents a specific heterogeneity in the data. We use two level of hierarchy to represent two different heterogeneities in EMG sensor datasets. The first level of hierarchy represents Sensor Level differences since the differences in calibration of different sensors and their positioning play a significant role in creating heterogeneity. The second level of hierarchy is used to represent person specific heterogeneities since for different person, a sensor with same calibrations and positioning will have different readings based on differences between the persons' biomechanics. We also incorporate heterogeneities for each activity (or clusters) separately. This is done to reduce the number of parameters since our approach enables us to consider heterogeneity only for clustering parameters and allow non-clustering parameters to fit independent of heterogeneity.

The top-most level of hierarchy for a specific cluster k is:

$$\gamma_{root} \sim N(0, \Sigma_\beta) \times N(0, \Sigma_\theta) \tag{3}$$

where γ_{root} is considered product of two clustering parameters i.e. $\beta_{root} \times \theta_{root}$. The prior for covariance of β parameter is diagonal matrix with IGa prior while for θ it is adapted from [12] to handle stationarity.

For the second level, which is the sensor specific heterogeneities incorporating level, we have R branches where R is the total number of sensor types present in the data. We sample a mean value from the parent which is *root* and use the covariance matrix from that level attached to the specific component to sample the clustering components that incorporates sensor level of heterogeneities. Specifically, for a sensor r data belonging to cluster k, we obtain the clustering component as:

$$\begin{aligned} \gamma_{r,k} \sim N(\overline{\beta_{r,k}}, \Sigma_{\beta,r,k}) \times N(\overline{\theta_{r,k}}, \Sigma_{\theta,r,k}) \\ \overline{\beta_{r,k}} \sim N(0, \Sigma_{\beta,k}) \\ \overline{\theta_{r,k}} \sim N(0, \Sigma_{\theta,k}) \end{aligned} \tag{4}$$

where $\Sigma_{\beta,r,k}$ and $\Sigma_{\theta,r,k}$ are covariances for β and θ parameters of cluster k, level 2 and branch r.

Similarly, for person specific heterogeneities incorporating level (which is level 3), we consider S persons for each sensor branch r in the previous level and obtain a person's heterogeneity incorporating clustering parameters as:

$$\begin{aligned} \gamma_{s,r,k} \sim N(\overline{\beta_{s,r,k}}, \Sigma_{\beta,s,r,k}) \times N(\overline{\theta_{s,r,k}}, \Sigma_{\theta,s,r,k}) \\ \overline{\beta_{s,r,k}} \sim N(\overline{\beta_{r,k}}, \Sigma_{\beta,r,k}) \\ \overline{\theta_{s,r,k}} \sim N(\overline{\theta_{r,k}}, \Sigma_{\theta,r,k}) \end{aligned} \tag{5}$$

Here, k represents the k^{th} cluster from K clusters, r represents the r^{th} sensor from R sensors and s represents the s^{th} person among S persons.

The block diagram of HNM for heterogeneities is given in Fig. 2 for more clarity.

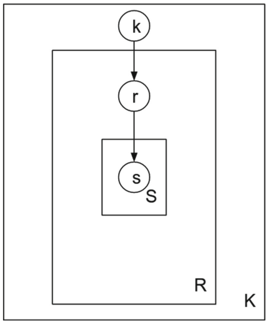

Fig. 2. Block HNMs for heterogeneity. Each k cluster is represented by one such block which incorporates two level of heterogeneities. For Sensor data, two level of heterogeneities represent sensor (r) and person (s) based heterogeneities.

For clustering, we use the Generalised Dirichlet Process (GDD) based prior as explained in [6]. The selection of GDD based prior is for two main reasons - (i) It enables us to learn number of clusters from the data itself. (ii) The posterior of GDD is conjugate with multinomial sampling and thus we can easily run Gibbs Sampling for cluster inference. The number of clusters that can sufficiently represent the data is obtained by using Adequate Truncation Value method explained in [6] during the first few iterations of Gibbs Sampling.

The prior characterization of our approach is now complete. The posterior for every random variables introduced in our approach can be obtained analytically based upon the likelihood function given as:

$$\boldsymbol{f(y) = \prod_{i=1}^{N} N_T(Z\alpha_i + X\beta_i + \theta_i, \Sigma_y)} \tag{6}$$

where, $\Sigma_y = \sigma_\epsilon^2 I$ and N is the total data count.

For the sake of clarity and brevity, we include posterior characterizations of our approach in Appendix A.

3.3 Inference and Feature Clustering

Gibbs Sampling algorithm is used for posterior inference, with Metropolis within Gibbs sampler being used for sampling ρ and σ_θ^2. The Gibbs Sampler algorithm used is same as given in [7] except for sampling from the hierarchical model and Metropolis steps for θ parameters. The hierarchical model's posterior sampling is done in bottom-up approach. The hierarchy is sampled beginning from the Sampling model until the top level is reached.

Features extraction from EMG sensor data is necessary to obtain better classification and clustering results. We extend our approach to multi-dimensional time series to handle EMG sensor data where we extract features for every 50 ms window. Each features extracted creates a series of their own (also called a vector).

We call such feature vector of EMG sensor data a feature series. We assume each feature series is independent of one another in order to reduce the complexity of the model. Then each feature is considered as an independent EMG sensor data and the above model is applied to all the features.

The most significant aspect during handling such multiple features is to consider how each feature series impacts the overall clustering aspect our approach. We propose two approaches for such cases:

- The Generalized Dirichlet Distribution (GDD) is used for combining different feature series. A single cluster label is selected for all the feature series. The parameters of GDD are updated with each feature series likelihood w.r.t. the data. Rest of the model is kept same as explained above.
- Different GDD is used for each feature series. The final cluster membership is based on majority voting of cluster assignments in individual feature series.

We perform experiments to evaluate each approach and find out that using a single GDD for combining different feature series works best for our dataset.

3.4 Cluster Selection

Each iteration of Gibbs Sampling produces a cluster assignment of data points, which is then filtered to select one cluster assignment as the best fit. One way of selecting a cluster membership used by [12] is Heterogeneity Measure (HM), which can be calculated as:

$$HM(G_1, .., G_m) = \sum_{k=1}^{m} \frac{2}{n_k - 1} \sum_{i<j \in G_k} \sum_{t=1}^{T} (y_{it} - y_{jt})^2 \qquad (7)$$

where, m is the number of clusters, G_k is k^{th} cluster, n_k is the number of data point belonging to cluster G_k, T is the length of time series representation of sensor data and y_{it} is the t^{th} value in time series y_i.

The larger the value of HM, the more heterogeneous a clustering is. It is preferable to have a cluster with small HM and small m.

4 Experiments

Our dataset consists of 9 normal human subjects performing 5 different gait events which are measured by eight different sensors placed in their bodies. In this experiment, we attempt to cluster each sensor data into individual gait events (level ground walking, stair ascent, stair descent, ramp ascent and ramp descent). We consider a data point to be a sensor reading of a single gait cycle. We conduct several experiments with different configuration of hyper parameters in order to determine the best configuration for the dataset. The number of clusters is determined initially using the Adequate Truncation Value during the initial Gibbs Sampling phase but we found that using the same number of clusters as in original dataset gives the best result. We found out 15 is sufficient number of clusters for this data. In order to eliminate the disproportions of data in different classes, we use subsampling to get the equal number of input data for each class.

The dimension of design matrices $Z_{T \times p}$ and $X_{T \times d}$ is an important design decision. We use $p = 1$ since we found that the value of p doesn't play a significant role for our dataset. For X, we set $d = 7$ with first three columns representing the polynomial trend of degree 3, with remaining four columns being used as a latent trait indicator for four gait phases (Before Heel Strike, After Heel Strike, Before Toe Off, After Toe Off). We also set the inverse gamma prior fixed to $[2, 1]$ throughout the experiment. The results presented is based on the heterogeneity score of sampled cluster membership, with the lowest score being selected as best clustering assignment. The best cluster membership obtained from the sampler based on Heterogeneity Measure is then used to obtain a confusion matrix. The confusion matrix is used to compute the accuracy of the clustering approach with labelled examples. The Heterogeneity Measure for every result obtained is between 0.5 and 1.95, with random impact on the accuracy of the clustering. For all the experiments, we run Gibbs sampler upto 5000 iterations, with 3000 as burn-in phase and collect a sample every 200 iterations after the burn-in phase.

First we conduct experiments to determine the different aspects of the model. The results are presented in Fig. 3. The first result in Fig. 3 shows that EMG data by itself is not meaningful for classifying or clustering purposes.

We obtain best result with configurations given as second bar in Fig. 3 where we specify the number of clusters same as the number of labels. The comparison between third bar and fifth bar of Fig. 3 along with fourth and sixth bar of Fig. 3 illustrate that majority voting for feature series performs slightly worse than a single cluster membership for entire feature series approach. Also, the use of

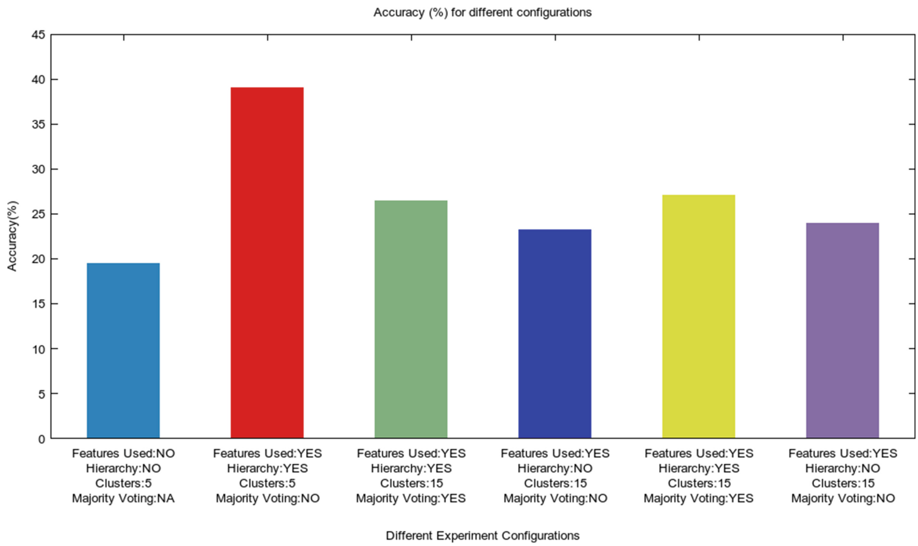

Fig. 3. Results obtained for different configuration of Model

hierarchy helps in obtaining the better performance compared to the non-usage of hierarchy as evident from third and fifth bar comparison.

Next, we compare the performance of our approach with other popular time series clustering algorithms. The result is presented in Table 1. It is evident that our approach outperforms other approaches for time series clustering in case of our dataset which consists of heterogeneity. For k-means based clustering experiment, we use TSclust package [11]. We refer the reader to [11] for more information about different distance metrices.

Table 1. Performance measure of different time series clustering approaches

Method	**Accuracy (%)**
Bayesian Nonparametrics Time Series Clustering (BNPTSclust)[1] [12]	26.0
Rest of the algorithms are k-means clustering algorithms implemented in [11]	
Autocorrelation based Dissimilarity (ACF)	26.0
Periodogram-based distances (PER)	25.3
Normalized Compression Distance (NCD)	23.3
Euclidean Distance (EUCL)	36.0
Compression-based dissimilarity measure (CDM)	24.7
Dynamic Time Warping (DTW) measure	31.3
Discrete Wavelet Transform (DWT)	30.7
Correlation Based Dissimilarity (COR)	29.3
Partial Autocorrelation based Dissimilarity (PACF)	28.0
Complexity Invariant Distance (CID)	30.7
Permutation Distribution Clustering (PDC)[2]	18.7
Our Approach[3]	**39.1**

[1] This approach is based on Bayesian non parametrics where the number of clusters is inferred from the data itself. This approach favours a single cluster most of the time.
[2] Used default configuration provided in TSclust package for clustering.
[3] The best accuracy is obtained when not considering Majority Voting, while specifying the number of clusters to be only 5.

5 Conclusion and Future Work

We study the feasibility of clustering approach for Human Activity Recognition using sensor dataset. Our approach introduces hierarchy-based heterogeneity for clustering time series where the number of clusters is not known in advance. Experimental result shows that introducing hierarchy helps in clustering sensor-based time series more accurately. Though the accuracy of our approach for EMG sensor data is low, comparison with other time series clustering approaches show that our method performs better than other approaches. The current method expresses the time series as a linear model only, future work will involve extension to non-linear models to handle more complex time series, along with using more datasets for further experiments.

Acknowledgements. This work is partially supported by the NIH grant R01GM1 03309. We acknowledge Deepak Joshi and Michel Kinsy for their inputs.

Appendix A Posterior Characterization

- α_i: The posterior for α_i is:

$$\begin{aligned} f(\alpha_i|rest) &\propto N_P(\mu_a, \Sigma_a) \\ \Sigma_a &= (\Sigma_\alpha^{-1} + Z^T \Sigma_y^{-1} Z)^{-1} \\ \mu_a &= \Sigma_a Z^T \Sigma_y^{-1}(y_i - X\beta_i - \theta_i) \\ f(\sigma^2_{\alpha_j}|rest) &= IGa(c_0^\alpha + \frac{n}{2}, c_1^\alpha + \frac{1}{2}\sum_{i=1}^{n}\alpha_{ij}^2), j = 1, .., p \end{aligned} \tag{8}$$

- β_i: The posterior for β_i (or $\beta_{s,r,k}$) is:

$$\begin{aligned} f(\beta_i|rest) &\propto N_D(\mu_b, \Sigma_b) \\ \Sigma_b &= (\Sigma^{-1}_{\beta,s,r,k} + X^T \Sigma_y^{-1} X)^{-1} \\ \mu_b &= \Sigma_b[X^T \Sigma_y^{-1}(y_i - Z\alpha_i - \theta_i) + \Sigma_{\beta,s,r,k}\overline{\beta_{s,r,k}}] \\ f(\sigma^2_{\beta_{s,r,k,i}}|rest) &= IGa(c_0^{\beta_{s,r,k,i}} + \frac{m}{2}, c_1^{\beta_{s,r,k,i}} + \frac{1}{2}\sum_{j=1}^{m}\beta^2_{s,r,k,i}) \\ i &= 1, .., p \end{aligned} \tag{9}$$

where m is the number of data points belonging to that cluster.

- θ_i: The posterior for θ_i (or $\theta_{s,r,k}$) is:

$$\begin{aligned} f(\theta_i|rest) &\propto N_T(\mu_c, \Sigma_c) \\ \Sigma_c &= (\Sigma^{-1}_{\theta,s,r,k} + \Sigma_y^{-1})^{-1} \\ \mu_c &= \Sigma_c[\Sigma_y^{-1}(y_i - Z\alpha_i - X\beta_i) + \Sigma_{\theta,s,r,k}\overline{\theta_{s,r,k}}] \\ f(\sigma^2_{\theta_{s,r,k,i}}|rest) &= IGa(c_0^{\theta_{s,r,k,i}} + \frac{m}{2}, c_1^{\theta_{s,r,k,i}} + \frac{1}{2}\sum_{j=1}^{m}\theta^2_{s,r,k,i}) \\ i &= 1, .., T \end{aligned} \tag{10}$$

where m is the number of data points belonging to that cluster.

- $\sigma^2_{\epsilon_i}$: The posterior for $\sigma^2_{\epsilon_i}$ is:

$$\begin{aligned} f(\sigma^2_{\epsilon_i}|rest) &\propto IGa(c_0^\epsilon + \frac{T}{2}, c_1^\epsilon + \frac{1}{2}M_i^{'}M_i) \\ M_i &= (y_i - Z\alpha_i - X\beta_i - \theta_i) \end{aligned} \tag{11}$$

- *Level k posterior*: The posterior for any level of hierarchy except for top-most level consists of following updates:

$$
\begin{aligned}
f(\beta_k|rest) &\propto N_D(\mu_g, \Sigma_g) \\
\Sigma_g &= (\Sigma_{\beta,r,k}^{-1} + \Sigma_{\beta,k}^{-1})^{-1} \\
\mu_g &= \Sigma_g(\Sigma_{\beta,k}\overline{\beta_k} + \Sigma_{\beta,r,k}^{-1}\beta_{r,k}) \\
f(\sigma^2_{\beta_{k,i}}|rest) &= IGa(c_0^{\beta_{k,i}}\frac{R}{2}, c_1^{\beta_{k,i}}\sum_{j=1}^{S}\beta^2_{k,i}) \\
f(\theta_k|rest) &\propto N_D(\mu_h, \Sigma_h) \\
\Sigma_h &= (\Sigma_{\theta,r,k}^{-1} + \Sigma_{\theta,k}^{-1})^{-1} \\
\mu_h &= \Sigma_h(\Sigma_{\theta,k}\overline{\theta_k} + \Sigma_{\theta,r,k}^{-1}\theta_{r,k}) \\
f(\sigma^2_{\theta_{k,i}}|rest) &= IGa(c_0^{\theta_{k,i}}\frac{R}{2}, c_1^{\theta_{k,i}}\sum_{j=1}^{R}\beta^2_{k,i})
\end{aligned}
\tag{12}
$$

- *Top level posterior*: The posterior at top-most level is:

$$
\begin{aligned}
f(\beta|rest) &\propto N_D(\mu_e, \Sigma_e) \\
\Sigma_e &= (\Sigma_{\beta,k}^{-1} + \Sigma_{\beta}^{-1})^{-1} \\
\mu_e &= \Sigma_e(\Sigma_{\beta,k}^{-1}\beta_k) \\
f(\sigma^2_{\beta_i}|rest) &= IGa(c_0^{\beta_i}\frac{K}{2}, c_1^{\beta_i}\sum_{j=1}^{K}\beta_i^2) \\
f(\theta|rest) &\propto N_D(\mu_f, \Sigma_f) \\
\Sigma_f &= (\Sigma_{\theta,k}^{-1} + \Sigma_{\theta}^{-1})^{-1} \\
\mu_f &= \Sigma_f(\Sigma_{\theta,k}^{-1}\theta_k) \\
f(\sigma^2_{\theta}|rest) &= IGa(\frac{KT}{2}, \frac{1}{2}\sum_{j=1}^{K}\theta_j^{'}Q^{-1}\theta_j) \\
f(\rho|rest) &\propto |Q|^{-K/2}\exp\frac{-1}{2\sigma^2_{\theta}}\sum_{j=1}^{K}\theta_j^{'}Q^{-1}\theta_j\frac{\sqrt{1+\rho^2}}{1-\rho^2}
\end{aligned}
\tag{13}
$$

  where $Q_{ij} = \rho^{|i-j|}$ for $i, j = 1, .., T$.
- *Posterior for GDD and p*: The posterior for GDD is conjugate with multinomial sampling. The probability p is updated based on the fit of the data with respect to the individual clusters lowest level mean using the likelihood function. The complete detail for GDD posterior characterization can be found in [6].

This completes the posterior characterization of our approach.

References

1. Graupe, D., Cline, W.K.: Functional separation of emg signals via arma identification methods for prosthesis control purposes. IEEE Trans. Syst. Man Cybern. **5**(2), 252–259 (1975)
2. Griffin, L.Y., Albohm, M.J., Arendt, E.A., Bahr, R., Beynnon, B.D., DeMaio, M., Dick, R.W., Engebretsen, L., Garrett, W.E., Hannafin, J.A., et al.: Understanding and preventing noncontact anterior cruciate ligament injuries: a review of the Hunt Valley II meeting, January 2005. Am. J. Sports Med. **34**(9), 1512–1532 (2006)
3. Gupta, P., Dallas, T.: Feature selection and activity recognition system using a single triaxial accelerometer. IEEE Trans. Biomed. Eng. **61**(6), 1780–1786 (2014)
4. Huang, H., Kuiken, T., Lipschutz, R.D., et al.: A strategy for identifying locomotion modes using surface electromyography. IEEE Trans. Biomed. Eng. **56**(1), 65–73 (2009)
5. Huang, H., Zhang, F., Hargrove, L.J., Dou, Z., Rogers, D.R., Englehart, K.B.: Continuous locomotion-mode identification for prosthetic legs based on neuromuscular-mechanical fusion. IEEE Trans. Biomed. Eng. **58**(10), 2867–2875 (2011)
6. Ishwaran, H., James, L.F.: Gibbs sampling methods for stick-breaking priors. J. Am. Stat. Assoc. **96**(453), 161–173 (2001)
7. Ishwaran, H., Zarepour, M.: Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models. Biometrika **87**(2), 371–390 (2000)
8. Kumar, D.K., Pah, N.D., Bradley, A.: Wavelet analysis of surface electromyography. IEEE Trans. Neural Syst. Rehabil. Eng. **11**(4), 400–406 (2003)
9. Liao, T.W.: Clustering of time series data–a survey. Pattern Recogn. **38**(11), 1857–1874 (2005)
10. Lindley, D.V., Smith, A.F.: Bayes estimates for the linear model. J. Roy. Stat. Soc.: Ser. B (Methodol.) **34**, 1–41 (1972)
11. Montero, P., Vilar, J.A.: TSclust: An R package for time series clustering. J. Stat. Softw. **62**(1), 1–43 (2014). http://www.jstatsoft.org/v62/i01/
12. Nieto-Barajas, L.E., Contreras-Cristan, A.: A bayesian nonparametric approach for time series clustering. Bayesian Anal. **9**(1), 147–170 (2014)
13. Peeraer, L., Aeyels, B., Van der Perre, G.: Development of emg-based mode and intent recognition algorithms for a computer-controlled above-knee prosthesis. J. Biomed. Eng. **12**(3), 178–182 (1990)
14. Reaz, M., Hussain, M., Mohd-Yasin, F.: Techniques of EMG signal analysis: detection, processing, classification and applications. Biolog. Proc. Online **8**(1), 11–35 (2006)
15. Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T.S., Kjærgaard, M.B., Dey, A., Sonne, T., Jensen, M.M.: Smart devices are different: assessing and mitigating mobile sensing heterogeneities for activity recognition. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, pp. 127–140 (2015)