# Investigation of Cluster Validity Indices for Unsupervised Human Activity Discovery

**Wee-Hong Ong**[1], **Takafumi Koseki**[1] and **Leon Palafox**[2]
[1]Department of Electrical Engineering and Information Systems, The University of Tokyo, Tokyo, Japan
[2]Department of Radiology, University of California, Los Angeles, Los Angeles, USA

**Abstract -** *An approach for unsupervised human activity discovery has been proposed in this paper. The approach automatically discover unknown activities from unlabeled data and has the ability to reject random activities. This ability will enable intelligent systems to discover and learn new activities autonomously. K-means is used to cluster a pool of unlabeled activity observations into groups of different activities. The system requires no prior knowledge of how many activities to be discovered. It uses cluster validity indices to automatically estimate the required number of clusters and further evaluate cluster homogeneity to accept clusters with homogenous activity and reject clusters with random activities. Experimental results showed the potential of the approach and identified suitable validity indices to achieve unsupervised human activity discovery.*

**Keywords**: human activity detection; human activity discovery; unsupervised learning; clustering; RGBD sensor

## 1 Introduction

From the detailed survey on human activity analysis by Aggarwal and Ryoo [4], we can see that there have been significant efforts to accurately capture human motion, i.e., solving computer vision problem and recognize the activity from the motion, i.e., solving modeling and learning problems. Systems with high accuracy in capturing human motion are expensive and require infrastructure setup. Many such systems require markers attached to the subjects. Further, those systems use supervised learning algorithms to learn models of activities. In these systems, pre-labeled examples are provided to the learning algorithm. However, to enable intelligent systems to autonomously learn new activities, they will be required to deal with unlabeled data. For this reason, there have been increasing interest to investigate human activity discovery using unsupervised learning. In the knowledge of the authors, most of the works relating unsupervised learning with activity recognition have been either focusing on solving computer vision problems or they require alternative form of pre-labeled data from wearable sensors or other sources. For examples, Song et al. [17] developed an EM-like algorithm on decomposable triangulated graphs to extract human as foreground from background clutter. Huynh et al. [12] used clustering to generate a vocabulary of labels from sensor data, which are then used for pattern extraction using topic models to recognize daily routines. They used data from custom made wearable sensors. Stikic et al. [9] applied two weakly supervised methods to discover activities from two published datasets obtained from wearable sensors. Wyatt et al. [2] described their techniques for mining object models from the web and use the information to recognize activities based on the interaction of user with objects. They attached RFID tags to the objects. We observed that current state-of-art of human activity recognition technologies is not cost effective and not suitable in our natural living environment. They either require expensive setup, wearable sensors or pre-labeled data.

Our work is motivated by the intend to create a fully autonomous personal intelligent agent, for example a personal robot, that is capable of understanding what its user, or owner, is doing and consequently provide appropriate support. Therefore, human activity recognition lies at the core of such system. The system should be able to be deployed in the natural human living environment with minimal changes, at low-cost and work with unlabeled data. Further, the system should not require users to wear markers or sensors. These requirements call for the use of low-cost components, lightweight algorithms, marker-less vision and/or audio sensors, and unsupervised learning. Recently, the availability of low-cost RGBD (RGB-Depth) sensors has enabled accurate capture of human poses. In our earlier work [15], [16], we demonstrated the ability of K-means to distinguish different activities using just the skeleton data obtained from low-cost RGBD sensor, Microsoft Kinect. The problems with the use of K-means are the need to specify the number of clusters, $k$ value, apriori and such system do not have the ability to reject random activities. Random activities will be assigned to a cluster anyway. In this paper, we propose an approach to address these two problems. We propose an approach to perform unsupervised human activity discovery. The approach uses cluster validity indices to evaluate the quality of clustering outcome and automatically determines suitable number of clusters and evaluates each cluster to identify highly homogeneous clusters as new activities.

# 2 Proposed approach

Assuming an intelligent system or agent has been observing daily activities of a subject. An example situation is a personal robot accompanying its owner. The system has collected a large pool of unlabeled activity observations (data). To be low cost with minimal requirements, we allow the system to be a best-effort one. The approach resembles how children learn activities around them. Children have the ability to distinguish different activities and ask adults to label the activities, i.e., they learn from unlabeled data (observations) and post-label the model they have formed. Children can learn from one subject, e.g., a parent, and adapt the model to different subjects. While there are many activities going on, children don't learn all of them at once. Finally, there are times that children are confused with similar activities.

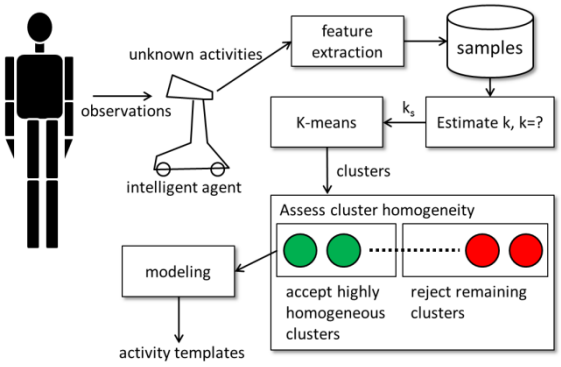


Figure 1. Proposed approach for unsupervised activity discovery in human activity recognition.

We propose the following steps to autonomously discover new activities, as illustrated in Fig. 1. The different cluster validity indices are described in Section 3.

1. Collect observations of unknown activities.
2. Use global internal cluster validity index to estimate the number of clusters, *k*.
3. Do clustering using the suggested number of clusters, $k_s$, from above.
4. Assess homogeneity of each of the $k_s$ clusters using local internal cluster validity measure.
5. Accept high ranking clusters and reject clusters ranked low by the validity measure.

# 3 Unsupervised learning

## 3.1 K-means clustering

K-means [5] clustering is one of the simplest unsupervised learning algorithms. It looks for similarity among the examples in the dataset by using simple distance measurement. Given the required number of clusters, *k*, K-means group the points (examples) in the dataset by minimizing the distance from each data point to a cluster center (centroid). We have used K-means clustering that minimizes the regular cost function given in Eq. (1).

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left\| x_i^{(j)} - c_j \right\|^2 \quad (1)$$

where *k* is the number of clusters, $x_i^{(j)}$ is *i*th data point in Cluster *j* and $c_j$ is the centroid of Cluster *j*, i.e., $C_j$, $n_j$ is the number of data points in $C_j$. Note that each data point is a row vector in all equations presented in this paper.

## 3.2 Cluster validation

We expect the intelligent system to autonomously discover new activities and the number of activities to be found is unknown to the system. Various cluster validity indices [13] have been proposed to assess the quality of clustering outcome and determine the appropriate *k*. Since we will deal with unlabeled data, we consider only internal cluster validity indices that do not require labeled data. In this paper, we tested five indices as given below.

1. Silhouette (Sil) [10]. For each data point, $a(x_i)$ is the average distance from the point to other points in its cluster, and $b(x_i)$ is the average distance from it to all points in nearest cluster. The objective is to maximize *Sil(k)*.

$$per\ point\ Sil(x_i) = \frac{b(x_i) - a(x_i)}{\max\{b(x_i), a(x_i)\}} \quad (2)$$

$$per\ cluster\ Sil(C_j) = \frac{1}{|C_j|} \sum_{x_i \in C_j} Sil\left(x_i^{(j)}\right) \quad (3)$$

$$overall\ Sil(k) = \frac{1}{k} \sum_{r=1}^{k} Sil(C_j) \quad (4)$$

2. Davies-Bouldin (DB) index [1]. The objective is to minimize the *DB(k)* index.

$$DB(k) = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left\{ \frac{s_i + s_j}{d_{ij}} \right\} \quad (5)$$

$$\boldsymbol{s_j} = \frac{1}{n_j} \sum_{x_i \in C_j} \left\| x_i^{(j)} - c_j \right\|, \ \boldsymbol{d_{ij}} = \|c_i - c_j\| \quad (6)$$

3. Calinski-Harabasz (CH) index [11]. The objective is to maximize *CH(k)*.

$$CH(k) = \frac{trace(SSW(k))/k-1}{trace(SSB(k))/n-k} \quad (7)$$

where $SSW(k)$ is the (sum-of-square) within-cluster scatter matrix and, $SSB(k)$ is the (sum-of-square) between-cluster scatter matrix as given below:

$$SSW(k) = \sum_{j=1}^{k} \sum_{x_i \in C_j} \left(x_i^{(j)} - c_j\right)^T \left(x_i^{(j)} - c_j\right) \quad (8)$$

$$SSB(k) = \sum_{j=1}^{k} n_j (c_j - \mu)^T (c_j - \mu) \quad (9)$$

where $\mu$ is the mean of the whole dataset.

4. Krzanowski-Lai (KL) index [14]. The objective is to maximize *KL(k)*.

$$KL(k) = \left| \frac{DIFF(k)}{DIFF(k+1)} \right| \quad (10)$$

$$DIFF(k) = (k-1)^{2/p} trace(SSW(k-1)) - k^{2/p} trace(SSW(k)) \quad (11)$$

where $p$ is the dimension (number of variables/features) of the data point; $p = 630$ in our case, see Section 4.

5. Hartigan (Ha) index [3]. The objective is to add cluster until Ha is below a threshold. $Ha \leq 10$ is typically used and we have used this value in our work reported in this paper.

$$Ha(k) = \left( \frac{trace(SSW(k))}{trace(SSW(k+1))} - 1 \right) (n - k - 1) \quad (12)$$

where *n* is the number of data points in whole data set.

The above indices are global as they consider all clusters in the validation. Even with the value of *k* given, there is no guarantee that K-means as well as any other clustering algorithm will group all observations of the same activity into same cluster. To assess the homogeneity, i.e., cohesion and compactness of individual cluster, we require local internal cluster validity indices. To assess the homogeneity of individual cluster and rank them accordingly, we define two measures: the intra-cluster mean variance ($\bar{\sigma}^2$) and mean joint probability density function ($\bar{P}_f$). Low value of $\bar{\sigma}^2$ indicates compactness of the cluster. The joint probability density function assumes that observations of a non-random activity should be normally distributed within its cluster around the cluster centroid with the standard deviation of the cluster. High value of $\bar{P}_f$ indicates good cohesion of the cluster based on assumption of normal distribution. The equations for the two measures are given in Equations (13) to (18). Logarithm is used in Equation (17) to compress the range of values.

$$mean\ variance\ \bar{\sigma}^2(\boldsymbol{C_j}) = mean\left(var(C_j)\right) \quad (13)$$

$$var(C_j) = \frac{1}{n_j - 1} \sum_{x_i \in C_j} \left( x_i^{(j)} - \bar{x}^{(j)} \right)^{\wedge 2} \quad (14)$$

$$\bar{\boldsymbol{x}}^{(j)} = mean(x_i \in C_j) \quad (15)$$

where $x_i^{(j)} = [\, x_{i1} \cdots x_{ip} \,]$ is a data point in Cluster $C_j$ with dimension $p$, $n_j$ is the number of points in Cluster $C_j$, $\bar{x}^{(j)}$ (dimension $p$) is the mean of all points in Cluster $C_j$, $\wedge 2$ is element-wise square.

$$mean\ joint\ probability\ \bar{P}_f(\boldsymbol{C_j}) = mean\left(P(C_j)\right) \quad (16)$$

$$P(C_j) = \sum_{f=1}^{p} ln\left(pdf\left(x_{if}^{(j)}\right)\right) \quad (17)$$

$$pdf\left(x_{if}^{(j)}\right) = \frac{1}{\sigma_f^{(j)}\sqrt{2\pi}} e^{-\frac{\left(x_{if}^{(j)} - \mu_f^{(j)}\right)^2}{\left(2\sigma_f^{(j)}\right)^2}} \quad (18)$$

where $x_{if}^{(j)}$ is the *f*th dimension of $x_i^{(j)}$, $\mu_f^{(j)}$ is the mean of values of *f*th dimension of all points in Cluster $C_j$, $\sigma_f^{(j)}$ is the standard deviation of values of *f*th dimension of all points in Cluster $C_j$.

# 4 Feature extraction

Feature extraction in the context of this paper is not about image processing. The raw data were coordinates of 15 joints in human skeleton provided by the application of Microsoft Kinect, a low-cost RGBD (RGB-Depth) camera, as shown in Fig. 2. Each activity example was sampled for a window of 2 seconds comprising 15 frames. We found empirically [16] that reducing the frames from full 60 frames to 15 frames did not degrade clustering performance. For each frame, the following features were extracted from the coordinates of the joint positions: four vectors describing body flexion, four vectors describing arms abduction, four vectors describing leg abduction and flexion and, two vectors describing interaction between hands and head. The vectors were formed locally (between joints) and normalized to shoulder width making them view invariant to camera and scale invariant to the size of the subject. There were 14 3-dimentional vectors giving 14×3=42 features per frame. With 42 features per frame, the total number of features was 42×15=630 features per activity observation.

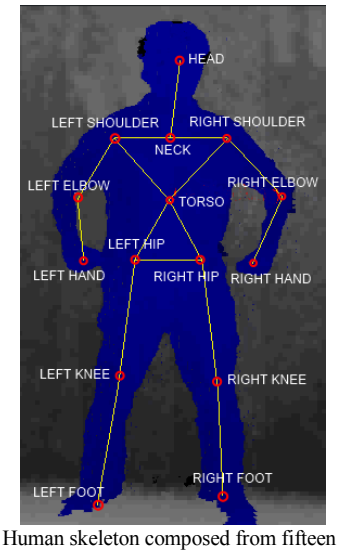


Figure 2. Human skeleton composed from fifteen (15) joint.

# 5 Data & experiment

## 5.1 Data

Currently, there are a few [6], [7] publicly available skeleton (coordinates of joints) datasets obtained from Microsoft Kinect sensor on human activities. None of these datasets have been adapted as benchmarking dataset for evaluation of human activity research works. The work presented in this paper used the dataset "Cornell Activity Dataset CAD-60" [6]. CAD-60 consisted of twelve daily activities: rinsing mouth, brushing teeth, wearing contact lens, talking on the phone, drinking water, opening pill container, cooking (chopping), cooking (stirring), talking on couch, relaxing on couch, writing on whiteboard, working on computer. The data was collected from four different subjects: two males (referred as Person 1 and Person 4) and two females (referred as Person 2 and Person 3). One of the females is left-handed (Person 3). Still (standing) and random activity samples by each subject are also included in the dataset. All of the data were collected in a regular household setting with no occlusion of body from the view of sensor. The CAD-60 dataset comprises of RGB images, depth images and skeleton data (coordinates of joint positions and orientations). We have only used the skeleton data of the joint positions in our experiment. We considered eight of them as listed in Table I (1 to 4, 6 to 9). The other three activities are not atomic in their dataset. We refer an atomic activity as one that cannot be further decomposed into sequence of smaller activities. For examples, the drinking comprised of picking up the cup and drink; rinsing mouth comprised of sipping water, gargle and spit; opening pill container comprised of lifting the pill box, twist the cap. At this stage, we are interested to discover atomic actions or lower-level activities, which will be used to discover higher-level activities eventually. We also considered the still (standing) as one activity. In total, we considered nine activities. The random activity samples were also used to test the ability of our approach to reject random activities. Two datasets were composed for each subject as following, giving a total of eight datasets in the experiment: (1) P1, P2, P3, P4: 50 observations of each of the 9 activities for Person 1, 2, 3, 4 respectively. (2) P1R, P2R, P3R, P4R: 100 observations of random activities in addition to 50 observations of each of the 9 activities for Person 1, 2, 3, 4 respectively.

Table I. List of activities

| | |
|---|---|
| 1. | brushing teeth |
| 2. | cooking (chopping) |
| 3. | cooking (stirring) |
| 4. | relaxing on couch |
| 5. | still (standing) |
| 6. | talking on couch (sitting) |
| 7. | talking on the phone |
| 8. | working on computer |
| 9. | writing on whiteboard |
| 10. | random |

## 5.2 Experiment

We conduct the following experiment on each of the eight datasets described in Section 5.

1. Estimate number of clusters, i.e., find $k_s$.
   1.1. Run K-means from $k$=2 to $k_{max}$. There is no concrete guideline for the choice of $k_{max}$, however many researchers had referred to Mardia et al. [8] as stating the rule of thumb for setting $k = \sqrt[2]{n/2}$ where $n$ is the number of data points (observations) in the dataset. We have chosen $k_{max} = \sqrt[2]{n}$ which include the value of $k$ suggested by the said rule of thumb. The actual value of $k_{max}$ is not crucial as we will only accept a few highly ranked clusters. $k_{max}$ can be up to $n$. However, it helps to restrict the computation time by setting reasonable value of $k_{max}$.
   1.2. For each value of $k$, K-means was run for three rounds with random initialization (seeded for comparison with other values of $k$) and the clustering result with lowest total sum-of-squared-Euclidean distance from all members to their centroid was taken as the result.
   1.3. Compute the global cluster validity indices for all values of $k$ in above.
   1.4. Determine the suggested number of clusters, $k_s$, based on the global cluster validity test above.
2. Assess clustering quality (homogeneity) based on $k_s$ above. Compute local cluster validity measures for each of the $k_s$ clusters to accept or reject cluster(s).

# 6 Results & discussion

Table II shows the result of estimating number of clusters (value of $k_s$) for K-means using the five cluster validity indices for the eight datasets described in Section 5. For the purpose of comparison, we compute the overall error for each index as given in Eq. (19). The index with lowest error is considered giving best value of $k_s$.

$$Overall\ error\ e_t = e_{t(1)} + e_{t(2)} \tag{19}$$

where

$$e_{t(1)} = \left( \sum_{P \in \{P1,P2,P3,P4\}} (k_s^P - k_e^P)^2 \right)^{1/2} \tag{20}$$

$$e_{t(2)} = \left( \sum_{P \in \{P1R,P2R,PR3,P4R\}} (k_s^P - k_e^P)^2 \right)^{1/2} \tag{21}$$

where $k_s^P$ is the $k$ suggested, for dataset $P$, $k_e^P$ is the expected $k$ for dataset $P$, $e_{t(1)}$ is the total error for datasets without random activities, $e_{t(2)}$ is the total error for datasets with random activities.

Table II. Estimated number of clusters, ks, for K-means using DB, CH, KL, Ha and Sil indices for eight datasets.

| | DB | CH | KL | Ha | Sil |
|---|---|---|---|---|---|
| P1 | 6 | 8 | 8 | 8 | 6 |
| P2 | 6 | 6 | 4 | 14 | 6 |
| P3 | 6 | 10 | 20 | 10 | 9 |
| P4 | 9 | 9 | 20 | 12 | 9 |
| $e_{t(1)}$ | 5.2 | 3.3 | 16.4 | 6.0 | 4.2 |
| P1R | 5 | 3 | 17 | 11 | 5 |
| P2R | 17 | 3 | 11 | 11 | 4 |
| P3R | 10 | 3 | 16 | 12 | 10 |
| P4R | 3 | 3 | 17 | 10 | 3 |
| $e_{t(2)}$ | 13.7 | 20.0 | 6.7 | 4.2 | 15.9 |
| $e_t$ | 18.9 | 23.3 | 23.1 | 10.2 | 20.1 |

For datasets without random activities, $k_e^P$ is 9. For datasets with random activities, we expect $k_e^P$ to be more than 9, however we do not know the exact number. For the purpose of comparison, we have used the average value of all $k_s$ above 9 for datasets with random activities. The value is 12.9. While Sil, CH and DB did well on datasets without random activities, they performed poorly on datasets with random activities that have high variances. The result suggests that Hartigan index was the best choice among the five indices. The values of $k_s$ given by Hartigan index were used to cluster each dataset and the results are shown in Fig. 3 and 4. Fig. 3 shows the confusion matrices for the clustering result from the datasets without random activities. The rows are the nine activities as listed in Table I (1 to 9). The columns are the clusters. The columns are not merged and not sorted to provide a complete picture of the clustering result. Fig. 4 shows the confusion matrices for the clustering results from the datasets with random activities. Activity 10 is the group of random activities. Table III gives the homogeneity evaluation of the clusters for datasets without random activities, i.e., the results shown in Fig. 3. To explain the interpretation of the table, we take an example of the results for P1. The result shows that Cluster 3 has been ranked top by both evaluation measures. This means Cluster 3 is the most homogeneous according to these measures. Both measures had ranked Cluster 5 at second. Comparing this ranking with the corresponding confusion matrix for P1 in Fig. 3, we see that Cluster 3 and

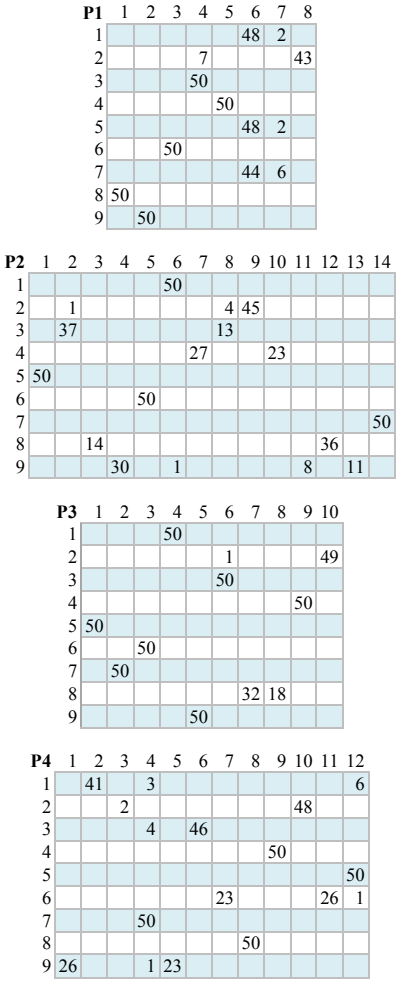


Figure 3. Confusion matrix without random activities for four subjects.

Table III. Mean variance ($\bar{\sigma}^2$) and mean joint probability ($\bar{P}_f$) for clustering result in Fig. 3, i.e., without random activities, for each subject

| P1 | $\bar{\sigma}^2$ | | $\bar{P}_f$ |
|---|---|---|---|
| 3 | 0.0087 | 3 | 347.19 |
| 5 | 0.0860 | 5 | 101.34 |
| 2 | 0.0883 | 1 | 63.40 |
| 1 | 0.129 | 2 | 57.23 |
| 6 | 0.223 | 8 | -2.88 |
| 8 | 0.230 | 6 | -29.87 |
| 4 | 0.298 | 4 | -91.74 |
| 7 | 0.410 | 7 | -124.28 |
| **P2** | $\bar{\sigma}^2$ | | $\bar{P}_f$ |
| 1 | 0.0570 | 14 | 127.46 |
| 14 | 0.0642 | 1 | 125.90 |
| 5 | 0.0935 | 6 | 34.74 |
| 6 | 0.119 | 12 | 31.77 |
| 10 | 0.122 | 5 | 20.92 |
| 12 | 0.132 | 7 | 3.77 |
| 7 | 0.136 | 2 | -7.01 |
| 2 | 0.145 | 10 | -12.34 |
| 3 | 0.206 | 3 | -33.63 |
| 13 | 0.333 | 11 | -43.07 |
| 4 | 0.336 | 13 | -117.00 |
| 11 | 0.338 | 4 | -123.47 |
| 9 | 0.362 | 8 | -148.66 |
| 8 | 0.452 | 9 | -150.45 |
| **P3** | $\bar{\sigma}^2$ | | $\bar{P}_f$ |
| 1 | 0.0122 | 1 | 373.36 |
| 9 | 0.0429 | 9 | 200.09 |
| 5 | 0.0582 | 5 | 143.98 |
| 7 | 0.0702 | 7 | 132.11 |
| 8 | 0.144 | 8 | 30.58 |
| 6 | 0.186 | 6 | -0.26 |
| 3 | 0.234 | 4 | -17.88 |
| 4 | 0.293 | 3 | -34.13 |
| 2 | 0.320 | 2 | -70.46 |
| 10 | 0.480 | 10 | -172.85 |
| **P4** | $\bar{\sigma}^2$ | | $\bar{P}_f$ |
| 11 | 0.0568 | 3 | 224.33 |
| 8 | 0.0786 | 11 | 162.60 |
| 7 | 0.114 | 8 | 103.68 |
| 6 | 0.117 | 6 | 72.56 |
| 2 | 0.174 | 7 | 30.75 |
| 1 | 0.179 | 2 | 16.10 |
| 5 | 0.199 | 5 | -37.77 |
| 10 | 0.249 | 1 | -38.34 |
| 9 | 0.251 | 9 | -43.79 |
| 4 | 0.275 | 10 | -64.10 |
| 3 | 0.358 | 4 | -92.64 |
| 12 | 0.396 | 12 | -160.00 |

5 concisely captured all 50 observations of Activity 6 (talking on couch) and 4 (relaxing on couch) respectively. In this confusion matrix, Cluster 4, 6 and 7 are weak clusters as they contained more than one activity. Incidentally, these clusters are ranked low in both measures in Table III. Looking at all results in Table III, we can observe that highly ranked clusters are generally homogeneous with a few exceptions. This will not be a problem if the objective is to select one or a few highly ranked clusters to perform learning. We also notice that $\bar{\sigma}^2$ ranked Cluster 3 in P4 low, while $\bar{P}_f$ ranked Cluster 3 on top. Cluster 3 is homogeneous, however, it has only two members; logically it should be rejected.

It will be more interesting to look at the results for datasets containing random activities in Table IV and Fig. 4. It is encouraging to see that clusters containing significant number of random activities have been consistently ranked at the bottom. For P1R, Cluster 3, 9 and 10 contain significant number of random activities and have been ranked at the bottom in corresponding section (P1R) in Table IV. For P2R, Cluster 7, 8, 9 and 11 are ranked at the bottom. Cluster 2 should be rejected as well, and has been ranked low just above those clusters with random activities. For P3R, Cluster 4, 5, 8, and 12 are ranked at the bottom, while Cluster 3 come slightly above them. Ideally, Cluster 3 should be ranked low as it should be rejected. Nevertheless, clusters ranked on top remain correctly identified from homogenous clusters. For P4R, Cluster 3, 5, 6, 7 and 9 are ranked at the bottom. In all datasets, the first few clusters on top are homogenous, apart from Cluster 8 in P4R. However, for Cluster 8 in P4R, having 2 random activities with 50 observations of Activity 3 (stirring) is not expected to cause problem in learning phase.

Referring to Table IV, for each of the homogeneity measure, is there any threshold we can use to identify homogeneous clusters confidently, i.e., to accept the clusters? We find the value for each measure in each dataset where the first cluster from the top is to be rejected. For P1R, the highest cluster to reject is Cluster 2 and the corresponding values for each measure are: $\bar{\sigma}^2$ less than 0.0881 and $\bar{P}_f$ higher than 13.98. For P2R, the highest cluster to reject is Cluster 2 and the corresponding values for each measure are: $\bar{\sigma}^2$ less than 0.268 and $\bar{P}_f$ higher than −112. For P3R, the highest cluster to reject is Cluster 3 and the corresponding values for each measure are: $\bar{\sigma}^2$ less than 0.255 and $\bar{P}_f$ higher than −139.15. For P4R, the highest cluster to reject is Cluster 5 and the corresponding values for each measure are: $\bar{\sigma}^2$ less than 0.458 and $\bar{P}_f$ higher than −183. To obtain a threshold value applicable to all subjects so that non-homogeneous clusters will be rejected, we require that $\bar{\sigma}^2$ to be less than 0.0881 (lowest of all) and $\bar{P}_f$ to be higher than 13.98 (highest of all). Applying these threshold to all datasets in Table IV, we identified the number of top few homogeneous clusters. The result is shown in Table V. The $\bar{P}_f$ measure has identified more clusters to accept.

**P1R**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | 50 | | | | | | | | | |
| 2 | | | 50 | | | | | | | | |
| 3 | | | 50 | | | | | | | | |
| 4 | | | | | | | | | | | 50 |
| 5 | | | | | | 50 | | | | | |
| 6 | | | | 50 | | | | | | | |
| 7 | | 50 | | | | | | | | | |
| 8 | | | | | | | 27 | 23 | | | |
| 9 | 24 | | | | 26 | | | | | | |
| 10 | | | 20 | | | 40 | | | 12 | 28 | |

**P2R**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | 50 | | | | | |
| 2 | | 50 | | | | | | | | | |
| 3 | | 41 | | | | | | | 9 | | |
| 4 | | | | | 50 | | | | | | |
| 5 | | | | | | | | 50 | | | |
| 6 | | | | | | | | | | 50 | |
| 7 | 50 | | | | | | | | | | |
| 8 | | | 50 | | | | | | | | |
| 9 | | | | 45 | | 5 | | | | | |
| 10 | | 1 | | | | 1 | 35 | 25 | 13 | | 25 |

**P3R**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | 50 | |
| 2 | | | | 50 | | | | | | | | |
| 3 | | | | 50 | | | | | | | | |
| 4 | | | | | | 50 | | | | | | |
| 5 | | | 50 | | | | | | | | | |
| 6 | | | | | | | | | 50 | | | |
| 7 | | | | | 50 | | | | | | | |
| 8 | 32 | | | | | | | 18 | | | | |
| 9 | | 50 | | | | | | | | | | |
| 10 | | | 24 | 5 | 21 | 3 | | 22 | 1 | | 1 | 23 |

**P4R**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | | | | | 47 | | | | |
| 2 | | 50 | | | | | | | | |
| 3 | | | | | | | | 50 | | |
| 4 | | | | | 50 | | | | | |
| 5 | | | 50 | | | | | | | |
| 6 | | | 1 | | 49 | | | | | |
| 7 | 50 | | | | | | | | | |
| 8 | | | | | | | | | | 50 |
| 9 | 1 | | | 49 | | | | | | |
| 10 | | | 38 | | | 6 | 28 | 2 | 26 | |

Figure 4. Confusion matrix with random activities for four subjects.

# 7 Conclusion

In this paper, we proposed an approach to automatically discover new activities without priori. We demonstrated the feasibility of the approach through experimental investigation on daily activity datasets from third party. The results showed that Hartigan index could assist to estimate the possible number of clusters, or $k$ value, in a pool of unlabeled observations of activities for each subject. Given the estimation of $k$, $k$ clusters were obtained using K-means. The clustering outcomes were assessed using two measures of cluster homogeneity. The results showed both measures consistently ranked highly homogeneous clusters on top while ranking clusters with significant number of random activities at the bottom. It was also observed that

the mean joint probability density function $\bar{P}_f$ measure has the potential to use a threshold value to assist in deciding how many highly homogeneous clusters can be accepted for subsequent learning phase. This is an ability to discover new activities (clusters) autonomously. The model of each of the discovered new activities (clusters) can then be learned using supervised learning algorithm. Armed with this ability, an intelligent system can self-learn and perform unsupervised human activity recognition.

Table IV. Mean variance ($\bar{\sigma}^2$) and mean joint probability ($\bar{P}_f$) for clustering result in Fig. 4, i.e., with random activities, for each subject.

| **P1R** | $\bar{\sigma}^2$ | | $\bar{P}_f$ |
|---|---|---|---|
| 4 | 0.00568 | 4 | 393.91 |
| 5 | 0.0374 | 5 | 157.31 |
| 7 | 0.0548 | 11 | 148.07 |
| 1 | 0.0623 | 7 | 142.36 |
| 8 | 0.0733 | 8 | 111.28 |
| 11 | 0.0768 | 1 | 104.60 |
| 2 | 0.0881 | 2 | 13.98 |
| 6 | 0.537 | 6 | -236.60 |
| 3 | 0.599 | 3 | -258.12 |
| 9 | 1.04 | 9 | -274.26 |
| 10 | 1.20 | 10 | -344.00 |
| **P2R** | $\bar{\sigma}^2$ | | $\bar{P}_f$ |
| 1 | 0.0341 | 1 | 180.34 |
| 10 | 0.0651 | 10 | 73.80 |
| 5 | 0.113 | 5 | 15.84 |
| 6 | 0.159 | 6 | -46.06 |
| 4 | 0.207 | 3 | -46.25 |
| 3 | 0.240 | 4 | -96.32 |
| 2 | 0.268 | 2 | -112.31 |
| 8 | 0.313 | 8 | -185.44 |
| 9 | 0.939 | 9 | -295.41 |
| 11 | 1.15 | 11 | -342.36 |
| 7 | 1.26 | 7 | -352.73 |
| **P3R** | $\bar{\sigma}^2$ | | $\bar{P}_f$ |
| 7 | 0.0249 | 7 | 258.57 |
| 2 | 0.0336 | 2 | 202.46 |
| 1 | 0.0404 | 1 | 190.59 |
| 10 | 0.144 | 10 | 24.35 |
| 9 | 0.225 | 11 | -6.07 |
| 11 | 0.252 | 9 | -115.99 |
| 3 | 0.255 | 6 | -133.27 |
| 6 | 0.291 | 3 | -139.15 |
| 4 | 0.308 | 4 | -161.39 |
| 12 | 1.12 | 12 | -317.39 |
| 8 | 1.18 | 8 | -333.79 |
| 5 | 1.35 | 5 | -361.30 |
| **P4R** | $\bar{\sigma}^2$ | | $\bar{P}_f$ |
| 10 | 0.0482 | 10 | 157.73 |
| 4 | 0.131 | 1 | -0.51 |
| 8 | 0.135 | 4 | -11.41 |
| 1 | 0.151 | 8 | -13.94 |
| 2 | 0.159 | 2 | -23.20 |
| 6 | 0.367 | 6 | -167.71 |
| 5 | 0.458 | 5 | -182.96 |
| 3 | 0.628 | 3 | -267.60 |
| 7 | 1.14 | 7 | -325.54 |
| 9 | 1.32 | 9 | -357.62 |

Table V. Number of accepted clusters based on threshold value for each measure.

| | Number of accepted clusters | |
|---|---|---|
| | $\bar{\sigma}^2 < 0.0881$ | $\bar{P}_f > 13.98$ |
| P1R | 6 | 7 |
| P2R | 2 | 3 |
| P3R | 3 | 4 |
| P4R | 1 | 1 |

# 8 References


[1] D. L. Davies, and D. W. Bouldin, D. W, "A cluster separation measure," in Pattern Analysis and Machine Intelligence, IEEE Transactions on, (2), 224-227, 1979.

[2] D. Wyatt, M. Philipose, and T. Choudhury, "Unsupervised activity recognition using automatically mined common sense," in Proceedings of the National Conference on Artificial Intelligence, July 2005, Vol. 20, No. 1, p. 21.

[3] J. A. Hartigan, "Clustering algorithms," John Wiley & Sons, Inc., 1975.

[4] J. K. Aggarwal, and M.S. Ryoo, "Human activity analysis: A review," in ACM Comput. Surv. 43, 3, Article 16, April 2011.

[5] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, June 1967, Vol. 1, No. 281-297, p. 14.

[6] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Human Activity Detection from RGBD Images," in Association for the Advancedment of Artificial Intelligence Workshop on Pattern, Activity and Intent Recognition (PAIR), 2011, pp. 47-55.

[7] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, June 2012, pp. 1290-1297.

[8] K. V. Mardia, J. T. Kent, and J. M. Bibby, "Multivariate analysis," 1980, pp. 365.

[9] M. Stikic, D. Larlus, S. Ebert, and B. Schiele, "Weakly Supervised Recognition of Daily Life Activities with Wearbale Sensors," in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 12, December 2011, pp. 2521-2537.

[10] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," in Journal of computational and applied mathematics, 1987, 20, 53-65.

[11] T. Caliński, and J. Harabasz, "A dendrite method for cluster analysis," in Communications in Statistics-theory and Methods, 1974, 3(1), 1-27.

[12] T. Huynh, M. Fritz, and B. Schiele, "Discovery of Activity Patterns using Topic Models," in UbiComp '08 Proceedings of the 10th International Conference on Ubiquitous Computing, 2088, pp. 10-19.

[13] U. Maulik, and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," inPattern Analysis and Machine Intelligence, IEEE Transactions on, 2002, 24(12), 1650-1654.

[14] W. J. Krzanowski, and Y. T. Lai " A criterion for determining the number of groups in a data set using sum-of-squares clustering," Biometrics, 1988, 23-34.

[15] W. Ong, L. Palafox, and T. Koseki, "Investigation of Feature Extraction for Unsupervised Learning in Human Activity Detection," in Bulletin of Networking, Computing, Systems, and Software, North America, 2, jan. 2013.

[16] W. Ong, T. Koseki, and L. Palafox, "Unsupervised Activity Detection with Skeleton Data From RGB-D Sensor," to be published in Computational Intelligence, Communication Systems and Networks (CICSyN), 2013 Fifth International Conference on IEEE, 2013, in press.

[17] Y. Song, L. Goncalves, and P. Perona, "Unsupervised learning of human motion," in Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2003, 25(7), 814-827.