**Problem statement**:

- To build a recognition model from raw sensor readings to high-level activities, it mainly consists of two steps.
    - Segment continuous streaming sensor readings automatically or manually (Yin et al. 2005; Janidarmian et al. 2017).
        - Each segment contains sensor readings received from a set of sensors in a specific period of various lengths, and is supposed to correspond to one activity category.
    - Learn a predictive model to map each segment to its corresponding activity label.

- Key words**: Higher-order, Various (1-2) , Hierarchical (3-4), Intrinsic (5-8)**

| | Key words | Title | Year | Common | Pros | Cons |
|---|---|---|---|---|---|---|
| 1 | Higher-Order | Sensor-based activity recognition via learning from distributions | AAAI 18 | 1. Statistical feature, e.g. mean, variance, etc is very important.<br>2.Low order moment can not distinguish different activities.<br>3. Higher order is better but how to decide suitable orders ?<br>4. Mapping to Reproducing Kernel Hilbert Space by kernel embedding | RKHS is a high-dimensional or even infinite-dimensional feature space, which is able to capture any order of moments of the probability distribution. The mean map operations in such higher order can distinguish the feature effecitvely | Use Fixed Gaussian Kernel, hence parameter tuning of proper bandwidth for kernel is required in advance |
| 2 | | A Novel Distribution-Embedded Neural Network for Sensor-Based Activity Recognition | IJCAI 19 | | Inspired by above, The Kernal embedding process is replaced by a Neural Network f(.). Meanwhile, a decoder is implemented to make the NN more injective. Instead of the Single Fixed Kernel, multiple kernels are combined together. Temporal and spatial information are also involved,. | More computational, Only the mean operation is involved |
| 3 | Hierarchical | Deep activity recognition models with triaxial accelerometers. | AAAI 16 | The Deep Generative model is implemented to extracting hierarchical features, which can analyze the inner relationship between visible (input) and hidden state. | This is a prior work that implement the generative model in sensor feature extraction. The model is DBNs and the hidden units are formed from RBMs. After training, the hidden states can represent the visible layer accurately and reconstruct the visible in the meanwhile. | The input of this work are 2d spectrum img. However, the DBNs cannot form the 2D structure because of the input is vectoral form. Also, the DBNs does not deal with temporal information clearly. |
| 4 | | Locomotion activity recognition using stacked denoising autoencoders | IoT 18 | | This work is similar as above. They use Stacked Denoising Autoencoders as the hidden units to replace the RBMs in DBNs. This method can make the neural network deeper and maintain the same training criteria. SD-Auto-EC is also good at noise reduction which make the extracted feature more robustness. | Compared with DBNs, stacked denoising autoencoders are discrimination model, which means it is difficult to sample the input space . The feature may not be intrinsic as generative model. The temporal information are poorly involved |
| 5 | Latent representation | Motion2Vector: unsupervised learning in human activity recognition using wrist-sensing data | Ubicom 19 | Unsupervised learning method with auto-encoder to embed the input to a multidimensional space, as a feature representation of a certain activity type. | The variational autoeccoder composed of Bidirectional LSTM layer can effectively embed the temporal information of the input. | They directly implement the VAE to the data embedding. The size of hidden vector is 128 which highly constrain the time period of the movement. The segmentation is very important |
| 6 | | Deep Auto-Set: A Deep Auto-Encoder-Set Network for Activity Recognition Using Wearables | MobiQ 18 | | The work implement a normal AutoEC to extract the hidden feature. The extracted hidden feature is used to distinguish the multi-activities in a same Window. | The temporal information is poorly involved. The ratio of muti-activities in a same window is highly influence the results |
| 7 | Intrinsic | Distributionally Robust Semi-Supervised Learning for People-Centric Sensing | AAAI 19 | These two work both aim at removing the person-centric information and maintain the task-content information. Hierarchical structure are implemented to solve the problems. | They use different auto-ec extract the task-specific consistency and remove the person-specific discrepancy. The adversarial loss is implemeted. | Assume the labeled data and unlabeled are in a different distribution. The shift problem are not well defined in Sensor errors. They do not distinguish the wearable error and user-centric error |
| 8 | | Collective Protection: Preventing Sensitive Inferences via Integrative Transformation | ICDM 19 | | Transform the data into a "random noise" style which can protect the sensitive information. | The wearable error will influence the performance. High computation. It is difficult to balance the content loss and user-privacy loss. |

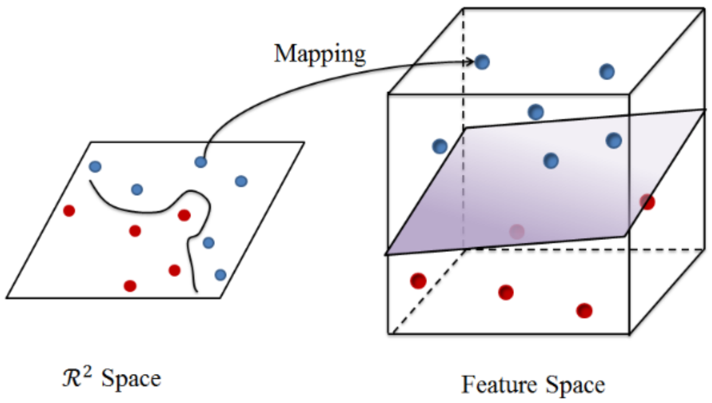
Hangwei Qian, Sinno Jialin Pan, and Chunyan Miao. Sensor-based activity recognition via learning from distributions. In AAAI, 2018.

- In the past, feature extraction approaches is to calculate statistical metrics, e.g., **mean, variance**, etc. one needs to **predefine** what statistical metrics
- Using predefined orders of moments ,e.g., statistical information captured by higher-order moments, may be discarded when constructing features.
- We consider sensor readings received within a period as a sample, which can be represented by a feature vector of infinite dimensions in a **Reproducing Kernel Hilbert Space (RKHS)** using **kernel embedding** techniques.

A RKHS is a **high-dimensional** or even **infinite-dimensional feature space**, which is able to capture **any order of moments of the probability distribution** from which the sample is drawn.

$$k(x, x') = \exp(-\gamma \|x - x'\|^2)$$

with fixed ***γ***, parameter tuning of proper bandwidth for kernel is required in advance.




$$\boldsymbol{\mu}_i = \frac{1}{n_i} \sum_{p=1}^{n_i} \phi(\mathbf{x}_{ip}).$$ mean map operations {(μ1, y1), ..., (μn, yn)}

Then our goal is to learn a classifier ***f*** **:** *H* -> *H˜* such that f(μi) = yi

$$f = \sum_{i=1}^{n} \alpha_i \psi(\boldsymbol{\mu}_i),$$

$$f = \sum_{i=1}^{n} \alpha_i \boldsymbol{\mu}_i, \text{ where } \alpha_i \in \mathbb{R}.$$

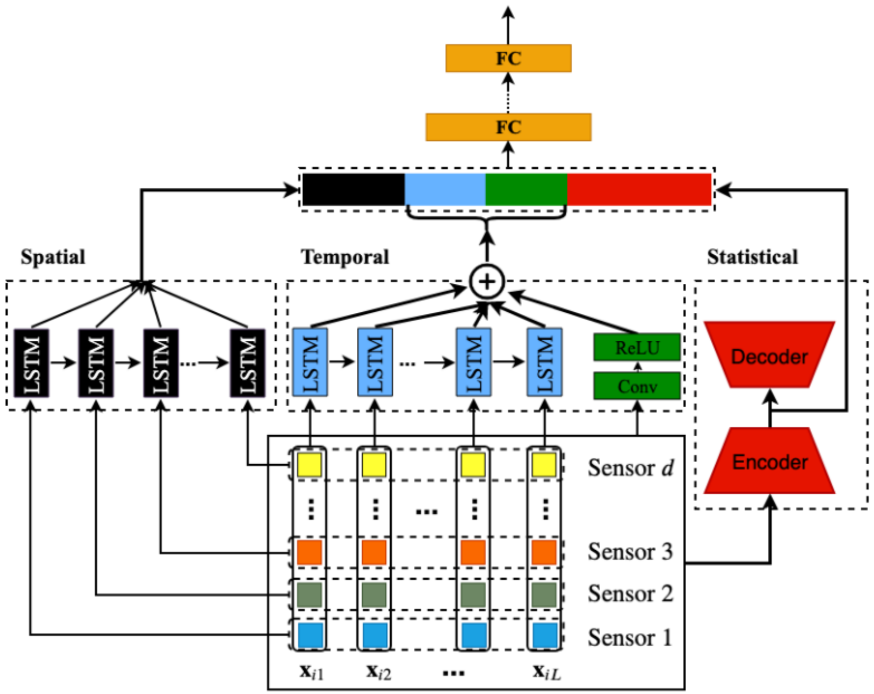
$$\begin{aligned} \min_f \quad & \frac{1}{2}\|f\|^2_{\tilde{\mathcal{H}}} + C\sum_{i=1}^{n} \xi_i, \\ s.t. \quad & y_i f(\boldsymbol{\mu}_i) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \\ & 1 \leq i \leq n, \end{aligned}$$

21. Hangwei Qian, Sinno Jialin Pan, Bingshui Da, and Chunyan Miao. 2019. A Novel Distribution-Embedded Neural Network for Sensor-Based Activity Recognition. In Proceedings ofthe Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019. 5614–5620

Fixed Gaussian kernel

$$k(x, x') = \exp(-\gamma \|x - x'\|^2)$$

with fixed $\gamma$, parameter tuning of proper bandwidth for kernel is required in advance.




Design a neural network f1 to learn the statistical feature mapping φf1(·) automatically

$$f_1(\mathbf{X}_i) = \phi_{f_1}(\mathbf{X}_i).$$

$$\phi_{f_1}(\mathbf{X}_i) = \frac{1}{L}\sum_{j=1}^{L} \phi_k(\mathbf{x}_{ij}).$$

To learn the best kernel automatically from different possible characteristic kernels **k** ∈ *K*.

$$f_1^*(\mathbf{X}_i) = \max_{f_1} \phi_{f_1}(\mathbf{X}_i) = \max_{k \in \mathcal{K}} \frac{1}{L}\sum_{j=1}^{L} \phi_k(\mathbf{x}_{ij}).$$

Autoencoder to guarantee the injectivity of the feature mapping

The encoder is the desired f1 module, and fd = f−1

reconstruction error $\ell_{ae} = \|x - f_d(f_e(\tilde{x}))\|$

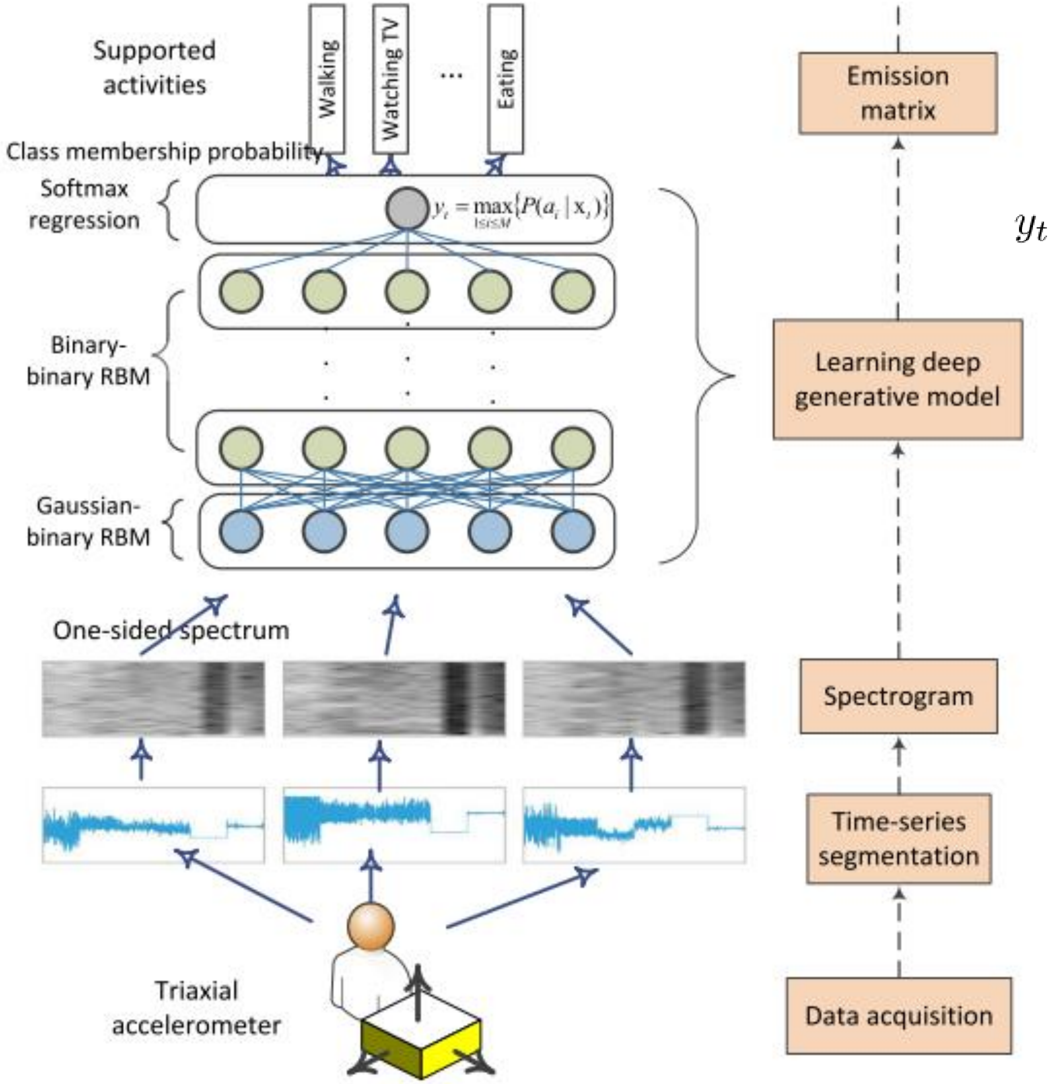
$$\ell_{\mathrm{MMD}}(\mathbf{X}_i, f_d(f_e(\mathbf{X}_i))) = \frac{1}{L}\left\|\sum_{j=1}^{L} f_e(\mathbf{x}_{ij}) - f_e(f_d(f_e(\mathbf{x}_{ij})))\right\|_2$$

21. Hangwei Qian, Sinno Jialin Pan, Bingshui Da, and Chunyan Miao. 2019. A Novel Distribution-Embedded Neural Network for Sensor-Based Activity Recognition. In Proceedings ofthe Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019. 5614–5620

| | DG | | OPPOR | | UCIHAR | | PAMAP2 | |
|---|---|---|---|---|---|---|---|---|
| Methods | miF | maF | miF | maF | miF | maF | miF | maF |
| DDNN | **92.59** | **91.61** | **83.66** | 86.01 | **90.53** | **90.58** | **93.23** | **93.38** |
| DDNN$-f_1$ | 91.38 | 90.67 | 81.27 | 84.51 | 89.96 | 89.93 | 87.49 | 86.84 |
| DDNN$-f_2$ | 89.67 | 88.97 | 77.96 | 82.27 | 88.60 | 88.58 | 89.37 | 89.43 |
| CNN_Yang | 87.96 | 86.65 | 9.98 | 2.95 | 88.12 | 88.11 | 70.17 | 70.46 |
| DeepConvLSTM | 87.21 | 84.28 | 75.47 | 78.92 | 89.05 | 89.07 | 84.31 | 82.73 |
| DNN | 88.91 | 86.47 | 77.05 | 80.25 | 87.65 | 87.72 | 80.31 | 79.82 |
| CNN | 89.23 | 88.85 | 10.66 | 3.56 | 86.66 | 86.77 | 89.75 | 89.72 |
| LSTM | 88.34 | 86.93 | 63.17 | 69.92 | 74.52 | 74.75 | 90.38 | 90.29 |
| LSTM-f$*$ | 67.3 | - | 67.2 | 90.8 | - | - | 92.9 | - |
| LSTM-S$*$ | 76.0 | - | 69.8 | 91.2 | - | - | 88.2 | - |
| b-LSTM-S$*$ | 74.1 | - | 74.5 | **92.7** | - | - | 86.8 | - |

1. Mohammad Abu Alsheikh, Ahmed Selim, Dusit Niyato, Linda Doyle, Shaowei Lin, and Hwee-Pink Tan. 2016. Deep activity recognition models with triaxial accelerometers. In Workshops at the Thirtieth AAAI Conference on Artificial Intelligence

## Extracting hierarchical features from triaxial acceleration data




$$y_t = \arg\max_{1\le i\le M}\left\{P\left(a_i|\mathbf{x}_t,\theta\right)\right\}.$$ $\mathcal{A} = \{a_1, a_2, \ldots, a_M\}$ be a finite activity space

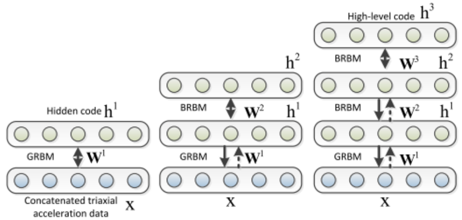
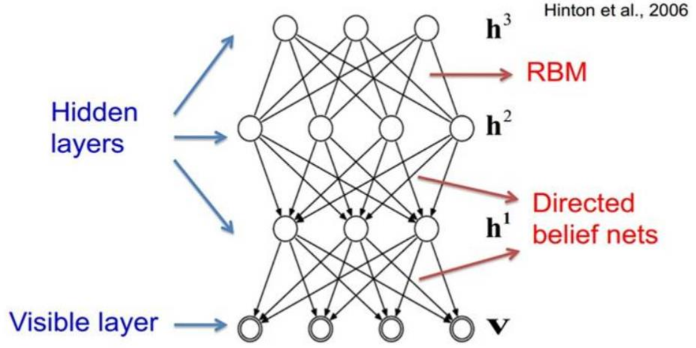



Figure 2: **The greedy layer-wise training of DBNs**. The first level is trained on triaxial acceleration data. Then, more RBMs are repeatedly stacked to form a deep activity recognition model until forming a high-level representation.

DBN structure

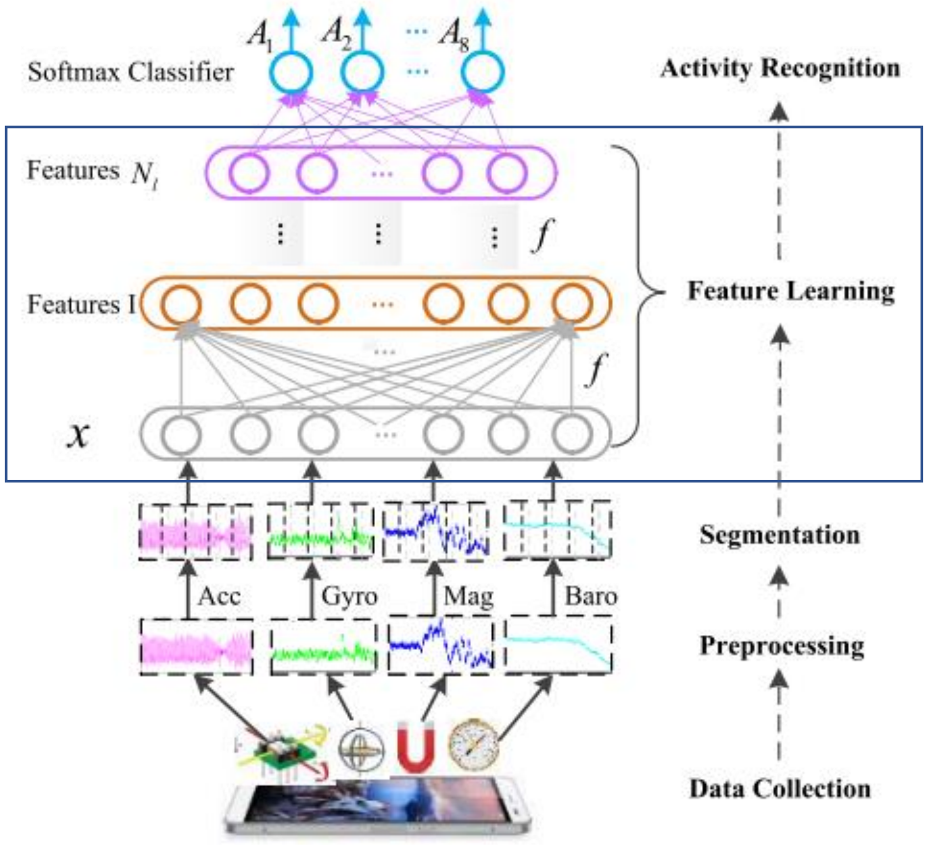



$$P(\mathbf{v},\mathbf{h}^1,\mathbf{h}^2,\ldots,\mathbf{h}^l) = P(\mathbf{v}\mid\mathbf{h}^1)P(\mathbf{h}^1\mid\mathbf{h}^2)\ldots P(\mathbf{h}^{l-2}\mid\mathbf{h}^{l-1})P(\mathbf{h}^{l-1},\mathbf{h}^l)$$

$$\begin{aligned}
\mathbf{s}_t^x &= [\ r_t^x \ \cdots \ r_{t+N-1}^x\ ]^T, && (2)\\
\mathbf{s}_t^x &= [\ r_t^y \ \cdots \ r_{t+N-1}^y\ ]^T, && (3)\\
\mathbf{s}_t^z &= [\ r_t^z \ \cdots \ r_{t+N-1}^z\ ]^T. && (4)
\end{aligned}$$

7. Fuqiang Gu, Kourosh Khoshelham, Shahrokh Valaee, Jianga Shang, and Rui Zhang. 2018. Locomotion activity recognition using stacked denoising autoencoders. IEEE Internet of Things Journal 5, 3 (2018), 2085–2093.

- Stacked autoencoders is a commonly used feature learning method, which is capable of learning useful features in an unsupervised manner




Sparsity parameter 0.05

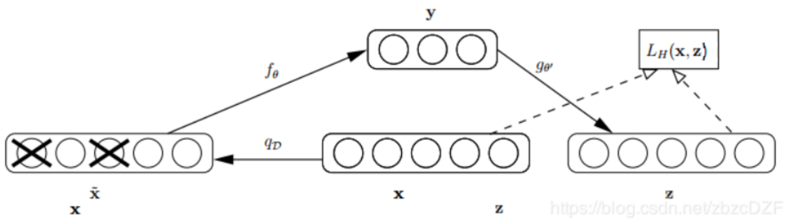
Average activation of hidden unit j

$$J_{\text{dae}} = L(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}_i) + \beta \sum_{j=1}^{N} KL\left(\rho \middle\| \hat{\rho}_j\right)$$

sparsity penalty

稀疏约束

$$L(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}_i) = \frac{1}{2} \sum_{j=1}^{M} \left\| x_j - \widehat{x}_j \right\|^2.$$




$$\boldsymbol{x}_i' = q(\boldsymbol{x}_i).$$

stochastic mapping

forcing a fraction ν of xi to be 0

$$\boldsymbol{x}_i = \left[ \boldsymbol{s}_i^{\text{acc}_x}, \boldsymbol{s}_i^{\text{acc}_y}, \boldsymbol{s}_i^{\text{acc}_z}, \boldsymbol{s}_i^{\text{gyro}_x}, \boldsymbol{s}_i^{\text{gyro}_y}, \boldsymbol{s}_i^{\text{gyro}_z}, \boldsymbol{s}_i^{\text{mag}_x}, \boldsymbol{s}_i^{\text{mag}_y}, \boldsymbol{s}_i^{\text{mag}_z}, \boldsymbol{s}_i^{\text{pres}} \right]^T \quad (4)$$

where $\boldsymbol{x}_i$ is an $M \times 1$ vector ($M$ equals to 608 in this paper).

2. Lu Bai, Chris Yeung, Christos Efstratiou, and Moyra Chikomo. 2019. Motion2Vector: unsupervised learning in human activity recognition using wrist-sensing data. In Proceedings ofthe 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings ofthe 2019ACM International Symposium on Wearable Computers. ACM, 537–542.

- Motion2Vector -- The model is trained using large amounts of unlabelled human activity data to learn a representation of a time period of activity data.

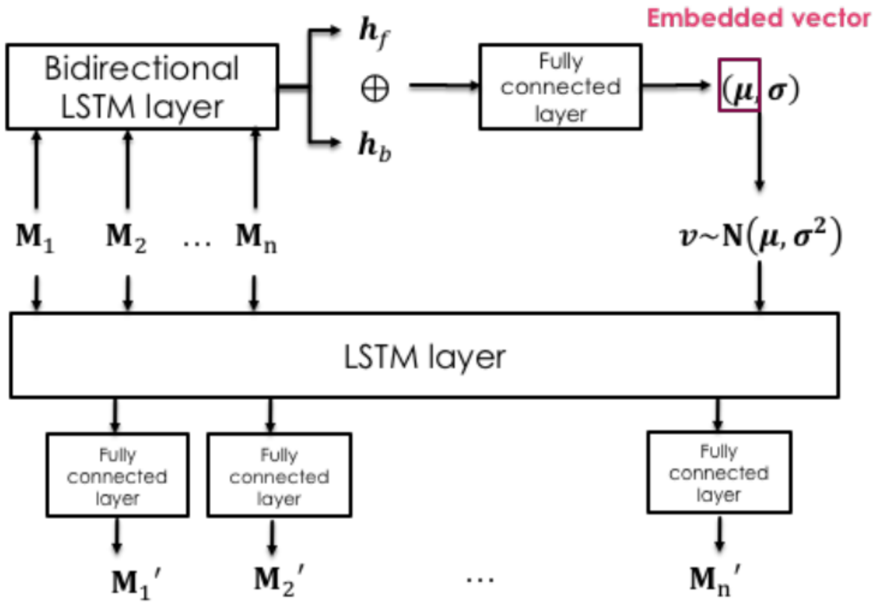



**Figure 2: bidirectional LSTM**

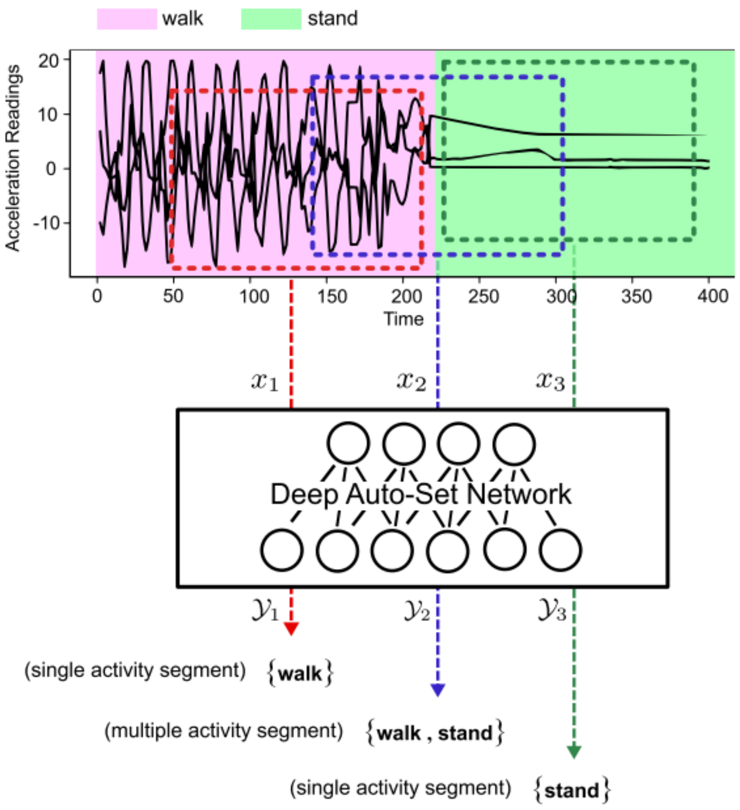
The embedded vector $\boldsymbol{\mu}$ is target in this study, it is a 128 dim vector （an appropriate representation of the input movement within a particular time period）.

$$L_c = \frac{1}{N}\sum_{i=1}^{N}(M' - M)^2 \tag{1}$$

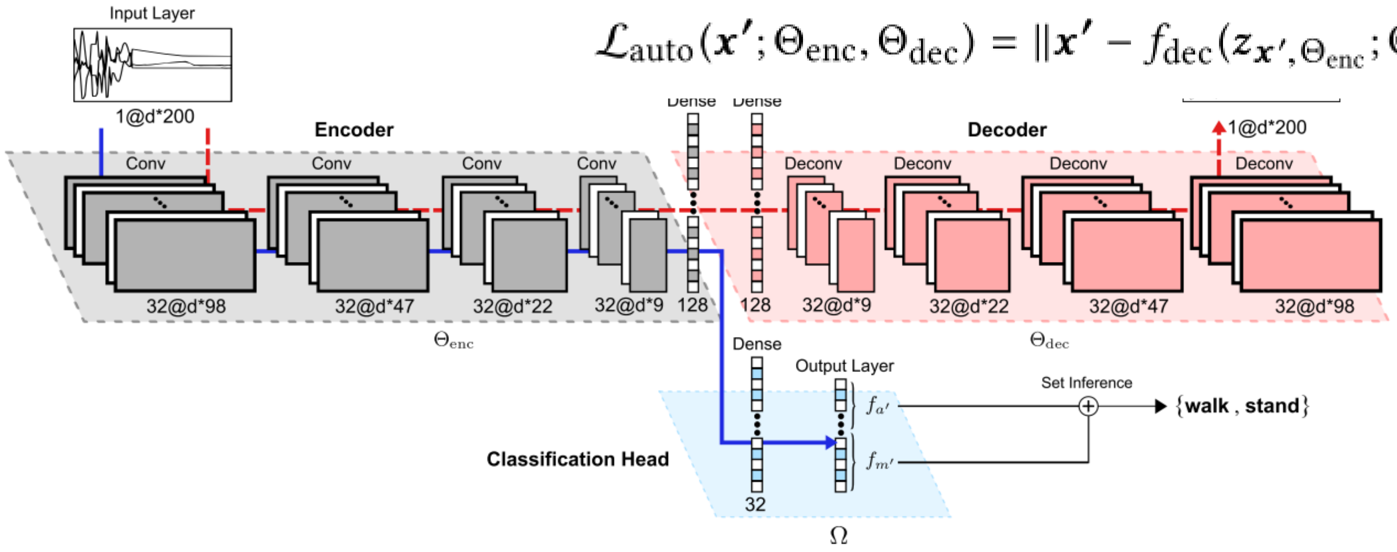
$$L_{KL} = -\frac{1}{2N}\sum_{i=1}^{N}(1 + ln\sigma_i - \mu_i^2 - \sigma_i) \tag{2}$$

$$L = L_c + \gamma L_{KL} \tag{3}$$

26. Alireza Abedin Varamin, Ehsan Abbasnejad, Qinfeng Shi, Damith C Ranasinghe, and Hamid Rezatofighi. 2018. Deep Auto-Set: A Deep Auto-Encoder-Set Network for Activity Recognition Using Wearables. In Proceedings ofthe 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. ACM, 246–253




$$\mathcal{L}_{\text{auto}}(\boldsymbol{x}'; \Theta_{\text{enc}}, \Theta_{\text{dec}}) = \|\boldsymbol{x}' - f_{\text{dec}}(\boldsymbol{z}_{\boldsymbol{x}', \Theta_{\text{enc}}}; \Theta_{\text{dec}})\|^2.$$




i) a set cardinality term fm′(x) with log softmax activation which produces cardinality scores;
ii) a set element term fa′(x) with sigmoid activation which produces scores for the (activity types).

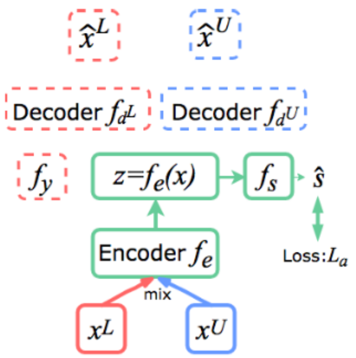
$$\mathcal{L}_{\text{set}}(\boldsymbol{x}, \mathcal{Y}^m; \Theta_{\text{enc}}, \Omega) = \sum_{a \in \mathcal{Y}} \ell_{\text{bce}}(a, f_{a'}(\boldsymbol{x}; \Theta_{\text{enc}}, \Omega)) + \ell_{\text{nll}}(m, f_{m'}(\boldsymbol{x}; \Theta_{\text{enc}}, \Omega)), \tag{4}$$

4. Kaixuan Chen, Lina Yao, Dalin Zhang, Xiaojun Chang, Guodong Long, and Sen Wang. 2019. Distributionally Robust Semi-Supervised Learning for People-Centric Sensing. In The Thirty-Third AAAIConference on Artificial Intelligence, AAAI, Honolulu, Hawaii USA, January 27 February 1, 2019. 3321–3328.

- Distribution shift in semi-supervised learning due to the **diverse biological conditions and behavior patterns of humans**

## Reducing Person-Specific Discrepancy

Force the feature extractor $\boldsymbol{f_e}$ to map $X_L$ and $X_U$ to a **unified distribution** which cannot be distinguished by the distribution classifier



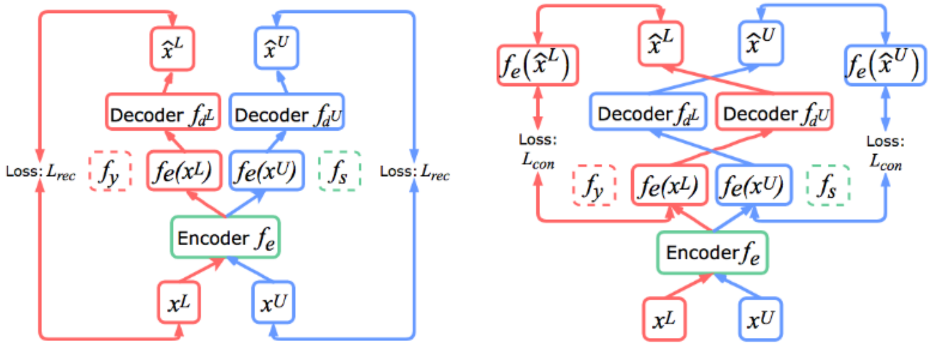

(a) Person-Specific Discrepancy

$$L_a = \frac{1}{N^L}\sum_{n=1}^{N^L} \log f_s(f_e(x_n^L)) + \frac{1}{N^U}\sum_{n=1}^{N^U} \log(1 - f_s(f_e(x_n^U)))$$

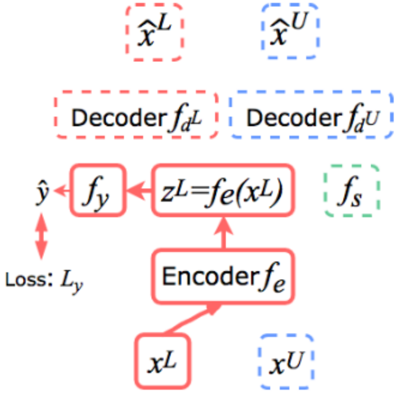
$$\min_{\theta_e}\max_{\theta_s}[L_a(x^L, x^U, \theta_e, \theta_s)]$$

## Preserving Task-Specific Consistency

fdL and fdU are able to reconstruct input vectors ˆx from the corresponding latent features z. Therefore, we generate ˆxU with xL: ˆxU = fdU (fe(xL)), and similarly for the reverse: ˆxL = fdL(fe(xU)).




(b) Task-Specific Consistency



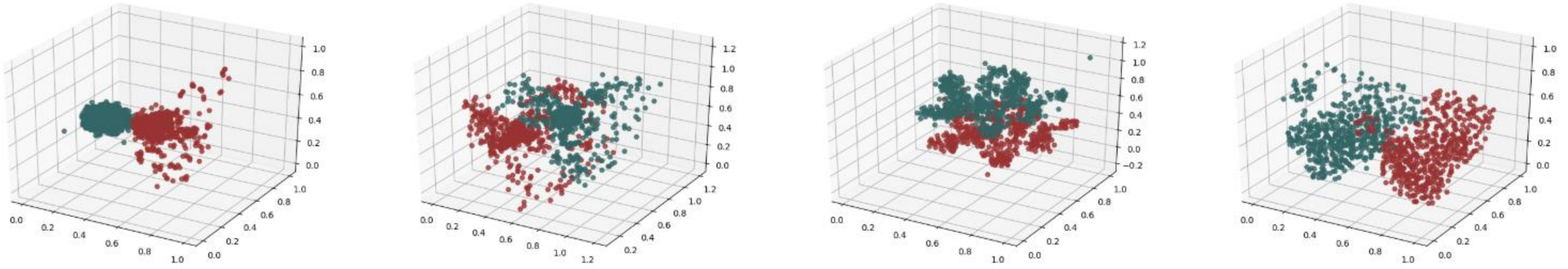

(c) Prediction

$$L_{rec} = \frac{1}{N^L}\sum_{n=1}^{N^L} \| x_n^L - f_{d^L}(f_e(x_n^L)) \|^2 + \frac{1}{N^U}\sum_{n=1}^{N^U} \| x_n^U - f_{d^U}(f_e(x_n^U)) \|^2, \quad (4)$$

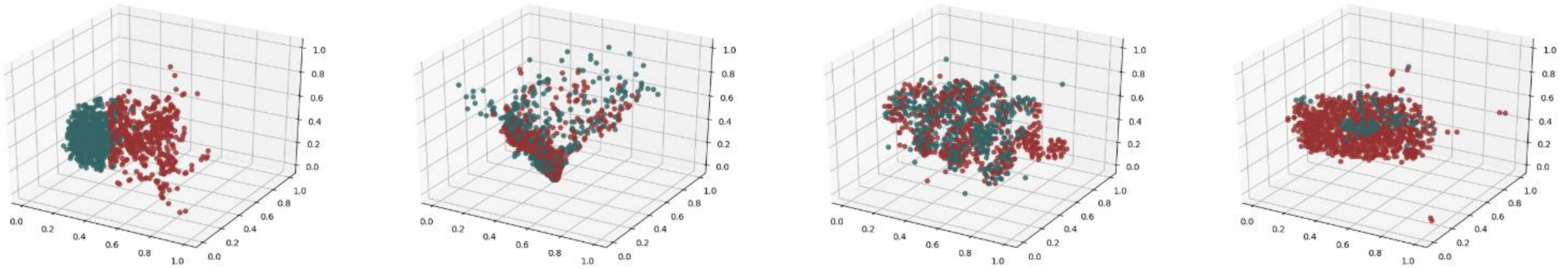
$$L_{con} = \frac{1}{N^L}\sum_{n=1}^{N^L} \| f_e(x_n^L) - f_e(f_{d^U}(f_e(x_n^L))) \|^2 + \frac{1}{N^U}\sum_{n=1}^{N^U} \| f_e(x_n^U) - f_e(f_{d^L}(f_e(x_n^U))) \|^2 \quad (5)$$

$$L_y = -\frac{1}{N^L}\sum_{n=1}^{N^L}\sum_{m=1}^{M} y_n(m)\log f_y(f_e(x_n^L)), \quad (6)$$

4. Kaixuan Chen, Lina Yao, Dalin Zhang, Xiaojun Chang, Guodong Long, and Sen Wang. 2019. Distributionally Robust Semi-Supervised Learning for People-Centric Sensing. In The Thirty-Third AAAIConference on Artificial Intelligence, AAAI, Honolulu, Hawaii USA, January 27 February 1, 2019. 3321–3328.



(a) EEG raw (b) MHEALTH raw (c) EMG raw (d) OPPORTUNITY raw



(e) EEG features (f) MHEALTH features (g) EMG features (h) OPPORTUNITY features

34. Dalin Zhang, Lina Yao, Kaixuan Chen, Guodong Long, and SenWang. 2019. Collective Protection: Preventing Sensitive Inferences via Integrative Transformation. In The 19th IEEE International Conference on Data Mining (ICDM). IEEE, 1–6

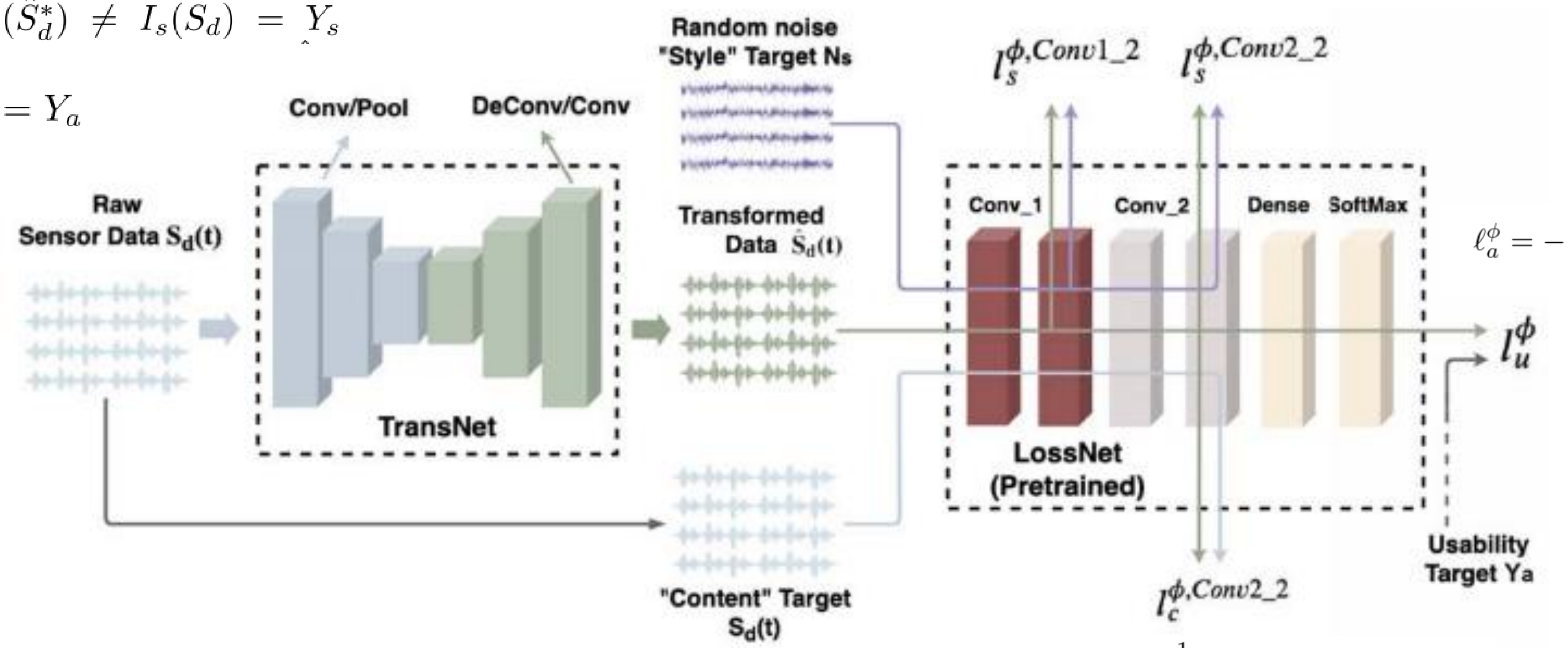
- Transforms raw sensor data into a new format that has a “style” (sensitive information) of random noise and a “content” (desired information) of the raw sensor data
- Transforming the data “style” comparable to random noise but keeping the data “content” exact as raw data

$$I_a(S_d) = Y_a \qquad I_s(S_d) = Y_s.$$

$$\hat{S}_d^* = f^*(S_d): I_s(\hat{S}_d^*) \neq I_s(S_d) = Y_s$$

$$I_a(\hat{S}_d^*) = I_a(S_d) = Y_a$$

$$\ell_s^{\phi,j}(\hat{S}_d, N_s) = ||G_j^{\phi}(\hat{S}_d) - G_j^{\phi}(N_s)||_F^2 \qquad \ell_s^{\phi} = \ell_s^{\phi,Conv1\_2}(\hat{S}_d, N_s) + \ell_s^{\phi,Conv2\_2}(\hat{S}_d, N_s).$$




$$\ell_a^{\phi} = -\sum_e Y_{a,e} log(\phi(S_d)_e),$$

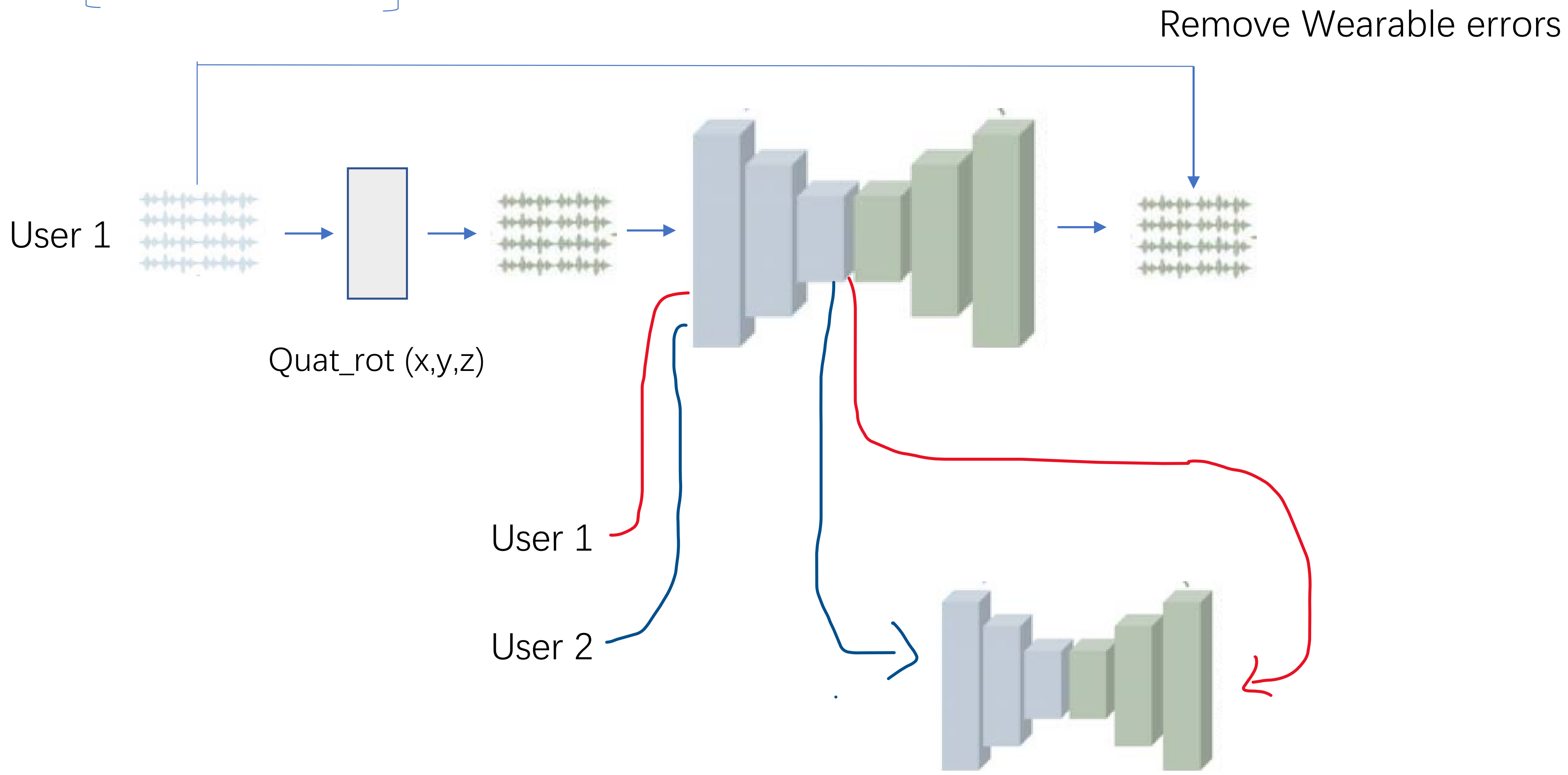
* The transformed data to keep the “content” similar to the raw data, but does not force them to match exactly

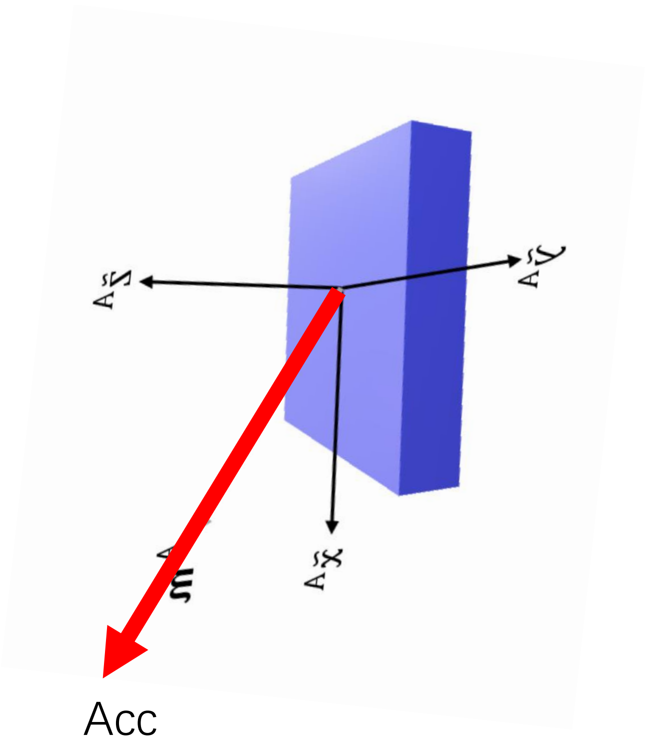
$$\ell_c^{\phi,j} = \frac{1}{C_j H_j W_j} ||\phi_j(\hat{S}_d) - \phi_j(S_d)||_2^2$$

Idea:

User heterogeneity

- **Biological conditions and behavior patterns**
- Wearable errors
- Sensor errors
- Random noise

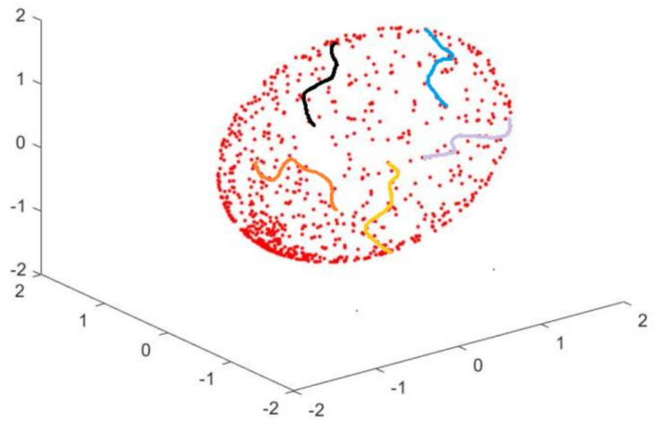




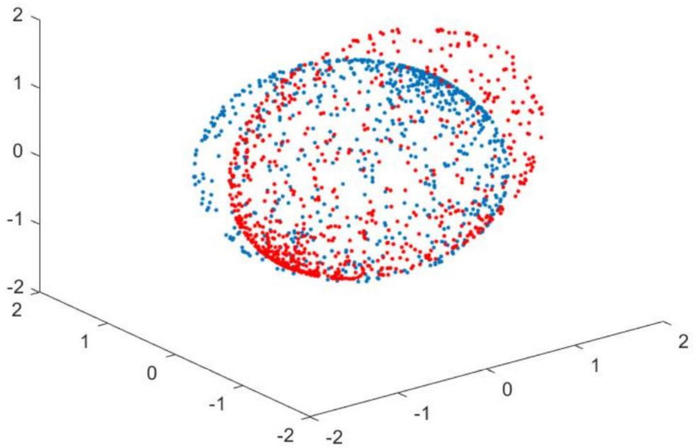

身体受到的合加速度在传感器坐标系下的投影

Errors type:

1. 佩戴
2. User （速度，习惯，时长）
3. 传感器固有误差
4. random noise



合加速度的所有情况构成了空间的椭球，运动的产生的加速度序列就是在椭球表面取轨迹



不同人，不同运动速度，佩戴误差，传感器误差

椭球面

$$a_1x^2 + a_2y^2 + a_3z^2 + a_4xy + a_5xz + a_6yz + a_7x + a_8y + a_9z = 1$$

球面

$$x^2 + y^2 + z^2 + 2fx + 2gy + 2hz + d = 0$$

M1 -> (V1 + Ve1) -> M1'

M2 -> (V2 + Ve2) -> M2'

$V1 \approx V2 = V$

V1 + Ve2 -> M2'

V2 + Ve1 -> M1'