Project 2. Human Action Recognition using Hidden Markov Model

Sangjun Son
Department of Computer Science & Engineering, Seoul National University
lucetre@snu.ac.kr

Backgrounds

- Human Activity Recognition is nowadays an active research field which aims to understand human.
- Human Activities and Postural Transitions datasets (HAPT) → Data selection
  - Performs basic activities and postural transitions while carrying a waist-mounted smartphone with embedded inertial sensors.
  - Includes 61 experiments and was measured 50Hz in 400s duration.
- Temporal models like HMM enables to learn the observation patterns from time serial dataset. → Time-serial

Action Type	
Walking	Stand to Sit
Walking Upstairs	Sit to Stand
Walking Downstairs	Sit to Lie
Sitting	Lie to Sit
Standing	Stand to Lie
Laying	Lie to Stand

Research Goals

- **To study the Hidden Markov Model and its applications**
  - Build a model finding a latent temporal pattern.
- **Fast and accurate Human Action Recognition system**
  - Extract relevant features from sensor data (acceleration, gyro sensors)
  - Use the multinomial HMM package
  - Model complexity vs. Accuracy trade-offs

Preliminaries

- **Hidden Markov Model**
  - Probabilistic model that infers hidden states with observation events as causal factors.

Markov process :
$$\begin{tikzcd} X_0 \arrow[r, "A"] \arrow[d, "B"] & X_1 \arrow[r, "A"] \arrow[d, "B"] & X_2 \arrow[r, "A"] \arrow[d, "B"] & \cdots \arrow[r, "A"] & X_{T-1} \arrow[d, "B"] \\ \mathcal{O}_0 & \mathcal{O}_1 & \mathcal{O}_2 & \cdots & \mathcal{O}_{T-1} \end{tikzcd}$$
Observations :

- Three Fundamental Problems in HMM

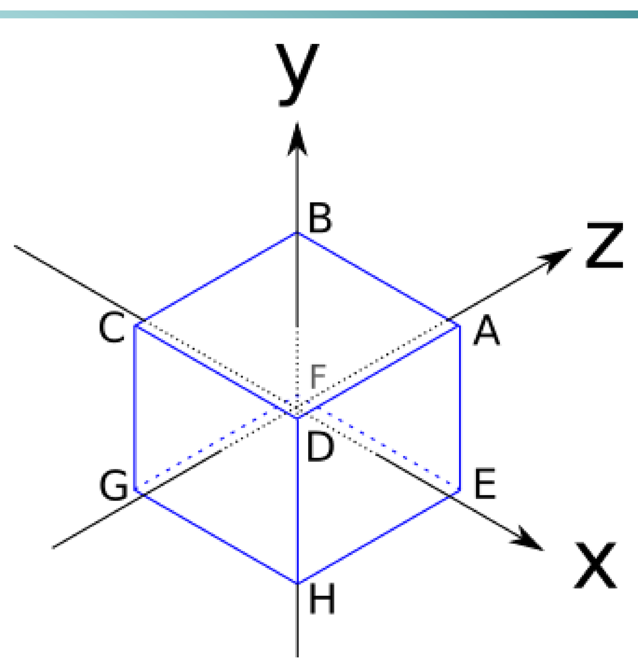
1. Likelihood 2. Decoding 3. Learning

Methodology

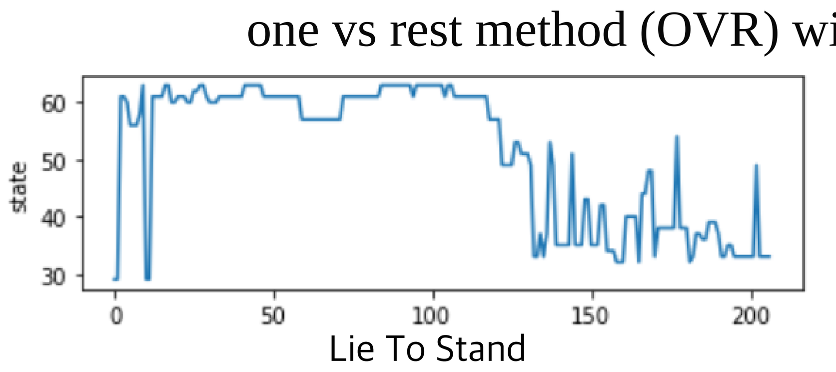
- **Feature Extraction**
  - Feature metric describing our hidden states.
    - **Jerk** and **Angular velocity**

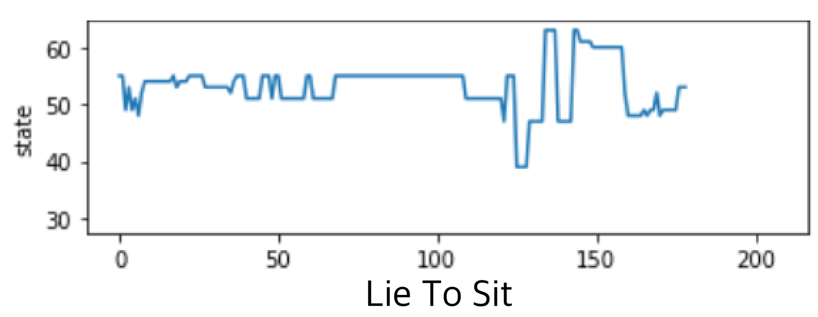
$$\vec{j}(t) = \frac{\mathrm{d}\vec{a}(t)}{\mathrm{d}t} = \frac{\mathrm{d}^2\vec{v}(t)}{\mathrm{d}t^2} = \frac{\mathrm{d}^3\vec{r}(t)}{\mathrm{d}t^3}$$

    - Direction separated into 8 sections



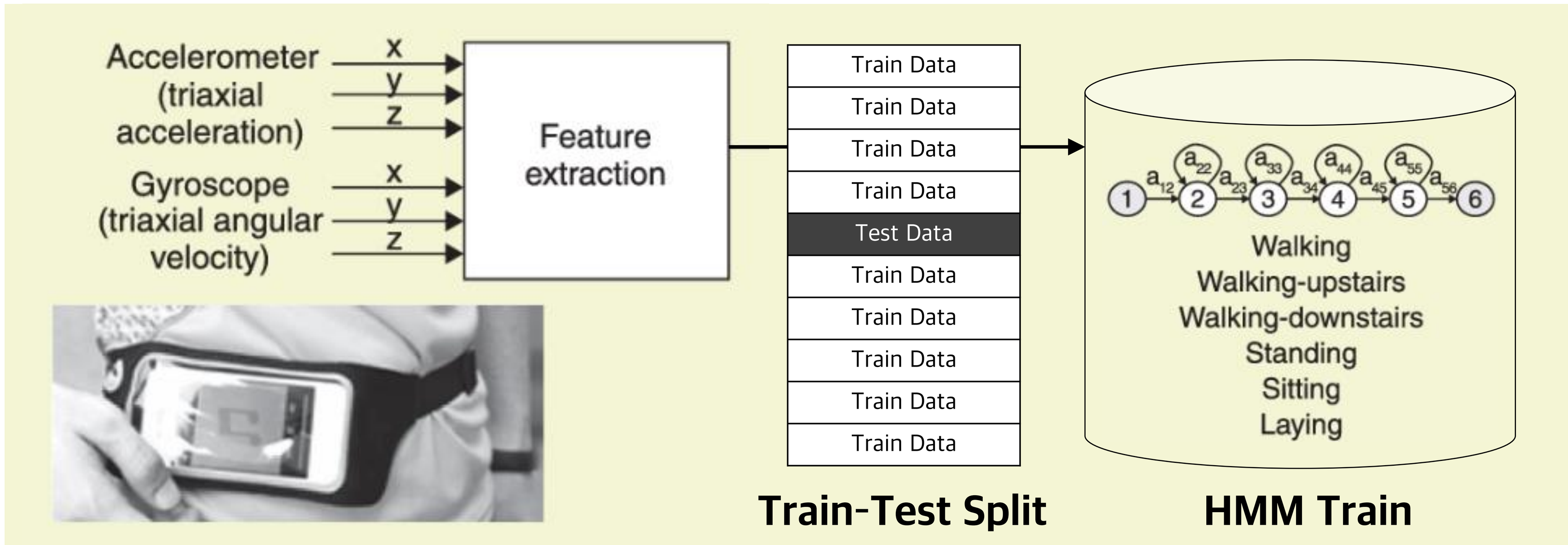
- **Multinomial HMM**
  - Observation values are all in discrete number.
  - 12 Activity Models
    - Each model learns corresponding activity and classifies test data set with one vs rest method (OVR) with maximum likelihood.





Experimental Results

- **Data Manipulation Flow**



- **One-iteration Trained Model Comparison**

Model Name	jerk j	acc a	vel α	Accuracy
jerk8	✓			50.81%
acc8		✓		35.24%
vel8			✓	60.65%
acc8vel8		✓	✓	78.68%
jerk8acc8	✓	✓		40.16%
jerk2vel8	✓		✓	63.93%
jerk4vel8	✓		✓	79.50%
jerk8vel8	✓		✓	**84.42%**
jerk8acc2vel8	✓	✓	✓	63.11%

- **Feature Metric Approach**
  - Angular velocity as significant evaluation metric

jerk8 vs. ***vel8*** vs. ***acc8***

  - Combination of jerk and velocity directions were used as observations.

acc8vel8 vs. ***jerk8vel8*** vs. ***jerk8acc8***

- **Model Complexity Approach**
  - Avoid overfitting issues for constructing model with too many states.

jerk4vel8 vs. ***jerk8vel8*** vs. ***jerk2vel8, jerk8acc2vel8***

→ Selected our model as *jerk8vel8 and* trained for 50 iterations, ***jerk8vel8iter50***

- **Model *jerk8vel8iter50***
  - 92.62% as test accuracy and 24623.64s as total training time

jerk8vel8iter50

Action Types	Log Likelihood	Training Time
WALKING	-202473.76	3876.99s
WALKING UPSTAIRS	-140582.82	3704.82s
WALKING DOWNSTAIRS	-186022.69	3408.61s
SITTING	-156591.76	3681.86s
STANDING	-161618.97	4011.79s
LAYING	-188243.51	3956.03s
STAND TO SIT	-8980.82	305.05s
SIT TO STAND	-6546.16	221.56s
SIT TO LIE	-11448.79	370.29s
LIE TO SIT	-9051.13	331.27s
STAND TO LIE	-14043.55	428.47s
LIE TO STAND	-10045.41	326.90s

Concluding Remarks

- **Constructed a fast and accurate HMM model to recognize human action**
  - A probabilistic model simply learning the observation patterns
  - Found latent states by appropriate feature extraction.
- **Future works to improve our model**
  1. Train with numerous instances created from more experiments.
  2. Use other additional sensors to extract various states features.