Dynamic Data Modeling, Recognition, and Synthesis

Rui Zhao
Thesis Defense
Advisor: Professor Qiang Ji

Contents

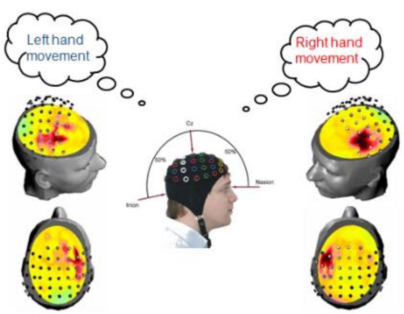
Overview

Speech Processing

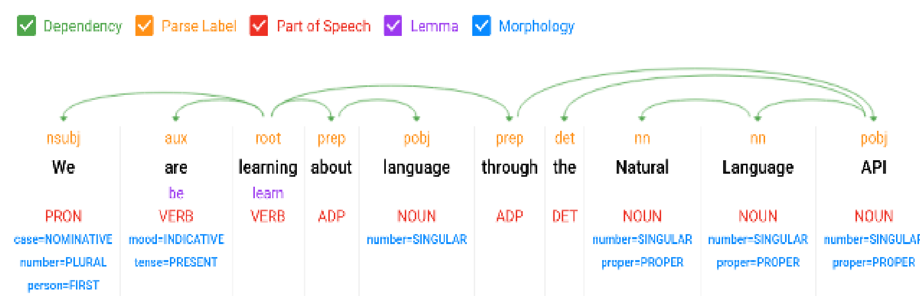
Computer Vision

Dynamic Data

Natural Language Processing



Neurology



Major Tasks

- Analyze Dynamic Data

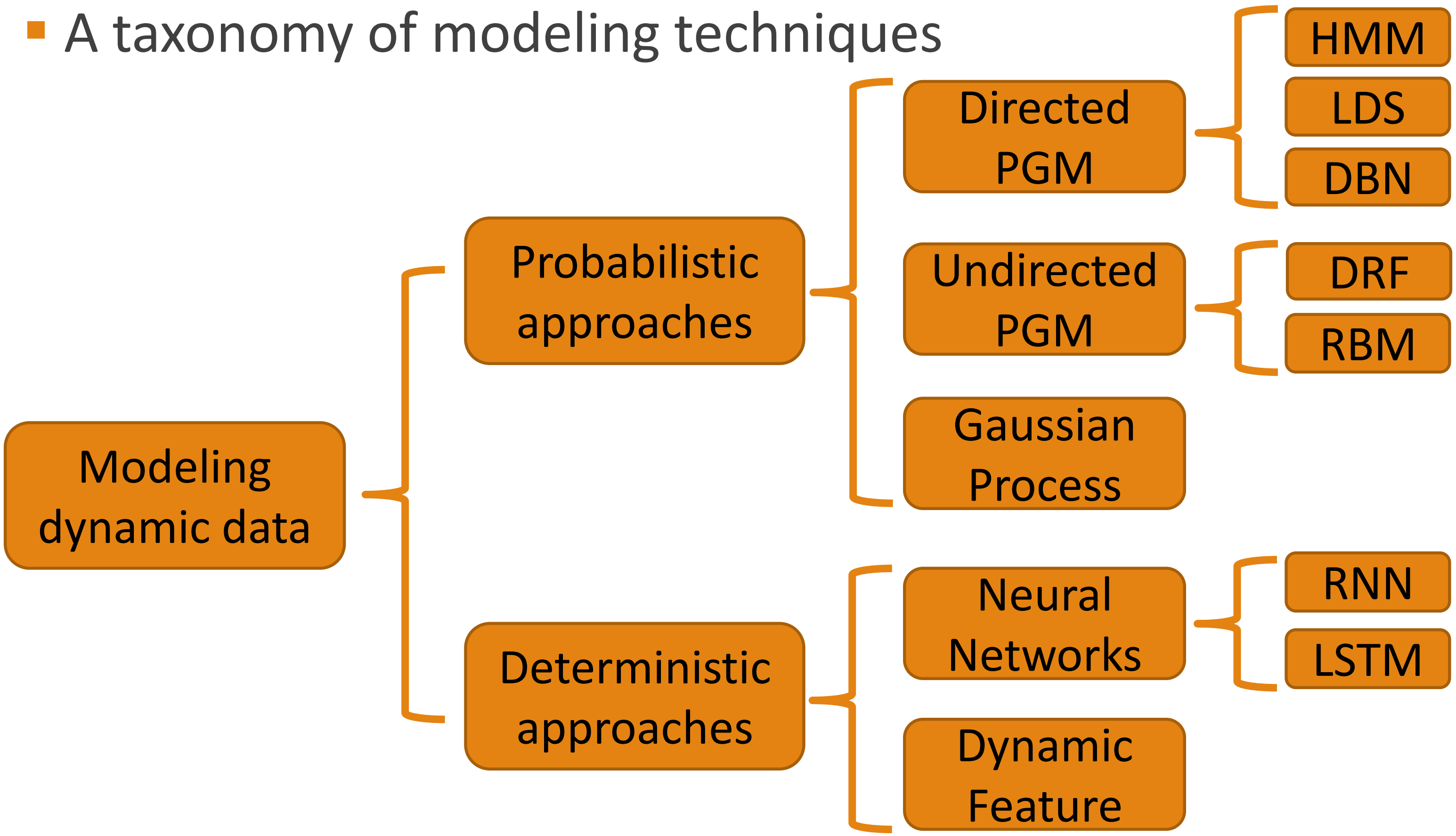
- Modeling
  - Provide a mathematical description of the dynamic process
- Analysis
  - Prediction: Forecast future values of dynamic data
  - Regression: Estimate a target value given dynamic data
  - Classification: Divide dynamic data into different categories
  - Synthesis: Generate new dynamic data

Challenges

- Insufficient annotation
- Uncertainty in data and model
- Intra-class variation
- Complex dynamics

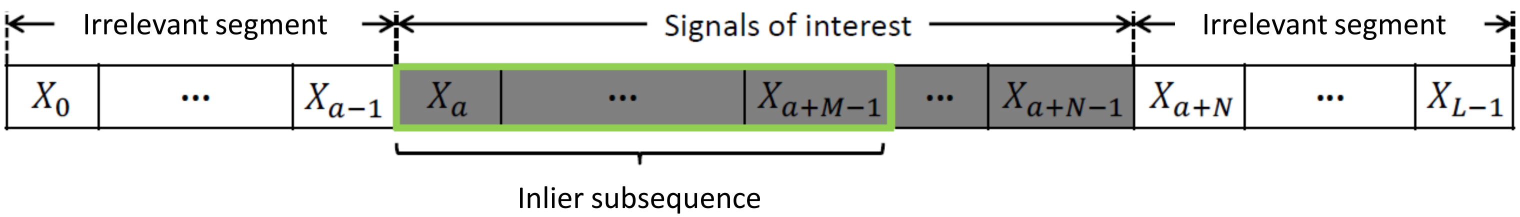
Related Work

- A taxonomy of modeling techniques



Part 1: Dynamic Pattern Localization

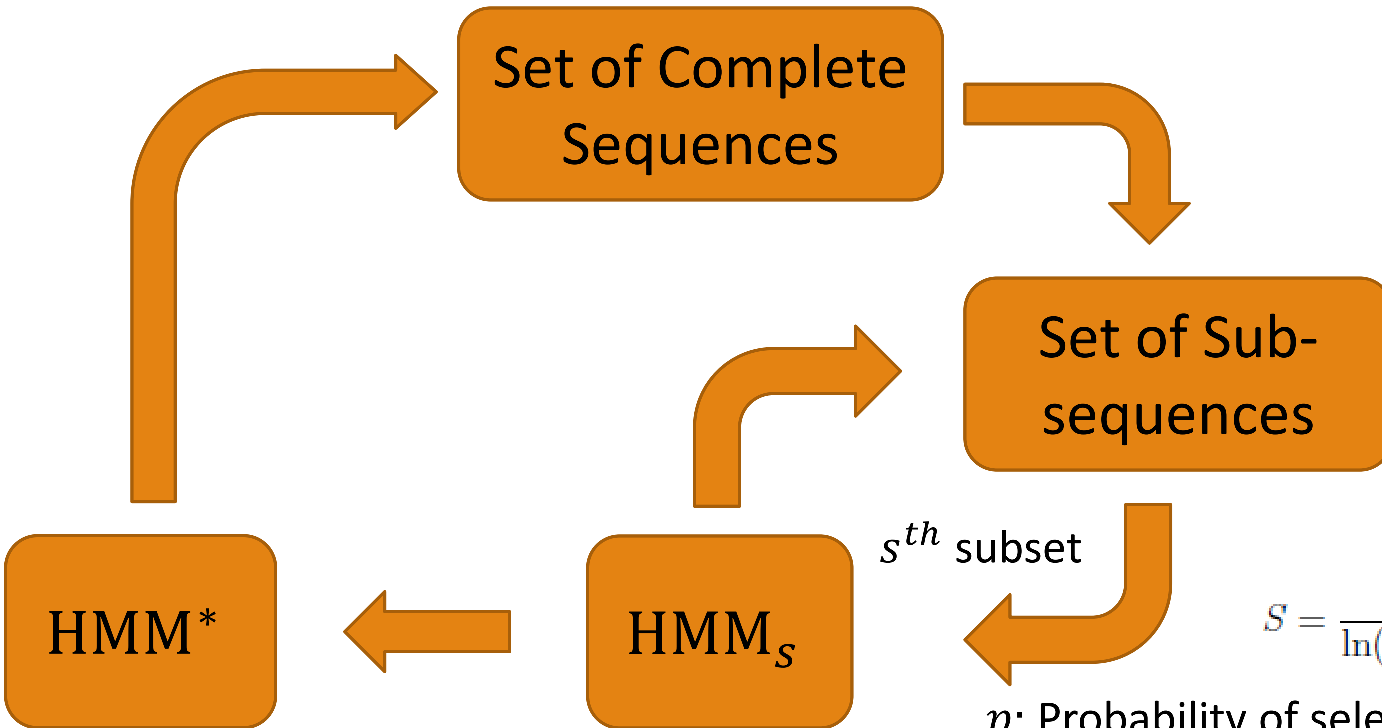
- Problem: Temporally localize the dynamic pattern in a time series by determining its starting and ending time



- Applications:
  - Brain computer interface (BCI)
  - Speech recognition
  - Event recognition
- Our Solution:
  - Combine dynamic model (HMM) with robust estimation (RANSAC)

Methods

- Overview

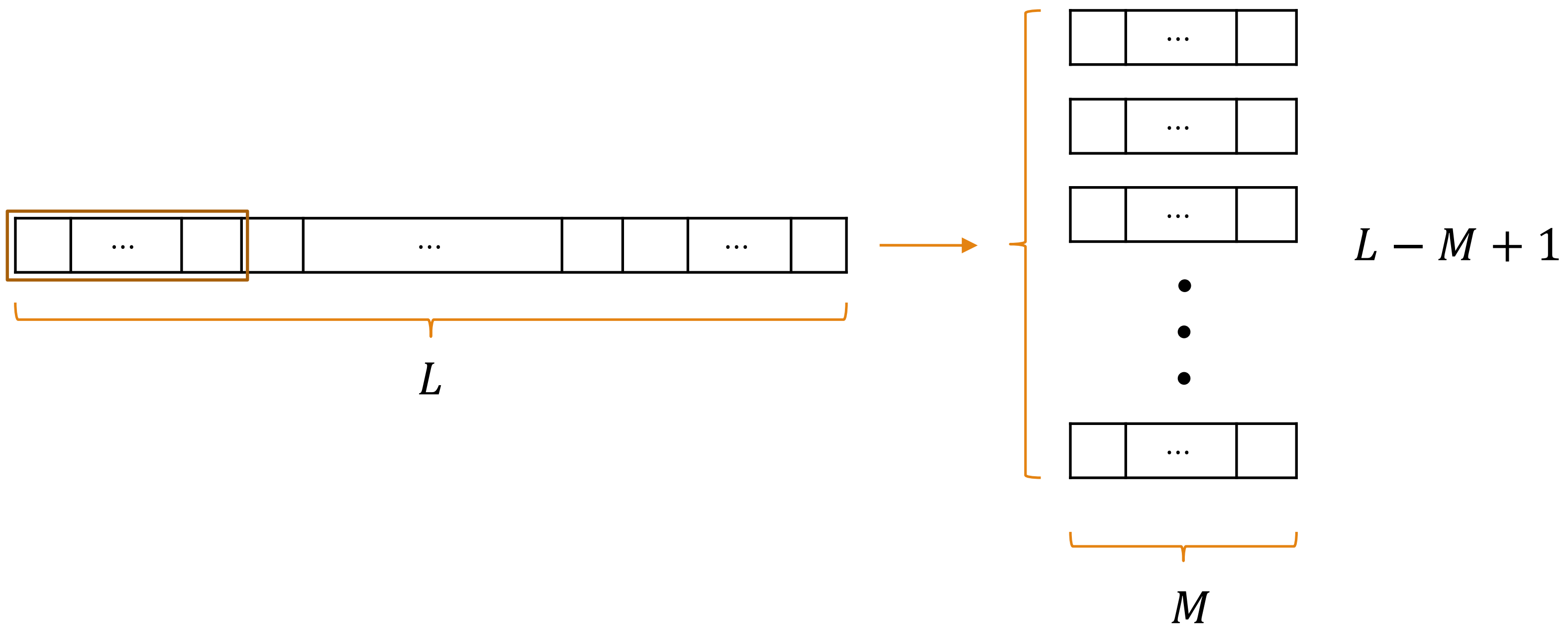


$$S = \frac{\ln(1-p)}{\ln(1-(1-\epsilon)^K)}$$

p: Probability of selecting all inliers
ϵ: Proportion of outliers
K: Total number complete sequences

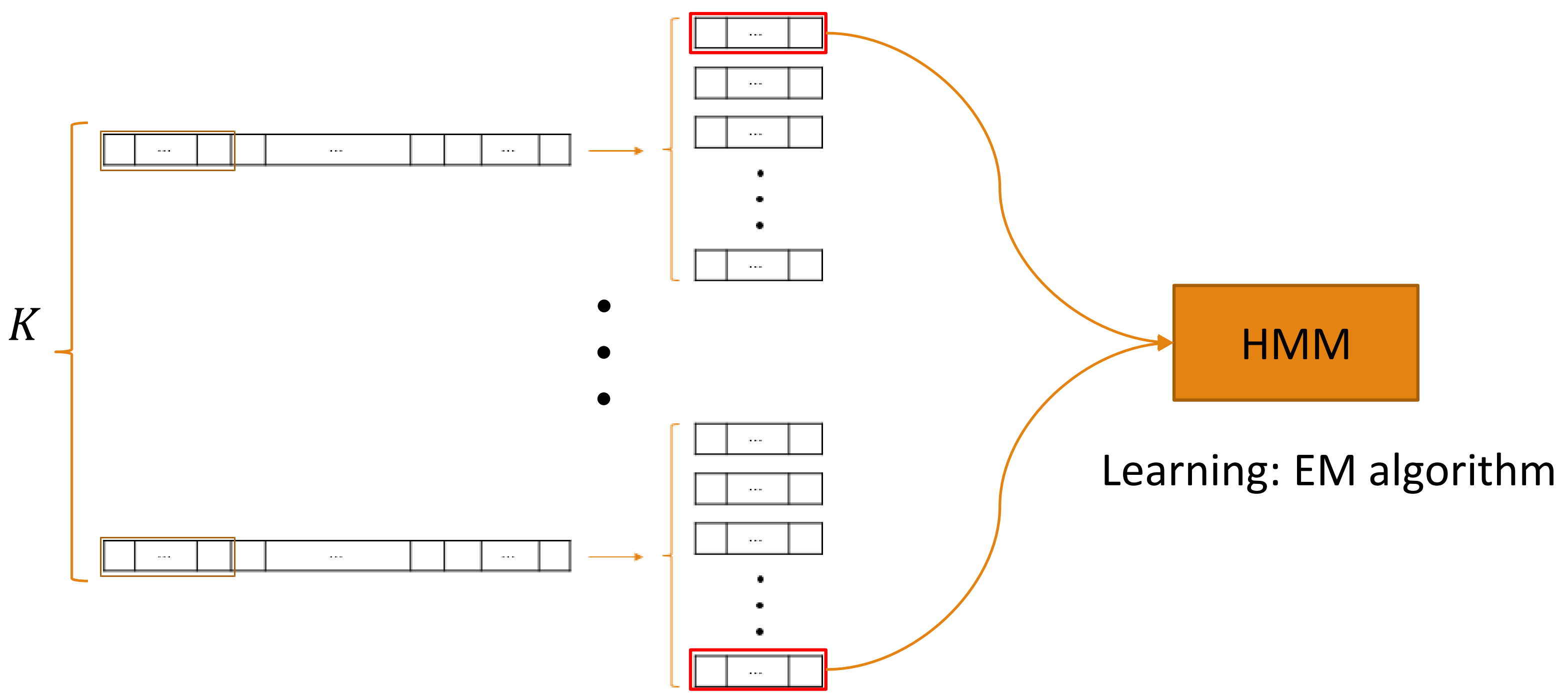
Methods

- Step 1: Divide complete sequences into overlapping subsequences



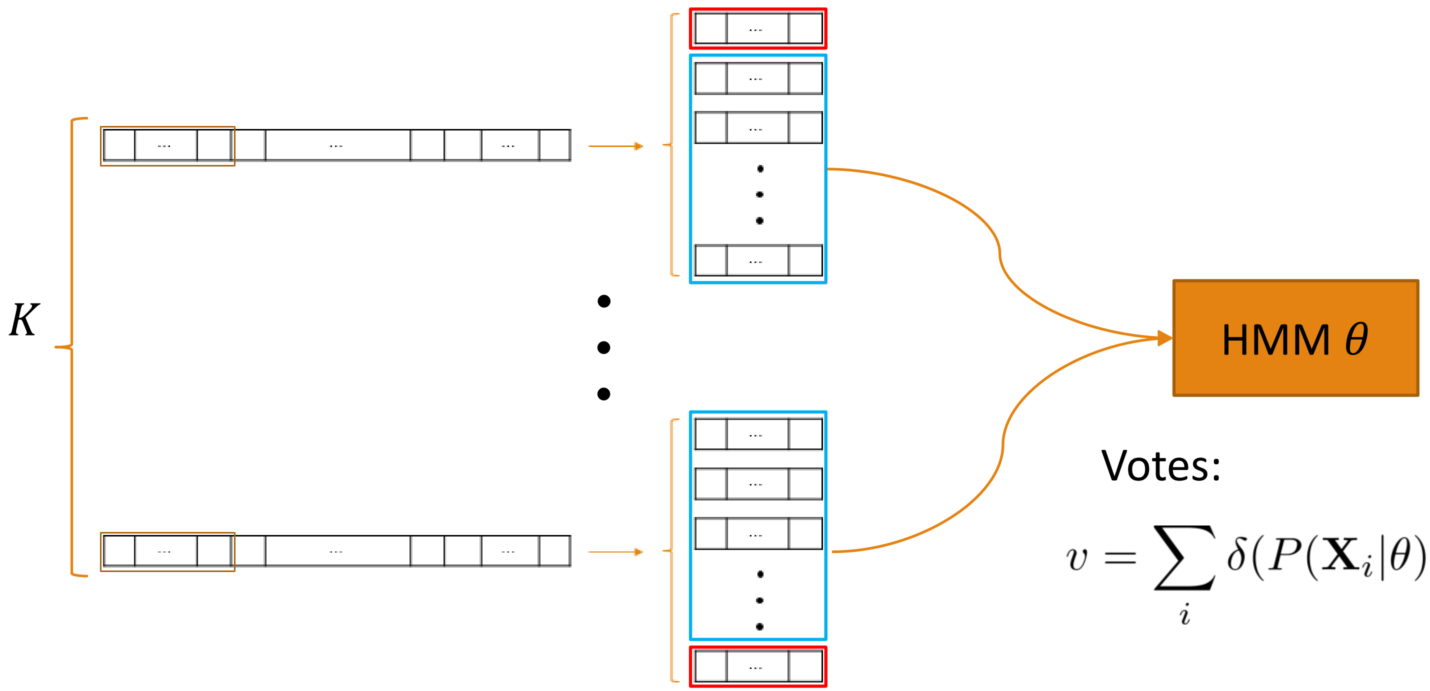
Methods

- Step 2: Randomly select a subsequence from each complete sequence to train HMM



Methods

- Step 3: Vote for the learned HMM using remaining subsequences



Votes:

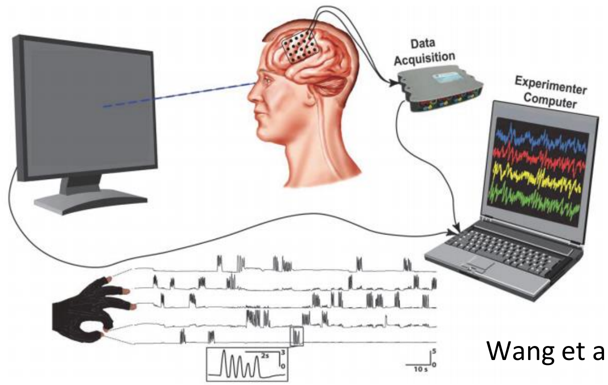
$$v = \sum_i \delta(P(\mathbf{X}_i|\theta) > \bar{P}_{\text{train}})$$

Methods

- Step 1: Divide complete sequences into overlapping subsequences
- Step 2: Randomly select a subsequence from each complete sequence to train HMM
- Step 3: Vote for the learned HMM using remaining subsequences
- Step 4: Go back to Step 2 and repeat for enough times $S = \frac{\ln(1-p)}{\ln(1-(1-\epsilon)^K)}$
- Step 5: Find the HMM with the highest vote and use it to identify the inlier subsequences
- Step 6: Retrain the HMM using inlier subsequences
- Step 7: Identify inlier subsequences and merge to get signals of interest

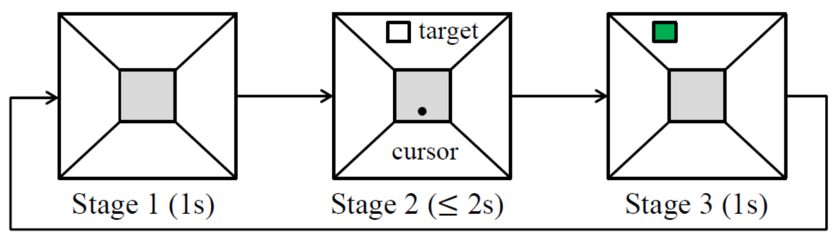
Experiments

- Electrocorticographic (ECoG) data
  - Data collection

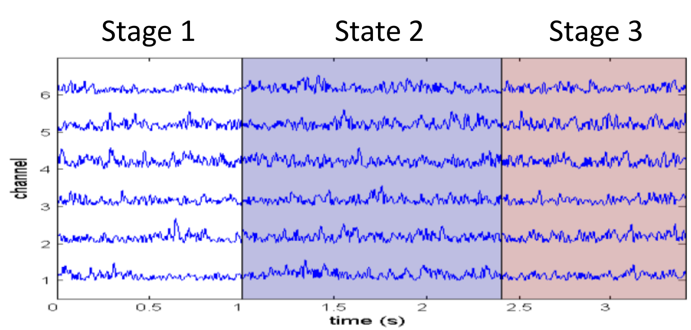


Wang et al. IJCAI2013

Experiment protocol



  - Preprocessing & Feature Extraction: Spectrum filter (70-170Hz) + Hilbert transform



Experiments

- Data Statistics
  - 4 subjects
  - Each has 10 trials, each of length 800
  - Average hit time

Subject	A	B	C	D	Average
Duration (s)	1.43	1.80	1.48	0.92	1.41

- Classification Results:

Subject	A	B	C	D	Average (CI95)
Manual	32.5	65.4	74.5	42.5	53.7 (42.9,64.2)
ACA	63.0	58.1	58.8	41.8	55.4 (44.5,65.8)
SC	**63.8**	60.6	80.4	38.8	60.9 (49.9,70.9)
Ours	**63.8**	**67.3**	**100**	**50.0**	**70.3** (59.5,79.2)

Part 2: Dynamic Regression under Insufficient Annotation

- Problem: Given sequential data $\mathrm{x}_1, \ldots \mathrm{x}_T$, we want to compute a regression function $y_t = f(\mathrm{x}_t)$, for some target value y_t
- Challenge:
  - Only part of the sequential data are annotated
- Applications:
  - Facial expression intensity estimation
  - Sensor fault detection
  - Part-of-speech tagging
- Our solution: Incorporate temporal information as additional constraints

Problem Statement

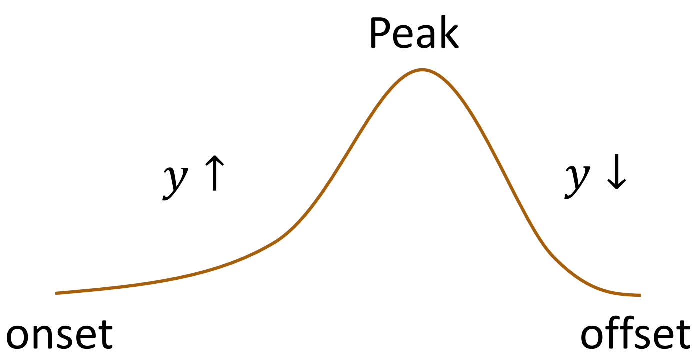
- Goal: Given input set with partial labels,

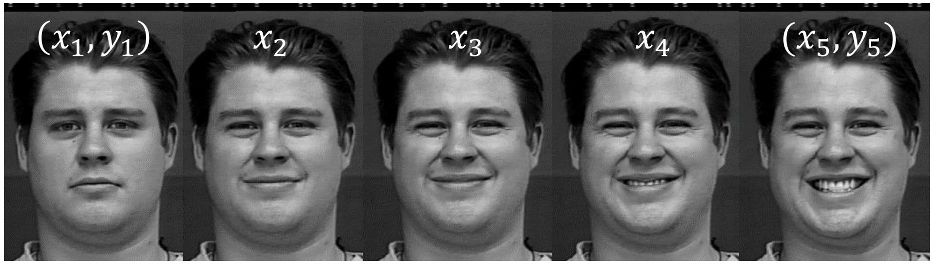
$$\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^d | i = 1, \dots, |\mathbf{X}|\}$$
$$\mathbf{Y} = \{y_i \in \mathbb{R} | i \in \mathbf{V}\}$$
$$\mathbf{V} \subseteq \{1, \dots, |\mathbf{X}|\}$$

find a regression function from $\mathbf{x}$ to y

$$f : \mathbb{R}^d \mapsto \mathbb{R} \quad y = f(\mathbf{x}; \theta)$$

- Example:





$$\mathbf{V} = \{1,5\} \quad \mathbf{E} = \left\{ \begin{matrix} (1,2), (1,3), (1,4), (1,5), (2,3) \\ (2,4), (2,5), (3,4), (3,5), (4,5) \end{matrix} \right\}$$

Ordinal Support Vector Regression (OSVR)

- Regression model $f(\mathbf{x};\theta) = \mathbf{w}^T\mathbf{x} + b$ with parameters $\theta = \{\mathbf{w}, b\}$
- Dataset with weak labels: $\mathcal{D} = \{\mathbf{X}_n, \mathbf{Y}_n, \mathbf{V}_n, \mathbf{E}_n\}, n = 1, ..., N$
- Optimization problem

Objective = Regularization + Regression Loss + Ordinal Loss

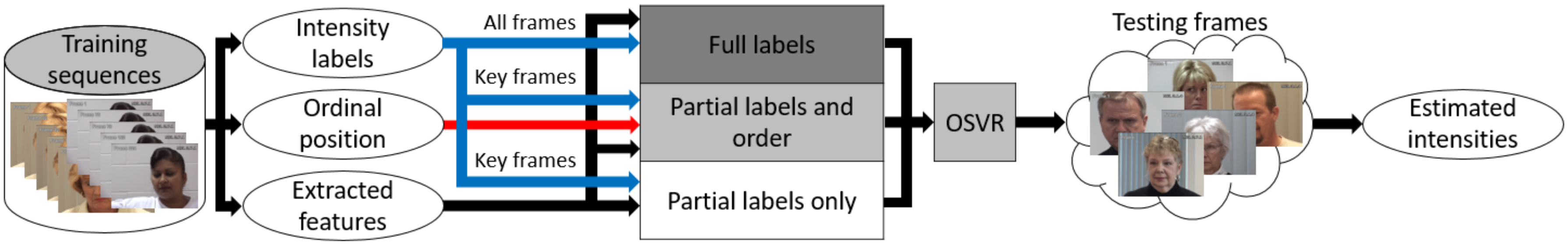
$$\min_{\theta,\eta,\xi} \boxed{\frac{1}{2}||\mathbf{w}||^2} + \gamma_1 \sum_{n=1}^{N} \sum_{k\in\mathbf{V}_n} \boxed{(l_1(\eta_k^{(n)+}) + l_1(\eta_k^{(n)-}))} + \gamma_2 \sum_{n=1}^{N} \sum_{(i,j)\in\mathbf{E}_n} \boxed{l_2(\xi_{ij}^{(n)})}$$

$$\begin{aligned}
\text{s.t. } & \mathbf{w}^T\mathbf{x}_k^{(n)} + b - y_k^{(n)} \le \epsilon + \eta_k^{(n)+} \\
& y_k^{(n)} - \mathbf{w}^T\mathbf{x}_k^{(n)} - b \le \epsilon + \eta_k^{(n)-} \\
& \mathbf{w}^T(\mathbf{x}_i^{(n)} - \mathbf{x}_j^{(n)}) \ge 1 - \alpha_{ij}\xi_{ij}^{(n)} \\
& \eta_k^{(n)+}, \eta_k^{(n)-}, \xi_{ij}^{(n)} \ge 0 \\
& \forall k \in \mathbf{V}_n, (i,j) \in \mathbf{E}_n,\ n = 1, ..., N
\end{aligned}$$

- Optimization method: alternating direction method of multipliers (ADMM)

Experiments: Facial Expression Intensity Estimation

- Overview



- Dataset: PAIN
- Feature: Gabor features, landmark, LBP + PCA
- Evaluation Criteria:
  - Pearson correlation coefficient (PCC)
  - Intra-class correlation (ICC)
  - Mean absolute error (MAE)

Experiments

- PAIN dataset
  - Experiments under different annotation settings
  - Partial labels: about 8.8% of the total number of frames
  - Use of ordinal information is very helpful

Setting	PCC	ICC	MAE
Full labels	**0.5659**	**0.5045**	**0.8538**
Partial labels + order	0.5441	0.4955	0.9519
Partial labels only	0.4766	0.4511	1.3895

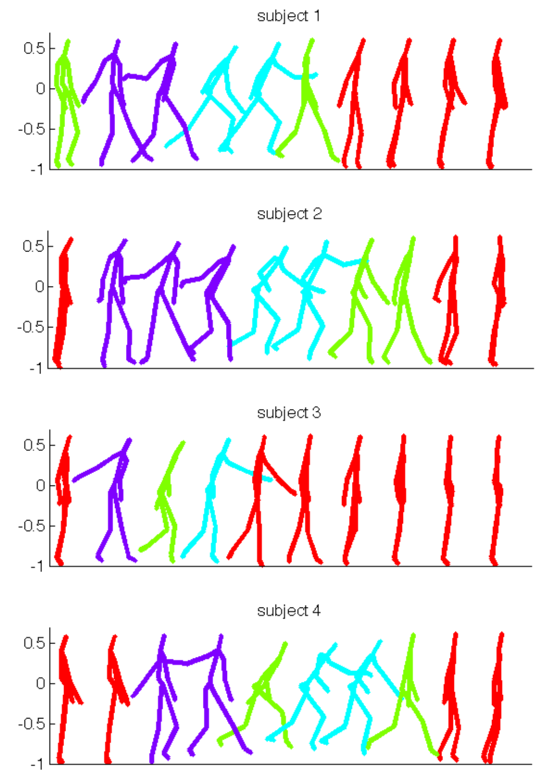
Experiments

- PAIN dataset
  - Comparison with state-of-the-art

Method	PCC	ICC	MAE
SVR [10]	0.4766	0.4511	1.3895
SVOR [4]	0.5051	0.4240	2.9801
RVR[6]	0.4823	0.4365	1.1122
Ours	**0.5441**	**0.4955**	**0.9519**

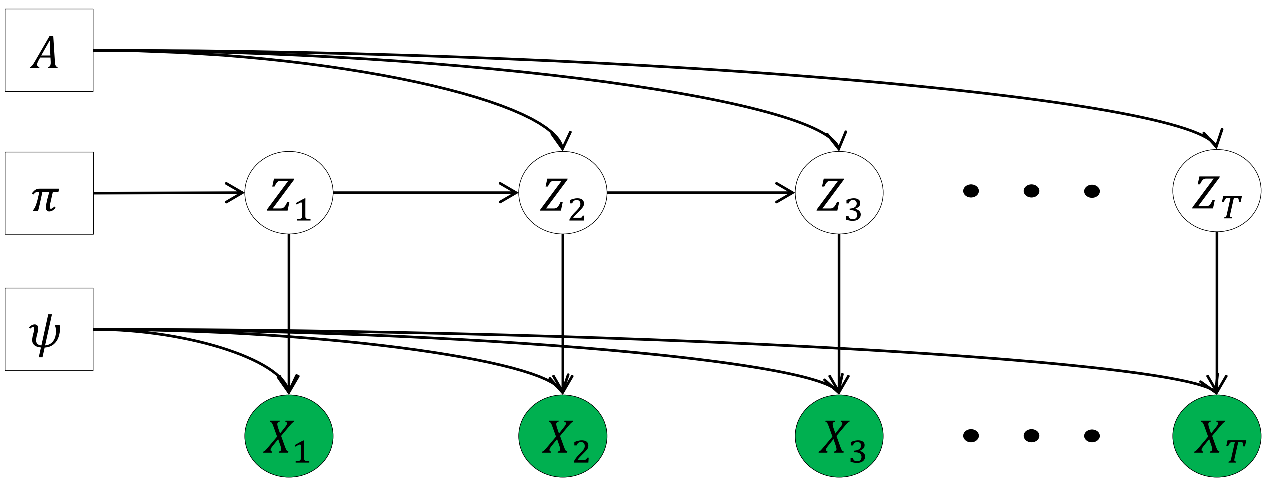
Part 3: Classification and Synthesis under Large Intra-class Variation

- **Challenge:**
  - Same underlying dynamic pattern can manifest significant intra-class variation
- **Applications:**
  - Human action recognition
  - Human motion synthesis
  - Speech recognition
  - Language translation
- **Our Solution:**
  - Hidden Semi-Markov Model
  - Bayesian Hierarchical Modeling



Methods

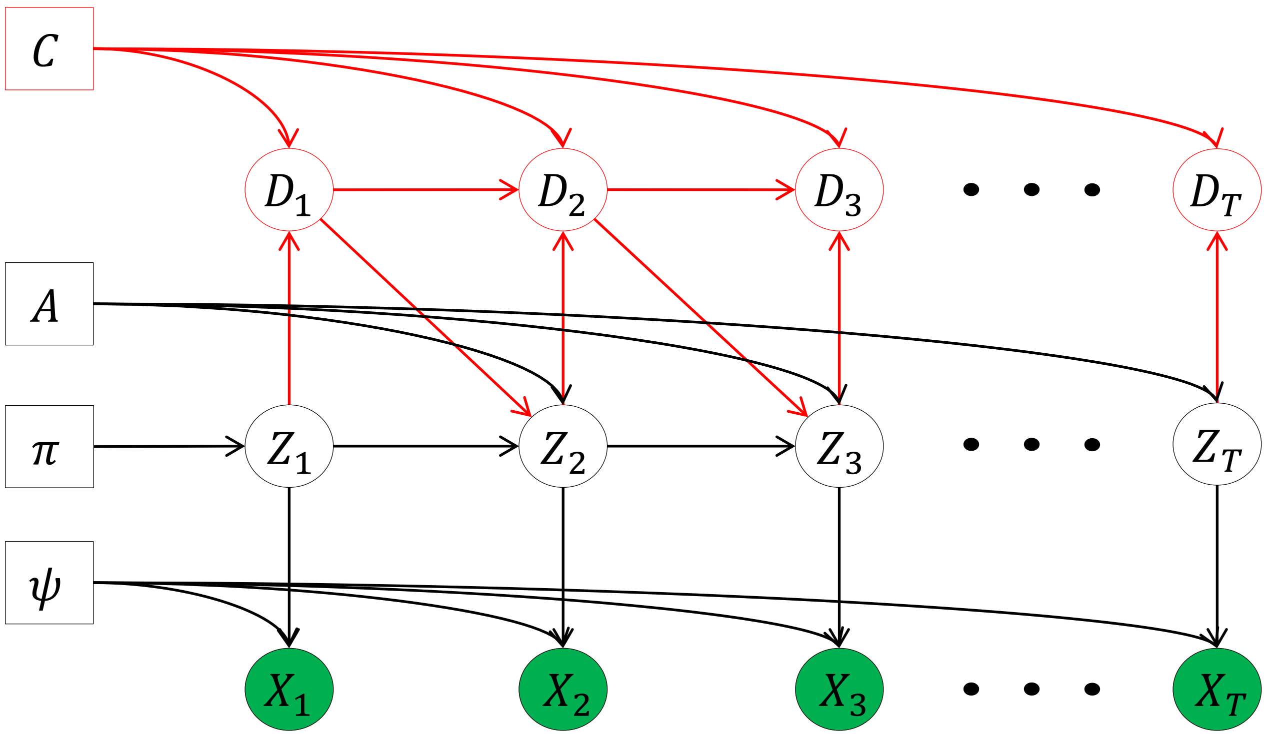
- Hidden Markov Model (HMM)



$$P(\mathbf{X}, \mathbf{Z}|\theta) = \underbrace{P(Z_1)}_{\text{initial}} \underbrace{\prod_{t=2}^{T} P(Z_t|Z_{t-1})}_{\text{transition}} \underbrace{\prod_{t=1}^{T} P(X_t|Z_t)}_{\text{emission}}$$

Methods

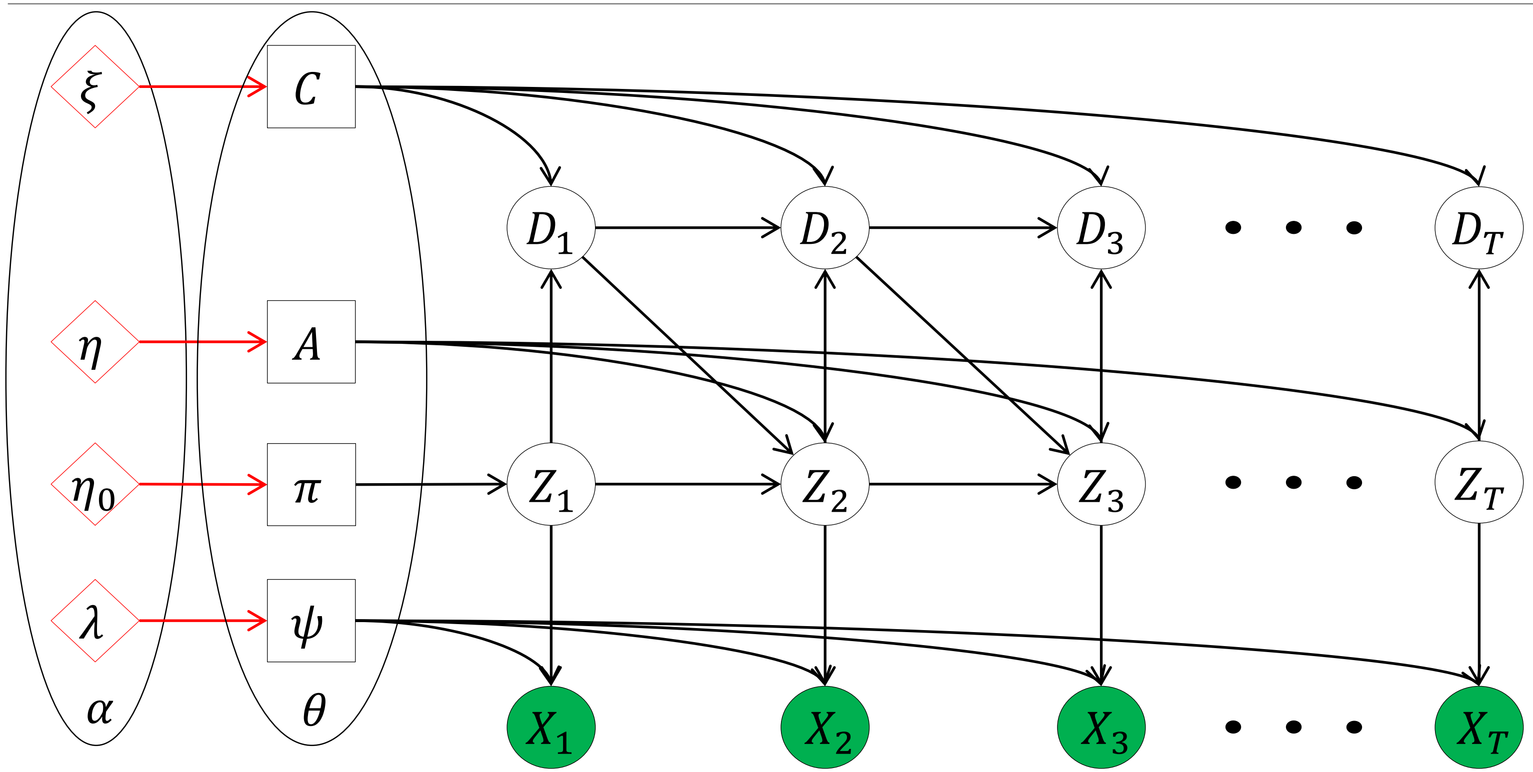
- Hidden Semi-Markov Model (HSMM)



$$P(\mathbf{X},\mathbf{Z},\mathbf{D}|\theta) = \underbrace{P(Z_1)P(D_1|Z_1)}_{\text{initial}} \prod_{t=2}^{T} \underbrace{P(Z_t|Z_{t-1},D_{t-1})}_{\text{transition}}\underbrace{P(D_t|D_{t-1},Z_t)}_{\text{duration}} \underbrace{\prod_{t=1}^{T} P(X_t|Z_t)}_{\text{emission}}$$

Methods

- Bayesian Hierarchical Dynamic Model (HDM)



$$P(\mathbf{X},\mathbf{Z},\mathbf{D},\theta|\alpha) = P(\mathbf{X},\mathbf{Z},\mathbf{D}|\theta)P(\theta|\alpha)$$

likelihood prior

Methods

- Learning: estimating hyperparameter α
  - Overall objective

$$\alpha^* = \arg\max_{\alpha} \log P(\mathbf{X}|\alpha)$$
$$= \arg\max_{\alpha} \log \int_{\theta} \sum_{\mathbf{Z},\mathbf{D}} P(\mathbf{X},\mathbf{Z},\mathbf{D}|\theta) P(\theta|\alpha) d\theta$$

Intractable

  - Approximate objective

$$\alpha^* = \arg\max_{\alpha} \log \left[\arg\max_{\theta} \sum_{\mathbf{Z},\mathbf{D}} P(\mathbf{X},\mathbf{Z},\mathbf{D}|\theta) P(\theta|\alpha) \right]$$

  - An alternating strategy

$$\theta^* = \arg\max_{\theta} \log P(\mathbf{X}|\theta) + \log P(\theta|\alpha) \longrightarrow \text{MAP-EM}$$
$$\alpha^* = \arg\max_{\alpha} \log P(\theta^*|\alpha) \longrightarrow \text{ML}$$

- Optimization methods:

Methods

- Bayesian Inference: predictive likelihood

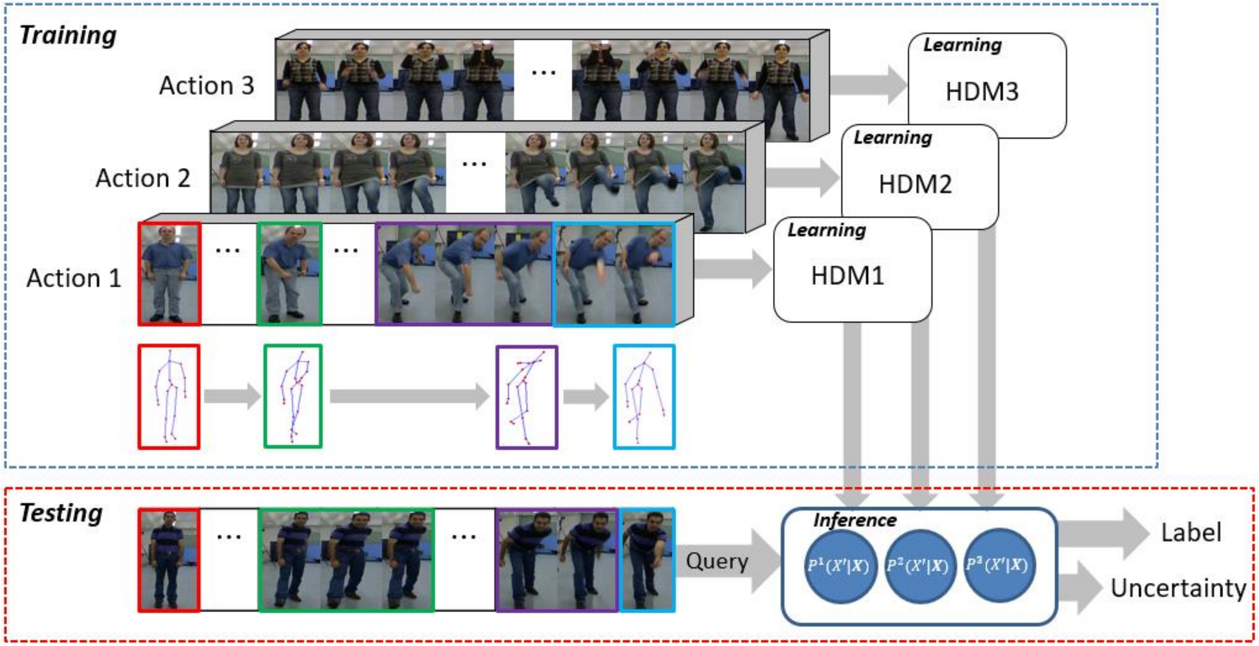
$$P(\mathbf{X}|\mathcal{D},\alpha_i^*) = \int_\theta \sum_{\mathbf{Z},\mathbf{D}} P(\mathbf{X},\mathbf{Z},\mathbf{D}|\theta)P(\theta|\mathcal{D},\alpha_i^*)d\theta$$

$$\approx \frac{1}{L}\sum_{l=1}^{L}\sum_{\mathbf{Z},\mathbf{D}} P(\mathbf{X},\mathbf{Z},\mathbf{D}|\theta^{(l)}), \theta^{(l)} \sim P(\theta|\mathcal{D},\alpha_i^*)$$

  - Advantage: reduce overfitting and improve generalization
  - Posterior inference: $\theta^{(l)} \sim P(\theta|\mathcal{D},\alpha_i^*)$
    - Method: Gibbs Sampling
  - Classification: $y^* = \arg\max_i P(\mathbf{X}|\mathcal{D},\alpha_i^*)$

Experiments: Action Recognition

- Overview:



- Dataset: MSR-Action3D, UTD-MHAD, G3D, Penn
- Feature: Joint position and motion in 3D or 2D

Experiments: Action Recognition

- Individual dataset: comparison with different baselines

Model	MSRA	UTD	G3D	Penn	Avg.
HMM	67.8	82.8	68.1	82.3	75.3
HSMM	66.3	82.3	77.5	78.9	76.3
Ours	**86.1**	**92.8**	**92.0**	**93.4**	**91.1**

Experiments: Action Recognition

- Individual dataset: comparison with different state-of-the-art

MSRA		UTD	
Method	Acc.	Method	Acc.
AS[9]	83.5	Fusion[3]	79.1
AL[12]	88.2	DMM[1]	84.2
MT[5]	**92.0**	CNN[13]	85.8
Ours	86.1	Ours	**92.8**

G3D		Penn	
Method	Acc.	Method	Acc.
LRBM[7]	90.5	Actemes[14]	86.5
R3DG[11]	91.1	AOG[8]	84.8
CNN[13]	**94.2**	JDD[2]	93.2
Ours	92.0	Ours	**93.4**

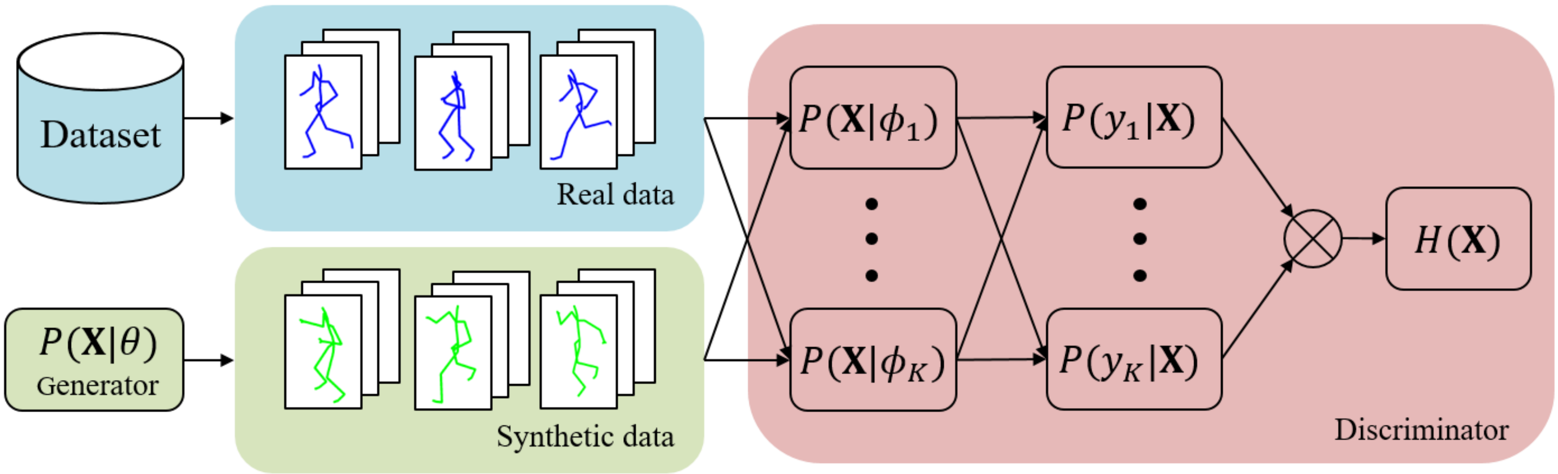
Experiments: Action Recognition

- Cross dataset classification accuracy
  - A: MSR
  - B: UTD
  - C: G3D

Train	Test	HMM	HSMM	R3DG	DLSTM	Ours
B,C	A	62.59	65.31	76.19	70.75	**89.21**
A,C	B	66.25	61.88	**86.25**	85.00	75.00
A,B	C	30.94	42.45	51.08	38.85	**61.15**
Average		53.26	56.55	71.17	64.87	**75.12**

Methods: Adversarial Learning

- Adversarial Learning: a better criterion for data synthesis

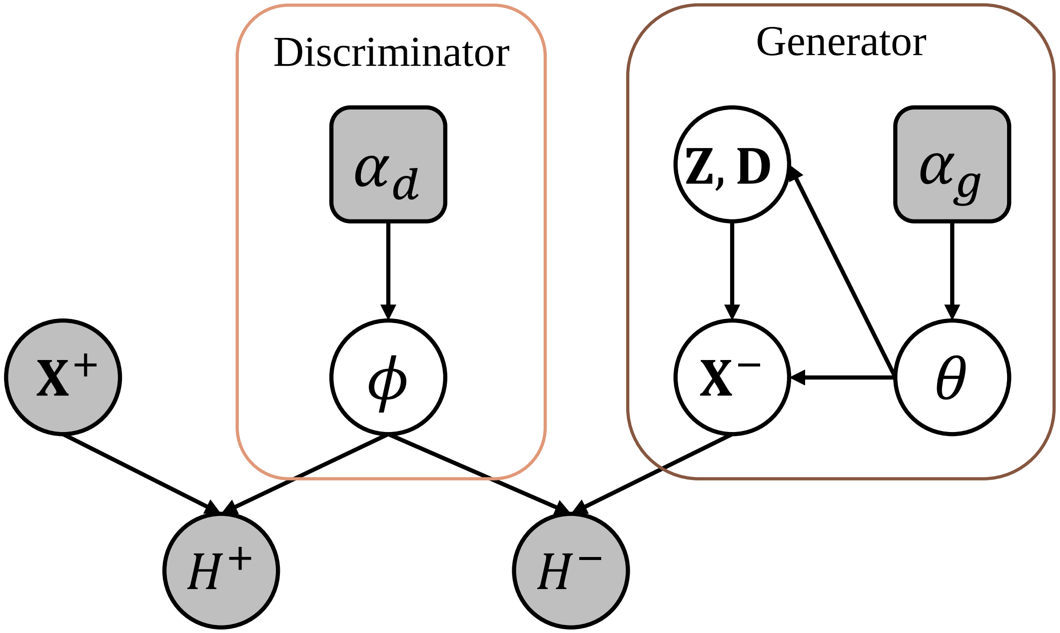


  - Overall objective

$$\min_{\theta} \max_{\phi} \; -\mathbb{E}_{P_{data}(\mathbf{X})}[H(\mathbf{X}|\phi)] + \mathbb{E}_{P(\mathbf{X}|\theta)}[H(\mathbf{X}|\phi)]$$

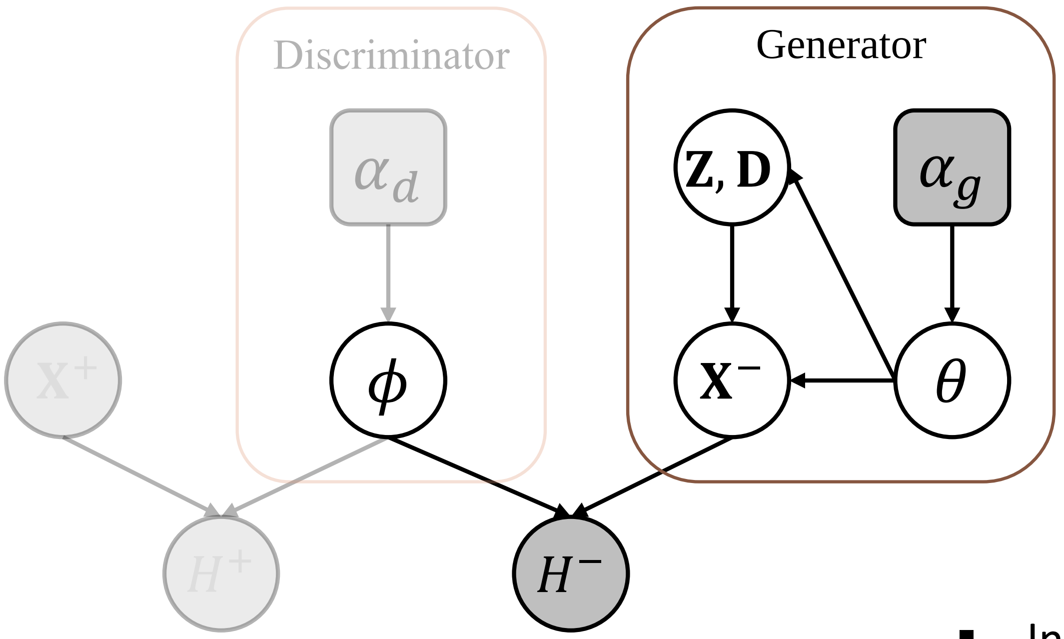
Methods: Adversarial Inference

- Bayesian adversarial inference $\theta, \phi \sim P(\theta, \phi | \mathcal{D}^+, \alpha_g, \alpha_d)$
  - Sample both generator θ and discriminator ϕ



Methods: Adversarial Inference

- Bayesian adversarial inference $\theta, \phi \sim P(\theta, \phi | \mathcal{D}^+, \alpha_g, \alpha_d)$
  - Sample generator θ conditioned on ϕ

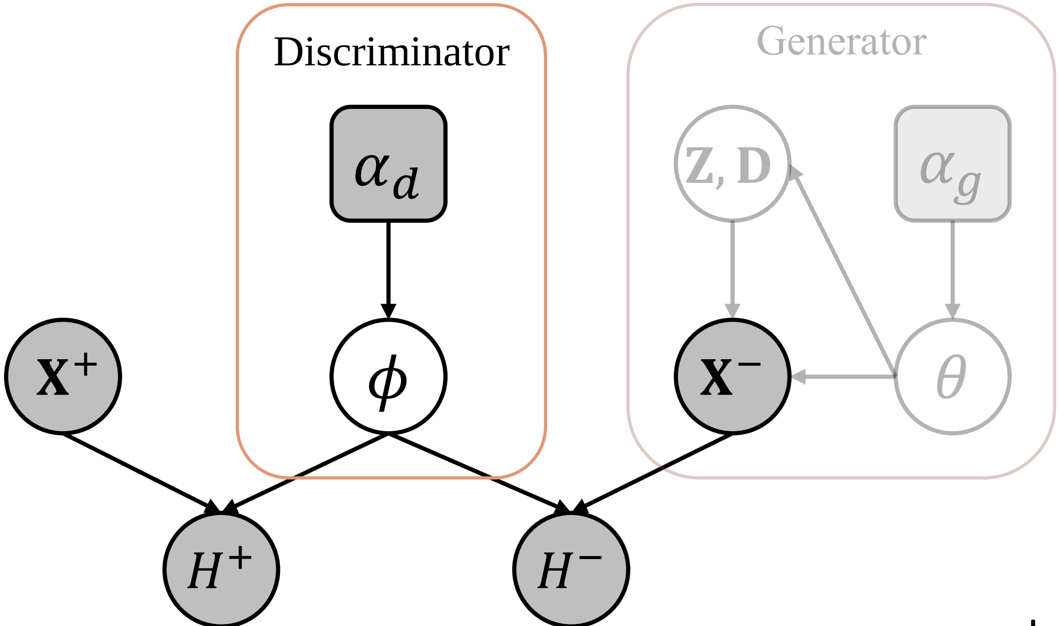


$$\theta \sim P(\theta | \mathcal{D}^+, \alpha_g, \alpha_d, \phi)$$
$$\propto \underbrace{\prod_i \exp\{-H(\mathbf{X}_i^- | \phi)\}}_{\text{likelihood}} \underbrace{P(\theta | \alpha_g)}_{\text{prior}}$$

- Inference method: Stochastic Gradient Hamiltonian Monte Carlo (SGHMC)

Methods: Adversarial Inference

- Bayesian adversarial inference $\theta, \phi \sim P(\theta, \phi|\mathcal{D}^+, \alpha_g, \alpha_d)$
  - Sample discriminator ϕ conditioned on θ



$$\phi \sim P(\phi|\mathcal{D}^+, \alpha_g, \alpha_d, \theta) \propto \underbrace{\prod_i \exp\{-H(\mathbf{X}_i^+|\phi)\} \prod_j \exp\{H(\mathbf{X}_j^-|\phi)\}}_{\text{Like-lihood}} \underbrace{P(\phi|\alpha_d)}_{\text{prior}}$$

- Inference method: Stochastic Gradient Hamiltonian Monte Carlo (SGHMC)

Methods

- Bayesian adversarial inference: data synthesis
  - Overall synthesis target:

$$\mathbf{X} \sim P(\mathbf{X}|\mathcal{D}^+, \alpha) = \int_\theta P(\mathbf{X}|\theta)P(\theta|\mathcal{D}^+, \alpha)d\theta$$

  - Steps:

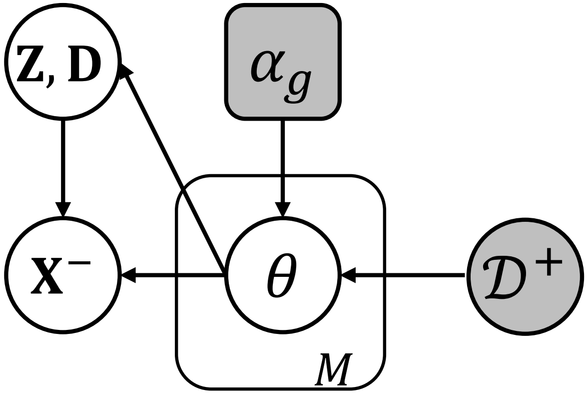
Posterior sampling

$$\{\theta_m, \phi_m\} \sim P(\theta, \phi|\mathcal{D}^+, \alpha)$$

Generate new data using all the $\{\theta_m\}$

$$\{\mathbf{D}_i, \mathbf{Z}_i\} \sim P(\mathbf{D}, \mathbf{Z}|\theta_m)$$

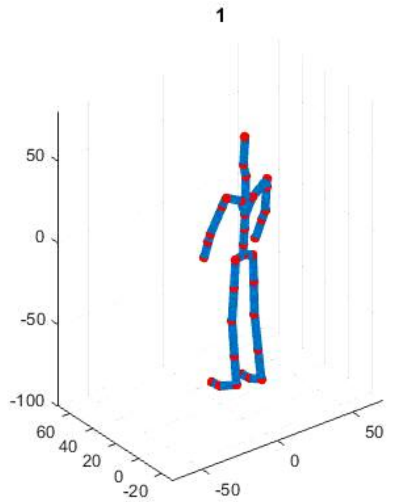
$$\mathbf{X}_i \sim P(\mathbf{X}|\mathbf{D}_i, \mathbf{Z}_i, \theta_m)$$



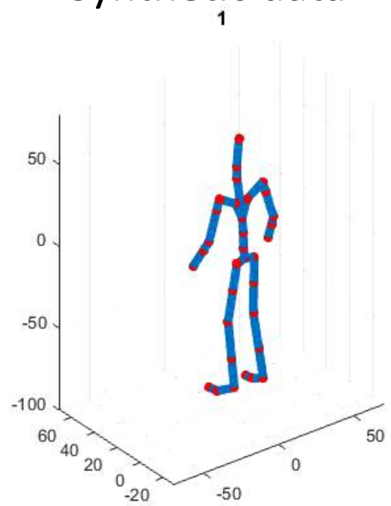
Experiments: Motion Synthesis

- Dataset
  - CMU Motion capture
  - Berkeley Motion capture
- Feature: Joint angles
- Qualitative Results

Real data

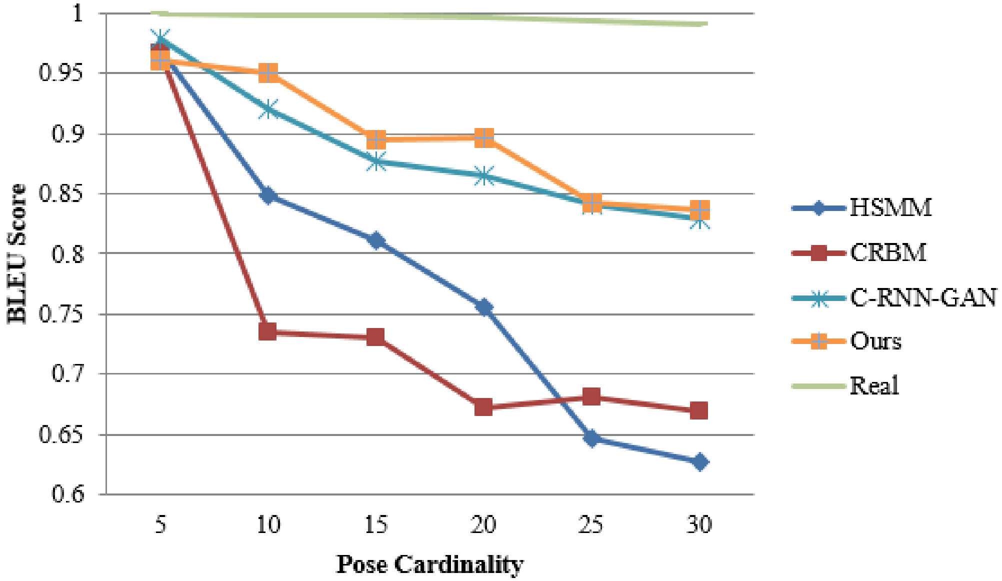


Synthetic data

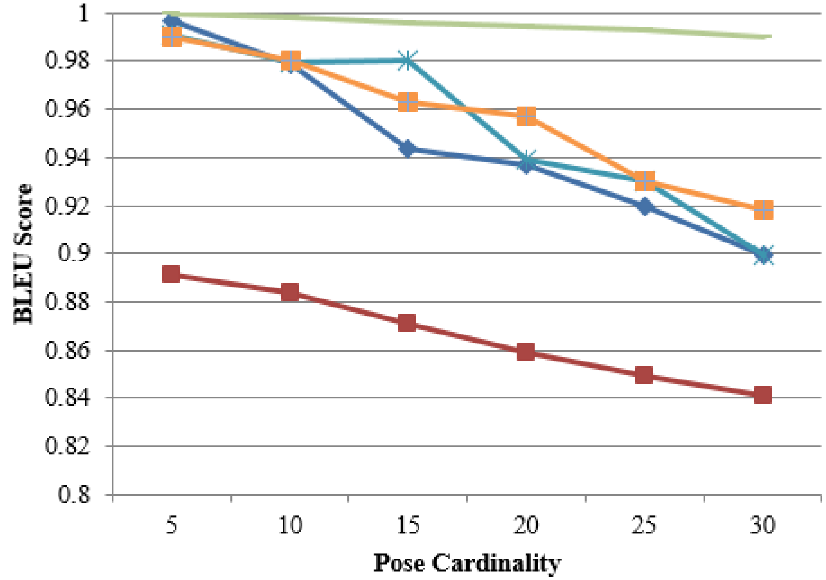


Experiments: Motion Synthesis

- Quantitative Results
  - Metric: BLEU score: fidelity



CMU



Berkeley

Experiments: Motion Synthesis

- Quantitative Results

Inception score: diversity

Method	CMU	Berkeley
HSMM	1.86 ± 0.07	4.99 ± 0.27
CRBM[14]	2.65 ± 0.09	5.24 ± 0.39
TSBN[6]	2.58 ± 0.04	2.57 ± 0.14
C-RNN-GAN[9]	1.95 ± 0.03	4.56 ± 0.37
Ours	$\mathbf{2.86} \pm 0.10$	$\mathbf{6.49} \pm 0.23$
Real	*2.96*	*8.79*

MMD: distribution-level similarity

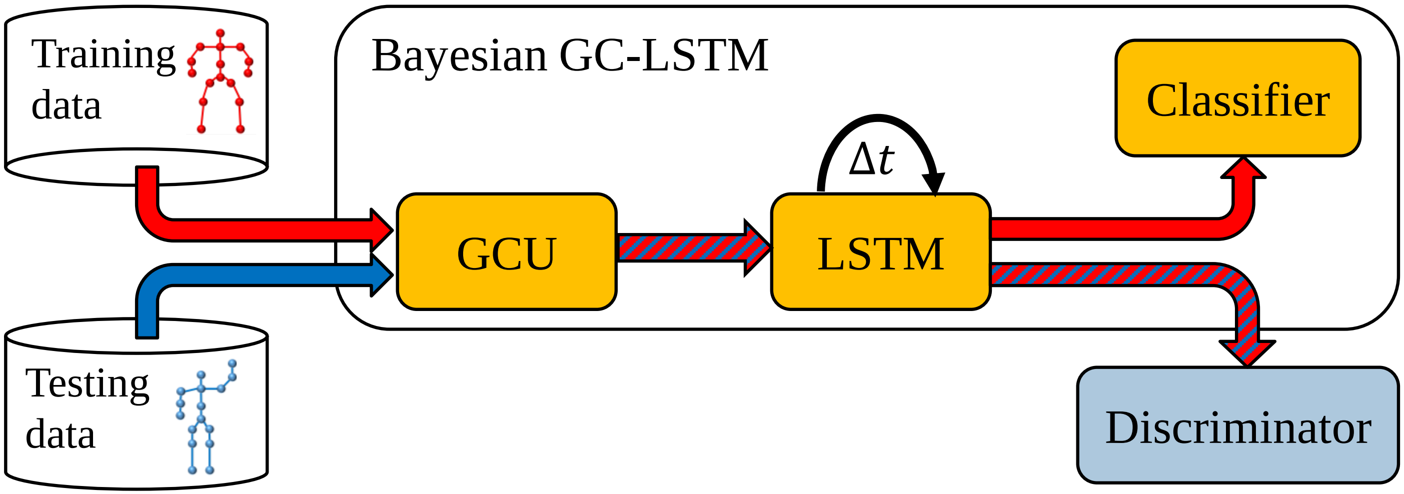
Method	CMU	Berkeley
HSMM	5.46 ± 0.62	432.25 ± 0.78
CRBM[14]	7.43 ± 0.97	55.39 ± 0.75
TSBN[6]	12.74 ± 0.10	110.55 ± 0.64
C-RNN-GAN[9]	10.58 ± 0.35	83.25 ± 0.96
Ours	$\mathbf{2.41} \pm 0.35$	$\mathbf{48.70} \pm 0.11$
Random	$\mathit{176.27 \pm 0.05}$	$\mathit{1089.91 \pm 0.10}$

Part 4: Modeling Complex Dynamics

- Challenge:
  - Structural dependency
  - Long-term temporal dependency
  - Uncertainty and large variations
- Application:
  - Action recognition
- Our solution: Bayesian Graph Convolution LSTM

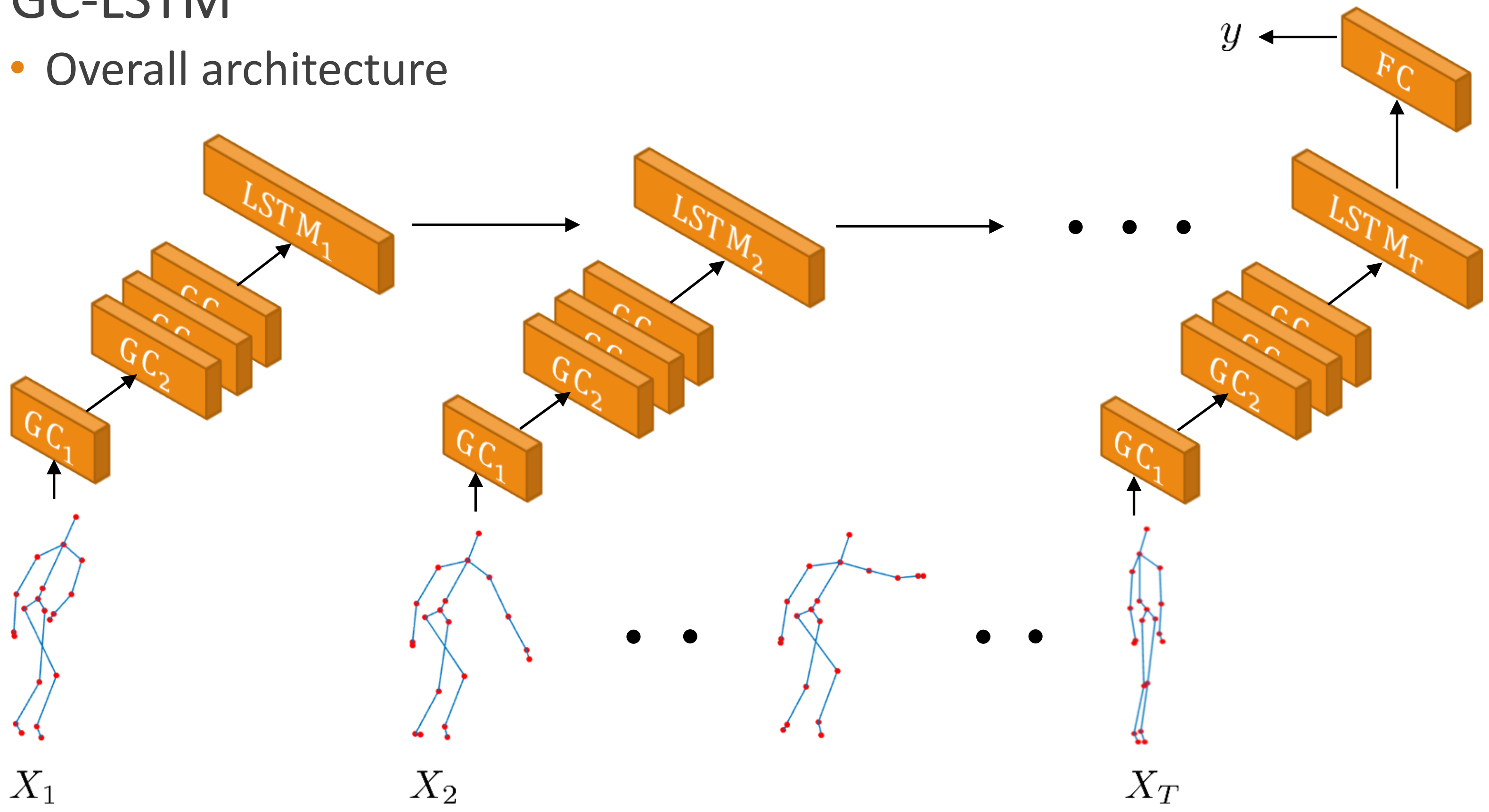
Methods

- Overall framework:



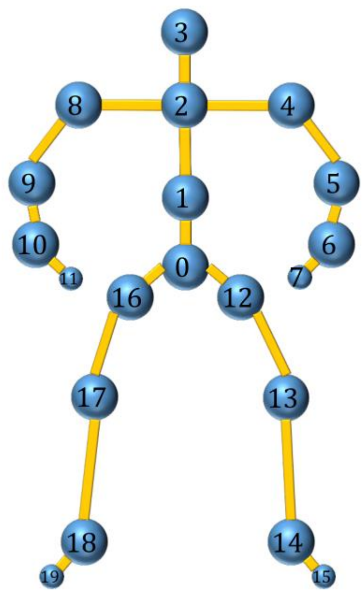
Methods

- GC-LSTM
  - Overall architecture

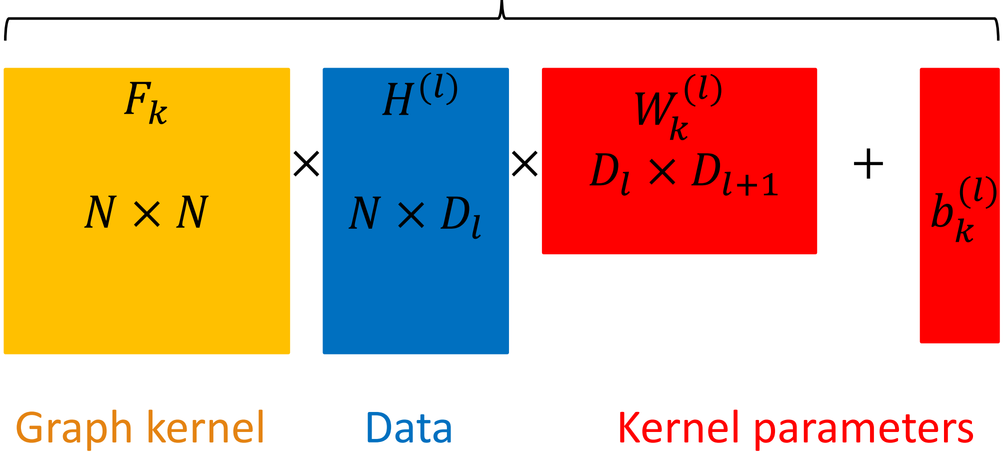


Methods

- GC-LSTM
  - Graph convolution: generalize convolution to arbitrary graph structured data



$$H^{(l+1)} = \sigma(\sum_{k=1}^{K} F_k H^{(l)} W_k^{(l)} + b_k^{(l)})$$

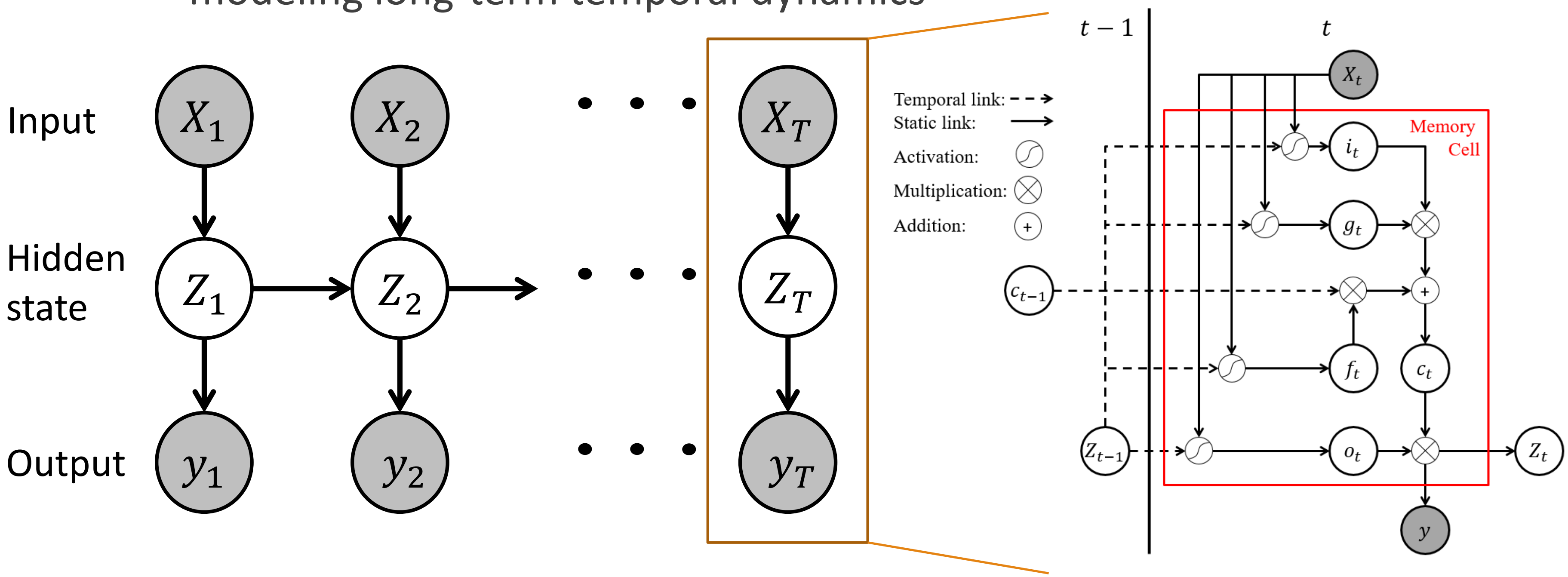


N: number of nodes
D: dimension of each node

Methods

$$i_t = \sigma(W_{xi}X_t + W_{zi}Z_{t-1} + b_i)$$
$$f_t = \sigma(W_{xf}X_t + W_{zf}Z_{t-1} + b_f)$$
$$o_t = \sigma(W_{xo}X_t + W_{zo}Z_{t-1} + b_o)$$
$$g_t = \phi(W_{xg}X_t + W_{zg}Z_{t-1} + b_g)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$
$$Z_t = o_t \odot \phi(c_t)$$

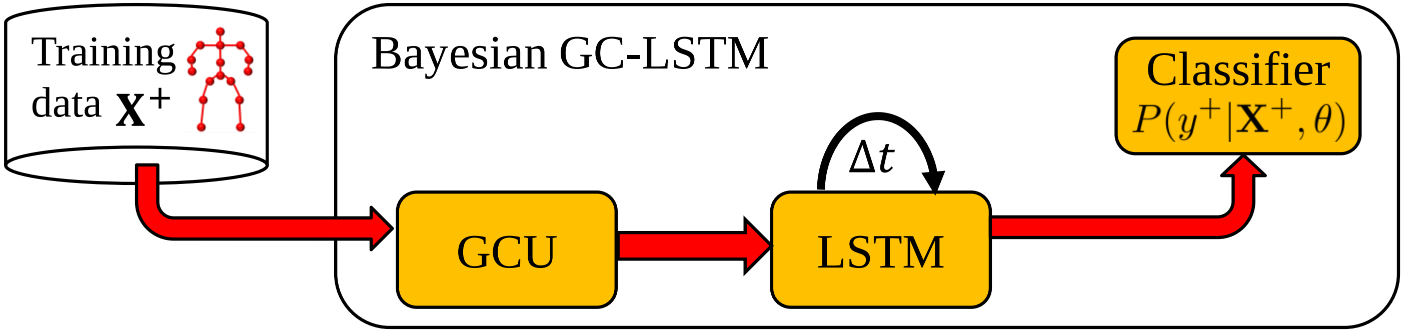
- GC-LSTM
  - Long short-term memory (LSTM) network: modeling long-term temporal dynamics



Methods

- Bayesian GC-LSTM
  - Extend GC-LSTM to a probabilistic model: $\theta \sim P(\theta|\alpha_\theta)$
  - Infer the posterior distribution of parameters

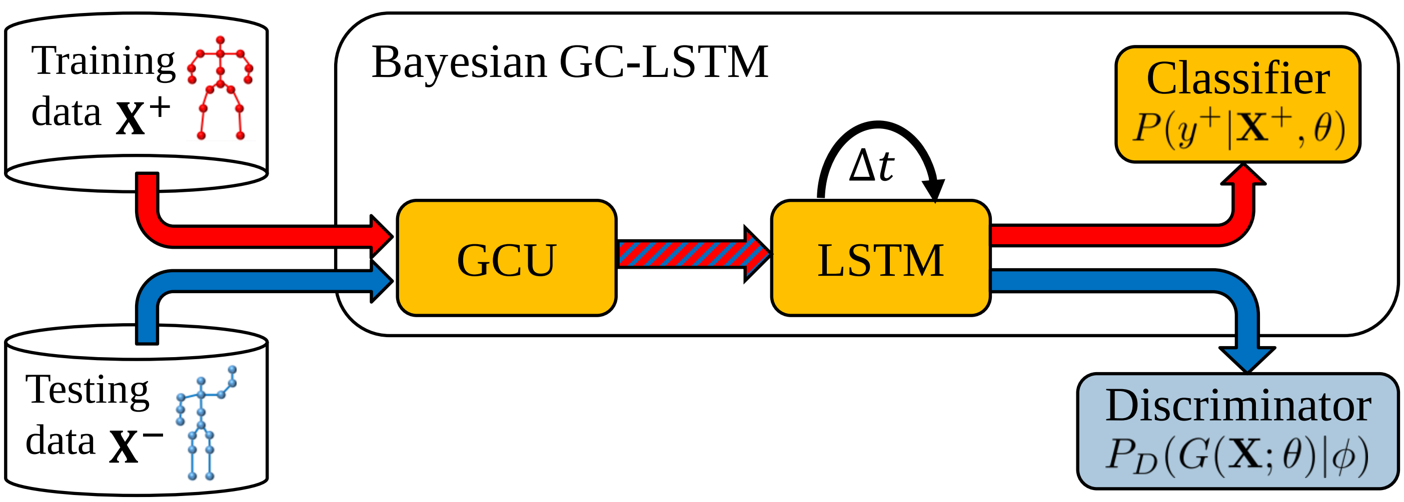
$$\log P(\theta|\mathcal{D}, \alpha_\theta) = \underbrace{\log P(y^+|\mathbf{X}^+, \theta)}_{\text{likelihood}} + \underbrace{\log P(\theta|\alpha_\theta)}_{\text{prior}} + C$$



Methods

- Bayesian GC-LSTM
  - Adversarial Prior: use additional discriminator to regularize parameters
  - Intuition: promote a feature representation to be invariant of subject

$$\log P(\theta|\mathcal{D}, \phi, \alpha_\theta) = \log P(y^+|\mathbf{X}^+, \theta) + \log P(\theta|\alpha_\theta) + \log P_D(G(\mathbf{X}^-; \theta)|\phi) + C$$



Methods

- Bayesian Inference

$$P(y'|\mathbf{X}', \mathcal{D}, \alpha) = \int_\theta P(y'|\mathbf{X}', \theta)P(\theta|\mathcal{D}, \alpha)d\theta$$
$$\approx \frac{1}{M}\sum_{m=1}^{M} P(y|\mathbf{X}', \theta_m), \theta_m \sim P(\theta|\mathcal{D}, \alpha)$$

  - Classification:

$$y^* = \arg\max_{y'} \frac{1}{M}\sum_{m=1}^{M} P(y'|\mathbf{X}', \theta_m)$$

Experiments

- Ablation study

Effect of graph convolution

Configuration	# of edges	Accuracy
No graph	N/A	85.3
Mean-field	0	82.5
Local graph	19	87.5
Global graph	10	81.8
Joint graph	29	**92.3**

Effect of Bayesian inference

Perturbation	Clean	Only R	Only N	R + N
ML	86.2	62.8	77.7	65.1
MAP	85.2	78.1	86.1	77.9
Bayesian	87.4	78.8	86.9	82.8
Bayesian + AP	**92.3**	**86.0**	**87.9**	**86.1**

R: random rotation
N: random noise

Experiments

- Comparison with state-of-the-art

MSR Action3D

Method	Accuracy
SC [10]	88.3
HBRNN [6]	94.5
Composition [12]	93.0
ST-LSTM [13]	**94.8**
Ours	94.6

SYSU

Method	Accuracy
D-Skeleton [9]	75.5
ST-LSTM [13]	76.5
DPRL [20]	76.9
SR-TSL [18]	80.7
Ours	**81.7**

UTD MHAD

Method	Accuracy
Sensor Fusion [3]	79.1
DMM-LBP [1]	84.2
3DHoT-MBC [26]	84.4
SOS-CNN [8]	87.0
Ours	**92.3**

Experiments

- Generalization across different datasets

Train	MSR	UTD	Avg.
Test	UTD	MSR	
R3DG [15]	66.5	59.9	63.2
DLSTM [19]	66.8	50.0	58.4
Ours	**82.8**	**70.2**	**76.5**

Thesis Summary

- Localization of dynamic pattern
  - Propose a method that combines robust estimation and dynamic model for localization
- Dynamic pattern regression under insufficient annotation
  - Incorporate ordinal information as addition constraints for model learning
  - Develop an optimization algorithm for parameter estimation
- Dynamic pattern classification and synthesis under large intra-class variation
  - Propose a Bayesian hierarchical model
  - Develop two Bayesian inference algorithms
- Modeling complex dynamics
  - Propose a Bayesian neural network model
  - Develop a Bayesian inference algorithm

Thank You!

- Related Publications:
  - **Rui Zhao**, Quan Gan, Shangfei Wang and Qiang Ji, Facial Expression Intensity Estimation Using Ordinal Information, CVPR 2016.
  - **Rui Zhao**, Md Ridwan Al Iqbal, Kristin Bennett and Qiang Ji, Wind Turbine Fault Prediction Using Soft Label SVM, ICPR 2016.
  - **Rui Zhao**, Gerwin Schalk, Qiang Ji, Robust Signal Identification for Dynamic Pattern Classification, ICPR 2016.
  - **Rui Zhao,** Qiang Ji, An Adversarial Hierarchical Hidden Markov Model for Human Pose Modeling and Generation, AAAI 2018.
  - **Rui Zhao**, Gerwin Schalk, Qiang Ji, Temporal Pattern Localization using Mixed Integer Linear Programming, ICPR 2018.
  - **Rui Zhao,** Qiang Ji, An Empirical Evaluation of Bayesian Inference Methods for Bayesian Neural Networks, NIPS Workshop 2018 (To appear)
  - **Rui Zhao,** Wanru Xu, Hui Su, Qiang Ji, Bayesian Hierarchical Dynamic Model for Human Action Recognition (Under review)
  - **Rui Zhao,** Hui Su, Qiang Ji, Bayesian Adversarial Human Motion Synthesis (Under review)
  - **Rui Zhao,** Kang Wang, Hui Su, Qiang Ji, Bayesian Graph Convolution LSTM for Skeleton based Action Recognition (Under review)