UNSUPERVISED AUTO-ENCODING MULTIPLE-OBJECT TRACKER FOR CONSTRAINT-CONSISTENT COMBINATORIAL PROBLEM

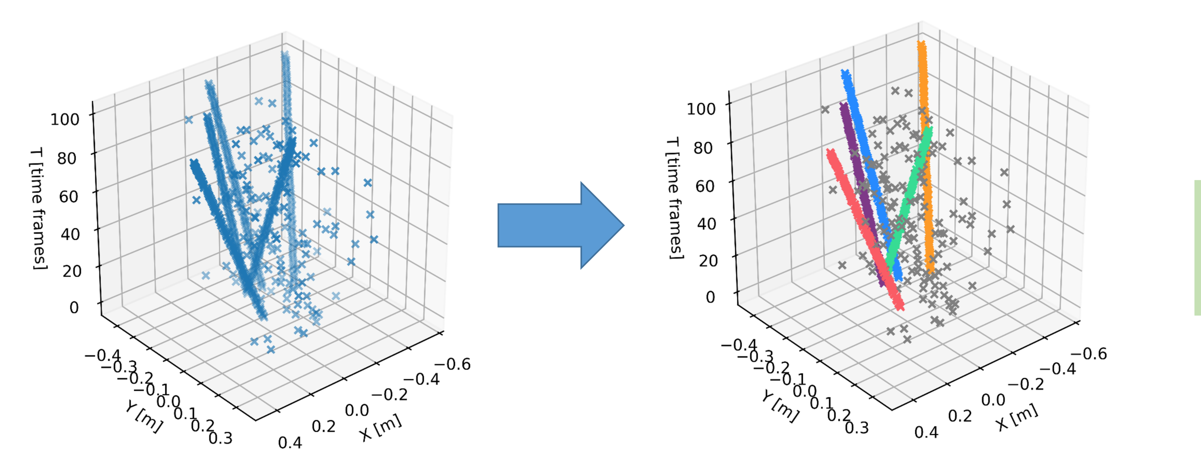
ICASSP2020

Yuta Kawachi and Teppei Suzuki (DENSO IT Laboratory, Inc., Tokyo, Japan)

Introduction

Multiple-object tracking (MOT) is a key component for RADAR or LIDAR point cloud applications. MOT includes an association problem which binds the observation order and the tracker order under the existence of missing and noisy observation points. This condition is expressed as the (at-most-) one-to-one constraint of the assignment matrix.

Multiple object tracking (MOT)



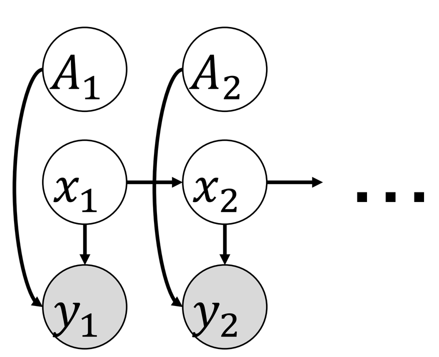
Raw observation point cloud sample

Answer labels (colors are tracks)

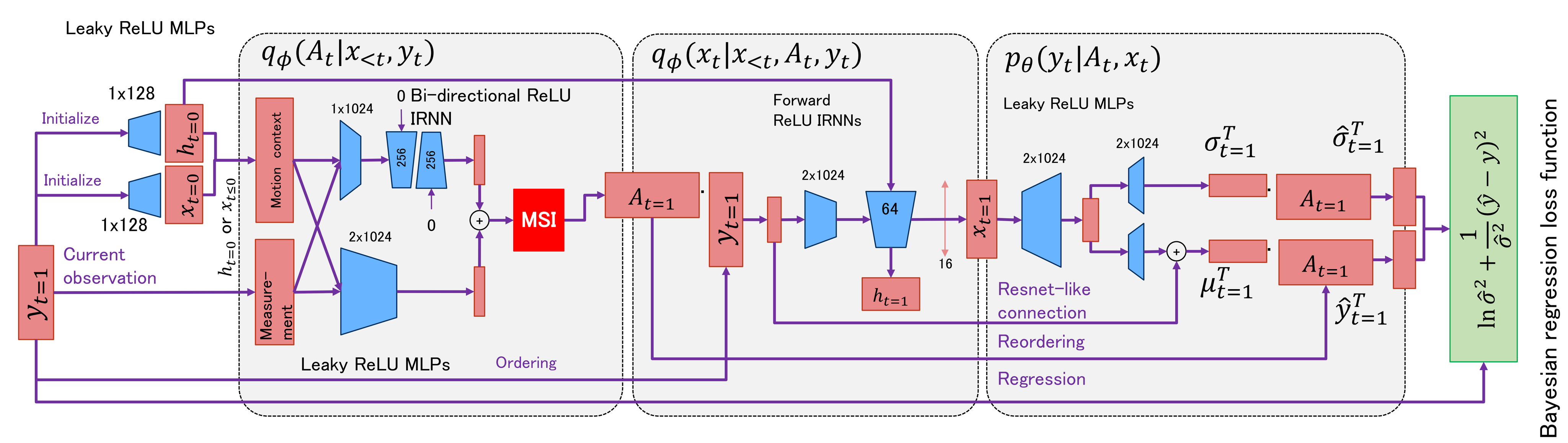
In this work,

- We derive a probability-based MOT formulation for learning unsupervised auto-encoding neural networks to solve the tracker-observation association problem.
- We propose a modified Sinkhorn iteration algorithm to obtain a differentiable delegate of assignment matrix for the tracker-observation association problem.

Model



Generative model of MOT



Derived auto-encoding NW of MOT (at t=1)

The joint probability is defined as:

$$p(\mathrm{A}, \mathrm{X}, \mathrm{Y}) = \prod_t p(\mathrm{A}_t) \prod_t p(\mathrm{x}_t|\mathrm{x}_{t-1}) \prod_t p(\mathrm{y}_t|\mathrm{A}_t, \mathrm{x}_t)$$

Maximum marginal likelihood criterion is:

$$\mathrm{argmax}_\theta\, p_\theta(Y)$$

We introduce amortized posterior $q_\phi(A, X|Y)$, continuous relaxation to A, omit the KL term to obtain an auto-encoder, then the objective is:

$$\theta_{AE}, \phi_{AE} = \mathrm{argmin}_{\theta,\phi} \mathcal{L}_{RL}(\theta, \phi)$$

$$\mathcal{L}_{RL}(\theta, \phi) = -\int_A \int_X q_\phi(A, X|Y) \log p_\theta(Y|A, X)\, d\mathrm{X} d\mathrm{A}$$

Time index $\mathrm{t} \in \mathbb{N}$
Max. time index $\mathrm{T} \in \mathbb{N}$
Number of features $\mathrm{D} \in \mathbb{N}$
Number of observations $\mathrm{M_t} \in \mathbb{N}$
Max. number of observations $M \in \mathbb{N}$
Number of trackers $\mathrm{K} \in \mathbb{N}$
Observations $\mathrm{y_t} \in \mathbb{R}^{\mathrm{M_t} \times \mathrm{D}}$, $Y \in \mathbb{R}^{\mathrm{T} \times \mathrm{M_t} \times \mathrm{D}}$
Trackers $\mathrm{x_t} \in \mathbb{R}^{\mathrm{K} \times \mathrm{D}}$, $\mathrm{X} \in \mathbb{R}^{\mathrm{T} \times \mathrm{K} \times \mathrm{D}}$
Assignment $\mathrm{A_t} \in \{0,1\}^{(\mathrm{K}+1) \times (\mathrm{M_t}+1)}$
s.t. $\sum_{\mathrm{i} \neq 0} \mathrm{A_t[i,j]} = 1$, $\sum_{\mathrm{j} \neq 0} \mathrm{A_t[i,j]} = 1$,
$A \in \{0,1\}^{T \times (\mathrm{K}+1) \times (M+1)}$

We further approximate the posterior and omit the sampling procedure, then the final forms are: $A_t \leftarrow q_\phi(A_t|x_{<t}, y_t)$, $\quad x_t \leftarrow q_\phi(x_t|x_{<t}, A_t \cdot y_t)$

$$\mathcal{L}_{RL}(\theta, \phi) = -\sum_t \log p_\theta(\tilde{y}_t|x_t), y_t = {A_t}^T \cdot \tilde{y}_t$$

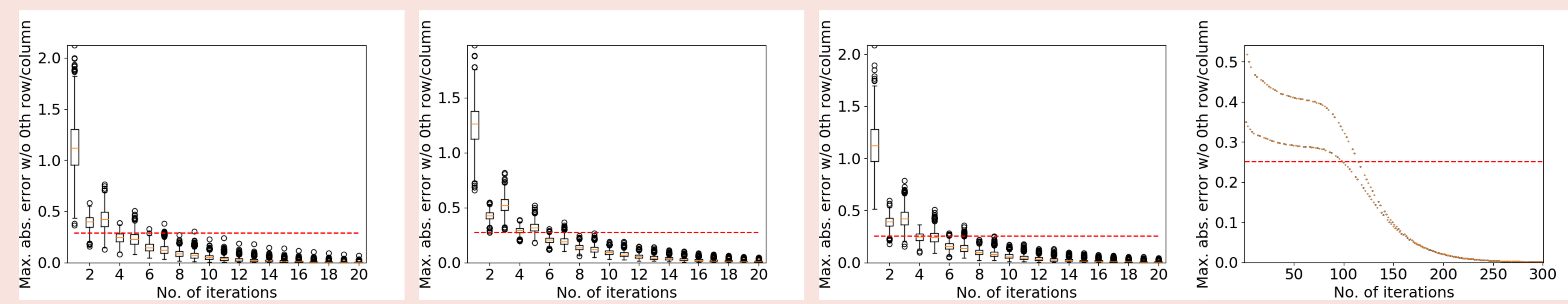
Modified Sinkhorn iteration (MSI)

MSI iteratively normalizes row and column in order to comply with the (at-most-) one-to-one constraint.

$$\mathrm{A_t}[i,j] \leftarrow \frac{\mathrm{A_t}[i,j]}{\sum_{\mathrm{i}} \mathrm{A_t}[i,j]}\ (j \neq 0),\ \ \mathrm{A_t}[i,j] \leftarrow \frac{\mathrm{A_t}[i,j]}{\sum_{\mathrm{j}} \mathrm{A_t}[i,j]}\ (i \neq 0)$$

Experiment

1. MSI verification experiment using random matrices



5x7 8x11 6x8 6x8 (corner case)

MSI empirically converges to constraint-consistent solution. The red dotted line indicates the lower bound of the conventional rectangular Sinkhorn iteration. The vertical axes correspond to the maximum number of false multiple or no associations which surpass or do not reach the constraint.

2. Assignment generation explerirment

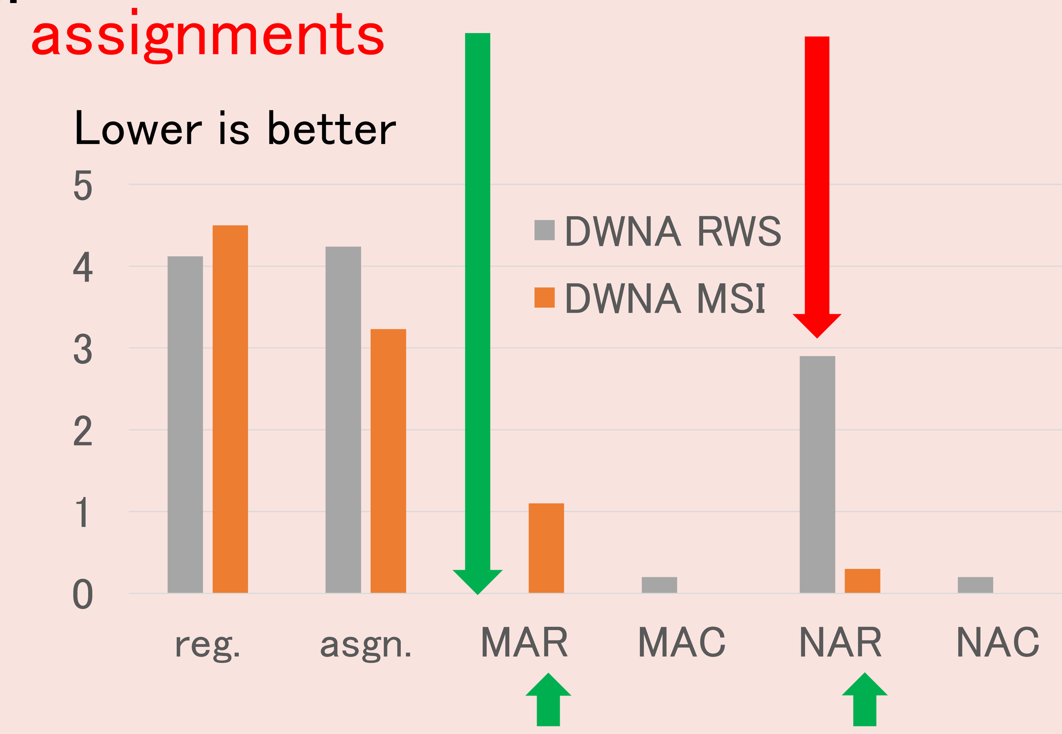
- Synthetic point-cloud from three motion models with clutters and missing detections.
- Comparing MSI and row-wise softmax (RWS) under the same model.
- Evaluation metrics: regression error (reg.), assignment L2 error (asgn.), no assignment error for row/column (NAR/ NAC), multiple assignment error for row/column (MAR/MAC)

The evaluation result under the dataset of discrete white noise acceleration model (DWNA) is shown in the right graph.

Conventional row-wise softmax

✓reduces multiple assignments

✕produces lots of false no assignments



✓Proposed MSI emits more balanced solutions

Conclusion

We proposed an unsupervised neural MOT algorithm for accurate assignment generation application. Experimental results demonstrated that our modified Sinkhorn iteration outputs a more (at-most-) one-to-one constraint-consistent rectangular assignment matrix than the previous row-wise softmax method.