2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Tutorial on Deep Probabilistic Generative Models for Robotics

Introduction

2020.10.25 IROS2020 on demand

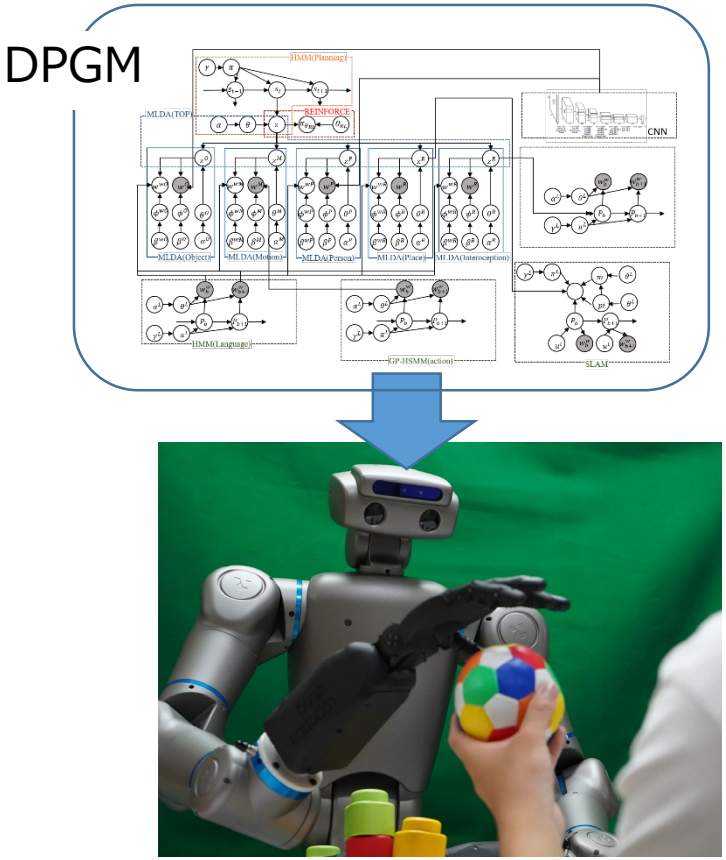
Organized by

Takayuki Nagai, **Osaka University**
Tadahiro Taniguchi, Ritsumeikan University
Takato Horii, Osaka University
Chie Hieida, Nara Institute of Science and Technology
Kaede Hayashi, Ritsumeikan University

978-1-7281-6211-9/20/$31.00 ©2020 IEEE

Welcome!

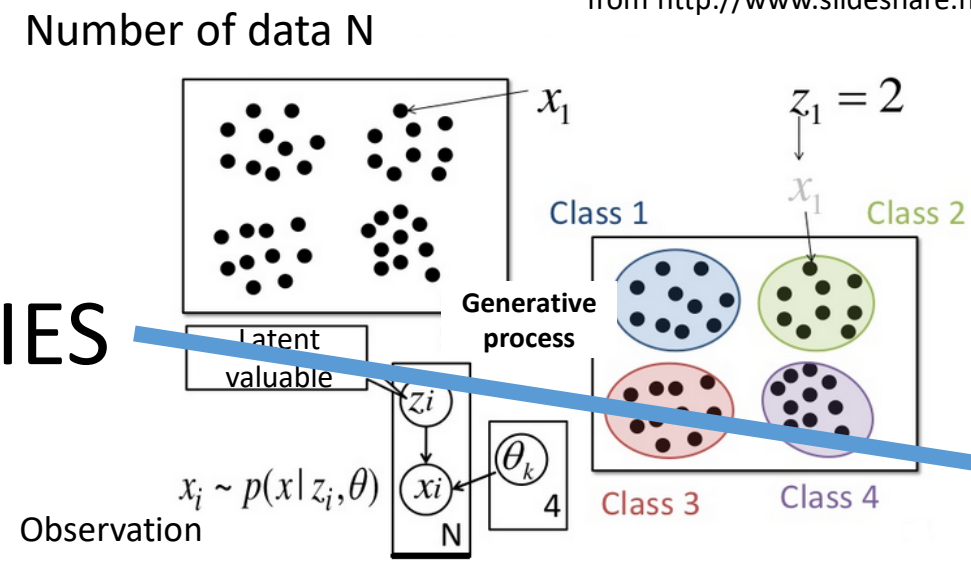
- Introduction
  - What is Deep Probabilistic Generative Models (DPGM)?
  - Why should we learn DPGM
  - What to learn?
    - Theoretical side (2 talks)
      - Prof. Taniguchi
      - Dr. Okada
    - Implementation side (2 talks)
      - Prof. Suzuki
      - Prof. Nakamura
- Information on this tutorial
  - HP, supplemental materials



Probabilistic generative models

Use these "generative models" for developing intelligent robots!

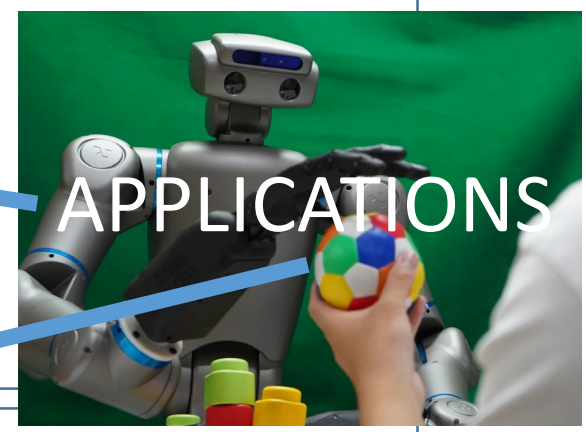
from http://www.slideshare.net/issei_sato/deim2012-issei-sato



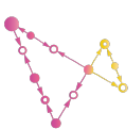
Probabilistic Generative Models

THEORIES

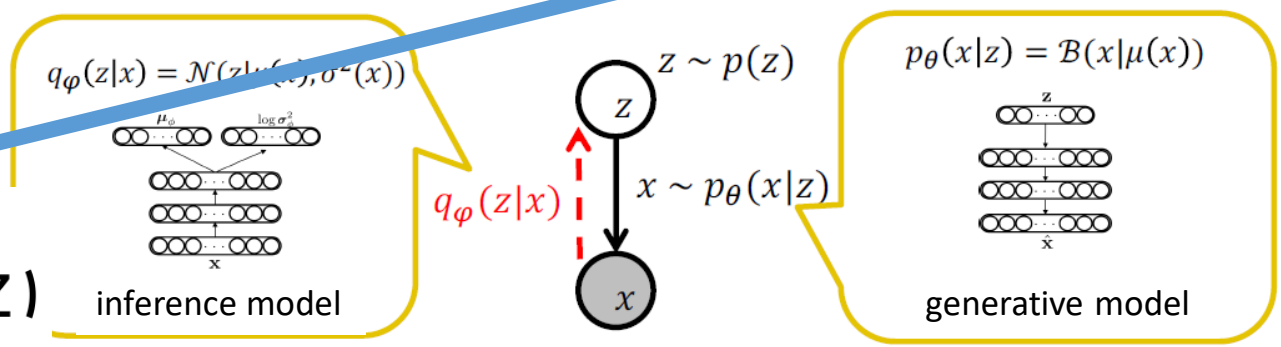
APPLICATIONS



TOOLS



Pixyz

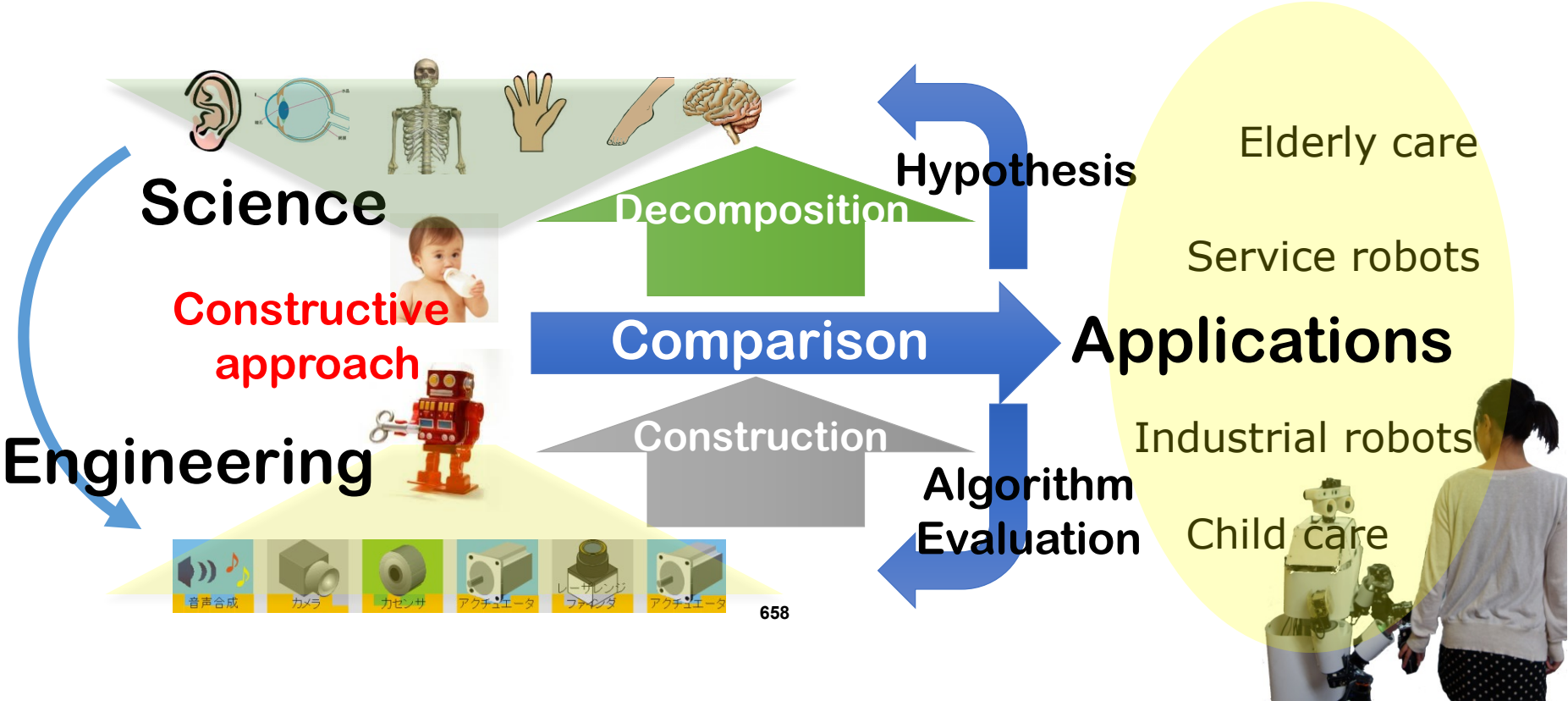


657

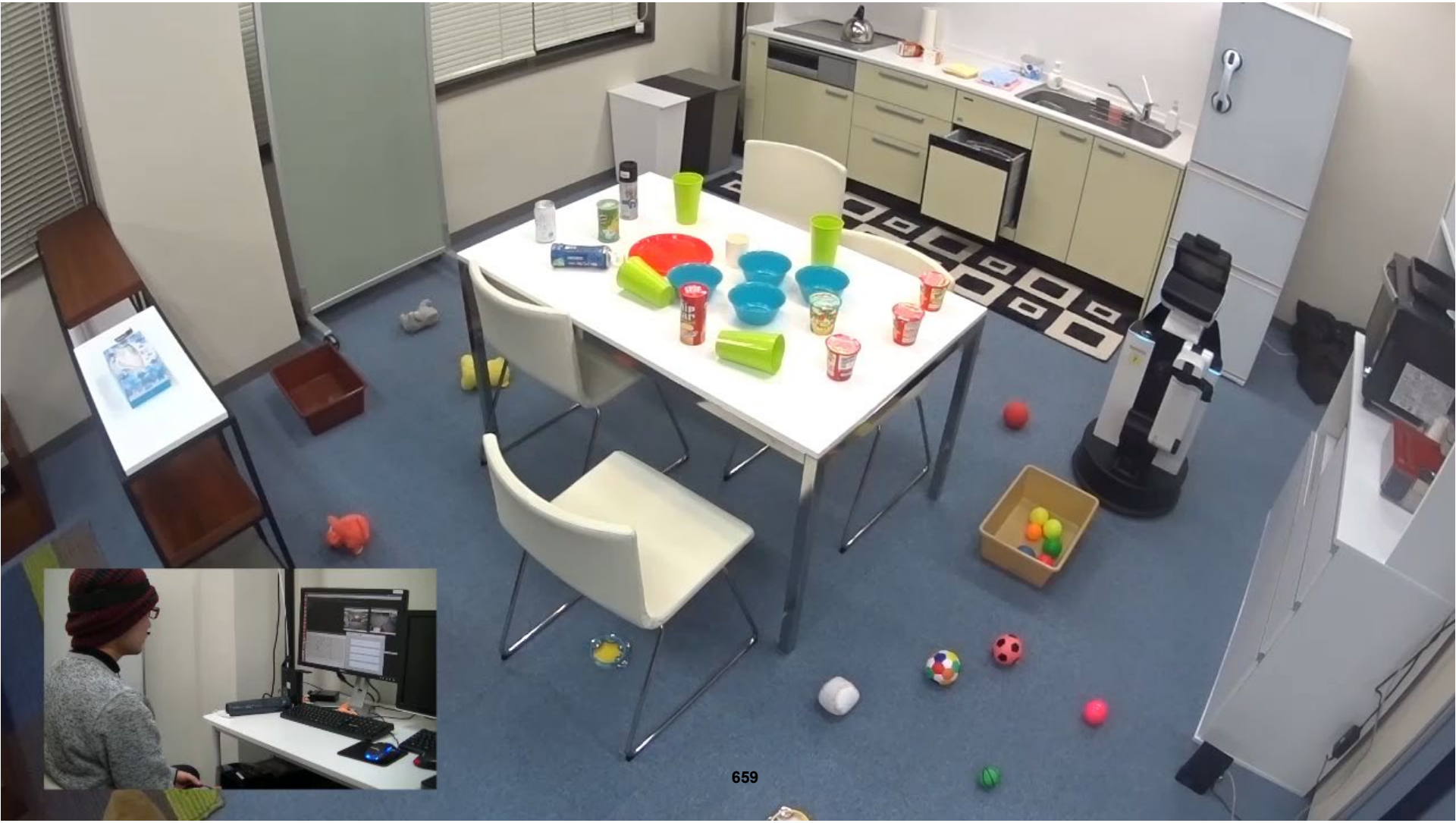
"Deep" Probabilistic Generative Models (DPGMs)

We are interested in cognitive robotics
Constructive approach

- **<u>Constructive Human Science</u>**
  - ⇒ Construct to know mechanisms behind our mind
  - ⇒ Construct to use it for specific applications



Domestic service applications

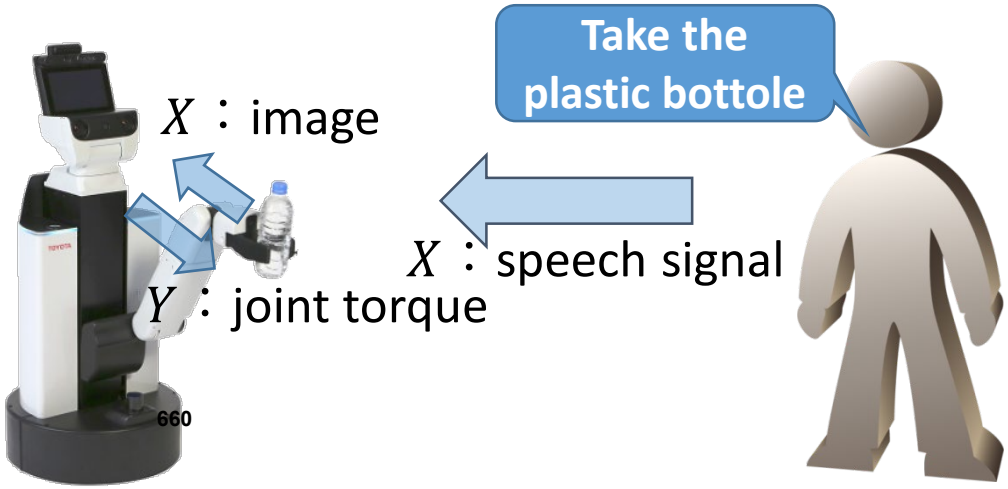


What is the problem ?

- Problem of building general intelligence
  - What is the essence of intelligence based on its own body?
  - We want to build a computational model and to implement

The ability required for the robot is to predict the (continuous) pattern *Y* that generates its own action from the input (continuous) pattern *X*

$$p(Y|X)$$



Approaches

◻ Neural Nets

$$p(Y|X)$$

X

Supervised
End to end learning

Y

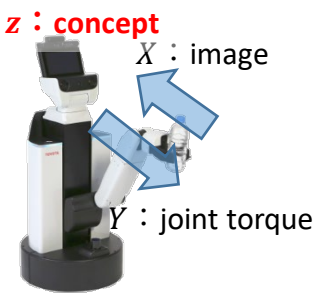
◻ Bayesian model

$$p(Y|X) = \sum_z p(Y,z|X) = \sum_z \underbrace{p(Y|z)}_{\text{action}}\underbrace{p(z|X)}_{\text{Rec./Rep.}}$$

X

z is a latent variable (concept),
which is acqired by the robot itself
=> **unsupervised learning**

Y



◻ Pipeline

$$p(Y|X) \approx p(Y|\mathrm{argmax}\, p(z|X))$$
$$= p(Y|\bar{z})$$

AI and/or Robot

$$\bar{z} = \mathrm{argmax}\, \frac{p(X|z)p(z)}{p(X)}$$

Pattern recognition

Problem of finding the label z corresponding to X
z is defined by hand⇒ Training the classfier $p(X|z)$
Required labeled data

When $\bar{z}$ is selected by the classifier, an action must be selected according to the result
Hard to design $p(Y|\bar{z})$

What is understanding?

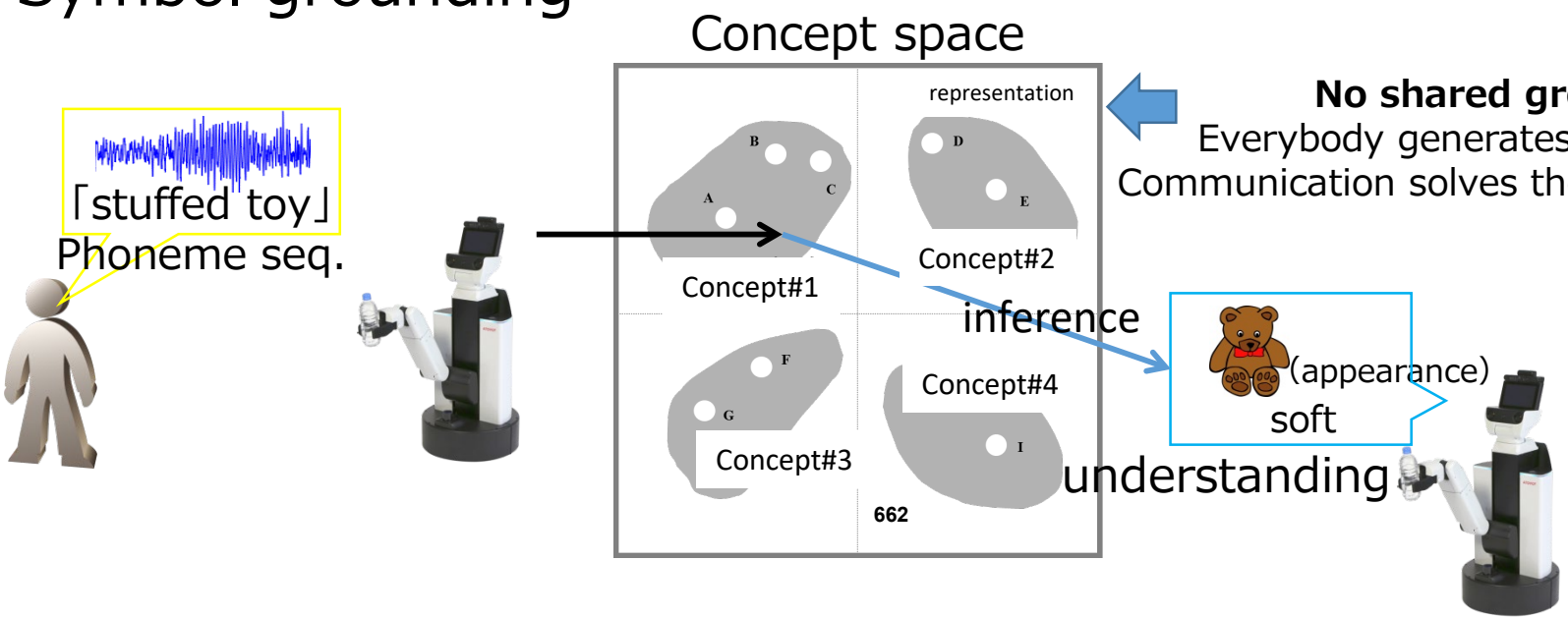
- Robot understanding of real world

  "Understanding" : prediction of unobservable information through concepts

  "Meaning" : predicted contents

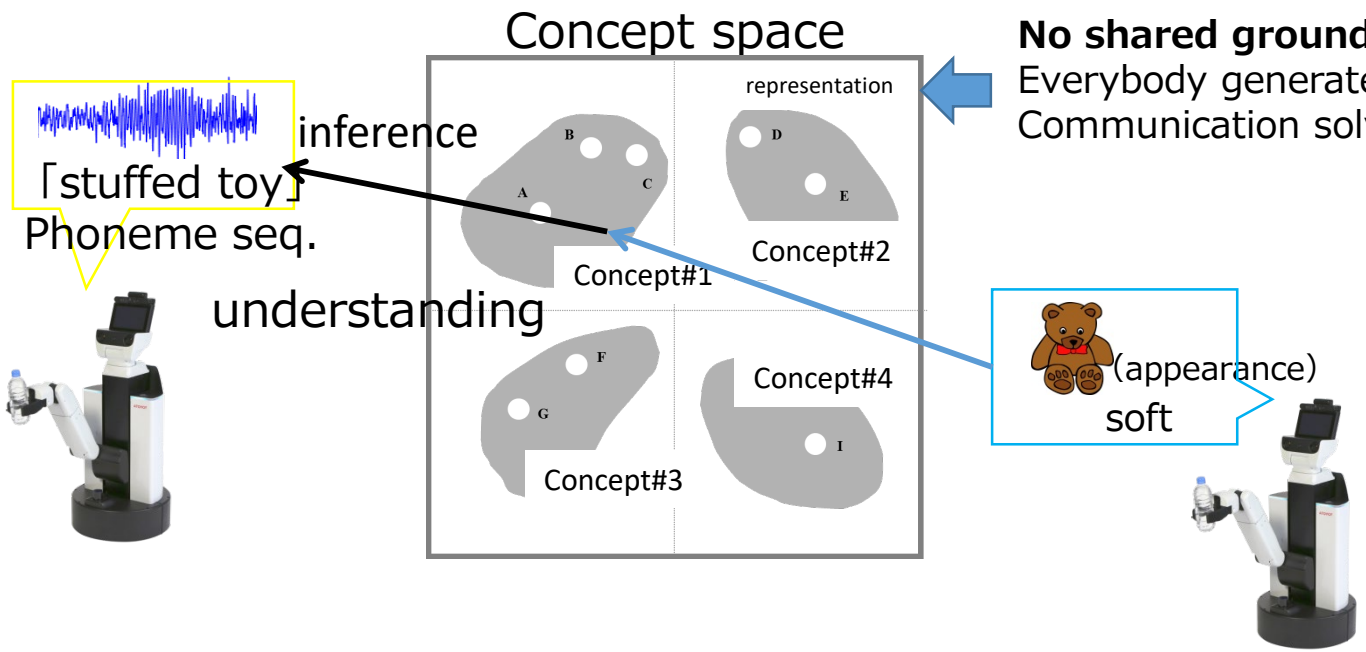
  "Concept" : multimodal categorization generates concepts (categories)

- Symbol grounding



No shared ground truth
Everybody generates own space
Communication solves the mismatch

Counter direction



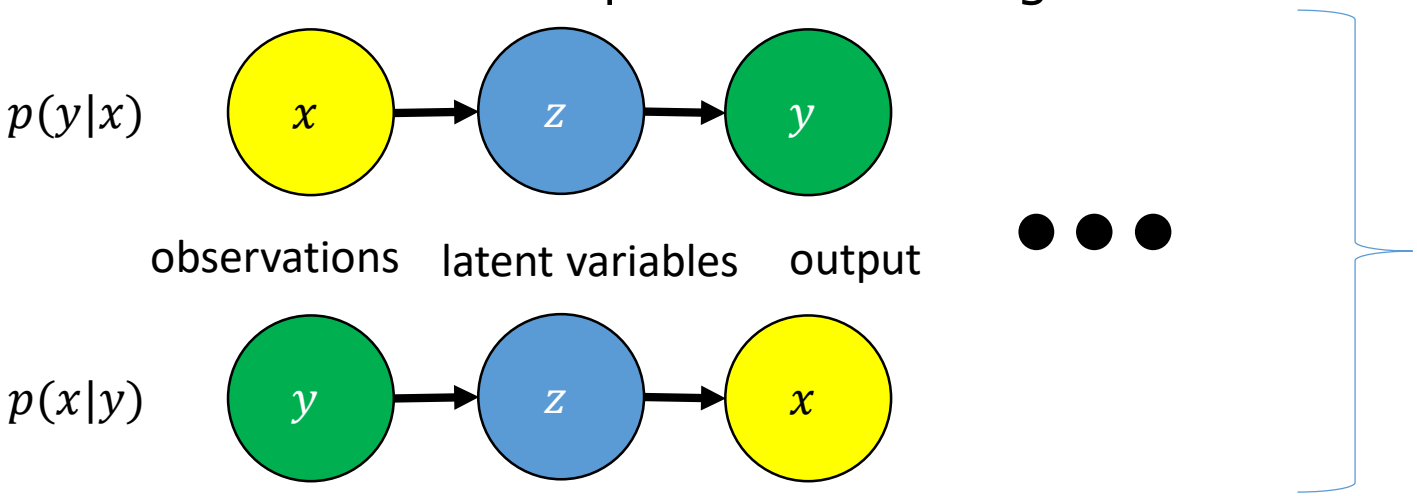
No shared ground truth
Everybody generates own space
Communication solves the mismatch

- Not modeling $p(y|x)$ alone, but modeling joint probabilities

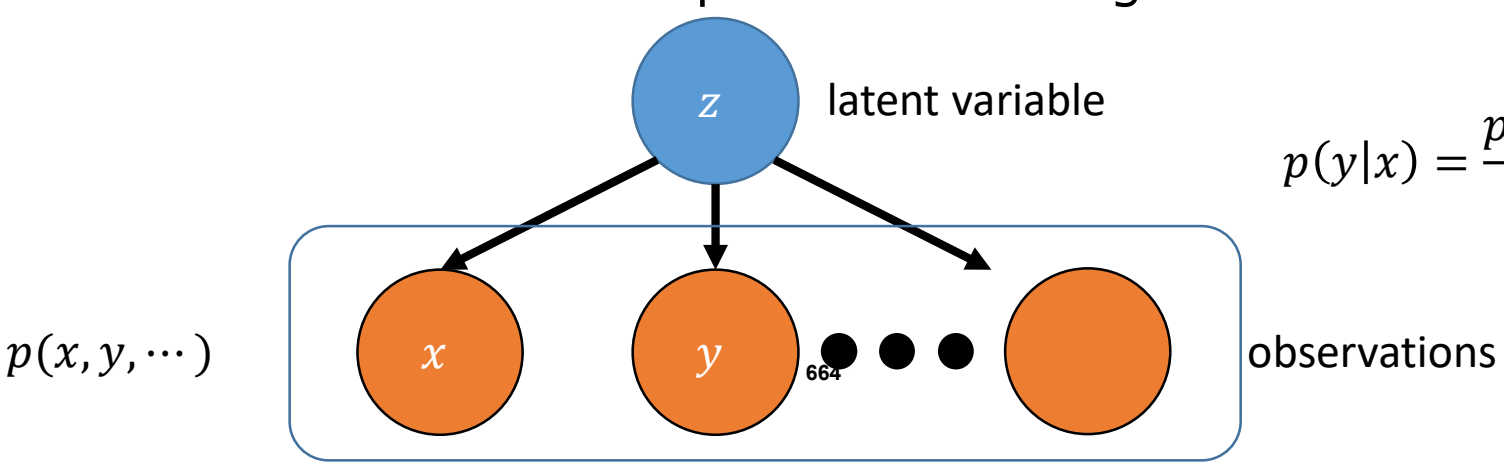
$$p(x, y, \cdots) = \sum_{z} p(x, y, \cdots | z) p(z)$$

Multimodal Generative Models

Multimodal supervised learning



Multimodal unsupervised learning



$$p(y|x) = \frac{p(x,y)}{p(x)}$$

Basics of DPGMfR

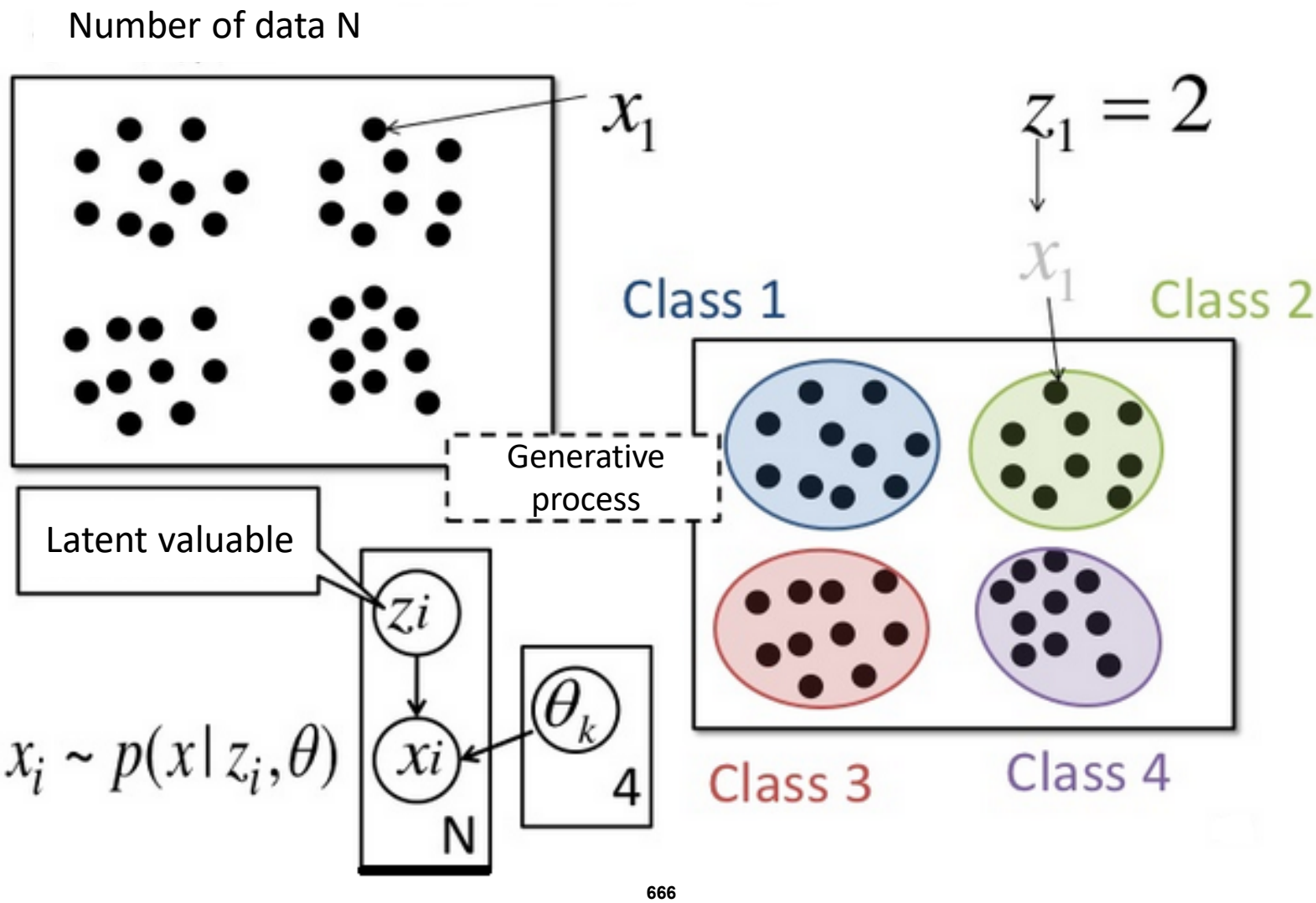
1st tutorial talk

Prof. Tadahiro Taniguchi

Ritsumeikan University

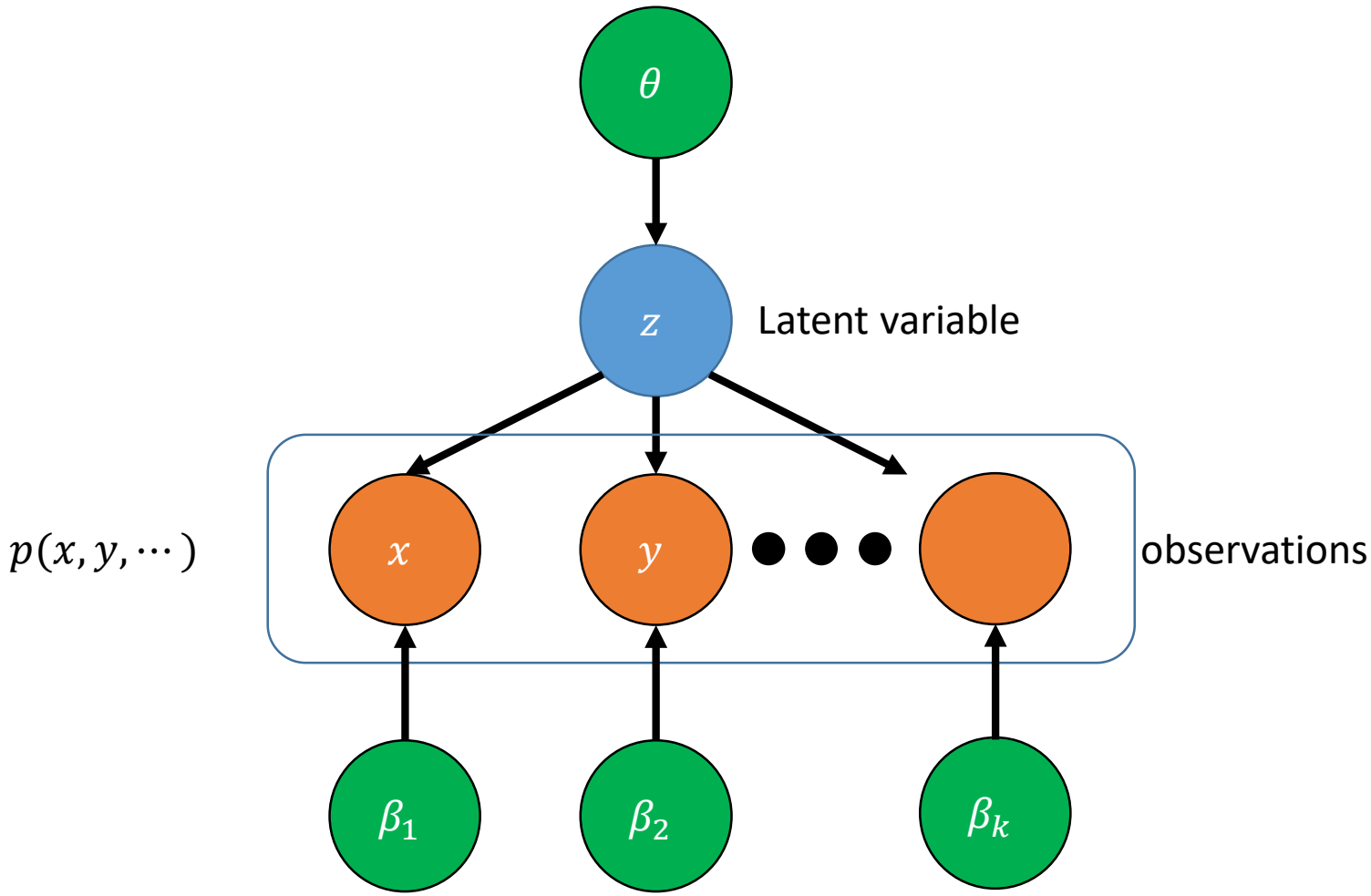
Basics of Probabilistic Generative Models for Robotics

Probabilistic generative model



from http://www.slideshare.net/issei_sato/deim2012-issei-sato

Multimodal generative model

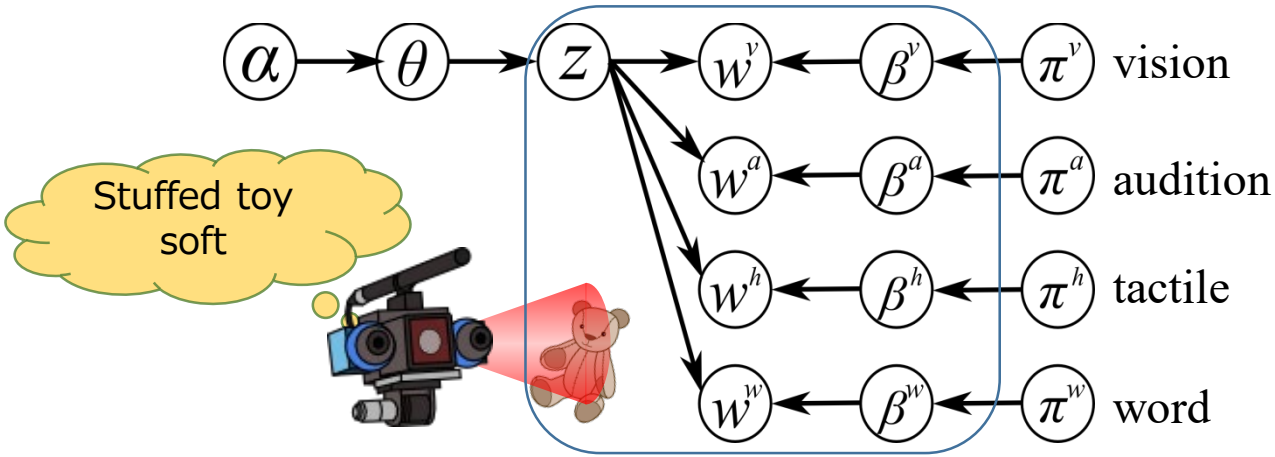


Multimodal categorization

◻Categorization of multimodal data

- **Multimodal Latent Dirichlet Allocation** (MLDA, MHDP, ⋯)

[Nakamura+ 09]



α : Dirichlet prior

θ : multinomial parameters

z : categories

w^*: multimodal infromation

β^*: multinomial parameters

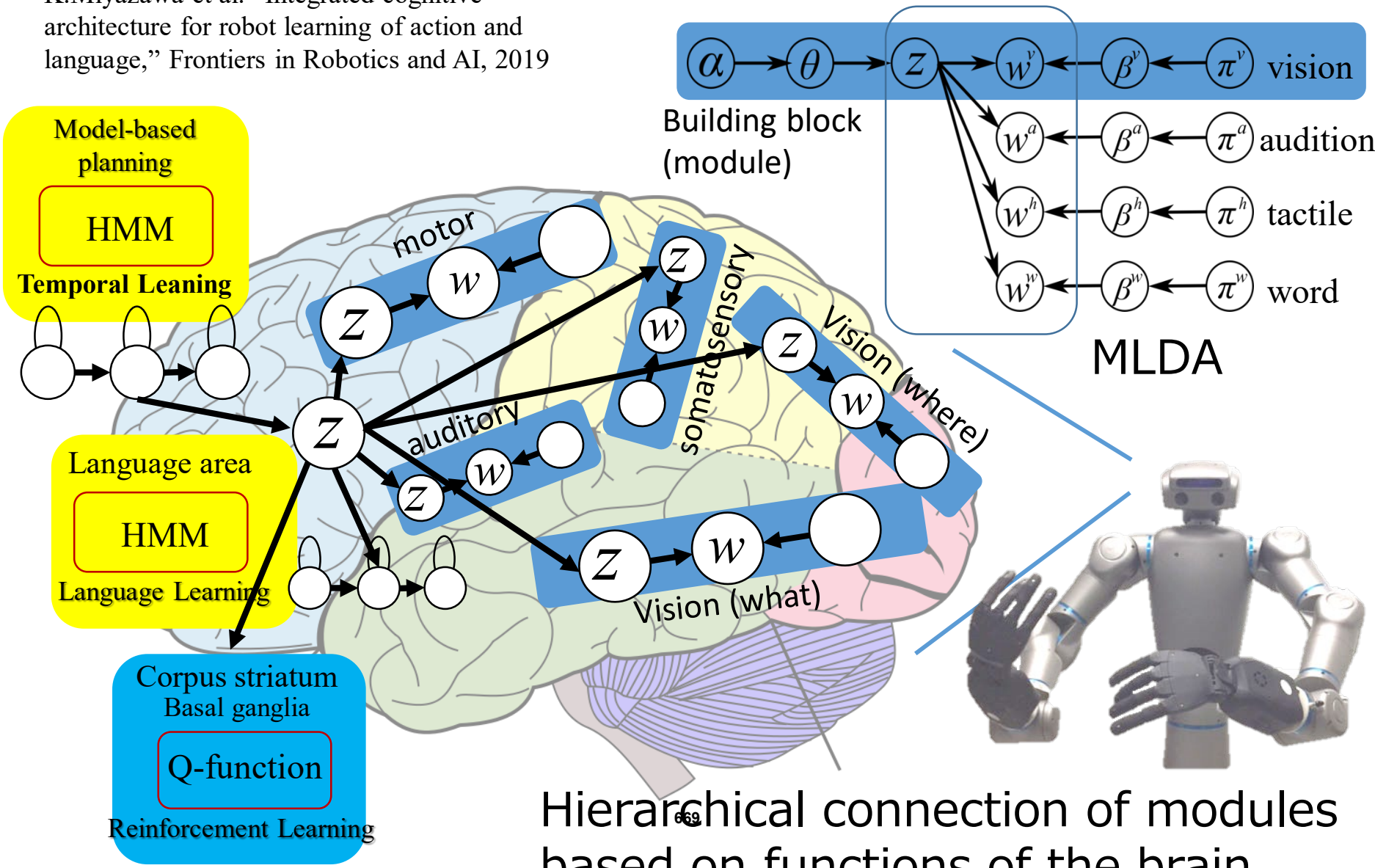
π^*: Dirichlet prior

Inference of the parameters β^* and θ by Gibbs Sampling

[Nakamura + 09] Nakamura,T. *et al.*, Grounding of word meanings in multimodal concepts using LDA, in Proc. IROS2009, pp.3943–3948, 2009

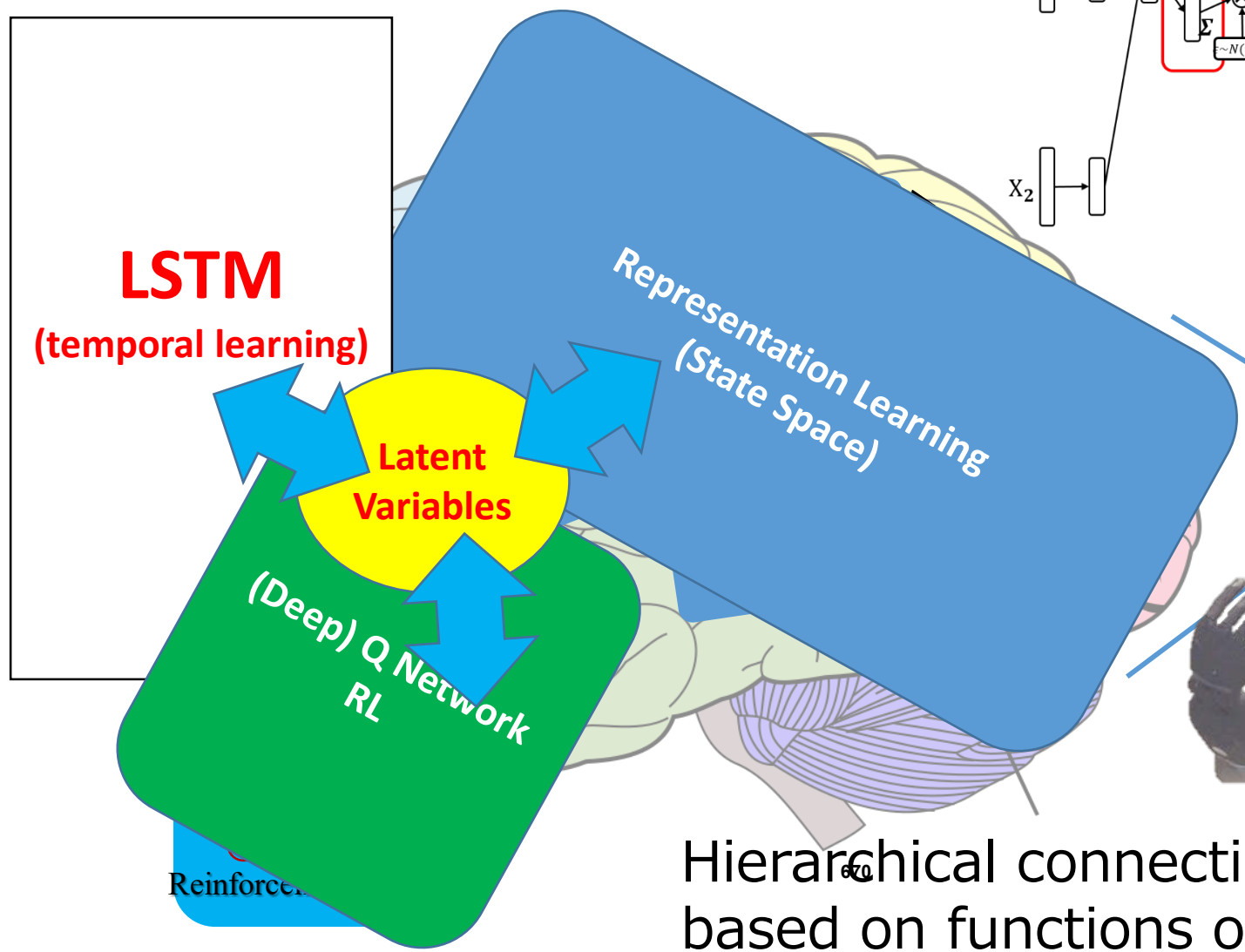
Integrated cognitive model

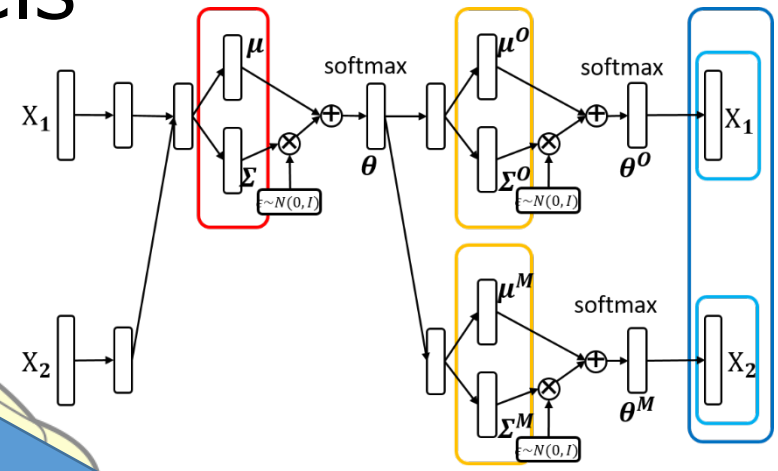
K.Miyazawa et al. "Integrated cognitive architecture for robot learning of action and language," Frontiers in Robotics and AI, 2019



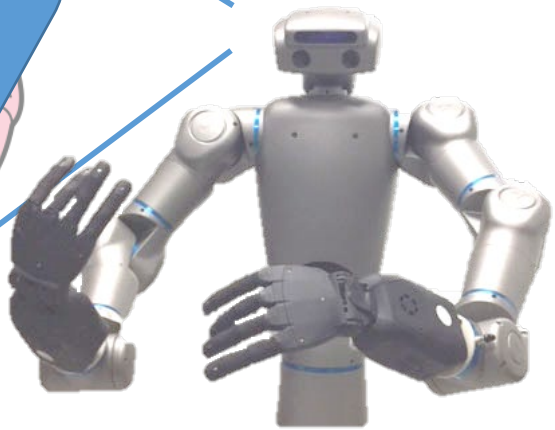
Hierarchical connection of modules based on functions of the brain

Integrated cognitive model w/ deep generative models





Deep mMLDA

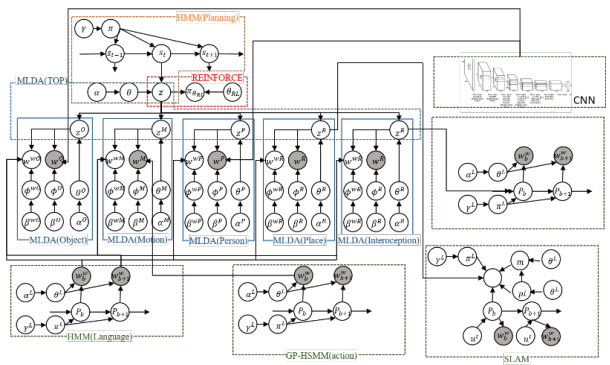


Hierarchical connection of modules based on functions of the brain

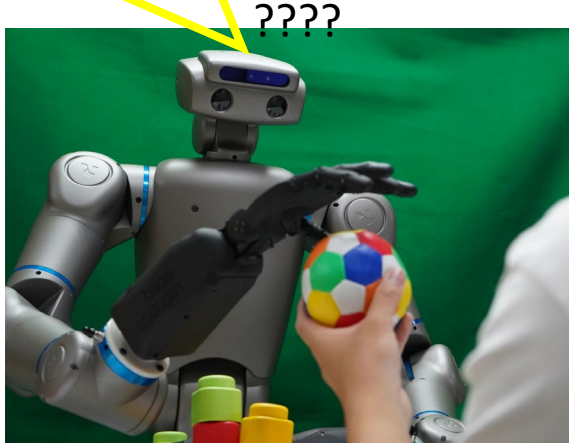
How does the robot use DPGMs?

- Planning/Control as probabilistic inference
- Relationship between DPGMs and MPC

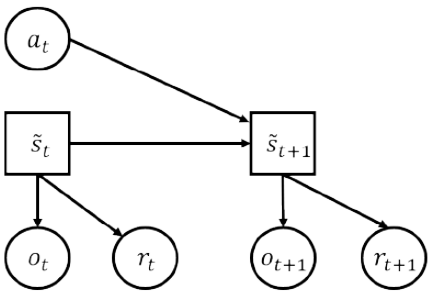
Complex cognitive model by DPGMs



????



Equivalent to



POMDP (World Model)

Planning/Control problems can be solved as probabilistic inference on the PGM

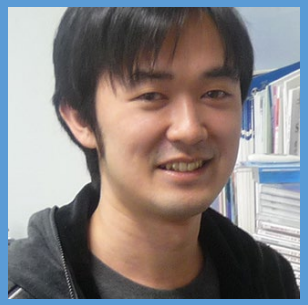
*Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review
*Variational Inference MPC for Bayesian Model-based Reinforcement Learning
*PlaNet of the Bayesians: Reconsidering and Improving Deep Planning Network by Incorporating Bayesian Inference

Planning/Control as inference

2nd tutorial talk

Dr. Masashi Okada

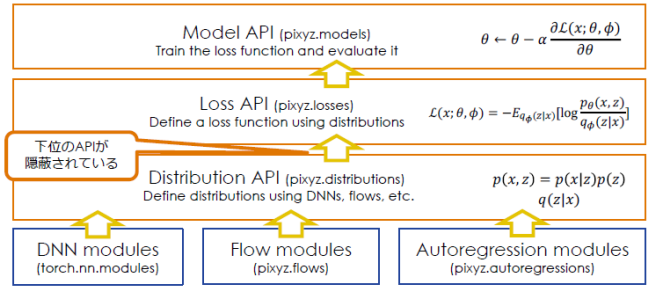
Panasonic Corp.



Theories of planning/control as probabilistic inference

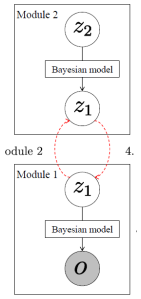
Tools

- We need to implement very complex models in practice
  - We have a very useful programing language for developing DPGMs!
    - **Pixyz**



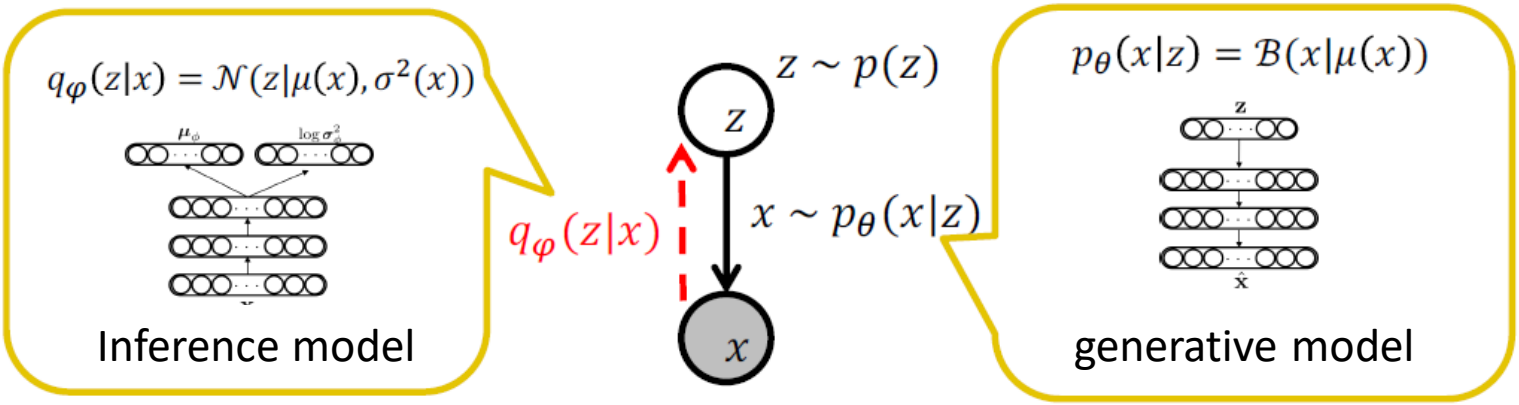
  - We have a framework for integrating multiple DPGMs (modules)
    - **SERKET/Neuro SERKET**

*Nakamura T, Nagai T and Taniguchi T (2018) SERKET: An Architecture for Connecting Stochastic Models to Realize a Large-Scale Cognitive Model. Front. Neurorobot. 12:25. doi: 10.3389/fnbot.2018.00025
*T. Taniguchi, T. Nakamura, M. Suzuki, R. Kuniyasu, K. Hayashi, A. Taniguchi, T. Horii, T. Nagai, **Neuro-SERKET: Development of Integrative Cognitive System Through the Composition of Deep Probabilistic Generative Models**, New Generation Computing. 38. 10.1007/s00354-019-00084-w



VAE（An example of DGMs）

- Variational autoencoder（VAE）[Kingma+ 13]



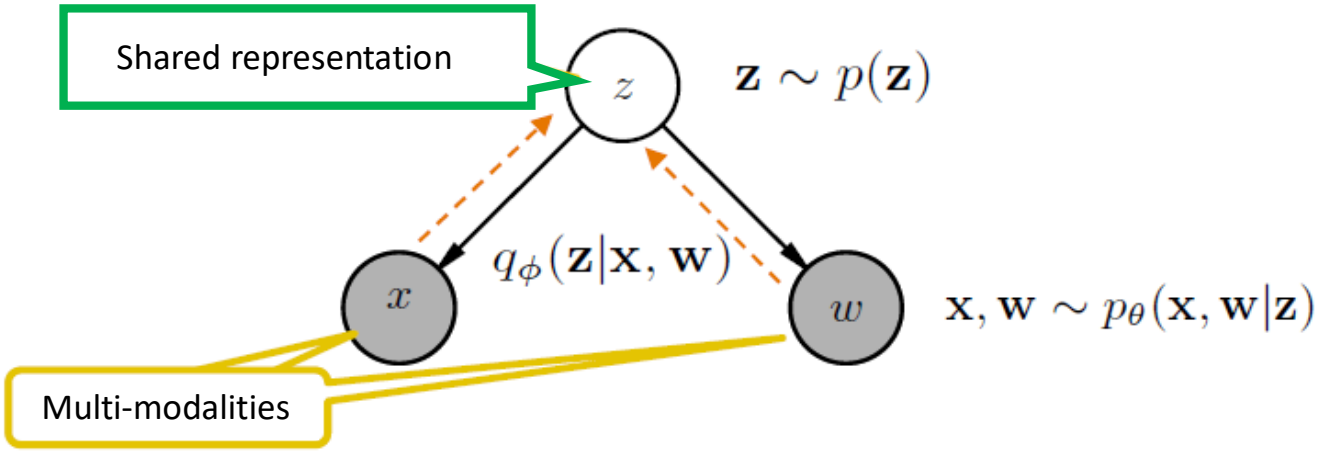
- Loss function : ELBO

$$\mathrm{E}_{q_\varphi(z|x)}[\log p_\theta(x|z)] - \mathrm{KL}[q_\varphi(z|x) \parallel p_\theta(z)]$$

This slide was provided by Dr. Suzuki

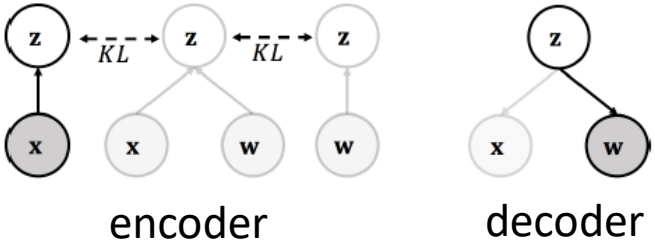
Multimodal deep generative models

- Encoder-decoder architecture is problematic in this case
  - Information cannot be predicted fro the other input



$$\mathcal{L}_{JM_{kl}}(x, w) = \mathcal{L}_{JM}(x, w) - [D_{KL}(q_\phi(z|x, w)||q_\lambda(z|x)) + D_{KL}(q_\phi(z|x, w)||q_\lambda(z|w))]$$

- JMVAE [Suzuki+ 16]
- PoE [Wu+ 18]
- Use associater between Z [Jo+ 19]



This slide was provided by Dr. Suzuki

Pixyz: programming language for DPGMs

3rd tutorial talk

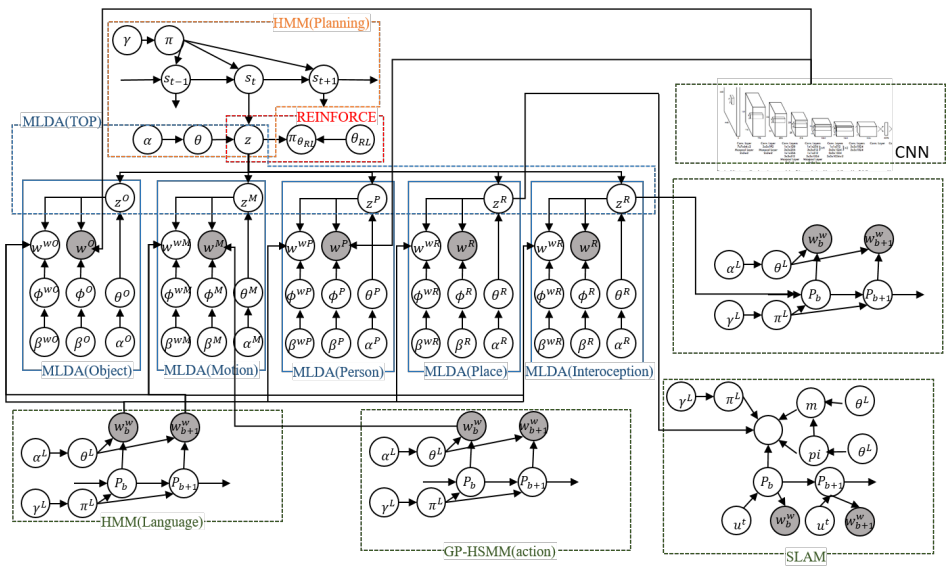
Prof. Masahiro Suzuki

The University of Tokyo

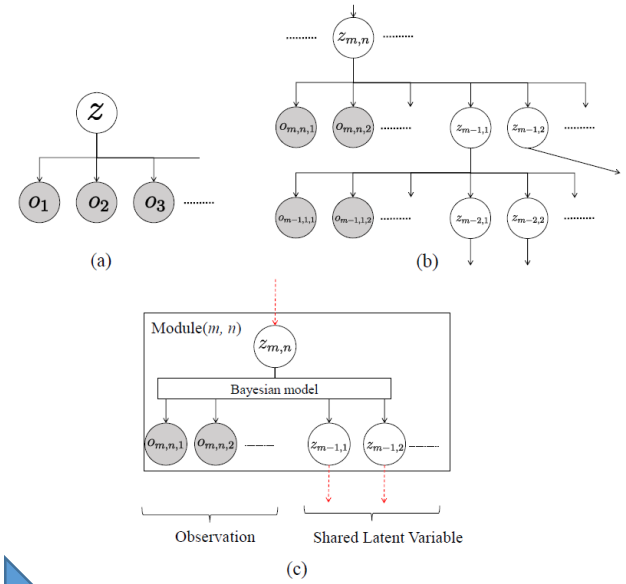
Pixyz: a framework for developing complex deep generative models

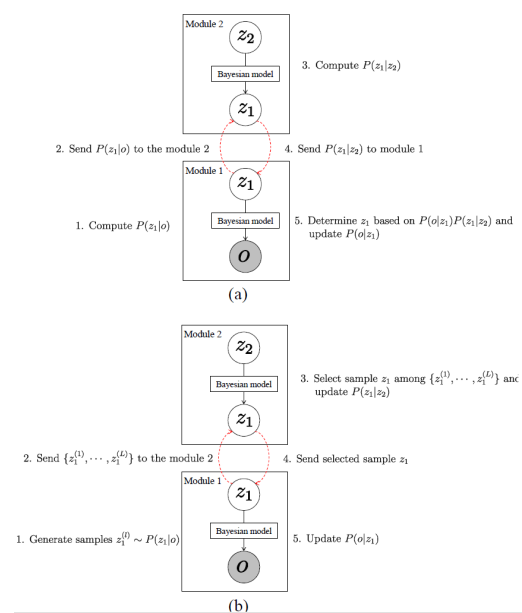
Even more complex generative models

- Integration of modules
- Optimization as a whole



decomposition

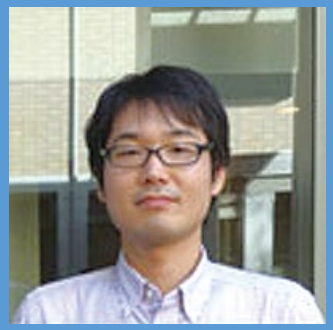




T.Nakamura, T.Nagai, T.Taniguchi, **SERKET: An Architecture for Connecting Stochastic Models to Realize a Large-Scale Cognitive Model**, Front. Neurorobot., 26 June 2018

T. Taniguchi, T. Nakamura, M. Suzuki, R. Kuniyasu, K. Hayashi, A. Taniguchi, T. Horii, T. Nagai, **Neuro-SERKET: Development of Integrative Cognitive System Through the Composition of Deep Probabilistic Generative Models**, New Generation Computing. 38. 10.1007/s00354-019-00084-w

SERKET: integration of multiple models



4th tutorial talk

Prof. Tomoaki Nakamura

The University of Electro-Communications

A Framework for constructing multimodal learning models: SERKET

Recap

- 4 tutorial talks
  - Theoretical side (2 talks)
    - by Prof. Taniguchi
    - by Dr. Okada
  - Implementation side (2 talks)
    - by Prof. Nakamura
    - by Prof. Suzuki
- Supplemental materials

  https://sites.google.com/view/dpgmfr/home
  - Slides, GitHhub, sample codes, papers, past workshops

Enjoy!

This tutorial is presented by **RSJ**
and
JST CREST "Symbol Emergence in Robotics for Future Human-Machine Collaboration"

Thanks for endorsing this tutorial !

- IEEE RAS TC on Robot Learning
- IEEE RAS TC on Cognitive Robotics
- IEEE CDS TC Task Force on Robotics