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
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!

Number of data $N$  
Latent valuable  
Generative process

Observation

$x_i \sim p(x_i | z_i, \theta)$

Probabilistic Generative Models

APPLICATIONS
THEORIES
TOOLS

“Deep” Probabilistic Generative Models (DPGMs)
We are interested in cognitive robotics
Constructive approach

- **Constructive Human Science**
  - Construct to know mechanisms behind our mind
  - Construct to use it for specific applications

Science

Constructive approach

Engineering

Applications

- Elderly care
- Service robots
- Industrial robots
- Child care
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**
  
  \[ p(Y|X) \]

  \[ X \rightarrow Y \]

  **Supervised**

  **End to end learning**

- **Bayesian model**
  
  \[ p(Y|X) = \sum_z p(Y, z|X) = \sum_z p(Y|z)p(z|X) \]

  \[ X \rightarrow Y \]

  \[ z \text{ is a latent variable (concept), which is acquired by the robot itself} \]

  \[ \Rightarrow \text{unsupervised learning} \]

- **Pipeline**
  
  \[ p(Y|X) \approx p(Y|\text{argmax } p(z|X)) \]

  \[ = p(Y|\bar{z}) \]

  **AI and/or Robot**

  \[ \bar{z} = \text{argmax } \frac{p(X|z)p(z)}{p(X)} \]

  **Pattern recognition**

  Problem of finding the label \( z \) corresponding to \( X \)

  \( z \) is defined by hand \( \Rightarrow \) Training the classifier \( p(X|z) \)

  Required labeled data

  When \( \bar{z} \) is selected by the classifier, an action must be selected according to the result

  \[ \text{Hard to design } p(Y|\bar{z}) \]
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

『stuffed toy』
Phoneme seq.

Concept space

representation

Concept#1

Concept#2

Concept#3

Concept#4

inference

soft
understanding

662
Counter direction

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

$$p(x, y, \cdots) = \sum_{z \in \mathcal{Z}} p(x, y, \cdots | z)p(z)$$
Multimodal Generative Models

Multimodal supervised learning

\[ p(y|x) \]

\[ x \rightarrow z \rightarrow y \]

Observations  Latent Variables  Output

\[ p(x|y) \]

\[ y \rightarrow z \rightarrow x \]

Multimodal unsupervised learning

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

\[ p(x, y, \cdots) \]

Observations

Latent Variable
1st tutorial talk
Prof. Tadahiro Taniguchi
Ritsumeikan University
Basics of Probabilistic Generative Models for Robotics
Probabilistic generative model

Number of data $N$

Latent valuable

Generative process

$x_i \sim p(x | z_i, \theta)$

Class 1

Class 2

Class 3

Class 4
Multimodal generative model
Multimodal categorization

Categorization of multimodal data
- Multimodal Latent Dirichlet Allocation (MLDA, MHDP, ...)

\[ \alpha, \beta, \theta \]

\[ \pi^v \text{ vision}, \pi^a \text{ audition}, \pi^h \text{ tactile}, \pi^w \text{ word} \]

\[ \alpha^* : \text{ Dirichlet prior}, \theta^* : \text{ multinomial parameters}, z^* : \text{ categories}, w^* : \text{ multimodal information}, \beta^* : \text{ multinomial parameters}, \pi^* : \text{ Dirichlet prior} \]

Inference of the parameters \( \beta^* \) and \( \theta \) by Gibbs Sampling

Integrated cognitive model


Hierarchical connection of modules based on functions of the brain
Integrated cognitive model w/ deep generative models

LSTM (temporal learning)

Hierarchical connection of modules based on functions of the brain

Deep mMLDA

Latent Variables

(Deep) Q Network

Reinforcement Learning (RL)

Representation Learning (State Space)
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

VAE (An example of DGMs)

- Variational autoencoder (VAE) [Kingma+ 13]

\[ q_\phi(z|x) = N(z|\mu(x), \sigma^2(x)) \]

\[ p_\theta(x|z) = B(x|\mu(x)) \]

- Loss function : ELBO

\[ E_{q_\phi(z|x)}[\log p_\theta(x|z)] - KL[q_\phi(z|x) \parallel p_\theta(z)] \]
Multimodal deep generative models

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

- 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


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, GitHub, 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