ROYAL INSTITUTE
OF TECHNOLOGY

Learning from Structures for Long Term Autonomous Agents

AKSHAYA THIPPUR SRIDATTA

Doctoral Thesis
Stockholm, Sweden 2020

TRITA-CSC-XXXXX
ISSN-XXXXXXXXXX
ISRN-KTH/CSC/XXXX
ISBN XXXXXXXXXXXXX

Robotics, Perception and Learning
School of Computer Science and Communication
KTH Royal Institute of Technology
SE-100 44 Stockholm, Sweden

Tryck: Universitetsservice US-AB 2020

ROYAL INSTITUTE
OF TECHNOLOGY

Learning from Structures for Long Term Autonomous Agents

AKSHAYA THIPPUR SRIDATTA

Doctoral Thesis
Stockholm, Sweden 2020

TRITA-CSC-XXXXX
ISSN-XXXXXXXXX
ISRN-KTH/CSC/XXXX
ISBN XXXXXXXXXXXX

Robotics, Perception and Learning
School of Computer Science and Communication
KTH Royal Institute of Technology
SE-100 44 Stockholm, Sweden

Tryck: Universitetsservice US-AB 2020

Abstract

Here is my Abstract

Sammanfattning

The Swedish Abstract

Acknowledgments

Thank everyone here

List of Papers

The thesis is based on the following papers:

[A] Alessandro Pieropan and Hedvig Kjellström. Unsupervised Object Exploration Using Context. In *Proceedings of the 2014 IEEE International Symposium on Robot and Human Interactive Communication (ROMAN'14)*, Edinburgh, UK, August 2014.

Contents

List of Figures

Chapter 1

Introduction

1.1 On the structures in our universe

- Organised vs. Search view of the world. Think folders in reality and on our machines.
- Where is structure?
  - Morphology
  - Carnatic Music
  - Human constructed situations: spanning environments to creations and object clusters. why so? To help memory? Transfer of knowledge, linguistics, communication ease, data compression?
  - languages, programming or human
- Where is there no structure?
  - Physics models: brownian motion
  - Environment, nature
  - Random chance situations
  - But we still tried to model them with some structure. Or did we?
- Is there really some structure or do we seek it out?
- We assume some all of us know some coordinates, some conventions and some functional knowledge about structures, then we go about trying to model our current circumstance or technical problem in terms of structures we already know, to be able to convey that across to others.

1.2 Where is structure?

1.3 Why care for structure?

1.4 Machine Learning: Attention to structures - An interpretation

1. Structure Learning
2. All regression and Classification are structures to predict or to bound. Structure matching?
3. What is contained in structure? Features, clarity, information
   a) In-filling deep learning
   b) word predictor NLP
   c) Space partitioning for object search
   d) Mapping, topological especially. Combination of planes, blobs and so on: Basis structures.
   e) Compression of data: Speech coders, Spectral Band Replication
   f) GP based regression
   g) Decompositions?

The reference lies here TEST [1]

Chapter 2

Finding Structure

Towards Spatial Relationships: KTH-3D-Total Dataset

2.1 Long-term observations of human indoor environments

2.2 RELATED WORK

2.3 Interesting surfaces - Why desks?

2.4 KTH Long Term Dataset

2.5 Matterport Dataset

2.6 KTH 3D Total Dataset

2.6.1 Need and construction

2.6.2 Noticing Spatial structures

2.6.3 INSERT PAPER KTH 3D TOTAL

2.7 Spatial Relations: Language of Spatial Structures

2.8 TODO

- See if you can convert Matterport pointclouds into 2D projections
- Matterport arXiv paper and hosting?
- KTH 3D Total hosting

Chapter 3

Describing Spatial Structure

Qualitative and Quantitative Spatial Relations

3.1 Spatial Relations - Intro

- why use it?
- language and description
- compression - topological descriptions
- mapping and directions

3.2 RELATED WORKS

3.3 Spatial Relations for our problem

3.4 Task description - object recognition

Talk about aiding the vision system and why? Extrinsic cues.

3.5 INSERT PAPER - AAAI QSR

3.6 Spatial Relations recommendations

Take the discussion section from paper and elaborate. When to use what SR and why should QSRs be measureable?

3.7 SR for STRANDS?

How to go forward from this analysis?

3.8 INSERT PAPER - IROS QSR

Discuss practical issues

Chapter 4

Designing Spatial Relations

Joint Object Classification with Intrinsic Frame of Reference Calculi IFRC

4.1 What kind of spatial relations to design?

4.2 Why IFRC?

TODO: Insert the linguistic experiment data. Conclusions drawn from it – elaborate.

4.3 INSERT PAPER- IFRC

4.4 Discussion

Difficulties of Joint Object Classification.

Elaborate discussion of joint object classification with IFRC. What can be improved? What are the pitfalls?

Chapter 5

Object Estimation

Bayesian Optimisation based Multiple Instance Estimation

5.1 Difficulties of Object Estimation

- Multiple instances
- Only extrinsic features
- small data
- multiple instances – explain different location hits

5.2 RELATED WORKS - Object Estimation

Focus: Non parametric methods for Object Estimation

5.3 INSERT PAPER – Bayesian Optimisation for Object Estimation

5.4 Discussion

Chapter 6

Making the Environment Continuous

Finite Mixture Models for Stochastic Kronecker Graphs

6.1 Environment Recognition

- Why constrain to objects?
- Generalise to all kinds of environments. Football scenes, Chess boards, Crime scenes
- Key points - large network graphs
- Use large networks to analyse
- place in this context environment recognition in office places
- elaborate on nodes, graph construction, edges what ARE nodes in interpretation.

6.2 Kronecker graph theory

6.3 Related works for large network analysis

6.4 INSERT PAPER - FMM for SKG

6.5 Discussion on generative model

What are the main takeaways from the generative model?

6.6 TODO

- convert couple scenes into large networks for examples - run inference experiments

Chapter 7

Deep Learning and Demo

Possible chapter

7.1 DNN problem formulations

- Bar code type of problem
- fly on object type of problem - Who am I?
- In painting
- Include IFRC data as metadata

Why not deep learning explored? Model importance for interpretability. Data constraints. Data feeding constraints.

7.2 Parallels to graph neural networks

refer to work from Talukdar tutorial

7.3 SORHACK Demo

Possible to include this chapter 5

7.3.1 Problem statement

- motivation
- resources
- hidden object estimation

success measurement. Precision Recall why not used? Refer papers - Sanne Elena

7.4 Further challenges

problems with object identification with integration to vision system.

7.5 TODO

- Report? arXiv paper on system integration?
- Make up the video

Chapter 8

Final Discussion

Discussion

8.1 Contextual learning - general discussion

- Saikat - What is it learning to learn? Dog example
- Kid learns about dogs only by cartoons - small representative subset of dogs. Then when exposed to various dogs in a park, there is not a single error type 1 or type 2 of detecting dogs. Why?
- Generalisability with no common examples. No supervision either of the generalising nature. No guidelines even by explicit explanations. Not even possibly a one shot learning!
- the kid learnt by context of activity, the context of environment and spatial conformation of sub objects such as leash, collar, stick or ball in the mouth?
- When is an aritficial agent intelligent?
  - Test by hardest example accuracy to the easiest. does performance decrease?
  - Robustness to Noise
  - Can maintain robustness across multiconditioning tests, simultaneously and individually tested.
- Fruits and veggies behind plastic sheets, fails because of lack of contextual support or even biased contextual knowledge. Most visual illusions work this way. Class boundary specification needs to be explicit, otherwise what is in our head needs to map to the machine through data examples! We as humans know "what happens when there is plastic sheets interfering with vision?" The machine does not.
- Unrelated: What is Selection bias? Ping pong theorem?

Bibliography

[1] Ruslan Salakhutdinov and Geoffrey E Hinton. Deep boltzmann machines. In *International Conference on Artificial Intelligence and Statistics*, pages 448–455, 2009.