Human-like Visual Learning and Reasoning

Extended Abstract[†]

Peng Cui Department of Computer Science Tsinghua University China cuip@tsinghua.edu.cn Wenwu Zhu Department of Computer Science Tsinghua University China wwzhu@tsinghua.edu.cn

ABSTRACT

Compared with machines, people are far better learners as they are capable of learning models¹ from very limited samples of a new category and make accurate predictions and reasoning accordingly. To learn from human learning process and investigate human-like visual learning and reasoning mechanism is a promising way to develop new effective learning methods. In this tutorial, several basic problems in human-like visual learning and reasoning will be discussed, including why we need to learn from human learning process, what are the research goals of human-like visual learning and reasoning, and how to embed human learning characteristics into learning models.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics • **Networks** → Network reliability

KEYWORDS

Visual reasoning, human-like learning

1 MOTIVATION

The recent progress of machine learning, especially the emergence of deep learning, has advanced the multimedia content understanding into an unprecedented level. The error rates on large-scale benchmark datasets has been halved and halved again, even approaching human-level performance on some object recognition benchmarks. Despite the success, the state-of-the-art models are notoriously data hungry, requiring tons of samples for parameter learning. In real cases, however, the visual phenomena follow a long-tail distribution where only a few sub-categories are data-rich and the rest are with limited training samples. How to learn a concept (or object) from as

MM '17, October 23–27, 2017, Mountain View, CA, USA © 2017 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-4906-2/17/10.

https://doi.org/10.1145/3123266.3130144

fewer samples as possible is critical for real applications and fundamental for exploring new learning mechanisms. Compared with machines, people are far better learners as they are capable of learning models from very limited samples of a new category and make accurate predictions and reasoning accordingly. To learn from human learning process and investigate human-like visual learning and reasoning mechanism is a promising way to develop new effective learning methods. Compared with most of performance-driven deep learning works, this line of research focus more on fundamental and insightful understanding on the learning mechanisms, and have large potential to lead to new emerging directions in multimedia field.

In this tutorial, we will review the recent thoughts and achievements on human-like visual learning and reasoning. More specifically, a series of fundamental problems in humanlike visual learning and reasoning will be discussed, including why we need to learn from human learning process, what are the research goals of human-like visual learning and reasoning, how to embed human learning characteristics into learning models, and the major future directions of human-like visual learning and reasoning.

2 TUTORIAL OUTLINE

The tutorial is planned and organized into 6 segments. Its detailed outline is presented below.

(1) The landscape of AI and Multimedia

(2) Motivation and background of human-like visual learning and reasoning

(3) Causal reasoning in visual space

(4) Learning to learn for visual reasoning

(5) Discussions and future directions

3 TARGET AUDIENCE AND PREREQUISITIES

This tutorial will be accessible to all multimedia researchers, students and practitioners who are interested in multimedia content analysis, computer vision, machine learning and other multimedia related directions. No special prerequisite knowledge is needed to attend this tutorial. We expect 60-80 attendees for this tutorial. The tutorial ppt will be shared online before the tutorial day.

¹ The representation and inference models

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

4 TUTORS' SHORT BIO AND EXPERTISE

Peng Cui is an Associate Professor in Tsinghua University. He got his PhD degree from Tsinghua University in 2010. He is keen to promote the convergence of social media data mining and multimedia computing technologies. His research interests include network representation learning, human behavioral modeling, and social-sensed multimedia computing. He has published more than 60 papers in prestigious conferences and journals in data mining and multimedia. His recent research won the SIGKDD 2016 Best Paper Finalist, ICDM 2015 Best Student Paper Award, SIGKDD 2014 Best Paper Finalist, IEEE ICME 2014 Best Paper Award, ACM MM12 Grand Challenge Multimodal Award, and MMM13 Best Paper Award. He is the Area Chair of ICDM 2016, ACM MM 2014-2015, IEEE ICME 2014-2015, ICASSP 2013, Associate Editors of IEEE TKDE, ACM TOMM, Elsevier Journal on Neurocomputing, and Guest Editors of IEEE Intelligent Systems, Information Retrieval Journal, Machine Vision and Applications, etc. He was the recipient of ACM China Rising Star Award in 2015.

Wenwu Zhu is with Computer Science Department of Tsinghua University as Professor of "1000 People Plan" of China. Prior to his current post, he was a Senior Researcher and Research Manager at Microsoft Research Asia. He was the Chief Scientist and the Director at Intel Research China from 2004 to 2008. He worked at Bell Labs New Jersey as Member of Technical Staff during 1996-1999. Wenwu Zhu is an IEEE Fellow, SPIE Fellow and ACM Distinguished Scientist. He has published over 200 referred papers in the areas of multimedia computing, communications and networking. He is inventor or co-inventor of over 40 patents. His current research interests are in the area of social media computing and multimedia communications and networking. He served(s) on various editorial boards, such as Guest Editor for the Proceedings of the IEEE, IEEE T-CSVT, and IEEE JSAC; Associate Editor for IEEE Transactions on Mobile Computing, IEEE Transactions on Multimedia, and IEEE Transactions on Circuits and Systems for Video Technology. He served as TPC Co-Chair of IEEE ISCAS 2013 and serves as TPC Co-Chair for ACM Multimedia 2014.

REFERENCES

- Kun Kuang, Peng Cui, Bo Li, Meng Jiang, Fei Wang, Shiqiang Yang. Treatment Effect Estimation with Data-Driven Variable Decomposition. AAAI, 2017.
- [2] Daixin Wang, Peng Cui, Wenwu Zhu. Structural Deep Network Embedding. KDD, 2016.
- [3] Peng Cui, Wenwu Zhu, Tat-Seng Chua, Ramesh Jain. Social-Sensed Multimedia Computing. IEEE Multimedia, 2016.
- [4] Shaowei Liu, Peng Cui, Wenwu Zhu, Shiqiang Yang. Learning Socially Embedded Visual Representation from Scratch. ACM Multimedia, 2015.
- [5] Daixin Wang, Peng Cui, Mingdong Ou, Wenwu Zhu. Deep Multimodal Hashing with Orthogonal Regularization. IJCAI, 2015.