

Course Description



KeyNote and Courses

Keynotes

 [Maria-Florina Balcan \(Carnegie Mellon University\)](#) [introductory/intermediate]
[Data Driven Clustering](#)

 [Mark Gales \(University of Cambridge\)](#) [introductory/intermediate]
[Use of Deep Learning in Non-native Spoken English Assessment](#)

 [Mihaela van der Schaar \(University of Cambridge\)](#) [introductory/intermediate]
[Learning Engines for Healthcare: Using Machine Learning to Transform Clinical Practice and Discovery](#)

Courses

 [Aaron Courville \(University of Montréal\)](#) [introductory/intermediate]
[Deep Generative Models](#)

 [Issam El Naqa \(University of Michigan\)](#) [introductory/intermediate]
[Deep Learning for Biomedicine](#)

 [Sergei V. Gleyzer \(University of Florida\)](#) [introductory/intermediate]
[Feature Extraction, End-end Deep Learning and Applications to Very Large Scientific Data: Rare Signal Extraction, Uncertainty Estimation and Realtime Machine Learning Applications in Software and Hardware](#)

 [Vasant Honavar \(Pennsylvania State University\)](#) [introductory/intermediate]
[Causal Models for Making Sense of Data](#)

 [Qiang Ji \(Rensselaer Polytechnic Institute\)](#) [introductory/intermediate]
[Probabilistic Deep Learning for Computer Vision](#)

 [James Kwok \(Hong Kong University of Science and Technology\)](#) [introductory/intermediate]
[Compressing Neural Networks](#)

 [Tomas Mikolov \(Facebook\)](#) [introductory]
[Using Neural Networks for Modeling and Representing Natural Languages \(with Piotr Bojanowski and Armand Joulin\)](#)

 [Hermann Ney \(RWTH Aachen University\)](#) [intermediate/advanced]
[Speech Recognition and Machine Translation: From Statistical Decision Theory to Machine Learning and Deep Neural Networks](#)

 [Jose C. Principe \(University of Florida\)](#) [intermediate/advanced]
[Cognitive Architectures for Object Recognition in Video](#)



[Fabio Roli \(University of Cagliari\)](#) [introductory/intermediate]
Adversarial Machine Learning



[Björn Schuller \(Imperial College London\)](#) [introductory/intermediate]
Deep Learning for Intelligent Signal Processing



[Alex Smola \(Amazon\)](#) [introductory]
Dive into Deep Learning



[Sargur Srihari \(University at Buffalo\)](#) [intermediate/advanced]
Explainable Artificial Intelligence



[Ponnuthurai N Suganthan \(Nanyang Technological University\)](#) [introductory/intermediate]
Learning Algorithms for Classification, Forecasting and Visual Tracking



[Johan Suykens \(KU Leuven\)](#) [introductory/intermediate]
Deep Learning, Neural Networks and Kernel Machines



[Bertrand Thirion \(INRIA\)](#) [introductory]
Understanding the Brain with Machine Learning



[Gaël Varoquaux \(INRIA\)](#) [intermediate]
Representation Learning in Limited Data Settings



[René Vidal \(Johns Hopkins University\)](#) [intermediate/advanced]
Mathematics of Deep Learning



[Haixun Wang \(WeWork\)](#) [intermediate]
Abstractions, Concepts, and Machine Learning



[Xiaowei Xu \(University of Arkansas, Little Rock\)](#) [introductory/advanced]
Multi-resolution Models for Learning Multilevel Abstract Representations of Text



[Ming-Hsuan Yang \(University of California, Merced\)](#) [intermediate/advanced]
Learning to Track Objects



[Zhongfei Zhang \(Binghamton University\)](#) [introductory/advanced]
Knowledge Discovery from Complex Data with Deep Learning



Maria-Florina Balcan

(Carnegie Mellon University) []

Data Driven Clustering Summary

Clustering is a fundamental problem in data science, used in myriad of applications. Despite significant research in different fields, clustering remains a major challenge. Most traditional approaches to designing and analyzing clustering algorithms have mainly focused on one shot clustering, where the goal is to design an algorithm to cluster a one-time dataset well. Unfortunately, from a theoretical standpoint, there are major impossibility results for such scenarios; first, in most applications it is not clear what notion of similarity or what objective function to use in order to recover a good clustering for a given dataset; second even in cases where the similarity function and the objectives can be naturally specified, optimally solving the underlying combinatorial clustering problems is typically intractable.

In this talk, I will describe a lifelong transfer clustering approach to address these challenges. Motivated by the fact that in many modern settings, we often need to solve not only one, but many clustering problems arising in a given application domain, we consider algorithms that adaptively learn to cluster. In particular, given a series of clustering instances to be solved from the same domain, we show how to learn a good parameter setting for clustering algorithms that perform well on instances coming from that domain. We provide formal guarantees on the number of typical problem instances that are sufficient to ensure that a clustering algorithm that does well on these typical instances, will also do well on new instances coming from the same domain, as a function of the complexity of the underlying parametrized family of clustering algorithms. We also show a significant benefit of our approach experimentally on datasets such as MNIST, CIFAR, and Omniglot.

Short Bio

Maria Florina Balcan is an Associate Professor in the School of Computer Science at Carnegie Mellon University. Her main research interests are machine learning, computational aspects in economics and game theory, and algorithms. Her honors include the CMU SCS Distinguished Dissertation Award, an NSF CAREER Award, a Microsoft Faculty Research Fellowship, a Sloan Research Fellowship, and several paper awards. She was a program committee co-chair for the Conference on Learning Theory in 2014 and for the International Conference on Machine Learning in 2016. She is currently board member of the International Machine Learning Society (since 2011), a Tutorial Chair for ICML 2019, and a Workshop Chair for FOCS 2019.



Mark Gales

(University of Cambridge) [\[\]](#)

Use of Deep Learning in Non-native Spoken English Assessment Summary

There is a high global demand for the learning of English as an additional language. Automatic assessment systems can help meet this need by reducing human assessment effort and enabling learners to independently monitor their progress when/wherever they choose. To properly determine a candidate's spoken English proficiency the auto-marker should be able to accurately assess the learner's ability level from spontaneous, prompted, speech. This should be independent of L1 language and audio recording quality which vary considerably making this a challenging task. This talk will look at the application of deep learning to spontaneous spoken English assessment. Examples of tasks that will be discussed include:

- i) efficient ASR systems, and ensemble combination, for non-native English;
- ii) task-specific phone "distance" features for assessment and L1 detection;
- iii) prompt-response relevance for off-topic response detection;
- iv) grammatical error detection and correction for learner English.

These tasks make use of a range of deep-learning techniques including: recurrent sequence models; sequence ensemble distillation (teacher-student training); attention mechanisms; and Siamese networks.

Short Bio

Mark Gales is a Professor of Information Engineering in the Machine Intelligence Laboratory (formerly the Speech Vision and Robotics (SVR) group) and a Fellow of Emmanuel College. He is a member of the Speech Research Group together with faculty staff members Phil Woodland and Bill Byrne. Recent past members of the group include Milica Gasic and Steve Young.

<http://mi.eng.cam.ac.uk/~mjfg/bio.html>



Mihaela van der Schaar

(University of Cambridge) []

Learning Engines for Healthcare: Using Machine Learning to Transform Clinical Practice and Discovery

Abstract

In this talk, I will discuss recent machine learning and AI theory, methods, algorithms and systems which we developed in our lab to understand the basis of health and disease, to catalyze clinical research, to support clinical decisions through individualized medicine, to inform clinical pathways, to better utilize resources & reduce costs and to inform public health.

To do this, we are creating what I call Learning Engines for Healthcare (LEH's). An LEH is an integrated ecosystem that uses machine learning, AI and operations research to provide clinical insights and healthcare intelligence to all the stakeholders (patients, clinicians, hospitals, administrators). In contrast to an Electronic Health Record, which provides a static, passive, isolated display of information, an LEH provides dynamic, active, holistic & individualized display of information including alerts.

In this talk I will focus on 3 steps in the development of LEH's:

- » 1. Building a comprehensive model that accommodates irregularly sampled, temporally correlated, informatively censored and non-stationary processes in order to understand and predict the longitudinal trajectories of diseases.
- » 2. Establishing the theoretical limits of causal inference and using what has been established to create a new approach that makes it possible to better estimate individualized treatment effects.
- » 3. Using Machine Learning itself to automate the design and construction of entire pipelines of Machine Learning algorithms for risk prediction, screening, diagnosis and prognosis.

Short Bio

Professor van der Schaar is John Humphrey Plummer Professor of Machine Learning, Artificial Intelligence and Medicine at the University of Cambridge and a Turing Faculty Fellow at The Alan Turing Institute in London, where she leads the effort on data science and machine learning for personalized medicine. Prior to this, she was a Chancellor's Professor at UCLA and MAN Professor of Quantitative Finance at University of Oxford. She is an IEEE Fellow (2009). She has received the Oon Prize on Preventative Medicine from the University of Cambridge (2018). She has also been the recipient of an NSF Career Award, 3 IBM Faculty Awards, the IBM Exploratory Stream Analytics Innovation Award, the Philips Make a Difference Award and several best paper awards, including the IEEE Darlington Award. She holds 35 granted USA patents. Her current research focus is on data science, machine learning, AI and operations research for medicine.

**Aaron Courville**

(University of Montréal) [introductory/intermediate]

Deep Generative Models Summary

The past few years have seen intensive research effort and dramatic improvements in neural network-based generative models. Modern generative models can generate large-scale photorealistic images that would have been unthinkable a few short years ago. More recent efforts have demonstrated how we can exploit these generative models for improved classification performance, transfer data from one domain to another, gain a 3D world from single 2D images and even how these models can be used to support reinforcement learning of AI agents.

This lecture will cover modern deep-learning based generative models with an emphasis on breadth. I will present four categories of deep generative models: autoregressive models, normalizing flows, variational autoencoders, and adversarial generative models. My goal is to highlight how the relative strengths and weaknesses of these diverse approaches and to show how neural networks are being leveraged in each of the these modelling paradigms to achieve impressive performance.

Syllabus

Lecture I: Autoregressive generative models (PixelCNN), WaveNet, and Normalizing Flows (Planar Flows, NICE, RealNVP)

Lecture II: Variational auto-encoders, Variational inference, importance weighted autoencoders (IWAE), Inverse Autoregressive Flows for VAE inference, and VAE models of video.

Lecture III: Generative adversarial networks and extensions (GANs, Wasserstein GAN, ALI, CycleGAN, BigGAN, StyleGAN, SPADE, HoloGAN, DVD GAN)

References

Autoregressive generative models:

The Neural Autoregressive Distribution Estimator by Hugo Larochelle and Iain Murray (AISTAT2011)

MADE: Masked Autoencoder for Distribution Estimation by Mathieu Germain, Karol Gregor, Iain Murray, Hugo Larochelle (ICML2015).

Pixel Recurrent Neural Networks by Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu (ICML2016)

Conditional Image Generation with PixelCNN Decoders by Aaron van den Oord, Nal Kalchbrenner, Oriol Vinyals, Lasse Espeholt, Alex Graves, Koray Kavukcuoglu (NIPS2016)

Normalizing Flows:

Chapter 20.10.2 of the Deep Learning textbook.

Sections 3.1, and 4 of Variational Inference with Normalizing Flows by Rezende and Mohamed, 2015

NFs as generative models: see Section 3.1 of Density Estimation using RealNVP by Laurent Dinh, 2016

For a unifying view, see Section 2 of Neural Autoregressive Flows by Huang et al., 2018

See the blog posts by Eric Jang: part 1, and part 2

Variational Autoencoders:

Chapter 20.10.3 of the Deep Learning textbook.

- Variational Inference, lecture note by David Blei. Section 1-6.
- Auto-Encoding Variational Bayes by Diederik Kingma (ICLR 2014) or Stochastic Backpropagation and Approximate Inference in Deep Generative Models by Danilo Rezende (ICML 2014).
- Importance Weighted Autoencoders by Yuri Burda (ICLR 2016)
- Inference Suboptimality in Variational Autoencoders by Chris Cremer (ICML 2018)
- Blog post Variational Autoencoder Explained by Goker Erdogan
- Blog post Families of Generative Models by Andre Cianflone

Generative Adversarial Networks:

- Sections 20.10.4 of the Deep Learning textbook.
- Generative Adversarial Networks by Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio (NIPS 2014).
- f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization by Sebastian Nowozin, Botond Cseke and Ryota Tomioka (NIPS 2016).
- NIPS 2016 Tutorial: Generative Adversarial Networks by Ian Goodfellow, arXiv:1701.00160v1, 2016
- Adversarially Learned Inference by Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Olivier Mastropietro, Alex Lamb, Martin Arjovsky and Aaron Courville (ICLR 2017).
- Wasserstein GAN (Arjovsky et al; 2017) Improved Training of Wasserstein GANs (Gulrajani et al; 2017)
- Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks (Zhu et al; 2017)
- Large Scale GAN Training for High Fidelity Natural Image Synthesis (Brock et al; 2018)

Pre-requisites

- Basic knowledge of probability theory, linear algebra and calculus.
- Basic knowledge of neural networks (architectures and training via back-propagation).
- Knowledge of convolutional neural networks would be a strong advantage.

Short Bio

Aaron Courville is an Associate Professor at the University of Montreal and founding member of the Mila institute. His research include generative models, generalization of machine learning models and applications of deep learning to natural language processing, computer vision and reinforcement learning. Together with Ian Goodfellow and Yoshua Bengio, he co-authored the widely read Deep Learning textbook (2016). He is a Fellow of the CIFAR Learning in Machines and Brains program and holder of a CIFAR Canadian AI chair. His research is supported by focused research awards from MSR and Google.



Issam El Naqa

(University of Michigan) [introductory/intermediate]

Deep Learning for Biomedicine Summary

Artificial intelligence (AI) and algorithms based on machine/deep learning (ML/DL) are witnessing tremendous resurgence in healthcare with the promise to transform the practice of medicine reducing its cost and improving treatment outcomes. Biomedicine is considered to be the leading venue of AI/ML/DL effort in the foreseeable future with applications ranging from process automation, diagnosis, and prognosis using imaging and genetic information. However, progress has been slower than its anticipated promise thus far. In this course, we will discuss the current applications of DL in biomedicine and draw attention to its specific challenges and prospects. We will present example applications and highlight future potentials.

Syllabus

- Overview of AI/ML/DL in medicine and its history
- AI hype versus current reality in healthcare
- Brief review of DL algorithms used in medicine and their categorization
- General applications of DL in medicine
 - Bioinformatics
 - Analysis of medical images
 - Computer aided diagnosis and detection
 - Clinical decision support systems
- Challenges of DL in biomedicine
 - Limited and noisy datasets
 - Privacy and patient protection concerns
 - Acceptance and commissioning
 - Clinical trial design
- Example applications
 - Drug discovery
 - Oncology
- Prospects and the clinical of the future

References

- Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, Basic Books, 2019.
- Machine Learning in Radiation Oncology: Theory and Applications. El Naqa, Li, Murphy (editors), Springer, 2015. (New edition is forth coming).
- Emerging Developments and Practices in Oncology: El Naqa (Editor), IGI Global, Hershey, PA, 2018.
- A Guide to Outcome Modeling In Radiotherapy and Oncology: Listening to the Data, El Naqa (Editor): CRC press: Taylor and Francis Group, Boca Raton, FL, USA, 2018.

Pre-requisites

None.

Short Bio

Issam El Naqa received his B.Sc. (1992) and M.Sc. (1995) in Electrical and Communication Engineering from the University of Jordan, Jordan. He worked as a software engineer at the Computer Engineering Bureau (CEB), Jordan, 1995-1996. He was awarded a DAAD scholarship to Germany, where he was a visiting scholar at the RWTH Aachen, 1996-1998. He completed his Ph.D. (2002) in Electrical and Computer Engineering from Illinois Institute of Technology, Chicago, IL, USA, receiving highest academic distinction award for his PhD work. He completed an M.A. (2007) in Biology Science from Washington University in St. Louis, St. Louis, MO, USA, where he was pursuing a post-doctoral fellowship in medical physics and was subsequently hired as a Instructor (2005-2007) and then an Assistant Professor (2007-2010) at the departments of radiation oncology and the division of biomedical and biological sciences and was an adjunct faculty at the department of Electrical engineering. He became an Associate Professor at McGill University Health Centre/Medical Physics Unit (2010-2015) and associate member of at the departments of Physics, Biomedical Engineering, and Experimental medicine, where he was a designated scholar. He is currently an Associate Professor of Radiation Oncology at the University of Michigan at Ann Arbor and associate member in Applied Physics. He is a certified Medical Physicist by the American Board of Radiology. He is a recognized expert in the fields of image processing, bioinformatics, computational radiobiology, and treatment outcomes modeling and has published extensively in these areas with more than 150 peer-reviewed journal publications and 3 edited textbooks. He has been an acting member of several academic and professional societies. His research has been funded by several federal and private grants and serves as a peer-reviewer and editorial board member for several leading international journals in his areas of expertise.



Sergei V. Gleyzer

(University of Florida) [introductory/intermediate]

Feature Extraction, End-end Deep Learning and Applications to Very Large Scientific Data: Rare Signal Extraction, Uncertainty Estimation and Realtime Machine Learning Applications in Software and Hardware Summary

Deep learning, and machine learning in general, has become one of the most widely used tools in modern science and engineering, leading to breakthroughs in a number of areas and disciplines ranging from computer vision to natural language processing to medical outcome analysis. This mini-course will introduce the basics of machine learning theory and classification theory based on statistical learning and describe two classes of popular algorithms in depth: decision-based methods (decision trees, decision rules, bagging and boosting, random forests) and deep neural network-based models of various types. The course will focus on practical applications in analysis of large scientific data, interpretability, uncertainty estimation and how to best extract meaningful features with autonomous feature extraction and feature engineering and how to implement realtime deep learning in software and hardware. No previous machine learning background is required.

Syllabus

- Introduction to Machine Learning: Theoretical Foundation, Classification Theory
- Practical Applications and Examples in Sciences and Engineering with Large Scientific Data (LHC/LSST)
- Tree-based Algorithms: decision trees, rules, bagging, boosting, random forests
- Deep Learning Methods: theory, fully-connected networks, convolutional, recurrent and recursive networks, graph networks and geometric deep learning
- Fundamentals of Feature Extraction and End-to-end Deep Learning
- Uncertainty Estimation and Machine Learning Model Interpretations
- Realtime Implementation of Deep Learning in Software and Hardware

References

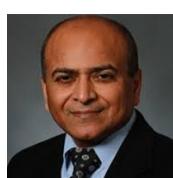
- I. Goodfellow, Y. Bengio and A. Courville, “Deep Learning” MIT Press 2016
- G. James et al., “Introduction to Statistical Learning” Springer 2013
- C.M. Bishop “Pattern Recognition and Machine Learning” Springer 2006
- J. R. Quinlan “C4.5: Programs for Machine Learning” Morgan Kaufmann 1992

Pre-requisites

None

Short Bio

Sergei Gleyzer is a particle physicist and university professor, working at the interface of particle physics and machine learning towards more intelligent systems to extract meaningful information from the data collected by the Large Hadron Collider (LHC), the world’s highest-energy particle physics experiment located at the CERN laboratory, near Geneva Switzerland. He is the a co-discover of the Higgs Boson and founder of several major machine learning initiatives such as the Inter-experimental Machine Learning Working Group and Compact Muon Solenoid experiment’s Machine Learning Forum. Professor Gleyzer is working on applying advanced machine learning methods to searches for new physics, such as dark matter.



Vasant Honavar

(Pennsylvania State University) [introductory/intermediate]

Causal Models for Making Sense of Data Summary

Some have argued that the advent of big data and machine learning spells the end of theory, and that big data makes scientific method obsolete. If only we let powerful computers to crunch through big data, statistical machine learning algorithms will find patterns where science cannot, they claim. In this course, we will show that, big data, instead of making the scientific method obsolete, makes ever increasing demands on it. We will argue for the importance of causal models in making sense of big data. We will show how causal models can be used to identify (and hence adjust for confounders). We will show how causal models can be learned from data. We will see how the causal lens allows us to peer into black box models produced by machine learning, to explain the predictions they make, and thus provide powerful tools for enhancing the transparency, explainability, and fairness of complex predictive models trained using machine learning. If time permits, we will touch upon additional topics such as causal transportability.

Syllabus

Importance of causal models in making sense of data, big and small. What can we do with causal models that we cannot using predictive models trained using traditional machine learning methods? Counterfactual inference. Identifying and adjusting for confounders. Learning causal models from data. Sample applications: Explaining black box predictive models using causal inference. Algorithmic fairness.

Pre-requisites

Basic probability and statistics. Some exposure to machine learning.

References

- ▶ Pearl, J., Glymour, M. and Jewell, N.P., 2016. Causal inference in statistics: A primer. John Wiley & Sons.
- ▶ Shipley, B., 2016. Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference with R. Cambridge University Press.
- ▶ Spirtes, P., Glymour, C.N., Scheines, R., Heckerman, D., Meek, C., Cooper, G. and Richardson, T., 2000. Causation, prediction, and search. MIT press.
- ▶ Pearl, J., 2000. Causality: models, reasoning and inference, Cambridge: MIT press.
- ▶ Hernan, M.A. and Robins, J.M., 2010. Causal inference. Boca Raton, FL: CRC.
- ▶ Spirtes, Peter. "Introduction to causal inference." Journal of Machine Learning Research (2010) 11:1643-1662.
- ▶ Khademi, A., Lee, S., Foley, D. and Honavar, V., 2019, Fairness in Algorithmic Decision Making: An Excursion Through the Lens of Causality. In The World Wide Web Conference (pp. 2907-2914). ACM.
- ▶ Kusner, M.J., Loftus, J., Russell, C. and Silva, R., 2017. Counterfactual fairness. In Advances in Neural Information Processing Systems (pp. 4066-4076).
- ▶ Bareinboim, E., Lee, S., Honavar, V. and Pearl, J., 2013. Transportability from multiple environments with limited experiments. In Advances in Neural Information Processing Systems (pp. 136-144).
- ▶ Lee, S. and Honavar, V., 2016, On learning causal models from relational data. In Thirtieth AAAI Conference on Artificial Intelligence.
- ▶ Chattopadhyay, A., Manupriya, P., Sarkar, A. and Balasubramanian, V.N., 2019. Neural Network Attributions: A Causal Perspective. arXiv preprint arXiv:1902.02302.

Short Bio

Vasant Honavar received the B.E. degree in electronics engineering from Bangalore University, Bangalore, India, and the M.S. and Ph.D. degrees in computer science from the University of Wisconsin–Madison, Madison, WI, USA. Honavar was on the faculty of Computer Science, Iowa State University, Ames, IA, USA, from 1990 to 2013. From 2010 to 2013, he served as a Program Director of the Information and Intelligent Systems Division with the National Science Foundation (NSF), where he led the Big Data Program, and contributed to the Smart and Connected Health, Information Integration and Informatics, and Expeditions in Computing programs. He currently holds the Edward Frymoyer Endowed Professorship in Information Sciences and Technology, and serves as a Professor of Computer Science, Bioinformatics and Genomics, Data Sciences, Informatics, and Neuroscience with The Pennsylvania State University (PSU), University Park, PA, USA. At PSU, he directs the Artificial Intelligence Research Laboratory and the Center for Big Data Analytics and Discovery Informatics, co-directs the NIH-Funded Biomedical Big Data to Knowledge Pre-Doctoral Training Program, and serves as an Associate Director of the Penn State Institute for CyberScience. He also serves as the Sudha Murty Distinguished Visiting Chair of Neurocomputing and Data Science at the Indian Institute of Science, Bangalore, India. Honavar's work (documented in over 250 peer-reviewed publications, with over 13,400 citations during 1990 to 2019) has resulted in foundational contributions in scalable approaches to learning predictive models from very large, richly structured data, including tabular, sequence, network, relational, and time series data; eliciting causal information from multiple sources of observational and experimental data; selective sharing of knowledge across disparate knowledge bases; representing and reasoning about preferences; composing complex software services from components; and applications in bioinformatics and computational molecular and systems biology, including characterization, analysis, and prediction of sequence and structural correlates of protein–protein, and protein–RNA interfaces and interactions, comparative analysis of biological networks (network alignment). His current research focuses on: Computational abstractions scientific artifacts (e.g., data, knowledge, hypotheses), and universes of scientific discourse (e.g., biology), and scientific processes (e.g., hypothesis generation, predictive modeling, experimentation, simulation, and hypothesis testing), cognitive tools that augment and extend human intellect; and human-machine infrastructure (including data and computational infrastructure and organizational structures and processes) to accelerate science; Design and analysis of algorithms for predictive modeling from very large, high dimensional, richly structured, multi-modal, longitudinal data; Representation learning from richly structured data; Elucidation of causal relationships from disparate experimental and observational studies; Elucidation of causal relationships from relational, temporal, and temporal-relational data; Design and analyses of accountable, explainable, and fair AI systems; Analysis and prediction of macromolecular interactions, elucidation of complex biological pathways e.g., those involved in immune response, development, and disease; Predictive and causal modeling of individual and population health outcomes from behavioral, biomedical, clinical, environmental, socio-demographic data; Accelerating materials discovery using machine learning; and Modeling the structure, activity, and function of brain networks from fMRI and other types of data. Honavar is a Fellow of the American Association for the Advancement of Science, a Distinguished Member of the Association for Computing Machinery (ACM), a Senior Member of the Association for the Advancement of Artificial Intelligence, and a Senior Member of IEEE. He has received many awards and honors during his career, including the NSF Director's Award for Superior Accomplishment in 2013 for his leadership of the NSF Big Data Program, the Iowa Board of Regents Award for Faculty Excellence in 2007, the Iowa State University College of Liberal Arts and Sciences Award for Career Excellence in Research in 2008, the Iowa State University Margaret Ellen White Graduate Faculty Award in 2011, and the Pennsylvania State University College of Information Sciences and Technology Research Excellence Award in 2016.



Qiang Ji

(Rensselaer Polytechnic Institute) [introductory/intermediate]

Probabilistic Deep Learning for Computer Vision Summary

Deep learning has become a major enabling technology for computer vision. By exploiting its multi-level representation and the availability of big data, deep learning has led to dramatic performance improvements for certain tasks. Despite the significant progresses they have made, existing deep learning methods are deterministic and cannot effectively quantify their prediction uncertainty. Uncertainty quantification is not only important to improve the algorithm's performance but is also essential for my practical applications. Furthermore, existing deep learning methods perform point-based prediction. Point-based prediction not only requires a time-consuming and heuristic training process but also suffers from overfitting and poor adaptation to novel conditions. Through this lecture, I will introduce probabilistic deep learning, where the deep models capture the probabilistic distribution of inputs and outputs, and produce not only a prediction but also its probability distribution. The lecture consists of 4 parts. In part 1, I will review the basic probability calculus and fundamental concepts in machine learning. Part 2 will cover deep probabilistic neural networks and deep Bayesian neural networks. In part 3, I will discuss deep probabilistic graphical models. The lecture will conclude with a discussion of applications of probabilistic deep learning to different computer vision and image processing tasks.

Syllabus

- Basics and Fundamentals
 - a. Probability calculus
 - b. Machine learning fundamentals
- Probabilistic Neural networks
 - a. Neural networks, deep neural networks, and convolutional neural networks
 - b. Deep probabilistic neural networks
 - c. Deep Bayesian neural networks
- Deep Probabilistic Graphical Models
 - a. Probabilistic graphical models (PGMs)
 - i. Directed PGMs
 - ii. Undirected PGMs
 - iii. PGM learning and inference
 - b. Deep probabilistic graphical models
 - i. Deep Boltzmann machine
 - ii. Deep belief network
 - iii. Deep Bayesian networks
- Deep Probabilistic Models Applications in Computer Vision
 - a. Image restoration and denoising
 - b. Object detection and recognition
 - c. Object pose estimation and prediction
 - d. Image reconstruction and synthesis

References

- Deep Bayesian Neural Networks <https://medium.com/@joeDiHare/deep-bayesian-neural-networks-952763a9537>
- Bayesian Deep Learning
https://www.cs.ox.ac.uk/people/yarin.gal/website/PDFs/2017_OReilly_talk.pdf
- Bayesian Deep Learning Workshop <http://bayesiandeeplearning.org/>
- Deep Regression Bayesian Network <https://arxiv.org/abs/1710.04809>
- Deep Boltzmann Machine
<http://proceedings.mlr.press/v5/salakhutdinov09a/salakhutdinov09a.pdf>

- Deep Belief Networks https://en.wikipedia.org/wiki/Deep_belief_network
<https://codeburst.io/deep-learning-deep-belief-network-fundamentals-d0dcfd80d7d4>
<https://www.cs.toronto.edu/~hinton/nipstutorial/nipstut3.pdf>
- Probabilistic Graphical Models: Principles and Techniques, Daphne Koller and Nir Friedman
- Probabilistic Graphical Models for Computer Vision, Qiang Ji, Manuscript, available upon request

Pre-requisites

Prior knowledge in probability, calculus, linear algebra, and optimization methods. Familiarity with basic machine learning and computer vision techniques.

Short Bio

Qiang Ji received his Ph.D degree in Electrical Engineering from the University of Washington. He is currently a Professor with the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute (RPI), Troy, NY, USA. He previously served as a program director at the US National Science Foundation (NSF), where he managed NSF's computer vision and machine learning programs. He also held teaching and research positions with the Beckman Institute at University of Illinois at Urbana-Champaign, Urbana, IL, USA; the Robotics Institute at Carnegie Mellon University, Pittsburgh, PA, USA; the Dept. of Computer Science at University of Nevada, Reno, Nevada, USA; and the Air Force Research Laboratory, Rome, NY, USA. Prof. Ji currently serves as the director of the Intelligent Systems Laboratory (ISL) at RPI.

Prof. Ji's research interests are in computer vision, probabilistic graphical models, machine learning, and their applications in various fields. He has published over 300 papers in peer-reviewed journals and conferences, and has received multiple awards for his work. Prof. Ji is has served as an editor on several related IEEE and international journals and as a general chair, program chair, technical area chair, and program committee member for numerous international conferences/workshops. Prof. Ji is a fellow of the IEEE and the IAPR.



James Kwok

(Hong Kong University of Science and Technology) [introductory/intermediate]

Compressing Neural Networks Summary

Deep neural networks have been hugely successful in various domains, such as computer vision, speech recognition, and natural language processing. Though powerful, the large number of network weights leads to space and time inefficiencies in both training and storage. Recently, attempts have been made to reduce the model size. They include sparsification using pruning and sparsity regularization, quantization to replace the weights and activations with fewer number of bits, low-rank approximation, distillation and the use of more compact structures. These attempts greatly reduce the network size, and allows possibility of deploying deep models in resource-constrained environments, such as embedded systems, smart phones and other portable devices.

Syllabus

- introduction to neural networks and deep learning
- network sparsification using pruning and sparsity regularizers
- network quantization using fewer bits
- low-rank approximation
- distillation and more compact models

References

- S. Han, J. Pool, J. Tran, and W.J. Dally. Learning both weights and connections for efficient neural networks. NIPS, 2015.
- W. Wen, C. Wu, Y. Wang, and Y. Chen. Learning structured sparsity in deep neural networks. NIPS, 2016.
- M. Courbariaux, Y. Bengio, and J.P. David. BinaryConnect: Training deep neural networks with binary weights during propagations. NIPS, 2015.
- L. Hou, Q. Yao, J.T. Kwok. Loss-aware binarization of deep networks. ICLR, 2017.
- G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. arXiv preprint, 2015.

Pre-requisites

A general knowledge of machine learning and neural networks is requested.

Short Bio

Prof. Kwok is a Professor in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. He received his B.Sc. degree in Electrical and Electronic Engineering from the University of Hong Kong and his Ph.D. degree in computer science from the Hong Kong University of Science and Technology. Prof. Kwok served/is serving as an Associate Editor for the IEEE Transactions on Neural Networks and Learning Systems, Neurocomputing and the International Journal of Data Science and Analytics. He has also served as Program Co-chair of a number of international conferences, and as Area Chairs in conferences such as NIPS, ICML, ECML, AAAI and IJCAI. He is an IEEE Fellow.



Tomas Mikolov

(Facebook) [introductory]

Using Neural Networks for Modeling and Representing Natural Languages (with Piotr Bojanowski and Armand Joulin) Summary

In this tutorial, we will describe the basics of machine learning and artificial neural networks when applied to the natural language processing tasks. The tutorial will have three parts: first, we will introduce neural networks and discuss popular training algorithms such as stochastic gradient descent, as well as related topics such as learning rate, regularization, backpropagation, and various neural architectures.

In the second part, we will cover representational learning from text: this includes algorithms such as word2vec and fastText. We will describe the differences between various algorithms for learning representations. Efficient supervised text classification with the fastText algorithm will also be discussed. In the last part of the tutorial, statistical language models based on neural networks will be introduced, and certain advanced topics such as vanishing and exploding gradients, as well as learning the longer term memory in recurrent networks will be explained. We will also talk about the limitations of the current learning algorithms, and discuss the limitations of generalization in the context of sequential data, and learning from language in general.

Syllabus

- Introduction to neural networks, deep learning and NLP (1.5 hours)
- Representational learning in NLP (1.5 hours)
- Neural language models (1.5 hours)

References

- Statistical Language Models Based on Neural Networks, T Mikolov, 2012
- Enriching word vectors with subword information, P Bojanowski et al, 2016
- Bag of tricks for efficient text classification, A Joulin et al, 2016
- fasttext.cc

Pre-requisites

Basic knowledge in linear algebra and programming.

Short Bio

Tomas Mikolov: I am a research scientist at Facebook AI Research since May 2014. Previously I have been member of Google Brain team, where I developed and implemented efficient algorithms for computing distributed representations of words (word2vec project). I have obtained my PhD from Brno University of Technology (Czech Republic) for my work on recurrent neural network based language models (RNNLM). My long term research goal is to develop intelligent machines capable of learning and communication with people using natural language.

Armand Joulin: I am a research scientist at Facebook Artificial Intelligence Research. I obtained my PhD in 2012 from the INRIA and the Ecole Normale Supérieure. My advisors were Francis Bach and Jean Ponce. Before joining Facebook, I was a postdoctoral fellow at Stanford University, working with Daphne Koller and Fei-Fei Li.

Piotr Bojanowski: I am a research scientist at Facebook AI Research, working on machine learning applied to computer vision and natural language processing. My main research interest revolve around large-scale unsupervised learning. Before joining Facebook, in 2016, he got a PhD in Computer Science at the Willow team (INRIA Paris) under the supervision of Jean Ponce, Cordelia Schmid, Ivan Laptev and Josef Sivic. He graduated from Ecole polytechnique in 2013 and received a Masters Degree in Mathematics, Machine Learning and Computer Vision (MVA).



Hermann Ney

(RWTH Aachen University) [intermediate/advanced]

Speech Recognition and Machine Translation: From Statistical Decision Theory to Machine Learning and Deep Neural Networks Summary

The last 40 years have seen a dramatic progress in machine learning and statistical methods for speech and language processing like speech recognition, handwriting recognition and machine translation. Many of the key statistical concepts had originally been developed for speech recognition and language translation. Examples of such key concepts are the Bayes decision rule for minimum error rate and sequence-to-sequence processing using approaches like the alignment mechanism based on hidden Markov models and the attention mechanism based on neural networks. Recently the accuracy of speech recognition and machine translation could be improved significantly by the use of artificial neural networks, such as deep feedforward multi-layer perceptrons and recurrent neural networks (incl. long short-term memory extension). We will discuss these approaches in detail and how they form part of the probabilistic approach.

Syllabus

- Part 1: Statistical Decision Theory, Machine Learning and Neural Networks.
- Part 2: Speech Recognition (Time Alignment, Hidden Markov models, sequence-to-sequence processing, neural nets, attention models).
- Part 3: Machine Translation (Word Alignment, Hidden Markov models, sequence-to-sequence processing, neural nets, attention models).

References

- Bourlard, H. and Morgan, N., Connectionist Speech Recognition - A Hybrid Approach, Kluwer Academic Publishers, ISBN 0-7923-9396-1, 1994.
- L. Deng, D. Yu: Deep learning: methods and applications. Foundations and Trends in Signal Processing, Vol. 7, No. 3–4, pp. 197–387, 2014.
- D. Jurafsky, J. H. Martin: Speech and Language Processing. Third edition draft, pdf; August 28, 2017.
- Y. Goldberg: Neural Network Methods in Natural Language Processing. Morgan & Claypool Publishers, Draft, pdf; August 2016.
- P. Koehn: Statistical Machine Translation, Cambridge University Press, 2010. In addition: Draft of Chapter 13: Neural Machine Translation, pdf, September 22, 2017.

Pre-requisites

Familiarity with linear algebra, numerical mathematics, probability and statistics, elementary machine learning.

Short Bio

Hermann Ney is a full professor of computer science at RWTH Aachen University, Germany. His main research interests lie in the area of statistical classification, machine learning, neural networks and human language technology and specific applications to speech recognition, machine translation and handwriting recognition.

In particular, he has worked on dynamic programming and discriminative training for speech recognition, on language modelling and on machine translation. His work has resulted in more than 700 conference and journal papers (h-index 95, 50000+ citations; estimated using Google scholar). He and his team contributed to a large number of European (e.g. TC-STAR, QUAERO, TRANSLECTURES, EU-BRIDGE) and American (e.g. GALE, BOLT, BABEL) large-scale joint projects.

Hermann Ney is a fellow of both IEEE and ISCA (Int. Speech Communication Association). In 2005, he was the recipient of the Technical Achievement Award of the IEEE Signal Processing Society. In 2010, he was awarded a senior DIGITEO chair at LIMIS/CNRS in Paris, France. In 2013, he received the award of honour of the International Association for Machine Translation. In 2016, he was awarded an advanced grant of the European Research Council (ERC).



Jose C. Principe

(University of Florida) [intermediate/advanced]

Cognitive Architectures for Object Recognition in Video Summary

I-Requisites for a Cognitive Architecture (intermediate)

- Processing in space
- Processing in time with memory

- Top down and bottom processing
- Extraction of information from data with generative models
- Attention

II- Putting it all together (intermediate)

- Empirical Bayes with generative models
- Clustering of time series with linear state models

III- Current work (advanced)

- Information Theoretic Autoencoders
- Attention Based video recognition
- Augmenting Deep Learning with memory

Short Bio

Jose C. Principe is a Distinguished Professor of Electrical and Computer Engineering at the University of Florida where he teaches advanced signal processing, machine learning and artificial neural networks (ANNs). He is Eckis Professor and the Founder and Director of the University of Florida Computational NeuroEngineering Laboratory (CNEL) www.cnel.ufl.edu. The CNEL Lab innovated signal and pattern recognition principles based on information theoretic criteria, as well as filtering in functional spaces. His secondary area of interest has focused in applications to computational neuroscience, Brain Machine Interfaces and brain dynamics. Dr. Principe is a Fellow of the IEEE, AIMBE, and IAMBE. Dr. Principe received the Gabor Award, from the INNS, the Career Achievement Award from the IEEE EMBS and the Neural Network Pioneer Award, of the IEEE CIS. He has more than 38 patents awarded over 800 publications in the areas of adaptive signal processing, control of nonlinear dynamical systems, machine learning and neural networks, information theoretic learning, with applications to neurotechnology and brain computer interfaces. He directed 97 Ph.D. dissertations and 65 Master theses. He wrote in 2000 an interactive electronic book entitled "Neural and Adaptive Systems" published by John Wiley and Sons and more recently co-authored several books on "Brain Machine Interface Engineering" Morgan and Claypool, "Information Theoretic Learning", Springer, "Kernel Adaptive Filtering", Wiley and "System Parameter Adaption: Information Theoretic Criteria and Algorithms", Elsevier. He has received four Honorary Doctor Degrees, from Finland, Italy, Brazil and Colombia, and routinely serves in international scientific advisory boards of Universities and Companies. He has received extensive funding from NSF, NIH and DOD (ONR, DARPA, AFOSR).



Fabio Roli

(University of Cagliari) [introductory/intermediate]

Adversarial Machine Learning Summary

This tutorial aims to introduce the fundamentals of adversarial machine learning, presenting a well-structured review of recently-proposed techniques to assess the vulnerability of machine-learning algorithms to adversarial attacks (both at training and at test time), and some of the most effective countermeasures proposed to date. We consider these threats in different application domains, including object recognition in images, biometric identity recognition, spam and malware detection.

This tutorial motivates and explains a topic of emerging importance for AI, and it is particularly devoted to:

- people who want to become aware of the new research field of adversarial machine learning and learn the fundamentals;
- people doing research in machine learning, AI safety, and pattern recognition applications which have a potential adversarial component, and wish to learn how the techniques of adversarial classification can be effectively used in such applications.

Syllabus

- Introduction to adversarial machine learning. Introduction by practical examples from computer vision, biometrics, spam and malware detection. Previous work on adversarial learning and recognition. Basic concepts and terminology. The concept of adversary-aware classifier. Definitions of attack and defense.
- Design of learning-based pattern classifiers in adversarial environments. Modelling adversarial tasks. The two-player model (the attacker and the classifier). Levels of reciprocal knowledge of the two players (perfect knowledge, limited knowledge, knowledge by queries and feedback). The concepts of security by design and security by obscurity.
- System design: vulnerability assessment and defense strategies. Attack models against pattern classifiers. The influence of attacks on the classifier: causative and exploratory attacks. Type of security violation: integrity, availability and privacy attacks. The specificity of the attack: targeted and indiscriminate attacks. Vulnerability assessment by performance evaluation. Taxonomy of possible defense strategies. Examples from computer vision, biometrics, spam and malware detection. Hands-on web demo on adversarial examples, "Deep Learning security".
- Summary and outlook. Current state of this research field and future perspectives.

References

- Biggio, B., Roli, F. Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning. ArXiv, 2017. (tutorial-related article)
- Barreno, M., Nelson, B., Sears, R., Joseph, A. D., Tygar, J. D. Can machine learning be secure? ASIACCS, 2006.
- Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B., Tygar, J. D. Adversarial machine learning. AISec, 2011.
- Biggio, B., Nelson, B., Laskov, P. Poisoning attacks against SVMs. ICML, 2012.
- Biggio, B., Corona, I., Maiorca, D., Nelson, B., Srndic, N., Laskov, P., Giacinto, G., Roli, F. Evasion attacks against machine learning at test time. ECML-PKDD, 2013.
- Biggio, B., Fumera, G., Roli, F. Security evaluation of pattern classifiers under attack. IEEE Trans. Knowl. Data Eng., 2014.
- Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., Fergus, R. Intriguing properties of neural networks. ICLR, 2014.
- Xiao, H., Biggio, B., Brown, G., Fumera, G., Eckert, C., Roli, F. Is feature selection secure against training data poisoning? ICML, 2015.
- Nguyen, A. M., Yosinski, J., Clune, J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. CVPR, 2015.
- Goodfellow, I., Shlens, J., Szegedy, C. Explaining and harnessing adversarial examples. ICLR, 2015.
- Moosavi-Dezfooli, S.-M., Fawzi, A., Frossard, P. Deepfool: a simple and accurate method to fool deep neural networks. CVPR, 2016.
- Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B., Swami, A. The limitations of deep learning in adversarial settings. IEEE Euro S&P, 2016.
- Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., Swami, A. Practical black-box attacks against machine learning. ASIACCS, 2017.
- Carlini N., Wagner, D. Towards evaluating the robustness of neural networks. IEEE Symp. SP, 2017.
- Demontis, A., Melis, M., Biggio, B., Maiorca, D., Arp, D., Rieck, K., Corona, I., Giacinto, G., Roli, F. Yes, machine learning can be more secure! a case study on Android malware detection. IEEE Trans. Dependable and Secure Comp., 2017.
- Melis, M., Demontis, A., Biggio, B., Brown, G., Fumera, G., Roli, F. Is deep learning safe for robot vision? Adversarial examples against the iCub humanoid. In ICCV Workshop ViPAR, 2017.
- Athalye, A., Carlini, N., Wagner, D. Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples. ArXiv, 2018.

- Athalye, A., Engstrom, L., Ilyas, A., Kwok, K. Synthesizing robust adversarial examples. ICLR, 2018.
- Jagielski, M., Oprea, A., Biggio, B., Liu, C., Nita-Rotaru, C., Li, B. Manipulating machine learning: Poisoning attacks and countermeasures for regression learning. IEEE Symp. SP, 2018.

Pre-requisites

No knowledge of the tutorial topics is assumed. A basic knowledge of machine learning and statistical pattern classification is requested.

Short Bio

Fabio Roli is a Professor of Computer Engineering at the University of Cagliari, Italy, and Director of the Pattern Recognition and Applications laboratory that he founded from scratch in 1995 and it is now a world-class research lab with 30 staff members, including five tenured Faculty members. He is the R&D manager of the company Pluribus One that he co-founded. He has been doing research on the design of pattern recognition systems for 30 years. Prof. Roli has published 86 journal articles and more than 250 conference articles on pattern recognition and machine learning and many of his papers are frequently cited. His current h-index is 59 according to Google Scholar (March 2019). He has been appointed Fellow of the IEEE and Fellow of the IAPR. He was the President of the Italian Group of Researchers in Pattern Recognition and the Chairman of the IAPR Technical Committee on Statistical Techniques in Pattern Recognition. He was a member of the NATO advisory panel for Information and Communications Security, NATO Science for Peace and Security (2008 – 2011). Prof. Roli is one of the pioneers of the use of pattern recognition and machine learning for computer security. He is often invited to give keynote speeches and tutorials on adversarial machine learning and data-driven technologies for security applications. He is (or has been) the PI of dozens of R&D projects, including the leading European projects on Security & Privacy CyberRoad and ILLBuster.



Björn Schuller

(Imperial College London) [introductory/intermediate]

Deep Learning for Intelligent Signal Processing Summary

This course will deal with deep learning algorithms into multimodal and multisensorial signal analysis such as from audio, video, text, or physiological signals. Methods shown will, however, be applicable to a broad range of further signal types. We will first deal with pre-processing for denoising or dereverberation. This will be followed by representation learning such as by convolutional neural networks or sequence-to-sequence encoder-decoder architectures as basis for end-to-end learning from raw signals or symbolic representation. Then, we shall discuss modelling for decision making such as by recurrent neural networks with long-short-term memory or gated recurrent units including compensation of dynamics by connectionist temporal classification. This will also include discussion of the usage of attention on different levels. We will also elaborate on the impact of topologies including multiple targets with shared layers and bottlenecks, and how to move towards self-shaping networks in the sense of Automatic Machine Learning. In a last part, we will deal with data efficiency, such as by weak supervision with the human in the loop based on data augmentation, active and semi-supervised learning, transfer learning, or generative adversarial networks. The content shown will be accompanied by open-source implementations of according toolkits available on github. Application examples will come from the domains of Affective Computing, Multimedia Retrieval, and mHealth.

Syllabus

Pre-Processing and Representation Learning (CNNs, S2S, end-to-end)
 Modelling for Decision Making (Attention, Feature Space Optimisation, RNNs, LSTM, GRUs, CTC, AutoML)
 Data Efficiency (GANs, Transfer Learning, Data Augmentation, Weak Supervision, Cooperative Learning)

References

- ▶ The Handbook of Multimodal-Multisensor Interfaces. Vol. 2, S. Oviatt, B. Schuller, P.R. Cohen, D. Sonntag, G. Potamianos, A. Krüger (eds.), 2018 <https://github.com/end2you/end2you>
- ▶ <https://github.com/openXBOW/openXBOW>
- ▶ <https://github.com/auDeep/auDeep>
- ▶ <https://github.com/DeepSpectrum/DeepSpectrum>
- ▶ <https://www.audeering.com/opensmile/>
- ▶ <https://www.ihearuplay.eu/>

Pre-requisites

Attendees should be generally familiar with Machine Learning and Neural Networks in general. They should further have basic knowledge of Signal Processing.

Short Bio

Björn W. Schuller received his diploma, doctoral degree, habilitation, and Adjunct Teaching Professor in Machine Intelligence and Signal Processing all in EE/IT from TUM in Munich/Germany. He is Full Professor of Artificial Intelligence and the Head of GLAM - the Group on Language Audio & Music - at Imperial College London/UK, Full Professor and ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing at the University of Augsburg/Germany, co-founding CEO and current CSO of audEERING - an Audio Intelligence company based near Munich and in Berlin/Germany, and permanent Visiting Professor at HIT/China amongst other Professorships and Affiliations. Before, he was Full Professor at the University of Passau/Germany, with Joanneum Research in Graz/Austria, and the CNRS-LIMSI in Orsay/France. He is a Fellow of the IEEE, President-Emeritus of the AAC, and Senior Member of the ACM. He (co-)authored 800+ publications (23000+ citations, h-index=70), and was Editor in Chief of the IEEE Transactions on Affective Computing, is General Chair of ACII 2019, ACII Asia 2018, and ACM ICMI 2014, and a Program Chair of Interspeech 2019, ACM ICMI 2019/2013, ACII 2015/2011, and IEEE SocialCom 2012 amongst manifold further commitments and service to the community. His 30+ awards include having been honoured as one of 40 extraordinary scientists under the age of 40 by the WEF in 2015. He served as Coordinator/PI in 10+ European Projects, is an ERC Starting Grantee, and consultant of companies such as Barclays, GN, Huawei, or Samsung.



Alex Smola

(Amazon) [introductory]

Dive into Deep Learning Summary

Dive into Deep Learning

Syllabus

- Backpropagation and automatic differentiation
- Multilayer perceptrons, regularization and model selection
- Convolutional neural networks (LeNet, AlexNet, GoogLeNet, ResNet, ResNext, etc.)
- Sequence models (RNN, GRU, LSTMs)

Distributed optimization (time permitting)

References

For details see <http://www.d2l.ai> and the associated UC Berkeley STAT-157 class at <http://courses.d2l.ai>. The course will cover a set of chapters from the book / website. All slides, notebooks and text are available on Github at <https://github.com/diveintodeeplearning/d2l-en.P>

Pre-requisites

- Python (experience with NumPy)
- Basic Linear Algebra and Statistics
- Basic optimization knowledge is a plus but not required

Short-Bio

Alex Smola is Distinguished Scientist/VP at Amazon Web Services in Palo Alto and Adjunct Professor at UC Berkeley. Prior to that he was full professor at Carnegie Mellon University. He worked at Google, Yahoo, the Australian National University and National ICT Australia. Alex received his PhD at TU Berlin. His research interests are Kernel Methods, Bayesian Nonparametrics, Deep Learning, Systems and Machine Learning, and algorithms on Graphs. He has written (or edited) 5 books and has published over 200 papers.



Sargur Srihari

(University at Buffalo) [intermediate/advanced]

Explainable Artificial Intelligence Summary

Today's AI approaches based on deep learning perform perceptual and other tasks exceedingly well. However the methods optimize the solution to each task, without considering the interpretability of the solution by humans. In tasks where human judgment is a necessary component, as in medicine and the courtroom, it is necessary for the decision by the AI system to be accompanied by an explanation. We will describe different types of explainability. Then go over several approaches to explainability in AI, emphasizing probabilistic approaches. We will take the example of forensic comparison and show how a high performance deep learning system and an explainable system can coexist.

Syllabus

- Definitions of Explainability
- Performance-Explainability trade-off of machine learning techniques
- Probabilistic AI and Most Probable Explanation
- Deep Learning: Architectures and Opaqueness
- The forensic comparison problem: coexistence of deep learning and explainability

References

- <https://www.darpa.mil/attachments/XAIProgramUpdate.pdf>
- <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8466590>
- <https://arxiv.org/pdf/1612.04757v1.pdf>

Pre-requisites

A course on Introductory machine learning covering the main topics such as described in <https://cedar.buffalo.edu/~srihari/CSE574/index.html>

Short Bio

Srihari is a SUNY Distinguished Professor in the Department of Computer Science and Engineering at the University at Buffalo, The State University of New York. He teaches a sequence of three courses in artificial intelligence and machine learning: (i) introduction to machine learning, (ii) probabilistic graphical models and (iii) deep learning. Srihari's work led to the world's first automated system for reading handwritten postal addresses. It was deployed by the United States Postal Service saving hundreds of millions of dollars in labor costs. A side-effect was that it led to the task of recognizing handwritten digits to be considered the fruit-fly of AI methods. Srihari also spent a decade developing AI and machine learning methods for forensic pattern evidence such as latent prints, handwriting and footwear impressions. In particular, quantifying the value of handwriting evidence-- to allow presenting such testimony in US courts. Srihari's honors include: Fellow of the IEEE, Fellow of the International Association for Pattern Recognition and distinguished alumnus of the Ohio State University College of Engineering Srihari received a B.Sc. from the Bangalore University, a B.E. from the Indian Institute of Science and a Ph.D. in Computer and Information Science from the Ohio State University.



Ponnuthurai N Suganthan

(Nanyang Technological University) [introductory/intermediate]

Learning Algorithms for Classification, Forecasting and Visual Tracking Summary

This presentation will primarily focus on learning algorithms with reduced iterations or no iterations at all. Some of the algorithms have closed form solutions while some of the algorithms do not adjust the structures once constructed. The main algorithms considered in this talk are randomized neural networks, kernel ridge regression and random forest. These non-iterative methods have attracted attention of researchers due to their high performance in terms of accuracy as well as their ability to train fast due to their non-iterative properties or closed form solutions. For example, the random forest delivers the top classification performance. The presentation will also include the basic methods as well as their state of the art variants. These algorithms will be benchmarked using classification, time series forecasting and visual tracking datasets. Future research directions will also be suggested.

Syllabus

- Non-iterative algorithms or algorithms with closed-form training solutions
- Randomization based neural networks and their variants
- Kernel Ridge Regression and their variants
- Random Forest and their variants
- Applications of the above methods in classification, time series and visual tracking
- Benchmarking of these methods

References

(Additional References will be included in the lecture materials)

- R. Katuwal, PN Suganthan, Multi-layer Random Vector Functional Link for Classification, submitted in Dec. 2018.
- R. Katuwal, PN Suganthan, L. Zhang, Heterogeneous Oblique Random Forest, submitted in Nov. 2018.

- PN Suganthan, "On non-iterative learning algorithms with closed-form solution," Applied Soft Computing 70, 1078-1082, 2018
- X Qiu, PN Suganthan, GAJ Amaratunga, Ensemble incremental learning Random Vector Functional Link network for short-term electric load forecasting Knowledge-Based Systems 145, 182-196, 2018.
- L Zhang, PN Suganthan, Benchmarking Ensemble Classifiers with Novel Co-Trained Kernel Ridge Regression and Random Vector Functional Link Ensembles [Research Frontier], IEEE Computational Intelligence Magazine 12 (4), 61-72, 2017.
- L Zhang, PN Suganthan, Visual tracking with convolutional random vector functional link network, IEEE Transactions on Cybernetics 47 (10), 3243-3253, 2017.
- L Zhang, PN Suganthan, Robust visual tracking via co-trained Kernelized correlation filters, Pattern Recognition 69, 82-93, 2017.
- L Zhang, PN Suganthan, A survey of randomized algorithms for training neural networks, Information Sciences 364, 146-155, 2016.
- L Zhang, P.N. Suganthan, Oblique decision tree ensemble via multisurface proximal support vector machine, IEEE Transactions on Cybernetics 45 (10), 2165-2176, 2015.

Pre-requisites

Basic knowledge of neural networks, pattern classification, kernel methods, and decision trees will be advantageous.

Short Bio

Ponnuthurai Nagaratnam Suganthan (or P N Suganthan) received the B.A degree, Postgraduate Certificate and M.A degree in Electrical and Information Engineering from the University of Cambridge, UK in 1990, 1992 and 1994, respectively. After completing his PhD research in 1995, he served as a pre-doctoral Research Assistant in the Dept of Electrical Engineering, University of Sydney in 1995–96 and a lecturer in the Dept of Computer Science and Electrical Engineering, University of Queensland in 1996–99. He moved to NTU in 1999. He was an Editorial Board Member of the Evolutionary Computation Journal, MIT Press (2013-2018) and an associate editor of the IEEE Trans on Cybernetics (2012 - 2018). He is an associate editor of an IEEE Trans on Evolutionary Computation (2005 -), Information Sciences (Elsevier) (2009 -), Pattern Recognition (Elsevier) (2001 -) and Int. J. of Swarm Intelligence Research (2009 -) Journals. He is a founding co-editor-in-chief of Swarm and Evolutionary Computation (2010 -), an SCI Indexed Elsevier Journal. His co-authored SaDE paper (published in April 2009) won the "IEEE Trans. on Evolutionary Computation outstanding paper award" in 2012. His former PhD student, Dr Jane Jing Liang, won the IEEE CIS Outstanding PhD dissertation award, in 2014. His research interests include swarm and evolutionary algorithms, pattern recognition, big data, deep learning and applications of swarm, evolutionary & machine learning algorithms. He was selected as one of the highly cited researchers by Thomson Reuters in 2015, 2016, 2017 and 2018 in computer science. He served as the General Chair of the IEEE SSCI 2013. He has been a member of the IEEE since 1991 and Fellow since 2015. He is an IEEE CIS distinguished lecturer (DLP) in 2018-2020. He was an elected AdCom member of the IEEE Computational Intelligence Society (CIS) in 2014-2016. Google Scholar: http://scholar.google.com.sg/citations?hl=en&user=yZNzBU0AAAAJ&view_op=list_works&pagesize=100



Johan Suykens

(KU Leuven) [introductory/intermediate]

Deep Learning, Neural Networks and Kernel Machines Summary

Neural networks & Deep learning and Support vector machines & Kernel methods have been among the most powerful and successful techniques in machine learning and data driven modelling. Initially, in artificial neural networks, the use of one hidden layer feedforward networks was common because of their universal approximation property. However, the existence of many local minima solutions in the training process was encountered as a drawback. Therefore, support vector machines and kernel methods became widely used, relying on solving convex optimization problems in classification and regression. In the meantime, computing power has increased and data have become abundantly available in many applications. As a result, currently one can afford training deep models consisting of (many) more layers and interconnection weights. Examples of successful deep learning models are convolutional neural networks, residual neural networks, stacked autoencoders, deep Boltzmann machines, deep generative models and generative adversarial networks. However, recent works related to understanding generalization, efficient training and adversarial networks, indicate that achieving complementary insights from kernel-based approaches and deep learning will become increasingly important.

In this course we will explain several synergies between neural networks, deep learning, least squares support vector machines and kernel methods. A key role at this point is played by primal and dual model representations and different duality principles. Recent developments on Restricted Kernel Machines will be highlighted, revealing new insights between neural networks, deep learning and kernel methods. In this way the bigger and unifying picture will be obtained and future perspectives will be outlined.

Syllabus

The material is organized into 3 parts:

- Part I Neural networks, support vector machines and kernel methods
- Part II Restricted Boltzmann machines, kernel machines and deep learning
- Part III Deep restricted kernel machines and future perspectives

In Part I a basic introduction is given to support vector machines (SVM) and kernel methods with emphasis on their artificial neural networks (ANN) interpretations. The latter can be understood in view of primal and dual model representations, expressed in terms of the feature map and the kernel function, respectively. Related to least squares support vector machines (LS-SVM) such characterizations exist for supervised and unsupervised learning, including classification, regression, kernel principal component analysis (KPCA), kernel spectral clustering (KSC), kernel canonical correlation analysis (KCCA), and other. Primal and dual representations are also relevant in order to obtain efficient training algorithms, tailored to the nature of the given application (high dimensional input spaces versus large data sizes). Application examples are given e.g. in black-box weather forecasting, pollution modelling, prediction of energy consumption, and community detection in networks.

In Part II we explain how to obtain a so-called restricted kernel machine (RKM) representation for least squares support vector machine related models. By using a principle of conjugate feature duality it is possible to obtain a similar representation as in restricted Boltzmann machines (RBM) (with visible and hidden units), which are used in deep belief networks (DBN) and deep Boltzmann machines (DBM). The principle is explained both for supervised and unsupervised learning. Related to kernel principal component analysis a generative model is obtained within the restricted kernel machine framework. In such a generative model the trained model is able to generate new data examples. The use of tensor-based models is also very natural within this new RKM framework.

In Part III deep restricted kernel machines (Deep RKM) are explained which consist of restricted kernel machines taken in a deep architecture. In these models a distinction is made between depth in a layer sense and depth in a level sense. Links and differences with stacked autoencoders and deep Boltzmann machines are given. The framework enables to conceive both deep feedforward neural networks (DNN) and deep kernel machines, through primal and dual model

representations. In this case one has multiple feature maps over the different levels in companion with multiple kernel functions. By fusing the objectives of the different levels (e.g. several KPCA levels followed by an LS-SVM classifier) in the deep architecture, the training process becomes faster and gives improved solutions. Different training algorithms and methods for large data sets will be discussed.

Finally, based on the newly obtained insights, future perspectives and challenges will be outlined.

References

- Belkin M., Ma S., Mandal S., To understand deep learning we need to understand kernel learning, Proceedings of Machine Learning Research, 80:541-549, 2018.
- Bengio Y., Learning deep architectures for AI, Boston: Now, 2009.
- Bietti A., Mialon G., Chen D., Mairal J., A Kernel Perspective for Regularizing Deep Neural Networks, arXiv:1810.00363.
- Binkowski M., Sutherland D.J., Arbel M., Gretton A., Demystifying MMD GANs, ICLR 2018.
- Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., Generative Adversarial Networks, pp. 2672-2680, NIPS 2014.
- Goodfellow I., Bengio Y., Courville A., Deep learning, Cambridge, MA: MIT Press, 2016.
- Hinton G.E., What kind of graphical model is the brain?, In Proc. 19th International Joint Conference on Artificial Intelligence, pp. 1765-1775, 2005.
- Hinton G.E., Osindero S., Teh Y.-W., A fast learning algorithm for deep belief nets, Neural Computation, 18, 1527-1554, 2006.
- Houthuys L., Suykens J.A.K., Tensor Learning in Multi-View Kernel PCA, in Proc. of the 27th International Conference on Artificial Neural Networks (ICANN), Rhodes, Greece, pp. 205-215, Oct. 2018.
- LeCun Y., Bengio Y., Hinton G., Deep learning, Nature, 521, 436-444, 2015.
- Mall R., Langone R., Suykens J.A.K., Multilevel Hierarchical Kernel Spectral Clustering for Real-Life Large Scale Complex Networks, PLOS ONE, e99966, 9(6), 1-18, 2014.
- Mehrkanoon S., Suykens J.A.K., Deep hybrid neural-kernel networks using random Fourier features, Neurocomputing, Vol. 298, pp. 46-54, July 2018.
- Mhaskar H., Liao Q., Poggio T., Learning Functions: When is Deep Better than Shallow, CBMM Memo No. 045, 2016.
- Montavon G., Muller K.-R., Cuturi M., Wasserstein Training of Restricted Boltzmann Machines, pp. 3718-3726, NIPS 2016.
- Salakhutdinov R., Hinton G.E., Deep Boltzmann machines, Proceedings of Machine Learning Research, 5:448-455, 2009.
- Salakhutdinov R., Learning deep generative models, Annu. Rev. Stat. Appl., 2, 361-385, 2015.
- Scholkopf B., Smola A., Muller K.-R., Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation, 10:1299-1319, 1998.

Scholkopf B., Smola A., Learning with kernels, Cambridge, MA: MIT Press,

Schreurs J., Suykens J.A.K., Generative Kernel PCA, ESANN 2018.

Suykens J.A.K., Vandewalle J., Training multilayer perceptron classifiers based on a modified support vector method, IEEE Transactions on Neural Networks, vol. 10, no. 4, pp. 907-911, Jul. 1999.

Suykens J.A.K., Vandewalle J., Least squares support vector machine classifiers, Neural Processing Letters, vol. 9, no. 3, pp. 293-300, Jun. 1999

Suykens J.A.K., Van Gestel T., De Brabanter J., De Moor B., Vandewalle J., Least squares support vector machines, Singapore: World Scientific,

Suykens J.A.K., Alzate C., Pelckmans K., Primal and dual model representations in kernel-based learning, Statistics Surveys, vol. 4, pp. 148-183, Aug. 2010.

Suykens J.A.K., Deep Restricted Kernel Machines using Conjugate Feature Duality, Neural Computation, vol. 29, no. 8, pp. 2123-2163, Aug. 2017.

Vapnik V., Statistical learning theory, New York: Wiley, 1998.

Zhang C., Bengio S., Hardt M., Recht B., Vinyals O., Understanding deep learning requires rethinking generalization, ICLR 2017.

Pre-requisites

Basics of linear algebra

Short Bio

Johan A.K. Suykens was born in Willebroek Belgium, May 18 1966. He received the master degree in Electro-Mechanical Engineering and the PhD degree in Applied Sciences from the Katholieke Universiteit Leuven, in 1989 and 1995, respectively. In 1996 he has been a Visiting Postdoctoral Researcher at the University of California, Berkeley. He has been a Postdoctoral Researcher with the Fund for Scientific Research FWO Flanders and is currently a full Professor with KU Leuven. He is author of the books "Artificial Neural Networks for Modelling and Control of Non-linear Systems" (Kluwer Academic Publishers) and "Least Squares Support Vector Machines" (World Scientific), co-author of the book "Cellular Neural Networks, Multi-Scroll Chaos and Synchronization" (World Scientific) and editor of the books "Nonlinear Modeling: Advanced Black-Box Techniques" (Kluwer Academic Publishers), "Advances in Learning Theory: Methods, Models and Applications" (IOS Press) and "Regularization, Optimization, Kernels, and Support Vector Machines" (Chapman & Hall/CRC). In 1998 he organized an International Workshop on Nonlinear Modelling with Time-series Prediction Competition. He has served as associate editor for the IEEE Transactions on Circuits and Systems (1997-1999 and 2004-2007), the IEEE Transactions on Neural Networks (1998-2009) and the IEEE Transactions on Neural Networks and Learning Systems (from 2017). He received an IEEE Signal Processing Society 1999 Best Paper Award and several Best Paper Awards at International Conferences. He is a recipient of the International Neural Networks Society INNS 2000 Young Investigator Award for significant contributions in the field of neural networks. He has served as a Director and Organizer of the NATO Advanced Study Institute on Learning Theory and Practice (Leuven 2002), as a program co-chair for the International Joint Conference on Neural Networks 2004 and the International Symposium on Nonlinear Theory and its Applications 2005, as an organizer of the International Symposium on Synchronization in Complex Networks 2007, a co-organizer of the NIPS 2010 workshop on

Tensors, Kernels and Machine Learning, and chair of ROKS 2013. He has been awarded an ERC Advanced Grant 2011 and 2017, and has been elevated IEEE Fellow 2015 for developing least squares support vector machines.

<https://www.esat.kuleuven.be/stadius/person.php?id=16>



Bertrand Thirion

(INRIA) [introductory]

Understanding the Brain with Machine Learning Summary

Neuroscience and artificial intelligence have long benefited from strong mutual interactions. Neuroscience has provided initial models of neural architectures to solve cognitive problems, such as vision, sound or language processing. Conversely, the spectacular development of artificial architectures in the last decade has brought candidates model to analyse the functional architecture of several brain systems. Importantly, leveraging neuroimaging data requires the analysis of large amounts of brain images or signals, making it a canonical example of larg-escalé structured signal analysis. In this course, we will review the interactions between neuroscience and AI, then discuss current challenges regarding the interactions of these fields.

Syllabus

- Introduction to brain imaging modalities
- Introduction to cognitive neurosciences
- Encoding models: AI models for neuroscience
- Encoding and model comparison
- inference from encoding models
- Ongoing work and open issues
- Brain activity decoding
- Application to vision
- Large-scale activity decoding
- Causal perspective

References

- Eickenberg M, Gramfort A, Varoquaux G, Thirion B. Seeing it all: Convolutional network layers map the function of the human visual system. *Neuroimage*. 2017 May 15;152:184-194.
- Artificial neural networks as models of neural information processing. M van Gerven, S Bohte Frontiers Media SA 2018.
- Natural speech reveals the semantic maps that tile human cerebral cortex AG Huth, WA de Heer, TL Griffiths, FE Theunissen, JL Gallant. *Nature* 532 (7600), 453
- Using goal-driven deep learning models to understand sensory cortex DLK Yamins, JJ DiCarlo. *Nature neuroscience* 19 (3), 356

Pre-requisites

general knowledge of Machine learning.

Short Bio

Bertrand Thirion is the leader of the Parietal team, part of INRIA research institute, Saclay, France, that addresses the development of statistics and machine learning techniques for brain imaging. He contributes both algorithms and software, with a special focus on functional neuroimaging

applications. He is involved in the Neurospin (CEA) neuroimaging center, one of the leading places on the use of high-field MRI for brain imaging. Bertrand Thirion is also leader of the DATAIA initiative that coordinates data science and AI research in the main French campus (Paris Saclay).



Gaël Varoquaux

(INRIA) [intermediate]

Representation Learning in Limited Data Settings Summary

The success of deep-learning hinges on intermediate representations: transformations of the data on which statistical learning is easier. Deep architectures can extract very rich and powerful representations, but it needs huge volumes of data. In this course, we will study the fundamentals of simple representations. Simple representations are interesting because they can be learned in limited data settings. We will also use them to provide didactic cases to understand how to build statistical models from data. The goal of the course is to provide the basic mathematical concepts that underly successful representation extracted in limited data settings.

Syllabus

- Shallow representations: what and why?
- Matrix factorizations and its variants:
 - From PCA to ICA
 - Sparse dictionary learning: formulation and efficient solvers
 - Word vectors demystified
- Fisher kernels: vector representations from a data model
 - Theory: from likelihood to representation
 - Encoding strings and text
 - Encoding covariances

References

- [1] Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5), 411-430.
- [2] Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11(Jan), 19-60.
- [3] Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In *Advances in neural information processing systems* (pp. 2177-2185).
- [4] Jaakkola, T., & Haussler, D. (1999). Exploiting generative models in discriminative classifiers. In *Advances in neural information processing systems* (pp. 487-493).

Pre-requisites

- General knowledge of statistical learning
- Basic knowledge of probability
- Basic knowledge of linear algebra

Short Bio

Gaël Varoquaux is a computer-science researcher at Inria. His research focuses on statistical learning tools for data science and scientific inference. He has pioneered the use of machine learning on brain images to map cognition and pathologies. More generally, he develops tools to make machine learning easier, with statistical models suited for real-life, uncurated data, and software for data science. He co-funded scikit-learn, one of the reference machine-learning toolboxes, and helped build various central tools for data analysis in Python. Varoquaux has contributed key methods for learning on spatial data, matrix factorizations, and modeling covariance matrices. He has a PhD in quantum physics and is a graduate from Ecole Normale Supérieure, Paris.



René Vidal

(Johns Hopkins University) [intermediate/advanced]

Mathematics of Deep Learning Summary

The past few years have seen a dramatic increase in the performance of recognition systems thanks to the introduction of deep networks for representation learning. However, the mathematical reasons for this success remain elusive. For example, a key issue is that the neural network training problem is nonconvex, hence optimization algorithms are not guaranteed to return a global minima. The first part of this tutorial will overview recent work on the theory of deep learning that aims to understand how to design the network architecture, how to regularize the network weights, and how to guarantee global optimality. The second part of this tutorial will present sufficient conditions to guarantee that local minima are globally optimal and that a local descent strategy can reach a global minima from any initialization. Such conditions apply to problems in matrix factorization, tensor factorization and deep learning. The third part of this tutorial will present an analysis of dropout for matrix factorization, and establish connections

Syllabus

- Introduction to Deep Learning Theory: Optimization, Regularization and Architecture Design
- Global Optimality in Matrix Factorization
- Global Optimality in Tensor Factorization and Deep Learning
- Dropout as a Low-Rank Regularizer for Matrix Factorization

References

- <https://arxiv.org/abs/1712.04741>
- <https://arxiv.org/abs/1708.07850>
- http://openaccess.thecvf.com/content_cvpr_2017/papers/Haeffele_Global_Optimality_in_CVPR_2017_paper.pdf
- <https://arxiv.org/abs/1506.07540>
- <https://arxiv.org/abs/1710.05092>

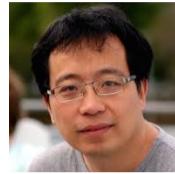
Pre-requisites

Basic understanding of sparse and low-rank representation and non-convex optimization.

Short Bio

Rene Vidal is a Professor of Biomedical Engineering and the Inaugural Director of the Mathematical Institute for Data Science at The Johns Hopkins University. His research focuses on the development of theory and algorithms for the analysis of complex high-dimensional datasets such as images, videos, time-series and biomedical data. Dr. Vidal has been Associate Editor of TPAMI and CVIU, Program Chair of ICCV and CVPR, co-author of the book 'Generalized Principal

Component Analysis' (2016), and co-author of more than 200 articles in machine learning, computer vision, biomedical image analysis, hybrid systems, robotics and signal processing. He is a fellow of the IEEE, IAPR and Sloan Foundation, a ONR Young Investigator, and has received numerous awards for his work, including the 2012 J.K. Aggarwal Prize for "outstanding contributions to generalized principal component analysis (GPCA) and subspace clustering in computer vision and pattern recognition" as well as best paper awards in machine learning, computer vision, controls, and medical robotics.



Haixun Wang

(WeWork) [intermediate]

Abstractions, Concepts, and Machine Learning Summary

Big data holds the potential to solve many challenging problems, and one of them is natural language understanding. As an example, big data has enabled the breakthrough in machine translation. However, natural language understanding still faces tremendous challenges. It has been shown that in areas such as question answering and conversation, domain knowledge is indispensable. Thus, how to acquire, represent, and apply domain knowledge for text understanding is of critical importance. In this short course, I will focus on understanding short text, which is crucial to many applications. Short texts do not always observe the syntax of a written language. As a result, traditional natural language processing methods cannot be easily applied. Second, short texts usually do not contain sufficient statistical signals to support many state-of-the-art approaches for text processing such as topic modeling. Third, short texts are usually more ambiguous. I will go over various techniques in knowledge acquisition, representation, and inferencing has been proposed for text understanding, and will describe massive structured and semi-structured data that have been made available in the recent decade that directly or indirectly encode human knowledge, turning the knowledge representation problems into a computational grand challenge with feasible solutions insight.

Syllabus

- Big data and statistical inference
- The rise and fall of the semantic network
- Knowledge of language
- Conceptual knowledge for text understanding
- Knowledge Extraction / Acquisition
- Knowledge Reasoning / Modeling
- Conclusion and Future work

References

- Kenneth Church, A Pendulum Swung Too Far, Linguistic Issues in Language Technology – LiLT Volume 2, Issue 4 May 2007
- Gregory Murphy, The Big Book of Concepts, MIT Press
- George Lakoff, Women, Fire and Dangerous Things: What Categories Reveal About the Mind, University of Chicago Press (1990)

Pre-requisites

Nothing.

Short Bio

Haixun Wang is an IEEE fellow and a VP of Engineering and Distinguished Scientist at WeWork, where he leads the Research and Applied Science division. He was Director of Natural Language Processing at Amazon. Before Amazon, he led the NLP Infra team in Facebook working on Query and Document Understanding. From 2013 to 2015, he was with Google Research, working on natural language processing. From 2009 to 2013, he led research in semantic search, graph data processing systems, and distributed query processing at Microsoft Research Asia. His knowledge base project Probbase has created significant impact in industry and academia. He had been a research staff member at IBM T. J. Watson Research Center from 2000 – 2009. He was Technical Assistant to Stuart Feldman (Vice President of Computer Science of IBM Research) from 2006 to 2007, and Technical Assistant to Mark Wegman (Head of Computer Science of IBM Research) from 2007 to 2009. He received the Ph.D. degree in Computer Science from the University of California, Los Angeles in 2000. He has published more than 150 research papers in referred international journals and conference proceedings. He served PC Chair of conferences such as CIKM'12, and he is on the editorial board of journals such as IEEE Transactions of Knowledge and Data Engineering (TKDE) and Journal of Computer Science and Technology (JCST). He won the best paper award in ICDE 2015, 10-year best paper award in ICDM 2013, and best paper award of ER 2009.



Xiaowei Xu

(University of Arkansas, Little Rock) [introductory/advanced]

Multi-resolution Models for Learning Multilevel Abstract Representations of Text Summary

Complex semantic meaning in natural language is hard to be mined using computational approach. Deep language models learning a hierarchical representation proved to be a powerful tool for natural language processing, text mining and information retrieval. This course will cover the models for word embedding and learning representations of text for information retrieval and text mining. The topic includes an introduction of language models for word embedding. It is followed by a presentation of recent multi-resolution models that represent documents at multiple resolutions in term of abstract levels. More specifically, we first form a mixture of weighted representations across the whole hierarchy of a given word embedding model, so that all resolutions of the hierarchical representation are preserved for the downstream model. In addition we combine all mixture representations from various models as an ensemble representation. Finally, the application for information retrieval and other text mining tasks is presented in the course.

Syllabus

Introduction (1.5 hours)

1.1. Vector space model

1.2. Word2vec

1.3. GloVe

1.4. FastText

1.5. ELMo

Multi-resolution models (1.5 hours)

2.1. Multi-resolution word embedding

2.2. Ensemble models

Applications (1.5 hours)

References

- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems 26, pp. 3111?3119. 2013.
- Pennington, J., Socher, R., and Manning, C. D. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pp. 1532?1543, 2014.
- Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., and Joulin, A. Advances in pre-training distributed word representations. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), 2018.
- Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. Deep contextualized word representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 2227?2237, 2018.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- Cakaloglu, T., and Xu, X. A Multi-Resolution Word Embedding for Document Retrieval from Large Unstructured Knowledge Bases. arXiv preprint arXiv:1902.00663

Pre-requisites

Basic knowledge of linear algebra and machine learning.

Short-bio

Xiaowei Xu, a professor of Information Science at the University of Arkansas, Little Rock (UALR), received his Ph.D. degree in Computer Science at the University of Munich in 1998. Before his appointment in UALR, he was a senior research scientist in Siemens, Munich, Germany. His research spans data mining, machine learning, bioinformatics, database management systems and high-performance computing. Dr. Xu is a recipient of 2014 ACM SIGKDD Test of Time award for his contribution to the density-based clustering algorithm DBSCAN.



Ming-Hsuan Yang

(University of California, Merced) [intermediate/advanced]

Learning to Track Objects Summary

The goal is to introduce the recent advances in object tracking based on deep learning and related approaches. Performance evaluation and challenging factors in this field will be discussed.

Syllabus

- Brief history of visual tracking
- Generative approach
- Discriminative approach

- Deep learning methods
- Performance evaluation
- Challenges and future research directions

References

Y. Wu, J. Lim, and M.-H. Yang, Object Tracking Benchmark, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.

H. Nam and B. Han, Learning Multi-domain Convolutional Neural Networks for Visual Tracking, CVPR, 2016.

M. Danelljan, G. Bhat, F. Khan, M. Felsberg, ECO: Efficient Convolution Operators for Tracking, CVPR, 2017.

Pre-requisites

Basic knowledge in computer vision and intermediate knowledge in deep learning

Short Bio

Ming-Hsuan Yang is a Professor of Electrical Engineering and Computer Science at University of California, Merced, and a Research Scientist at Google Cloud. He serves as a program co-chair of IEEE International Conference on Computer Vision (ICCV) in 2019, program co-chair of Asian Conference on Computer Vision (ACCV) in 2014, and general co-chair of ACCV 2016. He has served as an associate editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) from 2007 to 2011, and currently serves as an associate editor of the International Journal of Computer Vision (IJCV), Computer Vision and Image Understanding (CVIU), Image and Vision Computing (IVC) and Journal of Artificial Intelligence (JAIR). Yang received the Google Faculty Award in 2009 and the Faculty Early Career Development (CAREER) award from the National Science Foundation in 2012. In 2015. He received paper awards from UIST 2017, CVPR 2018 and ACCV 2018. He is an IEEE Fellow.



Zhongfei Zhang

(Binghamton University) [introductory/advanced]

Knowledge Discovery from Complex Data with Deep Learning Summary

This course aims at exposing the audience an introduction to the fundamental theories and advanced methods on knowledge discovery from complex data with deep learning. It is a well-accepted fact that today we are drown with data. Further typically the data we face are complex data. Complex data refer to the most comprehensive data formats we encounter. By complex data, it is meant that the data may be non-structural, media data such as text, imagery, video, audio, and graphics/animation, and/or that the data may be in the other modalities including time-series data, sequential data, and relational data that violate the i.i.d. assumption such as the social network data, e-commerce data, financial interaction/transaction data, and cyber communication/attack data where the data can be represented as a multi-type node graph involving multiple players, and further that the data may be noisy, meaning that not only the data per se can be noisy, but also the given training labels, if there are any, can also be noisy, such as in the scenario of image annotation or classification where the given training labels can be imperfect (incorrect and/or incomplete).

Consequently, complex data represent the most commonly encountered data in our daily life and also in almost all the real-world applications and thus, it is also extremely challenging to develop theories on complex data learning.

The course begins with an extensive introduction to the fundamental concepts and theories of knowledge discovery from complex data, as well as the relevant deep learning theories required for knowledge discovery from complex data; then the course showcases several important applications as case studies in the real-world as examples for knowledge discovery from complex data with deep learning.

Syllabus

The course consists of three an hour and half sessions. The syllabus is as follows:

First session: Introduction to the fundamental concepts and theories on knowledge discovery from complex data.

Second session: Introduction to the related deep learning theories to knowledge discovery from complex data.

Third session: Introduction to and discussions on several case studies as examples of developing advanced methods on knowledge discovery from complex data with deep learning.

References:

- Bo Long, Zhongfei (Mark) Zhang, and Philip S. Yu, Relational Data Clustering: Models, Algorithms, and Applications, Taylor & Francis/CRC Press, 2010, ISBN: 9781420072617
- Zhongfei (Mark) Zhang and Ruofei Zhang, Multimedia Data Mining -- A Systematic Introduction to Concepts and Theory, Taylor & Francis Group/CRC Press, 2008, ISBN: 9781584889663
- Zhongfei (Mark) Zhang, Bo Long, Zhen Guo, Tianbing Xu, and Philip S. Yu, Machine Learning Approaches to Link-Based Clustering, in Link Mining: Models, Algorithms and Applications, Edited by Philip S. Yu, Christos Faloutsos, and Jiawei Han, Springer, 2010
- Zhen Guo, Zhongfei Zhang, Eric P. Xing, and Christos Faloutsos, Multimodal Data Mining in a Multimedia Database Based on Structured Max Margin Learning, ACM Transactions on Knowledge Discovery and Data Mining, ACM Press, 2015
- Shuangfei Zhai, Yu Cheng, and Zhongfei (Mark) Zhang, Doubly Convolutional Neural Networks, Proc. Advances in Neural Information Processing Systems, (NIPS 2016), Barcelona, Spain, December, 2016
- Shuangfei Zhai, Hui Wu, Abhishek Kumar, Yu Cheng, Yongxi Lu, Zhongfei Zhang, Rogerio Feris, S3Pool: Pooling with Stochastic Spatial Sampling, Proc. IEEE International Conference on Computer Vision and Pattern Recognition, (CVPR 2017), Honolulu, Hawaii, USA, July, 2017

Pre-requisites

College math, fundamentals about computer science

Short Bio

Zhongfei (Mark) Zhang is a full professor of Computer Science at State University of New York (SUNY) at Binghamton, and directs the Multimedia Research Computing Laboratory in the University. He has also served as a QiuShi Chair Professor at Zhejiang University, China, and as the Director of the Data Science and Engineering Research Center at the university, and as a CNRS Chair Professor at the University of Lille 1, France, while he was on leave from State University of New York (SUNY) at Binghamton, USA. He has received a B.S. in Electronics Engineering (with Honors), an M.S. in Information Sciences, both from Zhejiang University, China, and a PhD in Computer Science from the University of Massachusetts at Amherst, USA. His research interests include machine learning and artificial intelligence, data mining and knowledge discovery, multimedia information indexing and retrieval, computer vision, and pattern recognition. He is the

author and co-author of the first monograph on multimedia data mining and the first monograph on relational data clustering, respectively. His research is sponsored by a wide spectrum of government funding agencies, industrial labs, as well as private agencies. He has published over 200 papers in premier venues in his areas and is an inventor for more than 30 patents. He has served in several journal editorial boards and received several professional awards including best paper awards in the premier conferences in his areas.

Active links:

[BigDat 2020 - 6th International Winter School on Big Data](#)

[TPNC 2019 - 8th International Conference on the Theory and Practice of Natural Computing](#)

[SLSP 2019 - 7th International Conference on Statistical Language and Speech Processing](#)

[AlCoB 2019 - 6th International Conference on Algorithms for Computational Biology](#)

[LATA 2019 - 13th International Conference on Language and Automata Theory and Applications](#)

© IRDTA 2019. All Rights Reserved.



Past links:

[DeepLearn 2018](#)

[DeepLearn 2017](#)