[1] MixHop: Higher-Order Graph Convolutional Architectures via Sparsified Neighborhood Mixing

[2] Multi-objective training of Generative Adversarial Networks with multiple discriminators

[3] Graph Element Networks: adaptive, structured computation and memory

[4] Analogies Explained: Towards Understanding Word Embeddings

[5] Fine-Grained Analysis of Optimization and Generalization for Overparameterized Two-Layer Neural Networks

[6] Entropic GANs meet VAEs: A Statistical Approach to Compute Sample Likelihoods in GANs

[7] Provable Guarantees for Gradient-Based Meta-Learning

[8] Concrete Autoencoders: Differentiable Feature Selection and Reconstruction

[9] Learning to Route in Similarity Graphs

[10] Active Learning for Probabilistic Structured Prediction of Cuts and Matchings

[11] Invertible Residual Networks

[12] Overcoming Multi-model Forgetting

[13] Optimal Kronecker-Sum Approximation of Real Time Recurrent Learning

[14] A Kernel Perspective for Regularizing Deep Neural Networks

[15] Adversarial Attacks on Node Embeddings via Graph Poisoning

[16] Compositional Fairness Constraints for Graph Embeddings

[17] Understanding the Origins of Bias in Word Embeddings

[18] Why do Larger Models Generalize Better? A Theoretical Perspective via the XOR Problem

[19] What is the Effect of Importance Weighting in Deep Learning?

[20] Learning Action Representations for Reinforcement Learning

[21] Dynamic Measurement Scheduling for Event Forecasting using Deep RL

[22] Neural Network Attributions: A Causal Perspective

[23] Information-Theoretic Considerations in Batch Reinforcement Learning

[24] Generative Adversarial User Model for Reinforcement Learning Based Recommendation System

[25] Understanding and Utilizing Deep Neural Networks Trained with Noisy Labels

[26] A Gradual, Semi-Discrete Approach to Generative Network Training via Explicit Wasserstein Minimization

[27] Transferability vs. Discriminability: Batch Spectral Penalization for Adversarial Domain Adaptation

[28] Fast Incremental von Neumann Graph Entropy Computation: Theory, Algorithm, and Applications

[29] Multivariate-Information Adversarial Ensemble for Scalable Joint Distribution Matching

[30] Robust Decision Trees Against Adversarial Examples

[31] RaFM: Rank-Aware Factorization Machines

[32] Control Regularization for Reduced Variance Reinforcement Learning

[33] Random Walks on Hypergraphs with Edge-Dependent Vertex Weights

[34] Probability Functional Descent: A Unifying Perspective on GANs, Variational Inference, and Reinforcement Learning

[35] MeanSum: A Neural Model for Unsupervised Multi-Document Abstractive

Summarization

[36] Sensitivity Analysis of Linear Structural Causal Models

[37] Quantifying Generalization in Reinforcement Learning

[38] Empirical Analysis of Beam Search Performance Degradation in Neural Sequence Models

[39] Certified Adversarial Robustness via Randomized Smoothing

[40] Gauge Equivariant Convolutional Networks and the Icosahedral CNN

[41] CURIOUS: Intrinsically Motivated Modular Multi-Goal Reinforcement Learning

[42] A fully differentiable beam search decoder

[43] Monge blunts Bayes: Hardness Results for Adversarial Training

[44] Flexibly Fair Representation Learning by Disentanglement

[45] Teaching a black-box learner

[46] Stochastic Deep Networks

[47] Learning to Convolve: A Generalized Weight-Tying Approach

[48] Generalized No Free Lunch Theorem for Adversarial Robustness

[49] Wasserstein of Wasserstein Loss for Learning Generative Models

[50] GDPP: Learning Diverse Generations using Determinantal Point Processes

[51] On the Connection Between Adversarial Robustness and Saliency Map Interpretability

[52] Dead-ends and Secure Exploration in Reinforcement Learning

[53] DL2: Training and Querying Neural Networks with Logic

[54] On discriminative learning of prediction uncertainty

[55] Learning Discrete Structures for Graph Neural Networks

[56] Transfer Learning for Related Reinforcement Learning Tasks via Image-to-Image Translation

[57] Graph U-Nets

[58] Deep Generative Learning via Variational Gradient Flow

[59] Geometric Scattering for Graph Data Analysis

[60] Multi-Frequency Phase Synchronization

[61] SelectiveNet: A Deep Neural Network with an Integrated Reject Option

[62] Recursive Sketches for Modular Deep Learning

[63] Adversarial Examples Are a Natural Consequence of Test Error in Noise

[64] Estimating Information Flow in Deep Neural Networks

[65] Efficient Training of BERT by Progressively Stacking

[66] A Statistical Investigation of Long Memory in Language and Music

[67] Graphite: Iterative Generative Modeling of Graphs

[68] Fast Algorithm for Generalized Multinomial Models with Ranking Data

[69] Towards a Deep and Unified Understanding of Deep Neural Models in NLP

[70] Humor in Word Embeddings: Cockamamie Gobbledegook for Nincompoops

[71] Exploring interpretable LSTM neural networks over multi-variable data

[72] Learning to Exploit Long-term Relational Dependencies in Knowledge Graphs

[73] IMEXnet A Forward Stable Deep Neural Network

[74] Understanding and Controlling Memory in Recurrent Neural Networks

[75] On the Impact of the Activation function on Deep Neural Networks Training

[76] Using Pre-Training Can Improve Model Robustness and Uncertainty

[77] Parameter-Efficient Transfer Learning for NLP

[78] Finding Mixed Nash Equilibria of Generative Adversarial Networks

[79] Hierarchical Importance Weighted Autoencoders

[80] HexaGAN: Generative Adversarial Nets for Real World Classification

[81] Learning What and Where to Transfer

[82] Graph Neural Network for Music Score Data and Modeling Expressive Piano Performance

[83] Ladder Capsule Network

[84] Learning Discrete and Continuous Factors of Data via Alternating Disentanglement

[85] Neural Logic Reinforcement Learning

[86] Kernel Mean Matching for Content Addressability of GANs

[87] Trainable Decoding of Sets of Sequences for Neural Sequence Models

[88] POPQORN: Quantifying Robustness of Recurrent Neural Networks

[89] Stochastic Beams and Where To Find Them: The Gumbel-Top-k Trick for Sampling Sequences Without Replacement

[90] Similarity of Neural Network Representations Revisited

[91] Characterizing Well-Behaved vs. Pathological Deep Neural Networks

[92] State-Reification Networks: Improving Generalization by Modeling the Distribution of Hidden Representations

[93] Self-Attention Graph Pooling

[94] Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks

[95] Are Generative Classifiers More Robust to Adversarial Attacks?

[96] LGM-Net: Learning to Generate Matching Networks for Few-Shot Learning

[97] Graph Matching Networks for Learning the Similarity of Graph Structured Objects

[98] Learn to Grow: A Continual Structure Learning Framework for Overcoming Catastrophic Forgetting

[99] Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations

[100]   Disentangled Graph Convolutional Networks

[101]   Traditional and Heavy Tailed Self Regularization in Neural Network Models

[102]   Disentangling Disentanglement in Variational Autoencoders

[103]   MIWAE: Deep Generative Modelling and Imputation of Incomplete Data Sets

[104]   Imputing Missing Events in Continuous-Time Event Streams

[105]   The Wasserstein Transform

[106]   Relational Pooling for Graph Representations

[107]   Dropout as a Structured Shrinkage Prior

[108]   Zero-Shot Knowledge Distillation in Deep Networks

[109]   Improving Adversarial Robustness via Promoting Ensemble Diversity

[110]   Deep Residual Output Layers for Neural Language Generation

[111]   Domain Agnostic Learning with Disentangled Representations

[112]   Towards Understanding Knowledge Distillation

[113]   GMNN: Graph Markov Neural Networks

[114]   HyperGAN: A Generative Model for Diverse, Performant Neural Networks

[115]   Mixture Models for Diverse Machine Translation: Tricks of the Trade

[116]   Revisiting precision recall definition for generative modeling

[117]   The Evolved Transformer

[118]   MASS: Masked Sequence to Sequence Pre-training for Language Generation

[119]   Insertion Transformer: Flexible Sequence Generation via Insertion Operations

[120]   BERT and PALs: Projected Attention Layers for Efficient Adaptation in Multi-Task Learning

[121]   Contextual Memory Trees

[122]   Robustly Disentangled Causal Mechanisms: Validating Deep Representations for Interventional Robustness

[123]   Hyperbolic Disk Embeddings for Directed Acyclic Graphs

[124]   Equivariant Transformer Networks

[125]   Making Deep Q-learning methods robust to time discretization

[126]   The Natural Language of Actions

[127]   Optimal Transport for structured data with application on graphs

[128]   Probabilistic Neural Symbolic Models for Interpretable Visual Question Answering

[129]   SATNet: Bridging deep learning and logical reasoning using a differentiable satisfiability solver

[130]   Improving Neural Language Modeling via Adversarial Training

[131]   State-Regularized Recurrent Neural Networks

[132]   Bias Also Matters: Bias Attribution for Deep Neural Network Explanation

[133]   Jumpout : Improved Dropout for Deep Neural Networks with ReLUs

[134]   Non-Monotonic Sequential Text Generation

[135]   Simplifying Graph Convolutional Networks

[136]   Gromov-Wasserstein Learning for Graph Matching and Node Embedding

[137]   Position-aware Graph Neural Networks

[138]   Learning Neurosymbolic Generative Models via Program Synthesis

[139]   DAG-GNN: DAG Structure Learning with Graph Neural Networks

[140]   Self-Attention Generative Adversarial Networks