Institute for AI and Beyond Research Report 2020-23

Automatic Learning of High Accurate Prediction Models from Limited Supervised Data

Tatsuya Harada (Research Center for Advanced Science and Technology)

Masashi Sugiyama (Graduate School of Frontier Sciences)

Introduction

The goal of this project is to develop methods for learning prediction models from limited supervised information and for automatic construction of such models. To achieve this goal, we have developed 1) theories and algorithms for learning from limited supervised information, 2) theories and algorithms for knowledge transfer, and 3) a framework for automatic construction of highly accurate prediction models. In addition, we have worked on 4) spatio-temporal modeling with the goal of constructing world models based on the real environment. Sugiyama group is mainly responsible for developing the theory and algorithms for learning from limited supervised information, while Harada group is mainly responsible for the other topics. Details of the results are given below.

1. Theories and Algorithms for Learning from Limited Supervised Information

1-1. Weakly Supervised Learning

One of the biggest bottlenecks of current machine learning is the requirement of a large number of high-quality labeled data. However, such abundant labeled data is often not available in applications such as medicine, robot, and disaster. To cope with problem, various different approaches were investigated so far, e.g., reducing label annotation costs by crowdsourcing, using a simulator to generate pseudo labels, and engineering models and incorporating domain knowledge to improve the sample efficiency. In this subproject, we have been exploring another approach called weakly supervised learning: Instead of fully labeled high-quality data, we use a “weakly” labeled data that is insufficient but can be collected easily.

A typical setup of weakly supervised learning is positive-unlabeled learning, which tries to train a binary classifier only from positive and unlabeled data, without negative data. Our group developed a method of positive-unlabeled learning in the empirical risk minimization framework, which allows unbiased estimation of the classification risk. We also developed a non-negative risk estimator correction method that allows us to mitigate overfitting particularly for complex models such as deep networks. In this subproject, we extended these methodologies to various weakly supervised learning scenarios and develop a unified framework of empirical risk minimization for weakly supervised learning.
So far, we showed that empirical risk minimization based weakly supervised learning can be extended to partial labels (multi-class labels that contain multiple candidate classes), pairwise comparison labels (a pair of samples in which one has a larger positive-class probability than the other), pairwise similarity confidence labels (the probability that a pair of samples are from the same class), and multiple sets of unlabeled data having different class priors. We summarized such a unified framework in a monograph and published it from the MIT Press.

Another important achievement is an application of weakly supervised learning, and we tackled an audio signal enhancement problem. Existing approaches require a so-called parallel training data that correspond to noisy and clean versions of the same signal. However, such parallel training data can be rarely obtained in the real-world; instead, synthesized noise is used to generate parallel training data. On the other hand, we can easily collect noisy signals (unlabeled) and noise-only signals (labeled) in real-world settings. Then a positive-unlabeled learning method allows us to train a classifier to separate positives (noise) and negatives (signals).

1-2. Noisy-Label Learning

In the case of regression, learning from noisy output is rather straightforward and the use of a large number of noisy data suffices since output noise is typically independent and additive and they cancel each other automatically. On the other hand, in the case of classification, learning from noisy output still requires challenges since label-flipping noise cannot be automatically cancelled even if they are independent. Thus, we need a specific mechanism to combat label noise.

A systematic approach is to correct the loss function based on the noise transition matrix, which describe the label flipping probability. However, the noise transition matrix is often unknown in practice and is not identifiable only from noisy training data. To cope with this problem, we explored noise transition estimation and developed a method that allows us to estimate the noise transition matrix under a mild assumption. Note that almost all existing methods require clean label, which is often unavailable in practice; while our method can still estimate the noise transition matrix as long as noisy training data at hand satisfy a certain “sufficiently scattered” condition.

Furthermore, we investigated a more challenging scenario of instance-dependent noise, where the noise property depends on the instance and is thus extremely challenging. For

this hard problem, we developed heuristic techniques to estimate the noise transition matrix function and demonstrated their practical usefulness in experiments.

Another topic that has been investigated in this subproject is coping with adversarial noise. It was shown that a machine learning solution is vulnerable to a tiny change in instances---a well-known example is image classification where a panda image corrupted by invisibly small noise is misclassified as a gibbon. Since this can be a critical security problem for real-world machine learning systems, it is an urgent challenge to cope with such adversarial noise. We explored several different scenarios of adversarial noise, proposed robust training methods, and demonstrated their usefulness in experiments.

1-3. Learning with Biased Data

Collecting high-quality labels is often expensive, so we want to reuse data that were collected previously for other tasks. However, since tasks are different, such previously collected data are biased to the current task. Or more recently, privacy issues are major concerns, and we are not allowed to collect data as we wish; then our training data is often biased to the test distribution. Overcoming such a data bias problem is an essential challenge and we have been taking a transfer learning approach based on importance weighting---we weigh training samples according to their “importance” in the test distribution. Such an importance weighting approach has been demonstrated to be quite general and effective.

In this subproject, we considered continuous distribution shift, where data distributions change over time and we want to update our model to follow the changing distributions. Key technical challenges are how to efficiently update our model over time and how to guarantee the prediction performance.

We first considered continuous label shift, where only the class-prior probabilities change over time and only a set of unlabeled data is available in each adaptation phase. For this problem, we developed an online gradient method for which we gave an unbiased estimator of the gradient of the classification risk in each adaptation phase. In this online gradient method, the choice of the step size is critical. The optimal step size can be computed if we know the true distribution shift, but such knowledge is usually not available. To cope with this problem, we proposed using ensemble learning over learners with different step sizes. We theoretically showed that the proposed method achieves the same dynamic regret as the method that uses the optimal step size with the knowledge of distribution shifts. Furthermore, we also provided a method to incorporate prior knowledge on the distribution shift to improve the performance.

We then extended the above approach to continuous covariate shift, where only the input distribution changes over time. The methodology itself is similar to the label shift case, and we use online gradient descent and ensemble learning. However, their ideas and usages are very different. We perform online gradient descent for importance estimation since the dynamic regret of the prediction risk can be upper-bounded by the importance estimation error. Then an ensemble is taken over learners with different intervals of the time horizon.

2. Theories and Algorithms for Knowledge Transfer

When a model trained in one domain is applied to another domain, the expected prediction accuracy may not be achieved due to differences in domain characteristics (domain shift). Unsupervised domain adaptation (UDA) is an effective approach to solve the domain shift problem by using information from both the source domain where the model was trained and the target domain where the model is applied.

Many recent domain adaptation methods have been based on the theory of Ben-David et al. that simultaneously minimizes the error in the source domain and the distance between marginal distributions, but the joint error is typically ignored due to its difficulty in estimation. To solve this problem, we proposed a new objective function associated with an upper bound on the joint error, and to tighten this bound, we used a hypothesis space induced by the source/pseudo-target labels that can reduce the search space. To measure the dissimilarity between hypotheses, we proposed a new cross-margin discrepancy measure that mitigates instability during adversarial training.

As mentioned above, domain adaptation sometimes learns domain-invariant features in an adversarial fashion to match the source and target marginal distributions. In this case, the domain and label classifier networks are often trained separately, and the two classifier networks rarely interact. In this study, we proposed a new weighting method for the feature space induced by error backpropagation. This approach has the advantage that domain classifiers can focus on features that are important for classification, and that classifiers and adversarial classifiers can be more closely coupled.

Most existing UDA methods have assumed that raw source data is directly available during training in the target domain when transferring knowledge from the source to the target domain. In recent years, with growing concerns about data privacy, direct availability of source data is not always possible when applying UDA methods in a new target domain. To solve this problem, we first analyzed the cross-domain feature representation in source-free unsupervised domain adaptation (SF-UDA), which does not use source data when training the target domain. Then, we derived a new theory for obtaining an upper bound on the prediction error of the target domain by using the learned source model instead of the source data. Based on this theory, we introduced an information bottleneck theory to realize SF-UDA by minimizing the generalized upper bound of the prediction error in the target domain.

Since SF-UDA does not have labeled source data, it is often difficult for SF-UDA methods to provide reliable class representations for target data. To solve this problem, we proposed the idea of Confidence-based Subsets Feature Alignment (CSFA), a feature alignment method that focuses on reliable distributions. CSFA divides the target data into two subsets: a confidence subset whose class predictions from the source model have low entropy, and a non-confidence subset whose class predictions do not. By using pseudo-labels from the confidence subset, the original SF-UDA problem is viewed as a Universal Domain

Adaptation (UniDA) problem. This provides a reliable class representation for the target data by aligning the feature distributions of the two subsets.

3. Framework for Automatic Construction of Highly Accurate Prediction Models

Various methods have been proposed for the automatic construction of models, but all of them are general-purpose and require a large amount of computational cost when the search space is expanded, so a dramatic increase in model efficiency is required. Therefore, we focused on hyperbolic space neural networks, which can adequately model hierarchically structured data even in low dimensions. Shifting the stage of neural network operations to a non-Euclidean geometry such as hyperbolic space is a promising method to find a geometrical structure more suitable for data representation and processing. Since volume increases exponentially with radius, hyperbolic spaces have the ability to continuously embed tree structures with arbitrarily low distortion. Compared to Euclidean spaces, hyperbolic spaces exhibit higher embedding accuracy with fewer dimensions in such cases. Therefore, this study generalized the basic components of neural networks with a single hyperbolic geometric model, the Poincaré ball model. This new methodology constructed multinomial logistic regression, fully connected layers, convolutional layers, and attention mechanisms under a unified mathematical interpretation without increasing parameters.

To dramatically increase the speed and reduce the power consumption of the model to cope with the enormous search cost, we also focused on generative models based on spiking neural networks. Spiking neural networks (SNNs) are binary and event-driven, and can run on neuromorphic devices with ultra-high speed and ultra-low power consumption. Therefore, SNNs are expected to have various applications, such as generating high-quality images as a generative model running on edge devices. In this study, we constructed a variational autoencoder (VAE) based on SNNs to enable image generation, which is known for its high stability among generative models, and its quality has been improving in recent years. In a simple VAE, the latent space is represented by a normal distribution and floating-point operations are required for sampling, but this is not possible in SNNs because all features must be binary time-series data. Therefore, in this study, an autoregressive SNN model is used to construct the latent space, and the latent variables are sampled from the output. This allows the latent variables to follow a Bernoulli process and enables variational learning.

4. Spatio-Temporal Modeling of the Real World

Image data as it is would have several million dimensions, and it would be extremely difficult to collect enough training images to satisfy this high dimensionality. However, the real world has a low-dimensional structure of 3 or 4 (if time is included), which can lead to a significant reduction in annotation cost by making good use of this low-dimensional structure as prior knowledge. Therefore, with the goal of reducing the number of data required by exploiting this real-world low-dimensional structure, we have worked on partial point cloud matching and registration, and learning controllable 3D models from sparse observational images.

First, we focused on the problem of partial point cloud matching in rigid and deformed scenes and proposed Lepard, a new learning-based approach to this problem. The features of this method are: 1) an architecture that separates the point cloud representation into a feature space and a 3D position space; 2) a position encoding method that explicitly expresses 3D relative distance information by the inner product of vectors; and 3) a repositioning technique that changes the relative position between point clouds. For deformable objects, the proposed Lepard achieved higher non-rigid feature matching recall than previous studies on the newly constructed 4DMatch / 4DLoMatch benchmark.

Non-rigid point cloud registration is also an important component of 4D reconstruction. This task is challenging due to the high complexity of the unknown non-rigid motions. In this study, we solved this problem by hierarchical motion decomposition. The proposed method, called Neural Deformation Pyramid (NDP), uses a pyramid structure to represent non-rigid motions. Each level of the pyramid, represented by an MLP, takes a sinusoidally encoded 3D point as input and outputs the increment of the motion from the previous level. The sinusoidal function starts at a low input frequency and gradually increases as the pyramid level decreases. This allows multi-step decomposition of rigid-body to non-rigid-body motion and speeds up model training.

When reconstructing the shape of an object from multiple object primitives, it is desirable to treat these primitives as if they were a single shape, and to be able to immediately access basic shape properties such as volume and surface of the combined object as a whole. This is made possible by a primitive representation that unifies implicit and explicit representations. Therefore, we proposed a new implicit and explicit shape primitive representation of 3D shapes, called Neural Star Domain (NSD), which learns the primitive shape of star domains.

Artificial articulated objects are ubiquitous in the real world. However, conventional unsupervised part decomposition methods have not been suitable for such objects because the part positions are spatially fixed, resulting in inconsistent part analysis. Therefore, in this study, we proposed an unsupervised pose-aware part decomposition (PPD) method that explicitly targets artificial articulated objects with mechanical joints.

Recent advances in 3D implicit function representation have made it possible to learn models of complex objects. However, current methods have required supervised information on the 3D geometry, making it difficult to learn representations of articulated objects whose deformation can be controlled. To solve this problem, this study proposed Neural Articulated Radiance Field (NARF), a method that considers only the rigid transformations of the object parts most relevant to the point of interest. By doing so, the proposed method is able to represent object deformations that depend on the pose of each joint without a significant increase in computational complexity.

Although the above method can acquire a 3D model of a controllable human, it assumes that the structure is known and targets only the same structure as the human, and unfortunately cannot handle new object categories. Therefore, we proposed a novel method to learn both the appearance and the structure of an articulated object of unknown category by observing the motion of the articulated object from multiple viewpoints without supervised information

on joint annotation or structure. The method is based on a combination of explicit and implicit 3D representations.

Photorealistic images of articulated objects can be explicitly pose-controlled using existing 3D neural networks, but these methods have required annotated 3D pose and foreground masks for learning, which are difficult to obtain. In this study, we bypassed the need for these by using an adversarial generative network (GAN) and developed a novel method, called ENARF-GAN, for learning representations that take into account the 3D shape of articulated objects.

We have controlled the radiance field by estimating the skeletal structure of the object and deforming the skeletal structure, but in this study we proposed a new method of deforming the radiance field, namely free-form deformation of the radiance field. This method uses a triangular mesh surrounding a foreground object, called a cage, as an interface, and allows free-form deformation of the radiance field by manipulating the vertices of the cage. The core of the approach is the cage-based deformation commonly used in mesh deformation. In this approach, we extended this deformation to radiance fields and proposed a new formulation that allows rendering the deformed scene by mapping the positions and viewing directions of the sampling points from the deformation space to the canonical space.

Future Challenges

In the first subproject, we successfully developed useful algorithms. Nevertheless, the biggest challenge in the current machine learning community is an appropriate choice of algorithms for the target data analysis task at hand. Each method requires some assumptions to work well, but it is often difficult to know the properties of data in advance. Therefore, our important next challenge is to develop a universal method that works reasonably and broadly well without imposing strong assumptions. We believe such a universal method should be the first method for many practitioners to use and see whether machine learning is potentially useful and worth investing their resources more.

In the second subproject, Harada's and Sugiyama's groups will work together to extend the weakly supervised domain adaptation algorithm. We will integrate our algorithms to build a theory of domain adaptation that can be used universally in a variety of situations. In spatio-temporal modeling, we will continue our research to develop scalable methods that can handle a wide range of data and long time periods. We will also develop multimodal models that are grounded in the real world by integrating 4D reconstruction methods, semantic understanding methods, language models, and so on. Furthermore, we will build a distributed computational framework that can train models with high predictive ability even when computational resources are not abundant or data cannot be collected at a central location. On top of this framework, we will implement and validate the machine learning methods developed by our team.

References

1. Zhang, J., Zhu, J., Niu, G., Han, B., Sugiyama, M., & Kankanhalli, M. Geometry-aware instance-reweighted adversarial training. International Conference on Learning Representations (ICLR2021), 29 pages, 2021. (Oral presentation: top 1.8% =53/2997)
2. Li, X., Liu, T., Han, B., Niu, G., & Sugiyama, M. Provably end-to-end label-noise learning without anchor points. International Conference on Machine Learning (ICML2021), pp. 6403-6413, 2021.
3. Bai, Y., Zhang, Y.-J., Zhao, P., Sugiyama, M., & Zhou, Z.-H. Adapting to online label shift with provable guarantees. Neural Information Processing Systems Conference (NeurIPS2022), pp. 29960-29974, 2022.
4. Sugiyama, M., Bao, H., Ishida, T., Lu, N., Sakai, T., & Niu, G. Machine Learning from Weak Supervision: An Empirical Risk Minimization Approach, 320 pages, MIT Press, 2022.
5. Ito, N. & Sugiyama, M. Audio signal enhancement with learning from positive and unlabeled data. International Conference on Acoustics, Speech, and Signal Processing (ICASSP2023), pp. 1-5, 2023. (Best paper award, top 0.02% =1/6127)
6. Zhang, Y.-J., Zhang, Z.-Y., Zhao, P., & Sugyiama, M. Adapting to continuous covariate shift via online density ratio estimation. Neural Information Processing Systems Conference (NeurIPS2023), 2023. (to appear)
7. Dexuan Zhang, Thomas Westfechtel, Tatsuya Harada. Unsupervised Domain Adaptation via Minimized Joint Error. Transactions on Machine Learning Research (TMLR), 2023. https://openreview.net/forum?id=kiPsMct7vL
8. Thomas Westfechtel, Hao-Wei Yeh, Qier Meng, Yusuke Mukuta, Tatsuya Harada. Backprop Induced Feature Weighting for Adversarial Domain Adaptation with Iterative Label Distribution Alignment. Winter Conference on Applications of Computer Vision (WACV), pp. 392-401, 2023.
9. Baoyao Yang, Hao-Wei Yeh, Tatsuya Harada, Pong C. Yuen. Model-Induced Generalization Error Bound for Information-Theoretic Representation Learning in Source-Data-Free Unsupervised Domain Adaptation. IEEE Transactions on Image Processing. pp.419-432, 2022. Digital Object Identifier: 10.1109/TIP.2021.3130530
10. Hao-Wei Yeh, Baoyao Yang, PongChi Yuen, Tatsuya Harada. SoFA: Source-data-free Feature Alignment for Unsupervised Domain Adaptation. Winter Conference on Applications of Computer Vision 2021 (WACV), pp.474-483, 2021.
11. Hao-Wei Yeh, Thomas Westfechtel, Jia-Bin Huang, Tatsuya Harada. Boosting Source-free Domain Adaptation via Confidence-based Subsets Feature Alignment. 26th International Conference on Pattern Recognition (ICPR), pp. 2857-2863, 2022. doi: 10.1109/ICPR56361.2022.9956719
12. Ryohei Shimizu, Yusuke Mukuta, Tatsuya Harada. Hyperbolic Neural Networks++. International Conference on Learning Representations (ICLR), 25 pages, online, May 4-8, 2021. https://openreview.net/forum?id=kiPsMct7vL
13. Hiromichi Kamata, Yusuke Mukuta, Tatsuya Harada. Fully Spiking Variational Autoencoder. Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI), pp.7059-7067, Feb. 22 - Mar. 1, 2022.

14. Yang Li, Tatsuya Harada. Lepard: Learning partial point cloud matching in rigid and deformable scenes. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5554-5564, 2022. (oral)
15. Yang Li, Tatsuya Harada. Non-rigid Point Cloud Registration with Neural Deformation Pyramid. Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS), pp. 27757-27768, 2022.
16. Yuki Kawana, Yusuke Mukuta, Tatsuya Harada. Neural Star Domain as Primitive Representation. Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS), pp.7875-7886, 2020.
17. Yuki Kawana, Yusuke Mukuta, Tatsuya Harada. Unsupervised Pose-aware Part Decomposition for Man-made Articulated Objects. The European Conference on Computer Vision (ECCV), pp. 558-575, 2022.
18. Atsuhiro Noguchi, Xiao Sun, Stephen Lin, Tatsuya Harada. Neural Articulated Radiance Field. International Conference on Computer Vision (ICCV), pp. 5762-5772, 2021.
19. Atsuhiro Noguchi, Umar Iqbal, Jonathan Tremblay, Tatsuya Harada, Orazio Gallo. Watch It Move: Unsupervised Discovery of 3D Joints for Re-Posing of Articulated Objects. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3677-3687, Jun. 21-24, 2022.
20. Atsuhiro Noguchi, Xiao Sun, Stephen Lin, Tatsuya Harada. Unsupervised Learning of Efficient Geometry-Aware Neural Articulated Representations. The European Conference on Computer Vision (ECCV), pp. 597-614, 2022.
21. Tianhan Xu, Tatsuya Harada. Deforming Radiance Fields with Cages. The European Conference on Computer Vision (ECCV), pp. 159-175, 2022.