

Learning with Limited Labeled Data

by Yanbin Liu

Thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy

under the supervision of Yi Yang

University of Technology Sydney
Faculty of Engineering and Information Technology
Jan 2021

Certificate of Authorship/Originality

I, Yanbin Liu declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Yanbin Liu

Signature:

Production Note:

Signature removed prior to publication.

Date: Jan 19, 2021

ABSTRACT

Learning with Limited Labeled Data

by

Yanbin Liu

The recent success of convolutional neural networks (CNNs) relies on a large amount of annotated training data. However, many research problems suffer from the scarcity of labeled data, since annotating a large number of data is time-consuming or infeasible. This dissertation focuses on learning with limited labeled data and addresses two problems: few-shot classification and object matching.

For few-shot classification, a transductive propagation network (TPN) is first proposed to deal with the low-data issue. The idea is learning to propagate labels from labeled instances to unlabeled ones by exploiting the manifold structure of the data. Then, online feature selection with imbalanced streaming data, as a special few-shot problem, is tackled by the proposed adaptive sparse confidence-weighted (ASCW) algorithm. This algorithm utilizes the confidence-weighted (CW) learning to explore the feature correlation and maintains multiple confidence-weighted learners with different costs to address the imbalanced issue.

For object matching, since the labeled matching pairs are usually scarce, finding the potential matching among unpaired objects is important. Based on this idea, two models are proposed to solve object matching with limited labeled data. First, a squared-loss mutual information (SMI) estimator is proposed to utilize a small number of paired samples and the available unpaired ones. The estimator is formulated with optimal transport and quadratic programming in an iterative way. Second, the specific object matching problem, namely semantic correspondence, can be solved in a unified optimal transport framework. The many to one matching and background matching issues are well addressed in the proposed framework.

To evaluate the effectiveness of the aforementioned algorithms with limited labeled data, extensive experiments are conducted on various benchmark datasets, ranging from the UCI machine learning repository, few-shot image classification datasets, semantic correspondence datasets, etc.

Dissertation directed by Professor Yi Yang

AAII - Australian Artificial Intelligence Institute, UTS

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Yi Yang for the continuous support of my Ph.D. study. I first met Yi in 2014 and started pursuing Ph.D. under his supervision from 2017. Without him, I would not make up my mind to start my Ph.D. journey. During my study, I learned scientific thoughts, communication skills, and research vision from him. He also encourages me to visit companies and research institutions for collaborations, which helped me a lot during my study.

I would like to thank Linchao Zhu, who is a senior and also my friend and co-supervisor. He is enthusiastic to share ideas with me and help me revise papers. His passion and perseverance for research inspire me.

I would like to thank Juho Lee, Saehoon Kim, Sung Ju Hwang, Eunho Yang for their support during my intern at AITRICS.

I would like to thank Makoto Yamada for his help during my visiting at RIKEN AIP. His expertise, patience, and communication skills inspired me.

I would also like to thank all the members in the ReLER lab. During the study, I was fortunate to take part in many illuminating discussions with these smart people. Particularly, I would like to thank Yan Yan who helped me with my first submission.

Lastly, this thesis is dedicated to my parents for their selfless support and love.

Yanbin Liu
Sydney, Australia, 2021.

List of Publications

Conference Papers

- C-1. **Liu, Y.**, Lee, J., Park, M., Kim, S., Yang, E., Hwang, S. and Yang, Y., Learning to Propagate Labels: Transductive Propagation Network for Few-shot Learning. International Conference on Learning Representations (ICLR), 2019.
- C-2. **Liu, Y.**, Yan, Y., Chen, L., Han, Y. and Yang, Y., Adaptive Sparse Confidence-Weighted Learning for Online Feature Selection. AAAI Conference on Artificial Intelligence (AAAI), 2019.
- C-3. **Liu, Y.**, Zhu, L., Yamada, M. and Yang, Y., Semantic Correspondence as an Optimal Transport Problem. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

Preprint Papers

- P-1. **Liu, Y.**, Yamada, M., Tsai, Y.H.H., Le, T., Salakhutdinov, R. and Yang, Y., LSMI-Sinkhorn: Semi-supervised Squared-Loss Mutual Information Estimation with Optimal Transport. arXiv preprint arXiv:1909.02373, 2019.

Contents

Certificate	ii
Abstract	iii
Acknowledgments	v
List of Publications	vi
List of Figures	xi
1 Introduction	1
1.1 Background and Research Problem	1
1.2 Thesis Organization	4
2 Literature Survey	6
2.1 Few-shot Learning	6
2.2 Online Learning	8
2.3 Mutual Information Estimation	10
2.4 Semantic Correspondence	12
3 Learning to Propagate Labels: Transductive Propagation	
Network for Few-shot Learning	15
3.1 Introduction	15
3.2 Main approach	18
3.2.1 Problem definition	18
3.2.2 Transductive Propagation Network (TPN)	20

3.3	Experiments	24
3.3.1	Datasets	24
3.3.2	Experimental Setup	25
3.3.3	Few-shot Learning Results	25
3.3.4	Comparison with semi-supervised few-shot learning	29
3.4	Conclusion	31
4	Adaptive Sparse Confidence-Weighted Learning for On-	
	line Feature Selection	32
4.1	Introduction	32
4.2	Imbalanced Online-Batch CW Learning	34
4.2.1	Notations	34
4.2.2	Cost-Sensitive Learning for Imbalanced Data	35
4.2.3	Online-Batch CW Learning	35
4.3	Sparse CW for Feature Selection	36
4.3.1	Feature Selection by Sparsity Index η	36
4.3.2	Optimization	38
4.3.3	Proximal Dual Coordinate Ascent for Subproblem (4.15)	39
4.3.4	Discussions	40
4.4	Multiple Cost-Sensitive Learning	42
4.4.1	Theoretical Analysis in F-measure	43
4.5	Experiments	44
4.5.1	Experimental Testbed	44
4.5.2	Comparison Algorithms	45
4.5.3	Experimental Results	46

4.5.4	Optimal Cost Vector	49
4.6	Conclusion	50
5	LSMI-Sinkhorn: Semi-supervised Squared-Loss Mutual Information Estimation with Optimal Transport	51
5.1	Introduction	51
5.2	Problem Formulation	53
5.3	Proposed Method	54
5.3.1	Least-Squares Mutual Information with Sinkhorn (LSMI-Sinkhorn)	54
5.3.2	Optimization	56
5.3.3	Discussion	59
5.4	Experiments	60
5.4.1	Setup	61
5.4.2	Convergence and Runtime	62
5.4.3	SMI Estimation	63
5.4.4	Deep Image Matching	65
5.4.5	Photo Album Summarization	67
5.5	Conclusion	68
6	Semantic Correspondence as an Optimal Transport Problem	69
6.1	Introduction	69
6.2	Proposed Algorithm	72
6.2.1	Preliminary	72
6.2.2	Semantic correspondence as an OT problem	73

6.2.3	Pre- and post-processing	78
6.3	Experiments	79
6.3.1	Benchmarks and evaluation metric	80
6.3.2	Implementation details	81
6.3.3	Evaluation results on SPair-71k	81
6.3.4	TSS, PF-PASCAL, and PF-WILLOW	85
6.4	Conclusion	87
7	Conclusion and Future Work	90
	Bibliography	92

List of Figures

3.1	A conceptual illustration of our transductive meta-learning framework, where lines between nodes represent graph connections and their colors represent the potential direction of label propagation. The neighborhood graph is episodic-wisely trained for transductive inference.	17
3.2	The overall framework of our algorithm in which the manifold structure of the entire query set helps to learn better decision boundary. The proposed algorithm is composed of four components: feature embedding, graph construction, label propagation, and loss generation.	19
3.3	Detailed architecture of the graph construction module, in which the length-scale parameter is example-wisely determined.	21
3.4	5-way performance with various training/test shots.	29
4.1	Online performance with imbalance ration 1:10 for different performance measures	49
5.1	Convergence curves of the loss and SMI values.	61
5.2	Runtime of LSMI-Sinkhorn and Gromov-Wasserstein. A base-10 log scale is used for the Y axis.	61
5.3	SMI estimation on synthetic data ($n_x = n_y = 500$).	62
5.4	Visualization of the matrix $\mathbf{\Pi}$	63

5.5	SMI estimation on synthetic data ($n_x = 1000, n_y = 500$).	65
5.6	SMI estimation on UCI datasets.	66
5.7	Deep image matching.	66
5.8	Photo album summarization on Flickr dataset. For (a) and (b), we fixed the corners with <i>green</i> , <i>orange</i> , <i>black</i> (triangle), and <i>blue</i> (rectangle) images. For (c), we fixed the center of each character with different images.	68
5.9	Photo album summarization on CIFAR10 dataset. For (a) and (b), we fixed the corners with <i>automobile</i> , <i>airplane</i> , <i>horse</i> (triangle), and <i>dog</i> (rectangle) images. For (c), we fixed the center of each character with different images.	68
6.1	We solve two problems caused by current approaches, such as HPF [1]. (a) Many pixels in a source image are assigned to one target pixel. (b) Some object pixels are assigned to the background pixels. Note that the two different results are only part of the whole results which reflect the many to one matching and background matching.	70
6.2	The proposed framework. A pair of images are input to the pre-trained CNN to get the multi-layer feature maps ($\mathbf{f}_s, \mathbf{f}_t$) and the class activation maps in a single forward pass. \mathbf{f}_s and \mathbf{f}_t are used to compute a cost matrix \mathbf{M} representing the matching difference. Activation maps undergo a staircase function and are then normalized as the empirical probability distributions μ_s and μ_t . We deal with semantic correspondence by solving optimal transport problem to get the optimal match T^* , which is further post-processed to ensure geometric consistency.	73
6.3	Left: Many source pixels are assigned to one target pixel. Right: The row sum and column sum constraints of the matching matrix suppress the many to one matching.	75

6.4	Qualitative results on SPari-71k. The source images are warped to align with target images using correspondences. For HPF [1], NC-net [2], and our method, we first use the source keypoints and the predicted target keypoints to estimate the thin-plate spline (TPS) parameters, then apply TPS transformation on the source image. For A2Net [1] and WeakAlign [3], they are global alignment methods that directly predict the global transformation parameters from the CNN models. We show image pairs with large intra-class, scale, and view-point changes. Our method performs better in complex conditions due to our global matching and background suppressing strategies.	86
-----	---	----