Visual Analysis with Limited Supervision

Wenhe Liu

Supervisor: Professor Yi Yang

Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney

A thesis submitted for the degree of
Doctor of Philosophy

November 2018

I would like to dedicate this thesis to my loving parents and my beloved grandparents.

Declaration

I, Wenhe Liu declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Centre for Artificial Intelligence, Faculty of Engineering and Information Technology at the University of Technology Sydney. This thesis is wholly my own work unless otherwise reference 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.

Wenhe Liu
November 2018

Acknowledgements

This thesis cannot be completed without important people in my study and life. And I would like to acknowledge my supervisor, Dr. Yi Yang, who led me into the realm of machine learning and visual analysis. His passion for exploring new ideas and rigorous attention to detail affected me profoundly. Dr. Xiaojun Chang is my associate supervisor. I am always impressed by the way he thinks about every research problem. His self-motivation, hard-working spirit and rigorous research attitude have set an excellent example for me. Dr. Alexander G. Hauptmann was my supervisor when I visited Carnegie Mellon University. I have appreciated his guidance immensely. I would like to thank many friends at University of Technology Sydney and Carnegie Mellon University. Yan Yan, Guoliang Kang have helped me so much when I was in the group. Their sincerity and kind heart impressed me to give warm care to every other person. I would also like to thank my friends at University of Adelaide. Dong Gong, a good friend who has taught me a lot of mathematical theory and skills and support me all the time. Lastly, I would like to thank my parents for their continuous support for my academic pursuit. I wish them health and happiness forever.

Abstract

Visual analysis is an attractive research topic in the field of computer vision. In the visual analysis, there are two critical directions, visual retrieval and visual classification. In recent years, visual retrieval has been investigated and developed in many real-world applications, for instance, in person re-identification. On the other hand, visual classification is also widely studied, such as in image classification. Typical visual analysis methods are supervised learning algorithms. In such algorithms, extensive labeled data is demanded for training supervised models in order to achieve acceptable performance. However, it is difficult to collect and generate annotated data in the real world due to the limited resources, such as human labor for annotation. Therefore, it is urgent to develop methods to complete the visual analysis mission with limited supervision.

In this thesis, we propose to address the visual analysis problem with limited supervision. Specifically, we treat limited supervision problem in three scenarios according to the amount of labeled data. In the first scenario, no labeled data are provided and only limited human labor for annotation is available; In the second scenario, scarce labeled data and abundant unlabeled data are accessible. In the third scenario, only few instances in the target dataset are labeled and there are multiple sources of labeled data from different domains.

In Chapter 2 and Chapter 3, we discuss the first scenario, when no labeled data are provided, and only limited human labor for annotation is available. We propose to solve the problem via active learning. Unlike conventional active learning, which usually starts with a set of labeled data as the reference, in this thesis, we adopt the active learning algorithm with no pre-given labeled data. We refer these algorithms as the Early Active Learning. In this thesis, first, we attempt to select the most contributive instances for annotation and later being utilized for training supervised models. We demonstrate that even by annotating a few selected instances, the proposed method can achieve comparable performance in the visual retrieval. Second, we further extend the instance based active learning to pair-based early active learning. Other than

select instances for annotation, the pair-based early active learning selects the most informative pairs for annotation, which is essential in the visual retrieval.

In Chapter 4, in the second scenario, we address the visual retrieval problem when there are scarce labeled data and abundant unlabeled data. In this thesis, we propose to utilize both the labeled and the unlabeled data in a semi-supervised attribute learning schema. The proposed method could jointly learn the latent attributes with appropriate dimensions and estimate the pairwise probability of the data simultaneously.

In Chapter 5 and Chapter 6, in the third scenario, we focus on visual classification with few or no labels, but there are pre-known labeled data from other domains. To improve the performance in the target domain, we adopt transfer learning algorithms to transfer helpful knowledge from the pre-known (source) domain with labeled data. First, in Chapter 5, the few-shot visual classification problem is considered. We have access to multiple source datasets with well-labeled data but can only access a limited set of labeled data in the target dataset. An Analogical Transfer Learning schema is proposed for this problem. It attempts to transfer the knowledge from the source domains to enhance the performance of the target domain models. In the algorithm, an analogy-revision schema is designed to select only the helpful source instances to enhance the target domain models. Second, in Chapter 6, we challenge a more difficult problem when there is no labeled data in the target domain in the visual retrieval problem. A Domain-aware Unsupervised Cross-dataset Transfer Learning algorithm is proposed to address this problem. The importance of universal and domain-unique appearances are valued simultaneously and jointly contribute to the representation learning. It manages to leverage the common and domain-unique representations across datasets in the unsupervised visual retrieval.

Contents

List of Figures **xv**

List of Tables **xix**

Publications **xxi**

1 Introduction **1**
1.1 Background 1
1.1.1 Video Analysis 1
1.1.2 Learning with Limited Supervision 2
1.2 Related Works 3
1.2.1 Active Learning 3
1.2.2 Semi-supervised Attribute Learning in Re-id 4
1.2.3 Transfer Learning and Few-shot Learning 5
1.3 Contributions 6

2 Early Active Learning with Pairwise Constraint **9**
2.1 Background 9
2.2 The Proposed Method 12
2.2.1 Early Active Learning 12
2.2.2 Early Active Learning with Pairwise Constraint 13
2.2.3 Kernelization 15
2.2.4 Optimization 16
2.3 Convergence Analysis 16
2.4 Experimental Study 18
2.5 Summary 24

3 Pair-based Early Active Learning **27**
3.1 Background 27

3.2 The Proposed Framework 28
3.2.1 Pairwise Uncertainty and Diversity 28
3.2.2 Pair-based Early Active Learning 29
3.2.3 Optimization 30
3.3 Experimental Study 32
3.3.1 Experimental Settings 32
3.3.2 Results Analysis 36
3.4 Summary 40

4 Semi-supervised Bayesian Attribute Learning 41
4.1 Background 41
4.2 The Proposed Method 43
4.2.1 Formulation 43
4.2.2 Optimization 46
4.3 Experiments 48
4.3.1 Datasets 48
4.3.2 Evaluation Metrics and Preprocessing 49
4.3.3 Supervised Person Re-ID 49
4.3.4 Semi-supervised Person Re-ID 51
4.3.5 Attributes Learning in Re-ID 52
4.3.6 The Influence of Latent Factor Dimensionality 53
4.4 Summary 53

5 Analogical Transfer learning 55
5.1 Background 55
5.2 Problem Statement 58
5.3 The Proposed Method 60
5.3.1 Single-Source ATL 60
5.3.2 Multi-Source ATL 64
5.3.3 Optimization 64
5.4 Experiments 65
5.4.1 Baseline Algorithms 66
5.4.2 Datasets and Setting 67
5.4.3 Performance Evaluation 69
5.5 Summary 74

6 Domain-aware Unsupervised Cross-dataset Transfer Learning 75
6.1 Background 75
6.2 The Proposed Framework 77
6.2.1 Formulation 77
6.2.2 Optimization 82
6.3 Experiments 85
6.3.1 Experimental Settings 85
6.3.2 Experimental Analysis 86
6.4 Summary 89

7 Conclusion and Future Work 91
7.1 Conclusion 91
7.2 Future Work 92

Bibliography 93

List of Figures

2.1 Procedures of QUIRE [40] (*upper*) and our Early Active Learning with Pairwise Constraint (EALPC) (*lower*). In QUIRE, pre-labeled samples $\mathbf{X}_l$ are used for the uncertainty evaluation on the unlabeled samples $\mathbf{X}_u$. Then, it selects a subset samples $\mathbf{V} \subset \mathbf{X}_u$ for labeling. At last, both $\mathbf{X}_u$ and $\mathbf{V}$ along with their labels are used for supervised learning. In EALPC, unlabeled data $\mathbf{X}$ is analyzed without pre-labeled data. Meanwhile, pairwise constraint $\mathbf{\Psi}$ is introduced to enhance the performance of early active learning for Re-ID . More details are in Section. 2.2. . . 11

2.2 CMC Performance Comparison of Active Learning algorithms. XQDA is chosen as the Re-ID algorithm. The percentage of selected samples is set to 10% of all samples. 22

2.3 Rank One Matching Accuracy(%) w.r.t. Number of Selected Instances. We use XQDA as the Re-ID algorithm and train it with varying numbers of samples selected by the active learning methods. 23

2.4 Convergence Analysis of EALPC on Benchmark Datasets. The parameters are set as $\alpha = 0.1$ and $\beta = 1$. The percentage of selected samples is 20%. 25

3.1 Rank One Matching Accuracy(%) w.r.t. Number of Selected Samples. XQDA is chosen as the Re-ID algorithm and trained with samples determined by the active learning methods with different amounts. . . 37

3.2 Convergence of the Proposed Algorithm on CAVIAR Dataset 38

3.3 Diversity Analysis of the Proposed Algorithm.The leverage parameter is $\alpha = 0.1$ and 20% of the samples are selected for annotation. 39

4.1 The plate notation of our model. 45

4.2 Influence of K_+ w.r.t. CMC Rank One accuracy. The automatically leaned attribute numbers are $K_+^* = 1740$ for VIPeR dataset and $K_+^* = 1588$ for PRID dataset(marked with asterisk symbol (*)). 53

5.1 An example of our algorithm: 'Learning Guava from other fruits'. First, in the **Revision** stage, when a new (target) genus of fruit 'guava' comes, the learner selects the source instances related to the target instances (i.e., fruits related to guava) and learn the revised source hypotheses. Second, in the **Transfer** stage, an *Analogical Hypothesis* is concluded based on the observations of guava and the revised source hypotheses. Detailed discussion is in Chapter 2.1. 57

5.2 Parameter Sensitivity of self paced regularizer. Soft weight drops when loss values increasing (solid lines). When β becomes larger, the function becomes more sensitive, i.e. weights drop more dramatically even loss values are small. Hard weight (dash lines) do not sensitive while loss increases. 62

5.3 Performances on synthetic data. ***Left***: In the target domain. Given only a few target samples (magenta points), the standard SVM classifier fails to classify the two half circles distribution (the decision boundary is marked as solid lines). ***Middle***: In the source domain. Given sufficient source samples (yellow and green points), the standard SVM classifier can classify two half circles distribution (The decision boundary is marked as dash line). ***Right***: If use entire source samples as in the Middle figure, it will fall to divide target examples that near to each other as shown in the center of graph. (The decision boundary is marked as dash line). Given selected source samples (dark blue points) and few target samples (magenta points), in our algorithm, the classifier can classify two half circles by a decision boundary (The decision boundary is marked as solid line). 70

5.4 Performance on Image Data (1) Baseline algorithms SVM is not affected much by the number of target training samples; (2) Overall, all algorithms perform better in Inner-family test than Cross-family test; (3) Algorithms that use target domain knowledge generally perform better when given more target training samples; (4) Our algorithm outperform others early when given over 10 target training samples. 72

5.5 Parameter Sensitivity of learning rate γ on Text Data. Our algorithm ATL (in red lines) performs consistently better than baseline SVM (green lines) at about 10% in Accuracy. Global optimal γ differs on different data sets. 73

6.1 Examples of common and unique appearances on four datasets. Better being watched in color. **Common appearances** shared by four datasets: (a) wearing dark coat and pants; (b) wearing red upper cloth; (c) walking forward; (d) wearing dark cloth and light color pants. The images in (a) to (d) from different datasets and they are belong to the different persons. **Unique appearances**: (e)VIPeR: carrying backpack; (f)PRID: carrying bags in hand. 76

6.2 Overview of our algorithm for person Re-ID . Source datasets are datasets with labels, target dataset is a dataset with no label. A dictionary is learned with all datasets. With the dictionary, all images are encoded in one low dimensional space. The linked samples in the source dataset are labeled as the same person. 78

6.3 Performances of UDML and Ours algorithm on PRID dataset. Probe images are provided in the left column. Top-5 candidate images are sorted in descent order according to their score. The ground truth images are marked with red bounding boxes. (a) Woman with pink handbag and dark coat. (b) Man with dark coat and white pants (and without handbag). (c) Woman with white handbag and dark coat. 84

6.4 Re-id performance for ablation study. 88

List of Tables

2.1 Rank One Matching Accuracy(%) on four benchmarks. Percentage of selected instances for labeling is 20% of all samples. Each column is an active learning algorithm and each row is a Re-ID algorithm. The best result of each Re-ID algorithm is marked in bold numbers. The best result of the algorithms overall is marked with an asterisk(∗). 21

3.1 Rank One Matching Accuracy(%) on the benchmark datasets. 20% of the samples are selected for labeling. The rows are Re-ID algorithms. The columns are active learning algorithms. For each Re-ID methods, the best result w.r.t active learning algorithms is marked in bold. The best result in all the Re-ID methods is marked with an asterisk(∗). . . 33

4.1 Supervised Re-ID result of Rank One Matching Accuracy(%) on two benchmarks. Best result of each Re-ID algorithm is marked as bold numbers. 50

4.2 Semi-supervised Re-ID results in terms of rank-1 matching accuracy(%) for VIPeR and PRID datasets. The best result from each Re-ID algorithm is shown in bold. 51

4.3 Attribute learning results on DukeMTMC-reID dataset. (1) Learning with predefined attributes (2) Learning with no pre-defined attributes. The best result for each category is in bold. The overall best results are marked with an asterisk (*) 52

5.1 List of Important Notations and Descriptions 59

5.2 Data Description of Newsgroup 68

5.3 Performance (ACC ± standard deviation) (%) of single-source test . . 71

5.4 Performance (ACC± standard deviation) (%) of multi-source test . . . 72

6.1 Rank One Matching Accuracy(%) on Unsupervised Re-ID. (a) Single-task methods. (b) Multi-task methods. 87
6.2 Inter-dataset Similarity Index with Target Dataset PRID ($\times 10^2$) 89

Publications

From this work:

- Chapter 2: **Wenhe Liu**,Xiaojun Chang, Ling Chen, Yi Yang: Early Active Learning with Pairwise Constraint for Person Re-identification. *ECML/PKDD*, 2017
- Chapter 3: **Wenhe Liu**,Xiaojun Chang, Yi Yang:Pair-based Early Active Learning for Person Re-identification. (submitted 2018)
- Chapter 4: **Wenhe Liu**, Xiaojun Chang, Ling Chen, Yi Yang: Semi-Supervised Bayesian Attribute Learning for Person Re-Identification. *AAAI*, 2018.
- Chapter 5: **Wenhe Liu**, Xiaojun Chang , Yan Yan , Yi Yang , Alexander G. Hauptmann Few-Shot Text and Image Classification via Analogical Transfer Learning. *ACM Transactions on Intelligent Systems and Technology (TIST)*, Volume 9 Issue 6, October 2018.
- Chapter 6: **Wenhe Liu**,Xiaojun Chang, Yi Yang:Domain-aware Unsupervised Cross-dataset Transfer Learning for Person Re-identification. (submitted 2018)

Referred:

- **Wenhe Liu**, Dong Gong, Mingkui Tan, Javen Qinfeng Shi, Yi Yang and Alexander G. Hauptmann: Learning Distilled Graph for Large-scale Social Network Data Clustering. *Transactions on Knowledge and Data Engineering (TKDE).* (submitted 2017)
- **Wenhe Liu**, Chenqiang Gao, Xiaojun Chang, Qun Wu: Unified discriminating feature analysis for visual category recognition. *J. Visual Communication and Image Representation*,2016.